Local Motion Is Sufficient for Human Body Movement Perception in High-Level Visual Cortex

Baichen Li[1], Rufin Vogels[2], Giuseppe Marrazzo[1], Marta Poyo Solanas[1], and Beatrice de Gelder[1,]

[1] Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht 6200 MD, The Netherlands

[2] Leuven Brain Institute, KU Leuven, Leuven 3000, Belgium

Correspondence: *Beatrice de Gelder <b.degelder@maastrichtuniversity.nl>*

Abstract

Perceiving the movements of other living organisms is a fundamental ability across species. Although several cortical regions are selective for static body features, it is unclear whether these areas also encode motion cues independently of body shapes. Here, we isolated local motion signals from naturalistic body movements using synthesized videos that eliminate explicit body shape information. We demonstrate that local motion alone is sufficient for perceiving human body movement. Using ultra-high field fMRI, we found that three body-motion-related areas, fusiform gyrus (FG), lateral occipitotemporal cortex (LOTC), and posterior superior temporal sulcus (pSTS), were sensitive to both the amount and the spatiotemporal structure of local motion. Critically, the local motion effects were significantly correlated to voxel-wise body motion selectivity in FG and LOTC, but not in pSTS. These findings identify local motion as a distinct and behaviorally sufficient cue for body perception, which can be encoded in FG and LOTC in parallel with body shape processing.

Introduction

Perceiving the movements of other living organisms is a fundamental survival skill shared across species. Extensive research has shown that human observers can readily recognize others' intentions, actions, and emotions from brief whole-body movements (De Gelder, 2006; De Meijer, 1989; Roether, Omlor, Christensen, & Giese, 2009; Wallbott, 1998). They can do so even from highly simplified kinematic stimuli, such as point-light displays (PLDs) composed of a few dots representing the movements from major joints of an actor (Brownlow, Dixon, Egbert, & Radcliffe, 1997; Cutting & Kozlowski, 1977; Dittrich, Troscianko, Lea, & Morgan, 1996; Johansson, 1973; Roether et al., 2009). There is a broad consensus that body movement perception engages the cortical regions involved in body processing. These include the extrastriate body area (EBA; Downing, Jiang, Shuman, & Kanwisher, 2001; Downing, Peelen, Wiggett, & Tew, 2006; Jastorff & Orban, 2009; Taylor, Wiggett, & Downing, 2007; Vangeneugden, Peelen, Tadin, & Battelli, 2014), the fusiform body area (FBA; Jastorff & Orban, 2009; Peelen & Downing, 2005; Taylor et al., 2007), and the posterior superior temporal sulcus (pSTS; Grossman & Blake, 2002; Jastorff & Orban, 2009; Vangeneugden et al., 2014). Despite this progress, it remains unclear whether motion related activations in these regions primarily reflect the temporal dynamics of global body shape, or the integration of local, low-level coherent motion signals.

Evidence in support of a shape-based account comes from studies demonstrating that EBA and FBA respond more strongly to videos composed of frame sequences randomly drawn from different actions than to videos depicting intact actions, whereas the motion-selective area, middle temporal cortex (MT), shows no such effect (Downing et al., 2006). Similarly, Michels et al. (2005) compared classical PLDs with sequential-position PLDs (Beintema & Lappe, 2002), which eliminate local image motion while preserving form-based structural information. Their results revealed robust activations in the FG and pSTS for sequential-position PLD videos, regardless of whether the underlying body structure was moving or static (Michels et al., 2005). Another fMRI study found that voxel-wise selectivity for PLD body movements in pSTS and EBA was more correlated with selectivity for static body images than with selectivity for coherent dot motion (Peelen, Wiggett, & Downing, 2006). Taken together, these findings suggested that body-selective regions rely predominantly on shape-based information ("motion-

from-shape"; Beintema & Lappe, 2002), with low-level motion contributing only to the segmentation of body structure from background noise.

However, it would be premature to conclude that the low-level motion does not contribute to cortical processing of body movements. Giese and Poggio (2003) proposed an alternative model where body movements are extracted from the local motion sequence. Instead of analyzing the body shapes, the model starts with local motion detectors which respond to the coherent motion within small receptive fields and develops into a hierarchy to analysis complex motion fields or optical flow (OF) patterns. Similar mechanisms have been proposed in the field of general motion perception, termed as "first-order / low-level motion" (Z.-L. Lu & Sperling, 2001), "surface motion" (Bigelow, Kim, Namima, Bair, & Pasupathy, 2023), or "short-range / local motion" (Braddick, 1974; Hedges et al., 2011). Neurophysiology studies in macaques have revealed distinct brain regions for low-level motion and shape-based motion analysis (Bigelow et al., 2023; Hedges et al., 2011), as well as the encoding of low-level motion features in the dorsal bank of anterior STS body patches (Raman et al., 2023). A recent framework highlighted a motion-driven pathway for social perception through MT and STS (Pitcher, 2025; Pitcher & Ungerleider, 2021). These results raise the question about the roles of pSTS, EBA and FBA in low-level motion analysis. The involvement of pSTS is reported for body movement perception (Giese & Poggio, 2003; Grossman & Blake, 2002; Jastorff & Orban, 2009; Pitcher, 2025; Pitcher & Ungerleider, 2021; Vangeneugden et al., 2014), but the functions of EBA and FBA are still under debate. Human fMRI studies have revealed substantial anatomical and functional overlap between EBA and MT, including voxels that are selective for body parts as well as for coherent motion (Ferri, Kolster, Jastorff, & Orban, 2013). Similarly, joint representation of motion and body shape has been reported in the EBA-homologous body patches of macaque brain (Bognar, Nejad, Rens, Raman, & Vogels, 2025). Thus, low-level motion modulations are potentially expected also in EBA. In contrast to the other two regions, FBA had less prominent motion encoding both in human (Robert, Ungerleider, & Vaziri-Pashkam, 2023) and in monkey homologous regions (Bognar et al., 2025). Yet empirical evidence for low-level motion modulations in these body-selective regions in human brain remains very limited. Determining whether motion- and shape-based representations depend on shared or dissociable neural mechanisms is therefore crucial for refining the hierarchical models of cortical processing underlying body-movement perception.

The present study investigated the contribution of low-level motion processing to cortical regions for body processing. Regions of interest (ROIs) were defined based on responses to PLD body movements. We then assessed how these ROIs were modulated by synthesized stimuli specifically designed to isolate local motion cues. These stimuli preserved the spatial distribution of transient, pixel-wise coherent motion extracted from naturalistic body movements while eliminating all explicit body shape information. Critically, these stimuli enabled precise control over both the amount of local motion and its spatiotemporal consistency. Using ultra–high-field fMRI, we examined how local-motion properties influence temporal integration in body-selective cortex and how these effects relate to voxelwise selectivity for static body shapes, PLD body movements, and coherent dot motion. We aimed to adjudicate between two competing functional accounts of the role of low-level motion in body-movement perception: (1) low-level motion only supplements the recovery of body shape sequences (Beintema & Lappe, 2002; Koerfer & Lappe, 2020), or (2) the spatiotemporal structure of local motion itself is sufficient for accurate action recognition (Giese & Poggio, 2003). If local-motion effects reflect shape-sequence analysis, the effects should correlate more strongly with body shape selectivity (Peelen et al., 2006). Alternatively, if local motion provides independent diagnostic information, the effects should be more closely associated with motion-related selectivity.

Methods

Participants

Thirteen healthy participants (age = 21.31 ± 2.66 years; 4 males, all right-handed) participated in the experiment. All participants had normal or corrected-to-normal vision and no medical history of any psychiatric or neurological disorders. During the recruitment, the original body movement videos (Supplementary Video 1) were provided to the potential participants. Only participants who could clearly differentiate the walking directions were finally included into the experiment. All participants provided informed written consent before the start of the experiment and received a monetary reward (vouchers) or course credits for their participation. The experiment was approved by the Ethical Committee at Maastricht University and was performed in accordance with the Declaration of Helsinki.

Main experiment

Stimuli

Original animation and optical flow extraction

The original video was downloaded and edited from Adobe's Mixamo animation library (https://www.mixamo.com/). The video depicted a human actor moving to the right while keeping the upper body facing forward ("Jog Strafe Right"). We chose this specific strafe animation because it can also be played backward to present a leftward moving pattern. Thus, the forward- and backward-played videos contain the same set of body postures, while the moving direction can only be inferred from the motion features. The full video presented three walking cycles in 3.6 seconds at a frame rate of 30 Hz.

The original animation was further edited in Blender (https://www.blender.org/), including removing the global translation, reducing keyframes, and enlarging the body area. A rendered video is shown in Supplementary Video 1 with a frame rate of 30 fps. After the editing, the framewise optical flow (OF) patterns were extracted during rendering using a third-party plugin in Blender (Cartucho, Tukra, Li, S. Elson, & Giannarou, 2021). For each frame, the motion information was removed from 50% of the pixels with the lowest speed. This ensured that only the most salient local motion was preserved in the OF patterns.

To reduce the segmentation of body shape from motion against the static background, random background motion was added to each OF pattern. For each individual OF pattern, the original body area was first binarized and then filtered with a 2D low-pass filter with a cutoff spatial frequency of 2 cycles/image. The filtered image then served as a probability map for adding random motion, resulting in a noise distribution with higher density closer to the body area and lower density in the periphery of the visual field. The body OF was then overlaid on top of the noise motion and used as a reference to synthesize the body motion videos.

Video synthesizing procedure

The synthesized videos were generated by continuously applying the OF patterns frame by frame (Figure 1). Each video started from a salt-and-pepper noise image, and each consecutive frame was generated by applying the corresponding OF pattern as the pixel-wise displacement map to

the current frame. Importantly, the low-level motion was constrained to the static body area in each frame, so that the generated motion reflected only local pixel movements rather than actual body-part displacements. The high spatial frequency components (> 55 cycles / image) were removed from each generated frame to reduce the sharp boundary of each pixel.

We created different experimental conditions by modifying the reference OF patterns. The first configuration controlled the amount of presented body motion (Supplementary Figure 1). Instead of applying the full sequence of the walking OFs, framewise OFs were applied every 2, 4, or 6 frames, interleaved with the noise motion OF described above. Thus, instead of a continuous body motion flow, each video consisted of a discrete sequence of transient local motion segments. At an overall frame rate of 30 Hz, the temporal frequency of body motion presentation became 15 Hz, 7.5 Hz, and 5 Hz.

Secondly, the body-motion optical flows were presented either in their original form or after reversing the pixel-velocity directions by 180°, while preserving their magnitudes and spatial layouts (Supplementary Figure 2). Unlike the original flows, the reversed flows disrupted the spatiotemporal continuity of motion across successive local motion segments. We hypothesized that such incongruence may further impair the integration of a local motion sequence into global motion perception. Moreover, because the spatial distribution of moving pixels remained the same between congruent and incongruent versions, the sequence of motion-induced shape was well-controlled. This configuration, therefore allowed us to isolate local motion modulation while controlling for residual shape-induced motion.

Finally, the synthesized frames were combined in either forward or backward-played order. Since the original animation depicted a rightward walking pattern, the backward-played version showed a leftward direction instead. Except for this walking direction difference, the backward videos with congruent OF motion contained the same set of motion segments as the forward videos with incongruent OF motion. Thus, by pooling forward and backward trials, the effects of motion direction were balanced between congruent and incongruent OF conditions.

All synthesized videos were resized to 1200 x 1200 pixels underneath a circle aperture spanning 10 degrees in visual angle. The overall body area was 7 x 5 degrees in visual angle. Examples of synthesized videos are shown in Supplementary Video 2.

Experimental design

The experiment had 2 (congruent/incongruent) x 3 (15 Hz/7.5 Hz/5 Hz) = 6 conditions, presented with an event-related design. Each trial started with a video of one condition playing for 3.6 s, either forward-played or backward-played, followed by a jittered intertrial interval of 10.7 s, 12 s, or 13.3 s. The participants were instructed to watch the videos and indicate the walking direction after each video by a button press. A white "+" fixation (0.5 degree visual angle) was overlaid on the center of the display throughout the experiment. It turned to green after each video as a signal for the participant to respond and turned back to white after the button press was received. No feedback was given regarding the response correctness during the experiment.

For each main experimental run, each condition was repeated eight times (four forward-played and four backward-played) and presented in a pseudorandom order, resulting in 48 trials spanning thirteen minutes. Each participant finished three main experimental runs, resulting in 144 trials (24 repetitions per condition) in total.

Localizer experiment

Stimuli

PLD body motion videos were generated from five different whole-body movements, adapted from Mixamo (https://www.mixamo.com/) animations including ball throwing, dancing, and kicking. All selected animations were cut to a length of 1 second, centered to the field of view, and duplicated with a horizontally mirrored version, resulting in ten unique videos in total. For each video, trajectories of joint markers (head, neck, shoulders, elbows, wrists, hands, hop, groins, knees, ankles, and feet) were extracted and formatted into a PLD animation at a frame rate of 30 Hz.

Based on the joint dynamics of PLD body motion, scrambled motion and flickering dot videos were created. The scrambled motion videos maintained the trajectory of each joint point while randomizing the onset horizontal position of the joint points. Thus, in the scrambled motion condition, the overall motion energy remained the same while the body configuration was disrupted. In the flickering dot condition, in addition to the onset randomization, the motion was

further removed by randomly displacing half of the joint points in a random direction and keeping the rest static in each consecutive frame.

