PPD 2008: Workshop on designing multi-touch interaction techniques for coupled public and private displays

May 31st 2008, Naples, Italy, as part of AVI 2008
(the International Working Conference on Advanced Visual Interfaces)

Dear Attendee,

The PPD 08 Workshop on designing multi-touch interaction techniques for coupled public and private displays focuses on the research challenges and opportunities afforded by the combination of touch sensitive small private input displays coupled with large touch sensitive public displays. Different touch-enabled devices rely on different types of touches (passive stylus, active stylus, fingers and tangible objects), the motivating question for this workshop is how do users switch between these devices and how to facilitate fluid transition from a collection of multiple displays to a single integrated multi-display environment.

Recent developments have seen the wide spread proliferation of both large shared displays and small display technologies. In parallel we have seen the emergence of new classes of device which support both touch or multi touch interaction. Examples of small touch driven devices include PDAs, Tablets and iPhones and examples of large interactive surfaces (multi-touch driven displays) include the Diamondtouch and Surface Computing. Interactive surfaces offer great potential for face-to-face work and social interaction and provide natural ways to directly manipulate virtual objects whereas small devices afford the individual a personal workspace or "scratch space" to formulate ideas before bringing them to a wider audience. Advanced visual interfaces can be built around a combination of both private and public touch driven displays. Such computer mediated multi-device interaction between local touch-driven displays and shared public ones presents a number of novel and challenging research problems.

Our aim with this workshop will be to focus on the research challenges in designing touch interaction techniques for the combination of small touch driven private input displays such as iPhones coupled with large touch driven public displays such as the Diamondtouch or Microsoft Surface.

We look forward to your hearing about your work in this area and brainstorming during the workshop.

Shahram Izadi - Microsoft Research Cambridge, UK
Aaron Quigley - University College Dublin, Ireland
Sriram Subramanian - University of Bristol, UK

Program

9.00:	Welcome
9.10:	Workshop Theme A - PPD Multi-touch interaction technologies: from Prototyping to Taxonomy **Shahram Izadi**
9.20:	Workshop Theme B - PPD Multi-touch interaction applications: from Use Cases to Scenarios **Sriram Subramanian**
9.30:	Workshop Theme C - PPD Multi-touch interaction evaluations: from Methodologies to Guidelines **Aaron Quigley**
9.40 - 10.30:	Five Workshop Presentations (10 mins each) *Using Public and Private Displays in a collaborative learning environment,* Edward Tse, Gerald Morrison, Smart Technologies, 1207 – 11 Avenue SW, Suite 300, Calgary, AB T3C 0M5. *A Touch Interaction Model for Tabletops and PDAs*, Xiangan Heng, Songyang Lao, Hyowon Lee, Alan F. Smeaton, Centre for Digital Video Processing & AIC, Dublin City University, Dublin, Ireland and National University of Defense Technology, Changsha, China. *Palm Interface: a display personally to show information in public spaces, by using image processing*, Yoko Ishii, Minoru Kobayashi, Mutsuhiro Nakashige, Hideki Koike, NTT Cyber Solutions Laboratories, 1-1 Hikarinooka Yokosuka-si, Kanagawa 2390847 JAPAN and University of Electro-Communications, Tokyo, 1-5-1 Chofugaoka Chofu-si, Tokyo 1820021, JAPAN. *Natural Gesture-Based Techniques for Sharing Documents from a Private to a Public Display*, Md. Mahfuzur Rahman, Pourang Irani, Peter Graham Department of Computer Science, University of Manitoba, Winnipeg, Canada. *Analysing Fluid Interaction across Multiple Displays*, Richard Morris, Paul Marshall, Yvonne Rogers, Open University, Milton Keynes, UK.
10.30 - 11.00:	Coffee
11.00 - 12.20:	Eight Workshop Presentations (10 mins each) *3-D Interaction with Wall-Sized Display and Information Transportation using Mobile Phones*, Hideki Koike, Masataka Toyoura, Kenji Oka, Yoichi Sato Graduate School of Information Systems University of Electro-Communications, 1-5-1, Chofugaoka, Chofu, Tokyo 182-8585, Japan and Institute of Industrial Science, University of Tokyo, 4-6-1 Komaba, Meguro-ku Tokyo 153-8505, Japan *The Joy of Use and Supporting Group Interaction in Multi-Touch Environments*, Ann Morrison, Giulio Jacucci, Peter Peltonen,

Helsinki Institute for Information Technology (HIIT) Helsinki University of Technology and
University of Helsinki, P.O. Box 9800, 02015 HUT, Finland.

Collaborative problem solving on mobile hand-held devices and stationary multi-touch interfaces, Simon Nestler, Florian Echtler, Andreas Dippon, Gudrun Klinker,
Technische Universität München, Institut für Informatik.

Using Mobile Phones to Spontaneously Authenticate and Interact with Multi-Touch Surfaces, Johannes Schöning, Michael Rohs, Antonio Krüger,
Institute for Geoinformatics, University of Münster, Robert-Koch-Str. 26-28, 48149 Münster, Germany and
Deutsche Telekom Laboratories TU Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany.

Multi–Touching 3D Data: Towards Direct Interaction in Stereoscopic Display, Environments coupled with Mobile Devices, Frank Steinicke, Klaus Hinrichs, Johannes Schöning, Antonio Krüger,
Visualization and Computer Graphics, Westfälische Wilhelms-Universität Münster, Einsteinstraße 62, Münster, Germany and
Institute for Geoinformatics, Westfälische Wilhelms-Universität Münster, Robert-Koch-Str. 26-28, 48149 Münster.

Sociality, Physicality and Spatiality: touching private and public displays, Alan Dix, Corina Sas, Devina Ramduny-Ellis, Steve Gill, Joanna Hare,
Computing Department, InfoLab21 Lancaster University, Lancaster, LA1 4WA, UK and
Cardiff School of Art & Design, University of Wales Institute, Cardiff, Wales, CF5 2YB, UK

"tune_eile": A platform for social interactions through handheld musical devices, Nora O Murchú,
Interaction Design Centre, Dept of Computer Science & Information Systems, University of Limerick

A Taxonomy for Exploring the Design Space of User-Display-Display Interactions, Lucia Terrenghi,
Vodafone GROUP R&D, Chiemgaustrasse 116 81549, Munich.

12.20 - 12.30: Overview of breakout/working sessions
12.30 - 2.00: Lunch (not provided but venue will be booked)
2.00 - 3.30: Three Parallel Breakout / Work-Group Sessions

Technologies
Applications
Evaluations

3.30 - 4.00: Coffee
4.00 - 4.30: Report from each breakout group (10 mins each)
4.30 - 5.00: Discussion
6.30 - 9.00: Dinner (not provided but venue will be booked)

Using Public and Private Displays in a Collaborative Learning Environment

Edward Tse, Gerald Morrison
Smart Technologies
1207 – 11 Avenue SW, Suite 300
Calgary, AB T3C 0M5
1 (403) 245-0333
[edwardtse, geraldmorrison]@smarttech.com

ABSTRACT

While the pervasive use of interactive whiteboards in education has emphasized the need for large public displays, a growing number of schools are moving towards a one to one (student to computer) ratio. Thus there is significant interest in understanding how personal devices (e.g. laptops, clickers) can be coupled with large public displays in a technology enabled classroom.

We describe a correlation between shifting educational practice (from teacher-centric to learner-centric) and its effect on education technology. In particular we explore how interactive response systems, networked computers, small shared displays, and large public displays have been used to create collaborative learning environments.

Categories and Subject Descriptors

H.5.2 [**Information interfaces and presentation**]: User Interfaces – Interaction Styles

General Terms

Design, Human Factors

Keywords

Public Displays, Private Displays, Shared Displays, Education.

1. INTRODUCTION

Educational pedagogy is shifting from teacher-centric towards active teacher/learner participation and even further to learner-centric collaboration where learners are expected to coach and collaborate with their peers.

Traditional educational practices have focused on a teacher-centric "sage on the stage" metaphor where the teacher transmits knowledge through large public surfaces such as blackboard or interactive whiteboards and learners play the role of passive recipients.

Contemporary educational practices such as constructivism [Vygotsky, 1978] emphasize the notion of learners as an active participant in the learning process. In constructivism, learning is enabled by expanding upon one's current knowledge through discussion and engagement; this is referred to as the zone of proximal development (ZPD). Feedback from instructors and peers is the crucial in constructivist learning.

To encourage this discussion and negotiation, constructivist educators try to engage the learner through small group activities with peers (typically 2-4) and collaborative exercises. Their goal is to create an environment for collaborative learning using prepared content and peer support.

Teachers apply constructivism in the classroom by manually assigning groups of learners and providing them with collaborative exercises typically printed on paper. There are several reasons why this is not practical in a typical classroom: The preparation is cumbersome because the collaborative exercises typically cannot be reused and need to be printed for each collaborative class. The organization of groups consumes a significant amount of class time providing less time for learners to focus on the taught materials. It is difficult for teachers to examine the progress of each individual learner as classroom sizes become larger. The logistics of delivering content and soliciting feedback significantly increases the burden upon teachers and learners in the collaborative classroom.

2. COLLABORATIVE TECHNOLOGY IN THE CLASSROOM

Educational technology has provided a means to simplify the delivery of content and solicit feedback in the classroom. These technologies vary in the amount of personal work space and visibility to other collaborators. We describe four main scenarios: interactive response systems, networked computers, small shared displays, and large public displays.

Figure 1. An interactive response system in a classroom (left) an individual clicker (right) from smarttech.com/Senteo

2.1 Interactive Response Systems

Interactive response systems (e.g. Figure 1) typically involve a clicker for each learner in the classroom. Some clickers (such as Senteo by Smart Technologies, Figure 1, right) contain a private text displays that can be used to review one's answer prior to submitting it. A teacher can post a question on a large interactive whiteboard and have the question deployed to each learner wirelessly. Feedback can be provided to each learner through their private displays or through the interactive whiteboard using a pie chart or bar chart visualization. The clicker is not only useful for broadcasting questions from the teacher, but also for eliciting questions from learners. By pressing a help button on the clicker any learner can initiate a discussion in the classroom.

While Figure 1 (left) shows the most use scenario for interactive response systems in the classroom, teachers can also use these systems to facilitate group discussion and negotiation. One method is to provide each group with a single clicker and ask groups to agree on an answer before submitting a response. Generally these groups depend on the interactive whiteboard for viewing shared content as the private clicker display is not readily visible to all collaborators. The limited personal space available on clickers means that most of the students express their thought process through external materials such as pencils and paper.

2.2 Networked Computers

In a one to one school, there is one computer available for each learner. These computers can be desktop computers or personal laptops such as the OLPC XO, Intel's Classmate PC, and the Asus EEE PC. By networking these computers together it is possible to deliver content and elicit responses in a very similar fashion to interactive response systems. One advantage is that networked computers provide a much larger personal space that expands the types of questions that can be asked (sorting, matching, open answer, drawing) and allows the thought process of learners to be expressed through the computer. Another advantage is that this configuration bridges a student's personal work to the collaborative setting, where the work they have done individually (and perhaps previously) can be brought into the collaborative space and vice versa, e.g., through file-sharing.

This provides an additional channel of feedback through the digital system. In addition to seeing the correct response on their personal computer, and the overall response rate on the interactive whiteboard, teachers can provide illustrative feedback about a particular student's thought process. For example, a digital recording of a student's manipulations could be shown on the interactive whiteboard to provide insights for the rest of the class.

Face to face small group collaboration is a bit more cumbersome as students may not be able to move their computers (e.g. a class with desktop computers for each student). Virtual teams can be automatically assigned to allow learns to work in small groups. However, since each display is tilted towards the learner it is generally considered private. Thus awareness of a collaborator's actions must be provided through software such as telepointers [Greenberg et al., 1996], split screen and radar views [Gutwin et al., 1996].

2.3 Small Shared Displays

Another approach is to provide learners with a single shared display specifically designed for groupwork. An example is the tabletop in Figure 2. On a shared display a small display size allows each learner to view and gesture over the entire surface, and a horizontal layout allows collaborators to see the on screen manipulations as well as the facial expressions of collaborators. These benefits facilitate the negotiation and discussion that is core to constructivist learning.

Figure 2. A small tabletop display for children

If the shared display runs standard commercial software, they are limited by the fact that most underlying educational software only supports the keyboard and mouse inputs of a single individual. For example, if multiple mice are connected to a single computer, their inputs are merged into a single stream of keyboard and mouse events. Studies of shared displays in an educational setting have emphasized the need for multiple simultaneous inputs on shared displays as children have been known to lose interest in the task at hand when they are waiting to use the computer [Inkpen et al., 1999].

Modern multi user, multitouch, and multi input technologies are able to circumvent current limitations but require content providers to write software for these specific hardware configurations. One lightweight input method is to multiple mice for each student using a shared display. This scenario has been tested in rural classrooms [Pawar et al., 2006]. However, the cursors from multiple mice are small and can be ineffective for monitoring the activities of collaborators. Another approach is to provide multitouch tabletop for groups of 2-4 learners. The tabletop is beneficial because the consequential communication caused by manipulating the tabletop also serves as communication to other collaborators [Tse et al., 2006]. Information on the shared display is readily visible to immediate collaborators but hidden to other groups. This allows groups to develop very independent solutions that can be efficiently reviewed by the instructor on a large vertical display. Since there is limited personal space in a small tabletop, the work done on the tabletop is more of a reflection of the group's thinking process over that of a single individual.

2.4 Large Public Displays

A large public display such as an interactive whiteboard can be used for collaborative learning if they support multiple simultaneous inputs. If a single interactive whiteboard does not provide sufficient personal space, multiple whiteboards can be tiled to provide a very large interactive canvas that classmates can easily see.

If the personal space of each participant is large, students will be able to see the work of collaborators while also being able to do independent work in their own personal space. This is particularly useful when people are engaged in learning that can divided into separate sub sections that can be synchronized when completed.

	Small Personal Space	**Large Personal Space**
Private Display	Interactive Response Systems	Networked Computers
Public Display	Small Shared Displays	Large Public Displays

Table 1. Comparison between display visibility and personal space for technology enabled collaborative learning.

3. Conclusion

Table 1 describes two axes for describing collaborative learning: personal space and display visibility. As learners have more personal space in a digital system they have an increased ability to do independent work and show their work on the digital system. This increased personal space comes at the cost of a learner's ability the monitor and aid the learning of co-located peers. Small personal spaces are beneficial for close knit collaboration but are less effective for supporting individual work. As the display becomes more publicly visible to collaborators, it becomes easier for groups to share a common view and monitor the activities and expressions of other learners. Private displays allow learners to work on content that they would not like to reveal to other members of the group (such as quiz answers). An understanding of the effects of personal space and display visibility will be crucial in supporting next-generation educational pedagogy.

4. REFERENCES

[1] Greenberg, S., Gutwin, C., and Roseman, M. (1996) Semantic Telepointers for Groupware. Proc. OZCHI '96. IEEE Computer Society Press, 54-61.

[2] Gutwin, C., Roseman, R. and Greenberg, S. (1996) Workspace Awareness Support With Radar Views. Proc. ACM CHI '96, 210-211.

[3] Inkpen, K.M., Ho-Ching, W., Kuederle, O., Scott, S.D., and Shoemaker, G.B.D. (1999). "This is fun! We're all best friends and we're all playing.": Supporting children's synchronous collaboration. Proc. CSCL '99, 252-259.

[4] Pawar, U. S., Pal, J., and Toyama, K. (2006) Multiple mice for computers in education in developing countries, IEEE ICTD 2006. 64-71.

[5] Smart Technologies Senteo www.smarttech.com/Senteo

[6] Tse, E., Shen, C., Greenberg, S. and Forlines, C. (2007) How Pairs Interact Over a Multimodal Digital Table. Proc. CHI 07. ACM Press, 215-218.

[7] Vygotsky, L. (1978). Mind in Society. London: Harvard University Press.

A Touch Interaction Model for Tabletops and PDAs

Xiangan Heng[1, 2], Songyang Lao[1], Hyowon Lee[2], Alan F. Smeaton[2]

[1]National University of Defense Technology
Changsha,
China
laosongyang@vip.sina.com

[2]Centre for Digital Video Processing &
Adaptive Information Cluster,
Dublin City University, Dublin, Ireland
{xheng, hlee, asmeaton}@computing.dcu.ie

ABSTRACT
Currently the definition of touch interactions in touch-based interfaces is application- and device-specific. Here we present a model for touch interaction which gives a deeper understanding of touch types for different devices. The model is composed of three levels – action, motivation and computing – and mappings rules between them. It is used to illustrate touch interaction in a tabletop and a mobile application and allows us to re-use touch types in different platforms and applications in a more systematic manner than how touch types have been designed to date.

1. INTRODUCTION
Currently, multi-touch technology is a popular research area in the field of Human-Computer Interaction, gaining momentum with the appearance of commercial products such as Apple's iPhone and Microsoft's Surface. Previous research on touch/gestural interaction has concentrated on gesture definition and recognition. However, there are issues that need to be addressed if we want to re-use touch types and styles to provide consistency for end-users. For example, the meaning of a touch type can vary according to different applications such as two hands moving closely on a surface meaning *zoom out* in a GIS-based application and *gather scattered items* together in a game interface. This is important when we have users simultaneously using public and private display devices. Different users and/or different cultures may have different ways they operate and interpret a touch, calling in the possible re-mapping of touch types and their meanings for different users/cultures. In the context of this workshop, touch may need to be interpreted differently depending on the situation of private or public display.

We have established a model for touch interaction in order to allow a systematic approach in defining a touch and its meaning, and ultimately to allow re-use of touch for different applications, platforms, and use contexts. The model is comprised of three levels, the action, motivation, and computing levels. We describe each separately, define *mapping rules* and we apply the model to tabletops and to PDAs as public and private displays. By defining and classifying gestures to be used on touch platforms and suggesting a foundation for the mapping between a human gesture and the action that it causes, the model can serve as a useful guideline for designers of touch applications.

This paper is organised as follows. In the next section we summarise related work in categorising touch/gesture actions. We then describe our interaction model, components, their properties and relationships (mapping rules). In Section 4 we illustrate how this model can be applied and interpreted in the practical cases of public and private devices, and we conclude the paper with our perspective and future work.

2. RELATED WORK
While there are many studies on developing different kinds of novel Tabletop/PDA platforms and applications, there is little effort or organised activity on generalising or standardising touch interaction other than some definition of available gestures/touch for specific applications.

With recent advances in input sensing technology, researchers have begun to design freehand gestures on direct-touch surfaces. Yee *et al.* [1] augmented a tablet computer with a touch screen to enable hand and stylus interaction. Wobbrock *et al.* [2] presented a "$1 recognizer" to enable novice programmers to incorporate gestures into their user-interface prototypes. Rekimoto [3] described interactions using shape-based manipulation and finger tracking using the SmartSkin prototypes. Wu and Balakrishnan [4] presented multi-finger and whole handle gestural interaction techniques for multi-user tabletop displays. Morris *et al.* [5] presented multi-user gestural interactions for co-located groupware. Finally, Shruti et al. [6] explored the user's perceptions to a novel interaction method with mobile phones. They studied responses and reactions of participants towards gestures as a mode of input with the help of a low fidelity prototype of a camera mobile phone. The study used an approach inspired by participatory design to gauge the acceptance of gestures as an interaction mode. These are all very useful contributions to the touch/gesture interaction field, and more amount and variety of such work is required to advance it considering the early stage of our understanding in this area. However, a study that will be particularly useful at this point is a more generalised understanding of the kinds, types, and styles of touch interaction beyond the specific realisation of an application/device, and consequently how designers should map different touch to different functions.

Elias et al. [7] presented a multi-touch gesture dictionary which includes a plurality of entries, each corresponding to a particular chord. The dictionary entries can include a variety of motions associated with the chord and the meanings of gestures formed from the chord and the motions. The gesture dictionary may take the form of a dedicated computer application that may be used to look up the meaning of gestures. It may also take the form of a computer application that may be easily accessed from other applications. And it may also be used to assign user-selected meanings to gestures. Wu *et al.* [8] developed a set of design principles for building multi-hand gestures on touch surfaces in a systematic and extensible manner. They proposed the concepts of gesture "registration", "relaxation", and "reuse", allowing many gestures with a consistent interaction vocabulary to be constructed using different semantic definitions of the same touch. While this is in line with the direction of our work, we attempt to standardise and generalize the overall picture of the touch interaction where the user's intentions, touch actions, and their mapping to system functionality are understood and specified.

3. A TOUCH INTERACTION MODEL

We structured touch into three levels in our model, as illustrated in Figure 1. The first is the *action* level which is independent of applications or platforms, and only explains what touch types/styles are available (e.g. tapping with a finger or wiping with a palm). The second level is *motivation*, also independent of platforms but specific to applications. This level explains a user's motivation of what they want to do when interacting (e.g. annotate a photo or send an email). This level can be reused by different platforms if they have the same application domain. The third level is the *computing* level, including hardware and software. It is specific to platforms and applications, and links people's actions to functionality in order to react to perform a specific set of tasks. The three levels make up the structural layout in our touch interaction model. When we design a touch interactive interface, we only need to design touch at the action level once, and can reuse in other applications and platforms. Then we define different mapping rules from the action level to the motivation level according to the application domain.

3.1 Action Level

The action level describes various touch actions. We distinguish between two touch types: simple and complex touch. A simple touch may combine with others by, for example, another hand joining in or by the same hand doing multiple touches, in order to make up a complex touch.

We define a simple touch as the basic unit and as being a single hand action with no repetition and not containing other simple touch. We consider that there are two kinds of touch styles according to the contacting part between hand and

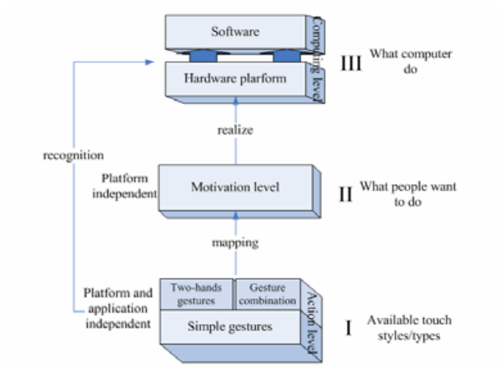


Figure 1: A touch interaction model

surface: single and multiple contacts. In a single contact the touch area between hand and surface is consecutive. In this style, touch can be with one finger, palm, half-palm (four closed fingers except thumb), fist and vertical hand, as illustrated in Figure 2. When touching with one finger, we don't distinguish which finger.

Figure 2: Single contact

In a multiple contact the touch area between hand and the surface is more than one. Touch with two fingers, three fingers, four fingers and five fingers are all included, as illustrated in Figure 3. We don't distinguish which combinations of fingers.

Figure 3: Multiple contacts

We now define the movement types as press, tap and drag, shown in Figure 4. Press means touching the surface and remain touched. Tap means touching the surface and lifting again rapidly. Drag means touching and then moving on the surface.

Press Tap Drag

Figure 4: Movement types

We describe a touch by combining touch styles and hand movements. For multiple contacts, movements are complex because each single contact can have its own direction and speed. However, the human hand has physical limitations, so possible movements are limited. These include one finger press while others tap, all fingers drag in two-direction, all fingers drag apart and all fingers drag in together. Simple movements such as press, tap and drag simultaneously are all possible for multiple contacts and these are summarised in Table 1 as a taxonomy. When we define a touch action, we choose a style and movement from here. For example, a case of two-finger touch is shown in Figure 5. For five fingers, Figure 6 shows possible gestures including the thumb pressing and the others tap, all fingers dragging in two directions, and all fingers dragging in together or apart.

Figure 5: Two finger gestures

Figure 6: Five finger gestures

Table 1: Touch taxonomy

Touch styles		Movement types	
Single contact	1 finger	Tap Press Drag (towards same direction)	
	Palm		
	Half-palm		
	Fist		
	Vertical hand		
Multiple contact	2 fingers		One press and the others tap Drag in bi-direction Drag apart Drag close
	3 fingers		
	4 fingers		
	5 fingers		

Complex touch is a combination of simple touches by adding another hand spatially or forming a sequence of simple touches temporally. Two handed action is usually symmetrical, however people are also used to fixing one hand and moving the other. It is difficult for people to do different actions with two hands synchronously. Most of the possible touch combinations are not easy for people to do but it provides choices in case an application requires a multitude of functionality distinctions with variety of finger touches.

3.2 Motivation Level

The motivation level addresses what people want to do and describes people's motivation according to the functions of the application. It is specific to a given application only and independent of platform. When an application is defined, all the motivations people can have when they interact with that particular application are confirmed as well. We take two examples below to illustrate this. In a map browsing application, people usually have the motivations of zooming in, zooming out, measuring distance, selecting a region, etc. (see Figure 7).

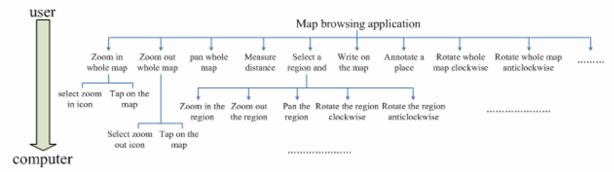


Figure 7: Motivations in a map browsing application

In a personal photo browsing application, people usually have

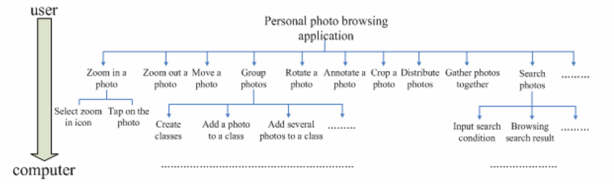


Figure 8: Motivations in a photo browsing application

the motivations of zooming in/out a photo, grouping photos, annotating photos, etc (see Figure 8). As this level is independent of platforms, it can be reused for different platforms. Some motivations need be further divided into sub-motivations in order to get any mapping to action level. For example, "zoom in whole map" can be divided into "press the zoom-in icon" followed by "tap the map" in a map browsing application. Thus, in a hierarchy of motivation, only the bottom-level motivations (leaf nodes) are assigned to a specific touch.

3.3 Computing Level

This level is concerned with how the device can detect and respond to a touch action. We divide this into two parts – hardware and software. A touch-enabled device reads the locations of touch points from the hardware and recognises its touch style and movement type according to these locations. The hardware provides runtime touch locations to the software. Although there are many different hardware platforms, algorithms that recognise a touch are in most cases the same. We use a toolkit of gesture recognition algorithms suitable for any platform. Such algorithms will handle the distinction between available and illegible gestures and ignore noises caused by the environment, inaccurate touch, or any other interferences that might occur during an interaction. Such work is out of the scope of this paper.

3.4 Mapping Rules

In order to complete the touch interaction design process, we need to define mapping rules between the action and motivation levels and also the recognition and realization algorithm between action, motivation and computing levels. We define general principles for mappings between the action and motivation levels as following:

- *Intuitive* - We have specific cognition in our real lives, for example shaking hands for friendship and nodding for agreement. This is similar for touch gestures so we can't define the mapping rules randomly and we should make them consistent with our intuition. For example, we usually map "two hands moving apart" to "zoom in the map" and "two hands moving close together" to "zoom out the map". If we swap these two around, it will be unintuitive.
- *Unambiguous* – When the mapping is done, there should not be misunderstandings either for human or for computers.
- *Minimal gesture as priority* – when assigning a touch to a motivation, simpler touch should be chosen if there are no other conflicts in the choice. That is, it is better for a user to accomplish the motivation by fewer steps or in a simpler way.

The above mapping rules should guide the designer in deciding which touch among the many listed in the action level taxonomy should be chosen for each of the motivation.

Sometimes there is a situation where we need to map the same touch gesture to several motivations. For example, people feel it convenient to pan a map by dragging with the index finger. However, we also feel it convenient to draw a path on a map by the same dragging action with the index finger. In this case, we need to make the computer register the same gesture as different motivations, possibly by setting different modes for the interface at the time of interaction. On the other hand, we may need to map several touches to one motivation, because different people or a same person at different situations might have different touch preferences. For example, some people like to use two fingers to rotate a map on a surface and sometimes they like to use three fingers or five fingers to do the same. Thus, between gestures and motivations there can be one-to-many as well as many-to-one mappings.

We introduce an *interactive context* in order to know which of more than one motivation should be mapped in response to a gesture. This can be considered as a mode or condition at the time of user interaction. For example, when we design a map browsing application, we can use a switch icon to distinguish between panning and path drawing. While the "drawing icon" is on (which the user can switch to "panning icon" if wished), a dragging gesture will register as a drawing action otherwise it will register as panning action.

Traditionally, we design windows, icons, menus and pointer (WIMP) to construct the interface to an application. This can be unnatural for us because we first have the motivation of click some icons or menus and then we need to know what will happen when we click. When we divide motivations into sub-motivations, the sub-motivations are usually exploring specific WIMP elements. We should try to reduce the WIMP elements to make the interaction and interface simple and clear.

Finally, we define the mapping rules from the action and motivation levels to the computing level using calls to the API. As we mentioned above, the computing level can recognise touch gestures, so what we need to do is to make the appropriate responses to each touch interaction.

