UNIVERSITÀ DEGLI STUDI DI PALERMO
Dottorato di Ricerca in Ingegneria dell'Innovazione Tecnologica
Dipartimento dell'Innovazione Industriale e Digitale (DIID)
Ingegneria Chimica, Gestionale, Informatica e Meccanica
SSD: ING-INF/05

Designing Touchless Gestural Interfaces for Public Displays

IL DOTTORE
VITO GENTILE

IL COORDINATORE
PROF. SALVATORE GAGLIO

IL TUTOR
PROF. ANTONIO GENTILE

IL COTUTOR
ING. SALVATORE SORCE

CICLO XXIX

ANNO CONSEGUIMENTO TITOLO: 2017

Designing Touchless Gestural Interfaces for Public Displays

Ph.D. Thesis

Vito Gentile

Abstract

In the last decade, many authors have investigated and studied touchless and gestural interactions as a novel tool for interacting with computers. Moreover, technological innovations have allowed for installations of interactive displays in private and public places. However, interactivity is usually implemented by touchscreens, whereas technologies able to recognize body gestures are more rarely adopted, especially in integration with commercial public displays. Nowadays, the opportunity to investigate touchless interfaces for such systems has become concrete and studied by many researchers. Indeed, this interaction modality offers the possibility to overcome several issues that cannot be solved by touch-based solutions, e.g. keeping a high hygiene level of the screen surface, as well as providing big displays with interactive capabilities.

The main goal of this thesis is to describe the design process for implementing touchless gestural interfaces for public displays. This implies the need for overcoming several typical issues of both public displays (e.g. interaction blindness, immediate usability) and touchless interfaces (e.g. communicating touchless interactivity). To this end, a novel Avatar-based Touchless Gestural Interface (or ABaToGI) has been developed, and its design process is described in the thesis, along with the user studies conducted for its evaluation.

Moreover, the thesis analyzes how the presence of the Avatar may affect user interactions in terms of perceived cognitive workload, and if it may be able to foster bimanual interactions.

Then, as ABaToGI was designed for public displays, it has been installed in an actual deployment in order to be evaluated in-the-wild (i.e. not in a lab setting). The resulting outcomes, along with the previously described studies, have been used to introduce a set of design guidelines for developing future touchless gestural interfaces, with a particular focus on Avatar-based ones.

The results of this thesis provide also a basis for future research, which concludes this work.

Sommario

Nell'ultimo decennio, molti autori hanno studiato la possibilità di utilizzare le interfacce a gesti come strumento innovativo per supportare l'interazione con i computer. Inoltre, le recenti innovazioni tecnologiche hanno permesso di installare display interattivi in ambienti privati e pubblici. Tuttavia, l'interattività di tali display è spesso basata sull'uso di touchscreen, mentre tecnologie come i dispositivi Kinect-like vengono adottate molto più raramente, soprattutto se si considera l'ambito dei display pubblici. Al giorno d'oggi, l'opportunità di studiare le interfacce touchless per i display pubblici è diventata concreta, e rappresenta il campo di studio di diversi ricercatori.

L'obiettivo principale di questa tesi è quello di descrivere e studiare i problemi legati alla progettazione e all'implementazione di un'interfaccia grafica dedicata all'interazione touchless a gesti con display pubblici. Ciò implica la necessità di superare alcuni problemi tipici, sia dei display pubblici (ad esempio, l'interaction blindness e l'usabilità immediata), che delle interfacce touchless (per esempio, comunicare che l'interattività è gestuale).

La tesi, inoltre, include uno studio che analizza quanto la presenza dell'Avatar possa influire sulle interazioni degli utenti, in termini di carico di lavoro percepito, e quanto essa sia in grado di incoraggiare le interazioni a due mani.

Poiché ABaToGI è stata progettata per i display pubblici, l'interfaccia è stata anche inclusa in un'installazione pubblica per essere valutata sul campo. I risultati di questo studio (e di quelli precedenti) sono stati quindi riassunti al fine di sviluppare una serie di linee guida per lo sviluppo di nuove interfacce touchless a gesti basata sull'uso di un Avatar.

La tesi si conclude con alcuni spunti di ricerca per il futuro.

Acknowledgements

During the last three years, I had the pleasure of working with several great persons, who contributed to my research and my personal development. I would like to thank all of them, but as it would be impossible, I sincerely apologize to everybody I will miss to mention.

First, I would like to thank my supervisor, **Prof. Antonio Gentile**, for having let me embark this hard path to the PhD. Among his numerous and precious advices, I particularly thank him for having been always clear and transparent with me.

I would like to express my sincere and special thanks to **Salvatore Sorce**, who shared with me a plethora of fundamental advices, trends and experiences on doing research in Italy (and in particular Palermo). In addition, I strongly thank him for having become a friend, probably more than a colleague.

Second, my sincere thanks go to **Dr. Alessio Malizia**, who supervised me during my experience at Brunel University London. His "method" helped me a lot in understanding how to do research in HCI. Moreover, I thank him, as well as his PhD student **Tommaso Turchi**, for the coffees, beers, lunches and dinners we had together in London, as well as for having become two dear friends of mine.

I would also like to thank all the other colleagues from the CHILab and USI group at the University of Palermo: **Fabrizio Milazzo**, **Giuseppe Tirone**, **Andres Camilo Duarte Eraso**, **Wilson Arteaga**, **Arianna Pipitone**, **Marta Paliaga**, **Francesco Gugliuzza**. All of them contributed to my research with advices and hints, as well as to brighten my days spent at the University through terrific coffee breaks, lunches and the countless jokes that still make me laugh.

Further thanks go to all the other persons who collaborated with me during these years: **Maurizio Schifano** (for his help with the GUIs design), **Dario Pirrello** (for his work on the HIG-based interface used for a study described in this thesis), **Cristina Enea** (for having helped me in collecting users' opinions) and **Alessandro Bruno** (for his precious work on our co-authored paper).

I want to thank also my family: my mother **Carla Lombardo**, my father **Agostino Gentile** and my brother **Giacomo Gentile**. They have always supported me, not only during the PhD but, in general, during my whole life. Thanks for having helped me during prior studies, for encouraging me in my choices and for having never trust on my tireless pessimism.

Last, but certainly not least, my wife **Vita Bonfiglio** probably had to struggle more with all these PhD-related stuff than I had to. Your support and love helped me more than anything else in getting through a lot of hard times. I sincerely apologize for all the time spent talking about computer science stuff, the weeks of travels, and having been away from you for a whole year. Thank you for always standing by me.

Table of Contents

List of Figures

List of Tables

Glossary

ABaToGI	Avatar-Based Touchless Gestural Interface
API	Application Programming Interface
FOV	Field of View
GUI	Graphical User Interface
HCI	Human-Computer Interaction
HIG	Human Interface Guidelines
LCD	Liquid Crystal Display
NASA	National Aeronautics and Space Administration
NUI	Natural User Interface
PC	Personal Computer
SDK	Software Development Kit
TLX	Task Load Index
UI	User Interface
WIMP	Windows-Icons-Menus-Pointer

Chapter 1
Introduction

In the last twenty years, after the visionary ideas of Weiser about ubiquitous computing [1], a lot of research contributions have been implemented and transformed into real products, deployed and used outside of controlled environments, such as laboratories or ad-hoc experimental test beds. Moreover, technological innovations have allowed for installations of interactive displays in private, semi-public and (more in line with the scope of this thesis) public places, like fairs, shop windows, malls, workspaces, and public institutions. In most cases, interactivity is usually implemented by equipping displays with touch sensing, whereas cameras are less often used. This means that new technologies, e.g. Kinect-like devices [2] and the related features, are still not commonly adopted, especially in integration with commercial public displays. Therefore, interactive displays often do not exploit all the interaction possibilities provided by cutting-edge technologies available today. Furthermore, other interactive solutions may allow for overcoming issues that cannot be solved by traditional touch-based solution, e.g. the need for keeping a high hygiene level of the screen surface, or the possibility of providing wall-sized displays with interactive capabilities. Nowadays there is the opportunity to integrate Kinect-like devices, as well as other new technologies, in artifacts like public displays in order to implement touchless (or mid-air) gestural interactions.

By using such novel interaction techniques, it is possible to imagine and design for new scenarios: users will interact with wall-sized or remote displays, and they will be able to use gestures to get information from them, independently from their abilities (e.g. impairments, such as wheel-chaired people). However, the main problem to be tackled is to find a valid design methodology for gestures and interaction modalities. Moreover, in order to make a product effective in a wide range of social settings, studies cannot be conducted only in controlled environments, but they should take place directly “in-the-wild” (i.e. in appropriate social contexts where public displays are typically deployed). This is one of the few ways in which users’ behavior can be observed by taking into account different context factors, and how they influence users’ attitudes.

The main goal of this thesis is thus to describe the most relevant problems and to propose possible general solutions in the process of design, implementation and deployment of touchless gestural interfaces. This document will refer to a case study used to test and validate the design choices. It is based on a graphical interface to enable gestural interactions with information provision systems. As those systems are usually deployed in public [3], the terms “information provision systems” and “public displays” will refer to the same artifacts in the rest of this document, as a sort of metonymy.

In the following Chapters, several studies (and the resulting outcomes) are described, and all of them contributed to the definition of a set of guidelines for the design of touchless interfaces that use gestures as a mean for interacting with displays in public.

The rest of this Section is divided into three parts: the first one introduces the main research questions addressed in this thesis; the second one provides a brief overview of the methodology adopted during this research work; the third one describes the contents of the thesis, and how they are organized in the following Chapters.

1.1 Research Questions

In order to design a touchless gestural interface for public displays, there are several issues that need to be overcome. Most of them are described in detail in the following Sections; however, a brief description is provided here too, in order to introduce the main research questions addressed in this work. Moreover, an overview of the Research Questions tackled in this document is available in Table 1.

Table 1. Summary of Research Questions.

Research Questions		Chapters
R1	Is the presence of an Avatar that replays user movements able to communicate the touchless gestural interactivity?	2; 4
R2	Do users need an activation gestures while interacting with a touchless gestural interface?	2
R3	Is the presence of an Avatar that replays user movements able to foster bimanual touchless interactions?	3
R4	Is there a correlation between the perceived mental workload and the presence of an interactive Avatar, while interacting with a touchless gestural interface?	3

In the starting phase of this research, a prototype of an Avatar-based Touchless Gestural Interface has been designed (named ABaToGI), and its capabilities have been tested with users. In particular, this interface allowed users to interact with no specific activation gestures. The aim of the initial exploratory studies was basically to understand users' needs while interacting via gestures (e.g. if they miss some sort of "click", as no activation gestures were supported), as well as the capability of the interface to overcome the *interaction blindness* (i.e. "the inability of the users to recognize the interactive capabilities of those displays" [4]). These issues are summarized by the research questions R1 and R2.

During the aforementioned studies, some users seemed to interact more frequently with both hands when an Avatar was displayed. This intuition acted as a basis for several new potential research questions (summarized by R3): is the presence of the Avatar able to foster bimanual interactions? Do users interact with two hands because this seems to be more "natural" to them? Do users perceive bimanual interaction as a usability problem? Most of these issues are addressed in the following Chapters.

After some studies conducted in controlled environments, the effectiveness of ABaToGI (which, in the meanwhile, has been improved according to users' opinions and feedbacks) has been tested in-the-wild. The main goal of this study was to obtain additional knowledge in order to summarize the lessons learnt in a set of general guidelines for the design of touchless gestural interfaces, with a special focus on Avatar-based ones.

1.2 Methodology

The research questions introduced in the previous Section have been elicited and then addressed by following the three cycles of design science research presented by Hevner in [5]:

- the *Relevance Cycle*, which includes the identification and representation of opportunities and problems related to an actual application context. The latter "not only provides the requirements for the research [...] as inputs but also defines acceptance criteria for the ultimate evaluation of the research results";
- the *Rigor Cycle*, which "provides past knowledge to the research project to ensure its innovation";
- the *Design Cycle*, which uses knowledge, opportunities and problems from the aforementioned cycles to iterate between the two basic activities upon which the design science is built: build (e.g. prototypes, design artifacts) and evaluate [5] [6].

According to this three-cycle view, building and evaluating artifacts (i.e. Design Cycle) may actually produce new problems or opportunities (i.e. Relevance Cycle), which in turn may require new knowledge from the available literature (i.e. Rigor Cycle). This cyclic process is represented in Figure 1.

The contents presented in this thesis have been built upon this view. The preliminary literature review conducted before the research project presented here has been used both for identifying problems and opportunities and for acquiring the required knowledge on the domain. A concept map of the main keywords is available in Figure 2.

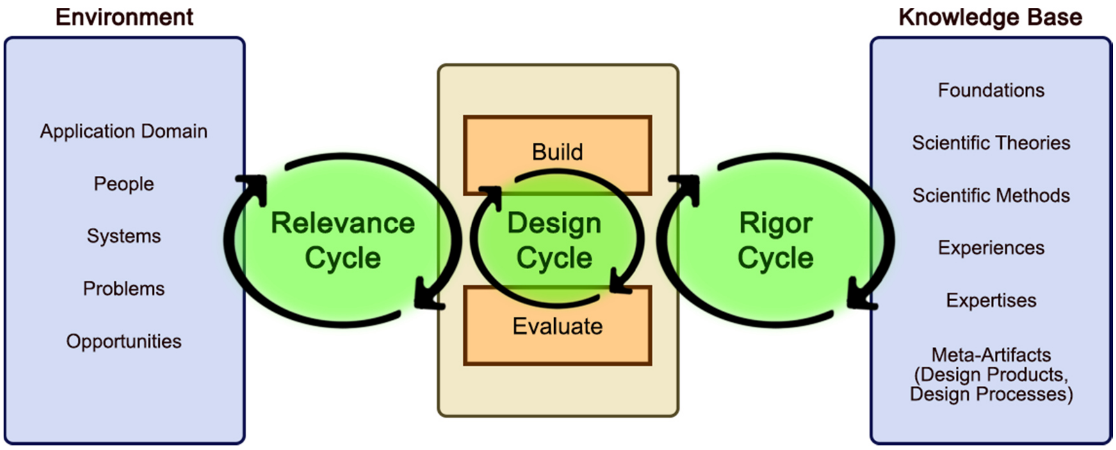


Figure 1. The three-cycle view of design science research.

1.3 Thesis Outline

After this brief introductory Section, the thesis includes three main Chapters:

- Chapter 2 describes the design process of ABaToGI, the studies conducted for collecting users' opinions and behaviors, and the results of a comparison study between ABaToGI and another touchless gestural interface;
- Chapter 3 describes how the presence of the Avatar in ABaToGI affects user interactions in terms of perceived workload, and if it may be able to foster bimanual interactions;
- Chapter 4 reports the results of observations of ABaToGI in-the-wild. A discussion follows, leading to a set of guidelines for designing touchless gestural interfaces for public displays.

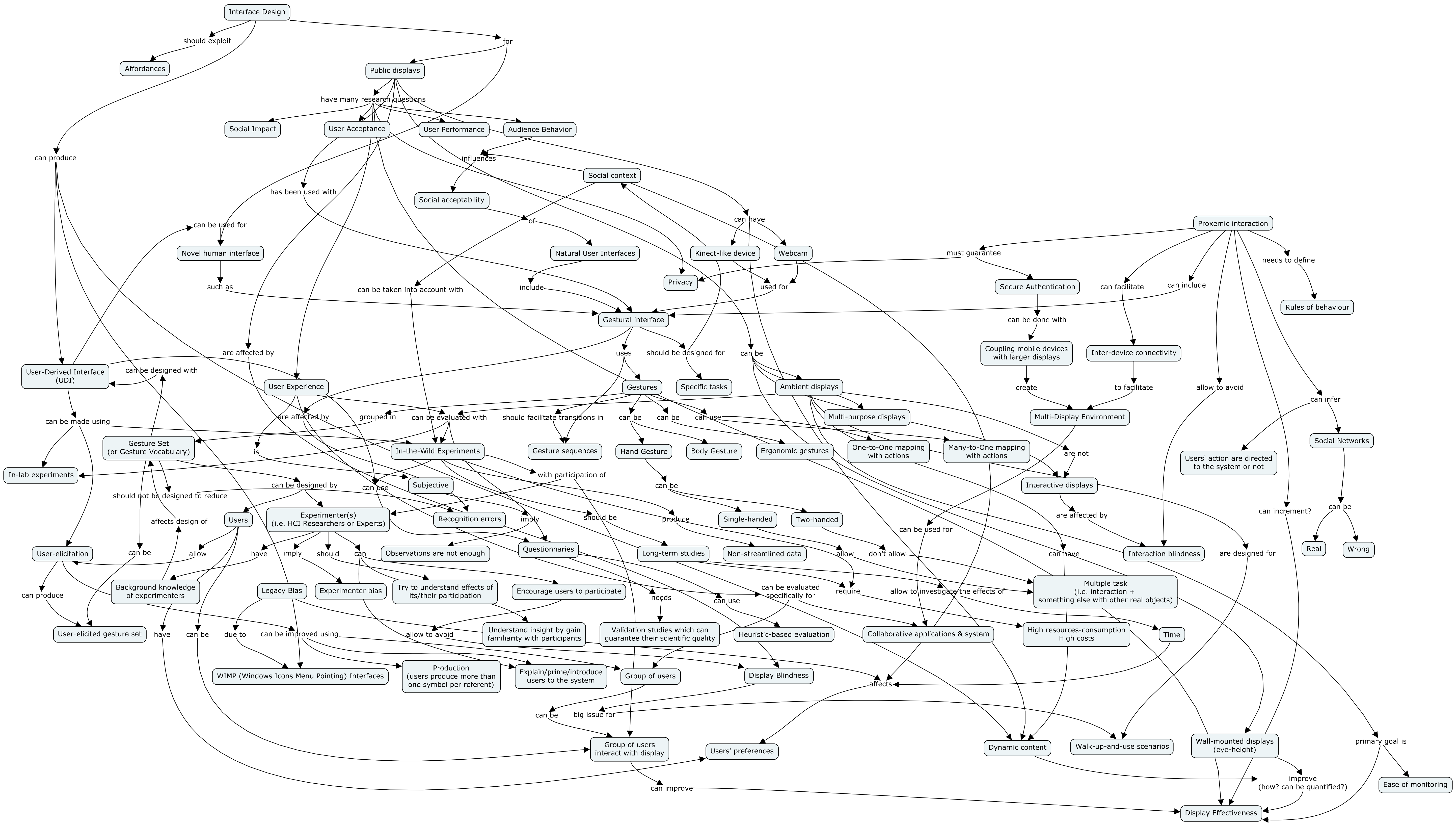


Figure 2. Concept map produced during the literature review (as part of the Rigor Cycle).

Each of the three aforementioned Chapters includes a first subsection to summarize the related works about the covered topics. Then, each Chapter includes at least an experimental study (with qualitative and/or quantitative data collection and analysis), followed by a discussion.

Then, Chapter 5 concludes the document, by summing up the results of the previous Chapters. Moreover, as this thesis provides a common ground for future research, this chapter will also summarize some possible future works, both in the short and in the long-term.

Chapter 2
An Avatar-based Touchless Gestural Interface

In the last decade, many authors have investigated and studied touchless and gestural interactions as a novel tool for interacting with computers. According to the definition by de la Barré et al. [7], "interaction is said to be touchless if it can take place without mechanical contact between the human and any part of the artificial system". This means that, for instance, interacting with a system using some controller, such as the Nintendo Wiimote or any other similar device, cannot be considered touchless interaction. On the contrary, eye trackers (such as Tobii EyeX and similar) or Kinect-like devices [2] have been widely accepted as valid examples of devices that enable for touchless interactions.

Recently, several authors proposed touchless interaction as a new way for interacting with public displays [8] [9]. One of the main advantages of this idea is the possibility of offering interactive solutions to users also if the display is placed in a non-touchable or non-reachable area. This approach can be useful, for instance, in order to provide wheelchair users with accessibility and/or to prevent vandalisms. Very large displays (e.g. media façades [8] [10]) can become interactive via touchless-enabled technologies too. Furthermore, whilst the increasing number of interactive displays in cities and other public urban places is contributing to change and improve the everyday life of people [11], it also result in a growing number of issues from users' point of view. For instance, users are becoming aware about bacterial contamination of touchscreens [12], which can be solved by enabling touchless gestural interactions.

