Journal Pre-proof

Predicting Mid-Air Gestural Interaction with Public Displays based on Audience Behaviour

Vito Gentile, Mohamed Khamis, Fabrizio Milazzo, Salvatore Sorce, Alessio Malizia, Florian Alt

PII: S1071-5819(20)30099-9
DOI: https://doi.org/10.1016/j.ijhcs.2020.102497
Reference: YIJHC 102497

To appear in: *International Journal of Human-Computer Studies*

Received date: 22 December 2018
Revised date: 10 June 2020
Accepted date: 12 June 2020

Please cite this article as: Vito Gentile, Mohamed Khamis, Fabrizio Milazzo, Salvatore Sorce, Alessio Malizia, Florian Alt, Predicting Mid-Air Gestural Interaction with Public Displays based on Audience Behaviour, *International Journal of Human-Computer Studies* (2020), doi: https://doi.org/10.1016/j.ijhcs.2020.102497

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Ltd.

- We report on a 35-days long field study focused on a public display deployment
- We show that audience size and behavior significantly influence user(s) interactions
- We build predictor models able to forecast users' interaction duration and distance
- We provide a tool to visualize predictions based on audience-related input variables
- We discuss how to adapt the tool and the predictor model to other deployments

Journal Pre-proof

Journal Pre-proof

Predicting Mid-Air Gestural Interaction with Public Displays based on Audience Behaviour

Vito Gentile[a,b,*], Mohamed Khamis[c], Fabrizio Milazzo[d], Salvatore Sorce[b], Alessio Malizia[a,e], Florian Alt[f]

[a] *University of Hertfordshire, School of Creative Arts, United Kingdom*
[b] *University of Palermo, Department of Physics and Chemistry, Italy*
[c] *University of Glasgow, School of Computing Science, United Kingdom*
[d] *University of Palermo, Department of Engineering, Italy*
[e] *Molde University College, Faculty of Logistics, Norway*
[f] *Bundeswehr University Munich, Research Institute for Cyber Defense (CODE), Germany*

Abstract

Knowledge about the expected interaction duration and expected distance from which users will interact with public displays can be useful in many ways. For example, knowing upfront that a certain setup will lead to shorter interactions can nudge space owners to alter the setup. If a system can predict that incoming users will interact at a long distance for a short amount of time, it can accordingly show shorter versions of content (e.g., videos/advertisements) and employ at-a-distance interaction modalities (e.g., mid-air gestures). In this work, we propose a method to build models for predicting users' interaction duration and distance in public display environments, focusing on mid-air gestural interactive displays. First, we report our findings from a field study showing that multiple variables, such as audience size and behaviour, significantly influence interaction duration and distance. We then train predictor models using contextual data, based on the same variables. By applying our method to a mid-air gestural interactive public display deployment, we build a model that predicts interaction duration with an average error of about 8 s, and interaction distance with an average error of about 35 cm. We discuss how researchers and practitioners can use our work to build their own predictor models, and how they can use them to optimise their deployment.

Keywords: Pervasive Displays, Users Behaviour, Audience Behaviour

*Corresponding author
Email addresses: vito.gentile@unipa.it (Vito Gentile), Mohamed.Khamis@glasgow.ac.uk (Mohamed Khamis), fabrizio.milazzo@unipa.it (Fabrizio Milazzo), salvatore.sorce@unipa.it (Salvatore Sorce), a.malizia@herts.ac.uk (Alessio Malizia), florian.alt@unibw.de (Florian Alt)

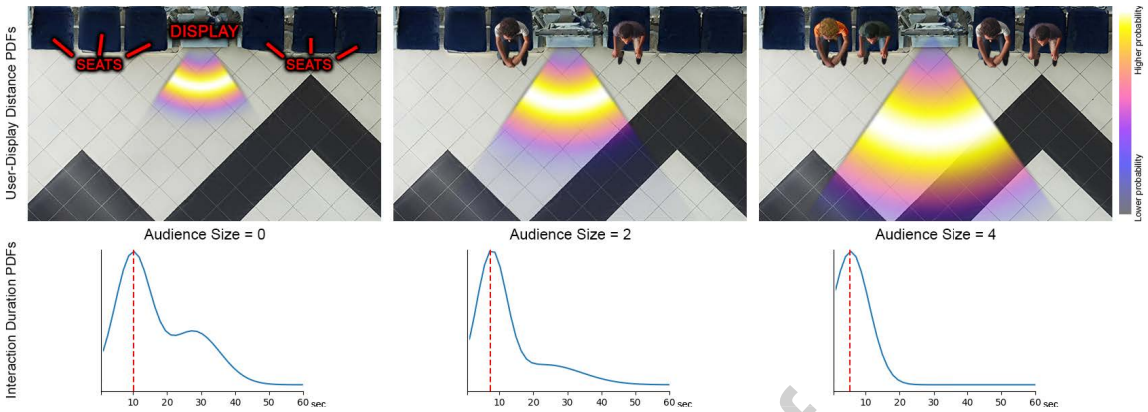

Figure 1: In this work we study the influence of the passive audience's size and behaviour on interactions with a public display, and accordingly build models to predict the distance from which users will interact, as well as the interaction duration. The figure shows a sample of our prediction model: users position themselves close to the display in the absence of an audience (A), a bit farther away in case of an audience of size two (B), and even further in case of an audience of size four (C). The graphs represent the predicted probability density functions of the interaction duration in seconds with the respective audience sizes.

1. Introduction

Interactive public displays can today be found in airports, universities, shopping malls and more. Although it is generally known that users interact with public displays for very short amounts of time, interaction durations vary widely [1, 2, 3]. The same is true for interaction distances, particularly when using interaction techniques such as mid-air gestures [2, 4]. Knowing the expected interaction duration and distance upfront can bring in a plethora of benefits to the different stakeholders of a public display. For instance, if a system is aware that the current situation will result in the user interacting at a certain distance, it can dynamically determine which interaction modality to employ [5]. Furthermore, depending on the expected interaction duration, the system could, for example, dynamically choose which version of an advertisement video to show according to the video length. In addition to run-time benefits, knowing which factors influence interaction duration and distance can be of great value to stakeholders who would then be able to tweak their setups to achieve the optimal user experience.

To address such issues, designers developed methods that try to keep users interacting longer by, for example, showing autopoiesic content [6], or presenting new content immediately after the user has finished interacting with current content [7]. Moreover, previous works reported that users interact at different distances, particularly when using at-a-distance interaction techniques such as mid-air gestures [2, 4]. The surrounding environment also affects how users position themselves: for instance, benches surrounding the display potentially result in larger audiences, which in turn discourage users from positioning themselves close to the display [8]. These facts have led researchers to study how to guide users into positioning themselves in the optimal sweet spot [9, 10].

In contrast to display-centred approaches discussed in previous work, in this paper we

Journal Pre-proof

Journal Pre-proof

focus on the people near the display, and in particular on estimating how the audience could affect possible interactions with the display itself. In more detail, here we present a method for building models to predict the duration and distance of interactions based on the behaviour of users (i.e., people who actually interact with the display) and audience (i.e., people who sit or stay around the display), and on the relationship between them.

To this end, we first analysed the behaviour of users and audience in a real-world deployment of a public display. We chose a deployment in which mid-air gestures were employed for interaction because this modality often results in varying interaction durations [11] and interaction distances [2]. Data analysis revealed that the users' interaction duration and distance from the display are significantly influenced by the number of users, the size of the audience, the user-audience relationship, and the audience's gaze towards the user(s). We then used the collected data to train predictor models based on our method, able to estimate the probability density functions (PDFs) for interaction duration and distance (see Figure 1), using audience-related information. Our solution estimates the interaction duration with a mean absolute error (MAE) of about 8 seconds, and the interaction distance with a MAE of about 35 cm. We describe our approach in details and discuss how the outcomes of our study, which include the predictor model and a visualisation tool, can be reused by researchers and designers to build predictor models for their public display setups.

We discuss how this work can help space owners and designers to optimise user engagement and experience, e.g., by (a) building similar models and (b) using them to predict how settings will influence users to (c) ultimately optimise the setup. We also explain how to leverage this knowledge to adapt content and interaction modalities in real time.

The contribution of this work is threefold:

- we report our findings from a 35-day long observation of our deployment to identify factors that influence users' behaviour;
- we propose and evaluate a machine learning approach based on expectation maximisation, aimed at generating predictor models using the collected data;
- we describe how similar models can be developed for predicting user behaviour on public displays.

To help visualise the predictions, we implemented a visualisation tool that can be adapted to other similar deployments. The tool is freely available and open source[1].

2. Related Work

Our work builds on previous work on the analysis, modelling, and prediction of the behaviour of users of interactive public displays. In particular, we focused on display applications that use mid-air gestures in order to provide interactivity to the users.

In this section, we provide an overview of the most significant previous works on these main topics.

[1] Available on `https://usi.unipa.it/MLPredictor.zip`.

Journal Pre-proof

2.1. Interacting with Public Displays

Today's public displays feature interactivity in many forms. Some displays offer implicit interaction, such as reacting to the user's natural behaviour when they approach the display. For example, many previous deployments aimed at making the display noticeable by showing silhouettes that mimic users' movements [2, 12].

Apart from implicit interaction, a large body of work investigated explicit input through modalities beyond keys and buttons, such as touch [13], mid-air gestures [14], feet-input [15], mobile devices [16], eye gaze [17], or multi-modal combinations of them [18].

2.1.1. Touch-based Interaction

Touch interfaces were a significant improvement over physical hardware (e.g., keypads, buttons and joysticks) since they expand the entropy of interaction possibilities, and allow faster software-based updates of the user interface [1]. A downside of touch interfaces is that they have to be physically reachable, while public displays are often mounted above user's height for visibility, or placed behind shop windows. As a result, distances from which users interact with displays via touch do not change significantly.

2.1.2. Mobile Device Interaction

Mobile devices, such as smartphones or smartwatches, can be used to interact with displays. Researchers explored their use for gesture-based interaction with remote displays [19, 20]. Some displays allow for interaction using mobile devices via Bluetooth [21], HTTP [22], or NFC [23]. Interaction distances could vary depending on the used technology. For example, NFC requires close proximity to the display, while Bluetooth's range can span several meters around the display. HTTP allows for interaction from anywhere as long as the user is connected to the Internet.

2.1.3. Touchless Interaction

Interaction is said to be touchless if does not require mechanical contact between the human and any part of the artificial system [24]. According to this definition, both eye gaze and mid-air gestures are touchless interaction modalities.

Eye gaze has been gaining attention as an alternative modality for interaction with public displays. Gaze has been considered a form of hygienic interaction since it can be used at-a-distance. Although gaze is fast [25], designing highly accurate gaze interfaces often results in significantly longer input time [26], thereby increasing interaction durations. Interaction distances are governed by the range of the cameras used for tracking the eyes. With the exception of active eye-tracking systems [17], most commercial eye trackers require users to position themselves 60cm-90cm away from the display [27].

The use of mid-air gestures provides an alternative at-a-distance interaction modality. Compared to other modalities, mid-air gestures were shown to extend interaction durations due to their playful nature [11]. Opposed to gaze, mid-air gestures often require larger distances between the user and the display to allow the sensors to capture the user's movement (e.g., Microsoft Kinect requires users to be between 0.5 and 2.5 meters [28]).

Journal Pre-proof

2.2. Mid-air Gesture Interaction with Public Displays

The variance in durations and distances of interactions is not the only reason why we focused on mid-air gestures. Indeed, especially in the context of public displays, this interaction modality has been widely adopted and used due to many advantages. Among them, touchless gestural interaction limits vandalism by placing displays in unreachable places [29], maintains hygiene as users no longer need to touch the display [30], and removes constraints to the display size (see, for instance, works on media façades [31, 32]). Several authors focused their work on user representation in mid-air gestural applications. Many prior work suggested to adopt users' silhouettes or avatars [2, 33, 34, 35], since they have proven to be very effective in solving some common pervasive display issues, namely *interaction blindness* (i.e. the inability of the users to recognise the interactive capabilities of a display [36]) and *affordance blindness* (i.e., the inability to understand the interaction modality of the display [37]). Gentile et al. showed also that the presence of an avatar makes two-handed interactions more "natural" in the sense that it contributes to a reduction of the cognitive workload while interacting with public displays [14].

On the other hand, mid-air gestures have also some drawbacks. If not properly designed, such interaction modality might indeed require teaching passersby how to perform gestures [38, 39]. In terms of ergonomics, using gestures for long periods of time might lead to arm fatigue, an issue often referred as the Gorilla-arm problem [40]. To solve these issues, it is crucial to incorporate users' preference when defining the gesture sets [41]. In this sense, gesture elicitation studies might help, along with specific measures of arm fatigue, used to filter out potentially bad gestures [42, 43]. Social acceptability of gestural interfaces has also been investigated by prior work. In particular, Ahlström et al. showed that users are sensitive and selective regarding where and in front of whom they would feel comfortable using mid-air gestures [44]. They also showed that acceptance and comfort are strongly linked to gesture characteristics, such as gesture size, duration and in-air position. Finally, an additional limitation of mid-air gestures is in the technologies that enable for mid-air gesture recognition, which are usually not suitable for being used in outdoor deployments. This is especially the case in very sunny environments as most depth sensors rely on infrared reflections that are impacted by sunlight [28, 45].

