Hierarchical visual processing of stimuli with varying complexity

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Yijun Ge

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Dr. Sheng He
Co-advisor: Dr. Daniel Kersten

June 2021

Acknowledgements

Foremost, I would like to express my sincere gratitude to my primary advisor Dr. Sheng He and co-advisor Dr. Daniel Kersten for all their guidance and support. I have learned so much from Sheng throughout the years. He has inspired me to pursue my academic goals, sharpen my thinking, and encouraged me to explore the unknown. I am also very honored to have had the opportunity to work with Daniel. He has unique insights and a great passion for scientific research and has provided me enormous support and insightful advice during the past two years. I could not have imagined having better advisors for my Ph.D. study.

Besides my advisors, I am very grateful to my other committee members Dr. Stephen Engel and Dr. Yang Zhang, for their helpful feedback and insightful comments.

My sincere thanks also go to Hongru Zhu (Johns Hopkins University) and Dr. Alexander Bratch, who made great contributions to the project in the chapter 4. I would not have been able to complete this study without their help and assistant.

I would like to thank all my colleagues and friends Chen Chen, Dr, Chencan Qian, Dr. Ruyuan Zhang, Dr. Quan Lei, Dr. Nilsu Atilgan, Dr. Yingzi Xiong, Yanjun Li, Dr. Juraj Mesik, Walter Wu, Xinyu Liu and Dr. Hao Zhou for all their valuable help and collaboration.

Finally, my deep and sincere gratitude to my parents and my wife, Zhouyuan Sun, for their unconditional love and continuous encouragement. Without their support in my back, I would never achieve where I am today.

Abstract

Our visual system samples external information, adjusts its sensitivity and constructs a stable representation of the world that allows us to perceive and interact with objects in our environments. Visual information with different levels of complexity is processed through the hierarchically organized visual cortical areas. This dissertation presents three studies exploring the neural processing of visual stimuli at different cortical levels associated with feedforward and feedback processes. Study 1 investigates whether cortical neurons adjust their sensitivity based on stimulus-driven feedforward or perception-related feedback signals when they are discrepant, using psychophysical and neuroimaging techniques. We found that feedback signals associated with perception dominantly contribute to neural sensitivity control. Study 2 explores the properties of viewpoint-independent spatiotopic reference frame transformation for simple and complex visual stimuli using a trans-saccade adaptation paradigm. The results showed that both simple (orientation) and complex (face gender) visual features could be transformed into the viewpoint-independent spatiotopic reference frame, even in the absence of visual awareness of the target objects. Study 3 examines the viewpoint-dependent and viewpoint-independent neural representation of a complex stimulus feature (human pose information in natural images) in distinct dimensions (2D vs. 3D) using representational similarity analysis of 7T-fMRI data. The results revealed a distributed neural representation encoding different aspects of human pose features, with the 3D viewpoint-independent pose information captured at the posterior superior temporal sulcus, and body viewpoint information mainly encoded near the extrastriate visual cortex. Together, these studies help us to understand the importance of feedback signals in cortical sensitivity control, the awareness-independent transformation of visual objects from retinotopic to spatiotopic reference frame, and the distributed representation of body pose features in the visual cortical hierarchy.

Table of Contents

List of Tables

List of Figures

Chapter 1. Overview

The human visual system obtains rich and dynamic information from our environment that enables our perception and supports our actions. The visual cortical areas are hierarchically organized both anatomically and functionally, from lower-level cortical areas (like primary visual cortex (V1), specialized for simple stimulus features like orientation and spatial frequency), to the intermediate-level (such as V4, tuned to shape and form) (Nandy, Sharpee, Reynolds, & Mitchell, 2013) and higher-level cortical areas (including inferotemporal (IT) cortex that are sensitive to the complex stimulus features like face and body). Across hierarchical cortical areas, visual information processing involves feedforward (bottom-up) and feedback (top-down) connections. The feedforward visual cortical processing begins in the V1, which receives subcortical input from the lateral geniculate nucleus (LGN), and ascend through a ventral pathway into the temporal lobe ('what/perception pathway', associated with object recognition) and through a dorsal pathway into the parietal and prefrontal cortex ('where/action pathway', associated with spatial locations, visually guided actions, and attentional control). On the other hand, the reciprocal feedback connections carry information about top-down predictions (Kveraga, Ghuman, & Bar, 2007), influences of attention (Noudoost, Chang, Steinmetz, & Moore, 2010), awareness (Ro, Breitmeyer, Burton, Singhal, & Lane, 2003) and behavior context (Gilbert & Li, 2013). Although accumulating evidence indicated that feedforward and feedback processes play important roles in visual processing, how they interact with each other in supporting our visual perception. In addition, how do our brain constructs a stable and viewpoint-independent representation of objects in our environment remains unclear. The neural representation of different visual features (from simple to complex) also requires critical attention.

This dissertation project investigated three (among many) impressive feats achieved by the visual system. First, a ubiquitous feature of the sensory nervous system is its ability to adapt to the state of the environment. We asked whether the feedforward or feedback-driven representation determines the outcome of cortical neuronal

adaptation when they are discrepant. Second, we have a stable representation of the visual world, despite the constant motion of our eyes and body. We studied whether the orientation and face information could be transformed into a viewpoint-independent reference frame and whether visual awareness is a prerequisite during this reference frame transformation. Third, as social animals, humans need to quickly estimate poses from others around us. We performed the representational similarity analysis using natural scene stimuli to delineate viewpoint-dependent and viewpoint-independent neural representation of human pose information in two- and three-dimensional space.

In general, the feedforward signal, which more directly represents the sensory input, is consistent with the feedback signal that is more tightly linked with the perceptual representation of the stimulus. However, sometimes the feedforward- and feedback-driven representation of the stimulus could be dissociated with each other. To study the relative contributions of feedforward and feedback signals to various aspects of cortical neural processing, we need tools and paradigms to probe and measure the corresponding neural responses. In this project, reported in Chapter 2, we addressed the question of whether cortical neurons adjust their sensitivity based on stimulus-driven feedforward or perception-related feedback signals. More specifically, we adopted the orientation adaptation paradigm to investigate whether adaption would be based on the original retinal or perceived stimulus orientation. A visual illusion, flash-grab effect (FGE), was used to dissociate the perceived and retinal orientation of the adapting stimulus. Results showed that the orientation adaptation is exclusively dependent on the perceived rather than the retinal orientation of the adaptor. The combined fMRI and EEG results also indicated that the perceived orientation of the FGE is indeed supported by feedback signals in the visual cortex.

With rich visual inputs, our brain builds representations of the external world which allow us to navigate through and interact with our environment. Despite continuous and frequent eye movements (up to three times per second), our perceptual

representation of the visual world remains stable. Given the retinotopic (coordinates centered on the retina) representation in the early visual cortex, the neural representation of the visual objects in their environment needs to be transformed into a viewpoint-independent spatiotopic (coordinates centered on the outside world) reference frame. In chapter 3, we reported a project investigating the properties of spatiotopic reference frame transformation for simple (orientation) and complex (face gender) visual stimuli using a trans-saccade adaptation paradigm. Results showed that both orientation and face gender adaptation occurred at the same spatiotopic location (but different retinotopic location). We further asked whether the reference frame transformation requires awareness of the target object. Interestingly, when the adapting stimuli were rendered invisible by continuous flash suppression (CFS), both tilt and face gender aftereffects could still be observed at the spatiotopic location. Thus, our results indicated that visual awareness of objects is not a prerequisite for their transformation to the spatiotopic reference frame.

Understanding visual processing eventually amounts to understanding the processing of daily visual scenes. In the past, the majority of vision research relied on using simplified artificial laboratory stimuli, which were based on the assumption that neural processing of visual stimuli could be understood based on the responses to simple constituent patterns of stimuli (Nelken, 2004). But recent studies showed that the responses to natural visual scenes might not simply be described by the combination of responses to simplified stimuli (Hasson & Honey, 2012). Using a large set of natural scene stimuli also has an advantage in studying the complex and high-level visual information (like human pose), compared to using limited simplified stimuli. Understanding human pose information is crucial for understanding other people's actions, emotions, and social interactions, but is also challenging because of high variations between body parts, and appearance changes due to occlusion, viewpoint, and lighting. In chapter 4, we investigate the neural representation of viewpoint-dependent and viewpoint-independent human pose information in two- and three-dimensional spaces using representational similarity analysis with 7T-fMRI Natural

Scene Dataset (NSD). The results showed that posterior superior temporal sulcus (pSTS) and supramarginal gyrus specifically encode the 3D viewpoint-independent pose information. We also found explicit encodings of body viewpoint information mainly near the extrastriate visual cortex.

To summarize, the experimental projects reported in this thesis addressed questions related to neural representations of visual stimuli at different cortical levels and associated with feedforward and feedback processes. Three major conclusions emerge from this thesis:

1). When the perceptual representation of a stimulus is dissociated with the retinal representation, cortical neurons recalibrate their sensitivity primarily based on the feedback signals associated with perception;

2). Both simple (orientation) and complex (face gender) visual features could be transformed from retinotopic to spatiotopic reference frame, even in the absence of visual awareness of the target objects;

3). Distributed neural representations encode the different aspects of human pose information (including 2D/3D viewpoint-dependent and viewpoint-independent).

Collectively, these results shed light on how feedforward and feedback visual processing contribute to neural sensitivity control, facilitate the interpretation of visual scenes, and enable the construction of a stable object representation in our environment.

Chapter 2

Adaptation to the feedback representation of illusory orientation produced from flash grab effect

Adaptation is a ubiquitous property of sensory systems. It is typically considered that neurons adapt to dominant energy in ambient environment to function optimally. However, perceptual representation of the stimulus, often modulated by feedback signals, sometimes do not correspond to the input state of the stimulus, which tend to be more linked with feedforward signals. Here we investigated the relative contributions to cortical adaptation from feedforward and feedback signals, taking advantage of a visual illusion, the Flash-Grab Effect, to disassociate the feedforward and feedback representation of an adaptor. Results reveal that orientation adaptation is exclusively dependent on the perceived rather than the retinal orientation of the adaptor. Combined fMRI and EEG measurements demonstrate that the perceived orientation of the Flash-Grab Effect is indeed supported by feedback signals in the cortex. These findings highlight the important contribution of feedback signals for cortical neurons to recalibrate their sensitivity.

This chapter is a reproduction of Ge, Y., Zhou, H., Qian, C., Zhang, P., Wang, L., & He, S. (2020). Adaptation to feedback representation of illusory orientation produced from flash grab effect. Nature communications, 11(1), 1-12.

INTRODUCTION

Though adaptation is typically considered to be neurons adjusting their sensitivity to accommodate to the state of the "world" (Colin W.G. Clifford & Rhodes, 2005; Schwartz, Hsu, & Dayan, 2007), it is necessarily the case that the state of the "world" is reflected in neural representations. However, neural processing involves both feedforward as well as feedback signals, typically with the feedforward signal more directly representing the proximal stimulus (Pizlo, 2001) while the feedback signal, influenced by spatiotemporal contextual factors, leading to the perceptual representation of the distal stimulus. In sensory information processing, contextual modulation and feedforward-feedback interactions are very common (Albright & Stoner, 2002; Gilbert & Li, 2013; Lamme & Roelfsema, 2000). An important unresolved question is whether the feedforward or feedback driven representation determines the outcome of cortical neuronal adaptation, especially when they are discrepant.

To address this question, it is necessary to dissociate the input feedforward signals from cortical feedback signals in the brain. A recently discovered visual illusion, Flash-Grab Effect (FGE) (Cavanagh & Anstis, 2013), provides such an opportunity. The FGE occurs when a bar is briefly flashed on the light-dark boundary of a sectored background moving back and forth, at the time-point of direction reversal of background motion. The "flashed" bar could be perceived as tilted by more than 10 degrees away from its original orientation, as what would be perceived without the moving background inducer (Cavanagh & Anstis, 2013).

Since the FGE can alter perceived orientation, an orientation-specific adaptation was adopted to investigate whether adaptation would be based on the original retinal or perceived orientation. The tilt-aftereffect (TAE) is a robust visual phenomenon that results from orientation selective adaptation of visual neurons (Jin, Dragoi, Sur, & Seung, 2005). After prolonged exposure to an adaptor slightly tilted from vertical, a

vertical test is perceived as tilted away from the adapting orientation (Gibson & Radner, 1937). The underlying mechanism of this aftereffect was thought to be that cortical orientation-selective neurons in the visual system adjust or recalibrate their sensitivity based on the prevalent orientation and contrast of incoming signals, often in a population coding context, and with the goal of achieving more efficient coding (Benucci, Saleem, & Carandini, 2013; Blakemore & Tobin, 1972; C. W.G. Clifford, Wenderoth, & Spehar, 2000; Colin W.G. Clifford, 2014; Colin W.G. Clifford & Rhodes, 2005; Fang, Murray, Kersten, & He, 2005; Forte & Clifford, 2005; Jin et al., 2005; Liu, Larsson, & Carrasco, 2007; Schwartz et al., 2007; Thompson & Burr, 2009).

Testing of the tilt aftereffect with the flash grab effect will inform us about the relative contribution to orientation adaptation from the input retinal orientation and the contextual modulated perceived orientation. However, for our goal, we would also need to establish a close link between the perceived orientation of FGE and the feedback signals. While previous neuroimaging experiments showed that the perceived orientation in the FGE could be decoded in the retinotopic cortex (Kohler, Cavanagh, & Tse, 2017), it remains unclear how the neural signals dynamically support the perceived orientation of the flashed bar (Hogendoorn, Verstraten, & Cavanagh, 2015). Thus, we performed high spatial and temporal resolution human brain imaging experiments to delineate the dynamic contribution of feedforward and feedback signals to the perceived orientation in FGE. As shown in the results section, we obtained strong evidence that the perceived orientation in FGE was indeed supported by feedback signals. With this link established, a demonstration of tilt-aftereffect from the perceived orientation would indicate that the feedback signals dominate cortical adaptation.

In the following sections, we first present behavioral data showing that perceived orientation dominates the tilt-aftereffect. Then, we show results from high spatial-temporal resolution measurements of the cortical representation of the perceived

orientation in FGE. The time-resolved EEG data and layer-resolved fMRI data provide clear evidence that the perceived tilt in FGE is driven by late onset feedback signals, primarily targeting the superficial layers of the retinotopic cortex. These results together strongly suggest that perceived orientation in FGE is supported by feedback signals in the early visual cortex, which dominate orientation-selective adaptation in spite of the available feedforward signal corresponding to the original orientation of the flashed bar stimulus on the retina.

RESULTS

TAE depends on the perceived rather than retinal input orientation

In two psychophysics experiments, we investigated the relative contribution of the perceived vs. retinal orientation of FGE to the tilt-aftereffect. In the first experiment, the adapting bars vertical at the retinal level were perceived as tilted away from vertical orientation; in the second experiment, the adapting bars tilted at the retinal level were perceived as vertical. In both experiments, the testing bars were presented around the vertical orientation.

Subjects viewed a pair of vertical bars that were repeatedly and briefly flashed on top of two patterned disks that oscillated clockwise and counter-clockwise, with the flashed bars presented at the moment of the rotation reversals. The adapting bars, which would be perceived as vertical if presented without the moving background inducer, were perceived as tilted away from vertical due to the FGE (Figure 2.1a, and left column of 1b). On each trial of the main experiment condition, subjects were presented with 11 flashes (10.6 s) of adaptation, followed by 33.3 ms of blank screen, then the test bars for 33.3 ms (Figure 2.1a). Subjects were asked to judge whether two test bars converged upward or downward using a two-alternative forced choice (2AFC) method. Three control adaptation conditions were also included in the experiment: (a) the vertical flashed bars only, without the rotating background disks; (b) the rotating background disks only; (c) tilted (5.71 degrees from vertical) flashed

bars only, without the rotating background disks. The four conditions were presented in separate blocks.

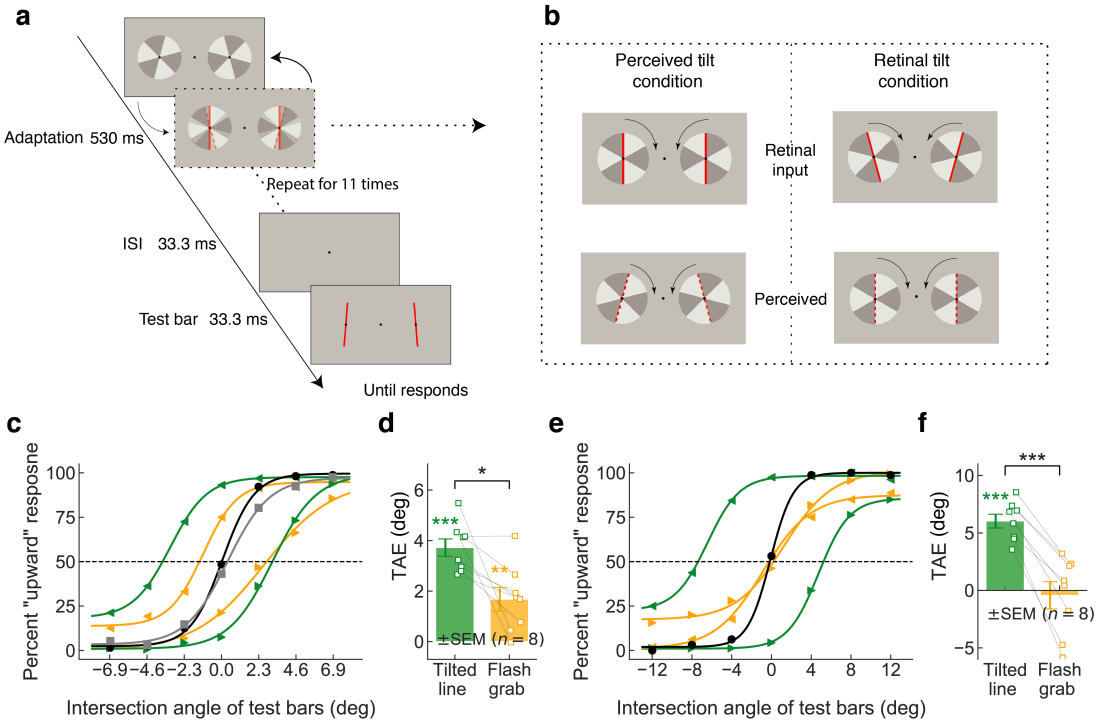

Figure 2.1. Psychophysics Stimuli and Results. (a) Stimulus presentation sequence for the TAE measurement. (b) The stimuli used in the two orientation adaptation conditions and demonstrations of subjects' illusory perception. Left part: original vertical but perceived tilt with flash grab inducer (tilt angle is by 15.55 degrees on average (SD = 7.54)); Right part: original tilt but perceived vertical with flash grab inducer. (c) Fitted psychometric functions using logistic regression in experiment 1; (d) Averaged TAE sizes across subjects (n=8) for conventional tilted bars condition ($p<0.001$ Holm-corrected) and flash grab condition ($p=0.010$ Holm corrected) in experiment 1 (two-sided paired t-test, $p=0.013$, $d=1.17$); (e) Fitted psychometric functions using logistic regression in experiment 2; (f) Averaged TAE sizes across subjects (n=8) for conventional tilted bars condition ($p<0.001$ Holm corrected) and flash grab condition ($p=0.730$ Holm corrected) in experiment 2 (two-sided paired t-test, $p<0.001$, $d=2.80$). Black curve: vertical bars only condition; grey curve: rotating background only condition; green curves/bars: flashed tilted bars only condition (conventional tilt adaptation); orange curves/bars: flash grab illusion condition. Error bars indicate standard errors of the mean (n=8 (individual subject)). Source data are provided as a Source Data file.

Results show that a significant TAE was generated by perceptually tilted bars: both the tilted bars without moving background and the FGE-induced tilted bars. Figure 2.1c shows the psychometric functions for adaptation to the FGE and the other three control conditions. Not surprisingly, there was no TAE in both the no background, vertical bar only condition ($p = 0.956$) and the background-only condition ($p = 0.534$). The strength of the TAE could be measured as half the difference on the x-axis between the two points of subjective equality (PSEs) following adaptation in two opposite orientations, i.e., the distance between the two green or two orange fitted curves (Equation (1)). Figure 2.1d plots the magnitude of the TAE from the flashed tilted bar adaptors (conventional TAE) and the TAE from the FGE condition. As expected, the conventional tilt adaptation condition generated strong TAE ($M = 3.72$ deg, $SD = 0.97$, $t(7) = 10.80$, Holm-corrected $p<0.001$, $d = 3.84$). The key result here is that a significant TAE was observed in the flash grab condition ($M = 1.67$ deg, $SD = 1.34$, $t(7) = 3.53$, $p = 0.010$ corrected, $d = 1.25$), though it was weaker than the conventional TAE (two-sided paired sample t-test, $t(7) = 3.31$, $p = 0.013$, $d = 1.17$).

The first experiment demonstrates that perceived tilted orientation could induce a TAE even though the input retinal orientation was vertical. Does the input orientation contribute to the TAE separately from the perceived orientation? To address this question, we tested subjects who adapted to bars with tilted input orientation but were perceptually vertical due to FGE (Figure 2.1b, right panel). At the beginning of this experiment, each individual subject adjusted the orientation of the flashed bars in FGE condition so that the bars were perceived as vertical. The adjusted retinal orientation was then set as the input orientation of adapting condition under FGE. Similar to the first experiment, we also included two control conditions: the vertical bars only condition and the tilted bars only condition.

Figure 2.1e/f shows the results of this experiment. The two control conditions generated results as expected: without the moving background inducer, the vertical bars by themselves did not generate the TAE ($p = 0.367$), and the tilted bars

generated very robust TAE (M = 6.02 deg, SD = 1.69, $t(7) = 10.08$, $p < 0.001$ corrected, $d = 3.56$). However, with the moving background inducer, the key result is that the originally tilted but perceptually vertical bars (due to FGE) generated no measurable TAE (M = -0.43 deg, SD = 3.36, $t(7) = -0.36$, $p = 0.730$ corrected), which is significantly weaker than the conventional TAE (two-sided paired sample t-test, $t(7) = 7.91$, $p < 0.001$, $d = 2.80$) as shown in Figure 2.1f.

Results from the two psychophysics experiments clearly show that, when the adaptor's perceived orientation is dissociated from its input orientation, the TAE is induced by the perceived rather than the input orientation itself. In other words, orientation selective adaptation seems to be primarily based on the eventual perceptual representation of the stimuli rather than simply on the neural representation directly linked to the input signals. To further understand the contribution of feedforward and feedback signals to FGE and in turn to orientation-selective adaptation, we conducted fMRI and ERP studies investigating the spatial and temporal neural correlates of the FGE.

Representation of FGE in the retinotopic visual cortex

We investigated the neural representation of the FGE in retinotopic visual areas in two fMRI experiments. The first experiment was conducted on a 3T scanner, with a focus on the retinotopic representation of the flashed bar under FGE. The second experiment was performed at high spatial resolution on a 7T scanner, which allowed us to obtain layer-resolved response signals to FGE in the retinotopic visual cortex. With known biases of feedforward and feedback signals in different cortical layers, the 7T data could inform us about the relationship between feedback signals and perceptual representation.

In the 3T fMRI experiment, we obtained the BOLD signal activated by the flashed bar in the FGE with block-designed fMRI scans (Figure 2.2a shows the stimuli and procedure of the experiment). Subjects' retinotopic maps were also obtained using the standard rotating wedge and expanding/contracting ring stimuli (Engel, Glover, &

Wandell, 1997) in two separate scans. The retinotopic map provides, for each voxel in the early visual cortex, the polar angle coordinate of its population receptive field. fMRI responses to the flashed bar for voxels with the same polar angle preference were averaged and used as the radial coordinate, plotted as a function of polar angle across the visual field (figure 2.2b). From V1 to V3, the fMRI response to the clockwise FGE was stronger in the upper right and lower left quadrants of the visual field in comparison with the counter-clockwise illusion, which showed stronger responses in the upper left and lower right quadrants. Therefore, the retinotopic representation of FGE in the early visual cortex is qualitatively consistent with the perceived tilt of the flashed bar. We further estimated the angular difference between the two polar angle representations of fMRI signals in the visual cortex. Note that the angular difference represents the summed effect of clockwise and counter-clockwise tilts. The estimated angular difference was smaller in V1 (17 and 13 degrees for upper and lower visual field, respectively) compared to V2 (41 and 27 degrees) and V3 (36 and 37 degrees) (Figure 2.2b). One-way ANOVA showed that the illusory effect significantly varied across visual cortical areas ($F(2, 16) = 22.24$, $p < 0.001$, $\eta_p^2 = 0.735$). Post hoc analysis showed that the illusory effect was significantly stronger in extra-striate than in striate visual cortex (for V2, $t(8) = 5.47$, $p < 0.002$, $d = 1.824$; for V3, $t(8) = 5.50$, $p = 0.002$, $d = 1.834$), while no significant difference was observed between V2 and V3 ($t(8) = 2.07$, $p = 0.072$). An important consideration is that BOLD responses reflected both the feedforward and feedback influences, and the reason for the smaller estimated tilt representation in V1 could be that V1 activity had a greater contribution from feedforward input signals (corresponding to the retinal orientation). The relative contribution of feedforward vs. feedback signals in different areas was investigated further with layer-resolved imaging (De Martino et al., 2015; Klein et al., 2018; Kok, Bains, Van Mourik, Norris, & De Lange, 2016; Muckli et al., 2015) as described in the following 7T high-resolution fMRI experiment.

In the follow-up 7T fMRI experiment, we obtained high-resolution layer-specific representation of the Flash Grab Effect in different layers of V1 to V3. The paradigm

was essentially the same as the 3T experiment, with the exception that the flashed bar was presented on the horizontal rather than vertical meridian due to the limited vertical field of view imposed by the 7T coil (Appendix Figure A1.2). In three independent scans, subjects were presented with a rotating bar (centered on the fixation point) to map the polar angle retinotopy of early visual areas (Engel et al., 1997). In the following layer-resolved analysis, the original fMRI data were resampled from 0.85 or 0.8 mm to 0.4 mm isotropic voxel size. Voxels were separated based on their distances from cortical surfaces into three separate layers: from 0% to 40% the superficial layers (S), from 40% to 80% the middle layers (M), and from 80% to 100% the deep layers (D) (Kok et al., 2016; Muckli et al., 2015; Wagstyl et al., 2018). Responses in each ROI to the clockwise and counter-clockwise tilted illusory orientations under FGE were plotted for voxels tuned to different orientations. For each layer (S, M, or D), there were two response curves, one corresponding to the perceived clockwise tilted and the other to the counter-clockwise tilted bars (Appendix Figure A1.3). To alleviate the bias of BOLD response towards superficial layers, the response curves were normalized across conditions within each cortical layer.