Two more static image conditions were created to test body shape selectivity. The first and middle frames of each body motion video were drawn as stick figures, resulting in twenty different images displaying a static body posture. As a control, scrambled images were generated by inverting the stick figures, randomizing the onset horizontal position the joint points and randomizing the joint point connections, resulting in twenty corresponding abstract figures with no body configuration.

All videos were resized to 1200 x 1200 pixel spanning 10 degrees in visual angle. The overall stimulus area was 7 x 5 degrees in visual angle. Examples of localizer stimuli are shown in Supplementary Video 3.

Experimental design

The functional localizer used a blocked design with five conditions of videos / images. Within each video condition block, five videos were presented for 1 second each and separated by a fixed 200-ms inter-trial interval. And for each image condition block, ten images were presented for 500 ms, interleaved by a fixed 100-ms inter-trial interval. The total length of each block was 6 seconds for all conditions. The order of block conditions was randomized for each participant, and each condition was repeated nine times within each run. Two consecutive blocks were interleaved by a jittered inter-block interval of 9.6, 10.9, or 12.2 seconds. A white “+” fixation was overlaid at the center of the display throughout the run. Each run contained five catch blocks (one catch per condition) where, in one of the trials, the fixation point changed its shape from a “+” to an “x”. Participants were instructed to press a button when they detected a change of the fixation shape. Each localizer run lasted for 13.5 minutes, and each participant finished two runs (18 repetitions per condition) during the scanning.

Stimulus presentation

All experiments were programmed using Psychtoolbox (https://www.psychtoolbox.net), implemented in Matlab 2018b (https://www.mathworks.com). Stimuli were projected onto a screen positioned at the end of the scanner bore with a Panasonic PT-EZ57OEL projector (screen

size = 30 x 18 cm, resolution = 1920 * 1200 pixels). Participants viewed the stimuli through a mirror attached to the head coil (screen-to-eye distance = 99 cm, visual angle = 17.23 x 10.38 degrees).

MRI acquisition

All images were acquired with a Siemens 7T MAGNETOM scanner with a 1-transmitter/32-receiver head coil (Nova Medical) at the Maastricht Brain Imaging Centre (MBIC) of Maastricht University, the Netherlands. Functional images were collected using the T2*-weighted multi-band accelerated EPI 2D BOLD sequence (TR/TE = 1300/20 ms; multiband acceleration factor = 3; flip angle = 59°; in-plane isotropic resolution = 1.6 mm, number of slices per volume = 66; matrix size = 128 x 128; volume number = 594 for the main experiment and 603 for the localizer). T1-weighted anatomical images were obtained using the 3D-MP2RAGE sequence (TR/TE = 5000/2.47 ms; Inverse time TI1/TI2 = 900/2750 ms; flip angle FA1/FA2 = 5/3°; in-plane isotropic resolution = 0.7 mm; matrix size = 320 x 320; slice number = 240).

Data analysis

fMRI preprocessing

For anatomical images, preprocessing was conducted with a third-party SPM pipeline (PreSurfer; Kashyap et al., 2021) under Matlab 2018b. Steps included MP2RAGE denoising, biasfield correction, and brain stripping. The resolution was then downsampled to 0.8 mm for better alignment to the 1.6 mm resolution of functional images. For functional images, the preprocessing steps were conducted by Brainvoyager 22.2 (https://www.brainvoyager.com/), including EPI distortion correction (Breman et al., 2020), slice scan time correction, 3D head motion correction, and high-pass temporal filtering (GLM with a Fourier basis set of 6 cycles / run, including linear trend). Trilinear/sinc interpolation was used in the motion correction step, and sinc interpolation was used in all other steps.

Quality assessment was conducted on each preprocessed functional run using two additional measures: the residual mean square error to reference frame (REFRMS) and the global mean amplitude change over time. The REFRMS algorithm was adapted from the “fsl_motion_outliers” function implemented in FSL

(https://web.mit.edu/fsl_v5.0.10/fsl/doc/wiki/FSLMotionOutliers.html), without calculating the REFRMS difference between consecutive frames. The measure thus quantifies the normalized intensity difference between each volume and the first volume, accounting for the residual errors after head motion correction. Similarly, global mean amplitude was calculated for each volume, accounting for the overall signal changes due to head motion or magnetic field inhomogeneity across time. The two measures were stored and served as indices for volume censoring in the later stage data analysis.

The first volume of the first functional run was chosen as a reference image for the following coregistration procedure. Displacement fields were estimated for the first volumes of all other functional runs with SyN deformation implemented in ANTs toolbox (Avants, Tustison, & Song, 2009), and further corrected with Lanczos-windowed-sinc interpolation. The preprocessed anatomical image was also warped to the reference functional image to account for the residual EPI distortion, resulting in a warped T1-weighted anatomical image. The aligned functional and anatomical images were then normalized to MNI space based on the SyN deformation estimated between the warped anatomical image and the MNI reference image. No spatial smoothing was applied in any of the steps to preserve the fine-grained voxel responses.

Noise and suspicious volume estimation

Physiological noise was estimated for each functional run using the CompCor algorithm (Behzadi, Restom, Liau, & Liu, 2007). Each participant's normalized anatomical image was segmented with a deep learning based pipeline (FastSurfer; Henschel et al., 2020) under FreeSurfer (Fischl, 2012). White matter and CSF masks were generated and eroded for 2 mm to extract non-gray matter signals. Five components were extracted for each mask and for each run with the aCompCor approach (Behzadi et al., 2007), then orthogonalized against the experiment design, resulting in ten independent confound time-courses. In addition, for each run, suspicious volumes were marked based on the REFRMS and global amplitude measures using box plot outlier limits for each measure.

Subject-level regions of interest (ROI) definition

ROIs were defined for each participant separately. A fixed-effects general linear model (GLM) was performed on the localizer runs to find voxel-wise body motion selectivity. In the design

matrix, each condition predictor was modeled as a boxcar function with the same duration of the block and convolved with the canonical hemodynamic response function (HRF). CompCor and motion confounds were added as nuisance regressors. Cortical ROIs were defined by the conjunction contrast analysis of [Body motion > other conditions] and [All conditions > 0]. No volume censoring was applied at this stage. The resulting statistical map was corrected using a cluster-threshold statistical procedure based on Monte-Carlo simulation (initial $p < 0.001$, alpha level = 0.05, iterations = 3000). The clusters were grouped based on each participant's anatomy. For bilateral clusters, only the one with higher conjunction t-value was included in the group-level ROI analysis.

Localizer ROI analysis

Group-level analysis was conducted on the localizer runs to further investigate each ROI's response profile. For each participant and each ROI, an ROI time-course was extracted by averaging the voxel time-courses and was normalized to percentage of change. A fixed-effects GLM at subject-level was conducted on the extracted ROI time-courses with the localizer conditions as predictors and the CompCor/head motion as confounds. Volume censoring was applied before GLM by removing each run's outlier volumes from both the time-course and from the design matrix. Contrast analysis was conducted on each participant's estimated betas for body motion selectivity, dot motion selectivity, and static body shape selectivity as shown in Supplementary Table 1. The beta differences were then averaged across participants, resulting in a group mean beta difference for each contrast.

Group-level statistical significance was tested by a permutation test with a bootstrapping procedure (Stelzer, Chen, & Turner, 2013). At the subject-level, the condition labels of blocks were shuffled within each run, after which the fixed-effects GLM and the contrast analysis were recalculated. The procedure was repeated for 2000 times per participant, resulting in a subject-level null-distribution of the contrast differences. Next, at group-level, bootstrapping was conducted on the participants by randomly drawing one sample with replacement from each participant's null-distribution and recalculating the group mean. This procedure was repeated 10000 times, generating a group-level null-distribution of the contrast beta difference. The p-values of the contrasts were calculated as the percentage of the absolute values of bootstrapped mean that are larger than the absolute value of experimental mean (two-tailed test).

Main experiment behavioral data analysis

The group-level statistics were first conducted in a non-parametric way to test the accuracy of each condition against the chance accuracy of 0.5. The accuracy of each condition was averaged across participants after subtracting the chance accuracy. A permutation null-distribution was generated by recalculating the group mean for 3000 times after randomly flipping the sign of each participant's accuracy difference from chance. The p-values of the contrasts were calculated as the percentage of the absolute values of bootstrapped mean that are larger than the absolute value of experimental mean (two-tailed test).

The accuracy differences between conditions were then tested with three planned contrasts (Supplementary Table 2), including the main effect of congruence, the main effect of motion frequency, and the interaction between the two factors. The accuracy difference was calculated for each contrast and averaged across participants. Permutation test was conducted for each contrast separately following the sign-flipping approach, resulting in a null-distribution with 1000 samples each for the two-tailed significance test.

Main experiment ROI analysis

The group-level ROI analysis for the main experiment followed the same pipeline as in the localizer ROI analysis. The contrasts of interest for the main experiment, included the congruence effect, the motion frequency effect, and the interaction effect as shown in the Supplementary Table 2. Permutation and bootstrapping steps were conducted to test the statistical significance, with 2000 samples for each subject-level null-distribution and 10000 samples for the group-level null-distribution.

Voxel-wise correlation

To further clarify the relationship between the main experiment effects and the localizer condition selectivity profiles, we conducted voxel-wise correlation analysis on each ROI (Peelen et al., 2006).For each participant and each ROI, a voxel-wise fixed-effects GLM was conducted on localizer and main experiment runs, followed by the corresponding contrast analysis as shown in (Supplementary Table 1 & 2). The resulting beta differences for each contrast were z-scored across voxels within each ROI. Partial correlation was conducted on each ROI's statistic map

between each main experiment effect and the dot motion selectivity, the body shape selectivity, and the body motion selectivity controlled by each other two selectivity maps. The subject-level partial correlation r-values were averaged across participants for each corresponding ROI and entered into a group-level statistical test.

The group-level statistics followed the permutation and bootstrapping approach as mentioned above. First, the GLM and ROI statistic maps were recalculated for the main experiment runs after shuffling the trial labels within each run. The partial correlation was recalculated for each pair of the maps, resulting in subject-level null-distributions of r-values, with 2000 samples generated for each pair of maps. Next, group means of the r-values were calculated after bootstrapping the subject-level null-distribution for 10000 times, resulting in the group-level null-distributions. A two-tailed test was conducted between each experimental r-value and the corresponding group null-distribution.

Results

Participant exclusion and data quality assessment

The group consisted of thirteen participants. Two participants were excluded from further fMRI analysis due to excessive head motion during scanning. For the rest of eleven participants, we detected suspicious volumes by REFRMS and Global mean amplitude measures. The percentage of marked volumes ranged from 0.66% to 8.86% (mean = 4.10%, SD = 2.44%) for the localizer runs and ranged from 1.12% to 8.14% (mean = 3.60%, SD = 2.25%) for the main experiment runs.

Subject-level ROIs

Based on the contrast analysis for body motion processing, three cortical areas were found to be robust across participants: fusiform gyrus (FG), lateral occipitotemporal cortex (LOTC), and pSTS. For FG and LOTC, large clusters were found in all participants for at least one hemisphere and for pSTS in nine out of eleven participants. The spatial coverage of the ROIs is shown in Figure 2. The details of each ROI and each participant are shown in Supplementary Table 3.

ROI response profiles for the localizer conditions

A ROI GLM analysis was conducted to investigate the response selectivity for the localizer conditions (Figure 3). Body motion selectivity was revealed for all ROIs (FG: N = 11, mean signal change % = 0.786, $p < 0.001$; LOTC: N = 11, mean signal change % = 0.583, $p < 0.001$; pSTS: N = 9, mean signal change % = 0.492, $p < 0.001$). Similarly, dot motion selectivity was also detected in all ROIs (FG: mean signal change % = 0.868, $p < 0.001$; LOTC: mean signal change % = 1.396, $p < 0.001$; pSTS: mean signal change % = 0.649, $p < 0.001$) as well as static body shape selectivity (FG: mean signal change % = 0.715, $p < 0.001$; LOTC: mean signal change % = 0.571, $p < 0.001$; pSTS: mean signal change % = 0.347, $p < 0.001$).

Behavioral results for the main experiment

The in-scanner behavioral data analysis was conducted with all thirteen participants, including the two excluded from the fMRI analysis (Figure 4). Above-chance accuracy was found for all congruent conditions at three frequency levels (15 Hz: mean = 0.965, $p < 0.001$; 7.5 Hz: mean = 0.813, $p < 0.001$; 5 Hz: mean = 0.732, $p < 0.001$), while significant below-chance accuracy (short-range motion bias) was found for all incongruent conditions (15 Hz: mean = 0.285, $p < 0.001$; 7.5 Hz: mean = 0.360, $p = 0.009$; 5 Hz: mean = 0.363, $p < 0.011$). Contrast analysis between conditions revealed a significant main effect of the congruence (mean difference = 0.500, $p < 0.001$) as well as the main effect of the frequency effect (mean difference = 0.078, $p = 0.002$). The interaction effect was significant (mean difference = 0.311, $p < 0.001$), showing a significant frequency effect only for the congruent conditions (mean difference = 0.233, $p < 0.001$) but not for the incongruent conditions (mean difference = -0.059, $p = 0.077$).

