Retinotopic encoding of the Ternus-Pikler display reflected in the early visual areas

Evelina Thunell

Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Wietske van der Zwaag

CIBM, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Haluk Öğmen

Department of Electrical and Computer Engineering, Center for Neuro-Engineering and Cognitive Science, University of Houston, Houston, TX, USA

Gijs Plomp

Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
Département de Psychologie, Université de Fribourg, Fribourg, Switzerland

Michael H. Herzog

Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

The visual representation of the world is often assumed to be retinotopic, and many visual brain areas are indeed organized retinotopically. Visual perception, however, is not based on a reference frame anchored in retinotopic coordinates. For example, when an object moves, motion of its constituent parts is perceived relative to the object rather than in retinotopic coordinates. The moving object thus serves as a nonretinotopic reference system for computing the properties of its parts. It is largely unknown how the brain accomplishes this feat. Here, we used the Ternus-Pikler display to pit retinotopic processing in a stationary reference system against nonretinotopic processing in a moving one. Using 7T fMRI, we found that the average blood-oxygen-level dependent activations in V1, V2, and V3 reflected the retinotopic properties, but not the nonretinotopic percepts, of the Ternus-Pikler display. In the human motion processing complex (hMT+), activations were compatible with both retinotopic and nonretinotopic encoding. Thus, hMT+ may be the first visual area encoding the nonretinotopic percepts of the Ternus-Pikler display.

Introduction

Vision starts with a retinotopic representation of the visual scene: Neighboring points in the scene are mapped onto neighboring neurons in the retina, and this retinotopic encoding prevails in many visual brain areas (Amano, Wandell, & Dumoulin, 2009; Gardner, Merriam, Movshon, & Heeger, 2008; Sereno et al., 1995). Our perception of the world, on the other hand, is clearly not anchored to retinotopic coordinates. For example, the image projected on the retina is shifted with every saccadic eye movement, causing a shift of the representation according to retinotopic coordinates. Yet these shifts are not perceived; our perception of the world remains stable across saccades. The key mechanism mediating perceptual stability across saccades is thought to be an efference copy of the motor command, which allows us to foresee and discount the displacements of the retinal image (Sperry, 1950; von Holst & Mittelstaedt, 1973). The transformation of the retinotopic representation by use of efference copy signals results in a representation based on a non-

retinotopic (more specifically, in this case, a spatio-topic) reference frame. Another aspect of vision relying on nonretinotopic processing is motion perception. For example, motion is often perceived relative to moving reference systems (Duncker, 1929; Johansson, 1950; for review, see Öğmen & Herzog, 2015). One example comes from gait perception: We perceive the limb movements of a walking person relative to his or her body. The body serves as a reference system for computing the relative motions, making the "true" retinotopic trajectories of the arms and legs invisible. The overall translational motion of the body is discounted from the motion of the limbs, similarly to the discounting of the retinal shifts caused by saccades, giving rise to a percept of pendular arm and leg motions. However, an important difference is that efference copies cannot aid the discounting of the motion. Perception of motion relative to moving reference systems has been studied extensively psycho-physically and modeled, for example, by perceptual vector analysis (see, for example, Johansson, 1973). However, almost nothing is known about the under-lying neural correlates.

Here, we adopted the Ternus-Pikler display to study the neural correlates of nonretinotopic, relative motion perception. This paradigm is a versatile tool for studying nonretinotopic processing without eye movements and has previously been used to show that form integration (Öğmen, Otto, & Herzog, 2006), visual search, and motion integration (Boi, Öğmen, Krummenacher, Otto, & Herzog, 2009) are nonretinotopic processes.

Materials and methods

Participants

Thirteen paid participants (six male) aged 18–34 years (mean 24 years) took part in the experiment after giving written informed consent. Data from one (female) participant were excluded due to excessive sleepiness, motion artifacts, and very low behavioral performance. All participants had normal or corrected-to-normal vision with an acuity of ≥ 1 in at least one eye as assessed with the Freiburg visual acuity test (Bach, 1996), and all reported being right-handed. All procedures were in accordance with the Declaration of Helsinki and approved by the local ethics committee.

Stimuli and task

In the No motion conditions, two checkerboards flickered on and off on the screen, interleaved by blank interstimulus intervals (ISIs; Figure 1a). The checkerboards were either identical in each image frame (No motion/Same) or changed contrast polarity from frame to frame (No motion/Alternating) resulting in a percept of vertical apparent motion. Observers perceived the veridical, retinotopic properties of the checkerboards (staying the same or alternating). When flanking checkerboards were added to the stimuli (Group motion conditions), three checkerboards were perceived to move back and forth as a group. This apparent motion created a nonretinotopic reference system in which the checkerboards in different image frames were integrated across spatial locations. Because of this nonretinotopic correspondence, checkerboards that alternated in contrast polarity from frame to frame (Group motion/Alternating) were perceived to stay the same, and checkerboards that were the same across frames (Group motion/Same) appeared to alternate. The veridical properties were no longer perceived. These four conditions made up a 2×2 factorial design with factors Reference system (No motion and Group motion) and Polarity (Same and Alternating). The data were analyzed using repeated-measures two-way AN-OVAs with participants as a random factor. In the Control conditions (Figure 1b), the checkerboards were the same across image frames, and neighboring checkerboards had aligned contrast polarity. The checkerboards were therefore perceived to stay the same across frames in both the No motion/Control and the Group motion/Control conditions.