Based on the above considerations, we decided to focus on a gestural interface based on the use of an avatar, continuously replaying user's movements. Although we are aware that this choice might lead to fewer interactions compared to touch-based displays (e.g. due to social embarrassment of performing gestures in public [46]), the use of gestural interfaces has many advantages. As further explained in section 3.1, this solution allows us to reduce interaction blindness [2, 33], makes interactions more natural [14] and does not require learning specific gestures in order to interact properly [34, 47] – thus supporting immediate usability. Moreover, technical limitations, such as sunlight sensitivity, do not apply in indoor deployments, like the one we took into account. Furthermore, given the greater variance in durations and distances of interaction via mid-air gestures compared to other modalities, prediction of interaction distance and duration of public displays that employ this modality is particularly valuable.

Journal Pre-proof

2.3. Behaviour of Public Display Users

Previous works studied how several aspects impact the behaviour of active public display users, passive audience, and passersby.

For example, researchers looked into how the setup influences users' behaviour. In Looking Glass, users were reported to be influenced by traffic lights; passersby were observed waiting expectantly at pedestrian crossings, and then interacting with the display as soon as they are on the other side [2]. In other words, the waiting situations make passersby notice the display and hence increase interactions. Similarly, ten Koppel et al. found that the configuration of multiple screens influences how users position themselves to interact with a display [48]. They found that arranging the displays in a hexagonal configuration encourages users to interact with neighbouring displays, while in flat configurations users maximise the distance to other users. The authors attributed this to the users' desire to maintain their personal space. As a result, configurations with a large interaction space had more simultaneous users. Fatah gen Schieck and colleagues found that the layout of the building influence the behaviour of spectators, passersby, and audience and suggested that spatial configuration can be manipulated to encourage certain behaviour [32, 49, 50]. Fatah gen Schieck et al. used displays as a socialising platform that triggers shared encounters among passersby, which in turn encouraged interactions [49]. Dalton et al. found that the architecture of the building in which a display is deployed has a strong influence on whether or not users notice displays [51]. For example, users followed ceilings and wall edges with their gaze before looking at a display. Gentile et al. found that the presence of seats around the display attracts a larger audience, which causes users to interact at a further distance from the display [8]. This was attributed to discomfort resulting from having a large audience gazing at users as they interact. CityWall [13] and MyPosition [52] showed that displays can encourage communication among strangers.

Social aspects were also shown to influence interaction in public. Research showed that people might interact only to be noticed by others [53], or resist interaction due to social embarrassment [54]. Multiple works observed and investigated the honeypot effect [2, 55, 56, 57, 58, 59], which refers to a social affordance in which the presence of users interacting with a display encourages surrounding passersby to come forward and interact with the display as well. Brignull and Rogers described the honeypot effect as a "social buzz" in which those interacting with a public display seemed to signal to others that they are willing to engage in discussions and meet new people [56]. Brignull and Rogers also discussed how attempts to reduce social embarrassment when interacting with a display, such as allowing users to interact remotely with the displays, might be counterproductive as they could reduce the honeypot effect which could in turn result in less interactions with the display. Observations from field deployments indicate the honeypot effect is a powerful cue to attract attention to displays; those who observed others interacting with a display often return to interact with it themselves, sometimes even days afterwards [2, 60]. It is expected that the honeypot effect evokes curiosity thus encouraging users to interact, and counteracts social embarrassment by suggesting engagement [59]. Indeed, in Dalton et al.'s eye tracking study [51], they found that 12% of fixations at displays were preceded by noticing someone interacting with the said display. In summary, as users interact, others are attracted to the display.

Journal Pre-proof

In these cases, the *user* influences the *audience*. In contrast, Gentile et al. studied how the *audience* influences the *user* [8]. They found that the larger the audience, and the more focused they are on interacting users, the farther users position themselves from the display. While the work of Gentile et al. simply shows an effect of the audience on user's behaviour, in this work we significantly build over that by leveraging knowledge about the audience behaviour to predict the user's interaction duration and interaction distance.

2.4. Modelling and Predicting Public Display Users' Behaviour

Several works modelled users' behaviour. For example, multiple spatial [61, 62] and temporal models [56, 63, 64] were proposed for public displays. While these works contributed to discovering the existence of interaction zones, our work builds significantly over that by identifying factors that impact the transition between these zones and leveraging these factors to predict user's interaction duration and distance.

Unlike our predictive model, existing models are static and mostly qualitative, and they cannot easily be adapted to, for example, the interaction modality, application, or display context. Moreover, our model aims to provide numerically quantifiable predictions, which would increase its practical applicability in many cases.

Prediction of users' behaviour has been researched extensively in HCI and psychology. For example, previous work predicted user's mouse [65] and touch interactions [66, 67, 68] to improve accuracy. Several works proposed predicting the user's gaze based on user's behaviour [69, 70]. Erazo and Pino proposed a model for predicting execution time of mid-air gestures [71]. Other works predicted the applications that will be used next [72, 73]. In the context of public displays, Huber et al. analysed passerby's feet position to detect the user's intention and accordingly adapt the rendered content [74].

In contrast to these previous works, our approach is the first to provide quantitative predictions of the behaviour of public display users in terms of interaction duration and distance, based on audience size and behaviour.

3. Audience's Impact on Interaction

To study the impact of audience presence and behaviour on interaction in our deployment, we conducted a longitudinal study in order to observe users' behaviour while interacting with a public display. In this section, we provide a brief description of the deployment we analysed, the data collection process and the outcome from our data analysis.

3.1. Description of the Deployment

The public display we studied is located in a 150 square-meters-large indoor space inside a building within the campus of the University of Palermo. It consists of a 32-inches LCD monitor placed at eye-level, with a Microsoft Kinect sensor placed below it. The display is situated next to several benches where students often sit while waiting for lectures to start.

The display acts as an information provision system, which is one of the most common applications for public displays [75, 76]. In particular, it provides university-related information for students of different disciplines and ages (mostly between 19 and 35 years), lecturers,

Journal Pre-proof

and other university staff members. The display runs an Avatar-Based Touchless Gestural Interface (named *ABaToGI* [47]), based on the prototype described in [34]. It consists of an animated avatar shown in the middle of the screen, with other interactive tiles arranged around it (see Figure 2).

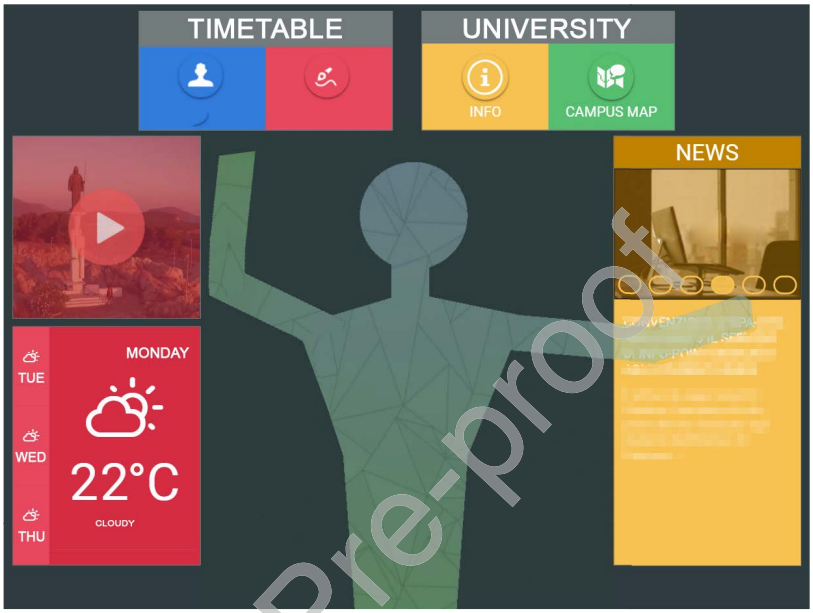

Figure 2: The display allows users to access different types of information via mid-air gestures reflected by an on-screen user representation.

To make the application interactive, the aforementioned avatar is shown every time a user approaches the display, and remains visible in the middle of the screen, continuously reflecting user's movements. This was done to reduce interaction blindness and affordance blindness [2, 33, 37], and to make interactions more natural [14]. Users can interact using in-air direct manipulations by placing a hand on a tile to trigger a selection event. By not requiring symbolic gestures (e.g., a gesture to convey a certain meaning), this system alleviates the need to learn which gestures to use in order to interact, greatly reducing the learning curve and allowing immediate usability [34].

3.2. Data Collection

We needed to collect as much information as possible about the area of interest for our study, such as the behaviour of people within it and, in particular, people's interest toward the display. For this reason, we needed to use an RGB camera to allow us to inspect the setup and the behaviour of the passersby. Since we were not allowed to install any new RGB cameras in the test environment, we had to use one of the available WiFi surveillance cameras. The WiFi camera we used was placed in front of the display, in an unreachable position, to observe users' interactions and audience behaviour. This way we were able to

Figure 3: The wide view of the camera enabled us to identify users who arrived in groups, or communicated with each other or with the audience before interacting with the display. This allowed us to identify relationships. The figure shows a sample of the variables and measurements that were logged in the recordings.

remotely observe: (1) users, (2) audience (e.g., people sitting on the benches next to the display), and (3) the display status. The University's institutional review board allowed us to access the camera's streaming video for the time the experiment has been carried out. It is worth noting that even though we had a Kinect device installed for gesture detection, we could not use it for data collection because its range is not wide enough and it does not allow for gathering contextual data (e.g. audience gaze, and whether the user interacted due to the honeypot effect).

Figure 3 shows an example of the camera view, and a subset of the variables and measurements we collected. We recorded interactions between 10 am and 5 pm on workdays.

Since we started collecting videos after one year from the first installation of the display, we can reasonably assume the absence of significant biases due to novelty effect.

For this work, we analysed 200 hours and 47 minutes of recorded video feed collected along 35 days. During this period of time, we observed 123 total interaction events. For our study, an event was relevant when there is at least an interaction attempt with the display by one or more people. Users interacted for 1 to 136 seconds ($M = 22.7s$, $SD = 26.5s$, see Figure 4b). It is worth noting that the average duration of interaction events is in line with data reported in other public display deployments [2]. Same considerations apply for the average number of interaction events per hour (0.615), which is in line with data reported for other indoor public display deployments [3].

Journal Pre-proof

Journal Pre-proof

	Variable name	Symbol	Type	Values	Description
Input	Number of Users	U_{count}	Discrete	≥ 1	Number of interacting users
	Audience Size	A_{size}	Discrete	≥ 0	Number of persons acting as passive audience while one or more users interact
	Audience-Display Distance	ADD	Continuous	≥ 0	Distance between the display and the closest person in the audience
	Audience Side	A_{side}	Categorical	left right both	Which side is the audience, with respect to the display
	Audience Gaze	A_{gaze}	Categorical	yes no	Whether or not someone in the audience looked at the user during interaction
	User(s) Gaze	U_{gaze}	Categorical	yes no	Whether or not a user noticed to be gazed from the audience
	User(s)-Audience Relationship	UAR	Categorical	acquainted strangers mixed	Relationship between user and the audience
	User(s) was/were in the Audience	U_{in}^{A}	Categorical	yes no	Whether or not the user(s) acted as part of the audience before interacting
	User(s) interacted due to Honeypot Effect	HE	Categorical	yes no	Whether or not the user(s) interacted due to the honeypot effect
Output	Interaction Distance	d	Continuous	≥ 0	Distance between the user(s) and the display
	Interaction Duration	t	Continuous	≥ 0	Duration of the interaction session

Table 1: Summary of variables that we tracked while annotating the videos. We used these variables both for the statistical analysis and for building our predictor model.

Journal Pre-proof

Journal Pre-proof

There are infinitely many factors that might impact users' behaviour around displays. We decided to focus more on the human factors than on technical ones, in particular on how some audience-related factors affects the behaviour of display's users-to-be. Furthermore, some of the audience-related variables, such as size, distance, side, and gaze, can be easily influenced by the deployment layout (presence/absence of benches, number and position of seats, etc.). Therefore, space owners might have more control on such changes, for instance in terms of cost, to achieve the desired values for duration and distance of interactions. Indeed it is likely that a space owner could easily add, remove or move some furniture around. We acknowledge that factors such as time (e.g., weekday and time of the day), weather, screen size as well as the user's gender, culture, age and body orientation might impact the users' behaviour [2, 77, 78]. Indeed we provided an overview of the most significant factors that impact the users' behaviour around displays in section 2.3. Here, we decided to take into account the variables on which space owners may have an easier, and cheaper, control, and leave the others for future work (see section 5.6). Thus, for each interaction event, we collected 11 variables, summarised in Table 1. Two researchers separately reviewed all the recorded videos, based on the following rules:

- Number of Users (U_{count}) was tracked because it influences interaction due to the honeypot effect, which we discussed at length in Section 2.3. U_{count} was coded by counting the number of users simultaneously involved in each interaction event.

