



Eye Gaze Tracking: A Breakthrough in Human Computer Interaction

Harsimrat Deo

Department of Computer Science, Mata Gujri College, Fatehgarh-Sahib, Punjab-140407, India

Abstract: Eye-gaze tracking holds the potential of becoming a very useful methodology for human computer interaction (HCI). However, integrating gaze tracking in everyday applications has been difficult for the factors ranging from invasiveness and robustness to availability and pricing. In past decade, gaze-tracking has been widely explored for usability in applications of HCI but it still has not been able to secure a place in general interface interaction in day-to-day life. Today, Eye-gaze tracking systems cost lakhs of Rupees, and thus their application is only limited to high-end niche products generally in research domain. It is important to note however that the bulk of this cost is not due to hardware, as the price of high-quality camera technology has dropped over the last ten years. Rather, the costs are mostly associated with customized software development, sometimes integrated with specialized, although inexpensive, digital processors, to obtain high-speed performance. In spite of many practical issues concerning the use of eye-movement tracking as input device in human computer interaction, it is a method that is increasingly being employed to study usability issues in HCI contexts. The objectives of this paper are to introduce the reader to the basics of eye-movement technology, and also present key aspects of practical guidance to those who might be interested in using eye tracking in HCI research, whether in usability-evaluation studies, or for capturing people's eye movements as an input mechanism to drive system interaction. Various opportunities for eye-movement studies in future HCI research and some of the challenges that need to be overcome to enable effective application of the technique in studying the complexities of advanced interactive-system use are also discussed.

Keywords: Eye gaze tracking, human computer interaction, input devices.

I. INTRODUCTION:

People continuously explore their environment by moving their eyes. They look around quickly and with little conscious effort. With tasks that are well-structured and speeded, research has shown that people look at what they are working on; the eyes do not wander randomly. Both normal and abnormal eye movements have been recorded and studied to understand processes like reading and diagnosing medical conditions (for example, a link between vestibular dysfunction and schizophrenia shows up in smooth pursuit). People naturally gaze at the world in conjunction with other activities such as manipulating objects; eye movements require little conscious effort; and eye gaze contains information about the current task and the wellbeing of

the individual. These facts suggest eye gaze is a good candidate computer input method.

A number of researchers have recognized the utility of using eye gaze for interacting with a graphical interface. Some have also made use of a person's natural ways of looking at the world, as we do. In particular, Bolt suggests that the computer should capture and understand a person's natural modes of expression [Bolt 1982]. The object is to create a comfortable way for decision-makers to deal with large quantities of information. A screen containing many windows covers one wall of an office. The observer sits comfortably in a chair and examines the display. The system organizes the display by using eye gaze as an indication of the user's attention. Windows that receive little attention disappear; those that receive more grow in size and loudness. Gaze as an indication of attention is also used in the self-disclosing system that tells the story of The Little Prince. A picture of a revolving world containing several features such as staircases is shown while the story is told. The order of the narration is determined by which features of the image capture the listener's attention as indicated by where they look.

Eye gaze combined with other modes helps disambiguate user input and enrich output. Questions of how to combine eye data with other input and output are important issues and require appropriate software strategies. Combining eye with speech using the OASIS system allows an operator's verbal commands to be directed to the appropriate receiver, simplifying complex system control. Goldberg and Schryver [1993] investigate whether there are consistent indications of a user's intent to zoom toward an object from characteristics of eye gaze, such as where the user is looking in the window. Their results are mixed but support the idea that information other than point-of-gaze is available from the behavior of the eyes. Zhai, Morimoto, and Ihde [1999] developed an innovative approach that combines eye movements with manual pointing.

Electro-oculography (EOG), limbus tracking and video-based tracking etc are some eye-gaze tracking techniques that have been available for many years, but they all suffer from number of limitations. Some techniques require equipment such as special contact lenses, electrodes, chin rests or other components that

must be physically intrusive to the user. These invasive techniques are uncomfortable for the user and are cumbersome to use. The expensiveness is another major issue. Recently, some video-based techniques for eye-gaze tracking have been developed that has minimized this invasiveness to some extent. Video based techniques capture an image of the eye from a camera either mounted on head gear worn by the user or mounted remotely, process it in frames, and map the pupil movement to the gaze location on display screen. It is clear that to reap the potential benefits of eye tracking in everyday human-computer interfaces, the development of low-cost, easy-to-use and robust eye-tracking systems will be necessary.

