Spatio-temporal dynamics of depth propagation on uniform region

Shigeaki Nishina [a,*], Masato Okada [b], Mitsuo Kawato [a]

[a] *ATR Computational Neuroscience Laboratories, 2-2-2 Hikaridai, Seika, Soraku, Kyoto 619-0288, Japan*
[b] *RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan*

Abstract

The depth of each point on a binocularly presented untextured horizontal bar is physically ambiguous except for the two vertical edges at both ends, since the correspondence between left and right images is not unique on such a uniform region. These depths, however, are unambiguously perceived, and this suggests the existence of some mechanism that interpolates the depth information from the two ends toward the center. Temporal properties of this integration process were examined by a phase-matching task, which allowed us to measure the phase of the perceived depth at the center of a horizontal bar when disparities at the ends were sinusoidally oscillated. We found that the perceived depth at the center of the bar was slightly temporally delayed for 7–60 ms relative to the physical depth at the ends. The difference increased with the length of the bar, decreased as the vertical position of the bar became farther from the fixation point, and increased in the presence of occluders. This finding indicates that depth information is propagated over an object to solve this ambiguity by using a time-consuming process. Accordingly, we suggest that depth propagation is accomplished by spatially local diffusion-like interactions of locally represented depth information.
© 2003 Published by Elsevier Ltd.

Keywords: Visual filling-in; Depth perception; Horizontal disparity; Computational time

1. Introduction

Binocular disparity is one of the most important cues for depth perception. The images of the external world from the left and right eyes slightly differ because the two eyes are located in spatially different places. The visual system integrates these two images and produces a single percept of the external world, using the difference between the two images as a depth cue (Wheatstone, 1838). However, because the mechanism of depth perception from horizontal disparity depends on the horizontal displacement of the two eyes, it fails to detect the depth of an object under certain spatial configurations of an observer and an object. As a typical case, consider observing an untextured horizontal bar. Fig. 1A shows a very simple stereogram that consists of exactly the same horizontal bar for left and right images. When this pattern is binocularly observed, the depth of each point on the bars is locally ambiguous, except for the left and right edges. These local ambiguities are re-

lated to the confined profiles of receptive fields of neurons in the early stages of the visual system. It is known that a large portion of the binocular neurons in V1 are disparity selective (Margaret & Doris, 1999; Poggio, Gonzales, & Krause, 1988). These neurons, however, cannot detect disparity when the midway portion of a horizontal bar is projected onto their receptive fields because the inputs from the two eyes are exactly the same in that case.

Notwithstanding these local ambiguities, the depths of interior portions of the bar are perceptually unambiguous. Actually, the depths of the physically ambiguous parts are perceived as the same depth as the endpoints at which the unique correspondence is obtained (Wurger & Landy, 1989). In other words, the ambiguity of the center area is interpolated using the depth information available at the endpoints (Blakemore, 1970; Halpern, Patterson, & Blake, 1987). A similar ambiguity has been shown for textured patterns in the case of stereograms with multiple correspondences (Julesz & Chang, 1976; Mitchison & McKee, 1985). Stereograms with spatially repeated texture patterns give multiple possible correspondences between left and right images. These studies showed that the

---

[*] Corresponding author. Fax: +81-774-95-2647.
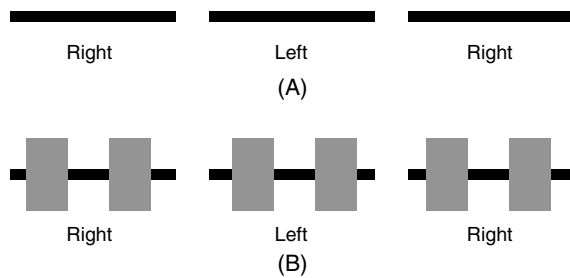*E-mail address:* nishina@atr.co.jp (S. Nishina).

Fig. 1. An example of a simple stereogram. Local depth ambiguity at the center of a horizontal bar is solved by integrating depth information available at the endpoints. The perceived depth at the center is the same as that at the endpoints (A). Such integration can also be observed even when parts of the horizontal bar are occluded (B). Left and center figures are used for cross-eyed viewing, and right and center for parallel viewing (although the three bars are exactly the same as those in A).

unique correspondence at the edges determined the correspondence of the ambiguous center area. This sort of depth integration is also the case with partially occluded objects (Fig. 1B). Even when an object is partly occluded, the depth information passes behind the occluder as long as the divided parts are perceived as a single object.

Many studies have demonstrated the phenomenon that the percept at one area is modified or even generated by nearby presented patterns. This phenomenon has been observed under various situations and is called perceptual filling-in (reviewed in Pessoa, Thompson, & Noë, 1998). Regarding brightness perception, it has been shown that the brightness of an area is affected by the brightness profile at the edge of the area (Cornsweet, 1970; Craik, 1966; O'Brien, 1958). This phenomenon is well-known as the Craik–O'Brien–Cornsweet effect (COCE), where the brightness of a uniform area is perceived at a brightness consistent with the steep discontinuity. Davey, Maddess, and Srinivasan (1998) investigated the temporal properties of the brightness changes due to this effect. They used a COCE pattern with a continuously reversing edge contrast and showed that changes in perceived brightness depended on the spreading rate and distance from the edge. The temporal properties of brightness perception have also been examined in terms of contextual modulation (De Valois, Webster, De Valois, & Lingelbach, 1986; Rossi & Paradiso, 1996), using visual stimuli that cause brightness assimilation or contrast. These recent studies have helped to clarify the neural representation of brightness. These results suggest lateral spreading processes for brightness perception. For depth perception, a phenomenon analogous to the brightness COCE has also been reported (Anstis, Howard, & Rogers, 1978). Their results showed that the perceived depth is determined not only by the local disparity but also by the surrounding context. The depth information obtained at the discontinuity affects the percept of the entire surface.

However, they examined only spatial properties of the phenomenon, so there is still a need for studies examining the temporal mechanism of depth filling-in.