A small red fixation dot was presented above the center of the stimuli. At random intervals, the dot was briefly displaced slightly up- or downward, and the participants reported this by pushing the right or left button, respectively, on an MR-compatible handheld device (Current Designs, Philadelphia, PA). The participants also had to detect targets on the checker-boards: At random intervals, a red dot appeared on one of the central checkerboards (in position 2 or 3 in Figure 1). The participants pushed the right or left button to indicate whether the dot was on the upper or lower half of the checkerboard, respectively (while constantly fixating their gaze on the fixation dot). The task on the fixation dot had to be performed even when no checkerboard stimulus was present. Targets appeared every 3 to 7 s. During stimulus presentation, the target appeared on a checkerboard in 75% and on the fixation dot in 25% of the cases. Half of the participants used their left hand and the other half their right hand for pushing the buttons. No feedback was given during data acquisition. Instead, the performance was displayed on the screen after each run and also communicated verbally via the scanner intercom. The participants were familiarized with the stimuli and the task on a separate occasion prior to the MR scanning session. They were asked to describe the stimuli, and all
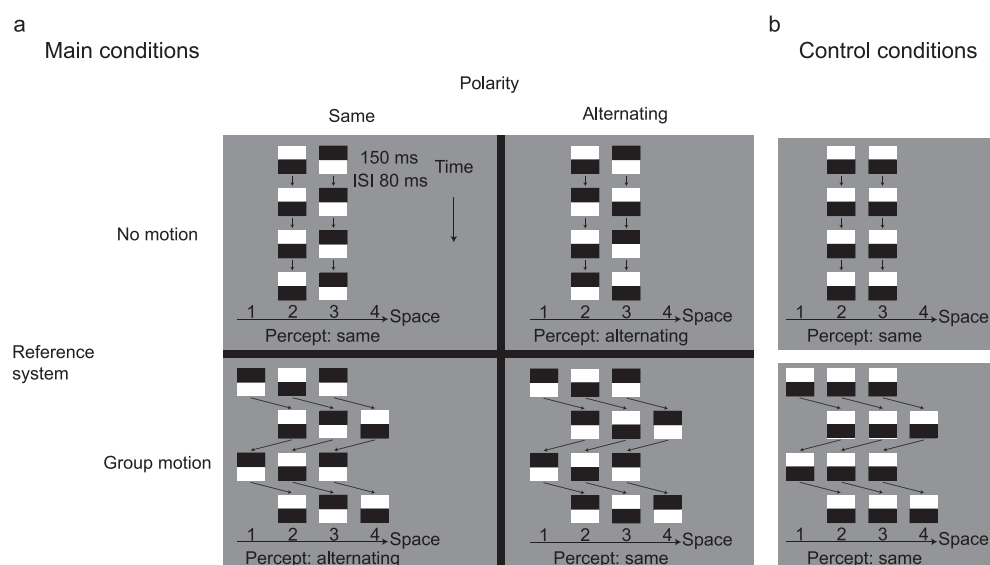
Figure 1. Ternus-Pikler stimuli (see also Movies 1–6). (a) Main conditions. In the No motion conditions (upper panel), two checkerboards with opposite contrast polarities are presented for 150 ms in positions 2 and 3, interleaved by blank ISIs of 80 ms. In the No motion/Same condition, the checkerboards are the same in each image frame. In the No motion/Alternating condition, the upper and lower halves of the checkerboards alternate from frame to frame. In both conditions, the checkerboards are integrated retinotopically across frames as indicated by the arrows (the arrows were not shown in the actual stimuli). When flanking elements are added in positions 1 and 4 (Group motion conditions; lower panel), horizontal group motion is perceived, causing the checkerboards in different image frames to be grouped across space according to the arrows. The checkerboards are integrated nonretinotopically, and the percepts are inverted with respect to the No motion conditions: Even though the checkerboards in positions 2 and 3 have the same contrast polarity in all frames in the Group motion/Same condition, the upper and lower halves are perceived to alternate. Similarly, the contrast polarity alternates across frames in the Group motion/Alternating condition in positions 2 and 3, yet it is perceived to be the same. (b) Control conditions. Neighboring checkerboards have the same contrast polarity and are the same in each image frame. They are perceived to be the same across frames in both the No motion/Control and the Group motion/Control conditions. The actual stimuli were light and dark gray checkerboards on a medium gray background.

reported the expected checkerboard percepts in all conditions as described in Figure 1. This is in accordance with previous work showing that for the timing used here, horizontal group motion is typically perceived in the Group motion conditions (Breitmeyer & Ritter, 1986a, 1986b; Pantle & Picciano, 1976).

The functional data for the Ternus-Pikler experiment were acquired in two runs of 296 volumes (12 min 20 s) each. We used a blocked design with stimulus presentations of 20 s interleaved by 20-s pauses during which only the fixation dot was present. Each run started and ended with a 20-s pause. A nonoverlapping set of random permutations of the six conditions was used for the presentation order. Three such permutations were concatenated in each run, resulting in a total of six presentations per condition and participant. The stimuli were randomized for polarity of the checkerboards in the first frame (dark gray upper half and light gray lower half or vice versa) and start position in the Group motion conditions (left: positions 1-2-3 or right: positions 2-3-4). The diameter of each checkerboard was 2.6° and the center-to-center spacing 3.9°. The edges of the checkerboards were blurred (0.2°). The dark and light gray levels of the checkerboards were

10% Michelson contrast apart, centered on the background luminance (dark 36 cd/m^2, light 44 cd/m^2, background 40 cd/m^2). These luminance values were chosen to give maximum sensitivity in V1/V2 based on contrast sensitivity curves estimated in a separate experiment with two participants who did not participate in the main experiment (Figure 2). The human motion processing complex (hMT+) was not localized in the contrast experiment. There was a clear percept of same or alternating checkerboards even with this low luminance contrast. A fixation disk was presented 3.9° above the vertical center of the stimuli. The disk had a diameter of 30′ and a luminance of 63 cd/m^2, and in the middle, there was a small red fixation dot with a diameter of 6′. The target dots appearing on the checkerboards were also red and had a diameter of 6′. Both the fixation dot and the checkerboard target dots had about the same luminance as the background (39.6 cd/m^2). The fixation target consisted of a brief 3′ up- or downward displacement of the red dot. The task was designed to assure that the participants kept fixation but at the same time attended to the stimuli. The performance on the task is reported for the Ternus-Pikler runs. The performance in the functional localizer
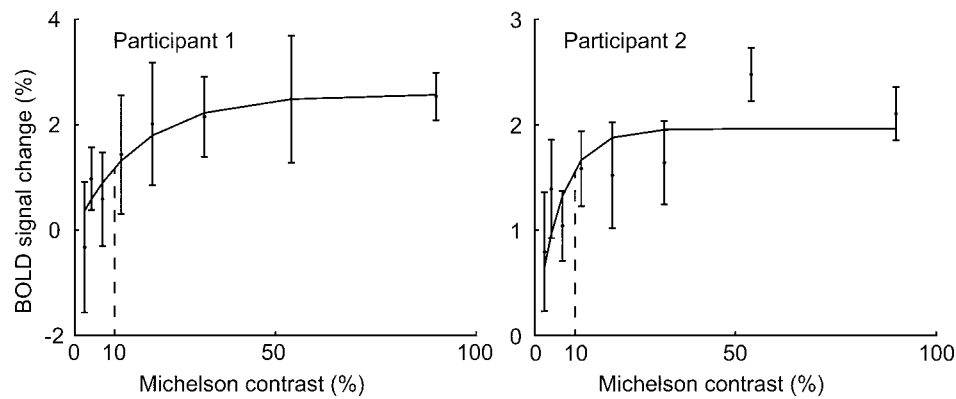
Figure 2. Contrast sensitivity measurements in the V1/V2 representations of positions 2 and 3 of the Ternus-Pikler display for two participants with fitted curves (solid lines). The stimuli were alternating checkerboards flickering at the same frequency as in the main experiment. The contrast level used for the main experiments (10%) is at a steep part of the curve for both participants. (The ROIs were larger than in the main experiment, hence the different absolute activation levels.)

