

Envisioning, designing, and implementing the user interface require a comprehensive understanding of interaction technologies. In this forum we scout trends and discuss new technologies with the potential to influence interaction design. — **Albrecht Schmidt, Editor**

What You Look at Is What You Get: Gaze-Based User Interfaces

Rob Jacob, Tufts University, **Sophie Stellmach**, Microsoft

Imagine a world in which you can seamlessly engage with a multifaceted interactive environment that includes both real-world appliances and virtual components. With the development of increasingly powerful computing machinery in various form factors, the diversity of interactive systems is tremendous, ranging from body-worn personal devices, networked ubiquitous appliances in smart homes, and public wall-filling displays to personal virtual- and augmented-reality head-up systems. Given the speed of technological advancement, the critical bottleneck is not so much in providing more powerful machinery but rather in creating appealing and intuitive ways for users to manage and interact with a vast amount of information. At the same time, an important goal of human-computer interaction research is to enable a higher communication bandwidth between the user and the machine. It is therefore critical to find suitable and applicable ways for orchestrating diverse input channels by carefully leveraging their unique characteristics. Eye movements and, more specifically, information about what a person is looking at, provide great opportunities for more engaging, seemingly magical user experiences. However, they also entail several design challenges, which if not considered carefully quickly result in overwhelming, aggravating experiences. In this article, we share some of our experiences and visions about using gaze as an input method

through which users and computers can communicate information.

A user interface based on eye movements provides several potential benefits. Two of the most commonly named ones are pointing-based interactions that are *faster* and *more effortless* than other interfaces, because we can move our eyes extremely fast and with little conscious effort. A simple thought experiment suggests the speed advantage: Before you operate any mechanical pointing device, you usually look at the destination to which you wish to move. Thus, your gaze implicitly indicates your intention before you're able to actuate an input device. In addition, since you naturally look at content that interests you, gaze input provides an implicit contextual cue about your current visual attention. Eye movements not only provide an interesting *complementary* input, but also allow for *fluent* interaction across diverse user contexts. For example, you could select a target simply by looking at it and confirming the

selection via a speech command or manual input when stepping in front of a large information display, lying on your couch and interacting with your TV, or engaging with content on your head-up display. Gaze can therefore be used as a universal pointing input. Finally, eye movements also enable more *attentive* systems tailored to the user's current focus and activity (see, for example, [1]).

A BRIEF HISTORY OF EYE-MOVEMENT-BASED INTERACTION

In the early 1980s, Richard Bolt [2] described a system in which a user could simply look at graphical items on a large screen and interact with them using gaze information as a subtle contextual cue. Unfortunately, at that time devices that could deliver real-time information about gaze positions were rare and expensive. When one of us (Jacob) worked on this line of research in the late 1980s, our premise was that by using the somewhat cumbersome and expensive (\$250K USD) eye-tracking equipment of the day (Figure 1), we could look into a future when tracking equipment would be cheap and ubiquitous. We could design, prototype, and evaluate interaction techniques that might work well with such an eye tracker. The equipment we used would serve as a time machine to let us develop and study future interfaces. We would develop new interactions, while the industry would develop better and cheaper eye trackers, and we would meet down the road in a few years with effective ways to use the new

Insights

- Using eye movements for human-computer interaction can be a rapid, effective, implicit way to communicate.
- Eye-based interfaces have a long history in research, but new developments are making them more practical and widespread.
- Since our eyes often move without our conscious intent, the design of eye-based interaction is particularly delicate.



equipment when it appeared. As it happened, it has taken far longer than expected for eye tracking to become more widely available for consumer use. But it is finally beginning to happen. For example, higher-resolution front-facing cellphone cameras can do a reasonable job of eye tracking. Wearables such as Google Glass can also provide a platform for eye tracking. Once a laboratory curiosity, eye-movement-based interaction is poised to become an everyday reality. However, while technological advancements in real-time eye tracking are important, this is only half of the equation. The other half is to design interaction techniques that incorporate eye movements in the user-computer dialogue in a convenient and natural way.

DESIGN CONSIDERATIONS FOR EYE-MOVEMENT-BASED INTERACTION

The design of fast and effortless eye-based input control is delicate for various reasons. While at first it sounds appealing to magically let the system respond to a user's eye movements, anticipating his or her next move, this can quickly become tedious and distracting. In the following discussion we summarize

Gaze input provides an implicit contextual cue about your current visual attention.

some of the most common challenges we faced in our research when designing eye-movement-based interactions—in particular, for applications that take into account the precise location of visual interest (usually as estimated gaze positions on a screen).

• **Midas Touch.** Everywhere you look, something is activated; you cannot look anywhere without issuing a command. This describes the “Midas Touch” problem that was coined by one of the authors in the early 1990s [3]. After all, our eyes are “always on” and always moving, and thus, careful consideration has to be given to ways to enable a quick, effortless, and smooth transition between engaging and disengaging from the interaction.

