PNAS
www.pnas.org

1

Supplementary Information for

Columnar clusters in the human motion complex reflect consciously perceived motion axis

Marian Schneider, Valentin Kemper, Thomas Emmerling, Federico De Martino and Rainer Goebel

Marian Schneider.
E-mail: marian.schneider@maastrichtuniversity.nl

This PDF file includes:

Supplementary text
Figs. S1 to S13
Tables S1 to S4
Caption for Movie S1
Caption for Database S1
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Movie S1
Database S1

Supporting Information Text

SI Materials and Methods

Participants. All participants were students or employees of Maastricht University and recruitment was limited to participants who were MRI compatible and had been in an MRI scanner at least once before to ensure high subject compliance. Four participants (S2, S4, S5, S6) were excluded from further data analysis (S2 did not show up for the second scanning session; S4 performed the task inappropriately during scanning, pressing buttons to indicate perceptual switches every 200 - 400 ms; S5 showed excessive head movement and S6 did not perceive stable states of horizontal and vertical apparent motion during the training session). Hence, five participants (3 females, 23-28 years old) were analyzed. All participants were experienced in fixation tasks.

Experimental Design and Stimuli. Visual stimuli were created and presented using the open-source application PsychoPy (version 1.82.01) (1, 2). Stimuli were projected on a frosted screen at the head end of the scanner (using Panasonic projector PT-EZ570; Newark, NJ, USA; resolution 1920 x 1200; nominal refresh rate: 60 Hz). Subjects viewed the screen via a tilted mirror attached to the head coil. Button responses were registered using an MR compatible button box (Current Designs, 8-button response device, HHSC-2 × 4-C; Philadelphia, USA). All scripts used for stimulus presentation are available as GitHub and zenodo repository.

For every subject we collected 3 runs to localize area hMT+ (240 volumes, 12 repetitions per condition per run), 5-6 runs of the ambiguous motion experiment (300 volumes, 22.30 ± 1.14 (mean $\pm$ standard deviation) repetitions per condition per run) and 5-6 runs of the physical motion experiment (300 volumes, 24 repetitions per condition per run). These runs were acquired across two different scanning sessions on separate days. Additionally, we performed three control experiments to exclude the possibility that observed differences between horizontal and vertical conditions were driven by retinotopic differences: For two subjects (S1 and S3), we acquired 3 runs (204 volumes, 6 repetitions per condition per run) to estimate axis-of-motion tuning curves (control experiment I) which were recorded during the first scanning session of the motion quartet experiments. Three subjects (S1, S3, S9) were recruited for an additional, third scanning session and we acquired 5-8 runs (222 volumes, 12 repetitions per condition per run) to map out responses to retinotopically identical horizontal and vertical motion conditions (control experiment II). Finally, for two subjects (S3 and S7) we organized a fourth scanning session to obtain 6 runs (172 volumes per run, in total 27 repetitions per aperture position) of data that allowed us to estimate population receptive field (pRF) parameters (control experiment III).

hMT+ localizer. Area hMT+ was localized using standard moving and static dot stimuli (3–5) presented in a circular aperture (10° of visual angle in diameter) that was either centered or displaced by 5° of visual angle to the left or right. All dots (200 dots; 0.2° in diameter; white on black background) moved at 8° per second and were either expanding or dilating, alternating direction every second. Every dot had a limited life time of 167 ms (10 frames) and was reborn at a randomized location within the aperture. 4s periods of moving dots were followed by periods of static dots (randomized duration: either 6s, 8s or 10s). The aperture position for static dots always matched the preceding aperture position for moving dots. Participants were asked to fixate on a centrally presented dot throughout the entire experimental run. The fixation dot consisted of a red circle surrounded by a yellow annulus (6). The fixation dot occasionally (20 targets per run) changed color from red to yellow for 0.3 s and participants indicated a color change by pressing a button.

Ambiguous and physical motion experiment. Participants were presented with two different motion quartet stimuli - an "ambiguous" and a disambiguated or "physical" version. In the ambiguous version (Experiment 1) we used the apparent motion quartet (7) (see upper panel Fig. 1A). The quartet is composed of four blinking squares (each 1° x 1° visual angle) in a rectangular configuration. Crucially, at any time point, only two squares at diagonally opposite corners are shown. A pair of squares was presented for 150 ms (9 frames) followed by an inter-stimulus interval of 67 ms (4 frames). Such a presentation frequency of 2.3 Hz was shown to lead to strong perception of apparent motion (8).

In the physical version (Experiment 2) squares moved unambiguously either horizontally or vertically. This was achieved by physically moving white squares along the motion paths that were previously only perceived in Experiment 1 (see lower panel Fig. 1A). Square positions were updated according to a harmonious oscillation, where the square location was updated as the sine of time (9). Stimulus parameters in Experiment 2 were matched to those in Experiment 1: Square size remained 1° x 1° of visual angle. Outermost horizontal and vertical points of the physical motion trajectory were equivalent to square positions in Experiment 1 (individually adjusted for every subject during a training session, see SI). Duration of a movement cycle either along the horizontal (left to right to left) or vertical trajectories (top to bottom to top) was 433 ms (26 frames), which matched the presentation frequency in Experiment 1.

The ambiguous motion quartet was shown during periods of 40s, which were interleaved with 16s baseline periods. A flicker motion quartet in which all four squares were shown synchronously served as the baseline. This local flicker stimulus did not induce apparent motion but still had the same stimulus energy as the regular motion quartet. Participants were instructed to fixate a red dot in the center of the screen throughout the experiment and to indicate the perceived axis of motion (vertical or horizontal) by button responses. The mapping of perceived motion axis to buttons (whether button "1" or button "2" was used to indicate horizontal or vertical motion) was counterbalanced across the two scanning sessions.

The physical motion stimulus was equally shown during periods of 40 s in total. However, this time the motion axis was unambiguous and alternated every 10 s between horizontal and vertical motion. During the 16 s baseline period, participants

Marian Schneider, Valentin Kemper, Thomas Emmerling, Federico De Martino and Rainer Goebel

again viewed the flicker motion quartet. They fixated the central red dot and indicated the perceived axis of motion by button responses. This was done although the motion axis was clearly determined by the stimulus, in order to engage participants in a task and to keep the task constant across Experiments 1 and 2.

Control experiment I: Axis-of-motion tuning. In order to determine the axis-of-motion tuning of voxels, we presented moving and static dot stimuli in a central aperture (aperture diameter 10° visual angle; 200 dots; 0.2° in diameter; white on black background) such that the aperture field covered the real and perceived motion trajectories in Experiments 1 and 2. There were four different conditions, each corresponding to one of four axes of motion (0°-180°, 45°-225°, 90°-270°, 135°-315°; thus including the vertical and horizontal motion axes occupied in Experiments 1 and 2). 6s periods of moving dots were interleaved with static rest periods that lasted 8s, 10s or 12s. The order of conditions was randomized. During axis-of-motion blocks the direction of dots alternated every second; e.g. during the 0°-180° condition leftward motion alternated with rightward motion. Dots were moving at a speed of 8° of visual angle per second. Participants performed the same fixation task as described for the hMT+ localizer.

Control experiment II: Moving dot pattern. In order to determine the tuning of voxels to motion either along the horizontal or vertical axis, we repeated control experiment I with two modifications. First, instead of four there were only two conditions, showing dots moving either along the horizontal (0°-180°) or vertical (90°-270°) motion axis. Second, visibility of the dots was restricted to a square annulus aperture. Shape and location of the aperture were defined as to reveal positions in visual space that had been occupied during the motion quartet experiments (simultaneously the locations of both horizontal and vertical motion trajectories as well as the four inducer squares, see Fig. S7C). Importantly, this resulted in the horizontal and vertical motion conditions being retinotopically identical and only differing with regard to the motion axis of the dots revealed by the aperture. Area hMT+, unlike areas V1-V3, does not display an aperture bias resulting in higher responses at retinotopic locations where moving dots enter the stimulus aperture, making this experimental set-up suitable to study the question of interest here (10). The aperture was adjusted for every control subject to match subject-specific vertical distances (see Table S4). Other parameters were identical to those reported for control experiment I.

Control experiment III: Population receptive field mapping. The pRF mapping stimuli consisted of apertures that took the shape of semi rings and were presented at systematically varied positions. A single semi ring aperture had a radial extent of 3° of visual angle and subtended 180 angular degrees. Apertures were presented at four orientations (horizontal, vertical and the two diagonal orientations in-between) and at four eccentricity levels (centered on eccentricities of 1.5°, 4.5°, 7.5°, and 10.5° of visual angle). Apertures at the innermost eccentricity level (1.5°) consisted of semi circles, not semi rings. All apertures were limited to a circular region of the display, 24° of visual angle in diameter, and presented against a mean luminance gray background.

The carrier pattern of the mapping stimuli consisted of a random texture pattern as described by (11) and was presented at 98.2% Michelson contrast (which was the highest possible contrast in the scanner environment). Aperture positions were updated with every TR (3s) and in a pseudo-random fashion such that two subsequent apertures never overlapped retinotopically. Every functional run started and ended with 15 s blank screen. To aid the estimation of large pRFs, 18 x 3s periods of blank screen were inserted throughout the functional run (4).

Participants performed the same fixation task as described for the hMT+ localizer. To facilitate central fixation, an extended grid of thin lines was presented throughout the experimental run (12). The grid consisted of two diagonal, orthogonal lines crossing behind the fixation dot as well as circles with radii of 1, 2.5, 5, 10 and 15 degrees of visual angle centered on the fixation dot.

Pre-scan training session. With central fixation, observers of the ambiguous motion quartet more frequently perceive vertical than horizontal motion (13). This is thought to reflect the cost of inter-hemispheric processing, which is necessary for horizontal but not for vertical motion (13, 14). To ensure that the motion quartet stimulus was bistable, i.e. that horizontal and vertical perceptual intervals were approximately equal in length, we scheduled a 20-30 minutes training session for every participant before the first scanning session. During this training session, we kept the horizontal distance between squares of the motion quartet constant (3° visual angle from square center to central fixation point), while we adjusted the vertical distance of the squares for every subject to obtain a bistable stimulus display. Table S4 shows the resulting subject-specific vertical distances. The ratio of vertical to horizontal distances was between 1.26 and 1.30 for all subjects, which is in agreement with previously reported ratios for motion quartet stimuli to achieve equal frequencies (13). Once the vertical distance had been calibrated for a subject, it was kept constant across Experiment 1 and 2 and the two scanning sessions.

MRI Acquisition. Data acquisition was performed on a whole-body Magnetom scanner (nominal field strength 7 Tesla, Siemens Medical Systems, Erlangen, Germany) at the Maastricht Brain Imaging center, The Netherlands. All images were acquired using a 32-channel head-coil (NovaMedical Inc.; Wilmington, MA, USA).

hMT+ localizer. To aid the positioning of the sub-millimeter slab for Experiments 1-2 and to determine our region of interest, we acquired one hMT+ localizer scan in the first session and two hMT+ localizer scans in the second session. We used a 2D gradient echo (GE) echo planar imaging (EPI) sequence (1.6 mm isotropic nominal resolution; TE/TR = 18/2000 ms; in-plane field of view (FoV) 150×150 mm; matrix size 94 x 94; 28 slices; nominal flip angle (FA) = 69°; echo spacing = 0.71 ms; GRAPPA factor = 2, partial Fourier = 7/8; phase encoding direction head - foot). We ensured that the area of acquisition had bilateral coverage of the posterior inferior temporal sulci, where we expected the hMT+ areas. Before acquisition of the

first functional run, we collected 10 volumes for distortion correction - 5 volumes with the settings specified above and 5 more volumes with identical settings but opposite phase encoding (foot - head).

Ambiguous and physical motion experiment. For the sub-millimeter measurements, we used a 2D GE EPI sequence (TE/TR = 25.6/2000 ms; in-plane FoV 148×148 mm; matrix size 186 x 186; slices = 28; nominal FA = 69°; echo spacing = 1.05 ms; GRAPPA factor = 3, partial Fourier = 6/8; phase encoding direction head - foot), yielding a nominal resolution of 0.8 mm isotropic (15, 16). Placement of the small functional slab was guided by online analysis of the hMT+ localizer data recorded immediately at the beginning of the first session. This allowed us to ensure bilateral coverage of area hMT+ for every subject. In the second scanning session, the slab was placed using Siemens auto-align functionality and manual corrections. Before acquisition of the first functional run, we collected 10 volumes for distortion correction (5 volumes with opposite phase encoding: foot - head). During acquisition, runs for the ambiguous and physical motion experiments were interleaved.

Control experiments. The acquisition parameters for control experiments I and II were identical to those reported for the motion quartet experiments. For the pRF mapping, we used the same 2D GE EPI sequence as in the motion quartet experiments with slightly adjusted parameters to obtain a larger field of view (in particular, multi band factor = 2; TE/TR = 23.2/3000 ms; in-plane FoV 140×140 mm; matrix size 176 x 176; slices = 82; nominal resolution = 0.8; nominal FA = 82°; echo spacing = 1.04 ms; GRAPPA factor = 3; partial Fourier = 6/8; phase encoding direction head - foot) (15–17).