II. EYE-TRACKING TECHNIQUE:

Eye tracking is a technique whereby an individual's eye movements are measured so that the researcher knows both where a person is looking at any given time and the sequence in which their eyes are shifting from one location to another. Tracking people's eye movements can help HCI researchers understand visual and display-based information processing and the factors that may impact upon the usability of system interfaces. In this way, eye-movement recordings can provide an objective source of interface-evaluation data that can inform the design of improved interfaces. Eye movements can also be captured and used as control signals to enable people to interact with interfaces directly without the need for mouse or keyboard input, which can be a major advantage for certain populations of users such as disabled individuals.

2.1 How Does an Eye Tracker Work?

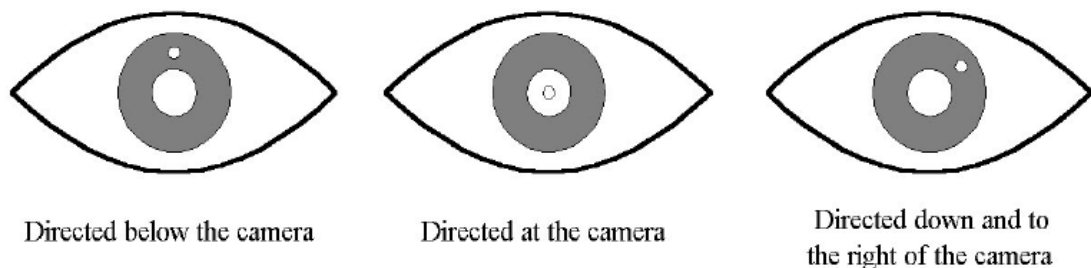


Figure 2. Corneal reflection position changing according to point of regard (Redline & Lankford, 2001).

In general, most eye tracking applications perform the following:

- 1) Connection: establish connection with the eye tracker (e.g., serial port or TCP/IP).
- 2) Calibration: display calibration points at the appropriate location and time.
- 3) Synchronization: display stimulus at the appropriate time (the eye tracker should be able to inform the application program of its state, or vice versa).
- 4) Data streaming: use eye tracker to capture data and/or update the stimulus scene in a gaze-contingent manner.

Most commercial eye-tracking systems available today measure point-of-regard by the “corneal-reflection/pupil-centre” method (Goldberg & Wichansky, 2003). These kinds of trackers usually consist of a standard desktop computer with an infrared camera mounted beneath (or next to) a display monitor, with image processing software to locate and identify the features of the eye used for tracking. In operation, infrared light from an LED embedded in the infrared camera is first directed into the eye to create strong reflections in target eye features to make them easier to track (infrared light is used to avoid dazzling the user with visible light). The light enters the retina and a large proportion of it is reflected back, making the pupil appear as a bright, well defined disc (known as the “bright pupil” effect). The corneal reflection (or first Purkinje image) is also generated by the infrared light, appearing as a small, but sharp, glint (see Figure 1).

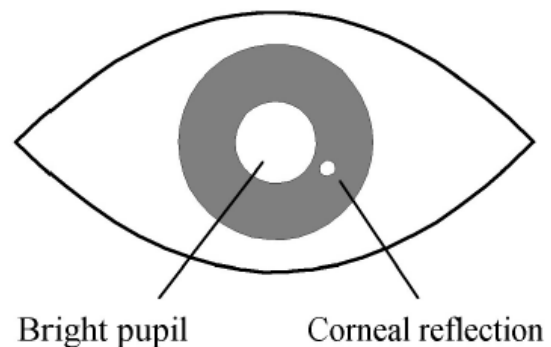


Figure 1. Corneal reflection and bright pupil as seen in the infrared camera image.

An example of a fourth-generation eye tracker is available from [TOBII, 2003]. The TOBII 1750 eye tracker can be configured in several ways, one of which is acting as a server for a (possibly remote) eye tracking client application. The benefit of this organization is platform independence since communication between client and server occurs over TCP/IP. Platform independence is true for an eye tracker communicating via a serial cable as well, although serial communication requires relatively closer proximity between the eye tracker and application computer. An example configuration with an application computer (e.g., Linux

PC) connected to the TOBII eye tracker is shown in Fig. 3.

