Understanding the Effects and Implications of Gesture-Based Interaction for Dynamic Presentations

Adam Fourney
afourney@cs.uwaterloo.ca

Michael Terry
mterry@cs.uwaterloo.ca

Richard Mann
mannr@uwaterloo.ca

David R. Cheriton School of Computer Science
University of Waterloo

Technical Report CS-2010-03

ABSTRACT

Gesture-based interaction has long been seen as a natural means of input for electronic presentation systems. However, gesture-based presentation systems have not been evaluated in real-world contexts. This paper presents the design and evaluation of Maestro, a gesture-based presentation system whose design was informed by observations of real-world practices. To understand the implications of gesture-based interaction, Maestro was deployed in a classroom setting for two weeks. The study results indicate that gestures that support interaction with content are valued most, as opposed to those that support slide navigation. Notably, past systems have only used gestures for slide navigation. Our study also revealed that the presenter would position, orient, and conduct himself in ways to more reliably perform gestures and perceive system feedback, and to avoid accidental gesture recognition. However, these behaviors negatively impacted presentation dynamics. Collectively, these results outline clear directions for future research.

Author Keywords

Gesture-based interface, electronic presentation, evaluation, computer vision

INTRODUCTION

The ubiquity of common web cameras, coupled with the capabilities of modern processors, is quickly making gesture-based interaction a realistic input modality for a variety of computing devices and applications. For example, gesture-based input has been demonstrated for robotic control [18], video game consoles [11, 8], and control of other home appliances [5].

In this paper, we focus on one of the most frequently proposed application spaces for this input modality, namely, electronic presentations. Electronic presentations are often offered as a natural domain for gesture-based input, in no small part because people naturally gesture at slides when giving presentations. It is also an obvious application space because an estimated 30 million presentations are given *daily* [12]. However, despite numerous prototype systems demonstrating this potential (e.g., [10, 17, 15, 4, 1]), little is known about how well this interaction style actually integrates with this activity. Instead, previous work has typically focused on the accuracy of the recognition technology, rather than the impact of this interaction style on presentations. Similarly, prior work has largely overlooked the question of how presenters naturally use gestures when giving presentations. Thus, past designs have not been informed by a more complete understanding of current real-world practices.

To address these gaps in the research literature, we used observations of the real-world practices of presenters to inform the design of a new computer vision-based gesture presentation system called *Maestro*. In contrast to previous systems which only supported slide navigation, Maestro enables presenters to use gestures to both navigate and interact with projected content. For example, presenters can dynamically highlight, expand, or collapse bullet points with Maestro. These capabilities were directly informed by observations that naturally occurring gestures are used to support *communication goals*, such as specifying what content is currently being discussed.

After an iterative design process, we evaluated Maestro in a two week field study. Maestro was used by one of the researchers to give 12 lectures to approximately 100 undergraduate students. These lectures were observed by a second researcher, who attended the lectures as an audience member. Students provided feedback through a questionnaire distributed at the end of the study. To the best of our knowledge, this study constitutes the first real-world, long-term evaluation of such a system.

The results of the field study confirm that gestures that support interaction with content (what we call *content-centric* gestures) appear to be the most effective and valued type of gestures in this context. In particular, the ability to zoom into figures and the ability to highlight talking points were well received by audience members. In contrast, *navigation gestures* were perceived to be less efficient than the use of a wireless remote. These findings suggest that previous systems have overlooked the most promising use of gestures in this domain, namely, to support rich interactions with projected content.

The field study also revealed how gestural input can noticeably alter the dynamics of a presentation. In particular, sensing needs encourage the presenter to position, orient, and conduct himself in ways that allow the presenter to more reliably perform gestures and perceive system feedback, while lessening the likelihood of false positive recognition errors. However, these behaviors can compromise the presenter's ability to fully engage the audience as desired. For example, the presenter must orient himself towards the projection screen to confirm that gestures reach their intended targets, and that the system has correctly interpreted any desired input. However, this orientation can cause the presenter to miss audience questions and feedback since it places a portion of the audience out of his field-of-view.

Collectively, these findings suggest clear paths forward for future work in this space. In particular, our results indicate that a promising area of research is examining how gestures can support richer interaction with projected content. For example, gestures could be used to control the parameters of plots or simulations to enable far more dynamic content than present-day animations allow.

Our findings also suggest the need to consider how these systems can be better designed to minimize potential side effects on presentation dynamics. For example, the observed need for reliable input with unobtrusive system feedback indicates that multimodal interfaces may be a promising way of addressing these issues, while retaining the benefits of gesture-based input. For instance, a wireless remote could be used for efficient navigation of slides, while gestures could be used for richer interaction with content.

The remainder of the paper is structured as follows. We begin by reviewing related work in the area of controlling presentations using hand gestures, then describe the types of natural gestures that arise when people give presentations. We present Maestro and describe its overall design, then describe our field study and its results. We conclude with directions for future work.

BACKGROUND

Numerous systems have been prototyped to demonstrate the possibility of using gestures to interact with presentations. In this section, we first describe early work performed by Baudel and Beaudouin-Lafon, which articulated the many challenges and needs associated with using gesture-based input in this context. We then survey previous systems demonstrating the application of this interaction technique in this context. Finally, we review results of a previous study that suggest the potential benefits of gestural input for controlling presentations.

