Vision Research 172 (2020) 11–26

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

A motion-in-depth model based on inter-ocular velocity to estimate direction in depth

Wei Wu[a], Yasuhiro Hatori[a,b], Chia-huei Tseng[a,b], Kazumichi Matsumiya[c], Ichiro Kuriki[a,b], Satoshi Shioiri[a,b,*]

[a] Graduate School of Information Science, Tohoku University, Sendai, Japan
[b] Research Institute of Electrical Communication, Tohoku University, Sendai, Japan
[c] Department of Applied Information Sciences, Tohoku University, Sendai, Japan

ARTICLE INFO

Keywords:
Model
Motion-in-depth
Interocular velocity difference
Motion channels
Motion direction perception

ABSTRACT

Perception of motion in depth is one of the most important visual functions for living in the three-dimensional world. Two binocular cues have been investigated for motion in depth: inter-ocular velocity difference (IOVD) and changing disparity (CD). IOVD provides direction information directly by comparing velocity signals from the two retinas. In this study, we propose for the first time a motion-in-depth model of IOVD that predicts motion-in-depth direction. The model is based on a psychophysical assumption that there are four channels tuned to different directions in depth (Journal of Physiology 235 (1973) 17–29). We modeled these channels by combining outputs of low-level motion detectors that are sensitive to left and right retinal stimulation. Using these channels, we constructed a model of motion in depth that successfully predicted a variety of psychophysical results including direction discrimination, perceived direction, spatial frequency tuning, effect of speed on rotation in depth, effect of lateral motion direction, and effect of binocular and temporal correlations.

1. Introduction

1.1. Available cues for MID perception

Depth perception is crucial to living and thriving in the three-dimensional (3D) world. Perceiving directions of moving objects in depth is important, for example, for catching a ball approaching the head or avoiding being hit by an approaching rock. There is no doubt of importance of investigation on the mechanism of motion in depth (MID). However, there are fewer studies on motion in depth than on static depth perception (Regan & Spekreijse, 1970; Horward & Rogers, 2002; Shioiri, Morinaga, & Yaguchi, 2002; Tidbury, Brooks, O'Connor, & Wuerger, 2016) in general. Moreover, while several models have been proposed to predict sensitivity to MID (Baker & Bair, 2016; Peng & Shi, 2014; Sabatini & Solari, 2004; Shioiri, Kakehi, Tashiro, & Yaguchi, 2009), there is no proposal of models that predict MID direction to our best knowledge.

There are three major cues for MID perception: inter-ocular velocity difference (IOVD), changing disparity (CD) over time, and changing size of retinal images (CS) (Fig. 1). IOVD is the difference in the velocity of the retinal images between the two eyes. CD refers to changes in binocular disparity over time: the disparity of 0 indicates the location of fixation, a change from uncrossed disparity to crossed disparity indicates an approaching motion and a change from crossed to uncrossed indicates a receding motion. CS is changes in retinal image size over time: retinal size increases as an object approaches and decreases as it recedes.

Among these three cues, IOVD can be used for estimating MID direction without further information. MID direction can be estimated directly by comparing left and right monocular motion signals through the process of IOVD calculation following a geometrical relationship. When an object is moving toward the middle of the eyes, the images of the object on the two retinas move in opposite directions at the same speed. When the object is moving laterally, motion directions are virtually identical on the two retinas. When the object is moving toward the head along the line of sight of left eye, the left retinal image remains stationary and the right retina receives motion signals (Fig. 1). This differs from the other two cues. For both CD and CS, lateral motion speeds are required to estimate MID direction in addition to the velocity along the depth axis. To enrich our understanding of how MID direction perception is achieved, we selected IOVD as the main focus because of the following reason. IOVD is better understood as a cue to extract direction in depth comparing with other two cues (i.e. CD and CS). Although MID motion detection can be achieved by each of the three

* Corresponding author at: Research Institute of Electrical Communication, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577 Japan.
E-mail address: shioiri@riec.tohoku.ac.jp (S. Shioiri).

https://doi.org/10.1016/j.visres.2020.04.002
Received 21 October 2019; Received in revised form 7 April 2020; Accepted 11 April 2020

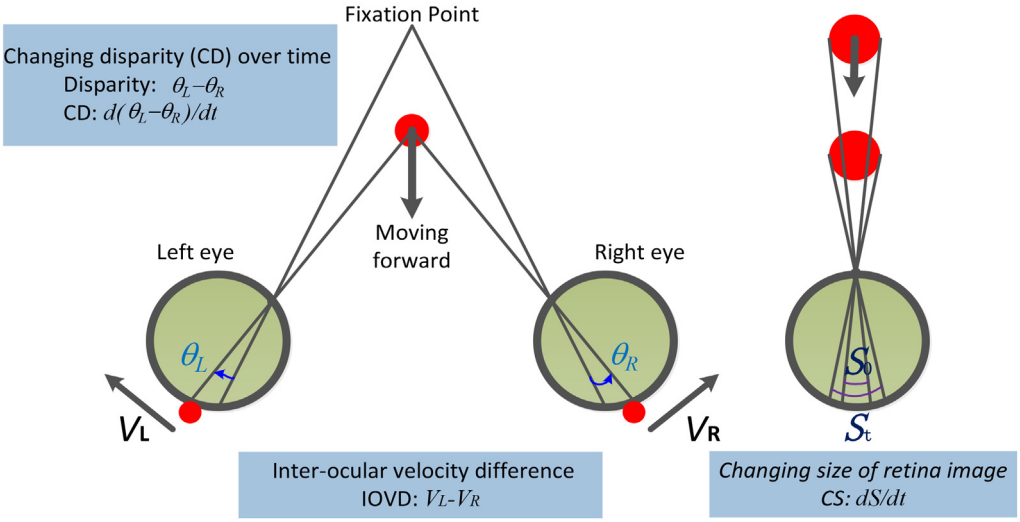


Fig. 1. Three possible visual cues for MID of a moving object: inter-ocular velocity difference (IOVD), changing disparity and changing size (CS).

cues alone (Erkelens & Regan, 1986; Gray & Regan, 1998; Shioiri et al., 2002), their separate contributions to MID direction are less well known. To date, more systematic investigations on IOVD cues are available than CD and CS cues, which prompts us to select it as the first step to formalize the model of MID direction perception. If the IOVD-based model does not well predict available psychophysical experiment results, it also provides important insights that other cues may be more crucial. Thus, we focus on IOVD in this study and aim to build an IOVD model that predicts MID direction.

1.2. Mechanism of IOVD cue for MID perception

IOVD has been identified psychophysically as a visual mechanism for analyzing MID (Shioiri, Saisho, & Yaguchi, 2000; Brooks, 2002a, 2002b; Fernandez & Farell, 2005; Rokers, Cormack, & Huk, 2008) while the relative contributions of IOVD and CD are still under debate (Cumming & Parker, 1997; Harris, Neefs, & Grafton, 2008; Shioiri, Saisho, & Yaguchi, 2000; Watanabe, Yasuoka, & Fujita, 2008; Shioiri, Matsumiya, & Matsubara, 2012; Sakano, Allison, & Horward, 2012; Kaestner, Maloney, Wailes-Newson, Bloj, & Harris, 2019). Physiological studies have recently revealed neurons sensitive to opposite directions between the two retinas in the middle temporal (MT) visual area of monkeys (Czuba, Rokers, Huk, Cormack, & Kohn, 2014; Sanada & DeAngelis, 2014). This is consistent with an fMRI study (Rokers, Cormark, & Huk, 2009) where the human MT was found to be responsible for MID perception. The neurons in the MT can be considered to calculate IOVD and have been suggested to be tuned to MID direction according to a comparison of their left and right retinal sensitivities.

Modeling is one useful method for investigating underlying visual mechanisms and for implementing applications such as 3D display design, 3D content creation and visual environment estimation. In this report, we propose a motion-in-depth model that predicts perception of MID direction as well as sensitivity to MID and apply it to predicting several psychophysical results related to perception of MID direction. Several MID models have been reported in the literature (Baker & Bair, 2016; Peng & Shi, 2014; Sabatini & Solari, 2004; Shioiri et al., 2009). Sabatini and Solari (2004) simulated MID sensitivity of cortical cells with combinations of responses from complex disparity energy units related to unbalanced ocular dominance and demonstrated that IOVD can be detected by the same mechanism sensitive to changing disparity. However, an additional lateral motion signal is required to estimate MID direction. Shioiri et al. (2009) and Peng and Shi (2014) proposed MID models based on IOVD and CD to estimate direction or speed judgments of motion perception along the depth axis. Baker and Bair (2016) proposed an MID model based on IOVD to simulate the activities of MT neurons, responding to motion also along the depth axis. However, none of previous studies attempted to predict MID direction except for Welchman, Lam, and Bülthoff (2008). They predicted perceived MID direction based on a Bayesian model with different reliabilities for horizontal motion and depth motion. Their model, however, is descriptive and based on assumptions of statistical distributions without assumption of the corresponding physiological process, which differs from the approach in the present study. Here, we propose a perception model of MID direction based on IOVD and predict the results of several psychophysical experiments of MID direction perception.

Electro-Physiological studies of monkey cortex have suggested that the basic motion selectivity of MT neurons is most plausibly inherited from V1. More than 90% of V1 neurons are direction selective (Zeki, 1974), and many of these direction-selectivity neurons project to area MT (Movshon & Newsome, 1996; Anderson, Binzegger, Martin, & Rockland, 1998; Sincich & Horton, 2003). MT receives the majority of its direct input from V1 layer 4 (Shipp & Zeki, 1989; Born & Bradley, 2005), where many of neurons are monocular (Hubel & Wiesel, 1968; Blasdel & Fitzpatrick, 1984). Many motion selective MT neurons are created by combinations of V1 neurons with different selectivities (Churchland, Priebe, & Lisberger, 2005). In our model, motion-in-depth detector is constructed from lateral motion signals from V1 motion sensitive neurons. Recent studies reported that many neurons in MT encode information about 3D motion direction on the basis of binocular cues (Czuba et al., 2014; Sanada & DeAngelis, 2014). Sanada and DeAngelis (2014) reported, for example, that in primate MT, 56% of 170 recorded neurons responded to the IOVD- based stimuli and 10% responded to the CD- isolating one. These studies suggest that the MID process in MT combined signals from V1 neurons sensitive to monocular motion.

1.3. Direction tuning of MID

Only a few experimental studies have investigated direction tuning of MID. Beverley and Regan (1973) reported experimental results that suggest four channels of MID directions. They used motion adaptation to isolate underlying MID channels with different direction tunings. After showing participants a bar moving in depth along one direction for 10 min, sensitivity was measured by adjusting peak-to-peak oscillations of disparity for the moving bar. The sensitivity loss after adaptation was analyzed for each of 13 motion directions. The threshold elevation curve against MID direction approximates the MID direction tuning of the underlying physiological channel sensitive to the

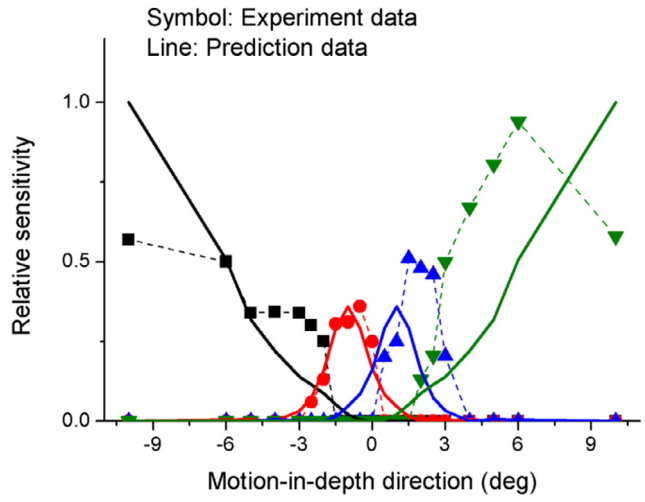


Fig. 2. Four MID channels estimated by direction adaptation experiment (dashed lines with symbols; Beverley & Regan, 1973) and sensitivity of model channels (solid lines).

adaptation direction. Measurements were repeated after several different adaptation directions revealed four MID channels tuned to different MID directions (Fig. 2). Following their report, we modeled a MID mechanism with four channels, using IOVDs; that is, the difference between lateral motion detectors (LMDs) for left and right retinal images.