- Audience Size (A_{size}) was tracked as it was shown to influence the behaviour of users of public displays [8]. A_{size} was coded by counting the number of persons that, during an interaction event, were not involved in the interaction

- Audience-Display Distance (ADD) is likely to impact the users' awareness that they are being observed, which could in turn cause social embarrassment [54] or, for some, even encourage interaction [53]. We coded ADD by counting the number of tiles on the floor from the display to the closest person in the audience, or the number of seats in case the closest person in the audience was seated. In cases where the audience were considerably moving during the interaction event (i.e., the distance between audience and the display varied, and it was not possible to identify a single constant value for ADD during the whole interaction event), we coded this value as N/A.

- Audience Side (A_{side}) refers to which side the audience was at, and was coded by checking whether the audience was on the left, on the right or on both sides with respect to the user(s). In those cases when the person(s) in the audience were not clearly placed on the left or on the right (i.e., they were moving around the joining line between the user and the display), we coded this value as N/A

- Audience Gaze (A_{gaze}) was suggested to influence interaction distance [8]. A_{gaze} was coded by looking whether at least one person from the audience looked at the user(s) during the interaction event. In those cases where this determination was dubious or unclear, and when such determination was useless (i.e., when $A_{size} = 0$) we coded this value as N/A.

Journal Pre-proof

- User(s) Gaze (U_{gaze}) can be an indicator of whether the user noticed the audience. This can be valuable in further distinguishing cases where the user was aware of the surrounding audience. U_{gaze} was coded by looking whether the user (or one of the users, in case of groups of users) looked back to a person from the audience who were looking at her/him during the interaction event. In other words, this variable represents if the user(s) noticed to be looked by someone in the audience. In those cases were this determination was dubious or unclear, we coded this value as N/A.

- User(s)-Audience Relationship (UAR) can be influential as previous work showed that users are sometimes more comfortable interacting in front of friends [57]. UAR was coded by looking at interactions between user(s) and persons in the audience before and after the interaction events. For instance, if a user and someone in the audience talked before or after the interactions, or if they arrived together next to the display, we assumed they were acquainted. Otherwise, if a user arrived alone at the display, and went away with no other social interactions with the audience, we assumed they were strangers. In those cases were this determination was dubious or unclear, we coded this value as N/A.

- User(s) was/were in the Audience (U^A_{in}): if the user was among the audience earlier, this can be considered a result of the honeypot effect as it was shown that users return, sometimes even days, after observing someone interacting with the display [2]. U^A_{in} was coded by checking if the user(s) was/were part of the audience during a previous interaction event.

- User(s) Interacted due to Honeypot Effect (HE): tracking this is useful in determining whether the user was motivated by the honeypot effect or not which is known to influence user behaviour (see Section 2.3). HE was coded by checking if an honeypot effect occurred (i.e., at least one user from the audience approached the user during an interaction event, and then started a new interaction event).

- Interaction Distance (d): this is one of the dependent variables that we wanted to investigate. Knowing which factors impact the interaction distance can be very valuable for space owners. For example, a display can dynamically determine which interaction modality to employ, or the size of the content depending on the anticipated interaction distance. d was coded by counting the number of tiles on the floor between the user(s) and the display.

- Interaction Duration (t): this is another dependent variables that we investigated. It is also very valuable to understand and be able to manipulate the factors that impact interaction duration. For example, if shorter interaction durations are expected, the display can play shorter versions of advertisements, or decide to employ more aggressive approaches for keeping the user. t was coded by measuring the time between an action that showed user(s)' intention of initiating the interaction (e.g. stopping in front of the display and raising arm or moving body), until user(s) left the display.

Journal Pre-proof

	U_{count}	A_{size}	ADD	A_{side}	A_{gaze}	U_{gaze}	UAR	U^A_{in}	HE	d	t
ICC or κ †	0.985	0.977	0.979	0.994	0.991	0.975	0.978	0.981	1.000	0.960	0.994

† $p < 0.001$ for all the above results

Table 2: Agreement scores: intra-class correlation coefficient (ICC) has been used for continuous variables (ADD, d, t), while Cohen's kappa has been computed for all the other variables (weighted in case of discrete variables U_{count} and A_{size}).

All distances coded by counting the number of tiles or seats were later converted to centimetres. In a few cases, when variables changed only slightly and shortly during an interaction event (e.g., changes in audience size amid interaction), we coded the most representative value for the specific interaction analysed (e.g., the audience size during the major part of the interaction session).

Both the researchers who were involved in the video reviewing and coding process did record the same number of interaction events. We also computed three different agreement scores, based on the variable type reported in Table 1. In particular, based on the recommendations by Raganathan et al. [79], we used Cohen's kappa for categorical variables, weighted Cohen's kappa for discrete variables, and intra-class correlation coefficient for continuous variables. Agreement scores are reported in Table 2. In those cases where the two researchers disagreed, both the researchers reviewed again the videos together. Then, we resolved the disagreements as follows:

- for continuous variables (i.e., ADD, d, t), if an agreement was not reached, we computed the average value between the values coded by the two researchers
- for categorical and discrete variables, if an agreement was not reached, the values were set to N/A (i.e., treated as dubious/unclear cases)

3.3. Limitations

The results of our statistical analysis are valid for the group of users that we considered in our longitudinal study. Results could be different in other deployments based on the setup, user groups, and interaction modality. However, the procedure described for data collection, statistical analysis and predictor training can be replicated in other deployments with different settings. Another limitation is that although we carefully reviewed the situation before and after each considered case to note any external influences on users' behaviour (e.g., indications that the audience and users know each other), however we are not fully aware of what happens beyond the camera's range. Finally, we do not know how often each user interacted with the display before each recorded interaction attempt. Consequently, we have not included information about previous experiences in interacting with the display. While an alternative would be to collect and recognise faces to identify users and collect such information, this was not feasible due to privacy concerns.

Journal Pre-proof

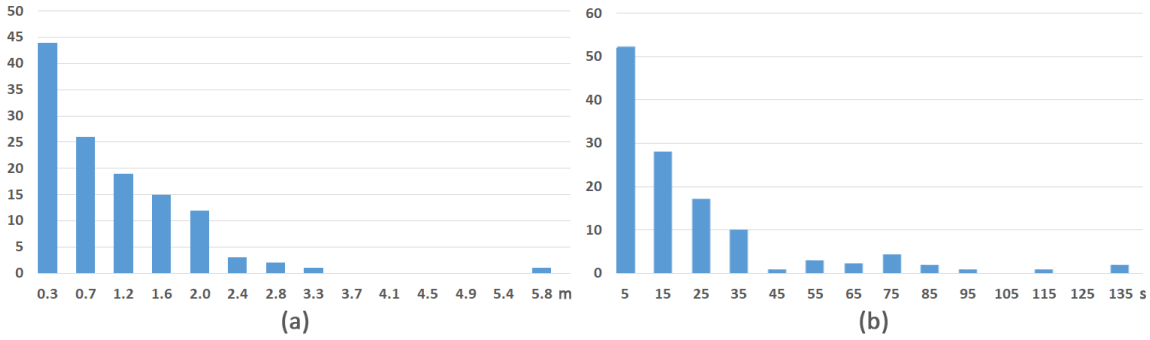

Figure 4: Histograms show the distributions of (a) interaction distance and (b) interaction duration.

3.4. Statistical Analysis

From the 123 collected interactions, we excluded exceptional cases that occurred very few times to be significant for drawing general conclusions. Namely, we excluded five interactions where there was an audience of six or more people, and three cases where the audience was more than 2.5 meters away from the display. This means we ended up with 115 valid interactions to analyse.

Out of the 115 considered interactions, 43 occurred in the presence of an audience that the users were acquainted to, 17 cases were logged where the audience and the users were strangers, 23 cases had a mixture of strangers and acquaintances among the audience, and in 32 cases there was no audience at all. A one sample binomial test confirmed that there is a statistically significant tendency for more interactions when there are acquaintances among the audience or when there is no audience, with a proportion of 0.85 (larger than expected 0.75, $p < 0.001$).

3.4.1. Interaction Duration

Our main findings with respect to interaction duration are illustrated in Figure 5. Namely, we found a significant effect of three variables on interaction duration.

We found an effect of the *number of users* on interaction duration. An ANOVA test revealed a significant effect of the number of users on interaction duration ($F_{2,113} = 9.84$, $p < 0.001$). A post-hoc Tukey test showed that interaction duration in the case of a single user ($M = 17.5\,s$, $SD = 18.4\,s$) is significantly shorter than in the case of two users ($M = 33.3\,s$, $SD = 34.9\,s$, $p < 0.05$) and three users ($M = 58.8\,s$, $SD = 45.5\,s$, $p < 0.005$). This means when multiple users interact, they are likely to interact for longer durations.

The *size of the audience* also had an influence on interaction duration. An ANOVA test showed that there is a significant effect of the audience size on interaction duration ($F_{2,109} = 3.15$, $p < 0.05$). Post-hoc Tukey tests revealed significant differences between interaction duration when there is an audience of size 4 or more ($M = 10.9\,s$, $SD = 9.0\,s$) compared to cases where there is no audience ($M = 27.6\,s$, $SD = 29.7\,s$, $p < 0.001$) and cases where there is a smaller audience of size 1 to 3 ($M = 20.8\,s$, $SD = 25.8\,s$, $p < 0.05$). This means that the bigger the audience, the shorter the duration users interact.

Journal Pre-proof

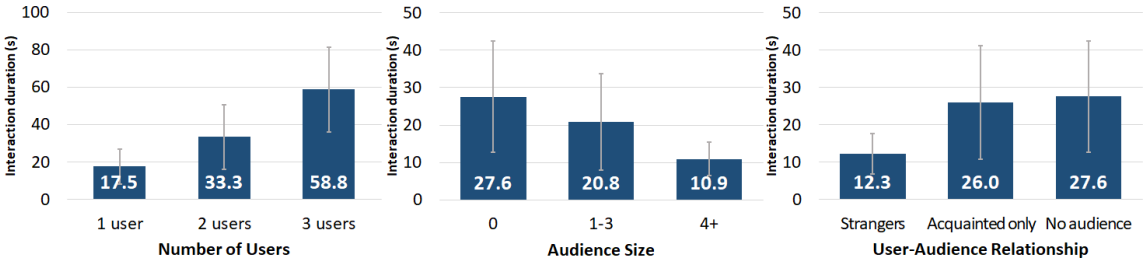

Figure 5: Users interact for longer durations when they are in groups. Interaction durations are significantly shorter when a larger audience surrounds the users, and particularly when it includes strangers rather than only acquaintances to the users.

We also found an effect of the *user-audience relationship* on the interaction duration. An ANOVA test showed a significant effect of the relationship on interaction duration ($F_{2,109} = 4.1$, $p < 0.05$). A post-hoc Tukey test showed that interaction duration in the presence of at least one stranger among the audience ($M = 12.3\,s$, $SD = 10.8\,s$) is significantly shorter than in the case of an audience composed of acquainted people ($M = 26.0\,s$, $SD = 30.6\,s$, $p < 0.05$), and also significantly shorter than the case of no audience ($M = 27.6\,s$, $SD = 29.7\,s$, $p < 0.005$). There is no significant difference in the interaction duration between the cases of no audience vs presence of acquaintances. In other words, this means that interaction durations in front of strangers are 1) shorter than in front of acquaintances, and 2) shorter than in the absence of an audience.

3.4.2. Interaction Distance

An independent-samples t-test was run to determine if there were differences in user-display distance between *presence of an audience* compared to the absence of an audience. A significant main effect was found ($t(113) = -2.84, p < 0.01$). Users position themselves significantly farther away from the display when an audience is present ($M = 100.1\,cm$, $SD = 84.9\,cm$), compared to cases where there is no audience ($M = 50.6\,cm$, $SD = 52.2\,cm$). When further analysing the cases where an audience is present, an ANOVA showed that the *audience size* has a significant effect on the interaction distance ($F_{2,113} = 8.8$, $p < 0.001$). Post-hoc Tukey tests revealed that users position themselves significantly farther away from the display in the presence of an audience of size four or more ($M = 140.4\,cm$, $SD = 121.9\,cm$) compared to cases where there is no audience ($M = 50.6\,cm$, $SD = 52.2\,cm$, $p < 0.001$) and compared to the presence of an audience of size between one and three ($M = 87.7\,cm$, $SD = 66.4\,cm$, $p < 0.05$). This means that users position themselves significantly farther away from the display in the presence of an audience, and in particular in the case of a large audience.

Finally, an ANOVA revealed that the *audience's gaze* at the user has a significant main effect on user-display distance ($F_{2,113} = 8.45$, $p < 0.001$). Post-hoc Tukey tests showed that users position themselves significantly farther away from the display when the audience gazes at them ($M = 112.7\,cm$, $SD = 91.0\,cm$) compared to when they do not gaze at the user ($M = 70.0\,cm$, $SD = 59.6\,cm$, $p < 0.05$), and compared to when there is no audience

Journal Pre-proof

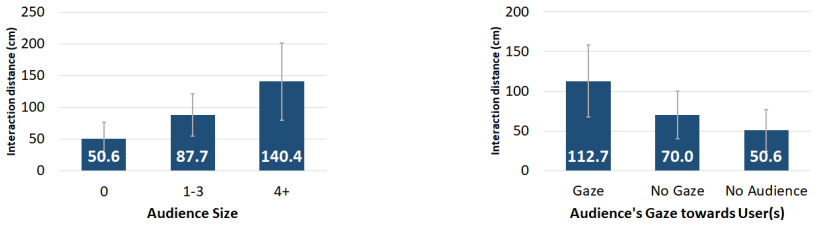

Figure 6: Users position themselves far from the display when a large audience is present, and significantly farther when the audience gazes at them.