Basic Needs and Constraints of Gesture-Based Presentations

Gesture-based interaction with presentations is not a new idea; the seminal research was conducted in 1993 by Baudel and Beaudouin-Lafon who developed the Charade presentation system [1]. Charade allowed presenters to navigate a HyperCard presentation via a DataGlove tracked by a Polhemus tracker. Using this system, presenters could advance slides, access a table of contents, and annotate slides with free-hand drawings.

One of the contributions of the Charade work was establishing a set of guidelines for the design of gesture-based presentation systems. In particular the authors noted that, since gestures are not "self-revealing", it is important to provide presenters with sufficient feedback to support their use of the system. Additionally, in order to support swift recovery from errors, the authors suggest that all gestures correspond to "fast, incremental, reversible" actions, and that the system provide a general "undo" operation. These issues are essential for constructing a useable gesture-based presentation system, but are often overlooked in other prototype systems.

The Charade work provides an important foundation for guiding the design of gesture-based presentation systems. However, few details are known about its effectiveness in a real-world setting. While Charade was used to present two sample presentations to an audience, only recognition rates were reported; no feedback from presenters or from the audience was reported. Nevertheless, Baudel and Beaudouin-Lafon's research remains one of the most thorough investigations of gesture-based interaction with presentations.

In summarizing the implications of their findings, the authors suggested that future work consider using computer-vision to enable gesture-based interaction, noting that the DataGlove was a major limiting factor in their system. More than 15 years have now passed since Charade was first developed. Inexpensive web cameras are widely available, and modern processors make computer vision a viable alternative for detecting hand gestures. As such, numerous gesture-based presentation systems have been prototyped using computer vision as the sensing technology. We review these systems next.

Computer Vision-Based Presentation Systems

The literature contains numerous examples of gesture-based presentation systems that use computer vision to detect the presenter's hand gestures. The majority of these systems are quite simple, and support linear navigation with the help of two onscreen buttons: one for the "next slide" command, and one for the "previous slide" command (e.g., [15, 14, 3]). To activate these buttons, presenters rest their hands over the buttons for a brief period of time (a gesture known as "dwelling"). Alternatively, the *FreeHandPresent* system by Von Hardenberg and Bérard [17] allows presenters to issue commands without buttons. Instead, it uses hand posture to differentiate between commands. For example, two outstretched fingers indicates the "next slide" command, while three outstretched fingers indicates the "previous slide" command. As with buttons, the presenter must hold their hand still for a brief period in order to issue the command. While a simple convention for issuing commands, these dwell-based gestures have a significant drawback: it is difficult to set a satisfactory duration for the dwelling. If the dwell duration is set too high, the system feels unresponsive. Conversely, setting the duration too low leads to the "Midas touch" problem

[7], where gestures may inadvertently be activated whenever the hands rest.

Gesture recognition strategies other than dwell detection have also been explored. For example, a number of systems detect dynamic gestures defined by a hand's path through space (e.g., [10, 2, 8]). Most notably, Lee and Kim's PowerGesture [10] system enables gesture-based control of a Microsoft PowerPoint presentation using ten separate gestures based on continuous hand motion. The various gestures allow users to navigate the presentation (e.g., advance slides, or quit the presentation), but do not support interaction with individual elements within the slides. In this sense, PowerGesture is similar to Charade, but uses computer vision rather than a DataGlove to sense hand gestures.

While each of these systems demonstrates the possibility of using computer vision to enable gestural control of a presentation, the literature describing these systems reports only the recognition rates of the various approaches. None of these systems were used to present actual slideshows to real audiences, and there are no details regarding the implications (either beneficial or detrimental) of this form of interaction. Furthermore, there is no indication that these systems' designs were informed by current real-world practices, calling into question the appropriateness of their various design choices. In fact, to the best of our knowledge, the literature contains only one study that begins to examine these issues, which we describe next.

Understanding the Potential Benefits of Gesture-Based Presentations

To understand the benefits and limitations of various presentation control modalities, Cao *et al.* conducted a Wizard-of-Oz study comparing 3 different input modalities [3]. 6 individuals presented talks in front of test audiences, using either a standard keyboard and mouse, a laser pointer with a button, or hand gestures and a touch-sensitive surface to control the presentation. The audience members were asked to rate each presentation for clearness, efficiency, and attractiveness using a numeric scale.

In their results, hand gesture interaction consistently received the highest score in all categories, beating the laser pointer and the keyboard by a wide margin: 70% of the audience and 83% of presenters stated that they preferred the use of hand gestures. Moreover, audience members indicated that hand gesture-based interaction resulted in "a more personalized, humanized, story-telling style."

Cao's findings argue for the overall benefit of gesture-based input when giving an electronic presentation, but a number of important research questions remain. For example, their study relied on a Wizard-of-Oz simulation, rather than a functioning prototype, and also assumed a touch-based interface. As such, it is unclear how well actual computer vision-based systems would fare in these contexts. More importantly, Cao's study was limited in scale: gesture-based interaction was evaluated for a total of six talks, each of which was only five minutes in duration. It remains unknown how well these systems fare in more regular, day-to-day use.

