Università degli Studi di Salerno

Dipartimento di Informatica

Dottorato di Ricerca in Informatica
XXXIV Ciclo

TESI DI DOTTORATO / PH.D. THESIS

The Influence of User Interface Design on User Experience - Focus on Accessibility and Situation Awareness

MARIANNA DI GREGORIO

SUPERVISOR: **PROF. GIULIANA VITIELLO**

PH.D. PROGRAM DIRECTOR: **PROF. ANDREA DE LUCIA**

A.A 2020/2021

Thesis Reviewers
Prof. Paolo Buono, University of Bari
Prof. Teresa Onorati, University Carlos III of Madrid

Defense Committee
Prof. Sandro Bimonte, National Research Institute for Agriculture - INRAE
Prof. Paolo Buono, University of Bari
Prof. Vincenzo Deufemia, University of Salerno

Date of defense
May 23rd, 2022

Marianna Di Gregorio
Department of Computer Science,
University of Salerno
madigregorio@unisa.it

Abstract

"User eXperience focuses on interaction between a person and something that has a user interface."[277]

User expectations, motivations, and feelings when using a system have prompted the need to investigate beyond traditional usability issues by evaluating and designing for the user experience. User eXperience (UX), therefore, has emerged from the recognition that usability alone does not take into account the more subjective emotional responses experienced when interacting with a system.
Although the term User eXperience has been widely accepted in the area of Human-Computer Interaction (HCI), its relationship to certain aspects of the User Interface (UI), such as accessibility and situational awareness, still remains unclear.

The main purpose of this research thesis was to determine how much influence the UI has on the UX when it is conditioned by accessibility and situational awareness.

First, the main goal during the UI design phase was to create a good experience for all users, regardless of skills. In most cases, people design a UI that in no way reflects the need for accessibility. However, many people live with disabilities and such projects make it difficult to use applications and systems and contribute to discrimination due to lack of inclusiveness. To achieve this, the focus has been on three main groups who live with disabilities and have difficulty accessing UIs due to inadequate UX design. These groups include people with hearing impairments, people with visual impairments and cognitive impairments (focusing on learning and behavioral disorders). Some of the solutions offered ensure that these groups, which are a large population, have access to UIs include proposals to help them in their daily life and, above all, to support them in their independence. In addition, the research found that most developers do not include factors such as accessibility in the design or development phase of UIs. This lack, especially with regard to mobile apps, is mainly caused by the lack of sufficient tools for the design and the lack of awareness of the accessibility guidelines by the developers. Making mobile app UIs highly inaccessible to a large population, including people with disabilities. In this case, the focus was on Android applications by examining the accessibility guidelines built into them.

The second focus was geared towards how UX is affected due to poorly designed UIs, making it difficult for users to maintain high SA. By focusing on Ground Control Stations for monitoring drone fleets, the research helped identify discriminating factors for SA levels. Most designers do not take SA as a factor in UX when designing UIs. UIs continue to grow in complexity and are becoming difficult for most operators due to too many features, some of them automated. In this regard, it was crucial to ensure that such UI designs are easy for operators to use to reduce the number of incidents and increase operator efficiency, thus improving UX. These changes are also proposed to be implemented in the UIs of manufacturing and industrial plants. Currently most of the factory managers use the Andon system to monitor the production process. In this case, guidelines and a UI solution have been proposed to help keep SA high.

The overall results of this research have provided some valuable insights and helped create a set of user experience heuristic practices that can be used to inform both future research and design practice.

Acknowledgements

"Spesso gli arcobaleni più luminosi seguono i temporali più bui"
A Nonna Betta
e
India

Premetto che non sono molto brava con le parole, ma proverò a scrivere brevemente qualcosa per ognuna di quelle persone che ho incontrato lungo il mio percorso.

Innanzitutto, vorrei ringraziare il mio supervisore, la Prof. Giuliana Vitiello. Non ho mai incontrato una persona così buona, gentile e paziente come lei. Ci tenevo ad esprimere la mia riconoscenza per avermi trasmesso parte della sua esperienza, e per avermi guidato nel mio percorso di ricerca. È stata sempre di supporto, talvolta come una madre, insegnandomi molte cose non solo nel campo della ricerca. Grazie per tutto ciò che mi ha insegnato in questi anni, ma soprattutto per come me lo ha insegnato. Spero che sia orgogliosa del mio percorso, grazie di cuore.

Un affettuoso ringraziamento va anche alla Prof. Monica Sebillo per la sua disponibilità. In molte occasioni è stata un punto di riferimento per me.

Grazie ai miei compagni di "laboratorio". Grazie di cuore a Marco per l'aiuto fondamentale offertomi, sia da un punto di vista scientifico che morale. Dimostrandosi sempre una brava persona, una delle migliori che ho incontrato lungo questo percorso.
Grazie a Pietro per questi tre anni, ma soprattutto grazie per avermi regalato sempre un sorriso.

Non posso dimenticare tutti i miei amici/colleghi che mi hanno accompagnato in questo percorso, ai quali va un grandissimo ringraziamento.

In particolare, un grazie di cuore e sincero è rivolto a Fabiano con cui ho iniziato questo percorso. Con lui ho condiviso tanti momenti in questi tre anni e si è dimostrato sempre una persona cara, vera e sincera. Ne approfitto per augurargli il meglio dalla vita, perché lo merita davvero. Spero che saremo sempre buoni amici.
Grazie a Gemma, la mia Amica. Grazie perché da quando l'ho conosciuta è diventata una persona su cui poter contare in qualsiasi occasione. Grazie perché in qualsiasi momento io abbia avuto bisogno di lei ci è sempre stata, anche se lontane chilometri. Grazie per aver condiviso con me momenti indimenticabili, dall'università al ktv bar. So che sarà così per sempre.
Grazie a Fabio per la sua amicizia, e per avermi aiutata sempre con preziosi consigli. La sua passione per la ricerca è stato un esempio da seguire per me.

Sentiti ringraziamenti sono per i miei genitori Pino e Lori che mi hanno permesso di raggiungere questo traguardo dandomi sempre opportuni consigli e non facendomi mai mancare il loro affetto e supporto. Sono la mia vita.

Un grande e particolare grazie va a mio fratello Angelo con cui ho sempre condiviso ogni aspetto della mia vita. Grazie alla mia súper-amiga Alba a cui voglio un bene che non si può spiegare. Grazie perché voi mi supportate in tutte le scelte, sempre. Inoltre, grazie per avermi regalato una delle gioie più grandi della mia vita. La nascita di India stata davvero come un arcobaleno dopo la pioggia.

Infine, ma non per importanza, vorrei ringraziare Nello. Grazie per tutta la gioia che mi dai, per le risate, i sorrisi, le emozioni, la comprensione, l'aiuto, la fiducia, la voglia di vivere che mi trasmetti, per la tua forza, la tua determinazione, il tuo amore, per come sei, perché ci sei. Grazie perché credi in me. Spero che questo sia solo uno dei tanti magnifici momenti che condivideremo insieme.
Ma soprattutto, grazie per avermi dato una seconda famiglia. Grazie a loro per l'affetto sincero e la stima nei miei confronti. Grazie a Mariafrancesca che è ormai per me complice e spalla.

Publications

Part of the ideas, results, images and data exposed in this thesis have been previously published, or are under review, in the following conferences and journals.

Published in Proceedings of International Conferences

[C1] Luongo Salvatore, Di Gregorio Marianna, Vitiello Giuliana, Vozella Angela.
Human Machine Interface Issues for Drone Fleet Management.
In Proceedings of the International Conference on Human Systems Engineering and Design: Future Trends and Applications (pp. 791-796). 2018. Springer, Cham.
Chapter 7

[C2] Vitiello Giuliana, Sebillo Monica, Fornaro Luigi, Di Gregorio Marianna, Cirillo Stefano, De Rosa Mattia, Fuccella Vittorio, Costagliola Gennaro.
Do you like my outfit?: Cromnia, a mobile assistant for blind users.
In Proceedings of the 4th EAI international conference on smart objects and technologies for social good (pp. 249-254). 2018. ACM.
Chapter 4

[C3] Di Gregorio Marianna, Sebillo Monica, Vitiello Giuliana, Pizza Antonio, Vitale Federico.
Prosign everywhere-addressing communication empowerment goals for deaf people.
In Proceedings of the 5th EAI International Conference on Smart Objects and Technologies for Social Good (pp. 207-212). 2019.

ACM.
Chapter 3

[C4] Battistoni Pietro, Di Gregorio Marianna, Sebillo Monica, Vitiello Giuliana.
AI at the edge for sign language learning support.
In ProceedingsIn of the International Conference on Humanized Computing and Communication (pp. 16-23). 2019. IEEE.
Chapter 3

[C5] Battistoni Pietro, Sebillo Monica, Di Gregorio Marianna, Vitiello Giuliana, Romano Marco.
ProSign+ a cloud-based platform supporting inclusiveness in public communication.
In Proceedings of the IEEE 17th Annual Consumer Communications & Networking Conference (pp. 1-5). 2020. IEEE.
Chapter 3

[C6] Battistoni Pietro, Sebillo Monica, Di Gregorio Marianna, Vitiello Giuliana, Romano Marco, Giandomenico Solimando.
Sign Language Interactive Learning-Measuring the User Engagement.
In Proceedings of the International Conference on Human-Computer Interaction (pp. 3-12). 2020. Springer, Cham.

[C7] Di Gregorio Marianna, Di Nucci Dario, Palomba Fabio, Vitiello Giuliana.
The Making of Accessible Android Applications: An Empirical Study on the State of the Practice.
In Proceedings of the IEEE International Conference on Software Maintenance and Evolution (ICSME) (pp. 857-861). 2020. IEEE.
Chapter 6

[C8] Di Gregorio Marianna, Nota Giancarlo, Romano Marco, Sebillo Monica, Vitiello Giuliana.
Designing usable interfaces for the Industry 4.0.

In Proceedings of the International Conference on Advanced Visual Interfaces (pp. 1-9). 2020. ACM.
Chapter 8

[C9] Battistoni Pietro, Di Gregorio Marianna, Giordano Domenico, Sebillo Monica, Tortora Genoveffa, Vitiello Giuliana.
Wearable Interfaces and Advanced Sensors to Enhance Firefighters Safety in Forest Fires.
In Proceedings of the International Conference on Advanced Visual Interfaces (pp. 1-3). ACM.

[C10] Amato Federica, Di Gregorio Marianna, Monaco Clara, Sebillo Monica, Tortora Genoveffa, Vitiello Giuliana.
The Therapeutic Use of Humanoid Robots for Behavioral Disorders.
In Proceedings of the International Conference on Advanced Visual Interfaces (pp. 1-3). 2020. ACM.

[C11] Amato Federica, Di Gregorio Marianna, Monaco Clara, Sebillo Monica, Tortora Genoveffa, Vitiello Giuliana.
The Therapeutic Use of Humanoid Robots for ADHD.
In Proceedings of the 7th ACM Celebration of Women in Computing. 2020. ACM.

[C12] Amato Federica, Di Gregorio Marianna, Monaco Clara, Sebillo Monica, Tortora Genoveffa, Vitiello Giuliana.
Socially Assistive Robotics combined with Artificial Intelligence for ADHD.
In Proceedings of the IEEE 18th Annual Consumer Communications \& Networking Conference (pp. 1-6). 2021. IEEE.
Chapter 5

[C13] Di Gregorio Marianna, Sebillo Monica, Tortora Genoveffa, Vitiello Giuliana.
Pandemics throughout history and geography–an interactive environment to educate the new generation of students.

In Proceedings of the 14th Biannual Conference of the Italian SIGCHI Chapter (pp. 1-7). 2021. ACM.

[C14] Di Gregorio Marianna, Sebillo Monica, Vitiello Giuliana.
Geo-education as a Valuable Support to Children with Learning Difficulties.
In Proceedings of the International Conference on Human-Computer Interaction (pp. 62-71). 2021. Springer, Cham.
Chapter 5 – Together with work [J1]

[C15] Battistoni Pietro, Di Gregorio Marianna, Romano Marco, Sebillo Monica, Vitiello Giuliana.
TactCube: Designing Mobile Interactions with Ambient Intelligence.
In Proceedings of the IFIP Conference on Human-Computer Interaction (pp. 599-609). 2021. Springer, Cham.

[C16] Di Gregorio Marianna.
Accessible applications - Study and design of user interfaces to support users with disabilities.
In Proceedings of the 20th ACM International Conference on Multimodal Interaction. 2021.
Doctoral Consortium.

Published in International Journals

[J1] Sebillo Monica, Vitiello Giuliana, Di Gregorio Marianna.
Maps4Learning: Enacting Geo-Education to Enhance Student Achievement.
JIEEE Access 8 (2020): 87633-87646.
Chapter 5

[J2] Di Gregorio Marianna, Romano Marco, Sebillo Monica, Vitiello Giuliana, Vozella Angela.
Improving Human Ground Control Performance in Unmanned Aerial Systems.
Future Internet 13.8 (2021): 188.
Chapter 7

Submitted to International Conferences

[SC1] Di Gregorio Marianna, Palomba Fabio, Vitiello Giuliana.
Did you spot it before falling into the trap? Dark Patterns in Web User Interfaces - An analysis of Eye Gaze Interaction and Dark Patterns Perception in Web User Interfaces.

Submitted to International Journals

[SJ1] Di Gregorio Marianna, Di Nucci Dario, Palomba Fabio, Vitiello Giuliana.
The Making of Accessible Android Applications: An Empirical Study on the State of the Practice.
Empirical Software Engineering, Springer, 2021.
Under Major Revision
Chapter 6 - Extension of the work [C7] - This paper will be discussed in this thesis.

Contents

Chapter 1

Introduction

This first chapter provides an overview of the context of the studied subject matter and the motivation behind it, the identified problems in connection to the topic, and the purpose and scope of the research. Additionally, research questions are formulated based on the problem discussion. The introduction concludes with a brief description of the essential contributions of the thesis.

1.1 Context and Motivation

User Interface (UI) is the link on which humans interact with the computer, a computer system, or the machine. It is the layer on which the computer users manage to complete their tasks. Therefore, the UI is part of the system acting as an intermediary between the computer and the user and facilitates the effective interaction between the computer and the users [264].

The purpose of the UI is for users to effectively control the computer or machine interacting with them. User eXperience (UX) is the user's interaction and the experience with a product or a system. According to ISO 9241-210, the UX sits "*the person's perception and the response which results in the intended use of a product or services*". The UX contains all the element which may influence how a person feels about a system. This includes emotions, behaviours, preferences, perceptions, and the user's physical and psychological response to the system [265]. It is determined by the outcomes that the user experienced before, during, and after using a system.

The complexity of the computer system determines how easy the users may find it to utilize them. Complex systems are very difficult to use. Therefore, the UX may be low when the complexity of the computer system is high and vice versa when the complexity of the system goes down [266]. To cope with the complexity of the system, appropriate UI development is needed. Furthermore, the development of a UI is difficult to process without understanding the characteristics of the users and the activities they perform with the system [267]. The good nature of the UX has to be contextual and subjective [268].

This research thesis aims to determine how much influence the UI has on the UX. In particular, it has already been found that the use of the UI contributes to the feeling of a positive UX. As a result, studies report strong correlations between the UI and aspects of the UX such as usability and credibility, but does this also apply to accessibility? Some believe that usability design prevents accessibility, while most designers perceive the accessibility initiative to be restrictive from a design standpoint. Furthermore, it is often thought that accessibility is only related to people with disabilities, but in reality, it affects everyone. These misconceptions have slowed the progress of inclusive UI design. This thesis deals with this problem in the first part.

Also, Situational Awareness (SA) is often ignored during UI design for some real-world work contexts. This lack of attention is misleading because we ignore the fact that 88% of human errors are due to Situation Awareness problems [269]. SA is one of the cornerstones of human-centered systems. Currently applied in many sectors, such as aeronautics [270, 271], mining [272], oil/gas [273], rail [274], and others [275], it involves the perception of elements in the environment, the understanding of those elements and the situation and the projection of future consequences and the state of the elements [269]. SA improves the decision-making ability of users by helping them view the right information at the right time and contributing to the UX. The goal of the second part of this thesis is to fill the gap in current studies in the area and help designers create human-centered UIs that provide SA to their users so as to improve their UX.

In summary, this research aims to understand how UIs influence UX in relation to underestimated factors such as accessibility and SA. This thesis takes a hybrid approach, using a blend of methods that combine quantitative measures with qualitative narratives to assess both objective and subjective interpretations of UX.

The ultimate goal of this research is to produce UX design heuristics that can help both the research and the design communities.

1.2 Research Statement

Starting from a careful analysis of the literature on the aspects of underestimated factors such as accessibility and Situational Awareness described above, in the context of our research, elements that could be improved, are summarized as follows.

From the point of view of accessibility:

- **Most developers do not integrate the needs of deaf people in the design of the UI.**
  Most of the people who are deaf or have hearing disabilities have difficulties accessing digital technologies whose UI has failed to include their needs in using the system. This makes the accessibility of digital technologies complicated for the deaf to use, and also contributes to limitations in the UX as they interact with such technologies.

- **Visually impaired/blind people have difficulty accessing or using most UI.**
  Most of the UI fail to recognize and incorporate the needs of the visually impaired/blind populations, making access to the technology difficult for people with this disability. In most cases, the designers make applications that do not reflect the needs of the visually impaired. As a result, many people with disabilities may not be able to access or enjoy the benefits of such technological advancement which also adversely affects their UX.

- **The needs of people with cognitive disabilities are not reflected in the design of the UI.**
  People with cognitive disabilities, particularly those with dyslexia, have problems accessing and fully utilizing the benefits of digital technologies used in training programs due to the lack of integration of their needs into UI design. As a result, this hugely affects their UX with such platforms and theoretical learning

outcomes compared to other standard learners. The appearance of the UX for this type of disability is essential when the deficit is behavioral, as is the case for children with behavioral disorders. Often these children during therapy appear to be fearful, showing little involvement.

- **The growing need to interact with UIs, especially mobile applications, makes accessibility essential.**
  Apps are now used by billions of users for almost all human interactions, and this trend is constantly increasing. For this reason, all users need to be able to interact with the UIs. Unfortunately, some of the UIs make UX bad with access to digital technologies due to their complexity and lack of ability to cater for the needs of people with disabilities. This is because some developers are unaware of people's needs or guiding principles that improve accessibility and lack tools to design such accessible tools.

While, with regards to the Situational Awareness the followign improvements should be achieved:

- **Designing innovative UIs for remote ground control could keep operators' SA levels high by reducing the negative impact of human misconduct on mission success.**
  The lack of attention to UX when designing UI to monitor and fly drones is one of the main causes of accidents and, therefore, the failure of missions. Specifically, the interfaces for the Ground Control Station that the operator must use to monitor the mission and communicate with the aircraft are not very efficient and adequate to allow him/her to maintain high levels of SA and to be able to act proactively throughout the mission, because sometimes designed without providing the operator with the appropriate tools to understand the altered state of the aircraft and take direct control of it. Moreover, in a collaborative work environment, it is necessary to improve the design of the UI so that it is simple and easy to use to improve operators UX to increase SA level.

- **UIs used in manufacturing and industrial plants are complex and difficult to use and understand.**
  In industry, many incidents occur due to the increasing complexity of UI which makes it difficult to obtain and interpret data in time to avoid problems. It is necessary to improve such monitoring UIs by first trying to improve their UX, and to maintain a good level of SA to improve effective responses in case of problems occurring in industrial and production plants.

All the aspects described above have been faced in this thesis or are still under investigation. Based on the issues described above, in the following section, the contributions of this thesis are presented.

1.2.1 Research Questions

One main research question is identified, which was used to guide the research:

> **RQ:** *Do key constructs, such as accessibility and SA, contribute to the UX?*
> Previous studies have shown that aesthetics, usability and content influence user's judgment on UIs; and therefore, they also influence the UX which depends on the context, on the framing of the activities and on user's background [305, 306, 307, 308]. Here, additional factors such as accessibility and SA will be used to explore UX.

This is the focal point of the research, which is split into two different **RQ**$_s$:

> **RQ**$_1$**:** *Does accessibility in UIs lead to a more positive UX?*
> **RQ**$_2$**:** *Do UIs with high SA levels improve UX?*

1.2.2 Research Approach

The approach selected to investigate the above **RQ**$_s$ is Design Science Research (DSR), also known as Constructive Research. It is a methodological approach that deals with designing: produce something that does not exist yet, or modify existing solutions for better results [309].

Design science research has been defined as "*research that invents a new intentional artefact to address a generalized type of problem and evaluates its usefulness in solving problems of that type*" [311]. It is a form of scientific knowledge production that involves the development of innovative constructions, intended to solve problems faced in the real world, and at the same time makes a kind of prescriptive scientific contribution. The interest in this scientific research approach has increasingly emerged in the field of computer systems and in [310] the authors emphasize its importance.

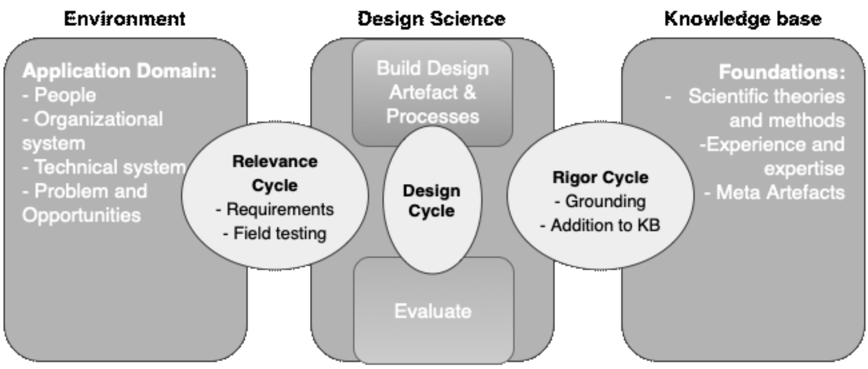


Figure 1.1: Design Science Research Cycles.

Hevner identifies three inherent research cycles in each design research project, namely: the *Relevance Cycle*, the *Rigor Cycle* and the *Design Cycle* as shown in Figure 1.1 [312]. The presence of these three cycles in a research project distinguishes design science from other research sciences. The *Relevance Cycle* connects the contextual environment of the research project with the activities of design science. The *Rigor Cycle* links design science activities with the knowledge base of scientific foundations, experiences and skills that inform the research project. The central *Design Cycle* iterates between core construction activities and artefact assessment and research design processes.
These three cycles must be present and clearly identifiable in a research project on the science of design.
In [316] the authors describe design research through a conceptual framework made up of guidelines or principles useful for conducting and evaluating good DSR.

Table 1.1: Guidelines for the Design Science Research approach.

n.	**Guideline**	**Description**
1	Design as an Artefact	Design science research must produce a viable artefact in the form of a construct, a model, a method, or an instantiation
2	Problem relevance	The objective of design science research is to develop technology-based solutions to important and relevant business problems
3	Design evaluation	The utility, quality, and efficacy of a design artefact must be rigorously demonstrated via well-executed evaluation methods
4	Research contributions	Effective design science research must provide clear and verifiable contributions in the areas of the design artefact, design foundations, and/or design methodologies
5	Research rigor	Design science research relies upon the application of rigorous methods in both the construction and evaluation of the design artefact
6	Design as a search process	The search for an effective artefact requires utilizing available means to reach desired ends while satisfying laws in the problem environment
7	Communication of research	Design science research must be presented effectively to both technology-oriented and management-oriented audiences

Subsequently, Henver et al. present a checklist derived from the previous guidelines [310]. The checklist is useful for making sure projects follow the design science research approach.

Table 1.2: Checklist for the Design Science Research approach.

Questions
1. What is the research question (design requirements)?
2. What is the artefact? How is the artefact represented?
3. What design processes (search heuristics) will be used to build the artefact?
4. How are the artefact and the design processes grounded by the knowledge base? What, if any, theories support the artefact design and the design process?
5. What evaluations are performed during the internal design cycles? What design improvements are identified during each design cycle?

6. How is the artefact introduced into the application environment and how is it field tested? What metrics are used to demonstrate artefact utility and improvement over previous artefacts?
7. What new knowledge is added to the knowledge base and in what form (e.g., peer-reviewed literature, meta-artefacts, new theory, new method)?
8. Has the research question been satisfactorily addressed?

Figure 1.2 shows how the eight questions are mapped into the three cycles of design science. This demonstrates the relationship between the checklist and the search cycles.

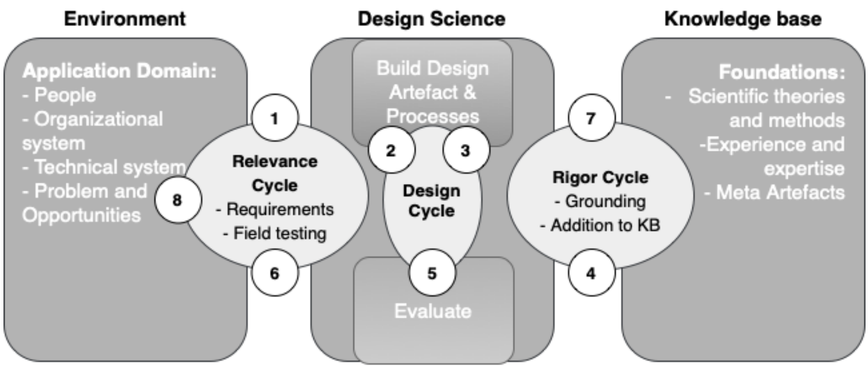


Figure 1.2: Questions mapped to three Design Science Research Cycles.

In the last chapter of this thesis, it is shown how the investigation performed in this thesis answered the questions of the DSR by highlighting the parts of the thesis in which each single question is answered.

1.3 Research Contributions

This thesis provides several contributions aimed at bridging the limitations described in section 1.2. In particular,

- *Inclusive solutions for deaf people.* We followed a detailed study on the difficulties deaf people face in their social life. Then, we leveraged them to overcome UX challenges by designing UIs and solutions that could be adopted anywhere by hearing-impaired/deaf users. The solutions proposed and explored through this thesis allow as to improve communication and inclusiveness in accessing information and services for deaf people. Also, because the barriers to communicating with deaf people are caused by unfamiliarity with sign language, we have provided a solution for anyone to learn sign language.

- *An assistive solution for blind people.* The main concerns of the blind community are those regarding their independence and their empowerment in daily activities. The need for innovative care solutions for the visually impaired community is increasing. The proposed solution is capable of increasing their autonomy in dressing. The aspect of autonomy in daily activities through the design of accessible UIs is fundamental to improve the UX of these people.

- *Solutions to support cognitive disabilities.* The design of UIs must also consider the limitations and cognitive disabilities of the user. The proposed solution improves the learning of students with cognitive disabilities, especially those with dyslexia. The goal was to improve the learning outcomes of children with learning difficulties. UX is essential for cognitive disabilities, especially for people with behavioral disorders. The proposed therapeutic solution for children with attention uses a humanoid robot, in order to increase their involvement. The challenge for this disability has been to ensure that users with cognitive disabilities have the same UX as close as normal people when interacting with UIs.

- *An exploratory empirical investigation into the accessibility of mobile apps were conducted.* In particular, we analyzed in a preliminary phase the needs of users with disabilities by identifying guidelines for developing an accessible application. Then, we tested how accessibility guidelines are implemented in mobile applications through a co-coding strategy. Finally, we investigated developer behavior to probe the problems and challenges of practical accessibility management.

- *A UI to remotely operate a fleet of drones.* We have designed a UI of a Ground Control Station for controlling drones to ensure better performance and reliability, as bad design is a major cause of accidents. We have analyzed the human factors that influence drone missions, and it has also emerged the need to design innovative UIs that are able to maintain high levels of SA. Especially for a single operator who has to monitor a fleet of drones. This UX challenge was subsequently moved from the single operator to the decision makers of an entire team. We have redesigned the UI to operate a fleet of drones in a collaborative environment. We tested the hypothesis that shared displays can improve the SA of the operator team and therefore their performance.

- *A set of guidelines for UIs in manufacturing plants.* The UX of such UIs is a fundamental challenge to increase the performance of production processes, and to reduce the number of emergencies and accidents. These guidelines are based on usability principles after a review of existing industrial interfaces. Next, we prepare a UI to test the guidelines identified to test their effectiveness via a realistic case study prepared or with the support of manufacturing experts.

- *A series of high-level heuristics.* All the studies discussed in the thesis, in the context of either accessibility or SA, rely on a fully participatory approach to the evaluation of the UX. They have been conducted with a variety of different, representative users and with different needs, in order to guarantee the validity of the results. Downstream of all those studies, a series of heuristics were derived to support the design and evaluation of UIs in both contexts analyzed within the thesis.

1.4 Thesis Organization

This PhD thesis is organized in 9 Chapters.

Chapter 1 provides an overview of the problem and the motivation behind this work and more general background required to follow the rest of the thesis.
Chapter 2 introduces the necessary foundations of this research. It provides a brief overview of the research on User eXperience Design, and subsequently on Accessibility and on Situation Awareness, as well as the existing literature in the context of the topics covered that lead to this work.
The rest of the Thesis is structured in three parts:

1. **Part I. Focus on Accessibility**
   In chapters 3-6 the first part of this research work is elaborated, which involved the study of the relationship between accessibility and UX as well as the variety of methods that can be used for the evaluation of UX itself. **Chapter 3** addresses some challenges of UX design for deaf users by proposing solutions - which can be adopted during their daily life to learn sign language, to communicate with others an to gain inclusion in information and services (e.g., public transportation companies, municipalities, hospitals, schools and higher education institutions). **Chapter 4** focuses on another very large community of visually impaired users and addresses the challenges of UX design providing them with greater autonomy and independence in daily life. **Chapter 5** concerns the design of a UI that reflects cognitive processes. In addressing this problem, we started from the belief that the design of a UI should reflect and fully consider the limits and capabilities of cognitive processing, especially in the case of cognitive disabilities. In summary, to get a well-designed UI, we should not only focus on task user tests, but also take an in-depth look at the cognitive processes entailed by those tasks. The specific domain dealt with is that of digital technologies for training and the challenge faced was that of providing a service that not only supports school teaching, but that can also motivate students more and more while learning, especially in cases of cognitive disability. In addition, a project for the use of a humanoid robot, through a UI on an integrated tablet, in the therapeutic treatment of children suffering from behavioral disorders is presented within the chapter. **Chapter 6** opens with a discussion on the lessons learned from the

experiences described in the design of interfaces for disabled users and continues by illustrating the results of a study aimed at evaluating the extent to which developers are now aware of and take into account the guidelines for the accessibility, which are the most followed and what are the possible reasons why some fundamental aspects of UX related to accessibility are still neglected.

2. **Part II. Focus on Situation Awareness.**
   In chapters 7-9, the thesis describes the second part of this research work, which focused on understanding how UI design influence the aspect of UX that affects situational awareness. **Chapter 7** presents the design of an interface for control rooms for monitoring fleets of drones. The same interface was then redesigned to be used in a collaborative environment, and therefore by a team of operators. A productive and inspiring interface helps ensure high-quality interactions. Therefore, it was necessary to explain the influence of UI design on human behavior and establish standards for measuring interface characteristics. Also, in the design of systems for Industry 4.0 it is essential to study how to interactively communicate, view and analyze the numerous data generated in the factory. **Chapter 8** discusses this case study and concludes with the presentation of a set of guidelines to help professional developers design usable interfaces for monitoring industrial manufacturing processes. My goal for this second part has been to improve awareness of the aspects that are of particular concern to UX.

3. **Part III. Conclusions and Future Research Direction**
   Chapter 9 contains a global review of the thesis work, revisiting the main conclusions drawn. These findings are reviewed in relation to the original research questions. The thesis objectives are then discussed in view of the results obtained, defining some heuristics, and discussing possible future lines of research.

Chapter 2

Background: UX - Accessibility and Situation Awareness

This chapter examines the literature surrounding UX, drawing on a broader research area beyond HCI. It outlines the significant aspects that shaped the development of UX and provides an overview of the main theoretical models used to aid the evaluation and design of interactive technologies. Broader influences affecting UX are also discussed in relation to how they can affect user quality judgment, particularly accessibility and situational awareness.

2.1 User Experience

The term User eXperience is well known to the HCI academic community and industry practitioners and has been related to a wide variety of meanings, ranging from usability, aesthetics, to emotions, affectivity, and experiential value. Yet it has been regularly criticized for being too vague, obscure and ambiguous. The UX landscape is fragmented with complicated models and theoretical approaches developed to help understand the experience, with contributions from design, business, philosophy, psychology, cognitive and computer science [276]. However, there is still no consistent theory with which to evaluate and design for the experiential qualities of the technology. Although, yet to be defined, many agree that UX is a dynamic, highly context-dependent and subjective account of human-machine interaction, which is strongly shaped by individual emotions [277].

This chapter examines the main literature surrounding UX by first outlining the major influences that prompted the advent of UX, drawing from a broader context beyond HCI.
The review begins by identifying the literature emerging from usability to UX, in Subsection 2.1.1, motivated by the need to look beyond traditional usability, as it did not take into account the subjective nature when interacting with technology.
In Subsection 2.1.2, due to the multidimensional aspect of the UX, the problem of getting a clear definition of UX is outlined, along with the most recent suggestions.
In Section 2.2, the nature of accessibility is discussed in relation to UX. In order to expand existing UX research, the area of situational awareness is discussed in Section 2.3, along with the growing area of research that UX is studying.

2.1.1 From Usability to Experience

Early UX research argued that usability concepts were too focused on task efficiency, productivity and learnability, and a more boarder notion of quality was needed. The problem with usability was its focus on the instrumental aspects, which was highlighted by Gaver & Martin [278], who argued for a wider range of alternative list of needs in which to evaluate technology.

What is Usability?

"*Usability of a system or equipment is the capability in human functional terms to be used easily and effectively by the specified range of users, given specified training and user support, to fulfill the specified range of tasks, within the specified range of environmental scenarios*" [279].
Usability is a fundamental concept within HCI that has been debated for decades, both in terms of its definition and its use as a measurement tool [280]. It was first introduced in the 1980s to replace the term 'user friendly' by compelling designers to take a more user-centered approach and consider the users all the way through the design stages [281]. This was a turning point in computer system design, which traditionally focused on a top-down design approach, where all requirements were outlined in the planning phase. Usability evolved through various stages of definition,

where initial expressions of 'ease of use' [298], were characterized by Eason, using four components, the user characteristic's, task characteristic, system function and the environment [299].
This was later expanded by Shackel & Richardson [282], who added four operational components (effectiveness, learnability, flexibility, and attitude) to enable usability evaluation during use, which paved the way for Nielsen's systematic evaluation methods [283], consisting of various measurable attributes (e.g., learnability, efficiency, memorability, errors and satisfaction). The first formal definition arose in 1998 by the International Organization for Standardization, (ISO 9241-11), defining usability as, "*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*" [296]. This definition places emphasis on the measurable aspects (effectiveness, efficiency and satisfaction) that should be assessed in terms of context of use (task, user and situation).

The Problem with Usability

Both effectiveness and efficiency are the objective measures of usability, so more easily measured, while satisfaction is a more subjective measure. Effectiveness refers to the accuracy and completeness with which users achieve their specified goals, typically measured through learning and completion time, while efficiency is measured by error rates, task completion and the quality of the outcome of a given task. However, satisfaction relates to the users' freedom from discomfort and their positive attitudes towards using the product, as measured through psychometric scales on overall preference, and general perceptions of product quality [284, 285]. It was assumed that these three measures of usability strongly correlate, as outlined in a study by Nielson & Levy [286], who conducted a meta-analysis of 57 studies and found a strong association with users average task performance and average subjective satisfaction. However, these findings have since been questioned by a number of subsequent studies [284], that suggest subjective perceptions of usability differ from objective performance measures, and can depend on the users past computer experience and preferences [287, 288]. Despite the usability standards being 'enhanced' in [297] (ISO 9126) to include a further two characteristics (learnability and security), both ISO standards lacked

harmony, as they were developed from two different perspectives (e.g., ISO 9241-11 originates from HCI experts, while ISO 9126 came from the Software Engineering field), making them difficult to interpret and therefore apply [289].
Due to the lack of a consistent standard for usability and the inconsistencies between the objective measures of performance and the subjective assessments of user judgement, evaluating usability proved a challenge. However, a variety of practical approaches to evaluating usability were proposed, such as Heuristic Evaluation [286], cognitive walkthroughs [290], GOMs [291] and user testing methods [292]. These usability evaluation approaches assume that it is possible to identify usability to varying levels of granularity. Nielsen [283], offers a wide range of design principles or "heuristics" that are conducted by small groups of evaluators to determine the usability of websites [293, 286]. According to Nielsen, usability comprised of five attributes (learnability, efficiency, memorability, errors and satisfaction), which can provide valuable diagnostic aid to practitioners to improve usability [283]. Although these methods are useful in ensuring products are more usable, they were found limited when assessing subjective aspects such the aesthetics and attractiveness [294] of a product, and in evaluating how these may influence the emotional experiences of users' while using a system [295]. Therefore, a broader prospective was needed, that extended the traditional usability concept to allow for a user's experience.

2.1.1 Defining User Experience

User experience evolved from the concerns that traditional usability focused on the objective measures of user performance and did not account for the emotional and affective responses to the aesthetics of an interactive product. Similar to usability, the ability to evaluate UX has become the focus of both the HCI research community and industry practitioners alike. Despite the term 'user experience' being widely used and readily accepted, there is still uncertainty about what it actually means. Due to the multidisciplinary nature of UX research, which draws from computing, psychology sociology, and design and beyond, it's not surprising that no clear unified theory of user experience exists. Only a few attempts at providing a definitive definition have been made. For example, a survey conducted by Law et al. [277], asked 275 researchers and practitioners within the field of UX to come up with a common agreement about what UX is. The result outcome was that UX is an individual experience that is "*dynamic, context-dependent, and subjective*". In a more recent white paper on UX, the term is defined as encompassing three different perspectives; 'experiencing' which refers to "*the individual and dynamic nature of experiencing the encounter with a system*"; 'a user experience', which is a retrospective perspective, based on the "*outcomes and memories of an experience*", and 'co-experience', which refers to "*experiencing a situation together*". Although this goes some way towards identifying variances within UX, there is still a lack of clarity in understanding what UX is. Despite this, there is a vast array of UX research with many suggested theories and frameworks, which have made some progress in teasing out some of the complex intertwined aspects that embody UX.

2.2 Accessibility in UX

Accessibility is a measure of a user's ability to use products/services, the extent and ease with which they can achieve their goals. Designing with accessibility in mind allows people with a range of abilities and disabilities to perceive, understand, navigate, interact with the UI. Doing so, therefore, has significant benefits, particularly for better UI design for everyone.
Accessibility is defined by [300] (ISO 9241-210) as "*the degree to which a product or system can be used by people with a wide range of characteristics and capabilities to achieve a particular goal in a specific context of use*". Therefore, Accessibility contributes as a facilitator in the use of technological systems.

There are accessibility laws to help people with disabilities, but designers should still try to accommodate all potential users in many contexts of use. Indeed, many jurisdictions, including the EU, have penalties for failing to create accessible designs.

Accessibility is not only the right thing to do, it often benefits all users as well. This is because accessibility features that help people with disabilities

often help other people as well. For example, video captions that help people with hearing difficulties also help a person who is watching the video with the sound off (for example, in a social media feed). The readable, high-contrast text that helps people with low vision also helps people with perfect eyesight who use the app outdoors in bright sunlight. Many users, regardless of their abilities, will face challenges due to challenging contexts. When designing for all skill levels, you can create UIs that anyone can use and enjoy.
While accessibility is a critical factor affecting design, many overlook it. We are all, designers and users, different. Some of us have dyslexia; others have partial hearing loss, for example. The areas of user needs that must be considered when designing accessible UIs are:

- **Visual:** low vision, blindness, color blindness, are all forms of visual impairment.
- **Motor/Mobility:** This category extends not only to problems with the use of the hands and arms, but also with other muscular or skeletal conditions.
- **Hearing:** Hearing impairments affect hearing and occur in varying degrees of severity, up to and including total deafness.
- **Convulsions:** some individuals can be affected by light, movement, flickering, etc. on the screen, thus triggering seizures. The most common problem in this category is photosensitive epilepsy.
- **Learning/cognitive:** It is also important to remember that not all disabilities are physical. Cognitive and learning disabilities can also affect accessibility.

Because they have similarities, accessibility is sometimes confused with usability. Both overlap and are vital parts of UX design, but there are also key distinctions between them.
In fact, another relevant quality criterion to consider in UI design is usability, defined by [300] (ISO 9241-210) as the "*extent to which a product can be used by specific users to achieve specific objectives effectively, efficiency and satisfaction in a specific context of use*". An interface with good Usability provides greater user satisfaction with the product. Finally,

UX has emerged as a new way to understand and study the quality of interactive products [302]. The UX "is considered as the emotions and expectations of the user with the context of use, where emotional factors and personal experiences stand out" [301]. Positive user experience is fundamental for adoption and acceptance success of an interface.

Usability, UX and Accessibility are related perspectives. For some researchers, usability is considered part of UX [304]. Furthermore, Accessibility appears as a subset of Usability [303], which means that usability includes accessibility.

However, this view may indicate that Usability should take priority over Accessibility. For example, a usability rating alone is not sufficient to meet accessibility requirements. The definition of ISO 9241-210 [300] suggests that usability is a subset of accessibility [303].

This vision can also lead to an understanding that it is only necessary to make technologies accessible, but without aspects of ease of use, learning, among others.

Interaction barriers can be solved by Accessibility, but there may still be an impediment when interfaces are not well designed. Therefore, the UI should be evaluated more comprehensively, analyzing different perspectives, such as Usability, UX and Accessibility, which are relevant factors for its system quality.

2.3 Situation Awareness in UX

When designing UIs, the goal is to develop an interface with which the user can interact effectively and efficiently. The wrong designs can lead to UIs that are not used, are ineffective, and in some cases dangerous for the end user.

In particular, human operators using systems with complex UIs must correctly perceive useful information while ignoring or neglecting other stimuli. Situation Awareness (SA) provides a framework for describing human performance in activities ranging, for example, from controlling drone fleets to monitoring industrial production processes. At a basic level, an operator who demonstrates perfect SA knows what information around

him/her is relevant to the task, what this information means for the present, and what this information will mean for the future. With these types of knowledge, the operator understands the current state and can effectively project his understanding into possible future states of the system.
Therefore, poorly designed UIs can also lead to low SA levels. One way to design effective UIs that have no problems is to focus the design process on maintaining and improving the end user's SA, so as to improve the UX as well. Specifically, the operator UX can be improved by incorporating usability into the UIs design. Improving the usability of a UI requires, in fact, the improvement of the SA of the operators who use the UI. The design of the UI will affect how operators can develop and maintain SA while on the job.

The definition includes several concepts that are important for understanding the SA construct. Endsley [217, 313] defines three levels of SA: perception, understanding, and projection. Perception is the basic level of situational awareness (SA level 1). This level of awareness is reached if the operators can perceive in the user interface the information necessary to carry out their work. The next level is understanding (SA level 2). Not only must information be perceived, but it must be combined with other information and interpreted correctly [220]. The third level (SA level 3) is the projection or the ability to predict what will happen next based on the current situation [314, 315].
So, by designing UIs that support all levels of SA and sharing SA among team members, we are more likely to develop effective and efficient UIs that support decision making and performance. This becomes even more critical, in fact, with complex UIs in which there is a large amount of information to keep up with, wich changes rapidly and in some cases is difficult to obtain [313].
In conclusion, the design of the UI will affect how quickly and easily operators can advance through the stage of SA's performance and how accurate and complete the understanding of the operator is at each stage.
The SA phases provide a framework for evaluating performance and identifying the activity and interface factors that can moderate SA performance, and consequently improve UX.

UX can influence the mental model that provides the structure used to process information related to the task. This model is stored in memory, which means it can be learned or partially forgotten and may not match the UI design.

The mental model of an operator situation provides the tools necessary to manage large amounts of information. They use their UX to stake out the intake of new information, noting what to pay attention to, what to discard, and what to remember for a given situation.

Therefore, situational awareness, awareness of the state of the world, what is happening and what will happen, is based on operator's UX.

When applied to UI design, this approach suggests that each stage of processing and operator response is essential for the system's proper functioning. The operator must be able to see and process the stimuli. They must have attention and time to understand them and the ability to recognize that the stimuli are important. They need to have an appropriate mental model to relate new information to previous information and current goals. They need to know what to do and how to respond. And they need the state of the world and a good mental model to predict what will happen in the world.

Situational awareness thus provides a way to organize UIs design for operators to improve their UX.

Part I
Accessibility

The first part of the thesis collects the studies that aim to obtain an improvement of the User eXperience in which the focus is given to the accessibility factor of the resulting User Interfaces.

Some challenges of UX design for deaf and blind users are addressed by proposing solutions that they can adopt during their daily life offering greater autonomy and independence during their activities. Also, in this part of the thesis the design of a UI that reflects the processes especially in the case of cognitive disabilities is presented. Finally, it concludes by illustrating the results of a study aimed at evaluating to what extent developers are aware of and take into account the guidelines for accessibility, and what are the possible reasons why some fundamental aspects of UX related to accessibility are still neglected.

Chapter 3

UX challenges for Deaf Users

Today the deaf community represents a population of over 70 million individuals worldwide. This chapter focuses on analyzing the difficulties that deaf people are still facing in their social life, and therefore on the challenges of UX design, which led us to design a mobile software solution that could be adopted everywhere by both deaf and hearing people as an effective tool. of communication and learning in different specific situations. Section 3.1 describes the current prototype and discusses the results of the usability testing activities carried out so far. Sign language is recognized as one of the most popular means of communication for the deaf. The Convention on the Rights of Persons with Hearing Disabilities recognizes and promotes the use of sign languages and states that their use is essential to ensure access to information and services, both in daily life and in emergencies. Nonetheless, most public information and service providers do not meet the goal of providing information and services through sign language interpreters. Therefore, leveraging the results of the work presented in Section 3.1, Section 3.2 proposed a cloud-based platform designed as an intermediary service for any person who should provide inclusive access to information and services even for deaf people. (e.g. public transport companies, municipalities, hospitals, schools and higher education institutions). The idea is to promote a cost-effective transformation of public service providers towards the goal of leaving no one behind, which is at the heart of the 2030 Agenda for Sustainable Development, adopted by all UN member states in 2015. However, hearing people are not familiar with sign language and this can pose unpleasant barriers to deaf people. One of the biggest challenges is raising awareness of the importance of sign language by providing learning support. Thanks to recent advances in the field of Deep Learning, it has been possible to develop a new method for recognizing the ASL alphabet, using convolutional neural networks (CNN), to monitor the user's learning progress. A robust model has been produced that correctly classifies gestures, and the experimental results have encouraged us to study the adoption of artificial intelligence techniques to support the learning of sign language, as a natural language with its syntax and vocabulary. Thus, the challenge presented in Section 3.3 was to provide a mobile solution for sign language training that users could adopt in their daily lives. To meet the necessary additional computational resources for locally connected end-user devices, the adoption of a Fog-Computing architecture is proposed.

3.1 ProSign Everywhere - Addressing Communication Empowerment Goals for Deaf People

3.1.1 Introduction

Severe and/or profound deafness, from any cause generated (before, during or immediately after birth) has as its primary consequence the loss of the period useful for the acquisition of the word, which corresponds to the first three years of life. As a result, damage due to acoustic sensory deprivation may occur and the associated effects will not only affect the learning process of the word but will also negatively affect the global perceptual mechanism and consequently the behavior of the subject [1]. Therefore, the identification of an effective alternative method to spoken language becomes crucial. The communication means accessible to people affected by deafness can be mainly divided into two broad categories: the lip reading and the mimic-gestural language. With lip-reading, the deaf person strives to understand what is said by reading the lips of his/her interlocutors and trying to relate to them speaking. This approach is very difficult for the deaf. Hearing people, for example, can speak too quickly and without articulating words properly or their face may not be sufficiently visible due to their position or low light. Moreover, it can be very hard for a deaf person to speak, because the learning of this practice is not natural and spontaneous. With the mimic-gestural language (or sign language) the deaf communicates using gestures. It has certainly happened to each of us to see two or more people communicating with each other in this way. What immediately strikes is that they move hands and arms, drawing repetitive

movements in the air, their faces assume various expressions, which help to clarify the message traced in the air with the dance of their hands. Unfortunately, such a communication method has its own drawbacks. The knowledge of sign language by hearing people is very rare and this does not allow deaf people to be adequately understood [317]. For the same reason, a deaf person with the only knowledge of sign language is not able to understand a person who communicates verbally. In this regard, the human translator is used in many circumstances to facilitate the exchange of information between hearing and deaf people (e.g. at school, on television, at conferences) and to increase the effectiveness of the transmission of information in a written form in the educational processes. Although science cannot completely overcome the hardships created by physical and natural compromises, researchers have worked hard in recent decades to give deaf people the chance to perceive the sound of the world around them [2]. Our work originates from the analysis of the difficulties that deaf people still encounter in their social life. The application, named *ProSign*, was conceived as a mobile software solution which could be adopted everywhere both by deaf people as well as by hearing people as an effective communication and learning tool in different specific situations.
Structure of the section. Subsection 3.1.2 discusses some relevant related work. Subsection 3.1.3 summarizes the problem domain analysis activities and the resulting functional and non-functional requirements. Subsection 3.1.4 introduces the main features of the ProSign app. Subsection 3.1.5 describes the initial usability testing activities carried out with the prototype. Subsection 3.1.6 concludes the section.

3.1.2 Related Work

Many recent efforts have focused on the search for a technological solution able to alleviate the communication difficulties of deaf people. In particular, several applications have been developed which apply methodologies for translating, more or less automatically, verbal language into sign language. Special attention has been put on the three-dimensional rendering of movements understood by a deaf person or "gestures" of sign language [3, 4, 5, 6, 7]. Some techniques are limited to show the movement of hand and fingers for fingerspelling, which is considered an important component of American Sign Language (ASL) and a necessary skill for complete communication sign. The University of Maryland and Texas A & M developed a website to show their system [3]. It shows the three-dimensional rendering of a hand for fingerspelling. In 2002 a similar example was created for the language of the South African signs. The VCom3D [4] produces an avatar-based tool that can read books for deaf children in sign language. Also, Sign4me is an app for learning sign language, which provides sign language instruction in 3D. An English project, TESSA (Text and Sign Support Assistant) [5] provides an assistant in the offices thanks to three-dimensional animations. This combines speech recognition and virtual animation with avatars. The English project SIMON uses an avatar to translate the texts in Teletext format received from a TV in English Signed. The PAULA project [6] was developed at the DePaul University of Chicago, led by Karen Kolby, a deaf student [7]. Normally, such approaches require systems with high computational capabilities and high graphics performance: they are not suitable for use on smartphones. Among the available mobile application, we mention FaceTime [9], only available for iOS, useful for communicating in real time via video calls, Google Voice app [10] features speech-to-text functionality, Dragon Anywhere [11] app transcribes everything that is said. Other apps were specifically targeted at deaf people. Far from being exhaustive, the multi-platform app P3 Mobile [8] facilitates video calling deaf and hard-of-hearing individuals, BuzzCards app [12] allows user to communicate through the Cards that are compiled and shown to the interlocutor, ASL Dictionary [13], a multi-platform app which contains a wide range of words translated into ASL, Marlee Signs [14] app helps people to learn ASL, available for iOS, created by the actress Marlee Matlin. One limitation of the existing apps is that they either focus on the language of sign learning functionality or on communication services, no one integrating the two aspects within a unique solution. The final goal of ProSign research project is instead to provide a complete and portable system integrates any functionality which may help to break down the barriers between deaf people and hearing people. The prototype we present is the result of the initial stage of the project.

3.1.3 Requirement Analysis

The United Nations Convention on the Rights of Persons with Disabilities recognizes and promotes the use of sign languages, establishing that sign languages are equal in status to spoken languages [1]. It recommends states parties to facilitate the learning of sign language

The problem domain analysis allowed us to get an insight into the most common issues affecting the autonomy of deaf people. As part of the contextual inquiry, we interviewed a group of 5 people from Ente Nazionale Sordi association in Salerno, Italy, with the help of an expert of the association, who acted as an interpreter from natural language to the Italian Language of Signs, and vice versa. Our goal was to understand what are the problems that a deaf/hard-hearing person encounters during his/her everyday activities when he/she needs to interact with others, in case he/she cannot rely on a companion, who acts as an interpreter between him/her and natural language speakers. We also conducted a focus group interview a group of 8 friends and family members of deaf people, as they were also considered important stakeholders and potential users for our design. We managed the interviews and the focus group in the form of informal 'cognitive meetings', starting from more general questions and then arriving at the specific questions of interest for the analysis activity. Structuring the interviews in that way we found a significant improvement in the clarity and response time of respondents, getting also information related to the possible adoption of assistive tools or software applications, without making any explicit questions. An empowerment-driven analysis method was adopted to elicit UX requirements, as illustrated in [16, 17, 20]. The method depicted in Figure 3.1 is a transformative process which starts from a contextual investigation meant to understand users, their behavior and capacities within a given community, and to identify potential improvements in their life quality, which are expressed in terms of human needs.

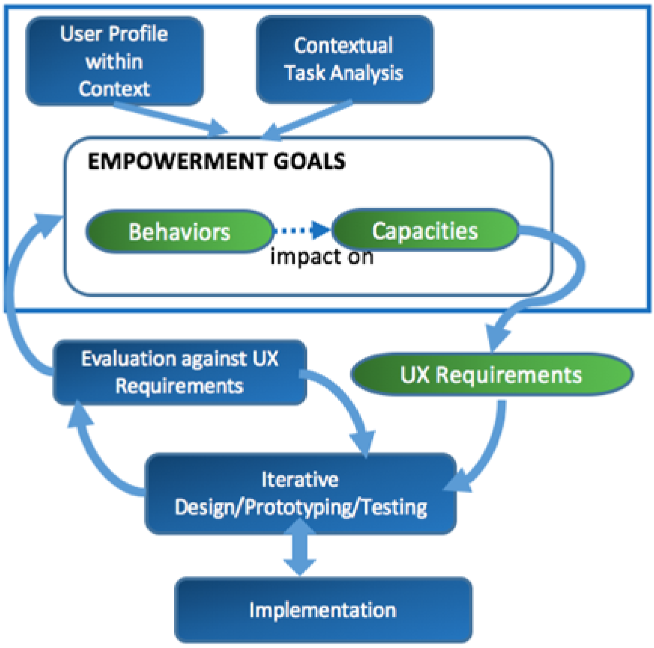


Figure 3.1: Empowerment-driven Requirements Engineering process.

The key factors which may contribute to the empowerment of the target users with respect to the identified needs are considered in order to formulate clear users' empowerment goals. The goals entail a set of UX requirements which may be iteratively modified/extended and ultimately drive usability designers. For each interview, an Mp3 file was generated and then cut using iMovie, in order to retain only the information considered useful for the analysis. As a result of the focus group, we found out that people do not have communication problems at home, especially in being understood. Most of them reported that when they go out, they happen to rely on someone else as an interpreter for certain activities (e.g, inquiring about public transportation, purchasing something in a shop, etc.). None of the interviewees use a specific app but they think that some IT support is desirable, especially when they attend conferences or frontal lessons where it becomes difficult to follow the speaker, who often is at a distance that does not allow to read lips. However, they insisted that they would not use laptops or computers and would rather prefer portable devices for convenience and size. Better support is also sought for rapid phone communications, where easy access of a set of predefined basic phrases is desirable.

In summary, we understood that deaf or hard-hearing individuals experience severe communication difficulties in everyday life. Comprehension of spoken language relies on their lip-reading capability, considering that it is not always easy to grasp the fast-flowing microscopic different positions that assume the mouth, the tongue, and the lips. It is also difficult for them to grasp the nuances of the timbre, of the intonation of the voice, which may suggest different concepts and ideas [18]. Such obstacles may cause a deaf individual to be involuntarily excluded from the social context unless some accompanying person acts as an interpreter. The interpreter may provide considerable support but cannot be considered the solution to communication problems. The presence of a translator forces the deaf person to make his handicap apparent to others, and, in case the translator is a stranger to him/her, harmony in communication can be hard to achieve. Other relevant information was derived from the focus group involving friends and family members. They reported that a common method to learn sign languages is watching online videos and they all agreed that some portable personalized solution for learning sign languages and facilitating communication with deaf people is still missing. Starting from those observations, we began the design of a practical and handy tool that would provide the deaf, and also the people around him/her, with all the information he/she needs in sign language, in a very discreet and confidential manner. We conceived a lightweight solution which could be used at school, on the road, during a conference, simply by installing it on one's smartphone. Therefore, as a result of the analysis phase, the following initial set of functional requirements were established for an integrated portable system fostering communication:

- Learning functionality for Sign Language alphabet, and basic words;
- Customized, an incremental vocabulary of Sign Language words learned so far;
- Real-time display of a GIF corresponding to a typed word;
- Text-to-speech functionality allowing remote communication from a deaf individual to his/her non-deaf interlocutor;

- Speech-to-text functionality allowing remote communication from a non-deaf individual to his/her deaf interlocutor;

The associated non-functional requirements were:

- High level of learnability which would encourage users to fully exploit the potentials of the tool in everyday communication activities;
- Cross-platform application available for all smartphones;
- Cheap and affordable – Core functionality available for free and advanced functionality at low cost;
- Real-time responsive.

