Investigating how User Avatar in Touchless Interfaces Affects Perceived Cognitive Load and Two-Handed Interactions

Vito Gentile[1], Salvatore Sorce[1], Alessio Malizia[2], Fabrizio Milazzo[1], Antonio Gentile[1]

[1] Ubiquitous Systems and Interfaces Group

Dipartimento dell'Innovazione Industriale e Digitale (DIID) – Università degli Studi di Palermo

Viale delle Scienze, edificio 6 – 90128 Palermo, Italy

{firstname.lastname}@unipa.it - http://usi.unipa.it

[2] Human Centred Design Institute (HCDI) – Brunel University London

UB8 3PH, Uxbridge, Middlesex, United Kingdom

Alessio.Malizia@brunel.ac.uk

ABSTRACT

In recent years, touchless-enabling technologies have been more and more adopted for providing public displays with gestural interactivity. This has led to the need for novel visual interfaces aimed at solving issues such as communicating interactivity to users, as well as supporting immediate usability and "natural" interactions. In this paper, we focus our investigation on a visual interface based only on the use of in-air direct manipulations. Our study aims at evaluating whether and how the presence of an Avatar that replays user's movements may decrease the perceived cognitive workload during interactions. Moreover, we conducted a brief evaluation of the relationship between the presence of the Avatar and the use of one or two hands during the interactions. To this end, we compared two versions of the same interface, differing only for the presence/absence of the user's Avatar. Our results showed that the Avatar contributes to lower the perceived cognitive workload during the interactions.

Author Keywords

Touchless Gestural Interfaces; Public Displays; Interface Evaluation.

ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces.

INTRODUCTION

According to the definition by de la Barré et al. [1], "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, is not considered touchless interaction. On the contrary, eye trackers or Kinect-like devices [2] have been widely accepted as valid examples of devices that enable for touchless interactions.

Despite the wide adoption of touchscreens as principal I/O devices for public displays, new interaction modalities have emerged to fulfill specific needs. For instance, the increasing number of interactive *media façades*, defined as installations in which displays are integrated into architectural structures [3], have implied the need of interacting from distance, and without any physical input device. Many authors proposed touchless interaction methods based on the detection of users' position and their body movements, as well as by using gestures or mobile devices (see for instances Aarhus by Light [4], Dynamically Transparent Window [5] and Climate Wall [6]).

Nevertheless, touchless interfaces have been more rarely studied for smaller and more traditional displays (ranging from TV- to billboard-sized screens). One of the reason is that those displays naturally afford touch-based interactions even when mid-air gestures are supported [7]. This contributes to worsening several typical issues of public displays, such as the need of communicating interactivity, as well as supporting immediate usability [8].

Although the above considerations may discourage designers to build touchless interfaces, there are many advantages in supporting them. For instance, supporting interaction from a distance when the display is not reachable may allow wheelchair users to access otherwise inaccessible information. Moreover, displays may be placed in not reachable locations in order to prevent vandalisms.

In this paper, we focus on a particular touchless gestural interface, based on the use of only in-air direct manipulations (i.e. without specific activation gestures).

The goal here is to understand how the presence of an Avatar that replays user's movements may influence two aspects of the interaction: 1) the perceived cognitive workload, and 2) the use of one or two hands.

Our results show that displaying an Avatar may help in reducing the cognitive load, by increasing the perceived performance and reducing effort and frustration levels. Moreover, we found that the presence of the Avatar allows users to perceive the use of both hands during the interactions as more "natural".

The rest of the paper is organized as follows. The next Section provides an overview of the related works, presenting a short list of selected solutions for user representation in public displays and the use of one or two hands. Then, a description of the interfaces designed for the aims of this paper is provided, followed by a detailed overview of the comparison study. Results are thus described, followed by a proper discussion. Finally, a summary of the findings, along with some possible future works, conclude the paper.

