Interacting With Laughter: A Case Study On Audio-Based Interactivity of Public Projections

Luke Hespanhol
Design Lab – Faculty of Architecture, Design and Planning
The University of Sydney NSW 2006, Australia
luke.hespanhol@sydney.edu.au

ABSTRACT
In this paper, I present a field study conducted with interactive projections deployed to an urban art festival and designed to give audio-visual feedback in response to audio input from the public. I describe the context for the study, the spatial layout of the urban space used, the methodology adopted, and the results of my observations of 900 people visiting the work over 5 days. Based on the results, I derive a series of insights about the utilization of audio-based interactive projections in public spaces, notably in regards to: making passers-by aware of the interaction; the placement of the microphones relative to the projections; the impact of the spatial layout of the urban precinct on participation; and pitfalls to avoid regarding the visual elements existing at or introduced to the urban space.

Author Keywords
Urban interaction design; responsive environments; audio-based interaction; urban interfaces.

ACM Classification Keywords
D H.5.2 [User Interfaces]: Interaction Styles; I.3.6 [Methodology and Techniques]: Interaction Techniques.

INTRODUCTION

Background
As digital environments increasingly become part of the urban landscape, so does the need for greater understanding about how making them interactive can assist with their integration into the broader built environment and local social dynamics. Significant field research has been conducted in the past decade, mostly involving visual systems that passers-by could engage with via full-body interaction or tangible user interfaces (Behrens et al., 2013; Fatah gen. Schieck et al., 2013; Hespanhol et al., 2012; Memarovic et al., 2012; Michelis and Mueller, 2011; Mueller et al., 2012; Walter et al., 2013). More specifically, recent studies such a the one by Kukka et al. (2016) have focused on the use of audio-based cues as a mechanism to increase awareness of passers-by about situated interactive public displays, attracting attention towards them and helping to overcome *display blindness* (Mueller et al., 2009; Memarovich et al., 2015).

OzCHI'16, November 29-December 02, 2016, Launceston, TAS, Australia

ACM 978-1-4503-4618-4/16/11...$15.00
DOI: http://dx.doi.org/10.1145/3010915.3010931

Figure 1. *LOL*'s audio-based interactive projections floating in the waters of Walsh Bay, Sydney, Australia.

However, the use of audio as main input modality for interaction with digitally augmented public environments remains, to the best of our knowledge, a lesser explored research scenario – at times dismissed outright due to conflict with other situated audio-based social interactions, as conversations around digital displays in cafés (Kray et al., 2008), for example. In an attempt to avoid such a conflict while still encouraging encounters between people sharing a public space, Jarusriboonchai et al. (2014) investigated audio-based interpersonal interactions via mobile devices. As the authors observed, however, such a use of mobile devices produced a contradictory effect: while successful in triggering a broad variety of social interactions, it also shifted participants' attention from each other to the devices themselves, thus hindering situated social interaction.

Research aims
In light of the limitations in the field outlined above, I propose, in this paper, to investigate audio as sole input for interaction with projections in public spaces. In particular, I aim at further understanding its potential for triggering social interactions, as well as any issues related to its utilisation in the context of a public digital media environment. To that end, I present a field study consisting of an interactive art installation (Figure 1) I developed for an urban festival, aimed at promoting collaborative interaction by giving *audio-visual* feedback in response to *audio-only* input from the crowd. In that

regard, I sought to establish a very low entry barrier for interaction by making the installation responsive purely to the volume and duration of audio produced by visitors, regardless of its nature (e.g. spoken words, screams, clapping, etc.). As feedback, I chose to display videos of laughing human mouths and a corresponding cacophony of bursting laughter, with the intention of enticing the audience to laugh back, therefore establishing a feedback loop. Key research goals included: (1) to gauge the *level of intuitiveness* of using audio as the only interaction mechanism with interactive projections; and (2) to understand the impact of the urban site's *spatial layout* on the way passers-by approached the installation.

RELATED WORK

Recent research in gestural interfaces and multimodal feedback has generally adopted public displays, projections or media façades as platforms for urban interaction. Studies have suggested that core factors influencing the experience of interaction in public spaces are (a) the extent to which the physical layout of the surrounding urban location facilitates the interactive system to be approached by passers-by; (b) how passers-by become aware of and learn the interaction mechanisms; and (c) the types of feedback given to input actions and how they help to make interfaces intuitive.

Spatial Layout and Interaction Patterns

Dalton et al. (2013) combined principles from space syntax theory with the concept of visibility catchment area (taken from the design of signage for emergency egress) in the development of software algorithms for determining the optimal position of displays in a public space. Likewise, Fatah gen. Schieck et al. (2013) used ethnographic and space syntax methods to understand properties of the urban environment, analyse the urban space in terms of visibility and accessibility, map local social groups and practices, and identify how spatial and visual properties of the targeted public display locations corresponded to social usability, co-presence and movement activity. The results of the observations helped the researchers to inform the selection of a range of screen placements, as well as to later observe the influence of spatial layout on the behaviour of actors, spectators and passers-by as they negotiated the space in front the screens. Fischer and Hornecker (2012) also focused on the crowd dynamics around an interactive media façade, investigating the influence of spatial layout on how people self-organize relatively to it as well as to each other. Analysing how social interactions evolved over time across the space, they classified specific zones according to the impact they created on public behaviour.