3.1.4 ProSign

As mentioned above, the idea to realize ProSign originated from the analysis of the difficulties that deaf people encounter in their social life, for example in free time, at work, during cultural meetings, at school, etc. Therefore, it was conceived as a flexible and extensible software solution which could be adopted by deaf people and by people who interact with them as a communication and learning tool in different specific situations. In fact, it can be integrated into a deaf' s smartphone and used as a valid, always available translator but it can also be helpful for a hearing person who does not know the Sign Language and wishes to break down barriers to communication with deaf, e.g. a family member or a friend or a colleague, or any other acquaintance. The informative immediacy obtained from the system accelerates the knowledge of the new language and allows better and optimized communication.

The Prototype

When the design phase started, we decided that we should target a wider population of potential users and therefore considered translation from/to the American Sign Language (ASL). The graphical interface design took into account general usability guidelines and was iteratively assessed by an automatic usability evaluation tool [19]. The visual information design process customized the popular Material Design system by Google with respect to the identified users' requirements. Let us briefly describe the resulting prototype. When the app is first accessed, four different usage modes are displayed, namely:

- The ASL Alphabet mode, for learning American Sign Language alphabet;
- The ASL Word mode, for learning American Sign Language basic words;
- The Speech mode, to access translation sessions from spoken to written text and vice versa;
- The Video Call mode, to make video calls.

The interface is very simple and appealing, with a lower tab bar which can be used to access the various modes.
In the *ASL Alphabet* mode, a 3D animated character is displayed who shows all the letters of the alphabet (Figure 3.2) so that the user can get appropriate training by imitating the character.

Figure 3.2: The 3D character showing the ASL alphabet.

In the *ASL Word* mode, the user can search the word/phrase via the search bar and the system will display in real-time an animated GIF representing the typed word (Figure 3.3).

Figure 3.3: The translation on the display through an animated GIF.

The function of the app is based on a database of GIFs, which are used to show the user the ASL animation of the searched word or phrase. The database was kindly made available by the creators of the appreciated DVD series ‘Sign with Robert’ [15]. This allowed us to exploit a comprehensive database of GIFs, which contains more than 2,000 educational sign language clips.
A customizable User Profile section has been designed, where Prosign user can add any new learnt word/phrase and update his/her personal learning progress (Figure 3.4). He/she may later directly access his/her personal dictionary for a facilitated communication.

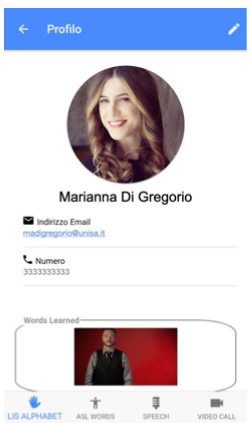
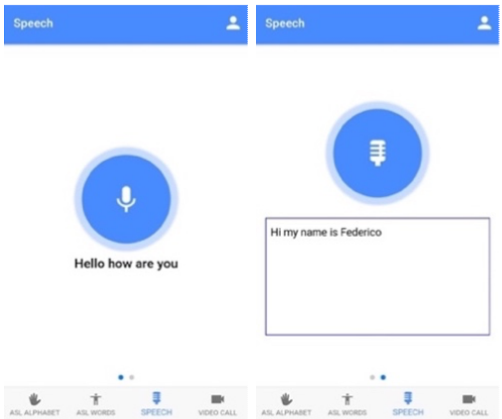


Figure 3.4: The user profile with the words learned.

Text-to-speech and Speech-to-text functionality are also integrated inside the application, corresponding to the Speech mode (Figure 3.5). The two functions are referred

- to the use of speech-to-text to allow deaf people to understand speeches in situations when it is hard/impossible to perform lip-reading (Figure 3.5.a) and
- to the use of text-to-speech to allow deaf people to communicate with hearing people who can't use ASL (Figure 3.5.b).



(a) (b)

Figure 3.5: Speech-to-text and Text-to-speech.

Finally, in the *Video Call* mode, the user can make video calls with any of the registered users through the use of the recipient's email, without the use of third-party apps, (Figure 3.6).

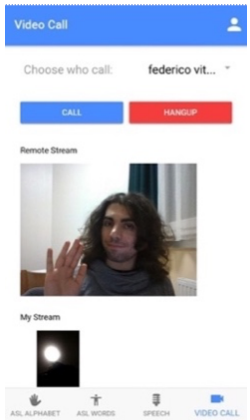


Figure 3.6: The Video Call functionality.

Implementation details

Fostering universal design philosophy, we wanted to create an app which could be accessible by all, usable on any device and therefore we decided to use a multiplatform solution rather than writing distinct applications in native codes, e.g. one for Android and one for IOS. So, we used two frameworks: Ionic for the frontend and Apache Cordova for the backend. Ionic Framework is an open source UI toolkit based on UX and the creation of graphical interfaces. Apache Cordova is an open-source mobile development framework that allows access to all the capabilities of the various devices. Both use web technologies (HTML, CSS, Javascript) for creating and running apps, which can be used on multiple devices. For storing all user information (access data, words/phrases in ASL learned, etc.) we used the services offered by the Firebase platform, provided by Google. Firebase is a serverless service that offers many useful tools for developing web applications. As a development methodology, we decided to use the architectural model MVC or Model/View/Controller. This allows separating the logic of data presentation from the business logic.

3.1.5 Preliminary Interface Usability Study

During the design phase, as a formative usability evaluation activity, we ran an initial usability study meant to test ProSign prototype effectiveness as a communication and as a learning tool. In order to avoid biases coming from existing knowledge of Italian Sign Language we did not involve the people from the initial analysis study but rather people who had no experience with sign languages at all. This was indeed necessary to correctly evaluate the system able to support training and learning. The group was composed of 12 participants (9 males, 3 women). Age ranged from 20-36 (average 25). All of them had a basic knowledge of English written and spoken language. Participants answered a background questionnaire and provided open-ended feedback about their knowledge about the community of deaf people. All participants reported having their phone almost always with them, everyone being curious and interested in learning the ASL language, and only 2 of them having experienced apps to communicate with deaf people. In the following subsections, we describe the study procedures and analyze the initial results gained.

Study Procedures

The initial session of the study took place in a lab setting. Participants had ProSign prototype installed on their smartphones and were initially instructed on its use and asked to watch a short demonstration tutorial. Then, they were asked to perform the following tasks:

- *T1.1 search a word you wish to learn in sign language;*
- *T1.2 enter the word learned in the appropriate section;*
- *T1.3 try the speech-to-text and the text-to-speech functionality.*

For each task, completion times were collected in terms of seconds as well as the number of errors. Upon task completion, participants answered specific questions about their experience and provided free-form feedback. They were also encouraged to ask questions to experimenters. In closing, we asked participants to talk about their experience. After the lab session, participants were asked to keep using the app on their own until the next meeting, which took place one month later. During that period participants were expected to use the application to learn new words/phrases in ASL, so that we could record any improvement in terms of efficiency and accuracy with progressive use of the ProSign but also get an insight into the average number of ASL words/phrases learned after one month time since the beginning. Thus, during the second experimental session, we first replicated the initial experiment and measured the meantime to complete tasks T1.1-T1.3 as well as the average number of errors. Then, participants were asked to perform the following tasks:

- *T2.1 recognize the execution of a word expressed in sign language, out of the personal database of learned words;*
- *T2.2 using the sign language to express a word out of the learned words.*

In order to evaluate participants' recognition/execution abilities, we were remotely supported by an ASL instructor. The tasks were repeated 3 times by each participant with words randomly chosen from their personal set of learned words. To further reduce biases, we tried to choose different words across the participants. For each execution of task T2.1, the expert would mimic a word in ASL. At each attempt, the participant had the chance to ask the ASL expert to repeat the gesture of the word. Similarly, for each execution of task T2.2, he could ask the experimenter to repeat the word he was supposed to replicate in ASL. The tasks were considered completed when all the repetitions were accomplished. To assess the efficiency of ProSign, we measured the meantime used to complete each task. To assess the accuracy, for each participant we registered the number of errors during task completion (wrong execution or recognition of the word).

Study Results

The first experimental study allowed us to evaluate usability considering participants' ability to use the app, and their qualitative feedback. All participants successfully performed the tasks. Eight participants declared to be strongly confident that this app can facilitate communication between deaf and normal-hearing people. Ten think that the application is an excellent support for learning sign language. Some of them suggested that a possible improvement may include a division into categories of the words one can learn as an alternative to the existing direct search bar. Participants generally liked the idea of using a GIF as a result of the research. Most participants used the Speech-to-text and Text-to-speech features correctly, experiencing no problem with those features. One participant forgot to click on the icon to activate voice recording and suggested to make the microphone icon more dynamic to better attract user's attention on the button. When asked, participants imagined many uses of our app in their lives. Most of them mentioned the choice of a word or phrase they would use during a conversation with a deaf person. Others noted that the app would be useful to a deaf user when the person speaking is too far away to be able to read his/her lips. Also, they imagined that deaf people would feel less frustrated if they may use the speech synthesis facility when communicating to people who are not knowledgeable of sign languages. Participants also suggested design improvements for real-world usability. One participant was concerned about the presence of background noise or competing sounds during the recording phase and suggested that a deaf user should be visually notified of the sound level (i.e. the volume) and quality. The second experimental session, not surprisingly, returned much shorter tasks completion times than the first session. Participants expressed no doubts or concerns about app usage.

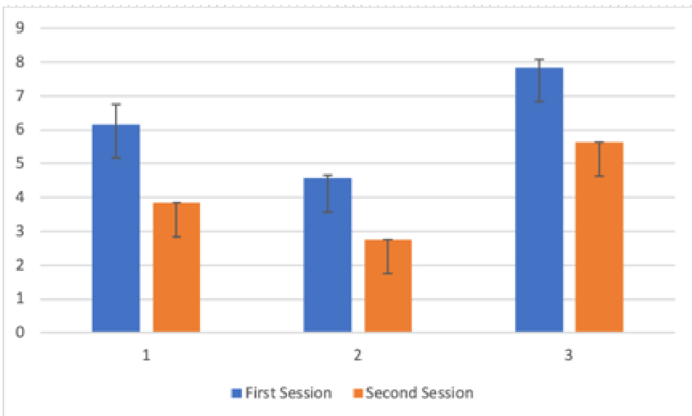


Figure 3.7: Average time to complete tasks and the average number of errors for the two sessions.

In the second experimental phase, we were also interested in measuring an average of the learning progress gained by participants. Our goal was to understand whether *ProSign* can be adopted as a valid 'pocket' tool for sign language learning. We realized that each participant had learned at least 10 new words in the sign language during the self-training period. Two of them had learned over 20 words. Only three made mistakes in the execution of a word in sign language. The ASL expert judged that all word executions were clear and fairly accurate. As for the word recognition task mimed by the ASL expert, four of them failed to guess the word immediately but succeeded after the expert repeated the corresponding gestures slowly. Overall, the expert judged the achieved results satisfactory, reporting that on the average, participants reached a discrete level of learning, albeit their short experience.

3.1.6 Conclusion

In the section, we have presented ProSign, a mobile software solution targeted both at deaf and hearing people, meant to be adopted as an effective communication and learning tool in different specific situations. During the stages of design, ProSign prototype was subject to constant supervision by the deaf people involved in the initial contextual inquiry. They actively collaborated with us in the definition and specification of the app functionality, and they also helped us set up the formative experimental study described in the section. Although the project is still in a prototype phase, they expressed considerable interest in future steps of the research: This will consist in the implementation of an Italian version of the app, based on the Italian Language of Signs. Although the focus of the contextual inquiry was understanding current communication issues, other interesting needs were discovered, which may address future directions of *ProSign* research project.

3.2 ProSign+ A Cloud-Based Platform Supporting Inclusiveness in Public Communication

3.2.1 Introduction

Many deaf people have significant reading problems. Written content is therefore not fully accessible for them. Embodied avatars have the potential to communicate in the mother tongue of this cultural group: sign language. To motivate the concept of signing avatars, we have to give a brief introduction to the culture and language of the deaf. Most deaf people communicate in sign language. Every country has its own specific sign language, and each sign language is a proper language in all its complexity [35] and is fundamentally different from a spoken language. In fact, it is a particularly hard-to-learn language for deaf individuals: it must be learnt based only on a set of written symbols and on observations of highly ambiguous mouth patterns, without any auditory cues – an almost impossible task. Anyway, interactive learning systems can facilitate this task as we will see below. Despite this, even when a sufficient level of confidence is achieved with sign language, natural language reading/writing still remains a hard task, and many deaf pupils leave school with significant writing and reading problems. Often, to make written material more accessible to deaf users, prerecorded videos of human signers are used. However, video contents cannot be modified after production, which makes it impossible to use them in dynamic or interactive scenarios (e.g., train station announcements).

Furthermore, the production cost is high, appearance parameters cannot be adjusted (clothes, gender, lighting), and videos cannot be anonymized. Therefore, signing avatars could complement the current range of human signer videos. With intuitive tools, production costs could be low, and post-production adjustments could be easily made (even automatically or at runtime). Sign language avatars could be used for the automatic translation of content, sign language lexicon visualization or simple train/flight announcement services.

The present work aims at achieving automatic translation from text/speech to sign language. A sentence in a written or spoken language is analyzed and transposed to a sequence of signs. These signs are displayed using animation. With this aim, we developed a system architecture in the form of a Platform As A Service (PAAS), we called ProSign+.

In the research presented in the previous section, we experienced the high acceptance gained by mobile applications within the deaf community as a valid communication means. Therefore, the architecture is designed so that the automatic translation service is accessible by users through a smartphone application, while it is available to content producers through a RESTful API, or a web page. ProSign+ is made of Web Services, a repository for the video messages with Sign Language and a Publisher/Subscriber (Pub/Sub) Broker. The use of ProSign+ isshown through three scenarios describing potential situations, where it can play an important role in the social inclusion of deaf people.

Structure of the section. In the next subsection, discusses some related work. In Subsection 3.2.3, introduces the system architecture that allows publishers to share their multimedia contents in sign language. Subsection 3.2.4 describes an application of the system architecture through three scenarios. In Subsection 3.2.5, describes the plan to evaluate the whole system with experts and real users. Finally, in the last subsection, a discussion and some conclusions are given.

3.2.2 Related Work

Signing avatar is a relatively young research area with some significant results. A major prerequisite for sign language synthesis is a representation system or notation. In gloss notation, each sign is denoted with a word (gloss) that most closely corresponds to the sign's meaning [29]. For instance, the sign sequence for "*What's your name?*" would be YOUR NAME WHAT. Gloss notation, however, does not describe how to execute a sign. The same gloss may be executed in various ways due to grammatical modifications and dialect variations. Historically, a milestone notation for the description of how to execute a sign was Stokoe notation [35] which formed the basis of modern notation systems like the widely used HamNoSys, the Hamburg Notation System for Sign Language [33].

In the research area of signing avatars, one can distinguish approaches along the articulatory vs concatenative axis [27]. While concatenative approaches piece together prerecorded chunks of human motion, articulatory approaches compute motion on-the-fly based on a sparse specification. Two influential European projects, ViSiCAST, and eSIGN, developed technology for signing avatars based on HamNoSys [22, 30], transitioning from a concatenative to an articulatory approach, and advancing HamNoSys to SiGML. The resulting avatar technology is called Animgen and was used, e.g. for the Virtual Guido avatar [22]. Drawbacks of Animgen are that it is not open source but also its dependence on HamNoSys, a relatively high-level language with no transparent way to modify animations. To overcome the limitations of HamNoSys, the LIMSI institute developed Zebedee, which is based on geometric constraints and allows parametrizable scripts [23]. However, the notation may result hard to read for humans and potentially hard to realize on the animation side. Another avatar project is called Paula with several interesting results for the synthesis of fingerspelling, nonverbal components and natural pose computation [37]. Both Paula and Zebedee are not yet mature and integrated enough to be used outside their original labs. On the other extreme, researchers have used commercial animation software (e.g., VCom3D) for their experiments (e.g. [26]). Those packages allow convenient timeline-based editing of animations but are closed toward any further external development. Lastly, there are general-purpose avatars that could be made usable for signing purposes. The Greta avatar has been used

for sign language research but only at an early prototype stage [32]. The SmartBody agent is another well-known avatar technology which focuses on coverbal behaviours [36]. Both general-purpose and signing avatars lack a clean animation interface with the notable exception of EMBR [25], which introduced the animation layer to decouple behaviour specification from low-level animation parameters [24, 31].
In most signing avatar projects, the actual comprehensibility of the produced animation by deaf users has been assessed. This is particularly important because most of the experts working in this field are not native speakers of sign language. Most evaluation studies establish a pure sign language environment (instructions and supervision by a native signer) and grant a dedicated warm-up time to get used to the avatar [34, 30, 28]. For questionnaires, it has been recommended to rely on non-written content like pictograms. In terms of assessment methodology, the subjective rating of understanding by the participant him/herself turns out to be highly unreliable [28]. Instead, outside judgments by experts are taken, based on questions about the content of the communicated sentences [34, 30]. Here, mere imitation without understanding may be a problem. Also, asking dedicated questions may give part of the answer away, especially in sign language. In [27], the authors propose a multiple-choice test, where similar spatial arrangements are shown. They explain that this method may not always be feasible, especially for more complex/abstract utterances, and requires careful decisions on what to ask and how to formulate the options. A more general challenge is to define a control condition, i.e. what is the avatar signing compared against? Therefore, they suggest Signed English (SE) as a control condition. SE is an artificial language that translates the words of spoken English into a one-to-one mapping to signs. Since Signed English and Sign Language are two distinct languages, the former might be even harder to understand than the latter, we do not deem this a good option. Instead, we suggest using the comprehensibility of the human signer as the control condition.

3.2.3 The Platform As A Service

The solution is offered in terms of a Platform As A Service (PAAS). The *platform* is accessible by users through a *smartphone application* while it is available to the content producer through a RESTful API, or a web page.

The *platform* is made of Web Services, a Repository for the video messages with Sign Language and a Publisher/Subscriber (Pub/Sub) Broker.

RESTFulService – Composition

First, a web service composition is a combination of web services to create a new service [21]. The structure can be made as an *orchestration*, where a central process maintains the operation state until the end or it can be made as a *choreography*, where the data are streamed to the web service components, and the focus is on communication among services that are distributed. In a *choreographic* solution, each component knows its role in the composition, without any centralised director. In our model, the composition is made of four components, four web services (Figure 3.8) in a *choreography* fashion.

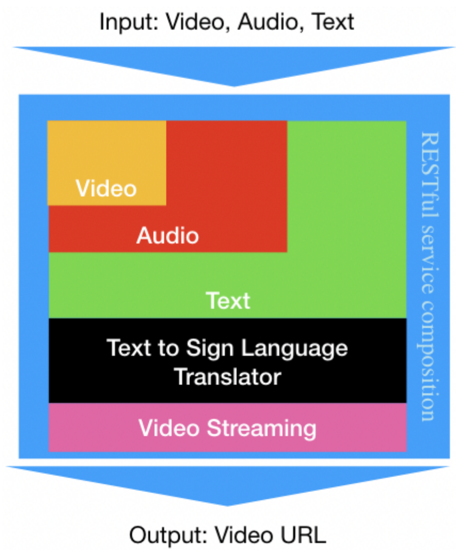


Figure 3.8: RESTful webservice.

Starting from the bottom, the last component receives as input a text stream and translates this in a video stream where an avatar performs the corresponding Sign Language, storing it in a repository and generating a unique URL to stream it. Above that, there is a web service which offers Speech to Text conversion; it takes an audio stream as input and converts it into a text stream, as output. At the top, there is a web service which offers the extraction of the audio track from a video stream; it takes a video stream as input, streaming as output its audio track. In this architecture, the Composed Web service can accept as input all the three kinds of media, and return as output an URL of the video stream, containing the performance of the translated Sign Language, the *text* message and some metadata related to its announcement, which can be used for further messages filtering. If as input it receives a *video stream*, the stream will cross all the web services until the bottom. If the input is an *audio stream*, it will skip the first web service and will flow through the other web services until the end. If it is a *text stream*, it will be routed directly to the last web service which produces the final output of our composed web service. Figure 3.9 depicts the full framework.

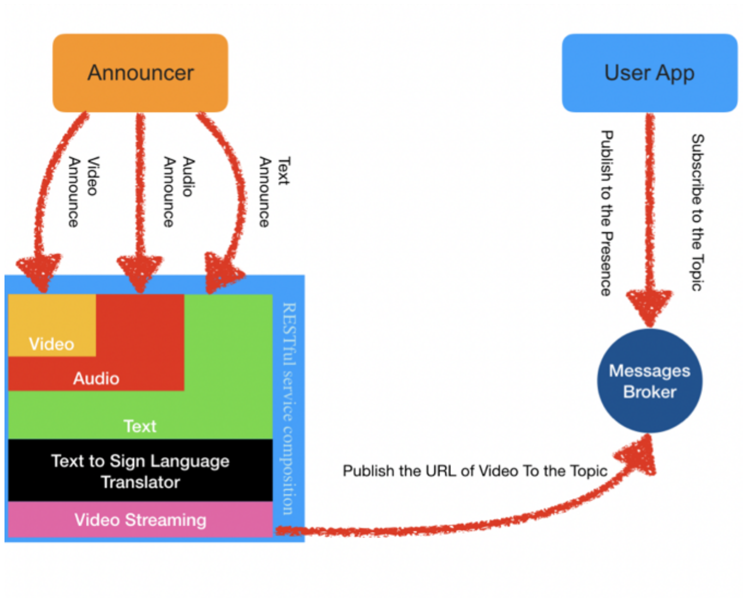


Figure 3.9: Full Framework.

Announcer

When a dispatcher of announcements (the *Announcer* in Figure 3.9) produces new messages, it can use the RESTful API Service to convert them into videos featuring the corresponding Sign Language announcements.

User

When a user wishes to receive the Sign Language translation of the messages produced by a specific Announcer, he/she can subscribe to the *topic* of interest and get it in real-time, provided that he/she is online. If not connected to the Internet, the user will receive the messages as soon as he/she gets connected, provided that the announcement is not expired yet.

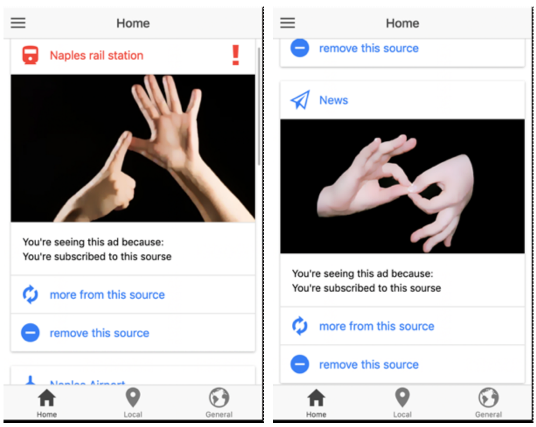
Publish-Subscribe protocol

A Publish-Subscribe (Pub/Sub) protocol is a client-server mechanism, where the clients can *publish* a message on a *topic* and *subscribe* to a *topic* to receive the related messages. The server part is called *broker* because it acts as a broker between the publishers and the subscribers. It routes all the messages to the right subscribers and performs other secondary tasks, such as verifying the identity of connected clients, managing QoS, giving a temporary persistence to the messages not received due to the lack of connection, and more.

1) *Client:* The *client*, usually has a small footprint of code and has just to know the IP address of the *broker* to which it should connect to subscribe or publish. A *topic* does not need to be defined before. It will be created the first time a *subscriber* subscribes to it or the first time a *publisher* publishes on it. A *topic* has a structure similar to a file path. For example, a topic could be defined as "\Station\Rome\Arrival". If a *client* subscribes to the topic "\Station\Rome", it will receive all and only the *messages* published as: *publish* “Message text” on the *topic* "\Station"\Rome". If a *client* subscribes to a *topic* as "\Station\Rome\#", then it will receive all the message published on the *topics* "Rome" and "Arrival".
2) *Broker:* The *user* can subscribe to the *topics* of interest so that whenever a new announcement is published on the same subscribed *topics*, he/she will receive the notification and the translated message. The *user* can unsubscribe in any moment not to receive any more notification. A *publisher/subscriber broker* will attend to this mechanism. One or more *publishers* can publish on the same *topic*, and one or more subscriber can subscribe to the same *topic*.

3.2.4 An Application of Prosign+ In Three Scenarios

In this subsection, we present three scenarios to explain the potentialities of a system using the architecture described beforehand. The scenarios are presented with an application that we designed to exploit the proposed platform. The aim is to provide deaf users with a tool to collect Geo-localized news, advertisements, and announcements, in the video, audio or text forms, and visualize them in the proper sign language. Figure 3.10 describes the home of the application. Here, the application shows a list of cards containing media translated into sign language. Each card properly contains: 1) a title indicating the category of the content, combined with one or two icons, one for representing the category and another to indicate if it is an urgent communication; 2) a video with an avatar translating the message; a brief text explaining why the content is shown; 3) a button to open a new page containing a list of contents belonging to that specific category; 4) a button to remove the source and stop receiving content from it.



(a) (b)

Figure 3.10: The home of the application showing the feed according to the user's preference.

Figure 3.11 describes the page from which the users can select Geo-localized services. By moving the map at the bottom of the page, the application displays in the upper part of the page, the categories of services available, which can be selected or deselected.

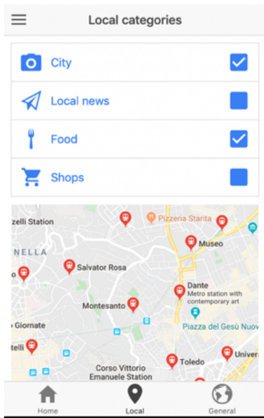


Figure 3.11: The page shows the category of content available in the map view.

Figure 3.12 shows the application page containing all the categories that cannot be geolocated. Each category contains news and announcements that can have general interest to the users and are not linked to a specific place. In this case, the users can scroll the list or filter it and select all the categories they are interested in. We exploited this application in the following scenarios.

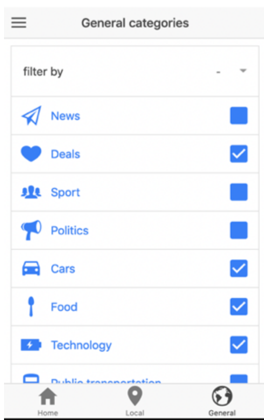


Figure 3.12: The page shows a list of categories not depending on the location view.

Scenario.1 – Strolling through a new city
The stakeholders involved are a deaf tourist visiting a city looking for points of interest, local news and activities, and managers of local commercial activities or local newspapers. The managers want to communicate their promotions and spread their news to foreign tourists, including also people speaking the sign language. So, they exploit the same content they have already created for some social networks by sending them to the ProSign+ system. The system elaborates this content by extracting the audio track from the videos, transforming audio into text and text into sign language animations. The tourist explores the zone using the map in Figure 3.11 planning her day. She wants to know local news, curiosities and find a promotion for dinner. Then, she selects the categories of food, city, and news. Starting from this moment, she receives on her smartphone app the content related to those categories.

Scenario.2 – Waiting at the train station
The stakeholders in this scenario are a traveller speaking only the sign language, the stationmaster and general news providers. Often, the stationmaster needs to announce problems related to trains or to the station. This kind of announcements is usually made by loudspeakers, so he uses the ProSign+ system to translate them to travellers in transit. The traveller goes to the rail station waiting for a train. He cannot hear the station announcements; he should pay constant attention to the information screens, and some messages could be communicated only through the loudspeakers. Moreover, while waiting, he wants to spend some time playing with his smartphone. Before starting to play, he opens the ProSign+ mobile app and goes to the map page. Here, he localizes the station and the services provided in the area. He selects the category ‘railway station’. In this way, the traveller can play with his smartphone while being notified of any announcement related to the station coming from ProSign+ app (Figure 3.10.b).

Scenario.3 – Broadcasting an important event
This scenario involves as stakeholders the general deaf population of a state and governmental authorities.

There are exceptional events of national interest that push the authorities to stop the television transmissions so that they can send a video message to the whole citizenry. During these events, sign language translators should also be present in the video message to speak to the deaf population. In case such a translator is missing, the TV editors can send the video message to ProSign+ platform, which extracts the audio track and converts it into a sign language message that can be embedded in a dedicated Web page. Deaf people can also receive a notification through the app and can watch the translated message at the top of their feed.

3.2.5 Evaluation Design

The objectives envisaged by the evaluation of the application are:

1) To assess the level of comprehension of notice, to specify any animation features that need to be improved before proceeding to the implementation of the application
2) To assess the usability of the app navigation, it should be intuitive and predictable
3) To measure the utility of the system from the point of views of deaf users and of content providers.

Method

We plan to carry on a qualitative and quantitative study involving both sign language experts and potential end-users. In particular, first, we plan to carry out a focus group with experts having the goal to get feedback about the utility of the system, its potential problems and benefits. Then, we plan an experiment with deaf users. The goal is to go deeper inside the points explored during the focus group and investigate the preferences and necessities of deaf people with respect to accessing information through an avatar or through a simple text. In the experiment, we are designing a scenario related to the rail station context. The participants will be asked to use the application, locate an alert notice and interpret it through sign language. We will measure the comprehension errors, such as problems with finger spelling, mouthing, and facial expression, the ability to navigate the interface and the satisfaction to use the system.

To get more information on the causes of errors, we will pay attention to:

- Whether the scores did not improve during a session

- Signs that are mainly dependent on the mouth will be less understood than signs that are not
- Signs that are mainly dependent on facial expression will be less well understood than signs that do not involve facial expressions.

The subjects will be tested individually in a controlled environment. They will be seated in a swivel chair with the device running the application. All the experiments will be recorded, in a way that the team of researchers can study each case painstakingly. After the last task, subjects will be asked for general comments on the gestures performed by an avatar and to rate their experience through a questionnaire.

3.2.6 Conclusion

In this work, we presented a new architecture designed for systems aimed at making multimedia content accessible to deaf people. It allows translating video, audio and text content into animation in the form of an "avatar" speaking in sign language. Moreover, we also designed an application able to exploit the platform service making the content pervasive and always available to deaf users. Sign language animation through an avatar is a potentially promising method for making services more accessible and user-friendly for deaf people. We also outlined the evaluation method with which we aim to measure the users' understanding and the utility of the system. As a concluding remark, we would like to stress that the involvement of deaf people is crucial not only for defining use cases and for evaluation but, even more so, for creating animations and developing animation methods. Hence, we argue that a stronger scientific involvement of deaf individuals should be encouraged, possibly applying a co-design approach.

3.3 AI at the Edge for Sign Language Learning Support

3.3.1 Introduction

The United Nations Convention on the Rights of Persons with Disabilities recognizes and promotes the use of sign languages, establishing that sign languages are equal in status to spoken languages [38]. It recommends states parties to facilitate the learning of sign language and promote the linguistic identity of the deaf community. One of the major challenges is to raise awareness of the importance of providing sign language learning support. Most deaf children are born to hearing parents and are not exposed to sign language until school age, missing a vital window of time for language acquisition, which corresponds to the first three years of life. As a result, damage due to acoustic sensory deprivation may occur and the associated effects will not only affect the learning process of the word but will also negatively affect the global perceptual mechanism and consequently the behavior of the subject [1]. Early access to sign language and services in sign language, including quality education available in sign language, is recognized to be vital to the growth and development of the deaf individual and critical to the achievement of personal goals. While special education promotes the integration of children with and without disabilities as the least restrictive environment (LRE), Deaf cultural perspective holds that a language-rich environment is best achieved through sign language. Therefore, an LRE for deaf children involves access to information through sign language and interaction with peers.

The goal of our research has been to provide IT support to the cumbersome process of sign language learning, relying on the use of deep learning techniques.

In the first part of the section, we present a recognition system that uses Convolutional Neural Networks (CNN), which was developed at the HCI-UsE laboratory of the University of Salerno to provide sign language learners with an advantageous interactive experience. The main objectives of the experiments we conducted on the system were the study of user's ability to efficiently and effectively reproduce the letters submitted to him/her. The results obtained from the tests carried out were positive about user learning. Most participants developed, during the various experimental sessions, a higher level of familiarity with the sign language, leading to an increase in the accuracy detected by the neural network. The system has proven to be a valid teaching tool for children or adults with no experience with this language, who may start learning the sign language alphabet. A natural step beyond has been to conceive a similar system that would deal with the complexities of the whole language of signs, as described in the second part of the section.

Sign languages are natural languages with their own grammar and lexicon. This raised some new challenges when conceiving the new system, not only because of the complexities of these signs, the high interclass similarities, the large interclass variation, and constant finger occlusion, but also for the computational resources required and the low-latency requirements to achieve answers in near real-time.

In the research presented in Section 3.1 we experienced the high acceptance gained by mobile applications within the deaf community as valid communication means. Therefore, we also aimed at delivering a deep learning solution on mobile smart devices to facilitate sign language training activities during everyday life. This gave rise to the idea of a Fog-Computing Architecture, which could provide additional computational resources to the locally connected end-user devices.

Structure of the section. Some related work is presented in Subsection 3.3.2. Subsection 3.3.3 presents the system we realized to recognize the American Sign Language (ASL) alphabet, and Subsection 3.3.4 describes the experiment performed and summarizes the results. Subsection 3.3.5 proposes the *fog-computing* architecture underlying the complete ASL learning support. Subsection 3.3.6 concludes the section.

3.3.2 Related Work

In general, the ASL alphabet recognition task is formulated as two subtasks: feature extraction and subsequent classification. Researchers have been using different methods to extract discriminative features and create powerful classifiers.

Pugeault and Bowden apply Gabor filters to extract features from both color and depth images at 4 different scales [39]. Then a multiclass random forest classifier is used to recognize the 24 static ASL alphabet signs. They report a 49% recognition rate in the leave-one-out experiment. Half of the signs could not be recognized, showing that Gabor filters cannot capture enough discriminative information for differentiating different signs. Also, Wang et al. used color and depth images for recognition [40]. They proposed a Superpixel Earth Mover's Distance (SPEMD) metric, and they reported a 75.8% recognition rate on the benchmark dataset. Using a Support Vector

Machine (SVM) classifier, Maqueda et al. gained 83.7% leave-one-out accuracy on the benchmark dataset [41]. Features were extracted from only depth images on randomly positioned line segments and a random forest was used for classification, with 81.1% accuracy [42]. Some studies attempted to exploit the 3D information embedded in the depth images (3D approach) [43, 44, 45, 46, 47]. Such 3D approaches are promising to achieve better performance than image representations due to the extra dimension. However, the 3D point cloud obtained from the depth image is sparse at the regions with large gradients and absent at the occluded areas, which affects the overall performance.

Due to the articulated structure of hands, some studies implemented a hand part segmentation step before the gesture recognition (bottom-up approach). Keskin et al. extracted depth comparison features from depth images following the method proposed in [49] and fed them into a per-pixel random forest classifier [48]. They reported their leave-one-out recognition rate as 84.3% on the benchmark dataset. This classifier was trained using synthetic depth images which have the parts' ground truth of a hand. To generate more realistic training data, a colored latex glove was employed by Dong et al., resulting in a 70% recognition rate on the benchmark dataset [50]. One of the major drawbacks for those bottom-up approaches is that the sign recognition performance is highly dependent upon the result of the hand part segmentation, and it is challenging to improve the performance of the hand part segmentation because of the high complexities and constant occlusions.

Recently, deep learning methods, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have demonstrated their extraordinary performance in various classification and recognition tasks.

From Pose to Gesture

The Sign language (SL), seen as a crucial language for hearing and speech impaired people, can be considered as the most grammatically structured gestural communication [51]. A Gesture can be seen as a continuous sequence of poses. Marcon et al. use Hidden Markov Models (HMMs), trained on different gestures, to identify a set of key postures and to classify their sequences over a set of possible actions [52]. Cui et al. in [53] proposed a CNN with temporal convolution and pooling for spatiotemporal representation learning from a video, and an RNN with a bidirectional Long Short Term Memory (LSTM) for the mapping of feature sequences to sequences of annotations. Madhuri et al. in [54], present a real-time vision-based system for recognizing finger spelling continuous Sign Language (SL) using a single camera to track user's unadorned hands. The goal is to help hearing or speech impaired people to communicate with people who do not know SL. Although facial expressions add relevant information to the emotional aspect of the sign, in [54], they are not considered since their analysis complicates the problem. They only focus on translating a one-handed sign of representation of alphabets (A-Z) and numbers (0-9).

Some researchers have proposed techniques to detect predefined signs from a continuous video stream (sign spotting), while others have handled the classification of isolated gestures into the correct category. Continuous SL recognition, instead, deals with transcribing videos of SL sentences into ordered sequences of annotations, possibly in real-time. Cui et al. in a recent work [55], focus on continuous SL recognition on videos, where learning the spatiotemporal representations as well as their temporal matching for the labels is crucial. They claim their framework based on recurrent convolutional neural networks shows a superior capability of learning temporal dependencies compared to HMMs.

Most related researches are using Markov models (HMM) or LSTM networks. Although both those networks take into account the spatiotemporal aspect of Gesture, they face different issues. The HMM solution has a problem with a more prolonged gesture, while the LSTM requires powerful computational resources, and it is slow on a small device like a smartphone.

Panzner e Cimiano in [56] compare a purely generative model based on Hidden Markov Models to a discriminatively trained recurrent LSTM network in terms of their properties and their suitability to learn and represent models of actions. They highlight the limitation of temporal context with HMM and the need for an extraordinary computational resource with LSTM networks.

3.3.3 Learning ASL Alphabet Relying on CNN Support

The proposed system has as its main objective to be a bridge of communication between deaf people and today's society. The system offers

a chance to those who are inexperienced in this language, a simple and interactive learning method.
Our classification of letters is carried out using a convolutional neural network (CNN or ConvNet). CNNs are machine learning algorithms that have been incredibly successful in managing a variety of activities related to video and image processing.
Our network, after being trained, allows the recognition of gestures, through the use of images made by the user. During the execution of the application, the user will have to take photos containing one of the 24 static letters of the American alphabet. After this, the classification of the latter will be carried out, i.e. the neural network will define the letter corresponding to the gesture made by the user. Finally, the number of correctly executed letters and the number of incorrect ones will be shown to the user, with the consequent level of learning developed during the use of the system.

Used Methods

1) *Transfer Learning:* is a machine learning technique where models are trained on (usually) larger data sets and refactored to fit more specific or niche data. This is done by recycling a portion of the weights from the pre-trained model and reinitializing or otherwise altering weights at shallower layers. The most basic example of this would be a fully trained network whose final classification layer weights have been reinitialized to be able to classify some new set of data. The primary benefits of such a technique are its less demanding time and data requirements. However, the challenge in transfer learning stems from the differences between the original data used to train and the new data being classified. Larger differences in those data sets often require re-initializing or increasing learning rates for deeper layers in the net.

2) *AlexNet*: Specifically, we used AlexNet. AlexNet is a Convolutional network that has had a great impact in the field of machine learning, designed by Alex Krizhevsky. The network acquires an image as input and generates a label for the object in the image along with the probabilities for each category of objects. AlexNet consists of 8 layers: the first 5 are Convolutional layers, and 3 layers are fully connected

and can classify images into 1000 categories of objects, such as keyboard, mouse, pencil and many animals. AlexNet is developed in Matlab.

The System

In order to obtain images of the user signing in real-time, we created a desk application that accesses a native camera. Image capture rate was a massive problem we struggled with.

Our desk application sends images to the neural network one by one. Each time, the net classifies the image and presents probabilities for each letter.

Our system, as we can see in Figure 3.13, presents on the left side the GIF of the letter to be reproduced by the user, and the consequent reproduction of the same. On the right, instead, we see the image acquired by the camera. When the user feels confident about the sign being made, he/she can take a photo. The letters are randomly generated (from 1 to 24) by the system. Once the photo is taken, it will be shown on the left side of the page. Moreover, it will be possible to change the photo just taken. Moreover, if the user is not able to reproduce the requested letter, he/she may decide to change it.

Figure 3.13: Core Page of the system.

Furthermore, our system presents another page containing the results of the tests performed by the user. In Figure 3.14 we can see a table divided into three columns: in the first one there are the letters that the user made during the test; in the second column the relative results are present for each letter, and that is how the neural network has classified the images sent to it; in the last column we can see a graphical response of the single test carried out. Marking the wrong result with a red sticker and a positive response with a green one. Finally, at the bottom of the page, we find the level of learning detected during the test. This result comes from the classification of all images made by the user during the experiment. The level of learning coming from the network is not based on an average of the correct or wrong letters, but the neural network produces this value based on the skill and precision that the user has used in reproducing the letters. A high percentage of learning assumes, therefore, that the user had better precision than the other users.

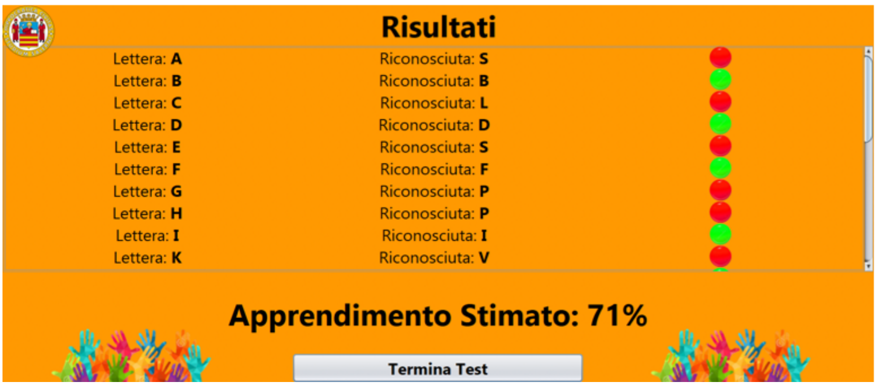


Figure 3.14: Result page of the system.

Dataset and Features

ASL Alphabet was the most dominant DataSet in the training of our network. It contains about 87,000 images of various kinds and depths. The dataset is divided into two directories.

Figure 3.15: Dataset example. Left: ASL Alphabet. Righ: Sign Language and Static gesture recognition.

The second dataset we used within the system, Sign Language and Static gesture recognition, was less predominant for training. It contains a very limited number of data and, consequently, it is inefficient for training an effective network. The total number of data present is 1,687.
With the use of those datasets, we have created a new dataset including all the data, about 73,488 images, and we have decomposed it to develop two happy directories: the first including 80% of the images, to perform training, the second, including the remaining 20%, for tests.
The network training was carried out on 58,944 images of various kinds. The images have been reduced in size to be consistent with the specifications of the network used. Each image was resized to 227x227 pixels.

Train and Test

In order to compare our results with state of the art, we considered two metrics, accuracy and confusion matrix.

1) *Accuracy and loss:* Accuracy in the validation set is the most popular criterion in the literature, defined as the percentage of correctly classified examples. Splitting the dataset into two for training and validation, we achieved a high validation accuracy, namely 99.96% on the alphabet gestures. The running time of the network train was 491

minutes and 24 seconds, performing 13,800 iterations. The network training has led, therefore, to the development of an efficient network, with very high accuracy in classifying the images, so that we can build a reliable sign recognition system.

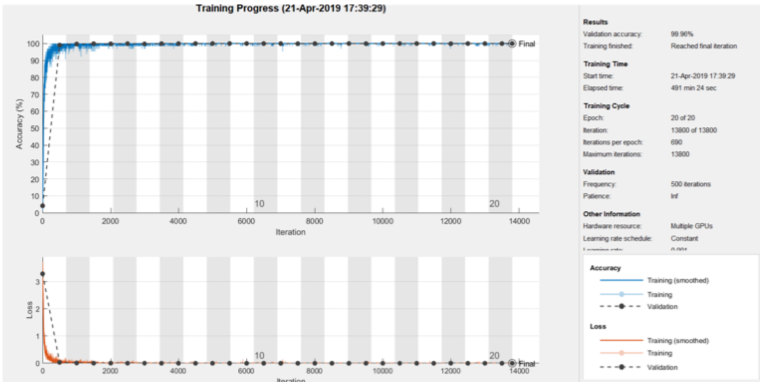


Figure 3.16: Train Convolutional Neural Network

2) *Confusion Matrix:* Additionally, we used a confusion matrix, which is a specific table layout that allows visualization of the performance of the classification model by class. This allowed us to evaluate which letters are the most misclassified and draw insights for future improvement.

	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
A	613	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
B	2	529	0	0	0	0	0	0	0	0	0	14	0	0	0	0	1	0	0	68	0	0	0	0
C	0	0	614	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	5	608	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	1	0	0	585	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0
F	0	0	0	0	0	613	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
G	0	0	0	0	0	0	602	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	599	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	614	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	611	0	0	0	0	0	0	0	0	0	3	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	614	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	614	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	614	0	0	0	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	614	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	614	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	614	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	614	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	614	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	600	0	0	0	0	0
U	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	613	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	614	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	614	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	614	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	614

Figure 3.17: Test Convolution Neural Network – Confusion Matrix.

3.3.4 Evaluation

This subsection is dedicated to the experiment performed on the developed system and to the results obtained from it.

Experiment

The main objective of our experiment was to carry out a study for the learning of sign language. Each participant used the application performing the gesture of the 24 letters of the American alphabet. The 12 participants of the experiment were divided into two groups: 5 participants were tested on the application once a day, for three days. This made it possible to detect improvements in user learning.
To accomplish that, two different metrics were chosen:

- Accuracy: percentage of gestures performed correctly;
- User satisfaction: the subjective impression of the user (SUS).

We focused on accuracy rates as a basic metric as we can immediately realize how many errors the user has committed. In addition to empirical data, it is also important to collect subjective information. As the opinions of users are very important to define any future developments that can improve our application.
The experiment started with an experimenter explaining the task and the system to the participants. It was explained to them that the purpose of the experiment was to test and evaluate this new intelligent sign language learning system, not their abilities. The participants were asked to replicate the gesture reproduced by the GIF, trying to simulate it as precisely as possible. The participants were positioned less than a meter from the computer on which the application was launched, comfortably seated in an armchair to create as much as possible a natural environment to be used.
At the end of the experiment, the participants were asked to complete a SUS (System Usability Scale) questionnaire [57]. It consisted of 10 statements to which the participant assigned a score on a scale from 1 (strongly discouraged) to 5 (strongly agreed). The final score of the SUS varies from 0 to 100. A higher score indicates greater usability by the participant. Furthermore, after the experiment, comments and suggestions were also collected from the participants.

The experiments were evaluated according to the Within-subject design, i.e., all the subjects participating in the experiment were tested in each condition (on each letter of the alphabet). The order of the letters to be executed was counterbalanced among the participants, to avoid problems of order effects.

Results

Each experimental session was allocated a 10 minutes time slot. All participants completed the experiment. To assess the accuracy of the method used for each participant, we recorded the level of learning achieved. As shown in Figure 3.18 the participants obtained an accuracy ranging between 17% and 65% (M = 33.9%, DS = 16.3). The figure refers to the percentage obtained after only one use of the system by the 12 participants.

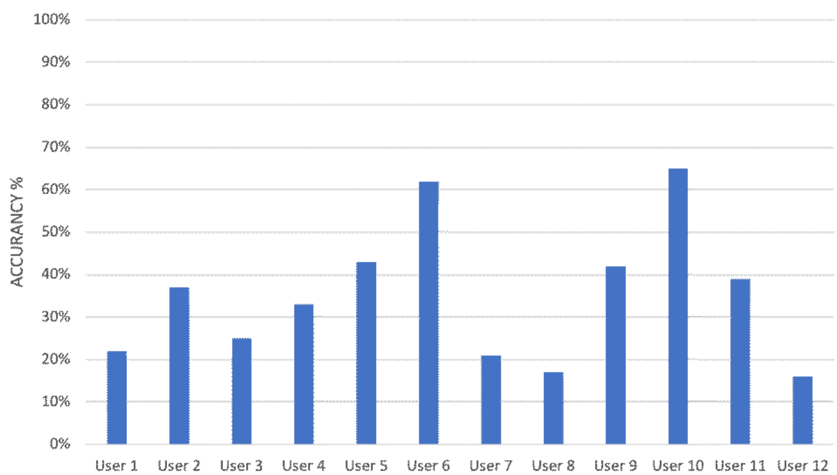


Figure 3.18: User learning achievements after a single experimental session.

As for the results gained over several days, performed by 5 of the participants, Figure 3.19 shows how the accuracy increases through the different sessions.

This result was very useful as the basic hypothesis of the experiment was verified, namely that the system can be used as a support for learning sign language.

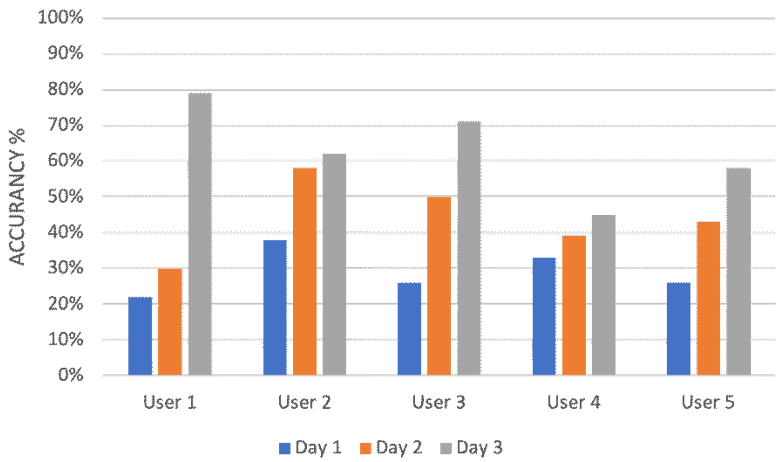


Figure 3.19: Learning progress through the three experimental sessions.

The average SUS score was 73.7 (SD = 29) for our system. All participants said that they find our application very stimulating for learning sign language. The use of the application by the participants was considered a choice that is always preferable to self-learning through video or other applications, above all because it is possible to have immediate feedback on one's level of learning. Moreover, since our application provides feedback for every gesture made, the user can also see his/her mistakes and become aware of the gesture to be improved.

3.3.5 Artificial Intelligence at The Edge

Sign languages are made of sequences of signs. Although, to use regular neural network for the machine learning of sequential data, as videos (a sequence of frames), could be possible feeding the CNN with entire sequence, the constraint of a fixed size of input could be an unacceptable limit. What we need, instead, is to feed an arbitrary length of images sequence, one element per time step and a neural network which has some kind of memory to remember events happened many step times in the past. This behavior is best performed by RNN with LSTM models. While these kinds of network are successfully used for Sequence Learning applications, they require extraordinary computational resource.

To offer additional computational resources to the locally connected end-user devices, this section proposes a *Fog-Computing* Architecture.

The Fog-Computing paradigm

Fog-Computing is an emerging distributed computing model aimed at bringing computation close to its data sources, which can reduce the latency and cost of delivering data to a remote cloud. This feature and related advantages have been profitably exploited in different application scenarios, especially for latency sensitive and mission intensive services [58, 59]. The definition and architecture of *Fog-Computing* model [60] are briefly presented in this section, while it proposes a new specific hardware platform.

Fog-Node is the core component of *Fog-Computing* architecture. They are between the edge of the network, and Cloud resources, distributed on layers, offering connectivity and computing resources to the smart end-devices (Figure 3.20). The paradigm is that, from Edge to the Cloud, there are layers which elaborate the data, forwarding the results to the upper layer, and eventually, collaborate with other *Fog-Nodes* to distribute the processes "in press" [61] and storage [62]. Usually, the lower layer offers connectivity to the end-user device and is part of the local network.

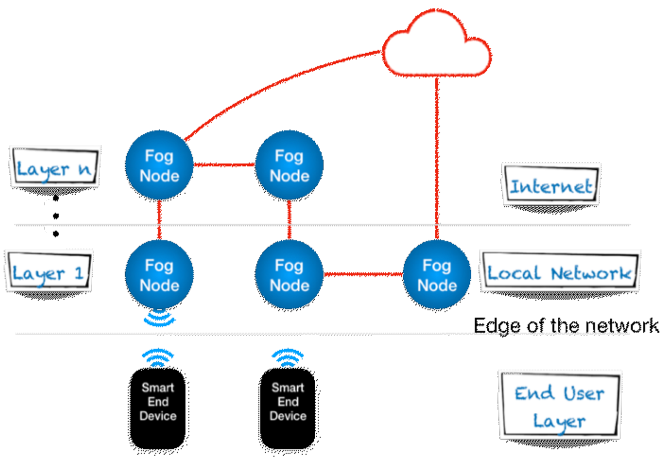


Figure 3.20: An exemple of Generic Fog-Computing Architecture.

Furthermore, the lower layer offers computation resources and limited storage to locally connected devices, with low latency and without the need of the Internet connection.

The Proposed Framework

In the Deep Neural Network (DNN) models, the large number of identical neurons, makes it natural to consider high parallelism in the computation. Actually, in 2012, the team led by Alex Krizhevsky, the creator of AlexNet, used for the first time a Graphics Processing Unit (GPU) -accelerated DNNs, winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) by a large margin [63].

The parallelism naturally fits with GPUs architecture, which provides significant speed-up over traditional CPU.

Currently, an increasing number of promising devices are available on the market, which claim to bring modern AI to the makers and to the embedded developers. We selected the new low-cost NVIDIA Jetson Nano module, which features a 128-core GPU and allows to use libraries like Python, C++, CUDA X and OpenGL.

A wide variety of deep neural network models that enable tasks like image recognition and object detection can be accelerated with support from NVIDIA CUDA® Deep Neural Network library (cuDNN) and TensorRT tm.

The cuDNN library has support for RNN, which are widely used for sequence learning in many fields.

TensorRT is a platform for high-performance deep learning inference. It contains a deep learning inference optimizer and a runtime that delivers low latency and high-throughput for deep learning inference applications. Applications based on TensorRT perform up to 40x faster than CPU-only platforms, during inference.

The idea is to optimize trained neural network models and calibrate them for lower precision with high accuracy on the Cloud data center, and finally deploy the optimized Inference Engine to the *Fog-Node*, at the edge of the network (Figure 3.21), deploying there the NVIDIA Jetson Nano module. Smart user devices, which could not have computational resources to deploy a fast enough inference engine, may demand the elaboration of image sequences to the locally networked *Fog-Node*. The optimized

Inference Engine of *Fog-Node* can then answer with low latency, and parallelism in the computation, to the end-user device requests.

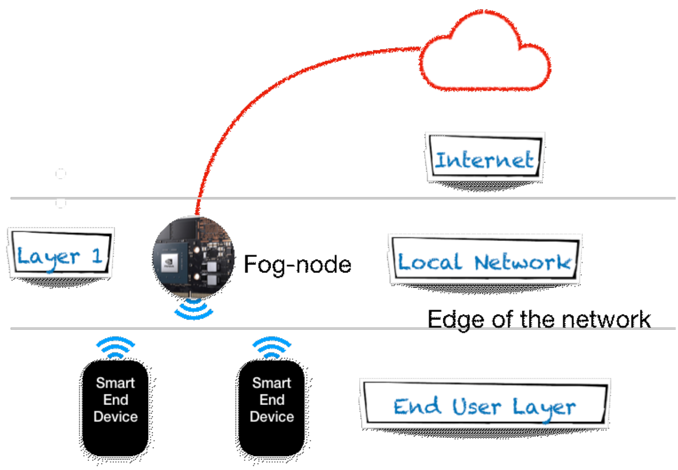


Figure 3.21: Simplified schema of the proposed framework.

Single *Fog-Node* can serve more than one end-user device, and more *Fog-Nodes* can be added whenever it needs to scale up for more computational resources.

3.3.6 Conclusions

This section proposed a simple and interactive learning method and application, which relies on the potentials of deep learning tools to offer a chance to those who are inexperienced in Sign Languages for hearing and speech impaired people. The method has been initially adopted to develop a system which supports ASL beginners during alphabet learning tasks. Calculating the accuracy of proposed system, the user's learning curve has been evaluated. The outcomes obtained from the tests have produced positive results, indicates that the solution may not only improve user's learning curve, but also encourage training, due to its efficient methodology. In order to apply a similar approach and support to learning of a complete sign language, we have adopted a new architecture that may deal with the increased complexities of sign languages, as natural languages with their own grammar and lexicon. So, to rely on more powerful computational resources while delivering the final support to learners on their mobile smart devices, our idea is to bring AI at the Edge, proposing the adoption of a modern *Fog-Computing* architecture with low cost small AI Computer hardware.