ROI response profiles for the main experiment conditions and voxel-wise correlation

With the ROI GLM and contrast analysis, a significant congruence effect was revealed in all three ROIs (Figure 5, Supplementary Figure 3), showing higher responses for congruent conditions than for incongruent conditions (FG: mean signal change % = 0.605, $p < 0.001$; LOTC: mean signal change % = 0.550, $p < 0.001$; pSTS: mean signal change % = 0.304, $p < 0.001$). The frequency effect was also significant for all ROIs, with higher responses for the 15

Hz conditions than for the 5 Hz conditions (FG: mean signal change % = 0.713, $p < 0.001$; LOTC: mean signal change % = 0.756, $p < 0.001$; pSTS: mean signal change % = 0.308 $p < 0.001$). However, the interaction effect was not significant for any of the ROIs (FG: mean signal change % = 0.091, $p = 0.290$; LOTC: mean signal change % = 0.029, $p = 0.544$; pSTS: mean signal change % = -0.073, $p = 0.263$).

Voxel-wise partial correlation analysis was conducted to investigate the relationship between the main experiment effects and localizer condition selectivity (Figure 6, Supplementary Table 4). For FG, the congruence effect was positively correlated with the body motion selectivity (mean partial $r = 0.268$, $p < 0.001$) and the dot motion selectivity (mean partial $r = 0.223$, $p = 0.004$), but was not significantly correlated with the static body shape selectivity (mean partial $r = -0.055$, $p = 0.469$). The frequency effect was positively correlated with the body motion selectivity (mean partial $r = 0.283$, $p < 0.001$) and the dot motion selectivity (mean partial $r = 0.349$, $p < 0.001$) and was weakly negatively correlated with the static body shape selectivity (mean partial $r = -0.168$, $p = 0.026$). The interaction effect was not significantly correlated with any of the localizer selectivities (Supplementary Table 4A).

For LOTC, the congruence effect was positively correlated with the body motion selectivity (mean partial $r = 0.174$, $p < 0.005$) and the dot motion selectivity (mean partial $r = 0.367$, $p < 0.001$), and was marginally negatively correlated with the static body shape selectivity (mean partial $r = -0.097$, $p = 0.043$). The frequency effect was positively correlated with the body motion selectivity (mean partial $r = 0.204$, $p < 0.001$) and the dot motion selectivity (mean partial $r = 0.598$, $p < 0.001$), and was marginally negatively correlated with the static body shape selectivity (mean partial $r = -0.096$, $p = 0.046$). The interaction effect was weakly positively correlated with the body motion selectivity (mean partial $r = 0.133$, $p = 0.035$), where body motion selectivity was positively correlated with the frequency effect for the congruent conditions (mean partial $r = 0.237$, $p < 0.001$) but not significant for the incongruent conditions (mean partial $r = 0.088$, $p = 0.165$). The interaction effect was not found to be correlated with the other two localizer selectivities (Supplementary Table 4B).

For pSTS, only the frequency effect was positively correlated with the body motion selectivity (mean partial $r = 0.237$, $p = 0.003$) and the dot motion selectivity (mean partial $r = 0.253$, $p < 0.001$), but was not significantly correlated with the static body shape selectivity (mean partial r

= -0.010, p = 0.882). No significant correlation was detected for the other pairs of effects (Supplementary Table 4C).

Discussion

We used synthesized pixel-wise optical flow videos to isolate the contribution of low-level motion extracted from naturalistic walking videos. This approach minimized body shape cues and enabled systematic manipulation of temporal congruence across sequences of local motion cues. Behavioral results showed that the body movement perceived from the synthesized videos was mainly driven by the local motion instead of the underlying sequence of body structure. At the neural level, multiple cortical areas, including LOTC, FG, and pSTS, were influenced by local motion properties. In LOTC and FG, voxel responses driven by local motion were positively correlated with dot motion and body motion selectivity and not with static body shape selectivity. The findings indicate that local motion serves as a parallel input for body-motion perception in LOTC and FG, offering a new perspective on the cortical pathways underlying body-motion analysis.

Behavioral Evidence for Local-Motion–Driven Perception

Behavioral results showed that participants could reliably perceive the walking direction from segments of local motion as revealed by significant above chance accuracy for the congruent condition. In contrast, when the local motion velocity was flipped in incongruent conditions, accuracy dropped below chance, even though the sequence of implied body postures remained unchanged. These findings confirmed that the recognition of movement direction in the synthesized videos was not driven by the trajectory of motion-induced body shapes but instead, was derived from the local motion patterns.

We further investigated how the spatiotemporal dependence of local motion segments may affect the integration of global body movements. A familiar assumption is that the perception of global motion requires the accumulation of segmented motion over time (Blake & Shiffrar, 2007; Oguz, Aydin, & Urgen, 2024)). If so, task performance should be dependent on the total amount of motion presented and the accumulation of motion signals would fail in the incongruent setting, as the motion information between consecutive segments is inconsistent. Thus, we compared the

task performance between videos with the lowest (5 Hz) and the highest (15 Hz) amount of body motion. Increasing the amount of motion improved task accuracy for the congruent conditions but not for the incongruent conditions, with a significant interaction effect observed between congruence and motion frequency. Together with the below-chance accuracy, the results suggest that in incongruent conditions, participants' decisions were biased by the individual motion segments and not based on integrated body motion.

Shared mechanism between local motion and body movement processing in FG and LOTC

For FG, LOTC, and pSTS, we found significant higher activations for congruent motion conditions than for incongruent ones. Since the congruent videos had the same underlying body shape sequences and total motion energy as the incongruent ones, the congruence effect observed in the ROIs reflects low-level modulations by the local motion signal. To further investigate the contribution from the shape-driven motion processing, we conducted partial correlation analysis between the low-level motion effects and the localizer selectivity. Various low-level motion effects were positively correlated with body-motion and dot-motion selectivity, whereas they showed nonsignificant or marginal negative correlations with body shape selectivity. The results contrast with Peelen et al. (2006), where the conventional PLD body motion was strongly correlated with the body shape selectivity instead of motion selectivity. Thus, the current results separated the low-level motion effects from the motion-from-shape mechanism (Beintema & Lappe, 2002). Instead, the results are in line with the computational framework of the optical flow motion pathway as proposed by Giese and Poggio (2003), which predicted the interference of motion integration due to the inconsistency between segments of the motion signal.

Our synthesized stimuli with minimal shape information provided the opportunity to revisit the roles of conventional body selective areas in body motion processing. We first found significant main effects of congruence and motion frequency in FG. Both effects significantly correlated with body motion and dot motion selectivity at the voxel level, while no interaction related effects or correlations were found. Thus, the two flow motion effects (congruence and frequency) and the two localizer selectivity (body motion and dot motion) results may be driven by similar factors. One candidate factor may be the general articulated structures implied by the motion

pattern (Taylor & Downing, 2011; Taylor et al., 2007). In the main experiment, inducing motion inconsistency can introduce ambiguity of the underlying structure through the interrupted global motion signals, while reducing the motion frequency directly decreases the amount of the shape-from-motion cues. Meanwhile in the localizer experiment, the structural cues decreased from the preferred one as in the body motion condition, to the random non-preferred structure as in the scrambled dot motion condition, and finally to the absence of structure as in the flickering dot condition. However, the configuration cues in our stimuli were derived from motion rather than from static body shapes, challenging the view that FG analyzes only individual frames, i.e. static body shape, rather than the overall temporal dynamics (Downing et al., 2006). Furthermore, the absence of a correlation with body shape selectivity indicates that the motion inputs served a role beyond merely cueing body shape segmentation. Thus, the current finding suggested a significant contribution of FG to the local coherent motion pattern analysis.

Within LOTC, dot motion selectivity was correlated with both the congruence effect and the motion frequency effect. Because LOTC region encompasses multiple motion selective areas including hMT, the dot motion correlation observed here may reflect differences in global and local motion energy manipulated by the motion consistency and temporal frequency. More revealing, however, were the patterns related to body-motion selectivity. Voxelwise body motion selectivity in LOTC was strongly correlated to the frequency effect only under congruent conditions but not under the incongruent ones. Together with LOTC's preference for congruent motion, this pattern suggests that body motion selectivity of the LOTC area reflects the spatiotemporal integration of the global motion pattern, a process that is disrupted when local motion segments are temporally incongruent. Importantly, rather than subdividing LOTC into hMT+ and EBA, we leveraged voxelwise selectivity to capture the functional heterogeneity across the LOTC region. These results may help explaining the prior reports of limited motion contributions in the static body defined EBA (Grossman & Blake, 2002; Peelen et al., 2006): we observed slightly negative correlations between local motion effects and static body shape selectivity, indicating that the voxels most responsive to static shape feature were less modulated by the local motion cues. Together with the findings in FG, these results argue against a strict hierarchical account, in which motion signals are sole extracted from body shape sequence. Instead, they support a model in which low-level motions and body shapes are encoded as partially separable dimensions within FG and LOTC.

Surprisingly, although pSTS is well known to be involved in body motion processing (Grossman, Battelli, & Pascual-Leone, 2005; Grossman & Blake, 2002; Jastorff & Orban, 2009; Vangeneugden et al., 2014), it showed no correlation between the flow motion effects and the body motion selectivity. Despite the pSTS preference toward the congruent conditions, the absence of correlation between the congruence effect and the body motion selectivity suggested that the PLD body motion selectivity may be driven by a different mechanism. Nevertheless, the results are well predicted by the study of Michels et al. (2005), in which the sequential position PLD was used to remove the low-level motion. In addition to the classic body motion selectivity, comparably high activation was found in pSTS for sequential position PLDs with static body postures (Michels et al., 2005). Thus, instead of motion itself, the pSTS may be involved in estimating and updating the body shape across time. These results underscored the importance to separate the motion features when studying body movement perception. While the PLD format is efficient for the laboratory control of visual features, it oversimplifies the naturalistic body movements by sacrificing the feature-rich motion field. This simplification may introduce redundant body structure computations to articulate the isolated joint points before extracting the global motion pattern (Duarte, Abreu, & Castelo-Branco, 2022; Hirai & Senju, 2020). Furthermore, due to the loss of motion features, such modification may misrepresent the cortical basis for naturalistic body movement recognition especially by overestimating the body shape contribution.

Reconsidering hierarchical models of body perception

Here, by taking the local motion signal as the primary visual input, we observed the sensitivity in LOTC to the global body movement. In line with the neuroimaging results, our behavioral results revealed stronger bias toward the low-level motion instead of the shape-from-motion sequence, suggesting that the computation within LOTC may serve as the major source for higher-level decision making in this task. Moreover, our data suggested that FG may analyze transient local motion patterns for an intermediate stage of motion processing, analogous to the computations of “OF pattern neurons” proposed by Giese and Poggio (2003). Thus, one remaining question is how this motion pathway may support the higher-level recognition for the body movements. In addition to the parallel processing view (Giese & Poggio, 2003), an alternative hypothesis is that body movement perception is achieved in a hierarchical manner, with FG and LOTC first

detecting body forms and pSTS integrating these signals into higher-level perceptions (Grossman & Blake, 2002; Jastorff & Orban, 2009; Landsiedel, Daughters, Downing, & Koldewyn, 2022). However, the current study indicates that the processing of isolated local motion cues can be already detected in FG and LOTC. In line with this result, other studies have reported that multiple categorical dimensions embedded in the low-level motion field can be decoded from temporal occipital areas (Robert et al., 2023; Schmid, Boyaci, & Doerschner, 2021). These results are compatible with the "third pathway" (Pitcher & Ungerleider, 2021), which proposed a motion-driven pathway through hMT+ and STS. Future studies may focus on whether those motion-based analyses will be merged into pSTS, or instead, will be directly read out through other parallel cortical pathways (Sokolov et al., 2018).

