Representation of auditory motion directions and sound source locations in the human planum temporale

Ceren Battal [*1,2], Mohamed Rezk[1,2], Stefania Mattioni[1,2], Jyothirmayi Vadlamudi[2] & Olivier Collignon [*1,2]

[1]Center of Mind/Brain Sciences, University of Trento, Italy
[2]Institut de recherche en Psychologie (IPSY), Institute of Neuroscience (IoNS), Université catholique de Louvain (UcL)

*Corresponding authors: Ceren Battal; ceren.battal@unitn.it; Olivier Collignon; Institut de recherche en Psychologie (IPSY), 10, place du cardinal Mercier, 1348 Louvain-la-Neuve ; olivier.collignon@uclouvain.be.

Key words: spatial hearing, planum temporale, auditory motion, auditory space, direction selectivity, fMRI, multivariate analyses.

Conflict of interest: The authors declare no competing financial interests.

Acknowledgements: We would also like to express our gratitude to Giorgia Bertonati, Marco Barilari, Stefania Benetti, Valeria Occelli, Stephanie Cattoir who helped with the data acquisition; Jorge Jovicich for helping setting-up the fMRI acquisition parameters and Pietro Chiesa for continuing support with the auditory hardware. The project was funded by the ERC starting grant MADVIS (Project: 337573) awarded to OC. MR and JV are PhD fellows and OC is research associate at the Fond National de la Recherche Scientifique de Belgique (FRS-FNRS).

ABSTRACT

The ability to precisely compute the location and direction of sounds is a crucial perceptual skill to efficiently interact with dynamic environments. How the human brain implements spatial hearing is however poorly understood. In our study, we used fMRI to characterize the brain activity of humans listening to left, right, up and down moving as well as static sounds. Whole brain univariate results contrasting moving and static sounds varying in their location revealed a robust functional preference for auditory motion in bilateral human Planum Temporale (hPT). Using independently localized hPT, we show that this region contains information about auditory motion directions and, to a lesser extent, sound source locations. Moreover, hPT showed an axis of motion organization reminiscent of the functional organization of the middle-temporal cortex (hMT+/V5) for vision. Importantly, whereas motion direction and location rely on partially shared pattern geometries in hPT, as demonstrated by successful cross-condition decoding, the responses elicited by static and moving sounds were however significantly distinct. Altogether our results demonstrate that the hPT codes for auditory motion and location but that the underlying neural computation linked to motion processing is more reliable and partially distinct from the one supporting sound source location.

INTRODUCTION

The ability to precisely locate and track moving information is a crucial perceptual skill for efficient interaction with the environment. While the brain mechanisms underlying the processing of visual localization and visual motion have received considerable attention (Braddick et al., 2001; Britten et al., 1996; Movshon and Newsome, 1996; Newsome and Park, 1988), much less is known about how the brain implements spatial hearing. The representation of auditory space relies on the computations and comparison of intensity, temporal and spectral cues that arise at each ear (Blauert, 1982; Searle et al., 1976). In the auditory pathway, these cues are both processed and integrated in the thalamus, brainstem and cortex in order to create an integrated neural representation of auditory space (Boudreau and Tsuchitani, 1968; Goldberg and Brown, 1969; Imig et al., 2000; Ingham et al., 2001; Young et al., 1992; Yin and Chan, 1990). At the cortical level, the acoustic space lacks point-to-point spatial representation (Derey et al., 2016; Middlebrooks, 2002; Middlebrooks and Bremen, 2013; Middlebrooks and Pettigrew, 1981; Ortiz-Rios et al., 2017; Rajan et al., 1990). Similar to the visual cortex dual-stream processing model (Ungerleider and Mishkin, 1982; Goodale and Milner, 1992), partially distinct ventral "what" and dorsal "where" auditory processing streams have been proposed for auditory processing (Barrett and Hall, 2006; Lomber and Malhotra, 2008; Rauschecker and Tian, 2000; Recanzone, 2000; Romanski et al., 1999; Tian et al., 2001; Warren and Griffiths, 2003). In particular, the dorsal stream is thought to process sound source location and motion both in animals and humans (Arnott et al., 2004; Maeder et al., 2001; Rauschecker and Tian 2000; Tian et al., 2001). However, it remains poorly understood whether the human brain implements the processing of auditory motion and location using distinct, similar, or partially shared neural codes (Poirier et al., 2017; Smith et al., 2010; Zatorre et al., 2002).

One candidate region that might integrate spatial cues to compute motion and location information in the human auditory cortex is the planum temporale (hPT) (Barrett and Hall, 2006; Baumgart and Gaschler-Markefski, 1999; Warren et al., 2002). hPT is located in the superior temporal gyrus, posterior to Helsch' gyrus, and is typically considered part of the dorsal auditory stream (Derey et al., 2016;

Poirier et al., 2017; Rauschecker and Tian, 2000; Warren et al., 2002). Some authors have suggested that hPT engages in both the processing of moving sounds and the location of static sound-sources (Barrett and Hall, 2006; Derey et al., 2016; Krumbholz et al., 2005; Smith et al., 2004, 2007, 2010; Zatorre et al., 2002). This proposition is supported by early animal electrophysiological studies demonstrating the neurons in the auditory cortex that are selective to sound source location and motion directions (Altman, 1968, 1994; Benson et al., 1981; Doan et al., 1999; Imig et al., 1990; Middlebrooks and Pettigrew, 1981; Poirier et al., 1997; Rajan et al., 1990), and these neurons display similar response profiles for moving and sound source location (Ahissar et al., 1992; Doan et al., 1999; Poirier et al., 1997). In contrast, other studies in animals (Poirier et al., 2017) and humans (Baumgart and Gaschler-Markefski, 1999; Bremmer et al., 2001; Griffiths et al., 1998; Hall and Moore, 2003; Krumbholz et al., 2005; Lewis et al., 2000; Pavani et al., 2002; Poirier et al., 2005) pointed toward a more specific role of hPT for auditory motion processing.