1.3.1. Model for predicting perception of MID direction

The proposed model is the first to predict perception of MID direction and provides a plausible connection between human behaviors and neural activities in the visual cortex. In the literature, there are at least four MID models based on IOVD (Baker & Bair, 2016; Peng & Shi, 2014; Sabatini & Solari, 2004; Shioiri et al., 2009; Welchman et al., 2008). All of these studies used monocular motion signals at the first stage. Although the present model also uses left and right monocular motion signals, it combines monocular motion signals with different speed tunings to construct four MID channels with different MID sensitivities, as reported in Beverley and Regan (1973). Fig. 3 summarizes the structure of the model. We chose certain combinations of four monocular motion detectors (left and right directions with two different speed tunings, 0.5 deg/s and 1 deg/s, respectively), in line with the report by Beverley and Regan (1973), Regan and Cynader (1982) (Fig. 3) at the second stage. The outputs of these lateral motion detectors are combined to construct four MID channels sensitive to approaching motion. There are also four MID channels sensitive to receding motions, which combine outputs of different sets of lateral motion directors from those used for approaching channels. To predict MID directions, a calculation to combine the outputs of these channels is required. This is a typical problem for other feature identifications, such as size perception (Gelb & Wilson, 1983), temporal frequency perception (Mandler & Makous, 1984), and lateral motion speed (Smith & Edgar, 1993). Among several types of calculations used for integrating channel outputs, here, we used the output difference between adjacent channels to estimate perceived direction, as for coding color through color opponent processes at the last stage. Details of each stage are described in the following sections.

At the first stage, the lateral motion detectors analyze monocular motion signals by referring to primary visual cortex (V1) neurons. At the second stage, the monocular motion signals are combined to construct four MID detectors by referring to middle temporal (MT) neurons. At the last stage (decision stage), the channels' responses are combined to estimate the direction. Here, these four channels are responsible for approaching motions.

1.4. Lateral motion detector (LMD)

Spatiotemporal filters (ST-filters) have been used to model low-level LMDs based on a motion energy model (Adelson & Bergen, 1985; Watson & Ahumada, 1985). In the present model, we used Gabor functions as ST-filters. The filter output models the response of motion-selective complex cells in the visual cortex and the function is known to approximate the spatiotemporal receptive field of direction-sensitive neurons in the primary visual cortex of monkeys (Livingstone & Conway, 2003; Rust, Schwartz, Movshon, & Simoncelli, 2005). The ST-filter is expressed as

$$f_{i,\phi}(x, t) = \exp(-\frac{x^2}{2\sigma_x^2} - \frac{t^2}{2\sigma_t^2}) \sin(\omega_x x + \omega_{i,t} t + \phi) \quad (1)$$

where i indicates the index of the MID channel to which the output of the filter feeds; x and t indicate horizontal position and time, respectively; ω and σ indicate the angular frequency and space constant of the Gabor function, respectively; and φ is the phase, taking either 0 (i.e., odd symmetric filters),

$$f_{i,even}(x, t) = \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{t^2}{2\sigma_t^2}\right) \cos(\omega_x x + \omega_{i,t} t) \quad (2)$$

or π/2 (i.e., even symmetric filters),

$$f_{i,odd}(x, t) = \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{t^2}{2\sigma_t^2}\right) \sin(\omega_x x + \omega_{i,t} t) \quad (3)$$

Here, we focus our argument on horizontal directions and disregard any changes in the vertical direction. The slope in the spatiotemporal coordinates (ST-filters in Fig. 3) indicates the preferred speed of the LMDs and we used two detectors with different speed tunings (0.5 and 1.0 deg/s; Fig. 3) for each direction (left or right). We set the values of ω for each ST-filter to a spatial frequency of 1 cycle per degree (cpd), which realizes these speeds. The output of each LMD is obtained by convolution with retinal stimuli. Outputs of four LMDs used as inputs to four MID channels are expressed as

$$r_{i,even,L}(x, t) = f_{i,even}(x, t) * I_L(x, t) \quad (4)$$

$$r_{i,odd,L}(x, t) = f_{i,odd}(x, t) * I_L(x, t) \quad (5)$$

$$r_{i,even,R}(x, t) = f_{i,even}(x, t) * I_R(x, t) \quad (6)$$

$$r_{i,odd,R}(x, t) = f_{i,odd}(x, t) * I_R(x, t) \quad (7)$$

where r indicates output to retinal stimulation I and * denotes convolution. Subscripts L and R indicate left and right retinas, respectively.

1.5. MID detector

We constructed a MID detector by combining the outputs of the left and right LMDs. To our knowledge, there has been only one study, by Beverley and Regan (1975), that has reported the discrimination threshold of the MID direction. To be consistent with their psychophysical data, we assumed there were four MID channels with different directional tunings, with two channels each the left and right sides. These four channels comprise the right-far, right-near, left-near and left-far MID channels and are symmetrical about the line of the straight-ahead direction (line Z in Fig. 4). The near channels are sensitive to directions between the line of the straight-ahead direction and the line of sight of one of the eyes, and the far channels are sensitive to outside the region between the lines of sight of the two eyes, as shown in Fig. 4. The near channels combine outputs of a pair of LMDs sensitive to opposite lateral directions between the two retinas and the latter combines outputs of the LMDs sensitive to the same lateral direction. The speed tuning differs between the two LMDs of the pairs for one channel. We used left-fast, left-slow, right-fast and right-slow LMDs for each retina to construct four approaching and four receding MID channels.

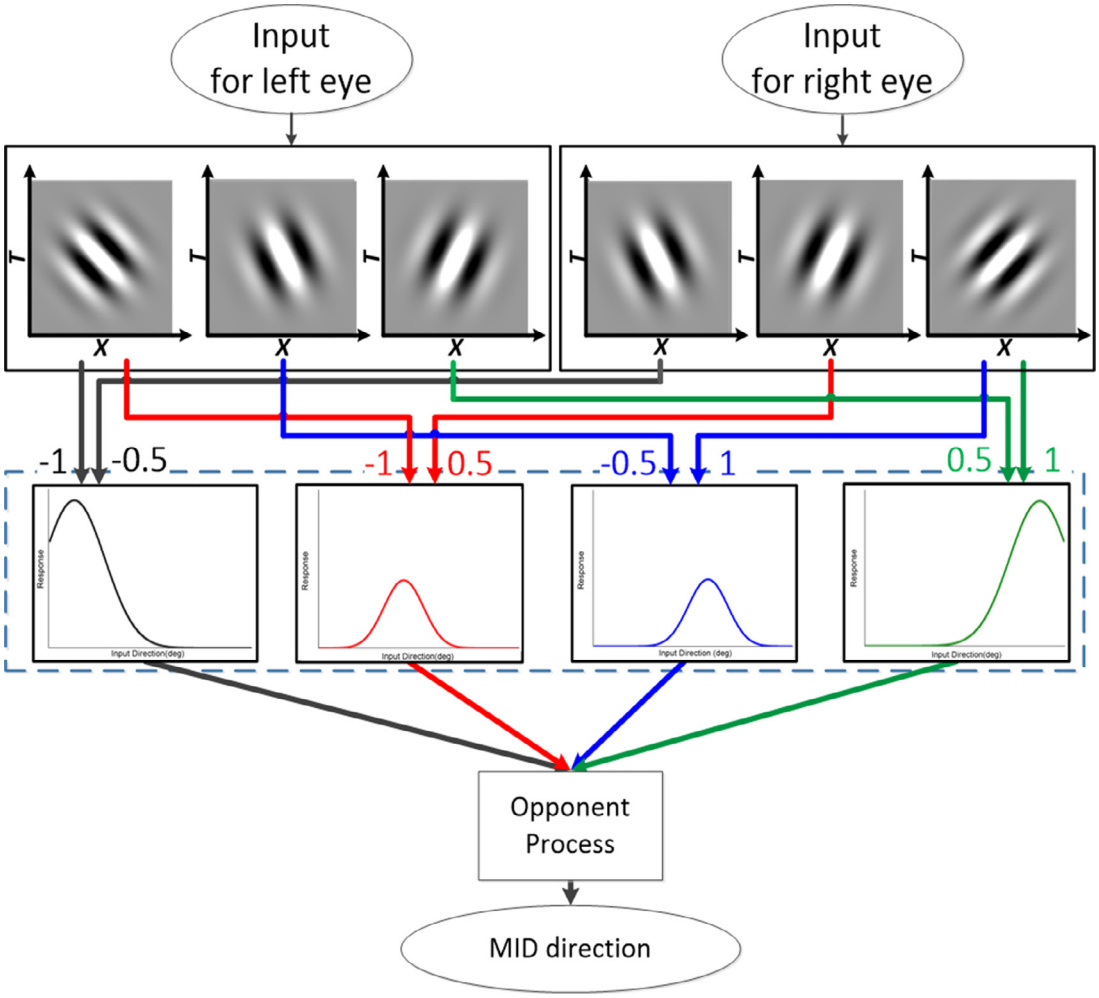


Fig. 3. Model structure for predicting perceived motion-in-depth (MID) direction.

The right-far, right-near, left-near, and left-far approaching MID channels are constructed as below.

$$r_{i,L}(x, t) = r_{i,even,L}(x, t)^2 + r_{i,odd,L}(x, t)^2 \quad (8)$$

$$r_{i,R}(x, t) = r_{i,even,R}(x, t)^2 + r_{i,odd,R}(x, t)^2 \quad (9)$$

$$r_i(x, t) = r_{i,L}(x, t) + r_{i,R}(x, t) \quad (10)$$

where the value of r_i is the MID output of channel i. The combinations of different LMD pairs provide different MID direction tunings. In addition to these four channels, which are sensitive to approaching motion, we constructed four channels sensitive to receding motion. Their sensitivity peaks were determined as the directions opposite to the direction peaks of corresponding approaching MID channels. For example, the left-far approaching channel adds the outputs of the left-fast LMD in the right eye and the left-slow LMD in the left eye as described above, and the left-near receding channel adds the outputs of the left-fast LMD in the right eye and the right-slow LMD in the left eye. For simulation purposes, we used positive and negative signs to code approaching and receding motion along trajectories after subtracting the output of a receding MID channel from the output of the corresponding approaching MID channel.

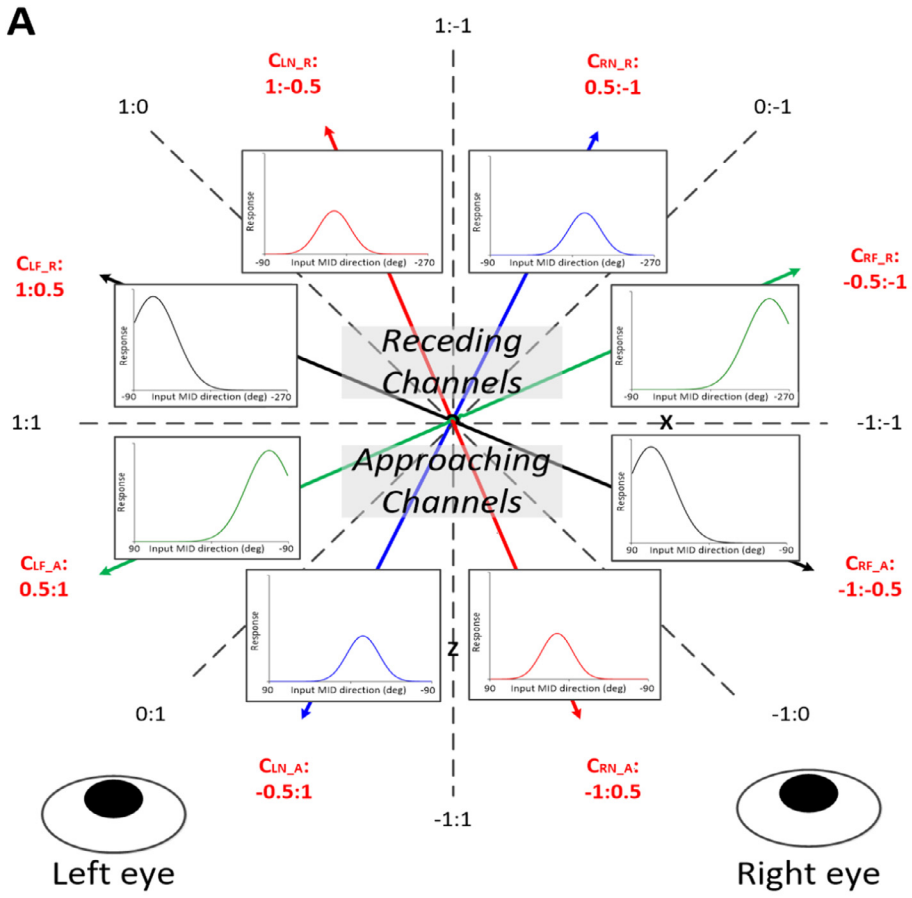
2. Model performance

2.1. Direction tuning

To estimate the direction tuning function of each constructed MID channel, we calculated outputs to stimulation with a moving bar as in Beverley and Regan (1975). However, unlike continuous motion in their study, two-frame motion sequences were used for simplicity. At 57 cm, the stimulus disparity changed from -0.6° to 0.6°, which corresponds to a depth change of 12 mm (± 6 mm from the fixation). The bar width and length were 2 arcmin and 7 deg, respectively. The presentation duration for each of two frames was 1600 ms and the two frames were replaced without a temporal gap. Solid lines in Fig. 2 show the outputs of each MID channel as a function of MID direction; that is, the direction tuning of each channel ($R^2 = 0.58$, $p < 0.01$). Parameter of spatial frequency used for all simulations was set as 1 cpd for all LMDs. The simulation results showed similar direction tunings to the psychophysical results (solid lines in Fig. 2) for all four MID channels. This quantitative evaluation of MID direction tunings provides support for using IOVD detectors to detect MID directions. MID detectors based on IOVD can predict features of direction tuning revealed by psychophysical experiment.

