

Gesturing in the Wild: 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

ABSTRACT

Driven by the increasing availability of low-cost sensing hardware, gesture-based input is quickly becoming a viable form of interaction for a variety of applications. Electronic presentations (e.g., PowerPoint, Keynote) have long been seen as a natural fit for this form of interaction. However, despite 20 years of prototyping such systems, little is known about how gesture-based input affects presentation dynamics, or how it can be best applied in this context. Instead, past work has focused almost exclusively on recognition algorithms. This paper explicitly addresses these gaps in the literature. Through observations of real-world practices, we first describe the types of gestures presenters naturally make and the purposes these gestures serve when presenting content. We then introduce *Maestro*, a gesture-based presentation system explicitly designed to support and enhance these existing practices. Finally, we describe the results of a real-world field study in which *Maestro* was evaluated in a classroom setting for several weeks. Our results indicate that gestures which enable direct interaction with slide content are the most natural fit for this input modality. In contrast, we found that using gestures to navigate slides (the most common implementation in all prior systems) has significant drawbacks. Our results also show how gesture-based input can noticeably alter presentation dynamics, often in ways that are not desirable.

Categories and Subject Descriptors

H5.m. [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Keywords

Gesture-based interface, electronic presentation, computer vision

1. INTRODUCTION

Low-cost sensing hardware is quickly making gesture-based interaction a realistic input modality for a variety of applications. For example, gesture-based input has been demonstrated for control of robots [18], video game consoles [8, 12], and 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 an obvious application space for gesture-based input, in no small part because people naturally gesture at slides when giving presentations. Presentation systems have also historically served as a testing ground for showcasing the capabilities of new recognition algorithms. However, over the 20 years in which gesture-based presentation systems have been prototyped (e.g., [1, 4, 11, 15, 17]), prior research has largely ignored the question of how well this style of interaction integrates with this activity. Instead, previous work has typically focused on the accuracy of the recognition technology, rather than on the impact that this mode of interaction has on presentations. Similarly, past research has not critically examined how presenters naturally use gestures when giving presentations. As such, past designs have not been informed by real-world needs or practices.

To address these gaps in the research literature, we studied the real-world practices of presenters to inform the design of a new computer vision-based gesture presentation system called *Maestro*. In our observations of real-world practices, we found that presenters naturally use gestures to complement their verbal and visual communications. Consequently, *Maestro* is explicitly designed with facilities that enable presenters to directly interact with the projected content, in order to enhance their presentation of that material. For example, presenters can dynamically highlight, expand, or collapse bullet points with *Maestro*. In contrast, past systems have only provided the capability to navigate slides.

After an iterative design process, we evaluated *Maestro* in a field study which spanned several weeks. *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 evaluation of such a system.

The results of the field study confirm that gestures which support interaction with content (what we call *content-centric* gestures) appear to be the most effective and valued type of gesture 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 that gestural input can noticeably alter the dynamics of a presentation, often in ways that are un-

desirable. As an example, we found that the presenter tended to position and orient himself towards the projected content to ensure his gestures were accurately recognized. However, this orientation increases the chance that the presenter will miss questions and feedback from the audience since a portion of the audience is out of his field-of-view.

Collectively, the results from our study highlight the unique challenges incumbent in designing interaction for this application space, as well as clear directions for future research. In particular, gesture-based presentation systems should provide a rich feature set that enhances *presentation of content*, rather than just navigation of that content. At the same time, the discoverability, learnability, and execution of the systems' features should not noticeably alter presentation dynamics. These sets of goals are at odds with one another in a gesture-based presentation system, presenting challenges for those doing research in gesture recognition as well as in interaction design.

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.

2. BACKGROUND

While numerous presentation systems have been prototyped to demonstrate gesture recognition algorithms, there is comparatively less research examining how this mode of interaction affects presentation dynamics. In this section, we focus primarily on research that explores the latter concerns.

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. Finally, we review results of a previous study that suggest the potential benefits of gestural input for controlling presentations.