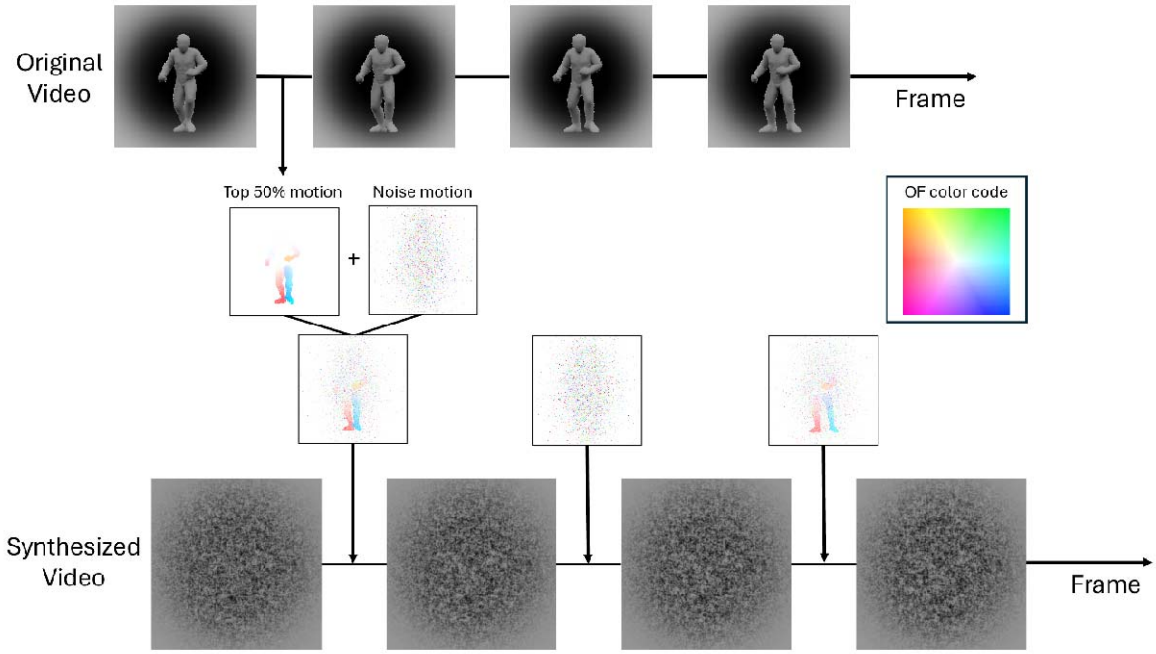
Another key question is whether the integration between shape and motion occurs within LOTC. Previous studies have shown that LOTC represents multiple shape-based body features including joint positions (Marrazzo, De Martino, Lage-Castellanos, Vaessen, & De Gelder, 2023), whole-body postures (Poyo Solanas, Vaessen, & de Gelder, 2020), and biomechanical dynamics (Marrazzo, De Martino, Mukovskiy, Giese, & de Gelder, 2025). While it has been proposed that the body movements are encoded in LOTC as "snapshots" or discrete frames (Downing et al., 2006), it remains unclear if these frame-wise representations can be modulated by the motion cues. Importantly, LOTC responses have been found sensitive to the temporal context. Temporally scrambled videos (Smekal, Poyo Solanas, & de Gelder, 2025) and videos with random body postures (Downing et al., 2006) elicit stronger activations than coherent, continuous body movement in LOTC. This pattern is compatible with the suppression effects observed in early visual cortex for shape-based apparent motion (Alink, Schwiedrzik, Kohler, Singer, & Muckli, 2010; Hidaka, Nagai, Sekuler, Bennett, & Gyoba, 2011; Shen, Han, & de Lange, 2020), which predicts stronger activation for violations to temporal contexts. However, such mechanism is incompatible with the local motion effects found in the current study, where the inconsistency across time resulted in decreased activities. This dissociation suggests that LOTC computations local motion differently from shape-based motion. Conceptually, shape-based motion analysis requires integrating multiple position changes across frames to infer global movements, the local motion reflects the motion state at each individual position (Kwon, Tadin, & Knill, 2015). Furthermore, the local motion states may also affect the efficiency when tracking the trajectories of multiple body parts (Clair, Huff, & Seiffert, 2010; Kwon et al., 2015)

and estimate the body part position (Maus, Fischer, & Whitney, 2013; Nishida & Johnston, 1999; Ramachandran & Anstis, 1990; Whitney, 2002). Interactions between static features and implied motion have been reported in LOTC (Kourtzi & Kanwisher, 2000; Lorteije et al., 2006; Pavan, Cuturi, Maniglia, Casco, & Campana, 2011). An important extension of the present work would be to test whether local motion states modulate LOTC encoding of static body postural features (e.g. joint position), for example by overlaying OF motion onto static body shapes.

Finally, the current results raise the question about the subcortical contributions for the body movement processing. Representations for complex motion patterns have been found in multiple subcortical structures including superior colliculus (SC; X. Lu et al., 2024; Schneider & Kastner, 2005; Wolf et al., 2015) and pulvinar (Cotton & Smith, 2007; Villeneuve, Kupers, Gjedde, Ptito, & Casanova, 2005), which have anatomical connectivity to hMT+ (Arcaro, Pinsk, & Kastner, 2015; Lanyon et al., 2009). Studies with fMRI and tractography have proposed the pathway between LGN and hMT+ that bypasses V1 and serves as the entrance to cortical process for motion (Ajina & Bridge, 2018; Ajina, Pestilli, Rokem, Kennard, & Bridge, 2015; Gaglianese, Costagli, Bernardi, Ricciardi, & Pietrini, 2012). These observations raise the possibility that motion signals may access body-sensitive cortex through pathways that are at least partly distinct from those carrying detailed shape information. Clarifying how subcortical and cortical motion pathways interact—and whether they converge or remain segregated before reaching body-selective regions—will be essential for understanding the neural architecture supporting biological motion perception and for establishing functional homologies between human and non-human primate systems.

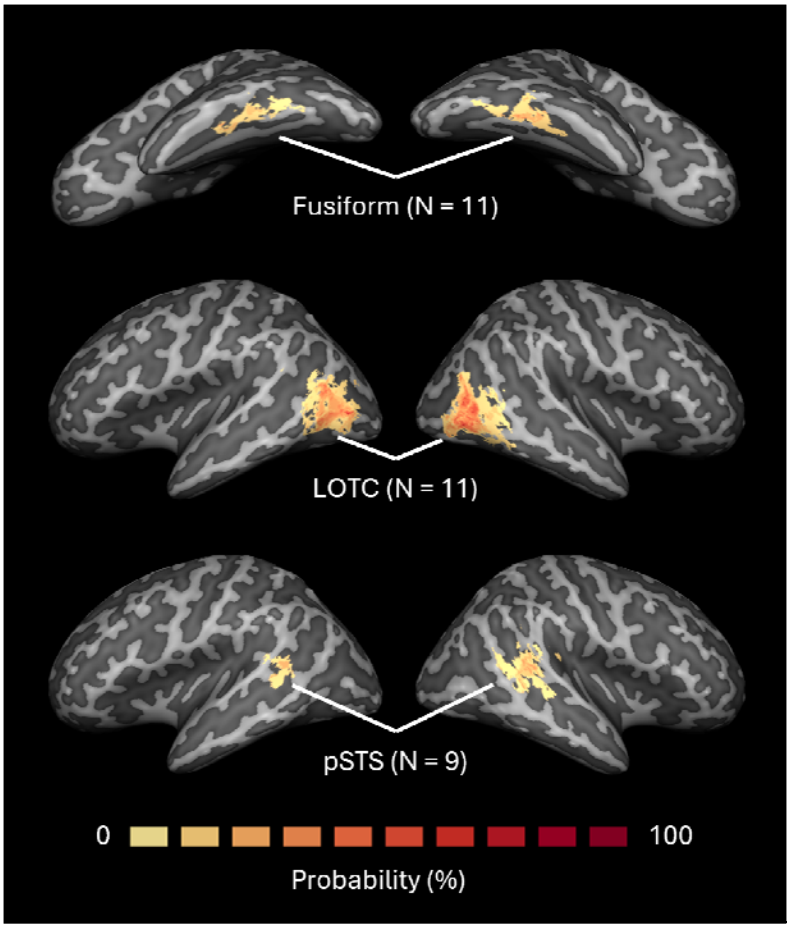
Figures

Figure 1. Illustration of OF video synthesizing pipeline.



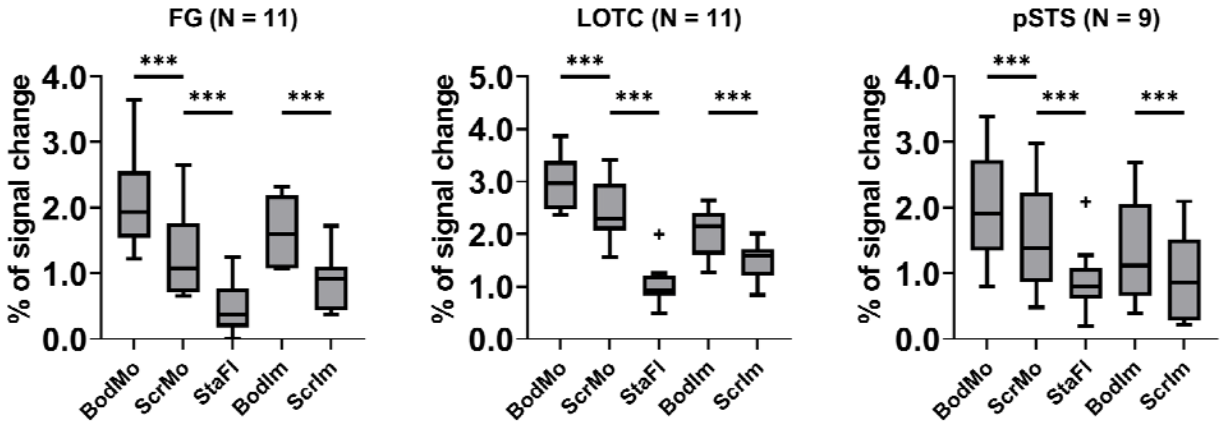
Pixel-wise velocity or optical flow (OF) between each two adjacent frames from the avatar body movement was extracted, filtered to keep the top 50% of motion, and combined with random background motion. A synthesized video started with a salt-and-pepper noise frame. Each subsequent frame was generated by displacing the pixels of the previous frame based on the corresponding OF map. According to different empirical conditions, the applied body motion OF can either be modified or be replaced by random noise motion. Example OF maps are visualized by HSV color code in the figure, with directions presented by the hue and magnitudes presented by the saturation.

Figure 2. Probability map for subject-level ROIs.



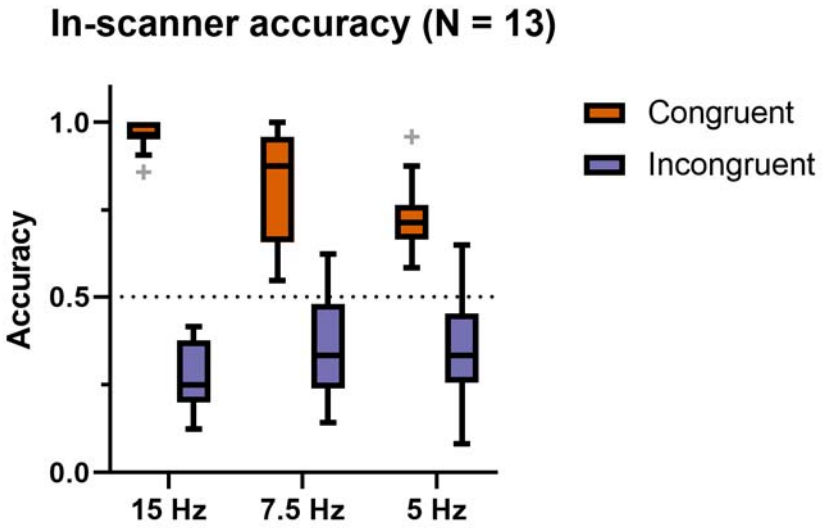
The probability maps demonstrate the overlap of significant localizer ROIs across participants. The ROIs were defined at a subject level by the conjunction contrast analysis of [Body motion > other conditions] and [All conditions > 0]. The resulting maps were projected to a normalized brain surface in MNI space and plotted as probability maps across participants. For bilateral clusters, only the one with higher conjunction t-value was included in the group-level ROI analysis (Supplementary Table 3).

Figure 3. ROI response profiles for localizer conditions.



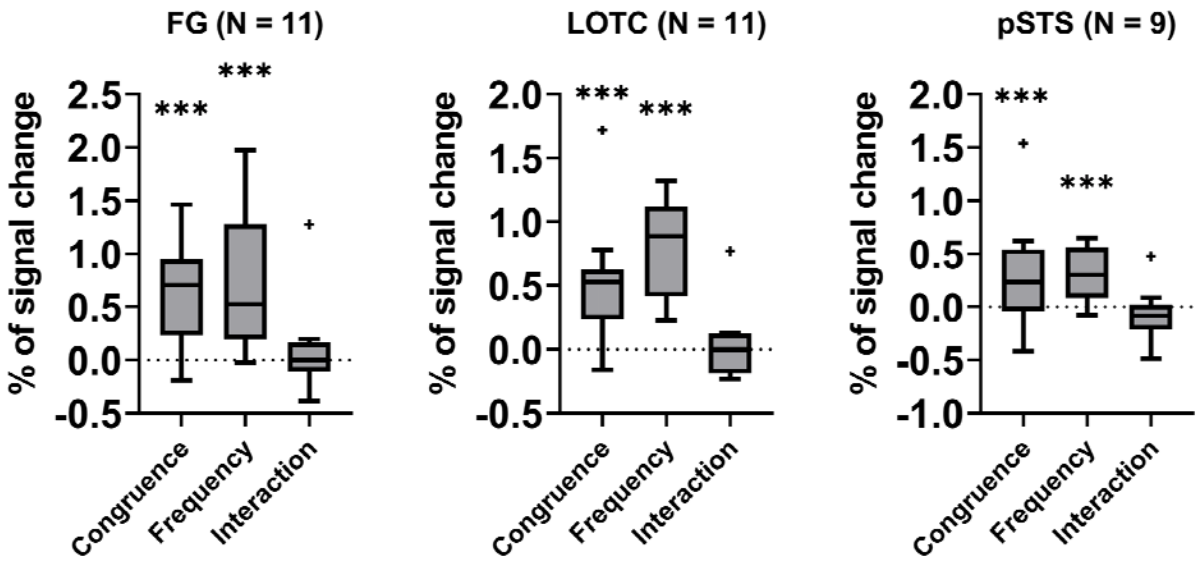
The boxplots show the beta values resulted from random effect ROI GLM on the localizer runs, with conditions of body motion (BodMo), scrambled dot motion (ScrMo), static flickering dots (StaFl), body image (BodIm), and scramble image (ScrIm). The whiskers (error bars) represent mininum / maximum values within 1.5 times the interquartile range from the lower or upper quartile and the scatters (+) are plotted for individual data beyond the upper and lower bounds. For all box plots, asterisks indicate the significant planned comparisons (Table 1) based on permutation-bootstrapping tests (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; two-tailed).

