Communicating deictic gestures through handheld multi-touch devices

Clinton Jorge[1], Jos P. van Leeuwen[2], Dennis Dams[3], Jan Bouwen[4]

[1]University of Madeira, Madeira-ITI, Funchal, Portugal;

[2]The Hague University of Applied Sciences, The Hague, Netherlands; [3,4]Bell Labs, Alcatel-Lucent, Antwerp, Belgium

[1]clinton.jorge@m-iti.org, [2]j.p.vleeuwen@hhs.nl, [3]dennis@research.bell-labs.com
[4]jan.bouwen@alcatel-lucent.com

Abstract: **Deictic gestures are gestures we make during communication to point at objects or persons. Indicative acts of *directing-to* guide the addressee to an object, while *placing-for* acts place an object for the addressee's attention. Commonly used presentation software tools, such as PowerPoint and Keynote, offer ample support for *placing-for* gestures, e.g. slide transitions, progressive disclosure of list items and animations. Such presentation tools, however, do not generally offer adequate support for the *directing-to* indicative act (i.e. pointing gestures). In this paper we argue the value of presenting deictic gestures to a remote audience. Our research approach is threefold: identify indicative acts that are naturally produced by presenters; design tangible gestures for multi-touch surfaces that replicate the intent of those indicative acts; and design a set of graphical effects for remote viewing that best represent these indicative acts for the audience.**

Keywords: Multi-touch device, deictic gestures, remote presentations, slide presentations, grounding.

1 INTRODUCTION

Increasingly, knowledge workers work outside the traditional office, and more and more teams are distributed over multiple physical locations. Teams often communicate their work through PowerPoint (and other) slide presentations. These presentations tend to follow a standardized path: single slides display information with very little interaction from the presenter or audience [4]. In presentations that are attended or viewed online, the remote presentation generally is displayed as one of three situations: solely the slides being presented; the slides and the presenter's voice; and in some occasions with the addition of a video feed of the presenter. Experimental studies have indicated that merely linking spaces through audio-video links does not improve performance to the levels observed between side-by-side collaborators [8].

Communication is a collective activity of the first order. Studies performed by Hindmarsh et al [7], have demonstrated how communication and collaboration depend upon the ability of individuals to invoke and refer to features of their immediate environment. Many activities within collocated working environments rely upon the participants talking with each other and monitoring each other's conduct. When A speaks to B, A must do more than merely plan and issue utterances while B must do more than just listen and understand. A, must speak only when A acknowledges B is attending, hearing and trying to understand what A is saying, and B must guide A by giving A evidence that B is doing just this [5]. This mutual acknowledgment of understanding between A and B is called *Grounding in Communication*. During a conversation people tend to utter back-channel responders such as "uh huh", "yeah". In Grounding, these confirmations or negations of understanding are named Evidence. Positive evidences become more noticeable while conversing over a telephone or during teleconferencing activities where there is a deficiency of visual cues, such as facial expressions.

Pointing is one of the mechanisms for grounding in communication that require least collaborative effort between the communicating parties. Clark and Brenan [5] argue that deictic gestures combined with communicative statements help establish common understanding and that appropriate gestures that are easily interpreted are preferable over complex sentence constructions. Pointing is a deictic gesture used to reorient the attention of another person so that an object becomes the shared focus of attention. There are four important stages for performing a successful pointing gesture: Mutual orientation; Preparation and staging; Production of the gesture; and Holding (until confirmation) [3].

Directing-to and *placing-for* are two basic techniques for indicating [6], *Directing-to* produces a signal that directs the addressee's attention to an object; *placing-for* places an object for the addressee's attention. Graphical user interfaces in computers demonstrate the extended notions of these basic indicating techniques. A click is a virtual form of *directing-to*, and dragging is a virtual form of *placing-for*.

Baecker et al [2] performed studies on a moving point such as a screen cursor and laser pointer that defines the remote person's reference space. Baecker described the results as "giv[ing] them the gestural and referential capability of a fruit fly." Similarly, Kirk et al [8] argue "laser pointers have lower bandwidth for the expression

Corresponding author: Clinton Jorge, University of Madeira, Madeira-ITI, Caminho da Penteada, Funchal 9020-105, Portugal, +351964452305, clinton.jorge@m-iti.org

of gestural information than the direct presentation of hand gestures or sketches.”