Awareness and Learnability

Arroyo et al. (2012) combined motion and proximity sensors with feedback from audio and projections to augment a water fountain installed in a corridor of a university campus. The main goal of the study was to investigate the effectiveness of three different feedback modalities – audio only, visual only, and audio-visual combined – in increasing awareness about and usage of the fountain over a long period of time. The study results revealed that embedded interactions can be more effective at engaging the public if they vary feedback and are responsive to the number of people within the space. Moreover, audio feedback is preferable when few are present in the environment, whereas visual feedback is preferable with a larger number of people in the space.

Mueller et al. (2012) ran a series of lab and field studies using displays to compare different levels of visual feedback on input actions from passers-by. They derived various important findings regarding the nature of the feedback to individuals, as well as the effects of the physical placement of screens, to cater for common behavioural patterns such as the *landing effect*: the observed behaviour when passers-by only become aware of the responsiveness of a public display a few moments after walking past it – by which time they then have to stop and walk back to check it for a second time. They also observed the emergence of multiple rows of participants in front of the displays, whereby spectators were able to learn how to interact on the spot by observing others interacting with the system.

Schroeder et al. (2012) investigated the influence of urban location, audience demographics and type of media content on participation levels. They defined as a "sweet spot" a design context where the choice of location facilitates audience members to find content that has, in turn, been adequately targeted at them. Conversely, where any of those factors are not optimal, participation levels may suffer in quantity (small number of interactions) as well as in quality (interactions that are inappropriate or erroneous). Following from that, Steinberger et al. (2014) leveraged on the affordances provided by a bus stop – in the sense that it is designed to accommodate people spending some idle time while waiting for transport – to transform it into a public survey interface. For that, they installed a public screen in the bus shelter structure to display yes/no questions about a variety of community topics, and placed two large push buttons on the floor – one labelled with "Yes", the other with "No" – connected to a computer system. By stepping on either button, an individual could therefore provide an answer for the question displayed on the screens. The interface articulated a minimal design that allowed very specific affordances: the buttons had clear instructions and were placed on the floor, allowing few actions other than being stepped on. The authors found that such a simple interface, by standing out from its surroundings, managed to attract attention to an otherwise peripheral display, while also revealing the importance of designing for effortless interaction: as pointed out by their study, users tend to value a low entry barrier over the drawbacks of being exposed to spectators (Steinberger et al., 2014).

However, not all public interactive settings can afford such a minimalistic interface and easily understandable feedback. When designing more complex scenarios, research by Bedwell and Caruana (2012) suggested that it might be helpful to have non-experts demonstrating the mechanisms of interaction to the public. O'Hara et al. (2008) also acknowledged the important role played by local hosts demonstrating the 'rules of engagement' with public interactive media, arguing that they enable the engagement to be enhanced by providing explanation,

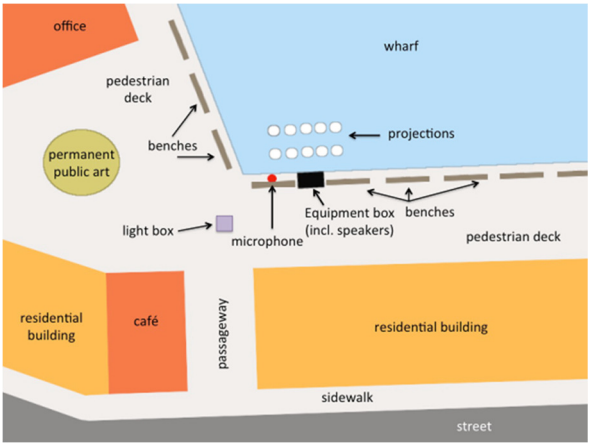

Figure 2. Spatial layout of the Walsh Bay precinct, where *LOL* was installed.

motivation and commentary. Yet, the also acknowledged that, even in complex settings, demonstrators may not be available all the time. To that end, they defined the notion of *accidental interaction* as one that occurs when a member of the public walks across the interactive space on the way to somewhere else with no intention of participating in the interaction, but is inadvertently tracked by the installation, which then responds accordingly. As the authors pointed out, accidental interaction can come to play a significant role in supporting participation in settings where demonstrators are not present, as an important strategy for both raising awareness about the interactive setting in public spaces, and enabling immediate learnability of the basic interactive mechanisms. As I will explain in the next section, in addition to an effortless interaction mechanism enabling accidental interaction, direct demonstration about how to interact with the installation was another of the strategies for engaging passers-by I evaluated during the execution of the field study described in this paper.

Feedback Modalities and Intuitiveness

To entice a smooth adoption by the public is a recurrent concern when it comes to designing public responsive environments. Antle et al. (2009) conducted studies on intuitiveness that investigated full body interaction in a responsive auditory environment. They compared two different mapping strategies, one where gestures were based on embodied metaphors and another where they were not. While results indicated that mental models derived from physical affordances led to greater non-conscious interaction, they were also not sufficient to imply intuitiveness, unless made easily discoverable. In other words, embodied metaphors only led to intuitively enacted input actions when accompanied by feedback clear and direct enough to be perceived without effort. One possible strategy for achieving such clarity is to take into account the context of the interaction and purpose of the feedback (Hespanhol et al., 2013). For example, when visual feedback is used to mirror the gestural input of participants, audio feedback can be employed to warn them about upcoming events. In that case, visual feedback is directed at individual participants, while audio is reserved for environment-wide cueing (provided that it is loud enough to be heard across the whole precinct, and requirements for ambient noise limits at the public space are observed). If cueing is unnecessary, both modalities can then be unified as feedback to the audience. This is an aspect I paid particular attention to during the design of my case study, as I expected that unifying audio and visual feedbacks could lead to quicker understanding of the interactive mechanisms by people casually strolling through my target public space.