The aforementioned scenarios, however, are often complicated by several typical issues of public displays and touchless interactions, e.g. the need for communicating interactivity [13], as well as for encouraging users to look at the display [3]. Researchers are still working on find solutions to all these issues, and this has contributed to a growing interest towards novel forms of interactivity in public display.

In this section, a novel interface for enabling touchless interactions with public displays is presented. After a brief overview of the related works, a description of the main functionalities of the proposed interface is provided, including the design process and the improvements added after a brief pilot study. Then, the proposed interface is compared with another one, based on the Microsoft Human Interface Guidelines (HIG) [14], which can be considered a *de facto* standard for applications developed using Microsoft Kinect devices[1]. The comparative study has been based on users' opinions, collected via observations and semi-structured interviews. As this was an initial (and thus exploratory) stage of the development, the choice of this form of opinions gathering provided for the exploration of topics in depth and breadth, letting also users to focus on what they found important [15]. The description of the aforementioned study is then followed by a discussion of the collected qualitative results.

2.1 Related Works

Nowadays public displays are almost everywhere. We can find them in squares, malls, and many other public places. Moreover, many of them are augmented with interactivity. Despite the wide adoption of touchscreens as main input (and output) devices, new interaction modalities have emerged to fulfill specific needs of public display systems. For instance, the increasing number of interactive *media façades*, defined as installations in which displays are integrated into architectural structures [8] [10], have implied the need of interacting from distance, and without any physical input device. Many authors proposed interaction methods based on detection of users' position and their body movements, as well as by using gestures or mobile devices (see for instances

[1] This statement is supported by the high number of existing interfaces based on the HIG, specifically for gesture-controlled games.

Aarhus by Light [16], Dynamically Transparent Window [17] and/or Climate Wall [18]).

However, such kind of interaction modalities is still rarely used in *situated public displays*. The latter term refers to smaller displays (size ranges from TV- to billboard-sized screens [8]), still placed in public spaces (both indoor and outdoor) that more often include touch-sensing features. Touchless gestural technologies have been less often studied for this kind of devices, probably due to the size and position of the screens that allow for interactions by touch. However, despite the widespread of touch-based technologies for enabling interactivity on situated public displays, some authors have investigated touchless interactions. In some circumstances, applications are very specific (see [19] and [8]), and it is difficult to design interfaces by following any kind of "standard" guidelines. This is the case of games and other forms of entertainment systems. For information provision systems, the definition and application of general guidelines may be more straightforward. In addition, they may result in different systems with the same interaction paradigm. Several general-purpose applications for public displays have been proposed by many authors, and quite a lot of them are based on the Microsoft HIG. For instance, most of features and controls used in the gestural system proposed by Cremonesi et al. [9], are implementations of HIG. Similar ideas have been adopted in [20] and [21].

More generally, in order to design a gesture-based graphical interface, one of the main problems to be solved is related to the question: "Which gestures should the interface support for each task?" In order to address such problem, several authors proposed to define proper gesture sets, conducting what have been referred as *gesture elicitation* studies [22]. In these works, researchers asked participants which gesture they would prefer to accomplish a specific task (e.g. scrolling a sliding photo gallery, or moving a given object on the screen). By asking participants to perform their favorite gestures, it is possible to create a user-defined gesture set, and therefore a user-derived touchless gestural interface. This means that the resulting gesture set fits the users' preferences as much as possible. Of course, the more accurate the participants sample, the more guessable the gestures will be. Although gesture elicitation studies have been mostly conducted for touch-based interfaces, recently some authors have documented touchless gesture sets elicited by users (see for instance [23] and [24]).

The main drawback of an elicited gesture set is that it will never include all the potential gestures that users will perform to interact with the system: without a training or some other mechanisms, during the starting interaction phase people will just try to interact "somehow", by using what they assume to be the right gesture. With a bit of luck, their gestures will be recognized (if belonging to the gesture set). If they are particularly lucky, their gestures will immediately produce the correct results (i.e. the linked task or action). Without training or explicit suggestions, however, it is quite impossible to avoid errors, especially during the initial phases of interactions.

Moreover, elicited gesture sets are also affected by another non-obvious issue. We all know that if a user sees an "X" on the top-right corner of a square-shaped item, she will feel natural to click on such icon to make the item disappear. But even if there is no "X", the user will try to find out where it could be instead of trying something completely different from the conventions used in the Windows-Icons-Menu-Pointer (WIMP) paradigm [25]. This behavior has been defined as *legacy bias* [26]: researchers observed that many user-defined gestures were manifestly legacy inspired and that those gestures turn out to be quickly guessable and learnable. The problem here is not in the legacy bias itself, but rather in the fact that it is not possible to avoid it: "Despite presenting participants with a large multitouch touchscreen without UI elements from traditional PC interfaces, most participants suggested mouse-like single-point or simple-path gestures" [26]. This implies that, in order to design touchless gestural interface, designers must take into account this bias, either to exploit its perceived naturalness when designing the interface [27], or to design an interface that must be robust to this issue.

Besides the applications and design guidelines, there is also the need of studying touchless interactions by keeping in mind other peculiar issues of public displays. First of all, experimenters must take into account all the influencing factors of a public place. The best results may be achieved conducting the study *in-the-wild* [28] (see Section 4.1), although this increases costs and efforts in terms of time and set up challenges. Furthermore, researchers must take into account issues like the need of overcoming the *display blindness* [29] and – probably more complicated – the *interaction blindness* [30]. The first term has been coined by Müller et al. in [29], and it - similarly to the *banner blindness* in web pages [31] – causes users not to look at the displays because of their prejudice about the content, which is expected to be an advertisement. Researchers must thus overcome this issue in order to study any aspect related to the interactions. Among the proposed solutions for attracting passers-by glances, visual animation effects and/or sounds have been demonstrated to be helpful in this direction [32]. More in detail, according to Müller et al., other factors that can mitigate display blindness are the colorfulness, the amount of time the display is potentially visible to passers-by, and the display size [29]. However, this problem is not simple to solve and can require applying some techniques from the persuasive computing area.

If the research is focused on issues related to interactions (as in this thesis), overcoming display blindness may not be enough. As noted by Ojala et al. in [30] indeed, even when users notice the display they often do not interact with it “because they simply do not know that they can”. This means that there is the need to explicitly communicate the interactivity in order to entice interactions. *Interaction blindness* (as this phenomenon has been called) was noticed also by other authors working on interactive public displays [20], and it generally refers to the inability of the public to recognize the interactive capabilities of those surfaces, also when looking at the display. Among the many solutions described in literature, one of the most commonly adopted is the use of explicit visual clues that ask users to perform some gesture. This approach has been described and evaluated in [33]. In this work, authors compared different presentation modes for such visual clues, i.e. integration, temporal division and spatial division. This study showed that spatial division results to be the most suited solution for public displays, although it implies the need of allocating part of the screen for showing such clues.

In [30], Ojala et al. suggest that "one way to overcome interaction blindness and entice interaction is to make the interface more natural. Proxemic interactions are emerging as a potential paradigm for realizing natural interfaces [...], but our simple visual proxemic cue [...] (the "Touch me!" animation) did not noticeably increase user interaction" [30]. *Proxemic interactions* were introduced by Ballendat et al. [34], and they are very related to (and actually based on) a previous work by Vogel and Balakrishnan [35]. In [34] and [35] authors propose systems that react on user's position and orientation, i.e. without any implicit interaction. Such idea seems promising in solving interaction blindness because users can easily see the interactivity of the display if its contents change in correspondence with their movements. Indeed, proxemic interactions allow the implementation of more sophisticated solutions than a simple "Touch me!" animation, and there is the need to better investigate how they can help to solve interaction blindness.

Moreover, proxemic interactions can help users to understand the features of an interactive public display, by modeling it as a sort of mirror (i.e. one of the four mental models proposed in [36]). The mirror mental model has been shown to have a strong potential to catch users' attention [36] [37], which suggest to use it also as a partial solution to display blindness, in addition to interaction blindness. A successful application of the mirror mental model is MirrorTouch [38], where authors studied the use of touch-based interactions combined with mid-air gestures. In this application, a user interacted with her silhouette shown in a public display, and this showed how effectively the mirror model communicates the touchless interactivity. Indeed, authors underlined the need of explicit call-to-action as the only effective way for letting users interact via touchscreen, instead of sticking on the gestural interaction modality only.

Furthermore, Müller et al. [19] studied the so-called *remote honeypot effect*, which can be observed when multiple public displays are interconnected.

Authors noticed that if the silhouette of a user who interacts with a display A is also shown in another display B on the same network, users in front of display B are encouraged to interact, guessing the interactivity of the display. Other mechanisms are based on giving users explicit indications about the display interactivity, e.g. using introductive video tutorials or posters. Moreover, in [39] authors show that displaying users' silhouettes may help in communicating display interactivity to passers-by, and this idea have been explored by other authors [40] [41].

2.2 ABaToGI: an Avatar-Based Touchless Gestural Interface

The discussion above implies the need for paying particular attention to some of the main problems that must be faced in the development and deployment of public interactive displays. The first one is the interaction blindness: it is crucial to build interfaces that are able to communicate their interactivity, as well as the specific kind of interactivity supported by them (which in this case is touchless-based).

The second problem arises from the need for novel visual interfaces expressly designed for touchless gesture-based and natural interactions, in order to outdo the WIMP paradigm commonly used in desktop-based systems. In other words, the designed interface should be also robust against legacy bias.

To address these issues, this Section presents a novel interface designed using only *in-air direct manipulation*, as defined in the following. According to [42] and [43], one promising solution to implement interactions that are more natural is the use of *direct manipulations*, instead of *symbolic gestures*. Such paradigm, however, is appropriate in touch-based or tangible systems, where "touching" actions allow for the direct manipulation of objects in the interface. This paradigm could be extended to touchless interfaces, thus becoming what here is referred as *in-air direct manipulations*. By means of body movements and in-air gestures, it is possible to imitate the direct manipulation of an object, as we would do in the real life, without actually grabbing or touching them.

To support this choice of using in-air direct manipulations, besides the above considerations, it is important to cite the similarities with *dontclick.it* [44], a website that allows its users to browse contents without the need of a single click (see Figure 3). In dontclick.it it is possible to open sections, select items and animate objects, just by moving the mouse over them and without pushing any buttons. Interestingly, the statistics of the website show that the majority of users do not miss the click. By observing dontclick.it, it can be noticed that if it is possible to interact naturally with a web page without a single click (that is an activation gesture for that interface), it should be possible to interact even more naturally with a touchless interface without any activation gesture. Indeed, it should be more difficult to avoid the use of the 'click' having a mouse in a hand, rather than avoid any activation gesture having nothing in the hand. In other words, the design of the proposed interface has been based on the hypothesis that in-air direct manipulation will improve the naturalness of touchless interactions.

Figure 3. Two screenshots from *dontclick.it*.

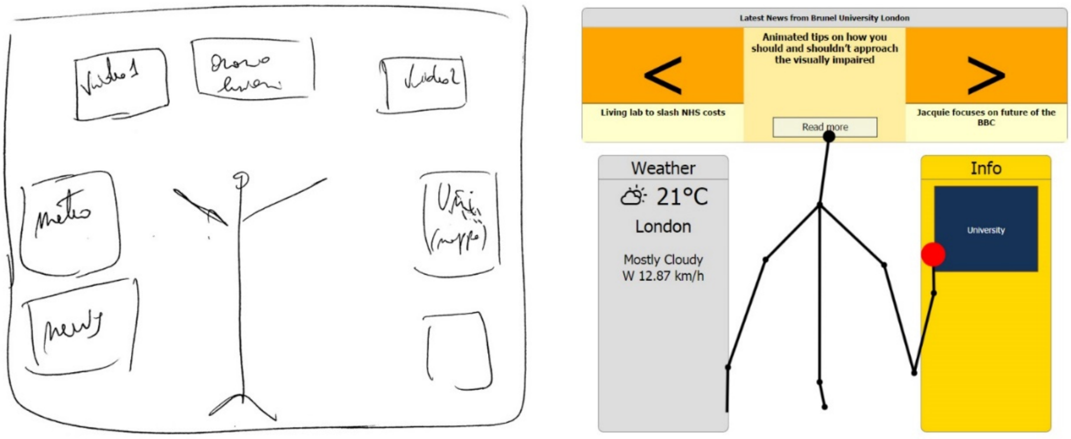


Figure 4. Mock-up and first implementation of ABaToGI.

2.2.1 Interface Description

The design choices aim at the definition of basic guidelines to overcome mainly the interaction blindness, and secondary the legacy bias.

The first proposed layout consists of an Avatar (i.e. just a stick man in this first prototype) placed in the middle of the screen, with all the other interface components arranged all around it (see Figure 4).

The Avatar appears whenever a user approaches the display, and remains permanently present in the middle of the screen, continuously replaying user's movements. As stated in the previous section, relevant works demonstrated that the presence of a predominant entity that continuously reproduces user movements significantly contribute to reduce the interaction blindness [4] [39] [41].

In this interface, a user can trigger the interaction events just by driving the Avatar's hands and placing them on top of the available tile-shaped components – with no activation gestures. As soon as an Avatar's hand (here represented as a red circle, as shown in Figure 4) enters inside the area of an interactive tile (e.g. the blue one on the right side), an event is immediately triggered and an animation acts as visual feedback for communicating the transition between two different pages. In other words, the user can interact with the interface using only in-air direct manipulations (i.e. without specific activation gestures), allowing designers to avoid the adoption of symbolic gestures to trigger events. A user should then better guess and learn how to interact by herself, since there is no need of any training about specific activation gestures.

Furthermore, the prototype also allows for using the so-called *grip gesture* (i.e. closing the hand into a fist [14]) to trigger specific (but not primary) actions, e.g. for zooming and panning. Although it is an activation gesture, this additional feature can help users affected by legacy bias to interact more easily with the interface.

Supported Tasks
By means of the interface, users were able to:

- read weather information (from the main page, see Figure 4);
- read general information about the University (Figure 5);
- read and navigate some news about the University (Figure 6);
- access and navigate a University campus map (Figure 10).

The last three tasks listed before can be achieved by accessing specific sections. All of them are accessible from the main page and, as stated before, once a user drives an Avatar's hand entering on an interactive tile, an animation is executed in order to communicate to the user the transition from a section to another one. All the animations last less than 1 second, and they consist in zooming to the center of the interactive tile, and the showing the content of the new section. Each section includes a "Back" button to return to the main page (see Figure 5 and Figure 6): in this case, the animation is simply reversed (i.e. zoom out from the section to the main page).

The following figures show the layout of the aforementioned sections and how they appear. Obviously, in this preliminary phase the interface looks simple and sketched, as the goal is to gather users' feedbacks about the interaction only.

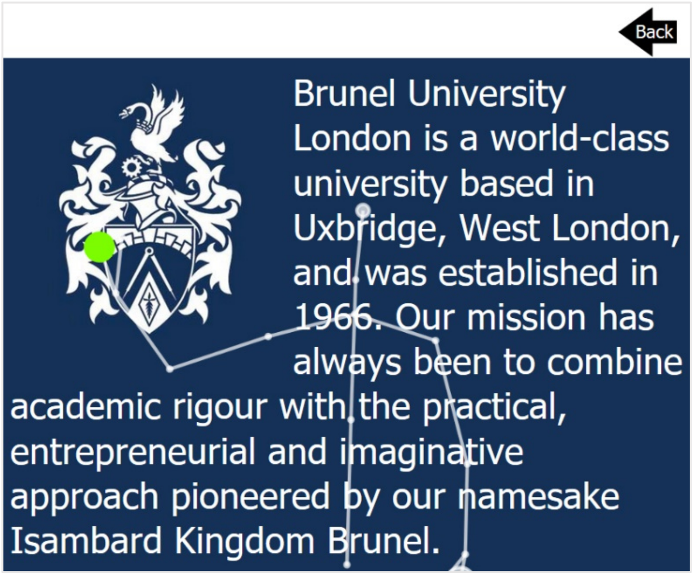


Figure 5. The University information page, using a semi-transparent Avatar.

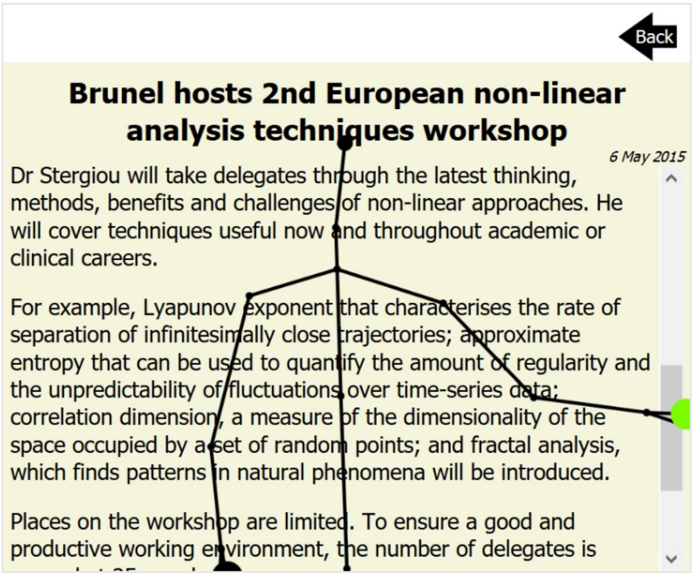


Figure 6. Layout of the news page used for the first prototype of ABaToGI.

Figure 7. A user interacts with ABaToGI: on the left, the user is interacting with the main screen, while on the right he is zooming a map.

2.2.2 Pilot Study

In order to collect and evaluate users' opinions on ABaToGI, a first pilot study was conducted with the very first prototype. The interface was developed as an information provision system, ideally designed for being deployed in a University foyer. The typical audience for this situation is then composed mostly of students, as well as (although in minor part) University staff members and lecturers.

This first user study represented a first attempt to answer the following questions:

- Is the Avatar able to let users guess the (touchless) interactivity in order to outdo the interaction blindness problem?
- Is this interface robust to legacy bias?
- Do the users feel comfortable while interacting via touchless gestures?

This pilot study was also aimed to find and fix bugs and usability issues, according to users' feedbacks. The study was conducted with 12 users, both researchers and students from HCI and other computer science areas. HCI researchers have been chosen in order to gather informed hints and suggestions about experienced issues, and three of them were previously informed about the touchless interactivity of the system. Obviously, students' opinions were still crucial as they represented the typical users of the system (which was developed to provide University-related information).

All users performed 10-minutes-long single-user interaction sessions: each of them was asked to find some specific news first, without any suggestions or hints on how to achieve the goal, especially in terms of way of interaction. Then, they were asked to switch to other tasks (e.g. find information about the University, try some free interaction), in order to observe their behavior.

At the end of each interaction session, a semi-structured per-user interview was conducted (see Appendix A for more details). In the following, the term "aware users" will refer to users that already knew something about the touchless nature of the system, while "unaware users" is used for all the others.

Since at this stage the focus was basically put on the interactions and the interface, this pilot study was conducted in the lab, for fast collection of users' feedbacks (at the cost of a lower ecological validity [3]).

The hardware used consisted of a projector (for turning a wall into a display) and a Microsoft Kinect sensor (clearly visible to all users) on a table, directly under the projected interface (see Figure 7). Both the projector and the Kinect were connected to a laptop, on which the whole business logic of the system was run.
The interface was composed of an interactive news slider on the upper area of the screen, a non-interactive weather widget on the left, and two smaller interactive tiles on the right to access some information about the University (see Figure 8).
The rest of this Section presents some interesting findings derived from the pilot study.