Anatomical scans. For visualization of the functional results, we acquired scans with structural information in the first scanning session. At high magnetic fields, MR images exhibit high signal intensity variations that result from heterogeneous RF coil profiles (18). We therefore acquired both T1w images and PDw images using a magnetization-prepared 3D rapid gradient-echo (3D MPRAGE) sequence (TR: 3100 ms (T1w) or 1440 ms (PDw), voxel size = 0.6 mm isotropic, FOV = 230 x 230 mm2, matrix = 384 x 384, slices = 256, TE = 2.52 ms, FA = 5°). Acquisition time was reduced by using 3× GRAPPA parallel imaging and 6/8 Partial Fourier in phase encoding direction (acquisition time (TA): 8 min 49 s (T1w) and 4 min 6 s (PDw)).

Behavioral data analysis. For every subject we calculated the mean length of horizontal and vertical perceptual periods during the ambiguous motion experiment. To test for differences between horizontal and vertical periods, we conducted an independent-samples t-test for every subject (p<.05, two-sided). We also calculated the mean, standard deviation and range across subjects for the two types of perceptual periods.

Structural data analysis and segmentation. Structural images were processed using advanced segmentation tools in BrainVoyager 20.0 (Brain Innovation, Maastricht, The Netherlands), SPM's bias correction (19), ITK-SNAP (20), FSL BET (21), morphological operations (22) and Segmentator (23). Where not specified otherwise, default settings were used. Since the processing included many different steps and programs, a diagram of the input/output relationships for all processing steps is provided in Fig. S11A. We first registered the PDw image for every subject to their T1w image, using ITK-SNAP's automatic co-registration tools. In order to reduce the B1 bias field, we divided the T1w image by the PDw image and obtained a ratio image (18). The ratio image was brain-masked with a masked obtained by inputting the co-registered PDw image to FSL BET. We then used SPM's anatomical bias correction to further reduce inhomogeneities in the ratio image.

The resulting image was input to BrainVoyager's advanced segmentation routine to obtain a white matter (WM) definition. This initial WM image was inspected and manually polished in ITK-SNAP using the adaptive paint brush mode in combination with a graphics tablet (Intuos Art; Wacom Co. Ltd; Kazo, Saitama, Japan). Special emphasis was placed on corrections in the region of interest (bilateral hMT+). In two additional steps, using the round paint brush mode, the parts of the cerebellum and sagittal sinus that were falsely included in the WM definition were removed and brain stem structures and ventricles were masked.

The thus polished WM definition and ratio image were input to BrainVoyager's advanced segmentation routine to obtain a gray matter (GM) definition. This GM definition tended to be too inclusive, containing dura mater and blood vessels. For this reason, we masked the ratio image with the WM and GM definitions and input the resulting image to Segmentator. This allowed us to further exclude non-brain voxels from the GM definition based on their 2D histogram profile along the dimensions of image intensity and gradient magnitude. The thus improved GM definition was inspected and further polished manually in ITK-SNAP using the round paint brush mode. For later mesh visualization, the WM-GM segmentation images were separated in two hemispheres using ITK-SNAP's scalpel tool. Remaining topological errors were corrected using the bridge removal option in BrainVoyager (24) and manual correction. Note that all segmentation steps were performed at the original resolution of the anatomical images (0.6 mm isotropic).

Functional data - preparation. Functional data were processed using BrainVoyager 20.0, SPM 12 (25), FSL 5.0 (26) as well as custom code in Python 2.7 (22, 27, 28) and in MATLAB R2014a (The MATHWORKS Inc.; Natick, MA, USA). Where not specified otherwise, default settings were used. All functional pre-processing steps were scripted and scripts are available as a GitHub and zenodo repository. Figure S11B provides an overview of all functional pre-processing steps.

Pre-processing steps. Pre-processing for all sub-millimeter images was performed in the following order: slice scan time correction (BrainVoyager), motion-correction (SPM 12), linear trend removal and high-pass filtering (5 cycles) using a general linear model (GLM) Fourier basis set (BrainVoyager) and distortion correction (FSL topup). For motion and distortion correction, we deviated from default settings. Functional images from the first session were motion-corrected using SPM 12 in three

steps. First, the first image of each run was realigned to the first image of the first run. Second, the images within each run were aligned to the first image of the run. Third, to avoid local minima, after these first two steps, a mean of all images was calculated and images were realigned to this mean. Motion correction was limited to voxels inside the brain based on an intensity-thresholded and manually corrected brain mask of the functional images. Note that the results of the three steps were combined into a single transformation that was applied to functional images to minimize interpolation artefacts. Functional images from the second scanning session were motion-corrected in a similar vein, with the difference that the first functional image of the second session served as reference image in the first correction step. Steps 2 and 3 were identical. For details on the alignment across scanning sessions, please see below. EPI distortions of the functional images were corrected using FSL topup (29, 30). The pairs of opposite phase encoding images acquired at the beginning of the first session were input to topup to estimate the susceptibility-induced off-resonance field. The estimated field was then used to correct the distortions for all functional images of the first and second scanning session. For the MT+ localizer images, pre-processing steps included slice scan time correction, motion-correction, linear trend removal combined with high-pass filtering (all performed in BrainVoyager) and distortion correction (FSL topup). In addition to the pre-processing steps described above, we applied Gaussian spatial smoothing (smoothing kernel = 0.8 mm) to the pRF mapping data, reflecting the prior assumption that pRF parameters vary smoothly over space.

Registration to functional images. All statistical analyses were conducted in the space of the sub-millimeter functional images in the first scanning session (hereafter called "voxel space"). To this end, the pre-processed anatomical T1w-divided-PDw ratio image was registered to voxel space by exploiting the scanner's positional information and fine-tuning co-registration with boundary-based registration (31) as implemented in BrainVoyager. The result was visually inspected for each subject by overlaying functional and anatomical images (anatomical images were displayed with inverted image intensities for better visualization). Co-registration and down-sampling of the anatomical image were performed in one step.

In order to register the high-resolution segmentation images (0.6 mm isotropic) to lower-resolution voxel space (0.8 mm isotropic) we proceeded as follows. Problematically, simply transforming an image with only two segmentation values (for WM and GM) results in interpolation artefacts and how to bin the resulting distribution of values to regain only two segmentation values is not obvious. For that reason, we performed a surface reconstruction of both the inner and outer GM border in BrainVoyager. This resulted in two different surface meshes to which we applied the established transformation from anatomical images to voxel space. Transformed meshes were smoothed, using advanced mesh smoothing tools (restricting vertex displacement to 0.5). Importantly, the advanced mesh smoothing is restricted to high-frequencies and leaves low-frequency changes such as cortical folds intact. Smoothed meshes were projected back into volume space and area filling tools were used to regain a segmentation image with only two values. We inspected the resulting WM-GM segmentation images in voxel space separately for every subject and found this procedure to preserve important features of the segmentation and to minimize the re-introduction of topological errors (see Fig. S12A). Re-introduced errors were manually corrected using itksnap. This was limited to 20-30 mislabeled voxels in our region of interest.

hMT+ localizer functional images acquired in the second session were first registered to the anatomical images in the first session based on the scanner's positional information, manual corrections, and boundary-based registration (31) for fine-tuning. Subsequently, the already established transformation from anatomical images to voxel space was used. MT+ localizer functional images were up-sampled from 1.6 mm isotropic to 0.8 mm isotropic using nearest-neighbor interpolation. All transformations were combined and applied in a single step to avoid unnecessary interpolation.

To align the sub-millimeter functional images acquired in the second scanning session (and in the control experiments) to voxel space, we chose a more complex procedure than for the hMT+ localizer functional images. This was because we found EPI distortions to differ across scanning sessions for sub-millimeter functional images and simply co-registering both (distortion-corrected) functional sessions to the anatomy did not meet our quality demands for sub-millimeter analyses. For that reason, we calculated a mean image across time, separately for every scanning session (before EPI-distortion correction). We then ran an initial affine registration between the mean image of the first session and each of the mean images of the other sessions using FSL FLIRT (32, 33) with 12 degrees of freedom (dof) and, since we noticed residual distortion differences, a subsequent non-linear registration using FSL FNIRT (34). The resulting linear and non-linear transformations were combined into a single transform, which was applied to all sub-millimeter images recorded outside the first scanning session. We visually inspected the quality of the resulting alignment between the first run in the first session and all runs in the other sessions. Only then EPI distortion correction was applied to the transformed images, using the off-resonance field estimated in session 1. We applied a mask to all images, excluding voxels that had a mean intensity value below 12 in at least one of the sessions. This excluded voxels at the fringes of the slabs (due to minimal differences in slab placement across the two sessions) and we verified that voxels in our region of interest (ROI) were not affected.

ROI definitions. In order to define our ROI (bilateral hMT+), we calculated a voxel-wise general linear model (GLM) on a single-subject level for the hMT+ localizer data. The GLM was corrected for temporal auto-correlation (AR2). The model contained a separate predictor for the three stimulus conditions (moving dots in the left, central and right aperture). We selected voxels that showed a significant response to the central condition (using a threshold (q) corrected for multiple comparisons using false discovery rate; q(FDR) < .05; S1: t(474)>3.00; S3: t(474)>2.80; S7: t(474)>2.68; S8: t(474)>2.81; S9: t(474)>2.75). For each hemisphere, we projected this selection of voxels onto the inflated surface reconstructed along the middle of gray matter (for every vertex, the maximum statistical value in a range from -2mm to +2mm from the middle of gray matter was displayed). This allowed us to delineate area hMT+ on each hemisphere by manual drawing. Although drawing introduced a component of

subjective judgment, the degree of subjectivity was minimal given that the areas were clearly outlined by significant responses (for an example, see Fig. S1). Significant responses needed to be located at the posterior/dorsal limb of the inferior temporal sulcus to be included in the ROI, in keeping with previous empirical work reporting the location of area hMT+ (3, 11, 35). Figure S2 shows the resulting definitions on the surface for left and right hemispheres. The surface patches were projected back into volume space (from -2mm to +2mm from the middle of gray matter) and restricted to GM. Delineations on the left and right hemisphere were grouped together to constitute a single ROI per subject. Table S1 lists the number of voxels included in the ROIs as well as the estimated areas of left and right hMT+ on the surface. In summary, three criteria thus determined the selection of voxels for our ROI: (1) significant responses to central condition of independent hMT+ localizer data, (2) anatomical position at the posterior/dorsal limb of ITS, (3) voxels needed to be in GM.

Event-related signal modulations during ambiguous and physical motion. In order to quantify the event-related signal increases and decreases observed during ambiguous and physical motion (Fig. 2), we extracted the 12s time periods (6 TRs) after subjects indicated a perceptual switch from one motion axis to the other. For ambiguous motion, the average signal across all voxels in either the horizontal or vertical cluster was extracted. Assignment of voxels to clusters was based on responses to the physical motion stimulus. In order to extract signal for physical motion without introducing circularity, we employed a leave-one-run-out cross-validation scheme. We used all but one run to assign voxels to either the horizontal or vertical cluster and extracted perceptual periods from the left-out run.

We expected average event-related signal to increase for horizontal/vertical motion percepts in the horizontal/vertical cluster. Reversely, we expected signal to decrease for vertical/horizontal motion percepts in the horizontal/vertical clusters. When we expected an increase, we calculated the difference between the maximum of the last two time points of the extracted time period (volume 5 or 6) and the minimum of the first two time points (volume 1 or 2). When we expected a decrease, we calculated the difference between the minimum of the last two time points of the extracted time period and the maximum of the first two time points. This amounts to $n = 4$ multiple comparisons. We chose the first two time points based on our expectation that signal in these volumes reflected the percept before the perceptual switch (due to the hemodynamic delay of the fMRI signal). We chose the last time points because we expected the signal to reflect the new percept after the switch.

Functional data - statistical analyses of control experiments.

Control experiment I. In order to obtain axis-of-motion tuning curves, we followed the same procedure as described in (36). Briefly, we fit a GLM with predictors for each of the four motion axes and calculated four contrasts testing for modulation in response to each of the motion axes, which resulted in four t-values per voxel. We demeaned the t-values and averaged responses per motion axis and cluster. To test for statistical significance of the obtained tuning curve, we proceeded in two steps. First, we converted the four responses in every cluster to ranks (from high to low) and compared these ranks to idealized ranks for either a horizontally or vertically tuned cluster. The idealized rank pattern for a horizontally tuned cluster was coded as [1, 2.5, 3, 2.5], expressing the expectation that the highest rank should be obtained for the horizontal motion axis, the lowest rank for the vertical motion axis, and tied ranks for diagonal motion axes. Reversely, the idealized rank pattern for a vertically tuned cluster was coded as [3, 2.5, 1, 2.5]. We calculated the Kendall rank correlation coefficient (Tau-b statistic) between observed and idealized ranks (empirical correlation). Second, we obtained a null distribution of tuning curves by randomly permuting condition labels and rerunning the analysis described above (1000-fold permutation testing). This resulted in 1000 axis-of-motion tuning curves as well as 1000 correlation coefficients. We compared the empirically observed correlation coefficient to the null distribution of coefficients. If the empirical correlation coefficient was above the 97.5th percentile of the null distribution, the motion tuning was declared significant. Ties in correlation coefficients were handled using the 'weak' method as implemented in the 'percentileofscore' function in scipy (22).

