Real World Responses to Interactive Gesture Based Public Displays

John Hardy
Lancaster University
Lancaster, UK
john@highwire-dtc.com

Enrico Rukzio
University of Duisburg-Essen
Essen, Germany
enrico.rukzio@uni-due.de

Nigel Davies
Lancaster University
Lancaster, UK
nigel@comp.lancs.ac.uk

ABSTRACT
Today, one does not have to travel far to find examples of digital signage, yet the adoption of interactive gesture based public displays remains quite rare. Subsequently, not much is known about them despite a large array of potential advantages. This paper contributes to our understanding of how people perceive, respond to and interact with such displays outside the controlled environment of a research lab. Unlike other works which have focused on isolated aspects of in-lab interaction, we present a detailed examination that addresses a wide range of responses to such a display - including those who ignore them completely. To facilitate our study we created an experimental coarse gesture based software suite and then deployed the system along with associated applications as part of an existing large scale public display network. Using this as a base, we executed four studies designed to passively observe the reactions of passers-by and followed these up with a fifth controlled experiment that compared the effectiveness of two different kinds of gesture in the context of menu item selection. To conclude, we present our key-findings and highlight possible avenues of further study for the future of gesture based digital signage.

Categories and Subject Descriptors
H.5.2 [**User Interfaces**]: Input Devices and Strategies (e.g, mouse, touchscreen)

General Terms
Design, Experimentation, Human Factors.

Keywords
Public display, gestures, interaction, experimentation, hand control, digital signage, deployments, attention, engagement.

1. INTRODUCTION
The falling cost of display hardware and increasing availability of network access is leading to the widespread deployment of digital public display systems. While these systems are primarily used as channels for delivering advertising material and providing live information such as train times, there is an intense interest in the development of novel usage models and new forms of interaction. In particular, a wide range of techniques for making these systems interactive (e.g. adding touch overlays and facilitating control via personal mobile devices) have been explored by researchers keen to harness them for a broad range of applications.

Despite significant advances in gesture recognition [1] and digital signage [14], it is rare to find scientific studies that address the combination of these systems within the context of an existing real world public display installation. Instead, most works have studied in-lab or overt deployments. Where real world studies of public displays have been undertaken [17] [5] [12] they typically focus on improving or facilitating an aspect of human behaviour or social interaction; placing a display conspicuously or in a position of prominent focus such that it cannot be avoided. This can limit the extent to which findings generalise to situations where the display is not an item of novelty. In other cases, researchers have applied obtrusive technology such as large IR tracking systems [24] [4] which has the side effect of predisposing participants to the knowledge they are being observed. Subsequently, this does not necessarily represent the complete picture of how people would respond to the same display if deployed without the technological constraints.

Figure 1. A photograph of a participant using the space game application at the deployment.

Our work aims to address these issues by contributing to the understanding of how users interact with real world deployments of gesture based public displays. Specifically, this is addressed through observing and analysing reactions to gesture based digital signage in terms of attention level, display content and social setting as well as comparing the effectiveness of different gestures in terms of task completion time and user experience.

Our experiments were supported by a software system that is able to detect a variety of simple gestures such as arm movements, spatial buttons and directional waving that requires no special markers, prior-knowledge or training. This creates a foundation for a simple walk-up-and-use model that allows people to behave freely around the system; helping to eliminate bias introduced by goal orientated scenarios and artificially inflated preconceptions about the capabilities of the display induced by controlled environments.

2. SYSTEM OVERVIEW

To facilitate our studies we built a gesture-controlled digital public display suitable for deployment in the real world (Figure 1). The display system, overviewed in Figure 2, is built using a commodity webcam and a 40" LCD public display. The webcam can be mounted to the top, bottom or front of the screen. Our software is responsible for managing content on the display, detecting gestures and then mapping them onto functions of a user interface.

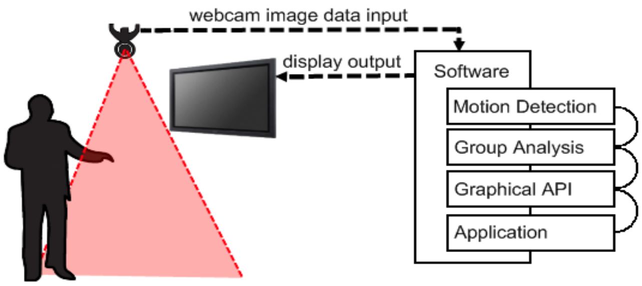

Figure 2. Conceptual system overview.

As the system is subject to real world evaluation, it was important that we took care in the presentation of both the content and the hardware in order to emulate the appearance of a professional interactive content display and not a prototype. To avoid complicating the initial stages of interaction we chose to implement gesture detection algorithms that are able to operate effectively without requiring a calibration stage for each user.

The kind of gestures the system detects are coarse[1] and do not require direct contact with the display. A benefit of coarse gestures is that they require little processing and resolution to detect. This coarseness also neither requires nor prohibits the use of props and can operate reliably in situations where users are wearing gloves, flamboyant clothes or other features known to require recalibration in similar systems. In terms of gesture detection, we consider misinterpreted gestures a normal part of operation that need to be accounted for in both system design and evaluation.

The system trades the accuracy offered by complex computer vision algorithms for more lightweight motion detection. The majority of the software, including visuals and interaction logic is written in Python. Components with performance constraints such as webcam capture and image processing are written in C++. In terms of processing, low resolution frames are at a rate of 15fps using the video capture feature of OpenCV. Each frame is forwarded to bespoke motion detection code which performs per-pixel operations on the data in order to determine groups of adjacent motion. From here, bespoke high-level group and feature analysis code offers different ways that the motion can be interpreted and handled by an application delivering content.

[1] We define coarse gestures as large motions that can be detected from a distance such as arm movements rather than small hand or finger movements.