We calculated indexes that reflect the signal strength corresponding to the input meridian orientation and perceived tilted orientation respectively, for different layers and separately for V1, V2, and V3 based on the normalized response curves. Specifically, the index for the perceived orientation was calculated based on the mean BOLD response differences between two experimental conditions (clockwise vs. counter-clockwise) over the range of -14 to -6 and 6 to 14 degrees polar angles. The index for input meridian orientation signal was calculated based on the mean BOLD response between -4 to 4 degrees. As shown in Figure 2.3, the main effects are: 1) The representation index for the “illusory orientation” was significantly stronger in V2 and V3 than in V1 ($F(2, 32) = 9.72$, $p < 0.001$, $\eta_p^2 = 0.378$); in contrast, the strength of signal corresponding to the input horizontal orientation was much more robust in V1 than in V2 and V3 ($F(2, 32) = 27.76$, $p < 0.001$, $\eta_p^2 = 0.634$). 2)

More importantly, when signals were analyzed from different layers, the illusory representation varied significantly across layers in V1 ($F(2, 32) = 3.91$, $p = 0.030$, $\eta_p^2 = 0.196$).

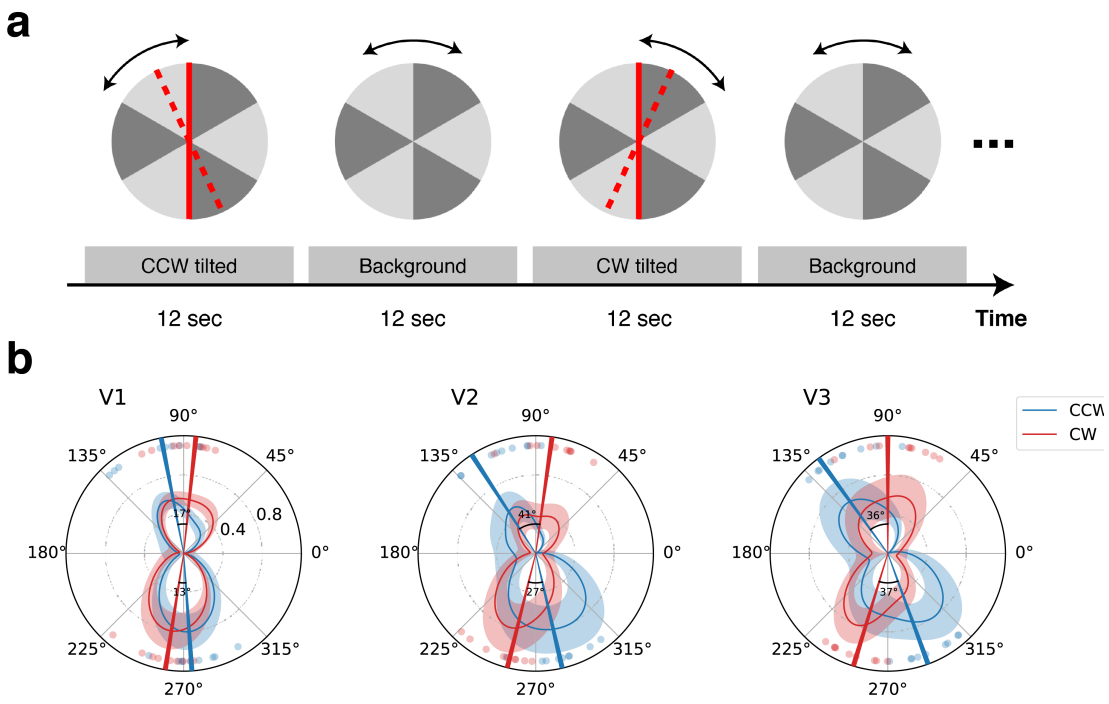

Figure 2.2. Stimuli and results of the 3T fMRI. (a) Schematic diagram of stimuli and procedures for the 3T fMRI experiment. A red bar flashed repeatedly for 12 seconds at the reversal point of the background motion, alternating with 12 seconds background only stimulation. The bar was presented at the vertical meridian but would be perceived as tilted clockwise or counter-clockwise from the vertical, depended on the direction of motion reversal. Red solid lines indicate the original position of the bar, while red dotted lines illustrate subjects' perception of the bar. (b) Polar angle representation of the flashed bar in FGE in the early visual cortex. Normalized fMRI response to the clockwise and counter-clockwise tilted illusions were plotted as a function of polar angle coordinates across the visual field. The minimal and maximum polar response for each subject were normalized to 0 and 1 (first subtracted the min and then divided by the max). Red and blue curves show the average polar response across subjects (low pass filtered by convolving with a 60-degree-width hamming window for illustration purposes). Shaded areas indicate standard errors of the mean (n = 9 (individual subject)). Red and blue bars illustrate the estimated average tilt from the curves, whereas the dots indicate the estimated tilt for individual subjects. Source data are provided as a Source Data file.

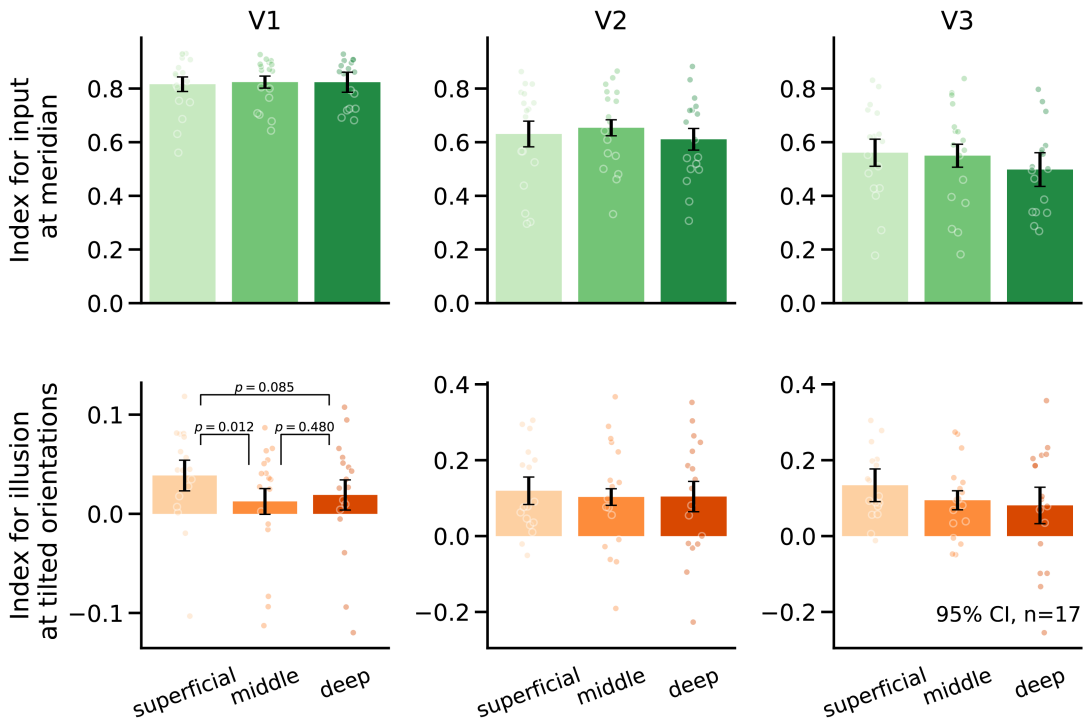

Figure 2.3. Representations corresponding to retinal input and illusory percepts across different layers of early visual cortices. fMRI response to the clockwise and counter-clockwise tilted illusions were calculated as a function of bar angle coordinates across the field of bar rotation. The computed response index for input representation at the horizontal meridian was based on the mean BOLD responses between -4 to 4 degrees polar angles, while the illusory index was based on the mean BOLD response differences between two experimental conditions (clockwise vs. counter-clockwise) at both -14 to -6 and 6 to 14 degrees polar angles. The representation index for the "illusory orientation" was significantly stronger in V2 and V3 than in V1 (two-way repeated measure ANOVA, $F(2, 32)=9.72$, $p<0.001$; Post-hoc Holm corrected, V1 vs. V2: $t(16)=-3.23$, $p=0.010$, $d=-0.783$, V1 vs. V3: $t(16)=-4.14$, $p=0.002$, $d=-1.005$, V2 vs. V3: $t(16)=0.31$, $p=0.755$), consistent with the 3T fMRI results; in contrast, the strength of signal corresponding to the input orientation (horizontal) was much stronger in V1 than in V2 and V3 (two-way repeated measure ANOVA, $F(2, 32)=27.76$, $p<0.001$; Holm corrected post-hoc, V1 vs. V2: $t(16)=4.84$, $p<0.001$, $d=1.176$, V1 vs. V3: $t(16)=-6.94$, $p<0.001$, $d=1.685$, V2 vs. V3: $t(16)=2.61$, $p=0.019$, $d=0.634$). When signals were analyzed from different layers, the illusory representation varied significantly across layers in V1 ($F(2, 32)=3.91$, $p=0.030$). Significant illusory representation was observed in V1 superficial layer ($t(16)=3.16$, $p=0.018$ Bonferroni corrected), but not in V1 middle layer ($t(16)=0.88$, $p=0.392$), and post hoc comparison showed that the illusory effect was significantly stronger in V1 superficial layer than in the middle layer ($t(16)=2.81$, $p=0.037$ Holm corrected). The statistical comparison across layers in V1 (one-way repeated measure ANOVA) was conducted by "within-subject" design. All the error bars represent within-subject 95% confidence interval of the mean index (n=17 (individual subject)). Source data are provided as a Source Data file.

Significant illusory representation was observed in V1 superficial layer ($t(16) = 3.16$, $p = 0.018$ Bonferroni corrected, Cohen's $d = 0.766$), but not in V1 middle layer ($t(16) = 0.88$, $p = 0.392$), and post hoc comparison showed that the illusory effect was significantly stronger in the superficial layer than in the middle layer ($t(16) = 2.81$, $p = 0.037$ Holm corrected, Cohen's $d = 0.682$). These layer-specific results indicate that the neural representation of FGE is primarily localized in the superficial layer for V1, but not the middle layer. This is consistent with previous studies showing that responses in the V1 middle layer reflect mainly bottom-up input signals, while responses in the V1 superficial layers are more related to feedback signals (Bastos et al., 2012; Felleman & Van Essen, 1991; Self, van Kerkoerle, Goebel, & Roelfsema, 2019; Self, van Kerkoerle, Supèr, & Roelfsema, 2013; van Kerkoerle, Self, & Roelfsema, 2017). In other words, the layer-resolved 7T data of FGE suggest that the representation of the perceived tilt was likely driven by feedback signals.

FGE correlates with late visual evoked potential signals

While the fMRI results suggest that early visual areas are closely involved in FGE representation, with the 7T layer-resolved data suggesting a dominant feedback contribution to the FGE, the temporal dynamics of feedforward and feedback processing in FGE remain unclear. Thus, we adopted EEG measurements to address this question.

Considering the limited spatial resolution of EEG, the flashed bar was only presented in the lower visual field so that perceptually with the influence of FGE, the flashed bar would fall onto either the left or right visual field (Figure 2.4b). This meant that an invoked ERP signal corresponding to the perceptual representation would be lateralized. In essence, the timing of the lateralized component in the ERP signal should indicate the timing of the neural representation of the perceptual effect. Trials with only a rotating background were included as a baseline condition, and trials with only a retinally tilted flashed bar without the rotating background were also included as a control condition. The orientation of the retinally tilted flashed bar were

individually adjusted to roughly match the perceived orientation in the FGE condition (Figure 2.4a).

Figures 2.4c and 2.4d show the differential ERP from posterior electrodes evoked by the contralateral versus ipsilateral bar in all three conditions. As expected, we observed a clear lateralized C1 component in retinally-tilted condition (Figure 2.4c), in response to the lateralized feedforward input. The cluster-based permutation test revealed an early positive peak (46-98ms) within C1 latency and a later negative peak (110-217ms). In contrast, after subtracting the background-only condition, no corresponding lateralized C1 was found in the illusory condition (Figure 2.4d), but only the later negative peak (118-161ms) remained, at which time window the rotating background generated a positive deflection. This is consistent with the lack of lateralized representation in the early visual cortex during the feedforward sweep.

We then performed multivariate pattern analysis to uncover the dynamic change of lateralized representation for the retinally or illusorily tilted stimuli from beyond the posterior electrodes. Linear classifiers were trained to predict whether the flashed bar was perceived to be tilted left or right at each time point (and for background-only trials, we were effectively predicting rotation direction). Retinally-tilted trials could be decoded significantly above chance about 50 ms after stimulus onset, reaching peak performance at C1 latency (Figure 2.4e). Illusory trials could also be successfully decoded starting from about 70 ms after stimulus onset. However, it outperformed the baseline condition (rotating background alone) only at a later stage, about 178 ms after stimulus onset (Figure 2.4f). We further characterized the nature of the lateralization information in the illusory condition using cross-decoding method. If the early lateralized representation before 100 ms reflected a mislocalized bar, similar to a retinally-tilted one, then a classifier trained using data from illusory condition in this time period should be able to decode data from retinally-tilted condition during C1 latency. The observed results did not support this hypothesis. Classifiers trained using illusory trials between 50-100 ms could not predict retinally-

tilted trials in the same period (Figure 2.5a, the decoding accuracy was actually significantly below chance level), but they did predict background-only trials significantly above chance level from 0 to around 150 ms (Figure 2.5b). Importantly, stimulus side in retinally-tilted trials between 50-100 ms could be predicted by classifiers trained using illusory trials between 180-220 ms (Figure 2.5a). This suggests that the lateralized representation for the illusory tilt appeared at a relatively late stage, and the early information about stimulus side was more closely associated with the rotating background.

We further asked whether and when lateralized EEG signals could predict the magnitude of the tilt perception in FGE. The inter-subject Pearson correlation between instantaneous amplitude of the difference wave (contralateral minus ipsilateral) in illusory condition (with background-only condition subtracted) and illusion size was calculated at each time point. Significant positive correlation emerged about 177 ms after stimulus onset (Figure 2.6a). Figure 2.6b, based on the same data as the shaded region in Figure 2.6a, more explicitly shows the clear relationship between the size of the perceptual illusion and the mean amplitude of the differential wave ($r(9) = 0.77$, $p = 0.004$). Notably, the onset of significant correlation matched well with the time when decoding performance in illusory condition overtook background-only condition, as well as when the scalp topography pattern in illusory condition became similar to that of retinally-tilted condition in C1 latency, convergently supporting that the main relevant component for the illusory effect appears rather late, consistent with the typical timing of feedback signals.

In contrast to a robust and clearly lateralized C1 signal from the retinally-tilted condition, no such lateralized signal was observed in the typical time window of C1 from the illusorily tilted bars under FGE. Only at a relatively late stage did the lateralized signal become prominent in the illusory condition, with its amplitude strongly correlated with the illusory effect size across individual subjects. These results support that the perceived tilt in FGE emerged later, likely a result of

feedback processing.

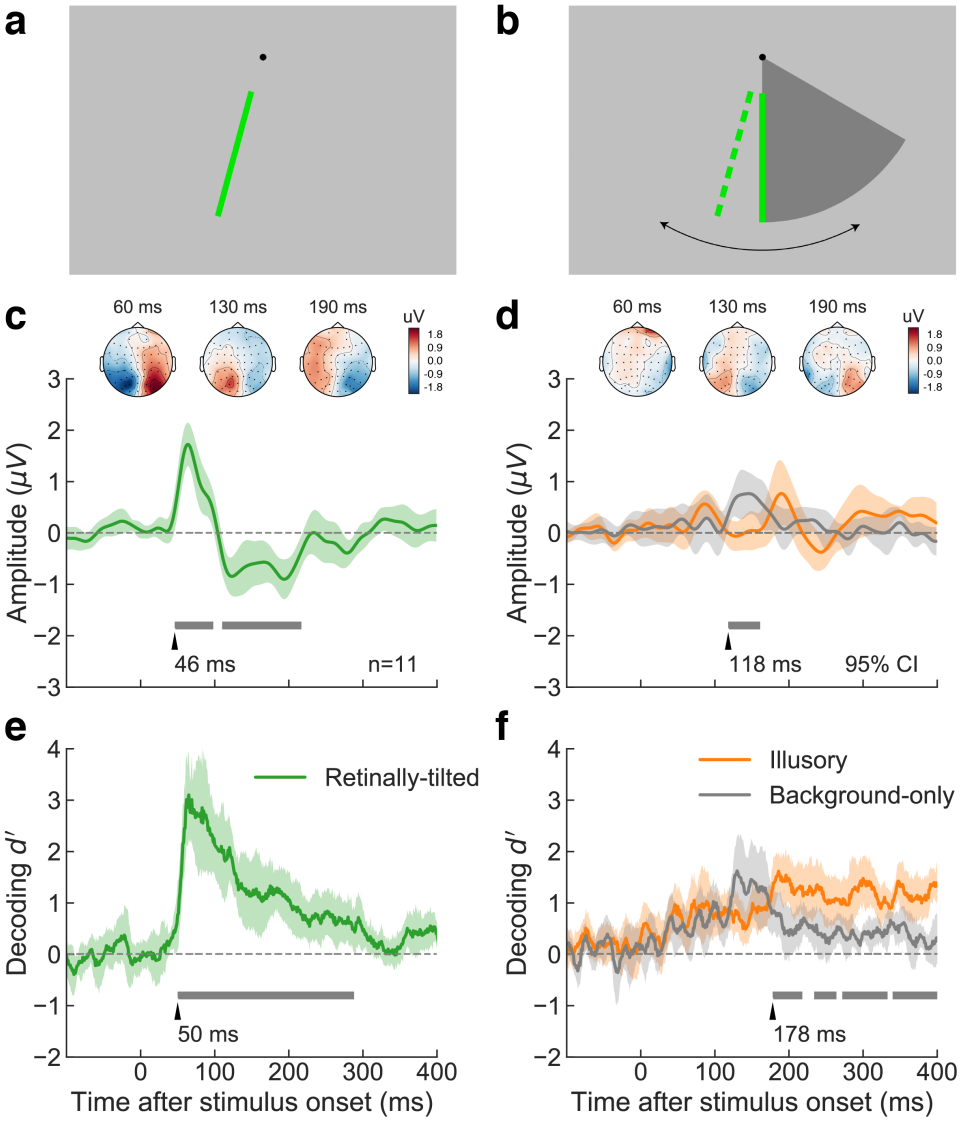

Figure 2.4. Stimuli and results of the EEG experiment. (a, b) Visual stimulus and perception (dashed bar) in tilted bar only (a) and illusory (b) conditions. (c, d) The differential ERPs (contralateral minus ipsilateral) averaged over five posterior electrodes and all subjects (n=12 (individual subject)) in the tilted bar only (green), illusory (orange), and background-only (gray) conditions. Insets show corresponding topography snapshots in three diagnostic time points, assuming the bar flashed to the left of the vertical meridian. Note the topography in (d) corresponds to differential wave between illusory and background-only conditions. (e, f) Cross-validation performance of linear classifiers trained at each time point to predict to which side the bar was flashed (green), was perceived (orange), or would have been perceived if flashed (gray) averaged across subjects. Gray bars indicate the significant time period after multiple comparison correction (p threshold = 0.05) Error bands indicate 95% confidence interval obtained by bootstrap. Source data are provided as a Source Data file.

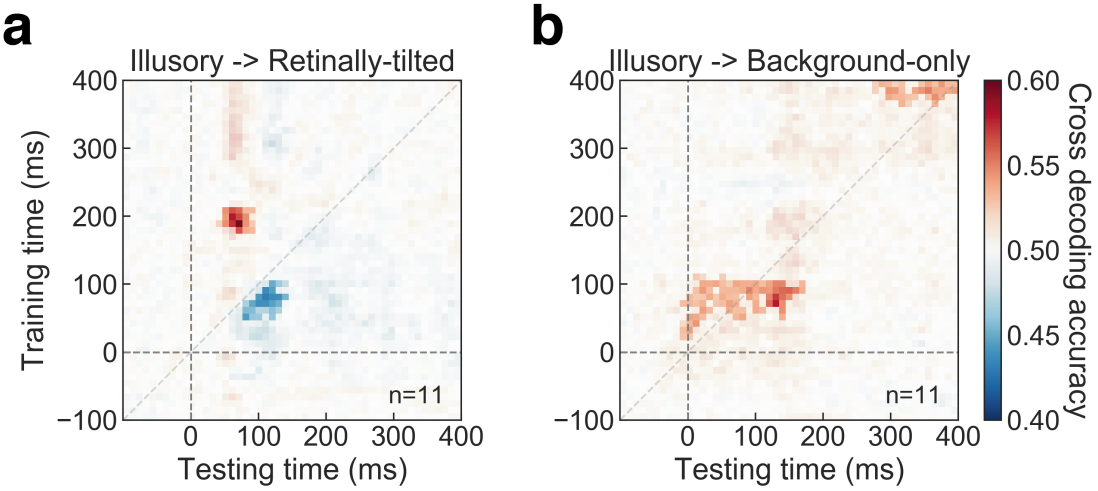

Figure 2.5. Cross-decoding analysis results. The color in each cell of the matrix indicates decoding accuracy if the classifier was trained with data from one time in one condition, and tested with data from another time in another condition. (a) Trained with illusory and tested with retinally tilted condition. (b) Trained with illusory and tested with background. The highlighted cells were significantly different (p threshold = 0.05) from chance level (50%) according to cluster-based permutation test, corrected for multiple comparisons. Source data are provided as a Source Data file.

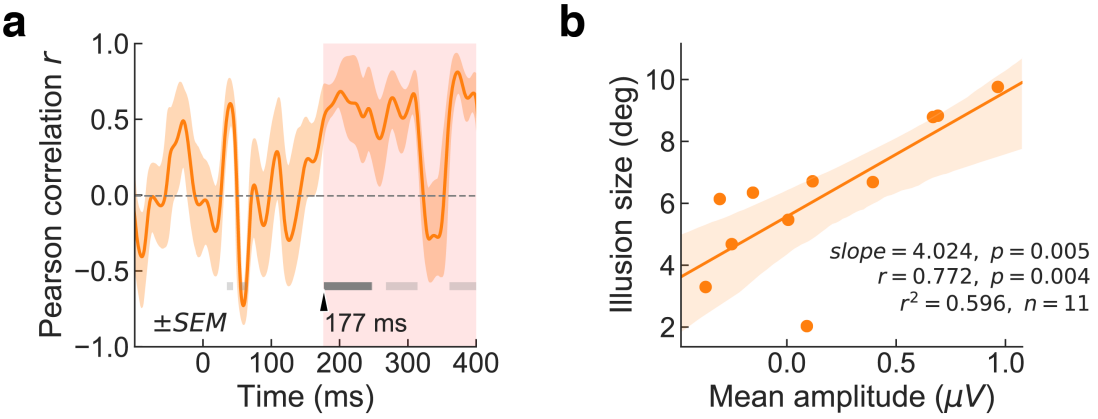

Figure 2.6. Inter-subject correlation Analysis. (a) The Pearson correlation between instantaneous amplitude of the difference wave (contralateral minus ipsilateral) in illusory condition (with background-only condition subtracted) and illusion size was calculated at each time point. Dark gray bars indicate period with significant correlation ($p < 0.05$, cluster-based permutation test). Light gray bars indicate period when the absolute value of correlation was greater than the cluster defining threshold $r = 0.5$. (b) The size of perceptual illusion (6.25 ± 2.35 deg) was well predicted by the mean amplitude, averaged within the shaded interval in (a). Source data are provided as a Source Data file.

Taken together, the fMRI results show that the distribution of fMRI BOLD signals in retinotopic visual cortical areas represented both the perceived and the input positions of the flashed bars. The 7T fMRI data further reveal that signals in the superficial layers were more influenced by the perceived illusory location of the flashed bars, especially in V1. Finally, a robust and behaviorally relevant lateralized EEG signature was only observed late in time, at around 170-180 ms after the onset of the flashed bars in the illusory condition. The combined spatio-temporal imaging results strongly suggest that the perceived tilt of the flashed bars in FGE was instigated by feedback signals.

DISCUSSION

The combined psychophysics, fMRI, and EEG results jointly support that cortical adaptation can be tuned to feedback-driven representations. In the case of orientation-selective adaptation investigated here, the tilt aftereffect was mainly dependent on the perceived illusory orientation from the FGE rather than the input orientation of the flashed bar. With spatiotemporal imaging results supporting a feedback origin of the perceived orientation in FGE, these results suggest that feedback signals play an important role in orientation adaptation and provide evidence that in the presence of discrepant feedforward and feedback supported representation of visual input, the feedback signal determines the adaptation outcome.

A recent fMRI decoding study showed that patterns of activation in early visual cortex could be used to classify the direction of perceived position shift of FGE (Kohler et al., 2017). Our study went beyond decoding and 1) generated direct estimates of the angular representations of FGE in early visual cortex (3T fMRI), 2) identified the relative contributions from different cortical layers to the perceptual illusion (7T fMRI), and 3) revealed that the neural correlates of the perceptual illusion arose relatively late (EEG). In addition, a noticeable aspect of the 3T fMRI results is

that BOLD signals showed stronger representation of the FGE in dorsal compared to ventral visual cortex (see in Appendix Figure A1.1). This might have resulted from asymmetric representation across the meridian of the visual field (Liu, Heeger, & Carrasco, 2006).

Perception has long been considered an inferential process (Hiebert, 1996; Pizlo, 2001), that retina inputs are modulated by spatiotemporal context and other priors to generate our perceptual experience. A number of neuroimaging studies have examined whether the neural signals in early visual cortex reflected the input properties or the perceived quality of the stimuli, with mixed results. Some studies showed that the BOLD signal in V1 reflected the perceived stimulus rather than the retinal input, such as activation reflecting distance scaling of perceived object size (Murray, Boyaci, & Kersten, 2006) and activation along apparent motion trajectory where there was no direct stimulation (Muckli, Kohler, Kriegeskorte, & Singer, 2005). Other studies have shown that local signals in V1 did not necessarily correspond to perceived brightness and color changes induced by modulating a surround field (Cornelissen, Wade, Vladusich, Dougherty, & Wandell, 2006). To reconcile the conflicting findings, an important point to consider is that BOLD responses are driven by both feedforward and feedback neural signals. In our study of FGE, the smaller estimated tilt angle based on fMRI signals in V1 could be due to a greater contribution from feedforward input signals in V1. In this regard, the layer-resolved 7T fMRI has a particular advantage, as shown in our results, in which the superficial layers tend to have more robust representations of the illusory tilt, compared to the middle layers that are more dominated by feedforward signals (De Martino et al., 2015; Kok et al., 2016; Muckli et al., 2015).

