Towards Learnable Gestures for Exploring Hierarchical Information Spaces at a Large Public Display

Christopher Ackad
Faculty of Engineering and Information Technologies
The University of Sydney, NSW, 2006, Australia
christopher.ackad@sydney.edu.au

Judy Kay
Faculty of Engineering and Information Technologies
The University of Sydney, NSW, 2006, Australia
judy.kay@sydney.edu.au

Martin Tomitsch
Faculty of Architecture, Design and Planning
The University of Sydney, NSW, 2006, Australia
martin.tomitsch@sydney.edu.au

ABSTRACT
Large displays are fast entering the public spaces, and some are beginning to support interaction based on mid-air gestures. At present, there is no set of standard gestures for that interaction; so people need to draw on existing mental models to learn such interaction. As a step towards closing this gap, we present the design and evaluation of learnable gestures for exploring a hierarchical information space on a public display. We draw on a series of in-the-wild studies, observing how people learn our 4 gestures, combined with explanatory icons and a skeleton representation for feedback. We conclude that people readily learnt the pair for exploring a linear information space, while having more difficulty learning the gestures to move up and down the hierarchy.

Author Keywords
Gesture-based Interaction; Interactive Wall; Public Displays;

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
With falling display costs and the availability of large screens and high-visibility projection technologies, public displays have found their ways into a wide range of public spaces. Uses span train departure times, to the broadcast of sports events in pubs, and advertising in foyers. To date, most public screens have not been interactive; rare exceptions are some SMS-based audience participation [10]. This is set to change, with the arrival of cheap and readily available technologies, such as the Microsoft Kinect. This can recognise mid-air gestures, creating unprecedented opportunities to transform these large public screens into interactive platforms. Mid-air gestures are suited to large screen interaction, so that people can view the content from a distance. Mid-air gestures enable immediate use, unlike approaches that use a secondary display such as a mobile phone [2].

We designed the Media Ribbon public display [4, 1], to enable users to explore a potentially large hierarchical information space on the wall. Our goal was to enable people to use mid-air gestures to browse, find areas of interest, and then drill down to further information. This makes the Media Ribbon an ideal testbed for exploring learnable gesture sets. Our design aimed to enhance learnability, by ensuring: the gesture set is small, gestures are easy to describe/teach, perform, remember and the technology should reliably recognise them. Taking account of the public setting; we aimed for gestures people would find acceptable to perform in public.

BACKGROUND
There has been considerable work on attracting users to public displays [4, 8], and exploring social behaviours [11, 5] around these displays. For example, the Chained Displays interface [11] used an interactive space invaders style game to gain the interest of passers-by. The game used a number of discoverable poses to control the strength and frequency of the player’s weapon. Similarly, the StrikeAPose interface [12] used a game to explore the learnability of a gesture through 3 different on-screen cues: a static instruction shown at the bottom of the display at all times, another where the instructions obstruct the display, and a third with the instruction integrated and displayed as a tool tip above the user’s representation on the screen.

Hespanhol et al. [6] explored the learnability of four gesture primitives (Dwell, Lasso, Push, Grabb) without any user prompts. They conclude that the most intuitive of these was “Dwell”, perhaps indicating a mental model based on using the hand like a mouse. To our knowledge, there are currently no standards or other primers for the gestures people would expect to use to navigate a hierarchical information space.

DESIGNING THE MEDIA RIBBON
We have deployed our “Media Ribbon” in a number of locations, from interior corridors during an Open Day event at our university to the exterior of buildings for our long term in-the-wild study. Our current installation is situated on the exterior wall of a glass building, adjacent to a theatre courtyard (Figure 1). The Kinect sensor reaches out to the pedestrian
Figure 1. The Media Ribbon running on the public display (left) facing a theatre courtyard (right).

Figure 2. The Media Ribbon showing the tutorial on an expanded top-level item that has sub items.

Figure 3. The on-screen help bar with the “Right” gesture highlighted.

walkway running between the two buildings, with the installation being clearly visible from within the courtyard. The installation consists of two high-intensity 1080p projectors, for rear projection on a film laminated onto the glass wall. A 2 centimetre gap separates the two projection panes. A single Kinect for Windows hangs in its own enclosure above and in front of the projection display.

The version of “Media Ribbon” shown in Figure 1 enables users to explore a hierarchical structured set of information, such as upcoming events, faculty news and a showcase of research projects. It is a wide interactive display, which first appears as a horizontal ribbon of media items (Figure 2). To aid engagement with the Media Ribbon, we followed the recommendations [7] to use a wide display, so that passers-by have time to react and engage while still in front of the display. It recognises four primary mid-air gestures (Left, Right, More and Back) which we now describe.

The central element of the Media Ribbon (Figure 2), is a ribbon showing the currently selected hierarchy level and all its items. The ribbon can be scrolled left or right. At any time the items closer to the centre are largest and as the user scrolls, items moving towards the centre become larger, and those moving from the centre become smaller. This means the central item is largest, and neighbouring items provide context. Once an item is centred, it automatically expands, displaying the multimedia content (an image or video) on the left and descriptive text on the right. This design approach was partly motivated by need to avoid having the focus content having a gap in the middle.

