PERGAMON
Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 43 (2003) 2439–2449
Vision Research
www.elsevier.com/locate/visres

# Integration of binocular disparity and monocular cues at near threshold level

Makoto Ichikawa [a,*], Shinya Saida [b], Atsushi Osa [a], Kohkichi Munechika [a]

[a] *Department of Perceptual Sciences and Design Engineering, Yamaguchi University, 2-16-1 Tokiwadai, Ube, Yamaguchi 755-8611, Japan*
[b] *Research Institute of Human Science and Biomedical Engineering, National Institute of Advanced Industrial Science and Technology, 1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan*

Received 31 July 2001; received in revised form 16 June 2003

**Abstract**

We examined the dependency of the integration of multiple depth cues upon the combined cues and upon the consistency of depth information from different cues. For each observer, depth thresholds were measured by the use of stimuli in which different depth cues (motion parallax, binocular disparity, and monocular configuration) specified the surface undulating sinusoidally with different spatial frequencies and different phases. Analysis of $d'$ showed that the performance was better than the prediction of probability summation only when parallax and disparity cues specified an undulation with the same spatial frequency and same phase. The probability summation model overestimated the performance for the other conditions of combination of disparity and parallax, and for all of the conditions of combination of disparity and monocular configuration. These results suggest that the improvement in depth perception caused by integration of multiple cues depends on the type of combined cues, and that the visual system possibly integrates the depth information from different cues at different stages of the visual processing.

*Keywords:* Depth; Integration; Disparity; Parallax; Monocular configuration

## 1. Introduction

Previous studies about the cue integration at suprathreshold level offer the model in which the information from different cues are integrated in a linear manner, using weighted averaging (Landy, Maloney, Johnston, & Young, 1995; Rogers & Collett, 1989). This model of linear integration assumes a weak fusion (Clark & Yuille, 1990) in which the processing of a given cue is independent of the processing of the other cues (Fig. 1(a)). A study at near threshold level, however, reports that the processing of binocular disparity is not independent of the processing of motion depth cues (motion parallax, and kinetic depth cues). That is, Bradshaw and Rogers (1996) show that observers' performance in a depth detection task is better than the performance predicted by probability summation when binocular disparity and motion parallax (due to head movement) cues specify a sinusoidal surface vertically undulating with the same spatial frequency. The performance predicted by probability summation corresponds to the performance of the integration in a linear manner. Therefore, the results of Bradshaw and Rogers (1996) suggest that the processing of binocular disparity and that of parallax are not independent of each other, and that they improve each other. Similar improvement caused by viewing both binocular disparity and motion depth cues is found by Cornilleau-Pérès and Droulez (1993) who measure the depth threshold by the use of binocular disparity and kinetic depth cues specifying the same spherical surfaces. These studies about depth perception at near threshold level suggest a strong fusion (Clark & Yuille, 1990) in which the binocular disparity processing and motion depth cue processing improve each other (Fig. 1(b)).

Whether this strong fusion (Fig. 1(b)) is common for a combination of any depth cues at near threshold level, or is restricted to the integration of disparity and motion depth cues is not clear. The above mentioned studies could not answer this question because they used stimuli which only combined disparity and motion depth cues. The first purpose of this study, therefore, was to find out whether the improvement in depth perception due to cue

---

[*] Corresponding author. Tel.: +81-836-85-9724; fax: +81-836-85-9701.
*E-mail address:* ichikawa@yamaguchi-u.ac.jp (M. Ichikawa).

0042-6989/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0042-6989(03)00432-2

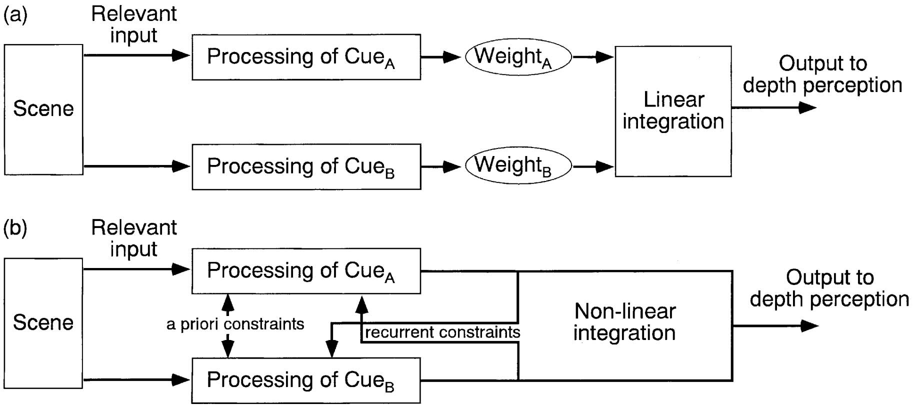

Fig. 1. Diagram of the fusion models in depth perception, adopted and modified from Clark and Yuille (1990). (a) Weak fusion model. The outputs of two or more modules for separate depth cues are averaged (linearly integrated) with the weights that are derived from measures of the relative reliabilities of each cue. The sum of the weights for multiple cues is assumed to be one. (b) Strong fusion model. The outputs of the component modules are no longer independent. A feedback loop is created, causing recurrent behavior.

combination depends upon the cues which are combined (Does the improvement due to non-linear integration depend on what Cue$_A$ and Cue$_B$ are in Fig. 1(b)?). To do this, we examined the improvement in depth perception not only for the stimuli in which binocular disparity was combined with motion depth cue but also for the stimuli in which binocular disparity was combined with monocular configuration cue. Our motion depth cue was motion parallax generated by yoking the observer's head movement to relative motion within a stimulus, as in Rogers and Graham (1979). Our monocular configuration cue was generated by curved lines, as in Stevens and Brookes (1988). Although there are other depth cues, we used this cue because it is known as a strong depth cue; it can overcome disparity in specifying apparent surface shape when it is inconsistent with disparity (Stevens & Brookes, 1988; Stevens, Lees, & Brookes, 1991).

As mentioned above, previous studies (Bradshaw & Rogers, 1996; Cornilleau-Pérès & Droulez, 1993) found that depth perception is improved for the stimulus which presented disparity with motion depth cues. Also, Rivest and Cavanagh (1996) demonstrate that contour detection is improved by presenting multiple sources (luminance, color, motion, and texture) compared to the detection due to a single source. Although Rivest and Cavanagh (1996) combined disparity with motion, the results of their study imply that combining the information from different attributes might improve signal detection. Therefore, we are facing a question. Does the strong fusion between disparity processing and motion depth cue processing found by the previous studies depend on the interaction between processing of different depth cues, rather than on the interaction between disparity processing and motion signal processing? The second purpose of this study was to answer to this question (Does the non-linear integration depend on the communication of depth signal rather than more primitive information from each cue processing in Fig. 1(b)?). For this purpose, we investigated whether the depth perception is improved when disparity was accompanied with stimulus motion although there is no head movement that is necessary to create a motion parallax depth cue. In this case, the retinal motion cannot specify any consistent depth pattern.

