Proposal of a Novel Multimodal Interaction Model for Users of Different Age Groups

By

Diana Carneiro Machado de Carvalho

Supervisor: Prof. Doutor Maximino Esteves Correia Bessa
Co-supervisor: Prof. Doutor Luís Gonzaga Mendes Magalhães &
Prof. Doutor Eurico Manuel Elias de Morais Carrapatoso

Thesis submitted to
UNIVERSITY OF TRÁS-OS-MONTES E ALTO DOURO
in partial fulfillment of the requirements
for the degree of DOCTOR PHILOSOPHY
in Computer Science, accordingly with the provisions of
DR – I A–series, n.º 74/2006, March 24th and accordingly with the
Regulation of the Postgraduate Studies of UTAD
DR, 2nd series – Deliberation n.º 2391/2007

Proposal of a Novel Multimodal Interaction Model for Users of Different Age Groups

By

Diana Carneiro Machado de Carvalho

Supervisor: Prof. Doutor Maximino Esteves Correia Bessa

Co-supervisor: Prof. Doutor Luís Gonzaga Mendes Magalhães &

Prof. Doutor Eurico Manuel Elias de Morais Carrapatoso

Thesis submitted to

UNIVERSITY OF TRÁS-OS-MONTES E ALTO DOURO

in partial fulfillment of the requirements

for the degree of DOCTOR PHILOSOPHY

in Computer Science, accordingly with the provisions of

DR – I A–series, n.º 74/2006, March 24th and accordingly with the

Regulation of the Postgraduate Studies of UTAD

DR, 2nd series – Deliberation n.º 2391/2007

Supervisors:

Prof. Doutor Maximino Esteves Correia Bessa

Assistant Professor
Engineering Department, Science and Technology School
University of Trás-os-Montes e Alto Douro

Prof. Doutor Luís Gonzaga Mendes Magalhães & Prof. Doutor Eurico Manuel Elias de Morais Carrapatoso

Assistant Professor with Habilitation / Assistant Professor
Department of Information Systems / Department of Electrical and Computer Engineering
University of Minho / University of Porto

$E = mc^2$

Albert Einstein

*To my **parents**, the best one could ever hope for.*

*To my husband, my soulmate and my best friend **João**.*

Proposal of a Novel Multimodal Interaction Model for Users of Different Age Groups

Diana Carneiro Machado de Carvalho

Submitted to the University of Trás-os-Montes e Alto Douro
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Abstract – In recent years there has been a boom of different natural interaction paradigms, such as touch or gesture-based interfaces, that make better use of human's innate skills rather than imposing new learning processes. However, no work has been reported that systematically evaluates how these interfaces influence users' performance with regard to their age, and when performing different tasks, such as data selection, insertion or manipulation; that is, which interface or combination of interfaces could be the best fit for each user profile and each task.

In this context, this research aims to understand: (1) if there are differences in performance or preference of three age groups (children, young adults and older-adults) during interaction using various input modalities in different types of tasks; (2) if there are synergies and gains that come from the combination of different interaction paradigms, recognizing at which point the addition of input modalities start to saturate and harm the users' performance.

We evaluate four interaction paradigms (computer mouse/keyboard, touch, tangibles and gestures) for executing elemental and compound tasks unimodally and multimodally. Results showed that the creation of distinctive user profiles regarding interaction with different interfaces could, indeed, be important. Thus, we propose an interaction model with guidelines for the three age groups regarding the input modalities that may be more suitable for the different user profiles, and also taking into account the specificities of each group.

Keywords: Age Groups, User Profiles, User Performance, Human-Computer Interaction, Natural User Interfaces, Multimodal User Interfaces, Multimodal Synergies, Interaction Paradigms, Input Modalities, Mouse/Keyboard, Touch, Tangibles, Gestures.

Resumo – Recentemente, temos assistido a um *boom* de novos paradigmas de interação natural, tais como o toque ou *interfaces* baseadas em gestos, que tiram melhor partido das capacidades inatas do Homem em vez de impôr novas aprendizagens. Contudo, ainda não foi documentado nenhum estudo que avalie sistematicamente a forma como estas novas *interfaces* influenciam o desempenho do utilizador no que concerne à sua idade, e aquando da execução de diferentes tarefas, tais como seleção, inserção ou manipulação; ou seja, qual a *interface* ou combinações de *interfaces* que poderiam ser mais adequadas para cada perfil de utilizador e cada tarefa.

Neste contexto, este estudo pretende compreender: (1) se existem diferenças no desempenho ou preferência de três grupos etários (crianças, jovens adultos e adultos mais velhos) durante a interação com várias modalidades de *input* em diferentes tipos de tarefas; (2) se existem sinergias e ganhos que poderão advir da combinação de diferentes paradigmas de interação, reconhecendo ainda até que ponto a constante adição de novas modalidades de *input* começam a saturar ou prejudicar o desempenho dos utilizadores.

Avaliamos quatro paradigmas de interação (rato/teclado, toque, peças tangíveis e gestos) para executar tarefas elementares e compostas for forma unimodal ou multimodal. Os resultados mostraram-nos que a criação de distintos perfis de utilizador referentes à interação com várias *interfaces* poderá ser, de facto, importante. Assim sendo, propomos um modelo de interação com diretrizes para os três grupos etários relativo às modalidades de *input* que poderão ser mais adequadas aos perfis do utilizador, tendo sempre em atenção as especificidades de cada grupo.

Palavras-chave: Grupos de Idades, Perfis de Utilizador, Desempenho do Utilizador, Interação Pessoa-Computador, Interfaces de Utilizador Naturais, Interfaces de Utilizador Multimodais, Sinergias Multimodais, Paradigma de Interação, Modalidades de Input, Rato/Teclado, Toque, Tangíveis, Gestos.

Acknowledgments

First and foremost, I would like to thank all those that directly or indirectly participated in this investigation and supported me in this phase of my life.

I would like to express my gratitude to my supervisors, Professor Maximino Bessa, Professor Luís Magalhães and Professor Eurico Carrapatoso, for their trust, expertise, guidance and encouragement during my PhD study.

I would also like to thank the Magnificent Rector of the University of Trás-os-Montes e Alto Douro, Professor António Augusto Fontaínhas Fernandes, the President of the Engineering Department, José Boaventura Cunha, the Manager of the Centre for Information Systems and Computer Graphics of INESC TEC António Gaspar, and to the President of the INESC-TEC José Mendonça for the means provided to me during the course of this work.

I would like to acknowledge the National Foundation for Science and Technology - FCT (Fundação para a Ciência e a Tecnologia) for supporting this PhD through the grant SFRH/BD/81541/2011. I would also like to acknowledge the European Union (COMPETE, QREN and FEDER) for the partial support through the project

REC I/EEI-SII/0360/2012 entitled “MASSIVE - Multimodal Acknowledgeable multiSenSorial Immersive Virtual Enviroments”.

A very special thanks must go to all the people that have participated in the conducted experiments and all the people who enabled me to complete this work. Without them this PhD was not possible. Hence, I would like to acknowledge the support and contribution of the schools that took part in this study: “Colégio Moderno de São José”, “Monsenhor Jerónimo do Amaral”, “Escola Secundária Morgado de Mateus”, and the studies center “Super-Heróis”, all in Vila Real, Portugal.

To my PhD colleagues, that would always point me in the right direction,

I would like to remember and also treasure those that are no longer with me, but continue by my side and give me strength to face life’s obstacles.

I would like to express my deep gratitude to my family, especially my parents and husband, who have always supported my decisions and have always shown their unconditional love. I can only say THANK YOU for your constant support, friendship, attention and patience during the most difficult times.

UTAD, Vila Real Diana Carvalho
July, 2016

General Index

Table Index

Figure Index

Glossary, acronyms and abbreviations

Glossary

Human-computer interaction: HCI research is concerned with the design, implementation and evaluation of interfaces on different contexts regarding the users' task and work at hand, allowing an easy and efficient use of the interactive systems.

Accessibility: The term signifies "the right of all people to obtain equitable access to, and maintain effective interaction with, a community-wide pool of information resources and artifacts" (Stephanidis e Salvendy, 1998).

Usability: The term signifies "the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments" (ISO, 2015b).

Design for all: The term represents an effort to build accessible and usable systems by anyone, anytime, anywhere, including able-bodied and disabled people, people of all ages, with different skills and levels of expertise, with different languages, cultures, education (Stephanidis, 2001).

Input device: A hardware computer peripheral through which the user interacts with the computer (Hinckley et al., 2004).

Direct input device: A device the user operates directly: on the screen or other display, such as a touch screen (see also indirect input device). Occlusion is one major challenge of these devices (see also occlusion effect).

Indirect input device: A device the user operates by moving a third-party peripheral that is located away from the screen or other display, such as a mouse or trackball (see also direct input device).

Discrete input device: A device that provides an "on/off" state, such as buttons or keyboards.

Continuous input device: A device that is manually operated, as the computer mouse, and is continuously sending a signal to the system.

Occlusion effect: Issue that occurs when the finger or device blocks the area at which the user is pointing, hiding the contents displayed (e.g. when using touch screens).

Acquisition time: The term represents the average time to move one's hand to a device, or rather, the time the user takes to pick up an input device.

Homing time: The term represents the average time the user takes to put down an input device, or rather, the time to return from a device to a "home" position: e.g. return from keyboard to mouse (Sears e Jacko, 2007).

Footprint: The physical movement space (area) required to operate an input device (Hinckley et al., 2004).

Cognitive load: Measure of the human working memory used while performing a task, that is, total amount of mental effort being used in the working memory. This capacity limits how many things a user can do at the same time. The term was first coined by John Sweller in the late 1980s (Sweller, 1988).

Interaction task: A low-level primitive input command sent to the system by the user, such as entering a text string. The entry of each symbol by the user is an interaction task, performed using an interaction technique (see also interaction technique). Each task can be implemented by many different techniques (Foley et al., 1984).

Elemental task: Basic interaction tasks, such as content selection, insertion or manipulation.

Compound task: Hierarchy of elemental sub-tasks performed together. For example, the navigate/select compound task consists of scrolling to view an item in a list, and then clicking on it to select it (Hinckley et al., 2004).

Interaction technique: The fusion of input and output, of and to the user, consisting of all hardware and software elements, that provides a particular way for the user to accomplish a low-level task with a physical input device (Hinckley et al., 2004).

Interaction modality/mode: Modality is the term that refers to a single independent channel of sensory input/output between a computer and a human (Karray et al., 2008). The term modality is also a synonym for mode.

Multimodal interface: Sensorial interface that adjusts to other interaction paradigms and allows a more intuitive use of the systems multimodally, taking into account several input/output mechanisms in order to create synergies amongst them (Hale e Stanney, 2004; Oviatt, 2007).

Acronyms list

Acronym	**Expansion**
HCI	Human-computer Interaction
NUI	Natural User Interface
MUI	Multimodal User Interface
WIMP	Windows, Icons, Menus and Pointing device
POST-WIMP	Post era of Windows, Icons, Menus and Pointing device
CHI	ACM Conference on Human Factors in Computing Systems
ICT	Information and Communication Technology

Abbreviations List

Abbreviation	**Meaning(s)**
e.g.	from latin *exempli gratia*; for example
et al.	from latin *et alii*; and others (co-authors)
i.e.	from latin *id est*; that is

1 Introduction

For many years, the traditional mode of interaction with computers was based on a WIMP interface (Windows, Icons, Menus, Pointing device), which allowed users to interact with the machine via specific pointing devices, usually a computer mouse and keyboard.

In that era, the traditional attitude towards user interface development and engineering had different perspectives. Figure 1.1 represents particular opinions of different human factors specialists and their regards about the responsibilities of interface design in the system development.

Curtis e Hefley (1994) described these quotes as being articulated in a changing age, where more and more importance would have to be given to interface design and how this discipline would have to embrace new demands from the public. Indeed, computer-based systems would have to undergo a change in their interaction paradigm. And so they have.

With the paradigm shift to Post-WIMP interfaces (van Dam, 1997), we have turned towards a user-oriented and task-oriented approach that attempts to simplify the usability of the interface (Nielsen, 1993), giving preference to the users' innate skills (Mayer e Moreno, 2003) and allowing them to take advantage of recognition-based

***Executive**—I'll worry about the user interface when someone can demonstrate that it makes a difference in my sales*
***Project manager**—Yes, I'm sure you would love to do some field testing with users, but there's no slack in the schedule or budget*
***System designer**—That's trivial, let them handle it in the user interface*
***Software engineer**—When I'm done they have somebody who comes around and makes the screen look pretty*
***Interface engineer**—Isn't it exciting, I get to design the user interface all by myself*
***Customer**—We require that any software we buy have a GUI, you know, a Generic User Interface*

Figure 1.1 – Quotes of different human factors specialists about the importance of an interface in the WIMP paradigm era (Curtis e Hefley, 1994).

technologies that interpret complex human behaviours, such as speech, eye gaze, body language or gestures.

These interfaces are considered natural as they require low cognitive loads from the users and prioritize their innate skills (Mayer e Moreno, 2003). According to Buxton (2010), the only thing truly natural for people is what they are born with and is instinctive, such as the act of breathing, walking, eating: innate attributes that do not depend on education. Later in life, there is another type of natural activity: the ability to learn and improve skills and knowledge. This situation may be triggered by innumerable factors, from curiosity to obligation.

Indeed, knowledge is expensive and difficult to acquire. Tasks that seem simple and natural, like tying laces, are not: if we look closely we can see that this act is not innate, but learned and practised. Technology should not make users learn other skills when they can take advantage of abilities already learned throughout life and not waste previously learned talents (Buxton, 2010). Accordingly, natural interfaces are an opportunity to harness skills acquired during a lifetime full of motor, tactile, cognitive or even social experiences, instead of forcing new learning processes (Blake, 2012). These interfaces should prioritize the basic communication modes users learn since birth (like touch, gestures, speech).

Natural interaction paradigms are, hence, an attempt to eliminate previous interferences, as third-party devices (e.g. the computer mouse), when interacting with digital systems. Moreover, new natural interfaces may also be multimodal, supporting several modes of interaction and privileging a communication with natural languages (e.g. 2D and 3D gestures recognition) and multiple users (e.g. multi-touch surfaces), as well as encouraging an ubiquitous computing that manages to connect real objects with the virtual world (e.g. tangible interfaces). The combination of these input techniques in a multimodal interface shows a lot of potential concerning the user experience with computers where technology is invisible, as this induces a multimodal interaction more similar to the one users have with the real world (Jain et al., 2011; Antonella De Angeli e Petrelli, 1998). A natural experience envisages a combination of certain types of input and respective output and, as such, this relationship is considered a mode of interaction that may include gestures, speech, tactile recognition, proximity awareness or objects positioning (Jain et al., 2011). These multimodal interfaces apply to a vast variety of purposes, ranging from virtual reality systems, to the fields of medicine, education, culture and collaborative work.

With the development and evolution of multimodal interfaces' research, it has become more important to comprehend the user's context and behaviors. Interfaces that perceive one or more modes of interaction become more flexible to what types of input they provide regarding the user's preferences, abilities or characteristics, such as age, skills, impairments or health conditions (Oviatt e Cohen, 2000). In addition, these interfaces are more aware of the user's spatial and geographical contexts (Oviatt, 1997), becoming more robust and efficient when dealing with visual-spatial information.

However, there seems to be a constant widening of new modes of interaction without the proper awareness as to which could be the most adequate for different user profiles (e.g., children, elderly users, people with different levels of digital literacy, people with disabilities) and also regarding distinct types of tasks (e.g. elemental or compound). Each task carried out can be influenced by the supported sense of naturalism, efficiency and degrees of freedom (DOF) (Bowman et al., 2001), and even though users prefer multimodal interaction, this is no guarantee they will issue every command to the system multimodally (Oviatt, 2003). So for different tasks there are different interfaces that can better assist the user.

Nevertheless, to our knowledge, a systematic study has not yet been made in order to understand this relation and throw some light as to which interfaces are considered best or worst for specific tasks.

It is important to understand these distinct interfaces individually, as well as their characteristics, benefits and limitations, in order to grasp the advantages they may bring to the community in virtue of a better human-computer interaction.

1.1 Motivation

The aim of HCI is to enable the universal access of interactive systems for all users regardless of the diversity of the group, their tasks, or their context of use (Stephanidis, 2001). And it has been recognized the importance of adapting the input method and presentation format according to different devices, as it can affect the user's perception and attitude when dealing with specific contexts and tasks (Jonsson et al., 2004; Sodnik et al., 2008). Thereby, it could be vital to grasp which interaction tasks may benefit from a specific interface to the detriment of others, or even from a combination of certain ones. For each of these user interface transactions it is essential to be aware of how the various modes of interaction can enhance or, quite on the contrary, impair the user's performance taking into account their profiles, the context, use and convenience of the interaction techniques (Wigdor et al., 2009).

State-of-the-art multimodal interfaces recognize up to two interaction modes within different fields of interest, whether through the combination of pen devices and speech (Piper et al., 2011; Watanabe et al., 2007), speech and gestures (Miki et al., 2014; Liu e Kavakli, 2010), or multi-touch and tangible interfaces (El-Glaly et al., 2013; Riedenklau et al., 2012; Lucchi et al., 2010). As Oviatt e Cohen (2000) allude, multimodal interfaces tend to become more flexible as to how they apprehend the user's commands and what forms of input they permit according to the user's abilities and/or preferences, for instance: age, skill levels, cognitive styles, sensory and motor impairments, native language, or temporary illness.

However, with the continuous addition of new input modalities in our daily lives there is no full conscience as to what paradigms can help or actually impair the user. That is to say, there is not much awareness of which interaction paradigms are the most adequate for different user profiles: children, people with different levels of digital literacy, elderly users, etc.

Multimodal systems incorporate different input modes that can either create complementarity or redundancy in one same interface (Oviatt, 2007). In this sense, it is our intention to understand this dynamic regarding human-computer interaction and understand which interface(s) could improve users' performance when considering different age brackets.

This work is particularly motivated by the absence of scientific data that shows a systematical evaluation on how age, skill levels, or sensory impairments could influence the time it takes for a user to understand how to use the multimodal interface and issue a command, and hence interact with the system.

Published studies have not yet provided an understanding of how different user groups perceive distinct tasks and if their performance is directly influenced by the interaction modality. Little is known about how the different interfaces affect one's performance when it comes to age-related issues. There are no transversal comparisons of different age groups in one same study where more than one natural interface is evaluated. Work has been developed in this area, but not as a systematic approach. Sambrooks e Wilkinson (2013) compare gestural, touch and mouse interaction with 15 participants aged between 22 and 34 years old. They reached the conclusion that touch and mouse presented better results, but this interaction performance was not compared between other groups of users and thus it just clarified that the gestural performance was indeed worse than the other interfaces regarding that specific niche of participants. Other studies (Findlater et al., 2013; Bobeth et al., 2014; Inai, 2013), compared the interaction performance when using traditional mouse inputs or touchscreens, but not between other natural recognition-based interfaces as a hole, and they only compare at the most two groups.

Indeed, studies are still immature when it comes to stating differences in terms of efficiency, effectiveness, and performance between users with distinct attributes, namely age. Also, there is no indication as to what unimodal interface could provide better interaction responses from users for which tasks, and also what multimodal interfaces could assist the interaction. Even when it comes to a multimodal interface, evidences lack on how many inputs (simultaneous or not) could improve the users' performance and under which combination (e.g. touch and gestures; touch and tangibles; and so on). In this sense, guidelines or interaction metaphors that could support the choice of specific multimodal interfaces for distinct user groups are still underdeveloped.

It is conceivable to believe that children or older-adults interact differently with technology, whether regarding task completion times or precision. Hence, our drive is to understand if this difference in performance is indeed real, and if it can be quantified for unimodal and multimodal interfaces. Also, scientific research lacks a systematic exploration around the level of efficiency and synergy that may result from the combination of two, three or even four inputs in one same interface.

In this sense, we believe it could be important to appreciate if different groups of users apprehend interaction similarly and even quantify a positive or negative synergy that comes with combining several interfaces.

We are encouraged to realize or even recommend if there could be an interaction model for users of three different age groups (children, young adults and older-adults) about distinctive tasks.

1.2 Objectives

This work aims at investigating if there are age-related differences in interaction regarding elemental and compound tasks, as well as understanding if any synergies emerge from combining different interaction paradigms or, on the other hand, if multimodal interfaces saturate the users and hinder their performance.

Our ultimate intention is to propose guidelines for users of three age groups regarding distinctive elemental and compound tasks in a specific context, whether using a unimodal interface or a multimodal one.

Therefore, we ascertained specific goals for our research:

- Understand how three different age groups relate to different input modalities (mouse/keyboard, touch, tangibles, gestures) when it comes to user preference and performance in elemental tasks;
- Understand if there are age-related differences regarding compound tasks when using distinct input modes of interaction, and which are better suited for which tasks in a specific context;
- Understand if there are any synergies when combining different modes of interaction in a multimodal interface and, if so, which ones and regarding which age group;
- Propose an interaction model with guidelines for each age group regarding the input modalities that may be more suitable for the different user profiles.

1.3 Thesis contributions

Aside from the objectives stipulated for this research, we also make other contributions to the field, namely:

- We present a literature review and relevant work accomplished in the area, where we discuss and explain particular foundations of the human-computer interaction field of expertise.
- We show different evaluations methods well-known and used in the field of human-computer interaction to analyze the results and enumerate their advantages, also explaining why they all have to be considered when assessing

human performance. Thus, we provide a thorough investigation as to what evaluation methods to use in the field when analyzing different natural and/or multimodal interfaces.

- We demonstrate how people with distinctive profiles interact with the systems differently: one interface may be good for a user in particular regarding a specific context, but that same interface may not be good for another person.

- We ascertain the participants' feedback when accomplishing compound tasks using more than one interface and retain certain conclusions that may help composing interaction metaphors and models in different fields of study.

- We propose an interaction model based on different user profiles that may provide an easier interaction experience and low learning curves when using natural and/or multimodal interfaces.

1.4 Outline of the thesis

This document is divided into 7 chapters and is structured as described next:

Chapter 1: The current chapter clarifies the theme at hand, as well as the motivation for this research, the intended goals and contribution of this work. Also, the methodology and research strategies taken into account for this study are highlighted and the target audience outlined is portrayed.

Chapter 2: Provides fundamental information on the field of human-computer interaction, beginning with the concept of user interfaces for all, and followed by a discussion on the current shift to natural and multimodal user interfaces, as well as relevant work carried out in the literature. Moreover, this chapter conveys some of the key characteristics and considerations of interaction design, thus informing the reader about interaction modalities and techniques, as well as active and passive interfaces, types of tasks and input devices. Interaction metaphors are also dealt with.

Chapter 3: Presents the methodological approach of the study and the various phases of the research. Also, this chapter thoroughly describes the evaluation methods taken into account and specifies the different strategies for data collection.

Chapter 4: Presents formal evaluations of the different elemental tasks with regard to distinct age groups, in order to throw some light on which interface may provide better user performance when it comes to selection, insertion and manipulation activities.

Chapter 5: Presents formal evaluations of compound tasks showing whether there are age-related differences in interaction when combining different interfaces and if there are synergies arising therefrom. Also, the case studies carried out attempt to corroborate previous conclusions and find characteristics that can be generalized into interaction guidelines for different age groups.

Chapter 6: Presents a proposal of a multimodal interaction model regarding different user profiles. Here, several guidelines are recommended accordingly for the three age groups studied when it comes to elemental and compound tasks in a controlled environment.

Chapter 7: This final chapter summarizes the results and contributions of this work, states the limitations of the study and also proposes areas for future work.

2 Background

The following chapter presents the concepts regarding human-computer interaction that are addressed throughout this thesis. This framing is important not only to understand the context of the theme, but is also relevant to justify the course of the study regarding the groundwork of the investigation, the methodology that was taken into account, and its contribution.

We thereby aim at (1) providing an overview regarding the field of human-computer interaction; (2) highlighting the most important characteristics of the concept "design for all"; (3) going through the theoretical and technological roots of interaction paradigms, being natural user interfaces briefly revised and special attention given to multimodal interfaces; (4) giving a prompt look to the concept of interaction design. Moreover, we summarize the extensive literary review in order to provide the reader with insights on the research field and a more straightforward explanation of the context.

2.1 Human-Computer Interaction

Human-computer interaction (HCI) is an area of research that began initially as a specialty area in the field of computer science, embracing cognitive science and human factors engineering (Soegaard e Dam, 2014).

Nowadays, HCI is a multidisciplinary field that has expanded rapidly, combining diverse concepts and approaches. That is why an exact definition about this field is complex and there is no agreed upon classification about the range of topics that form this area of research and practice. However, the Association for Computing Machinery (ACM) tried to define HCI stating: "Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them" (Hewett et al., 1992, pp. 5)

Indeed, HCI research is concerned with the design, evaluation and implementation of interfaces on different contexts regarding the users' task and work at hand, allowing an easy and efficient use of the interactive systems. The main goal of this field if to understand how interfaces can allow users to take advantage of high performance computing without retracting their focus from their work, rather than making them learn technological details necessary to handle the interfaces and complete the tasks (Marchionini e Sibert, 1991).

For this dynamic there are three major issues of concern in the field (Dix et al., 2004): (1) the user, considered as an individual or group of users working together for a specific work or for an organization, dealing with the tasks or processes in the interactive system; (2) the computer, meant as any sort of technological system, from the basic desktop computer to a large-scale system; (3) the interaction itself, being this factor represented by any type of communication, direct or not, between the user and the computer, either through direct manipulation of the data and respective feedback, or by involving intelligent sensors controlling the environment. On the other hand, Marchionini e Sibert (1991) further clarify this classification and recognize a broad set of topics in HCI that can be organized according to a similar assessment (Figure 2.1), only differing on the topic of interaction. Regarding their classification, interaction is influenced by the use for which the systems are prepared, the context of that use (I), and the process of design/development of the interface itself (IV). In fact, as more importance is being given to the user experience with the interface, for the interaction to be successful the system/product must be: (1) useful, as in it accomplishes what is required of it; (2) usable, as in it is easy to use, natural, and holds no danger for the user; and (3) used, as in it should make people want to use it and be attractive and engaging (Dix et al., 2004).

I. Use and Context Social Organization and Work Application Areas Human-Machine Fit and Adaptation	
II. Human Human Information Processing Language, Communication and Interaction Ergonomics	**III. Computer** Input and Output Devices Dialogue Types and Techniques Dialogue Genre Computer Graphics Dialogue System Architecture
IV. Development Design Techniques Implementation Techniques Evaluation Techniques	

Figure 2.1 – HCI topics (Marchionini e Sibert, 1991, pp. 20).

Thus, we can understand that this subject, being an interdisciplinary area, draws supporting knowledge on both the machine and the human side and embraces several disciplines with an emphasis on: computer science (engineering of human interfaces, techniques in computer graphics, operating systems, programming languages), psychology (cognitive processes and analysis of user behavior), sociology and anthropology (linguistics, communication theory, interactions between technology, work and organization), graphic and industrial design (designing interactive products), and human factors (ergonomics and user performance) (Dix et al., 2004; Hewett et al., 1992). Recently, in HCI, issues regarding motivation, enjoyment and a unified experience with the interface and task have also been increasingly noticed (Dix et al., 2004).

2.1.1 Designing for diversity

In recent years, interactive computer-based systems have become tools for communication, collaboration and social interaction amongst diverse user population with different abilities, skills, disorders, requirements and preferences in a variety of contexts of use (Stephanidis, 2001). As such, the needs of the users are becoming increasingly important and computers are considered integrated environments that should be accessible and usable by anyone, anytime, anywhere: "The aim of HCI is to ensure the safety, utility, effectiveness, efficiency, accessibility and usability (...)" of interactive systems by all (Stephanidis, 2001, pp. 3).

Indeed, a major challenge in the field of HCI is the concept of "user interfaces for all", firstly coined by Stephanidis *et al* in the late 90's (Stephanidis e Salvendy, 1998; Constantine Stephanidis e Pier Luigi Emiliani, 1999; Stephanidis et al., 1999). This approach entails three major fields of concern (Soegaard e Dam, 2014):

i **User-centered design:** being a design approach that focuses on an iterative process for the development of usable systems, involving direct user participation. Its main concern is to bring the ease of use concept into the total user experience with products and systems (Vredenburg et al., 2001). This subject will be discussed further below.

ii **Accessibility and assistive technologies:** being a concern increasingly implemented in the field to consider the access to the ICTs mainly by, but not restricted to, people with disabilities. Here, interaction with the interface may be affected in various ways by the user's individual abilities or functional limitations, whether they are permanent, temporary, situational or contextual (Bergman e Johnson, 1995; Cook e Polgar, 2015).

iii **Universal design:** being a term initially used in engineering disciplines, the objective is to design products and environments to be appealing and usable by all people, regardless of their age, ability, or status in life, without the need for adaptation (Mace et al., 1991).

The term "user interfaces for all" represents, indeed, an effort to overcome the aforementioned challenges and should be conceived as a new perspective on HCI that implies a reevaluation of conventional design principles such as accessibility and usability.

Accessibility "signifies the right of all citizens to obtain equitable access to, and maintain effective interaction with, a community-wide pool of information resources and artifacts" (Stephanidis e Salvendy, 1998, pp. 6). Although being a term initially associated with elderly people and individuals with disabilities or special needs, it currently has a more ample connotation, embracing all individuals with different levels of abilities, skills, requirements and preferences that should be able to access information technologies (Stephanidis et al., 1999).

On the other hand, usability refers to the capability of all supported methods towards task successful accomplishment to maximally fit distinctive users' needs and requirements in a particular context or situation of use (Soegaard e Dam, 2014). This principle is documented in the International Standards Organization (ISO) report, "ISO 9241-210:2010" (ISO, 2015a), which provides an approach for evaluating usable interfaces and suggests metrics for effectiveness, productivity, satisfaction and safety that can be used for this purpose. This evaluation method will be further discussed in Chapter 3.

Indeed, several efforts have been made to address the concern of designing for diversity. The two main contradictory approaches taken into account in this field of research are the reactive and proactive strategies (Stephanidis, 2001). The first, more directed to building accessibility features into interactive applications as a result of specific user requirements, could be very limited, especially when considering the constant changes in the technological systems and it could lead to the lack of important functionalities for the user. Contrarily, proactive strategies entail an effort to purposefully include features into a product early in the process (e.g. from its conception, to design and release) as a way to accommodate the broadest possible end-user population and minimize the need for *a posteriori* adaptations. This is the supporting motto of user-centered design.

Universal design rests on seven principles that were developed in 1997 by a working group of architects, product designers, engineers, and environmental design researchers, led by Ronald Mace at North Carolina State University. Their goal was to implement a strategy to design accessible solutions and define a certain code that "may be applied to evaluate existing designs, guide the design process and educate both designers and consumers about the characteristics of more usable products and environments" (NC State, 1997). These concepts were elaborated as a way to ensure access by anyone, anywhere and at any time to interactive products and services, to the greatest extent possible, without the need for adaptation or specialized design. The seven principles of universal design are not restricted to a usability standpoint, but also encourage a design process that incorporates other considerations such as economic, engineering, cultural, gender, and environmental concerns (NC State, 1997). These are:

1. **Equitable use:** The design should be useful and appealing to people with diverse abilities, and provide the same means of use by all - identical whenever possible; equivalent when not. Also, it should avoid excluding or stigmatizing groups of users and ensure privacy, security, and safety.

2. **Flexibility in use:** The design should accommodate a wide range of individual preferences and abilities, in order to facilitate the user's accuracy and precision: whether it is by allowing the user to choose the methods of use, providing adaptability to the user's own pace, or accommodating right- or left-handed access and use.

3. **Simple and intuitive use:** The design should eliminate unnecessary complexity and be easy to understand, regardless of the user's experience, knowledge, language skills, or current concentration level. Moreover, effective prompting and feedback should be provided during and after task completion, always being consistent with user expectations and awareness.

4. **Perceptible information:** The design should communicate the necessary information effectively to the user by using different modes (pictorial, verbal, tactile) for redundant presentation, regardless of ambient conditions or the user's sensory abilities, i.e. provide compatibility with a variety of techniques or devices that can also be used by people with sensory limitations. This can maximize "legibility" of essential information and make it easy to give instructions or directions.

5. **Tolerance for error:** The design should minimize hazards, in order to prevent accidental or unconscious actions, by providing warnings and fail safe features and arranging elements to reduce errors.

6. **Low physical effort:** The design should facilitate use and allow the user to maintain a neutral body position with low physical effort, encouraging an efficient and comfortable interaction with a minimum of fatigue and repetitive actions.

7. **Size and space for approach and use:** The design should accommodate adequate size and space for approaching, reaching, and manipulating information comfortably by any seated or standing user, regardless of his/her body size, posture, or mobility, allowing assistive devices or personal assistance.

Thus, the term universal design either incorporates, or is a synonym of, terminologies such as accessible design, inclusive design, or design for all, each highlighting different aspects of the concept (Stephanidis e Salvendy, 1998).

Clearly, several concerns are being attended to provide wider access to technology: from the challenging digital divide (OECD, 2001), the role of gender in HCI and the use of technology by elderly people and their acceptance of information technologies (Carvalho et al., 2012b,a); to the design process for and with children and people with cognitive/perceptual impairments or physical disabilities (Sears e Jacko, 2007). Some of these are briefly discussed next.

Knowledge is presently considered a fundamental driver to enhance global competitiveness and productivity, as well as innovation and wealth generation (Gomes, 2002). One must understand the different obstacles or handicaps of the technology in order to provide better solutions and close the gap between those who could keep up with the technological advances, and those that did not. Indeed, the gap is not only about the info-excluded users, but also about the reactions to the technology with which users interact, i.e. it is less about the total time using a computer, than about using a computer voluntarily for amusement and contentment. It is about their comfort, attitudes, and levels of anxiety using the digital systems (Cooper e Kugler, 2007). That is why it becomes important to acknowledge that different user groups interact with technology in different ways, and thereby various means of interaction are needed and desired.

Gender stereotypes are also an obstacle in HCI, as sometimes exists in our social world entrenched stereotypes of the behaviors and attitudes that are appropriate for children and adults of each gender. This situation is equally due to concerns about anxiety and fear of poorer performance on the male or female's part (Cooper e Kugler, 2007).

Linked with the technology explosion is also the aging of the worlds' population. There are a number of contexts where older people are likely to encounter computers and other forms of communication technologies. Nonetheless, despite a tendency toward increased computer use and other forms of technology among older people, it is still lower among elders as compared to younger people. The main reasons for this scenario are related to health and disabilities increased with age (Czaja e Lee, 2007). Indeed, health issues play an important part in the likelihood of elderly people using technology, both at home or at work, and thus have critical implications for the systems' design (Kaye, 2000). System designers need to understand the heterogeneity of the older adult population and ensure usability testing is done concerning their changes in motor skills, response speed, cognitive abilities and sensory/perceptual processes (Czaja e Lee, 2007). There should be hardware and software considerations in designing the system in order to accommodate older-adults. For example, attention should be paid to the placement and size of the screen, shape and labeling of controls, layout of instructions, and the modes of interaction to provide ease of use of the systems.

Likewise, attention should be given to interfaces designed for children. Even though these users tend to be creative, intelligent and capable of great things if they are given the tools and relevant support, they too have distinct physical and cognitive differences when compared to adults (Bruckman et al., 2007).

Jean Piaget, leading psychologist in understanding children's development and cognitive evolution (Piaget, 1970), divided children into development stages and uncovered that they lack knowledge and experience that are matured during growth. That being said, software for children needs to be designed differently and adapted from stage to stage of growth, according to dexterity, speech, reading capabilities, background knowledge and interaction style (Bruckman et al., 2007).

Firstly, children's fine motor control is different from that of the adults and they are physically smaller. As such, the technological devices themselves should be adapted or else they are difficult for children to use. Moreover, speech can also be an inconvenient when interacting with the systems, as their pronunciation is still being developed, and the interaction can become very frustrating with incorrect feedback (Nix et al., 1998). The same is valid for interaction that requires reading, as the system must be aware of children's level of reading skills.

Also, when it comes to interaction styles, the children's attention levels and patterns of interaction are not equal to those of adults. Even the visual feedback needs to be considered to seize the playful and spontaneous nature of the children, without discarding the importance of the informative/error notifications (Hanna et al., 1997). The interaction pattern itself used during interaction can produce different results (Inkpen, 2001). In a collaborative environment, children are more likely to cooperate with one another in a system with multiple input devices (like a collaborative game), each making use of an input modality to help each other. This dynamic between multiple users and devices increase their productivity and satisfaction (Stewart et al., 1998).

Computers are no longer considered mere business tools, but integrated environments, accessible anytime and anywhere by a diverse user population with different abilities, skills, requirements and preferences, regardless of the context and means of interaction. Technology is increasingly penetrating a wider range of human activities, even in non-traditional contexts, and is being used by people with different cultural, educational, training and employment backgrounds, novice or experienced users, very young and elderly, or people with different types of disabilities.

2.2 Paradigm shift to natural user interfaces

The human being has passed several technological revolutions over time, not continuously, in periods of cyclical evolution, but irregularly. A technological revolution is commonly portrayed as an era of crisis that gradually adjusts to a new pattern adopted by society. This moment of change was firstly coined by Thomas Kuhn as "paradigm shift" (Kuhn, 1962). Indeed, the human being has observed different paradigms shifts throughout times when it comes to technology.

According to Galitz (2007), the need for human beings to communicate was revealed early. The most basic and primeval means of communication were based on specific movements and gestures. In fact, these movements are independent of any idiom or lexicon, allowing interaction between people that do not speak the same language. The next level is speech itself, and the highest level of complexity in communication is written language.

During the development of interface patterns with the computer, the mode of interaction has been the opposite of the one that better fits an effective data transmission between entities. In a first instance, the mode of human-computer interaction followed the most complex level of communication: written language, that is, the command line interface, or CLI (:aa). Prior to these interface, the execution of a series of jobs was accomplished through a process of batch processing (:ab). Later, interaction was enhanced and evolved to a lower level of complexity to users: the recognition of icons as a way to convey information more intuitively - the graphical user interface (Nielsen, 1993).

Currently, society witnesses a new technological revolution regarding the interaction paradigm: natural interfaces. Figure 2.2 shows the several paradigm shifts that have been occurring throughout times regarding interaction. These allow for a direct interaction, taking advantage of human senses for a better human-computer communication. Instead of having a technology-centered context, we now turn to a human-centered one, where technology can “understand” the user and his/her context of use (Krippendorff, 1997).

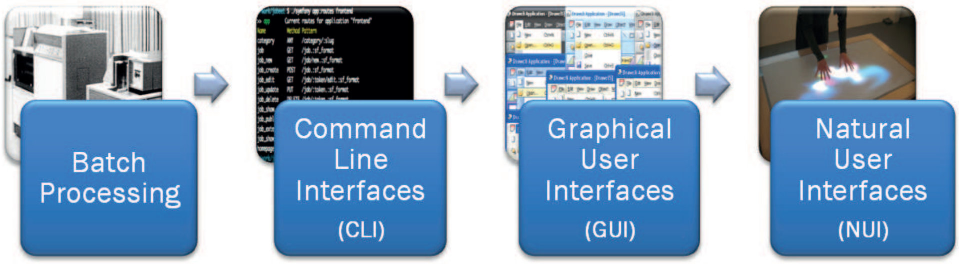


Figure 2.2 – Paradigm shift towards Natural User Interfaces.

Contrarily to previous technological generations, as command line interfaces and “point-and-click” WIMP (graphical user interfaces based on window, icon, menu, pointing device) interfaces, this paradigm does not have a pointing device as a demand, such as the computer mouse/keyboard. Instead, it is based on users’ innate behaviors and abilities, such as gestures or touch, for specification and execution of commands and tasks. This new era is also known as Post-WIMP, due to its status post-graphical interfaces, and focuses on the users’ needs and skills, facilitating the usability of the interface.

According to Buxton (Buxton, 2010), the only innate behavior is to eat, breathe, touch, hear. Knowledge is something expensive to acquire. That is why, when it comes to technology, people should not be obligated to learn the machine’s language and waste their own abilities, but it is technology who should learn the users’ language, allowing them to use their cognitive, motor, tactile or even social abilities. “The goal we strive for with today’s user interfaces is to minimize the mechanics of manipulation and the cognitive distance between an intent and the execution of that intent” (van Dam, 1997, pp. 64). Indeed, the user should focus on the task, not the technology, for specifying the task.

The main goal of a natural user interface is to remove other intermediary devices, like the computer mouse, to support information input (van Dam, 1997). In this regard, technology is focusing on the users and their needs and adopting a user-oriented and task-oriented approach (Nielsen, 1993). Presently, we witness a user interface paradigm that is contextual and enables the user to interact with it physically, with no need for previous acquaintance with the device or obligatory preparation (Towell, 2009). This concept is strongly supported by Blake (2011) that states that the interfaces will finally become more natural for users to explore their abilities acquired during a lifetime of socializing in the real world. Indeed, natural interfaces encourage the use of the users' own senses to communicate with the machine. The system should prioritize the most basic means of communication people learn since birth (speaking, gesturing, facial expression, and other forms of human communication), to the detriment of an interaction that requires third-party devices, unfamiliar to our innate skills, and forces new learning processes.

Indeed, some characteristics of this new generation of natural interfaces are the progresses made regarding a more direct and physical interaction between the human and the machine. Instead of a more textual and syntax-oriented approach, natural interfaces will embrace users' innate skills, allowing a greater physical control over the information. Data will be seen as a set of tangible information that can be directly manipulated instead of something ambiguous: the simultaneous control of input and output of information, the continuous real-time feedback of the machine, and the adoption of an increasing ubiquitous computing are known advantages (Wu, 2000).

The new interaction paradigm is not defined by the required input devices mouse/keyboard, but its mode of interaction (Blake, 2011). Any technological interface may be considered natural as long as it permits the users to take advantage of their innate skills, or the ones they have acquired by virtue of their own development as people in a social world. Hence the difficulty in defining a taxonomy that can distinguish the various modes of interaction.

Based upon the discussed innate skills that should be taken into account in interaction, guidelines have been proposed that help identify natural interfaces according to four concepts (Figure 2.3): instant expertise, progressive learning, direct interaction and cognitive load (Blake, 2012).

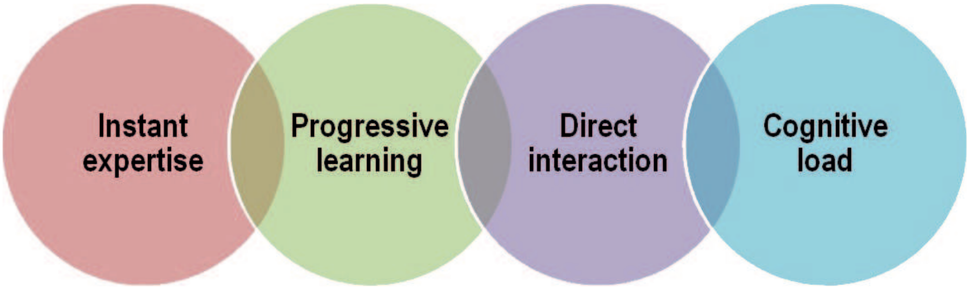


Figure 2.3 – Natural User Interfaces guidelines.

Instant expertise refers to the ability of reusing existing skills users already have, either by reusing domain-specific skills or reuse common human skills. If the interface requires the use of existing skills rather than learning new ones, users can get up to speed very quickly with little to no effort.

Also, a natural interface should enable the user to progressively learn and advance from novice to expert, providing a smooth learning path from basic tasks to more advanced activities. However, at the same time, the interface should not restrict the user or get in the way of expert users doing advanced tasks if they intend so at any given point in time.

The interface should provide a way to make sure interactions are fluid, direct and appropriate to the context, meaning they should allow users to access many features on demand without overwhelming him/her by presenting them all at once. In terms of direct interaction with content, different degrees are possible depending on the specific input modality (Figure 2.4).

When it comes to the required cognitive load, we have already mentioned that ideally the interface should minimize the effort required to the user, in order to benefit interaction and provide an ease of use interface, as well as becoming quick to learn. For example, using the computer mouse requires a higher cognitive load than using our own fingers.

Thereby, provided that the various paradigms can gather the same common aspects

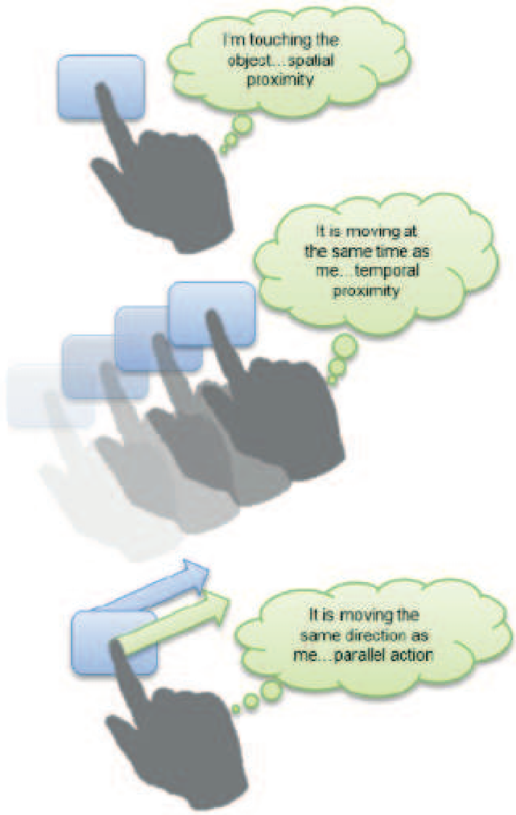


Figure 2.4 – Types of directness in natural interfaces (Blake, 2012).

to allow the users an interaction with a low cognitive load, using only their innate abilities and not forcing new apprenticeships (Mayer e Moreno, 2003), these can be considered modes of natural interaction.

On a broader perspective, input/output interfaces can be considered natural as long as they comprise gestures or body language, speech, proximity and location, eye gaze and expression, biometrics on the input side, and the full spectrum of audio and visual output, smell, tactile and object location, and other experiences on the output side, leveraging the full range human senses (Jain et al., 2011).

2.2.1 Speech-based interfaces

Many advances were carried out in the area of speech-based user interfaces, not only due to the progress in the field of linguistics itself, but also due to the technological progress regarding how the computer "understands" or, rather, seizes and analyses the human speech (Godwin-Jones, 2009). Indeed, several equipment have been emerging that support and encourage the use of such recognition technologies (Figure 2.5).

Figure 2.5 – Apple iPhone's SIRI application is an example of an interface that supports speech-recognition - Photograph taken by (Jonnalagadda, 2014).

Nowadays, the development of techniques that fall upon input/output of speech, that is to say, the introduction of speech commands and respective feedback in a system is growing (Myers et al, 2000). Such progress will shape the way interfaces are developed, emerging new models for the study of interaction, namely how data input is recognized and how it is processed, in order to send the corresponding output in real time and without long delays: "First, a conventional event-based model may no longer work, since recognition systems need to provide input continuously, rather than just discrete events (...). Similarly, in speech, the interpretation of the speech should begin immediately, and feedback should be provided as soon as possible" (Myers et al., 2000, pp. 19).

Indeed, for the new speech-based interface to have the intended success regarding the understanding of speech and human commands, it will be necessary to study and recognize users' linguistic patterns (Yankelovich e Lai, 1999). It is unlikely to expect users to adapt their linguistic abilities and speech patterns in order to bypass the systems' limitations. Instead, the systems will have to adjust to attend to a more robust recognition of the human voice through an improved capacity of error detection in speech (Oviatt et al., 1994). Human communication displays a tremendously inconstant and ambiguous nature: whether it is due to speech impairments, as broken sentences and reasoning, different accents and pauses in speech; or because of the convergence phenomenon that implies the influence of the users themselves amongst each other, where they unconsciously adopt other people's dialogue characteristics such as volatile choices of vocabulary, lexical and grammatical structures (Leiser, 1989). Also, human-human communication is more prone to off-topic comments and deviations of context (Godwin-Jones, 2009), and can be easily influenced by the users' own tone of voice, and even acoustics or ambient noise (Suhm et al., 1996). It is therefore important to recognize these limitations, emphasizing the relevance of predicting and correcting these errors upon the user's input, namely resorting to multimodal interfaces that support mutual disambiguation, a technique of creating redundancy in input in order to prevent and reduce errors during interaction (Oviatt, 2007).

Some of the strategies used to enhance error correction performance of such interfaces involve the development of input recognition algorithms that are more precise and efficient, taking advantage of multimodal interfaces that opt for the redundancy of data through the input of written commands (Suhm et al., 1996; Oviatt et al., 2008). On the other hand, Löhr e Brügge (2008) have demonstrated a system that resorts to a heuristic evaluation to infer the command when such is badly interpreted, and also incorporate the possibility to ask the user for additional clarifications in case of ambiguous situations, in order to avoid errors and grant the task a more informal nature of conversation. Also, Danis (1989) presented a study about the possibility to augment the efficiency of detecting speech commands through user's training, in order to "accustom" the system to the user's speech patterns. Simultaneously, this practice could even increment good speech habits by the repetition of the commands, thus eliminating erroneous patterns: e.g. rapid pace of the user's speech, lack of pauses between words, pronunciations overly strong or, on the contrary, incomplete regarding the first or either last phoneme of the syllable.

Suhm et al. (1996) summarize three of the most adopted strategies for increasing speech-based interfaces' productivity and efficacy: reduce the number of error by training the user according to a speech pattern better detected by the system; facilitate, through an interface's immediate feedback, error detection of textual interpretation; and also, involve the user in interactive dialogues with the system in order to avoid and prevent errors. In this study, the researchers determined that the most effective technique could be the error correction resorting to a multimodal interface based on speech and manual writing recognition, highlighting the importance of a richer interaction concerning different modes of data input.

However, even recognizing the limitations of systems controlled by speech, there are also several advantages of these interfaces regarding a natural interaction: they enable the input of information without the resort to a keyboard or even when a monitor is unavailable; they simplify tasks when the user's hands and/or eyes are occupied; they relieve the need for written commands for people with disabilities or motor impairments (Yankelovich e Lai, 1999). Indeed, some studies show that these systems may take advantage of speech recognition in order to replace quick commands given by the mouse or keyboard in graphical interfaces, with very positive outcomes (Gorniak e Roy, 2003; Nakano, 2008; Munteanu et al., 2013; Munteanu e Penn, 2015) although being this mode of interaction still at an early stage.

Regarding the field of education, other studies have also been performed, more specifically regarding language teaching, where students learning a second language other than their native one - or even people with disabilities trying to perfect interpersonal communication - resort to automatic speech-recognition systems aided by visual outputs (Godwin-Jones, 2009; Shrawankar e Thakare, 2010). Advances are being made regarding the use of such systems with low cost and low power equipment, in an attempt to embrace poorer nations that struggle to access information and technology (Nedevschi et al., 2005).

Undoubtedly, speech-based interfaces may have a big impact on man-machine interaction, being more context-aware and simplifying some tasks and allowing the computer to understand and infer which commands we intend to perform (Serafin, 2010). However, this mode of interaction is still not at its peak being necessary more studies to understand contexts where this interfaces could be more helpful and what models to implement to increase their efficiency and effectiveness (Aylett et al., 2014; Truschin et al., 2014; Ortiz, 2014; Schaffer et al., 2015), and level the system to comprehend users with low technological skills but also more advanced and experienced users.

2.2.2 Gesture-based interfaces

According to the different complexity levels of communication, the most basic mode of human interaction has always been through gestures (Galitz, 2007). "When people interact naturally with each other, it is common to see indications of acknowledgment, agreement, or disinterest given with a simple head gesture" (Morency e Darrell, 2006, pp. 32). These movements are natural gestures performed by humans during a face-to-face conversation, and they can be incorporated and interpreted by the computer as interaction facilitators when using gestural user interfaces.

Indeed, a lot of the information transmitted amongst humans is passed on through gestures. They are a non-verbal mode of interaction between people, and can be achieved by simple hand movements, such as pointing, dragging or moving objects, or by some more complex actions that are able to demonstrate one's feelings. Indeed, instead of relying on third-party devices, as the mouse, to interact with the systems, this new interaction paradigm intends to incorporate those every-day gestures into the own technology, providing the user with the choice of not using any other external devices and instead using his/her own body.

There are gestures that we can all identify with ease, intuitively, such as for a selection (click or touch): the action of touching an object to select it. Although similar at first glance, these two gestures are not the same (Figure 2.6): clicking is something learned throughout the years with the experience acquired by using a computer and, as such, is not a natural innate behavior for the user. As for the touch, this is a human ability with a low cognitive load. Despite the result being the same - the selection of an element on the screen - the action performed is not: clicking is an action learned, compound, whilst touch is part of the human senses (Blake, 2011). Other examples of gestures already being incorporated in interactive systems are dragging objects, rotating or resizing them using just our fingers. These types of bi-dimensional gestures are mostly related to multi-touch interfaces that allow several simultaneous touch points (this subject will be explored ahead in this section).

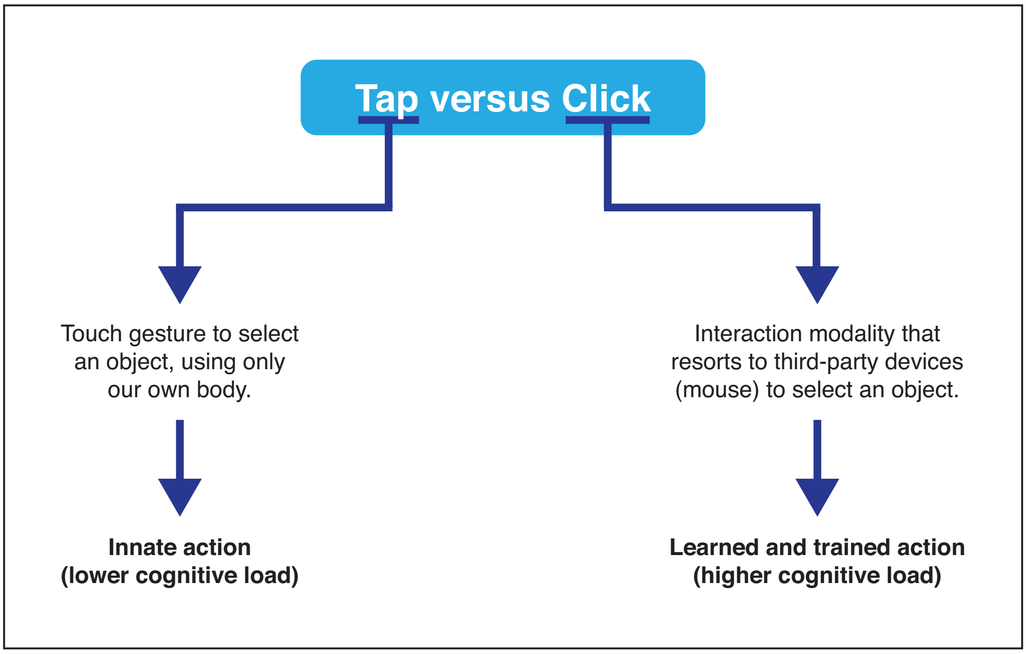


Figure 2.6 – Difference in cognitive loads regarding two gestures (tap and click).

Moreover, other than taking advantage of finger-based gestures, these interfaces may also recognize other parts of the body, as arms, legs or hands. The Microsoft Xbox

Kinect is an example of a commercially available input sensor that enables gesture-based interaction (Figure 2.7). This device resorts to depth sensing and computer vision to interpret human gestures as commands. Indeed, different attempts have been made to embrace even more gestures, rather than just hand movements (Figure 2.8).

Figure 2.7 – Microsoft Kinect: an example of a gesture recognition-based sensor.

Kjeldsen e Kender (1996) studied how a simple hand gesture could have the potential to be more efficient and intuitive than using the mouse for some tasks, even though it is necessary to calibrate the interface in this regard. In their studies, some users actually preferred resorting to gestures to point and manipulate virtual objects. On the other hand, they came to the conclusion that the learning curve for using gestures, how to perform them and which ones to choose, was harder than adjusting to the mouse: there were no auxiliary visual elements during the interaction to help the user remember the position sequence and the right movements to perform each task. This situation could, however, be avoided by providing the user with brief explanations and visual "on-screen" aids during an extended pause by the user (Kjeldsen e Kender, 1996).

It is also possible to benefit more from this mode of interaction when it allows the recognition of, for instance, the user's head gestures, such as a simple nod for approval or disagreement, in order not to compromise interaction with the main task at hand. Currently, the most frequent interruptions during interaction with

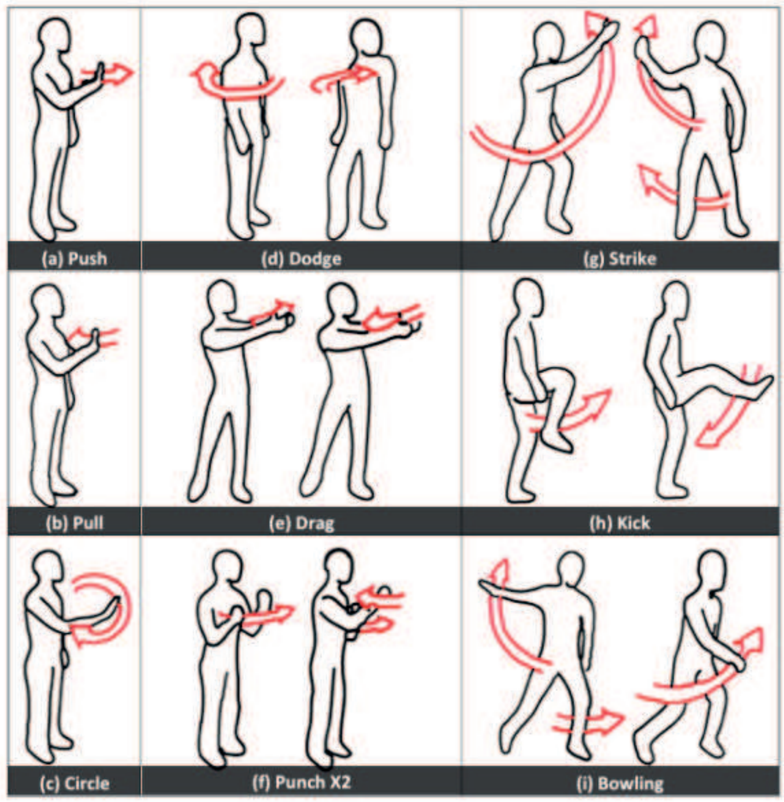


Figure 2.8 – Example of nine gesture sketches by gesture-recognition interfaces (Pu et al., 2013).

the interface are due to system notifications, about an event or condition that can actually be irrelevant for the main task. Several studies (Kjeldsen e Kender, 1997; Morency e Darrell, 2006) show that users tend to better accept interfaces that allow a multimodal interaction with notification windows, where the user can accept or deny the action required by a simple head or hand gesture, thus representing a "yes" or "no" response. This action would not disturb the user, as it is considered innate, and allows the interaction with the system without pauses to search for third-party peripheral devices (e.g. the mouse) to be able to give the system an answer for the notification.

However, the recognition of gestures is not merely a matter of head or hand movements. Indeed, the most used human gesture is gaze, or eye movements (Bridgeman, 1992). Like other body gestures, studies reveal that methods of tracking gaze can also be classified in accordance with two approaches: a 2D approach (bi-dimensional), that analyses the path of the user's gaze directed to the screen (Figure 2.9) by using simple algorithms to calculate the gaze's position; and a 3D approach (three-dimensional) that considers the three-dimensional structure and the own movement of the user's eye. The latter, despite requiring a more complex calibration and being more cumbersome in terms of the system's processing, can achieve more accurate and reliable tracking paths (Hwan et al., 2010).

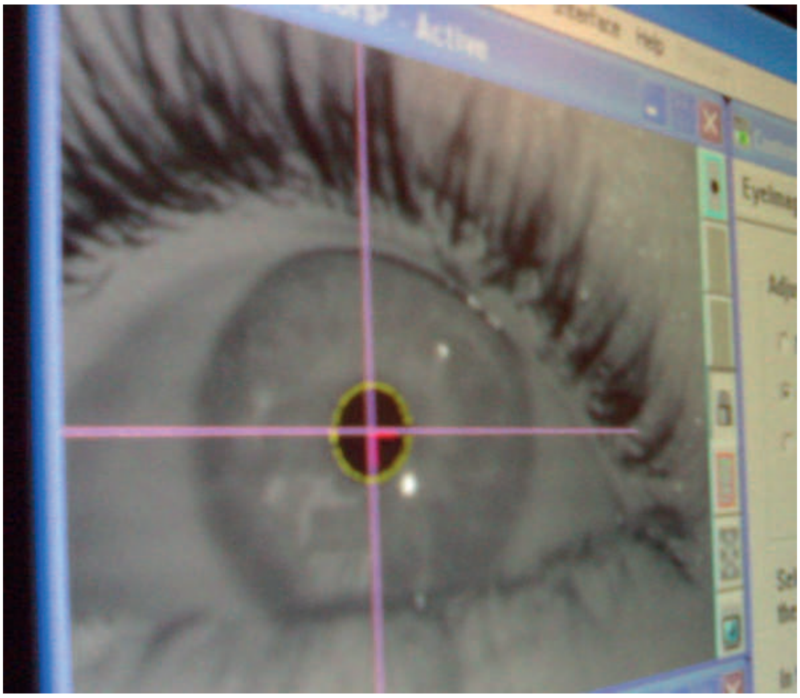


Figure 2.9 – Example of a calibration layout for using eye-tracking devices (Photograph taken by (Analisis e investigacion)).

Recognition systems based on eye tracking have been perfected and have become more used each day throughout the work in different fields of study that include, among others: studying users' attention focus as a way of communicating with other people/technological systems (Jokinen et al., 2010); attentive user interfaces that take advantage of the posture, gaze and proximity of the user in order to determine which task or device is preferred (Vertegaal, 2003); immersive environments that encourage collaborative work between users on different geographical locations (Murray et al., 2007; Pietinen et al., 2008); interfaces able to improve teaching and learning of contents (Lech e Kostek, 2010) and contribute, simultaneously, for the development of guidelines in this field (Van Gog e Scheiter, 2010). Moreover, gesture-based interfaces can be important to improve interaction regarding people with disabilities or impairments (Lee et al., 2005): "Face and gesture recognition provides enormous benefits in the design of augmentative and alternative communication systems for the disabled" (Reilly, 1998, pp. 20). People that can only control eye gaze may be potential users of this paradigm, being this the only one that can be considered by these users regarding interaction with digital systems (Kirbis e Kramberger, 2009). In fact, making use of brain waves using exclusively electric activity acquired from a brain computer interface device, by performing actions such as winking, could also mean an expected improvement in people's autonomy and quality of life (de Castro, 2013). Indeed, these gestural interfaces may be very important for users with autism and other similar psychological disorders, as a way to better explain their vision and attention focus (Riby e Hancock, 2009).

Technical issues have plagued eye tracking in the distant past, making it unreliable and time consuming, and there are still problems with eye tracking a considerable minority of participants - typically 10 to 20% cannot be tracked reliably (Goldberg et al., 2002). Also, the need to limit the physical relationship between some eye tracking systems and the participant is one of the most significant barriers to inclusion of eye tracking in more studies (Jacob e Karn, 2003).

One other disadvantage of this interaction paradigm when using bi-dimensional gestures is the incapability to provide the user with a more immersive environment (Yoo et al., 2010). Hence, other interfaces able to perceive three-dimensional gestures have surfaced, in order to give the user more freedom on manipulating virtual objects, and supporting spatial data and the sense of depth (Benko e Wilson, 2008; Hilliges et al., 2009; Yoo et al., 2010).

Indeed, both gestural and gaze movements are an incentive in the development of new interfaces intuitive for the user, allowing a greater flexibility in the achievement of simple or even complex tasks. However, there is currently a strong limitation regarding the implementation of such mode of interaction: the muscle fatigue that comes associated with this type of paradigm, also known as "gorilla arm". This difficulty arises from the fact that an arm, or other part of the body, must be elevated or held higher in order to interact with the system, thus causing discomfort if the action or task to perform is lengthy (Wachs et al., 2011).

2.2.3 Haptic interfaces

Haptic technology refers to the sense of touch, taking advantage of vibrations and/or forces being applied to the user's body. This designation contemplates data acquisition and object manipulation by the means of the user's touch, considered as manual interaction with environments that can be real or virtual (Srinivasan e Basdogan, 1997). An example of two devices that brought awareness to the masses was the commercial product Apple iPhone, that provided a haptic feedback coupled with the visual feedback of the screen from a tap (Figure 2.10a), and the Wiimote, a remote controller for a home video game console with haptic features (Figure 2.10b).


a)


b)

Figure 2.10 – Examples of haptic devices: a) Apple iPhone's haptic ability (image taken from (Smykil, 2008)); b) Nintendo Wiimote (image taken from (Nintendo, 2006))

According to Robles-De-La-Torre (2009), the haptic technology is able to provide a richer interaction with an object that we can manipulate directly with our hands or body and immediately feel the result, as this is a bi-dimensional stimulation between haptic objects generated by the machine (also called computer-generated Haptic Virtual Objects (HVO)) and the user. Haptic interaction can be coupled with other sensorial faculties, as vision and audition, as a way to create a sense of immersion through an interface that allows the user to touch, feel, manipulate objects, and also see them (or even hear them), in a way that would not otherwise be possible (Srinivasan e Basdogan, 1997). These sensorial interfaces should adjust to other interaction paradigms and allow a more intuitive use of the systems multimodally, taking into account several input/output mechanisms in order to create synergies amongst them (Hale e Stanney, 2004; Khan et al., 2010).

Srinivasan e Basdogan (1997) divide haptic senses into three sub-categories: the human haptic perception, or rather the study of sensation and manipulation through touch; the machine's haptic perception, or better the machine's design to replace or amplify human touch; and also the computer's haptic perception or, in this case, the algorithms and software associated to the interpretation of the touch perception. Indeed, the human haptic system consists of a combination between mechanical, sensory, motor and cognitive components of the human organism. The sensory information of the user's hands can be classified as: tactile data, referring to the sense that is triggered when the skin comes in contact with (touches) an object; and kinaesthetic - or proprioceptive - data, referring to the perception of the body, as position, orientation and force wielded by the limbs.

Indeed, when a user touches a real object, either directly or via a third-party device, forces are exerted on the skin itself. This is the sensory information that is sent to the brain by the human nervous system and subsequently leads to the haptic perception. In order to simulate this sense of touch regarding virtual objects, it is necessary, as such, to generate reaction forces being applied to the skin of the user. These forces are produced through collision calculations between the human hand and the virtual object, replicating the sense of touch as if the object were real regarding its physical and mechanical properties: "weight and shape of objects, object elasticity, object's surface texture (e.g., smooth or rough)" (Robles-De-La-Torre, 2009, pp. 1036).

Haptic computing is a field of rapid progress and development. There is a multiplicity of disciplines it can embrace, such as biomechanics, neurosciences, mathematics, software engineering, rehabilitation, product design, among others, and its research is two-fold: understanding how our haptic sensorimotor system works; and how to develop appropriate haptic interfaces that take advantage of this technology (Srinivasan, 2004). There are several prototypes with resort to this natural interaction paradigm, namely in the field of medicine, with surgical simulations (Basdogan et al., 2004; Kalicki et al., 2007), collaborative work (Kim et al., 2004; You et al., 2007; Hamza-Lup et al., 2009), education (Minogue e Jones, 2006), or even robotics (Wessberg et al., 2000). Moreover, this technology is also studied within the scope of people with disabilities, as visual or hearing impairments (Li et al., 2011; Johnson et al., 2011), since the touch is more intimately related to the users emotions than any other natural interaction paradigm (Ichiyanagi et al., 2011). Indeed, haptic sensations caused by a device as an answer to an action triggered by the user would be very helpful for people with auditory or visual disabilities that could otherwise not notice a sound of a notification alert, but could instead feel its vibration (Jaijongrak et al., 2011).

Currently, we can see this technology incorporated in many devices, although with some limitations, as is the case for joysticks' haptic feedback, haptic electronic gloves, gaming devices and controllers (like the Wiimote, previously mentioned), or even other mobile devices already in the market that show a lot of potential regarding haptic interaction (Chui Yin et al., 2010). However, some technical drawbacks of haptic interfaces include the limited workspace for desktop work, the large weight of force-feedback gloves, the lack of force feedback to the body, safety concerns, etc (Burdea, 2000).

In a way, haptic technology has made possible the study of how humans perceive the sense of touch, by allowing the creation of haptic virtual objects, used systematically to test the ability and sensitivity of human touch. Thus, haptic interfaces aim at augmenting and improving our perception of virtual realities and computer generated data.

2.2.4 Tangible interfaces

One if the first statements recognized in the field of tangible user interfaces (TUI) was at CHI Conference (ACM Conference on Human Factors in Computing Systems), when pioneers Ishii e Ullmer (1997) presented their concept of tangible interfaces as the ones that "augment the real physical world by coupling digital information to everyday physical objects and environments" (Ishii e Ullmer, 1997, pp. 235). Indeed, this was not the first entry on such concepts, also labeled as passive real-world props, embodied or graspable (Fitzmaurice et al., 1995), but it was this perspective presented at the conference that actually showed their advantages. Figure 2.11 demonstrates an example of a tabletop interaction surface that can track small objects on top of it, named PICO (Ishii et al., 2012).

Figure 2.11 – PICO, a tabletop interaction surface that recognizes and tracks small objects and give visual feedback accordingly (Photograph taken from (Ishii et al., 2012))

The criterion of interaction is simple: "a user uses their hands to manipulate some physical object(s) via physical gestures; a computer system detects this, alters its state, and gives feedback accordingly" (Fishkin, 2004, pp. 347). Figure 2.12 explains this process: firstly, an input event occurs when a physical object is manipulated, that is to say, shaken, pushed or merely moved. The system "understands" the event and alters the digital state of the virtual version of that same object, returning the respective output and changing, e.g., the ambient of the surface surrounding the physical object, the environment, emitting sounds, modifying colors, or even giving an haptic feedback. Acting as physical representations that can directly manipulate digital information, tangible objects (or tangibles) can work as parallel input/output devices, simultaneous and in real-time through a passive haptic feedback. Also, regarding the digital environment, the user can have a visual and/or auditory feedback that informs him/her of the intended commands being processed. There is not a separation between the visual representation of digital data (output) and the controllers (e.g. mouse) of that same information (input). Tangibles explore that conceptual gap by eliminating the distinction between input/output of data (Ullmer e Ishii, 2001).

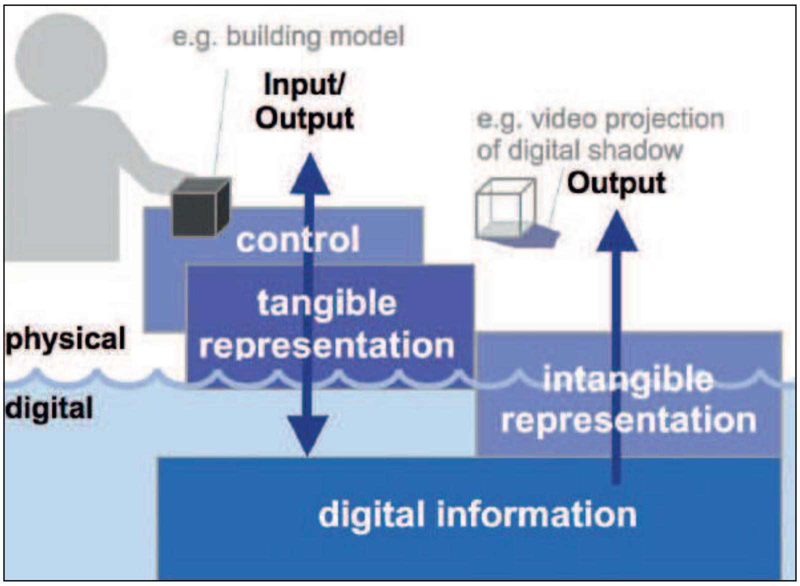


Figure 2.12 – Organization of a tangible user interface (diagram taken from (Ishii, 2008))

There is a plethora of different research fields based on tangible interfaces, that can be mutually affected, and they can be divided into four categories: tangible augmented reality, tangible tabletop interaction, ambient displays e embodied user interfaces (Shaer e Hornecker, 2010). These technologies have been used in diverse fields of work, but the most prominent are within the scope of data storage, manipulation and visualization; spatial modeling and simulation; system's management, control and configuration; education and entertainment (Ullmer e Ishii, 2001).

Regarding tangible augmented reality, virtual three-dimensional objects are combined with physical elements marked with a pattern traceable by the system. The user may manipulate them freely and the output can be visualized through a computer screen, special glasses, or others (Sinclair et al., 2002; Lee e Park, 2005; Sin e Zaman, 2009).

On the other hand, we can also observe a different type of tangible interaction through a touch surface, designated as tangible tabletop interaction, where the same interaction surface also gathers the features of multi-touch technology, and thus enables the user to manipulate simultaneously virtual data and physical objects with specific patterns (Waldner et al., 2006; Liarokapis et al., 2009; Antle et al., 2011; Pedersen e Hornb, 2011). We are bound to mention as a popular example of this type of tangible interaction the product "Reactable", a musical instrument with a revolutionary interface that enables direct manipulation of objects and transforms music into something tangible, that is to say, into visible data with the respective visual feedback (Figure 2.13). Basically, this new intuitive and robust instrument can be used by children and amateur adults (or even music professionals) as the interaction is accomplished through physical objects (small acrylic pieces) marked with specific patterns called fiducials that are recognized by the computer (Kaltenbrunner et al., 2006).

