Gesture Registration, Relaxation, and Reuse for Multi-Point Direct-Touch Surfaces

Mike Wu[1,2], Chia Shen[1], Kathy Ryall[1], Clifton Forlines[1], Ravin Balakrishnan[2]

[1]*Mitsubishi Electric Research Laboratories*
Cambridge, MA
shen | ryall | forlines @merl.com
www.merl.com

[2]*Department of Computer Science*
University of Toronto
mchi | ravin @dgp.toronto.edu
www.dgp.toronto.edu

Abstract

Freehand gestural interaction with direct-touch computation surfaces has been the focus of significant research activity recently. While many interesting gestural interaction techniques have been proposed, their design has been mostly ad-hoc and has not been presented within a constructive design framework. In this paper, we develop and articulate a set of design principles for constructing – in a systematic and extensible manner – multi-hand gestures on touch surfaces that can sense multiple points and shapes, and can also accommodate conventional point-based input. To illustrate the generality of these design principles, a set of bimanual continuous gestures that embody these principles are developed and explored within a prototype tabletop publishing application. We carried out a user evaluation to assess the usability of these gestures and use the results and observations to suggest future design guidelines.

1. INTRODUCTION

Computerized displays such as whiteboards, plasma displays, and tablet computers are increasingly available in offices, airports, classrooms, and even our homes. Currently, many of these displays serve solely as output only devices, while others employ limited means of input such as a stylus. However, recent advances in sensing technologies such as SmartSkin [18], DiamondTouch [8], and DViT [22], can transform these displays into multi-point touch sensitive surfaces that combine input *and* output in a co-located manner. This enables users to directly harness computational power through simple direct freehand gestural interaction involving fluid touches on a wall or tabletop surface, much like they might interact with physical artifacts in the real world.

Previous research on gestural interaction has concentrated on camera-based gesture recognition systems [11, 12, 16, 17, 19, 23, 24], virtual reality environments [25], special input gloves [3, 25] and mouse and pen-based gestural input [7, 14]. This body of research provides us with design insights, as well as empirical and experimental guidelines in their respective settings. However, interacting with table and wall surfaces through touch presents interesting new challenges. First, although general purpose stylus-based and single-finger touch interaction is well understood [14, 21], it is not clear how to seamlessly incorporate multi-finger and multi-hand

gestures into a computing environment that has been traditionally pointer-based where either the user's hand typically operates away from the surface itself, or a stylus is used for input interaction. Second, true direct-touch interfaces accentuate the occlusion problem; when both the display and input spaces are spatially coincident, the interacting hand may partially or entirely block digital objects from view. Third, the physical affordances of the display and interaction surface, such as height or angle of incline, can affect the contact shape and dynamics of a gesture. Finally, accessing areas of the surface that are physically distant can be uncomfortable or impractical.

Recent research on multi-input direct-touch interaction [18, 26] has developed interesting gestural interactions that address some of these challenges, but not in a systematic manner that is easily extensible to larger and more complex application domains. Furthermore, the resulting gestures typically bear little conceptual relationship to one another, making it difficult for users to understand the range of possibilities. Our goal is to develop design principles that can enable designers to construct new freehand, multi-point and multi-shape gestural interaction techniques whose invocation and action are easily understood and performed by users. Specifically, we propose 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 data. To illustrate the generality of these principles, we develop and evaluate example gestures within a tabletop publishing prototype (Figure 1) that acts as a vehicle for exploring our ideas.
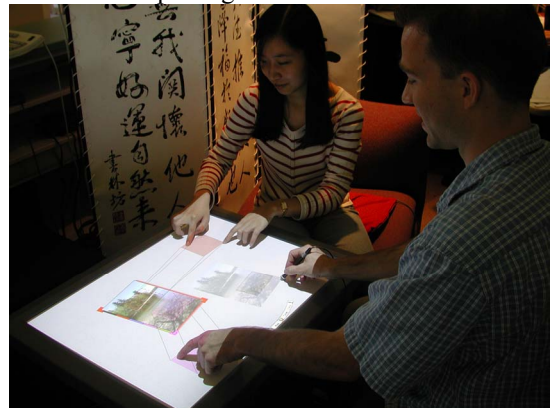

Figure 1: Two people using freehand gestures to interact with the same image document on a table.