($M = 50.6\,cm$, $SD = 52.2\,cm$, $p < 0.005$). This means that when the audience gazes at the user, the user keeps a larger distance to the display.

4. Predicting Interaction Duration and Distance

In the previous section, we described the process of observing users' behaviour and collecting data related to a set of surrounding information, e.g., the audience's presence and behaviour, in order to understand the relationship between such audience-related data and the users' behaviour while interacting with a display. Using this information, we concluded that some of the observed variables have a more significant influence than others.

In this section, our goal is to understand how to use such data in order to automatically predict the impact of audience and users' behaviour with respect to the deployment setup. To this end, here we present a machine learning approach for supporting display owners in predicting users' behaviour in a specific deployment. Such predictions can then be used for increasing the use of the display, improving the experience, and making public display deployments more successful.

4.1. Predictor Model: Definition

The predictor model proposed in this section estimates the probability density function (PDF) of a continuous output variable y, using a set X of N input variables x_i, i.e., $X = \{x_1, x_2, \ldots, x_N\}$. In particular, we estimate the PDF of an output variable given the observed variables (namely $p(y|X)$), which implies the estimation of the joint probability $p(y, X)$.

In the practice, a perfect joint probability estimation generally requires a number of samples that grows exponentially with the number of input variables. For instance, suppose that our predictor model would require nine input variables, each of which can assume only two values. Thus, a proper joint probability estimation would require at least 2^9 training samples, which is often not practical in actual deployments. This is particularly true when considering video-coded data, which is very often the case of observations in-the-wild and longitudinal studies in pervasive display research [1].

To overcome this limitation, we leverage a standard technique that implies the assumption of a Naive Bayesian dependence between output and input variables. This is a machine-learning stratagem that was proven effective and eases the estimation of complex joint probabilities, even with relatively few data samples, unlike other approaches (e.g.,

Journal Pre-proof

neural networks, SVMs, etc...) [80]. In particular, it allows considering the input variables conditionally independent given the output.

Thus we can write down the conditional as follows:

$$\begin{aligned} p(y|X) =& \frac{p(y,X)}{p(X)} = \frac{p(X|y)p(y)}{p(X)} = \\ =& \frac{p(x_1|y)p(x_2|y)\dots p(x_N|y)p(y)}{p(X)} = \\ =& \frac{p(y|x_1)p(x_1)\dots p(y|x_N)p(x_N)p(y)}{p(y)^N p(X)} \end{aligned} \tag{1}$$

We can group the constant parts $p(x_i)$ and $p(X)$, which form a constant scaling factor K, defined as follows:

$$K = \frac{\prod_{i=1}^{N} p(x_i)}{p(X)} \tag{2}$$

Thus, it holds that Equation 1 can be rewritten as follows:

$$\begin{aligned} p(y|X) =& \frac{K \cdot p(y|x_1)\dots p(y|x_n)p(y)}{p(y)^N} = \\ =& K \cdot p(y|x_1)\dots p(y|x_N)p(y)^{1-N} \end{aligned} \tag{3}$$

In practice, $p(y|x_i)$ and $p(y)$ can be estimated from the observed data (i.e., training samples) as Gaussian mixtures, using the expectation-maximisation (EM) algorithm [80].

The calculus of K can be avoided because it is equivalent to a simple normalisation operation of $p(y|X)$.

Thus, the $p(y|X)$ estimation all comes down to estimating $p(y|x_i)$ for each i-th input variable, and $p(y)$. As we explain in section 4.2, such estimations represent the whole training process of our model, which can be implemented by means of the available (free and open source) implementations of the EM algorithm [80, 81].

Estimating $p(y|X)$ would then allow to predict the most probable value of y, according to the input variables x_i.

4.1.1. Choice of Prediction Technique

This approach has several significant advantages compared to other machine learning approaches, such as regression analysis. Although the higher the number of samples are, the more accurate the estimation can be, unlike neural networks and SVMs this model can be used with fewer samples. Moreover, it features incremental training which allows improving precision with newly collected data, without the need to retrain the whole model.

Furthermore, unlike classical regression, this approach allows to train and use the model even if some of the input variables are not available. For instance, an automatic data collection process might provide uncertain values in some cases (e.g., due to occlusion or low video resolution). Even if some input values are missing, this approach would still be

Journal Pre-proof

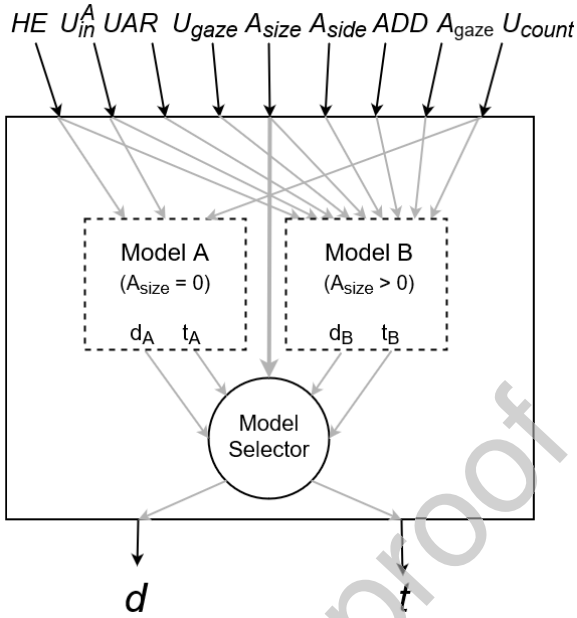

Figure 7: Overview of the variables and the predictor models. Because some variables, namely ADD, A_{side}, A_{gaze}, U_{gaze}, UAR, would be meaningless if there is no audience, we built a model per interaction duration (t) and interaction distance (d), and we further split each to two models to account for $A_{size} = 0$ and for $A_{size} > 0$.

useful for training a predictor model. While in this paper we have manually coded data from videos, we envision that future work will collect the data automatically in the wild, which means that it will be more likely that certain data points will be missing (e.g., due to inaccuracies, obstacles, etc.). We discuss how to automate the data collection process in Section 5.5.

The feature of managing missing values is also useful while using the model as a predictor (i.e., after the training stage): even if some of the input variables are not available, the PDF estimation is still possible. Of course, the more the known inputs are, the more precise the estimation will be.

4.2. Predictor Model in Practice

As stated before, our aim is to predict how the distance between the users and the display varies according to audience behaviour, as well as how long users interact with the display. In order to put the above methodology into practice, we have thus built two different predictor models: one for predicting the user-display distance (i.e., $p(d|X)$), and another one for predicting the interaction duration (i.e., $p(t|X)$). The variables d and t are therefore our outputs.

We have then considered a total of nine possible input variables x_i, which correspond to the data collected for the aforementioned data analysis (see Table 1).

Obviously, if there is no audience (i.e., $A_{size} = 0$) then many variables (namely ADD, A_{side}, A_{gaze}, U_{gaze}, UAR) are irrelevant, since they are subsumed by A_{size} in this particular

Journal Pre-proof

case. For this reason, we have not built two models (i.e., one for predicting $p(d|X)$, and another one for $p(t|X)$) but actually four, doubling the original two by splitting the cases for $A_{size} = 0$ and for $A_{size} > 0$. This distinction is depicted in Figure 7. However, for the sake of simplicity and clarity, in the rest of the paper we will often refer to a single model that predicts both $p(d|X)$ and $p(t|X)$.

Training such models basically means that we have to estimate $p(d)$, $p(t)$ and each $p(d|x_i)$ and $p(t|x_i)$ by applying EM-based Gaussian mixture estimation from the data collected during our observations.

We have implemented this procedure using the Python programming language, exploiting some of the features available from the scikit-learn package [81]. This allowed us to quickly estimate probability distribution as Gaussian mixtures, and to easily operate on them. Our predictor model was built using the data gathered during the observations described previously. In particular, we used the same dataset analysed during the statistical analysis.

In practice, interactions from distances of 2.5 meters or farther from the display are not detected reliably by the Kinect device used in our deployment [28]. This suggests that interaction experiences from such distances might be skewed. Furthermore, analysing the distribution of the interaction durations revealed that there is a significant drop in number of cases per interaction durations that are longer than 60 seconds (see Figure 4b). For these reasons, we used the 104 interactions with distances and durations below these criteria to train our model.

4.3. Performance Analysis

We evaluated the performance of our methodology by using the Leave-One-Out Cross Validation (LOOCV), over the 104 observed samples. Thus, we measured a mean absolute error (MAE), and its standard deviation σ^{MAE} for the two output variables (i.e., interaction distance and duration).

We computed the MAE by considering all the possible combinations of 8 input variables (i.e., all the variables except A_{size}, which was always included in order to select the right model, as depicted in Figure 7).

According to our evaluation, the best input variables combination for predicting user-display distance is $< A_{size}, U_{count}, UAR, U^A_{in} >$ ($MAE = 0.3518$, $\sigma^{MAE} = 0.3525$), while for the interaction duration, the best input variables combination is $< A_{size}$, U_{count}, A_{gaze}, U_{gaze}, UAR, U^A_{in}, $HE >$ ($MAE = 7.8786$, $\sigma^{MAE} = 5.5776$).

In order to rank all the input variables combinations for minimising the MAE of both the outputs, we computed the following metric for each k-th input variable combination:

$$M^E_k(y) = MAE_k(y) + \sigma^{MAE}_k(y) \tag{4}$$

where y can be d or t depending on the considered output. Thus, the error vector representing the k-th input variable combination can be defined as follows:

$$E_k = \langle M^E_k(d), M^E_k(t) \rangle \tag{5}$$

Journal Pre-proof

We adopted the Pareto dominance as an order relation [82] for all the error vectors E_k. In general, if E_1 and E_2 are two error vectors, then the Pareto dominance of E_1 with respect to E_2 is expressed by the following equation:

$$E_1 \preceq E_2 \Leftrightarrow \{\forall i = 0, \ldots, n \Rightarrow E_1(i) \leq E_2(i)\} \tag{6}$$

Thus, an error vector E^* can be considered Pareto optimal if it is not worse than (i.e., dominated by) any other error vector:

$$E^* = \{E_i : \forall k = 0, \ldots, n \wedge E_i \neq E_k \Rightarrow E_i \preceq E_k\} \tag{7}$$

Using the Pareto sorting algorithm proposed in [82], the input variable combination $< A_{size}, U_{count}, A_{gaze}, UAR, U^A_{in} >$ results in the best outcome with respect to both interaction distance and duration ($MAE(d) = 0.3575$, $\sigma^{MAE}(d) = 0.3530$, $MAE(t) = 7.8793$, $\sigma^{MAE}(t) = 5.9246$), since it produced a Pareto optimal error vector.

It is worth noting that, among all the input variables listed in Table 1, HE and U^A_{in} require knowledge on the behaviour of user(s) before the interaction. This means that such variables may result tricky to use, and somehow not particularly useful for display providers or space owners, that are the final users of this predictor model. For this reason, we report here also the best combination of input variables excluding HE and U^A_{in}, which is $< A_{size}, U_{count}, A_{gaze}, UAR >$ ($MAE(d) = 0.3497$, $\sigma^{MAE}(d) = 0.3588$, $MAE(t) = 7.9180$, $\sigma^{MAE}(t) = 6.0094$).

This means that our model can predict interaction duration with a MAE of 7.9180 seconds, and interaction distance with a MAE of 0.3497 meters.

4.3.1. Prediction Errors and Statistical Significance

The variables $< A_{size}, U_{count}, A_{gaze}, UAR >$ resulted to be the best combination to predict both user-display distance (d) and interaction duration (t). This is in accordance with the findings of the statistical analysis (see section 3.4), which showed a significant effect of audience size (A_{size}) and audience's gaze (A_{gaze}) on user-display distance, as well as a significant impact of audience size (A_{size}), number of users (U_{count}) and relationship between them and the audience (UAR) on the interaction duration. These results, along with the MAEs reported above, confirm the effectiveness of our proposed model in automatically (and correctly) predicting the impact of audience and user's behaviour with respect to the deployment setup.

4.4. Validation Against a Different Dataset

In order to further confirm the results discussed in the previous sections, we decided to perform an additional validation by using a different dataset. To this end, we collected an additional dataset from the same deployment, after several months from the first data collection period. In particular, we collected data for a total of 20 interaction events, coded over 2 days for 14 hours and 20 minutes. It is worth noting that such number of interaction events represents a sufficient percentage with respect to the size of the dataset, according to

Journal Pre-proof

the common practices adopted for sizing validation sets [83]. The data collection procedure followed the same rules explained in section 3.2.

Using this additional dataset, we computed the MAE by considering a combination of all the input variables, as well as using only the combination $< A_{size}, U_{count}, A_{gaze}, UAR >$, as this was the more suitable (based on considerations in section 4.3). For the combination of all the variables, we computed $MAE(d) = 0.3753$ m ($\sigma^{MAE}(d) = 0.3119$) and $MAE(t) = 7.700$ s ($\sigma^{MAE}(t) = 7.4364$). Considering instead the combination $< A_{size}, U_{count}, A_{gaze}, UAR >$, our results were $MAE(d) = 0.3654$ m ($\sigma^{MAE}(d) = 0.3147$) and $MAE(t) = 8.0500$ s ($\sigma^{MAE}(t) = 6.1192$). Since the results are in line with the ones reported in section 4.3 after the LOOCV analysis, we can reasonably confirm the average prediction errors.