Our third purpose was to examine the dependency of the way in which depth information is integrated from different cues upon the consistency between depth cues. The above mentioned studies at near threshold level have not examined how the consistency of depth information from different depth informative sources affects the integration processing because, in their stimuli, cues always specified the same depth order and surface shape. On the integration of disparity with motion depth cues at suprathreshold levels, several studies demonstrated that the visual system integrates the depth information from these cues even if the information from these cues were inconsistent to each other in terms of the surface shapes they defined. That is, observers perceived the surface shape as the combination of two different shapes specified by different cues (Ichikawa & Saida, 2002; Rogers & Collett, 1989; Uomori & Nishida, 1994). Whether depth cues specifying different surface shapes are integrated in a similar way at near threshold levels is not known. In the previous study (Ichikawa & Saida, 1998), we measured the depth threshold for the stimuli in which disparity and parallax specified sinusoidal undulations with different spatial frequencies and phases. In the study, we found that the sensitivity elevation was restricted for the case in which the spatial frequency and the phase specified by disparity were consistent with those specified by parallax. In this study, we aim to confirm the results of our previous study for the combination of disparity with parallax (Does the non-linear integration depend on the consistency of information from Cue$_A$ and Cue$_B$ in Fig. 1(b)?), and to examine how the visual system copes with consistency of information from disparity and monocular configuration in integrating these cues. To examine how the consistency of information from different cues affects the integration processing, we used stimuli in which the cues specified the sinusoidal undulation with different spatial frequencies and different phases.

## 2. Methods

### 2.1. Observers

The three authors (AO, KM, MI) and two naive observers (AM, MO) served as observers. All of them had corrected normal acuity and normal stereo. Before the experiments, they completed several training sessions to familiarize themselves with the equipment and stimuli.

### 2.2. Apparatus

A personal computer (IBM Aptiva T8C) and VSG 2/3 board (CRS) presented stimuli on a 17 in. display (Eizo T560-I). A liquid crystal screen (Nu Vision 17SX Stereoscopic Display Kits) and Polaroid filters presented images to each eye separately. In order to present parallax, we used a bendable head-movement-guide. It was fixed to the arm of a plotter (Roland DPX-2200) whose movement was controlled by the computer. The movement of the guide was horizontal (parallel to the surface of the display), and its range was 6.5 cm. At the center of the movable range, the distance between the guide and display was 45 cm. When viewing other stimulus conditions that did not present parallax, the observer's head was fixed to a chinrest that was at 45 cm just in front of the center of the display.

### 2.3. Stimuli

There were three types of stimuli. The first one was aimed to investigate the interaction between disparity processing and parallax processing (the Disparity with Parallax stimulus). Random dot stereograms (about 16° × 14°) specified a vertical sinusoidal undulation or a flat surface in terms of binocular disparity and/or motion parallax. In order to present motion parallax, the stimuli in the display was yoked to the movement of the head-movement-guide under the observer's chin. The second type of stimuli was similar to the first one; it was aimed to investigate the interaction between disparity processing and motion signal processing without head movement (the Disparity with Motion stimulus). This was a random dot stereogram the same as the one previously described. For this type of stimuli, the movement in the stimulus was independent of the position of observer's head because observers viewed this stimulus with their head fixed on a stationary chinrest. The third type of stimuli was aimed to investigate the interaction between disparity processing and pictorial line configuration cue processing (the Disparity with Monocular Configuration stimulus). This was a line-contoured stereogram (see Fig. 2) which was made of 18 sinusoidally curved lines. The lines were tilted by about 21° (similar to Stevens et al., 1991). In all stimuli, the spatial frequency of the sinusoidal undulation was 0.13 or 0.39 cpd.

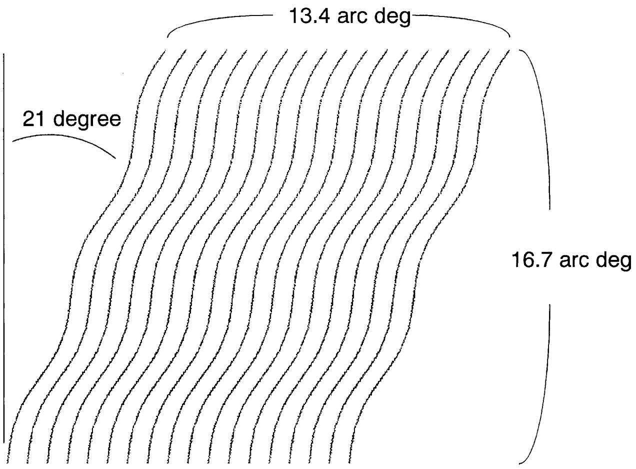

Fig. 2. Diagram of the Disparity with Monocular Configuration stimulus specifying sinusoidal undulations with a spatial frequency of 0.13 cpd, which is convex at center.

For the Disparity with Parallax stimulus and Disparity with Motion stimulus, the dots, each measuring $2.9' \times 23.2'$ of visual arc ($1 \times 8$ pixel), were distributed randomly so that they occupied 40% of the stimulus area. Each dot had a sinusoidal profile of the luminance in the horizontal dimension. The extent of disparity and parallax smaller than one pixel were presented by shifting the phase of the sinusoidal luminance profile of each dot by eight bits of depth by the use of a color lookup table. For the Disparity with Monocular Configuration stimulus, sinusoidal lines were composed by lining up the dots the same as the ones that composed the Disparity with Parallax stimulus and Disparity with Motion stimulus.