Across individuals, EEG signal lateralization about 180 ms after flash onset closely correlated with the magnitude of FGE. But while all subjects showed illusory tilt effect in consistent directions, the corresponding (contra-ipsi) lateralized ERP was not always positive, with subjects experiencing weaker illusion tending to have little or

reversed lateralization (Figure 2.6b). This is likely because the observed ERP during that interval was also influenced by other sensory and cognitive processes. For example, a stronger feedforward representation may induce a larger negative component in the P1/N1 range, reducing the potential lateralized ERP signals in the time window. Another interesting observation is that the background by itself induced a significant lateralized EEG signal at around 120 ms (Figure 2.4d), which was not observed when a vertically flashed bar was added to this background in the FGE condition. It is possible that the abruptly flashed bar attracted attention and reduced the signal from the rotating wedge background. Alternatively, it may have been canceled out by an oppositely lateralized signal from the perceived tilted bar, which means the illusory representation could have emerged as early as 120 ms after bar onset. The fact that the lateralized signal around 120 ms was not correlated with illusion size and did not outperform background-only condition in decoding implies that this signal was not intrinsically linked to the FGE. In any case, 120 ms is not typically considered in the temporal window of feedforward processing in early visual cortex. Overall, the temporal data strongly support a feedback interpretation of FGE.

An interesting observation is the below-chance level cross-decoding performance (from illusory to retinally-tilted condition) shown in Figure 2.5a. This was observed during a very early time window for the training stimulus. The implication is that the activity patterns of illusory (centered around 80 ms) and retinally-tilted (centered around 100 ms) trials were likely oppositely lateralized. It is possible the two patterns represented different features of the stimuli. Indeed, the activity patterns of the illusory condition around the same time window cross decoded the background-only condition significantly above-chance, suggesting that the former was more related to the moving background wedge (note that the wedge would always be at the opposite side of the perceived location of the flashed bar, Figure 2.4b). This below-chance decoding performance in the early time window of the illusory condition forms a clear contrast to the above-chance decoding in a later time window (~200 ms). Together,

they point to an early background based and late illusory bar position based cross-decoding performance.

With the results from spatiotemporal imaging supporting a feedback interpretation of the FGE, the behavioral data showing that the perceived tilt in FGE could generate a TAE implies that the visual cortical neurons adapted to orientation representation driven by the feedback signals. Given that the goal of adaptation is to adjust the system's sensitivity based on the statistics of the environment to process information more efficiently, this point becomes more interesting when the input driven feedforward representation and the feedback driven perceptual representation are in conflict and both are available in cortex. When input signals and perceptual representation agree, it is difficult to distinguish between adaptation to feedforward or feedback signals. Our previous demonstration that orientation-selective adaptation could occur to invisible gratings (S. He, Cavanagh, & Intriligator, 1996; Sheng He & MacLeod, 2001) constitutes support for adaptation to feedforward-dominated cortical representation of orientation. Our current results show that when the feedforward input orientation is different from perception, adaptation is primarily driven by the feedback-driven neural representation of the perceived property. These results also go beyond the demonstration of TAE from mentally generated bars (Mohr, Linder, Dennis, & Sireteanu, 2011; Mohr, Linder, Linden, Kaiser, & Sireteanu, 2009). Since no feedforward inputs were presented in those studies, there was no competition between the feedforward and feedback signals.

There were early experiments investigating the potential influence on adaptation effect resulting from dissociation between input and perceived properties of stimuli, with mixed results. For example, in the so-called flash-drag effect, where the perceived position of a flashed stimulus appears to be shifted in the direction of a nearby moving object, the perceived location biased the effectiveness of adaptation (Kosovicheva et al., 2012). However, other studies showed that those motion-induced position changes had little contribution to the adaptation aftereffect (Fukiage

& Murakami, 2010, 2013). The lack of clear results from these early studies could be due to weak adaptation effect (Fukiage & Murakami, 2010) or rather small size of perceptual mislocalization (Fukiage & Murakami, 2013). The FGE could induce a 10 times larger position shift compared with the flash-drag effect (Cavanagh & Anstis, 2013), by presenting the flashed target on top of the moving background at the time it reverses its motion trajectory, rather than adjacent to the moving object. The current results, with complete dissociation between retinal input and perceived orientation of the adapting stimuli, combined with the clear demonstration of the feedback origin of the perceptual effect, provide unequivocal evidence for neural adaptation to feedback representations.

Since information processing networks consist of both hierarchical stages and parallel pathways, naturally adaptation could occur at multiple stages of processing. Consequences of adaptation observed at later stages of processing could be based on inherited signals from other parts of the neural networks, or the adaptation effect could be itself inherited (Solomon & Kohn, 2014). For example, contrast adaptation effect could be observed in MT neurons or from the inheritance of contrast adaptation effect at early stages of processing (Kohn, 2007; Kohn & Movshon, 2003). Early studies have also demonstrated adaptation effect to biases in appearance in color and motion, which allowed the authors to conclude that these adaptation effects were cortical in origin (Goddard, Solomon, & Clifford, 2010; Krauskopf & Zaidi, 1986; Zaidi & Sachtler, 1991). In addition, attention could modulate the representational strength of attended features and in turn enhance its adaptation. While it is common that many factors modify the retinal input to generate perception, and these results are certainly consistent with adaptation to perception-linked neural representations, our current study has the advantage of explicitly contrasting the feedforward representation and feedback representation in their effectiveness for adaptation. Specifically, our study adds to the understanding of adaptation that when input signal and feedback representation are clearly different, the visual system can adjust its sensitivity based on the feedback-driven neural representation despite the discrepant feedforward representation. Although this point

is demonstrated with just one perceptual phenomenon here, our study prompts future neural adaptation models to take into account the different roles of feedforward and feedback signals, especially when they are discrepant.

In summary, our spatiotemporal imaging results reveal that the illusory orientation representation was temporally late and spatially biased to the superficial cortical layers, thus pointing to a feedback origin of the FGE. Combined with psychophysical results, this study provides evidence that when perceived and input stimulus orientations of the adapting bars are dissociated with each other, the orientation adaptation mainly depends on the feedback supported neural representation linked to perception. These results highlight the important contribution of feedback signals for cortical neurons to recalibrate their sensitivity.

METHODS

Participants

Eight healthy subjects (5 female, ages 21-27) participated in the psychophysics experiments; eleven (2 female, ages 21-27) participated in the 3T fMRI experiment (two subject was excluded due to head movement or failed to obtain clear retinotopy); seventeen (9 female, ages 22-35) participated in the 7T fMRI experiment; and twelve (4 female, ages 21-27) participated the EEG experiment (one subject was excluded due to excessive eye movement/blinks). Subjects were unaware of the purpose of the experiments. All observers had normal or corrected-to-normal vision and gave written consent. The protocol was approved by The Institutional Review Panel at the Institute of Biophysics (IBP), Chinese Academy of Sciences (CAS).

Psychophysics stimuli and procedures

Subjects' head position was stabilized with a chin-rest at a viewing distance of 57cm. Stimuli were presented in a dark room on a CRT monitor (NESO FS210A,

Nanchang, China), with a resolution of 1024×768 and a refresh rate of 120 Hz. The experiment was programed in MATLAB (The Math Works, Inc.) using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) extensions.

During the experiment, a small black fixation dot was presented at the center of the screen and a pair of rotating disks of 3.9 dva (degree of visual angle) radius were presented at the two sides of the fixation point, on a uniform gray background. The disks were patterned with 6-sectors (spanning 60 degree each sector). The distance between the fixation point and the center of each disk was 10.2 dva. The sectors had 25% Michelson Contrast (Michelson, 1995), which was defined by

$$C_m=((L_{max}-L_{min}))/((L_{max}+L_{min}))$$

Where L_{max} and L_{min} represent the luminance of brighter and darker sectors respectively.

The disks rotated 250° (degrees of rotation) every second and reversed direction every 240 ms (covering 60°, 1 sector, in that time). On each reversal a light-dark edge would be at the vertical orientation, and for every other rotation reversal (480 ms/cycle) two red vertical bars (0.3 dva width) were flashed on for 33 ms, aligned with the light-dark edges.

In the first experiment, we tested the tilt after effect to perceived tilted but retinally vertical condition. We first measured the size of the flash-grab effect. Subjects were presented with a pair of rotating sectored disks and two vertical bars were flashed briefly at the direction reversals. A pair of green pointers (0.3 dva) was presented around each of the two disks. Using the keyboard, the subjects adjusted the angles between the pointers until the pointers and bars appeared to be aligned. They had unlimited time to adjust the angles, and were asked to press the spacebar when they were satisfied with the angle alignment to record the setting and to start the next trial. The two rotation directions (left clockwise and right counter-clockwise, vice versa) were tested 5 times each for each subject. The mean perceived tilt (away from vertical) across subjects was 15.55° (n = 8, SD = 7.54) .

The adaptation trial sequence is depicted in Figure 2.1a. On each trial, subjects were presented with the same patterned disks as in the flash grab measurement part of the experiment and adapted to the two flash bars. The bars were perceived to be tilted due to the flash grab effect. The adaptation period included 11 flashes (5.3 s) in each trial, followed by a 33.3 ms blank period. Then a pair of test bars were presented for 33.3ms. The test bars were the same as the pair of red bars presented during the adaptation period except that the angle between two bars was varied ranging from -6.9° to +6.9° (7 variations, -6.9°, -2.3°, -1.1°, 0°, +1.1°, +2.3°, +6.9°, positive degree represents the two bars converging upward). Subjects were asked to judge whether the two test bars were converging upward or downward using a 2AFC method. The 7 different angular conditions of bars were tested 20 times each (selected in random order across trials).

Three control adaptation conditions were included in the experiment: (a) the vertical flashed bars only without the rotating background disks; (b) the rotating background disks only; (c) tilted flashed bars as in conventional TAE experiment (The bars were tilted 5.7 degrees away from vertical). The tilted flash bars conditions and the flash grab conditions are counterbalanced between blocks among the subjects.

In the second experiment, we tested the tilt after effect to perceived vertical but retinally tilted condition. The conditions were similar to that described above, except that subjects needed to adjust the reversal angle of disks until the two flashed bars appeared vertical using keyboard. Subjects had unlimited time to make the adjustment. When they were satisfied with the adjustment, they pressed spacebar to start another trial. Two rotation directions were tested 20 times each for each subject. The mean orientation away from vertical across subjects was 16.02° (n = 8, SD = 7.34) . The adaptation stimulus used in this experiment is demonstrated in Figure 2.1b (right column).

The tilt aftereffect was measured with similar procedure as described above, except that the adapting stimuli were retinally tilted but perceived vertical for each subject. Two control conditions were included as well, one is the vertical flashed bars without the background, and the other is the retinally tilted bars without the background as in conventional TAE experiments.

3T fMRI procedures and data acquisition

Stimuli were presented with an MRI safe projector (1024x768@60Hz) on a translucent screen behind the head coil. For the FGE experiment, the rotating pinwheel background (Figure 2.2a) was presented at 3.12% contrast, 36.87 degrees of visual angle in diameter, rotating at 180 degrees per second and changed motion direction every 0.67 seconds (120 degrees per rotation). A red vertical bar (36.87 and 0.96 degrees in length and width, respectively) was briefly presented for 67 ms at the boundary of two disc sectors, at the moment of background motion reversal. Subjects were instructed to keep fixation while passively viewed the stimuli. Four runs of functional data were collected for the FGE experiment, each consisted of 144 image volumes. Retinotopic localizer were rotating wedge and expanding ring checkerboard stimuli reversing contrast at 5 Hz. The wedge stimulus has a center angle of 22.5 degrees, rotating clockwise across the full visual field in 32 seconds. The ring stimulus expanded from fixation to the edge of the viewing aperture (47.93 degrees in diameter) in 32 seconds. Two runs of functional images were collected for the retinotopic localizer, 128 image volumes for each run.

MRI data were acquired with a 3T MRI scanner (Siemens Trio) using a 12-channel receive head coil at Beijing MRI Center for Brain Research (BMCBR), IBP, CAS. Functional images were acquired with a gradient echo planar imaging sequence (3 mm isotropic voxels, 30 axial slices of 3 mm thickness, 64×64 matrix with 3 mm in-plane resolution, TR/TE = 2000/28 ms, flip angle = 90°). High-resolution anatomical volume was obtained with a T1-MPRAGE sequence (1 mm isotropic voxels, 192

sagittal slices of 1mm thickness, 256×256 matrix with 1 mm in-plane resolution, TR/TE = 2600/3.02 ms, flip angle = 8°).

7T fMRI procedures and data acquisition

Viewing aperture of the 7T screen was 26.27 degrees horizontally and 19.85 degrees vertically. Fullfield rotating pinwheel background (Appendix Figure A1.2) was presented at 2.91% contrast, rotating at 240 degrees per second and changed motion direction every 0.5 seconds (120 degrees per rotation). A red horizontal bar (26.27° and 0.52° visual angle in length and width, respectively) was briefly presented for 67 ms at the boundary of two disc sectors, at the moment of reversal of background motion. Subjects were instructed to keep fixation while passively viewed the stimuli. Nine runs of functional images were collected for the FGE experiment, 144 volumes of images for each run. Retinotopic localizer was a rotating bar stimulus with checkerboard patterns reversing contrast at 5 Hz (26.27° and 0.52° visual angle in length and width, respectively). Centered on the fixation, the bar rotated counter-clockwise from -16 to +15 degrees in 32 seconds. Three runs of functional images were collected for the retinotopic localizer, each consisted of 128 volumes of images.

MRI data were acquired with a 7T whole body MRI scanner (Siemens Healthineers GmbH, Erlangen, Germany) using a 32 channels head coil (Nova Medical, Wilmington, USA) at BMCBR, IBP, CAS. For the first seven subjects, a reduced-FOV Gradient-echo EPI sequence was used to acquire functional images (0.85 mm isotropic voxels, 21 coronal slices of 0.85 mm thickness, 126 × 96 matrix with 0.85 mm in-plane resolution, TR/TE = 2000/21 ms, flip angle = 80°, 6/8 phase partial Fourier (GRAPPA acceleration factor 3). High-resolution anatomical volume was obtained with a T1-weighted MPRAGE sequence (0.7 mm isotropic voxels, 256 sagittal slices at 0.7 mm thickness, 320 × 320 matrix with 0.7 mm in-plane resolution, TR/TE = 3100/3.56 ms, TI = 1200ms, flip angle =5°) and a proton density or PD-weighted MPRAGE sequence (0.7 mm isotropic voxels, 256 sagittal slices at 0.7 mm

thickness, 320 × 320 matrix with 0.7 mm in-plane resolution, TR/TE = 2340/3.56 ms, flip angle = 5°). For the rest ten subjects, functional images were collected with a GE-EPI sequence with larger FOV (TR = 2000 ms, TE = 23 ms, 80° flip angle, voxel size 0.8 × 0.8 × 0.8 mm, FOV 128 × 128 mm, 31 oblique-coronal slices, 6/8 phase partial Fourier, GRAPPA acceleration factor 3). High-resolution anatomic volume was obtained with a T1-weighted MP2RAGE sequence (TR = 4000 ms, TE = 3.05 ms, voxel size 0.7 × 0.7 × 0.7 mm, field of view 224 × 224 mm, 256 sagittal slices, receiver bandwidth 240 Hz/pix, 7/8 phase partial Fourier, 7/8 slice partial Fourier, TI1 = 750 ms, 4° flip angle, TI2 = 2500 ms, 5° flip angle).

EEG procedures and data acquisition

Observers were tested individually in a dark testing room. Head position was stabilized with a chin rest at a viewing distance of 57 cm. Stimuli were presented on a CRT monitor (NESO FS210A, Nanchang, China) with a resolution of 800*600 and a refresh rate of 100 Hz. The experiment script was written in MATLAB (The Math Works, Inc.) using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) extensions.

As shown in Figure 2.4a and 2.4b, the screen was filled with a uniform gray background. A small, black fixation dot was 5.9 dva (degrees of visual angle) above the screen center and a 60-degree sector (6.3% contrast with background) of 15.6 dva radius rotated back and forth below the fixation point. The sector rotated 80° (degrees of rotation) every second and reversed direction every 1500 ms (covering 120°, from -60° to 60° around vertical meridian). When the reversal occurred, a green vertical bar (0.3 dva in width) might flash for 30 ms (3 frames) at the vertical meridian, aligning with one of the two edges of the sector.

In order to match the illusorily and retinally-tilted conditions, we first did a psychophysical experiment to measure the size of flash-grab effect. Within each trial, the flashed bar was always illusorily titled toward one direction. The oscillating sector

described above could be rotated clockwise or counter-clockwise using keyboard by the subjects, who were instructed to adjust the display so that the flashed bar appeared to be subjectively vertical. They had unlimited time to make this “subjective vertical” adjustment. When they were satisfied with the adjustment, they pressed spacebar to move on to the next trial. The two reversal directions were tested 20 times each for each subject.

In the EEG experiment, subjects were presented with the same rotating sector as in the psychophysics session, except that the bar always flashed at the vertical meridian (See Figure 2.4). The green vertical bar had 50% chance to flash on for 30 ms at the reversal. The flash grab effect biased the perceived location of the flash bar in the direction of the sector’s motion after the reversal. There were four situations after a reversal: (1) sector rotated to the left without bar flash; (2) sector rotated to the right without bar flash; (3) sector rotated to the left with the flashed bar perceived to be tilted to the left; (4) sector rotated to the right with the flashed bar perceived to be tilted to the right. (1) and (2) were termed “background-only” condition, whereas (3) and (4) were termed “illusory” condition. Stimuli were presented in runs that lasted ~120s. Data from 5 runs were collected, yielding 200 repetitions in each situation. In the control experiment, only the retinally-tilted flash bar was presented (adopting the angle obtained in the psychophysics session, 50% chance to flash), without the rotating background sector, termed “retinally-tilted” condition.

EEG data were acquired from 64 scalp electrodes (Neuroscan), digitized at 1000 Hz. Vertical electro-oculogram (VEO) was recorded by electrodes placed above and below the left eye. Horizontal electro-oculogram (HEO) was recorded by electrodes placed at the left and right outer canthi. The reference electrode was placed on the top of the midline between electrodes CZ and CPZ.

Psychophysics data analysis

Psychophysical data were analyzed using custom MATLAB scripts (MathWorks Inc.). The average behavioral performance was plotted separately for each condition as the percentage of upward responses against intersection angles of test bars (Figure 2.1c/e). Data points were fitted with the following logistic function to estimate the PSE (point of subject equality) where the test bars appeared parallel (both vertical).

$$(1)\ p(x) = \gamma + \frac{1-\lambda-\gamma}{1+e^{-\beta*(x-\alpha)}}$$

x is the intersection angle and p(x) is the percentage of upward response. α, β, λ and γ are free parameters that were fitted using least squares estimation.
The magnitude of TAE was measured as half the distance of PSEs following adaptation in two opposite orientations.

fMRI data analysis

3T MRI data were analyzed with Brain Voyager QX software package (Goebel, Esposito, & Formisano, 2006) and Matlab (MathWorks Inc.). Functional images were motion corrected, low and high pass temporal filtered, and slice timing corrected. The high-resolution T1 volume was co-registered to the first volume of functional images, and transformed to Talairach space. General linear model was used to estimate fMRI responses to the flashed bars with clockwise and counter-clockwise illusions. The retinotopic mapping data was analyzed using a cross-correlation method embedded in BrainVoyager QX software package. 16 phase lags (every 2 seconds) was used to find the best fit of polar angle or eccentricity representation for each voxel. ROIs of early visual cortices (V1, V2, V3d/VP) were defined according to the retinotopic maps on inflated cortical surface. For each ROI, voxels were sorted and resampled into 360 bins according to their polar angle representations. Then the BOLD response of the flashed bar was plotted as a function of polar angle. From this response curve, the angular representation of a flashed bar was estimated separately for the upper and lower visual fields, defined as the polar angle that splits the area under the curve into two equal halves.

7T MRI data were analyzed with AFNI (Cox, 1996), Freesurfer (Fischl, 2012), and custom Matlab/Python codes. Functional images were motion corrected and EPI distortion. The high-resolution T1 volume was co-registered to the mean volume of functional images. General linear model was used to estimate fMRI responses to the red bars with clockwise and counter-clockwise illusions. A cross-correlation method with 32 phase lags (every one second) was used to generate the polar angle retinotopic map of early visual areas V1/V2/V3. Pial and White Matter surfaces were reconstructed based on PD corrected T1 volume (Van de Moortele et al., 2009). An equi-distance method was used to estimate the relative cortical depth of a voxel. The voxels in a ROI were sorted and resampled into three depth bins: superficial depth (0-0.4), middle depth (0.4-0.8), and deep cortical depth (0.8-1.0). The partition ratio was selected based on the thickness of cortical layers of human visual cortex (De Sousa et al., 2010). Similar as the 3T data analysis, BOLD response to the flashed bar was plotted as a function of polar angel representation. To alleviate the draining veins effect of BOLD signal cross cortical layers, the min and max values of polar angle response curve was normalized to 0 and 1. The FGE illusory effect was calculated as the difference of normalized response between two illusory conditions (clockwise vs. counterclockwise), averaged across two polar angle windows (voxels identified through independent localizer scan with preferred orientation tuning to -14 to -6 degrees and 6 to 14 degrees). The input representation index was calculated as the mean of normalized responses centered on the horizontal meridian (where voxels had preferred orientation tuning ranging from -4 to 4 degrees). The polar angle windows were chosen to maximize the sensitivity of the index, because when pooling across all subjects/areas/layers, the difference between CW/CCW illusory conditions were most prominent around ±10 degrees (i.e., for voxels with preferred orientation tuning around 10 or -10 degrees). A small gap was left between these orientation windows to mitigate potential cross talk, and a slightly different gap did not qualitatively change the final results. The data with error bars are displayed as mean±SEM. The p values < 0.05 were considered statistically significant. Within-

subject confidence intervals were estimated according to the method described by Cousineau (Cousineau, 2005).

EEG data analysis

Data were analyzed using EEGLAB v13.3.2 (http://www.sccn.ucsd.edu/eeglab) and MNE v0.16.2 (https://martinos.org/mne/) (Gramfort et al., 2013). Raw data were first filtered off-line with a 1-35 Hz bandpass filter. Data excursions exceeding 75 μV at electrode VEO (-100 to +300 ms) were excluded from analysis. Remaining epochs were separately averaged according to the stimulus conditions. To select electrodes for the C1 amplitude and latency analysis, grand averaged ERPs were made for each electrode and each condition but pooling all subjects. Five electrodes showing the largest C1 amplitudes were chosen for further analysis (posterior electrodes including P3, P5, PO5, PO7, O1). To quantify the C1 amplitude and latency for each stimulus and each subject, the waveforms at these five electrodes were first averaged to obtain a mean waveform.

Multivariate pattern analysis (Grootswagers, Wardle, & Carlson, 2017) was conducted using scikit-learn 0.16.0 (http://scikit-learn.org/) (Pedregosa et al., 2015). Linear support vector machine classifiers were trained at each time point for each subject to predict to which side the flashed bar was retinally or perceived to be tilted, using preprocessed EEG data from all electrodes as features. For the background-only condition, we were predicting to which side a bar would be illusorily tilted if it was flashed as in the illusory condition, although the imaginary bar was not actually displayed. The decoding accuracy was estimated using a stratified 10-fold cross-validation procedure, and the regularization parameter C was set to 1.0. Each feature (electrode) was normalized to have zero mean and unitary standard deviation. To reduce the impact of random noise in single trials, we employed a mini-ERP approach. From all trials sharing the same label in the training set, k trials were randomly selected and averaged into a mini-ERP, which served as one training sample. The sampling process repeated until 1000 samples were generated and

used to train the classifier. Similar procedure was used at test time except that the mini-ERP samples were derived from test set. We chose k = 9 in current analysis, leading to a 3-fold boost in SNR and hence more accurate and robust decoding.

Cross decoding was performed across different conditions and different time points. A separate SVM was trained using all trials in condition A at time t_A, and tested using all trials in condition B at time t_B. The average prediction accuracy of all subjects was recorded in a matrix at row t_A and column t_B. To reduce computational burden, the EEG time series were decimated in time, and raw trial data instead of mini-ERP were used (i.e., k = 1) in this analysis.

The inter-subject correlation between either instantaneous or time-averaged ERP amplitude and TAE effect size was quantified with Pearson's linear correlation coefficient. The lateralization potential evoked by the vertical bar was calculated by first subtracting ERP signals in ipsilateral electrodes from corresponding contralateral electrodes, and then contrasting illusory condition with background-only condition. The same set of posterior electrodes were selected as with the ERP analysis. The illusion size for each subject was obtained by pooling all measurements for both directions from the adjustment experiment for both directions. The mean ERP amplitude was averaged within the interval between 177 ms and 400 ms after bar onset for visualization purpose. The time interval was chosen according to the onset of significant instantaneous correlation and the interval of significant higher decoding accuracy in illusory condition compared with background-only condition.

The difference in time series were tested for statistical significance at population level using cluster-based permutation test (Maris & Oostenveld, 2007; Nichols & Holmes, 2003) which corrected for multiple comparisons. Values at individual time points were first subjected to mass univariate t-test with cluster-defining threshold set to $p < 0.05$ (or $|r| > 0.5$ for correlation analysis). The resulted contiguous

suprathreshold intervals, in which statistics were of the same sign, were defined as clusters. For cross-decoding matrix, 2D clusters were defined on regular lattice. These clusters had to further pass a critical value in “cluster mass” before reported as significant. Cluster mass is the sum of t values in the cluster. The critical values were obtained with the following procedure: 1) randomly permute left or right labels for each subject, apply mass univariate t-test, calculate cluster mass for each cluster, and record the max and min cluster mass values; 2) repeat the above for 10000 times or all possible permutations, and construct the empirical distribution for max and min values; 3) take the 97.5 and 2.5 percentiles of the max and min distributions, respectively, as the critical values for a two-tailed test. The confidence interval of population mean time courses as well as instantaneous intersubject correlation was estimated using bootstrap technique by resampling the subjects with replacement for 1000 times.

