The Role of Early Vision in the Determination of Depth and Motion from Ambiguous Binocular Information

Ko Sakai[1] and Mitsuharu Ogiya[2]

[1] Institute of Information Sciences and Electronics,

[2] Graduate School of Systems and Information Engineering, University of Tsukuba 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8573, Japan

Corresponding author: Ko Sakai, sakai@is.tsukuba.ac.jp

ABSTRACT

The visual system can determine motion and depth from ambiguous information contained in images projected onto both retinas over space and time. The key to the way the system overcomes such ambiguity lies in dependency among multiple cues—such as spatial displacement over time, binocular disparity, and interocular time delay—which might be established based on prior knowledge or experience. We conducted a psychophysical investigation of whether a single ambiguous cue (specifically, interocular time delay) permits depth discrimination and motion perception. Data from this investigation are consistent with the predictions derived from the response profiles of V1 neurons, which show interdependency in their responses to each cue, indicating that spatial and temporal information is jointly encoded in early vision.

1. Introduction

Motion and depth are fundamental attributes for determining and predicting the location of an object in the environment. When an object is moving in the natural environment, three cues, spatial displacement over time, binocular spatial disparity (BSD), and interocular time delay (ITD), may be available to the visual system for determining the motion and depth of the object. The selectivity of V1 neurons to these three cues is represented in three space-time response profiles—two for the monocular domain and one for the binocular domain—in which dependency between the selectivity of each cue is observed. The monocular space-time response profiles represent selectivity to spatial displacement over time, namely motion direction and velocity; the binocular response profile represents BSD and ITD, obtained from the cellular responses to two bars or dots that are sequentially presented to each eye [1]. Anzai *et al* have recently analyzed the responses of binocular complex cells in the cat striate cortex to the stimuli of various interocular spatial and temporal shifts, and reported that most neurons exhibited space-time-oriented response profiles in binocular domains [2], as shown in Figure 1. When the binocular response profile is oriented diagonally in space-time, a similar response to ITD and BSD is predicted, and it is impossible to distinguish whether BSD or ITD evoked the response. It has been also reported that the response profiles in monocular domains are generally similar to those in a binocular domain [2]. These results demonstrate dependency among the three cues: BSD, ITD, and displacement over time.

Dependency among the three cues might be the key to the visual system's mechanism for inferring depth and motion from ambiguous information contained in images projected onto the retinas. This idea leads to the prediction that a single cue containing ambiguity, specifically ITD, can evoke the perception of depth and motion. The space-time-oriented profiles in a binocular domain predict that the depth of an object will be discriminated from ITD in a manner similar to that from BSD. The correspondence between monocular and binocular profiles predicts that ITD will evoke the perception of motion direction, and also the velocity of an object. Based on computational analyses, we investigated psychophysically whether ITD alone evokes a perception of depth and motion that is consistent with the characteristics of the response profiles reported physiologically. The dependency among three cues will also explain Pulfrich-like effects and Mach-Dvorak's phenomena [3,4].

2. Experimental Results

To investigate the perception of depth and motion evoked from ITD, we designed a series of psychophysical experiments in which subjects observed moving random dots through a single narrow slit presented on a dichoptic, stereo display system, as illustrated in Figure 2-A. A one-pixel wide slit was used so that no pictorial cue for motion direction was possible. Such a narrow slit also excluded the cue from Da Vinci stereopsis that gives correct depth perception from binocularly unpaired stimuli originating from occlusion [5].

2.1 Discrimination of depth and motion direction

The space-time-oriented response profiles in a binocular domain predict that fine depth discrimination from ITD is possible in a manner similar to that from BSD. We determined whether human subjects are able to correctly perceive the relative depth of random dots with distinct ITD. The subjects were able to perceive the relative depth in this condition, as shown in Figure 2-B. The correct rate for the perception of depth increased up to 85% as the difference in ITD increased. The perception of motion direction was almost perfect (around 95%) for the entire range of the ITD, as shown in Figure 2-C. The results indicate that fine discrimination of depth and motion direction are possible from ITD.