4. APPLICATIONS

What we presented so far is a general model for touch interactions derived from an extensive set of observations of touch applications on the DiamondTouch, iPhone, and other touch devices. Now we apply the model to two different touch platforms – a public tabletop and a private PDA.

4.1 Tabletops

A tabletop such as the DiamondTouch [1] is usually designed for group decision-making and its touch interaction has several characteristics including a large public shared screen so that all group members can gather around and each can use two-handed gestures. A tabletop is usually arranged so people can sit or stand around it so gestural input from different directions should have the same meaning. Applications running on tabletops are quite wide ranging so gestural interactions should be comprehensive. We have examined many tabletop applications in our own lab including games, map browsing, photo browsing and multimedia search but we take map browsing application as an example to explain the model for tabletops. Users can pan, zoom in, zoom out rotate the map, measure the distance between two points, get the location of a point, draw or annotate, etc. According to the general functions of map browsing, we describe the major motivations in Figure 9.

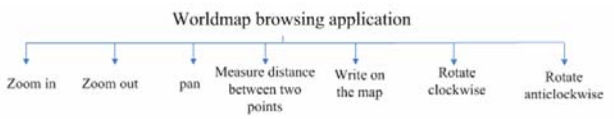


Figure 9: Motivations in a map browsing application

Some of these can be accomplished directly by a single touch gesture such as zooming in, zooming out, panning, rotating clockwise and anti-clockwise, but measuring distance and writing on the map should be divided into sub-motivations as in Figure 10.

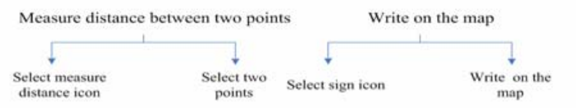


Figure 10: Map browsing sub-motivations

Once sub-motivations are defined the leaf nodes on the motivation hierarchy can be assigned to a particular touch gesture. Figure 11 illustrates our mapping decisions from touch gestures to motivations. The choice of touches on the left side was made based on the mapping rules, i.e. most intuitive to the motivation, unambiguous and simple. In this mapping we have a case where two touch gestures were mapped to one motivation (both 3rd and 4th gestures in the Action level pointing to the same "Pan the map" motivation in Figure 11). We also have a case where one touch gesture is mapped to two different motivations (3rd gesture in the Action level pointing to both "Pan the map" and "Sign on the map" motivations in Figure 11), requiring two different interactive contexts. For this we include switch icons so that when a user taps the pan icon, the gesture means panning and when he/she taps the sign icon, the same gesture will mean writing. More complex applications with more functionality can be designed and mapped in the similar way.

4.2 PDAs

PDAs are mainly designed for private applications such as personal contact, appointment, and entertainment. They have a small, private screen and the touch area is usually the display area. Users usually use one hand to hold the device and touch with the thumb of the holding hand and the fingers of the other. There are usually external inputs to the PDA, such as physical buttons.

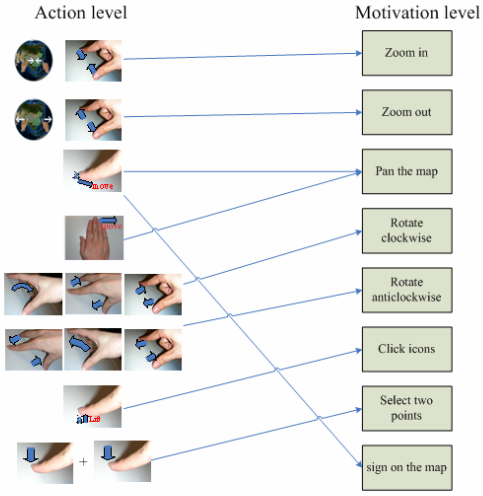


Figure 11: Mapping in a tabletop map browsing application

Because the screen is small, some touch styles between hands and PDAs can be unified. For example, we unify palm, half palm, fist and vertical hand as one style. At the same time, it is hard to distinguish between one-hand and two-hand gestures which use the same number of fingers and have the same movement types. Because users usually have one hand holding the PDA and only the thumb can move, two-handed gestures are limited. Thus in most cases the available touch gestures for PDAs are a subset of those for a tabletop.

We take a photo organiser application as an example to explain a model for PDA interactions. People can browse, resize, group and create classifications, search and rotate photos, etc. and we describe the motivations in Figure 12.

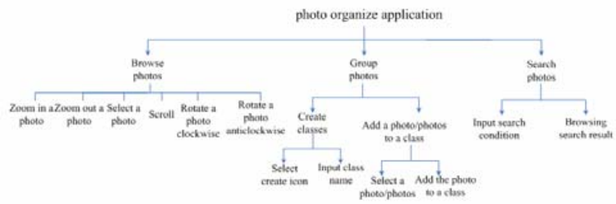


Figure 12: Motivations in a photo organiser application

We divided motivations until they are considered to be accomplished directly by a single touch gesture in Figure 12. Then we define the mapping between motivations and touch gestures in Figure 13. Here we have one touch gesture mapped to more than one motivation, such as one finger dragging (3rd gesture in the Action level in Figure 13). If there are photos selected, one finger dragging on top of any of the selected photos means adding selected photos to a group of photos; otherwise it means scrolling. As we know, there are often physical buttons on a PDA, so we can use these to indicate different interactive contexts. For example, if users press a button and drag on the screen, it means select photos; otherwise it means scrolling.

5. CONCLUSIONS AND FUTURE WORK

We have established an interaction model for touch interaction comprised of action, motivation and computing levels, in order to allow re-use of gestures and promote consistency in user interaction across applications and devices. By providing a set of available touch gestures and mapping rules to guide the designers in deciding which touch to map to which motivation, the touch interaction design for an application becomes more systematic. Tabletops and PDAs are private and public platforms which use multi-touch technologies. We discussed a particular touch interaction model for each of these according to their specific interactive characteristics and we described how touch gestures for the PDA are more limited than those for tabletops, effectively making the PDA touch gesture set a subset of the tabletop's.

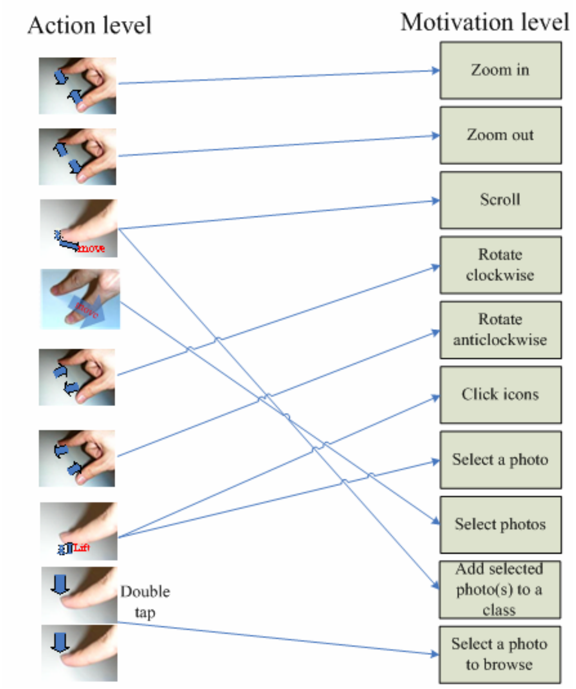


Figure 13: mapping rules for PDA interaction

In the future we hope to standardise each level and the mapping rules, and also the process of touch interactive interface design. We also plan to develop a number of combined applications for PDA, tabletop and touch wall with the touch gesture model and its mapping rules in mind from the start. Our ultimate aim is to build touch recognition middleware for all platforms, rather than retrofit as has been done heretofore.

6. ACKNOWLEDGMENTS

We thank Sinéad McGivney, Kirk Zhang, Guohua Zhang, Liang Bai, and Peng Wang for help. This work was supported by the Chinese Scholarship Council and by Science Foundation Ireland under grant 03/IN.3/I361.

7. REFERENCES

[1] Yee, K.-P. (2004) *Two-handed interaction on a tablet display*. Extended abstracts of ACM CHI. p. 1493-1496.

[2] Jacob O. Wobbrock, Andrew D. Wilson, Yang Li. (2007). *Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes*. ACM UIST. p. 159-168.

[3] Rekimoto, J. (2002). *SmartSkin: an infrastructure for freehand manipulation on interactive surfaces*. ACM CHI. p.113-120.

[4] Wu, M., & Balakrishnan, R. (2003). *Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays*. ACM UIST. p. 193-202.

[5] Morris, M. R., Huang, A., Paepcke, A., & Winograd, T. (2006). *Cooperative gestures: multi-user gestural interactions for co-located groupware*. ACM CHI. p. 1201-1210.

[6] Shruti, B., & Youn-Kyung, L. (2008). *Exploring gestural model of interaction with mobile phones. ACM CHI. p.2979-2984.*

[7] Elias, John, G., Westerman, Wayne, C., Haggerty, Myra Mary. (2007). *Multi-touch gesture directory*. United States Patent Application. 20070177803.

[8] Wu, M., Shen, C., Ryall, K., Forlines, C., & Balakrishnan, R. (2006). *Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces*. IEEE TableTop. p. 183-190.

Palm Interface: a display personally to show information in public spaces, by using image processing

Yoko Ishii, Minoru Kobayashi, Mutsuhiro Nakashige
NTT Cyber Solutions Laboratories
1-1 Hikarinooka Yokosuka-si,
Kanagawa 2390847 JAPAN
ishii.yoko@ntt-it.co.jp, minoru@acm.com,
nakashige.mutsuhiro@lab.ntt.co.jp

Hideki Koike
University of Electro-Communications, Tokyo
1-5-1 Chofugaoka Chofu-si,
Tokyo 1820021JAPAN
koike@acm.com

ABSTRACT
In this paper, we describe a system that allows a user to read information personally in public spaces without wearing any hardware devices. This paper presents the interface system ”InfoSnow”, which displays information on the palm of the user’s hand. This interface employs a snow fall metaphor to induce the user’s action to catch snowflakes of information, so that information display is focused on the user.

Categories and Subject Descriptors
H.5.2 [User Interfaces]: Interaction styles, H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities,

General Terms
Design, Verification.

Keywords
hand, interface, image processing.

1. INTRODUCTION
Various kinds of public information displays, such as posters and large screen displays are encountered in daily life. The advantage of these displays is that information can be distributed to many people at a time. However it is often the case that we do not perceive the information as being directed towards oneself, but rather to someone else. The display does not attract us thus often fails in properly delivering information. Reasons for this failure may be that, due to restrictions in the information presentation environment, we feel some physical distance from and a lack of control over the information.

Nowadays the use of mobile phones as information retrieving devices is popular. For example, we use our mobile phones to read email and to search for information on the web. Because the mobile phone is a personal device, information is retrieved upon the person’s own actions and perhaps better appreciated than that from public displays. On the other hand accessing information with a mobile phone can be troublesome; the phone must be taking out of a pocket or bag and many keys must be pressed to retrieve what might in the end be considered useless. If desire for the information is not strong enough, a person might not bother.

Given the above disadvantages of current information displays, the vision of our research is as follows. We aim to make an information presentation environment where:

- people can see and interact with digital information with the convenience of not having to wear or carry any hardware devices and

- people in public places can handle information in a personal manner and can easily feel as though the information is for him or herself.

This new form of information presentation could then be used for public space services such as wayfinding, information retrieval, message boards, personalized advertising, and so on.

Accordingly, we consider the implementation of our research vision as having three goals based on how people experience information in public displays. First, we aim for ambient-like design so that a person can be aware of information existing in a public space, but the information is not too aggressively presented. The purpose of this is that we strongly want to avoid creating an information system that is difficult or unattractive for a person to approach in order to interact with information. Second, we want to create an interface that allows a person to engage with information personally without the use of a device in public spaces occupied by many people. Finally, the person should not only just see, but also be able to interact with displayed information.

Figure 1. Snows InforSnow projecting information on a person’s palm..

The current embodiment of these goals is our new human interface system, InfoSnow. InfoSnow displays information on the palm of a person’s hand and allows him or her to interact with that information. Figure 1 shows a typical InfoSnow display. By presenting public information on the person’s palm, each person in the system’s space is able to individually see and experience one-on-one interaction with the information. The interface employs a snowfall metaphor to induce a person’s action to catch snowflakes of information allowing the information display to be focused on the person and be engaging. Snow, rather than rain, is our metaphor as it emphasizes that the information has some persistence and can be caught and seen as it lands on your hand. Each snowflake may be unique but is also disposable as there are other snowflakes to be captured.

2. InfoSnow

2.1 System Structure

The hardware components for this work include a projector, a video camera, and a PC. The image processing system locates the palm from the image captured by the video camera above the user's head. Then, an image is generated to fit the located palm, and this is projected onto the target by a projector also installed above the user (figure 2).

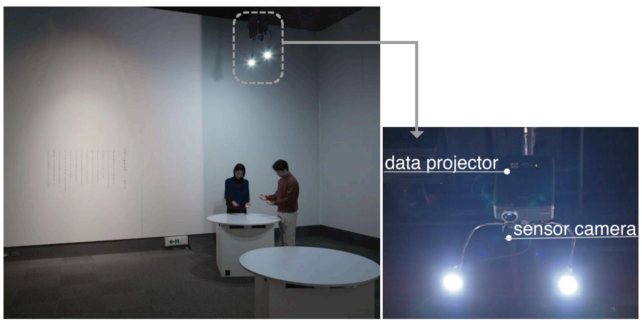


Figure 2. System Structure.

2.2 Interaction

InfoSnow creates a winter-like atmosphere by projecting snowflakes onto the floor. Animated snowflakes randomly appear and as they "fall" they gradually become smaller. Once a snowflake "hits" the ground it slowly disappears.

If a snowflake catches a person's attention and she holds out her palm, a snowflake image appears on the palm and thaws away to reveal another image (figure 3). This gesture of holding out the palm to receive information is called catching. When the system contains more than one image or piece of information to display, the image shown is chosen at random. Moreover, each person sees the image in a corrected direction and size; the orientation of the image changes according to the person's orientation and its size according to the size of the his or her palm.

Figure 3. Catching gesture: a snowflake lands on the palm and thaws to reveal a new image.

When the person closes her hand, the projected image disappears. This is called the grabbing gesture(figure 4).

Figure 4. Grabbing an image causes it to disappear.

When a person holds out both her left and right hands, images are projected onto both palms. By bringing both hands together, the two images can be combined into one associated image. For example, combining images of a flame and a star produces a sun (figure 5). This is called the "combining gesture".

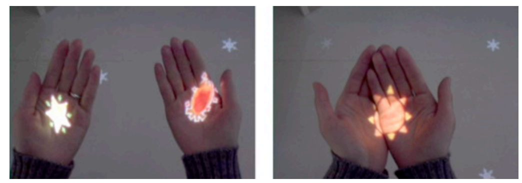


Figure 5. bringing images together combines and transforms them.

The above gestures allow for personal interaction and play with projected images so that one may feel closer to received information.

2.3 Palm Tracking Technology

To detect the position and sizes of multiple palms, the InfoSnow palm tracking software calculates the difference between sequential frames captured by the camera at 16fps. In the event that people are in the camera's view, the system will detect the areas around these people as regions holding human bodies. If a region's size exceeds a threshold value, the system attempts to split the region into two regions containing bodies. This process is iterated until no region exceeds the threshold value. At this point the center-of-mass is calculated for each region and this is taken as the position of a person's body.

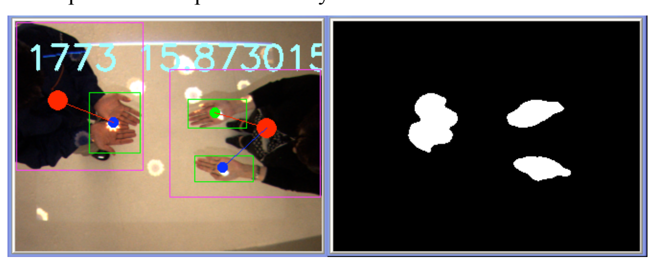


Figure 6. Image recognitnion results. Green square: palm orientation.

After body positions are calculated, skin color extraction is performed for each region containing a body to identify two possible palm candidates close to the center-of-mass. If the area of a palm candidate is larger than that of a threshold value the system interprets this area as including a person's arm. In this case the most distant point of the palm candidate from the body's center-of-mass can be assumed to be part of the palm. The system detects this point and identifies the area included in a circle of radius α around the point as the palm area. Finally, for each body candidate palm orientation is determined by a vector from the body's center-of-mass to the centers of the one or two palm areas. This is done to ensure that images are projected in a suitable orientation relative to the direction a user is facing. Figure 7 shows an example where the left"L" and right"R"hands orientations are different. Their orientation vectors are marked by lines extending from the center-of-mass(red circle).

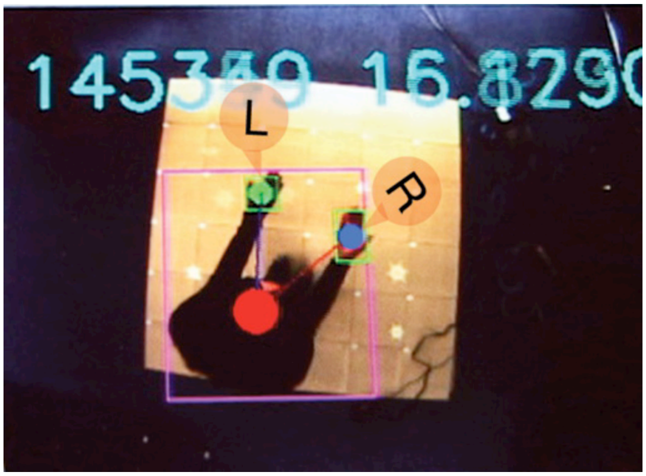


Figure 7. Image recognition results. Red circle: center-of-mass of a user's body, green circle: left palm orientation, blue circle: right palm orientation.

3. Technical Innovation of InfoSnow

As discussed in the following subsections, InfoSnow's technical innovation include: a new way to interact with large, public displays, a new robust palm tracking and video projection approach, and a new approach for combining display space and input space. Combined, these innovations provide an new interactive experience that is personal and intimate within a public space.

3.1 Peripheral Information Presentation

We aim to install InfoSnow in public spaces so that the manner with which it presents information blends into the characteristics of the space. Still, the system should have some kind of signal to induce and encourage a person's actions to get information. Until someone holds out his or her hand to receive information, InfoSnow only projects small falling snowflakes. Thus people can be aware of an ambient information space through a pleasing projection of snowflakes in their periphery, but they are not forced to view the information and the information does not consume the public space. InfoSnow provides a calming technology that supports interaction only when desired.

3.2 Robust Palm Tracking and Projection

The InfoSnow system recognizes both palms of multiple people in its space at the same time. Generally when there are many people in an area being processed for recognition, the most typical problem is occlusion. As discussed above, we use a threshold size for body-containing regions and region splitting to distinguish multiple bodies from each other. As a result, many people in the InfoSnow space can view information without frustration.

In order to avoid mixing up multiple projections of information onto palms, the system must also be able to discriminate between right and left palms and between the palms of different people. Discrimination is accomplished using the center-of-masses calculated for each body, as well as correspondences between consecutive frames. In order to display easily readable information to each person, the system must be able to detect his or her viewing direction. When capturing the image of a person with an overhead camera, the center-of-mass point is almost equal to the position of the person's head. Because of this we can assume a viewing direction based on the vector from the center-of-mass point to the center of the palm and project correctly orientated information.

Furthermore, two advantages of the InfoSnow system are that firstly, it can recognize the palms of a person even if he or she wears a short sleeved shirt and secondly, it employs a guessing algorithm to enhance its performance in following quickly moving bodies and palms.

3.3 Intutive Interaction Design with Hands in the Physical Environment

Interaction Using Hands

InfoSnow allows people to interact with and manipulate information using their hands. Interactive information control using gesture has been well researched[2], however usually the mediums for information presentation (display) and information control (hand gestures) are different[1][5]. An exception to this is research on flexible-displays and HMD[3][4]. In this research the information presentation medium (the palm) and the information control medium (hand gestures) are the same - the hand. Because of this, careful design is necessary to make use of the hand's best characteristics.

InfoSnow incorporates natural hand motions into the three gestures discussed above:

- catching gesture

  When an object that can fit in the hand is to be held, people naturally hold out their hand to catch it. In InfoSnow people naturally hold out their hands to catch falling snowflakes and they receive electronic information when they do.

- combining gesture

  If an object held in one hand appears interesting, there is a tendency to bring the object into both hands to manipulate and explore it more closely. InfoSnow takes advantage of this gesture to provide a function for combining information to provide new information upon exploration.

- grabbing gesture

  When an object is held, the fingers maybe closed over the object to hide or protect it. InfoSnow reacts to this gesture by making information held in the palm disappear.

PPD'08(AVI Workshops), May 31, 2008, Napoli, Italy.

Our gesture vocabulary at the moment is quite straightforward as we believe the affordances of the snowflakes metaphor will be sufficient for people to discover them through experience. As the vocabulary increases in complexity, we can use the display space to facilitate learning new means of interacting with the information space. Similar to marking menus, instructions could be displayed after a fixed delay so they help the novice, but do not interfere with proficient users.

4. InfoSnow Scenario

To depict the experience a person may have in a real world implementation of InfoSnow, we present a scenario of the system used in a grocery store. The system would be pre-loaded with various moving and still images of food items for sale. Catching these in the palm, a customer could combine food items using the combining gesture and receive suggestions for dishes that can be made from the food items. Combining dishes could reveal a menu accompanied with information on the amount of calories and nutritional balance. All of this would be capable without needing to use other devices.

5. Future Work

Presently InfoSnow can detect three users simultaneously and with ease. To accommodate a larger number of users we will increase the projection and image recognition area. We will also make improvements on the quality of image recognition and consider how the system may be adapted to suit various environments. Lastly, since InfoSnow allows one to view information in a private space, it is desirable to be able to customize this information. In order to do this, we would like to incorporate technology for personal authentication.

6. Conclusion

In this research we constructed a system, InfoSnow, that displays information on the palm of a person's hand upon their action to catch falling snowflakes. This is an agreeable manner in which people can actively take information from a public environment. Consequently, InfoSnow offers a new form for information presentation where people can interact with information personally and without the restrictions of other devices.

7. REFERENCES

[1] A. D. Wilson. "PlayAnywhere:A Compact Interactive Tabletop Projection-Vision System",UIST, pp.83-92. 2005

[2] Hardenberg, C. et al. "Bare-Hand Human-Computer Interaction", ACM PUI, pp.1-8., 2001.

[3] Lee, T. and Höllerer, T., "Handy AR: Markerless Inspection of Augmented Reality Objects Using Fingertip Tracking", I*IEEE ISWC*, 2007.

[4] Murakami, T., "Contact water", ACM SIGGRAPH, n-space gallery, 2001.

[5] Y. Sato, et al. "Real-time input of 3D pose and gestures of a user's hand and its applications for HCI", IEEE VR2001, pp. 79-86. 2001

Natural Gesture-Based Techniques for Sharing Documents from a Private to a Public Display

Md. Mahfuzur Rahman
Department of Computer Science
University of Manitoba
Winnipeg, Canada
+1 (204) 230-3851
mahfuz@cs.umanitoba.ca

Pourang Irani
Department of Computer Science
University of Manitoba
Winnipeg, Canada
+1 (204) 474-8995
irani@cs.umanitoba.ca

Peter Graham
Department of Computer Science
University of Manitoba
Winnipeg, Canada
+1 (204) 474-8837
pgraham@cs.umanitoba.ca

ABSTRACT

The proliferation of private devices and the more recent appearances of public displays has created numerous gaps in the manner in which users interact with these devices. In particular, we lack knowledge on how to seamlessly integrate and interface these devices such that users can share information from their private devices onto public displays, and vice versa. This paper is our first approach to investigating the types of paradigms necessary to fluidly allow users to exchange data, from their private devices to the public displays. We developed two gestural interaction techniques, Flicking and Chucking, based on natural human gestures that allow users to move objects from their private devices to the public display. We present a study showing that users required no training to operate the natural interaction metaphors. We want to continue our investigation by extending the new techniques to include multi-touch interactions in novel and interesting ways. This work will lead to guidelines for the design of next generation interfaces for sharing information between private and public displays.

Keywords

Flicking, Chucking, public-to-private sharing, MDEs.

1. Introduction

Private devices such as cell phones or personal digital assistants (PDAs) are ubiquitous and considered by some as extensions to our cognitive resources. More recently, we are witnessing the introduction of public displays in numerous environments, such as in schools, airports, museums, and shopping centres. Public displays are large in size and intended for use by several users simultaneously. The recent proliferation of public displays has led to the establishment of multi-display environments (MDEs), in which several private devices (or displays) can now interact and use information available on public displays. However, to date we know very little about how interfaces and interaction techniques should be designed for MDEs.

In this paper we discuss the element of sharing documents from private-to-public displays. We present two techniques, Flicking and Chucking for moving a document from a private device onto a public display. Based on our experience in implementing and evaluating these interactions, we intend on integrating multi-touch interactions in our proposed techniques for: (1) developing natural and seamless interaction techniques for moving documents from a private device to a public display, (2) designing interactions for controlling documents on public displays using the input mechanisms provided by private devices, and (3) interactions for moving documents back from public displays onto private devices. In a workshop setting, we hope to discuss the potential of this work for establishing a design framework and a set of design guidelines that will facilitate the future development of interaction techniques for sharing documents in MDEs.

Submitted to the AVI'08 Workshop on Private-Public Display Coupling

2. Background

The core related work to our investigation concerns device connectivity, gestural interactions, and object transfer in MDEs.

2.1 Device connectivity

Connecting with devices is a key component of the sharing paradigm in MDEs. Hinckley et al. [5] introduced 'stitching', a technique that allows users to connect pen-based devices using gestures that span multiple displays. 'Stitching' consists of a continuous stylus motion starting on one device and ending on the screen of another device. This helps users with mobile devices to collaborate with others and to share information with other persons. Pering et al. [12] introduce the Gesture Connect System which connects a user's personal mobile device with another device. The connection is established if both users are "shaking" their devices (cell-phones) with the same frequency. Swindells et al [14,15] evaluated gesturePen, a pointing-based technique which uses a gesture to select a device. They found that users took significantly longer to select a device from a list than using a simple pointing gesture [15]. Wilson et al. [17] introduced Bluetable, an interactive tabletop which allows users to establish the connection between mobile devices by simply placing the device on a table surface. Bluetable uses a combination of vision techniques and bluetooth, to determine the precise position of the device on the surface. We will adopt some of the similar mechanisms as these prior work, namely the possibility for seamless connection using bluetooth protocols [17].

2.2 Gestural interactions

Several systems have used human gestures, in particular Flicking for interactive systems. These are becoming common with the widespread use of devices such as Apple's iPod™. Flicking is analogous to a throwing motion in the real world. According to theories of naïve physics, the human perceptual and cognitive system simplifies occurrences of physical events that can lead to erroneous judgments about the estimated distance and trajectory traveled by an object [16]. Geiβler's throw technique [3] requires the user to make a short stroke over a document, in the direction opposite of the intended target, followed by a long stroke in the direction of the target. Wu et al. [18] describe a 'flick and catch' technique, in which an object is 'thrown' once it is dragged up to a certain speed. Kruger et al. [7] extend a rotation and translation technique to include Flicking for passing and moving items on a tabletop. Hinrichs et al. [6] utilize Flicking with a stylus to control the flow of documents moving on the periphery of a tabletop. Reetz et al. [13] demonstrate the benefits of Flicking as a method for passing documents over large surfaces. Flicking was designed to mimic the action of sliding documents over a table, and closely resembles the push-and-throw model designed by Hascoet [4]. Flicking was found to be much faster than other document passing techniques for tabletop systems [13]. While Flicking seems to have numerous advantages for various interactions, none of these systems have assessed the benefits of Flicking in MDEs or in particular for moving documents from one device onto a public display. Furthermore, other natural gestures such as 'Chucking', tilting or shaking have not been studied in the context of MDEs.

2.3 Object transfer in MDEs

To date very little work has been carried out with document transfers in MDEs. Nacenta et al. [10,11] developed the Perspective Cursor to view documents across MDEs. Drag-and-pop is a technique for copying objects on a public display, using object proxies [1]. Maunder et al [9] designed SnapAndGrab a technique to share between private and public displays using camera and photo processing techniques. None of these prior systems have explored the benefits of natural human gestures for cross display object transfer.

3. Flicking and Chucking

We designed two techniques, for moving documents from the PDA onto a public display. We developed an implementation of the well used Flicking [6,13,18] and developed a one-handed interaction technique, we refer to as Chucking. In Flicking, the user puts the pen down on the document, drags the pen towards the desired direction and then releases the pen to send a document (Figure 1). Flicking can be easily adopted but requires the use of two hands.

