Interaction Techniques for the Analysis of Complex Data on High-Resolution Displays

Chreston Miller, Ashley Robinson, Rongrong Wang, Pak Chung, and Francis Quek
Center for Human-Computer Interaction, Virginia Tech
2202 Kraft Drive, Blacksburg,Va, USA
{chmille3, arrobin, wangr06, pchung, quek}@cs.vt.edu

ABSTRACT

When combined with the organizational space provided by a simple table, physical notecards are a powerful organizational tool for information analysis. The physical presence of these cards affords many benefits but also is a source of disadvantages. For example, complex relationships among them are hard to represent. There have been a number of notecard software systems developed to address these problems. Unfortunately, the amount of visual details in such systems is lacking compared to real notecards on a large physical table; we look to alleviate this problem by providing a digital solution. One challenge with new display technology and systems is providing an efficient interface for its users. In this paper we look at comparing different interaction techniques of an emerging class of organizational systems that use high-resolution tabletop displays. The focus of these systems is to more easily and efficiently assist interaction with information. Using PDA, token, gesture, and voice interaction techniques, we conducted a within-subjects experiment comparing these techniques over a large high-resolution horizontal display. We found strengths and weaknesses for each technique. In addition, we noticed that some techniques build upon and complement others.

Categories and Subject Descriptors

H.5 [**INFORMATION INTERFACES AND PRESENTATION**]: H.5.2 User InterfacesInput Devices and Strategies

General Terms

Experimentation, Performance, Human Factors

Keywords

Human-Computer Interaction, Multimodal Interfaces, Tabletop Interaction, Horizontal Display, High-resolution Displays, PDA interaction, Tangible Interaction, Gesture Interaction, Voice Interaction, Embodied Interaction

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ICMI'08, October 20–22, 2008, Chania, Crete, Greece.

Copyright 2008 ACM 978-1-60558-198-9/08/10 ...$5.00.

1. INTRODUCTION

Organizing and maintaining a working set of information for analysis can be a daunting task. A venerated approach employed by scholars in diverse fields has been the use of physical index or notecards. Physical notecards can be a very useful means to categorize and organize information. The power of notecards for reasoning with information may be summarized in three terms related to interaction: portability, representative power, and configurability. Notecards are portable and easily manipulated because they fit nicely into the average hand, and are rigid enough to be held and read, stacked, shuffled, and sorted. When used correctly so that separable ideas are stored in separate cards, they represent individual ideas that may be assembled, conceptualized, and organized in a flexible manner. Finally, when laid out spatially, notecards support the use of spatial reasoning, memory, and association to organize, analyze, and synthesize higher level concepts.

Physical notecards have some shortcomings and disadvantages when compared against typical digital information formats. These shortcomings may be summarized, again, under three headings: One cannot 1. represent complex (multi-dimensional) relationships/associations using notecard configurations; 2. handle multiple alternative organizations (i.e. save physical layouts and compare different layouts); and 3. search notecards by content.

There have been a number of notecard software systems developed for organizing ideas and analyzing information, some being [10, 18, 1, 20, 14, 32]. These notecard systems were developed for use with personal displays or conceivably for projected displays. Systems such as these work well in aiding the organization of thoughts and ideas, however expanding to a larger viewing area with the same detail of a personal display may be difficult. Simply projecting the display of such systems onto larger screens only provide a low-resolution view over a large display surface, and does not afford the detail necessary for simultaneous spatial organization and inspection of notecard details. This simultaneity of organization and inspection (representation) is precisely one of the strengths of physical notecards. Our CardTable system [2] is an attempt to exploit the strengths of physical notecards and address some of their shortcomings by supporting the manipulation, viewing, and editing of graphical notecards on a high-resolution horizontal display.

In this paper, we investigate different interaction techniques for such a high-resolution horizontal display. We compare interaction using a personal data assistant (PDA), a tangible token, gesture, and voice. We conducted a within-

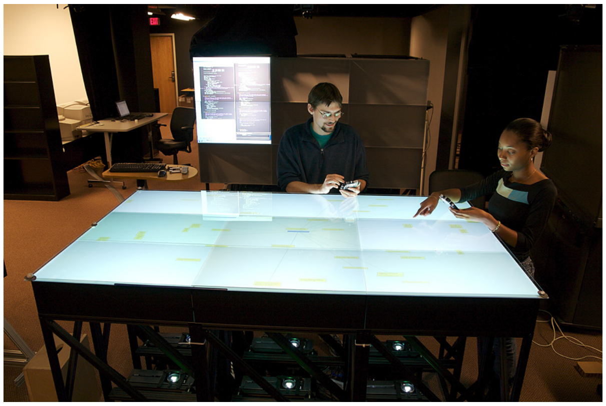

Figure 1: Two users interacting with CardTable.

subjects experiment comparing different combinations of techniques for performing various manipulation tasks using our CardTable system as a testbed. Even though we employed a testbed, our results are extensible to a broader set of applications for organizing and making sense of content on a high-resolution horizontal display (e.g. managing a filesystem). In such applications, there are objects with content that the user wants to browse, edit, create associations, and search.