2.2 Types of eye trackers

Eye trackers can be classified as given below:

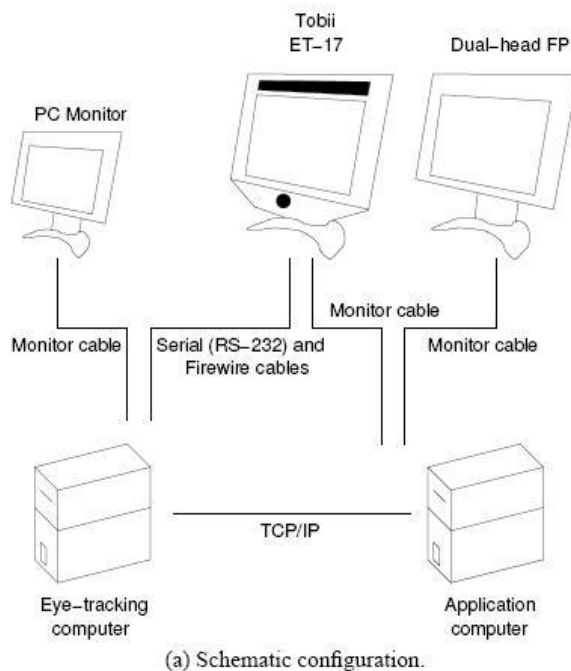
- 1) First generation: eye-in-head measurement of the eye consisting of techniques such as scleral contact lens/search coil, electro-oculography;
- 2) Second generation: photo- and video-oculography;
- 3) Third generation: analog video-based combined pupil/corneal reflection.
- 4) Fourth generation: digital optics coupled with on-chip Digital Signal Processing

The most salient form of eye tracking output is estimation of the projected Point of Regard (POR) of the viewer. First and second generation eye trackers generally did not provide this type of information. So-called video-based combined pupil/corneal reflection eye trackers easily provide POR calculation following calibration. Due to the availability of fast analog video processors, these third-generation eye trackers are capable of delivering the calculated POR in real-time. However, eye tracking technology is about to undergo its next evolution. Fourth-generation eye trackers, which

began to appear on the market, are starting to make use of digital optics. Coupled with on-chip Digital Signal Processors (DSPs), eye tracking technology stands to significantly increase in usability, accuracy, and speed while decreasing in cost. The state of today's technology can best be summarized by a brief functional comparison of equipment available about 5 years ago and today's state-of-the-art given in Table 1.

Table 1. Functional eye tracker comparison

Feature	Legacy systems	State of the art
Technology	Analog video	Digital video
Calibration	5 or 9 point, tracker controlled	Any number, application-controlled
Optics	Requires focusing/thresh holding	Automatic
Communication	Serial	TCP/IP (client/server)
Synchronization	Status word	byte API callback



(b) The ET-1750 flat panel.



Fig. 3. Single Tobii eye tracking station hardware setup (one possible configuration).

III. EYE TRACKER AS USABILITY EVALUATION TOOL

What a person is looking at is assumed to indicate the thought “on top of the stack” of cognitive processes. This “eye-mind” hypothesis means that eye-movement recordings can provide a dynamic trace of where a person’s *attention* is being directed in relation to a visual display. Measuring other aspects of eye movements,

such as fixations (moments when the eyes are relatively stationary, taking in or “encoding” information), can also reveal the *amount* of processing being applied to objects at the point-of-regard. In practice, the process of inferring useful information from eye-movement recordings involves the HCI researcher defining “areas of interest” over certain parts of a display or interface under evaluation, and analysing the eye movements which fall within such areas. In this way, the visibility,

meaningfulness and placement of specific interface elements can be objectively evaluated and the resulting findings can be used to improve the design of the interface (Goldberg & Kotval, 1999). For example, in a task scenario where participants are asked to search for an icon, longer-than-expected gaze on the icon before eventual selection would indicate that it lacks meaningfulness, and probably needs to be redesigned. A detailed description of eye-tracking metrics and their interpretation is provided in the following sections.