The Avatar Communicates Touchless Interactivity
During the interviews, unaware users were asked if they were able to guess that the interactivity of the system was based on touchless gestures and actions, and all of them asked that they could. In particular, all of them explained that their ability to understand immediately the kind of interactivity was quite

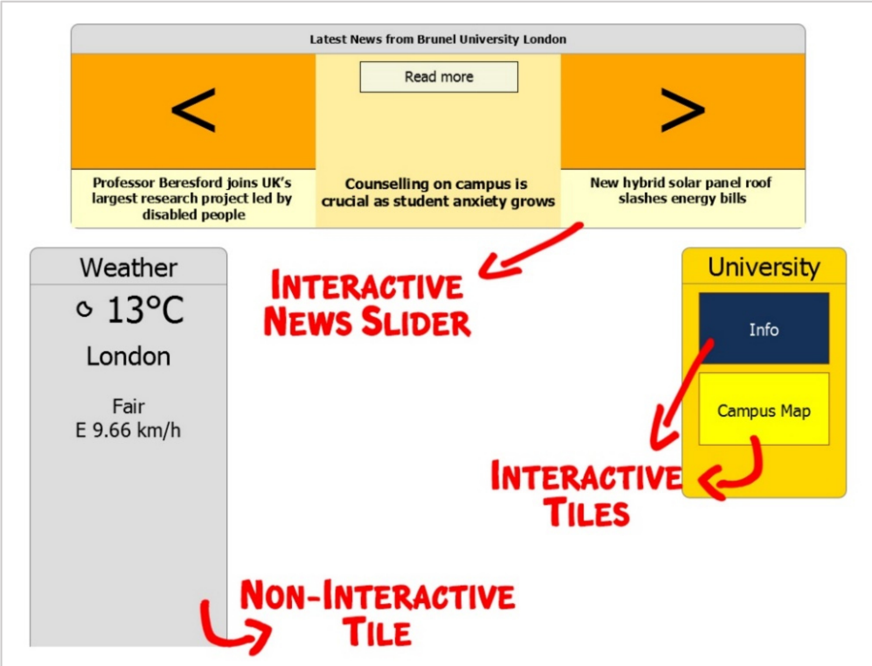


Figure 8. Basic structure of the first prototype of ABaToGI.

straightforward, due to the presence of the "repeater" Avatar in the middle of the screen. One user told that "seeing the little man suggested me that gestures were supported"; other interviewees used words like "stick man" or "puppet" to indicate the avatar, and all of them described it as the main interface component for explaining the touchless interactivity of the display. Unaware users were also asked about previous experiences with touchless gestural systems; however, no connections were noticed between previous experiences (that however were quite poor) and the ability to perceive the touchless interactive nature of the system.

These findings are in line with what Müller et al. described in [39], i.e. the importance of using user's silhouette that continuously mimics user movements. Putting an avatar in the middle of the screen, and use it for the whole interaction process, seems to serve as a valid hint for explaining to users the touchless interactive nature of the system.

Perceived Limitations

Although the use of an Avatar seems a good solution in order to communicate the touchless interactivity, its presence implies some constraints that should be taken into account during the design stage.

First of all, the Avatar limits the space available in the interface for interactive and non-interactive tiles. In particular, it is important to avoid positioning interactive tiles on the bottom part of the screen, as well as on the center of the screen. Indeed, the transition between a standing pose (i.e. user standing with her arms by her side) and any other pose aimed at interacting with a tile in the middle-upper part of the screen would require the user's hand to pass over tiles positioned in the bottom part of the screen. In fact, this has been also observed during user tests. Because the only way to avoid unintentional activations would consist in unnatural movements, only non-interactive tiles should be positioned

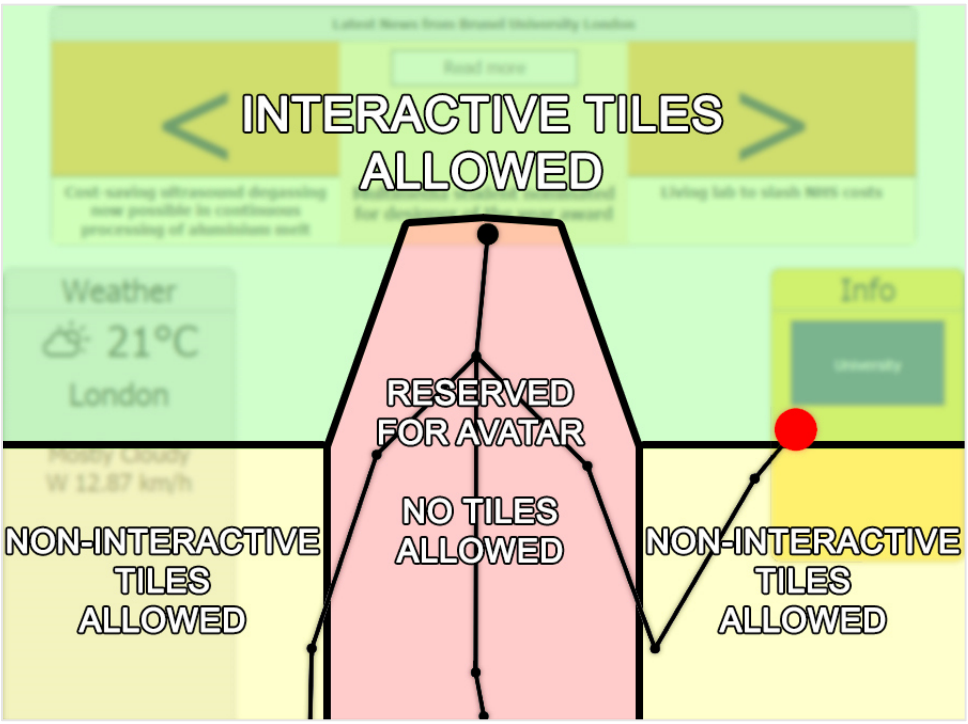


Figure 9. Tile positioning scheme.

at the bottom, while interactivity should be supported only in the upper part of the interface.

Moreover, because the Avatar would occlude the center of the screen, tiles should not be positioned behind, but instead all around it.

Figure 9 shows a visual scheme that helps in understanding how tiles should be positioned, according to the aforementioned considerations.

These constraints have been taken into account during the initial design stage. During this pilot study, some users seemed able to understand by themselves the reasons behind this apparently unconventional tile positioning. Furthermore, they perceived some difficulties in avoiding unintentional activation of interactive tiles. The latter issue is due to the so-called *live-mic problem* [42], and will be discussed later.

Click without a Mouse
As previously stated, *legacy bias* affects users, particularly when they interact with a new system. When designing a touchless interface for public displays, avoiding this problem is not trivial (if not impossible), since WIMP is deeply rooted in users' habits after years of classical interaction. In order to investigate how ABaToGI addresses this issue, all the participants have been observed during their interactions with our system, to see if they were affected by legacy bias and to assess if the proposed interface was robust to this issue or not.

The first noticeable result is that the "click" bias seems to be very rooted among people. Indeed, although interface was designed to have no activation gesture and, obviously, no mice and keyboards, during the interviews several users often referred to a "click", both in the answers and during the interaction. For instance, one user asked "what happens if I click here?", and another said "if I click here, something is strange". Similar sentences were uttered by HCI researchers and NUI experts too, despite their knowledge in the field. This can somehow be considered as a sign of the strength of legacy bias, which can result to be very hard to avoid.

The observed behavior of a participant seems to be a good point if related with the robustness against legacy bias. When this user tried to activate the task associated with one tile, he also rapidly closed and opened his hand, as it were a sort of his own activation gesture. This happened even after having expressly told him that no activation gestures were needed to trigger the task. Despite this unnecessary gesture, his interactions with the interface worked fine and as expected, and the user assessed that he liked the system.

These encouraging, although very preliminary, results suggested the need for a deeper investigation, in order to obtain useful suggestions for the design of post-WIMP touchless interfaces that are capable of accommodating both legacy biased and non-biased people.

The Live-Mic Problem

In their *Brave NUI World*, Wigdor & Wixon describe "the always-on nature of in-air gesturing", in the sense that the gesture recognition capabilities of hardware devices and software recognizers are continuously active. They refer to this issue as the *live mic problem* [42]. In general, this means that there is the need to "differentiate physical actions that are intended to drive the computing system from those that are not".

During the trials, few users experienced one of the most common issues due to the live mic problem. Sometimes, users' movements caused some unexpected visual outputs on the interface, resulting in a sort of touchless extension of the *Midas touch problem* [42] (i.e., in the context of touch-based interfaces, the unwanted activation of interactive objects because of the always-on nature of touch input devices). This happened because all the actions are triggered by hovering a hand cursor over the corresponding tile. For instance, if during a movement toward a tile A the hand cursor passes over another tile B placed along the path to A, the activation would be triggered on B even if it is not the user intent.

Such issues could be reduced by correctly structuring the interface layout. In order to verify if this opinion was true, users have been also asked to interact

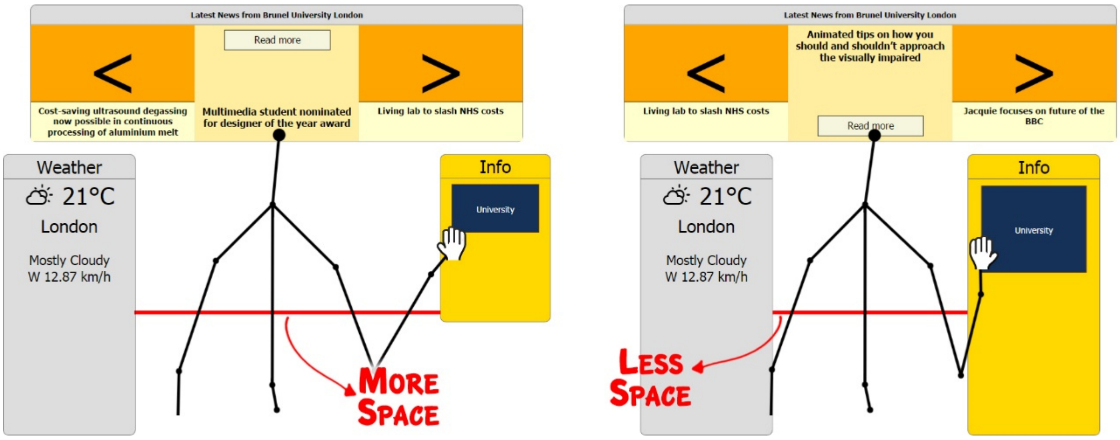


Figure 10. Visual comparison of two versions of ABaToGI: far vs. near tiles

with two slightly different versions of the same interface: the first version had some tiles placed nearer to the Avatar, whereas in the second one the same tiles (with the same functionalities) were placed more distant from the Avatar (see Figure 10).

As expected, users confirmed that using more spaced tiles resulted in a reduction of the Midas-touch/live-mic problem. Even if the issue was not completely solved by the different layout arrangement, this provided a useful hint on how to proceed in the next steps. Indeed, a possible and probably definitive solution seems to be the introduction of an appropriate time interval (e.g. 100 msec) before the hovering over a tile may trigger the corresponding task.

Other Findings

All the users explained that they liked the interface, and many of them assessed that they "felt in control" while interacting. Although this may be a direct consequence of the experimenter bias [28] (i.e. the unconscious user's attitude to adapt his/her behavior to a – presumed – experimenter's desired result), generally users' feedbacks seemed to be encouraging for future studies.

The pilot tests have also represented the opportunity to evaluate how users' behavior is affected by the form factor of displays. The projected display has been compared with a laptop version of the same interface. All the users told us that our interface was more suitable for big displays, mainly because of the text size (it were difficult to read it on a smaller display), but not just that. For instance, a user assessed that he felt dizzy after interacting for approximately ten minutes with the laptop version. One of the reasons for this unexpected effect might be the smaller form factor. Although just one user had this feeling, further investigations may help in understanding this phenomenon.

Another interesting consideration is about users' preferences on interaction modalities. At the end of his session, only one user asked why it was better to interact by touchless gestures rather than using a classic touchscreen. This question was made after a discussion on how this prototype can be used, and in which context it might be deployed when ready. It is evident that touchless interaction represents a useful (and in some case the only possible) solution for adding interactivity to public displays, and also an opportunity for brand new ones. For instance:

- wheelchair users may gain interactive access to information;
- displays may be placed in non-reachable (i.e. non-touchable) positions thus providing interactivity while avoiding vandalisms;
- big wall-sized displays can be made entirely interactive despite the fact that certain areas are not in the range of users' arm for a touch-based interaction.

2.2.3 Interface Improvements

After having collected and analyzed users' feedbacks, in this Section the issues previously discussed (along with other minor corrections), are further described and addressed in terms of improvements to the interface.

The first issue noted during the tests was due to the representation of the hand joints, depicted as colored circles. The color of such circles was red if the hand was closed, and became blue when it was open. However, because of the low resolution of depth images provided by the Kinect sensor, the hand pose recognition algorithm was not precise enough for letting users understand the

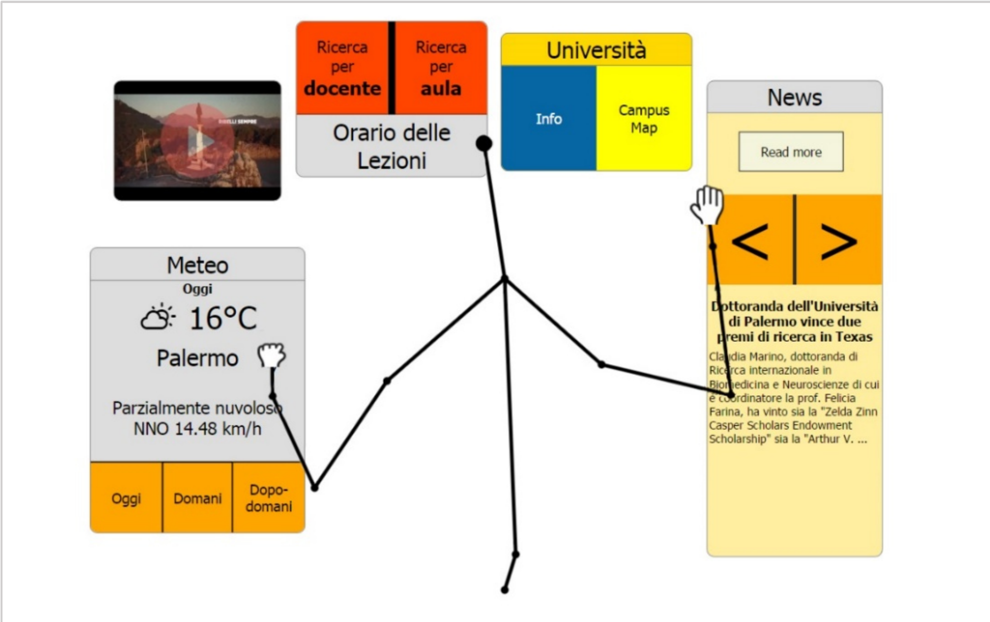


Figure 11. The layout of the improved version of ABaToGI.

meaning of such colors. The direct consequence was that users simply ignored them.

Another significant problem was about the usability of the interactive news slider. Several users were not able to understand how to use it, and they often faced with the Midas touch issue described above.

Such feedbacks (together with other minor requirements) have been used for implementing an improved version of ABaToGI (see Figure 11). In the new version, the Avatar's arms end in two hand-shaped cursors (similar to the Kinect Cursor described by Microsoft in [14]), which represent and replay the user's hands movements.

The news slider was modified too, in order to place it on the right side of the layout and also to reduce the aforementioned Midas-touch issues. Its new layout is depicted in Figure 11.

The interface still allows users for *opening sections* by placing the hand-shaped cursor on top of the available interactive tiles, exactly as described for the previous version. Users can also explore and read texts (see Figure 12), using two modalities:

- by placing the hand-cursor on top of the up-and-down arrows-shaped icons (no activation gestures);
- by using a grip gesture to "grab and scroll" the page vertically (activation gesture for legacy-biased users).

In this new version of ABaToGI, users can still manipulate images, by imitating what people usually do with touch-based systems (except that they use mid-air gestures): users can close both hands into fists to "grab" a document (e.g. a map), and then stretch, pan or rotate it according to distance and angle variations between the fists (see Figure 7).

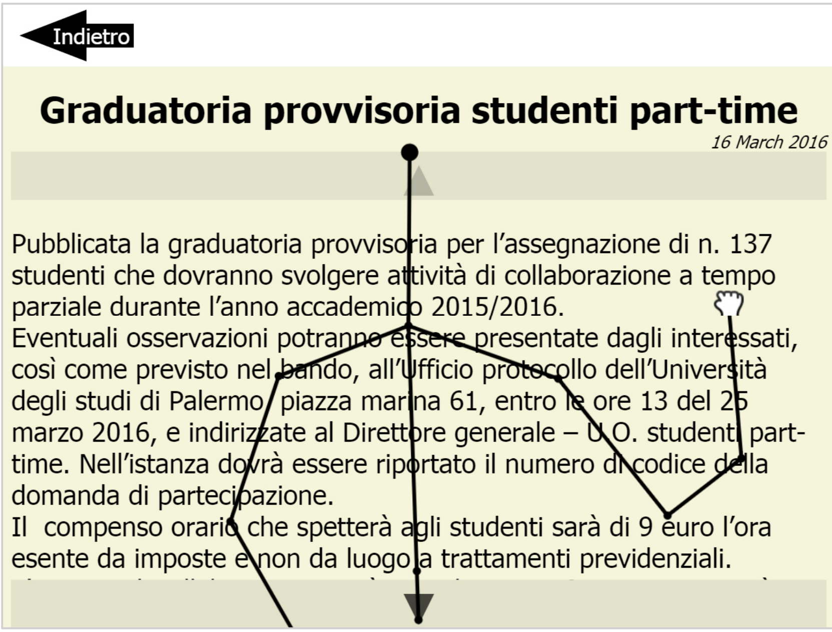


Figure 12. ABaToGI allows for exploring and reading texts by using "grab and scroll" gesture, or by placing the hand-cursors on top of a tile.

It is also possible to play multimedia content by using a tile for starting/stopping the reproduction. Furthermore, we implemented the *search* feature (see Figure 13) by hierarchically filtering out contents, while the query is visually composed, with no activation gestures needed.

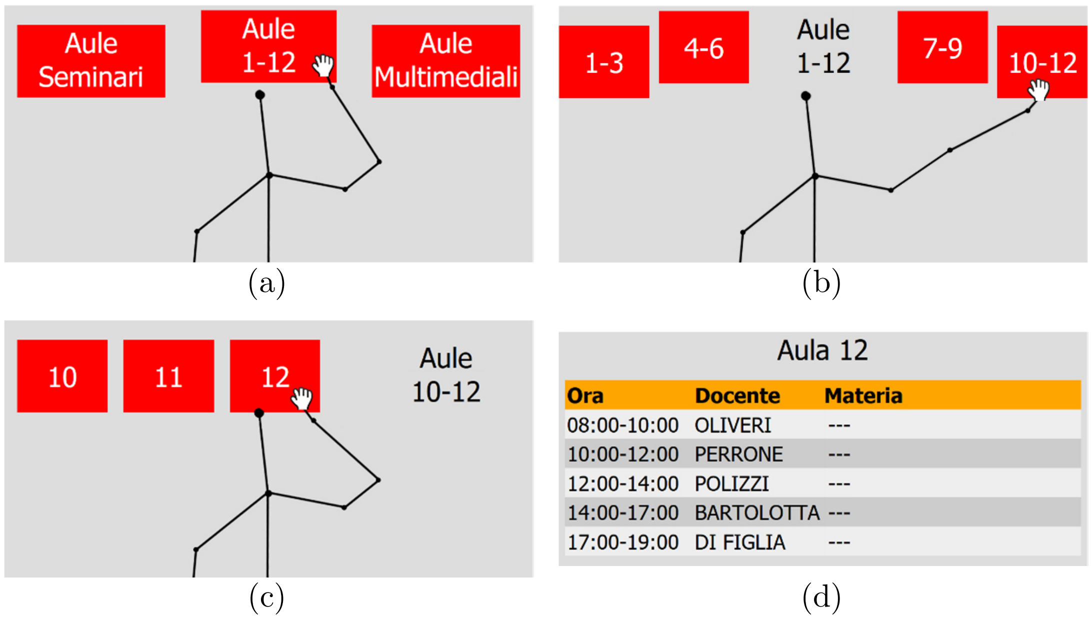


Figure 13. Hierarchical search via touchless interactions: pictures show the steps to display the lecture timetable of room 12 (i.e. "Aula 12" in Italian).