Chapter 4

UX challenges for Blind Users

The visually impaired community has been very active in recent decades with initiatives to raise awareness in society about their specific needs and encourage the adoption of innovative assistance solutions for their personal empowerment and to improve their UX. A contextual inquiry conducted in Europe revealed that some of the major concerns to those people refer to independence of living. According to recent studies, most existing remote control home automation systems UIs lack some specific usability and accessibility features that could meet the specific needs of blind users, thereby improving their UX. In the following section, an assistive mobile app is presented with a UI designed to allow autonomy of blind/visually impaired users in the everyday activity of getting dressed, thus improving their UX. The user testing activities conducted so far are described and the derived results are discussed.

4.1 Do you like my outfit? Cromnia, a mobile assistant for blind users

4.1.1 Introduction

The community of visually impaired people has been very active during the last decades with initiatives devoted to raise awareness about their specific needs in the society and encourage the adoption of any innovative assistive solutions for their personal empowerment [64, 69].

A cross-border contextual inquiry involving that community was performed as part of a recent European project aimed at defining an innovation procurement roadmap for partially sighted people [72]. The survey was initially conducted in the form of a questionnaire, to be distributed on a

European scale. Experts in the domain of low vision took part in the study, ranging from clinicians, to vocational rehabilitation professionals, to operators in the social care field. Then, a number of focus groups were also held in different partner countries, involving partially sighted and blind people. Among the needs that emerged, one especially inspired the present work, namely “maintaining appropriate environmental conditions” and “raising independence in personal care”. In a recent study investigating the needs of visually impaired people in the area of home automation for independent living, Leporini and Buzzi argue that most existing remote control home automation systems are generally targeted to people with physical/cognitive disabilities but miss certain specific usability and accessibility features which are necessary for blind users [70]. The purpose of our study was to design some assistive solution that would provide greater autonomy to a blind person, in everyday activities. A problem domain analysis was therefore performed to address those needs through some suitable mobile application, properly taking into account accessibility and usability requirements. A participative process started since the early stages of analysis and design, involving a group of 4 blind people from the Italian Blind Union, who volunteered to actively collaborate with the usability design team. Among the daily tasks for which higher autonomy is expected, getting dressed with a proper color match of clothes turned out to be a major concern for those people, who usually get help from their guardian. Recently, some assistive devices have been put on the market to address that need. A support to independent living is offered by OrCam MyEye, a lightweight text/color/face recognition device which can be mounted on a pair of glasses to help vision impaired/blind users in everyday activities, including clothes selection [73]. Another commercial tool is Colorino, a device which helps user with clothes selection, with washing procedure and with color combination, being able to distinguish more than 150 colors and tones, as well as shades of color [74]. However, the idea to purchase a specific tool was neglected by most of our interviewees, who observed that many people could not even afford them. Moreover, they insisted that even cheaper devices could be annoying, requiring users to bring them everywhere and check batteries status or regularly recharge them. Some software solutions, including mobile apps, are available for color and pattern detection but very few were especially conceived for blind

or visually impaired users [65, 66, 67, 71]. The mobile application BeMyEyes is one exception but it simply relies on the use of a video call to require attendance by some remote guardian [68].
As a result of our investigation, a mobile application, named Cromnia, was developed, which allows users to recognize colors, patterns and color combinations, also considering the brightness level of the surrounding environment. The latter functionality has a twofold usefulness. It can be used to make sure that the scanned clothing is sufficiently illuminated but also to let the user gain awareness of the brightness of the place he/she is in.
Encouraging results were drawn from the experimental user testing, which involved a group of 6 blind people and a group of 6 partially sighted people.
Structure of the section. Subsection 4.1.2 summarizes the problem domain analysis activities and the resulting functional and non-functional requirements. Subsection 4.1.3 introduces the main features of the *Cromnia* app. Subsection 4.1.4 describes the usability testing activities. Subsection 4.1.5 concludes the section.

4.1.2 Requirement Analysis

The problem domain analysis allowed us to get an insight into the most common issues affecting autonomy of blind people. The interviews we conducted with 10 blind people also aimed at understanding what are the problems that a visually impaired person encounters during his/her everyday activities when he/she is alone with respect to the case when he/she can rely on the sight of a companion/guardian. We found out that more meaningful responses could be achieved managing each interview in the form of an informal 'cognitive meeting', starting from more general questions and then arriving at the specific questions of interest for the analysis activity. Structuring the interview in that way we found a significant improvement in the clarity and response time of respondents, getting information directly related to the accessibility of certain applications, without making any explicit questions. For each interview an Mp3 file was generated and then cut using iMovie, in order to retain only the information considered useful for the analysis. Finally, the adequately analyzed information was reported as a form of 'Guiding Activities' in the Trello management system, in order to be easily shared with the design team. An empowerment-driven analysis method was adopted to elicit UX requirements, as illustrated in [16, 17]. The method, also explained in the previous chapter and depicted in Figure 3.1, is a transformative process which starts from a contextual investigation meant to understand users, their behavior and capacities within a given community, and to identify potential improvements in their life quality, which are expressed in terms of human needs. The key factors which may contribute to the empowerment of the target users with respect to the identified needs are considered in order to formulate clear users' empowerment goals. The goals entail a set of UX requirements which may be iteratively modified/extended and ultimately drive usability designers.

Among the daily tasks for which higher autonomy is expected, getting dressed with a proper color match of clothes turned out to

be one of the major concerns for those people, who usually rely on their guardian's advice.

As a result of the analysis phase the following functional requirements were established:

- Ambient light detection;
- Framed color detection;
- Framed texture detection;
- Appropriate color combination.

The associated non-functional requirements, also related to usability and accessibility, were:

- Full integration with the device voice synthetizing tool;
- Simple and immediate notification system;
- Adaptive fonts and label size with respect to the type of disability studied;
- Real-time response.

4.1.3 The Cromnia App

Cromnia is a mobile app designed and developed for Apple IoS platform, which is the most widely used by the community of blind/visually impaired people. The graphical interface design took into account general usability guidelines and was iteratively assessed by an automatic usability evaluation

tool [19]. Besides, Apple best practices for the design of accessible applications for blind users were considered so as to ensure consistency with consolidated interaction modalities, e.g., integrating Voiceover and implementing accessibility rules, ensuring ease of use by the end user. The application has a very simple interface. *Cromnia* consists in a single interface that looks like the interface of iOS systems stock camera.
Four different recognition modalities can be set when the app is accessed, namely:

- The *viewfinder* modality for continuous real-time recognition of colors (Figure 4.1.a)
- The *brightness detector*, to detect the brightness level of the environment (Figure 4.1.b)
- The *scanner*, to recognize the color and the texture of the cloth in the frame (Figure 4.1.c)
- The *color combination analyzer*, to verify the appropriateness of color matching between the last two recognized clothes. (Figure 4.1.d)

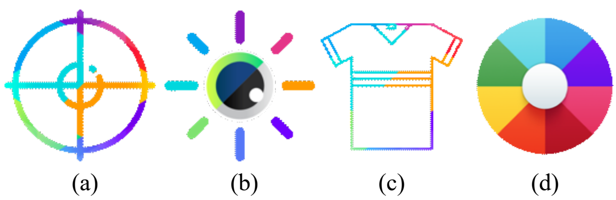


Figure 4.1.: Cromnia recognition modalities.

At the top there is the information related to the framed clothes, which is also communicated through voice synthesis. In spite of a number of existing color recognition algorithms, a simplifying algorithm was developed to associate one color to an interval of RGB values (e.g., red is the color associated from RGB [R=255, G=064, B=064] to RGB [R=161, G=031, B=018]). Such a simplification was necessary to address the specific needs of visually impaired people.
As for the texture recognition feature, it relies on the use of a convolutional neural network. We adopted the TensorFlow open source machine learning

framework to build an initial neural network creating specific datasets of textures and then training the network up to 10000 steps. As a result of the training phase, the neural network is capable of recognizing simple pattern like vertical stripes, horizontal stripes, squares, polka dots, rhombs, triangles and solid colors and also complex patterns like drawings. Finally, the surrounding brightness is calculated using a real-time analysis algorithm of grayscale obtained by converting the camera image. The combination of these analyses is then communicated, through the use of speech synthesis, with a completely accessible notification system to the user.
To illustrate the use of the application, let us consider the following scenario.
Pasquale, a native blind 35 man, has to dress up for an appointment with a friend he wishes to impress for his outfit. He uses Cromnia to decide what to wear. He puts a set of possible clothes on his bed and starts using the app. He detects the switch button through the VoiceOver feature and enters the viewfinder modality to scan the clothes. A vocal assistant informs Pasquale about the color of the current target as he directs the camera on the clothes. He checks for brightness in the room and discovers it is sufficient to get a reliable response. Then he switches to the scanner mode and gets information about the color and pattern of each cloth. He presses the color combination button and the assistant informs him that the last two scanned clothes form a bad color matching, so Pasquale tries a new one, and this time he gets approval.

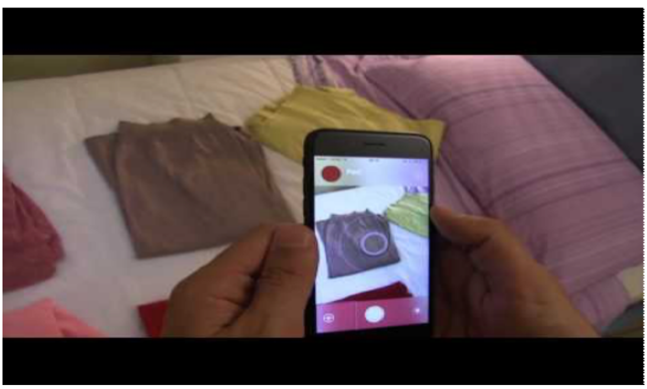


Figure 4.2: Scanning clothes with Cromnia.

4.1.4 The Cromnia App

A comparative user study was conducted to evaluate usability of *Cromnia* with respect to existing day-to-day methods adopted to get dressed. We were especially interested in evaluating efficiency as well as degree of acceptance and empowerment perception.

We recruited 12 visually impaired participants, 10 males, from the Italian Blind Union. Participants' age ranged from 27 to 72 (M= 45.8, SD=11.7), half of them being totally blind, and the others partially sighted. All the participants knew how to use a smartphone. Three participants had never used a smartphone as an aid. All the participants were volunteers. Figure 4.3 illustrates some participants performing their tasks.

The experimental tasks focused on

- Color and texture recognition from one cloth;
- Checking the combination of two items of clothing.

To perform the experiment, the zonal section of Italian Blind Union provided one of their laboratories. The laboratory was chosen in order to have an optimal light condition for the type of tests to be performed. The laboratory was equipped in order to recreate the typical home environment. The equipment used consisted of:

- 1 device (Apple iPhone 8) with Cromnia app preinstalled
- 1 chronometer to measure task completion time
- 1 empty chest of drawers in which participants must arrange clothes provided
- 1 basket to fill with clothes during the experiment to complete the selections
- 1 set of different clothing, consisting of 18 different garments:
    - 6 t-shirts
    - 6 pullovers
    - 6 pants

The evaluation was set up as a within-subjects design where all participants performed the required tasks with and without the use of *Cromnia*. The experiment started with a presentation session, explaining the purpose of the experiment to all participants. Before starting the experiment, the participants were asked to respond to a demographic questionnaire asked to sign a form for the treatment of data provided during the experiment. Then the participants were introduced individually to the laboratory environment, thus enabling them to become familiar with it. Participants were then shown the app *Cromnia* on the laboratory device, thus enabling them to become familiar with the equipment and make a quick 10 minutes training phase with regard to the use of the app. The training included:

- Unlocking the laboratory device
- Exploring the home interface of the laboratory device unlocked (identification of individual applications present by default on the main interface, until the discovery of the *Cromnia* icon, located in the lower right corner)
- Opening the application and listening to the app tutorial, which runs automatically with the first access
- Exploring the interface of the App *Cromnia*
- Checking for brightness level in the room
- Finding a label that reproduces the pattern and color of the garment detected (located in upper and central area of the interface)
- Finding the button designed for reading the clothing item (button located at the bottom center of the application)
- Finding the button for switching to viewfinder mode for continuous and exclusive reading of the framed color (located in the lower left corner)
- Acquisition of colors and patterns of clothing. For this testing phase all participants performed their tasks with the same set of clothes to train:
    - 1 yellow t-shirt with a black drawing
    - 1 black t-shirt with a white drawing
    - 1 violet pullover with white yellow stripes

Figure 4.3: Participants performing the experimental tasks.

Then the set of clothes used for the experiment were described to participants, who were given the opportunity to become familiar with them. We were careful to provide any necessary information about the characteristics of the experimental clothing. Finally, participants were asked to put into drawers the clothing items in the order they preferred provided that they kept in mind their positions.
One of the task was: *find 3 clothes which are communicated by the investigato*r. The task was repeated 6 times by each participant with random clothing. The tasks were divided into 3 repetitions with the *Cromnia* method and 3 repetitions with the day-to-day method. For each series of repetitions, a different set of clothes was used, so that the choice of the set could not influence the results. To further reduce biases, 2 sets of different clothes, each consisting of 4 t-shirts, 4 pullovers and 4 pants, chosen out of the totality of the clothing available for the experiment.

	T-shirt - M	**Pullover - N**	**Pant - P**
1	Yellow t-shirt with white stripes	Red solid pullover	Light blue solid jeans
2	Black t-shirt with white drawing	Violet solid pullover	Dark blue solid jeans
3	Blue t-shirt with a solid color	Blue solid pullover	Black solid pants
4	Blue t-shirt with red drawing	Black solid pullover	Brown solid pants
5	White t-shirt with violet drawing	Green solid pullover	Gray solid pants

6	White t-shirt with blue stripes	Green pullover with gray drawing	Dark blue jeans with white drawing

	T-shirt - M	**Pullover - N**
T-shirt	M1, M2, M3, M4	M3, M4, M5, M6
Pullover	N1, N2, N3, N4	N3, N4, N5, N6
Pants	P1, P2, P3, P4	P3, P4, P5, P6

To facilitate the balancing of repetitions and prevent their order might influence the results, participants were divided into two blocks, depending on the order of repetitions:

- Block A: Participants who make first 3 repetitions with the day-to-day method and then 3 repetitions with the *Cromnia* method.
- Block B: Participants who make first 3 repetitions with the *Cromnia* method and then 3 repetitions with the day-to-day method.

	Block	**Rep.1**	**Rep.2**	**Rep.3**	**Rep.4**	**Rep.5**	**Rep.6**
1	A	M1 N1 P4	M4 N3 P2	M2 N4 P4	M5 N1 P4	M1 N5 P6	M3 N3 P3
2	B	M4 N3 P1	M1 N2 P2	M2 N3 P4	M5 N6 P6	M6 N4 P3	M3 N6 P4
3	A	M3 N4 P3	M2 N2 P1	M3 N4 P2	M4 N6 P4	M6 N4 P6	M5 N6 P4
4	B	M1 N2 P4	M3 N4 P2	M2 N1 P1	M4 N3 P6	M3 N4 P3	M5 N3 P4
5	A	M2 N2 P2	M3 N4 P1	M4 N1 P2	M5 N5 P3	M3 N3 P4	M6 N5 P3
6	B	M4 N2 P3	M2 N4 P1	M4 N1 P4	M3 N5 P5	M6 N6 P4	M4 N4 P5
7	A	M1 N2 P3	M2 N4 P1	M1 N1 P4	M3 N5 P5	M6 N6 P4	M4 N4 P5
8	B	M4 N1 P1	M2 N3 P3	M3 N2 P4	M5 N5 P6	M6 N4 P4	M3 N3 P5
9	A	M1 N1 P2	M4 N3 P3	M3 N4 P4	M4 N6 P3	M3 N5 P4	M5 N6 P5
10	B	M2 N4 P4	M3 N3 P1	M4 N1 P2	M6 N4 P6	M3 N3 P3	M6 N6 P4
11	A	M2 N1 P2	M1 N4 P4	M4 N3 P2	M6 N5 P4	M3 N6 P5	M5 N4 P4
12	B	M1 N1 P4	M3 N3 P2	M4 N4 P1	M6 N5 P6	M3 N3 P5	M5 N5 P4

At the beginning of each repetition the experimenter would communicate the set of three garments to be selected for that exact repetition. Time count started immediately after the communication. A repetition was considered complete, and the timing was stopped, when there were exactly 3 clothing items in the basket. Of course, the 3 selected items would not necessarily match with the 3 clothes communicated. The task was considered completed when all the repetitions were completed. The order in which participants selected the clothes was considered irrelevant. At the end of each repetition the participant put back the clothes into the chest, except for the third and fourth repetitions when the set of clothes was changed. During the execution of a repetition the participant had the chance to ask the experimenter to repeat the clothes he was supposed to find.
For each participant, completion times have been reported in terms of seconds with relative accuracy (whenever a garment selected does not match with the garment communicated, an error is assigned to the repetition), for each repetition, separating them according to the system used for the selection of the clothes (*Cromnia* or traditional method). After the experiment the participants were asked to respond to a SUS questionnaire referring to the use of *Cromnia*.
We applied the statistical software Anova2 tests to all observed measures and conducted parametric statistical analysis. To assess efficiency of the method adopted, we measured the mean time used to complete the task across the repetitions. To assess accuracy of the used method, for each participant we registered the number of correct selections during task completion. Any participant could make at most 9 errors per series of repetitions with a specific method.
For the completion time, results grouped by method and type of disability are shown in Figure 4.4. As it can be seen, for both totally blind participants and partially sighted participants the Cromnia method was faster (217.0 seconds vs 247.3 seconds for totally blind participants, 163.8 seconds vs 172.5 seconds for partially sighted participants). The main effect of type of disability on task completion time was statistically significant ($F_{1,10} = 7.519$, $p < 0.05$). The effect of method was also high statistically significant ($F_{1,10} = 11.884$, $p < 0.01$).

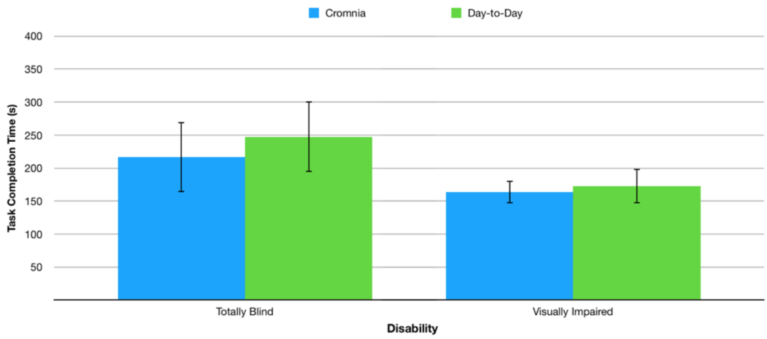


Figure 4.4: Task completion time for both methods, across the totally blind and the partially sighted participants. Error bars show standard deviation.

For the number of errors, results grouped by method and type of disability are shown in Figure 4.5. As it can be seen, for both groups of totally blind and partially sighted participants the Cromnia method was more accurate (0.17 errors vs 2.33 errors for totally blind participants, 0.00 errors vs 1.17 errors for partially sighted participants).
The grand mean for number of errors for participant was 0.92 errors. Cromnia was the most accurate at 0.08 errors, while the day-to-day method was the less accurate at 1.75 errors. The main effect of type of disability on number of errors was statistically significant ($F1,10 = 3.265$, $p < 0.05$). The effect of the method was also high statistically significant ($F1,10 = 34.483$, $p < 0.0001$). There was a not significant effect for disability interaction effect ($F1,10 = 3.103$).

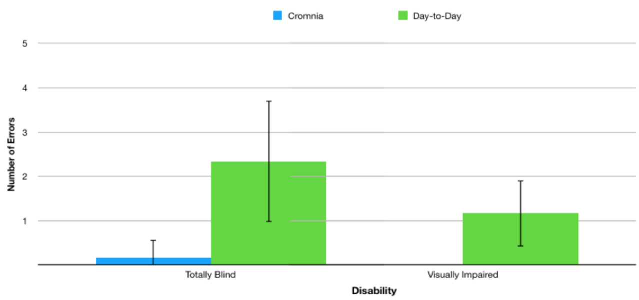


Figure 4.5: Number of errors for both methods, across the totally blind and the partially sighted participants. Error bars show standard deviation.

The average SUS score was 84.46 for the *Cromnia* method (SD=14.20). All the totally blind participants said they prefer the *Cromnia* method instead of the day-to-day method, while only half of the partially sighted participants said they prefer the *Cromnia* method instead of the usual method. Moreover, all the participants had already used other types of assistive software, but not all of then were using the smartphone as an aid, especially to resolve personal problems. Such result relates to the importance of social acceptance of assistive technologies. Using the smartphone as software aid was considered by participants always a preferable choice compared to hardware support, mostly because the software solution is more discrete and less bulky.

4.1.5 Concluding remarks

We have presented *Cromnia*, an assistive software which combines accessibility, empowerment-oriented design and neural networks to address problems of autonomy for visually impaired people. The initial experimental results suggest that *Cromnia* app is more efficient and accurate than traditional methods used by visually impaired people to dress up. Participants appreciated the benefits of a software aid for augmenting personal autonomy and they were keen to experience new versions of the application, in order to be able to eventually use it in their real lives. The current release of *Cromnia* is available from the AppStore and a high number of people have downloaded it so far.

Chapter 5

UX challenges for Cognitive Disabilities

In recent years, increasing emphasis has been placed on geoeducation in middle and high school curricula, transforming the way certain subjects are now taught to provide students with a thorough understanding of the dynamics and interconnections in the world, which are also based on relevant information. This is the case with history, geography, earth sciences and other subjects. However, so far, little has been done to provide adequate learning tools to support such a major transformation. Section 5.1 proposes the integration of advanced geospatial technology into traditional interactive learning tools as a way to describe experiences that help students, with and without learning difficulties, understand phenomena and improve their skills. One of the most common learning difficulties experienced by children and adults is dyslexia. Globally, an astounding 15% of the world's population is estimated to suffer from dyslexia. Much research is currently underway to explore the benefits of using information and communication technologies as a learning platform for individuals and especially children with such learning difficulties. The system presented is the result of a usability engineering process aimed at providing users with an effective learning experience, by iteratively analyzing their expectations and needs with respect to georeferenced content. The resulting learning environment allows users to manage content in terms of educational objects called geoLO +, which are extended with spatial and temporal components and built according to standardized metadata. The system was tested in the context of a middle school program. A usability study was conducted to analyze the impact of geoLO + resources in terms of perceived quality, engagement and learning performance of students, with and without learning difficulties, and the results of the resources were compared with the results of teaching methods traditional. The results were encouraging and showed that it is possible to achieve better learning, and consequently better UX, using the system. Remaining on the topic of the chapter, namely the UX challenges for cognitive disabilities, in Section 5.2 a project for therapy was presented through the use of a humanoid robot for children suffering from behavioral disorders, in particular for children with attention deficit. This new therapeutic methodology was created to support and make the therapist's work more attractive. The robot used is equipped with a tablet and two identical cameras. The tablet is used to let patients interact with the application, while the cameras are used to capture their emotions in real time to understand the degree of attention and any difficulties they may have. The interaction with the tablet takes place through some exercises in the form of games. The exercises performed by the patients are analyzed and combined with the data acquired by the cameras. The combination of these data is processed to propose adequate levels of therapeutic activity. This process leads to the digitization of the patient's therapeutic path so that any improvement (or worsening) is monitored and makes the robot a reliable and predictable technological intermediary for the child. The work was developed in collaboration with a diagnostic and therapeutic center, where it is being tested. By interacting with a humanoid robot, children show greater involvement and better UX, which can be explained, according to psychologists, by the fact that a robot is emotionally less rich than humans and the patient feels less fearful.

5.1 Maps4Learning: Enacting Geo-Education to Enhance Student Achievement

5.1.1 Introduction

Territory represents a unifying element for providing people with powerful heterogeneous information and services during their everyday lives. It is common indeed that data provided by a service also has embedded geospatial components in terms of either input parameters or geographic reference associated with the results. This feature allows users to handle phenomena and local dynamics through unique virtual access to related geographic information in a way that is also useful in other domains, such as cultural heritage, environmental protection and education. In particular, much emphasis has been given to the value of new learning methods, which allow students to acquire comprehensive, spatial-related knowledge in any subject and enhance their skills for their future roles in society. As a consequence, the meaning of the term geo-education has gradually evolved from ‘education in geography’ to the wider concept of ‘education about our world’ [202].

According to the well-known National Geographic community, geo-education is expected to happen across many subjects in traditional curricula to enable young people to analyse problems from different perspectives related to different disciplines. History is one subject that immediately suits this approach. Acquiring knowledge about a historical period becomes easier if it is strongly connected to the underlying territory, which guides data exploration and information visualization tasks. The same applies to environmental science and the earth sciences. Other disciplines instead show their relationship with the territory while exploring a topic of interest. This is the case for the life sciences, for example, which create connections between phenomena under investigation and their possible causes once the territory is taken into account.

The Motivation

One of the most common learning difficulties experienced by children and adults is dyslexia. Globally, an astounding 15% of the world's population is estimated to suffer from dyslexia [155].

Dyslexia is the most common learning disorder affecting the ability to read and write. Individuals with dyslexia typically read at significantly lower levels than expected, despite having normal or average intelligence scores. Much research is currently underway to explore the benefits of using information and communication technologies as a learning platform for individuals and especially children with such learning difficulties [156, 157].

Furthermore, following the trend described, a lot of research has recently been conducted in the field of e-learning to pursue the goal of greater education in geography and to devise tools that include functionalities related to the semantic web and ontologies.

Given the potential benefits, we focused on developing an interactive learning system built with the integration of advanced geospatial technologies that could foster learning and help children improve some of their core skills, such as understanding phenomena and improvement of their skills.

However, no attempts have been made to support a multidisciplinary approach to education through e-learning. Additionally, little attention has been paid to hiding the increased complexity of learning components behind usable interfaces, which could help instructors create learner-tailored learning pathways while improving the engagement of learners with learning disabilities on a given subject.

The Research Contribution

In this section, we address the challenges mentioned above and propose the following:

- a new model of spatio-temporal learning object
- an ontology specially conceived to enrich the knowledge base of such learning objects, and
- a usable, innovative geo-education environment that supports both instructors and learners in the adoption of the new model.

The geoLO+ model allows users to handle didactic units in terms of metadata used to describe several properties, including spatial and temporal properties.

We built an ontology that organizes the concepts of disciplines, topics, and (education) levels in a 3-dimensional space, enhanced by adding both a temporal and a spatial axis, altogether forming a new reference system for resource allocation. Such an ontology was conceived to allow learning content designers (e.g., schoolteachers) to organize resources based on semantic relationships, which may ultimately allow learners to browse interconnections among different didactic topics.

We show that an emphasis on UX during the design of a geo-education environment may allow both instructors and learners (with and without learning difficulties) to better exploit the high potential of geographic information and the semantic web. A usability engineering process was indeed followed to develop the proposed Maps4Learning educational environment, aimed at providing users with an effective learning

experience, iteratively analysing their expectations and needs with respect to geo-referenced content. The challenge we faced was integrating advanced geospatial technology inside traditional interactive learning tools. In this section, we present the methodology, design choices, implementation, and results of our preliminary assessment and evaluation of Map4Learnings, a geo education system which has turned out to be especially suitable for children with learning difficulties. The system was tested in the context of a middle school program. The preliminary results show the promising prospects that geo-learning has in such contexts.
To experiment with the approach, we developed the new geoLO+ resource in the Moodle learning environment and implemented the Maps4Learning web application, where students can navigate didactic units through their geographic references. The usability engineering process was completed by evaluating the system in the context of a middle school. A usability study was carried out to analyse the impact of geoLO+ resources in terms of perceived quality, engagement, and student learning performance; we also compared the results of the resources to those of traditional teaching methods. The results were encouraging and showed that a learning improvement can indeed be achieved using Maps4Learning.
Structure of the section. Subsection 5.1.2 recalls some relevant related work. In Subsection 5.1.3, the process of usability engineering that was adopted to design the *Maps4Learning* educational environment is described. In Subsection 5.1.4, we explain the formal process leading to the construction of a *geoLO+* knowledge base. In Subsection 5.1.5, the *Maps4Learning* visual environment is described, and the relevant interaction design choices are explained. Subsection 5.1.6 describes the experiment conducted in a middle school and analyses the results obtained. Some final remarks are made in Subsection 5.1.7.

5.1.2 Related Work

The goal of this subection is to provide a brief description of recent work dedicated to e-learning environments conceived to support teaching in geography, which includes features related to the semantic web and ontologies.

The literature about online learning is wide and includes different approaches used in investigations during the last three decades. Starting from well-established learning management systems, possibly enhanced by innovative technologies, such as augmented and virtual reality [196], the Internet's pervasiveness and the growing availability of open source platforms have notably diversified the way learning contents are distributed. Massive open online courses (MOOCs) are examples of an open online learning modality designed for a large number of participants [185] that is integrated with social media platforms. MOOCs have gained wide popularity and are producing a large amount of data that can be used to monitor both a single user's performance and trends resulting from groups of learners. To this aim, learning analytics has been introduced as an emerging research field to manage information gathered from massive platforms and use it to inform learning design decisions and better satisfy future learners' needs [171, 188].

The educational environment of *Maps4Learning* proposed in this section represents an innovative online learning tool for geo-education, which supports instructors in building multidisciplinary didactic units, the *geoLO*+, which are enriched with geographic references. *Maps4Learning* differs from a MOOC in that it was conceived for a limited number of learners, who may benefit from the multidisciplinary nature of didactic units. Indeed, much work has been done proving that e-learning tools improve students' performances if a multidisciplinary approach is followed when preparing the units, especially in those units dealing with scientific matters and disciplines [172, 182].

Moreover, unlike MOOCs, the geo-education tool we present was conceived to be used for closed online courses. Here, didactic units structured in terms of *geoLO*+ are distributed by training organizations or schools, are used for groups of regularly enrolled students, and are distributed on the *Maps4Learning* platform, which can also be part of a traditional training course. These courses are usually led by a teacher or a tutor, who allows communication with the participants and can perform live training in virtual classrooms.

The literature about e-learning includes different topics, such as e-learning systems, e-learning management systems, and mobile learning. In this field, the role of the learning objects is fundamental [179]. A learning object (LO) represents a coherent content unit with a standardized form, which embeds information on how to sequence and organize it into courses. The definition of a LO has been revised and extended to build different versions capable of being better adapted to complex multimedia content. In particular, in [177], LOs were extended to include a geographic component, which could be used, as an example, to reveal to the citizens of a territory the geographic information that is usually hidden in learning content, thus improving that content's searching and traceability.

Geographic information and e-learning also represent the basis for mobile learning (m-learning), a topic related to applications that support learning anywhere and anytime through the usage of mobile devices [198].

The introduction of Web 2.0 basic features has modified the way learning content is created and distributed. It has brought substantial changes to traditional e-learning methods. The introduction, for example, of wikis and blogs, has enhanced students' interaction, thus giving them a more active role in the training process and providing teachers with higher awareness of the results of their work. Moreover, in addition to pre-packaged e-learning content, students may aggregate units by means of embedded applications and may dynamically create customized content. To reach this aim, ontologies have provided geographic e-learning environments with relevant contributions.

In [170], Ali and Sah created an adaptive, quiz-based e-learning system by using ontologies that allow users of different levels and capabilities to learn their course content. The learning materials include physics, chemistry, and geography extracted from DBpedia and are presented to the users as a question/answer activity. The adaptive e-learning system is then compared with the baseline system. In [175], Bratsas *et al.* proposed an online game-based e-learning system derived from the Greek DBpedia that includes topics from geography, history, astronomy, politics, economics, and other general knowledge-related content.

Grivokostopoulou *et al.* [183] provided an adaptive, web-based e-learning system using ontology and semantic web technologies. The e-learning system allows users to learn different topics of the artificial intelligence system, which is personalized through rules that take into account users' preferences and level of knowledge.

In [181], El Fazazi *et al.* proposed a personalized e-learning platform by using the semantic web to provide a digital space for universities for academic, higher education and research purposes. The platform allows various students to appear for placement tests in their domain of expertise, which are then evaluated using machine learning models. Based on the classification of the machine learning algorithm, the system suggests specialized training using a meta-heuristic optimization algorithm. Similarly, Rani *et al.* [192] employed two ontologies to provide users with adaptive learning materials, track and manage their style, and change them accordingly. Specific adaptive agents are employed to monitor and change the user's styles of learning.

Mouromtsev *et al.* [190] developed an e-learning system using an ontology and the Information Workbench to rectify the list of problems that currently exist in the Russian education system. The authors employed semantic web technologies to cope with problems such as the weak structure of the learning materials and the poor connections between their components.

In [195], Rocha and Faron-Zucker employed semantic web technologies to automatically generate quizzes that contain questions about geography, answers to those questions, and distractors. The approach is implemented and investigated via the use case of the famous French game "Les Incollables".

Yarandi *et al.* [200] proposed an adaptive ontology-based approach to provide different users with course content according to their knowledge, capabilities, and preferences. The ontology is used to store the learning content as well as to keep track of users' progress. The content is updated for the users based on their learning progress. Similarly, in [199], Tunde e*t al.* designed an ontology-based model for an e-learning management system for course outlines, teaching methods, activities, and learning styles. The system also prevents the learners from skipping any courses that are designed as prerequisites for them.

In [173], Borges and Silveira exploit descriptors associated with metadata to improve students' searches in educational semantic portals on the web. In particular, they focus on educational videos and propose adding descriptive information to such resources once they have been properly segmented and associated with semantics for each segmented portion. This

approach allows users interested in specific video content to obtain more precise and easier access to the required section.
In [197], Shmelev *et al.* proposed a technique for sequencing learning objects for different users using a genetic algorithm, which is one of the most widely used meta-heuristic algorithms. The objective function makes use of the ontology and Bloom's taxonomy. The ontology can be used to describe the connection between the learning objects, while Bloom's taxonomy evaluates the quality of that connection.
The *geoLO+* learning objects and the interactive learning environment *Maps4Learning* were designed taking into account the demonstrated benefits coming from the use of ontologies. Each resource described as a *geoLO+* can be located in a 3-dimensional space expressed in terms of disciplines, topics, and (education) levels, thus improving the discovery functionality that can be executed by additional metadata. Adding semantic relationships enriches the capability of creating multidisciplinary contents, which can be searched through multiple geographic and temporal references, which are positioned in the 3-dimensional space.

Dyslexia is a lifelong condition, and its symptoms can vary at different stages in a person's life, but timely and appropriate intervention can provide significant results. Many methods of intervention are currently in use and more studies are needed to determine which interventions work best. Research is now focusing among others and on the potential benefits of using information and communication technologies (ICT) to develop interactive experiences and an optimistic learning environment, which can motivate and help children, thus helping them cope. their disability early on and possibly mitigate its various negative effects.
Digital technologies can be used to train, assist and even enable the learning process. Specifically, designed applications can stimulate student interest, but they can also help students with disabilities to fit and progress within traditional school environments [158]. Thomas considered ICT an enabling factor, as it can facilitate students' access by learning, increasing their motivation, fostering personal competition, improving their confidence and self-esteem [159]. Various implementations of ICT in education and learning have been researched, such as the use of websites as educational motivators for adults with learning disabilities [160], virtual environments

[161, 162], and computer games [163, 164], implementations of portable writing aids and configurable word processing environments to support people with writing difficulties [165, 166]. A major group of people with special educational needs, such as dyslexia, could potentially benefit from ICT [158, 167]. Keates [168] explained the need for dyslexic pupils to access ICT for learning and to be introduced to appropriate ICT, including hardware and software (such as different word processors) for these pupils. It is also believed that the use of multimedia helps dyslexic students [169]. Multimedia applications not only allow, but also strengthen the bimodal presentation of information through visual and auditory channels; thus, information processing is accelerated, and mnemonic recall is facilitated [163].

The Maps4 Learning interactive learning environment has been designed taking into account the proven benefits of using technology and in this section we show that it is especially useful for children with learning difficulties.

5.1.3 The Usability Engineering Process

In the educational domain, the adoption of user-centred approaches to design e-learning systems represents a promising innovative methodology with respect to traditional approaches, which were mainly focused on the technological quality of such systems. HCI theories and methodologies have been proven to support the design of appropriate e-learning settings responding to the complex and rapidly changing requirements of both the academic and business contexts of our society [178]. In the present subsection, we summarize the usability engineering process we followed for *the Maps4Learning* educational environment, aim at providing users with an effective learning experience, and iteratively analyse their expectations and needs with respect to geo-referenced content.

An initial contextual inquiry was carried out among a group of 10 teachers from primary and middle schools, with a twofold goal. First, we wished to understand the extent to which geographical references are used while teaching, to identify the subjects where they are most commonly used and to gain insight into the methods and tools that are currently adopted to enrich a given subject with geographic references. The second part of the survey was therefore dedicated to determining the extent to which the support of digital teaching at school and/or for homework is pursued, in particular, aimed at children with learning disabilities. The interviewees reported several examples of subjects taught with the support of geographic references, from history, biology, and physics to the Earth sciences, Italian literature, and more.

In many cases, the temporal dimension of the geographic information was also relevant (e.g., for biology, the bubonic plague in the territory of Milan, Italy, in 1630-1631, or, for history, the growth of the Roman Empire from 264 B.C. to 44 B.C.). For the instructional tools adopted in a classroom, the interviewees usually rely on the use of Internet-connected multimedia interactive whiteboards, and many of the interviewees adopt cloud platforms to share teaching materials with students and manage homework. The teachers consider such tools pedagogically valid, yet they admit that higher effectiveness could be achieved through the ability to connect materials to their geographical references in terms of historical map visualization.

As a result of the contextual inquiry, we started a scenario-based design process meant to build a geo-education component, which could enhance learners' experience with didactic content inside a traditional e-learning system. The challenge was, on the one hand, to provide instructors with some usable interfaces able to support geo-educational activities and, on the other hand, to improve students' engagement through innovative and customized learning paths.

The major requirements were as follows:

- management of geospatial didactic resources (instructor),
- design of a geo-educational topic (instructor), and
- interactive visualization of any didactic geo-educational topic in a given discipline for a given education level (student).

Satisfying the first requirement was a prerequisite for the instructors. An appropriate knowledge base was designed that consisted of didactic resources, which properly encompass any (possibly indirect) historical or spatial reference to other objects/events of interest inside a given topic and can be associated with multiple disciplines. A Web of Data-oriented approach was followed to realize such a knowledge base to connect ontologies and complex data on a world scale, as explained in detail in Subection 5.1.5.

A user-centred interaction design process was carried out to provide *Maps4Learning* with usable interfaces. As illustrated in the use-case diagram of Figure 5.1, the instructor may interact with the system to

- add/modify/delete didactic resources managing the underlying *geoLO*+ knowledge base, or
- create a geoLO+ starting from existing resources, or
- give a lesson organizing the available geoLO+ resources.

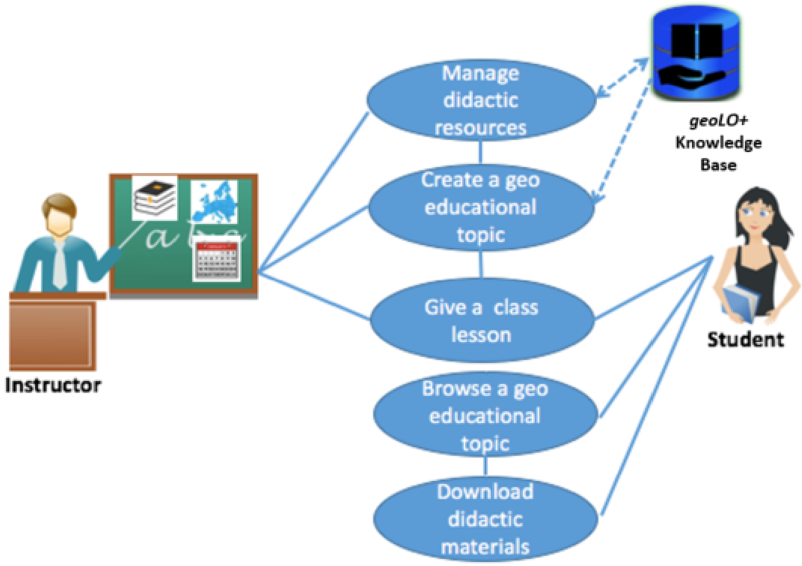


Figure 5.1: Maps4Learning users and their abilities.

The student, who represents the second type of primary user for *the Maps4Learning* platform, may browse a geoLO+ offered as part of a lesson and may download didactic materials that also include thematic maps associated with certain topics.

As anticipated, a participatory design approach was adopted. Among the 10 teachers who took part in the initial contextual inquiry, the 4 who appeared most motivated and keen to contribute were asked to participate and actively give their feedback in crucial design decisions. They were also involved in formative evaluation activities performed to test the interface prototypes throughout the design process.

During the "Investigation Phase", therefore, we developed a survey aimed at parents with children with dyslexia, aged between 7 and 12 years. The

process is based on carefully chosen questions; based on research and specialized methods, so that response data can be collected, and reliable results provided. We developed questions that involved the following areas: their children's demographics (age, gender, etc.), type of dyslexia the child struggled with, when and where symptoms were first noticed, with what tasks the children had difficulties, what were their treatment and their familiarity with ICT and cell phones.
The survey questions evaluate the views and opinions of parents, while offering a better understanding of the problems and difficulties their children face. The survey was conducted with 5 parents and the results gave us an overview of the main indicators of dyslexia and how parents come to observe and react to the first symptoms. Analyzing the responses, we noticed that most of the children did not have severe symptoms of dyslexia (the children were all diagnosed by specialists and speech therapists). The tasks that confused and upset most of the children were reading, writing and memorizing. All parents agreed that the design of a system based on a geo-educational component could improve the students' experience with didactic contents within a traditional e-learning system. The challenge was, on the one hand, to provide instructors with some usable interfaces capable of supporting geo-educational activities and, on the other, to improve the involvement of students with learning disabilities through innovative and personalized learning paths.

5.1.4 Designing a geoLO+

In the present subection, we describe the creation process of a geoLO+ knowledge base. Before that, some preliminaries are given on the set of metadata applied to model a document as a spatio-temporal resource.

Preliminaries

Each resource (i.e., text, photo, video, and didactic material) available on the web can be structured according to the Dublin Core (DC) Metadata Element Set [180]. This set represents a vocabulary of fifteen broad and generic properties to be used to facilitate information discovery on the rapidly growing web.

Table 5.1 lists the Dublin Core metadata elements. The fifteen Italic terms represent the basic elements to which sub-elements and qualifiers have been added.

Table 5.1: DC Metadata Elements.

abstract	educationLevel	license
accessRights	extent	mediator
accrualMethod	*format*	medium
accrualPeriodicity	hasFormat	modified
accrualPolicy	hasPart	provenance
alternative	hasVersion	publisher
audience	*identifier*	references
available	instructionalMethod	*relation*
bibliographicCitation	isFormatOf	replaces requires
conformsTo	isPartOf	*rights*
contributor	isReferencedBy	rightsHolder
coverage	isReplacedBy	*source*
created	isRequiredBy	spatial
creator	issued	subject
date	isVersionOf	tableOfContents
dateAccepted	*language*	temporal
dateCopyrighted		*title*
dateSubmitted		*type*
description		valid

Each term is, in turn, specified through a set of 5 attributes, namely, Name, Label, URI, Definition, and Type of Term and, when applicable, also by additional attributes that provide further information, namely, Comment, See, References, Refines, Broader Than, Narrower Than, Has Domain, Has Range, Member Of, Instance Of, Version, Equivalent Property. As an example, the DC metadata element *Coverage* is defined as follows:

Term Name: Coverage
URI: http://purl.org/dc/terms/coverage
LABEL: Coverage
DEFINITION: The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
COMMENT: Spatial topic and spatial applicability may be a named place or a location specified by its geographic coordinates. The temporal topic may be a named period, date, or date range. A

jurisdiction may be a named administrative entity or a geographic place to which the resource applies. The recommended best practice is to use a controlled vocabulary such as the Thesaurus of Geographic Names (TGN). Where appropriate, named places or time periods can be used in preference to numeric identifiers such as sets of coordinates or date ranges.
REFERENCES: [TGN] http://www.getty.edu/research/tools/vocabulary/tgn/index.html
TYPE OF TERM: PropertyREFINES: http://purl.org/dc/elements/1.1/coverageHASRANGE:http://purl.org/dc/terms/LocationPeriodOrJurisdictionVERSION:http://dublincore.org/usage/terms/history/#coverageT-001

The geoLO+ Knowledge Base Construction

To organize didactic units in terms of knowledge base elements, each resource can be associated with a subset of the above DC metadata elements. In particular, the term *Coverage* is used to express the spatial and temporal properties enriched, when necessary, by details in terms of geometry and format, which allow locating it in a 3-dimensional space according to the concepts of disciplines, topics and levels. The purpose is to build a knowledge base where the role that a document (resource) plays can be fully exploited, that is, it is automatically selected anytime it contains a direct or indirect and historical or spatial reference about any object/event of interest for a topic in a given discipline. This goal is achieved by adopting a Web of Data-oriented approach, where resources are also linked through semantic relationships between (parts of) the resources, thus enriching the knowledge base and projecting it on a world scale, always keeping in mind the goals of availability and accessibility by both humans and software applications [184].

Figure 5.2 shows an ancient document about the history of Naples, which can be used to illustrate the proposed approach. It describes an event involving a famous point of interest (POI), “Il Castel Nuovo e il Molo Grande”, an Angevin fortress with its pier, also known as “Il Maschio Angioino”, built by Carlo d’Angiò starting in 1279.

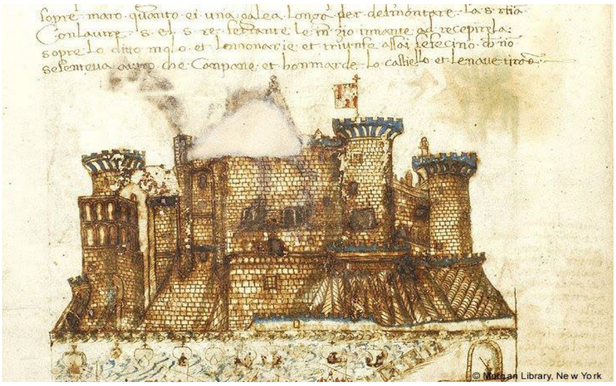


Figure 5.2: “A li 1477 che fo alli xi de sottiembro intrao in la cita de napole la illostrissima signora riina mugliere del signor re Ferrante. . .” source: www.napoliaragonese.it/castel-nuovo-nel-1477.

In this exceptional document, Castel Nuovo and Molo Grande are represented with the arrival of boats. Information describing the event is shown in Table 5.2.

Table 5.2: Information describing the document.

Melchiorre Ferraiolo, Cronaca della Napoli Aragonese, 1498 – 1503	Chronicle of the Aragonese Naples by Melchiorre Ferraiolo, 1498 - 1503
Foglio 90v del manoscritto MS M.801 presso la Pierpont Morgan Library di New York (unica copia esistente)	Worksheet 90v of the MS M.801 manuscript at the Pierpont Morgan Library in New York (the only existing copy)
L’11 Settembre 1477 la Regina Giovanna d’Aragona fa il suo ingresso in Napoli, via mare.	On 11 September 1477 the Queen Giovanna of Aragon enters Naples, by sea.

The resource has three different (implicit and explicit) historical references. The first two explicit links refer to the date of the described event, namely, 1477, and to the document date (i.e., a range 1498 – 1503), respectively. The third one, an implicit reference that can be added to the metadata of the document, corresponds to the year of monument construction, namely, 1279. To properly handle such a multivalued feature, the Coverage element has to be extended so that each historical reference is captured by a suitable format and is associated with a corresponding geometric primitive (e.g., polygon or point).

Semantic relationships are necessary for building the underlying knowledge base and establishing links among documents referring in some way to the same resources. The model adopted for semantic relationships is RDF [194]. Each statement is enunciated as a triple, subject – predicate – object, where subject and predicate are expressed as resources (i.e., they are specified as URIs/IRIs) while an object may be a resource or a literal object. In the previous example, many semantic relationships can be initially stated as RDF triples suitable to express relationships among individuals or statements with the *rdf:type* (*a*, for short) predicate. As an example, to state that the resource *Castel-nuovo-nel-1477* is a manuscript authored by Melchiorre Ferraiolo and dated between 1498 and 1503, the following triples (subject – predicate – object) can be specified:

Subject	Predicate	Object
gc : Castel-nuovo-nel-1477	a	dcterms : Manuscript.
gc : Castel-nuovo-nel-1477	dcterms : author	dbpedia : Melchiorre Ferraiolo .
gc : Castel-nuovo-nel-1477	dcterms : date	“1498 – 1503” .

where *dcterms* and *gc* correspond to the Dublin Core and *geoLO*+ prefixes, respectively.

Such an initial set of metadata and semantic/temporal/spatial relationships is enough to help users capture every basic feature associated with the item of interest. As an example, it is possible to select documents about POIs referring to the Aragonese period and obtain information about their location, authorship, state of conservation, etc.

This initial level of data structuring underlies the exploration of the whole knowledge base in terms of selections and conditions applied to specific known parameters. To enhance the capability of modelling concepts and relationships in the domain of interest, more expressive constructs and operators were adopted, namely, some properties from RDF Schema and OWL [201]. Then, an ontology was built that provides designers with a data model and vocabularies for describing the domain content. It is based on a model for the basic disciplines and organizes *geoLO*+ to let users (students) discover and explore content. In particular, the ontology organizes the concepts of disciplines, topics and levels in a 3-dimensional space that is

enhanced by adding both a temporal and a spatial axis, thus allowing document allocation with reference to a discipline, a topic relevant to it, an education level, one or more temporal tags and one or more spatial positions.
Once the triples have been specified, they can be serialized according to the RDF/XML syntax [193]. The ontology was built in Protegé [191]. Figure 5.3 displays part of the taxonomy and some classes of interest.

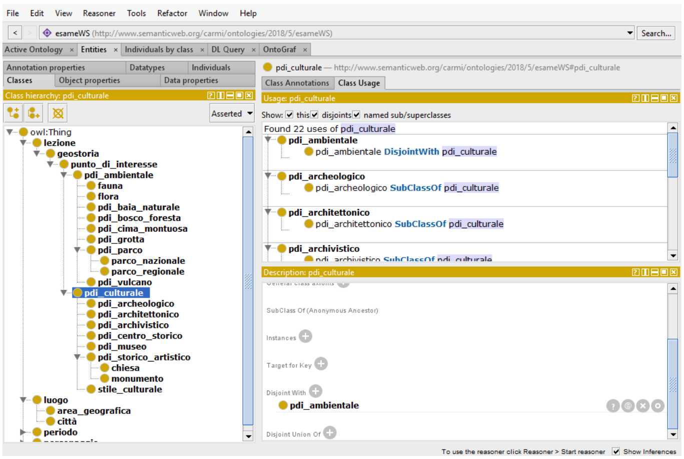


Figure 5.3: A screenshot showing some classes of the ontology.

Far from being complete, it contains basic topics concerning the geo-history ('geostoria') discipline, and the class hierarchy is organized in terms of lesson (lezione), place (luogo), period (period) and subject (argomento). In particular, place and period are structured according to some well-established online ontologies, such as GeoNames and Time, subject is organized in agreement with the curriculum and syllabus of the school's didactic offerings, and lesson is organized as a set of geo-history (geostoria) classes. This last component is then divided into subclasses of different points of interest (punto di interesse, POI), such as environmental POI (for example, natural bay and forest), cultural POI (for example, museum and archaeological site) and historical POI (for example, church and monument). These georeferenced POIs are acquired by posing SPARQL queries to DBpedia. Additional queries can be posed to populate the underlying knowledge base through the Virtuoso SPARQL Query Editor to

the endpoint for DBpedia. A procedure in Python was built that derives and translates results into csv files for open data-oriented management.
In the following subsection, some of the more expressive statements are specified; these were validated by the reasoner used to build the ontology. The indentation emphasizes the triple structure, thus avoiding redundancy in their components

Subject	Predicate	Object
gc : Worksheet 90v	rdfs : isPartOf	gc : Castel-nuovo-nel-1477 .
gc : Castel-nuovo-nel-1477	a	rdfs : Resource .
gc : Castel-nuovo-nel-1477	a	gc : Topic .
gc : Topic	rdfs : isPartOf	gc : Discipline .
gc : Castel-nuovo-nel-1477	gc : learning	dbpedia : History .
gc : learning	a rdfs : domain rdfs : range	rdf : Property. gc : Topic. dbpedia : Discipline .

where dbpedia corresponds to the DBpedia prefix.
Such an improvement in terms of relationships enhances users' exploration capability. As an example, it is possible to capture the locations of POIs whose historical descriptions cite the presence of Queen Giovanna of Aragon, thus visually creating a geographic link among places where she was in a period chosen by the user.

5.1.5 Maps4Learning - an environment for geo-education empowerment

The goal of this subsection is twofold. It describes how an expert in a discipline (e.g., a teacher) can create a geoLO+ structure according to the model described in Subsection 5.1.3. Then, this subsection shows how such resources are displayed in the visual environment for knowledge exploration by different typologies of users.

The Interaction Design for geoLO+ Creation

The following figures show the activities performed for resource creation and modification. Such a capability is offered as an additional functionality in Moodle, the popular modular open source platform used to support e-learning processes [189]. In Moodle, users can add content and general-purpose stuff and develop, plan and handle different kinds of learning evaluation activities, such as assignments and questionnaires. The open source nature of Moodle also allows for the creation of customized functionality. Figure 5.4 shows a *geoLO+* resource that was developed and added to the set of default activities and resources.

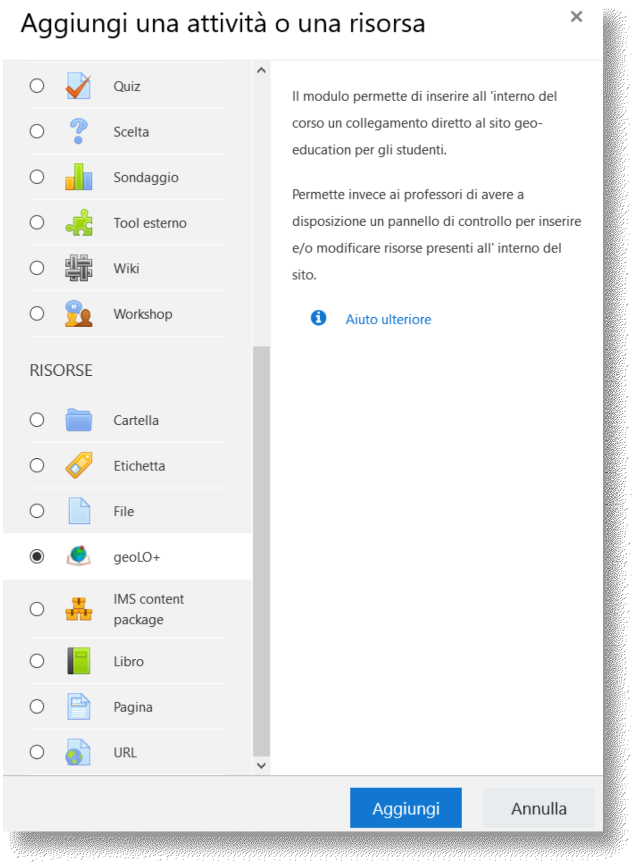


Figure 5.4: A geoLO+ resource as provided in Moodle.

A *geoLO+* is conceived to model multidisciplinary content according to the set of mandatory metadata described in Subsection 5.1.3. Once the *geoLO+* option is selected, the insertion of a new document, handled as a new resource, may start (Figure 5.5). The Coverage element can be managed either as a set of timestamps and a couple coordinates or in a more complex way, such as a hierarchy of temporal attributes and polygonal areas (set of vertices).

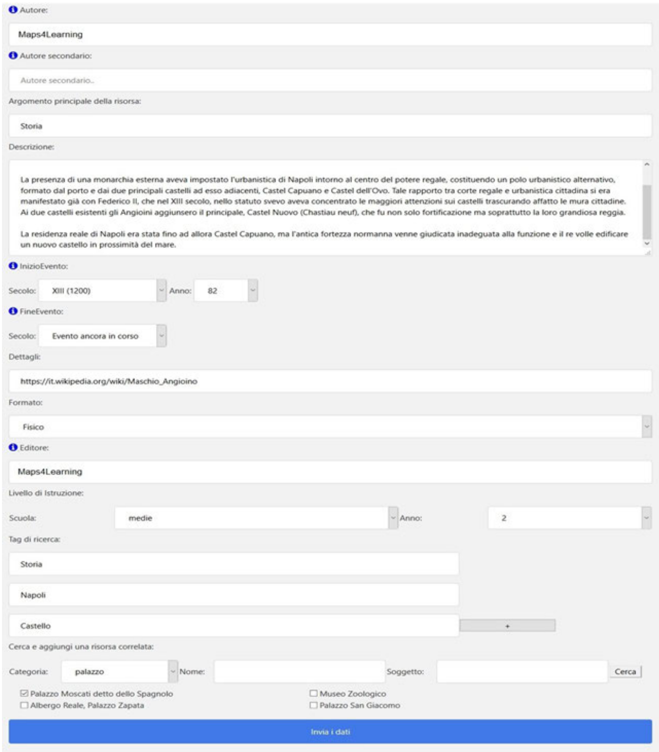


Figure 5.5: A geoLO+ creation.