runs (see below) is not reported. To further control for fixation, eye movements were monitored online by video during the session for 10 of the participants. The right eye was tracked at a frequency of 50 Hz with an MR-compatible eye tracker positioned at the rear of the scanner (SMI, Teltow, Germany; data not shown).

Functional localizers

The retinotopic representations of positions 1–4 in the Ternus-Pikler stimuli (Figure 1) were localized in the primary visual cortex (V1), V2, and V3 by presenting a black-and-white alternating checkerboard for 15 s at one of the four positions at a time. The same parameters as for the Ternus-Pikler experiment were used except that the checkerboards had sharp edges, the background was slightly lighter (45 cd/m^2), and the red checkerboard target had a diameter of 30′ and the same luminance as the background. The four positions were stimulated without interleaving pauses in such a way that neighboring checkerboards never appeared consecutively (i.e., the order was 2-4-1-3 or 3-1-4-2). When all positions had been stimulated once, a 15-s pause followed. The sequence was repeated five times, and the run started and ended with a 15-s pause, resulting in 156 volumes (6 min 30 s).

We used a standard hMT+ localizer paradigm (Saenz, Lewis, Huth, Fine, & Koch, 2008; Tootell et al., 1995), in which 15-s periods of static dots were interleaved with 15-s periods of radial in- or outward dot motion. This cycle was repeated eight times, resulting in 96 volumes (4 min). There was no task except for passive fixation on a dot in the middle of the screen. Meridian mapping was performed using 45°-wide radial checkerboard wedges flickering at 8 Hz, extending from 30′ away from fixation to the edge of the screen. In order to get reliable activation despite the

limited height of the screen, the lower and upper halves of the vertical meridian were stimulated separately and the fixation dot was put at the top or bottom of the screen, respectively. Twenty seconds of stimulation was interleaved with 5-s pauses to allow for fixation to stabilize. The three meridian mapping conditions (horizontal meridian, upper meridian, lower meridian) were each presented nine times, interleaved, resulting in 272 volumes (11 min 20 s). During the meridian mapping, participants indicated up- or downward displacements of 6′ of the fixation dot by means of the push buttons. As in the Ternus-Pikler runs, we tracked the right eye during all the localizer runs for 10 participants. The identified V1, V2, V3, and hMT+ regions in the left hemisphere of one participant are shown in Figure 3.

MRI data acquisition

All data were acquired on a short-bore 7T MR system (Siemens, Munich, Germany) with a head-gradient insert. An eight-channel RF coil (Rapid Biomedical GMBH, Germany) was used for RF transmission and reception (Salomon, Darulova, Narsude, & van der Zwaag, 2014). The stimuli were projected on a screen placed inside the scanner bore at the service end by a projector placed outside the shielded scanner room. Participants viewed the screen via a mirror placed inside the open-ended RF coil. fMRI data were acquired using an echo planar imaging (EPI) sequence with sinusoidal readout. The 56-mm-thick imaging slab was positioned coronally oblique with the phase encoding direction along the head–foot axis (repetition time [TR] = 2500 ms, echo time [TE] = 27 ms, flip angle = 63° [SAR limited], bandwidth/pixel = 1628 Hz/px, phase partial Fourier = 6/8). We used a matrix size of 128 × 128 × 34 voxels and a field of view
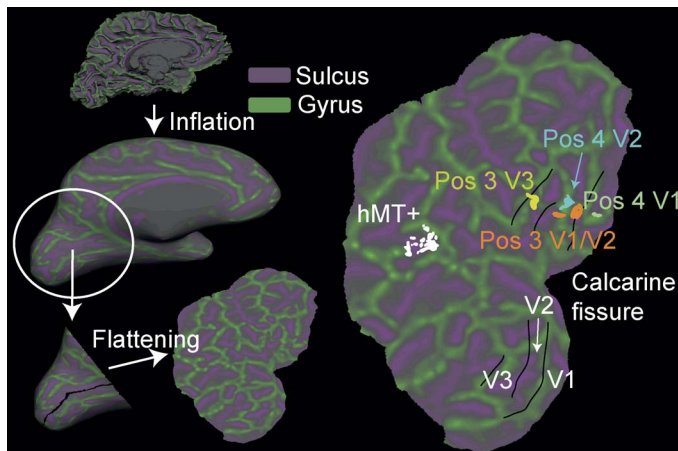
Figure 3. Left hemisphere of one participant. We applied standard inflation and flattening techniques (left) to obtain a flat representation of the occipitotemporal cortex (right). In a first step of the ROI analysis, we localized hMT+ and the cortical representations of positions 1–4 of the Ternus-Pikler display (Figure 1) in the early visual areas V1, V2, and V3. We first targeted the ROIs corresponding to positions 1–4 in V1/V2. In a second analysis, we used smaller ROIs comprising only positions 2 and 3 in both V1/V2 and V3. Note that the cortical representations of positions 1 and 2 are in the right hemisphere and thus not shown here. The V1 and V2 representations of position 3 (marked in orange) were indistinguishable in the three-dimensional data. The V1 representation of position 4 is marked by light green, and its V2 representation is shown in blue. The V3 representation of position 3 is marked in yellow. Left hMT+ is marked in white. The number of ROIs that could be identified varied between participants and hemispheres (24 out of 24 hemispheres for both the large and the small V1/V2 ROIs, 15 for the small V3 ROIs, and 21 for hMT+).