In this paper we approach the issue of limited information bandwidth of deictic gestures in remote slide presentations (e.g., pointing with laser pointers) that hinder natural (deictic) communication in these settings. We argue that handheld multi-touch devices are capable of enhancing the representation of a presenter’s deictic gestures without introducing a steep learning curve or high cognitive load. We describe our theoretical framework and present the results of our experiments. We conclude by discussing our design guidelines and future work suggestions.

2 STATE-OF-THE-ART

Commonly, the mouse cursor or physical laser pointer are the tools used within collocated presentations as an extension of the performer’s gestures. Figure 1 shows a collocated presentation that was recorded and then broadcast online. They recognized and approached two issues for recording the local presentation for online visualization: how to capture the presenter and the slide projection within the same frame with enough quality to perceive both; and how to capture the presenter’s indicative gestures towards the slide projection. The cameraman positions the presenter to one of the sides of the video frame while the content being discoursed is augmented in the remaining portion of the frame. In this specific scenario the presenter uses a laser pointer to point to referents on the slide. Since the slides are augmented on the video (Figure 1, left), there is no visual feedback to where the presenter is pointing. To repair this detachment between verbal utterances and gestures, the cameraman pans the camera to capture the projected slide presentation (Figure 1, right), thus showing where the presenter’s laser pointer is located. At this point the audience can link verbal utterances to the laser pointer but at the cost of removing the presenter from the frame and viewing the content (slides) at a much lower quality.

Pointing gestures made towards a display (e.g. slide projection) are in general not retrievable at remote sites and participants are unable to tell what object has been pointed at. Lucero et al [9], describe an interactive wall-mounted display named Funky Wall, to support designers in easily conveying messages or ideas in the form of an asynchronous visual presentation. The authors designed four different proximity regions to act as individual interactive triggers. The closest region allowed users to record their gestures by augmenting them onto the content as white translucent streaks. Cheng and Pulo [3] proposed extending the reach of the performer of the gesture with a physical laser pointer, not only for indicative purposes but also as a direct interaction device. The authors argued the form of interaction would thereby reduce the cognitive load of the user and improve users’ mobility while interacting and performing actions. In [11], Tan et al presented a system capable of visually detecting pointing gestures and estimating the 3D pointing direction in real-time. The system offered at best an 88% detection rate and a 75% precision.

Keynote, FuzeMeeting, and other web conferencing tools currently support virtual laser pointers on their tablet applications. This resolves the interaction issues encountered with physical laser pointers but does not address the lack of gestural expressiveness or capture the larger array of presenter gestural intentions.

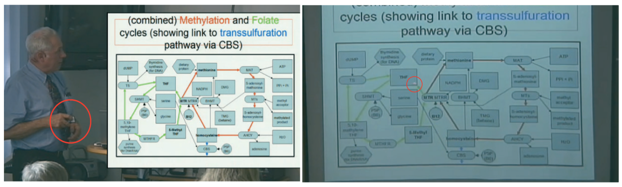

Figure 1: Two distinct repairs performed to enhance audience’s perception and understanding source: http://iaomt.media.fnf.nu/2/ skovde_2011_me_kroniskt_trotthetssyndrom

3 RESEARCH FRAMEWORK

We propose a theoretical framework (see Figure 2) for developing support for deictic gestures, which involves two entities: the presenter speaking and using the slides as a visual aid; and the audience to whom the presenter is speaking. The framework represents the presenter’s intention, which is to transmit a message to the audience. The gestures he performs are intentional, for example: directing the audience’s attention to a particular section of the slides. These intentions are exteriorized through gestures (in addition to utterances). The system recognizes the gestures and creates visual representations thereof as an effect for the audience to perceive. The audience then interprets their perception of the effects and creates their own mental model of what the presenter’s intention could be.

The framework is described in further detail during the subsequent subsections and guided our research methodology in this project.

3.1 Intent, Gesture, Effect, Perception

This project’s research activities were designed around the four key nodes of the theoretical framework: intent, gesture, effect, and perception.

3.1.1 Intent: Presenter

The *intent* node defines the high-level meaning for the performed gesture. The presenter has an intention and externalizes this by performing a gesture in order to, e.g., direct the audience’s attention to a specific part of a slide. Ideally, the addressees should easily understand the presenter’s intent and act accordingly.