5. Discussion and Uses of the Model

In this section, we discuss the results of the field study and the subsequent statistical analysis, and we describe how display providers and space owners can benefit from the proposed model.

5.1. Performance Assessment

In section 4.3 we computed a mean distance error of 0.3497 meters, which is roughly 1.5 times the average length of a human foot [84]. Moreover, based on Hall's work on proxemics and interpersonal interaction, this error is less than the size of intimate space [85]. As discussed in section 2.4, to the best of our knowledge prior works do not provide quantitative estimations that are comparable to these results, but only qualitative interaction zones [56, 61, 62, 63, 64]. Consequently, to our knowledge, our model is the first providing quantitative predictions, and therefore we cannot entirely compare our results to similar models. In future work we may expect that new predictor models will improve the aforementioned prediction error, but based on the above considerations, distance error we obtained here can be considered acceptable.

As for the duration error, our results showed a mean duration error of 7.9180 seconds. Prior work, albeit in a variety of different contexts and scenarios, showed an average interaction duration with public displays of about 26 seconds [2] for which, approximately 8 seconds, represents a non-negligible proportion. Nonetheless, such circa 8-sec an error might be useful to qualitatively predict if interactions will be "short" or "long", with respect to some threshold (e.g. the average duration defined in [2]). It is worth noting that up to our knowledge, no prior works have provided predictor models that are even vaguely able to predict interaction duration.

5.2. Impact of the Audience on Interaction

In section 3.4, we found that interaction duration increases as the number of users increases. This is in line with previous work which showed that when users interact, they encourage others to interact too in what is referred to as the honeypot effect [56]. On the other hand, we also found that an increase in audience size results in a decrease in interaction duration. This is also the case when people in the audience are strangers; if the users do

Journal Pre-proof

not know the audience, they interact for a short time compared to when they know them. This is a novel finding that we attribute to social embarrassment resulting from the presence of strangers and a large audience. Social embarrassment is known to influence interaction [56, 86]. Previous works have also reported cases where users felt more comfortable when interacting in front of friends [57], which is inline with our results that show that users interact longer when in front of an acquainted audience.

Another factor that was found to be influenced by the audience size and behaviour is the interaction distance. We found that the presence of an audience results in significantly larger distance between the user and the display. This effect is amplified in case of a large audience; users position themselves significantly farther as the size of the audience increases. Furthermore, the audience's gaze towards the user has a significant effect on the interaction distance as well. These results can be interpreted as follows: users *compensate* for the audience's eye contact, which is a form of "increasing closeness", by keeping a larger distance to the audience. This can be explained by the compensation model of interpersonal distancing, which states that individuals will increase the distance to others if others are too close in terms of physical proximity, eye contact, etc. in an attempt to reach an equilibrium [87].

Although previous work showed that there is a relationship between the audience size and how far users distance themselves from the display [8], our work is the first to find statistically significant results to support this finding. This is mainly due to our larger sample size; our sample covers 35 days while previous work coded only 5 days. Another difference is that previous work found a trend showing that audience-display distance might influence the user-display distance. Although we noticed a similar trend in our data, we could not find any significant effects to support the existence of an influence ($p > 0.05$).

5.3. Data Visualisation

In order to highlight the main advantages that our predictor model can provide to researchers and practitioners, we developed an interface to easily set the inputs and visualise the outputs of our predictions based on the model[2]. Through this visual tool, a user can easily set any of the input variables (see the upper part of the GUI, shown in Figure 8), in order to see their effects on the outputs. This provides users with a straightforward visualisation of the predictions.

This visualisation tool shows the most probable interaction areas in front of the display. The interface (shown in Figure 8) was implemented in Python, by means of some packages for GUI programming (Tkinter [88]), graph plotting (matplotlib [89]), and image processing (Pillow [90]). Our tool uses a dataset file to train the predictor model on startup, and then uses the model to generate distance heat maps based on the input values set by the user. Such heat maps overlay an orthogonal top view of the deployment.

Along with such visualisation, the probability density functions (PDFs) generated by our predictor model, both for user-display distance and interaction duration, are shown in the bottom area of this visual tool. The heat maps are generated by expanding the estimated

[2] The tool is available on `https://usi.unipa.it/MLPredictor.zip`.

Journal Pre-proof

user-display distance PDF within a 2-D wedge-shaped area, corresponding to the field of view of the Kinect installed below the display.

This means that, in order to adapt our model to different deployments and to suitably visualise the corresponding predicted values, users would only need to change two files: the deployment orthophoto, and the dataset used for training the model.

5.3.1. How to adapt the Visualisation Tool to different deployments

The tool can be run by means of the Python script file *gui.py*. When started, it loads the two files representing the current deployment, and that can be used for customisation in other contexts:

- a JPEG image file, representing an orthogonal top view of the deployment;
- a plain text file, formatted as CSV (i.e., comma-separated values), representing the whole dataset.

The CSV file contains all the coded data used for training the predictor model. The structure of this file is described in Figure 9. The training time depends on many factors, such as the number of entries in the dataset and the available computational resources.

The tool can be used by public display researchers and practitioners to visualise the estimated interaction distances and durations in different settings of their deployments. Adapting the tool to other deployments is straightforward; the tool's users only need to convert their data to CSV format in order for them to be used to train the model, and replace the JPEG file with a top view of the deployment. By visualising interaction distances, display owners can better understand the behaviour of their users, and the impact of their setup. Furthermore, by changing the variables set and values, they can easily estimate the effect of layout changes on distance and duration of interactions.

Journal Pre-proof

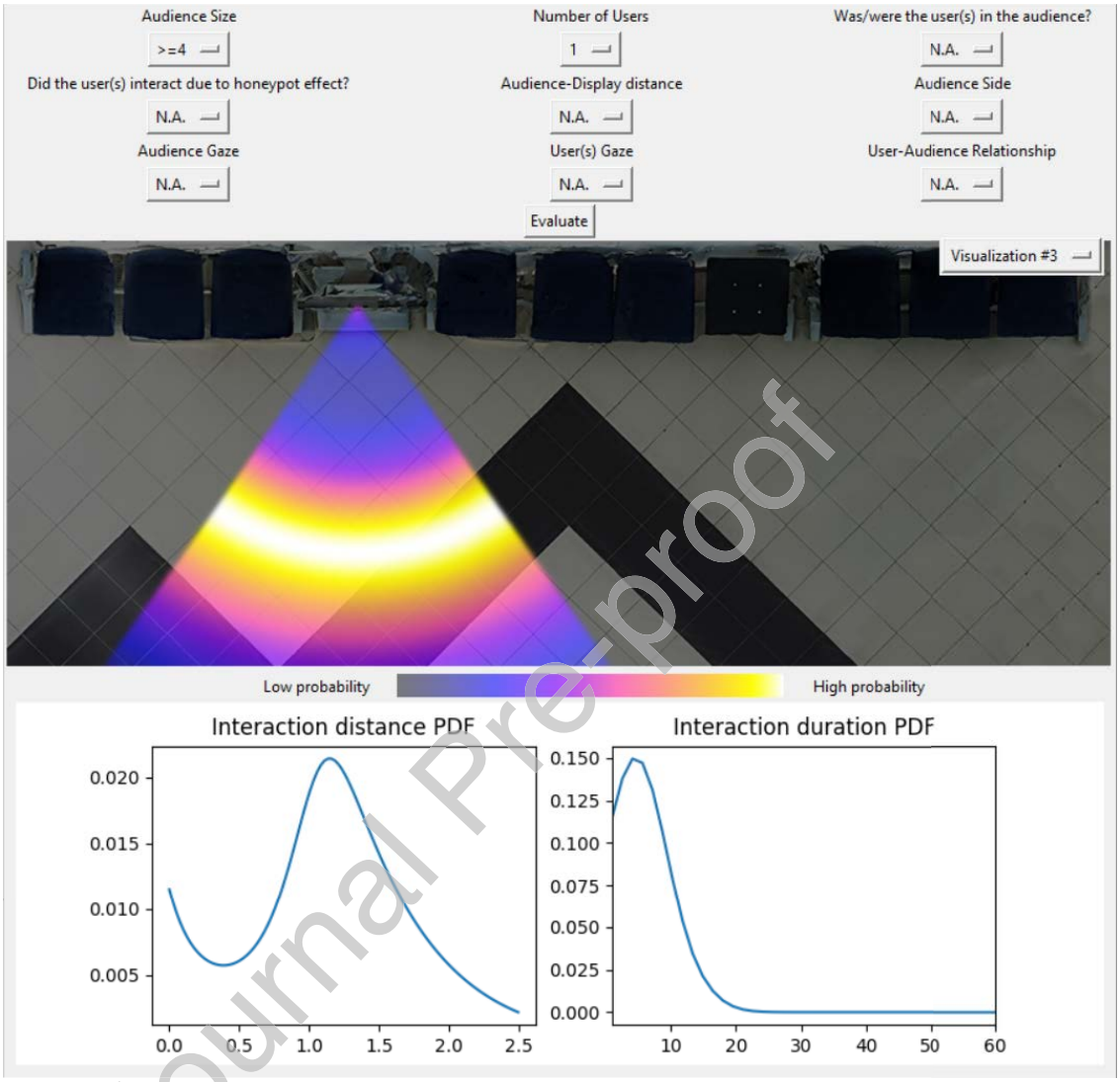

Figure 8: GUI of our proposed visualisation tool. Researchers and practitioners can use their data to feed the prediction model, and thus visualise the estimated interaction distance and duration through this tool.

Journal Pre-proof

Journal Pre-proof

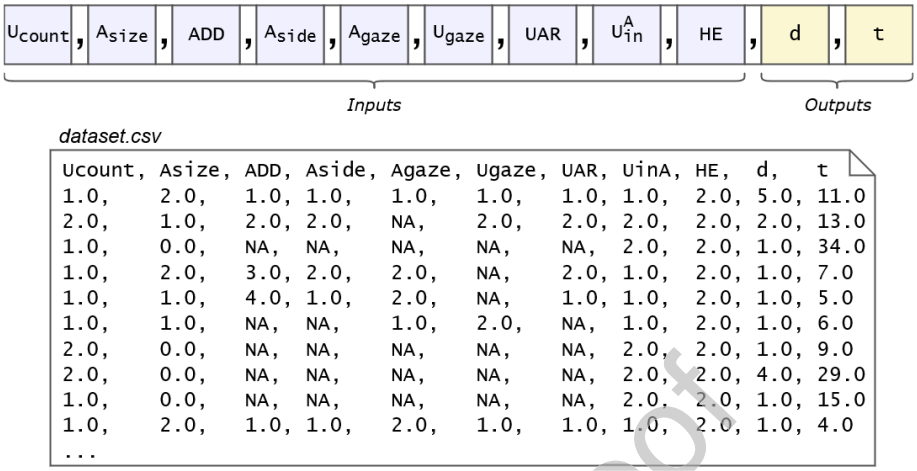

Figure 9: Structure of the CSV file, along with a sample file content. Note the presence of *NA* (i.e., not available) values: when A_{size} is equal to 0, many variables cannot be measured (namely ADD, A_{side}, A_{gaze}, U_{gaze}, UAR), which is in line with our choice of defining two models instead of one. Moreover, there are some cases where the right values of some variables cannot be correctly assessed (e.g., due to low video resolution or other uncertain situations): in such cases, *NA* values can still be used for training, due to the features of the machine learning approach adopted in this work.

5.4. Use Cases

We envision three possible use cases where our prediction model can be helpful.

Use Case 1: Planning and Improving Setups. Space owners can build a model for their layout according to our proposed method and then use it, along with the visual tool, for decision support when planning and improving public display setups. For example, based on the knowledge acquired from a model, a space owner who wants to achieve certain values for interaction durations or distances, could use the visual tool to experiment with different arrangements of seats around the display (which mainly affect the audience size), without actually making any changes in the real deployment, until the estimated values are satisfactory. This can be done as follows: the space owner should first collect data from a period of observations on the actual layout to generate a dataset, and use it to feed the model built according to our method.
The space owner can then use the visualisation tool to visualise PDF graphs and a heat map over the top-view picture of the layout, similar to those shown in Figure 8, which would help understanding how audience behaviour affects the interactions. Here the space owner uses the model to understand the behaviour resulting from the current layout, in terms of observed values for distance and duration of interactions.
The trained model can then be used to *predict* the performance of different setups *before* actually implementing them. For example, if a space owner would like to test the impact of changing the audience size, they can just set different values for the audience size vari-

Journal Pre-proof

able in the visualisation tool, thus simulating a different layout (e.g. a different number of benches around the display), and examining the resulting estimated values of interaction distance and duration. In other words, once the model has been trained for a given actual layout, which results in a given set of input variables and data, space owners can simulate different possible layouts, which means different arrangements of variables and data, without actually changing the layout itself. This allows for quicker and cheaper trials based on simulations and estimations from the proposed model, instead of expensive experimentation on different layout arrangements. This opportunity will empower the space owners to make key decisions to optimise their deployments. According to the result, the space owner can decide to add/remove seats around the display. Such a decision, which is based on the model predictions, would, in turn, affect the expected interaction distance and duration as desired.