Moving Forward

From this survey of related research, we note that while many systems have been prototyped, they have not been evaluated in real-world scenarios. Furthermore, past system designs do not appear to be extensively informed by real-world practices. Given the potential benefits of gesture-based interaction observed in the Cao study [3], we conducted research to understand how to design such systems to better integrate with real-world practices, and to explore the impact these systems have on real-world presentations. To inform the design of our system, we identified common practices of presenters when giving presentations. We describe these practices next.

UNDERSTANDING COMMON PRESENTATION PRACTICES

To understand common presentation practices, we observed 10 talks posted on Google's "Tech Talks at Google" website [6]. These videos consisted of 10 individuals lecturing for a total of approximately $7\frac{1}{2}$ hours. In each instance, the presenters used an electronic presentation that was front-projected onto a small projection screen typical of a classroom or boardroom. Importantly, most areas of the screen could be accessed by the presenters. When analyzing these videos, we focused our attention on gestures directed towards the projection screen.

Across the presentations, we observed a wide variety of deictic (pointing) gestures, which presenters used in order to draw the audience's attention to specific aspects of the visual presentation. The three most consistent and common practices are summarized below:

- Presenters frequently pointed to bullet points or phrases. In addition to emphasizing or reinforcing talking points, this gesture helped signal the transition from one subtopic to the next as the presentation progressed.
- Presenters were often observed pointing to numerous items in rapid succession. This gesture was used to group objects or to indicate membership in a set. When referring to the entire set, presenters often waved their hands over all items, without indicating any one item in particular.
- In at least four cases, we observed presenters using two hands to "crop" or "frame" portions of a figure. This gesture served to clarify the precise portion of the figure being discussed.

The aforementioned deictic gestures were not unexpected, but they served to remind us that natural gestures tend to co-occur with speech and tend to be highly contextualized by the contents and the layouts of the slides. This observation is important because past systems have not supported these common practices; previous systems have instead relegated gestures to issuing navigational commands such as moving between slides. Accordingly, this finding suggests a previously unexplored space in which gestures are used to interact with content, rather than merely to navigate slides.

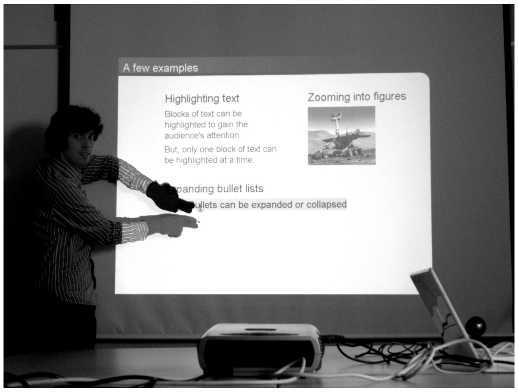


Figure 1. A user controlling the Maestro presentation system using only hand gestures.

In addition to the aforementioned deictic gestures, our observations also uncovered various trends in the formation of gestures. These trends permeate all gestures which we observed, and should directly influence the design of a gesture-based presentation system. Specifically:

- Presenters typically gestured from a position just outside the left or right edge of the projection screen, and rarely stood in front of the projected display. From this position, presenters avoided occluding the audience's view of the slide, and avoided the bright projector light which can distract or disorient the presenter [13, 16]. Gesture-based presentation systems can thus be designed to take advantage of presenters' tendencies to position and orient themselves in this manner.
- When forming gestures, a presenter's hand preferences appeared to depend mostly on his or her position relative to the screen. For example, when pointing, presenters used whichever hand allowed them to continue to face the audience while speaking. As a result, presenters used their hands interchangeably over the course of a presentation. Consequently, gesture-based presentation systems should allow gestures to be performed with either hand.
- Finally, presenters employed a wide variety of hand postures for the same gestures (e.g., pointing with one finger, two fingers, an open hand, or the hand seen edge-on). The choice of hand posture did not noticeably affect the apparent meaning of the gesture, suggesting that gesture-based presentation systems should avoid using hand postures to differentiate between commands as it would require one to significantly alter common behaviors.

As mentioned, previous systems have not recognized these common practices, nor their implications for design. In the next section, we describe the design of Maestro, which was directly informed by these observations of real-world practices.

MAESTRO

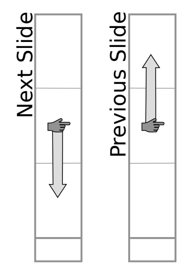
Maestro is a prototype gesture-based presentation system developed to explore the implications of gestural interaction with presentations (figure 1). Maestro relies on a *single* web camera for input, and a data projector for output. Together, these devices yield a highly portable presentation system that allows presenters to use hand gestures to control their presentations. This portability was essential in allowing Maestro to be evaluated in real-world contexts, such as multi-use classrooms.

Maestro's slideshows are structured similarly to those of PowerPoint, and are composed of a sequential deck of slides. Each slide can contain a combination of written text, bullet hierarchies, and embedded figures.