Figure 2.13 – Reactable, a tangible user interfaces that allows the user to interact and create music using specific pieces fiducials markers (Reactable, 2003)

Another concept of tangible interaction is ambient displays. These are mostly based on the exhibition of graphical representations on monitors or interactive walls; but they are also based on the showcase of alterations made to existing objects in some specific physical space, as is the case of phidgets (Greenberg e Fitchett, 2001). In this case, the goal is the attempt to create context-aware environments, where the user can manipulate everyday physical objects to alter features of the environment he/she is in, such as changing the state of digital equipment Greenberg e Fitchett (2001); Brewer et al. (2007).

Finally, there is also a concept of interfaces incorporated on the digital equipment itself, called embodied user interfaces. More and more popular, these interfaces claim that the current computing is increasingly imposing itself on the users' daily lives, being already integrated on physical and virtual devices that may help the users accomplishing their tasks (Gujar et al., 1999; Fallman, 2002; Ferscha et al., 2007; Cafaro et al., 2010)

Indeed, one of the major advantages of this natural interface is the immediate response of the given stimuli through physical objects to the social interactive experience it encourages. The communication between different users and the potential to embrace several objects to manipulate virtual data is an asset of this modality (Mi Jeong e Maher, 2007; Schneider et al., 2011). On the other hand, it is also beneficial for the user to be able to think about virtual and abstract information in a tangible, direct, manner. This is very useful, for example, when working or teaching children (Resnick et al., 1998; Xie et al., 2008; Sylla et al., 2009).

However, there are also clear limitations that arise. Shaer e Hornecker (2010) have defined some challenges about this mode of interaction and its implementation design, namely: the difficulty in resizing the stage of interaction and the possible problem of losing physical objects in such a situation; the struggle in creating a truly versatile and malleable interface that permits the adaptation, reproduction and distribution of information, as the physical objects are rigid and static; and also the possibility of muscular fatigue of the own user during interaction (a problem also shared by gesture-based and touch interfaces).

Hence, despite the challenges that tangible interfaces still face, there are great expectations about their capacity of adjusting to a myriad of applications regarding different fields of action, and also given their advantage of supporting several forms of input and/or output of information. Clearly, this paradigm is not only ruled by the sense of human vision and audition, but also touch.

2.2.5 Touch interfaces

Unlike tangible interfaces, where the user interacts with the virtual world through real objects, the concept of touch surfaces is based on the possibility to interact with digital surfaces through the user's own hands and fingers, allowing the manipulation of digital objects as if they were, indeed, real.

Touch technology has been undergoing significant advances: from technical demonstrations of its abilities, to rich interactive presentations and data manipulation, ever since the launch of the commercial products Apple iPhone and Microsoft Surface.

The multi-touch paradigm refers to the ability of certain devices and equipment to register several simultaneous touches on one same surface caused by the contact of one or more fingers, of one or more users, and thus allowing the recognition of intuitive bi-dimensional gestures. The users are only asked to use their hand to manipulate information direct and intuitively through touches and gestures (Figure 2.14).

Note that we have analyzed numerous advantages of gestural interaction when gesture-based interfaces were reviewed, and thus the same benefits are also applied in this interaction mode. Indeed, the touch paradigm is considered sophisticated and with great potential for expansion mainly due to its ease of use, efficiency and intuitiveness (Kin et al., 2009). Each user, whether or not he/she is familiarized with the technology and/or computers, is able to use multiple fingers to enter commands through touch or gestures. These can represent various actions that are effortlessly achieved by the users to perform a specific task (Figure 2.15), from grabbing virtual objects with the finger, to rotating them or dragging their borders to resize them,

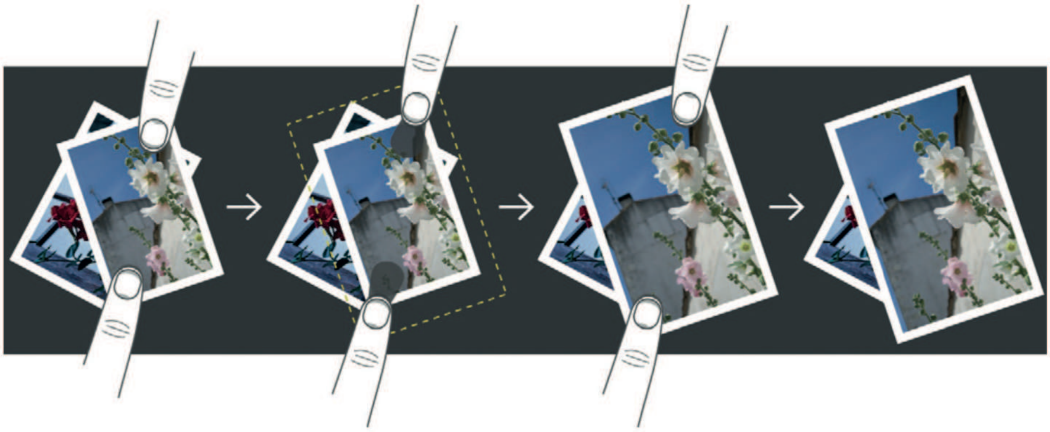


Figure 2.14 – Image scaling using two fingers directly on the surface (Wigdor e Wixon, 2011)

and thus having a physical perception of the virtual data (Towell, 2009).

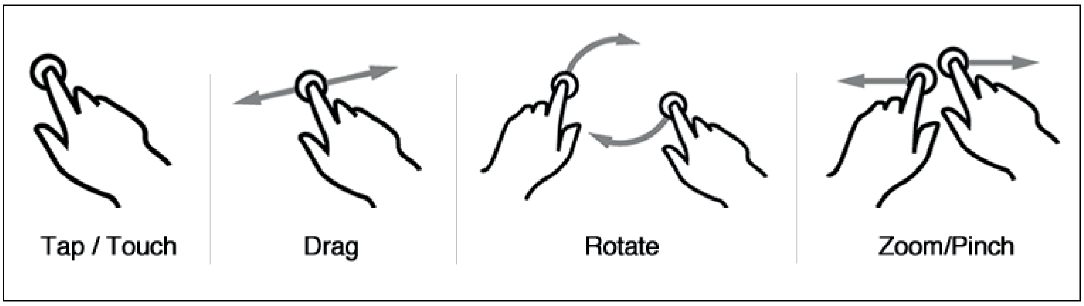


Figure 2.15 – Examples of actions performed by the user using touch and 2D gestures.

Beside the possibility of an intuitive and natural interaction (Wobbrock et al., 2009), this paradigm also enables a high degree of freedom in manipulating data: the major value in this direct interaction is the potential to augment the rhythm of production as far as information use (Kin et al., 2009), as well as its precision, enabling the user to have a high level of control over the virtual data when performing continuous and concurrent tasks (Wigdor e Wixon, 2011). Several other studies show us how this interface can take advantage of the sense of direction of the users' own hands, as well as the distinction between the left and right hand in the process of data manipulation (Dang et al., 2009; Don e Smith, 2010); but also regarding the input of text and content manipulation, suggesting frameworks and guidelines for the implementation of touch interfaces (Cuypers et al., 2008; Kammer et al., 2010; Scholliers et al., 2011; Antle et al., 2011).

Gestural patterns recognized by multi-touch technology are diverse and end up simplifying the users execution of a specific task due to the interface's easy interpretation and empathy with the tasks that exist in real life from the users' daily routines (Morris et al., 2010). Guiard (1987) confirmed that the motor dexterity of hand interaction with the surface resembles the one the user has with real physical objects, even preserving the singularities of content manipulation with the dominant hand, while the other complements the main process with secondary actions.

However, it is necessary to take into account the"gorilla arm" phenomenon: one of the major drawbacks when using vertical touch surfaces as walls and windows displays (Wachs et al., 2011). Indeed, the need to use hands as the interaction mode is not practical in situations of extended use of vertical surfaces, as the user tends to become tired of keeping the arms in an upright position. One other disadvantage of this input modality involves situations where precision is important: the mouse is a tool that easily supports movements that are both precise and rapid; however touch interfaces lack this accuracy due to the known "fat-finger" problem. Indeed, when the user touches the surface, a relatively large area of the finger comes into contact with it, whereas the user wants a small or larger area of selection. That situation can also cause another issue: the occlusion effect, where the user ends up covering unintentionally the area of interest with the finger (Figure 2.16).

As an attempt to perfect human-computer interaction, this paradigm has certainly the advantage of not requiring a manual or mandatory training, as the user can perform tasks more immediately and intuitively (Towell, 2009). This is an important asset that facilitates the interaction process between the user and the machine without the need for learning and familiarizing with third-party devices (e.g. computer mouse or optical pen), which may encourage the use of digital systems by digitally excluded people (Carvalho et al., 2012b), seniors (Leonardi et al., 2010), or users with motor disabilities (Annett et al., 2009).

Figure 2.16 – Issue of "fat fingers", that can cause the occlusion effect and trigger wrong selections.

Touch interfaces are widely used in various domains, being very influential in education (Xiaohua Yu et al., 2010; Martinez-Maldonado et al., 2014; Lievens e Van Daele, 2015) and health (McKee et al., 2015), with the development of applications that help users with special needs (Hourcade et al., 2012) and elderly people (Nawaz et al., 2014; Carvalho et al., 2012b,a). There are also benefits regarding the relationships between users, discouraging isolation and stimulating social collaboration and interaction (Clifton et al., 2011). Indeed, different users can share experiences, since the system recognizes multiple points of contact, enabling interaction with several fingers of one or more users: "The shift to multiple points of input also supports novel forms of interaction where people can share a single interface by gathering around it and interacting together" (Harper et al., 2008, pp. 17). The assets that touch technology brings in this context of collaborative work makes it a reference for many applications that are based on cooperation, regardless of the users' age (Gross et al., 2008; Harris et al., 2009; Scharf et al., 2010).

Therefore, we acknowledge that this interaction modality may even encourage the use of technology by people excluded from the digital world. However, it is necessary to analyze the type of multi-touch applications to develop, as well as the implementation model and how to use them.

2.3 Multimodal interfaces

Interfaces can be independent or combined in a multimodal system richer in terms of human-computer interaction, using one's strengths to compensate for the other's weaknesses (Cohen, 1992). The paradigm shift to multimodal user interfaces (MUIs) has taken place over a period of 40-50 years (Oviatt e Cohen, 2015), evolving from mouse-driven graphical user interfaces to the combination of multiple modalities.

This new era of multimodality turns the interaction with the computer into something much closer to the one users have dealing with other people in the real world. There is, indeed, an attempt to bring the digital world closer to the physical one with new familiar interaction metaphors emerging (Mundie, 2010). The combination of input and output of information defines a mode of interaction, and thus the term multimodal refers to the combination of various modes of human-computer interaction as a part of a natural interface. Indeed, human interaction is multimodal by nature and, as such, a rich multimodal interaction is part of what defines a natural experience (Jain et al., 2011). It is important to understand the potential of the technology that takes advantage of the human senses during interaction, whether on individual interfaces or even in a multimodal system, where they coexist together in one same interface (simultaneously or alternately). These input modalities may involve recognition-based technologies (e.g. gestures, speech), simpler discrete input (e.g. the keyboard, touch), or sensory-based information (e.g. haptic sense, tangibles) (Oviatt e Cohen, 2015). Therefore, the development of novel multimodal systems intends to recognize naturally occurring forms of human language and behavior, which incorporate at least one recognition-based technology (Sears e Jacko, 2007).

Since the appearance of Bolt's "Put That There" interface (Bolt, 1980), which recognized speech in parallel with a touch-pad for pointing for manipulating information (Figure 2.17), a variety of new multimodal systems have emerged. Indeed, the tendency towards multimodal interfaces that have recently been emerging aims at supporting more natural and flexible user inputs, compared to past keyboard-and-mouse interfaces that are limited to discrete input (Oviatt e Cohen, 2015). Multimodal interfaces are expected to be easier to learn and use, therefore being preferred by users for many applications, and also show the potential to work in a more robust and stable manner than unimodal recognition systems involving a single recognition-based technology (Oviatt, 2007). As these interfaces allow us to take advantage of recognition-based technologies that understand human behaviors, they have evolved to include new modality combinations, e.g. pen devices and speech (Watanabe et al., 2007; Piper et al., 2011), speech and gestures (Liu e Kavakli, 2010; Miki et al., 2014), or multi-touch and tangible interfaces (Lucchi et al., 2010; Riedenklau et al., 2012; El-Glaly et al., 2013).

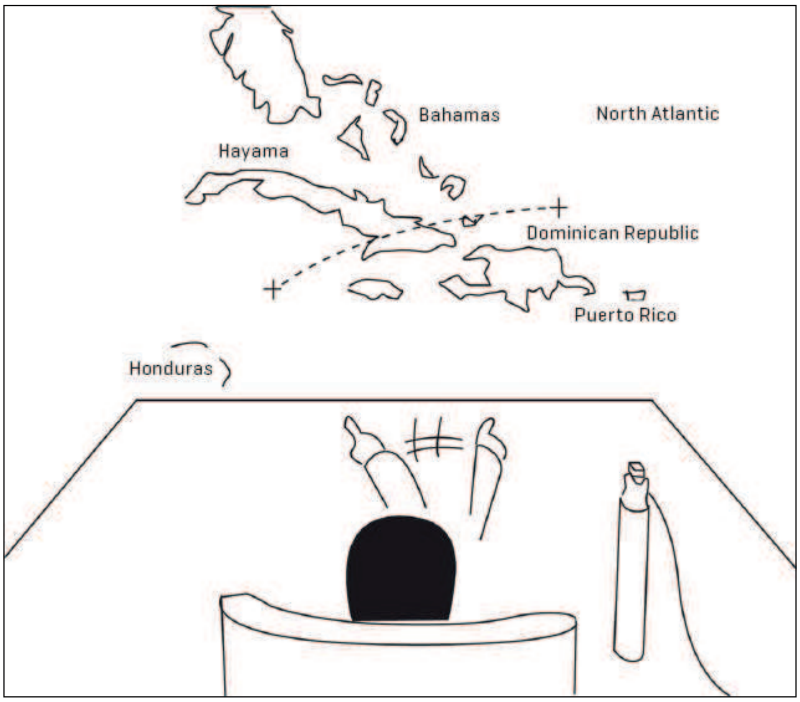


Figure 2.17 – Richard Bolt's "Put That There" interface combined speech and gestures as input modalities (Bolt, 1980).

Multimodal systems have focused on three major concerns that try to attend in order to permit a more transparent experience than ever before (Oviatt e Cohen, 2000).

- **Accessibility for diverse users and usage contexts:** One main concern of multimodality is the potential of the systems to greatly expand the accessibility of computing to diverse and non-expert users. It is one central goal of this novel interaction paradigm to increase accessibility of computing for users of different ages, skills, levels, cognitive styles, sensory or motor impairments, native languages or even temporary illnesses. All because multimodal interfaces permit users to have control over how they interact with the computer, or switch between modalities, and the systems should begin to understand the users' profile and context of use and provide the necessary and appropriate input modes.

- **Performance stability and robustness:** The second goal of multimodal architectures is to improve the performance stability and robustness of recognition-based systems (Oviatt, 1999), i.e. provide several input modes effectively so that errors are avoided. Thus the systems should provide parallel or duplicate functionalities, so that users can overcome problems found during interaction through a less efficient input modality for a specific task.

- **Expressive power and efficiency:** Point-and-click interfaces that rely on the keyboard and mouse input are limited or inappropriate for interacting in specific environments and contexts. On the other hand, systems that interpret multimodal input aim to give users a more powerful interface for accessing and physically manipulating information.

It has been demonstrated that multimodal interaction can be faster than point-and-click systems (Chen, 2006; Cohen et al., 1998). But the basic addition of one other input modality to a unimodal interface does not necessarily lead to improvement (Oviatt, 1999). A higher cognitive load due to more degrees of freedom may instead be the result (L Schomaker, 1995), and the different modalities may interfere with each other (Wechsung et al., 2009) when presenting similar information via two modalities due to synchronization problems that may occur (Schnotz et al., 2002). Also, if different modalities refer to the same cognitive resources, task performance may decrease due to redundancy of information (Wickens, 2002). Single input modalities, as well as their combinations, have to be taken into account in order to be able to bond the advantages of each and encourage the user to execute relevant commands (Wechsung et al., 2009).

In order to demystify multimodal assumptions of computationalists, researcher and expert in the area Sharon Oviatt (1999) enumerated a list of misconceptions associated to these interfaces as a way of replacing popularized myths with a more accurate foundation for guiding the design of next-generation multimodal systems:

- **Myth #1: If you build a multimodal system, users will interact multimodally.** Users like being able to interact multimodally, but they do not always do so: even though users prefer multimodal interaction, there is no guarantee they will issue every command to the system multimodally. They are most likely to interact multimodally when manipulating an object, e.g. for zooming.
- **Myth #2: Speech and pointing is the dominant multimodal integration pattern.** It is true that since the development of Bolt's "Put That There" system (Bolt, 1980) the concept of multimodal interaction as point-and-speak seemed limited to be used only as a new input mode for selecting objects - just as the mouse does, but this just represents the persistence of the old mouse-oriented metaphor. During interpersonal multimodal communication, linguistic analysis of spontaneous gesturing confirms that deictic gestures (e.g. pointing) account for less than 20% of all gestures (Feyereisen, 1994), which implies that multimodal systems designed exclusively to process speak-and-point will fail to provide users with many useful functionalities
- **Myth #3: Multimodal input involves simultaneous signals.** Although two input modes can be highly interdependent and synchronized during multimodal interaction, this does not necessarily imply that they co-occur temporally, i.e. input will be simultaneous. Indeed, synchrony does not imply

simultaneity: user's input of information does not frequently overlap at all during multimodal commands.

- **Myth #4: Speech is the primary input mode in any multimodal system that includes it.** Speech used to be viewed as a primary input mode, considering gestures, head and body movements, direction of gaze, and other input as secondary. This seemed the case in primitive speak-and-point integrations, where the secondary modality is used only for simple selection. However, depending on the contexts (e.g. noisy environment), the relevance of the modalities can be inverted, as happens when confidence in speech recognition is low. For example, pen and speech interfaces tend to have 97-99% of the time occupied by the pen when used with children and adults (Oviatt, 2003).

- **Myth #5: Multimodal language does not differ linguistically from unimodal language.** When interacting multimodally, users selectively eliminate many linguistic complexities and prefer not to speak utter error-prone commands, using instead other means of input. On the other hand, when used in unimodal interfaces, users do not have the desire to substitute their commands with other modalities, increasing the error rate and inefficiency.

- **Myth #6: Multimodal integration involves redundancy of content between modes.** The concept that multimodal interfaces contain a high degree of redundancy is commonly discussed. However, the dominant goal of such interfaces is actually complementarity, not redundancy: alternate input of information complements the content and does not add information redundancy. For example, speech and pen input regularly contribute distinctive and complementary information: the pen is normally used regarding spatial information, whereas speech is normally used regarding temporal information - speech and gestures complement each other (Oviatt, 2003).

- **Myth #7: Individual error-prone recognition technologies combine multimodally to produce even greater unreliability.** It is believed that combining two error-prone recognition technologies does not result in a more robust interface, but instead results in compounded errors and even greater performance unreliability. However, in a flexible multimodal system, users will avoid using an input modality that they consider to be error-prone for certain contents. Indeed, the increased robustness of any multimodal system depends on designing an architecture that integrates modes synergistically, where mutual disambiguation of two input signals can occur (Oviatt, 1999). The greater goal would be to integrate input modes that complement each other (creating a synergy amongst them) and where some interaction techniques may help overcome other weaknesses.

- **Myth #8: All users' multimodal commands are integrated in a uniform way.** During interaction, there can be large individual differences in modality integration patterns. Each user can have different patterns when selecting modalities, but they tend to remain consistent throughout interaction. Multimodal systems that can detect and adapt to a user's dominant integration pattern lead to considerably improved recognition rates.

- **Myth #9: Different input modes are capable of transmitting comparable content.** This assumption is based on the belief that different modes are basically interchangeable and are able to transmit comparable content, as well as on thinking that input and output modalities may suit any user's physical, perceptual, or cognitive limitations. However, we must acknowledge that different modes that recognize speech, handwriting, manual gesturing, head movements, and gaze are each strikingly unique. They differ in several ways: the type of information they transmit; their functionality during communication; the way they can integrate with other modes; and how suitable they are to be incorporated into different interface styles.

- **Myth #10: Enhanced efficiency is the main advantage of multimodal systems.** It is often assumed that the primary advantage of a multimodal system is enhanced speed and efficiency enabled by parallel input modes. However, this efficiency benefit may be limited to specific domains, and this has not been demonstrated yet when task content is verbal or quantitative in nature (Oviatt, 1997). Moreover, other advantages equally important also arise:

> the flexibility that multimodal systems permit users in selecting and alternating between input modes, avoiding physical exhaustion for any individual modality; the avoidance of error and easier error recovery; the convenience to accommodate a wide range of users, tasks, and environments.

The ability to develop a new generation of robust multimodal interfaces depends on knowledge of the integration and synchronization requirements for combining different modes into a strategic whole system, and on understanding the interaction patterns of different users (Oviatt, 1997). Multimodal systems incorporate different input modes that can either create complementarity or redundancy in one same interface. As such, Oviatt elaborated a terminology related to multimodal interfaces, which differentiates the types of input that are eligible to be integrated in one same system (Figure 2.18).

The most advanced multimodal interfaces are the fusion-based ones, where two or more input modes are co-processed and can be entered simultaneously (e.g. speech and lip movements, speech and gestures). On the other hand, the simplest multimodal interfaces have alternative input modes, rather than simultaneous (Oviatt e Cohen, 2015). However, existing multimodal systems are mainly bimodal and therefore constrained in the number and type of input modes they can recognize. Even more, some bimodal systems, as speak-and-point, only act as simple mouse replacements, like writing, gestures, gaze, or facial expressions, and therefore should not be considered as fully functioning multimodal systems (Oviatt, 2003). Indeed, these constrains are mostly observed in visual-spatial domains and even regarding temporal constrains (e.g. first, the users turn to the pen and only afterwards to the voice when specifying a command). Depending on the task at hand, spatial commands have better chances of being multimodal.

As multimodal interfaces advance towards the support of human behaviors in specific contexts, they have begun to incorporate more vision-based technologies that interpret manual gestures or body positions (Turk e Robertson, 2000). This has caused a growth in multimodal interfaces that passively monitor users' behavior

Multimodal interfaces process two or more combined user input modes— such as speech, pen, touch, manual gestures, gaze, and head and body movements— in a coordinated manner with multimedia system output. They are a new class of interfaces that aim to recognize naturally occurring forms of human language and behavior, and which incorporate one or more recognition-based technologies (e.g., speech, pen, vision).

Active input modes are ones that are deployed by the user intentionally as an explicit command to a computer system (e.g., speech).

Passive input modes refer to naturally occurring user behavior or actions that are recognized by a computer (e.g., facial expressions, manual gestures). They involve user input that is unobtrusively and passively monitored, without requiring any explicit command to a computer.

Blended multimodal interfaces are ones that incorporate system recognition of at least one passive and one active input mode. (e.g., speech and lip movement systems).

Temporally-cascaded multimodal interfaces are ones that process two or more user modalities that tend to be sequenced in a particular temporal order (e.g., gaze, gesture, speech), such that partial information supplied by recognition of an earlier mode (e.g., gaze) is available to constrain interpretation of a later mode (e.g., speech). Such interfaces may combine only active input modes, only passive ones, or they may be blended.

Mutual disambiguation involves disambiguation of signal or semantic-level information in one error-prone input mode from partial information supplied by another. Mutual disambiguation can occur in a multimodal architecture with two or more semantically rich recognition-based input modes. It leads to recovery from unimodal recognition errors within a multimodal architecture, with the net effect of suppressing errors experienced by the user.

Simultaneous integrator refers to a user who habitually presents two input signals (e.g., speech, pen) in a temporally overlapped manner when communicating multimodal commands to a system

Sequential integrator refers to a user who habitually separates their multimodal signals, presenting one before the other with a brief pause intervening

Multimodal hypertiming refers to the fact that both sequential and simultaneous integrators will further accentuate their basic multimodal integration pattern when under duress (e.g., as task difficulty or system recognition errors increase)

Visemes refers to the detailed classification of visible lip movements that correspond with consonants and vowels during articulated speech. A *viseme-phoneme mapping* refers to the correspondence between visible lip movements and audible phonemes during continuous speech.

Feature-level fusion is a method for fusing low-level feature information from parallel input signals within a multimodal architecture, which has been applied to processing closely synchronized input such as speech and lip movements.

Semantic-level fusion is a method for integrating semantic information derived from parallel input modes in a multimodal architecture, which has been used for processing speech and gesture input.

Figure 2.18 – Multimodal terminology (Oviatt, 1997).

(e.g. gestures) without requiring any explicit commands. Even more, some interfaces even encompass active input modes (e.g. speech) to intentionally issue commands (Oviatt, 2003). The latter are blended multimodal interfaces, that are typically temporally cascaded, meaning that information from the first input mode can influence the following mode and constrain it to a more accurate interpretation (e.g. combining gaze and speech). Indeed, passive input modes are more attentive to the users' behaviors and combine an understanding of natural human capabilities (e.g. motor, cognitive, and perceptual skills) with computer input/output devices and machine perception and reasoning (Turk e Robertson, 2000). Figure 2.19 illustrates the flow of information of these perceptual user interfaces. These devices and sensors are transparent and are passively monitoring the user in order to simulate the ways in which people naturally interact with each other and with the world - both verbally and nonverbally. Indeed, blended multimodal interfaces have the potential to improve robustness and offer broader application functionality.

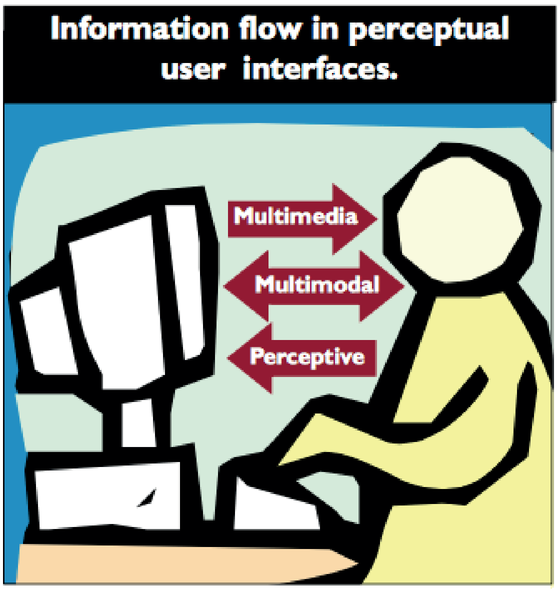


Figure 2.19 – Information flow in perceptual user interfaces (Turk e Robertson, 2000).

Moreover, since there are large individual differences regarding the type of users, the time a user takes to issue a command may be influenced by age, skill levels or sensory impairments. The context may also influence the type of input mode the user chooses.

Finally, the two major concerns in the development of multimodal systems - that have previously been reviewed when the 10 myths of multimodal interfaces development were discussed - fall upon the notion of: (1) complementary versus redundancy; and (2) error suppression (Oviatt, 2003).

Indeed, in multimodal systems, one explicit goal has been to integrate complementary modalities to provide a synergistic blend, so that each mode can be used to overcome weaknesses in the other mode (Cohen et al., 1989). By integrating more than one technology, it is simpler for the system to understand the users' queries, and therefore they will have an easier and more expressive means of achieving their goals. Indeed, the input should complement the content and not add information redundancy. This approach to system design has enforced the combination that allows for a mutual disambiguation associated with each mode of interaction, which means that the greatest advantage of multimodal interfaces is their ability to better handle errors than unimodal recognition-based interfaces, whether in terms of error avoidance, or recovery from errors (Suhm, 1998).

When users encounter system errors, they have a strong tendency to switch modes, facilitating error avoidance and recovery, as well as improving user satisfaction with error handling (Oviatt, 1999). Users tend to make less mistakes when they can choose the desired input mode, and even when they have different input modes they can choose from, having the possibility to vary from one to another during interaction. Therefore, there is a lot less frustration for the user regarding interaction. Therefore, a well-designed complex multimodal system should support mutual disambiguation of input signals: one input mode helps to complement another so that there are no doubts regarding the issued command. This technique of error suppression involves recovery from unimodal recognition errors within a multimodal architecture, where each mode used supplies partial disambiguation of the other mode, thus leading to a more robust and stable system performance (Oviatt, 1999).

In conclusion, multimodal interfaces recognize and identify users' natural actions and language, empowering systems to better understand human behaviors and better support them to avoid errors and interact more human-like with digital interfaces. Recent multimodal systems now accept a wider gamut of input integrations, which are no longer restricted to the simple point-and-speak combinations handled by earlier systems. And such interfaces will eventually interpret continuous input from a large number of different visual, auditory, and tactile input modes, supporting an intelligent adaptation to the user, task and environment.

2.4 Interaction Design

Recently, the notion of interaction design started to gain in popularity as a way to acknowledge the combination of two distinct areas of expertise: design and engineering. Thus, a more "designerly" approach goes beyond pure usefulness and efficiency to consider also aesthetic qualities of use, for example: it is interaction design that makes artifacts usable, useful, and fun, whether they are computers or other digital equipment (Soegaard e Dam, 2014).

It was the industrial designer Bill Moggridge, back in 1990, that first introduced the concept of interaction design, and since then this discipline has evolved and infiltrated in a lot of fields. In his 2007 book "Designing Interactions", Moggridge explains this new design discipline dedicated to creating attractive solutions in a virtual world and making them useful to the real world, where two distinct systems may look identical but "feel" so different (Moggridge, 2007). According to Saffer (2009), there are three ways of looking at interaction design: a technology-centered view; a behaviorist view; and a social interaction design view. All these major schools of thought have in common the view that this discipline is art combined with engineering, and that it has a contextual nature: it solves specific problems under a particular set of circumstances using the available materials. In fact, it has many roots in sister fields (Figure 2.20), such as information architecture (IA), industrial design (ID), visual (or graphic) design, user experience (UX) design, and human factors (Saffer, 2009).

The paradigm shift to a user-centered design has enabled new interaction processes that focus on the requirements of end-users, and provide early feedback to design, so as to reduce the cost of design defects and to meet specific usability objectives (Norman e Draper, 1986). This in turn implies five major characteristics for interaction

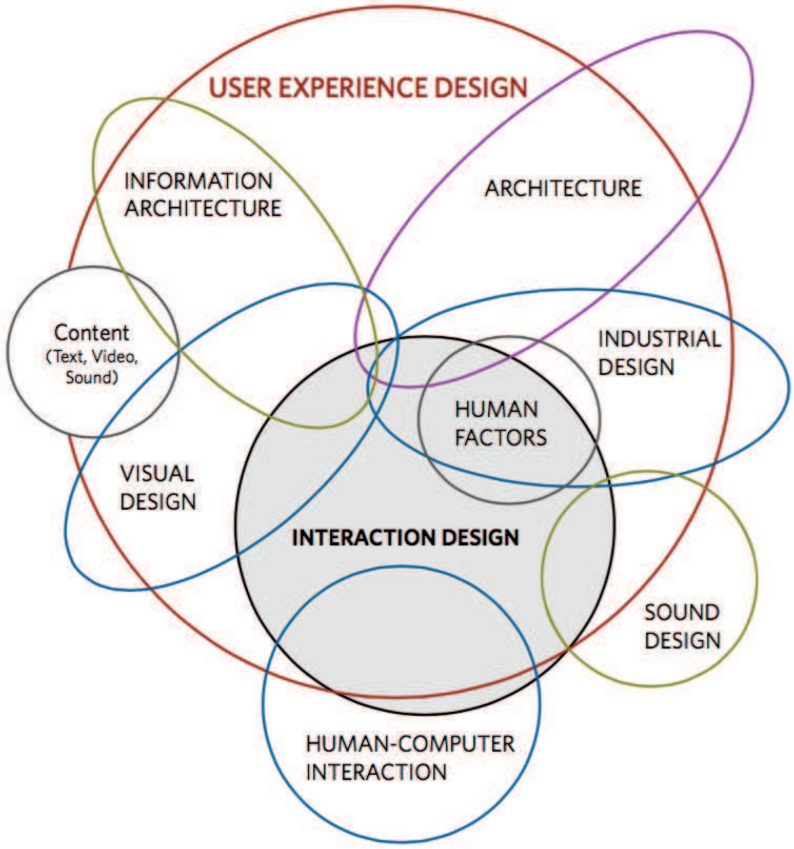


Figure 2.20 – The disciplines surrounding interaction design (Saffer, 2009).

design (Soegaard e Dam, 2014):

- Design involves changing situations by shaping and deploying artifacts;
- Design is about exploring possible futures;
- Design entails framing the "problem" in parallel with creating possible "solutions";
- Design involves thinking through sketching and other tangible representations;
- Design addresses instrumental, technical, aesthetical and ethical aspects throughout.

Indeed, interaction design is about transformation of digital materials - software, electronics, communication networks - and analytical and critical studies focus on what exists, whereas design concerns itself with what could be. This notion of changing situations and exploring possible futures resolves in the idea that when we have designed something, the situation in which it is used is no longer the same (Soegaard e Dam, 2014). Also, a key feature of interaction design addresses sketching possible aspects of the digital artifact, being these drawings micro-experiments and experience prototypes that provide insights regarding strengths, weaknesses and possible changes to overcome (Buxton, 2007). Technical assessments influence the aesthetic and ethical qualities of the resultant interaction, but choices on what features to offer may have ethical repercussions: "Whether something looks and feels good to use, and whether it makes you comfortable in terms of social accountability and moral standards, has a real impact not only on the overall user experience but also on measurable, instrumental outcomes" (Soegaard e Dam, 2014).

The main goal of interaction design is to provide the users with the ability to perform the tasks as quickly, efficiently and correctly as possible. Thus, the notion of user experience to apprehend the various features of non-instrumental, aesthetical, emotional qualities in the human use of a digital artifact is being developed alongside the discipline of interaction design (Buxton, 2007).

Also, interaction techniques and devices are important parts of the user-computer interface. The costs of poorly designed interfaces have been recognized: whether in the form of degraded user productivity, user frustration, increased training costs, the need to redesign and re-implement the user interface, etc (Foley et al., 1984). In this sense, a special attention must be given to the concept of interaction devices, techniques and tasks. As such, the following section is a presentation of the most relevant characteristics and properties of common input devices and types of interaction tasks and techniques.

2.4.1 Types of tasks

Interaction sequences can be decomposed into a series of basic interactions tasks. These are low-level primitive inputs required from the user, such as entering a text string or choosing a command, and thus, for each task, an appropriate interaction technique must be chosen.

Early efforts in human-computer interaction sought to identify user interface transactions that appear repeatedly during interaction. Foley et al. (1984) proposed six fundamental low-level types of interaction tasks, that is, primitive action units performed by a user:

i **Select:** Making a selection from a set of alternatives.

ii **Position:** Indicating a position within a range, e.g. picking a screen coordinate with a pointing device.

iii **Orient:** Specifying a 2D angle or three-dimensional orientation of an object, e.g. rotating a symbol or controlling the perspective of a 3D model.

iv **Path:** Specifying a series of positions and/or orientations over time.

v **Quantify:** Specifying an exact numeric value, e.g. the height of an object.

vi **Text:** Inputting a text string/entry of data.

Each of these tasks can be implemented through several interaction techniques, although the task requirements should limit the choice of techniques to those whose properties match these requirements, as will be addressed below in the section about interaction techniques.

None of the six primitive transactions described actually modifies the object itself being displayed. In this sense, Foley et al. (1984) also proposed a number of tasks whose basic purpose is to control the objects which are already visible on the screen, rather than just specifying something (as is the case of elemental tasks). Thus, these transactions are known as controlled tasks and are named for the type of modification they effect on the object (Foley et al., 1984):

i **Stretch:** Distorting the shape of the object.

ii **Sketch:** Manipulating a device like a brush, to create line structures or freehand sketching.

iii **Manipulate:** Moving an object by either dragging it, twisting it or scaling it, depending on whether the manipulation is based on translation, orientation or valuation techniques.

iv **Shape:** Creating a smooth, curved line or surface to change its general shape according to a positioning control.

However, it is not clear where new input devices that are appearing in the market fit in, as they can offer new "elemental" data types as location (e.g. GPS device), images (e.g. cameras), identity (e.g. fingerprint scanners) (Hinckley, 2007).

While elemental tasks are commonly occurring tasks in many direct-manipulation interfaces, a challenge with this taxonomy is that it is not very clear and well defined what actually entails "elemental" tasks. For example, considering position tasks, a coordinate on the screen could be indicated using a pointing device (such as a mouse), but might also be entered as a pair of numeric values (quantify task). Here, it is not clear what constitutes the elemental task, as two types of tasks are being considered in one same transaction.

A method to solve this issue is to view all tasks as hierarchies of sub-tasks (Figure 2.21). Indeed, the main tasks are known as compound tasks (Buxton, 1986), and they entail a series of elemental tasks that may seem like a single task, but are actually divided into different steps or sub-tasks. In actual applications, many of the transactions we perform consist of the so-called compound tasks: selecting an object on the screen and re-positioning it would be an example of a selection/positioning task; or scrolling a web page to click on a link would be considered as an elemental

1D positioning task followed by a 2D selection task, hence a compound navigation/selection task (Buxton e Myers, 1986). An interaction technique can also encourage the user to work resorting to a two-handed input, known as the "chunking" method, to complete a compound path: for example, by scrolling with one hand while pointing to the link with the other hand. The user models the compound task as a single entity (Buxton, 1986).

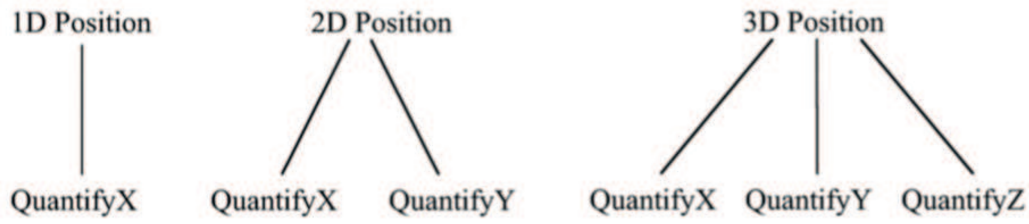


Figure 2.21 – Task hierarchies for 1D, 2D and 3D position tasks (Hinckley, 2007).

In this sense, it becomes important to recognize the different combinations of elemental tasks that can compose a compound task and, also, what forms exist to combine these tasks into one synchronous transaction.

According to (Allen, 1983), there are seven ways of synchronizing elements throughout the tasks based on temporal relations between two actions: before, equal, meets, overlaps, during, starts, ends. These relations have also been adopted not only for multimedia applications, by representing both the temporal and the spatial ordering of actors (Vazirgiannis et al., 1996), but also for multimodal interfaces that require more than one interaction modality to be used. In this kind of interface a very important aspect is the possibility of parallel use of multiple devices provided by the fusion of interactive events. Thus the importance of understanding what forms of action parallelism a compound task may provide. According to Figure 2.22, supposing that A is an elemental navigation task and B is an elemental selection task, the possibilities for combination in a compound task are: A is executed and after a certain time B is executed; A is executed and as soon as it ends B starts to be executed; A is executed in parallel with a part of B; a part of A is executed in parallel with a part of B; A starts to be executed in parallel with B; A ends execution in parallel with B; and A is executed in parallel with B.

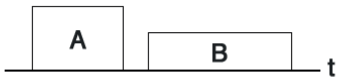
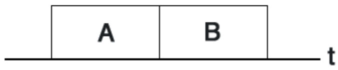
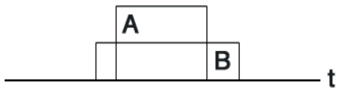
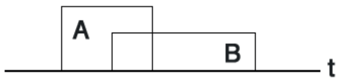
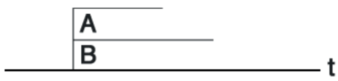
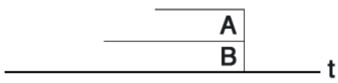
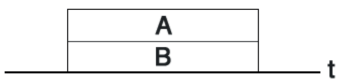
A before B		A < t B >
A ends, B starts		A < 0 B >
A is executed during B		A > t B >
A is executed in parallel with a part of B		A > t B >
A starts with B		A > 0 B >
A ends with B		A <0 B !
A is executed in parallel with B		A > 0 B >

Figure 2.22 – Possibilities to synchronize elements in compound tasks (adapted from (Vazirgiannis et al., 1996; da Silva, 2008)).

The full range of existing combinations allow the user to perform the tasks with different levels of parallelism, i.e. the user does not need to perform all actions simultaneously, being able to opt to perform the tasks without parallelism, with partial parallelism or with absolute parallelism. Mainly concerning multimodal activities, the user may be given the choice and freedom to execute and complete the compound task in the most appropriate manner for the type of interaction that is being performed at each time (da Silva, 2008; Trevisan et al., 2011).

Indeed, the choice of input devices influences the level at which the user must think about the different individual actions that must be performed in order to achieve a goal (Hinckley, 2007). Thus, the choice of device and interaction technique directly influences the steps required from the user and hence the complexity of the interface design.

2.4.2 Input devices

An input device can be considered as a hardware computer peripheral through which the user interacts with the computer (Hinckley et al., 2004). Nowadays, there is a vast amount of input devices and displays on the market used for interacting with the computer systems. Input devices present differences in their capabilities and supported interaction techniques (such as pointing, dragging) in their demands, and the variation in each can be both qualitative (type of signal) and quantitative (amount of signal, or bandwidth) (Buxton, 1990).

Indeed, input devices may sense physical properties of people, places or things, but they must be connected with some sort of corresponding feedback in order to be useful for the user (Hinckley, 2007): they have to provide a way for users to accomplish tasks by combining input with appropriate feedback. Some devices are intertwined between their input and output capabilities (e.g. small-screen devices). Thus, it is particularly challenging to describe properties of input devices without suitable references to output, since input devices are only useful as long as they support interaction techniques that allow the user to accomplish something (Hinckley et al., 2004). In this sense, an interaction designer must be aware of four very important aspects: (a) the physical controller; (b) the feedback provided to the user; (c) the ergonomics of the device; and (d) the relationship between all of the interaction techniques supported by a system (Hinckley, 2007).

Regarding the physical controller, there are a number of organizing properties and principles that can help to make sense of the design space needed for interaction and performance issues that can arise by using either discrete or continuous input devices (Hinckley et al., 2004). We consider discrete mechanisms those that provide an "on/off" state such as buttons or keyboards; and continuous devices as the ones that are manually operated pointing devices, as the computer mouse (Buxton, 1983), as well as sensors that can record continuous data (temperature, light, humidity, etc).

Indeed, the understanding of input technologies and task requirements provides the user with a more natural workflow. The variety of these pointing devices can be summarized according to a few important properties (Hinckley et al., 2004):

- **Physical property sensed:** Pointing devices typically sense linear position, motion, or force (e.g. a tablet senses position, a mouse measures motion - i.e. change in position - and an isometric joystick senses force); whereas rotary devices sense angle, change in angle, and torque (Buxton, 1983; Card et al., 1991). On a different matter, position sensing devices can also be known as absolute input devices, e.g. touch-screens, whereas motion sensing devices are relative input devices, e.g. the computer mouse, that requires a constant visual feedback to report its location (Hinckley, 2007).

- **Gain:** The physical property sensed determines the respective transfer function, that is, determines the most appropriate mapping from input to output of the specific device, and thus matching the physical properties sensed by that device. Appropriate mappings include force-to-velocity, position-to-position, and velocity-to-velocity functions: e.g., an isometric joystick senses force, so the mapping transforms this into a velocity of cursor movement. The ratio between the movement of the input device and the corresponding movement of the object it controls can also be measured, e.g. when the user moves a mouse (the control) 1 cm on the desk in order to move a cursor 2 cm on the screen (the display). Here, the device has a 1:2 control-display (C:D) ratio or gain

(Hinckley et al., 2004). This gain combines two distinct measurements (the device size and the display size) into one metric (Accot e Zhai, 2001) that can be defined as the distance moved by an input device divided by the distance moved on the display. However, on commercial pointing devices and operating systems, an acceleration function is often used to regulate the ratio depending on velocity. An acceleration function is a transfer function between velocity and gain, commonly used to reduce the footprint, or the physical movement space, required by an input device (Hinckley et al., 2002).

- **Number of dimensions:** Devices can sense one or more linear and angular dimensions. For example, a mouse measures two linear dimensions, whereas 3D input devices sense three or more simultaneous dimensions of spatial position or orientation (Hinckley et al., 2004).

- **Pointing speed and accuracy:** The standard way to characterize pointing device performance employs the Fitts' law paradigm (Fitts, 1954). This model associates: the movement time to point at a target, the amplitude of the movement (the distance to the target), and the width of the target (i.e., the precision requirement of the pointing movement). While not emphasized in this chapter, but in the following instead, Fitts' law is the single most important quantitative analysis, testing, and prediction tool available to input research and device evaluation (Hinckley et al., 2004).

- **Direct versus indirect control:** A direct input device is one that the user operates directly and has a unified input and display surface, such as a touch screen; whereas an indirect input device is a device that the user operates by moving a third-party peripheral that is located away from the screen or other display, such as a mouse or trackball. Although they carry a lot of advantages, direct input devices also have several unique issues: there is a possibility of parallax error resulting from a gap between the input and display surfaces, reduced transmittance of the screen introduced by a sensing layer, or the occlusion effect triggered by the user's fingers or hand. This situation may also make the user overlook pop-up menus, dialogs or status indicators (Hinckley, 2007).

- **Device acquisition time:** The term represents the average time to move one's hand to a device, or rather, the time the user takes to pick up an input device. By contrast, the time the user takes to put down an input device and return from a device to a "home" position (e.g. from mouse to keyboard) is known as homing time. This is often assumed to be a significant factor for user performance, but the Fitts' law performance index of a device tends to overshadow acquisition time unless switching occurs frequently (Douglas e Mithal, 1994).

These properties may help categorize input devices, but one must not forget other performance metrics also in place when evaluating pointing devices. Error rates, user preference, comfort and cost are also of utmost importance (Card et al., 1991).

Previous work on systematizing human-computer input devices has provided two major lines of development: toolkits and taxonomies. However, other attempts at categorizing such devices have also been published, concerning (a) performance studies of pointing devices, although they have not attempted to separate task, subject, and human performance variables (English e Douglas C. Engelbart, 1967; Epps, 1986); (b) morphological design space analysis (Card et al., 1991); or (c) classification according to which physical property they transduce, which of the six degrees of freedom they sense, and the measure of the input domain set (Mackinlay et al., 1990).

On one hand, user interface toolkits help to improve some issues, including the construction, run-time execution, and post-run-time analysis of a user interface (P. P. Tanner, 1985). They may help systematize the input device usage by providing libraries of prebuilt modules. However, the device models inherent in user interface toolkits only outline a limited picture of the design space of input devices and their properties. Even when considering the development of interfaces, they present many design alternatives, but do not support or mediate the design decisions themselves. As such, in order to succeed in a systematic framework for input devices, toolkits need to support concepts about the user, the devices themselves, and the task they are used to perform (Card et al., 1991).

On the other hand, taxonomies have also been reported that help categorize input devices. Two main classifications have been presented in this field: computer graphics subtasks (Foley et al., 1984) and input device states (Buxton, 1990).

The first approach classified input devices under the graphics subtasks they were capable of performing (e.g., the tablet and the pen are grouped together because they are both capable of character recognition). The taxonomy of (Foley et al., 1984) was aimed at ensuring that the designer would select the proper devices and techniques with which the user could achieve his elementary tasks more successfully and reliably. The purpose of their report was to provide systematic structure for the process of choosing the right device to accomplish a multitude of interaction techniques. Indeed, each technique has a specific purpose, such as specifying a command, designating a position, or selecting an object on the screen, and each is implemented with some input device, such as tablet, keyboard, pen, etc. Thus, this approach was based not only on the similar properties of the input devices, but also on the types of interaction tasks they would accomplish, that is, it organizes the techniques according to the task they fulfill (Figure 2.23). These basic tasks were gathered in six generic transactions and were already addressed on the previous section regarding interaction tasks. Nonetheless, the limitation of this scheme is that the categories are somewhat ad hoc, i.e. created regarding a very specific purpose, and there is no attempt at taking into account the actual design space (Card et al., 1991).

The second taxonomy of input devices, defended by Buxton (1990), is based on the physical properties and the number of spatial dimensions input devices sense. Indeed, it is difficult to find a way to join the vast demands of graphical interfaces primitive tasks, such as clicking, dragging, double-clicking, right-clicking, and the states and events supported by the input devices (Hinckley, 2007). There is no one simple solution to fit all requirements. In this sense, Buxton (1990) created a model that explains the different states and transitions an input device goes through: the three-state model. Here, it is explained that most input devices support three possible states (Table 2.1): (a) out-of-range, also referred to as State 0; (b) tracking, also known as State 1; and (c) dragging, also called State 2. This approach is useful

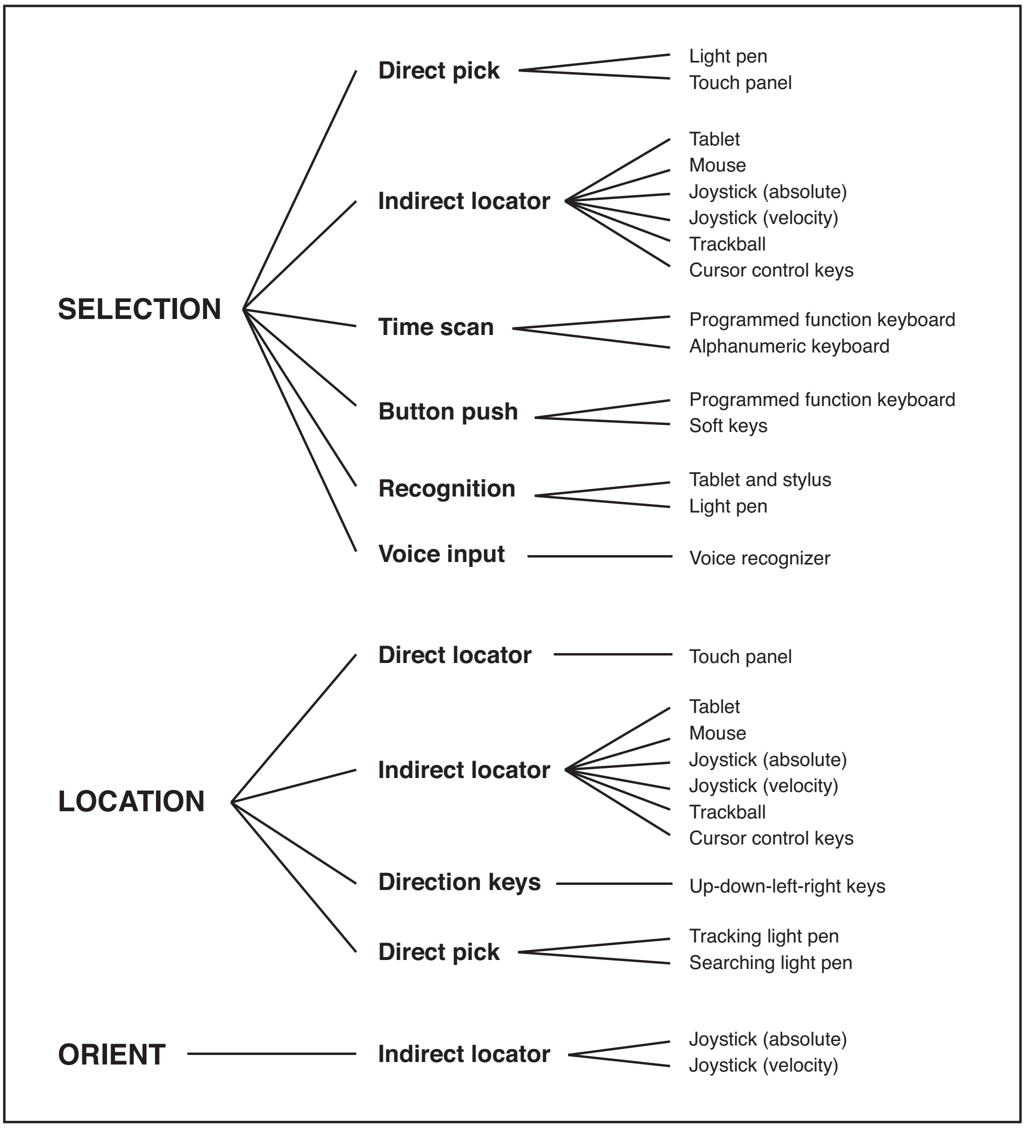


Figure 2.23 – Trees, adapted from (Foley et al., 1984), show the relationship between some computer graphic input applications and various input devices (adapted from (Mackinlay et al., 1990)).

to understand the relationship between the events sensed by an input device and the demands of interaction techniques. Indeed, to make a selection with an input device, users need a way to signal the system about when they are selecting something versus when they are just moving over something to reach the actual desired target. This signal of intention is very important and is thus valuable to acknowledge the different

Table 2.1 – State description of the three-state model (Buxton, 1990).

State	Description
0	**Out-of-range (OOR):** The device moves out of its physical tracking range and has no effect on the system.
1	**Tracking:** Device is in contact with the system and its motion moves the cursor.
2	**Dragging:** Device is in motion and moves objects on the screen by clicking and moving.

possible states of an input device to favor interaction.

Most pointing devices sense only two of these three states: for example, a mouse senses tracking (state 1) and dragging (state 2), but a touchpad senses tracking (state 1) and the out-of-range (state 0). Both are two-state devices, even though they sense different states. Regarding the mouse, state 1 provides cursor feedback of the screen position that the device will manipulate, while state 2 allows the user to drag an object by holding down the main left button while moving the mouse simultaneously. As seen in Figure 2.24, the input device senses the tracking and dragging states (dx, dy). On the contrary, touch-activated devices can sense when a finger is in contact with the surface - equivalent to the mouse dragging state (state 2) - but when the finger is removed from the surface the state falls back to out-of-range (state 0), where no motion is detected (being this the nil state in Figure 2.25).

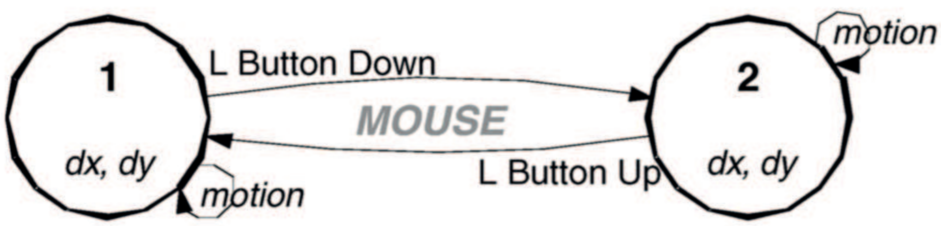


Figure 2.24 – States sensed by a computer mouse - tracking and dragging (Hinckley, 2007).

However, the limitation of this taxonomy is that it only includes continuous devices, maintaining the disadvantage of not being able to characterize all devices. As a result, (Hinckley, 2007) suggested an extended version of the three-state model to

Figure 2.25 – States sensed by a touch-operated device - out-of-range and dragging (Hinckley, 2007).

better characterize all the interaction techniques at the core of graphical user interfaces. The resulting five states were: (a) tracking, (b) hover, (c) left click, (d) dragging, (e) right click. In spite of still being very difficult for the taxonomy to encompass every input device on the market, these state transitions form fundamental indications of intent that are essential for a prompt and determined interaction.

The state-transition model and taxonomy captures many important aspects of input devices and techniques. As such, it provides a means of aiding the designer in evaluating the match between the two and thus creating a model that helps to bring together the device and the ideal technique for the right interaction task at hand.

2.4.3 Interaction modalities versus techniques

Having in the previous sections discussed interaction tasks and devices, we now turn our attention towards the interaction techniques and modalities used to implement the interaction tasks, and the differences between both terms.

There may be several different ways of accomplishing one same task. For example, one could use a mouse to select a command by using: a pop-up menu, a fixed menu (such as a command bar), multiple clicking, circling the desired command, or even writing the name of the command (Hinckley et al., 2004). All these actions are alternative ways to perform one same task using the same input device (the mouse), but resorting to different interaction techniques, that is, different ways of executing the task.

It is incorrect to include hardware device characteristics, such as length or diameter of an input device, as a parameter of an interaction technique. Instead, technique parameters should rely on aspects normally controlled by software (Foley et al., 1984). Also, each of the techniques has a set of hardware prerequisites with respect to both the display technology and the type of device used. Aside from describing a set of six elemental tasks during interaction (as discussed earlier), Foley et al. (1984) define the respective technique to have into account for each elemental transaction: selection techniques, for selection tasks (Figure 2.26); positioning techniques, for positioning tasks (Figure 2.27); orienting techniques, for orientation tasks (Figure 2.28); pathing techniques, for creating paths; quantifying techniques, for specifying values (Figure 2.29); and text entry techniques, for inserting content (Figure 2.30). Also, they specify certain controlled techniques for processing a continuous modification on an object: stretching techniques, sketching techniques, manipulating techniques and shaping techniques.

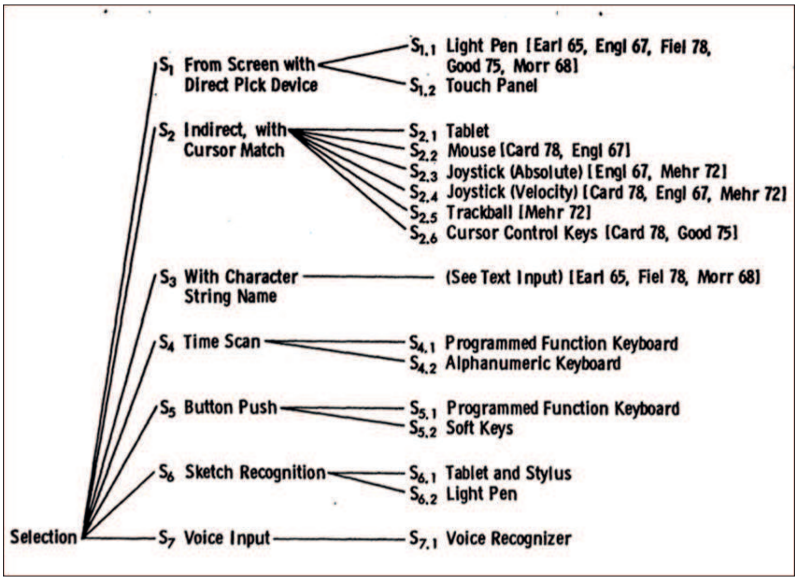


Figure 2.26 – Variety of selection techniques (Foley et al., 1984).

There have been other examples of techniques introduced regarding different contexts of use, such as: immersive virtual environments that present techniques for

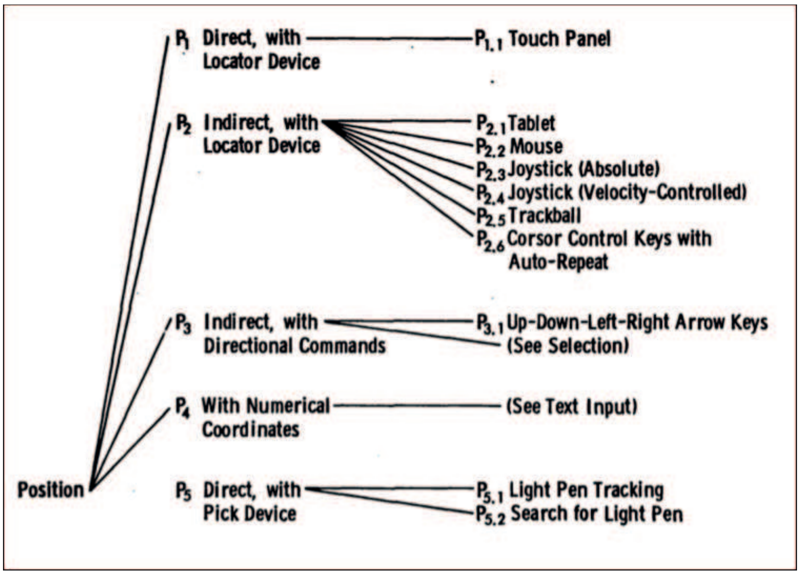


Figure 2.27 – Variety of position techniques (Foley et al., 1984).

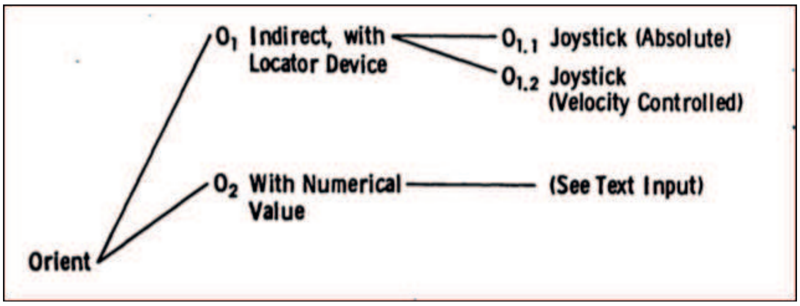


Figure 2.28 – Variety of orienting techniques (Foley et al., 1984).

travelling, selecting and manipulating virtual elements (Bowman, 1999); or physical mobile interaction that present techniques for touching, pointing and scanning (Rukzio et al., 2006).

However, one must not confound the terms interaction technique with modality. An interface mainly relies on the number and diversity of its inputs and outputs, which are communication channels that enable the user to interact with the system via this interface (Karray et al., 2008). Each of the different independent single channels is called an interaction modality. With the ever growing attention turned towards new interaction methods, different research branches have had different

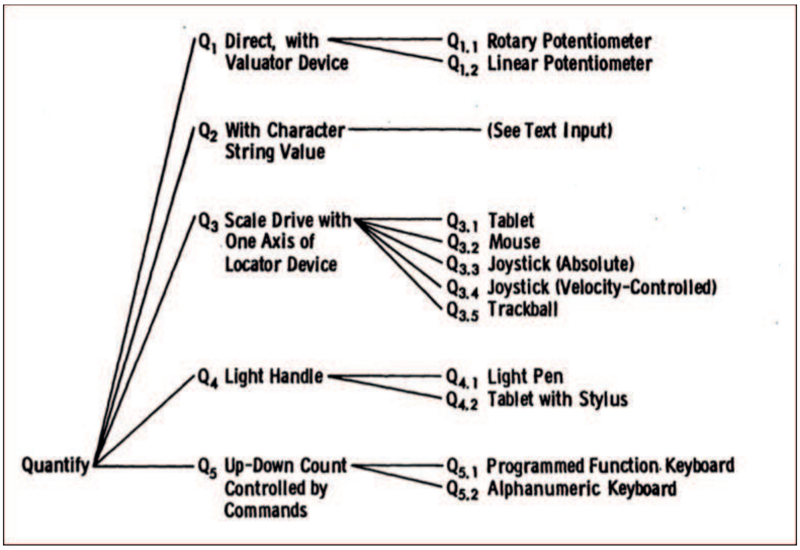


Figure 2.29 – Variety of quantifying techniques (Foley et al., 1984).

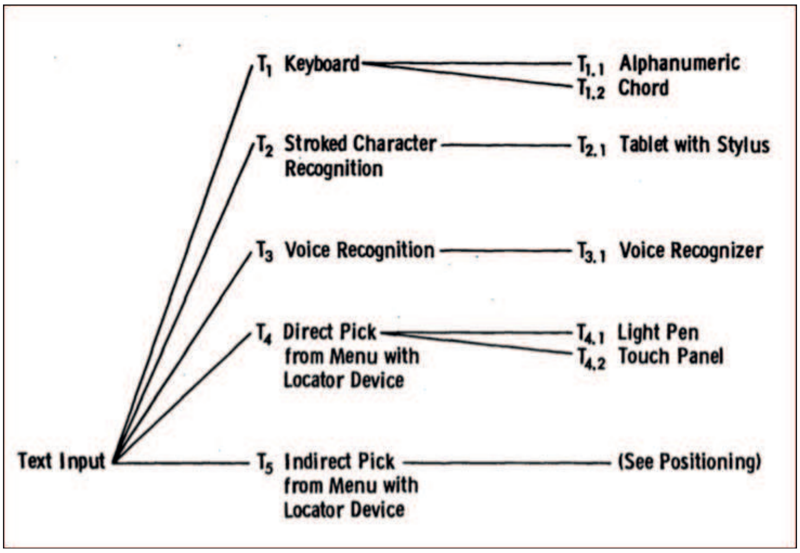


Figure 2.30 – Variety of text entry techniques (Foley et al., 1984).

focus on the concepts of multimodality rather than unimodality, intelligent adaptive interfaces rather than command/action based ones, and finally active rather than passive interfaces (Jaimes e Sebe, 2007). Note that these terms have been previously discussed in this Chapter.

As mentioned earlier, a system that is based on only one input modality is known as unimodal. Based on the nature of different modalities, they can be divided into three categories that have additional main research areas (Jaimes e Sebe, 2007):

- **Visual-based:**
  - Facial expression analysis
  - Body movement tracking (large-scale)
  - Gesture recognition
  - Gaze detection (eyes movement tracking)
- **Audio-based:**
  - Speech recognition
  - Auditory emotion analysis
  - Human-made noise/sign detections (gasp, sigh, laugh, cry, etc)
  - Musical interaction
- **Sensor-Based:**
  - Pen-based interaction
  - Mouse keyboard
  - Joysticks
  - Motion tracking sensors and digitizers
  - Haptic sensors
  - Pressure sensors
  - Taste/smell sensors

The visual-based human computer interaction is probably the most widespread area in HCI research. Furthermore, considering audio-based modalities, speech recognition has historically been the main focus of researchers (Rabiner e Juang, 1993). On the other hand, sensor-based modalities encompass input devices that can be very primitive (as the computer mouse) or very sophisticated (as sensors that recognize smell or taste).

According to (Grifoni, 2009), there are six types of possible cooperation between modalities, and they help define how a combination or fusion of modalities work together to convey information more effectively:

- **Equivalence:** data are presented in multiple ways but even so be interpreted as the same data.
- **Specialization:** a specific kind of information is always processed through the same modality.
- **Redundancy:** multiple modalities are able to process the same information.
- **Complementarity:** multiple modalities take separate information and merge it, turning it richer and more complete.
- **Transfer:** a modality produces information that another modality consumes.
- **Concurrency:** multiple modalities take in separate information that is not merged.

Indeed, designers should support the best modality or, combination of modalities, capable of coping with changing environments (for example, during office work vs. driving a car), but also understand that different techniques for interacting with the system may also be of importance, especially when dealing with different user profiles.

2.5 Concluding remarks

In this chapter, a quick review and analysis was made about the paradigm shift to natural and multimodal interfaces, and what changes that transition entails. The different interaction paradigms were here discussed with the purpose of understanding the interfaces' evolution process, and of grasping to what extent technology can take advantage of users' innate abilities, setting aside unknown or third-party elements needed for interaction and often inadequate to the human-computer interaction. Also, we have in the previous sections set forth a framework for the discussion and analysis of interaction modalities and different types of tasks.

Even considering new interaction paradigms that show several progresses regarding easiness of interaction, we should keep in mind their purpose and specific context to what they were developed for: "Along with considering the correct method of input for tasks, it's important for designers to take the context, use, and practicality of these systems into account" (Wigdor et al., 2009, pp. 2756). There are applications of which one may take better advantage if used according to one specific mode of interaction to the detriment of others. Buxton too refers to the importance of knowing what to use and how to proceed considering certain situations: "Everything is best for something and worst for something else" (Buxton, 2016). Even today there are applications in which the best mode of interaction may be the Command Line Interface.

Indeed, we should understand that the new interaction paradigms do not impose themselves as a way to replace every mode of interacting with the computer, but rather they intend to improve this interaction performance according to specific goals and contexts.

Thereby, we have witnessed several paradigm shifts throughout different technological advancements, being the main evolution the transition from a technology-centered to a human-centered design (Krippendorff, 1997).

3 Research Methodology

This chapter presents the strategy of the investigation considered for this study, the target-audience and interfaces elected, as well as the methods of evaluation and data collection used in the various phases of the research. We believe it is important to justify our methodology and explain our strategy, since it was a thorough research with unique stages of evaluation that required different approaches according to the conditions and tasks at hand.

3.1 Target-audience

For this investigation we intended to emphasize the distinct age groups and work with users that had noticeable differences concerning cognitive performance levels and, as such, we selected distinct groups. We find it important to describe the participants' attributes so early in the study, as the different phases of our research rely on the same stipulated criteria. We narrowed our sample to specific age brackets, contexts of technological use and computer proficiency in every research phase.

Indeed, we opted to study three different age brackets:

- **Children:** with ages varying from 8 to 12 years old, and attending the fifth and sixth year of basic education.
- **Young adults:** with ages varying from about 20 to 30 years old.
- **Older-adults:** with ages varying from about 45 to 60 years old.

We pondered these groups due to our previous experience and research in the literature considering: dexterity, levels of cognitive performance, and technological proficiency.

Firstly, we did not consider users of less than 8 years old because even though young children have proven to be capable of using the mouse to click within 3mm horizontally and 6mm vertically from the center of a 3mm target, not all movement procedures were suitable for them (Donker e Reitsma, 2007). Also, they are not so used to the computer and do not use it on a daily basis (for example, to do their homework). As such, they do not have the dexterity capabilities as improved as older children at this early stage, and therefore they are less prone to be able to master computer interfaces (Lane e Ziviani, 2010).

Indeed, studies indicate that age, visual-motor coordination or previous experience may all influence the successful use of the computer mouse by children, and older children are faster and commit fewer errors than younger ones (Donker e Reitsma, 2007; Lane e Ziviani, 2010). Also, improvements in speed and accuracy are noticed across the ages of four to twelve, although these differences become less significant at the age of eight, and there are no differences in performance between pointing and dragging tasks between eight and twelve year-old children (Joiner et al., 1998).

Moreover, these ages are encompassed by the fifth and sixth year of basic school, being these school years the ones that show greater variance when compared to previous ones in terms of reasoning by the children (de Lemos, 2006). Indeed, for this age group, cognitive maturation seems to have a stronger influence than age on the dexterity of using the mouse, and frequency of use of this input device is important for speed and accuracy (Lane e Ziviani, 2010). This variance in the cognitive process tends to significantly fade towards high school (de Lemos, 2006).

For the second age group considered in this study (young adults), we encompassed people of 20 up to 30 years old, being these users already at a different stage concerning cognitive performance when compared with the group of children (de Lemos, 2006). We restricted our sample to participants that were students and used the computer on a daily basis. As such, we selected subjects attending courses related to computer science: it was our purpose to limit the age brackets, but we did not intend to include people of other fields that would not require everyday computer use. This case helped us control our human sample in terms of digital literacy.

Finally, the group of older-adults consisted of active workers ranging from 45 to 60 years old. This group could present worse dexterity or consider gestures to be more challenging, but it was within range of our sample's characteristics. In previous studies, we acknowledged the fact that seniors did not approach technology with such ease, mainly users above 60 years old (Carvalho et al., 2012b,a). The ages that displayed more interest in technology and better dexterity in using natural interfaces were the ones between 45 to about 60 years old, and this was an important lead when selecting our third age group for this study. Nevertheless, they had to come from similar economical-financial backgrounds and be in touch with computer systems on a daily basis.

Having defined the three user groups that were the target of this investigation, we narrowed our sample within those groups according to specific requisites. We ascertained that all of the participants had the same level of computer proficiency and used the computer on a daily basis. Our intention was to choose users within the same range of experience with the computer mouse. Also, we were careful when narrowing our sample to users with similar economical-financial contexts and were thus surrounded by similar technological environments with the same access to technology.

We controlled the participants' characteristics by questionnaires organized specifically for this purpose at the beginning of each experiment, and restricted the users according to: (1) same level of digital literacy; (2) same daily experience with digital systems; (3) same economical-financial context; (4) same access to technology.

3.2 Input modalities

Scientific research lacks a systematic exploration of how children, young adults or older-adults interact with technology and the different input modalities currently available, whether when it comes to completion times, precision, or even preference. Indeed, it becomes important to analyze the various modalities and recognize which could be beneficial to study and join in a multimodal interface.

After a meticulous state-of-the art review concerning various interfaces commercially available nowadays and an analysis of their advantages and limitations (as described in the previous Chapter), we opted to study four: the graphical interface, the touch interface, the tangible interface and the gestural interface.

First and foremost, we chose to work with the graphical interface (computer mouse and keyboard) because it was our plan to understand if this was still suitable for any elemental or compound tasks in the current days. Also, we had interest in knowing if our target-audience would present similar performances with regard to this traditional interface that held a higher rate of proficiency amongst the participants.

We also chose touch as an interaction modality, since it has proven its benefits considering different contexts of use and fields of study, whether in education (Xiaohua Yu et al., 2010; Martinez-Maldonado et al., 2014; Lievens e Van Daele, 2015), health (McKee et al., 2015; Yu et al., 2015), working with users with special needs (Hourcade et al., 2012) or elderly users (Nawaz et al., 2014). This interface has been constantly undergoing significant advances, enabling the user to interact with the system without the need for learning and training with third-party devices. Indeed, this interaction modality may even encourage the use of technology by people excluded from the digital world and shows a wide range of possible applications and tasks that can be achieved. Thus, we felt that it was important to explore these benefits and understand if they are valid for different age groups.

Regarding the tangible interface, it would be noteworthy to analyze how different users from distinct age groups behave using physical objects that are able to couple digital information and consequently eliminate the conceptual gap from input/output of data (Ullmer e Ishii, 2001). We opted to follow the tangible tabletop interaction metaphor for our study, in which the user can manipulate the virtual data resorting to physical objects coupled with specific patterns. The potential already shown by this interface to embrace several objects to manipulate virtual data is an asset that can influence interaction and user preference.