2.3 Interfaces Comparison: HIG-based vs. Avatar-based

After having described ABaToGI and its features, this section presents a study in which the Avatar-based interface is compared with another touchless gestural interface (Figure 14), based on the Microsoft HIG, i.e. a *de facto* standard for designing Kinect-based touchless gestural applications.

The aim of this study was mainly to address the following research questions:

- Do users prefer (or accept) to have no activation gestures to activate command, such as the selection or the click on a button?

- Does the presence of the Avatar in the middle of the screen help in overcoming interaction blindness?

The interfaces used for the study described in the following are aimed at the same goals, and have been designed for providing the same functionalities. In more detail, the development of both interfaces has been conducted in order to allow their use in a public space inside a University campus. The main goal of both systems was to provide an easy access to useful information for students, such as news, events, weather data and lectures timetable.

2.3.1 HIG-based Interface

In their Human Interface Guidelines, Microsoft recommends the use of one or two Kinect cursors, i.e. hand-shaped cursors by means of which users can interact with several tiles in the interface. Microsoft also suggests the use of an activation gesture, consisting in emulating a pushing action that, if executed when a Kinect cursor overlays an interactive tile, triggers the corresponding event.

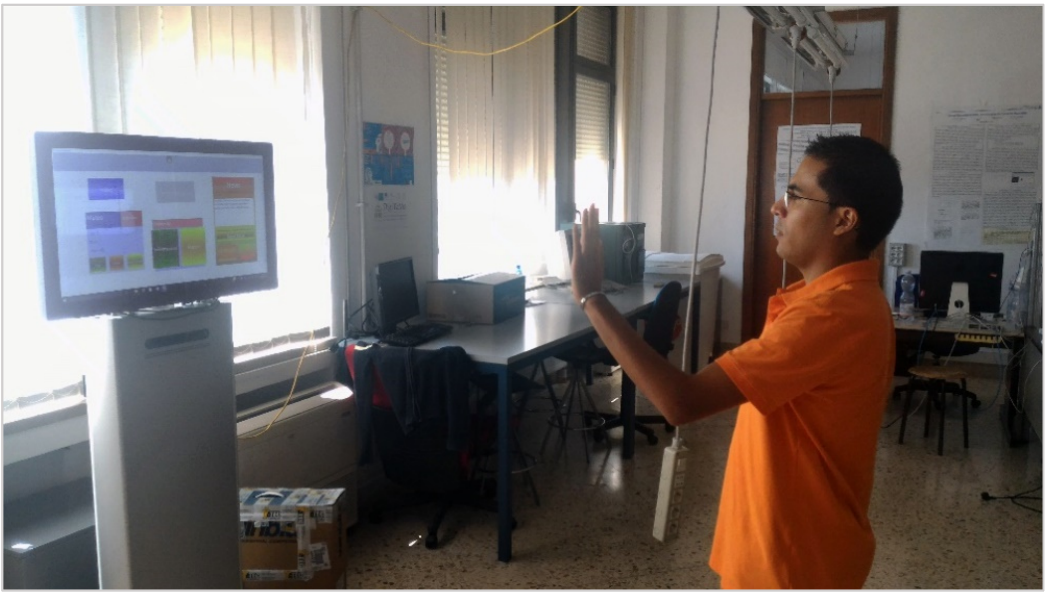


Figure 14. A user interacts with the HIG-based interface.

Another interesting feature is the presence of the *User Viewer* control, a small frame in the middle upper border of the window that shows the user silhouette (taken from Kinect depth camera). The presence or the absence of this silhouette suggests users if they are detected or not.

Figure 15 shows an interface developed by using the Microsoft Kinect SDK (which includes several controls ready to be used for implementing HIG). Again, this interface allows for the same operations of the previously described Avatar-based system.

2.3.2 Supported Tasks

Both the interfaces have been implemented to provide users with exactly the same functionalities and to accomplish the same tasks; in particular:

- reading news;
- reading university information;
- displaying and navigating campus map;
- displaying lecture timetable;

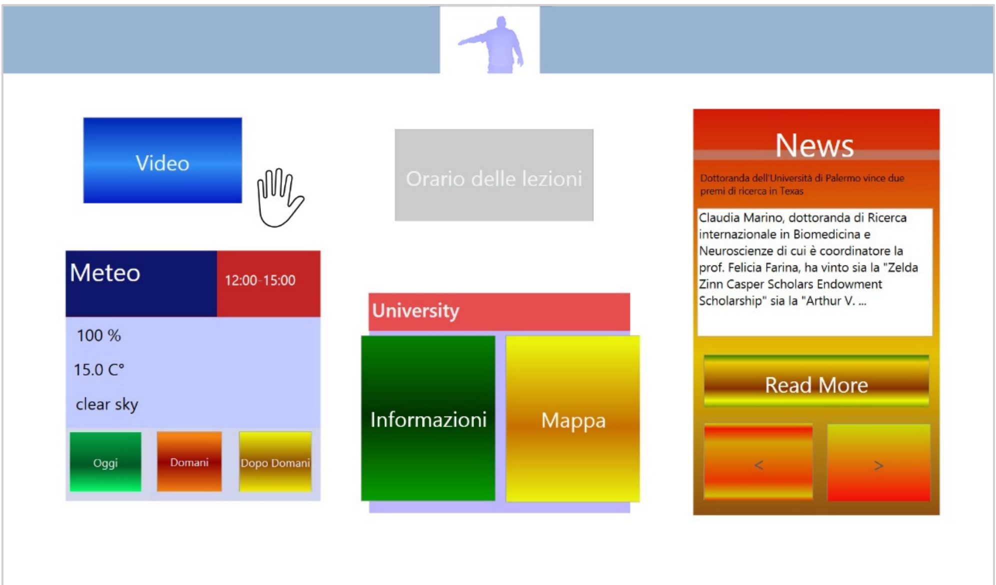


Figure 15. Layout of the HIG-based interface used for the comparison study.

- displaying weather data;
- displaying a video.

Of course, the layouts of the two interfaces were different, due to the presence of the avatar in the middle of the screen, and to reduce the Midas touch issue due to involuntary interactions with the tiles [4] [42].

2.3.3 Experimental Design

In order to evaluate which of the aforementioned interfaces better fits users' needs, this section will present the design of a user study for collecting users' opinions.

In particular, a public display has been installed in a transit area inside a building within the University campus in Palermo. This area is usually frequented by students of several disciplines and different ages (mostly from 19 to 35 years old), lecturers and other University staff members. The users sample included 17 students (10 male, 7 female) who have been asked to test both the interfaces ("within subject" set up [45]). However, some of the results discussed in the following have been obtained considering only the first tested interface, i.e. as if the study would be conducted in a "between subject" set up. Indeed, evaluating how an interface is able to communicate touchless interactivity may result trivial if the user was aware of its touchless capabilities before interacting with it. Since users were asked to perform two interaction sessions, only the first one was started without any knowledge about the interaction modality required, while the second one was unavoidably biased by the previous. Then, although users' opinions have been collected anyway (e.g. for understanding which was the preferred interface), some aspects are discussed by taking only into account the first interaction session.

Concerning the hardware, we used a 32-inch monitor placed at eye-sight, with a Microsoft Kinect sensor (clearly visible to all users) placed right above or below the monitor. We used these two different hardware arrangements in order to check if the sensor position significantly affects its recognition capabilities. Moreover, we wanted to verify if a proper positioning of the Kinect may increase the discoverability of the touchless capabilities of the interface. However, we did not notice any relationship between Kinect position and neither user recognition nor discoverability.

As anticipated, each user performed two 5-minutes-long interaction sessions (one per interface, tested in random order), followed by semi-structured individual interviews. In order to obtain the most valuable results, a diversified users sample have been chosen, with different levels of technology-related skills. We enrolled students attending various courses, from different disciplines.

In the interaction sessions, each participant was asked to carry out the following tasks:

1. find and read a specific news;
2. find and read university information;
3. find the timetable for a specific class;
4. play a video;
5. find and read the weather forecast for the next day.

Users were asked to perform these tasks without any suggestions or hints on how to achieve such goals, especially in terms of interaction modality.

The scheme used as basis for the semi-structured interviews was made by the following questions (see also Appendix A):

- Did you know that this system is based on touchless gestural input before starting the test?
    - If no, have you guessed that it was gestural?
        - If yes, which hints have suggested you that the system was/wasn't gestural? (e.g.: display size, presence of the Kinect sensor, the Avatar on the screen, etc.)
- Have you ever had previous experiences in interacting with gestural systems?
- Did you miss the touchscreen or the possibility of interacting using more conventional interactive modalities?
- Are there some other tasks that you would like to perform by using this system?

- Do you have any other suggestions or ideas to improve this system?

Other information asked to users was their sex, age, and current job, as well if they were right-handed, left-handed or ambidextrous.

The following section summarizes the findings, all based on the interviews and the observations during the interaction sessions.

2.3.4 Lessons Learnt

Among all the differences between the interfaces, the comparison study was focused on understanding users' preferences related to two crucial aspects:

- the use/non-use of an activation gesture to activate a command, such as the selection or the click on a button;
- if and how the presence of the Avatar in the middle of the screen helps in overcoming interaction blindness.

By observing users' interactions, several interesting behaviors have been also noted and described in the following.

Use of Activation Gestures

During the interviews after the interaction session with the Avatar-based interface, nine users assessed that they miss a gesture that allows to "click" on the tiles shown in the interface. Among the remaining ones, only three of them assessed they were comfortable in interacting without activation gestures. Users explained this necessity by explicitly referring to the habit of using mice and touch-based systems. Interestingly, others used some activation gestures also if they were not necessary, with the consequence of complaining about the fact that "some buttons activates by themselves".

On the other hand, guessing the activation gesture could result frustrating. None of the participants used immediately the "push" gesture on which the HIG-based interface was based, starting with other gestures (e.g. closing the hand, using a single finger, etc.). The difficulties in guessing the gesture to use may convince users to stop any further interactions (as noted also in [46]). These arguments suggest that the idea of avoiding activation gestures should be still pursued, or at least maintained in the initial interactions.

Fostering Bimanual Interactions

During the tests, another interesting behavior was observed: all users but one preferred the use of both hands while interacting with the Avatar-based interface, whilst all users but one interacted by a single hand with the HIG-based interface (Figure 16). Some of them explained this behavior because of their habit in using a mouse (which is always dragged and controlled by the same hand). Because of the presence of a cursor in the HIG-based interface, they used their gestures as if they were moving a mouse. On the other hand, being able to see the Avatar on the screen seemed to elicit the use of both arms.

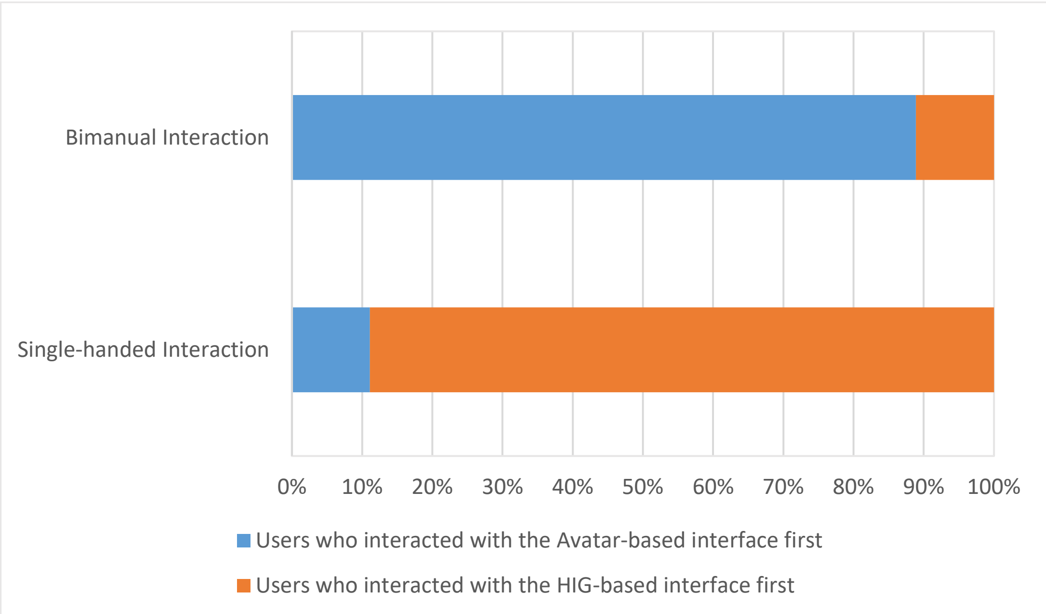


Figure 16. Use of hands for interacting with both the compared interfaces.

Communicating Touchless Gestural Interactivity

Another interesting characteristic of the Avatar observed during the tests was its ability to communicate the touchless interactivity supported by the interface. All the eight users who started the interaction session with the Avatar-based interface already knew the Kinect sensor; however, only two of them assessed that the presence of this device was the main clue for understanding the touchless gestural interactivity of the system. The remaining six users explained that they were immediately able to guess the touchless nature of the system, and that is was mainly due to the presence of the avatar in the middle of the screen. On the other hand, its presence was often perceived as annoying, muddler and useless when users interacted with the video or wanted to read a long text. In such cases, the Avatar continued to be visible in the middle of the interface (despite it became semi-transparent to let users read and see the contents through it). Although only few users explicitly assessed that displaying only hand-shaped cursors would be a better choice, several comments recorded during the interviews support this idea.

Time-to-Task Data Analysis

The interaction sessions have been exploited for gathering also quantitative data. In particular, the time required by each user to accomplish the previously listed tasks was measured. This was aimed to see if there could be a potential correlation between the time needed to carry out each task and the interface used. Considering the mean of the timings measured for all the tasks, users needed 15.5 seconds on average to perform a task on the Avatar-based interface, and 16.3 seconds on average using the HIG-based interface. Since the measured time difference is less than 1 second, which is reasonably considered insignificant if related with the average time to accomplish the tasks, it is not possible to conclude whether there is a clear correlation between the interface and the time required to accomplish a task.

However, this result proves that the Avatar-based interface proposed here is not worse than the HIG-based one. Indeed, any novel idea used in an interface paradigm always implies the risk of affect other aspects (e.g. the average task execution time).

By looking at the timings for each task (shown in Figure 17), it is possible to note some slight differences between the two interfaces. In particular, the "Read Uni Info" and "Play Video" tasks seemed to be faster accomplished through the Avatar-based interface, while the "Read Time schedule" and "Read Weather

Forecast" tasks are better accomplished with the HIG-based interface. In fact, the first two tasks are simpler (as they only require to activate a tile, and then read the content shown), while the other ones require more articulated actions. This supports the idea that the Avatar-based interface is more appropriated for providing rapid access to contents (i.e. it may be a proper solution for a starting page that acts as a main menu). More articulated tasks may instead require to use activation gestures (to provide more precise selection), as well as to hide the Avatar.

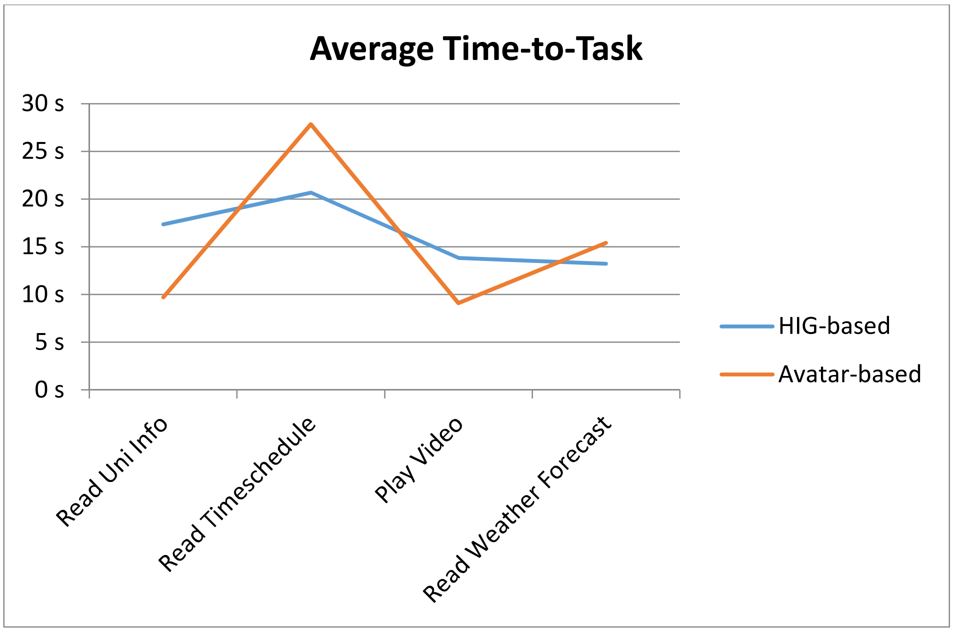


Figure 17. Average times to accomplish four tasks on both the interfaces.

Discussion

According to the above results, the best approach should probably be a sort of "fusion" of the two ideas described in this section. Using only hand-shaped cursors (as in HIG, except for dropping any activation gesture) could be a good choice for interacting with videos, images, texts and other contents. This way, users will not be annoyed by the presence of the Avatar that overlaps main contents, according to the opinions collected and discussed above.

On the other hand, the Avatar seems to be the best choice to interact with menus, allowing also for better communicating the touchless gestural interactivity.

Although the main goal was to verify if the proposed Avatar-based interface was able to communicate the interactivity more than the HIG-based one, a potentially useful feature of the Avatar was noted, although it still needs to be better investigated: the capability of fostering bimanual interaction. This may results useful, for instance, in some persuasive computing applications. Moreover, an interface that foster bimanual interactions may somehow increase the number of glances or (even better) interactions from passers-by/users. Because using both arms may result in wider (and thus more "visible") gestures, it is reasonable to think to a growing possibility of fostering honeypot effects around the interacting users.

Starting from this apparent potential of fostering bimanual interactions, the next Chapter aims at investigating how and if this capability is related to the presence of the Avatar in the middle of the interface.

Chapter 3
Investigate Bimanual Touchless Gestural Interactions

As described in the previous Chapter, during the comparison study, users showed an attitude of interacting with both hands more frequently when using the Avatar-based interface. In other words, while with the HIG-based interface users interacted mostly using a single hand, the presence of an Avatar in the middle of the screen that replays user's movements seemed to induce them to try both hands much more frequently (although not necessarily simultaneously).

This observed behavior opens several questions:

- What is the main element of a touchless interface that induces single-handed or bimanual interactions?
- While seeing the Avatar, do users interact with two hands because this seems to be more "natural" to them, or because some other interface elements affect their behavior?
- Do users perceive bimanual interaction as a usability problem?
- Do users feel more in control while interacting with two hands or one hand?
- From an ergonomic point of view, do users think it is more stressing to use two hands or only their preferred one?

In order to address some of the above questions, this Chapter summarizes the experimental design and the outcomes of another comparison study between two touchless gestural interfaces, one based on the presence of an Avatar (i.e. ABaToGI), and one being a modified version of ABaToGI that only displays the two hand cursors, with no Avatar at all.

After a brief overview of the related works in the field of bimanual interactions, in the following the experimental design is described (including also the visual improvements introduced on ABaToGI). The results and the related discussion close this Chapter.

3.1 Single-handed and Bimanual Interactions

Section 2.1 described an overview of the touchless gestural solutions adopted for introducing novel interactive capabilities in public displays. Most of the mentioned interfaces allowed for both single-handed and bimanual interaction; however, in many cases these possibilities have not been investigated in depth.

Differences in users' preference between single-handed and bimanual interactions have been documented in some gesture elicitation studies. Even if the main focus of this thesis is about touchless gestures, here it is important to consider also some noticeable findings related to other interaction modalities (e.g. touch-based ones).

In [22], Wobbrock et al. described the outcomes of a gesture elicitation study aimed at the development of a gesture set for interactive horizontal surfaces. In this study, 20 users were asked to perform their preferred gestures for 27 different referents (i.e. videos depicting the visual effects that should be the result of a hypothetical gesture). From their tests, authors showed that "participants preferred 1-hand gestures for 25 out of 27 referents [...], and were evenly divided for the other two". This attitude has been also confirmed by other similar elicitation studies for touch-based gestures, as in [47] and in [48].