In the present study, we examined whether spatio-temporal propagation of information is also observed in depth perception. Investigations were made principally on temporal properties in solving depth ambiguity, which have not been examined so far. We adopted a phase-matching task to measure the perceived depth of the center of an untextured horizontal bar that moves sinusoidally in depth. A vertical probe was presented at or above the center of the horizontal bar and oscillated sinusoidally at the same frequency as the horizontal bar. Subjects adjusted the phase of the probe oscillation to match it to the center of the horizontal bar in perceived depth. Examining temporal changes of depth at the center in detail should provide an important clue to the spatio-temporal characteristics of depth propagation.

2. General method

2.1. Apparatus and stimuli

A Silicon Graphics OCTANE MXI graphics workstation was used to create and present the stimuli. The stimuli were presented on a color display (Silicon Graphics GDM-20E21). Stereographics Crystaleyes LCD shutter glasses were used for the stereo presentation. The screen refresh rate was 60 Hz for each eye. The subjects sat in a chair and made responses with a keyboard. The view distance was 0.8 m and the horizontal resolution was 0.019°/pixel, which is the unit of disparity. A chin rest was used to maintain the head position during each session.

A horizontal bar was presented as a stimulus on a dark background. The thickness of the bar was 0.3°. The color of the bar was red (rgb = [1.0, 0.0, 0.0]) in order to minimize the crosstalk between the images presented to the two eyes. The luminance of the bar was 17.0 cd/m². The disparities at the endpoints of the bar oscillated with 0.1° amplitude. The frequency was 1.0 Hz in Experiments 1, 3, and 5, and 1.5 Hz in Experiments 2 and 4. When the disparity of the endpoints were periodically changed, subjects generally perceived the periodic depth motion of the entire bar with the same frequency as the oscillation of the endpoints' disparity.

A vertical bar was presented as a depth probe at or above the center of the horizontal bar (Fig. 2 for Experiments 1–4, and Fig. 12 for Experiment 5). Its color and luminance were the same as the horizontal bar's. Consequently, when the vertical probe intersects the horizontal bar, there is no edge, either vertical or horizontal, at the intersection region. In the entire elongated cross-like figure consisting of the horizontal and vertical bars, depth information was available only at the ver-
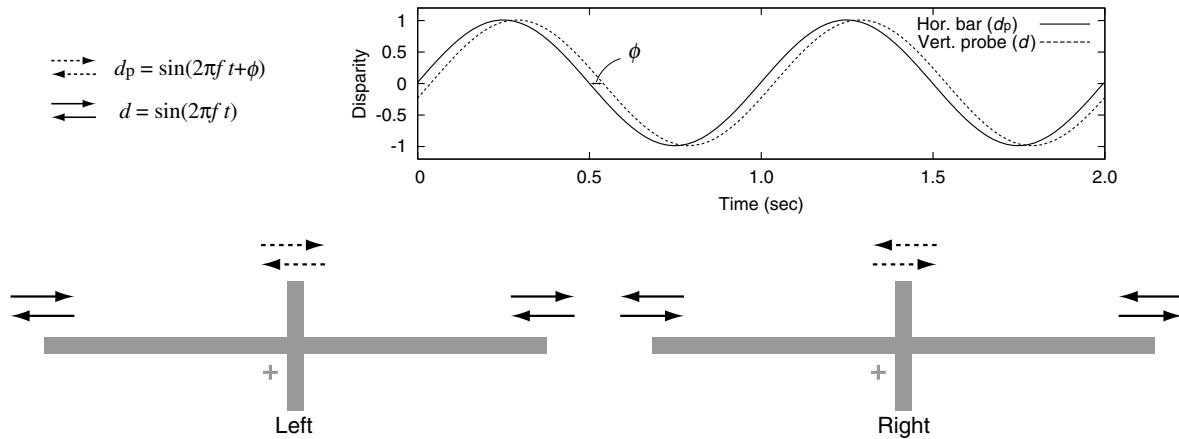
Fig. 2. Only the vertical edges of the stimulus have disparity information. In the experiments, the disparity of the edges were continuously updated according to the equations in this figure. t represents time, f represents frequency of oscillation, and ϕ represents phase difference between center and endpoints. The disparity change at the ends of the horizontal bar and that at the vertical edges of the vertical probe differed only in their phases. The frequency and amplitude were exactly the same.

tical edges of the horizontal bar and the vertical probe, since the correspondence between left and right images cannot be uniquely determined on horizontal edges or in uniform regions.

The probe also oscillated in depth with the same frequency and amplitude as the horizontal bar. The initial phase difference at the beginning of a trial was randomly set between the range of $-\pi$ and π. The horizontal and vertical bars started moving by a subject pressing a key. The subjects could change the phase difference between the endpoints of the horizontal bar and the vertical probe by pressing keys. The step size for disparity change was 0.019°/pixel for both horizontal and vertical bars. The step size for adjustment of the phase difference was 1.0° ($\pi/180$). Although the disparity was limited by the horizontal resolution of the display, 0.019°/pixel, a more precise measurement could be achieved by randomly setting the initial phase difference and averaging the results across multiple trials. The adjustment of the phase difference was made by pressing one of two keys. The phase difference can be any value between the range of $-\pi$ and π. If the phase difference exceeds π by the subject's pressing the key, 2π is subtracted from the phase difference, and vice versa, so that the value of the phase difference always remains between $-\pi$ and π. The subject was able to increase or decrease the phase difference freely and chose the point where the percept was closest to that shown in Fig. 3A. The direction of adjustment was arbitrarily determined by the subject. Because in most cases the subject went back and forth several times, the bias caused by adjustment in the ascendant or descendant direction must be small even within a trial, and it was nearly canceled by averaging across trials.