For the Disparity with Parallax stimulus, there were five conditions combining the two depth cues (Fig. 3(a)). In the Disparity with zero parallax condition, disparity specified the sinusoidal undulation while parallax specified a flat surface. In the Parallax with a zero disparity condition, parallax specified the sinusoidal undulation while disparity specified a flat surface. In the In-phase condition, the spatial frequency and phase of the sinusoidal undulation specified by disparity were the same as those specified by parallax. In the Out-of-phase condition, the phase of the sinusoidal undulation specified by disparity was opposite to that specified by the parallax cue while their spatial frequency was the same. In the

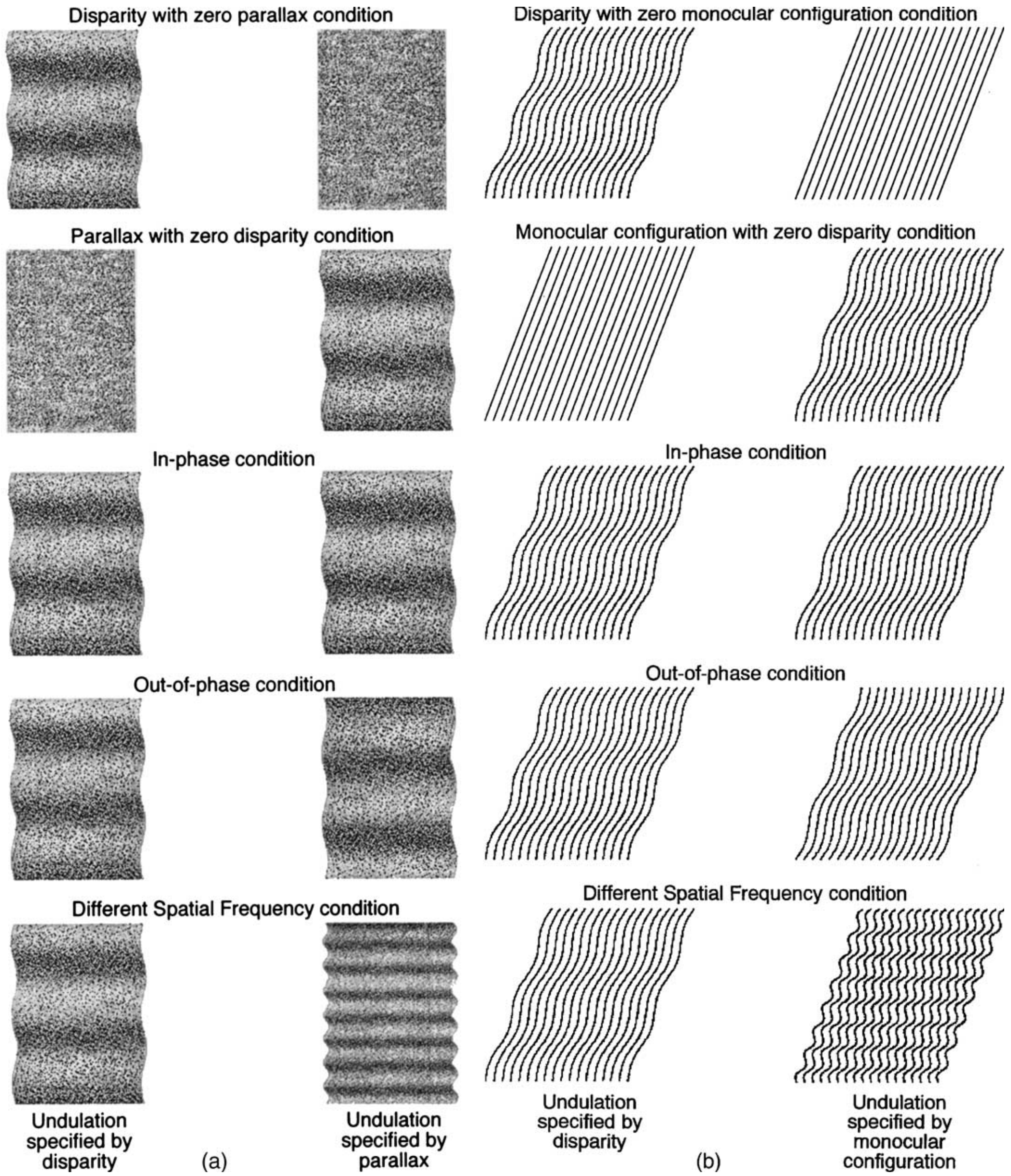

Fig. 3. Diagram of the conditions of cue combination, in which the spatial frequency of the fundamental is 0.13 cpd. (a) Disparity with Parallax stimulus, and (b) Disparity with Monocular configuration stimulus.

Different Spatial Frequency condition, the spatial frequency of the sinusoidal undulation specified by disparity was one-third or three times of that specified by parallax while their phase was the same at the center of the stimulus.

For the Disparity with Motion stimulus, we prepared three conditions to examine whether the visual system integrates disparity with motion signal in an additive way the same as it integrates disparity with parallax. The stimuli for these three conditions were the same as the stimuli for the Disparity with zero parallax condition, Parallax with zero disparity condition, and In-phase condition for the Disparity with Parallax stimulus, except they were observed without head movement. In these conditions, stimulus motion does not specify depth in terms of motion parallax because it is not accompanied by the observer's head movement. In this case, retinal motion is ambiguous in specifying the surface undulation. Therefore, although observers might perceive some undulation instead of a flat moving surface for this stimulus with disparity at subthreshold levels, the perception of the undulation is supposed to be inconsistent and unstable. The first condition, which was the counterpart of the Disparity with zero parallax condition for the Disparity with Parallax stimulus, was called the Disparity condition. The second condition, which was the counterpart of the Parallax with zero disparity condition for the Disparity with Parallax stimulus, was called the Motion condition. The third condition, which was a counterpart of the In-phase condition for the Disparity with Parallax stimulus, and in which disparity and motion signal without head movement specified the sinusoidal undulation, was called the Composite condition. We prepared only three conditions (Disparity, Motion, and Composite conditions) for the Disparity with Motion stimulus because the improvement due to cue combination was most likely in the Composite condition, as demonstrated by previous studies (Bradshaw & Rogers, 1996; Cornilleau-Pérès & Droulez, 1993).

For the Disparity with Monocular Configuration stimulus, there were five conditions that are similar to the ones in the Disparity with Parallax stimulus (Fig. 3(b)). In the Disparity with zero monocular configuration condition, disparity specified the sinusoidal undulation while monocular configuration specified a flat surface. In the Monocular Configuration with zero disparity condition, monocular configuration specified the sinusoidal undulation while disparity specified a flat surface. In the In-phase condition, the spatial frequency and phase of the sinusoidal undulation specified by disparity were the same as those specified by the monocular cue. In the Out-of-phase condition, the phase of the sinusoidal undulation specified by disparity was opposite to that specified by monocular cue while their spatial frequency was the same. In the Different Spatial Frequency condition, the spatial frequency of the sinusoidal undulation specified by disparity was one-third or three times of that specified by the monocular cue while their phase was the same at the center of the stimulus.