Figure 4. Task performance for OF videos



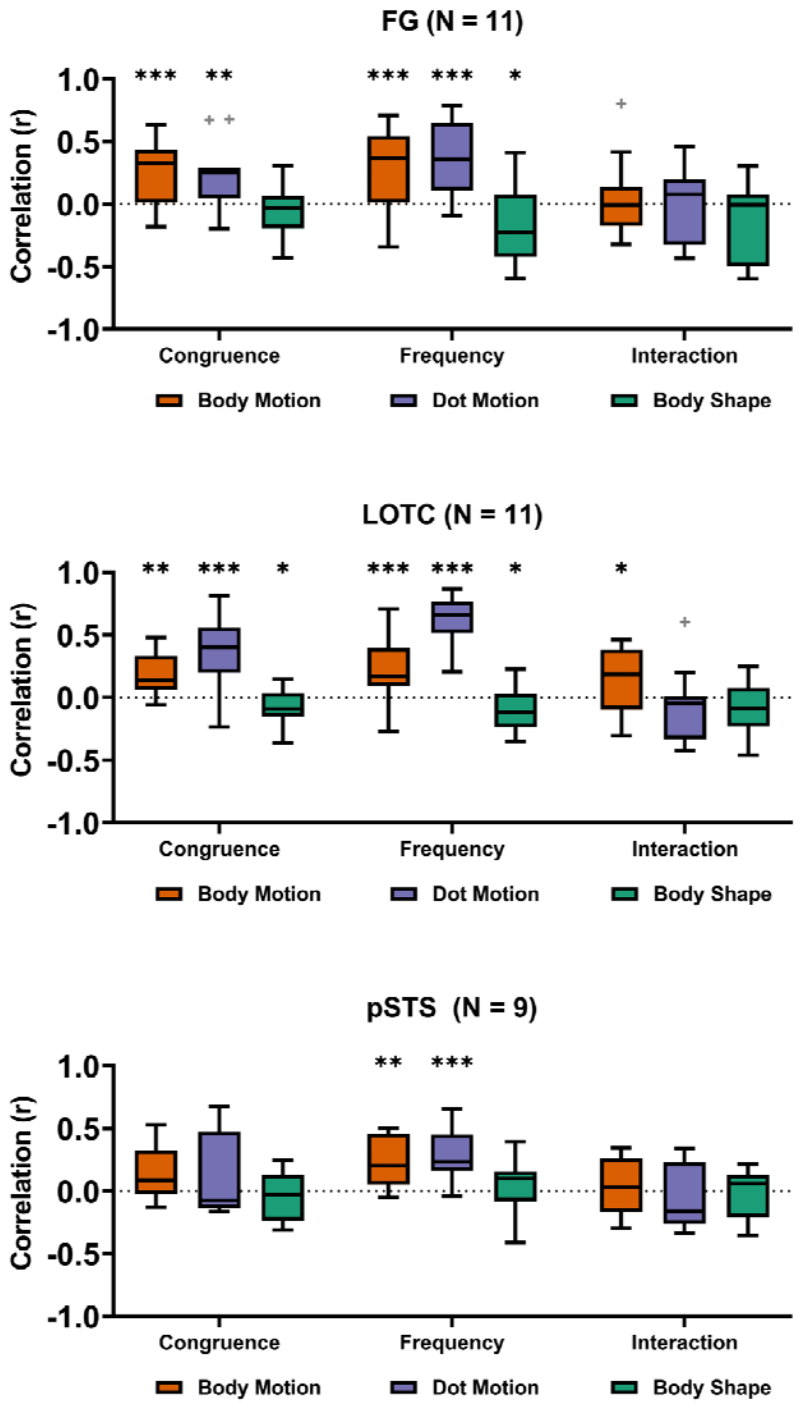
The boxplots show participants' accuracies for the in-scanner walking direction task, with whiskers (error bars) represent minimum / maximum values within 1.5 times interquartile range from the lower or upper quartiles and the scatters (+) plotted for individual data beyond the upper and lower bounds.

Figure 5. ROI response profiles for local motion effects.



The boxplots show the beta differences resulted from random effect ROI GLM and planned contrasts (Table 2) on the main experiment, with whiskers (error bars) represent minimum / maximum values within 1.5 times the interquartile range from the lower or upper quartiles and the scatters (+) plotted for individual data beyond the upper and lower bounds. For all boxplots, asterisks indicate the significant contrast based on permutation-bootstrapping tests (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; two-tailed).

Figure 6. Voxel-wise correlation between local motion effects and localizer selectivity within each ROI.



The boxplots show the partial correlation r-values of voxel-wise correlations between the local motion effects (congruence, frequency, and interaction) and the localizer-based selectivities

(body motion, dot motion, and body shape). Whiskers (error bars) represent minimum / maximum values within 1.5 times the interquartile range from the lower or upper quartiles and the scatters (+) are plotted for individual data beyond the upper and lower bounds. For all boxplots, asterisks indicate significant correlations based on permutation-bootstrapping tests (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; two-tailed).

References

Ajina, S., & Bridge, H. (2018). Blindsight relies on a functional connection between hMT+ and the lateral geniculate nucleus, not the pulvinar. *PLoS biology, 16*(7), e2005769.

Ajina, S., Pestilli, F., Rokem, A., Kennard, C., & Bridge, H. (2015). Human blindsight is mediated by an intact geniculo-extrastriate pathway. *Elife, 4*, e08935.

Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *Journal of Neuroscience, 30*(8), 2960-2966.

Arcaro, M. J., Pinsk, M. A., & Kastner, S. (2015). The anatomical and functional organization of the human visual pulvinar. *Journal of Neuroscience, 35*(27), 9848-9871.

Avants, B. B., Tustison, N., & Song, G. (2009). Advanced normalization tools (ANTS). *Insight j, 2*(365), 1-35.

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage, 37*(1), 90-101.

Beintema, J., & Lappe, M. (2002). Perception of biological motion without local image motion. *Proceedings of the National Academy of Sciences, 99*(8), 5661-5663.

Bigelow, A., Kim, T., Namima, T., Bair, W., & Pasupathy, A. (2023). Dissociation in neuronal encoding of object versus surface motion in the primate brain. *Current Biology, 33*(4), 711-719. e715.

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annu. Rev. Psychol., 58*(1), 47-73.

Bognar, A., Nejad, G. G., Rens, G., Raman, R., & Vogels, R. (2025). Expanding the stimulus domain: Co-occurrence of motion and body-category selectivity in the macaque ventral STS. *Progress in Neurobiology, 249*, 102769.

Braddick, O. (1974). A short-range process in apparent motion. *Vision research, 14*(7), 519-527.

Breman, H., Mulders, J., Fritz, L., Peters, J., Pyles, J., Eck, J., . . . Goebel, R. (2020). *An image registration-based method for EPI distortion correction based on opposite phase encoding (COPE).* Paper presented at the International Workshop on Biomedical Image Registration.

Brownlow, S., Dixon, A. R., Egbert, C. A., & Radcliffe, R. D. (1997). Perception of movement and dancer characteristics from point-light displays of dance. *The Psychological Record, 47*(3), 411-422.

Cartucho, J., Tukra, S., Li, Y., S. Elson, D., & Giannarou, S. (2021). VisionBlender: a tool to efficiently generate computer vision datasets for robotic surgery. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 9*(4), 331-338.

Clair, R. S., Huff, M., & Seiffert, A. E. (2010). Conflicting motion information impairs multiple object tracking. *Journal of Vision, 10*(4), 18-18.

Cotton, P. L., & Smith, A. T. (2007). Contralateral visual hemifield representations in the human pulvinar nucleus. *Journal of neurophysiology, 98*(3), 1600-1609.

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the psychonomic society, 9*(5), 353-356.

De Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience, 7*(3), 242-249.

De Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal behavior, 13*(4), 247-268.

Dittrich, W. H., Troscianko, T., Lea, S. E., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception, 25*(6), 727-738.
Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science, 293*(5539), 2470-2473.
Downing, P. E., Peelen, M. V., Wiggett, A. J., & Tew, B. D. (2006). The role of the extrastriate body area in action perception. *Social Neuroscience, 1*(1), 52-62.
Duarte, J. V., Abreu, R., & Castelo-Branco, M. (2022). A two-stage framework for neural processing of biological motion. *Neuroimage, 259*, 119403.
Ferri, S., Kolster, H., Jastorff, J., & Orban, G. A. (2013). The overlap of the EBA and the MT/V5 cluster. *Neuroimage, 66*, 412-425.
Fischl, B. (2012). FreeSurfer. *Neuroimage, 62*(2), 774-781.
Gaglianese, A., Costagli, M., Bernardi, G., Ricciardi, E., & Pietrini, P. (2012). Evidence of a direct influence between the thalamus and hMT+ independent of V1 in the human brain as measured by fMRI. *Neuroimage, 60*(2), 1440-1447.
Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience, 4*(3), 179-192.
Grossman, E. D., Battelli, L., & Pascual-Leone, A. (2005). Repetitive TMS over posterior STS disrupts perception of biological motion. *Vision research, 45*(22), 2847-2853.
Grossman, E. D., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron, 35*(6), 1167-1175.
Hedges, J. H., Gartshteyn, Y., Kohn, A., Rust, N. C., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (2011). Dissociation of neuronal and psychophysical responses to local and global motion. *Current Biology, 21*(23), 2023-2028.
Henschel, L., Conjeti, S., Estrada, S., Diers, K., Fischl, B., & Reuter, M. (2020). Fastsurfer-a fast and accurate deep learning based neuroimaging pipeline. *Neuroimage, 219*, 117012.
Hidaka, S., Nagai, M., Sekuler, A. B., Bennett, P. J., & Gyoba, J. (2011). Inhibition of target detection in apparent motion trajectory. *Journal of Vision, 11*(10), 2-2.
Hirai, M., & Senju, A. (2020). The two-process theory of biological motion processing. *Neuroscience & Biobehavioral Reviews, 111*, 114-124.
Jastorff, J., & Orban, G. A. (2009). Human functional magnetic resonance imaging reveals separation and integration of shape and motion cues in biological motion processing. *Journal of Neuroscience, 29*(22), 7315-7329.
Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & psychophysics, 14*(2), 201-211.
Kashyap, S., Ivanov, D., Havlicek, M., Huber, L., Poser, B. A., & Uludağ, K. (2021). Sub-millimetre resolution laminar fMRI using Arterial Spin Labelling in humans at 7 T. *Plos one, 16*(4), e0250504.
Koerfer, K., & Lappe, M. (2020). Pitting optic flow, object motion, and biological motion against each other. *Journal of Vision, 20*(8), 18-18.
Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience, 12*(1), 48-55.
Kwon, O.-S., Tadin, D., & Knill, D. C. (2015). Unifying account of visual motion and position perception. *Proceedings of the National Academy of Sciences, 112*(26), 8142-8147.
Landsiedel, J., Daughters, K., Downing, P. E., & Koldewyn, K. (2022). The role of motion in the neural representation of social interactions in the posterior temporal cortex. *Neuroimage, 262*, 119533.

Lanyon, L. J., Giaschi, D., Young, S. A., Fitzpatrick, K., Diao, L., Bjornson, B. H., & Barton, J. J. (2009). Combined functional MRI and diffusion tensor imaging analysis of visual motion pathways. *Journal of Neuro-Ophthalmology, 29*(2), 96-103.

Lorteije, J. A., Kenemans, J. L., Jellema, T., Van der Lubbe, R. H., De Heer, F., & Van Wezel, R. J. (2006). Delayed response to animate implied motion in human motion processing areas. *Journal of Cognitive Neuroscience, 18*(2), 158-168.

Lu, X., Hu, Z., Xin, Y., Yang, T., Wang, Y., Zhang, P., . . . Jiang, Y. (2024). Detecting biological motion signals in human and monkey superior colliculus: a subcortical-cortical pathway for biological motion perception. *Nature Communications, 15*(1), 9606.

Lu, Z.-L., & Sperling, G. (2001). Three-systems theory of human visual motion perception: review and update. *Journal of the Optical Society of America A, 18*(9), 2331-2370.

Marrazzo, G., De Martino, F., Lage-Castellanos, A., Vaessen, M. J., & De Gelder, B. (2023). Voxelwise encoding models of body stimuli reveal a representational gradient from low-level visual features to postural features in occipitotemporal cortex. *Neuroimage, 277*, 120240.

Marrazzo, G., De Martino, F., Mukovskiy, A., Giese, M. A., & de Gelder, B. (2025). Neural encoding of biomechanically (im) possible human movements in occipitotemporal cortex. *PLOS Computational Biology, 21*(12), e1013694.

Maus, G. W., Fischer, J., & Whitney, D. (2013). Motion-dependent representation of space in area MT+. *Neuron, 78*(3), 554-562.

Michels, L., Lappe, M., & Vaina, L. M. (2005). Visual areas involved in the perception of human movement from dynamic form analysis. *Neuroreport, 16*(10), 1037-1041.

Nishida, S. y., & Johnston, A. (1999). Influence of motion signals on the perceived position of spatial pattern. *Nature, 397*(6720), 610-612.

Oguz, O. C., Aydin, B., & Urgen, B. A. (2024). Biological motion perception in the theoretical framework of perceptual decision-making: An event-related potential study. *Vision research, 218*, 108380.