3. RELATED WORK

Perhaps the best known commercial gesture detection system is Microsoft's Kinect [1] which has since been applied primarily in the gaming domain[2]. In the public display domain, Apple and Orange were amongst the early adopters when they deployed interactive adverts attracting much press attention [2][3]. Since then interactive floor, wall and table-top displays have become more common and there are several businesses now specialising in this area.

Previous work in the public display domain has focused on understanding and supporting a sense of community [23] [6] or facilitating interaction and collaboration between groups of people [13] [19]. This style of study recognises the relative advantages that different interactive tools offer in different contexts. Rogers and Lindley [16] provide an interesting critique of this space where they focus on examining the nature of the physical-social relationship underpinning how situating displays in different places invites different kinds of social interactions. They illustrate this by describing differences in the ways groups coordinate their work when using an interactive table-based display versus a wall-based display.

In 2010, Rubegni et al. deployed a gesture based display (USIAlumni Faces) at a university alumni event to help "*break the ice*" amongst guests [17]. Participants would interact with a large board using a 'toy torch'. They found that the outward nature of gesture based interfaces assisted with the diffusion of interaction patterns in public spaces through an observe-and-learn model. Adding that: "*... sensory-motor patterns can aid social interaction in public, as they act as conversation starters between both strangers and acquaintances.*"

A small proportion of interactive public display research addresses the attention level people pay interactive public displays when they are not placed in prominent focus and simply designed to convey information. To this end, our work follows that of Huang et al. [8] and Müller et al. [15].

Huang et al. [8] address this gap through a comprehensive qualitative review of the traits exhibited by people around real world public displays. They observed that many such systems are built on the idea that passers-by will engage with interactive displays after first being drawn to them as ambient displays. In practice, these displays typically have a low rate of draw. They found that placing the display at eye-height and with dynamic content (such as video) was far more effective at attracting glances than static content or displays placed high up or low to the ground. Additionally, while the presence of something else to first catch someone's attention increased the number of glances, having a captive audience (such as people standing in a queue at a shop or sitting in a foyer) did not promote more or longer glances.

Müller et al. [15] build on the importance of these findings to add that even though knowledge of low attention levels is well established, explanations for this behaviour are still needed. To this end, they define 'Display Blindness' and explain it by saying: "*…expectations of uninteresting content leads to a tendency to ignore displays.*" They propose possible solutions to help focus audience attention towards public displays. We observed that 'Display Blindness' still applied to our gesture based interactive display until people became aware of its purpose. Without a

[2] This study was carried out in 2009, before the public deployment of Microsoft's Kinect.

longer term survey it is difficult to speculate if this effect would begin to diminish with novelty.

Huang et al. [8] also highlight that research systems are often designed partly to attract people but the evaluations focus largely on the reactions of those who interacted and not those who did not. Subsequently, the importance of observing each response (including those that completely ignored the display) played a key role in shaping this study.

There are several frameworks that can be applied to the interactive public display domain. Vogel and Balakrishnan [24] classified the role of a user's location with respect to the display. They describe 'four phases of interaction': *personal*, *subtle*, *implicit* and *ambient display*. This paper looks at the transition from *ambient display* through the *implicit*, to *subtle interaction*.

Tang et al. [22] classify 3 kinds of bystanders: those that are *passer-by*s with a purpose – these pay minimal attention to the display, *stander-by*s – people who are in the environment but not primarily to interact and lastly, *engaged bystander*s – people with a focused awareness of the display and are 'staring' or making use of the content. This is complemented by the work of Skog et al. [20], who offer 3 levels of comprehension that are used to assess a person's understanding of a public information visualisation. These are: *that something is visualised*, *what exactly is being visualised* and an *ability to interpret the visualised data*.

Brignull and Rogers [5] discuss a public display's role in encouraging people to cross the thresholds of focal-awareness and participation. They relate these thresholds to a conceptual framework for analysing public interaction. This consists of three activity spaces: *peripheral awareness* – people aware of the display but don't know much about it, *focal awareness* – people are talking about activities associated with the display and *direct interaction* – an individual or group using the display.

4. EXPERIMENT CONTEXT AND DESIGN

This section is describes the location and context of all the observational studies carried out in this paper (Section 5). We discuss the deployment location and review the rationale behind our attention level classification before summarising the data gathering techniques used.

4.1 Deployment Location

All the studies were carried out in a busy foyer at Lancaster University in the UK using a single piece of pre-existing display hardware [21]. This location is an important route for people travelling through the university, with people passing through at a rate of approximately three people per minute during working hours of term time.

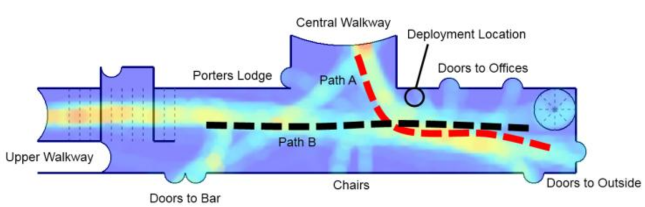

Figure 3. A floor map showing the deployment location with traffic intensity represented as a heatmap. Lines representing paths A and B indicate the main routes passing the display.

A floor plan of the foyer is illustrated in Figure 3. This is overlaid with the most common routes taken through the foyer. Of these routes, 2 pass the display:

A) Between the central campus walkway and doors leading outside – the red line.
B) Between the stairs or college bar to the doors leading outside – the black line.

In order to establish a baseline for the rate of flow, a preliminary survey collected data on the paths people take through the foyer and the rate at which people moved by the display. Two discrete markers were placed on the floor, 2.2 meters in front of the display, 4.3 meters apart. Over a working-hours period of 15 minutes, 43 people passed the display at an approximate speed of 6.1 km/h.

A Pearson's Chi-Square test over all the data collected for this paper showed that the attention level given to the display did not significantly differ with the path taken through the building, $\chi^2(2, N=615)=4.26, p > .05$.