In the Disparity with Parallax stimulus and Disparity with Monocular Configuration stimulus, the magnitude of depth specified by each cue was 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, or 1.4 times of the threshold value that was measured in a preliminary test (Table 1). In the preliminary test, the depth threshold in terms of disparity magnitude was measured without head movement. The depth thresholds in terms of magnitudes of parallax and monocular configuration were measured monocularly. The motion threshold in terms of extent of the relative motion between peak and trough were measured monocularly. Table 1 shows that the threshold for monocular configuration was always much larger than that for disparity. Therefore, discrepant corrugation of lines in the retinal images for each eye caused by presenting binocular disparity was supposed to be negligible as a monocular configuration cue. For Observer AO, there were two other additional magnitude conditions (1.6 and 1.8 times of the threshold value) for several stimulus conditions (0.39 cpd of the Parallax, 0.13 cpd of the Disparity, and the Different Spatial Frequency conditions for the Disparity with Parallax stimulus) because 1.4 times of the threshold value was not large enough to generate consistent depth perception in viewing the Different Spatial Frequency condition. In the Disparity with Motion stimulus, the magnitude of depth and motion was 0.2, 0.6, 1.0, or 1.4 times of the threshold value that was measured in a preliminary test.

## 2.4. Procedure

The observation was always binocular for the three types of stimuli. For the Disparity with Parallax stimulus and Disparity with Monocular Configuration stimulus, the presentation of each stimulus condition was divided into 28 blocks. For the Disparity with Motion stimulus, the presentation of each stimulus condition was divided into 16 blocks. In a given block, one stimulus condition was presented 40 times. In half of the trials, the stimuli specified an undulation. In the other half, the stimuli specified a flat surface. The order of the depth magnitude and phase condition, the order

Table 1
Thresholds measured in a preliminary test (arc sec)

| Observer | Spatial frequency (cpd) | Disparity | Parallax | Monocular configuration | Motion without head movement |
|---|---|---|---|---|---|
| MI | 0.13 | 6.5 | 20.8 | 12.3 | 13.1 |
| | 0.39 | 6.6 | 22.8 | 4.5 | 10.6 |
| AO | 0.13 | 10.5 | 36.7 | 11.0 | 25.4 |
| | 0.39 | | 25.0 | 8.9 | |
| KM | 0.13 | 6.8 | 38.3 | 25.5 | 15.5 |
| | 0.39 | | 37.1 | 10.5 | |
| AM | 0.13 | 18.8 | 46.1 | 4.3 | 30.5 |
| | 0.39 | | 38.4 | 4.1 | |
| MO | 0.13 | 9.5 | 38.5 | 7.0 | 36.8 |
| | 0.39 | | 44.5 | 8.2 | |

*Probit analysis* determined the 50% thresholds in terms of the amplitude between peaks and troughs in the sinusoidal undulation. The magnitude of parallax and motion are described in terms of equivalent disparity (Equivalent disparity is the unit used to describe the magnitude of motion parallax, introduced by Rogers and Graham (1982). It is defined by the visual angle of the retinal motion caused by the lateral head movement whose distance is the same as the interocular distance.).

of the flat and undulating surface presentation in a block, and the order of the stimulus condition for each observer were random.

In viewing the Disparity with Parallax stimulus, observers moved their heads by following the head movement guide. The movement on the display was yoked to the guide. In each trial for the stimuli, the observation was restricted to about 3.5 s. About 0.5 s after the beginning of stimulus presentation, the head movement guide moved to the right for about 0.5 s, and stayed at the end of the movable range for 1.5 s. Then, it moved to the left for about 0.5 s. About 0.5 s after the guide reached to the end of movable range (start point), stimulus disappeared. The head movement guide moved at constant velocity of 12 cm/s in which the magnitude of parallax has been shown to determine the sensitivity of depth perception from parallax (Ono & Ujike, 1993). About 0.1 s before the head movement guide moved, a short beep signalled observers the start of the motion of head movement guide and stimulus. In viewing the Disparity with Motion stimulus, the stimulus was the same as the Disparity with Parallax stimulus, except that the magnitude of motion was determined in terms of motion threshold measured in the preliminary test. About 0.5 s after the beginning of stimulus presentation, the stimulus moved for 0.5 s without head movement, and after 1.5 s of interval, the stimulus moved for 0.5 s in the opposite direction to the first one. About 0.1 s before the stimulus movement, a short beep for the observers signalled the start of the stimulus motion, the same as for the Disparity with Parallax stimulus. In viewing the Disparity with Monocular Configuration stimulus, the duration of stimulus presentation was about 3.5 s (the same as in the Disparity with Parallax stimulus).

In each trial for the Disparity with Parallax stimulus and Disparity with Monocular Configuration stimulus, after the stimulus disappeared, observers judged whether they saw sinusoidal undulations or a flat surface in the stereogram by pressing keys of a keyboard. In each trial for the Disparity with Motion stimulus, after the stimulus disappeared, observers judged whether they saw a stationary and flat surface or sinusoidal modulation specified by depth or surface flow by pressing keys of a keyboard. For all conditions, the judgment was defined as correct when observers judged that they saw sinusoidal undulations for the stimuli which specified sinusoidal undulations, and also when they judged that they saw a flat surface for the stimuli which specified a flat surface. Observers were given feedback on false alarms only, that is, when an observer mistakenly judged that there was an undulation in the stimuli that specified a flat surface.

Observer MI viewed all combinations of spatial frequencies for the In-phase, Out-of-phase, and Different Spatial Frequency conditions while the other four observers viewed only the conditions in which disparity specified the undulation with a spatial frequency of 0.13 cpd.

## 3. Results

In experimental sessions, every observer reported that they could not distinguish the depth perception generated by the two cues (disparity and parallax, or disparity and monocular configuration) from perception generated by only one of the cues. The performance in the depth detection task of the observers, however, varies regarding the numbers of the cues that specified the sinusoidal undulation in the stimulus. Fig. 4 shows the $z$ score (standard score) from the *percent correct* (summation of *Hit* and *Correct rejection* rates) for each condition. For the Disparity with Parallax stimuli (Fig. 4(a)), the $z$ scores tended to be higher for the In-phase condition than those of both the Disparity and Parallax conditions. For the Disparity with Monocular Configuration stimulus (Fig. 4(b)), the $z$ scores of the In-phase condition were not always higher than those of the Disparity and Parallax conditions. These imply that the extent of improvement in depth perception due to cue combination varies with the type of combined cues. For the Disparity with Motion stimuli (Fig. 4(c)), the $z$ score for the Composite condition was not always higher than those of the Disparity and Motion conditions.

