Vision Res. Vol. 23, No. 8, pp. 765–768, 1983
Printed in Great Britain. All rights reserved

0042-6989/83 $3.00 + 0.00

INTEGRATING VISUAL INFORMATION FROM SUCCESSIVE FIXATIONS: DOES TRANS-SACCADIC FUSION EXIST?

J. K. O'Regan and A. Lévy-Schoen
Laboratoire de Psychologie Experimentale, 28 rue Serpente, 75006 Paris, France
(Univ. Paris V, EPHE, CNRS)

(*Received* 28 *June* 1982; *in revised form* 19 *November* 1982)

Abstract—It has recently been suggested that there exists a level in the nervous system where a kind of photographic representation of our visual environment is constructed from "snapshots" taken by successive eye fixations. An experiment is presented that argues against this view, and an alternative explanation is put forward to explain why we see the environment as being stable and continuous despite eye movements.

Recently Jonides *et al.* (1982) presented evidence that when the eye explores a visual scene, the consecutive "snapshots" taken during successive eye fixations can be superimposed at an early, iconic, level of representation in the nervous system to form a coherent view of the environment. This combination of successive snapshots takes place, they suggest, despite the fact that the eye moves between the successive snapshots. The representation is therefore "trans-saccadic", that is, in spatial coordinates, not in retinal coordinates. Jonides *et al.*'s paper prompts us to publish in greater detail some contradictory data which we had previously abstracted in a French journal (Lévy-Schoen and O'Regan, 1979).

The experimental paradigm we used was very similar to Jonides *et al.*'s. The principle involved is to present one half of a visual stimulus before an eye movement, and another half after the eye movement. The stimulus halves are displayed in the same physical location in space, but because of the eye movement, they fall on different retinal locations. The critical question is: can the visual system fuse the two halves together into a coherent whole corresponding to their true physical proximity, despite the fact that they fell on different retinal locations?

As shown in Fig. 1(A) each of our stimulus halves consisted of a set of apparently random line segments. When two halves were superimposed, they formed one of three possible three-letter words which the subject had to name. However, each stimulus half individually was not sufficient to name the word. The subject sat 50 cm from a large CRT tube with fast decaying P15 phosphor, fixating a central fixation point. Figure 1(B) shows the sequence of events for each trial. A target point appearing either on the left or the right of the screen at an eccentricity of 8.2° was the signal for the subject to move his eyes to this target point. At a random moment before, during, or after the eye movement, the computer displayed the first stimulus half for 1 msec at a position midway between fixation point and target. 50 msec later, the second stimulus half was displayed for 1 msec in the same place on the screen. The stimuli subtended 2.9° horizontally and 1.7° vertically. By pressing a button the subject indicated which of the three stimulus words he thought had been displayed. A button was also provided for the response, "Don't know". Eye movements were measured using the photoelectric scleral reflection technique, and recorded by the computer at a sampling interval of 3 msec.

Probability of correct response for four subjects is plotted in Fig. 2. For each trial the computer calculated the interval between saccade onset and the start of the stimulus sequence. By convention, time will be measured from the moment of saccade onset to the moment of occurrence of the second stimulus half. A time of −30 msec for example corresponds to the second stimulus occurring 30 msec before saccade onset, and the first stimulus occurring 50 msec earlier, that is 80 msec before saccade onset. For such a trial, both stimulus halves occur before the eye had started moving, and so they impinge on the same retinal location, and accurate responses can be expected. Only for a certain critical range of times do the two stimulus halves not impinge on the same retinal location: this range goes from the time at which the second stimulus coincides with the saccade onset ($t = 0$), to the time the first stimulus coincides with the saccade end, i.e. at t = saccade duration + interstimulus interval. Since saccade durations were of the order of 30 msec, we have taken $t = 30 + 50 = 80$ msec for this critical time. Within this region the hatched region is the region of particular interest to the issue of trans-saccadic fusion, since it corresponds to trials where one stimulus occurred before and one after the saccade. If two stimulus halves occurring in the same physical location, but one before and one after the saccade, can be fused despite the fact that they im-

