Eye Movements, for A Bidirectional Human Interface

Richard Epworth
Technology Strategy Manager, STC Technology Ltd.,
London Road, Harlow, Essex CM17 9NA, UK

Abstract

When we look at a screen, we are unaware of precisely where our eyes are looking, yet our eye movements contain valuable information. An optoelectronic system which monitors our eye movements, may be used to communicate with a machine. Items on a screen may be selected simply by looking at them, and actions initiated by deliberate eye movements. Such a system can also be aware of our interests and difficulties, and may respond to our needs, even those of which we are unaware. This paper describes the techniques of eye movement measurement, possible applications of this information, and several practical demonstrations of an eye-controlled interface.

1 Introduction

Human computer communication is far less intimate than a meeting between two human beings, where eye movements, facial expressions and body language communicate a considerable amount of the information. As a greater percentage of the population becomes involved in the use of computers, it is natural to expect the manner of controlling computers to move away from the programming model and closer to the perceptual process we use to accomplish our goals in the physical world [Krueger et al, 1985]. A more intimate relationship with computers should bring a greater awareness of our real needs. The use of eye movement information is a step towards a closer relationship.

This paper will explain the meaning of eye movements and describe techniques for their measurement. The range of applications of eye movement controlled technology will be presented, and finally details will be given of several practical demonstrations.

2 The Human Eye, its capabilities and limitations

2.1 The Physiology of the Eye

The human visual system is very complex. For the purposes of this paper it is necessary to describe a little of the physiology of the eye. Those who wish greater detail, are recommended to read "Eye and Brain" by Richard Gregory [1966]. The eyeball is approximately spherical, and is rotated within its bearing socket by means of a number of muscles. An optical system images the external view onto the retina which is a dense matrix of light sensitive cells, inside the back of the eyeball (Fig. 1). The dominant imaging component is the bulge in the front surface of the eye, the cornea. The lens within the eye provides some additional focusing which is controllable, enabling the eye to focus on objects at different distances.

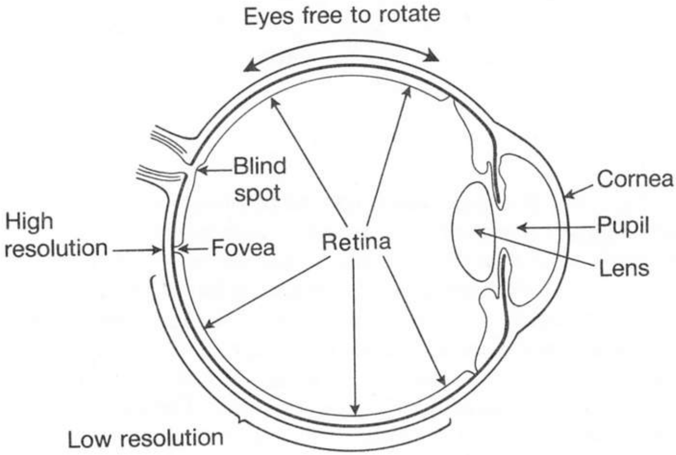

Fig. 1 The human eye. Note, that only that small part of the image which falls on the Fovea, can be seen with high resolution. The rest is "blurred"

Whilst it is attractive to imagine the eye as a sort of camera, this is not an appropriate analogy. Firstly it begs the question "What looks at the picture?". Secondly, the eye is in effect an extension of the surface of the brain, and much signal processing occurs within the back of the retina itself, in groups of cells optimised for specific purposes, such as moving edge detection. The image we see of reality is a software simulation updated by features sensed by our eyes.

2.2 The Resolution (Visual Acuity) of the Eye

The human eye has very high resolution, about one minute of arc, close to the diffraction limit for the size of the optics used. However the eye does not provide high resolution over a wide field of view simultaneously. Central vision is used for the acquisition of fine detail. Only coarse detail is discernible away from the central region (see 4.3.1.1 and [Yager and Davis, 1987]), and peripheral vision is used largely for motion detection. We achieve high resolution perception by sequentially moving the small central high resolution view of the eye, over the scene of interest.

2.3 Fovea Fixations and Point-of-Gaze

The fovea is the small, roughly disc shaped area on the retina, which provides the highest visual acuity. It is located close to the central optic axis (the line through the centre of the eye, lens and cornea). The diameter of the fovea corresponds to about 1 degree visual angle. In other words if a person looks at a disc whose diameter subtends one degree visual angle, the image of the disc that is focused on the retina will cover an area about equal to that of the fovea. The word is sometimes used as a verb, to "foveate" meaning to position the eyes so that the image of a certain target or element of the visual scene falls on the fovea. To "fixate" a point in the visual field, implies foveation of that point. In this paper, when we talk about the "point of fixation" or "point-of-gaze" we are indicating the part of the scene being imaged on the fovea.