LEARNING MORE ABOUT EYE TRACKING

If you were inspired by the article and would like to learn more about the topic we recommend the following further reading:

- Jacob, R.J.K. and Karn, K.S. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises (Section Commentary). In *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*. J. Hyona, R. Radach, and H. Deubel, eds. Elsevier Science, Amsterdam, 2003, 573-605; <http://www.cs.tufts.edu/~jacob/papers/ecem.pdf>
- Holmqvist, K., Nystrom, M., Andersson, R., Dewhurst, R., Jarodzka, H., van de Weijer, J. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford Univ. Press, 2011.
- Duchowski, A.T. *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag, London, 2007.

NEW FRONTIERS IN EYE INTERACTION

Over the past years researchers have pushed the boundaries of what you can do with gaze tracking into many new directions. Here are some examples:

- *Smooth Pursuit: Gaze Interaction Beyond Saccades and Fixations* See Vidal, M., Bulling, A., and Gellersen, H. Pursuits: Spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. *Proc. of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, New York, 2013, 439-448.
- *Implicit Use of Gaze: Cropping Images by Gaze* See Santella, A., Agrawala, A., DeCarlo, D., Salesin, D., and Cohen, M. Gaze-based interaction for semi-automatic photo cropping. *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, 2006, 771-780.
- *Eye Gestures: Performing Gestures with the Eyes* See Drewes, H. and Schmidt, A. Interacting with the computer using gaze gestures. *Proc. of the IFIP Conference on Human-Computer Interaction*. Springer, Berlin; Heidelberg, 2007, 475-488.

• *Unconscious eye movements.*

People do not normally move their eyes in the same slow and deliberate way in which they operate conventional computer input devices. Eyes continually dart from point to point, in rapid and sudden saccades. Even when a user thinks he or she is viewing a single object, the eyes do not remain still for long. It would therefore be inappropriate to simply substitute the mouse cursor (or some other explicit pointing modality) with the gaze signal. Wherever possible, it is more desirable to attempt to obtain information from the natural movements of the user's eye, or use the gaze signal as a supporting modality augmenting other more explicit inputs, than to require the user to make specific trained eye movements, such as gaze gestures, to actuate the system.

• *Inaccurate targeting.* Imprecise gaze data is mainly due to physiological and technological constraints. While tracking algorithms are continuously improving to provide more precise gaze estimations, the simple fact remains that our visual field combines high- and low-resolution perceptions (foveal and peripheral vision). Thus, to assume that you could perform the same precise and controlled movements that you could with a mouse is not practical. In addition, eye-tracking devices usually have a certain error margin for precisely tracking eye movements (current state-of-the-art systems range less than one degree in visual angle). This increases, for example, the difficulty in selecting and manipulating very small or closely positioned targets with your eyes [4].

• *Eye behaviors.* There are several additional physiological aspects that are important when considering the applicability of gaze as an input for a particular application. For example, our eyes are not able to perform controlled smooth curves if not guided by a visual stimulus. This means that while we are able to smoothly follow a moving target (e.g., a flying bird or a jumping ball) using smooth-pursuit eye movements, we cannot do the same deliberate type of movements on a blank canvas. This makes gaze less applicable for

interactions that require some smooth input control, such as drawing.

• *Synchronization of multimodal inputs.* Since we move our eyes so rapidly, and usually without even thinking about it, it sometimes seems as if we are led by our eyes rather than the other way around. Several studies in which gaze is conveniently combined with other more explicit inputs (e.g., button presses or touch input [5] as illustrated in Figure 2) report a "leave before click" issue. That is, once a user gets more and more acquainted with a system, he or she would often look toward the next target before finalizing the manual trigger action, resulting in false selections.

• *Double-role.* The main purpose of our eyes is to observe our environment. Thus, our gaze assumes a double role for visual observation and control if used as input. While this is not a problem for target selections, as we would normally look at a target we wish to select, it becomes an issue if we wanted to deliberately control continuous input parameters, such as moving a slider to a particular location or rotating and repositioning a graphical object [5].

• *Reliability of tracking.* The most common technique for eye tracking is to use cameras directed at the user's eyes to detect reflections of an infrared light source. This has several challenges, such as unreliable pupil detection due to changing lighting conditions and physiological adaptations (e.g., pupil-size changes).

• *Unfamiliarity.* People are not accustomed to operating devices simply by moving their eyes. However, the careful integration of implicit gaze input with more explicit inputs mitigates this problem.