2. RELATED WORK

Vision-based freehand gesture interactions have mostly focused on recognition [17], including tracking of human hands and 3D positions with multiple cameras [12, 24], and tracking of pointing directions and sweeping arm gestures with an infrared filter camera with multiple infrared lights [23]. The EnhancedDesk [11, 16] can track measured fingertip trajectories. Barehands [19] uses infrared cameras to enable tracking of hand postures on a back-projected SMARTBoard. Wexelblat [25] explored using position sensors, data gloves and eye gaze for gesticulative inputs in a virtual environment. Krueger's VIDEODESK [12] used image processing to track 2D hand and finger position and orientation, with a set of simple, self-revealing gestures. For example, a two-handed, four-finger technique is used as a continuous, but not compound, operation to stretch and squeeze an ellipse.

With the recent advances in input sensing technology, researchers have begun to design freehand gestures on direct-touch surfaces. Yee [27] augmented a tablet computer with a touchscreen to enable hand and stylus interaction. Rekimoto [18] described interactions using shape-based manipulation and finger tracking using the SmartSkin prototypes. Wu and Balakrishnan [26] presented a set of multi-finger and whole hand gestures on the DiamondTouch. Their gestures were categorized entirely by the shape of hand contact with the surface. While this work on interaction with multi-point direct-touch surfaces provides valuable insights, the designs have not been generalized to larger gesture sets where it would be infeasible to have a completely new gesture for each application command. We extend this prior art by introducing the ideas of gesture reuse and relaxation.

Researchers have also explored bimanual and compound interactions in the context of Guiard's Kinematic Chain (KC) model [9]. Leganchuk et al. [13] examined bimanual input from the perspective of both time-motion efficiency and possible cognitive benefits. The robustness of the KC model has also been studied, leading to empirical evidence that set guidelines for the design of bimanual interactions [1, 10]. Participants in these studies used two external input devices, typically two mice. Most important is that much of this research investigated setups with a displacement between input devices and output display. While our work focuses on direct-touch surfaces that integrate input and output, we nonetheless leverage the relevant insights from this important prior art. For example, several of our example gestures embody the KC model's principle of kinesthetic reference frames, where the dominant hand works within the context set by the non-dominant hand.

Buxton [4] suggested the exploration of gesture-based phrasing to chunk the human-computer dialogue into units meaningful to the application. He suggested that this could be the key to accelerate the acquisition of expert skills by novices, since experts and novices differ in the coarseness of granularity with which they view the constituent subtasks of the problem at hand. This work has inspired our strong emphasis on phrasing continuity in our proposed design principles and example gestures. In particular, we deliberately support the composition of simple gestures into a more complex compound gesture over time, thus increasing the size of the gestural phrasing as users gain expertise with the gestures.

Baudel and Beaudouin-Lafon [3] built Charade, a computer slide presentation system using data glove input. Charade relies on the classification of each gesture as a distinct posture to carry out a discrete invocation of a command. For example, commands for slide presentations can be "Next Chapter", "Previous Page", etc. In contrast, our gestures afford continuous operations, such as the continuous adjustment of the size of a selection box, while allowing the hand posture to be relaxed and varied after the initial gesture is recognized and registered by the system. We also advocate the reuse of gesture primitives to construct compound gestures, thus reducing the number of basic gestures that the user needs to learn. Charade also proposes a three stage model for gestural interaction, which we discuss in detail and extend significantly in the "gesture registration" section below.

3. DESIGN PRINCIPLES

In a general computational environment, the user must manage a plethora of tools and interaction methods. Thus, when designing a system with direct-touch surfaces consideration must be given to the style of interaction and the available tools since these factors can significantly influence design. In terms of interaction style, how should the gestures map to the various system functionalities? Should the gestures map to the most common tasks or the most complex? Should there be support for general point-based interaction in addition to the gestural interface? If so, how can the system support transitions between both interaction styles? As for available tools (e.g., fingers, hands, stylus), should each one control a different functionality or can we mix and match as desired depending on application context? While these decisions must be made by designers based on the particulars of their applications, they will nonetheless benefit from guidelines that can systematically introduce new gestures without overly complicating the overall interaction. In this paper, we begin to address the above research questions with a set of design principles: (1) gesture registration, (2) gesture relaxation, and (3) gesture and tool reuse. In the process of developing these principles for gesture design on direct-touch surfaces, we have drawn insights not only from the literature reviewed previously but also from our observations of current work practice.

Gesture Registration

Gesture registration is the beginning phase of every gesture operation, be it compound or simple, continuous or discrete; it sets the context for subsequent interactions. In their Charade system, Baudel and Beaudouin-Lafon [3] propose a model for gesture design where each command is described by three stages – a start position, a dynamic phase and an end position. In their model, start positions are not unique, thus gestures must be classified according to a combination of their start position and their dynamic phase. We extend this with an explicit gesture registration phase. The registration phase is entered by a distinctive posture that, once recognized, sets the context for the dynamic and end phases. The registration phase clearly delineates one context from another, enabling gesture reuse in various phases of the entire compound gesture.