Considering now touchless gestural interaction, an interesting elicitation study has been described by Koutsabasis and Domouzis in [49]. In this work, authors asked users to think and perform mid-air gestures for only two referents: browsing an image gallery and selecting an image from the gallery. Not surprisingly, results showed that users clearly prefer to use a single hand.

Another similar result has been presented by Walter et al. in [41]. In this paper, authors have observed users' behaviors while interacting with a public display by touchless gestures. They noted that "from those users that could potentially use both of their hands [...], 80% decide to use the same hand [...]. Even if they could use the left hand to better reach an item at the left side of their body, they would still use the right hand". Although this may seem surprising, the reason may be related to users' habits in using WIMP interfaces (i.e. legacy bias [26]).

According to the aforementioned outcomes, a touchless gestural interface should allow users to interact mainly with a single hand. This is also confirmed by Microsoft in their Human Interface Guidelines for developing touchless applications for Kinect [14]. In particular, they suggest to "use one-handed gestures for all critical-path tasks. They're efficient and accessible, and easier than two-handed gestures to discover, learn, and remember", as well as to "use two-handed gestures for noncritical tasks (for example, zooming) or for advanced users". Practically, several gestural applications by Microsoft (e.g. the samples provided with the Microsoft Kinect SDK) only allow for interactions with one hand, forcing users to only use a single hand-shaped cursor at a time.

This discussion suggests that, as a usable gestural interface have to be based on easily guessable gestures, these interfaces should allow users to interact mainly with a single hand. However, a designer may still have the need of fostering bimanual interactions, in order to reach some more significant or specific goals. For instance, an interface that fosters the use of both hands can result less tiring and more ergonomic for short interactions: users would change arm more frequently, so the fatigue should be more equally distributed among both arms instead of on a single one. Another advantage may relate to the *honeypot effect* [50], i.e. users are more interested in a public display (both in terms of interactions and number of glances) when other people interact with (or look at) it, rather than when nobody is in front of the display. Having this in mind, if users interact with two hands, her movements are more visible in the surrounding environment from other users, thus this situation may increase the possibility that honeypot effect occurs.

3.2 Experimental Design

In Section 2.3.4, the comparison study between an HIG-based interface and the Avatar-based one showed interesting hints about a possible relation between the presence of an Avatar in the middle of the screen and the user's attitude of using two hands. However, although both interfaces were built for allowing users to reach the same goals, other factors may influence user's behavior. For instance, the need of using a "press" gesture may convince the user to "stick" on a single hand, which would be something completely unrelated to the presence of the Avatar.

In order to verify if there exists a correlation between the presence of the Avatar and the use of two hands during the interaction, a comparison study has been conducted between two modified versions of ABaToGI: a first one that still used an Avatar, and another one where only hand cursors were displayed. The following Section introduces these two interfaces, along with the improvements applied on ABaToGI. Then, the comparison study is described, along with the results coming from the data analysis.

3.2.1 Touchless Gestural Interfaces Description

After the pilot and comparison studies, the first improvement on ABaToGI was aimed at making its appearance more attractive and visually prettier. Moreover, some of the supported functionalities, although still available, were slightly changed according to the previously discussed users' feedbacks, and in order to better integrate them with the new interface look. In particular, a web UI designer was asked to prepare some detailed mockups for this interface, to be used as implementation baseline. He was provided with the previous interface, and a brief discussion was conducted to explain him all the features of ABaToGI, and the main issues noted during the aforementioned studies. The first proposed mockup of the main page is shown in Figure 18.

This first proposal presented a couple of problems. First, tiles in the upper corners are unreachable without passing on other interactive tiles. Because

Figure 18. First mockup proposed for the second version of ABaToGI.

placing Avatar's hand on the tiles implies their activation, this layout was slightly changed by making the upper tiles more centered.

Another problem was about the interactivity of the weather widget. By placing an Avatar's hand on top of a weekday, the weather shows the related forecast information. However, the position of such weekdays was too near to the center of the interface: this implies user to inadvertently pass over such tiles, thus activating an unwanted behavior. This issue was solved by changing the position of the weekdays from the bottom to the left of the tile.

Finally, yet importantly, in this new version of ABaToGI the news slider has been completely redesigned. In the previous version of the interface, users assessed that they were not comfortable in using the arrow tiles under the "Read more" button (see Figure 11). This was mainly due to difficulties in avoiding unwanted activations of next/previous news commands (activated every time a hand went on top of the tile). After a discussion with the designer, the possibility to switch among the news was dropped from the first page, and the news slider became an automatic slideshow. These improvements are shown in Figure 19.

Figure 19. Layout of main page of ABaToGI.

The following Figures show the final layout used for the other sections of ABaToGI.

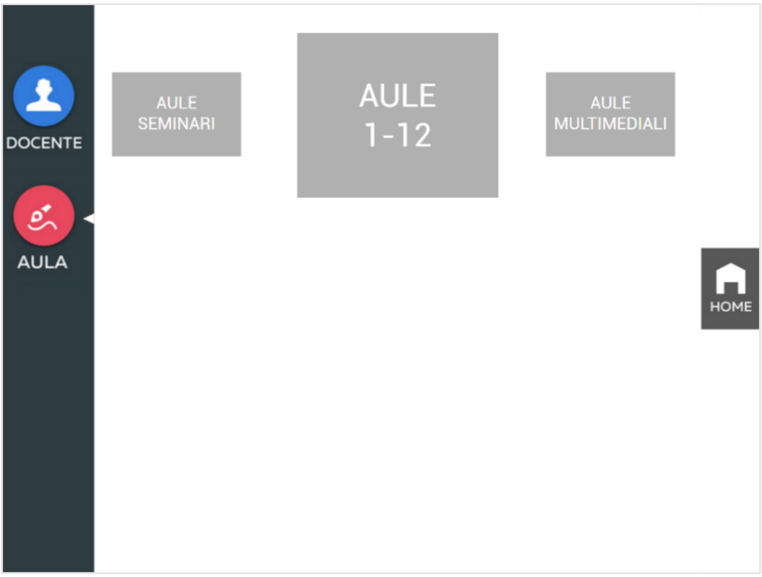


Figure 20. The layout of the hierarchical timetable search used for ABaToGI.

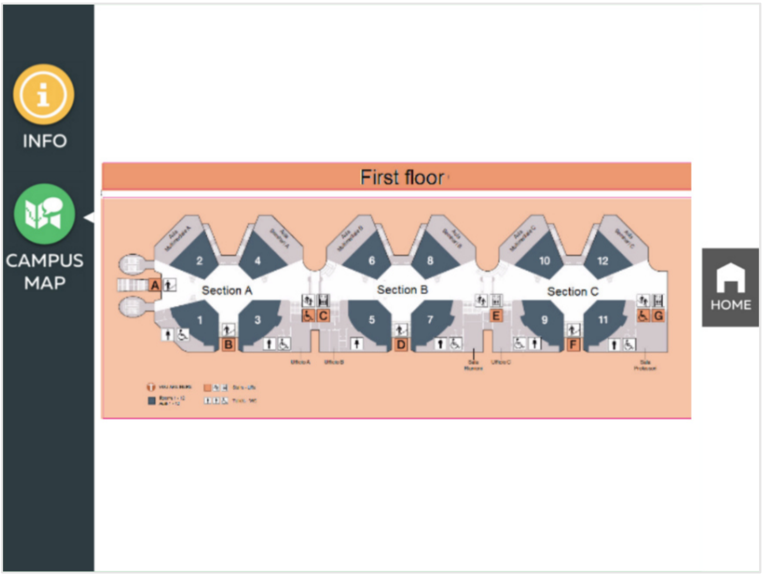


Figure 21. The layout of the campus map navigation used for ABaToGI.

Figure 22. The layout of the news page used for ABaToGI.

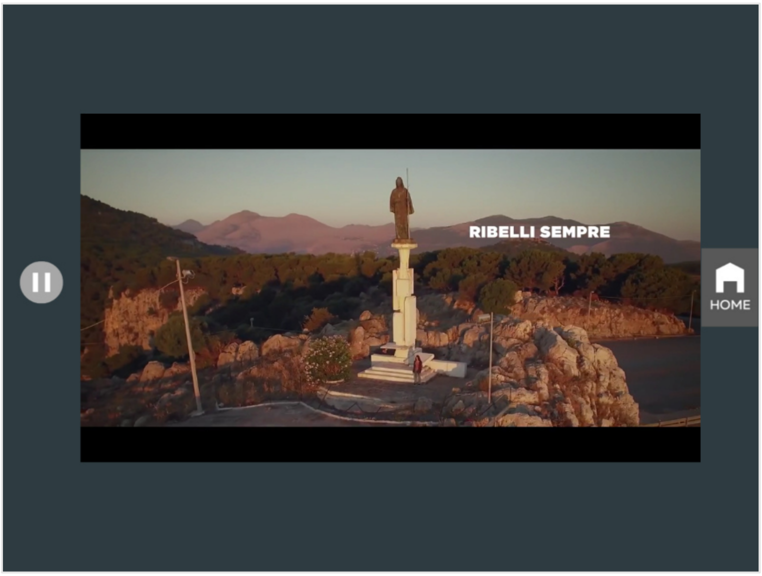
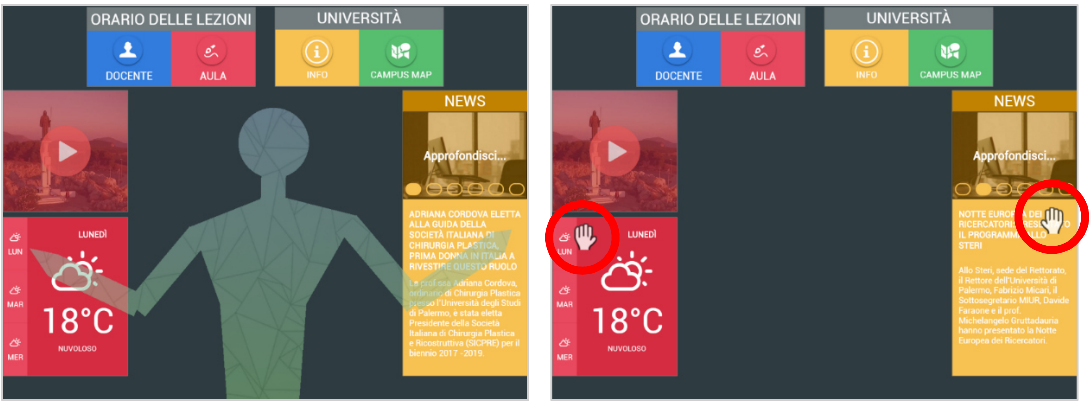


Figure 23. The layout of the video player used for ABaToGI.

This layout has been then used for producing two identical interfaces, except for the presence of the Avatar (see Figure 24). In the following, the interface with the Avatar will be referred as “interface A”, while the other one (where only two hand cursors are used, and no avatar at all is shown) will be referred as “interface B”.



(a) (b)

Figure 24. Visual comparison between the interfaces.

3.2.2 Study Description

The two interfaces described in the previous Section have been used for conducting a comparison study, consisting of a brief interaction session with one of the two interfaces, followed by a questionnaire that each user had to fill in. In particular, 50 users (31 males, 19 females) were asked to interact with the two interfaces, in a between-subjects set up (i.e. 25 users interacted only with interface A, and the remaining ones only with interface B). Before the interaction sessions, all the users were informed about the need of using touchless gestures to interact with the interface. As the interface was designed for being deployed in a University foyer, the users were chosen on the basis of their age (ranging from 19 to 50) and occupations (students, staff members, and lecturers).

The interaction sessions have been conducted in a controlled environment (i.e. a laboratory). The hardware consisted in a 42-inch LCD monitor placed at eye height. A laptop was connected to the monitor and a Microsoft Kinect sensor (clearly visible to all users) was placed right below the screen. Users were asked to interact from a distance of about 1.7 meters from the display.

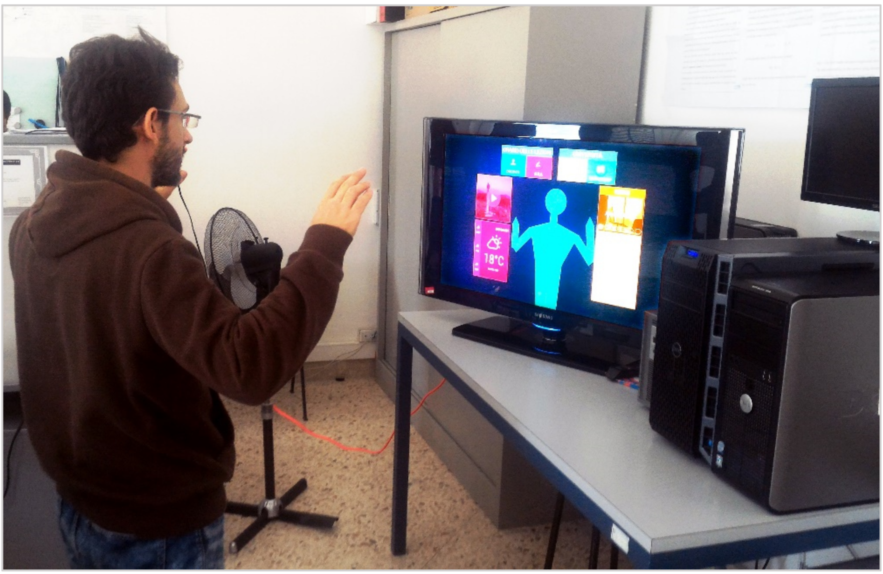


Figure 25. A user during the interaction session.

The goal of this study was to understand if there exists a relation between users' preferences in using one or two hands while interacting via gestures, and the presence/absence of their Avatar in the middle of the screen. Moreover, a secondary goal was to compare the perceived cognitive workload between the two interfaces. To this end, we used a slightly modified NASA-TLX questionnaire [51], with the addition of three simple questions focused on the use of one or two hands for the interaction session. The questionnaire is available in the Appendix B (translated in Italian as it was the native users' language, with additional English captions), and it was used as a "Raw TLX" [52] (i.e. by eliminating the weighting process of the subscales and then analyzing them individually).

Results are discussed in the following Section.

3.2.3 Results

The NASA-TLX subdivides the whole workload in six 20-points Likert subscales, each of which is intended to measure a specific aspect of the task execution:

- *Mental Demand*, intended to evaluate how much mental and perceptual activity was required to accomplish the task;
- *Physical Demand*, intended to evaluate how much physical activity was required to accomplish the task;
- *Temporal Demand*, intended to evaluate how much time pressure the user felt due to the pace at which the task occurred;
- *Overall Performance*, intended to evaluate how successful was the user in performing the task;
- *Frustration Level*, intended to evaluate how irritated, stressed, and annoyed versus content, relaxed, and complacent the user felt during the task;
- *Effort*, intended to evaluate how hard the user had to work (mentally and physically) to accomplish the task.

Figure 26 allows for comparing the means of the above characteristics. By observing the graph, it is clearly visible that both median and mean values are lower (i.e. better) for the interface A (i.e. the "Avatar" columns). In particular, a Fisher-Freeman-Halton test revealed that the interface probably affects mental demand (93% confidence interval, i.e. $p \cong 0.07$). Moreover, the same test showed that all the other measured values (physical and temporal demands, overall performance, frustration level and effort) are influenced by the used interface ($p < 0.04$). Generalizing these results, a plausible conclusion is that displaying an Avatar can decrease the workload when interacting with a touchless gestural interface.

Although the clear significance of the previously described outcomes, the main goal of this study was to investigate the capability of fostering bimanual interactions of an Avatar in the middle of the screen, replaying users' movements. During our tests, however, all the users decided to use both hands (even if often not simultaneously) while interacting. This is probably due to the

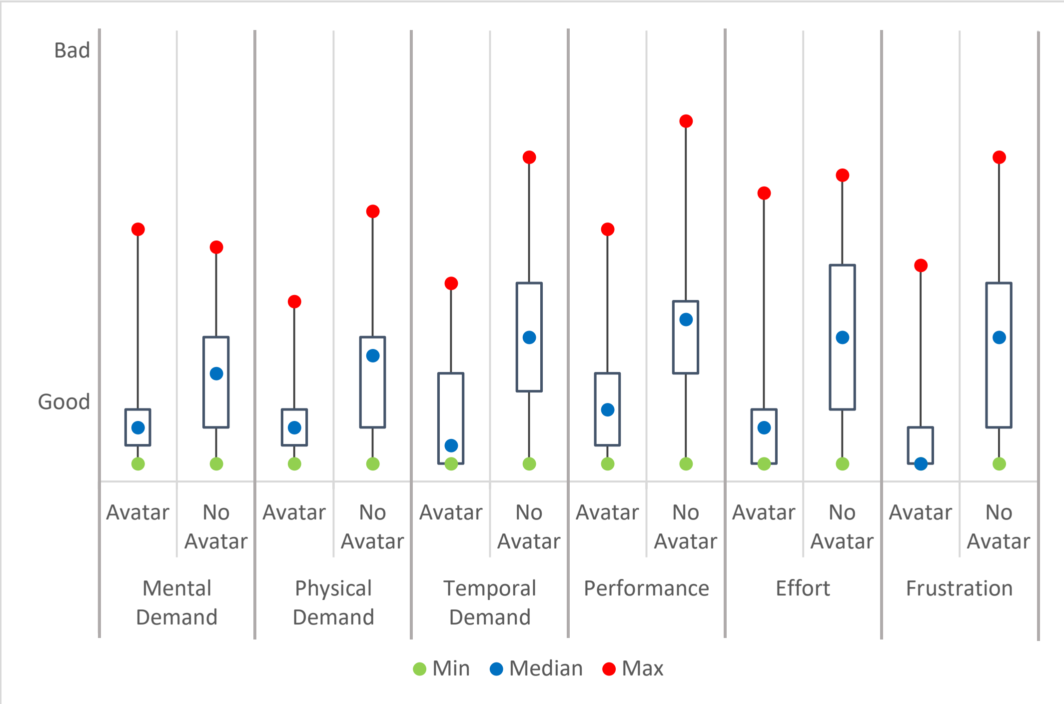


Figure 26. Results from NASA-TLX data.

interface layout: as Figure 24 suggests, reaching tiles in the left upper side can result very difficult with only the right hand (and this is the same for the opposite tiles with the left hand). In order to understand if users switched to bimanual interactions because of a perceived usability issue or not, we asked them the following two questions (see also Appendix B):

a) Have you used two hands because of a usability issue of the interface, or just because you prefer to use two hands?
b) Would you rather prefer to be able to use one hand only for the whole interaction session?

Table 2. Contingency table for question a)

		Why used two hands?	
		Usability Issue	**Preference**
Interface	**Avatar**	6	19
	No Avatar	20	5

Table 3. Contingency table for question b)

		Would prefer to use a single hand?	
		Yes	**No**
Interface	**Avatar**	6	19
	No Avatar	16	9

Users were thus asked to answer two multiple-choice questions. Their responses are summarized as contingency tables in Table 2 and Table 3. A Chi-squared test confirmed the statistical correlation between the interface used and users' responses ($p < 0.05$ for both the contingency tables).

3.2.4 Discussion

In conclusion, the results presented previously show that using an Avatar decreases the perceived cognitive workload if compared with the same interface with only hand cursors. Moreover, users seem to use both hands as a natural and intuitive way for interacting when the Avatar is displayed. The latter idea is supported by the fact that most of the users that interacted with interface A did not perceive the need of using two hands as a usability issue. On the contrary, the majority of users that interacted without seeing the Avatar (i.e. with interface B) judged the need for bimanual interactions as a usability issue (see Table 2). In addition, most of the users that interacted with interface A assessed that they do not prefer to use a single hand instead (see Table 3).

It is worth noting that the findings described in the previous Section do not demonstrate the originally expected relationship between the presence of an Avatar and the use of single or both hands during the interactions. However, further investigations may show that other factors relate with the ability of fostering bimanual interactions (e.g. the particular layout used in the interface, or the use/non-use of activation gestures).

