Gaze Interaction in the Post-WIMP World
CHI 2012 Special Interest Group Meeting

Andreas Bulling
Computer Laboratory
University of Cambidge
Cambridge, UK
andreas.bulling@acm.org

Robert J. K. Jacob
Computer Science
Tufts University
Medford, MA 02155 USA
jacob@cs.tufts.edu

Raimund Dachselt
User Interface and Software Engineering Group
University of Magdeburg
Magdeburg, Germany
dachselt@acm.org

Sophie Stellmach
User Interface and Software Engineering Group
University of Magdeburg
Magdeburg, Germany
stellmach@acm.org

Andrew T. Duchowski
School of Computing
Clemson University
Clemson, SC 29634 USA
duchowski@acm.org

Veronica Sundstedt
School of Computing
Blekinge Inst. of Technology
371 39 Karlskrona, Sweden
veronica.sundstedt@bth.se

CHI'12, May 5–10, 2012, Austin, Texas, USA.
ACM 978-1-4503-1016-1/12/05.

Abstract
With continuous progression away from desktop to post-WIMP applications, including multi-touch, gestural, or tangible interaction, there is high potential for eye gaze as a more natural human-computer interface in numerous contexts. Examples include attention-aware adaptations or the combination of gaze and hand gestures for interaction with distant displays.

This SIG meeting provides a discussion venue for researchers and practitioners interested in gaze interaction in the post-WIMP era. We wish to draw attention to this emerging field and eventually formulate fundamental research questions. We will discuss the potential of gaze interaction for diverse application areas, interaction tasks, and multimodal user interface combinations. Our aims are to promote this research field, foster a larger research community, and establish the basis for a workshop at CHI 2013.

Keywords
Eye tracking; gaze; multimodal; interaction; post-WIMP

ACM Classification Keywords
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces–input devices and strategies.

General Terms
Design, Human Factors

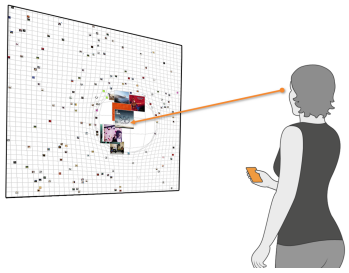

Figure 1: Gaze-supported exploration of a large image collection on a distant display with a handheld smartphone [10].

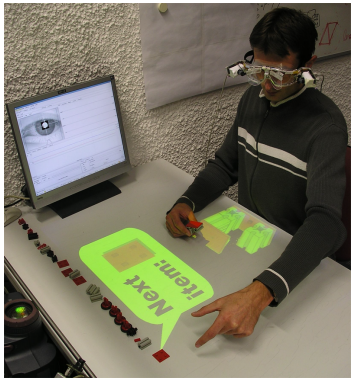

Figure 2: Gaze-based interaction in physical interaction contexts [1].

Introduction and Motivation

Using our eyes for interacting with digital systems is particularly appealing, because gaze provides a natural and fast input channel. Pointing tasks, for example, can be significantly accelerated by using gaze input (e.g., [14]). Moreover, eye gaze is an important cue from which to infer a user's intentions and current mental state, i.e., interaction context. However, until now research on gaze-based interaction has mainly focused on traditional desktop settings often trying to substitute mouse input, e.g., for selecting menu items. New approaches for integrated gaze input for multimodal interaction and for the application in diverse contexts are continually introduced (e.g., [1, 9, 10, 15, 16]).

The development of more natural interaction techniques integrating gaze data with other post-WIMP (*Windows, Icons, Menus, Pointer*) modalities [5] such as multi-touch, hand gestures, speech, or tangibles is a highly promising avenue of research. Since gaze has not been considered much in this context, we advocate a higher emphasis on the integration of gaze into the post-WIMP world. To better motivate this, below we describe some examples of gaze input based on Jacob et al.'s [5] four themes drawn from their reality-based interaction framework.

Naïve Physics is the common sense knowledge about the physical world. Considering the inherent change of view when looking around in the real world, the view at a virtual scene could adapt analogously depending on a user's gaze (e.g., [7, 8]). This can also benefit an automatic gaze-directed panning for the exploration of large information spaces in various application contexts (e.g., [10] and Figure 1). In addition, gaze vergence may drive stereoscopic display disparity [3].

