Voxel-wise binocular energy models for binocular disparity decoding

Hongna Zheng¹, Maoming Chen², Li Yao¹,², and Zhiying Long²*

¹ School of Artificial Intelligence, Beijing Normal University, Beijing, 100875, China.
² State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, 100875, China.
*Corresponding author’s e-mail: zlong@bnu.edu.cn

Abstract. Binocular disparity is a powerful cue for depth perception in three-dimensional (3D) space. Some neurophysiological studies proposed the binocular phase-shift and position-shift energy models to predict the responses of individual disparity-tuned neurons in cats and macaques. By far, it is unclear how to use binocular energy models to characterize the voxels’ responses in human visual cortex. In this study, we introduced the binocular energy models to the functional magnetic resonance imaging study and constructed the position-shift receptive-field model (Position-RFM) and the phase-shift receptive-field model (Phase-RFM) to predict voxel responses to disparity and to identify novel disparity levels from voxel responses. The results revealed that Phase-RFM outperformed Position-RFM in fitting the voxel responses for all the visual regions. Moreover, the novel disparity levels can be better identified from voxel’s responses in visual regions by Phase-RFM than Position-RFM. The findings may suggest that Phase-RFM can better encode the responses of disparity-tuned neuron populations than Position-RFM for human visual regions.

1. Introduction
Horizontal separation of human eyes results in the left and right eyes seeing an image at different locations. The differences in the two retinal images form binocular disparity, which is a powerful cue of depth perception in three-dimensional (3D) space [1]. Characterizing how the neurons in the visual system encode binocular disparity is important for understanding the neural mechanisms of stereopsis.

Neurophysiological studies have determined that neurons in the visual cortex in many mammals are sensitive to binocular disparity [2, 3]. The binocular disparity energy model proposed by Ohzawa et al. (1990) can explain the responses of individual disparity-tuned neurons. It was demonstrated that the predictions of this model match the measured data of cats and macaques [4, 5]. Although the model cannot well fit biological data in some studies [6, 7], it is still good for describing the neural encoding of binocular disparity.

The functional magnetic resonance imaging (fMRI) technique that indirectly reflects local pooled neural activity has been used to investigate the neural mechanism of the binocular disparity of the human brain. Many human brain regions, including V1, V2, V3, V3A, V3B, V7, hMT+/V5 (human motion complex), hV4 (human area V4), the lateral occipital cortex, parietal cortex, and premotor cortex, were found to engage in binocular disparities [8-10]. Although many previous studies have revealed the neural mechanism underlying binocular disparity processing using fMRI, it is unclear how to characterize the responses of disparity-tuned voxels in the human brain and build the relationship between 3D visual stimuli and fMRI activities.
In our previous study, the voxel-wise Gabor model was constructed to build a computational map between the disparity levels and the voxel responses to characterize voxels’ responses to disparity levels in the human cortex [11]. Due to the lack of receptive-field information of neurons in the model, Gabor model does not characterize the relationship between 3D visual stimuli and voxel response directly. The receptive field encoding model was demonstrated to be able to characterize the relationship between 2D visual stimuli and fMRI activity[12]. Because the binocular energy model contains receptive fields in each eye, it is essential to investigate how to use a binocular energy model to characterize the map between 3D visual stimuli and fMRI activity, and reveal the neural encoding mechanism of voxel population responses to binocular disparity.

In this study, we constructed two voxel-wise binocular receptive field energy encoding models that characterize the relationship between 3D visual stimuli and fMRI activity to predict the voxel responses and identify the novel disparity levels. The two voxel-wise binocular energy encoding models were the position-shift receptive field model (Position-RFM) and the phase-shift RFM model (Phase-RFM). Moreover, we compared the encoding and decoding performance of the Position-RFM, Phase-RFM. In contrast to Position-RFM, Phase-RFM showed a better accuracy in predicting the responses of disparity-tuned neuron populations and identifying the novel disparities from fMRI activities for all regions of interest (ROIs).

2. Materials and Methods
The fMRI datasets of eight subjects (6 males, aged 24-26 years) used in this study, the detailed experimental design was described in our previous study [11]. The disparity levels of 3D visual stimulus included ± 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, and 36 arcmin with a ±1 arcmin jitter.