2.4 The Blind Spot

The normal eye has a region which has no response to light, known as the "blind spot". This is where all the connections are made to the eye, i.e. where the blood supply enters and leaves, and where the optic nerve leaves the eye, carrying the visual information to the brain. What is intriguing about the blind spot is that we do not normally know it is there, we are blind to its blindness. It is not perceived as a hole in the visual field, but instead appears to be filled with a stimulus similar to whatever surrounds it. Blind spots caused by damage or disease also fill-in in this way. The blind spots are offset horizontally from our centre of vision by about 15 degrees. The blind spots of left and right eyes are displaced to the left and right sides of our centre of vision respectively, so that the image from one eye provides details of the image missing from the other. However, the fact that we are ignorant of our loss of information when using one eye alone suggests that vision (and in fact all our perception) is in effect an ongoing simulation of reality, which is continually updated by our senses. In Reference 2, Gregory writes: "The large brains of mammals, and particularly humans, allow past experience and anticipation of the future to play a part in augmenting sensory information, so that we do not perceive the world merely from the sensory information available at any given time ...". In other words, what we are familiar with, influences what we actually see.

3 Eye Movements

3.1 Point-of-Gaze as an Indicator of Point of Attention

When performing a visual task, the point-of-gaze is a reliable indicator of where our attention lies, especially when the task involves resolving some visual detail. It has been shown that, in the absence of peripheral stimulation, it is not possible to make an eye movement without making a corresponding shift in the focus of attention [Shepherd et al, 1986]. Conversely however, it is possible to shift one's attention without making an eye movement, for example one might start thinking about something else while staring at a word in this piece of text. To summarise, an eye movement usually indicates a shift of attention to that new location.

3.2 Types of Eye Movements

Because of the intimate connection between eye and brain, the study of eye movements is now providing a unique insight into human information processing [Groner et al]. There are two distinct types of eye movements, first a jerky motion jumping from one fixed pointing direction to another, and second a smooth tracking motion [Yarbus, 1967], [Carpenter, 1989]. These may occur independently or simultaneously, and perform very different functions. In addition, eye movements when reading text are highly distinctive.

3.2.1 Fixations and Saccades During normal scanning of a visual scene, eye movement is characterised by a series of stops and very rapid jumps between stopping points. These stops, which normally last at least 100 milliseconds, are called "fixations", and it is during these fixations that most visual information is acquired and processed. The rapid jumps or flicks between fixation points are called "saccades". Saccades are conjugate eye movements (both eyes move together) that can range from 1 to 50 degrees visual angle. They generally have durations from 30 to 120 milliseconds, and achieve angular velocities as high as 600 degrees per second [Carpenter, 1989]. Very little visual information is acquired during saccades, mainly due to blurring caused by the fast motion of the image across the retina, and because the brain partially suppresses information just prior to and during a saccade. When our gaze is attracted to a new spatial location, our gaze jumps towards the new location, but typically undershoots by about 10%, followed by a second or even third corrective saccade (see Fig. 2). Each of these jumps takes time, and delays the moment when any high resolution detail from the scene can be perceived. This highlights the need to reduce the number of events which cause eye movements if the speed of performing a task is important.

This jerky motion of the eye can be felt by placing ones fingers on the closed lid of one eye, whilst looking around with the other. Except for unusual mental states such as unconsciousness, the eyes are rarely completely still for more than a few moments. When we are thinking, we do fixate for quite long periods, but even then the eyes are not completely still. The eye exhibits "micro-saccades", tremor and drift, all of which maintain the image in motion on the retina. This motion is necessary for image perception, for if the image is perfectly stabilised on the retina, it will fade totally within a few seconds [Tulunay-Keesey, 1982]. It is possible to observe this fading, by staring fixedly at a scene which contains only low spatial frequencies.