Adopting the research in gestural interfaces as a model, I hoped that my study could provide some initial insights in the design of audio-based interactive public environments, particularly for scenarios when audio is used as only input mechanism, followed by audio-visual feedback. The next section will explain the context and spatial layout for my case study in greater detail, followed by conceptual and design considerations. I then present the methodology adopted, the results from my field observations and derive a series of recommendations for similar design scenarios.

CONTEXT AND SPATIAL LAYOUT

LOL (Figure 1) was a prominent interactive art installation I developed for the 2014 edition of Vivid Sydney, a large annual festival of arts in Sydney, Australia. The festival runs for 18 nights during winter and features around 60 light installations distributed around various venues in the city central district. The 2014 edition of the festival attracted 1.43 million visitors. The installation ran for 6 hours per night, during 18 nights (6pm to midnight), and was deployed to Walsh Bay, part of the city's original docklands. After an urban revitalization project, the area is now home to various residential apartments, restaurants, bars and corporate offices. The dominant wharf section runs parallel to the main local street, separated from it by apartment buildings; passageways are available between the buildings, connecting the street to the wharf area. The wharf is still functional, serving private boats and a public ferry service. A deck, accessible from the street, runs along the entire marine precinct, which is closed for car traffic. Figure 2 displays a diagram of the site's spatial layout, also marking the spot allocated for the installation. Figure 3 shows a photo of the interactive space viewed from the passageway connecting the street to the wharf, a

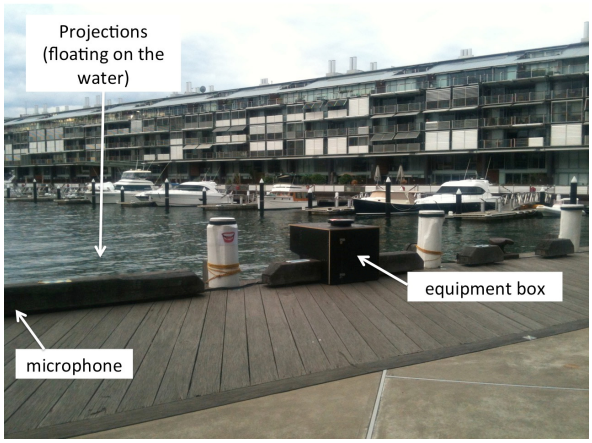

Figure 3. *LOL*'s interactive space, with the equipment box positioned along the pier's edge.

Figure 4. Pedestrians interacting from the deck around the spot where the microphone was installed (input side).

common route for pedestrians approaching the precinct. As I will describe later in the paper, this approaching direction – and the fact passers-by coming that way were not able to see the actual balls and projections at the water level, only potentially hear them – would prove to be highly valuable to explain some of the interactive behaviour observed.

CONCEPT AND DESIGN CONSIDERATIONS

LOL consisted of 10 inflatable balls resting on a platform, floating on the waters of the urban marina. Each ball would display a projection of an animated human mouth, which in the absence of external input would remain shut. When noise above a certain threshold was detected in the nearby environment, the mouths would burst into loud laughter, enticing the public to laugh back and, consequently, keep the interaction loop going for as much as they felt like. Conversely, once noise fell under the set threshold, the projections would return to their resting state. I assigned different noise thresholds for each projection, so that varying levels of input audio would make the installation to respond accordingly: occasional noise would cause only a couple of mouths to react; sustained, loud noise would make them all respond. The projector was fit within a waterproof equipment box installed on the pedestrian deck overlooking the water (Figures 2 and 3), from which passers-by could also appreciate and interact with the work (Figure 4).

As pointed out by Kukka et al. (2016), when considering using audio as an attractor or feedback mechanism to a public display, it is important to first consider the ambient noise level and the suitability of added audio in the given space. As Figures 2 illustrates, the spot allocated for the work was near a corner of the wharf site, surrounded by apartment and office buildings and close to one of the passageways connecting the main street to the marine precinct. Such a location posed strong technical and logistic constraints to the design. The main setting for the work – the noise threshold above which the projections would respond to people – was therefore determined by calibration sessions ran onsite prior to the festival start, considering factors such as: (a) average volume of noise captured from the platform; (b) echoes produced by the buildings surrounding the wharf; and (c) the trade-off between the volume produced by the balls laughing and the sensitivity of the microphone used to capture the input audio. The combination of all those settings should be such that the installation would be activated only by noise made by passers-by (intentionally or not), rather than by the audio produced by the installation itself. To that end, a crucial design step was to determine the relative positioning of the microphone capturing the input noise and the speakers producing the output audio. Based on discussions with the local public space managers, I decided to house the speakers alongside the projector inside the waterproof equipment box, therefore providing a less loud and more situated feedback to passers-by around the installation. Due to logistic limitations, a single microphone was used. Figures 3 and 4 illustrate the final setup *in situ*.