Gesture registration is an important phase in an interaction environment where multiple interaction styles and tools are present. Given a computational multi-point direct-touch surface where cursor-based and pointer-based interactions coexist with the possibilities of free hand gestures, gesture registration can be used to demarcate the functional transition of a tool from one interaction style to another. For example, a stylus can be transitioned between being a pointer for selection and dragging, and a writing tool with a simple gesture registration phase.

Gesture Relaxation

Most prior research on freehand gestures requires the hand posture to remain the same throughout the dynamic phase of the gesture. This imposes an undue burden on the user in having to maintain fairly precise hand postures with muscular tension. We propose the principle of gesture relaxation to allow a gesture to be performed with minimal constraints after it is registered. Relaxing the shape and dynamics of gestures after registration allows someone to more comfortably perform a gesture, as tension would only be required in the gesture registration phase.

While the concept of gesture relaxation is clearly applicable to any gestural interaction domain, it is particularly important for tabletop interaction because of the high variability in how users stand or sit at a table. For example, performing the same gesture with different body postures or on tables of different height (e.g., a coffee table versus a drafting table) can result in a significantly different signal from the tabletop sensing technology. While a user could adjust and perform a particular gesture such that it is easily recognized by the system during the short gesture registration phase, it would be difficult if not impossible to maintain that gesture during the much longer gesture interaction execution phase. Hence, the particular value of gesture relaxation in this domain.

Gesture and Tool Reuse

Gesture and tool reuse refers to employing the same gesture, including hand postures, finger touches or stylus, to accomplish different tasks. When using gesticulative inputs in virtual environments, Wexelblat [25] allowed the same gesture to mean two different commands depending on the application context. However, the context relied on an interpreter component that had to understand application scenarios case by case. We extend the notion of reuse to gesture primitives (i.e., the basic components that define a gesture, such as hand postures or gesture dynamics, given that a gesture can consist of continuous motions and be compounded with more than one gesture primitive). A large set of primitives both burdens the users in memorizing gestures, and the system in having to recognize many different patterns. Reuse of primitives enables larger sets of gestures to be constructed without requiring additional primitive gestures to be defined.

Combining the Principles

Taking the principles of registration, relaxation, and reuse as a whole, we can systematically create compound gestures from a sequence of multiple simple gestures each of which could be unique or reused. The registration and relaxation phases as a combined chunk act as the delineator between concatenated operations, since the system can easily distinguish between the relaxed posture and a new gesture registration. Further, the series of registration/relaxation sequences allows for a compound gestural interaction to be performed without requiring the user to lift their hands off the interaction surface. As such, gesture registration and relaxation can be thought of as being the essential enabling components of gesture composition. By taking the previous sub gestures within a compound gesture into account, gesture primitives can be reused for slightly different actions, thus requiring the user to learn only a small set of primitives that can be combined as needed for sophisticated actions.

It is important to note that gesture registration can take two forms. In the simplest, registration is achieved by recognition of the posture of the hand alone in a single static snapshot. A more sophisticated approach to registration considers both the hand posture and dynamic actions that occur immediately after the posture is recognized, within a predefined time window. This allows for a single posture to be used for multiple different registrations just by varying the dynamics of movement during the registration phase, thus enhancing gesture definitions beyond mere postural characteristics

4. PROTOTYPE APPLICATION with SAMPLE GESTURES

We have developed a tabletop publishing prototype as a vehicle to validate our design principles and explore new interface ideas. Although this application domain presents interesting challenges in and of itself, we focus on developing four key interaction techniques aimed at illustrating the use of our design principles. Through these four techniques, we show how the principles can be used to develop techniques with a range of complexities, ranging from simple application of a subset of the principles to more sophisticated application of all the principles working in concert.

When designing a magazine page or web page layout collaboratively, people often sit around a table and work with physical paper and photographs. In our experience with web page layout designs, we have observed the actions of writing, annotating, selecting, copying, arranging, and piling physical documents. Writing is usually carried out with the non-dominant hand holding down a piece of paper while the other hand annotates interesting details with a pen, in accordance with Guiard's KC model. Art and image clippings are frequently folded or cut up, and then spatially arranged on a table to reflect the design layout. These paper materials are also often grouped according to some theme, and piled when more table space is required. Such piles are spread out from time to time in order to browse their contents. A common theme in such scenarios is the use of both hands to manipulate these documents in various ways, and the use of a variety of tools such as scissors or pens. From these observations, we felt that a tabletop publishing scenario would provide interesting opportunities for using hand gestures to interact with digital documents. We created a set of gestures to organize, cut up, and mark documents on a computationally augmented table (Figure 2). This set was carefully designed to support the application context as well as exercise our design principles to the fullest.