Pavan, A., Cuturi, L. F., Maniglia, M., Casco, C., & Campana, G. (2011). Implied motion from static photographs influences the perceived position of stationary objects. *Vision research, 51*(1), 187-194.

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of neurophysiology, 93*(1), 603-608.

Peelen, M. V., Wiggett, A. J., & Downing, P. E. (2006). Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron, 49*(6), 815-822.

Pitcher, D. (2025). The Third Visual Pathway for Social Perception. *arXiv preprint arXiv:2512.09351*.

Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a third visual pathway specialized for social perception. *Trends in cognitive sciences, 25*(2), 100-110.

Poyo Solanas, M., Vaessen, M., & de Gelder, B. (2020). Computation-based feature representation of body expressions in the human brain. *Cerebral cortex, 30*(12), 6376-6390.

Ramachandran, V. S., & Anstis, S. M. (1990). Illusory displacement of equiluminous kinetic edges. *Perception, 19*(5), 611-616.

Raman, R., Bognar, A., Nejad, G. G., Taubert, N., Giese, M., & Vogels, R. (2023). Bodies in motion: Unraveling the distinct roles of motion and shape in dynamic body responses in the temporal cortex. *Cell reports, 42*(12).

Robert, S., Ungerleider, L. G., & Vaziri-Pashkam, M. (2023). Disentangling object category representations driven by dynamic and static visual input. *Journal of Neuroscience, 43*(4), 621-634.

Roether, C. L., Omlor, L., Christensen, A., & Giese, M. A. (2009). Critical features for the perception of emotion from gait. *Journal of Vision, 9*(6), 15-15.

Schmid, A. C., Boyaci, H., & Doerschner, K. (2021). Dynamic dot displays reveal material motion network in the human brain. *Neuroimage, 228*, 117688.

Schneider, K. A., & Kastner, S. (2005). Visual responses of the human superior colliculus: a high-resolution functional magnetic resonance imaging study. *Journal of neurophysiology, 94*(4), 2491-2503.

Shen, L., Han, B., & de Lange, F. P. (2020). Apparent motion induces activity suppression in early visual cortex and impairs visual detection. *Journal of Neuroscience, 40*(28), 5471-5479.

Smekal, V., Poyo Solanas, M., & de Gelder, B. (2025). Disentangling dynamic information and temporal order processing of human action perception. *PNAS nexus, 4*(3), pgaf067.

Sokolov, A. A., Zeidman, P., Erb, M., Ryvlin, P., Friston, K. J., & Pavlova, M. A. (2018). Structural and effective brain connectivity underlying biological motion detection. *Proceedings of the National Academy of Sciences, 115*(51), E12034-E12042.

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *Neuroimage, 65*, 69-82.

Taylor, J. C., & Downing, P. E. (2011). Division of labor between lateral and ventral extrastriate representations of faces, bodies, and objects. *Journal of Cognitive Neuroscience, 23*(12), 4122-4137.

Taylor, J. C., Wiggett, A. J., & Downing, P. E. (2007). Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas. *Journal of neurophysiology, 98*(3), 1626-1633.

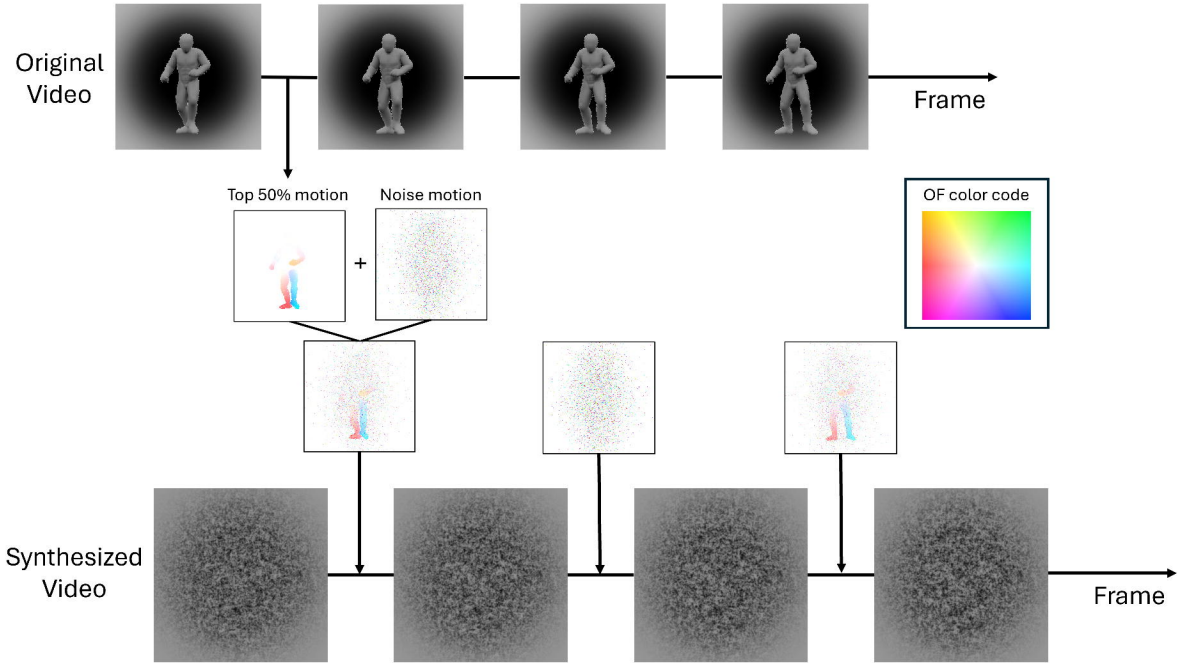
Vangeneugden, J., Peelen, M. V., Tadin, D., & Battelli, L. (2014). Distinct neural mechanisms for body form and body motion discriminations. *Journal of Neuroscience, 34*(2), 574-585.

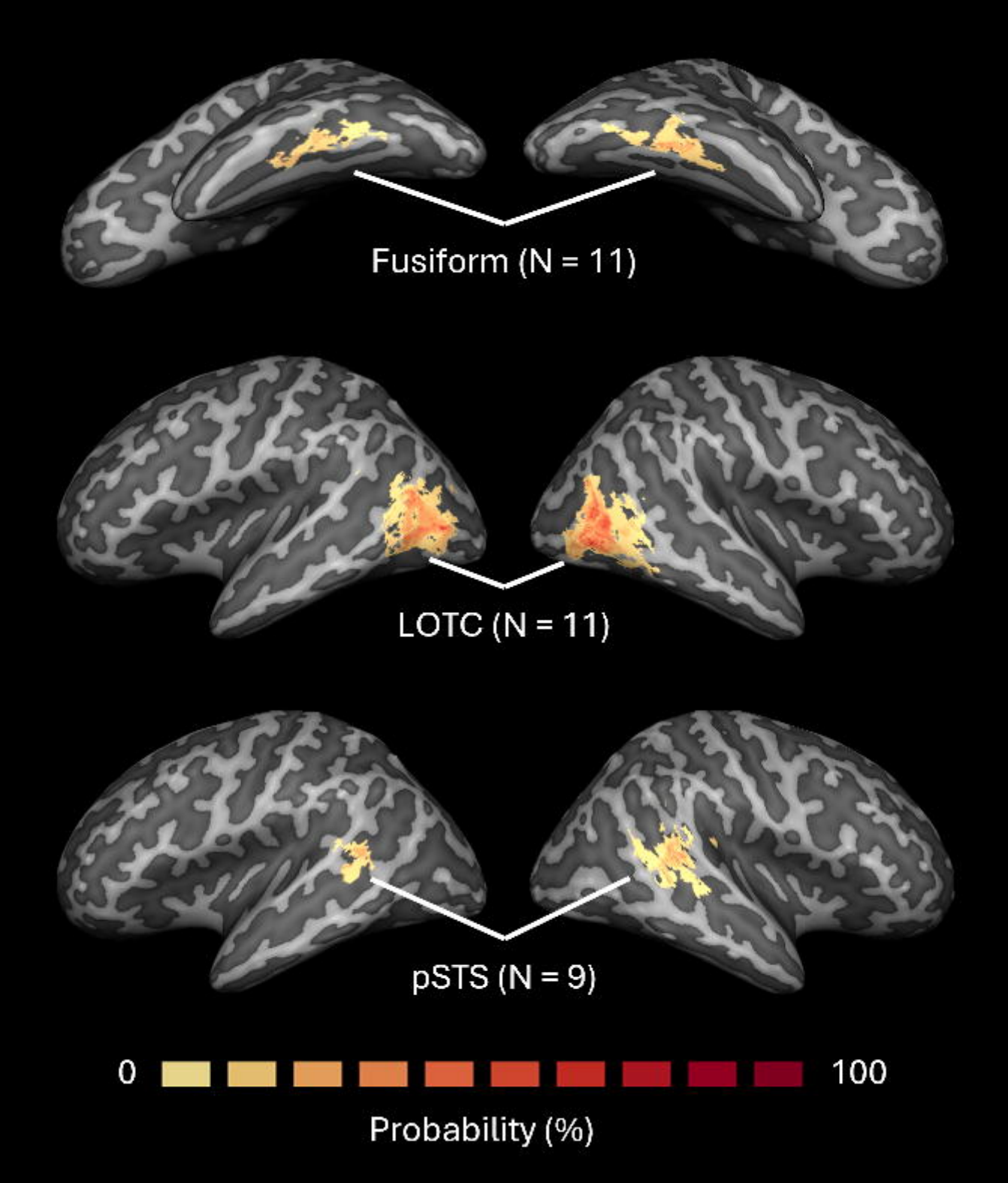
Villeneuve, M. Y., Kupers, R., Gjedde, A., Ptito, M., & Casanova, C. (2005). Pattern–motion selectivity in the human pulvinar. *Neuroimage, 28*(2), 474-480.

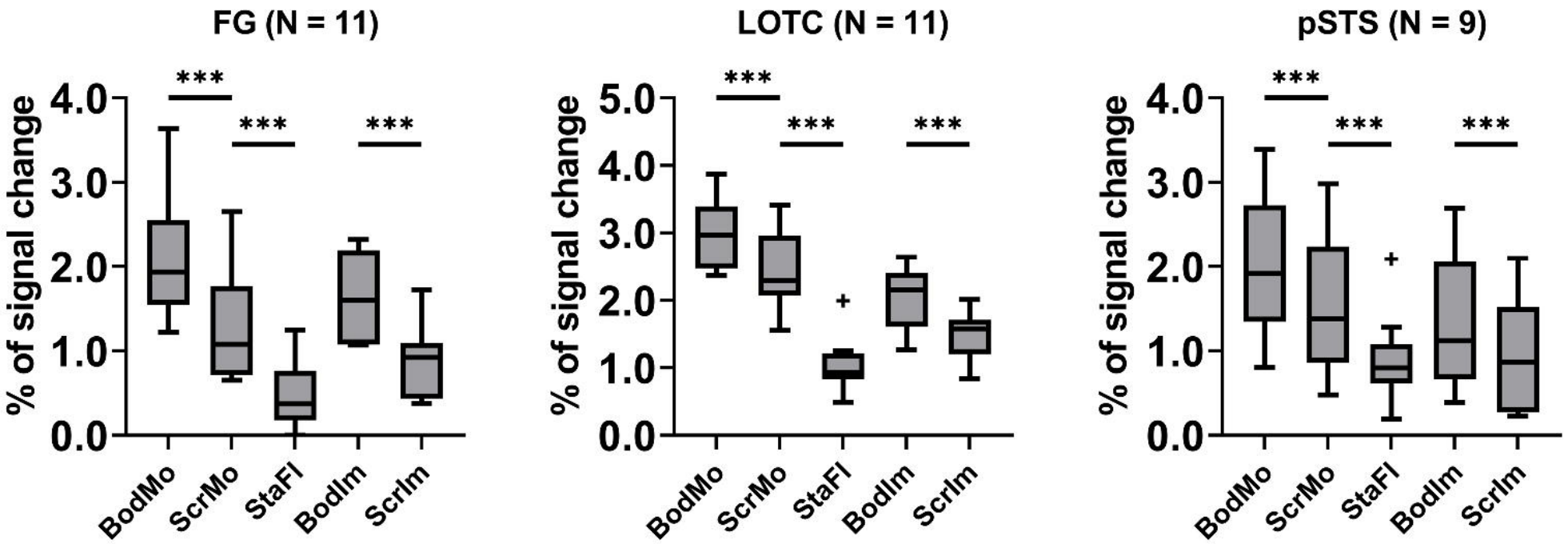
Wallbott, H. G. (1998). Bodily expression of emotion. *European journal of social psychology, 28*(6), 879-896.