4.2 Attention Level Classification

We considered several methods for ascertaining the amount of attention a person gives a display, such as observation based ethnographic methods and automated solutions like gaze tracking. Given the high flow rate it would have been difficult if not impossible to accost each person who passed by without interfering with the experiment or limiting the demographic to people who have the time to answer questions.

When observing people around public displays, Huang et al. [8] use a technique called micro-shadowing to note people nearby and record if they glance at the display or not. They point out the anecdotal nature of this technique means it should be used as a sampling method rather than an absolute measure. This technique is used as a basis for the observational studies undertaken for this work. In another study, Huang and Mynatt [9] note that: "*opportunistic glancing has prompted* [users of a public display] *to offer assistance and information to group members who posted information*". Despite our focus on high-traffic areas where people do not necessarily share a common goal (and are thus less likely to interact in such a manner), glancing remains an individual reaction and, as such, a useful measure of attention.

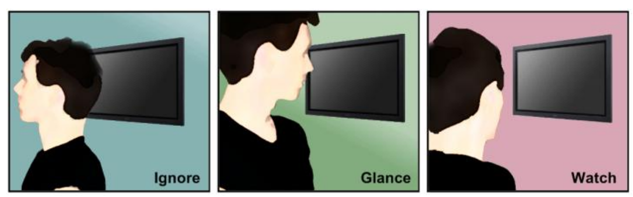

Figure 4. Levels of attention that people can pay to a display.

Subsequently, we formalised a ranking system (Figure 4) that distinguishes between three levels of attention. This classification is based on the work of Tang et al. [22], the observation techniques used by Huang et al. [8] and Rukzio [18]:

- **Ignored**: A person completely ignores the display as they pass. They do not alter their behaviour in response to the display.
- **Glanced**: These are people who, in some noticeable way, react to the display. Usually by turning their head or looking towards it. Glances last less than 3 seconds.
- **Watched**: These are people who alter their behaviour significantly in response to the display. Perhaps by stopping to look or turning round after they pass to get another look.

4.3 Data Gathering

In order to be able to quickly and accurately gather rich observational data, as people passed through the foyer a researcher used a bespoke data gathering application to take notes, record frequencies and index the events that unfolded. These records were supplemented with verbose notes taken by hand. Each observation session lasted 1 hour and all were carried out during university term time.

The dependant measures made for each person that passed the display were: the time of passing, the path taken through the building, their attention level regarding the display, gender, an approximation of age (younger or older than 25) and if they triggered a response from the display. Where people interacted with the display, interaction errors, corrections and common traits (such as any usability problems and gesture styles) were directly flagged and associated with verbose notes.

5. UNDERSTANDING ENGAGEMENT

In this section we describe four experiments designed to elicit understanding about how people engage with gesture based public displays in the real world. The first experiment reports on the general reactions of people in the vicinity of the display. The second compares the impact of content type and the third does the same for social setting. The final experiment describes an invited interaction session where people passing-by were told they could interact with the display.

5.1 Exploring User Attention

Our first deployment took place between 2pm and 3pm to sample behaviour during typical working hours. It was an exploratory study designed to observe the frequency and nature of people's responses to the interactive display. The content on the screen provided passers-by with an interactive photo-gallery that allowed them to cycle and view a selection of pictures.

When a person approached the display they triggered a transition from the title screen to a photo selection mode. By waving in a sweeping motion from left-to-right or right-to-left, the application would cycle a carousel of photographs in the direction the person waved. If the person left the display and no movement was detected for 10 seconds, it would then transition back to the title screen.

During the deployment, a total of 184 people passed the display. Of these people, 55% were female and, as one might expect of a university, 80% appeared younger than 25. The large majority, 84%, ignored the display, while 15% glanced at the display. This left only 1% stopping to watch it. These figures are consistent with those observed by Huang et al. [8].

Table 1. Cross tabulation of attention level and path taken.

Path Taken	Attention Level: Ignore	Glance	Watch	Total
A	86 (82%)	18 (17%)	1 (1%)	105
B	68 (86%)	10 (13%)	1 (2%)	79
Total	154 (84%)	28 (15%)	2 (1%)	184

Two Pearson's Chi-Square tests showed that a person's attention level did not significantly differ with gender or age: $\chi^2(2, N=184)=3.54$, $p > .05$ and $\chi^2(2, N=184)=2.12$, $p > .05$ respectively.

Of the whole, 34% passers-by were close enough to trigger a transition between the title screen and photo-carousel. We found a significant association between whether a person triggered the display and their attention level: $\chi^2(2, N=184)=2.12$, $p < .05$. However, it is worth noting that this association may be due to the person passing closer to the display and thus giving it greater consideration than as purely a response to the display itself.

In some cases, the behaviour of those who ignored the display was just as valuable as those who watched or glanced. We observed that as people passed by, many were talking on the phone or wearing headphones - presumably listening to music. This would impact the effectiveness of audio-based interactive digital signage. However, the ability to detect such features may be a valuable input that helps context sensitive displays ignorantly imposing themselves on busy people. We also observed a tendency of passers-by to watch or glance at people sat on chairs and that it was very common for people to walk with their head facing the floor. Interestingly, Huang et al. [8] report that displays mounted low to the ground received almost no attention.

5.2 Examination of Content Types

Our second deployment compared two different gesture-based application types: a weather information display (Figure 5) and an interactive asteroids game (Figure 6). The study took place between 2pm and 3pm during working hours.

We hypothesised that the eye-catching content of the game (lots of motion with coloured animated particles set against a contrasting background) would attract more attention than the static weather information display.

Figure 5. The interactive weather information point.

Figure 6. The interactive asteroids game.

The weather information display consisted of a ring menu with icons representing the weather for the next five days in addition to a full textual summary of the foremost item. The user interacted with the system by moving their arm to the far left or right in line with the display screen. This would then rotate the ring menu in the direction of the hand in order to display the weather for another day. Instructions were always available on-screen at the bottom.