By addressing these design considerations, you can create eye-movement-based user interfaces that feel fast, effortless—and yes, even magical. We advocate thinking of eye position more as a piece of information available to a user-computer dialogue involving a variety of input devices than as the intentional actuator of the principal input device. With this approach, we have the goal of enabling a fluent integration of diverse input modalities by effectively leveraging

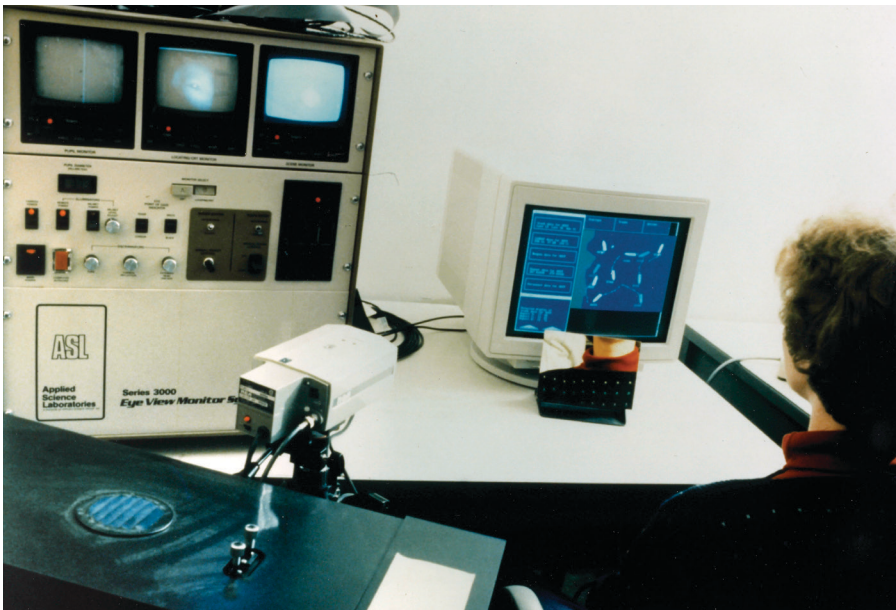


Figure 1. State-of-the-art eye-tracking equipment in the 1980s: the Eye View Monitor from Applied Science Laboratories (ASL).



Figure 2. Concept illustration for interacting with out-of-reach content via fast and effortless gaze-based targeting in combination with more explicit and controlled touch input [4,5].

their unique input characteristics, including gaze, touch, and speech, as well as hand and body postures and gestures. We strongly believe that gaze is most powerful in combination with other more explicit input modalities to address the challenges described here.

Aside from using gaze as a direct pointing modality, several researchers have considered how to use gaze in a more subtle and covert way. A classic example is foveated rendering, in which only the region that is currently looked at would be rendered in high resolution, while peripheral regions that are naturally perceived in lower resolution can be rendered at a lower quality. Ideally the result is that system performance

can be drastically increased while the user does not notice any differences in the interface. Other approaches to incorporating gaze input include analyzing gaze patterns to recognize a user's current activity or even mental state. Finally, the consideration of gaze can also help to create more immersive game experiences in which characters can react more naturally to your gaze.

LOOK INTO THE FUTURE

Coming back to our initial motivation, eye movements are an appealing and powerful input medium for various forms of innovative computing machinery. The applications are endless: foveated rendering in virtual reality environments, shared

attention between people and robots or virtual agents for better mutual understanding of conversational contexts, seamless target selections across multiple distributed displays, personal visual reminders based on what you look at, the seamless interaction with augmented and mixed reality content placed in the real world. It doesn't stop here, though. Eye movements can be well integrated with brain-computer interfaces, speech, and hand and body motions, as well as more explicit input from a personal smart device. The integration of these and potentially many more input channels is challenging, as it may quickly become overwhelming to the user; however, it also offers the tremendous possibility for rich and engaging interactions by providing higher communication bandwidth and a closer connection between users and their machines.

ENDNOTES

1. Vertegaal, R. Attentive user interfaces. *Communications of the ACM* 46, 3 (2003), 30–33.
2. Bolt, R.A. Gaze-orchestrated dynamic windows. *Computer Graphics* 15, 3 (Aug. 1981), 109–119.
3. Jacob, R.J.K. What you look at is what you get: Eye movement-based interaction techniques. *Proc. of ACM CHI '90*. Addison-Wesley/ACM Press, 1990, 11–18.
4. Stellmach, S. and Dachsel, R. Look & touch: gaze-supported target acquisition. *Proc. of CHI '12*. ACM, New York, 2012, 2981–2990.
5. Stellmach, S. and Dachsel, R. Still looking: Investigating seamless gaze-supported selection, positioning, and manipulation of distant targets. *Proc. of CHI '13*. ACM, New York, 2013, 285–294.

👤 **Robert Jacob** is a professor of computer science at Tufts University, where his research interests are new interaction modes and techniques and user interface software; his current work focuses on implicit brain-computer interfaces. He was elected to the ACM CHI Academy in 2007.
→ jacob@cs.tufts.edu

👤 **Sophie Stellmach**'s Ph.D. was dedicated to the design of fluent and convenient interactions with large-size displays by combining various inputs, such as gaze, touch, and foot input. Her current work at Microsoft involves the design and development of novel interaction techniques for AR, VR, and distributed displays.
→ stellmach@acm.org