Chapter 3

Spatiotopic updating across the saccades in the absence of awareness

Despite the continuously changing visual inputs due to eye movements, our perceptual representation of the visual world remains remarkably stable. Visual stability has been a major area of interest within the field of visual neuroscience. The early visual cortical areas are retinotopic-organized and presumably there is a retinotopic to spatiotopic transformation process that supports the stable representation of the visual world. In this study, we used a cross-saccadic adaptation paradigm to show that both the orientation adaptation and face gender adaptation could still be observed at the same spatiotopic (but different retinotopic) locations even when the adapting stimuli were rendered invisible. These results suggest that awareness of a visual object is not required for its transformation from the retinotopic to the spatiotopic reference frame.

This chapter is a reproduction of Ge, Y., Sun, Z., Qian, C., & He, S. (2021). Spatiotopic updating across saccades in the absence of awareness. Journal of Vision, 21(5), 7-7.

INTRODUCTION

Despite the continuous movements of the eyes and body, our visual world remains stable. In other words, an object could be imaged at very different positions on our retina (when eyes move), but our perceptual representation of that object remains stable in the visual world. Key to this visual stability is the transformation of visual object representation from the retinotopic (coordinates centered on the retina) to spatiotopic (coordinates centered on the outside world) reference frame across saccades (Cicchini, Binda, Burr, & Morrone, 2013; Crapse & Sommer, 2012; Fabius, Fracasso, Nijboer, & Van Der Stigchel, 2019). Previous studies showed that neurons in the extrastriate visual cortex (such as V4) and the lateral intraparietal cortex (LIP) could temporarily remap their receptive fields to compensate for an impending saccadic eye movement (Duhamel, Colby, & Goldberg, 1992; Tolias et al., 2001; Wurtz, Joiner, & Berman, 2011). Meanwhile, other studies also indicated explicit spatiotopic neural representation in middle temporal area (MT) and parietal areas (D'Avossa et al., 2007; Duhamel, Bremmer, BenHamed, & Graf, 1997), although this has remained a topic of debate (Gardner, Merriam, Movshon, & Heeger, 2008; Merriam, Gardner, Movshon, & Heeger, 2013). In any case, either by continuously updating or remapping the retinotopic maps, or by transforming the retinotopic representation to explicit spatiotopic representation, our brain would be able to keep track of the salient objects in the scene and achieve visual stability.

While the input visual information during saccades is suppressed, our conscious representation of the visual scene across saccades seems to be smooth and continuous, yet we typically do not keep track of the whole visual scene. Selective attention is one of the potential mechanisms to help to maintain visual stability (Crespi et al., 2011; Melcher, 2008, 2011; Szinte, Jonikaitis, Rangelov, & Deubel, 2018). Attentional selection contributes to visual stability by restricting information processing to salient or task-relevant objects. Thus the trans-saccadic spatiotopic updating of salient objects would allow the brain to track important features or items

in the scene. With multiple objects, the allocation of the selective attention would influence the spatiotopic updating and previous results showed that unattended stimuli could induce decreased but still measurable adaptation aftereffect in the spatiotopic location (Melcher, 2009; Melcher & Colby, 2008). However, while attention plays an important role in gating information to awareness, attention and awareness are not the same. Here we ask if the visual stimulus is invisible, could the spatiotopic updating process still happen? In other words, is spatiotopic updating so critical to our visual function that this process occurs even when we are not aware of the objects in the visual scene? Previous studies have shown that attention can be drawn to unconscious stimuli (Cohen, Cavanagh, Chun, & Nakayama, 2012; Jiang, Costello, Fang, Huang, & He, 2006) and the unconscious stimuli can still be processed to a certain level in the neural pathway (Axelrod, Bar, & Rees, 2015; Fang & He, 2005; Z. Lin & He, 2009; Sterzer, Stein, Ludwig, Rothkirch, & Hesselmann, 2014). Thus the key question addressed in this study is: is awareness of a visual object necessary for its reference frame transformation from retinotopic to spatiotopic across saccades?

Retinotopic vs. spatiotopic representations are dissociated by object locations pre- and post-saccadic eye movements. To investigate the question raised above, in addition to using eye movement that dissociates the object's retinotopic and spatiotopic locations, we also need a tool to probe the neural representation in the corresponding locations before and after the saccade. Adaptation paradigms are effective in studying neural representations in different reference frames for they allow a relatively long temporal delay in measuring the adaptation effect, so that if an object has achieved representation at the spatiotopic reference frame we would expect to see adaptation effect when the test probe is presented at the same spatiotopic location (even if its retinotopic location is different from that of the adapting stimulus). Adaptation paradigms also have the advantage of being able to target specific levels of neural representation in the visual pathway by selectively adapting to properties with different levels of complexity (Boynton & Finney, 2003; Colin W.G. Clifford & Rhodes, 2005; Georgeson, 2004; Kohn, 2007; Rushton, 1965).

In our study, we took advantage of two forms of visual aftereffects that were previously shown capable of generating spatiotopic aftereffects, namely the tilt aftereffect (TAE) and the face gender aftereffect (FGAE) (Cha & Chong, 2014; D. He, Mo, & Fang, 2017; T. He, Fritsche, & Lange de, 2018; Melcher, 2005, 2009; Nakashima & Sugita, 2017; Wolfe & Whitney, 2015; Zimmermann, Morrone, Fink, & Burr, 2013; Zirnsak, Gerhards, Kiani, Lappe, & Hamker, 2011). We first verified that both TAE and FGAE could be observed at the spatiotopic location, which implied that the adapting stimulus had undergone retinotopic to spatiotopic transformation.

Next, to render the adapting stimulus invisible so that we could investigate whether the aftereffects could still be observed at the spatiotopic location from the invisible adaptor, we adopted the continuous flash suppression (CFS) approach. CFS is an effective way to render adapting stimuli in one eye invisible by presenting a stream of rapidly changing noise to the other eye. CFS has the advantage of achieving prolonged suppression duration and being less influenced by visual properties of the to be suppressed stimulus (Fang & He, 2005; Kim & Blake, 2005; Tsuchiya & Koch, 2005). There is evidence showing that different types of adaptation aftereffects are differentially influenced by interocular suppression. Not surprisingly, more complex stimulus properties like face gender and identity information are more vulnerable to suppression, compared with simple stimulus features such as flicker, motion, or orientation (Alais & Melcher, 2007; Kaunitz, Fracasso, & Melcher, 2011; Tsuchiya & Koch, 2005; Yang, Hong, & Blake, 2010). In this study, we investigated the role of awareness in the retinotopic to spatiotopic reference frame transformation, by using CFS to suppress the awareness of the target visual objects. Our results show that for visual targets not consciously perceived, both local orientation information and face gender information could be transformed from retinotopic to spatiotopic reference frame.

METHODS

Participants

Twelve participants (7 females, mean age=23.2) took part in the main experiment. Half of the participants (n=6) also took part in the eye movement recording experiment. All participants had normal or corrected-to-normal vision. All participants provided written informed consent and were paid to take part in the study, which was approved by the Institutional Review Panel at the Institute of Biophysics (IBP), Chinese Academy of Sciences (CAS).

Stimuli

Stimuli were displayed on two synchronized 23.8-inch LCD displays (Dell U2414H, 1920*1080 at 60 Hz refresh rate) and viewed from a distance of 80 cm through stereo mirrors. All visual stimuli were generated using MATLAB Psychophysics Toolbox (Brainard, 1997). The presentation of a frame (18 * 12 dva) with dashed lines facilitated stable convergence of images in two eyes and also provided background coordination information for the saccade task. A cross (0.56 * 0.56 dva) presented in the left or right part of the frame served as the fixation point.

The adaptor for tilt aftereffect was a tilted (±15°) Gaussian-windowed sinusoidal luminance Gabor that subtended 5 dva (Figure 3.1b). The frequency of the Gabor was 0.8 c/deg. The test stimuli were similar to the adaptor, tilted from -4.5 to 4.5 degrees. For the face adaptation, male and female faces were used as adaptors subtending 5 dva. The morphs were generated using Morph 3.0 (Gryphon Software, San Diego, CA) with 100 intervening morphs. Morph number 50 was regarded as a neutral center point within the morphing space.

Procedure

There were two conditions, visible and invisible, for each adaptation stimulus type in separate sessions to avoid task complexity. A total of 2688 trials were obtained for each participant across all conditions. In the visible condition, after the initial adaptation period (25s), the participants first fixated at the left cross for 0.8 s. Then

the top-up adaptor was presented to the participant's non-dominant eye for 2 s at the upper-middle location of the monitor. Following a 0.8 s ($SD = 0.1\ s$) preview of the next fixation cross on the right side, while still maintaining fixation on the left cross, the participants made a saccade to the right fixation cross (6 dva from the left cross) prompted by the extinction of the current fixation cross on the left. Then a test probe was presented for 100 ms at one of four possible locations (retinotopic, spatiotopic, retinotopic-control, or spatiotopic-control) pseudo-randomly selected with equal probability (Figure 3.1). Participants needed to report the direction of tilt of the Gabor or the gender of the face.

The invisible condition was the same as the visible condition, except that dynamic Mondrian patterns (10 Hz, subtending 5 dva) were simultaneously presented to participants' dominant eye in both initial and top-up adaptation periods. To ensure that the dynamic Mondrian patterns could effectively suppress the adapting stimuli (Stein & Sterzer, 2014; Yang, Brascamp, Kang, & Blake, 2014), we first presented the adaptor at 80% contrast to test whether it could be suppressed in both initial adaptation (25s) and 20 trials of top-up adaptation (2s each trial) for each participant. Participants were asked to press a button if they detected the adaptor in the initial adapting period or in any trial. If the adaptor broke the suppression in more than 5% of trials, we then reduced the contrast of the adaptor by 5% and tested again. This process resulted in the adaptor been seen under CFS suppression in no more than 5% of the trials. The contrast of adaptor was recorded and used in the formal experiment (average contrast for Gabor patch: 79.7% $\pm$ 0.8%; average contrast for face: 78.3% $\pm$ 2.3%). During the adaptation period, if participants could see the Gabor or tell the gender of the face, they pressed a button (spacebar) to indicate the Mondrian patterns did not fully suppress awareness of the adaptor. These trials were excluded from further analysis.

In addition, we included a full adaptation condition in which participants maintained the fixation on the left without making a saccade during the whole period with the

test stimulus presented in the same location as the adaptor. The logic of the experiment is that if an aftereffect could be observed at the spatiotopic location, then it would imply that the adapting stimulus had achieved spatiotopic representation, in other words, had undergone retinotopic to spatiotopic transformation.

Eye movement measurements

To verify that the participants were generally able to follow the instructions, half of the participants (n=6) took part in an eye movement experiment, which was the same as the main experiment, but half in the number of trials (1344 trials). Eye movements of the participants were monitored by the Eyelink 1000 Plus system (SR Research), which sampled gaze positions with a frequency of 1000 Hz. Only the left eye was recorded. The system detected a start and an end of a saccade when eye velocity exceeded or fell below 22°/s and acceleration was above or below 3800°/s^2. At the beginning of each session during the experiment, a 9-point calibration and validation procedure was conducted. If the calibration did not meet the defined requirements, calibration was repeated until successful. The averaged horizontal eye positions over the time course of the trial for each participant were showed in Appendix Figure A2.1. The eye position traces were aligned with the midpoint of the saccade.

Analysis

MATLAB was used to analyze the data. The psychometric response curve was fitted with a Bayesian-based cumulative Gaussian function (psignifit toolbox in MATLAB) (Schütt, Harmeling, Macke, & Wichmann, 2016) to measure the aftereffects. The magnitude of the TAE was defined as half the difference of tilt to annul the effects of adapting clockwise, compared with counter-clockwise gratings. The FGAE was calculated with a similar method. Example fitting results for one participant were shown in Figure 3.2. It showed the tilt aftereffect in four different locations when the adaptor was visible. One-half of the distance between two fitted curves was the measured magnitude of the aftereffect.

RESULTS

Participants were well able to maintain their fixation and execute the required eye movements (Appendix Figure A2.1). The mean distances between eye position and fixation center were 0.11° (SD=0.09°) and 0.36° (SD=0.28°) before and after saccades. Saccades, which need to be executed within 500ms after the extinction of the left fixation cross, were on average accurate and prompt, with 143.6 ms (SD=117.3) mean saccade latency. In only 1.15% of all trials, the saccades were not executed before the test stimulus presentation. Due to the very small proportion of these delayed saccades, our results were not affected by whether we exclude these trials or not in the following statistical analysis.

For participants who finished separate sessions with and without eye movement recording, no significant differences were found between the two sessions (dependent sample t-tests for all conditions, $p>0.05$, Appendix Figure A2.2). There were also no significant differences between participants with and without eye movement recording (independent sample t-tests for all conditions, $p>0.05$, Appendix Figure A2.3). Thus we combined these data in the further statistical analysis.

The strength of TAE and FGAE for each participant was calculated as half of the difference on the x-axis between the two points of subjective equality (PSEs) based on the psychometric functions following adaptation in two opposite orientation (TAE) or gender (FGAE) (see Figure 3.2 for an example). Statistics were then performed on the group data.

We performed two-way ANOVA analyses to examine the effects of two factors (two levels of adaptor awareness and five different adapt-test relationships) on the magnitude of TAE and FGAE. For the TAE, both the main effects of adaptor awareness and adapt-test relationship are significant (adaptor awareness: $F(1,11)=61.48$, $p<0.001$, $\eta_p^2 = 0.848$; adapt-test relationship: $F(4,44)=61.61$, $p<0.001$,

$\eta_p^2 = 0.849$). The interaction between adaptor awareness and the adapt-test relationship was also significant ($F(4,44)=11.71$, $p<0.001$, $\eta_p^2 = 0.516$), indicating that the impact of adaptor awareness depended on the relationship between adapt-test locations. Post hoc analysis showed that the TAE in spatiotopic location is significantly larger than the control-spatiotopic location in both visible (t=5.91, p<0.001) and invisible condition (t=3.26, p<0.01), suggesting the existence of a spatially specific adaptation effect at the spatiotopic location, regardless of awareness state of the adapting stimulus.

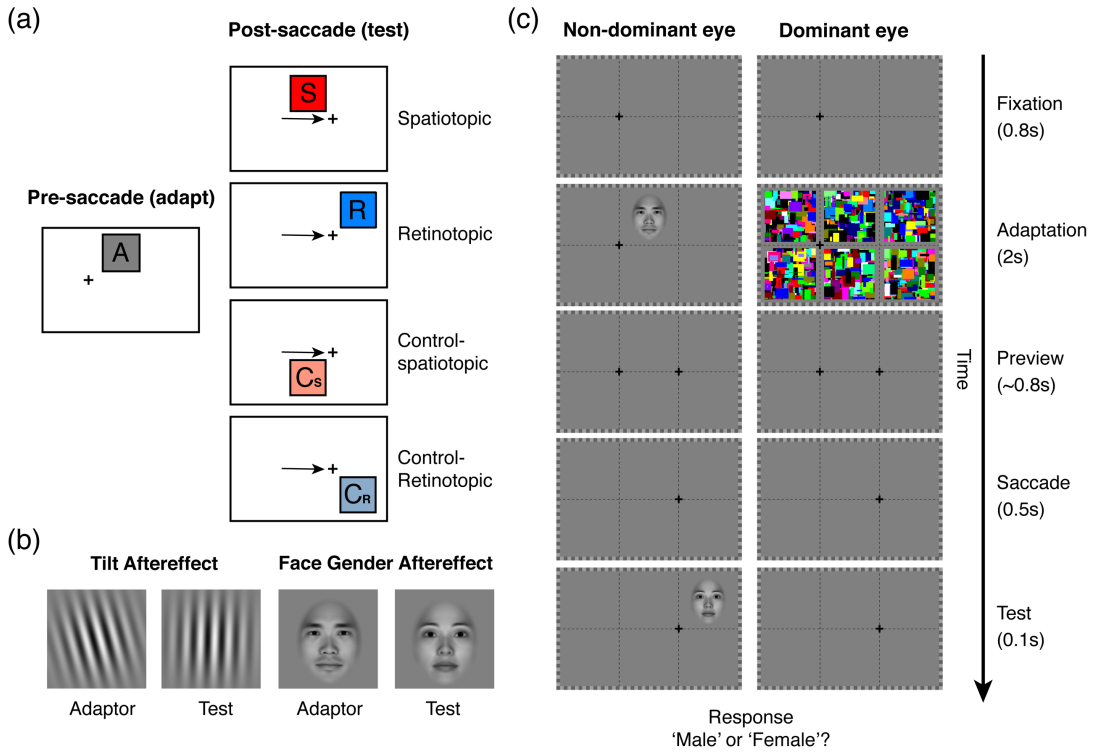

Figure 3.1. Experiment paradigms for different conditions. (a) The locations of adaptation and test stimuli before and after the saccade. The cross presents the fixation point. The black arrow represents the saccade direction (from left to right); The letters in the squares: A-adaptation location (also the full adapt test location); S-spatiotopic location; R-retinotopic location; Cs- control spatiotopic location; Cr- control retinotopic location; (b) Adaptor and test stimuli for tilt aftereffect and face gender aftereffect; (c) Time sequences in the experiment. The adapter was presented for 2 s (top-up adaptation) after 0.8 s fixation in the left cross. After a 0.8 s preview of the right cross, participants need to saccade to the right cross after the extinction of the left cross. Then a test stimulus was present for 0.1s in 1 out of 4 locations randomly.

For the FGAE, again both the main effects of adaptor awareness and adapt-test relationship are significant (adaptor awareness: $F(1,11)=14.49$, $p=0.003$, $\eta_p^2 = 0.568$; adapt-test relationship: $F(4,44)=12.15$, $p<0.001$, $\eta_p^2 = 0.525$). However, the interaction effect between adaptor awareness and adapt-test relationship is not significant ($F(4,44)=1.83$, $p=0.141$, $\eta_p^2 = 0.142$), suggesting that the impact of adaptor awareness was not dependent on the relationship between adapt-test locations. Post hoc analysis showed that the FGAE in spatiotopic location is not significantly larger than that in the control-spatiotopic location in both visible and invisible conditions ($p>0.05$).

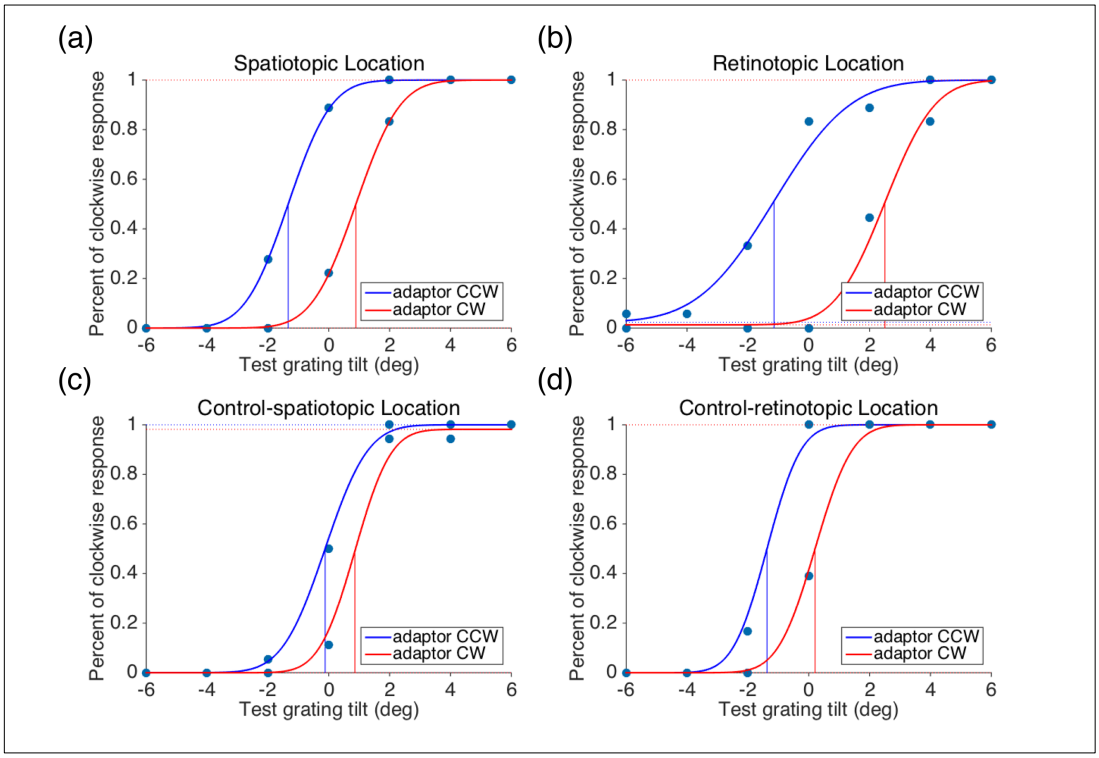

Figure 3.2. Fitted curves of Tilt Aftereffect results for one participant in four test locations without CFS stimuli (spatiotopic (a), retinotopic (b), control-spatiotopic (c), and control-retinotopic location (d)). Similar results were found for the other 11 participants. Red and blue curves represent clockwise and counterclockwise adaptors respectively. The vertical bars represent the estimated 50% threshold.

For the visible condition (without CFS), the one-sample t-tests with Holm correction (*N*=10, 5 locations* 2 state awareness (with(out) CFS) for TAE and FGAE respectively) indicate that both TAE and FGAE could be induced at the spatiotopic location (TAE: *M*=0.93˚, *p*<0.001; FGAE: *M*=7.56%, *p*<0.001), and not surprisingly, at the retinotopic location (TAE: *M*=2.26˚, *p*<0.001; FGAE: *M*=16.67%, *p*<0.001). Results show that the TAE and FGAE partially transfer to control-retinotopic location (TAE: *M*=0.48˚, *p*<0.01; FGAE: *M*=5.46%, *p*<0.05) and control-spatiotopic location (TAE: *M*=0.27˚, *p*<0.05; FGAE: *M*=7.85%, *p*<0.01). The full adaptation condition (no saccade) reveals the strength of the TAE (*M*=2.38˚, *p*<0.001) and FGAE (*M*=10.31%, *p*<0.001) in the classic condition (Figure 3.3, left panels) (also see normalized results in Appendix Figure A2.4).

For the invisible condition (with CFS), interestingly, results show that both stimuli could still generate robust aftereffects at the retinotopic (TAE: *M*=0.85°, *p*<0.02; FGAE: *M*=6.62%, *p*<0.02) and spatiotopic locations (TAE: *M*=0.25°, *p*<0.02; FGAE: *M*=3.88%, *p*<0.03), whereas no aftereffect was observed at the control-spatiotopic location (TAE: *M*=0.02°, *p*=0.97; FGAE: *M*=1.09%, *p*=0.88) nor at the control-retinotopic location (TAE: *M*=0.02°, *p*=0.97; FGAE: *M*=0.19%, *p*=0.88). For the full adaptation condition without saccade, significant TAE and FGAE were observed (TAE: *M*=0.69°, *p*<0.01; FGAE: *M*=6.83%, *p*<0.05) (Figure 3.3, right panels). Comparing with results in the visible adaptation condition, the spread of aftereffects to control locations did not occur when participants had no awareness of the adaptation stimulus, however, the adaptation effect remained robust at the spatiotopic location.

DISCUSSION

We used the adaptation paradigm to investigate whether visual objects could be transformed from retinotopic to spatiotopic reference frame while observers were not aware of their presence. We first established that both the orientation and the face gender adaptation were capable of generating tilt and face gender aftereffects,

respectively, when tested at different retinotopic but the same spatiotopic location. The critical observation is that when the adapting stimulus was rendered invisible, both aftereffects could still be observed at the spatiotopic location.

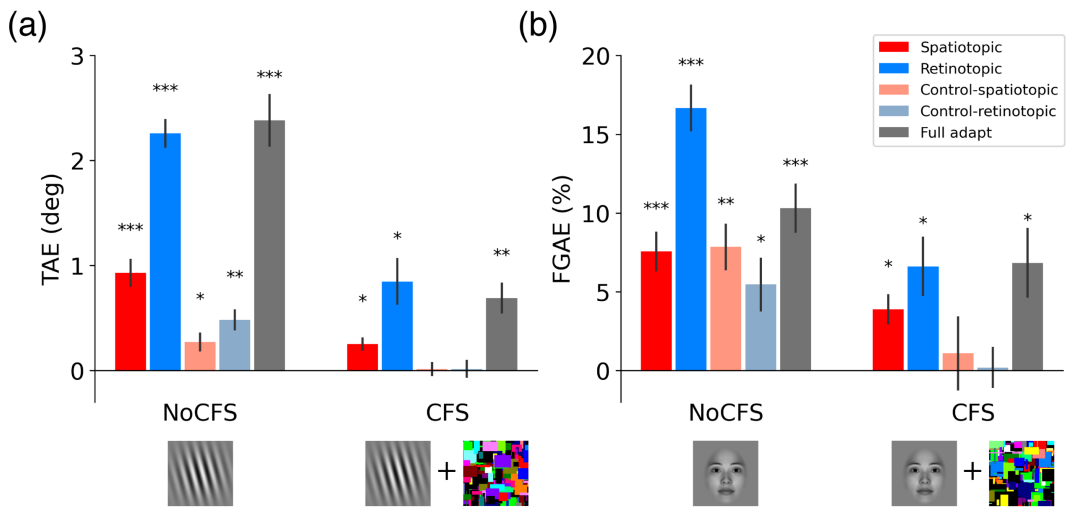

Figure 3.3. Adaptation aftereffects (a, TAE; b, FGAE) for the No-CFS and CFS conditions in different locations. Average results from 12 participants show significant TAE and FGAE effects in spatiotopic locations when the adaptors were visible. The effect partially transferred to the two control locations. For invisible adaptor, robust adaptation effects were observed in spatiotopic locations, but not in two control locations. Error bars show ±1 SE of the mean. Multiple comparisons were Holm corrected. (* adjusted $p<0.05$; ** adjusted $p<0.01$; *** adjusted $p<0.001$).

In contrast to awareness being not necessary for the spatiotopic updating, the buildup of spatiotopic neural representation requires spatial attention (Crespi et al., 2011; Melcher, 2008, 2009, 2011; Melcher & Colby, 2008; Szinte et al., 2018). Crespi et al. (2011) found that when participants were conducting a demanding attention task on the foveal stimuli, BOLD responses evoked by moving stimuli unrelated to the fovea task were mainly tuned in retinotopic coordinates. But the BOLD responses were tuned in spatiotopic coordinates when subjects could easily attend to the motion stimuli. In our study, when the adaptors were visible, the spatial attention to the adaptor location might help the buildup of the adaptation effect in the spatiotopic location. Previous studies showed that the stimuli under CFS could still

influence spatial attention (Jiang et al., 2006), which may enable our observation that both TAE and FGAE could occur at the spatiotopic location without visual awareness.