Fixations: Fixations can be interpreted quite differently depending on the context. In an encoding task (e.g., browsing a web page), higher *fixation frequency* on a particular area can be indicative of greater interest in the target, such as a photograph in a news report, or it can be a sign that the target is complex in some way and more difficult to encode (Jacob & Karn, 2003; Just & Carpenter, 1976). However, these interpretations may be reversed in a search task: A higher number of single fixations, or clusters of

fixations, are often an index of greater uncertainty in recognising a target item (Jacob & Karn, 2003). The *duration* of a fixation is also linked to the processing-time applied to the object being fixated (Just & Carpenter, 1976). It is widely accepted that external representations associated with long fixations are not as meaningful to the user as those associated with short fixations (Goldberg & Kotval, 1999). Fixation-derived metrics are described in Table 2.

Table 2. Fixation-derived metrics and how they can be interpreted in the context of interface design and usability evaluation.

Eye-Movement Metric	What it Measures
Number of fixations	More overall fixations indicate less efficient search (perhaps due to sub-optimal layout of the interface).
Fixations per area of interest	More fixations on a particular area indicate that it is more noticeable, or more important, to the viewer than other areas.
Fixations per area of interest and adjusted for text length	If areas of interest are comprised of text only, the mean number of fixations per area of interest should be divided by the mean number of words in the text. This is necessary to separate out: (i) a higher fixation count simply because there are more words to read, from (ii) a higher fixation count because an item is actually harder to recognize.
Fixation duration	A longer fixation duration indicates difficulty in extracting information, or it means that the object is more engaging in some

	way.
Gaze (also referred to as “dwell, fixation cluster” and “fixation cycle”)	Gaze is usually the sum of all fixation durations within a prescribed area. It is best used to compare attention distributed between targets. It can also be used as a measure of anticipation in situation awareness if longer gazes fall on an area of interest before a possible event occurring.
Fixation spatial density	Fixations concentrated in a small area indicate focused and efficient searching. Evenly spread fixations reflect widespread and inefficient search.
Repeat fixations	Higher numbers of fixations off-target after the target has been fixated indicate that it lacks meaningfulness or visibility.
Time to first fixation on-target	Faster times to first-fixation on an object or area mean that it has better attention-getting properties.
Percentage of participants fixating an area of interest	If a low proportion of participants is fixating an area that is important to the task, it may need to be highlighted or moved.
On-target (all target fixations)	Fixations on-target divided by total number of fixations. A lower ratio indicates lower search efficiency.

No encoding takes place during saccades, so they cannot tell us anything about the complexity or salience of an object in the interface. However, regressive saccades (i.e., backtracking eye-movements) can act as a measure of processing difficulty during encoding (Rayner & Pollatsek, 1989). Although most regressive saccades (or “regressions”) are very small, only skipping back two or three letters in reading tasks, much larger phrase-length regressions can represent confusion in higher-level processing of the text (Rayner & Pollatsek, 1989). Regressions could equally be used as a measure of recognition value, in that there should be an inverse relationship between the number of regressions and the salience of the phrase. Saccade-derived metrics are described in Table 3.

Table 3. Saccade-derived metrics and how they can be interpreted in the context of interface design and usability evaluation.

Eye-Movement Metric	What it Measures
Number of saccades	More saccades indicate more searching.
Saccade amplitude	Larger saccades indicate more

	meaningful cues, as attention is drawn from a distance.
Regressive saccades (regressions)	Regressions indicate the presence of less meaningful cues.
Saccades revealing marked directional shifts	Any saccade larger than 90 degrees from the saccade that preceded it shows a rapid change in direction. This could mean that the user's goals have changed or the interface layout does not match the user's expectations.

Scanpaths: A scanpath describes a complete saccade-fixate-saccade sequence. In a search task, an optimal scan path is viewed as being a straight line to a desired target, with relatively short fixation duration at the target (Goldberg & Kotval, 1999). Scanpaths can be analysed quantitatively with the derived measures described in Table 4.

Table 4. Scanpath-derived metrics and how they can be interpreted in the context of interface design and usability evaluation.