of 180×180 mm^2, leading to a spatial resolution of 1.4 $\times$ 1.4 mm^2 in-plane with a slice thickness of 1.5 mm and a 10% slice gap. No parallel imaging acceleration was applied. For the meridian mapping, a slice thickness of 1.4 mm was instead used, leading to a 52-mm-thick imaging slab. A single whole-head EPI volume with 128 slices and TR = 12 s but otherwise identical parameters to the functional data (slice thickness 1.5 mm) was acquired to aid coregistration of the anatomical to the functional data. Whole-brain T1- and T2*-weighted anatomical data were acquired using the MP2RAGE sequence (Marques et al., 2010) with $256 \times 256 \times 176$ voxels and an isotropic resolution of 1.0 mm (TR = 5500 ms, TE 2.84 ms, inversion time [TI1] = 750 ms, TI2 = 2350 ms). Phase partial Fourier = 6/8 and a generalized autocalibrating partially parallel acquisitions acceleration factor of two were used to speed up image acquisition.

MRI data processing

Standard preprocessing was carried out separately for each participant using SPM8 (http://www.fil.ion. ucl.ac.uk/spm/), including slice timing correction, motion correction, and coregistration of the anatomical to the functional data. The functional localizer data were smoothed with a 1.5 * voxel size full width at half maximum Gaussian kernel. No normalization or intersubject coregistration was performed.

Inflation and flattening of the anatomical images was carried out using the Freesurfer image analysis suite, which is documented and freely available for download online (http://surfer.nmr.mgh.harvard.edu). The skull-stripping step was done manually. The activation maps from the meridian mapping were overlaid on the flattened occipital patches in order to delineate the V1/V2 borders (the upper or lower vertical meridian contrasted against the horizontal meridian) as well as the V2/V3 borders (the horizontal meridian contrasted against the upper and lower vertical meridians). The activations from the position localizer scan (each position in the Ternus-Pikler display contrasted against its within-hemisphere neighbor) were overlaid on the same flattened patch in order to identify the cortical representations of positions 1–4. The regions of interest (ROIs) were then identified in the three-dimensional data and extracted using the MarsBaR toolbox (http://marsbar.sourceforge.net). Thresholds were chosen individually to ensure consistent ROI sizes (t values between two and 11 were used).

hMT+ was localized bilaterally (coherent contrasted against random dot motion) with a size of 443 ± 118 mm^3 (mean $\pm$ *SD*; t values between two and seven). The MNI coordinates were $(-44.4, -77.4, 2.0) \pm (3.1, 7.0, 8.7)$ for the left hMT+ and $(43.9, -69.8, 2.0) \pm (4.7, 7.5, 3.6)$ for the right hMT+. This is consistent with previously reported sizes and coordinates of hMT+ (see, for example, d'Avossa et al., 2007; Kolster, Peeters, & Orban, 2010; Tootell et al., 1995).

The size of the V1/V2 ROI covering positions 1–4 was 914 ± 217 mm^3 (mean $\pm$ *SD*). For the smaller ROI covering only positions 2 and 3, the size was 519 ± 115 mm^3 for V1/V2 and 154 ± 81 mm^3 for V3. None of the cortical representations of the central positions (2 and 3) were overlapping with or bordering on the representations of the outer positions (1 and 4) in the three-dimensional data. The number of ROIs that could be identified varied between participants and hemispheres.

The voxel-average blood-oxygen-level dependent (BOLD) signal time series were extracted for each ROI. Subsequently, linear drift was removed, and the average signal changes for each condition and participant were computed using Matlab. The signal change for each stimulus presentation was calculated as the

average BOLD signal from 7.5 s after stimulus onset, to account for the lag of the hemodynamic response, until the offset of the stimulus. These values were then compared with the average baseline level in the run (from 10 s after stimulus offset to stimulus onset) to obtain relative signal changes. For hemispheres in which separate V1 and V2 representations were found, the signal was averaged over these two regions. In the analysis of the ROIs covering screen positions 1–4, only hemispheres in which we could identify representations of both the central position (2 or 3) and the outer position (1 or 4) were taken into account. The signal was averaged per hemisphere with equal weight to the representations of the central and outer positions. The left and right ROIs were analyzed together because the stimulation did not differ between the two hemifields and the activation patterns were similar.

Results

Behavior

The participants fixated on a dot above the Ternus-Pikler display, which was displaced slightly up or down at random time points. The task was to report these displacements as well as dots appearing in the lower or upper half of one of the central checkerboards (in positions 2 or 3, see Figure 1). The participants correctly discriminated 90.9% ± 9.3% (mean ± *SD*) of the fixation dot displacements and 87.1% ± 8.6% of the dots on the checkerboards. There were only 1.3 ± 0.96 false positive responses. The high performance on both tasks confirms that the participants kept fixation while attending to the stimuli. Online monitoring of the right eye (of 10 of the 12 participants) showed no apparent systematic eye movements, such as tracking of the stimuli in the Group motion conditions (the quality of the recorded eye-movement data was too poor for quantitative analysis). In fact, tracking of these stimuli would require more than four saccades per second, which is not a comfortable rate. Observers have previously been shown to be able to fixate in similar Ternus-Pikler paradigms even without a fixation dot (Boi et al., 2009; Noory, Herzog, & Öğmen, 2015).

fMRI: Large V1/V2 ROIs

In the Ternus-Pikler display, the checkerboards are perceived differently depending on how they are grouped across time, i.e., depending on the *reference system* (Figure 1). The checkerboards integrate retinotopically and are therefore "veridically" perceived in the No motion conditions (stationary reference sys-

tem). In the Group motion conditions (moving reference system), the percepts are *opposite* to the veridical properties because of the nonretinotopic integration: Checkerboards that alternate in contrast polarity across images are perceived to stay the same, and checkerboards that stay the same are perceived to alternate. The retinotopic properties are in this case invisible. This dissociation between the retinotopic stimulation and the subjective percepts allows for testing to what extent a given brain region encodes each of these two aspects.