2.2. MID direction

To predict MID directions from the outputs of the MID channels, additional consideration is required. We used the concept of the opponent process to model the process of identifying MID direction, where the output of one of two adjacent channels is subtracted by that of the other, as has been used to model color perception. Our selection of opponent system was based on the most relevant study by Regan and Beverley (1975), who conducted a MID direction discrimination study



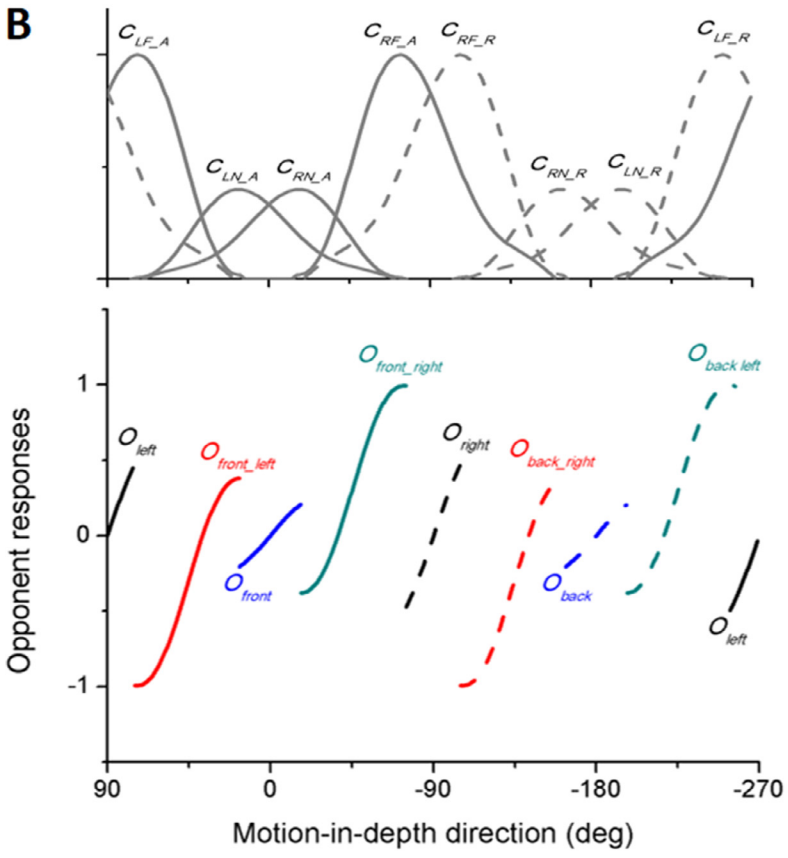


Fig. 4. (A) Arrangement of motion-in-depth detectors. The dashed line with X represents lateral motion with identical motion signals in two retinas, and the dashed line with Z represents motion along depth axis with motion signals in opposite directions between the two retinas. Arrows show peak sensitivity directions of eight MID channels (four approaching and four receding). C_{RN_A} indicates the approaching MID channel sensitive to right-near, C_{RF_A} indicates the approaching MID channel sensitive to right-far, C_{RF_R} indicates the receding MID channel sensitive to right-far, and so on. The numbers at each arrow end indicate the LMDs combinations: e.g., (−1:0.5) indicates right-fast in the right eye and left-slow in the left eye. (B) Eight opponents (named as O_{left}, O_{front_left}, O_{front}, O_{front_right}, O_{right}, O_{back_right}, O_{back} and O_{back_left}) resulted from these eight channels. Solid and dashed gray lines represent approaching and receding channels, respectively. Colored lines represent the outputs of opponent process. Solid and dashed black lines represent opponents resulted from adjacent approaching and receding channels.

Table 1
Opponent (i) combined from two adjacent channels (j and $j + 1$).

i or j	1	2	3	4	5	6	7	8
O (i)	O_{front_right}	O_{front}	O_{front_left}	O_{left}	O_{back_left}	O_{back}	O_{back_right}	O_{right}
C (j)	C_{RF_A}	C_{RN_A}	C_{LN_A}	C_{LF_A}	C_{LF_R}	C_{LN_R}	C_{RN_R}	C_{RF_A}
C ($j + 1$)	C_{RN_A}	C_{LN_A}	C_{LF_A}	C_{LF_R}	C_{LN_R}	C_{RN_R}	C_{RF_R}	C_{RF_R}

and obtained discrimination thresholds. An opponent process, based on 4-MID direction channels (psychophysically estimated from adaptation experiment in Beverley & Regan, 1973), was suggested to predict the human thresholds in MID discrimination. Regan and Beverley (1975) selected the most activated adjacent channel pair from three possible pairs for each MID stimulus, and subtracted the output of one of the channel from that of the other (opponency) to explain the discrimination performance results.

With four approaching and four receding motion channels, the operation results in eight opponents, covering 360° direction in depth (Fig. 4B). We explain the process, focusing on three opponents with four approaching channels while actual calculation was performed with all of the eight channels (the results are the same for these three opponent because there is no contribution from receding channels). In order to make combination of these opponents to predict MID direction, each opponent as O_i, is given a weight as r_i, which is based on these two related channels' responses (C_j and C_{j+1}), as Eqs. (11) and (12). Here, i represents each opponent derived from a pair of channels, as listed in Table 1.

For example, opponent O_{front_left} is the response difference between two adjacent channels of C_{LF_A} and $C_{LN\_A;}$ in this case, r_i, the output of O_{front_left}, is expressed as the square sum of responses from channel C_{LF_A} and C_{LN_A}.

$$r_i = C_j^2 + C_{j+1}^2 \tag{11}$$

$$r_{Ni} = \frac{r_i}{\sum r_i} \tag{12}$$

Fig. 5A shows the outputs of three opponents calculated from four approaching channels and Fig. 5B shows the outputs normalized as expressed in Eq. (12). The model selects the opponent with the largest output as expressed in Eq. (13), and uses it to estimate the MID direction. In the case of Fig. 5, opponent O_{front} will be selected when the input MID direction are between −1 to 1 deg.

$$S(\theta) = O_i(\theta),\ \theta \in Max\{r_i(\theta)\} \tag{13}$$

Each of these segments corresponds to direction signals between the peaks of the two adjacent channels (Fig. 2). Here, we simply connected the end points of the outputs for flanking segments, assuming that the visual system recognizes that the outputs of the two differential systems correspond to the same direction in each case. It should be noted that the outputs of the differential process are relative responses in arbitrary units so that this operation does not lose or distort any information in question here, as expressed by

$$PD(\theta) = K \times \{O_i(\theta) + \Delta\phi_i\},\ \theta_{T,j} < \theta < \theta_{T,j+1} \tag{14}$$

where $PD(\theta)$ is the perceived direction under input direction for the range between −90 and 90 deg; K is the scaling constant; $O_i(\theta)$ is the opponent response combined by channel j and $j + 1$; $\theta_{T,j}$ is the tuning direction of MID detector j; $\Delta\varphi_i$ is the supplement value for connection as a continuous line for a central segment; and $\Delta\varphi_2$ is set as 0 for reference.

Fig. 6(B) shows the continuous function with signed output for perceived direction after connecting these segments. The prediction deviates from the veridical line. Interestingly, the deviation curve is similar to that of error in human judgment (see Prediction 4.2). Overall, we constructed a complete model that could account for human perception of MID direction, based on monocular motion signals and MID detection channels. We, next used the model to predict a variety of psychophysical results for human perception of MID.

3. Prediction of psychophysical results

We applied our model to five different aspects of MID perception: discrimination of MID direction, perceived angle of MID, sensitivity for rotation in depth, temporal tuning of rotation perception and the effect of lateral motion direction on MID perception. In addition, we also examined whether the model could predict MID perception with different types of random-dot stimuli. Parameter of spatial frequency used for all predictions below was set as 1 cpd for all LMDs, and parameters of temporal frequency tuning was set as 1 Hz or 0.5 Hz for fast and slow LMDs.

3.1. Direction discrimination of MID

Directional discrimination was measured for different MID directions by Beverley and Regan (1975). A black bar was used as stimulus against the light background and the bar moved in depth in a specific direction sinusoidally, controlled by the ratio of lateral motion speeds of bars for two eyes. The bar stimuli were presented trough a mirror stereoscope, and the moving direction in depth was controlled by difference between the bars for the left and right eyes. There were two sequential presentations of bar movements in a trial. Observers judged whether the movement of the second presentation was on the left or right side of the first presentation. Thresholds for direction discrimination were measured by the method of constant stimuli, with varying direction difference between two sequential movements. The results showed three peaks: corresponding motion toward either of the two eyes and toward the center of the two eyes (squares in Fig. 7), and four channels that were estimated experimentally (Fig. 2) predicted the

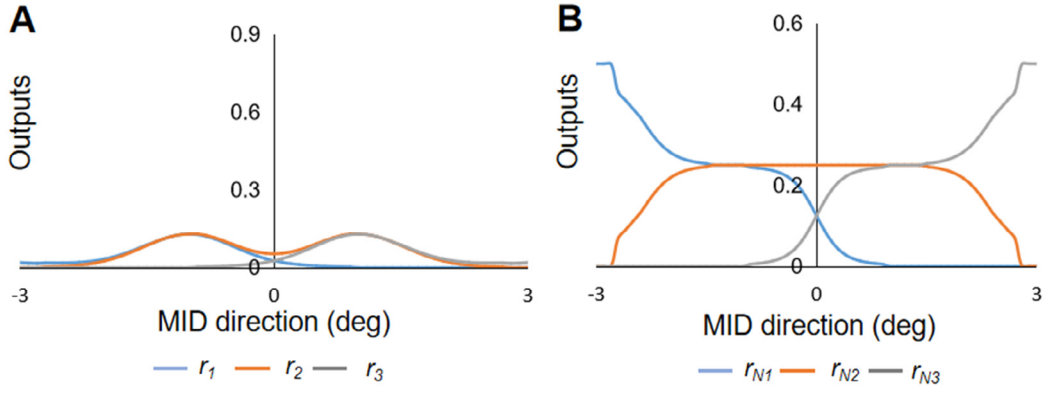


Fig. 5. (A) Original outputs of these opponent (B) Normalized outputs of these opponent.

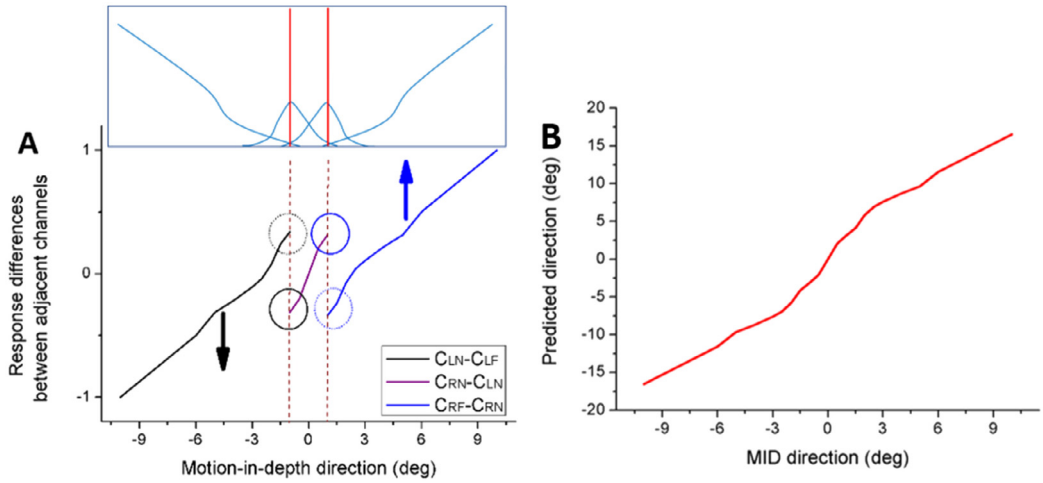


Fig. 6. Model simulation responses. (A) Response difference between adjacent channels as a function of motion-in-depth (MID) direction. Colored lines represent outputs of the opponent process. (B) Prediction of perceived MID direction. Predicted perceived direction is plotted against physical MID direction.

MID directional discrimination results well. We expected that our model would also predict the results because the model assumes four MID channels similar to those proposed by Beverly and Regan (see Fig. 2).