Eye-Movement Metric	What it Measures
Scanpath duration	A longer-lasting scanpath indicates less efficient scanning.
Scanpath length	A longer scanpath indicates less efficient searching (perhaps due to a sub-optimal layout).
Spatial density	Smaller spatial density indicates more direct search.
Transition matrix	The transition matrix reveals search order in terms of transitions from one area to another. Scanpaths with an identical spatial density and convex hull area can have completely different transition values – one is efficient and direct whilst the other goes back and forth between areas, indicating uncertainty.
Scanpath regularity	Once “cyclic scanning behaviour” is defined, deviation from a “normal” scanpath can indicate search problems due to lack of user training or bad interface layout.
Spatial coverage	Scanpath length plus

calculated with convex hull area	convex hull area define scanning in a localised or larger area.
Scanpath direction	This can determine a participant's search strategy with menus, lists and other interface elements (e.g. top-down vs. bottom-up scanpaths). “Sweep” denotes a scanpath progressing in the same direction.
Saccade/fixation ratio	This compares time spent searching (saccades) to time spent processing (fixating). A higher ratio indicates more processing or less searching.

Blink rate and pupil size: Blink rate and pupil size can be used as an index of cognitive workload. A lower blink rate is assumed to indicate a higher workload, and a higher blink rate may indicate fatigue. Larger pupils may also indicate more cognitive effort. However, pupil size and blink rate can be determined by many other factors, such as ambient light levels, so are open to contamination (Goldberg & Wichansky, 2003). For these reasons, pupil size and blink rate are less often used in eye tracking research.

IV. EYE TRACKING AS INPUT DEVICE

Eye movements can be measured and used to enable an individual actually to interact with an interface. Users could position a cursor by simply looking at where they want it to go, or “click” an icon by gazing at it for a certain amount of time or by blinking. The first obvious application of this capability is for disabled users who cannot make use of their hands to control a mouse or keyboard (Jacob & Karn, 2003). However, intention can often be hard to interpret; many eye movements are involuntary, leading to a

certain “Midas Touch” (see Jacob & Karn, 2003), in that you cannot look at anything without immediately activating some part of the interface. One solution to this problem is to use eye movements in combination with other input devices to make intentions clear. Speech commands can add extra context to users' intentions when eye movements may be vague, and vice versa. Virtual reality environments can also be controlled by the use of eye movements.

4.1 Relevant case studies

Case study 1: (Sibert et. al., 2001) Presented two experiments to demonstrate the advantages of using a person's natural eye gaze as a source of computer input. The circle experiment attempted to measure raw performance, while the letter experiment simulated a real task in which the user first decides which object to select and then finds it. Their experimental results show

that selecting with eye gaze technique is faster than selecting with a mouse. The speed difference with the eye was most evident in the circle experiment. In it, selecting a sequence of targets is so quick and effortless that it almost feels like watching a moving target. The Fitts analysis points out that, within the range they have tested, the farther you need to move, the greater the advantage of the eye because its cost is nearly constant. While the resolution of the eye makes it impractical for positioning tasks that require precision, it is excellent for jumping to distant regions of the screen quickly (where the hand might then be used for detailed work).

Case study 2: (Biswas & Langdon, 2014) Conducted a case study for using eye-gaze tracking based interaction for common computing tasks. They choose to conduct this study in India as a lot of Indian middle aged and elderly users were not proficient with existing user interfaces but keen to use computers and other digital devices if it is easy to use and learn. They found that with a user friendly interface and intelligent eye-gaze tracking system, users can not only outperform the conventional computer mouse but also learn to use this system within 20 minutes. They hoped results from their studies will be useful to develop algorithms and applications for future eye-gaze tracking system.

V. SOME TECHNICAL ISSUES IN EYE GAZE TRACKING

Experimenters looking to conduct their own eye-tracking research should bear in mind the limits of the technology and how these limits impact the data that they will want to collect. For example, they should ensure that if they are interested in analyzing fixations, that the equipment is optimized to detect fixations. The minimum time for a fixation is also highly significant. Interpretations of cognitive processing can vary dramatically according to the time set to detect a fixation in the eye-tracking system. Researchers have to work with limits of accuracy and resolution. A sampling rate of 60hz is good enough for usability studies, but inadequate for reading research, which requires sampling rates of around 500hz or more (Rayner & Pollatsek, 1989). It is also imperative to define areas of interest that are large enough to capture all relevant eye movements.