The reason for this surprising fact can be traced to our evolutionary past. Early creatures used their eyes simply as motion sensors, they lacked the

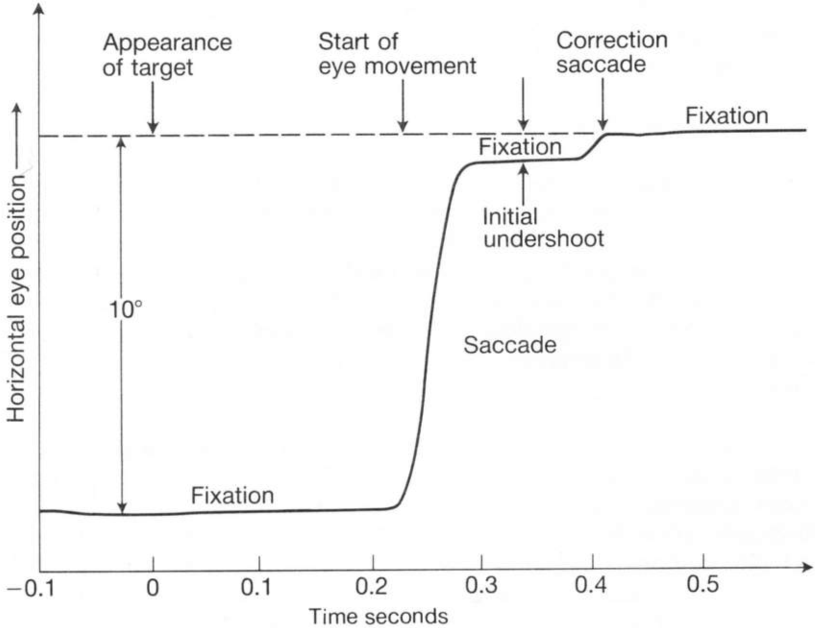

Fig. 2 Plot of the eye pointing direction versus time, when a target suddenly jumps 10 degrees to the right.

This clearly illustrates the abrupt jumps (or Saccades) between Fixations. Note that it is 0.2 of a second before anything happens, and more than 0.4 seconds before the point-of-gaze finally falls on the target. The subject however, thinks he has responded "instantly"

computing hardware to perform any more sophisticated image processing such as object recognition. They kept their eyes still with respect to the surroundings, and the sensitive cells in their eyes rapidly adapted to whatever light fell on them. Cells would then only be triggered if part of the scene were moving, alerting them to pounce or run in the direction determined by the particular cells triggered.

Many higher creatures still stabilise their head and eyes when specifically looking for movement, to suppress all the stationary part of the image (e.g. chickens looking for worms, kestrels looking for mice). We are generally above that sort of thing, and are capable of sophisticated image processing; however we still have light sensitive cells which have no DC response. The solution is to "chop" the signal to get it to fall within the passband of our eyes, and we do this by moving our eyes, well before the image fades. However, we are not normally conscious of this strategy.

3.2.2 Smooth Pursuit The eye can smoothly track targets that are moving in the range of 1 to 30 degrees a second. These conjugate slow tracking eye movements are usually called "smooth pursuit" and their function is partially to stabilise slowly moving features of interest, onto the retina. This is necessary if we are to extract high resolution detail from a moving image. These slow smooth eye movements cannot in general be executed without a slowly moving target. We do however possess another mechanism which compensates for movements of the head. Inertial information from the balance organs in the inner ear is used to provide an equal and opposite rotation of the eyes when the head rotates, to ensure that the eye tracks the image even when the head is not stationary[7]. A consequence of this is that we can also produce slow pursuit eye movements by gazing at a fixed object and turning our head.

3.3 Eye Movements During Reading

Though normal eye movements are not taught, the act of reading is unique, in that it requires a specific eye movement strategy to be trained. Except for very short words, or those anticipated by context, words can only be read by fixating the point-of-gaze upon them [Rayner, 1983, McConkie, 1983, O'Regan, 1987]. When reading at a reasonable speed, the eyes must make a sequence of saccades, most of which are towards the text that is as yet unread. A smaller number of saccades, called "regressions", move the eyes backwards towards a location that was passed earlier in the course of reading. Early readers train their eye movements, usually by following a finger which points at the words in sequence. This provides a target to direct the next saccade. With practice, these directed eye movements become programmed in the brain, and the pointing finger is no longer required.

When reading multi-line text, the eyes make a large saccade to the left and down slightly, to the start of the new line. More often than not, this saccade undershoots, and an additional correction saccade is required to locate the point-of-gaze on the first word of the next line. This is in part a consequence of the poor resolution of the eye off-axis. We cannot accurately predict what size a large saccade should be, because we have no accurate detail about the target until the point-of-gaze gets near to it.