Attentional facilitation to the saccade destination may also influence the adaptation effects. In our study, the saccade target did not overlap with test locations and eccentricity-matched control locations were included for both spatiotopic and retinotopic conditions. Thus, the possible effects of attention facilitation to the saccade target were avoided due to the equal probability of test presence among four different locations (Afraz & Cavanagh, 2009). Besides, since the adaptation and test stimuli were always presented in the periphery, there was no switch between foveal and peripheral locations in testing the aftereffects, presumably generating more stable aftereffect measurements.

It has been debated whether visual feature information or just the spatial information is transferred in the trans-saccadic remapping. Recent studies demonstrated that feature information like orientation (Ganmor, Landy, & Simoncelli, 2015; Wutz, Drewes, & Melcher, 2016; Zimmermann, Weidner, & Fink, 2017), shape (Demeyer, De Graef, Wagemans, & Verfaillie, 2009), motion (Fabius, Fracasso, & Van Der Stigchel, 2016; Fracasso, Caramazza, & Melcher, 2010; Melcher & Fracasso, 2012; Turi & Burr, 2012), and facial expressions (Wolfe & Whitney, 2015), could be remapped across saccades. Our results provide further support that trans-saccadic remapping takes place at the feature level. The process of feature remapping would enabling the construction of spatiotopic representations of visual features.

The time course of spatiotopic updating might also influence the adaptation effects among different locations across saccades (Burr, Tozzi, & Morrone, 2007; Melcher & Morrone, 2003). There is evidence showing that the preview duration is a necessary requirement for the spatiotopic representation to fully build up (Golomb, Marino, Chun, & Mazer, 2011; Golomb, Nguyen-Phuc, Mazer, McCarthy, & Chun, 2010; Golomb, Pulido, Albrecht, Chun, & Mazer, 2010; Mathôt & Theeuwes, 2010; Morrone, Cicchini, & Burr, 2010; Zimmermann, Morrone, & Burr, 2015, 2014;

Zimmermann et al., 2013). Thus, the relatively long target-preview duration (0.8 s) used in our study likely contributed to a stronger object representation at the spatiotopic location. It is also possible that spatiotopic updating may have different temporal dynamics for different stimulus types and states of awareness. For example, a recent study using rotating motion illusion suggested that spatiotopic updating could occur rapidly (e.g., within 150 ms) (Fabius et al., 2019).

Recent fMRI adaptation studies showed reduced BOLD response in the extrastriate visual cortex when two repeated gratings were presented at the same spatiotopic location before and after a saccade (Dunkley, Baltaretu, & Crawford, 2016; Fairhall, Schwarzbach, Lingnau, Van Koningsbruggen, & Melcher, 2017; Zimmermann, Weidner, Abdollahi, & Fink, 2016). These repetition suppression effects indicate a transfer of representation (and consequently adaptation effect) from retinotopic to spatiotopic reference frame, which is in accord with our finding of spatiotopic adaptation effect with visible grating adaptors.

Our results show that when the adaptor was visible, a robust tilt aftereffect could be observed in the spatiotopic location (with the largest effect in the retinotopic location and smaller effects in the control locations). For the face gender adaptation, the magnitude of aftereffects was similar among the spatiotopic and other two control locations (smaller than the retinotopic location), which is consistent with a previous study that showed no significant difference between spatiotopic and control locations (Afraz & Cavanagh, 2009). Such results indicate that, in addition to the transformation from retinotopic to spatiotopic reference frame, when the adapting face was visible, there was a spatially non-local adaptation effect. In other words, there was a more spatially invariant representation when an object was consciously perceived, in contrast to a more spatially local object representation in the absence of awareness. The role of awareness in spatially invariant representation was also revealed for object viewpoint in a recent study using Necker cubes as stimuli (Cho & He, 2019). With awareness, the spatially non-specific effect was also observed for

TAE, but quite a bit weaker, presumably due to the intrinsic local nature of orientation processing in the visual cortex.

More interestingly, when the adaptor was rendered invisible, our results show that there was still a significant representation of the adaptor at its spatiotopic location for both orientation and face gender information, but not in the two eccentricity-matched control locations. In other words, both local orientation and face gender information could be transformed from the retinotopic to spatiotopic reference frame without awareness. The spatiotopic updating of an object from its retinotopic reference frame, a process that is critical for achieving a stable perceptual representation of the visual world, can occur even when the object is not explicitly perceived.

Chapter 4

Neural representation of human pose information in natural images

The human body is a stimulus that occurs frequently in real life, and the pose, defined as the spatial relationships between body parts, carries a great deal of information about the underlying motion and action of a person. While there has been literature on the neural representation of some human pose variations, the enormous pose space experienced in natural images is largely unexplored. Here we examined the cortical sensitivity to a broad range of natural poses with a high degree of appearance variations from complex natural images of people. With recent advances in 3D human pose recovery from natural images, we developed several pose models to parameterize natural pose images and characterize the structure of the natural pose space from different aspects (viewpoint-dependent vs. viewpoint-independent) in distinct dimensions (2D vs. 3D). Using representational similarity analysis of fMRI data, we found several cortical regions, including areas of lateral occipital-temporal cortex (LOTC), fusiform gyrus, and superior parietal cortex that captured the structures of the pose space from both viewpoint-independent and viewpoint-dependent parameterizations. We also found that the right superior temporal sulcus captures only the intrinsic, viewpoint-independent 3D pose dissimilarity structure. Together, our results revealed distributed representations of different aspects of human pose information from a broad range of natural poses and appearances.

** This study was done in collaboration with Hongru Zhu and Alexander Bratch. Hongru Zhu developed the various model parameterizations, including the application of computer vision to extend the annotations to 3D. He also drafted the Introduction and Discussion sections. Alex Bratch provided advice on the localization of EBA/FBA ROIs and other cortical areas. In particular, Alex worked with Kendrick Kay of the CMRR to identify and standardize these ROIs for the NSD project. My primary role was in the analysis of the NSD fMRI data reported in the results, figures, and tables, including the searchlight analysis and comparisons of voxel RDMs with model RDMs. The development of the hypotheses and the interpretation of the data was the synergistic result of the collaboration between all three of us and our advisor.*

INTRODUCTION

As highly social creatures, our visual world is filled with a prevalent and complex stimulus—the human body in the natural world. The perception of the human body provides crucial support for the understanding of other people's emotions, actions, and social interactions. More specifically, pose, defined as the spatial relationships between body parts, carries a great deal of information about the underlying motion and action of a person. Further, human vision can draw inferences about both motion and action from even a single glance. However, computing human pose from a single natural image is computationally challenging (Wang, Wang, Lin, & Yuille, 2019). For one thing, human bodies have non-rigid forms with various joint articulations, making them prone to self-occlusion. For another, there is inherently a high degree of appearance variations in natural body stimuli from changes due to occlusion, clothing, lighting, and viewpoint. Given the complexity and importance of body pose information, we investigated the cortical representation of static, natural human poses defined by the local body parts and their spatial configurations in two dimensions and three dimensions.

An important line of research work has revealed specialized neural mechanisms for processing human body stimuli. Early fMRI studies found distinct cortical regions that are preferentially activated for human bodies, including the extrastriate body area (EBA) (Downing & Kanwisher, 2001) as well as the fusiform body area (FBA) (Peelen & Downing, 2005). Subsequent studies identified body part maps in the occipitotemporal cortex (OTC) with dissociable responses to individual body parts, and suggested that their organization was related to the action-related properties of body parts (Bracci, Caramazza, & Peelen, 2015; Orlov, Makin, & Zohary, 2010). Following from these previous findings which connect representations of individual body parts with action-related information, we focused on the cortical representation of human poses defined as the spatial configurations of body parts. Such intermediate pose representations have been relatively little studied but are effective

for motion and action understanding from a computational perspective (Campbell & Bobick, 1995; Wang, Wang, & Yuille, 2013; Yacoob & Black, 1999).

Along the line of human pose representations, previous work has investigated several human brain regions and their roles in pose discrimination using static images of a few pre-selected poses. Studies found that repetitive transcranial magnetic stimulation (rTMS) of EBA disrupts the perception of bodily form while rTMS of the premotor cortex disrupts the perception of bodily action (Urgesi, Candidi, Ionta, & Aglioti, 2007). Another fMRI study suggested viewpoint-independent encoding of contorted and ordinary postures in the fusiform gyrus, posterior superior temporal sulcus (pSTS), inferior frontal gyrus (IFG) and, inferior parietal lobule (IPL), including regions classically associated with action observation (Cross, MacKie, Wolford, & Antonia, 2010). Together, these studies measured viewpoint-independent cortical responses to pose variations in static images. However, given the vast range of legitimate, natural pose variations, prior work has not addressed how the enormous pose space experienced in natural images is represented. Further, findings from the use of simplified stimuli may not generalize to complex, real visual scenes (Hasson & Honey, 2012). Considering the high degree of appearance variations for human poses in real life and the highly simplified pose stimuli used in the prior work, it raises the question of cortical sensitivity to the broad range of human poses from complex natural images of people.

In light of this, we have developed several pose models that capture the structure of the pose space over a large range of natural poses. From a computational point of view, different pose parameterizations are arguably utilized to extract different aspects of pose information as needed. For example, viewpoint-dependent 3D pose representations make explicit body part depth and body orientation information with respect to the viewer, and thus are useful in the computation of relationships between a person and other objects/people. Whereas viewpoint-independent 3D pose representations are likely to be computationally more efficient for action

categorization by making explicit configurations of body parts of a person independent of viewpoint. Other possible pose parameterizations include viewpoint-dependent 2D pose representations, which, though requiring less computation, ignore relative depth information. These different pose parameterizations present trade-offs between computations and representations required for different tasks. In this work, we investigated cortical representations of pose information given three different parameterizations using (1) viewpoint-independent 3D pose representations, (2) viewpoint-dependent 3D pose representations as well as (3) viewpoint-dependent 2D pose representations. As a direct comparison, we also investigated another (4) viewpoint representations that were purely and explicitly based on body orientation with respect to the viewer.

To parameterize poses, we need to solve the problem of extracting pose information from natural scene images. Such information usually takes the form of joint locations in three dimensions. Traditionally, it is often complicated to obtain three-dimensional pose information from natural images because human subjects have to wear markers for motion capture (mo-cap) systems when the images are taken. Even with existing natural image datasets with three-dimensional pose annotations, it is still hard to extract the viewpoint of human body images – a necessity to produce viewpoint-independent pose parameters. Benefitting from the recent advances in computer vision, we made use of an off-the-shelf human 3D mesh reconstruction model (Kanazawa, Black, Jacobs, & Malik, 2018) to extract a corresponding 3D human mesh for each human body in natural images. The 3D human mesh comes with 3D body joint rotation and 3D body global rotation parameters, namely the viewpoint parameters, and can be transformed into 3D joint locations. With 3D joint locations and global rotation parameters, we produced the desired, different parameterizations for each pose. We subsequently built separate pose models to parameterize the broad range of human poses and characterized the pose space structures with different parameterizations.

Our adopted pose parameterization approach enabled us to extract 2D and 3D pose information from a large set of natural human images, allowing for our analysis of cortical activations obtained from the Natural Scene Dataset (NSD) (Allen et al., 2021). This is a massive high-resolution dataset containing 7T fMRI responses to natural scene images. For the scope of our analysis, we selected a subset of 4,450 natural scene images containing only single persons engaged in different activities including sports, household activities, eating and drinking, etc. Despite the additional complexity, variations, and nuisance factors inherent to natural images, the use of this large set of NSD images complements previous studies which have used highly simplified body images with much smaller variations in pose articulations and appearances.

To compare model predictions with the patterns of cortical activity, we used representational similarity analysis (RSA) and search-light mapping (Kriegeskorte, Goebel, & Bandettini, 2006; Kriegeskorte, Mur, & Bandettini, 2008). RSA enabled us to identify cortical regions whose responses correlate with the pose dissimilarity structure characterized by different pose parameterizations. This allows for a flexible form of pattern analysis and the plug-in use of different representational dissimilarity matrices (RDMs) from different models. Furthermore, RSA can also benefit from a data-driven perspective as we used search-light mapping to discover spatial clusters of voxels that may be distributed across the whole brain. We tested four different RDMs – three built on the dissimilarity measurements from three different pose parameterizations, and a fourth one built on the dissimilarities of the associated viewpoint from pairs of natural pose images. If any part of the cortical regions is sensitive to the auxiliary, relative depth information, we would expect distinct results from 2D and 3D viewpoint-dependent pose parameterizations. If viewpoint-independent pose information is automatically computed for NSD subjects in the continuous recognition task, which was to indicate whether they have seen each presented image at any point in the past, we expect some cortical regions to show greater sensitivity from 3D viewpoint-independent pose parameterizations.

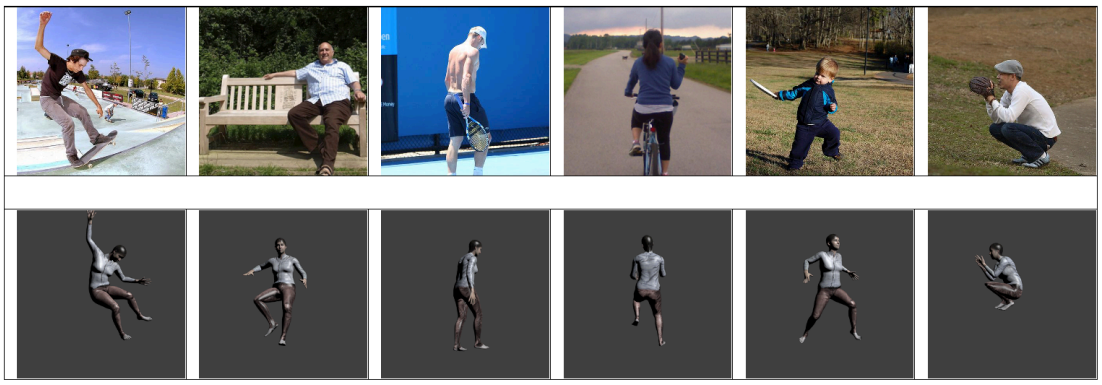

Figure 4.1. Example natural single human pose images from Natural Scene Dataset that were used in our analysis (first row), together with reconstructed 3D meshes (second row).

RESULTS

Pose parameterization and RDM construction

With the off-the-shelf human 3D mesh reconstruction model, we extracted human 3D mesh to further parameterize poses. Figure 4.1 shows examples of natural pose images sampled from the Natural Scene Dataset together with reconstructed meshes. These reconstructed meshes were reasonable and captured the major characteristics of different poses. It is thus feasible to make use of such mesh reconstruction results to parameterize complex natural poses.

With reconstructed meshes, we built different pose models to capture the structure of the pose space. We first obtained 3D joint locations and global rotation parameters, which were subsequently converted into (1) viewpoint-independent aligned 3D joint locations by reversing the global rotation in three-dimensions, (2) viewpoint-dependent 3D joint locations, (3) viewpoint-dependent 2D joint locations by discarding depth coordinates, and (4) explicit viewpoint information from global rotations. Four different models were built with these different aspects of pose information, and different RDMs were subsequently constructed in accordance with

dissimilarity measurements on different parameterizations (Figure 4.2). These RDMs showed that our pose models can capture dissociable pose and viewpoint information.

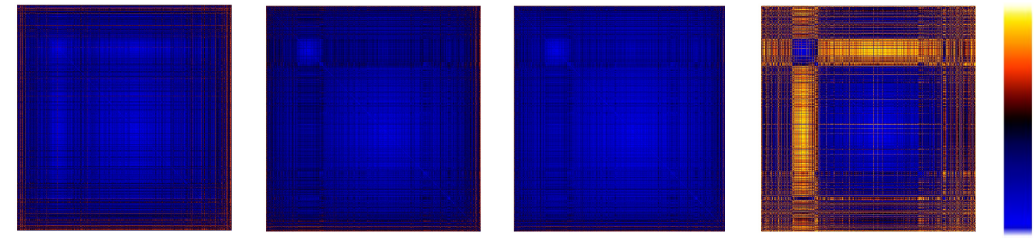

Figure 4.2. RDMs from (1) the view-independent 3D pose model, (2) the view-dependent 3D pose model, (3) the view-dependent 2D pose model, and (4) the viewpoint model. RDMs were calculated using 4450 natural pose images from NSD with the same image ordering.

RSA Searchlight

To investigate the spatial organization of cortical regions encoding different type of pose information, we performed searchlight-based representational similarity analyses using four different RDMs: (1) viewpoint-independent pose RDM, (2) viewpoint-dependent 3D pose RDM, (3) viewpoint-dependent 2D pose RDM, and (4) viewpoint RDM. Results were compared with cortical parcellation atlas (Desikan et al., 2006) as well as several regions of interest (ROIs) from functional localizers in NSD.

RSA results from different RDMs were shown in Figure 4.3 and Figure 4.4. Using the viewpoint RDM, we identified significant clusters that correlate with body viewpoint dissimilarity structures in the lateral occipital cortex, right fusiform gyrus, inferior parietal cortex, and superior parietal cortex. For the 3D viewpoint-independent pose RDM, we found distributed clusters across lateral occipital-temporal cortex (LOTC), fusiform gyrus, and temporal-parietal junction (including posterior superior temporal sulcus (pSTS), supramarginal gyrus).

For the 3D viewpoint-dependent pose RDM, we find little or no significant clusters around the pSTS and supramarginal gyrus. But we found a more distributed pattern in LOTC, fusiform gyrus, posterior frontal cortex, and cingulate cortex. Particularly, the activated areas contain the pericalcarine cortex, lateral occipital cortex, lingual gyrus, fusiform gyrus, parahippocampal gyrus, inferior and middle temporal gyrus, anterior supramarginal gyrus, inferior and superior parietal cortex, precuneus cortex, left precentral and paracentral gyrus, left caudal middle frontal gyrus, and posterior cingulate cortex. The 2D viewpoint-dependent pose model showed similar activation with the 3D viewpoint-dependent pose model, with overlapped clusters in LOTC, fusiform gyrus, and inferior parietal cortex (Figure 4.5).

For the viewpoint RDM, significant clusters were found mainly near the extrastriate visual cortex and posterior parietal cortex.

As a result, 2D and 3D pose RDMs produced overlapping clusters mainly in areas near LOTC, fusiform gyrus, and the superior parietal cortex. Appendix Tables A3.1-A3.4 provide further details about cluster size, location, and other information from the use of different RDMs in cortical parcellation atlas (Desikan et al., 2006). We further compared the RSA searchlight results with NSD functional localizer results as shown separately in Appendix Figure A3.1. Results show that our distributed pose clusters also overlap with several ROIs that are associated with body or face processing, including OFA, FFA, FBA, and EBA.

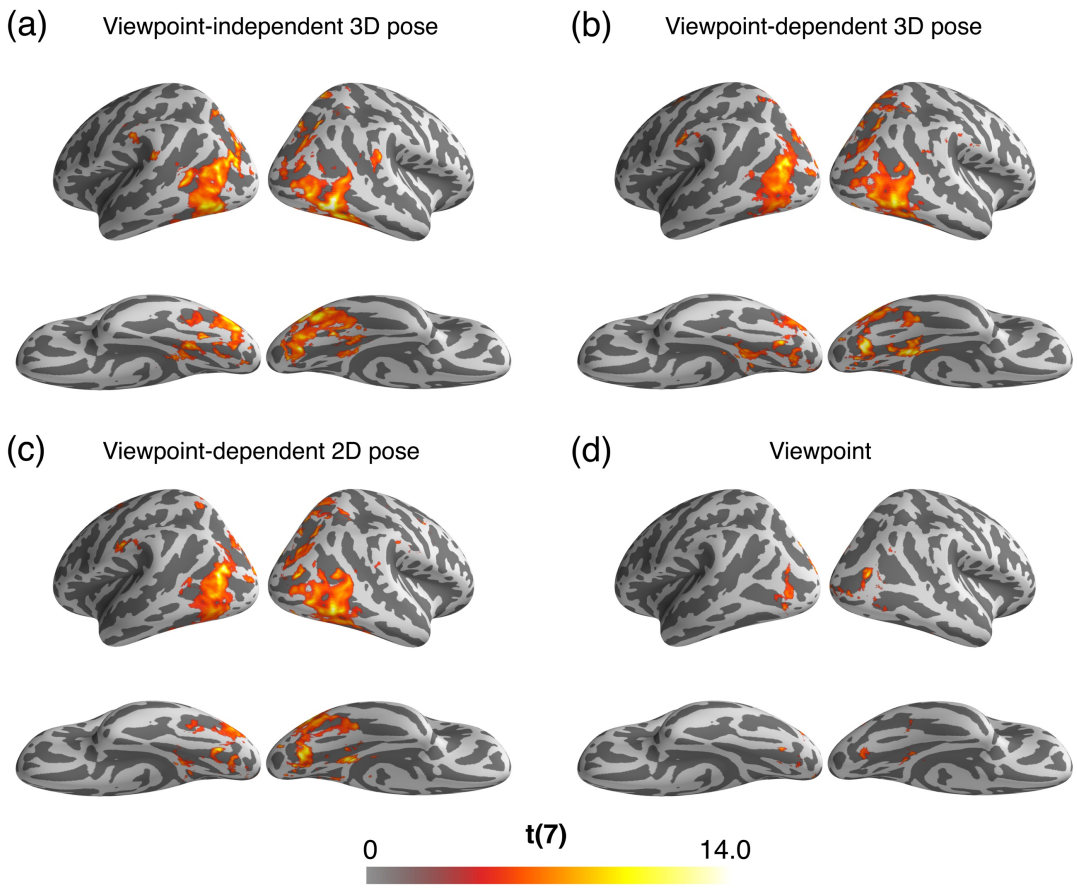

Figure 4.3. The group results of the searchlight-based RSA for the (a) 3D viewpoint-independent pose, (b) 3D viewpoint-dependent pose, (c) 2D viewpoint-dependent pose, and (d) viewpoint models. Color maps show the t values of significant clusters survived a cluster-based nonparametric Monte Carlo permutation test (cluster stat: max sum; initial threshold $p<0.001$, $n=10000$). The correlation maps for each participant were first Fisher transformed to the normal distribution and then the t scores were measured. After the correction for multiple comparisons, the results were converted into z scores (z-value > 1.65 were defined as significant).

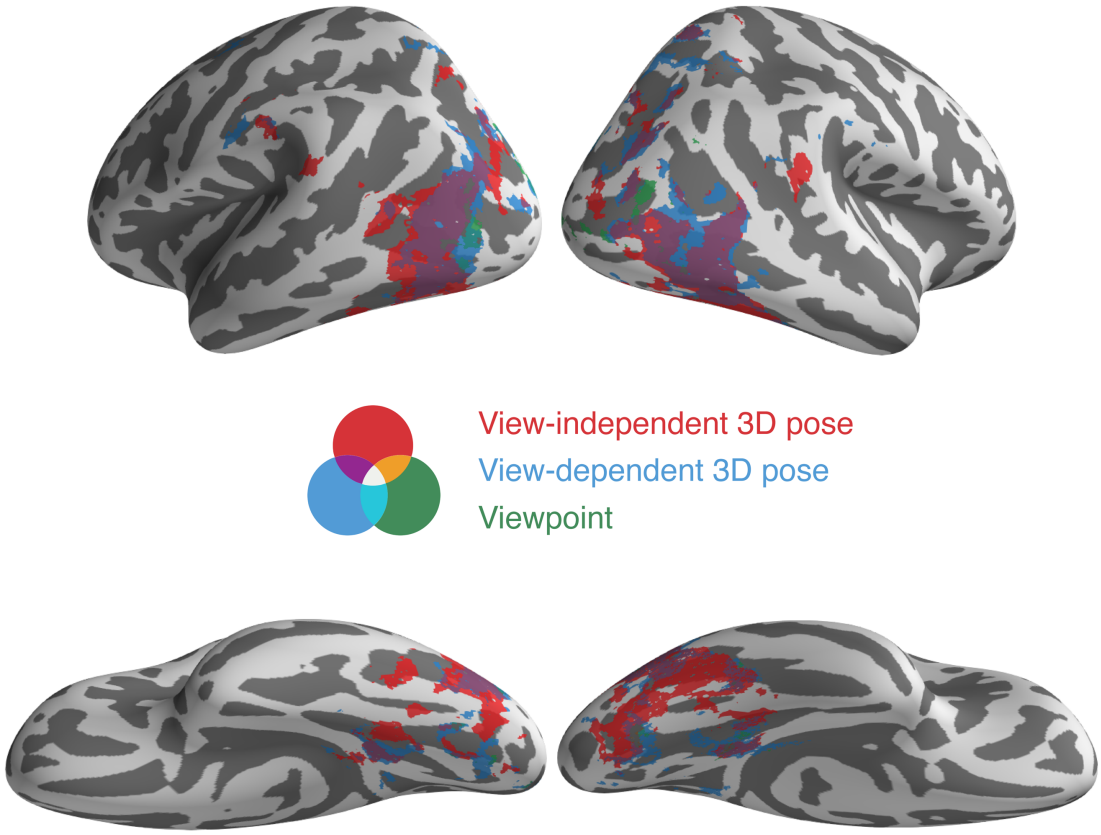

Figure 4.4. Conjunction plots for the significant clusters (z value > 1.65) in view-independent 3D pose (red), view-dependent 3D pose (blue) and viewpoint (green) models. The results showed significant clusters in pSTS for the 3D view-independent pose model only. For the viewpoint model, the significant clusters were mainly near the extrastriate visual cortex and posterior parietal cortex. The results for 3D viewpoint-independent and 3D viewpoint-dependent pose models also revealed significant clusters that overlapped in LOTC, fusiform gyrus, and superior parietal cortex.