Hence, our goal in this study is to 1. provide empirical guidelines for interaction choices in large high-resolution horizontal display systems and 2. provide usage pros and cons for each interaction technique.

2. DESIGN RATIONAL

To motivate and provide a framework for our discussion, we present the rationale for design alternatives that we shall test. We first discuss the general decision to employ the large high-resolution horizontal configuration and proceed to the choices we made for interacting with the entities in our CardTable system via the PDA, token, gesture, and voice.

2.1 Large High-Res. Horizontal Displays

While the focus of this paper is not on large high-resolution displays, we discuss these design choices to contextualize the interaction methodologies that constitute our current focus.

The use of physical notecards is naturally paired with organization on large horizontal surfaces (e.g. tabletop, or floor). The reason for this is obvious: horizontal surfaces afford support for physical objects and provide the spatial resource for organization. Since we utilize various physical objects like the PDA to organize our notecards, we chose to use a horizontal display for our system. The power of notecards to support concept formation is the simultaneous capacity for representation and configuration. To retain this power in our digital CardTable, we implemented the system on a high-resolution table with a surface resolution of over 8 million pixels. This enables the user to both organize the cards over a large surface, and to view the contents of the cards during the interaction simultaneously. Figure 1 shows the CardTable system we use in this study.

2.2 Interaction Methodologies

Recall that our premise concerning the power of notecards for concept formation derives from the properties of portability, representative power, and configurability. Since notecards have the same portable form factor as a typical PDA, it may seem obvious that a PDA may serve as a proxy for our digital notecards. It also supports convenient entry and editing of rich data, such as text. This choice of a PDA follows from its many uses as a portable viewing and editing device in previous ubiquitous and embodied interaction systems [16, 31, 17, 2]. The use of a PDA in these systems gives the user a powerful control tool by which to interact with the system and to view private information pertinent to the user's current task [17].

In our prior study using CardTable, we investigated the use of a tracked PDA as the sole interaction device for our notecards [2]. One may decompose the role of the device into two components: access to a card's content (to enter, view, and edit content) and manipulation of the card on the table (to select, move, copy a card, and to make associations between cards). In our original design, when the user moved the tracked PDA over the table, a cursor is projected onto the tabletop permitting manipulation of the objects on the table. We saw that some interactions solely using a PDA were cumbersome. This results from the conflating of both interaction components within one device. Users found it hard to view the contents of the PDA while manipulating its projected cursor. This is consistent with the observations of Klemmer et. al. in their work "Books with voices" [16], where a PDA is used to scan barcodes to replay recordings of oral histories. The PDA was required to be held perpendicular so the scanner on top of the PDA could scan vertically to read the bar-codes properly. This caused the users to hold the PDA in awkward positions making the use of the system unnatural and uncomfortable. They proposed the separation of operation between PDA for control, playback and viewing, and the barcode scanner for content selection.

Our current study investigates a similar decoupling of interaction, leaving the data editing, entry, and viewing aspects on the PDA and consigning manipulation to a tracked token.

The manipulative component of interacting with CardTable entities may further be decomposed into two aspects: the categorical function (e.g., *what* we want to do with the object, such as 'copy', 'move', and 'select'), and the analog operation (e.g., *where* we want to move an object to). When the PDA is coupled with a tracked token for manipulation, PDA buttons may specify the categorical aspect, while the location of the token furnishes the analog aspect of the operation. The analog aspect of the operation may be readily replaced by the use of pointing hand gestures, and the categorical aspect may be replaced by voice (i.e., saying what we want to do). The studies reported in this paper investigates the relative benefits of these methodologies.

2.3 Related Interaction Methodologies

We review interaction methodologies related to our design to situate the contribution of this research. We first overview interaction with data-rich environments within the framework of large high-resolution displays. We then visit the various methodologies related to our design.

2.3.1 Interaction with Large Displays Via a PDA

Large high-resolution displays have recently become increasingly prominent [30]. A number of configurations [30, 4, 25, 3] and interactions [13, 3] have been investigated. However, there have been few studies in the area of interaction with large high-resolution tabletop display systems.

That tabletop displays seem to lend themselves to touch interaction and use in tandem with portable computation devices such as laptops and PDAs, is supported by recent research on such systems. The DiamondTouch system [8] and the associated DiamondSpin toolkit [29], for example, have spawned a number of interesting applications like SIDES [23] and UbiTable [28]. SIDES [23] takes advantage of the collaborative aspects of DiamondTouch to support learning by Asperger's subjects using touch-based interaction. UbiTable [28] extends the touch interaction by including laptop systems to address the rich data problem. In UbiTable, the tabletop becomes the medium across which two collaborators with laptops can share information. Similarly, i-Land [31] was a room-based ubiquitous computing environment that allowed users to easily move information back and forth between local machines like laptops and PDAs and shared surfaces like walls and tables. In the context of our work, i-Land represents an example where portable devices serve as interactive entities within a larger display environment. Magerkurth et. al. [17] also use PDAs with their large display STARS platform, but the PDA plays a different role. Rather than using the PDA as a peer that connects to the system, it serves as a smart controller in combination with tracked games pieces and voice commands.