Control experiment II. The data obtained in control experiment II allowed us to conduct two types of analysis. First, we again investigated tuning of responses in the horizontal and vertical clusters. For this analysis, we proceeded in a similar fashion as for control experiment I, with the following modifications. We fit a GLM with only two predictors for horizontal and vertical motion conditions and calculated two contrasts, resulting in two t-values per voxel. To test for statistical significance of the tuning, in each cluster we first formed differences between empirically observed average t-values for horizontal and vertical motion (empirical difference). We then obtained a null distribution of differences by randomly permuting condition labels and re-running the analysis described above (1000-fold permutation testing). If the empirical difference was above the 97.5th percentile of the null distribution, the motion tuning was declared significant.

Second, we investigated whether amplitude modulations during ambiguous motion would still reflect the perceived motion axis even if voxel selection was based on retinotopically identical conditions. To this end, we assigned voxels within the ROI used in the main experiment to either horizontal or vertical clusters, depending on their responses during control experiment II. We formed clusters based on the 100 voxels with the largest absolute t-values for the contrast horizontal > vertical. For each cluster, we obtained average time courses for the ambiguous motion experiment and calculated t-values for the contrast horizontal > vertical in the horizontal clusters and for the contrast vertical > horizontal in the vertical clusters. These empirical values were compared to a null distribution from a 1000-fold permutation analysis and modulation of a cluster's time course was considered significant if its empirical t-value was above the 97.5th percentile of the null distribution. To see if our results were robust to a varying number of voxels included in the cluster, we systematically varied the number of included voxels (from 25 to 50, 75, 100, 250 or 500).

Marian Schneider, Valentin Kemper, Thomas Emmerling, Federico De Martino and Rainer Goebel

Control experiment III. The pRF mapping data was analyzed as described in (37). Briefly, every voxel time course was assumed to be generated by an isotropic 2D Gaussian in visual space. Model responses were calculated by transforming the visual stimuli presented during the pRF mapping experiment to binary apertures (indicating presence or absence of a stimulus in space) and spatially convolving the apertures with an instance of the 2D Gaussian model. The Gaussian model includes parameters for its position (x- and y-position) and its size (sigma; standard deviation of the 2D Gaussian) in visual space and the predicted model time course will vary with assumed parameters. To account for hemodynamic delay in the fMRI response, we convolved every model time course with a canonical double-gamma function (25). We determined x, y and sigma parameters for every voxel by minimizing the sum of squared errors (SSE) between the predicted model and the observed fMRI time course. In particular, we systematically varied parameters and created candidate model time courses on a polar grid: 48 radial position (0°-12° eccentricity) x 24 polar positions (0-2π) x 40 pRF sizes (0°-22°). We then scaled the model time courses to every voxel time course using a GLM and minimized SSE in a brute force search across candidate models. In order to avoid over-fitting, a leave-one-run-out cross-validation procedure was employed: the GLM was fit on $n-1$ runs and the SSE were calculated between the predicted and the fMRI time series of the left-out run. The winner model for every voxel was determined by minimizing the mean SSE across all cross-validation folds. Exhaustively searching a very dense grid was achieved in a timely manner using multi threading and optimized Cython code, which is publicly available as part of the PyPrf feature package (38).

Based on the resulting pRF parameters, we created a representation of the visual field coverage, separately for voxels in the vertical and horizontal cluster. Clusters included voxels from both left and right hemispheres. For every voxel we recreated its estimated 2D Gaussian pRF in visual space (modeled as a 512 x 512 pixels square). For this analysis, the 2D Gaussians are normalized such that the maximum value of the Gaussian was equal to 1. For every cluster, we summed up the resulting values across all voxels at every pixel and divided by the total number of voxels included in the cluster. This results in coverage values between 0 and 1, where 1 would indicate that all voxels in a cluster had their receptive field center on this particular pixel of the visual field representation. These steps were implemented in the openly available ViFiCov package (39).

We also calculated for every voxel in the clusters by how much their estimated 2D Gaussian pRF overlapped with either the horizontal or vertical motion trajectories that were shown or perceived in the motion quartet experiments. Motion trajectories were represented as binary images. This resulted in two overlap values per voxel, for overlap with either horizontal or vertical trajectories. Using a paired samples t-test, we tested for every cluster separately for a difference in overlap with the horizontal and the vertical motion trajectory. We also formed the difference of overlap values for every voxel and compared the mean difference values between horizontal and vertical clusters using an independent samples t-test. For visualization of the visual field coverage as well as the statistical tests, we employed an R2 threshold such that voxels with an R2 value lower than .05 were excluded. We also report the statistical results with all voxels included in the analysis.

Calculation of directional spatial autocorrelation. As an additional test for columnarity, we compared spatial autocorrelation of physical motion preferences in the cortical depth and in the cortical plane direction. If motion preferences were organized in a columnar fashion, we would expect that motion preferences display a higher autocorrelation coefficient in the cortical depth than in the cortical plane direction. To quantify the degree of spatial autocorrelation, we calculated Moran's I (40, 41) as:

$$I = \frac{n \sum\limits_{i \neq j} w_{ij} z_i z_j}{\sum\limits_{i \neq j} w_{ij} \sum\limits_i z_i^2} \tag{1}$$

where n is the total number of grid points; $\sum\limits_{i \neq j}$ is the double summation of all grid points i from 1 to n and j from 1 to n where $i \neq j$; $z_i = Y_i - \overline{Y}$, where Y_i is the value of motion preference at grid point i and $\overline{Y}$ is the mean of Y for all grid points; finally, w_{ij} is the weight describing the relationship between grid point i and grid point j and was based on the distance between the two grid points. In particular, we constructed several distance bins. Bins ranged from 1mm to 4mm cortical distance with a bin width of 0.6mm. Below 1mm and above 4mm we had insufficient data points given that cortical depth levels were separated by at least 0.4 * relative cortical thickness (lower bound) and at most 0.8* cortical thickness (upper bound). We then iterated over the distance bins and assigned a weight of 1 if the distance of a pair of grid points fell within the bin and a zero otherwise and calculated the autocorrelation coefficients for each bin separately.

To determine whether there was a difference in autocorrelation between the cortical depth and the cortical plane direction, we calculated directional autocorrelation (41, 42). This was achieved by updating the weights w_{ij} as follows. To determine autocorrelation in the cortical plane direction, we only considered relationships between grid points that were located at the same depth level (i.e. the same grid surface) by setting all weights to zero for relationships between grid points located on different depth levels. Conversely, in order to calculate autocorrelation in the cortical depth direction, we only considered relationships between grid points that were in the same grid column and set all weights to zero for relationships between grid points that were not located in the same grid column.

SI Results

Behavioral results. All participants reported regular perceptual switches between horizontal and vertical percepts during the ambiguous motion experiment. Across all participants, mean perceptual periods ranged from 7.2s to 13.3s. The average length of horizontal and vertical perceptual periods was 10.8s ± 1.9s and 10.8s ± 2.4s (mean ± standard deviation across subjects),

respectively. Goal of the calibration of the motion quartet stimulus during the training session had been 10s long percepts. Figure S13 shows the average length of horizontal and vertical perceptual periods individually for every subject. For subjects S3 and S9 we found a statistical difference between the length of horizontal and vertical periods (t(249)=-3.78, p<.001 and t(336)=4.80, p<.001, respectively). For subjects S1, S7 and S8 we found no statistical difference (S1: t(252)=-0.41, p=.685; S7: t(223)=-0.07, p=.947; S8: t(179)=1.14, p=.255). We thus did not observe bias towards one of the two motion axes that was consistent across participants.

Control experiments.

Control experiment I. Figure S7A shows the axis-of-motion tuning curves derived from control experiment I in clusters used in the main experiment for two control participants. We observed that clusters showed expected axis-of-motion tuning. Horizontal clusters showed the largest average t-value for the horizontal motion condition, lower t-values for the diagonal motion conditions and the lowest t-value for the vertical motion condition. For vertical clusters we observed a reversed pattern. The permutation testing showed significant correlation for ranks in the horizontal cluster and an idealized horizontally tuned cluster (S1: τ=0.91, p<.001; S3: τ=0.91, p<.001). Ranks of the vertical cluster correlated significantly with ranks of an idealized vertically tuned cluster (S1: τ=0.91, p<.001; S3: τ=0.91, p<.001).

Control experiment II. Figure S7B shows motion tuning derived from control experiment II for three control participants. All clusters showed tuning in the expected direction such that horizontal clusters show larger average t-values for the horizontal than for the vertical condition and vertical clusters show a reversed pattern. The permutation testing showed significant differences between average t-values for the horizontal and vertical motion condition in both the horizontal (S1: Δt=0.43, p<.001; S3: Δt=0.25, p<.001; S9: Δt=0.27, p<.001) and vertical cluster (S1: Δt=-0.65, p<.001; S3: Δt =-1.21, p<.001; S9: Δt =-0.68, p<.001).

Control experiment III. Upon visual inspection, we did not find a difference in visual field coverage between horizontal and vertical clusters (Fig. S8A). All clusters showed coverage of the area between and foveal to the inducer squares. We found that clusters had more coverage of the vertical than the horizontal trajectory (Fig. S8B), which was significant in all clusters (S3 horizontal cluster: t(165)=-4.21, p<.001; S3 vertical cluster: t(40)=-6.35, p<.001; S7 horizontal cluster: t(522)=-22.15, p<.001; S7 vertical cluster: t(207)=-9.40, p<.001). This is partially to be expected since the retinotopic trajectory is larger for the vertical than for the horizontal condition (by a factor of 1.26 for both control participants). We did not find a difference between horizontal and vertical clusters in the amount they overlapped with either the horizontal or vertical motion trajectory (S3: t(205)=0.066, p=.948; S7: t(729)=1.26, p=.209). We obtained similar findings when no R2 thresholding was performed and all voxels were included in the analysis (S3: t(549)=-1.56, p=.119; S7: t(1213)=1.40, p=.163).

Directional spatial autocorrelation. Figure S10 shows how spatial autocorrelation of preferences for physical motion changes with distance both in the cortical depth and the cortical plane direction. Spatial autocorrelation is consistently higher in the cortical depth than in the cortical plane direction, across participants and across a wide range of cortical distances (with the exception of small distances for participant S9). At low cortical distances, the autocorrelation is high both in the cortical depth and plane direction. Autocorrelation decreases with cortical distance for both directions; however, in several participants (S1, S3, S9) remains high in the cortical depth direction even at large cortical distances. Taken together, these observations indicate that motion preferences were more similar in the cortical depth than the cortical plane direction and often remained similar over large cortical distances in the cortical depth direction.

SI Discussion

Physical bottom-up and perceptual lateral and top-down processes. The observed signal modulations that reflected the conscious percept were based on average cluster responses. It is therefore possible that, although as a whole the cluster showed modulation in the expected direction, a subgroup of voxels did not (or only to a lesser degree) modulate its signal with the percept. Indeed, when we compared amplitude modulations for ambiguous and physical motion (Fig. 2, Table S2), we found larger modulations in response to the latter. One explanation for this difference is that the physical motion stimulus gives rise to both physical bottom-up and perceptual lateral and top-down processes, while the ambiguous motion display invokes primarily perceptual lateral and top-down processes. Earlier electro-physiology findings that only about 40% of identified MT neurons modulated their spiking with the dominant motion direction in binocular rivalry (43) are in line with this interpretation. One idea inspired by animal physiology (44, 45) is that bottom-up and lateral / top-down processes recruit different laminar parts of a column. This idea could account for our observation that preferences during ambiguous motion displayed less consistent columnarity than during physical motion. To fully illuminate this possibility, one would need to distinguish the specific laminar contributions to activity in a column, which is beyond the scope of this paper and subject to future studies.