RELATED WORKS
Many authors have studied the design of visual interfaces for supporting touchless gestural interactions for public displays. This Section summarizes some prior works, organizing them according to the objectives of this paper. In particular, here we focus on how the visual representation of the users influences their behavior, and also on the use of one or two hands during the interactions.

User Representation
The idea of representing the user as an Avatar (or similar variations) has been implemented in many public display touchless applications. This is the case of StrikeAPose [9], as well as MirrorTouch [7], where the use of mid-air gestures was coupled with touchscreens. In both these cases, user representation consists in displaying user's silhouette (based on depth information obtained from a Kinect sensor), by means of which it was possible to interact with the available graphical items.

The advantages of using such users' representations have been investigated in many prior works, reflecting the benefits of the mirror mental model. In [10], authors explain how extending this metaphor to public displays may help in communicating interactivity to users and facilitating implicit interactions.

A more detailed investigation on specific types of "Avatars" for representing users has been conducted by Müller et al. [11]. In their work, authors measured how the number of elicited interactions varies if the interface uses user's silhouette (based on depth data, as seen before) instead of user's whole body picture. Interestingly, they noticed that the more abstract the user representation, the lower the number of interactions. Although these results may seem useful for overcoming the so-called interaction blindness [12], the main drawback of using users' pictures in public relates to the perceived privacy issues [11] [8].

While the use of an Avatar has been shown to be useful in communicating interactivity, here we focus on the benefits of such approach in order to reduce the perceived cognitive workload.

Single-handed Vs Two-handed Interactions
Most of the aforementioned interfaces allow for interactions with both single and two hands. However, in many cases, these possibilities have not been investigated in depth.

Some gesture elicitation studies have documented the differences in users' preference between single-handed and two-handed interactions. Even if the focus of this paper is on touchless gestures, here it is worth considering also some noticeable findings related to other interaction modalities (e.g. the touch-based ones).

In [13], 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, 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 [14] [15].

Considering now touchless gestural interaction, an interesting elicitation study has been described by Koutsabasis and Domouzis in [16]. 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 [17]. 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 on 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. the so-called legacy bias [18]).

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 [19]. 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 at a time, showing one single hand-shaped cursor at a time.

However, a designer may still have the need of fostering interactions with two hands, 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: users would change arm more frequently, so the fatigue could be more equally distributed between both arms. Another advantage may relate to the *honeypot effect* [20], 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. Indeed, if users interact with two hands, their movements are more visible by other users in the surrounding environment, and this may result in a higher probability the honeypot effect may occur.

INTERFACES DESCRIPTION

We based our investigation on a comparison of two similar visual interfaces, aimed at supporting touchless gestural interactions. In particular, both the interfaces (and the supported tasks) were designed for being deployed in a public display, situated inside a building at the university campus in Palermo (Italy).

The first interface is built upon a previous work described in [21], consisting of an Avatar placed in the middle of the screen that replays user's movements, with all the other interface components arranged all around it (see Figure 1a). In this case, the Avatar appears whenever a user approaches the display, and remains permanently present in the middle of the screen, continuously replaying user's movements.

We used this interface layout as a basis to produce a second interface, in which we kept the tiles arrangement and functionality, and replaced the entire Avatar with two hand-shaped cursors only (see Figure 1b). In the following, the interface that includes the Avatar will be referred as "interface A", while the other one will be referred as "interface B".

In both cases, a user can trigger the interaction events just by driving the Avatar's hands (interface A) or the hand-shaped cursors (interface B) and placing them on top of the available tile-shaped components – with no activation gestures. As soon as the visual representation of a user's hand enters the area of an interactive tile, the corresponding event is immediately triggered. In other words, the user can interact with the interface using only *in-air direct manipulations*, allowing designers to avoid the adoption of symbolic gestures to trigger events. This way a user should better guess and learn how to interact by herself since there is no need of any training about specific activation gestures.