Lastly, we considered gesture-based interfaces for this study because of their potential when it comes to human natural communication. Indeed, a lot of information transmitted amongst humans is passed through gestures (Galitz, 2007) and this interaction modality can thus be explored by different fields of study and regarding several purposes: attentive and immersive environments (Vertegaal, 2003; Pietinen et al., 2008), education (Lech e Kostek, 2010) and as alternative communication systems for people with disabilities or impairments (Reilly, 1998; Lee et al., 2005). Indeed, instead of relying on third-party devices to interact with the systems, gestural interfaces intend to incorporate every-day gestures into the own technology, providing the user with the choice of not using any other external devices and instead using his/her own body, thus decreasing the cognitive load necessary during interaction. This is an important feature to evaluate when considering different groups of users that should not have to be experts in one specific interface in order to interact with it.

On the other hand, we did not consider other interfaces, such as speech-based ones or haptic, on two accounts: because of their still existing limitations, and inappropriateness for everyday tasks. We acknowledge the convenience of, for instance, speech-based interfaces in certain contexts of use (e.g. driving), but it was not our goal to study such specific contexts. Instead, we opted to study input modalities that could be beneficial when interacting with the systems for a wider range of activities. The limitations and characteristics of each modality were further discussed in Chapter 2, where we presented the paradigm shift to natural user interfaces.

3.3 Evaluation approach

We were dealing, in this research, with very specific variables and conditions, each being dependent on the different stages of the investigation and according to each goal set for our work. We thus divided the research into three phases (Figure 3.1). Also, for each phase, we resorted to different evaluation methods (described below in Section 3.4), as no specific method is sufficient when it comes to input modalities and devices: if the conditions of the experiments change, the method may not lead to the same results (Zhai et al., 2004).

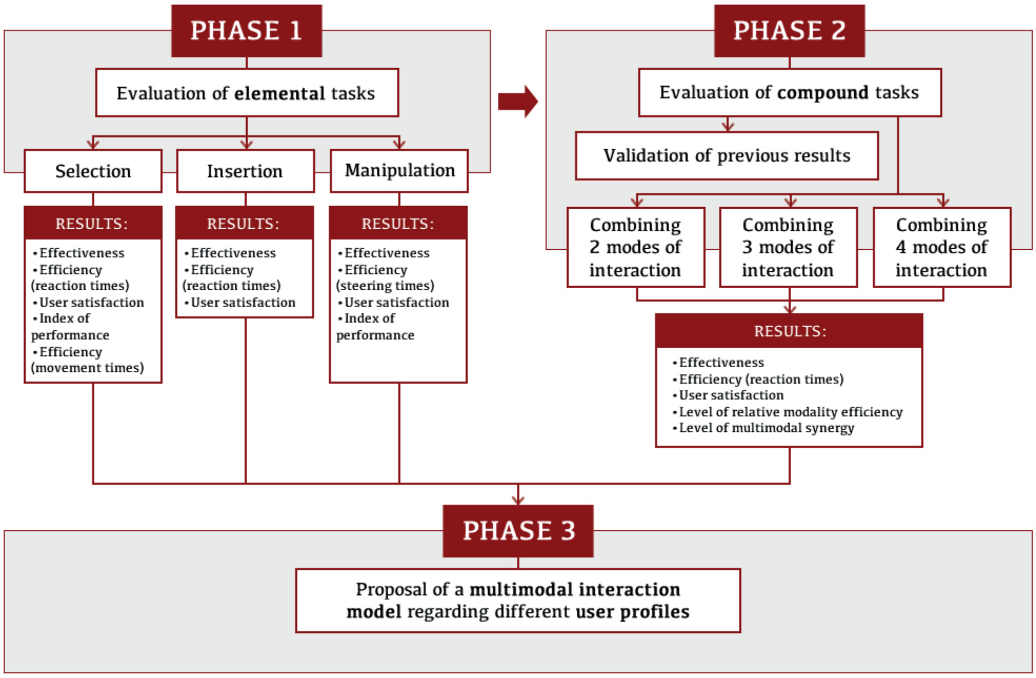


Figure 3.1 – Scheme of our investigation's strategy.

In the first phase, our intention was to evaluate how the three age groups performed primitive tasks, whether we are talking about data selection, insertion or manipulation, with resort to the different input modalities provided. Here, we wanted to understand: (1) how the users relate to the different interfaces individually in terms of satisfaction and behavior; (2) if the process of interaction is influenced by the users' age; and (3) which elemental interaction tasks may benefit from a specific interface to the detriment of others.

For that purpose, regarding the task of content selection: we performed a usability evaluation, collecting data on effectiveness, efficiency (reaction times) and user satisfaction; and we also evaluated their movement times and index of performance based on Fitts' law, a well-known human performance evaluation model. When assessing the insertion task, we based ourselves on a usability evaluation. Finally, when analyzing manipulation activities based on trajectory (e.g. drawing or navigating), we resorted to the Steering law and analyzed movement times of the different age groups and respective index of performance in two different steering tasks: one circular and one linear.

In the second phase, we intended to understand: (1) if the results gathered on the previous phase were also valid for compound tasks or, on the other hand, if a change in the type of task affected interaction; (2) if there were any synergies when combining more than one interface to perform compound tasks, namely when using two, three or four input modalities together in the same activity. Here, we resorted to several case studies, with different evaluation methods specifically dedicated to this type of interaction analysis: relative modality efficiency and multimodal synergy. Each of these assessment approaches studied a specific attribute regarding interaction. Also, we performed usability tests on each case study in order to fairly compare human performances.

Every result achieved in the first two phases was relevant for the last stage: the proposal of a multimodal interaction model based on different user age profiles. Here, we intended to suggest certain guidelines regarding which input modalities to use on which tasks, and which interfaces provide positive synergies for an improvement in interaction.

3.4 Evaluation methods

With the goal of outlining our methodological strategy during our research, we present in this Section the evaluation methods used in the investigation, as well

Table 3.1 – Characteristics of quantitative and qualitative methodology (Lor, 2011).

Characteristic	Quantitative	Qualitative
Metatheory	Positivist, Postpositivist	Interpretivist
Nature of reality	Singular, stable, independent of observer; external reality	Multifarious, culturally determined, socially constructed; holistic reality
Relation of investigator to what is studied	External, observing from outside; in artificial setting	In the study setting, observing from within; in real-life setting
Relation to social phenomenon	Neutral; empirical	Engaged; normative
Research aim	Nomothetic; hypothesis testing; generalizing	Idiographic; hypothesis generating; contextualizing
Strategies	Structured, theory-derived variables identified beforehand; controls; operationalization & measurement	Unstructured, open-ended, theory developed during research; concepts that are rich in meaning
Typical methods	Experiments, surveys	Participant observation, case studies
Criteria for judging research	Validity & reliability; objectivity	Credibility, transferability, dependability; authenticity

as the approach for data collection and resulting analysis. Throughout our study we resorted to a comparative research methodology and case study evaluations, resorting this strategy to both quantitative and qualitative variables.

Pickard (2013) supports that there are only two basic methodologies: quantitative and qualitative. The choice between these two is the highest level methodological decision. Table 3.1 describes the characteristics of these two distinct evaluation methods. Quantitative methods emphasize the formulation of hypotheses, operationalization of concepts, measurement (a metaphor derived from the physical sciences), the design of experiments or surveys, sampling, and the statistical testing of hypotheses (Lor, 2011). On the other hand, qualitative methods are committed to the naturalistic perspective, and to the interpretive understanding of human experience (ethnographic methodology).

There is a huge literature discussing the pros and cons of quantitative versus qualitative approaches, but this division is becoming unimportant and even excessive, as the acceptance of "methodological pluralism", i.e. mixed methods in comparative studies has been growing (Hantrais, 2009). Indeed, Hantrais (2009) discusses that two or more different research strategies may be used to investigate the same phenomenon. When quantitative and qualitative approaches are used in parallel, findings or insights from one strategy can be corroborated by the other.

As we intended to pursue a method that enabled us to complement our research and provide us with evidence of differences between audiences regarding the technology use, but also study characteristics within each targeted group, we resorted to a multidisciplinary research approach. On one hand, we intended to analyze the relation of the participants of each group with each interface thoroughly, which meant that we resorted to a case study evaluation, focusing on one group and understanding if there was any relation between the variables, and not making any comparisons between age groups. This was important for us to grasp which interface could help the participant during interaction, or otherwise harm the process. The case study as an evaluation method allows us to observe an individual or group's characteristics in order to understand the context of the individual, organizational, social or political phenomena (Robert K. Yin, 2009). Also, the case study requires a systematic, detailed, intensive, thorough and interactive approach.

However, aside from understanding the characteristics and preferences of interaction within each specific age group elected for our research, we also sough out the differences in interaction between groups. Our cases are similar in some respects (otherwise, it would not be meaningful to compare them), e.g. the interface used and the task at hand, but they differ in some respects, e.g. age group. These differences become the focus of examination. The goal is to find if the cases are different, and try to reveal the variable that allows such a variation, if any exists. Indeed, comparison is inherent in all science. It can help the researcher ascending from an initial level of exploratory case studies to more advanced levels of general theoretical models.

Thus, in order to analyze the collected variables (both quantitative and qualitative), we chose a comparative research strategy when considering the comparison between age groups and their results concerning interaction with the different interfaces.

Each evaluation method used in this research is explained below. The approach used in our study and how we analyzed the results considering the various methods will be also explained in each case study.

3.4.1 Usability evaluation

Usability testing is one of the most used evaluation methods, having a large impact on product improvement (Rosenbaum et al., 2000). It defines itself as "the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments", according to the norm 9241-11 of the International Organization for Standardization (ISO): "Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts" (ISO, 2015b).

Thus, usability refers to the resultant measurements of a certain task and the process used to obtain these values (effectiveness, efficiency and satisfaction). Indeed, effectiveness indicates how the users achieve specific goals in particular environments and is evaluated by the accuracy and completeness of the task's results. For its part, efficiency measures the resources expended during interaction to achieve the goals, in relation to the number of errors committed and successful completion of the task. Finally, user satisfaction comprehends the subjective phenomena during interaction and is the most difficult measure of a usability evaluation: it refers to the comfort and acceptability of the users towards the system/interface.

Usability has multiple components and applies to all aspects of a system with which the user interacts, including installation and maintenance procedures. Based on the measures of usability evaluation previously discussed, there are five attributes that determine the systems' good use and their usefulness (Nielsen, 1993). By usefulness one refers to the aptitude of the system regarding both utility - if the functionality of the system can do what is needed and meant to do - and usability - how well users can perform and use that functionality (Grudin, 1992). Indeed, usability is traditionally associated with the following parameters (Nielsen, 1993).

- **Learnability:** The user should be able to learn the system with ease so that he/she can get rapidly started using it.
- **Efficiency:** Once the user is familiarized with the system, he/she can achieve a higher level of productivity.
- **Memorability:** The user should be able to remember how to use the system

without effort, so that he/she does not need to learn it all over again after longer periods of inactivity.

- **Errors:** The system must simplify interaction so that users do not incur in errors and, if so, they can easily and quickly recover from them.

- **Satisfaction:** The system should provide a pleasant experience, so that users like using it and are subjectively satisfied with it.

Evaluating usability can be considered a key point of user-centered technologies. The term evaluation refers to the process of gathering all data about usability of a system, design or software, relative to a group of users for a specific task in a specific environment (Preece et al., 1995), in order to look for any problems. On its turn, a usability problem can be considered as anything that interferes with the user's ability to complete the task at hand efficiently and effectively (Karat et al., 1992).

Overall, there are two methods for usability evaluation: user tests and inspection (Neto e Campos, 2014). The first is optimal to identify usability problems, although being very expensive and time-consuming for it requires gathering users to test the systems is specific and controlled physical locations. Usability inspection can be achieved through heuristic evaluations on the system performed by experts, often used because it is cheaper and easy to be executed. Here, expert evaluators systematically inspect and evaluate an interface according to specific usability principles, or heuristics.

Dumas e Fox (2008) attest that valid usability tests have six main characteristics to take into account:

- **The focus is on usability:** An apparently obvious trait of the evaluation test, but indeed important to emphasize, as the purpose of the test should provide an opportunity for the users to actually evaluate the product itself and their experience, and not answer questions relevant for fields of study other than that of usability evaluation (e.g. marketing or promotion issues, sales, etc).

- **The participants are end users or potential end users:** For a test to be valid, the participants must be part of the target market for the product, and not random evaluators, as testing other populations may prevent the test to be generalized to the target population. As such, it is important to find people who are potential candidates for the test, tracing a proper user profile (Branaghan, 1997), or a persona (Pruitt e Adlin, 2005), to capture the target-audience characteristics that are relevant for the tests.

- **There is a product or system to evaluate:** Tests can be performed with almost any kind of product or system, namely user interfaces, software, consumers, components that are embedded or accompany a product, hardware, etc.

- **The participants perform tasks, usually while thinking aloud:** This step is part of the execution phase of the test, where participants interact with the product and data are collected. Here, the test administrator instructs the user about the test and the procedures, and assists the user with the tasks, that should be selected for being frequently used or critical for the entire interaction process.

- **The data are recorded and analyzed:** The test should measure different variables and generate results that help the system improve in terms of error avoidance, task times, statements of frustration, and the need for assistance from the moderator. The process of collecting and combining different measures is called "triangulation" (Dumas e Redish, 1999).

- **The results of the test are communicated to appropriate audiences:** Recently, it has become common for the tests to be reported more informally, e.g. with the resort to video highlighting the main issues, facilitating the communication between teams of development. Also, more formal reports are also important to justify the tests.

Throughout the usability evaluation process it is extremely important for diagnosis purposes to record and take into account the commentaries of the participants with the “thinking aloud” strategy. Thinking aloud has been used in psychological research and it is considered the best cognitive method for studying short-term memory (Ericsson e Simon, 1993). This method is relevant for our research and, as such, we took advantage of it during the tests, always trying to analyze and annotate the participants’ attitudes and commentaries.

Thus, when considering the evaluation method there are certain aspects to be aware of: the physical environment, the evaluator and the user (Bowman et al., 2002). The environment refers to aspects used during the test and the procedures for evaluation: the different input/output devices used during interaction, the position and body movements of the user, or the evaluation practices. On the other hand, the evaluator (or evaluators) should participate in the test as to provide information and help to the user considering the overview of the test and detailed instructions of the process. The user also needs to be taken into account during the test’s preparation, following the user profile outlined and considering the users’ level of frustration, tiredness and mood that could occur during the trials.

3.4.2 Fitts’ law

It is naïve to try to measure input performance merely by analyzing user’s task completion times and the average of all trials, including, at times, the number of errors that took place; that is to say, measure input performance only through usability evaluation (Zhai, 2004). If the circumstances of the experiments change (such as target sizes or distances in the screen) the method may not lead to the same results. Thus, the generalization of those results is not valid.

This scenario is relevant for Fitts’ law (Fitts, 1954). This original experimental paradigm aims at measuring and characterizing the performance of various input devices, regarding different levels of difficulty and movement times. Indeed, it measures the rate of transmission in “bits per second” of a pre-established movement with different indices of difficulty. This human performance model has been widely applied to describe the relationship between movement time, distance and accuracy within specific pointing tasks, as a way to formally evaluate pointing devices (MacKenzie,

1992) and compare their efficiency (Soukoreff e MacKenzie, 2004). Also, ISO 9241-9 “Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices” (ISO, 2000, 2007) provides a standard approach for input device evaluation based on Fitts’ law.

Fitts’ law has been used for target acquisition performance evaluation and optimization, mainly in two ways: as a predictive model and as a speed-accuracy tradeoff model. The concept of the predictive model can be explained as a way to predict the time a user takes to move the cursor from an initial position to a target, like a button. This can be applied to user interface design, e.g. interface’s preference measurement (Toomim et al., 2011), difference in target-icon sizes depending on the movements’ distance (Newell e Card, 1985); or even in real-world situations (Rohs et al., 2011).

On the other hand, Fitts’ performance model can also be applied when comparing and evaluating pointing devices. Task differences, selection techniques, response irregularity and test conditions applied may cause experimental variations. Understanding these variables increases the validity of across-studies comparisons regarding input techniques (MacKenzie, 1992). In this sense, researchers use this model to measure multiple movement times and then determine how the different conditions or devices affect interaction, that is, the performance of the user or device. The combination of several measurements of movement time throughout the tasks into a single statistic value yields an Index of Performance (IP), or throughput, of the pointing device. This index comprises both speed and accuracy in a pointing task regulated by specific parameters: Fitts’ Index of Difficulty (ID).

Fitts’ task evaluation paradigm and variations:

Researchers in the HCI field use Fitts’ Law as a way to evaluate the time a user takes to carry out a movement task, according to different levels of difficulty. The movement tasks’ difficulty, also known as Index of Difficulty (ID), can be quantified by the metric “bits” and is calculated by using the value of the distance, or amplitude (A), between two specific targets and their width (W) or tolerance area (as shown in Figure 3.2):

$$ID = log_2(A/W + 1) \tag{3.1}$$

Although not in it's original form, this formulation is a more stable model for the Fitts' law suggested by MacKenzie (1992). Also, the index of difficulty can be obtained by varying the values of the width of the targets and their distance, providing thus a range of task difficulties.

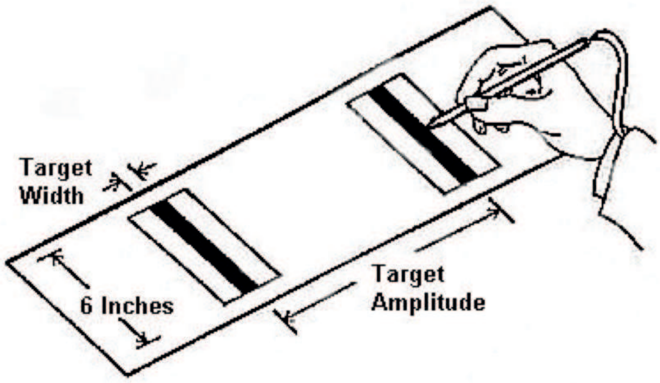


Figure 3.2 – Fitts' task paradigm (MacKenzie, 1992).

This index of difficulty is important for calculating the time it takes for a user to complete the task. As supported by Zhai et al. (2004), the more accurately the task at hand must be accomplished, the longer it takes for the user to perform it, and vice versa: movement time for hard tasks is longer than for easy tasks (see Figure 3.3). The movement time taken to hit the target is predicted using a linear equation, where movement time (MT) is a linear function of ID. Therefore, a movement model based on Fitts' law is an equation predicting movement time from a task's index of difficulty (MacKenzie, 1995):

$$MT = a + b \times ID \tag{3.2}$$

Or

$$MT = a + blog_2(A/W + 1) \tag{3.3}$$

The empirical constants “a” (intercept) and “b” (slope) are found using a regression analysis on the movement time data. In addition, the constant “a” has proven to be significant since it can be affected by the learning curve of the input device and the task (Card et al., 1991). For example, in target acquisition tasks involving selection with an extra button push, additional pressure or waiting time, there is an additive factor to the index of difficulty, which contributes to the increase of the intercept of the regression line, but not the slope (MacKenzie, 1992). As movement time is measured in seconds and the index of difficulty in bits, the slope of the line is in “seconds/bit”. Thus, the reciprocal of the slope is in “bits/s”. This measure is considered the rate of transmission known as Index of Performance (or Throughput - TP). Since the Index of Performance is in bits/s, the term “bandwidth” is also considered. The higher the value of the throughput the higher the rate of human performance, given that more information is being managed per unit time (MacKenzie, 1995).

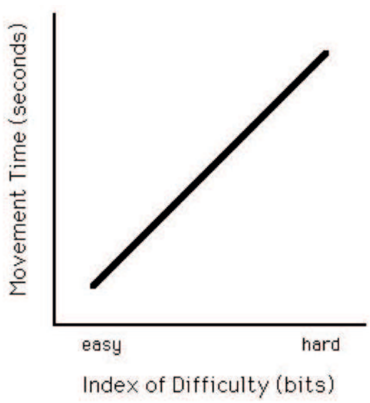


Figure 3.3 – Movement time prediction (MacKenzie, 1995).

Refinements to extend the model to 2 dimensions:

The original experiments performed according to Fitts’ law were considered only for movements along one axis: x. Thus, since both target amplitude and width were defined along the same axis, the model reflects merely one-dimensional experiments.

With the advent of new input devices and interaction modes, researchers must consider a performance model that can mirror results on 2 and eventually 3 dimensions. Likewise, it is critical to also consider the width as well as the height of the target, when it is not a circle or square-shaped object. Hence, different formulations to the Fitts' task paradigm arose, considering both the shape of the target and the angle of approach. MacKenzie (MacKenzie, 1995) noted distinct strategies to correct these problems, as for example: a substitution of the target width for the smaller value of either its height or width, known as "Smaller-of model"; or a constraint in the model by measuring the width of the target along the approach angle and thus allowing a 1D interpretation of a 2D task.

Device and task differences:

Fitts' law considers spatial variability, which means that the target width itself may be adjusted based on the distribution of the actual target selections. This approach may reduce the Index of Performance, as users usually take longer to complete the task when they have to be more accurate, and thus demonstrates less endpoint variability.

This performance model measures the movement time (MT) a user takes to cover the distance between two targets, always contemplating a width of error tolerance. Since this model is meant to evaluate the performance of several input devices, the only disparities that should occur would be the between-device ones (MacKenzie, 1992). However, this is an idealistic scenario. In this regard, the sources of variation end up being considered as factors during experiments, since studies tend to compare different input devices using within-study rankings and not absolute measures in cross-studies. Amongst the differences that can influence experiments' results are the characteristics of the devices, such as properties or number of spatial dimensions sensed (MacKenzie, 1992).

All the same, the user cannot perform a generic task that accommodates adjustments for the devices' disparities. This task should consider these different device factors.

Hence, the tasks might follow two different strategies: as continuous or discrete tasks. Discrete tasks tend to be more appropriate for devices that provide direct manipulation (MacKenzie, 1992). This type of task requires the user to perform an action at a time, stopping when the movement ends, and the time is measured between the start and end points. On the other hand, if the user resumes the movement without stopping and performs another continued action, as starting from one point and having to select more than one target uninterruptedly with different amplitudes between them, then this is considered a continuous task. Here, the total time taken during the exercise is divided by the number of movements in order to determine the movement time (MT) for the specific target sizes and amplitudes.

3.4.3 Steering law

Although being the best known performance model and the most frequently used to make empirical comparisons between input devices, Fitts' law lacks the capability to generalize the task parameters regarding the devices when it comes to tasks other than just target acquisition or pointing. Indeed, capabilities such as drawing, writing, navigating through menus, fall short of such analysis.

Choosing the input modalities for interactive systems that best suit the users' needs is of utmost importance, but for that we must consider the rich dimensionality of input devices and go beyond target acquisition evaluations alone. In particular, during interaction with the system the user frequently needs to perform trajectory-based tasks, i.e. produce movement trajectories, and not just point to targets.

In that sense, (Accot e Zhai, 1999) proposed a speed-accuracy trade-off model for trajectory tasks, and thus overcoming the lack of trajectory-based testing paradigms. These tasks had been studied before in various contexts (Rashevsky, 1960; Drury, 1971). The action of steering was proposed, and it rests on moving along curves and through narrow tunnels in order to evaluate the index of performance in steering, and therefore mirroring the Fitts' law index of performance for pointing. Regarding the various input devices, the steering law can be applied as an evaluation paradigm complementary to Fitts' law (Accot e Zhai, 2001).

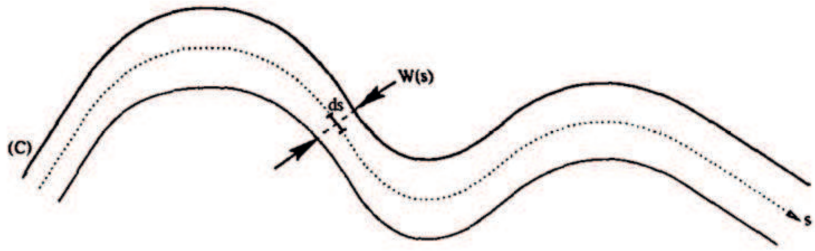


Figure 3.4 – Steering along a curve (Accot e Zhai, 1999, pp.467).

A steering task is based on moving the cursor along a constrained trajectory, without touching or leaving its boundaries (Figure 3.4). For a generic path C that allows the calculation of steering difficulty for a wide range of tunnel shapes (straight, narrowing and spiral paths), the steering law relates the completion time T to the task parameters and can be expressed in the following equation:

$$T_C = a + b \times ID_C \tag{3.4}$$

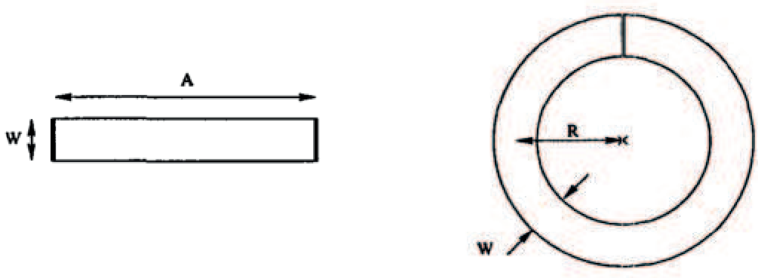
Here, a and b are constants; also, ID_C is the index of difficulty of the task (the same definition as in the Fitts' law), and is found by integrating the inverse of the path width along the trajectory. The integration variable s stands for the curvilinear abscissa and W(s) for the path width at abscissa s:

$$ID_C = \int_C \frac{ds}{W(s)} \tag{3.5}$$

However, analogous to the Fitts' law, the steering law also has its limitations: the independent variables in both performance metrics are the same. Indeed, the path length and width are controlled and thus the movements that can be performed are restrained for evaluation's validity purposes. As such, the own shape of the path can affect performance and therefore it must be controlled, as it is not possible to generalize the performance prediction across different shapes of paths. Regarding the different movements a user can produce when performing a manipulation activity (e.g. drawing or writing), the most popular one is navigation (Figure 3.5).

Figure 3.5 – A sequence of straight steering movements during navigation (Accot e Zhai, 1999, pp. 468).

Hence, the steering law focuses mainly on two types of paths/tunnels that are most popular, in order to maintain its validity across multiple devices and evaluate both linear and non-linear movements: straight tunnels and circular tunnels (Figure 3.6). The first allows the user to interact with the system with linear motions that are mostly found in everyday tasks, such as moving or dragging an object from one place to another. The latter resembles curved trajectories that require more motor coordination. Here, as the radius of the circle changes, so does the curvature of the path, providing a wider range of trajectories and paths' variety.



a) b)

Figure 3.6 – Two steering paths (Accot e Zhai, 1999, pp. 468): (a) Straight tunnel steering; (b) Circle tunnel steering

According to the equations presented above, the calculations can be simplified for both tunnel shapes if the parameters of width and amplitude (distance between the initial and final positions) are preserved. Therefore, for straight paths, the equations can be explained as:

$$T = a + bA/W \tag{3.6}$$

Here, the A stands for the amplitude (length) of the tunnel, while W is its width. Through the same procedure, the steering law equation for circle paths can take the form of the following equation, where the movement amplitude A is equal to the circle circumference 2π R, and R is the circle radius:

$$T = a + b2\pi R/W$$

Indeed, the simplicity of these two steering paths shows that the tasks can be considered representative for standard evaluation input devices in trajectory-based movements during interaction.

3.4.4 Relative modality efficiency and multimodal synergy

A fundamental issue that has to be considered when developing multimodal interfaces is the suitability of the different input methods for the tasks at hand (Bernsen e Dybkjaer, 1998). Also, merging multiple modalities efficiently and determining the appropriate modality combination is complex, such as the issue of multimodal turn-taking (switching between modalities) and the selection of the most efficient input device (Perakakis e Potamianos, 2008a).

Traditionally, the evaluation of multimodal systems is based on objective metrics such as speed, number of errors and task completion, along with specific subjective metrics (Hone e Graham, 2000), mainly grounded on surveys and user opinions (Dumas e Fox, 2008), that are then statistically analyzed (Mason et al., 2003; Myers et al., 2010). However, there are various difficulties and issues in applying these methods (Larsen, 2003; Perakakis e Potamianos, 2008a). Also, most of these traditional objective metrics undergo a poor correlation with subjective usability metrics. The goal in evaluating multimodal interfaces should go beyond traditional metrics and propose methods to assess unimodal efficiency, input modality usage, inactivity times, and exploit synergies that affect the multimodal system performance. Each of these concepts will be approached below.

Indeed, the evaluation of multimodal systems should also encounter two other principals as to overcome the difficulties in combining the different interfaces: relative modality efficiency and multimodal synergy (Perakakis e Potamianos, 2008b). These objective metrics are meant to provide information and help identify usability problems during interaction.

Relative modality efficiency:

Evaluating the relative modality efficiency of an interface in comparison to others helps identify underutilization of a specific input modality due to poor interface design or data communication asymmetries (Perakakis e Potamianos, 2008b). This metric only makes sense when based on the modality usage during interaction, that is, the amount of time a user is actually making use of a specific input to the detriment of the other, whether in the number of turns or total duration. This analysis dictates the actual contribution of that input modality to the usability of the system.

Basically, relative modality efficiency calculates the amount of information communicated in unit time for each modality or, as in the previous performance evaluation paradigms (e.g. Fitts' law), the information bandwidth. As such, this metric should correlate well with the relative modality usage unless there is information asymmetry between the user and the system: the modality efficiency should be proportional to the inverse of the time required by that modality to complete a task, that is, efficiency is proportional to the number of attributes (or actions) correctly completed in unit time (the information bandwidth of each modality).

For example, when considering two input modalities in one same multimodal system, the overall time spent using both (T_1 and T_2) is associated to the amount of correctly completed attributes (N_1 and N_2) using each, respectively. Thus, relative modality efficiency (E) is defined as:

$$E = N_1T_2/(N_1T_2 + T_1N_2) \tag{3.8}$$

On the other hand, relative modality usage (U) is defined as the percent of time sent using one specific modality over the total interaction time:

$$U = T_1/T_1 + T_2 \tag{3.9}$$

When considering an evaluation uniquely based on efficiency aspects, the ratio of modality efficiency and usage (E / U) should be approximately 1. In this sense, ratios >1 means an underuse of the input modality during interaction, while <1 means overuse. These equations may be used and adapted for all available modalities of the system.

Also of interest to this evaluation metric are inactivity and interaction times (Perakakis e Potamianos, 2008a). Indeed, it is important to separate the input efficiency (related to actual interaction times) from the output efficiency and cognitive load (related to inactivity times).

The input efficiency is defined as being inversely proportional to task duration, so if the activity takes a lot of time to be completed, it must be analyzed if the amount of time taken was interface-related or, on the other hand, if the user was not responding for different reasons other than the use of the input device, such as inactivity time. This time refers to the time the user appears inactive to the system, i.e. the idle time interval that starts at the beginning of the activity, and finishes on the moment the user actually interacts with the system using one of the devices. During this interval, the user must to comprehend the system's response and respond accordingly (Figure 3.7). This period is also known as output efficiency or cognitive load time, as it represents the time the user needs to apprehend the output of the system and react. Whereas, interaction time is related to the actual time the user is interacting with the system and completing the activities. Also, it could be important to analyze which input modality leads to superior cognitive loads and, thus, higher inactivity times (Perakakis e Potamianos, 2008b).

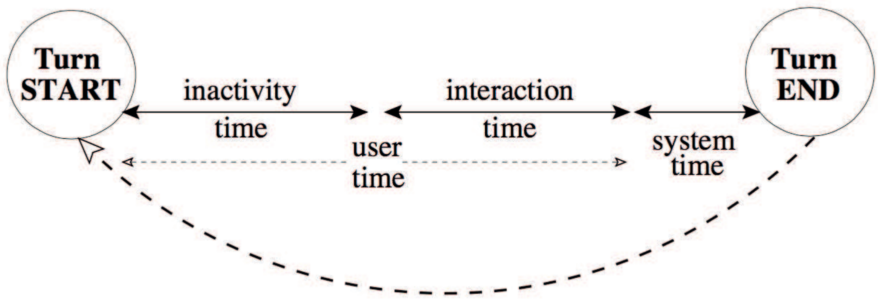


Figure 3.7 – Inactivity versus interaction times when using an interface (adapted from (Perakakis e Potamianos, 2008b)).

Multimodal synergy:

Multimodal synergy measures the added value from an efficient combination of multiple interaction modalities and can thus be used as a measure of the quality of modality fusion in a multimodal system.

Indeed, synergy is a design principle that applies to multimodal systems that support more than one input/output modality and should be considered as another step in the evaluation of such systems (Perakakis e Potamianos, 2008b). The principal relates to the concept that a synergistic multimodal interface is better than the sum of its parts, that is, unimodal interfaces (Bilici et al., 2000).

This metric is meant to translate into a number the percent of interface efficiency improvement in terms of time-to-completion achieved by the multimodal system when compared to the average of the unimodal interface efficiency (Perakakis e Potamianos, 2008b). Thus, multimodal synergy can help identify problems in effectively combining various input modalities in one same system, consequently creating more synergistic interfaces.

When calculating the multimodal synergy of a system, the time-to-completion and relative usage of each modality must be accounted for. For example, D_1, D_2 and D_m are the times taken by two separate interfaces to complete the task, and the time of the respective multimodal system that combines both. Also, U_1 and U_2 are the relative usage of each separate interface (normalized in [0,1] and summing to 1, as explained previously). Then, the time-to-completion of the multimodal system D_r is:

$$D_r = U_1 D_1 + U_2 D_2 \tag{3.10}$$

Or, where i sums over all of the available modalities (more than two):

$$D_r = \sum_i U_i D_i \tag{3.11}$$

Modality synergy S_m for a multimodal system m, where i sums over all input modalities and corresponding unimodal systems, is defined as:

$$S_m = 1 - D_m / \sum_i U_i D_i \tag{3.12}$$

Furthermore, these equations take into account output synergies, which means that the system supports visual feedback or graphical output to help the user during interaction.

It is also important to acknowledge that multimodal synergy expresses the relative improvement in terms of total time to complete the task achieved by multimodal interfaces over the sum of its unimodal parts, and that synergy may also be negative.

Indeed, when a multimodal system combines input modalities inefficiently, it does not exploit synergies well, becoming difficult to use and inducing high cognitive loads. Thus, the result may be a negative multimodal synergy. In these situations, the interface may become over complex to use and require changes to improve the modality efficiency.

3.5 Data collection and analysis

We resorted to several strategies for data collection: questionnaires, interviews, non-participant direct observation, photographic records, a field diary and error maps. Aside from photographs taken during the activities, with the express statement of consent of all the participants, the sessions were documented through annotations as well.

In the beginning of the experiment, the participants were asked to complete a questionnaire about their previous experience with the different interfaces used in the different activities. Our intention was to select and control our sample group in order to be able to generalize results without the target-audience being incongruent. Therefore we could have certainty that our sample was consistent and valid for comparison and analysis.

The several studies were monitored at all times by the investigators resorting to direct and non-participant observation, with the intent of systematically collecting data on qualitative variables. Regarding the qualitative data, our main goal was to be able to apprehend the participants' behaviors, reactions, feelings and even commentaries or opinions, and thus we resorted to the creation of field diaries for each phase of the study. During the tests, the researcher would also take photographs using a camera that was set before the experiment began, so the participant would not be disturbed and distracted by the photographic camera. Regarding quantitative analysis, all the values were recorded during the tasks by the computer, being the results organized into a participant ID for further statistical evaluation.

After the tests were concluded, enquiries were made to gather opinions on the quality of the interaction and their user experience. Therefore, aside from the discrete data gathered throughout the tests by the system, we also requested the fulfillment of a questionnaire at the end of each test, with closed-end questions that relied on qualitative Likert Scales (Likert, 1932) and ranking lists to understand, for all the input modalities: the ease of use, ease of learning, fatigue effect, naturalism of interaction, level of user comfort/frustration, user's degree of presence and concentration, and also participants' preference. The questionnaires' layout can be found in the appendix of this document. Also, an informal interview would also take place, in order for the researcher to take notes on their opinion and compare them to the questionnaires to validate their responses.

For the last phase of our study (the analysis of compound tasks and multimodal activities), we resorted to a visualization method: error maps. These are graphical representations of data and are used to illustrate specific areas of interest to analyze patterns in the users' performance, resorting to colors and spatial locations. Indeed, we wanted to understand not only how many errors the participants would cause, but also where exactly they occurred in the display. This could help us recognize the patterns of interaction and what paths or techniques to avoid in the interface design.

Finally, we analyzed all data accordingly, resorting to descriptive statistics where we consider discrete values that the system would record and describe measures that are commonly used, such as central tendency and measures of variability or dispersion: mean values, standard deviation (or variance), and the minimum and maximum values of the variables. Also, we performed the appropriate statistical analysis of the values in the software package "IBM SPSS Statistics" and conclude their significance using different methods depending on the normalization of the data. When a presence of outliers was noticed on the recorded times, the results were removed following the outlier-labeling rule (Beyer, 1981; Hoaglin e Iglewicz, 1987), in order to prevent error rates and substantial distortion of estimates.

3.6 Concluding remarks

In our study, every phase presented different quantitative and qualitative dependent variables for evaluation due to the intricacy of the different activities. In that regard, tests were conducted following specific methods of assessment well-known in the literature: usability evaluation, Fitts' law, steering law, relative modality efficiency and multimodal synergy.

Also, due to the complex approach of this investigation, we also stipulated strategies for data collection: questionnaires, non-participant direct observation, photographic records, and a field diary. Please note also that the researcher was always present at the experiments, but acted only as a non-participant observer and thus did not help the participants during the tests.

All tests were performed in controlled environments familiar to the participants, in order to make them feel comfortable and at ease. Also, all subjects were volunteers that accomplished the tests with their express consent and conscious of all the conditions of the experiment (the consent form is in the appendix).

4 Evaluation of elemental tasks

The aim of HCI is to enable the universal access of interactive systems for all users regardless of the diversity of the group, their tasks, or their context of use (Stephanidis, 2001). And it has been recognized the importance of adapting the input method and presentation format according to different devices, as it can affect the user's perception and attitude when dealing with specific tasks (Jonsson et al., 2004; Sodnik et al., 2008). Thereby, it becomes vital to grasp which elemental interaction tasks may benefit from a specific interface to the detriment of others, whether we are talking about data selection, insertion or manipulation (Foley et al., 1984). For each of these elemental user interface transactions it is essential to be aware of how the various modes of interaction can enhance or, quite on the contrary, impair the user's performance taking into account their profiles, the context, use and convenience of the interaction techniques (Wigdor et al., 2009).

This study is particularly motivated by the absence of scientific data that systematically evaluates which interaction mode could be the most appropriate for users with age-related differences when performing elemental tasks. Taking into account our audience attributes and characteristics, we determined a sample for the first phase of this study according to the directives stipulated on Chapter 3. For the first group we chose children attending the basic school and worked with three different schools in the city of Vila Real: "Colégio Moderno de São José", "Escola Monsenhor Jerónimo do Amaral", and"Centro de Estudos Super-Heróis". Regarding the group of young adults, we restricted our sample to participants attending courses related to computer science in "Universidade de Trás-os-Montes e Alto Douro". It was our purpose to limit the age brackets, but we did not intend to include people of other fields that would not require everyday computer use. This case helped us control our human sample in terms of digital literacy. Finally, for the third group, we chose older-adults that would work with the computer mouse as the primary professional activity and thus contemplated workers of secretariat departments of two different schools in the city of Vila Real: "Escola Monsenhor Jerónimo do Amaral" and "Escola Secundária Morgado de Mateus".

This chapter describes a study that was carried out in order to determine which input modalities are better suited for which elemental tasks (selection, insertion and manipulation) in a specific context, regarding the three groups of users considered throughout this study (previously described in Chapter 3). Part of the results, more precisely the case studies regarding selection and insertion tasks, was published in (Carvalho et al., 2014, 2015a,b), and accepted for publication in (Carvalho et al., 2016).

4.1 Evaluation of selection tasks

With this study we intend to understand: (1) how the users relate to the different interfaces individually, and which yields better results when it comes to usability; and (2) if the process of interaction is influenced by the users' age, that is, if age has an impact in the usability results. Our goal is to compare results between age groups, as well as between interfaces.

We thus divided the experiment into two phases: for phase one we adopted a methodology based on usability testing in terms of speed, error rate and user preference; phase two consisted on exploring target acquisition performances, resorting to Fitts' law.

4.1.1 Usability evaluation

Considering the scope of our study, we intended to compare four different interaction modes (the computer mouse, touch, tangibles and gestures) in terms of speed, error rate and user preference, and thus a usability analysis was adopted.

4.1.1.1 Materials and Methods

For the usability analysis, we resorted to discrete selection tasks, where a user begins and ends only one task independently, and does not have a series of continuous actions. Notice, also, that we have mentioned this type of discrete task on Chapter 2.

Apparatus:

The experiment was conducted in a closed room with artificial light and the tests were performed in a specific setup assembled for the purpose of this research (Figure 4.1). The system consisted on a 22” touchscreen placed in front of the user, with a resolution of 1280x800 pixels; a Microsoft Kinect sensor mounted on a tripod behind the screen, about 25 cm above it and facing the user; a computer mouse; a webcam; and 8 tangible pieces located in front of the screen, following a specific order and with even spaces in-between them (Figure 4.2).

Materials:

A purpose-built application was created with support for all modes of interaction, in order to keep the testing environment as coherent as possible. The tasks’ software was developed in Python with the support of: the Kivy Framework, the open-source computer-vision Framework reacTIVision (to track the fiducials) and TuioKinect (to track the hand gestures).

In order to reduce the impact of the system’s feedback across the different input

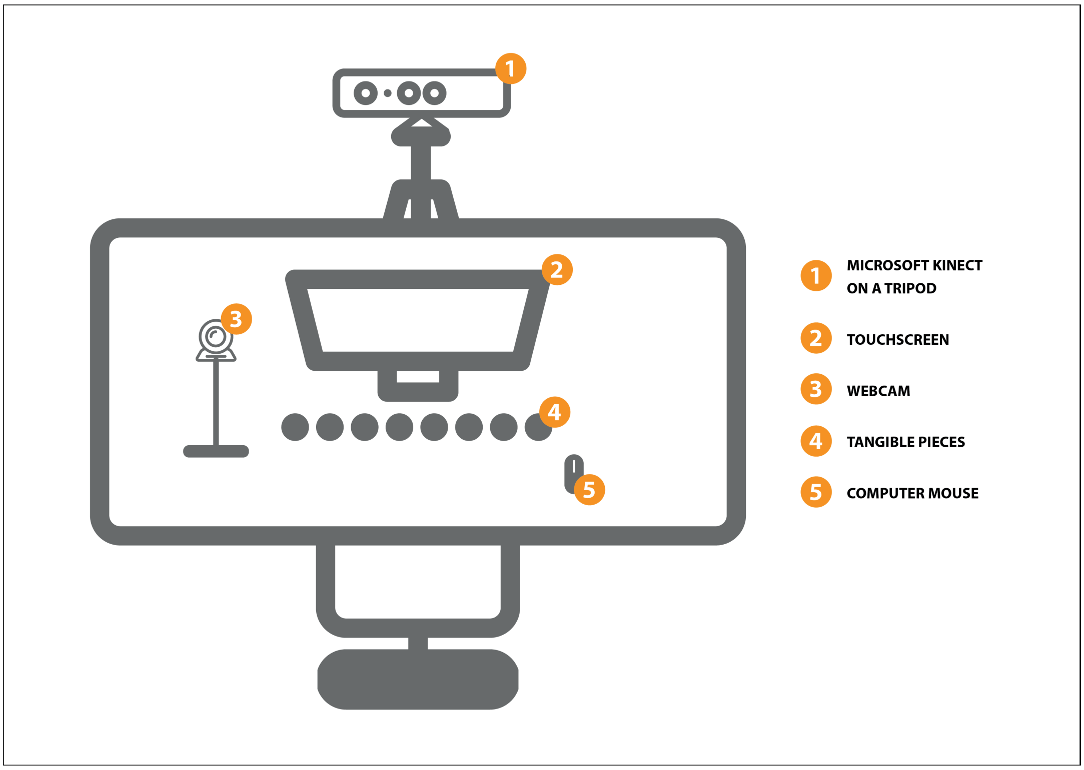


Figure 4.1 – Setup used in the experiment.

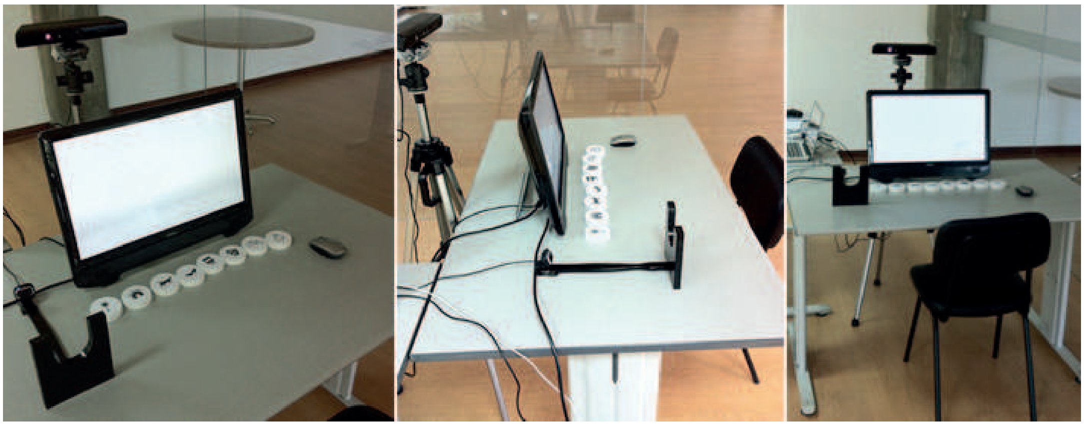


Figure 4.2 – Real setup used in the experiment.

modalities, some relevant adjustments were made regarding the cursor representation. The computer mouse was naturally represented by the classic white cursor, whereas the tangible interface triggered an instant visual feedback of the corresponding button's selection and thus had not a de facto representation of the cursor on the screen. Additionally, the gestural-based pointing illustration was a target badge grounded on the validated "point and wait" strategy (Schapira e Sharma, 2001) for selection. The touch input, however, led to a slight adjustment made to the cursor position on the screen in order to avoid the occlusion effect (Vogel e Baudisch, 2007). Therefore, for touch inputs we calibrated the system to register a contact point at a corrected position of minus 1 mm on the x axis, a slight change that enhanced the accuracy of the users for low objects' widths, as was understood after a preliminary pilot study.

Aside from the discrete data collected during the tests, we also gathered qualitative observational analysis on the participants' attitudes. At the beginning of each phase of this study, we performed a straightforward user questionnaire regarding the participants' previous experience with the interfaces at issue. Also, at the end of each test we proposed a questionnaire with closed-end questions, qualitative Likert Scales and ranking lists, in order to understand the user's preferences and their views regarding: ease of use of the input modalities, ease of learning, fatigue effect, naturalism of interaction, level of user comfort/frustration, and user's degree of presence and concentration.

Experimental design:

We dealt with two independent variables in this study: the age groups; and the input modalities. In turn, our dependent variables were: effectiveness, efficiency and user preference.

Altogether there were 1200 experimental trials during the first phase of the study: each of the 60 participants completed four required tasks five times each, which gives a total of 20 trials per subject. Regarding the evaluation within groups using the different input modalities, we used a repeated-measures within-participants design, whereas regarding the evaluation between age groups we resorted to a repeated-measures between-participants design.

We have approached how we handled outliers in Chapter 3 of this document.

Participants:

The study included 60 volunteers who were naïve to the purpose of the experiment. We divided them into three groups, according to their age, and we thereby had 20 users for each age group participating in the tests:

- **Children:** 20 participants of 9 to 12 years old ($M = 10.4$; $SD = 1.01$).
- **Young adults:** 20 participants of 20 to 30 years old ($M = 25.3$; $SD = 3.44$).
- **Older-adults:** 20 participants of 45 to 60 years old ($M = 53.3$; $SD = 6.43$).

These individuals were selected according to the directives stipulated on Chapter 3. Of all of the participants, two were left-handed and the rest was right handed. All the participants reported normal or corrected to normal vision.

We assessed their level of awareness in terms of the different modes of interaction used in the experiment (Table 4.1), so that we could perceive if their experience could actually influence the interaction process.

Table 4.1 – Ratio of participants with prior experience using the different interfaces.

Interfaces	Children	Young adults	Older-adults
Graphical	100%	100%	100%
Touch	95%	100%	65%
Tangible	15%	35%	15%
Gestural	40%	60%	15%

The interface with the lowest experience amongst all the participants was the tangible one. The graphical and touch interfaces were the most familiar, although 35% of the older-adults confirmed having never used a display with touch support. The gesture-based interface was familiar to most groups but mainly for the young adults, due to their time spent playing console games with support to this modality. Also, we acknowledge that the percentage relating to tangibles regarding the young adults was much higher than the other groups for this same reason.

Procedure:

The aim of each task was to use one of the four input modalities available (mouse, touch, tangibles, gestures) to make a selection of the button on the screen. After given an overview about the range of available modes of interaction, participants were instructed about the goal of each task's level at hand: make a button selection out of eight buttons displayed on the screen that were randomly arranged into two columns (Figure 4.3).

Figure 4.3 – Layout of the buttons displayed on the screen.

We divided the task into four levels, each making use of a different mode of interaction: (1) computer mouse; (2) touch; (3) tangible pieces; and (4) hand gestures. Over the course of the experiment the researcher was always present in order to explain the levels and respond to any questions the participants could have.

Each level would start with a countdown to 1 before the grid of buttons was presented and the time started counting. At this moment, participants were requested to select a specific button as quickly as possible while avoiding the amount of errors. If the right button was selected the task would be complete and the corresponding message displayed on the screen, after which the user could advance to the next level. On the contrary, if the user selected the wrong button a “try again” message would appear and the level would be resumed 5 seconds later to give the participant another chance to successfully complete it. The task completion time was measured between the appearance of the group of buttons displayed on-screen and the final notification of “task successfully completed”.

The levels were arranged as follows. Level 1 and 2 involved the use of the computer mouse and the touchscreen, respectively, to make the button selection. For level 3 the system imposed the use of the tangible pieces labelled with icons related to the buttons’ names. Here, the users were told to hold the objects against the webcam so that the fiducials engraved at their bottom could be recognized (Figure 4.4). Lastly, level 4 consisted in a hand gesture-based selection strategy. To this end, the participants were requested to hold the arm in an upright position within a certain area in front of the screen and hold it for a short period of time. The system’s visual feedback of the hand detection was a target on the screen that would be controlled by the hand movement. Also, an animation was integrated to the target to inform the user about a selection being in progress. Figure 4.5 shows the details of the graphic feedback when the system detected the hand to be motionless. To make a selection the participants would thus move the target to the front of the desired button and wait 0.8 seconds for the system to recognize their intent. For evaluation purposes this interval was removed from each participant’s interaction time.

Considering the stress of the experiment with some audiences, mostly children, there were no time constraints to successfully complete the tasks, but the participants were aware of the importance of completing them as fast as they could. Also, the tests could be paused at any moment between the tasks in order for the participants to rest. Furthermore, training trials were not provided, as we wanted to understand the

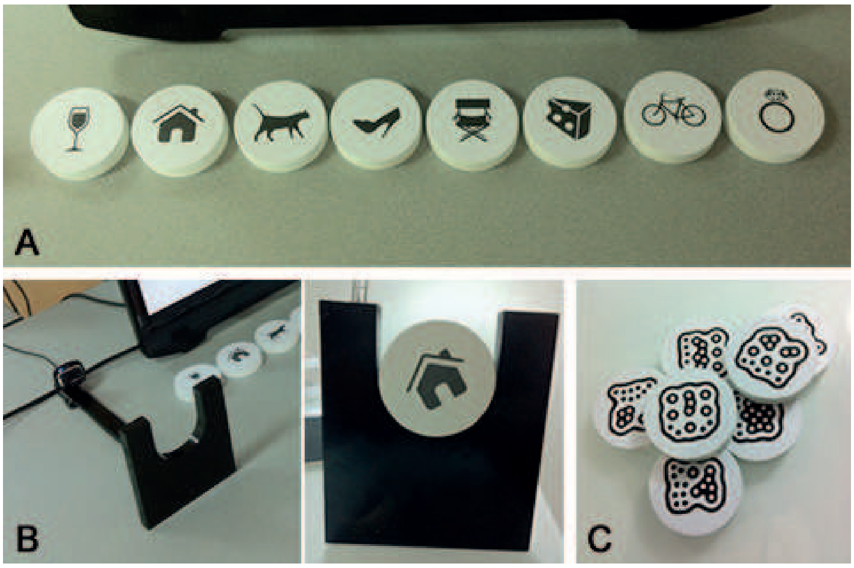


Figure 4.4 – (a) Tangible pieces used in the experiment; (b) The webcam device used to recognize the fiducials; (c) Fiducials engraved at the back of the pieces.

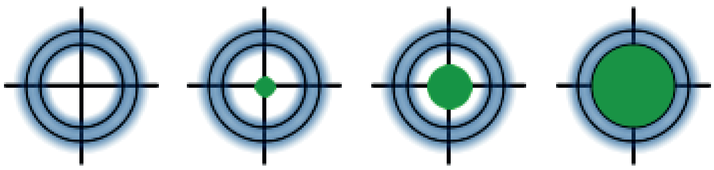


Figure 4.5 – Feedback for the selection strategy.

reaction time and adaptability of the participants who had no previous experience with certain modalities. However, we asserted the recurrence of the task (five times each) to perceive if performance would improve with practice or not.

Once the participants had finished the experiment, they were asked to fill a questionnaire. The purpose of this questionnaire was to gather their opinion about the different interaction modalities and preferences, as already described, and to evaluate the perceived performance and efficiency of those modalities from their own perspective.

4.1.1.2 Results

This section presents the results of the experimental tests for this study's first phase. The results are divided into subsections for clarity of presentation. The effectiveness of the selection trials (error rate and completion success) is dealt with, as well as efficiency (completion time) and qualitative results about the user preference for each interaction mode. We believe that this could be the first step in understanding how different user groups, regarding age brackets, perceive different input modalities and which could be considered the most efficient / effective in terms of a basic discrete selection task.

Effectiveness:

All of the participants managed to complete the tasks, which means that we recorded a 100% of success rate in task completion. According to accuracy, we considered an error when the button that was selected by the participant was not the correct one. The only input modality that registered faults during the task was the gesture-based one, with a total of 2 errors from the children's group, 8 errors from the adults' and 14 errors from the older-adults'.

The errors recorded were all due to: the users leaving their hand still in front of a button without realizing it while searching the screen for the right one; and moving their hand so slowly that the target on the screen would continue to be on top of a button when the selection animation terminated.

Efficiency:

Regarding the different interaction modalities, we analyzed the completion time within each group of participants and also between them. Table 4.2 shows each group's average completion times, the maximum and minimum time, as well as the standard deviation for the five trials with the four different input modalities, in seconds. Also worth noting is that regarding the overall efficiency of participants, the completion times recorded in each trial using each interface did not decrease,

Table 4.2 – Each group's mean, minimum, maximum and standard deviation times for the selection task regarding the four input modalities (in seconds).

Groups	Interaction mode	Mean time	Maximum time	Minimum time	Standard deviation
Children	Computer mouse	2.69	4.67	0.79	0.82
	Touch	2.27	3.77	0.98	0.66
	Tangibles	4.16	7.75	1.98	1.26
	Gestures	9.12	16.67	2,72	2.72
Young adults	Computer mouse	2.12	3.45	0.78	0.65
	Touch	2.10	2.98	1.13	0.46
	Tangibles	3.61	4.74	2.32	0.65
	Gestures	8.00	11.75	2.49	0.65
Older-adults	Computer mouse	3.25	5.74	0.90	1.00
	Touch	2.08	3.73	1.11	0.57
	Tangibles	3.84	5.94	2.72	1.01
	Gestures	9.63	12.14	2.72	3.21

and thus the recurrence of the trials did not enhance the results regarding any group.

The groups were more efficient using touch as the interaction modality. The group that registered the mean fastest completion time regarding the selection task was the older-adults, followed closely by the young adults and then the children. Curiously, all groups recorded the minimum times when using the computer mouse (children: 0.79 s; young adults: 0.78 s; older-adults: 0.90 s). However, regarding the mean completion time throughout all the groups, touch was the interface that showed faster results. The maximum time recorded during the experiment was using gestures, especially regarding children (16.67 s).

Figure 4.6 shows the mean time taken to complete the tasks for all three groups in each interface. Also, the data were normally distributed for each group, as we assessed with the Kolmogorov-Smirnov test.

Indeed, a one-way ANOVA determined there was no statistically significant difference between groups for each interface, except for the mouse input (p = .003). Significant correlation exists if the value of p is <0.05. Post-hoc comparison tests revealed that this modality registered a significant difference between the groups of adults and older-adults (p = .002).

Overall, the adults' group presented the lowest mean time in most of the interfaces,

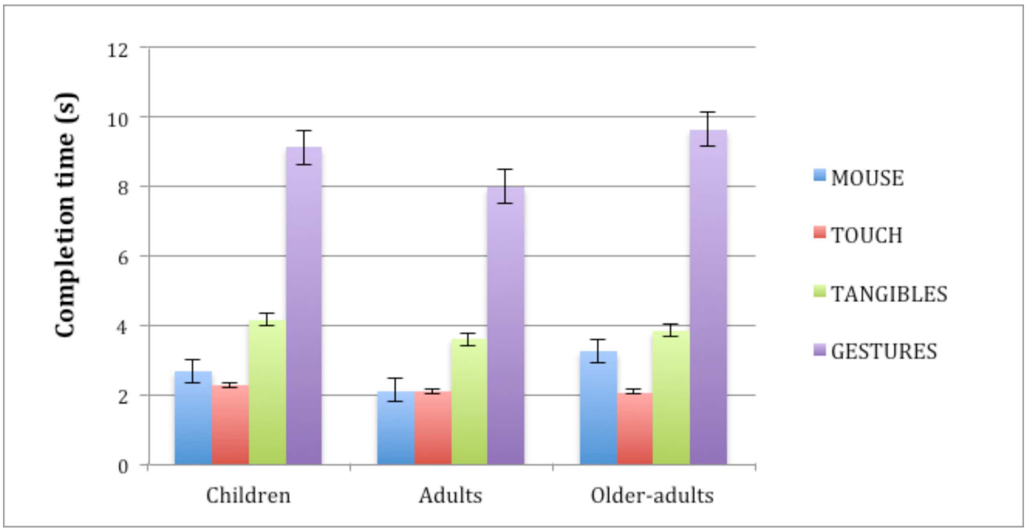


Figure 4.6 – Mean time taken to complete the task (in seconds).

comparing to the other groups. Also, the mean selection time for the several input modalities did follow a pattern throughout the three user groups: (1) the gesture-based interface was not nearly as efficient as the rest, showing much discrepancy when compared to the other interfaces in terms of mean time; and (2) touch was the fastest and proved to be more consistent throughout all the groups, with little variations in mean time.

Additionally to comparing the efficiency between each group regarding the different interfaces, we also analyzed the efficiency inside each group.

Regarding the group of the children, in terms of mean time, the input with the fastest results was touch (2.27 s), followed by the mouse (2.69 s), tangibles (4.16 s) and finally gestures (9.12 s). A repeated measures ANOVA with a Greenhouse-Geisser correction determined there was a significant difference between the mean time taken to complete the task in each interface ($F(1.108, 21.044) = 19.904$, $p <.000$). Post-hoc pairwise comparisons tests using the Bonferroni correction showed that, except for the mouse and touch inputs ($p = .353$), the mean times were statistically significantly different: mouse and tangibles ($p = .001$); mouse and gestures ($p = .000$); touch versus tangibles or gestures ($p = .000$); tangibles and gestures ($p = .000$).

As for the results of the young adults, the touch input registered again the lowest mean time (2.10 s) throughout the trials, followed very narrowly by the mouse (2.12 s). Also, the input modalities that registered the higher values were tangibles (3.61 s) and gestures (8.00 s). Indeed, this group registered such an insignificant difference between mouse and touch inputs (.02 s) that we can assume that it is indifferent for this group the use of the mouse or touch to perform a simple selection task. As the Mauchly's sphericity was not assumed, like in the children's group, we applied the Greenhouse-Geisser adjustment to the repeated measures ANOVA ($F(1.010, 19.190) = 12.520$, $p <.002$). Regarding the mouse and touch input modalities this group also elicited a slight reduction in terms of mean time, which was not statistically significant as demonstrated by pairwise comparisons tests using the Bonferroni correction ($p = 1.000$). This difference registered as even lower than the previous group. Moreover, the difference between gestures and tangibles was also not statistically significant ($p = .077$), as opposed to the other combinations: mouse and gestures ($p = .008$); touch and gestures ($p = .003$); and tangibles versus mouse or touch ($p = .000$).

The older-adults' efficiency also achieved a statistically significant difference between the mean times in each interface ($F(1.137, 21.606) = 32.007$, $p <.000$), as determined by a repeated measures ANOVA with the Greenhouse-Geisser correction. According to post-hoc pairwise comparisons tests using the Bonferroni correction, the mean times were statistically significantly different, except for the tangibles and mouse ($p = .373$): mouse and gestures ($p = .000$); touch versus tangibles or gestures ($p = .000$); tangibles and gestures ($p = .001$). Also, contrarily to the other groups, the difference between mouse and touch input was statically significant in the older-adults' group ($p = .003$). Indeed, this group registered a different pattern concerning performance when compared to the other two groups: mouse and tangibles did not register a statistically significant difference with .59 seconds apart from each other in terms of mean time; and, also very dissimilar to the other groups, mouse and touch registered a statistically significant difference with 1.17 seconds apart from each other in terms of mean time. These results led to a contrast between the groups concerning the different interfaces. The interface that scored the lowest mean time was touch (2.08 s), followed by the mouse (3.25 s), tangibles (3.84 s) and gestures

(9.63 s).

Participants' attitudes and preferences:

We observed different behaviors in the various groups concerning each modality. When using the computer mouse, all of the participants showed a regular approach of looking for the right button and clicking on it. However, we noticed that children and the older-adults first took their time searching for the cursor and only after finding its position they started searching for the right button. When resorting to touch and tangibles all groups demonstrated a similar behavior, and did not demonstrate an abnormal effort in completing the tasks.

The gesture-based interface was the one that proved to be more challenging, especially when selecting the buttons located at the top left and right corners. The most perceptible behavior of the participants in this situation was that they would push the arm forward to try to select the button, and thus exiting the detection area, instead of lifting it upwards. This situation would lead the participant to a higher level of frustration and thus ending up affecting the rest of the task until successfully concluding it.

On a different matter, there was no apparent difference in attitude concerning users right or left-handed. They performed the task with equal comfort.

At the end of the experiment, when asked about the interface they liked the most regarding ease of use and intuitiveness, the majority chose touch. Even for those who had never used touch as the interaction mode, this was the most popular and every participant said it was easy to use and practical for an everyday use. As for comfort in interaction, all of the subjects designated the gestural interface as the least comfortable and more challenging of all. Also on this interface, the groups of adults and older-adults thought it was also more demanding in terms of concentration.

In terms of user preference, the children chose the touch modality and the tangibles as the top two favorites, and the groups of adults and older-adults chose the touch interface. This was registered during the final survey where the participants ought to pick their favorite type of interface and also their least preferred. Indeed, the majority of children chose touch as favorite, with 60% of the answers putting this modality in the first place, followed by the tangible interface (25%). At the bottom two places were the gestures and the mouse. Likewise, the majority of the young adults elected touch as the favorite, with 65% of the replies, and pointed gestures as the least favorite. The older-adults chose touch as the favorite with 75% of the answers, leaving behind the other input modalities with just 10% for mouse and the same for gestures, and 5% for the tangibles.

4.1.1.3 Discussion

We followed a quantitative method of evaluation with specific performance metrics: speed in successfully completing a task and the respective effectiveness (number of errors) for the task completion. Also, qualitative survey was made in order to understand the participants' attitudes and preferences.

The group of the older-adults registered the higher error rate and also the higher mean time of all groups. In the case of gestural input, during the trials the older-adults seemed not to have the motor reflexes to divert the hand from the button being incorrectly selected, that is, move the hand away from the button before the animation of the target selection was completed. It could also be the case that the cognitive process of understanding the hand position, while thinking about the task, was compromised, and thus a mistake was not avoided. All the same, we may infer that the acuity of the group of older-adults could decline along the years: the motor reflexes could have been influenced by age, as the motion required to prevent unintentional selections was slower.

Regarding the overall efficiency of participants, the repetition of the task with the different input modalities did not enhance the users' efficiency. This was an important factor, as we did not have a random task order and thus the learning factor was kept under close attention. Indeed, it was not a priority to understand if the mean times improved with experience, as it might happen, but to understand the time taken by the user to adapt to each interface and perceive if the latter could be intuitive and not require previous training (for that same reason we did not offer a preparation period). Such circumstance is latent regarding the computer mouse that, curiously, even with a similar experience level amongst the participants (it was the input device every participant used on a daily basis), proved to be the only input modality with statistically significantly different results between the groups of young adults and older-adults. We might infer that analyzing these two groups it seems that instead of increasing their performance with the experience over the years, it could end up suffering, that is, performance might get compromised as the years go by. We think that this fact requires special attention during the research. Observing the worst completion times when using the mouse, and knowing that this modality presented parallel levels of experience throughout the groups, the group of older-adults shows there could be a loss of performance due to age, instead of an improved performance due to increased experience.

For a simple selection task, touch appears to be more efficient amongst the three groups in terms of mean time, even considering that their awareness in terms of this input modality was distinctive (as concluded in the initial questionnaire). Even though users assumed to have different levels of experience with touch surfaces, they showed the best results overall using this interface. Therefore, evidences suggest that touch could actually be more intuitive and provide an easy adaptation.

4.1.1.4 Summary

In this first study, we intended to understand which input modality could be the most efficient for a specific target-audience in a discrete selection task. The modalities considered were the computer mouse, touch, tangibles and gestures; and the three age groups were organized as children, young adults and older-adults.

Overall, gestures presented the worst results amongst the three groups, followed by tangibles. On the other hand, touch and the mouse performed the best, being touch the interface with the lower completion times in all groups. However, each group presented significant differences in interaction when comparing the input modalities side-by-side.

Also, in terms of user preference, it seems important to be aware that specific groups of users may perform better with one interface, and actually prefer / like to use another. This is the case of children that enjoyed using tangibles for selection and chose this input modality as one of their favorites. Hence, this decision may throw some light on which interface could be better accepted and welcomed regarding particular age groups.

Even though this is the first stage of our study, according to the results presented it appears to be evidence that for a specific selection task, the age group does not influence the performance but, instead, the type of interface that is used does. That is to say, according to the type of task, the use of different interfaces may influence the users' performance and preference. Even when comparing the groups side by side, we may reach the conclusion that not all of the interfaces produced the same results.

Furthermore, usability results have highlighted the need for further investigation in order to understand key issues related to input modalities and user preferences. Hence, it seems valid and important to understand performance according to different levels of difficulty a continuous selection task may involve, and so a Fitts' Law evaluation was the next step to have into account in our study.

4.1.2 Target acquisition performance evaluation

This second phase of our study regarding selection tasks was designed to characterize the index of performance of the three age-based user groups regarding three different input modalities (mouse, touch and gestures) for an elemental activity: selection (or target acquisition). In this phase we did not cover tangibles as they are not a pointing device and thus could not be considered.

4.1.2.1 Materials and Methods

To compare the three input modalities for pointing we made use of Fitts' Law, a frequently used model for measuring movement performance.

Indeed, it is naïve to try to measure input performance merely by analyzing user's task completion times and the average of all trials, including, at times, the number of errors that took place; that is to say, measure input performance only through usability evaluation. If the circumstances of the experiments change (such as target sizes or distances in the screen) the method may not lead to the same results. This scenario is relevant for Fitts' law, as one of its goals is to measure and characterize the performance of various input devices, regarding different levels of difficulty and movement times. More details about this performance model can be found in Chapter 3.

For this second analysis, we resorted to continuous selection tasks, where the participants resume a selection movement without stopping and perform multiple continued actions (for more information about this type of discrete task, please refer back to Chapter 2).

Apparatus:

The experiment was conducted in a closed room with artificial light and the tests were performed in a similar setup assembled for the previous testing (Figure 4.7): the system consisted on a 22" touchscreen placed in front of the user, with a resolution of 1280x800 pixels; a Microsoft Kinect sensor mounted on a tripod behind the screen, about 25 cm above it and facing the user; and a computer mouse.

Materials:

A purpose-built application was created with support for all modes of interaction, in order to keep the testing environment as coherent as possible. The tasks' software was developed in Python with the support of: the Kivy Framework and TuioKinect

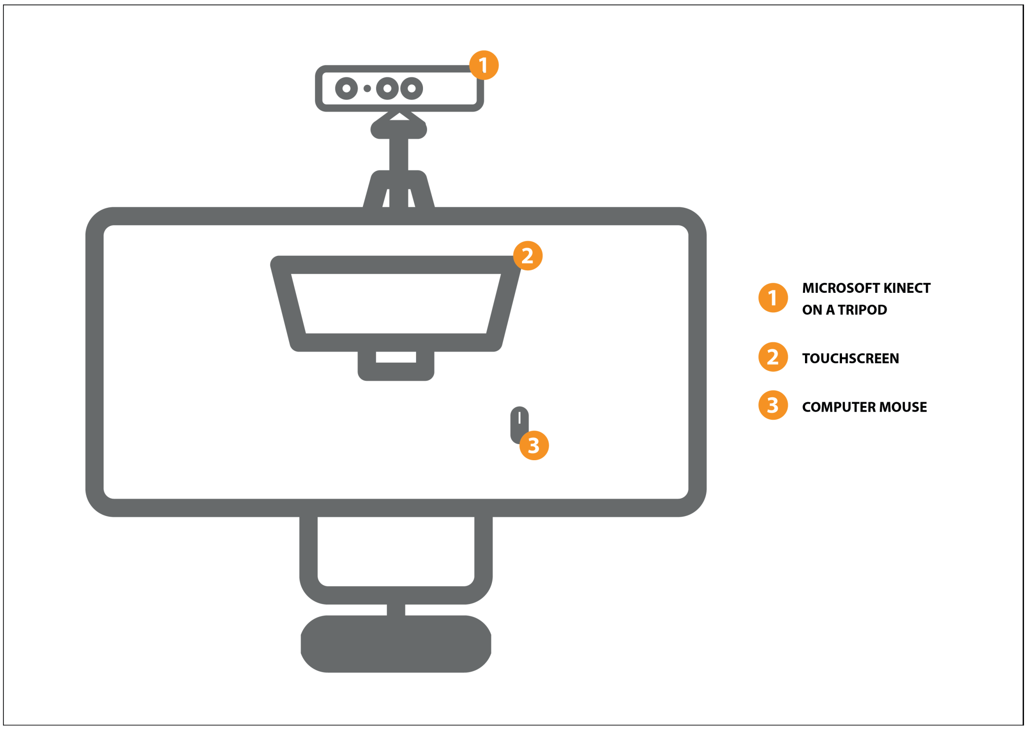


Figure 4.7 – Experimental setup.

(to track the hand gestures).

In order to reduce the impact of the system's feedback across the different input modalities, identical adjustments were made to this application as on the previous usability evaluation: the gestural-based pointing illustration was a target badge grounded on the validated "point and wait" strategy (Schapira e Sharma, 2001) for selection; and the touch input underwent a slight adjustment made to the cursor position on the screen in order to avoid the occlusion effect (Vogel e Baudisch, 2007).

We also performed a straightforward user questionnaire regarding the participants' previous experience with the interfaces at issue at the beginning of the tests. During the tests, we gathered qualitative observational analysis on the participants' attitudes and, at the end of each test, we proposed a questionnaire with closed-end questions, qualitative Likert Scales and ranking lists, in order to understand the user's preferences and their views regarding: ease of use of the input modalities, ease of learning, fatigue effect, naturalism of interaction, level of user comfort/frustration, and user's degree of presence and concentration.

Experimental design:

We dealt with three independent variables in this study: the age groups; the input modalities; and the indices of difficulty (ID). In turn, our dependent variables were: effectiveness, efficiency, index of performance and user preference.

Our goal is to compare the age groups and interaction modalities resorting to Fitts' law. As such, we used six different amplitudes: A = 200, 460, 750, 770, 930, 940 pixels; and five different target widths: W = 16, 26, 50, 120, 200 pixels; which attain seven levels of Fitts' law index of difficulty, from 1.72 bits to 5.20 bits. The order of appearance of the seven indices of difficulty was randomized. We also analyzed the speed-accuracy tradeoff (Zhai et al., 2004) to understand the performance of the users.

Thus, the target acquisition data consisted of: 3 age groups (60 participants); three input modalities (mouse, touch, gestures); two target directions (left, right); 25 repetitions per index of difficulty; and seven indices of difficulty (1.72, 2.25, 3.12, 3.14, 4.04, 4.89, and 5.20 bits). Thereby, altogether there were 42 experimental conditions and an overall of 63 000 correct trials.

Regarding the evaluation within groups using the different input modalities, we used a repeated-measures within-participants design, whereas regarding the evaluation between age groups we resorted to a repeated-measures between-participants design. Also, we have approached how we handled outliers in Chapter 3 of this document.

Participants:

We invited 60 participants who were naïve to the purpose of the experiment and divided them into the three groups considered for our study:

- **Children:** 20 participants of 10 to 12 years old ($M = 11.5$; $SD = 0.87$).

- **Young adults:** 20 participants of 20 to 30 years old ($M = 27.1$; $SD = 3.56$).
- **Older-adults:** 20 participants of 45 to 60 years old ($M = 50.2$; $SD = 5.85$).

As the previous experiment, these individuals were selected according to the directives stipulated on Chapter 3. Of all of the participants, only one was left-handed. All the participants reported normal or corrected to normal vision.