Body Awareness and Skills is a person's awareness of his own physical body and how to coordinate it, including one's own gaze. Considering this, gaze input may play an important role in the context of hand and eye coordinated tasks, such as performing throw gestures with a handheld device. In addition, awareness of one's own visual attention can be used to directly retrieve or place information from or at a user's point-of-regard (e.g., [9, 10]). An example is transferring previously selected information from a smartphone directly to a currently fixated location on a distant display. In return, targets on a distant display could be flexibly selected using a combination of gaze input and a handheld for confirmation (e.g., [9]).

Environment Awareness and Skills is a person's sense of their surroundings and how to interact within their environment. From the system's perspective, the user's fixation of a particular item facilitates its selection [4]. Furthermore, gaze can be used to infer intention in manual picking tasks (e.g., [1] and Figure 2), or other forms of physical interaction with attentive objects (e.g., [12]). Beyond traditional WIMP items, gaze can also be used for entry of graphical passwords by looking at image features (e.g., [2] and Figure 3).

Social Awareness and Skills is a person's awareness of others in his environment and how to interact with them. In this respect, gaze helps to disambiguate who is looking at whom and when (e.g., [11]). Additionally, expert systems can respond to novice users' gaze when learning search strategies (e.g., [13] and Figure 4), and virtual agents could better assess what a user is currently doing by considering their gaze (e.g., [6]).

SIG Objectives

The intent of this SIG is to raise interest in this emerging topic and bring together researchers and practitioners who are interested in the exploration and development of new applications and services using a combination of gaze and post-WIMP interfaces. We are interested in exploring the following issues:

Figure 3: Gaze-based graphical password entry [2].

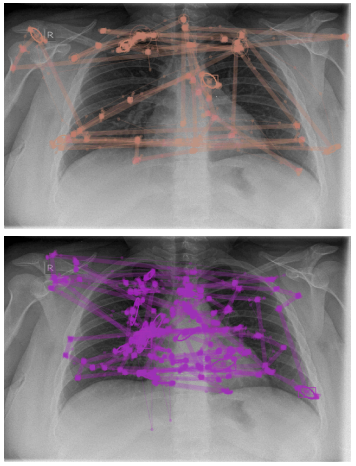

Figure 4: Expert (above) and novice (below) visual search of Chest X-Rays (CXR); an expert system can act as a learning aid [13].

- **What examples for gaze-supported interaction do already exist?** What went well and what went wrong? What are the lessons learned?
- **How can post-WIMP interaction benefit from eye gaze?** What can a user's visual attention indicate? How could particular user groups, tasks, and contexts benefit? What new application areas for gaze-supported interaction can emerge? How could gaze input facilitate interaction with diverse output devices? How could a seamless integration of gaze data with other modalities look like?
- **What needs to be considered for the design of gaze-supported interaction?** What are general design guidelines, methods, and models that have to be taken into account for gaze-supported interaction? What are particular challenges and how could they be tackled?

The overall goal of this SIG is to lay the foundation for a community of researchers and practitioners interested in exploring novel ways to incorporate gaze input beyond the traditional WIMP setup. Our SIG offers a forum for exchanging information and experiences, bringing forward novel ideas, and forming collaborations on specific projects. Based on the obtained feedback, a practical objective of this SIG meeting is to develop a workshop outline for CHI 2013.

Organization

In preparation for this SIG meeting, we will establish a website for this community. It will offer an overview of the discussed topics at this SIG meeting and will also serve as a continuing online venue for further exchanging ideas including updated news about relevant topics and a forum in which further discussion is encouraged.