2.1 Basic Needs and Constraints of Gesture-Based Presentations

The seminal research investigating interaction issues with gesture-based presentation systems was conducted in 1993 by Baudel and Beaudouin-Lafon with their *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 primary 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 suggested that all gestures should correspond to “fast, incremental, reversible” actions, and that the system provide a general “undo” operation. These issues are essential for constructing a usable gesture-based presentation system, but are often overlooked in other systems.

While the *Charade* work provides an important set of design guidelines for gesture-based presentation systems, few details are known about its effectiveness in a real-world setting. Two sample presentations were given to an audience, but only recognition rates were reported; no feedback from presenters or from the audience was discussed.

In summarizing the implications of their findings, Baudel and Beaudouin-Lafon 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.

2.2 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 often only support linear navigation of slides. For example, with the help of two onscreen buttons, the presenter can navigate to the next or previous slide by resting their hand over the appropriate button for a brief period of time (a gesture known as “dwelling”) (e.g., see [3, 14, 15]). 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., [2, 8, 11]). Most notably, Lee and Kim's *PowerGesture* [11] system enables gesture-based control of a Microsoft PowerPoint presentation using ten such gestures. 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.

2.3 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]. Six 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.

2.4 Why People Gesture

While gesture-based presentation systems are often motivated by the observation that individuals tend to gesture when giving presentations, there is little work in this space that examines *why* presenters gesture. Perhaps most relevant to this latter question is a recent study examining presentation practices with the goal of understanding how best to leverage multiple data projectors in large lecture halls [9]. While this study did not focus on gestures specifically, it provides a good foundation for understanding the contexts in which gestures are most likely to arise. For example, the study reported that presenters gesture to approximately 17% of slides containing only text, increasing to 88% when slides contain figures or tables. The researchers also noted that gestures tend to draw the audience’s focus, and are used to “connect the audio and visual parts of a presentation”. While these findings help contextualize the use of gestures in presentations, few details were provided regarding the *specific* gestures used by presenters. This information is critical to help create designs that naturally integrate with current presentation practices.

2.5 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. Finally, while there is some work that sheds light on how gestures are used in presentations, there is still a need to more clearly understand the types of gestures people actually employ. Given these open questions, along with 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.

3. UNDERSTANDING COMMON PRESENTATION PRACTICES

To better understand common presentation-associated gestures, we examined 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.

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:

- **Simple pointing.** 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.
- **Enumerating.** 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.
- **Framing.** 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 highlight the notion that natural gestures tend to co-occur with speech and are highly related to the *content* of the slides. This observation is important because past systems have not supported interaction with content, and have instead focused on using gestures to issue navigational commands. Accordingly, there is a need to consider how gestures can be used to interact with content to enhance its presentation.

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. These trends are as follows:

- **Gesturing from the sidelines.** 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.
- **Ambidextrous gesturing.** 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.
- **Hand posture variability.** 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.

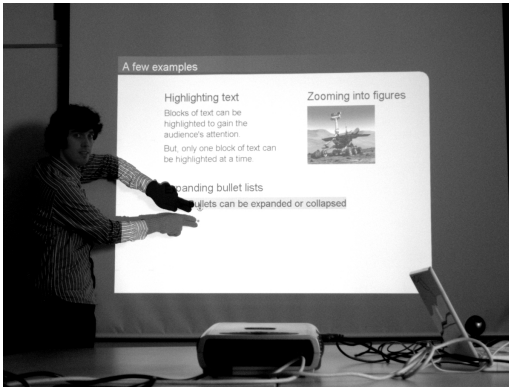


Figure 1: A user controlling the Maestro presentation system using only hand gestures.

These three findings, together with the observation that natural gestures are heavily contextualized by the screen contents, suggest the following design implications for a gesture-based presentation system:

- Systems should be designed to take advantage of presenters' tendencies to gesture from the sidelines.
- Systems should allow gestures to be performed with either hand.
- Systems should avoid using hand postures to differentiate between commands.
- Systems should be designed to leverage a presenter's tendency to gesture towards slide content in order to enhance the presentation of that content.

As mentioned, previous systems have not recognized these common practices, nor their implications for design. Mindful of these design implications, we developed the Maestro presentation system to better explore gestural interaction with presentations. We describe the design of Maestro in the section that follows.