The hPT, selective to process auditory motion/location in a dorsal "where" pathway is reminiscent of the dorsal region in the middle temporal cortex (hMT+/V5) in the visual system. hMT+/V5 is dedicated to process visual motion (Movshon and Newsome, 1996; Watson et al., 1993; Tootell et al., 1995) and displays a columnar organization in particular of a preferred axis of motion directions (Albright et al., 1984; Zimmermann et al., 2011). Whether hPT disclose similar characteristic tuning remains however unknown.

The main goals of the present study were threefold. First, using multivariate pattern analysis (MVPA), we investigated whether information about auditory motion direction and sound-source location can be retrieved from the pattern of activity in hPT. Further, we asked whether the spatial distribution of the neural representation is in the format of "preferred axis of motion/location" as observed in the visual motion selective regions (Albright et al., 1984; Zimmermann et al., 2011). Finally, we aimed at characterizing whether the processing of motion direction (e.g. going to the left) and sound-source location (e.g. being in the left) rely on partially common neural representations in the hPT.

MATERIALS AND METHODS

Participants

Eighteen participants with no reported auditory problems were recruited for the study. Two participants were excluded due to poor spatial hearing performance in the task. The final sample included 16 right-handed participants (8 females, age range: 20 to 42, mean ± SD = 32 ± 5.7 years). Participants were blindfolded and instructed to keep their eyes closed throughout the experiments and practice runs. All the procedures were approved by the research ethics boards of the Centre for Mind/Brain Sciences (CIMeC) and University of Trento. Experiments were undertaken with the understanding and written consent of each participant.

Auditory stimuli

Our limited knowledge of the auditory space processing in the cortex of humans might be a consequence of the technical challenge of evoking vivid perceptual experience of auditory space inside fMRI. To create an externalized ecological sensation of sound location and motion, we relied on individual in-ear stereo recordings that were recorded in a semi-anechoic room and from 30 loudspeakers on horizontal and vertical planes, mounted on two semicircular wooden structures with a radius of 1.1m (see Fig. 1A). Participants were seated in the center of the apparatus with their head on a chin-rest, such that the speakers on the horizontal and vertical planes were equally distant from participants' ears. Then, these recordings were re-played to the participants when they were inside the functional MRI (fMRI). By using such sound system with in-ear recordings, auditory stimuli automatically convolved with each individuals' own pinna and head related transfer function to produce a salient auditory perception in external space.

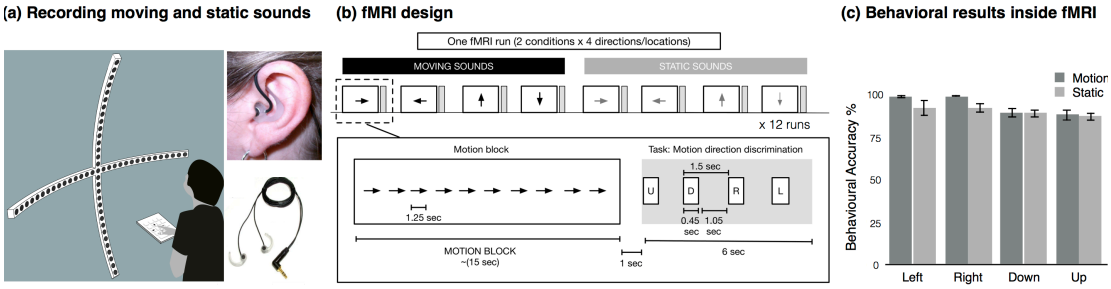

Figure 1. Stimuli and Experimental Design. (A) The acoustic apparatus used to present auditory moving and static sounds while binaural recordings were carried out for each participant before the fMRI session. **(B)** Auditory stimuli presented inside the fMRI consisted of 8 conditions: leftward, rightward, downward and upward moving stimuli and left, right, down and up static stimuli. Each condition was presented for 15 s (12 repetition of 1250 ms sound, no ISI) and followed by 7 s gap for indicating the corresponding direction/location in space and 8 s of silence (total inter-block interval was 15 s). Sound presentation and response button press were pseudo-randomized. Subjects were asked to respond as accurately as possible during the gap period. **(C)** The behavioural performance inside the scanner.

The auditory stimuli were prepared using custom MATLAB scripts (r2013b; Matworks). Auditory stimuli were recorded using binaural in-ear omni-directional microphones (Sound Professionals-TFB-2; 'flat' frequency range 20–20,000 Hz) connected to a portable Zoom H4n digital wave recorder (16-bit, stereo, 44.1 kHz sampling rate). Microphones were positioned at the opening of participant's left and right auditory ear canals. While auditory stimuli were played, participants were listening without performing any task with head fixed to the chin-rest in front of them. Binaural in-ear recordings allowed combining binaural properties such as interaural time and intensity differences, and participant specific monaural filtering cues to create reliable and ecological auditory space sensation (Pavani et al., 2002).