We used our model to simulate the experiment by Beverley and Regan (1975) and predicted their psychophysical results. We assumed that a direction discrimination threshold ($Thresh_{Dis}(\theta)$) is determined by the direction difference that provides a fixed value K_{Dis} and the value is obtained by least square error fitting, as expressed by Eq. (15).

$$K_{Dis} = |PD(\theta + Thresh_{Dis}(\theta)/2) - PD(\theta - Thresh_{Dis}(\theta)/2)| \quad (15)$$

where $Thresh_{Dis}(\theta)$ is the discrimination angle at threshold under input MID direction θ; $PD(\theta)$ is the perceived direction; and K_{Dis} is a fixed value obtained by least square error fitting as the only one free parameter. Fig. 7 compares the predicted values and the experimental results ($R^2 = 0.43$, $p < 0.01$), which are similar in the following two aspects: 1) there are three peaks located at the trajectories to the left eye, between the two eyes and to the right eye, and 2) gradual sensitivity reduction with the distance from the center of the head for directions beyond ± 3°. This is consistent with Beverley and Regan's (1975) attempt to predict the discrimination results with the four channels obtained empirically (symbols in Fig. 2) in Beverley and Regan (1973), which also indicates the two features.

3.2. Perceived MID direction

There are several reports of perceived MID direction (Harris & Dean, 2003; Welchman, Tuck, & Harris, 2004; Welchman et al., 2008). They show that the visual system is biased toward overestimating angles from the head direction when the movement angle is small as shown in Fig. 8. Here, we predicted the results of Welchman et al. (2008), which measured perceived MID direction systematically for a wide range of directions. The stimulus was a small square target surrounded by a square reference plane. In each trial, the target moved towards the observer along a certain trajectory on the transverse plane at the height of the eyes. To report the perceived direction in depth, the observer adjusted the angle of a bar (pointer) to match the perceived direction. We calculated the output of our MID model to a moving small dot in different directions using Eq. (14) above.

Fig. 8 shows predictions of perceived MID direction ($R^2 = 0.99$,

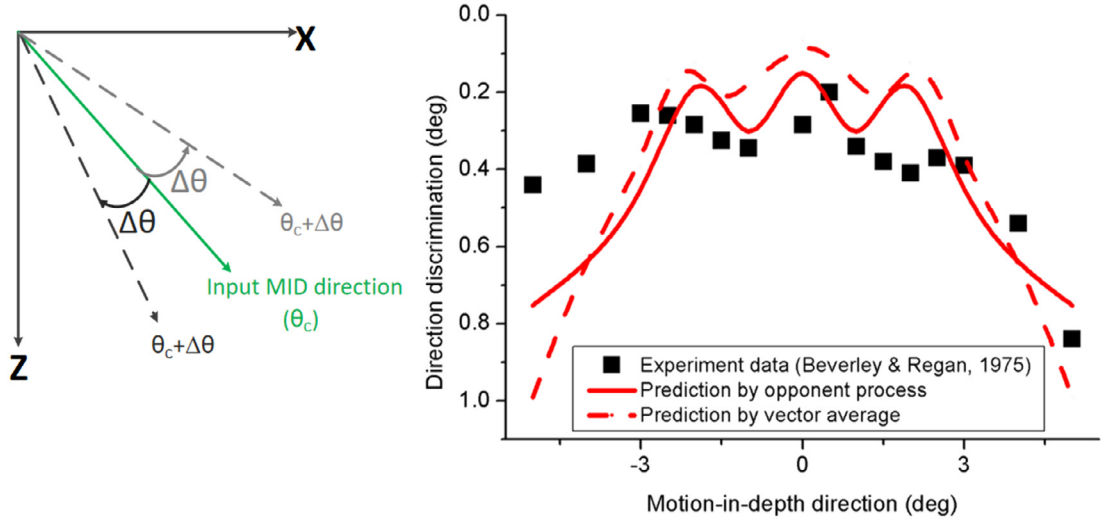


Fig. 7. MID direction discrimintion between predictions and psychophysical data. Left figure illustrates the MID direction, which is the angle along the depth dimension and the direction discrimination ($\Delta\theta$, y-axis in the right figure), which is the threshold in terms of angle on the transverse plane. Right figure shows the prediction compared with experimental data. The horizontal axis represents the input direction of motion in depth (θc in the left figure) and the vertical axis represents discrimination threshold. Squares represent human psychophysical data obtained by Beverley and Regan (1975). The solid line represents predicted discrimination thresholds by the present model. Dashed line represents predictions based on vector average (see Discussion 5.2).

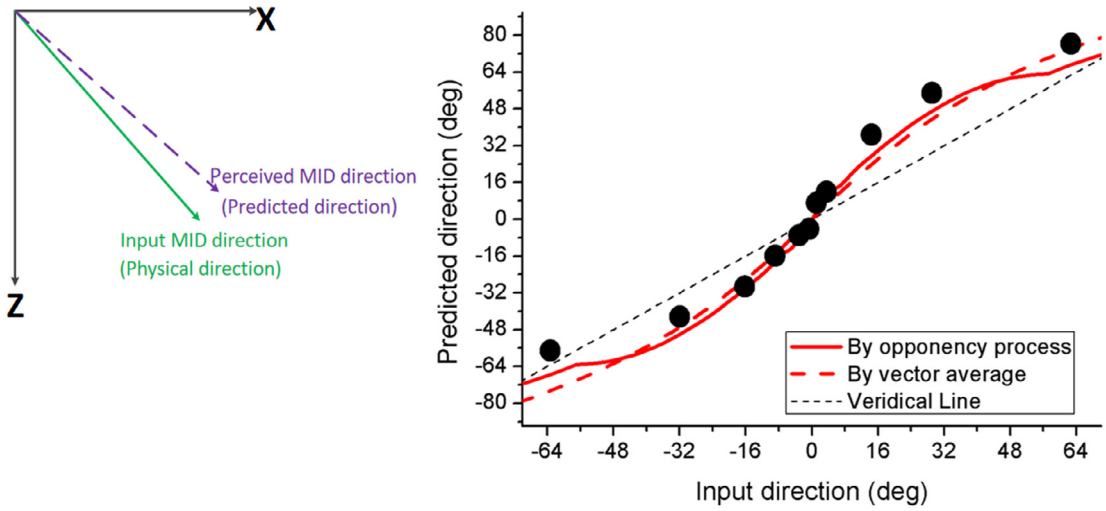


Fig. 8. Predictions of perceived direction in depth by the MID model. The left figure illustrates the physical or input direction (green solid line) and perceived direction (violet dotted line) on the transverse plane. The right figure shows input direction on x-axis, and experimental and predicted results of perceived directions on y-axis. The circles represent human psychophysical data obtained by Welchman et al. (2008), averaged over 10 observers. The red line represents predicted MID direction. The dashed line represents veridical directions. Dashed line represents predictions based on vector average. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$p < 0.01$) with experimentally measured data (Welchman et al., 2008). Similar deviation from the veridical direction was evident in both predictions (red solid line) and experimental results (filled circles). Overestimation was observed for directions closer to head orientation (around 0 deg), appearing as a slope steeper than the veridical line. The slope depends on the shapes of the two channels used for subtraction to obtain the differential responses and the steep slope is related to the small distance between the peaks of the two central channels (C_{RN} and C_{LN}).

3.3. Spatial and temporal frequency tuning of MID perception

Rotation in depth was investigated with different spatial and temporal frequency stimuli using sinusoidally oscillating sine-wave gratings (Lages, Mamassian, & Graf, 2003). In this study, they measured the discrimination threshold for the direction of rotation in depth (clockwise [CW] or counterclockwise [CCW]) while controlling depth amplitude by the inter-ocular phase difference. Vertically oriented sine-wave gratings moving sinusoidally were presented within a Gaussian spatial envelop (i.e. Gabor patch drifting sinusoidally). Gabor patches for the two eyes were separately presented through a mirror stereoscope, and controlling the phase between the two Gabor patches changes the amplitude of motion in depth. Thresholds for MID discrimination were measured by the method of constant stimuli, varying interocular phase difference between Gabor patches presented to two eyes. Their results showed a U-shaped function with temporal frequency for all four spatial frequencies and the sensitivity decreases as spatial frequency increase (symbol in Fig. 9). In order to predict this experimental data, sensitivity, S, was predicted by the proposed model as below.

$$S = Max\left\{\frac{K_{ST}}{A_{DM}(\theta, \omega_{in}) \times \cos\theta}\right\} \tag{16}$$

where *Max* takes the maximum value over the range $0 < \theta < 2\pi$ with different MID direction θ. The parameter, ω_{in} is the temporal frequency of stimulus oscillation; $A_{DM}(\theta, \omega_{in})$ is the amplitude of perceived depth motion; and K_{ST} is a free parameter to set the absolute value of sensitivity. K_{ST} was the only free parameter for least square fitting, and the same value was used to fit for all combinations of ω_{in} and spatial frequencies.

Prediction results showed similar U-shaped functions with temporal frequency for all four spatial frequencies (Fig. 9, $R2 = 0.20$, $p = 0.15$). To see the effect of spatial frequency, we plotted the lowest threshold value among different temporal frequencies for each spatial frequency in Fig. 10 with the prediction from the present model ($R^2 = 0.91$, $p = 0.032$). According to the model structure, the tuning functions are determined by spatiotemporal properties of the LMD subunits. The simulation result indicates, therefore, that the parameters of the LMDs selected to construct the four MID channels are also suitable to explain sensitivities for a variety of spatiotemporal conditions.

3.4. Effect of oscillation frequency on perceived rotation in depth

We investigated the effect of speed on rotation in depth as in Shioiri, Nakajima, Kakehi, and Yaguchi (2008). Random-dot patterns (correlated and uncorrelated between the eyes) were used as stimuli. The Random-dot planes were rotated in depth without changing the surface normal (always front parallel). It is assumed that both the IOVD and CD cue are available in the correlated condition, but only the IOVD cue is available in the uncorrelated condition. The dot contrast against the background was controlled to measure the contrast threshold for discriminating rotation directions in depth (CW or CCW). They found that the sensitivity of MID has a peak rotation with oscillation frequency at around 1 cyc/s.

Because the experiment used a contrast threshold without amplitude changes, we could not use the previous prediction method to determine the amplitude threshold. So, it was necessary to assume a process to determine the rotation direction to predict the results. The direction in depth changed continuously while stimulus rotating in depth. To discriminate CW and CCW rotations, how direction changes in time should be estimated. For example, during the approaching half of CW rotation, motion direction changes from 90 deg (rightwards) to −90 deg (leftwards) through 0 deg, whereas the motion direction changes from −90 deg (leftwards) to 90 deg (rightwards) through 0 deg. The direction change can be estimated by taking a temporal derivative of the moving direction in depth. The direction is CW if the sign of the derivative is negative and CCW if it is positive in our model. We assumed that strength of CW and CCW rotation are coded based on one cycle of the temporal derivatives of direction in depth as expressed in Eq. (17).

$$K_{Out} = C \times \int_{\theta=0}^{\theta=2\pi} d\{PD(\theta)\}/dt \tag{17}$$

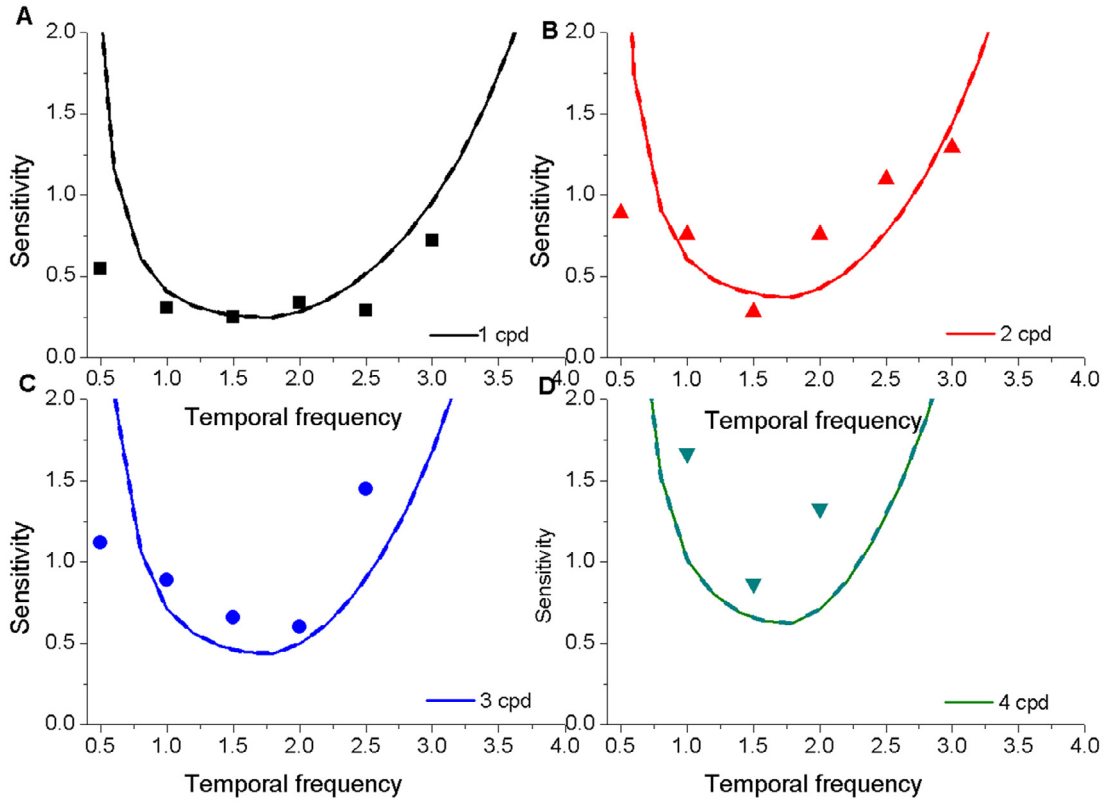


Fig. 9. Sensitivity to rotation in depth as a function of temporal frequency. (A) 1 cpd, (B) 2 cpd, (C) 3 cpd, (D) 4 cpd. Solid lines represent predicted sensitivity to rotation in depth. Symbols represent human psychophysical data obtained by Lages et al. (2003). Dashed line represents predictions based on vector average (same with predictions based on opponent process).