The functional images were preprocessed and the 15 ROIs included V1, V2d, V3d, V3A complex, V7, V2v, V3v, hV4, LO, MTC (MT/V5 cluster), POIPSd, DIPSMd, DIPSAd, phAIPd, and LOC were defined in the same way as our previous study.

2.1. Binocular phase-shift and position-shift energy model
Complex cells have been modeled as energy neurons that sum the squared responses of a quadrature pair of linear neurons that are 90 deg out-of-phase [13]. Let L(x) and R(x) be the complex-valued linear responses for the left eye and right eye, respectively. The energy response E(x) of a binocular energy neuron can be expressed by equation (1) [13].

\[
E(x) = (Re(L(x)) + Re(R(x)))^2 + (Im(L(x)) + Im(R(x)))^2
\]  

(1)

For a quadrature pair of binocular linear neurons, the responses to a stereo image pair \( I(x, y) \) and \( I(x + D, y) \) can be written as

\[
Re(L(x)) + Re(R(x)) = \int \{f_l(x, y)I(x, y) + f_R(x, y)I(x + D, y)\}dxdy \quad (2)
\]

\[
Im(L(x)) + Im(R(x)) = \int \{f_l^*(x, y)I(x, y) + f_R^*(x, y)I(x + D, y)\}dxdy \quad (3)
\]

The phase-shift model assumes that the receptive fields between left and right eyes have different phases, but centers at the same retinal position. Let \( \Delta \varphi \) represent the phase shift between the left and right eyes’ receptive fields of a binocular neuron. In the phase-shift model, the left and right receptive fields of a quadrature pair of binocular linear neurons can be expressed as

\[
f_L(x, y) = f_L(x, y, \varphi) \quad f_R(x, y) = f_R(x, y, \varphi + \Delta \varphi)
\]

\[
f_L^*(x, y) = f_L^*(x, y, \varphi) \quad f_R^*(x, y) = f_R^*(x, y, \varphi + \Delta \varphi)
\]  

(4)

(5)

The binocular energy function in equation (1) of the phase-shift model has a peak for a particular receptive field phase shift that is related to the stimulus disparity [13].

The position-shift model assumes that the left and right receptive fields of a binocular neuron have identical shapes, but there is an overall shift between their horizontal positions. Suppose a binocular energy neuron whose right receptive field is shifted by a distance \( d \) compared to the retinal position of the left receptive field. In the position-shift model, the left and right receptive fields of a quadrature pair of binocular linear neurons can be expressed as
\[
f_{L,\text{pos}} = f_L(x, y) \quad f_{R,\text{pos}} = f_R(x + d, y) \quad f'_{L,\text{pos}} = f'_L(x, y) \quad f'_{R,\text{pos}} = f'_R(x + d, y)
\]  
(6)

The binocular energy function of the position-shift model has a peak when the receptive field position shift equals the stimulus disparity [13].

### 2.2. Estimation of the voxel-wise Phase-RFM and Position-RFM encoding models

We constructed simplified Phase-RFMs by using a fixed spatial frequency, orientation, and phase of the Gabor wavelets. The spatial frequency of the Gabor wavelets was set to 16 cycles/FOV (field of view), and the orientation was set to 0°. The FOV of each 3D image is 128×128 mm². Moreover, the phases of \( f_{L,\text{pha}} \) and \( f_{R,\text{pha}} \) were set to 0°, and those of \( f'_{L,\text{pha}}\) and \( f'_{R,\text{pha}}\) were set to 90°. The phase differences \( \Delta \phi \) were set to ±20°, 40°, 60°, 80° and 0°. 9 quadrature pairs of Gabor wavelets corresponding to 9 phase differences between the left and right receptive fields were used.

For the model-fitting runs, the images with fine disparities (±6, 12, and 18 arcmin) were used in the fine Phase-RFM estimation, and the images with coarse disparities (±24, 30, and 36 arcmin) were used in the coarse Phase-RFM estimation. The projections for each quadrature pair of Gabor wavelets of the two eyes are squared and summed based on equation (1), yielding a measurement of the energy response.