Stimuli recordings

Sound stimuli consisted of 1250 ms pink noise (50 ms rise/fall time). In the motion condition, the pink noise was presented moving in 4 directions: leftward, rightward, upward and downward. Moving stimuli covered 120° of space/visual field in horizontal and vertical axes. To create the perception of smooth motion, the 1250 ms of pink noise was fragmented into 15 equal length pieces with each 83.333 ms fragment being played every two speakers, and moved one speaker at a time, from outer left to outer right (rightward motion), or vice-versa for the leftward

motion. For example, for the rightward sweep, sound was played through speakers located at -60° and -52° consecutively, followed by -44°, and so on. A similar design was used for the vertical axis. This resulted in participants perceiving moving sweeps covering an arc of 120° in 1250 ms (speed = 96°/s; 50 ms fade in/out) containing the same sounds for all four directions. The choice of the movement speed of the motion stimuli aimed to create listening experience relevant to everyday-life conditions. Moreover, at such velocity it has been demonstrated that human listeners are not able to make the differences between concatenated static stimuli from motion stimuli elicited by a single moving object (Poirier et al., 2017), supporting the subject's report that our stimuli were perceived as smoothly moving (no perception of successive snapshots). In the static condition, the same pink noise was presented separately at one of 4 locations: left, right, up, and down. Static sounds were presented at the second most outer speakers (-56° and +56° in the horizontal axis, and +56° and -56° in the vertical axis) in order to avoid possible reverberation difference at the outermost speakers. The static sounds were fixed at one location at a time instead of presented in multiple locations (Smith et al., 2008; 2010; Poirier et al., 2017 Krumbholz et al., 2005). This strategy was purposely adopted for three main reasons. First, randomly presented static sounds can evoke auditory apparent motion (Strybel & Neale 1994; Lakatos et al., 1997; see Carlile 2016 for review). Second, in the visual domain, reducing the coherence of motion signals sometimes evoke higher activity compared to 100% coherent moving stimuli in hMT+/V5 region (Mckeefry et al., 1997; Vachon et al., 2009). One possible explanation is that incoherent moving stimuli are actually more difficult to process and therefore engage more attention compared to the 100% coherent moving directions which in turns triggers enhanced activity in motion selective regions (Vachon et al., 2009). Lastly, presenting static sounds located in a given position and moving sounds directed toward the same position allowed us to investigate a possible shared computation between moving and static sounds using cross-condition decoding (see below), which would have been impossible if the static sounds where randomly moving in space.

Before the recordings, the sound pressure levels (SPL) were measured from the subject's head position and ensured that each speaker conveys 65dB-A SPL. All participants reported strong sensation of auditory motion and were able to detect locations with high accuracy (see Fig 1C). Throughout the experiment, participants were blindfolded. Stimuli recordings were conducted in a session that lasted approximately 10 minutes, requiring the participant to remain still during this period.

Auditory experiment

Auditory stimuli were presented via MR-compatible closed-ear headphones (Serene Sound, Resonance Technology; 500-10KHz frequency response) that provided average ambient noise cancellation of about 30 dB-A. Sound amplitude was adjusted according to each participant's comfort level. To familiarize the participants with the task, they completed a practice session outside of the scanner until they reached above 80% accuracy.

Each run consisted of the 8 conditions (4 motion and 4 static) randomly presented using a block-design. Each condition was presented for 15 s (12 repetition of 1250 ms sound, no ISI) and followed by 7 s gap for indicating the corresponding direction/location in space and 8 s of silence (total inter-block interval was 15 s). The ramp applied at the beginning and at the end of each sound creates static bursts and minimized adaptation to the static sounds. During the response gap, participants heard a voice saying "left", "right", "up", and "down" in pseudo-randomized order. Participants were asked to press a button with their right index finger when the auditory block's direction or location was matching with the auditory cue (Figure 1B). The number of targets and the order (position 1-4) of the correct button press were balanced across conditions. This procedure was adopted to ensure that the participants gave their response using equal motor command for each condition and to ensure the response is produced after the end of the stimulation period for each condition. Each scan consisted of one block of each condition, resulting in a total of 8 blocks per run, with each run lasting 4 m 10

s. Participants completed a total of 12 runs. The order of the blocks was pseudo-randomized within each run, and across participants.

Based on pilot experiments, we decided to not rely on a sparse-sampling design as sometimes done in the auditory literature in order to present the sounds without the scanner background noise (Hall et al., 1999). These pilot experiments showed that the increase in the signal to noise ratio potentially provided by sparse sampling did not compensate for the loss in the number of volume acquisitions. Indeed, pilot recordings on participants not included in the current sample showed that, given a similar acquisition time between sparse-sampling designs (several options tested) and continuous acquisition, the activity maps elicited by our spatial sounds contained higher and more reliable beta values using continuous acquisition.

fMRI data acquisition and analyses

Imaging parameters

Functional and structural data were acquired with a 4T Bruker MedSpec Biospin MR scanner, equipped with an 8-channel head coil. Functional images were acquired with T2*-weighted gradient echo-planar sequence. Acquisition parameters were: repetition time of 2500 ms, echo time of 26 ms, flip angle of 73°, a field of view of 192 mm, a matrix size of 64 x 64, and voxel size of 3 x 3 x 3 mm^3. A total of 39 slices were acquired in ascending feet-to-head interleaved order with no gap. The three initial scans of each acquisition run were discarded to allow for steady-state magnetization. Before every two EPI run, we performed an additional scan to measure the point-spread function (PSF) of the acquired sequence, including fat saturation, which served for distortion correction that is expected with high-field imaging (Zeng and Constable, 2002).

High-resolution anatomical scan was acquired for each subject using a T1-weighted 3D MP-RAGE sequence (176 sagittal slices, voxel size of 1×1×1 mm^3; field of view 256 x 224 mm; repetition time = 2700 ms; TE = 4.18 ms; FA: 7°; inversion time: 1020 ms). Participants were blindfolded and instructed to lie still

during acquisition and foam padding was used to minimize scanner noise and head movement.

Univariate fMRI analysis

Whole brain

Raw functional images were pre-processed and analysed with SPM8 (Welcome Trust Centre for Neuroimaging London, UK; http://www.fil.ion.ucl.ac.uk/spm/software/spm/) implemented in MATLAB R2014b (MathWorks). Before the statistical analysis, our preprocessing steps included slice time correction with reference to the middle temporal slice, realignment of functional time series, the coregistration of functional and anatomical data, spatial normalization to an echo planar imaging template conforming to the Montreal Neurological Institute space, and spatial smoothing (Gaussian kernel, 6 mm FWHM) were performed.