Both interfaces allow for one event at a time, no simultaneous activations are possible. This means that users can use equally one or two hands during the interaction, but
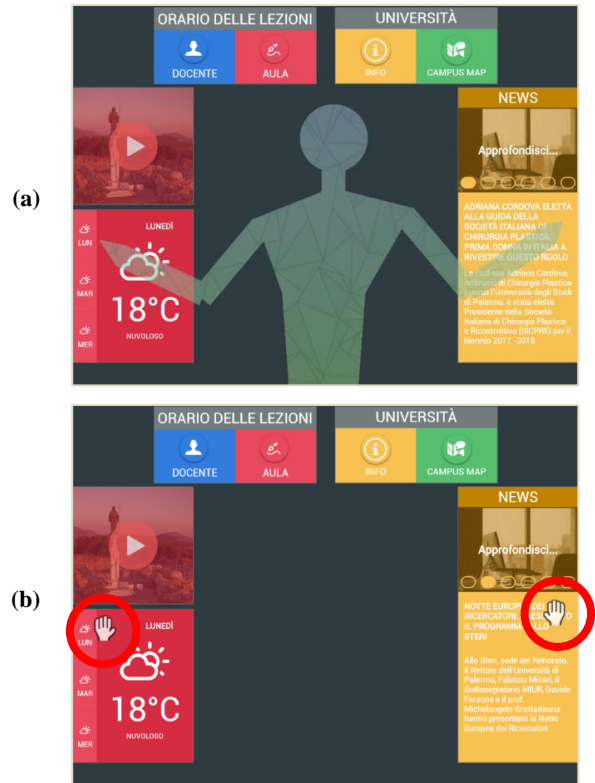
Figure 1. Visual comparison between the interfaces.

there will always be only one hand to trigger an event, that is the first detected one in temporal order.

A suitable animation acts as a visual feedback to communicate the event activation. All the animations last less than 1 second, and they consist in a zoom-in or zoom-out transition depending on whether the navigation event leads to a deeper layer in the information tree or not.

Both interfaces allow people to accomplish the same tasks, and in particular:

- read weather information;
- read general information about the University;
- read and navigate some news about the University;
- play a promotional video of the University;
- access and navigate a University campus map.

All the visual components that allow for the above tasks are accessible from the main page of the interface (Figure 1).

STUDY DESCRIPTION

We used the two interfaces described in the previous Section to conduct a twofold comparison study.

The first goal of this study was to compare the perceived cognitive workload between the two interfaces, in order to understand how it relates to the presence/absence of the Avatar.
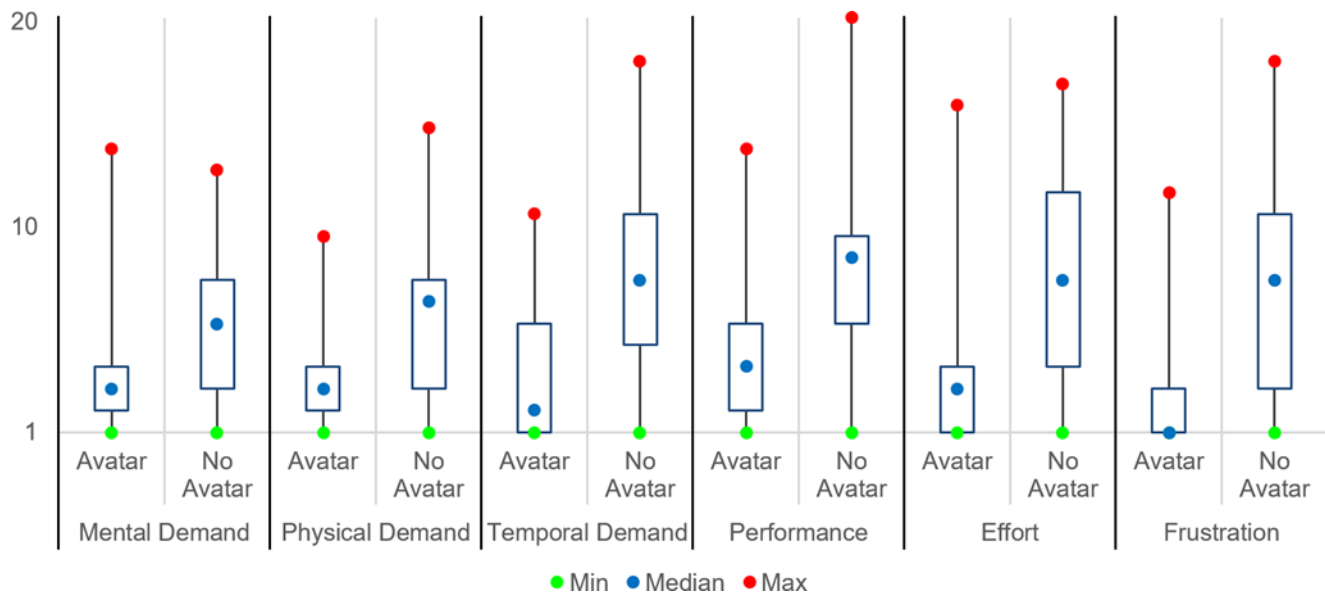
Figure 2. Results from NASA-TLX data.