Functional significance of columnar organization. The functional significance of columnar organization is currently debated. Some authors have argued that cortical columns are structures devoid of function (46) based on the observation that no difference in visual performance is apparent between species that exhibit columnar organization in visual areas and those that do not. By contrast, other authors have successfully used columns as a fundamental functional unit to account for information processing in cortical computations (47). The column offers a way to integrate feedforward and feedback information processing

Marian Schneider, Valentin Kemper, Thomas Emmerling, Federico De Martino and Rainer Goebel

around a computational feature. For example, feedforward and feedback processing can be integrated for a particular orientation in V1 or a particular frequency in A1, which offers an efficient manner to process attentional or contextual effects (48). We think that the best way to advance the debate is by performing studies that map out columnar structures in species that have them and to test their relation to perception and cognition. In the current study, we did not pre-suppose columnar organization. However, when mapping horizontal and vertical preferences in response to physical motion, we found them to be organized in a columnar fashion. These distinct columnar structures appear to be relevant for perception as each of them is involved in tracking a conscious percept of a distinct particular motion axis.
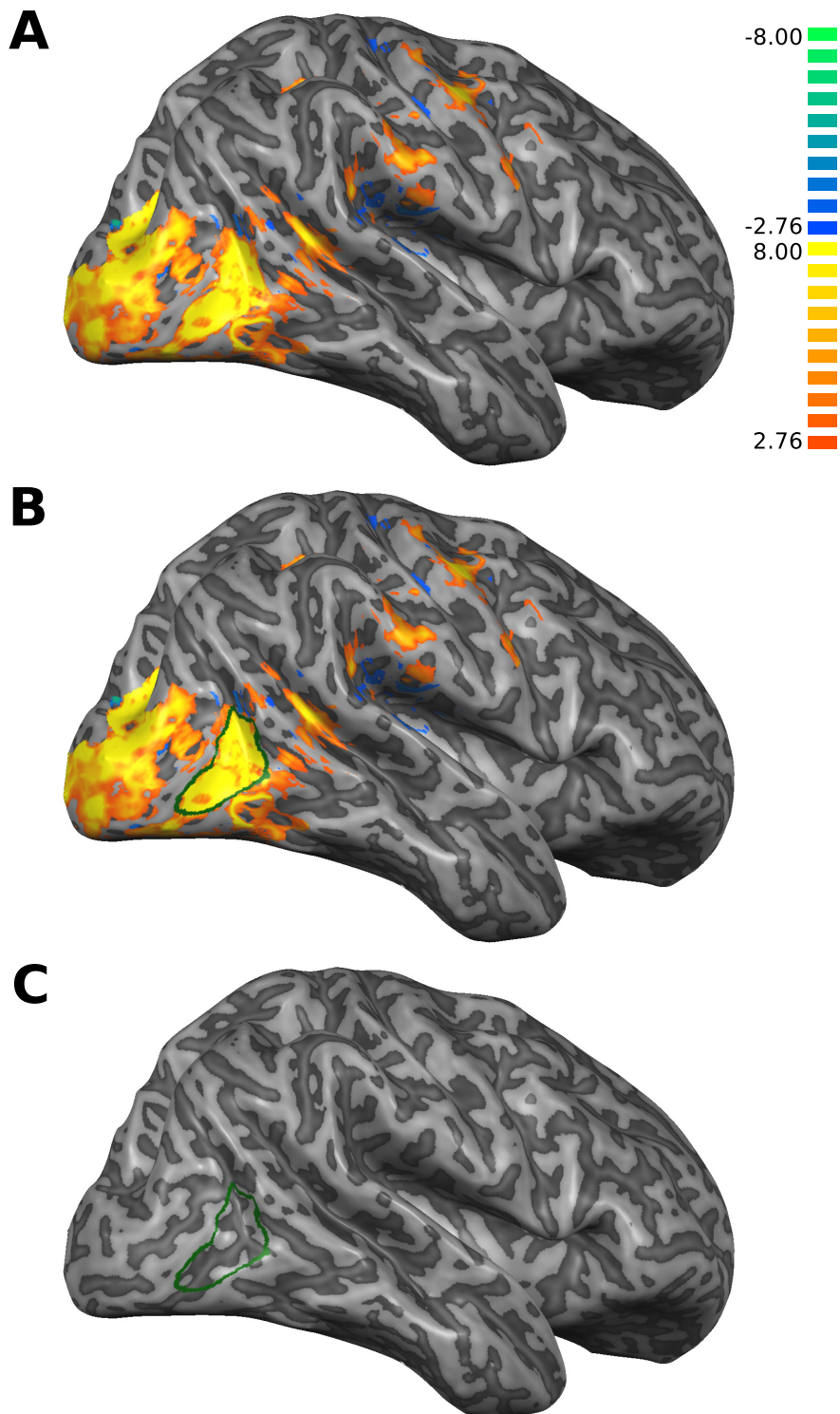
Fig. S1. Example of delineation process for area hMT+. Displayed is the smoothed and inflated right hemisphere of one representative subject (S3). Dark-gray color indicates sulci and light-gray color indicates gyri. (A) Statistical map of responses (t-values) to moving stimuli presented in central apertures (thresholded at q(FDR) < .05). (B) Delineation of area hMT+ (green) in the right hemisphere overlaid on top of the statistical map. Selection of area by manual drawing was clearly constrained by the activation map, leaving little room for subjective judgment. (C) Delineation of area hMT+ (green) overlaid without statistical map.
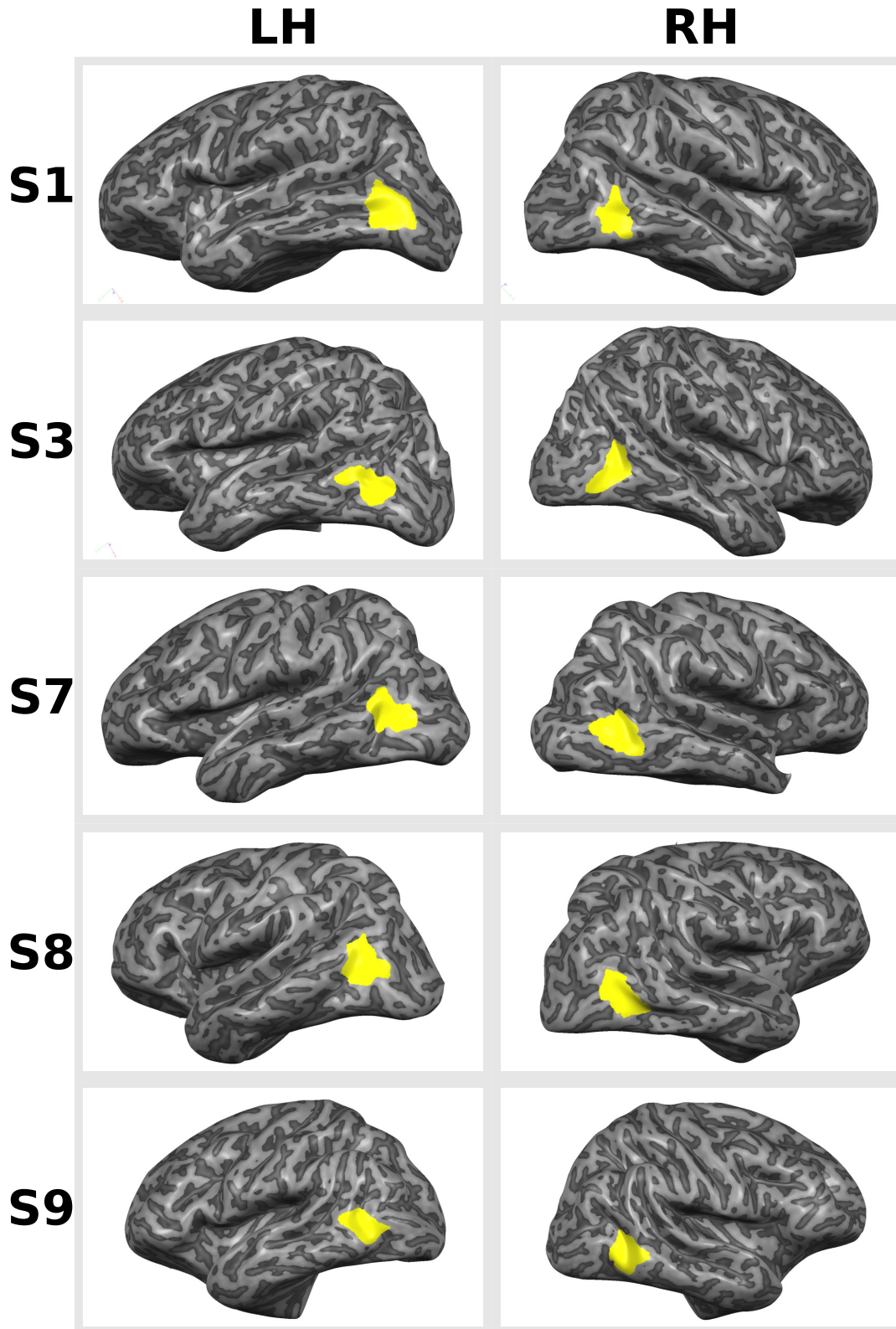
Marian Schneider, Valentin Kemper, Thomas Emmerling, Federico De Martino and Rainer Goebel