To obtain blood oxygen level-dependent (BOLD) activity related to auditory spatial processing, we computed single subject statistical comparisons with fixed-effect general linear model (GLM). In the GLM, we used eight regressors from each condition (four motion direction, four sound source location). The canonical double-gamma hemodynamic response function implemented in SPM8 was convolved with a box-car function to model the above mentioned regressors. Motion parameters derived from realignment of the functional volumes (3 translational motion and 3 rotational motion parameters), button press, and the four auditory response cue events were modeled as regressors of no interest. During the model estimation, the data were high-pass filtered with cut-off 128s to remove the slow drifts/ low-frequency fluctuations from the time series. To account for serial correlation due to noise in fMRI signal, autoregressive (AR (1)) was used.

In order to obtain activity related to auditory processing in the whole brain, the contrasts tested the main effect of each condition ([Left Motion], [Right Motion], [Up Motion], [Down Motion], [Left Static], [Right Static], [Up Static], [Down Static]). To find brain regions responding preferentially to the auditory motion and static, we combined all motion conditions [Motion] and all static conditions [Static]. The

contrasts tested the main effect of each condition ([Motion], [Static]), and comparison between the conditions ([Motion > Static], and [Static > Motion]). These linear contrasts generated statistical parametric maps (SPM[T]) which were further spatially smoothed (Gaussian kernel 8 mm FWHM) and entered in a second-level analysis, corresponding to a random effects model, accounting for inter-subject variance. One-sample t-tests were run to characterize the main effect of each condition ([Motion], [Static]), and the main effect of motion processing ([Motion > Static]) and static location processing ([Static > Motion]). Statistical inferences were performed at a threshold of $p<0.05$ corrected for multiple comparisons (Family-Wise Error corrected; FWE) either over the entire brain volume or after correction for multiple comparisons over small spherical volumes (12 mm radius) located in regions of interest (SVC). Significant clusters were anatomically labeled using the xjView Matlab toolbox (http://www.alivelearn.net/xjview) or structural neuroanatomy information provided in the Anatomy Toolbox (Eickhoff et al., 2007).

Region of interest analysis

ROI Definition

Due to the hypothesis-driven nature of our study we defined hPT as an *a priori* region of interest for statistical comparisons and in order to define the volume in which we performed multivariate pattern classification analyses.

To avoid any form of double dipping that may arise when defining the ROI based on our own data, we decided to independently define hPT, using a meta-analysis method of quantitative reverse inference, implemented via the online tool Neurosynth (Yarkoni et al., 2011) using the term query "Planum Temporale". Rather than showing which regions are disproportionately reported by studies where a certain term is dominant (forward inference; P (activation | term)), this method identifies regions whose report in a neuroimaging study is diagnostic of a certain term being dominant in the study (reverse inference; P (term | activation)). As such, the definition of this ROI was based on a set of 85 neuroimaging studies at the moment of the query (September 2017). This method provides an independent method to obtain masks for further region-of-interest analysis. The peak coordinate

from the meta-analysis map was used to create a 6 mm spheres (117 voxels) around the peak z-values of hPT (peak MNI coordinates [-56 -28 8] and [60 -28 8]; lhPT and rhPT hereafter, respectively).

Additionally, hPT regions of interest were individually defined using anatomical parcellation with FreeSurfer (http://surfer.nmr.mgh.harvard.edu). The individual anatomical scan was used to perform cortical anatomical segmentation according to the Destrieux (Destrieux et al., 2010) atlas. We selected planum temporale label defined as [G_temp_sup-Plan_tempo, 36] bilaterally. We equated the size of ROI across participants and across hemispheres to 110 voxels (each voxel being 3mm isotropic). For anatomically defined hPT ROIs, all further analyses were carried out in subject space for enhanced anatomical precision and to avoid spatial normalization across subjects. We replicated our pattern of results in anatomically defined parcels (left and right hPT) obtained from the single subject brain segmentation (see Supplemental Info, Figure S1-3).

ROI Analyses

Univariate

The beta parameter estimates of the 4 motion directions and 4 sound source locations were extracted from lhPT and rhPT regions (Fig 2C). In order to investigate the presence of motion directions/sound source locations selectivity and condition effect in hPT regions, we performed a 2 Conditions (motion, static) x 4 Orientations (left, right, down, and up) repeated measures ANOVA in each hemisphere separately on these beta parameter estimates. Statistical results were then corrected for multiple comparisons (number of ROIs x number of tests) using the false discovery rate (FDR) method (Benjamini and Yekutieli, 2001). A Greenhouse–Geisser correction was applied to the degrees of freedom and significance levels whenever an assumption of sphericity was violated.

ROI - Multivariate pattern analyses

Within Condition Classification

Four-class and binary classification analyses were conducted within the hPT region in order to investigate the presence of auditory motion direction and sound source location information in this area. To ensure that the number of voxels was identical across subjects an ANOVA-based feature selection was performed to select the 110 voxels within each ROI, which are most informative/discriminative across all motion and static conditions (Cox & Savoy 2003; Haxby et al., 2001; Norman et al., 2006).