2.2 Invariance to occluding direction

In the second experiment, we investigated whether occluding direction affects the perception of depth and motion direction evoked from ITD. It has been widely believed that occlusion cues are crucial for perception from ITD, and that the neural correlates might include intermediate-level processing, such as occurs in V2, V4 and MT [6,7]. However, our hypothesis predicts that depth and motion direction will be evoked from an ITD even if an occlusion cue is inconsistent. We designed a 'contradictory' stimulus configuration in which the direction of ITD was orthogonal to the direction of occlusion. In this condition, the slit was rotated 90° to the horizontal and the dots moved upward or downward, as illustrated in Figure 3-A. However, the time delay between left and right eyes was identical to that in the natural condition. The directions of ITD (left–right) and the occlusion (up–down) were therefore inconsistent. This condition is artificial because occlusion-evoked ITD should occur only if an object moves laterally behind a vertical slit.

The results, as shown in Figure 3-B, indicate that subjects perceived depth with a correct rate of up to 85%, suggesting that ITD is capable of yielding depth perception, despite being inconsistent with the occlusion direction in the absence of BSD. The apparent motion direction was measured in two successive blocks. Although the correct rate for the up–down choice was close to the 50% chance rate for the entire range of ITDs

(Figure 4-A), the correct rate for left–right choice was about 70% (Figure 4-B). This indicates that the apparent direction of motion is consistent with the ITD, but inconsistent with the occlusion. This result is surprising because the horizontal slit could be expected to have given the impression of vertical motion, as a result of occlusion.

2.3. Depth from time delay and spatial disparity

Finally, we compared the apparent depths derived from ITD and BSD cues. Positive, linear correlation between the depths from ITD and BSD is expected from the space-time-oriented response profiles in binocular domains. The similarity of the response profiles of each neuron between monocular and binocular domains predicts that the velocity of an object could also be inferred from the ITD. The stimulus configuration was identical to that of the first experiment with the vertical slits, except that the lower slit was wider (1.1°) and the lower rectangle was replaced by a stationary solid bar as a reference, as illustrated in Figure 5-A. The apparent depths of the rectangle and the bar were compared using a constant-stimuli method. The results for each of three subjects are plotted in Figure 5-B, together with the results for a solid square instead of random dots, showing the positive, linear correlation between apparent depths evoked by ITD and BSD (correlation coefficient was 0.90). The divergence between the apparent depths evoked from BSD and ITD might increase as the ITD increases. One reason is that an ITD larger than 100 ms typically causes difficulty in binocular fusion of the random dot stimuli; observers tend to see two moving objects rather than one. A similar ceiling for interocular delay has been reported for Mach-Dvorak phenomena [8] and Da Vinci stereopsis [7]. This ceiling of about 100 ms for depth perception is consistent with the typical space-time-oriented response profiles in the binocular domains of striate complex cells [2].

3. Discussion

The independence of the apparent depth from the occluding direction is consistent with the notion that V1 neurons are responsible for perception from ITD. It has been widely believed that occlusion, as well as surface segmentation and the determination of figure direction, is processed in intermediate-level vision such as V2 and V4 [9]. If the coherent perception of motion and depth from ITD originates from the grouping of the responses of V1 neurons with space-time-oriented receptive-field structure, the occlusion process that takes place in later stages will not alter the neuronal responses that have already been grouped [10,11]; it is thus natural to observe the independence of apparent depth from the occluding direction. These results indicate the joint encoding of temporal and spatial information in the visual system, the independence of such coding from other cues such as occlusion and pictorial cues, and the crucial role of V1 neurons in this process.

References

1) Anzai, A., Ohzawa, I. & Freeman, R. D. Neural mechanisms for processing binocular information I. Simple cells. *J. Neurophyiol.* **82**, 891-908 (1999)

2) Anzai, A., Ohzawa, I. & Freeman, R. D. Joint-encoding of motion and depth by visual cortical neurons: neural bases of the Pulfrich effect. *Nature Neurosci.* **4**, 513-518 (2001)

3) Ross, J. Stereopsis by binocular delay. *Nature* **248**, 363-364 (1974)

4) Tyler, C. W. Stereopsis in dynamic visual noise. *Nature* **250**, 781-782 (1974)

5) Nakayama, K. & Shimojo, S. Da Vinci stereopsis: depth and subjective occluding contours from unpaired image points. *Vision Res*. **30**, 1811-1825 (1990)

6) Burr, D. C. & Ross, J. How does binocular delay give information about depth. *Vision Res*. **19**, 523-532 (1979)