We first targeted the cortical representation of screen positions 1–4 in V1 and V2. In these areas as well as in V3, which we will consider later, BOLD responses are known to increase with temporal luminance contrast (Avidan et al., 2002; Boynton, Demb, Glover, & Heeger, 1999; Boynton, Engel, Glover, & Heeger, 1996; Buracas & Boynton, 2007; Tootell et al., 1995; Yan et al., 2014). From a retinotopic encoding, we would therefore expect higher BOLD activations in the Alternating than in the Same conditions because the temporal luminance range in positions 2 and 3 is larger in the former: In the Same conditions, the luminance changes from dark gray in one frame to medium gray during the ISI and then back to dark gray in the following frame in one half of the checkerboard (and *light* gray to medium gray to *light* gray in the other half), and in the Alternating conditions, the luminance changes from dark gray to medium gray during the ISI and then to *light* gray, etc. (Figure 1). The temporal contrast is thus twice as large in the Alternating conditions compared to the Same conditions. If V1, V2, and V3 encode the subjective percepts rather than the veridical stimulation, we would instead expect an interaction effect in the direction of stronger activations in the conditions in which the checkerboards are *perceived* to alternate (No motion/Alternating and Group motion/Same) than in the conditions in which the checkerboards are *perceived* to be the same (No motion/Same and Group motion/Alternating). If an area encodes partially the retinotopic properties and partially the percepts, we would expect both effects to be present.

In the combined V1/V2 ROIs, we found higher BOLD activations in the Alternating conditions (3.5% on average) than in the Same conditions (3.1% on average), main effect of polarity: $F(1, 23) = 31.7$, $p < 1 \times 10^{-5}$ (Figure 4a), and no interaction effect ($p = 0.66$). Hence, the activations reflect the retinotopic properties but not the percepts. In addition, we found stronger activations in the Group motion (3.9%) than in the No motion conditions (2.8%), main effect of motion: $F(1, 23) = 84.7$, $p < 4 \times 10^{-9}$, as expected considering that there are only two checkerboards in the No motion conditions but three in the Group motion conditions.
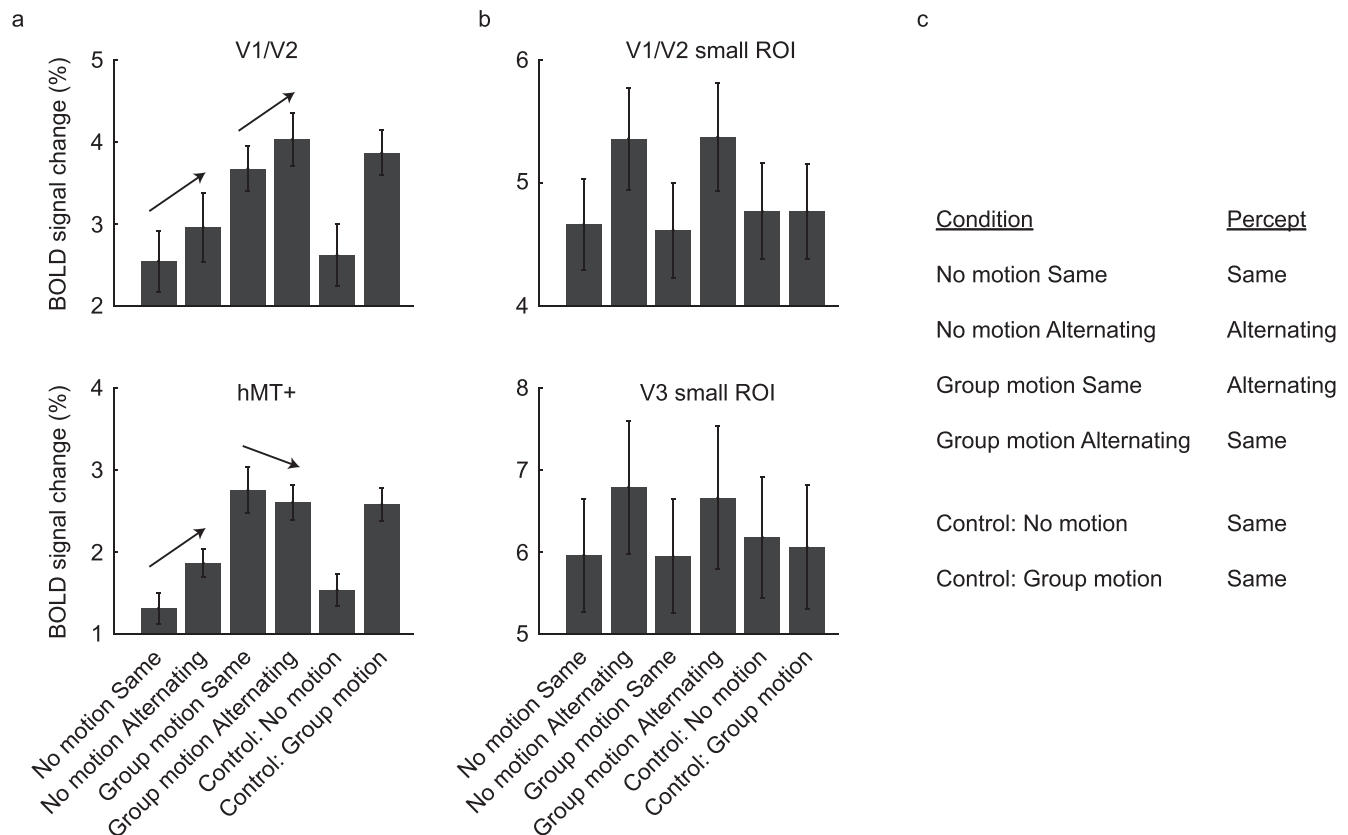
Figure 4. Relative BOLD signal changes compared to baseline (average interstimulus activity) in the main and control conditions. (a) BOLD activity in the V1/V2 ROIs covering the entire Ternus-Pikler display (positions 1–4 in Figure 1) and in hMT+. In both areas, activations are higher in the Group motion than in the No motion conditions. V1/V2 responds more strongly to Alternating than Same checkerboards (arrows), reflecting the larger temporal luminance range in positions 2 and 3 in the Alternating conditions. This main effect of polarity is absent in hMT+, in which instead there is a significant interaction effect (arrows). (b) In a second analysis, we identified V1/V2 and V3 ROIs covering only positions 2 and 3 in the Ternus-Pikler display. In both V1/V2 and V3, activations are higher in the Alternating than in the Same conditions. There is no significant difference between the No motion and Group motion conditions and no interaction effect. The Control conditions show similar activation levels as the No motion/Same and Group motion/Same conditions. Error bars indicate the standard error of the mean (*SEM*) across participants. (c) Checkerboard percepts in all conditions.