Procedure:

As in the first phase of the study for selection tasks, we divided the task into levels, each making use of a different mode of interaction: (1) computer mouse; (2) touch; and (3) hand gestures. The researcher was always present as well. The entire procedure described below was repeated for each interaction mode and their order was shuffled between participants.

The users' performance was analyzed considering only movements along one axis and thus we used a reciprocal one-dimensional pointing task based on Fitts' original experiments. This consisted on a horizontal movement between two vertical bars displayed on the screen, representing an initial location and its target (as shown in Figure 4.8).

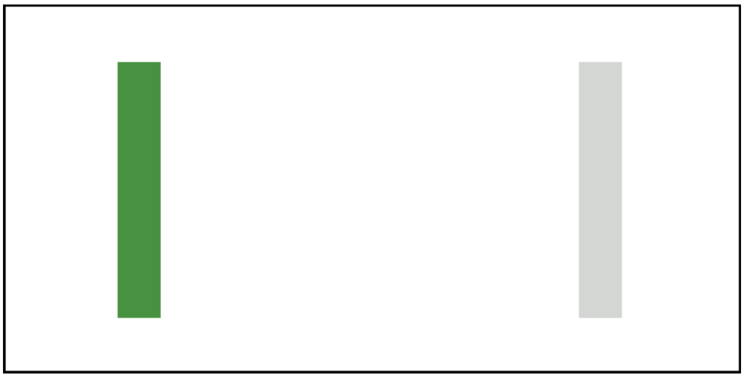


Figure 4.8 – Layout of the targets displayed on the screen.

We only intend to understand the one-dimensional movement with homogeneous shapes for a simple selection task. Each trial proceeded as follows. Prior to each trial, the participant would be required to select a “start” button shown on the screen, and thus keeping a consistent original position. The two vertical bars would then be revealed: the initial position bar being grey and the target bar green. Participants then selected the target bar in each successive trial, which was at all times organized in opposing directions (from left to right and from right to left). Also, there was a visual feedback to indicate a hover effect on the target bar: it changed its color to blue. The start time was recorded from each target bar to the next and stopped when the next target was selected. At the end of each interface’s trial, the bars disappeared and a “successfully completed task” message appeared on the screen.

Considering the target bar selection with the computer mouse and touch, the interaction process was kept original: a mouse click with the cursor on top of the bar; and a finger touch on the screen regarding the bar’s position. Here, the participants were asked to choose their index finger to interact with the touchscreen. For gestural selection the process was identical to the one implemented on phase one of this study, described earlier.

Participants were given practice attempts for each interaction mode before starting the experiment. This period consisted on completing a continuous task (MacKenzie, 1992) with 10 consecutive trials for each of the input mode, in order for the participant to feel at ease with the interaction.

There were no time constraints to successfully complete the tasks, but the participants were asked to complete them as fast as they could. Also, they could pause between the input modalities and rest.

After the experiment, participants were asked to fill a questionnaire. The purpose of this questionnaire was to gather their opinion about the different interaction modalities and preferences. The topics found on the questionnaire matched the ones included in the previous usability evaluation’s survey.

4.1.2.2 Results

This section presents the results of the experimental tests for this study's second phase. The movement time of the continuous tasks and indices of performance of each group regarding the different interaction modalities are presented and compared. Also, we examine the participants' attitudes during the tests and their preference towards the interfaces. The results are divided into subsections for clarity of presentation. Each experimental session lasted approximately 25 min.

Effectiveness:

We assumed a speed-accuracy tradeoff approach (Zhai et al., 2004): the more precisely the task was performed, the longer it took to be completed, and vice-versa. When dealing with errors in gestures, the selection rested on the "point and wait" strategy and thus did not trigger false selections caused by user mistakes. Here, the selection would only be triggered when the system detected the gesture to be motionless and precisely on top of the target bar. Otherwise, no selection was made.

Regarding touch and the computer mouse, error rates were higher for touch than for the mouse for the three groups. For children, error rates were on average 7.14% when using touch versus 5% for the mouse. Young adults made 6,71% errors on average using touch, compared to 6% on the mouse, and older-adults made 7.57% errors on average regarding touch, compared to 4,57% using the computer mouse. The correlation between speed and accuracy will be furthered next, when we analyze and compare the indices of performance (IP) of the three age groups.

Efficiency:

We present the movement times recorded during each trial of the experimental tests. First and foremost, we examine the trials' movement times recorded by each group of participants, and then we compare the age groups in order to determine if there is a significant difference regarding the interaction modalities.

Figure 4.9 shows the mean movement time of each age group when using the four different interaction modes, considering all trials. Keep in mind that the total time taken during the exercise is divided by the number of movements in order to determine the movement time (MT) for the specific target sizes and amplitudes. The input modality that registered the fastest mean results throughout all the three groups was touch. Here, the young adults registered a mean movement time of 0.64 s, followed by the children (0.71 s) and older-adults (0.75 s) with a slight difference between them. All groups registered the longer movement times when using gestures, especially considering children (2.29 s), that took 21% more time than young-adults (1.83 s) to complete the continuous task.

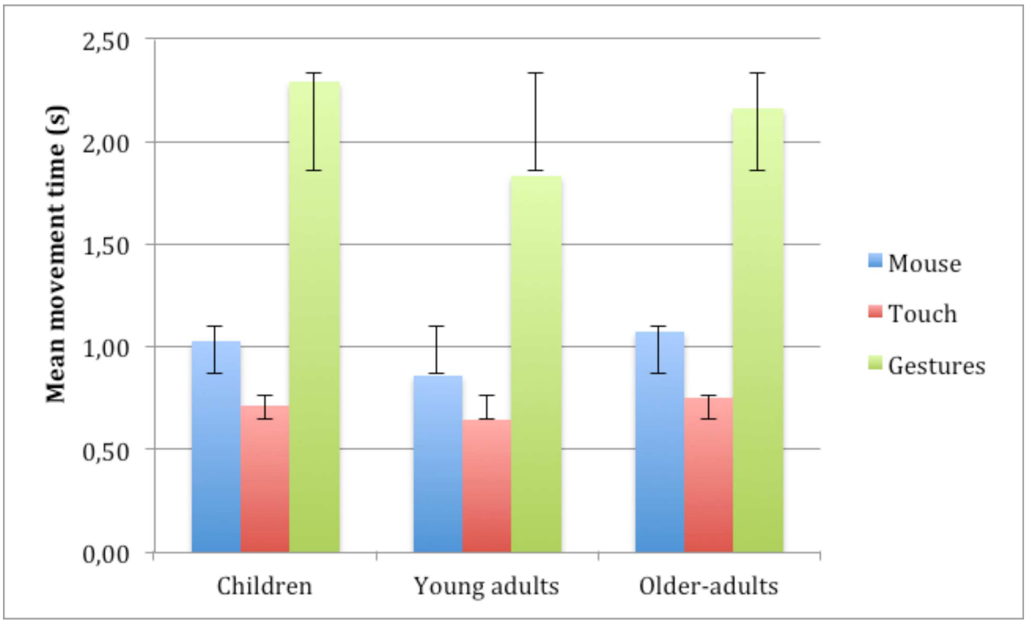


Figure 4.9 – Mean movement time of each age group.

A Kolmogorov-Smirnov test showed that the data were normally distributed. As such, we performed a one-way ANOVA test to determine if these values were statistically different. Indeed, all groups proved to have significant differences regarding the movement times using the three input modalities: mouse ($F(2{,}57) = 15.011$, $p = .000$); touch ($F(2{,}57) = 8.294$, $p = .001$); gestures ($F(2{,}57) = 15.172$, $p = .000$).

Also, Tukey post-hoc comparison tests revealed which groups presented significant differences, and about which input modality. Regarding the computer mouse, touch

Table 4.3 – Movement times recorded on ID 1 and ID 7 by all groups using the computer mouse, touch and gestures.

	Mouse		Touch		Gestures	
	ID 1	ID 7	ID 1	ID 7	ID 1	ID 7
Children	0.78	1.21	0.60	0.83	1.59	3.16
Young adults	0.62	1.10	0.55	0.76	1.32	2.20
Older-adults	0.81	1.29	0.63	0.86	1.66	2.66

and gestures, the only groups that were not statistically different were the children versus older-adults (p = .760; p = .274; p = .327; respectively). On the other hand, considering the groups of young adults versus older-adults, the differences in movement times were always significant: mouse and touch (p = .000); gestures (p = .001). Considering the computer mouse, the children and young adults performed significantly different (p = .000). Likewise, regarding gestures, the groups of children versus young adults registered a significant difference (p = .000). On the other hand, the difference between movement times were statistically significantly lower considering the groups of children and young adults when resorting to touch (p = .042).

On another subject, we also present the groups' movement time per index of difficulty for each interface (Figure 4.10), since it could be helpful to understand if the movement times are influenced when increasing the difficulty stage.

Indeed, as the difficulty increases, so do the mean movement times of all three groups. Thus, when considering the input modalities, we can see that movement times tend to be aggravated as the index of difficulty (ID) rises. Also, gestures tend to degrade more exponentially as difficulty increases than the other two input modes, which shows that they are more susceptible to variances considering the difficulty of the task. The same situation happens when using the computer mouse, but on a smaller scale. Once more, we can understand that touch does show lower movement times regarding all groups. In order to better perceive the values, Table 4.3 shows the movement times recorded on ID 1 and ID 7 by the three groups with the different input modalities.

Touch seems to suffer little changes from the first index of difficulty to the seventh.

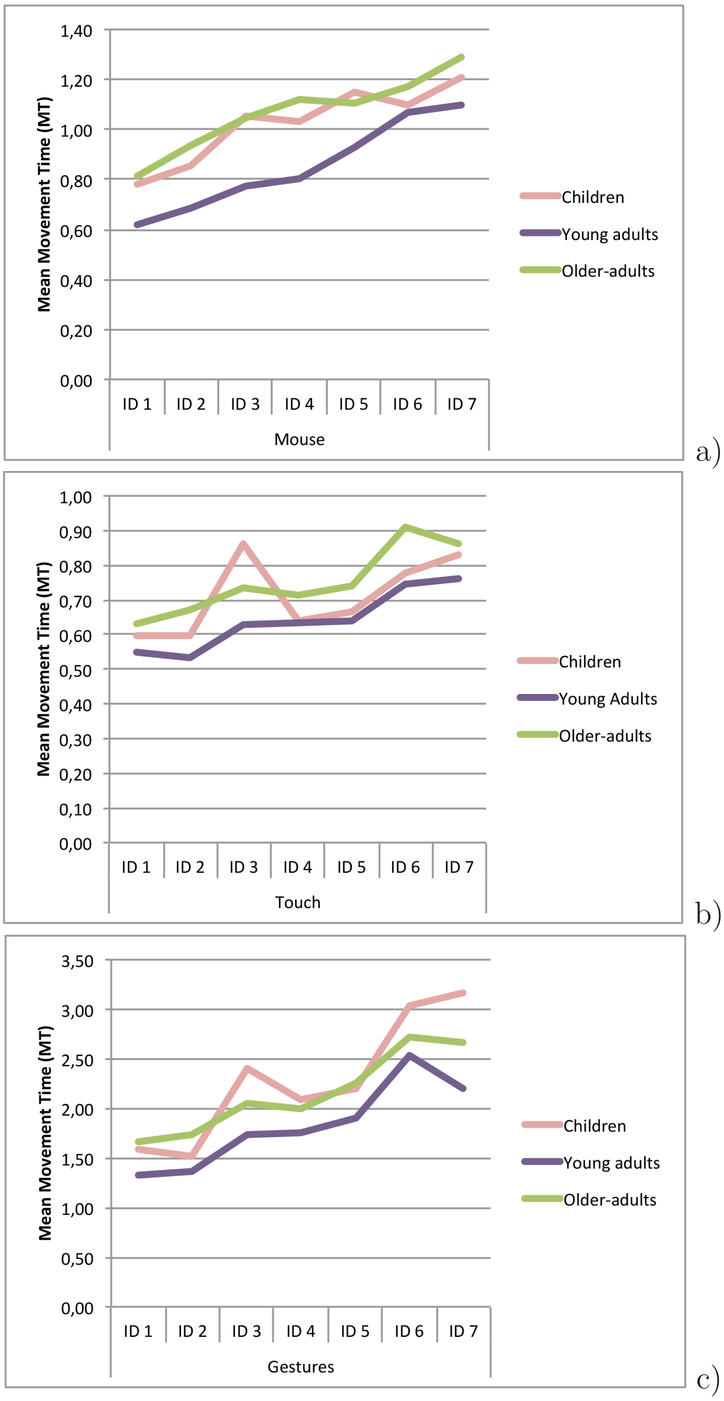


Figure 4.10 – Mean movement time per index of difficulty (in seconds): a) Using the mouse; b) Using touch; c) Using gestures.

Also, it shows fewer discrepancies, as happens with the computer mouse, when comparing the values of the three age groups. On the other hand, the gesture-based interface shows higher variances in values mainly when considering the last level of difficulty, which does not happen with the other interfaces.

Additionally to comparing the movement times between each group regarding the different interfaces, we also analyzed these values inside each group.

Regarding the group of the children, the input modality with the fastest mean results was touch (0.71 s / SD = 0.08 s), followed by the computer mouse (1.03 s / SD = 0.13 s) and then gestures (2.29 s / SD = 0.39 s). This indicates that this age group performed 31% faster when using touch comparing to mouse, and 69% faster when compared to gestures. A repeated measures ANOVA with a Greenhouse-Geisser correction determined there was a significant difference between the mean time taken to complete the task in each interface ($F(1.150, 21.844) = 247.770$, $p < .000$). Post-hoc pairwise comparisons tests using the Bonferroni correction revealed that this age group shows significant differences when using distinctive interaction modalities, whether it is mouse, touch or gestures ($p = .000$), regarding continuous selection tasks.

The young adults' movement time also revealed to be statistically significantly different between the mean movement times in each interface ($F(1.296, 24.620) = 688.371$, $p < .000$), as determined by a repeated measures ANOVA with the Greenhouse-Geisser correction. Also, according to post-hoc pairwise comparisons tests using the Bonferroni correction, these values were indeed significantly different comparing all interfaces ($p = .000$). Indeed, the young adults registered mean movement times of 0.85 s using the computer mouse (SD = 0.072 s), 0.64 s when using touch (SD = 0.074 s), and 1.83 s when using gestures (SD = 0.17 s), which denotes a 25% improvement in performance using touch rather than mouse, and 65% comparing to gestures.

As for the results of the older-adults, touch (0.75 s / SD = 0.10 s) was 30% faster than the computer mouse (1.07 s / SD = 0.18 s), and 65% than gestures (2.16 s / SD = 0.19 s). As the Mauchly's sphericity was not assumed, we applied the

Greenhouse-Geisser adjustment to the repeated measures ANOVA ($F(1.754, 33.327) = 637.926$, $p < .000$), which showed us that, as the other groups, the differences in movement times were, also, statistically significantly different ($p = .000$).

Performance:

We present the values of the Index of Performance (IP) of each group regarding the different input modalities, and then we analyze these values more closely within each group, in order to determine if these report significant differences.

It is important to recall that the combination of several measurements of movement time throughout the tasks with increased difficulty into a single statistic value reveals an Index of Performance, or throughput, of the pointing device. As such, this index comprises both speed and accuracy in a pointing task. As movement time is measured in seconds and the index of difficulty in bits, the reciprocal of the slope of the line is in "bits/s". This measure is considered the rate of transmission known as Index of Performance. The higher the value of the throughput, the higher the rate of human performance.

Figure 4.11 shows the mean IPs registered during the experiment for each age group when using the four different interaction modes.

Indeed, users can manage more information per unit time using touch as the interaction modality, especially children, that appear to have a higher IP (IP = 20,42 bits/s; SD = 8.42 bits/s) than young adults (IP = 15.46 bits/s; SD = 5.08 bits/s) and older-adults (IP = 17.31 bits/s; SD = 8.67 bits/s).

On the other hand, when using gestures for interaction, the groups presented the worst IPs of all: children presented an IP of 2.17 bits/s (SD = 0.69 bits/s); young adults, 3.25 bits/s (SD = 0.83 bits/s); and older-adults, 3.62 bits/s (SD = 1.43 bits/s).

When using the computer mouse as the pointing device: children registered the highest throughput of 9.29 bits/s (SD = 3.29 bits/s); the young adults, 6.79 bits/s (SD = 1.29 bits/s); and the older-adults, 9.22 bits/s (SD = 2.49 bits/s).

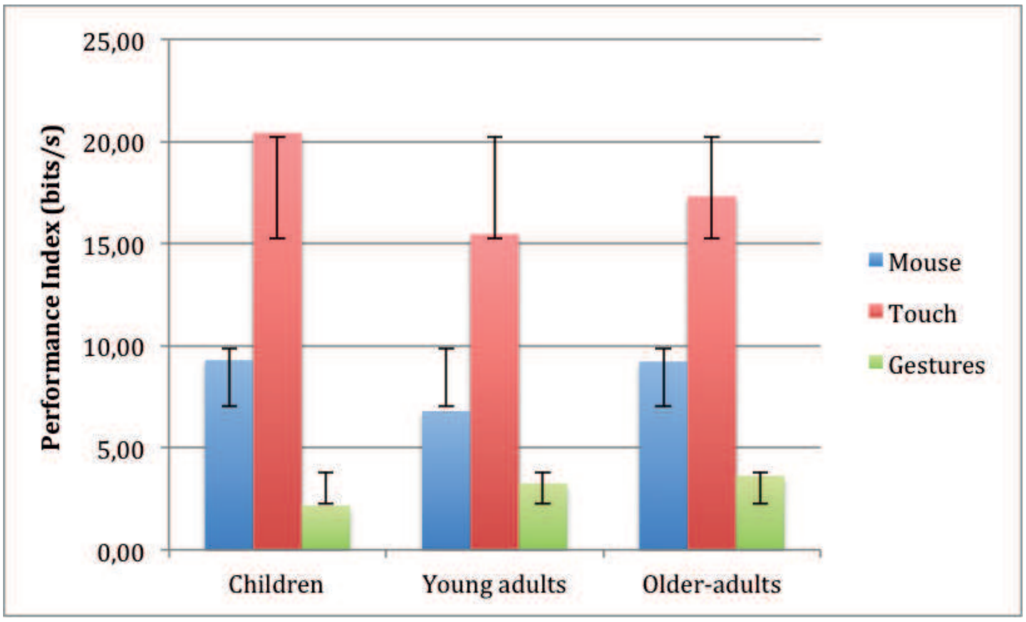


Figure 4.11 – Index of Performance (IP) of the three groups regarding the different input modalities (in bits/s).

A Kolmogorov-Smirnov test showed that the data were not normally distributed. As such, a non-parametric independent-samples Kruskal-Wallis 1-way ANOVA was run to determine whether the difference in terms of IP throughout the groups was indeed significant. In this regard, there were significant differences in IPs when using the computer mouse ($\chi2(2) = 11.124$, p =.004), concerning the group of young adults versus children ($p = .018$) and older-adults ($p = .007$), but not between children and older-adults ($p = 1.000$). The same is true for gestures ($\chi2(2) = 17.929$, p =.000), but regarding children versus young adults ($p = .001$) and older-adults ($p = .000$) and not between young adults versus older-adults ($p = 1.000$). On the other hand, touch did not present significant differences between the groups ($\chi2(2) = 2.595$, p =.273).

Additionally to comparing the IPs between each group regarding the different interfaces, we also analyzed these values inside each group.

Regarding the group of the children, the IP was 54,51% greater with touch than using the computer mouse, and 89,37% higher when compared to gestures. A Friedman test determined that there was a significant difference between the IPs of the input modalities ($\chi 2(2) = 30.125$, $p = .000$). Also, a post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p <.017$. Median perceived throughputs for the mouse, touch and gestures were 8.7, 17.03 and 2.07, respectively. There were, indeed, significant differences regarding all of the input modalities when it comes to IP: touch and mouse (Z = -3.506, $p = .000$); gestures and touch (Z = -3.724, $p = .000$); gestures and mouse (Z = -3.516, $p = .000$).

As for young adults, their throughput was 56,08% greater with touch than using the computer mouse, and 79,5% higher when compared to gestures. There was a statistically significant difference in perceived IPs according to Friedman test, $\chi 2(2) = 26.143$, $p = .000$. A Bonferroni correction was applied to post hoc analysis with Wilcoxon signed-rank tests, resulting in a significance level set at p <.017. Median throughputs for the mouse, touch and gestures were 6.80, 14.50 and 3.84, respectively. Thus, there were statistically significant differences in performance: touch and mouse (Z = -3.294, $p = .001$); gestures and touch (Z = -3.516, $p = .000$); gestures and mouse (Z = -3.516, $p = .000$).

All the same, older-adults increased their IP by 46,74% using touch rather than the mouse, and 81,74% comparing to gestures. The Friedman test validated a significant difference between the IPs of the input modalities ($\chi 2(2) = 30.632$, $p = .000$). As in the previous groups, the Bonferroni correction was applied to post hoc analysis with Wilcoxon signed-rank tests, resulting in a significance level set at p ¡ .017. Median throughputs for the mouse, touch and gestures were 8.80, 13.07 and 3.01, respectively. Indeed, the different input modalities significantly influenced performance: touch and mouse (Z = -3.018, $p = .003$); gestures and touch (Z = -3.920, $p = .000$); gestures and mouse (Z = -3.823, $p = .000$).

Participants' attitudes and preferences:

Overall, the participants understood what was asked and the procedure. Also, there were no significant events that could be considered relevant during this study in the group of children and young adults, other than the fact that gestural interaction did show a tendency to tire the participants. Indeed, we noticed that the participants would become very uncomfortable over time and get more tired using gestures in this continuous task activity. As they could not rest until all of the trials ended using one specific modality, they usually complained about discomfort in the arm being used to interact. This was also noticed in the group of older-adults.

Moreover, we observed other different attitudes in the group of older-adults concerning touch and gestures when compared to the other groups. Indeed, older-adults showed a tendency of following the same pace throughout the trials when using touch, that is, whether the index of difficulty changes or not, they maintain the same rhythm selecting the intended targets. They do not go faster when targets and easier (larger) and they do not slow down when they are more difficult (thinner). This fact is interesting, because we can see that using touch, they seem to follow a pace and maintain it until the end of the trials. This was only noticed using touch. With the other input modalities, they would go faster and slower according to the amplitude of the objects.

The same problem noticed in the previous usability evaluation regarding gestures was also observed in this activity: the older-adults would tend to push the arm upwards instead of forward and thus leaving the device's detection area. Also, when the targets were thinner, the participants would complain about the difficulty in selecting them.

Regarding the final questionnaire that the participants were asked to answer, 40% of the children said the computer mouse and touch were extremely easy to use. The young and older-adults had similar opinions: 50% of the young and older-adults said the computer mouse was extremely easy to use, and they had the same opinion towards touch (70% and 40%, respectively). On the other hand, 35% of older-adults and children said gestures were relatively difficult to handle, as well as 50% of young adults saying they were relatively easy to use. This input modality was indeed the least liked one. Moreover, all age groups acknowledged that touch and the computer mouse were extremely comfortable to use, whereas gestures were mostly considered as relatively tiring and frustrating.

4.1.2.3 Discussion

Regarding the movement times recorded using the computer mouse, touch and gestures, the only groups that were not statistically different were the children versus older-adults. This insight may indicate us that children and older-adults show a similar interaction behavior regarding movement times when performing continuous selection tasks. On the other hand, considering the groups of young adults versus older-adults, the differences in movement times were always significant. This consideration may denote that despite the input modality used, these age groups perform very differently.

The difference between movement times of the groups of children versus young adults were statistically significantly lower when resorting to touch as the input modality, which may suggest that these age groups have close interaction characteristics when using touch surfaces for interaction, although still meaningful.

Like the previous usability evaluation, we can see a pattern in the results of this experiment: (1) the gesture-based interface remained the least efficient throughout all three groups; and (2) touch was the one to display faster movement times. Therefore, we may come to the conclusion that regarding simple reaction times, these interfaces allow similar efficiency whether on discrete or continuous selection tasks.

In contrast, considering continuous selection tasks, age seems to be of importance. Indeed, not only the input modality influenced the efficiency of the users, but also their age seems to have played a part in the significant differences in interaction. This is somewhat different than we had experienced during the usability evaluation of the discrete selection task, when we acknowledged that the age had not influenced interaction, but the type of interface had. This disparity could have something to do with cognitive processes and dexterity when performing multiple non-stop actions, although this discussion will not be further examined yet. More types of tasks will have to be considered to understand this situation, as we will see in the next sections of this document.

The young adults' movement time also revealed to be statistically significantly different between the mean movement times of each interface, which indicates that young adults display considerable differences when resorting to distinctive interaction modalities to perform continuous selections tasks.

Overall, in spite of being faster using touch to accomplish the continuous task, all groups were less precise when selecting the targets, and thus made more errors, which can imply that touching the surface to select objects may reduce accuracy. However, users can manage more information per unit time using touch as the interaction modality, especially children, that appear to have a higher IP than young adults and older-adults. This is interesting to analyze, as they were not the ones to present faster movement times. Nonetheless, they seem to better balance precision versus speed in interaction comparing to the other age groups. Also, we can see that young-adults are the ones to display a worse IP regarding touch. This is interesting, since they displayed the faster results in terms of movement times, which leads us to think that they could not balance speed and accuracy as well as the other groups. Certainly, they registered the lowest movement times but still could not take advantage of that situation and ended up decreasing their accuracy and, consequently, their throughput. Such fact allowed the older-adults to demonstrate a better IP than the young adults. Also, touch proved to be constant throughout the three age groups regarding throughput, as opposed to the other input modalities.

When using the computer mouse as the pointing device, children registered the highest throughput of all groups. Again, children performed the best, closely followed by older-adults. Curiously, the young adults demonstrated once again the worst throughput. This can be an indication that although they are faster than the other groups, they aren't so precise when it comes to continuous selection tasks.

Also, regarding touch and the computer mouse, young adults proved to be faster than the other groups, but still could not take advantage of that situation and ended up decreasing their accuracy and, consequently, their throughput, being surpassed by children and older-adults.

On the other hand, when using gestures for interaction, the groups presented the worst IPs of all. Even so, older-adults were a positive surprise and demonstrated the best IP of all the groups, which may indicate that, even though they present one of the lowest movement times of all the groups, they are capable of managing their movement time and accuracy accordingly, resulting in a higher throughput.

Indeed, when it comes to gestural interaction, the results were the worst in terms of movement time and throughput. In this sense, we had come to believe that it was important to experiment with other gesture-based devices in order to understand if these values are due to actual interaction difficulties by the participants, or rather the device itself was not the most appropriate. We thus acknowledged that the device itself used for the gesture-recognition - the Microsoft Kinect - could be negatively influencing the user due to the accuracy required by the task.

Different optical sensors, which allow human body acquisition with respected accuracy, have been released and comparable controllers in the same price range include the Leap Motion, a sensor with declared sub-millimeter precision that claims to obtain higher levels of accuracy than the Microsoft Kinect (Weichert et al., 2013). Therefore we believed to be important to understand if another device could perform better than the one used in the former experiments. After all, when it comes to gestural interfaces the precision of the sensor is said to be vital.

In this context, we performed a case study that aimed at understanding the indices of performance of the three age groups comparing the use of two different gestural sensors: Microsoft Kinect and Leap Motion. We intended to understand: (1) if the gestural interaction presents significantly better results on specific groups opposed to others with regards to their age, and thus confirming our previous findings; (2) if the devices used and their declared precision do influence or not the target acquisition times and indices of performance in each group.

The study was published in the paper entitled "Performance evaluation of gesture-based interaction between different age groups using Fitts' Law" (Carvalho et al., 2015b). Also, a deeper study has been accepted for publication in (Carvalho et al., 2016). Hence, we direct the reader to those papers to further analyze our study, as we did not intend to turn away from the main focus of our study and thus do not explain it in this document.

In short, we came to the conclusion that: (1) in terms of index of performance the groups displayed statistically different results (the group of older-adults held the best gestural performance results, and the children the worst, when compared between each other); (2) both devices behave in a similar manner for selection tasks and there are no statistically significant differences concerning their accuracy when comparing each of the three groups individually.

Indeed, it appears to be evident that for continuous selection and target acquisition tasks a gestural interface may not be the best approach, as it shows low indices of performance throughout the three groups. We may presume that for selection tasks that require a higher level of precision neither device displayed proof that it influenced the user to complete the trials with better performance. In fact, the devices' accuracy did not appear to have a direct relation to improved performance.

As stated earlier, mean movement times should not be considered the only variable when exploring the users' performance, and here we can relate to why: for example, when comparing the movement times and respective index of performance, children exhibited mean times similar to the other groups, but not nearly as good performance as the others. Children have as good response times as other groups, but they do not present a consistent interaction, which may reveal that they are faster but not as precise when interacting with gesture-based interfaces. On the other hand, the group of older-adults was indeed the one that exhibited the best results concerning performance with both devices. However, it was also the one group to present higher movement times, which may indicate that they have a consistent interaction. Regardless of their movement times being not as fast as the other groups, they tend to be more precise throughout the several levels of difficulty of each set of trials.

These findings also attest the results formerly debated. However, in terms of user preference, some participants felt that the Leap Motion sensor was easier to use and less demanding in terms of concentration than the Kinect. In that regard, we thus opted to continue our study with the Leap Motion instead of the Kinect, as the survey regarding the users' preference pointed us in that direction.

4.1.2.4 Summary

In this second study, we intended to understand which input modality could positively influence the indices of performance of a specific target-audience in a continuous selection task, considering Fitts' Law. Comparing these results with the ones registered on the first phase, we may presume that the creation of distinctive user profiles regarding interaction with different interfaces could, indeed, be important.

Indeed, the experiment presented considerable differences within each group when resorting to distinctive interaction modalities to perform continuous selections tasks. Regarding the movement times, as well as performance indices of each group, we have realized that when undertaking continuous selection tasks, not only age but also the choice of the input modality could induce significant differences in interaction. In a way, this contradicts the previous results of the usability evaluation, which may suggest that discrete and continuous selection tasks may also influence the users' performance. When accomplishing discrete tasks, users are less volatile to the interface used, as opposed to continuous tasks, where they seem to show more susceptibility to the input modality.

4.2 Evaluation of insertion tasks

After analyzing the results obtained from our study regarding selection tasks, we came to the conclusion that a simple shift between discrete to continuous tasks could influence interaction when it comes to users of different age groups.

We think it is, thus, important to understand if the diverse input modalities may also cause significant differences in interaction regarding elemental insertion tasks. We intended to compare four different input devices (the physical keyboard, touch, tangibles and gestures) in terms of speed, error rate and user preference.

4.2.1 Materials and Methods

With this study we intend to recognize: (1) the relation users may have with the different interfaces on a content insertion activity; (2) if age could affect interaction with the system. Our goal is to compare results between age groups, as well as between interfaces.

Apparatus:

The conditions of the experiment were similar to the ones adopted in our previous studies, apart from the fact that, instead of using the Microsoft Kinect as our gesture-recognition input device, we used Leap Motion, due to our feedback on the case study where we compared the two devices side by side.

Therefore, we conducted this study in a closed room, and the tests were performed in a specific setup assembled for the purpose of this research (Figure 4.12). The system (Figure 4.13) consisted on: a 22" touchscreen placed in front of the user, with a resolution of 1280x800 pixels; a Leap Motion sensor placed on top of the desk, between the user and the screen, and facing upwards; a physical keyboard placed on top of the desk and in front of the user, next to the leap motion; a webcam; and 10 tangible pieces located in front of the screen, following the specific order of the numbers (from 1 - 0) and with even spaces in-between them.

Materials:

A purpose-built application was created with support for all modes of interaction, in order to keep the testing environment as coherent as possible. The tasks' software was developed in Python with the support of: the Kivy Framework, the open-source computer-vision Framework reacTIVision (to track the fiducials) and the Leap Motion SDK. Also, in order to reduce the impact of the system's feedback across the different input modalities, the same adjustments of the previous evaluation regarding selection were made to this application.

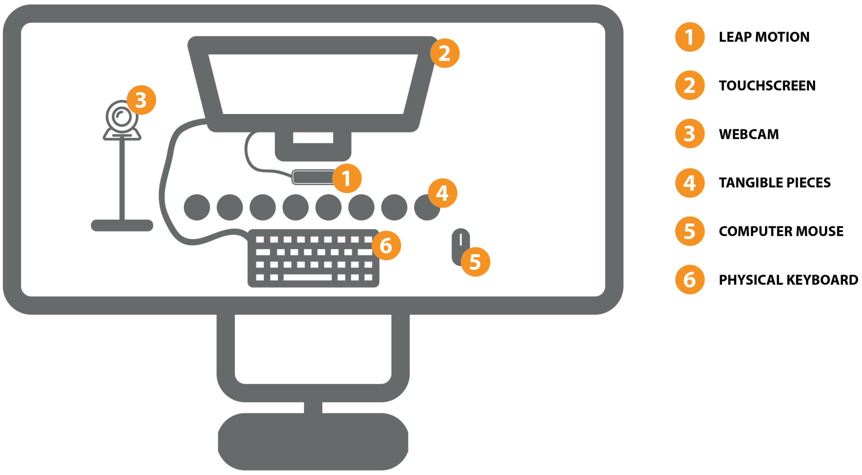


Figure 4.12 – Experimental setup.

In order to reduce the impact of the system's feedback across the different input modalities, identical adjustments were made to this application as on the previous usability evaluation: the gestural-based pointing illustration was a target badge grounded on the validated "point and wait" strategy (Schapira e Sharma, 2001) for selection; and the touch input underwent a slight adjustment made to the cursor position on the screen in order to avoid the occlusion effect (Vogel e Baudisch, 2007).

At the beginning of each test, we performed a minor survey to determine the participants' previous experience with the interfaces at issue. Also, at the end of each test we proposed a questionnaire with closed-end questions, qualitative Likert Scales and ranking lists, in order to understand the user's preferences and their views regarding: ease of use of the input modalities, ease of learning, fatigue effect, naturalism of interaction, level of user comfort/frustration, and user's degree of presence and concentration.

Figure 4.13 – Setup assembled for the insertion task.

Experimental design:

We dealt with two independent variables in this study: the age groups; and the input modalities. In turn, our dependent variables were: effectiveness, efficiency and user preference. Thus, we followed two methods of evaluation: a quantitative one with specific performance metrics -number of errors and efficiency - and a qualitative one through user preference and observational analysis of the participants' attitudes.

Altogether there were 720 experimental trials during the study. Each of the 60 participants completed four required tasks randomly, three times each, which gives a total of 12 trials per subject. Regarding the evaluation within groups using the different input modalities, we used a repeated-measures within-participants design, whereas regarding the evaluation between age groups we resorted to a repeated-measures between-participants design. This is due to this study having two independent variables to study.

We have approached how we handled outliers in Chapter 3 of this document.

Participants:

We invited 60 participants who were naïve to the purpose of the experiment, and distributed them into three age groups:

- **Children:** 20 participants of 9 to 11 years old ($M = 10.3$; $SD = 0.71$).
- **Young adults:** 20 participants of 22 to 30 years old ($M = 25.7$; $SD = 3.14$).
- **Older-adults:** 20 participants of 47 to 58 years old ($M = 49.7$; $SD = 3.86$).

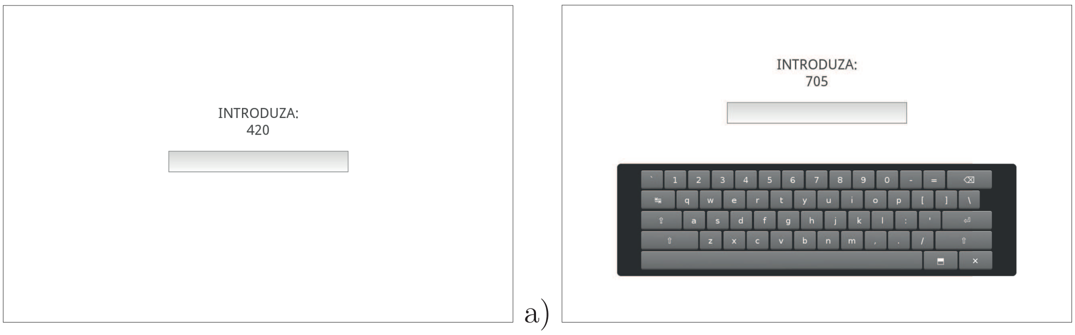
These individuals were selected according to the directives stipulated on Chapter 3. Of all of the participants, four were left-handed. All the participants reported normal or corrected to normal vision.

Procedure:

Firstly, in the beginning of each test, we provided the users with an overview about the range of available input modalities: the keyboard, touch, tangibles and gestures. Participants were instructed about the goal of each activity at hand: correctly insert a code provided on the screen in the beginning of each trial. The aim of each task was to use one of the four input modalities available (keyboard, touch, tangibles, gestures) to insert the requested content.

We divided the task into four random activities, each making use of one interaction modality. Over the course of the experiment, the researcher was always present in order to explain any doubts that could occur and respond to any questions the participants could have. Also, we acknowledged training trials in order for the participants to adapt to the interface and understand their reaction time.

Similar to the previous usability evaluation of a selection task, each activity would start with a countdown from 3 to 1 to prepare the participant for the test, and then an input field would appear on the screen, along with a numerical code. When using touch or gestures for interaction, a soft-keyboard would also appear on the bottom of the screen (Figure 4.14).



a) b)

Figure 4.14 – (a) Layout of the screen when using the keyboard and the tangible pieces; (b) Layout of the screen with the soft-keyboard when using touch and gestures.

Participants were asked to insert the provided code using one of four input modalities, according to what was requested, as quickly as possible while avoiding any errors. The four activities were randomly placed.

As such, touch and gestures made use of the soft-keyboard in the screen to insert the numbers. On the other hand, regarding the graphical interface, users would insert the code with the physical keyboard, and with the tangible interface they used the numbered pieces on top of the table. In this case, when a participant made an error and inserted the wrong digit, they could delete it using a specific black piece. All the other pieces were white, as illustrated in the previous Figure 4.13.

When a digit was inserted, it would appear inside the text field and the user would keep typing until the three numbers were inserted. The user did not need to select any buttons to submit, as we wanted this task to remain elemental and not composed in order to control the variables of the experiment. When the last digit was inserted, the system would automatically check if the code was correct. If so, it would automatically advance to the next trial, until it reached the third and last trial, and then a “successfully completed” message would appear on the screen, after which the user could continue to the next activity. If the code was incorrect, the numbers inside the text field would turn red and the trial would not advance until the code was corrected. For that, participants could use the “delete” button on the soft and physical keyboards. We registered an error each time the delete button would be used.

There were no time constraints to successfully complete the tasks, since this could cause stress and influence performance. However, the participants were asked to complete the activities as fast as they could. Also, the tests could be paused at any moment between the tasks in order for the participants to rest. At the end of the tests, participants were asked to fill in a survey about their preferences.

4.2.2 Results

Below are presented the results regarding the usability evaluation of a basic insertion task: effectiveness (error rate and completion success), efficiency (completion time) and satisfaction are presented and analyzed accordingly. The results are divided into subsections for clarity of presentation. In this case, each experimental session lasted approximately 15 minutes.

Effectiveness:

We considered an error when a wrong number was inserted in the system. Figure 4.15 presents the errors registered during the experiment. In total, we registered 102 errors: 3 with the physical keyboard as the input device, 6 using touch, 10 using tangibles and 83 resorting to gestures for content insertion, being most of this errors committed by older-adults.

Efficiency:

We analyzed the completion time within each group of participants and also between them, considering the different input modalities. The task completion time was measured between the appearance of the input field displayed on-screen and the

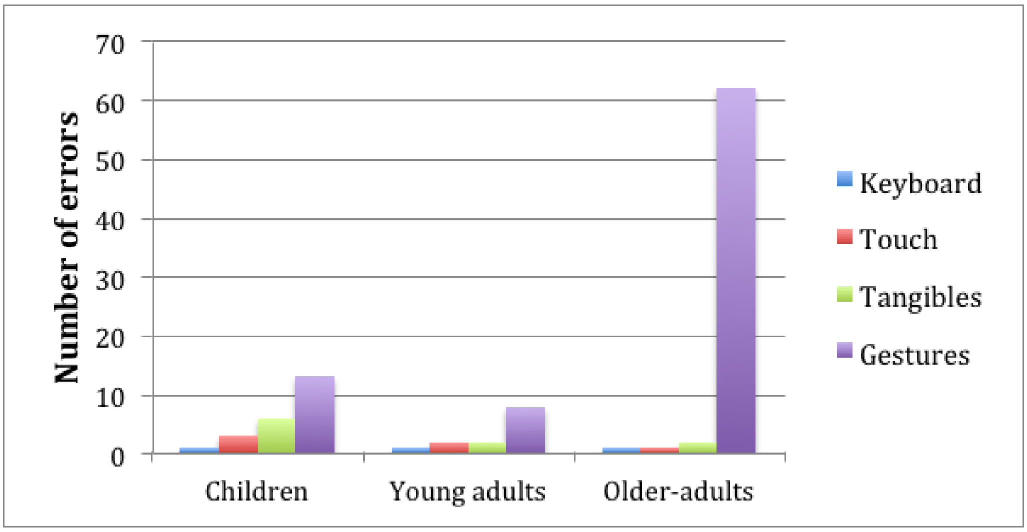


Figure 4.15 – Number of errors each group committed per interface.

successful insertion of the code. Table 4.4 shows each group's average completion times, the maximum and minimum time, as well as the standard deviation (in seconds) for the three trials with the four different interface conditions.

In terms of completion times recorded, the input modality that registered the fastest mean results throughout all the three groups was the graphical interface: children (1.28 s); young adults (0.61 s); and older-adults (0.95 s). On the other hand, the gestural interface registered higher mean completion times: children (2.91 s); young adults (1.81 s); and older-adults (6.09 s). The lowest time recorded was by young adults using the keyboard (0.21 s) and the highest time was registered by older-adults using gestures (19.85 s). Notice also that the mean minimum time recorded using gestures (0.69 s) was actually very close to the mean maximum value documented using the keyboard (0.96 s).

Figure 4.16 presents the tasks' completion times recorded during the experimental tests with the different interfaces by each age group.

Overall, the children's group presented the highest mean times in most of the interfaces, comparing to the other groups, except regarding the gestural input modality. In contrast, the group of young adults performed better with all interfaces and thus

Table 4.4 – Each group's mean, minimum, maximum and standard deviation times for the insertion task regarding the four input modalities (in seconds).

Groups	Interaction mode	Mean time	Maximum time	Minimum time	Standard deviation
Children	Keyboard	1.28	3.27	0.24	0.50
	Touch	1.61	6.62	0.28	0.60
	Tangibles	2.41	5.15	0.20	0.78
	Gestures	2.91	8.23	0.96	2.80
Young adults	Keyboard	0.61	0.96	0.21	0.15
	Touch	0.92	1.74	0.34	0.23
	Tangibles	1.62	2.57	0.34	0.31
	Gestures	1.81	2.85	0.69	0.38
Older-adults	Keyboard	0.95	1.91	0.28	0.34
	Touch	1.26	2.83	0.44	0.38
	Tangibles	2.06	3.33	0.84	0.42
	Gestures	6.09	19.85	0.95	6.81

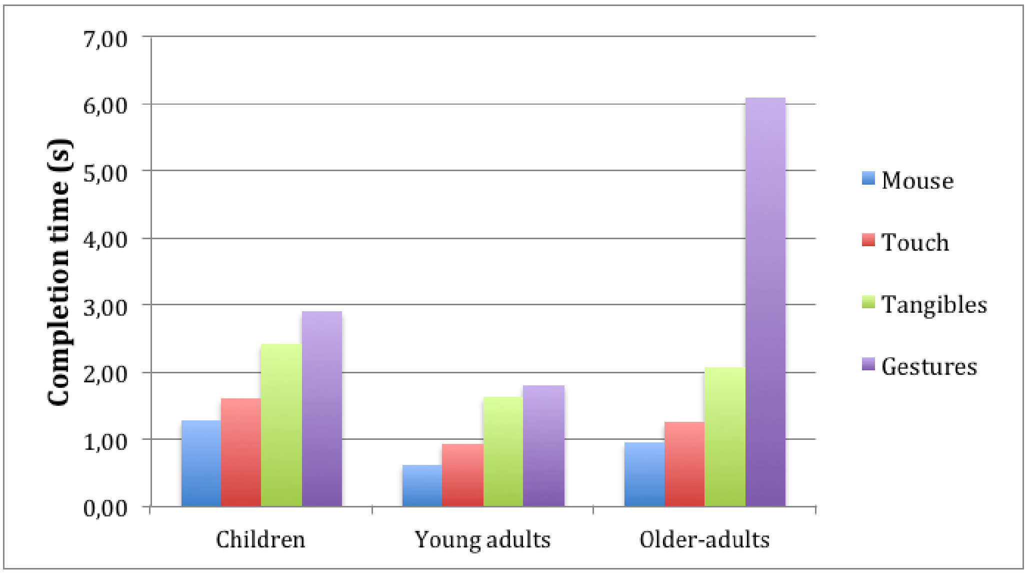


Figure 4.16 – Mean time taken to complete the task (in seconds).

showed the lowest mean completion times. It may be also important to highlight the completion time recorded by the older-adults with the gestural interface, which was about twice the children's group.

Nevertheless, the mean selection time for the several input modalities did follow a pattern throughout the three user groups: (1) the gesture-based interface was not as efficient as the rest, especially considering the group of older-adults; and (2) using the physical keyboard was the fastest. It is noticed that the soft-keyboard was the second best regarding completion times, but for an elemental insertion task the use of a physical keyboard appears to be more efficient amongst the three groups in terms of mean time.

We assessed the normality of data with the Shapiro-Wilk Test in order to understand if the data were normally distributed and could thus be considered for statistical analysis. We elected this test over the Kolmogorov-Smirnov due to the size of our sample. Except for the gesture-based interface used by the older-adults ($p =$ .008), the data were normally distributed. Indeed, there was a statistically significant difference between all the groups and each interface as determined by one-way ANOVA: graphical interface ($F(2, 57) = 17.539$, $p = .000$); touch interface ($F(2, 57) = 12.526$, $p = .000$); tangible interface ($F(2, 57) = 10.821$, $p = .000$); gestural interface($F(2, 57) = 16.053$, $p = .000$).

Regarding the use of the physical keyboard, a Tukey post-hoc multiple comparisons test showed that the mean completion times were significantly different between children and young adults ($p = .000$), children and older-adults ($p = .012$), young adults and older-adults ($p = .013$). As for the soft-keyboard, the results were also significantly different: children versus young adults ($p = .000$), children and older-adults ($p = .037$), young adults and older-adults ($p = .042$).

However, not all the groups showed a significant difference in the results considering the tangible and gestural interfaces: the tangibles only presented a significant difference between children and young adults ($p = .000$), and young adults versus older adults ($p = .032$); and a not significant difference between the children and older-adults ($p = .110$). As for the gesture-based interface, the results were significantly different between the group of children and older-adults ($p = .000$), and young adults versus older-adults ($p = .000$). On the other hand, the children's group and the young adults did not present statistically significant differences regarding the use of the gestural interface ($p = .349$). To sum up, there are not significant differences between the groups of children versus older-adults regarding the tangible interface; and between the children's group versus the young adults regarding the gestural interface.

Additionally to comparing the performance between each group regarding the different interfaces, we also analyzed the performance within each group resorting to a within-participant repeated-measures design.

A repeated measures ANOVA with a Greenhouse-Geisser correction determined there was a significant difference between the mean time taken by children to complete the task in each interface ($F(2.393, 45.470) = 23.493$, $p < .000$). The input device with the fastest results was the physical keyboard (1.28 s), followed by touch (1.61 s), tangibles (2.41 s) and finally gestures (2.91 s). Post-hoc pairwise comparisons tests using the Bonferroni correction showed that, except for the physical versus soft keyboard as input ($p = .305$) and tangibles versus gestures ($p = .534$), the mean times were statistically significantly different throughout the other comparisons. Using the graphical interface against tangibles ($p = .000$) and gestures ($p = .000$) was indeed significantly faster; as was the touch interface against these same interfaces: tangibles ($p = .000$) and gestures ($p = .000$). To sum up, the physical and soft keyboards display significantly faster results for insertion tasks regarding the group of children, but do not show much of a discrepancy between each other. Although the graphical interface presents a difference of 1.28 seconds comparing to touch, it is not significant.

As the Mauchly's sphericity was not assumed, like in the children's group, we applied the Greenhouse-Geisser adjustment to the repeated measures ANOVA. This group also presented statistically significantly different completion times ($F(2.381, 45.241) = 95.497$, $p < .000$). Post hoc tests using the Bonferroni correction revealed that the mean times returned by the young adults using the tangible interface compared to the gesture-based one (1.62s and 1.81s, respectively) were not statistically significant ($p = .491$). However, there were significant differences in using the other interfaces ($p = .000$), being the graphical interfaces the fastest for this insertion tasks (0.61 s), followed by touch (0.92 s), as observed in the group of children.

The older-adults' efficiency also achieved a statistically significant difference between the mean times in each interface ($F(1.034, 19.643) = 26.766$, $p < .000$). Although the analysis returns a statistically significant difference regarding the use of tangibles versus gestures ($p = .001$) and the other interfaces ($p = .000$), using the physical or soft keyboard is near to not being significantly different ($p = .045$). In this situation, more tests would have to be performed to really understand this correlation. The graphical interface scored the lowest mean time (0.95 s), followed by touch (1.26 s), tangibles (2.06 s) and gestures (6.09 s). Indeed, the gesture-based interface presented steeper results, proving to be the worst option for older-adults when it comes to insertion tasks.

Participants' attitudes and preferences:

Overall, there were no significant events that could be considered relevant during this study. When using the physical and the soft keyboards, all of the participants showed a regular approach of looking for the right key and pressing/touching it. Regarding the tangibles, all participants understood the procedure and had no difficulty with the pieces, and gestures, in spite of being considered more challenging by the participants, did not reveal problematic behaviors.

However, using gestures as the input modality, the group of older-adults did present a tendency: they would put the arm extremely close to the screen, almost as if they were going to select the key with touch, and so the occlusion effect would occur. Also, they would read the numbers that were asked for with the hand over the soft-keyboard and thus an unintentional selection was made.

We noticed another curious behavior in this group. When using the physical keyboard, they would read all of the numbers on the screen and then type them sequentially, all three at the same time. On the other hand, when using gestures, they would look at the screen and type the digits one by one, thus causing more errors and false insertions, that is, they would read one digit, then type, then read another digit, then type it, and finally read the last digit and type it also.

Another aspect worth noting is that both children and older-adults tended to mix up the pieces when using them to insert the digits, that is, they would not return the piece to its original place after using it, and so they consequently got confused over time. However, this was the most relevant interface for children, that responded in the final questionnaires that this input modality was their favorite for insertion tasks. Indeed, for this age group, gestures were relatively difficult, whereas tangibles and the mouse were considered relatively easy to use. Touch was the most easy input modality to use, with 40% of the children saying it was extremely easy to use. This indicates that, although this age group admits that tangibles are harder to use than touch, they still prefer it for insertion activities.

Regarding the other two groups, the participants thought the task was relatively or extremely easy to use resorting to all the input modalities, except for gestures.

4.2.3 Discussion

Being the group of older-adults the one to register the higher error rate in most of the available interfaces and also the higher mean times of all the groups, specially using gestures, this situation could imply that over the years the users' age might decrease the dexterity. This situation was also visible in selection activities.

The difference in efficiency between children versus young adults using the touch interface was higher than when comparing with the other groups, which may imply that they tend to have more disparate performance results regarding interaction with the soft-keyboard.

We have reached the conclusion that for elemental insertion tasks, the use of different interfaces may also influence the users' performance and preference, such as in selection activities. The pattern for the interface that enables the fastest or the slowest completion times appears to be constant (being the physical keyboard the fastest and gestures the slowest), although displaying significantly different times depending on the interface. This may imply that the type of interface that is used can influence efficiency in insertion tasks, more even than other factors like age.

Indeed, our findings showed that: (1) more errors were detected when using gestures, regarding all age groups; (2) the physical keyboard proved to be more efficient, followed by the soft-keyboard using touch; (3) although tangibles proved not to be as efficient as other interaction modalities, children had a great response in terms of preference, which may suggest that this type of input could be interesting to use with this age group not because of efficiency results, but because this group had such an empathy with the pieces.

4.2.4 Summary

With this analysis, we wanted to understand which input modality could induce better results considering basic content insertion. We believe that we have indications that the type of interface that is used can influence efficiency in insertion tasks, and not so much other factors like age.

Indeed, the graphical interface was the best in these specific insertion tasks and all groups performed better with the physical keyboard. However, we could observe a curiosity regarding this study, as we did with the selection task: children are not as efficient with tangibles, but they chose it to be their favorite, which could eventually influence and improve their interest in the task at hand.

4.3 Evaluation of manipulation tasks

It is increasingly important to understand which input device could best suit the users' needs regarding different contexts. We often need to perform trajectory-based tasks, such as drawing or navigating. These actions are part of an elemental task of manipulation of content, where the user needs to perform a continued action with an input device in order to complete an activity.

The Steering Law can be applied as an evaluation paradigm complementary to Fitts' law, to understand if the different input modalities can influence interaction in manipulation tasks based on trajectory. More information on this well-known human performance evaluation model can be found on Chapter 3.

We expect these results can offer us an insight on: (1) how users interact with the different input modalities in a trajectory-based task; (2) if age could play any role in their performance and thus influence the results.

4.3.1 Materials and Methods

In this study we intend to analyses movement times of the different age groups and the respective index of performance in two different steering tasks: one circular and one linear.

As such, we intend to compare three different interaction modes (the computer mouse, touch and gestures). In this phase we did not cover tangibles as they are not a pointing device and thus could not be considered.

Apparatus:

The conditions of the experiment were identical to the ones adopted in our previous studies. The only difference in the setup was that neither tangibles nor the physical keyboard were used. Instead, a computer mouse was available for the activity (Figure 4.17).

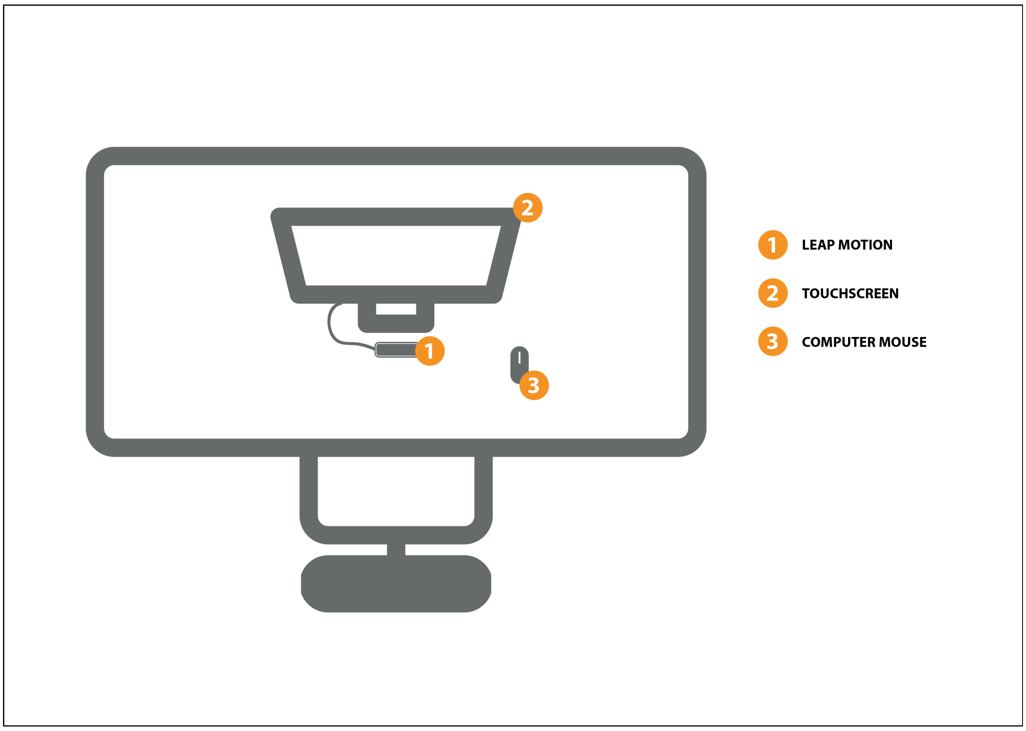


Figure 4.17 – Experimental setup.

Materials:

A purpose-built application was created with support for all modes of interaction, using the same frameworks and programming language as the previous studies, and also with the same adjustments made to the system's feedback across the different input modalities.

This study makes use of two different trajectory-based tasks: one circular and one linear (Figure 4.18). These tasks were based on the Steering Law, and we wanted to use these specific shapes as they are valid to simulate common linear and non-linear movements when manipulating an object, or even during navigation: the straight tunnel (linear task) is meant as a way to imitate navigation in hierarchical menus, perform an action of resizing an objects, or dragging data across the screen. On the other hand, the circular tunnel mimics the ability to move along curved paths to perform manipulation activities that require more coordination in multiple dimensions. These curve-based paths are important to analyze because as the radius of the circle changes, so does its curvature, and thus we can cover a wide range of curvature trajectories. Also recall that on Chapter 3 we better explained this human performance evaluation model.

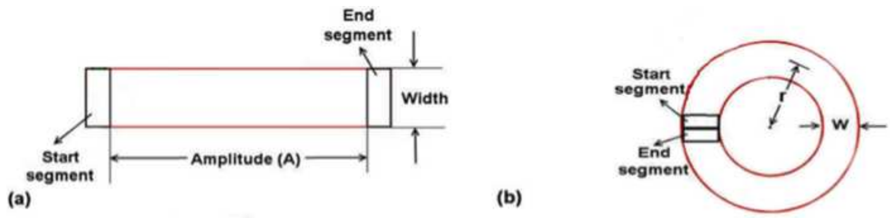


Figure 4.18 – Two steering tasks: (a) Linear path; (b) Circular path.

Like the previous studies, during the tests we gathered qualitative observational analysis on the participants' attitudes. Afterwards, we asked the participants to complete a survey with closed-end questions, qualitative Likert Scales and ranking lists, in order to understand the user's preferences considering: ease of use of the input modalities, ease of learning, fatigue effect, naturalism of interaction, level of user comfort/frustration, and user's degree of presence and concentration.

Experimental design:

We dealt with four independent variables in this study: the age groups; the input modalities; the indices of difficulty; and the steering path types. In turn, our dependent variables were: effectiveness, efficiency, index of performance and user preference.

We used 2 different amplitudes: A = 600, 1000; and three different target widths: W = 60, 100, 160 pixels; which provide six levels of difficulty, from 3.75 bits to 16.67 bits. We did not create tunnels under 60 pixels of width because it has been proven that this should be the minimum value when using touch as an input modality (Yin et al. 2015). Also, the order of the tasks according to the several indices of difficulty was randomized. So, the experiment conditions were the following: 60 participants, three input modalities (mouse, touch, gestures), two path types (linear and circular), 3 repetitions per index of difficulty, and six indices of difficulty (3.75, 6, 6.25, 10, 16, 16.67 bits). Thereby, altogether there were 36 experimental conditions and an overall of 6 480 correct trials.

Regarding the evaluation within groups using the different input modalities, we used a repeated-measures within-participants design, whereas regarding the evaluation between age groups we resorted to a repeated-measures between-participants design. Also, we have approached how we handled outliers in Chapter 3 of this document.

Participants:

The study included 60 volunteers who were naïve to the purpose of the experiment. We divided them into three groups, according to their age, and we thereby had 20 users for each age group participating in the tests:

- **Children:** 20 participants of 8 to 12 years old (M = 10.6; SD = 1.24).
- **Young adults:** 20 participants of 20 to 30 years old (M = 25.1; SD = 4.38).
- **Older-adults:** 20 participants of 45 to 55 years old (M = 49.1; SD = 4.14).

These individuals were selected according to the directives stipulated on Chapter 3. Of all of the participants, two were left-handed. All the participants reported normal or corrected to normal vision.

Procedure:

Before the test began, all participants were allowed to practice a set of trials of each level of difficulty for as long as they wanted, since they needed to get used to the input modalities and gain adequate practical skill.

Subjects would perform two types of steering tasks: a linear (straight) task and a circular one. The entire procedure described below was repeated for each interaction mode.

Firstly, participants would execute the linear path task, as the level of difficulty was lower, giving the participants time and experience before accomplishing the circular path. The goal was to begin drawing a path on one end of a tunnel and finish it on the other end, from left to right, attempting to do this throughout different levels of difficulty (Figure 4.19). The initial and end position would be highlighted with the green color, and the tunnel to follow was a light grey. Also, there was a visual feedback to indicate the path being drawn, being this path yellow (19/b and 19/d). When reaching the end of the path, the yellow line turned dark green to indicate the participant that the trial had ended and automatically advanced to the next one. During the test, if the path went out of the boundaries of the tunnel before the completion of a trial, an error was recorded and the trial would restart. The time would only start counting when the user crossed the green area in the starting position, and would stop recording when the user got to the final position. At the end of each interface's trial, the layout disappeared and a "successfully completed task" message appeared on the screen.

Secondly, participants were asked to complete the circular steering task. Likewise, for all levels of difficulty, the goal was to draw a path from one specific point and then circle inside the tunnel until reaching the initial position again (Figure 4.20). The initial/end position would be highlighted in green, and the tunnel to follow was

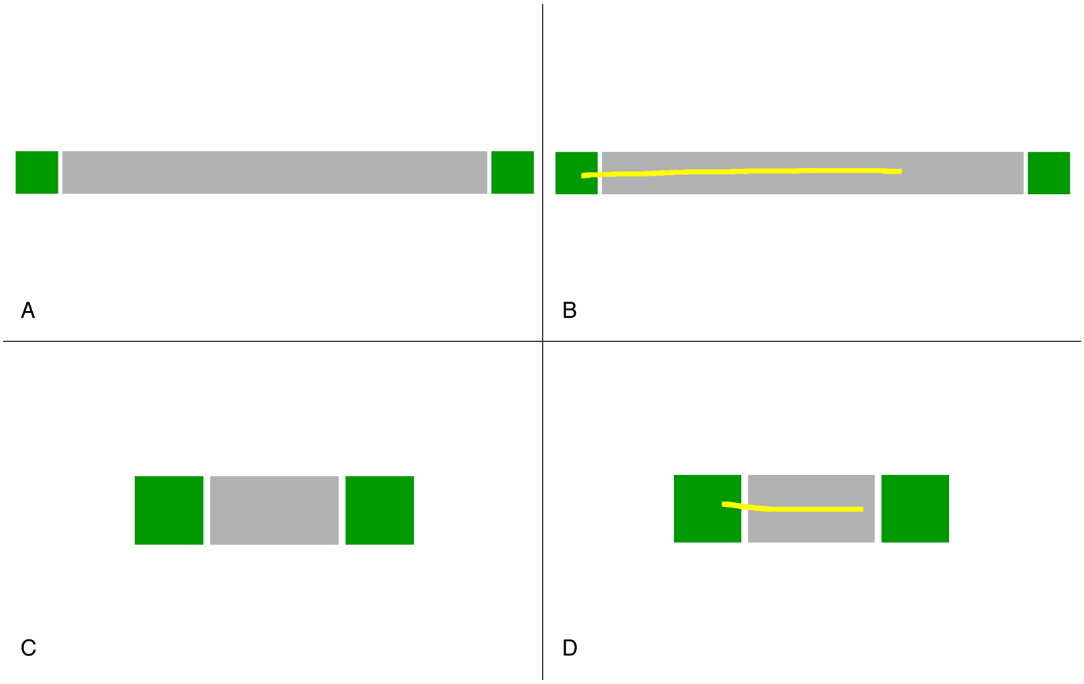


Figure 4.19 – Example of two linear steering trials: (a / c) two levels of difficult as they would appear on the screen; (b / d) the path drawn by the user from an initial to an end point.

a light grey, just like the linear task. Also, if the path went out of the boundaries of the tunnel before the completion of a trial, an error was recorded and the trial would restart. The same visual feedback from the linear steering is used and the same procedure on task completion was also maintained.

Considering the computer mouse and touch, the interaction process was standard: a mouse click with the cursor and a finger touch on the screen on top of the initial highlighted position. The participants were asked to choose their dominant hand and use the index finger to interact with the touchscreen. For gestural selection, the process was identical to the one implemented on previous studies, described earlier on this chapter.

There were no time constraints to successfully complete the tasks, but the participants were asked to complete them as fast as they could and minimize the errors. Every time the path drawn hit or crossed the borders of the tunnel, an error was recorded and the trial was cancelled, forcing the user to return to the initial position

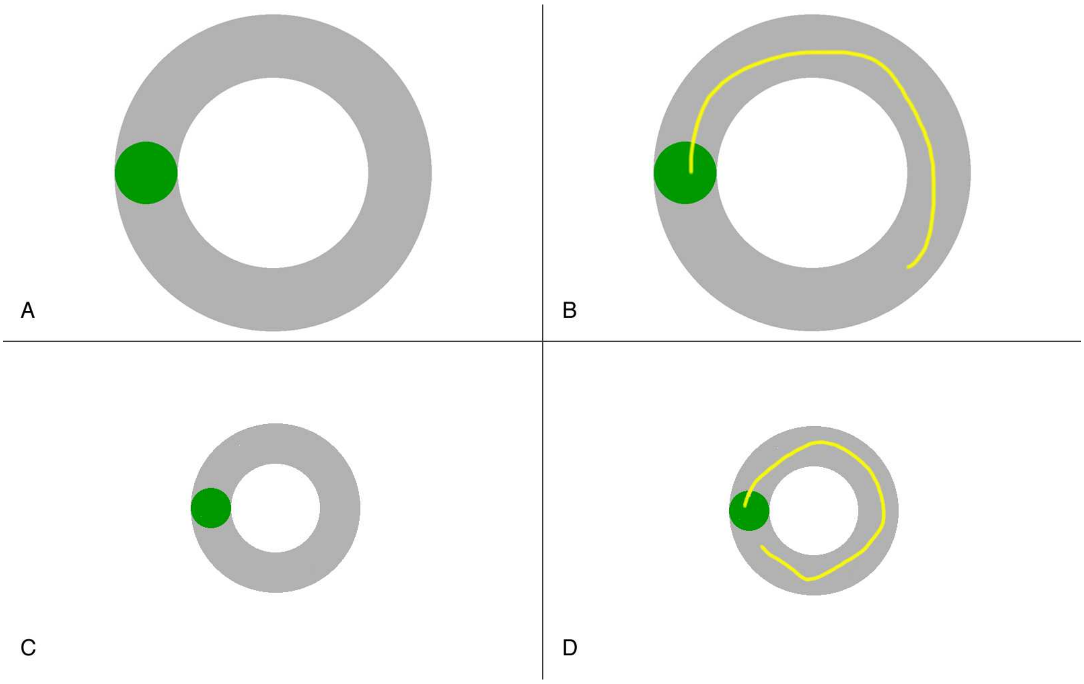


Figure 4.20 – Example of two circular steering trials: (a / c) two levels of difficult as they would appear on the screen; (b / d) the path drawn by the user from an initial to an end point.

and restart the path. The time would also restart counting. Whenever the participants felt tired, they were allowed to rest between the levels. After the experiment, participants were asked to fill a questionnaire with the purpose of gathering their opinion about the different interaction modalities.

4.3.2 Results

This section presents the results and discussion of the experimental tests for both tasks (linear and circular). Steering times, number of errors and indices of performance of each age group regarding the different interaction modalities are presented and compared. Also, we examine the participants' attitudes during the tests and their preference towards the interfaces. In this case, each experimental session lasted approximately 20 minutes.

Effectiveness:

For the linear steering task, regarding children, error rates were on average 22.75% when using the mouse, 10.00% with touch, and 31.56% with gestures. Young adults had an error rate of 4% when using the mouse, 2.44% with touch, and 19.46 with gestures. On the other hand, older-adults presented an error rate of 9.09% with the mouse, 2.17% with touch, and 33.82% with gestures. As one can see, on straight movements, the interaction modality that registered fewer errors was touch, and the age group to have fewer faults was the older-adults.

For the circular steering path: children registered a 37.5% error rate using the mouse, 10.22% with touch and 80.1% with gestures; the young adults registered a 7.22% error rate using the mouse, 3.49% with touch and 35.71% with gestures; and older-adults registered a 12.41% error rate using the mouse, 6.98% with touch and 59.78% with gestures. Indeed, touch was again the interaction mode to present fewer errors.

Efficiency:

We present the steering times recorded during each trial of the experimental tests. First and foremost, we examine both linear and circular times recorded by each group of participants, and then we compare the age groups in order to determine if there is a significant difference regarding the interaction modalities.

Concerning the linear steering task, children took longer to complete the task with the mouse than gestures, taking respectively 1.56 s versus 1.39 s (Figure 4.21), very closely followed by touch (1.37 s). The best times were recorded by the young adults in all input modalities, respectively mouse (0.80 s); touch (0.79 s); gestures (1.20 s). Also, older-adults were the ones that took longer to complete the task with gestures (1.98 s), but on the other hand they tend to be faster than children with touch and the computer mouse, respectively 1.15 s and 1.53 s.

A Kolmogorov-Smirnov test showed that the data were normally distributed. As such, we performed a one-way ANOVA test and determined there was a statistical difference in steering times: mouse ($F(2,57) = 14.649$, $p = .000$); touch ($F(2,57) =$

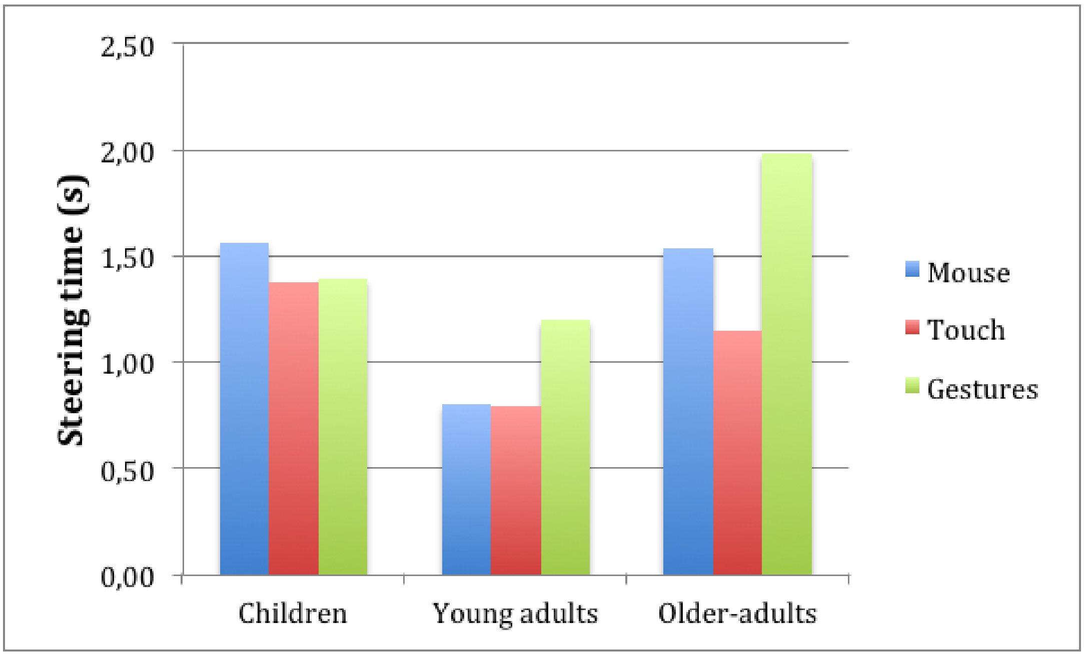


Figure 4.21 – Average steering time of the three age groups for the linear task.

10.104, $p = .000$); gestures ($F(2{,}57) = 12.257$, $p = .000$). Tukey post-hoc comparison tests unveiled that children did not presented significant differences using the mouse and touch when comparing to older-adults ($p = .997$; $p = .156$; respectively), but was indeed significantly different comparing to young adults ($p = .000$). Moreover, young adults versus older-adults performed with statistical differences with all modalities: mouse ($p = .000$), touch ($p = .031$) and gestures ($p = .000$); as did children compared to older-adults using gestures ($p = .001$). However, young adults compared to children did not show significant differences when using gestures ($p = .423$).

Concerning the circular steering task, Figure 4.22 illustrates the completion times of all groups when using the different input modalities. Gestures registered the highest mean values with all age groups: children (5 s), young adults (4.69 s), older-adults (5.31 s). On the other hand, touch presented the fastest steering times: children (2 s), young adults (1.74 s), older-adults (2.09 s); followed by the computer mouse: children (3.13 s), young adults (2.14 s), older-adults (2.76 s). As can be seen by the chart, the young adults managed to complete the task faster than the other groups with all of the interaction modalities.