Our gestures are prototyped within the DiamondSpin Java Toolkit [21] using the DiamondTouch table [8]. Both technologies support an around-the-table setup, in which people may be seated on any side of the table. Our gestures are designed to be orientation-invariant in that they can be used from any direction of the surface. They also support ambidextrous interaction with no assumption about handedness in any of the interaction techniques.

Our gestures are part of a larger shared-display groupware system that supports multiple people and includes standard cursor document manipulation. While a discussion of these features and interactions are outside the scope of this paper (see [21] for details), it is important to note that our system utilizes a stylus for cursor point-based interaction such as dragging or resizing a document.

Although direct finger input could be used instead, we chose to use a stylus based on observations that during publishing activities, people needed to write, annotate and sketch – tasks suited to stylus input. As a person switched between gesture performance and stylus use, it gradually became clear that it was inefficient to repeatedly grasp and release the stylus throughout collaborative activities. We thus explored gestural interactions in which the stylus can be held in the hand comfortably most of the time.

Annotate Gesture

This illustrates a basic application of our registration and relaxation principles. The goal of the `Annotate` gesture is to allow freeform marks to be written with digital ink on the table surface. When any two fingers are placed onto the table (Figure 2a), the gesture is registered and the stylus behaves as a writing tool. This interaction is similar to how people hold down a piece of paper with their non-dominant hand while writing [2, 9]. The action is continued as long as either the hand or stylus is touching the table, regardless of the shape of hand contact; thus illustrating gesture relaxation (Figure 2b) where the non-dominant hand relaxes after the initial registration).

Wipe Gesture

Another example of a basic application of our registration and relaxation principles is the `Wipe` gesture, which allows us to erase annotations. We modeled this gesture on the physical actions used when erasing chalk from a blackboard (Figure 2, `Wipe`). Unlike the traditional discrete delete function, `Wipe` erases over time. This was a decision based on our observation that designers often need to fade and selectively erase portions of ink marks, instead of simply performing a crude delete operation. This prompts us to explore continuous gestures with subtle variations on the effect of the operations, rather than designing for time-motion efficiency.