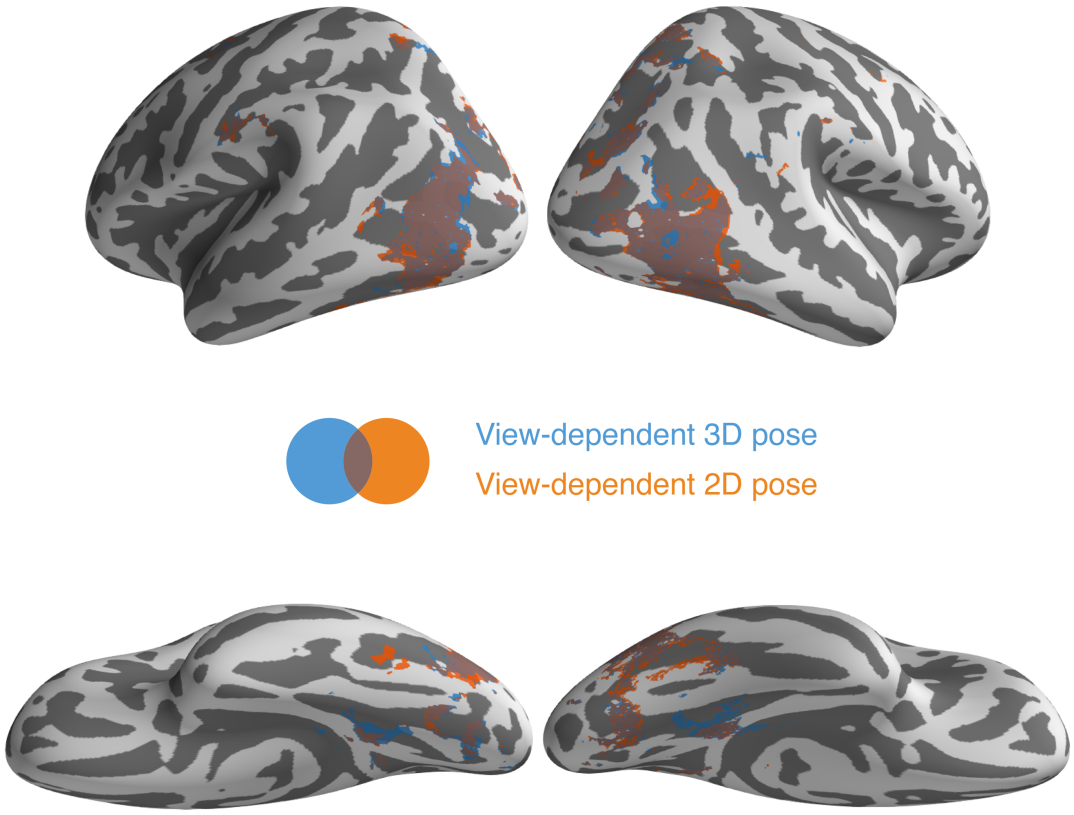

Figure 4.5. Conjunction plots for the significant clusters (z value > 1.65) in view-dependent 3D pose (blue) and view-dependent 2D pose (orange) models. The 3D and 2D view-dependent pose models are highly overlapped in LOTC, fusiform gyrus, and inferior parietal cortex.

DISCUSSION

The representation of the human pose is a central aspect in the computation and interpretation of body actions. While existing research has examined cortical responses to a limited range of human poses from simplified stimuli, here we focused on the spatial organization of cortical sensitivity to a broad range of human poses from complex natural scenes. By introducing an off-the-shelf human 3D mesh reconstruction model, we parameterized natural human poses in a large set of complex natural scene images and built 2D/3D viewpoint-independent and viewpoint-dependent pose models as well as a viewpoint model. RDM analysis

showed that our pose models captured dissociable pose and viewpoint information. Using RSA searchlight, we showed that the dissimilarity structure of a broad range of natural poses was best captured in a set of distributed clusters across the brain, primarily including areas of lateral occipital-temporal cortex (LOTC), fusiform gyrus, and pSTS as well as supramarginal gyrus.

Distributed representation of pose information

The distributed clusters encoding natural pose dissimilarity structure found in our analysis converges with previously reported cortical network encoding viewpoint-independent postures from a limited range of poses (Cross et al., 2010; Urgesi et al., 2007). For example, the 3D viewpoint-independent pose model produced significant clusters in LOTC, covering what is traditionally thought to be specialized for body parts and bodies (Bracci et al., 2015; Orlov et al., 2010; Peelen & Downing, 2005). We observed right-lateralized pose clusters in the fusiform gyrus, consistent with prior work that reported right-lateralized FBA responses to bodies (Hodzic, Kaas, Muckli, Stirn, & Singer, 2009).

However, our results diverge from previous work regarding the type of pose information encoded in the cortical network. As both 2D and 3D viewpoint-dependent pose models produced overlapped clusters with the 3D viewpoint-independent pose model in LOTC, fusiform gyrus, and superior parietal cortex, the pose information encoded in these cortical regions is not necessarily viewpoint independent. Whereas pSTS may indeed encode viewpoint-independent 3D pose information as they captured only the dissimilarity structure from 3D viewpoint-independent pose models. This is in contrast to the previous work that suggested viewpoint-independent encoding of postures across multiple regions including fusiform gyrus, posterior superior temporal sulcus, inferior frontal gyrus, and inferior parietal lobule. For one thing, our results suggested several candidate cortical regions that are likely to encode structured information about human poses. These regions include areas that are traditionally associated with the processing of bodies and body parts as well

as the processing of motion and actions (Grèzes & Decety, 2000; Isik, Koldewyn, Beeler, & Kanwisher, 2017; Peelen & Downing, 2005; Pelphrey et al., 2003; Saxe, Xiao, Kovacs, Perrett, & Kanwisher, 2004). For another, our results also suggested that within this likely distributed cortical network encoding pose structures, 3D viewpoint-independent pose information is likely to be automatically computed and that different aspects of pose information (viewpoint-dependent vs. viewpoint-independent) were encoded in different regions.

Representation of viewpoint information

Besides the distributed representation of pose information, our RSA searchlight using viewpoint RDM identified cortical encoding of viewpoint for bodies mainly near the extrastriate visual cortex with a few extending into the posterior parietal cortex. These clusters bearing explicit viewpoint information are rather localized compared to the distributed pose clusters. Although we found clusters encoding 2D and 3D viewpoint-dependent pose information in some distributed pose clusters, they do not seem to explicitly encode body viewpoint information. Further, both 2D and 3D viewpoint-dependent pose clusters did not emerge near pSTS, where only the 3D viewpoint-independent pose clusters were situated. In the line with these findings, several behavioral studies have shown that human pose representations have more viewpoint invariance when crossing different poses and viewpoints (Sekunova, Black, Parkinson, & Barton, 2013). Our results added evidence suggesting a possible increase in view-tolerant representations along with human pose processing. Given the degree of articulation and wide range of potential viewpoints, it seems plausible to maintain sensitivity to features irrespective of changes in viewpoint and orientation.

We noted that we used the body trunk as the reference frame to determine viewpoint. Hence two poses will be deemed from the same viewpoint as long as their trunks are facing the same direction. Future experiments will be needed to study different reference frames for assessing body orientation and viewpoints, and to

determine the sensitivity to viewpoints across different cortical regions encoding pose structures.

Computational role of pose representation

One strength of our approach is that we structured the pose space with a vast range of parameterized poses covering different ways of parameterization. One future direction is to pin down the specific use of the different aspects of pose information regarding different perception tasks. As the computation of pose information serves as an essential step in the computation of motion and action of a person, the distributed nature of pose representation may be attributed to the various perception tasks (motion, emotion, action, etc.) that pose information supports. It will be an important direction to investigate the role of each local pose cluster in the computation of pose information, and the relationship between the type of pose information encoded and subsequent computation it supports.

CONCLUSION

In conclusion, we present an approach to parameterize three-dimensional human poses from single static images, making explicit different aspects of pose information (e.g. viewpoint-dependent vs. viewpoint independent). With different pose parameterizations, we built several pose models to capture pose dissimilarity structures from a broad range of natural poses. We applied our pose models to a large set of complex natural body images from the Natural Scene Dataset and used searchlight RSA to find cortical regions encoding pose dissimilarity structures. As a result, we found distributed pose clusters encoding pose information in LOTC, pSTS, and superior parietal cortex. In particular, our results suggested that viewpoint-independent pose information is likely to be computed automatically and that pSTS specifically encodes such 3D viewpoint-independent aspects of pose information. Furthermore, we found explicit encodings of body viewpoint information mainly near the extrastriate visual cortex, suggesting the possibly increasing view-tolerant representations along with the human pose processing. Future experiments are

needed to determine the differential contribution of each pose cluster in the computation of different aspects of pose information.

METHODS

Stimulus selection

Natural Scene Dataset (Allen et al., 2021) contains 73,000 cropped color natural scene images from the MS COCO dataset (T. Y. Lin et al., 2014). We aimed to select a subset of images that contain only single persons and cover a broad range of legitimate human body poses. To this aim, we used the ground truth person keypoint annotations provided by the MS COCO dataset. For each person in each image, the annotations consist of an enclosing person bounding box together with two-dimensional image coordinates and visibility flags for 17 defined body keypoints, including 5 face keypoints (L/R eyes, nose, and L/R ears) as well as 12 limb keypoints (L/R shoulders, L/R elbows, L/R wrists, L/R hips, L/R knees, L/R ankles). We selected images with keypoint annotations for one and only one person inside the cropped image regions. As a next step, we further excluded single-person images under partial body presence, namely, where the persons were partially truncated by the image boundary. Specifically, we selected single-person images with 12 limb keypoints fully annotated. Face keypoints (eyes, nose, and ears) were not considered because these annotations were sometimes missing for persons with smaller areas in the images. Finally, we selected a subset of 4450 images of full single persons under different poses.

Pose parameterization

To parameterize natural poses, we first extracted 3D pose information from complex natural scene images. MS COCO dataset does not provide ground truth person keypoint annotations or viewpoint parameters in three dimensions. Therefore, we adopted an approach to use an off-the-shelf human 3D mesh reconstruction model (Kanazawa et al., 2018) to extract 3D pose information. Given a single RGB image

in the wild, this model can reconstruct a full 3D human body mesh. The model was quantitatively evaluated on standard 3D joint estimation benchmarks and outperformed previous approaches that output 3D meshes (Kanazawa et al., 2018). The viewpoint parameter for the 3D human body mesh is an axis-angle representation for the 3D body global rotation in SMPL format (Loper, Mahmood, Romero, Pons-Moll, & Black, 2015). The 3D rotation was transformed into a rotation matrix $R \in \mathbb{R}^{3\times3}$ for further processing. For body pose parameters, we transformed the 3D body mesh into a list of 3D joint locations with a trained joint location regressor (Kanazawa et al., 2018). This joint list includes 19 joints (L/R ankles, L/R knees, L/R hips, L/R wrists, L/R elbows, L/R shoulders, neck, head, nose, L/R eyes, L/R ears). Thus, for each pose, we obtained a rotation matrix R for the body global rotation and a list of $K = 19$ joint locations $p = [J_1, J_2, \dots, J_K]$ where $J_k \in \mathbb{R}^3$. We did not perform additional normalization on these 3D joint coordinates because they were already in the same 3D body mesh reference frame.

To parameterize pose using 3D view-dependent joint locations $p_{3d\_v} = [J_1^{3d\_v}, J_2^{3d\_v}, \dots, J_k^{3d\_v}]$, we simply used these 3D joint coordinates $J_k^{3d\_v} = J_k \in \mathbb{R}^3$.

To parameterize pose using 2D view-dependent joint locations $p_{2d} = [J_1^{(2d)}, J_2^{(2d)}, \dots, J_K^{(2d)}]$, we simply discarded the depth coordinate to make $J_k^{(2d)} \in \mathbb{R}^2$.

To parameterize pose using 3D view-independent aligned joint locations $p_{3d\_vi} = [J_1^{(3d\_vi)}, J_2^{(3d\_vi)}, \dots, J_K^{(3d\_vi)}]$, we reversed the global rotation to align poses to the same, original orientation

$$J_k^{(3d\_vi)} = R^{-1} J_k$$

where R^{-1} is the inverse of the rotation matrix for the 3D global body rotation.

Construction of representational dissimilarity matrices (RDMs)

Once we parameterized each natural pose and obtained 2D view-dependent joint locations and 3D view-independent joint locations as well as 3D global rotations, we construct representational dissimilarity matrices by measuring dissimilarity under different metrics. To construct 3D view-independent aligned pose RDMs, we measured the dissimilarity between two aligned poses using *Mean Per Joint Position Error* which is used in much of the literature on 3D joint estimation. It measures the Euclidean distance averaged on all joints after aligning two poses. Specifically, the dissimilarity between two poses $p_i^{(3d\_vi)}$ and $p_j^{(3d\_vi)}$ is measured as

$$d^{(3d\_vi)}\left(p_i^{(3d\_vi)}, p_j^{(3d\_vi)}\right) = \frac{1}{K}\sum_{k=1}^{K}\left\|J_{ik}^{(3d\_vi)} - J_{jk}^{(3d\_vi)}\right\|_2$$

Similarly, we can construct the 2D and 3D view-dependent pose RDM by measuring dissimilarity as

$$d^{(2d)}\left(p_i^{(2d)}, p_j^{(2d)}\right) = \frac{1}{K}\sum_{k=1}^{K}\left\|J_{ik}^{(2d)} - J_{jk}^{(2d)}\right\|_2$$

$$d^{(3d\_v)}\left(p_i^{(3d\_v)}, p_j^{(3d\_v)}\right) = \frac{1}{K}\sum_{k=1}^{K}\left\|J_{ik}^{(3d\_v)} - J_{jk}^{(3d\_v)}\right\|_2$$

To construct the viewpoint RDM, we measured the viewpoint dissimilarity between pairs of bodies as the distance between the body global rotations in three dimensions. We first transformed the associated 3D rotation matrix $R \in \mathbb{R}^{3\times3}$ into a unit quaternion q. Following (Huynh, 2009), we used the distance metric below to assess the dissimilarity of two 3D body global rotations

$$d^{(v)}\left(q_i, q_j\right) = \cos^{-1}\left|q_i \cdot q_j\right|$$

Representational Similarity Analysis

We carried out the representational similarity analyses (RSA) using a searchlight approach in the individual volume space to investigate the relationship between the computed features and the brain activity. A spherical neighborhood of 100 voxels (approximately 10mm in radius) were used for each searchlight. The multivariate analyses were performed using CosMoMVPA (Oosterhof, Connolly, & Haxby, 2016)

and custom-written MATLAB functions (ver2017b, The MathWorks Inc.). The beta maps with the same images were first averaged, and then the beta maps were normalized across features. The neural RDM was derived using 1-correlation as the distance metric. We selected 3D viewpoint-independent pose RDM, 2D viewpoint-dependent pose RDM, and viewpoint RDM as three target RDMs in our analysis. The neural RDM was correlated with the normalized target RDM for each searchlight. Then, the beta values were assigned to the central voxel of each searchlight in each participant, which resulting in beta maps for each model. The beta maps in individual volume space were then resampled to the standard MNI space and used in a group level analysis to compare the individual beta maps against zero using a one-tailed t-test at each voxel. The t maps were then corrected with a nonparametric cluster-based Monte Carlo permutation test (initial threshold $p<0.001$; 10000 iterations).

Bibliography

Afraz, A., & Cavanagh, P. (2009). The gender-specific face after effect is based in retinotopic not spatiotopic coordinates across several natural image transformations. *Journal of Vision*, *9*(10), 1–17. https://doi.org/10.1167/9.10.10

Alais, D., & Melcher, D. (2007). Strength and coherence of binocular rivalry depends on shared stimulus complexity. *Vision Research*, *47*(2), 269–279. https://doi.org/10.1016/j.visres.2006.09.003

Albright, T. D., & Stoner, G. R. (2002). Contextual influences on visual processing. *Annual Review of Neuroscience*, *25*(1), 339–379. https://doi.org/10.1146/annurev.neuro.25.112701.142900

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Dowdle, L. T., Caron, B., … Kay, K. N. (2021). A massive 7T fMRI dataset to bridge cognitive and computational neuroscience. *BioRxiv*, 1–70. Retrieved from https://doi.org/10.1101/2021.02.22.432340

Axelrod, V., Bar, M., & Rees, G. (2015). Exploring the unconscious using faces. *Trends in Cognitive Sciences*, *19*(1), 35–45. https://doi.org/10.1016/j.tics.2014.11.003

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical Microcircuits for Predictive Coding. *Neuron*, *76*(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038

Benucci, A., Saleem, A. B., & Carandini, M. (2013). Adaptation maintains population homeostasis in primary visual cortex. *Nature Neuroscience*, *16*(6), 724–729. https://doi.org/10.1038/nn.3382

Blakemore, C., & Tobin, E. A. (1972). Lateral inhibition between orientation detectors in the cat's visual cortex. *Experimental Brain Research*, *15*(4), 439–440. https://doi.org/10.1007/BF00234129

Boynton, G. M., & Finney, E. M. (2003). Orientation-specific adaptation in human visual cortex. *Journal of Neuroscience*, *23*(25), 8781–8787. https://doi.org/10.1523/jneurosci.23-25-08781.2003

Bracci, S., Caramazza, A., & Peelen, M. V. (2015). Representational similarity of body parts in human occipitotemporal cortex. *Journal of Neuroscience*, *35*(38), 12977–12985. https://doi.org/10.1523/JNEUROSCI.4698-14.2015

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Burr, D., Tozzi, A., & Morrone, M. C. (2007). Neural mechanisms for timing visual events are spatially selective in real-world coordinates. *Nature Neuroscience*, *10*(4), 423.

Campbell, L. W., & Bobick, A. F. (1995). Recognition of human body motion using phase space constraints. In *IEEE International Conference on Computer Vision* (pp. 624–630). IEEE. https://doi.org/10.1109/iccv.1995.466880

Cavanagh, P., & Anstis, S. (2013). The flash grab effect. *Vision Research*, *91*, 8–20. https://doi.org/10.1016/j.visres.2013.07.007

Cha, O., & Chong, S. C. (2014). The background is remapped across saccades. *Experimental Brain Research*, *232*(2), 609–618. https://doi.org/10.1007/s00221-013-3769-9

Cho, S., & He, S. (2019). Size-invariant but location-specific object-viewpoint adaptation in the absence of awareness. *Cognition*, *192*, 104035. https://doi.org/10.1016/j.cognition.2019.104035

Cicchini, G. M., Binda, P., Burr, D. C., & Morrone, M. C. (2013). Transient spatiotopic integration across saccadic eye movements mediates visual stability. *Journal of Neurophysiology*, *109*(4), 1117–1125. https://doi.org/10.1152/jn.00478.2012

Clifford, C. W.G., Wenderoth, P., & Spehar, B. (2000). A functional angle on some after-effects in cortical vision. *Proceedings of the Royal Society B: Biological Sciences*, *267*(1454), 1705–1710. https://doi.org/10.1098/rspb.2000.1198

Clifford, Colin W.G. (2014). The tilt illusion: Phenomenology and functional implications. *Vision Research*, *104*, 3–11. https://doi.org/10.1016/j.visres.2014.06.009

Clifford, Colin W.G., & Rhodes, G. (2005). *Fitting the Mind to the World: Adaptation and After-Effects in High-Level Vision*. *Fitting the Mind to the World: Adaptation and After-Effects in High-Level Vision* (Vol. 2). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198529699.001.0001

Cohen, M. A., Cavanagh, P., Chun, M. M., & Nakayama, K. (2012). The attentional requirements of consciousness. *Trends in Cognitive Sciences*, *16*(8), 411–417. https://doi.org/10.1016/j.tics.2012.06.013

Cornelissen, F. W., Wade, A. R., Vladusich, T., Dougherty, R. F., & Wandell, B. A. (2006). No functional magnetic resonance imaging evidence for brightness and

color filling-in in early human visual cortex. *Journal of Neuroscience*, *26*(14), 3634–3641. https://doi.org/10.1523/JNEUROSCI.4382-05.2006

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

Crapse, T. B., & Sommer, M. A. (2012). Frontal eye field neurons assess visual stability across saccades. *Journal of Neuroscience*, *32*(8), 2835–2845. https://doi.org/10.1523/JNEUROSCI.1320-11.2012

Crespi, S., Biagi, L., d'Avossa, G., Burr, D. C., Tosetti, M., & Morrone, M. C. (2011). Spatiotopic coding of BOLD signal in human visual cortex depends on spatial attention. *PLoS ONE*, *6*(7), e21661. https://doi.org/10.1371/journal.pone.0021661

Cross, E. S., MacKie, E. C., Wolford, G., & Antonia, A. F. (2010). Contorted and ordinary body postures in the human brain. *Experimental Brain Research*, *204*(3), 397–407. https://doi.org/10.1007/s00221-009-2093-x

D'Avossa, G., Tosetti, M., Crespi, S., Biagi, L., Burr, D. C., & Morrone, M. C. (2007). Spatiotopic selectivity of BOLD responses to visual motion in human area MT. *Nature Neuroscience*, *10*(2), 249–255. https://doi.org/10.1038/nn1824

De Martino, F., Moerel, M., Ugurbil, K., Goebel, R., Yacoub, E., & Formisano, E. (2015). Frequency preference and attention effects across cortical depths in the human primary auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(52), 16036–16041. https://doi.org/10.1073/pnas.1507552112

De Sousa, A. A., Sherwood, C. C., Schleicher, A., Amunts, K., MacLeod, C. E., Hof, P. R., & Zilles, K. (2010). Comparative cytoarchitectural analyses of striate and extrastriate areas in hominoids. *Cerebral Cortex*, *20*(4), 966–981. https://doi.org/10.1093/cercor/bhp158

Demeyer, M., De Graef, P., Wagemans, J., & Verfaillie, K. (2009). Transsaccadic identification of highly similar artificial shapes. *Journal of Vision*, *9*(4), 28. https://doi.org/10.1167/9.4.28

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., … Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Downing, P., & Kanwisher, N. (2001). A cortical area specialized for visual processing of the human body. *Journal of Vision*, *1*(3), 2470–2473. https://doi.org/10.1167/1.3.341

Duhamel, J. R., Bremmer, F., BenHamed, S., & Graf, W. (1997). Spatial invariance of visual receptive fields in parietal cortex neurons. *Nature*, *389*(6653), 845–848. https://doi.org/10.1038/39865

Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, *255*(5040), 90–92. https://doi.org/10.1126/science.1553535

Dunkley, B. T., Baltaretu, B., & Crawford, J. D. (2016). Trans-saccadic interactions in human parietal and occipital cortex during the retention and comparison of object orientation. *Cortex*, *82*, 263–276. https://doi.org/10.1016/j.cortex.2016.06.012

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex*, *7*(2), 181–192. https://doi.org/10.1093/cercor/7.2.181

Fabius, J. H., Fracasso, A., Nijboer, T. C. W., & Van Der Stigchel, S. (2019). Time course of spatiotopic updating across saccades. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(6), 2027–2032. https://doi.org/10.1073/pnas.1812210116

Fabius, J. H., Fracasso, A., & Van Der Stigchel, S. (2016). Spatiotopic updating facilitates perception immediately after saccades. *Scientific Reports*, *6*(1), 1–11. https://doi.org/10.1038/srep34488

Fairhall, S. L., Schwarzbach, J., Lingnau, A., Van Koningsbruggen, M. G., & Melcher, D. (2017). Spatiotopic updating across saccades revealed by spatially-specific fMRI adaptation. *NeuroImage*, *147*, 339–345. https://doi.org/10.1016/j.neuroimage.2016.11.071

Fang, F., & He, S. (2005). Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nature Neuroscience*, *8*(10), 1380–1385.

https://doi.org/10.1038/nn1537

Fang, F., Murray, S. O., Kersten, D., & He, S. (2005). Orientation-tuned fMRI adaptation in human visual cortex. *Journal of Neurophysiology*, *94*(6), 4188–4195. https://doi.org/10.1152/jn.00378.2005

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1–47. https://doi.org/10.1093/cercor/1.1.1-a

Forte, J. D., & Clifford, C. W. G. (2005). Inter-ocular transfer of the tilt illusion shows that monocular orientation mechanisms are colour selective. *Vision Research*, *45*(20), 2715–2721. https://doi.org/10.1016/j.visres.2005.05.001

Fracasso, A., Caramazza, A., & Melcher, D. (2010). Continuous perception of motion and shape across saccadic eye movements. *Journal of Vision*, *10*(13), 14. https://doi.org/10.1167/10.13.14

Fukiage, T., & Murakami, I. (2010). The tilt aftereffect occurs independently of the flash-lag effect. *Vision Research*, *50*(19), 1949–1956. https://doi.org/10.1016/j.visres.2010.07.002

Fukiage, T., & Murakami, I. (2013). Adaptation to a spatial offset occurs independently of the flash-drag effect. *Journal of Vision*, *13*(2), 7. https://doi.org/10.1167/13.2.7

Ganmor, E., Landy, M. S., & Simoncelli, E. P. (2015). Near-optimal integration of orientation information across saccades. *Journal of Vision*, *15*(16), 8. https://doi.org/10.1167/15.16.8

Gardner, J. L., Merriam, E. P., Movshon, J. A., & Heeger, D. J. (2008). Maps of visual space in human occipital cortex are retinotopic, not spatiotopic. *Journal of Neuroscience*, *28*(15), 3988–3999. https://doi.org/10.1523/JNEUROSCI.5476-07.2008

Georgeson, M. (2004). Visual aftereffects: Cortical neurons change their tune. *Current Biology*, *14*(18), R751–R753. https://doi.org/10.1016/j.cub.2004.09.011

Gibson, J. J., & Radner, M. (1937). Adaptation, after-effect and contrast in the perception of tilted lines. *Journal of Experimental Psychology*, *20*(5), 453–467. https://doi.org/10.1037/h0059826

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, *14*(5), 350–363. https://doi.org/10.1038/nrn3476

Goddard, E., Solomon, S., & Clifford, C. (2010). Adaptable mechanisms sensitive to surface color in human vision. *Journal of Vision*, *10*(9), 17. https://doi.org/10.1167/10.9.17

Goebel, R., Esposito, F., & Formisano, E. (2006). Analysis of Functional Image Analysis Contest (FIAC) data with BrainVoyager QX: From single-subject to cortically aligned group General Linear Model analysis and self-organizing group Independent Component Analysis. *Human Brain Mapping*, *27*(5), 392–401. https://doi.org/10.1002/hbm.20249

Golomb, J. D., Marino, A. C., Chun, M. M., & Mazer, J. A. (2011). Attention doesn't slide: Spatiotopic updating after eye movements instantiates a new, discrete attentional locus. *Attention, Perception, and Psychophysics*, *73*(1), 7–14. https://doi.org/10.3758/s13414-010-0016-3

Golomb, J. D., Nguyen-Phuc, A. Y., Mazer, J. A., McCarthy, G., & Chun, M. M. (2010). Attentional facilitation throughout human visual cortex lingers in retinotopic coordinates after eye movements. *Journal of Neuroscience*, *30*(31), 10493–10506. https://doi.org/10.1523/JNEUROSCI.1546-10.2010