3.1.2 Gesture: Presenter/Computer

The *gesture* node describes interactions gestured by the presenter based on his intentions and captured by the system—the handheld multi-touch device. These gestures may be triggers for events (navigation) or to communicate deictic gestures. In section 4.1 we present an experiment designed to understand what interaction can transform the presenter’s intent into gestures.

3.1.3 Effect: Computer

The *effect* node is the result of recognizing the gestures performed by the presenter and translating them into graphical effects displayed to the audience. Different effects are associated to different gestures (and therefore intentions), influencing the audience's interpretation of the effects and thereby of the presenter's intents.

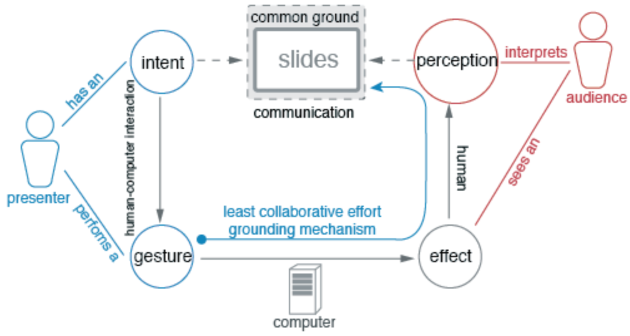

Figure 2. Proposed theoretical framework representing how the presenter through a multi-touch device transmits his intentions to the audience.

3.1.4 Perception: Audience

The *perception* node is the result of the effect (the remote representation of the gesture) being perceived and interpreted by the audience whose members create their own mental model of the presenter's intention. It is at this node that the effectiveness and value of our research is evaluated (see sections 4.2 and 4.3).

4 USER NEEDS STUDY

In our project we performed small-scale studies on each of the four nodes of the framework, focusing mostly on the *perception* node—in order to understand whether pointing effects added any meaningful information to a remote slide presentation. The latter study (see section 4.2) was designed involving thirteen subjects and performed in a lab setting at The Hague University of Applied Sciences. Subsequent refinements to the user experiment led to an online experiment (section 4.3) that involved nineteen participants.

An initial experiment was carried out that was related to the *intent* and *gesture* nodes of our framework, this experiment is described next.

4.1 Mapping Intentions to Gestures

The objective for this user experiment was to understand the connection between some common gestural intentions identified through observations and literature reviewing. The experiment, required subjects to perform the first gesture that came to mind when the researcher read out a pre defined "intent" (e.g., "point out the second bullet point"). A list of intents was created for each of four slides shown, where each intent required the subject to perform a gesture. The intents are categorized as being *pointing*, *indicating*, *highlighting*, or *grouping*. Subjects were seated in front of an iPad displaying a single slide in full screen running on the drawing application Adobe Ideas. Interactions were recorded, overlaying the displayed slide with a pen tool (50 pixel (similar size to finger tip) 50% transparency and red in color).

4.1.1 Findings

Twelve subjects participated in the experiment held in a lab environment. Subjects worked at Bell Labs in technology related positions and were over 35 years old. Four were novices and never used an iPad or multi-touch device, eight owned iPhones or were familiar with the technology. The results where analyzed individually and then compared to identify similarities or patterns.

A total of 134 gestures were recorded and observed. 31.34% of all recorded gestures were 1-finger pointing gestures (e.g., tap or touch on the device). 17.91% of all recorded gestures were grouping 1-finger gestures (such as circular gestures). For *pointing*, 11/12 subjects performed an index-finger indicative gesture equivalent, a tap or touch. For *indicating*, 9/12 subjects perceived this intent to be similar to pointing and performed an equivalent tap or touch gesture. For *highlighting*, often interpreted as a persistence technique using a semi-transparent coloring tool, 8/12 performed 1-finger dragged gestures to highlight text and 7/12 subjects performed a circular gesture to highlight individual artifacts. For *grouping*, 9/12 subjects grouped objects with a circular gesture.

Similar results (gestures) were found in Lucero et al's experiment [9] and can be categorized as "standardized multi-touch gestures".

We found that experienced users tend to simplify gestures, while novice users perform more personal, embodied gestures and techniques—especially for highlighting and relating content on paper.