The `Wipe` gesture is initially registered by placing a contiguous portion of the hand, such as a palm or closed fist that is larger than one fingertip onto the table. Once the gesture is registered, the user can change how the hand contacts the surface (i.e., the gesture is relaxed) (Figure 2d). As the gesture is continued, annotations under the hand are slowly removed by becoming increasingly transparent. This change is based on a function of two touch parameters: the amount of surface contact and the speed of hand motion. The less surface contact there is, the slower the change in transparency, and the less speed involved with the wipe, the longer it takes for the stroke to disappear. For visual feedback, three concentric unfilled circles are displayed, centered on the touch location. Strokes within these circles grow fainter.
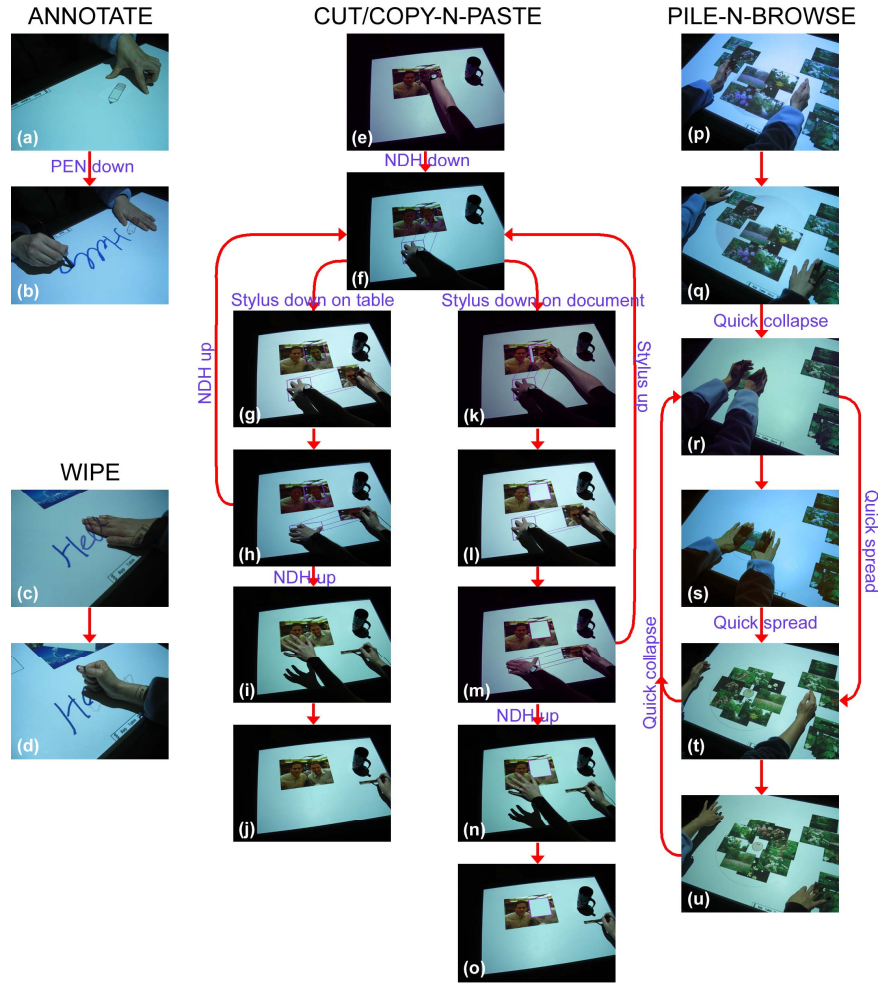
Figure 2: The four gestures. **Annotate** (a-b) relaxation of hand while annotating with the stylus. **Wipe** (c-d) relaxation allows various hand postures to be used to fade the marks left by the stylus. `Cut/Copy-n-Paste` (e) grabbing a document to cut/copy, (f) indirect adjustment of selection box location/size using non-dominant hand, (g) stylus down on table to indicate intention to copy, or (k) stylus down onto document to indicate intention to cut, (l) stylus dragging cut item away from source location, (h/m) indirect scaling using non-dominant hand, (i/n) lifting non-dominant hand indicates intention to paste, (j/o) stylus moves the copied/cut portion to the appropriate place before committing the paste operation by lifting. `Pile-n-Browse` (p-q) choose documents (q-r) both hands are quickly pulled together to create a pile (s-t) both hands are quickly spread to browse the pile, (q/s/u) illustrate gesture relaxation.

Cut/Copy-n-Paste Gesture

This illustrates multiple invocations of the registration principle in concert with relaxation and sequencing of multiple primitive gestures into a complex whole.

This gesture is for cutting or copying a region of an image (Figure 2, `Cut/Copy-n-Paste`). A table surface can be quite large, so this gesture affords the placement of the copied object at a location that may be far from the original document. In addition, this allows many people to collaboratively work together on the same document simultaneously without having to move the original document back and forth or requiring a proxy object between them. In Figure 1, two people are simultaneously using our `Copy-n-Paste` gesture on the same image.

The `Cut/Copy-n-Paste` gesture is a variation of the conventional desktop cut/copy and paste procedure that involves multiple disjoint steps, carried out in some serial order: choosing the item on which to perform the action, making a selection, refining the size and shape of the selection as necessary, copying the selection, and then pasting it. Inspired by Buxton's notion of phrasing [4], we combine the steps into a set of fluid motions, many of which can be carried out in parallel. Unlike the disjoint phrases of standard desktop copy/paste, our technique is executed with one continuous complex phrase, held by relaxed kinesthetic tension. Such kinesthetically held modes have been shown to significantly reduce mode errors as compared to standard persistent modes [20]

To copy a portion of a document, one grabs the desired portion using three or more fingers (Figure 2e). The system recognizes this contact as the gesture registration phase. A rectangular box illustrates what region of that document is selected, and its size can be changed with the expanding and shrinking of the finger spread. This is the gesture relaxation phase during which hand poses can vary from using one finger to five fingers, and from using a single hand to using two hands. The gesture terminates when the user stops touching the table.

Sliding the hand away from the document, while still touching the table, transitions to *indirect* adjustment of the selection box's location and size. Four visual lines provide feedback to indicate how the control and display regions are related (Figure 2f). This is a visually-tethered indirect distant operation, which is a possible solution to the occlusion problem on direct-touch surfaces identified earlier. A user can thus control a document from a location from which he/she feels comfortable. This is also useful when multiple people simultaneously copy different portions of the same document from different sides of a table in that physical interference is also mitigated.