4. MAESTRO

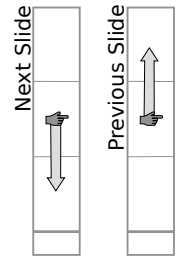
Maestro is a gesture-based presentation system comprised of an off-the-shelf web camera and data projector, and our custom software (figure 1). Maestro's slideshows are similar to those of PowerPoint, and are composed of a sequential deck of slides. Each slide can contain a combination of written text, 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.

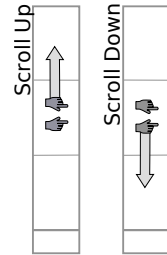
In the following subsections, we describe Maestro's navigation and content gestures in more detail. We then briefly describe how Maestro provides feedback and affordances to presenters, and how Maestro was implemented.

4.1 Presentation Navigation

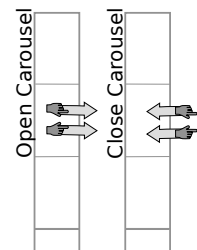
Maestro's navigation gestures allow presenters to move between slides, to scroll slides, and to bring up a slide carousel (described later). These gestures are independent of slide content, and are thus performed in the left margin of each slide – a region we call the *staging area*. To move to the next 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.



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.



4.2 Interactions With Slide Content

Maestro also affords gestural interaction with the actual content of the slides (figure 2). 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 practices identified in the observational study.

Finally, presenters can 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.

The gestures described above were the result of an iterative design process that sought to balance the need for learnability, ef-

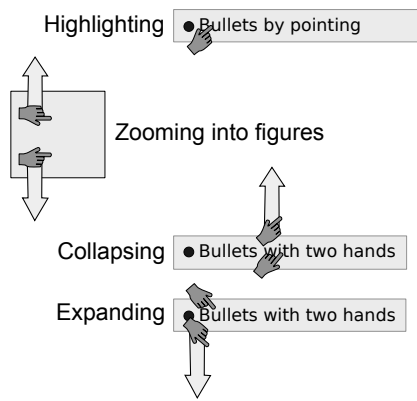


Figure 2: Gestures performed *directly* on slide content.

efficiency, rememberability, and expressivity of individual gestures, while maintaining the ability to robustly and reliably recognize gestures. These two sets of goals are often at odds with one another, which is one of the primary reasons why interaction design is challenging in this space.

4.3 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. To accomplish these goals, Maestro provides feedback through the display itself. However, these cues are designed to be subtle and unobtrusive to reduce their impact on the presentation content.

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. One of the most important functions of these cursors is to simply indicate 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.

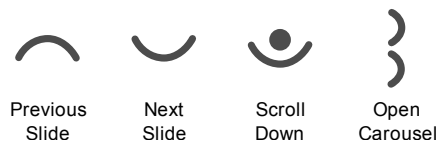


Figure 3: Several gesture mnemonics used by Maestro. The dot in the “scroll down” mnemonic indicates the presence of a stationary hand.

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.

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.

4.4 Implementation

Maestro was developed with the expressed purpose of quickly and inexpensively exploring the implications of gesture-based interactions with presentations. Given this focus on interaction, we were less concerned with developing a fully optimized recognition system and were more concerned with creating a system that was reliable and robust enough to examine questions related to the use of such systems. With this in mind, we briefly describe the implementation of Maestro.

Maestro’s 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 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 the expert user. 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 [11].

5. EVALUATION

To understand the effects of gesture-based input on presentations, we deployed Maestro in a classroom for several weeks. 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? How does it alter the dynamics of presentations?

5.1 Procedure

To answer these questions, one of the researchers used the system to give lectures to approximately 100 students for several weeks. 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 lectures, each approximately 50 minutes in length (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.

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. After Maestro’s deployment, lectures were given for two weeks using PowerPoint and a wireless remote control. The blackboard was also used occasionally throughout the whole term. 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.

5.2 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 study to provide both structured and open-ended feedback. The questionnaire consisted of 40 questions that compared the various presentation media (the blackboard, PowerPoint, and Maestro) across a range of dimensions. It also included questions that allowed participants to 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 [10]; 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 we found the sign test to be more conservative in measuring statistical significance.