An important design goal for the study was to investigate the intuitiveness of an audio-based interactive environment, as well as to what extent the spatial layout and visibility of the projections would influence the way passers-by would approach the installation. For that reason, I decided that there should be no signage indicating the exact placement of the microphone within the space; rather, I was interested in finding out whether a "sweet spot" for the interaction, as described by Schroeder et al. (2012), could emerge spontaneously once the local audience started to make sense of the setting. In order to increase the chances of passers-by engaging in interaction with the work, I positioned the microphone slightly to the left-hand side of the projections, near the corner of the wharf and towards the passageway most pedestrians used to enter the marine precinct. My hypothesis was that people would most often approach the installation from that side; therefore placing the microphone at that location would result in greater number of interactions. Following the same hypothesis, the festival organization installed at the end of the passageway (Figure 2) a light box with captions about the concept and explicit instructions on how to interact with it. In addition to that, following festival standards, the organizers also arranged a volunteer to mind the installation during most of the time it was operational. Usually, that involved mere passive surveillance, but occasionally volunteers would actively demonstrate to passers-by how to engage in interaction with the installation, particularly when very few people were around, in order to boost participation from the public. I purposely avoided deploying additional explanatory elements other than those already made available by the festival, i.e. captions and occasional demonstration of the work to passers-by. Given my research interest in probing intuitiveness and the influence of spatial layout over the interaction, I refrained from designing clear physical affordances (Norman, 2004) or call to actions that might indicate to passers-by how to interact with the work. I could thus observe how significant the influence of either the captions or demonstrations would be over participation. Likewise, I was able to observe whether those approaching the installation spontaneously – i.e. not consulting the captions nor watching demonstrations in the first place – could perceive the interactivity of the projections at all.

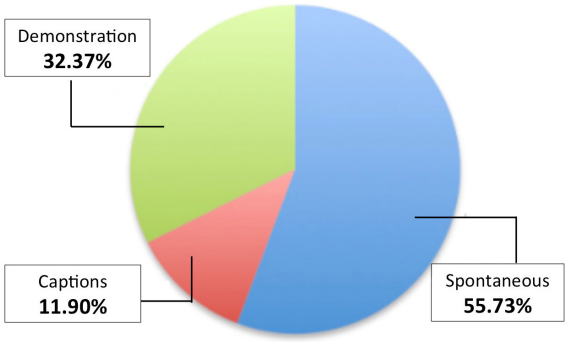

Figure 5. Motivational factors and percentage of total interactions informed by each of them.

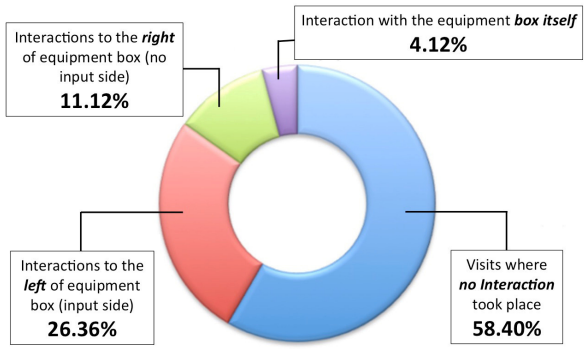

Figure 6. Overall spots where interactions were attempted, in relation to the installation elements.

METHODOLOGY

I observed a total of 900 passers-by approaching the installation over 5 nights, between 6:30pm and 7:30pm. I selected this time slot since it corresponded to peak time for visits to the festival, allowing my observations to reflect the natural flow of people through the public space. Each night, I attempted to observe (1) motivational factors for the interaction, and (2) the spots where interactions unfolded. *Motivational factors* refer to elements that passers-by could be exposed to on their way to the installation, informing them about the interactive nature of the work while also teaching them about how to interact. In that sense, participation could be motivated by (a) reading the *captions* about the work; (b) witnessing other people (passers-by or festival volunteers) *demonstrating* how to interact; or (c) just by approaching the installation and making *spontaneous* noise around it. By capturing those metrics, I expected to address my goals of understanding the impact of the local spatial layout on the way passers-by approached the installation, and of gauging the intuitiveness of using audio as sole mechanism to interact with projections in a public space.

In regards to the actual spots where the interactions took place, an obvious strategy was to consider the equipment box as dividing element for the space, given the asymmetry between the output audio (produced by speakers stored in the box) and the input audio (captured by the microphone). While the speakers were centralized in relation to the projections, I positioned the input microphone towards the passageway giving access to the precinct, which was also next to the spot where the festival organizers positioned the captions. For the purposes of my analysis, I shall refer to the side of the box where the microphone was positioned as the input (or left-hand) side, with the other area (away from the microphone) constituting the no input (or right-hand) side (Figures 2 and 4). Counting the occurrence of interactions on each side could therefore provide me with insights into the effect of the spatial layout on the way people approached the installation. For each attempted interaction, I also counted whether or not it was successful (i.e. triggered a response).

FIELD OBSERVATIONS AND RESULTS

Figure 5 illustrates the results of my observations on motivational factors, considering only the visits that resulted on interactions (374/900, or 41.60%), i.e. visitors who I observed trying to interact with the work, even if unsuccessfully. In the absence of clear visual affordances, some passers-by spontaneously explored the interface in a variety of ways. While some would result in successful interactions, others would trigger no response from the work, e.g. waving to the projections, or touching the equipment box. A common form of "calling out" to the work would be by clapping hands. While the captions explicitly instructed visitors to laugh at the work (as per the original concept), I often observed that festival volunteers combined laughter with clapping hands to activate the installation. In order to simplify data gathering and analysis, I did not count the occurrences of each of those different strategies used by people while exploring the interface. However, preliminary observations revealed that some people trying to interact spontaneously would misinterpret the interface and try to touch or talk to the equipment box itself. That prompted me to break down the observed interactions into three spatial zones: (1) the input side; (2) the no input side; and (3) the equipment box itself. In that regard, it is important to point out that in the scenarios where visitors both read the captions and watched other people interacting, or interacted spontaneously then proceeded to read the captions, I counted only whatever occurred first. Figure 7 displays the percentage of the total visits driven by each motivational factor and on each side of the installation, broken down by outcome (no interaction, success or failure. Figure 6 shows the results aggregated by location. I also observed that people felt more inclined to join in collaborative interaction either (a) when others started to clap; or (b) when they were among acquaintances and someone laughed or made loud noises.