Studies show that one-handed use is the preferred method for operating a handheld device [Karlson et al]. In one-handed interactions, the user commonly grips onto the device and interacts using the thumb or other auxiliary fingers. While this mode of operation works well for interacting with buttons on a cell-phone for example, it does not work conveniently for touch-input. The main reason is that the distance covered by the thumb is not sufficient to manipulate objects in the extreme and opposite corners of the device

Chucking is a one-handed document sharing interaction. In Chucking, the user "gestures" the device as in Chucking cards on a table (not throwing). Active documents on the private device get transferred onto the public display. The motivation behind each of these techniques was to provide the user with a natural interaction to perform the sharing. Chucking is performed by means of a TiltControl™, a small device that we attach to a Pocket PC (can also be attached to cell-phone or other mobile devices). The TiltControl is able to detect the precise angle that the mobile device is being held at, and communicate this information with its host mobile device. Thus it allows to control any application on the device with motion e.g. turn the device into a wireless mouse, measure vehicle performance, automatically change screen orientation depending on the rotation and movement of the TiltControl.

Figure 1. Flicking – from a steady state the user flicks the active object on the PDA onto the large display.

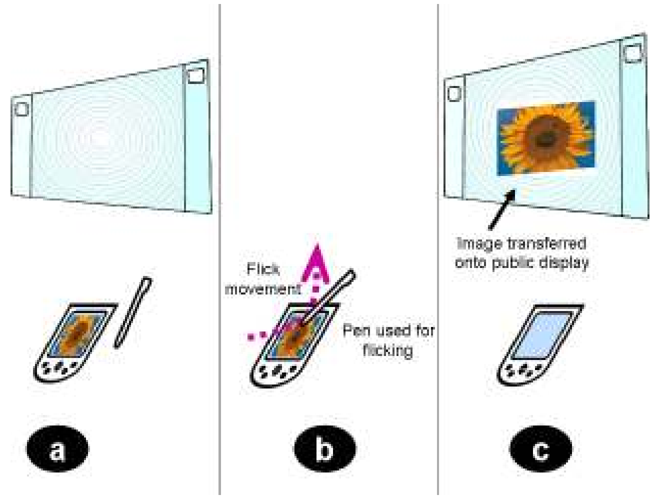
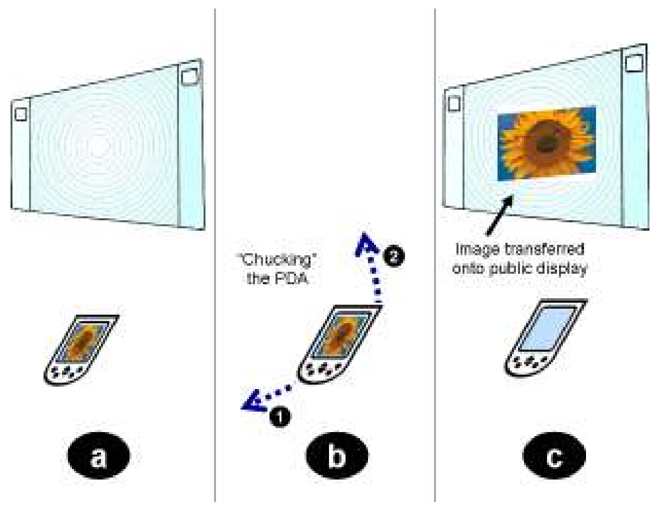


Figure 2. Chucking – with one fluid back-to-front, analogous to dealing cards like a black jack dealer, the user can move the active object from private to public display.



We use the parameters of the TiltControl to identify a Chucking motion. When the adequate gesture is invoked, Chucking communicates with the public display and sends the object. In the process of Chucking, the user can also tilt the device in the direction or position they wish to move the object to. In this manner, it allows control of position and sharing in one smooth and fluid movement (Figure 2).

The implementation of Chucking is multifaceted and therefore we briefly explain some of its details. The implementation of Chucking consists of an application running on the mobile device and continuously polling the tilt readings (horizontal and vertical angle) from the attached TiltControl. To identify and accurately capture the gesture motion, we measure at frequent intervals the changes in vertical angles. We primarily use the direction and position of the changes with respect to the horizontal and vertical angles. When the desire angular movement is identified Chucking transfers the object to the public display.

3.1 Bluetooth connectivity

Sharing and connectivity in our system was based on Bluetooth pairing, similar to that used in [17]. We developed a simple protocol to assist in the document transfer. The Bluetooth connectivity imposed the significant bottleneck in our implementation and therefore several iterations were necessary to make the interaction seem natural.

3.2 Prototype

We designed a prototypical application to determined user satisfaction and ease of use with each of the two techniques. We designed a photo sharing prototype in C#.Net. In the application, user can select various images to make them active on the PDA. With the Chucking metaphor users can flip through images simply by rapidly tilting the device in a given direction. With Flicking, the user is given control buttons to iterate through the photo album. Our public display is simulated with a large projector display onto a wall. The application accepts images in any of three pre-defined locations. With he private PDA, the user can flick or chuck an image in any one of the three locations.

Figure 3. Public display consisting of a projected application onto a wall. The public display can accept documents in various locations.

4. Evaluation

One question that bears some attention concerns how users perceive these various techniques as natural and how they interact with these. Since the ability to position was developed in our system, we choose object positioning as our experimental task. Chucking appears to be a natural form of interacting, we therefore hypothesized that participants will subjectively prefer Chucking to Flicking. However, we also hypothesized the Flicking will be more accurate than the Chucking metaphor.

4.1 Task and Stimuli

Participants were asked to send a list of images to any of the three locations of the public display from the PDA. The user could flip through the list of images to assign an active image for sharing. An image would appear on the public display in one of three positions (see Figure 3). The user would have to select that image on the device (with Flicking they would simply scroll and with Chucking they would tilt to select an active image) and send it to the public display in the location specified by the experimental system. We measured the accuracy at which the user was able to place the document during the transfer.

4.2 Apparatus

The experiment was conducted using a Dell AximX30 PDA with a TiltControl. The public display consisted of a workstation attached to a multimedia projector display. The workstation ran in single-user mode with its bluetooth device on and disconnected from all other network traffic.

4.3 Participants

Six right-handed volunteers, students in Computer Science department, participated in the experiment.

4.4 Design

All participants performed the experiment using both techniques. The presentation order of the two techniques was counterbalanced across participants. For each technique, the participant placed the image n one of 3 positions. For a particular position or target, each participant performed a total of 30 trials. Participants were given eight practice trials to familiarize themselves with the task. The experiment consisted of total 1080 trials as follows:

6 participants x 2 techniques x 3 galleries for each technique x 30 trials per gallery = 1080 trials.

For each subject, the experiment was conducted in one sitting. Subjects were alternatively assigned to one of two experimental orders: Flicking technique followed by Chucking or Chucking first. A short questionnaire designed to elicit participants' subjective preferences for the two techniques was completed by participants at the end of the experiment.

5. Results

We were primarily interested in the user experience with both of these techniques. Since completion times would be dominated by the Bluetooth connectivity module we only used subject preferences and accuracy as measures.

5.1 Accuracy Measurement

Table 1 contains the average success rate for accurately positioning a document on the public display with both Flicking and Chucking for all participants.

Table 1. Average Success (no. of times) for locating galleries

Technique	Direction	Average Success
Flicking	Left	27
	Middle	29
	Right	26
Chucking	Left	25
	Middle	28
	Right	24

We observe that participants are overall slightly more accurate with Flicking than with Chucking. There are two primary reasons for this performance. The first involves the angular movement of Chucking and the recognition of this gesture by our system. Where as we are able to fluidly "chuck" objects in the physical world, such action cannot be easily discretized into its elemental aspects and replicated in virtual environments. The second reason for slightly poorer accuracy with Chucking resulted from the inability to see objects on the screen once it was being titled in various directions. This led participants to send the wrong object to the public display. An improved system would necessitate mechanisms for improving selection of objects and sharing with tilt such that the gesture matches our real-world actions and that the user is not inhibited in the process.

5.2 Preferences

At the end of the experiment, participants were asked to rate their preference for each technique on a scale of -2 (very low) to 2 (very high). The results summarized in Table 2 validate our second hypothesis and is consistent with the Accuracy measurement.

Table 2. Each cell contains participants' preferences. -2 is a very low acceptance, and 2 very high. 0 is neutral.

	- 2	-1	0	1	2
Flicking		*1*		*1*	*4*
Chucking			*1*	*3*	*2*

From the experimental result, we have found that Flicking's performance is a slightly better than Chucking. Since Chucking also requires a gesture that may not work entirely the same as its physical counterpart, the participants required time to get acquainted with the technique. Some participants reported that because of the gesture in Chucking, they felt tired in their wrist at the end of the trials. These problems were not present in Flicking as it is fairly flexible in accepting a large range of movements from the participants.

6. Multi-touch input

Our initial implementation has focused on only one aspect of the private-to-public coupling, i.e. natural interactions to share objects. We believe that integrating multi-touch interactions with our system will improve our systems in several ways.

Control of document position: using a multi-touch mechanism users can place their finger in one of several locations before performing the flick or the chuck. Since Chucking does not require fingers in its core interaction, this technique would require only one touch. However Flicking would necessitate multi-touch capability.

Control of document orientation: in our current scenario we concentrated primarily on a vertical public display. However Chucking and Flicking could be utilized in horizontal displays, which would necessitate appropriate object orientation. Multi-touch input could be sued to specify the orientation for the document before it even gets placed on the public display.

Figure 4.a depicts the position of a finger (thumb for example) to position the document in one of three locations. While Figure 4.b depicts the position of the thumb again for orienting the object on the public display. We observe that multi-touch is not a requirement when the technique takes advantage of viewing the device as an additional input mechanism (as in the case of Chucking). However multi-touch is necessary for the case where already the input mechanism is dedicated to an action, in the case of Flicking for example.

7. Conclusion

It is very important to design natural interfaces and interaction techniques for MDEs. Chucking adopts one handed interaction which seems to be more natural for sharing documents from a private display to a public display. Our experiment and informal study have shown that Flicking and Chucking seem adequate for placing and sharing documents on public displays. Future work and workshop discussion will revolve around successful methods for integrating multi-touch input to extend that capabilities of Chucking and Flicking.

Figure 4. Chucking and Flicking benefit touch input modes, for (a) specifying the position of the object on the public display, and for (b) identifying the orientation of the object.

8. References

1. Baudisch, P., Cutrell, E., Robbins, D., Czerwinski, M., Tandler, P., Bederson, B., and Zierlinger, A. (2003). Drag-and-pop and drag-and-pick: techniques for accessing remote screen content on touch- and pen-operated systems. *In proceedings of Interact'03*, 57-64.
2. Buxton, B. (2007) Sketching User Experiences: Getting the Design Right and the Right Design, California: Morgan Kauffman, pp. 443.
3. Geißler, J. (1998) Shuffle, throw or take it! working efficiently with an interactive wall. *In the ACM Proceedings of CHI'98 Extended Abstracts*, 265-266.
4. Hascoët, M. (2003) Throwing models for large displays, *In the Proceedings of British HCI'2003*, 73-77.
5. Hinckley, K., Ramos, G., Guimbretière, F., Baudisch, P., Smith, M. (2004) Stitching: pen gestures that span multiple displays. . *In the ACM Proceedings on Advanced Visual Interfaces (AVI'04)*, 23-31.
6. Hinrichs, U., Carpendale, S., and Scott, S. D. (2006) Evaluating the effects of fluid interface components on tabletop collaboration. *In the ACM Proceedings on Advanced Visual Interfaces (AVI '06)*, 27-34.
7. Kruger, R., Carpendale, S., Scott, S. D., and Tang, A. (2005) Fluid integration of rotation and translation. *In the ACM Proceedings of CHI '05*, 601-610.
8. Karlson, A., Bederson, B. , and SanGiovanni, J. (2005). Applens and launchtile: two designs for one-handed thumb use on small devices. In Proc. CHI'05, pages 201–210.
9. Maunder, A., Marsden, G., Harper, R. (2008) SnapAndGrab: accessing and sharing contextual multi-media content using bluetooth enabled camera phones and large situated displays. CHI Extended Abstracts, pp. 2319-2324.
10. Nacenta, M., Sallam, S., Champoux, B., Subramanian, S., Gutwin, C. (2006) Perspective cursor: perspective-based interaction for multi-display environments. *In the ACM Proceedings of CHI'06,* 289-298.
11. Nacenta, M., Sakurai, S., Yamaguchi, T., Miki, Y., Itoh, Y., Kitamura, Y., Subramanian, S., Gutwin, C. (2007) E-conic: a perspective-aware interface for multi-display environments. *In the ACM Proceedings of UIST'07,* 279-288.
12. Pering, T., Anokwa, Y., Want, R. (2007) Gesture connect: facilitating tangible interaction with a flick of the wrist. *In Proceedings of Tangible and Embedded Interaction'07*, 259-262.
13. Reetz, A., Gutwin, C., Stach, T., Nacenta, M., and Subramanian, S. (2006) Superflick: a natural and efficient technique for long-distance object placement on digital tables. *In the ACM Proceeding of Graphics Interface'06*, 163-170.
14. Swindells, C., Inkpen, K. (2002) What is that?: gesturing to determine device identity. *In the ACM Proceedings of CHI'02 Extended Abstracts*, 588-589.
15. Swindells, C., Inkpen, K., Dill, J. Tory, M. (2002) That one there! Pointing to establish device identity. *In the ACM Proceedings of UIST'02*, 151-160.
16. Weld, D., and de Kleer, J. (1989) Readings in Qualitative Reasoning about Physical Systems. San Francisco, California: Morgan Kaufmann.
17. Wilson, A., Sarin, R. (2007) BlueTable: connecting wireless mobile devices on interactive surfaces using vision-based handshaking. *In the ACM Proceedings of Graphics Interface'07*, 119-125.
18. Wu, M. and Balakrishnan, R. (2003) Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. *In the ACM Proceedings of UIST '03*, 193-202.

Analysing Fluid Interaction across Multiple Displays

Richard Morris, Paul Marshall and Yvonne Rogers
Open University
Milton Keynes, UK

ABSTRACT

Interaction with groups carrying out tasks across multiple displays and devices can be complex. Users have to switch their attention from controlling one device to another while continuing with their ongoing activity and conversations. This raises questions about how to support and evaluate interface design which facilitate fluid interaction. This paper provides a nascent framework of fluidity as a way of analysing interactions across multiple displays and tasks. Three fluidity heuristics are outlined illustrating how they can be used to aid the design and evaluation of interactions with multi-display systems.

1. INTRODUCTION

Shareable and personal devices are providing designers with new opportunities for creating a wide range of rich technology-augmented spaces that can support collaborative working, learning or playing. However, there are significant challenges in doing so: infrastructure and interfaces must be developed to share information, representations and interactions across an increasingly diverse ecology of devices. Furthermore, this diversification leads to a combinatorial explosion of factors that the designer must take into account when developing such a system for a user group, task or context. Such factors include the number of devices available to the users; what kinds of information should be shared and what should be private; what mechanism or metaphor should be used to move information between devices; and in what orientation should shared displays be placed. As pointed out by Tan *et al*. [6] there is a dearth of evaluation methods, tasks and metrics that could be used in evaluating multi-device collaborative environments.

A key problem is managing the flow of work between displays, be they personal/small or shared/large displays, specifically how one addresses the other displays, and transfers work, from the one currently in use. Will they be controlled through gestures (if touch-enabled) or menus? Will animation help in reducing the cognitive overhead of switching between screens? How will the users be given feedback or retrieve their work if something goes wrong? Our research seeks to help designers address these questions by providing conceptual tools of analysis.

2. BACKGROUND

Fluidity is a concept that is increasingly being used to describe a desired state for new forms of interaction. This would be manifest in ways such as users being able to move smoothly between displays, devices and tasks without having to exert too much cognitive effort. In particular, users should not have to constantly switch their attention between control operations and the goals of the task. The aim is to enable a group's actions and interactions with a system to be invisible (cognitively), ordinary and to flow smoothly. While this is an important goal, the concept has yet to be operationalized so that it is possible to assess the fluidity of the diversity of interactions when using multiple displays.

Fluidity has been used to describe the various transitions that are needed to enable collaboration [7] and the obstacles that can hinder interactions, such as dialog boxes popping up [1] and as Isenberg *et al*. [3] have noted that these guidelines can be expressed in the positive sense of supporting high-level cognitive aspects of a task without forcing the user to deal with low-level objects. The benefit of such fluidity of interaction is that users can bring more of their attention and creativity to bear on their ultimate goals, or other demands such as collaboration, leading to more productivity and higher quality work.

One approach to fluid interface design is in terms of reality-based interaction [4]. This seeks to model real-world themes and to reduce the gap between a user's goals and the means of execution. The real-world themes are naïve physics, body awareness, environmental awareness and social awareness. By designing interfaces, based on the rules of these dynamics, the need for low-level operational expertise is reduced, affording the user the opportunity to focus on higher-order goals and more focused creativity. Also, it should be easier for users to return to where they were previously when interrupted, as the cognitive effort of getting back into the framework of the interaction is reduced. This also affords the benefit of encouraging reflection and viewing the bigger picture for a fresh perspective or learning. As these interfaces provide more natural interaction it is also hypothesised that they will lead to better social interaction when working in groups.

It follows that multiple display and device systems should not be unnecessarily complicated, and should employ reality-based interaction where possible, except where certain explicit trade-offs are made to add further functionality. Jacob uses the analogy of the character Superman: when he is performing simple tasks he walks and talks like a regular human, but when the situation requires it he uses his powers to increase his efficiency in completing his task.

The concept of fluidity is appropriate for analysing the complex development of multi-user, multi-device interactions. One challenge is to provide a way for users to get the most out of the technology at novice and expert levels. Too little help or signposting and the novice cannot engage with the system: too much and the expert user becomes frustrated. Guimbretière argues that dialog boxes, tool selections, object handles etc. are "inevitable to provide complex functionality" [1, pg. 3]. His FlowMenu [2] gives visual feedback without permanent menu bars or palettes by using a pen-addressed radial layout menu, which encircles the pointer whenever the menu is summoned but also allows experts to use gestural memory without feedback.

However, collaboration is not governed solely by the quality of the interaction that the user has with the interface but also the interactions between the user and others, and other users and the interface. A successful collaborative task may depend on the ability of individuals to work singly in personal spaces while carefully choosing their interactions with the other users at various stages. Given the intricacy of group interactions, another challenge is to design computer interfaces which can support them while being simple enough to use that all group members can contribute effectively.

3. FLUIDITY HEURISTICS

Below we propose three heuristics that can be used to analyse how systems of multiple displays and devices are able to support users in achieving their task goals. These are ready-presence ratio, cognitive focus maps and interaction matrices.

3.1 Ready-presence Ratio

The first heuristic, *ready-presence ratio*, is based on the idea of measuring interactions when moving between subjective states of involvement: our starting point is Heidegger's well known concepts of readiness-to-hand and presence-at-hand (see also [8]). The canonical example of using a concrete tool such as a hammer exemplifies what it means to switch between 'present-at-hand' and 'ready-to-hand' depending on the user's awareness of the hammer. When hammering away at a nail one is often not aware of the hammer as being distinct from one's own arm and hand or part of our 'totality of involvements'. The tool becomes an extension of ourselves in the expression of our task. In this state the hammer is ready-to-hand. However, should the hammer break or hit our thumb we would become aware of the interruption to our task and the hammer would become present-at-hand.

In terms of user interactions, we employ this idea to conceptualise when a user is interrupted in the flow of completing their task. *Higher-order* user actions are those directly related to dealing creatively with a task; those which are directed at dealing with the state of the computer are *lower-order*. Expressed as a ratio of higher- to lower-order action, fluidity is essentially the property of being in a higher cognitive state and focused on the task, not the tool. Thus:

$$\textit{fluidity} = \frac{\textit{higher-order} - \textit{lower-order}}{\textit{total operations}}$$

The key feature of fluidity is that it is a measure of the proportion of task-specific actions and cognition. For example, if a user is to draw a circle and label it with text, they might perform 15 operations dealing with low level aspects of the machine such as opening the program, selecting the appropriate view and palette, selecting the right tool, and changing to the text tool, and the operations which are related to the higher-order goal such as drawing the circle or typing the text would amount to two. This would give a fluidity score of F=-0.77 (2-15)/17).

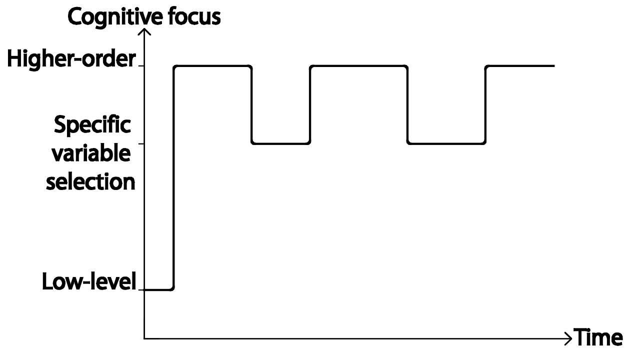
Compare this to performing a similar task on a drawing surface such as Guimbretière's PostBrainstom interface [1]. The lower-order task would be picking up the pen, but drawing the circle and writing the text would be done directly as two higher-order goal-centred operations, giving a fluidity score of F=0.33. Compared to the previous example the fluidity score F is large, and in a more positive direction, indicating that it leads to a more fluid interaction.

As well as comparing across interfaces, this heuristic is also intended to be applied across experience levels. Supposing that a new interface is highly reality-based then experience level should have less of an effect on the F score. Any difference in F could indicate that experienced users are employing shortcuts, which could indicate an area for further study.

When defining and analysing fluid human-computer interactions, therefore, it is important to take into account the users' level of expertise with the task and the technology. It may be possible to design interfaces that are fluid to use by experts for a task but not for novices (e.g., a games console). There is a distinction also between expertise at lower and higher levels of action. For example, being an expert typist may not automatically confer an advantage to a player in a strategy game if they are not also expert at the higher-level goals and conventions of the game. Conversely, an expert tennis player might be at a disadvantage in a game of Wii Tennis against someone who has more expertise in using the WiiMote controller.

3.2 Cognitive Focus Maps

The second heuristic, *cognitive focus maps,* graphically project *cognitive focus* over time in an interaction. Figure 1 (top) shows an example of how an experienced user might interact with a complicated application like AutoCAD. After launching the application the user can begin outlining whilst in a high-order cognitive state and considering their design goals. Next the user has to specify a certain variable and a specific dialogue must be sought where the user can input a variable e.g. wall thickness, or material type. Because the user is experienced and knows what to expect they can interact smoothly and without feedback or cogitation. Like Jacob's Superman the architect must make a small but useful interruption to their flow to make an explicit input.



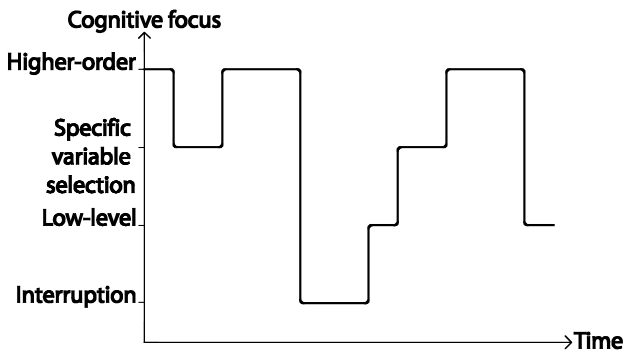


Figure 1. *Cognitive focus* over time in an interaction for (top) an experienced user and (bottom) during an interruption.

Figure 1 (bottom) describes a difference scenario where an individual is sharing photos with someone else using a tabletop display such as a Microsoft Surface with an inturruption in the middle of the task. The figure is intended to highlight the difference between the users' experience of interacting with the table at times when low-level objects must be dealt with, such as waiting for data transfer or resuming the machine after it goes into standby during the interruption, and being able to operate on the higher-order goals of the task such as the actual photo sharing and discussion.

Following the interruption and resuming the machine from its standby state, a short period of time is spent by both users looking back over the photos in the stack. This is an example of how the user experience can be 'buffered' when moving back into an interaction, whereby remembering the state of the interface before the interruption and the position of photos relative to each other can aid the users' memories and help in resuming the conversational thread. This could be enhanced further by, for example, replaying recorded audio from before the interruption to assist recollection.

3.3 Interaction Matrices

Our third heuristic, *interaction matrices*, describes the interactions between groups of users with various interfaces. Supporting a collaborative design task requires the ability to move from working one-on-one with the computer, to social interaction, and multi-user interaction with the interface. In this context, fluidity impacts on the quality of an interaction that extends beyond the user-interface, as the properties of interaction 'inside the interface' can have an effect on social interactions 'outside', collaboration and the flow of ideas. Thus a user who is experiencing a fluid interaction with an interface will find it easier to take part in the social level of interaction, theoretically leading to better collaboration.

Figure 2 depicts several modes of interaction using a short-hand notation, or interaction matrix, taking the form {('outside' interactions):(interface interactions)}. Situation 'A' is the simplest: one user and one interface are having one interaction {1:1}. In 'B' there are three users all interacting with both the interface and each other. The dotted lines on the interface are meant to denote that there are different ways to divide the work area. All three users could be sharing the one interface together {(3*3):1} or they could be working in separate spaces and sharing between each others' spaces {(3*3):(3*3)}, or simply working on their private spaces alone {(3*3):(1*3)}. In 'C' the users are interacting with each other but one user is mainly interacting with the interface.

Situation 'D' is a special situation where an expert user is interacting with the interface in a way the other group cannot and the output of this interaction is used by the group {(3*3):1:1}, such as when using a facilitator.

The interaction matrices can be used to describe how different user / interface combinations can lead to different design goals and expectations about fluidity. By separating the interaction matrices inside and outside the interface a clearer understanding can be reached of the true nature of interaction occurring. All these situations have different modes of interaction, but a fluid interaction between the user and the interface always benefits the entire goal, whether the user is in a group, alone, novice or expert. In 'D' the user is required to be highly expert as creating real-time visualisations of discussions is a complicated task. However, in 'B' simpler interface actions should be used to ensure all users have a similar level of control. Also, the interface should avoid dialog boxes, as it may be unclear which user it corresponds to. In 'A' the user can be novice or expert, depending on their level of experience and the necessity for complex 'superpower' operations. 'C' is in-between as the main user can fall on a range of expertise but other users may wish to input directly.

4. USING THE HEURISTICS

Our fluidity heuristics are intended to assist both in the design and evaluation of interfaces and the various types of interactions, and group modes, by expressing different aspects of the fluidity of these interactions. The ready-presence ratio is intended to focus the designer on the way a user experiences readiness-to-hand, when focused on the higher-order goals of the task, and presence-at-hand – seeing the user and the tool (interface) separately. This heuristic can be used in tandem with the guidelines produced by other authors (e.g. [1],[5]) to assist understanding of users' shifts in conscious awareness at key points. It assists in evaluation of the overall interaction quality and in comparing across interfaces or user experience levels.

The cognitive focus map can help in highlighting the transitions between users' states of awareness and 'presence' in the interaction, to help identify key areas in the design of the interface to enhance the user experience. The area under the graph also gives an evaluative indication of the overall fluidity of the interface, where a larger area indicates greater time spent in goal-focused states of mind. By adjusting for the total length of time of the interaction, it could be possible to analyse interactions in a way which is less skewed by experience level, in terms of dealing with dialog boxes etc., than the ready-presence ratio.

The interaction matrices heuristic can be useful in designing an interface by highlighting the ways that groups and single users can interact with it and with each other. By separating the interactions inside and outside of the interface it can be seen where design goals, such as removing visual clutter, will be most effective. It also provides a shorthand way of expressing specific interaction modes to help facilitate discussion and evaluation.

To illustrate how these heuristics can be used together to analyse how fluid the interactions are for users moving between displays consider the scenario of how scheduling work meetings could be enhanced through having a system of shared and personal displays. People in organisations use shared software calendars to arrange projects, meetings and schedules of work. However, it can be very time consuming to arrange a meeting, especially when it depends on email response. If a shared calendar application was made available whereby a large touchscreen could display an overall work schedule (i.e. a Gantt chart), representatives from each team could work either on the overview schedule or on small tablet or handheld devices to make fine-scale adjustments or to rearrange outside commitments around the emerging work schedule. The application could be analysed by using the three heuristics above. The interaction matrices would help in describing the different permutations of interaction possible in this arrangement, i.e. whether the users are all interacting with the large screen, their small screens or any combination between. This could assist a designer focus their methods for moving

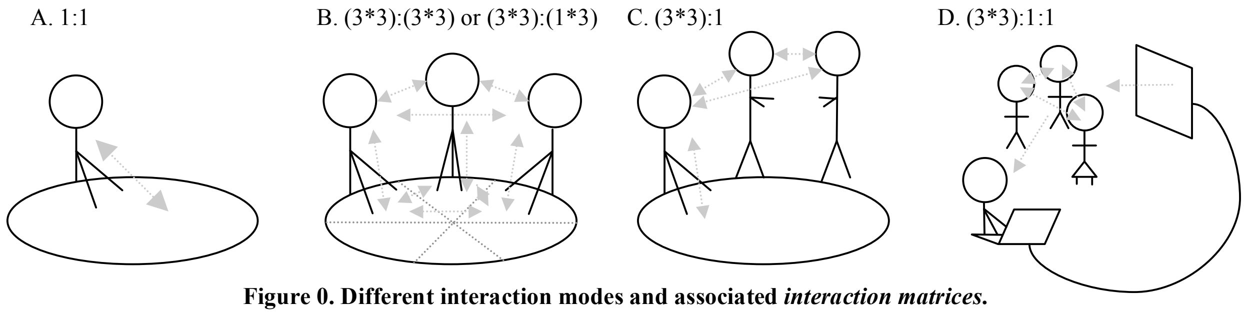


Figure 0. Different interaction modes and associated *interaction matrices*.

data between screens at the most appropriate times.