In order to evaluate how presenting multiple cues improve depth perception, we analyzed the responses in each condition by the use of the $d'$ value. As mentioned in Section 1, if the visual system linearly integrates the depth information from the two cues served by independent processing, the improvement of depth perception could be predicted by probability summation. Or, if the processing of disparity and that of other cues improved each other in a non-linear manner, the performance could be better than the prediction based on probability summation.

We obtained $d'$ by the use of two cues ($d'_{\text{combined}}$) from five observers for the condition presenting undulation, that is, for each magnitude condition in the In-phase, Out-of-phase, and Different Spatial Frequency conditions of the Disparity with Parallax stimuli and Disparity with Monocular Configuration stimuli. In order to examine whether the visual system integrated the depth information from different cues in a linear or in a non-linear manner, we plotted the obtained $d'_{\text{combined}}$ against predicted values of $d'_{\text{combined}}$ with two different predictions (Green & Swets, 1966). The abscissa of the left column of Fig. 5 is the value of $d'_{\text{combined}}$ predicted by the model of linear integration, that is, probability summation. In this case, the $d'_{\text{combined}}$ is predicted by square-root improvement of $d'$ for the condition presenting undulation by the use of only one cue

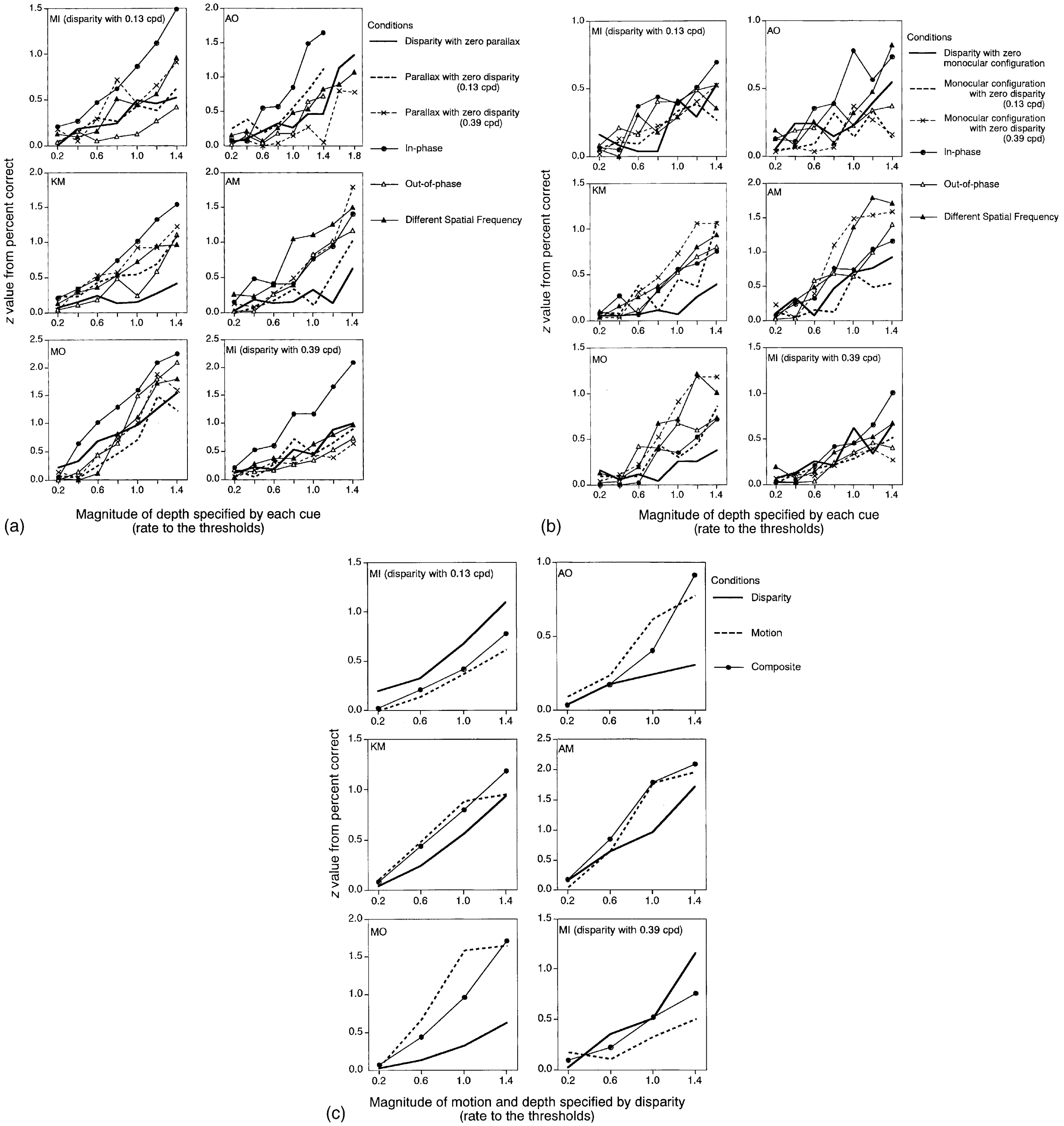

Fig. 4. Plot of $z$ score for each condition from five individuals. (a) Disparity with Motion Parallax stimulus, (b) Disparity with Monocular Configuration stimulus, and (c) Disparity with Motion stimulus.

($d'_{\text{cue1}}, d'_{\text{cue2}}$); the integration of the two cues based upon probability summation is described as $d'_{\text{combined}} = \sqrt{(d'_{\text{cue1}})^2 + (d'_{\text{cue2}})^2}$. The abscissa of the right column of Fig. 5 is the value of $d'_{\text{combined}}$ predicted by the model of additive integration of two cues. This prediction is based on the assumption that the processing of a cue is improved by the processing of another cue, and that the performance of the visual system is better than the probability summation of independent modules. This simple, ad hoc addition was introduced by Green and Swets to assess the lack of independence between two processing, that is, non-linear processing. Although the equation was not based on any exact theoretical model of cue integration, we show the value derived from the equation as a tentative prediction in terms of the non-linear integration, which causes a better performance in depth perception than the performance predicted based