As we discussed in Section 2.2, portable computation platforms may function with large display or physical interaction spaces to provide information access (e.g., visualization, and editing) or for object manipulation (e.g., selecting and moving data entities). Our previous work [2] suggests that mixing these two functions can lead to awkward interaction. Similarly, Klemmer et. al. [16] also found that conflating access and manipulation in the same device was problematic.

2.3.2 Physical Object Interaction

For tabletop systems, one interaction technique that is employed is the use of actual physical objects as manipulators [15, 33, 22, 24]. Tangible devices can take advantage of the natural affordances of flat, horizontal surfaces to support 'natural' interaction with data entities through actual physical manipulation. The chief claim of such tangible interaction approaches is that the tactile nature of the physical manipulators engages more of our senses than typical virtual interaction without physical manipulators [34].

The token-based interaction we propose in this paper can be thought of as a tangible proxy by which tabletop entities are manipulated. The difference in our work is that this manipulator is used multimodally in conjunction with either a PDA or voice interaction, such that token-manipulation provides the analog component of the interaction while the PDA or voice furnishes the categorical aspect of the operation.

2.3.3 Touch-Based Interaction

Touch-based interaction is another popular technique in tabletop applications [35, 27, 9, 8]. One of the most often argued benefits of such interaction is the notion that interacting with an application by directly touching graphical elements is a more natural or compelling approach than working indirectly with a mouse or other pointing device. However, in traditional desktop display settings, there is some evidence [19] that indirect mouse input may equal or outperform direct-touch input when the task requires just a single point of contact. For tabletop displays, Forlines et. al. [11] did a systematic investigation to examine the difference between direct-touch and mouse input for unimanual and bimanual tasks. The results not only show that users benefit from a direct-touch tabletop when performing bimanual tasks, but also raise questions as to the appropriateness of a direct-touch tabletop interface for a single user working on tasks requiring only single-point interaction. While a direct-touch input modality may not lead to greater performance in terms of speed and accuracy for unimanual tasks, other considerations, such as fatigue, spatial memory, and awareness of other's actions in a multi-user setting, might convince a system designer to choose single-finger input over single-mouse input in a tabletop environment. Although, the work of Forlines et. al. [11] looks at the benefits of direct-touch, such investigations were completed using a low-resolution display. We build upon this knowledge with our investigation using a large high-resolution display.

2.4 Voice Interaction

Voice interaction provides a natural way for the user to interact with a system. Since the use of voice is natural, it can lighten the cognitive load of the user, enabling them to focus on the task at hand [6]. It is often complementary to other input devices [5, 21]. In the case of multimodal interaction, the use of voice frees up ones hands because it can be used at a distance [12]. By not confining the user to a particular piece of hardware when using voice interaction, the ability to multitask is enhanced[26].

While voice interfaces have these advantages, they also have some disadvantages. Voice recognition systems can sometimes misinterpret voice input because of background noise and similar sounding words [21]. Many voice recognition systems require training, which can be quite cumbersome [26]. Also, there is lag time associated with voice recognition so users often do not get immediate feedback after speaking to the system [7].

We chose the use of voice for input because it provides a natural way for the user to communicate with the system [26]. Voice interaction compliments other interaction devices by supplying categorical information such as the change of a mode of operation, freeing the manual device to provide analog information such as distance of motion or degree of an operation. The portability of voice frees up the user's hands to use other devices [12].

2.5 Mouse Interaction

Although the use of a mouse as an interaction medium is common with most personal computer systems, it has been shown that using an unaltered mouse on large high-resolution displays poses problems for interaction with the display [30]. While Springer et. al. [30] reviews vertical large-format displays, it seems feasible that the problems of using a mouse on a vertical high-resolution display would be the same or even worse on a horizontal high-resolution display. As discussed earlier, Forlines et. al. [11] studied the relative benefits of mouse vs direct-touch interactions on a low-resolution tabletop display. However, Ball [4] notes that large form-factor displays need to be treated differently when it comes to interaction, which is why we do not test the use of a mouse in our experiment.

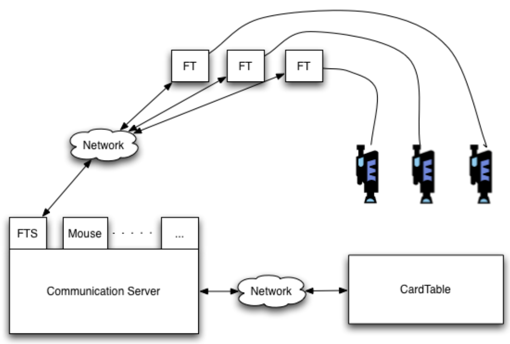

Figure 2: Layout of the communication infrastructure. The finger tracker server (FTS) is a service registered with the communication server. The FTS receives finger positions from the finger trackers (FT) each connected to a camera hanging above CardTable.