The fluidity of the interaction could be assessed for each individual user using the ready-presence ratio. This would give an impression of how different styles of interface would support or hinder fluid interaction for any given situation. For example, when working on a small personal screen the user may have to make more low-level actions due to the size constraint of the interface, but this may lead to more rapid progression of the overall goal of organisation on the main chart.

The cognitive focus maps can be used to analyse the interaction over time and to bring attention to key moments, such as when a user switches between working at the big screen to their individual screen, or to help design ways for users to collaborate or resume work after an interruption. Explicitly considering where the user is focusing their attention at certain points can help the interface designer support key actions.

One problem which may arise when collaboratively creating schedules is that a clash may arise. Being able to work on their own sub-schedules individually, the team members involved can work in parallel to make fine adjustments and compromise to make the overall schedule work, and this could be expressed in an interaction matrix. Key points in this interaction would be the identifying of the clash on the main screen. Then the users would have to use the interface to edit their schedules individually and then return their change to the main schedule. How this is accomplished through interface design choices can be readily assessed using the ready-presence ratio and cognitive focus maps. Experimental studies could then be performed on different interface prototypes to evaluate their fluidity.

5. SUMMARY

We propose that in order for groups to effectively utilise multiple displays by switching work between screens, interfaces and interaction styles and be able to do so without interrupting the flow of their ongoing tasks, the interactions have to be fluid. However, fluidity can be a nebulous term that is difficult to define. In this paper we propose three heuristics intended to aid in the analysis of interface and task interactions, which can provide an indication of fluidity and clarify the processes involved. In so doing, they can highlight how to design for users so they can easily transition between multiple interfaces, tasks and conversation whilst keeping their creative thoughts and expressions ‘flowing’.

6. ACKNOWLEDGMENTS

This work is supported by the EPSRC who are funding the ShareIT project (shareitproject.org).

7. REFERENCES

[1] Guimbretiere, F. V. 2002. *Fluid Interaction for High Resolution Wall-Size Displays*. Doctoral Thesis. UMI Order Number: AAI3040022. Stanford University.

[2] Guimbretiére, F. and Winograd, T. 2000. FlowMenu: combining command, text, and data entry. In *Proc. of the 13th Annual ACM Symposium on User interface Software and Technology* (San Diego, California, United States, November 06 - 08, 2000). UIST '00. ACM, New York, NY, 213-216.

[3] Isenberg, T., Nix, S., Schwarz, M., Miede, A., Scott, S. D. and Carpendale, S. 2007. Mobile spatial tools for fluid interaction. In *Technical Report* (Department of Computer Science, University of Calgary, Jul 2007).

[4] Jacob, R.J.K., Girouard, A., Hirshfield, L.M., Horn, M.S., Shaer, O., Treacy, E.S. and Zigelbaum, J. 2008. Reality-based interaction: a framework for post-WIMP interfaces. In *Proc. of ACM CHI 2008 Human Factors in Computing Systems Conference*, (ACM Press, 2008).

[5] Scott, S. D., Grant, K. D., and Mandryk, R. L. 2003. System guidelines for co-located, collaborative work on a tabletop display. In *Proc. of the Eighth Conference on European Conference on Computer Supported Cooperative Work* (Helsinki, Finland, September 14 - 18, 2003). K. Kuutti, E. H. Karsten, G. Fitzpatrick, P. Dourish, and K. Schmidt, Eds. ECSCW. Kluwer Academic Publishers, Norwell, MA, 159-178.

[6] Tan, D. S., Gergle, D., Mandryk, R., Inkpen, K., Kellar, M., Hawkey, K., and Czerwinski, M. 2008. Using job-shop scheduling tasks for evaluating collocated collaboration. *Personal Ubiquitous Comput.* 12, 3 (Jan. 2008), 255-267.

[7] Tang, A., Tory, M., Po, B., Neumann, P., and Carpendale, S. 2006. Collaborative coupling over tabletop displays. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada, April 22 - 27, 2006). R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, and G. Olson, Eds. CHI '06. ACM, New York, NY, 1181-1190.

[8] Winograd, T. and Flores, F., Eds. 1985 *Understanding Computers and Cognition*. Ablex Publishing Corp.

3-D Interaction with Wall-Sized Display and Information Transportation using Mobile Phones

Hideki Koike Masataka Toyoura
Graduate School of Information Systems
University of Electro-Communications
1-5-1, Chofugaoka, Chofu
Tokyo 182-8585, Japan
koike@acm.org
masataka@vogue.is.uec.ac.jp

Kenji Oka Yoichi Sato
Institute of Industrial Science
University of Tokyo
4-6-1 Komaba, Meguro-ku
Tokyo 153-8505, Japan
oka@hci.iis.u-tokyo.ac.jp
ysato@hci.iis.u-tokyo.ac.jp

ABSTRACT

Recently large wall-sized displays are often seen in public environment. Such wall-sized displays only show information to users, but do not allow users to interact information on the wall. Some input devices, such as touch panels, could be used, but it is hard to develop larger touch panels and they are extremely expensive. On the different perspective, one of the main roles of such wall-sized displays is to deliver information. However, few displays allow users to obtain detailed information. This paper describes a wall-sized display system which allows users to manipulate displayed information by using their hand or fingers. Moreover, by using 2-D barcode which can be recognized by recent mobile phones, the system allows users to receive displayed information to their mobile phones and to send information in the mobile phones to the display.

1. INTRODUCTION

Recently we often see large wall-sized displays such as plasma displays or LCD projectors in public environment. These wall-sized displays are expected to be used as interactive advertising posters or as interactive bulletin board.

In order for such wall-sized displays to function as an interactive media, two issues should be solved. One is a natural and intuitive interaction method which is suitable for wall-sized displays. Current wall-sized displays only output information and do not allow inputs from users. For this issue, traditional input devices such as mouse and keyboard are inadequate. As well, when we imagine the scenes where the wall-sized displays are used such as in public environment, special devices such as a data glove which is popular in VR researches or additional markers for motion capture is also inadequate. One of the promising approach is computer vision based method such as body recognition or hand/finger recognition. However, it must be a real-time recognition and it must be robust for lighting condition.

Second issue is information transformation between the wall-sized display and its users. For example, in case of advertising posters, people might want to obtain a link to the detailed information about the merchandise from the wall and access to it later. In case of bulletin board, people might want to publish their message or photos to the wall. For this issue, using wireless LAN or bluetooth is a candidate. However, it is not easy to send information to the user who selected the information when there are some users interacting with the wall. The system need to recognize who selects and which information, authenticate the user (or his/her information device), establish communication channel, and send the information.

Only after these two issues are solved, the wall-sized display works as an interactive media for information publication and sharing.

This paper describes the wall-sized display system which enables natural and intuitive interaction with users' bare hands and information transformation between the wall and user's mobile phone.

2. RELATED WORK

Interaction with Wall. As an input device for wall-sized displays, a touch panel is a candidate. However, it is extremely expensive to develop larger touch panels. GestPoint by GestureTek[5] uses two cameras to capture user's hand and enables to point an object displayed on the wall. However, it recognizes a hand position on a 2-D plane which is perpendicular to the wall. It does not support the recognition of 3-D recognition and gestures. Holowall[4] uses infrared light and infrared camera. It recognizes position and shape of objects near the wall. Therefore, interaction is limited on the wall or very close to the wall. Much worse, Holowall requires larger space behind the wall to set up light and cameras.

Information Transportation. There are some researches which demonstrates transformation of information between information terminals and personal devices. However, these systems work only in closed environment. i-Land[1, 2] demonstrated transportation of information by using small cube. Users can transport information displayed on the wall to the table. Pick-and-drop[3] showed intuitive information transportation between PCs using wireless LAN. The user selects information on one PC by pointing with the stylus and then moves the information to another PC by pointing the PC.

3. VISION-BASED INTERACTION USING HAND/FINGER

This section describes vision-based interaction with the wall-sized display using hand and fingers.

3.1 Hardware

Figure 1 illustrates the hardware setup of the wall system. On top of the display, a stereo camera unit (Point Grey Research Digiclops[7]) is installed as these cameras look down the floor. The camera unit is connected to the PC and the PC's screen is displayed on the wall.

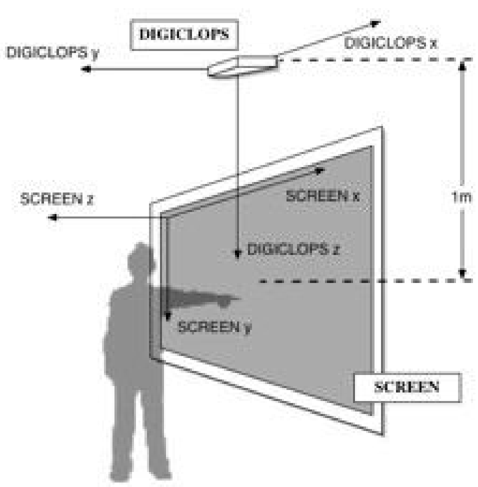


Figure 1: Hardware setups and coordinates

3.2 Hand/finger Recognition

The hand/finger recognition is based on our previous work on augmented desk system[8]. In [8], template matching with normalized correlation were used in order to track fingertips on 2-D surfaces. We extended this hand/finger recognition to 3-D.

First, hand regions are segmented by using skin color in each image captured by color CCD cameras in the camera unit. Unlike most of other hand recognition researches, the background subtraction is relatively easier because the background is the floor. There is no moving object such as other humans and the floor's color is dark and stable.

After the hand regions are segmented, an image shrinking operation is applied to the hand regions in order to calculate the center of each palm in camera view (i.e. 2-D).

Then, the template matching operation is applied to detect fingertips. Figure 2 shows its result. Some fingertips are occluded by the own hand. However, if we do not expect complicated gestures, this is enough.

After the 2-D positions of each hand and fingertips are decided for one camera view, their 3-D positions are calculated by using stereo matching technology as seen in Figure **??**.

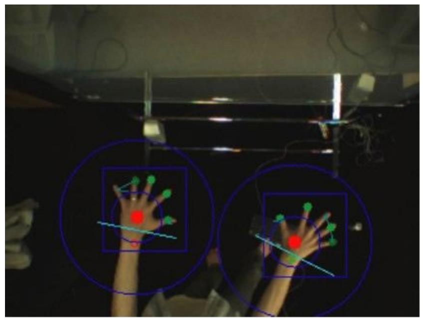


Figure 2: Detection of hand and fingers.

3.3 Coordinate transformation

Let camera coordinate and screen coordinate be (D_x, D_y, D_z) and (S_x, S_y, S_z), respectively such as shown in Figure 1. We assume that the plane $D_x - D_y$ is parallel to the planes $S_x - S_y$.

The screen coordinate has to be calibrated when the system is used for the first time. There is no need for calibration unless the camera or the screen is moved.

Let the coordinates of the top-left corner and the bottom-right corner of the screen in camera coordinate be (P_{x0}, P_{y0}, P_{z0}) and (P_{x1}, P_{y1}, P_{z1}), respectively. The coordinates of hand center in camera coordinate (H_x, H_y, H_z) is transformed to those in screen coordinate (S_x, S_y, S_z) as follows.

$$\begin{bmatrix} S_x \\ S_y \\ S_z \end{bmatrix} = \begin{bmatrix} \frac{1}{P_{x1}-P_{x0}} & 0 & 0 \\ 0 & 0 & \frac{1}{P_{z1}-P_{z0}} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} H_x \\ H_z \\ H_y \end{bmatrix} - \begin{bmatrix} \frac{P_{x0}}{P_{x1}-P_{x0}} \\ \frac{P_{z0}}{P_{z1}-P_{z0}} \\ \frac{P_{y1}+P_{y0}}{2} \end{bmatrix} \quad (1)$$

3.4 Gesture Recognition

The system recognize some gestures. Most important one is a touch gesture. The system determines as the user touches the screen if the distance between the user's fingertip and the screen is less than a certain threshold (e.g. 10 cm).

Other gestures are determined by counting the number of recognized fingertips. When one finger for one hand is recognized, it is recognized as a pointing gesture. When no fingertip is recognized, it is recognized as a grabbing gesture. When three or more fingers are recognized, it is recognized as a releasing gesture. Although these are very simple gestures, they are enough to perform actions which are common in GUI environment, such as click, drag, and release.

3.5 Performance Evaluation

The system tracks two hands simultaneously at 12 fps with Pentium 4 3.2 GHz PC. This limitation is mainly because of the limitation of the camera's frame rate.

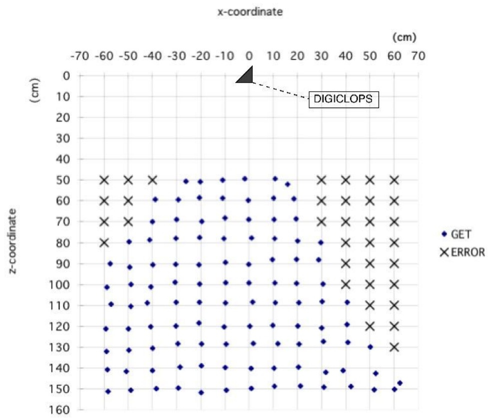


Figure 3: Accuracy of pointing.

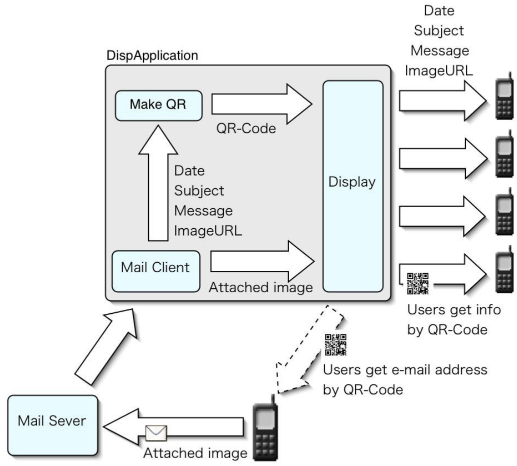


Figure 4: Information transformation system.

We measured the accuracy of pointing of the system as below. First, a 10x10 cm lattice was made on the wall. Then, the user touched each grid point with his middle finger. Figure 3 shows its result. In the figure, the camera unit is at the origin of the graph. Black dots show the measured coordinates. Cross markers shows that the hand position was not recognized correctly because the hand was out of the camera's sight. As seen in the figure, the errors are within 3 cm in the camera's sight. We think it enough accuracy for our large display applications.

4. INFORMATION TRANSFORMATION USING MOBILE PHONES

As we described in Section 1, it is relatively difficult to exchange information between the wall and the user's device because of the authentication issue. We solved this by using 2-D barcode and its reader on the mobile phone.

The mobile phones which are able to read a special 2-D barcode called QR-code are very popular in Japan. At the first quarter of 2005, at least 30 million mobile phones, which is one third of all phones in Japan, have such function. Some text such as messages, URL address, or email address, are encoded in this 2-D barcode. When the user captures the 2-D barcode using a camera on the phone, the encoded text is decoded. If the decoded text is an URL, the user can jump to the site just by clicking the URL. If the decoded text is an email address, the user can send a message and attachment file to the address. Thanks to the generic mobile phones and 2-D barcode, our system becomes so scalable that it would work in the real world.

4.1 System

Figure 4 shows an overview of the information transformation system. The system composed of three modules, mail client, barcode generator, and display client. We described them in detail.

Mail client. When the user send an email with an attachment file to the specified address encoded in the 2-D barcode, the mail client accesses to the mail server and download the email. The email is divided into the text part and the attachment file. The attachment file is sent to the display client. The text part is used to get the date, subject, and body of the email. The mail client is written by using Java Mail API.

Barcode generator. The URL address where the attachment file is saved is passed to the QR-Code generator[9] with the date, subject, and body of the email. The QR-code generator generates a QR-code in which these information are encoded, and then passes the QR-code to the display client.

Display client. Display client shows images attached to the emails. when the user selects the image on the screen, the display client shows the image and the QR-code generated by the QR-code generator. By capturing the QR-code by mobile phones, the user can obtain URL address where the detailed information on the image are saved.

4.2 Application: Interactive Bulletin Board

Figure 5 shows Interactive Bulletin Board System which demonstrates the information exchange between the wall and the personal device. On the right, Icons which represent each information are shown. At the bottom left corner, a 2-D barcode is displayed. This barcode encodes an email address of a mail server corresponding to this bulletin board. If the user want to upload a file to this board, he or she captures the barcode and send the file to the email address.

When the user wants to see the detail of each displayed icon, he or she points the icon with his or her finger. The user does not need to touch the screen. Figure 5 shows a snapshot of the wall when the user selected one icon and the associated barcode is shown on the screen. In order to download the information, the user may read the barcode by his/her phone.

5. DISCUSSION

Figure 5: Interactive Wall. Detailed information and QR-code are shown.

Figure 6: Interactive Wall. A user is downloading detailed information using a mobile phone.

hand/finger recognition. The vision-based input looks similar to GestPoint[5]. However, there are some essential differences between them. First, GestPoint recognizes 2-D movement of a hand, but our system recognizes 3-D movement. GestPoint recognizes one hand at a time, but our system recognizes more than one hand (depending on the CPU power). Also, our system recognizes simple gestures by counting the number of fingertips. Accuracy of pointing is not described in GestPoint specification.

The camera unit is very small and it could be used not only for projection screens but also for plasma displays.

On the other hand, there are some issues which should be solved. Since our hand recognition system currently uses skin color to detect hand region, it is sometimes unstable for different light conditions. This issue could be solved by using depth information which is obtained by stereo vision instead of color information. Also, use of infrared cameras would make it easier to detect human skins.

information transformation. As we described in Section 1, one of the issues when we try to exchange information between public terminal and personal devices when there are multiple users is how to identify and authenticate the user who select information on the terminal. Our system solved this issue by using 2-D barcode.

Currently our data transformation uses email system which is based on SMTP protocol. Therefore, it is not adequate for transferring large sized data such as movie clips.

6. CONCLUSIONS

We developed an interactive wall system which allows its user to manipulate displayed information with or without touching by his or her own hand. The system also allows the user to exchange information between the wall and the user's mobile phone. and information transportation between the wall and user's mobile phones. Because of the use of 2-D barcode and mobile phones, the system has become generic and could be used in the real world.

7. REFERENCES

[1] N. A. Streitz, J. GeiBler, T. Holmer, S. Konomi, C. MullerTomfelde, W. Reischl, P. Rexroth, P. Seitz, R. Steinmetz, "i-LAND: An interactive Landscape for Creativity and Innovation", In: ACM Conference on Human Factors in Computing Systems(CHI ' 99), May 15-20, 1999. pp.120-127.

[2] N. A. Streitz, Th. Prante, C. Muller-Tomfelde, P. Tandler, C. Magerkurth, "Roomware: The Second Generation", In: Video Proceedings and Extended Abstracts of the ACM Conference on Human Factors in Computing Systems (CHI ' 02), Minneapolis, MN, USA, ACM Press, April 20-25, 2002. pp. 506-507.

[3] Jun Rekimoto,"Pick-and-Drop: A Direct Manipulation Technique for Multiple Computer Environments", Proceedings of UIST'97, pp. 31-39, 1997

[4] J. Rekimoto et al. "Perceptual surfaces: Towards a human and object sensitive interactive display." In Proceedings of Workshop on Perceptual User Interfaces (PUI ' 97), October 1997.

[5] GestureTek http://www.gesturetek.com/gestpoint/

[6] Nebojsa Jojic, Thomas Huang, Barry Brumitt, Brian Meyers, Steve Harris,"Detection and Estimation of Pointing Gestures in Dense Disparity Maps " Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition March 2000

[7] Digiclops http://www.ptgrey.com/products/digiclops/

[8] Kenji Oka et al., "Real-Time Fingertip Tracking and Gesture Recognition", IEEE Computer Graphics and Applications, vol22, Issue6, Nov.2002, pp.64-71

[9] Psytec http://www.psytec.co.jp/

CityWall: Limitations of a Multi-Touch Environment

Ann Morrison, Giulio Jacucci, Peter Peltonen

Helsinki Institute for Information Technology (HIIT)
Helsinki University of Technology and University of Helsinki
P.O. Box 9800, 02015 HUT, Finland
firstname.lastname@hiit.fi

ABSTRACT

In this paper we discuss some of the successes of current multi-touch surfaces and look at what these interfaces enable. We work specifically with CityWall as a case study—a multi-touch display installed in the center of Helsinki. We then discuss some of the shortfalls, focusing on the limitations of technologies that unintentionally support novelty use and/ or disregard for content. We briefly touch-on some of the ideas under consideration for the next stages of development to overcome these perceived shortcomings.

Categories and Subject Descriptors

H.5.1. Multimedia Information Systems:

General Terms

Design, Experimentation, Human Factors

Keywords

Situated public displays, urban environments, multi-user interfaces, group interaction, multi-touch, gestural interfaces, experience-design.

1. INTRODUCTION

In this paper we discuss some of the limitations and affordances that a multi-touch display provides. We have installed a large multi-touch display called CityWall in the centre of Helsinki to observe how group interaction happens naturally there. We conducted extensive field trials and based on these findings we look at what such a system enables in such a context. We look briefly at other related works—the breadth of these though is too wide for the scope of this short paper. The main focus then is to outline what works, what does not work, and to explore further why this is and finally to look at what kind of improvements might then be tackled. We found that users do not process the actual information on the wall, rather the activity is as if learning to 'play ball' in a new medium. We are currently exploring future developments for CityWall and in this paper we touch on some of the improvements under consideration. We are exploring in particular how CityWall as an environment can go beyond its novelty factor and truly address the user experience. There are many potential solutions we see in other works, particularly ones that work directly with the situated community. However simplicity of use has also shown itself to be a key factor.

PPD'08, May 31, 2008, Naples, Italy.

2. RELATED WORK

Previously the social dimension of large display use has been studied in tabletop, ambient and large display research. *Tabletop displays* have been used primarily in collaborative work spaces. Research has presented new kinds of collaborative touch-based interaction techniques that also support multi-hand use [1, 2, 3]. *Ambient displays* do not usually involve direct interaction on their surface as they have been developed to investigate the ways in which displays can be situated in physical settings, representing movements of people in a space, displaying information that requires only peripheral attention, and increasing awareness of other users [4, 5, 6]. In this section we will briefly introduce some aspects of multi-touch systems that are currently in use, which relate to the kinds of issues we look at with the CityWall case study.

The settings of *large multi-user wall display* research have ranged from collaborative workspaces in office environments to more public settings such as schools. A study on BlueBoard, a touch-screen display that can identify its users, highlighted the benefits of visible physical actions facilitating learning from others, difficulties in developing turn-taking practices, and supporting ways to collaborate without necessitating anyone taking a leader role [7]. While CityWall does not identify its users, as a system it does readily support turn-taking and collaboration. A study on eyeCanvas, an interactive single user public display and bulletin board installed in a gallery café, highlighted the richness that messages containing not just plain text but also user contributed pictures and sketches can have and discussed ways to better enable 'conversations' [8]. Support for conversation is important for engaging users and we will discuss how this is missing in the current implementation of CityWall.

Another system, Dynamo, was installed in a school as a multi-user public display for multimedia sharing. This system supported the use of private content with dedicated spaces on the screen for personal purposes. During the user study various use patterns evolved, including ways to draw other people's attention through "upsizing" one's pictures, and staging video performances in the display [9]. We found similar activities in our user studies with the enactment of performative roles, upstaging, upsizing etc and also aim to include private data in the future development of CityWall.

Furthering the integration of multiple devices, a system such as iRoom (http://iwork.stanford.edu/) operates as a meeting, research and work space, combining large displays, wireless/multimodal I/O and mobile devices such as handheld PCs. Unfortunately it falls prey to the necessity for a 'wizard user' that is needed for solving problems and conflicts caused by the setup of changing multiple devices. This is one of the pitfalls of connecting many devices to a system and needs to be considered when integrating private data and devices.

Another problem with dealing with private data has been encountered with Braccetto (http://www.hxi.org.au/). This large-scale multi-touch system, can be used both as a tabletop and as a vertical display, making this a very flexible system where groups can video-conference, as well as file share and problem-solve simultaneously. The system is designed for use by groups working in emergency situations like fires and floods. However, due to strict government security policies users can work only with limited access to restricted information at any one time. This makes the system as a portable environment—one of its aims—cumbersome to use, and stalls the immediacy of team work.

CityWall as a multi-touch gestural system (http://CityWall.org) is a vertical surface that works for multiple users 'playing' and sharing information remotely as well as locally. This system operates well in changing lighting conditions, both indoors and outdoors. The display is set as a permanent installation in an urban environment and initially aligned its ‘openings’ with local festivals and events. We will concentrate on this display—and the results of our field work—in the confines of this paper.

3. AFFORDANCES AND LIMITATIONS

For the remaining discussion we will concentrate on our case study, CityWall that was setup to investigate the interaction and situatedness of displays in an urban setting. CityWall can be used by people who take part in different events happening around Helsinki, as well as for daily life ‘events’. It shows the digital media content people have captured in those events and then have submitted to Internet media sharing services.

We studied the use of the CityWall using two approaches. In a first series of studies at city events we recruited groups of visitors (around 6 participants per event) equipping them with mobile phones and applications to publish their pictures on the CityWall. These studies lasted a long weekend and were aimed at exploring how the CityWall supports groups at events [10]. A second approach was used to study passers by interaction at the display. The core of the study included observing interactions for eight days during summer 2007 [11]. A total of 1199 persons were observed to interact with the system in various social configurations. Videos of these encounters were examined qualitatively as well as quantitatively based on human coding of events. Many different types of interaction were observed during this time: crowding, massively parallel inter-action, teamwork, games, negotiations of transitions and handovers, conflict management, gestures and overt remarks to co-present people, and “marking” the display for others.

The multi-touch feature, gesture-based interaction, and the physical display size contributed to these uses. Unlike in most of the settings in which public displays have been studied in previous research, a real urban environment is populated by individuals and groups that are strangers to each other. In our study it was shown how people were configured in groups of users and crowds of spectators rather than as individual users. They were able to use the display both in parallel and collectively by adopting different roles. Learning from other users may be one of the key explanations for this: seeing someone else using the display made people aware that it was an interactive installation and when standing behind the earlier users people learned more about its interactive properties.

Figure 1. Two people using CityWall.

The public location and size of CityWall created a sufficient space for a “stage” for multiple users who were able to adopt different roles, such as being teachers, apprentices, clowns, or members of the audience. In some cases, multiple activities were taking place at the same time at the display. Content on the wall and features of the interface were used as resources to coordinate the activity and to create events or interactions so they were meaningful in front of others: interaction could be perceived as a performance to others. The multi-touch feature of the interface was central, as it supported expressive gestures that helped participants in coordinating, communicating and acting out different roles.

3.1 Novelty and “superficial usage”

The CityWall project aimed at giving access to present and past events of the city by engaging passers by with tagged images. It became clear after on-site interviews that users were not always interested in the pictures but where mostly exploring the playful interface. The groups recruited at city events of course were interested mostly in the pictures they created and published on the CityWall. While it did occur that also passers by reported being interested in the pictures, the higher interest in the novelty and playfulness of the interface poses several challenges:

Novelty “factor”: What happens when the novelty factor wears out? How do we keep users engaged with the installation? There have been cases of users coming back to the installation, to try it out again, sometimes this was users who were also professionally interested in the installation, or others that came back to show their friends how it works. Also a scenario where an installation constantly seduces passers by with its newly developed engaging interaction techniques is not feasible. Rapid design changes are not easily accomplished in such an environment.

Application design: It is difficult to distinguish in our study the contribution of the application to the success in terms of usage but also the non-success in terms of users not paying attention to the content. What if, for example, pictures would have been organized not only chronologically but also using more thematic groupings? Users might have found a more “meaningful” way to browse the content. The problem of evaluating a multi-touch installation then shows how it might be difficult to distinguish between the contribution of the engaging multi-touch technique and how the application makes use of it.

3.2 Limitations

Our field studies indicate that one of the limitations of CityWall is that users only interact with the display after seeing it in use. So unless a more adventurous person, or somebody who knows how to work with the wall is active, then the display may be viewed as if it is a shop window or an advertising space and is not interacted with at all. As a work-around, we are looking to put in a time-out default setting displaying a life-size demonstration movie of people using the wall.