The Control conditions elicited similar activation levels as the corresponding Same main conditions, also in accordance with a retinotopic encoding. We used combined V1/V2 ROIs because the proximity of positions 2 and 3 to the vertical meridian (which delineates the border between V1 and V2) made the V1 and V2 representations indistinguishable.

To maximize the sensitivity in V1/V2, the luminance values for the checkerboards were chosen based on contrast response curves for single alternating checkerboards measured for two additional participants (Figure 2). We can therefore be sure that the BOLD responses to the Ternus-Pikler stimuli in these ROIs are not at ceiling level. The overall high percentage of BOLD signal changes in our results are typical for the high field strength (7 T) of the magnet (van der Zwaag et al., 2009).

fMRI: Small V1–V3 ROIs

In a second analysis, we used smaller ROIs covering only positions 2 and 3 of the Ternus-Pikler display (Figure 1). Because the flanking elements in the Group motion conditions appear in positions 1 and 4, this manipulation eliminates the trivial difference between the No motion and Group motion reference systems and thereby gives higher sensitivity for interaction effects. It was possible to localize the small ROIs both in V1/V2 and in V3. In line with retinotopic encoding, in V1/V2 we again found higher activity in the Alternating conditions (5.4% on average) than in the Same conditions (4.6% on average), main effect of polarity: $F(1, 23) = 53.2$, $p < 3 \times 10^{-7}$ (Figure 4b), and no interaction effect ($p = 0.67$). Thus, there is no influence of whether or not the checkerboards are perceived to alternate. We found no main effect of

reference system ($p = 0.86$), assuring that the ROIs are well separated from the cortical representations of screen positions 1 and 4. ("Leakage" from positions 1 and 4 would have resulted in higher activations in the Group motion than in the No motion conditions.) The Control conditions elicited similar activation levels as the Same main conditions, also in accordance with a retinotopic encoding.

In V3, the pattern of BOLD activations is similar to that in V1/V2: Activations are higher in the Alternating (6.7% on average) than in the Same conditions (6.0% on average), main effect of polarity: $F(1, 14) = 14.6$, $p < 2 \times 10^{-3}$ (Figure 4b). There are no other significant effects (main effect of reference system: $p = 0.58$, interaction effect: $p = 0.64$). Again, the responses in the Control conditions are similar to the Same main conditions. Thus, also the V3 activations reflect the retinotopic properties of the stimuli without any significant influence of the percepts.

fMRI: hMT+

Next, we targeted hMT+, which is involved in the processing of real as well as apparent motion (Goebel, Khorram-Sefat, Muckli, Hacker, & Singer, 1998; Liu, Slotnick, & Yantis, 2004; Muckli et al., 2002; Tse, 2006). The same hypotheses as for the early visual areas about veridical, retinotopic encoding versus percept-based, nonretinotopic encoding are valid also for this area, albeit for different reasons. Unlike V1–V3, hMT+ activation for flickering stimuli typically saturates already at low contrast (Tootell et al., 1995), and we therefore expect little response modulation due to the different temporal contrasts in the different conditions. However, because the contrast polarity alternation elicits a percept of vertical apparent motion, hMT+ should still be more sensitive to alternating than same checkerboards. Indeed, a post hoc *t* test showed significantly higher activation in the No motion/ Alternating than in the No motion/Same condition (two-tailed *t* test, $p < 0.01$). As in the early visual areas, we thus expect a purely retinotopic encoding to result in higher BOLD activations in the Alternating than in the Same conditions, and we expect an influence of the percepts to yield an interaction effect in the direction of higher activations in the conditions in which the checkerboards are *perceived* to alternate compared to when they are *perceived* to stay the same across image frames.

Contrary to the results in V1/V2, in hMT+ we found no significant main effect of polarity (Figure 4a, lower graph, $p = 0.12$). Instead, there is a significant interaction effect, $F(1, 20) = 8.9$, $p < 0.01$. There are several possible accounts for these results, including both retinotopic and nonretinotopic ones. For exam-

ple, even though the average BOLD signal change in hMT+ during the localizer run was 6.5% ± 1.9%, i.e., considerably higher than the responses elicited by the Ternus-Pikler stimuli (<3%), we cannot be sure that the signal did not reach ceiling level in the Group motion conditions. A ceiling effect could mask a retinotopic main effect of polarity such as the one found in V1/V2. On the other hand, assuming that the activations are *not* at ceiling, the pattern of hMT+ activations might be the result of an encoding of the percepts. In line with this interpretation, the interaction effect is in the direction of higher BOLD activity in the conditions in which the checkerboards are perceived to alternate than in the conditions in which they are perceived to stay the same. The activations in the Control conditions (Figure 1b) also speak for the percepts being encoded in hMT+. As opposed to the main conditions, here the checkerboard percepts do *not* change between the No motion and Group motion conditions. Therefore, an influence of the percepts should result in the difference in BOLD activation between the Control/Group motion and the Control/ No motion condition lying in between the corresponding main condition differences. Indeed, for the Alternating main conditions, the difference between Group motion and No motion is 0.7 percentage points (pp), for the Same main conditions 1.4 pp, and for the Control conditions 1.0 pp.

As in V1/V2, we found a main effect of reference system in hMT+: the Group motion conditions elicited a higher response (2.7% on average) than the No motion conditions (1.6% on average), $F(1, 20) = 34.5$, $p < 1 \times 10^{-5}$. This was expected because there are only two checkerboards in the No motion conditions but three in the Group motion ones, which, in addition, move back and forth in apparent motion.

Discussion

We found that the average BOLD activations in the stimulated regions of visual areas V1–V3 reflect the retinotopic properties of the checkerboards in the Ternus-Pikler display without any influence of the subjective percepts. The hMT+ activations are compatible with both veridical (retinotopic) and percept-based (nonretinotopic) encoding. Hence, although on the level of V1–V3 the nonretinotopic percepts are not reflected in the average BOLD activations, it remains an open question whether hMT+ is the first visual area where the activations reflect the percepts.