Multivariate pattern analyses (MVPA) were performed in the lhPT and rhPT. Preprocessing steps were identical to the steps performed for univariate analyses, except for functional time series that were smoothed with a Gaussian kernel of 2 mm (FWHM). MVPA was performed in CoSMoMVPA (http://www.cosmomvpa.org/; (Oosterhof et al., 2016), which implements LIBSVM software (http://www.csie.ntu.edu.tw/~cjlin/libsvm). A general linear model was implemented in SPM8, where each block was defined as a regressor of interest. A beta map was calculated for each block separately. Two multi-class and six binary linear support vector machine (SVM) classifiers with a linear kernel with a fixed regularization parameter of C = 1 were trained and tested for each participant separately. The two multi-class classifiers were trained and tested to discriminate between the response patterns of the 4 auditory motion directions and locations, respectively. Four binary classifiers were used to discriminate brain activity patterns for motion and location within axes (left vs. right motion, left vs. right static, up vs. down motion, up vs. down static, hereafter within axis classification). We used 8 additional classifiers to discriminate across axes (Left vs. Up, Left vs. Down, Right vs. Up, and Right vs. Down motion directions, Left vs. Up, Left vs. Down, Right vs. Up, and Right vs. Down sound source locations, hereafter across axes classification).

For each participant, the classifier was trained using a cross-validation leave-one-out procedure where training was performed with n-1 runs and testing was then applied to the remaining one run. In each cross-validation fold, the beta maps in the training set were normalized (z-scored) across conditions, and the estimated parameters were applied to the test set. To evaluate the performance of the classifier and its generalization across all the data, the previous step was repeated

12 times where in each fold a different run was used as the testing data and the classifier was trained on the other 11 runs. For each region per subject, a single classification accuracy was obtained by averaging the accuracies of all cross-validation folds.

Cross-condition Classification

To test whether motion directions and sound source locations share a similar neural representation in hPT region, we performed cross-condition classification. We carried out the same steps as for the within-condition classification as described above but trained the classifier on sound source locations and tested on motion directions, and vice versa. The accuracies from the two cross-condition classification analyses were averaged. For interpretability reasons, cross-condition classification was only interpreted on the stimuli categories that the classifiers discriminated reliably (above chance level) for both motion and static conditions (e.g. if discrimination of left vs. right was not successful in one condition, either static or motion, then the left vs. right cross-condition classification analysis was not carried out).

Across-condition Classification

To foreshadow our results, cross-condition classification analyses (see previous section) showed that motion directions and sound source locations share, at least partially, a similar neural representation in hPT region. To further investigate the similarities/differences between the neural patterns evoked by motion directions and sound source locations in the hPT, we performed 4 binary classifications: leftward motion vs. left static, rightward motion vs. right static, upward motion vs. up static, and downward motion vs. down static. The mean of the four binary classifications was computed to produce one accuracy score per ROI. Prior to performing the across-condition and cross-condition MVPA, each individual pattern was normalised separately across voxels so that any cross or across-condition classification could not be due to global univariate activation differences across the conditions.

Statistical significance

Statistical significance in the multivariate classification analyses was assessed using non-parametric tests permuting condition labels and bootstrapping (Stelzer et al., 2013). Each permutation step included shuffling of the condition labels and re-running the classification 100 times on the single-subject level. Next, we applied bootstrapping procedure in order to obtain a group-level null distribution that is representative of whole group. From each subject's null distribution one value was randomly chosen and averaged across all the subjects. This step was repeated 100,000 times resulting in a group level null distribution of 100,000 values. The classification accuracies across subjects we considered as significant if the $p<0.05$ after corrections for multiple comparisons using the FDR method (Benjamini and Yekutieli, 2001).

Similar approach was adopted to assess significance difference between the two auditory conditions (e.g. left motion vs left static). We performed additional permutation tests (100,000 iterations) by building a null distribution for t-stats after randomly shuffling the classification accuracy values across two conditions, and re-calculating the two-tail t-test between the classification accuracies of motion and static conditions. All p-values were corrected for multiple comparisons using the FDR method (Benjamini and Yekutieli, 2001).

Representation Similarity analysis

Neural Dissimilarity matrices

We employed representation similarity analysis (RSA; Kriegeskorte et al., 2008) to characterize the degree of shared representation between motion directions and sound source locations in hPT region. The RSA was performed using CosmoMVPA toolbox (Oosterhof et al., 2016) implemented in MATLAB. To perform this analysis we first extracted in each subject the activity patterns associated with each condition (Edelman et al., 1998; Haxby et al., 2001). Then, we averaged individual subject statistical maps (i.e. activity patterns) in order to have a mean pattern of activity for each condition across runs. Finally, we used Pearson's linear

correlation as the similarity measure to compare each possible pair of the activity patterns evoked by the four different motion directions and four different sound source locations. This resulted in an 8 x 8 correlation matrix for each participant that was then converted into a representational dissimilarity matrix (RDMs) by computing 1 – correlation. Each square of the RDM contains the dissimilarity index between the patterns of activity generated by two conditions, in other words the RDM represents how different is the neural representation of each condition from the neural representations of all the other condition in the selected ROI. The 16 neural RDMs (1 per subject) for each of the 2 ROIs were used as neural input for RSA.

Computational models

To investigate shared representations between auditory motion directions and sound source locations, we created multiple computational models ranging from a fully condition-distinct model to a fully condition-invariant model with intermediate gradients in between (Zabicki et al., 2016).

Condition-Distinct model

The condition-distinct models assume that dissimilarities between motion and static condition is 1 (i.e. highly dissimilar), meaning that neural responses/patterns generated by motion and static conditions are totally unrelated. For instance, there would be no similarity between any motion directions with any sound source location. The dissimilarity values in the diagonal were set to 0, simply reflecting that neural responses for the same direction/location are identical to themselves.

Condition-Invariant model

The condition-invariant models assume a fully shared representation for specific/corresponding static and motion conditions. For example, the models consider the neural representation for the left sound source location and the left motion direction highly similar. All within-condition (left, right, up and down)

comparisons are set to 0 (i.e. highly similar) regardless of their auditory condition. The dissimilarities between different directions/locations are set to 1 meaning that each within condition sound (motion or static) is different from all the other within conditions.