As it stands, the interface is designed for intuitive use, and so that novice users can and do easily participate. We have already discussed the novelty factor in section 3.1. As Csikszentmihalyi argues, to maintain optimal engagement tasks need to be within the realm of the possible, but must still stretch the participant [19]. Here once the participants learn to use the timeline; rotate, enlarge, shrink, slide, and throw the images; perhaps even bringing friends, or showing others how to use—all the while enacting the roles as discussed in section 3—they have achieved the finite potential available. There are no more tasks to stretch the participant. If we want something to be taken up on a continuous basis, it is important to consider the addition of varied levels of difficulty to continue to 'stretch' the participants beyond this initial learning curve.

Our field studies also indicated that other peoples images have limited relevance, unless the participants have some level of engagement with the place or activity. The tags and annotations to the images mainly give limited descriptors, so there are no stories that can be readily associated with the image, there are no reply comments that can be added onto at the site, so the only further discussions that happen at CityWall do so at an oral level, that is not then translated onto the digital display. Conversations and stories evoked at the wall are then lost. As well we found that most of the participants at CityWall were tourists to Helsinki, who had come there for either a festival or some event, so CityWall as an environment had no great 'sense of place' or on-going community engagement for them.

One of the major limitations we find with the current interface design is that if one user moves the timeline it, then the timeline is moved for all users and people lose the content they were 'working' with. So for example, if the current timeline is positioned at e.g. 21st May 2007 and 2 groups are playing with two different sets of images, and one user moves the timeline forward to e.g. 13th June 2008, then this move completely disrupts what the other group of users were doing with their images at 21st May 2007 and they can no longer access these. While this can facilitate interesting negotiations for groups at the wall (see Peltonen et al [11] and section 3) this severely limits the ability to engage in multiple interactive spaces and restricts entirely any threading of images—as *photographic-type conversations*—through time [8]. CityWall supports "one conversation space and that without threading" [8, p. 9] so there is no way—without extensive scrolling back through time—to link images that are responses to other images in this current interface. During Helsinki festival we did see the participants and organizers enacting image-based 'conversation' about the event on CityWall. There is of course further discussion on what constitutes 'a conversation' and if either of eyeSpace or CityWall allow this.

While we have analyzed the group dynamics and interactions at the wall, our studies have not analyzed the content there, nor the *persistent photographic conversations* that may be in play from the community who do regularly upload images there.

However, it may be that the participants to CityWall engage via Flickr only. For example, a search on Flickr for one of the tags used for CityWall, shows 319,049 results with one randomly selected image from the collection having 55 comments or exchanges (http://www.flickr.com/photos/doc18/262860166/), another with six comments, another again with eight. These participants may or may not be aware that their images (because of how they are tagged) are also selected and displayed on CityWall, so while the picture-based exchanges and comments are occurring in an online environment, their relevance in the large display in an urban environment is at this stage unexplored and unknown. Another limitation for the participants who do actively and knowingly engage with posting images to CityWall is that while they may post these images online or via their phone, they have to physically return to the wall in order to see their images large-scale. To enable follow-up comments on a multi-user multi-touch display requires consideration of the interface design. Replication of an online environment is not a feasible solution.

3.3 Affordances

Above, we have detailed some of the limitations that our field studies have revealed, so what then are the successes we have found with this multi-touch display? Some are already discussed in section 3, but we will add detail in this section to enable further discussions and to ensure we bear these in mind for any future re-design considerations.

Firstly, what we find is that it is that novice users to technology generally can easily participate with this interactive display. This is a no-nonsense system of use, with no need for difficult drill-down procedures or an intensive learning period to access the features. While we have discussed already this factor in terms of its limitations, it is important to consider technology proficiency. Where the intended audience is a passing general public, and an unknown quantity, novice-level use acts as a draw card for initial engagement, attracting the public in. Secondly, evidence from our video footage and interviews revealed an evident sense of achievement amongst these first-time users, of which the longer-term impact cannot be gauged. Thirdly, for participants who added images and saw these on CityWall, there was interest in the content—and a palpable sense of ownership and involvement in the 'event'. Their contribution was evident to themselves and to others at the time—they often bought others along to show-off their images—as well as being written into the archival history of the event.

Finally, gesture and play as a pleasurable means to be able to interact was also evident, both with the display itself and with others interacting at the display. As Don Norman, a prominent interaction design researcher comments: gestural systems are "agreeable to the senses… [and] pleasurable to use. This [gestural system] engenders good feelings in users". [13] This is an important feature of the CityWall display and should not be under-estimated—as is a facility in an urban environment that affords play and social activity. Video footage showed much laughter and enjoyment by our participants. However it is important to note, that while for some participants we witnessed an initial reluctance to participate, this rapidly evaporated after a few moves had been successfully negotiated. Confidence in use then, was fast achieved, and advanced rapidly into playful, experimental activity. In this environment it was observed that participants felt free to play and explore without fear of making errors. [14] Our field studies revealed our participants interacting in playful ways—they go to the city centre for leisure and entertainment (e.g. for a festival) and then look for things to do. The beauty of a work like CityWall is that it affords a place for just such ludic activity. CityWall offers people something to do, something to be involved with and a place and means to easily meet and interact with other people through.

4. EXPANSION AND EXPERIMENTATION

How then do we work with the limitations and affordances in a meaningful way to extend and improve the current implementation? We discuss here some of the considerations under review. Of course the implementation requires we are as critical of our technical practices, as we are of the kinds of activities we are looking to support. Some of what is under discussion here is in the early stages of experimentation and discussion in a design-development-iterate cycle that the small CityWall team deploys as its process.

To support user's engagement to the content presented on a public multi-touch display, we need to add more interaction techniques other than just simple browsing, moving and/or resizing of the content and items. This could be facilitated by looking at how to further support the collaborative interaction we have witnessed happening at multi-user public displays, for example considering how to develop additional interaction techniques that go beyond multi-touch. We note that CityWall makes use only of multi-touch gestures and visual output. This could be extended to combine gestures, the acoustic feature of speech and multi-touch, as we see with an interactive gestural and voice-activated environment such as PuppetWall [15]. New interaction techniques need not include only additional modalities, but could also contain new gestures that would for example help to go beyond the current 2D paradigm of resizing, moving and rotating objects. With three dimensional objects and space the current interaction interface and paradigms could be extended to enable a more sophisticated and in-depth level of access to information and content.

In addition, current tabletop prototypes demonstrate the inclusion of several other interaction techniques. Mobile devices, physical artifacts and even something as simple as sound can dramatically change the user experience. For example, a mobile phone gives access to one's own data and/or habitual ways of working (e.g. personal devices). Adding the capability for users to access their own data creates design opportunities to extend for users how they may then interact with a multi-touch installation, and also for defining what the display might be used for (e.g. they may process the information, or act more efficiently with all resources at hand).

Having multiple users and both parallel and collaborative use happening at a public display raises the question of how to deal with conflicts that are unavoidable when users have different goals. For more in-depth use, where access to content on a deeper level is required, the ways to access multiple contents at different levels at the same time need re-consideration. A discussion around a design that enables and implements multiple timelines, allowing multi-use by groups and individuals is one potential way to resolve this. Recently work was published that proposes group gestures on multi-touch table top, which however was evaluated only in a laboratory setting [2]. According to our experience with CityWall, any design idea that structures group interaction should be trialed in a naturalistic setting, so that all elements affecting the interaction between the users and the system are present and can be observed.

5. CONCLUSIONS

The evaluation of CityWall either recruiting groups at events or analyzing passers by interaction resulted in several findings. The studies showed the difficulty in prolonging engagement of specific groups or communities. The CityWall does not provide support for conversation threads, persistent conversations or other thematic groupings other than chronological. We believe these would be beneficial to target usage beyond sporadic or ephemeral interactions of passers-by.

The key issue is to how the CityWall is useful to users. Generally either a specific practice is studied and becomes the design target of the technology or the technology enables a new practice. We feel that we have not yet found a clear hypothesis for neither of these approaches. For example if the CityWall is to address public-picture-sharing it should take this practice more seriously.

The CityWall did support serendipitous social interaction in public space and a more conscious designlong these lines could be attempted. That an urban environment provides for its citizens a place for play activity, for genuine exchange, and unplanned interactions with strangers cannot be under-estimated. Despite some of the limitations coupled with these features, these aspects need to be maintained in future developments—so we look to add more features without taking away the affordances that genuinely work well.

Currently the design does not concretely support any specific "practice". One aspect of practice is of including specific communities and groups. How then can the interface—and the types of interactions this allows—be expanded to allow for more in-depth and meaningful exploration of content? How can this content have meaning to the inhabitants of the place itself—and the citizens of the city claim CityWall as their own? To do this well we need to ensure we 'ask' the stakeholders in the space—the shop owners, the passers-by, the city-dwellers—what they need, and integrate the kinds of technologies and activities that fit well with their everyday activities, to allow them to meaningfully participate and co-author in their place. This requires we employ current critical cultural, technical, and community practices and ask what do our citizens want with their CityWall?

6. ACKNOWLEDGMENTS

This work has been co-funded by the 6th Framework Research Programme of the EU, through the IPCity project (FP-2004-IST-4-27571).

7. REFERENCES

[1] Morris, M.R., Huang, A., Paepcke, A., Winograd, T. Cooperative gestures: multi-user gestural interactions for co-located groupware. In Proc. CHI 2006, ACM Press (2006), New York, 1201–1210.

[2] Rogers, Y., Hazlewood, W., Blevis, E. and Lim, Y.-K. Finger talk: collaborative decision-making using talk and fingertip interaction around a tabletop display. In Proc. CHI 2004, ACM Press (2004), New York, 1271–1274.

[3] Wu, M. and Balakrishnan, R. Multi-finger and whole hand gestural interaction techniques for multi-user tabletop

displays. In Proc. UIST 2003, ACM Press (2003), New York, 193–202.

[4] Skog, T., Ljungblad, S. and Holmquist, L.E. Between aesthetics and utility: designing ambient information visualizations. In Proc. InfoVis 2003, IEEE Computer Society (2003), New York, 233–240.

[5] Vogel, D. and Balakrishnan, R. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In Proc. UIST 2004, ACM Press (2004), New York, 137–146.

[6] Wisneski, C., Ishii, H., Dahley, A., Gorbet, M., Brave, S., Ullmer, B., Yarin, P. Ambient displays: turning architectural space into an interface between people and digital information. In Proc. CoBuild 1998, Lecture Notes in Computer Science 1370, Springer (1998), Berlin, 22–33.

[7] Russell, D.M., Drews, D. and Sue, A. Social Aspects of Using Large Public Interactive Displays for Collaboration. In Proc. UbiComp 2002, Lecture Notes in Computer Science 2498, Springer (2002), Berlin, 229–236.

[8] Churchill, E. and Nelson, L, Interactive Community Bulletin Boards as Conversational Hubs and Sites for Playful Visual Reparteee, Annual Hawaii International Conference on System Sciences, 2007. HICSS 2007. 40th Volume , Issue , Jan. 2007 Page(s):76 – 76.

[9] Brignull, H., Izadi, S., Fitzpatrick, G., Rogers, Y., and Rodden, T. The introduction of a shared interactive surface into a communal space. In Proc. CSCW 2004, ACM Press (2004), New York, 49–58.

[10] Peltonen, P., Salovaara, A., Jacucci, G., Ilmonen, T., Ardito, C., Saarikko, P. & Batra, V. (2007). Extending large-scale event participation with user-created mobile media on a public display. In Proc. of the 6th international conference on Mobile and ubiquitous multimedia (MUM2007), pp.131-138.

[11] Peltonen, P., Kurvinen, E., Salovaara, A. Jacucci, G. Ilmonen, T. Evans, J. Oulasvirta, A., Saarikko, P. (2008). "It's mine, don't touch": Interactions at a large multitouch display in a city Center. Full paper to be presented in CHI2008.

[12] Csikszentmihalyi, M. Flow: The Psychology of Optimal Experience, New York: Harper and Row, 1990.

[13] Norman, (2005). D Emotional Design: Why We Love (or Hate) Everyday Things, Basic Books, NY.

[14] Saffer, D. Interactive Gestures: Designing Gestural Interfaces, O'Reilly Media, Inc, 1005 Gravenstein Highway North, Sebastopol, CA 95472 (2008)

[15] Liikkanen, L., Jacucci,G., Huvio, E., Andre, E., Exploring Emotions and Multimodality in Digitally Augmented Puppeteering, to appear in AVI 2008 Advanced visual Interfaces, ACM, Napoli.

Collaborative problem solving on mobile hand-held devices and stationary multi-touch interfaces

Simon Nestler Florian Echtler Andreas Dippon Gudrun Klinker

Technische Universität München
Institut für Informatik, I16
Boltzmannstraße 3
85748 Garching b. München
{nestler|klinker}@in.tum.de

ABSTRACT

This paper focuses on the coupling of mobile hand-helds with a stationary multi-touch table top device for collaborative purposes. For different fields of application, such as the health care domain, the coupling of these two technologies is promising. For the example of sudoku puzzles we evaluated the collaboration between multi-touch table top devices and mobile hand-helds. During the small-scale evaluation we focused on the differences between face-to-face collaboration and remote collaboration when solving problems collaboratively on table top devices and hand-helds.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces; H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces

Keywords

Multi-touch user interfaces, Mobile user interfaces, Multi-touch devices, Group interfaces, Collaboration

1. INTRODUCTION

In various computer applications the user has or wants to collaborate with other users. The role of the user interface is to optimally assist the collaboration of users. Whereas broader research has been performed on collaboration via desktop computers, there exists less experience on the collaboration using mobile or table top devices. For practical applications the combination of these two technologies is of special interest.

The challenge is, that the way of interacting with the mobile devices typically differs significantly from the interaction with table top devices. Whereas mobile devices are typically only used by a single user at a time, simultaneous multi-user input is possible on table top devices, especially when multi-touch devices are used. Moreover, the possibilities to present all relevant information to the user are substantially more limited on mobile devices than on table top interfaces.

1.1 Mass casualty incidents

In different fields of application the coupling of a table top device with mobile hand-helds makes sense. We plan to improve the collaboration of paramedics or doctors with the operation control center in mass casualty incidents by coupling mobile hand-helds with a multi-touch table top. Whereas paramedics and doctors require high mobility in order to be able to move around in the field, the operation controllers require an overview of the overall situation. Therefore equipping the paramedics with mobile devices and equipping the operation control center with a table top device would make sense. The operation control center as well as the paramedics retrieve from and store to the system all patient related information [10].

The way of presenting the information on mobile devices on the one hand and on the table top device on the other hand differ slightly. Whereas on the table top device information on the overall situation is presented, on the mobile devices the information on specific patients is of primary importance.

1.2 Modalities of collaboration

When coupling mobile devices with a multi-touch table top two entirely different ways of collaboration are possible. Either the table top device facilitates direct collaboration or the mobile devices facilitate remote collaboration. The collaboration on the table top device includes the possibility to directly keep track of all users' interactions, to point at problematic areas and to discuss face-to-face. When using mobile hand-helds, the users not necessarily have to be in the same room, they can freely move around during the collaboration. By combining the table top device with mobile hand-helds we expect to take the advantages of both technologies.

1.3 Sudoku

The concrete problem which we chose for the first evaluation of the collaboration between multi-touch table top devices and mobile hand-helds was the sudoku puzzle. We chose the

Table 1: Sudoku solution (Start values are written in bold).

7	**9**	**4**	5	8	2	**1**	**3**	6
2	6	8	9	3	1	7	4	5
3	1	5	4	**7**	**6**	9	8	**2**
6	**8**	9	7	**1**	5	3	2	4
4	**3**	**2**	8	6	9	5	7	1
1	5	7	**2**	4	3	8	**6**	9
8	2	1	6	**5**	7	**4**	9	3
9	4	3	1	2	**8**	6	5	**7**
5	7	**6**	**3**	9	**4**	2	1	**8**

sudoku puzzle because previous research has been performed on the exploration of relationships at the example of sudoku games by Klinker et al. [7]. First of all we want to focus on the question of collaboration between mobile and table-top devices, this subproblem can be represented at the example of sudoku.

We implemented this puzzle on the table top device as well as on the mobile devices. This puzzle consists of a 9*9 grid, the grid consists of nine 3*3 sub-grids. The puzzle has to be filled with numbers from 1 to 9 in a way, that each row, column and sub-grid contains every number exactly once, as shown in Table 1. On the basis of given start values the sudoku puzzle typically is uniquely solvable.

At first glance sudoku seems to be an absolute single player game. This is not true, in fact there are extended possibilities for collaborative solving. Especially because of the indirect dependence of the numbers from 1 to 9, the game can be solved collaboratively by assigning one or more numbers to each player. This simple subdivision of the entire problem facilitates the collaborative solving by up to nine players. Note that a similar subdivision exists in mass casualty incidents when assigning one or more patients to each paramedic or doctor.

2. RELATED WORK

Previous research on coupling mobile hand-held devices with public displays has been performed. The approach of Greenberg et al. [3, 4] bases on hand-held devices with personal information and large displays with public information. During a real time meeting the participants can share personal information and modify all public information. Carter et al. [1] proposed a combination of public displays with hand-held devices for public annotation of multimedia content. They used hand-held devices to augment, comment and annotate public content which is displayed on public displays. In the health-care domain public and private displays were used by Favela et al. [2]. They supported the decision making of doctors and nurses with mobile computing technologies. Furthermore they proposed a concept to integrate public displays in this ubiquitous application. Semi-public displays for collaboration within smaller groups have been developed by Huang et al. [6]. Their concept focuses on sharing information on activities within certain user groups. Information shared by group members is not fully public, it can be only viewed and modified by group members.

3. SYSTEM SETUP

For playing the sudoku puzzle collaboratively on the mobile hand-helds and the stationary table top device we designed a simple system architecture. The state of the sudoku game can be described by a string of 81 characters (assuming that a standard sudoku puzzle with a 9*9 grid is played). Starting in the upper left corner of the grid, all fields of the grid are listed row-by-row. In summary each field can take on one of 19 different states, besides the *empty* state (represented by *0*) it can contain a *user state* from 1 to 9 (represented by *1-9*) or a *start state* from 1 to 9 (represented by *A-I*). The current system architecture bases on a client-server model. The table top serves as the server to which the mobile hand-helds are connected via a wireless network. The current communication protocol is restricted to the commands which are compulsorily necessary for the collaborative solving of a sudoku puzzle:

- **State?**
  Client request for sending the current state of the sudoku puzzle
- **State! <valueString>**
  Client request for setting the current state of the sudoku puzzle to the state which is described by the *valueString*
- **State <valueString>**
  Server response on both state requests with the state contained in the *valueString*
- **Action? <x> <y> <value>**
  Client request for setting the field in column x and row y to *value*
- **Action <x> <y> <value>**
  Server response on an action request containing the *value* for the field in column x and row y

The requests for changing the server state typically succeed, provided that the *valueString* is syntactically and semantically correct. The string has to contain 81 characters from 0-9 or A-I to be syntactically correct. In order to suceeed the test on semantical correctness, the *start states* in the sudoku puzzle must be arranged in a way that the sudoku is solvable. For instance, each of the characters A-I may occur only once in each row, column and sub-grid. The fields filled with *user states*, however, are not tested during the semantical test because the sudoku remains solvable even if the user states are semantically inconsistent (assumed that the user interface contains the functionality to go back). On the one hand clients can join a running game by sending the *State?* request and on the other hand the clients can share their game to other clients by sending the *State!* request.

The requests for performing actions are slightly more complicated. An action which a client wants to perform can fail for two reasons: The client tries to overwrite a field filled with a *start state* with a *user state* or a other client tries to change the field at the same time. When one of these conflicts occurs, the server sends the current field state (which differs from the state requested by the client) in his *action* response to inform the client that his action failed. This

Figure 1: Sudoku puzzle on the multi-touch table top

concept is generally completely resistant against state inconsistencies because of the fact that a central server decides whose action succeeds and whose fails.

4. USER INTERFACE

As stated above, two different user interfaces are necessary for the two interaction modalities. The table top system has to support multiple concurrent users, while the mobile UI should be easily usable with a stylus.

4.1 Table top device

Multi-touch technologies for public displays have first been developed by Lee et al. in 1985 [9]. The multi-touch table top which we used for our implementation is based on the technology proposed by Han [5]. The table top user interface is presented in Figure 1. It was inspired by the JigSawDoku browser game [8]. On the left and right side of the grid, users are presented a selection of colored number tiles. Fixed numbers are shown with a white background. Users can drag and drop the colored tiles into the free fields of the sudoku grid by simply touching and moving them with their fingers. As the table top system provides multi-touch input, several users can concurrently move and place tiles. As the users can view the table from any side, the tiles show each number in four different orientations. To ease correct placement, the tiles snap into the free fields below a certain distance. During the game, users can quickly determine the approximate number of fields left for a certain number by looking at the tile colors. When the grid has been filled correctly, a message is displayed that the game has finished. The time which users took to complete the puzzle is displayed on top of the screen as well as logged to a file for later evaluation.

Figure 2: Sudoku puzzle on mobile devices

All tiles which are placed in the table top interface are wirelessly transmitted to the hand-helds and also displayed there. Vice versa, when a number is set on the hand-held, one of the free tiles on the table top is moved to the correct cell with a short animation.

4.2 Mobile hand-helds

The user interface for the mobile hand-held devices is shown in Figure 2. Due to the fact that screen space is highly limited when developing for mobile hand-helds the visualization differs from the one for the table top device. For the benefit of overview we had to do without displaying all unset tiles separately. Otherwise the space would have been too limited to show the complete sudoku grid at once. Thus the user interface then would have to contain intuitive metaphors to scroll, pan and zoom. Therefore we alternatively sorted all unset tiles on 9 different stacks and indicated the height of these stacks numerically.

The metaphor for moving tiles slightly differs from the one for the table top. During a review with experts we found out that the movement of tiles by the "'stick-to-finger"' metaphor is very inaccurate for hand-held devices. However, separat-

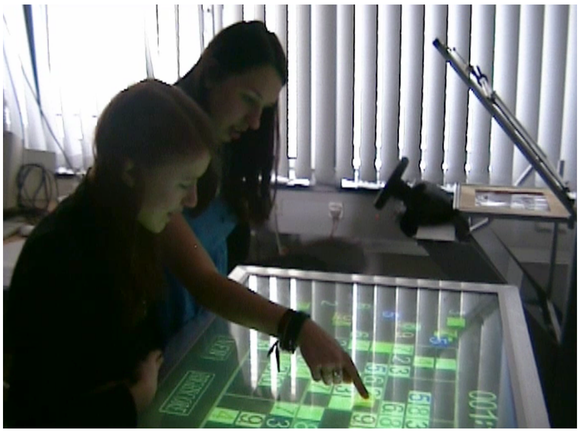


Figure 3: Evaluation of the sudoku puzzle in a small-scale user-study

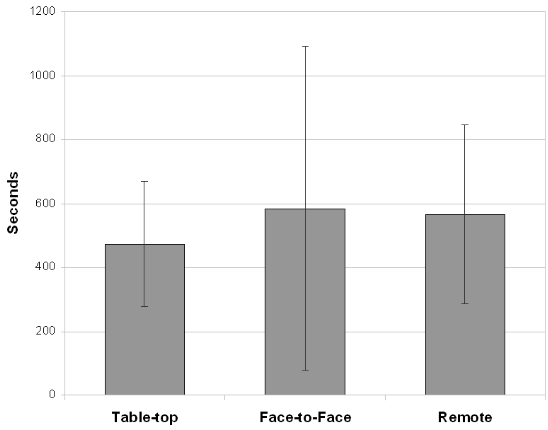


Figure 4: Quantitative evaluation results

ing the tile movement into the two steps *tile selection* and *tile placement* worked quite well, when performing both of these sub-actions with a separate click. First the user clicks on the tile which he wants to place and afterward he clicks on the field which he wants to fill with that tile. Furthermore when the user wants to place several tiles from the same stack, the first click is not required because the tile stacks remain selected. Additionally a tenth stack was included, the "'empty stack"' which can be used to clear *user state* fields. The metaphor for clearing fields works in analogy to the one for filling fields: first the empty stack and then the field which has to be cleared is selected. The height of the "'empty stack"' indicates the number of tiles which have still to be set in the current game.

5. EVALUATION

In addition to the expert review we performed a small-scale evaluation to determine the advantages of coupling mobile hand-helds with table top devices. In the user study shown in Figure 3 we focused on the impact of physical presence on the effectiveness of collaboration. The better the two user-interfaces support collaborative problem solving the less face-to-face discussions are essential for successful problem solving. Therefore we compared in the evaluation the effectiveness of face-to-face and remote collaboration in a quantitative manner. The subjective impression of the participants was identified by a questionnaire.

In total 16 people participated in our small-scale user study. Their objective was to solve five different sudoku puzzles collaboratively in teams of four. We evaluated three different alternatives of collaboration:

- **Table top.** All four people are collaborating at the table top
- **Face-to-face.** Two people are collaborating at the table top, two people are equipped with hand-helds. All participants are in the same room.
- **Remote.** Similar to *face-to-face*, but all participants are in different rooms (except the two at the table top).

Due to the fact that we wanted every participant to evaluate the *face-to-face* collaboration and the *remote* collaboration on the hand-held as well as on the table top we needed two cycles for these two alternatives. In summary five alternatives had to be evaluated by our four teams. We permuted the order of the alternatives to avoid training effects and to compensate potential differences in the difficulty of the five sudoku puzzles.

The quantitative results of the user-study are shown in Figure 4. When using the table top device the users solved the sudoku puzzle within 473 seconds in average (SD: 194 s), whereas the face-to-face collaboration needed 585 seconds (SD: 506 s) and the remote collaboration needed 566 seconds in average (SD: 280 s). As a consequence the null hypothesis could not be rejected in this small scale user-study. Face-to-face collaboration, however, seems not to be faster than remote collaboration when using hand held devices. This is a quite remarkable result when it can be approved in a larger user-study. On basis of this first small-scale evaluation we can assume that collaboration works best when the users are not only in the same room but also working on the same device.

In addition to the quantitative evaluation the subjective impression of the 16 participants was documented by a simple questionnaire which consisted of six questions:

- Which interface you did enjoy more? (1..table top – 5..hand-held): **2,4 (SD: 1,4)**
- Which interface was more efficient? (1..table top – 5..hand-held): **2,6 (SD: 1,3)**
- Have you been disturbed by the actions of other players when you played on the hand-held? (1..very often – 5..never): **3,1 (SD: 1,2)**
- Have you been disturbed by the actions of other players when you played at the table top? (1..very often – 5..never): **1,9 (SD: 0,7)**
- How present were the other players when you played

on the hand-held? (1..very present – 5..not present): **2,5 (SD: 1,0)**

- How present were the other players when you played at the table top? (1..very present – 5..not present): **2,1 (SD: 0,9)**

Regarding the interface the participants could not clearly decide between the table top and the hand-held device. The users on the table top were often disturbed by the hand-held users whereas they were not that much disturbing for the hand-held players. On the other hand the high disturbance leads to a high presence of the hand-held players.

6. CONCLUSION AND FUTURE WORK

We presented an approach to couple mobile hand-helds with a stationary multi-touch table top device. The evaluation showed that mobile hand-helds enable the users to remotely collaborate with users playing on the table top. Whereas a table top offers possibilities for direct collaboration, the physical presence of all participants can not be guaranteed in all applications. Therefore the extension of existing table top applications with mobile user-interfaces leads to an enrichment for the hole application. The future work will be to find out how the different modalities of collaboration work in detail. For instance it is interesting, whether the contribution of every single player depends on the used input device.

7. REFERENCES

[1] S. Carter, E. Churchill, L. Denoue, J. Helfman, and L. Nelson. Digital graffiti: public annotation of multimedia content. In *CHI '04 extended abstracts on Human factors in computing systems*, pages 1207–1210, 2004.

[2] J. Favela, M. Rodriguez, A. Preciado, and V. Gonzalez. Integrating context-aware public displays into a mobile hospital information system. *Information Technology in Biomedicine, IEEE Transactions on*, 8:279–286, 2004.

[3] Greenberg, Boyle, and Laberge. Pdas and shared public displays: Making personal information public, and public information personal. *Personal and Ubiquitous Computing*, 3:54–64, Mar. 1999.

[4] S. Greenberg and M. Boyle. Moving between personal devices and public displays. In *Workshop on Handheld CSCW*, Nov. 1998.

[5] J. Han. Low-cost multi-touch sensing through frustrated total internal reflection. In *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 115–118, New York, NY, USA, 2005. ACM Press.

[6] E. M. Huang and E. D. Mynatt. Semi-public displays for small, co-located groups. pages 49–56, 2003.

[7] G. Klinker and F. Echtler. 3d visualization and exploration of relationships and constraints at the example of sudoku games. Technical Report TUM-I0722, Technische Universität München, Department of Computer Science, Nov. 2007.

[8] R. Lee and G. Greenspan. Jigsawdoku. `http://www.jigsawdoku.com`, 2008.

[9] S. Lee, W. Buxton, and K. Smith. A multi-touch three dimensional touch-sensitive tablet. In *CHI '85: Proceedings of the ACM Human Factors in Computing Systems Conference*, pages 21–25, San Francisco, California, USA, 1985. ACM Press.