7) Shimojo, S., Silverman, G. H. & Nakayama, K. An occlusion-related mechanism of depth perception based on motion and interocular sequence. *Nature* **333**, 265-266 (1988)

8) Michaels, C. F. *et al,* An onset to onset rule for binocular integration in the Mach-Dvorak illusion. *Vision Res*. **17**, 1107-1113 (1977)

9) Zhou, H., Friedman, H. S. & von der Heydt, R. Coding of border ownership in monkey visual cortex. *J. Neurosci*. **20**, 6594-6611 (2000)

10) Sakai, K. & Tanaka, S. Spatial pooling in the second-order structure of cortical complex cells - computational analysis. *Vision Res*. **40**, 855-871 (2000)

11) Sakai, K. & Hirai, Y. Neural grouping and geometric effect in the determination of apparent orientation. *J. Opt. Soc. Am, A* , **19**, 1049-1062 (2002)

Figure Captions

Figure 1

The response profiles of a striate neuron in monocular and binocular domains (reproduced from [2]), in which space-time-oriented profiles are visible in both binocular and monocular domains (top row). Green (light grey) and orange(dark grey) regions indicate positive and negative values respectively. The neuron is tuned to *near,* moving to the *left*, which is depicted in simple illustrations in the bottom row.

Figure 2

The stimulus configuration, moving dots observed binocularly through a narrow slit, is shown in (A) as a simplified schematic illustration. Note that ITD alone cannot correctly yield depth and velocity simultaneously. The mean correct rate among the three subjects for the determination of relative depth is plotted in (B) as a function of the difference in ITD between the two rectangles, with error bars indicating the standard deviation. The temporal disparity that is a product of ITD and the designed velocity of dots (2.8 °/s) also appears on the abscissa. The estimated correct rate of around 95% for motion direction is shown in (C). These results show that ITD evokes the perception of both depth and motion direction. The results for solid squares as opposed to random dots are also plotted, with triangular symbols.

Figure 3

The perception of relative depth in a contradictory condition in which temporal and occlusion cues are inconsistent. The correct rate for depth judgment is shown in (B). The conventions used are the same as those for Figure 2. There was no significant difference in results between the natural (vertical slit) and contradictory (horizontal slit) conditions (ANOVA p=0.72).

Figure 4

The perception of relative motion direction in a contradictory condition in which the estimated correct rate for motion direction, up or down, approximates the 50% chance rate, as shown in (A). The result of right–left motion judgment is shown in (B), indicating a correct rate of about 70% independent of the amount of ITD. This suggests that ITD is capable of yielding both depth and motion direction even if the occlusion cue is contradictory.

Figure 5

The quantitative comparison of apparent depth derived from ITD and that from BSD. Stimulus configuration is illustrated in (A). The 50% thresholds for the three subjects are plotted in (B). The error bars represent the residual standard deviation for curve fitting. The apparent depth from temporal disparity corresponds almost equally to that from spatial disparity.