The second goal was to understand if there exists a relationship between the users' preferences in using one or two hands while interacting via touchless gestures, and the presence/absence of their Avatar in the middle of the screen.

To this end, we asked 50 users to interact with the two interfaces described before. Since the interfaces are designed to be deployed in a University foyer, we recruited the users among our friends, colleagues, and by word of mouth. Then, we selected the 50-users sample according to the expected audience composition in a corresponding real setting. In more details, in order to achieve the best significance, our sample was arranged as follows:

- gender: 31 males, 19 females;
- age: 25 in the range 20-24 y.o., 16 in the range 25-35 y.o., 9 in the range 36-50 y.o.;
- occupation: 6 lecturers/professors, 40 students, 4 staff;
- users with experience in gestural interactions: 22.

All of them agreed to participate in our study on a friendly base.

We conducted the study in a controlled environment, i.e. our laboratory at the University. We are aware that in this case the ecological validity of the results is somehow decreased [8]. However, since the aim was to investigate the effect of the interface layout on two interaction issues, we reasonably assume that other external factors (typical of uncontrolled environments) should equally affect both the conditions we tested. For these reasons, the results can still be considered sufficiently valid.

The system to interact with was a mock-up of a typical public display. In particular, it consisted in a 42-inch LCD monitor placed at eye height, a laptop, and a Microsoft Kinect sensor placed right below the screen and clearly visible to all users. We asked users to interact from a distance of about 1.7 meters from the display in order to match the best recognition conditions of the sensor.

For our study, we used a *between-subjects* setup [22], i.e. 25 users interacted only with interface A, and the remaining 25 interacted only with interface B. In the two sub-groups, we kept the same age/gender/occupation distribution of the whole group, rounding the numbers to the nearest integer.

Despite we were aware of 22 users with previous experiences in gestural interactions, we did not care about this or other technology skills, since the focus of our study was set on interface-related issues. This is the reason why before each session, we informed the user about the touchless interaction modality.

In each interaction session, we asked the user to perform the following tasks:

- find and read a specific news;
- find and read university information;
- find the timetable for a specific class;
- play a video;
- find and read the weather forecast for the next day.

We used these tasks as a mean for our study. In other words, for the purposes of this study we were not interested in the success rate, so here we do not report any detail about the interaction paths users followed to accomplish the tasks. We only observed their behavior during the interaction sessions for the goals described above.

Anyway, all users successfully accomplished the tasks they were assigned. Despite the interaction time was not constrained, we noted that the longest session took 6m20s, and the shortest took 1m32s, with an average of 3m51s.