Use Case 2: Real-time Prediction and Uses in Applications. A second use case is that the model could be used to forecast the users' behaviour before they approach the display. For example, assuming the input variables are collected automatically (e.g., a camera estimates the size of the audience and their gaze direction – as described in section 5.5), the system could dynamically decide to show the user different types of content (e.g., shorter videos) depending on the expected interaction durations. The system could also dynamically decide which input modality is better to enable – if users are likely to come close to the display, the system could allow for input via touch and disable other at-a-distance supported modalities (e.g., mid-air gestures).

Use Case 3: Facilitating Qualitative Comparisons across Deployments. Another possible use case is of particular interest for display providers. Typically, providers manage several deployments with different setups, and usually need tools for measuring performance of their deployments. Our visualisation tool would allow them to train multiple models (one per deployment), and easily compare the outputs by visually inspecting the differences in the heatmaps. Display providers can then use these estimations to provide qualitative data to their clients about their users' behaviour, allowing them to take informed decisions about arrangements setup or displayed content as explained in the previous use cases.

5.5. Data Collection Strategies

In the previous sections, we described the process of collecting data and how to use them in order to train our models and make predictions. However, such a data collection process may be non-trivial and very time consuming, in particular if the coding is done by watching videos.

However, all the variables we coded can be collected automatically or semi-automatically. In particular, the number of users, the audience size and the side at which they were present, as well as the audience-display distance, can be all collected using video processing tools specifically aimed at monitoring audience's behaviour, such as the AudienceMonitor [91] and/or the pedestrian tracker [92]. The same information can be collected over time to estimate if the honeypot effect occurs (e.g., by looking for peaks in the number of interactions), or if users were present during previous interactions (i.e., estimating variables HE and U_{in}^{A}). Users' and audience's gaze can be also captured by visual computing techniques such as

Journal Pre-proof

head pose estimation [93], and appearance-based gaze estimation [94, 95]. Furthermore, the relationship between the user(s) and the audience can be estimated by tracking the arrival of users to the display in a way similar to how Block et al. identified if a user is a member of a group [96]. Indeed, recognising such situation would allow to code an interaction event where the audience and the user are acquainted. Another possibility is to detect if a group of users interact with each other. Marin-Jimenez et al. demonstrated that this can be done using computer vision algorithms [97]. In those cases where some of the aforementioned variables cannot be identified with sufficient precision, they can be treated as N/A values, which means they are ignored during the fitting process due to the EM algorithm.

5.6. Extending the Use of the Model

The proposed approach is straightforward to extend by including other types of variables. We discuss some of the most promising ones below.

5.6.1. Time of the Day

Müller et al. observed a very diverse audience over the day [2]: school children in the morning, business people during lunch, and people shopping in the evening. In such cases, including the time of the day as an additional input variable would likely result in more accurate predictions. This can be done in two ways: one approach is to build different models for different periods of the days (e.g., one for the morning, one for the afternoon, and another one for the evening). In this case, the additional variable can be used as we did with the audience size A_{size} for selecting the right model accordingly (see Figure 7). Alternatively, the time of the day can be added as an additional discrete variable, and the EM algorithm can be used to estimate its probability density function as we described and did for all the other variables.

5.6.2. Interaction Modality

Another interesting consideration is about the interaction modalities. In our work, we considered a display that employs interaction via mid-air gestures. Hence we had the possibility of focusing on both the duration and distance. However, the same approach can be used for estimating interaction duration only, for instance in the case of touch-based interfaces, or eye gaze-based ones (where usually the user-display distance is required to be constant). Moreover, multimodal interface designers can use predictions in order to understand users' behaviours, and thus design different attracting sequence for different supported modalities, e.g., based on audience size and/or audience-display distance.

Similarly, instead of guiding users to the sweet spot [9, 10], a system that can estimate the interaction distance could dynamically determine the sweet spot, according to the input variables used for the predictions. For example, EyeScout adapts the position of an eye tracker based on the position of the user of a large public display as detected by a Kinect [17]. This can be improved by preemptively adapting the system according to the expected position of the incoming user.

Journal Pre-proof

5.6.3. Promising Variables to Explore in Future Work

Previous work suggested that variables that could impact the user's behaviour around displays include the screen's size and position [78], user's properties like gender, age and body orientation [77], as well as the time of the day [2], the weather, and the presence of nearby events [45]. Future work should explore incorporating these factors into the model building. Note that this would require very long deployments to collect enough data (e.g., how weather over the years impact attention to displays). Additionally, prior work showed that building layouts [32] and building architecture [51] influence user behaviour. Thus, future work can investigate how models can be trained to predict the impact of a building's layout and architecture on the behaviour of passersby and users of public displays.

6. Future Work

As discussed in section 3.2, the data collection process can be very time consuming if not automated. A direction for future work is to automatically detect the situation around the display. Tools such as those proposed by Williamson and Williamson [92] and by Elhart et al. [91] already track some of the relevant aspects of audience behaviour (e.g., audience size). These tools can be extended to collect additional information such as the audience's gaze by, for example, using head pose estimation [93, 98] or appearance-based gaze estimation [95]. Another tool proposed by Block et al. [96] infers the relationship between user groups, which can also be extended to estimate the relationship between the user and the audience in public display deployments. Collecting the aforementioned data automatically and feeding it into the model has the potential to result in more accurate models.

In our work we focused on predicting interaction distance and duration, which neither requires nor produces information about the number or frequency of interactions. Often, researchers and practitioners are interested in interaction frequency since it provides quantitative information about the passersby interest in the display's content. Another interesting direction for future work is to predict the number of expected interactions. This requires continuous tracking of the audience to pinpoint situations that led to interactions, and would thereby be practically feasible by using an automated tracking tool like the ones discussed earlier.

Finally, interaction distance and duration are not the only variables that can be predicted. The model may be extended to be trained in order to predict, for instance, the kind of gestures used (e.g. two-handed vs single-handed, or large movements vs subtle ones) according to a particular audience configuration.

7. Conclusion

In this work, we studied the impact of audience size and behaviour, along with contextual information, to predict duration and distance of interactions via mid-air gesture on public displays. Through a field study, we found that the interaction duration is influenced by the number of users, the audience size and the relationship between the users and the audience. We also found that the interaction distance is influenced by the audience size and

Journal Pre-proof

whether they gaze at the user(s). We used the collected data along with the expectation maximisation algorithm to build two predictor models to estimate the probability density functions of interaction duration and distance. Our results show that our models predict the interaction duration with a mean absolute error (MAE) of about 8 seconds and the interaction distance with a MAE of about 35 cm. We found that more accurate predictions can be achieved by using the variables that showed to have a significant influence on user behaviour. We also developed and presented a publicly available tool for visualising the results of our predictor models, making them more useful for researchers and practitioners. Our methodology can be also extended for including other contextual information (e.g., time of the day), and can be applied to many other situations by integrating it with many available systems for automatic data collection.

Acknowledgements

This research was supported by the Bavarian State Ministry of Education, Science and the Arts in the framework of the Center Digitization.Bavaria (ZD.B; Grant no. M7426.6.4) in Germany, and the German Research Foundation (DFG), Grant No. AL 1899/2-1. Moreover, this work was partially funded on two research grants by the Italian Ministry of University and Research (MIUR), namely project BookAlive (Grant no. PAC02L2_00068) and project NEPTIS (Grant no. PON03PE_00214_3). This work was supported, in part, by the Royal Society of Edinburgh (Award number 65040).

The authors would also like to thank our colleagues Andrea Scianna and Marcello La Guardia for their help with the acquisition of the ortophoto of our deployment in Palermo.

Finally, a special thanks goes to Maurizio Schifano for his invaluable support in the design of graphical assets used in our visual interface.

References

[1] N. Davies, S. Clinch, F. Alt, Pervasive Displays: Understanding the Future of Digital Signage, 1st Edition, Morgan & Claypool Publishers, 2014.

[2] J. Müller, R. Walter, G. Bailly, M. Nischt, F. Alt, Looking glass: A field study on noticing interactivity of a shop window, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12, ACM, New York, NY, USA, 2012, pp. 297–306. doi:10.1145/2207676.2207718.
URL http://doi.acm.org/10.1145/2207676.2207718

[3] C. Parker, M. Tomitsch, J. Kay, Does the public still look at public displays?: A field observation of public displays in the wild, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2 (2) (2018) 73:1–73:24. doi:10.1145/3214276.
URL http://doi.acm.org/10.1145/3214276

[4] M. Nancel, E. Pietriga, O. Chapuis, M. Beaudouin-Lafon, Mid-air pointing on ultra-walls, ACM Transactions on Computer-Human Interaction 22 (5) (2015) 21:1–21:62. doi:10.1145/2766448.
URL http://doi.acm.org/10.1145/2766448

[5] T. Dingler, M. Funk, F. Alt, Interaction proxemics: Combining physical spaces for seamless gesture interaction, in: Proceedings of the 4th International Symposium on Pervasive Displays, PerDis '15, ACM, New York, NY, USA, 2015, pp. 107–114. doi:10.1145/2757710.2757722.
URL http://doi.acm.org/10.1145/2757710.2757722

Journal Pre-proof

[6] N. Memarovic, I. Elhart, M. Langheinrich, Funsquare: First experiences with autopoiesic content, in: Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia, MUM '11, ACM, New York, NY, USA, 2011, pp. 175–184. doi:10.1145/2107596.2107619.
URL http://doi.acm.org/10.1145/2107596.2107619

[7] F. Alt, S. Torma, D. Buschek, Don't disturb me: Understanding secondary tasks on public displays, in: Proceedings of the 5th ACM International Symposium on Pervasive Displays, PerDis '16, ACM, New York, NY, USA, 2016, pp. 1–12. doi:10.1145/2914920.2915023.
URL http://doi.acm.org/10.1145/2914920.2915023

[8] V. Gentile, M. Khamis, S. Sorce, F. Alt, They are looking at me!: Understanding how audience presence impacts on public display users, in: Proceedings of the 6th ACM International Symposium on Pervasive Displays, PerDis '17, ACM, New York, NY, USA, 2017, pp. 11:1–11:7. doi:10.1145/3078810.3078822.
URL http://doi.acm.org/10.1145/3078810.3078822

[9] Y. Zhang, J. Müller, M. K. Chong, A. Bulling, H. Gellersen, Gazehorizon: Enabling passers-by to interact with public displays by gaze, in: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '14, ACM, New York, NY, USA, 2014, pp. 559–563. doi:10.1145/2632048.2636071.
URL http://doi.acm.org/10.1145/2632048.2636071

[10] F. Alt, A. Bulling, G. Gravanis, D. Buschek, Gravityspot: Guiding users in front of public displays using on-screen visual cues, in: Proceedings of the 28th Annual ACM Symposium on User Interface Software &; Technology, UIST '15, ACM, New York, NY, USA, 2015, pp. 47–56. doi:10.1145/2807442.2807490.
URL http://doi.acm.org/10.1145/2807442.2807490

[11] C. Ackad, M. Tomitsch, J. Kay, Skeletons and silhouettes: Comparing user representations at a gesture-based large display, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, ACM, New York, NY, USA, 2016, pp. 2343–2347. doi:10.1145/2858036.2858427.
URL http://doi.acm.org/10.1145/2858036.2858427

[12] M. Khamis, C. Becker, A. Bulling, F. Alt, Which one is me?: Identifying oneself on public displays, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, ACM, New York, NY, USA, 2018, pp. 287:1–287:12. doi:10.1145/3173574.3173861.
URL http://doi.acm.org/10.1145/3173574.3173861

[13] P. Peltonen, E. Kurvinen, A. Salovaara, G. Jacucci, T. Ilmonen, J. Evans, A. Oulasvirta, P. Saarikko, It's mine, don't touch!: Interactions at a large multi-touch display in a city centre, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '08, ACM, New York, NY, USA, 2008, pp. 1285–1294. doi:10.1145/1357054.1357255.
URL http://doi.acm.org/10.1145/1357054.1357255

[14] V. Gentile, S. Sorce, A. Malizia, F. Milazzo, A. Gentile, Investigating how user avatar in touchless interfaces affects perceived cognitive load and two-handed interactions, in: Proceedings of the 6th ACM International Symposium on Pervasive Displays, PerDis '17, ACM, New York, NY, USA, 2017, pp. 21:1–21:7. doi:10.1145/3078810.3078831.
URL http://doi.acm.org/10.1145/3078810.3078831

[15] F. Steinberger, M. Foth, F. Alt, Vote with your feet: Local community polling on urban screens, in: Proceedings of The International Symposium on Pervasive Displays, PerDis '14, ACM, New York, NY, USA, 2014, pp. 44:44–44:49. doi:10.1145/2611009.2611015.
URL http://doi.acm.org/10.1145/2611009.2611015

[16] P. C. Ng, J. She, K. E. Jeon, M. Baldauf, When smart devices interact with pervasive screens: A survey, ACM Transactions on Multimedia Computing, Communications, and Applications 13 (4) (2017) 55:1–55:23. doi:10.1145/3115933.
URL https://doi.org/10.1145/3115933