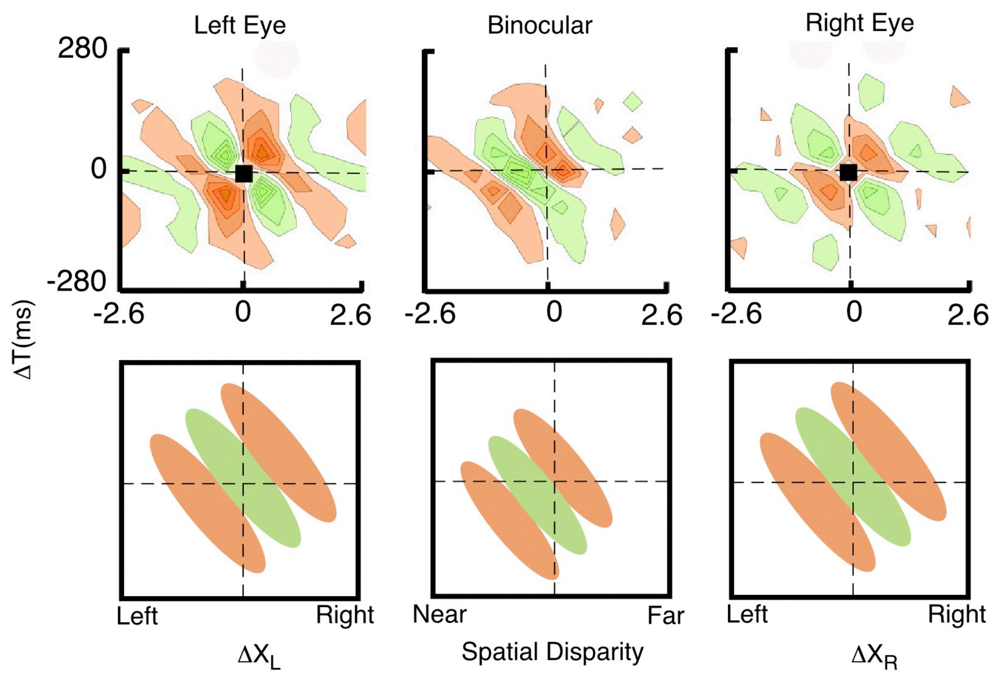
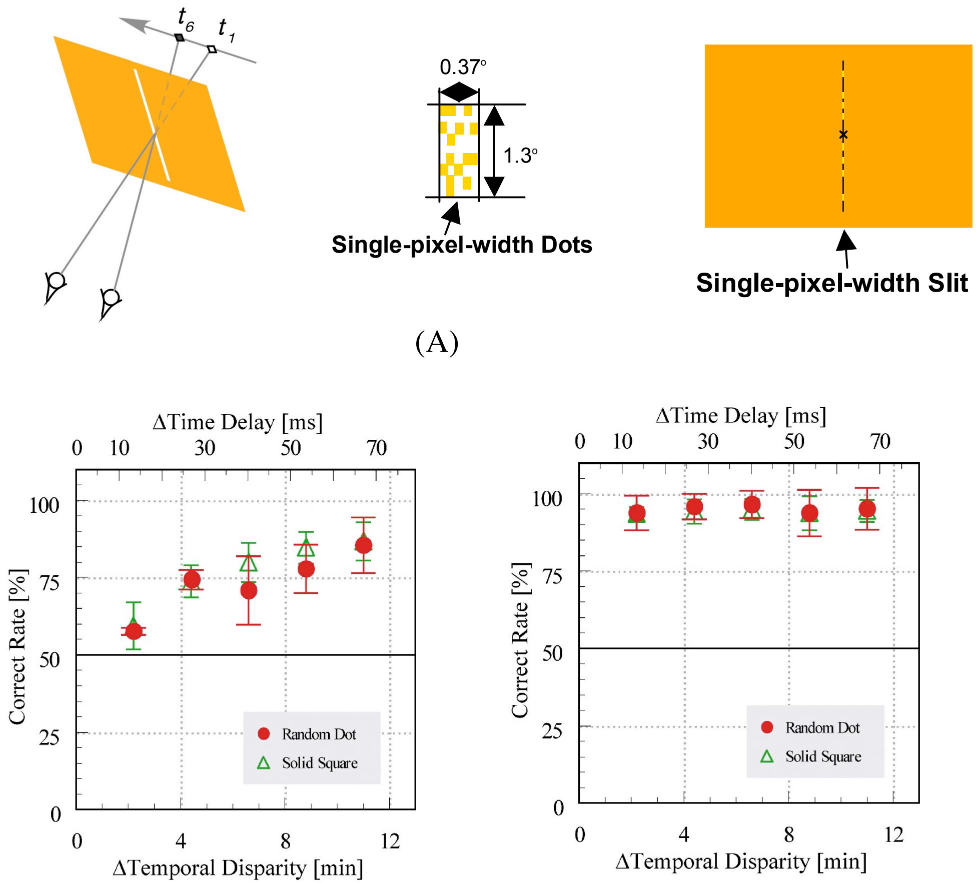