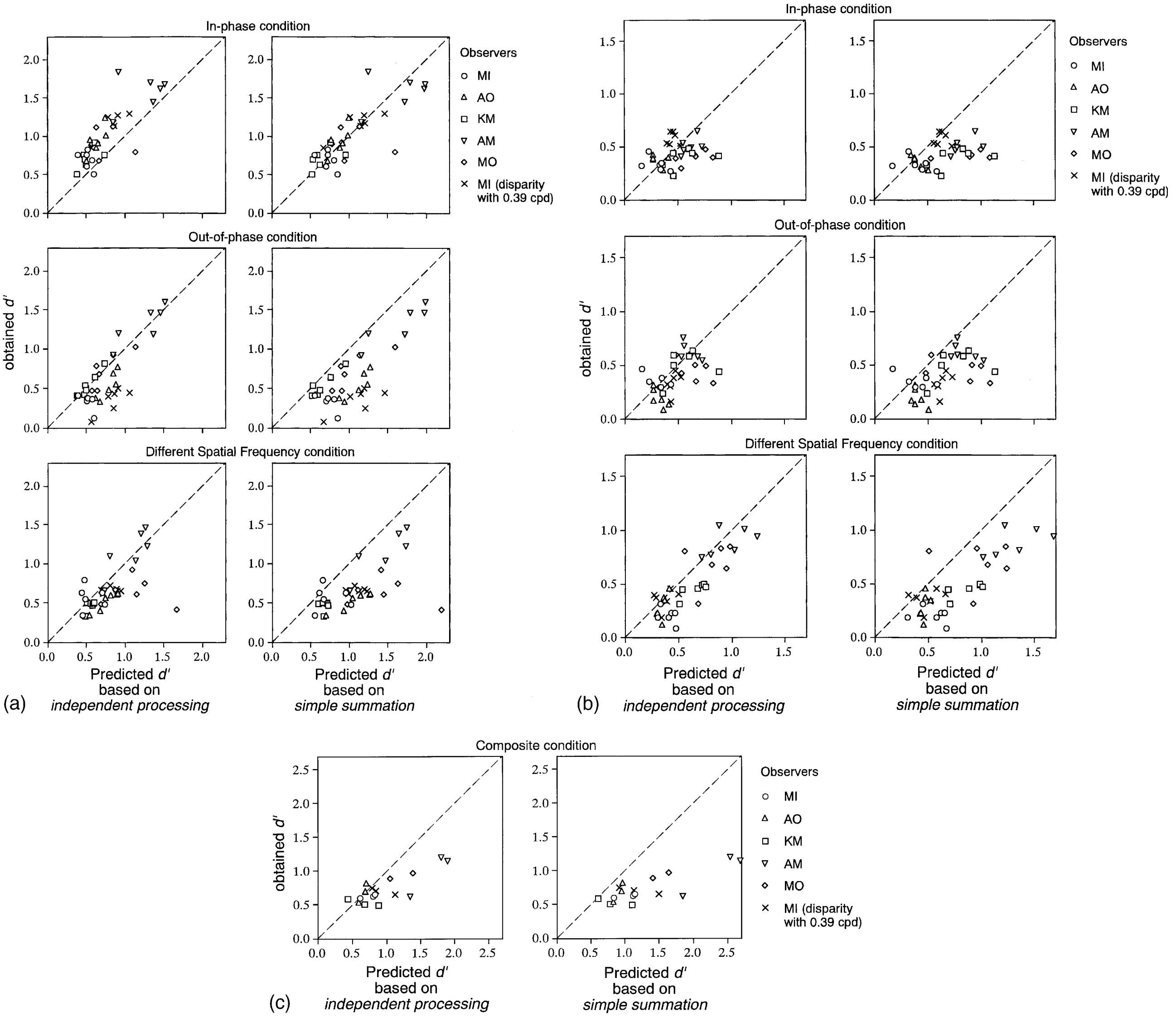

Fig. 5. Obtained and predicted $d'$ for five individuals. (a) $d'$ for the Disparity with Parallax stimulus. (b) $d'$ for the Disparity with Monocular Configuration stimulus. (c) $d'$ for the Disparity with Motion stimulus for five individuals. In each panel, the ordinate shows the predicted value based on the linear integration in the left column, and the predicted value based on the non-linear, additive integration in the right column. If obtained $d'$ fits to the predicted value, data would be on the orthogonal broken line.

on probability summation. For this case, we evaluate the $d'_{\text{combined}}$ based upon simple summation of $d'_{\text{cue1}}$ and $d'_{\text{cue2}}$; the integration of the two cues is described as $d'_{\text{combined}} = d'_{\text{cue1}} + d'_{\text{cue2}}$. If the obtained $d'_{\text{combined}}$ was identical to predicted $d'_{\text{combined}}$, the plotted data would be on the broken diagonal line in each panel. To determine which of the predictions was true for each condition, we counted the data point above and below the diagonal line in each panel as did Green and Swets (1966).

Fig. 5(a) shows that, in the In-phase condition of the Disparity with Parallax stimulus, nearly half of the 36 plotted data (15 points) are below the diagonal line in the right panel while only two points were below the diagonal line in the left panel. That is, in the In-phase condition of the Disparity with Parallax stimuli, the model of probability summation tends to underestimate the actual performance; the obtained $d'_{\text{combined}}$ was better fitted in the right panel than in the left panel. This suggests that the processing of disparity and that of parallax improved each other in the In-phase condition. This result is compatible with the results of Bradshaw and Rogers (1996). In other conditions of the Disparity with Parallax stimulus, the data points which were located above the diagonal line were much fewer than half of the 36 data points; 12 points for the Out-of-phase condition, and seven points for the Different Spatial Frequency condition. These indicate the model of probability summation overestimate the actual performance for these conditions.

For the Disparity with Monocular Configuration stimulus (Fig. 5(b)), in the right panel, the number of the plotted data above the diagonal line, was at most, seven

(In-phase condition). In the left panels, the data points plotted above the diagonal line were less than the half of the 36 data points; 10 for the In-phase condition, 12 for Out-of-phase condition, and eight for Different Spatial Frequency condition. These indicate that both models of probability summation and additive integration tend to overestimate the actual performance for the Disparity with Monocular Configuration stimulus.

We obtained $d'$ for the Disparity with Motion stimulus in order to examine whether the improving interaction shown in the In-phase condition of Fig. 5(a) is based on the integration of disparity with motion signal, or on the integration of the two depth cues, disparity and parallax. In Fig. 5(c), we plotted the obtained $d'_{\text{combined}}$ against predicted values of $d'_{\text{combined}}$ with the two predictions that were the same as in Fig. 5(a). Fig. 5(c) shows that both models of probability summation and additive integration overestimate the actual performance of the visual system; in the left panel, the data points above the diagonal line were three and those below the diagonal line were 15, while, in the right panel, all data points were below the diagonal line. These results indicate that the improving interaction shown in the In-phase condition of Fig. 5(a) was based on the integration of depth information from disparity and parallax, not on the integration of disparity and motion signal, which is ambiguous in specifying an object's depth. Unambiguous information about surface undulation would be necessary for retinal motion to be integrated with disparity to improve depth perception. Ambiguous information about surface undulation cannot improve depth perception in terms of integration with disparity.