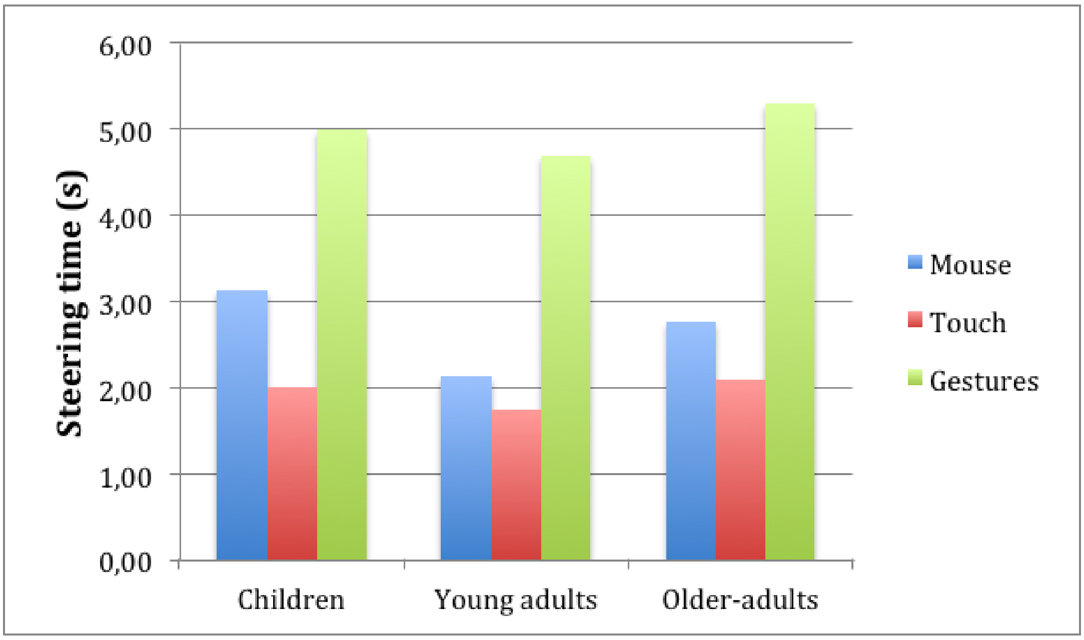


Figure 4.22 – Average steering time of the three age groups for the circular task.

An independent-samples Kruskal-Wallis 1-way ANOVA test indicated that touch ($\chi2(2) = 3.866$, $p = .145$) and gestures ($\chi2(2) = 2.146$, $p = .342$) did not significantly influence interaction. We performed this non-parametric analysis since the Kolmogorov-Smirnov test indicated that the results were not normally distributed. On the other hand, using the mouse recorded significant differences in interaction between the groups ($\chi2(2) = 15.280$, $p = .000$). Moreover, a post-hoc multiple comparisons test showed that the steering times with this device were statistically different between young adults and older-adults ($p = .018$), and young adults and children ($p = .000$); however, not between children and older-adults ($p = .912$).

Additionally to comparing the steering times between each group, we also analyzed them within each group.

Firstly, we present the result analysis of the linear tunnel task, and secondly the circular tunnel one. Children were on average 12.18% faster using touch rather than the computer mouse, but only a merely 1.44% better comparing to gestures. A Friedman test showed a statistical difference in their steering times ($\chi2(2) = 7.620$, $p = .022$). We were aware of a significant difference between the modalities mouse versus touch (Z = -2.502, p = .012) and gestures (Z = -2.110, p = .035), specified by a post hoc analysis with Wilcoxon signed-rank tests with a Bonferroni correction. However, there were no significant differences between gestures and touch (Z = -.242, p = .809).

As for young adults, there was a 33.33% increase in the steering time when using gestures compared to the mouse, and 34.17% comparing to touch. Also, there was only 1.25% improvement in interaction when using touch against the mouse. Indeed, the Friedman test showed a significant difference in the mean steering time when comparing the use of the three input modalities ($\chi 2(2)$ = 16.025, p = .000). After a post hoc analysis with Wilcoxon signed-rank tests with a Bonferroni correction applied, we could observe no statistical difference between touch and mouse (Z = -.411, p = .681). However, there was significance with gestures versus touch (Z = -3.342, p = .001) and mouse (Z = -3.562, p = .000).

Considering older-adults, the touch had an improvement of 41.92% in speed with regard to steering times when compared to gestures and 24.84% against the mouse. As before, the Friedman test showed significant differences in mean times ($\chi 2(2)$ = 33.899, p = .000). Post hoc analysis with Wilcoxon signed-rank tests with a Bonferroni correction also indicated statistical differences between: touch and mouse (Z = -3.548, p = .000), and gestures and touch (Z = -3.921, p = .000), and mouse and gestures (Z = -3.461, p = .001).

On a different subject, regarding the circular steering motion, children showed differences in interaction with the several input devices ($\chi 2(2)$ = 33.899, p = .000), as presented by a Friedman test. Indeed, touch displayed an improvement of 60% in speed when compared to gestures, and 37.40% regarding the mouse. Post hoc analysis with Wilcoxon signed-rank tests with a Bonferroni correction also indicated statistical differences with all combination (Z = -3.920, p = .000).

Young adults demonstrated touch could be faster by 62.90% rather than gestures and 18.69% in comparison to the computer mouse. This difference in speed was significant, as shown by the Friedman test ($\chi 2(2)$ = 40.000, p = .000). Just the same, here a post hoc analysis with Wilcoxon signed-rank tests with a Bonferroni correction also indicated statistical differences with all combination (Z = -3.920, p = .000).

As for older-adults, the Friedman test showed significant differences in mean times ($\chi 2(2)$ = 38.100, p = .000). With regard to touch, they managed to complete the circular tunnel 60.64% faster than gestures and 48.02% faster than with the mouse. Post hoc analysis with Wilcoxon signed-rank tests with a Bonferroni correction proved statistical differences between: mouse and touch (Z = -3.659, p = .000), gestures and touch (Z = -3.920, p = .000), and gestures and mouse (Z = -3.920, p = .000.)

Performance:

We present the values of the Index of Performance (IP) of each group regarding the different input modalities in both linear and circular tasks, and then we analyze these values more closely within each group, in order to determine if they are significantly different.

Figure 4.23 shows the mean IPs registered during the experiment for each age group when using the three interaction modes during the linear steering task. As we can observe, young adults can manage more information per unit time than the other age groups, since they presented the best mean values of IP using the mouse (IP = 10.76 bits/s; SD = 3.88 bits/s) and touch (IP = 14.89 bits/s; SD = 6.12 bits/s). They registered high values with gestures (IP = 13.80 bits/s; SD = 7.95 bits/s), but still children performed slightly better with this modality (IP = 13.83; SD = 11.97 bits/s). For this group, gestures provided the best result and the mouse caused the worst IP value of 6.01 bits/s with a standard deviation of 2.75 bits/s, and touch presented and IP of 9.95 bits/s (SD = 6.10 bits/s), close to the IP registered by young adults but using the mouse. This means that children performed this steering task 56.54% better with gestures than the mouse, and 28.05% better than touch. The older-adults performed better than children using the computer mouse (IP = 6.39 bits/s; SD = 2.61 bits/s), but poorer when using touch (IP = 9.33 bits/s; SD = 3.16 bits/s) and gestures (IP = 5.80 bits/s; SD = 2.51 bits/s).

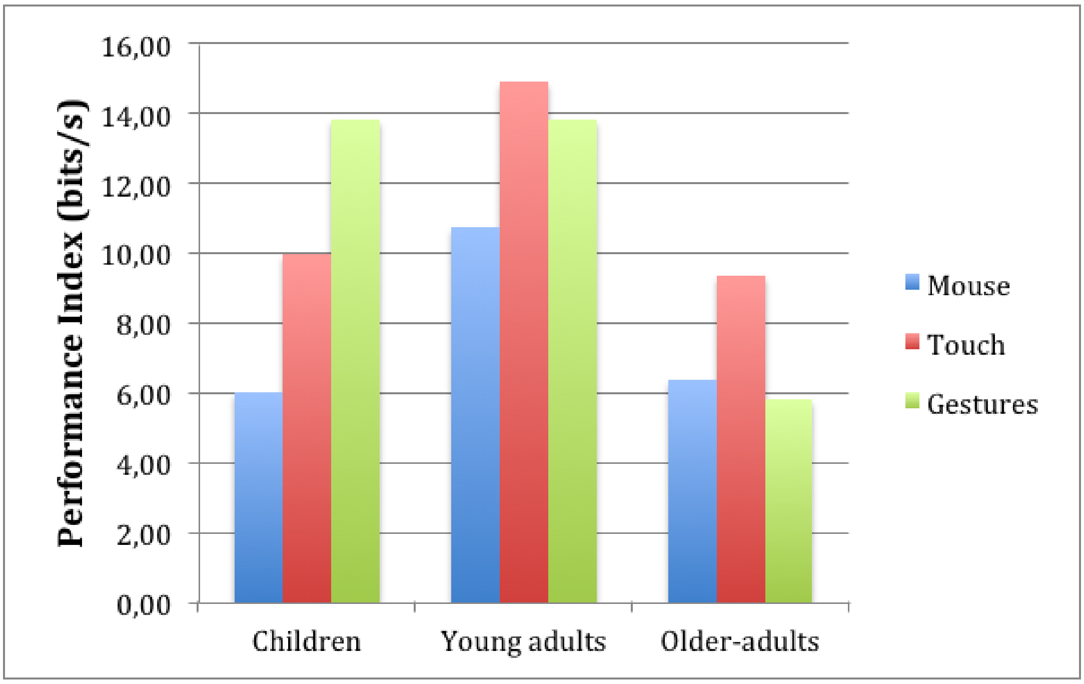


Figure 4.23 – Average IP of the three age groups for the linear task.

A Kolmogorov-Smirnov test showed that the data were normally distributed, and a one-way ANOVA test determined that there were significant differences in interaction in linear manipulation tasks: mouse ($F(2{,}55) = 13.651$, $p = .000$), touch ($F(2{,}55) = 6.258$, $p = .004$), gestures ($F(2{,}51) = 5.662$, $p = .006$). Moreover, Tukey post-hoc comparison tests revealed that children and older-adults only registered significant differences when using gestures ($p = .014$), and not with the mouse ($p = .926$) or touch ($p = .930$). As for young adults and children, they did not prove significant disparities either with gestures ($p = 1.000$), but when compared to older-adults they did ($p = .016$). All the more, young adults also differentiated from children ($p = .015$) and older-adults ($p = .006$) when using touch. Regarding the mouse, young adults also displayed statistical variances compared to children ($p = .000$) and older-adults ($p = .000$).

Regarding the circular steering task, Figure 4.24 clarifies mean IPs registered during the experiment for each age group when using the three interaction modes. In this case, touch was the one to present higher performance indices from all groups:

children (IP = 5.50 bits/s; SD = 2.30 bits/s), young adults (IP = 5.81 bits/s; SD = 1.88 bits/s) and older-adults (IP = 6.37 bits/s; SD = 2.89 bits/s).

About this task, we can observe the older-adults did have higher performance values with touch than young adults. On the other hand, gestures, contrarily to the linear task, presented the worst results in all groups: children (IP = 2.24 bits/s; SD = 0.72 bits/s), young adults (IP = 2.93 bits/s; SD = 1.17 bits/s) and older-adults (IP = 2.64 bits/s; SD = 1.08 bits/s). In this regard, young adults demonstrated to have the highest performance index and children the lowest. The same happened with touch: children (IP = 5.50 bits/s; SD = 2.30 bits/s), young adults (IP = 5.81 bits/s; SD = 1.88 bits/s) and older-adults (IP = 6.37 bits/s; SD = 2.89 bits/s). Here, a Kolmogorov-Smirnov test showed that the data were normally distributed, and a one-way ANOVA test determined that there were no significant differences in the three interaction modalities regarding circular manipulation tasks: mouse ($F(2{,}54) = 3.029$, $p = .057$), touch ($F(2{,}56) = 0.678$, $p = .512$), gestures ($F(2{,}55) = 2.194$, $p = .121$).

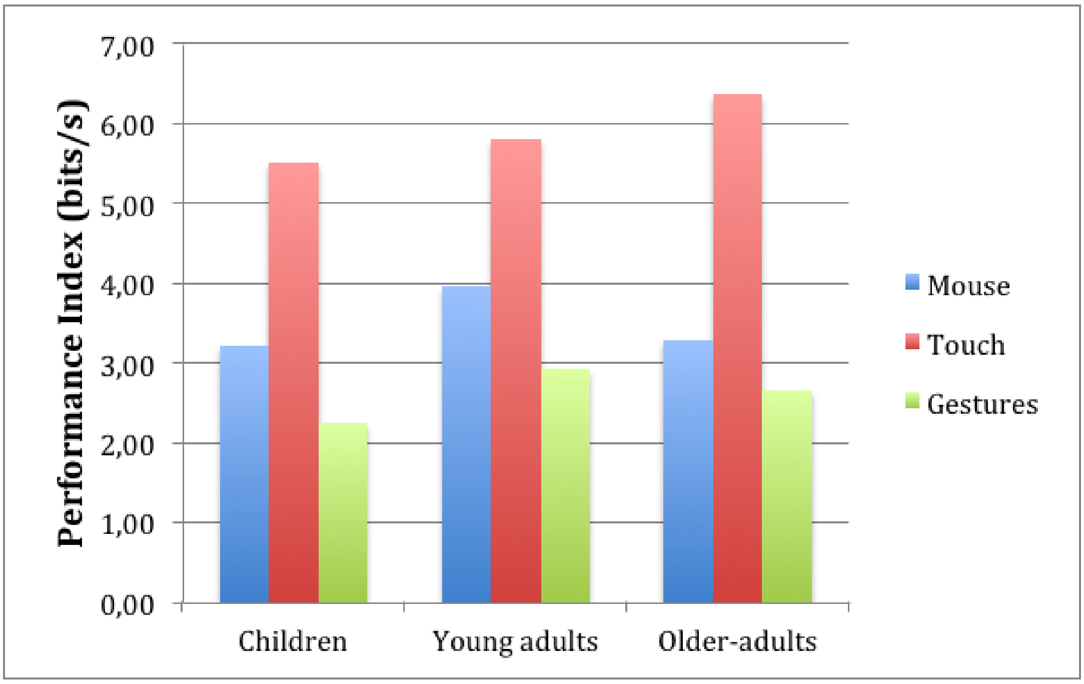


Figure 4.24 – Average IP of the three age groups for the circular task.

As in other tests, additionally to comparing IPs between each group regarding the different interfaces, we also analyzed these values within each group. When it comes to the linear task performed by children, a repeated measures ANOVA with a Greenhouse-Geisser correction determined there was a significant difference between IPs ($F(1.350, 21.599) = 6.615$, $p <.011$), but only between the mouse and gestures ($p = .032$) according to post-hoc pairwise comparisons tests using the Bonferroni correction. Comparisons between mouse and touch ($p = .068$), or touch and gestures ($p = .179$) were not statistically different.

On the other hand, the young adults' performance was not significantly different between the interfaces ($F(1.416, 21.241) = 3.174$, $p <.077$), according to the repeated measures ANOVA with a Greenhouse-Geisser correction.

As for the results of the older-adults, they are significantly distinct between each other ($F(1.904, 32.371) = 23.344$, $p <.000$). Moreover, post-hoc pairwise comparisons tests show us which comparisons are indeed different: touch against mouse ($p = .000$), and touch against gestures ($p = .000$). By contrast, the computer mouse and gestures did not prove to be statistically different ($p = 1.000$).

Regarding the circular steering motion, a repeated measures ANOVA with a Greenhouse-Geisser correction indicated that children also displayed statistical differences in all combinations ($F(1.397, 23.754) = 26.045$, $p <.000$): mouse and touch ($p = .001$), touch and gestures ($p = .000$), mouse and gestures ($p = .015$).

The young adults' performance was also significantly different ($F(1.659, 28.204) = 28.261$, $p <.000$) according to the repeated measures ANOVA with a Greenhouse-Geisser correction. This groups shows variances between touch and the mouse ($p = .000$) plus touch and gestures ($p = .000$). Mouse and gestures were also significantly different ($p = .017$).

As for the older-adults, they display significant differences in performance as well ($F(1.258, 22.651) = 28.179$, $p <.000$), namely between touch and mouse ($p = .000$), touch and gestures ($p = .000$), and mouse and gestures ($p = .017$).

Participants' attitudes and preferences:

We observed different behaviors in the various groups concerning each modality, mainly in the groups of children and older-adults. Indeed, the young adults were constant during the trials and did not present any behavior worth noting. Also, they did not provide comments during interaction that could be relevant for the investigation. However, both children and older-adults exhibited certain aspects that should be taken into account.

Regarding the children's group, linear motions did not cause any relevant attitude, other than the fact that when the participants would get comfortable with using gestures in the activity, they would perform sudden movements with the arms instead of following with cautious the path drawn from the initial point to the destination, as if they were doing an shortened swipe movement. On the other hand, considering circular motions, after a few trials some would perform a gesture with the finger as to create a mid-air circular shape, not having to move their arm, but only the finger. At a certain point during the test, one female child affirmed that in order to complete the small circles she preferred gestures and it was easier. However, with bigger circular movements most of the children ended up drawing weary paths, possibly due to not having such good control over the movement in bigger circles (Figure 4.25). This was also true for the older-adults. Moreover, once they leave the boundaries of the circle and have to return to the initial position to reinitiate the trial, it is curious that they do not go directly to the initial position, but instead circle back all the way around the opposite direction to begin the motion once again.

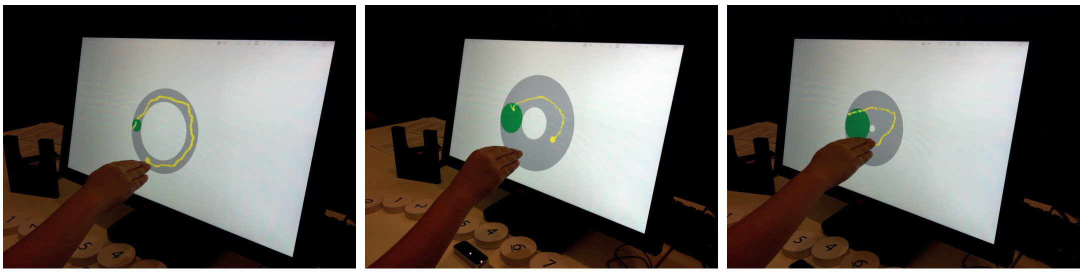


Figure 4.25 – Child performing a circular steering task with gestures.

When asked about the interface they thought they were faster with, children often answer the computer mouse. However, as we can see from the quantitative results, touch was the fastest modality.

Children and older-adults shared another behavior: they tended to make specific shapes within the boundaries of the circle when using the computer mouse, which did not happen with the other input modalities. For instance, Figure 4.26 represents two shapes that were common when completing the steering path: a rectangle or a triangle. This could indicate that when using the computer mouse, these groups tend to make straight lines as opposed to organic or curved ones. This behavior was not noticed in the group of young adults.

Figure 4.26 – Shapes drawn by some participants to complete the circular steering task.

Regarding the group of older-adults, we noticed they would have the tendency of leaning over the screen unintentionally, which at times caused an occlusion effect and thus the participant would not see the circle behind his/her hand. This could aggravate the movement times of the participants and consequently their index of performance. Also, one female participant even assumed that her eyes were getting dry over the needed concentration to complete the task.

For this steering task, children and young adults considered the mouse and touch as extremely easy to use (50% and 80%, respectively), whereas 60% of older-adults thought they were relatively easy to use. Also, all age groups considered touch and mouse as very comfortable to use. On the other hand, gestures were their least favorite as the participants even discussed this input modality was tiring and frustrating to handle.

4.3.3 Discussion

Considering the error rate, the interaction modality that registered fewer errors on straight movements was touch, and the age group to have fewer faults was the older-adults. Comparing this result with our previous studies, it may seem older-adults are more careful when interacting with the system using touch than the other groups. Moreover, gestures negatively influenced the subjects, presenting the worse error rates of all modalities in the three age groups. Regarding circular paths, it seems they require more skills to accomplish, especially considering gestures that caused the error rate to escalate in all groups. On the other hand, touch was once again the interaction mode to present fewer errors.

When observing linear steering efficiency, gestures did not present the poorest mean results in all groups, contrarily to what has happened on previous tests: children took longer to complete the task with the mouse than gestures. However, keep in mind that gestures was also the one to register more errors, which means that children may be faster but they may not be very precise with gestures. Also, young adults did not show significant differences when compared to children using gestures, whereas there were differences recorded when compared to older-adults. This analysis could present some insights on age groups' patterns during interaction: (1) older-adults complete linear steering tasks with significant difference when compared to young adults, regarding all modalities; (2) they interact in a similar manner as children when using the computer mouse and touch. Also, as in previous studies, they take longer to complete the tasks with gestures than the other groups. Therefore, it is valid to acknowledge that gestures may negatively influence interaction, as the groups presented an increase in movement times when resorting to this input modality.

Considering the circular steering task, using the mouse produced curious results: the difference in interaction was indeed significant between all the groups. This situation had happened also in the previous usability study on elemental selection tasks. This device, in spite of being the one all participants had experience with, showed again statistical differences regarding efficiency, this time on circular steering paths.

When considering mean IPs registered during the experiment for each age group, the results of the linear steering task showed that young adults can manage more information per unit time than the other age groups, since they presented the best mean values of IP using the mouse.

Also, children and older-adults only registered significant differences when using gestures, and not with the mouse or touch. That may indicate that children actually perform gestures very well, contrarily to older-adults. As for young adults and children, they did not prove significant disparities with gestures, but when compared to older-adults they did. All the more, young adults also differentiated from children and older-adults when using touch. Regarding the mouse, young adults also displayed statistical variances compared to children and older-adults. It becomes increasingly curious the fact that the computer mouse, widely used by all in daily activities and with supposedly consistent levels of experience, actually has continually demonstrated that it does not have similar performances regarding the different age groups.

Regarding the circular steering task, touch was the one to present higher performance indices from all groups. Also, using touch, older-adults did have higher performance values than young adults. This may be curious, since the steering time of young adults in this task was lower and thus they took less time to complete the task than older-adults. However, keep in mind that the performance index is not only measure by movement time, but also by number of mistakes committed and precision in the task. This may indicate that older-adults, when it comes to steering movements, perform better with touch than young adults and children. On the other hand, gestures, contrarily to the linear task, presented the worst results in all groups.

4.3.4 Summary

In this experiment, we intended to understand which input modality could positively influence the indices of performance of three specific age groups in a linear and circular steering task, considering the Steering Law. We considered this human performance evaluation method because it is well-known and validated for analyzing trajectory-based tasks, a widely used action for manipulating and navigating through content.

The modalities considered were the computer mouse, touch and gestures; and the three age groups were organized as children, young adults and older-adults.

Certainly, the experiment presented considerable differences in each group when resorting to distinctive interaction modalities to perform steering tasks. Especially considering linear manipulation paths, children showed that they could manage more information per unit time using gestures rather than other input modalities. This was, indeed, a surprise, as gestures tended to present the poorest results in previous studies. Moreover, touch proved to be the paramount interaction modality between all groups, regarding both linear and circular activities.

When it comes to the circular movement, we could observe the older-adults did present higher performance values with touch than young adults, but were slower on completing the task. This could indicate, yet again, that in spite of young adults being able to complete the tasks more quickly, they could be less accurate than the other groups, and thus influencing interaction performances. This situation is actually important to justify why we thought relevant to evaluate not only the usability of the interface, but also analyze performance indices.

4.4 Concluding remarks

The goal of this chapter was to understand if there could be differences in reaction times and performance regarding three distinct age groups. It was also particularly motivated by the absence of scientific data that systematically evaluates which interaction mode could be the most appropriate for users with age-related differences when performing the three most common elemental tasks, whether it is content selection, insertion or manipulation.

We intended to comprehend if different input modalities could enhance or, quite on the contrary, impair the user's performance and reaction times. Also, we have become to understand that, aside from detecting the most or least efficient interface, we should take notice on the user preferences, since sometimes they do not match to the interface that would be more logical to improve performance. This is an example of children that, whilst performing better with other interfaces, declared several times in the surveys that their favorite was the tangible one. Although it may indicate that they do not perform as good with tangibles, it could be relevant for children to use this input modality for interacting with the systems, as it may improve their level of satisfaction and keep their interest in the task at hand.

Considering all surveys, the most popular interface was touch for all groups and for most tasks. Indeed, selection and manipulation presented superior results when using touch. However, insertion did not and the physical keyboard still managed to obtain better results in the content insertion activity.

Curiously, considering the selection and manipulation activities, the computer mouse proved to have statistically significantly different results between the groups. This can be an indication of a limitation of this graphical user interface paradigm, since this situation occurred with groups that had a similar experience level with the device, but were on different age brackets. This could throw some light on the fact that age could influence interaction, and not only the input device.

Discrete and continuous selection tasks may also influence the users' performance. When accomplishing discrete tasks, users are less volatile to the interface used, as opposed to continuous tasks, where they seem to show more susceptibility to the input modality.

When it comes to the circular movement, especially considering linear manipulation paths, children showed that they could manage more information per unit time using gestures rather than other input modalities. This was, indeed, a surprise, as gestures tended to present the poorest results in previous studies. All the more, gestures were the modality the groups found to be more difficult and struggling in almost all activities. Additionally, we could observe the older-adults did present higher performance values with touch than young adults, but were slower on completing the task. This could indicate that in spite of young adults being able to complete the tasks more quickly, they could be less precise than the other groups, and thus influencing interaction performances. This situation is actually important to justify why we thought relevant to evaluate not only the usability of the interface, but also analyze performance indices.

When doing an overall analysis, young adults presented the best reaction times in most activities, but their performance index diminished, which could indicate there are faster, but not as precise.

Therefore, we may presume that the creation of distinctive user profiles regarding interaction with different interfaces could, indeed, be important. Thus far, touch seems better for all groups when it comes to selection, and the keyboard seems to improve the three groups' performance in insertion activities. For manipulation tasks, linear movements are better executed by gestures when it comes to children, and touch when it comes to young and older-adults. As for circular steering movements, touch could once again be the best option.

This investigation was important to understand the input modalities' individualities and if they could present age-related differences in interaction. This was a first step in grasping which could present better results and positively influence interaction and which ones actually impaired the user on elemental activities. Thanks to these results, we could begin to understand which combinations could help improve performance. This is this investigation's next step.

5 Evaluation of compound tasks

The results obtained in the previous chapter show evidence that the creation of distinctive user profiles regarding interaction with different interfaces could, indeed, be important when it comes to elemental tasks. All groups in the study (children, young adults and older-adults) showed differences in interaction regarding the use of different input modalities for selection, insertion and manipulation tasks, and also considering linear and circular steering movements.

With this concept in mind, it is now important to acknowledge if the differences still remain when it comes to compound tasks, whether using one input modality or combining them in a multimodal system. Multimodal interface design has been inspired largely by the goal of supporting more flexible, natural and transparent means of interacting with computers. These novel interfaces are said to reduce cognitive load, stimulate cognition and improve performance in a variety of tasks (Oviatt e Cohen, 2015). As such, they are better for accommodating individual differences among users and thus may provide a universal access to computation.

Therefore, in this second phase of our investigation, we intend to understand: (1) if the results gathered on the previous phase were also valid for compound tasks or, on the other hand, if a change in the type of task (being a compound transaction) affected interaction; (2) if there were any synergies when combining more than one interface to perform compound tasks, namely when using two, three or four input modalities together in the same activity.

For this purpose, we created a simple game that could gather all the features we were interested in studying. We did not opt to create everyday-based formal applications, such as the task of browsing the web, or consulting the mailbox, because we had to embrace different groups of users with different day-to-day routines concerning the computer. Indeed, it was necessary to create an application that would arouse the interest of children, instigating them to play and be captivated to be willing to keep playing throughout the several stages of the game. This would be more difficult if we considered formal applications, and could cause the children to perform worse due to the lack of interest in the tasks, rather than the interaction modality itself. Also, the groups of young and older-adults would be more accustomed to the tasks performed on a daily basis, and could find it confusing to execute them using modalities other than the ones they were used to. This could enhance error susceptibility and trigger non-intended results due to recalling the wrong input modality that they had more experience with in doing that same routine task, and thus confuse the user as to what interaction modalities to actually use.

Therefore, we resolved to create an informal and playful application (a simple game) that would still resort to the same elemental tasks of the daily routine using a computer, but with a different feel to the groups. We did not want to recreate specific daily tasks, as checking the email, or navigating through the web, because we did not want to create bias groups that would more easily relate to those activities. Instead, we created a simple game that required the same types of elemental or compound tasks, but in a more casual manner.

The three groups of users and their characteristics were previously described in Chapter 3, and were maintained throughout the whole study. The sample for this phase is also in accordance to the preceding experiments.

5.1 Performance with one mode of interaction

In order to validate previous results with regard to interaction with individual modalities, and before combining different interaction modalities and observe their synergies, we performed a specific case study with a compound activity that makes use of each modality individually in order to complete a simple game.

5.1.1 Materials and Methods

Three input modalities are used in this case study: the computer mouse, touch and gestures. We alienated tangibles, as they are not pointing devices and the game created for this purpose has a path specification activity.

Apparatus:

The experiment was conducted in a closed room with artificial light and the tests were performed in a specific setup assembled for the purpose of this research (Figure 5.1). The conditions of the experiment were identical to the ones adopted in our previous phase, regarding the final elemental tasks.

Materials:

A purpose-built game was created with support for all modes of interaction, using the same frameworks and programming language as the previous studies, and also with the same adjustments made to the system's feedback across the different input modalities. The efficiency and efficacy of the participants were recorded automatically by the game, for each step and each trial.

Regarding the user preferences, we asked the participants to complete a finishing survey with closed-end questions, qualitative Likert Scales and ranking lists, in order to understand the user's preferences considering: ease of use of the input modalities,

Figure 5.1 – Experimental setup.

ease of learning, fatigue effect, naturalism of interaction, level of user comfort/frustration, and user's degree of presence and concentration. Also, like the previous studies, during the tests we gathered qualitative observational analysis on the participants' behavior.

Experimental design:

We dealt with two independent variables in this study: the age groups; and the input modalities. In turn, our dependent variables were the traditional metrics: effectiveness, efficiency, inactivity time and user preference.

Altogether, there were 180 experimental trials. The aim of each trial was to observe the participants execute the compound task with each of the three available input modalities. Regarding the evaluation within groups using the different input modalities, we used a repeated-measures within-participants design, whereas regarding the evaluation between age groups we resorted to a repeated-measures between-participants design.

Participants:

The study included 60 volunteers who were naïve to the purpose of the experiment. We divided them into three groups, according to their age, and we thereby had 20 users for each age group participating in the tests:

- **Children:** 20 participants of 8 to 12 years old ($M = 10.5$; $SD = 1.41$).
- **Young adults:** 20 participants of 25 to 30 years old ($M = 28.5$; $SD = 2.20$).
- **Older-adults:** 20 participants of 45 to 60 years old ($M = 51.5$; $SD = 5.04$).

These individuals were selected according to the directives stipulated on Chapter 3. All participants were right handed and reported normal or corrected to normal vision.

Procedure:

After being given an overview of the steps of the game and the different obstacles to overcome, participants were instructed about the goal of game. Before the test began, all participants were allowed to practice the game, since they needed to get used to the input modalities and gain adequate practical skills for the different actions required. Participants were the ones that indicated the researcher that they were comfortable and ready to begin the tests, as the practice attempts were not accounted for nor had a limited time.

The game was organized into two phases with two different goals: (1) overcome a labyrinth and its obstacles; and (2) insert a code to unlock the prize (Figure 5.2). Here, the participants were asked to complete the first phase before continuing to the second one.

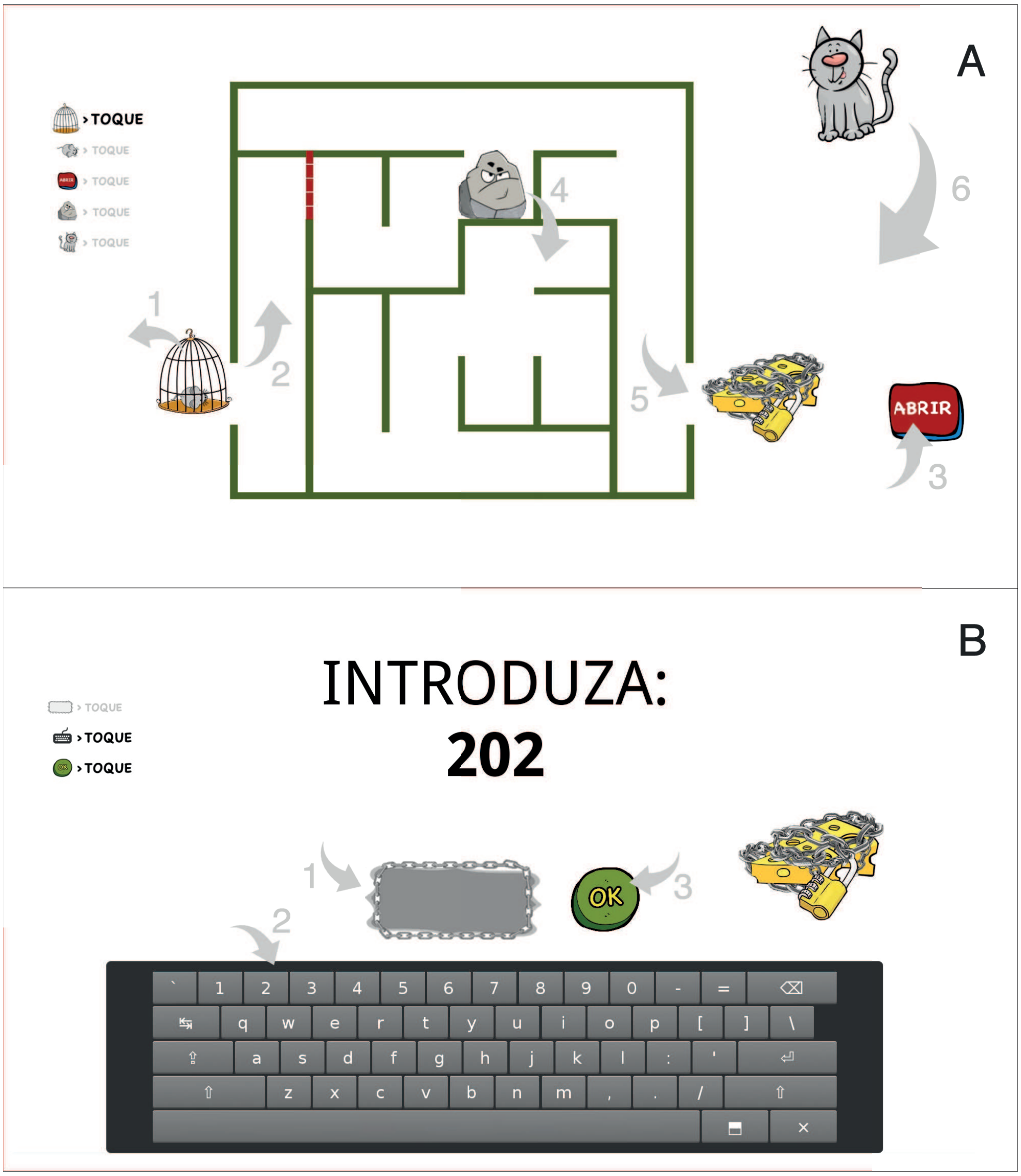


Figure 5.2 – Layout of the game: a) first phase; b) second phase.

During the game, users had to complete the first phase by performing different continuous steps: (1) free a mouse from the cage; (2) direct the mouse through the labyrinth without touching its sides; (3) open the red brick door; (4) remove the rock from the way; (5) get to the end of the labyrinth to get the cheese; (6) scare a moving cat that appears in the screen. For the second phase, users had to: (1) select the input field; (2) insert the code; (3) submit the code.

The participant had to complete the steps using only one input modality at each time. The intended modality was given at the beginning of the game and would be always visible in the top left corner of the display. As this phase of the study only encompassed one modality at a time, all the icons regarding the different obstacles would have the same caption, as is the case of the Figure 1 that maintains the touch modality throughout the game. The order of the modalities was random in each test.

The steps are as follows: to free the mouse, the user had to select the cage and move it out of the way, to wherever the user wanted. The final position of the cage was not important, as long as it was out of the way for the mouse to be able to enter the labyrinth. Then, the user had to lead the mouse through the labyrinth without letting it touch the walls. When this situation happened, an error was recorded and a position of the impact. This would be useful for the evaluation of heat maps, in order to understand where most errors occurred. When getting close to the brick wall, the participant would have to open it. However, instead of directly selecting it, as with the cage, the participant had to select a button on the bottom right corner of the display. Our intention was to save the different times both these actions took the user and compare them: selecting an object directly or manipulating it by selecting a third-party icon. Afterwards, the user would have to move a rock out of the way, using the same technique as for moving the initial cage. With the path clear, the participant would direct the mouse towards the exit of the labyrinth to get the cheese. But as soon as the mouse would get out of the maze, a cat that was not in the display originally would appear from the top right corner, moving towards the mouse in a controlled speed. The cat would take approximately 2 seconds to get near the mouse. In this instant, the participant would have to scare the cat away, or else a fault was recorded and the participant would get a visual feedback that the mouse had been eaten (the image of the mouse would disappear).

This first phase of the study, using one modality at a time, did not embrace parallel tasking, as there were no simultaneous activities. The user would terminate the previous action and only then continue with the following ones. Please remember that parallel tasking was addressed on Chapter 2 of this document.

Regarding the technique necessary to complete the steps: when using the computer mouse, the participant would click on top of the necessary obstacle and drag it to a new position on the screen to make it move. This is the case for the cage, the rock and the little mouse. As for the brick wall, all that was necessary was for the participant to click on the red button to open the wall and let the mouse pass. When it came to the cat, the participant had to click on it and drag it upwards to make him go away and disappear to where he first came. When resorting to touch, the participant had to touch the intended object and drag it directly on the screen. The same is true for scaring the cat and opening the brick wall. On the other hand, when using gestures, the participant would have to select the intended objects resorting to the same technique explored in the previous experiments (explained in Chapter 4): get the target badge on top of the object and wait for selection. However, as these new experiments require the user to drop objects as well, the participant learned that in order to release any object that he/she had picked up, it was necessary to shake their hand. The system would detect the gesture and let go of the object. When an object was being dragged by gestures, it would get semi-transparent, for the user to have a visual feedback of the action being performed (Figure 5.3). In the last step of this phase, to scare the cat away, by using gestures the user did not have to take the time to select the mouse and drag it, as it was a moving target. As such, the participants learned that they would only be required to perform a gesture for the system to recognize their intent: a swipe gesture to scare off the animal. As soon as the gesture was recognized, the cat would leave. On the contrary, if the participant did not perform the gesture, or did not perform it correctly, the cat would continue to advance towards the mouse.

We considered an error in these situations: for the first phase, the participant led the mouse to collide with the labyrinth walls; and for the second phase, a wrong number was inserted. Regarding the different obstacles in the maze, the participant could not cause an error, but instead the time taken to overcome each obstacle would escalate. To a certain degree, this method is equivalent to the speed-accuracy trade-off of Fitts' Law (for more depth on this subject, please refer to Chapter 3).

After completing the first phase of the game, the cheese would be locked, and so

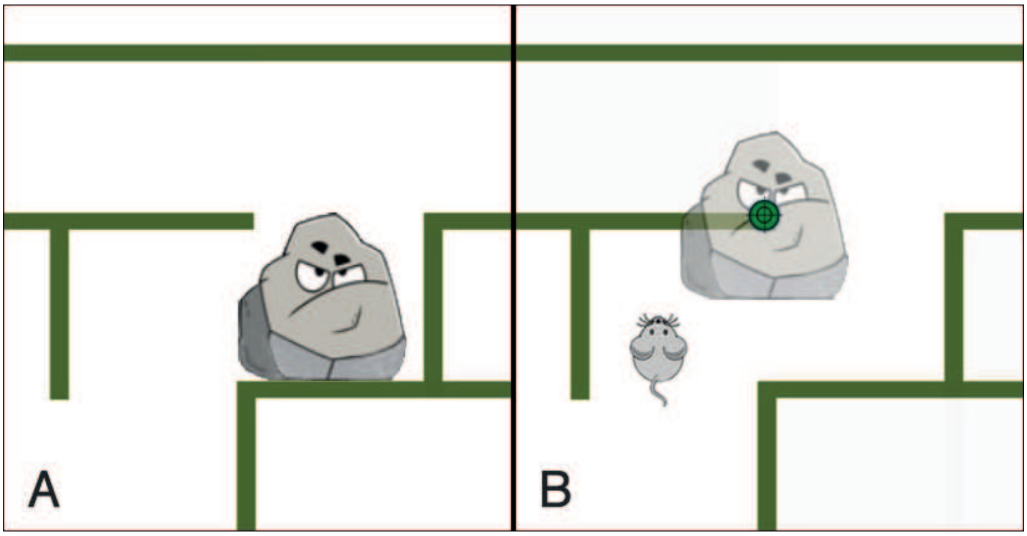


Figure 5.3 – Visual feedback of the dragging action with gestures as the input modality: a) object standing still; b) object being dragged using gestures.

the participant had to insert the correct code to unlock it and feed the starving mouse. They had to insert the provided code using one of the three input modalities, according to what was requested, as quickly as possible while avoiding any errors. First, the participant selected the input text box, then inserted the numbers one by one, and finally selected the submit button. If the code was the right one, a message of congratulations would appear on the screen (Figure 5.4) and, if the participant did not want to rest, the game would restart with a different interaction modality. If the code was incorrect, the numbers would turn red and the user would have to delete the numbers inserted and insert the right ones again, until the code was accurate. For this compound selection/insertion transaction the participant performed the same way as explained in the previous Chapter 4, regarding the elemental tasks.

Regarding the time recorded, the system would only start recording it when the user started the first action of each phase (removing the cage, and selecting the input box), and not before, and would stop recording at the end of each action of each phase (scaring the cat, and submitting the code). Thus, each phase had the times recorded individually: time taken to complete phase 1 and time taken to complete phase 2. This is due to our interest in comparing the times without the period of inactivity, i.e. when the participant was just looking at the game layout and not

Figure 5.4 – Message of successfully completed level of the game.

performing any action. Thus, we also recorded the inactivity periods, in order to compare them and understand which modality presented the highest values. We consider an inactivity period all the time that the user does not perform any action and is just looking at the layout, inert. An example of this condition could be when the user scares the cat away and the screen to insert the code appears: the user's reflects to immediately select the insertion box could be slow or the user could be reading the code before starting to insert it. Also, all the times of each object manipulation were recorded, as well as the entire time for the game.

There were no time constraints to successfully complete the tasks, since this could cause stress and influence performance. However, the participants were asked to complete the activities as fast as they could.

The game completion time was measured between the fist action performed by the user (removing the cage and thus freeing the mouse) and the successful insertion of the code, minus the inactivity period.

Once the participants had finished the experiment, they were asked to fill a questionnaire. The purpose of this questionnaire was to gather their opinion about the different interaction modalities and preferences, as already described, and to evaluate the perceived performance and efficiency of those modalities from their own perspective.

5.1.2 Results

Below are presented the results regarding the usability evaluation of the compound selection/navigation and selection/insertion tasks: effectiveness (error rate and completion success), efficiency (completion time) and satisfaction are presented and analyzed accordingly.

With this case study we intend to validate previous findings and also understand which modality can more positively influence interaction regarding compound tasks, and which one can lead to worse results when it comes to the three age groups. In this case, each experimental session lasted approximately 10 minutes.

Effectiveness:

Figure 5.5 presents the total number of errors registered during the experiment. In total, there were 1960 errors in both phases, being the majority committed by children and older-adults. The gestural-based interaction modality was the one to present most faults in all three groups. However, as we can see, the groups of older-adults suffered the most with this input modality.

Indeed, most of the errors occurred during the first phase, in the labyrinth, as the errors recorded by the wrong insertion of the code consisted of only about 3.4% of the entire value. Also worth noting is that the computer mouse also produced more errors than touch, and that this interface was the most constant between the three groups: with 9 errors from the children, 10 from the young adults, and 11 errors from the older-adults' group.

Efficiency:

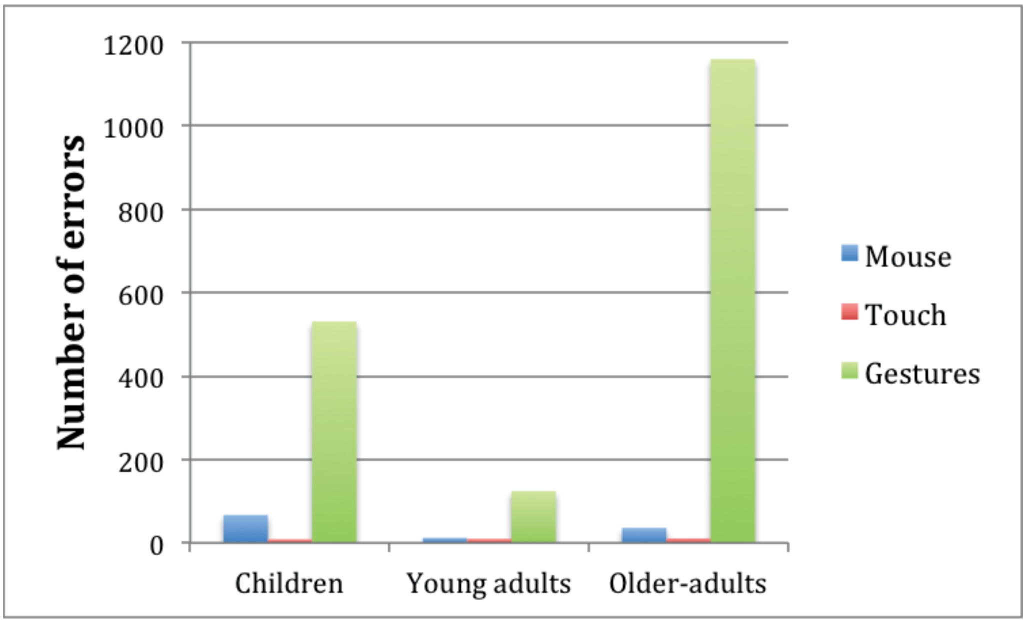


Figure 5.5 – Total number of errors each group committed per interface.

Table 5.1 shows each group's average completion time, the maximum and the minimum time, as well as the standard deviation (in seconds) for the three groups with the three input modalities.

In terms of completion times recorded, the input modality that registered the fastest mean results throughout all the three groups was touch: children (21.49 s); young adults (15.40 s); and older adults (26.47 s). On the other hand, the gestural interface registered higher mean completion times: children (95.52 s), young adults (40.24 s), and older-adults (70.53 s). The lowest time recorded was by young adults with touch (10.06 s) and the highest time was registered by a child using gestures (188.64 s). This value was 228% higher than the same maximum value registered by young adults, and 63% higher than the older-adults.

Figure 5.6 presents the game's mean completion time recorded during the experimental tests with the different interfaces by each age group. Overall, the children's group presented the highest mean times using both the mouse and the gesture-based input. In contrast, the group of young adults performed better with all interfaces and thus showed the lowest mean times. It may also be important to highlight that

Table 5.1 – Each group's mean, minimum, maximum and standard deviation times for the game regarding the three input modalities (in seconds).

Groups	Interaction mode	Mean time	Maximum time	Minimum time	Standard deviation
Children	Mouse	30.81	51.32	16.46	9.22
	Touch	21.49	32.06	13.73	4.85
	Gestures	95.52	188.64	48.88	39.34
Young adults	Mouse	19.45	26.66	13.74	3.98
	Touch	15.40	23.19	10.06	3.78
	Gestures	40.24	60.55	25.25	9.14
Older-adults	Mouse	28.36	40.60	16.97	6.19
	Touch	26.47	42.91	15.36	8.10
	Gestures	70.53	122.77	35.42	24.22

the completion time recorded by the children with the gestural interface was about twice the young adults' group.

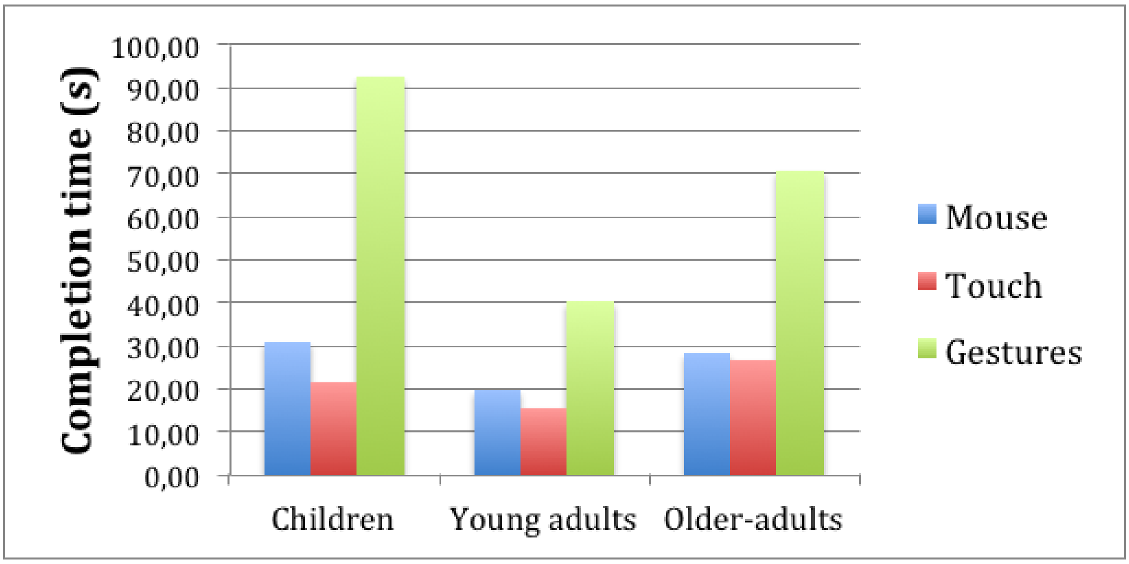


Figure 5.6 – Mean time taken to complete the game (in seconds).

We recorded the mean time of each action individually as well, using each modality. Figure 5.7 demonstrates the times recorded during the completion of the different activities individually by the three groups using each input modality.

Gestures had the highest results, except for the manipulation task that was required

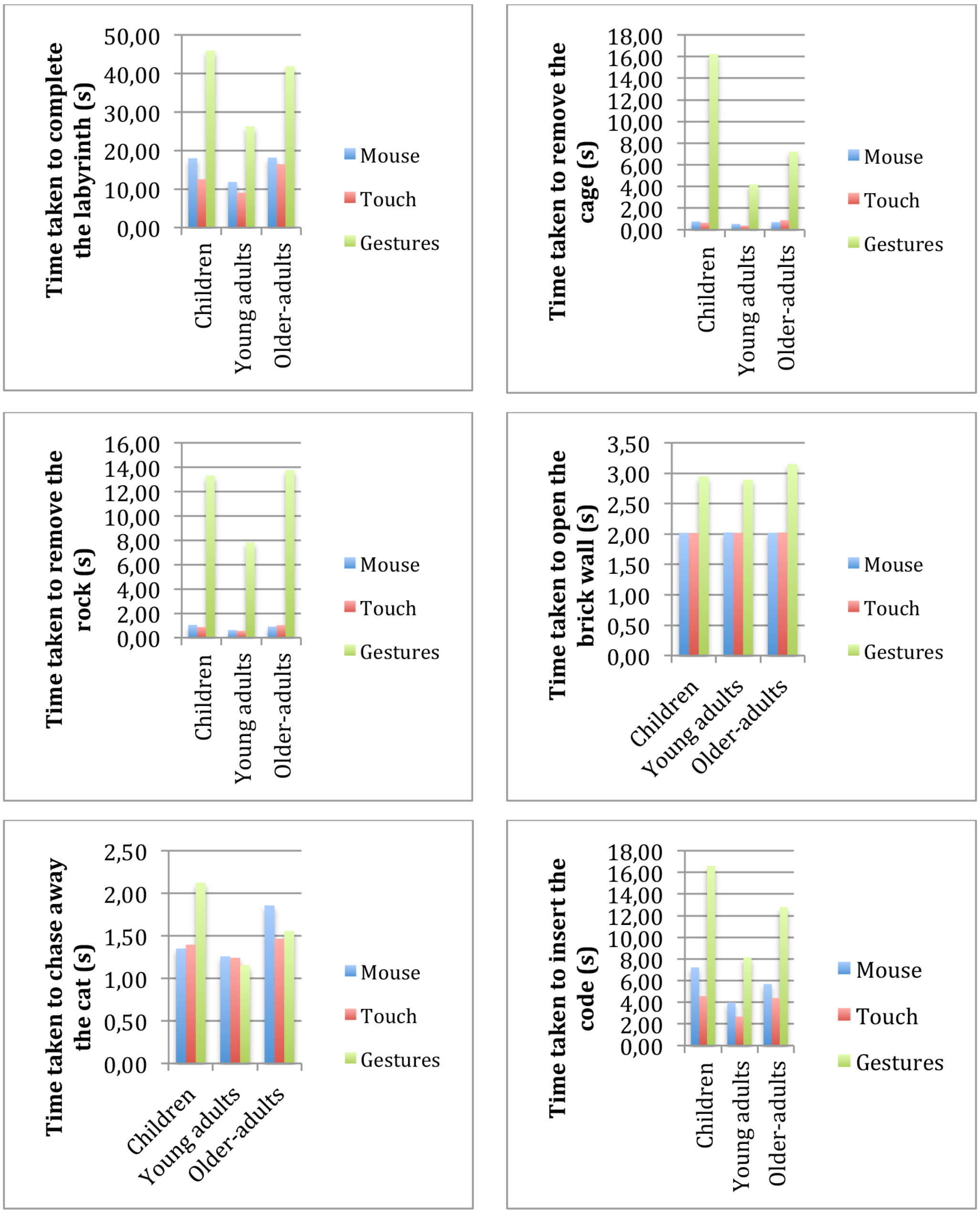


Figure 5.7 – Time taken by the groups to complete the several actions in the game using the three modalities (in seconds).

by chasing the cat away. Here, the results were distinct from the other tasks (selection and insertion). The remaining values appear to follow the same conclusions as previous studies.

Also, we measured inactivity periods as well (Figure 5.8). As we can see, the young adults were the ones to record fewer inert times using all of the modalities, and these seem to share similar values of inactivity. In the children's group, gestural interaction did indeed record the higher inactivity times, and the computer mouse did not create so many inert times as the others. This is also valid for the group of older-adults that presented very similar inertia results for both touch and gestures.

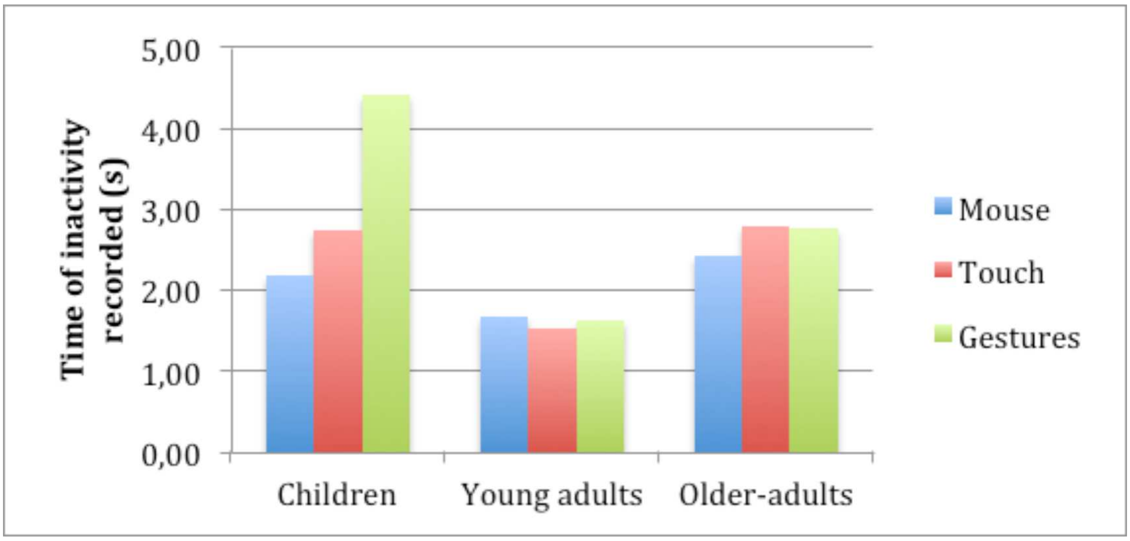


Figure 5.8 – Time of inactivity recorded during the entire game by each modality (in seconds).

Regarding the mean completion time taken by the groups to complete the game, we analyzed the data and their statistical validity. We assessed the normality of data with the Shapiro-Wilk Test in order to understand if the data were normally distributed and could thus be considered for statistical analysis. We elected this test over the Kolmogorov-Smirnov due to the size of our sample. Except for the gesture-based interface in the group of older-adults ($p = .002$) and touch in the group of children ($p = .020$), the data were normally distributed. Analyzing the significance of the results by a one-way ANOVA, we observed a statistical difference at the $p < .05$ level between all the groups and each modality: mouse ($F(2, 53) = 13.993$, $p = .000$); touch ($F(2, 54) = 16.026$, $p = .000$); gestures ($F(2, 55) = 16.176$, $p = .000$).

A Tukey post-hoc multiple comparison test revealed that the only input modality that did not present a significant difference between the groups of children and older-adults was the mouse (p = .809). On the contrary, every other combination presented significant differences in terms of mean completion time by the three groups. As such, using the mouse proved to be statistically different between children and young adults (p = .000), and young adults versus older-adults (p = .000). As for touch, young and older-adults were significantly different when interacting with this modality (p = .000), as were the children compared to young adults (p = .008). However, in spite of being statistically unlike, the differences in interaction with touch in the groups of children and older-adults were not as significant as the others (p = .036). The same situation happened when using gestures: the group of children versus older-adults were statistically different (p = .045), but not at the same level as the other groups: children and young adults (p = .000), or young adults versus older-adults (p = .005). Thus, children and older-adults do not present such a disparate interaction performance when compared to other combinations regarding compound tasks.

Additionally to comparing the performance between each group regarding the different interfaces, we also analyzed the performance within each group resorting to a within-participant repeated-measures design.

Regarding the group of children, a repeated measures ANOVA with a Greenhouse-Geisser correction determined there was a significant difference between the main time taken to complete the game in each interface ($F(1.095,\ 18.615) = 44.659$, p = .000). The input modality with the fastest results was touch (21.49 s), followed by the computer mouse (30.81 s), and finally gestures (92.52 s). Post-hoc pairwise comparisons tests using the Bonferroni correction showed that the mean times were statistically significantly different. Using touch was indeed significantly faster than using the computer mouse (p = .007) or gestures (p = .000), as was using the mouse compared to gestures (p = .000).

As the Mauchly's sphericity was not assumed in the young adults' group, like in the children's group, we applied the Greenhouse-Geisser adjustment to the repeated measures ANOVA. This group also presented statistically significantly different completion times ($F(1.228,20.879) = 130.570$, $p = .000$). Post hoc tests using the Bonferroni correction revealed that the mean times returned by the young adults using gestures were significantly higher compared to using touch ($p = .000$) or the mouse ($p = .000$) to complete the game, considering that this group recorded a high mean completion time of 40.24 seconds using gestures. The same goes for using the computer mouse versus touch ($p = .002$), being the touch modality the fastest (15.40 s), followed by the computer mouse (19.45 s).

Older-adults' efficiency also achieved a statistical significant difference between the mean times in each interface ($F(1.103,19.850) = 57.128$, $p = .000$). However, although the analysis returns a statistical difference regarding the use of gestures compared to touch ($p = .000$) and the mouse ($p = .000$), using the computer mouse and touch to execute the game did not lead to significant differences ($p = .844$). In this situation, the touch modality also scored the lowest mean time (26,47 s), closely followed by the computer mouse (28.36 s), and finally gestures (70.53 s).

Participants' attitudes and preferences:

During the tests, all of the participants expressed with vigor their discontent towards using gestures to complete the labyrinth. Indeed, in previous experiments participants had already commented their dislike towards this input modality. Now, during a compound task, their opinions were even stronger.

Some participants even discussed their unwillingness to complete the task when they were asked to go through the labyrinth using gestures. This was a fact for all age groups. Also, the same attitude was expressed in the screen to insert the code, where participants were asked to select the proper digit in the soft-keyboard.

We also noticed that participants would mistake the gestures they had to perform during the game. Indeed, they would shake their hand for more than the intended action: whereas to let go of an object or select the objects. Instead, the procedure was to stand still while selecting and object, and shake the hand to drop it. This proved to be very confusing for the participants, mostly children. In turn, the other input modalities did not cause any uncommon attitudes in the participants and the tests continued without special behaviors.

5.1.3 Discussion

For a compound task involving three transactions (selection, manipulation and insertion), the gestural interface appears to show worse results in terms of completion times recorded. On the contrary, touch seems to be the most coherent modality among the groups and returned the lowest values.

Indeed, the compound tasks followed a pattern coincidental with the respective elementary activities approached in the previous Chapter: (1) the gestural-based interface was not as efficient as the rest considering the game as a whole; and (2) touch was the modality that achieved the best results among the three user groups.

We recorded the times of each action individually as well, using each modality. This can be important to establish if the results obtained are caused by the composition of different actions in one compound task, and if the results are still valid and concurrent with the previous experimental phase of this study: elemental tasks.

It is important to highlight that gestures had the highest results, except for the manipulation task that was required by chasing the cat away. Here, the results were distinct from the other tasks (selection and insertion). Notice that this result was discussed in the previous Chapter, where we came to realize that, for manipulation tasks, especially for linear ones, gestures was a modality to consider in terms of speed. That being said, we presume that this situation continues valid when the actions are combined in a compound task, once the gesture-based interface showed better results than the mouse in both the young and older-adults' groups, and even proved to be better than touch for young adults. Curiously, previous tests showed that children took less time to complete the task than older-adults in linear steering tasks. However, in this situation they took longer than the other group. This could indicate that when it comes to compound tasks, gestural inputs may degrade interaction, as opposed to elementary transactions, for the group of children.

All the same, we find it interesting to also note that for removing the obstacles out of the way (remove the cage and the rock; open the door), we had two techniques: (a) select object and drag; (b) select button; respectively. As we can see in the charts (Figure 5.7), both removing the cage and the rock shared the same technique. However, the gestural interface showed different results: the time taken to remove the cage was far less in the groups of young and older-adults than the time taken to remove the rock. One might think that this could be due to the order of the actions: removing the cage was the first action to be performed, while removing the rock was a task executed right after another one. This could indicate that gestural interaction during compound tasks may suffer if the action is not required at the beginning of the transaction.

Regarding the periods of inactivity, the young adults were the ones to record fewer inert times using all of the modalities, and these seem to share similar values of inactivity. This situation could also imply that young adults have the fastest reflexes during interaction, or at least they do not seem to take as long as the other groups to act upon the task.

Regarding the mean completion time taken by the groups to complete the game, we observed a statistical difference in the values. Indeed, we have reached the conclusion that considering compound tasks that are performed unimodally, the usage of different interfaces may influence the users' performance. This also proved to be true for elemental tasks, as described in the previous Chapter.

5.1.4 Summary

This experiment was intended to validate previous findings on which modality would be better for which interaction task. Considering a compound task that aggregated several elemental activities, we noticed that the completion times of each elemental activity regarding the different modes of interaction are more exacerbated, that is, we did not present such statistically significantly different results in the previous experiments as we noted in this purpose-built game.

All participants presented better efficiency values with touch. On the other hand, all participants thought the gestures were the worst interaction modality for precision activities, but not when used to scare the cat, as it was a linear steering task and thus it was considered practical to use. Also, more errors were detected when using gestures regarding all age groups, mainly during the steering circular movements required by the labyrinth. However, we can see that unlike previous tests that showed children took less time to complete the task in linear steering tasks using gestures, this compound task did not follow the case: children took longer than the other groups to execute the action with gestures. This could indicate that when it comes to compound tasks, gestural inputs may degrade interaction, as opposed to elementary transactions, for the group of children.

Indeed, we may come to realize that: (1) compound tasks may augment the discrepancies of each modality considering the different age groups; (2) the type of interface that is used can influence efficiency in compound tasks, even aggravating the efficiency of each group when compared amongst each other.

5.2 Performance with several modes of interaction

Results presented in the previous section show that different input modalities can influence performance in compound tasks. However, this analysis was only directed at performing the purpose-built game unimodally. Indeed, it is increasingly important to understand the synergies that can occur if one same compound task is performed multimodally.

Hence, in this section we intend to combine the different interaction modalities addressed throughout this systematic study and observe their synergies. Therefore, four input modalities are used in this case study: the computer mouse, touch, tangibles and gestures.

5.2.1 Materials and Methods

We conducted a study using different combinations/sequences of modalities, and requested participants to execute the same game created in the previous section, in order to be able to compare results and understand if the users' performance improved. The experiment was divided into three levels:

- **Level 1:** Combining two modes of interaction;
- **Level 2:** Combining three modes of interaction;
- **Level 3:** Combining four modes of interaction.

Each level presented a combination of input modalities that were randomly grouped and the participant had to execute the complete set of combinations provided in order to advance to the next level. However, the levels were not organized randomly, that is, they would respect the order of level 1 first, then level 2 and finally level 3. We did not intend for the users to get overwhelmed by all the possible combinations and preferred to keep increasing the number of combinations gradually. Also, another aspect that could be important to analyze is if the learning process influences the results.

As explained in Chapter 2 regarding interaction tasks, it was important for us to understand the performance of the participants by executing the game with parallel tasking (where two activities are performed simultaneously) and non-parallel tasking (where there are no simultaneous tasks). Thus, the participants performed both types of tasks. In the case of being a parallel compound task, the only change the level would suffer would be to integrate one other obstacle in the path of the labyrinth: the rock (as described in the previous experiment). This rock could be removed in parallel with the dragging of the mouse, if the participant saw fit. This parallel activity was also significant to recognize if the participants would embrace this possibility and which input modalities would be more useful to carry out this action.

Apparatus:

The experiment was conducted in a closed room with artificial light and the tests were performed in a specific setup assembled for the purpose of this research (Figure 5.9). The conditions of the experiment were identical to the ones adopted in our previous phases.

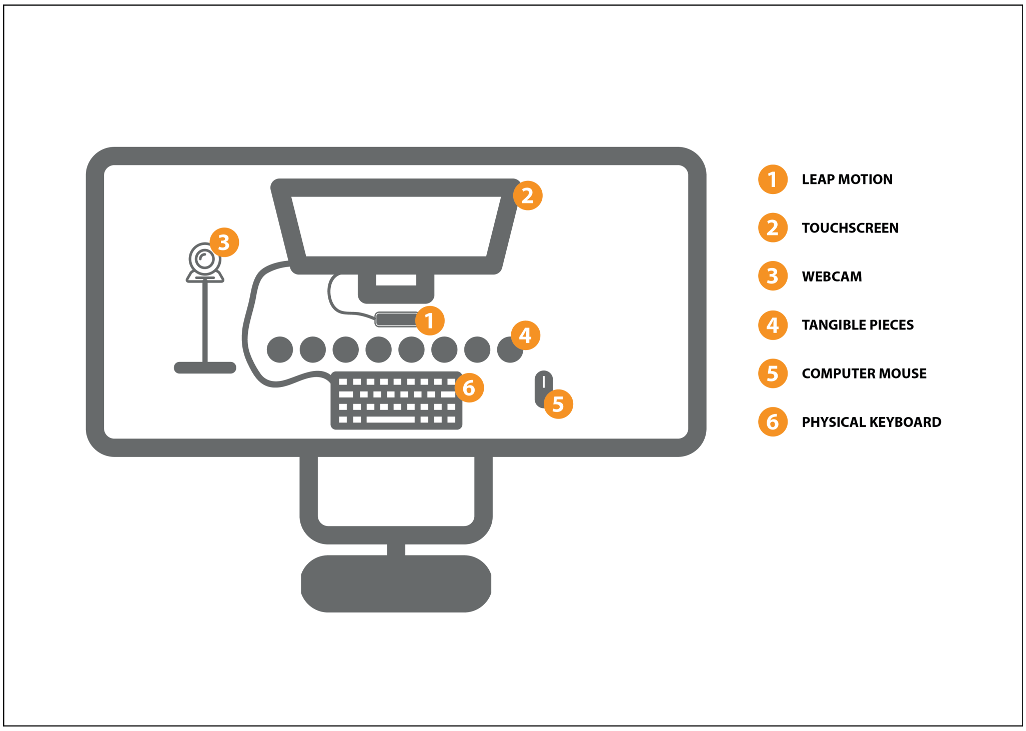


Figure 5.9 – Experimental setup.

Materials:

The purpose-built game created, with support for all modes of interaction, was used in this phase as well. The efficiency and efficacy of the participants were recorded automatically by the game, for each step and each trial. We also saved a heat map for the errors that were registered throughout the first phase of the game, as will be described later on. This map is a graphical representation of the faults detected by each modality.

Also, in order to analyze the synergies that came from combining the different input modalities we resorted to two different metrics: relative modality efficiency and multimodal synergy. The first metric (when compared with modality usage) can identify an underutilization of modalities due to poor interface design or information asymmetries. On the other hand, multimodal synergy measures the added value from efficiently combining multiple input modalities, and can be used as a single measure of the quality of modality fusion in a multimodal system. As already further described on Chapter 3, relative modality efficiency calculates the amount of information being communicated in unit time for each modality, i.e., the information bandwidth, and should correlate with relative modality usage unless there is information asymmetry between the user and the system. Indeed, relative modality usage is defined as the percent of time spent using a certain modality over the total interaction time, and is thus important to define the ratio of modality efficiency. This value is essential to calculate the multimodal interfaces with the "sum" of its unimodal parts and measure how "synergetic" the interface design is.

Regarding the user preferences, we asked the participants to complete a finishing survey with closed-end questions, qualitative Likert Scales and ranking lists, in order to understand the user's preferences considering: ease of use of the input modalities, ease of learning, fatigue effect, naturalism of interaction, level of user comfort/frustration, and user's degree of presence and concentration. Also, like the previous studies, during the tests we gathered qualitative observational analysis on the participants' attitudes.

Experimental design:

We dealt with two independent variables in this study: the age groups; and the combinations of the various input modalities. In turn, our dependent variables were: effectiveness, efficiency, inactivity time, multimodal synergy and user preference.

Regarding the evaluation within groups using the different input modalities, we used a repeated-measures within-participants design, whereas regarding the evaluation between age groups we resorted to a repeated-measures between-participants design.

We have approached how we handled outliers in Chapter 3 of this document.

Each experimental session lasted approximately 50 minutes.

Participants:

The study included 60 volunteers who were naïve to the purpose of the experiment. It is important to highlight that these were not the same subjects that participated in the previous case study. We divided them into three groups, according to their age, and we thereby had 20 users for each age group participating in the tests:

- **Children:** 20 participants of 8 to 12 years old ($M = 9.3$; $SD = 1.28$).
- **Young adults:** 20 participants of 20 to 27 years old ($M = 22.8$; $SD = 3.01$).
- **Older-adults:** 20 participants of 45 to 57 years old ($M = 50.13$; $SD = 4.67$).

These individuals were selected according to the directives stipulated on Chapter 3. Of all of the participants, three were left-handed and the rest was right handed. All the participants reported normal or corrected to normal vision.

Procedure:

Level 1:

The procedure was maintained during the different levels of the game. Also, all participants were allowed to practice the game. In this first level, the labyrinth was simpler than in the first case study. The only actions to perform were: (1) get the mouse through the labyrinth; (2) scare the cat away; (3) insert the code. Depending on the parallelism of the tasks, there could be a rock in the middle of the labyrinth (as in the previous study) or no obstacles between the mouse and the cheese. Also, the participants had to begin with a specific input mode and change to another one at some point, as instructed. This indication was at all times present in the upper left corner of the display, as in the previous study.

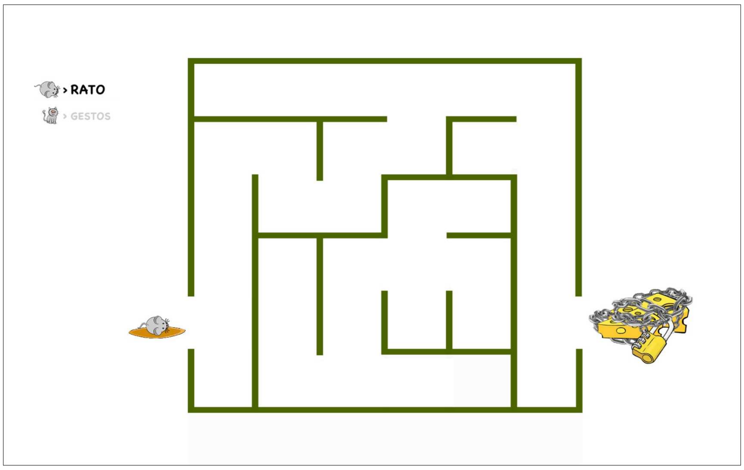


Figure 5.10 – Layout of the first level (task without parallelism) and indication of two modalities.

When it came to the parallel tasking, participants would encounter the rock in the middle of the labyrinth and had to remove it. As such, as we can see in the Figure 5.10, the maze would be performed with the first modality of the bimodal interface (in this case using the computer mouse), and chasing the cat away with the second modality (gestures). Whenever a parallel action was required, the user would intercalate with the one already in use. In this case, if the crossing the labyrinth was performed using the computer mouse, the rock would have to be removed using gestures.

The completion time of each task was measured between the first action performed by the user (as soon as entering the labyrinth) and the successful insertion of the code, minus the inactivity period.

Level 2:

The procedure was maintained during the different levels of the game. Also, all participants were allowed to practice the game. The second level of the game had one more obstacle than in the first one: the mouse was trapped inside a cage that the participant had to release. Depending on the parallelism of the tasks, there could be a rock in the middle of the labyrinth (as in the previous level) or no obstacles between the mouse and the cheese. The actions to perform were: (1) release the mouse from the cage; (2) get the mouse through the labyrinth; (3) scare the cat away; (4) insert the code. Also, the participants had to begin with a specific input mode and change to another at some point, as instructed. This indication was at all times present in the upper left corner of the display (Figure 5.11). When it came to the parallel tasking, participants would encounter the rock in the middle of the labyrinth and had to remove it.