At the end of the interaction session, each user was asked to fill in a NASA-TLX questionnaire [23], used as a "Raw TLX" [24] (i.e. by eliminating the weighting process of the subscales and then analyzing them individually). The NASA-TLX questionnaire allows for measuring the perceived cognitive workload in six 20-points Likert subscales, meant to measure specific aspects of the related tasks. For reader's convenience, we recap below these aspects, along with a brief explanation:

- *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.

Furthermore, in order to understand why people who interacted with two hands decided to do this, we added three simple two-choice questions focused on the use of one or two hands for the interaction session:

Q1. Have you used both hands while interacting with the interface?

[Answers: *Yes | No*]

Q2. (only for users who asked *Yes* to Q1): Have you used two hands because of a usability issue of the interface, or just because you prefer to use two hands?

[Answers: *Usability Issue | Prefer Two Hands*]

Q3. (only for users who asked *Yes* to Q1): would you rather prefer to be able to use only one hand for the whole interaction session?

[Answers: *Yes | No*]

Question Q1 is self-explaining. With question Q2, we would understand whether the use of two hands was perceived as the natural way to interact with the interface (*Prefer Two Hands*) or it was perceived as uncomfortable but unavoidable, e.g. due to the interface layout (*Usability Issue*). The aim of question Q3 was to understand whether people perceived the use of two hands as uncomfortable or avoidable. It is worth noting that both questionnaires were submitted by experimenters and filled in under their supervision, thus users could ask for explanations about the questions before to express their answers.

RESULTS

Figure 2 allows for comparing the quantities measured by means of the NASA-TLX questionnaires. 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 Wilcoxon-Mann-Whitney 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 interface ($p < 0.04$). Generalizing these results, a plausible conclusion is that displaying an Avatar can decrease the perceived workload when interacting with a touchless gestural interface.

Regarding the three questions about the interaction with one or two hands, during our tests all the users decided to interact with both hands (i.e. all 50 users answered *Yes* to question Q1). In more details, some users moved the both hands together, whereas some others used one hand for a while and then the other one, even with multiple switches from one hand to another. We think that this result might be due to the interfaces layout. Indeed, considering the interfaces layout (Figure 1), and that they are shown on a 42-inch monitor, reaching tiles on the left upper side can turn out to be very difficult with only the right hand, and the same for the opposite tiles with the left hand. It is worth noting that even in the case of two-handed interactions, users could only activate one tile at a time.

Consequently, we expected that most users would have answered the question Q2 with *Usability Issue*, and the question Q3 with *Yes*. Conversely, we had a different distribution. Table and Table summarize users' answers to questions Q2 and Q3. A Chi-squared test confirmed the statistical correlation between the interface used and users' responses ($p < 0.05$ for both the contingency tables).

		Why used two hands?	
		Usability Issue	Preference
Interface	A	6	19
	B	20	5

Table 1. Contingency table for question Q2.

		Would prefer to use a single hand?	
		Yes	No
Interface	A	6	19
	B	16	9

Table 2. Contingency table for question Q3.

DISCUSSION

The results of NASA-TLX questionnaires presented in the previous Section show that using an Avatar decreases the perceived cognitive workload if compared with the same interface with only hand cursors. These findings are in line and enforce what Müller et al. noted in [11]: not only the Avatar (or a user's silhouette) increases the elicited interactions, but also it seems to help in making the interface more easy-to-use (Figure 2: *mental demand*, *performance*, *effort*, *frustration*) and less tiring (Figure 2: *physical demand*, *temporal demand*, *effort*).

Furthermore, it seems that users used both hands as a more natural and intuitive way for interacting with the interface A. This 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 who interacted with interface B judged the need for two-handed interactions as a usability issue (see Table). In addition, most of the users that interacted with interface A assessed that they would not prefer to use a single hand (see Table). We think that this result may partly rely on the fact that the Avatar shows the full body postures, even the less natural ones. In other words, if a user tries to activate a tile in the upper-left side of the interface with the right hand, the Avatar will show the right arm "stretching" while crossing the silhouette (Figure 3a). This is a strong visual feedback of such an odd posture, and it suggests switching the hand (Figure 3b).