I should also note that when a crowd gathered in the precinct, many participants would make noise simultaneously, making difficult to tell who was actually triggering the response. This corroborates similar observations by Mueller et al. (2012) about the emergence of *overattribution* around a public interactive environment: when observing people interacting with the work, people tend to watch and copy the behavior of other users, assuming they are also affecting the interactive system even when they are actually not.

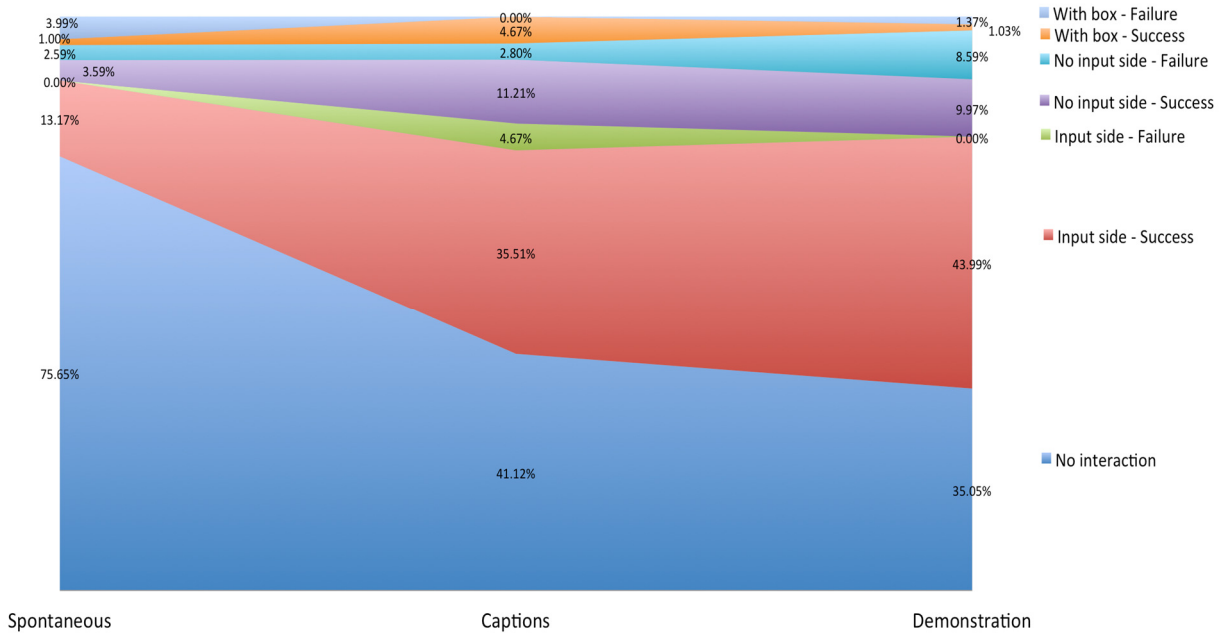

Figure 7. Percentage of interactions driven by each motivational factor on each side of the installation, broken down by outcome (no interaction, success or failure).

ANALYSIS AND DISCUSSION

Effects of Explicit Information on Interactivity

When analyzing the results, it is important to make a distinction between *visits* and *active interactions*. In the analysis below, I will refer to *visits* to the total number of people approaching the site where the installation was deployed at, regardless of whether they engaged or not in *active interaction* with the artwork. Active interactions, in turn, correspond to the number of visitors who actively attempted (successfully or not) to engage in interaction with the work. For each scenario, active interactions correspond, therefore, to a subset of the total visits registered for that scenario.

Figure 5 shows that the majority of active interactions (55.73%) were spontaneously motivated, with the rest promoted by explicit information situated in the space. Among the explicit factors promoting participation, watching other people interacting led to greater number of participants (32.37%), while only 11.90% of the interactions took place as a result of people consulting the captions. However, when considering the total number of visits and looking at those which became active interactions (Figure 7), it is clear that, proportionally, both captions and demonstrations resulted in much higher success rate than spontaneous visits, despite the latter being more numerous. Only under 25% (122/501 visitors) of spontaneous visits translated into active interactions, with less than 18% of visits (89/501) becoming successful interactions. Among those who consulted the captions, the rate of active interactions reached about 59% (63/107 visits), most of them successful (more than 51% of total visits, or 55/107). Yet, even more effective seemed to be watching others to perform: when that happened, 65% of visits (189/291) got converted into active interactions, with a success rate of 55% (160/291).

Figure 8. Full-body interaction game installation at shopping mall in Sydney, Australia.