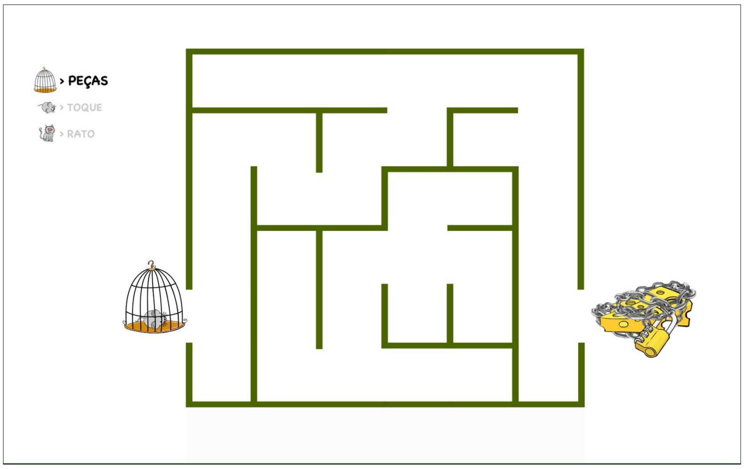


Figure 5.11 – Layout of the second level (task without parallelism) and indication of three modalities.

Level 3:

The procedure was maintained during the different levels of the game. Also, all participants were allowed to practice the game. The third level of the game had one more obstacle than in the second one: there was a brick wall inside the labyrinth that the participant had to open. Depending on the parallelism of the tasks, there could be a rock in the middle of the labyrinth (as in the previous level) or no obstacles between the mouse and the cheese. The actions to perform were: (1) release the mouse from the cage; (2) get the mouse through the labyrinth; (3) scare the cat away; (4) insert the code. Also, the participants had to begin with a specific input mode and change to another at some point, as instructed. This indication was at all times present in the upper left corner of the display (Figure 5.12). When it came to the parallel tasking, participants would encounter the rock in the middle of the labyrinth and had to remove it.

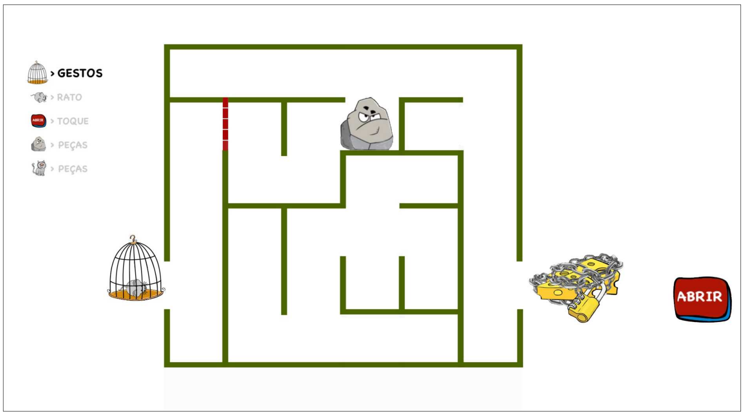


Figure 5.12 – Layout of the third and final level and indication of four modalities.

5.2.2 Synergies between two modes of interaction

For this case study, participants were asked to complete the activities resorting to two different interaction modalities required. The combinations are as follows:

- Computer mouse and gestures;
- Touch and gestures;
- Gestures and tangibles;
- Touch and computer mouse;
- Mouse and tangibles;
- Touch and tangibles.

5.2.2.1 Results

This section presents the results and discussion of the experimental tests for the two compound bimodal tasks. Thus, we examine effectiveness (error rate and completion success), efficiency (completion time) and user preference, regarding the usability of the interface, but also the inactivity periods and multimodal synergy of the interface. Moreover, such as the previous experiments, we review the participants' attitudes during the tests and their satisfaction towards the different interfaces.

Effectiveness:

Please remember that a fault was recorded under the same circumstances as the previous unimodal study. Thus, it posed as an error whenever the participant led the mouse to collide with the labyrinth walls, or a wrong number of the code was inserted. Considering the obstacle in the maze (the rock), during the parallel tasking, any issue removing it would cause the time to escalate, augmenting the modality usage of the active mode of interaction and thus reducing the modality efficiency of that specific combination.

- ***Task without parallelism:***

Figure 5.13 presents the total number of errors registered during the experiment. In total, there were 782 errors, being the majority committed by children (348 errors). As we can see, the combination of gestures and tangibles presented the most errors in all three groups, accounting for about 50% of all the errors committed.

The second combination to present more errors is the computer mouse and gestures. Considering the other bimodal interface that uses gestures to scare the cat away (touch and gestures), it seems that touch copes better with gestures than using the computer mouse.

The combination that showed fewer errors was touch and the computer mouse, with children making 13 mistakes, young adults making 4, and older-adults making 9 mistakes during the non-parallel task.

Aside from collecting the number of errors of the groups in the labyrinth, we also

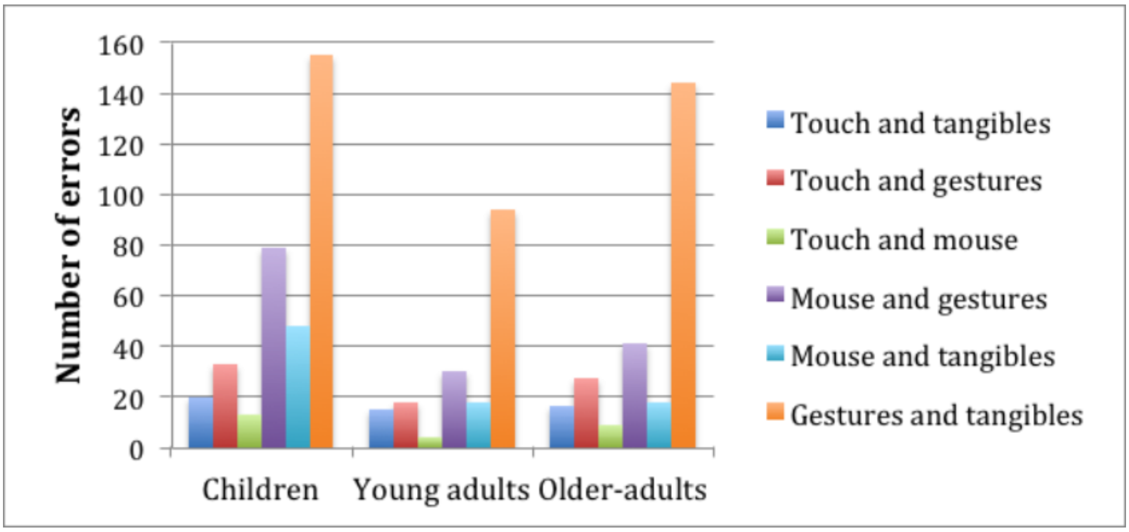


Figure 5.13 – Total number of errors presented during the experiment (task without parallelism).

recorded their positions, in order to understand where exactly in the path the participants had most difficulty. Figure 5.14 shows the position of the errors committed by the three age groups in the labyrinth using the various input modalities. The dots in the images relate to the place where a collision against the labyrinth walls occurred. As we can see, most errors were performed by gestures, and children collided against certain locations in the wall that the other participants did not.

- ***Task with parallelism:***

Regarding the task with parallelism, there were a total of 781 errors (Figure 5.15), being the majority also committed by children (433 errors).

Moreover, the following situations are similar to the task that does not require parallelism: (a) the combination of gestures and tangibles presented the most errors in all three groups, accounting for about 57% of all the errors committed; (b) the combination of mouse and gestures was the second worst in all groups; and (c) the combination of touch and mouse was the one to present fewer errors. In this case, children made 21 mistakes, the young adults 4, and older-adults made 6 mistakes.

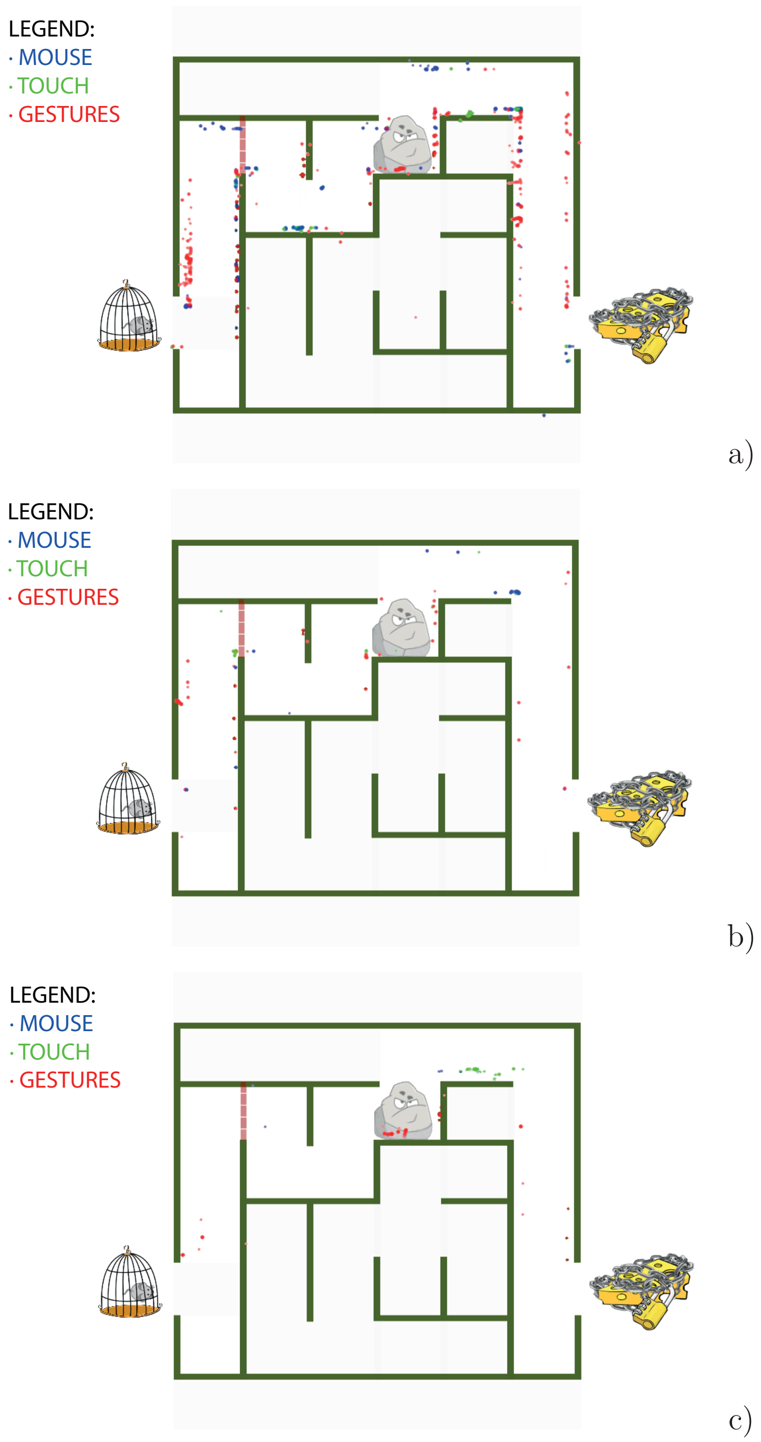


Figure 5.14 – Position of the errors committed in the labyrinth using the various combinations (in a non-parallel task): a) Children; b) Young adults; c) Older-adults.

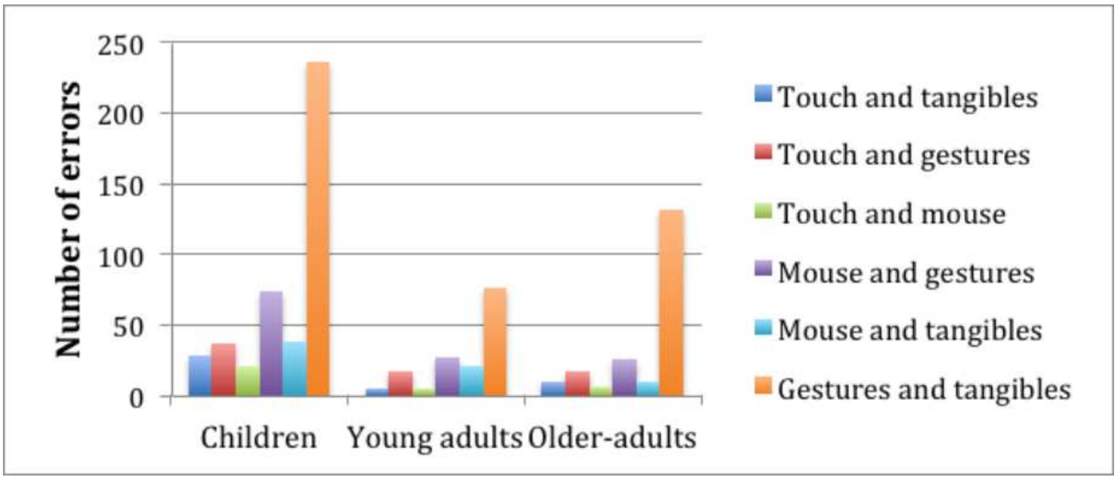


Figure 5.15 – Total number of errors presented during the experiment (task with parallelism).

Also, we recorded the position of the errors committed by the groups in the labyrinth (Figure 5.16). The same pattern as in the non-parallel activity occurs.

Efficiency:

We analyzed the completion time of the tasks with each combination within each group of participants and also between them.

- ***Task without parallelism:***

Figure 5.17 presents the mean completion time of each age group registered during the experiment when using the different bimodal interfaces. The combination that registered the fastest results among the three age groups was touch and mouse: children (18.69 s); young adults (14.06 s); older-adults (23.58 s). On the other hand, the combination with worse results was gestures and tangibles in all groups: children (49.09 s); young adults (33.82 s); older-adults (46.67 s). As we can observe, aside from these interfaces being the ones to present the highest and lowest number of errors, they are also the slowest and fastest ones, respectively, to complete the tasks without parallelism.

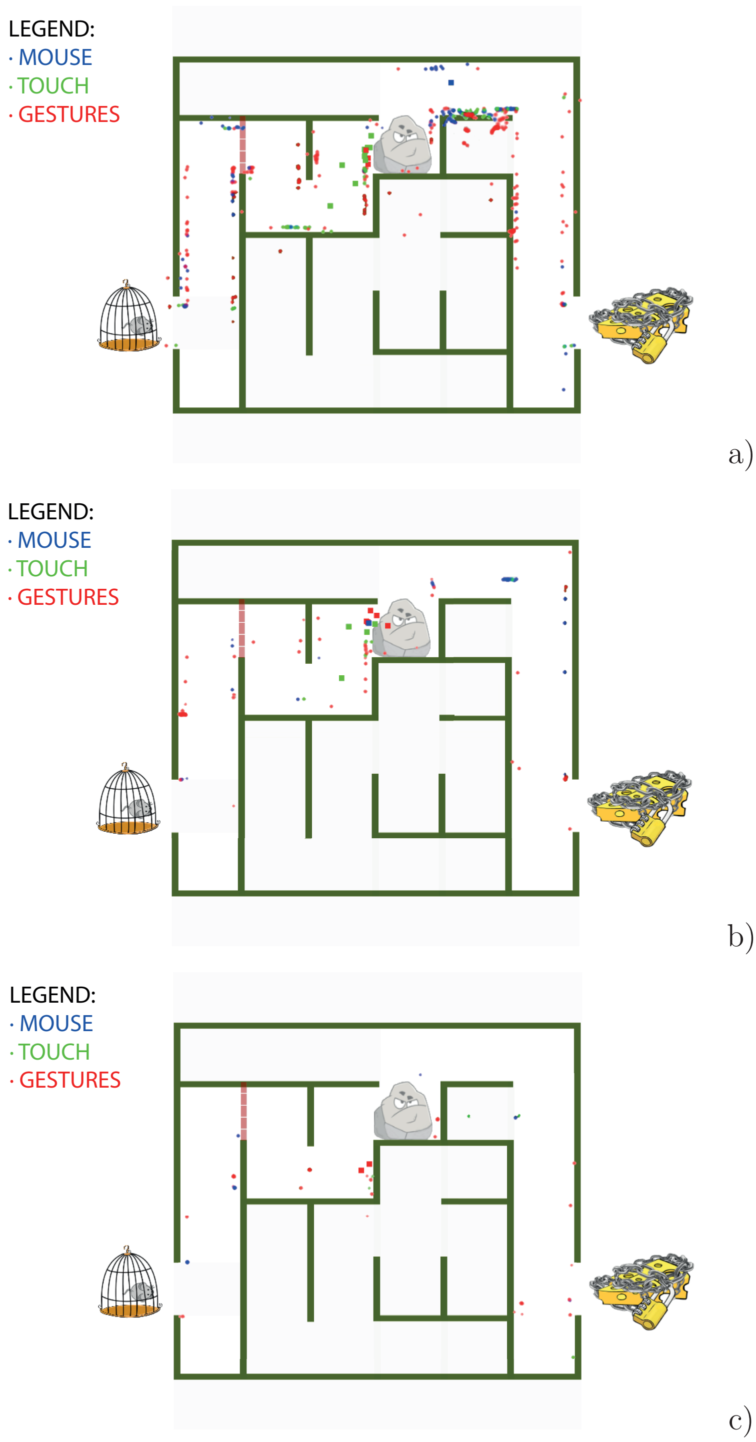


Figure 5.16 – Position of the errors committed in the labyrinth using the various combinations (in a parallel task): a) Children; b) Young adults; c) Older-adults.

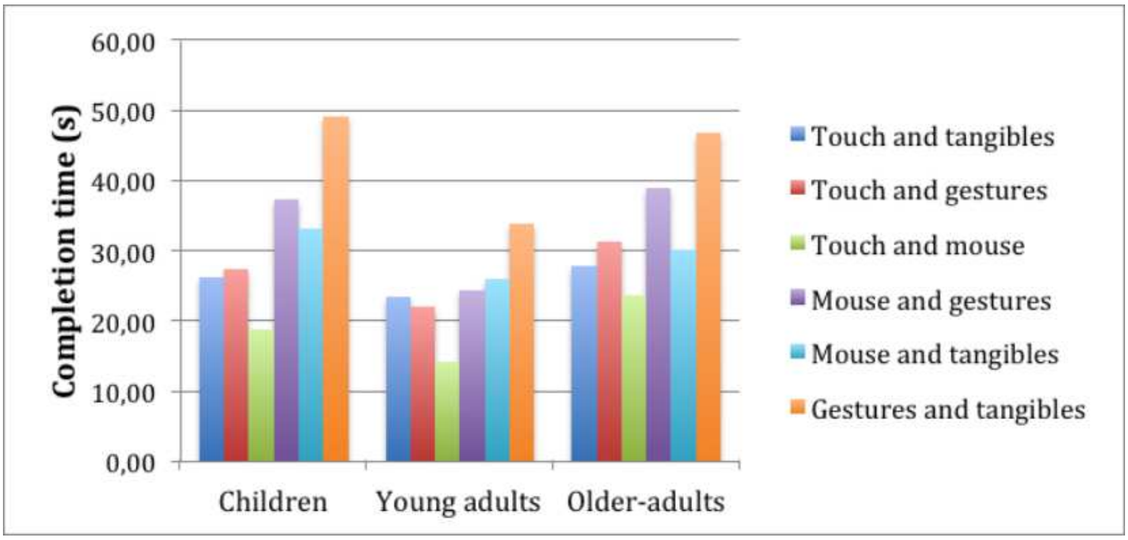


Figure 5.17 – Mean time taken to complete the task without parallelism (in seconds).

A Shapiro-Wilk test showed us most of the data was normally distributed, except for the combinations of touch and tangibles/touch and gestures/mouse and tangibles in the group of young adults (p = .000). Thus, we proceeded with a one-way ANOVA analysis as it requires approximately normal data and is quite "robust" to violations of normality, meaning that assumption can be a slightly violated and still provide valid results (Schmider et al. 2010). There was a significant difference in completion times between all groups when using the combinations: touch and gestures (F(2,57) = 4.339, p = .018); touch and mouse (F(2,56) = 6.998, p = .002); and mouse and gestures (F(2,57) = 9.885, p = .000). On the contrary, regarding the other combinations, there was no statistical difference: touch and tangibles (F(2,55) = .216, p = .806); mouse and tangibles (F(2,57) = 2.675, p = .078); and gestures and tangibles (F(2,57) = 2.921, p = .062).

Regarding the combination touch and tangibles, no significant differences were detected among the groups: children and young adults (p = .790); children and older-adults (p = .931); young and older-adults (p = .958). As for the touch and gestures bimodal interface, the results were only significantly different between the group of young versus older-adults (p = .015), whereas the other groups did not show significance: children and young adults were just as fast (p = .129), and also children and older-adults (p = .634).

Considering the combination that obtained the best results in completing the task (touch and mouse), the analysis showed us that young adults were significantly faster than the other groups: versus children (p = .038) and also versus older-adults (p = .002). However, as in the previous combination, children were quite similar to older-adults (p = .484).

The combination of the computer mouse and gestures proved to cause similar efficiency results in the groups of children versus older-adults (p = 1). However, when comparing the other two groups, the difference is significant (p = .001), which means that young adults are indeed faster than the other two groups using this combination. On the other hand, young adults were quicker than the other groups using the combination of the computer mouse and tangibles, but this difference was not significant: children versus young adults (p = .069); children versus older-adults (p = .745); young versus older-adults (p = .283). Finally, regarding the gestures and tangibles combination, this was the worst one for all groups, not existing any significant differences between them: children versus young adults (p = .063); children versus older-adults (p = .857); young versus older-adults (p = .186).

Additionally to comparing efficiency between each group regarding the different interfaces, we also analyzed efficiency within each group resorting to a within-participant repeated-measures design. Regarding the group of the children, a repeated measures ANOVA with a Greenhouse-Geisser correction determined there was a significant difference between the bimodal interfaces ($F(1.902, 36.130)$ = 15.875, p <.000). For this group, touch and mouse was the fastest combination (18.69 s). Post-hoc pairwise comparisons tests using the Bonferroni correction showed that the mean completion times using this bimodal interface were indeed statistically significantly lower when comparing to the other interfaces (p = .000), except for the touch and tangibles where, in spite of still being significant, the difference was not as high (p = .041). Regarding the worst mean completion time (49.09 s), recorded by gestures and tangibles, this value was also significantly higher than the other interfaces, except for the bimodal interfaces mouse and gestures (p = 1.000), and mouse and tangibles (p = .117).

A repeated measures ANOVA with a Greenhouse-Geisser correction also determined that the group of young adults presented statistically different results regarding the interfaces ($F(2.245,\ 42.655) = 10.976$, $p < .000$). Post-hoc pairwise comparisons tests using the Bonferroni correction confirmed that the mean times using touch and mouse were significantly lower than using the other interfaces ($p < .05$). When using the combination of gestures and tangibles, the analysis reported a lack of significant differences in efficiency between this interface and two others: touch and tangibles ($p = .279$); and mouse and tangibles ($p = .332$). Every other combination was significantly different ($p < .05$).

Just the same as the other groups, older-adults seem to present a statistically significant difference between the mean times using the interfaces ($F(3.188,\ 54.203) = 21.742$, $p < .000$). Also, similar to the group of children, the fastest interface (touch and mouse) was significantly different when compared with the others ($p < .01$), except for the combination of touch and tangibles ($p = .217$). On the other hand, considering the worst values, gestures and tangibles versus mouse and gestures did not present significant differences of time ($p = 1.000$), whereas the other combinations did ($p < .007$).

Also, we measured inactivity periods as well (Figure 5.20). All the groups appear to be very constant in terms of the periods of inertia. However, we can see that the combination mouse and tangibles, and also touch and tangibles, caused higher values on young adults. Curiously, in this group, gestures and tangibles (the combination that presented the worst results in terms of mean time) did not record such high values of inactivity. Also, the combination to have the lowest period of inactivity was the same that provided the best mean time results for this group: touch and mouse. Regarding the group of older-adults, the combination that presented longer intervals of inactivity was a different one than the previous group: it was mouse and gestures. Note that this was the second worst interface regarding efficiency for this group.

- ***Task with parallelism:***

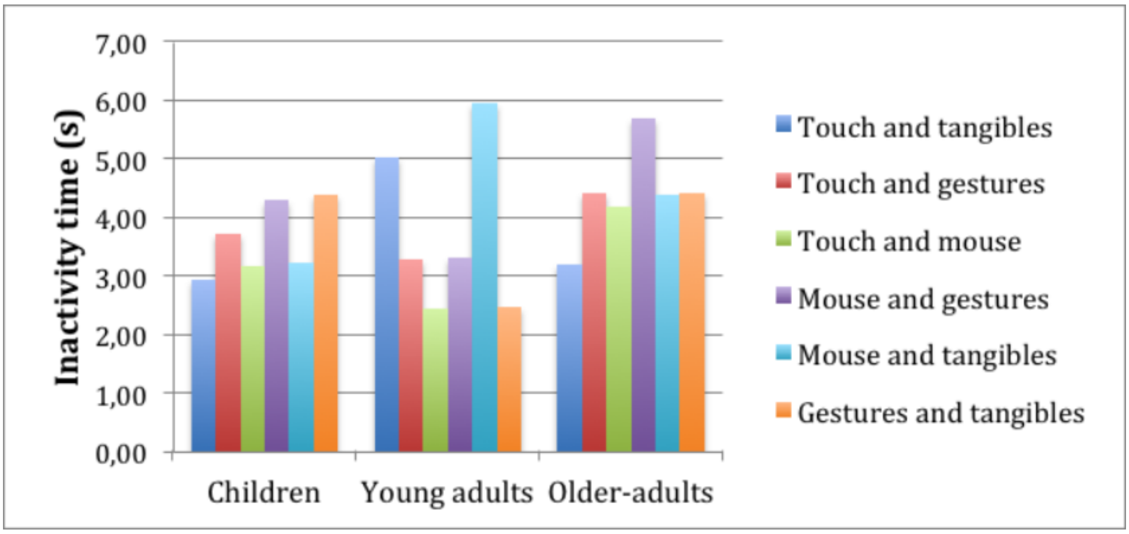


Figure 5.18 – Time of inactivity recorded during the non-parallel task by each sequence (in seconds).

Figure 5.19 presents the mean completion time of each age group registered during the experiment when using the different bimodal interfaces to complete the task with parallelism. The combination that registered the fastest results among the three age groups was, again, touch and mouse: children (23.39 s); young adults (16.85 s); older-adults (23.30 s). This situation appears to be very identical to the values recorded on the previous non-parallel activity. The combination with worse results was, just like the previous results, gestures and tangibles in all groups: children (50.50 s); young adults (37.70 s); older-adults (46.71 s). Also, these results were very close to the previous. As we can observe, aside from these interfaces being the ones to present the highest and lowest number of errors, they are also the slowest and fastest ones, respectively, to complete the tasks with parallelism.

We proceeded with a one-way ANOVA analysis of the data (which were mostly normally distributed, as in the previous case), that determined a statistically significant difference in the values: touch and tangibles ($F(2, 54) = 6.113$, $p = .004$); touch and gestures ($F(2, 53) = 6.191$, $p = .004$); touch and mouse ($F(2, 57) = 3.769$, $p = .029$); mouse and gestures ($F(2, 51) = 14.304$, $p = .000$); mouse and tangibles ($F(2, 54) = 10.066$, $p = .000$); and gestures and tangibles ($F(2, 53) = 3.775$, $p = .029$).

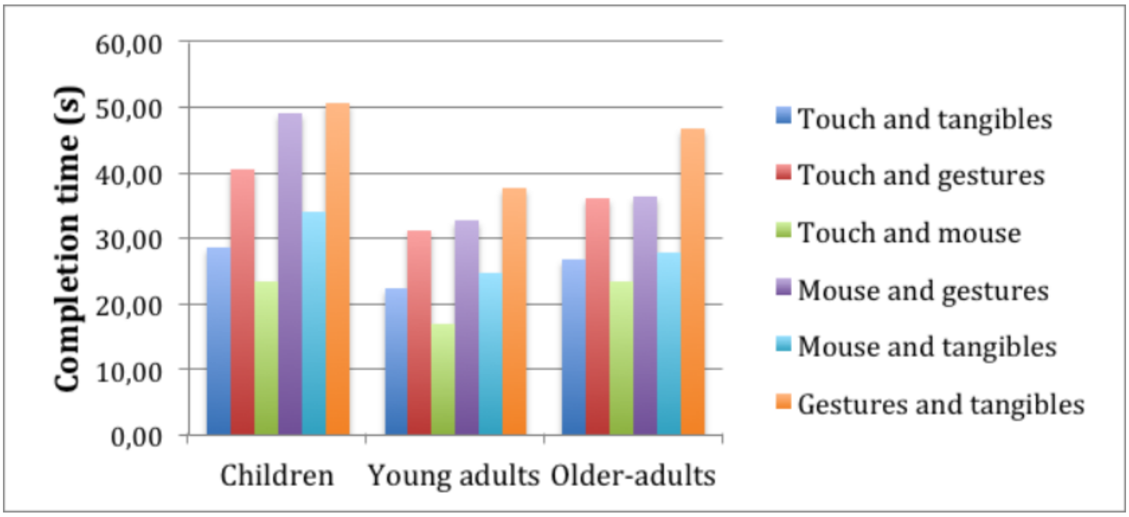


Figure 5.19 – Mean time taken to complete the task with parallelism (in seconds).

The bimodal interface touch and tangibles presented significant differences in mean time among the groups of children and young adults (p = .003), as indicated by Tukey post-hoc multiple comparisons test. However, this interface did not register any differences regarding the other groups: children versus older-adults (v = .485) and young versus older-adults (p = .059). Regarding the combination touch and gestures, children were significantly slower than young adults (p = .003), but there were no other significant differences regarding interaction by children versus older-adults (p = .305), nor by young versus older-adults (p = .112).

When it comes to touch and mouse, the combination to display the fastest mean results, no significant differences were found in neither the groups: children versus young adults (p = .051), children versus older-adults (p = .999), young versus older-adults (p = 056).

Regarding the combination of the computer mouse and tangibles, the group of young versus older-adults were statistically similar (p = .175), but children were significantly slower than young adults (p = .000) and older-adults (p = .023). As for the worst combination of all, gestures and tangibles, the group of older-adults versus children (p = .438) and young adults (p = .286) did not present statistical differences. However, young adults were significantly more efficient than children (p =

.023).

We also analyzed efficiency within each group resorting to a within-participant repeated-measures design, in order to understand if there were any significant differences in interaction within each group.

A repeated measures ANOVA with a Greenhouse-Geisser correction determined the group of children presented statistical differences in mean completion times ($F(2.826, 16.407) = 16.407$, $p <.000$). Just like in the previous non-parallel task, this activity also presented the same best and worst interfaces regarding mean completion times: touch and mouse (23.39 s), gestures and tangibles (50.50 s), respectively. Post-hoc pairwise comparisons tests using the Bonferroni correction showed that the mean completion times using touch and mouse were indeed statistically significantly lower when comparing to the other interfaces ($p <.010$), except for the touch and tangibles, where the difference was not significant ($p = .148$). Regarding the worst mean completion time, recorded by gestures and tangibles, this value was significantly higher than using touch and mouse ($p = .001$), the computer mouse and tangibles ($p = .023$), or even touch and tangibles ($p = .011$). On the contrary, the bimodal interfaces mouse and gestures ($p = 1.000$), and touch and gestures ($p = .197$) were not significantly faster than gestures and tangibles.

Regarding the young adults, as the Mauchly's sphericity was not assumed, like in the children's group, we applied the Greenhouse-Geisser adjustment to the repeated measures ANOVA. This group also presented statistically significantly different completion times ($F(3.582, 42.985) = 35.847$, $p <.000$). Post hoc tests using the Bonferroni correction revealed that the best mean time returned by the young adults using the combination touch and mouse (16.85 s) was significantly more efficient than any other bimodal interface: whether versus touch and tangibles ($p = .016$) or any other ($p = .000$). On the other hand, the combination with the worst results, gestures and tangibles, was indeed significantly more inferior than touch and tangibles ($p = .000$), touch and mouse ($p = .000$), or even mouse and tangibles ($p = .001$). However, this modality did not present statistically significant differences when compared to touch and gestures ($p = .408$) or mouse and gestures ($p = .288$).

The older-adults' efficiency in the parallel compound task also achieved a statistically significant difference between the mean times in each interface ($F(3.336, 60.043) = 24.744$, $p < .000$). Also similar to the other groups, the fastest interface (touch and mouse) was significantly different when compared with the others ($p < .01$), except comparing touch and tangibles ($p = 1.000$) or mouse and tangibles ($p = 1.000$). On the other hand, considering the lower value, gestures and tangibles did present significant differences of time when compared to the other combinations ($p < .05$).

On a different matter, we measured inactivity periods as well (Figure 5.20). All the groups appear to be constant in terms of the periods of inertia, although we can see the children used up slightly more time. This parallel activity presented lower values regarding inactivity when compared to the previous non-parallel activity. Here, the participants missed a maximum of approximately 4 seconds, by standing still, whereas in the previous case, participants missed approximately 6 seconds.

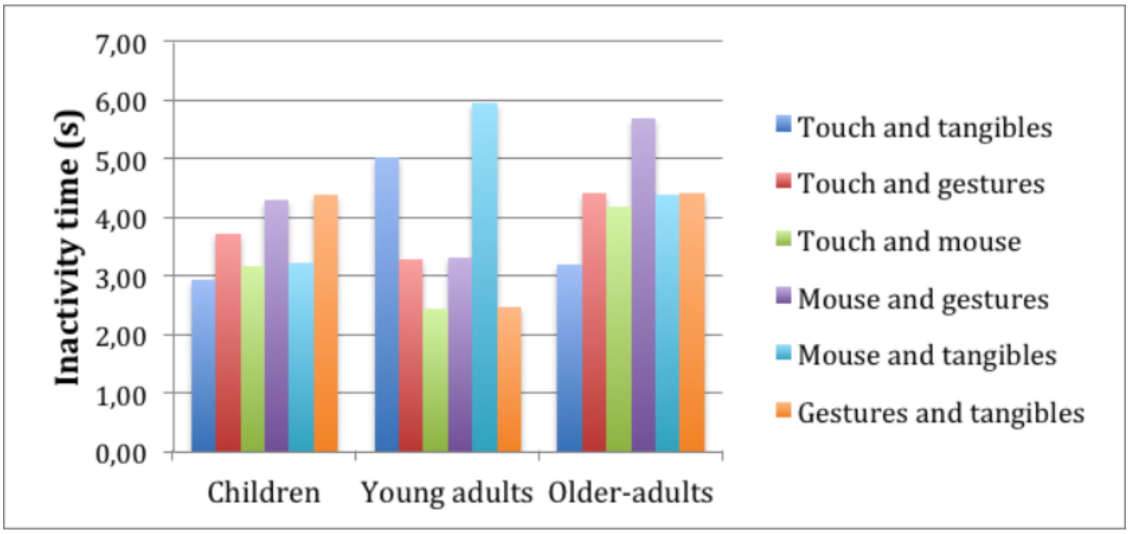


Figure 5.20 – Time of inactivity recorded during the parallel task by each sequence (in seconds).

Multimodal synergy:

Aside from the previous analyzed objective metrics such as speed, number of errors and task completion that are usually computed, we also find it important to evaluate the combination of the input modalities according to their relative modality efficiency and thus multimodal synergy.

As in the previous sections, this metric will be analyzed considering both the parallel and non-parallel activities of the trials.

- ***Task without parallelism:***

Figure 5.21 presents the multimodal synergy of the several combinations regarding each age group performing a non-parallel activity. As shown in the Figure, there are synergy fluctuations regarding the different combinations between the three groups.

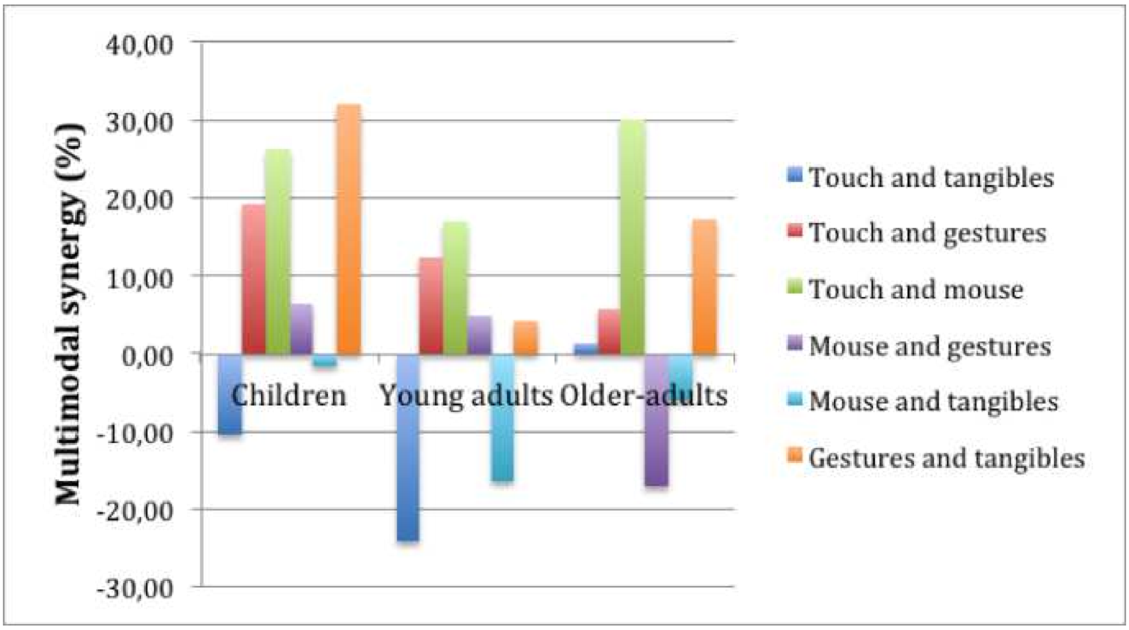


Figure 5.21 – Multimodal synergy (%) for the six multimodal interaction sequences in the non-parallel task.

We find it important to point the very low synergies of some combinations. These values are negative, and thus imply that there is no added value from efficiently combining these specific input modalities. These are the cases of: (a) touch and tangibles, in the groups of children and young adults; (b) mouse and tangibles, in all groups; and (c) mouse and gestures, in the group of older-adults. Indeed, touch and tangibles proved to have the lowest values for young adults, and was also the worst synergy of all.

On the contrary, some interfaces appear to display high multimodal synergies: such as touch and mouse. However, please note that even though the young adults exhibited the fastest mean completion times with this combination, their displayed multimodal synergy was the lowest of all groups. The synergy of the bimodal interface touch and gestures seems to decrease with relation to age. The older the groups are, the lower the value of synergy is: from 19% (by children), and 12% (by young adults), to 6% (by older-adults).

- ***Task with parallelism:***

Figure 5.22 presents the multimodal synergy of the several combinations regarding each age group performing a parallel activity. As shown in the Figure, there are synergy fluctuations regarding the different combinations between the three groups.

As we can see, the only group that does not present negative synergies is the older-adults' one. On the other hand, both children and young adults presented negative synergies, as is the case of the combination touch and tangibles in both groups. Moreover, children presented a high synergy value for the combination of gestures and tangibles, as opposed to the other interfaces in this group.

Participants' attitudes:

After overviewing the main results of this study's phase, we present the major difficulties sensed by the participants. These behaviors were annotated in a field diary created for the purpose of the experiment.

Regarding the group of the children, they were very excited during the tests. They wanted to perform the trials as fast as possible, and sometimes their accuracy and

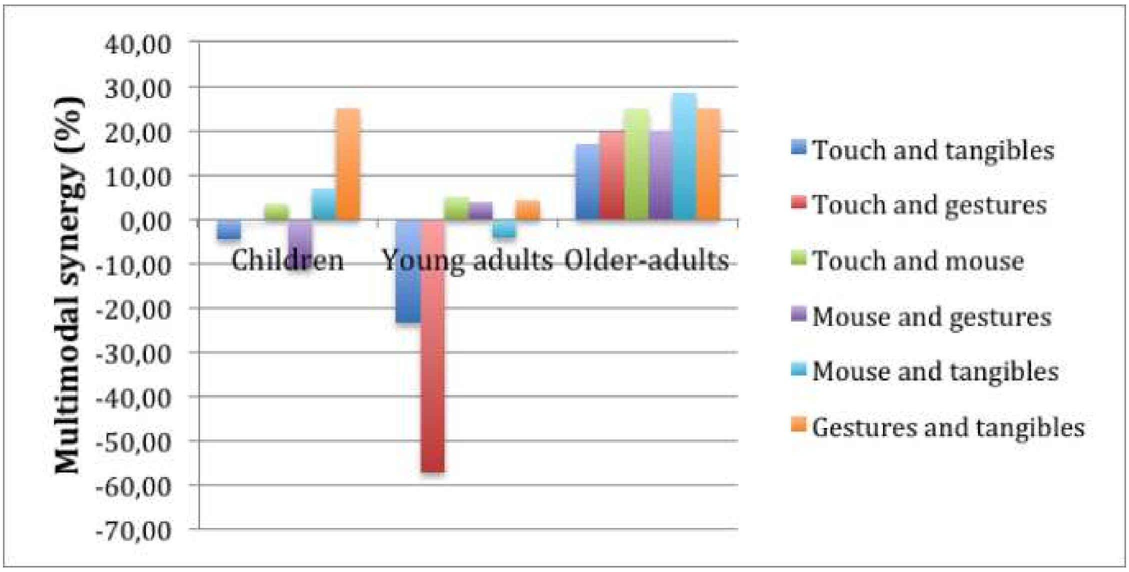


Figure 5.22 – Multimodal synergy (%) for the six multimodal interaction sequences in the parallel task.

attention would suffer. Also, they kept discussing the tangibles as being different and “cool”. The group of young adults tried to remain focused on the tasks, and did not speak much, whereas some participants in the group of older-adults would from time to time giggle or smile when they made an error or got confused about which input modality to use.

At first, participants were more confused with having to adapt to using two interaction modes, as was commented during the activities. However, after the trial period when they were practicing the game, they got more comfortable about it.

5.2.2.2 Discussion

We intend to understand which combination of two modalities could benefit interaction, causing a positive synergy, or on the other hand, which could be harmful, regarding the three age groups. For that purpose, we performed two types of compound tasks: with and without parallelism during interaction.

Regarding the task without parallelism, the combination of gestures and tangibles presented the most errors in all three groups. This is important to highlight, because this combination appears to cause more errors than all the other combinations united. This is also a validation of our previous findings, as we have discussed that gestures are not an advised modality for steering complex paths (regardless of the group), and note that the participants would cross the maze with gestures in this specific combination.

Considering the task with parallelism, the majority of errors was committed by children. Also, the total number of errors is extremely close to the one of the task without parallelism, which may indicate that in a bimodal interface, having parallelism in an activity may not influence effectiveness regarding all the age groups.

On the other hand, when analyzing the completion time of the non-parallel tasks with each combination within each group of participants, it seams that: (1) touch and mouse is the fastest and more accurate bimodal interface to accomplish these types of tasks; whereas (2) gestures and tangibles is both the slowest and less accurate bimodal interface regarding compound non-parallel tasks.

Considering the combination that obtained the best results in completing the task (touch and mouse), young adults were superior to the other groups when combining the two technologies that are most common for them during daily interaction. In turn, when analyzing efficiency within each group, the mean time using touch and mouse was significantly lower than using the other interfaces in the group of young adults. This may indicate that this age group can indeed take better advantage of this combination regarding efficiency in tasks without parallelism.

Regarding inactivity periods, for young adults gestures and tangibles (the combination that presented the worst results in terms of mean time) did not record such high values of inactivity. This may implicate that although the users take longer to perform the task, they are more constant and thus have fewer periods of inactivity.

Considering the task with parallelism, the pattern of values recorded regarding the different combinations was very similar to the ones presented by non-parallel tasking. This could indicate that, regarding efficiency (completion times), the users do not seem to have fluctuations in performance when executing parallel or non-parallel activities. However, the interfaces show more inconsistencies among the groups. Also, when it comes to touch and mouse, the combination to display the fastest mean results, no significant differences were found in neither the groups, which may indicate that this combination, regardless of the group, presents the best efficiency in parallel tasks.

Regarding inactivity periods, the task with parallelism presented lower values when compared to the previous non-parallel activity. Here, the participants missed a maximum of approximately 4 seconds, by standing still. In the previous case, participants missed approximately 6 seconds. This may indicate that users may waste less time during interaction when dealing with parallel activities. This point could be interesting to further explore, as we could try to understand if productivity increases using bimodal interfaces to complete parallel tasks. Also, the mean times taken to complete the task do not directly relate to the respective inactivity time of each interface, that is, the longer times of completion certain interfaces display are not in line with inertia periods: a task may take longer to complete with a certain interface, but that fact does not mean that the user will also present higher inactivity periods using that same interface.

On a different matter, when analyzing how "synergetic" the interface is in a non-parallel task, some interfaces appear to have a high multimodal synergy, and other display negative values. Touch and mouse, being the combination to show the best efficiency values when it comes to completion times regarding all three groups, also displayed high multimodal synergy values. However, please note that even though the young adults exhibited the fastest mean completion times with this interface, their displayed multimodal synergy was the lowest of all groups. This may indicate that, even though this group is able to quickly execute the tasks with the two modalities, they do not take advantage of their synergies as the other groups. There is another interesting fact to highlight, regarding the combination of gestures and tangibles. This combination presented the worst results considering mean completion times of this activity. However, its multimodal synergy is higher in all groups when compared to other interfaces. Indeed, in the case of children, this interface presents the best synergy value of all interfaces among all groups. The same is valid for older-adults, which showed better synergy values than young-adults.

We might ponder that, even though those age groups performed slower with this interface, their ratio of relative modality efficiency and usage is very symmetric and, thus, they may be able to take better advantage of the synergies. This is not the case for the young adults that have lower synergy values regarding this bimodal interface. Also, the synergy of the bimodal interface touch and gestures seems to decrease with relation to age. Moreover, if we look at the values for mouse and gestures (being the latter the same modality for both interfaces), we can observe that even though one modality is the same, it does not mean the synergies will be. Indeed, the gesture-based interface influences the previous modality used, whether touch or the computer mouse. Here, touch is positively influenced by the gestures, whereas the computer mouse does not display such positive synergy ratios.

Therefore, we might assume that: (a) the age groups may affect multimodal synergy values, i.e. the same interface can present very different values depending on the age group; (b) two different combinations with one same modality may cause very different multimodal synergy values, as is the case of touch and gestures / mouse and gestures; and (c) fastest mean results do not necessarily mean high synergy ratio, e.g. the combination gestures and tangibles presented high mean completion times, and also high synergy ratios.

Regarding the task with parallelism, all interfaces have a quite consistent synergy ratio amongst them. If we relate to these participants' behavior during the tests (as will be further discussed in the next topic), they were calmer during interaction and not as nervous to complete the task. This could indicate that they can manage both interaction modalities better when they are used in parallel. Also, we find it important to highlight the negative synergy of the bimodal interface touch and gestures recorded by young adults. Indeed, this combination seems to suffer immensely with its adaptation to a parallel activity, as this value is far different from the one displayed in the previous non-parallel task.

Moreover, children presented a high synergy value for the combination of gestures and tangibles, as opposed to the other interfaces in this group. This is indeed interesting, because we are finding that in non-parallel as well as in parallel tasks, children are able to take better advantage of the features of both modalities to improve their synergy: when compared, evidently, to their unimodal parts. They may not perform faster or more efficiently, but they seem to know how to combine them better as to improve their synergy ratio.

5.2.2.3 Summary

In this phase of the study, it was our goal to understand which combination of two modalities could cause positive synergies and present the best results regarding effectiveness and efficiency in two types of compound tasks: with and without parallelism during interaction.

Regarding the number of error committed during interaction, the combination of gestures and tangibles was the one to present the highest number of faults, regardless of the type of task. Also, the group of children made the most mistakes.

During interaction, all groups displayed similar results: (1) touch and mouse is the fastest and more accurate bimodal interface to accomplish the tasks; (2) gestures and tangibles was both the slowest and less accurate bimodal interface; (3) their mean completion times were all significantly different when compared among groups; and (4) the multimodal synergy analyzed for the combinations were distinct depending on the age group, which may indicate that the group's characteristics may influence interaction, and not only the interaction modality used.

5.2.3 Synergies between three modes of interaction

In this stage of the case study, participants were asked to complete the activities resorting to three different interaction modalities required. The combinations are as follows:

- Computer mouse, touch and gestures;

- Gestures, computer mouse and tangibles;
- Tangibles, computer mouse and gestures;
- Touch, computer mouse and tangibles;
- Tangibles, touch and computer mouse;

5.2.3.1 Results

In this level of the game, we aim at understanding which combination of three modalities could benefit interaction. As in the previous phase, we also performed two types of compound tasks: with and without parallelism during interaction.

This section presents the results and discussion of the experimental tests for the two compound tasks, regarding effectiveness (error rate and completion success), efficiency (completion time), user preference and multimodal synergy of the interface. Moreover, such as the previous experiments, we review the participants' behavior during the tests and their satisfaction towards the different interfaces.

Effectiveness:

The same circumstances of error evaluation from the previous test apply to this phase of the study.

- ***Task without parallelism:***

Figure 5.23 presents the total number of errors registered during the experiment. In total, there were 356 errors in both phases, being approximately 71% committed by children. The combination that presented more errors regarding all the groups was tangibles, mouse and gestures, with a total of 126 faults. The fewest errors were recorded using the combination tangibles, touch and mouse: children (25 errors), young adults (only 2 errors), and older-adults (5 errors).

Comparing to the previous bimodal interfaces, participants made far less errors when combining three input modalities. However, this may be explained by our own experimental design, as in this case, we avoided the combination of the interfaces with the previous worst results. Therefore, the specified the best-case scenarios for the combination of the three interfaces, and so we can see that, as far as effectiveness, the results improved.

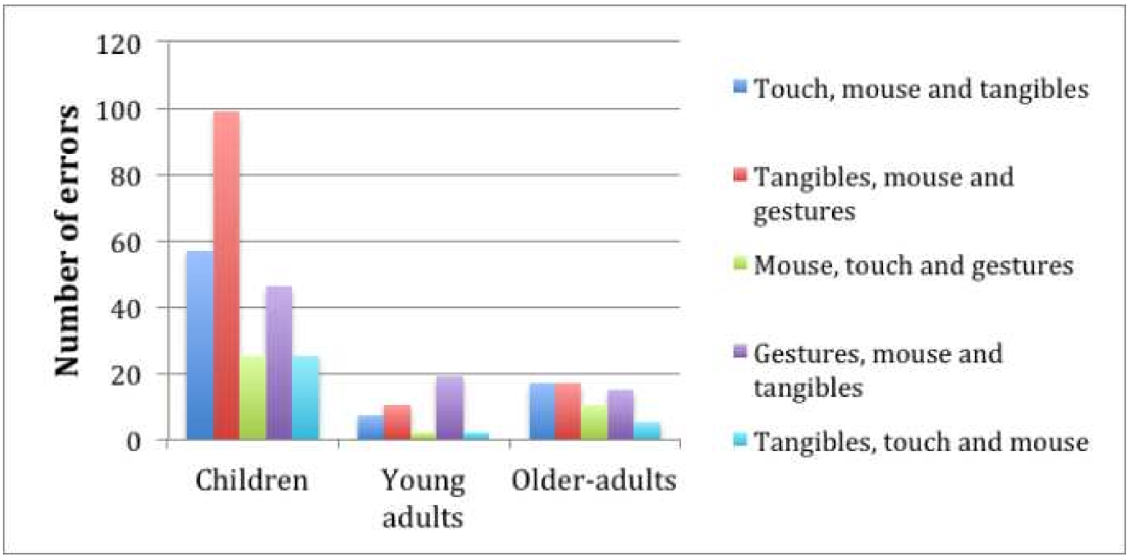


Figure 5.23 – Total number of errors presented during the experiment (task without parallelism).

As we did on the previous test with bimodal interfaces, aside from collecting the number of errors of the groups in the labyrinth, we also recorded their positions. Figure 5.24 shows the position of the errors committed by the three age groups in the labyrinth using the various input modalities during the non-parallel task.

- ***Task with parallelism:***

Regarding the task with parallelism, there were a total of 306 errors (Figure 5.25), being the majority also committed by children (154 errors).

Moreover, resorting to the combination of tangibles, touch and mouse (the same as

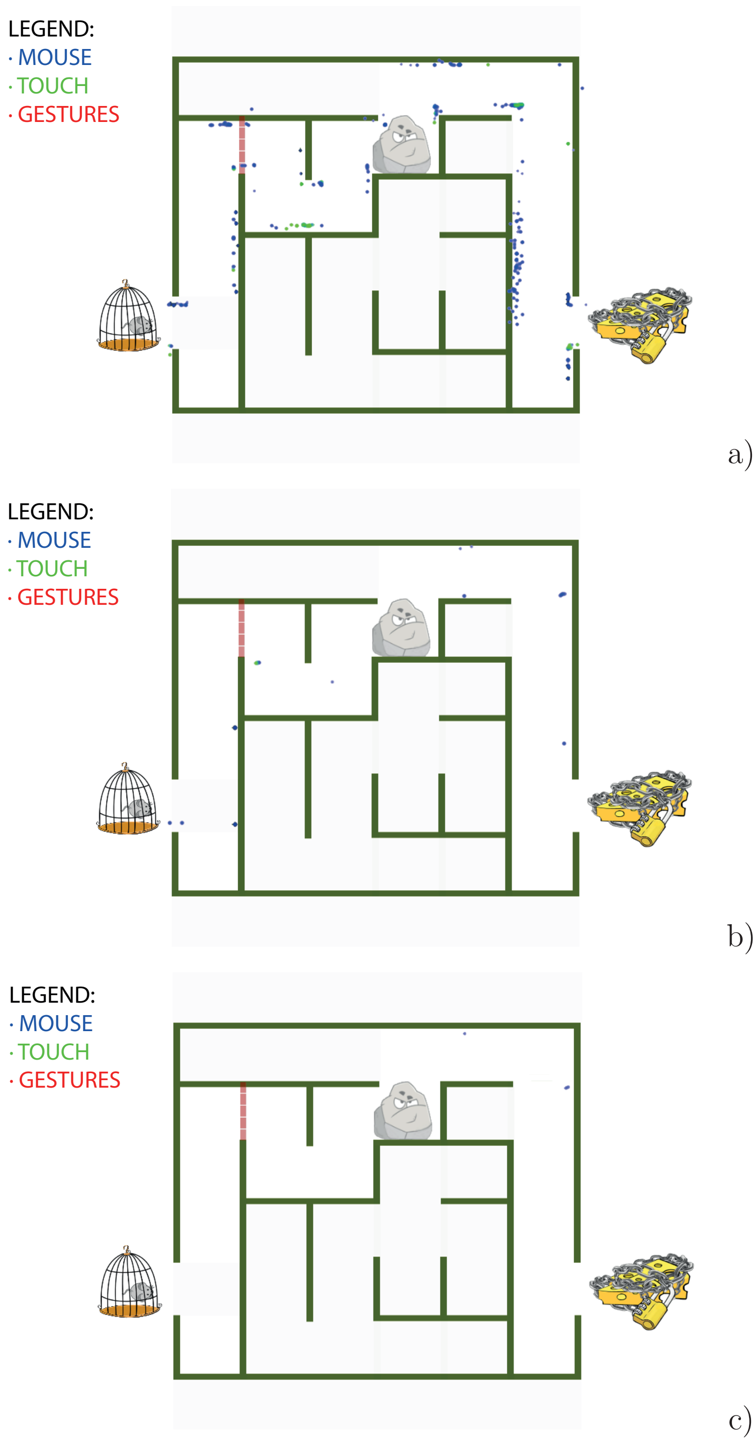


Figure 5.24 – Position of the errors committed in the labyrinth using the various combinations (in a non-parallel task): a) Children; b) Young adults; c) Older-adults.

the non-parallel task) the participants recorded fewer faults: children (11 errors), young adults (only 1 error) and older-adults (7 errors).

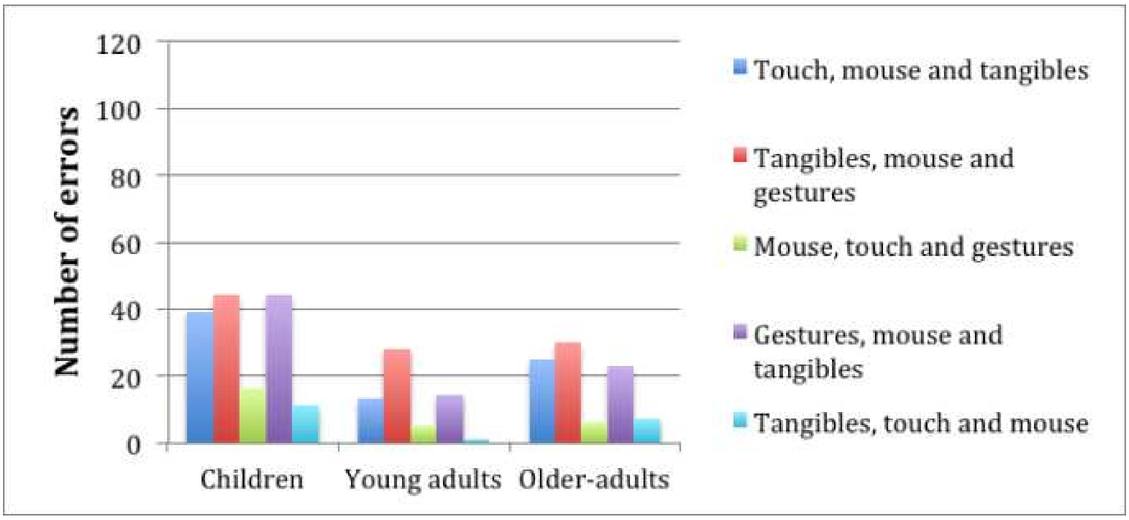


Figure 5.25 – Total number of errors presented during the experiment (task with parallelism).

Also, Figure 5.26 shows the position of the errors committed by the three age groups in the labyrinth using the various input modalities during the parallel task.

Efficiency:

The completion time of each task follows the same principles of the previous experiment: the time of each task was measured between the first action performed by the user (as soon as entering the labyrinth) and the successful insertion of the code, minus the inactivity period.

- ***Task without parallelism:***

Figure 5.27 presents the mean completion time that each age group registered during the experiment when using the combination of three interfaces. As we can see in the chart, the fastest combination used by children was mouse, touch and gestures (21.97 s). However, this was not the fastest combination for the other age groups.

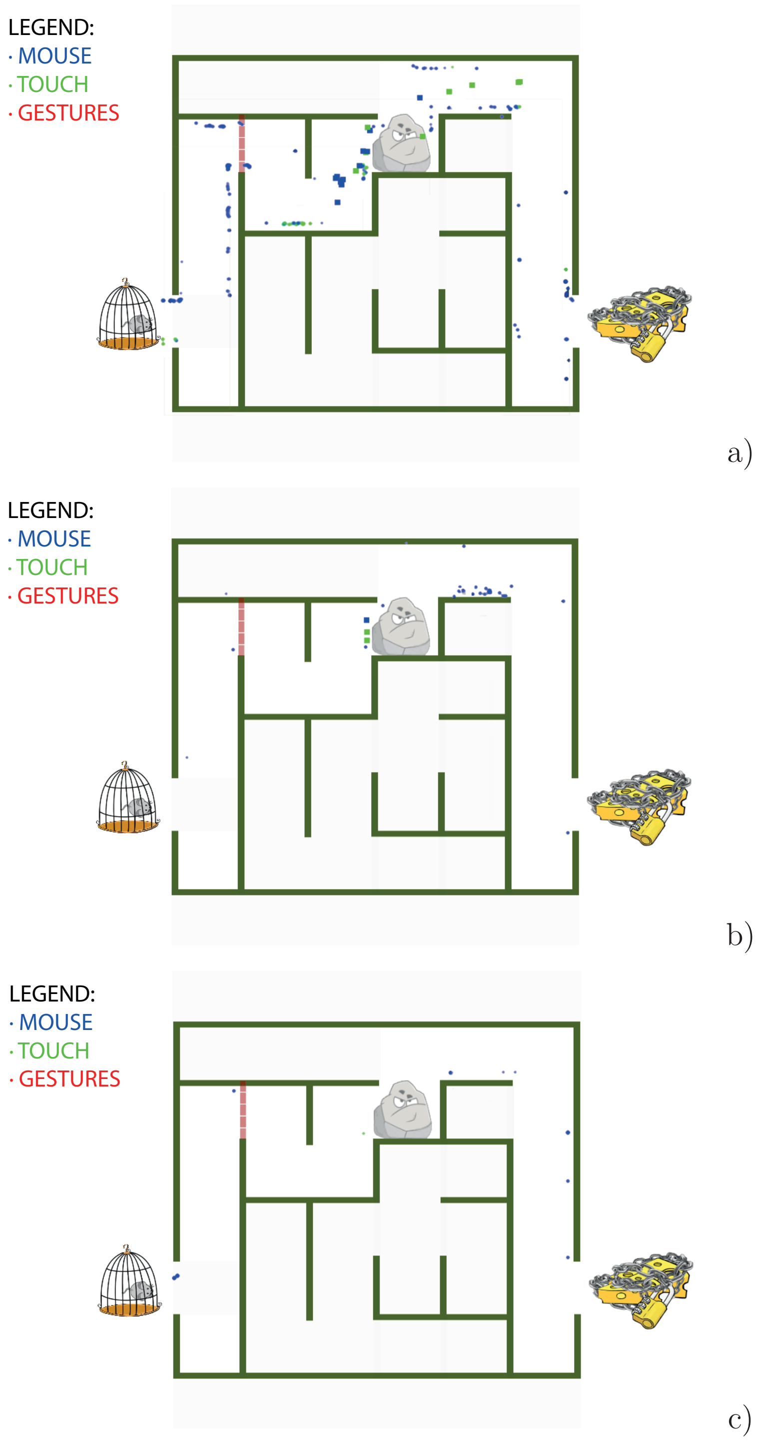


Figure 5.26 – Position of the errors committed in the labyrinth using the various combinations (in a parallel task): a) Children; b) Young adults; c) Older-adults.

Instead, touch, mouse and tangibles were the most efficient for young (15.32 s) and older-adults (19.53 s). On the other hand, the three groups presented the worst results with the same combination (gestures, mouse and tangibles) with 36.52 s, 23.28 s and 27.89 s achieved by children, young and older-adults, respectively.

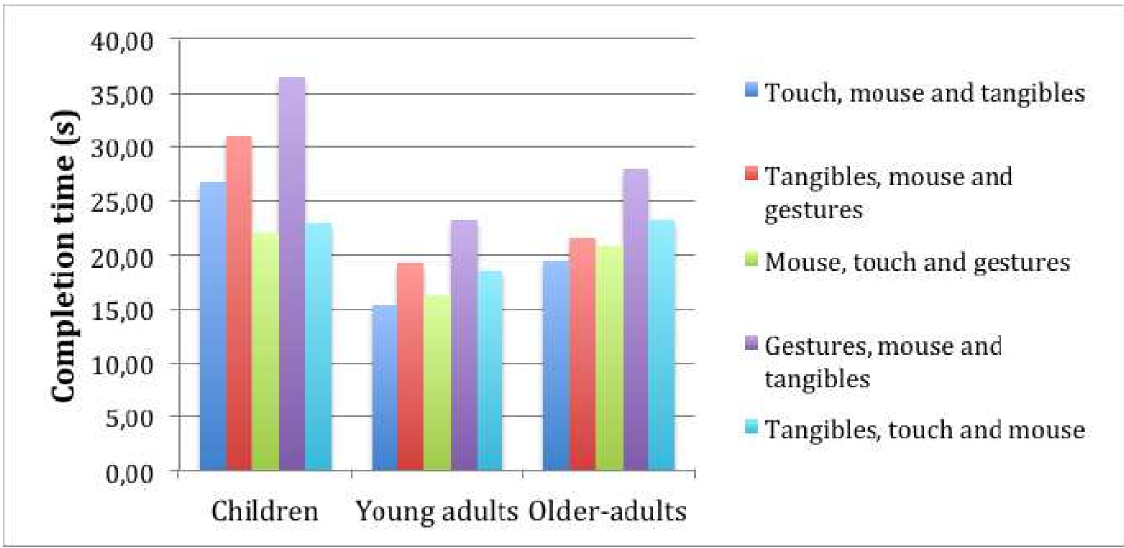


Figure 5.27 – Mean time taken to complete the task without parallelism (in seconds).

Indeed, the worst three-interface combination is common for the three age groups, but the best fit for the participants is not the same for the three groups, which indicates that when it comes to efficiency, gestures, mouse and tangibles are definitely a wrong choice to use as a multimodal combination.

A Shapiro-Wilk test showed us most of the data were normally distributed, except for the combinations of: touch, mouse and tangibles (p = .002) / tangibles, mouse and gestures (p = .002) in the group of children; and also tangibles, touch and mouse in all age groups (p = .003). Next, a one-way ANOVA analysis showed that there are statistically significant differences in all groups regarding most combinations: touch, mouse and tangibles ($F(2,57) = 10.300$, $p = .000$); tangibles, mouse and gestures ($F(2,56) = 13.065$, $p = .000$); mouse, tangibles, gestures ($F(2,56) = 8.146$, $p = .001$); gestures, mouse and tangibles ($F(2,57) = 8.783$, $p = .000$). The only combination that was not significantly different was tangibles, touch and mouse ($F(2,57) = 1.272$, $p = .288$).

When considering the three age groups, results of efficiency were mainly significantly different. Regarding the combination of touch, mouse and tangibles, between age groups, results show that this combination is not significantly different amongst young and older-adults (p = .227), but it does present differences between children and young adults (p = .000), and children and older-adults (p = .018). Indeed, this combination was significantly slower for children when compared to young and older adults, but in turn showed to be the fastest for these two age groups. As for the combination tangibles, mouse and gestures, it presents the same pattern: significantly different between children versus young adults (p = .000), and children and older-adults (p = .001); but not between young and older-adults (p = .610). The sequence mouse, touch and gestures was indeed the fastest for children and showed significant differences when compared to young adults (p = .002), but not to older-adults (p = .968). Understandably, young and older-adults also presented statistical differences (p = .004). Statistical differences were also encountered in the combination of gestures, mouse and tangibles: children versus young adults (p = .000); children versus older-adults (p = .025); but not between young and older-adults (p = .328).

Indeed, the only combination that did not show any statistically significant differences between age groups was tangibles, touch and mouse: children versus young adults (p = .391) and older-adults (p = .992); and young versus older-adults (p = .329). This indicates that this sequence is not the fastest or the slowest of each group, but it is also not statistically different between groups.

Additionally to comparing efficiency between each group regarding the different combinations, we also analyzed efficiency within each group resorting to a within-participant repeated-measures design.

Regarding the group of the children, a repeated measures ANOVA with a Greenhouse-Geisser correction determined there was a significant difference between the tested sequences ($F(3.027, 54.489) = 7.382$, p <.000). Post-hoc pairwise comparisons tests using the Bonferroni correction showed that the mean completion time using the combination mouse, touch and gestures (21.97 s) was indeed statistically significantly lower when comparing to tangibles, mouse and gestures (p = .034); and gestures, mouse and tangibles (p = .004). However, it did not display significantly different results when compared to the following sequences: touch, mouse and tangibles (p = .697); or tangibles, touch and mouse (p = 1.000).

A repeated measures ANOVA with a Greenhouse-Geisser correction also determined that the group of young adults presented statistically different results regarding the various combinations ($F(2.257, 42.886) = 10.942$, p <.000). Post-hoc pairwise comparisons tests using the Bonferroni correction confirmed that the mean completion time using the sequence touch, mouse and tangibles (15.32 s) was significantly lower than when using tangibles, mouse and gestures (p = .001), or gestures, mouse and tangibles (p = .001); but now when using mouse, touch and gestures (p = 1.000), or tangibles, touch and mouse (p = .174).

Just the same as the other groups, older-adults seem to present a statistically significant difference between the mean times using the combinations ($F(1.352, 24.332)$ = 4.151, p <.042). Also, similar to the group of young adults, the fastest combination (touch, mouse and tangibles, with 19.53 s) was significantly different when compared with the sequence gestures, mouse and tangibles (p <.000). However, it did not present statistical differences between the other sequences: tangibles, mouse and gestures (p = .112); mouse, touch and gestures (p = .373); or tangibles, touch and mouse (p = 1.000).

Also, we recorded the periods of inactivity during interaction. Figure 5.28 shows that all groups presented periods of inactivity regarding all sequences, although the three groups presented lower inactivity periods using touch, mouse and tangibles. This sequence was also the best in terms of mean completion time for the groups of young and older-adults. Two sequences recorded the highest inactivity periods: tangibles, touch and mouse; and tangibles, mouse and gestures. However, these combinations were not the ones to present the highest mean completion time. On the contrary, gestures, mouse and tangibles (the worst combination regarding efficiency) did not present the highest inactivity time.

- ***Task with parallelism:***

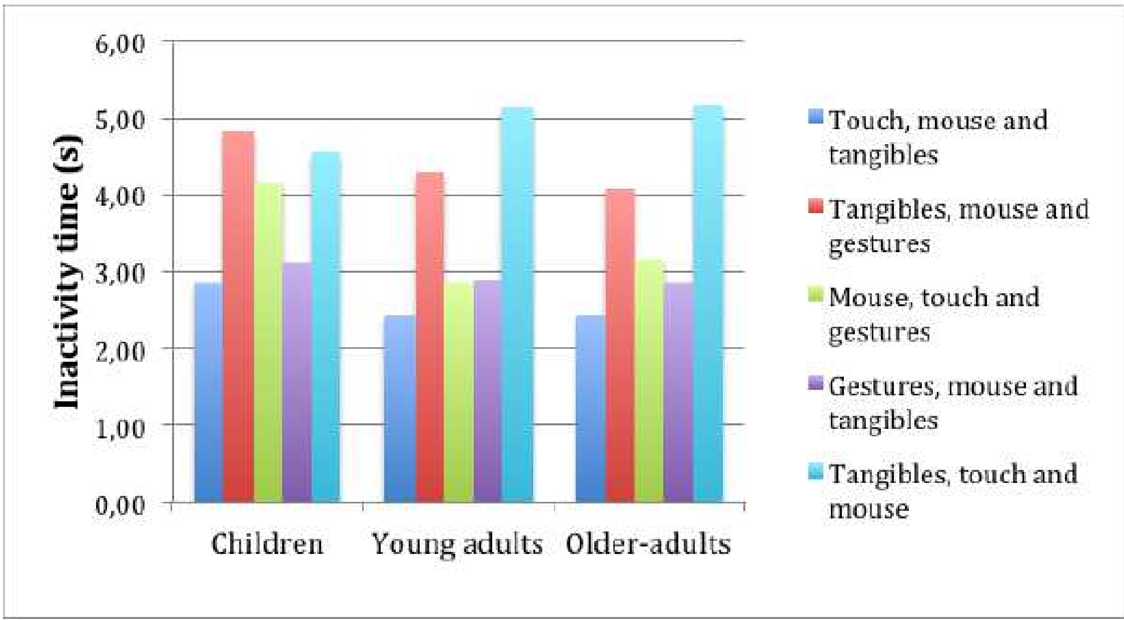


Figure 5.28 – Time of inactivity recorded during the non-parallel task by each sequence (in seconds).

The mean completion time of each age group registered during the experiment when using the different sequences to complete the task with parallelism is shown in Figure 5.29.

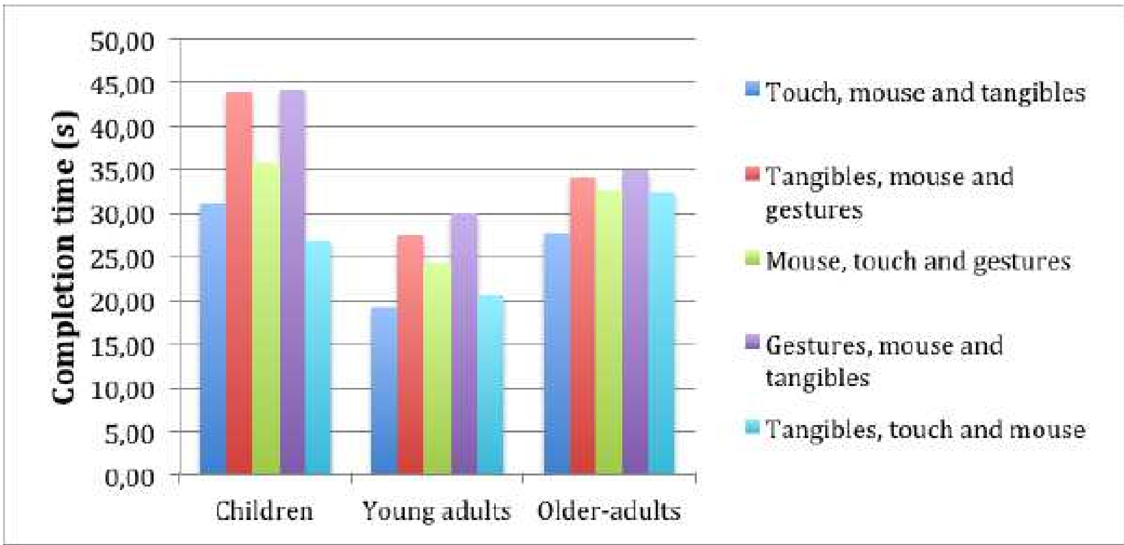


Figure 5.29 – Mean time taken to complete the task with parallelism (in seconds).

The combination that registered the fastest results was not the same for the three age groups. The group of children managed to complete the parallel tasks quicker with tangibles, touch and mouse (26.91 s). On the other hand, the other two groups performed better with touch, mouse and tangibles (young adults recorded 19.21 s and older-adults, 27.69 s). The same pattern is also presented in the worst mean completion time results. Children took longer to complete the parallel task using the following sequences: tangibles, mouse and gestures (44.05 s); and gestures, mouse and tangibles (44.09 s). In turn, the latter sequence was also the one that presented worse results with young and older-adults (30.09 s and 34.93 s, respectively).