Relative motion perception has been proposed to rely on a two-stage process in which first the reference system is computed, and relative motions are then computed with respect to this system (Clarke, Öǧmen,

& Herzog, 2015; Duncker, 1929; Johansson, 1973; for review, see Öğmen & Herzog, 2015). Recently, Poor-esmaeili, Morrone, Cicchini, and Burr (2012) sparked a discussion about this topic, proposing a one-stage model to account for the percepts in a Ternus-Pikler–type stimulus based on sinusoidal gratings. Detailed analysis showed that this one-stage model might be restricted to these specific simplified Ternus-Pikler stimuli, and it is unclear whether it generalizes to our stimuli (Clarke & Herzog, 2013; but see also Poor-esmaeili, Cicchini, Morrone, & Burr, 2013). However, it cannot be ruled out that a one-stage process could account for nonretinotopic motion processing. In the present study, we merely looked for correlates of the percepts elicited in the Ternus-Pikler display, and the results do not allow us to draw conclusions about whether the underlying mechanisms are based on one- or two-stage processing in general. Our recent EEG study, however, seems to speak for a two-stage process, showing small early effects of reference system followed by large effects of relative motion (Thunell, Plomp, Öğmen, & Herzog, 2015). These results can be interpreted as an early encoding and discounting of the reference system, allowing a reference system–independent encoding of relative motion throughout most of the visual processing.

Could the checkerboard percepts in the different conditions and any BOLD correlates of these percepts occur as a mere result of receptive field (RF) size? In the primate visual system, RFs are smallest in V1 and increase in size with the level in the visual hierarchy. For example, in the macaque brain, RFs are about 10 times larger in MT than in V1 (Albright & Desimone, 1987). Although large RFs are likely necessary for integrating visual information nonretinotopically, the RF size alone does not specify whether the encoding is purely retinotopic or influenced by the subjective percepts. The percepts elicited by the Ternus-Pikler display are contingent on specific nonretinotopic temporal integration of elements and do not follow as a trivial consequence of general integration over large RFs.

Several areas in the human visual system have previously been found to encode subjective percepts rather than the retinotopic stimulation. For example, using induced motion, Takemura, Ashida, Amano, Kitaoka, and Murakami (2012) showed that illusory perceived speed rather than actual speed is represented in hMT+. Induced motion appears when a central stationary pattern is surrounded by a moving pattern (Duncker, 1929); the central pattern appears to move in the opposite direction. The central pattern eliciting the weakest response in hMT+ was not a static one, but one moving in the same direction as the surround with a speed cancelling the illusory motion percept. Even V1 has been found to encode subjective percepts: When

apparent motion is perceived, BOLD activations are found in the representation of the illusory motion path (Akselrod, Herzog, & Öğmen, 2014; Muckli, Kohler, Kriegeskorte, & Singer, 2005; but see also Liu et al., 2004). Apparent motion filling-in activity has been suggested to be supported by hMT+ (Sterzer, Haynes, & Rees, 2006), making this area a candidate for percept-based encoding. Although the hMT+ activations in our results seem to point to an influence of the percepts, they are compatible also with a purely retinotopic encoding. Further research is needed to determine which visual area first encodes the non-retinotopic percepts of the Ternus-Pikler display.

Acknowledgments

Commercial relationships: none.
Corresponding author: Evelina Thunell.
Email: evelina.thunell@cerco.ups-tlse.fr.
Address: Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

References

Akselrod, M., Herzog, M. H., & Öğmen, H. (2014). Tracing path-guided apparent motion in human primary visual cortex V1. *Scientific Reports, 4*, 6063, doi:10.1038/srep06063.

Albright, T. D., & Desimone, R. (1987). Local precision of visuotopic organization in the middle temporal area (MT) of the macaque. *Experimental Brain Research, 65*, 582–592, doi:10.1007/BF00235981.

Amano, K., Wandell, B. A., & Dumoulin, S. O. (2009). Visual field maps, population receptive field sizes, and visual field coverage in the human MT+

complex. *Journal of Neurophysiology, 102*(5), 2704–2718, doi:10.1152/jn.00102.2009.

Avidan, G., Harel, M., Hendler, T., Ben-bashat, D., Zohary, E., & Malech, R. (2002). Contrast sensitivity in human visual areas and its relationship to object recognition. *Journal of Neurophysiology, 87*, 3102–3116.

Bach, M. (1996). The "Freiburg Visual Acuity Test" – Automatic measurement of the visual acuity. *Optometry and Vision Science, 73*(1), 49–53.

Boi, M., Öğmen, H., Krummenacher, J., Otto, T. U., & Herzog, M. H. (2009). A (fascinating) litmus test for human retino- vs. non-retinotopic processing. *Journal of Vision, 9*(13):5, 1–11, doi:10.1167/9.13.5. [PubMed] [Article]

Boynton, G. M., Demb, J. B., Glover, G. H., & Heeger, D. J. (1999). Neuronal basis of contrast discrimination. *Vision Research, 39*(2), 257–269, doi:10.1016/S0042-6989(98)00113-8.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 16*(13), 4207–4221.

Breitmeyer, B. G., & Ritter, A. (1986a). The role of visual pattern persistence in bistable stroboscopic motion. *Vision Research, 26*(11), 1801–1806.

Breitmeyer, B. G., & Ritter, A. (1986b). Visual persistence and the effect of eccentric viewing, element size, and frame duration on bistable stroboscopic motion percepts. *Perception and Psychophysics, 39*(4), 275–280.

Buracas, G. T., & Boynton, G. M. (2007). The effect of spatial attention on contrast response functions in human visual cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 27*(1), 93–97, doi:10.1523/JNEUROSCI.3162-06.2007.

Clarke, A. M., & Herzog, M. H. (2013). Does spatiotemporal filtering account for nonretinotopic motion perception? Comment on Pooresmaeili, Cicchini. *Journal of Vision, 13*(10):19, 1–14, doi:10.1167/13.10.19. [PubMed] [Article]

Clarke, A. M., Öğmen, H., & Herzog, M. H. (2015). A computational model for reference-frame synthesis with applications to motion perception. *Vision Research,* in press, doi:10.1016/j.visres.2015.08.018.