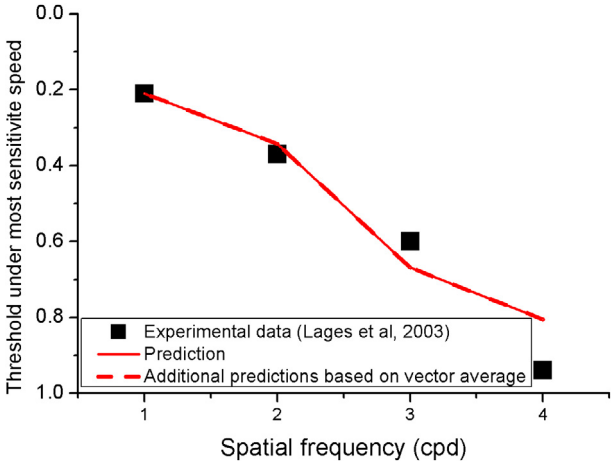


Fig. 10. Threshold comparisons for rotation in depth between predictions and experimental data. The minimum value of the temporal frequency tuning curve is plotted against spatial frequency. The red line represents predictions and squares represent experimental data from Lages et al. (2003). Dashed line represents predictions based on vector average (same with predictions based on opponent process). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where ω_t is the oscillation frequency of input stimuli; K_{Out} is the coded output of rotation speed where direction is determined by the output sign; $PD(\theta)$ is the perceived direction as previously described and C is the contrast of the input stimuli. Contrast value is multiplied after integration and differentiation because the responses of the MID channels are linear to the input luminance contrast. Using the equation, we can determine the threshold for rotation detection under stimuli with different oscillation frequencies of ω $Thresh_{TF}(\omega_t)$) by the contrast that provides fixed value K_{Out}. K_{Out} is positive or negative dependently on rotation direction, but, here, we simply use one direction for simulation since the outputs are the same for CW and CCW except for the sign. K_{Out} is the only free parameter for fitting to the experiment results.

$$Thresh_C = \frac{K_{Out}}{\int_{\theta=0}^{\theta=2\pi} d\{PD(\theta)\}/dt} \tag{18}$$

Fig. 11(A) shows the time course of stimulated perceived MID direction for rotation in depth after temporally averaging with a 150-ms window (Fig. 8). The derivatives were always positive for CW rotations and negative for CCW rotations. Threshold was obtained as the contrast that gives a certain value of the difference between the derivative responses for CW and CCW directions averaging one cycle (Fig. 11(B)). The value was a fitting parameter as in previous cases.

Fig. 12 compares psychophysical data and predictions for different oscillation frequencies of rotation in depth ($R^2 = 0.98$, $p < 0.01$). The predictions under both IOVD cue and full (IOVD + CD) cue conditions show similar oscillation frequency sensitivity with a peak at around 1cps, and only difference is the estimated parameter of K_{Out}. The sensitivity is about 2 times higher with full cue stimulus than with IOVD stimulus.

3.5. Effect of lateral motion direction on MID perception

The effect of direction sensitivity to lateral motion on MID has been reported (Rokers et al., 2008). They measured MID sensitivity by changing motion coherence in random-dot stimuli under several directions on the front parallel plane from horizontal to vertical directions. Anti-correlated random-dot patterns were used in their experiment to isolate IOVD cue from CD cue. Two volumes, composed by anti-correlated RDP, were shown, moving in opposite directions in depth,

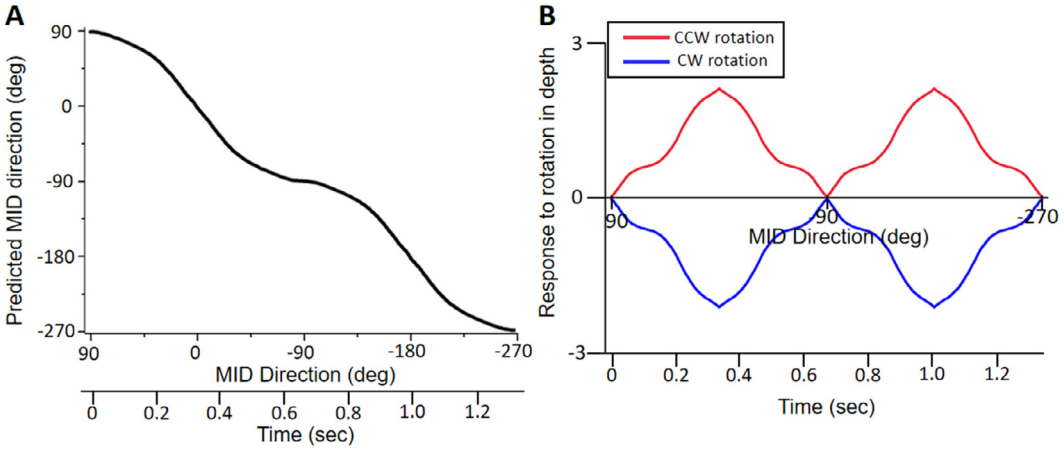


Fig. 11. (A) Time course for predicted direction of clockwise (CW) rotation in depth. (B) Time course for responses to CW and counterclockwise (CCW) rotations. Blue and pink lines represent the CW and CCW rotations, respectively. Oscillation frequency is 0.7 cyc/s in this figure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

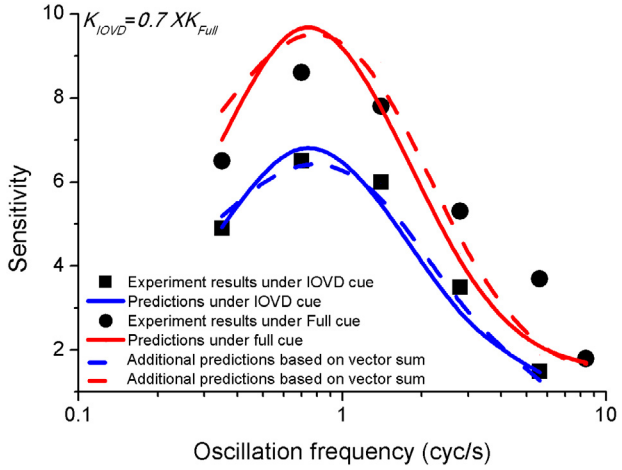


Fig. 12. Comparison of predictions and experimental data for different rotation speeds. Sensitivity to rotation in depth is plotted against oscillation frequency. Solid red and blue lines represent predictions for IOVD cue condition and Full cue condition, respectively. Squares and circles represent human psychophysical data under IOVD and Full cue conditions obtained by Shioiri et al. (2008). Dashed line represents predictions based on vector average. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

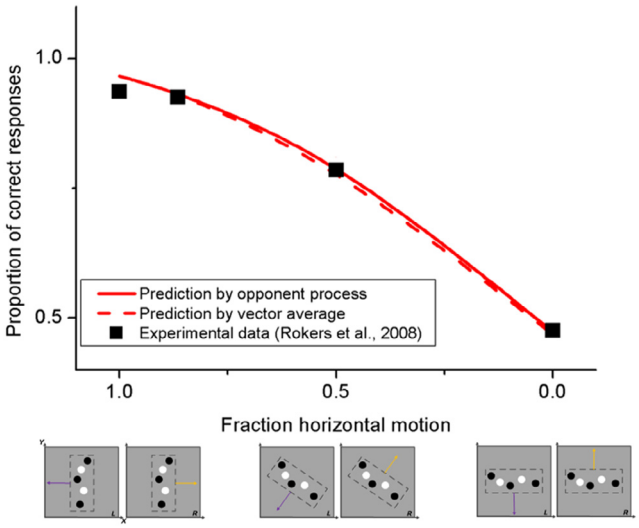


Fig. 13. Correct response rates for MID as a function of moving direction on the front parallel plane. Bottom figures illustrate the moving conditions used. The two arrows in each figure represent motion directions for two eyes respectively. The red solid line represents predictions and the squares represent experimental data. Dashed line represents predictions based on vector average. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

i.e. approaching or receding. The orientation of the volumes, thus the direction of dot motion on the front parallel plane, was varied. Thresholds for MID direction discrimination were measured by the method of constant stimuli, varying coherence of moving dots under different dot-motion-direction conditions. Their results showed that sensitivity to MID decreased with changing direction from horizontal to vertical (the average of four participants is shown Fig. 13). This suggests that the perception of MID is basically determined by the horizontal motion signals. This is consistent with our implicit assumption that only LMDs sensitive to horizontal motion contribute to MID perception.

To predict performance of human observers in Rokers et al. (2008), we simple defined output strength as summed responses of four MID detectors, as expressed below.

$$St_{out}(\gamma) = \sum r_i(\gamma) \tag{19}$$

where $St_{out}(\gamma)$ is the output strength from MID detectors under moving direction of γ. γ represents the moving direction on the front parallel plane, and $r_i(\gamma)$ is the response of channel i for stimulus moving with that direction. After calculating the output for each condition, we transformed the output strength to a percentage using a cumulative Gaussian function (changing from 50% to 100% for correct responses). This was done to relate the model output with the percentage of correct responses, as expressed in Eq. (18). The space constant, σ, was used to scale the model output as a free parameter to fit experimental results as in the previous cases.

$$PC(\gamma) = \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{St_{out}^2(\gamma)}{2\sigma^2}\right) \tag{20}$$

where $St_{out}(\gamma)$ is the output strength from the MID detectors; σ is a free parameter to fit the experimental data; $PC(\gamma)$ is the prediction of correct response rates under moving direction of γ. Fig. 13 shows that the model predicts the results well ($R^2 = 1.0$, $p < 0.01$). This suggests that only horizontal component of motion signals on the front parallel plane contributes to MID perception.