3. SYSTEM DESIGN

3.1 Hardware

The hardware setup is centered on nine upturned VisBox VisBlocks (`http://www.visbox.com`) arranged in a 3x3 grid to form a horizontal workspace. Each VisBlock is a freestanding, rear projected unit providing a 24"x13.5" surface capable of a WXGA resolution of 1280x720. This produces an approximately 6'x3.5' table with a resolution of over 8 million pixels – far greater than could be accomplished with a single projector. The table is managed by a cluster of six PCs running Fedora Core 5. PDA interaction with the system is implemented on a wireless HP iPAQ hx2795 Pocket PC running Windows Mobile 5.0. A near-infrared Vicon MX motion capture system (`http://www.vicon.com`) provides positional information by tracking small retro-reflective balls that are attached to the sides of the PDA and the edges of the table. This allows the PDA to be tracked not just on the table, but anywhere in the near vicinity as well. The token is a simple piece of square cardboard fitted also with retro-reflective balls for tracking.

We use three overhead Sony DCR-VX2100 cameras for the gesture recognition system each driven by a separate Apple desktop: one dual-processor G5 2.5GHz with 6GB RAM running OSX 10.4.11, one dual-processor dual-core Xeon 2.66GHz Mac Pro with 2GB RAM running OSX 10.4.10, and one dual-processor quad-core Xeon 2.8GHz Mac Pro with 2GB RAM running OSX 10.5.2. The camera orientations over the table are unimportant since we use a configuration program that creates transformation matrices from the camera spaces to the table space in order to provide correct mapping between the cameras and the table.

3.2 Software

3.2.1 Gesture System

To provide users a means to manipulate cards gesturally, we developed a vision-based system to track 2D positions of fingertips on the horizontal display surface. The gestural interaction comprises four kinds of operations: 1. An object is selected when a pointing finger hovers over it for more than a second; 2. A selected object is moved with a smoothly moving fingertip; 3. An object is deselected when the finger is retracted quickly (e.g., by pulling a finger in to form a fist); and, 4. An object is 'pushed' onto a PDA when the PDA is brought close to the object and the finger flicks the object in the direction of the PDA.

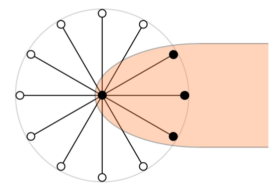

Figure 3: Star morphological operator for fingertip extraction.

The vision-based finger tracking algorithm first segments the hand from the rest of the image. We exploit the fact that the interactive surface is back-illuminated such that objects over the surface appear much darker by comparison. This allows us to extract the silhouettes of objects on top of the display by simple thresholding. We use a shape-filtering method to extract the tips of fingers.

For efficiency, we abstract our morphological structuring element as a 'star' configuration (see Figure 3). Fingertips are located where the center and three endpoints of the star are 'in' the foreground and the other points are background pixels. The system tracks fingertips over time by modeling the trajectory of the fingertip movement with a polynomial fitting curve over the previous k frames. This permits us to estimate the location of the fingertip in the $k + 1^{th}$ frame. A least squares method is used to estimate the polynomial parametrically in both x and y.

3.2.2 Architecture for Interaction

We use the CardTable software system described in [2]. We also developed an *interaction service* architecture to accommodate additional interaction methodologies. The system features a *communication server* that presents *service provider* systems with a simple API. The finger tracking system, for example, functions independently from the CardTable software, and simply communicates tracked positions through the *communication server*. Hence, interface components can be added to the testbed without modifying the main code structure. New interaction systems may be added by a simple *service registration* procedure. An overview of this layout can be seen in Figure 2. In accordance to our categorical/analog interaction decomposition, services may provide analog data (e.g., tracked locations) or categorical information (e.g., mnemonics indicating mode change).

For voice, we used the Microsoft Speech SDK. The system uses a small set of categorical commands, *grab*, *drop*, and *link*, which makes it easy for the participants to remember. The *CardTable* and *Escape* keywords activate and deactivate the voice recognition system.

4. USER STUDY

4.1 Experiment Design

The study consisted of five components where fifteen participants were asked to use a PDA, a tangible token, gesture, or voice to complete five tasks. The tasks were 1. to select a card, 2. move a card and place it below another card, 3. create a link between two cards, 4. arrange cards in a specified order, and 5. edit the text on three cards. Prior to each component of the study, participants were given time to familiarize themselves with the interaction techniques for the specific component.

In the first three components of the study, participants were asked to use a PDA, a tangible token, and gesture to interact with the horizontal display to complete the five tasks. The order of the use of these three interaction techniques was randomized. The fourth component enabled participants to choose between using a PDA, tangible token, and/or gesture to complete each of the same five tasks. The fifth component was similar to the fourth except that voice was used for mode changes in place of the PDA buttons.

Participants were timed as they completed each task and were given a questionnaire at the conclusion of each component. A five-point Likert scale was given to participants to rate how they liked using the various interaction techniques for the different tasks. They were also asked to choose the interaction techniques that they preferred to use when performing each task. This excluded voice since voice was only used for mode changes.