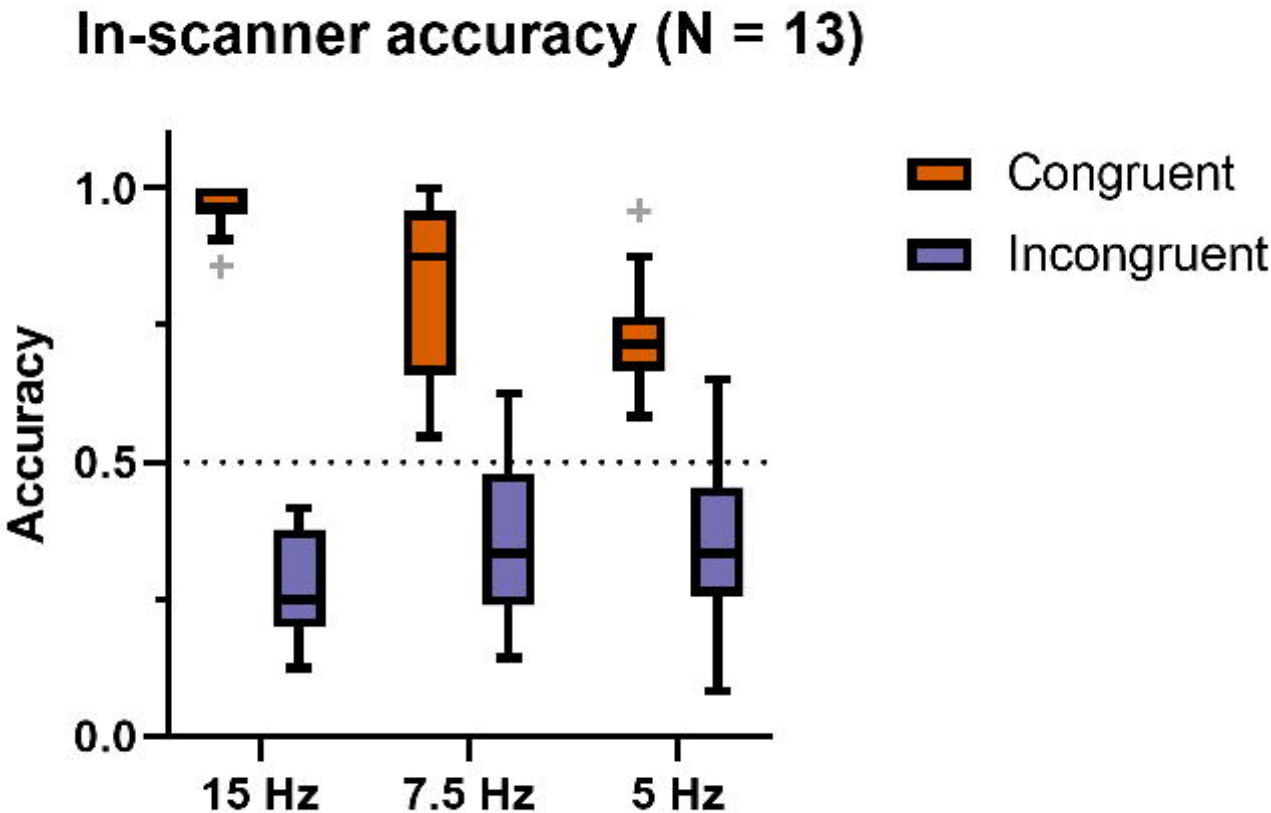
Whitney, D. (2002). The influence of visual motion on perceived position. *Trends in cognitive sciences, 6*(5), 211-216.

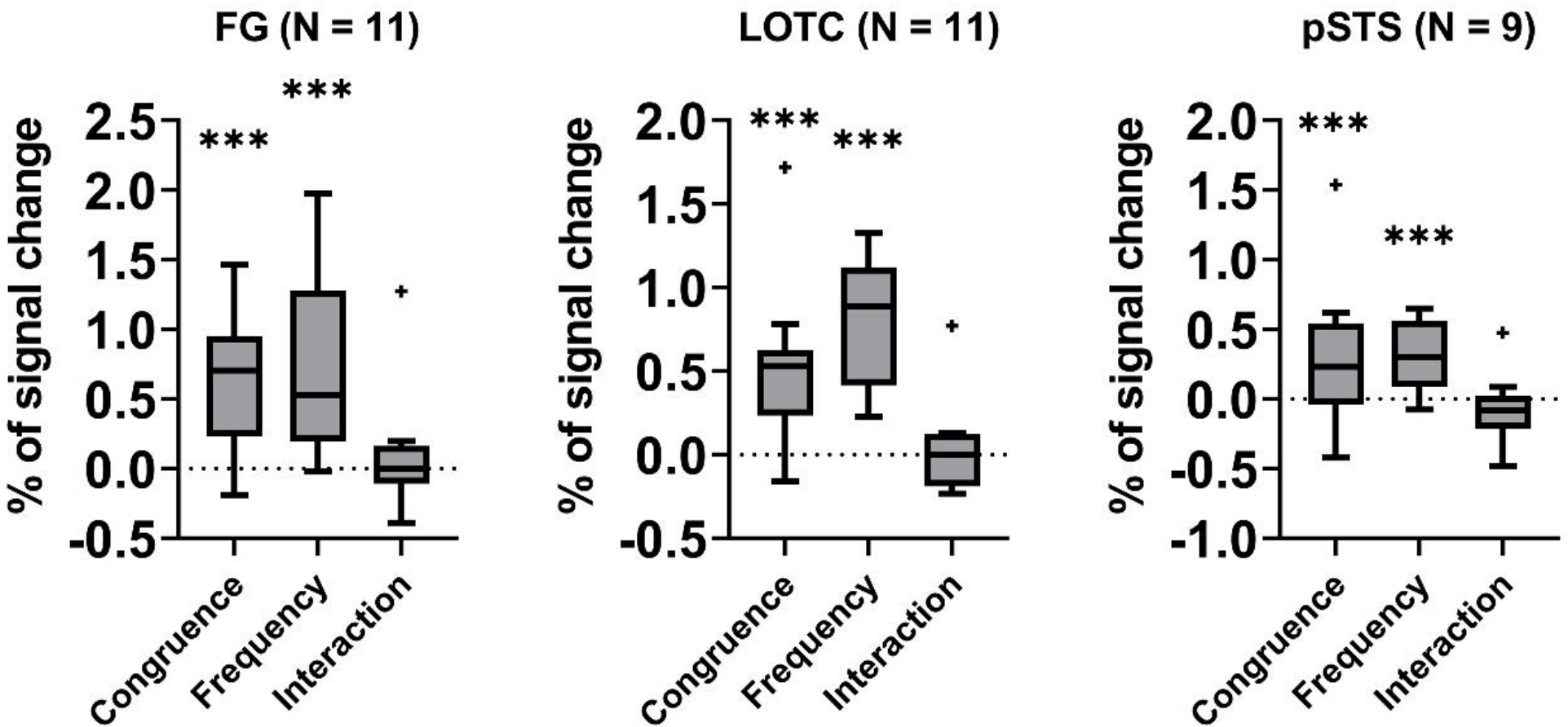
Wolf, A. B., Lintz, M. J., Costabile, J. D., Thompson, J. A., Stubblefield, E. A., & Felsen, G. (2015). An integrative role for the superior colliculus in selecting targets for movements. *Journal of neurophysiology, 114*(4), 2118-2131.

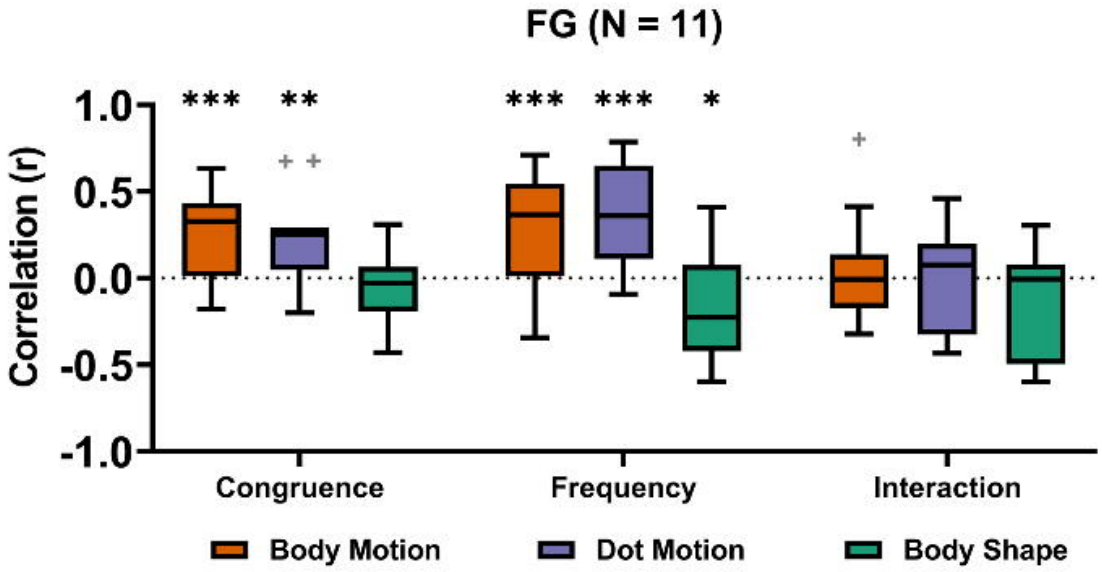

Original Video
Frame
Top 50% motion
Noise motion
+
OF color code
Synthesized Video
Frame

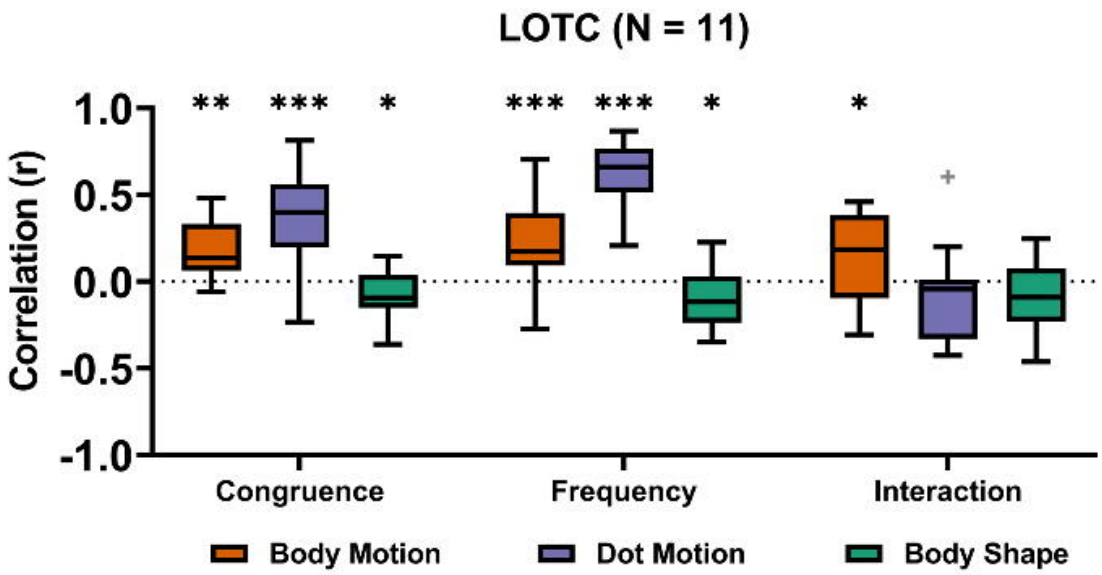

Fusiform (N = 11)
LOTC (N = 11)
pSTS (N = 9)
0
100
Probability (%)

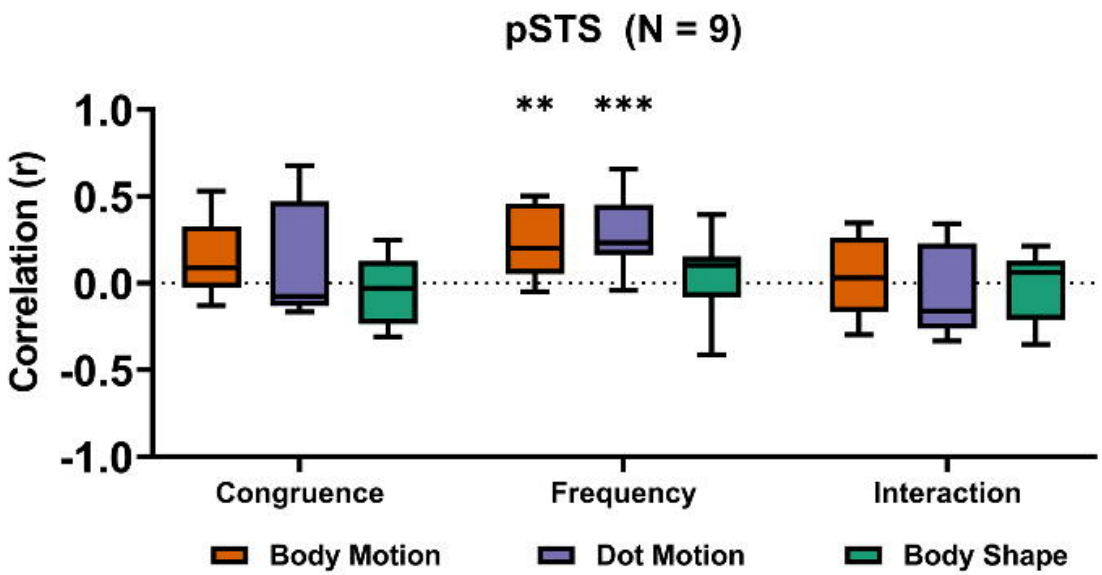

FG (N = 11)
LOTC (N = 11)
pSTS (N = 9)
% of signal change
0.0
1.0
2.0
3.0
4.0
5.0
***
BodMo
ScrMo
StaFl
BodIm
ScrIm


In-scanner accuracy (N = 13)
Accuracy
1.0
0.5
0.0
15 Hz
7.5 Hz
5 Hz
Congruent
Incongruent