Figure 5.6 shows the geo-referencing task and the corresponding parameters for the new resource. Coordinates are automatically acquired by positioning the *geoLO*+ on the map. To establish semantic relationships, it is possible to connect a *geoLO*+ and every resource semantically connected to it, thus allowing semantic queries in terms of the graph matching of triples (Figure 5.7).

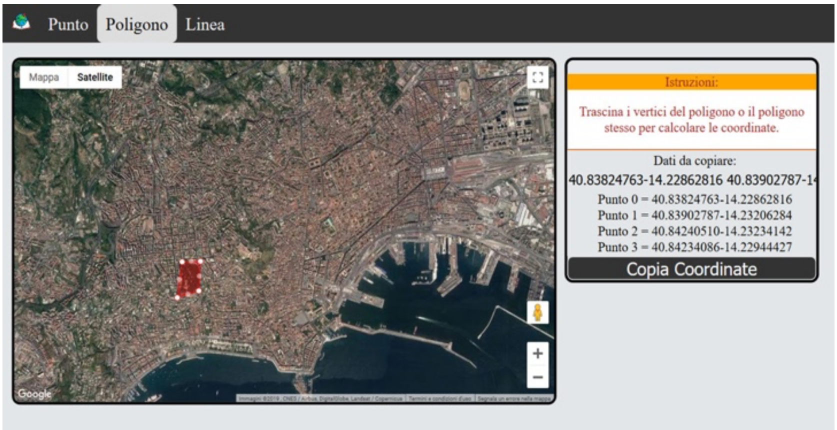


Figure 5.6: Geo-referencing a resource.

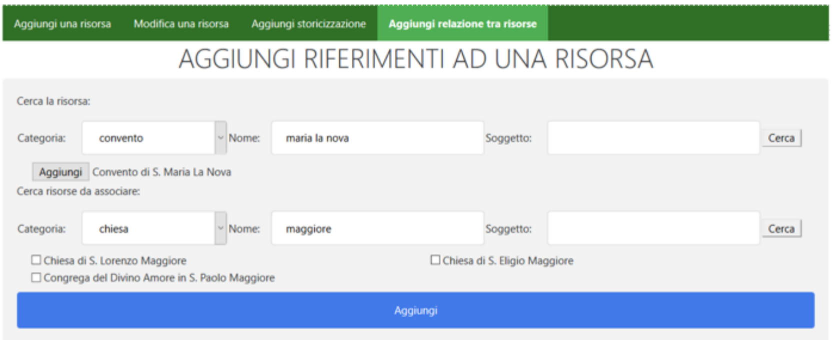


Figure 5.7: Semantically linking geoLO+s.

Interactive Visualization of any Didactic Educational Topic

Figure 5.8 shows the Maps4Learning visual environment used to query and explore the knowledge base that has been populated to date. In addition to using a legend that lists the POI categories stored in the knowledge base, it is possible to choose a temporal range and a geographic area through which POIs can be filtered. Additionally, criteria involving specific parameters related to didactic aspects, such as teaching level and disciplines, can be applied.

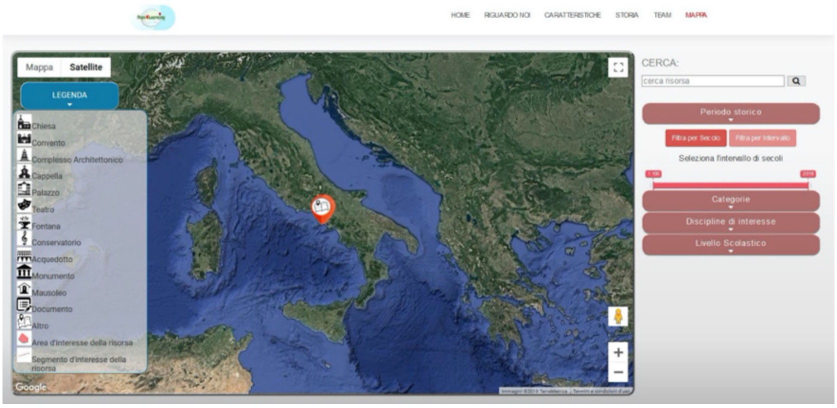


Figure 5.8: The Maps4Learning environment.

Once a query is executed, a historical map may be overlapped that contains POIs referring to the period selected through the temporal slider. On the bottom, a brief description of the period is recalled. By clicking on a POI, a link to DBpedia is shown with a preliminary description embedded, as shown in Figure 5.9.

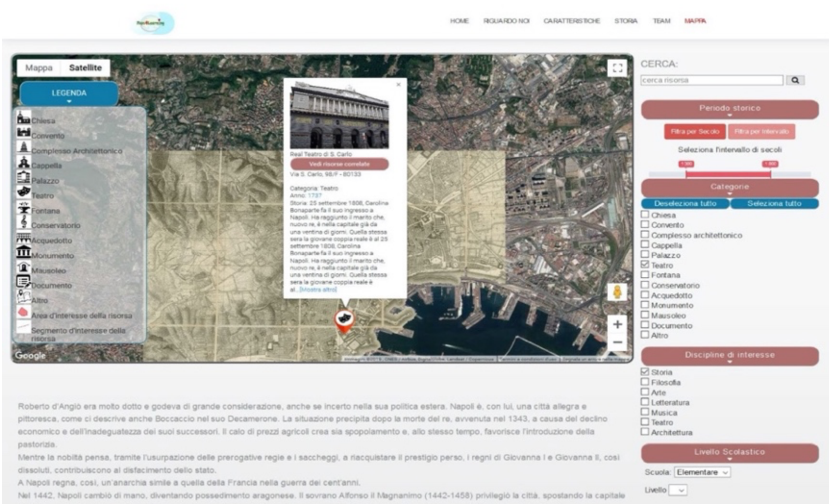


Figure 5.9: Exploring a geoLO+ in Maps4Learning.

Finally, starting from a specific POI, semantically related topics can be accessed by simply invoking them from the preliminary description of the POI itself. Figure 5.10 shows POIs related to the Veiled Christ (Il Cristo

Velato) sculpted by Giuseppe Sanmartino. Four POIs are displayed, each of them sharing a property with the statue, thus creating a multidisciplinary network of didactic units. Of particular note are the following:

- the Sansevero Chapel, where the masterpiece is currently located and where bio-chemical studies on human bodies are stored;
- the Historical Archive of the Bank of Naples (Archivio del Banco di Napoli), where one document refers to a down-payment of fifty ducats to Giuseppe Sanmartino signed by Raimondo di Sangro;
- the San Martino Museum (Museo di San Martino), where a terracotta scale model of the Christ by Corradini, the sculptor initially commissioned to complete the work, is preserved, and finally;
- Palazzo di Sangro, which belonged to the Sansevero princes.

When one of these POIs is selected, historically, geographically and semantically related POIs can be immediately displayed to emphasize their relationships and allow users to capture both different aspects of the same topic and the presence of a feature in different topics/disciplines.

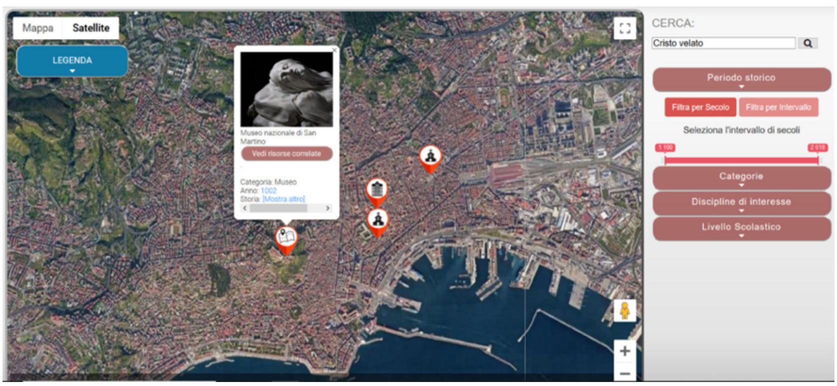


Figure 5.10: Exploring a topic through Maps4Lerning.

Maps4Learning Architecture

As for the underlying architecture, Maps4Learning is a client server platform, as shown in Figure 5.11. Clients can be different thanks to the usage of software for building responsive applications. In particular, the following software components have been used:

- NodeJS for running JavaScript applications outside the browser;
- Angular JS and IONIC to create graphic interfaces for mobile hybrid interfaces;
- Apache Cordova for building native applications for mobile multi platforms embedding web applications;
- Jena, a Java framework, for building semantic web and linked data applications.

HTTP requests are posed to and resolved by Jena through NodeJS. JSON is also in charge of exchanging data between clients and servers.

As for the server, both Altervista and the ones underlying the UNISA e-learning platform have been used. The former is used by students developing modules for topics. The latter is used to embed and provide users with the validated versions of documents.

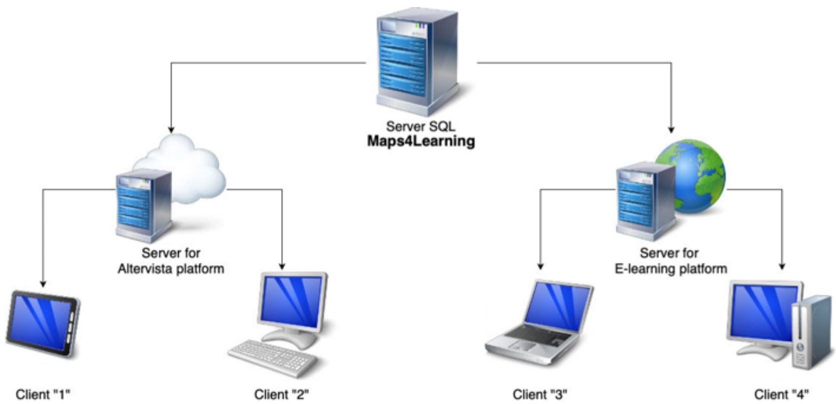


Figure 5.11: Maps4Learning architecture.

5.1.6 Usability Evaluation of Maps4Learning

Usability testing is recognized as a crucial activity to improve the quality and effectiveness of e-learning systems [176]. Usability concerns for such systems include not only traditional properties related to the ease of use of the hosting environment but also to the efficacy of the environment as a learning tool and to students' engagement with the provided content. The final goal of our usability study was therefore to analyse the impact of geoLO+ resources in middle schools in terms of perceived quality,

engagement, and student learning performance. The study was carried out in two different phases. In the first phase, we performed an experiment meant to compare the adoption of geoLO+ resources with a traditional teaching method. In the second phase, we assessed the instructors' and students' perspectives, asking questions of both groups of stakeholders.
In this phase we have divided the students with and without learning difficulties as one of the challenges was to improve the engagement of students with learning difficulties through innovative and personalized learning pathways. From the outset, it was obvious that for the design of such an application, we had to collect as much input as possible from students, teachers, and parents of students with learning disabilities. In an effort to develop conceptualizations and assessment methods for our learning application, a personal and contextual learning process was considered. It is important that the system is adapted to each student's learning level, recognizing their diversity and needs. It is also necessary to understand if a minimum initial level of knowledge is required, in order to acquire and learn from the application itself.
We have chosen to develop the application in Italian in order not to create difficulties in completing certain levels due to their knowledge of the language. Finally, we interviewed the parents of the children to record their views and opinions, offering us useful information that helped us better understand the environment where children with dyslexia operate and the difficulty they encounter every day.

Participants

1) Instructors

The teachers who took part in the study came from three middle schools. The sample consisted of 21 teachers, 16 women and 5 men. All of them had at least 8 years of teaching experience. The subject areas taught by the teachers were history and geography (n=7), science (n=6), foreign language (n=5), and technology (n=3). Before forming the sample, we ensured that all of the selected participants were well acquainted with computer technology and had a positive attitude towards digital learning tools, having used multimedia interactive whiteboards and/or cloud platforms to share teaching materials.

2) Students

The students involved in the study were distributed over 7 middle school classes, chosen on the basis of the selected teachers. Thus, the sample of students consisted of 4 classes from year 2 and 3 classes from year 3, with an overall population of 179 students (92 girls and 87 boys), aged from 11 to 13. As digital natives, all of them were regular web users and had experienced the use of interactive whiteboards and cloud-based platforms for downloading teachers' didactic materials at least once.

Furthermore, five students with dyslexia participated in this evaluation study. We classified them based on their level of dyslexia symptoms, gender, age, and treatment period in which they visited a therapist. Furthermore, the classification of dyslexia was based on the students' diagnosis by an expert, based on a psycho-educational assessment. The criteria used for the assessment of each individual were based on the age of the users and the level of learning difficulties.

We also established a "control group" by recruiting 5 students of the same age group who were not assessed as having learning difficulties. The comparative use of the application and the subsequent data analysis also allowed us to determine whether the results of our assessment were actually related to learning difficulties or other factors (design choices, technology, IT experience, etc.). For example, knowledge of Italian has been isolated as an important prerequisite for users of this application.

Procedure

During the first phase of the experiment, teachers were distributed among 7 teams, each with 3 people teaching different subjects but sharing the same class of students. Each team was asked to create a geoLO+ multidisciplinary resource and to choose a leader, who was expected to give an interactive lesson starting from his/her subject (the first subject in Table 5.3). Most of the participants had never used Moodle, so they were preliminarily trained on the use of the Moodle-based platform Maps4Learning. Specifically, the team of experimenters spent one day training participants on how to create geoLO+ resources through the specification of appropriate metadata attributes and how to assess students' learning progress on related topics through evaluation tests.

Then, the teachers' teams were instructed to start by choosing a basic resource (e.g., a document, a video, an image) and specifying its metadata properties while also performing the georeferencing task. Each team member was also asked to add some evaluation tests that could allow student learning assessment on the specific subject each teacher taught. After the creation session, each participant was asked to respond to a survey on the use of the platform to create geoLO+ resources and on the motivation for the adoption of the created geoLO+ inside a classroom lesson, which follows the consolidated approaches available in the literature [186].

Table 5.3: Results of students' learning assessment tests.

		Mean Scores (*Standard Deviation)*		**Success Rate**	
Participants	**Subjects**	μ_{CG}	μ_{EG}	**CG**	**EG**
Class1 (n_1=13, n_2=14)	His-FL-Scie	4.23 (*1.96*)	5.00 (*2.42*)	23%	38%
Class2 (n_1=13, n_2=12)	Scie-Geo-FL	3.08 (*2.95*)	4.54 (*1.85*)	15%	31%
Class3 (n_1=13, n_2=13)	Geo-Scie-FL	3.69 (*1.89*)	5.77 (*2.13*)	23%	62%
Class4 (n_1=14, n_2=13)	FL-His-Tech	4.29 (*1.68*)	5.50 (*1.79*)	36%	64%
Class5 (n_1=12, n_2=13)	His-Scie-Tech	4.08 (*1.68*)	5.0 (*2.50*)	25%	25%
Class6 (n_1=13, n_2=13)	Scie-Geo-FL	3.50 (*2.64*)	5.25 (*2.14*)	23%	54%
Class7 (n_1=11, n_2=12)	Tech-Scie-His	3.73 (*2.37*)	5.27 (*2.41*)	18%	55%
Total		**3.85 (*2.01*)**	**5.11 (*2.16*)**	**24%**	**47%**

To involve students in the experiment, we prepared a written formal consent form, which parents were asked to sign. Then, each class was split into two groups of students of approximately equal size, forming the control group (CG) and the experimental group (EG). The former group received traditional lessons on the chosen topic; the latter took part in an interactive Maps4Learning lesson. In regard to the validity of the experiment, in order to avoid biases coming from the quality of students in each group, students were distributed by their teachers so that the average grades were comparable between the two groups. Considering the chosen topic, the first group of each class received three separate traditional lessons by the teachers of the associated team, and the second group received a comprehensive lesson by the lead teacher using the created geoLO+ resource. All of the students were asked to undertake a class test prepared by the team of teachers. Grades below 6/10 were considered failures.

After evaluating the post-lesson test taken by students, teachers were asked to respond to a questionnaire meant to determine how much they perceived students had learned and the level of the students' engagement with the geoLO+ used while also considering the hosting Maps4Learning visual environment.

Students who were in the second group of each class (forming the experimental group), in turn, had to complete an anonymous online survey about their use of the geoLO+ resource embedded in the given visual environment, responding to questions about the degree of personal engagement and also with reference to the multidisciplinary nature of the explored resource and the degree of difficulty encountered when performing the tests [174, 187].

Overall, it took approximately three months to complete the experiment: from March to May 2019.

Results

The experimental results show that a learning improvement can be achieved using Maps4Learning.

1) The construction OF geoLO+ resources

The first part of the study was conducted in the HCI-UsE usability engineering lab of the Department of Computer Science, University of

Salerno. Teams were given two hours for brainstorming to decide the topics they would embedin the geoLO+ resource, they were expected to build and to prepare a list of questions that could be used for the learning assessment tests. When the construction session started, an observer sat behind each team, annotating the errors, the time to complete the tasks, and any comments from the participants, who were encouraged to think aloud while collaborating on the tasks. The creation task was divided into two subtasks, the *geoLO+ construction* and the *learning assessment test* definition. For the former task, each teacher in the group was asked to interact with the system in turn to specify metadata properties relevant to his/her own subject. The performance of each teacher in the creation of the 7 *geoLO+* resources was recorded in terms of task completion time (see Table 5.4). When asked about it later, all of the teachers judged that the time required to construct the geoLO+ was reasonable, especially as they were all experiencing the system as novice users. In addition, they were aware that they would be able to repeatedly use the resulting learning object again in the future and hence to capitalize on their effort. For the second subtask, to ensure uniformity, we asked each team to build an assessment module made of 9 T/F questions, 3 per subject, chosen from the list they had prepared during the brainstorming session. Table 5.4 summarizes the results of the teacher group performances on the creation of the 7 *geoLO+* resources. In particular, the overall time to complete the task was recorded as the sum of the single actions taken.

Table 5.4: GeoLO+ construction.

Team of instructors (triplet of subjects)	Overall Time
Team 1 (His-FL-Scie)	12 min 56 sec
Team 2 (Scie-Geo-FL)	9 min 15 sec
Team 3 (Geo-Scie-FL)	11 min 17 sec
Team 4 (FL-His-Tech)	8 min 36 sec
Team 5 (His-Scie-Tech)	12 min 14 sec
Team 6 (Scie-Geo-FL)	14 min 45 sec
Team 7 (Tech-Scie-His)	10 min 39 sec

2) Students' Learning Assessment

The second part of the experiment, performed during regular school activities, was focused on students' learning assessment and on their reaction to the use of geoLO+ inside the *Maps4Learning* platform. Students in each class took the test consisting of 9 T/F questions as devised by their

teachers in the first part of the experiment. The test consisted of 3 questions for each of the 3 subjects chosen for the class (see Table 5.4)[1]. For each class, Table 5.3 reports the average scores achieved by each group of students on the three subjects and the corresponding success rate. In all cases, a better learning effect was recorded for the group of students who used the geoLO+ resources (EG in Table 5.3).
Figure 5.12 indicates that given a fair distribution of students between the EG and CG groups, a higher success rate was achieved in each class by EG students.

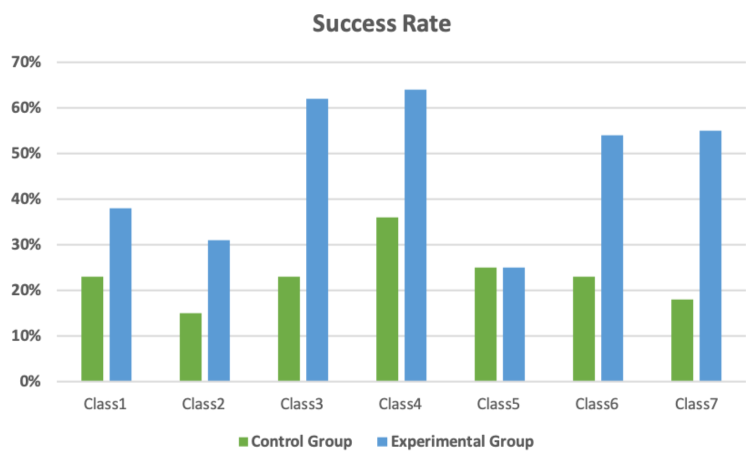


Figure 5.12: Comparing success rates in each class.

To prove that a statistically significant improvement is achieved in terms of learning effect by using Maps4Learning instead of the traditional method, we conducted an independent sample t-test (see Table 5.5).
Before applying the t-test, we verified that the right conditions existed, namely, we tested for the following:

- a normal distribution of data for each CG and EG group using the Kolmogorov-Smirnov test and
- a homogeneity of variances tested using Levene’s test.

[1] An example of class questionnaire can be found at https://forms.gle/5cEh6gikUjcCqVXE7

As reported in Table 5.5, the Levene test returned an f-ratio value of 0.459, and the p-value was 0.499. That means that the result is not significant at $p < 0.05$ and that the hypothesis that variances are not equal is rejected.

The hypotheses for the t-test were expressed as:

H0: $\mu EG - \mu CG = 0$ ("the difference of the means between the two groups is equal to zero")

H1: $\mu EG - \mu CG \neq 0$ ("the difference of the means between the two groups is not equal to zero").

In the table, t=-4.029 is the computed test statistic, df is the degrees of freedom, and Sig (2-tailed) is the p-value corresponding to the given test statistic and degrees of freedom. Since $p < 0.001$ is less than the chosen significance level $\alpha = 0.05$, we can reject the null hypothesis $\mathbf{H_0}$ in favour of the alternative hypothesis H_1 and conclude that the mean scores for CG students and EG students are significantly different. This proves that the learning improvement of EG students over that of CG students is not due to the specific case of the sample.

Table 5.5: GeoLO+ construction. Independent Samples t-Test.

Levene's Test for Equality of Variances		***t-test for Equality of Means***						
F	Sig	t	df	Sig. (2 - tailed)	Mean Difference	Std. Error Difference	Lower	Upper
.459	.499	-4.029	176	.000	-1.258	.312	-1.875	-.642

Even after the experimental assessment with children with learning difficulties was conducted, we collected the results and analyzed the students' progress. Our first significant observation was that all students showed their preference for learning through Map4Learning rather than traditional methods. The system helps children with dyslexia to focus and keep them focused, avoiding distractions. A result that indicates the importance of technology in today's learning methods. Students with different levels of dyslexia indicated differences in the duration of each test, while students with mild dyslexia symptoms completed the tests much faster than the others. Most of the students indicated a higher performance (based on the score obtained) than previous evaluations. Experimental results show that improved learning can be achieved using Maps4Learning.

3) The Survey results

The teachers' feedback on the perceived teaching efficacy of the proposed tool was considered paramount for deriving a qualitative measure for the tool's validity. Therefore, teachers were asked to respond to a brief questionnaire meant to evaluate students' ability to work in the hosting environment, the perceived learning progress and their engagement with the geoLO+ they used during the experiment. Table 5.6 reports the average responses gained from the interviews carried out with the 21 teachers to understand how they perceived the effect of the proposed geoLO+ on students in terms of system usability, learning progress and engagement. Responses were based on a 5-point Likert scale to derive a qualitative measure of the three considered aspects.

Table 5.6: Independent Samples t-Test Survey on teachers' perceived usability of the tool.

QUESTION	**Teachers (*N*=21) Mean (*SD*)**
Maps4Learning Usability	
Was it easy for students to locate and access the geoLO+ associated with the lesson you gave?	3.00 (*1.07*)
Did students find it easy to use the geoLO+ inside the environment?	2.14 (*0.89*)
Was it easy for students to explore the multidisciplinary resources associated with the geoLO+?	3.59 (*0.85*)
	2.91 (*0.79*)
Learning progress	
Did the interactive, geo-referenced nature of the geoLO+ help students learn?	3.7
Were students able to learn from the geoLO+?	2.83
Did the multidisciplinary nature of the geoLO+ help students learn the given topic?	3.58
	3.37 (*0,47*)
Engagement	
Did students like interacting with the geoLO+?	4.3
Were students focused while using the geoLO+?	3.66
Were students motivated while exploring the related didactic materials?	3.21
	3.98 (*0.45*)

Open-ended questions	
What was the overall impact of the geoLO+ on your lesson? (open ended, sample comments)	**Overall positive 56%** *'Students kept focused and reactive during the class.'* *'The interactivity of the geoLO+ helped the students view interesting time and space relationships.'* **Some issues 27%** *'My weakest students were not able to relate the topic of the lesson* [the hop plant] *to historical events using the geoLO+ and associated materials.'* **Overall negative 17%** *'Some material associated with the geoLO+ was not appropriate for my students' education level.'*

To gain insights into how the students perceived the use of geoLO+ within the given environment, we asked the 90 students in the experiment group to respond to a questionnaire, again covering the aspects of usability, learning achievement, and engagement. Table 5.7 reports the average responses received from students. As in the first questionnaire, responses were based on a 5-point Likert scale to derive a qualitative measure.

Table 5.7: Survey on students' perceived usability of the tool.

QUESTION	Students (*N*=89) Mean (*SD*)
Maps4Learning Usability	
How easy was it for you to locate and access the geoLO+ associated with the topic discussed by the teacher during the lesson?	3.28 (*1.35*)
How easy was it for you to use the geoLO+ inside the environment?	3.27 (*1.55*)
How easy was it for you to explore the multidisciplinary resources associated with the geoLO+?	3.97 (*1.24*)
	3.51 (*1.38*)
Learning progress	
Did the interactive nature of the geoLO+ help you learn?	3.92 (*0.86*)
Were you able to learn from the geoLO+?	3.77 (*0.90*)
Did the references to other disciplines (e.g., historical or geographical references) in the geoLO+ help you gain a better understanding of the given topic?	2.65 (*1.12*)
	4.01 (*0.96*)
Engagement	
Did you like interacting with the geoLO+?	4.42 (*0.76*)
Were you focused while using the geoLO+?	3.88 (*0.66*)
Were you curious about the related didactic materials you could discover?	3.64 (*0.70*)
	3.98 (*0.71*)

The bar chart in Figure 5.13 summarizes the survey results and compares the means of perceived usability, learning progress, and engagement between the group of teachers and the group of students.

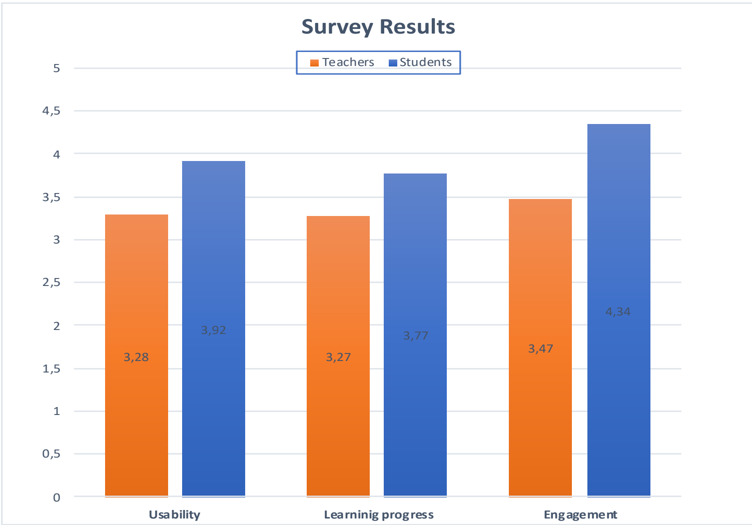


Figure 5.13: Comparing the means of perceived usability, learning progress and engagement.

While, the bar graph in Figure 5.14 summarizes the survey results on perceived usability, on progress in learning and on the involvement of students with learning difficulties.

In all cases, the mean threshold value of 3 (the middle point in the Likert scale used) was exceeded, indicating that the platform was indeed recognized as a valuable tool by both groups of stakeholders.

In summary, the described experimental study allowed us to infer the following:

1. a statistically significant improvement was achieved, in terms of learning effect, using *Maps4Learning* with respect to traditional teaching methods
2. the learning progress was positively perceived by both groups of stakeholders, i.e., teachers and students, and
3. the proposed platform gained a satisfactory level of user acceptance.

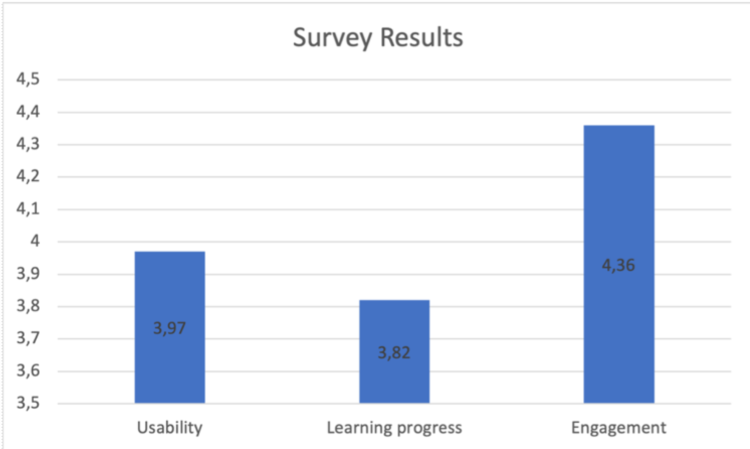


Figure 5.14: Means of perceived usability, learning progress and engagement of students with learning difficulties

5.1.7 Conclusion

The term geo-education is related to education about the world. Its goal is to let people build knowledge about human and natural worlds and interactions between them. To reach that goal, geo-educators stimulate students to engage in geographical thinking, which is based on the examination of the Earth's past and the assessment of the present, thus acquiring information from different points of view useful in studying possible future scenarios. Fostering training on geographical thinking produces an equally important priority effect, namely, that geo-education represents a new approach to facing the "geographically illiterate" state of many people, a critical issue that is due to the shift of priority from geography to the English language and STEM subjects (science, technology, engineering, and math).

Maps4Learning was also developed to pursue this goal. Using Maps4Learning, students can both follow didactic paths by searching among geoLO+ and discover further relationships among them by invoking appropriate semantic web functionality.

In this section, moreover, we have shown that also students with learning disabilities can benefit from its adoption.

The usability testing described in Subsection 5.6 has also been useful in this aim. It showed that analysing territorial elements, searching information, and studying past events through geography can help students (with and without learning difficulties) reach a deeper comprehension of world phenomena. Similarly, students who experienced the use of Maps4Learning have had the territory. This can help them memorize connections among topics more easily, thus improving their geographical literacy and preparing them for a geographical thinking model.
The developed system based on geo-information is quite universal and can be meaningfully adapted to any subject. The technology is implemented on specific examples but can be used to train students of various fields and different levels, thereby ensuring the continuity of education.

5.2 Socially Assistive Robotics combined with Artificial Intelligence for ADHD

5.2.1 Introduction

Socially Assistive Robotics (SAR) [75] aims to address critical areas of medical care by automating the supervision, coaching, motivation, and companionship aspects of interactions with vulnerable individuals from various large and growing populations, such as stroke survivors, individuals with dementia, and/or Alzheimer's disease, children with autism spectrum disorders (ASD) and others. The goal of SAR is to improve the quality of life of these individuals by providing services for daily use, therapeutic support, and a way for individual diagnosis through constant monitoring and interaction. This will allow a continuous longitudinal objective observation, with analysis and evaluation. In the field of education and childcare, SAR has been successfully applied to help children with diabetes [76], to help teachers with storytelling [77], or parents in home education [78]. In the clinical setting, Beran et al. [79] introduced a small humanoid robot for influenza vaccination. In another study [80], it was shown how human-robot interaction could be used as a means of pediatric care. Robots also have the advantage of overcoming concerns related to physical inactivity and isolation of children, typically associated with the use of computers [81] because they involve them in interactions and encourage movement [82]. Recent studies also show that a humanoid robot can be successfully used in physical rehabilitation. Indeed, it is essential to motivate children during rehabilitation, as physical therapy sessions include repetitive tasks [83]. Monotonous actions cause a reduction in the performance of children's behavior due to reduced attention. Additionally, repetition decreases focus on current activity and the motivation for repetitive exercise. Play is, therefore, a concrete way for children to communicate and express themselves in a more natural way [84] and children's motivation to participate in activities can improve if the child perceives the robot as a playmate. One of the main fields of application of current SAR research is in the clinical setting for children with behavioral disabilities. Usually, the problems of behavioral disorders concern isolated episodes or delicate and temporary evolutionary phases. However, in some cases, they can represent the prelude to subsequent psychopathological disorders. It is essential to investigate the psychological and relational aspects of the child to prevent the onset of major problems. Behavioral disorders may include attention/hyperactivity disorder (ADHD), opposition disorder (OD), or conduct disorder (CD) [90]. In particular, attention deficit hyperactivity disorder (ADHD), a neurodevelopmental disorder characterized by inattention, hyperactivity and impulsivity, has recently attracted significant research attention [94, 95]. Furthermore, working memory deficits are known to be a major factor in ADHD [95]. ADHD is one of the most common childhood behavioral disorders [96]. To develop effective intervention strategies, careful observation of this behavior is necessary in order to fully understand its causes and identify the most suitable resources to deploy. In connection with their deficits in social and communication skills, children with attention deficit often have difficulty managing their emotions. It often happens that therapists struggle to communicate with these children, as the human being is very expressive, and also the tone of voice, never constant, affects the child's perception of feeling at ease. Also, in play scenarios, ordinary toys are often unable to stimulate responses and involvement in those children. The contextual analysis carried out by us in collaboration with a diagnostic and therapeutic center led us to deepen the adoption of SAR solutions to improve the clinical contexts in which children undertake a therapeutic path. In this research, we design and develop SAR technology to support children's psychologists in therapeutic activities related to behavioral disorders, in particular ADHD. We use a toy/game robotic approach thanks to the use of Pepper. It is a humanoid robot designed by SoftBank Robotics [93] to interact with humans. This will provide the therapist with a new tool to achieve predefined therapeutic goals through social robotic interaction. Furthermore, we want to exploit the ability of artificial intelligence algorithms to derive facets, such as emotions, that we humans can escape. The Pepper-based prototype system we have developed has undergone an initial experimental study before the COVID-19 events of February 2020 at the Diagnostic and Therapeutic Center. The results of the initial study are reported here and indicate that the system is a useful tool when a therapist is trying to help, teach, communicate, or interact with patients with behavioral disorders.

Structure of the section. Related work will be presented in Subsection 5.2.2. The contextual investigation performed will be illustrated in Subsection 5.2.3. In Subsection 5.2.4, we illustrate the system and its architecture. How artificial intelligence can support therapy is explained in Subsection 5.2.5, which describes the use of neural networks for emotion recognition. The experiment is presented in Subsection 5.2.6 and finally the conclusions are presented in Subsection 5.2.7.

5.2.2 Related Work

Assistive Robotics (AR) is a broad class of robots whose function is to provide assistance to users, ranging from getting out of bed, brushing teeth, locomotion, and rehabilitation. Socially Assistive Robotics (SAR) is the class of robots that provide diverse assistance to different vulnerable populations through social interaction. The application of robotics in the treatment of children with behavioral disorders should aim to teach children basic social skills, communication, and interaction. Currently, researchers have focused on achieving these goals primarily in the area of the autistic disorder [91] [92]. Scassellati [85] indicates that robotic devices can provide quantitative data that could be useful to clinicians in diagnosing, monitoring patient progress, and comparing patients. Robots have been shown to generate a high degree of motivation and involvement in subjects, including those with difficulty or unwilling to interact socially with human therapists [86]. Some individuals with behavioral disorders even prefer robots to humans [87]. Robins et al. have shown this preference in children with autism spectrum disorder (ASD) and limited verbal skills [88]. Fridin, Angel and Azery [84] in their research studied the acceptance, interaction, and authority of robots over children in kindergarten, or in other words, they studied assisted robotics in kindergarten. They conducted their research on 19 children between the ages of 4 and 8. In their experiments, the researchers measured eye contact, face, body, expression of vocal emotions, and proxemics. After the experiment, the researchers concluded that the children accept the robot's presence, even if the robot intrudes into their personal space. Furthermore, children like to interact with the robot and can even follow the robot's instructions and accept its authority. It is also worth mentioning that although their study was conducted with typically developing children, their proposed technology was intended to be implemented for children with ADHD. Humanoid robots, which look like a human but are much less complicated than humans, could allow a child with behavioral disorders to relate better. Prins et al [97] in their research on the impact of play elements on child motivation conducted an experiment with 51 children with ADHD, aged from 7 to 12 years. All subjects were divided into two randomly selected groups and exposed to three weekly memory courses. One group had a play version of the training, while the other had a normal one. The results suggest that the children, who were assigned a game version of memory training, showed greater motivation and better performance, making fewer errors on tests. The main objective of this work is to use humanoid robotic technology together with psychological and engineering sciences, combining it with play elements, to improve the social skills of children with behavioral disorders. The designed system focuses on different aspects, such as the recognition of facial expression, patient's attention, and the emotion aroused. In addition to the beneficial effects deriving from the interaction with a humanoid robot, the data collected during the therapeutic sessions can be used to monitor the progress of the patients suffering from behavioral disorders.

5.2.3 Contextual Investigation

An empowerment-based analysis method was adopted to achieve UX requirements, as illustrated in [17]. The method, previously illustrated in

Figure 3.1, is a transformative process that starts with a contextual investigation aimed at understanding users, their behaviors, and skills within a given community and identifying potential improvements in their quality of life, which are expressed in terms of human needs.

The key factors that can contribute to the empowerment of the target users with respect to the identified needs are taken into consideration in order to formulate clear user empowerment goals. The goals involve a number of UX requirements that can be iteratively modified/extended and ultimately guide usability designers. The contextual analysis started in September 2019 and allowed us to have a vision of the most common questions regarding the problems related to people with ADHD. Our work was carried out in collaboration with a diagnostic and therapeutic center in Senigallia, which allowed us to meet and interview a developmental psychiatrist and a psychologist. During the interview, we focused on analyzing the domain of ADHD and, in particular, on understanding the associated therapeutic process. An Mp3 file was generated for each interview in order to keep only the information deemed useful for the analysis. Finally, the properly analyzed information was reported on a document so that it could be easily shared with the design team. The therapists explained that the therapeutic pathways for children with ADHD must be determined based on the specific patient, as each subject can present different symptomatology. Surely, the protagonist is never just the patient but the whole context and the people around which a multilateral path is necessary.

Stimulus control is a fundamental part of therapy, as these children often get distracted, precisely because they are constantly looking for stimuli. They reported that for a child with attention deficit predictability is a watchword, these children need to predict what happens. They agreed that the humanoid robot could become a reliable and predictable technological middleman for the child. Among the therapeutic activities for which greater attention is expected, the personalization of the therapeutic path has proven to be the innovative element of our system.

Following the analysis phase, the following functional requirements were established:

- Recognition of facial expression;
- Personalization of the therapeutic path;
- Data acquisition in real-time;

- Data processing;
- Monitoring of the therapeutic progress.

The research work we carried out capitalizing the results of the contextual inquiry aimed to provide therapists with a complete tool, capable of simplifying the collection and analysis of data thanks to an ad hoc structure, but above all, we wished to give them the possibility to experience a new approach to cognitive behavioral therapies with the help of artificial intelligence. Our goal was to demonstrate that the use of artificial intelligence and the use of innovative technologies can positively affect what therapy for ADHD is today.

5.2.4 The System

Our system was conceived to provide support to therapists in the treatment of ADHD patients by making the therapy less boring and more attractive, by fully capturing patients' attention. It is perfectly integrated with a humanoid robot equipped with a touchscreen that allows interaction with patients through our application. The system is connected to a remote server, and the patient receives customized exercises in the form of games for therapeutic purposes. The humanoid robot used is Pepper. The system, through one of the two cameras integrated on Pepper (one on the forehead and one on the mouth), observes the child interacting with the application and uses emotion recognition algorithms to capture and understand the patient's degree of attention and any difficulties. Analyzing the child's expressions and recognizing his state of mind, through artificial intelligence, helps the therapist build the right therapeutic path and the right behavioral measures to take. The data collected by the camera are combined with the results of the exercise. This combination is subsequently analyzed through evaluation metrics with automatic data analysis algorithms to propose adequate levels of therapeutic activity, and above all to provide the therapist with the degree of patient participation. This process leads to the digitization of the patients' healing path so that any improvement (or worsening) is monitored. Children perceive the presence of a robot as a game. Therefore, children are aware that they are facing therapy but do not perceive stress. Pepper becomes a reliable and predictable technological intermediary for the child, making the patient active, protagonist, and autonomous during a therapeutic session.

Architecture

As mentioned in the previous paragraph, Pepper is equipped with a touch screen and video camera. The architecture of the system is presented in Figure 5.15.

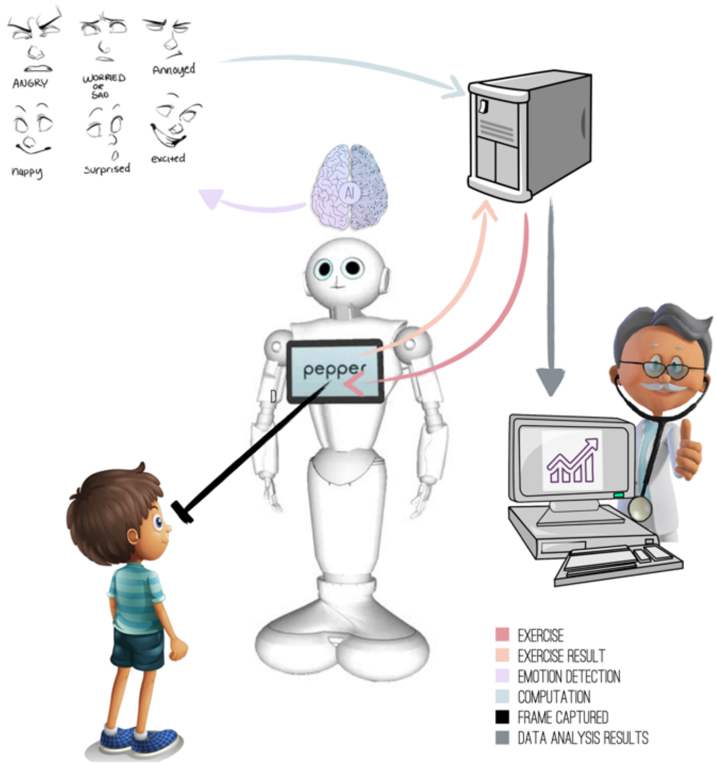


Figure 5.15: System Workflow.

During the therapy session, the therapist starts the system via their PC so that the application is launched on Pepper's touchscreen and the child can face the first exercise in terms of play. During therapy, the child will (unknowingly) be supervised by the humanoid robot's video camera, which will periodically capture frames. It was decided to carry out a sampling in order to reduce the computational load of the system. The captured frames are analyzed by an emotion classification model so that an estimate of the child's emotionality can be made. After the exercise, the results and frames are sent to the remote server which, through ad hoc metrics and algorithms, generates the next level of therapeutic activity. The combination of data is fundamental as we cannot rely solely on the outcome of the exercise but also on the patient's specific emotional state, which can be influenced by various factors (such as boredom, hyperactivity, anger). With the acquisition of the frames, we can understand the degree of attention and this allows us to analyze the session efficiently. Once the data has been analyzed, the therapist can view the patient's progress on his PC and thus understand in real-time how to continue the exercise session. Each therapy session will be recorded and will provide an estimate of the patient's progress.

5.2.5 Artificial Intelligence in the System

This subsection will show how artificial intelligence can support the therapist on this path. As illustrated in the previous subsection, during the therapy session the child will be monitored by the humanoid robot's webcams to analyze his emotions.

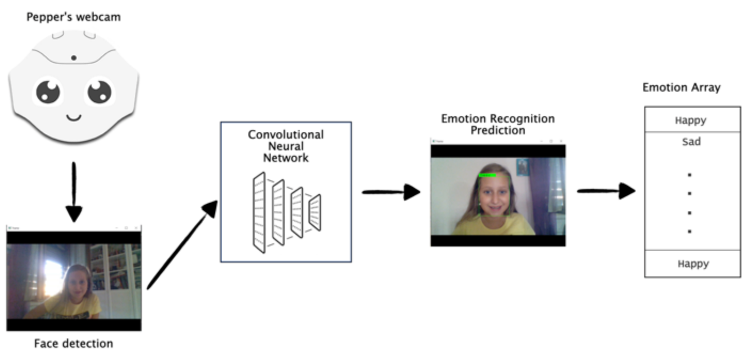


Figure 5.16: Emotion Recognition Workflow.

The classification of emotions has been carried out using convolutional neural networks (CNN or ConvNet). CNNs are machine learning algorithms that have been incredibly successful in handling a variety of tasks related to video and image processing. The captured images, therefore, pass through these algorithms which first detect a face and then recognize the emotion of that face. Finally, the trend of the captured data will be analyzed in order to obtain a qualitative measure, which will be used both to understand the progress of the therapy and to let Pepper give a message of encouragement to the child. Figure 5.16 shows a schematic representation of the various parts that make up the recognition of emotions in the system.

Face Detection

A Multi-task Cascaded (MTCNN) was used to perform the face detection. The MTCNN uses three convolutional networks in cascade, capable of recognizing the faces and positions of the five points of reference, namely:

- 2 reference points in the center of the eyes;
- 1 reference point on the tip of the nose;
- reference points at the end of the mouth.

The three phases of face detection are divided in this way:

1. in the first phase a shallow CNN is used in order to quickly capture the candidate faces;
2. in the second phase a selection of the candidate faces is made, or false positives are eliminated, through a more complex CNN;
3. in the third phase, another CNN fine-tunes the previous results, also identifying the reference points of the face.

Let's analyze in detail the three phases listed above.

1) *Proposal Network (P-Net):* This network is used to obtain the bounding boxes relating to faces, through a technique called "bounding box regression", which has the purpose of locating the bounding boxes relating to the objects sought, in this case the faces. Finally, refinements are made to combine the overlapping bounding boxes.
2) *Refine Network (R-Net):* The output of the P-Net goes to the R-Net input. R-Net reduces the number of candidates by performing bounding box regression calibration and non-max suppression to merge overlapping candidates. For each candidate, R-Net establishes whether it is a face or not and outputs a 4-element vector representing the

bounding box and a 10-element vector for locating the facial reference points.

3) *Output Network (O-Net):* The last step is similar to R-Net, but O-Net aims to describe the face in more detail and visualize the positions of the five facial reference points for the eyes, nose, and mouth.

Emotion Recognition

For the classification of emotions, we use a CNN which takes as input the faces returned in output by the MTCNN. The network is structured as follows:

1. convolutional layer: 32 output features, 3x3 kernel, stride 1, batch normalization, and ReLu;
2. convolutional layer: 64 output features, 3x3 kernel, stride 2, batch normalization, and ReLu;
3. convolutional layer: 128 output features, 3x3 kernel, stride 2, batch normalization, and ReLu;
4. convolutional layer: 256 output features, 3x3 kernel, stride 2, batch normalization, and ReLu;
5. fully connected layer:
    a) 256 features 6x6 in input, 256 neurons, ReLu;
    b) 256 input connections, one output neuron for each emotion to be classified (7), sigmoid function.

The output of this CNN will be the emotion classified for the input image; the bounding box around the identified face and a label bearing the classified emotion will be visible on the video stream.

Dataset

The network illustrated above was trained on the Toronto Face Database dataset [99]. This dataset comes from a subset of images present in the data set called Facial Expression Recognition 2013 (FER-2013) [100]. FER-2013 was created by Pierre Luc Carrier and Aaron Courville using the Google Image Search API to search for face images that match a set of 184 emotion-related keywords such as "happy", "angry" and so on. Toronto Face Database was then created on the basis of FER-2013 by Mehdi Mirza and Ian Goodfellow. The latter prepared a subset of 35887 images and mapped the keywords of the emotions in order to obtain seven broad categories:

- 4953 images for "Anger";
- 547 images for "Disgust";
- 5121 images for "Fear";
- 8989 images for "Happy";
- 6077 images for "Sad";
- 4002 images for "Surprise";
- 6198 images for "Neutral".

Capture of the predominant emotion

Analyzing Figure 5.17 it can be seen that the emotions captured during the therapy session are inserted into an array and then processed. The therapy session is divided into several established time intervals and for each interval the various emotions are captured. To carry out this operation, the underlying algorithm relies on a support array where all the emotions relating to an interval of therapy are first saved. After the interval, the array is analyzed. Since emotion is a qualitative measure, the mode is extrapolated, corresponding to the most frequent value in the observed distribution. We can have either unimodal or bimodal distributions. The mode is saved in another array and the support array is emptied to be ready for the next interval. Carrying out the analysis of emotions during the therapy phase is very important as it allows us to observe the subject's moods. In fact, the therapist can understand if the child is able to keep attention during the session (as if he/she looks away, the "distracted" label will be marked) or if he/she is getting bored and therefore some new strategy must be adopted. At the same time, Pepper uses the latest mode value to provide appropriate feedback to the little patient, pronouncing an encouraging message to the subject. To achieve that, a dictionary of encouragement phrases has been designed and created for each emotion so that Pepper may randomly "choose" what to say in order to properly encourage the child, considering the latest captured emotion.

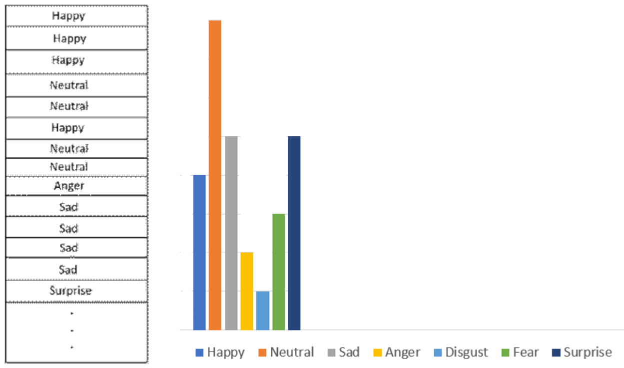


Figure 5.17: Example of a support array with related emotions Humanoid.

5.2.6 Prototype Evaluation

Our system has been designed to provide children with a playful way to approach therapy, giving the traditional work of therapists an extra boost. The prototype was tested within a therapeutic session of a group of children to assess the degree of acceptance of the humanoid robot therapist. At the outset, a laboratory experiment was conducted with a child with ADHD to make sure that adequate therapeutic stimuli came from Pepper at any time.

Preliminary experiment

The preliminary experiment happened in January 2020 and was conducted in the HCI-UsE laboratory of the University of Salerno. During the child-robot interaction, the child was accompanied by his therapist. The participant in the preliminary experiment is an eight-year-old boy named Francesco who is in the third year of primary school. From an early age, he showed difficulty managing his emotions. After several tests it emerged that Francesco has behavioral disorders; in particular, he has difficulty in staying focused even during mundane activities. Plus, he has a hard time dealing with frustrations. Initially, Francesco was afraid to face the therapies with the therapist and appeared extremely introverted. The therapist also found it difficult to attract his attention and let him do the exercises. With the introduction of the Pepper based application, Francesco was able to carry out the therapy in a fun way. During the experimental session, Francesco started playing a game, which aimed to train him to stay focused. Pepper told and illustrated a story through his display and, in real-time, captured the child's expressions. Francesco was initially distracted from discovering the origin of the sound but was intrigued by the story he observed on the display. When Pepper finished illustrating the story, he asked Francesco to answer the questions in order to analyze the little patient's degree of understanding. The data acquired from the responses plus the frames were processed to obtain the percentage of attention received. The result was displayed on the therapist's terminal. The next game was then chosen based on the information gathered and was submitted to Francesco. After the session, the therapist, thanks the data processed by the system, was able to obtain a quote on the progress of Francesco's therapy.

Figure 5.18: The child while experimenting with the system.

Assessment of the degree of acceptance

The first experimental phase continued through the observation of group therapy activities, that occurred in the center, after gaining consensus by parents. The goal was to evaluate the degree of acceptance of the introduced technological support. The group therapy allows children to manage their emotions through social relationships while collaborating with other children [89]. A group of children (N = 5), aged between 7 and 10 years, was followed by the therapist, who interacted with them through games and exercises. The observation spread over three subsequent sessions lasting an average of 50 minutes for three weeks. During the first session, the therapist introduced Pepper to the children and explained how to interact with the humanoid robot. Children immediately accepted the presence of the humanoid robot, starting to collaborate and showing a high degree of attention regarding the exercise they had to perform. After the end of the third session, a set of 10 questions extracted from the UTAUT questionnaire [98] was submitted to individual children to measure their acceptance of the system. Children could indicate their level of agreement on a five-point smiley face scale, selecting the appropriate smiley face corresponding to totally disagree (1) - disagree (2) - neither agree nor disagree (3) - agree (4) - totally agree (5). Although a rigorous assessment of the degree of acceptance, based on the UTAUT model is part of our current work, the feedback gained by the initial experiment was encouraging, promising a good acceptance and intention of using a humanoid robot for therapeutic use in children with behavioral disorders. This was also confirmed by the therapist who reported a favorable reaction to this playful Pepper-based therapy.

5.2.7 Conclusion and Future Work

In conclusion, the idea to rely on a humanoid robot in the therapeutic context has proven very attractive for both therapists and children, who are better motivated in their therapeutic path. The prototype we have described allows to build personalized therapeutic paths based on the profile of the patients with ADHD, also thanks to the support of artificial intelligence.

Chapter 6

The Influence of the User Interface Design on Accessibility

As we have seen in the previous chapters, the design of UIs for disabled users is fundamental for the accessibility. The UX is one of the fundamental aspects in the design and can make the difference in obtaining an accessible system. In my research path I have always focused on analyzing the needs, behaviors and experiences of users' digital services. I have tried to vary the methods I work with to discover them based on the needs of the project. Once we gathered this data, I identified patterns, user journeys, and a roadmap for what we created. In this way, it was possible to develop hypotheses to offer the solution that creates the greatest value for users. So I developed the design with the new insights. I am convinced that accessibility must be integrated right from the start in all digital products, so we always involve users with different needs in the design process: in interviews, workshops and use tests. We need to design accessible UIs that work best for all users, regardless of disability. Unfortunately, these activities are almost never performed by developers or requested by those who commission the system, and this often leads to the development of inaccessible systems. The best way to have a site or an app accessible to all and inclusive is to start from the beginning, that is to take into account the needs of the project already in the preliminary phase following at least the guidelines. Nowadays, mobile applications are the primary means of enabling human interaction. Being so pervasive, these applications should be made usable for all users: accessibility collects guidelines that developers should follow to include features that allow users with disabilities (e.g., visual impairments) to better interact with an application.

While research in this field is gaining interest, there is still a notable lack of knowledge on how developers practically approach the problem: (i) whether they are aware of and take into account accessibility guidelines when developing apps, (ii) which guidelines are more difficult for them to implement and (iii) which tools they use to be supported in this task. To bridge the knowledge gap on the state of the practice regarding the accessibility of mobile applications, in the following section we adopt a mixed-method research approach with a dual objective. My goal is (i) to test how accessibility guidelines are implemented in mobile applications through a coding strategy and (ii) to probe mobile developers about the problems and challenges of practical accessibility management. Key findings from the study show that most accessibility guidelines are ignored when developing mobile apps. This behavior is mainly due to developers' lack of awareness of accessibility issues and the lack of tools to support them during development.

6.1 The Making of Accessible Android Applications: An Empirical Study on the State of the Practice

6.1.1 Introduction

Mobile applications, a.k.a. apps, are nowadays used by billion users for any social and emergency connectivity [150]. The trend is tremendously and continuously increasing these days: the rise of social distancing has indeed changed the way people communicate and interact with each other [128, 142]. In such a context, mobile apps represent one of the primary means of allowing human interaction. Therefore, an ever-increasing population of users needs to interact with the functionalities they implement. This aspect does not only represent a challenge for researchers in the field of computer-human interaction (CHI) but also for software maintenance and evolution research, which is called to devise novel instruments to support developers when evolving successful mobile apps that all types of users can use [128, 153].

Applications that are not accessible or are only partially accessible are an obstacle for both individuals and businesses [108, 153]. For a single user, a hard-to-use app will either be a source of stress and frustration or be entirely sidelined in favor of a more accessible alternative [39]. For a business, the fewer users can use their mobile application, the lower the translated revenue stream will be [151]. The pervasiveness of mobile applications has led researchers to reason more and more in terms of accessibility. This trend is giving rise to a research field that aims at developing mobile apps usable by those affected by disabilities (e.g., visual impairments) [127], which represent over one billion (around 15%) of the world's population. Ensuring the accessibility of the app functionalities has become more crucial than ever [153] when people affected by disabilities are more dependent on their mobile devices.