A Shapiro-Wilk test showed us that the data were normally distributed. As such, a one-way ANOVA analysis of the data determined a statistically significant difference in the values: touch, mouse and tangibles ($F(2, 57) = 16.192$, $p = .000$); tangibles, mouse and gestures ($F(2, 57) = 9.295$, $p = .000$); mouse, touch and gestures ($F(2, 57) = 9.415$, $p = .000$); gestures, mouse and tangibles ($F(2, 57) = 6.260$, $p = .0043$); and tangibles, touch and mouse ($F(2, 57) = 6.993$, $p = .002$).

A Tukey post-hoc multiple comparisons test showed the significant differences in the three groups regarding the sequences. The sequence touch, mouse and tangibles presented statistical differences between the groups of young adults versus children ($p = .000$) and older-adults ($p = .001$). However, the children and older-adults presented similar values ($p = .240$). Regarding the combination tangibles, mouse and gestures, young and older-adults present statistically similar values ($p = .206$), although there are significant differences between children versus young adults ($p = .000$), and versus older-adults (p .034).

Moreover, the combination mouse, touch and gestures presents statistical differences in results between the groups of young adults versus children ($p = .000$) and older-adults ($p = .009$), but not between children and older-adults ($p = .515$).

As for one of the worst sequences (gestures, mouse and tangibles), children present significant differences in efficiency when compared to young adults ($p = .003$). However, the other groups do not: children versus older-adults ($p = .067$); and young versus older-adults ($p = .456$).

Finally, concerning the sequence tangibles, touch and mouse, results were mainly not significantly different (children versus young adults: p = .125; and versus older-adults: p =.196), except for the groups of young versus older-adults (p = .001).

We also analyzed efficiency within each group resorting to a within-participant repeated-measures design, in order to understand if there were any significant differences in interaction within each group.

A repeated measures ANOVA with a Greenhouse-Geisser correction determined the group of children presented statistical differences in mean completion times (F(2.651, 50.376) = 15.137, p <.000). Post-hoc pairwise comparisons tests using the Bonferroni correction showed that the mean completion times using the sequence tangibles, touch and mouse (26.91 s) was statistically faster than three other sequences: tangibles, mouse and gestures (p = .000); mouse, touch and gestures (p = .006); and gestures, mouse and tangibles (p = .001). However, it was not significantly faster than the sequence touch, mouse and tangibles (p = .053).

Regarding the young adults, as the Mauchly's sphericity was not assumed, like in the children's group, we applied the Greenhouse-Geisser adjustment to the repeated measures ANOVA. This group also presented statistically significantly different completion times (F(2.526, 47.988) = 8.953, p <.000). Post hoc tests using the Bonferroni correction revealed that the best mean time returned by the young adults using the combination touch, mouse and tangibles (19.21 s) was significantly more efficient than two other sequences: mouse, touch and gestures (p = .029), and gestures, mouse and tangibles (p = .000). However, it was not significantly more efficient than tangibles, mouse and gestures (p = .077), or tangibles, touch and mouse (p = 1.000).

On the contrary, the older-adults' efficiency in the parallel compound task did not achieve a statistically significant difference between the mean times in each interface (F(2.481, 47.137) = 2.904, p <.054). Indeed, the sequence touch, mouse and tangibles only displayed significant differences when compared to tangibles, mouse and gestures (p = .016), and to gestures, mouse and tangibles (p = .001). Indeed, there were not statistical differences between the first sequence and mouse, touch and gestures ($p = .056$), or tangibles, touch and mouse ($p = 1.000$).

On a different matter, we measured the periods of inertia during interaction as well (Figure 5.30). Overall, the children used up slightly more time. However, when comparing this period with the previous type of non-parallel task, the groups seem to have been more active during this activity. For children, they were more passive with the sequence gestures, mouse and tangibles (5.59 s), and more dynamic with touch, mouse and tangibles (3.15 s). Similarly to children, young and older-adults did not waste as much time with touch, mouse and tangibles (2.25 s and 2.39 s, respectively). However, these two groups were different when it comes to the sequences that displayed the longest periods of inactivity: young adults with tangibles, mouse and gestures (4.15 s); and older-adults with tangibles, touch and mouse (5.06 s).

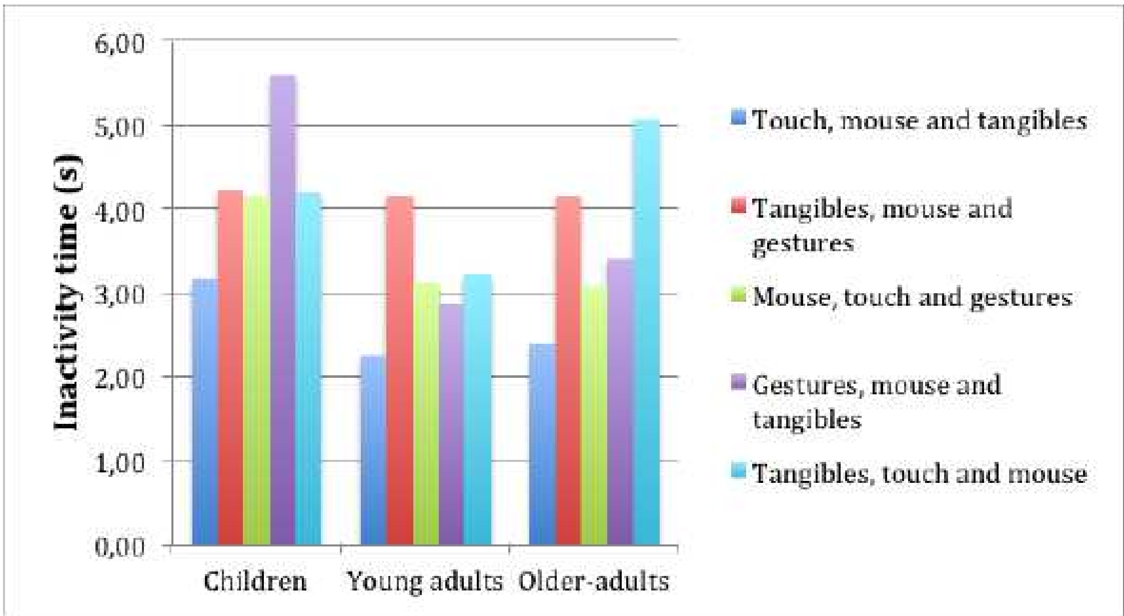


Figure 5.30 – Time of inactivity recorded during the parallel task by each sequence (in seconds).

Multimodal synergy:

Aside from analyzing efficiency and number of errors committed, we also evaluated the combination of the three input modalities according to their relative modality efficiency and thus multimodal synergy. As in the previous sections, this metric will be analyzed considering both the parallel and non-parallel activities of the trials.

- ***Task without parallelism:***

Figure 5.31 presents the multimodal synergy of the several combinations regarding the three age groups performing the non-parallel activity. As we can see in the chart, not all sequences return positive synergies. This same situation happened in the previous experiment when analyzing bimodal interfaces. Indeed, all groups present the same difficulties when using two sequences: touch, mouse and tangibles; and tangibles, touch and mouse. Moreover, older-adults also showed negative synergies when combining gestures, mouse and tangibles. On the other hand, mouse, touch and gestures is the sequence that presents the higher synergetic values in all groups.

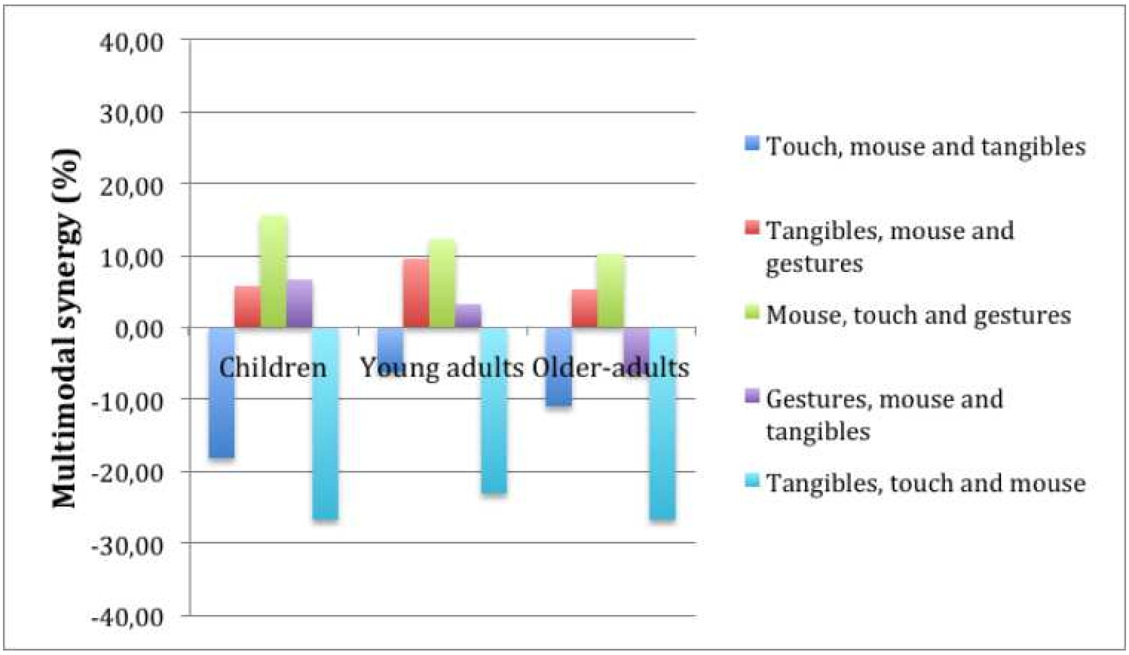


Figure 5.31 – Multimodal synergy (%) for the five multimodal interaction sequences in the non-parallel task.

- ***Task with parallelism:***

Figure 5.32 presents the multimodal synergy of the several combinations regarding each age group performing a parallel activity. As shown in the chart, there are synergy fluctuations regarding the different combinations between the three groups. All the groups present positive and negative synergy levels, depending on the sequence.

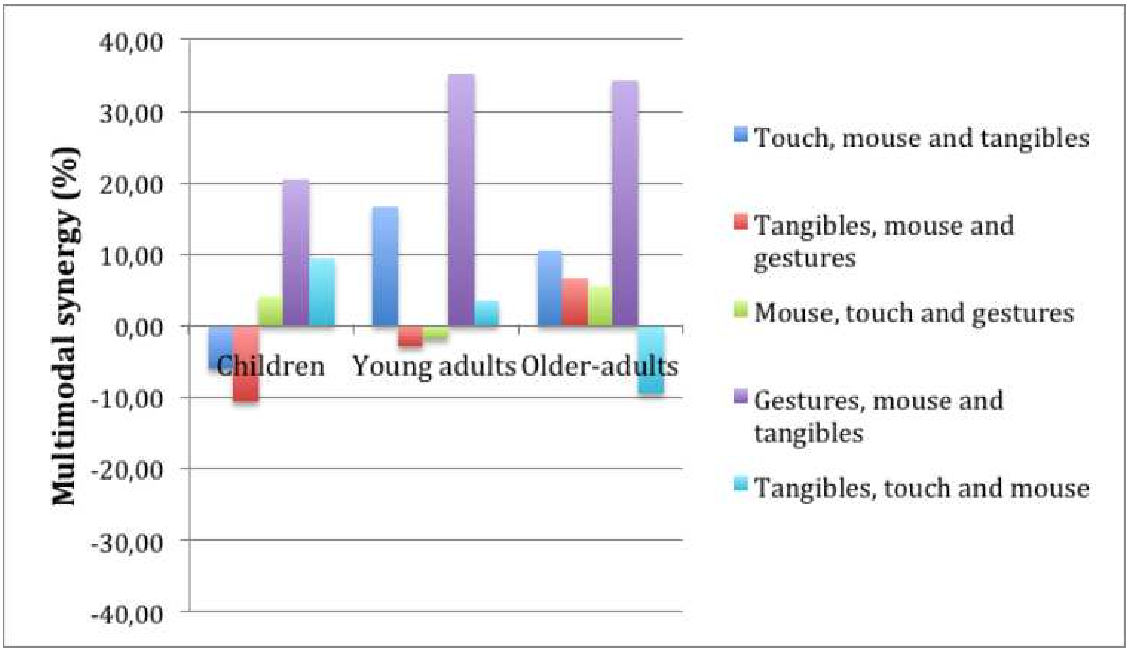


Figure 5.32 – Multimodal synergy (%) for the five multimodal interaction sequences in the parallel task.

First, children do not present positive synergies with touch, mouse and tangibles, neither with tangibles, mouse and gestures. Young adults present negative synergies with tangibles, mouse and gestures, as well as mouse, touch and gestures. On the other hand, older-adults only present negative synergies with one sequence: tangibles, touch and mouse. Overall, the combination with the best synergy results in all three groups was gestures, mouse and tangibles.

Participants' attitudes:

We present the major difficulties sensed by the participants during this test. These attitudes were annotated in a field diary created for the purpose of the experiment.

We were not aware of significantly different attitudes towards the use of three input modalities to complete the tasks. The participants behaved similarly to the previous activities, and did not present any commentaries worth of discussion.

However, we did notice that the participants were more anxious during the tests. They did make more sounds of frustration and often complained that they were lost about the task and which input modality to use.

5.2.3.2 Discussion

Like in the previous study of bimodal interfaces, the overall number of errors is similar between both types of tasks: with and without parallelism. This could indicate once more that in multimodal interface, parallelism may not influence effectiveness regarding all the age groups.

On a different matter, by analyzing the number of errors committed and the participants' efficiency in the non-parallel task, we can acknowledge that the these two variables are not related, that is, the combination that led to more errors did not turn out to be the least efficient.

Regarding the different periods of inactivity during interaction, the sequence touch, mouse and tangibles was the best in terms of mean completion time for the groups of young and older-adults. This suggests that this sequence can be presented as the most efficient and the most dynamic, or "active", for these two groups. Similar to the interaction with bimodal interfaces, these results may suggest that although the users take longer to perform the task, thus having higher mean times, they are more constant and have fewer periods of inactivity.

Unlike the previous experiment that analyzed bimodal interfaces, the pattern of the various sequences was not similar between parallel and non-parallel tasking. This may indicate that when putting three interfaces together, the type of compound task can also influence interaction. This is clear when comparing children's efficiency times of the sequence mouse, touch and gestures in both types of tasks. Indeed, the results were different as in non-parallel tasks this sequence was one of the best, and in parallel tasks, it is the third worst sequence of all.

Also, most of the combinations do not display significant differences between the groups of children and older-adults in parallel activities. This could suggest that these two groups show similar efficiency results in this type of tasks.

On the other hand, regarding periods of inertia during interaction, parallel tasking encourages the users to be more active and do not waste as much time. This is also interesting to explore because multimodal interaction favors the active use of different modalities and thus does not encourage inactivity periods.

In turn, considering multimodal synergy of the several combinations regarding the three age groups performing the non-parallel activity, we can assume that when using three input modalities multimodally, there are also combinations that do not help the user during interaction and thus do not generate positive fusion levels. The sequence that showed the best results in terms of efficiency in the groups of young and older-adults presents negative synergies: this indicates that although the results of interaction is positive, the relative modality efficiency considering the input modalities is not. On the other hand, mouse, touch and gestures is the sequence that presents the higher synergetic values in all groups. As this combination was one of the best regarding mean completion times, this could be a good choice for the three groups.

Moreover, overall there are sequences that seem to present decreasing values over time. Indeed, if we look at two specific sequences (mouse, touch and gestures; and gestures, mouse and tangibles), they present worse results in relation to age, that is, the older the groups are, the lower the level of synergy presented. Please remember that this scenario also happened in the previous experiment.

Therefore, we might assume that (similarly to the previous results concerning bi-modal interaction): (a) the age groups may affect multimodal synergy values; (b) a change in the sequence and not the chosen input modalities can affect the multimodal synergy values; and (c) fastest mean results do not necessarily mean high synergy ratios.

Regarding the task with parallelism, as we could observe in the previous experiment about bimodal interaction, in parallel activities older-adults tend to display better synergy levels in the majority of combinations than the other groups. Again, this could indicate that they can manage the three interaction modalities better when they are used in parallel.

Overall, the combination with the best synergy results in all three groups was gestures, mouse and tangibles. Remember that this combination was one of the ones to present a higher mean completion time in bot parallel and non-parallel activities. However, it does present higher relative modality efficiency levels and, thus, improved multimodal synergy.

The groups may not perform faster or more efficiently using this combination, but they seem to know how to combine the input modalities better as to improve their synergy ratio. This could also be explained by the acquisition and homing time being better managed by the groups when using this sequence.

5.2.3.3 Summary

During this phase of the study, we intended to understand which combination of three input modalities could cause positive synergies and present the best results regarding effectiveness and efficiency in two types of compound tasks: with and without parallelism.

Regarding the number of error committed during interaction, children were the ones that presented more faults during the activities, and tangibles, mouse and gestures was the combination that cause those faults.

Also, the differences in each group became more apparent that in the previous experiment about bimodal interaction: (1) the differences in interaction were significant for the three age groups; (2) gestures, mouse and tangibles was the sequence to present the slowest and less accurate values in all groups; and (3) the multimodal synergy analyzed for the combinations were distinct depending on the age group, which may indicate that the group's characteristics may influence interaction, and not only the multimodal sequence used.

5.2.4 Synergies between four modes of interaction

In this stage of the case study, participants were asked to complete the activities resorting to four different interaction modalities required. The combinations are as follows:

- Tangibles, touch, computer mouse and gestures;
- Computer mouse, touch, tangibles and gestures;
- Gestures, computer mouse, touch, tangibles;
- Touch, computer mouse, gestures and tangibles;

5.2.4.1 Results

In this phase of our study, we aim at understanding which, if any, combination of four input modalities could benefit interaction. Contrarily to previous phases, we only developed one compound task, as in this case the evaluation of the non-parallel task did not make sense nor added any new results other than the ones already reviewed.

This section presents the results of the experimental tests regarding effectiveness (error rate and completion success), efficiency (completion time), user preference and multimodal synergy of the interface. We also assess the participants' attitudes during the tests and their satisfaction towards the different interfaces.

Effectiveness:

The same circumstances of error evaluation from previous tests apply to this phase of the study.

Figure 5.33 presents the total number of errors registered during the experiment. In total, there were 204 errors, being the majority committed by children (127 errors), regardless of the multimodal interface. This situation was constant in all tests: children made the majority of errors.

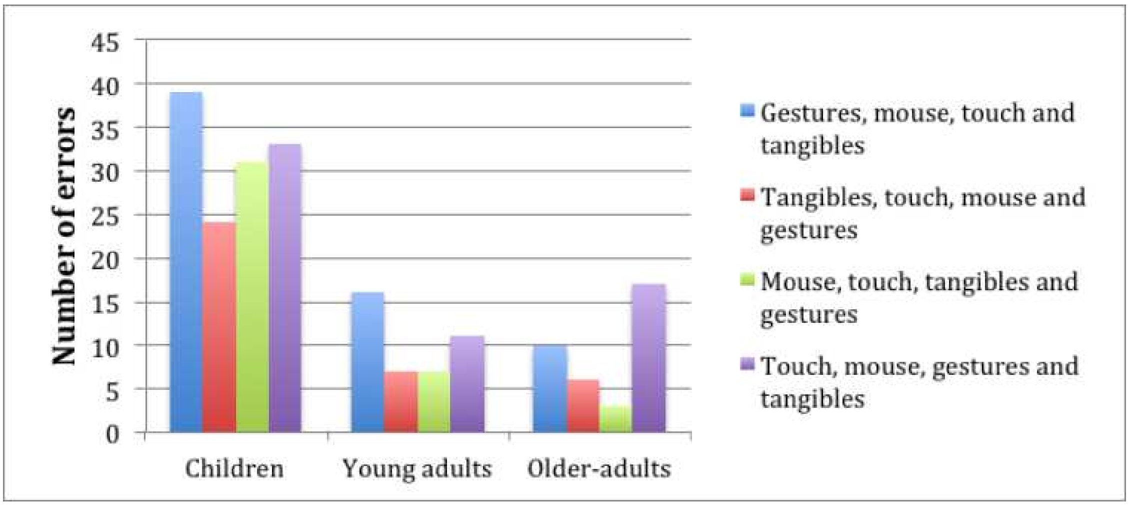


Figure 5.33 – Total number of errors presented in this level.

Overall, the interface to present more errors was using gestures, mouse, touch and tangibles (a total of 65 errors). On the other hand, the interface to present fewer errors was: tangibles, touch, mouse and gestures (37 errors).

Also, we recorded the position of the collisions in the labyrinth caused by the three age groups using the various input modalities during the task (Figure 5.34).

Efficiency:

We analyzed the completion time of the tasks with each combination within each group of participants and also between them. The completion time of each task was measured between the first action performed by the user (as soon as the cat was freed from the cage) and the successful insertion of the code, minus the inactivity

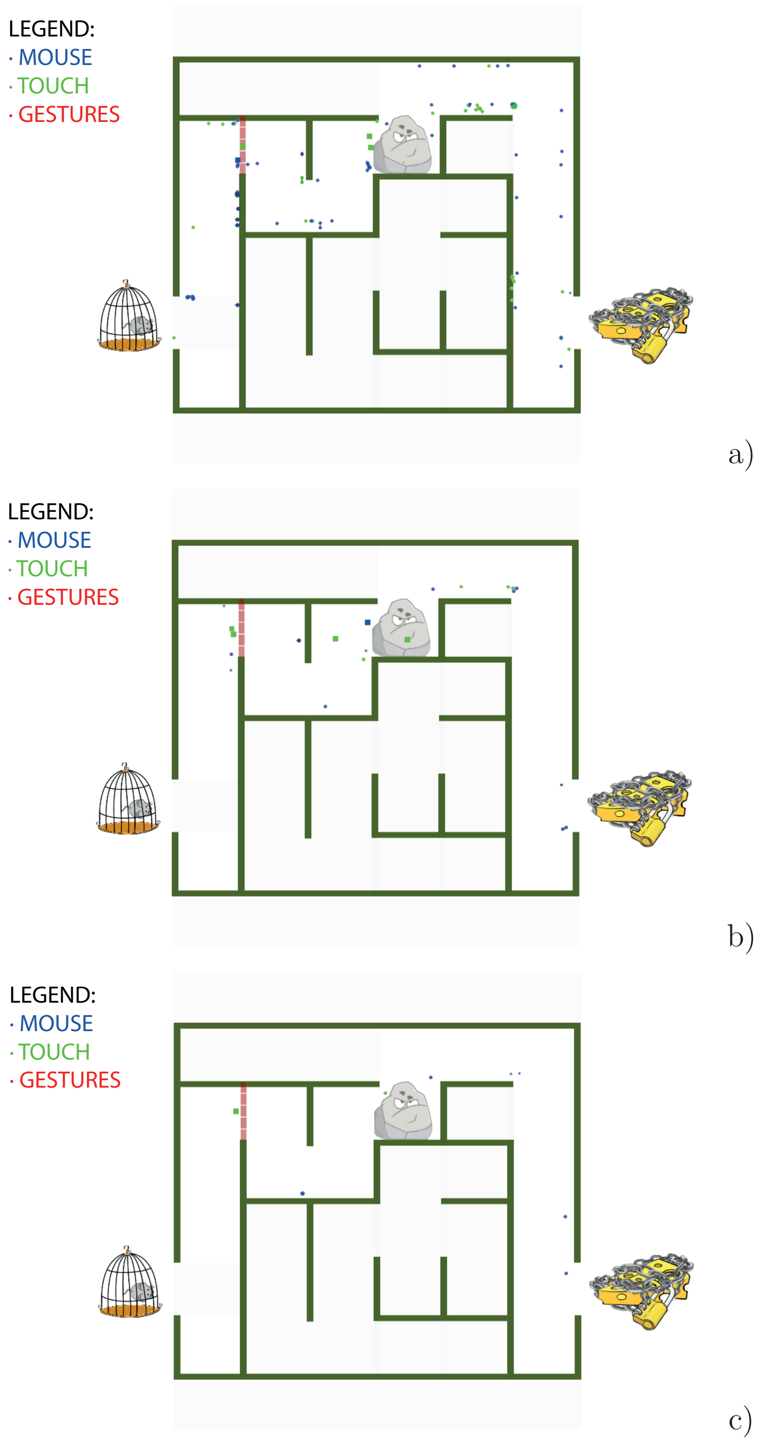


Figure 5.34 – Position of the errors committed in the labyrinth using the various combinations: a) Children; b) Young adults; c) Older-adults.

period.

Figure 5.35 presents the mean completion time of each age group registered during the experiment when using the different interfaces. The combination that registered the fastest results among the three age groups was touch, mouse, gestures and tangibles: children (37.83 s); young adults (26.26 s); older-adults (35.76 s). On the other hand, the combination with worse results was gestures, mouse, touch and tangibles in the group of children (46.21 s); and both groups of young and older-adults completed the task using more time with the combination mouse, touch, tangibles and gestures, with 32.09 s and 41.93 s, respectively.

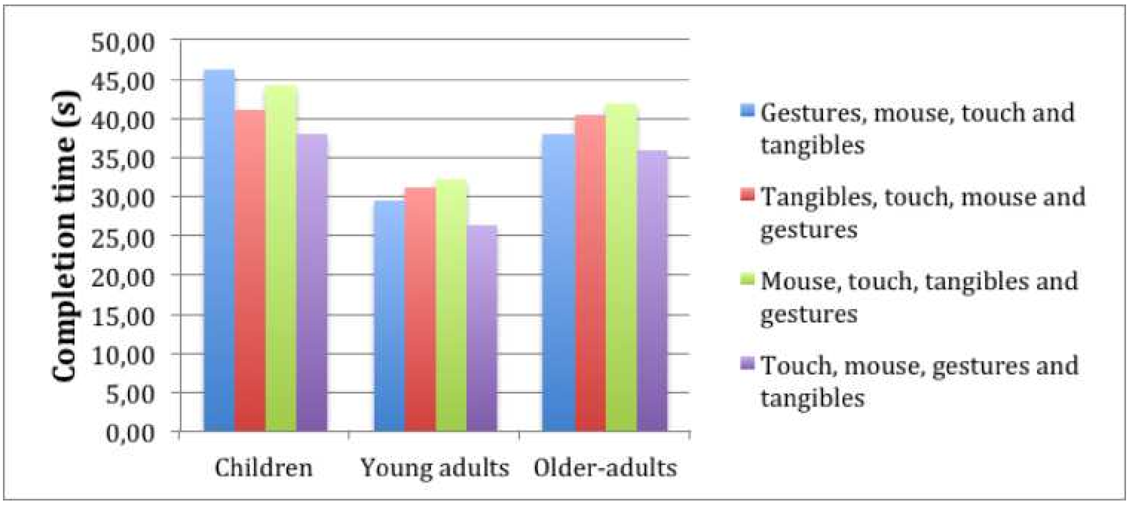


Figure 5.35 – Mean time taken to complete the task (in seconds).

A Shapiro-Wilk test showed us most of the data were normally distributed. Thus, we proceeded with a one-way ANOVA analysis as it requires approximately normal data and is quite "robust" to violations of normality. There were significant differences in all groups regarding the various combinations: gestures, mouse, touch and tangibles ($F(2{,}57) = 13.007$, $p = .000$); tangibles, touch, mouse and gestures ($F(2{,}57) = 7.896$, $p = .001$); mouse, touch, tangibles and gestures ($F(2{,}57) = 4.887$, $p = .011$); and touch, mouse, gestures and tangibles ($F(2{,}57) = 10.305$, $p = .000$).

Regarding the first sequence, all groups displayed statistically significant differences in efficiency: children versus young adults ($p = .000$); children versus older-adults (p

= .037); and young versus older-adults (p = .034). On the other hand, the sequence tangibles, touch, mouse and gestures also displayed differences in efficiency in the groups of children versus young adults (p = .002), and young versus older-adults (p = .005), but not between children and older-adults (p = .966).

Considering the other two combinations, the same pattern in values occurs: with the sequence mouse, touch, tangibles and gestures, children versus young adults display significant differences (p = .013), as well as young versus older-adults (p = .049), but not between children and older-adults (p = .864); with the sequence touch, mouse, gestures and tangibles, children versus young adults display significant differences (p = .000), as well as young versus older-adults (p = .003), but not between children and older-adults (p = .727).

Additionally to comparing efficiency between each group regarding the different interfaces, we also analyzed efficiency within each group by resorting to a within-participant repeated-measures design. Regarding the group of the children, a repeated measures ANOVA with a Greenhouse-Geisser correction determined there was a significant difference between the sequences (F(2.334, 44.338) = 4.070, p <.019). Post-hoc pairwise comparisons tests using the Bonferroni correction showed that, for this group, the fastest combination (37.83 s) was not significantly better than the other sequences: comparing with gestures, mouse, touch and tangibles (p = .060); tangibles, touch, mouse and gestures (p = .296); or mouse, touch, tangibles and gestures (p = .072).

A repeated measures ANOVA with a Greenhouse-Geisser correction determined that the group of young adults did not presented statistically significantly different results regarding the combinations (F(2.334, 44.347) = 3.675, p <.070). Indeed, Post-hoc pairwise comparisons tests using the Bonferroni correction confirmed that the mean efficiency time was not significant between any sequence.

Just the same as the young adults, older-adults do not present a statistically significant difference between the mean times using the combinations (F(2.265, 43.038) = 2.203, p <.117).

Indeed, with a four-interface multimodal combination, young and older-adults do not display significant differences in performance and, as such, they are indifferent towards which sequence they use.

We also found it important to measure inactivity periods for these combinations (Figure 5.36). As in the previous phase, all the groups appeared to be constant in terms of the periods of inertia, although the children may present slightly higher values. Moreover, the combination to display longer inactivity times if the one that supports tangibles, touch, mouse and gestures, while the least amount of inertia was supported by gestures, mouse, touch and tangibles, for the young and older-adults; and touch, mouse, gestures and tangibles regarding children.

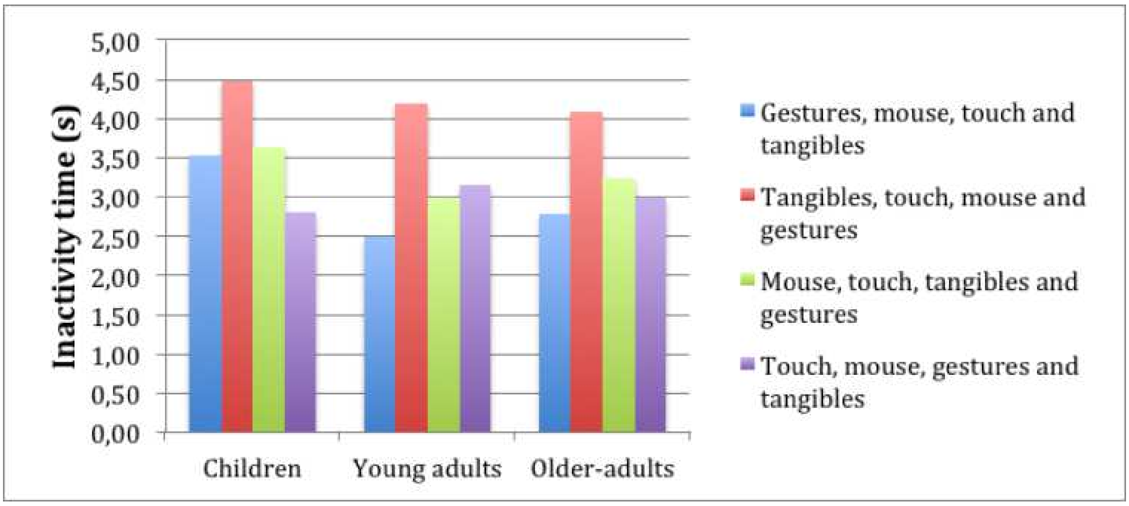


Figure 5.36 – Time of inactivity recorded during the task by each sequence (in seconds).

Multimodal synergy:

Aside from the previous analyzed objective metrics such as speed, number of errors and task completion that are usually computed, we also evaluate the combination of the input modalities according to their relative modality efficiency and thus multimodal synergy.

Figure 5.37 presents the percentage of multimodal synergy of the several combinations regarding each age group. As shown in the chart, there are synergy fluctuations regarding the different combinations between the three groups. However, all values are negative, thus showing that the combination of four interaction modalities may be excessive for all groups of users.

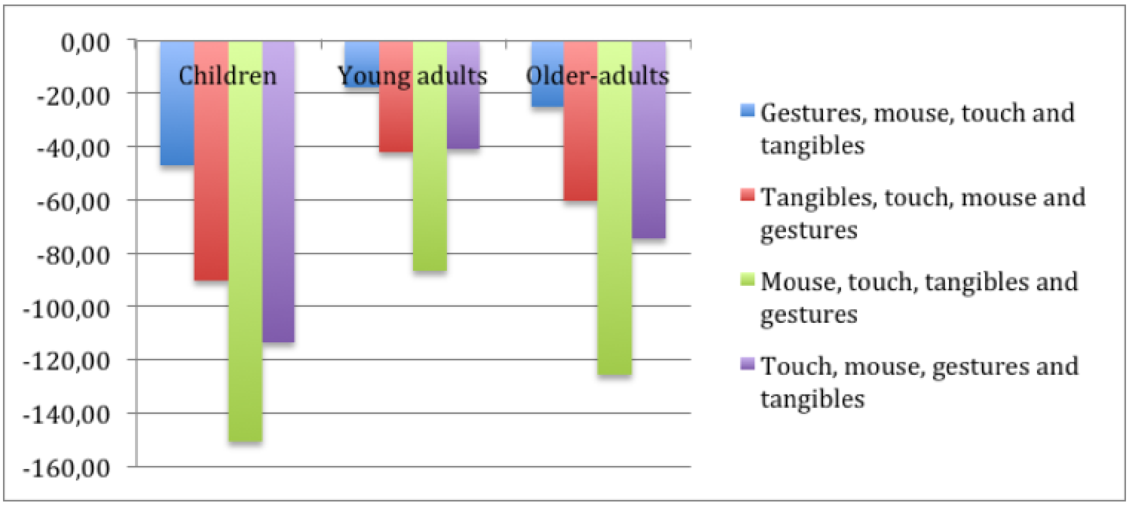


Figure 5.37 – Multimodal synergy (%) of the four multimodal interfaces.

The lowest values relate to the combination of mouse, touch, tangibles and gestures, in all groups. On the contrary, gestures, mouse, touch and tangibles managed to have a less negative value.

Participants' attitudes:

We present the major difficulties sensed by the participants and discuss their most prominent attitudes.

Indeed, the observation of the participants' interaction validates the results already presented and discussed above. Every participant, regardless of the group, would get confused during the trials. We could notice that they did not stop playing the game, but rather gesticulated a lot more and almost every participant would begin to talk at some point, as if they were explaining the steps to themselves in order not to get lost. This method may have helped maintaining the rhythm of the interaction, because we can see in the efficiency results that they did not scatter when compared to the other simpler combinations.

A sign of confusion was also the participants' reactions regarding the tangibles, mainly the children. Indeed, when picking up a tangible to execute the action, a lot of participants did not put it down afterwards to the same position as where they got it. They would begin to mess the order of the tangibles and then they would become more confused.

5.2.4.2 Discussion

Regarding error evaluation, overall the combination to present more errors was gestures, mouse, touch and tangibles, whereas the interface to present fewer errors was: tangibles, touch, mouse and gestures. Comparing the groups of young and older-adults, the number of errors was rather consistent and quite similar.

On a different matter, when it comes to efficiency the values of mean completion time using four input modalities do not present so many fluctuations as in the other combinations of two and three interaction paradigms. Also, children and older-adults display similar results in terms of efficiency, which can imply that these participants have similar interaction patterns when dealing with these types of multimodal interfaces.

Considering the inactivity time recorded, the time gaps are similar to the previous studies. That being said, we believe the inactivity time may not increase with the addition of more input modalities to the same interface.

Also, considering the multimodal synergy that occurs with the combination of four input modalities, we may imply that users tend to: (a) become overwhelmed with the different input options; (b) there are no positive synergies generated during interaction and thus no input modality helps the other; (c) it seems that the order of the modalities, i.e. the elemental task they perform does not influence synergies at all.

5.2.4.3 Summary

This experiment was intended to understand if four modes of interaction in a multimodal interface could cause positive synergies and improve the efficiency, effectiveness and preference of the participants.

Considering the third level of the game, we noticed that the synergies of the combinations become negative in all age groups, which may indicate that the participants cannot handle that many input devices in one same task.

Indeed, we may come to realize that: (1) four input modalities augment the discrepancies of possible fusion levels; (2) the order of the sequence does not influence interaction, as all combinations present negative synergetic values; (3) every participant, regardless of the group, would get confused during the trials with so many modality options.

5.3 Concluding remarks

The goal of this chapter was to understand if there could be differences in effectiveness, efficiency and preference regarding three distinct age groups using multimodal interfaces. Also, one other goal was to identify synergies that could come from combining various modes of interaction, specifically two, three and four input modalities. Out study was particularly motivated by the absence of scientific data that systematically evaluates which combination of input modes could be the most appropriate for compound tasks and users with age-related differences.

Indeed, we have come to realize that: (1) compound tasks may augment the discrepancies of each modality considering the different age groups; (2) the type of interface that is used (unimodal or multimodal) can influence efficiency in compound tasks; (3) the differences in each group became more apparent in a three-interface combination than in a bimodal interface; (4) and the multimodal synergy analyzed for the various combinations were distinct depending on the age group, which may indicate that the group's characteristics may influence interaction, and not only the interaction modality used;

Thus, according to the results presented, it appears to be evident that the creation of distinctive user profiles regarding interaction with different interfaces could be important. Furthermore, these indications have highlighted the need for further investigation to understand key issues related to input modalities and user preferences.

I

6 Proposal of a multimodal interaction model regarding different user profiles

In this Chapter, we propose a multimodal interaction model for compound tasks regarding different user age groups: children, young adults and older-adults.

Throughout our study we came to realize that the use of different interfaces that better accommodate the users' needs or individualities is important, that is, the creation of distinctive user profiles regarding interaction with different interfaces. Thus, several guidelines are recommended accordingly for the three age groups studied in a controlled environment. Our intention is to convey information that helps the reader get an understanding of the most adequate interfaces that should be used in certain contexts.

Alongside the guidelines suggested, we provide examples that should help the reader understand the different contexts that can be considered for developing interactive systems. As such, the specifications given intend to aid the designer/developer in creating more adequate interfaces that can harness the interaction potential of the three age groups considered for this research.

Indeed, there are some considerations we find important to relate forthwith, considering the age groups behavior, and the types of tasks.

First, the age groups revealed certain tendencies throughout the study when using the various input modalities, whether considering them for elemental or compound tasks. The participants kept a consistent opinion regarding the different input modalities during the different phases of the study, whether about their preferences or even difficulties. This surely indicates that the three age groups have individual opinions and display a consistent behavior about the input modalities. The participants maintained their opinions and showed not to be volatile before different tasks, sharing a very strong judgment about which input modalities they preferred. Of course, this does not indicate which input modalities are better suited for them, as other key aspects like speed, accuracy or performance indices must be taken into account, but reveals that when considering tasks that do not require great control over the quantitative values of participants, their preference should be taken into account. Please note, however, that the participants' opinion could be influenced by their familiarization with their daily input modalities and hence some training could help them favor other interaction modes as well.

On a different matter, regarding the types of tasks studied and the four input modalities contemplated, we must acknowledge that a multimodal interaction only makes sense when considering compound tasks. Indeed, elemental tasks require a much basic and direct feedback from the participants and, as such, a multimodal system would not provide a gain in performance. In this context, if we consider low-level types of interaction tasks, i.e. primitive actions as hierarchies of compound sub-tasks, they only require single interactions. A unimodal scenario in that situation is more desirable and beneficial for the user. However, in a compound task with sub-tasks and different actions that can be performed with or without parallelism, a multimodal system does influence interaction.

It is, indeed, important to acknowledge that elemental actions are a part of a compound activity that benefits from multimodal interaction. Therefore, we find it essential to provide an overview of the results gathered throughout the study regarding elemental tasks for the three age groups.

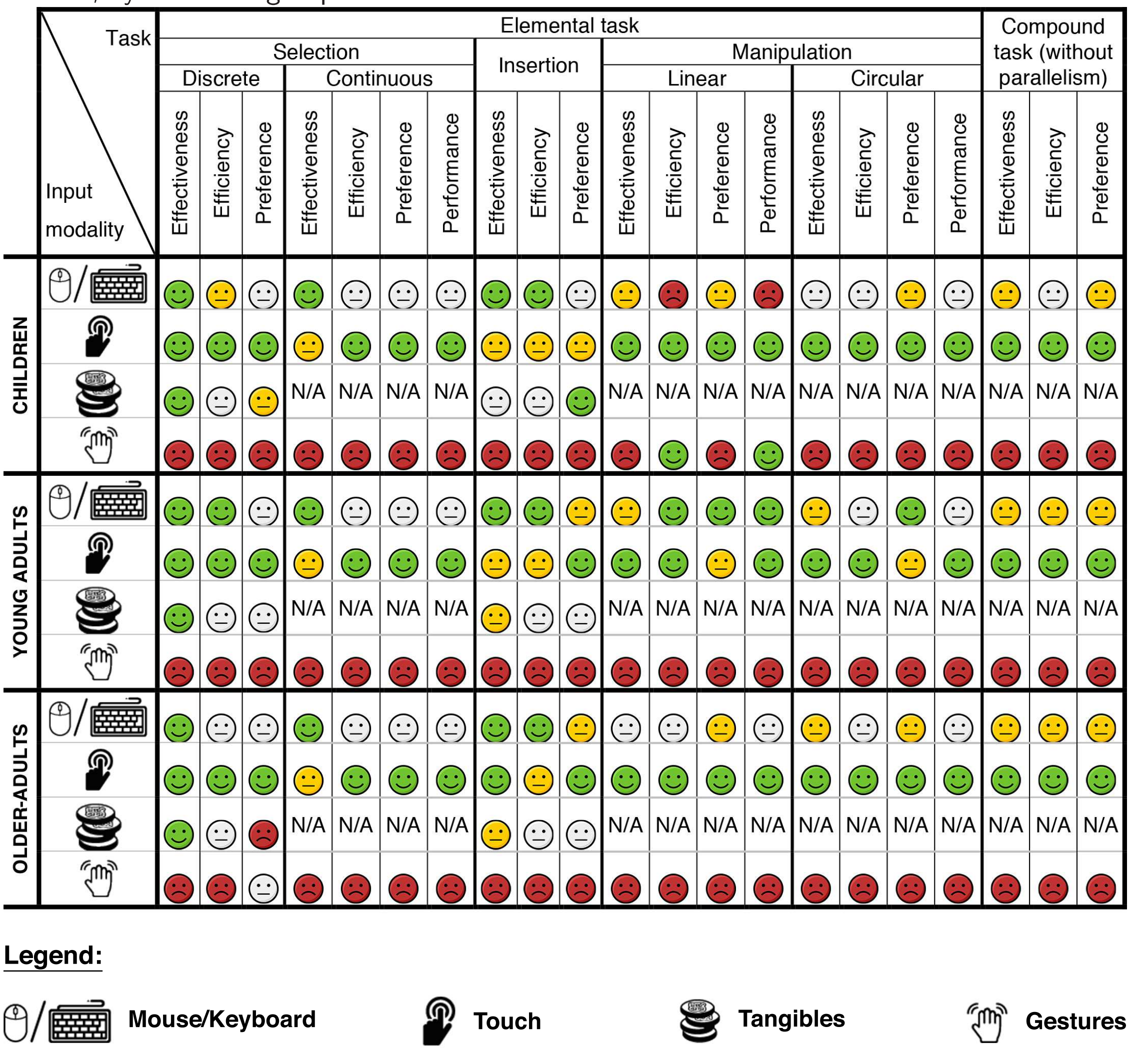
Table 6.1 shows an overview of the qualitative and quantitative results gathered throughout the study on elemental tasks, and also provides an overall look at the unimodal interfaces when used in compound tasks. In the chart, we can see: the best modality choices for performing the different tasks studied, represented by a green smiley face; the second best choices that, in spite of having worse results, were not statistically significantly different from the best modality, represented by a yellow neutral face; the modalities that were not considered the best nor worst in the evaluation, represented by a grey neutral face; and finally the worst modality choices for the respective tasks, represented by a red sad face. Due to the high amount of data retrieved from the various evaluation phases, we opted to present the information in this format, as this assessment benefits the reader to have a wider and more comprehensive perspective of the study.

The presented guidelines are organized into various categories, depending on the applicability of the task: effectiveness, efficiency, user preference and performance. Moreover, we make an overall analysis of the results with a more global perspective in some final considerations. Please note that these results were extensively presented and discussed in Chapter 4 and Chapter 5 of this document.

First and foremost, we can observe that the tangible interface is not applicable in various situations. This is due to this type of input modality not being a pointing device, as thus cannot be used to point to objects on the screen. Whenever the tasks required a direct action that influences a particular object on the screen relating to positioning, tangibles cannot be considered, that is, for activities that require direct manipulation and simultaneous steering capacities, this input modality may not be applicable. We are aware that other strategies exist in order to not set aside this input modality, but we do not consider them in our study. Note that this subject and other limitations of our study will be discussed later, in the next Chapter.

Thus far, we are already recognizing a limitation of tangibles: they cannot be widely used as the other input modalities, hence the need to accurately define the type of task being performed when developing an interactive system with support to a tangible interface, and its requirements. Indeed, when we refer to elemental tasks, we need to consider the type of basic action required. Hence, tangibles are not a good choice for manipulating objects that require pointing, that is, the objects that can be manipulated must be already configured in the system with the respective tangible

Table 6.1 – Overview of results gathered on elemental and compound tasks using a unimodal interface, by the three groups.



piece, or else the users' performance may be undermined. Hence, working with tangibles can be limiting when using them as the primary modality for interaction. However, in compound tasks, when considered as a complementary input to be used alongside other modalities, this can be a good tactic, as will be explored ahead in the Chapter. Regarding the other input modalities, they can all be considered for the various elemental tasks, whether they involve selection, insertion or manipulation.

We describe below how the different modalities scored in terms of effectiveness, efficiency, preference and, when applicable, performance during the different elemental tasks. This overview is important to demonstrate variations in the three age groups, and also understand with which input modalities the disparities occurred. Indeed, when developing interactive systems, it is important to follow the main requisites imposed and thus select which modality could help the user when it comes to the various contexts: if the system values accuracy, or rather speed, different input modalities will be used for the three age groups, according to the task at hand.

Effectiveness:

When the interactive system values mostly the effectiveness in performance, and whether the user completes the tasks with the least errors, the input modality that should not be used to complete any type of task is gestures, regardless of the age group in question.

Indeed, gestures presented the worst results of all modalities when it comes to errors made during interaction. This interface shows that even in different age groups, the outcome is the same: gestures cause more errors than any other input modality.

When performing selection tasks, for instance, we can see that there is a difference on the best modality to be used for discrete versus continuous tasks. This difference indicates that the change in the type of task itself influences the users' effectiveness. Continuous tasks require more accuracy than discrete tasks, as the user needs to finish one selection and move on to the next continuously and without breaks. This implies that the user needs to be more precise in selecting the object and does not have the same amount of time as he/she would have in discrete tasks. Indeed, for these discrete selection tasks, all the input modalities behave alike and results showed that neither caused the users to make mistakes during interaction. On the other hand, continuous selection tasks did not display the same results: the age groups made fewer errors using the computer mouse, followed by touch as the second best input for this task. This reveals: (1) the computer mouse is a good option for both discrete and continuous tasks, but touch should be the second choice for continuous tasks; (2) gestures should not be considered for selection tasks at all. Also, even though the type of selection did make a difference in effectiveness considering the various input modalities, the age groups maintained the same behavior, as the three groups presented similar conclusions. This indicates that, for basic selection tasks, age does not influence interaction.

Analyzing the results for insertion tasks, the effectiveness of the three age groups is similar, although we can see that the soft-keyboard is considered the second best choice for the groups of children and young adults, and only the older-adults showed no differences whatsoever in terms of the number of errors registered. This indicates: (1) older-adults can use the soft- or physical keyboard as a first choice; (2) children and young adults should use the physical keyboard as the first choice, followed by touch if the first is not available; (3) young and older-adults can consider tangibles as a second choice, since the difference in the number of errors was not statistically significant between the physical keyboard; and (4) children should only consider using tangibles if the two best options are not offered, as they make more errors using the tangible pieces. Again, gestures should not be considered for this task in any age group.

For manipulation tasks, the best options indeed vary according to age. Thus it is important to understand which inputs to use and which ones to avoid according to the age group. First, the type of manipulation being performed (linear or circular) proves to influence the users' effectiveness. Hence, for linear or circular tasks we highlight: (1) touch is the best option for all age groups; (2) gestures remain the worst, causing more errors. Specifically regarding linear tasks, the computer mouse can be considered a second choice for children and young adults, but not for older-adults, that made significantly more errors with this input modality, and thus should also be avoided in this age group. On the contrary, for circular tasks, the computer mouse can be considered a second choice for young and older-adults, but not for children that recorded significantly more errors using this input modality.

Considering effectiveness in compound tasks, the age groups follow the same pattern: users should interact with the system via touch, use the computer mouse only if the first option is not possible, and definitely avoid gestures.

Efficiency:

Next, we analyze the best input modality choice for interactive systems that mostly value efficiency, where it is important for the user to finish the task as fast as possible regardless of mistakes that can occur.

For all three age groups performing selection tasks, touch proves to be the fastest choice and gestures the one to avoid. For discrete selection tasks, tangibles are not a good choice either and should not be considered for any age group. Also, the computer mouse should be used with caution, as it presents different results by the groups: (1) children should not use the computer mouse for discrete, let alone continuous tasks; (2) young adults may use the computer mouse with the same confidence as touch for discrete tasks, but not for continuous, as it is significantly less efficient than touch; (3) older-adults should only be required to use the computer mouse if there is no possibility to resort to touch, either for discrete or continuous tasks.

Curiously, when considering insertion tasks, the interactive system can follow the same guidelines for the three age groups if speed is the main requisite: the physical keyboard should be the required input, touch should just be considered as the second choice, tangibles are significantly slower to use and thus should not be provided, and gestures should definitely not be considered for interaction.

When it comes to manipulation activities, linear actions should be carefully planned for the three age groups. Indeed, even tough systems are mostly discouraged to provide gestural interaction, when children perform linear paths this situation reverses:

gestures become the first choice to be considered for interactive systems that value linear movements, along with touch, and the computer mouse should be avoided. For young and older-adults, though, gestures are still the worst choice for interactive systems. Also, young adults are able to use the computer mouse with the same efficiency as touch, but older-adults should use touch and avoid the computer mouse. On the other hand, for performing circular steering movements, the system can give the same relevance towards the input modalities for all the groups: choose touch as the first option; if this is not possible, provide the use of the computer mouse; and the last resort should be gestures.

Lastly, if we analyze overall compound tasks, when the system must emphasize speed we recommend: (1) do not provide gestural interaction; (2) favor the use of touch; (3) consider the use of the computer mouse by young and older-adults as a second choice, but do not enforce children to use this modality unless touch is not available

Preference:

When the interactive system being developed has the intention of providing the input modalities the users prefer, that is, value their preference over speed or accuracy, the following guidelines should be taken into account.

Regardless of the age group in question, for selection tasks (discrete or continuous) the users prefer to use touch. If this input modality is not available, the second choice to consider if the system is being built for children should be tangibles (for discrete selections). For this group, the computer mouse is not advised, and gestures should be avoided. For young adults, both the computer mouse or tangibles (where applicable) can be considered, but are not advised, and gestures should be also avoided. On the other hand, older-adults prefer using the computer mouse or gestures for discrete selections, and instead tangibles should be avoided, but for continuous tasks they also dislike gestures.

When considering insertion tasks, young and older-adults share opinions: touch is the favorite and should be highlighted, the physical keyboard can be considered a second choice, tangibles are not advised, and gestures should be avoided. However, even though children also dislike gestures for this type of task, they prefer using the tangible pieces, being these considered the first choice for the interactive systems when user preference is the higher purpose. Next, children choose the soft-keyboard, i.e. touch, and the actual physical keyboard. Indeed, when developing interactive systems that privilege the children's preference, for example, for applications that depend on the interest of children, tangibles are the best option to maintain their interest and enthusiasm. Finally, gestures is the least liked input modality.

For manipulation purposes, both children and older-adults share their preference: touch is the favorite, the computer mouse should be used only as a second choice, and gestures is the least liked input modality and is thus not recommended. The group of young adults shares their lack of interest in gestures, but prefer to use the computer mouse over touch.

Finally, when performing overall compound tasks, all the groups agree: (1) touch should be given privilege; (2) the computer mouse can be seen as a second choice; (3) gestures should be avoided.

Performance:

Regarding the index of performance analyzes in previous the Chapters, we have seen that both speed and accuracy influence results. If the interactive system being developed has the purpose of harnessing the performance potential of the users, different input modalities should be pondered.

When performing continuous selection activities, all the groups behave similarly: (1) touch is the best option for interaction; (2) the computer mouse should only be considered in the impossibility to provide touch; (3) gestures should be avoided.

Attention should be given to manipulation activities if performance is of interest. Indeed, the three age groups show contrasting options. For linear movements, children show a better index of performance using touch or gestures. This is very important, as the gestural interface tends to be the most avoided one for the other activities. On the other hand, the computer mouse should not be provided for interaction when the interactive system is being built for children. Contrarily to this group, young adults show better performance rates with touch, and thus this input modality should be prioritized, followed by the computer mouse, that does not show significant differences in performance. Gestures should be avoided if possible. For the group of older-adults, touch is the best option, the computer mouse is not recommended and gestures should be avoided. When considering circular steering movements, the groups' performances are more consistent: (1) touch should be privileged; (2) the computer mouse can be used but as a second choice; (3) gestures should be avoided.

Overall:

As we can see in Table 1, the input modalities have different results for the three age groups, and do not demonstrate the same aptitude for performing basic actions. However, we can see a pattern in the chart: the input modality with the majority of green smiley faces is touch, whereas the one with the majority of red sad faces is gestures. This is true for the three age groups, and this means that these two input modalities are consistently the best and worst choices for the majority of the tasks in an interactive system.

6.1 Multimodal interaction model

Thus far, we have come to realize that the creation of distinctive user profiles regarding interaction with different interfaces could, indeed, be important.

Now that we have acknowledge the age group's differences regarding elemental activities and the compound tasks that can be performed resorting to a unimodal interface, it becomes important to provide guidelines that can be followed when developing multimodal interfaces. The previous analysis was important to provide a baseline for the multimodal study and must be regarded as a stepping-stone for the combinations chosen and studied during this research.

Ultimately, our goal was to understand at which point the redundancy of commands, and constant addition of new input modalities, start to saturate and harm the user's overall performance. As such, we believe it is important to analyze side-by-side the combinations chosen for this study, as we intend to propose guidelines about the best possible ratio between quantities of input modalities provided in a multimodal system versus the positive benefits for the age groups according to the system's various purposes and contexts.

First, we present a series of observations and guidelines that assure the best fit for the three age groups regarding the number of input modalities provided in a multimodal interface. Secondly, we explore the guidelines and deepen the model by providing an overview of each combination and their influence on the three age groups. This is important for the reader to understand the best options that should be provided in a specific multimodal system, for each type of combination: two, three or four input modalities.

Table 6.2 presents an overview of the qualitative and quantitative results gathered throughout the study on compound tasks regarding the combination of two, three or four interfaces. The same strategy is used to represent the best modality combinations for performing overall compound tasks: happy, neutral and sad faces with the corresponding coloring. Additionally, note that these results were extensively presented and discussed in Chapter 5 of this document.

In the chart we analyze compound tasks with and without resorting to parallelism to complete the activities. Note that parallelism in multimodal interaction may cause differences in the users' performance, as different actions are being executed simultaneously.

Our presented guidelines are organized into four categories: effectiveness, efficiency, synergy and user preference. Note that in this situation we present one more category, synergy, which was not presented in the previous guidelines. Please remember that this factor is about the measure of the quality of modality fusion in a multimodal system, and relates to the concept that a synergistic multimodal interface is better than the sum of its parts, that is, unimodal interfaces. However, we alert for

Table 6.2 – Overview of results gathered on compound tasks by the three age groups using three types of combinations.

	Task / Combinations	Compound task: Without parallelism – Effectiveness	Without parallelism – Efficiency	Without parallelism – Synergy	Without parallelism – Preference	With parallelism – Effectiveness	With parallelism – Efficiency	With parallelism – Synergy	With parallelism – Preference
CHILDREN	Combination of two interfaces	🙂	🙂	🙂	🙂	🙂	🙂	🙂	🙂
	Combination of three interfaces	🙁	🙂	🙁	🙁	🙂	🙂	😐	🙁
	Combination of four interfaces	N/A	N/A	N/A	🙁	🙁	🙁	🙁	🙁
YOUNG ADULTS	Combination of two interfaces	🙁	🙂	🙂	🙂	🙂	🙂	😑	🙂
	Combination of three interfaces	🙂	🙂	🙁	😑	🙂	🙂	🙂	😑
	Combination of four interfaces	N/A	N/A	N/A	🙁	🙁	🙁	🙁	🙁
OLDER-ADULTS	Combination of two interfaces	🙁	🙂	🙂	🙁	🙂	🙂	🙂	🙁
	Combination of three interfaces	🙂	🙂	🙁	🙁	😐	🙂	😑	🙁
	Combination of four interfaces	N/A	N/A	N/A	🙁	🙁	🙁	🙁	🙁

the fact that a better synergy between input modalities may not necessarily convert to better effectiveness or efficiency. Even tough two modalities can improve the relative modality efficiency when brought together, it may not be sufficient to provide better results and overcome other combinations that, even though may present lower synergy rates, still manage to continue being the best choice in terms of speed or success levels. Moreover, we make an overall analysis of the results with a more global perspective in some final considerations.

Similarly to the elemental tasks analysis, this chart demonstrates at times more than one first option to be considered when taking into account the desired feature. This is due to the results being equivalent for the various input modalities and thus cannot be ranked on a hierarchical basis.

Effectiveness:

When the multimodal interactive system values mostly the effectiveness in compound tasks without parallelism, that is, the number of error committed while executing actions that are not simultaneous, it should privilege two different combinations, regarding the age groups: children are more effective when using a bimodal interface, whereas young and older adults are more successful using an interface that combines three input modalities.

Indeed, young and older-adults executed fewer errors during interaction with three input options for compound tasks. This indicates that they are more effective when using three interaction modalities and thus commit fewer errors.

For tasks that require synchronous actions, a new pattern emerges considering the group of children and young adults: (1) they display similar results, and so the combination of three interfaces should be used by this age groups to prevent more errors; (2) the bimodal interface can also be considered for these two groups, as it does not display significant differences; (3) four interfaces prove to be overwhelming and thus should not be considered at all by these two groups. On the other hand, this model changes for older-adults, as they perform more accurately using the bimodal interface, closely followed by three combinations, and finally the interface with four input modalities should not be considered.

Efficiency:

If the multimodal system being developed has the purpose of exploiting the users' efficiency and values speed in completing compound tasks without parallelism, the three age groups share the same model: they are more efficient using combination of three input modalities, closely followed by the bimodal paradigm.

In this situation we cannot consider that there is a combination that should be avoided, as the groups did not display significant differences in speed, and so we can acknowledge that both types of combinations can be presented.

Regarding the compound tasks with parallelism, we also encounter similar results for the three groups: the combinations of two and three input modalities can be considered as the first choice when building the multimodal system, whereas the combination of four interaction modalities should not be considered.

Accordingly, note that for compound tasks, having simultaneous actions or not does not influence efficiency when comparing the three age groups. Indeed, the same guidelines can be followed, since the groups display analogous efficiency results.

Synergy:

Next, we analyze the best combinations for multimodal systems regarding their declared synergy levels. Our intention is to provide guidelines that help develop interfaces that are more synergetic, that is, manage to integrate more efficiently the different input modalities available in order to create positive cooperation between modalities.

When it comes to compound tasks without parallel actions, the synergies that come from combining the various input modalities by the three age groups are similar. Thus, we recommend that, when building multimodal systems that value the positive integration between input modalities, bimodal interfaces should be prioritized, since the synergies that come from combining three interfaces are negative. This indicates that combining more than two interfaces for compound tasks without parallel actions, leads to negative synergies, and thus the input modalities cannot provide cooperative work between each other. In other words, regardless of the age group, the input modalities hinder each other and do not function together to take advantage of each other's benefits nor improve each other's limitations, regardless of the age group.

Considering compound tasks with parallelism, the recommended guidelines about multimodal synergy levels for each age group are divergent, except for the four input modalities combination, which should not be used under any circumstances by any age group.

On one hand, bimodal interfaces should be prioritized when the system is being built for children, as they can exploit the two modalities' advantages and set forth positive synergies. The combination of three input modalities can also be considered, but four input options prove to be overwhelming and damage the integration. On the other hand, a multimodal system built to harness synergies for young adults should not follow these same guidelines. In this situation, the combination used should be with three input modalities as the best choice, and a bimodal interface should only be considered if three input modalities are not available for the same system. Also, the combination of four input options should not be provided. In turn, older adults display the opposite result: the bimodal interface is the best option and three input modalities should only be offered as a second choice. Again, four input modalities should be avoided at all times.

As we can see, as more input modalities are brought together and combined into a more complex multimodal interface, the existing synergies that could come from this grouping begin to fade. Indeed, the more input modalities available, the more negative the synergies become, especially considering the group of children and older-adults.

Preference:

In terms of user preference, it is indifferent whether the compound tasks are performed with parallel actions or not. The three age groups maintain identical opinions about both situations: (1) children prefer to use a bimodal interface and have a strong negative opinion about the other combinations, which means that more than two input modalities are not advised if the system intends to follow the users' preference; (2) young adults also prefer a bimodal interactive system, having less favoritism about a three-interface combination and even less about four; (3) older-adults firmly expressed their unwillingness about using more than one single input modality. As such, when the system developed is built on the user preferences alone, it should not use multimodal interaction at all for this group, and instead prioritize a unimodal interface. This guideline is very important, as it is the fiercest regarding the group of older-adults when considering multimodal interaction and their advantages. Aside from the benefits that multimodal systems can provide, older-adults show that they are reluctant to use them.

Overall:

It is important to acknowledge a wider perspective of the three types of combinations proposed. Even tough the amount of input modalities that should be combined are not the same for all the groups, we can assume a broaden awareness that considering all the factor, generally children and older-adults work best using a bimodal interface, whereas young adults can withstand three input modalities in the same multimodal system.

6.1.1 Guidelines for combining two interfaces

In this section, we present guidelines that may help the development of multimodal interfaces that are more aware of the users' characteristics. Here, we provide an overview of the results for a two-interface combination and its influence on the three age groups, in order to inform the reader about which bimodal interface is the best fit for each age group.

Table 6.3 presents an overview of the quantitative results gathered throughout the study on compound tasks regarding the combination of two interfaces. The same strategy is used to represent the best modality combinations for performing overall compound tasks: happy, neutral and sad faces with the corresponding coloring. Furthermore, remember that these results were extensively presented and discussed in Chapter 5 of this document.

Our guidelines are organized into three categories: effectiveness, efficiency and synergy; all analyzed with and without parallel activities. Moreover, it is important to mention that the worst fit regarding synergy levels indicate that that specific combination displayed negative synergies and, thus, should not be used by any means. Note that, contrarily to the above guidelines, where we compare the type of combinations side-by-side, we do not involve user preference in this assessment. As explained in the previous Chapter, this situation is due to the lack of feedback from the participants during the tests about which combination of two input modalities was their favorite. Remember that the participants were somewhat confused about the multimodal combinations and, thus, we do not have concrete opinions about their preference. We also make an overall analysis of the results with a more global perspective in some final considerations.

Effectiveness:

If the interface being developed provides two choices for interaction, the following guidelines may be helpful to understand which combinations are best suited for the three age groups.

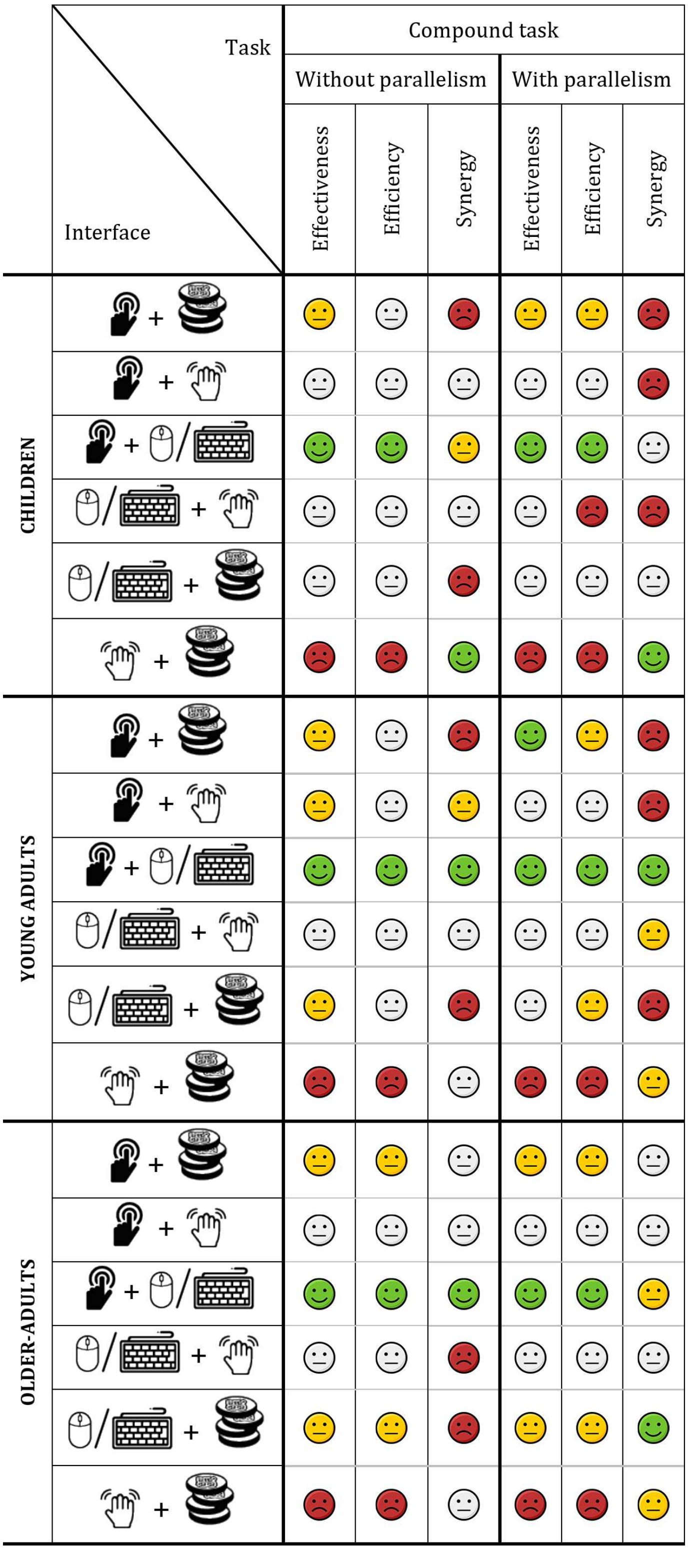
Provided that the bimodal interface must offer a smoother interaction without errors, the best combination for all the age groups is touch and mouse, whether the task has parallel activities or not. This combination leads back to fewer errors during interaction and is thus safer to adopt when the amount of errors is relevant. On the other hand, the combination to avoid at all times is gestures and tangibles, regardless of the age group. Indeed, this conclusion does make sense when we ponder that both of these input modalities cause difficulties in users if used together, because gestures alone can be daunting for some users, let alone combining them with physical pieces that have to be carefully picked.

The remaining combinations are also relevant for some groups, but not equally by all. Regarding parallel and non-parallel activities, children and older-adults may use touch and tangibles as a second best fit; mouse and tangibles may also be considered for older-adults. For young adults performing non-parallel activities, mouse and gestures are best to be avoided for interaction, and the other combinations may be considered as second best; and when performing parallel activities touch and tangibles help the group's effectiveness and thus can also be pointed out as a first choice. The other combinations not specified should be given neutral preference.

Efficiency:

When it comes down to efficiency and the tasks must be completed as fast as possible, the best combination is again touch and mouse, regardless of age group or task

Table 6.3 – Multimodal interaction guidelines for the three age groups using a two-interface combination.



parallelism. Furthermore, gestures and tangibles should not be considered for the multimodal system.

Older-adults are also efficient when using touch and tangibles, or mouse and tangibles for both types of parallelism in tasks. However, there are differences for the other two groups when performing non-parallel or parallel activities. For the first situation, all the other combinations not mentioned yet should be used with care, as they cannot be even considered second choices in interaction. But they are not the worst, either. Considering the latter situation, young adults reveal the same pattern as older-adults. Furthermore, mouse and gestures should also be entirely avoided; touch and tangibles may be considered a second best fit; and the other two remaining combinations are neutral.

Synergy:

Next, we provide guidelines that help develop interfaces that are more synergetic and create positive cooperation between modalities. In this regard, groups behave very differently.

First, the combinations touch and tangibles or mouse and tangibles should not be considered for interaction by children or young adults, as they display negative synergies working together, regardless of the type of task parallelism. The grouping of mouse and tangibles should also not be considered in a multimodal interface when it comes to young adults (in parallel or non-parallel tasks), and children (in non-parallel tasks). Still dealing with these two groups, parallel tasks should not support the combination of touch and gestures or mouse and gestures, for they too present negative synergies. Moreover, children should not be provided with the combination of mouse and gestures either. Regarding the execution of parallel activities, the group of older adults did not register any negative synergies in any combination.

On a different matter, the best fit for the three age groups is distinct. The multimodal interface should always favor the following combinations if its purpose is having higher synergy levels: gestures and tangibles with children; touch and mouse with young adults; mouse and tangibles with older-adults. Indeed, we can once again understand the need for creating distinct user profiles in multimodal interaction, as the age groups may have very distinct performances/behaviors towards the input modalities combinations.

Lastly, the second best fit for young adults is mouse and gestures, or gestures and tangibles; and the second best choice for older-adults is touch and mouse, or also gestures and tangibles, as the previous group.

The remaining combinations are neutral as should be provided with attention.

Overall:

Regarding a wider perspective of the combinations proposed, generally the three age groups have a tendency towards touch and mouse as the best multimodal scenario, and gestures and tangibles as the worst.

6.1.2 Guidelines for combining three interfaces

As in the previous section, we present guidelines that may help the development of multimodal interfaces that are more aware of the users' characteristics regarding multimodal interfaces that combine three input modalities. Our intention is to inform the reader about which combination is the best fit for each age group. Furthermore, remember that these results were extensively presented and discussed in Chapter 5 of this document.

We present an overview of quantitative results of the three-interface combinations in Table 6.4, following the same visual representation for the best, second best, neutral and worst fits. Similarly, our guidelines are organized into three categories: effectiveness, efficiency and synergy, for the exact same reasons as the previous analysis. We also make an overall analysis of the results with a more global perspective in some final considerations.

Take under consideration that, at this stage, the order in which the input modalities are used also influences end results. Remember that the acquisition and homing times are also of interest during multimodal interaction and can influence results. This is an indication that the input modalities by themselves are not the only influential factors during interaction, but also their order of use and how well they can merge with the previous or next modality being used in the combination.

Effectiveness:

Whenever the interface being developed provides three choices for interaction, the following guidelines may be helpful to understand which combinations are best suited for the three age groups.

Provided that the bimodal interface prizes the user effectiveness, the best combinations are divert for all three three groups. To ensure the least amount of errors is made, children should be provided with the combination of mouse, touch and gestures, or tangibles, touch and gestures, for non-parallel tasks. On the other hand, for parallel tasks, they should use tangibles, touch and mouse. The combinations to avoid at all times are gestures, mouse and tangibles (for parallel tasks), along with tangibles, mouse and gestures (for both types of tasks). Still considering parallel activities, children could also use mouse, touch and gestures as a second choice. The other combinations are neutral.

Young adults also favor two combinations for non-parallel tasks, namely mouse, touch and gestures, or tangibles, touch and mouse. In this types of tasks, gestures, mouse and tangibles are the worst fit. Contrarily, for parallel tasks, the worst fit is tangibles, mouse and gestures; but the best combination still is tangibles, touch and mouse, followed by mouse, touch and gestures as a second choice.

The worst combinations for older-adults are touch, mouse and tangible (for non-parallel tasks), or tangibles, mouse and gestures, regardless of the type of task. In turn, the best combination is tangibles, touch and mouse, followed by mouse, touch and gestures (for non-parallel tasks), and mouse, touch and gestures, followed by tangibles, touch and gestures (for parallel tasks). In this situation, keep in mind

Table 6.4 – Multimodal interaction guidelines for the three age groups using a three-interface combination.

that the best combinations are the same, but their relevance changes contingent on the type of task.

Efficiency:

Next, we analyze the best combinations for multimodal systems that mostly value efficiency, where it is important for the user to finish the task as fast as possible regardless of mistakes that can occur.

Indeed, mouse, touch and gestures prove to be the best fit for children when performing non-parallel tasks, whereas tangibles, touch and mouse are the best combination for parallel-tasks. Tangibles, touch and mouse can also be considered as second best in non-parallel tasks. In turn, young and older-adults share the best combination, regardless of the type of task: touch, mouse and tangibles. They can also consider mouse, touch and gestures as a second choice, but only in non-parallel tasks. With regard to parallel tasks, tangibles, touch and mouse can also be considered as a second best for the group of young adults.

On the other hand, for both types of tasks, gestures, mouse and tangibles are the worst fit for all age groups, and so that combination should be avoided at all times. Additionally, tangibles, mouse and gestures should also not be provided for children or older-adults. The other combinations are neutral.