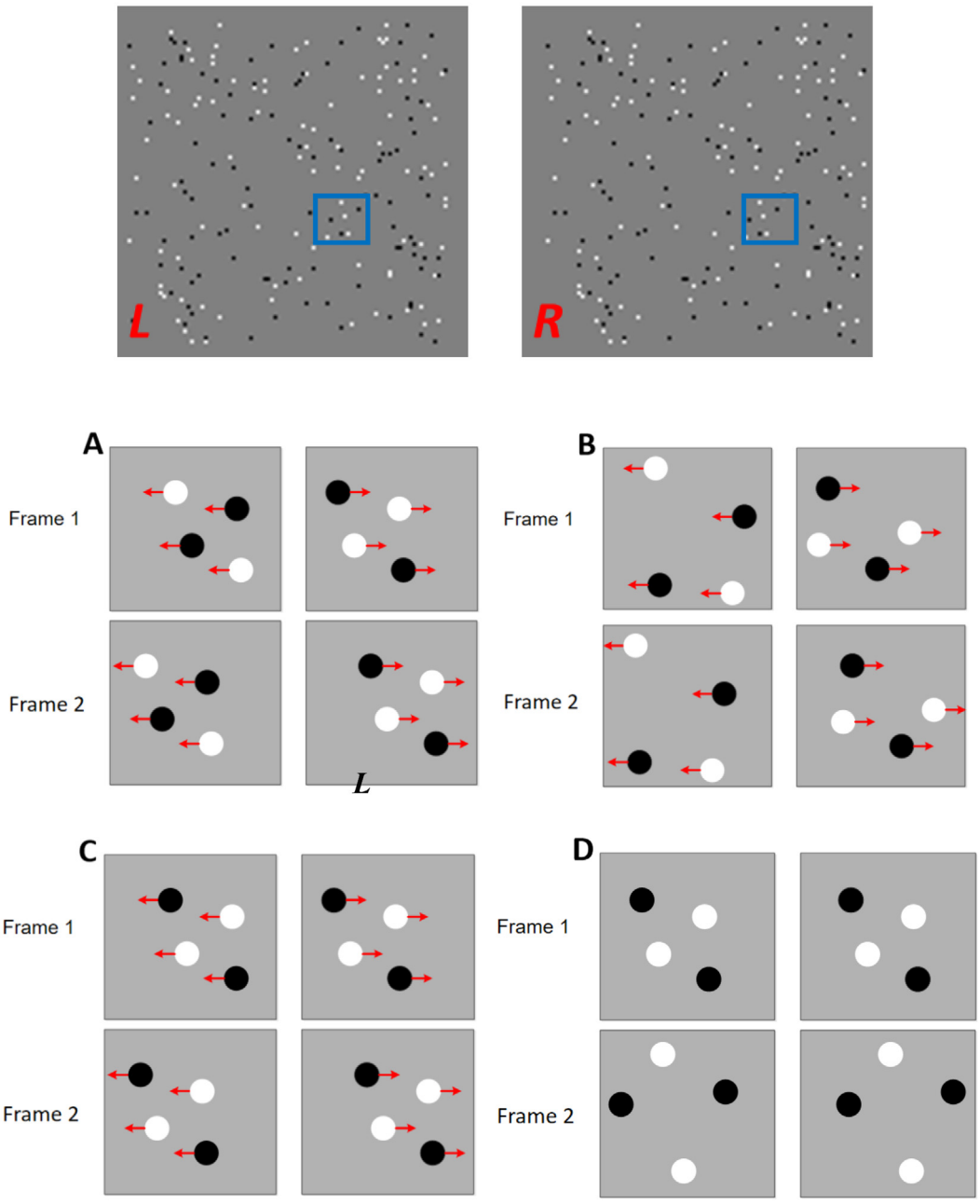


Fig. 14. Four types of random-dot pattern as stimulus for motion in depth. Binocular correspondence and dot motion for each stimulus is depicted for the small region of the blue square within each of the left and right random-dot patterns. (A) anti-correlated random-dot-patterns (RDPs), (B) de-correlated RDPs, (C) full-cue RDPs and (D) CD-cue RDPs. Left (L) and right (R) eye half images for each RDP are shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.6. Effect of binocular and temporal correlations on MID perception

We investigated the effect of cue conditions on different types of random-dot stimuli. There are several psychophysical experiments that have used different types of random-dot patterns (RDPs) as stimuli to investigate MID perception. We used four types of RDPs that are typically used for MID studies. The first type is full-cue RDP (Fig. 14(C), where there is a corresponding dot in the right image for each dot in the left image and change in disparity with time provides coherent monocular motion by shifting all the dots horizontally. Thus, both IOVD and CD cues are available in the full-cue RDP. The second type is anticorrelated RDP (anti-RDP; Fig. 14(A)), where the contrast of a dot in the left image is opposite to that of the corresponding dot in the right image and change in disparity provides coherent monocular motion by shifting all the dots without changing the contrast of each image (Rokers et al., 2008, 2009). The third type is decorrelated RDP (de-RDP, Fig. 14(B)), where the random dots in the left and right images are generated independently so that there are no dot correlations between the two images and monocular motion is observed at each retinal image (Shioiri et al., 2000; Shioiri et al., 2008; Harris et al., 2008). The fourth type is dynamic random-dot stereogram (d-RDS, Fig. 14(D)), where the left and right RDPs are spatially correlated and no temporal correlation is provided between frames to see monocular motion (Julesz, 1960, 1971; Shioiri et al., 2000; Tseng, Gobell, Lu, & Sperling, 2006). Both anti-RDP and de-RDP were designed to isolate the IOVD cue from the CD cue and the d-RDS was designed to isolate the CD cue from the IOVD cue. Debate surrounds the isolation of cues among the anti-RDPs, de-RDPs, and d-RDSs (Cogan, Lomakin, & Rossi, 1993; Cogan, Kontsevich,

Lomakin, & Halpern, 1995; Neri, Parker, & Blackemore, 1999; Shioiri et al., 2012; Giesel, Wade, Bloj, & Harris, 2018) and model simulation is one approach for examining the contribution of each cue to each stimulus type (Shioiri et al., 2008; Giesel et al., 2018).

The effect of de-RDP and anti-RDP was investigated experimentally by Giesel et al. (2018). Three types of random-dot pattern stimuli were used in their experiment: (1) correlated, (2) anti-correlated and (3) de-correlated condition. The dots moved to generate motion in depth perception, approaching or receding, when binocularly viewed. Thresholds for direction discrimination were measured by the method of constant stimuli, varying motion coherence of dots. Their results suggested that both IOVD and CD are used to see MID. Furthermore, they found that performance with anti-RDP is better than that with de-RDP, which indicates that CD possibly contributes to anti-RDP stimulation in addition to IOVD. Thus, sensitivity is perhaps higher with two cues than with one cue. To confirm this, we examined whether there was any difference between anti- and de-RDPs for MID perception based on IOVD, which could differentiate the sensitivity between the two anti- and de-RDPs. We simulated responses to each of four types of RDPs used in Giesel et al.

Similarly to previous section, we defined the output strength ($St_{out}(\rho)$) as the summation of responses from the MID detectors, as expressed below.

$$St_{out}(\rho) = \sum r_i(\rho) \tag{21}$$

where $St_{out}(\rho)$ is the output strength from the MID detectors; ρ is the coherence between the input images for the two eyes; and $r_i(\rho)$ is the response of channel i for stimulus with coherence of ρ.

To transform the channels' outputs to a coherence threshold, we used the cumulative Gaussian function described in Eq. (22), to obtain the coherent threshold as the coherence value that provides 75% correct responses as expressed by the following equation.

$$Thresh_{coh} = \{\rho | f_{psy}\{St_{out}(\rho)\} = 0.75\} \tag{22}$$

where $St_{out}(\rho)$ is the output strength; $Thresh_{coh}$ is the coherence threshold at 75%; and f_{psy} is the cumulative Gaussian function. We calculated 100 times with each of coherence leavels (0, 0.2, 0.4, 0.6, 0.8, and 1) used.

The predictions showed similar sensitivity for anti-RPD, de-RDP and full-cue RDP (R^2 = 0.93, p = 0.023), but the sensitivity of d-RDS (or CD condition) was almost none (Fig. 15). This does not agree with the behavioral results, which show better performance (lower threshold) for full-cue and anti-RPD than for de-RDP. Since our IOVD model indicates the same threshold for these three conditions, the different sensitivities of anti-RDP and de-RDP is difficult to be attributed to potential differences in responses to IOVD among different stimulus conditions. This finding agrees with suggestion of contribution of CD cue in anti-RDP by Giesel et al. (2018). This is also similar to static depth perception with contrast reversal stereograms (Cumming & Parker, 1997; Masson, Busettini, & Miles, 1997; Watanabe et al., 2008). Contrast reversal may not be an appropriate method to isolate the IOVD cue from the CD cue as suggested before also by a model study that showed the potential contribution of the CD cue in anti-RDP but not in de-RDP (Shioiri et al., 2012).

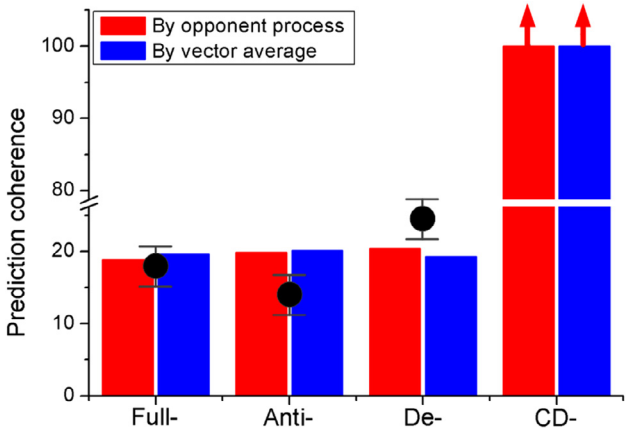


Fig. 15. Red columns show predicted response to each of the four types of random dot patterns (RDPs): Full-RDP, anti-RDP, de-RDP and cd-RDP. Filled circles represent average results of human data in Giesel et al. (2018). Blue columns represent predictions based on vector average. Error bars represent ± 1 standard error. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

4.1. Comparison with other models

We proposed an IOVD-based MID model that successfully predicts a variety of psychophysical results. To our knowledge, our study is the first computational models that predicts the perceived MID direction for given retinal input images. We compared our model with other MID models. So far, several studies have presented valid evidence for IOVD as an effective cue for human perception of MID and have proposed models of MID based on IOVD (Baker & Bair, 2016; Peng & Shi, 2014; Sabatini & Solari, 2004; Shioiri et al., 2009; Welchman et al., 2008). All these models have a similar initial stage, where monocular motion signals are obtained by LMDs and comparisons are made between the left and right motion signals to provide the MID signals. Their predictions should be similar to those made by the present model.

However, only one of the previously proposed models focused on MID direction. Welchman, Lam and Bülthoff predicted the depth direction bias estimated by human judgement based on a Bayesian model (Welchman et al., 2008). Assumptions made in this model are as follows. First, perception of motion direction for an oblique trajectory is separated into the one along the depth axis and the other along the lateral axis. Second, the reliability of lateral motion signals is higher than that of MID signals. Velocities on each direction (horizontal and depth) are estimated using the Bayesian model based on reliability on each direction. The combination of these two components represents the MID direction. The prediction showed bias toward enhancing changes in direction around the head direction (Fig. 16(A)) as does our model (Fig. 8). Similar prediction results for perceived direction between the two models suggest that our model can be regarded as a decisive form for realization of the Bayesian process. However, we found that the Bayesian model cannot predict the psychophysical results of the direction discrimination (Beverley and Regan, 1975). This contrasts to our model, which shows better predictions for direction discrimination as well as perceived MID direction.

4.2. Comparison between opponent process and vector average

To predict perceived direction, combining four channels' outputs, we used an opponent type of calculation, where direction is determined based on the output difference between adjacent channels. It is an alternative method to use vector averaging outputs of all four channel. Assuming that each channel has a label of tuning direction (e.g., the direction with peak sensitivity), perceived direction can be estimated by a combination of these channels' outputs as in Eq. (23).

$$PD(\theta) = a\tan\left(\frac{\sum r_i(\theta) \times \sin\phi_i}{\sum r_i(\theta) \times \cos\phi_i}\right) \tag{23}$$

where θ is the input direction; $r_i(\theta)$ is the response of the ith channel; and φ_i is the tuning direction of the ith channel. The prediction results based on vector summation for each psychophysical experiment were shown (dashed red lines in Figs. 7–10, 12 and 13 and blue columns in Fig. 15). The vector averaging showed similar performance to the

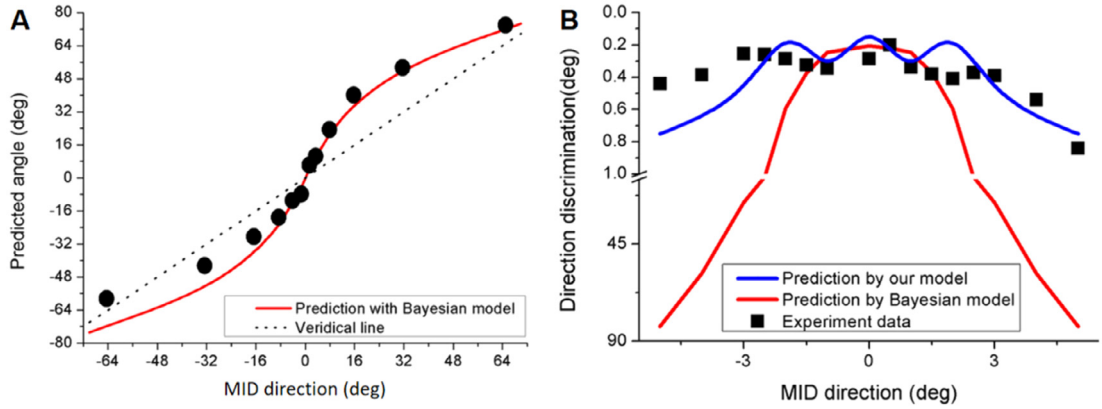


Fig. 16. Prediction results based on the Bayesian model. (A) Perceived MID direction. The red line represents the prediction result and the circles represent experimental data from Welchman et al. (2008). (B) Prediction of direction discrimination based on perceived direction derived from (A). The red line represents the prediction result using the Bayesian model of Welchman, Lam, and Bülthoff and blue line represents the prediction of our model. Squares represent experimental data from Beverley and Regan (1975). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Comparison of RMS errors based on opponent process and vector average (OP: opponent process; VA: vector average).