Fig. S2. Delineations of area hMT+ for all subjects. Displayed are the smoothed and inflated left (left side) and right (right side) hemispheres for all subjects (different rows: S1, S3, S7, S8, S9). Dark-gray color indicates sulci and light-gray color indicates gyri. The delineations of areas hMT+ are overlaid in yellow.
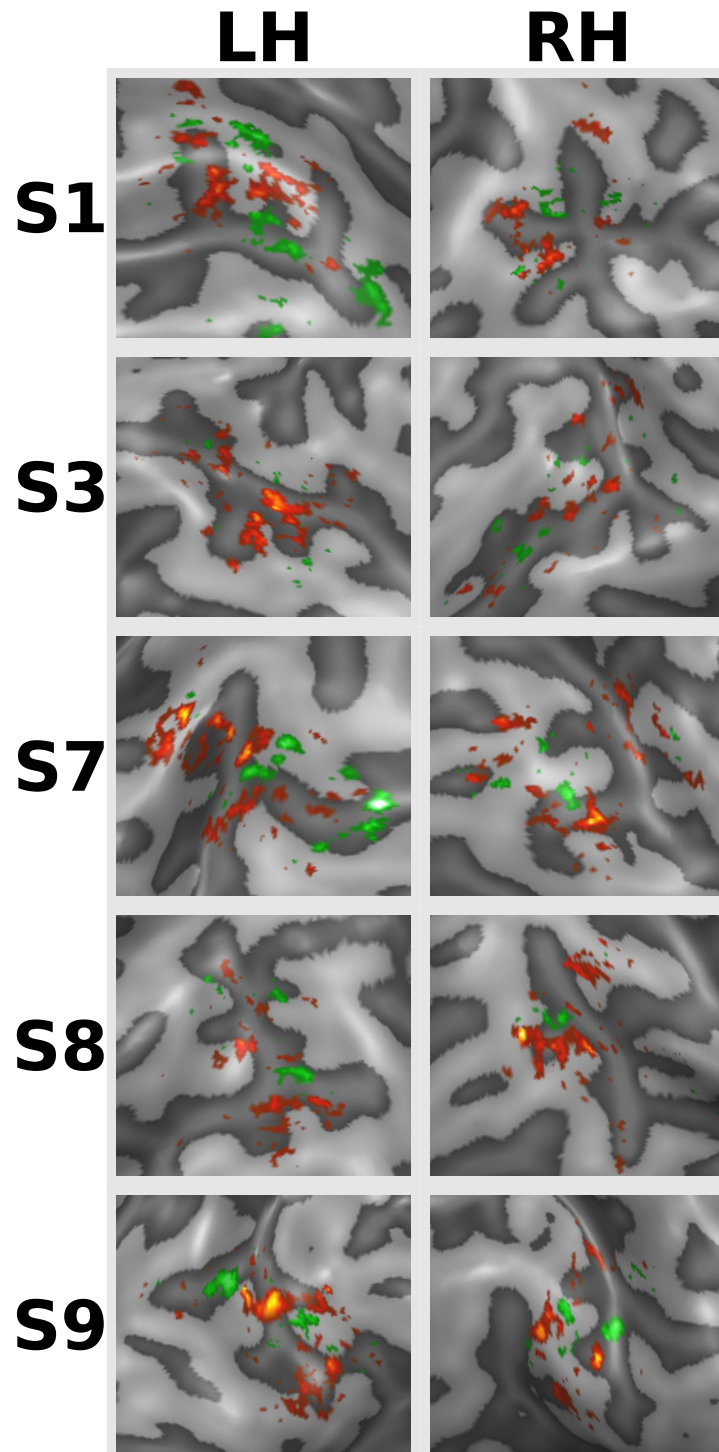
Fig. S3. Horizontal and vertical clusters for physical motion. Horizontal (red) and vertical (green) clusters determined by responses during the physical motion experiment are displayed on the smoothed and inflated left (left side) and right (right side) mid-GM reconstructions of area hMT+ for all subjects (different rows: S1, S3, S7, S8, S9). Dark-gray indicates sulci and light-gray indicates gyri.
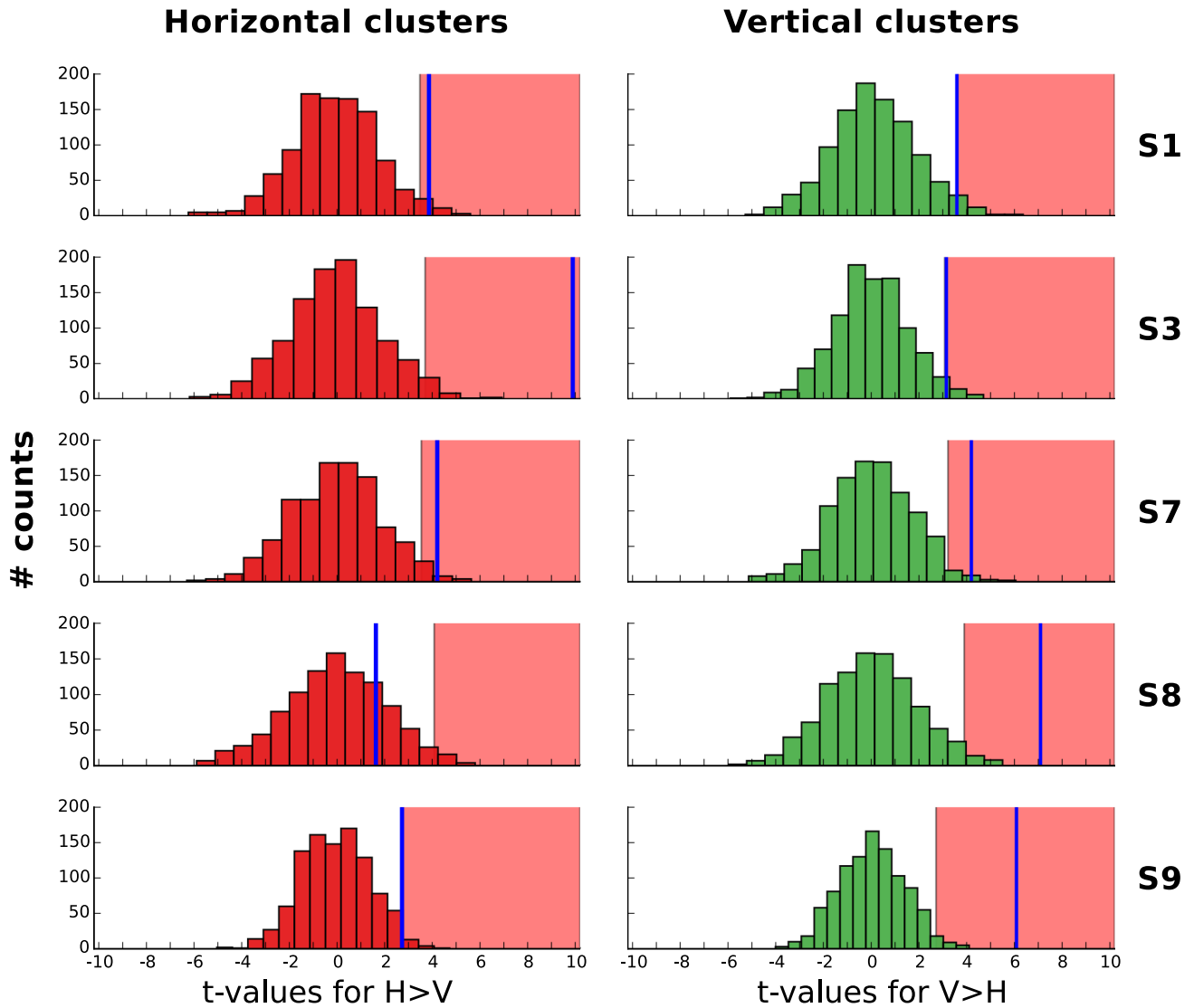
Fig. S4. Results for statistical significance testing of the amplitude modulations during ambiguous motion. The histograms represent a null distribution of t-values obtained by randomly permuting condition labels (1000-fold permutation testing) for either the horizontal (left side, red) or vertical (right side, green) cluster. Blue lines indicate the empirical t-value. Red shading indicates areas above the 97.5th percentile of the null population obtained with 1000-fold permutation testing (rejection area). "H" stands for horizontal, "V" for vertical. Each row represents the results from a different subject.
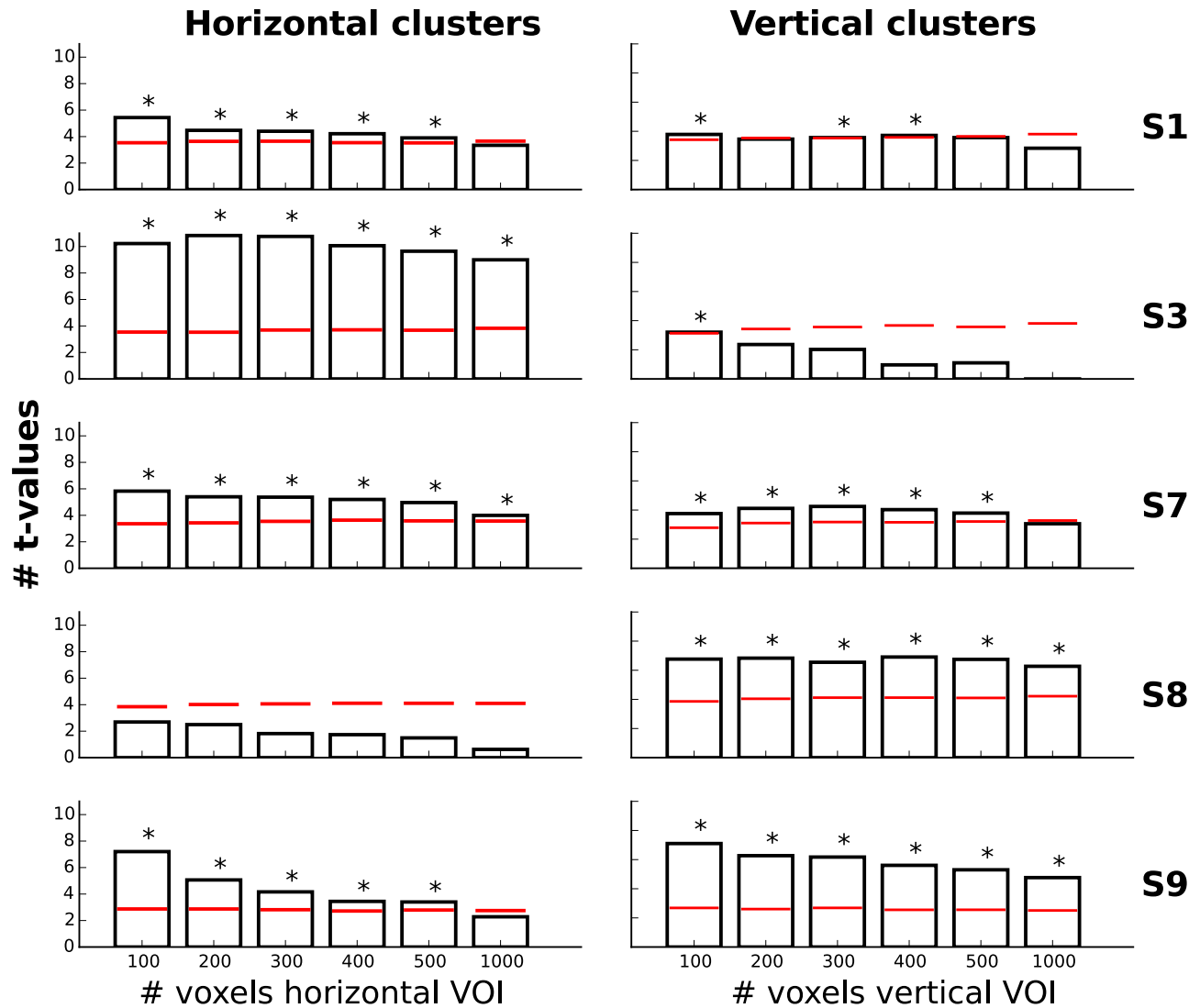
Horizontal clusters

Vertical clusters

Fig. S5. Observed effects of amplitude modulation during ambiguous motion are robust to changing numbers of voxels included in the ROI. Black bars indicate empirical t-values either for the horizontal (left side) or vertical (right side) cluster. Different bars indicate the results for a particular number of voxels included in the ROI (100, 200, 300, 400, 500, 1000). Red values indicate the 97.5th percentile of a null population obtained with 1000-fold permutation testing. Stars indicate that empirical values fell above the 97.5th percentile of the null distribution. Each row represents the results from a different subject.
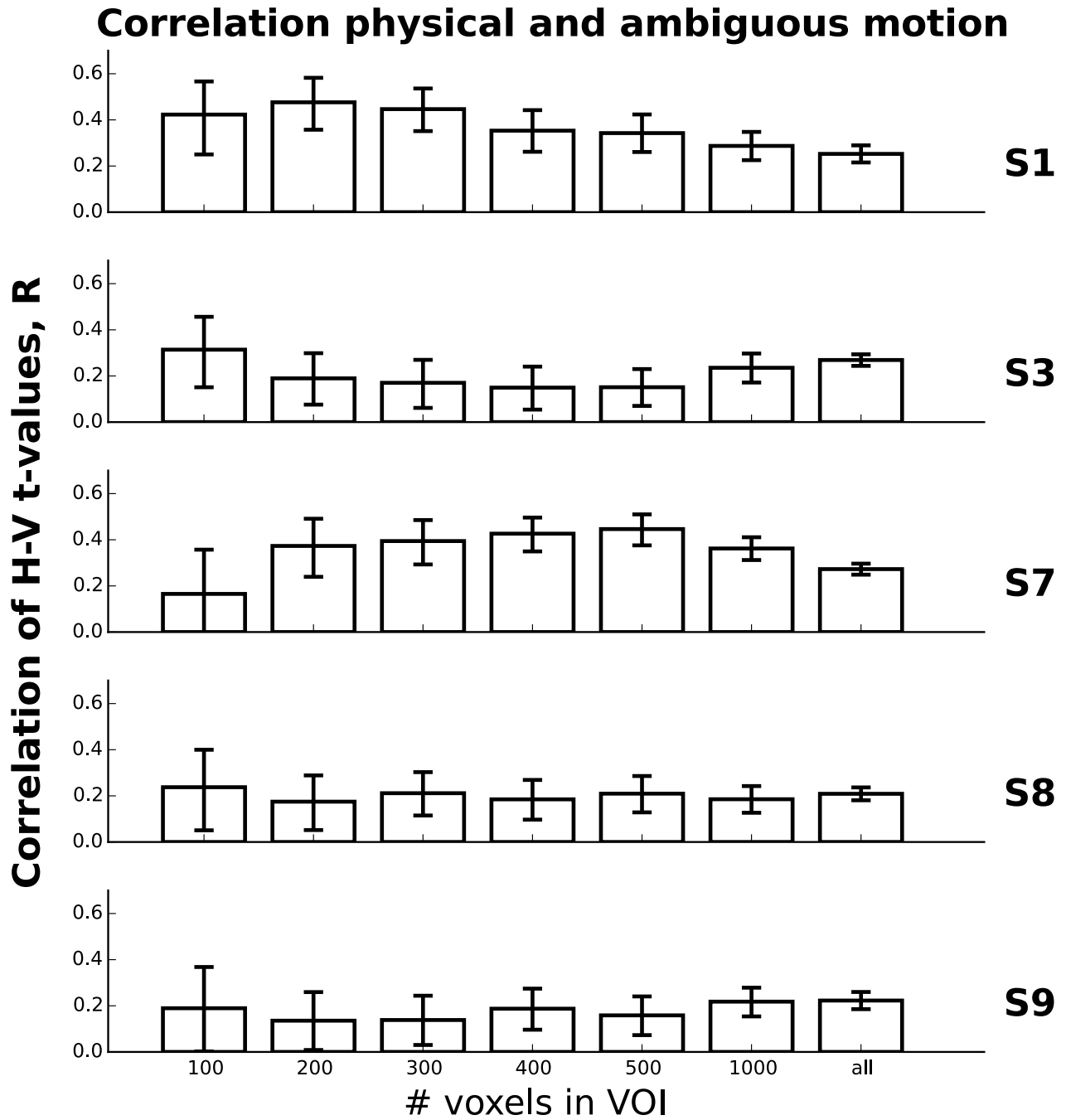
Marian Schneider, Valentin Kemper, Thomas Emmerling, Federico De Martino and Rainer Goebel