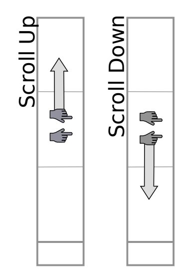
As with other gesture-based presentation systems in the literature, Maestro allows presenters to use hand gestures to navigate the slide deck (e.g., to advance slides). However, Maestro is distinguished by the fact that it also allows presenters to interact directly with the *content* of their slides (e.g., to zoom into figures, or to expand bullet hierarchies). These latter capabilities were directly inspired by the observations described in the previous section. We refer to these two classes of gestures as *navigation gestures* and *content-centric gestures*, respectively. Later, we will show that content-centric gestures are the most promising of the two classes of gestures in this context.

We now describe Maestro's navigation and content gestures in more detail. We then briefly describe how Maestro provides feedback and affordances to presenters, and how Maestro was implemented in software.

Presentation navigation

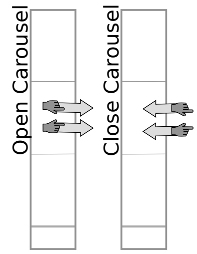


Maestro's navigation gestures allow presenters to move between slides, to scroll slides, and to bring up the slide carousel. These gestures are independent of slide content, and are thus performed in the left margin of each slide, a region we call the *staging area*. To move to the next slide, a presenter places one hand in the center of the staging area and moves the hand straight down. Likewise, to move to the previous slide, a presenter need only move their hand straight up, again starting from the center of the staging area. A set of horizontal ruled lines delineates the areas for invoking these gestures, but these visual guides appear only when the presenter rests their hand within the margin for a short period of time. Gestures can be performed even when the guidelines are not visible.



Unique to Maestro is the ability to navigate *within* slides: Maestro allows presenters to author slides whose content is longer than the height of the projection screen. This content can be scrolled by placing both hands in the stage's center region, and then moving one of the hands straight down. The slide responds by immediately scrolling down, and continues to scroll down as long as the hands remain in that particular configuration. The scroll speed is determined by the distance between the hands. Scrolling up is performed with a similar gesture.

Finally, Maestro allows presenters to open a carousel containing thumbnails of all slides in the presentation. To access the carousel, the presenter places both hands in the stage's center section, and then pushes the hands away from their body. Using other gestures, the presenter is then able to randomly access any slide.



Interactions with slide content

Maestro also affords gestural interaction with the actual content of the slides. Blocks of text can be highlighted by pointing to them with one hand. Presenters can also selectively enlarge figures embedded alongside text. When enlarged, a figure occupies the entire screen. To zoom into a figure, the presenter moves both hands into the figure, then pulls them apart vertically. These capabilities – highlighting points and enlarging figures to introduce more detail – were directly inspired by gestures identified in the observational study.

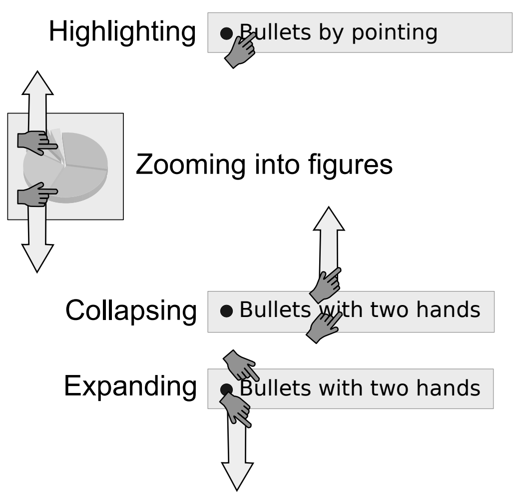


Figure 2. Gestures performed *directly* on slide content.

Finally, presenters can also author slides with hierarchical lists of bullets, with child bullets initially hidden. To reveal child bullets, the presenter places both hands next to the bullet point of interest, and slides one hand down, similar to the scroll gesture. The reverse motion hides the child bullet point. This capability allows the presenter to cater the detail of the presentation to the particular needs of the audience.

Affordances, Feedback, and Error Recovery

As noted by Baudel and Beaudouin-Lafon in [1], it is imperative that a gesture-based presentation system provide sufficient affordances and feedback to support the presenter's use of the system. Maestro renders all feedback to the display. Since the system display is shared between the audience and the presenter, and because affordances and feedback should only be directed at the presenter, all visual feedback is kept quite subtle. To achieve this balance, Maestro communicates command affordances via cursors that follow the hands as they move around onscreen. These cursors are small and translucent to reduce the chance that audience members can see them. At the most basic level, these cursors provide feedback by revealing where the system believes the presenter is pointing. The cursors are also augmented with *gesture mnemonics*, which serve to indicate *which* commands are available in a particular context (similar to context-sensitive mouse cursors), and to remind users *how* to perform the gestures (as in figure 3). Mnemonics are not meant as detailed gesture instructions, but instead serve to indicate the basic form and direction of the gesture.

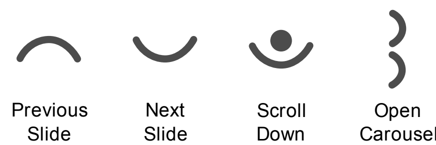


Figure 3. Several gesture mnemonics used by Maestro. The dot in the "'scroll down" mnemonic indicates the presence of a stationary hand.

Similarly, Maestro provides command feedback using translucent icons which appear in the staging area to reassure the presenter that a command has been received. These icons remain displayed for several seconds allowing the presenter ample time to confirm that the command has been invoked.