To indicate the intention to copy (rather than cut), the user touches the stylus tip with their other hand onto an open area of the table. To indicate the intention to cut, the user touches the tip of the stylus onto the document itself (see Figure 2k). The selected portion of the document then follows the movements of the stylus point, and allows the user to comfortably position and view it from a convenient location, before committing to pasting it.

Before issuing the paste command, the user must lift the hand controlling the selection box. The stylus can still drag the document to a desired location. Once the stylus is lifted, the paste command is issued.

This example illustrates the use of multiple gesture registrations – first by the use of three or more fingers to set up the copy area, and second by the use of the stylus tip to specify the paste location; gesture relaxation where the copy gesture is relaxed to allow for manipulation of the bounding box; and composition of two separate primitive gestures into a sequenced complex gesture to achieve the compound task of selecting and copying an item from one location and pasting at another.

Pile-n-Browse Gesture

In this last example, we show all three principles used in concert to create a sophisticated interaction composed of several distinct but related gestural phrases. This example enables piling and browsing of items to aid organization, and uses continuous motions to transition through three subtasks: choose, pile, and browse (Figure 2, `Pile-n-Browse`).

When two hands are placed onto the table, the `Pile-n-Browse` gesture is registered. A filled circle visually appears between the hands (Figure 2p) indicating which documents will be part of the pile. This circle can be adjusted by moving the hands. The shape of the hand contact can change during the gesture continuation (i.e., gesture relaxation is afforded). At this point, the user can lift his/her hands to cancel the operation.

When the selection scope is satisfactory, a pile can be created by quickly bringing in both hands together to "scoop" the items (see transition in Figure 1, from (p) to (q)). A speed threshold marks the registration of this gesture, but the scooping speed is relaxed once the gesture is registered. This is an illustration of gesture reuse, in that the same gesture is used first as a static posture to indicate the selection scope, and then is reused with dynamic characteristics to perform the scooping action. The sequencing of similar gestures determines the resulting actions, thus allowing for powerful complex commands to be formulated from a single primitive gesture.

Once the pile is created, it follows the hands as they move together, while the gesture itself can be relaxed. Lifting both hands leaves the pile on the table. Elements of a pile are stacked with incremental offsets so that one can visually see the approximate number of items in the pile [15]. A visual icon, labeled "Pile", is left overtop of the collected documents that can be used to move the pile.

To browse a pile, two hands are pulled away from each other quickly (i.e., see transition in Figure 2, from (s) to (t)). Again, a speed threshold marks the gesture registration, but there are no speed or shape constraints thereafter, once the gesture is relaxed. Documents within the pile spread out in a circular manner and animate by slowly moving clockwise. The distance between the hands controls how far apart the documents should be displaced from the centre. The browse gesture can be applied to an already existing pile by placing the hands on the pile before spreading them. Removing both hands from the table will cancel the browse action and leave the pile on the table, spread out as displayed.

5. USER EVALUATION

We conducted an observational study to evaluate the usability of our example gesture set. We note that at the current state of development of tabletop interfaces, there is yet to emerge any semblance of standard interface elements that would serve as a baseline comparison for our current designs. As such, the typical comparative experiments with time and error metrics that one might perform when evaluating new interface designs for more established interaction platforms are simply not feasible in the current context, Accordingly, our current study focuses on user reaction to the overall interface designs and their ability to understand, learn, and execute the various phases of the interaction techniques. We also believe that while the typically reported quantitative task completion time

and error measures reflect an important aspect of the usability of interface techniques, it is important to acknowledge that it is often the more subtle subjective elements that make or break a design's acceptability and usability. Indeed, perhaps the most appealing aspect of tabletop interaction is that it affords more expressive manipulation styles than is possible with devices like mice and pens which may be more time or error efficient, but do not have the aesthetic expressivity that tabletops afford.

Ten people (5 female, 5 male, ages 19-30 years) from outside our lab participated in an hour-long session each. None had experience with tabletop or gesture interaction.

Each session started with instructions on how to gesture to perform actions on the tabletop. These instructions took the form of watch and repeat, with the experimenter performing a gesture and the participant immediately imitating it. The order of presentation was as follows: annotate, wipe, moving images, copy-n-paste, pile-n-browse. Once the participants felt comfortable performing the gestures, they were given two tasks, each making use of a different set of images. To aid our observations, we had participants talk aloud during the experiment, a standard usability evaluation protocol.

The first task involved positioning six photos before making copies of just the faces of people within the photos. Participants were then asked to make a pile with the original images, and another pile with the new face images. Finally, after participants marked each of the piles with annotations, they were asked to erase those labels and write new ones. In the second task, each participant organized 18 photos into groups (either by location, person, activity, etc.), created piles for these groups, and annotated a name for each group.