Simulated psychophysical experiment	OP	VA
Direction discrimination of MID	0.15	0.21
Perceived MID direction	4.96	7.12
Spatial and temporal frequency tuning	0.077	0.077
Effect of speed on rotation in depth	0.34	0.38
Effect of lateral motion direction	0.017	0.024
Effect of binocular and temporal correlations	4.19	5.73

prediction results based on the opponent process. The comparison of RMS (root mean square) errors between the predictions by the opponent process and vector averaging were shown in Table 2. Although a slightly better prediction was found with the opponent process model than with the vector averaging model, the difference was not large enough to conclude that the opponent process model is better than the vector averaging model. Besides, there was no clear evidence for the neuronal process based on the opponent process or vector average. We used the opponent process model in this study simply to be consistent with Beverley and Regan (1975) as described above.

4.3. MT neurons

There are physiological studies on neurons sensitive to MID in the visual cortex of cat and monkey. It has been reported that there are neurons with sensitivity to MID in area 18 of the cat visual cortex (Cynader & Regan, 1978; Regan & Cynader, 1982; Spileers, Orban, Gulyas, & Maes, 1990). In these studies, neuron activities were recorded when bar stimuli were separately presented to the two eyes with different speeds and directions (rightward or leftward). Some neurons were tuned to monocular motions with opposite directions whereas other neurons were tuned to monocular motions with the same directions. Although these findings may be related to neurons sensitive to IOVD, such responses can be interpreted as increasing/decreasing responses with disparity changes around their optimal disparity (Cumming, Shapiro, & Parker, 1998; DeAngelis & Newsome, 1999; DeAngelis & Uka, 2003). However, recent studies have revealed neurons sensitive to certain combinations of monocular velocities corresponding to certain binocular motion directions in the macaque MT neurons (Sanada & DeAngelis, 2014; Czuba et al., 2014). Czuba et al. found MT neurons with selectivity to MID, many of which were sensitive to motion direction toward the head center (i.e., between the two eyes). The distribution of direction tuning of the MID neurons covers a large range from left to right for the directions toward the head. The four MID channels of the present model may be considered as results of grouping some of these neurons into four classes.

It is not uncommon to designate a limited number of channels in visual systems for passively computational efficiency. Both psychophysical and physiological studies have shared a successful history in this regard of human motion perception. In psychophysics, Smith and Edgar (1994) proposed that a process by two psychophysically-plausible, board-tuned temporal channels, interpreted as low-speed-motion detector and fast-speed-motion detector, could well match the human performance of perceived speed after adaptation to motion. Alais, Verstraten, and Blurr (2005) used MAE to isolate two temporal channels- a broad low-pass temporal channel and a band-pass high temporal channel, by dissociating the duration of a single component's MAE as a function of adapters' temporal frequencies. For another example, the perceived MAE was the longest when the test and adaptation spatial frequencies matched (Bex, Verstraten, & Mareschal, 1996; Nishida, Ledgeway, & Edwards, 1997; Shioiri et al., 2009), implying the existence spatial-selective channels in motion perception. Britten, Shadlen, Newsome, and Movshon (1992), Britten, Newsome, Shadlen, and Celebrini (1996) founded that monkeys' 2-D motion direction discrimination can be most accurately simulated if psychophysical decisions are based on pools of at least 100 weakly correlated sensory neurons.

We attempted to construct four MID channels from neural outputs with much narrower direction tunings. Suppose there are 100 neurons whose direction sensitivities are distributed equally for all directions in depth from left to right. We classified them into four groups: (1) from the left fronto-parallel motion (LP) to the left-eye (LE) hitting direction, (2) from the left-eye (LE) hitting to the nose-hitting (NH) direction, (3) from the nose-hitting (NH) to right-eye (RE) hitting direction from right-eye (RE) hitting to right fronto-parallel (RP) motion direction. We assumed that neuron outputs are simply pooled or averaged within each direction group to construct each of the four MID channels (dotted line in Fig. 17) with the direction sensitivity of each neuron calculated according to the following equation.

$$r_i(\theta) = \exp\left(-\frac{(\theta - \theta^T{}_i)^2}{2\sigma^2}\right) \tag{24}$$

where θ and $\theta^T{}_i$ represent input and tuning directions and i indicates number of neurons (labeled from 1 to 100), deviation, σ, was set as 15 deg. Outputs of i between 1 and 38 (LP–LE), those of i between 39 and 50 (LE–NH), those of i between 51 and 88 (NH–RE) and those of i between 88 and 100 (RE–RP) are integrated for the four channels. There were 38, 12, 12, and 38 neurons in these direction groups, for (LP–LE), (LE–NH), (NH–RE), and (RE–RP), respectively. These numbers

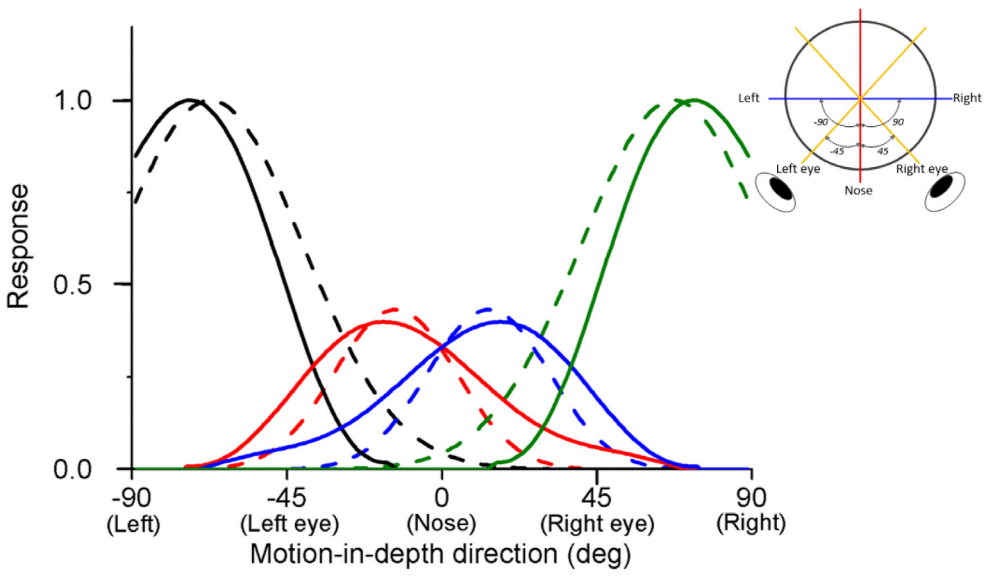


Fig. 17. Constructing four MID channels by pooling outputs of model MT neurons with the direction tuning distribution reported by Czuba et al. (2014). The solid lines represent simulation of constructed MID channels and the dashed lines represent our four channels. Each color represents a different channel of the motion-in-depth detectors.

were decided based on the number of recorded neurons in Czuba et al. (2014), with 17% sensitive to motion direction in the LE–RE range and 53% sensitive to motion direction in the LP-LE and RE-RP range. The resultant sensitivities are shown by solid lines in Fig. 17. They showed good agreement with the MID channels of the present model although the sensitivity is simply assuming the pooling ranges based on psychophysically speculated MID channels. Whether there are physiological MID channels with such sensitivities is a question for future investigation.

5. Limitations and future directions of this model

The experimental data simulated by our model were collected using stimuli that contains CD cue as well as IOVD cue, except for data obtained with de- or anti-RDP, while our current model only considers IOVD cue. A desirable MID model will be one that also analyze CD and changing size cues and it is desirable to investigate the integration process of the all three cues. Another issue is the effect of vertical direction. The model proposed here is restricted to transverse plane, without considering the vertical direction. Since a study (Portfors-yeomans & Regan, 1997) reported that the sensitivity on the sagittal place is as good as on the transverse plane, extending the MID model to vertical direction is important. In addition, this model could be expanded to predict speed of motion in depth as several studies estimated perceived speed under different stimulus conditions. Harris and Watamanuik (1995) investigated speed discrimination of motion in depth; Brooks and Stone (2004) found that IOVD cue provided a cue to speed estimation along with CD cue; and MID speed could be explained by response ratio of two motion channels (Harris, 1980; Smith & Edgar, 1994). It is also important to consider the effect of spatial and temporal tuning and effect of color as well, considering related psychophysical measurements (Tyler & Cavanagh, 1991; Shioiri et al., 2009; Shioiri, Yoshizawa, Ogiya, Matumiya, & Yaguchi, 2012) as well as the effect of visual salience (Harris, McKee, & Watamanuik, 1998) upon MID. Interestingly, a recent study reported a new cue for MID perception (Brooks & Gillam, 2006) based on half-occlusion or half-camouflage (Gillam, Balackburn, & Nakayama, 1999; Pianta & Gillam, 2003a, 2003b). This cue, without binocular matches to compare either disparity or velocity, is suggested to differ from well-known binocular cues, IOVD and CD although some argued that it could be another form of CD (Assee & Qian, 2007). Our proposed model may be a possible investigation tool to provide new insights to this debate as we could provide a quantified IOVD contribution to MID perception under these half-occlusion conditions.

6. Conclusion

We propose a MID model based on IOVD cue comprising four channels sensitive to different directions. The model succeeded in predicting several related psychophysical results. The model assumes physiologically plausible LMDs and provides a connection between perception and neural processes for perception of MID. The contribution of other cues to MID and integration with other cues are issues to be addressed in further studies.

Appendix A. Methods and procedures of psychophysical experiments

MID direction discrimination (Beverley & Regan, 1975)

Stimulus

A black bar was oscillating in depth at 0.8 cyc/s along a certain direction in depth seen through a stereoscope.

Procedure

Two motion-in-depth stimuli (comparison and test) were presented sequentially in one trial. Subjects were asked to answer whether the second stimulus was moving on the left or right side of the first stimulus.

Measurement

Threshold in terms of direction difference between the two stimuli was determined with 75% correct direction judgments. The threshold was measured for the left and right separately and averaged.

Perceived MID direction (Welchman et al., 2008)

Stimulus

A small square target (black) surrounded by a gray reference plane, seen through stereoscope to display motion in depth. The target moved in depth from the screen plane towards the observer on a variety of trajectories (2, 4, 8, 16, 32, 64 deg from the direction of toward the head).

Procedure

After viewing a trajectory, subjects moved the pointer (a stick on a

2D plane) to match the angle of stimulus motion and the angle of pointer was recorded.

Measurement

The angle of pointer indicates subjects' perceived direction.

Spatiotemporal frequency tuning (Lages et al., 2003)

Stimulus

Vertically oriented Gabor patches (space constant, σ of 0.69 deg) drifted horizontally, with different combinations of spatial and temporal frequencies (ten temporal and four spatial frequencies). Increase of phase difference in gratings between the left and right Gabor patches increased amplitude of depth component of stimulus motion so that phase thresholds can be converted to amplitude thresholds.

Procedure

Subjects discriminated between clockwise and counter-clockwise motion in depth.

Measurement

Threshold in terms of phase difference was determined with 75% correct responses.

Effect of oscillation frequency on perceived rotation in depth (Shioiri et al., 2008)

Stimulus

Random-dot patterns (RDP) with full cue and binocularly decorrelated RDP (IOVD) conditions. Front parallel planes rotated without changing the surface normal. The trajectory of rotation in depth was determined by lateral shift of 11 min and disparity change of 16 min. The speed of rotation was controlled by the cycle of rotation per second (0.35, 0.7, 1.4, 2.8 and 5.6 cyc/s).

Procedure

Subjects responded the direction of rotation in depth (clockwise or clockwise).

Measurement

Threshold in terms of contrast was determined as the contrast for 75% correct responses.

Effect of motion direction on the front parallel plane (Rokers et al., 2008)

Stimulus

Anti-correlated random-dot patterns. Two volumes, composed by anti-correlated RDP, were shown, moving in opposite directions in depth, i.e. approaching or receding. The dot motion direction was varied from horizontal to vertical.

Procedure

Subjects judged whether dots in the volume moved towards or away.

Measurement

Response accuracy was collected under different orientation conditions.

Effect of binocular and temporal correlations on MID perception (Giesel et al., 2018)

Stimulus

Three types of random-dot pattern stimuli were used in their experiment, which were moving in depth, approaching or receding, when binocularly viewed. Motion coherence of dots was varied from 0 to 100% in step of 10%.

Procedure

Subjects judged whether dots in the volume moved towards or away by pressing one of the two keys on a keyboard.

Measurement

Threshold for MID perception under different stimuli conditions, in terms of motion coherence, was collected with 75% correctness.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America, 2*, 284–321.

Alais, D., Verstraten, F. A. J., & Blurr, D. C. (2005). The motion aftereffect of transparent motion: Two temporal channels accounts for perceived direction. *Vison Research, 45*(4), 403–412.

Anderson, J. C., Binzegger, T., Martin, K. A. C., & Rockland, K. S. (1998). The connection from cortical area V1 to V5: A light and electron microscopic study. *The Journal of Neuroscience, 18*(24), 10525–10540.