A fixation point was presented near the intersection point of two bars. The fixation point was always pre-
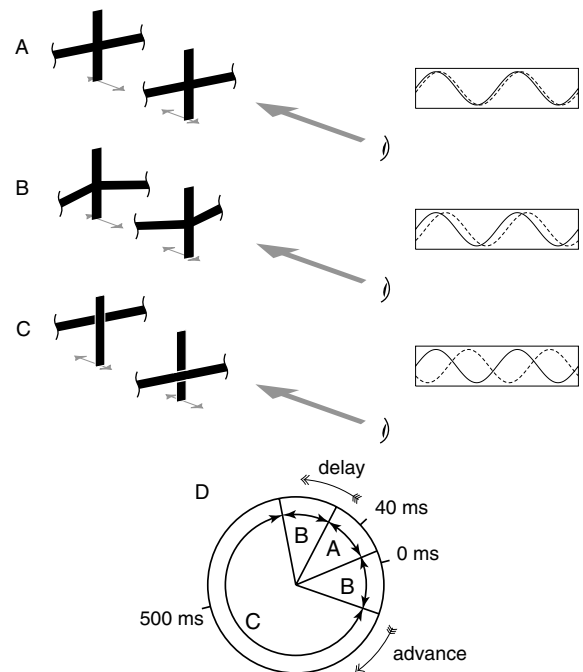

Fig. 3. A, B, C: Three possible percepts at the center of a horizontal bar with different phase gaps. Each figure shows a percept at a moment under a certain phase difference. A: The vertical bar is perceived to be moving precisely with the horizontal bar when the phase difference is a relatively small delay. B: The vertical bar is perceived to be on the horizontal bar, but the horizontal bar is perceived to be flapping and bending at the point of intersection when the phase difference is either more advanced or delayed relative to the phase difference that gives the percept (A). C: The vertical and horizontal bars are perceived to be separately moving when the phase difference is very large. D: A typical example of phase difference with the corresponding three percepts shown as a schema.

sented with zero disparity. A cross was used as the fixation point in Experiments 1 and 2, and nonius figures were used in Experiments 3–5 to control the eye movement.

The nonius pattern consisted of dichoptically presented T-shaped and inverse T-shaped patterns (Fig. 8). When the nonius figures were used as the fixation point, subjects were required to keep the two vertical lines of the T's collinear during the task.

2.2. Subjects

Five subjects participated in Experiments 1–4, and two subjects in Experiment 5. Three of the five subjects who participated in Experiments 1 and 2 also participated in Experiments 3 and 4. One author participated as a subject in all experiments, but all of the others were naïve as to the purpose of the experiments. All of the subjects had normal or corrected to normal vision.

3. Experiment 1

3.1. Procedure

As shown in Fig. 2, the horizontal bar and the vertical probe crossed each other in Experiments 1–4, that is, they were overlapping. The percept at the intersection of the two bars varies according to the phase difference between the vertical probe and the ends of the horizontal bar. As a result of a pilot experiment, we confirmed that there exist three kinds of percepts at the center area according to the phase difference between the probe and the endpoints of the horizontal bar. In the first case (Fig. 3A), the vertical bar is perceived to be moving seamlessly and synchronously with the horizontal bar. The two bars were perceived as if they were glued together, and the horizontal bar appears flat with the same depth as the vertical probe at any moment of the oscillation cycle. In the second case (Fig. 3B), the vertical bar is perceived to be glued on the horizontal bar, but the horizontal bar appears to be flapping and bending at the point of intersection in the manner of a butterfly (the wings are the two segments of the horizontal bar and the body is the vertical probe). In the last case (Fig. 3C), the vertical and horizontal bars are perceived to be separately moving. When the two bars are perceived separately, the two bars are basically perceived at different locations in depth during most of an oscillation cycle. Because disparity is available only at the vertical edges, the perceived depth of the region other than those edges is the result of interpolation. The interior of the two bars is interpolated, and the intersecting region is perceived as a part of both the vertical bar and the horizontal bar. This means that two surfaces are perceived simultaneously at the intersecting point. The bar perceived in front was modally interpolated, and the other bar was amodally interpolated. In this case, because the two bars oscillate at the same frequency and amplitude,

one *cuts through* the other twice in one cycle. Although the percept shown in Fig. 3B clearly indicates interaction between the two bars, this is not an obstacle to our measurement method for the following reasons. The subjects were required to adjust the phase difference so that the percept became that shown in Fig. 3A. Specifically, they were asked to choose a phase difference at which the vertical probe and the horizontal bar oscillated together with the least local perceptual distortion of the surface at the center of the horizontal bar. Since this phase difference should give a minimum interaction between the two bars, we could objectively measure the time difference between the center and the endpoints of the horizontal bar. Although the ambiguous area of the horizontal bar does oscillate with the endpoints, the amplitude of the center oscillation may not be the same as that at the endpoints (Wurger & Landy, 1989). Actually, we also observed a percept where the amplitude at the central part was generally smaller than that at the endpoints, which is equivalent to the reports by Wurger and Landy (1989). However, even when the amplitude is not equal to that at the endpoints, phase matching can be done. A subject was still required to make the percept closest to that shown in Fig. 3A, which should minimize the perceptual phase difference. Therefore, the attenuation of the amplitude is not a substantial difficulty in this method. In addition, the computational model we describe in the discussion predicts very small attenuation within our experimental range, which was only 1–5% of the amplitude at the edges.

There were three experimental parameters (Fig. 4), i.e., length of the horizontal bar (8°, 10°, 12°, 14°, or
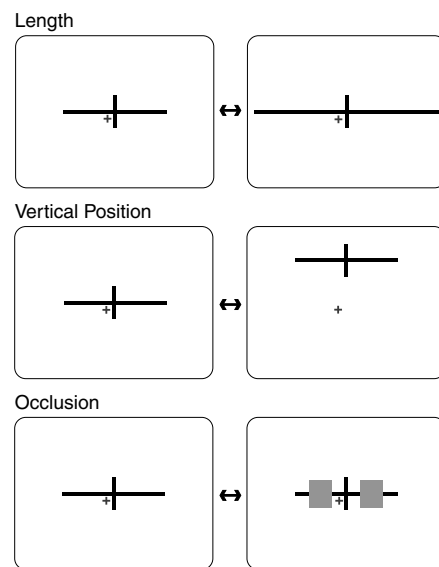

Fig. 4. In Experiment 1, we examined the effects of the length of the horizontal bar, vertical stimulus position, and occlusion.