A challenge for our comparison study was to have a finger tracking system that could work over such a large display surface. Conventional tracking systems are not designed to track over a surface this large with the necessary accuracy. Our system must track an object as small as a finger with three cameras each having a 320x240 resolution covering an approximate 2' by 3.5' area. Despite its need to be more robust, the system allowed us to complete our experiment albeit with some halting and inaccuracies due to its prototype state.

4.2 Results

4.2.1 Statistical Analysis

We analyzed questionnaire data, chosen interaction techniques, and task completion times to determine the best interaction techniques for interacting with large horizontal displays. The data were analyzed using a one-way Anova test at a 95% confidence interval.

Using the questionnaire data, we first compared each device based on their overall ratings. Participants were asked to rate the devices on a five-point Likert scale. A rating of one was given for a device that was very difficult to use and a rating of five was given for a device that was very easy to use. Task 5 was not included in this analysis since the PDA was the only device that could be used to edit text. The means were 2.93, 4.52, and 4.45 for gesture, PDA, and token, respectively. We found that there was a significant difference between gesture ($p<0.0001$) and the PDA and token. The difference that places gesture apart from PDA and token is probably due to the reliability issues we had with our prototype gesture system. Interestingly enough, the PDA and token were, on average, rated about the same.

We then compared each device based on their ratings, but sorted them by task. Task 5 was not included in this analysis for the same reason as previously stated. The results can be seen in Figure 4. For task 1, the means were 3.13, 4.67, and 4.60 for gesture, PDA, and token, respectively. For task 2, the means were 2.73, 4.80, and 4.67 for gesture, PDA, and token, respectively. For task 3, the means were 3.53,

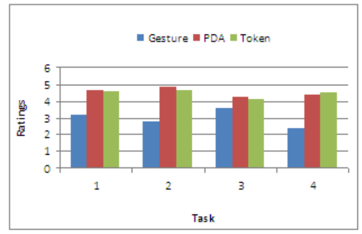

Figure 4: Ratings of each device sorted by task.

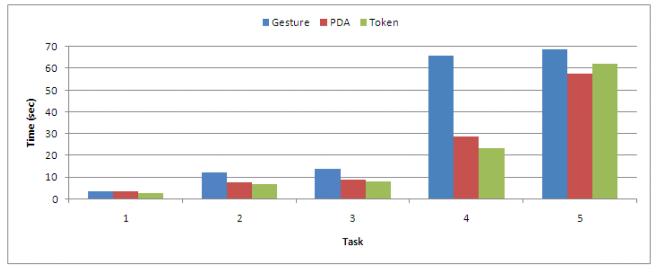

Figure 5: Comparison of each device based on their completion time sorted by task.

4.20, and 4.07 for gesture, PDA, and token, respectively. For task 4, the means were 2.33, 4.40, and 4.47 for gesture, PDA, and token, respectively. We found that there was a significant difference between gesture ($p<0.0001$) and the PDA and token for tasks 1,2, and 4; however, for task 3 there was not a significant difference between gesture, PDA, and token. The reason for the lack of difference in task 3 may be due to the fact that the user was required to put the card on the PDA. For gesture, this was simply accomplished by selecting a card with a finger and pushing the card into the PDA (or vice versa) and then using the more stable tracking of the PDA to accomplish the task. The difference that places gesture apart from PDA and token in tasks 1,2, and 4 is again probably due to the reliability issues we had with our prototype gesture system. Interestingly, the PDA and token were still rated similarly for all tasks with only slight differences.

Next, we compared each device based on their completion time (in seconds) but sorted them by task. The results can be seen in Figure 5. For task 1, the means were 3.49, 3.44, and 2.75 for gesture, PDA, and token, respectively. For task 2, the means were 12.35, 7.66, and 6.68 for gesture, PDA, and token, respectively. For task 3, the means were 13.65, 8.84, and 8.02 for gesture, PDA, and token, respectively. For task 4, the means were 65.94, 28.87, and 23.10 for gesture, PDA, and token, respectively. For task 5, the means were 68.59, 57.57, and 62.27 for gesture, PDA, and token, respectively. There was a significant difference between gesture ($p<0.0001$) and the PDA and token for tasks 2,3, and 4; however, for tasks 1 and 5 there was not a significant difference between gesture, PDA, and token.

Even though there was a problem with the robustness of the gesture system, the selection of a card still showed no difference with the other techniques. This explains why there was no difference for tasks 1 and 5 since this task only required a selection from the gesture system and then the user pushed the cards onto the PDA. When the tasks required

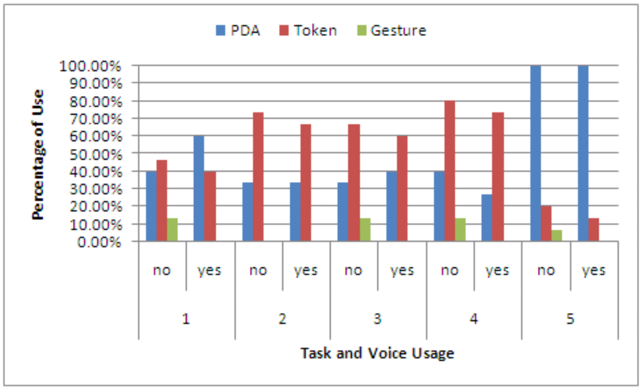

Figure 6: Effect of device usage based on using or not using voice.

more extensive movement (tasks 2, 3, and 4), the ill-effects of the gesture system can be seen effecting performance time. Despite this, for all tasks the PDA and token did not have any significant differences in completion times.