At the end of the session, participants were given a questionnaire asking them to rank the difficulty of their actions and to rate their agreement with a collection of statements. They were also asked to list the "best three things" and "worst three things" about the interface.

Results and Observations

The numerical results from the questionnaire are summarized in Tables 1 and 2. Participants were able to quickly learn how to perform the gestures. During the tutorial, most were able to accurately perform each of the gestures after only one demonstration. Four participants listed ease of learning or performing the gestures as one of the "best three things" about the interface.

Table 1. Average difficulty ranking for each gesture. Lower ranks equate to lower difficulty.

Gesture	Mean (SD)
Moving images around	1.3 (0.7)
Annotate (writing marks on the table)	2.5 (0.7)
Wipe (erasing marks on the table)	2.9 (1.4)
Pile-n-Browse (piling images together)	3.7 (1.2)
Copy-n-Paste (copying and pasting images)	4.5 (0.9)

Table 2. Participants rated their agreement with statements on a 7-point Likert scale. An answer of 1 corresponded to "strongly disagree" and 7 to "strongly agree"

Statement	Mean (SD)
Adjusting the size of the selection box when copying was easy to do.	3.4 (2.1)
Adjusting the location of the selection box when copying was easy to do.	4.2 (1.9)
Erasing marks on the table was easy to do.	4.5 (1.9)
The machine understood what I wanted to do.	4.5 (1.0)
Piling images together was easy to do.	5.1 (1.4)
Browsing the images in a pile was easy to do.	5.1 (1.2)
Selecting an image to copy was easy to do.	5.2 (1.4)
Pasting the selection was easy to do.	5.2 (1.6)
Selecting images to pile together was easy to do.	5.2 (1.2)
The gestures to get the behavior I wanted were obvious.	5.3 (1.1)
Writing marks on the table was easy to do.	5.4 (1.4)
Overall I think that the table understood me well.	5.4 (0.7)
It was easy to remember how to do what I wanted to do.	5.6 (0.8)
Canceling the copying of an image was easy to do.	6.2 (1.0)

Participants were able to quickly complete the given tasks, although one had trouble erasing marks on the table and a second had a lot of trouble piling images together. Additionally, four listed trouble with the piling gesture as one of the "worst three things" about the interface.

Visual feedback and the ability to cancel an operation were important for the continuous actions. For each user, the system exhibited at least one misrecognition of the intended gesture during gesture registration. However, in almost all cases, the participant was able to correct themselves by canceling out before any side-effect occurred, resulting in low error rates for the given tasks. From these observations, we note that side-effect free cancellation is important for gestures that combine a series of commands. We noticed that pixel-accurate selection was difficult. Some participants had trouble or expressed concern over the accuracy with which they could select a region of an image during `Copy-n-Paste`. Oftentimes, as a participant lifted their hand to complete the paste operation, the pasted image was slightly shifted in one or both dimensions. Three participants listed this issue as one of the "worst three things" about the interface. Table 2 shows that participants felt they had more difficulty with adjusting the size (*#1*) and position (*#2*) of the selection rectangle than any other facet of the system. Such "jitters" [5] occured when the hand was lifted from the surface.

6. CONCLUSIONS and FUTURE WORK

Drawing from prior research on gestural interaction, we have developed and evaluated a novel set of design principles that support multi-hand gestural interaction on direct-touch surfaces. These design principles address a number of unique challenges that arise from working with direct-touch surfaces and in environments where both conventional point-based input and free hand touch gestures can co-exist. Gesture reuse reduces the number of gesture primitives that a user must learn, tool reuse allows

input devices to be multi-purpose, gesture relaxation enables transitioning from explicit postures to arbitrary relaxed freehand interaction, and gesture registration supports both static and dynamic gesture definitions, as well as tool reuse. With these principles, we developed and evaluated sample gestural interaction techniques within the context of a tabletop publishing application.

The work presented in this paper also raises a number of new areas of research. We have started to develop 'self-revealing' gestural interaction designs in which a user is visually shown available options at each step of a multi-stage gestural interaction. This 'self-revealing' concept enlarges the reusability of component elementary gestures.