Correlation physical and ambiguous motion

Fig. S6. Consistency in voxel preference was robust to a varying number of voxels included in our ROI. Black bars indicate the median bootstrapped correlation coefficient between t-values for physical and ambiguous motion for a bootstrapped population of voxel values (20,000 re-samples). Error bars represent the 2.5th and 97.5th percentile of the bootstrapped correlation coefficients. Different bars indicate the results for a particular number of voxels included in the ROI (100, 200, 300, 400, 500, 1000). Each row represents the results from a different subject.
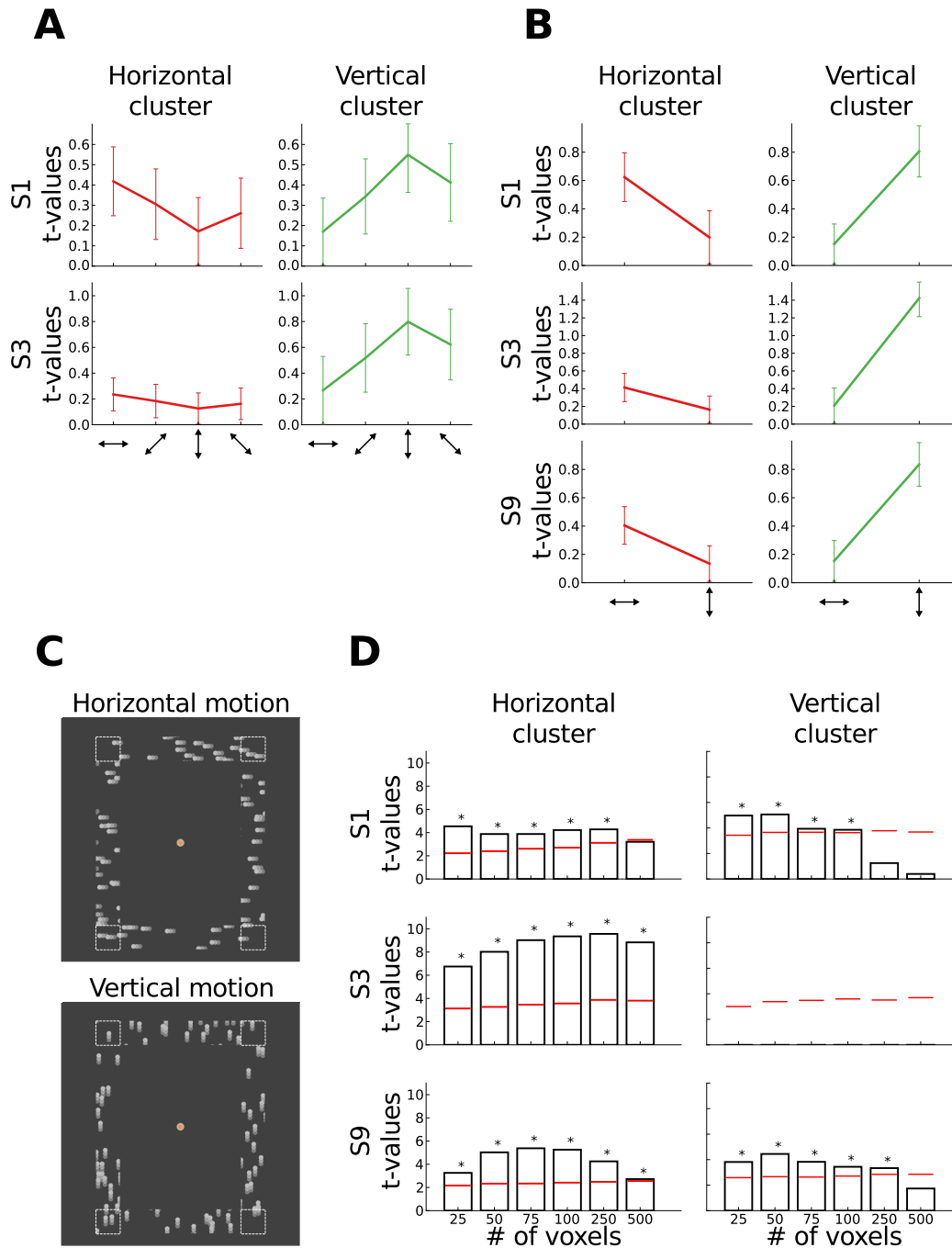
Fig. S7. Clusters show expected motion tuning. (A) Experimental results for control experiment I. Plots show axis-of-motion tuning curves for horizontal (left) and vertical (right) clusters, as used in the main experiment. Lines depict average t-values in response to four presented axes of motion. Error bars represent standard error of the mean. To facilitate visual comparison of tuning curves between clusters, all t-values were normalized such that the lowest response (including its standard error) in a given cluster was equal to zero. This did not change differences between t-values and this step was not performed for data entering statistical analyses. The clusters for the two control participants (S1 and S3, different rows) show expected motion tuning with highest responses to the preferred axis of motion and gradually lower responses to non-preferred axes. (C) Time-lapsed image showing sum of three frames for stimuli used in control experiment II. Retinotopically identical dot fields were moving either horizontally (upper row) or vertically (lower row). Visibility of dots was restricted to the aperture defined by the inducer squares and motion trajectories in the motion quartet experiments. White dashes indicate positions of inducer squares in the motion quartet experiments and were not shown during the actual experiment. (B) Experimental results for control experiment II. Same conventions as for panel (A), just that in this experiment only horizontal and vertical motion axes were presented and three subjects were recorded (S1, S3 and S9, different rows). Although the underlying data were obtained in independent experiments, the same tuning as in (A) was observed, with highest t-values to the preferred axis of motion. (D) Amplitude modulations during ambiguous motion when data from control experiment II were used to assign voxels to either the horizontal (left) or vertical (right) cluster. Black bars indicate empirical t-values. Different bars indicate results for different numbers of voxels included in the ROI (25, 50, 75, 100, 250, 500). Red bars indicate the 97.5th percentile of a null population obtained with 1000-fold permutation testing. Stars indicate empirical values above the 97.5th percentile of the null distribution. With the exception of one cluster, we can replicate the modulations observed in the main experiment, even though assignment of voxels was based on retinotopically identical conditions.
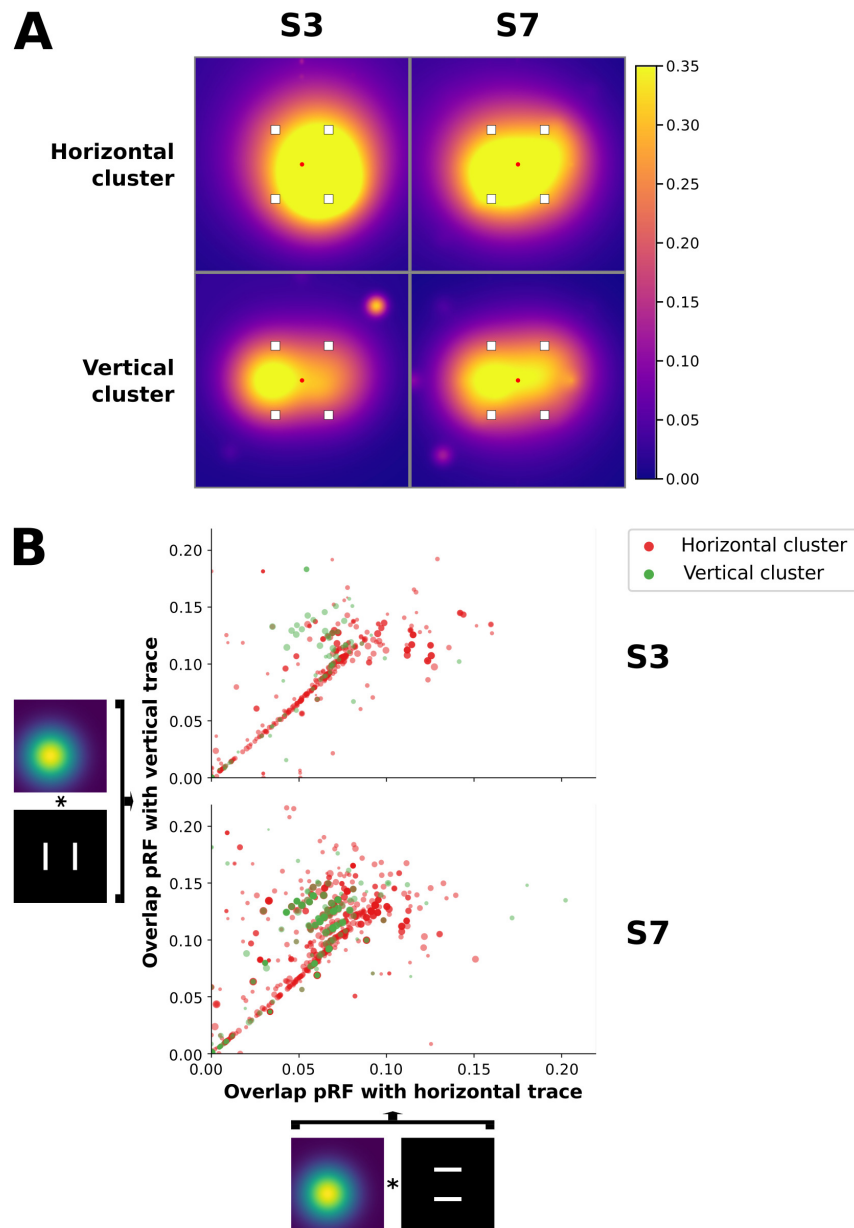
Marian Schneider, Valentin Kemper, Thomas Emmerling, Federico De Martino and Rainer Goebel