4.2 Personal Perceptions of Pointing

The designed experiment required test subjects to view three video presentations on a laptop. The Repertory Grid Technique (Kelly 1955) [1] was used to elicit subjects' personal constructs (perceptions) and scoring without researcher bias, and was followed-up with a semi-structured interview. Each experiment required around 45 minutes to complete (depending on the interview).

Subjects viewed three videos subsequently, each a part of the same presentation. Each video was shown in a different visualization style, randomly ordered: slides and audio (A); slides and audio with an additional video feed of the presenter (V); and slides and audio combined with a virtual laser pointer – representing gestures (P).

Having viewed the three videos, subjects were asked to choose two presentation styles and compare these to the third, writing down the similarities or differences in their experiences, in the form of constructs. This was repeated for all possible combinations. Subjects then scored the three styles for each of these constructs, on a 7-point Likert scale.

The experiment was held at The Hague University of Applied Sciences over the course of a day. Thirteen subjects participated in the experiment, including students in design and engineering as well as professors.

4.2.1 Findings

Eight male and five female, subjects participated in the Repertory Grid (RGT) experiment and generated 96 construct pairs (e.g., "*helps concentrate* versus *distracting*"). These constructs were analyzed and subject preference (for a single presentation variant) was obtained based on the sum of scores: highest as the preferred presentation variant.

These participants scored the three variants as follows:

- 5/13 scored slides, audio and pointer (P) highest
- 5/13 scored slides, audio and video (V) highest
- 3/13 scored slides and audio (A) highest

Key differences were found between male subjects, who preferred the pointing (five-out-of-eight, 5/8) and disliked the video, and female subjects, who preferred the opposite (4/5). These results were consistent with the outcome of the semi-structured interviews that followed with each subject (see following subsections).

4.2.2 Male Subjects

Eight male subjects took part of the semi-structured interviews. During the interviews subjects were not bound to the three presentation variants thus 4/8 subject commented on preferring the combination of pointing and video (VP) (a style not included in the study). The interviews confirmed the disliking of the slide and audio (A) 1/8 and video (V) 1/8 variant. The pointing (P) variant received highest score of the displayed variants in the experiment with 2/8. Comments (6/8) about the pointing (P) included how pointing helped them "think like the presenter," because their "eyes are guided through the constructions" and "pointing directs you to important stuff on the slides." Two-out-of-eight subjects did not see the immediate benefit of pointing in remote presentations.

4.2.3 Female Subjects

Five female subjects took part of the semi-structured interviews. Female interview results were consistent with the RGT experiment. Four-out-of-five (4/5) female subjects preferred the video (V) variant while 1/5 preferred audio (A). Similar to the male interviews, 2/5 expressed preference for pointing and video integrated (this style was not included in the experiment). One subject commented on how pointing (P) was useful while three found pointing useful only for complicated or complex presentations, when guidance is needed.

During the semi-structured interview 4/5 of females preferred pointing for complicated presentations. They commented on the visual and kinetic aspect of the pointing cursor, that the drag effect was distracting and the motion erratic.

4.2.4 Discussion

Three–out-of-thirteen subjects that disliked the pointing (P) variant were professors. They commented on not liking to be guided and how they preferred to think for themselves.

During the interviews these similar comments arose on how pointing helped better understand the content and the thought process of the presenter in more complex scenarios such as graphs.

4.3 Significance of Pointing in Presentations

Another, online, experiment was performed with the objective of expanding on the findings of the previous studies. By refining the videos (shorter duration) and the pointing effect (improved effect and movement) we aimed to find further evidence of the benefits of pointing in remote presentation scenarios.

In this experiment, subjects accessed a webpage to view the three video presentation variants, again in randomly ordered styles. Skipping videos or parts of the video was disabled. Subjects were then asked to score each presentation style based on constructs resulting from the previous user study. The audio variant (A) was replaced by video and pointing (VP). None of the previous subjects participated in the online experiment.

4.3.1 Findings

Nineteen subjects completed the online experiment; eight females and eleven males, aged between 21 and 51. Subjects were recruited from three universities: University of Madeira, Eindhoven University of Technology and The Hague University of Applied Sciences.