Intermediate models

To detect the degree of similarity/shared representation between motion direction and sound source location patterns, we additionally tested 2 classes of 5 different intermediate models. The two classes were used to deepen the understanding of characteristic tuning of hPT for separate direction/location or axis of motion/location. The two model classes represent 2 different possibilities. The first scenario was labeled as Within-Axis Distinct, and these models assume that each of the 4 directions/locations (i.e. left, right, up, down) would generate a distinctive neural representation different from all of the other within-condition sounds (e.g. the patterns of activity produced by the left conditions are highly different from the patterns produced by right, up and down conditions) (see Figure 4C, upper panel). To foreshadow our results, we observed preference for axis of motion in MVP-classification, therefore we created another class of models to further investigate neural representations of within-axis and across-axes of auditory motion/space. The second scenario was labeled with Within-Axis Combined, and these models assume that opposite direction/locations within the same axis would generate similar patterns of activity (e.g. the pattern of activity of horizontal (left and right) conditions are different from the patterns of activity of vertical conditions (up and down) (see Figure 4C, lower panel).

In all intermediate models, the values corresponding to the dissimilarity between same auditory spaces (e.g. left motion and left location) were gradually modified from 0.9 (motion and static conditions are mostly distinct) to 0.1 (motion and static conditions mostly similar). These models were labeled M9, 7, 5, 3, and 1 respectively.

In all condition-distinct and intermediate models, the dissimilarity of within-condition sounds was fixed to 0.5 and dissimilarity of across-condition sounds was fixed to 1. Across all models, the diagonal values were set to 0.

Performing RSA

We computed Pearson's correlation to compare neural RDMs and computational model RDMs. The resulting correlation captures which computational model better explains the neural dissimilarity patterns between motion direction and sound source location conditions. To compare computational models, we performed Mann-Whitney-Wilcoxon rank-sum test between for every pair of models. All p-values were then corrected for multiple comparisons using the FDR method (Benjamini and Yekutieli, 2001). To test differences between two classes of models (Within-Axis Combined vs. Within-Axis Distinct), within each class, the correlation coefficient values were averaged across hemispheres and across models. Next, we performed permutation tests (100,000 iterations) by building a null distribution for differences between classes of models after randomly shuffling the correlation coefficient values across two classes, and re-calculating the subtraction between the correlation coefficients of Within-Axis Combined and Within-Axis Distinct classes.

To visualize the distance between the patterns of the motion directions and sound source locations, we used multi-dimensional scaling (MDS) to project the high-dimensional RDM space onto 2 dimensions with the neural RDMs that were obtained from both lhPT and rhPT.

Additionally, the single-subject 8 x 8 correlation matrices were used to calculate the reliability of the data considering the signal-to-noise ratio of the data (Kriegeskorte et al., 2007). For each participant and each ROI, the RDM was correlated with the averaged RDM of the rest of the group. The correlation values were then averaged across participants. This provided the maximum correlation that can be expected from the data.

RESULTS

Behavioral results

During the experiment, we collected target direction/location discrimination responses (see Figure 1C). The overall accuracy scores were entered into 2 x 4 (Condition, Orientation) repeated measures ANOVA. No main effect of Condition ($F_{1,15}$ = 2.22; p = 0.157) was observed, indicating that the overall accuracy while detecting direction of motion or sound source location did not differ. There was a significant main effect of orientation ($F_{1,15}$ = 11.688; p < 0.001), caused by greater accuracy in the horizontal orientations (left and right) as compared to the vertical orientations (up and down). Post-hoc two-tailed t-tests (Bonferroni corrected) revealed that accuracies did not reveal significant difference within horizontal orientations (left vs right; t_{15} = -0.15, p=1), and within vertical orientations (up vs down; t_{15} = 0.89, p=1). However, left orientation accuracy was greater as compared to down (t_{15} = 3.613, p=0.005), and up (t_{15} = 4.51, p<0.001) orientations and right orientation accuracy was greater as compared to the down (t_{15} = 3.76, p=0.003) and up (t_{15} = 4.66, p<0.001) orientation accuracies. No interaction between Condition x Orientation was observed, pointing out that differences between orientations in terms of performance expresses both for static and motion.

fMRI results – whole-brain univariate analyses

To identify brain regions that are preferentially recruited for auditory motion processing, we performed a univariate RFX- GLM contrast [Motion > Static] (Figure 2A). Consistent with previous studies (Dormal et al., 2016; Getzmann and Lewald, 2012; Pavani et al., 2002; Poirier et al., 2005; Warren et al., 2002), whole-brain univariate analysis revealed activation in the superior temporal gyri, bilateral hPT, precentral gyri, and anterior portion of middle temporal gyrus in both hemispheres (Figure 2A, Table 1). The most robust activation (resisting whole brain FWE correction, p<0.05) was observed in the bilateral hPT (peak MNI coordinates [-46 -32 10] and [60 -32 12]). We also observed significant activation in occipito-temporal regions (in the vicinity of anterior hMT+/V5) as suggested by previous studies (Dormal et al., 2016; Poirier et al., 2005; Warren et al., 2002).

fMRI results – ROI univariate analyses

Beta parameter estimates were extracted from the pre-defined ROIs (see methods) for the four motion directions and four sound source locations from the auditory experiment (Figure 2C). We investigated the condition effect and the presence of direction/location selectivity in lhPT and rhPT regions separately by performing 2 x 4 (Conditions, Orientations) repeated measures of ANOVA with beta parameter estimates. In lhPT, main effect of Conditions was significant ($F_{1,15}$ = 37.28, $p < 0.001$), indicating that auditory motion evoked higher response compared to static sounds. There was no significant main effect of Orientations ($F_{1.5,22.5}$ = 0.771, $p = 0.4$), and no interaction ($F_{3,45}$ = 2.21, $p = 0.11$). Similarly, in rhPT, only main effect of Conditions was significant ($F_{1,15}$ = 37.02, $p < 0.001$). No main effect of Orientation ($F_{1.5,23.2}$ = 1.43, $p = 0.3$) or interaction ($F_{3,45}$ = 1.73, $p = 0.19$) was observed. Overall, brain activity in the hPT as measured with beta parameter estimate extracted from univariate analysis did not provide evidence of motion direction or sound source location selectivity.