The two main operating systems for tablets and smartphones, i.e., iOS and Android, are equipped with pre-installed accessible functions, including screen reading functionalities as in the case of TalkBack for Android. The unique needs of individuals with disabilities and their right to participate in the digital age cannot be ignored by developers. However, differently from iOS, accessibility work in Android apps is very limited [128, 153] and, as such, it is unclear to what extent developers implement universal design principles or use accessibility features in their mobile applications.

So far, most of the research on accessibility has focused on the web and mainly provided guidelines and instruments that developers can employ to implement accessible websites (e.g., [116,117,139]). On the contrary, the accessibility of mobile applications has not been examined so thoroughly [153] and, as a matter of fact, it still represents an open research challenge to face.

In the recent past, empirical investigations have been conducted to study how developers discuss the matter on StackOverflow [147] and how existing accessibility features support users with disabilities [123,149]. Nevertheless, there is still a notable lack of knowledge on the way developers approach the problem of accessibility and whether they implement the available guidelines to develop accessible applications. An improved understanding of these aspects is crucial to guide future software maintenance and evolution research efforts toward the definition of design, evolutionary, and testing techniques that can better support practitioners while developing mobile accessible applications.

In our previous registered report [112], we designed an exploratory empirical investigation into the making of mobile apps from the perspective of accessibility to bridge the current gap of knowledge concerning the relation between mobile apps and accessibility. We focused on Android for a two-fold reason. On the one hand, it has been the subject of previous accessibility studies. On the other hand, although a vast number of apps is developed worldwide on this platform, still little is known on how to best engineer the problem in Android devices—as opposite to iOS and Apple, which provide an integrated set of devices and features to handle accessibility [111]. More particularly, we discussed our plan toward this goal by defining two research questions to understand (i) whether and to what extent the available accessibility guidelines are implemented in Android applications and (ii) the developer's opinions about the matter. We sought to elicit the state of the practice and the key issues and challenges faced by developers when dealing with accessibility.

In this work, we follow up on the registered report and present the results of our study. The study has adopted a two-step methodology. We first conducted manual coding activities to quantify how existing accessibility guidelines are implemented in the context of 50 top-rated Android applications. Then, we conducted a survey study with 70 mobile developers and ten semi-structured interviews to gather insights into the issues and challenges of developing accessible apps and understand the extent to which developers are implementing accessibility support in Android apps.

The key results yielded by our study are that only a subset of the available guidelines is typically implemented in Android apps, and these mainly relate to aspects like color contrast and interactive content. While surveying developers, we could recognize a general lack of awareness of accessibility concerns; furthermore, developers indicated the lack of (semi-)automated support to control accessibility while developing mobile applications.

The findings of the work allow us to provide the research community with a set of open issues and challenges that represent the next research avenues that should be addressed to provide developers with usable accessibility tools.

To sum up, our study provides the following contributions:

1. An empirical study reporting the accessibility guidelines that are and are not implemented in Android applications, which can be used by researchers and tool vendors as a basis for either prioritizing accessibility concerns within techniques/tools or

conducting further analyses into the specific reasons why certain guidelines are more/less applied in practice;

2. Insights into the developer's perception that researchers can use to understand the underlying motivations leading practitioners not to apply accessibility guidelines as well as by tool vendors to tune current tools based on the opinions that developers have of specific accessibility concerns;
3. A list of current issues and challenges that developers face when dealing with accessibility in practice that can be useful for researchers to motivate and conduct additional studies into the matter;
4. A publicly available replication package [113] containing the data to address the goals of our study—data are anonymized whenever needed. The package includes a novel dataset reporting the accessibility guidelines implemented in a set of 50 Android apps—researchers can use it as ground truth to evaluate novel accessibility tools.

Structure of the section. Subsection 6.1.2 discusses the background on accessibility guidelines and overviews the related literature. In Subsection 6.1.3 we describe the research questions and methodology employed to address our study, while Subsection 6.1.4 reports the achieved results. In Subsection 6.1.5, we summarize the main findings, discuss the limitations of the study, and outline the key implications that our work has for the research community. Subsection 6.1.6 overviews and discusses how we mitigated possible threats to validity. Finally, Subsection 6.1.7 concludes the section and presents our future research agenda on the matter.

6.1.2 Background and Related Work

In this subsection, we define accessibility, provide an overview of the currently available guidelines to create accessible apps, and discuss the related literature, comparing it with the methodology employed in our study.

Accessibility: Definition and Guidelines Overview

According to Iwarsson and Ståhl [122], accessibility is defined as "*the quality of being easily reached, entered or used by people with disabilities*". Mobile accessibility refers to making websites and apps more accessible to people with disabilities when using smartphones and other mobile devices [148].

While the definition explicitly targets people suffering from disabilities, it is worth mentioning that accessibility is a desirable property for other groups of people as well, since these could also benefit from the availability of accessibility functionalities [126]. As such, we could generalize the concept of accessibility and state that it is the practice of making websites and apps usable *by as many people as possible* [126]. Indeed, there are two key elements driving accessibility: (1) The attention to the problems of accessing websites and apps for people with disabilities; and (2) The attention to guaranteeing *universality* of access, that is, not to exclude anyone: not only people with disabilities in the strict sense but, for example, also those suffering from temporary disabilities, those with obsolete equipment or slow connections.

Various mobile accessibility standards have been proposed, including those defined by W3C [149] and by the UK BBC Standards and Guidelines for Mobile Accessibility [109]. Within these standards, several recommendations have been formulated to provide better support for people with different types of disabilities, including motor, hearing, and vision problems. Several companies have also created their list of developer guidelines based on standards such as the Android Accessibility Developer Guidelines [102], Apple's Accessibility Developer Guidelines [103] and the IBM accessibility checklist [121]. In our work, we focused on accessibility issues in Android apps and considered the recommendations provided by Android, W3C, and BCC.

Standard controls, objects, and elements should be used to ensure a higher level of accessibility as custom controls tend not to implement accessibility as fully as standard platform controls. When standards and guidelines are implemented using non-standard techniques, there is a risk that users who depend on platform-specific accessibility features such as accessibility settings and speech output are excluded from accessing the content.

Progressive enhancement is recommended to ensure that users with accessibility settings or assistive technology enabled using older phones and platforms can access the content. This mechanism ensures that the content and features are accessible even if the experience is slightly altered. All content must be accessible and navigable using the platform navigation paradigm for assistive technology. For example, the directional controller must be supported on Android to allow TalkBack screen reader users to review and navigate the page content. Android requires that all elements be accessible from the keyboard to be accessed with a D-pad or trackball. In this respect, Android 4.0 has reduced this requirement somewhat by including an "*Explore by touch*" method.

When applications or sites block, disable, or interfere with platform-specific accessibility features or technology, users with disabilities may not be able to use the site or the app. Potential problems include zoom suppression, garbled screen content, or the inability to run assistive technology. This behavior could occur when the technology directly controls audio, video, or CPU resources preventing assistive technology from accessing these resources promptly.

Some users with disabilities may require more accessibility features because they may have more disabilities. For example, a user may be deaf and blind or may have poor eyesight and may be unable to use a pointing device or touchscreen. More modes of operation should be supported that allow users to access content according to their preferences. On Android, for example, built-in keyboard support should not prevent other standard touch events.

Accessibility: State of the Art

The topic of accessibility is rapidly gaining interest in software engineering and closely related research communities, like computer-human interaction and computer-supported cooperative work. These multidisciplinary research opportunities allow for treating the problem from different perspectives. However, although accessibility has long been investigated in the context of web applications [119, 140, 141], and many books are now available to drive developers toward the construction of accessible websites [120, 132, 137], the definition of accessibility principles and guidelines for mobile applications can be still considered a work in progress.

From an empirical standpoint, Kocielinski et al. [123] investigated the currently available accessibility features with virtual QWERTY keyboards on mobile devices, comparing them against the use of an integrated Braille notetaker. The controlled study involved visually impaired users who were called to complete tasks using the two treatments. The main findings reported were that the existing mobile support is not sufficient to assist visually impaired users. Indeed, the integration of a Braille notetaker led to better results in terms of task completion time. Zhong et al. [154] augmented the touch feature of Android devices in order to facilitate less precision to a targeted location and assist with gestures that might not be suitable for individuals with tremors, e.g., double-tapping the same location. Mehta et al. [130] conducted an empirical study—involving 12 blind users—to study the accessibility of date-pickers and proposed to augment the current capabilities of mobile devices with additional features able to support blind users. On a similar note, Xie et al. [152] focused on the understanding and improvement of the support provided for GUIs responsiveness when connecting smartphones to external displays, while Milne et al. [131] studied the accessibility of mobile health sensors for blind users: interestingly, they found that most of the accessibility problems identified could be solved with a small amount of effort.

Case studies have also been performed to study specific types of disabilities. Serra et al. [138] assessed four Brazilian government applications against the W3C guidelines, discovering that most of them were not applied. Walker et al. [149] evaluated several weather apps and their usability/accessibility for blind and sighted users: as a result, they discovered that most of the considered apps were not designed to be universally accessible. Al-Subaihin et al. [101] reported that, if appropriately used, structural HTML elements can make the functionality of TalkBack and VoiceOver similar in mobile web apps and native applications. Also, Krainz et al. [124] proposed a change in the mobile app development to support accessibility, concluding that a model-driven approach with automated code generation might potentially avoid many of the accessibility problems experienced by visually impaired users.

Some more recent studies focused on defining specific instruments and methods to support users with special needs. For instance, Araújo et al. [104] defined a manual test that can assess whether mobile audio games meet the need of visually impaired users. Similarly, Díaz-Bossini [115, 114] proposed guidelines to make mobile applications closer to the needs and requirements of older users. Park et al. [133] and Ross et al. [135] analyzed the image-based button labeling problem by focusing on missing and alternative text labels, respectively.

When turning the attention to the software engineering research community, Armaly et al. [105, 106, 107] conducted several studies aiming at comparing program comprehension tactics applied by blind and sighted programmers. Their key findings reported that, despite having different reading processes, both tend to prioritize the understanding of method signatures; furthermore, audio highlight facilities might provide additional support to blind programmers when skimming code on the web. McMillan and Rodda-Tyler [129] also reported on a didactic software engineering course setting that allows blind and sighted programmers to collaborate more effectively and improve their capabilities to share programming expertise and knowledge.

The above-mentioned papers are different from the study proposed herein. Most of them focused on accessibility from the perspective of specific users and aim at characterizing the limitations of currently available guidelines compared to the needs and requirements of such users. On the contrary, our focus is on developers and how they act when it turns to keep accessibility into account, highlighting the challenges they face when implementing accessibility guidelines and the additional instruments that they would need to build more accessible mobile applications.

The closest work is the one by Vendome et al. [147]. The authors first performed a mining study showing the limited usage of accessibility APIs in a set of 13,817 apps. Then, they focused on the developer's perspective by mining StackOverflow posts related to accessibility. From this analysis, they identified the aspects that developers mainly implement in their apps and those requiring more effort. Compared to that study, ours can be seen as complementary, mainly because we adopted a mixed-research method that allowed us to gain more precise information on the extent to which accessibility features are applied during development. On the one hand, we conducted manual analyses to verify which accessibility guidelines are implemented in Android applications. In this way, we can be more precise in indicating the specific guidelines that developers tend to care about more when including accessibility concepts. Furthermore, such an analysis leads to additional insights into whether and how developers implement critical guidelines that are relevant for users (e.g., the MUST ones). On the other hand, we directly inquired developers having two key advantages. First, we could ask more complete and specific questions on their perceptions and opinions of accessibility in practice, rather than letting them emerge from the analysis of posts. Second, we could involve a broader population of developers rather than focusing on those subscribed to StackOverflow—which might provide a limited view on the matter.

Accessibility: State of Practice

When considering the tools to assess accessibility of Android apps in practice, there exist three officially supported instruments, namely Accessibility Scanner (AS), Lint, and Node Tree Debugging (NTD).

The former takes a snapshot of an application as input. It scans each GUI component to identify accessibility issues related to content labels, touch target size, clickable items, and text/image contrast. The tool is based on dynamic analysis; therefore, it requires the app under investigation to be installed on a device. Lint is instead based on static analysis and runs as part of the Android SDK, even though it is also integrated within the Android Studio IDE. It has a broader scope than AS, as it reports micro-optimization opportunities to security, performance, and other non-functional aspects of source code. Lint also operates in terms of accessibility/usability and can detect problems related to missing content descriptions and accessibility labels. Finally, NTD is a testing tool for Android apps that can be employed to test for accessibility concerns. In particular, the tool describes how an `AccessibilityService` in the app interprets the GUI components and provides information as well as improvement recommendation related to focusable elements and their assistive descriptions.

It is also worth mentioning the existence of unofficial tools that are not integrated within the Android SDK but can provide developers with additional insights into mobile app accessibility. One of the most popular tools in this category is Enhanced UI Automator Viewer [134]: it extends the standard UI Automator in order to verify unlabeled UI elements and color contrast.

Our work has an empirical connotation and, therefore, does not aim to improve the capabilities of the above-mentioned tools directly. Nevertheless, we provide pieces of information concerning accessibility guidelines and developers' takes that are actionable for both tool vendors and researchers. The former can exploit them to tune the available tools, while the latter can devise novel approaches that better assist practitioners. We elaborate on these points when distilling the concrete implications of our work in Subsection 6.1.5.

6.1.3 Research Methodology

The goal of our empirical study is to understand the state of the practice of accessibility in mobile applications, with the purpose of providing an overview of how mobile developers currently deal with this problem as well as the issues and challenges they face when implementing accessibility guidelines. The perspective is of both researchers and tool vendors. The former are interested in gathering insights into the current state of the practice on accessibility to devise novel possible instruments to support mobile developers when dealing with accessibility in practice. The latter are interested in tuning and providing new features that might further assist developers in assessing and improving accessibility aspects in mobile apps.

Research Questions

We structure our investigation around two main research questions (**RQ**$_s$). In the first place, we seek to understand how the existing accessibility guidelines are implemented in mobile applications, namely the extent to which developers adopt these guidelines when developing their apps. This leads to our first **RQ**:

RQ$_1$. *How are existing accessibility guidelines implemented in mobile applications?*

Once established how the accessibility guidelines are implemented, we proceed with a finer-grained understanding of developers' perspective regarding the problem, particularly collecting their opinions on (i) the issues and challenges of implementing accessible applications and (ii) the tools currently supporting them. An improved understanding of those aspects would allow the research community to understand the developer's needs to support further. Hence, we pose our second **RQ**:

RQ2. *What are the developer's take on implementing accessibility guidelines in mobile applications?*

To address our **RQ**s, we conducted mixed-method research [110], combining manual coding analyses with surveys and semi-structured interviews with developers [136]. It is important to note that the empirical study has an exploratory connotation and, as such, it must be seen as an investigation whose outcome produces a number of implications that further research can exploit to generate research hypotheses.

Material and Objects

The objects of the empirical study are represented by (i) mobile applications and (ii) accessibility guidelines.

As for the former, we focus on the 50 top-rated Android apps coming from the *AndroidTimeMachine* dataset [118], which collects a reliable set of real open-source Android apps. We focus on these apps for two main reasons. On the one hand, we seek to analyze popular apps used by thousands, if not millions, users worldwide: this allows us to verify the behavior of developers who should be more sensitive to accessibility issues given the number of users they can potentially attract. On the other hand, we have to limit the number of applications to consider because of the time- and effort-intensive manual activities that we need to perform to address our research questions.

As for the latter, Table 6.1 reports the entire set of accessibility guideline categories currently available for the design of Android applications. Each category groups a set of guidelines to account for when considering a specific aspect of the mobile application (e.g., 'Audio and Video').

Table 6.1: Categories of the Accessibility guidelines for Android applications considered in the study.

Guideline	Description
Audio and Video	When creating interactive content, consider font size, style/position of controls, and how content is presented. If there is a strong need for the content to auto-play, the user

	should be aware of it and be able to set preferences to prevent it.
Design	For the best user experience, aspects such as clarity on color contrast, color and meaning, touch target sizes, content resizing, actionable elements, visible focus, content consistency, and adjustability should be properly designed.
Editorial	Use of consistent labeling for buttons, links, and headings. Work closely with editorial colleagues to maintain consistency.
Focus	How content is visually presented can impact the order in which content is coded and, subsequently, the content order and focus order in which a user experiences the content, especially users with alternative input methods such as keyboard or screen reader users.
Forms	Provide labels for all form inputs and ensure form layout and order are clear. Related form inputs should follow each other, and, if needed, the visual design should be applied to imply grouping.
Images	Avoid the use of images of text and those that do not covey key information solely through a background image.
Links	Design content layouts that facilitate grouping text and images as one link.
Notifications	Design notifications to be inclusive and perceivable by all users. Where appropriate, include other feedback and assistance cues and prompts that might guide or encourage a user when needed.
Scripts and dynamic content	Work from a basic core experience and progressively enhance this for more capable users.
Structure	The design of the interface should convey the intended structure of the content. Identify headings, containers, and landmarks, working closely with UI/UX designers if needed.
Text Equivalents	The design of the non-textual content should describe their intent and not be used to convey meanings.

For the sake of understandability, we report in Table 6.2 a list of all the guidelines and their description. The identified accessibility guidelines are a set of technological agnostic best practices for mobile web content, hybrid, and native apps. The guidelines are based on the content requirements of three de-facto standard providers of information on the matter, i.e., the Android developer's documentation, the World Wide Web

Consortium (W3C) community, and the BBC Standards and Guidelines academy. We combine the three providers to create a comprehensive set of accessibility guidelines organized into 11 categories. Some guidelines are marked as 'MUST' or 'MUST NOT' (highlighted in green and red in the Table 6.2, respectively) depending on whether their implementation must be ensured or avoided. These guidelines are associated with specific, objective criteria that can assess their presence in a mobile app and can be implemented using the available mobile device technologies. As an example, audio-only or video-only content MUST be accompanied by a text transcript. Indeed, audio-only or video-only content would not be available to users who cannot hear or see, respectively. Transcripts of the audio and/or video must be provided as equivalent to allow users with disabilities to use and interact with an application properly. On the other hand, audio MUST NOT be played automatically unless the user is warned and can control the audio. This can be indeed harmful to users relying on assistive technologies such as speech output software to hear the page content, i.e., the auto-play feature would enable multiple audios at the same time, hence not allowing the user to properly interact with the app.
Other guidelines are marked as 'SHOULD' or 'SHOULD NOT' (highlighted in blue and orange in the Table 6.2, respectively): these represent less critical, yet important accessibility principles that should or should not be implemented. These guidelines are generally less testable and can be more subjectively interpreted by a user. For instance, separate volume controls should be provided for background music, ambient sounds, narrative, and editorially significant sound effects. Instead, narrative audio in games or interactive media should not talk over or conflict with native assistive technology.

Table 6.2: Accessibility guidelines for Android applications considered in the study.

Categories	Guidelines	Description
Audio and Video	Alternatives for audio and visual content	Alternative formats, such as subtitles, sign language, audio description and transcripts, must be provided with embedded media when available.
	Autoplay	Audio must not play automatically unless the user is made aware this will happen or a pause/stop/mute button is provided.
	Metadata	Relevant metadata should be provided for all media.
	Volume control	Separate volume controls should be provided for background music, ambient sounds, narrative and editorially significant sound effects.
	Audio conflict	Narrative audio in games or interactive media should not talk over or conflict with native assistive technology.
Design	Colour contrast	The colour of text and background content must have sufficient contrast.
	Colour and meaning	Information or meaning must not be conveyed by colour only.
	Styling and readability	Core content must still be accessible when styling is unsupported or removed.

	Target touch size	Touch targets must be large enough to touch accurately.
	Spacing	An inactive space should be provided around actionable elements.
	Content resizing	Users must be able to control font sizing and user interface (UI) scaling.
	Actionable elements	The colour of text and background content must have sufficient contrast.
	Visible focus	When focused, all actionable and focusable elements must have a visible statechange.
	Consistency	A user's experience should be consistent.
	Choice	Interfaces must provide multiple ways to interact with content.
	Adjustability	Interactive media, especially games, should be adjustable to accommodateuser ability and preference.
	Flicker	Content must not visibly or intentionally flicker or flash more than three times in any one-second period.
Editorial	Consistent labelling	Consistent labelling should be used across websites and native applications as well as within websites and applications.
	Indicating language	The language of a page or app must be specified, and changes in language must be indicated.
	Instructions	When needed, additional instructions should be provided to supplement visual and audio cues.
Focus	Focusable elements	Interactive media, especially games, should be adjustable to accommodateuser ability and preference.
	Keyboard trap	There must not be a keyboard trap. Content and focus order must be logical.
	Content order	Actionable content must be navigable in a meaningful sequence. Focus or context must not automatically change during user input.
	Focus order	Actions must be triggered when appropriate for the type of user interaction.
	Changing focus	Alternative input methods must be supported.
	Appropriate triggers	Interactive media, especially games, should be adjustable to accommodateuser ability and preference.
	Alternative input methods	There must not be a keyboard trap. Content and focus order must be logical.
Forms	Labelling form controls	All form controls must be labelled.
	Input format	A default input format must be indicated and supported.
	Form layout	Labels must be placed close to the relevant form control, and laid out appro-priately.
	Grouping form elements	Controls, labels, and other form elements must be properly grouped.
Images	Images of text	Images of text should be avoided.
	Background images	Background images that convey information or meaning must have an acces-sible alternative.
Links	Descriptive links	Link and navigation text must uniquely describe the target or function of the link or item.
	Links to alternative formats	Links to alternative formats must indicate that an alternative is opening.
	Combining repeated links	Repeated links to the same resource must be combined within a single link.
Notification	Inclusive notifications	Notifications must be both visible and audible.
	Standard operating system notifications	Standard operating system notifications should be used where available andappropriate.
	Error messages and correction	Clear error messages and a way to correct errors must be provided.
	Feedback and assistance	Non-critical feedback or assistance should be provided when appropriate.
Scripts and dynamic content	Progressive functionality	Apps and websites must be built to work in a progressive manner that ensures a functional experience for all users.
	Controlling media	Media that updates or animated content must have a pause, stop or hide control.
	Page refreshes	Automatic page refreshes must not be used without warning.
	Timeouts	A timed response must be adjustable.
	Input control	Interaction input control should be adaptable.
Structure	Unique page/screen titles	All pages or screens must be uniquely and clearly identifiable.
	Headings	Content must provide a logical and hierarchical heading structure, as sup-ported by the platform.
	Containers and landmarks	Containers should be used to describe page/screen structure, as supported bythe platform.
	Grouped elements	Controls, objects and grouped interface elements must be represented as a
Text Equivalent	Alternatives for non-text content	Alternatives must briefly describe the editorial intent or purpose of the image, object, or element.
	Decorative content	Decorative images must be hidden from assistive technology.
	Tooltips and supplementary information	Tooltips must not repeat link text or other alternatives
	Roles, traits and properties	Elements must have accessibility properties set appropriately.
	Visual formatting	Visual formatting alone must not be used to convey meaning.

Subjects

The subjects of the study are developers of Android applications. We have involved both original and external developers of the applications that are the objects of the study. While the former can provide us with feedback on implementing the accessibility guidelines in their applications and their view of the problem, surveying a larger population of developers may provide additional insights into the issues and challenges of dealing with accessibility in practice. We collected participants' background and demographic information to understand the representativeness of our results. We followed the sampling strategies defined in literature [143] to define a sample that meets our goals. More details on the recruitment strategies applied in our empirical study are reported in the next subection.

Execution of the Empirical Study

In this subsection, we report the methodological details that describe the execution of the empirical study—we discuss the two **RQ**s independently.

RQ₁. Accessibility guidelines in practice. To address **RQ**1, we manually tested the considered applications to verify the implementation of accessibility guidelines—this strategy allows us to interact with an app and its accessibility services directly, much like a user would normally do. Overall, the guidelines to be verified were 54, divided into the 11 categories presented in Table 6.2. To perform such a manual test, we adopted a closed-coding strategy [146]: this is a systematic methodology that, in our case, involves the analysis of all the graphical UIs of an application and the subsequent labeling of the guidelines implemented as functionalities of the app, starting from a pre-established coding scheme, which is our case is represented by the set of guidelines available for Android applications.

More specifically, we have created a data extraction form, implemented using an Excel sheet, to facilitate the verification of the guidelines. For each of them, the form contains four pieces of information: (i) the name of the guideline to verify, (ii) the procedure to follow to discover whether the guideline is implemented, e.g., activate the notifications to verify that they are both visible and audible, (iii) the excepted visual/audio effect to observe in case the guideline is implemented, and (iv) the outcome to add once evaluating the guideline. The extraction process of an app was conducted by me and consisted of the following steps:

<u>Step 1 - Download.</u> The author downloaded the app from the Google Play Store on a Huawei Y5 smartphone.

Step 2 - Guideline identification. The author selected the next guideline to test and the corresponding instructions from the data extraction form.
Step 3 - Activation of accessibility features. Depending on the selected guideline, she has activated the accessibility function required to verify it if needed. Otherwise, she has gone straight to the next validation step.
Step 4 - Element identification. The author has been exercising the app to identify the feature connected to the accessibility guideline, if available. For instance, this concerns the identification of the app's media in case the guideline refers to *'Audio and Video'* accessibility aspects. If identified, the author proceeded with the next step; otherwise, she went back to Step 2 and continued with another guideline.
Step 5 - Verification of the guideline. Once the element is identified, the author determined if the guideline is implemented in the app. If so, she annotated the data extraction form by putting, in the row corresponding to the considered guideline, a *'true'* in the fourth column. Otherwise, she annotated the column with *'false'*.

Using the above-described methodology, we have collected 50 Excel sheets, one for each application considered. These were later analyzed to address the first research question.

***RQ_2* - Surveying mobile developers.** To address **RQ_2**, we conduct a survey study aiming at gathering insights regarding accessibility concerns from a broad audience of Android developers. The survey is composed of three main sections—we report the full list of questions in Table 6.3. The first one presents a total of nine questions about accessibility and how developers consider it in practice. We ask questions on the relevance of the problem, i.e., how important is accessibility for the participants, what reasons would make them willing to implement accessibility features in their applications, and whether they are aware of the existence of guidelines to make an app accessible. Afterward, we continue with questions more related to the implementation of accessibility guidelines. In particular, how often developers implement them in their applications, how difficult they are to apply, and why. Finally, we ask participants to report up to five challenges they usually face when dealing with accessibility concerns and report whether and which tools they currently use when performing the task.

Table 6.3: Full list of survey questions.

n.	**Question**	**Evaluation Criterion**
Section I. Accessibility of Android applications.		
1	In your opinion, how relevant is the problem of accessibility?	Likert scale from 1 (Not at all) to 5 (Very important)
2	Please, tell us more about your answer.	Open answer.
3	What makes you willing (or not) to implement accessibility guidelines?	Multiple Choice - it includes the *'Other'* option.
4	To what extent are you aware of the accessibility guidelines available for Android applications?	Likert scale from 1 (Not at all) to 5 (Very much)
5	To what extent do you follow accessibility guidelines when developing Android applications?	Likert scale from 1 (Not at all) to 5 (Very much)
6	Can you please rate how difficult it is for you to implement the following guidelines?	Likert scale from 1 (Not at all) to 5 (Very much) for each guideline.
7	For each guideline rated by the participant with 3/4/5 to question #6: 7.1. Can you please explain more about what makes it harder for you to implement the guideline?	Open answer.
8	What are the top 3 problems of dealing with accessibility in Android development?	Open answer.
9	What are the top 3 to 5 challenges you face when dealing with accessibility concerns?	Open answer
10	Do you use any tool to verify the implementation of accessibility guidelines?	Open answer
Section II. Further opinions.		
11	If you have further comments on the accessibility of Android applications and how you deal with the problem, feel free to comment more on it.	Open answer.

12	If you would like to receive a summary of our research results, please leave your e-mail.	Open answer.
13	Would you be willing to participate in a follow-up interview to better discuss the problem of accessibility in Android development?	Yes/No.
Section III. Background.		
14	What is your current job?	Multiple Choice - it includes the *'Other'* option.
15	What if your gender?	Multiple Choice - it includes the *'Other'* option.
16	How do you rate your expertise with programming?	Likert scale from 1 (Very poor) to 5 (Very high).
17	How do you rate your expertise with Android programming?	Likert scale from 1 (Very poor) to 5 (Very high).
18	What is your company size?	Multiple Choice - it includes the *'Other'* option.
19	What is your team size?	Multiple Choice - it includes the *'Other'* option.

In the second part of the survey, we allow participants to provide us with additional insights and feedback. They can leave their e-mail address if they are interested in receiving a summary of our findings and can express their consent to a follow-up interview to discuss the problem of accessibility in practice further. Finally, the third section of the survey concerns background information that we collect to understand better our sample of developers and possibly analyze the generalizability of our results.

The survey is designed to last 15/20 minutes and is created using a Google survey module. Before releasing the survey on a large scale, we ran a pilot with two developers of our contact network to evaluate if the survey is short and understandable enough to reduce the risk of having a low response rate and be appropriately filled out. Based on the pilot results, we have changed the text of some questions, add/remove some of them, or change the response type to make the questionnaire easier to understand or quicker to be compiled.

To gather insights from the original developers, we extract the e-mail addresses from the Github repositories of the considered applications. Then, we invite developers to fill the survey out, first asking whether they would like to participate. In other words, we recruit only volunteers to avoid privacy issues or other developer concerns. In addition, to gather insights from external mobile developers, we advertise the online survey using the personal social network accounts of the authors (i.e., Facebook, Twitter, and LinkedIn). It is worth remarking that we were aware that the reliance on social media might negatively impact selecting a valid sample. Therefore, we integrated social media with other sources to ensure the quality/completeness of the information gathered when addressing **RQ**$_2$, still relying on a large sample of developers for our study. On the one hand, we involved additional developers from our private contacts (e.g., former University students or other practitioners who are currently mobile developers). On the other hand, we advertised the survey on a specialized practitioners' blog such as Reddit to acquire information from developers who have a solid knowledge of programming [145]. In particular, Reddit contains more than 100 different subreddits dedicated to Android development that we exploit to potentially reach thousands of Android developers. We track the source used by participants to access the survey to better comment on the validity of the sample. To further stimulate the participation, we allow participants to indicate a non-profit organization of their choice to which we would donate 2 USD for the research against COVID19.

The answers are anonymized to preserve the privacy of participants. As a result of this study, we have a clearer view of the relevance of accessibility in practice and the major challenges developers face when dealing with the problem. Based on the answers received to question #13, we also planned follow-up semi-structured interviews with Android developers. Their main goal is to clarify ambiguous or contrasting answers received during the survey and to have a better picture of the current practices, issues, and challenges experienced by developers when dealing with accessibility in Android environments. From a practical perspective, we summarize the survey results to the interviewees and ask them to comment on the answers from which we could not derive a definitive outcome. The semi-structured interviews are conducted through Skype, last 30/40 minutes, and are transcribed for further analysis.

Data Analysis

Once we gathered data from the closed-coding exercise and the survey study, we proceeded with their analysis.

As for $\mathbf{RQ}_1$, we first provided descriptive statistics on the extent to which accessibility guidelines are implemented in the sample of Android applications. We computed minimum, mean, median, standard deviation, and maximum number of accessibility guidelines implemented in the considered apps. Secondly, we provided a finer-grained overview of each specific category of guidelines. We discussed (i) to what extent each of them is present in the sample by reporting descriptive statistics, i.e., minimum, mean, median, standard deviation, and maximum number accessibility guidelines for each category, and (ii) the relative and absolute frequency of implementation of the guidelines included in each category. Then, we focused on the guideline requirements, i.e., 'MUST', 'MUST NOT', 'SHOULD', and 'SHOULD NOT': in this case, we aimed at understanding whether developers take them into account, e.g., if the 'MUST' guidelines are implemented in the considered apps. Finally, we verified the relation between the guidelines and the type of application considered. We grouped the apps by category, as provided by the Google Play Store, and we computed descriptive statistics to grasp if some categories are more prone to accessibility concerns.

As for $\mathbf{RQ}_2$, we first described the background of survey participants by discussing the answers they provide in Section III of the survey. This detail allowed us to understand the sample of developers and reason about the generalizability of our findings. In the second place, we distinguished the analysis procedures to use when considering closed and open questions. The former was analyzed employing statistics: we plotted and discussed the distribution of answers provided by participants through the Likert scale evaluations. The latter was subject of an iterative content analysis: in particular, we conducted the following methodological steps:

<u>Step 1 - Microanalysis.</u> I went through the content of the participant's answers and the possible semi-structured interviews. She split sentences using standard text separators (e.g., commas) and assigned initial labels to each sentence: these labels represent the main concepts discussed by participants. Then, the initial assigned labels were validated and feedback provided on how to improve them, for instance, by proposing to aggregate

two semantically similar labels. When this step was accomplished, we computed a measure of agreement between the labels assigned by me.
Step 2 - Categorization. I used the suggestions and feedback received in the first step to conduct a second iteration over the labels assigned. This step resulted in a set of themes deemed important by participants when addressing each survey question.
Step 3 - Saturation. Final agreement was reached concerning the names and meanings of each label. This step led to a *theoretical saturation*, i.e., the point in which no further labels are required because the existing ones already correctly represent the concepts expressed by the study participants.
The themes coming from this data analysis procedure concern each specific open question posed in the survey. We discussed each theme and provided qualitative insights by presenting the most significant answers for a specific theme. In addition, when analyzed the answers to questions #8, #9, and #10 of the survey, we also provided statistical data reflecting the number of times a specific issue/challenge/tool named by the participants, hence providing a kind of prioritization of the concerns and tools that developers have concerning the problem of accessibility.

Verifiability and Replicability
The data generated from our study are made persistently publicly available through Figshare [113]. In particular, we release raw data about the accessibility guidelines implemented in our dataset, the survey structure, the anonymized responses, and all scripts used for data analysis.

6.1.4 Analysis of the Results

This subsection presents the results of the empirical study, which we discuss by addressing the two research questions independently.

RQ_1. Accessibility guidelines in practice
In the context of RQ1, an iterative manual verification was performed to evaluate which accessibility guidelines were implemented within the mobile applications that are the subject of the study. As explained in Subsection 6.1.3, these were evaluated for their general applicability verifying whether each guideline was implemented or not in the application.

According to the results obtained, we observe that no application implemented all the guidelines. This result was somehow expected, other than reasonable, since the accessibility guidelines do not represent fixed rules. Their applications must therefore be considered based on the specific application domain and context. Nonetheless, we noticed that most of the guidelines were applied at least once in our dataset: as such, we can report that mobile developers sometimes take care of them.

Looking deeper into the considered apps, we observed that 94% of the guidelines (51/54) could be assessed, i.e., the apps contained features that might have enhanced through the implementation of accessibility mechanisms. Other guidelines were instead not applicable. For instance, this is the case of the *'Metadata'* guideline, which cannot be currently applicable in Android. Indeed, it does not support a mechanism for navigating between containers within native applications. A user can only navigate through a single item at a time. As a consequence, the *'Containers and landmarks'* guideline is also not applicable. Finally, Android does not provide tooltips or additional hint text other than `aria:contentDescription`. Therefore, the *'Tooltips and supplementary information'* guideline is not applicable—however, users can still obtain tooltips by long-pressing on icons in the Action Bar.

Based on the considerations above, we considered the number of guidelines that could be assessed while measuring the total amount of guidelines implemented within the considered applications. For instance, let consider the Budget application. In this case, 30 guidelines were assessable and, among these, ten were violated (i.e., 1/3 of them).

Figure 6.1 shows the percentage of guidelines implemented by the developers of the 50 apps considered. In particular, the x-axis represents apps (i.e., each bar is an app) while the y-axis reports their accessibility coverage level (i.e., the percentage of guidelines implemented). From the figure, we could immediately understand that the number of accessibility guidelines implemented in the considered apps was typically low, with a minimum of 24% and an average of 41%. In the best case, 63% of the guidelines were implemented. Consequently, we could first conclude that, overall, mobile developers tend not to implement accessibility guidelines while developing and maintaining their apps, even though they might have the chance to do that.

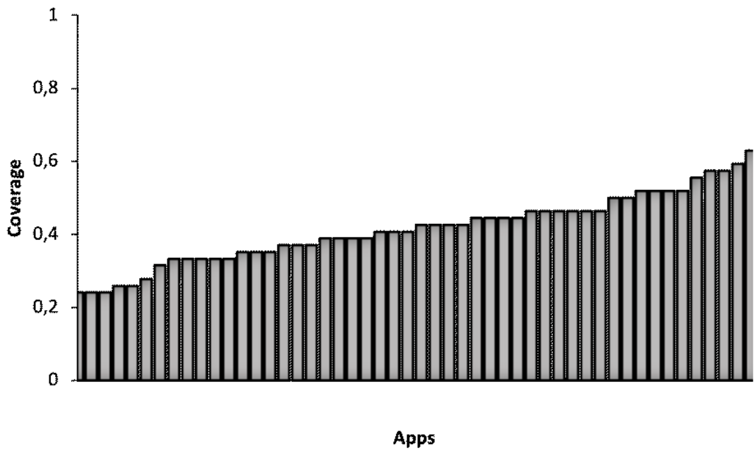


Figure 6.1: The child while experimenting with the system.

Figure 6.2 provides a more detailed overview of which are the individual accessibility guideline categories implemented in our dataset.
From the figure, we could observe that some guideline categories seem to be more considered by developers. For instance, the *'Images'* category showed the highest ratio between the number of accessibility guidelines implemented and those actually assessable (37/41). This category refers to evocative visual content that allows the user to interpret the meaning of the features implemented in the applications. The highest implementation ratio is somehow reasonable and expected since the use of images to reflect the content of a piece of text is something that human beings typically do to convey meaningful messages [144]. As such, independently from the knowledge that developers might have of the specific guidelines ruling the usage of images, we could have expected to observe the category to be highly implemented.

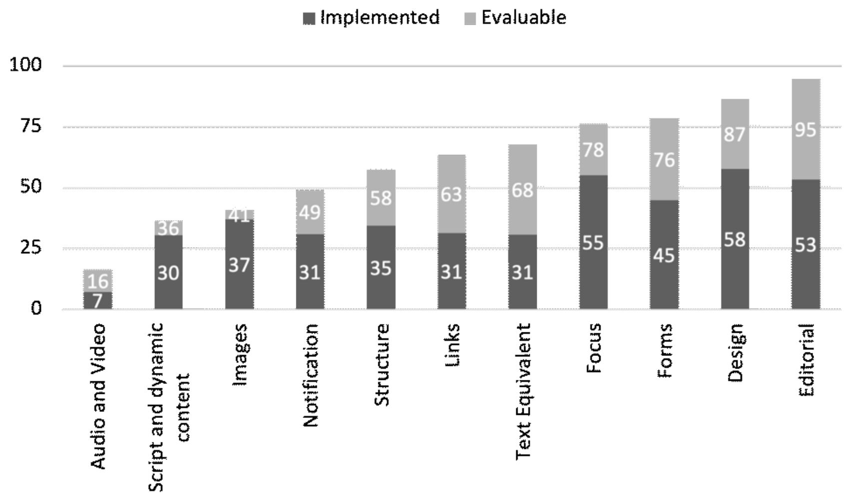


Figure 6.2: Percentage of guidelines assessable against percentage of guidelines actually implemented in the considered mobile applications, grouped by category.

The category having the highest ratio of guidelines violated was *'Audio and Video'* (only 44% of the guidelines were implemented over the total assessable). This indicates that mobile developers do not often take care of the characteristics that the interactive content should have in terms of font size, style/position of controls, and so on. In this case, it is likely that developers are not keen nor aware of the need to put themselves in the others shoes and offer functionalities that facilitate users to interact with the app.

As for the other guideline categories, we could delineate a general trend from the analysis of Figure 6.2. A notable percentage of guidelines were violated: while we could not speculate on the reasons making them less implemented at this stage, we sought to understand this aspect further in the context of our second research question.

When lowering the granularity of our investigation to the individual guidelines within each category, we could first observe that the highest amount of guidelines implemented pertained to *'Design'*, with an average of 28.8 out of a maximum of 50, i.e., around 29 applications contained implementation of accessibility guidelines related to the design of the application. More specifically, the *'Style and readability'* design guideline, with a value of 47, appeared to be the most implemented, followed by the *'Spacing'* guideline with a value of 45. On the contrary, the least evaluable category was *'Audio and Video'*, which was also the least implemented with an average of 3.6. In this case, the 'Volume control' guideline was implemented only two times out of 50 applications.

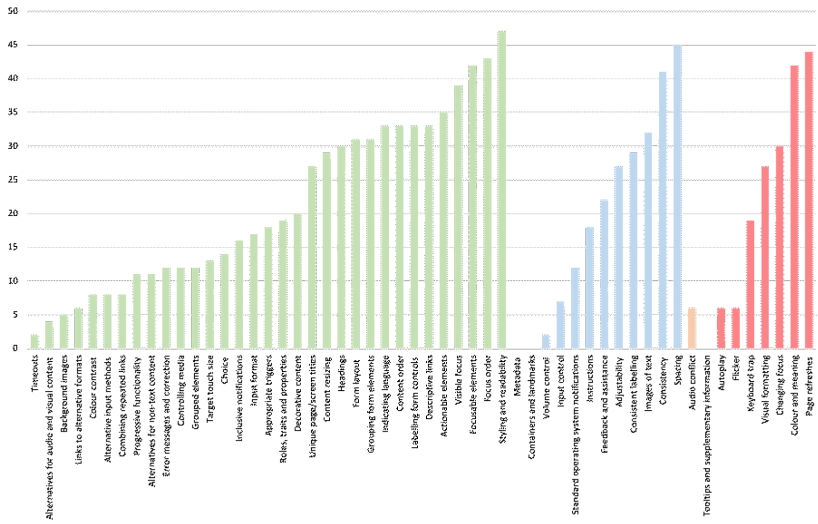


Figure 6.3: Distribution of accessibility guidelines implemented, grouped by requirements.

When it turns to the guidelines requirements, i.e., 'MUST', 'MUST NOT', 'SHOULD' and 'SHOULD NOT', Figure 6.3 shows the distribution of guidelines applied grouped by their associated requirement—we visualize the distribution in ascending order based on the number of guidelines abide by. Such an assessment was intended to understand whether developers consider the guideline requirement while deciding which accessibility guidelines to apply.

However, as depicted in the figure, we could not find any relation between those requirements and the application of the guidelines, meaning that developers do not likely consider whether a certain guideline must/should or not be applied. This aspect is further considered in our **RQ**$_2$, where we inquired mobile developers on their expertise on accessibility concerns.

Finally, we verified the relationship between the implementation of guidelines and mobile app types, as indicated by the category assigned by the Google Play Store to the considered applications. During this analysis, we merged two of the Google Play Store categories, namely *'Music & Audio'* and *'Video Players & Editors'*. These indeed refer to very similar applications that deal with multimedia content. For the sake of our analysis, we decided to assign these apps to a macro-category called *'Audio & Video'*, which better aligned with the homonymous accessibility guideline category we considered.

Figure 6.4 depicts the descriptive statistics for the ten categories identified. For the sake of interpretability, we plot bar charts reporting the minimum, mean, median, and maximum number of accessibility guidelines implemented in each category of the apps in our dataset.

As shown, the *'Books'* category has the greatest value of implemented guidelines—this result is consistent for all the reported statistics. Nonetheless, we did not observe a significant difference for the guidelines implemented in other categories. Consequently, it seems that the application category does not consistently influence the developer's willingness to adopt accessibility mechanisms. The only exception to this general discussion is the *'Game'* category.

As the reader can see, this category has the lowest amount of guidelines implemented when considering all statistics. A possible reason behind this result may be that developers are more focused on making an application attractive for users rather than accessible.

To conclude the discussion, from the first research question, we could observe that developers often tend not to implement accessibility guidelines. This result is neither connected to whether the guideline must/should (not) be applied nor the application category. In our discussion, we also identified possible reasons behind the way developers operate in terms of accessibility. The next research question aims to elicit directly from the developer's experience the main problems and challenges they face when dealing with the problem of accessibility of mobile applications.

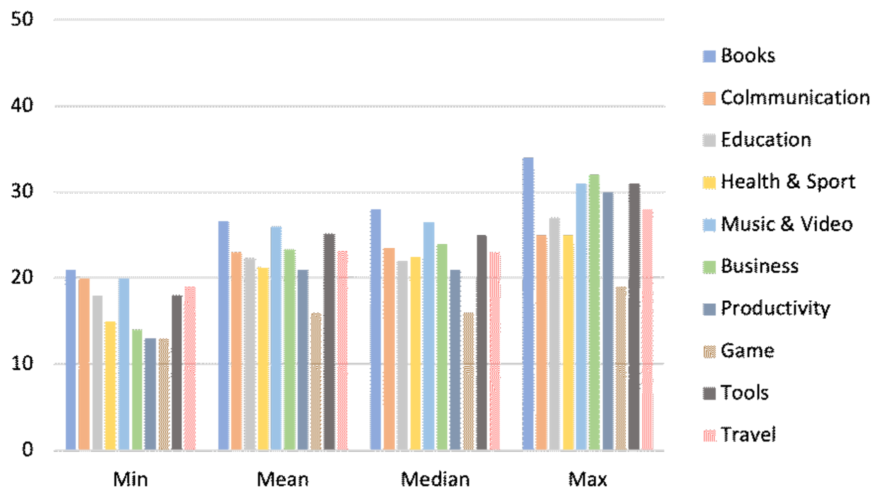


Figure 6.4: Descriptive statistics reporting the implementation of guidelines by mobile app categories, as indicated by the Google Play Store.

Main findings for RQ_1

Developers do not often consider accessibility when developing their apps, i.e., only 41% of guidelines is implemented, on average. The most respected category is *'Design'*, whereas the least amount of implemented guidelines pertains to *'Audio and Video'*. The *'Images'* category has the highest ratio between the number of accessibility guidelines implemented and those actually assessable, while developers seem to have more difficulties with the guidelines connected to interactive content. Finally, there is no relation between the application of guidelines and categories of the apps.

RQ_2. The developer's take on accessibility

While the previous research question allowed us to understand, from a quantitative perspective, how accessibility guidelines are implemented in Android applications, we could only delineate some conjectures on the reasons why developers decide to apply or not specific guidelines. The survey study conducted in **RQ_2** aimed at shedding light on the developer's perspective of the accessibility problem. In particular, our survey was answered by a total of 75 developers, of which 63 male, four female, one transgender person, and seven who preferred not to declare it. Given the nature of the dissemination mechanisms, we cannot estimate the response rate—we are not aware of how many potential developers were reached over the various social networks and blogs considered. Nonetheless, we can report that 65% of the participants had access to the survey via personal contact, 11% via Telegram, 9% via Reddit, 8% via Facebook 5% via Twitter and 1% via Tandem.

Developers' background

Figure 6.5 shows the background of our participants. Among the 75 respondents, 60% (45 participants) declared to have a high level of experience in programming, and 43% (32 participants) had high experience in Android programming. About 42% of the participants (mainly) work as developers and 24% (18) work in medium-sized companies with more than 100 employees. From these descriptive statistics, we can say that our sample consisted of various developers with sufficient programming experience and whose opinions may provide us with valid and reliable insights on how they deal with the accessibility problem. Furthermore, 15% of participants work in a large team of 10 to 200 people, 24% within a team of 5-10 people, while the majority (43%) in a small team (i.e., 2-5 people).

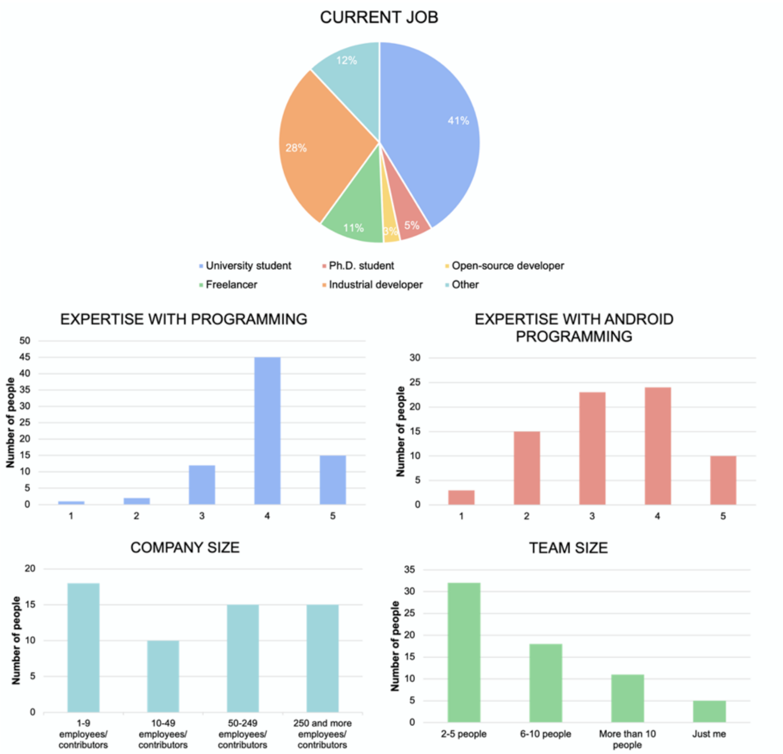


Figure 6.5: Background of our participants.

Relevance of the problem

In the first part of the questionnaire, we aimed at understanding how developers consider accessibility in practice when developing mobile apps. Figure 6.6 shows how many participants responded with values between the minimum and maximum of the Likert scale to the first question of the survey. As shown, most of the participants (60%) considered the problem of accessibility as *very important* for mobile app development. At the same time, only three developers (4% of our sample) claimed that this is negligible. Hence, as expected, we can confirm that accessibility is a significant concern for most developers. As an example, one of the participants commented:

#26 -All users need to have the same possibilities.

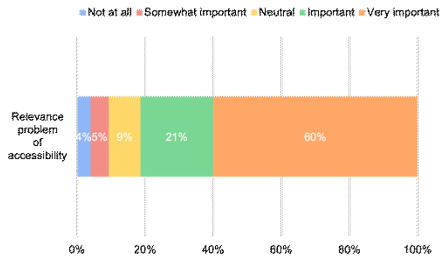


Figure 6.6: Relevance of the accessibility problem from the developer's perspective.

Reasons for implementing (or not) accessibility guidelines

Personal ethics (39%) and the widening of the potential user base (37%) were mentioned as the main reasons making developers willing (or unwilling) to implement the accessibility guidelines, as depicted in Figure 6.7. A smaller percentage of participants (17%) declared that their applications are dedicated to people with disabilities and, therefore, they have to follow accessibility guidelines. Only 7% of the developers reported that their companies implement policies aimed at ensuring mobile accessibility. Looking at these results, we can observe that the main driver for the implementation of accessibility guidelines is the personal willingness of developers to provide additional functionalities that would enable the usage of the app to a wider variety of users.

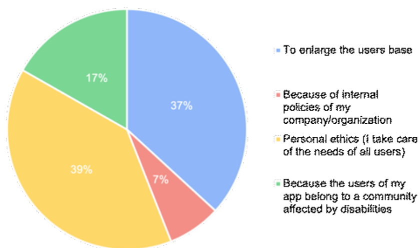


Figure 6.7: Results for Question n.3 - What makes you willing (or unwilling) to implement the accessibility guidelines.

By matching these observations with the poor implementation of accessibility guidelines discovered in **RQ**$_1$, our findings seem to suggest that more work should be done on motivating developers and stimulating their willingness to apply accessibility guidelines while developing their apps. This result is confirmed by the analysis of questions #4 and #5 of our survey (Figure 6.8): although most of the participants have a medium-high knowledge of accessibility guidelines, a large majority of participants apply them only in a few cases.

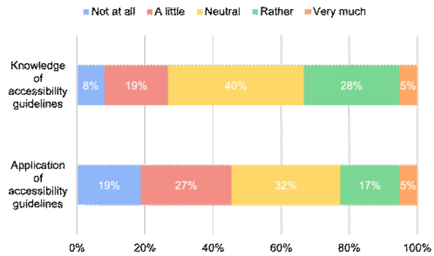


Figure 6.8: Results for Questions n.4 and n.5 - Awareness and implementation of accessibility guidelines.

The opinion on the individual categories

Once investigated the general behavior of developers, our survey aimed at seeking their opinions and experience with the implementation of the individual categories of guidelines. Figure 6.9 reports the results obtained in this respect. It is worth remarking that our survey allowed participants to express additional opinions on the factors making hard the implementation of specific guidelines. While the comments left were helpful in most cases, others were not clear enough to elicit those factors. In these cases, we took advantage of the follow-up semi-structured interviews conducted with the participants to discuss them further. In particular, eight developers left their email addresses in response to question #13 of our survey and were later interviewed. For the sake of readability and conciseness, we discuss the results by guideline, reporting data from the survey and accompanying the discussion with the insights coming from the semi-structured interviews whenever needed.

In the first place, our analysis revealed that most developers encounter little or any difficulty in implementing the guidelines. As a matter of fact, for the *'Audio and Video'*, *'Forms'*, and *'Text equivalent'* guidelines, no developer found it very difficult to implement the category of guidelines. This result was quite surprising and, at the same time, interesting: while developers consider the vast majority of accessibility guidelines as easy (or fairly easy) to implement, they are still reluctant to implement them—hence, confirming that the problem is strictly connected to the willingness or, perhaps, a limited understanding of how significant might be implementing those guidelines for users with disabilities. However, there are exceptions to this general conclusion.

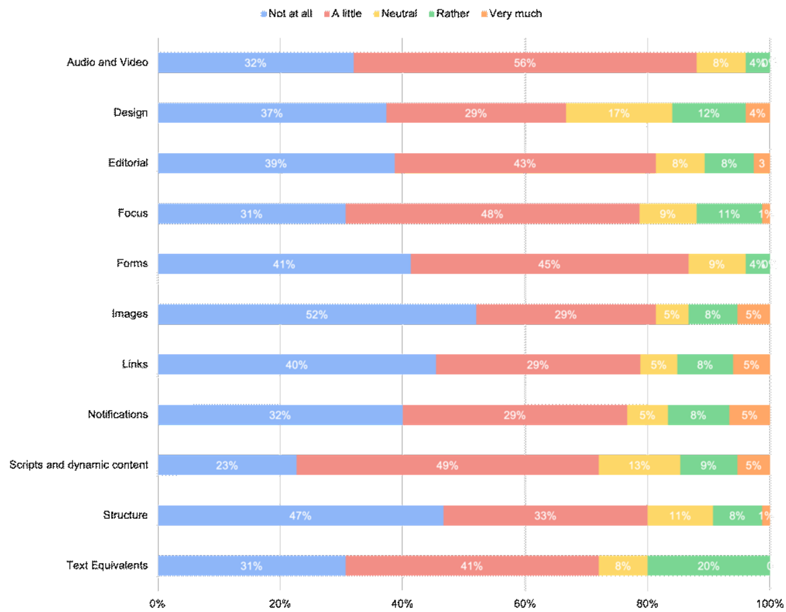


Figure 6.9: Results Questions n.7 and n.7.1 - Difficulty in implementing the specific guidelines.

Design. Diving deeper into the individual guideline categories identified by at least one developer as hard to implement, three participants declared the 'Design' category to be a difficult one. From the survey analysis, we found that developers mostly focus (or need to focus) on the aesthetics of the application rather than on its accessibility. As such, they would need more precise guidelines for implementing the design principles that address accessibility concerns. This consideration is also common to other developers, who commented, for instance, saying that:

#23 -Standards are not defined precisely.

The semi-structured interviews confirmed that the guideline definitions are sometimes vague and not easily interpretable, potentially complicating their implementation. For this reason, Interviewee #3 claimed that an improved accessibility guideline should provide informal definitions and concrete examples on how to integrate them within various types of applications. This tooling would help developers to learn by examples, simplifying and speeding up the implementation process.

Editorial. 82% of the participants considered the implementation of the guideline to be little and not difficult. The remaining 14 developers, instead, rated this guideline as complex or very complex to apply. By looking at the open comments left by those participants, we could understand that the guideline is not complex. However, the time required for implementing it is too high and/or there is a lack of resources available. Two developers commented, indeed, that:

#57 -A lot of businesses just don't have enough resources to comply with all such consistency across all clients.

#63 -It is very tedious and takes a significant amount of time to label interactive elements and images.

Focus. The vast majority of the participants did not consider this guideline hard to implement. Only one participant justified the complexity of the implementation by saying that:

#17 -I saw many examples of buggy focusability in android development and sometimes providing for example good keyboard navigation on the screen is really really hard due to these bugs.

In other words, the developer suggested that the Android APIs to use for implementing this guideline may sometimes be defective, increasing the time and effort required to ensure the focusability of the app. Once again, this seems not to be strictly connected to the guideline itself but to the surrounding environment required to implement accessibility guidelines.

Images. 13% of the participants reported that ensuring the accessibility of images can be hard or very hard. By looking at the open answers provided, we could understand that this is mainly due to the role played by images in the graphical UI of mobile applications. One of the developers commented as follows:

#56 -Difficult as the images play an important graphic role.