No male subject preferred the video and slides style (V), while 4/11 preferred the pointing style (P). The combination of pointing and video (VP) scored the highest with 7/11. Interestingly, only one female subject preferred video and slides (V), while 4/8 (50%) of female subjects scored the pointing style (P) the highest. 3/8 preferred the combination of pointing and video. The contradiction in the female results with the previous study is remarkable. The female subjects seem very susceptible to the pleasantness of the effect and movement of the pointing cursor. These two attributes were refined for this study and the pointing was used only when required with the deictic utterances. Video was clearly less scored with only 1 out of 19 subjects preferring it. Video and Pointing scored the highest with 10 out of 19, while Pointing appeared second best, with 8 out of 19.

From our analysis of the scores on constructs, it appears that pointing (P) helped subjects to concentrate (high scores on the construct *concentrate*), while video (V) did not. Also, the combination of pointing and video scored low on the aspect of concentration, meaning that the added video is experienced as distracting. The same negative effect of adding video to pointing leads to reduced scores for *helpful* and *better understanding*.

Emotional, *personal* and *presence* constructs were scored lowest for pointing (P) with some exceptions of individual high scores. When analyzing the combination of video and pointing, these constructs—that scored highest in the video style (V)—suffer little to no reduction in their scoring. This led us to conclude that, while pointing does not add as much social presence, personal information and emotion as the video feed of the presenter does, it also does not negatively affect the qualities in the presentation as adding video does for the *concentration* construct.

5 DESIGN GUIDELINES

The *persistence* of the visual effects augmented to slides during presentations influences how the audience perceives gestures. Regarding the persistence as a spectrum, running from transient to persistent, we identified artifacts for both extremities of this spectrum. At the most transient extreme, the mouse cursor and laser pointer are located, which convey very little information (current location only). At the other extreme, we find persistent graphics, e.g. notes, highlights and annotations. These artifacts convey increased information but their persistency may not at all times be useful. Our contribution is to the intermediate spectrum that has not been fully explored: between transient, user cancelled events and slide exposure duration.

Indicative gestures are related in time to utterances and to referents (objects), thus no pointing cursor should exceed the duration of a slide exposure or be too transient to be missed due to late glances by attendees. We propose the following gestures and effects (pointing cursors, see Figure 3) for some of the most common gestural intents identified in slide presentations.

The *touch* cursor is similar to the laser pointer. It allows for referencing a single referent easily by moving around or by tapping at a location. The ripple effect provides an "epicenter-like" event, and provides a brief persistency, enough for late glancing addressees to view.

The *drag* cursor leaves behind a trail similar to a heat surface concept. This should allow for late glances to get enough feedback to follow the presenter's chain of thought throughout the slide and easily identify past referents and present ones.

The *sticky* cursor derives from the notion of the fourth stage of deictic pointing: holding. A little wiggle gesture places a cursor (a fingerprint) remaining there until the user cancels it or until the end of the slide exposure;, no continuous interaction is needed. Multiple objects can be referenced through multiple sticky cursors with different colors or shapes.

The *region* cursor surrounds a group of objects or an area of the slide, whereas the *shape* cursor (a repetition of the same gesture) highlights that area and is more persistent (during slide exposure). The *highlight* cursor is a two-finger gesture for highlighting text.

We argue that these effects should represent the majority of presenters' deictic gesturing needs and subsequently aid addressees' focus attention and follow the presenter especially in more complex (visual) slides.

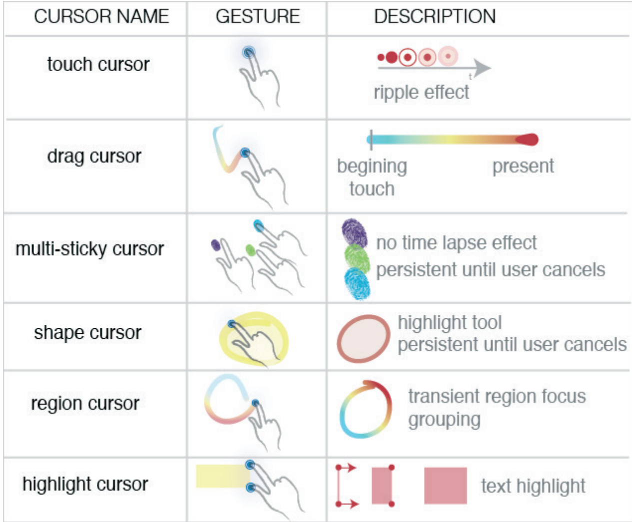

Figure 3. Gestures and effects for the most common gestural intentions in slide presentations.