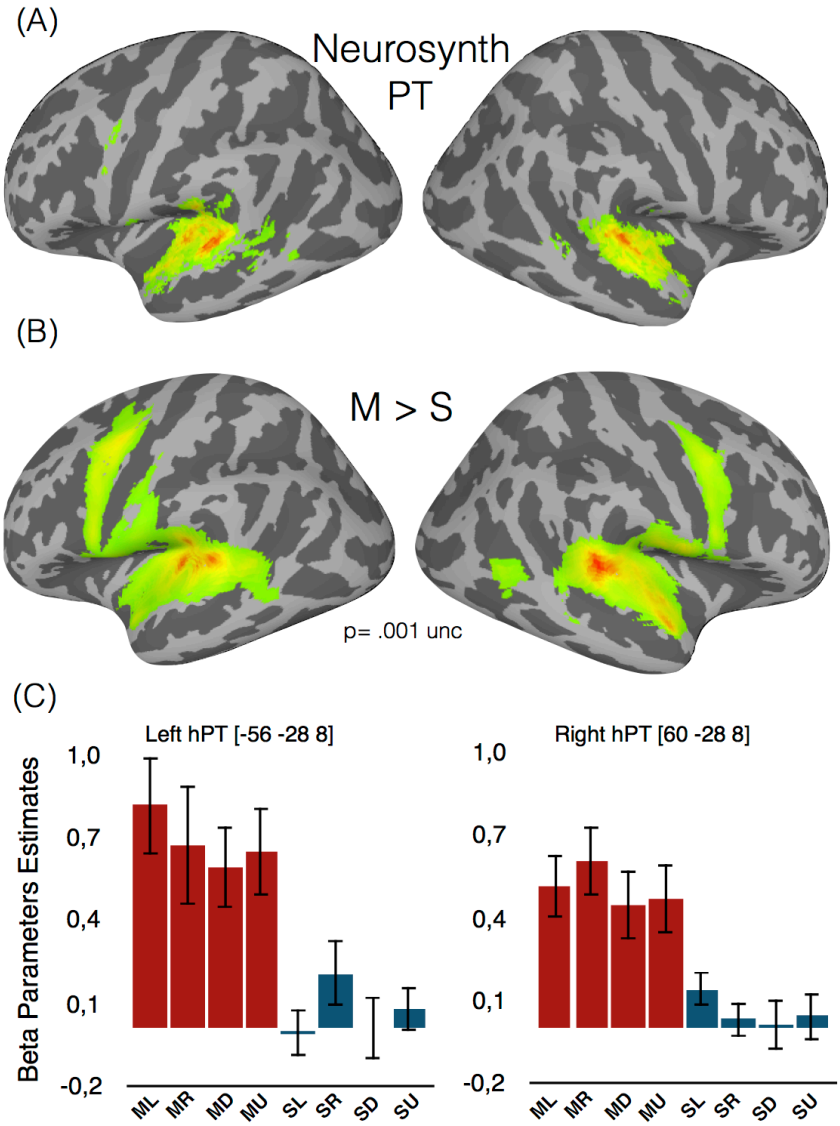

Figure 2. Univariate whole brain results. (A). Reverse inference map was obtained from the online tool Neurosynth using the term "Planum Temporale" (FDR corrected p <0.05). The black spheres are illustration of drawn mask (radius = 6mm, 117 voxels) around the peak coordinate from Neurosynth (search term "planum temporale", meta-analysis of 85 studies). **(B).** Auditory motion processing [motion > static] at p <0.001 uncorrected. **(C).** Mean activity estimates (arbitrary units ± SEM) associated with the perception of auditory motion direction (red) and sound-source location (blue). ML: motion left, MR: motion right, MD: motion down, MU: motion up, SL: static left, SR: static right, SD: static down, and SU: static up.

Table 1. Results of the univariate analyses for the main effect of auditory motion processing [motion > static], and auditory localization processing [static > motion]. Coordinates reported in this table are significant ($p < 0.05$ FWE) after correction over small spherical volumes (SVC, 12 mm radius) of interest (#) or over the whole brain (*). Coordinates used for correction over small spherical volumes are as follows (x, y, z, in MNI space): left middle temporal gyrus (hMT+/V5) [-42 -64 4] (Dormal et al., 2016), right middle temporal gyrus (hMT +/V5) [42 − 60 4] (Dormal et al., 2016), right superior frontal sulcus [32 0 48] (Collignon et al., 2011), right middle occipital gyrus [48 -76 6] (Collignon et al., 2011). K represents the number of voxels when displayed at p(unc) < 0.001. L: left, R: right, G: gyrus, S: sulcus.

Area	k	x (mm)	y (mm)	z (mm)	Z	p
MOTION > STATIC						
L planum temporale	10506	-46	-32	10	6.63	0.000*
L Middle Temporal G		-56	-38	14	6.10	0.000*
L Precentral G		-46	-4	52	5.25	0.004*
L Putamen		-22	2	2	4.98	0.011*
L Middle Temporal G	43	-50	-52	8	3.79	0.01#
R Superior Temporal G	7074	66	-36	12	6.44	0.000*
R Superior Temporal G		62	-2	-4	5.73	0.000*
R Superior Temporal G		52	-14	0	5.56	0.001*
R Precentral G		50	2	50	4.70	0.032*
R Superior Frontal S	159	46	0	50	4.40	0.001#
R Middle Temporal G	136	42	-60	6	4.31	0.001#
R Middle Occipital G	24	44	-62	6	3.97	0.006#

fMRI results – ROI multivariate pattern analyses

To further investigate the presence of information about auditory motion direction and sound source location in hPT, we ran multi-class and binary multivariate pattern classification. Figure 3A-D shows the mean classification accuracy across categories in each ROI.