Unfortunately, the comment could not provide us with a clear understanding of the key issues connected to the implementation of the guideline. For this reason, we have further elaborated the question in our follow-up interviews with the developers. From the discussion, it turned out that this is due to the lack of proper usability skills, which might lead to complex implementations of this guideline. Indeed, optimizing the use of images while keeping accessibility under control is not easy, as implementing the guidelines risk affecting the overall aesthetics and look-and-feel of the app, potentially creating more issues than benefits. This result suggests that the definition of recommendation approaches that may suggest how to best implement the GUI of mobile apps by balancing usability and accessibility might be a nice addition for mobile developers.

Links. This category presents a very similar situation as for the previous guideline, with a lower percentage (40%) of users who did not rate the implementation difficult. Nonetheless, we were surprised to see some open comments like the one shown below:

#72 -Didn't even know.

By discussing this further in the semi-structured interviews, we understood that some developers were not even aware of the existence of this guideline. In particular, Interviewee #6 explained that most of the developers with whom s/he worked were not only unaware of accessibility guidelines but also unable to find helpful information on the web. As a result, the implementation difficulty is sometimes due to the retrieval of appropriate information on usability and accessibility, making it hard for developers to correct the problem in their apps.

Notifications. As shown in Figure 6.9, 5% of the respondents to our survey (4) declared that implementing notification-related accessibility guidelines is hard in practice. In this case, the comments left in the open answers were already clear enough to elicit the main reasons behind this result. One of the participants stated:

#70 -Notifications in Android are often very finicky and device-dependent so we can't expect them to conform reliably to specific behaviors.

As reported, the mechanisms enabling notifications in Android are not always easy to use. This challenge may be related to the different notification types that can be visualized differently, and developers should apply the guidelines to the specific implementations. Among all the guidelines discussed so far, this seems to be the most source code-related one: in this sense, the definition of smart mechanisms may potentially address the problems raised by the participants, e.g., dynamic wizards helping developers to select the most appropriate set of notifications along with the accessibility rules to implement.

Scripts and dynamic content. As for this category, 14% of developers rated its implementation as fairly or very difficult. As stated by one of the developers, the implementation of this guideline:

#64 -Would harm the simple user interface, adding effort and making the development of simple apps unfeasible.

As reported, some developers would not find enough benefits from the implementation of this guideline, as it may have possible negative effects on the aesthetics and overall usability of the app. While discussing this issue further in the semi-structured interviews, Interviewee #4 commented on the statement above by discussing the exemplary case of *Progressive Functionality*. This fine-grained guideline is related to creating graphical UIs that allow users to do actions in a stepwise, progressive manner. S/he reported that:

Interviewee #4 - The fundamental problem with progressive content is that it takes what was previously one requirement (do x when user enters the screen) and turns it into three or four requirements (do x or do y or do z based on condition A B or C). These types of multi-requirements are actually quite difficult to communicate about (conversations will be full of confusion and miscommunication). Naturally they increase the workload but if the requirements are clear it is actually not all that much work. The problem is getting the requirements clear in the first place. Additionally, it can be very difficult for the quality team to actually exercise all of these different pathways, so it increases the work there too. Most developers are going to advise against these type of progressive interfaces and instead promote that you create one interface that works for all situations, perhaps with the ability to have some content hidden by default.

In other words, the implementation of scripts and dynamic content enforces developers to increase the number of app requirements, requiring them to create more test cases or develop more code review activities. In addition, the definition of the requirements might be a source of miscommunication that possibly leads to the introduction of undesired defects. Based on our results and discussion, implementing this guideline seems to be related to multiple aspects and levels of expertise covering the entire software life cycle. As such, developers might be more reluctant to consider its actual application.

Structure. The last guideline we discuss is concerned with the structured content. More particularly, one participant reported that:

#74 -Sometimes it is difficult to maintain the original structure and refactoring is required.

This answer was later discussed in the semi-structured interviews. The main point here is concerned with the moment in which accessibility is considered. If an application is not designed to be accessible in the first place, refactoring for accessibility can be effort-prone and costly since it may imply the re-design of entire screens of the app. In addition, Interviewee #7 pointed out the lack of automated mechanisms and integrated tools that can provide accessibility feedback directly within the IDE. In her opinion, the availability of these tools might help to address the problem of accessibility from the start, hence avoiding costly refactoring that is rarely implemented.

The challenges of accessibility

The last part of our survey was reserved for the issues and challenges of accessibility in practice. Table 6.4 reports the top-5 list of challenges identified when analyzing the participants' answers.

Table 6.4: 5 top challenges to face when accessibility problems are encountered.

n.	**Challenge**
1	Raise awareness of accessibility and the needs of disabled users.
2	Standardize accessibility guidelines during app implementation.
3	Implement accessibility guidelines without compromising the aesthetics and functionality of the application.
4	Involve more disabled users during application development.
5	Raise awareness of companies and customers on the accessibility problem of universal inclusion.

As shown in the table, two of these challenges are related to accessibility awareness. In the first place, developers expressed their inability to understand the exact needs of users with disabilities: this represented the main, most popular challenge mentioned in our survey. Participants also reported that one of the challenges concerns the involvement of those users during the development: this is made complicated by identifying the right target audience and the mechanisms to use for involving them. For example, developers mentioned the complexity of requirement elicitation, which naturally leads to ineffective solutions. Our participants (and our interviewee) suggested using a user-centered methodology to develop mobile apps, where real users are surveyed and involved throughout the application development process.

At the same time, our participants mentioned the awareness of companies and customers. When discussing this further, the developers told us that only a small percentage of users need accessibility in mobile applications and, therefore, companies tend to underestimate the problem. In addition, accessibility is often considered non-portable and essential only for large apps. Perhaps more importantly, all interviewees raised another social issue connected to accessibility: they argued that their customers sometimes dictate not to follow accessibility guidelines to get the product up and running in the least amount of time. Consequently, they found it difficult to convince the customers of the additional time required to implement a universally usable product.

Additional challenges are more on the technical side. On the one hand, standardizing accessibility guidelines is related to defining techniques that help developers implement the guidelines while the app is still under development: accessibility should be considered a first-class citizen. On the other hand, developers need mechanisms that allow for a trade-off between the aesthetics of the graphical UI and its accessibility. As discussed in Subection 5, these represent calls for software maintenance and evolution researchers.

The current accessibility support

As a final step of the survey, we asked participants if they use tools to verify the implementation of the accessibility guidelines. 77% of the respondents do not use tools. The remaining 13% reported the usage of the Accessibility Scanner app, the Google Play pre-launch report, and the definition of beta tests with users. Our results clearly show that it is not very common for developers to rely on tools to verify the accessibility of the apps being developed. Three of the respondents reported that the missing usage of tools is because they provide minimal information while lacking a more detailed and careful analysis of both the different categories of disabilities that should be considered and the specific guidelines that should be implemented. In addition, the interviewees highlighted that different types of devices must be taken into consideration; often, different brands/models of devices behave differently. Therefore, the implementation of some accessibility UIs requires complex logic to include all target devices. Last but not least, some brands' battery-saving policies may affect or even suspend accessibility services.

Main findings for RQ_2
Although considered necessary, developers rarely implement accessibility guidelines for various reasons, including personal choices, company policies, vague standard definitions, socio-technical concerns, lack of automated support, and others. The currently available tools provide a limited amount of information and limited automation. The involved participants argued that more innovative and sophisticated mechanisms would be helpful to understand the user's needs and make their apps more accessible since the beginning of the development process.

6.1.5 Discussion and Implications

The results achieved when addressing our research questions provided several insights that need to be further discussed and implications for both researchers and tool vendors, which we elaborate on in the following.

Finding 1 - Accessibility problems are widespread in apps. One of the main results coming from our analyses refers to the poor adoption of accessibility guidelines in practice. We discovered that, in each app, most of the guidelines were not considered by developers in the implementation phase or incorrectly applied. Furthermore, our results show that *'Design'*, *'Script and Dynamic Content'*, and *'Text Equivalents'* are the most problematic categories of guidelines to implement. The feedback received by developers through surveys and follow-up semi-structured interviews allowed us to elicit the main reasons behind such difficulties. We conclude that several technical aspects connected to the implementation of accessibility guidelines should deserve further attention in the future. On the one hand, more research is needed around this subject. We hope that the reported results might serve as a basis for stimulating software engineering researchers to proactively consider novel mechanisms to support mobile developers. A critical challenge here concerns the definition of (semi-)automated techniques that might support developers while evaluating the level of accessibility of their applications and developing accessible mobile apps. On the other hand, the lack of standardized methods to apply accessibility guidelines represents a challenge for the designers of these guidelines. While there exist catalogs that suggest the general universal design principles to apply, the results of our study pointed out the lack of practical recommendations and patterns to follow during the development. This latter challenge is, in particular, one of the main points of our future research agenda: the definition of *accessibility design patterns* will be the next challenge to face.

Finding 2 - Lack of developer's awareness of different types of disabilities. Not only the problem of accessibility is widespread in practice, but also developers generally lack knowledge of the different disabilities of users and how they should be considered from a software engineering perspective. This clear issue represents a call to action for researchers working in multiple fields, from medical branches to software engineering and human-computer interaction. Indeed, the results of our study revealed the need for multidisciplinary research that can formulate novel instruments and methods to increase the sensibility of developers around the matter. Educators should consider the various forms of disabilities while forming the next generation of mobile developers. At the same time, tool vendors and app store providers should introduce instruments to solicit developers in taking care of accessibility issues. For example, we envision the introduction of some form of accessibility rewards that would rank accessible mobile apps higher in the ranking provided by the Google Play Store. Last but not least, this is also a matter that involves developers, who might want to consider additional user-centered usability testing phases with a more diverse set of users, which would enable the identification of accessibility criticalities before issuing new releases.

Finding 3 - Software engineering meets human-computer interaction. As pointed out by several developers involved in $\mathbf{RQ}_2$, the accessibility perspective of the mobile engineering process is all but defined. Developers argued the difficulties they experience with accessibility requirement elicitation and management, other than the challenges concerning design, refactoring, and testing for accessibility. Besides the actions that developers might take individually, e.g., the user-centered usability testing processes mentioned above, our findings are more general and recall the need of *software engineering methods for accessibility*. This aspect is one of the main implications of our work. We argue the definition of symbiotic methods that would allow human-computer interaction and pure software engineering to more closely collaborate to define unified processes that may enable improved engineering of mobile applications that take accessibility and usability into account. The rise of a tighter collaboration between the two research communities would enable the definition of combined methods that optimize the quality of mobile apps simultaneously with usability, possibly leading to the production of better software.

Finding 4 - On the accessibility technical debt. Among the various points raised by developers in the context of $\mathbf{RQ}_2$, we identified traces of a new type of technical debt [125] connected to the management of accessibility concerns. In particular, one approach to mitigate accessibility issues is to plan for accessibility early in the design phase rather than managing it as an afterthought at the end of the development phase. In other words, our results allow us to define an *accessibility technical debt* as the accumulated long-term cost caused by choosing an early, sub-optimal UI or design solution. So far, this unknown debt has been neglected by the research community. We argue that additional analyses and researches would be needed and desirable to devise new accessibility technical debt detectors and refactoring recommenders. These tools could work at different granularities (e.g., in the IDE rather than during code review) and stages (e.g., designing early prototypes, such as UI sketches, refactoring existing functionalities).

6.1.6 Threats to Validity

Several threats might have affected the validity of our results and the conclusions drawn. This subsection discusses how we mitigated them.

Threats to Construct Validity. Possible issues in this category refer to the methods used to set up the empirical study. The first point of discussion concerns the dataset of mobile apps exploited in the study. We focused on 50 top- rated Android apps coming from the AndroidTimeMachine dataset [118]: the decision was made to collect and analyze data of real open-source Android apps that are used by thousands of users worldwide.
Another possible concern is connected to how we elicit the set of issues and challenges developers face when dealing with accessibility. We opted for a

survey-based investigation through which participants could share their past/current experiences with the matter. Of course, those participants performed the task in a remote setting. While we could not completely avoid the lack of conscientious responses, the follow-up data analysis allowed us to verify the meaningfulness of the answers. It could have possibly detected data to be removed for the sake of reliability. In addition, we performed semi-structured interviews to complement the survey study and discuss questions for which we obtained contrasting outcomes.

Threats to Conclusion Validity. In the context of **RQ**$_1$, we conducted an iterative manual analysis to verify the presence of accessibility guidelines. Similarly, in **RQ**$_2$, we conducted an additional coding procedure to analyze the developer's open answers. In both processes, I was the main responsible. Nevertheless, to double-check my actions and mitigate possible misinterpretation, was constant intervention whenever necessary. This continuous collaboration and the level of agreement reached make us confident of the results reported in the study.
When reporting the results, we also computed statistics to enlarge the discussion and provide more detailed insights into the accessibility problem in mobile applications. Nonetheless, we recognize that further replications of our study would still be desirable.

Threats to External Validity. Threats in this category refer to the generalizability of the conclusions drawn from our study. We targeted 50 top-rated open-source mobile applications and involved 75 survey respondents. As for the former, the considered apps belonged to different domains and had various characteristics that enabled us to investigate accessibility under different perspectives. As for the latter, the participants had previous experience and expertise with Android development, hence being able to provide us with meaningful insights into the problem of accessibility. In any case, we cannot claim the generalizability of our results to mobile applications having a different connotation, e.g., closed-source or industrially developed apps. As such, further replications of our study would be desirable and are already part of our future research agenda.

6.1.7 Conclusion

The growing popularity of mobile devices, coupled with constant technological improvements in the field, has led to an increasing number of mobile applications. In this context, usability aspects play a pivotal role both when considering the design and implementation phases. Although usability is already recognized as a crucial aspect of mobile development, only a few studies analyzed the accessibility of mobile applications. In this research, we aimed at advancing the state of the art by analyzing (1) the extent to which a set of known accessibility guidelines are applied in practice and (2) the developer's take on the accessibility problem.

We conducted a quantitative investigation of 50 Android applications finding that most of the guidelines available are not implemented within applications. Afterward, we interviewed 75 developers, conducting eight semi-structured interviews, showing that accessibility is perceived necessary, but several socio-technical barriers often prevent developers from applying the accessibility guidelines. The overall output of our research identified several challenges that must not only be considered by the software engineering research community but also by experts of other disciplines like human-computer interaction, medicine, and others.

Part II
Situation Awareness

The second part of the thesis focuses on the collection of studies that aim to obtain an improvement of the User eXperience through the design of User Interfaces able to keep the situation awareness factor high.

Some challenges of UX design for control rooms for monitoring fleets of drones, by an operator or by a team of operators, are faced by proposing a usable UI capable of guaranteeing high levels of SA. Furthermore, this second part deals with the design of a UI for Industry 4.0, providing a set of guidelines to help design UI for monitoring industrial processes able to guarantee high levels of SA and improve UX.

Chapter 7

UX challenges for life critical decisional interfaces

Part of my research also involved the use of unmanned aerial systems, commonly called drones. It is worth highlighting that today the demand for applications that require the use of fleets of drones and a related man-machine interface, which affects the UX, is growing enormously, to ensure better performance and reliability. A major cause of accidents involving drones is ground operator error, due to poorly designed UIs. Therefore, the interest in human factors arises with the aim of contributing to the maintenance of high safety standards of flight missions through the systematic application of usability principles during the development of interfaces. Section 7.1 reports an ongoing study aimed at designing Ground Control Station interfaces for controlling multiple drones, with the aim of improving operator performance and increasing the chances of mission success. In particular, human interaction with the aircraft can be based on various levels of automation, thus requiring high workloads, when the aircraft is manually controlled, and low workloads, when the aircraft autonomously performs missions. From the analysis of the human factors that influence missions, it has also emerged the need to design innovative UIs that also include new levels of automation, as they could result in more reliable systems capable of reducing the negative impact of wrong human behavior during the mission. Consequently, these interfaces could improve the UX of operators. In particular, for a single operator who must effectively control multiple aircraft, the SA of the operator is a fundamental component of the man-aircraft system. Subsequently, the human-machine interface, presented in Section 7.1, was redesigned for a Ground Control Station used to remotely operate a fleet of drones but in a collaborative environment, therefore by a team of multiple operators. In such a collaborative environment, one of the main challenges of UX interface design has been to maximize the situational awareness of the team, shifting the focus from the single operator to the decision makers of the whole group. Of particular interest was testing the hypothesis that shared displays can improve team situational awareness and therefore overall performance. The experimental study presented in Section 7.2 shows that there is no difference in performance between shared and unshared displays. However, in tests where unexpected events occurred, teams using shared displays maintained good performance, while teams using unshared displays performance was reduced.

7.1 Human Machine Interface Issues for Drone Fleet Management

7.1.1 Introduction

The number of civil applications where Unmanned Aerial Systems (UAS's) can be both effectively and efficiently used has increased in recent years, thanks to a considerable reduction in costs. The main use of drones is for surveillance but other areas of application in science and research can be found and many private companies have been created to offer paid services for the most disparate business applications. There are applications that can be carried out with a single drone, but recently applications are emerging which require the use of fleet of drones (both formation and swarm flights) and a related human machine interface to ensure better performance and reliability. The present work reports on an ongoing study meant to design Ground Control Station (GCS) interfaces for vehicle remote control, with the goal to enhance operator's performance and, consequently, increase both the probability of a mission success and the safety.

UAS's can be remotely piloted (RPAS - Remotely Piloted Aircraft System) or they can automatically fly. As a single powerful UAS equipped with a large array of different sensors is limited to a single point of view, in the last years, the multi-UAS paradigm seems to be a more suitable approach for many applications since it can guarantee:

- Multiple simultaneous interventions: a multi-UAS team can simultaneously collect data from multiple locations;

- Greater efficiency: a multi-UAS team can split up in order to efficiently cover a large area, optimizing available resources;
- Complementarity: a multi-UAS team can perform different tasks with growing accuracy;
- Reliability: a multi-UAS team assures fault tolerant missions by providing redundancy and capability of reconfiguration in the case of a failure of a vehicle;
- Safety: for a permit to fly, a fleet of mini-UASs is safer than a single great and heavy UAV;
- Cost Efficiency: a single vehicle to execute some tasks could be an expensive solution when comparing to several low cost vehicles.

Safety is a very important aspect and in recent years the number of accidents in which drones are involved is high [203]. Studies reveal that a major cause is ground operators' error, due to poorly designed UIs [204]. The human interaction with the aircraft can be based on varying levels of automation thus requiring high workloads, when the aircraft is controlled manually and low workloads, when the aircraft autonomously performs missions. The level of implemented automation, whatever it is, affects the performance of the operator. From the analysis of human factors affecting RPAS operations, the need has emerged to design innovative UIs, also encompassing new levels of automation, which could result in more dependable systems able to reduce the negative impact of erroneous human behaviors on the success of a mission.
This section intends to set out a vision on how to design a human machine interface for a GCS (Ground Control Station) of a fleet of drones performing a cooperative mission of persistent surveillance. The design challenge for the fleet management interface is to capitalize all the available heterogeneous amount of data and to set up a mechanism to allow efficient and safe interaction between the pilot/operator and the fleet. As the interface should also enable the management of scientific data, coming from the payload, one of its key objectives is to let the user perform different choices, which may support flexibility and real-time decisions, all for the sake of safety and mission success. Obviously strong time and usability constraints exist. A User Centered Design methodology and a

participatory approach are being adopted to design appropriate interaction scenarios, involving multiple operators of GCS, who may provide feedback on the appropriateness of certain choices. In the future such UI will allow to train pilots also for off nominal scenarios.

Structure of the section. Subsection 7.1.2 gives an overview of the specific system for managing the fleet of drones which perform persistent surveillance mission. Such a system needs a HMI which should allow the pilot to properly manage all the related parameters to perform the mission. Subsection 7.1.3 gives an overview of the state of the art in terms of interfaces for drone piloting, and summarizes the interface requirements for the fleet management system. Subsection 7.1.4 discusses future work to be carried out following the user centered design process.

7.1.2 The Fleet Mission

The persistent surveillance consists in monitoring "n" geographic targets with a fleet of "m" drones (with m<n) for an indefinite time. It is achieved by maximizing an objective function for each individual aircraft that assures for the fleet both a maximum global and the equilibrium of Nash, this formalization of the problem is based on game theory. In order to perform a safe test of this complex system consisting of "n" aircraft and a monitoring and control GCS, it was necessary to realize a multi-aircraft simulator capable of modeling the dynamics of the aircraft, the behavior of the autopilots and the communication channel.

Therefore the design, development, implementation and testing of a dedicated GCS interface was required. Such GCS can run the simulation of the fleet mission as well as the experimental missions involving real drones.

7.1.3 HMI for Drone Piloting

Several human factors have been identified as relevant factors in drone operations [205]. These may in turn depend on several system characteristics:

- Level of automation / allocation of functions
- Sensory isolation
- Bandwidth / Latency
- Information / bore saturation

- Simulation illness
- Spatial disorientation.

Looking at the previous list, we observe that the negative influence that human factors may have on the performance of the operator and, therefore, on the whole system can be mitigated if appropriate UIs are designed, to effectively support communication between the operator and the aircraft during mission planning and execution.

This is true not only in the field of UAS drone remote control systems but more generally whenever there is interaction between a man and an automated system, because often human error depends on it [206].

The role that the interface plays within the UAS system becomes even more critical if we consider that the interface is supposed to be the only communication means between the human operator and the aircraft. For that reason it should be designed to overcome or compensate the limitations coming from their physical separation.

Our approach, will focus on the previous system characteristics which can turn out to negatively impact on human correct behavior using them as driver for the HMI design. The system will be used in simulated off-nominal scenarios, related to those specific system characteristics (e.g. sensor isolation, communication latency, ...) collecting usage data by the pilot.

The pilot, involved in the requirement definition phase, will have a crucial role in using the system in off-nominal scenarios thus providing information about the capability of the HMI to support him.

State of The Art Analysis of GCS User Interfaces

A survey was carried out on the current state of the GCS UI for remote control of drones.

There are tools supporting the user to set up the system, procedures to ensure correct configuration and correct operation of the related software, procedure for planning the mission (design of the route that the drone should perform), tools allowing users to monitor flight in real time and tools to analyze and debug flight mission data [210].

Some GCS interfaces already communicate to users crucial information concerning the operation of drones: incompatibility alerts, usage instructions or safety checks, that prevent the user from basic errors. Some

apps already provide learning / support material. Most of them appear in mobile apps targeting basic drone users.
However, such apps were conceived for use in missions only, focusing on video recording or mapping [209].
Besides displaying graphs, some apps have rerun mission features that are very interesting for visually representing the data collected during flights [211]. However, they often result in hard-to-use UIs for mission planning activities as well as for alerting and recovery actions and for drone flight data analysis. Moreover, most of those interfaces do not support fleet mission management and hence do not properly address issues related to situation awareness and safety.
Our case study, involving a fleet of drones which perform cooperative missions, builds upon a more complex system especially conceived to support remote control of multiple drones operating synergistically. For such a system the design of a usable interface raises even more challenges and becomes paramount for the properly management of a fleet mission.

The List Interface Requirements

In order to experiment with different solutions to the problems outlined above, a HMI has been developed for the fleet management system following a user centered design process and adopting a participatory approach.
Since it must enable the control of more drones simultaneously, one of its characteristics is to maximize the situation awareness of the operator. The following are some representative requirements of the related GCS which the HMI has incorporated:

- management of 4 types of mission parameters: communication parameters, scenario parameters, algorithmic parameters, and mission targets;
- loading of mission parameters from files, which the operator can subsequently modify if necessary;
- ability to locate the position of take-off of the fleet, by typing numerical coordinates or by clicking on a point on the suitable map;
- choice for each aircraft of the committed target, i.e. the identifier of the first mission target to be reached;

- saving of the mission parameters in a file for future loading, with a preliminary verification of the completeness and consistency of the data, allowing for any necessary correction;
- 2D map visualization of:
    - all the targets (the committed ones with different colors) or a subset of them appropriately identified;
    - all or a subset of the aircraft with appropriate identification (also informing on distance and time from the committed target);
    - all the trajectories or just a subset of them;
    - all future aircraft trajectories or just a subset of them;
- additional simultaneous display of:
    - Artificial horizon and compass, for all aircraft or only a subset of them;
    - Graph of aircraft altitude over time, for all aircrafts or only a subset of them (with different identification);
    - 3D view (which also displays any obstacles such as structures, vegetation, et cetera) where one can choose the point and angle of view with the same functionality as the 2D view also with observer on board the chosen aircraft;
    - Telemetry of the main info for all the aircrafts or just a subset of them: navigation, status of the aircraft, data payload.
- warnings and/or errors display, with appropriate visual/auditory clues;
- reception of telemetry data of the aircraft via MavLink protocol or through COM port or (if using a mission simulator) on proprietary protocol;
- mission parameters transmission to the aircraft, with a preliminary verification of the completeness and consistency of the data, allowing possible corrections;
- transmission of the mission start signal that is activated by a special key, after verifying that all the aircrafts are ready to take off and connected;

- transmission at any time of the end of mission signal, that is activated by a special key always available to the operator.

7.1.4 Future Work for the Ground Operator in the Design Cycle

The approach followed for the design of the interface and, therefore, for the realization of the whole system, was centered on the operator (User-Centered Design).

Based on UCD principles, designing the interaction between the ground operator and the fleet of drones implies to analyze, understand and identify the interaction specifications that make the human-aircraft communication safe as well as functional to the objectives of the mission, through a usable interface which is effective, efficient and satisfactory for the operator. The interface design will therefore combine the knowledge on the fleet management system functional requirements with the knowledge on human physical and cognitive constraints.

7.2 Improving Human Ground Control Performance in Unmanned Aerial Systems

7.2.1 Introduction

Unmanned Aerial Systems (UAS), commonly known as drones, are being employed in more and more civilian settings, from crowd-monitoring activities, to road traffic control, to agricultural crop monitoring. The number of civil applications where UAS can be both effectively and efficiently used has increased in recent years, thanks to a considerable reduction in costs. UAS can be remotely piloted (RPAS—Remotely Piloted Aircraft System) or they can automatically fly. As a single powerful UAS equipped with a large array of different sensors is limited to a single point of view, in recent years, the multi-UAS paradigm seems to be a more suitable approach for many applications requiring the observation of wider areas. One of the emerging areas of the civilian use of UAS is public safety and services [212].

Moreover, the multi-UAS paradigm can guarantee:

- Multiple simultaneous interventions: a UAS fleet can simultaneously collect data from multiple locations
- Greater efficiency: a UAS fleet can split up to efficiently cover a large area, optimizing available resources
- Complementarity: a UAS fleet can perform different tasks with growing accuracy
- Reliability: a UAS fleet assures fault-tolerant missions by providing redundancy and capability of reconfiguration in the case of a failure of a vehicle
- Safety: for a permit to fly, a fleet of mini-UAS is safer than a single big and heavy UAV
- Cost efficiency: a single vehicle to execute some tasks could be an expensive solution when comparing to several low-cost vehicles.

Safety is a very important aspect and in recent years the number of accidents in which drones are involved is high [213]. Studies reveal that a major cause is ground operators' error, due to poorly designed UIs [214, 215].

Therefore, applications have emerged which require the use of a fleet of drones and a related human-machine interface to ensure better Situation Awareness (SA) by the operator. In fact, the lack of situation awareness (SA) is also often cited as a major cause of accidents [216] and therefore understanding SA and improving it with adequate system design is an important research target [217]. As a result, much work has been published on the subject.

In a previous section, we set out a vision on how to design a human machine interface for a GCS (Ground Control Station) of a fleet of drones performing a mission of persistent surveillance [218]. The design challenge for the fleet management interface was to capitalize all the available heterogeneous amount of data coming from drones and to set up a mechanism to allow efficient and safe interaction between the ground operator and the fleet, thanks to an enhanced SA. We focused on the design of a UI which would allow a single remote ground operator to control a fleet of multiple drones. In the resulting prototype, scientific data coming from drones' payload are timely transformed into useful information on the interface in a way that supports an operator's SA, reducing the risk of being overwhelmed by information which may hinder his/her decision-making

capability. The prototype can manage and show relevant data about the status of the drones such as battery life, altitude, coordinates, id, distance from the next target. All for the sake of safety and mission success.
In the present section we consider another realistic situation, which may occur when a fleet of drones is remotely operated from a ground control room in a collaborative setting, i.e., by a team of multiple operators. Control rooms for mission critical operations and monitoring networks have changed considerably over the years, relying today on smart technology able to better support human–machine allocation tasks and decision support activities as well as collaborative tasks. In particular, we adopted a MultiTouch wall display to design a human–machine interface for a GCS (Ground Control Station) of a fleet of drones, conceived to allow collaboration among multiple operators. The working conditions of the control room operators are characterized by high cognitive workload in terms of user's attention and decision processes both in daily routine and extraordinary/unexpected circumstances. Therefore, the concept of situation awareness acquires a collective meaning, shifting the focus from the SA of the individual operator to that of the entire team of decision makers, resulting in a multidimensional feature that researchers call Team Situation Awareness (TSA) [219, 220, 221]. Maximizing TSA has been our major interface design challenge. The study we present is meant to test the hypothesis that shared displays may improve team performance [222, 223]. To do this, in the experiment the displays are manipulated within a simulation of a mission, and the performance assessed in routine operating conditions and in the event of unexpected events.
The results show that the shared displays do not universally improve team performance, but they help maintain performance in off-nominal situations. Hence, it can be assumed that in those situations the operators will deliver better performance, relying on a higher level of situation awareness.
Structure of the section. Subsection 7.2.2 illustrates some related work on the Ground Control Station interfaces, on the team situation awareness and the shared displays and team performances. Subsection 7.2.3 illustrates the system requirements that may impact the interface design choices, the description of the GCS interface prototype, and the system architecture. Subsection 7.2.4 discusses the study carried out to evaluate the team

performance and situation awareness of the use of the prototype on a shared screen, and the related results. Subsection 7.2.5 concludes the document.

7.2.2 State of the Art

Ground Control Station Interfaces

A survey was carried out on the current state of the GCS UI for remote control of drones. There are tools supporting the user to set up the system, procedures to ensure correct configuration and correct operation of the related software, the procedure for planning the mission (design of the route that the drone should perform), tools allowing users to monitor flight in real-time and tools to analyze and debug flight data mission [224].

Some GCS interfaces already communicate to user's crucial information concerning the operation of drones: incompatibility alerts, usage instructions or safety checks, that prevent the user from basic errors. Some apps already provide learning/support material. Most of them appear in mobile apps targeting basic drone users. However, such apps were conceived for use in missions only, focusing on video recording or mapping [225]. In addition to displaying graphs, some apps have rerun mission features that are very interesting for visually representing the data collected during flights [226]. However, they often result in hard-to-use UIs for mission planning activities as well as for alerting and recovery actions and for drone flight data analysis. Moreover, most of those interfaces do not support fleet mission management and hence do not properly address issues related to situation awareness and safety.

Our case study, for detecting an area and transmitting relevant information to the ground control station, builds upon a more complex system especially conceived to support remote control of multiple drones operating synergistically. In addition, the system was designed for use on a MultiTouch wall where multiple users can interact with it simultaneously. Currently, in literature, it is not possible to find this type of system.

For such a system the design of a usable interface raises even more challenges and becomes paramount for the proper management of a fleet mission.

Team Situation Awareness

Situation Awareness is an individual's awareness in a situation—one person's understanding of "what is going on" in that exact situation. The SA is different from person to person even if people have the same information available and the same working conditions. This is because cognitive factors such as experience, mental models, schemata, and qualifications differ between individuals.

A team can be characterized as a group of people with a shared goal. Working in a team may give several advantages over single operators, such as sharing the workload between operators, contributing with expertise on subtasks and there may be an advantage in safety considering that the operators can check each other's work.

Several papers have been published with good theoretical accounts of situation awareness in teams [219, 220, 221, 227]. In particular, Endsley defines situation awareness as "*the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future*." She defines TSA as "*the degree to which every team member possesses the SA required for his or her responsibility*," distinguishing this from SA which is shared [219]. Salas et al. [220] and Stout et al. [221] place a greater emphasis on the shared understanding of a situation in TSA and also consider the issue of how team process influences situation assessment as an intrinsic component of TSA. The application of the concept of our study combines elements of the two approaches. It is close to the one by Endsley in that shared understandings are not emphasized, however, the role of team communication in developing TSA is explicitly considered, in line with Salas et al. [220]. In short, this experiment draws on both theories and does not seek to discriminate between them. As a result, the findings are readily interpretable within either approach.

Shared Displays and Team Performance

Endsley suggests that SA directly affects performance, particularly poor SA leads to poor performance [219]. This idea is implicitly and explicitly supported throughout the literature and applies to both teams and people. For example, Cooke et al. found a positive correlation between TSA and team performance [228]. Endsley also proposes that the interface design (especially the displays) will influence the SA by determining the amount and accuracy of the information that can be acquired. Fewer articles focus on the impact of displays on SA, while the majority comment more on training [220, 221]. However, there have been some. Of these, there are three empirical studies linking display sharing to team performance through the influence of TSA. Unfortunately, the conclusions of such experiments are ambiguous both for the pattern of results and for the disorders within the experimental design. These studies will be briefly reviewed to place this study in the context of existing work and to demonstrate the need for further experimentation to clarify the relationship between shared displays and team performance.

Sebok used high-fidelity nuclear power plant simulations to compare teams of nuclear power plant operators using conventional or advanced interfaces [227]. Advanced interfaces were designed to improve TSA (along with other factors) and it was found that performance improved when they were used compared to conventional interfaces. However, it is not possible to say which feature of the interfaces caused this improvement. The advanced interfaces provided integrated information and a common overview of the system (shared display), while conventional interfaces did not. Therefore, they confused the advantages of an ecological display [229] with the presentation of all the information to all team members. The latter manipulation was intended to increase TSA, but due to the confusion, it is unclear whether it was actually influential. Measures of improvements in TSA with the advanced interface interacted with the size of the crew, further confusing the interpretation of the results. The study also looked at the effect of various defects that occur in the plant. They showed that the failures caused a temporary decrease in TSA, but did not investigate the reverse, i.e., does TSA improve fault response? Still in [227], the author highlights the relevance for survey drone fleet management applications on shared displays as a means of improving TSA and demonstrates the importance of comparing routine performance with that in case of off-nominal situations. Therefore, both these factors have been manipulated in the experiment we present in this section, and simulations have been designed accordingly.

Shared displays and TSA were explicitly examined in [222] using a task that required two-person teams to defend a departure base from arriving planes. One team member was responsible for classifying the aircraft (e.g., friend or foe) and the other was responsible for appropriate action (e.g., destroying them). The TSA was manipulated by allowing both team members to see each other's displays (good TSA) or by allowing them to see only their own displays (poor TSA). Unfortunately, the results of this manipulation were unclear. Performance only deteriorated in the conditions of the non-shared display when it was completed before the shared condition without any training on the role of the other team members. In all other counterbalanced presentation performance orders, it was equivalent. This model of results suggests that the difference in performance arose from a lack of understanding of the task because so little information was presented about it (there was no practical evidence) rather than the manipulation of the TSA. Therefore, no definitive conclusions can be drawn from this study about the effects of TSA on team performance. Further work should include training to ensure that the results are due to the displays themselves, not the lack of training. The same group of researchers later conducted another experiment with an improved design in which all teams were trained on each other's roles [223]. They used shared and non-shared displays again and abstract displays that showed only the most relevant parts of the display for the other team members. In addition, they manipulated the workload. With low workloads, there was no difference in performance between conditions, but with higher workloads, teams with non-shared and abstract displays performed better than those with shared displays. This is surprising, especially the performance of teams with non-shared displays, as it suggests that TSA is inversely related to performance, which is contrary to all predictions in the existing literature. From this model of results, it is not clear how displays affect performance in this task. Further work is needed to clarify the outcome of this experiment.

Overall, the described experiments demonstrate the relevance of display sharing for TSA and team performance, but do not offer convincing evidence of the impact of shared displays and TSA on team performance.

Our study was meant to further investigate this relationship by comparing shared and non-shared displays and routine and off-nominal situations. An experimental design will be used that offers greater experimental control and will therefore allow for more solid conclusions to be drawn.

7.2.3 The Ground Control Station Prototype

Interface Requirements

Designing the interaction between an operator and a multi-UAS system means setting the communication mechanisms between the human and each drone in the swarm, understanding their specificities and making the human–system relationship effective, efficient and satisfactory for the operator, in addition to being mission-oriented and safe.

It is, therefore, necessary to consider the mental processes that are activated in the operator during a remote mission control and to understand which dimensional, configurative, cognitive, and functional features the UI should guarantee to be consistent with its user's expected behavior.

Considering the case when the user must manage multiple events, within a limited time lapse, it is paramount that the interface keeps the operator's workload below a given threshold while maintaining a high level of Situational Awareness. The design of the interface was already presented in [218], this section intends to set out a vision on how to design a human–machine interface for a GCS (Ground Control Station) of a fleet of drones performing a cooperative mission of persistent surveillance.

The following is a list of requirements which may have a direct impact on the interface design choices:

- The GCS should allow for the inclusion or removal of a UAS to/from the mission, for efficient management of the whole swarm. In [230, 231], the authors showed that a single operator could manage a maximum of five drones, the operator's workload being the limiting factor as regards the maximum number of UAS to be managed by the GCS.
- The GCS should notify any event which may occur during the mission, such as a low battery charge or a problem with video transmission from a video camera. The interface should provide suitable visual clues as a response to those events. For example, a visual/sound alarm could be activated when the battery reaches a critical level.
- The GCS should provide timely information on the status of each UAS in the swarm. The information on the interface must be carefully selected to avoid the operator's overload and draw his/her

attention on relevant events. The standard GCS can be used for a maximum number of five drones, and it is recommended to have a list of drones in a control panel from which the operator can view the parameters associated with specific drones.

- A certain level of autonomy should be implemented in the system so that the operator can rely on (semi-)automatic procedures meant to take control over an aircraft featuring an off-nominal behavior and terminate its participation in the mission as soon as possible. Establishing the right balance between decision-making autonomy and human intervention in re-planning or homing actions may have serious effects on mission success.

The Ground Control Station Interface

In this subsection, we explain how we took into account the previous requirements to design the remote-control interface of the persistent surveillance system. Analyzing the UI in Figure 7.1, we immediately distinguish two areas: on the left we have a sidebar that allows the operator to interact with the application; at the top there are buttons to create or set the mission, information on the targets, and acquisition of the mission log; while at the bottom we see the possibility of connecting to drones with LEDs indicating the connection, to start and end the mission; most of the screen is intended for the map.

Once the connection is made, the LED highlights the new connection status on the map, the targets are added to the mission and the drones at their starting point are indicated.

In addition, at the bottom for each drone a card is created with the data related to the drone:

- drone identification,
- indication of next target,
- coordinates of the drone,
- battery status,
- buttons to send the drone back to the take-off,
- button to view the artificial horizon related to the drone.

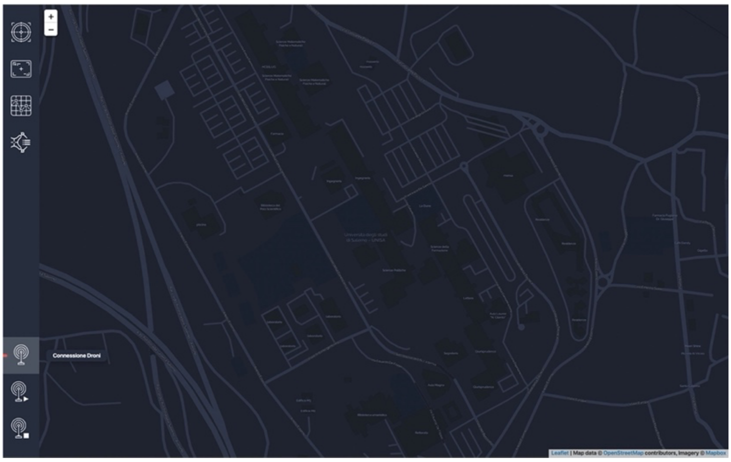


Figure 7.1: The GCS interface before establishing the connection.

Once given the start mission command, one will see the drones move towards the second target the algorithm set. In the event that the drone's battery status is 30%, a notification is displayed in the upper right corner (Figure 7.2). If it is 10%, the drone returns to the take-off automatically and a notification is shown in the upper right corner (Figure 7.3). Finally, near 0% the drone lands automatically and the interface shows the landing position.

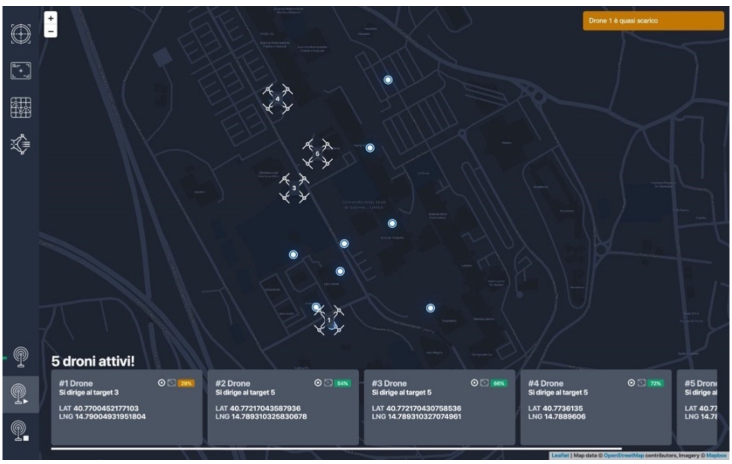


Figure 7.2: The GCS interface during the mission execution in case the drone battery is ≤30%.

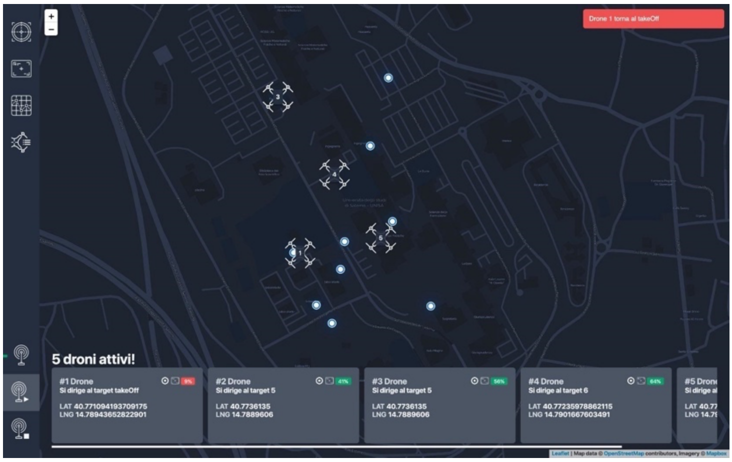


Figure 7.3: The GCS interface during the mission execution in case the drone battery is <= 10%.

In Figure 7.4 we see how the interface provides the operator with the view of the artificial horizon associated with any aircraft selected in the swarm, to facilitate the display of all the information at the same time, if this component is opened in the same position, all other information regarding the drone is displayed.

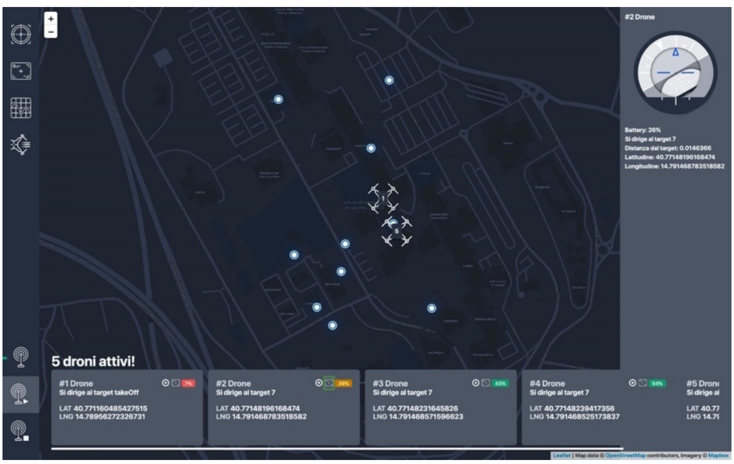


Figure 7.4: The GCS interface during the execution of the mission with the visualization of the artificial horizon.

Ground Control Station Architecture

For the development of the project, we opted for the use of NodeJS technology, this allowed us to work in an innovative environment and take advantage of the flexibility allowed us to work in an innovative environment and take advantage of the flexibility that that this technology exhibits. Indeed, a web app was created with all the pro deriving from this technology exhibits.

Indeed, a web app was created with all the pro deriving from it: we can in fact use it from any operating system through a browser but what interests it: we can in fact use it from any operating system through a browser but what interests us even more is the capability to use it on different devices; more specifically it has been us even more is the capability to use it on different devices; more specifically it has been designed for use on normal desktop or MultiTouch wall systems.

The application is real-time: it receives updates on drones' status data in real-time and updates the interface to allow the operator to react immediately to changes. We decided to adopt a three-tier architecture:

- The first level consists of the data source; in the experiments, a JSON-coded scenario was used which contains all the initial information on the mission and how the system should behave but in real cases, the data source will be the drones that communicate through MavLink protocol;
- The second level consists of the controllers; we need a controller that receives events from the drones and updates the interface and a controller that instead receives events from the operators and communicates them to the level below to send direct commands to the drone;
- The third level takes care of the graphical interface and sends the user inputs in the form of events to the associated controller.

The word "event" has often been used; this is because the NodeJS environment is asynchronous and allows event-driven programming.

For the UI of the application, we opted for the use of HTML5, Sass (compiled in CSS3), and JavaScript.; this stack of technologies allows us to develop a working prototype in a short time (just think the limitations posed by the development of a graphical interface in java).

Another technology used for the project is Webpack: it is a module bundler that has allowed us to implement a workflow such as: speeding up development, ensuring simple and intelligent code organization, and optimizing the size of the app in production.

7.2.4 The Study

The study presents uses of the interface to simulate a scenario in which detection of an area takes place by five drones and a transmission of relevant information to the ground control station where a team of operators is present. The aim of this study is to assess the effects of TSA on team performance.

A team member controls the drones on the map. The second team member checks the drone's battery level.

If the drone is discharged below 30%, it is indicated on the display with a notification on the map "Drone almost discharged". If the drone has a 10% battery status the drone is immediately homing. This is indicated on the display with a "Drone return to takeOff" notification.

This task is interdependent as the higher the drone battery rate, the higher the success rate of the mission.

The drone batteries are limited and require management during the mission. Two independent variables were considered, referring to display *sharing* and *event* occurrence, respectively. In the "Shared" condition, the teams had the screen shared on the MultiTouch wall so that both team members could see the interface which represented the factors that vary in the simulation for their specific task. In the "Non-shared" condition, the teams had screens that were not shared and had to see all the factors of the simulation, without distinction of task.

In the "No event" condition, the mission occurred normally. An unexpected event occurred on the mission in the "Event" condition. In the simulation scenarios that signaled an unexpected event, this is the drone battery without load.

During this activity, the communication of the team members was also evaluated. Salas et al. [220] proposed that group processes are related to TSA. For example, team members can exchange information to develop TSA. To test this idea, team communication was analyzed. Specifically, the frequency with which team members asked questions about the status of the

mission was measured. The questions were speculated to provide a good indication of the extent to which team members tried to improve their SA through communication with the team.
The following hypotheses were derived. In any case, the performance was measured by the success of the mission and of the frequency of communication. Endsley [219] predicts that poor AS will lead to poor performance, and Cooke et al. [228] provides an empirical demonstration of this. A shared view is expected to enhance TSA by providing all team members with direct access to the complete plant status. Therefore, it is expected that:

Hypothesis 1: *Operator's **performance** are higher in the "Shared" condition than in the "Non-shared" one:*

Hypothesis 1.1: *In normal conditions, the **performance** is higher in the shared setting compared to the non-shared setting.*

Since, Sebok observes that SA is reduced when unexpected events occur (see [227]), we are interested in exploring which setting can better deal with it.

Hypothesis 1.2: *In case of unexpected events, the **performance** deteriorates more in the non-shared setting compared to the shared setting.*

Method
For the development of the project, we opted for the use of NodeJS technology, this allowed us to work in an innovative environment and take advantage of the flexibility allowed us to work in an innovative environment

Participants
A total of 40 students from the University of Salerno were recruited to participate in this experiment. There were 25 male and 15 female participants. Their average age was 23.7 years. They were all volunteers and were UAS drivers at amateur level.

Design

A between-subject design experiment was performed. We considered two independent variables that are respectively: the display setting, which can be “Shared” or “Non-Shared”, and the presence of unexpected events, “No Event” and “Event”.

The dependent variables express the operator’s performance. They are the situation awareness (SA), the communication among the team members (CT) and the number of drones back home (DBH).

SA is measured through the situation awareness global assessment technique (SAGAT).

SAGAT is a tool that measures situation awareness in terms of three levels: *perception*, *comprehension*, and *projection*. For each level of situation awareness, SAGAT poses a number of questions whereby only one of the possible answers is correct.

SAGAT scores provide a detailed collection of information that can directly be compared to reality. Hence, SAGAT is a more objective method that is not dependent on a subjective evaluation because there is one right answer that is defined from the outset.

The SAGAT questionnaire is filled out while the simulation is at rest. In this way, participants can only report what they really perceived and remembered from the last seen situation. The questions refer to objects or events from the just presented situation [230].

The communication among the team members was assessed during the activity as it is also related to TSA [220]. The team members can exchange information to develop TSA. To test this idea, the team’s communication was analyzed in [220]. In particular, the frequency with which team members asked questions (NQ) about the status of the mission was measured. The authors observed that the questions provided a good indication of the extent to which team members tried to improve their SA through communication with the team.

During the mission, the lack of situational awareness can cause the drones to be lost. Then, the number of drones returning home (DBH) is an additional parameter that measures the operator’s performance.

Tasks

The task scenario and experiment setting were designed collaborating with the Italian Aerospace Research Center experts; the operators and the system share responsibility following the liability allocation [231]. In this case the system informs operators about the drone status while the operator is in charge to decide when the drone must land or go back home, except for the minimum battery level in which the system controls the drone landing automatically. Figure 7.5 shows two users experimenting with the “Shared” setting.

The goal of the tasks is to correctly monitor the situation during drones’ exploration mission. Two operators sit in front of the system in the “Shared” or “Non-shared” setting. A team member controls the drones on the map. The second team member checks the drone’s battery level.

If the drone is discharged below 30%, a message is displayed on the map. If the drone has a 10% battery status the drone immediately tries to go back home.

The task is run twice: once in a normal situation and once with the unexpected event of the battery. The starting task is selected randomly.

During the task an unexpected event is triggered. A drone’s battery suddenly gets down to 30% causing troubles to the team.

Figure 7.5: Two users operating in the “Shared” setting during the experiment.

Procedure

Before completing the experimental tests, the operation of the simulated mission was explained to each participant. Participants completed a practical test during which they could ask questions about the experiment and the system to make sure they fully understood it. They were then divided into two groups. The teams were then randomly assigned to "Shared" or "Non-shared" conditions. Within the shared view group, they were randomly assigned to the specific control role. The participants were not advised that this could happen. In a simulation, a drone was discharged, causing a low battery event (>30%). In the second error test, the battery indicator signaled a low battery event less than 30% but larger than 10%.
The order of these simulations was counterbalanced between the teams. Each simulation lasted 4 min. During this period all their communications were recorded for subsequent analysis. After 2 min a facilitator asked the participants the questions coming from the SAGAT questionnaire. The questionnaire consists of three items. Each of the three items belongs to one of the three levels of SA: perception, understanding and projection. Participants were free to respond openly. The right answers were defined from the start and were rated as correct or incorrect based on industry expert suggestions. Since each participant received two questionnaires made of three questions, respectively, in the two settings of the experiment, it was possible to obtain a maximum score of 6 correct answers in total. A summary of the questions is provided in the Table 7.1.
Questions asked during the test to assess operator situational awareness.

Table 7.1: Questions asked during the test to assess operator situational awareness.

Question
Q1. How many UAVs are you controlling?
Q2. What is the altitude and battery level of each of the UAVs?
Q3. What is the next UAV target on the screen?

Results

Since in our scenario the performance consists of SA, CT and DBH, we formulate the following sub-hypotheses:

Sub-Hypothesis 1.1.1: *In normal conditions, SA is higher in the shared setting compared to the non-shared setting.*
Sub-Hypothesis 1.2.1: *In case of unexpected events, SA deteriorates more in the non-shared setting compared to the shared setting.*

Sub-Hypothesis 1.1.2: *In normal conditions, DBH is larger in the shared setting compared to the non-shared setting.*

Sub-Hypothesis 1.2.2: *In case of unexpected events, DBH deteriorates more in the non-shared setting compared to the shared setting.*

Sub-Hypothesis 1.1.3: *In normal conditions, **NQ** is smaller in the shared setting compared to the non-shared setting.*

Sub-Hypothesis 1.2.3: *In case of unexpected events, **NQ** increases more in the non-shared setting compared to the shared setting.*

As for SA, users answered three questions while performing the two tasks. The number of correct answers are reported in Table 7.2 next to the relative *p*-value calculated by performing a non-parametric Chi-squared 2-by-2 test. This is to statistically analyze the frequency of correct and incorrect answers for each group. The test shows that there are no significant differences between the correct response frequencies among the two groups, except for Q3 which is related to the operators' prediction factor. Therefore SH1.1.1 is rejected and we can state that there is no significant difference in situational awareness between the two settings when there are no unexpected events. On the other hand, SH1.1.2 is supported by data, so we can affirm that when an unexpected event occurs, the predictive ability of the operators deteriorates, particularly in the "Non-shared" setting.

Table 7.2: Number of correct answers per question as measured with SAGAT for each setting.

	Correct Answers in Shared Setting Group	**Correct Answers in Non-Shared Setting Group**	***p*-Value**
	No Event task		
Q1	20	20	/

Q2	18	16	0.376
Q3	17	13	0.144
	Event task		
Q1	19	15	0.077
Q2	18	14	0.114
Q3	18	10	0.003

The number of drones that potentially can return home at the end of the mission can vary between 0 and 5. Table 7.3 shows for each task how many drones each pair of operators managed to bring home.
The Mann–Whitney parametric test analysis was performed. It is adequate to compare two datasets collected from two different groups. The analysis produced a *p*-value of 0.271 in the task without an unexpected event and 0.045 in the other case. This indicates that there are no particular differences in the drones returned to base in the first task, rejecting SH1.1.2, while there is statistical evidence in the second case, confirming SH1.2.2.

Table 7.3: The number of drones back home (DBH) at the end of each task.

	Number of Drones Back Home in Shared Setting Group	**Number of Drones Back Home in Non-Shared Setting Group**
	No Event task	
Team 1	5	4
Team 2	5	5
Team 3	5	5
Team 4	5	4
Team 5	5	5
Team 6	5	4
Team 7	4	4
Team 8	5	5
Team 9	5	5
Team 10	5	5
	Event task	
Team 1	4	3
Team 2	3	3
Team 3	3	2
Team 4	4	3
Team 5	3	3
Team 6	3	2
Team 7	4	1

Team 8	3	4
Team 9	4	3
Team 10	3	2

Finally, Table 7.4 shows the number of questions about the mission status asked between the pairs of operators during each task. For each task, the means of the two groups were compared by applying a two tailed independent-samples t-test. For the "No event" task, the *p*-value is 0.268, thus rejecting hypothesis SH1.1.3. For the second task the *p*-value is 0.002 showing that when an unexpected event occurs, the number of questions increases in the "Non-Shared" setting.

Table 7.4: The number of questions among the team members (NQ) about the status of the mission.

	Number of Questions in Shared Setting Group	**Number of Questions in Non-Shared Setting Group**
No Event task		
Team 1	3	2
Team 2	2	3
Team 3	2	2
Team 4	1	1
Team 5	2	4
Team 6	1	2
Team 7	2	2
Team 8	2	1
Team 9	1	2
Team 10	1	2
Event task		
Team 1	4	5
Team 2	3	6
Team 3	5	4
Team 4	4	6
Team 5	5	7
Team 6	4	6
Team 7	4	5
Team 8	3	6
Team 9	5	4
Team 10	4	7

Discussion and Implication

The purpose of this study is to carry out predictions derived from the situation awareness of the team operating the ground control station that sharing displays between team members would improve team performance. The results show that the shared displays do not universally improve team performance, but they help to maintain performance in off-nominal situations. These results are discussed below.

There was no overall visualization effect on team performance. Shared displays were predicted to cause better TSA than non-shared displays and thus lead to better team performance by controlling what information team members could easily access. The analysis of team communication clarifies the picture. Teams with unshared displays asked many more questions about the simulation. This implies that they integrated the limited information presented to them with the information aroused by the other team member. It is entirely possible that this allowed them to develop TSAs comparable to teams in shared conditions. This explanation of the results is interesting in that it supports the conjecture of Salas et al. [220] that states that the group process is an intrinsic part of situation assessment and development of the TSA.

Unexpected events in the simulation reduced performance, showing that they had a measurable impact on the activity. This leads to the second interesting discovery, that performance deteriorated significantly when unforeseen events occurred in the "Non-shared" setting, more than in the "Shared" setting. The frequency of mission status questions demonstrated a similar interaction, implying that the team process offset the negative impact on TSA caused by the unshared view and an unexpected event.