[17] M. Khamis, A. Klimczak, M. Reiss, F. Alt, A. Bulling, Eyescout: Active eye tracking for position and movement independent gaze interaction with large public displays, in: Proceedings of the 30th Annual ACM Symposium on User Interface Software & Technology, UIST '17, ACM, New York, NY, USA,

Journal Pre-proof

2017, pp. 155–166. doi:10.1145/3126594.3126630.
URL https://doi.org/10.1145/3126594.3126630

[18] V. Mäkelä, M. Khamis, L. Mecke, J. James, M. Turunen, F. Alt, Pocket transfers: Interaction techniques for transferring content from situated displays to mobile devices., in: Proceedings of the 36th Annual ACM Conference on Human Factors in Computing Systems, CHI '18, ACM, New York, NY, USA, 2018. doi:10.1145/3173574.3173709.
URL http://dx.doi.org/10.1145/3173574.3173709

[19] T. Dingler, T. Bagg, Y. Grau, N. Henze, A. Schmidt, ucanvas: A web framework for spontaneous smartphone interaction with ubiquitous displays, in: J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque, M. Winckler (Eds.), Human-Computer Interaction – INTERACT 2015, Vol. 9298 of Lecture Notes in Computer Science, Springer International Publishing, 2015, pp. 402–409. doi:10.1007/978-3-319-22698-9_27.
URL http://dx.doi.org/10.1007/978-3-319-22698-9_27

[20] G. Reyes, J. Wu, N. Juneja, M. Goldshtein, W. K. Edwards, G. D. Abowd, T. Starner, Synchrowatch: One-handed synchronous smartwatch gestures using correlation and magnetic sensing, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1 (4) (2018) 158:1–158:26. doi:10.1145/3161162.
URL http://doi.acm.org/10.1145/3161162

[21] V. Mäkelä, J. James, T. Keskinen, J. Hakulinen, M. Turunen, "it's natural to grab and pull": Retrieving content from large displays using mid-air gestures, IEEE Pervasive Computing 16 (3) (2017) 70–77. doi:10.1109/MPRV.2017.2940966.

[22] M. Khamis, R. Hasholzner, A. Bulling, F. Alt, Gtmopass: Two-factor authentication on public displays using gazetouch passwords and personal mobile devices, in: Proceedings of the 6th International Symposium on Pervasive Displays, PerDis '17, ACM, New York, NY, USA, 2017. doi:10.1145/3078810.3078815.
URL http://doi.acm.org/10.1145/3078810.3078815

[23] G. Broll, W. Reithmeier, P. Holleis, M. Wagner, Design and evaluation of techniques for mobile interaction with dynamic nfc-displays, in: Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction, TEI '11, ACM, New York, NY, USA, 2011, pp. 205–212. doi:10.1145/1935701.1935743.
URL http://doi.acm.org/10.1145/1935701.1935743

[24] R. de la Barré, P. Chojecki, U. Leiner, L. Mühlbach, D. Ruschin, Touchless interaction-novel chances and challenges, in: J. A. Jacko (Ed.), Human-Computer Interaction. Novel Interaction Methods and Techniques, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 161–169.

[25] L. E. Sibert, R. J. K. Jacob, Evaluation of eye gaze interaction, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '00, ACM, New York, NY, USA, 2000, pp. 281–288. doi:10.1145/332040.332445.
URL http://doi.acm.org/10.1145/332040.332445

[26] M. Khamis, O. Saltuk, A. Hang, K. Stolz, A. Bulling, F. Alt, Textpursuits: Using text for pursuits-based interaction and calibration on public displays, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '16, ACM, New York, NY, USA, 2016. doi:10.1145/2971648.2971679.
URL http://dx.doi.org/10.1145/2971648.2971679

[27] M. Khamis, L. Trotter, V. Mäkelä, E. von Zezschwitz, J. Le, A. Bulling, F. Alt, Cueauth: Comparing touch, mid-air gestures, and gaze for cue-based authentication on situated displays., Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2 (4). doi:10.1145/3287052.
URL https://doi.org/10.1145/3287052

[28] Microsoft, Human Interface Guidelines 2.0, http://download.microsoft.com/download/6/7/6/676611B4-1982-47A4-A42E-4CF84E1095A8/KinectHIG.2.0.pdf, accessed: 2018-06-19 (2014).

[29] S. Sorce, V. Gentile, C. Enea, A. Gentile, A. Malizia, F. Milazzo, A Touchless Gestural System for Extended Information Access Within a Campus, in: Proceedings of the 2017 ACM Annual Conference

Journal Pre-proof

on SIGUCCS, ACM, 2017, pp. 37–43. doi:10.1145/3123458.3123459.
URL https://dl.acm.org/citation.cfm?doid=3123458.3123459

[30] C. P. Gerba, A. L. Wuollet, P. Raisanen, G. U. Lopez, Bacterial contamination of computer touch screens, American Journal of Infection Control 44 (3) (2016) 358 – 360. doi:https://doi.org/10.1016/j.ajic.2015.10.013.
URL http://www.sciencedirect.com/science/article/pii/S0196655315010688

[31] P. Dalsgaard, K. Halskov, Designing urban media façades: Cases and challenges, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, ACM, New York, NY, USA, 2010, pp. 2277–2286. doi:10.1145/1753326.1753670.
URL http://doi.acm.org/10.1145/1753326.1753670

[32] A. Fatah gen Schieck, K. Al-Sayed, E. Kostopoulou, M. Behrens, W. Motta, Networked architectural interfaces: Exploring the effect of spatial configuration on urban screen placement, in: Proceedings of the Ninth International Space Syntax Symposium, SSS '13, SSS9, 2013.
URL https://discovery.ucl.ac.uk/id/eprint/1410314/

[33] R. Walter, G. Bailly, N. Valkanova, J. Müller, Cuenesics: Using mid-air gestures to select items on interactive public displays, in: Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services, MobileHCI '14, ACM, New York, NY, USA, 2014, pp. 299–308. doi:10.1145/2628363.2628368.
URL http://doi.acm.org/10.1145/2628363.2628368

[34] V. Gentile, S. Sorce, A. Malizia, D. Pirrello, 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, AVI '16, ACM, New York, NY, USA, 2016, pp. 40–43. doi:10.1145/2909132.2909282.
URL http://doi.acm.org/10.1145/2909132.2909282

[35] P. Bottoni, S. Faralli, A. Labella, A. Malizia, M. Pierro, S. Ryu, Copuppet : Collaborative interaction in virtual puppetry, in: R. Adams, S. Gibson, S. M. Arisona (Eds.), Transdisciplinary Digital Art. Sound, Vision and the New Screen, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 326–341.

[36] T. Ojala, V. Kostakos, H. Kukka, T. Heikkinen, T. Linden, M. Jurmu, S. Hosio, F. Kruger, D. Zanni, Multipurpose interactive public displays in the wild: Three years later, Computer 45 (5) (2012) 42–49. doi:10.1109/MC.2012.115.

[37] J. Coenen, S. Claes, A. V. Moere, The concurrent use of touch and mid-air gestures or floor mat interaction on a public display, in: Proceedings of the 6th ACM International Symposium on Pervasive Displays, PerDis '17, ACM, New York, NY, USA, 2017, pp. 9:1–9:9. doi:10.1145/3078810.3078819.
URL http://doi.acm.org/10.1145/3078810.3078819

[38] C. Ackad, A. Clayphan, M. Tomitsch, J. Kay, An in-the-wild study of learning mid-air gestures to browse hierarchical information at a large interactive public display, in: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '15, ACM, New York, NY, USA, 2015, pp. 1227–1238. doi:10.1145/2750858.2807532.
URL http://doi.acm.org/10.1145/2750858.2807532

[39] R. Walter, G. Bailly, J. Müller, Strikeapose: Revealing mid-air gestures on public displays, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, ACM, New York, NY, USA, 2013, pp. 841–850. doi:10.1145/2470654.2470774.
URL http://doi.acm.org/10.1145/2470654.2470774

[40] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, P. Irani, Consumed endurance: A metric to quantify arm fatigue of mid-air interactions, in: Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems, CHI '14, ACM, New York, NY, USA, 2014, pp. 1063–1072. doi:10.1145/2556288.2557130.
URL http://doi.acm.org/10.1145/2556288.2557130

[41] R. Aigner, D. Wigdor, H. Benko, M. Haller, D. Lindbauer, A. Ion, S. Zhao, J. Koh, Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci, Microsoft Research TechReport MSR-TR-2012-111 2 (2012) 30.

Journal Pre-proof

[42] J. Ruiz, D. Vogel, Soft-constraints to reduce legacy and performance bias to elicit whole-body gestures with low arm fatigue, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, ACM, New York, NY, USA, 2015, pp. 3347–3350. doi:10.1145/2702123.2702583.
URL http://doi.acm.org/10.1145/2702123.2702583

[43] M. R. Morris, J. O. Wobbrock, A. D. Wilson, Understanding users' preferences for surface gestures, in: Proceedings of Graphics Interface 2010, GI '10, Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 2010, pp. 261–268.
URL http://dl.acm.org/citation.cfm?id=1839214.1839260

[44] D. Ahlström, K. Hasan, P. Irani, Are you comfortable doing that?: Acceptance studies of around-device gestures in and for public settings, in: Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services, MobileHCI '14, ACM, New York, NY, USA, 2014, pp. 193–202. doi:10.1145/2628363.2628381.
URL http://doi.acm.org/10.1145/2628363.2628381

[45] V. Mäkelä, S. Sharma, J. Hakulinen, T. Heimonen, M. Turunen, Challenges in public display deployments: A taxonomy of external factors, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17, ACM, New York, NY, USA, 2017, pp. 3426–3475. doi:10.1145/3025453.3025798.
URL http://doi.acm.org/10.1145/3025453.3025798

[46] O. Mubin, T. Lashina, E. van Loenen, How not to become a buffoon in front of a shop window: A solution allowing natural head movement for interaction with a public display, in: Proc. of INTERACT '09, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 250–263. doi:10.1007/978-3-642-03658-3_32.
URL https://doi.org/10.1007/978-3-642-03658-3_32

[47] V. Gentile, Designing Touchless Gestural Interfaces for Public Displays, Ph.D. thesis, Università degli Studi di Palermo (2017).
URL https://iris.unipa.it/handle/10447/220972

[48] M. Ten Koppel, G. Bailly, J. Müller, R. Walter, Chained displays: Configurations of public displays can be used to influence actor-, audience-, and passer-by behavior, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12, ACM, New York, NY, USA, 2012, pp. 317–326. doi:10.1145/2207676.2207720.
URL http://doi.acm.org/10.1145/2207676.2207720

[49] A. Fatah gen Schieck, C. Biones, C. Mottram, The urban screen as a socialising platform: exploring the role of place within the urban space, in: MEDIACITY: Situations, Practices and Encounters, Frank & Timme GmbH, 2008, pp. 285–305.

[50] M. Behrens, A. Fatah gen Schieck, E. Kostopoulou, S. North, W. Motta, L. Ye, H. Schnadelbach, Exploring the effect of spatial layout on mediated urban interactions, in: Proceedings of the 2Nd ACM International Symposium on Pervasive Displays, PerDis '13, ACM, New York, NY, USA, 2013, pp. 79–84. doi:10.1145/2491568.2491586.
URL http://doi.acm.org/10.1145/2491568.2491586

[51] N. S. Dalton, E. Collins, P. Marshall, Display blindness?: Looking again at the visibility of situated displays using eye-tracking, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, ACM, New York, NY, USA, 2015, pp. 3889–3898. doi:10.1145/2702123.2702150.
URL http://doi.acm.org/10.1145/2702123.2702150

[52] N. Valkanova, R. Walter, A. Vande Moere, J. Müller, Myposition: Sparking civic discourse by a public interactive poll visualization, in: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '14, ACM, New York, NY, USA, 2014, pp. 1323–1332. doi:10.1145/2531602.2531639.
URL http://doi.acm.org/10.1145/2531602.2531639

[53] P. Dalsgaard, L. K. Hansen, Performing perception—staging aesthetics of interaction, ACM

Journal Pre-proof

Transactions on Computer-Human Interaction 15 (3) (2008) 13:1–13:33. doi:10.1145/1453152.1453156.
URL http://doi.acm.org/10.1145/1453152.1453156

[54] P. Dalsgaard, K. Halskov, O. S. Iversen, Participation gestalt: Analysing participatory qualities of interaction in public space, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, ACM, New York, NY, USA, 2016, pp. 4435–4446. doi:10.1145/2858036.2858147.
URL http://doi.acm.org/10.1145/2858036.2858147

[55] G. Beyer, V. Binder, N. Jäger, A. Butz, The puppeteer display: Attracting and actively shaping the audience with an interactive public banner display, in: Proceedings of the 2014 Conference on Designing Interactive Systems, DIS '14, ACM, New York, NY, USA, 2014, pp. 935–944. doi:10.1145/2598510.2598575.
URL http://doi.acm.org/10.1145/2598510.2598575

[56] H. Brignull, Y. Rogers, Enticing people to interact with large public displays in public spaces, in: In Proceedings of the IFIP International Conference on Human-Computer Interaction (INTERACT 2003, 2003, pp. 17–24.