## 4. General discussion

Our $d'$ analysis, only in the In-phase condition of the Disparity with Parallax stimulus, showed that the processing of multiple cues improve each other in a nonlinear manner. The observers' performances for the other conditions, including the In-phase condition of the Disparity with Monocular Configuration stimulus, remained below the level predicted by probability summation. These suggest that only when disparity was combined with consistent parallax, could the integration be conducted as a strong fusion (Fig. 1(b)), and that in other conditions, the integration is conducted as a weak fusion (Fig. 1(a)).

The dependency of the strong fusion on the combined cues suggests that the visual system integrates depth information from different cues in different ways. It also indicates that the disparity processing would be more closely linked with parallax processing than with processing of pictorial cues including monocular configuration. This notion is compatible with the findings of functional brain researches. On one hand, research on monkeys has found that binocular disparity is coded in the middle temporal visual area (MT/V5) (Bradley, Qian, & Andersen, 1995; DeAngelis & Newsome, 1999; Maunsell & van Essen, 1983) and the medial superior temporal visual area (MST) (Roy, Komatsu, & Wurtz, 1992). These areas are known to be involved in motion perception (e.g. for MT, Albright, Desimone, & Gross, 1984, for MST, Komatsu & Wurtz, 1988). These imply that the disparity processing and motion processing are conducted in brain sites that are very near each other, and that they could interact with each other at a very early stage after depth signal detection. On the other hand, human brain research using fMRI have found that the processing of pictorial depth cues is accompanied by the excitation of other area, for example, the lateral occipital complex (LOC) for contours, shading and pictorial cues (Kourtzi & Kanwisher, 2000), and V1 for shading (Humphrey et al., 1997). These sites are apart from MT and MST that are supposed to be involved in disparity and motion processing. We assume, therefore, that interaction of disparity processing and parallax processing might be conducted in an earlier stage than the one in which disparity processing interacts with pictorial cue processing.

Combining disparity with motion (without head movement) did not show improvement beyond what probability summation would permit in depth perception in $d'$. This indicates that the prominent improvement in depth perception shown in the In-phase condition of the Disparity with Parallax stimulus was not a consequence of the interaction of disparity and motion signal even though both of them are supposed to be processed in very near brain sites. We propose that the strong fusion of disparity with parallax is conducted after the visual system converts the motion signal into the consistent depth signal. This integration as a strong fusion should be based on the communication in terms of the depth token. Roy et al. (1992) considered that MST would be involved in self-motion perception because it reacts to the motion of the whole visual field. MST could convert the motion information into depth signal in terms of self-motion information, and it also might be involved in the integration of depth information from disparity and motion depth cues. Their proposal is compatible with our present data.

We found that the improvement in depth perception due to cue combination depends upon the consistency in the information specified by these cues. These results are compatible with our previous study (Ichikawa & Saida, 1998). In the previous study, we compared the depth threshold for the stimuli in which the disparity specifying a sinusoidal undulation was fixed at a subthreshold level and combined with variable parallax that specified a sinusoidal undulation with different spatial frequencies and phases. In the present study, the dependency of improvement in depth perception on the consistency in the spatial frequency of the undulation specified by each cue suggests that, before integrating depth information from different cues, the visual system assesses the congruence of the different cues. Previous studies have assumed that a channel-like mechanism which is tuned to the spatial frequency of the undulation underlies the disparity processing. For example, the studies using a masking method (Cobo-Lewis & Yeh, 1994; Tyler & Julesz, 1978) and a selective adaptation method (Schumer & Ganz, 1979) have demonstrated the presence of multiple spatial frequency-tuned channels underlying the processing of binocular disparity that specify sinusoidal undulation. Rogers and Graham (1982) demonstrated that stereopsis based on parallax has very similar sensitivity functions to that based on disparity for a given spatial frequency of surface undulation. In accordance with this finding, Rogers and Graham (1982) claimed that stereopsis based on parallax and that based on disparity are founded on a similar mechanism with regard to the spatial frequency of surface undulation. The spatial frequency-tuned channel could be a candidate for processing which determines the consistency of the information from disparity and parallax, and which integrates them in an additive manner (see Fig. 6).

At the suprathreshold level, the additive integration is not restricted to the in-phase combination of disparity and parallax. That is, the visual system can combine depth information from disparity and parallax even when these cues specify the sinusoidal undulation with different spatial frequency (Ichikawa & Saida, 2002; Rogers & Collett, 1989), and when the phase of sinusoidal undulation specified by one of the cues were counter phase to that specified by the other cue (Ichikawa & Saida, 1998). Moreover, the visual system can combine depth information from disparity and that from monocular configuration cue (Stevens & Brookes, 1988; Stevens et al., 1991). The results of these previous studies at suprathreshold level seem to be different from the present results on integration of depth information at near threshold level. To understand why there are differences between the processing at near threshold and at suprathreshold levels in the cue integration, we should notice the difference in the task for the visual system at each level. On the one hand, at near threshold level, the task of the visual system is depth detection. In this task, the consistency among cues should be important because, to avoid false detection, the visual system must distinguish the depth signal not only from internal noise, but also external noise derived from multiple cues. Therefore, the visual system should take into account the cue consistency, in the decision to integrate the cues. On the other hand, at suprathreshold level, it pertains to the task of completion of the surface shape representation. In this task, the consistency among cues might be less important because the depth information from each cue is sufficiently larger than the noise level. In order to complete the elaborate representation of the undulating surface, the visual system would not worry about false detection. Then, the visual system would use information from any available cues in an additive manner, regardless of the consistency of the surface shape specified by those cues. The visual system would integrate the depth information from different cues in a different way, and maybe at different stages, depending upon the combined cues and upon the stimulus level.