The implication of these results is that shared displays are particularly useful for performance in off-nominal situations. This experiment does not reveal how this effect occurs. It is possible that with shared displays there are more people who can see the status of the mission and therefore they are more likely to quickly detect an unexpected event. It may be that with a good TSA they are more able to get around the problem once it has been detected.

Although the "Shared" setting demonstrated generally more advantages in situations involving sudden events, and it does not generate difficulties to operators in normal situations, it also reveals the need to have physical spaces enough large to allow collaboration among operators. Moreover, the experiment studied only the collaboration among two operators while in the wild there is also the possibility to have several operators collaborating at the same time. Thus, more investigation is needed in this sense.

7.2.5 Conclusion

In the section, we analyzed relevant human factor implications for remote ground control of multiple unmanned aerial systems.

During the stages of design, the prototype was subject to supervision by the ground operators involved since the initial contextual inquiry. They actively collaborated with us in the definition and specification of the interface requirements. The formative evaluation results obtained so far are considered encouraging. The sensorial isolation of the operator and the problems related to spatial disorientation of existing multi-UAS systems are likely to be solved by using a single virtual screen that provides the operator with all the information regarding the mission and all the UAS involved while allowing the horizon view of any single aircraft.

In conclusion, this study provided qualified support for the prediction that shared displays would lead to better team performance than non-shared displays.

In particular, the advantages of display sharing arise in off-nominal situations, for example when an unexpected event occurs but not during routine operation.

The study also suggests the important role of team process (communication) in maintaining TSA. However, the results indicate that the team process cannot overcome the poor TSA derived from non-shared displays in all situations. In this task, the team process only seemed to provide sufficient compensation in routine situations.

Chapter 8

UX challenges for Industry 4.0 process managers

Also, in the design of systems for Industry 4.0 it is essential to study the UX challenges of the design of Human-Machine Interfaces, as they are widely used in order to increase the performance of production processes while reducing the number of emergencies and accidents. In production, the most typical system used to monitor production is the Andon. It is a graphic system used in the plants to alert the operators who deal with management, maintenance and production performance of the presence of a problem. Of course, the usability of such interfaces is critical to allow an operator to identify and react more effectively to potentially critical situations. Improving the usability of such interfaces, and therefore also the UX, is a great challenge due to the growing complexity of data that must be processed and understood quickly by operators. In this chapter, a set of guidelines are presented to help professional developers design usable interfaces for monitoring industrial production in manufacturing. These guidelines are based on usability principles and formalized by reviewing existing industrial interfaces. Using a realistic case study prepared with manufacturing experts, an Andon interface was proposed which was developed to test the effectiveness of these guidelines on a state-of-the-art touch-wall device.

8.1 Design usable interfaces for Industry 4.0

8.1.1 Introduction

Human-Machine Interaction or Interface (HMI) is defined as the way we humans interact with machines, being any machine any mechanical or electrical device that transmits or modifies energy to perform or assist

people in their tasks [235]. In literature, the definition of HMI often overlaps with that of the best-known Human-Computer Interaction (HCI). Nevertheless, HMI is mainly associated with applications in industrial and robotic context [235, 242, 243, 236]. HMIs are largely used in the industry 4.0 [253]. It is formed by numerous associated technologies and paradigms, including for example RFID, corporate resource planning (ERP), and Internet of Things (IoT). Here, HMIs are largely exploited to monitor the production processes. These interfaces are typical for any sort of monitoring process which includes, for example, the sectors of water flow management, of waste disposal or the manufacturing sector. They help industrial plants to reduce the number of emergencies and accidents. This is possible by providing operators with a global view of the situation and thus allowing their rapid reaction [245].
In manufacturing, the most typical system used to monitor the production is the Andon. It is a typical HMI exploited in plants to notify operators who deal with management, maintenance and production performance of the presence of a problem [251]. It can be used to display an alarm, a warning or any other situation through data collected either by plant workers or by sensors present in the machines. Typically, an Andon system can provide not only information about the kind of problem but also a description to help operators to interpret and localize the problem. Traditional Andon systems are based on signboards usually including a little text with colors. Modern systems are based on large displays and can provide more information in terms of text, images, and sounds. The information shown in such a system is generally processed using a Key Performance Indicator (KPI). The most common indicator is the *Overall Equipment Effectiveness* (OEE) [250] that can express an instantaneous view of the production performance. It is important to highlight that these KPIs can express the contemporaneity of the events and help to solve problems that have just occurred or to prevent new ones.
Of course, the usability of such interfaces is paramount. In [240, 261] authors point out the usability issues due to big amount of data to display on a screen while Hollifield in [244] states that the correct design of a usability-focused HMI is essential to allow an operator to identify and intervene effectively and quickly on anomalies within a process before they turn into real emergencies. In [239], the author explains that usable

interfaces would help plants save up to $ 800,000 a year allowing operators to identify and react more effectively to various critical situations. However, supporting the usability of such HMIs is a big challenge due to the increasing complexity of the data that must be processed and understood quickly by operators. Moreover, as stated in [263] the design of HMIs in the industrial context does not take into account the many technological advances in the field of HCI, such as touch-screen devices or the tangible UIs that can further support usability [232].

In this section, we present a collection of guidelines for designing usable Andon interfaces for manufacturing production. Using a realistic case study prepared in collaboration with manufacturing experts, we also propose an Andon interface that we developed to test the efficacy of these guidelines on a last generation touch-wall device.

Structure of the section. In Subsection 8.1.2, we describe the manufacturing production context showing the different kinds of plant layouts and the Key Performance Indicators used in the Andon systems. In Subsection 8.1.3, we formalize a set of design guidelines for Andon interfaces based on usability principles. Subsection 8.1.4 is dedicated to the case study for which we developed and experimented an Andon interface based on the proposed guidelines. In the last subsection, we present 4 *lessons learnt* from our research and give some conclusions.

8.1.2 The context of manufacturing production

In the present subsection we briefly describe the different kinds of plant layouts used in manufacturing and the Key Performance Indicators commonly used in Andon systems.

Classification of manufacturing plant layouts

In the manufacturing industry, there are many ways of organizing production facilities depending on nature (intrinsic characteristics) of the product, the production volume and the repetitiveness of the processing cycle. The aim is to obtain process efficiency ensuring, at the same time, a satisfactory level of safety for the personnel carrying out the work [254]. Over the years, certain types of production facilities have established themselves as the most appropriate ways to organize manufacturing: *fixed*, *position*, *process*, *cellular* and *product* layout [234, 240]. In [260] a detailed definition of the fundamental plant layout can be found. The authors highlight how the objectives of production efficiency and safety can be declined in terms of efficient use of resources (men, machinery, and materials), the flexibility of manufacturing, ease of maintenance, reduction of work-in-progress and storage costs. We show in Figure 8.1 the types of production structures most frequently used. During the manufacturing process that takes place with the *fixed-position* layout, the product remains stationary in a predetermined position. It is the workers, materials, tools and work tools that move in close proximity to the good to be produced. This configuration is chosen when the product is large and complex to manufacture (e.g. a ship or an airplane).

In the *process* layout, departments are organized in such a way that they contain similar machines. For example, in a metalworking industry, the milling department will contain several milling machines possibly with different machining capacities, the laser cutting department will contain machines capable of performing cuts, perforations, etc. on metallic material and so on. The workpieces will follow different flows during the processing. In fig. 1(b), it is highlighted how two different parts can follow two separate processing paths; the first part will be machined by the M1, M8, and M7 machines while the second part will be machined by the M2, M4, M5, M6, and M8 machines.

When the parts to be produced are different but are quite similar, then it is possible to group the parts into homogeneous families and configure the machines in such a way that groups of parts or similar products can be made on the same machines without significant waste of time for changeovers. In other words, similar or sufficiently homogeneous pieces can follow the same processing path. In such cases, the *cellular* layout is used. In comparison with the *process* layout, the *cellular* layout allows us to increase production efficiency and simplify its management aspects. On the other hand, the structure loses the typical flexibility of the *process* layout. In the fourth fundamental pattern, the *product* layout, the workstations are organized in sequence so that each station can perform a part of the work on the workpiece. The workstations consist of machinery, people and tools and are designed to maximize the efficiency of the production process. Workpieces move from one workstation to the next, usually by means of a powered conveyor. This type of layout is chosen when the production volumes are high.

Although the layouts described are those most frequently implemented, these must be understood as basic schemes that can also be combined within a single industry or varied in order to search the most performing scheme for the considered production type. In Subsection 8.1.5 we will present a case study that refers to a variant of the *process* layout called identical machines in parallel where the jobs in progress require a single operation and can be processed on each of the machines in a department or on anyone that belongs to a given subset [261].

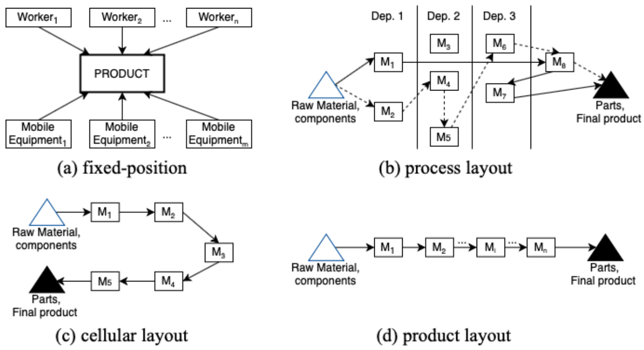


Figure 8.1: The fundamental types of plant layout.

Each of the layouts discussed above has its own needs in terms of presenting data concerning the state of the production processes. In Subsection 8.1.4 we will describe the design principles that can be taken as a reference both for the scenarios outlined above and for their variants.

The OEE as an Aggregate Key Performance Indicator

The *Overall Equipment Effectiveness* is a key performance indicator (KPI) widely used in manufacturing industries [246]. The OEE can be defined as a roadmap that helps plant workers and managers to view and eliminate equipment losses and other problems. The OEE was first proposed by Seiichi Nakajima in the 1980s in the context of total productivity maintenance (TPM) as a metric of equipment efficiency and was considered to be the primary measurement for evaluating production processes and identifying potential problems affecting the operation of the equipment

[246]. The OEE is made up of other three KPIs that are: quality rate, performance rate and availability rate. For a plant, 100% of quality means that it only produced valid parts, 100% performance means it produced as fast as possible and 100% availability means that no machine has had any downtime. The three KPIs can be expressed as followed

Availability= Run time / Planned production time

Performance= (Ideal Cycle Time X Total Count) / Run time

Quality= Good Count / Total Count

According to Nakajima, the OEE equation is:

OEE = (Availability) * (Performance) * (Quality)

Therefore, when the OEE is at 100% it means that the machine is producing only good parts, as fast as possible and without downtime.

One of the factors behind the popularity of the OEE metric is its simple calculation method and its widespread impact on production efficiency [249]. However, when it is necessary to score the entire factory's production or a single entire line, different modified versions of the OEE should be applied. The OTE [255] or OFE [259] can be used to score the factory's production and the OLE [256] can measure a production line. However, due to the variety of the industry layouts, putting the equations into practice could be quite complicated then its calculation formula often is customized to meet particular production requirements [257]. One of the best practices is to identify and measure the OEE only at the machine that governs the throughput. When all the machines run at identical speed, it is possible to monitor the OEE only of the machine that does the primary work. On the other side, when it is necessary to score a set of machines working in a single line, the OEE output can be determined by the machine with the lowest processing speed in the line (HE).

Typically, the KPIs are displayed in different points of the plants, both in the plant floor, offices and other places. This is a simple and effective way to a) help operators and maintenance staff understand and quickly respond to production problems, b) monitor the activity of the factory in real-time from any point of it, c) automate the information collection process, allowing more time for analysis and improvement and more time to use the equipment effectively.

8.1.3 HMI design guidelines for Andon interfaces

The correct design of a usability-focused HMI is essential to allow an operator to identify and intervene effectively and quickly on anomalies within a process before they turn into real emergencies [244]. Therefore, HMI design guidelines focused on usability are a necessity that acquires big relevance in the design of interfaces for the manufacturing context. In fact, the right representation of data, processes and KPIs not only helps operators monitor the state of the production but can also be crucial to prevent possible emergencies. Starting from the analysis of existing solutions used in industry to design Andon interfaces, we identified and formalized six design guidelines corresponding to common visual management activities: *State of the machines* (Table 8.2), *OEE representation* (Table 8.3), *View all OEE together* (Table 8.4), *Temporal representation of the machine production* (Table 8.5), *Temporal representation of the machine states* (Table 8.6), *Temporal representation of KPIs over a period* (Table 8.7).

Each guideline is structured according to the following scheme: Title, Problem, Context (To be used when), Solution, and Example. Table 8.1 gives a brief explanation of each field in the structure. The guidelines were built on the basis of usability principles that are: Visibility of system status, Error prevention, Recognition rather than recall, Aesthetic and minimalist design, Accessibility, Hierarchical navigation, Consistency and Standards [237, 258].

Table 8.1: The structure used to describe the guidelines.

Explanatory title	
Problem	It is the description of the problems related to the use of the system relevant to the usability and therefore to the experience of the stakeholders
To be used when	This section describes a situation where the guideline is appropriate to address the problem
Guideline	The guideline that describes the design solution addressing the problem
Example	A picture showing an example of application of the guideline

Table 8.2: Guideline to display the states of the machines in a plant.

State of the machines	
Problem	Represent the state of each machine so that each state is immediately intelligible without overloading the interface
To be used when	When it is necessary to monitor several machines simultaneously and there is a risk to overload the interface

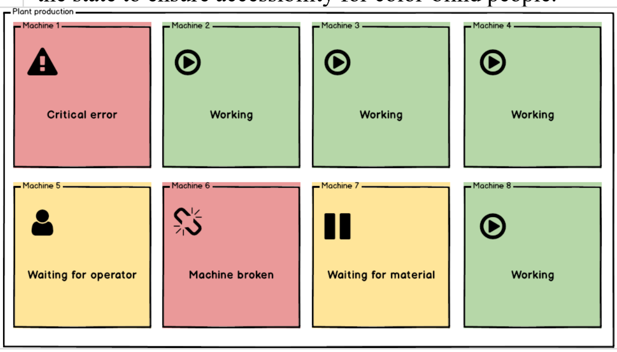
Guideline	Use a different and distinguishable color for each machine state, such as working, waiting, broken. For each state, it is recommended to present an icon to indicate a state specification. The colors can be applied as a frame, a background or as a color of the font provided that readability is kept. The colors must always be accompanied by a brief descriptive text of the state to ensure accessibility for color-blind people.
	

Table 8.3: The guideline to allow the analysis of the KPIs of a specific machine.

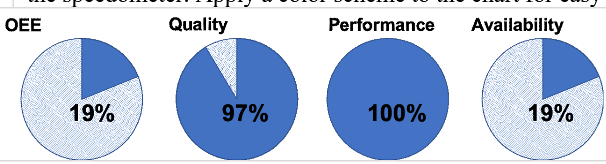
OEE representation	
Problem	Represent the value of the OEE to operators and production managers in a way that is immediately understandable and significant
To be used when	During the whole production process, when it is necessary to offer information at a glance about all the KPIs in an intuitive but yet effective way
Guideline	Use a specific interface to display the OEE. Represent the OEE through its percentage value and accompany it with the values of the other indicators that make it up. Make it possible to highlight the indicator considered most important within the production. In addition to the percentage value, represent the indicators using simple diagrams such as circle diagrams or based on metaphors such as the speedometer. Apply a color scheme to the chart for easy reading.
	

Table 8.4: The guideline to display the OEE for a set of machines.

View all OEE together	
Problem	View the OEE of all machines on a single display
To be used when	When the display is used to analyze the situation of the entire production and there is a risk to overload the interface
Guideline	Clearly separate each machine using separator graphics such as cards. Each card will only contain elements related to the OEE for a specific

	machine. Limit the elements to one KPI associated to a representative graph, so as to avoid information overloading in each card.

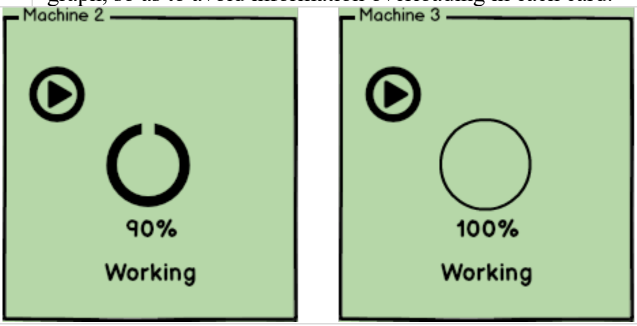


Table 8.5: The guideline to analyze the production of a specific machine.

Temporal representation of the machine production	
Problem	Represent the trend of the production during the working day
To be used when	During analysis tasks, when it is necessary to let operators pay particular attention to the evolution of a machine production
Guideline	In a dedicated display, show the production trend through a diagram in a way to make it possible to visually analyze the number of pieces produced, compared to those expected. Use text as an alternative to colors to guarantee the accessibility to color-blind users. Consider the possibility to apply a time filter.

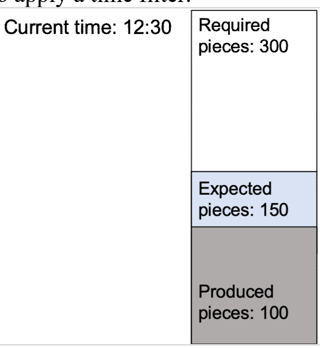


Table 8.6: The guideline to analyze the evolution of the states of a specific machine.

Temporal representation of the machine states	
Problem	Represent all the states of a machine during the working day
To be used when	During analysis tasks, when it is necessary to let operators pay attention to the evolution of a machine state
Guideline	In a dedicated display, show all the machine states in diagram so that they can be visually analyzed over a given time lapse. Overlay the states representation with the planned stops. Consider the possibility to apply a time filter.

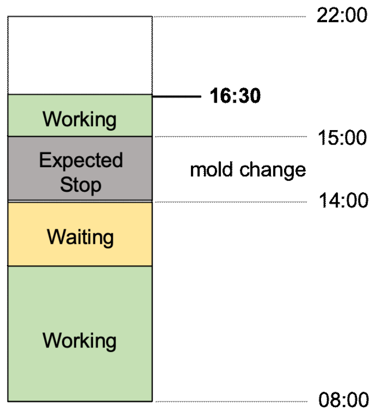
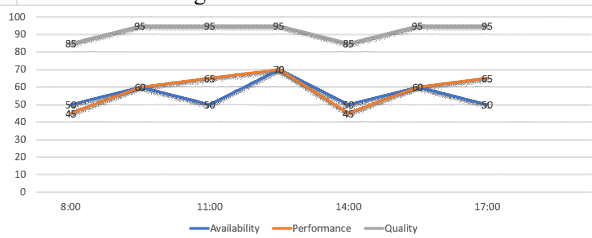


Table 8.7: The guideline to analyze the evolution of the KPIs.

Temporal representation of KPIs over a period	
Problem	Represent the trend of the indicators during a certain production period
To be used when	During analysis tasks, when it is necessary to allow operators to compare the trends of the KPIs over a certain period
Guideline	This can be implemented in a separate interface or in the same interface where the KPIs of a machine are displayed. Represent the KPIs through a percentage / time graph. Consider enabling a selector to view all the indicators simultaneously or one at a time. Consider enabling a time selector.



8.1.4 Case Study

Here we introduce the prototype we developed to experiment with the proposed guidelines. It was designed to deal with the following scenario formalized with experts of manufacturing plants. **Manufacturing plant scenario**: an industrial plant that produces automotive plastic parts needs to monitor the progress of its production. The plant is based on a variant of the *process* layout in which machines producing the same family of

components are arranged in the same sector. The plant has an industrial pavilion containing 26 machines independent of each other, running at the same speed and producing plastic pieces through a press with an interchangeable mold. Some machines will print the same piece throughout the day, other machines will print a certain number of pieces and then the mold will be replaced to change the type of product. Operators are in charge of monitoring the progress of production keeping in mind the independence of each individual machine.
In particular, the interface was designed to run on a large interactive multitouch wall to be placed on the plant floor. The multitouch feature allows operators, on the one hand, to interact with the interface in a more intuitive and quicker way, avoiding using other computer peripherals, and on the other, to work simultaneously on the same device. The device is a touchwall consisting of 4 46 "LED modules that allow 32 simultaneous touches, and with a 4320x1920 pixel resolution. The software was developed in HTML 5 to allow the portability on different platforms. Figure 8.2 shows the main page of the Andon interface. The interface displays all the machines each one separated in a single card.

Figure 8.2: The main page of the Andon interface.

The color background of the card visually represents the machine general state and an icon adds more details about the precise state. A descriptive text helps also color-blind users to interpret the states properly. There is a total of 7 states described in Table 8.8.

Table 8.8: The 7 states of a machine.

Main color	State
Green	Running but not ready for production
	Running and in production
	Running but awaiting mold change
Yellow	Unknown warning
	Machine waiting for material
Red	Broken
Dark grey	Off

Apart from the state, through a circular graph and its percentage. This interface was built using the guidelines *State of the machines* and *View all OEE together*.

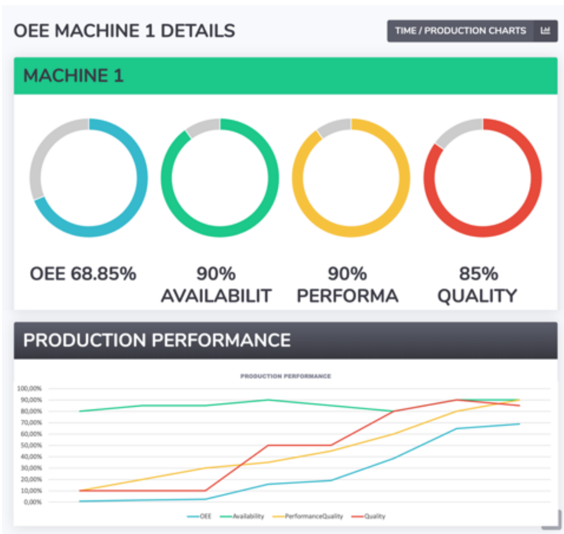


Figure 8.3: The page of the KPI of a machine.

The navigation through the linked pages is designed to always pop up a new window. This is to take advantage of the large display and allow each operator to view the new information without losing the focus on the present. Therefore, following the hierarchical navigation principle, tapping the single card, the user opens a new window showing more details of the machines. The page is shown in Figure 8.3. The graphs show an instant view of the four KPIs (*State of the machines* guideline), while at the bottom of the page the linear graph shows the percentage evolution of the KPIs during the day. Still following the hierarchical navigation principle, from this page it is possible to access another window containing a visual analysis of the machine production.

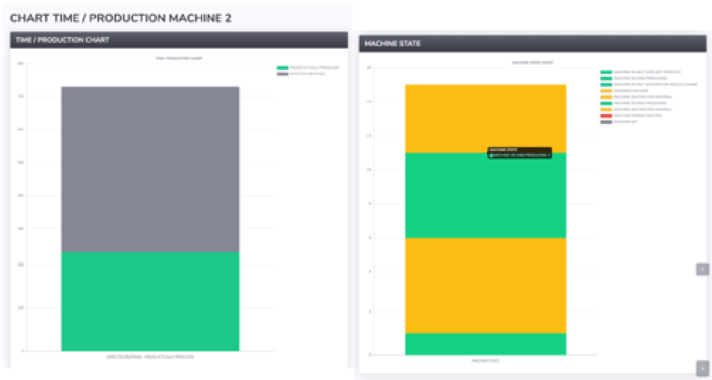


Figure 8.4: The visual analysis of the production and states of a machine.

In particular, in Figure 8.4, the two graphs developed on the *Temporal representation of the machine states* and *Temporal representation of the machine production* support operators to analyze the machine performance and possible problems.
The OEE can be calculated for the entire factory's production. In this case, because the plant is based on a variant of the *process* layout, as default the OEE is measured at the first machine, but it is possible to select the machine that one considers more appropriate.

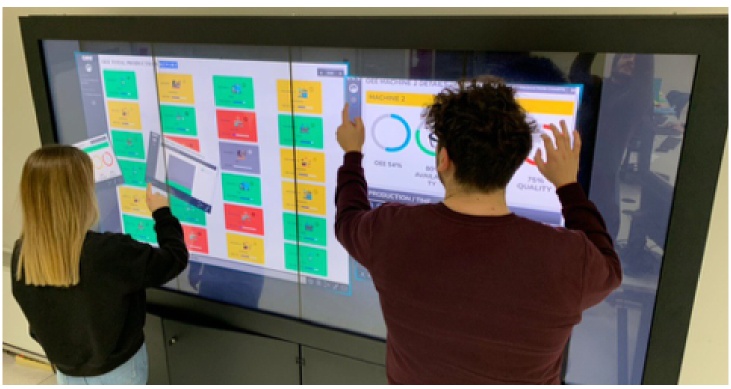


Figure 8.5: Users interacting with the touchwall.

Figure 8.5 shows two users interacting at the same time with the application running on the multitouch wall. In order to experiment with the prototype, we also developed a factory simulator (Figure 8.6). The simulator generates data typically captured by the factory's equipment during the production. More in particular, it is possible to define the number of the machines in the plant and for each printing press, the system generates the effective number of beatings per hour, the number of good and bad pieces per hour, the different states that the machine can take according to the pieces produced. Once the data are generated, it is possible to review and change them to build some specific situations. Starting from the generated data, the system calculates and visualizes the KPIs along a 24 hours period. To make easier to navigate through the 24 hours, a graphical controller was added to the home page. To see the plant situation in a certain time it is enough to move the slider at the specific hour.

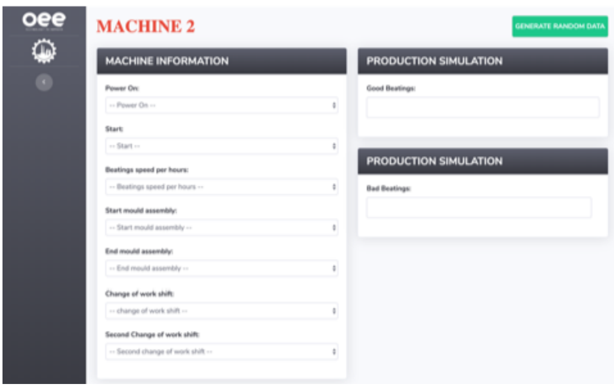


Figure 8.6: The first screen of the simulator**.**

Usability testing

To evaluate the goodness of the proposed guidelines we performed a between-subjects experiment involving 18 participants divided into two groups. Participants were asked to play the role of plant operators in charge to monitor the production using an Andon system. To each group a different Andon system was assigned, the experimental group interacting with our multitouch wall interface, the control group interacting with a popular Andon system, as suggested by our experts. We developed a clone interface of such system instead of using the original one to be able to connect it to our simulator. The interface is based on a led panel, which can be placed on the production floor to monitor the general performance of the entire production and a display showing more information, which can be used both in a control room and in the production floor. The panel shows through text the measured KPIs and some of the data scheduled or collected by the machine such as the number of pieces produced and of those expected. The background color represents the machine state. The second display is used to monitor the performance of all the machines in one device. In the left column of the interfaces, the machines are grouped by production lines and the operator can access the information of only one machine at a time. The KPIs are shown through a bar diagram and through a cartesian graph showing their evolution along the time. Other information is reported in text areas, such as the numbers of expected and produced pieces, the remaining number of pieces, the estimated remaining time, etc... The states of a machine are not directly expressed but its uptime and downtime are reported. Figure 8.7 shows a user experimenting the cloned interface. In this experiment, we decided to adopt a between-subjects design with the aim to limit the effects of the bias. 18 engineering students with basic knowledge in the plant production were involved in the experiment. This is so because we wanted to avoid that skilled subjects could distort the experiment results. Participants were divided into two groups of 9: Gr1 and Gr2. The experimental group Gr1 used the system based on our guidelines (henceforth AndonA), whereas the control group Gr2 used the traditional Andon system (henceforth AndonB).

Figure 8.7: The cloned interface of a common Andon system.

Additionally, to limit biases due to users' variability, participants were randomly assigned to each group. The participants were introduced to the context and to the manufacturing plant scenario presented in Section 5. Three tasks of increasing complexity were proposed for testing. For each task we asked to identify potential problems and their causes. The tasks were presented as a temporal evolution of the same scenario, as explained in the following.

T1: The OEE of machine 5 is steadily decreasing over the last two hours but no obvious change can be detected in the state of the machine. Look for the possible causes. **T2**: The panel shows a high OEE score for machine 1 but the current production is really low. Look for the possible causes. **T3**: Consider machine 3 and machine 10, which are the same kind of machine, fabricate the same product but have different OEE, look for the factors that can explain such a situation.

Quantitative (correct and not correct, and time) and qualitative measurements were collected for each experimental session. The participants were required to justify their answers, which were evaluated by the observers as correct or not. Moreover, we asked them to use the *Think-aloud* protocol, to bring to light possible hidden aspects of the usage of the system. Table 8.9 resumes the three tasks, the corresponding number of correct answers (NoCA), and the consumed average time (CAT) for each group with the relative standard deviation. Gr1 answered correctly to almost all the situations, while some members of Gr2 showed some difficulties. In particular, 2 testers belonging to Gr2 failed in all the situations, while another failed only in the second and third. As for the time consumed to carry out the tasks, for each task Gr1 was quicker than Gr2 on average. The standard deviations for Gr2's means are significatively higher than those of Gr1 because the users who could not accomplish the tasks correctly consumed more time to explore the interface.

Table 8.9: Quantitative data collected through each task.

Task id	Gr1's NoCA	Gr1's CAT	SD	Gr2's NoCA	Gr2's CAT	SD
1	9	294 s	52	7	333 s	71
2	8	253 s	40	6	380 s	92
3	9	274 s	58	6	458 s	125

After the experiment, the participants were asked to perform the NASA Task Load Index (NASA-TLX) test [247]. This test is largely used to measure the physical and digital experiences in terms of the perceived workload needed to solve a particular task. Precisely, the TLX measures: *Mental Demand* (MD), *Physical Demand* (PD), *Temporal Demand* (TD), *Performance* (P*), Effort* (E), *Frustration level* (FL). The TLX provides a general workload score on a 100- point scale based on the average of the six subscales. Each subscale is evaluated by a user from 1 to 100 points: the higher is the score, the higher the perceived workload. Table 8.10 resumes the test results through the means of each item for each group and the relative *coefficient of variation* (CV). The CV for each item is significantly small, indicating a relatively small variance among the responses and therefore a similitude of the workload perception among the users of the same group. The test revealed that the general cognitive workload was lower for users belonging to Gr1, properly 37 score for Gr1 against 49.7 for Gr2. Between the items' means, the biggest difference is noticed in the *Frustration level* that is decidedly higher for Gr2, properly 21.7 for Gr1 and 58.3 for Gr2. In the test, the higher is the *Performance* the higher is the degree of difficulty perceived when accomplishing a task. Therefore, a value of 31.7 for the experimental group G1 is better than the value 59.4 associated to the control group G2 responses. Finally, it can be noted that the physical effort was slightly greater for Gr1 who used the touchwall system, but this had a little or no influence on the general workload.

Table 8.10: The means of TLX items**.**

Item	**Gr1**	**CV**	**Gr2**	**CV**
MD	49.4	0.1	55	0.1
PD	43.3	0.2	39.4	0.3
TD	41.7	0.3	47.2	0.2
P	31.7	0.2	59.4	0.1
E	34.4	0.2	38.9	0.3
FL	21.7	0.3	58.3	0.2
Total workload	**37.0**		**49.7**	

Analyzing the testers' comments collected through the *Think-aloud* protocol, we can state that users in Gr1 appreciated the possibility to hierarchically access the information and to visualize it in separate windows that can be rearranged as needed. For example, one said: "*I need to access the information of the states of a specific machine, but I am repositioning the window representing the machine's OEE to avoid losing the focus on it*". "*To understand the critical issues in this scenario I can open and arrange the windows containing the diagrams of the suspect machines to make a visual comparison*". Moreover, although the *Physical Demand* is higher for Gr1, this can be explained by the initial difficulties observed during the experiment in interacting with a new device. Indeed, a user said: "*At the beginning, I thought it was difficult and tiring to move and rearrange the windows using gestures but, after a while, I got soon familiar with it and now it looks something more natural and conformable*". Finally, the users showed appreciation also for the diagrams: *"It is easy to understand the trend of the states and performance of a machine thanks to the graphical representation*". On the other side, Gr2 found the interface of the AndonB to be overloaded with information. Generally, the comments were about the difficulty to find the information needed to analyze the situation, and this can explain the lower number of correct answers shown in Table 8.9. Some users commented "*There is too much information to navigate, I got lost in it*", "*I am not sure if the information that I am reading is related to the right machine or to another one*", "*Comparing two machines is too demanding, I am often losing the focus of the box I am looking at*".

The experiment showed that the interface based on the guidelines proposed in this work can improve both the effectiveness and efficiency of a common Andon system. However, the main outcomes are related to the effects on the users' cognitive workload. Indeed, the experimental group showed a significantly smaller workload than the control group. This is attributable not only to the goodness of the proposed design guidelines but also to the multitouch interaction that provides an intuitive and familiar way to interact in a *multi-window* environment.

8.1.5 Conclusions

In this work, we laid down the foundations to develop more usable HMIs for monitoring the production process of an industrial plant. For this, we formalized a set of design guidelines starting from usability principles and the analysis of existing solutions corresponding to common visual management activities. To experiment with them, we designed and developed a prototype for a multitouch interactive wall to deal with a realistic case study prepared in collaboration with manufacturing experts. As a result of the experimental study, some considerations can be made. 1) the information displayed in a hierarchical way permits more efficient and accurate analyses allowing access only to the information needed in a given situation. 2) The *multi-window* interaction experience is recommended to allow operators to arrange the interface along with their own mental model and create arrangements that better support their visual comparison skills. 3) Furthermore, the multitouch interaction allows to quickly and easily arrange the windows without increasing significantly the users' workload. 4) Finally, the self-explanatory diagrams, instead of using raw data, relieve the cognitive workload of users during the monitoring of a situation.

Part III
Conclusion and
Future Research Direction

Chapter 9

Lessons Learned and Open Issues

This research thesis sought to establish how user interface design can affect the user experience, its various aspects, and how it is perceived by exploring related issues and challenges and identifying possible opportunities. The research objective was addressed through a series of studies to detect the importance of accessibility and situational awareness in positively influencing the user experience. Studies have found that accessibility and situational awareness are important elements in the user experience with the system interface. The results of these studies show that the quality of the user interface can solve the challenges of the user experience. A good user interface, capable of ensuring accessibility and high levels of situational awareness, will help users obtain information based on their needs and improve their experience. This chapter begins with a summary of the key findings from the studies presented in Chapters 3 to 8, which are grouped into 6 major research contributions. The findings are reviewed along with the aims and objectives and the original research questions outlined at the beginning of this research. The contribution outlines the main insights gained from the methodological approach used within this research. These six key contributions identified within this thesis are then discussed in reference to the relevant literature within the HCI field and beyond. The chapter concludes with a series of UX design heuristics generated from the main findings, along with some suggestions for future research.

9.1 Conclusions

The motivation for this research was guided by the main goals and objectives set. The influence of UI on a variety of UX constructs was

investigated. I looked at several design features of UIs in relation to accessibility and SA and how these can affect UX. User diversity was explored through the use of a variety of methods to understand how accessibility and SA can affect UX and overall user satisfaction.
The main conclusions for each topic covered in this thesis are summarized below:

- **Involving deaf people for inclusive UIs design (Chapter 3).**
  The results of Chapter 3 indicated and emphasized that the involvement of deaf people was and is fundamental not only for the definition of use cases and for the evaluation, but even more so for the creation and development of inclusive UIs. In addition, they also expressed considerable interest in the future stages of the research. The results showed how a highly usable and accessible UI can arouse the trust of deaf users in using the proposed solutions, considering them an excellent support for communication and learning. Finally, the preliminary longitudinal studies carried out showed how the use of the UI over time can not only improve the learning curve but also optimize the UX.

- **Design of accessible UIs to address the autonomy problems of blind people (Chapter 4).**
  In Chapter 4, an assistive UI was presented that combines accessibility, empowerment-oriented design and neural networks to address the autonomy problems of people with visual impairments. The results of the chapter show the high usability, and the high UX, of the proposed solution to support the daily activity of dressing, with a statistically significant difference between visually impaired and blind participants. Specifically, the results point to the more efficient and accurate proposed solution than traditional methods of dressing for both types of disabilities, although the visually impaired performed better than the blind. Blind people appreciate the benefits of helping one UI increase personal autonomy and are eager to experiment with other UIs.

- **Estimation of the impact in the use of UIs through the UX of people with cognitive disabilities (Chapter 5).**
  In Chapter 5, innovative UIs were presented to improve the UX of people with behavioral disorders. Specifically, a UI for learning based on geo-information was presented, which can be significantly adopted in any topic. The results showed that students with learning difficulties can benefit from this solution, helping them to reach a deeper understanding. In fact, a statistically significant improvement was obtained in terms of learning. Additionally, a solution is presented in the chapter to offer children with behavioral disorders a playful way to approach therapy using a UI on a humanoid robot, giving an extra boost to traditional therapists' work. The feedback obtained from the study was encouraging, promising good acceptance and intention to use the solution for therapeutic use in these children with behavioral disorders.

- **Definition of the accessibility perspective in the design process of mobile UIs (Chapter 6).**
  The results of Chapter 6 highlight that although usability is already recognized as a crucial aspect of mobile UIs development, only a few studies have analyzed accessibility. It turned out that, in each app, most of the accessibility guidelines were either not considered by the developers in the implementation phase or applied incorrectly. On the other hand, the lack of standardized methods for applying accessibility guidelines presents a challenge. Although catalogs exist that suggest general universal design principles to be applied, the results highlighted the lack of practical recommendations and templates to follow during development.
  Finally, the results highlight a lack of awareness on the part of developers of the different types of disabilities.

- **Estimation of the impact of the UI to manage fleets of drones through the SA of the operators (Chapter 7).**
  The design of a usable UI, and therefore effective, efficient, and satisfactory for the operator, allows to maintain high levels of SA.

In Chapter 7, an interface was presented that combines knowledge on the functional requirements of the UI for managing the fleet of drones with knowledge on the physical and cognitive constraints of humans. The results obtained are considered encouraging. The sensory isolation of the operator and the problems related to the spatial disorientation of UIs for the management of drone fleets can be solved using a UI that provides the operator with all the information regarding the mission and all the drones involved. In particular, one goal was to make predictions derived from the SA of a team of operators who manage the UI, and that sharing the UI would improve the UX. The results show that shared UI does not universally improve UX, but it does help maintain high levels of SA in off-nominal situations.

- **UIs in manufacturing plants based on guidelines to improve UXs (Chapter 8).**
  In Chapter 8, the foundations were laid to develop UIs for monitoring the production process of an industrial plant, capable of providing a better UX. A series of design guidelines were formalized, and to experiment with them, a UI was developed on an interactive multitouch wall. The results showed that the UI based on the proposed guidelines can improve the UX of a UI in manufacturing plants.

9.1.1 Main Contributions

The main contributions of this thesis are shown above, which identify six key results of each of the 2 main studies reported in Chapters 3 to 8, which are grouped into studies that deal with accessibility and SA respectively.
This research has clearly demonstrated the importance of UI in positively influencing UX.
The quantitative and qualitative results of the first six studies (Chapters 3, 4, 5, and 6) on accessibility confirmed that the involvement of disabled users is fundamental in the UIs design process in two ways; First, the needs and UX of users must reflect in the construction of the UI requirements and the definition of use cases; second, evaluation with a variety of different users helps identify accessibility problems. However, the importance of

designing accessible UIs is not yet a priority for developers. Defining methods for applying design guidelines would provide crucial support, and presently represent a challenge. The lack of awareness, together with the lack of interest, in accessibility, has not yet revealed the particular preferences in the design of accessible UIs, as demonstrated by the three user groups of the study. Deaf users, for example, have shown a preference for animations and in particular for the avatar. It was also shown that the users' judgment was also influenced by the comparison with the tools they use daily and the ability to identify with the situations of use, which indicates that previous experience was important. Interactivity has also proven to be a contributing factor to satisfaction and improving UX.
While, the results of the last two studies (Chapters 7 and 8) highlight that the factors that influenced UX in UI intended for viewing and monitoring, were usability and usefulness, where SA levels remained high. . A negative perception of UI and its elements are the main reasons for human errors, further exacerbated by the lack of SA.
Finally, the mixed methods approach revealed that the user's judgment can change, depending on the type of measures used, as the responses to the quantitative questionnaires allow us to analyze the more instantaneous effects of the UI, while the interviews stimulate more reflective judgments.

The main contributions are now reviewed in relation to the original research questions posed in Chapter 1:

RQ: *Key constructs such as accessibility and SA contribute to the UX?*
It was found that factors such as accessibility and SA contribute to the UX, also positively influence the users' overall judgment on the quality of the UI.

RQ_1: *Does accessibility in UIs lead to a more positive UX?*
Accessibility in UI is clearly important; firstly, the general experience of interaction was demonstrated by an increase in UX; and secondly, the accessibility differences in the UIs found that the higher the inclusiveness, the more positive the users' opinion. Clearly, UIs contributed the most to positive UX.

RQ$_2$: *Do UIs with high SA levels improve UX?*
Perceived usability and usefulness were the most important determinants of the adoption of UI, where it is necessary to maintain high levels of SA. Negative perception about UI was the main cause of human errors, along with low SA levels and high levels of cognitive workload. In conclusion, UIs that guarantee high levels of SA clearly lead to the improvement of UX

Furthermore, to understand how the RQs were answered, Figure 9.1 shows how I conducted the design science research following the framework of Hevner et al. [312].

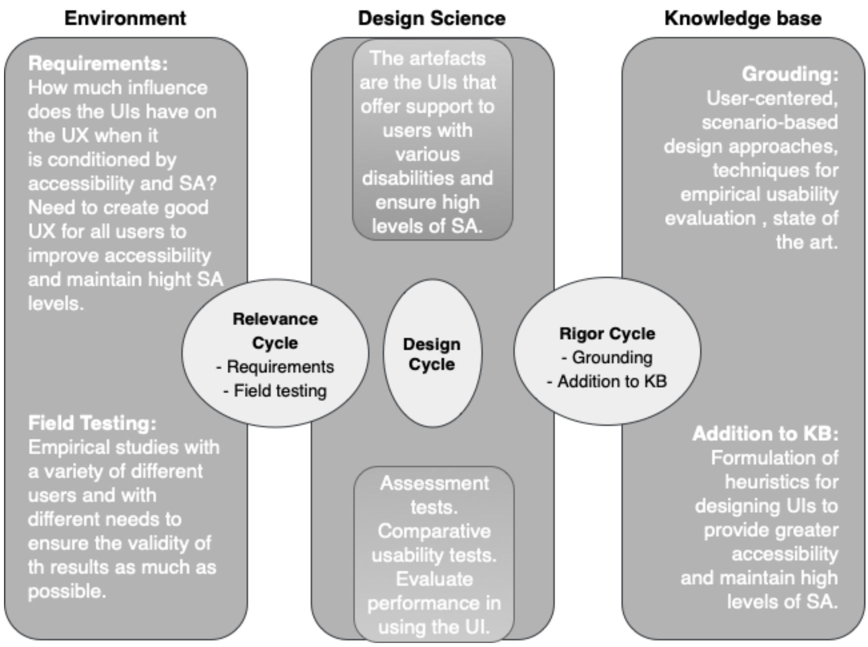


Figure 9.1: Design Science Research process for Thesis.

9.1.2 Design Science Checklist

In this section I am providing answers to the eight questions of the design science research approach. The following table associates each question to the chapters where they are satisfied and a short description.

Table 9.1: Final checklist for the Design Science Research approach.

Questions	**Chapters**	**Short Descriptions**
1. What is the research question (design requirements)?	Chapter 1	The research question **"*Key constructs such as accessibility and SA contribute to the UX?*"** is introduced in the first chapter of the thesis. This main research question is further divided into two parts: 1. *Does accessibility in UIs lead to a more positive UX?* 2. *Do UIs with high SA levels improve UX?*
2. What is the artefact? How is the artefact represented?	Chapter 3, 4, 5, 7, 8	The artefacts, which address the first research topic, are the UIs that offer support to users with various disabilities. They have been developed so that accessibility is integrated. The artefacts developed to address the second research problem are UIs to ensure high levels of SA.
3. What design processes (search heuristics) will be used to build the artefact?	Chapter 3, 4, 5, 7, 8	The design process followed a user-centered design approach. This approach starts by studying potential users and existing solutions giving attention to the needs of users at each stage of the design process.
4. How are the artefact and the design processes grounded by the knowledge base? What, if any, theories support the artefact design and the design process?	Chapter 3, 4, 5, 7, 8	The artefacts were grounded on the following: user-centered design, scenario-based design approaches, well settled techniques for empirical usability evaluation and the state of the art, also on existing relevant design patterns.
5. What evaluations are performed during the internal design cycles? What design improvements are identified during each design cycle?	Chapter 3, 4, 5, 6, 7, 8	Assessment tests were performed. Comparative usability tests have shown that the new artefacts offer significant advantages in terms of efficiency, effectiveness and satisfaction over existing common solutions. In addition, the performance in the use of the UI was also evaluated. In addition, high support for UX has been found following empirical studies on UI.
6. How is the artefact introduced into the application environment and how is it field tested? What metrics are used to	Chapter 3, 4, 5, 7, 8	The artefacts were tested in different contexts by developing UIs based on the reference environment. Furthermore, comparative tests were carried out to demonstrate the advantages

demonstrate artefact utility and improvement over previous artefacts?		in terms of effectiveness, efficiency, satisfaction or performance of the new artefacts compared to current solutions.
7. What new knowledge is added to the knowledge base and in what form (e.g., peer- reviewed literature, meta-artefacts, new theory, new method)?	Chapter 9	The knowledge base has been expanded with the formulation of heuristics for designing UIs to provide greater accessibility and maintain high levels of SA. Based on this work 11 articles were published.
8. Has the research question been satisfactorily addressed?	Chapter 9	The research question has been satisfactorily addressed. The studies performed have shown how the UIs of my solutions positively influence the UX, focusing on aspects such as accessibility and SA. The international scientific community has recognized their value as demonstrated by publications in book chapters, journals, and conference proceedings. [C1], [C2], [C3], [C4], [C5], [C7], [C8], [C12], [C14], [J1], [J2])

9.2 Lessons Learned

The main lesson learned from this thesis can be summarized as follows.

- *A participatory design and inclusive UIs improve UX for the deaf people.*
  First, as mentioned above, I argue that greater scientific involvement of deaf people should be encouraged, preferably by applying a co-design approach.
  Furthermore, the results showed that the use of GIFs and avatars within the UI is appreciated by these users. The use of sign language animation has proved to be a potentially promising method for making UIs more accessible and easier to use for deaf people. Text-to-speech and text-to-speech features were also found to be useful and intuitive, improving their UX. I consider the results obtained to be satisfactory, as the use of educational solutions contributed to the achievement of a fair level of learning by users, albeit with little experience.

- *Blind people appreciate the advantages of a mobile UI to increase their autonomy.*
  The results showed that blind people preferred the proposed solution instead of the daily method, and this preference is almost totally supported also by the visually impaired. I can say that this result refers to the importance of the acceptance by this community of UIs as support for UX. However, during the research it was found that all blind people use other assistive UIs, but not all use UIs on their smartphones as an aid. Nonetheless, mobile UIs solutions are always considered a preferable choice, even with respect to hardware support, as it is more discreet and less bulky.

- *The UX of people with cognitive disabilities represents a fundamental factor for the adoption of UIs.*
  The results showed that the proposed solutions were recognized as a valuable tool for these people. In particular, the learning progress, in using the solution designed for people with learning disabilities, was positively perceived. This perception triggered an enhancement of their UX leading to an appreciable level of UI acceptance by people with learning disabilities.
  The analysis of territorial elements, the search for information, and the study of past events through geography can contribute to a better UX by students with learning difficulties. The solution devised for the humanoid robot therapy of children suffering from behavioral disorders was also immediately accepted, attracting a high degree of attention for the proposed therapy on UI. The idea of relying on humanoid robots in the therapeutic context proved to be very interesting, as it proved to be a factor in improving UX in children with behavioral disorders.

- *The accessibility problem is not only overlooked in practice but also by developers.*
  The results showed that the lack of awareness of the different types of disabilities on the part of developers is the main challenge for the problem of accessibility. This issue, combined with the study findings, reveals the need to consider further user-centric usability

testing steps with a more diverse set of users to help identify accessibility challenges. In addition to the actions that developers could take individually, such as the user-centric usability testing processes mentioned above, the results are more general and call for the need to formulate methods for accessibility. Also, the definition of (semi) automated techniques could support developers when assessing the accessibility level of their applications and developing accessible mobile UIs.
Finally, I believe that one approach to mitigating accessibility issues is to plan for accessibility early in the design phase rather than managing it as an afterthought at the end of the development phase.

- *Choosing a shared UI to manage fleets of drones in off-nominal situations has an impact on UX.*
  Shared UI was expected to cause better SA than unshared UIs, thus improving UX as well. But there was no general boosting effect on UX. Only during unforeseen events in the use of the UI was the SA reduced, showing that it had an impact on the UX. This leads me to the conclusion that UX has worsened significantly when unexpected events occurred in the unshared UI setting. The main implication of the results obtained is that shared UIs are particularly useful for UX in off-nominal situations. I believe that this result is dictated by the fact that with the shared UI there are more users who can monitor and therefore the anomaly is more likely to be detected quickly. However, it is possible that a good SA will be able to circumvent this problem.

- *Appropriate Information Design and Interaction Paradigm may result in a better SA*
  The main findings are related to the effects on users' cognitive workload, in turn influencing SA. This is attributable not only to the appropriateness of the proposed design guidelines, but also to the multitouch interaction that provides an intuitive and familiar way to interact while improving the UX.

The results of the study showed that the information displayed hierarchically on the UI allows for more efficient and accurate analysis by allowing access only to the information needed in a given situation. Furthermore, the UX in multitouch interaction is recommended to allow you to organize the UI together with your mental model. Finally, the presentation of self-explanatory diagrams in the UI lighten the cognitive workload of users when monitoring a situation, ensuring high levels of SA.

9.2.1 Heuristics

Heuristics are often associated with usability evaluation [293], where evaluative judgments are based on a set of pre-established guidelines or "heuristics" [286]. The third and final goal of this research was to identify generic design principles and/or evaluation heuristics.
The results of this thesis show that UX is complex, where context, task, UI quality criteria, individual difference play a role in shaping the UX, which is subject to change as the experience progresses over time. To facilitate understanding, a set of eight high-level principles or heuristics are proposed. These are based on the main findings reported in this thesis with the main purpose of assisting in the design of UIs and guiding the evaluation.

i. *Task framing effects:* UX is subject to task framing effects both with tasks pursuing inclusiveness (where - in addition to usability and accessibility - interactivity and involvement are favored) and with tasks useful for guarantee a high level of SA (which favor usability and usefulness).

ii. *UI framing effect:* users prefer UIs with different elements and tools depending on the context of application and use. The integration of context specific UI elements enables better UX and raises user satisfaction.

iii. *Supporting the individual difference:* user predisposition can impact user judgment. This heuristic emphasizes the need to know users' preferences, expectations, and diversity to provide optimal UI design.

iv. *Alternative interactions:* advanced interactive features that provide alternative UI navigation methods have shown to increase user engagement but can cause frustration due to usability issues if the UI is poorly designed.

v. *Customization:* adaptability refers to the ability of a UI to customize and adapt user's needs, which can be an essential factor for long-term use, improving user engagement and UX.

vi. *User control:* high priority should be given to user control, allowing users the freedom to control their own experience. User control can be an essential aspect of the UI and can lead to satisfaction and improvement of the UX.

vii. *Provide training:* offering adequate and timely training while implementing new UIs can increase user engagement and improve UX, leading to greater user satisfaction. Training sessions should focus on how the new UIs can provide benefits to the user.

viii. *Avoid forced use:* user satisfaction can be influenced by forced mandatory use that leads to negative UX. Providing flexibility, user choice, and alternative ways of interacting with a UI can improve users' perception of control.

9.3 Open Issues

Despite the dedicated effort by the research community and despite the advances proposed in this thesis, below I propose some open questions and challenges that need to be addressed in the future.

- *Understanding additional demanding and UI elements that improve the UX of deaf people.*
  Understand additional needs of the deaf community in interacting with UIs in order to increase their UX. Our studies, in Chapter 3, showed how the proposed mobile solutions are much more appreciated by deaf users; therefore, our future goal is to provide

full final support on their mobile devices, integrating multiple sign languages. Finally, we are also interested in creating a further study dedicated to UX benchmarking, in particular for the learning activity.

- *Investigate the degree of accountability of blind people through the use of mobile UIs.*
  The study in Chapter 4 showed how blind people prefer mobile UIs. We intend, in the near future, to exploit the comments coming from this community of users and use them to improve the usability and accessibility of the mobile UIs. Therefore, we plan to conduct further studies with the blind to measure UX and, specifically, the degree of empowerment that a blind person can achieve through the use of a mobile UI.

- *Provide additional UIs to improve UX and the inclusion of people with cognitive disabilities.*
  Given the results of the solution for people with learning disabilities, we also plan to bring the solution to mobile devices. We think a mobile version of the UI can enable learning that can take place anywhere, anytime. In addition, we also plan to use the solution proposed for children with behavioral disorders based on robots within the school environment. This is because we believe that the robot can help the child with these disorders to socialize with his classmates. These UIs solutions, shown in Chapter 5, could be key to increasing UX and the inclusion of people with cognitive disabilities. And this is precisely the focus of our current research into this disability.

- *Further explore the area of accessibility to provide concrete support.*
  The results obtained from the study in Chapter 6 revealed several aspects that need to be further discussed. In particular, the different technical aspects related to the implementation of the accessibility guidelines should deserve further attention in the future. On the other hand, our goal is to further explore accessibility on a wider

set of systems, possibly considering how the same application on different operating systems can generate a different level of accessibility. Finally, perhaps most importantly, we will seek to obtain a set of accessibility design patterns that enable developers to more practically address accessibility guidelines and define new automated tools to facilitate the adoption of the accessibility guidelines.

- *Provide complex scenarios to assess the impact of UI on UX and improve it.*
  In Chapter 7, we analyzed the human factor-relevant UX implications for remote control of multiple drones.
  UX considerations must be built on historical usage data; therefore, one of the key elements of our future work will be to correctly structure and collect data in order to assess the impact of UI on UX and work towards its improvement. In addition, the next steps will involve defining new, more complex scenarios to allow the user to validate the effectiveness of the UI to inform them and to enable them to take the best action. Finally, the experiment studied only the collaboration between two operators while in nature there is also the possibility of having more operators collaborating at the same time. Therefore, further investigations are needed in this regard.

- *Experiment and formalize the UIs design guidelines in Industry 4.0.*
  Chapter 8 shows how improving the design of UIs in manufacturing plants is fundamental to let operators identify and react more effectively to potentially critical situations. This is a vital aspect for ensuring high levels of SA. For the future, we intend to test the guidelines with real plant managers and in different fields. The goal is to formalize more guidelines and, if successful, transform them into interaction design patterns for the design of UIs to be used in Industry 4.0.

In conclusion, the global contributions of this research were to demonstrate the importance of accessibility and SA in UIs, evaluating different user groups, and studying the influence on the UX. Eight high-level heuristics have been proposed as a result of this research with the aim of providing generic principles to aid the design and evaluation of UIs. However, they are provisional and therefore subject to further evaluation and open to future research.

Bibliography

[1] de Quirós, J. B., & Schrager, O. L. (1979). Neuropsychological fundamentals in learning disabilities. Academic Therapy Publications.

[2] House, W. F. (1976). Cochlear implants. Annals of Otology, Rhinology & Laryngology, 85(3_suppl), 3-3.

[3] Su, S. A., & Furuta, R. K. (1998, January). VRML-based representations of ASL fingerspelling on the World Wide Web. In Proceedings of the third international ACM conference on Assistive technologies (pp. 43-45).

[4] Vcom 3D, Inc. Accessed: Feb., 2019. [Online]. Available: www.vcom3d.com

[5] Sedgwick, E., Alkoby, K., Davidson, M. J., Roymieco, C., Juliet, C., Craft, B., ... & Wolfe, R. (2001). Toward the effective animation of American Sign Language.

[6] DePaul ASL Synthesizer. Accessed: Feb., 2019. Availabe: www.asl.cs.depaul.edu

[7] Computer Program Translates Spoken English into Sign Language. National Geographic, August, 12.