For each voxel in each ROI, the energy responses of 9 quadrature pairs of Gabor wavelets were linearly weighted and summed by using the equation \( r = EW^T \), where \( r_{N \times 1} \) is the voxels' responses to different disparities and can be estimated by the general linear model (GLM), \( E_{9 \times 9} = \begin{bmatrix} e_1, e_2, \ldots, e_9 \end{bmatrix} \) contains 9 energy responses vectors, \( W = [w_1, w_2, \ldots, w_9] \) contains the weights of the energy responses, and \( N \) represents the number of training samples. Lasso regulation was used to estimate the regression weight \( W \) by minimizing the cost function in equation (7).

\[
J(W) = \frac{1}{2} \| r - EW^T \|_2^2 + \lambda \| W \|_1
\]

where, hyperparameter \( \lambda \) controls the amount of penalization. Cross-validation was performed to determine the optimal regularization parameter \( \lambda \).

Simplified Position-RFM was constructed using the same spatial frequency, orientation, and phase of the Gabor wavelets as the Phase-RFM. The position shift \( d \) between the left and right receptive fields was set to ±1, 2, 3, 4 for the fine Position-RFM and ±3, 4, 5, 6 for the coarse Position-RFM. Thus, 8 quadrature pairs of Gabor wavelets corresponding to 8 position shifts between the left and right receptive fields were used. The Position-RFMs were estimated using the same method as the Phase-RFMs.

### 2.3. Disparity identification

Disparity identification was performed in the same way as in our previous study [11]. All 3D images with the same disparity level in the model-testing runs were averaged. Thus, 12 mean images with 12 disparity levels were generated. For each voxel, the 12 mean images were inputted into the Phase-RFM and Position-RFM to predict the responses to the 12 novel disparity levels in the model-testing runs.

To identify the disparity level of a mean image in the testing runs, the Pearson correlation between the observed pattern that was induced by the disparity level of the mean image and the 12 predicted patterns was calculated. The predicted pattern that showed a maximum correlation with the observed pattern was selected as the best matching. Therefore, the disparity level of each observed pattern was identified as the disparity level of the predicted pattern with the best matching.

### 2.4. Model Prediction Error

For each ROI, the first 150 voxels were selected to calculate the model prediction errors of the Phase-RFM and Position-RFM. The model prediction error of each voxel was defined as the L2 norm of the difference between the predicted voxel's responses and the measured voxel's responses (beta-values) to all of the disparity levels in the model-testing runs. The model prediction error of each ROI was calculated by averaging the prediction errors of the 150 selected voxels for each encoding model.

Moreover, For the identification accuracy and model prediction error a two-way repeated measures ANOVA test using ROIs (15 levels) and the encoding model (two levels) as the two within-subject factors were applied to reveal the main effects and the interaction effect between the two factors.
3. Results

Figure 1 shows each ROI's mean identification accuracies of the Phase-RFM and Position-RFM. The results of the two-way repeated measure ANOVA revealed a significant model effect ($F[1, 7] = 78.346$, $p < 0.001$), it means the identification accuracies of Phase-RFM was significantly higher than Position-RFM. A further simple effect analysis was performed to examine the differences between each pair of ROIs/models. The results of the simple effect were corrected by the Bonferroni method. In contrast to the Position-RFM, the Phase-RFM produced a significantly higher mean identification accuracy in most ROIs, except for phAIpd. The better identification performance of Phase-RFM in this study is consistent with the findings of previous studies and further supports the existence of phase disparity-tuned neurons in biological neural systems[13].

Figure 2 shows the mean model prediction errors of each ROI of Phase-RFM and Position-RFM. The results of the two-way repeated measure ANOVA revealed a significant ROI effect ($F[2.913, 20.393] = 4.689$, $p=0.013$), a significant model effect ($F[1, 7] = 90.618$, $p < 0.001$), and a significant interaction effect ($F[2.823, 19.760] = 3.453$, $p=0.038$). A further simple effect analysis was performed to examine the differences between each pair of ROIs/models. The results of the simple effect were corrected by the Bonferroni method. For Position-RFM, V1, V3A complex, and V7 had significantly lower prediction errors than MTC and DIPSAd. Moreover, V1, V3A complex, V7, V2v and POIPSd had significantly lower prediction errors than phAIpd. In contrast to Position-RFM, Phase-RFM produced significantly lower mean prediction errors in all ROIs. The results indicated that the Phase-RFM predictions matched the voxels’ responses better than those of Position-RFM, that may be attributed to the robustness in the presence of neuronal variability[13].