Finally, Baudel and Beaudouin-Lafon's work also stressed the importance of providing facilities to allow presenters to quickly recover from recognition errors. As such, Maestro supports a general "undo" command. In addition to supporting the undo operation, Maestro allows slides to be navigated using common keyboard commands (e.g., left-arrow, right-arrow), and allows context-sensitive commands to be issued using the mouse. These capabilities allow presenters to continue a presentation in case of technical difficulties with Maestro's computer vision and gesture recognition machinery.

Implementation

Maestro was developed with the expressed purpose of quickly and inexpensively exploring the implications of gesture-based interactions with presentations. The software consists of two separate processes that run concurrently on a single laptop computer. The first process, written in C, interfaces with a standard web camera and is responsible for detecting and tracking the presenter's hands. To simplify hand tracking, Maestro requires users to wear a mismatched pair of brightly colored gloves (one red glove and one blue glove). Hand detection and tracking can then be achieved using simple color thresholding techniques that are computationally inexpensive. Maestro's second process, written in Java, renders the presentations and performs gesture recognition. A simple template-based approach is used for motion recognition; various features of the hand trajectories (e.g., start/end location, path length, general direction of travel, moment of inertia, etc.) are measured, and are tested against manually tuned gesture templates.

While Maestro's hand detection and gesture recognition techniques are by no means state-of-the-art, together they yield a reasonably accurate and efficient prototype amenable to the needs of rapid iterative interaction design – an important criteria for this work. In terms of efficiency, Maestro's two processes each operate at between 15 and 30 frames per second. This allows for real-time interactions with the system. Maestro's gesture recognizer is also reasonably accurate. To establish the overall robustness and accuracy of the recognition system, five new users and one expert (one of Maestro's researchers) were asked to perform ten instances of each gesture. The system accurately recognized 86% of gestures for new users, increasing to 96% for expert users. In both cases, false positives accounted for fewer than 1% of all detections. These error rates compare favorably with those of similar systems. For example, Charade achieved an accuracy ranging from 72% to 84% for inexperienced users, increasing to between 90% to 98% for expert users [1] (Charade used a modified Rubine gesture recognizer). Similarly, Lee's PowerGesture system achieved an accuracy of 93% when using hidden Markov models to recognize gestures [10].

EVALUATION

To understand the effects of gesture-based input on presentations, we deployed Maestro in a classroom for two weeks. During this time, Maestro replaced PowerPoint as the main presentation system. Our field study was motivated to answer the following questions:

- How does gesture-based input compare to more traditional input modalities such as keyboards, mice and presentation remotes?
- What software features are most useful, and which need further refinement?
- How does gesture-based input fit in with current presentation practices? Does gesture-based input noticeably alter the dynamics of presentations?

Procedure

To answer these questions, one of the researchers used the system to give lectures to approximately 100 students over a two-week period. The lectures were part of a third-year university course unrelated to the research project. During this period, Maestro was used a total of 12 times to deliver six unique one-hour lectures (lectures were given three times a week, with the same lecture given twice a day). For each lecture, the lecturer carried in, set up, and calibrated the necessary equipment for deploying Maestro (lectures were given in two separate rooms). The specific equipment included a laptop, an external web camera, and the colored gloves (the rooms were already equipped with non-portable data projectors). Since the classrooms were used by other courses, Maestro's portability and ease of deployment was a necessary precursor to these trials.

As mentioned above, lectures were taught by one of Maestro's researchers. This researcher functioned in a supervisory role during Maestro's development, but he was not familiar with its specific implementation. Accordingly, he had to learn how to set up, calibrate, and use the system, as well as author content. Thus, while he was involved in the project, his experiences in using the system were closer to those of a first-time user. (In fact, there were many times when he needed to ask what features were available and how they were used.)

Before deploying Maestro, lectures were given for approximately eight weeks using PowerPoint controlled by a laptop keyboard. The laptop was located at a lectern in a corner of the classroom. The blackboard was also used occasionally during this time. After Maestro's deployment, lectures were given for two weeks using PowerPoint and a wireless remote control. While this evaluation did not attempt to perfectly balance the use of the various interaction mechanisms, it nonetheless serves to provide the first real-world comparison of three distinct control mechanisms, and includes the first longitudinal evaluation of a gesture-based interface to a presentation system.

Methods

For data collection, three of the six unique lectures were videotaped by a second researcher, who also took notes. Students were encouraged to provide feedback during lectures and were given a questionnaire at the end of the term to provide both structured and open-ended feedback. The questionnaire consisted of 40 questions that enabled students to compare the various presentation media (the blackboard, PowerPoint, and Maestro) across a range of dimensions, and to specifically evaluate Maestro's gestures and software features.

The questionnaire employed a 4-level Likert scale with responses ranging from 1 (strongly disagree) to 4 (strongly agree). While a five-point scale is more common (which includes a "neutral" option), a 4-point scale forces participants to indicate either a positive or negative expression of agreement to each statement. For data analysis, nonparametric statistical tests are most appropriate for Likert responses [9]; thus, this paper uses the *sign test* as the main statistical method for data analysis. A Student t-test is also commonly applied to Likert data. Both the t-test and sign test reveal similar trends in our data set, but the sign test is more conservative in measuring statistical significance.