The game consisted of a scene of asteroids, hoops and a spaceship with particle streams that emit from the engines. When the display did not detect a person stood in front of it, a large pulsing "Approach To Play" message was displayed (Figure 6). When a person approached, this faded out and the ship began to accelerate. Like the weather display, textual instructions were always available on-screen. The objective of the game was to dodge oncoming asteroids whilst flying a ship through as many hoops as possible. Users were able to strafe the ship by smoothly waving their arm from left to right in line with the display screen.

The speed and distance of the hand movements was directly reflected in the distance and rate at which the ship moved.

Over both hour-long observation periods, a total of 252 people passed the display (Table 2). During this deployment, nobody interacted with the weather information point and a single person interacted with the game. Additionally, 1 person stopped to watch the interactive weather display in contrast to 12 who stopped to watch the game. This suggests that the game attracted significantly more attention, and indeed, $\chi^2(2, N=252)=11.65$, $p < .05$ offers statistical support.

Table 2. Cross tabulation of deployment and attention level.

Content	**Attention Level**			
	Ignore	Glance	Watch	Total
Weather	113(86%)	17(13%)	1(1%)	131
Game	90(74%)	19(16%)	12(10%)	121
Total	203(80%)	36(15%)	13(5%)	252

These results question the assumption that glancing acts as a prelude to turning to watch the display – suggesting the decision to transition between them is made consciously. Given that conscious decision making takes longer than a natural reaction, a person's decision to watch the display would not always be made before they had finished walking past. If the first assumption were true we would have expected to observe the game attracting a proportionate increase in the number of people glancing due to the eye-catching nature of the content. However, this was not the case. People seemed equally likely to glance at the display regardless of the content (we observed only a 3% increase in the number of glances received by the game). This lack of a significant increase suggests that the decision to watch the display is made around the same time as the decision to glance. While it is impossible to say for certain without further study, it may be that natural reactions dominate the decision to devote attention to a display such as ours, especially in cases where users are simply passing by.

The person who watched the weather application appeared to notice something move on the screen and went to back to investigate for approximately 2 seconds before carrying on her way. In contrast, the people who watched the game tended to notice the content as they approached the screen – turning their heads and slowing to look at it as they walked past.

The man who interacted with the game did so whilst talking on his mobile phone. He appeared to notice the display in his periphery as he passed it, slowed down and then turned to face it. He then lent in towards the display and extended his free arm to play for approximately thirty seconds before carrying on through the foyer. Ten seconds into gameplay he appeared to notice that when he reached further forward, the display became more responsive and thus adjusted his position to match. After doing so he successfully avoided the asteroids and collected a number of rings.

5.3 Examination of Time and Social Setting

This experiment was designed to determine if people's behaviour in different social environments at a different times of day affected the amount of attention they paid the display. We hypothesised that the display would receive more attention later in the day during social activities.

The experiment was conducted during the evening of a campus social event (8pm to 9pm) where people were passing through the foyer, to and from the bar and often in groups. At times people could be observed waiting in the foyer for friends to arrive and leaving the bar for various reasons such as to take a phone call.

As with the previous surveys, the format of this experiment followed a standard series of observations. The application in use was the asteroids space game because it previously attracted the most attention. The results were compared with those collected in the previous experiment conducted earlier in the day between 2pm and 3pm.

During the observation period, 141 people passed the display. Table 3 shows only slight differences between treatments. A Pearson's Chi-Squared test $\chi^2(2, N=262)=.70$, $p > .05$ ($p=.71$) showed no significant association between the two which implies that there is no noteworthy difference in the amount of attention people pay the display on a social evening or working hours.

Table 3. Cross tabulation of social setting and attention level.

Social Setting	**Attention Level**			
	Ignore	Glance	Watch	Total
Event	111 (79%)	18 (13%)	12 (8%)	141
Normal	90 (74%)	19 (16%)	12 (10%)	121
Total	201(77%)	37(14%)	24(9%)	262

5.4 Invited Interaction

In this experiment, individuals and groups were invited to interact with the display. Every effort was made to ensure the selection process was free of bias and the overwhelming majority of those asked, took part. This deployment was conducted informally and intended to inform the analysis with qualitative information regarding behavior and interaction styles.

During this hour long deployment, a wide variety of interaction types were observed over a total of 46 participants. Our first observation was that passers-by would quite often stop walking past to watch those interacting. At one point, this resulted in a crowd of approximately ten people. Those in groups naturally took turns to operate the system and appeared to copy each other's interaction methods for a few seconds before trying their own.

Notable interaction strategies included: reaching upwards towards the screen; holding a flat palm towards and perpendicular to the display; ducking and diving with the arm (and in one case, the entire body) to try and make the ship go up and down as well as holding out both hands and using a 'window washing' movement while standing on tip-toes. Even when given prior verbal instruction, people appeared to like to try their own techniques.

An important observation was that people did not expect the system to be interactive. A group of eight young men did not believe that the system was real. They asked the experimenter to stop taking notes on the laptop while they played, reasoning the experimenter must be controlling it.

People appeared to have more inhibitions about interacting with the display when alone compared to when they were in a group. When promoting interaction it was much easier to attract people from other social groups if they had seen another set of people previously interacting with the display. Of the 46 people who were initially invited to interact with the display, on separate occasions three groups later came back and interacted with the display without invitation. This suggests that once the users had experienced the capabilities of the display and were aware of its function they were more likely to engage with it again and were willing to tell others of their discovery.

When people used the display in groups, the focus of the interacting user's attention tended to be on the group interaction rather than the display application. People would pay more attention to on-going discussions at the expense of accurately steering the ship. This suggests that in the context of interactive public displays, the relationship between Brignull and Rogers' *direct interaction* and *focal-awareness* [5] is exaggerated, perhaps because of the large and novel nature of the physical activity.