At CHI 2012, we envision a 4-part schedule for our 80-minute meeting with the main emphasis on a diversified discussion:

1. **Introduction**: Briefly explain what the SIG is about, establish the post-WIMP context within which gaze-based interaction will be examined.
2. **State-of-the-art**: Give a brief review of current research in the area, highlighting gaze-supported interaction.
3. **Brainstorming for the future**: Hold a discussion fleshing out potential future gaze-based interaction scenarios. Here we will focus on a breadth of applications instead of a depth-based search for the killer app. In this context we will explore four key aspects:
   (a) *User contexts*, e.g., public displays at airports, multi-display settings in meetings, multi-touch tables in museums, etc.;
   (b) *User groups*, e.g., machine operators, experts vs. novices, accessibility issues, etc.;
   (c) *Interaction tasks*, e.g., target acquisition, visual search, etc.;
   (d) *Multimodal combinations*, e.g., gaze and hand gestures, speech, handhelds.
4. **Structure for CHI 2013 workshop**: To sum up, we present what we had in mind for CHI 2013, gather feedback from attendees regarding what should be presented and discuss further activities.

Audience

The SIG meeting invites PhD students, researchers, as well as practitioners and industrial partners alike. We aim at creating a strong interdisciplinary research community linking various interest groups including (but not limited to) interaction design, computer science, cognitive science, psychology, computer supported cooperative work, and eye tracking.

We do not assume a particular attendee background but rather hope for high open-minded interest in exploring novel

ways to incorporate gaze input with post-WIMP user interfaces. At the beginning of our SIG meeting, we will provide a brief introduction to gaze-based interaction and post-WIMP interfaces, so that those who have not worked with either so far should have a sufficient overview for a fruitful discussion on this topic.

References

[1] Bardins, S., Poitschke, T., and Kohlbecher, S. Gaze-based interaction in various environments. In *Proc. VNBA '08* (2008), 47–54.

[2] Bulling, A., Alt, F., and Schmidt, A. Increasing the security of gaze-based cue-recall graphical passwords using saliency masks. In *Proc. CHI '12* (2012).

[3] Duchowski, A., Pelfrey, B., House, D., and Wang, R. Measuring gaze depth with an eye tracker during stereoscopic display. In *Proc. APGV '11* (2011).

[4] Jacob, R. What you look at is what you get: Eye movement-based interaction techniques. In *Proc. CHI '90* (1990).

[5] Jacob, R., Girouard, A., Hirshfield, L., Horn, M., Shaer, O., Solovey, E., and Zigelbaum, J. Reality-based interaction: a framework for post-WIMP interfaces. In *Proc. CHI '08* (2008), 201–210.

[6] Pfeiffer, T., Latoschik, M., and Wachsmuth, I. Evaluation of binocular eye trackers and algorithms for 3D gaze interaction in virtual reality environments. *Journal of VR and Broadcasting 5*, 16 (January 2009).

[7] Smith, J., and Graham, T. Use of eye movements for video game control. In *Proc. ACE'06* (2006), 20.

[8] Stellmach, S., and Dachselt, R. Designing gaze-based user interfaces for steering in virtual environments. In *Proc. ETRA '12* (2012).

[9] Stellmach, S., and Dachselt, R. Look & touch: Gaze-supported target acquisition. In *Proc. CHI '12* (2012).

[10] Stellmach, S., Stober, S., Nürnberger, A., and Dachselt, R. Designing gaze-supported multimodal interactions for the exploration of large image collections. In *Proc. NGCA '11* (2011), 1–8.

[11] Vertegaal, R. The GAZE groupware system: mediating joint attention in multiparty communication and collaboration. In *Proc. CHI '99* (1999), 294–301.

[12] Vertegaal, R., Shell, J., Chen, D., and Mamuji, A. Designing for augmented attention: Towards a framework for attentive user interfaces. *Computers in Human Behavior 22*, 4 (July 2006), 771–789.

[13] Vitak, S., Ingram, J., Duchowski, A., Ellis, S., and Gramopadhye, A. Gaze-augmented think-aloud as an aid to learning. In *Proc. CHI '12* (2012).

[14] Ware, C., and Mikaelian, H. An evaluation of an eye tracker as a device for computer input. In *Proc. SIGCHI+GI '87* (1987), 183–188.

[15] Wilcox, T., Evans, M., Pearce, C., Pollard, N., and Sundstedt, V. Gaze and voice based game interaction: the revenge of the killer penguins. In *SIGGRAPH Posters* (2008), 81.

[16] Yoo, B., Han, J.-J., Choi, C., Yi, K., Suh, S., Park, D., and Kim, C. 3D user interface combining gaze and hand gestures for large-scale display. In *CHI EA '10* (2010), 3709–3714.