Those observations point to two interesting features of audio as sole input mechanism for interaction with wider responsive environments: (1) it is not intuitive; but (2) it is engaging and easily learned. As pointed out by earlier research in public displays (Mueller et al., 2012; Mueller et al., 2009) people generally do not expect public displays or projections to be interactive. For that reason, recent implementations in public spaces have generally relied upon more familiar input mechanisms, such as touch-based screens, gestures or full body interaction (like the public game implementation displayed in Figure 8). Employing novel input strategies would, therefore, require prior explanation. When passers-by did check captions or watch others performing, they not only seemed more likely to engage in interaction but also quickly learned how to do it successfully: 44.27% of the active interactions (Figure 5) – or about 28% (252/900) of the total number of visits – occurred after people either checked the captions or watched demonstrations. Thus, according to my observations, presenting passers-by with explicit instructions about how to interact – either written

or (and especially) demonstrated – might be interpreted as an effective strategy for increasing participation rates, which could corroborate the findings by Mueller et al. (2012) and Bedwell and Caruana (2012).

More likely, however, the results highlight the unfamiliarity of the proposed interactive mechanisms. Interacting with projections through laughter is, admittedly, an unorthodox scenario, particularly in an urban public space. However, the proposed audio-based interaction was also supported by an interface less intuitive than anticipated. As Figure 7 shows, succeeding in getting a response from the installation when interacting at the no input side was a very unlikely outcome for people exploring the interface spontaneously: only 3.59% of those visits in that side of the space resulted in successful interactions. However, visitors exposed to explicit information about the interactivity were much more likely to get a successful interaction even when attempting it away from the microphone: 9.97% of the visitors who watched demonstrations managed to successfully interact at the no input side, a number even higher (11.21%) among those who checked the captions. Yet, occasional failure to interact near the microphone only happened when people consulted the captions (4.67% visits for that scenario). Although the captions made no mention to where the microphone was located (in line with our goal of probing intuitiveness), they did specifically instruct visitors to interact by laughing out loud. It seems, therefore, that once people read the instructions, they would get biased towards laughing near where the captions were placed, which at times might not have been loud enough. Those results indicate that both the existence and placement of written captions influenced the behavior of passers-by when interpreting an interactive installation that seemingly lacked familiar affordances.

Influence of Spatial Layout

As Figure 6 indicates, the majority of visitors who interacted did so on the input side (26.36% of total visits, or 63.36% of active interactions), despite the complete absence of signage indicating the microphone placement. This was likely due to the fact this section of the precinct was directly in front of the passageway giving access to the wharf zone and, consequently, to the spot where both the microphone and the captions had been installed. Worth of notice is that such a percentage does not refer to successful interactions only, although of course most of those near the microphone were successful. This was the preferred side regardless of motivational factors: 43.99% of visits for demonstrations, 40.18% for the light box and 13.17% for spontaneous visits (Figure 7). In terms of active interactions (i.e. the subsets of visitors who actively engaged in interaction with the installation), the rates rose to 67.73%, 62.45% and 54.10%, respectively. These data suggest, therefore, that the participation rate can be increased significantly by positioning the microphone in line with the direction the crowd naturally moves through the space.

Figure 7 also shows that a large proportion of the visits following demonstrations (18.56%) took place on the no input side. That can be explained by the dynamics of demonstrations: people arriving when participants were already interacting, usually found them standing at the side closer to the microphone, and thus were naturally 'pushed' towards to the other side of the precinct. That could be verified even when the festival volunteers were those performing the demonstrations, given that they were aware of the microphone's position: late arrivals would naturally gravitate towards the no input side. Another factor contributing to this natural movement is the 'landing zone' prompted by the audience funnel (Michelis and Mueller, 2011), a well-known behavioral pattern observed in front of public displays. Under those circumstances, passers-by coming in contact with a large display (or with other performers) may keep walking while trying to make sense of it until eventually stopping and trying to interact. In my target spatial layout, people would mostly come from the passageway towards the captions and microphone, and then away from it. The 'landing zone' pattern can thus help to explain the occurrence of attempted interactions in the no input side.

Lack of Visual Cues

Figure 6 illustrates that the majority of visits (58.40%) did not result in interaction. As Mueller et al. (2012) pointed out, people generally do not expect public environments to be interactive. Despite that being certainly less true in the context of a public art festival, interactive displays or projections are still not common scenarios. The large number of visits not leading to further interaction can be partially explained by the lack of visual cues about the interactivity. Since metaphors constructed from embodied schemata can significantly increase the occurrence of non-conscious interaction (Antle et al., 2009), I attempted to address that concern by designing a direct mapping between volume and duration of noise produced by participants and the corresponding magnitude of the response from the installation. However, as Antle et al. (2009) demonstrated, embodied metaphors should be easily discoverable to be perceived as intuitive. I found relying on audio alone for that purpose to be a challenge: in the absence of visual feedback to their audio input, participants interacting spontaneously had to resort to trial and error to be sure the responses they were getting were really triggered by them and not just random reactions. Despite somewhat playful, this solution might have pose issues should the interface have been designed for a more utilitarian purpose. In that case, I would recommend the use of visual feedback to highlight the embodied metaphor suggested by the auditory mapping.

Impact of False Affordances

As indicated by Figure 7, both in the case of spontaneous interaction and consultation to the captions I could observe just under 5% of the visits resulting in misguided attempts to use the equipment box as interface (25/501 for spontaneous interactions, 5/107 for captions). However, while most people who spontaneously explored the interface failed to get a response (20/25, or 80%), all (5/5) who read the captions succeeded. That indicates the latter, having learnt that laughing out loud would trigger a response from the work, behaved accordingly by producing a noise loud enough to be captured by the

microphone installed just a few meters from the equipment box (notwithstanding their belief that it was actually installed within the box). Those exploring the interface spontaneously would mostly try to touch or press parts of the equipment box in an attempt to trigger responses. In particular, an air vent at the top of the box (Figure 3) – arguably resembling a large push button – would prompt many visitors to try exploring it through repetitive (and at times even aggressive) touch.