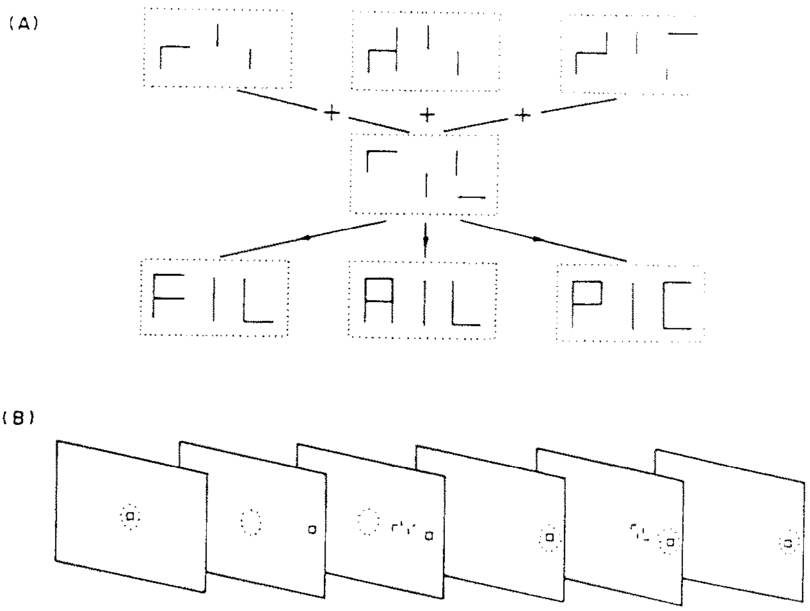

Fig. 1. (a) There were three possible stimuli, each consisting of two halves. The half that occurred second was always the same. The two halves, when superimposed, formed a three letter word. The word subtended 2.9° horizontally and 1.7° vertically. (b) The sequence of events for an individual trial. The dotted circle shows the behavior of the eye in a critical trial where one stimulus half appears before the saccade, and one after. The eye is fixating a fixation point in the center of the screen. A target point appears 8.2° to the right or to the left. During the eye's latency period the first stimulus half appears for 1 msec midway between initial fixation point and fixation target. The eye makes a saccade and arrives at the fixation target. The second stimulus half appears for 1 msec. Its moment of occurrence is always 50 msec after the first stimulus half.

pinge on different retinal locations, then we expect accurate responses in this hatched region.

Looking at the curves in Fig. 2 it is evident that trans-saccadic fusion does not occur: in the critical hatched regions, subjects' responses were inaccurate. In fact they were inaccurate precisely in the region where at least one of the two stimulus halves occurred during the saccade. This is consistent with the hypothesis that the subjects could identify the stimuli only when the two halves impinged on the same retinal location.

The results are in contradiction with those of Jonides *et al.* Several interesting differences between our experiment and Jonides *et al.*'s could be at the root of this difference. The most striking of these is the fact that in our experiment the two retinal locations stimulated are both peripheral, and are symmetrically placed with respect to the fovea. We used this method so that the two stimulus halves would fall on retinal regions having about the same acuity. If we had done as Jonides *et al.*, that is if the first stimulus half had impinged on the peripheral retina, and the second half on the fovea, then there would have been a difference in the quality of the information available to the visual system about the first and second stimulus halves. It seemed to us that we were improving the chances that trans-saccadic fusion would occur by making the stimuli of comparable quality. However, an interesting alternative presents itself. It could be that trans-saccadic fusion exists, but works only to integrate previously occurring peripheral information with presently available foveal information.

Another difference between the two experiments may be important. In our experiment, the direction of the eye movement the subject was required to make depended on the side on which the target point appeared on the screen. This changed randomly from trial to trial, sometimes being on the right, sometimes on the left. In Jonides *et al.*'s experiment, the stimulus always occurred on the right of the initial fixation point. It may be that fusion is facilitated by greater certainty of the spatial location of the stimulus.

Several further differences related to the stimuli that were used in the two experiments may also be related to the difference in results.

In our experiment, the stimuli were made of line segments instead of dots as Jonides *et al.*'s. It may be that the precision of trans-saccadic fusion is not very great, and that the precision of alignment of the line segments required to recognize the stimuli in our experiment was greater than in Jonides *et al.*'s. However, a comparison of the two tasks suggests this is not the case. In our experiment the precision required was about one third of a letter, that is, 0.32°. In Jonides *et al.*'s experiment a 5 × 5 dot matrix subtending 3° was used. Assuming that the task could be done providing the two stimulus halves were not displaced by more than one half the dot spacing from