In each case where groups of more than 3 people congregated around the display, the people who were not interacting would lose interest quicker than the person engaged by the system. In more than one instance, the group walked off and when the person playing the game noticed, they immediately stopped playing and returned to the group. From this we assert that interactive displays design for those not being directly engaged, they may be able to alleviate social pressures that call for a premature end to the interaction session. Possible strategies for this include recognising people in the vicinity of the display and inferring the nature of the relationship between them. The display would then be able to automatically suggest team structures, prescribing turn taking strategies and offer competitive schemes. However, the cost of being wrong may be prohibitively high as while we observed people were happy to watch others interact, they appeared more reluctant to join them if they had no social tie.

6. GESTURE TECHNIQUE COMPARISON

The goal of this experiment was to conduct a controlled examination of how effective two different gesture techniques were at selecting items from a ring menu on a wall mounted public display (Figure 7).

Figure 7. The experimental setup used to compare gestures.

The hypothesis was that relative kinaesthetic gestures are significantly more effective as a means of interaction than absolute location gestures. These are illustrated in Figure 8. Effectiveness is evaluated in terms of the time taken to complete a certain task and a user experience rating. The type of gesture served as our independent variable which consisted of two levels:

- **Absolute location gestures** (ALGs): a directed pulling or pushing motion (with either arm) in a certain area triggers the menu to move.
- **Relative kinaesthetic gestures** (RKGs): a sweeping motion with the arm (in either direction) allows a user to build up and reduce momentum in the menus movement.

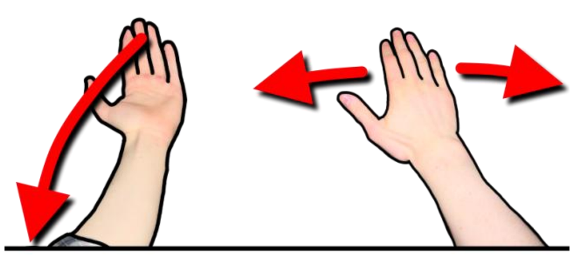

Figure 8. Illustration of ALGs: a pulling or pushing gesture (left) and RKGs: a graduated waving motion (right).

The experiment involved 70 participants that were randomly sampled from the people passing through the foyer (35 for each interaction technique). The subjects were people who were walking past the display and agreed to participate. None of them had any prior experience using the system or previous knowledge of its existence.

The goal of the task was for a participant to select a specific item from a ring menu when prompted. Using the intended gesture, users could cycle though 6 menu items until the target item (a number displayed in the lower left corner of the screen) matched the foremost menu item. Once a target was selected, it turned green and the target item would cycle. The change of selection target was emphasised by a brief fade transition and animation of the instructional text.

The sequence of numbers the participants were asked to select followed a repeating pattern, ensuring that everyone was exposed to the same repeating sequence of numbers shown in Table 4. The sequence was specifically selected in order to offer participants the opportunity to use shortcuts by reversing the direction of the menu. In Table 4, complexity is a measure of the distance between a target menu item and the preceding target in the repeating sequence.

Table 4. Showing the (repeating) sequence of numbers and their respective indices and differences.

Index	0	1	2	3	4	5
Number	1	4	3	6	2	4
Complexity	3	3	1	3	3	2

The dependant variables measured were task competition time and a user experience ranking:

- Task completion time is defined as the time taken for a user to select a specific item on the menu. This statistic was recorded automatically for each participant by the display application.
- User experience is measured as an interval ranking which asked participants to rate usability between 1 (really easy) and 10 (really hard).

Counterbalance was provided through the use of a between-subjects design; using a different set of participants for each treatment to maintain the real world setting and prevent learning effects being carried between interaction modes. The interaction mode was swapped every ten minutes. This was intended to minimise any artefacts of the typical university pedestrian timetable. On average it took each participant approximately five minutes to complete the entire process.

Participants were approached individually to avoid creating a sense of competition or other social pressures that might skew results. Furthermore, participants received no feedback from either the researcher or display regarding their performance relative to the other participants.

Once a participant agreed to take part, the researcher explained the task: "*Using this gesture* [act out gesture], *you can rotate the menu. The idea is to bring the number in the bottom left to the front of the menu. When you've done this, it will turn green and the bottom number will change. Do this a few times and I'll tell you when to stop.*" After completing a full sequence of number selections they were asked to rate the usability of the system, were encouraged to give feedback and ask any questions.

6.1.1 Quantitative Analysis

The usability ratings are shown as side-by-side histograms in Figure 9. These results indicate that users perceive the ALG based interface (M=2.97, SD=1.52) as easier to use than the RKG based interface (M=3.86, SD=1.57). An independent samples t-test, $t(68)=-2.39$, $p < .05$, shows that the means are significantly different.

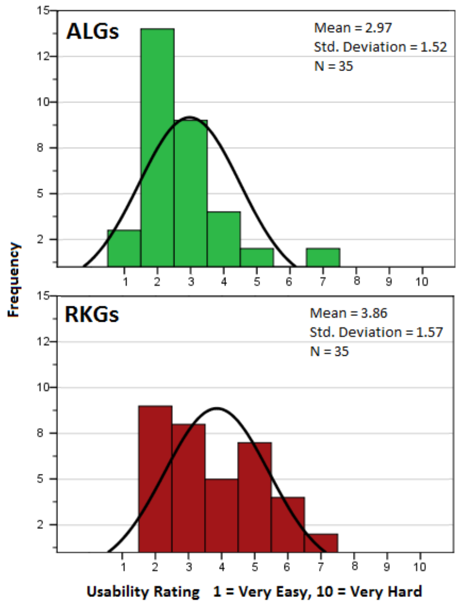

Figure 9. Aligned histograms that compare usability ratings for ALGs (top) and RKGs (bottom).