In hindsight, and taking the example from Gaver et al. (2009), I can partially attribute my failure to anticipate the impact of those false affordances to having largely assumed, in this study, a stance of *design for research* as opposed to *design as research* (Gaver et al., 2009). My design activities were pursued primarily in service of a theoretical concern – namely, creating a minimalist interface that could allow the audio and projections to stand out, thus (in theory) attracting passers-by and articulating the interactive experience. Yet, I failed to appreciate the cognitive mechanisms people normally employ when attempting to make sense of interactive interfaces. As Blackler and Hurtienne (2007) described, technological familiarity is a strong factor determining the intuitiveness of an interface: as humans, when trying to make sense of new technological settings, we seek potential clues in experiences we had with previous interfaces. In my installation setting, the equipment box was the only visible element on the deck close enough to the projections to be perceived as an affordance (Norman, 2004). Until people became aware of the audio interactive nature of the work, it would make good sense to regard the equipment box as some sort of console or kiosk.

That is especially true when one considers recent similar implementations using tangible user interfaces as mediators between media façades and the general public (Behrens et al., 2014; Fischer and Hornecker, 2012). As pointed out by my own previous research (Hespanhol and Tomitsch, 2015), when faced with an interactive public environment, people have become familiar with receiving direct feedback to their actions, usually in the form of mirrored images that can convey agency and identity; in the absence of visual feedback matching their individual input, the environment may be intuitively perceived either as ambient (i.e. little responsive or not interactive at all) or triggered by remote devices. The somewhat imposing presence of the equipment box right in front of the projections may explain why many people therefore interpreted it as such a remote device – something I had not considered as a potential issue until actually testing the interface with members of the general public, *in situ*.

LIMITATIONS

I acknowledge that the case study presented in this paper may be slightly unorthodox as an interactive public environment, requiring participants to stand on an observation deck and look at it from above. Yet, having been invited to deploy a projection-based installation in a high profile public art festival, I approached it as an opportunity for carrying out research in and through design (Dalsgaard, 2010). Field studies in the wild are notoriously filled with challenges (Memarovic et al., 2013; Ojala et al., 2011), and I acknowledge the limitations inherent to this type of study, particularly when deployed within a context that may potentially bias people towards a particular set of social norms, as may be the case with art festivals.

The strict setting requested by the festival environment also limited the possibilities to introduce more comprehensive test cases or vary controlled variables – such as placement of input elements, the required decibel threshold, or the types of feedback – beyond the scenarios presented in the paper. That would have been desirable, since it would have allowed results beyond qualitative observations and with greater statistical significance. Likewise, interviewing participants might have easily compromised ecological validity, given that the object of the study was the perception of interactivity. Yet, running the study as part of a special event also brings advantages: the fact that the installation was attached to a special event and its existing social infrastructure allows the observation of interactions without setting up signifiers, attractors or active involvement of the researcher (Behrens et al., 2013).

CONCLUSIONS

This paper presented a field study in noticing and making sense of audio-based interactivity of public projections. It describes observations from an interactive art installation deployed to a prominent urban festival, using laughter and noise as main input for the interaction. From the results obtained, I derived a number of insights, notably: (a) presenting passers-by with clear instructions about how to interact is an effective strategy for increasing participation with an unfamiliar platform, but care must be taken in order to avoid unnecessary bias regarding the type of audio accepted; (b) situated demonstrations (including watching others performing) are particularly persuasive; (c) as people will generally attempt to start interacting around the spots where they can firstly see the projections, placing microphones on those spots increases the chances of successful interaction; (d) feedback about the audio interactivity should be complemented by visual cues guiding participants towards the “sweet spots” for interaction, i.e. the locations where the microphones are located; (e) in the absence of visual feedback unambiguously related to the interaction, other visual elements in the immediately surrounding area become perceived affordances, even if just part of the local infrastructure; extra care should thus be taken regarding their placement within the context of the interactive environment. Despite the limitations outlined above, I am nonetheless confident that those are largely outweighed by the insights gained, and hopeful this study can be considered as an important step forward in getting a better understanding of multi-modal mechanisms of interaction making predominant use of audio for input.

ACKNOWLEDGMENTS

LOL was developed with support from Vivid Sydney, Destination NSW and the Design Lab – Faculty of Architecture, Design and Planning, University of Sydney, Australia.

REFERENCES

Antle, A. N., Corness, G., & Droumeva, M. (2009). Human–-computer intuition? Exploring the cognitive basis for intuition in embodied interaction. Int. J. Arts and Technology, 2(3), 235-254.

Arroyo, E., Bonanni, L., & Valkanova, N. (2012). Embedded interaction in a Water Fountain for Motivating Behavior Change in Public Space. In Proc. CHI'12, Austin, Texas, USA.

Bedwell, B., & Caruana, T. (2012). Encouraging Spectacle to Create Self-Sustaining Interactions at Public Displays. In Proc. PerDis'12, Porto, Portugal.

Behrens, M., Fatah gen. Schieck, A., Kostopoulou, E., North, S., Motta, W., Ye, L., & Schnädelbach, H. (2013). Exploring the Effect of Spatial Layout on Mediated Urban Interactions. In Proc. PerDis'13, MountainView, California, USA.