16°), vertical position of the horizontal bar (0.5° or 3.0°), and the presence of occluders. Each subject completed 300 trials for each condition. The occluders were presented with 0.1° disparity so that they would always be perceived front-most. The fact that a percept like that in Fig. 3B occurs indicates that there is a certain interaction between the horizontal and vertical bars. A case of statically presented cross-shapes with different disparities at the vertical bar and the ends of the horizontal bar is mentioned by Nakayama and Shimojo (1992). They showed two possible interpretations for that stimulus: two separate flat horizontal and vertical bars, or a flat vertical bar and a horizontal bar bent at the center and adhered to the vertical bar. Although they argued that the former percept takes place much more frequently because of the generic viewpoint principle, they also mentioned that a small number of observers reported the latter percept. In our condition, the percept of a bent horizontal bar would take place more easily because the horizontal bar was much longer and disparity was relatively smaller than those used in their experiments; moreover, the effect of the generic viewpoint should be weaker. Although this kind of interaction may be an obstacle to objective measurement, it is practically impossible to completely eliminate such interaction. In the present experiments, we pursued a method that utilizes the interaction. That is, we utilized the fact that the percept varied according to the amount of interaction. In the case of C, the interaction is small or even zero because the bars are oscillating with two very different phases. It is also small in the case of A for the different reason that the oscillation phases of the two bars are close enough. In the case of B, the bars are not too close but not too far, a condition where the interaction is maximum. By observing these qualitative changes in percept, subjects would be able to report the phase that gives the minimum perceived depth difference between the two bars without quantitatively matching perceived depth. However, to further confirm that the interaction does not actually interfere with the measurement, we performed a similar experiment with physically reduced interaction between the vertical probe and the horizontal bar in Experiment 5. That is, the vertical probe was displayed above the horizontal bar without any intersection (Fig. 12). In this modified method, subjects were asked to adjust the phase of the vertical probe so that it moved with the center of the horizontal bar at the same depth.

3.2. Results

Fig. 5 shows the averaged results of five subjects. The time delay was calculated from measured phase differences between the center and the edges of the horizontal bar. A positive value means that the perceived depth at
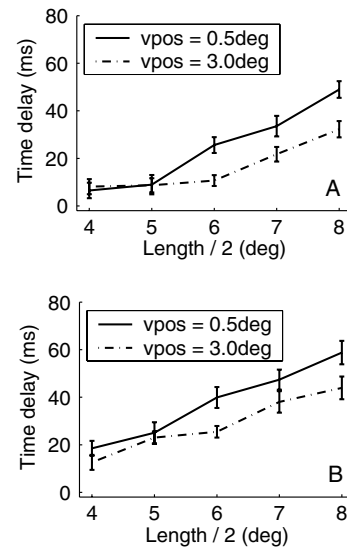

Fig. 5. Effect of line length from the results of Experiment 1 is shown for each vertical position (vpos). Graph A shows results with no occluders, and graph B shows results with occluders. The error bars show the standard errors of the means.

the center lagged behind the ends by that amount. The solid and broken lines show the results of vertical positions of 0.5° and 3.0°, respectively. Graph A shows results with no occluders, and graph B shows results with occluders. The abscissa represents the half size of the bar length (i.e., the distance from the center to one of the endpoints). As the frequency was 1.0 Hz, 50 ms corresponds to $\pi/10$ in phase. In all cases, the perceived depth at the center was delayed relative to that at the endpoints.

The delay at the center increased as the bar became longer. A two-factor analysis of variance (ANOVA) over the subjects was performed on the time delay and revealed a significant effect of the bar length ($F_{4,8} = 15.3$, $p < 0.001$). As for the vertical position, the delay was shorter when the stimulus was presented farther from the fixation point ($F_{1,4} = 82.7$, $p < 0.001$). Concerning the occluders, the time difference was larger when part of the bar was occluded (Fig. 6). The presence of the occluders showed a significant effect. With the occluders, the perceptual delay was larger ($F_{1,2} = 142.1$, $p < 0.01$).
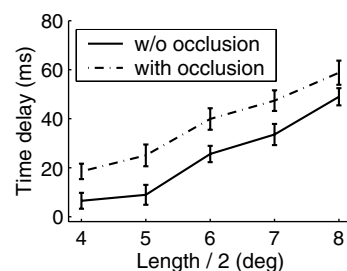

Fig. 6. Effect of occlusion.

4. Experiment 2

4.1. Procedure

We attributed the obtained phase difference to time delay, but it may not be clear whether the phase difference actually reflects time delay because the stimuli were periodically presented and the psychophysical measurements were performed while the subjects were allowed to view the stimuli for as many cycles as they wanted. The phase difference might be a result of some temporal dynamical characteristic of the visual system. In Experiment 2, we tested the effect of the oscillation frequency to confirm that the phase difference can be treated as time delay.

The method and procedure were the same as those of Experiment 1, except that the oscillation frequency was 1.5 Hz.

4.2. Results

Fig. 7 shows the results of both Experiment 1 and Experiment 2 with a 0.5° vertical position and no occluders. The solid line shows the 1.0-Hz frequency results (Experiment 1) and the broken line shows the 1.5-Hz results. Fig. 7A is plotted for the time difference, and Fig. 7B is plotted for the phase difference. For all combinations of occluders and vertical positions, ANOVAs were performed to examine the significance of the effect of phase difference and that of time difference. As a result, the effect of the oscillation frequency was significant for the phase difference ($F_{1,2} = 213.2$, $p < 0.005$) but insignificant for the time difference ($F_{1,2} = 4.04$, $p = 0.182$).
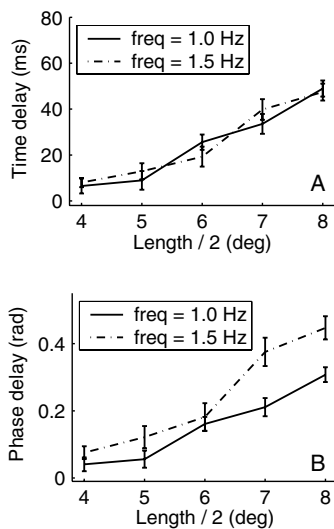
5. Experiment 3