Chapter 4
ABaToGI in-the-wild: a Case Study

In the previous Chapters, ABaToGI has been presented along with its features. Moreover, it has been used for conducting two controlled studies, where the ecological validity was unavoidably limited by the presence of an interviewer (i.e. the experimenter), or by the need of conducting the study inside a laboratory (i.e. not in public, where the interface should be actually deployed for its goals) [3] [53]. While the experimenter intervention can be somehow limited or “measured” when evaluating users’ behaviors and feedbacks [54], this is not always true when analyzing the results of lab studies. In particular, the latter does not allow for taking into account most of the issues related to public displays, e.g. social interactions among users, social acceptability of gestures in public, display blindness, etc.

This Section presents a case study where ABaToGI has been evaluated in-the-wild. After a brief overview of the issues related to conducting longitudinal studies in uncontrolled environments, the case study will be presented, and it will be followed by a discussion that takes into account observations and interviews.

4.1 Evaluations in-the-wild

In 2005, Richard Sharp and Kasim Rehman coordinated the UbiApp Workshop [55], which had the aim to define new practices for application-led research in the area of ubiquitous computing. In particular, experts in this research field agreed on how to evaluate ubiquitous applications, “arguing that the only way to evaluate an application against the ideals of ubiquitous computing [...] is through long-term deployment in the wild. [...] Small-scale lab studies still have a place - everyone agreed that they’re very useful in the early stages of user-centered design”, but “once researchers have performed lab-scale trials, they [...]

should use this data to continue to design, deploy, and evaluate similar applications on a larger scale" [55]. In other words, a fundamental result of the UbiApp Workshop was the need to evaluate applications outside of controlled environments, i.e. *in-the-wild.* This is particularly important in evaluating applications for public displays, because of the strong difference between a laboratory and the real settings in which these systems are deployed.

The necessity to carry out longitudinal studies in-the-wild has been then underlined by various authors [3] [28] [56] [57]. Ojala et al. [30] have deployed various displays in public places (naming them UBI-hotspots), and published results from a three-years-long study. During this period, they continuously observed behavioral changes in users and collected new insights for improving display functionalities and contents. Their findings demonstrate how important are longitudinal studies in order to follow users' preferences; in fact, UBI-hotspots are still the subject of ongoing longitudinal studies. The need of conducting such kind of data analysis in-the-wild is particularly important when studying public displays, because of their implicit "wild" nature: this is probably the only way to collect ecologically valid data [3], and the related results.

In order to correctly acquire users' preferences while interacting with public displays, the possible presence of researchers or experimenters should be taken into account. In real situations, users are not invited by anyone to interact, and they do not know which is the interaction modality. The presence of an experimenter who asks users to interact and explain how to do it, allow him to collect much more data than a totally uncontrolled situation. However, the intervention of an external agent on the users' behavior introduces a bias (what we refer as *experimenter bias*). Solutions to this problem fall into two categories [28]: 1) allow experimenter intervention, and study the arising bias; 2) avoid experimenter intervention and keep the environment uncontrolled.

Johnson et al. [54], who participated in the activities to be evaluated with the users, investigated the first option. They derived several dimensions in which the role of the researcher can be described, in terms of the abilities to facilitating or encouraging users, explaining the system, but also the level of authority and familiarity with participants, and the experimenter's relationship with the research. They concluded that participating and building a friendship with users can improve knowledge about how they see the system or the prototype object of the study. Another class of approaches that require experimenter intervention are the ones in which users are asked to fill up questionnaires directly provided by the researcher during the experiment. In [58], questionnaires are provided to users before and after interactions, in order to evaluate their expectations and experience. In this way, the behavior is biased, but there is a margin to evaluate this bias, by comparing expectations (questionnaires provided before the experiment) with experience (questionnaires provided after the experiment). Furthermore, it is possible to collect much more data and it is relatively easy to analyze them based on the answers to the questionnaires.

According to [54] and [58], allowing experimenter intervention provides researchers with several advantages. However, this approach inevitably introduces biases in users' behavior. This is the main reason why avoid experimenter intervention should be the preferred option. With their three-years-long deployment, Ojala et al. [30] have demonstrated how much information is available without the experimenter's intervention. They explicitly assessed that laboratory, single-location and campus-wide deployment cannot capture location influence, and so it is important to reproduce experiments and evaluation in a wider area, and in different locations. They are still continuing to evaluate their public displays, trying to solve the users' hesitancy to use technology in public (indeed, it is another finding which can be only discovered and studied with no experimenter intervention). Non-intrusive methods were used also by Messeter and Molenaar [59] in order to evaluate non-interactive ambient displays, basing all observation on data gathered from cameras and a Wizard-of-Oz approach to edit the displayed content. They also directly interviewed users, but only at the end of the experiment, when their intervention did not constitute a bias anymore. In [60], researchers evaluated gestures used to interact with a tabletop computer. They blended themselves in the crowd by using casual clothes, and collect data using cameras.

However, the main drawback of experiments in which researcher's intervention is not allowed is the requirement of long-term studies, which usually implies high costs and efforts. This is probably the main reason why several public displays are often not evaluated using this approach, but usually with explicit researcher's intervention. The main challenge here is to find methods to reduce costs and time. Imitating the "experimenter blending in the crowd" solution proposed in [60], as well as the use of cameras may help in this direction. Such methods are not simple as they seem, because of ethical issues like the privacy of users or the need to inform them before executing any personal information gathering.

4.2 Deployment Description

In order to evaluate ABaToGI in-the-wild, a public display has been deployed in a public transit area, similarly to the installation described in Section 2.3.3. In particular, the display was installed in a little indoor square inside a building within the University campus in Palermo (see Figure 27), next to a couple of benches where students often sit while waiting for lectures starting times.
As explained previously in this thesis, this area is usually frequented by students of several disciplines and different ages (ranging from 19 to 35), as well as by lecturers and other University staff members.

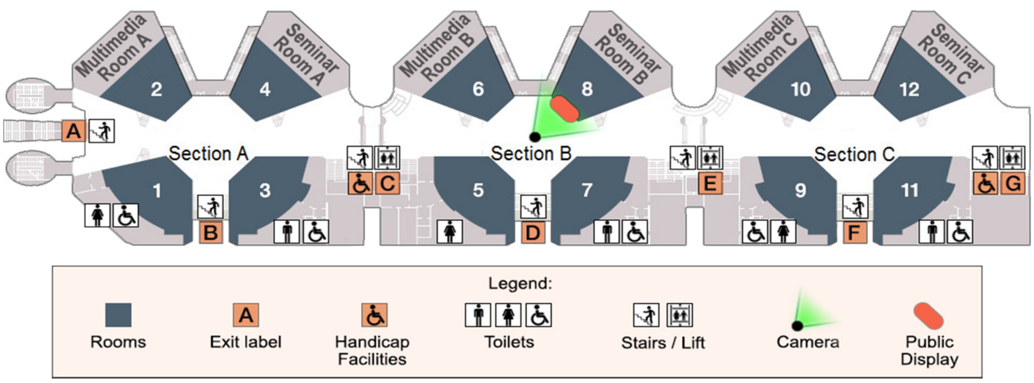


Figure 27. Map of the building where the display was installed.

The public display consisted in a 32-inch LCD monitor placed at eye height, with a Microsoft Kinect sensor placed right below the monitor. In particular, the Kinect was fixed inside a solid case (made of plastic and metal), together with all the other hardware components, in order to secure them and avoid theft attempts. The case acted also as display holder; its detailed scheme, along with all the relevant measures, is available in Figure 30.

Along with the display, on which ABaToGI was installed, there was the need of observing (and possibly recording) users' behaviors. To this end, a WiFi camera was installed in front of the display (see Figure 27), in a non-reachable position. This allowed both for remotely observing users, but also to be able to observe the actual display status. Indeed, Davies et al. noted that "the only thing that is really important is what is on the screen of a public display, yet this is the most difficult thing to monitor. Monitoring software that reports on the status of the player hardware is typically of limited value [...] If, for example, the display surface is damaged, then the system may not be displaying content

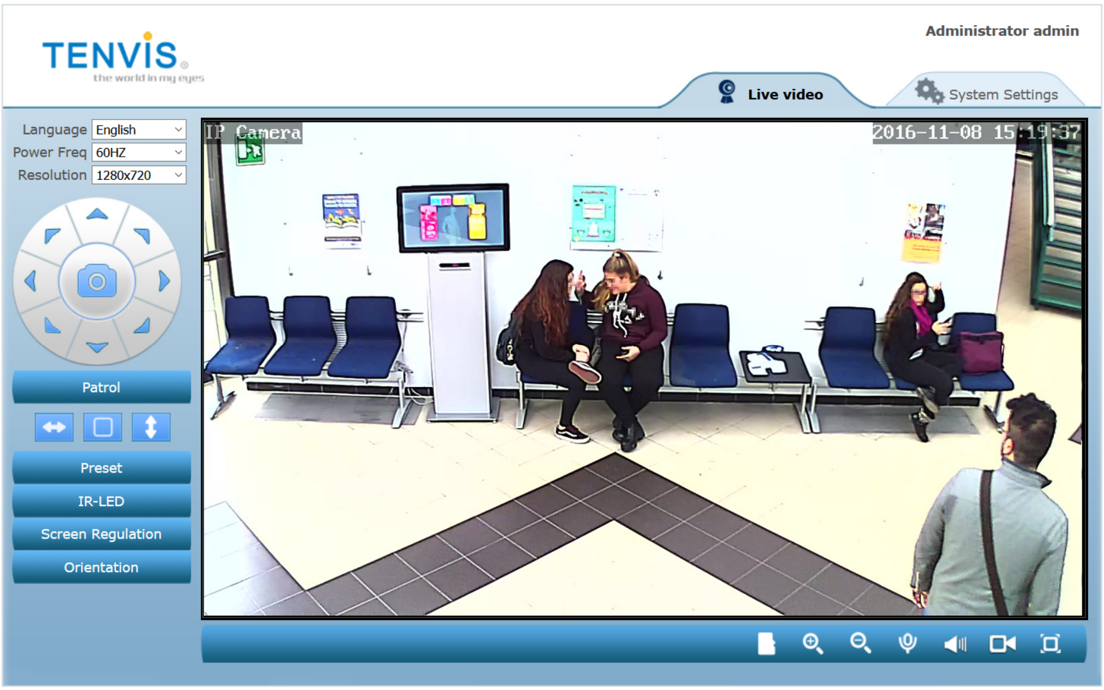


Figure 28. Web GUI for monitoring users' behaviors and display status.

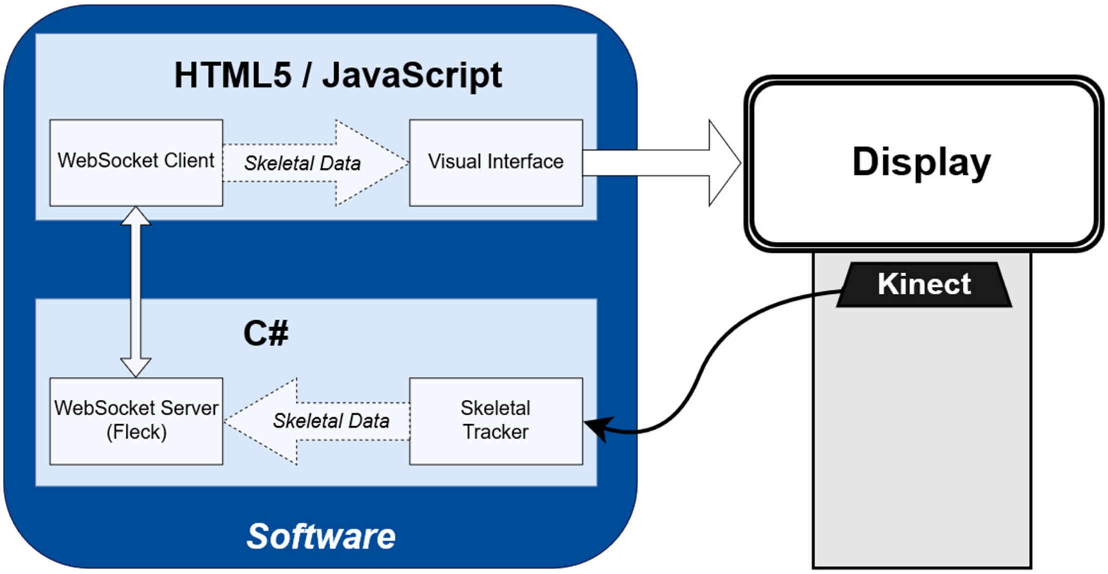


Figure 29. System workflow.

even though no errors are reported" [3]. Display status may then be monitored remotely by means of a camera, which streams video via the Internet. Figure 28 shows a web GUI used to check both users' behaviors and display status.

For this deployment, ABaToGI has been developed as an HTML5 / Javascript web application, connected to a C# server that read data from the Kinect. Those data were thus sent to a web socket and read from the rendering client in order to build and show the Avatar (see Figure 29). The whole application run on Google Chrome (and it was also Firefox-compliant), and it was executed automatically on start-up. The display was configured to turn on automatically every weekday at 8 a.m., and to shutdown at 8 p.m. (i.e. at opening and closing hours of the building). Any crash, errors or any other issues had been managed remotely using Team Viewer [61].

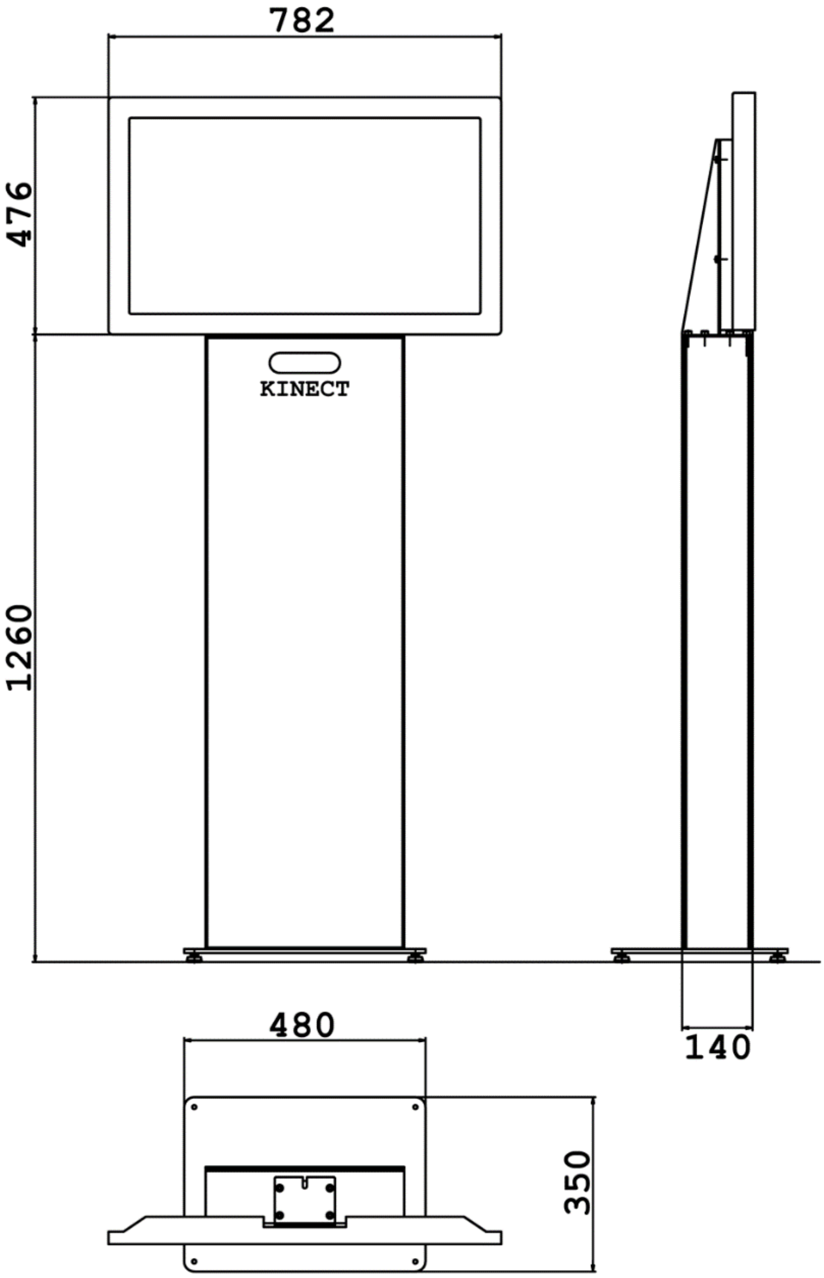


Figure 30. Technical scheme of display and case/holder used.

4.3 Case Study

ABaToGI has been deployed for two months, and during this time users' opinions have been collected by means of observations (both manual and automatic [3] [62]), questionnaires (NASA-TLX, see Appendix B) and semi-structured interviews.

Observations have been conducted for a total of about 30 hours, distributed across 40 days. During this time, an experimenter went to the deployment place, sit on a bench next to the display and observed users' behaviors, taking handwritten notes. Among the aforementioned observations, some consisted of analyzing video recordings (offline) and live streaming (remotely).

During observations on-site, the experimenter also invited some users to interact with the interface and then asked to compile a questionnaire. The number of observed interacting users was about 50, and about 50% of them have been directly invited by the experimenter. A total of 29 users accepted also to compile the questionnaire, while only 6 users accepted to conduct a semi-structured interview (using the scheme available in Appendix A).

The rest of this Section will describe the main findings noted during the aforementioned observations.

4.3.1 Communicating Interactivity

The main rationale behind the interactive Avatar placed in the middle of the interface was to help users in understanding the touchless gestural interactivity supported by the interface. During the observations, indeed, the experimenter noted this capability several times:

> *"A user was attracted by the Avatar, and he approaches the display. He observed the screen (and in particular the Avatar) for several seconds, clearly curious but without interacting. Then I approached him and explained that he could interact by gestures. He explained that he had understood the interactivity, but he did not interact because embarrassed.*

> *Moreover, he did not seem very skilled, but after a while, he showed his appreciation for the novelty".*

In this situation, the Avatar showed its capability of communicating the interactivity; however, it was not clear if the user was able to understand how to interact with the interface (i.e. to use touchless gestures as interaction mean). Other users seemed to understand quite easily the gestural interactive capabilities, and most of them (interviewed after the interactions) explained that it was due to their previous experience with Kinect-based applications. This is the case of the user interaction described in the following extract:

> *"At the beginning of this observation session, I noticed a user who was using the interface. He seemed very skilled, so when he finished I approached him and asked some opinions about the experience. He told me that he had previous experiences with gestural systems (he used one abroad in the past), and he immediately noted the Kinect. He also suggested some improvements [...], but he was definitely satisfied with the currently supported features".*

However, not all the users were able to understand the gestural capabilities, and some of them did not understand them correctly. In many cases, users' prejudice about the supported interaction modality was clear: the display size probably naturally affords more touch-based interactions than touchless ones. Moreover, using gestures in public seems to be something not easily acceptable for all. In the experimenter's notes, several excerpts demonstrate both this prejudice and the low acceptability of gestural interactions; for instance:

> *"Two users interacted with the interface.*
>
> *The first one approached the display and tried to use its touchscreen. After some attempts, she figured out that the system was not responding, so I tell her that it was touchless. She did not know any similar systems and stated that touchscreens are, in her opinion, more practical.*
>
> *The second user interacted by means of gestures (but after having understood how to interact by observing the first one). After a brief*

interaction session, she explained her embarrassment in using mid-air gestures, and that she would prefer to use a more traditional touchscreen".