Golomb, J. D., Pulido, V. Z., Albrecht, A. R., Chun, M. M., & Mazer, J. A. (2010). Robustness of the retinotopic attentional trace after eye movements. *Journal of Vision*, *10*(3), 1–12. https://doi.org/10.1167/10.3.19

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., … Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*(7 DEC), 267. https://doi.org/10.3389/fnins.2013.00267

Grèzes, J., & Decety, J. (2000). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping*, *12*(1), 1–19. https://doi.org/10.1002/1097-0193(200101)12:1<1::AID-HBM10>3.0.CO;2-V

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data. *Journal of Cognitive Neuroscience*, *29*(4), 677–697. https://doi.org/10.1162/jocn_a_01068

Hasson, U., & Honey, C. J. (2012). Future trends in Neuroimaging: Neural processes as expressed within real-life contexts. *NeuroImage*, *62*(2), 1272–1278.

https://doi.org/10.1016/j.neuroimage.2012.02.004

He, D., Mo, C., & Fang, F. (2017). Predictive feature remapping before saccadic eye movements. *Journal of Vision*, *17*(5), 14. https://doi.org/10.1167/17.5.14

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, *383*(6598), 334–337. https://doi.org/10.1038/383334a0

He, Sheng, & MacLeod, D. I. A. (2001). Orientation-selective adaptation and tilt after-effect from invisible patterns. *Nature*, *411*(6836), 473–476. https://doi.org/10.1038/35078072

He, T., Fritsche, M., & Lange de, F. P. (2018). Predictive remapping of visual features beyond saccadic targets. *BioRxiv*, *18*(13), 20. https://doi.org/10.1101/297481

Hiebert, E. N. (1996). *Science and Culture: Popular and Philosophical Essays . Hermann von Helmholtz , David Cahan* . *Isis* (Vol. 87). University of Chicago Press. https://doi.org/10.1086/357539

Hodzic, A., Kaas, A., Muckli, L., Stirn, A., & Singer, W. (2009). Distinct cortical networks for the detection and identification of human body. *NeuroImage*, *45*(4), 1264–1271. https://doi.org/10.1016/j.neuroimage.2009.01.027

Hogendoorn, H., Verstraten, F. A. J., & Cavanagh, P. (2015). Strikingly rapid neural basis of motion-induced position shifts revealed by high temporal-resolution EEG pattern classification. *Vision Research*, *113*(PA), 1–10. https://doi.org/10.1016/j.visres.2015.05.005

Huynh, D. Q. (2009). Metrics for 3D rotations: Comparison and analysis. *Journal of Mathematical Imaging and Vision*, *35*(2), 155–164. https://doi.org/10.1007/s10851-009-0161-2

Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(43), E9145–E9152. https://doi.org/10.1073/pnas.1714471114

Jiang, Y., Costello, P., Fang, F., Huang, M., & He, S. (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(45), 17048–17052. https://doi.org/10.1073/pnas.0605678103

Jin, D. Z., Dragoi, V., Sur, M., & Seung, H. S. (2005). Tilt aftereffect and adaptation-

induced changes in orientation tuning in visual cortex. *Journal of Neurophysiology*, *94*(6), 4038–4050. https://doi.org/10.1152/jn.00571.2004

Kanazawa, A., Black, M. J., Jacobs, D. W., & Malik, J. (2018). End-to-End Recovery of Human Shape and Pose. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 7122–7131). https://doi.org/10.1109/CVPR.2018.00744

Kaunitz, L., Fracasso, A., & Melcher, D. (2011). Unseen complex motion is modulated by attention and generates a visible aftereffect. *Journal of Vision*, *11*(13), 10. https://doi.org/10.1167/11.13.10

Kim, C. Y., & Blake, R. (2005). Psychophysical magic: Rendering the visible "invisible." *Trends in Cognitive Sciences*, *9*(8), 381–388. https://doi.org/10.1016/j.tics.2005.06.012

Klein, B. P., Fracasso, A., van Dijk, J. A., Paffen, C. L. E., te Pas, S. F., & Dumoulin, S. O. (2018). Cortical depth dependent population receptive field attraction by spatial attention in human V1. *NeuroImage*, *176*, 301–312. https://doi.org/10.1016/j.neuroimage.2018.04.055

Kohler, P. J., Cavanagh, P., & Tse, P. U. (2017). Motion-induced position shifts activate early visual cortex. *Frontiers in Neuroscience*, *11*(APR), 168. https://doi.org/10.3389/fnins.2017.00168

Kohn, A. (2007). Visual adaptation: Physiology, mechanisms, and functional benefits. *Journal of Neurophysiology*, *97*(5), 3155–3164. https://doi.org/10.1152/jn.00086.2007

Kohn, A., & Movshon, J. A. (2003). Neuronal adaptation to visual motion in area MT of the macaque. *Neuron*, *39*(4), 681–691. https://doi.org/10.1016/S0896-6273(03)00438-0

Kok, P., Bains, L. J., Van Mourik, T., Norris, D. G., & De Lange, F. P. (2016). Selective activation of the deep layers of the human primary visual cortex by top-down feedback. *Current Biology*, *26*(3), 371–376. https://doi.org/10.1016/j.cub.2015.12.038

Kosovicheva, A. A., Maus, G. W., Anstis, S., Cavanagh, P., Tse, P. U., & Whitney, D. (2012). The motion-induced shift in the perceived location of a grating also shifts its aftereffect. *Journal of Vision*, *12*(8), 1–4. https://doi.org/10.1167/12.8.7

Krauskopf, J., & Zaidi, Q. (1986). Induced desensitization. *Vision Research*, *26*(5), 759–762. https://doi.org/10.1016/0042-6989(86)90090-8

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(NOV), 4. https://doi.org/10.3389/neuro.06.004.2008

Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition*, *65*(2), 145–168. https://doi.org/10.1016/j.bandc.2007.06.007

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579. https://doi.org/10.1016/S0166-2236(00)01657-X

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., … Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 8693 LNCS, pp. 740–755). Springer. https://doi.org/10.1007/978-3-319-10602-1_48

Lin, Z., & He, S. (2009). Seeing the invisible: The scope and limits of unconscious processing in binocular rivalry. *Progress in Neurobiology*, *87*(4), 195–211. https://doi.org/10.1016/j.pneurobio.2008.09.002

Liu, T., Heeger, D. J., & Carrasco, M. (2006). Neural correlates of the visual vertical meridian asymmetry. *Journal of Vision*, *6*(11), 12. https://doi.org/10.1167/6.11.12

Liu, T., Larsson, J., & Carrasco, M. (2007). Feature-Based Attention Modulates Orientation-Selective Responses in Human Visual Cortex. *Neuron*, *55*(2), 313–323. https://doi.org/10.1016/j.neuron.2007.06.030

Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., & Black, M. J. (2015). SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics*, *34*(6), 1–16. https://doi.org/10.1145/2816795.2818013

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data ಽ , ಽಽ, *164*, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Mathôt, S., & Theeuwes, J. (2010). Gradual remapping results in early retinotopic and late spatiotopic inhibition of return. *Psychological Science*, *21*(12), 1793–1798. https://doi.org/10.1177/0956797610388813

Melcher, D. (2005). Spatiotopic transfer of visual-form adaptation across saccadic eye movements. *Current Biology*, *15*(19), 1745–1748. https://doi.org/10.1016/j.cub.2005.08.044

Melcher, D. (2008). Dynamic, object-based remapping of visual features in trans-saccadic perception. *Journal of Vision*, *8*(14), 2. https://doi.org/10.1167/8.14.2

Melcher, D. (2009). Selective attention and the active remapping of object features in trans-saccadic perception. *Vision Research*, *49*(10), 1249–1255. https://doi.org/10.1016/j.visres.2008.03.014

Melcher, D. (2011). Visual stability. *Philosophical Transactions of the Royal Society B: Biological Sciences*. article, England: The Royal Society. https://doi.org/10.1098/rstb.2010.0277

Melcher, D., & Colby, C. L. (2008). Trans-saccadic perception. *Trends in Cognitive Sciences*, *12*(12), 466–473.

Melcher, D., & Fracasso, A. (2012). Remapping of the line motion illusion across eye movements. *Experimental Brain Research*, *218*(4), 503–514. https://doi.org/10.1007/s00221-012-3043-6

Melcher, D., & Morrone, C. (2003). Spatiotopic temporal integration of motion across saccades. *Journal of Vision*, *3*(9), 877–881. https://doi.org/10.1167/3.9.172

Merriam, E. P., Gardner, J. L., Movshon, J. A., & Heeger, D. J. (2013). Modulation of visual responses by gaze direction in human visual cortex. *Journal of Neuroscience*, *33*(24), 9879–9889. https://doi.org/10.1523/JNEUROSCI.0500-12.2013

Michelson, A. A. (1995). *Studies in optics TT -. Dover Books on Physics; Dover Books on Physics. TA -*. Courier Corporation.

Mohr, H. M., Linder, N. S., Dennis, H., & Sireteanu, R. (2011). Orientation-specific aftereffects to mentally generated lines. *Perception*, *40*(3), 272–290. https://doi.org/10.1068/p6781

Mohr, H. M., Linder, N. S., Linden, D. E. J., Kaiser, J., & Sireteanu, R. (2009). Orientation-specific adaptation to mentally generated lines in human visual cortex.

NeuroImage, *47*(1), 384–391. https://doi.org/10.1016/j.neuroimage.2009.03.045

Morrone, M. C., Cicchini, M., & Burr, D. C. (2010). Spatial maps for time and motion. *Experimental Brain Research*, *206*(2), 121–128. https://doi.org/10.1007/s00221-010-2334-z

Muckli, L., De Martino, F., Vizioli, L., Petro, L. S., Smith, F. W., Ugurbil, K., … Yacoub, E. (2015). Contextual Feedback to Superficial Layers of V1. *Current Biology*, *25*(20), 2690–2695. https://doi.org/10.1016/j.cub.2015.08.057

Muckli, L., Kohler, A., Kriegeskorte, N., & Singer, W. (2005). Primary visual cortex activity along the apparent-motion trace reflects illusory perception. *PLoS Biology*, *3*(8), e265. https://doi.org/10.1371/journal.pbio.0030265

Murray, S. O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nature Neuroscience*, *9*(3), 429–434. https://doi.org/10.1038/nn1641

Nakashima, Y., & Sugita, Y. (2017). The reference frame of the tilt aftereffect measured by differential Pavlovian conditioning. *Scientific Reports*, *7*(1), 1–11. https://doi.org/10.1038/srep40525

Nandy, A. S., Sharpee, T. O., Reynolds, J. H., & Mitchell, J. F. (2013). The Fine Structure of Shape Tuning in Area V4. *Neuron*, *78*(6), 1102–1115. https://doi.org/10.1016/j.neuron.2013.04.016

Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Current Opinion in Neurobiology*, *14*(4), 474–480. https://doi.org/10.1016/j.conb.2004.06.005

Nichols, T., & Holmes, A. (2003). Nonparametric Permutation Tests for Functional Neuroimaging. *Human Brain Function: Second Edition*, *15*(1), 887–910. https://doi.org/10.1016/B978-012264841-0/50048-2

Noudoost, B., Chang, M. H., Steinmetz, N. A., & Moore, T. (2010). Top-down control of visual attention. *Current Opinion in Neurobiology*, *20*(2), 183–190. https://doi.org/10.1016/j.conb.2010.02.003

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-modal multivariate pattern analysis of neuroimaging data in matlab/GNU octave. *Frontiers in Neuroinformatics*, *10*(JUL), 27. https://doi.org/10.3389/fninf.2016.00027

Orlov, T., Makin, T. R., & Zohary, E. (2010). Topographic Representation of the Human

Body in the Occipitotemporal Cortex. *Neuron*, *68*(3), 586–600. https://doi.org/10.1016/j.neuron.2010.09.032

Pedregosa, F., Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., & Mueller, A. (2015). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research 12*, *19*(1), 29–33.

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology*, *93*(1), 603–608. https://doi.org/10.1152/jn.00513.2004

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Pelphrey, K. A., Mitchell, T. V., McKeown, M. J., Goldstein, J., Allison, T., & McCarthy, G. (2003). Brain activity evoked by the perception of human walking: Controlling for meaningful coherent motion. *Journal of Neuroscience*, *23*(17), 6819–6825. https://doi.org/10.1523/jneurosci.23-17-06819.2003

Pizlo, Z. (2001). Perception viewed as an inverse problem. *Vision Research*, *41*(24), 3145–3161. https://doi.org/10.1016/S0042-6989(01)00173-0

Ro, T., Breitmeyer, B., Burton, P., Singhal, N. S., & Lane, D. (2003). Feedback contributions to visual awareness in human occipital cortex. *Current Biology*, *13*(12), 1038–1041. https://doi.org/10.1016/S0960-9822(03)00337-3

Rushton, W. A. H. (1965). The Ferrier Lecture, 1962 Visual adaptation. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *162*(986), 20–46. https://doi.org/10.1098/rspb.1965.0024

Saxe, R., Xiao, D. K., Kovacs, G., Perrett, D. I., & Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia*, *42*(11), 1435–1446. https://doi.org/10.1016/j.neuropsychologia.2004.04.015

Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research*, *122*, 105–123. https://doi.org/10.1016/j.visres.2016.02.002

Schwartz, O., Hsu, A., & Dayan, P. (2007). Space and time in visual context. *Nature*

Reviews Neuroscience, *8*(7), 522–535. https://doi.org/10.1038/nrn2155

Sekunova, A., Black, M., Parkinson, L., & Barton, J. J. S. (2013). Viewpoint and pose in body-form adaptation. *Perception*, *42*(2), 176–186. https://doi.org/10.1068/p7265

Self, M. W., van Kerkoerle, T., Goebel, R., & Roelfsema, P. R. (2019). Benchmarking laminar fMRI: Neuronal spiking and synaptic activity during top-down and bottom-up processing in the different layers of cortex. *NeuroImage*, *197*, 806–817. https://doi.org/10.1016/j.neuroimage.2017.06.045

Self, M. W., van Kerkoerle, T., Supèr, H., & Roelfsema, P. R. (2013). Distinct Roles of the Cortical Layers of Area V1 in Figure-Ground Segregation. *Current Biology*, *23*(21), 2121–2129. https://doi.org/10.1016/j.cub.2013.09.013

Solomon, S. G., & Kohn, A. (2014). Moving sensory adaptation beyond suppressive effects in single neurons. *Current Biology*, *24*(20), R1012–R1022. https://doi.org/10.1016/j.cub.2014.09.001

Stein, T., & Sterzer, P. (2014). Unconscious processing under interocular suppression: Getting the right measure. *Frontiers in Psychology*, *5*(MAY), 387. https://doi.org/10.3389/fpsyg.2014.00387

Sterzer, P., Stein, T., Ludwig, K., Rothkirch, M., & Hesselmann, G. (2014). Neural processing of visual information under interocular suppression: A critical review. *Frontiers in Psychology*, *5*(MAY), 453. https://doi.org/10.3389/fpsyg.2014.00453

Szinte, M., Jonikaitis, D., Rangelov, D., & Deubel, H. (2018). Pre-saccadic remapping relies on dynamics of spatial attention. *Elife*, *7*, e37598.

Thompson, P., & Burr, D. (2009). Visual aftereffects. *Current Biology*, *19*(1), R11–R14. https://doi.org/10.1016/j.cub.2008.10.014

Tolias, A. S., Moore, T., Smirnakis, S. M., Tehovnik, E. J., Siapas, A. G., & Schiller, P. H. (2001). Eye movements modulate visual receptive fields of V4 neurons. *Neuron*, *29*(3), 757–767. https://doi.org/10.1016/S0896-6273(01)00250-1

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience*, *8*(8), 1096–1101. https://doi.org/10.1038/nn1500

Turi, M., & Burr, D. (2012). Spatiotopic perceptual maps in humans: Evidence from motion adaptation. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1740), 3091–3097. https://doi.org/10.1098/rspb.2012.0637

Urgesi, C., Candidi, M., Ionta, S., & Aglioti, S. M. (2007). Representation of body identity and body actions in extrastriate body area and ventral premotor cortex. *Nature Neuroscience*, *10*(1), 30–31. https://doi.org/10.1038/nn1815

Van de Moortele, P. F., Auerbach, E. J., Olman, C., Yacoub, E., Uğurbil, K., & Moeller, S. (2009). T1 weighted brain images at 7 Tesla unbiased for Proton Density, T2* contrast and RF coil receive B1 sensitivity with simultaneous vessel visualization. *NeuroImage*, *46*(2), 432–446. https://doi.org/10.1016/j.neuroimage.2009.02.009

van Kerkoerle, T., Self, M. W., & Roelfsema, P. R. (2017). Erratum: Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nature Communications*, *8*(1), 15555. https://doi.org/10.1038/ncomms15555

Wagstyl, K., Lepage, C., Bludau, S., Zilles, K., Fletcher, P. C., Amunts, K., & Evans, A. C. (2018). Mapping cortical laminar structure in the 3D bigbrain. *Cerebral Cortex*, *28*(7), 2551–2562. https://doi.org/10.1093/cercor/bhy074

Wang, C., Wang, Y., Lin, Z., & Yuille, A. L. (2019). Robust 3D human pose estimation from single images or video sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(5), 1227–1241. https://doi.org/10.1109/TPAMI.2018.2828427

Wang, C., Wang, Y., & Yuille, A. L. (2013). An approach to pose-based action recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 915–922). https://doi.org/10.1109/CVPR.2013.123

Wolfe, B. A., & Whitney, D. (2015). Saccadic remapping of object-selective information. *Attention, Perception, and Psychophysics*, *77*(7), 2260–2269. https://doi.org/10.3758/s13414-015-0944-z

Wurtz, R. H., Joiner, W. M., & Berman, R. A. (2011). Neuronal mechanisms for visual stability: Progress and problems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1564), 492–503. https://doi.org/10.1098/rstb.2010.0186

Wutz, A., Drewes, J., & Melcher, D. (2016). Nonretinotopic perception of orientation: Temporal integration of basic features operates in object-based coordinates. *Journal of Vision*, *16*(10), 3. https://doi.org/10.1167/16.10.3

Yacoob, Y., & Black, M. J. (1999). Parameterized Modeling and Recognition of Activities. *Computer Vision and Image Understanding*, *73*(2), 232–247.

https://doi.org/10.1006/cviu.1998.0726

Yang, E., Brascamp, J., Kang, M. S., & Blake, R. (2014). On the use of continuous flash suppression for the study of visual processing outside of awareness. *Frontiers in Psychology*, *5*(JUL), 724. https://doi.org/10.3389/fpsyg.2014.00724

Yang, E., Hong, S. W., & Blake, R. (2010). Adaptation aftereffects to facial expressions suppressed from visual awareness. *Journal of Vision*, *10*(12), 1–13. https://doi.org/10.1167/10.12.24

Zaidi, Q., & Sachtler, W. L. (1991). Motion adaptation from surrounding stimuli. *Perception*, *20*(6), 703–714. https://doi.org/10.1068/p200703

Zimmermann, E., Morrone, M. C., & Burr, D. (2015). Visual mislocalization during saccade sequences. *Experimental Brain Research*, *233*(2), 577–585.

Zimmermann, E., Morrone, M. C., & Burr, D. C. (2014). Buildup of spatial information over time and across eye-movements. *Behavioural Brain Research*, *275*, 281–287. https://doi.org/10.1016/j.bbr.2014.09.013

Zimmermann, E., Morrone, M. C., Fink, G. R., & Burr, D. (2013). Spatiotopic neural representations develop slowly across saccades. *Current Biology*, *23*(5), R193–R194. https://doi.org/10.1016/j.cub.2013.01.065

Zimmermann, E., Weidner, R., Abdollahi, R. O., & Fink, G. R. (2016). Spatiotopic adaptation in visual areas. *Journal of Neuroscience*, *36*(37), 9526–9534. https://doi.org/10.1523/JNEUROSCI.0052-16.2016

Zimmermann, E., Weidner, R., & Fink, G. R. (2017). Spatiotopic updating of visual feature information. *Journal of Vision*, *17*(12), 6. https://doi.org/10.1167/17.12.6

Zirnsak, M., Gerhards, R. G. K., Kiani, R., Lappe, M., & Hamker, F. H. (2011). Anticipatory saccade target processing and the presaccadic transfer of visual features. *Journal of Neuroscience*, *31*(49), 17887–17891. https://doi.org/10.1523/JNEUROSCI.2465-11.2011

Appendix 1: Supplemental Information for Chapter 2

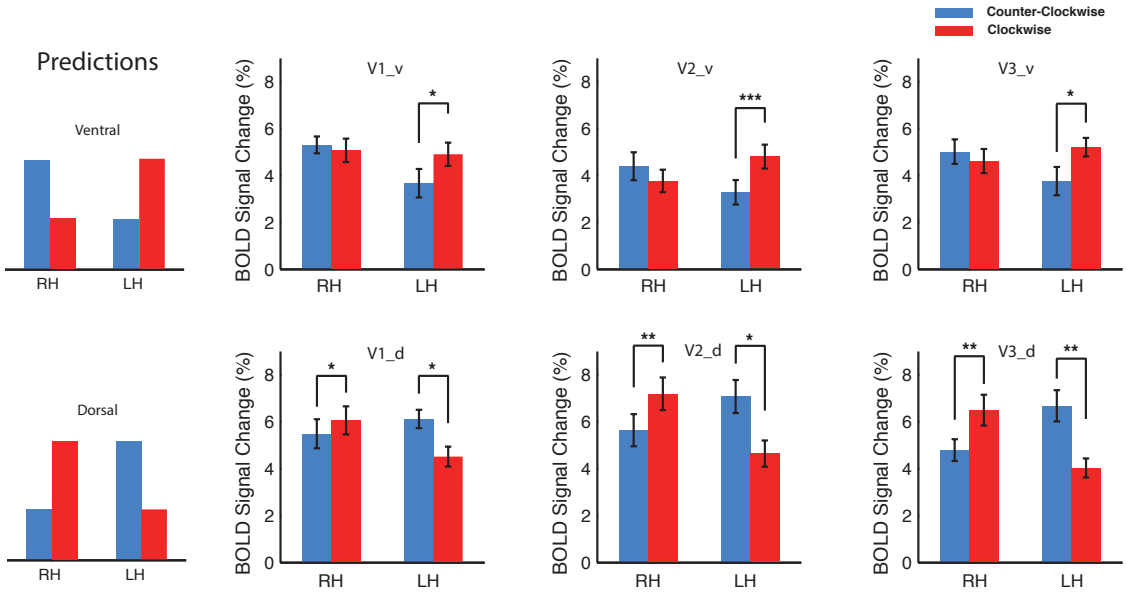

Figure A1.1. Hemi-visual field fMRI response to the flash grab illusion. Upper row shows data from the upper visual field (ventral part of visual cortex); lower row shows data from the lower visual field (dorsal part of visual cortex). Compared to the counter-clockwise illusion, fMRI response to the clockwise tilted illusion was stronger in the left ventral and right dorsal visual cortex, but weaker in the left dorsal and right ventral visual cortex. The left column shows predictions for the illusory representation in the visual retinotopic cortex. The right three columns show the fMRI responses to the clockwise and counter-clockwise tilted illusion from different quadrants of the visual field in early visual cortices from V1 to V3. Three-way repeated measures ANOVA revealed a significant three-way interaction in V1 across dorsal/ventral, left/right hemisphere, and clockwise/counter-clockwise illusion ($F(1, 8) = 38.11$, $p < 0.001$). In the ventral part of V1 (corresponding to the upper visual field), compared to the counter-clockwise condition, clockwise illusion produced stronger fMRI signals in the left hemisphere (corresponding to the right visual field), and weaker response in the right hemisphere, resulting in a significant interaction between left/right hemisphere and clockwise/counter-clockwise illusion ($F(1, 8) = 5.55$, $p = 0.046$) in a two-way repeated measures ANOVA. The opposite was true for the dorsal part of V1: BOLD response to the clockwise condition was weaker in the left hemisphere and stronger in the right hemisphere ($F(1, 8) = 14.16$, $p = 0.006$). Similar results were found for V2 (V2_v/upper: $F(1, 8) = 39.36$, $p < 0.001$; V2_d/lower: $F(1, 8) = 11.60$, $p = 0.009$; three-way interaction: $F(1, 8) = 39.83$, $p < 0.001$), and V3 (V3_v/upper: $F(1, 8) = 12.50$, $p = 0.077$; V3_v/lower: $F(1, 8) = 23.09$, $p = 0.001$; three-way interaction: $F(1, 8) = 28.56$, $p < 0.001$). Simple effects of CW vs CCW conditions (stars in the figure, two-sided paired t-test) were not further corrected beyond the protection of a significant ANOVA. The error bars indicate standard error of mean (n=9 individuals). Source data are provided as a Source Data file.

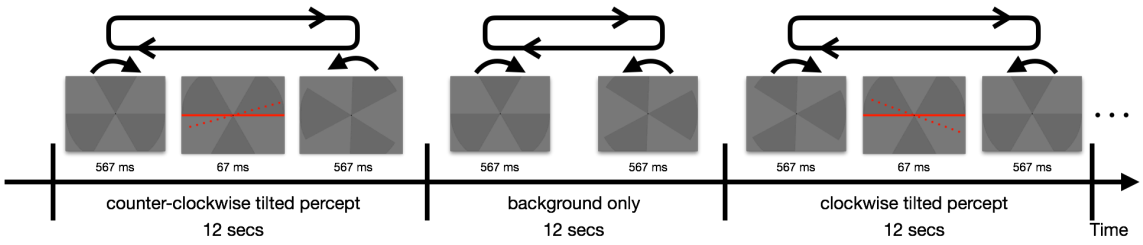

Figure A1.2. Schematic diagram of stimuli and procedures for the 7T fMRI experiment. In each block, a red bar repeatedly presented(flashed) at the reversal point of the pinwheel disc which is rotating back and forth for 12 seconds as a constant background , alternating with 12 seconds rotating background-only stimulus section. The bar would be percieved as tilted clockwise or counter-clockwise from the horizontal meridian, depended on the direction of motion reversal. Red solid lines indicate the presented position of the bar, while red dotted lines illustrate the perceived position. The bar rotation covered a section of both left and right visual fields between 16 degrees clockwise and 15 degrees counter-clockwise from the horizontal meridian.