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, doi:10.1038/nn1824.

Duncker, K. (1938). Uber induzierte Bewegung (Bin Beitrag zur Theorie optisch wahrgenommener Bewegung) [Concerning induced movement]. In W D. Ellis (Ed. and Trans.), *Source book of Gestalt psychology* (pp. 161–172). London: Routledge & Kegan Paul. (Reprinted from *Psychologische Forschung.* 1929, *12,* pp. 180–259.)

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. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 28*(15), 3988–3999, doi:10.1523/JNEUROSCI.5476-07.2008.

Goebel, R., Khorram-Sefat, D., Muckli, L., Hacker, H., & Singer, W. (1998). The constructive nature of vision: Direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *The European Journal of Neuroscience, 10*(5), 1563–1573. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9751129

Johansson, G. (1950). *Configurations in event perception, an experimental study*. Uppsala, Sweden: Uppsala, Almqvist & Wiksells Boktryckeri AB.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics, 14*(2), 201–211.

Kolster, H., Peeters, R., & Orban, G. A. (2010). The retinotopic organization of the human middle temporal area MT/V5 and its cortical neighbors. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 30*(29), 9801–9820, doi:10.1523/JNEUROSCI.2069-10.2010.

Liu, T., Slotnick, S. D., & Yantis, S. (2004). Human MT+ mediates perceptual filling-in during apparent motion. *NeuroImage, 21*(4), 1772–1780, doi:10.1016/j.neuroimage.2003.12.025.

Marques, J. P., Kober, T., Krueger, G., van der Zwaag, W., Van de Moortele, P.-F., & Gruetter, R. (2010). MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *NeuroImage, 49*(2), 1271–1281, doi:10.1016/j.neuroimage.2009.10.002.

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, doi:10.1371/journal.pbio.0030265.

Muckli, L., Kriegeskorte, N., Lanfermann, H., Zanella, F. E., Singer, W., & Goebel, R. (2002). Apparent motion: Event-related functional magnetic resonance imaging of perceptual switches and states. *The Journal of Neuroscience: The Official Journal of*

the Society for Neuroscience, 22(9), RC219, doi: 20026362.

Noory, B., Herzog, M. H., & Öğmen, H. (2015). Retinotopy of visual masking and non-retinotopic perception during masking. *Attention, Perception & Psychophysics, 77*, 1263–1284.

Öğmen, H., & Herzog, M. H. (2015). Apparent motion and reference frames. In J. Wagemans (Ed.), *Oxford handbook of perceptual organization* (pp. 487–503). Oxford, UK: Oxford University Press.

Öğmen, H., Otto, T. U., & Herzog, M. H. (2006). Perceptual grouping induces non-retinotopic feature attribution in human vision. *Vision Research, 46*(19), 3234–3242, doi:10.1016/j.visres.2006.04.007.

Pantle, A., & Picciano, L. (1976, Aug 6). A multistable movement display: Evidence for two separate motion systems in human vision. *Science, 193*, 500–502.

Pooresmaeili, A., Cicchini, G. M., Morrone, M. C., & Burr, D. (2013). Spatiotemporal filtering and motion illusions. *Journal of Vision, 13*(10):21, 1–4, doi:10.1167/13.10.21. [PubMed] [Article]

Pooresmaeili, A., Morrone, M. C., Cicchini, G. M., & Burr, D. (2012). "Non-retinotopic processing" in Ternus motion displays modeled by spatiotemporal filters. *Journal of Vision, 12*(1):10, 1–15, doi:10.1167/12.1.10. [PubMed] [Article]

Saenz, M., Lewis, L. B., Huth, A. G., Fine, I., & Koch, C. (2008). Visual motion area MT+/V5 responds to auditory motion in human sight-recovery subjects. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 28*(20), 5141–5148, doi:10.1523/JNEUROSCI.0803-08.2008.

Salomon, R., Darulova, J., Narsude, M., & Van Der Zwaag, W. (2014). Comparison of an 8-channel and a 32-channel coil for high-resolution FMRI at 7 T. *Brain Topography, 27*(2), 209–212.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., & Tootell, R. B. (1995, May 12). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science, 268*(5212), 889–893. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7754376

Sperry, R. W. (1950). Neural basis of the spontaneous

optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology, 43*(6), 482–489.

Sterzer, P., Haynes, J.-D., & Rees, G. (2006). Primary visual cortex activation on the path of apparent motion is mediated by feedback from hMT+/V5. *NeuroImage, 32*(3), 1308–1316, doi:10.1016/j.neuroimage.2006.05.029.

Takemura, H., Ashida, H., Amano, K., Kitaoka, A., & Murakami, I. (2012). Neural correlates of induced motion perception in the human brain. *The Journal of Neuroscience, 32*(41), 14344–14354, doi:10.1523/JNEUROSCI.0570-12.2012.

Thunell, E., Plomp, G., Öğmen, H., & Herzog, M. H. (2015). EEG correlates of relative motion encoding. *Brain Topography*, E-pub ahead of print, doi:10.1007/s10548-015-0458-y.

Tootell, R. B., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., & Belliveau, J. W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 15*(4), 3215–3230. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7722658

Tse, P. U. (2006). Neural correlates of transformational apparent motion. *NeuroImage, 31*(2), 766–773, doi:10.1016/j.neuroimage.2005.12.029.

van der Zwaag, W., Francis, S., Head, K., Peters, A., Gowland, P., Morris, P., & Bowtell, R. (2009). fMRI at 1.5, 3 and 7 T: Characterising BOLD signal changes. *NeuroImage, 47*(4), 1425–1434, doi:10.1016/j.neuroimage.2009.05.015.

von Holst, E., & Mittelstaedt, H. (1973). Das Reafferenzprinzip Wechselwirkungen zwischen Zentralnervensystem und Peripherie. *Naturwissenschaften, 37*, 464–476. (Translated in: *The behavioral physiology of animals and man. Selected papers from E. von Holst*. 1973. Coral Gables, FL: University of Miami Press.)

Yan, T., Wang, B., Geng, Y., Yan, Y., Mu, N., Wu, J., & Peng, Y. (2014). Contrast response functions with wide-view stimuli in the human visual cortex. *Perception, 43*(7), 677–693, doi:10.1068/p7640.