Because the system was used by one of the researchers involved in the project, it is natural to question what biases are present in the data, as well as how generalizable the results are. With a researcher taking on the role of lecturer, the deployment study can be viewed as a best-case scenario, where the system is commanded by an expert user. Ultimately, however, the system was evaluated not by the presenter, but by the student audience. To ensure candid audience responses, questionnaires were submitted anonymously, and remained sealed to the researchers until the course marks were finalized. Approximately 70% of the students completed the voluntary questionnaire. The data collected from students clearly indicate the ways in which Maestro succeeds and fails, and suggest a number of ways in which future designs can learn from our experiences.

5.3 Results

In this section, we present the results derived from the audience questionnaire, beginning with results comparing the various presentation media. We then present the students’ evaluations of Maestro’s design. As we summarize students’ responses, we also include open-ended feedback to complement the quantitative data.

5.3.1 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 ($p < 0.001$), and PowerPoint with 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.

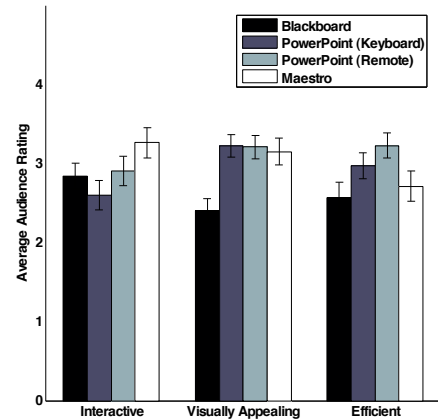


Figure 4: Mean scores for each of the presentation media. Error bars represent a 95% confidence interval about the sample means.

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.

Open-ended feedback from students highlights the perceived strengths and limitations of Maestro found in the quantitative data discussed above. 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.

5.3.2 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 welcomed by 94% of the 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 overwhelmingly 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). 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.”

Students 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, and this feature was used only occasionally during the evaluation period. Similarly, while the slide carousel was always available for use, this feature was never utilized during the field study; 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.

5.4 Potential Impacts on Presentations

As noted above, the evaluation was performed with one of the researchers lecturing to a class of undergraduate students. As such, we cannot expect the lecturer’s opinions of the system to be impartial. However, he experienced various difficulties with the system

that are worth discussing since they are likely to be experienced by other users of such systems. Specific instances of these difficulties were both noticed by the research-in-the-moment and through review of the lecture videos.

5.4.1 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], we found the issue to be further amplified by the tendency of the presenter to face the screen to ensure that gestures are performed on their intended targets. As a result, 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. 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 when standing next to the laptop to control PowerPoint presentations.

5.4.2 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.

5.4.3 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 activity created an obvious distraction for the class. After noting 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.

5.5 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” in which 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 that “avoidance” behaviors will arise. 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 could also be employed when building gesture-based interfaces in other application domains.

6. 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 have shown that content-centric gestures provide the clearest motivation for such systems, in contrast to previous work that has focused exclusively on using gestures for navigation of slides. We have also shown 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.

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 use of traditional input mechanisms. Extending this type of system to afford gesture-based control has yet to be explored.

In the larger context of general gesture-based interaction, our work reinforces the importance of carefully assessing the appropriateness and potential implications that this input modality might have on the user’s core task. While researchers have proposed that electronic presentations are an ideal application of gesture-based input for decades, our work has shown that, even in this ideal problem domain, gesture-based input is appropriate only for a subset of the interaction. With technologies enabling gesture-based *input*

becoming increasingly available, it is crucial to begin to develop a set of gesture-based *interactions* tuned to the unique needs and constraints inherent in this input modality.

7. 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. Lertsichai, 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] J. Lanir, K. S. Booth, and L. Findlater. Observing presenters’ use of visual aids to inform the design of classroom presentation software. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 695–704, New York, NY, USA, 2008. ACM.
- [10] P. J. Lavrakas, editor. *Encyclopedia of Survey Research Methods*, volume 2, page 429. Sage Publications, 2008.
- [11] 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.
- [12] Microsoft Press. Xbox unveils entertainment experiences that put everyone center stage. <http://www.microsoft.com/Presspass/press/2009/jun09/06-01E3PR.msp>, September 2009.
- [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. Podlasek. 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.