Results

In this section, we first present results comparing the various presentation media. We then present students' evaluations of Maestro's design. Finally, we describe observations of the impact Maestro had on the presentations.

Approximately 70% of the students completed the voluntary questionnaire. As we summarize students' responses, we also incorporate open-ended feedback to complement the quantitative data.

Comparing Presentation Media

The first section in the questionnaire sought to compare Maestro with: a blackboard; PowerPoint with a keyboard and mouse; and, PowerPoint controlled by a wireless remote.

Participants rated each system independently in terms of interactivity, visual appeal, and efficiency. This portion of the questionnaire was very similar to the one used by Cao *et al.* in [3].

To analyze the data from this section of the questionnaire, the paired version of the sign test was used. This statistical test directly compares an individual's perceptions of one presentation medium to their perceptions of another presentation medium. In comparing the competing presentation technologies to Maestro, we found the following results (at a significance level of $\alpha = 0.05$):

- Maestro is considered more interactive than using the blackboard ($p < 0.001$), PowerPoint with a keyboard ($p < 0.001$), or PowerPoint and a remote ($p < 0.001$).
- Maestro is considered more visually appealing than using the blackboard ($p < 0.001$). However, no statistical difference was found when comparing the visual appeal of Maestro to that of PowerPoint using a keyboard ($p = 0.664$) or PowerPoint and a remote ($p = 0.832$).
- Maestro is seen as less efficient than PowerPoint using a remote ($p < 0.001$). No statistical difference was found between Maestro and the blackboard ($p = 0.627$); or between Maestro and PowerPoint controlled using a keyboard ($p = 0.076$). However, a low p-value in the latter case suggests a trend towards finding Maestro less efficient.

The mean scores across these dimensions and presentation media are presented in figure 4.

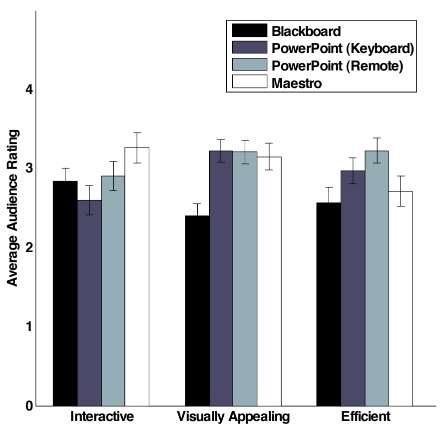


Figure 4. Mean scores for each of the presentation media. Error bars represent a 95% confidence interval about the sample means.

From these results we find that Maestro is considered to be more interactive than the other presentation media and input modalities. This result validates the notion that gesture-based input can positively enhance presentations. At the same time, Maestro was found to be less efficient than PowerPoint. This lower efficiency score is worthy of further investigation, but there are a number of potential reasons for this lower score. First, advancing slides requires a relatively large physical action: The presenter must orient himself next to the projected content, position the hand, then swipe it downward. This takes quite a bit longer than pushing a button on a remote that is already in hand. Also, the lower perceived efficiency could be partially attributed to delays caused by occasional gesture recognition errors. More work is required to determine the importance of these contributing factors.

Open-ended feedback from students highlights the perceived strengths and limitations of Maestro found in the quantitative data. For example, students appreciated the additional functionality afforded by Maestro:

> (Maestro) has advantages over remote devices because of an increase in the range of functions.

At the same time, students noted that navigating slides using gestures does not necessarily confer any particular advantages:

> Having a remote to switch slides is sufficient (...) and just as effective as Maestro.

These findings indicate that content-centric gestures, as opposed to those which enable navigation, are likely the most compelling reason to consider gesture-based interaction for presentations.

Evaluating Maestro's Features

The second section of the questionnaire asked participants to rate Maestro's specific software features. Maestro's ability to present figures in full-screen mode (i.e., by zooming in) was overwhelmingly welcomed by participants: 42% of the students *agreed*, and 52% *strongly agreed*, with the statement that "it is often useful to view figures in full-screen mode." The positive response to this feature is highly statistically significant, and represents the most positive response to any of Maestro's features.

The majority of participants also responded positively to the automatic highlighting of bullet points, with 64% indicating it was a useful feature. Again this result is statistically significant at the 5% level (p-value of 0.033). However, although the feature elicited a positive response, there are many opportunities for improvement. For example, one student commented that the bullet highlighting decreased readability because it placed a gray background behind black text, thereby reducing contrast. This is a legitimate concern, especially for audience members with visual disabilities. One way to address this issue would be to reverse foreground and background colors when highlighting bullet points. Manipulation of the text's typography is also a potential solution to this problem.

Students' open-ended feedback also support the data above. For example, one student stated:

> The system appeared to work fairly well, with some obvious issues with precision. (...) I did like the ability to zoom in on images and highlight points.

Audience members were also asked to evaluate the usefulness of scrollable slides. Here, only about 42% of participants thought that the ability to scroll slides up and down was advantageous. However, this feature was used only occasionally during the evaluation period. In part, this limited use of scrolling was due to slides being transcribed from existing PowerPoint presentations, which don't explicitly support this feature. Thus, more research is required to understand the potential benefits and drawbacks of this capability.