Next, we wanted to evaluate participant preference of voice for mode specification in place of the PDA buttons. Task 5 was not included in this analysis since voice was not used for text editing. Overall, the participants thought that voice was easy to use for tasks 1 and 2, and neutral towards using voice for tasks 3 and 4.

Finally, we investigated how the availability of voice interaction affected user choice of other interaction modalities. The results are shown in Figure 6. The x-axis of the chart shows each task and whether or not voice was used. The y-axis shows the percentage of times each modality was used in each task (i.e., if a modality was used by all participants in a task, its usage is 100%). Note that more than one modality may be used simultaneously. From the graph, we see that with the introduction of voice, token use decreased for each task and the use of the PDA increased or stayed the same except for task 4 where it decreased. These differences were not significant but analytically noticeable.

Another trend seen in this data is that in the absence of speech, users prefer bimanual interaction, using the token for the analog component, and the PDA for categorical (mode) selection (as opposed to combining both analog and categorical aspects in the PDA unimanually). This supports our division of these aspects.

Another observation for voice usage (the 'yes' columns in our chart) is that it assumes the categorical responsibility, and the PDA ceases to convey categorical information (i.e., the mode button is not pressed). Hence, the PDA becomes identical to the token, conveying only analog (where) information. Our data shows, however, that the same preference for the token applies to more complex manipulative tasks (tasks 2 - 4). Only in the simple selection task (task 1) did the PDA supersede the token for analog information. Our interpretation for this is the added weight of the PDA and the ability to set the token down on the tabletop biased the selection of device for manipulation to the token. For the simple selection (task 1), the weight and 'set down' influence was not significant, and some participants simply used the device (PDA) that was already in hand.

Task 5 required editing of data for which the PDA was needed. Since editing dominated this task (the only manipulation on the table was card selection as in Task 1), most users just used the PDA whether or not voice was available.

Gesture was generally neglected because of the reliability issues in our prototype (one participant reported that she ignored gesture because she "did not like" the modality in an earlier test because of its unreliability).

Overall, the PDA and token seemed to rank equally in time and by user ratings. This is interesting since our observational data, which is discussed next, shows obvious differences in the uses of these interaction techniques. The differences seen with the gesture system are hard to analyze as being meaningful due to some of the instabilities of its prototype state. However, we asked a few users if they would be more inclined to use gestures or rate it higher if it was more stable and they said they would. With the introduction of voice we saw an increase in PDA usage and a decrease in token usage, but the dominant technique used for each task was not affected significantly.

4.2.2 *Further Observations*

During our user study, we noticed some strengths and weaknesses of each interaction technique. We also saw how some techniques complemented others and worked together. Here we discuss our observations.

PDA: The PDA was the only device by which the user could solely perform all the specified tasks (i.e., movement, mode change, selection and editing). A stronger association is possible with the PDA since the user can see the information of the selected (associated) object on the screen of the PDA and also control its location on the display. Only one hand is required for control which leaves the user's other hand free for other uses such as text input.

There are a few drawbacks to the use of the PDA, most due to its physical characteristics. The PDA can block the crosshairs cursor that is projected on the table. The button's physical configuration is static on the PDA, requiring the developer to choose the buttons wisely that will be least awkward to the user. Some user's were afraid of dropping the PDA, as opposed to the token, when stretching over the table. This affected how they handled the PDA and their performance time negatively. Users also commented that the PDA was heavy to hold out over the table. Lastly, the metaphor of the PDA becoming the card and the card following the PDA can be a problem if the user wants to keep the card in the same location on the table while editing text. Hence, a strong association between the PDA and the card may or may not be desirable.

Token: For the token, there was one main positive aspect that most user's commented on. The token gives the user the ability to decouple the manipulation of the card on the table from the interaction with the card's content on the PDA. The users also liked the idea of having one hand on the PDA to specify the kind of operation, and one hand for movement of objects (holding token). This allowed the users to keep the viewing interface of the PDA close to them during all interactions. The light weight of the token, compared to the PDA, was very favorable among the participants. The token was also able to give a physical presence to a digital object and allow the user to keep the associated digital object in one location on the table as they viewed or edited the object's content on the PDA. This allowed users to preserve the organizational layout of the digital objects while interacting with their content. One can also slide the token across the table so that users can pass it around as we saw one user demonstrate.

User's were also less afraid of breaking the token because it was made of inexpensive materials. As a result of this construction, the physical configuration of the token is dynamic in the respect that it can be crafted to meet the desires of the user. If the user wants to be able to see through the token to the display so as to not hide the crosshairs cursor, a token of translucent material can be created.