[10] S. Nestler, , and G. Klinker. Towards adaptive user-interfaces: Developing mobile user-interfaces for the health care domain. In *Mobiles Computing in der Medizin (MoCoMed)*, Augsburg, Germany, September 2007.

Using Mobile Phones to Spontaneously Authenticate and Interact with Multi-Touch Surfaces

Johannes Schöning
Institute for Geoinformatics
University of Münster
Robert-Koch-Str. 26-28
48149 Münster, Germany
j.schoening@uni-muenster.de

Michael Rohs
Deutsche Telekom Laboratories
TU Berlin
Ernst-Reuter-Platz 7
10587 Berlin, Germany
michael.rohs@telekom.de

Antonio Krüger
Institute for Geoinformatics
University of Münster
Robert-Koch-Str. 26-28
48149 Münster, Germany
antonio.krueger@uni-muenster.de

ABSTRACT
The development of FTIR (Frustrated Total Internal Reflection) technology has enabled the construction of large-scale, low-cost, multi-touch displays. These displays—capable of sensing fingers, hands, and whole arms—have great potential for exploring complex data in a natural manner and easily scale in size and the number of simultaneous users. In this context, access and security problems arise if a larger team operates the surface with different access rights. The team members might have different levels of authority or specific roles, which determines what functions they are allowed to access via the multi-touch surface. In this paper we present first concepts and strategies to use a mobile phone to spontaneously authenticate and interact with sub-regions of a large-scale multi-touch wall.

Categories and Subject Descriptors
H.5.2 [**User Interfaces**]: Input devices and strategies, interaction styles.

Keywords
Multi-touch interaction, frustrated total internal reflection, large displays, mobile devices, input strategies, authentication, emergency scenario, CSCW.

1. INTRODUCTION & MOTIVATION
Multi-touch interaction with computationally enhanced surfaces has received considerable attention in the last few years. The rediscovery of the FTIR principle, which allows for building such surfaces at low cost, has pushed the development of new large-scale multi-touch applications fast forward. These walls are well suited for multi-user collaboration with large data sets, such as geographical or time-stamped data. In scenarios with large surfaces (i.e. more than 2 meters) and large groups of users (i.e., more than two) controlling access to content and functionality made available through the multi-touch surface is often an important requirement. However, although FTIR allows identifying a large number of contact points on the wall, it does not discriminate between different users. This makes it difficult to control who is issuing a command. This can lead to severe security problems if the multi-touch wall is used for triggering real-world events, as is the case in control room scenarios. For example, in an emergency response to a flooding event (cf. [9]), where a team of experts needs to coordinate mobile forces on the ground (e.g., fire brigades) and monitor data on a geographical representation (e.g., flood level and degree of pollution of air and water), not all users should be able to manipulate all data presented on the multi-touch wall. Depending on the particular policy, only the commander of the fire brigade forces might be allowed to send a mobile unit to a new target (e.g., by pointing to the unit and the new destination). Authentication concepts known from desktop computing are not well suited for these settings, since they usually grant access to an application or the whole computer, rather than to a local area of the screen.

In this paper we are addressing the problem that in some collaborative work situations the group of users of a multi-touch wall varies greatly in competence, hierarchical level, and decision-making authority, demanding a dedicated authentication and access mechanism for small regions of a multi-touch surface. We present a first solution for how to authenticate a user who wants to interact with a sub-region of a multi-touch wall. We present novel concepts that enrich the interaction with multi-touch surfaces by using a personal mobile device to spontaneously authenticate and interact with the multi-touch wall.

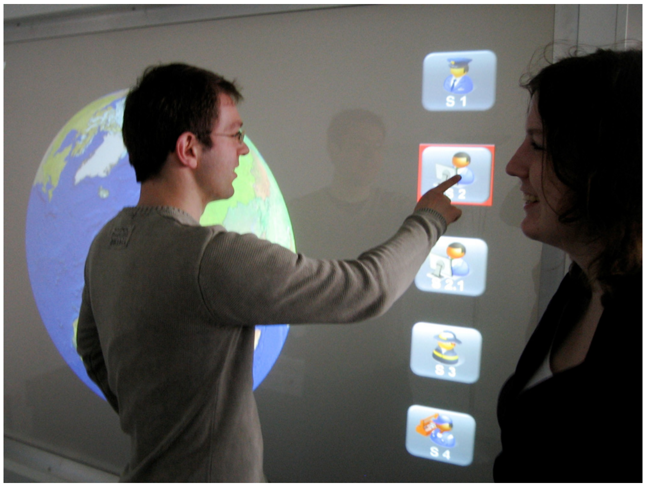


Figure 1: Multi-user interaction with a multi-touch wall in an emergency scenario without dedicated access control: The user is selecting an authentication level by pressing a button representing a certain role.

The paper is structured as follows: First we briefly give an overview of related work. In Section 3 we introduce an authentication concept using the flashlight and Bluetooth unit of a mobile device as response channels. Due to the fact that we did not yet run user tests on the interaction we discuss some possible variants of the basic concept, which we intend to evaluate in the future. We also present more general ideas for how to enrich the functionalities of a large scale multi-touch wall using mobile devices in an emergency setting. In Section 3 we briefly summarize the state of implementation. In the last section we present our conclusion and ideas for future work.

2. RELATED WORK

Collaborative visualization and decision-making tools for crisis response has been a classical field of the Digital Cartography, Visualization and GIS communities. In addition, other disciplines, such as the HCI and Ubiquitous Computing communities, have tried to tackle various aspects of this problem. Most of the existing work focuses on large format map applications that support decision-making, for example, in an emergency operation center (EOC). McEachren [5] et al. provide a good overview of these large format map applications that support collaborative visualization and decision-making. The GIS wallboard [3] is a conceptual example of an electronic white board envisioned to support sketch-based gestures to interact with geospatial data. Sharma et al. [8] concentrate on multi-modal interaction (speech and gesture) with a large dynamic map display and evaluated that system in a crisis response scenario with real users. All this work concentrates on supporting decision-making and group collaboration in an EOC, but does not concentrate on the problem of multi-user interaction with different levels of authority. An interesting alternative to classical input devices, like mice and keyboards, especially in emergency scenarios is multi-touch technology, which allows multi-finger and bimanual operation [1], because in such scenarios users have to make large-scale decisions very quickly and definitely. Several hardware solutions exist that allow the realization of multi-touch input on surfaces of different sizes. Buxton[1] gives a thorough overview of current technologies as well as the history of multi-touch surfaces.

Jeff Han presented the original FTIR multi-touch sensing work in February 2006 at the Technology Entertainment Design (TED) Conference [4]. This technology has the advantages in that it can be constructed from readily available components, is cheap and can be scaled without problems to a large scale multi-touch wall. Using this technology, multi-touch surfaces can be easily integrated into EOC where users often interact with geospatial information. However, FTIR surfaces just detect touch events and do not provide the identity of the users, per se. If multi-touch applications need to distinguish between different users, the *Diamond Touch* concept from MERL [2] could be used, with the drawback that the users either need to be wired or stay in specially prepared locations. Because an EOC is a very dynamic work setting and users have to be flexible and switch between different work stations, such a technology is not useful for an emergency scenario. We have determined that the benefits of using FTIR far outweigh the disadvantage that it does not identify users.

Mayrhofer et al. [6] present a method for establishing and securing spontaneous interactions on the basis of spatial references which are obtained by accurate sensing of relative device positions. In their work they implemented an interlocked protocol using radio frequency messages and ultrasonic pulses for verifying that two devices share a secret.

3. USER IDENTIFICATION & AUTHENTICATION

As already motivated in the introduction, collaborative work at a multi-touch surface often involves users with different roles, competencies, and scopes of expertise. In an emergency response scenario, for example, a media contact person may be allowed to visualize statistical data on the wall to get an up-to-date picture of the situation, while only the officer-in-charge may command emergency troops at the real emergency site. It would thus increase safety and security if the system could distinguish between users or if individual input events could be authenticated. This would also help in a later analysis of the events that took place, since critical operations could be attributed to individual users.

Even in such a scenario we would like to retain the direct-touch interaction scheme of FTIR multi-touch surfaces as much as possible. We assume that most interactions are allowed for every user and that only a small subset of interactions are critical, e.g., because they trigger external real-world events such as sending troops to a specific position. It therefore seems to be acceptable if these critical operations require a slightly higher interaction effort than the other operations.

The minimum requirement to support the above scenario is to identify the user who generates the critical input event. The system could then check whether the identified user is

[1]http://www.billbuxton.com/multitouchOverview.html (2008)

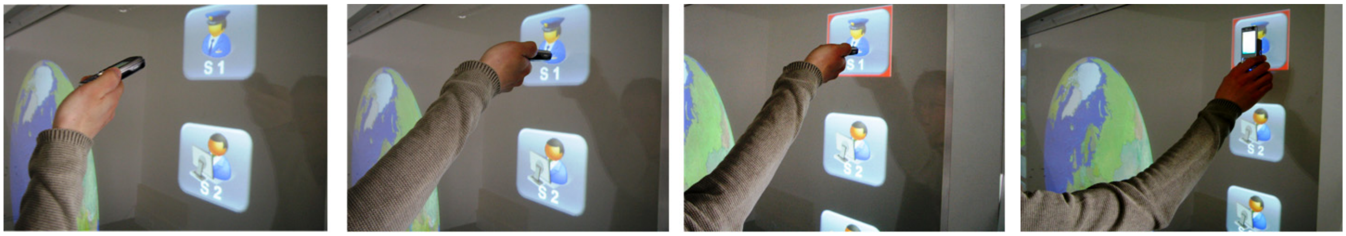


Figure 2: Interaction scheme to authenticate with a specific user role on an FTIR multi-touch surface: (i) The user touches the wall with the phone. (ii) The mobile phone flash light sends a light flash (or a camera flash) to indicate the region the user wants to interact with and at the same time initializes the authentication process. (iii+iv) The user can interact in his/her assigned role with the wall and do critical actions.

authorized to trigger the associated action. A better solution would be to also cryptographically authenticate the user attempting the input action instead of mere identification. Of course, it would be best to continuously authenticate each individual contact point, e.g., each contact point during a dragging operation. However, this is not possible given bare finger input and current FTIR technology. It is also not necessary for enabling scenarios like the one outlined above. A solution in which a user "logs in" to a small region in order to gain exclusive access to the region until the user releases that region again does not seem to be adequate, because we assume that, in general, quick access to all parts of the multi-touch surface is required.

We therefore propose to identify—and if possible also authenticate—users in the case of critical operations by using a mobile device as a mediator. We assume that the device contains a flash light and Bluetooth connectivity, and is able to detect touch events with an integrated microphone or accelerometer. We further assume that the FTIR system has a second camera that detects light flashes in the visible range. The basic identification scheme (without cryptographic authentication) works as follows:

1. The user touches a region of the wall with the phone.
2. The phone detects the touch event with its built-in accelerometer or microphone and generates a light flash. Simultaneously it sends the user ID via Bluetooth. (Optionally, microphones can be installed at the multi-touch surface as proposed in [7] to determine the position of touch event on the surface.)
3. The surface detects the light flash at a certain position and receives the user ID via Bluetooth. The light flash can be distinguished from finger touch events, because it produces a bright light strobe in the visible range, whereas finger touch events are detectable mainly in the infrared range.
4. The surface either detects the light flash first or receives the user ID via Bluetooth first. Both events have to be received within a short time window Δt. If either one is missing or if they are more than Δt apart, the protocol is aborted. If more than one flash event and one ID event are detected during a time window extending from Δt before the first event and Δt after the second event, this is considered as a collision.
5. If a collision was detected the server asks one of the devices that have sent an ID to repeat the procedure. Here also random backoff procedures could be used to resolve the collision, in which the device waits a random amount of time before a retransmission is attempted (c.f. Ethernet media access).
6. If a unique association of position and user ID is found the server looks up the authorization data for the object at the respective position and checks whether the user is allowed to perform the action. If so, a positive response is sent via Bluetooth and the action is executed. In addition, visible feedback on the region is given to indicate success or failure.

The above algorithm uniquely identifies input events on individual regions, even with multiple simultaneous users generating finger input events and multiple users generating phone touch events. If a user touches some other object this will generate only a Bluetooth ID event, but no flash event will be detected by the surface, so the algorithm will abort or a collision with another user will happen. The algorithm is guaranteed to uniquely associate user identities to regions if both events are generated and sensed within Δt.

A shortcoming of this algorithm is that it is not cryptographically secure. An attacker could forge a user ID and thus execute unauthorized critical operations on behalf of another user. We identified the following requirements for an algorithm that authenticates input on a sub-region of the multi-touch wall to support the above scenario:

- The main goal is to ensure that critical operations are only executed by authorized users. The authentication scheme thus has to prove the identity as well as the input position of the user who attempts the operation.
- The system should log all critical interactions for later analysis and documentation. Ideally, the system should also ensure non-repudiation of critical interactions. It should be possible to reconstruct who was responsible for which interaction.
- The system should allow for easy and spontaneous authentication without requiring too much effort and without interfering with other simultaneous users who perform non-critical operations.

Figure 3: General interaction scheme to identify a user with a certain area on an FTIR multi-touch surface: (i) The user touches the wall with the phone. (ii) The mobile phone flash light sends a light flash (or a camera flash) to indicate the region the user wants to interact with and at the same time initializes the authentication process. (iii+iv) The user is identified can interact in his/her assigned role with the wall and do critical actions. The more detail scheme is described in the body of that paper.

With Bluetooth we have a high bandwidth connection but we cannot determine the position on the multi-touch surface where the user actually touched the surface. With the flash light we have a very low bandwidth data channel and way to detect the input position. We assume that the multi-touch surface server and all mobile devices that are allowed to interact with the surface have a pair of cryptographic keys—a public key, a private key, and a corresponding certificate.

We propose the following preliminary authentication scheme. In order to prevent forging, the user ID is signed with the private key of the mobile device before sending it to the server. To prevent replay attacks a timestamp and a sequence number are included in the authentication request. The authentication protocol proceeds as follows:

1. The user touches a region of the wall with the phone.

2. The phone detects the touch event with its built-in accelerometer or microphone and generates a light flash. Simultaneously it sends the message m via Bluetooth:

   $$m = enc(R', pubKey_{server})$$

   with

   $$R' = (R, sign(hash(R), privKey_{device}))$$

   $$R = (opcode, userID, time, seq.nr., rand.delay)$$

   $$opcode = inputrequest$$

   We assume that only the device knows $privKey_{device}$ and thus only it is able to generate a valid "input request" message.

3. The surface detects the light flash at a certain position and receives m via Bluetooth. If the content of m cannot be verified it is discarded. Verification includes the signature, the timestamp, and the sequence number for that device.

4. As above, if more than one flash event and one ID event are detected during a time window extending from Δt before the first event and Δt after the second event, this is considered as a collision.

5. As above, if a collision was detected the process is repeated.

6. As above, authorization is performed and feedback is given accordingly.

We assume that a valid signature of the message sent via Bluetooth can only be generated by the device containing the private key. Therefore the server can be sure that a successfully verified ID stems from an authentic input request. If an attacker produces or replays an input request, verification will fail at the server. However, an attacker can produce flash events. If we assume that the authentic device produces a flash event as well, the attacker can only produce a collision.

A problem occurs, if a device generates an input request, but the corresponding light flash is not detected by the surface. This could happen if a touch event is triggered while not facing the surface. In this case the light flash would never reach the surface and an attacker could produce a light flash on some random display region.

To solve this problem, a second light flash could be produced after a random delay whose duration is sent in m (see step 2 above). The attacker would then have to guess the right delay and produce the second flash at exactly the right moment. If the server detects a flash before the indicated delay, the procedure is aborted. The security of this approach depends on the accuracy with which the camera can detect the light flashes. In the current setup, the camera runs at 30 Hz, which severly limits the bandwidth of the visual channel. An obvious way to get a higher bandwidth is to increase the frame rate of the camera. We are also working on other solutions. One idea is to introduce a light back channel. A challenge could be sent by projecting a pattern on the surface next to the detected light spot. The camera of a mobile device is normally located next to the light flash and could detect the challenge and send it back to the server (signed and encrypted). This approach has the advantages that the back channel via the mobile device camera has a higher bandwidth and we can be sure that the user is actually interacting with the right sub-regions of a large-scale multi-touch wall.

For the implementation we use a Nokia 5500 with a built-in flash light and the Nokia N95 using its built-in camera flash. A camera image (recorded by a DragonFly camera with an infrared filter) of the raw camera image and the N95 touching the multi-touch surface can be seen in Figure 4.

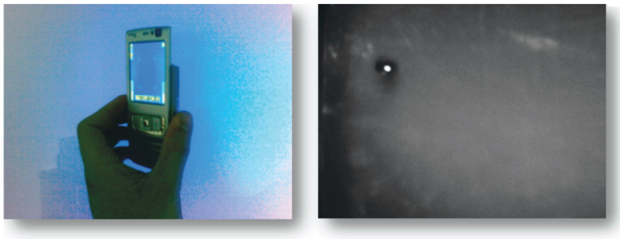


Figure 4: User is touching the multi-touch wall with a mobile device. Raw camera image of the phone flash using a DragonFly camera with an infrared filter.

4. CONCLUSIONS AND FUTURE WORK

We addressed the problem of spontaneous authentication of individual input actions in the context of large-scale multi-touch FTIR surfaces. We described an access mechanism for small sub-regions of the surface that is capable of authenticating multiple simultaneous users. Users have to touch the wall with their personal mobile device for spontaneous authentication and interaction. We still have to do user tests on the usability and general acceptability of the proposed scheme. We intend to do a formal security analysis of the method and to evaluate it with real users in an emergency operation center. As future work, we plan to add additional functionality to our prototype. As an example, while the user touches the surface, the front camera can take a photo of the user and we can verify if the right user acts with the mobile device. Other functionalities beyond the authentication problem can be easily added. For example, in our emergency response scenario, a secured voice call connection could be easily established by touching the icon of a first responder troop on the surface. We also experiment with other output modalities like the display light or, if available, the IrDA port.

5. ACKNOWLEDGMENTS

The authors would like to thank Chris Kray for fruitful discussions during CHI 2008.

6. REFERENCES

[1] W. Buxton and B. Myers. A Study in Two-handed Input. *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 321–326, 1986.

[2] P. Dietz and D. Leigh. DiamondTouch: A Multi-user Touch Technology. *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 219–226, 2001.

[3] J. Florence, K. Hornsby, and M. Egenhofer. The GIS wallboard: Interactions with Spatial Information on large-scale Displays. *International Symposium on Spatial Data Handling*, 7:449–463, 1996.

[4] J. Y. Han. Low-cost Multi-touch Sensing through Frustrated Total Internal Reflection. In *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 115–118, New York, NY, USA, 2005. ACM.

[5] A. Maceachren and I. Brewer. Developing a Conceptual Framework for Visually-enabled Geocollaboration. *International Journal of Geographical Information Science*, 18(1):1–34, 2004.

[6] R. Mayrhofer, H. Gellersen, and M. Hazas. An Authentication Protocol using Ultrasonic Ranging. Technical Report COMP-002-2006, Lancaster University, October 2006.

[7] J. Paradiso and C. Leo. Tracking and Characterizing Knocks Atop Large Interactive Displays. *Sensor Review*, 25(2):134–143, 2005.

[8] R. Sharma, I. Poddar, E. Ozyildiz, S. Kettebekov, H. Kim, and T. Huang. Toward Interpretation of Natural Speech/Gesture: Spatial Planning on a Virtual Map. *Proceedings of ARL Advanced Displays Annual Symposium*, pages 35–9, 1999.

[9] SoKNOS Project. www.soknos.de, 2008.

Multi–Touching 3D Data: Towards Direct Interaction in Stereoscopic Display Environments coupled with Mobile Devices

Frank Steinicke, Klaus Hinrichs
Visualization and Computer Graphics
Westfälische Wilhelms-Universität Münster
Einsteinstraße 62, Münster, Germany
{fsteini,khh}@uni-muenster.de

Johannes Schöning, Antonio Krüger
Institute for Geoinformatics
Westfälische Wilhelms-Universität Münster
Robert-Koch-Str. 26-28, 48149 Münster,
{j.schoening,krugera}@uni-muenster.de

ABSTRACT

In many different application domains the use of 3D visualization is accelerating. If the complexity of 3D data increases often stereoscopic display provides a better insight for domain experts as well as ordinary users. Usually, interaction with and visualization of the 3D data is decoupled because manipulation of stereoscopic content is still a challenging task. Hence, 3D data is visualized stereoscopically whereas interaction is performed via 2D graphical user interfaces. Although such *interscopic* interaction between stereoscopic and monoscopic content is of major interest in many application domains it has not been sufficiently investigated. Recently emerging multi–touch interfaces promise an alternative approach to this challenge. While multi–touch has shown its usefulness for 2D interfaces by providing more natural and intuitive interaction, it has not been considered if and how these concepts can be extended to 3D multi–touch interfaces, in particular in combination with stereoscopic display. In this paper we discuss the potentials and the limitations as well as possible solutions for the interaction with interscopic data via multi–touch interfaces.

1. BACKGROUND AND MOTIVATION

In recent years virtual environments (VEs) have become more and more popular and widespread due to the requirements of numerous application areas. Two–dimensional desktop systems are often limited in cases where natural interfaces are desired. In these cases virtual reality (VR) systems using tracking technologies and stereoscopic projections of three–dimensional synthetic worlds support a better exploration of complex data sets. Although costs as well as the effort to acquire and maintain VR systems have decreased to a moderate level, these setups are only used in highly specific application scenarios within some VR laboratories. In most human-computer interaction processes VR systems are only rarely applied by ordinary users or by experts – even when 3D tasks have to be accomplished [1]. One reason for this is the inconvenient instrumentation required to allow immersive interactions in such VR systems, i. e., the user is forced to wear stereo glasses, tracked devices, gloves etc. Furthermore the most effective ways for humans to interact with synthetic 3D environments have not finally been determined [1, 3]. Even the WIMP metaphor [14], which is used for 2D-Desktop interaction, has its limitations when it comes to direct manipulation of 3D data sets [6], e. g., via 3D widgets [7]. Devices with three or more degrees of freedom (DoFs) may provide a more direct interface to 3D manipulations than their 2D counterparts, but using multiple DoFs simultaneously still involves problems [3]. Most 3D applications also include 2D user interface elements, such as menus, texts and images, in combination with 3D content. While 3D content usually benefits from stereoscopic visualization 2D GUI items often do not have associated depth information. Therefore, interactions between monoscopic and stereoscopic elements, so–called *interscopic interactions*, have not been fully examined with special consideration of the interrelations between the elements.

Multi–touch interaction with computationally enhanced surfaces has received considerable attention in recent years. When talking about multi-touch surfaces we think of surfaces that support multi-finger and multi-hand operation (in analogy to the seminal work by Bill Buxton [5]). Multi–touch surfaces can be realised by using different technologies, ranging from capacitive sensing to video analysis of infrared or full color video images. Recently the promising FTIR (frustrated total internal reflection) technology has been rediscovered by Jeff Han [12]. Its cheap footprint has accelerated the usage of multi–touch in the last two years. If multi–touch applications need to distinguish between different users, the *Diamond Touch* concept from MERL [8] could be used, with the drawback that the users either need to be wired or stay in specially prepared locations. Another benefit of multi–touch technology is that the user does not have to wear inconvenient devices in order to interact in an intuitive way [16]. Furthermore, the DoF are restricted by the physical constraints of the touch screen. In combination with autostereoscopic displays such a system can avoid any instrumentation of the user, while providing an advanced user experience. However, the benefits and limitations of using multi–touch in combination with stereoscopic display have not been examined in-depth and are not well understood. Our experiences make us believe that mobile devices

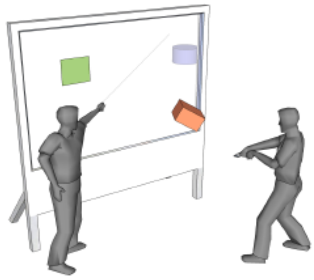


Figure 1: Illustration of two users interacting with stereo- as well as monoscopic content.

with multi–touch enabled surfaces, such as the iPhone/iPod touch, have great potential to support and enrich the interaction with large scale stereoscopic projection screens or even in immersive virtual reality. In this position paper we discuss challenges of such user interfaces for stereoscopic display setups and in particular the role multi–touch enabled mobile devices could play in those environments.

The paper is structured as follows: In section two we discuss issues related to the parallax-dependent selection and direct manipulation of 3D objects as well as issues related to navigation in 3D data sets. These issues have to be taken into account when designing a multi–touch user interface for 3D interaction. In addition, we will illustrate how the combination of a mobile multi–touch device and a stereoscopic multi–touch wall can enrich the interaction and solve existing interaction problems. Furthermore, we discuss application areas that show the potential for the interaction with stereoscopic content via multi–touch interfaces, in particular multi-touch enabled mobile devices. Section 3 concludes the paper.

2. MULTI–TOUCHING 3D DATA

In this section we discuss aspects which have to be taken into account when designing a multi–touch user interface for interscopic interaction.

2.1 Parallax Paradigms

When stereoscopic display is used each eye of the user perceives a different perspective of the same scene. This can be achieved by using different technologies, either by having the user wear special glasses or by using special 3D displays. Although the resulting binocular disparity provides an additional depth cue, in a stereoscopic representation of a 3D scene it may be hard to access distant objects [3]. This applies in particular if the interaction is restricted to a 2D touch surface. Objects might be displayed with different parallax paradigms, i. e., negative, zero, and positive parallax, resulting in different stereoscopic effects. Interaction with objects that are displayed with different parallaxes is still a challenging task in VR–based environments.

2.1.1 Negative Parallax

When stereoscopic content is displayed with negative parallax the data appears to be in front of the projection screen (see orange-colored box in Figure 1). Hence, when the user wants to interact with data objects by touching, s/he is limited to touch the area behind the objects since multi–touch screens capture only direct contacts. Therefore, the user virtually has to move fingers or her/himself through virtual objects, and the stereoscopic projection is disturbed. Consequently, immersion may get lost. This problem is a common issue known from two-dimensional representation of the mouse cursor within a stereoscopic image. While the mouse cursor can be displayed stereoscopically on top of stereoscopic objects [18], movements of real objects in the physical space, e. g., the user's hands, cannot be constrained such that they appear only on top of virtual objects.

2.1.2 Zero Parallax

If stereoscopic content is displayed with zero parallax an object appears to be aligned with the projection screen (see green-colored rectangle in Figure 1). Hence both eyes perceive the same image which causes a two-dimensional impression. As mentioned in Section 1, for such a situation multi–touch interfaces have considerable potential to enhance the interaction process, in particular when 2D manipulations are intended.

2.1.3 Positive Parallax

When stereoscopic content is displayed with positive parallax the data appears to be behind the projection screen (see purple-colored cylinder in Figure 1). These distant objects can not be accessed directly via virtual touch since the projection screen limits the reach of the user's arms. This is a common problem known from VR-based environments, and several approaches address this issue [15, 3]. Some of these approaches, in particular image plane techniques, are even applicable with multi–touch displays. When using image-based approaches, the interaction is performed on the projection screen analogous to a 2D mouse. Selection can be performed by casting a ray from the dominant eye of the user through the touch position on the screen (see Figure 1). The first object hit by the ray is the active object the user can select, e. g., by pressing a button. On a multi–touch screen even a pinch gesture can be used to perform the selection of the object underneath the fingers.

Possible Solution of Parallax Problems

One solution might be to allow a user to interactively change the parallax of objects by using a mobile device attached to the user's body as a "soft slider". If the touch–surface is portable the screen can be moved through the VE (analog to a 3D window metaphor) until desired objects are displayed with zero or negative parallax and interaction can be performed as described in Sections 2.1.2 and 2.1.3. An interesting alternative proposed by Zadow et al. [20] recognizes the positions of the user's hands not only on the surface but also above it.

2.2 3D Manipulation

3D manipulation (*selection*, *translation*, *rotation* and *scaling*) of objects on stereoscopic displays is a complex task. A major goal when performing direct 3D interaction is to manipulate an object in terms of its position and orientation in space. For two–dimensional manipulation multi–touch has proved to be a very powerful interface paradigm. Objects can be manipulated by means of a single or multiple fingers or with the palm or edge of one or both hands; even different levels of pressure can be applied [5]. However, when the user's interaction is restricted to a two–dimensional touch surface the specification of six DoF gets non-intuitive and complicated gestures may be required [9, 10, 21].

2.2.1 Selection

Before a user can interact with virtual objects the desired targets have to be identified. This task has to be accomplished by an interaction technique itself. When interaction is restricted to a 2D surface, selection can be implemented by using image-plane techniques [15, 2, 19] (see Section 2.1).

2.2.2 Translation

When a 3D object is selected and translation is intended, a movement in the plane parallel to the surface can be implemented easily. For example, contacts on the projection screen's local x and y direction can by mapped one-to-one to the virtual object. Translations are constrained to the orientation of the touch–screen surface. Since perspective projection is usually applied when stereoscopy is used, this mapping may be disadvantageous because distant objects appear to move more slowly in image-space than objects close to the projection screen. Therefore, different mapping strategies may be applied, for instance, a projected distance can be mapped [19]. However, when translation along the z direction of the screen's coordinate system is desired, different approaches have to be considered. For instance, gestures can be used to specify a translation along the depth axis, but users need to learn different non-intuitive gestures.