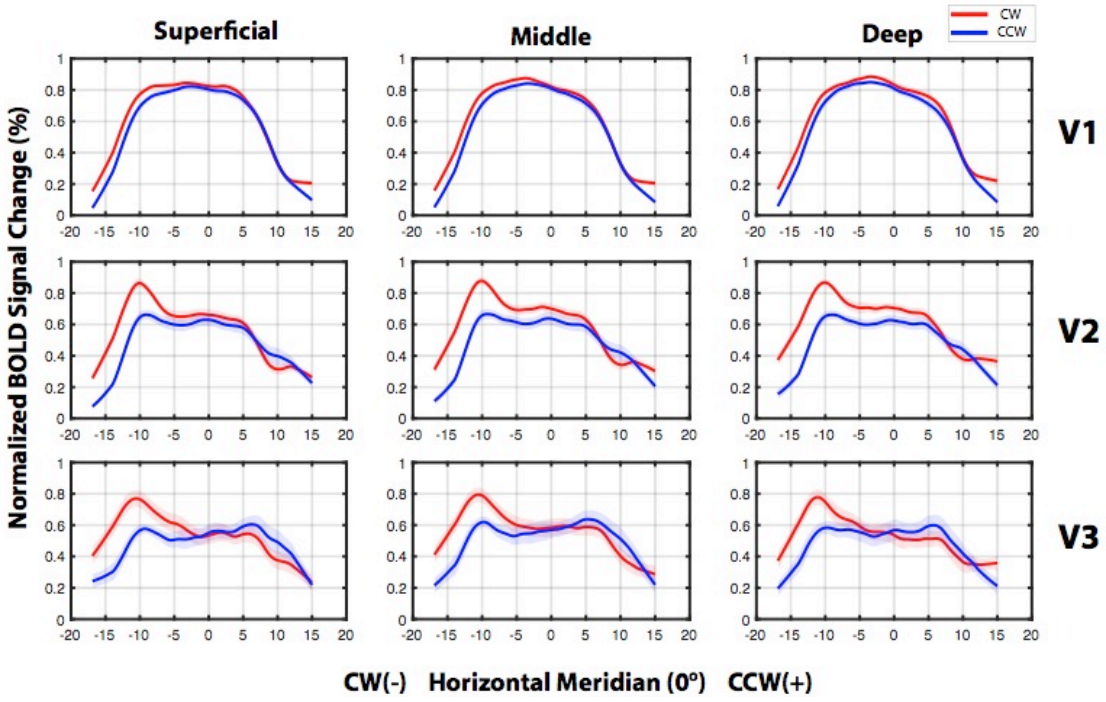

Figure A1.3. Layer-specific bar angle representation of the flash grab illusion in each retinotopic visual area. In layers of V1, V2, V3, fMRI responses to the clockwise and counter-clockwise tilted illusions were plotted as a function of bar angle coordinates across the field of bar rotation. The red and blue curves represent mean retinotopic responses for clockwise and counter-clockwise conditions across seventeen subjects. The shading color indicate between-subject standard error. Source data are provided as a Source Data file.

Appendix 2: Supplemental Information for Chapter 3

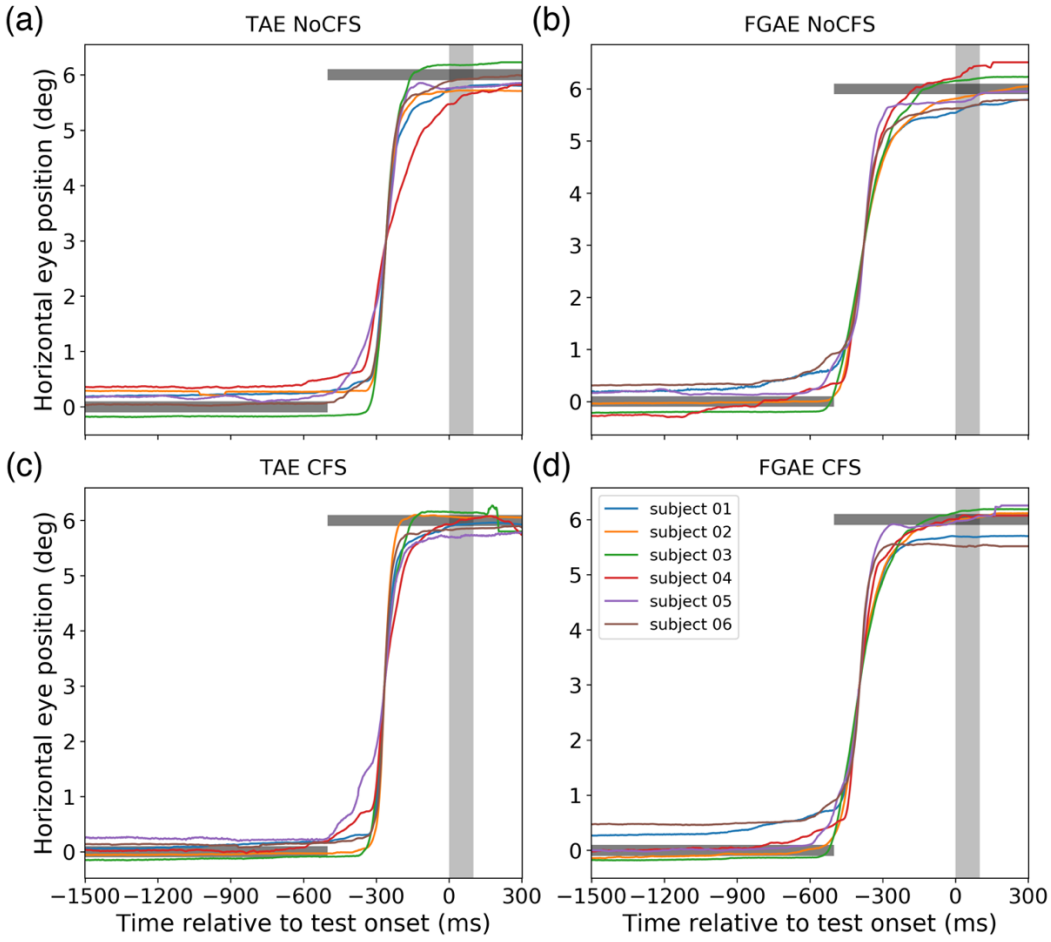

Figure A2.1. Averaged horizontal eye position over the time course of the trials (time relative to the test probe onset (ms)) for six subjects, aligned with the midpoint of the saccade. (a) TAE without CFS condition; (c) TAE with CFS condition; (b) FGAE without CFS condition and (d) FGAE with CFS condition. Different colored curves represent horizontal eye positions for each individual (N=6). Dark gray horizontal bars represent the positions of the fixation point (at 0 degree) and the saccade target (at 6 degree). Light gray vertical bar represents the time course of the test presentation (0-100 ms).

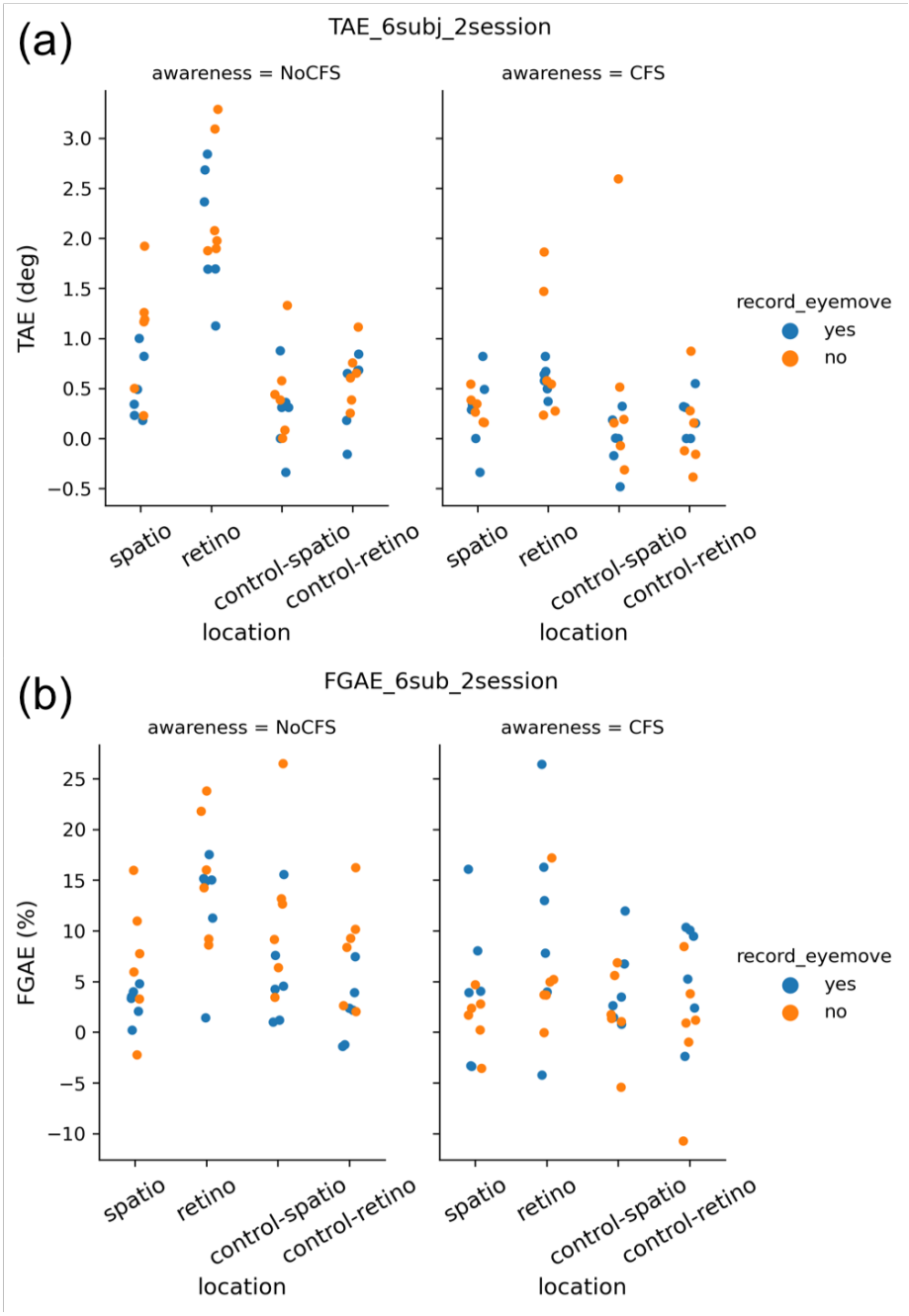

Figure A2.2. Scatter plots for two sessions results with and without eye movement recording for six participants (a, TAE; b, FGAE). The dependent sample t-tests showed that there were no significant differences between two sessions in each condition for both TAE and FGAE ($p>0.05$, Holm corrected).

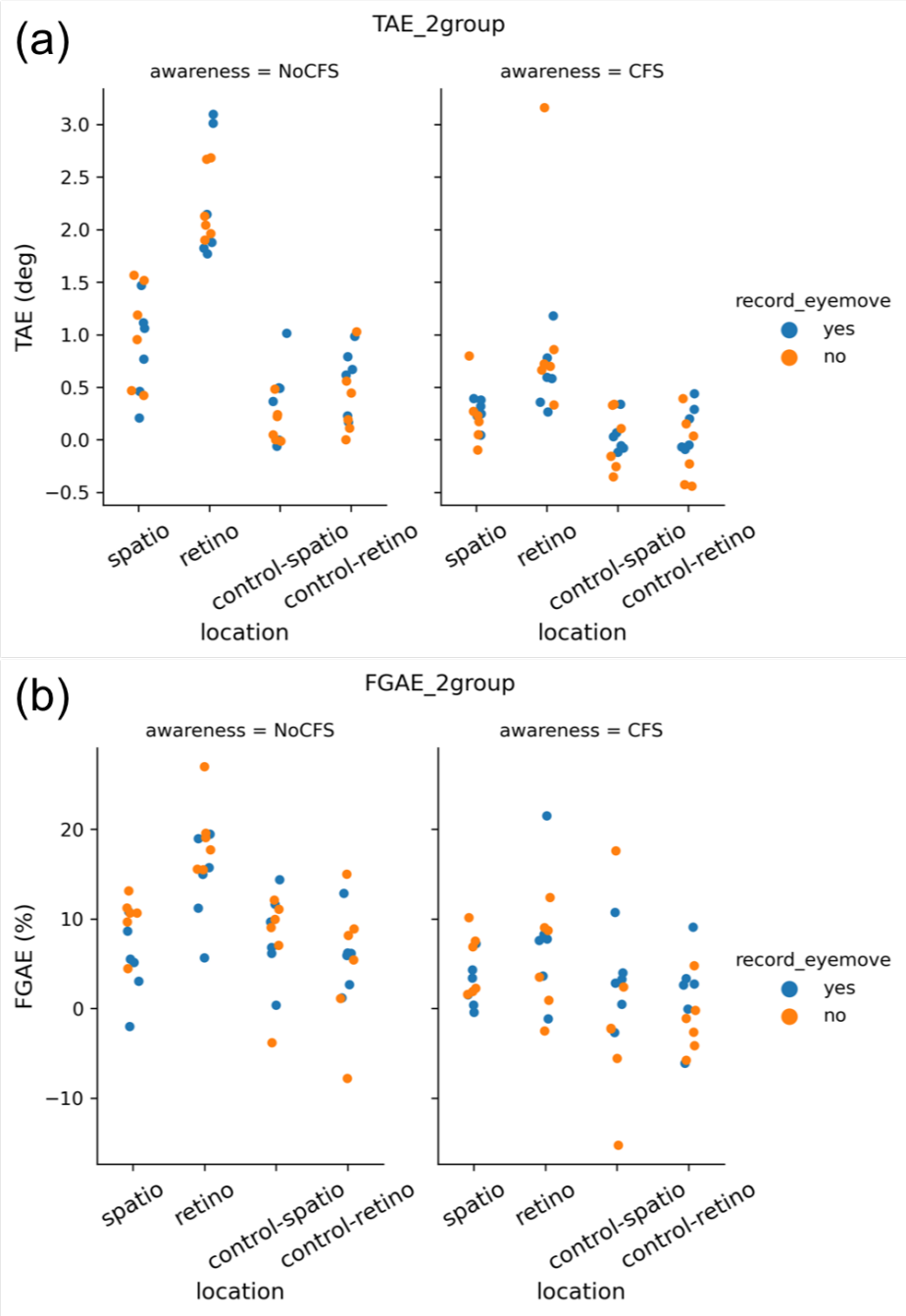

Figure A2.3. Scatter plots for the two groups of participants (n=12) with and without eye movement recording (a, TAE; b, FGAE). The independent sample t-tests showed that there were no significant differences between two groups in each condition for both TAE and FGAE ($p>0.05$, Holm corrected).

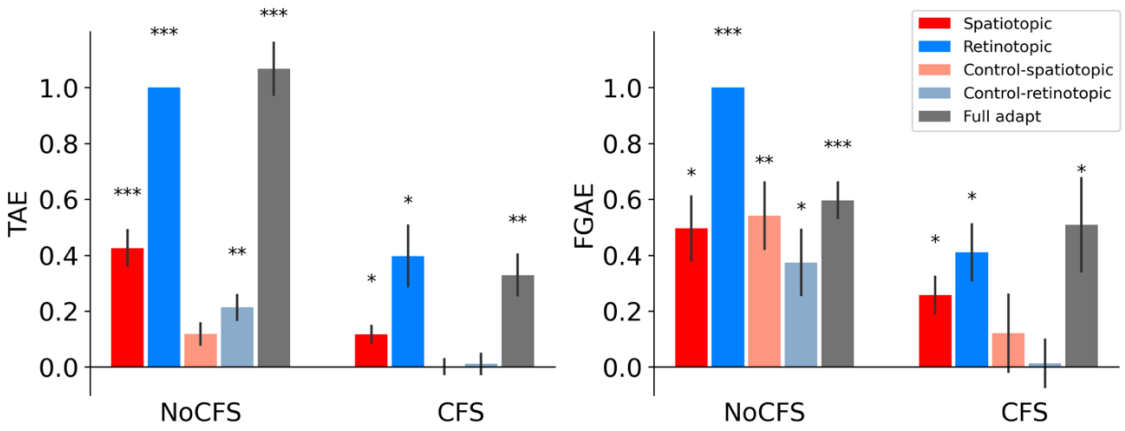

Figure A2.4. Normalized adaptation aftereffects (a, TAE; b, FGAE) for different conditions. The data were normalized against the NoCFS retinotopic condition in TAE and FGAE for each participant (dividing the aftereffect value by that in the NoCFS retinotopic condition). Following the normalization, the pattern of results is similar to that of the main results (Figure 3). Error bars show ±1 SE of the mean. Multiple comparisons were Holm corrected. (* adjusted $p<0.05$; ** adjusted $p<0.01$; *** adjusted $p<0.001$).

Appendix 3: Supplemental Information for Chapter 4

Abbreviation	Full roi name	Num of voxels	Total Num of voxels	Voxel percent in the roi (%)
L lateraloccipital	Left lateral occipital cortex	644	6379	10.09562627
R lateraloccipital	Right lateral occipital cortex	436	5963	7.311755828
L lingual	Right lingual gyrus	61	4205	1.450653983
R fusiform	Right fusiform gyrus	119	4661	2.553100193
R parahippocampal	Right parahippocampal gyrus	134	1742	7.692307692
L inferiorparietal	Left inferior parietal cortex	81	7871	1.029094143
R inferiorparietal	Right inferior parietal cortex	176	9676	1.818933444
L superiorparietal	Left superior parietal cortex	123	10456	1.176358072
R superiorparietal	Right superior parietal cortex	84	10222	0.821756995

Table A3.1. List of ROI activation for viewpoint RDM

Abbreviation	Full roi name	Num of voxels	Total Num of voxels	Voxel percent in the roi (%)
L lateraloccipital	Left lateral occipital cortex	1863	6379	29.20520458
R lateraloccipital	Right lateral occipital cortex	1794	5963	30.08552742
L lingual	Left lingual gyrus	419	4205	9.964328181
R lingual	Right lingual gyrus	304	3894	7.806882383
L parahippocampal	Left parahippocampal gyrus	324	1838	17.62785637
R parahippocampal	Right parahippocampal gyrus	395	1742	22.67508611
L fusiform	Left fusiform gyrus	1114	4714	23.63173526
R fusiform	Right fusiform gyrus	1838	4661	39.43359794
L inferiortemporal	Left inferior temporal gyrus	330	4415	7.474518686
R inferiortemporal	Right inferior temporal gyrus	792	4198	18.86612673
L middletemporal	Left middle temporal gyrus	160	4452	3.593890386
R middletemporal	Right middle temporal gyrus	432	5057	8.542614198
L superiortemporal	Left superior temporal gyrus	81	7271	1.114014578
R bankssts	Right banks of the superior temporal sulcus	148	2196	6.739526412
L supramarginal	Left supramarginal gyrus	361	8600	4.197674419
R supramarginal	Right supramarginal gyrus	227	8150	2.785276074
L inferiorparietal	Left inferior parietal cortex	1223	7871	15.53805107
R inferiorparietal	Right inferior parietal cortex	912	9676	9.425382389
L superiorparietal	Left superior parietal cortex	698	10456	6.675592961
R superiorparietal	Right superior parietal cortex	748	10222	7.317550382
R precuneus	Right precuneus cortex	64	7975	0.802507837

Table A3.2. List of ROI activation for 3D viewpoint-independent pose RDM

Abbreviation	Full roi name	Num of voxels	Total Num of voxels	Voxel percent in the roi (%)
L pericalcarine	Left pericalcarine cortex	101	1912	5.282426778
L lateraloccipital	Left lateral occipital cortex	2095	6379	32.84213827
R lateraloccipital	Right lateral occipital cortex	1425	5963	23.8973671
L lingual	Left lingual gyrus	836	4205	19.88109394
R lingual	Right lingual gyrus	574	3894	14.74062661
L parahippocampal	Left parahippocampal gyrus	293	1838	15.94124048
R parahippocampal	Right parahippocampal gyrus	620	1742	35.5912744
L fusiform	Left fusiform gyrus	437	4714	9.270258804
R fusiform	Right fusiform gyrus	1028	4661	22.05535293
L inferiortemporal	Left inferior temporal gyrus	82	4415	1.857304643
R inferiortemporal	Right inferior temporal gyrus	746	4198	17.77036684
R middletemporal	Right middle temporal gyrus	505	5057	9.986157801
L supramarginal	Left supramarginal gyrus	271	8600	3.151162791
R supramarginal	Right supramarginal gyrus	101	8150	1.239263804
L inferiorparietal	Left inferior parietal cortex	1128	7871	14.33108881
R inferiorparietal	Right inferior parietal cortex	1376	9676	14.22075238
L superiorparietal	Left superiorparietal	800	10456	7.651109411
R superiorparietal	Right superior parietal cortex	1586	10222	15.51555469
L precuneus	Left precuneus cortex	206	7308	2.818828681
R precuneus	Right precuneus cortex	515	7975	6.457680251
L postcentral	Left postcentral gyrus	54	9519	0.56728648
L paracentral	Left paracentral gyrus	132	3294	4.007285974
L precentral	Left precentral gyrus	159	10740	1.480446927
L caudalmiddlefrontal	Left caudal middle frontal gyrus	69	3736	1.846895075
L superiorfrontal	Left superior frontal gyrus	85	12179	0.697922654
L isthmuscingulate	Left isthmus cingulate cortex	60	2531	2.370604504
R isthmuscingulate	Right isthmus cingulate cortex	199	2388	8.333333333
L posteriorcingulate	Left posterior cingulate cortex	145	3266	4.439681568

Table A3.3. List of ROI activation for 3D viewpoint-dependent pose RDM

Abbreviation	Full roi name	Num of voxels	Total Num of voxels	Voxel percent in the roi (%)
L pericalcarine	Left pericalcarine cortex	91	1912	4.759414226
L lateraloccipital	Left lateral occipital cortex	2093	6379	32.81078539
R lateraloccipital	Right lateral occipital cortex	1505	5963	25.23897367
L lingual	Left lingual gyrus	485	4205	11.53388823
R lingual	Right lingual gyrus	468	3894	12.01848998
L parahippocampal	Left parahippocampal gyrus	204	1838	11.09902067
R parahippocampal	Right parahippocampal gyrus	392	1742	22.50287026
L fusiform	Left fusiform gyrus	545	4714	11.56130675
R fusiform	Right fusiform gyrus	1138	4661	24.41536151
L inferiortemporal	Left inferior temporal gyrus	123	4415	2.785956965
R inferiortemporal	Right inferior temporal gyrus	857	4198	20.41448309
L middletemporal	Left middle temporal gyrus	61	4452	1.37017071
R middletemporal	Right middle temporal gyrus	571	5057	11.29127941
L supramarginal	Left supramarginal gyrus	318	8600	3.697674419
R supramarginal	Right supramarginal gyrus	102	8150	1.251533742
L inferiorparietal	Left inferior parietal cortex	995	7871	12.64134163
R inferiorparietal	Right inferior parietal cortex	1363	9676	14.08639934
L superiorparietal	Left superior parietal cortex	931	10456	8.903978577
R superiorparietal	Right superior parietal cortex	1590	10222	15.55468597
L precuneus	Left precuneus cortex	232	7308	3.174603175
R precuneus	Right precuneus cortex	642	7975	8.05015674
L postcentral	Left postcentral gyrus	60	9519	0.630318311
L precentral	Left precentral gyrus	104	10740	0.968342644
R precentral	Right precentral gyrus	64	10705	0.597851471
L paracentral	Left paracentral lobule	139	3294	4.219793564
R paracentral	Right paracentral lobule	128	3831	3.341164187
L superiorfrontal	Left superior frontal gyrus	130	12179	1.067411117
L caudalmiddlefrontal	Left caudal middle frontal gyrus	90	3736	2.408993576
R isthmuscingulate	Right isthmus cingulate cortex	184	2388	7.70519263
L posteriorcingulate	Left posterior cingulate cortex	195	3266	5.970606246

Table A3.4. List of ROI activation for 2D viewpoint-dependent pose RDM

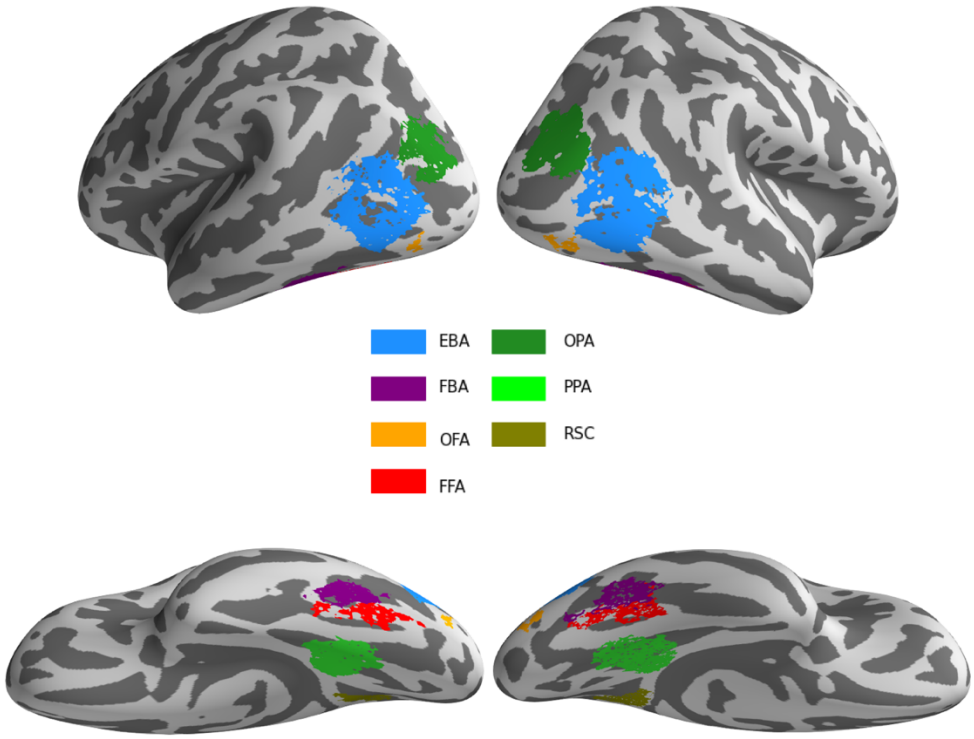

Figure A3.1. Group-level functional localizer results for body-, face-, and place- selective areas across all the subjects (color map threshold is 62.5% (five out of eight subjects)). The body-selective areas include EBA and FBA. The face-selective areas include FFA and OFA. The place selective areas include PPA, OPA, and RSC.