The biggest problem that was observed is the inability to enter rich data, such as text, with the token alone. This requires a second device, such as a PDA, in another hand in order to provide this functionality. Due to this, two hands are necessary for some interactions and requires the user to juggle the token and PDA, especially if text input is required. Some users noted that the token could use buttons so the use of the PDA would be unnecessary for some actions, which would make the token more autonomous.

Gesture: Gesture provides benefits not easily replicated by the previous methods, such as decoupling movement from the PDA and placing it on the user's fingers without any extra user hardware. This leaves one hand free for data input on the PDA *and* for movement. Many users noted that the gesture interaction is simple and familiar and since their hand is used, this means the interaction is light with respect to weight.

Some drawbacks to using gesture is more extensive training may be required, depending on the gesture system. Our system, since the tracking is difficult, is sensitive to finger position variations and hence, can lose track if the user is not careful. The sensitivity may also give false positives during interaction. Additionally, unless more complex gestures are defined, another device or voice interaction is needed in conjunction with gesture for mode specification.

Voice: A number of participants indicated that voice is a good channel for conveying mode information. The problems we encountered with the use of voice in our experiments came largely from recognition errors and delayed response, which were already known problems as noted in Section 2.4. Some participants expressed frustration over these issues.

Device Complements: There were some interesting ways in which the techniques complemented each other in different task situations and uses.

PDA and gesture were able to complemented each other well. Cards could be brought to the PDA using gesture without needing to move the heavier PDA around the table. Since the user could reach further with their finger than with the PDA, the number of cards in the user's reach increased. Since a card can be picked up on the PDA using a gesture or the PDA, then if one technique fails the other one can substitute. This is true for the PDA paired with either token, gesture, or voice and is the same observation made by Oviatt in[21].

One interesting observation made when the users were interacting with the voice commands is that most of the tasks could have been completed solely with voice and token. The PDA was only needed for the last task which required text input. Hence, no specialized device, like a PDA, is needed for simple interactions with objects.

5. CONCLUSION

We investigated the comparison of four interaction techniques, PDA, token, gesture, and voice, for use with a large high-resolution tabletop display. Our study shows the relative strengths and weaknesses of each device/interaction methodology, how some complement others, and the affordances that are most beneficial.

From our observations, we can draw some conclusions as to the benefits of the use of each device when interacting with a large high-resolution display. The PDA is a well-rounded interaction device that can handle simple (i.e. selection, dropping, linking) and more involved interaction (i.e. text editing), which is not as easily attainable by the other interaction devices. The token provides an excellent means of movement control because it is lightweight, robust, and easily configurable. The token also decouples the editing/viewing functions relating to card content from the tabletop operations of card manipulation. Due to these affordances, participants enjoyed pushing the token across the horizontal table as the card followed. Gestures give the user control over the system without needing any additional interaction devices, such as a token or PDA (with the exception of text editing). Lastly, voice allows users to capitalize on a natural communication channel, which frees up other resources for multi-tasking.

When looking at the devices in terms of categorical and analog information, this study shows that there are advantages in separating categorical and analog information specification into different modes (along the line of the Bolt 'put-that-there' multimodal adage [5]). This is proven by the preference of bimanual use of the PDA and token within the study. The PDA buttons were the categorical device and the token was the analog device. When categorical specification was available through speech, the PDA became an analog interaction device.

6. ACKNOWLEDGMENTS

This work has been partially supported by NSF grants This research has been supported by the U.S. National Science Foundation NSF ITR program, Grant No. ITR-0219875, "Beyond the Talking Head and Animated Icon: Behaviorally Situated Avatars for Tutoring," IIS-0624701, CRI-0551610, "Embodied Communication: Vivid Interaction with History and Literature," and, NSF-IIS- 0451843, "Interacting with the Embodied Mind," and "Embodiment Awareness, Mathematics Discourse and the Blind." We thank Ms. Michelle McLeese for assistance in editing.

7. REFERENCES

[1] Mori. *http://apokalypsesoftware.com/products/mori*, 2007.

[2] C. Andrews, T. Henry, C. Miller, and F. Quek. Cardtable: An embodied tool for analysis of historical information. 2007.

[3] R. Ball and C. North. Visual analytics: Realizing embodied interaction for visual analytics through large displays. *Comput. Graph.*, 31(3):380–400, 2007.

[4] R. G. Ball. *Effects of large, high-resolution displays for geospatial information visualization.* PhD thesis, Virginia Tech, 2006.

[5] R. A. Bolt. "put-that-there": Voice and gesture at the graphics interface. *SIGGRAPH Comput. Graph.*, 14(3):262–270, 1980.

[6] B. L. Chalfonte, R. S. Fish, and R. E. Kraut. Expressive richness: a comparison of speech and text as media for revision. In *CHI '91*, pp. 21–26, 1991.

[7] N. T. Dang, M. Tavanti, I. Rankin, and M. Cooper. A comparison of different input devices for a 3d environment. In *Proc. 14th European conf. on Cog. ergonomics*, pp. 153–160, 2007.

[8] P. Dietz and D. Leigh. Diamondtouch: a multi-user touch technology. In *UIST '01*, pp. 219–226, 2001.