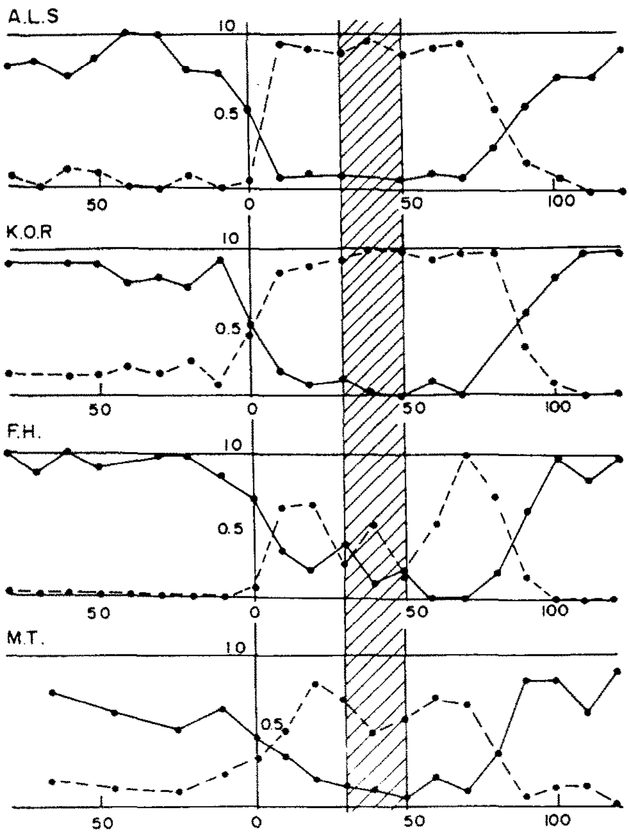

Fig. 2. Results for 4 subjects. The top two were the authors. The bottom two were naive. Each subject did 500 trials except M.T., who did 400. The solid line shows probability of correct response, the dashed lines are "Don't know" responses. The abscissa shows the time with respect to the saccade onset at which the second stimulus occurred. The hatched region between 30 and 50 msec is the critical region in which one stimulus half appeared before the saccade, and one after the saccade.

their true position, this implies an accuracy of 0.3°. It therefore seems like the accuracy required of the fusion mechanism was of the same order of magnitude in the two experiments.

Two other possibly relevant differences between our and Jonides *et al.*'s experiments concern the timing of the stimuli. The first difference is that the duration of the stimulus halves in our experiment was 1 msec, and in Jonides *et al.*'s it was 17 msec. Second, the blank interval between stimulus halves was 50 msec in our experiment and 37 msec in Jonides *et al.*'s. We believe that neither of these differences account for the difference in results, because we have done experiments using comparable durations in which trans-saccadic fusion also did not occur. In these experiments the stimulus halves consisted of a square with a small vertical bar either at "12 o'clock" or at "6 o'clock". When the two squares were superimposed, they formed a square with one vertical bar. The subject's task was to say when he saw this bar. Thinking that perhaps fusion did not occur because the stimuli were of too brief duration, we varied this duration from 1 to 200 msec. However fusion never occurred. We also thought that fusion would be favored if the stimulus halves had common parts: this would allow the two halves falling on different retinal locations to be superimposed somewhat in the way satellite pictures of the earth are superimposed, that is, making use of common boundaries. The stimulus halves already had the square in common, but we added two larger squares around each stimulus half so that it lay within a kind of picture frame subtending about 5°. We again found no fusion.

The conclusion seems to be that the only vital differences in the two experiments are the fact that in Jonides *et al.*'s work the second stimulus half was displayed foveally whereas in our work it was displayed peripherally, and the fact that in Jonides *et al.*'s experiment the spatial location of the stimulus was more predictable. Awaiting further work where these factors are investigated, the evidence from our present experiment suggests that trans-saccadic fusion probably does not exist*. Other more theoretical arguments in favor of this idea are the following.

*Since submitting this paper we have received personal communications from D. E. Irwin and from G. J. Mitchison, C. I. Baker and G. E. Hinton suggesting that the apparent trans-saccadic fusion found in the Jonides *et al.* (1982) report was probably an artefact of the remanence of the CRT phosphor used.

If trans-saccadic fusion existed, then the nervous system would have to have some way of knowing how to properly align the successive snapshots of the visual environment taken at each fixation. There are two ways this might be done: using extraretinal information about the distance through which the eye has moved in its orbit, and using what visual information is common to the successive snapshots.

The first method would require the existence of accurate extraretinal information about the eye's position in the orbit. Measurements done in the dark, that is, where no retinal, only extraretinal information is present, have shown that this extraretinal information is not very accurate (accuracy less than 1°), and is not properly time-locked to the saccade (errors of at least from 300 msec before to 1500 msec after the saccade) (cf. Matin, 1976).