Behrens, M., Valkanova, N., Fatah gen Schieck, A., & Brumby, D. P. (2014, June 03-04). Smart Citizen Sentiment Dashboard: A Case Study Into Media Architectural Interfaces. In Proc. PerDis'14, Copenhagen, Denmark.

Blackler, A., & Hurtienne, J. (2007). Towards a unified view of intuitive interaction: definitions, models and tools across the world. MMI-Interaktiv, 13, 37-55.

Dalsgaard, P. (2010). Research in and through design: an interaction design research approach. In Proc. OzCHI'10, Brisbane, Australia.

Dalton, N., Marshall, P., & Dalton, A. (2013). Extending architectural theories of space syntax to understand the effect of environment on the salience of situated displays. In Proc. PerDis'13, Mountain View, CA, USA.

Fatah gen. Schieck, A., Al-Sayed, K., Kostopoulou, E., Behrens, M., & Motta, W. (2013). Networked architectural interfaces: exploring the effect of spatial configuration on urban screen placement. In Proc. 9th International Space Syntax Symposium, Seoul, Korea.

Fischer, P. T., & Hornecker, E. (2012). Urban HCI: Spatial Aspects in the Design of Shared Encounters for Media Façades. In Proc. CHI 2012, Austin, TX, USA.

Gaver, W., Bowers, J., Kerridge, T., Boucher, A., & Jarvis, N. (2009). Anatomy of a Failure: How We Knew When Our Design Went Wrong, and What We Learned From It. In Proc. CHI'09, Boston, Massachusetts, USA.

Hespanhol, L., Bown, O., Cao, J., & Tomitsch, M. (2013). Evaluating the Effectiveness of Audio-Visual Cues in Immersive User Interfaces. In Proc. OzCHI'13, Adelaide, SA, Australia.

Hespanhol, L., & Tomitsch, M. (2015). Strategies for Intuitive Interaction in Public Urban Spaces. Interacting with Computers. doi: 10.1093/iwc/iwu051

Hespanhol, L., Tomitsch, M., Grace, K., Collins, A., & Kay, J. (2012). Investigating intuitiveness and effectiveness of gestures for free spatial interaction with large displays. In Proc. PerDis'12, Porto, Portugal.

Jarusriboonchai, P., Olsson, T., & Väänänen-Vainio-Mattila, K. (2014). User experience of proactive audio-based social devices: a wizard-of-oz study. In Proc. MUM'14, Melbourne, Australia.

Kray, C., Galani, A., & Rohs, M. (2008). Facilitating Opportunistic Interaction with Ambient Displays. In Proc. CHI'08, Florence, Italy.

Kukka, H., Goncalves, J., Wang, K., Puolamaa, T., Louis, J., Mazouzi, M., & Barco, L. R. (2016). Utilizing Audio Cues to Raise Awareness and Entice Interaction on Public Displays. In Proc. DIS'16, Brisbane, Australia.

Memarovic, N., Langheinrich, M., Alt, F., Elhart, I., Hosio, S., & Rubegni, E. (2012). Using Public Displays to Stimulate Passive Engagement, Active Engagement, and Discovery in Public Spaces. In Proc. MAB'12, Aarhus, Denmark.

Memarovic, N., Langheinrich, M., Cheverst, K., Taylor, N., & Alt, F. (2013). P-LAYERS - A Layered Framework Addressing the Multifaceted Issues Facing Community-Supporting Public Display Deployments. ACM Trans. Comput.-Hum. Interact., 20(3), Article 17.

Memarovic, N., Clinch, S., & Alt, F. (2015). Understanding display blindness in future display deployments. In Proc. PerDis'15, Saarbrücken, Germany.

Michelis, D., & Mueller, J. (2011). The Audience Funnel: Observations of Gesture Based Interaction with Multiple Large Displays in a City Center. International Journal of Human-Computer Interaction, 27(6).

Mueller, J., Walter, R., Bailly, G., Nischt, M., & Alt, F. (2012). Looking glass: A field study on noticing interactivity of a shop window. In Proc. CHI'12, Austin, TX, USA.

Mueller, J., Wilmsmann, D., Exeler, J., Buzeck, M., Schmidt, A., Jay, T., & Krueger, A. (2009). Display blindness: The effect of expectations on attention towards digital signage. In Proc. Pervasive'09, Galveston, TX, USA.

Norman, D. (2004). Affordances and design. Retrieved 25/08/2014, from http://www.jnd.org/dn.mss/affordances_and_desi.html

O'Hara, K., Glancy, M., & Robertshaw, S. (2008). Understanding collective play in an urban screen game. In Proc. CSCW'08, San Diego, CA, USA.

Ojala, T., Kostakos, V., & Kukka, H. (2011). It's a jungle out there: Fantasy and reality of evaluating public displays "in the wild". In Proc. Workshop on Large Displays in Urban Life - from Exhibition Halls to Media Façades. Held in conjunction with CHI'11, Vancouver, BC, Canada.

Schroeter, R., Foth, M., & Satchell, C. (2012). People, Content, Location: Sweet Spotting Urban Screens for Situated Engagement. In Proc. DIS'12, Newcastle, UK.

Steinberger, F., Foth, M., & Alt, F. (2014, June 03-04, 2014). Vote With Your Feet: Local Community Polling

on Urban Screens. In Proc. PerDis'14, Copenhagen, Denmark.

Walter, R., Bailly, G., & Mueller, J. (2013). StrikeAPose: Revealing Mid-Air Gestures on Public Displays. In Proc. CHI'13, Paris, France.