Finally, while the slide carousel was always available for use, this feature was never utilized during the two-week trial; there was little need to randomly access slides once the presentation was started.

From this questionnaire data, we now turn to observations of the effects Maestro had on presentations.

Emergent Behaviors

The two week field study enabled us to identify a set of emergent behaviors that result when using gesture-based interaction with presentation systems. In this section, we describe how this interaction style led to *grouping through highlighting*, the *anchor problem*, the *field-of-view problem*, and the introduction of a *no-fly zone*. As evidenced by the names, most of these behaviors highlight side effects that point to areas in need of future research for this problem space.

Grouping Through Highlighting

While the questionnaire revealed that the audience responded positively to bullet highlighting, the presenter also found Maestro's implementation useful in bringing attention to a *set* of bullet points all at once. In particular, Maestro implements a gradual fade-out of highlighted points, enabling the presenter to sweep his hand across a range of points, highlighting them all at once. This same waving or sweeping gesture, for grouping objects, was also noted in the observational study as serving a similar purpose. This mass-highlighting of bullet points was not planned for, but became a welcome emergent feature of the system.

The Anchor Problem

One of the most visible effects of utilizing gestural input was that it tended to "anchor" the presenter next to the screen so he could navigate the presentation (e.g., advance slides). While this side effect was previously noted by Cao *et al.* in [3], this anchoring led to a number of unexpected outcomes, which we expand upon.

Maestro's placement of the staging area caused the presenter to locate himself next to the left side of the screen. However, because the presenter must frequently face the screen to ensure that gestures are performed on their intended targets, the presenter found himself angling his body away from part of the class. This pivoting was not always corrected, leading the presenter to miss questions from students not in his field of view. Since the staging area was always incorporated into the left margin of the slides, it was always the same portion of the class whose questions were missed. In contrast, the location of the lectern (and, hence the laptop) in the corner of the room provided a clear view of the entire class.

The Field-of-view Problem

The tendency for the presenter to anchor himself next to the screen also made it difficult for the presenter to see all of the content being projected – what we term the *field-of-view problem*. After advancing to the next slide, the presenter would sometimes need to step back 4-5 feet from the screen to be able to see all of the slide's contents. From the audience's perspective, this behavior caused an obvious break from the presentation flow, and could be interpreted as the presenter being unprepared (when in fact the presenter simply needed to recall the points he wished to make). In contrast, when giving a presentation using a keyboard or remote control, the presenter was typically in a position to easily view each new slide in its entirety, whether it was on the laptop or projection screen. Glancing at a slide in these latter contexts is far less distracting since the presenter does not need to make a visible effort to look at the slide content.

The No-fly Zone Problem

Finally, the design of the gesture recognition system also created a "no-fly zone" – an area of the room that the presenter could not enter without the risk of distracting the audience. Maestro was designed with the assumption that the presenter normally stands to the side of the projected content, only occasionally entering the projected content to selectively interact with elements in the slide. This is a safe assumption to make, since the presenter typically wishes the audience to be able to view the content without interference. However, after the first day of lecturing, the presenter found himself forgetting about the system and fully immersing himself in the act of lecturing. At times, he would wander in front of the projected content to address the class, gesticulating as he did so. This would lead to constant activity in the slides behind him, with bullet points automatically highlighting and un-highlighting as the presenter's hands unknowingly moved over these objects. This created an obvious distraction for the class. Recognizing this issue, the presenter consciously reduced his travel into and *through* this area. Accordingly, this "no-fly zone" served to further limit the presenter's movements.

Discussion

As noted in the previous section, Maestro's exclusive use of gestures for issuing commands resulted in a number of side effects. Importantly, many of these effects are *not* specific to Maestro's particular design, but are more generally applicable to a wide range of systems which make use of hand gestures for interaction. For example, the anchoring and field-of-view problems are likely to occur with any gesture-based input mechanism that uses a large screen as a focal point, including those with touch-sensitive surfaces. Similarly, the no-fly zone problem is likely to arise with any system which uses computer vision for detecting gestures; recognition errors are always possible in such systems (even if they are rare), and users will almost certainly learn that error rates increase when they enter or gesticulate within the "active regions" where gestures are sensed (e.g., the area in front of Maestro's screen). Accordingly, users will learn to avoid those actions, and this may negatively affect the users' ability to perform their primary tasks. For example, in our specific case of presentations, these avoidance behaviors changed the presenter's basic presentation style in ways that were not always positive (e.g., by discouraging the presenter from gesticulating when standing in front of the screen).

These findings suggest that future systems must carefully design their interfaces to not only allow for reliable, fluid interaction, but also to reduce the likelihood of avoidance behaviors arising. While it is unreasonable to think that one can completely prevent such behaviors from forming, one can nonetheless make design decisions such that these behaviors will minimally impact the primary task. For example, a presentation system can be designed so that the presenters need not face the screen when issuing commands that do not depend on displayed content (e.g., by using a remote to advance slides, and by using audible or tactile feedback to inform the presenter that commands have been received). Additionally, it may be beneficial to allow presenters to selectively enable/disable portions of the gesture recognition subsystem. When gestures are disabled, the presenter will not need to worry about a "no fly zone" or other problems related to false positives. As an example, the system could disable content gestures when the presenter steps towards the audience (and thus away from the projection screen). Similar strategies can also be employed when building gesture-based interfaces in other application domains.