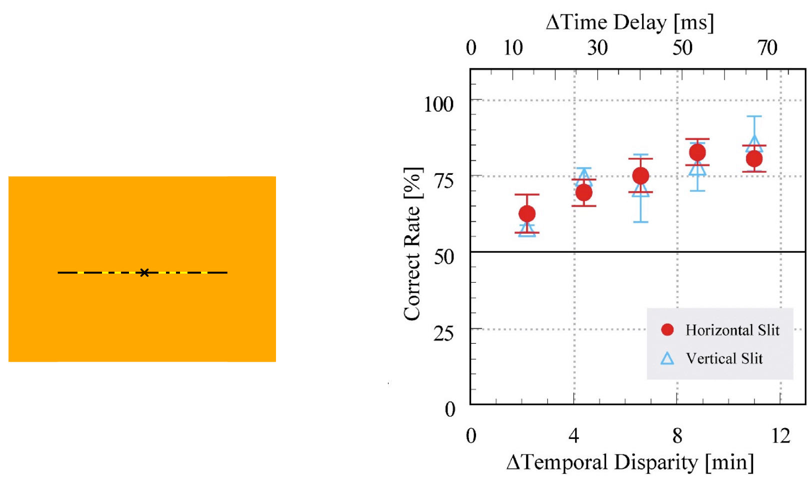
Figure 1



(A)

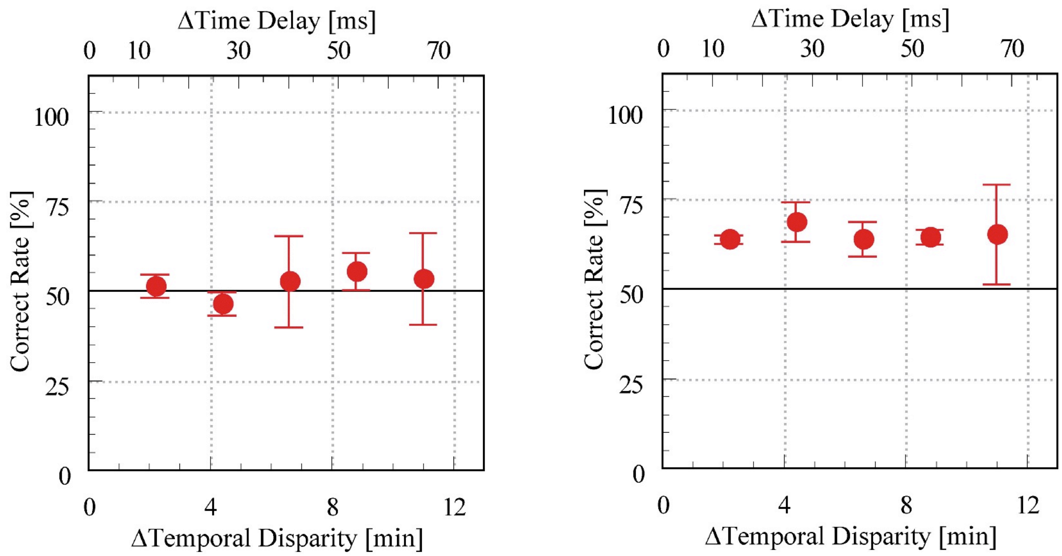
(B) (C)

Figure 2



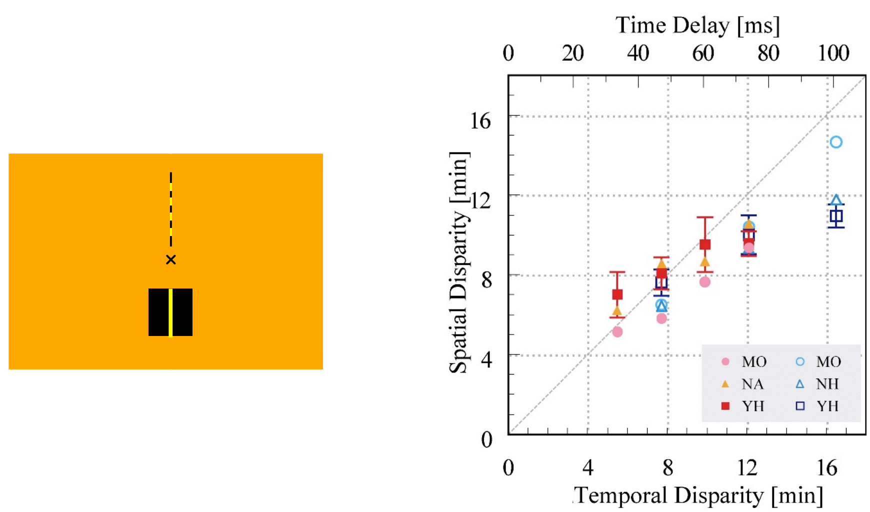
(A) (B)

Figure 3



(A) (B)

Figure 4



(A) (B)

Figure 5