In terms of task completion time, while Figure 10 shows that on average participants completed the task approximately 2 seconds faster using ALGs (M=6104ms, SD=2571ms) when compared to RKGs (M=8057ms, SD=3485ms). However, these results are also subject to high variance relative to the size of the measure.

In order to determine if the gesture type (ALG or RKG) had a significant impact on task completion time we computed a series of linear models. As one would expect, the first model showed task completion time had a significant association with task complexity ($p<0.01$, $R^2 \approx 0.02$). Using a step-wise model selection scheme, we then added gesture type as an indicator variable to see if it explained any more patterns in the residuals. Controlling for task complexity, this reported a significant association between task completion time and gesture type ($p<0.001$, $R^2 \approx 0.12$) allowing us to conclude that using ALGs is statistically more likely to be faster than using RKG based alternatives.

To ensure that skill level was not a significant residual anomaly we conducted one more variation on the model which asserted that when controlling for gesture type and complexity, there is insignificant variation across individuals and task completion times ($p<0.10$). That is to say, specific people are not especially quicker or slower overall at completing tasks.

The slower task completion times observed with RKGs may be due to participants being more likely to encounter usability problems as a result of the extra time required to gauge the responsiveness of the momentum equations controlling the menu. In the case of the ALGs, the fast onscreen reaction to a discrete physical event alludes to participants perceiving a reliable action-response mechanism and therefore not requiring as much time for self-calibration.

A 1-way Bonferoni's ANOVA conducted for ALGs shows that task completion time varied significantly between different task complexities $F(5, 175)=2.52$, $p < .05$. None of the post-hoc tests revealed individual significance beyond the .05 level.

The same ANOVA test performed for the RKGs, revealed a significant difference between task complexity and completion time: $F(5, 175)=7.13$, $p<.05$. In particular, post-hoc tests showed a significant difference between sequence index 2 and all the other sequence indices (1, 3, 4 and 5) where $p < .05$. This selection only required 1 movement if the shortcut were used. This suggests that participants would take the particularly obvious shortcuts with the ALGs more frequently than with the RKGs.

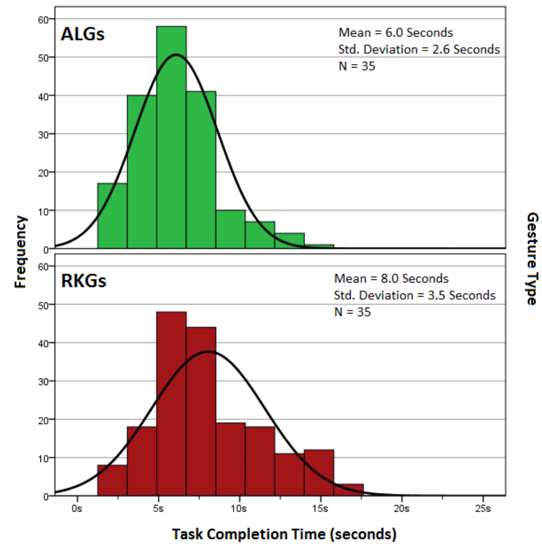

Figure 10. Side-by-side histograms that compare task completion time for ALGs (top) and RKGs (bottom).

6.1.2 Qualitative Analysis

Generally, the system was received very positively by participants and the overwhelming majority of people were willing to engage with the system when invited.

For both interaction techniques, several participants remarked on the initial learning curve through comments like "*At the beginning I would rate it a 10* [very hard], *but once you get the hang of it, it is really easy.*", "*takes a lot of getting used to, but once you have got it, it's fine*" and "*it gets much easier as you go along*". Given the low number of interactions each participant was required to do, this illustrates that it took a relatively short amount of time for people to become accustomed to the system.

During the experiments, passers-by would turn their heads and sometimes even stop for a few seconds to watch, much more so than in the previously discussed studies in Section 5. One man said: "*I'm intrigued to know what it is*". Such observations and comments, coupled with the low-awareness statistics discussed in the prior deployments suggest that initial understanding of similar displays in the wild would be low. Raising awareness of the system, perhaps through systems that encourage early-adopters who can serve as demonstrators would help address this issue.

A particularly interesting observation was the variety of subtle gestures participants made with their hands. These included: finger pointing, fists and flat palms. If the participant was able to successfully elicit a response from the display using a given hand pose, they would try changing it only when they encountered errors - conceivably associating it directly to the response.

With the ALGs, all but one participant made extensive use of the shortcuts throughout the study. The majority tended to complete 1 or 2 tasks - perhaps getting comfortable with the gesture - before trying out the other hand to use the shortcut. The person who did not use shortcuts employed a technique whereby they would keep spinning the menu round. This was not observed in any of the other participants for either interaction mode.

When using RKGs, participants used shortcuts right from the start which suggests that the ability to change the direction of the menu was an expected response from the system. Further, some found it easier to keep the momentum of the menu going rather than to reverse the direction.

When errors such as a missed gesture were experienced using ALGs, participants tended to gesture with more force and pause between motions to look for confirmation. However, when errors were encountered using RKGs, they tended to pose a more serious usability concern. At first, two participants struggled to move the menu at all. This observation correlates with the statistics in Figure 9 which show ALGs receiving a higher frequency of similar usability ratings around the mean and a shorter mean task completion time (Figure 10).

Unlike ALGs, those using RKGs often made superfluous subtle corrections to the number they had just brought to the front, ensuring that it was completely central, even after it had turned green. However, when correcting large errors such as an overshot number, participants using both ALGs and RKGs tended to gesture faster and more erratically – sometimes with their entire body. If the display did not respond as they expected, they did not change tactic back to the gestures that were previously working, instead increasing the tenacity of their gesturing.

7. KEY FINDINGS

People appear more interested in the display when they see it being used by another person. In addition, people taking photographs of their friends and other social motivators like competition were useful in helping to raise the profile of the display. This builds on the 'gradual buy in' and 'tutors' observed by Izadi et al. [11] and supports both the 'observe-and-learn' model [17] and Honeypot effect [5].