FG (N = 11)
LOTC (N = 11)
pSTS (N = 9)
% of signal change
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
***
Congruence
Frequency
Interaction

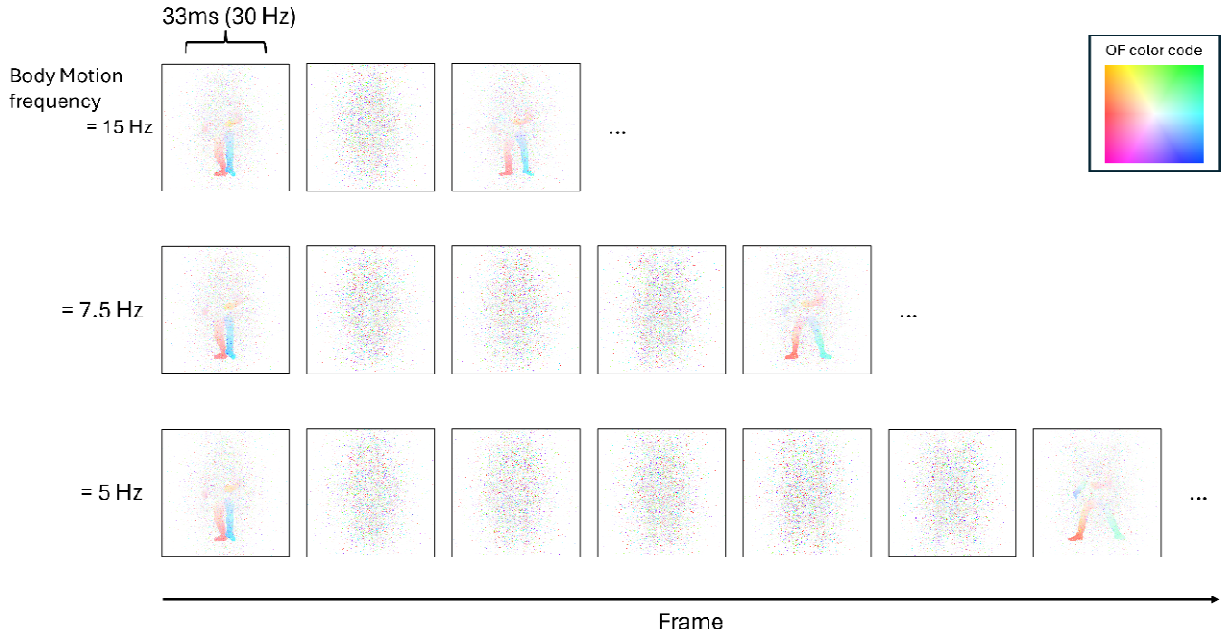

FG (N = 11)
Correlation (r)
1.0
0.5
0.0
-0.5
-1.0
***
**
***
***
*
Congruence
Frequency
Interaction
Body Motion
Dot Motion
Body Shape


LOTC (N = 11)
Correlation (r)
1.0
0.5
0.0
-0.5
-1.0
**
***
*
***
***
*
*
Congruence
Frequency
Interaction
Body Motion
Dot Motion
Body Shape


pSTS (N = 9)
Correlation (r)
1.0
0.5
0.0
-0.5
-1.0
**
***
Congruence
Frequency
Interaction
Body Motion
Dot Motion
Body Shape

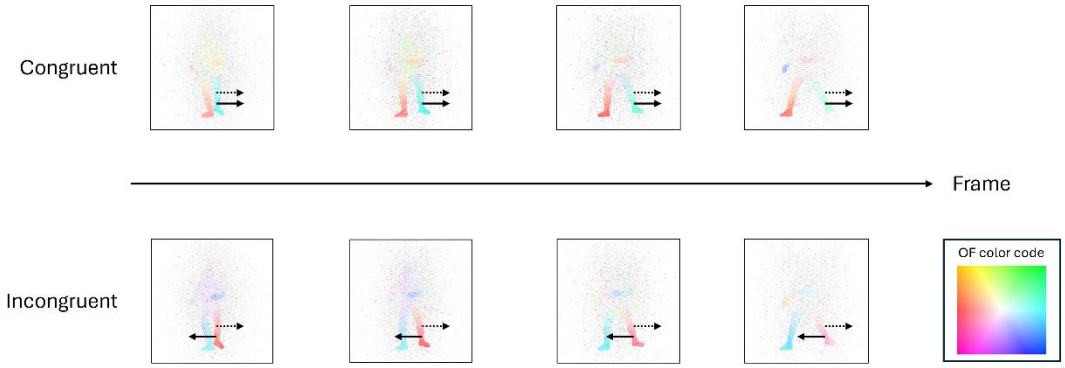
Supplementary Figures

Supp. Figure 1. Illustration of applied OF sequence in different frequency settings.



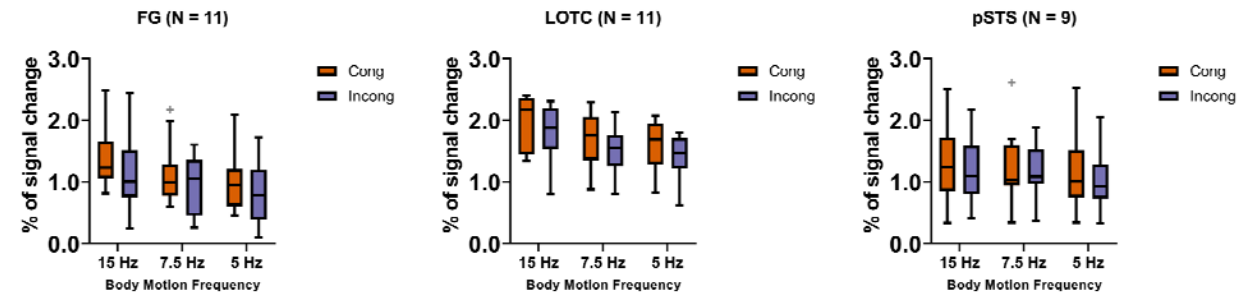
The pixel-wise OF motion was applied to generate each subsequent frame, creating a transient motion (33 ms) between each two frames at a frame rate of 30 Hz. The OF motions contained either body motion patterns or only noise motion. The body motion OFs were applied every two, four, or six frames, interleaved by noise motions. At a frame rate of 30 Hz, the presence of body motion became 15 Hz, 7.5Hz, and 5 Hz accordingly.

Supp. Figure 2. Illustration of applied OF sequence in different congruence settings.



The congruence factor was controlled to modify the spatiotemporal dependence between the local motion vectors (solid arrow) and the displacement of motion field across time (dotted arrow). As shown in the plot, for both conditions, the motion field for one leg move to the right across time. For the congruent condition, each segment of local motion within the leg area has the same rightward velocity which matches the motion field displacement. However, for the incongruent condition, the local motion always has the opposite direction against the motion field displacement.

Supp. Figure 3. ROI response profiles for main experiment conditions.



The boxplots show the beta values resulted from random effect ROI GLM on the main experiment runs, with whiskers (error bars) representing minimum / maximum values within 1.5 times the interquartile range from the lower or upper quartile and the scatters (+) plotted for individual data beyond the upper and lower bounds.

Supplementary Tables

Supp. Table 1

	Body motion	Scrambled motion	Flickering dots	Body shape	Scrambled shape
Body motion selectivity	1	-1	0	0	0
Dot motion selectivity	0	1	-1	0	0
Body shape selectivity	0	0	0	1	-1

Supp. Table 2

	Cong / 15 Hz	Cong / 7.5 Hz	Cong / 5 Hz	Incong / 15 Hz	Incong / 7.5 Hz	Incong / 5 Hz
Congruence effect	1	1	1	-1	-1	-1
Frequency effect	1	0	-1	1	0	-1
Interaction	1	0	-1	-1	0	1

Supp. Table 3

ROI	Subject	Hemisphere	MNI coordinates of center			voxeL number	t(df)
			x	y	z		
FG	1	L*	-44.31	-36.44	-20.32	79	5.69(1052)
		R	45.34	-58.03	-14.21	94	3.26(1052)
	2	L	-47.53	-65.22	-15.21	391	7.42(1134)
		R*	45.5	-56.34	-17.04	706	8.02(1134)
	3	R*	45.16	-54.71	-16.94	419	5.30(1145)
	4	L	-44.91	-47.53	-14.73	170	4.14(1117)
		R*	42.52	-44.96	-15.48	112	5.92(1117)
	5	L	-41.51	-48.43	-12.32	84	4.21(1094)
		R*	43.92	-44.67	-16.57	96	5.04(1094)
	6	L	-45.47	-45.27	-18.33	627	7.13(1085)
		R*	42.35	-34.61	-19.49	299	7.24(1085)
	7	L*	-42.41	-44.59	-18.4	67	5.80(1104)
	8	L*	-44.29	-34.17	-20.67	172	8.05(1129)
		R	46.18	-53.01	-15.31	953	7.15(1129)
	9	L*	-44.43	-52.98	-17.89	67	6.34(1143)
		R	41.93	-48.37	-14.93	294	6.15(1143)
	10	L*	-44.68	-55.65	-14.44	660	13.65(1123)
		R	42.49	-47.69	-19.47	774	8.41(1123)
	11	L*	-43.67	-42.42	-18.21	174	6.60(1159)
LOTC	1	L	-54.84	-64.35	5.02	933	6.18(1052)
		R*	49.4	-64.13	9.27	1309	6.81(1052)
	2	L	-46.18	-73.27	10.41	3563	11.54(1134)
		R*	46.37	-69.55	5.46	3337	11.83(1134)
	3	L	-46.46	-68.27	13.93	2362	8.55(1145)
		R*	46.87	-66.63	12.89	2378	9.71(1145)
	4	L	-46.08	-75.83	9.91	2890	9.93(1117)
		R*	48.88	-69.08	6.16	3835	11.05(1117)
	5	L	-50.11	-70.56	7.78	2174	7.78(1094)

ROI	Subject	Hemisphere	MNI coordinates of center			voxeL number	t(df)
		R*	48.5	-68.48	8.42	2162	9.65(1094)
	6	L	-51.03	-71.7	11.42	2406	7.82(1085)
		R*	51.81	-63.48	6.02	3103	8.07(1085)
	7	L	-48.77	-70.35	4.51	2499	9.36(1104)
		R*	50.78	-62.83	3.14	3145	9.50(1104)
	8	L*	-52.21	-66.37	9.45	1958	8.88(1129)
		R	51.47	-66.02	7.48	1152	7.47(1129)
	9	L	-49.77	-71.69	2.16	1405	7.31(1143)
		R*	47.64	-66.81	3.83	3028	8.79(1143)
	10	L	-48.49	-69.16	9.02	4018	12.56(1123)
		R*	50.49	-61.65	0.91	6034	15.42(1123)
	11	L	-45.09	-73.82	6.33	1328	8.85(1159)
		R*	45.88	-66.28	1.36	2185	10.74(1159)
pSTS	1	R*	54.22	-46.72	11.26	382	4.57(1052)
	2	L*	-55.82	-46.1	15.06	447	5.30(1134)
		R	48.63	-40.72	13.11	365	4.52(1134)
	3	R*	51.88	-38.72	13.95	177	6.79(1145)
	6	L	-51.07	-46.99	11.43	353	5.79(1085)
		R*	53.97	-40.15	13.33	1092	6.68(1085)
	7	L*	-48.76	-45.95	6.49	240	4.81(1104)
		R	53	-37.26	8.61	212	4.69(1104)
	8	L	-54.64	-45.45	12.14	403	7.25(1129)
		R*	50.51	-43.84	11.49	1975	9.18(1129)
	9	L	-52.62	-48.24	16.18	500	5.40(1143)
		R*	52.37	-38.1	18.2	524	7.28(1143)
	10	L*	-53.03	-52.79	18.93	2466	11.91(1123)
		R	54.76	-37.86	9.57	2493	9.47(1123)
	11	L*	-55.83	-39.08	9.08	557	7.52(1159)
		R	50.07	-36.75	6.85	264	6.63(1159)

Supp. Table 4

A.

			Flow motion effect	
		Congruence	Frequency	Interaction
	Biological motion	r = 0.268, p < 0.001	r = 0.283, p < 0.001	r = 0.042, p = 0.612
Localizer selectivity	Dot motion	r = 0.223, p = 0.004	r = 0.349, p < 0.001	r = 0.005, p = 0.956
	Body shape	r = -0.055, p = 0.469	r = -0.168, p = 0.026	r = -0.122, p = 0.104

B.

			Flow motion effect	
		Congruence	Frequency	Interaction
	Biological motion	r = 0.174, p = 0.005	r = 0.204, p < 0.001	r = 0.133, p = 0.035
Localizer selectivity	Dot motion	r = 0.367, p < 0.001	r = 0.598, p < 0.001	r = -0.055, p = 0.561
	Body shape	r = -0.097, p = 0.043	r = -0.096, p = 0.046	r = -0.064, p = 0.184

C.

			Flow motion effect	
		Congruence	Frequency	Interaction
	Biological motion	r = 0.141, p = 0.076	r = 0.237, p = 0.003	r = 0.037, p = 0.648
Localizer selectivity	Dot motion	r = 0.130, p = 0.113	r = 0.253, p < 0.001	r = -0.034, p = 0.668
	Body shape	r = -0.106, p = 0.104	r = -0.010, p = 0.882	r = -0.053, p = 0.424