Fig. S8. Clusters show no difference in retinotopic preference. (A) Visual field coverage for either the horizontal (upper row) or vertical (lower row) cluster, as used in the main experiment, for subject S3 (left) and S7 (right). Coverage is computed as the average value across all voxels' population receptive fields (pRF) in a given cluster at every pixel of the visual field (theoretical maximum equal to 1). The subject-specific configuration of the motion quartet is overlaid on top to indicate where inducer squares were presented. Generally, clusters show coverage of the area between inducer squares and foveal to inducer squares. Both horizontal and vertical clusters show a retinotopic bias to vertical motion trajectories but retinotopic bias did not differ between horizontal and vertical clusters. (B) Calculated overlap between the pRF of every voxel in the horizontal (red) and vertical (green) cluster for the horizontal (x-axis) and vertical (y-axis) motion trajectories. Every dot represents a single voxel and dot size is scaled with the fitting accuracy of the pRF model. Results are shown for subject S3 (upper panel) and S7 (lower panel). Many voxels are close to or even on the diagonal, indicating similar or equal overlap with both motion trajectories. The bias towards vertical trajectories observed in (A) is apparent as a cluster of voxels above the diagonal but horizontal and vertical clusters show no difference in this bias.
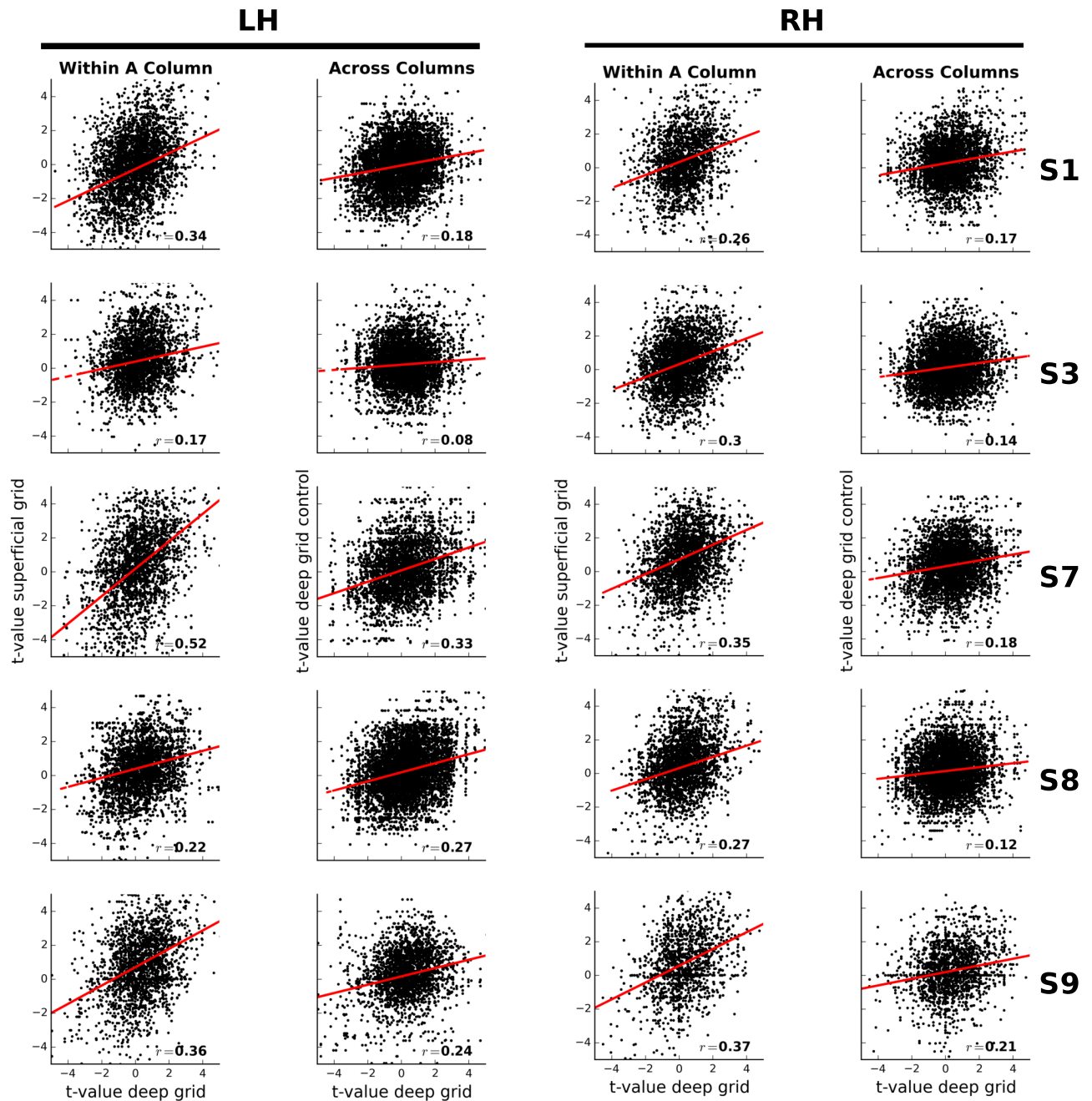
Fig. S9. Motion preferences were more stable in the direction of cortical depth than along the cortical surface. Results are shown for all subjects (rows) and two hemispheres (left hemisphere on the left). Scatter plots show the correlation of axis preference sampled for deep and corresponding superficial depth level ("within a column", respective left side) compared to the correlation for different nearby locations in deep depth level only ("across columns", respective right side).
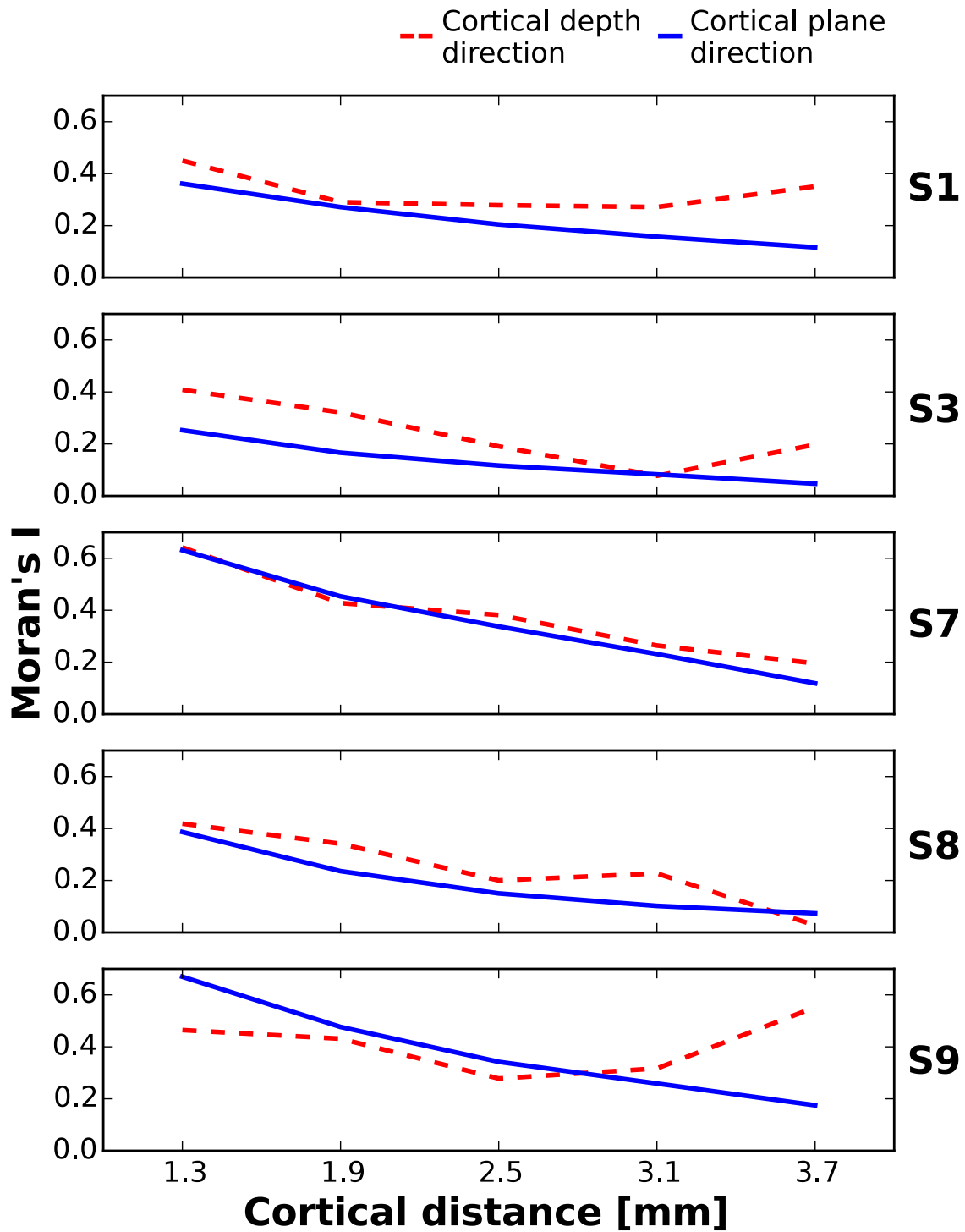
Marian Schneider, Valentin Kemper, Thomas Emmerling, Federico De Martino and Rainer Goebel

Fig. S10. Greater autocorrelation in cortical depth than in cortical plane direction. Moran's I, a measure of spatial autocorrelation, is plotted over cortical distance (in mm). The red, dashed line shows autocorrelation coefficients in cortical depth (columnar) direction; the blue, solid line shows coefficients in cortical plane direction. Each row shows the results from a different subject.
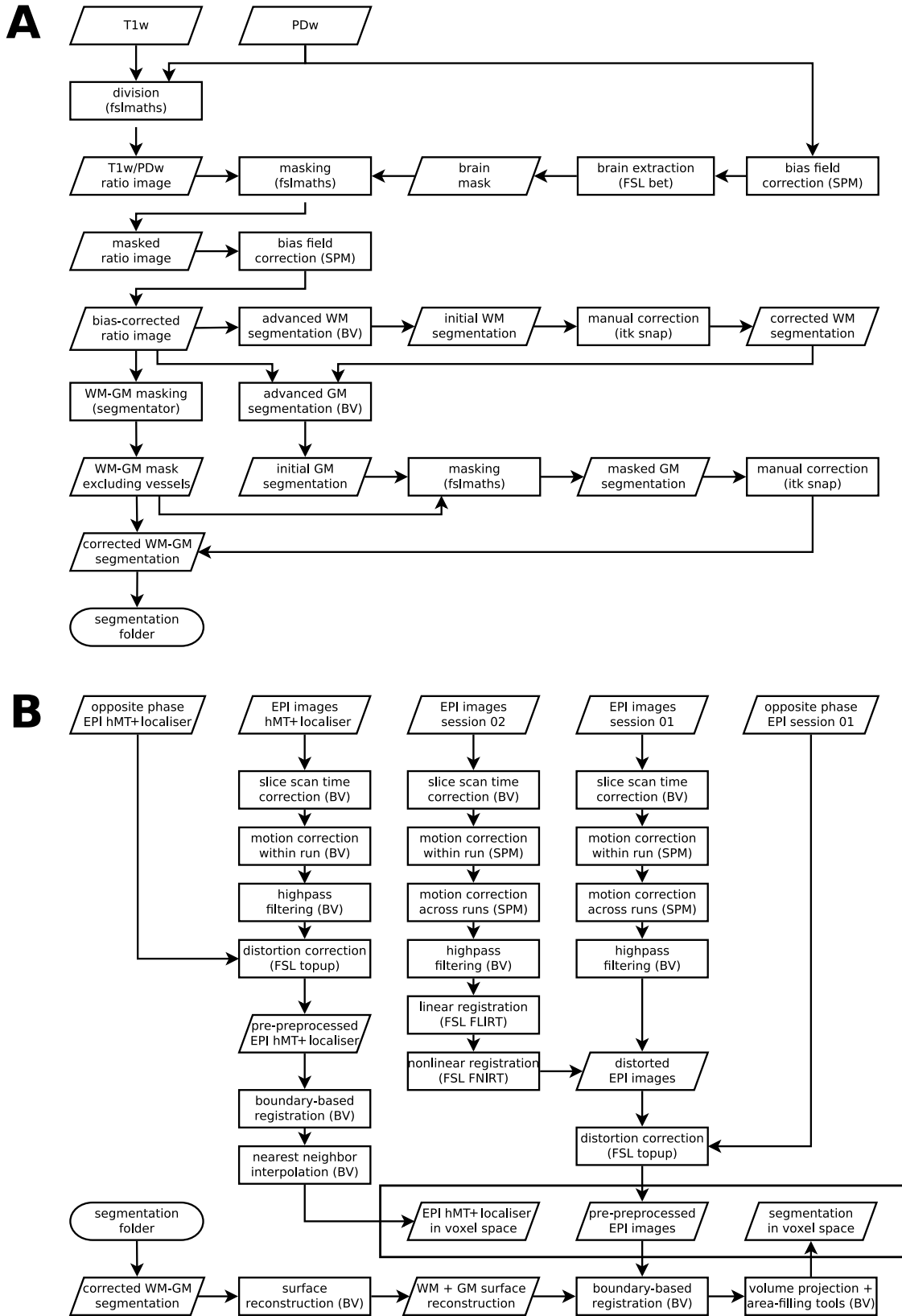
Fig. S11. Overview of pre-processing pipeline. (A) Pre-processing pipeline for structural images. (B) Pre-processing pipeline for functional images. Rectangular shapes represent processing steps, rhombic shapes represent input or outputs and cylindrical shapes represent input or output locations.
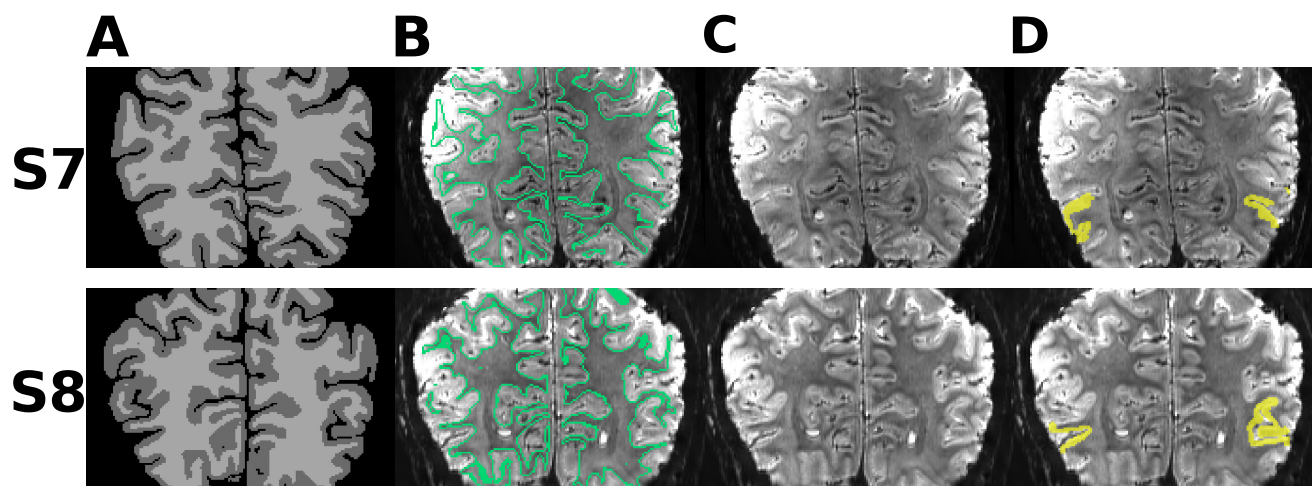
Marian Schneider, Valentin Kemper, Thomas Emmerling, Federico De Martino and Rainer Goebel

Fig. S12. Segmentation and co-registration quality. All images show coronal slices in radiological convention. (A) Segmentation image showing WM (light gray) and GM (dark gray). Although the segmentation has been down-sampled to the slightly lower resolution of the functional space (0.6 mm isotropic to 0.8 mm isotropic), desirable segmentation features are preserved due to our surface reconstruction pipeline. (B) Overlay of the WM-GM boundary (mint) on the mean functional image in voxel space. High congruence between projected WM-GM boundary (mint) and inherent contrast of functional image (lower intensity values in WM, higher intensities in GM) indicate little segmentation and co-registration error. (C) Mean functional image without segmentation overlay. (D) Mean functional image overlaid with ROI (yellow).