A novel concept that emerged from our work is allowing a user to transition between control and display spaces on a direct-touch surface (as illustrated in the design of our `Copy-n-Paste` gesture). This technique can be particularly useful for large-display settings and/or multi-user settings, and is an interesting area of future work. Direct-touch surfaces also raise the issue of input granularity. Our user study indicated that some users had difficulty with pixel-accurate selection in our techniques. The study also highlighted the importance of visual feedback throughout a gesture interaction. We hope others will benefit from, and add to, the set of design principles in this paper. Direct-touch surfaces are becoming more prevalent, and continuous multi-hand gesture interactions provide a powerful interaction paradigm.

REFERENCES

1. Balakrishnan, R., & Hinckley, K. (1999). The role of kinesthetic reference frames in two-handed input performance. *ACM UIST*. p. 171-178.

2. Balakrishnan, R., & Hinckley, K. (2002). Symmetric bimanual interaction. *ACM CHI*. p. 33-40.

3. Baudel, T., & Beaudouin-Lafon, M. (1993). Charade: remote control of objects using free-hand gestures. *Communications of ACM*, 36(7). p. 28-35.

4. Buxton, W. (1986). Chunking and phrasing and the design of human-computer dialogues. *IFIP Conference*. p. 475-480.

5. Buxton, W. (1986). There's more to interaction than meets the eye: Some issues in manual input. In Norman, D.A., Draper, S.W. (editors), *User Centered Systems Design*, Lawrence Erlbaum Associates. p. 319-337.

6. Buxton, W., Hill, R., & Rowley, P. (1985). Issues and techniques in touch sensitive tablet input. *ACM SIGGRAPH*. p. 215-223.

7. Citrin, W.V., & Gross, M.D. (1996). Distributed architecture for pen-based input and diagram recognition. *AVI Conference on Advanced Visual Interfaces*. p. 132-140.

8. Dietz, P., & Leigh, D. (2001). DiamondTouch: A multi-user touch technology. *ACM UIST*. p. 219-226.

9. Guiard, Y. (1987). Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of Motor Behaviour*, 19 (4). p. 486-517.

10. Kabbash, P., Buxton, W., & Sellen, A. (1994). Two-handed input in a compound task. *ACM CHI*. p. 417-423.

11. Koike, H., Sato, Y., & Kobayashi, Y. (2001). Integrating paper and digital information on EnhancedDesk. *ACM TOCHI*, 8 (4). p. 307-322.

12. Krueger, M., Gionfriddo, T., & Hinrichsen, K. (1985). VIDEOPLACE - An artificial reality. *ACM CHI*. p. 35-40.

13. Leganchuk, A., Zhai, S., & Buxton, W. (1998). Manual and cognitive benefits of two-handed input: an experimental study. *ACM TOCHI*, 5 (4). p. 326-359.

14. Long, A.C., Landay, J.A., Rowe, L.A., & Michiels, J. (2000). Visual similarity of pen gestures. *ACM CHI*. p. 360-367.

15. Mander, R., Salomon, G., & Wong, Y. Y. (1992). A 'pile' metaphor for supporting casual organization of information. *ACM CHI*. p. 627-634.

16. Oka, K., Sato, Y., & Koike, H. (2002). Real-time tracking of multiple fingertips and gesture recognition for augmented desk interface systems. *IEEE International Conference on Automatic Face and Gesture Recognition*. p. 429-434.

17. Pavlovic, V., Sharma, R., & Huang, T. (1997.) Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Trans. on PAMI*, 19 (7). p. 677-695.

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

19. Ringel, M., Berg, H., Jin, Y., & Winograd, T. (2001)). Barehands: Implement-free interaction with a wall-mounted display. *ACM CHI*. p. 367-368.

20. Sellen, A., Kurtenbach, G., and Buxton, W. (1992). The prevention of mode errors through sensory feedback. *Human Computer Interaction*, 7(2). p. 141-164.

21. Shen, C., Vernier, F.D., Forlines, C., & Ringel, M. (2004). DiamondSpin: An extensible toolkit for around-the-table interaction. *ACM CHI*. p. 167-174.

22. Smart Technologies Inc. Digital Vision Touch Technology. http://www.smarttech.com/dvit/

23. Starner, T., Leibe, B., Minnen, D., Westyn, T., Hurst, A., & Weeks, J. (2003). The perceptive workbench: Computer-vision-based gesture tracking, object tracking, and 3D reconstruction of augmented desks. *Machine Vision and Applications*. 14. p. 59-71.

24. Utsumi, A., & Ohya, J. (1999). Multiple-hand-gesture tracking using multiple cameras. *IEEE Conference on Computer Vision and Pattern Recognition*. p. 473-478.

25. Wexelblat, A. (1995). An approach to natural gesture in virtual environments. *ACM TOCHI*, 2 (3). p 179-200.

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

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