The previous excerpt is in line with the analysis of the NASA-TLX questionnaires. If compared with the in-lab study described in Section 3.2, mental and physical demands, as well as performances, are more or less in line with the expectation (see Figure 31). However, a Fisher-Freeman-Halton test revealed that the context probably affects the general cognitive workload perceived by users: after the interactions in public, frustration level resulted in being slightly higher if compared with the in-lab condition (>95% confidence interval, i.e. $p < 0.05$). Moreover, users' opinions revealed that, on average, interacting in-the-wild implies an increasing temporal demand. In other words, users feel more comfortable and less frustrated in interacting via gestures in a

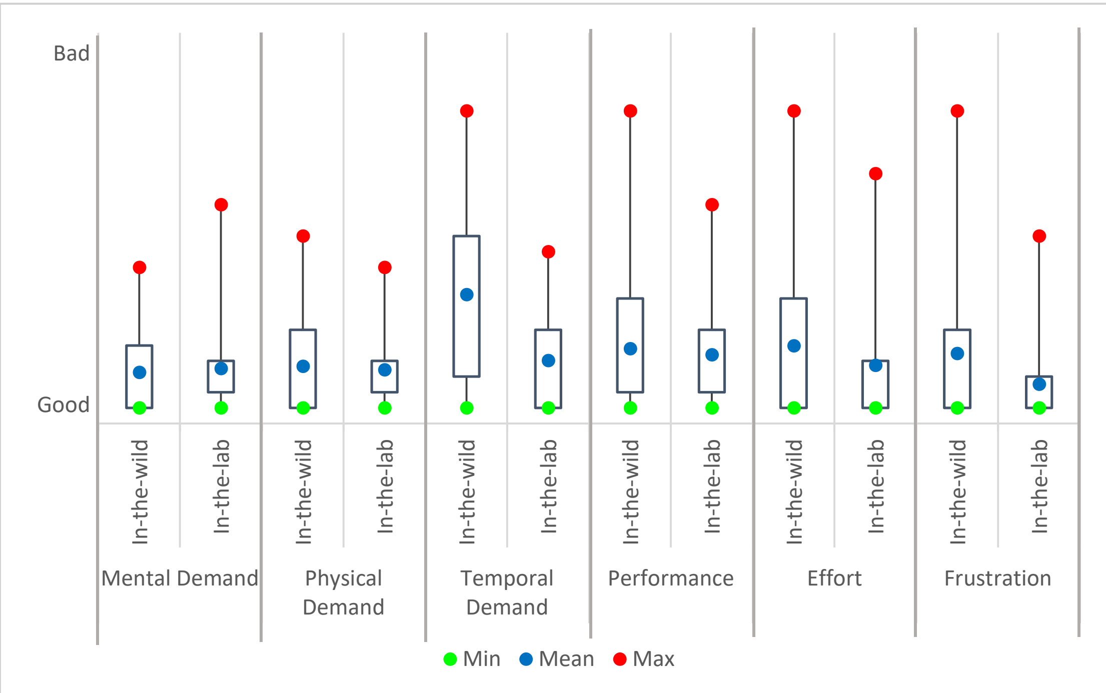


Figure 31. Comparison of NASA-TLX data gathered during in-the-lab and in-the-wild studies.

controlled environment (probably because of the lower level of embarrassment due to the absence of stranger observers, as well as other external factors).

The attempts of using the screen by means of touch interactions were demonstrated also by an interesting observation:

> *"I tried to interact with the interface [...] and I noticed a really curious fact. The backlight showed me several fingerprints on the screen surface, and this can only mean one thing: during these days, several users guessed that the screen was able to detect touches, so they used it as a traditional touchscreen".*

A possible and partial explanation of such issues may be sought in the screen size and the previous experiences of users in interacting with situated public displays. As stated before in this thesis, situated public displays are most commonly equipped with touchscreens; so it is plausible to think that users expectations about supported interactivity are more oriented on touch-based ones than on other alternatives.

It is also important to note that some users approached the display from left or right, arriving in front of the Kinect (i.e. within its field of view) at about 10-20 centimeters from the sensor. At this distance, the Avatar is not shown at all, due to the sensor capabilities [14]. A possible solution may be to place a marker on the floor, which indicates users the correct distance at which they should stay for a correct interaction. Moreover, introducing some sort of mechanism to alert users about the required distance may help in this direction.

Obviously, such kind of issue should disappear if the system is deployed with bigger and/or not reachable displays, where touch-based interaction can be neither afforded nor supported.

4.3.2 Single-handed vs. Bimanual Interactions

As noted in the previous Chapter, ABaToGI users seemed to prefer to use both hands/arms for their interactions. However, the quantitative results analyzed in Section 3.2.3 have been collected in a controlled environment, which lower their ecological validity. In order to confirm such results in an actual deployment, the experimenter focused particularly on the way users interacted, and collected some interesting notes:

> *"I have invited a user to interact with the interface. He used only the right hand for more or less the first 20 seconds. Then, he started using also the other hand, and said that he felt much more comfortable and in control".*

Similar behaviors have been recorded in other excerpts, with different users:

> *"At the beginning, the user interacted by means of a single hand, but then he tried both hands and this made easier to accomplish his tasks. I interviewed him just after he finished, and he clearly seemed to be an enthusiast of such system; he stated that he will use it again in the future".*
>
> [...]
>
> *"A girl used the interface and seemed to be quite skilled in using it. She immediately used both hands. In the follow-up interview, she confirmed that she found the interface intuitive and easy-to-use".*

It is important to note that not all the users want to (or can) use both hands. For instance, a user could have a hand full for carrying bags or other stuff, or she may just prefer to stick on a single hand. Interestingly, during the observations some users decided to use only a single hand; all these users assessed that they did not like touchless interactions, preferring touch-based ones:

> *"She used the interface by means of gestures [...]. After a brief interaction session, she explained [...] that she would prefer to use a more traditional touchscreen. She used mostly the right hand during the whole interaction session. As time passes, it is increasingly clear to me that using a single hand is improper for this interface, and it is certainly less practical".*

According to the above considerations, the proper use of ABaToGI seems to be based on bimanual interactions. This, however, is in contrast with some users' preferences, who sometimes prefer to "stick to one hand" (as noted also by Walter et al. in [41]). The choice of redesigning the whole interface for correctly supporting single-handed interactions should be made by analyzing pros and cons of both the paradigms, which can depend on the specific use case. For instance, bimanual interactions can make the interacting users more "visible" from passers-by, and this could increase the probability that a honeypot effect would start, or more simply the number of glances toward the display. These considerations may be very useful for an advertising company, while may result useless from a usability point of view.

4.3.3 Honeypot Effect and Novelty Factor

During the observations, the experimenter noted several times that users behavior contributed to some sort of honeypot effect. Sometimes it happens spontaneously, just because interactions by a user attracted others passers-by; in other cases, the experimenter acted as a user, and then elicited the honeypot effect. This is the case of the following excerpt from the experimenter's notes:

> *"No users had been approaching the display for 7 minutes, so I decided to interact with it in order to attract some user's attention. After less than 30 seconds of interaction, I noticed a user approaching me and looking at the interface. He started asking me about the interface, and I explained him that it was touchless (despite he had clearly understood this capability by observing me). While the user was ending his interaction session, a girl (who had been looked at the display for about a minute) approached the display and tried to use large body gestures in front of the display".*

The experimenter also noted that users sometimes came back to the display, for using it again. However, it seems that they came back because of the novelty factor [30], instead of because of an actually perceived need of gathering information from the display. The following excerpt, for instance, shows that users go to the display for discussing the interaction modalities (i.e. touchless vs. touch), which clearly has nothing in common with the information offered by the display:

> *"Two users approached the display, started interacting via gestures and they seemed quite skilled. I then approached them and discovered that they already knew the display, because they have used it the day before. Both users showed their appreciation for the novel interface and started talking about the gestures as a useful alternative interactive mean.*
>
> *While we were talking about the interface, a third user stepped in our discussion and assessed that using a touchscreen would be better for this display as it is an easier interaction mean. All the three users continued discussing the interaction modalities for a couple of minutes and then went away".*

There was the need of investigating if users were interested in the display mainly because of the novelty factor, or because an actual interest in gathering the provided information. To this end, during the first 7 days the lectures timetables available on the display were intentionally outdated. This choice was taken in order to check if some users complain about the incorrect information. The experimenter recorded three users complaining about this incorrectness, over a total of four users that read the timetables not just to try this functionality, but to obtain such information. This confirmed that not all the users were using the display just for curiosity, and that the wrong information was correctly transferred to the end users.

4.3.4 Multiple Users

For concluding this Section, it is useful to describe some additional issues noted during the observations.

As stated multiple times before in this thesis, ABaToGI shows an Avatar that continuously replays the movement of a user in front of the display, using Kinect skeletal data. With a frequency of 30 frames per second, the Microsoft API are able to recognize up to six distinct users skeletons, and one of the issue that

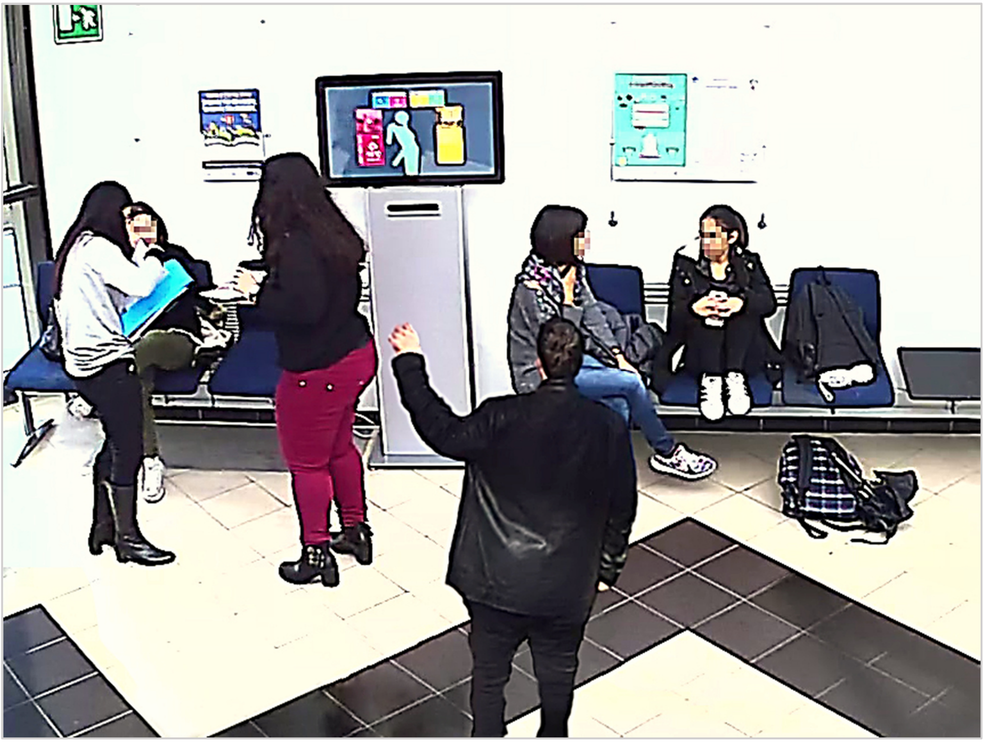


Figure 32. Wrong selection of the tracked skeleton making an active user unable to interact.

emerge from this functionality is how to select the skeleton to be used for animating the Avatar when multiple users are tracked. This has been solved by selecting the nearest one in relation with the Kinect position, and in most of the situations this seems to be a good choice. However, sometimes users tried to interact from distance while another passive user (i.e. inside the FOV of the Kinect, but not interested in its content) was actually the nearest one (see Figure 32). In such situation, interactions became frustrating and, often, not possible at all. In general, it seemed that the field of action of the sensor is too large for this use case, as explained by the experimenter in the following excerpt:

> *"The Avatar seems to appear shortly whenever someone passes in front of the display, also if they are not interested and still quite far from the screen. I had also noted a similar issue yesterday. Probably the system must be improved by limiting the field of action of the interface".*

Figure 33. Display is made inaccessible by the crowd.

The above issue is further exacerbated whenever the area inside the field of view of the Kinect sensor is particularly crowded (see Figure 33). During one of the observation day, the deployment place was used for hosting an event, and some tables were placed next to the display, in order to make stands for several companies to present their products. In this occasion, the display was still active and able to recognize people in front of it; however, the crowd did not allow users to interact with it. This was due mainly to the stands that partially hid the display, but also to the users next to the table that "confused" the skeletal tracking algorithm. The experimenter commented as follows:

> *"Today the building is very crowded, and it is impossible to use the interface. The Avatar seems to be unresponsive when there are too many people in front of it".*

More generally, the interface should communicate better which user is being tracked, and which ones are not. A possible solution could be, for instance, to draw some secondary (and smaller) Avatars behind the main one, so that users can see their own movements and figure out which of the users is tracked or not. This idea has something in common with the User Viewer control described by the Microsoft HIG [14].

4.4 Lessons Learnt

Using the findings derived from the observations described above, along with all the considerations and findings described in the previous Chapters, this Section summarizes some implications. The goal is to provide some guidelines for designing touchless gestural interfaces for public displays.

(1) *Using an Avatar-based interface helps in both communicating its interactive capabilities, and overcoming display blindness. However, showing an Avatar may not be enough.*

Using an Avatar has been shown to be helpful (as discussed also in Section 2.3.4); however, one should also take into account that, in some situations, users may be wrongly recognized by the software (or not recognized at all, i.e. not displaying any Avatar). Designers should thus try other solutions (together with the Avatar) for better communicating touchless interactivity (e.g. placing a marker on the floor to clarify the distance, or adding explicit instructions on the screen or next to the display).

(2) *Always take into account the actual capabilities of the sensorial devices used for gesture recognition in relation with typical passers-by use cases.*

If the gesture recognition capabilities are based on the use of Kinect-like devices, then designers should take into account all the possible walking directions of passers-by: if they may arrive from left or right, then the Avatar may not be displayed at all because of the limitation of the device (i.e. because users are too near to the RGB-D sensor). Using multiple cameras, pointing at different directions, may instead allow for a more robust Avatar visualization.

(3) *Prefer touch over touchless gestures when designing for reachable screens (i.e. touchable and placed at eye-height ones).*

According to the observations described in Section 4.3, such screens seem to afford mainly touch-based interactions. While several users enjoyed touchless interactions, observations have shown that some users discussed more about the novel interaction mean instead of the information provided by the display. This suggests that their interest might be strongly biased from the novelty effect. In addition, some users complained about the inability of using touch-based interactions as the main way for getting information from the system.

(4) *Use an Avatar-based interface allows for better user experiences when interactions must be based on the use of both hands/arms. However, remember that many users do not like to interact with both hands/arms.*

As stated in Section 4.3, not all the users want to use both hands: they can just prefer single-handed interactions, or have one hand busy for carrying a bag. The choice of designing for single-handed or bimanual interactions should be made by analyzing the specific use case, and thus the pros and cons of both the approaches.

(5) *Always make a touchless gestural interactive display accessible regardless of the number of users in front of it, supporting as much as possible crowded situations.*

Crowded situations like those depicted in Figure 33 make the display not accessible by any users. To avoid such problem, the display position has a fundamental role; however, while touch-based interactive displays always require being reachable, touchless capabilities allow for wider ranges of positions to keep information accessible. For instance, it would be possible to increase the display size and height and still maintain it visible and accessible. Furthermore, it is crucial to take into account situations where multiple users try to interact together (as observed in Section 4.3.2), also if the interface supports single user only: it is important to implement mechanisms to communicate which user is recognized as the active one, and which users are not.

Chapter 5
Conclusions

This thesis described the design of ABaToGI, an Avatar-based touchless gestural interface for public displays, and evaluated its features with several studies, in order to address different research questions.

At the very beginning, the initial design has been guided by a pilot study for gathering users' feedback on the interface. Such feedbacks have been thus translated in improvements on the visual interface, which in the end was based on the presence of an Avatar in the middle of the screen that replays user's movements. The interaction was supported by means of in-air direct manipulations, i.e. with no activation gestures: the user moves the arm in order to place the Avatar's hand on top of an interactive tile, and that tile is immediately activated.

The first version of ABaToGI was compared with another interface, developed using the guidelines provided by Microsoft Human Interface Guidelines (HIG). This comparison study revealed that:

- the use of **in-air direct manipulations** only (i.e. no activation gestures) allows for building interfaces to be robust to legacy-biased users, as well as for supporting immediate usability (as users do not need to guess a gesture). On the other hand, several users complained about the inability of using a "click", which implies that several of them preferred the HIG-based solution;
- the proposed Avatar-based interface seemed to elicit a greater number of **bimanual interactions** if compared to the HIG-based one. However, further investigations are certainly needed for gathering more valid results, because of the small size of users sample interviewed during the study;

- showing an Avatar in the middle of the screen that replays user's movements may help in **communicating touchless gestural interactivity**.

After this study, a better version of ABaToGI was developed and compared with a similar interface, which differs only for the absence of the Avatar (replaced by two hand-shaped cursors). This comparison study showed that the **interaction cognitive workload is lowered by the presence of the Avatar**; this result upholds the design choices made during the previous studies. Moreover, **bimanual interactions are generally perceived as a usability issue only if no Avatar is shown** on the interface; on the contrary, showing the Avatar seems to foster bimanual interactions as a more "natural" way for interacting with the system.

The last (and ongoing) study was based on the deployment of ABaToGI in-the-wild. Observing users' behaviors, and gathering their feedbacks through questionnaires and semi-structured interviews allowed to define a set of five **guidelines for designing touchless gestural interfaces for public displays**:

(1) Using an Avatar-based interface helps in both communicating its interactive capabilities, and overcoming display blindness. However, showing an Avatar may not be enough.

(2) Always take into account the actual capabilities of the sensorial devices used for gesture recognition in relation with typical passers-by use cases.

(3) Prefer touch over touchless gestures when designing for reachable screens (i.e. touchable and placed at eye-height ones).

(4) Use an Avatar-based interface allows for better user experiences when interactions must be based on the use of both hands/arms. However, remember that many users do not like to interact with both hands/arms.

(5) Always make a touchless gestural interactive display accessible regardless of the number of users in front of it, supporting as much as possible crowded situations.

5.1 Future Works

This thesis provides a common ground for future research on designing of touchless gestural interactions for public displays. Furthermore, the previous Chapters have described many open challenges to be tackled in further investigations. The goal of this chapter is to summarize such possible future work, both in the short and in the long term.

5.1.1 Improving ABaToGI

In Section 4.4, a series of guidelines have been described for designing gestural interfaces for public displays. Since those guidelines have been derived from users' behaviors while interacting with ABaToGI, as well as from the observed usability issues, the first possible future work is to improve the proposed interface by exploiting these guidelines.

At now, ABaToGI only supports single-user interaction. This may be considered acceptable in many cases, but there is the need for communicating which user is recognized as the interacting one (i.e. which user is to be considered "active" from the system's point of view). This improvement would be in line with guideline (5), as one of the main causes that discourage users from further interacting in crowded situations seems to be their inability to understand which user is the "active" one. To this end, users' silhouette (or other Avatars) should be displayed in the background, together with the main Avatar. While the latter would continue to be the only with interactive capabilities, other users should be able to understand much better if they are recognized as "active" or not. This idea of using users' silhouettes along with the Avatar is depicted in Figure 34.

Together with the above solution, and in order to address guideline (1) (i.e. "showing an Avatar may not be enough"), other contrivances should be used. A possibility would to evaluate the impact of placing a marker on the floor at

Figure 34. Multiple users support in future version of ABaToGI.

about 1.5 meters of distance from the display. This may help some users in understanding the proper distance for interacting with the display. At the same time, other explicit visual clues or messages may be shown on the interface, to communicate users about the proper distance for interacting. These different solutions, together with different degrees of Kinect visibility, should be thus compared among each other, for evaluating which one is the most effective for communicating touchless gestural interactivity in public.

5.1.2 Bimanual Interactions

The studies described in Chapter 2 showed a potential interesting feature of Avatar-based interfaces, i.e. the ability of fostering bimanual interactions. As those studies were based on a small users sample, and since the compared interfaces differed also for other characteristics (e.g. the activation gesture), another study was designed and conducted. The latter, described in Chapter 3, only showed that users perceived a lower workload when interacting with an Avatar, but did not show anything about its ability of fostering bimanual interactions. In the final study described in Chapter 4, however, users that interacted with two hands seemed generally more satisfied when used two hands.

With these results, it is impossible to conclude if bimanual interactions are fostered by Avatar-based interfaces or not. However, they might be fostered by the layout arrangement, or by the use/non-use of activation gestures. These different factors need to be tested and compared, and some of them (alone or in combination) could turn out to foster bimanual interactions.

Nonetheless, it would be very interesting to investigate if an interface that foster bimanual interactions may somehow increase the number of glances or (even better) interactions from passers-by/users. Since using two hands/arms, although less acceptable for some users, may result in wider and (thus) more "visible" gestures, it is reasonable to think in a growing possibility of fostering the honeypot effect around the users.

Clearly, all the above are just hypothesis, and many investigations are required for gathering valid results.