In contrast to previous evaluations which place displays as primary artefacts in goal orientated scenarios, we find that deployment alone is not enough for a gesture based public display to attract attention and thus easily move from the periphery into active social space. It would be interesting to maps social motivators and physically distinct actions (such as taking a photograph, or standing in a small / large group) onto varying degrees of success at propagating awareness via this Honeypot effect [5].

Wall mounted displays have a relatively short window of opportunity in which to engage passers-by. The relatively fast pace at which people passed our display leaves the person a small amount of time to decide if they are going to watch, glance at, or ignore the display. This raises interesting questions regarding the optimal point at which to attract an approaching person's attention. It may be possible to establish an interest falloff curve where the probability of success diminishes relative to distance and speed of approach.

Environmental security may affect and to a certain extent define the interaction techniques people choose. Of those who used the system, several were carrying bags and proceeded to place them on the floor before interacting. In less secure environments, or outdoors, the system would ostensibly not reach its full potential because people would not be willing to risk their possessions. It may be possible to cater for this by simply prescribing a single handed interaction technique. This would also allow people on mobile phones to continue their conversation while interacting. It would be valuable to see more longitudinal studies that examine relationship between such factors and the reasons why people do or do not interact with a display.

It is difficult to conclude if content design matters more than social setting. We observed that content type (static, animated) affected people's attention levels more than the social context the people were in. Different models of attention have been cited by the public display field before [10] [7], yet research still lacks a way to map characteristics and behaviours of displays onto the different understandings offered by the models. To that end, attention levels and perception remain relatively unexplored beyond anecdotal evidence. Works which measure and quantify the effects of different content features (such as striking changes in graphical contrast, sounds, pulsing) on attention levels would be an extremely useful reference in the design of future systems and a ground truth in their evaluation.

Different gesture models are more appropriate for different forms of application interaction. If content interaction consists of discrete events such as item selection on a menu, absolute location gestures (ALGs) were shown to be faster, easier to learn and more user friendly than relative kinaesthetic gestures (RKGs). The discreteness, repeatability and dependability of the ALGs offer a more suitable means of correcting for large errors in menu item selection.

RKGs offer an intuitive interaction method for small and continuous movements such as error corrections. For example, ensuring a menu item was fully central. However, users expected more detailed functionality from this interaction mode such as the ability to rotate the menu both ways. This observation was mirrored in the space game where people attempted to fly the ship up and down as well as from side-to-side.

People would gesture and move more vigorously when encountering problems. Using ALGs this manifested itself as gesturing with more force. When the system did not respond how users of RKGs expected; instead of slowing down, they made larger, faster and more erratic motions. Error management functionality on real world displays should focus on responding to such behaviours.

As a guideline, if users are repeatedly executing fast, large gestures, it often means the system is not responding quickly enough or in the way that the user is expecting.

People do not always successfully adopt the correct interaction technique, even when presented with clear onscreen instructions. Carefully communicating interaction issues will naturally help avoid the propagation of bad interaction techniques between contiguous users; something that is especially important given the aforementioned Honeypot effect [5].

The short amount of time that public displays have to engage passers-by means that it is challenging to select the best time and medium over which to communicating better interaction techniques. We observed that even with a human demonstrator to copy, some people had problems adopting the desired gestures. With this in mind, any time spent educating the user must be traded off against design factors. Subsequently, gesture based public displays are a prime candidate for user interfaces that communicate using natural metaphors. For example: rendering a hand (representing the users) interacting with virtual objects on screen.

An adaptive gesture detection algorithm could conceivably profile the types of gesture the user is making, taking into account different natural responses such as correction though big, small, fast and slow movements. This would hopefully lead to a more intuitive experience that would in turn promote increased usage.

People did not expect wall mounted digital displays to be interactive. Vogel and Balakrishnan observed that users immediately understood that their body position was controlling the display [24]. However, we observed that when people were invited to interact they would often not make the association until it was explicitly explained. In some cases, the system's capabilities were so far outside normal expectations that people did not believe it was real. Unless this mismatch between expectations and capabilities is addressed, it may well limit the adoption of similar systems.

Once made aware of the display's purpose and the potential for interaction, participants happily integrated the display into their social space as groups and individuals. This raises questions about how active a role they should take in engaging people. For instance, would public displays overcome 'display blindness' [15] and successfully engage more people if they actively targeted passers-by rather than passively waiting to be approached?

8. CONCLUSIONS

This paper presents a real-world-ready example of applied gesture based digital signage. It consists of a simple, low-cost hardware and software architecture that provides satisfactory performance in real world conditions. The system is deployable in many situations without the need for calibration and has since been used by large numbers of visitors at two major science fair exhibitions.

While this work demonstrates that gestures are a viable method of interaction with public digital signage, such displays still suffer from naturally low audience engagement figures due to a wide variety of reasons. We note that the Honeypot effect [5] is one of the most readily reproducible effects in digital signage and discuss its impact on the propagation of interaction techniques between users who are socially connected. Our observations raise questions about the nature of the decision making process governing the amount of attention people pay displays and comment on challenges posed by the short window of opportunity available to engage passers-by.

From a comparison of gesture types in the context of menu item selection our results show that absolute location gestures offered the best overall user experience and task completion time. We have discussed the relative subtleties of these gesture types given an analysis of learning times and error corrections based on their use in a real world setting.

An immediate challenge in the future of these gesture based displays is achieving a significant level of system awareness leading to sustainable usage. Indeed, appropriate content and functionality with respect to a user's situation and expectations will be vital in assisting their adoption.

9. ACKNOWLEDGMENTS

Many thanks to Stephen Wattam, Carl Ellis, Peter Newman and Andrew Wise for their kind support.