Figure 1. Mean identification accuracies of all ROIs for Phase-RFM and Position-RFM across eight subjects. V3A* indicates the V3A complex. * indicates $p < 0.05$ (Bonferroni corrected). The error bars represent the standard deviation.
Figure 2. Mean model prediction errors of all ROIs of Phase-RFM and Position-RFM across eight subjects. V3A* indicates the V3a complex. * indicates p < 0.05, • indicates 0.05 < p < 0.1 (Bonferroni corrected). The error bars represent the standard deviation.

4. Conclusions
In this study, we explored how to characterize the relationship between 3D visual stimuli and voxel responses by constructing two voxel-wise decoding models, namely, Phase-RFM and Position-RFM. The two models were applied to identify the novel disparity levels from voxel responses. Phase-RFM showed significantly better encoding and decoding performances than Position-RFM for all ROIs. The results may suggest that the phase-shift mechanism is more appropriate for characterizing the disparity-tuned voxels in the human cortex than the position-shift mechanism.

Acknowledgements
This work was supported by the National Natural Science Foundation of China [61671067, 62071050].

References
[1] Parker, A. J. (2007) Binocular depth perception and the cerebral cortex. Nature Reviews Neuroscience, 8: 379.
[2] Barlow, H. B., Blakemore, C., Pettigrew, J. D. (1967) The neural mechanism of binocular depth discrimination. J Physiol, 193: 327-42.
[3] DeAngelis, G. C., Ohzawa, I., Freeman, R. D. (1995) Neuronal Mechanisms Underlying Stereopsis - How Do Simple Cells In the Visual-Cortex Encode Binocular Disparity. Perception, 24: 3-31.
[4] Cumming, B. G., Parker, A. J. (1997) Responses of primary visual cortical neurons to binocular disparity without depth perception. Nature, 389: 280-283.
[5] Ohzawa, I., DeAngelis, G. C., Freeman, R. D. (1990) Stereoscopic Depth Discrimination In the Visual-Cortex - Neurons Ideally Suited as Disparity Detectors. Science, 249: 1037-1041.
[6] Prince, S. J. D., Cumming, B. G., Parker, A. J. (2002) Range and mechanism of encoding of horizontal disparity in macaque V1. Journal Of Neurophysiology, 87: 209-221.
[7] Ohzawa, I., DeAngelis, G. C., Freeman, R. D. (1997) Encoding of binocular disparity by complex cells in the cat's visual cortex. Journal Of Neurophysiology, 77: 2879-2909.
[8] Backus, B. T., Fleet, D. J., Parker, A. J., Heeger, D. J. (2001) Human cortical activity correlates with stereoscopic depth perception. Journal of neurophysiology, 86: 2054-2068.
[9] Welchman, A. E., Deubelius, A., Conrad, V., Bültthoff, H. H., Kourtzi, Z. (2005) 3D shape perception from combined depth cues in human visual cortex. Nature neuroscience, 8: 820.
[10] Li, Y., Zhang, C., Hou, C., Yao, L., Zhang, J., Long, Z. (2017) Stereoscopic processing of crossed...
and uncrossed disparities in the human visual cortex. BMC neuroscience, 18: 80.

[11] Li, Y., Hou, C., Yao, L., Zhang, C., Zheng, H., Zhang, J., Long, Z. (2019) Disparity level identification using the voxel-wise Gabor model of fMRI data. Human brain mapping, 40: 2596-2610.

[12] Kay, K. N., Naselaris, T., Prenger, R. J., Gallant, J. L. (2008) Identifying natural images from human brain activity. Nature, 452: 352.

[13] Fleet, D. J., Wagner, H., Heeger, D. J. (1996) Neural encoding of binocular disparity: Energy models, position shifts and phase shifts. Vision Research, 36: 1839-1857.