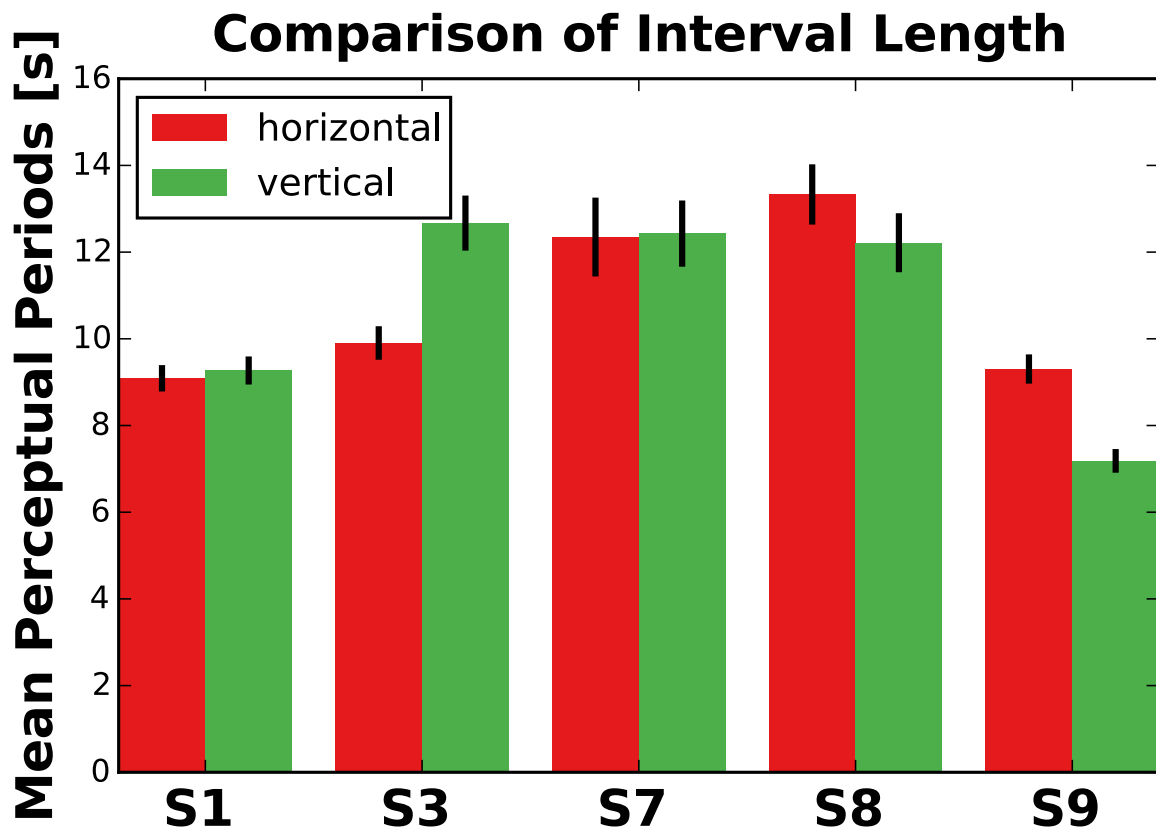**Fig. S13.** Average length of perceptual periods. Mean length of perceptual periods in seconds (s) during the ambiguous motion experiment are shown for every analyzed subject (S1, S3, S7, S8, S9), separately for horizontal (red) and vertical (green) perceptual periods. Error bars represent standard error across perceptual periods.

Marian Schneider, Valentin Kemper, Thomas Emmerling, Federico De Martino and Rainer Goebel

Table S1. Size of the hMT+ ROIs for every subject. Surface areas are in square mm. LH and RH indicate left and right hemisphere, respectively.

Subject	Number of voxels	Surface area LH	Surface area RH
S1	8341	855	517
S3	10267	781	783
S7	8394	710	857
S8	7923	720	632
S9	7361	626	720

Table S2. Event-related signal modulations during ambiguous and physical motion. Mean trough-to-peak differences $\pm$ standard error of the mean. Positive values indicate that signal increased from the minimum of the first two time points to the maximum of the last two time points. Negative values indicate a signal decrease from the maximum of the first two time points to the minimum of the last two time points. H and V cluster represent the horizontal and vertical clusters defined in the main experiment. H and V percept indicate time periods during which subjects indicated a horizontal or vertical percept. S1, S3, S7, S8 and S9 indicate the different subjects.

		Ambiguous Motion		Physical Motion	
		H percept	V Percept	H percept	V percept
S1					
	H cluster	1.07±0.17	-0.42±0.26	1.92±0.20	-1.23±0.21
	V cluster	-0.15±0.17	0.68±0.22	-0.50±0.25	1.73±0.16
S3					
	H cluster	1.27±0.16	-0.89±0.15	1.85±0.18	-1.08±0.16
	V cluster	0.11±0.15	0.30±0.13	-0.96±0.26	1.51±0.21
S7					
	H cluster	0.89±0.16	-0.65±0.19	2.09±0.12	-1.23±0.15
	V cluster	-0.37±0.15	0.46±0.18	-1.13±0.19	1.91±0.20
S8					
	H cluster	0.77±0.17	-0.43±0.18	1.82±0.12	-0.97±0.12
	V cluster	-0.12±0.14	0.29±0.20	-0.79±0.15	1.42±0.13
S9					
	H cluster	0.24±0.13	-0.75±0.16	2.11±0.13	-0.98±0.13
	V cluster	-0.53±0.14	0.79±0.16	-0.66±0.15	2.14±0.12

Marian Schneider, Valentin Kemper, Thomas Emmerling, Federico De Martino and Rainer Goebel

Table S3. Average cortical thickness of hMT+ ROIs for every subject. LH and RH indicate left and right hemisphere, respectively. 2H means that both hemispheres were included. Std indicates standard deviation.

Subject	$Mean \pm std$ thickness LH	$Mean \pm std$ thickness RH	$Mean \pm std$ thickness 2H
S1	2.74 ± 0.55	2.73 ± 0.66	2.74 ± 0.59
S3	2.72 ± 0.56	2.84 ± 0.65	2.77 ± 0.61
S7	2.78 ± 0.56	2.82 ± 0.59	2.80 ± 0.57
S8	2.58 ± 0.53	2.53 ± 0.52	2.56 ± 0.53
S9	3.35 ± 1.18	2.47 ± 0.46	2.97 ± 1.04

Table S4. Subject-specific vertical distances for the motion quartet display. Distances are in degrees of visual angle from the fixation point to the center of the square in the motion quartet. Horizontal distance was kept constant at 3 degrees of visual angle.

Subject	Vertical distance
S1	3.8
S3	3.9
S7	3.9
S8	3.9
S9	3.8

Marian Schneider, Valentin Kemper, Thomas Emmerling, Federico De Martino and Rainer Goebel

Movie S1. The movie visualizes the dynamic amplitude modulations during ambiguous motion for an exemplary subject (S1) in response to changing horizontal and vertical motion perception. The bistable motion stimulus is shown in the inset in the upper right corner. The conscious percept at any moment is indicated by the inset on the upper left. Horizontal clusters are shown in red, vertical clusters in green.

Additional data table S1 (motion_quartet_data.zip)

Supplementary data consist of structural and functional MRI images for the five analyzed participants. The data are provided as compressed nifti files and are organized according to the Brain Imaging Data Structure (BIDS). The data is deposited in a zenodo repository.

References

1. Peirce JW (2007) PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods* 162:8–13.
2. Peirce JW (2008) Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics* 2.
3. Huk AC, Dougherty RF, Heeger DJ (2002) Retinotopy and functional subdivision of human areas MT and MST. *The Journal of Neuroscience* 22(16):7195–7205.
4. Amano K, Wandell BA, Dumoulin SO (2009) Visual field maps, population receptive field sizes, and visual field coverage in the human MT+ complex. *Journal of Neurophysiology* 102(5):2704–18.
5. Emmerling TC, Zimmermann J, Sorger B, Frost MA, Goebel R (2016) Decoding the direction of imagined visual motion using 7 T ultra-high field fMRI. *NeuroImage* 125(10):61–73.
6. Marquardt I, Schneider M, Gulban OF, Ivanov D, Uludağ K (2018) Cortical depth profiles of luminance contrast responses in human V1 and V2 using 7 T fMRI. *Human Brain Mapping* 39(7):2812–2827.
7. Ramachandran VS, Anstis SM (1985) Perceptual organization in multistable apparent motion. *Perception* 14(2):135–143.
8. Finlay D, von Grünau M (1987) Some experiments on the breakdown effect in apparent motion. *Perception & Psychophysics* 42(6):526–534.
9. 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).
10. Wang HX, Merriam EP, Freeman J, Heeger DJ (2014) Motion Direction Biases and Decoding in Human Visual Cortex. *Journal of Neuroscience* 34(37):12601–12615.
11. Kolster H, Peeters R, Orban Ga (2010) The retinotopic organization of the human middle temporal area MT/V5 and its cortical neighbors. *The Journal of Neuroscience* 30(29):9801–20.
12. Tyler CW, et al. (2005) Extended Concepts of Occipital Retinotopy. *Current Medical Imaging Reviews* 1:319–329.
13. Chaudhuri A, Glaser DA (1991) Metastable Motion Anisotropy. *Visual Neuroscience* 7(5):397–407.
14. Gen E, Bergmann J, Singer W, Kohler A (2011) Interhemispheric connections shape subjective experience of bistable motion. *Current Biology* 21(17):1494–1499.
15. Moeller S, et al. (2010) Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain FMRI. *Magnetic Resonance in Medicine.*
16. Setsompop K, et al. (2012) Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magnetic Resonance in Medicine.*
17. Feinberg DA, et al. (2010) Multiplexed echo planar imaging for sub-second whole brain fmri and fast diffusion imaging. *PLoS ONE.*
18. Van de Moortele PF, et al. (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.
19. Ashburner J, Friston KJ (2005) Unified segmentation. *NeuroImage* 26(3):839–851.
20. Yushkevich PA, et al. (2006) User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage* 31(3):1116–1128.
21. Smith SM (2002) Fast robust automated brain extraction. *Human Brain Mapping* 17(3):143–155.
22. Jones E, Oliphant T, Peterson P, , et al. (2007) SciPy: Open source scientific tools for Python. *Computing in Science and Engineering.*
23. Gulban OF, Schneider M, Marquardt I, Haast RAM, De Martino F (2018) A scalable method to improve gray matter segmentation at ultra high field MRI. *PloS one.*
24. Kriegeskorte N, Goebel R (2001) An efficient algorithm for topologically correct segmentation of the cortical sheet in anatomical MR volumes. *NeuroImage* 14(2):329–346.
25. Friston KJ (2006) *Statistical Parametric Mapping: The Analysis of Functional Brain Images.* p. 656.
26. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM (2012) FSL. *NeuroImage* 62(2):782–790.
27. Van Der Walt S, Colbert SC, Varoquaux G (2011) The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering* 13(2):22–30.
28. Hunter JD (2007) Matplotlib: A 2d graphics environment. *Computing In Science & Engineering* 9(3):90–95.
29. Andersson JL, Skare S, Ashburner J (2003) How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage* 20(2):870–888.
30. Smith S, et al. (2004) Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage.*

31. Greve DN, Fischl B (2009) Accurate and robust brain image alignment using boundary-based registration. *NeuroImage* 48(1):63–72.
32. Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain images. *Medical Image Analysis* 5(2):143–156.
33. Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* 17(2):825–841.
34. Andersson JLR, Jenkinson M, Smith S (2007) Non-linear registration, aka spatial normalisation. FMRIB Technial Report TR07JA2., Technical Report June.
35. Dumoulin SO (2000) A New Anatomical Landmark for Reliable Identification of Human Area V5/MT: a Quantitative Analysis of Sulcal Patterning. *Cerebral Cortex* 10(5):454–463.
36. Emmerling T (2016) Ph.D. thesis (Maastricht University).
37. Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex. *NeuroImage* 39(2):647–660.
38. Schneider M, Marquardt I, Gulban OF, Isik I (2018) Pyprf feature - population receptive field estimation with feature weights. *https://doi.org/10.5281/zenodo.1401465*.
39. Schneider M (2018) Vificov - visual field coverage visualization in python. *https://doi.org/10.5281/zenodo.1401468*.
40. Moran PA (1950) Notes on continuous stochastic phenomena. *Biometrika*.
41. Rosenberg MS (2010) The Bearing Correlogram: A New Method of Analyzing Directional Spatial Autocorrelation. *Geographical Analysis* 32(3):267–278.
42. Oden NL, Sokal RR (1986) Directional Autocorrelation - an Extension of Spatial Correlograms to 2 Dimensions. *Systematic Zoology*.
43. Logothetis NK (1989) Neuronal correlates of subjective visual perception. *Science* 245(4919):761–763.
44. Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* 1(1):1–47.
45. Markov NT, et al. (2014) Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. *Journal of Comparative Neurology* 522(1):225–259.
46. Horton JC, Adams DL (2005) The cortical column: a structure without a function. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360(1456):837–862.
47. Bastos AM, et al. (2012) Canonical Microcircuits for Predictive Coding. *Neuron* 76(4):695–711.
48. De Martino F, et al. (2015) Frequency preference and attention effects across cortical depths in the human primary auditory cortex. *Proceedings of the National Academy of Sciences* 112(52):16036–16041.