CONCLUSION

This paper has critically examined the intersection of computer vision-based gestural input and the application space of electronic presentations. In this research, we showed that content-centric gestures provide the clearest motivation for such systems, in contrast to previous work that has focused exclusively on using gestures for presentation navigation. We also showed how gesture-based input can lead to a number of side effects on presentation dynamics. These effects are most often caused by a presenter's desire to more accurately perform gestures and perceive system feedback, while minimizing the probability of false positives.

The results of this work suggest several areas for future research. First, and foremost, multimodal interaction seems to hold great promise in this area. For navigating slides, a wireless remote might be the optimal solution. This traditional form of interaction is reliable and it leads to efficient presentations. On the other hand, gestures seem well suited for supporting rich, direct interaction with slide content. Creating a system that elegantly balances the multiple input modalities should result in a more optimal experience for both presenters and audience members.

Given the value we found for interacting with slide content, there is a need to more fully explore this design space. For example, gestures could be used to manipulate the parameters of a mathematical plot or simulation. The benefits of such manipulations are well articulated by Douglas Zongker and David Salesin in their SLITHY presentation system [19]. SLITHY makes heavy use of parameterized diagrams and interactive objects using traditional input mechanisms. Extending this type of system to afford gesture-based control has yet to be explored.

REFERENCES

1. T. Baudel and M. Beaudouin-Lafon. Charade: remote control of objects using free-hand gestures. *Commun. ACM*, 36(7):28–35, 1993.
2. M. J. Black and A. D. Jepson. Recognizing temporal trajectories using the condensation algorithm. In *FG '98: Proceedings of the 3rd. International Conference on Face & Gesture Recognition*, page 16, Washington, DC, USA, 1998. IEEE Computer Society.
3. X. Cao, E. Ofek, and D. Vronay. Evaluation of alternative presentation control techniques. In *CHI '05: CHI '05 extended abstracts on Human factors in computing systems*, pages 1248–1251, New York, NY, USA, 2005. ACM.
4. P. Chiu, Q. Liu, J. Boreczky, J. Foote, T. Fuse, D. Kimber, S. Lertsihichai, and C. Liao. Manipulating and annotating slides in a multi-display environment. In *Proceedings of INTERACT 2003*, 2003.
5. J. A. Fails and D. Olsen Jr. Light widgets: interacting in every-day spaces. In *IUI '02: Proceedings of the 7th international conference on Intelligent user interfaces*, pages 63–69, New York, NY, USA, 2002. ACM.
6. Google. Tech talks at google. `http://research.google.com/video.html`, September 2008.
7. R. J. K. Jacob. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Trans. Inf. Syst.*, 9(2):152–169, 1991.
8. H. Kang, C. W. Lee, and K. Jung. Recognition-based gesture spotting in video games. *Pattern Recogn. Lett.*, 25(15):1701–1714, 2004.
9. P. J. Lavrakas, editor. *Encyclopedia of Survey Research Methods*, volume 2, page 429. Sage Publications, 2008.
10. H.-K. Lee and J. H. Kim. An hmm-based threshold model approach for gesture recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(10):961–973, 1999.
11. Microsoft Press. Xbox unveils entertainment experiences that put everyone center stage. `http://www.microsoft.com/Presspass/press/2009/jun09/06-01E3PR.mspx`, September 2009.
12. I. Parker. Absolute powerpoint: Can a software package edit our thoughts? In *The New Yorker: Annals of Business Section*, pages 76–87. Conde Nast Publications, Inc., May 28 2001.

13. G. Shoemaker, A. Tang, and K. S. Booth. Shadow reaching: a new perspective on interaction for large displays. In *UIST '07: Proceedings of the 20th annual ACM symposium on User interface software and technology*, pages 53–56, New York, NY, USA, 2007. ACM.
14. N. Sukaviriya, R. Kjeldsen, C. Pinhanez, L. Tang, A. Levas, G. Pingali, and M. Podlaseck. A portable system for anywhere interactions. In *CHI '04: CHI '04 extended abstracts on Human factors in computing systems*, pages 789–790, New York, NY, USA, 2004. ACM.
15. R. Sukthankar, R. Stockton, and M. Mullin. Self-calibrating camera-assisted presentation interface. In *Proceedings of International Conference on Control, Automation, Robotics and Computer Vision*, December 2000.
16. D. S. Tan and R. Pausch. Pre-emptive shadows: eliminating the blinding light from projectors. In *CHI '02: CHI '02 extended abstracts on Human factors in computing systems*, pages 682–683, New York, NY, USA, 2002. ACM.
17. C. von Hardenberg and F. Bérard. Bare-hand human-computer interaction. In *PUI '01: Proceedings of the 2001 workshop on Perceptive user interfaces*, pages 1–8, New York, NY, USA, 2001. ACM.
18. A. Xu, G. Dudek, and J. Sattar. A natural gesture interface for operating robotic systems. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 3557–3563, May 2008.
19. D. E. Zongker and D. H. Salesin. On creating animated presentations. In *SCA '03: Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 298–308, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association.