The second method of "gluing together" successive snapshots, in which they are aligned on the basis of information common to both, is problematic. First, the quality of the information coming from different parts of the retina is different, the peripheral regions containing less color and detail. More important, changes in eye and head position provoke complex perspective changes within successive retinal images. For these reasons, any simple superposition of successive images could not work.

How then do we perceive the environment as stable and continuous despite eye movements? We suggest that our mental representation of the visual scene before us is not like a photograph, that is, it does not contain information in a code indicating the distribution of light in different parts of space. Rather, the representation we have is of a more semantic nature: a blue chair is coded as a "blue chair", with possibly other peculiarities we have noticed, like "wooden legs". It is not coded as an assembly of bluish light points at various points in space. In addition, the absolute position in space of the blue chair is not coded. Rather its relative position to the body and to other objects is coded, again in semantic terms (e.g. "in front of the red table"), and generally not in terms of a distance metric.

This kind of representation has the advantage of not requiring complex shifting or aligning of successive snapshots to compensate for eye or body movements. It has the disadvantage that position information is only retained to the extent to which it is semantically coded ("near", "far", "in front", "a few centimeters", etc.). The visual details of objects, as they would be recorded in a photograph, are also not retained unless they happen to be coded semantically. Several experiments support this idea. In reading, if the sentence being read is suddenly shifted during the time the subject makes a saccade, subjects are unaware of the shift and their reading is often not perturbed (O'Regan, 1981). This suggests that position of objects in space is not being coded at each fixation. Other experiments suggest that the visual identity of objects is also not being coded. For example, sentences are displayed on a computer screen in AlTeRnAtE cAsE lIkE tHiS. As the subject reads, during each saccade he makes, the computer changes all the capital letters into small letters and vice versa. Subjects' reading is unperturbed by this manipulation (McConkie, 1979). In other studies it is shown that even though information in parafoveal vision is contributing to processing, when this information is suddenly changed during the saccade, the subject in many cases is unaware of these changes (Rayner, 1975; Rayner *et al.*, 1980; Lévy-Schoen, 1981).

If our internal representation of space is semantic and not photographic we may further ask: how is it that we have the impression of the visual environment being rich in detail and located accurately in space, that is, we feel it is like a photograph? The answer we suggest is that this photographic representation is available on the retina. We have no need to make another version of it which is independent of eye location. We do have a continuously present internal representation of our environment, but it is semantic and does not contain detail of a pictorial nature. It does, however, give enough relative-position information so that we know approximately how to move our eyes if we want to obtain information with more "photographic" visual detail. It is interesting to note that in this view, the visual scene acts as a kind of *external memory buffer* whose unclear parts can be activated by making an eye movement. Just as by an effort of attention we can bring to mind the details of remembered events, by moving our eyes we can cause parts of the visual field which momentarily interest us to have a more "photographic" quality.

REFERENCES

Jonides J., Irwin D. E. and Yantis S. (1982) Integrating visual information from successive fixations. *Science* **215**, 192–194.

Lévy-Schoen A. (1981) Flexible and or rigid control of oculomotor scanning behavior. In *Eye Movements: Cognition and Visual Perception* (Edited by Fisher D. F., Monty R. A. and Senders J. W.). Erlbaum, Hillsdale, NJ.

Lévy-Schoen A. and O'Regan J. K. (1979) Comment voit-on en bougeant les yeux? Expériences sur l'intégration des images rétiniennes successives (Résumé) *Psychologie Française* **25**, 76–77.

Matin L. (1976) Saccades and extraretinal signal for visual direction. In *Eye Movements and Psychological Processes* (Edited by Monty R. A. and Senders J. W.). Erlbaum, Hillsdale, NJ.

McConkie G. (1979) On the role and control of eye movements in reading. In *Processing of Visual Language* (Edited by Kolers P. A., Wrolstad M. E. and Bouma H.). Vol. I. Plenum Press, New York.

O'Regan J. K. (1981) The convenient viewing position hypothesis. In *Eye Movements: Cognition and Visual Perception.* (Edited by Fisher D. F., Monty R. A. and Senders J. W.) Erlbaum, Hillsdale, NJ.

Rayner K. (1975) The perceptual span and peripheral cues in reading. *Cog. Psychol.* 7, 65–81.

Rayner K., McConkie G. and Zola D. (1980) Integrating information across eye movements. *Cog. Psychol.* **12**, 206–226.