MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Memo. No. 123.
Vision Memo.

March, 1967.

Computer Tracking of Eye Motions

Seymour Papert

Marvin Minsky.

This memo is to explain why the Artificial Intelligence group of Project MAC is developing methods for on-line tracking of human eye movements. It also gives a brief resume of results to date and the next steps.

1. Reasons for developing real-time tracking systems.

The study of eye motions is important already in perceptual psychology and in human engineering of many kinds of systems. Standard methods are based on recording, in one way or another, and analysing the results later. These methods do not permit experiments making the visual environment conditional on the oculomotor behavior, and are thus limited in the extent of experimental control of the stimulus.

Standard methods of eye-tracking all require either immobilization of the subject's head, by clamps or bite-bars, or wearing special helmets or contact lenses. One of our goals is to allow the subject to operate without any special apparatus. Thus, we have to reject special contact lenses, which otherwise would make remote tracking fairly easy.

Now we list reasons for eye-tracking more specifically.

1.1. Eye-motions are the only observable intra-perceptual processes.

The Artificial Intelligence group is concerned with making perceptual machinery, and naturally wants to know about human visual mechanisms. Unfortunately, too little is known, and this is one instance where it appears that we will have to do much of the work ourselves to find out what we need. Eye-motions report to us the effect of certain decisions in the scene-analysis process; we have no such information in, say, audition. (We do, in tactile perception, but the problems here aren't so interesting, at least yet). Thus, we ought to be able to find out a great deal about the relation between central and peripheral visual field cues, by watching eye motions . While this has often been done to some extent with off-line recordings, the results of such analyses are only interesting, and not very decisive. With on-line information, we can make the visual stimulus---on a display scope, for example--- depend on the sequence of fixations in such ways as to make decisive determinations of which peripheral cues has what effects.

1.2 Clues about internal thought processes.

Allen Newell has suggested a number of ways that eye-motions might be analysed in conjunction with problem-solving "protocols" so as to help trace a subject's execution through his problem-solving flow diagram. Although no really conclusive techniques for this are established, the possibilities seem important, for the reasons discussed in 1.4 below.

1.3 Use as control system input.

It seems likely that one could use eye-motion for control of complex systems, in some very comfortable and precise ways. The program has to infer the subjects intentions from his fixations on a possible very complex control console, and we can expect serious problems in classifying such intentions and recognizing them. Tracking, in particular, should be particularly natural. It should be noted that there is no reason to expect visual tracking control to be comparable to manual control in precision; one must expect to have to need Sketchpad-like inferences about the proximity of the fixation from a control point, and from the local behavior of the fixation.

1.4 Hypothesis about smooth integration with memory.

Another reason for supposing that on-line eye-tracking will be important is based on the following introspective notion. Normally, one is unconcious of details about one's memory of visual scenes, especially when the scene is present. For then, one can either remember what is in a portion of the visual environment or one can look there. It is possible that these alternatives are integrated into scene analysis so smoothly that instructions to move the fixation point to a point have practically the same status as requests to memory for information about what is (supposed to) be there. If something like that holds, then we may be able to make dramatically effective and theoretically interesting information retrieval systems according to the following schema: Imagine a display with only some very general categories of information on it. Assume that the user is quite familiar with this coarse system. When he fixates on a region, this means that he is interested in it, and that part of the display is replaced by an expansion in greater detail. The subject feels that he knew this material already, in his mind's eye, and if he is really familiar with the second-level he will find the next sub-category in the expanded portion quickly, again without conscious effort. The effect, I hope, would be that subjectively the system will feel as though an immense amount of detail is available, in familiar forms, in the immediate visual environment, although this is not really true, of course. We certainly ought to be able to get simple first-order effects such as television pictures that appear to have full bandwidth, though only a small output data rate is really operating. We cannot expect to make the system provide new information in forms that make the subject feel that he already "knows" it thoroughly, for we can't expect to get the effect of full networks of relations, but one ought to get some such effects and it seems worth finding how far it goes.

2. Simple immediate applications.

The following experiments could be immensely important in education, if they yield good results.

2.1 Reading without learning to read.

Suppose that a pre-literate child, adult or blind person (with oculomotor control) could look at a page, and have the word or phrase in fixation read to him. The eye-tracking program would have to be developed to a point of some refinement, of course, to approximate the subject's attention. Then,

(i) One would expect some learning.

(ii) One would in fact be able to "read". For the child, this is valuable enough so that he would want to keep at it. The control experience in reading oneself is different from the experience of being read to. Indeed, it is a special skill in itself, so that

(iii) the control system for reading can be developed independent of the character-and word-reading learning. One need not degrade the child back to the "primer" vocabulary.

2.2 <u>Language learning</u>.

The problem of learning to read a foreign lanuage is complicated by the need simultaneously to acquire several skills---vocabulary, inflection, grammatical construction and word-order, pronunciation, and even new character sets. Several of these, particularly vocabulary knowledge, can be minimized by having the word under fixation persists longer than threshold. (Or instead of threshold, some other property of eye-motion, or any other subject-output could be used.) Instead of pronunciation, one might replace the word by its English equivalent, or if necessary the phrase, or one might apply some other transformations. One could guide the reader by programmed cues, like brightening, moving, enlarging, deleting, etc., sections or phrases ahead or behind, to encourage desired scanning behavior or other reading skills.

2.3 <u>Blind</u>.

Many blind persons have good oculomotor control, or could acquire it, given appropriate feedback. Although a blind person can train himself to understand text read at higher than normal rates, this is next-to-useless for intricate things like mathematical reading, where one doesn't normally scan the page in a "regular" manner. The same is really true for any difficult reading, and it is possible that with oculomotor control over a page, blind people could absorb the material, read-alound fast but locally, at very respectable rates. It would be fun to see if anything could be done in the line of making aural picture images, with special sounds.