5.1.3 Audience Influence on Users

Another curious effect that could be investigated is about the specific deployment described in Chapter 3. As shown in Figure 35, during the study the display has been placed next to two benches, where usually students stay while waiting for lectures. This choice has been taken because that place is usually frequented by many people, so it seemed to be a good solution to test the system. However, the presence of people that look at a user may discourage her/him for starting to interact with the interface (see Figure 35a). Moreover, even if the user starts interacting, she may be discouraged by people looking at her/him.

From another point of view, curious users may be more encouraged in trying the system if no other people are on the benches looking at her/him (see Figure 35b).

(a)

(b)

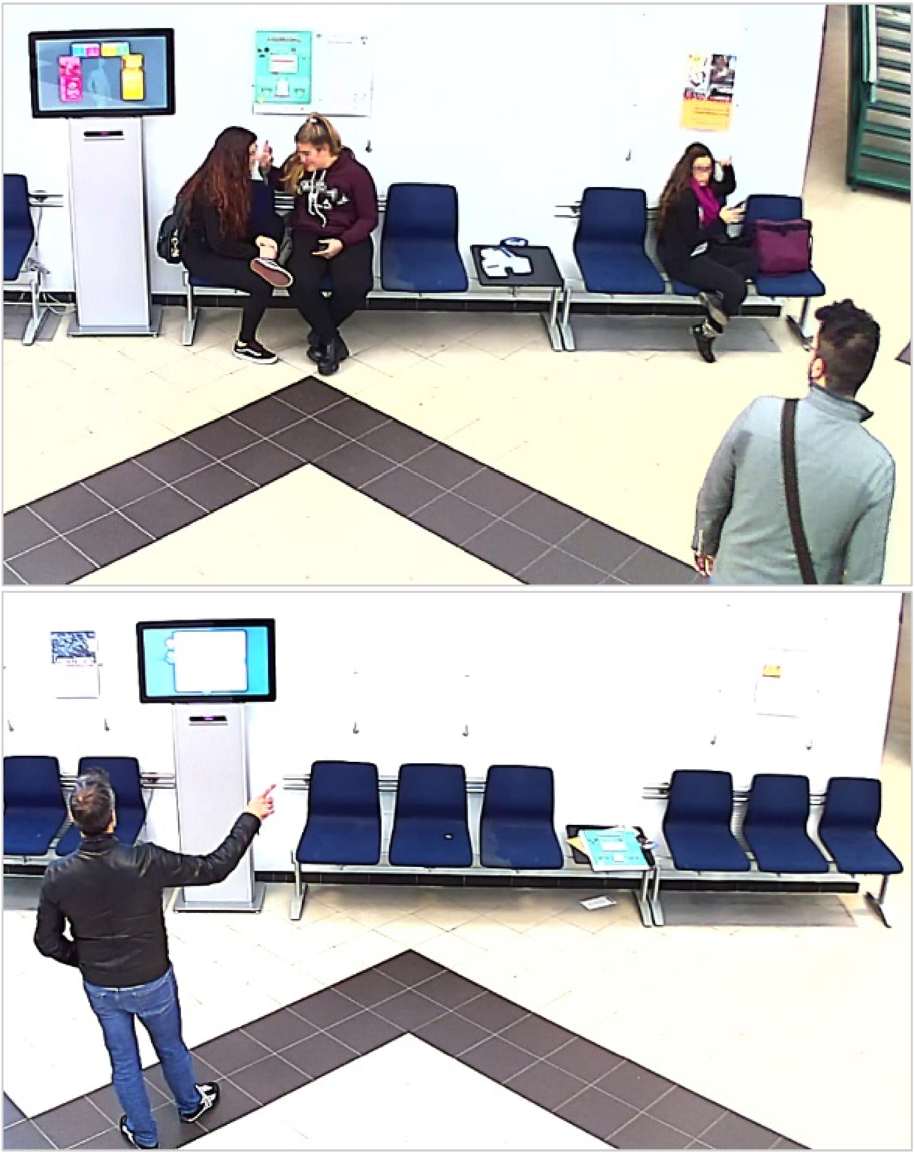


Figure 35. Audience next to the interactive display.

The above considerations suggest that a proper long-term study may help in evaluating possible relations between the presence and the number of persons sit on the benches, and users' performances while interacting.

5.1.4 Ongoing Deployment

As of today, the deployment described in Chapter 3 is still active, and data on users' behaviors are being collected and analyzed. This could produce several outcomes, and some of them have been briefly described above. However, it is reasonable to guess that some other outcomes may be discovered as part of an ongoing exploratory study on touchless gestural interaction in public.

One of the goal of the aforementioned ongoing study is also to get users' feedbacks on ABaToGI, and use them in order to improve the interface, as well as to refine and enhance the content accessible by the system.

References

[1] M. Weiser, "The computer for the 21st century," *Scientific american,* vol. 265, no. 3, pp. 94-104, 1991.

[2] V. Gentile, S. Sorce and A. Gentile, "Continuous Hand Openness Detection Using a Kinect-Like Device," in *Eighth International Conference on Complex, Intelligent and Software Intensive Systems (CISIS)*, Birmingham, UK, 2014.

[3] N. Davies, S. Clinch and F. Alt, Pervasive Displays: Understanding the Future of Digital Signage, Morgan & Claypool, 2014.

[4] V. Gentile, S. Sorce, A. Malizia, D. Pirrello and A. Gentile, "Touchless Interfaces For Public Displays: Can We Deliver Interface Designers From Introducing Artificial Push Button Gestures?," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, 2016.

[5] A. R. Hevner, "A Three Cycle View of Design Science Research," *Scandinavian Journal of Information Systems,* vol. 19, no. 2, pp. 87-92, 2007.

[6] S. T. March and G. F. Smith, "Design and natural science research on information technology," *Decision support systems,* vol. 15, no. 4, pp. 251-266, 1995.

[7] R. de la Barré, P. Chojecki, U. Leiner, L. Mühlbach and D. Ruschin, "Touchless Interaction-Novel Chances and Challenges," in *Human-Computer Interaction. Novel Interaction Methods and Techniques*, J. A. Jacko, Ed., Springer Berlin Heidelberg, 2009, pp. 161-169.

[8] S. Gehring and A. Wiethoff, "Interaction with Media Façades," *Informatik-Spektrum,* vol. 37, no. 5, pp. 474-482, 2014.

[9] P. Cremonesi, A. Di Rienzo, C. Frà, F. Garzotto, L. Oliveto and M. Valla, "Personalized interaction on large displays," in *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces - AVI '14*, 2014.

[10] M. H. Haeusler, Media facades: history, technology, content., Ludwigsburg: Avedition, 2009.

[11] K. Kuikkaniemi, G. Jacucci, M. Turpeinen, E. Hoggan and J. Müller, "From Space to Stage: How Interactive Screens Will Change Urban Life," *Computer,* vol. 44, no. 6, pp. 40-47, 2011.

[12] C. P. Gerba, A. L. Wuollet, P. Raisanen and G. U. Lopez, "Bacterial contamination of computer touch screens," *American Journal of Infection Control,* vol. 44, no. 3, p. 358–360, February 2016.

[13] S. Sorce, A. Malizia, V. Gentile and A. Gentile, "Touchless gestural interfaces for networked public displays: overcoming interaction blindness and performing evaluations in-the-wild," in *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, Osaka, Japan, 2015.

[14] Microsoft, [Online]. Available: http://download.microsoft.com/download/6/7/6/676611B4-1982-47A4-A42E-4CF84E1095A8/KinectHIG.2.0.pdf.

[15] J. Lazar, J. H. Feng and H. Hochheiser, Research Methods in Human-Computer Interaction, Wiley & Sons Ltd, 2010.

[16] M. Brynskov, P. Dalsgaard, T. Ebsen, J. Fritsch, K. Halskov and R. Nielsen, "Staging Urban Interactions with Media Facades," in *INTERACT*, 2009.

[17] P. Dalsgaard and K. Halskov, "Dynamically transparent window," in *CHI EA '09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, 2009.

[18] J. Fritsch and P. Dalsgaard, "Media Facades Beyond Interaction," in *Proceedings of OzCHI 2008*, 2008.

[19] J. Müller, D. Eberle and K. Tollmar, "Communiplay: a field study of a public display mediaspace," in *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, 2014.

[20] J. Hardy, E. Rukzio and N. Davies, "Real world responses to interactive gesture based public displays," in *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia - MUM '11*, 2011.

[21] V. Gentile, S. Sorce, G. Russo, D. Pirrone and A. Gentile, "A Multimodal Fruition Model for Graphical Contents in Ancient Books," in *Proceedings of the 17th International Conference on Computer Systems and Technologies 2016 (CompSysTech '16)*, Palermo, Italy, 2016.

[22] J. O. Wobbrock, M. R. Morris and A. D. Wilson, "User-defined gestures for surface computing," in *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, 2009.

[23] P. Koutsabasis and C. K. Domouzis, "Mid-Air Browsing and Selection in Image Collections," in *International Working Conference on Advanced Visual Interfaces - AVI 2016*, 2016.

[24] R.-D. Vatavu, "User-Defined Gestures for Free-Hand TV Control," in *Proceedings of the 10th European conference on Interactive tv and video - EuroiTV '12*, 2012.

[25] A. van Dam, "Post-WIMP user interfaces," *Communications of the ACM,* vol. 40, no. 2, pp. 63-67, February 1997.

[26] M. Ringel Morris, A. Danielescu, S. Drucker, D. Fisher, B. Lee, C. Schraefel and J. O. Wobbrock, "Reducing legacy bias in gesture elicitation studies," *Interactions,* vol. 21, no. 3, pp. 40-45, 2014.

[27] A. Köpsel and N. Bubalo, "Benefiting from legacy bias," *Interactions,* vol. 22, no. 5, pp. 44-47, 2015.

[28] V. Gentile, A. Malizia, S. Sorce and A. Gentile, "Designing Touchless Gestural Interactions for Public Displays In-the-Wild," in *Human-Computer Interaction: Interaction Technologies*, M. Kurosu, Ed., Springer International Publishing, 2015, pp. 24-34.

[29] J. Müller, D. Wilmsmann, J. Exeler, M. Buzeck, A. Schmidt, T. Jay and A. Krüger, "Display Blindness: The Effect of Expectations on Attention towards Digital Signage," in *Proceedings of the 7th International Conference on Pervasive Computing Pervasive '08*, 2009.

[30] T. Ojala, V. Kostakos, H. Kukka, T. Heikkinen, T. Linden, M. Jurmu, S. Hosio, F. Kruger and D. Zanni, "Multipurpose Interactive Public Displays in the Wild: Three Years Later," *Computer,* vol. 45, no. 5, pp. 42-49, 2012.

[31] J. P. Benway, "Banner blindness: The irony of attention grabbing on the World Wide Web," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1998.

[32] E. M. Huang, A. Koster and J. Borchers, "Overcoming assumptions and uncovering practices: When does the public really look at public displays?,"

in *Proceedings of the 6th International Conference on Pervasive Computing - Pervasive '08*, 2008.

[33] R. Walter, G. Bailly and J. Müller, "StrikeAPose: revealing mid-air gestures on public displays," in *CHI '13 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Paris, France, 2013.

[34] T. Ballendat, N. Marquardt and S. Greenberg, "Proxemic interaction: designing for a proximity and orientation-aware environment," in *ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)*, 2010.

[35] D. Vogel and R. Balakrishnan, "Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users," in *Proceedings of the 17th annual ACM symposium on User interface software and technology (UIST '04)*, 2004.

[36] J. Müller, F. Alt, D. Michelis and A. Schmidt, "Requirements and design space for interactive public displays," in *Proceedings of the 18th ACM international conference on Multimedia (MM '10)*, Firenze, Italy, 2010.

[37] J. Schönböck, F. König, G. Kotsis, D. Gruber, E. Zaim and A. Schmidt, "MirrorBoard – An Interactive Billboard," in *Proceedings of Mensch und Computer 2008*, 2008.

[38] J. Müller, G. Bailly, T. Bossuyt and N. Hillgreen, "MirrorTouch: Combining Touch and Mid-air Gestures for Public Displays," in *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14*, Toronto, Canada, 2014.

[39] J. Müller, R. Walter, G. Bailly, M. Nischt and F. Alt, "Looking glass: a field study on noticing interactivity of a shop window," in *Proceedings of*

the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12, 2012.

[40] I. Elhart, F. Scacchi, E. Niforatos and M. Langheinrich, "ShadowTouch: a Multi-user Application Selection Interface for Interactive Public Displays," in *Proceedings of the 4th International Symposium on Pervasive Displays - PerDis '15*, 2015.

[41] R. Walter, G. Bailly, N. Valkanova and J. Müller, "Cuenesics: using mid-air gestures to select items on interactive public displays," in *16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14*, 2014.

[42] D. Wigdor and D. Wixon, Brave NUI World: Designing Natural User Interfaces for Touch and Gesture, Morgan Kaufmann Publishers Inc., 2011.

[43] E. Hutchins, J. Hollan and D. Norman, "Direct Manipulation Interfaces," *Human-Computer Interaction,* vol. 1, no. 4, pp. 311-338, 1985.

[44] A. Frank, "dontclick.it," [Online]. Available: http://dontclick.it. [Accessed 17 March 2016].

[45] A. Dix, J. Finlay, G. D. Abowd and R. Beale, Human Computer Interaction, III ed., Upper Saddle River: Prentice-Hall, 2003.

[46] B. Kules, H. Kang, C. Plaisant, A. Rose and B. Shneiderman, "Immediate usability: a case study of public access design for a community photo library," *Interacting with Computers,* vol. 16, no. 6, pp. 1171-1193, 2004.

[47] D. Grijincu, M. A. Nacenta and P. O. Kristensson, "User-defined Interface Gestures," in *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces - ITS '14*, 2014.

[48] T. Seyed, C. Burns, M. Costa Sousa, F. Maurer and A. Tang, "Eliciting usable gestures for multi-display environments," in *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces - ITS '12*, 2012.

[49] P. Koutsabasis and C. K. Domouzis, "Mid-Air Browsing and Selection in Image Collections," in *AVI '16 Proceedings of the International Working Conference on Advanced Visual Interfaces*, 2016.

[50] H. Brignull and Y. Rogers, "Enticing people to interact with large public displays in public spaces," in *Proceedings of the IFIP International Conference on Human-Computer Interaction (INTERACT 2003)*, 2003.

[51] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Advances in Psychology,* vol. 52, p. 139–183, 1988.

[52] S. G. Hart, "NASA-task load index (NASA-TLX); 20 years later," in *Proceedings of the human factors and ergonomics society annual meeting*, 2006.

[53] F. Alt, S. Schneegaß, A. Schmidt, J. Müller and N. Memarovic, "How to evaluate public displays," in *Proceedings of the 2012 International Symposium on Pervasive Displays - PerDis '12*, 2012.

[54] R. Johnson, Y. Rogers, J. van der Linden and N. Bianchi-Berthouze, "Being in the thick of in-the-wild studies: The Challenges and Insights of Researcher Participation," in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, 2012.

[55] R. Sharp and K. Rehman, "The 2005 UbiApp Workshop: What Makes Good Application-Led Research?," *IEEE Pervasive Computing,* vol. 4, no. 3, pp. 80-82, July 2005.

[56] M. Jurmu, "Towards engaging multipurpose public displays," PhD Thesis, Oulu, 2014.

[57] A. Chamberlain, A. Crabtree, T. Rodden, M. Jones and Y. Rogers, "Research in the wild: understanding 'in the wild' approaches to design and development," in *Proceedings of the Designing Interactive Systems Conference on - DIS '12*, 2012.

[58] T. Keskinen, J. Hakulinen, T. Heimonen, M. Turunen, S. Sharma, T. Miettinen and M. Luhtala, "Evaluating the experiential user experience of public display applications in the wild," in *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia - MUM '13*, 2013.

[59] J. Messeter and D. Molenaar, "Evaluating ambient displays in the wild: highlighting social aspects of use in public settings," in *Proceedings of the Designing Interactive Systems Conference on - DIS '12*, 2012.

[60] U. Hinrichs and S. Carpendale, "Gestures in the wild: studying multi-touch gesture sequences on interactive tabletop exhibits," in *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, 2011.

[61] Team Viewer, 2016. [Online]. Available: https://www.teamviewer.com/.

[62] G. Beyer, F. Alt, J. Müller, A. Schmidt, K. Isakovic, S. Klose, M. Schiewe and I. Haulsen, "Audience behavior around large interactive cylindrical screens," in *CHI '11 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011.

Appendix

A. Interview Guide for Pilot and Comparison Studies

USER ID: ____________________ Date: ________________________

Interface

☐ Avatar-based ☐ HIG-based

Setup

1. Display size:

   W: __________________ cm H: ____________________ cm

2. Is the Kinect on top or bottom of the display? ☐ Top ☐ Bottom

Privacy

3. May I record your depth and audio data?

   ☐ Yes ☐ No

4. If yes, may I also record a video of you interacting with the system?

   ☐ Yes ☐ No

Personal and background information

5. Sex:

   ☐ Male ☐ Other

   ☐ Female ☐ Unspecified

6. What's your job? ____________________

   a. If it is a research-related one, which is your research area?

   ____

   ____

7. Are you right-handed, left-handed or ambidextrous? __________

   b. If ambidextrous, do you had a preferred hand for interacting?

   ____

      i. If yes, which one? ____________________

8. Have you ever had previous experiences with gestural interfaces? ______

   - If yes, describe them: ____________________

     ____

Interaction Stage: time-measured tasks

Task	Time to task
9. There is a news about “________”. Find it, read it and sum it up to me.	
10. Go back, find university information, read them and sum them up to me.	
11. Find the lecture timetable of room 7, and sum it up to me.	
12. Play the video	
13. Tell me the weather forecast for tomorrow (just the temperature)	

Interaction Stage: additional tasks

14. Interact with the system *ad lib*, and talk aloud about the system, according to your thoughts and feelings.

Other questions

15. Did you know that this system is based on touchless gestural input before starting the test? ____

    - If no, have you guessed that it was based on gestures? __________

i. If yes, which hints have suggested you that the system was/wasn't gestural? (e.g.: display size, presence of Kinect, the avatar on the screen...)

____

____

16. What about responsiveness of the interface? Was it too slow in reproducing your movements? Have you noticed something strange while interacting?

____

____

17. Are there some other tasks that you would like to perform by using this system?

____

____

18. Did you miss the touchscreen?

____

____

19. Any other suggestions or ideas to improve this system?

Additional notes

B. Extended NASA Task Load Index (NASA-TLX)

The following questionnaire is an Italian version of the NASA-TLX questionnaire [51]; the red labels are the original ones in English.

Carico Mentale — Quanto è stato mentalmente impegnativo il compito?
Mental Demand — How mentally demanding was the task?

Molto basso Very Low — Very High Molto alto

Carico Fisico — Quanto è stato fisicamente impegnativo il compito?
Physical Demand — How physically demanding was the task?

Molto basso Very Low — Very High Molto alto

Carico Temporale — Quanto è stato frenetico il ritmo di esecuzione del compito?
Temporal Demand — How hurried or rushed was the pace of the task?

Poco Very Low — Very High Molto

Performance — In che modo è stato svolto il compito?
Performance — How successfull were you in accomplishing what you were asked to do?

Perfettamente Perfect — Failure Malissimo

Sforzo — Quanto è stato difficile svolgere il compito?
Effort — How hard did you have to work to accomplish your level of performance?

Poco Very Low — Very High Molto

Frustazione — Quanto ti sei sentito insicuro, scoraggiato, irritato e/o stressato?
Frustration — How insecure, discouraged, irritated, stressed, and annoyed were you?

Poco Very Low — Very High Molto

The above questionnaire has been extended by adding the following questions (here listed in English only), which were specific for evaluating the influence of Avatar in fostering bimanual interactions.

- Have you used both hands (even not simultaneously) while interacting with the interface?

  ☐ Yes ☐ No

  - If yes, have you used two hands because of a usability issue of the interface, or just because you prefer to use two hands?

    ☐ Usability Issue ☐ Prefer Two Hands

  - If yes, would you rather prefer to be able to use one hand only for the whole interaction session?

    ☐ Yes ☐ No