5.1. Procedure

The purpose of this experiment and the next experiment was to more rigorously suppress the effect of eye movement. The phenomenon observed in Experiments 1 and 2 indicates that disparity information available at the ends might be propagated toward the center by some time-consuming mechanism. However, our interpretation may not be straightforward because eye movements might be initiated by the moving stimuli in the previous experiments. It is essential to control eye movements if we want to properly understand disparity propagation. In this experiment, we asked subjects to keep the vergence as stable as possible while performing the task. We used nonius figures as the fixation point to help the subjects monitor and actively minimize their vergence eye movement (Fig. 8). During the task, subjects were
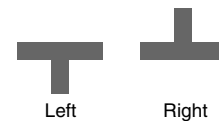
Fig. 8. The nonius figures used as the fixation point in Experiments 3–5. When each T-shaped figure was presented to each eye, the observer perceived an overlapped shape. Subjects were asked to keep the vertical lines of the nonius figures collinear while performing the task in order to minimize the effect of vergence eye movement.
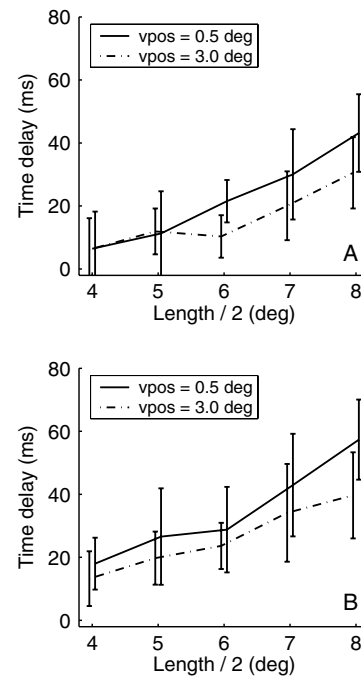
Fig. 9. Effect of line length from the results of Experiment 3 is shown for each vertical position (vpos). Graph A shows results with no occluders, and graph B shows results with occluders. The error bars show the standard errors of the means.
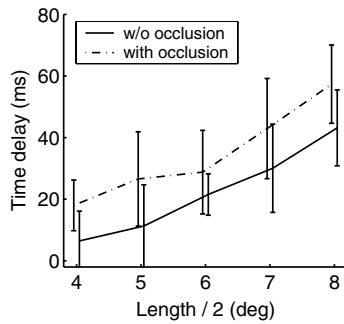
Fig. 7. Effect of frequency from Experiments 1 and 2. The same data are plotted as time delay (A) and phase delay (B).

Fig. 10. Effect of occlusion.

Fig. 11. Effect of frequency from Experiments 3 and 4. The same data are plotted as time delay (A) and phase delay (B).

required to fixate on the nonius pattern at the center of the screen and to keep the vertical lines of the nonius figures collinear. All other conditions were exactly the same as those in Experiment 1.

5.2. Results

We observed time delay in exactly the same way as in Experiment 1. The delay at the center increased as the bar became longer. A two-factor ANOVA revealed a significant effect of bar length ($F_{4,8} = 7.01$, $p < 0.01$). As for the vertical position, the delay was shorter when the stimulus was presented farther from the fixation point ($F_{1,4} = 10.3$, $p < 0.05$). The presence of the occluders also showed a significant effect. With the occluders, the perceptual delay was larger ($F_{1,2} = 51.1$, $p < 0.02$) (Figs. 9 and 10).

6. Experiment 4

6.1. Procedure

In this experiment, we again used nonius figures as the fixation point and confirmed the effect of frequency observed in Experiment 2. Conditions were the same as those in Experiment 2 except for using nonius figures as the fixation point.

6.2. Results

The obtained effect of frequency was very similar to that observed in Experiment 2. An ANOVA showed that the effect of oscillation frequency was significant for
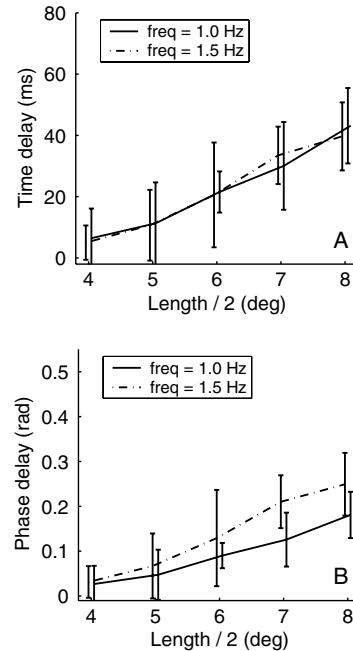
the phase difference ($F_{1,2} = 108.8$, $p < 0.01$) but insignificant for the time difference ($F_{1,2} = 1.34$, $p = 0.367$) (Fig. 11).

7. Experiment 5

7.1. Procedure

A vertical bar crossing the horizontal bar was used as a depth probe in Experiments 1–4. As shown in Fig. 3, however, a strong interaction between the two bars was observed for some ranges of phase differences. In Experiment 5, we examined whether this interaction helped or hindered the measurement of the center depth. In order to minimize the interaction, a shorter probe presented apart from the horizontal bar was used (Fig. 12). Only the effect of the horizontal bar's length was examined. Each subject completed 75 trials. The vertical position was 0.5° from the fixation point, the occluders were not presented, and the oscillation frequency was 1.0 Hz. All other conditions were exactly the same as those of Experiments 1 and 3.
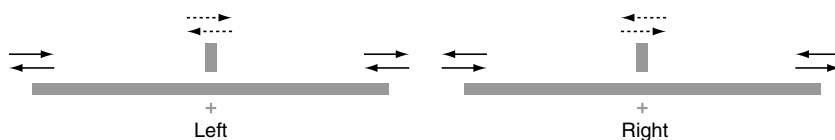

Fig. 12. Stimulus used in Experiment 5. A vertical bar presented above the horizontal bar oscillated at the same frequency as the endpoints of the horizontal bar, but its phase could be different from theirs (see Fig. 2).
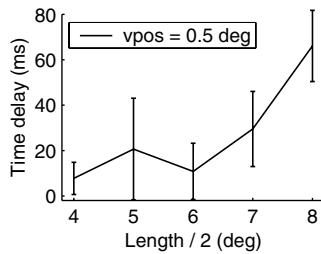
Fig. 13. In the minimized interaction condition, a similar effect of bar length was observed. However, the relatively larger deviation indicates that the task was more difficult than that with a crossed vertical probe.