Eye trackers are quite sensitive instruments and can have difficulty tracking participants who have eye-wear that interrupts the normal path of a reflection, such as hard contact lenses, bifocal and trifocal glasses, and glasses with super-condensed lenses. There may also be problems tracking people with very large pupils or "lazy eye", such that their eyelid obscures part of the pupil and makes it difficult to identify. Once a person is successfully calibrated, the calibration procedure should than be repeated at regular intervals during a test session to maintain an accurate point-of regard measurement.

VI. FUTURE DEVELOPMENTS IN EYE TRACKING

Future developments in eye tracking should centre on standardizing what eye movement metrics are used, how they are referred to, and how they should be interpreted in the context of interface design. Eye-tracking technology also needs to be improved to increase the validity and reliability of the recorded data. The robustness and accuracy of data capture needs to be increased, so that point-of-regard measurement stays accurate without the need for frequent re-calibration. Data collection, filtering and analysis software should be streamlined so that they can work together without user intervention. The intrusiveness of equipment should be decreased to make users feel more comfortable, perhaps through the development of smaller and lighter head-mounted trackers. Finally, eye-tracking systems need to become cheaper in order to make them a viable usability tool for smaller commercial agencies and research labs. Once eye tracking achieves these improvements in technology, methodology, and cost, it can take its place as part of a standard HCI toolkit.

VII. CONCLUSION

Our contention is that eye-movement tracking represents an important, objective technique that can afford useful advantages for the in-depth analysis of interface usability. Eye-tracking studies in HCI are beginning to burgeon, and the technique seems set to become an established addition to the current battery of usability-testing methods employed by commercial and academic HCI researchers. This continued growth in the use of the method in HCI studies looks likely to continue as the technology becomes increasingly more affordable, less invasive, and easier to use. The future seems rich for eye tracking and HCI.

REFERENCES

- [1] Biswas, P., & Langdon, P. (2014). Eye gaze tracking based interaction in India, *Procedia Computer Science* 39 (2014) 59-66.
- [2] Bolt, R. A. 1982. Eyes at the interface. *Proc. ACM CHI'82*, 360-362.
- [3] Chennamma, H. R., & Yaun, X. (2013). A survey on eye-gaze tracking technique, *Indian Journal of Computer Science and Engineering*, Vol. 4, No. 5 pp 388-393.
- [4] Duchowski, A. T. 2003. *Eye Tracking Methodology: Theory & Practice*. Springer-Verlag, Inc., London, UK.
- [5] Goldberg, J. H. and Schryver, J. C. 1993. Eye-gaze control of the computer interface: discrimination of zoom intent. *Proc. Human Factors and Ergonomics Society*, 1370-1374.
- [6] Goldberg, H. J., & Kotval, X. P. (1999). *Computer interface evaluation using eye*

- movements: Methods and constructs. *International Journal of Industrial Ergonomics*, 24, 631-645.
- [7] Goldberg, H. J., & Wichansky, A. M. (2003). Eye tracking in usability evaluation: A practitioner's guide. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 493-516). Amsterdam: Elsevier.
- [8] Jacob, R. J. K., & Karn, K. S. (2003). Eye tracking in Human-Computer Interaction and usability research: Ready to deliver the promises, In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 573-605). Amsterdam: Elsevier.
- [9] Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, 8, 441-480.
- [10] Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.
- [11] Redline, C. D., & Lankford, C. P. (2001). Eye-movement analysis: A new tool for evaluating the design of visually administered instruments (paper and web). In *Proceedings of the Section on Survey Research Methods of the American Statistical Association*.
- [12] Sibert, L. E., Templeman, J. N., Jacob, R. J. K. (2001), *Evaluation and analysis of eye-gaze interaction*, Naval Research Laboratory, Washington DC.
- [13] Starker, I., and Bolt, R. A. 1990. A gaze-responsive self-disclosing display. *Proc. ACM CHI'90*, 3-9.
- [14] TOBII Technology AB. 2003. Tobii ET-17 eye-tracker product description (v1.1). <<http://www.tobii.se/>> (last accessed January 2007).
- [15] Zhai, S., Morimoto, C., and Ihde, S. 1999. Manual and gaze input cascaded (MAGIC) pointing. *Proc. ACM CHI'99*, 246-253.