Figure 3. The Avatar shows non-natural body poses and suggests switching the interacting hand.

These outcomes do not allow us to state that the presence or the absence of the Avatar affects the number of elicited two-handed interactions. Indeed, in our case, it seems that the tile layout itself almost forces users to use both hands, whether or not the interface shows the Avatar. Rather, our results show that the Avatar affects the users' perception of this potential usability issue.

CONCLUSION AND FUTURE WORKS

In this paper, we have described a comparison study between two different touchless gestural interfaces for public displays. Our goal was to investigate how the presence of an Avatar that replays user's movements may affect the perceived cognitive workload. Moreover, we wanted to understand if the Avatar might also foster bimanual interactions.

Our study showed that the presence of an Avatar decreases the perceived cognitive workload, particularly in terms of performance, effort and frustration levels.

Concerning the two-handed interactions, since all users of both the compared interfaces used two hands, we cannot assess how the Avatar affects the number of such kind of interactions. Anyway, we can say that users who interacted with the Avatar-based interface perceived the use of two hands as natural instead of a potential usability issue.

Since we want to understand which interface element is able to foster bimanual interactions, in the next future we are planning to design further studies for comparing other interface elements and aspects than the Avatar. For instance, the layout or the need of using activation gestures may influence users in stick on a single hand instead of both.

ACKNOWLEDGMENTS

REFERENCES

1. R. de la Barré, P. Chojecki, U. Leiner, L. Mühlbach e D. Ruschin, «Touchless Interaction-Novel Chances and Challenges,» in *Human-Computer Interaction. Novel Interaction Methods and Techniques*, J. A. Jacko, A cura di, Springer Berlin Heidelberg, 2009, pp. 161-169.

2. V. Gentile, S. Sorce e 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. M. H. Haeusler, Media facades: history, technology, content., Ludwigsburg: Avedition, 2009.

4. M. Brynskov, P. Dalsgaard, T. Ebsen, J. Fritsch, K. Halskov e R. Nielsen, «Staging Urban Interactions with Media Facades,» in *INTERACT*, 2009.

5. P. Dalsgaard e K. Halskov, «Dynamically transparent window,» in *CHI EA '09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, 2009.

6. J. Fritsch e P. Dalsgaard, «Media Facades Beyond Interaction,» in *Proceedings of OzCHI 2008*, 2008.

7. J. Müller, G. Bailly, T. Bossuyt e 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.

8. N. Davies, S. Clinch e F. Alt, Pervasive Displays: Understanding the Future of Digital Signage, Morgan & Claypool, 2014.

9. R. Walter, G. Bailly e 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.

10. J. Müller, F. Alt, D. Michelis e 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.

11. J. Müller, R. Walter, G. Bailly, M. Nischt e 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.

12. V. Gentile, A. Malizia, S. Sorce e A. Gentile, «Designing Touchless Gestural Interactions for Public Displays In-the-Wild,» in *Human-Computer Interaction: Interaction Technologies*, M. Kurosu, A cura di, Springer International Publishing, 2015, pp. 24-34.

13. J. O. Wobbrock, M. R. Morris e 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.

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

15. T. Seyed, C. Burns, M. Costa Sousa, F. Maurer e 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.

16. P. Koutsabasis e 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.

17. R. Walter, G. Bailly, N. Valkanova e 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.

18. M. Ringel Morris, A. Danielescu, S. Drucker, D. Fisher, B. Lee, C. Schraefel e J. O. Wobbrock, «Reducing legacy bias in gesture elicitation studies,» *Interactions,* vol. 21, n. 3, pp. 40-45, 2014.

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

20. H. Brignull e 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.

21. V. Gentile, S. Sorce, A. Malizia, D. Pirrello e 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.

22. A. Dix, J. Finlay, G. D. Abowd e R. Beale, Human Computer Interaction, III a cura di, Upper Saddle River: Prentice-Hall, 2003.

23. S. G. Hart e 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.

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