7.2. Results

The effect observed in this condition was very similar to that in Experiments 1 and 3 (Fig. 13). However, the trial-by-trial variation of each subject's response was larger than that of those experiments, indicating that the interaction between the horizontal and vertical bars did not interfere with the measurement of the center depth but rather helped to achieve a more reliable measurement. In Experiments 1 and 3, because the subjects had to concentrate only on the qualitative change of percept at the center in the case of an intersecting vertical probe, the task should be much more easily performed. Although the variance was large, the existence of delay was significant ($t_{14} = 2.45$, $p < 0.05$).

8. Experiment 6

8.1. Procedure

The phase delay observed in Experiments 1–5 was ranged approximately from 7 to 60 ms. In this experiment, we show that such small phase differences can be measured with the phase-matching task even when using a 60-Hz CRT display. Although the step size of the phase is 6° for 1-Hz oscillation and 9° for 1.5-Hz oscillation because of the temporal limitations of the CRT, more precise measurement is possible because the observers perceive interpolated smooth motion from discretely presented stimuli. Consequently, in this task, we can check the pure precision of the phase matching without any influence of other factors such as horizontal distance or propagation. A fixation point and two dots were presented on the screen. The target dot was 0.5° above the fixation point, and a probe dot was 0.5° above the target. The size of the dots was 0.1°×0.1°. The dots were both moving in depth by temporally changing horizontal disparity in a sinusoidal manner in the same way as in Experiments 1–5. The amplitude of the disparity change was 0.1°, and the frequency was 1 Hz. At the beginning of a trial, the phase of the oscillation of the target dot relative to the probe dot was randomly set
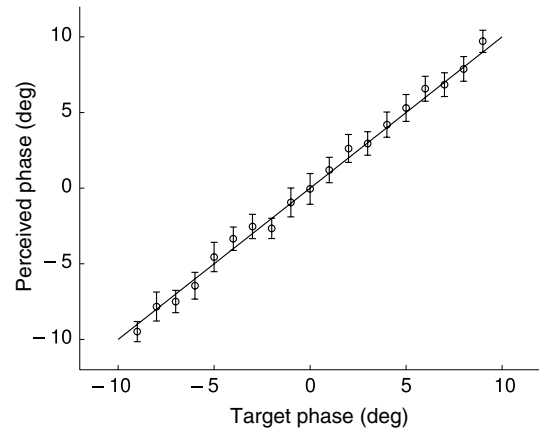

Fig. 14. Results of Experiment 6. Phase difference between the target and the probe was perceived in finer resolution than the refresh rate of the CRT.

as an integer value between the range of −9° and 9°. The step size, 1°, is much smaller than the phase resolution in a 60-Hz CRT, which is 6° in 1-Hz oscillation. The subject was asked to adjust the phase of the probe dot to make the two dots move perceptually in phase.

8.2. Results

Fig. 14 shows the relationship between the phase of the target dot and the subject's response. The phase of the target dot is the phase difference between the target dot and the probe dot at the beginning of the trials. The response is the amount of phase that the subject has changed to make the two dots perceptually in phase. The results show that the subject was able to match the phase of the two dots with finer resolution than the temporal resolution of the CRT. Although the presented depth motion was quantized both spatially and temporally, the subject could detect phase change much smaller than the temporal resolution of the CRT. This is natural because the observer perceived smoothly interpolated motion from the quantized stimuli and the percept is not directly related to the temporal limitations of a CRT. This result shows that the phase-matching task was appropriate for measuring the phase delay observed in Experiments 1–5.

9. Discussion

We have presented experiments demonstrating that the perceived depth at a physically ambiguous region of a horizontal bar may systematically lag behind that of the endpoints, depending on the length of the bar, vertical position of the bar, and presence of occluders. The oscillation frequency did not show a significant effect

when the temporal difference between the center and the ends was considered to be a time delay.

From the results of Experiment 1 (and also of Experiment 3), the delay was larger for a longer horizontal bar. The difference in processing time between fovea and periphery cannot explain our results for two reasons. One is that shorter processing time in periphery predicts the opposite result, that is, the probe should be moving physically ahead of the endpoints to be perceptually in phase with the center of the horizontal bar. The other reason is that the result of the experiment with occluders showed that the delay depended not only on the spatial location of the endpoints but also on the existence of the occluders between the center and the endpoints. If the delay were solely caused by the variation in processing time due to the retinal position, the occluders would not have had any effect. Our results indicate that the entire depth is not just labeled at the endpoints' depth but solved by some size dependent time-consuming process. Furthermore, the delay was shorter when the bar was presented vertically farther from the fixation point. The cortical sizes of objects of a physically equal size differ when presented at different retinal locations. The density of the receptive fields is the highest at the fovea and becomes lower toward the periphery. Accordingly, a stimulus with a fixed size covers fewer neurons in the cortex when presented peripherally than when presented foveally. If the depth information were propagated via local mutual interactions of neurons, the time for the endpoints' depth information to reach the center would depend on the number of neurons. Furthermore, the number of neurons is expected to be smaller for the periphery. Our results were qualitatively consistent with this assumption. The results suggest that depth is represented by neurons in the visual system maintaining spatial configurations of objects, and ambiguities are resolved by the mutual interactions of the neurons.

For this kind of interpretation, it is essential to ensure that the eyes are not moving during the task because disparity is defined based on the fixated depth. We used a nonius pattern as the fixation point in Experiments 3 and 4 and ensured stability of vergence by asking the subjects to maintain their vergence. The subjects were required to make the vertical lines in the two T-shaped patterns collinear during the task. The experimental results using this method were almost equivalent to the results obtained using an ordinary fixation point in averages, although the variance of the results were much smaller in Experiments 1 and 2. The larger variances in Experiments 3 and 4 can be readily explained by the larger difficulty due to their dual task characteristics. Accordingly, the results of Experiments 1 and 2 also support the idea of depth propagation.