Synergy:

Next, we analyze the best combinations for multimodal systems regarding their synergy levels.

The best three-input combination for all age groups is mouse, touch and gestures (for non-parallel tasks), whereas gestures, mouse and tangibles are the best fit for all of them when performing parallel tasks. In turn, the combinations that match the worst are touch, mouse and tangibles, or tangibles, touch and mouse, for all age groups. Additionally, gestures, mouse and tangibles are also the worst fit for older-adults. Regarding tasks with parallel actions, the groups present different results.

The combinations of touch, mouse and tangibles, or tangibles, mouse and gestures, should not be provided in a multimodal system targeted at children. For the group of young adults, the worst combinations are: tangibles, mouse and gestures, or mouse, touch and gestures. Lastly, older-adults should not be provided with the combination of tangibles, touch and mouse for interaction.

Also, note that for the most tasks, there are not second best choices, which means that either the groups are provided with the best synergetic combinations, or the results are drastically worse.

Overall:

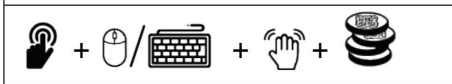
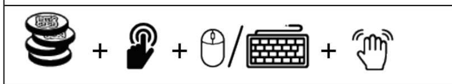
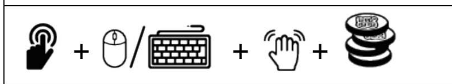
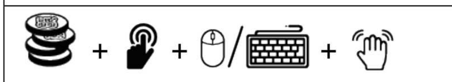
As we have seen by the proposed guidelines above that the best and worst choices for input modalities combinations are scattered through the various multimodal options and age groups. As such, in this situation we cannot provide an overall analysis of the combinations, and thus the designer/developer should try to select their main requisites for the multimodal system based on what age group they intend to target.

6.1.3 Guidelines for combining four interfaces

Regarding the combination of four input modalities in a multimodal interface, we hereupon present the respective guidelines that have the goal to help the development of multimodal interfaces and indicate the best and worst combination choices. Remember that these results were extensively presented and discussed in Chapter 5 of this document.

We present an overview of quantitative results of the three-interface combinations in Table 6.5, following the same visual representation for the best, second best, neutral and worst fits. Similarly, our guidelines are organized into three categories: effectiveness, efficiency and synergy. Also, expect an overall analysis of the results with a more global perspective in some final considerations.

Table 6.5 – Multimodal interaction guidelines for the three age groups using a four-interface combination.

	Interface \ Task	Compound task (with parallelism)		
		Effectiveness	Efficiency	Synergy
CHLDREN		☹️	☹️	☹️
		🙂	😐	☹️
		😐	😐	☹️
		😐	🙂	☹️
YOUNG ADULTS		☹️	😐	☹️
		🙂	😐	☹️
		🙂	☹️	☹️
		😐	🙂	☹️
OLDER-ADULTS		😐	😐	☹️
		😐	😐	☹️
		🙂	☹️	☹️
		☹️	🙂	☹️

Effectiveness:

Provided that the multimodal interface intends to deliver an error-free interaction, the best combination for the age groups are distinct. Children privilege interaction with tangibles, touch, mouse and gestures. This is also true for young adults. Additionally for this group, the combination of mouse, touch, tangibles and gestures can also provide the best results out of the four combinations. Regarding the group of older-adults, the combination that returns fewer errors is mouse, touch, tangibles and gestures.

On the other hand, the worst possible combination for the groups of children and young adults are gestures, mouse, touch and tangibles. Contrarily, the combination that should be avoided by older-adults is touch, mouse, gestures and tangibles.

Also, there are no second best combinations, as all the other options have significantly worse results.

Efficiency:

If the multimodal system being developed has the purpose of exploiting the users' efficiency and prizes speed in completing compound tasks, the three age groups share the same best combination: touch, mouse, gestures and tangibles.

Similarly to effectiveness, the multimodal interface should not consider second choices when targeting users' efficiency, as the remaining combinations are significantly worse. Therefore, children are the least efficient when using gestures, mouse, touch and tangibles; and young and older-adults take more time completing the tasks using mouse, touch, tangibles and gestures as the multimodal combination.

Synergy:

According to the synergetic levels of the different combinations, we have come to realize that, when considering four input modalities working together in one same interface, there are no positive synergies that come from the merging.

Whether we are talking about children, young adults, or older-adults, the various input modalities impair each other during interaction and, thus, negatively influence the modality's relative efficiency. This incurs in a loss on the quality of modality fusion of a multimodal system.

Overall:

On the whole, there is no combination that is more consistent among the three age groups when using four input modalities. As we can observe in the chart, each combination favors different aspects of the system for the three age groups. However, we can see that children reveal a worst overall performance when using gestures, mouse, touch and tangibles.

6.2 Concluding remarks

Even though we provide guidelines for which multimodal interface could be the best fit for the three age groups, we point out that, according to the results obtained during our study, combining more than two input modalities into one multimodal interface can causes negative synergies to occur in the majority of situations. This means that instead of helping the user during interaction, the input modalities start to degrade the users' performance.

As such, it appears to be evident that the creation of distinctive user profiles regarding interaction with different interfaces could be important. Furthermore, these indications have highlighted the need for further investigation to understand key issues related to input modalities and user preferences.

We understand that our study presents certain limitations, mainly because we did not assess the wide spectrum of available input modalities, such as speech-based or haptic interfaces. Clearly, we could also have extended the number of possible combinations for the multimodal evaluation. However, we do consider that our study provides early guidelines that can help and improve the development of multimodal interfaces.

Upon formulating some guidelines that intend to be useful in the development on multimodal systems, there are some considerations that we find important to discuss. Indeed, the combination of touch or the computer mouse with other input modalities can show differences in interaction by the three groups. However, the most noticed changes occurred when using tangibles or gestures. These input modalities need to be carefully chosen and integrated with caution. Thus, we have come to realize that there are a few examples where these interfaces could be of relevance in interaction and positively influence results.

For instance, tangible pieces and gestures have demonstrated to be good choices when the user needs to take a parallel action in the course of the main task, that is, they could help the user during certain actions that complement the main task, and not be considered for the main activity.

A good example where tangible pieces could be used is during a compound task, where the user is constantly switching from one specific task to another, or needs to keep choosing different tools for completing the task. With tangibles, the user could swap tools in a specific drawing software, could switch between perspectives on a 3D environment, or alternate between different operating desktops.

On the other hand, gestures show the best performance with linear movements. Therefore, they could also represent a good choice for complementary activities, where the user does not need to step aside the main task to complete a secondary activity. For instance, when the user is undergoing a main activity and a notification appears, it could be interesting to use gestures to accept or discard that notification with a swipe. This could mean that the user did not have to leave the main activity and loose their focus, and still manage to take action about incoming messages.

These are just a few examples of how the modalities that usually take part in the worst combinations could be used for positive activities and thus enhance the users' productivity.

7 Conclusions and future work

The recent advances made in human-computer interaction have allowed us to manipulate digital contents more intuitively, exploiting recognition-based technologies that interpret complex human behaviors, such as speech, eye gaze, body language or gestures. The way we grow up, live together and grow old is intimately entangled with computers.

This field of research is still evolving and much more attention is being given to different areas of interest. Technology is becoming even more pervasive, ubiquitous, existing everywhere: devices are becoming completely connected and constantly available; and the computer is "disappearing", as coined by the European Initiative for this wave in computing (The Disappearing Computer Initiative (:ac)). This has induced user to a hyper-connectivity with technology and the world (Harper et al., 2008).

However, these engineering and scientific challenges are not isolated. In an interview with several researches in the field of HCI, more challenges were pointed out that should be looked after, whether in terms of educational approaches, health, collaborative work, or even organizational / political challenges (Sears e Jacko, 2007). Also, the social acceptability of computer-based products and services is of utmost importance, being needed more attention to adapting / anticipating new user requirements and thus preventing digital exclusion: e.g. providing more natural and activity-oriented, rather than application-oriented, forms of interaction (Stephanidis, 2007). Stephanidis also believes that the next step is evolving beyond usability and efficiency in design towards a more emotionally pleasant user experience, making the systems more appealing and fun to users as well.

A considerable trend pointed by researches in the field of HCI is the revitalization of intelligent interfaces and computer systems, i.e. user interface personalization, more context- and user-aware (Sears e Jacko, 2007). This is an inspiring area, where adaptation methods and techniques have been investigated, as well as new forms of interaction that have emerged through multimodality and multidimensionality of user interfaces (Stephanidis, 2007), not only for disabled users but also for the population at large.

With the development and evolution of multimodal interfaces' research, it has become more important to comprehend the user's context and behaviors. However, there seems to be a constant addition of new modes of interaction without the proper awareness as to which could be the most adequate for different user profiles (e.g., children, young adults, elderly users) and also regarding which types of tasks (elemental or compound, with or without parallelism).

The aim of this thesis was to study if there were age-related differences in interaction regarding elemental or compound tasks, as well as understand if any synergies emerged from combining different interaction paradigms or, on the other hand, if multimodal interfaces impaired their performance. Indeed, as interfaces that perceive several modes of interaction become more flexible, there should be an interaction model that proposes what types of input they provide regarding the users' preferences or age group.

We have performed a systematic study with the goal of understanding this relation and point out which interfaces are considered best or worst for specific tasks, and also regarding distinctive age groups. Thus, we have come to realize that the creation of distinctive user profiles regarding interaction with different interfaces could, indeed, be important.

First, we researched how users complete three types of elemental tasks (selection, insertion and manipulation) to have a basis for comparison with the remaining compound tasks. We studied different activities the users could perform using four input modalities: the computer mouse / keyboard, touch, tangibles and gestures. Considering all surveys, the most popular interface was touch for all groups and for most tasks: selection and manipulation presented superior results when using touch; but insertion did not and the physical keyboard still managed to obtain better results in the content insertion activity. For steering linear paths, children displayed improved performance using gestures, although this modality was the one groups found to be more difficult and struggling in almost all activities. Discrete and continuous selection tasks may also influence the users' performance: when accomplishing discrete tasks, users are less volatile to the interface used, as opposed to continuous tasks, where they seem to show more susceptibility to the input modality. Also, we came to realize that compound tasks may augment the discrepancies of each modality considering the different age groups, and the type of interface that is used can influence efficiency.

Second, regarding multimodal interfaces, as more interaction modalities are introduced in one same interface, the more volatile it becomes to generate negative synergies. Generally, the number of modalities simultaneously supported by an interface should be two - a bimodal interface - even though young adults can withstand three input modalities in the same multimodal system. However, in terms of user preference, children and young adults favor a bimodal interaction, whereas older-adults point out their unwillingness to use more than one single input modality. As we have discovered, the best and worst choices for input modalities combinations are scattered throughout the various multimodal options and age groups. There is not one specific combination that would please every age group, nor every task. As such, in this situation the designer/developer should try to select their main requisites for the multimodal system based on what age group they intend to target and the main goal, whether it is effectiveness, efficiency, synergy levels or preference.

Finally, we recommended a multimodal interaction model based on certain guidelines, regarding the context of the tasks and also the age group, giving some examples on how to better exploit the input modalities that presented the worst quality results in terms of modality fusion and relative modality efficiency, namely the tangible and gestural interfaces.

7.1 Limitations of the study

We understand that our study comes with certain limitations. Even though we tried to overcome a series of obstacles, there are issues that need to be addressed.

First, the number and type of input modality chosen does not cover the whole range of existing possibilities. As we have seen in the literature review, there are also interfaces based on speech recognition or haptic perception that become increasingly recognized. They are capable of aiding the users in specific contexts, such as while driving or on the move, where users do not have the ability to control the interactive systems otherwise. However, they were not accounted for in our study. Even though these input modalities are becoming more widely used, they were not considered due to technical limitations or applicability constrains. Thus, we have to acknowledge that the number of input modalities that took part in our study was restricted and can be enhanced.

Moreover, the number of combinations of input modalities analyzed in our study could be increased. We have followed certain specifications to help us combine and order the interfaces being used, but there are more combinations that could have been possible and valid for analysis. Indeed, factors such as the acquisition and homing times also influence results, which means that more combinations with different sequences are possible but were not covered in this study. As such, further tests would have to be performed, in order to understand if other possible combinations could provide different results.

One other limitation of our study is certainly the possibility of the participants learning or adjusting to the interaction modalities during the tests. Indeed, we have tried to narrow our sample to very specific groups, and we were aware of the levels of digital literacy or acquaintance with the various input modalities. However, the existence of a learning effect throughout the tests cannot be discarded. Still on this subject, the participants' familiarity with some input modalities to the detriment of others could also have influenced their preferences and conduct during the experiments, as they could be more inclined towards favoring an interaction modality because they were more accustomed to it, and not because it was more helpful or easy to use. Also, more training could have enhanced some results. Further tests would have to be performed in order to attest the validity of results over time, that is, with more training participants could improve performance and even change their opinions about their most and least favorite input modalities.

The context of the tasks studied can be seen as a limitation, as well. We have created specific tasks for controlled environments that may not be scalable to daily activities. We need to acknowledge that other more realistic, every day tasks could provide different results: real contexts of use could influence interaction performance, such as executing conventional or recurring activities with different input modalities. Indeed, we have studied the interfaces with specific requirements and in controlled contexts of use; hence a wider research that encompasses other contexts could be important.

Finally, we did not validate the proposed model. In order for the results to be scalable, we need to corroborate the proposed guidelines with a controlled test in real-world contexts of use. This situation was not covered in this study, but will be explored in future work.

7.2 Future work

Interaction with digital systems is changing the way users live, whether it is within the family, work, or how we grow up and grow older. It is changing lives, and societies. And technology also needs to understand its surroundings, incorporating human values, as moral, social and ethical aspects, in interface design and interaction

(Harper et al., 2008).

In the future, we will direct our attention to validating the proposed multimodal interaction model, thus validating the recommended guidelines. This situation was not covered in this study, but will be explored ahead, in real-world contexts of use and other environments other than in controlled spaces, such as school environments for children, office circumstances, on-the-go situations, among others.

Thus, other specific tasks related to everyday use of digital systems will be pondered in our study as well, and more multimodal combinations will be addressed.

Finally, we should consider different target-audiences and propose guidelines for multimodal interaction with regard to other groups, such as people with disabilities or elderly users.

In summary, our goal is to contribute for the development of more user-friendly and context-aware digital systems, capable of understanding the user profiles and adjust the interfaces accordingly.

References

Batch processing. 20

The Disappearing Computer Initiative. 277

What is a Command Line Interface (CLI)? 20

Accot, J. e Zhai, S. (1999). Performance evaluation of input devices in trajectory-based tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99*, páginas 466–472, New York, New York, USA. ACM Press. 94, 95, 96

Accot, J. e Zhai, S. (2001). Scale effects in steering law tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '01*, páginas 1–8, New York, New York, USA. ACM Press. 64, 94

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843. 60

Analisis e investigacion, S. Aplicaciones del eye tracking. 32

Annett, M., Anderson, F., Goertzen, D., Halton, J., Ranson, Q., Bischof, W. F., e Boulanger, P. (2009). Using a multi-touch tabletop for upper extremity motor rehabilitation. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*, OZCHI '09, páginas 261–264, New York, NY, USA. ACM. 43

Antle, A. N., Bevans, A., Tanenbaum, J., Seaborn, K., e Wang, S. (2011). Futura: design for collaborative learning and game play on a multi-touch digital tabletop. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*, TEI '11, páginas 93–100, New York, NY, USA. ACM. 39, 43

Antonella De Angeli, Walter Gerbino, G. C. e Petrelli, D. (1998). Visual display, pointing, and natural language: the power of multimodal interaction. In *Proceedings of the working conference on Advanced visual interfaces*, páginas 164–173. ACM. 3

Aylett, M. P., Kristensson, P. O., Whittaker, S., e Vazquez-Alvarez, Y. (2014). None of a CHInd. In *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14*, páginas 749–760, New York, New York, USA. ACM Press. 28

Basdogan, C., De, S., Kim, J., Manivannan, M., Kim, H., e Srinivasan, M. A. (2004). Haptics in minimally invasive surgical simulation and training. *Computer Graphics and Applications, IEEE*, 24(2):56–64. 36

Benko, H. e Wilson, A. D. (2008). DepthTouch: Using depth-sensing camera to enable freehand interactions on and above the interactive surface. In *IEEE Workshop on Tabletops and Interactive Surfaces*. 33

Bergman, E. e Johnson, E. (1995). Towards Accessible Human-Computer Interaction. In *in Advances in Human-Computer Interaction*. 14

Bernsen, N. O. e Dybkjaer, L. (1998). Is Speech The Right Thing For Your Application? In *Proc. 5fth Int. Conf. Spoken Lang. Process.*, páginas 3209–3212, Sydney, Australia. 97

Beyer, H. (1981). Tukey, John W.: Exploratory Data Analysis. Addison-Wesley Publishing Company Reading, Mass. — Menlo Park, Cal., London, Amsterdam, Don Mills, Ontario, Sydney 1977, XVI, 688 S. *Biometrical Journal*, 23(4):413–414. 103

Bilici, V., Krahmer, E., te Riele, S., e Veldhuis, R. N. J. (2000). Preferred modalities in dialogue systems. In *Proceedings of ICSLP2000*. 100

Blake, J. (2011). *Natural User Interfaces in .NET*. Manning Publications Co. 22, 29

Blake, J. (2012). The natural user interface revolution. In *Natural User Interfaces in .NET*, capítulo 1, páginas 1–43. Manning Publications Co. 2, 23, 24

Bobeth, J., Schrammel, J., Deutsch, S., Klein, M., Drobics, M., Hochleitner, C., e Tscheligi, M. (2014). Tablet, gestures, remote control? In *Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video - TVX '14*, páginas 139–146, New York, New York, USA. ACM Press. 5

Bolt, R. A. (1980). "Put-that-there": Voice and gesture at the graphics interface. In *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, {SIGGRAPH} '80, páginas 262–270, New York, NY, USA. ACM. 46, 48

Bowman, D. A. (1999). Interaction techniques for common tasks in immersive virtual environments: design, evaluation, and application. 71

Bowman, D. A., Gabbard, J. L., e Hix, D. (2002). A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. *Presence: Teleoperators and Virtual Environments*, 11(4):404–424. 89

Bowman, D. A., Kruijff, E., LaViola, J. J., e Poupyrev, I. (2001). An Introduction to 3-D User Interface Design. *Presence: Teleoperators and Virtual Environments*, 10(1):96–108. 3

Branaghan, R. (1997). Ten tips for selecting usability test participants. *Common Ground*, (7):3–6. 88

Brewer, J., Williams, A., e Dourish, P. (2007). A handle on what's going on: combining tangible interfaces and ambient displays for collaborative groups. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, páginas 3–10, Baton Rouge, Louisiana. ACM. 40

Bridgeman, B. (1992). Conscious vs Unconscious Processes: The Case of Vision. *Theory & Psychology*, 2(1):73–88. 31

Bruckman, A., Bandlow, A., e Forte, A. (2007). HCI FOR KIDS. In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Second Edition (Human Factors and Ergonomics)*, páginas 793–809. 2 edition. 19

Burdea, G. (2000). Haptics issues in virtual environments. In *Proceedings Computer Graphics International 2000*, páginas 295–302. IEEE Comput. Soc. 36

Buxton, B. (2007). *Sketching User Experiences: Getting the Design Right and the Right Design (Interactive Technologies)*. Morgan Kaufmann Publishers Inc., 1 edition. 57

Buxton, B. (2010). CES 2010: NUI with Bill Buxton. 2, 21

Buxton, B. (2016). Bio. 75

Buxton, W. (1983). Lexical and pragmatic considerations of input structures. *ACM SIGGRAPH Computer Graphics*, 17(1):31–37. 63

Buxton, W. (1986). Chunking and Phrasing and the Design of Human-Computer Dialogues. In *Proceedings of the 10th IFIP World Computer Congress*, páginas 475–480. 59, 60

Buxton, W. (1990). A three-state model of graphical input. páginas 449–456. 62, 66, 68

Buxton, W. e Myers, B. (1986). A study in two-handed input. *ACM SIGCHI Bulletin*, 17(4):321–326. 60

Cafaro, F., Lyons, L., Radinsky, J., e Roberts, J. (2010). RFID localization for tangible and embodied multi-user interaction with museum exhibits. In *Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing*, páginas 397–398, Copenhagen, Denmark. ACM. 40

Card, S. K., Mackinlay, J. D., e Robertson, G. G. (1991). A morphological analysis of the design space of input devices. *ACM Transactions on Information Systems*, 9(2):99–122. 63, 65, 66, 92

Carvalho, D., Bessa, M., e Magalhães, L. (2014). Different interaction paradigms for different user groups: an evaluation regarding content selection. In *Proceedings of the XV International Conference on Human Computer Interaction - Interaccion'14*, página 40, Puerto de la Cruz, Tenerife, Spain. ACM. 106

Carvalho, D., Bessa, M., Magalhaes, L., e Carrapatoso, E. (2015a). Interaction Paradigms Versus Age-Related User Profiles: an Evaluation on Content Selection. *IEEE Latin America Transactions*, 13(2):532–539. 106

Carvalho, D., Bessa, M., Magalhães, L., e Carrapatoso, E. (2015b). Performance evaluation of gesture-based interaction between different age groups using Fitts' Law. In *Proceedings of the XVI International Conference on Human Computer Interaction - Interacción '15*, páginas 1–7, New York, New York, USA. ACM Press. 106, 138

Carvalho, D., Bessa, M., Magalhães, L., e Carrapatoso, E. (2016). Performance evaluation of different age groups for gestural interaction: a case study with Microsoft Kinect and Leap Motion. *International Journal Universal Access in the Information Society UAIS*. 106, 138

Carvalho, D., Bessa, M., Oliveira, L., Guedes, C., Peres, E., e Magalhães, L. (2012a). New Interaction Paradigms to Fight the Digital Divide: A Pilot Case Study Regarding Multi-Touch Technology. *Procedia Computer Science*, 14:128–137. 17, 44, 79

Carvalho, D., Bessa, M., Peres, E., Magalhães, L., Guedes, C., e Oliveira, L. (2012b). Developing a multi-touch serious game to fight the digital divide: The Portuguese

ATM: A pilot case study. In *7th Iberian Conference on Information Systems and Technologies (CISTI)*, páginas 1–6, Madrid. IEEE. 17, 43, 44, 79

Chen, F. (2006). *Designing Human Interface in Speech Technology*. Springer US, 1 edition. 47

Chui Yin, W., Chu, K., Chee Weng, K., e Tek Yong, L. (2010). Evaluating playability on haptic user interface for mobile gaming. In *Information Technology (ITSim), 2010 International Symposium in*, volume 3, páginas 1093–1098. 36

Clifton, P., Mazalek, A., Sanford, J., Rébola, C., Lee, S., e Powell, N. (2011). SketchTop: design collaboration on a multi-touch tabletop. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*, páginas 333–336, Funchal, Portugal. ACM. 44

Cohen, P. R. (1992). The role of natural language in a multimodal interface. In *Proceedings of the 5th annual ACM symposium on User interface software and technology*, páginas 143–149, Monteray, California, United States. ACM. 45

Cohen, P. R., Dalrymple, M., Moran, D. B., Pereira, F. C., e Sullivan, J. W. (1989). Synergistic use of direct manipulation and natural language. *ACM SIGCHI Bulletin*, 20(SI):227–233. 54

Cohen, P. R., Johnston, M., Mcgee, D., Oviatt, S. L., Clow, J., e Smith, I. A. (1998). The efficiency of multimodal interaction: a case study. In *Proceedings of the International Conference on Spoken Language Processing*, páginas 249–252, Sydney, Australia. 47

Constantine Stephanidis e Pier Luigi Emiliani (1999). 'Connecting' to the information society: a European perspective. 14

Cook, A. M. e Polgar, J. M. (2015). *Assistive Technologies: Principles and Practice*. Mosby, 4th editio edition. 14

Cooper, J. e Kugler, M. B. (2007). THE DIGITAL DIVIDE: THE ROLE OF GENDER IN HUMAN COMPUTER INTERACTION. In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Second Edition (Human Factors and Ergonomics)*, páginas 763–775. 2 edition. 18

Curtis, B. e Hefley, B. (1994). A wimp no more: the maturing of user interface engineering. *interactions*, 1(1):22–34. 1, 2

Cuypers, T., Schneider, J., Taelman, J., Luyten, K., e Bekaert, P. (2008). Eunomia: toward a framework for multi-touch information displays in public spaces. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction - Volume 2*, BCS-HCI '08, páginas 31–34, Swinton, UK, UK. British Computer Society. 43

Czaja, S. J. e Lee, C. C. (2007). INFORMATION TECHNOLOGY AND OLDER ADULTS. In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Second Edition (Human Factors and Ergonomics)*, páginas 777–792. 2 edition. 18

da Silva, A. B. (2008). *Interação multimodal na execução de tarefas em paralelo*. PhD thesis, Universidade do Estado de Santa Catarina – UDESC. 61, 62

Dang, C. T., Straub, M., e André, E. (2009). Hand distinction for multi-touch tabletop interaction. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, páginas 101–108, New York, NY, USA. ACM. 42

Danis, C. M. (1989). DEVELOPING SUCCESSFUL SPEAKERS FOR AN AUTOMATIC SPEECH RECOGNITION SYSTEM. volume 33, páginas 301—-304(4). Computer Systems. 26

de Castro, L. B. (2013). *Human-Computer Interaction via brainwaves for disabled people*. PhD thesis, University of Porto. 33

de Lemos, G. C. E. M. P. (2006). *Habilidades cognitivas e rendimento escolar entre o 5.Âº e 12.Âº anos de escolaridade*. PhD thesis, Universidade do Minho. 78, 79

Dix, A., Finlay, J. E., Abowd, G. D., e Beale, R. (2004). *Human-Computer Interaction (3rd Edition)*. Pearson Prentice Hall, 3rd editio edition. 12, 13

Don, L. e Smith, S. P. (2010). Applying bimanual interaction principles to text input on multi-touch surfaces and tabletops. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, páginas 253–254, New York, NY, USA. ACM. 42

Donker, A. e Reitsma, P. (2007). Young children's ability to use a computer mouse. *Computers & Education*, 48(4):602–617. 78

Douglas, S. A. e Mithal, A. K. (1994). The effect of reducing homing time on the speed of a finger-controlled isometric pointing device. In *Proceedings of the SIGCHI conference on Human factors in computing systems celebrating interdependence - CHI '94*, páginas 411–416, New York, New York, USA. ACM Press. 65

Drury, C. G. (1971). Movements with Lateral Constraint. *Ergonomics*, 14(2):293–305. 94

Dumas, J. e Fox, J. (2008). Usability Testing: Current Practice and Future Directions. In Sears, A. e Jacko, J., editors, *The Human-Computer Interaction Handbook*, páginas 1129– 1149. Taylor and Francis, USA, 2nd editio edition. 87, 97

Dumas, J. S. e Redish, J. (1999). *A practical guide to usability testing*. Intellect Books. 88

El-Glaly, Y. N., Quek, F., Smith-Jackson, T., e Dhillon, G. (2013). Touch-screens are not tangible. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction - TEI '13*, página 245, New York, New York, USA. ACM Press. 4, 46

English, W. K. e Douglas C. Engelbart, M. L. B. (1967). Display-Selection Techniques for Text Manipulation. *IEEE Transactions on Human Factors in Electronics*, HFE-8(1):5–15. 65

Epps, B. W. (1986). Comparison of Six Cursor Control Devices Based on Fitts' Law Models. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 30(4):327–331. 65

Ericsson, K. A. K. A. e Simon, H. A. H. A. (1993). *Protocol analysis : verbal reports as data*. MIT Press. 89

Fallman, D. (2002). Wear, point, and tilt: designing support for mobile service and maintenance in industrial settings. In *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*, páginas 293–302, London, England. ACM. 40

Ferscha, A., Vogl, S., Emsenhuber, B., e Wally, B. (2007). Physical shortcuts for media remote controls. In *Proceedings of the 2nd international conference on INtelligent TEchnologies for interactive enterTAINment*, páginas 1–8, Cancun, Mexico. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering). 40

Feyereisen, P. (1994). Hand and Mind: What Gestures Reveal about Thought. *The American Journal of Psychology*, 107(1):149–155. 48

Findlater, L., Froehlich, J. E., Fattal, K., Wobbrock, J. O., e Dastyar, T. (2013). Age-related differences in performance with touchscreens compared to traditional mouse input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, página 343, New York, New York, USA. ACM Press. 5

Fishkin, K. P. (2004). A taxonomy for and analysis of tangible interfaces. *Personal Ubiquitous Comput.*, 8(5):347–358. 37

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6):381–391. 64, 89

Fitzmaurice, G. W., Ishii, H., e Buxton, W. A. S. (1995). Bricks: laying the foundations for graspable user interfaces. In *Proceedings of the SIGCHI conference* *on Human factors in computing systems - CHI '95*, páginas 442–449, New York, New York, USA. ACM Press. 37

Foley, J. D., Wallace, V. L., e Chan, P. (1984). The human factors of computer graphics interaction techniques. *IEEE Computer Graphics and Applications*, 4(11):13–48. xxvii, 57, 58, 59, 66, 67, 70, 71, 72, 105

Galitz, W. O. (2007). *The essential guide to user interface design: An introduction to GUI design principles and techniques*. Wiley. 20, 28, 81

Godwin-Jones, R. (2009). Emerging technologies: speech tools ans technologies. *Language Learning & Technology*, 13(3):4–11. 25, 26, 27

Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N., e Wichansky, A. M. (2002). Eye tracking in web search tasks. In *Proceedings of the symposium on Eye tracking research & applications - ETRA '02*, página 51, New York, New York, USA. ACM Press. 33

Gomes, E. (2002). Exclusão digital: um problema tecnológico ou social? *Rio de Janeiro: Trabalho e Sociedade*, Especial 2002 - Conhecimento e Desigualdade:8. 17

Gorniak, P. e Roy, D. (2003). Augmenting user interfaces with adaptive speech commands. In *Proceedings of the 5th international conference on Multimodal interfaces - ICMI '03*, página 176, New York, New York, USA. ACM Press. 27

Greenberg, S. e Fitchett, C. (2001). Phidgets. In *Proceedings of the 14th annual ACM symposium on User interface software and technology - UIST '01*, página 209, New York, New York, USA. ACM Press. 40

Grifoni, P. (2009). *Multimodal Human Computer Interaction and Pervasive Services*. Istituto di Ricerca sulla Popolazione - Consiglio Nazionale delle Ricerche, Italy. 74

Gross, T., Fetter, M., e Liebsch, S. (2008). The cuetable: cooperative and competitive multi-touch interaction on a tabletop. In *CHI '08 extended abstracts on*

Human factors in computing systems, CHI EA '08, páginas 3465–3470, New York, NY, USA. ACM. 44

Grudin, J. (1992). Utility and usability: research issues and development contexts. *Interacting with Computers*, 4(2):209–217. 86

Guiard, Y. (1987). Asymmetric Division of Labor in Human Skilled Bimanual Action: The Kinematic Chain as a Model. *Journal of Motor Behavior*, 19:486–517. 43

Gujar, A., Fishkin, K. P., Harrison, B. L., e Want, R. (1999). Manipulative user interfaces: exploring physically embodied user interfaces. In *CHI '99 extended abstracts on Human factors in computing systems*, páginas 33–34, Pittsburgh, Pennsylvania. ACM. 40

Hale, K. S. e Stanney, K. M. (2004). Deriving haptic design guidelines from human physiological, psychophysical, and neurological foundations. *Computer Graphics and Applications, IEEE*, 24(2):33–39. xxvii, 35

Hamza-Lup, F. G., Lambeth, B. M., e LaPlant, J. (2009). Collaborative haptic environment assessment. In *EuroHaptics conference, 2009 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2009. Third Joint*, páginas 397–398. 36

Hanna, L., Risden, K., e Alexander, K. (1997). Guidelines for usability testing with children. *interactions*, 4(5):9–14. 19

Hantrais, L. (2009). *International Comparative Research: Theory, Methods and Practice*. Basingstoke (England). 84

Harper, R., Rodden, T., Rogers, Y., e Sellen, A. (2008). *Being Human: Human-Computer Interaction in the year 2020*. Microsoft Research Ltd, Cambridge. 44, 277, 282

Harris, A., Rick, J., Bonnett, V., Yuill, N., Fleck, R., Marshall, P., e Rogers, Y. (2009). Around the table: are multiple-touch surfaces better than single-touch for children's collaborative interactions? In *Proceedings of the 9th international* *conference on Computer supported collaborative learning - Volume 1*, CSCL'09, páginas 335–344. International Society of the Learning Sciences. 44

Hewett, T. T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., Perlman, G., Strong, G., e Verplank, W. (1992). ACM SIGCHI Curricula for Human-Computer Interaction. Relatório técnico, New York, NY, USA. 12, 13

Hilliges, O., Izadi, S., Wilson, A. D., Hodges, S., Garcia-Mendoza, A., e Butz, A. (2009). Interactions in the air. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology - UIST '09*, página 139, New York, New York, USA. ACM Press. 33

Hinckley, K. (2007). Input technologies and techniques. In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Second Edition*, capítulo 9, páginas 161–195. 59, 60, 62, 63, 64, 66, 68, 69

Hinckley, K., Cutrell, E., Bathiche, S., e Muss, T. (2002). Quantitative analysis of scrolling techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems Changing our world, changing ourselves - CHI '02*, página 65, New York, New York, USA. ACM Press. 64

Hinckley, K., Jacob, R., e Ware, C. (2004). Input/Output Devices and Interaction Techniques. In A.B. Tucker, C. e Hall, editors, *The Computer Science Handbook*, capítulo 20, páginas 20.1 – 20.32. CRC Press, 2nd edition. xxvi, xxvii, 62, 63, 64, 69

Hoaglin, D. C. e Iglewicz, B. (1987). Fine-Tuning Some Resistant Rules for Outlier Labeling. *Journal of the American Statistical Association*, 82(400):1147–1149. 103

Hone, K. S. e Graham, R. (2000). Towards a tool for the Subjective Assessment of Speech System Interfaces (SASSI). *Natural Language Engineering*, 6(3&4):S1351324900002497. 97

Hourcade, J. P., Bullock-Rest, N. E., e Hansen, T. E. (2012). Multitouch tablet applications and activities to enhance the social skills of children with autism spectrum disorders. *Personal Ubiquitous Comput.*, 16(2):157–168. 44, 80

Hwan, H., Eui Chul, L., Kang Ryoung, P., Chi Jung, K., e Mincheol, W. (2010). A realistic game system using multi-modal user interfaces. *Consumer Electronics, IEEE Transactions on*, 56(3):1364–1372. 32

Ichiyanagi, Y., Cooper, E., Kryssanov, V., e Ogawa, H. (2011). A Haptic Emotional Model for Audio System Interface Human-Computer Interaction. Towards Mobile and Intelligent Interaction Environments. volume 6763 of *Lecture Notes in Computer Science*, páginas 535–542. Springer Berlin / Heidelberg. 36

Inai, Y. (2013). Age-related differences in pointing movements in restricted visual tasks and their design implication. In *2013 ICME International Conference on Complex Medical Engineering*, páginas 439–443. IEEE. 5

Inkpen, K. M. (2001). Drag-and-drop versus point-and-click mouse interaction styles for children. *ACM Transactions on Computer-Human Interaction*, 8(1):1–33. 19

Ishii, H. (2008). The Tangible User Interface and Its Evolution. *Commun. ACM*, 51(6):32–36. 38

Ishii, H., Lakatos, D., Bonanni, L., e Labrune, J.-B. (2012). Radical Atoms: Beyond Tangible Bits, Toward Transformable Materials. *interactions*, 19(1):38–51. 37

Ishii, H. e Ullmer, B. (1997). Tangible bits. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '97*, páginas 234–241, New York, New York, USA. ACM Press. 37

ISO (2000). ISO 9241-9 International standard: Ergonomic requirements for office work with visual display terminals (VDTs) – Part 9: Requirements for non-keyboard input devices: International Organization for Standardization. Relatório técnico. 90

ISO (2007). ISO 9241-400:2007 - Ergonomics of human–system interaction – Part 400: Principles and requirements for physical input devices: International Organization for Standardization. Relatório técnico, International Organization for Standardization. 90

ISO (2015a). ISO 9241-210:2010 - Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems. Relatório técnico. 15

ISO (2015b). ISO/DIS 9241-11 : Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts. Relatório técnico, International Organization for Standardization. xxv, 86

Jacob, R. J. e Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In Hyona, Radach, e Deubel, editors, *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, capítulo 4. Elsevier Science BV, Oxford, England. 33

Jaijongrak, V., Kumazawa, I., e Thiemjarus, S. (2011). A haptic and auditory assistive user interface: Helping the blinds on their computer operations. In *Rehabilitation Robotics (ICORR), 2011 IEEE International Conference on*, páginas 1–6. 36

Jaimes, A. e Sebe, N. (2007). Multimodal human–computer interaction: A survey. *Computer Vision and Image Understanding*, 108(1-2):116–134. 72, 73

Jain, J., Lund, A., e Wixon, D. (2011). The future of natural user interfaces. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA '11*, página 211, New York, New York, USA. ACM Press. 3, 24, 45

Johnson, S., Li, Y., Nam, C., e Yamaguchi, T. (2011). Analyzing User Behavior within a Haptic System Human-Computer Interaction. Interaction Techniques and Environments. volume 6762 of *Lecture Notes in Computer Science*, páginas 62–70. Springer Berlin / Heidelberg. 36

Joiner, R., Messer, D., Light, P., e Littleton, K. (1998). It is best to point for young children: a comparison of children's pointing and dragging. *Computers in Human Behavior*, 14(3):513–529. 78

Jokinen, K., Nishida, M., e Yamamoto, S. (2010). On eye-gaze and turn-taking. In *Proceedings of the 2010 workshop on Eye gaze in intelligent human machine interaction*, páginas 118–123, Hong Kong, China. ACM. 32

Jonnalagadda, H. (2014). Apple reportedly developing in-house speech recognition solution for Siri. 25

Jonsson, I.-M., Nass, C., e Min Lee, K. (2004). Mixing personal computer and hand-held interfaces and devices: effects on perceptions and attitudes. *International Journal of Human-Computer Studies*, 61(1):71–83. 4, 105

Kalicki, K., Starzynski, F., Jenerowicz, A., e Marasek, K. (2007). Simple ossiculoplasty surgery simulation using haptic device. In *Multimedia and Ubiquitous Engineering, 2007. MUE '07. International Conference on*, páginas 932–936. 36

Kaltenbrunner, M., Jorda, S., Geiger, G., e Alonso, M. (2006). The reacTable*: A Collaborative Musical Instrument. In *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2006. WETICE '06. 15th IEEE International Workshops on*, páginas 406–411. 39

Kammer, D., Keck, M., Freitag, G., e Wacker, M. (2010). Taxonomy and Overview of Multi-touch Frameworks: Architecture, Scope and Features. Berlin, Germany. 43

Karat, C.-M., Campbell, R., e Fiegel, T. (1992). Comparison of empirical testing and walkthrough methods in user interface evaluation. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '92*, páginas 397–404, New York, New York, USA. ACM Press. 87

Karray, F., Alemzadeh, M., Saleh, J. A., e Arab, M. N. (2008). Human-Computer Interaction: Overview on State of the Art. *INTERNATIONAL JOURNAL ON SMART SENSING AND INTELLIGENT SYSTEMS*, 1:137–159. xxvii, 71

Kaye, H. S. (2000). Computer and Internet Use among People with Disabilities. Disability Statistics Report 13. 18

Khan, M., Sulaiman, S., Said, M. D. A., e Tahir, M. (2010). Exploring the quantitative and qualitative measures for haptic systems. In *Information Technology (ITSim), 2010 International Symposium in*, volume 1, páginas 31–36. 35

Kim, J., Kim, H., Tay, B. K., Muniyandi, M., Srinivasan, M. A., Jordan, J., Mortensen, J., Oliveira, M., e Slater, M. (2004). Transatlantic touch: a study of haptic collaboration over long distance. *Presence: Teleoper. Virtual Environ.*, 13(3):328–337. 36

Kin, K., Agrawala, M., e DeRose, T. (2009). Determining the benefits of direct-touch, bimanual, and multifinger input on a multitouch workstation. páginas 119–124. 41, 42

Kirbis, M. e Kramberger, I. (2009). Mobile device for electronic eye gesture recognition. *Consumer Electronics, IEEE Transactions on*, 55(4):2127–2133. 33

Kjeldsen, R. e Kender, J. (1996). Toward the use of gesture in traditional user interfaces. In *Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on*, páginas 151–156. 30

Kjeldsen, R. e Kender, J. (1997). Interaction with on-screen objects using visual gesture recognition. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, páginas 788–793. 31

Krippendorff, K. (1997). Human-Centeredness: A paradigm shift invoked by the emerging cyberspaces. página 8, Zentrum für Kunst und Medientechnologie, Karlsruhe, Germany. University of Pennsylvania. 21, 75

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. 20

L Schomaker, J Nijtmans, A. C. F. L. P. M. C. B. T. G.-M. B. L. G. J. R.-R. A. A. I. D. S. M. K. H. J. B. (1995). A taxonomy of multimodal interaction in the human information processing system. Relatório técnico, Esprit Project. 47

Lane, A. E. e Ziviani, J. M. (2010). Factors influencing skilled use of the computer mouse by school-aged children. *Computers & Education*, 55(3):1112–1122. 78

Larsen, L. (2003). Issues in the evaluation of spoken dialogue systems using objective and subjective measures. In *2003 IEEE Workshop on Automatic Speech Recognition and Understanding (IEEE Cat. No.03EX721)*, páginas 209–214. IEEE. 98

Lech, M. e Kostek, B. (2010). Gesture-based computer control system applied to the interactive whiteboard. 33, 81

Lee, S., Henderson, V., Hamilton, H., Starner, T., Brashear, H., e Hamilton, S. (2005). A gesture-based american sign language game for deaf children. In *CHI '05 extended abstracts on Human factors in computing systems - CHI '05*, página 1589, New York, New York, USA. ACM Press. 33, 81

Lee, W. e Park, J. (2005). Augmented Foam: A Tangible Augmented Reality for Product Design. In *Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality*, páginas 106–109. IEEE Computer Society. 39

Leiser, R. (1989). Exploiting convergence to improve natural language understanding. *Interacting with Computers*, 1(3):284–298. 26

Leonardi, C., Albertini, A., Pianesi, F., e Zancanaro, M. (2010). An exploratory study of a touch-based gestural interface for elderly. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, NordiCHI '10, páginas 845–850, New York, NY, USA. ACM. 43

Li, Y., Johnson, S., e Nam, C. (2011). Haptically Enhanced User Interface to Support Science Learning of Visually Impaired Human-Computer Interaction. Users and Applications. volume 6764 of *Lecture Notes in Computer Science*, páginas 68–76. Springer Berlin / Heidelberg. 36

Liarokapis, F., Macan, L., Malone, G., Rebolledo-Mendez, G., e Freitas, S. d. (2009). Multimodal augmented reality tangible gaming. *Vis. Comput.*, 25(12):1109–1120. 39

Lievens, J. e Van Daele, T. (2015). Touch Teach Learn. Unlocking the potential of touch enabled mobile devices for higher education. In *INTED 2015 Proceedings 9th International Technology, Education and Development Conference*, páginas 5298–5304. 44, 80

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22 140:55. 102

Liu, J. e Kavakli, M. (2010). A survey of speech-hand gesture recognition for the development of multimodal interfaces in computer games. In *2010 IEEE International Conference on Multimedia and Expo*, páginas 1564–1569. IEEE. 4, 46

Löhr, A. e Brügge, B. (2008). Mixed-initiative dialog management for speech-based interaction with graphical user interfaces. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, CHI '08, páginas 979–988, New York, NY, USA. ACM. 26

Lor, P. J. (2011). Methodology in comparative studies. In *International and Comparative Librarianship*, capítulo 4. 84

Lucchi, A., Jermann, P., Zufferey, G., e Dillenbourg, P. (2010). An empirical evaluation of touch and tangible interfaces for tabletop displays. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction - TEI '10*, página 177, New York, New York, USA. ACM Press. 4, 46

Mace, R. L., Hardie, G. J., e Place, J. P. (1991). Accessible Environments: Toward Universal Design. In Preiser, W., Vischer, J., e White, E., editors, *Design Intervention: Toward a More Humane Architecture*, página 32. Van Nostrand Reinhold, New York, NY, USA. 15

MacKenzie, I. S. (1992). Fitts' Law as a Research and Design Tool in Human-Computer Interaction. *Human-Computer Interaction*, 7(1):91–139. 89, 90, 91, 92, 93, 94, 126

MacKenzie, I. S. (1995). Movement time prediction in human-computer interfaces. páginas 483–492. 91, 92, 93

Mackinlay, J., Card, S., e Robertson, G. (1990). A Semantic Analysis of the Design Space of Input Devices. *Human-Computer Interaction*, 5(2):145–190. 65, 67

Marchionini, G. e Sibert, J. (1991). An agenda for human-computer interaction: science and engineering serving human needs. *ACM SIGCHI Bulletin*, 23(4):17–32. 12, 13

Martinez-Maldonado, R., Clayphan, A., Ackad, C., e Kay, J. (2014). Multi-touch technology in a higher-education classroom. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures the Future of Design - OzCHI '14*, páginas 220–229, New York, New York, USA. ACM Press. 44, 80

Mason, R. L., Gunst, R. F., e Hess, J. L. (2003). *Statistical design and analysis of experiments*. John Wiley & Sons, Ltd. 97

Mayer, R. E. e Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, 38(1):43–52. 1, 2, 24

McKee, J. H., Yetter, F. L., Klaus, J. A. B., e Malik, N. A. (2015). Integrated Electronic Health Care Coordination System with Touch Screen Tablets and Smartphones. 44, 80

Mi Jeong, K. e Maher, M. L. (2007). Collaborative Design in a Tabletop System employing Tangible User Interfaces. In *Computer Supported Cooperative Work in Design, 2007. CSCWD 2007. 11th International Conference on*, páginas 192–197. 40

Miki, M., Kitaoka, N., Miyajima, C., Nishino, T., e Takeda, K. (2014). Improvement of multimodal gesture and speech recognition performance using time intervals between gestures and accompanying speech. *EURASIP Journal on Audio, Speech, and Music Processing*, 2014(1):2. 4, 46

Minogue, J. e Jones, M. G. (2006). Haptics in Education: Exploring an Untapped Sensory Modality. *Review of Educational Research*, 76(3):317–348. 36

Moggridge, B. (2007). *Designing Interactions*. The MIT Press, 1 edition. 55

Morency, L.-P. e Darrell, T. (2006). Head gesture recognition in intelligent interfaces: the role of context in improving recognition. In *Proceedings of the 11th international conference on Intelligent user interfaces*, páginas 32–38, Sydney, Australia. ACM. 28, 31

Morris, M. R., Wobbrock, J. O., e Wilson, A. D. (2010). Understanding users' preferences for surface gestures. In *Proceedings of Graphics Interface 2010*, GI '10, páginas 261–268, Toronto, Ont., Canada, Canada. Canadian Information Processing Society. 43

Mundie, C. (2010). Computing, Naturally. 45

Munteanu, C., Jones, M., Oviatt, S., Brewster, S., Penn, G., Whittaker, S., Rajput, N., e Nanavati, A. (2013). We need to talk. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*, página 2459, New York, New York, USA. ACM Press. 27

Munteanu, C. e Penn, G. (2015). Speech-based Interaction. In *Proceedings of the 20th International Conference on Intelligent User Interfaces - IUI '15*, páginas 437–438, New York, New York, USA. ACM Press. 27

Murray, N., Roberts, D., Steed, A., Sharkey, P., Dickerson, P., e Rae, J. (2007). An assessment of eye-gaze potential within immersive virtual environments. *ACM Trans. Multimedia Comput. Commun. Appl.*, 3(4):1–17. 32

Myers, B., Hudson, S. E., e Pausch, R. (2000). Past, present, and future of user interface software tools. *ACM Trans. Comput.-Hum. Interact.*, 7(1):3–28. 26

Myers, J. L., Well, A. D., e Jr, R. F. L. (2010). *Research Design and Statistical Analysis*. Routledge, 3rd editio edition. 97

Nakano, T. (2008). Flexible shortcuts. In *Proceeding of the twenty-sixth annual CHI conference extended abstracts on Human factors in computing systems - CHI '08*, página 2621, New York, New York, USA. ACM Press. 27

Nawaz, A., Helbostad, J. L., Skjæret, N., Vereijken, B., Bourke, A., Dahl, Y., e Mellone, S. (2014). Designing Smart Home Technology for Fall Prevention in Older People. In Stephanidis, C., editor, *Communications in Computer and Information Science*, volume 435 of *Communications in Computer and Information Science*, capítulo HCI Intern, páginas 485–490. Springer International Publishing, Cham. 44, 80

NC State, U. (1997). The Center for Universal Design - Universal Design Principles. 16

Nedevschi, S., Patra, R. K., e Brewer, E. A. (2005). Hardware speech recognition for user interfaces in low cost, low power devices. In *2005 Design Automation Conference. 42nd Proceedings*, páginas 684–689. 28

Neto, E. V. e Campos, F. F. C. (2014). Evaluating the Usability on Multimodal Interfaces: A Case Study on Tablets Applications. In Marcus, A., editor, *Lecture Notes in Computer Science*, volume 8517 of *Lecture Notes in Computer Science*, capítulo Design, Us, páginas 484–495. Springer International Publishing, Cham. 87

Newell, A. e Card, S. (1985). The Prospects for Psychological Science in Human-Computer Interaction. *Human-Computer Interaction*, 1(3):209–242. 90

Nielsen, J. (1993). Noncommand user interfaces. *Communications of the ACM*, 36(4):83–99. 1, 20, 22, 86

Nintendo (2006). Nintendo. 34

Nix, D., Fairweather, P., e Adams, B. (1998). Speech recognition, children, and reading. In *CHI 98 conference summary on Human factors in computing systems - CHI '98*, páginas 245–246, New York, New York, USA. ACM Press. 19

Norman, D. A. e Draper, S. W. (1986). User Centered System Design; New Perspectives on Human-Computer Interaction. 55

OECD (2001). Understanding the Digital Divide. Relatório técnico, OECD Publications. 17

Ortiz, C. L. (2014). The Road to Natural Conversational Speech Interfaces. *IEEE Internet Computing*, 18(2):74–78. 28

Oviatt, S. (1997). Multimodal Interactive Maps: Designing for Human Performance. *Human-Computer Interaction*, 12(1):93–129. 3, 50, 51, 52

Oviatt, S. (1999). Mutual disambiguation of recognition errors in a multimodel architecture. In *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99*, páginas 576–583, New York, New York, USA. ACM Press. 47, 48, 50, 54

Oviatt, S. (2003). User-centered modeling and evaluation of multimodal interfaces. *Proceedings of the IEEE*, 91(9):1457–1468. 4, 49, 51, 53, 54

Oviatt, S. e Cohen, P. (2000). Perceptual user interfaces: multimodal interfaces that process what comes naturally. *Communications of the ACM*, 43(3):45–53. 3, 4, 47

Oviatt, S. e Cohen, P. R. (2015). The Paradigm Shift to Multimodality in Contemporary Computer Interfaces. *Synthesis Lectures on Human-Centered Informatics*, 8(3):1–243. 45, 46, 51, 175

Oviatt, S., Swindells, C., e Arthur, A. (2008). Implicit user-adaptive system engagement in speech and pen interfaces. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, páginas 969–978, Florence, Italy. ACM. 26

Oviatt, S. L. (2007). Multimodal Interfaces. In Sears, A. e Jacko, J. A., editors, *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Second Edition*, Multimodal Interfaces, páginas 413–432. CRC Press, 2 edition. xxvii, 5, 26, 46

Oviatt, S. L., Cohen, P. R., e Wang, M. (1994). Toward interface design for human language technology: modality and structure as determinants of linguistic complexity. *Speech Communication*, 15(3-4):283–300. 26

P. P. Tanner, W. A. S. B. (1985). Some Issues in Future User Interface Management System (UIMS) Development. In *Part of the series Eurographic Seminars*, capítulo User Inter, páginas 67–79. 65

Pedersen, E. W. e Hornb, K. (2011). Tangible bots: interaction with active tangibles in tabletop interfaces. In *Proceedings of the 2011 annual conference on Human* *factors in computing systems*, páginas 2975–2984, Vancouver, BC, Canada. ACM. 39

Perakakis, M. e Potamianos, A. (2008a). A Study in Efficiency and Modality Usage in Multimodal Form Filling Systems. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(6):1194–1206. 97, 98, 99

Perakakis, M. e Potamianos, A. (2008b). Multimodal system evaluation using modality efficiency and synergy metrics. In *Proceedings of the 10th international conference on Multimodal interfaces - IMCI '08*, página 9, New York, New York, USA. ACM Press. 98, 100

Piaget, J. (1970). *Science of Education and the Psychology of the Child*. Orion Press, New York, NY, USA. 19

Pickard, A. J. (2013). *Research Methods in Information*. ALA Neal-Schuman, 2nd edition. 84

Pietinen, S., Bednarik, R., Glotova, T., Tenhunen, V., e Tukiainen, M. (2008). A method to study visual attention aspects of collaboration: eye-tracking pair programmers simultaneously. páginas 39–42. ACM. 32, 81

Piper, A. M., Weibel, N., e Hollan, J. D. (2011). Write-N-Speak. *ACM Transactions on Accessible Computing*, 4(1):1–20. 4, 46

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., e Carey, T. (1995). *Human-computer interaction*. Addison-Wesley Pub. Co. 87

Pruitt, J. e Adlin, T. (2005). *The Persona Lifecycle: Keeping People in Mind Throughout Product Design*. Morgan Kaufmann Publishers Inc. 88

Pu, Q., Gupta, S., Gollakota, S., e Patel, S. (2013). Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th annual international conference on Mobile computing & networking - MobiCom '13*, página 27, New York, New York, USA. ACM Press. 31

Rabiner, L. e Juang, B.-H. (1993). Fundamentals of speech recognition. 73

Rashevsky, N. (1960). Further contributions to the mathematical biophysics of automobile driving. *The bulletin of mathematical biophysics*, 22(3):257–262. 94

Reactable (2003). Reactable. 39

Reilly, R. B. (1998). Applications of face and gesture recognition for human-computer interaction. MULTIMEDIA '98, páginas 20–27. ACM. 33, 81

Resnick, M., Martin, F., Berg, R., Borovoy, R., Colella, V., Kramer, K., e Silverman, B. (1998). Digital manipulatives: new toys to think with. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, páginas 281–287, Los Angeles, California, United States. ACM Press/Addison-Wesley Publishing Co. 40

Riby, D. e Hancock, P. (2009). Do Faces Capture the Attention of Individuals with Williams Syndrome or Autism? Evidence from Tracking Eye Movements. *Journal of Autism and Developmental Disorders*, 39(3):421–431. 33

Riedenklau, E., Hermann, T., e Ritter, H. (2012). An integrated multi-modal actuated tangible user interface for distributed collaborative planning. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction - TEI '12*, página 169, New York, New York, USA. ACM Press. 4, 46

Robert K. Yin (2009). *Case Study Research: Design and Methods*. SAGE Publications, applied so edition. 85

Robles-De-La-Torre, G. (2009). Virtual Reality: Touch / Haptics. *Sage Encyclopedia of Perception. Thousand Oaks CA, Sage Publications*, 2:1036–1038. 34, 35

Rohs, M., Oulasvirta, A., e Suomalainen, T. (2011). Interaction with magic lenses. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, página 2725, New York, New York, USA. ACM Press. 90

Rosenbaum, S., Rohn, J. A., e Humburg, J. (2000). A toolkit for strategic usability: results from workshops, panels, and surveys. In *Proceedings of the SIGCHI* *conference on Human factors in computing systems - CHI '00*, páginas 337–344, New York, New York, USA. ACM Press. 86

Rukzio, E., Leichtenstern, K., Callaghan, V., Holleis, P., Schmidt, A., e Chin, J. (2006). An Experimental Comparison of Physical Mobile Interaction Techniques: Touching, Pointing and Scanning. In *UbiComp 2006: Ubiquitous Computing*, capítulo Lecture No, páginas 87–104. 71

Saffer, D. (2009). *Designing for interaction: Creating innovative applications and devices*. New Riders, 2 edition. 55, 56

Sambrooks, L. e Wilkinson, B. (2013). Comparison of gestural, touch, and mouse interaction with Fitts' law. In *Proceedings of the 25th Australian Computer-Human Interaction Conference on Augmentation, Application, Innovation, Collaboration - OzCHI '13*, páginas 119–122, New York, New York, USA. ACM Press. 5

Schaffer, S., Schleicher, R., e Möller, S. (2015). Modeling input modality choice in mobile graphical and speech interfaces. *International Journal of Human-Computer Studies*, 75(C):21–34. 28

Schapira, E. e Sharma, R. (2001). Experimental evaluation of vision and speech based multimodal interfaces. In *Proceedings of the 2001 workshop on Percetive user interfaces - PUI '01*, página 1, New York, New York, USA. ACM Press. 109, 123, 142

Scharf, F., Günther, S., Winkler, T., e Herczeg, M. (2010). SpelLit: Development of a multi-touch application to foster literacy skills at elementary schools. In *2010 IEEE Frontiers in Education Conference (FIE)*, páginas T4D–1–T4D–6. IEEE. 44

Schneider, B., Jermann, P., Zufferey, G., e Dillenbourg, P. (2011). Benefits of a Tangible Interface for Collaborative Learning and Interaction. *Learning Technologies, IEEE Transactions on*, 4(3):222–232. 40

Schnotz, W., Bannert, M., e Seufert, T. (2002). Toward an integrative view of text and picture comprehension: visualization effects on the construction of mental models. In Otero, J., León, J. A., e Graesser, A. C., editors, *The psychology of science text comprehension*. Lawrence Erlbaum Associates, Hillsdale, NJ. 48

Scholliers, C., Hoste, L., Signer, B., e De Meuter, W. (2011). Midas: a declarative multi-touch interaction framework. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*, TEI '11, páginas 49–56, New York, NY, USA. ACM. 43

Sears, A. e Jacko, J. A. (2007). *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Second Edition (Human Factors and Ergonomics)*. CRC Press, 2 edition. xxvi, 17, 45, 277, 278

Serafin, Z. (2010). Speech NUI with Zig Serafin. 28

Shaer, O. e Hornecker, E. (2010). *Tangible User Interfaces: Past, Present and Future Directions*. Now Publishers Inc. 38, 40

Shrawankar, U. e Thakare, V. (2010). Speech user interface for computer based education system. In *2010 International Conference on Signal and Image Processing (CSIP)*, páginas 148–152. 27

Sin, A. K. e Zaman, H. B. (2009). Tangible Interaction in Learning Astronomy through Augmented Reality Book-Based Educational Tool. In *Proceedings of the 1st International Visual Informatics Conference on Visual Informatics: Bridging Research and Practice*, páginas 302–313, Kuala Lumpur, Malaysia. Springer-Verlag. 39

Sinclair, P., Martinez, K., Millard, D. E., e Weal, M. J. (2002). Links in the palm of your hand: tangible hypermedia using augmented reality. In *Proceedings of the thirteenth ACM conference on Hypertext and hypermedia*, páginas 127–136, College Park, Maryland, USA. ACM. 39

Smykil, J. (2008). Rumor: Apple in talks with Immersion over Haptic tech. 34

Sodnik, J., Dicke, C., Tomažič, S., e Billinghurst, M. (2008). A user study of auditory versus visual interfaces for use while driving. *International Journal of Human-Computer Studies*, 66(5):318–332. 4, 105

Soegaard, M. e Dam, R. F. (2014). *The Encyclopedia of Human-Computer Interaction, 2nd Ed.* The Interaction Design Foundation, 2nd edition. 12, 14, 15, 55, 56, 57

Soukoreff, R. W. e MacKenzie, I. S. (2004). Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies*, 61(6):751–789. 90

Srinivasan, M. A. (2004). What is Haptics? Relatório técnico, Laboratory for Human and Machine Haptics: The Touch Lab - Massachusetts Institute of Technology. 36

Srinivasan, M. A. e Basdogan, C. (1997). Haptics in virtual environments: Taxonomy, research status, and challenges. *Computers & Graphics*, 21(4):393–404. 34, 35

Stephanidis, C. (2001). User Interfaces for All: New Perspectives into Human-Computer Interaction. In Stephanidis, C., editor, *User Interfaces for All - Concepts, Methods, and Tools*, capítulo 1, página 760. Lawrence Erlbaum Associates, Mahwah, NJ. xxv, 4, 14, 15, 105

Stephanidis, C. (2007). FUTURE TRENDS IN HUMAN-COMPUTER INTERACTION. In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Second Edition (Human Factors and Ergonomics)*. 278

Stephanidis, C. e Salvendy, G. (1998). Toward an Information Society for All: An International Research and Development Agenda. *International Journal of Human-Computer Interaction*, 10(2):107–134. xxv, 14, 15, 17

Stephanidis, C., Salvendy, G., Akoumianakis, D., Arnold, A., Bevan, N., Dardailler, D., Emiliani, P. L., Iakovidis, I., Jenkins, P., Karshmer, A., Korn, P., Marcus, A., Murphy, H., Oppermann, C., Stary, C., Tamura, H., Tscheligi, M., Ueda, H., Weber, G., e Ziegler, J. (1999). Toward an Information Society for All: HCI Challenges and R&D Recommendations. *International Journal of Human-Computer Interaction*, 11(1):1–28. 14, 15

Stewart, J., Raybourn, E. M., Bederson, B., e Druin, A. (1998). When two hands are better than one: enhancing collaboration using single display groupware. In *CHI 98 conference summary on Human factors in computing systems - CHI '98*, páginas 287–288, New York, New York, USA. ACM Press. 19

Suhm, B. (1998). *Multimodal interactive error recovery for non-conversational speech user interfaces*. PhD thesis, Fredericiana Uni- versity, Karlsruhe, Germany. 54

Suhm, B., Myers, B., e Waibel, A. (1996). Interactive recovery from speech recognition errors in speech user interfaces. In *Fourth International Conference on Spoken Language, 1996. ICSLP 96. Proceedings.*, volume 2, páginas 865—-868 vol.2. 26, 27

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, 12(2):257–285. xxvii

Sylla, C., Branco, P., Coutinho, C., e Coquet, M. E. (2009). Storytelling through drawings: evaluating tangible interfaces for children. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, páginas 3461–3466, Boston, MA, USA. ACM. 40

Toomim, M., Kriplean, T., Pörtner, C., e Landay, J. (2011). Utility of human-computer interactions. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, página 2275, New York, New York, USA. ACM Press. 90

Towell, L. (2009). The Student-Teacher Digital Divide and Six New Technology Rollercoaster Rides. *Distance Learning*, 6(2):53–57. 22, 42, 43

Trevisan, D. G., Nedel, L. P., e da SIlva, A. B. (2011). Evaluation of Multimodal Interaction in Parallel Tasks Execution. In *2011 XIII Symposium on Virtual Reality*, páginas 38–47. IEEE. 62

Truschin, S., Schermann, M., Goswami, S., e Krcmar, H. (2014). Designing interfaces for multiple-goal environments. *ACM Transactions on Computer-Human Interaction*, 21(1):1–24. 28

Turk, M. e Robertson, G. (2000). Perceptual user interfaces (introduction). *Communications of the ACM*, 43(3):32–34. 51, 53

Ullmer, B. e Ishii, H. (2001). Emerging Frameworks for Tangible User Interfaces. In *Human-Computer Interaction in the New Millenium*, páginas 579–601. Addison-Wesley Professional. 38, 81

van Dam, A. (1997). Post-WIMP user interfaces. *Commun. ACM*, 40(2):63–67. 1, 21, 22

Van Gog, T. e Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction*, 20(2):95–99. 33

Vazirgiannis, M., Theodoridis, Y., e Sellis, T. (1996). Spatio-temporal composition in multimedia applications. In *Proceedings International Workshop on Multimedia Software Development*, páginas 120–127. IEEE Comput. Soc. Press. 60, 61

Vertegaal, R. (2003). Attentive User Interfaces. *Communications of ACM*, 46(3)(Editorial, Special Issue on Attentive User Interfaces). 32, 81

Vogel, D. e Baudisch, P. (2007). Shift: a technique for operating pen-based interfaces using touch. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*, página 657, New York, New York, USA. ACM Press. 109, 123, 142

Vredenburg, K., Isensee, S., e Righi, C. (2001). *User-Centered Design: An Integrated Approach*. Prentice Hall. 14

Wachs, J. P., Kölsch, M., Stern, H., e Edan, Y. (2011). Vision-based hand-gesture applications. *Communications of the ACM*, 54(2):60. 34, 43

Waldner, M., Hauber, J., Zauner, J., Haller, M., e Billinghurst, M. (2006). Tangible tiles: design and evaluation of a tangible user interface in a collaborative tabletop setup. In *Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*, páginas 151–158, Sydney, Australia. ACM. 39

Watanabe, Y., Iwata, K., Nakagawa, R., Shinoda, K., e Furui, S. (2007). Semi-Synchronous Speech and Pen Input. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, volume 4, páginas IV–409–IV–412. IEEE. 4, 46

Wechsung, I., Engelbrecht, K.-P., Schaffer, S., Seebode, J., Metze, F., e Möller, S. (2009). *Human-Computer Interaction. Novel Interaction Methods and Techniques: 13th International Conference, HCI International 2009, San Diego, CA, USA, July 19-24, 2009, Proceedings, Part II*, capítulo Usability, páginas 113–119. Springer Berlin Heidelberg, Berlin, Heidelberg. 47, 48

Weichert, F., Bachmann, D., Rudak, B., e Fisseler, D. (2013). Analysis of the accuracy and robustness of the leap motion controller. *Sensors (Basel, Switzerland)*, 13(5):6380–93. 138

Wessberg, J., Stambaugh, C. R., Kralik, J. D., Beck, P. D., Laubach, M., Chapin, J. K., Kim, J., Biggs, S. J., Srinivasan, M. A., e Nicolelis, M. A. L. (2000). Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*, 408(6810):361–365. 36

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2):159–177. 48

Wigdor, D., Fletcher, J., e Morrison, G. (2009). Designing user interfaces for multi-touch and gesture devices. páginas 2755–2758. ACM. 4, 75, 105

Wigdor, D. e Wixon, D. (2011). *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. Morgan Kaufmann Publishers Inc. 42

Wobbrock, J. O., Morris, M. R., e Wilson, A. D. (2009). User-defined gestures for surface computing. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, páginas 1083–1092, New York, NY, USA. ACM. 42

Wu, Y. (2000). Design patterns and framework for WIMP user interfaces design. In *CHI '00 extended abstracts on Human factors in computing systems*, páginas 359–360, The Hague, The Netherlands. ACM. 22

Xiaohua Yu, Mian Zhang, Yaofeng Xue, e Zhiting Zhu (2010). An exploration of developing multi-touch virtual learning tools for young children. In *2010 2nd International Conference on Education Technology and Computer*, volume 3, páginas V3–4–V3–7. IEEE. 44, 80

Xie, L., Antle, A. N., e Motamedi, N. (2008). Are tangibles more fun?: comparing children's enjoyment and engagement using physical, graphical and tangible user interfaces. In *Proceedings of the 2nd international conference on Tangible and embedded interaction*, páginas 191–198, Bonn, Germany. ACM. 40

Yankelovich, N. e Lai, J. (1999). Designing speech user interfaces. In *CHI '99 extended abstracts on Human factors in computing systems - CHI '99*, página 124, New York, New York, USA. ACM Press. 26, 27

Yoo, B., Han, J.-J., Choi, C., Yi, K., Suh, S., Park, D., e Kim, C. (2010). 3d user interface combining gaze and hand gestures for large-scale display. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, páginas 3709–3714, Atlanta, Georgia, USA. ACM. 33

You, Y., Sung, M. Y., Kim, N. J., e Jun, K. K. (2007). An Experimental Study on the Performance of Haptic Data Transmission in Networked Haptic Collaboration. In *Advanced Communication Technology, The 9th International Conference on*, volume 1, páginas 657–662. 36

Yu, B., Feijs, L., Funk, M., e Hu, J. (2015). Breathe with Touch: A Tactile Interface for Breathing Assistance System. In Abascal, J., Barbosa, S., Fetter, M., Gross, T., Palanque, P., e Winckler, M., editors, *Human-Computer Interaction – INTERACT 2015 SE - 4*, volume 9298 of *Lecture Notes in Computer Science*, páginas 45–52. Springer International Publishing. 80

Zhai, S. (2004). Characterizing computer input with Fitts' law parameters—the information and non-information aspects of pointing. *International Journal of Human-Computer Studies*, 61(6):791–809. 89

Zhai, S., Kong, J., e Ren, X. (2004). Speed–accuracy tradeoff in Fitts' law tasks—on the equivalency of actual and nominal pointing precision. *International Journal of Human-Computer Studies*, 61(6):823–856. 82, 91, 124, 127

A Appendix:

.1 Consent form

.2 Final questionnaire

Folha de Consentimento

Investigador: Diana Carvalho, Doutoramento em Informática (UTAD)

Propósito: O propósito não pode ser divulgado, pois esta informação pode alterar os dados recolhidos. No final da experiência, o sujeito é esclarecido.

Participantes: Os participantes são voluntários que se ofereceram para efetuar a experiência.

Confidencialidade: A informação recolhida apenas é identificada por um "ID"atribuído aleatoriamente.

Procedimento: O procedimento pode variar de teste para teste. As instruções vão sendo dadas ao longo do desenvolvimento das tarefas.

Benefícios: Os participantes não vão receber quaisquer benefícios pela realização da experiência.

Riscos: O procedimento não causa qualquer tipo de lesões e já foi utilizado em estudos previamente realizados.

Compensação: O correio eletrónico dos participantes vai ser introduzido numa base de dados. Quando o estudo tiver terminado os participantes vão receber um email com todos os dados sobre a experiência.

Desistência: A participação neste estudo é completamente voluntária. Os voluntários não estão vinculados a nenhuma obrigação, estes podem terminar o teste quando o desejarem.

Questões ou Dúvidas posteriores: Em caso de dúvidas sobre qualquer um dos tópicos anteriormente descritos, pode enviar um email ao investigador para: dianac@utad.pt

Primeiro e último nome: ____

Email: ____

Declaração: Fui informado sobre os objetivos deste projeto, assim como dos procedimentos envolvidos nesta experiência. Eu reservo o direito de abandonar a experiência em qualquer fase da mesma e a não ser com o meu consentimento para a sua utilização, toda a informação previamente obtida é destruída ou a minha identidade retirada.

Assinatura: ____

Data: ___/___/___

Inquérito final:

1) Tinha conhecimento de algum dos modos de interação aqui experimentados? Se sim, qual?

___

2) Selecione com uma cruz, para cada uma das interfaces e mediante os critérios abaixo apresentados, a opção que melhor reflete a sua opinião:

a) Facilidade de utilização:

	Extremamente difícil	Muito difícil	Relativamente difícil	Neutro	Relativamente fácil	Muito fácil	Extremamente fácil
Interface Gráfica	○	○	○	○	○	○	○
Interface Multi-toque	○	○	○	○	○	○	○
Interface Tangível	○	○	○	○	○	○	○
Interface Gestual	○	○	○	○	○	○	○

b) Facilidade de aprendizagem:

	Extremamente difícil	Muito difícil	Relativamente difícil	Neutro	Relativamente fácil	Muito fácil	Extremamente fácil
Interface Gráfica	○	○	○	○	○	○	○
Interface Multi-toque	○	○	○	○	○	○	○
Interface Tangível	○	○	○	○	○	○	○
Interface Gestual	○	○	○	○	○	○	○

c) Efeito de fadiga:

	Extremamente cansativa	Muito cansativa	Relativamente cansativa	Neutro	Relativamente fácil	Muito fácil	Extremamente fácil
Interface Gráfica	○	○	○	○	○	○	○
Interface Multi-toque	○	○	○	○	○	○	○
Interface Tangível	○	○	○	○	○	○	○
Interface Gestual	○	○	○	○	○	○	○

d) Naturalidade na interação:

	Extremamente artificial	Muito artificial	Relativamente artificial	Neutro	Relativamente natural	Muito natural	Extremamente natural
Interface Gráfica	○	○	○	○	○	○	○
Interface Multi-toque	○	○	○	○	○	○	○
Interface Tangível	○	○	○	○	○	○	○
Interface Gestual	○	○	○	○	○	○	○

e) Conforto da utilização:

	Extremamente frustrante	Muito frustrante	Relativamente frustrante	Neutro	Relativamente confortável	Muito confortável	Extremamente confortável
Interface Gráfica	○	○	○	○	○	○	○
Interface Multi-toque	○	○	○	○	○	○	○
Interface Tangível	○	○	○	○	○	○	○
Interface Gestual	○	○	○	○	○	○	○

f) Concentração necessária para completar a tarefa:

	Extremamente difícil	Muito difícil	Relativamente difícil	Neutro	Relativamente fácil	Muito fácil	Extremamente fácil
Interface Gráfica	○	○	○	○	○	○	○
Interface Multi-toque	○	○	○	○	○	○	○
Interface Tangível	○	○	○	○	○	○	○
Interface Gestual	○	○	○	○	○	○	○

3) Organize por prioridades todas as interfaces mediante a sua preferência de utilização, ou seja, a que mais gostou de utilizar (1 sendo a que mais gostou – 4 sendo a que menos gostou).

Interface Gráfica	Interface Multi-toque	Interface Tangível	Interface Gestual

4) Organize por prioridades todas as interfaces mediante a sua preferência de utilização a nível de rapidez, ou seja, a qual foi mais rápida de utilizar (1 sendo a mais rápida – 4 sendo a menos rápida).

Interface Gráfica	Interface Multi-toque	Interface Tangível	Interface Gestual