6 DISCUSSION

Through the work presented here we argue that handheld multi-touch devices are a low-cost solution capable of facilitating deictic gestures. Our designs support more gestural intentions than the common laser pointer and mouse cursor thus increasing the expressiveness of the gestures in these communicative activities. While the cursors presented here have not yet been subject to user testing, we expect to find that they are representative of user intentions. The drag cursor could work in conjunction with face or eye gaze tracking software used for single person audience, allowing the system to recognize when the remote user is not looking at the presentation, triggering the drag effect.

Our studies show that pointing is considered a helpful tool for addressees in concentrating and understanding a presentation – in particular remote, distributed presentations. A combination of pointing with a video feed of the presenter provides the best of both worlds for some individuals. Our studies also indicate a substantial variation in the appreciation of these tools; a result that is not unusual when analyzing sex difference data from experiments [10]. This suggests that options to disable and show each one of these modal communication tools would be required.

7 FUTURE WORK

Our first prototype multi-touch app for presentations to remote audiences does not support all designed cursors. Future work would involve implementation of our cursor designs and further user studies for confirming and refining the gestures and the related effects. Deployment in real work environments would provide invaluable feedback from the presenters' perspective.

8 CONCLUSION

This paper focused on remote slide presentations and the lack of gestural expressiveness perceived by remote audiences. Deictic gestures are part of our natural language and are not fully supported in these scenarios.

Our objective was to explore this issue and present some guidelines to aid future research in the area. We present six cursor designs (deictic gesture representations) that we argue are representative of most deictic gestures and can be captured on a handheld multi-touch device.

9 ACKNOWLEGMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. ICT-2011-287760. This project received financial and operational support from Alcatel-Lucent Bell Labs, Belgium and Madeira Interactive Technologies Institute, Portugal, and operational support from The Hague University of Applied Sciences. We thank all participating subjects.

10 REFERENCES

[1] Alexander, P. and Loggerenberg, J. Van. The repertory grid: discovering a 50-year-old research technique. *... of the 2005 annual research ...*, (2005), 192 – 199.

[2] Baecker, R., Harrison, S., and Buxton, B. Media spaces: past visions, current realities, future promise. *CHI'08 extended*, (2008), 2245–2248.

[3] Cheng, K. and Pulo, K. Direct interaction with large-scale display systems using infrared laser tracking devices. *In Proceedings of the Asia-Pacific symposium on Information visualisation - Volume 24 (APVis '03), Tim Pattison and Bruce Thomas (Eds.), Vol. 24. Australian Computer Society, Inc., Darlinghurst, Australia, Australia, 67-74.*,

[4] Chiu, P., Liu, Q., Boreczky, J., et al. Manipulating and annotating slides in a multi-display environment. *Proceedings of INTERACT*, Citeseer (2003), 2.

[5] Clark, H. and Brennan, S. grounding in communication. *Perspectives on socially shared cognition*, (1991).

[6] Clark, H. 2003. "Pointing and placing." In: Kita S. Pointing: A foundational building block of human communication. *Pointing: Where Language, Culture, and Cognition Meet, Lawrence Erlbaum Associates, Inc., Mahwah, New Jersey 1 (2003): 1-8.*, (2003).

[7] Hindmarsh, J.O.N., Benford, S., and Greenhalgh, C. Object-Focused Interaction in Collaborative Virtual Environments University of Nottingham. *Design 7*, 4 (2001), 477–509.

[8] Kirk, D.S. and Fraser, D.S. The effects of remote gesturing on distance instruction. *Proceedings of th 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!*, International Society of the Learning Sciences (2005), 301–310.

[9] Lucero, A., Aliakseyeu, D., Overbeeke, K., and Martens, J.-B. An interactive support tool to convey the intended message in asynchronous presentations. *Proceedings of the International Conference on Advances in Computer Enterntainment Technology - ACE '09*, (2009), 11.

[10] Passig, D. and Levin, H. Gender preferences for multimedia interfaces. *Journal of Computer Assisted Learning 16*, 1 (2001), 64–71.

[11] Tan, K., Gelb, D., Samadani, R., Robinson, I., Culbertson, B., and Apostolopoulos, J. Gaze Awareness and Interaction Support in Presentations. *ACM MUltimedia*, (2010), 643–646.