In our previous research, we examined the temporal dynamics of motion propagation on edges and reported that the propagation of motion information takes more time when the occluded part of an object is larger (Okada, Nishina, & Kawato, 2000). This occlusion effect can be explained by assuming an amodal completion process, in which the processing time depends on the amount of occlusions. It is controversial whether the mechanisms of modal completion and amodal completion share the same brain functions (Kellman, Yin, & Shipley, 1998). The answer to this question may be explored by examining the dynamic properties of completions.

In the present experiment, the results of using the occluders indicated that depth information spreads more slowly over occluded regions. Sekuler and Palmer compared presentation times of priming stimuli, which show maximum priming effects in a shape recognition task, and argued that it takes more time to build a representation of an occluded surface than a visible surface (Sekuler & Palmer, 1992). Our results for depth propagation is consistent with theirs.

In the past two decades there has been considerable research into the computational understanding of visual perception. One important viewpoint conceives of vision as a kind of inverse optics (Kawato, Hayakawa, & Inui, 1993; Marr, 1982; Poggio, Torre, & Koch, 1985). Computationally, solving an ambiguity by using sparsely available unambiguous information is considered an 'inverse problem'. Mathematically, it is characterized as a so-called ill-posed problem that can be solved by being regularized by a certain a priori constraint. It is known that this sort of process can be implemented as an iterative calculation process based on locally represented information and local interaction of the information, that is, as an information diffusion process (Gerrits & Vendrik, 1970; Grossberg & Mingolla, 1985). Watanabe and Cole (1995) showed that the direction of apparent motion spatially propagates and affects the motion direction at another location in the scene. They investigated both spatial and temporal properties of the effects and suggested that motion correspondence information propagates by using a locally connected iterative network model. The spatio-temporal properties demonstrated in the present study are quite compatible with theories like these.

The motion direction of an edge behind an aperture is another typical example of a local ambiguity. Ambiguous disparity is caused by the impossibility of determining the spatial correspondence of the left and right images, while an ambiguous motion direction is caused by the impossibility of determining the temporal correspondence of sequential inputs. In this pattern, the tangential component of the motion is absolutely unavailable. We earlier reported that an ambiguous motion direction can be gradually solved (Okada et al., 2000). We also showed a computational model that consists of two processes, i.e., line completion and motion propagation over the line, and that can predict

experimental results quantitatively well. The results of the experiment and the simulation strongly suggested that the representation of the motion direction in the brain is spatially local, and a whole percept is achieved by gradually modifying the representation through mutual local interactions.

In the proposed model, the interpolation of visual information is performed by local information being sent to adjacent neurons in order. As for the depth interpolation we investigated in this paper, information is acquired only at endpoints of the horizontal bar. When the model is applied to depth interpolation, depth information is gradually transferred toward the center from both ends of the horizontal bar. In this case, the model predicts that the delay at the center of the horizontal bar becomes a quadratic function of the length of the bar. The model is based on the heat conduction equation and hypothesizes that the depth information of a surface is spatially and locally represented by multiple neurons. The neurons have a relatively small receptive field on incoming images and connect only with vicinal neurons. Therefore, the depth information of the neurons for ambiguous regions is determined by propagating the unambiguous depth information available at the vertical edges via short-range connections. Communication between neurons takes a certain amount of time, so the depth of the central part of the horizontal bar changes with delay. The model predicts that the delay is proportional to the square of the distance between one endpoint and the center. Although the model is based on local representation and local connection of neurons, a global percept can be obtained with the model because of the iterative processing. For example, when two bars with different lengths are presented to each eye, two endpoints are perceived at different depths. Although unique correspondence is available only at the endpoints and the depth of the intermediate area is locally ambiguous, a slanted surface is perceived. When our model is applied to this stimulus, the depth of the intermediate area, which is initially ambiguous, gradually varies as a result of the depth propagation from the endpoints. After a certain amount of time elapses, a flat surface linearly interpolating the two endpoints is formed. This is an exact reproduction of the human percept, and it shows that a local process alone is enough to obtain a global percept.

The ability of perceptual filling-in for blind spots has been extensively investigated. Some experiments have shown that this sort of percept is caused not by ignoring the area but by actually filling-in the area with surrounding visual information (Murakami, 1995; Paradiso & Hahn, 1996). In this paper, we took this a step further and examined the temporal dynamics of the mechanism for solving depth ambiguity. As a result, we found that the perceived depth at the center of a horizontal bar is slightly delayed from the oscillating depth of the endpoints. Although such stimuli are usually perceived as if the entire bar is oscillating together, we clearly showed an existing delay by using a variable phase probe. In our experiments, the depth of the stimuli was periodically oscillated. This was mainly because the task was very difficult to accomplish when the depth was changed only once for a trial. However, the results might be different under periodical and step changes (Moutoussis & Zeki, 1997; Nishida & Johnston, 1999). Considering these reports, we compared the results under two oscillating frequencies in Experiments 2 and 4. The results showed that the temporal difference was almost constant when it was considered to be a time delay. This indicates that the phase difference can be treated as a time delay. We also checked the effect of the amplitude and found that the effect was small. Our model predicts that the phase delay will not be affected by either the frequency or by the amplitude of depth motion, and this is consistent with the results of experiments.