We provide three ways to help users learn the gestures. First is a real time representation of the user, as a skeleton on top of the content. Second is the four icons shown in Figure 3. We place this help bar below the content, right in the user’s direct line of sight. Third, the help content gives real-time feedback on the user’s interactions; when a gesture is recognised, its icon is highlighted, as shown for the “right” gesture in Figure 3. This less intrusive method of presenting the available gestures, was shown to be effective in a recent study [12].

To navigate through the Media Ribbon’s hierarchical information structure, the user needs to swipe left or swipe right for horizontal navigation within the current level of the hierarchy. The “More” gesture calls for the user to hold one arm up; this moves the current item up, displaying more information about that topic. Items with sub-items have the same “More” icon and text telling the user to “Hold your Arm Up to find out more” as in Figure 2. The “Back” gesture is achieved by holding one arm at a 45 degree angle from the waist, similar to the Xbox 360 pause menu gesture (Figure 3). For media such as photos or videos, users can interact to “like” them, using the “More” gesture. For such items, the arm-up icon appears indicating the number of likes and whether this user has “liked” the item (see Figure 4). A user can undo their vote, by repeating the “More” gesture.
3. The gesture set had to be socially acceptable for use in a public space; by contract our gesture language must be learnt quickly and without investment in learning sign language to use it for the long term. We were aware that sign language is fundamentally different from our gesture language in that people expect to find gestures required only one hand [1].

Our first design for these were two-handed. The matching “More” and “Back” gestures were difficult for participants to learn; often one of the researchers had to intervene to explain and demonstrate the gestures. Overloading the gesture is a clear risk for learning. But it also limits the gesture set that needs to be learned. We felt that these more expressive gestures worked well in an entertainment context (such as games) but may not be appropriate for an information exploration context. We also needed to consider whether people would be likely to find gestures inappropriate for a public space [9] making them uncomfortable and reluctant to use them with the Media Ribbon.

Based on these findings and explorations we arrived at the current gestures described above. We note that the arm-up “More” matches the familiar hands-up from the classroom, when asking a question. Another factor in our gesture design was to limit it to single arm gestures; we would expect that people often have one arm occupied, for example, carrying a drink or bag.

We then decided to add a voting gesture and reused the “More” gesture. Overloading the gesture is a clear risk for learning. But it also limits the gesture set that needs to be recognised by the system and learnt by the user. This is in line with work [3] indicating that such overloading, in distinct contexts has lower cognitive load than adding additional unique gestures.

**DISCUSSION**

We designed the “Media Ribbon” with just four gestures, aiming for a minimal but effective set. Our experience in the design process established the difficulty in communicating and reliably recognising more complex gestures. Both our studies [4, 1] and other work [8] indicate that people are likely to interact with these displays for very short periods, typically 30-60 seconds, making for very tight constraints in learning.

We found that our swiping gestures were easy to learn; perhaps due to the very direct mapping to the on-screen results. By contrast, our semaphoric gestures, “More” and “Back” seemed harder to learn, perhaps reflecting a mismatch between the action, description and result. In our theatre study [1], many users seemed to ignore the interactive tutorial. Our results indicate that the real-time feedback of the icons do facilitate learning the gestures, without an explicit tutorial. For our context, it also turned out to be quite important that gestures required only one hand [1].

Another important aspect of designing interactive displays is providing real-time feedback. Previous work in this area [8] has demonstrated that presenting passers-by with a silhouette
or video of themselves is more successful for attracting users than a written call to action. Our current “Media Ribbon” has a skeleton as the user representation. In future work, we plan to compare this against the effectiveness of a silhouette representation, both for learning and perceptions of the interface, particularly in terms of the sense of fun. Our skeletal representation seems important for feedback on users’ interactions [4]. Equally, the changing icons of the last recognised gesture seems important. We are also exploring additional means of feedback; as we can now recognise partial gestures, we can begin to respond and make the display make gradual changes so that a swipe left gesture will start moving the ribbon left a little early in the gesture and then continue as the user finishes the gesture; this approach may aid the learning of the up-down gestures as well. Of course, it also brings a risk of introducing confusion if very small arm movements that the user did not intend as a gesture produces an unexpected effect.

CONCLUSION

This paper has presented the design of the “Media Ribbon” mid-air gesture-controlled interface. We presented our design rational for the four gestures to navigate and interact with information hierarchies. Our in-the-wild studies provided insights into people’s behaviours and expectations, beyond those possible in the lab.

Our work suggests that a small number of gestures is effective, perhaps because it reduces the cognitive load [3]. We observed that our manipulative gestures, involving swiping, were more intuitive than our semaphoric gestures, “More” and “Back”. Providing real-time and responsive feedback is critical to reinforcing gestures and making users aware of the display’s interactive nature. Overloading gestures helps to minimise the total number of gestures that need to be learnt and potentially reinforces them in users’ memories from reuse. Environmental and social factors [9] affect the likelihood of people using the gesture vocabulary and it is important to consider even the very mundane matter of designing for use with one hand.

Our work provides a body of empirical, in-the-wild evidence about the design of learnable gestures for public display interaction for browsing through a hierarchical information space. Our mid-air gestures are quite different from most user’s experience of interaction and cannot build upon existing mental models of mouse interaction.

REFERENCES