[57] M. Khamis, F. Alt, A. Bulling, A field study on spontaneous gaze-based interaction with a public display using pursuits, 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, UbiComp/ISWC'15 Adjunct, ACM, New York, NY, USA, 2015, pp. 863–872. doi:10.1145/2800835.2804335.
URL http://doi.acm.org/10.1145/2800835.2804335

[58] P. Marshall, R. Morris, Y. Rogers, S. Kreitmayer, M. Davies, Rethinking 'multi-user': An in-the-wild study of how groups approach a walk-up-and-use tabletop interface, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, ACM, New York, NY, USA, 2011, pp. 3033–3042. doi:10.1145/1978942.1979392.
URL http://doi.acm.org/10.1145/1978942.1979392

[59] N. Wouters, J. Downs, M. Harrop, T. Cox, E. Oliveira, S. Webber, F. Vetere, A. Vande Moere, Uncovering the honeypot effect: How audiences engage with public interactive systems, in: Proceedings of the 2016 ACM Conference on Designing Interactive Systems, DIS '16, ACM, New York, NY, USA, 2016, pp. 5–16. doi:10.1145/2901790.2901796.
URL http://doi.acm.org/10.1145/2901790.2901796

[60] N. Memarovic, A. Fatah gen Schieck, H. Schnädelbach, E. Kostopoulou, S. North, L. Ye, Longitudinal, cross-site and "in the wild": A study of public displays user communities' situated snapshots, in: Proceedings of the 3rd Conference on Media Architecture Biennale, MAB, Association for Computing Machinery, New York, NY, USA, 2016. doi:10.1145/2946803.2946804.
URL https://doi.org/10.1145/2946803.2946804

[61] N. Streitz, C. Rocker, T. Prante, R. Stenzel, D. van Alphen, Situated interaction with ambient information: Facilitating awareness and communication in ubiquitous work environments (2003).

[62] D. Vogel, 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, ACM, New York, NY, USA, 2004, pp. 137–146. doi:10.1145/1029632.1029656.
URL http://doi.acm.org/10.1145/1029632.1029656

[63] D. Michelis, J. Müller, The audience funnel: Observations of gesture based interaction with multiple large displays in a city center, International Journal of Human-Computer Interaction 27 (6) (2011) 562–579. doi:10.1080/10447318.2011.555299.
URL http://dx.doi.org/10.1080/10447318.2011.555299

[64] J. Müller, F. Alt, D. Michelis, A. Schmidt, Requirements and design space for interactive public displays, in: Proceedings of the 18th ACM International Conference on Multimedia, MM '10, ACM, New York, NY, USA, 2010, pp. 1285–1294. doi:10.1145/1873951.1874203.
URL http://doi.acm.org/10.1145/1873951.1874203

Journal Pre-proof

[65] P. T. Pasqual, J. O. Wobbrock, Mouse pointing endpoint prediction using kinematic template matching, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14, ACM, New York, NY, USA, 2014, pp. 743–752. doi:10.1145/2556288.2557406.
URL http://doi.acm.org/10.1145/2556288.2557406

[66] D. Buschek, A. De Luca, F. Alt, Improving accuracy, applicability and usability of keystroke biometrics on mobile touchscreen devices, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, ACM, New York, NY, USA, 2015, pp. 1393–1402. doi:10.1145/2702123.2702252.
URL http://doi.acm.org/10.1145/2702123.2702252

[67] D. Buschek, O. Schoenleben, A. Oulasvirta, Improving accuracy in back-of-device multitouch typing: A clustering-based approach to keyboard updating, in: Proceedings of the 19th International Conference on Intelligent User Interfaces, IUI '14, ACM, New York, NY, USA, 2014, pp. 57–66. doi:10.1145/2557500.2557501.
URL http://doi.acm.org/10.1145/2557500.2557501

[68] K. Hinckley, S. Heo, M. Pahud, C. Holz, H. Benko, A. Sellen, R. Banks, K. O'Hara, G. Smyth, W. Buxton, Pre-touch sensing for mobile interaction, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, ACM, New York, NY, USA, 2016, pp. 2869–2881. doi:10.1145/2858036.2858095.
URL http://doi.acm.org/10.1145/2858036.2858095

[69] S. Frintrop, E. Rome, H. I. Christensen, Computational visual attention systems and their cognitive foundations: A survey, ACM Transactions on Applied Perception 7 (1) (2010) 6:1–6:39. doi:10.1145/1658349.1658355.
URL http://doi.acm.org/10.1145/1658349.1658355

[70] J. Huang, R. White, G. Buscher, User see, user point: Gaze and cursor alignment in web search, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12, ACM, New York, NY, USA, 2012, pp. 1341–1350. doi:10.1145/2207676.2208591.
URL http://doi.acm.org/10.1145/2207676.2208591

[71] O. Erazo, J. A. Pino, Predicting task execution time on natural user interfaces based on touchless hand gestures, in: Proceedings of the 20th International Conference on Intelligent User Interfaces, IUI '15, ACM, New York, NY, USA, 2015, pp. 97–109. doi:10.1145/2678025.2701394.
URL http://doi.acm.org/10.1145/2678025.2701394

[72] C. Shin, J.-H. Hong, A. K. Dey, Understanding and prediction of mobile application usage for smart phones, in: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, UbiComp '12, ACM, New York, NY, USA, 2012, pp. 173–182. doi:10.1145/2370216.2370243.
URL http://doi.acm.org/10.1145/2370216.2370243

[73] Y. Xu, M. Lin, H. Lu, G. Cardone, N. Lane, Z. Chen, A. Campbell, T. Choudhury, Preference, context and communities: A multi-faceted approach to predicting smartphone app usage patterns, in: Proceedings of the 2013 International Symposium on Wearable Computers, ISWC '13, ACM, New York, NY, USA, 2013, pp. 69–76. doi:10.1145/2493988.2494333.
URL http://doi.acm.org/10.1145/2493988.2494333

[74] B. Huber, J. H. Lee, J.-H. Park, Detecting user intention at public displays from foot positions, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, ACM, New York, NY, USA, 2015, pp. 3899–3902. doi:10.1145/2702123.2702148.
URL http://doi.acm.org/10.1145/2702123.2702148

[75] S. Clinch, J. Alexander, S. Gehring, A survey of pervasive displays for information presentation, IEEE Pervasive Computing 15 (3) (2016) 14–22. doi:10.1109/MPRV.2016.55.

[76] N. Valkanova, S. Jorda, A. V. Moere, Public visualization displays of citizen data: Design, impact and implications, International Journal of Human-Computer Studies 81 (2015) 4 – 16, transdisciplinary Approaches to Urban Computing. doi:https://doi.org/10.1016/j.ijhcs.2015.02.005.
URL http://www.sciencedirect.com/science/article/pii/S1071581915000282

[77] T. Ballendat, N. Marquardt, S. Greenberg, Proxemic interaction: Designing for a proximity and

Journal Pre-proof

orientation-aware environment, in: ACM International Conference on Interactive Tabletops and Surfaces, ITS '10, ACM, New York, NY, USA, 2010, pp. 121–130. doi:10.1145/1936652.1936676.
URL http://doi.acm.org/10.1145/1936652.1936676

[78] K. M. Zielinska-Dabkowska, A. Fatah gen Schieck, Designing digital displays and interactive media in today's cities by night. do we know enough about attracting attention to do so?, Conscious Citiesdoi:10.33797/CCA18.01.

[79] P. Ranganathan, C. S. Pramesh, R. Aggarwal, Common pitfalls in statistical analysis: Measures of agreement, Perspectives in clinical research 8 (4) (2017) 187–191. doi:10.4103/picr.PICR_123_17.

[80] C. M. Bishop, Machine learning and pattern recognition, Springer, 2006.

[81] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[82] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: Nsga-ii, IEEE Transactions on Evolutionary Computation 6 (2) (2002) 182–197. doi:10.1109/4235.996017.

[83] A. Krogh, J. Vedelsby, Neural network ensembles, cross validation and active learning, in: Proceedings of the 7th International Conference on Neural Information Processing Systems, NIPS'94, MIT Press, Cambridge, MA, USA, 1994, pp. 231–238.
URL http://dl.acm.org/citation.cfm?id=2998687.2998716

[84] R. E. Wunderlich, P. R. Cavanagh, Gender differences in adult foot shape: implications for shoe design, Medicine & Science in Sports & Exercise 33 (4). doi:10.1097/00005768-200104000-00015.

[85] E. Hall, The Hidden Dimension, Anchor Books, 1992.

[86] M. Perry, S. Beckett, K. O'Hara, S. Subramanian, Wavewindow: Public, performative gestural interaction, in: ACM International Conference on Interactive Tabletops and Surfaces, ITS '10, ACM, New York, NY, USA, 2010, pp. 109–112. doi:10.1145/1936652.1936672.
URL http://doi.acm.org/10.1145/1936652.1936672

[87] M. Argyle, J. Dean, Eye-contact, distance and affiliation, Sociometry 28 (3) (1965) 289–304.
URL http://www.jstor.org/stable/2786027

[88] F. Lundh, An introduction to tkinter, http://www.pythonware.com/library/tkinter/introduction/index.htm (1999).

[89] J. D. Hunter, Matplotlib: A 2d graphics environment, Computing in Science Engineering 9 (3) (2007) 90–95. doi:10.1109/MCSE.2007.55.

[90] F. Lundh, Pillow, https://python-pillow.org/ (2011).

[91] I. Elhart, M. Mikusz, C. G. Mora, M. Langheinrich, N. Davies, Audience monitor: An open source tool for tracking audience mobility in front of pervasive displays, in: Proceedings of the 6th ACM International Symposium on Pervasive Displays, PerDis '17, ACM, New York, NY, USA, 2017, pp. 10:1–10:8. doi:10.1145/3078810.3078823.
URL http://doi.acm.org/10.1145/3078810.3078823

[92] J. R. Williamson, J. Williamson, Analysing pedestrian traffic around public displays, in: Proceedings of The International Symposium on Pervasive Displays, PerDis '14, ACM, New York, NY, USA, 2014, pp. 13:13–13:18. doi:10.1145/2611009.2611022.
URL http://doi.acm.org/10.1145/2611009.2611022

[93] G. Fanelli, J. Gall, L. V. Gool, Real time head pose estimation with random regression forests, in: CVPR 2011, 2011, pp. 617–624. doi:10.1109/CVPR.2011.5995458.

[94] E. Wood, T. Baltruaitis, X. Zhang, Y. Sugano, P. Robinson, A. Bulling, Rendering of eyes for eye-shape registration and gaze estimation, in: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), ICCV '15, IEEE Computer Society, Washington, DC, USA, 2015, pp. 3756–3764. doi:10.1109/ICCV.2015.428.
URL http://dx.doi.org/10.1109/ICCV.2015.428

[95] X. Zhang, Y. Sugano, M. Fritz, A. Bulling, Appearance-based gaze estimation in the wild, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 4511–4520. doi:10.1109/CVPR.2015.7299081.

Journal Pre-proof

[96] F. Block, J. Hammerman, M. Horn, A. Spiegel, J. Christiansen, B. Phillips, J. Diamond, E. M. Evans, C. Shen, Fluid grouping: Quantifying group engagement around interactive tabletop exhibits in the wild, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, ACM, New York, NY, USA, 2015, pp. 867–876. doi:10.1145/2702123.2702231.
URL http://doi.acm.org/10.1145/2702123.2702231

[97] M. J. Marin-Jimenez, A. Zisserman, M. Eichner, V. Ferrari, Detecting people looking at each other in videos, International Journal of Computer Vision 106 (3) (2014) 282–296. doi:10.1007/s11263-013-0655-7.
URL https://doi.org/10.1007/s11263-013-0655-7

[98] E. Murphy-Chutorian, M. M. Trivedi, Head pose estimation in computer vision: A survey, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (4) (2009) 607–626. doi:10.1109/TPAMI.2008.106.

Journal Pre-proof

Journal Pre-proof

DECLARATION OF ABSENCE OF CONFICT OF INTEREST

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address (vito.gentile@unipa.it), which is accessible by the Corresponding Author and which has been configured to accept incoming emails from the journal editorial board and staff.

Signed by all authors as follows:

Vito Gentile

Mohamed Khamis

Fabrizio Milazzo

Salvatore Sorce

Alessio Malizia

Florian Alt

Date: December 22nd, 2018

Journal Pre-proof

Journal Pre-proof

CRediT author statement

Vito Gentile:

- Conceptualization
- Methodology
- Software
- Validation
- Formal analysis
- Investigation
- Data Curation
- Writing - Original Draft
- Writing - Review & Editing
- Visualization

Mohamed Khamis:

- Conceptualization
- Methodology
- Validation
- Formal analysis
- Investigation
- Data Curation
- Writing - Original Draft
- Writing - Review & Editing
- Visualization

Fabrizio Milazzo:

- Conceptualization
- Methodology
- Software

Salvatore Sorce:

- Conceptualization
- Validation
- Investigation
- Data Curation
- Writing - Original Draft
- Writing - Review & Editing
- Supervision
- Funding acquisition
- Project administration

Alessio Malizia:

- Conceptualization
- Writing - Review & Editing
- Supervision
- Funding acquisition

Florian Alt:

- Conceptualization
- Methodology
- Resources
- Writing - Review & Editing
- Supervision
- Funding acquisition

Journal Pre-proof