Paradiso and Nakayama (1991) investigated the temporal dynamics of brightness filling-in by using a visual masking paradigm. Their results were consistent with the hypothesis that brightness signals are generated at the borders of the target stimuli and propagate inward at some rate. Grossberg and Todorovic (1988) proposed a filling-in model of brightness, and Arrington (1994) showed that the results of Paradiso and Nakayama (1991) fit excellently by using the model. Brightness filling-in has also been investigated by using dynamically changing stimuli (Paradiso & Hahn, 1996). These authors showed that when the brightness of an entire region is temporally changed, the perceived brightness is not uniform. This phenomenon is also important evidence of the propagation of visual information. However, there exists local brightness input at the center of the region in their experiments, and the brightness around the center is not completely ambiguous. Accordingly the nonuniform brightness observed in this pattern is the result of unknown interaction between the locally obtained brightness information and the information propagated from the edge. This is also the case in COCE experiments. The two regions divided by the edge have local inputs, so the difference in the perceived brightness is the result of the modulation carried out by the information propagated from the edge. In brightness experiments, local inputs of visual information are basically inevitable, except in an experiment using a blind spot as an induced area (Paradiso & Hahn, 1996). In this paper, we used horizontal disparity as the target visual information and were able to completely eliminate local inputs at ambiguous areas, allowing us to measure the discrepancy in the perceived depth between the center and the endpoints very precisely.

The results showed that the time for depth filling-in depends on the length of the bar, positions on the retinae, and presence of occlusions. This finding strongly

supports a propagation mechanism based on a kind of neural spreading mechanism, which we have already proposed for the perception of the motion direction of an edge moving behind an aperture. In the visual cortex, depth information is very likely represented locally, and the local interactions of that information play an important role in creating consistent percepts. Local representations, local interactions, and iterative calculations appear to form a fundamental mechanism of visual information processing in the brain.

References

Anstis, S. M., Howard, I. P., & Rogers, B. (1978). A Craik–O'Brien–Cornsweet illusion for visual depth. *Vision Research, 18*(2), 213–217.

Arrington, K. F. (1994). The temporal dynamics of brightness filling-in. *Vision Research, 34*(24), 3371–3387.

Blakemore, C. (1970). A new kind of stereoscopic vision. *Vision Research, 10*, 1181–1199.

Cornsweet, T. N. (1970). *Visual perception*. New York, NY: Academic Press.

Craik, K. J. W. (1966). *The nature of psychology: A selection of papers, essays and other writings*. Cambridge, MA: Cambridge University Press.

Davey, M. P., Maddess, T., & Srinivasan, M. V. (1998). The spatiotemporal properties of the Craik–O'Brien–Cornsweet effect are consistent with 'filling-in'. *Vision Research, 38*, 2037–2046.

De Valois, R. L., Webster, M. A., De Valois, K. K., & Lingelbach, B. (1986). Temporal properties of brightness and color induction. *Vision Research, 26*(6), 887–897.

Gerrits, H. J., & Vendrik, A. J. (1970). Simultaneous contrast, filling-in process and information processing in man's visual system. *Experimental Brain Research, 11*(4), 411–430.

Grossberg, S., & Mingolla, E. (1985). Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychological Review, 92*(2), 173–211.

Grossberg, S., & Todorovic, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: a unified model of classical and recent phenomena. *Perception and Psychophysics, 43*(3), 241–277.

Halpern, D. L., Patterson, R., & Blake, R. (1987). What causes stereoscopic tilt from spatial frequency disparity. *Vision Research, 27*, 1619–1629.

Julesz, B., & Chang, J. J. (1976). Interaction between pools of binocular disparity detectors tuned to different disparities. *Biological Cybernetics, 22*, 107–120.

Kawato, M., Hayakawa, H., & Inui, T. (1993). A forward-inverse optics model of reciprocal connections between visual areas. *Network: Computation in Neural Systems, 4*, 415–422.

Kellman, P. J., Yin, C., & Shipley, T. F. (1998). A common mechanism for illusory and occluded object completion. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 859–869.

Margaret, S. L., & Doris, Y. T. (1999). Receptive fields of disparity-selective neurons in macaque striate cortex. *Nature Neuroscience, 2*(9), 825–832.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: W.H. Freeman.

Mitchison, G. J., & McKee, S. P. (1985). Interpolation in stereoscopic matching. *Nature, 315*(6018), 402–404.

Moutoussis, K., & Zeki, S. (1997). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society of London B, 264*, 393–399.

Murakami, I. (1995). Motion aftereffect after monocular adaptation to filled-in motion at the blind spot. *Vision Research, 35*, 1041–1045.

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science, 257*(5075), 1357–1363.

Nishida, S., & Johnston, A. (1999). Perceptual asynchrony of colour and motion results from repetitive alternation. *Investigative Ophthalmology and Visual Science Abstracts, 40*(4), 1003.

O'Brien, V. (1958). Contour perception, illusion and reality. *Journal of Optical Society of America, 48*, 112–119.

Okada, M., Nishina, S., & Kawato, M. (2000). Dynamics of visual computation in the brain: experimental and theoretical evidence for iterative calculation. *Society for Neuroscience Abstracts, 30*, 670.

Paradiso, M. A., & Hahn, S. (1996). Filling-in percepts produced by luminance modulation. *Vision Research, 36*(17), 2657–2663.

Paradiso, M. A., & Nakayama, K. (1991). Brightness perception and filling-in. *Vision Research, 31*(7–8), 1221–1236.

Pessoa, L., Thompson, E., & Noë, A. (1998). Finding out about filling in: a guide to perceptual completion for visual science and the philosophy of perception. *Behavioral and Brain Sciences, 21*, 723–802.

Poggio, G. F., Gonzales, F., & Krause, F. (1988). Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *Journal of Neuroscience, 9*, 4531–4550.

Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature, 317*, 314–319.

Rossi, A. F., & Paradiso, M. A. (1996). Temporal limits of brightness induction and mechanisms of brightness perception. *Vision Research, 36*, 1391–1398.

Sekuler, A. B., & Palmer, S. E. (1992). Perception of partly occluded objects: a microgenetic analysis. *Journal of Experimental Psychology: General, 121*, 95–111.

Watanabe, T., & Cole, R. (1995). Propagation of local motion correspondence. *Vision Research, 35*(20), 2853–2861.

Wheatstone, C. (1838). Contributions to the physiology of vision—part the first. *Philosophical Transactions of the Royal Society, 128*, 371–394.

Wurger, S. M., & Landy, M. S. (1989). Depth interpolation with sparse disparity cues. *Perception, 18*(1), 39–54.