[9] A. Esenther and K. Ryall. Fluid dtmouse: better mouse support for touch-based interactions. In *Proc. Working Conf. on Advanced Visual Interfaces*, pp. 112–115, 2006.

[10] EverNote. *http://www.evernote.com/*, 2007.

[11] C. Forlines, D. Wigdor, C. Shen, and R. Balakrishnan. Direct-touch vs. mouse input for tabletop displays. In *CHI '07*, pp. 647–656,2007.

[12] M. A. Grasso, D. S. Ebert, and T. W. Finin. The integrality of speech in multimodal interfaces. *ACM Trans. Comput.-Hum. Interact.*, 5(4):303–325, 1998.

[13] F. Guimbretière, M. Stone, and T. Winograd. Fluid interaction with high-resolution wall-size displays. In *UIST '01*, pp. 21–30, 2001.

[14] F. G. Halasz, T. P. Moran, and R. H. Trigg. Notecards in a nutshell. In *CHI '87*, pp. 45–52, 1987.

[15] H. Ishii and B. Ullmer. Tangible bits: towards seamless interfaces between people, bits and atoms. In *CHI '97*, pp. 234–241, 1997.

[16] S. R. Klemmer, J. Graham, G. J. Wolff, and J. A. Landay. Books with voices: paper transcripts as a physical interface to oral histories. In *CHI '03*, pp. 89–96, 2003.

[17] C. Magerkurth, M. Memisoglu, T. Engelke, and N. Streitz. Towards the next generation of tabletop gaming experiences. In *Proc. Conf. on Graphics Interface*, pp. 73–80, 2004.

[18] C. C. Marshall, F. G. Halasz, R. A. Rogers, and J. William C. Janssen. Aquanet: a hypertext tool to hold your knowledge in place. In *HYPERTEXT '91*, pp. 261–275, 1991.

[19] S. Meyer, O. Cohen, and E. Nilsen. Device comparisons for goal-directed drawing tasks. In *CHI '94*, pp. 251–252, New York, NY, USA, 1994. ACM.

[20] C. Neuwirth, D. Kaufer, R. Chimera, and T. Gillespie. The notes program: a hypertext application for writing from source texts. In *HYPERTEXT '87: Proceeding of the ACM conference on Hypertext*, pp. 121–141,1987.

[21] S. Oviatt. Taming recognition errors with a multimodal interface. *Commun. ACM*, 43(9):45–51, 2000.

[22] J. Patten, H. Ishii, J. Hines, and G. Pangaro. Sensetable: a wireless object tracking platform for tangible user interfaces. In *CHI '01*, pp. 253–260, 2001.

[23] A. M. Piper, E. O'Brien, M. R. Morris, and T. Winograd. Sides: a cooperative tabletop computer game for social skills development. In *CSCW '06*, pp. 1–10, 2006.

[24] J. Rekimoto, B. Ullmer, and H. Oba. Datatiles: a modular platform for mixed physical and graphical interactions. In *CHI '01*, pp. 269–276, 2001.

[25] A. J. Sabri, R. G. Ball, A. Fabian, S. Bhatia, and C. North. High-resolution gaming: Interfaces, notifications, and the user experience. *Interact. Comput.*, 19(2):151–166, 2007.

[26] T. W. Schneider and O. Balci. Vtquest: a voice-based multimodal web-based software system for maps and directions. In *Proc. 44th annual Southeast Regional Conf.*, pp. 300–305, 2006.

[27] S. D. Scott, M. Sheelagh, T. Carpendale, and K. M. Inkpen. Territoriality in collaborative tabletop workspaces. In *CSCW '04*, pp. 294–303, 2004.

[28] C. Shen. Ubitable: Impromptu face-to-face collaboration on horizontal interactive surfaces, 2003.

[29] C. Shen, F. D. Vernier, C. Forlines, and M. Ringel. Diamondspin: an extensible toolkit for around-the-table interaction. In *CHI '04*, pp. 167–174, 2004.

[30] J. P. Springer, C. Sladeczek, M. Scheffler, J. Hochstrate, B. Frohlich, and F. Melchior. A survey of large high-resolution display technologies, techniques, and applications. In *VR '06*, p. 31, 2006.

[31] N. A. Streitz, J. Geißler, T. Holmer, S. Konomi, C. Müller-Tomfelde, W. Reischl, P. Rexroth, P. Seitz, and R. Steinmetz. i-land: an interactive landscape for creativity and innovation. In *CHI '99*, pp. 120–127, 1999.

[32] SuperNoteCard. *http://www.mindola.com/snc/index.html*, 2007.

[33] B. Ullmer and H. Ishii. The metadesk: models and prototypes for tangible user interfaces. In *UIST '97*, pp. 223–232, 1997.

[34] B. Ullmer, H. Ishii, and R. J. K. Jacob. Token+constraint systems for tangible interaction with digital information. *ACM Trans. Comput.-Hum. Interact.*, 12(1):81–118, 2005.

[35] M. Wu and R. Balakrishnan. Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In *UIST '03*, pp. 193–202, 2003.