This work was conducted as part of the post-disciplinary HighWire programme at Lancaster University in collaboration with the PD-Net project. HighWire is funded by the Digital Economy Programme: a Research Councils UK cross council initiative led by the EPSRC and contributed to be AHRC, ESRC and MRC. Part of this work has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 244011.

10. REFERENCES

[1] Microsoft. XBOX Kinect Homepage. [Online] 2011. [Cited: October 26, 2011.] http://www.xbox.com/kinect.

[2] dhiram via NewLaunches.com. Gesture based Interaction screen make the iPhone look from the 19th century. NewLaunches.com. [Online] January 19, 2008. [Cited: October 26, 2011.] http://www.newlaunches.com /archives/gesture_based_interaction_screen_make_the_iphone_look_from_the_19th_century.php.

[3] Technovelgy LLC. Interactive Apple Store Has Reactive Walls. Technovelgy. [Online] Technovelgy LLC, November 24, 2007. [Cited: October 26, 2011.] http://www.technovelgy.com/ct/Science-Fiction-News.asp?NewsNum=1321.

[4] Ahn, S, Lee, T, Kim, I, Kwon, Y, and Kim, H. Large Display Interaction Using Video Avatar and Hand Gesture Recognition. In Campilho, Aurélio and Kamel, Mohamed, eds., *Image Analysis and Recognition*. Springer Berlin / Heidelberg.

[5] Brignull, H and Rogers, Y. Enticing People to Interact with Large Public Displays in Public Spaces. *INTERACT'03* (2003), 17-24.

[6] Churchhill, E, Nelson, L, and Denoue, L. Designing digital bulletin boards for social networking. In *In Workshop on Public, Community and Situated Displays at CSCW 2000* (New Orleans 2002).

[7] Hamker, F. The emergence of attention by population-based inference and its role in distributed processing and cognitive control of vision. *Computer Vision and Image Understanding - Special issue: Attention and performance in computer vision*, 100, 1-2 (October 2005), 64--106.

[8] Huang, E, Koster, A, and Borchers, J. Overcoming Assumptions and Uncovering Practices: When Does the Public Really Look at Public Displays? *Pervasive Computing*, 5013/2008 (2008), 228-243.

[9] Huang, E and Mynatt, E. Semi-Public Displays for Small, Co-located Groups. In *Conference on Human Factors in Computing Systems Proceedings of the SIGCHI conference on Human factors in computing systems* (Ft. Lauderdale, Florida 2003), ACM, 49-56.

[10] Itti, L, Koch, C, and Niebur, E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 11 (November 1998), 1254--1259.

[11] Izadi, S, Fitzpatrick, G, Rodden, T, Brignull, H, Rogers, Y, and Lindley, S. The iterative design and study of a large display for shared and sociable spaces. In *Proceedings of the 2005 conference on Designing for User eXperience* (San Francisco, California 2005), AIGA: American Institute of Graphic Arts.

[12] Lee, J, Lee, J, Kim, H, and Kim, J. Gesture-Based Interactions on Multiple Large Displays with a Tabletop Interface. In Stephanidis, Constantine, ed., *Universal*

Access in Human-Computer Interaction. Ambient Interaction. Springer Berlin / Heidelberg, 2007.

[13] Mccarthy, J, Costa, Tony J. and Liongosari, Edy S. UniCast, OutCast & GroupCast: Three Steps Toward Ubiquitous, Peripheral Displays. In *UbiComp '01: Proceedings of the 3rd international conference on Ubiquitous Computing* (2001), 332-345.

[14] Müller, J, Alt, F, Michelis, D, and Schmidt, A. Requirements and design space for interactive public displays. In *Proceedings of the international conference on Multimedia* (Firenze, Italy 2010), 1285-1294.

[15] Müller, J, Wilmsmann, D, Exeler, J, Buzeck, M, Schmidt, A, Jay, T, and Krüger, A. Display Blindness: The Effect of Expectations on Attention towards Digital Signage. In Tokuda, Hideyuki et al., eds., *Pervasive Computing*. Springer Berlin / Heidelberg, 2009.

[16] Rogers, Y. and Lindley, S. Collaborating Around Large Interactive Displays: Which Way is Best to Meet? *Interacting with Computers* (2004).

[17] Rubegni, E, Memarovic, N, and Langheinrich, M. Talking to Strangers: Using Large Public Displays to Facilitate Social Interaction. In *14th International Conference on Human-Computer Interaction (HCII 2011)* (2011).

[18] Rukzio, E. *Physical Mobile Interactions: Mobile Devices as Pervasive Mediators for Interactions with the Real World*. München, 2006.

[19] Russell, D. M., Drews, C, and Sue, A. Social Aspects of Using Large Public Interactive Displays for Collaboration. *UbiComp 2002: Ubiquitous Computing*, 2498 (2002), 663-670.

[20] Skog, T, Ljungblad, S, and Holmquist, L.E. Between Aesthetics and Utility: Designing Ambient Information Visualizations. *INFOVIS 2003. IEEE Symposium* (2003), 233-240.

[21] Storz, O., Friday, A., Davies, N., Finney, J., Sas, C., and Sheridan, J.G. Public Ubiquitous Computing Systems: Lessons from the e-Campus Display Deployments. *Pervasive Computing, IEEE* , 5, 3 (July-Sept 2006), 40--47.

[22] Tang, A, Finke, M, Blackstock, M, Leung, R, Deutscher, M, and Tain, G. Designing for Bystanders: Reflections on Building a Public Digital Forum. In *Conference on Human Factors in Computing Systems Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (Florence, Italy 2008), ACM, 879-882.

[23] Taylor, N and Cheverst, K. Social interaction around a rural community photo display. *International Journal of Human-Computer Studies*, 67, 12 (December 2009), 1037-1047.

[24] Vogel, D and Balakrishnan, R. 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* (Santa Fe, NM 2004), ACM, 137 - 146.