[8] P3 Mobile. Accessed: Feb., 2019. [Online]. Available: https://www.purplevrs.com/p3mobile

[9] FaceTime. Accessed: Feb., 2019. [Online]. Available: https://support.apple.com/it-it/HT204380

[10] GoogleVoice. Accessed: Feb., 2019. [Online]. Available: https://voice.google.com/u/0/about

[11] DragonAnywhere. Accessed: Feb., 2019. [Online]. Available: https://www.nuance.com/dragon/dragon-anywhere.html

[12] BuzzCards. Accessed: Feb., 2019. [Online]. Available: http://www.buzzcard.gatech.edu/

[13] ASLDictionary. Accessed: Feb., 2019. [Online]. Available: https://www.aslsignlanguagedictionary.com/

[14] MarleeSigns. Accessed: Feb., 2019. [Online]. Available: https://www.marleematlin.net/app/

[15] Sign With Robert. Accessed: Feb., 2019. [Online]. Available: http://www.signwithrobert.com/

[16] Sebillo, M., Tortora, G., Tucci, M., Vitiello, G., Ginige, A., & Di Giovanni, P. (2015). Combining personal diaries with territorial intelligence to empower diabetic patients. Journal of Visual Languages & Computing, 29, 1-14.

[17] Vitiello, G., Francese, R., Sebillo, M., Tortora, G., & Tucci, M. (2017, September). UX-Requirements for Patient's Empowerment—The Case of Multiple Pharmacological Treatments: A Case Study of IT Support to Chronic Disease Management. In 2017 IEEE 25th International Requirements Engineering Conference Workshops (REW) (pp. 139-145). IEEE.

[18] Holmes, M. (2009). English in international deaf communication. Edited by Cynthia J. Kellet Bidoli and Elana Ochse, Verlag Peter Lang, Bern, 2008, 444pp, ISBN 978-3-03911.

[19] Cassino, R., Tucci, M., Vitiello, G., & Francese, R. (2015). Empirical validation of an automatic usability evaluation method. Journal of Visual Languages & Computing, 28, 1-22.

[20] Di Giovanni, P., Romano, M., Sebillo, M., Tortora, G., Vitiello, G., Ginige, T., ... & Ginige, A. (2013). Building social life networks through mobile interfaces: the case study of Sri Lanka farmers. In Organizational Change and Information Systems (pp. 399-408). Springer, Berlin, Heidelberg.

[21] Di Giovanni, P., Bertolotto, M., Vitiello, G., & Sebillo, M. (2014). Web services composition and geographic information.

Geographical Information Systems: Trends and Technologies, 104-141.

[22] Elliott, R., Glauert, J. R., Kennaway, J. R., Marshall, I., & Safar, E. (2008). Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. Universal Access in the Information Society, 6(4), 375-391.

[23] Filhol, M. (2009). Zebedee: a lexical description model for sign language synthesis. Internal, LIMSI.

[24] Hanke, T. (2002, June). iLex-A tool for Sign Language Lexicography and Corpus Analysis. In LREC.

[25] Heloir, A., & Kipp, M. (2010). Real-time animation of interactive agents: Specification and realization. Applied Artificial Intelligence, 24(6), 510-529.

[26] Huenerfauth, M. (2009). A linguistically motivated model for speed and pausing in animations of american sign language. ACM Transactions on Accessible Computing (TACCESS), 2(2), 1-31.

[27] Huenerfauth, M., Zhao, L., Gu, E., & Allbeck, J. (2007, October). Evaluating American Sign Language generation through the participation of native ASL signers. In Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility (pp. 211-218).

[28] Huenerfauth, M., Zhao, L., Gu, E., & Allbeck, J. (2008). Evaluation of American sign language generation by native ASL signers. ACM Transactions on Accessible Computing (TACCESS), 1(1), 1-27.

[29] Johnston, T., & Schembri, A. (2007). Australian Sign Language (Auslan): An introduction to sign language linguistics. Cambridge University Press.

[30] Kennaway, J. R., Glauert, J. R., & Zwitserlood, I. (2007). Providing signed content on the Internet by synthesized animation. ACM Transactions on Computer-Human Interaction (TOCHI), 14(3), 15-es.

[31] Kipp, M., Heloir, A., Schröder, M., & Gebhard, P. (2010, September). Realizing multimodal behavior. In International Conference on Intelligent Virtual Agents (pp. 57-63). Springer, Berlin, Heidelberg.

[32] Niewiadomski, R., Bevacqua, E., Mancini, M., & Pelachaud, C. (2009, May). Greta: an interactive expressive ECA system. In Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2 (pp. 1399-1400).

[33] Prillwitz, S., Leven, R., Zienert, H., Hanke, T., & Henning, J. (1989). Hamburg Notation System for sign languages: an introductory guide, HamNoSys version 2.0. Signum, Seedorf, Germany.

[34] Sheard, M., Schoot, S., Zwitserlood, I., Verlinden, M., & Weber, I. (2004). Evaluation reports 1 and 2 of the EU project essential sign language information on government networks. Deliverable D6, 2.

[35] Stokoe Jr, W. C. (2005). Sign language structure: An outline of the visual communication systems of the American deaf. Journal of deaf studies and deaf education, 10(1), 3-37.

[36] Thiebaux, M., Marsella, S., Marshall, A. N., & Kallmann, M. (2008, May). Smartbody: Behavior realization for embodied conversational agents. In Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1 (pp. 151-158).

[37] Wolfe, R., McDonald, J., Davidson, M. J., & Frank, C. (2007, March). Using an animation-based technology to support reading curricula for deaf elementary schoolchildren. In The 22nd annual international technology & persons with disabilities conference (pp. 1-8).

[38] Convention on the Rights of Persons with Disabilities (CRPD). Accessed: Aug., 2019. [Online]. Available: https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html.

[39] Pugeault, N., & Bowden, R. (2011, November). Spelling it out: Real-time ASL fingerspelling recognition. In 2011 IEEE International conference on computer vision workshops (ICCV workshops) (pp. 1114-1119). IEEE.

[40] Wang, C., Liu, Z., & Chan, S. C. (2014). Superpixel-based hand gesture recognition with kinect depth camera. IEEE transactions on multimedia, 17(1), 29-39.

[41] Maqueda, A. I., del-Blanco, C. R., Jaureguizar, F., & García, N. (2015). Human–computer interaction based on visual hand-gesture recognition using volumetric spatiograms of local binary patterns. Computer Vision and Image Understanding, 141, 126-137.

[42] Nai, W., Liu, Y., Rempel, D., & Wang, Y. (2017). Fast hand posture classification using depth features extracted from random line segments. Pattern Recognition, 65, 1-10.

[43] Kuznetsova, A., Leal-Taixé, L., & Rosenhahn, B. (2013). Real-time sign language recognition using a consumer depth camera. In Proceedings of the IEEE international conference on computer vision workshops (pp. 83-90).

[44] Wohlkinger, W., & Vincze, M. (2011, December). Ensemble of shape functions for 3d object classification. In 2011 IEEE international conference on robotics and biomimetics (pp. 2987-2992). IEEE.

[45] Zhang, C., & Tian, Y. (2015). Histogram of 3D facets: A depth descriptor for human action and hand gesture recognition. Computer Vision and Image Understanding, 139, 29-39.

[46] Zhang, C., Yang, X., & Tian, Y. (2013, April). Histogram of 3D facets: A characteristic descriptor for hand gesture recognition. In 2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG) (pp. 1-8). IEEE.

[47] Rioux-Maldague, L., & Giguere, P. (2014, May). Sign language fingerspelling classification from depth and color images using a

deep belief network. In 2014 Canadian Conference on Computer and Robot Vision (pp. 92-97). IEEE.

[48] Keskin, C., Kıraç, F., Kara, Y. E., & Akarun, L. (2012, October). Hand pose estimation and hand shape classification using multi-layered randomized decision forests. In European Conference on Computer Vision (pp. 852-863). Springer, Berlin, Heidelberg.

[49] Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., ... & Blake, A. (2011, June). Real-time human pose recognition in parts from single depth images. In CVPR 2011 (pp. 1297-1304). Ieee.

[50] Dong, C., Leu, M. C., & Yin, Z. (2015). American sign language alphabet recognition using microsoft kinect. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 44-52).

[51] Ong, S. C., & Ranganath, S. (2005). Automatic sign language analysis: A survey and the future beyond lexical meaning. IEEE Transactions on Pattern Analysis & Machine Intelligence, 27(06), 873-891.

[52] Marcon, M., Paracchini, M. B. M., & Tubaro, S. (2019). A framework for interpreting, modeling and recognizing human body gestures through 3D eigenpostures. International Journal of Machine Learning and Cybernetics, 10(5), 1205-1226.

[53] Cui, R., Liu, H., & Zhang, C. (2017). Recurrent convolutional neural networks for continuous sign language recognition by staged optimization. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7361-7369).

[54] Madhuri, Y., Anitha, G., & Anburajan, M. (2013, February). Vision-based sign language translation device. In 2013 International Conference on Information Communication and Embedded Systems (ICICES) (pp. 565-568). IEEE.

[55] Cui, R., Liu, H., & Zhang, C. (2019). A deep neural framework for continuous sign language recognition by iterative training. IEEE Transactions on Multimedia, 21(7), 1880-1891.

[56] Panzner, M., & Cimiano, P. (2016, August). Comparing hidden markov models and long short term memory neural networks for learning action representations. In International workshop on machine learning, optimization, and big data (pp. 94-105). Springer, Cham.

[57] System Usability Scale. Accessed: Feb., 2019. [Online]. Available: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html

[58] Battistoni, P., Sebillo, M., & Vitiello, G. (2019, June). Experimenting with a fog-computing architecture for indoor navigation. In 2019 Fourth International Conference on Fog and Mobile Edge Computing (FMEC) (pp. 161-165). IEEE.

[59] Liu, Y., Fieldsend, J. E., & Min, G. (2017). A framework of fog computing: Architecture, challenges, and optimization. IEEE Access, 5, 25445-25454.

[60] Iorga, M., Feldman, L., Barton, R., Martin, M. J., Goren, N. S., & Mahmoudi, C. (2018). Fog computing conceptual model.

[61] Battistoni, P., Sebillo, M., & Vitiello, G. (2019, October). Computation offloading with MQTT protocol on a Fog-Mist Computing Framework. In International Conference on Internet and Distributed Computing Systems (pp. 140-147). Springer, Cham.

[62] Moysiadis, V., Sarigiannidis, P., & Moscholios, I. (2018). Towards distributed data management in fog computing. Wireless Communications and Mobile Computing, 2018.

[63] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

[64] Brady, E., Morris, M. R., Zhong, Y., White, S., & Bigham, J. P. (2013, April). Visual challenges in the everyday lives of blind

people. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 2117-2126).

[65] Chhaya, S. V., Khera, S., & Kumar, P. (2015). Basic geometric shape and primary color detection using image processing on Matlab. International Journal of Research in Engineering and Technology, 4(5), 505-509.

[66] Elrefaei, L. A., Al-musawa, M. O., & Al-gohany, N. A. (2017). Development of an android application for object detection based on color, shape, or local features. arXiv preprint arXiv:1703.03848.

[67] Grauman, K., & Leibe, B. (2011). Visual object recognition. Synthesis lectures on artificial intelligence and machine learning, 5(2), 1-181.

[68] Kaur, P., Ganore, M., Doiphode, R., Garud, A., & Ghuge, T. (2017). Be My Eyes: Android App for visually impaired people.

[69] L. Jones. Lessons for Living with Vision Loss. American Foundation for the Blind Accessed: Oct., 2018. Available: http://www.afb.org/info/programs-and-services/professional-development/aging/lessons-for-living-with-vision-loss/1234.

[70] Leporini, B., & Buzzi, M. (2018, April). Home automation for an independent living: investigating the needs of visually impaired people. In Proceedings of the 15th International Web for All Conference (pp. 1-9).

[71] Rege, S., Memane, R., Phatak, M., & Agarwal, P. (2013). 2D geometric shape and color recognition using digital image processing. International journal of advanced research in electrical, electronics and instrumentation engineering, 2(6), 2479-2487.

[72] V. Vella, J. P. Mathieu, M. O. Elias, S. Bedin, K. Segatti, R. Ferretti, J. Hüllen, A. Borrell. (2016). Elicitation of needs and definition of relevant ontologies report. Accessed: Sept., 2018. Available: https://www.bfw-dueren.de/projekte/abgeschlosseneprojekte/pro4vip/pro4vip_survey.pdf

[73] OrCam MyEye 2.0. Accessed: Sept., 2018. [Online]. Available: https://www.orcam.com/en

[74] Colorino - Talking Color Identifier and Light Probe. Accessed: Sept., 2018. [Online]. Available: https://independentliving.com/colorino-talking-color-identifier-and-light-probe/

[75] Feil-Seifer, D., & Mataric, M. J. (2005, June). Defining socially assistive robotics. In 9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005. (pp. 465-468). IEEE.

[76] Henkemans, O. A. B., Bierman, B. P., Janssen, J., Neerincx, M. A., Looije, R., van der Bosch, H., & van der Giessen, J. A. (2013). Using a robot to personalise health education for children with diabetes type 1: A pilot study. Patient education and counseling, 92(2), 174-181.

[77] Fridin, M. (2014). Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education. Computers & education, 70, 53-64.

[78] Han, J., Jo, M., Park, S., & Kim, S. (2005, August). The educational use of home robots for children. In ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005. (pp. 378-383). IEEE.

[79] Beran, T. N., Ramirez-Serrano, A., Vanderkooi, O. G., & Kuhn, S. (2013). Reducing children's pain and distress towards flu vaccinations: a novel and effective application of humanoid robotics. Vaccine, 31(25), 2772-2777.

[80] Beran, T. N., Ramirez-Serrano, A., Vanderkooi, O. G., & Kuhn, S. (2015). Humanoid robotics in health care: An exploration of children’s and parents’ emotional reactions. Journal of health psychology, 20(7), 984-989.

[81] Dockrell, S., Earle, D., & Galvin, R. (2010). Computer-related posture and discomfort in primary school children: The effects of a school-based ergonomic intervention. Computers & Education, 55(1), 276-284.

[82] Tanaka, F., Movellan, J. R., Fortenberry, B., & Aisaka, K. (2006, March). Daily HRI evaluation at a classroom environment: reports from dance interaction experiments. In Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction (pp. 3-9).

[83] Brooks, D. A., & Howard, A. M. (2012). Quantifying upper-arm rehabilitation metrics for children through interaction with a humanoid robot. Applied Bionics and Biomechanics, 9(2), 157-172.

[84] Landreth, G. L. (2012). Play therapy: The art of the relationship. Routledge.

[85] Scassellati, B. (2005). Using social robots to study abnormal social development.

[86] Scassellati, B. (2002). Theory of mind for a humanoid robot. Autonomous Robots, 12(1), 13-24.

[87] Vanderborght, B., Simut, R., Saldien, J., Pop, C., Rusu, A. S., Pintea, S., ... & David, D. O. (2012). Using the social robot probo as a social story telling agent for children with ASD. Interaction Studies, 13(3), 348-372.

[88] Robins, B., Dautenhahn, K., Boekhorst, R. T., & Billard, A. (2005). Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills?. Universal access in the information society, 4(2), 105-120.

[89] Prizant, B. M., Audet, L. R., Burke, G. M., Hummel, L. J., Maher, S. R., & Theadore, G. (1990). Communication disorders and emotional/behavioral disorders in children and adolescents. Journal of Speech and Hearing Disorders, 55(2), 179-192.

[90] Schachar, R., & Wachsmuth, R. (1990). Oppositional disorder in children: a validation study comparing conduct disorder, oppositional disorder and normal control children. Journal of child psychology and psychiatry, 31(7), 1089-1102.

[91] Tapus, A., Maja, M., & Scassellatti, B. (2007). The grand challenges in socially assistive robotics. IEEE Robotics and Automation Magazine, 14(1), N-A.

[92] Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R., & Scassellati, B. (2013). Social robots as embedded reinforcers of social behavior in children with autism. Journal of autism and developmental disorders, 43(5), 1038-1049.

[93] SoftBank Pepper. Accessed: Oct., 2020. [Online]. Available: https://www.softbankrobotics.com/emea/en/pepper

[94] Clarke, A. R., Barry, R. J., McCarthy, R., & Selikowitz, M. (1998). EEG analysis in attention-deficit/hyperactivity disorder: a comparative study of two subtypes. Psychiatry research, 81(1), 19-29.

[95] Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., ... & Westerberg, H. (2005). Computerized training of working memory in children with ADHD-a randomized, controlled trial. Journal of the American Academy of child & adolescent psychiatry, 44(2), 177-186.

[96] Peng, X., Lin, P., Zhang, T., & Wang, J. (2013). Extreme learning machine-based classification of ADHD using brain structural MRI data. PloS one, 8(11), e79476.

[97] Prins, P. J., Dovis, S., Ponsioen, A., Ten Brink, E., & Van Der Oord, S. (2011). Does computerized working memory training with game elements enhance motivation and training efficacy in children with ADHD?. Cyberpsychology, behavior, and social networking, 14(3), 115-122.

[98] Heerink, M., Krose, B., Evers, V., & Wielinga, B. (2009, September). Measuring acceptance of an assistive social robot: a suggested toolkit. In RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication (pp. 528-533). IEEE.

[99] Susskind, J., Anderson, A., & Hinton, G. E. (2010). The Toronto face dataset. Technical Report UTML TR 2010-001, U. Toronto.

[100] Fer2013 Dataset. Accessed: Jun., 2020. [Online]. Available: https://www.kaggle.com/deadskull7/fer2013

[101] Al-Subaihin, A. A., Al-Khalifa, A. S., & Al-Khalifa, H. S. (2013, August). Accessibility of mobile web apps by screen readers of touch-based mobile phones. In International Conference on Mobile Web and Information Systems (pp. 35-43). Springer, Cham.

[102] Android. Accessibility Developer Guidelines. Accessed: Mar., 2020. [Online]. Available: https://developer.android.com/guide/topics/ui/accessibility

[103] Apple. Accessibility Developer Guideline. Accessed: Mar., 2020. [Online]. Available: https://developer.apple.com/ accessibility/

[104] Araújo, M. C., Façanha, A. R., Darin, T. G., Sánchez, J., Andrade, R., & Viana, W. (2017, July). Mobile audio games accessibility evaluation for users who are blind. In International Conference on Universal Access in Human-Computer Interaction (pp. 242-259). Springer, Cham.

[105] Armaly, A., & McMillan, C. (2016, May). An empirical study of blindness and program comprehension. In Proceedings of the 38th International Conference on Software Engineering Companion (pp. 683-685).

[106] Armaly, A., Rodeghero, P., & McMillan, C. (2017). A comparison of program comprehension strategies by blind and sighted programmers. IEEE Transactions on Software Engineering, 44(8), 712-724.

[107] Armaly, A., Rodeghero, P., & McMillan, C. (2018, September). AudioHighlight: Code skimming for blind programmers. In 2018 IEEE International Conference on Software Maintenance and Evolution (ICSME) (pp. 206-216). IEEE.

[108] Ballantyne, M., Jha, A., Jacobsen, A., Hawker, J. S., & El-Glaly, Y. N. (2018, November). Study of accessibility guidelines of mobile applications. In Proceedings of the 17th international conference on mobile and ubiquitous multimedia (pp. 305-315).

[109] BBC. Mobile Accessibility Guidelines. Accessed: Mar., 2020. [Online]. Available: https://www.bbc.co.uk/accessibility/forproducts/guides/mobile

[110] Creswell, J. W., & Creswell, J. D. (2017). Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

[111] Darvishy, A. (2014, June). Accessibility of mobile platforms. In International Conference of Design, User Experience, and Usability (pp. 133-140). Springer, Cham.

[112] Di Gregorio, M., Di Nucci, D., Palomba, F., & Vitiello, G. (2020, September). The Making of Accessible Android Applications: An Empirical Study on the State of the Practice. In 2020 IEEE International Conference on Software Maintenance and Evolution (ICSME) (pp. 857-861). IEEE.

[113] M. Di Gregorio, D. Di Nucci, F. Palomba, and G. Vitiello. The making of accessible android applications: An empirical study on the state of the practice—replication package. https://tinyurl.com/accessible-mobile-apps, 2021.

[114] Díaz-Bossini, J. M., & Moreno, L. (2014). Accessibility to mobile interfaces for older people. Procedia Computer Science, 27, 57-66.

[115] Díaz-Bossini, J. M., Moreno, L., & Martínez, P. (2014, June). Towards mobile accessibility for older people: a user centered evaluation. In International Conference on Universal Access in Human-Computer Interaction (pp. 58-68). Springer, Cham.

[116] Flatla, D. R. (2011, October). Accessibility for individuals with color vision deficiency. In Proceedings of the 24th annual ACM symposium adjunct on User interface software and technology (pp. 31-34).

[117] Friedman, M. G., & Bryen, D. N. (2007). Web accessibility design recommendations for people with cognitive disabilities. Technology and disability, 19(4), 205-212.

[118] Geiger, F. X., Malavolta, I., Pascarella, L., Palomba, F., Di Nucci, D., & Bacchelli, A. (2018, May). A graph-based dataset of commit history of real-world android apps. In Proceedings of the 15th International Conference on Mining Software Repositories (pp. 30-33).

[119] Harper, S., & Chen, A. Q. (2012). Web accessibility guidelines. World Wide Web, 15(1), 61-88.

[120] Harper, S., & Yesilada, Y. (Eds.). (2008). Web accessibility: a foundation for research. Springer Science & Business Media.

[121] IBM. IBM Accessibility Requirements. Accessed: Mar., 2020. [Online]. Available: https://www.ibm.com/able/guidelines/ci162/accessibility_checklist.html

[122] Iwarsson, S., & Ståhl, A. (2003). Accessibility, usability and universal design—positioning and definition of concepts describing person-environment relationships. Disability and rehabilitation, 25(2), 57-66.

[123] Kocieliński, D., & Brzostek-Pawłowska, J. (2013, September). Improving the accessibility of touchscreen-based mobile devices: Integrating Android-based devices and Braille notetakers. In 2013 Federated Conference on Computer Science and Information Systems (pp. 655-658). IEEE.

[124] Krainz, E., Miesenberger, K., & Feiner, J. (2018, July). Can we improve app accessibility with advanced development methods?. In International Conference on Computers Helping People with Special Needs (pp. 64-70). Springer, Cham.

[125] Kruchten, P., Nord, R. L., & Ozkaya, I. (2012). Technical debt: From metaphor to theory and practice. Ieee software, 29(6), 18-21.

[126] Lawrence, S., & Giles, C. L. (2000). Accessibility of information on the web. intelligence, 11(1), 32-39.

[127] Leporini, B., Buzzi, M. C., & Buzzi, M. (2012, November). Interacting with mobile devices via VoiceOver: usability and

accessibility issues. In Proceedings of the 24th Australian Computer-Human Interaction Conference (pp. 339-348).

[128] Martin, W., Sarro, F., Jia, Y., Zhang, Y., & Harman, M. (2016). A survey of app store analysis for software engineering. IEEE transactions on software engineering, 43(9), 817-847.

[129] McMillan, C., & Rodda-Tyler, A. (2016, May). Collaborative software engineering education between college seniors and blind high school students. In 2016 IEEE/ACM 38th International Conference on Software Engineering Companion (ICSE-C) (pp. 360-363). IEEE.

[130] Mehta, Y., Joshi, A., Joshi, M., & Jadhav, C. (2016, December). Accessibility of date picker for touchscreens. In Proceedings of the 8th Indian Conference on Human Computer Interaction (pp. 64-69).

[131] Milne, L. R., Bennett, C. L., & Ladner, R. E. (2014, December). The accessibility of mobile health sensors for blind users. In International Technology and Persons with Disabilities Conference Scientific/Research Proceedings (CSUN 2014) (pp. 166-175).

[132] Paciello, M. (2000). Web accessibility for people with disabilities. Crc Press.

[133] Park, K., Goh, T., & So, H. J. (2014). Toward accessible mobile application design: developing mobile application accessibility guidelines for people with visual impairment. 한국 HCI 학회 학술대회, 31-38.

[134] Patil, N., Bhole, D., & Shete, P. (2016, September). Enhanced UI Automator Viewer with improved Android accessibility evaluation features. In 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT) (pp. 977-983). IEEE.

[135] Ross, A. S., Zhang, X., Fogarty, J., & Wobbrock, J. O. (2018, October). Examining image-based button labeling for accessibility in Android apps through large-scale analysis. In Proceedings of the

20th International ACM SIGACCESS Conference on Computers and Accessibility (pp. 119-130).

[136] Rossi, P. H., Wright, J. D., & Anderson, A. B. (Eds.). (2013). Handbook of survey research. Academic press.

[137] Rutter, R., Lauke, P. H., Waddell, C., Thatcher, J., Henry, S. L., Lawson, B., ... & Urban, M. (2007). Web accessibility: Web standards and regulatory compliance. Apress.

[138] Serra, L. C., Carvalho, L. P., Ferreira, L. P., Vaz, J. B. S., & Freire, A. P. (2015). Accessibility evaluation of e-government mobile applications in Brazil. Procedia Computer Science, 67, 348-357.

[139] Sevilla, J., Herrera, G., Martínez, B., & Alcantud, F. (2007). Web accessibility for individuals with cognitive deficits: A comparative study between an existing commercial web and its cognitively accessible equivalent. ACM Transactions on Computer-Human Interaction (TOCHI), 14(3), 12-es.

[140] Sierkowski, B. (2002, November). Achieving web accessibility. In Proceedings of the 30th annual ACM SIGUCCS conference on User services (pp. 288-291).

[141] Sloan, D., Heath, A., Hamilton, F., Kelly, B., Petrie, H., & Phipps, L. (2006, May). Contextual web accessibility-maximizing the benefit of accessibility guidelines. In Proceedings of the 2006 international cross-disciplinary workshop on Web accessibility (W4A): Building the mobile web: rediscovering accessibility? (pp. 121-131).

[142] Statista. Number of smartphone users worldwide. Accessed: Mar., 2020. [Online]. Available: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

[143] Topp, L., Barker, B., & Degenhardt, L. (2004). The external validity of results derived from ecstasy users recruited using purposive sampling strategies. Drug and alcohol dependence, 73(1), 33-40.

[144] Turkle, S. (Ed.). (2011). Evocative objects: Things we think with. MIT press.

[145] Vassallo, C., Panichella, S., Palomba, F., Proksch, S., Gall, H. C., & Zaidman, A. (2020). How developers engage with static analysis tools in different contexts. Empirical Software Engineering, 25(2), 1419-1457.

[146] Vaughn, P., & Turner, C. (2016). Decoding via coding: Analyzing qualitative text data through thematic coding and survey methodologies. Journal of Library Administration, 56(1), 41-51.

[147] Vendome, C., Solano, D., Liñán, S., & Linares-Vásquez, M. (2019, September). Can everyone use my app? an empirical study on accessibility in android apps. In 2019 IEEE International Conference on Software Maintenance and Evolution (ICSME) (pp. 41-52). IEEE.

[148] W3C. Mobile Accessibility at W3C. Accessed: Mar., 2020. [Online]. Available: https://www.w3.org/WAI/standards-guidelines/

[149] Walker, B. N., Tomlinson, B. J., & Schuett, J. H. (2017, July). Universal design of mobile apps: Making weather information accessible. In International Conference on Universal Access in Human-Computer Interaction (pp. 113-122). Springer, Cham.

[150] Wasserman, A. I. (2010, November). Software engineering issues for mobile application development. In Proceedings of the FSE/SDP workshop on Future of software engineering research (pp. 397-400).

[151] Wentz, B., & Tressler, K. (2017). Exploring the accessibility of banking and finance systems for blind users. First Monday.

[152] Xie, Z., Li, N., & Luo, L. (2015, January). A study and implementation of vga multi-resolution on android platform. In 2015 International Conference on Computer and Computational Sciences (ICCCS) (pp. 110-115). IEEE.

[153] Yan, S., & Ramachandran, P. G. (2019). The current status of accessibility in mobile apps. ACM Transactions on Accessible Computing (TACCESS), 12(1), 1-31.

[154] Zhong, Y., Weber, A., Burkhardt, C., Weaver, P., & Bigham, J. P. (2015, May). Enhancing Android accessibility for users with hand

tremor by reducing fine pointing and steady tapping. In Proceedings of the 12th International Web for All Conference (pp. 1-10).

[155] Dyslexia International (2017) – Better Training, Better Teaching. Accessed: Sept., 2018. Available: https://www.dyslexia-international.org/wp-content/uploads/2016/04/DI-Duke-Report-final-4-29-14.pdf

[156] Cidrim, L., & Madeiro, F. (2017). Information and Communication Technology (ICT) applied to dyslexia: literature review. Revista CEFAC, 19, 99-108.

[157] Adam, T., & Tatnall, A. (2017). The value of using ICT in the education of school students with learning difficulties. Education and Information Technologies, 22(6), 2711-2726.

[158] Williams, P., Jamali, H. R., & Nicholas, D. (2006, July). Using ICT with people with special education needs: what the literature tells us. In Aslib Proceedings. Emerald Group Publishing Limited.

[159] Thomas, M. (Ed.). (1992). IT and students with emotional and behavioural difficulties. Coventry: National Council for Educational Technology.

[160] Johnson, R., & Hegarty, J. R. (2003). Websites as educational motivators for adults with learning disability. British Journal of Educational Technology, 34(4), 479-486.

[161] Brooks, B. M., Rose, F. D., Attree, E. A., & Elliot-Square, A. (2002). An evaluation of the efficacy of training people with learning disabilities in a virtual environment. Disability and rehabilitation, 24(11-12), 622-626.

[162] Rose, F. D., Brooks, B. M., & Attree, E. A. (2002). An exploratory investigation into the usability and usefulness of training people with learning disabilities in a virtual environment. Disability and Rehabilitation, 24(11-12), 627-633.

[163] Kazakou, M., Soulis, S., Morfidi, E., & Mikropoulos, T. A. (2011). Phonological Awareness Software for Dyslexic Children. themes in science and technology education, 4(1), 33-51.

[164] Larcher, J. (2000). Information technology for children with language difficulties. Language Difficulties in an Educational Context, 13-146.

[165] Nisbet, P., & Poon, P. (1998). Special access technology. CALL Centre, University of Edinburgh.

[166] Nisbet, P., Spooner, R. W. S., Arthur, E., & Whittaker, P. (1999). Supportive writing technology.

[167] Dickinson, A., Gregor, P., & Newell, A. F. (2002, July). Ongoing investigation of the ways in which some of the problems encountered by some dyslexics can be alleviated using computer techniques. In Proceedings of the fifth international ACM conference on Assistive technologies (pp. 97-103).

[168] Keates, A. (2013). Dyslexia and information and communications technology: A guide for teachers and parents. Routledge.

[169] Rahman, F. A., Mokhtar, F., Alias, N. A., & Saleh, R. (2012). Multimedia elements as instructions for dyslexic children. International Journal of Education and Information Technologies, 6(2), 193-200.

[170] Ali, A., & Sah, M. (2017, December). Adaptive game-based e-leaming using semantic web technologies. In 2017 International Conference on Open Source Systems & Technologies (ICOSST) (pp. 15-23). IEEE.

[171] Anshari, M., Alas, Y., Sabtu, N. P. H., & Hamid, M. S. A. (2016). Online Learning: trends, issues and challenges in the Big Data Era. Journal of e-Learning and Knowledge Society, 12(1).

[172] Ruiz-Primo, M. A., Briggs, D., Iverson, H., Talbot, R., & Shepard, L. A. (2011). Impact of undergraduate science course innovations on learning. Science, 331(6022), 1269-1270.

[173] Borges, P. R. S., & Silveira, I. F. (2019). Adding and segmenting educational videos: Experiences of teacher users in an educational portal. IEEE Access, 7, 87996-88011.

[174] Bradley, C., & Boyle, T. (2004). The design, development, and use of multimedia learning objects. Journal of Educational Multimedia and Hypermedia, 13(4), 371-389.

[175] Bratsas, C., Chrysou, D. E., Eftychiadou, E., Kontokostas, D., Bamidis, P. D., & Antoniou, I. (2012, January). Semantic Web game based learning: An i18n approach with Greek DBpedia. In LiLe@ WWW.

[176] Crowther, M. S., Keller, C. C., & Waddoups, G. L. (2004). Improving the quality and effectiveness of computer-mediated instruction through usability evaluations. British Journal of Educational Technology, 35(3), 289-303.

[177] Dodero, G. M., & Del Fatto, V. (2013). Geographic learning objects in smart cities context. Interaction Design and Architecture (s), 17, 53-66.

[178] Dix, A., Roselli, T., & Sutinen, E. (2006). E-learning and human-computer interaction: Exploring design synergies for more effective learning experiences. Journal of Educational Technology & Society, 9(4), 1-2.

[179] Downes, S. (2005). E-learning 2.0. ELearn, 2005(10), 1.

[180] DCMI. Dublin Core Metadata Initiative. Accessed: Mar., 2019. [Online]. Available: http://Dublincore.org

[181] El Fazazi, H., Qbadou, M., & Mansouri, K. (2016). Using semantic Web technologies for e-learning system in higher education: Case of scheduling of upgrade trainings. Int. J. Multi-Disciplinary Sci., 3, 65-69.

[182] Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. Proceedings of the national academy of sciences, 111(23), 8410-8415.

[183] Grivokostopoulou, F., Perikos, I., & Hatzilygeroudis, I. (2014, December). Using semantic web Technologies in a web Based

System for personalized learning AI course. In 2014 IEEE Sixth International Conference on Technology for Education (pp. 257-260). IEEE.

[184] Hitzler, P., Krotzsch, M., & Rudolph, S. (2009). Foundations of semantic web technologies. Chapman and Hall/CRC.

[185] Kaplan, A. M., & Haenlein, M. (2016). Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster. Business horizons, 59(4), 441-450.

[186] Kay, R. (2007). A systematic evaluation of learning objects for secondary school students. Journal of Educational Technology Systems, 35(4), 411-448.

[187] Lim, C. P., Lee, S. L., & Richards, C. (2006). Developing interactive learning objects for a computing mathematics module. International Journal on E-Learning, 5(2), 221-244.

[188] Mangaroska, K., & Giannakos, M. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. IEEE Transactions on Learning Technologies, 12(4), 516-534.

[189] Modular Object-Oriented Dynamic Learning Environment (Moodle), Moodle Pty Ltd, West Perth, WA, Australia, 2020. Accessed: Mar., 2019. [Online]. Available: http://Moodle.org

[190] Mouromtsev, D., Kozlov, F., Parkhimovich, O., & Zelenina, M. (2013, October). Development of an ontology-based e-learning system. In International Conference on Knowledge Engineering and the Semantic Web (pp. 273-280). Springer, Berlin, Heidelberg.

[191] Protégé 5.5.0. Accessed: Mar., 2019. [Online]. Available: https://protege.stanford.edu/

[192] Rani, M., Nayak, R., & Vyas, O. P. (2015). An ontology-based adaptive personalized e-learning system, assisted by software agents on cloud storage. Knowledge-Based Systems, 90, 33-48.

[193] RDF/XML Syntax. Accessed: Mar., 2018. [Online]. Available: https://www.w3.org/TR/rdf-syntax-grammar/

[194] Resource Description Framework (RDF). Accessed: Mar., 2018. [Online]. Available: https://www.w3.org/RDF/

[195] Rocha, O. R., & Zucker, C. F. (2017, July). Automatic generation of educational quizzes from domain ontologies. In EDULEARN 2017-9th International Conference on Education and New Learning Technologies (pp. 4024-4030).

[196] Santos, M. E. C., Chen, A., Taketomi, T., Yamamoto, G., Miyazaki, J., & Kato, H. (2013). Augmented reality learning experiences: Survey of prototype design and evaluation. IEEE Transactions on learning technologies, 7(1), 38-56.

[197] Shmelev, V., Karpova, M., & Dukhanov, A. (2015). An approach of learning path sequencing based on revised Bloom's taxonomy and domain ontologies with the use of genetic algorithms. Procedia Computer Science, 66, 711-719.

[198] Tatar, D., Roschelle, J., Vahey, P., & Penuel, W. R. (2003). Handhelds go to school: Lessons learned. Computer, 36(09), 30-37.

[199] Tunde, F., Sunday, A., & Perpetual, O. (2015). Ontology-based model for E-learning management system (O-Bmems). International Journal of Computer Science Issues (IJCSI), 12(3), 118.

[200] Yarandi, M., Jahankhani, H., & Tawil, A. (2013). A personalized adaptive e-learning approach based on semantic web technology. webology, 10(2), Art-110.

[201] Web Ontology Language (OWL). Accessed: Mar., 2018. [Online]. Available: https://www.w3.org/OWL/

[202] Geo-Education: Essential Preparation for an Interconnected World. Accessed: Dec., 2018. [Online]. Available: https://www. nationalgeographic.org/education/geo-education-essential-preparation- interconnected-world/

[203] Cooke, N. J. (2006, October). Human factors of remotely operated vehicles. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 50, No. 1, pp. 166-169). Sage CA: Los Angeles, CA: SAGE Publications.

[204] McCarley, J. S., & Wickens, C. D. (2004). Human factors concerns in UAV flight. University of Illinois at Urbana-Champaign Institute of Aviation, Aviation Human Factors Division.

[205] Williams, K. W. (2004). A summary of unmanned aircraft accident/incident data: Human factors implications. Federal Aviation Administration Oklahoma City OK Civil Aeromedical Inst.

[206] T.B. Sheridan, "Human and Automation", a John Wiley & Son, Inc., Santa Monica, Publications, 2002. Sheridan, T. B. (2002). Human and Automation. A John Willey & Sons. Inc., Santa Monica.

[207] Tso, K. S., Tharp, G. K., Zhang, W., & Tai, A. T. (1999, October). A multi-agent operator interface for unmanned aerial vehicles. In Gateway to the New Millennium. 18th Digital Avionics Systems Conference. Proceedings (Cat. No. 99CH37033) (Vol. 2, pp. 6-A). IEEE.

[208] Wickens, C. D., Gordon, S. E., Liu, Y., & Lee, J. (2004). An introduction to human factors engineering (Vol. 2). Upper Saddle River, NJ: Pearson Prentice Hall.

[209] Draper, M., Calhoun, G., Ruff, H., Williamson, D., & Barry, T. (2003, October). Manual versus speech input for unmanned aerial vehicle control station operations. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 47, No. 1, pp. 109-113). Sage CA: Los Angeles, CA: SAGE Publications.

[210] Sabikan, S., & Nawawi, S. W. (2016). Open-source project (OSPs) platform for outdoor quadcopter. Journal of Advanced Research Design, 24, 13-27.

[211] Qian, G. M., Pebrianti, D., Chun, Y. W., Hao, Y. H., & Bayuaji, L. (2017, October). Waypoint navigation of quad-rotor MAV. In 2017 7th IEEE International Conference on System Engineering and Technology (ICSET) (pp. 38-42). IEEE.

[212] Mohammed, F., Idries, A., Mohamed, N., Al-Jaroodi, J., & Jawhar, I. (2014, May). UAVs for smart cities: Opportunities and challenges.

In 2014 International Conference on Unmanned Aircraft Systems (ICUAS) (pp. 267-273). IEEE.

[213] Cooke, N. J. (2006, October). Human factors of remotely operated vehicles. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 50, No. 1, pp. 166-169). Sage CA: Los Angeles, CA: SAGE Publications.

[214] McCarley, J. S., & Wickens, C. D. (2004). Human factors concerns in UAV flight. University of Illinois at Urbana-Champaign Institute of Aviation, Aviation Human Factors Division.

[215] Barr, L. C., Newman, R., Ancel, E., Belcastro, C. M., Foster, J. V., Evans, J., & Klyde, D. H. (2017). Preliminary risk assessment for small unmanned aircraft systems. In 17th AIAA Aviation Technology, Integration, and Operations Conference (p. 3272).

[216] Cak, S., Say, B., & Misirlisoy, M. (2020). Effects of working memory, attention, and expertise on pilots' situation awareness. Cognition, Technology & Work, 22(1), 85-94.

[217] Endsley, M.R. (2000) Theoretical underpinnings of situation awareness: A critical review. In Situation Awareness Analysis and Measurement; Endsley, M.R., Garland, D.J., Eds.; LEA: Mahwah, NJ, USA, 2000; pp. 3–32.

[218] Vicente, K. J. (2002). Ecological interface design: Progress and challenges. Human factors, 44(1), 62-78.

[219] Ensley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. Human factors, 37, 85-104.

[220] Salas, E., Prince, C., Baker, D. P., & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. Human factors, 37(1), 123-136.

[221] Stout, R. J., Cannon-Bowers, J. A., & Salas, E. (2017). The role of shared mental models in developing team situational awareness: Implications for training. In Situational awareness (pp. 287-318). Routledge.

[222] Bolstad, C. A., & Endsley, M. R. (1999, September). Shared mental models and shared displays: An empirical evaluation of team performance. In proceedings of the human factors and ergonomics society annual meeting (Vol. 43, No. 3, pp. 213-217). Sage CA: Los Angeles, CA: SAGE Publications.

[223] Bolstad, C. A., & Endsley, M. R. (2000, July). The effect of task load and shared displays on team situation awareness. In Proceedings of the Human factors and ergonomics society annual meeting (Vol. 44, No. 1, pp. 189-192). Sage CA: Los Angeles, CA: SAGE Publications.

[224] Sabikan, S., & Nawawi, S. W. (2016). Open-source project (OSPs) platform for outdoor quadcopter. Journal of Advanced Research Design, 24, 13-27.

[225] Zhou, Y., Hou, J., & Gong, Y. (2020, December). Research and Application of Human-computer Interaction Technology based on Voice Control in Ground Control Station of UAV. In 2020 IEEE 6th International Conference on Computer and Communications (ICCC) (pp. 1257-1262). IEEE.

[226] Qian, G. M., Pebrianti, D., Chun, Y. W., Hao, Y. H., & Bayuaji, L. (2017, October). Waypoint navigation of quad-rotor MAV. In 2017 7th IEEE International Conference on System Engineering and Technology (ICSET) (pp. 38-42). IEEE.

[227] Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. J. (2000). Measuring team knowledge. Human factors, 42(1), 151-173.

[228] Cooke, N. J., Kiekel, P. A., & Helm, E. E. (2001). Measuring team knowledge during skill acquisition of a complex task. International Journal of Cognitive Ergonomics, 5(3), 297-315.

[229] Sebok, A. (2000). Team performance in process control: influences of interface design and staffing levels. Ergonomics, 43(8), 1210-1236.

[230] Dixon, S. R., Wickens, C. D., & Chang, D. (2005). Mission control of multiple unmanned aerial vehicles: A workload analysis. Human factors, 47(3), 479-487.

[231] Friedrich, M., & Lieb, J. (2019, September). A novel human machine interface to support supervision and guidance of multiple highly automated unmanned aircraft. In 2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC) (pp. 1-7). IEEE.

[232] Bellucci, A., Romano, M., Aedo, I., & Díaz, P. (2016). Software support for multitouch interaction: the end-user programming perspective. IEEE Pervasive Computing, 15(1), 78-86.

[233] Blackmon, M. H., Polson, P. G., Kitajima, M., & Lewis, C. (2002, April). Cognitive walkthrough for the web. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 463-470).

[234] Brandolese, A. M. N., Pozzetti, A., & Sianesi, A. (2004). Gestione della Produzione Industriale - principi, metodologie applicazioni e misure di prestazione. Hoepli, 2004.

[235] Cannan, J., & Hu, H. (2011). Human-machine interaction (HMI): A survey. University of Essex.

[236] Cruz-Benito, J., Garcia-Penalvo, F. J., & Theron, R. (2019). Analyzing the software architectures supporting HCI/HMI processes through a systematic review of the literature. Telematics and Informatics, 38, 118-132.

[237] Di Chiara, G., Paolino, L., Romano, M., Sebillo, M., Tortora, G., Vitiello, G., & Ginige, A. (2011, September). The framy user interface for visually-impaired users. In 2011 Sixth International Conference on Digital Information Management (pp. 36-41). IEEE.

[238] Dix, A. (2017). Human–computer interaction, foundations and new paradigms. Journal of Visual Languages & Computing, 42, 122-134.

[239] Errington, J., Reising, D. V. C., Bullemer, P., DeMaere, T., Coppard, D., Doe, K., & Bloom, C. (2005, September). Establishing human performance improvements and economic benefit for a human-

centered operator interface: An industrial evaluation. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 49, No. 23, pp. 2036-2040). Sage CA: Los Angeles, CA: SAGE Publications.

[240] Ginige, A., Romano, M., Sebillo, M., Vitiello, G., & Di Giovanni, P. (2012, May). Spatial data and mobile applications: general solutions for interface design. In Proceedings of the International Working Conference on Advanced Visual Interfaces (pp. 189-196).

[241] Groover, M. P. (2016). Automation, production systems, and computer-integrated manufacturing. Pearson Education India.

[242] Hewett, T. T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., ... & Verplank, W. (1992). ACM SIGCHI curricula for human-computer interaction. ACM.

[243] Hoc, J. M. (2000). From human–machine interaction to human–machine cooperation. Ergonomics, 43(7), 833-843.

[244] Hollifield, B. (2012, August). A High Performance HMI: Better Graphics for Operations Effectiveness. In 2012 Water/Wastewater and Automation Controls Symposium.

[245] Hollifield, B. (2013). High Performance HMI–Proof Testing in Real-World Trials. In 2013 ISA Water/Wastewater and Automatic Controls Symposium.

[246] Hansen, R. C. (2001). Overall equipment effectiveness: a powerful production/maintenance tool for increased profits. Industrial Press Inc..

[247] Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage publications.

[248] He, F., Shen, K., Lu, L., & Tong, Y. (2018). Model for improvement of overall equipment effectiveness of beerfilling lines. Advances in Mechanical Engineering, 10(8), 1687814018789247.

[249] Hedman, R., Subramaniyan, M., & Almström, P. (2016). Analysis of critical factors for automatic measurement of OEE. Procedia Cirp, 57, 128-133.

[250] Purba, H. H., Wijayanto, E., & Aristiara, N. (2018). Analysis of overall equipment effectiveness (OEE) with total productive maintenance method on jig cutting: a case study in manufacturing industry. J. Sci. Eng. Res, 5, 397-406.

[251] Liker, J. (2004). The toyota way. Esensi.

[252] Liff, S., & Posey, P. A. (2004). Seeing is believing: how the new art of visual management can boost performance throughout your organization. AMACOM/American Management Association.

[253] Lu, Y. (2017). Industry 4.0: A survey on technologies, applications and open research issues. Journal of industrial information integration, 6, 1-10.

[254] Monga, R., & Khurana, V. (2015). Facility layout planning: a review. International Journal of Innovative Research in Science, Engineering and Technology, 4(3), 2347-6710.

[255] Muthiah, K. M. N., & Huang, S. H. (2007). Overall throughput effectiveness (OTE) metric for factory-level performance monitoring and bottleneck detection. International Journal of Production Research, 45(20), 4753-4769.

[256] Nachiappan, R. M., & Anantharaman, N. (2006). Evaluation of overall line effectiveness (OLE) in a continuous product line manufacturing system. Journal of Manufacturing Technology Management.

[257] Nayak, D. M., Vijaya Kumar, M., Naidu, G. S., & Shankar, V. (2013). Evaluation of OEE in a continuous process industry on an insulation line in a cable manufacturing unit. International Journal of Innovative Research in Science, Engineering and Technology, 2(5).

[258] Nielsen, J. (2005). Ten usability heuristics.

[259] Oechsner, R., Pfeffer, M., Pfitzner, L., Binder, H., Müller, E., & Vonderstrass, T. (2002). From overall equipment efficiency (OEE)

to overall Fab effectiveness (OFE). Materials science in semiconductor processing, 5(4-5), 333-339.

[260] Chikwendu, C., & Okechukwu, C. (2016). Plant layouts' analysis and design. Int. J. Adv. Eng. Technol. E-Int J Adv Engg Tech, 201-206.

[261] Paolino, L., Romano, M., Tortora, G., & Vitiello, G. (2013, July). Spatial data visualization on mobile interface-A usability study. In 2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC) (pp. 959-963). IEEE.

[262] Pinedo, M. L. (2016). Scheduling: theory, algorithms, and systems. 5-th ed. Cham.

[263] Wächter, M., Hoffmann, H., & Bullinger, A. C. (2018, August). Towards an engineering process to design usable tangible human-machine interfaces. In Congress of the International Ergonomics Association (pp. 136-147). Springer, Cham.

[264] Meixner, G., Paternò, F., & Vanderdonckt, J. (2011). Past, present, and future of model-based user interface development.

[265] Rusu, C., Rusu, V., Roncagliolo, S., & González, C. (2015). Usability and user experience: what should we care about?. International Journal of Information Technologies and Systems Approach (IJITSA), 8(2), 1-12.

[266] Mubin, O., Tejlavwala, D., Arsalan, M., Ahmad, M., & Simoff, S. (2018). An assessment into the characteristics of award winning papers at CHI. Scientometrics, 116(2), 1181-1201.

[267] Ruiz, J., Sedrakyan, G., & Snoeck, M. (2015, September). Generating user interface from conceptual, presentation and user models with JMermaid in a learning approach. In Proceedings of the XVI International Conference on Human Computer Interaction (pp. 1-8).

[268] Mirnig, A. G., Meschtscherjakov, A., Wurhofer, D., Meneweger, T., & Tscheligi, M. (2015, April). A formal analysis of the ISO 9241-210 definition of user experience. In Proceedings of the 33rd annual

ACM conference extended abstracts on human factors in computing systems (pp. 437-450).

[269] Tukker, A., & Tischner, U. (2006). Product-services as a research field: past, present and future. Reflections from a decade of research. Journal of cleaner production, 14(17), 1552-1556.

[270] Tischner, U., Verkuijl, M., & Tukker, A. (2002). First Draft PSS Review. SusProNet Report. Netherlands, TNOSTB Delft.

[271] Shehab, E., Tiwari, A., Vasantha, G. A., Lightfoot, H., & Roy, R. (2011). Industrial Product-Service Systems (IPS2): Think Tank. Decision Engineering Report Series, Cranfield University.

[272] Uei, K., Nemoto, Y., & Shimomura, Y. (2013, October). EDIPS: Effective and Enjoyable Product–Service System Design Education Through Active Thinking. In International Conference on Serviceology (pp. 107-116). Springer, Tokyo.

[273] McAloone, T. C., & Andreasen, M. M. (2004). Design for utility, sustainability and societal virtues: developing product service systems. In DS 32: Proceedings of DESIGN 2004, the 8th International Design Conference, Dubrovnik, Croatia (pp. 1545-1552).

[274] Ikejiri, R., Fujimoto, T., Tsubakimoto, M., & Yamauchi, Y. (2012). Designing and Evaluating a Card Game to Support High School Students in Applying Their Knowledge of Wo rld History to Solve Modern Political Issues.

[275] Gagne, R. M., Wager, W. W., Golas, K. C., Keller, J. M., & Russell, J. D. (2005). Principles of instructional design.

[276] Forlizzi, J., & Battarbee, K. (2004, August). Understanding experience in interactive systems. In Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques (pp. 261-268).

[277] Law, E. L. C., Roto, V., Hassenzahl, M., Vermeeren, A. P., & Kort, J. (2009, April). Understanding, scoping and defining user

experience: a survey approach. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 719-728).

[278] Gaver, B., & Martin, H. (2000, April). Alternatives: exploring information appliances through conceptual design proposals. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 209-216).

[279] Shackel, B., & Richardson, S. J. (Eds.). (1991). Human Factors for Informatics Usability. Cambridge University Press.

[280] Dillon, A., & Watson, C. (1996). User analysis in HCI: the historical lesson from individual differences research. International Journal of Human-Computer Interaction, 45(6), 619-637.

[281] Stevens, G. C. (1983). User-friendly computer systems?: a critical examination of the concept. Behaviour & Information Technology, 2(1), 3–16.

[282] Shackel, B., & Richardson, S. J. (Eds.). (1991). Human Factors for Informatics Usability. Cambridge University Press.

[283] Nielsen, J. (1993). Usability Engineering. Academic Press.

[284] Hornbæk, K., & Law, E. L. C. (2007). Meta-analysis of correlations among usability measures. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 617– 626). ACM.

[285] Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. International Journalof Human-Computer Studies, 64(2), 79–102.

[286] Nielsen, J. (1994). Usability inspection methods. In Conference companion on Human factors in computing systems (pp. 413–414). ACM.

[287] Bailey, R. W. (1993). Performance vs. Preference. In Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting (pp. 282–286).

[288] Kissel, G. V. (1995). The effect of computer experience on subjective and objective software usability measures. In Conference companion on Human factors in computing systems (pp. 284–285). ACM.

[289] Abran, A., Khelifi, A., & Suryn, W. (2003). Usability Meanings and Interpretations in ISO Standards. Software Quality Journal,, 11, 325–338.

[290] Lewis, C., Polson, P. G., Wharton, C., & Rieman, J. (1990). Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 235–242). ACM.

[291] Card, S. K., Newell, A., & Moran, T. P. (1983). The Psychology ofHuman-Computer Interaction. Erlbaum Associates, Hillsdale, NJ.

[292] Agarwal, R., & Venkatesh, V. (2002). Assessing a Firm's Web Presence : A Heuristic Evaluation Procedure for the Measurementof Usability. Information Systems Research,, 13(2), 168–186.

[293] Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 249–256). ACM.

[294] Sutcliffe, A. (2002). Assessing the reliability of heuristic evaluation for website attractiveness and usability. In Proceedings of the 35th Hawaii International Conference on System Sciences (pp. 1838–1847).

[295] Hassenzahl, M., & Tractinsky, N. (2006). User experience - a research agenda. Behaviour & Information Technology, 25(2), 91-97.

[296] ISO (1998). 9241-11. In Ergonomic requirements for office work with visual display materials (VDTs) - Part 11, Guidance on Usability. Geneva Switzerland: International Standards Organizations.

[297] ISO (2001). 9126 DIS. In Software Product Evaluations-Quality characteristics and guidelines for their use.

[298] Miller, R. B. (1971). Human Ease of Use Criteria and Their Tradeoffs. In IBM, Systems Development Division, Poughkeepsie Lab.

[299] Eason, K. D. (1984). Towards the experimental study of usability. Behaviour & Information Technology, 3(2), 133–143.

[300] ISO (2019). 9241-210. In Ergonomics of human-system interaction. Part 210: Human-centred design for interactive systems. The international organization for standardization.

[301] Arhippainen, L. (2003). Capturing user experience for product design. The 26th, 1-10.

[302] Bargas-Avila, J. A., & Hornbæk, K. (2011, May). Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 2689-2698).

[303] Stephanidis, C. (2009). The universal access handbook. CRC Press.

[304] Vermeeren, A. P., Law, E. L. C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010, October). User experience evaluation methods: current state and development needs. In Proceedings of the 6th Nordic conference on human-computer interaction: Extending boundaries (pp. 521-530).

[305] Angeli, A. D., Hartmann, J., & Sutcliffe, A. (2009, August). The effect of brand on the evaluation of websites. In IFIP Conference on Human-Computer Interaction (pp. 638-651). Springer, Berlin, Heidelberg.

[306] Hartmann, J., Sutcliffe, A., & Angeli, A. D. (2008). Towards a theory of user judgment of aesthetics and user interface quality. ACM Transactions on Computer-Human Interaction (TOCHI), 15(4), 1-30.

[307] Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. Human-Computer Interaction, 19(4), 319–349.

[308] Tuch, A. N., Roth, S. P., Hornbæk, K., Opwis, K., & Bargas-Avila, J. a. (2012). Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI. Computers in Human Behavior, 28(5), 1596–1607.

[309] Dresch, A., Lacerda, D. P., & Antunes, J. A. V. (2015). Design science research. In Design science research (pp. 67-102). Springer, Cham.

[310] Hevner, A., & Chatterjee, S. (2010). Design science research in information systems. In Design research in information systems (pp. 9-22). Springer, Boston, MA.

[311] Venable, J., & Baskerville, R. (2012). Eating our own cooking: Toward a more rigorous design science of research methods. Electronic Journal of Business Research Methods, 10(2), pp141-153.

[312] Hevner, A. R. (2007). A three cycle view of design science research. Scandinavian journal of information systems, 19(2), 4.

[313] Endsley, M. R., Bolté, B., & Jones, D. G. (2003). Designing for situation awareness: An approach to user-centered design. CRC press.

[314] Endsley, M. R. (1988, October). Design and evaluation for situation awareness enhancement. In Proceedings of the Human Factors Society annual meeting (Vol. 32, No. 2, pp. 97-101). Sage CA: Los Angeles, CA: Sage Publications.

[315] Endsley, M. R. (1993). A survey of situation awareness requirements in air-to-air combat fighters. The International Journal of Aviation Psychology, 3(2), 157-168.

[316] MISQ, Special issue on design science research, MIS Quarterly 32 (4), pp. 725–868, 2008.

[317] Jalal, M. A., Chen, R., Moore, R. K., & Mihaylova, L. (2018, July). American sign language posture understanding with deep neural networks. In 2018 21st International Conference on Information Fusion (FUSION) (pp. 573-579). IEEE.