Finally, we summarized what we found in the present study as Fig. 6. It shows that the type of data fusion (strong or weak) in depth perception depends on the

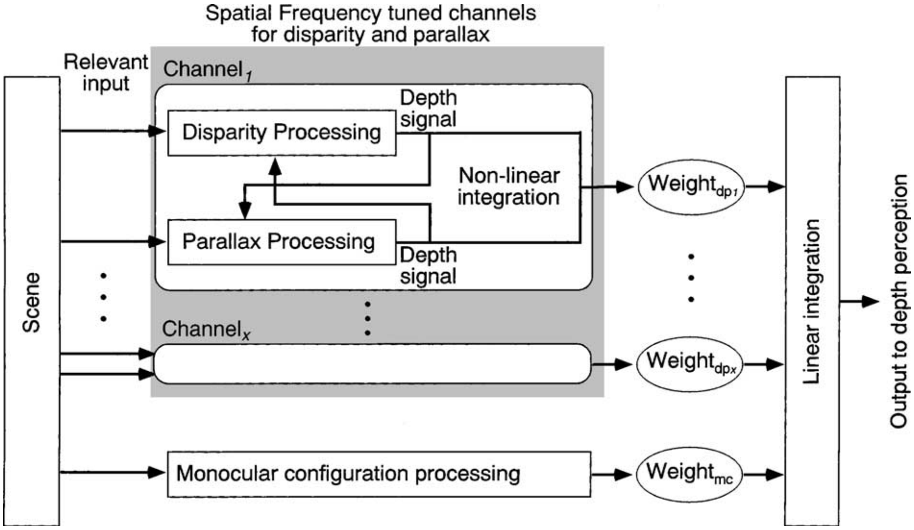

Fig. 6. Diagram of the fusion models for disparity, parallax, and monocular configuration. The inputs from disparity processing and parallax processing are integrated in a non-linear manner after their signals are converted into depth signal within each spatial frequency-tuned channel. The outputs of each channel and the module for monocular configuration are averaged (linearly integrated) with the weights derived from measurements of the relative reliabilities of each source. The sum of the weights for multiple cues is assumed to be one, as in the weak fusion model.

combined source and consistency among the sources. The integration of output from the processing for disparity and parallax with the output from the monocular configuration processing is conducted as a weighted addition (integrated in a linear manner), as the weak fusion model proposed. The integration of disparity with parallax would be conducted by spatial frequency-tuned channel-like processing that considers the consistency between the undulations specified by disparity and parallax. When the spatial frequency of the undulation specified by disparity is the same as that specified by parallax, the inputs from disparity processing and parallax processing are integrated so that they improve each other (integrated in a non-linear manner), as the strong fusion model proposed. This integration is conducted after the retinal motion signal is converted into consistent depth signal, and it depends on the acquisition of unambiguous depth information from motion. In contrast, when the spatial frequency of the undulation specified by disparity is different from that specified by parallax, the output of each channel is averaged (linearly integrated) with the weights derived from measures of the relative reliabilities of each channel, as the weak fusion model proposed.

## Acknowledgements

A preliminary report on this research was presented at the annual meeting of the Association for Research in Vision and Ophthalmology, Fort Lauderdale, FL, May, 2000. We thank Dr. Jennifer Steeves, Dr. Philip Grove, Dr. Hiroshi Ono, and Dr. Michael Higgins for their helpful comments and suggestions on an earlier version of this article.

## References

Albright, T. D., Desimone, R., & Gross, C. G. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. *Journal of Neurophysiology, 51*, 16–31.

Bradley, D. C., Qian, N., & Andersen, R. A. (1995). Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature, 373*, 609–611.

Bradshaw, M. F., & Rogers, B. J. (1996). The interaction of binocular disparity and motion parallax in the computation of depth. *Vision Research, 36*, 3457–3468.

Clark, J. J., & Yuille, A. L. (1990). *Data fusion for sensory information processing system*. Boston, MA: Kluwer.

Cobo-Lewis, A. B., & Yeh, Y. (1994). Selectivity of cyclopean masking for the spatial frequency of binocular disparity modulation. *Vision Research, 34*, 607–620.

Cornilleau-Pérès, V., & Droulez, J. (1993). Stereomotion cooperation and the use of motion disparity in the visual perception of 3-D structure. *Perception & Psychophysics, 54*, 223–239.

DeAngelis, G. C., & Newsome, W. (1999). Organization of disparity selective neurons in macaque area MT. *Journal of Neuroscience, 19*, 1298–1415.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Los Altos, CA: Peninsula.

Humphrey, G. K., Goodale, M. A., Bowen, C. V., Gati, J. S., Vilis, T., Rutt, B. K., & Menon, R. S. (1997). Differences in perceived shape from shading correlate with activity in early visual areas. *Current Biology, 7*, 144–147.

Ichikawa, M., & Saida, S. (1998). Integration of motion and binocular disparity depth cues at near depth threshold level. *Japanese Journal of Psychonomic Science, 17*, 1–11.

Ichikawa, M., & Saida, S. (2002). Integration of motion parallax with binocular disparity specifying different surface shapes. *Japanese Psychological Research, 44*, 34–44.

Komatsu, H., & Wurtz, R. H. (1988). Relation of cortical areas MT and MST to pursuit eye movements. I. Localization and visual properties of neurons. *Journal of Neurophysiology, 60*, 580–603.

Kourtzi, Z., & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *Journal of Neuroscience, 20*, 3310–3318.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research, 35*, 389–412.

Maunsell, J. H. R., & van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology, 49*, 1148–1167.

Ono, H., & Ujike, H. (1993). Zone in which motion parallax is completely effective. *Investigative Ophthalmology & Visual Science, 34*, 1052.

Rivest, J., & Cavanagh, P. (1996). Localizing contours defined by more than one attribute. *Vision Research, 36*, 53–66.

Rogers, B. J., & Collett, T. S. (1989). The appearance of surfaces specified by motion parallax and binocular disparity. *The Quarterly Journal of Experimental Psychology, 41*, 697–717.

Rogers, B. J., & Graham, M. E. (1979). Motion parallax as an independent cue for depth perception. *Perception, 8*, 125–134.

Rogers, B. J., & Graham, M. E. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research, 22*, 261–270.

Roy, J.-P., Komatsu, H., & Wurtz, R. H. (1992). Disparity sensitivity of neurons in monkey extrastriate area MST. *Journal of Neuroscience, 12*, 2478–2492.

Schumer, R. A., & Ganz, L. (1979). Independent stereoscopic channels for different extents of spatial pooling. *Vision Research, 19*, 1303–1314.

Stevens, K. A., & Brookes, A. (1988). Integrating stereopsis with monocular interpretations of planar surfaces. *Vision Research, 28*, 371–386.

Stevens, K. A., Lees, M., & Brookes, A. (1991). Combining binocular and monocular curvature features. *Perception, 20*, 425–440.

Tyler, C. W., & Julesz, B. (1978). Spatial frequency tuning for disparity gratings in the cyclopean retina. *Journal of the Optical Society of America, 68*, 1365.

Uomori, K., & Nishida, S. (1994). The dynamics of the visual system in combining conflicting KDE and binocular stereopsis cues. *Perception & Psychophysics, 55*, 526–536.