Reading aloud is a slow process limited by the speed of speech, and almost every word is fixated in turn. In contrast, normal reading is for comprehension, not for translation into speech. Reading speed is vitally important and surprising strategies are employed. On average there is one fixation per word, however only 60% of the words are fixated, the rest, often short words like "the" are not fixated at all. This accounts for errors in proof reading short words. In addition, short function words such as "and" and "for", are skipped more often than short content words such as "ate" and "fog" [Hogaboam, 1981].

Long words, of ten or more characters in length, receive on average two fixations per word. Inexperienced readers make more fixations and regressions than fast readers. The fixation durations are about 250 milliseconds on average, and show little reduction with increased reading speed. Fast, and so called "speed" reading are accomplished by fixating fewer words, not by spending significantly shorter time per word. The fact that the reader can "understand" the text, despite acquiring such a small fraction of the image, is a consequence of the high redundancy in language. To be understandable, text must conform to the rules of spatial structure, sensibility, spelling, and context. The rules, known to both writer and reader in advance, reduce uncertainty, making messages partially predictable.

The understanding of reading behaviour is a new science, the study of which is providing valuable insight into the architecture of the brain. There is much debate over the influence which words not being fixated have on the reading process, and to what extent the mental processing of a word is completed prior to the next saccade [Rayner et al, 1987]. There are however measurable factors which can be correlated with reading and comprehension difficulties. Multiple fixations and regressions are frequently a sign of reading difficulty. Multiple fixations and unusually long fixations occur on words which are misspelt, less contextually predictable, or unfamiliar [McConkie, 1983, Rayner et al, 1987].

3.4 Are we aware of our eye movements?

We do not need to be taught how to fixate, make saccades or slow pursuit eye movements, and many who read this paper may previously have been unaware of the very nature of their own eye movements. When we perceive our world, our eye movements secretly assist in acquiring data on the reflectance and spectral properties of key features in our surroundings.

We are also unaware of many of the limitations of our senses, these include the poor resolution away from the centre of vision, the long delay times in responding to visual stimuli, and the blind spot. Presumably, it has been of little evolutionary value to be distracted by knowledge of those limitations which cannot readily be overcome. The processing speed associated with vision, has been at a premium for ensuring our survival. Many of the programs are "hard-wired" to ensure the minimum latency (delay).

Note that in everyday life, we very rarely use eye movements for sending information, apart from the occasional raising of the eyes as a sign of exasperation, or looking away in avoidance of eye contact.

4 Applications of Eye Movement Information

4.1 Eye Movements as a Measurement Tool

4.1.1 Medical Simple eye movement measurement systems, have been used quite widely in the medical field for investigating visual impairments and reading difficulties. Until recently, their use for non medical purposes has been comparatively minor.

4.1.2 Human factors

4.1.2.1 Sports Eye movement analysis has provided a new insight into sports training. For example, it has revealed that in fast ball games, good players do not keep their "eye on the ball", contrary to advice by their coaches. Instead of tracking the ball, they take a few "snapshot" fixations, which enable them to regularly update their predictive model of the ball's trajectory.

4.1.2.2 Usability In order to develop new software and hardware systems which are easy to use, it is important to be able to measure their "usability". In a present day usability evaluation laboratory, video recordings are made of the subject, display, keyboard etc., and keystrokes and mouse movements are recorded, all for subsequent analysis. This provides a record of what the subject is doing physically, but one would really like to know where the subject's attention is located.

During a typical trial, there are significant periods of time when nothing appears to be happening, when there are no body movements or keystrokes. During these times it is vital to know whether the subject is simply daydreaming or whether they are struggling with the system. One would ideally like to know where their attention is focused. The subject can be asked to give a verbal commentary, but speaking interferes with the thinking processes, and is subject to significant delay errors.

There is no record of the point-of-gaze to provide an indication of the point of attention. With appropriately positioned cameras it is possible to judge which quadrant of the screen the subject is looking at, but only if head movements rarely occur. However the video image of the subject's face does not provide sufficient resolution to show what the eyes are looking at.

4.1.2.3 Application to usability measurement Point of gaze measurement systems are increasingly being used for the study and evaluation of human factors [Grat, 1987]. This information may be used at two levels:

i) Simply as an indicator of the location of the subjects attention.
ii) From analysis of the sequence of the fixations and their durations, one may make detailed inferences about the mental processes taking place.

Eye movement measurements can provide a strong indication of which mental processes are taking place. Eye movements during reading are distinctly different from other eye movements, such as searching or staring vacantly (thinking?). Thinking produces long fixations, therefore we can discriminate thinking from reading.

Point-of-gaze measurement provides a technique for following the moment by moment processing of visual information during a task. In tasks where the acquisition of visual information is associated with the spatial arrangement