2.2.3 Rotation

Rotation in 2D can be implemented very naturally. For instance, one touch point determines the center of rotation, while the amount of rotation is specified by circular movements around the center. Thus objects can be rotated via two contacts only. In 3D the center of rotation and the rotation axis have to be determined by means of a 3D point and a vector. Since a touch surface constrains the user's action to 2D, rotations in 3D are difficult to realize.

2.2.4 Scaling

While scaling in 2D can be implemented very intuitively using a multi–touch interface, for example, by means of one– or two–handed pinch gestures, scaling in 3D is complicated. In particular, if non–uniform scaling is intended, an intuitive specification of the scaling vector to be applied to the virtual object is a challenging task. Even in VR-based environments non–uniform scaling is often implemented via indirect approaches, e. g., GUI widgets.

Approaches for Multi–Touch 3D Manipulation

Position and orientation of mobile multi–touch surfaces can be tracked very accurately and could therefore be used for specifying fine-grained input data. The orientation of the device could be used to provide precise data, in particular 3D vectors which could otherwise not be specified by the rather coarse multi–touch input alone. Such separation between precise and coarse interaction performed with the dominant and non-domination hand, respectively, is also applied in 2D multi–touch interfaces [4]. Likewise translation in space can be implemented by using the mobile device's orientation that determines the axis along which a translation of an object is to be performed. In the same way the device's orientation can define a rotation axis or a non-uniform scaling vector.

2.3 Navigation

Since navigation is the most common interaction task in VEs it is essential to provide intuitive mechanisms to enable users to explore large and complex virtual environments. Essentially navigation is similar to performing 3D object manipulation, whereas when exploring a VE manipulations are applied to the virtual camera. Current navigation techniques exploiting multi–touch devices are limited to simple panning, zooming or rotation approaches [12]. Usually, the non-dominant hand poses a predefined gesture that determines the navigation mode, while the dominant hand specifies the amount of movement. Since the touch is only used to define certain modi multi–touch is actually degraded to single touch. It has not been examined how multi–touch, for instance by using the entire hand surface, can enhance this process.

Possible Solution of Navigation problems

For single touch interfaces there are already intuitive mechanisms to implement certain camera movements [11]. Such traveling metaphors can be realized by means of specifying direction, speed, velocity, etc. of the virtual camera. As mentioned above mobile devices equipped with orientation sensors may be exploited to define the orientation of the virtual camera. All movements of the camera may be determined by the touch interface of the mobile devices. This is especially beneficial for presentation scenarios, where the presenter is using a mobile device to guide other viewers through a VE. Alternatively the touch surface itself can be used as a navigation device. Camera movements can be initiated by virtually pushing the touch screen. For instance, pressing the touch screen on the right side yields a camera rotation to the left, touching the screen at the top moves the camera downwards and vice versa. Furthermore, these concepts can be combined such that an omni-directional flying metaphor can be implemented. Hence the user gets the impression of navigating a vehicle via the window to the virtual world.

3. CONCLUSIONS

In this position paper we have discussed problems and potentials related to the use of multi–touch interfaces for the interaction with interscopic data. Figure 2 summarizes, from our point of view, the potentials of multi–touch interfaces for the interaction with stereoscopic content as well as the possibilities when using multi-touch enabled mobile devices. We are working on the realization of such a system for formal evaluation. The icons indicate whether we believe that an interaction in this combination is beneficial (green/smile), possible (yellow/neutral), or impracticable (red/sad). Of course, not all problems are covered or can be solved with such a device setup. We have mentioned some problems which might occur in such scenarios.

3D widgets can be used to integrate solutions for desktop-based environments [7]. If direct interaction is not required, users can specify 3D manipulations by means of constraint-based techniques. These widgets provide several interaction handles which themselves support different interaction tasks such as translation, rotation or scaling. The available DoFs are reduced with respect to the degrees provided by the input device. Currently multi–touch walls are horizontally or vertically mounted. VR-based display devices such as the responsive workbench allow to turn the display from horizontal to vertical. In contrast to vertical multi–touch surfaces, horizontal ones provide the possibility to place physical objects on the surface [13]. In order to enhance the perception

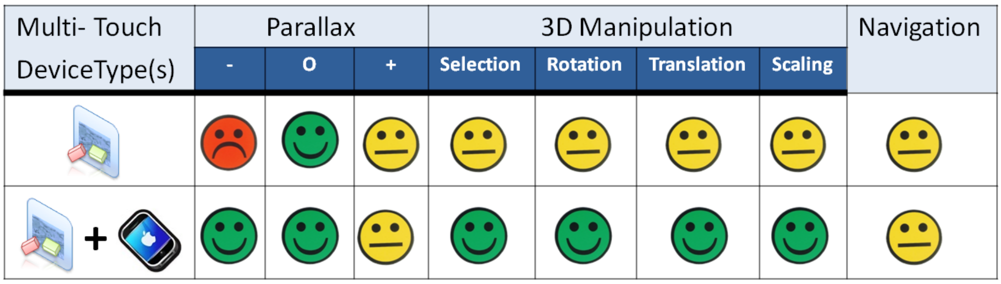


Figure 2: Potentials and limitations as well as possible solutions for using multi–touch interfaces with/without mobile devices (having a multi–touch enabled surface) to interact with stereoscopic content.

of spatial data 3D multi–touch or at least 2.5D projection screens can be exploited [17].

ACKNOWLEDGEMENTS

We thank the Deutsche Telekom Laboratories, which partially funded this research.

4. REFERENCES

[1] R. Balakrishnan. A Grant of 3D. Keynote speach Symposium on 3D User Interfaces, 2006.

[2] D. Bowman. *Interaction Techniques for Common Tasks in Immersive Virtual Environments: Design, Evaluation, and Application*. PhD thesis, Georgia Institute of Technology, 1999.

[3] D. Bowman, E. Kruijff, J. LaViola, and I. Poupyrev. *3D User Interfaces: Theory and Practice*. Addison-Wesley, 2004.

[4] P. Brandl, M. Haller, J. Oberngruber, and C. Schafleitner. Combining and Measuring the Benefits of Bimanual Pen and Direct-Touch Interaction on Horizontal Interfaces. In *Proceedings of Advanced Visual Interfaces*, page accepted, 2008.

[5] W. Buxton and B. Myers. A study in two-handed input. *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 321–326, 1986.

[6] W.-S. Chun, J. Napoli, O. S. Cossairt, R. K. Dorval, D. M. Hall, T. J. P. II, J. F. Schooler, Y. Banker, and G. E. Favalora. Spatial 3-D Infrastructure: Display-Independent Software Framework, High-Speed Rendering Electronics, and Several New Displays. In *In Stereoscopic Displays and Virtual Reality Systems XII*, volume 5664, pages 302–312, 2005.

[7] D. B. Conner, S. C. Snibbe, K. P. Herndon, D. C. Robbins, R. C. Zeleznik, and A. van Dam. Three-Dimensional Widgets. In *Symposium on Interactive 3D Graphics*, 1992.

[8] P. Dietz and D. Leigh. DiamondTouch: a multi-user touch technology. *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 219–226, 2001.

[9] J. Epps, S. Lichman, and M. Wu. A study of hand shape use in tabletop gesture interaction. In *CHI '06: CHI '06 extended abstracts on Human factors in computing systems*, pages 748–753, New York, NY, USA, 2006. ACM.

[10] C. Forlines and C. Shen. Dtlens: multi-user tabletop spatial data exploration. In *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 119–122, New York, NY, USA, 2005. ACM.

[11] M. Hachet, F. Decle, S. Knödel, and P. Guitton. Navidget for Easy 3D Camera Positioning from 2D Inputs. In *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI)*, pages 83–89, 2008.

[12] J. Y. Han. Low-cost multi-touch sensing through frustrated total internal reflection. In *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 115–118, New York, NY, USA, 2005. ACM.

[13] Microsoft Surface. `http://www.microsoft.com/surface/`, 2008.

[14] B. A. Myers. A taxonomy of window manager user interfaces. *IEEE Comput. Graph. Appl.*, 8(5):65–84, 1988.

[15] J. Pierce, A. Forsberg, M. Conway, S. Hong, R. Zeleznik, and M. Mine. Image Plane Interaction Techniques in 3D Immersive Environments. In *ACM Symposium on Interactive 3D Graphics*, pages 39–44, 1997.

[16] J. Schöning, B. Hecht, M. Raubal, A. Krüger, M. Marsh, and M. Rohs. Improving Interaction with Virtual Globes through Spatial Thinking: Helping users Ask "Why?". In *IUI '08: Proceedings of the 13th annual ACM conference on Intelligent User Interfaces*, New York, NY, USA, 2008. ACM.

[17] R. Smith, B. Thomas, and W. Piekarski. Techn Note: Digital Foam. In *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI)*, pages 35–38, 2008.

[18] F. Steinicke, T. Ropinski, G. Bruder, and K. Hinrichs. Interscopic User Interface Concepts for Fish Tank Virtual Reality Systems. In *Proceedings of the Virtual Reality*, pages 27–34, 2007.

[19] F. Steinicke, T. Ropinski, and K. Hinrichs. VR and Laser-based Interaction in Virtual Environments using a Dual-Purpose Interaction Metaphor. In *VR2005 Workshop Proceedings on New Directions in 3D User Interfaces*, pages 61–64. IEEE, 2005.

[20] U. von Zadow. http://www.multi-touch.de/.

[21] M. Wu and R. Balakrishnan. Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. *Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 193–202, 2003.

Sociality, Physicality and Spatiality: touching private and public displays

Alan Dix, Corina Sas, Devina Ramduny-Ellis
Computing Department, InfoLab21
Lancaster University
Lancaster, LA1 4WA, UK

Steve Gill, Joanna Hare
Cardiff School of Art & Design
University of Wales Institute, Cardiff
Wales, CF5 2YB, UK

alan@hcibook.com, {corina, devina}@comp.lancs.ac.uk, {sjgill, juhare-pdr}@uwic.ac.uk

http://www.hcibook.com/alan/papers/PPD08/

ABSTRACT
This paper considers two strands of research that each contributes to an understanding of touch-based interaction with private and public displays. The first is based on general frameworks for private device–public display interaction, which is driven by the growing body of work in the area, but focuses on the level of integration of public and private devices and the importance of understanding social setting and bystanders. The second strand is centred on physicality; how different kinds of physical device impact interaction and how modelling of touch-based devices causes particular problems that require notations and formalisms of continuous and bodily interaction.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation (e.g., HCI)**]: User Interfaces – *graphical user interfaces, interaction styles.*

General Terms
Design, Human Factors

Keywords
public displays, touch interaction, spatial interaction, physicality

1. INTRODUCTION
This position paper is drawing from two strands of work involving studies, models and frameworks for understanding:

(i) interaction with personal devices and public displays

(ii) physicality and spatiality of human interaction

For the first of these we will draw on our own experience and analysis and also from a recent workshop at CHI2008 [15]. For the second we will draw on ongoing studies and formal analysis, and also work of the DEPtH project and its associated Physicality workshops (http://www.physicality.org/).

Our framework for personal device–public display interaction covers various dimensions, but here we will address two in particular: the level of integration between devices and the social setting. Similarly physicality covers many issues, but we will focus on two of these: the issues of space and movement, and bodily interaction with the devices. We are partly presenting some of our work that is relevant to the issue of touch-based interaction with private and public displays, but doing so in the knowledge that our models and frameworks need to be adapted in order to address these emerging technologies.

2. MOVEMENT AND CONTACT
In the context of this workshop there are two obvious kinds of touch (depicted in figure 1):

(i) touching a display in a private device

(ii) touching a public display

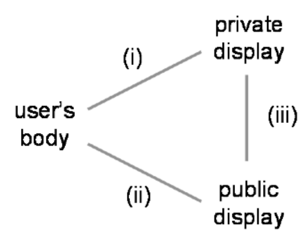


Figure 1. Ways to touch and connect

One of the aspects that emerged from the recent CHI workshop was the ways in which mobile phones could be used to gesture and move in multiple kinds of space:

- body-relative space – For example, using the accelerometers built into some phones.
- walking / absolute space – For example, using GPS tracking or Bluetooth signal strength location techniques.
- screen-relative space – Where the phone is positioned near or on the screen.

One example uses phone with built-in Near Field Communication (NFC) tag readers (like RFID), so an array of tags are placed behind a screen onto which a map and interactive content is projected, and the phone is touched against the display in order to select content [12]. This suggests that, as well as the direct physical touch of a finger (or other part of the body) on a public screen, we should also consider indirect touch using the device itself. This is shown as link (iii) in figure 1. There are other technologies for achieving the same effect including visually tracking the phone or using the phone's camera to detect visual codes (e.g. [14]).

In a public setting there can be several advantages to this form of indirect touch. In a restaurant or busy place personal hygiene may be important, so the act of physically touching a screen that others have also touched or perhaps be dirty may not be acceptable. On the other hand, if the users' hands are expected to be dirty we may not wish them to dirty the screen (greasy fingers in a fish and chip shop!).

In addition, the use of a proxy device effectively creates a very clear minimum granularity for selection. This can be a problem if fine section is needed, but sometimes may be advantageous especially where the tracking mechanism is not accurate and a more direct interaction might encourage incorrect expectations.

This form of proxy interaction does not readily admit straightforward multi-touch interaction as the device itself makes a single point of contact. However, one can imagine various forms of multi-user multi-touch interaction where several users cooperatively use their personal devices. Also in the NFC tag system described above, the user combines touching the phone against the screen with keypad-based interactions. It is easy to imagine systems that combine placing a personal device against a public screen and then simultaneously using a (probably single touch) finger interaction on the device screen whist moving it across the public screen. For example, placing a photograph on a public display where the position is indicated by the device location and finger gestures are used on the device display for sizing and rotating .

3. LEVEL OF INTEGRATION

When considering multi-display interactions, one of the first dimensions to bear in mind is the level of coupling between the public and private displays.

alternative interface (no coupling) – For example, a public display may show the same news feed as is available on a mobile phone. In the Hermes system at Lancaster, small screens are placed beside office door. Visitors leave messages on the doorplate, which the door owner can subsequently read *either* on the door plate itself *or* via a web interface [3].

secondary interface (weak coupling) – The Hermes web interface or its SMS interface can be used to update the display that is subsequently seen by someone at the door. Although both displays are clearly part of a single interaction, they function as two single display systems interacting with a common information store.

coherent interface (strong coupling) – In a public photo display developed as part of the CASIDE project at Lancaster, users can navigate using the phone to find an image and then upload it to the screen, so this feels like a single interaction [4].

The proxy interactions in the previous section are an extreme form of coherence as the two displays are not just digitally, but physically brought together. More commonly coherent interaction involves using the personal device for input and maybe some personal feedback. Controlled experiments on distributing interfaces over public and private devices have confirmed more widespread deployment experience. They have shown that the impact of combining the public and private displays can indeed increase interaction efficiency in terms of task-completion time, and also increase satisfaction in terms of perceived ease of use and speed [10]. However, the qualitative analysis of these experiments revealed that switching of attention could be problematic.

4. SETTING AND AUDIENCE

Public displays by definition are in public spaces where there are likely to be other people around as well as those directly interacting: some watching the display, others totally unaware of its existence.

Urban artistic performances, such as street theatre, similarly include members of the public with various levels of engagement and an analysis of these events [6] divided people into several categories: performers, witting and unwitting participants and witting and unwitting bystanders. In non-artistic setting there is no 'performer', but we do find the other categories:

unwitting participant – triggers sensors to have some effect, but does not know it

participant – actively engaged with the system doing some form of input/interaction

unwitting bystander – sees the screen but does not realise interaction is occurring

witting bystander – sees the screen and realises interaction is occurring

passer-by – may know the screen is there, but does not watch or interact with it

These categories clearly allow many possibilities. Figure 2 looks at some of these combinations, focusing on active/witting participants and "bystanders" (this general heading includes unwitting and witting bystanders and passers-by). Here are some of the issues that can arise in each combination.

		audience	
		no bystanders	some bystanders
active participants	none	turn off display?	standard broadcast
	1	individual multi-display	public/ individual conflicts?
	2 or more	collaborative or interfering?	ditto + are group themselves part of 'display'

Figure 2. Interactions between participants and audience on public screens

The above table can be interpreted in two ways (i) as a set of possibilities of a particular system, *what may happen* and (ii) at any particular moment, *what is happening*. So a particular system may allow multiple active participants and an audience but at a specific moment there may be one or no participants, or no audience. Often it is the momentary situation (ii) that is crucial, but in some case the dynamics is significant – it is the fact that the use of a particular display moves between situations that can be important.

In particular we may want to encourage people to use a public display, what Brignull and Rogers [2] call the 'honeypot effect', enticing people from being passers-by to being active participants.

If active participants are seen to be actively interacting with a public interface, then this may encourage bystanders to (a) become aware that the display is interactive, i.e. move from being an unwitting to a witting bystander and (b) be encouraged to interact themselves, i.e. transition from witting bystander to participant.

To encourage these transitions, interactions (ii) and (iii) from figure 1 are particularly important. However, even individual interaction with a personal display (link (i) in figure 1), while in a sense is still 'private', in that others cannot see the display, is nonetheless 'public', in that others may see that the individual is interacting. For example, the active participant may be standing in a pose that suggests interaction with the screen or may be shifting gaze to and from the personal device and public display. Depending on the balance between privacy and desire to engage bystanders, fine choices of interface design may be able to subtly change the 'performance' of using the device.

5. PHYSICALITY OF DEVICES

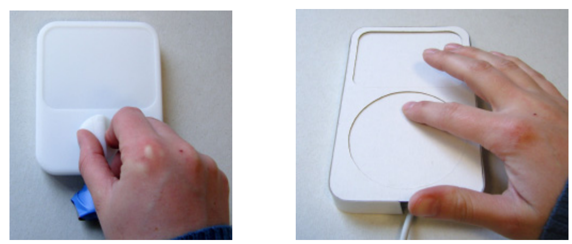
Two of the authors are product designers, part of a research group attempting to create a suite of systems for the development of computer embedded products sympathetic to the designer's mindset and methods. In particular they have been using low-tech keyboard emulation boxes called IE Units alongside software building blocks allowing rapid prototyping without electronics or programming skills. [8]. There are a number of other groups working in this area including Phidgets [11], DTools [13], and Switcharoo [1], although these mostly come from a computing or electronics background. The IE system has been used to empirically measure the performance of real products against physical and virtual prototypes and this research found that the link between the physical act of holding a product and interaction was more marked than has previously been understood [8].

In the context of touch-based interaction one particular series of experiments was most interesting. Mobile phone prototypes were produced at various levels of fidelity: from a real handset with solely the display rendered on screen to a completely screen-based emulation. For the purpose of the experiments, the separate screen was intended as an emulation of an 'on device' screen but in the context of this position paper, it effectively became a personal device interaction with a larger (although private) display. Instead of a smooth change in user responses to the gradually less physical prototype, a sharp change was observed. The 'break point' was reached when the keyboard became smooth (paper over a soft keypad). While an emulated keyboard is not the same as a touch-based screen, still this suggests the physical impact of not having tactile elements is significant to interactive experience.

In section 3, we noted that experiments on distributing information between public and private displays could lead to problems related to switching of attention [10]. The participants in these experiments were young (average age 30), but for older users switching attention between hand-held devices and distant screens (e.g. remote control and TV) can be difficult and for some people may even require changing spectacles. Even feeling for buttons such as cursor keys can be problematic for older users and so what may appear to be a 'heads-up' interaction actually involves switching visual attention. While this sounds like any form of multi-device interaction with displays is more problematic for older users, it can also be seen as an opportunity for touch-based interaction as the gains of subtle tactile feedback often disappear. Speculating, larger touch-based devices for interacting with remote displays may have advantages in this context, not just for public displays but also in the home.

We have in addition been involved in formally modelling the nature of physical interaction. In particular we take a stance where we separately model the device's physical states and interaction effectively 'unplugged', i.e. totally ignoring any digital functionality and then as a separate exercise, map this to digital behaviour [7]. We have modelled a variety of consumer devices using separate state transition networks (STNs) to model the physical and digital/logical aspects of the devices. While many devices have quite complex mappings, there are some simple devices, such as some light switches, where there is a one-to-one correspondence between physical device states and logical states (in the case of the light switch electricity flowing). These simple mappings, which we call *exposed state* devices, are particularly easy to understand as the device itself physically encodes everything.

This technique has been applied to device prototypes within the product design setting using IE units. Alternative physical devices have been developed for the same digital functionality – a media player. Figure 3 shows two of these devices: (i) is a dial with exposed state and (ii) is touchpad. The latter is of course similar to many touch-based personal devices.



(i) dial (ii) touchpad

Figure 3. Physical prototypes

The raw physical model of the touchpad is in fact trivial – there is no visual (or tactile) difference between states in the device itself. In addition, while your finger moves over the surface, there is no intrinsic haptic feedback as it traverses critical regions (in this case changing menu selection on the media player). As with mouse-based interaction, users have to use their imagination in order to construct the virtual world behind the device. It is perhaps odd that touch-based interaction, which, on the one hand, is far more physical than pressing keys, on the other hand, it has less tactile feedback.

6. MODELS AND ARCHITECTURE

So far there appears to be little systematic modelling or user interface architecture for interactions between personal devices and public displays, although there is certainly interest in the area.

In the *single* device modelling above, we chose STNs to model the physical device as these are well understood in computing science and even used in end-user documentation. However, we were aware from previous work on status–event interaction that a purely discrete notation would have limitations [5]. Indeed this has turned out to be the case and detailed analysis of even simple switches requires such extensions to describe the 'bounce' found when one initially tries out the switch to see which ways it moves.

Again even in simple switches, we have found that a thorough analysis really requires, at least simplified, modelling of the human body, in particular the forces exerted by a finger. This is even more important for touch-based devices as the device itself is stateless and the trajectory of interaction is driven by the sensing of the body alone. This is also evident in the explicit role of the human body in figure 1.

For UI architecture, there are various models for multi-user and multi-modal systems, which should be useful as they already deal with multiple input streams and non-standard inputs such as gestures. However, for public screen interactions there are also many issues relating to security and trust that need to be reflected in the architecture. Whereas in a 'normal' application, all the devices are typically owned by the user, with public displays, there are multiple 'owners' and many stakeholders.

7. SUMMARY

In this position paper, we have considered several aspects of two strands of work focused on interaction with public displays using personal devices and on the issue of physicality in design.

We have seen that 'touching' in such contexts may include mediated touch using the device itself, potentially powerful in certain types of public place. This form of proxy interaction entails a high degree of coupling between devices, although other forms exist involving either pure heads-up interaction with fingers on a personal device or interactions dividing visual attention between personal device and public screen.

An understanding of 'audience' is also important; both bystanders watching the screen and passers-by, whom we might wish to attract. So, whilst in some situations we may wish to have unobtrusive interactions in order to preserve privacy, in others, more expansive gestures may be appropriate in order to create a form of ad hoc 'performance'.

Finally, we considered the modelling of physical devices and saw how effective modelling of touch-based interaction is likely to require both notations for continuous phenomena and also modelling of aspects of the human body.

These various factors from the two strands do not yet make a single coherent view of touch-based interaction with multiple devices. However, there are threads of integration, notably the way all the topic covered inform or are informed by the different relationship sin figure 1. While this is still a developing area, we believe both the separate strands and the emerging threads between them offer initial insights.

8. ACKNOWLEDGMENTS

Work related to this position paper was supported by a donation from Nokia Global University Donations Group; HEFCE SRIF funding of the eCampus public display infrastructure at Lancaster University (http://ecampus.lancs.ac.uk/); AHRC/EPSRC funding of the DEPtH project (http://www.physicality.org/depth/), part of the Designing the 21st Century programme; and EPSRC funding of the CASIDE project (http://www.caside.lancs.ac.uk/).

9. REFERENCES

[1] Avrahami, D. and S. Hudson, Forming interactivity: a tool for rapid prototyping of physical interactive products, in: Proceedings of DIS '02. ACM, New York, (2002) 141–146.

[2] Brignull, H., and Rogers, Y. (2003). Enticing people to interact with large public displays in public spaces. Proceedings of INTERACT'03 (pp. 17–24). Zurich.

[3] Cheverst, K., D. Fitton, and A. Dix. "Exploring the Evolution of Office Door Displays", in: Public and Situated Displays: Social and Interactional aspects of shared display technologies. K. O'Hara, M. Perry, et al (Eds). Chapter 6, pp. 141-169, Kluwer. ISBN 1-4020-1677-8. 2003.

[4] Cheverst, K. Dix, A. Fitton, D., Kray, C., Rouncefield, M., Sas, C., Saslis-Lagoudakis, G. and Sheridan, G. "Exploring Bluetooth based Mobile Phone Interaction with the Hermes Photo Display", in Proc. of the seventh ACM International Symposium on Human Computer Interaction with Mobile Devices and Services (MobileHCI '05) 2005. Salzburg, Austria, pp. 47-54, Sept 2005.

[5] Dix, A., and G. Abowd, Modelling Status and Event Behaviour of Interactive Systems}. Software Engineering Journal, 11:6, (1996), 334–346.

[6] Dix, A., Sheridan, J., Reeves, S., Benford, S., & O'Malley, C. (2005). Formalising Performative Interaction. *Proceedings of DSVIS'2005*. (Newcastle, UK, 13-15 July 2005). Springer, LNCS 3941, pub. 2006. pp 15-25.

[7] Dix, A., M. Ghazali and D. Ramduny-Ellis Model ling Devices for Natural Interaction. in: Proceedings of Second Intnl. Workshop on Formal Methods for Interactive Systems, FMIS2007. ENTCS, Elsevier, (2007).

[8] Evans, M. and Gill, S., (2006) "Rapid Development of Information Appliances" in Proceedings of International Design Conference – Design 2006, Croatia, May 15 – 18

[9] Gill, S., Loudon, G., Hewett, B. & Barham, G. (2005) "How to Design and Prototype an Information Appliance in 24 Hours – Integrating Product & Interface Design Processes" The proceedings of the 6th International conference on Computer Aided Industrial Design and Concept Design conference, University of Delft, The Netherlands

[10] R. Gostner, H. Gellersen, C. Kray and C. Sas. Reading with Mobile Phone & Large Display. http://www.comp.lancs.ac.uk/~corina/CHI08Workshop/

[11] Greenberg, S. and Fitchett, C. 2001. Phidgets: easy development of physical interfaces through physical widgets. In Proceedings of the 14th Annual ACM Symposium on User interface Software and Technology (Orlando, Florida, November 11 - 14, 2001). UIST '01. 209-218.

[12] R. Hardy and E. Rukzio. Direct Touch-based Mobile Interaction with Dynamic Displays. Proceedings of Workshop on Designing and Evaluating Mobile Phone-Based Interaction with Public Displays. http://www.comp.lancs.ac.uk/~corina/CHI08Workshop/

[13] Hartman, B., S. Klemmer, M. Bernstein and N. Mehta, d.tools: Visually Prototyping Physical UIs through State-charts, in Extended Abstracts of UIST 2005, ACM Press, (2005).

[14] Rohs, M., Real-World Interaction with Camera-Phones, 2nd International Symposium on Ubiquitous Computing Systems (UCS 2004), Tokyo, Japan, November 2004.

[15] Sas, C. and Dix, A. 2008. Designing and evaluating mobile phone-based interaction with public displays. In CHI '08 Extended Abstracts on Human Factors in Computing Systems (Florence, Italy, April 05 - 10, 2008). CHI '08. ACM, New York, NY, 3941-3944.

“tune_eile”: A platform for social interactions through handheld musical devices

Nora O Murchú
Interaction Design Centre
Dept of Computer Science & Information Systems
University of Limerick
+353 61 234280
nora.omurchu@ul.ie

ABSTRACT

This paper presents ‘tune_eile’ an interactive tabletop installation for public spaces that allows the dynamic participation of users in sharing their music choice. tune_eile is a novel musical jukebox system where users generate playlist content from their own portable music players. The tune_eile jukebox explores how personal music players can be used as a mechanism for facilitating musical expression in public spaces such as bars and cafes; it explores techniques for interacting with multi touch surfaces by means of portable music devices. The jukebox allows multiple communications with portable musical devices, thus enabling users’ personal music to become part of a more social experience by allowing users to play music for those around them.

Keywords

multi-touch, music sharing, social interaction, multi-user, public spaces

1. INTRODUCTION

The human need for social connectedness signals opportunities for technology development - with particular regard to ubiquitous computing - to facilitate novel forms of cooperative use. Rapid changes in technology, combined with an increasingly mobile society, ensure that the average person is continually challenged to use unfamiliar devices. Ubiquitous computing technologies, in particular multi touch, not only enable new ways of acting and interacting, but also stimulate fundamental reassessment of the meaning of human action and interaction with technology. These technologies bring about new ways of working, completing tasks and facilitating new social behaviors between individuals.