Assee, A., & Qian, N. (2007). Solving da Vinci stereopsis with depth-edge-selective V2 cells. *Vision Research, 47*, 2585–2602.

Baker, P. M., & Bair, W. (2016). A model of Binocular motion integration in MT neurons. *The Journal of Neuroscience, 36*, 6563–6582.

Beverley, K. I., & Regan, D. (1973). Evidence for the existence of neural mechanisms selectively sensitive to the direction of movement in space. *The Journal of Physiology, 235*, 17–29.

Beverley, K. I., & Regan, D. (1975). The relation between discrimination and sensitivity in the perception of motion in depth. *The Journal of Physiology, 249*, 387–398.

Bex, J. P., Verstraten, F. A. J., & Mareschal, I. (1996). Temporal and spatial frequency tuning of the flicker motion aftereffect. *Vision Research, 36*(17), 2721–2727.

Blasdel, G. G., & Fitzpatrick, D. (1984). Physiological organization of layer 4 in macaque striate cortex. *The Journal of Neuroscience, 4*(3), 880–895.

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience, 28*, 157–189.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *The Journal of Neuroscience, 12*(12), 4745–4765.

Britten, K. H., Newsome, W. T., Shadlen, M. N., & Celebrini, S. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neuroscience, 13*(1), 87–100.

Brooks, K. R. (2002a). Interocular velocity difference contributes to stereomotion speed perception. *Journal of Vision, 2*, 218–231.

Brooks, K. R. (2002b). Monocular motion adaptation affects the perceived trajectory of stereomotion. *Journal of Experimental Psychology, 28*, 1470–1482.

Brooks, K. R., & Stone, L. S. (2004). Stereomotion speed perception: Contributions from both changing disparity and interocular velocity difference over a range of relative disparity. *Journal of Vision, 4*, 1061–1079.

Brooks, K. R., & Gillam, B. J. (2006). The swinging doors of perception: Stereomotion without binocular match. *Journal of Vision, 6*, 685–695.

Churchland, M. M., Priebe, N. J., & Lisberger, S. G. (2005). Comparison of the spatial limits on direction selectivity in visual areas MT and V1. *The Journal of Neurophysiology, 93*(3), 1235–1245.

Cogan, A. I., Lomakin, A. J., & Rossi, A. F. (1993). Depth in anticorrelated stereograms: Effects of spatial density and interocular deley. *Vision Research, 33*, 1959–1975.

Cogan, A. I., Kontsevich, L. L., Lomakin, A. J., & Halpern, D. L. (1995). Binocular disparity processing with opposite-contrast stimuli. *Perception, 24*, 33–47.

Cumming, B. G., & Parker, A. J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth. *Letters to Nature, 389*, 280–283.

Cumming, B. G., Shapiro, S. E., & Parker, A. J. (1998). Disparity detection in anticorrelated stereograms. *Perception, 27*, 1367–1377.

Cynader, M., & Regan, D. (1978). Neurons in cat parastriate cortex sensitive to the direction of motion in three-dimensional space. *Journal of Physiology, 274*, 549–569.

Czuba, T. B., Rokers, B., Huk, A. C., Cormack, L. K., & Kohn, A. (2014). Area MT encodes three-dimensional motion. *The Journal of Neuroscience, 34*, 15522–15533.

DeAngelis, G. C., & Newsome, W. T. (1999). Organization of disparity-selective neurons in macaque area MT. *The Journal of Neuroscience, 19*, 1398–1415.

DeAngelis, G. C., & Uka, T. (2003). Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *Journal of Neurophysiology, 89*, 1094–1111.

Giesel, M., Wade, A. R., Bloj, M., & Harris, J. M. (2018). Investigating human visual sensitivity to binocular motion-in-depth. *Vision, 2*, 41.

Erkelens, C. J., & Regan, D. (1986). Human ocular vergence movements induced by changing size and disparity. *Journal of Physiology, 379*, 145–160.

Fernandez, J. M., & Farell, B. (2005). Seeing motion-in-depth using inter-ocular velocity difference. *Vision Research, 45*, 2786–2798.

Gelb, D. J., & Wilson, H. R. (1983). Shifts in perceived size as a function of contrast and temporal modulation. *Vision Research, 23*, 71–82.

Harris, M. G. (1980). Velocity specificity of the flicker to pattern sensitivity ratio in human vision. *Vision Research, 20*, 687–691.

Harris, J. M., & Watamanuik, S. N. J. (1995). Speed discrimination of motion-in-depth using binocular cues. *Vision Research, 35*(7), 885–896.

Harris, J. M., McKee, S. P., & Watamanuik, S. N. J. (1998). Visual search for motion-in-

depth: Stereomotion does not 'pop out' from disparity noise. *Nature, 1*(2), 165–168.
Gillam, B., Balackburn, S., & Nakayama, K. (1999). Stereopsis based on monocular gaps: Metrical encoding of depth and slant without matching contours. *Vision Research, 39*, 493–502.
Gray, R., & Regan, D. (1998). Accuracy of estimating time to collision using binocular and monocular information. *Vision Research, 38*(4), 499–512.
Harris, J. M., & Dean, P. J. A. (2003). Accuracy and precision of binocular 3-D motion perception. *Journal of Experimental Psychology, 29*, 869–881.
Harris, J. M., Neefs, H. T., & Grafton, C. E. (2008). Binocular vision and motion in depth. *Spatial Vision, 21*, 531–554.
Zeki, S. M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *Journal of Physiology, 236*, 549–573.
Horward, I. P., & Rogers, B. J. (2002). Seeing in depth, Vol. 2: Depth perception. University of Toronto Press. Toronto, Canada.
Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology, 195*, 215–243.
Julesz, B. (1960). Binocular depth perception of computer-generated patterns. *Bell System Technical Journal, 39*, 1125–1162.
Julesz, B. (1971). *Foundations of cyclopean perception.* Chicago, IL: University of Chicago Press.
Kaestner, M., Maloney, R. T., Wailes-Newson, K. H., Bloj, M., & Harris, J. M. (2019). Asymmetries between achromatic and chromatic extractions of 3D motion signals. Proceedings of the National Academy of Sciences of the United States of America.
Lages, M., Mamassian, P., & Graf, E. W. (2003). Spatial and temporal tuning of motion in depth. *Vision Research, 43*, 2861–2873.
Livingstone, M. S., & Conway, B. R. (2003). Substructure of direction-selective receptive fields in macaque V1. *Journal of Neurophysiology, 89*, 2743–2759.
Mandler, M. B., & Makous, M. (1984). A three channel model of temporal frequency perception. *Vision Research, 24*, 1881–1887.
Masson, G. S., Busettini, C., & Miles, F. A. (1997). Vergence eye movements in response to binocular disparity without depth perception. *Nature, 389*, 283–286.
Movshon, J. A., & Newsome, W. T. (1996). Visual response properties of striate cortical neurons projecting to are MT in macaque monkeys. *The Journal of Neuroscience, 16*(23), 7733–7741.
Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature, 401*, 695–698.
Nishida, S., Ledgeway, T., & Edwards, M. (1997). Dual-multiple-scale processing for motion in the human system. *Vision research, 37*, 2685–2698.
Peng, Q., & Shi, B. E. (2014). Neural population models for perception of motion in depth. *Vision Research, 101*, 11–31.
Pianta, M. J., & Gillam, B. J. (2003). Monocular gap stereopsis: Manipulation of the outer edge disparity and the shape of the gap. *Vision Research, 43*, 1937–1950.
Pianta, M. J., & Gillam, B. J. (2003). Paired and unpaired features can be equally effective in human depth perception. *Vision Research, 43*, 1–6.
Portfors-yeomans, C. V., & Regan, D. (1997). Cyclopean discrimination thresholds for the direction and speed of motion in depth. *Vision Research, 36*(20), 3265–3279.
Regan, D., & Spekreijse, H. (1970). Electrophysiological correlate of binocular depth perception in man. *Nature, 225*, 92–94.
Regan, D., & Beverley, K. I. (1975). The relation between discrimination and sensitivity in the perception of motion in depth. *Journal of Physiology, 249*, 387–398.
Regan, D., & Cynader, M. (1982). Neurons in cat visual cortex tuned to the direction of motion in depth: Effect of stimuli speed. *Investigate Ophthalmology & Visual Science, 22*, 535–550.
Rokers, B., Cormack, L. K., & Huk, A. C. (2008). Strong percepts of motion through depth without strong percepts of position in depth. *Journal of Vision, 8*, 1–10.
Rokers, B., Cormack, L. K., & Huk, A. C. (2009). Disparity- and velocity- based signals for three-dimensional motion perception in human MT+. *Nature Neuroscience, 12*, 1050–1055.
Rust, N. C., Schwartz, O., Movshon, J. A., & Simoncelli, E. P. (2005). Spatiotemporal elements of macaque V1 receptive fields. *Neuron, 46*, 945–956.
Sanada, T. M., & DeAngelis, G. C. (2014). Neural representation of motion-in-depth in area MT. *The Journal of Neuroscience, 34*, 15508–15521.
Sabatini, S. P., & Solari, F. (2004). Emergence of motion-in-depth selectivity in the visual cortex through linear combination of binocular energy complex cells with different ocular dominance. *Neurocomputing, 58–60*, 865–872.
Sakano, Y., Allison, R. S., & Howard, I. P. (2012). Motion aftereffects in depth based on binocular information. *Journal of Vision, 12*(11), 1–15.
Smith, A. T., & Edgar, G. K. (1993). Antagonistic comparison of temporal frequency filter outputs as a bias for speed perception. *Vision Research, 34*, 253–265.
Smith, A. T., & Edgar, G. K. (1994). Antagonistic comparison of temporal frequency filter outputs as a basis for speed perception. *Vision Research, 34*(2), 253–265.
Shioiri, S., Saisho, H., & Yaguchi, H. (2000). Motion in depth based on inter-ocular velocity differences. *Vision Research, 40*, 2565–2572.
Shioiri, S., Morinaga, A., & Yaguchi, H. (2002). Depth perception of moving objects. In B. Javidi, & F. Okano (Eds.). *3D television, video and display technology* (pp. 397–427). Berlin: Springer-Verlag.
Shioiri, S., Nakajima, T., Kakehi, D., & Yaguchi, H. (2008). Differences in temporal frequency tuning between the two binocular mechanisms for seeing motion in depth. *Journal of the Optical Society of America: Optics, Image Science, and Vision, 25*, 1574–1585.
Shioiri, S., Kakehi, D., Tashiro, T., & Yaguchi, H. (2009). Integration of monocular motion signals and the analysis of interocular velocity differences for the perception of motion-in-depth. *Journal of Vision, 9*(10), 11–17.
Shioiri, S., Matsumiya, K., & Matsubara, K. (2012). Isolation of two binocular mechanisms for motion in depth: A model and psychophysics. *Japanese Psychological Research, 54*, 16–26.
Shioiri, S., Yoshizawa, M., Ogiya, M., Matumiya, K., & Yaguchi, H. (2012). Low-level motion analysis of color and luminance for perception of 2D and 3D motion. *Journal of Vision, 12*(6), 1–14.
Shipp, S., & Zeki, S. (1989). The organization of connections between area V5 and V1 in macaque monkey visual cortex. *European Journal of Neuroscience, 1*(4), 309–332.
Sincich, L. C., & Horton, J. C. (2003). Independent projection streams from macaque striate cortex to the second visual area and middle temporal area. *The Journal of Neuroscience, 23*(13), 5684–5692.
Spileers, W., Orban, G. A., Gulyas, B., & Maes, H. (1990). Selectivity of cat area 18 neurons for direction and speed in depth. *Journal of Neurophysiology, 63*, 936–954.
Tidbury, P. L., Brooks, K. R., O'Connor, A. R., & Wuerger, S. M. (2016). A systematic comparison of static and dynamic cues for depth perception. *Investigative Ophthalmology & Visual Science, 57*, 3545–3553.
Tseng, C. H., Gobell, J. L., Lu, Z.-L., & Sperling, G. (2006). When motion appears stopped: Stereo motion standstill. *Proceedings of National Science Academy, USA, 103*, 14953–14958.
Tyler, C. W., & Cavanagh, P. (1991). Purely chromatic perception of motion in depth: Two eyes as sensitive as one. *Perception and Psychophysics, 49*(1), 53–61.
Watanabe, S., Yasuoka, S., & Fujita, I. (2008). Disparity-energy signals in perceived stereoscopic depth. *Journal of Vision, 8*, 1–10.
Watson, A. B., & Ahumada, A. J., Jr (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A, 2*, 322–341.
Welchman, A. E., Tuck, V. L., & Harris, J. M. (2004). Human observers are biased in judging the angular approach of a projectile. *Vision Research, 44*, 2027–2042.
Welchman, A. E., Lam, J. M., & Bülthoff, H. H. (2008). Bayesian motion estimation accounts for a surprising bias in 3D vision. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 12087–12092.