Music has often been utilised as a medium for social interactions and has long been considered an important aspect of social environments. According to DeNora [4], [5] listening and interacting around music are an immensely important part of everyday life and an integral part of the cultural material through which social interactions can be constructed and organised [4], [5]. The digitisation of music has lead to the development of new devices and services for finding, obtaining, viewing, listening and sharing this medium, at home, at work, or on the move [2]. This change in music consumption has allowed people to listen to music anywhere. Thus the widespread availability of digital music players has opened new opportunities for social exploration and interaction around music sharing.

The basic motivation for a system such as tune_eile is that it aims to stimulate social interactions between unique individuals through the medium of music. The tune_eile system allows users to generate play list content within a public space from personal handheld musical players, and has been designed to act as a platform for encouraging social interactions and musical exploration between transient individuals in public spaces.

2. TUNE_EILE

tune_eile is a multi touch jukebox which incorporates the user's own digital music player as a source from which they can generate playlist content. In contrast to a traditional jukebox, tune_eile combines the user's mp3 player and an interactive tabletop. Standard jukeboxes are designed for single-user interaction, and do not support multiple music selection in a social setting. They also force the user to browse through an unfamiliar selection of music. In contrast, the users' personal mp3 player is familiar to them and thus does not require substantial investment of interaction time nor a departure from the social atmosphere. Hence, music that is formally within the user's private sphere is now transported into the public domain of the chosen space, which in this case is a pub/café. This emphasises and reinforces the sense of belonging in that social setting for the user.

Users place their music player on the tabletop screen and the content of their music collection becomes graphically displayed on the screen. The musical content is displayed in a circular formation of album covers' thumbnails surrounding the users music player, as illustrated in Figure 1. Users browse through the menu by touching the album art and then selecting the song they wish to hear the menu appearing below.

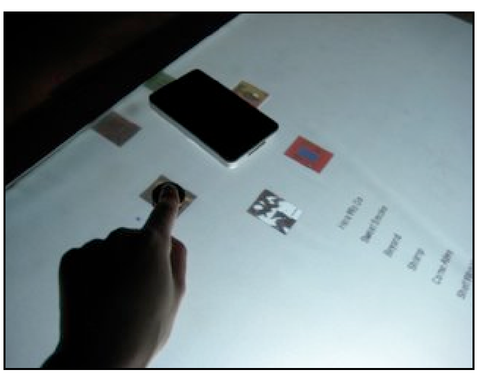


Figure 1: Graphical Interface of tune_eile

Users select the song that they wish to add to the playlist from the display. The multi-touch interface enables several users to use tune_eile concurrently without interfering with each other. This multiple access distributes control of the system, prevents individuals from taking over, and also lower thresholds for shy people. One of the main objectives of tune_eile is to create an open and unrestrictive environment, which would encourage natural social interactions between people in public spaces. The intuitive interface affords easy interaction for first time users.

By using the space in and around the music player provides a rich visual feedback for the user [3]. These graphics are also essential for the user to intuitively understand the system. When the user's music player is on the surface of tune_eile, the surrounding space becomes a individual space through which the user can browse and select music from their personal music player.

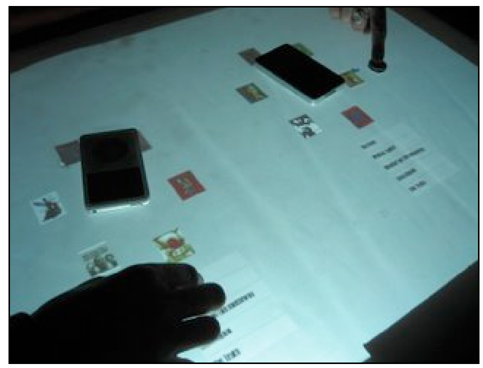


Figure 2: Multiple users interacting with tune_eile

If two or more users have their music players on tune_eile, if they share common interest in music than a graphical connection between the two will be displayed. This graphical connection can serve as a conversational prop between two users of tune_eile and thus creating a common ground that both informs and coordinates their activities. Upon selecting a song from their music player, the track selected enters into a queue. The queue operates on a first selected first played facility. If the same song is selected twice by two different users it is 'bumped up' on the playlist. This results in more popular songs being played faster.

3. USER TESTING

The testing sessions aimed to introduce unknown users to one another in an informal environment. The testing session aimed to explore whether tune_eile could be utilised in the support of social interactions between transient individuals in a public space. For the participants to act more naturally it was deemed necessary to test in an environment in which they felt comfortable in. Therefore the designers staged a pub environment in the living room of a house. A bar was set up and table and chairs were placed in different areas in the room. The tune_eile prototype was situated in the centre of the room. Participants were also free to ask for drinks from the "bar".

Figure 3: tune_eile prototype

The multi touch technology of tune_eile aims to become a catalyst to initiate and develop social networks. We designed these informal testing sessions to develop an insight into how the user would adopt the system into everyday life. Each session was videotaped. The recorded video acted as a narrative to the evaluation session and a commentary of activities. We also used questionnaires and group discussions to capture additional information. The questionnaires and discussions facilitated us in acquiring feedback on tune_eile as a system while focusing users to think about specific aspects of the system. These evaluation sessions tested the entertainment factor of the system and how that system might integrate into public environments.

These sessions had an informal tone and users were free to interrupt and ask questions at any time. Interaction between users was encouraged and at the end of each workshop group interviews were conducted where users were free to comment on their experience. The sessions were initiated with a brief introductory to tune_eile, which was followed by a discussion and concise evaluation of the interface. The session followed with a scenario and role-play performance and then concluded with a "play time" with tune_eile. An end of session group interview was conducted with all participants.

At the start of the session users were debriefed and given a detailed overview of the system and how it operated. The aim of testing was to develop the methods further, find new ideas and discover problems. The session began with a walkthrough of the user interface. Participants conducted the evaluation of the interface individually. Each participant used "thinking aloud" as they completed a simple task list. When a task was completed each user was asked to comment on their initial reactions to the task and were asked to rate its difficulty. The result of our initial user testing showed that all users reported an easy or medium level of difficulty in communicating with the interface but those who reported medium said that they "*got the hang of it after one go*". The participants considered the idea of displaying the album artwork effective and said "*it not only makes the interface look very attractive but also made it easier to find an album*", as most people are familiar with the graphics. The participants stated that the concept was "*novel and would be great to have at a house party*."

After the interface evaluation the participants were split into two groups. Each group was given two descriptive scenarios and were asked to act them out. The groups were given twenty minutes to prepare the scenarios and coordinate their activities. Each group acted out a performance. After the scenarios were performed, a reflection/ discussion session occurred. These group discussions enabled the evaluators' impressions to be captured immediately. The participation and observation of the evaluators in each scenario allowed time for reflection and idea generation. Scenarios and role playing combined with informal discussions allowed for reflection over and understanding of the system and enabled the designers to observe how social interactions could possibly take place.

The group discussion, which followed the scenario improvisation, was centered on the interface and the social interactions that took place. The participants remarked "*it was nice to see the system in a context of use and it the interactions that took place around it made sense and seemed quite close to reality*." Participants stated that the multi user aspect of tune_eile was unique and catered for all types of users (shy to confident). The participants felt that shyer users would be encouraged to use the system due to its "*open nature*" and the manner in which it facilitated more than one user at a time.

Participants agreed that the music could be a potential "icebreaker" between individuals. One participant stated that music has the capacity to "*create a new dynamic between* *people*". Music is an inherently social activity. All types of music have certain connotations and physical characteristics that suggest appropriate ways to behave, relate, and even appropriate topics of conversation. Music can embody strong personal links to times, events and other people in our lives. Shared interests are a way in which people can connect to one another, and they can sometimes act as a catalyst to encourage social interactions. Participants stated that "*people remember music and can often you can become associated with certain songs which can be a starting point in conversation*."

The final part of the workshop concluded with a "play-time" session where the participants were free to interact with tune_eile. At many times during this part of the workshop all users were gathered around tune_eile, which mirrors a typical situation in a real life pub/café. At these times it was noted that the person making the choice at a given time was the central focus of the group. Other users waited to see their selection. Users' reaction to the choice was also interesting as it was met with either approval or disapproval. Either way the selection of music created a ground for conversation and friendly teasing. Participants agreed that music selection reveals a lot about one's personal identity. The choice of music denotes certain aspects of one's perceived "self" to others present while those present, in turn, judge the musical selection to some extent.

In the testing sessions with tune_eile we were able to collect insights and feedback on its design in addition to how it might operate in a 'real life 'situation. The analysis of video recordings of the session and of the interviews indicated that the participants understood that tune_eile provides a platform for social interaction in a bar/café while also allowing them to play and select music from their own personal mp3 players. Music choice can be used to establish, reinforce or undermine group belonging and other social relationships. Hence an mp3 player can be seen as an extension of one's personality, reflecting personal music tastes and reinforcing the owner's identity. This personal choice of music by the user can be used to engage others in the social environment. We have found that music selection is a very special medium for transmission and exhibits traits that act as channel to connect people based on shared "meanings". From our initial study it was discovered that a number of participants could identify with one another through the same interest in music and in the testing session the music choices acted as a catalyst for social interaction.

The informal session was also an effective tool in observing and understanding how potential users might interact with one another. Participants in the evaluation sessions would converse with each other and discussed their musical preferences. Given the intended social use of tune_eile it was advantageous for the designers to observe this type of interactions taking place. Not only did the designers evaluate the prototype form a functional perspective but were also able to observe the prototype within a social context.

By evaluating in an informal environment, the designers quickly developed an understanding of the intended user of tune_eile. By placing the users at centre stage the designers learnt by observation how tune_eile could function in a social environment. The role-playing and acting exercises were of particular importance and value in involving multiple users in the evaluation sessions. The workshop provided a forum where users discussed their reactions to tune_eile. As a result of these workshops designers developed an insight into our intended user and can aid in the prediction of how they may employ tune_eile in the public space. These sessions were particularly effective in exposing the users to the system and thus the users were able to fully understand the system in a context of use.

4. DISCUSSION

The rising trends in ubiquitous technologies [7] are not only producing significant changes in the way we develop and view technology, but also in how we interact with it. Although these technologies do not change social behaviors they can however change the context in which these social interactions take place [1]. Multi touch technology allows multiple, simultaneous users to interact in an intuitive fashion. Embedding this technology in everyday objects such as tables serves to promote social ice breaking without disrupting existing behavioral patterns [8]. These familiar everyday objects reinforce existing metaphors in the interaction of the individual with an interface. Social aspects surrounding the use of tables make them appealing for use as displays [6]. The multi touch surface of tune_eile, which allows multiple users to add music simultaneously caters for a distribution of control and access amongst users, and thus aims to encourage social interactions in the environment.

5. CONCLUSION

Music is a medium that communicates with people on an individual level and has personal meaning to everyone. By using music as a channel for communication and coupling it with ubiquitous technologies within the tune_eile system, we offer a new novelty jukebox that that explores music distribution and the consequent social activities in a public space that are a by-product of the system.

Through evaluation and testing we discovered the entertainment possibilities of this type of technology. The multi touch surface of tune_eile allows users to simultaneously add music to the jukebox while gathered around the table. The system addresses the ways in which music is increasingly being used in public spaces to create place and social identities. The use of multi touch technology coupled with smaller displays is appealing, since it is a simple concept that replicates out daily interaction with real objects.

We are currently developing the prototype to integrate with other input devices such as Wi-Fi and USB. We believe that multi touch offers an intuitive method to interact with our surrounding environment.

6. ACKNOWLEDGEMENTS

The author would like to thank Deirdre Casey, the co-designer of tune_eile. Many thanks to all at the Interaction Design Centre at the University of Limerick for their support in the development of this project. Tune_eile is funded under cycle 4 of the HEA Programme for Research in Third Level Institutions (PRTLI).

7. REFERENCES

[1] Abowd, G. D. & Mynatt, E. D. (2000) Charting past, present, and future research in ubiquitous computing. ACM Trans. Comput.-Hum. Interact., 7, 29-58.

[2] Brown, B., Sellen, A. J. & Geelhoed, E. (2001) Music sharing as a computer supported collaborative application. Proceedings of the seventh conference on European Conference on Computer Supported Cooperative Work. Bonn, Germany, Kluwer Academic Publishers.

[3] Davidson, P. L. & Han, J. Y. (2006) Synthesis and control on large scale multi-touch sensing displays. Proceedings of the 2006 conference on New interfaces for musical expression. Paris, France, IRCAM, Centre Pompidou.

[4] DeNora, T (1986) How is Extra-Musical Meaning Possible? Music as a Place and Space for" Work". On-line: http://www.jstor.org/view/07352751/ap010004/01a00070/0

[5] DeNora, T (2000), Music in Everyday life. On-line: http://assets.cambridge.org/97805216/22066/copyright - /9780521622066_copyright.pdf

[6] Stavness, I., Gluck, J., Vilhan, L. & Fels, S. (2005) The MUSICtable: A Map-based Ubiquitous System for Social Interaction with a Digital Music Collection. International Conference on Entertainment Computing (ICEC05).

[7] Weiser, M. (1991) The Computer for the 21st Century. Scientific American.

[8] Yoon, J., Oishi, J., Nawyn, J., Kobayashi, K. & Gupta, N. (2004) FishPong: encouraging human-to-human interaction in informal social environments. Proceedings of the 2004 ACM conference on Computer supported cooperative work. Chicago, Illinois, USA, ACM Press.

A Taxonomy for Exploring the Design Space of User-Display-Display Interactions

Lucia Terrenghi
Vodafone GROUP R&D
Chiemgaustrasse 116
81549 Munich

lucia.terrenghi@vodafone.com

ABSTRACT

Based on a review of the related work, this paper proposes a taxonomy for a systematic approach to the classification of hybrid interactions at two different levels: First, it considers the relationship between user and display at the pragmatic level. Second, it proposes an extension of such a taxonomy to user-display-display interaction, considering a personal device as transducer and suggesting the adoption of such a framework for a systematic exploration of the design space of multi-display interactions.

Categories and Subject Descriptors

H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General Terms

Design, Economics, Human Factors.

Keywords

Surface computing, direct input, taxonomy

1. INTRODUCTION

Advances in display and input technologies bring digital information and interaction possibilities to the surfaces of the very artifacts of our physical space, such as tables and walls. For input and navigation into the digital space, we obviously need physical handles in the analogue one, be they tangible (e.g., a mouse or a personal mobile device), or not (e.g., speech). In this sense, one can consider every kind of interaction with digital media as “hybrid” in nature, since it involves a physical as well as a digital component. Furthermore, different layers of interaction are possible: for example, by interacting on a personal device such as a tablet PC we can affect the information landscape displayed on a large shared display and vice versa (e.g., [17]).

Within such a broad class of physical-digital interactions, this paper focuses on interactions characterized by direct input. This can be affected using either a physical transducer, such as a stylus or some other physical device, or with fingers, by direct touch. In fact, as display technologies have acquired novel sensing capabilities (such as touch and proximity), a diversity of interaction possibilities emerge which affect the spatial relationship between the user and the displayed information at the pragmatic level: This deals with gestures, spatial, as well as device issues, and is the first level of contact between the user and the system. In other words, it is the handle for hybrid interaction with the digital space. Such a level has an impact on the whole experience of interaction and it is in the scope of this paper to drive the attention on this aspect.

From a design perspective, in this paper I first propose a taxonomy of physical-digital interactions based on a survey of existing work [19] and considering the relationship between the user and the display at the pragmatic level. Drawing upon such a taxonomy, I propose an extension thereof to the level of user-display-display interaction, in which a personal device can be considered as transducer itself, enabling the interaction between a user and a larger display.

2. DIRECTNESS AND SURFACE COMPUTING

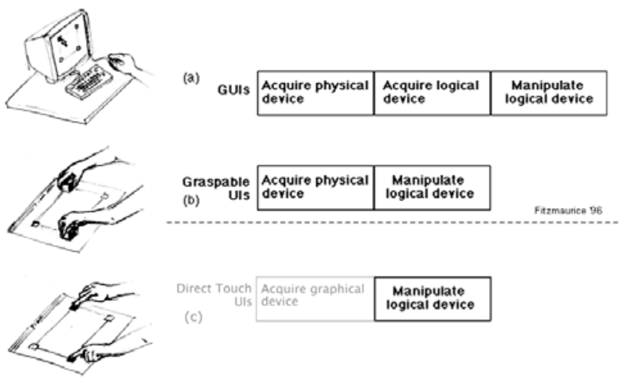


Figure 1: The upper part of the figure illustrates the reduction of the number of interaction phases as represented and articulated by Fitzmaurice [5]. The lower part (c) shows an extension of the same concept, enabled by direct touch surface computing.

			direct	indirect
space-multiplex	physical shape	semantic continuity	- MetaDesk (Ullmer and Ishii, 1997) [21] - HabilisDraw (Butler and St. Amant, 2004)[2] - Desigers' Outpost (Klemmer et al., 2001)[8] - 2D and 3D PhotoLens (Terrenghi et al., '08)[20]	- Props (Hinckley et al., 1994)[7]
		diverse	- Bricks (Fitzmaurice et al., 1995)[6] - Flatland (Mynatt et al., 1999)[10] - DataTiles (Rekimoto et al., 2001)[16]	- Toolstone (Rekimoto et al., 2000)[15] - Navigation Blocks (Camarata et al., 2002)[3]
		malleable	- Illuminating clay (Piper et al., 2002)[11]	- ShapeTape (Balakrishnan et al., 1999)[1]
time-multiplex	physical shape	diverse	- light pen on tablet PC	- mouse - Wacom graphic tablet - touch pad

Figure 2: A taxonomy of hybrid interaction paradigms for user-display interaction referring to related work.

Fitzmaurice's work on work on Graspable User Interfaces [5] had already exhibited the potential of reducing the stages of interaction by adopting physical objects as transducers, thus affording a more direct manipulation of the logical device (see Fig. 1, a vs. b). As described by Fitzmaurice [5] "A graspable function consists of a specialized physical input device which is bound to a virtual function and can serve as a functional manipulator."

Thanks to the persistent association between a physical object and its function, graspable UIs reduce the number of phases of interaction: Indeed, while the mouse needs to be alternatively associated with different functions in different moments in time (i.e., it is a generic, time-multiplex input device), graspable UIs are specialized tools embodying a certain function, which has its physical representation in the space (i.e., they are space-multiplex input devices, thus implying that multiple input points are possible simultaneously). Direct touch interfaces for surface computing (such as Diamond Touch [4] or Microsoft Surface [9] for example) can allow for a similar mapping between the acquisition of the interface (i.e., the handle for manipulation in the physical world) and the logical device (see Fig. 1, c). When multi-touch is enabled, space-multiplex input is possible. Furthermore, if the shapes of the graphical UIs of direct touch interfaces suggest their functions, domain-specific tools can be designed that afford space-multiplex input as graspable UIs do.

Considering these dimensions, a taxonomy of physical-digital interactions can be drawn as described in the next section and illustrated in Fig. 2.

3. A TAXONOMY OF USER-DISPLAY INTERACTIONS

The previous paragraph has anticipated one of the dimensions, directness, that one can consider in order to characterize different types of interactions based on the type of physical (spatial) relationship between the user and the interface at the pragmatic level.

Another dimension is the persistence of the association of a transducer with a virtual function (i.e., space-multiplex vs. time-multiplex input). Space-multiplex interfaces can provide handles which are specific for the task at hand. In these cases, the transducer - be it graspable (e.g., Ullmer and Ishii's [21] models and lenses in the MetaDesk interface) or graphical (e.g., Butler and St. Amant's HabilisDraw [2]) - can have a shape and/or perform its function consistently with its use and manipulation vocabulary in the physical space. One can then talk of semantic continuity of the transducer. On the other hand, a physical cube like in the case of the Bricks project [6], for example, can alternatively be associated with different functions depending on the context. In this case, its interaction vocabulary is diverse, and the binding between physical shape and virtual function is looser. Additionally, a transducer can be malleable, thus implying that the user can change its shape, as for example in the cases of Piper et al.'s Illuminating Clay [11] and Balakrishnan et al.'s ShapeTape [1]. Based on these dimensions, one can then define a taxonomy of interaction paradigms as illustrated in Fig. 2, which describes the design space in consideration of the related work.

Previous research [19] has indicated that depending whether the interfaces for hybrid interactions are 2D graphical ones for direct touch (e.g., the 2D PhotoLens [20]), or 3D graspable ones (3D PhotoLens [20]), different affordances can be perceived and experiences can emerge, according to the context (see Fig. 3 for an example and Fig. 4 for a summarizing overview). This implies that this additional dimension should be considered when designing the transducer for interaction. Furthermore, the use of a transducer such as a stylus can be considered as an example of semantic continuity in some cases (e.g., for scribbling and writing) and diverse in others. In this latter case, as in most of the tablet PCs, a stylus could alternatively be used for handwriting and taking notes, as well as for moving and selecting items, thus borrowing alternatively from a physical pen-like interaction vocabulary and from a light pen for direct manipulation of graphical digital media. This reinforces the idea that for an understanding (and design) of the semantics of the transducer, one needs to carefully consider the context of use and possible ambivalent meanings of the transducer across physical and digital worlds.

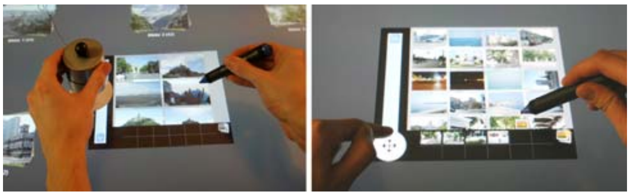


Figure 3: A 3D vs. a 2D handle used in the PhotoLens UI [20].

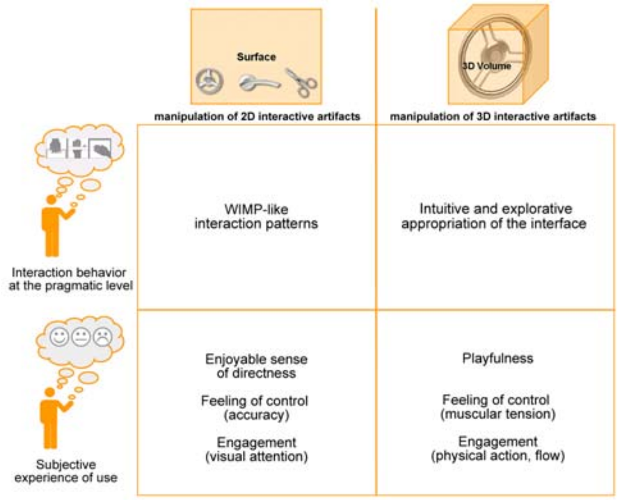


Figure 4: Different implications on the interaction behaviour and on the subjective user experience of the two approaches (2D vs. 3D handles) to the design of space-multiplex interfaces for direct input.

4. TOWARDS A TAXONOMY OF USER-DISPLAY-DISPLAY INTERACTIONS

The proliferation of displays in different sizes and with different features makes likely a scenario in which personal mobile devices are going to play the role of transducers for interaction with larger, and possibly shared displays. This opens up a new design space, in which the manipulation vocabulary of the personal device in relationship to a larger one needs to be defined. The taxonomy previously described can possibly be extended to the level of user-display-display interaction. Existing work has considered for example a 3D gesture vocabulary for interaction with a mobile device [13]. Novel display technologies emerge which make malleable surfaces [18] and flexible displays [12] feasible. Mapping these particular kinds of transducers to novel semantics of interaction which combine smaller and larger displays can inform the exploration of the design space for user-display-display interaction. In this respect, additional aspects like the use of a stylus for interaction across multiple displays (similarly to the case of Pick and Drop [14], for example) imply further layers of complexity to be considered. Furthermore, the implications of 2D vs. 3D transducers in the user-display interaction previously described (see Fig. 4) promise to have an impact in this context as well. Finally, while in the case of TUIs the transducer is usually considered as a purely input device, in the case of mobile devices used as transducers users' physical handle can simultaneously convey input as well as provide output (i.e., on the mobile's display). How will the borders between private and public information will be managed then?

This paper aims at contributing to the exploration of such a design space by proposing a systematization of user-display-display interaction paradigms: the described taxonomy can possibly lay the ground for discussion during the workshop, so as to explore the different ways in which mobile devices can be used as transducers in relationship to other displays.

5. REFERENCES

[1] Balakrishnan, R., Fitzmaurice, G., Kurtenbach, G., and Singh, K. (1999a). Exploring interactive curve and surface manipulation using a bend and twist sensitive input strip. In I3D ’99: Proceedings of the 1999 symposium on Interactive 3D graphics, pages 111–118, New York, NY, USA. ACM Press.

[2] Butler, C. G. and St. Amant, R. (2004). Habilisdraw DT: a bimanual toolbased direct manipulation drawing environment. In CHI ’04: extended abstracts on Human factors in computing systems, pages 1301–1304, New York, NY, USA. ACM Press.

[3] Camarata, K., Do, E. Y.-L., Johnson, B. R., and Gross, M. D. (2002). Navigational blocks: navigating information space with tangible media. In IUI ’02: Proceedings of the 7th international conference on Intelligent user interfaces, pages 31–38, New York, NY, USA. ACM Press.

[4] Dietz, P. and Leigh, D. (2001). Diamondtouch: a multi-user touch technology. In UIST ’01: Proceedings of the 14th annual ACM symposium on User interface software and technology, pages 219–226, New York, NY, USA. ACM Press.

[5] Fitzmaurice, G. W. (1996). Graspable User Interfaces. PhD thesis.

[6] Fitzmaurice, G. W., Ishii, H., and Buxton, W. (1995). Bricks: laying the foundations for graspable user interfaces. In Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI’95), pages 442–449. ACM Press/Addison-Wesley Publishing Co.

[7] Hinckley, K., Pausch, R., Goble, J. C., and Kassell, N. F. (1994). Passive real-world interface props for neurosurgical visualization. In CHI ’94: Proceedings of the SIGCHI conference on Human factors in computing stems, pages 452–458, New York, NY, USA. ACM Press.

[8] Klemmer, S. R., Newman, M. W., Farrell, R., Bilezikjian, M., and Landay, J. A. (2001). The designers’ outpost: a tangible interface for collaborative web site. In UIST ’01: Proceedings of the 14th annual ACM symposium on User interface software and technology, pages 1–10, New York, NY, USA. ACM Press.

[9] Microsoft Surface: http://www.microsof.com/surface/

[10] Mynatt, E. D., Igarashi, T., Edwards, W. K., and LaMarca, A. (1999). Flatland: new dimensions in office whiteboards. In CHI ’99: Proceedings of the SIGCHI conference on Human factors in computing systems, pages 346–353, New York, NY, USA. ACM Press.

[11] Piper, B., Ratti, C., and Ishii, H. (2002). Illuminating clay: a 3-d tangible interface for landscape analysis. In CHI ’02: Proceedings of the SIGCHI conference on Human factors in computing systems, pages 355–362, New York, NY, USA. ACM Press.

[12] Polymervision: http://www.polymervision.com

[13] Rekimoto, J. (1996). Tilting operations for small screen interfaces. In UIST ’96: Proceedings of the 9th annual ACM symposium on User interface software and technology, pages 167–168, New York, NY, USA. ACM Press.

[14] Rekimoto, J. (1997). Pick-and-drop: a direct manipulation technique for multiple computer environments. In UIST ’97: Proceedings of the 10th annual ACM symposium on User interface software and technology, pages 31–39, New York, NY, USA. ACM Press.

[15] Rekimoto, J. and Sciammarella, E. (2000). Toolstone: effective use of the physical manipulation vocabularies of input devices. In UIST ’00: Proceedings of the 13th annual ACM symposium on User interface software.

[16] Rekimoto, J., Ullmer, B., and Oba, H. (2001). Datatiles: a modular platform for mixed physical and graphical interactions. In CHI ’01: Proceedings of the SIGCHI conference on Human factors in computing systems, pages 269–276, New York, NY, USA. ACM Press.

[17] Sanneblad, J. and Holmquist, L. E. (2006). Ubiquitous graphics: combining hand-held and wall-size displays to interact with large images. In AVI ’06: Proceedings of the working conference on Advanced visual interfaces, pages 373–377, New York, NY, USA. ACM Press.

[18] Smith, D., Nicholas Graham, T.C., Holman, D., and Jan Borchers (2007) Low-Cost Malleable Surfaces with Multi-Touch Pressure Sensitivity. In Proceedings of TABLETOP ’07.

[19] Terrenghi, L. (2007) Designing Hybrid Interactions through an Understanding of the Affordances of Physical and Digital Technologies. PhD Thesis.

[20] Terrenghi, L., Kirk, D., Richter, H., Krämer, S., Hilliges, O., and Butz, A. (2008) Physical Handles at the Interactive Surface: Exploring Tangibility and its Benefits. In the Proceedings of AVI’08.

[21] Ullmer, B. and Ishii, H. (1997). The metadesk: models and prototypes for tangible user interfaces. In UIST ’97: Proceedings of the 10th annual ACM symposium on User interface software and technology, pages 223–232, New York, NY, USA. ACM Press.