MVPA – Within Condition

Multi-class across four conditions classification accuracy in the hPT was significantly above chance (chance level = 25%) in both hemispheres for motion direction (lhPT: mean ± SD = 38.4 ±7, $p<0.001$; rhPT: mean ± SD = 37.1 ± 6.5, $p<0.001$), and sound source location (lhPT: mean ± SD = 32.4 ±6.7, $p<0.001$; rhPT: mean ± SD = 31.2 ± 7.5, $p<0.001$). In addition, we assessed the differences between classification accuracies for motion and static stimuli by using permutation tests in lhPT ($p = 0.0238$) and rhPT ($p = 0.0236$), indicating greater accuracies for classifying motion

direction than sound source location across all regions.

MVPA – Binary Within-Axis

Binary horizontal (left vs. right) within-axis classification showed significant results in both lhPT and rhPT for static sounds (lhPT: mean ± SD = 58.6 ±14.5, $p<0.001$; rhPT: mean ± SD = 56.5 ± 11.9, $p = 0.008$), while motion classification was significant only in the rhPT (mean ± SD = 55.5 ± 13.9, $p = 0.018$). Moreover, binary vertical (up vs down) within-axis classification was significant only in the lhPT for both motion (mean ± SD = 55.7 ± 7.1, $p = 0.01$), and static (mean ± SD = 54.9 ± 11.9, $p = 0.03$) conditions. For multiple-comparisons, FDR-correction performed on p values.

MVPA – Binary Across-Axis

8 additional classifiers to discriminate across axes

We used 8 additional binary classifiers to discriminate across axes moving and static sounds. Binary across-axes (Left vs. Up, Left vs. Down, Right vs. Up, and Right vs. Down) classifications were significantly above chance level in the bilateral hPT for motion directions (Left vs. Up lhPT: mean ± SD = 65.8 ±14.8, $p<0.001$; rhPT: mean ± SD = 64.8 ± 9.4, $p<0.001$; Left vs. Down lhPT: mean ± SD = 74 ±15.9, $p<0.001$; rhPT: mean ± SD = 66.9 ± 10.4, $p<0.001$; Right vs. Up lhPT: mean ± SD = 72.4 ±15.8, $p<0.001$; rhPT: mean ± SD = 71.4 ± 13.3, $p<0.001$; Right vs. Down lhPT: mean ± SD = 73.4 ±11.8, $p<0.001$; rhPT: mean ± SD = 68.2 ± 15.9, $p<0.001$). However, sound source locations across-axes classifications were significant only in certain conditions (Left vs. Up rhPT: mean ± SD = 56 ± 12.7, $p=0.018$; Left vs. Down lhPT: mean ± SD = 55.2 ±11.2, $p=0.024$; rhPT: mean ± SD = 57.3 ± 17.2, $p=0.003$; Right vs. Down lhPT: mean ± SD = 57.6 ±8.4, $p=0.005$). For multiple-comparisons, FDR-correction performed on p values.

MVPA – "Axis of Motion" Preference

In order to test whether neural patterns within hPT contain information about opposite directions/locations within an axis, we performed four binary within-axis

classifications. Similar multivariate analyses were performed to investigate the presence of information about across-axes directions/locations. The classification accuracies were plotted in Figure 3B-C.

In motion direction classifications, to assess the statistical difference between classification accuracies of across axes (left vs. up, left vs. down, and right vs. up, right vs. down) and within axes (left vs. right, and up vs. down) directions, we performed pairwise permutation tests and FDR-corrected for multiple comparisons. Across-axes classification accuracies in lhPT ([left vs. up] vs. [left vs. right]: $p=0.006$, [left vs. down] vs. [left vs. right]: $p<0.001$, [right vs. down] vs. [left vs. right]: $p<0.001$, [right vs. up] vs. [left vs. right]: $p=0.001$), and rhPT ([left vs. up] vs. [left vs. right]: $p=0.029$, [left vs. down] vs. [left vs. right]: $p=0.014$, [right vs. down] vs. [left vs. right]: $p=0.02$, [right vs. up] vs. [left vs. right]: $p=0.003$) were significantly higher compared to the horizontal within-axis classification accuracies. Similarly, across-axes classification accuracies were significantly higher when compared with vertical within-axis classification accuracies in lhPT ([up vs. down] vs. [left vs. up], $p=0.02$; [up vs. down] vs. [left vs. down], $p=0.001$; [up vs. down] vs. [right vs. up], $p=0.001$; [up vs. down] vs. [right vs. down], $p <0.001$) and rhPT ([up vs. down] vs. [left vs. up], $p= 0.001$; [up vs. down] vs. [left vs. down], $p=0.001$; [up vs. down] vs. [right vs. up], $p= 0.001$; [up vs. down] vs. [right vs. down], $p=0.002$). No significant difference was observed between the within-axis classifications in lhPT ([left vs. right] vs. [up vs. down], $p=0.24$) and rhPT ([left vs. right] vs. [up vs. down], $p=0.31$). Similarly, no significance difference was observed among the across-axes classification accuracies in the bilateral hPT.

In static sound location classifications, no significant difference was observed between across-axes and within-axes classification accuracies, indicating that the classifiers did not perform better when discriminating sound source locations across axes compared to the opposite locations.

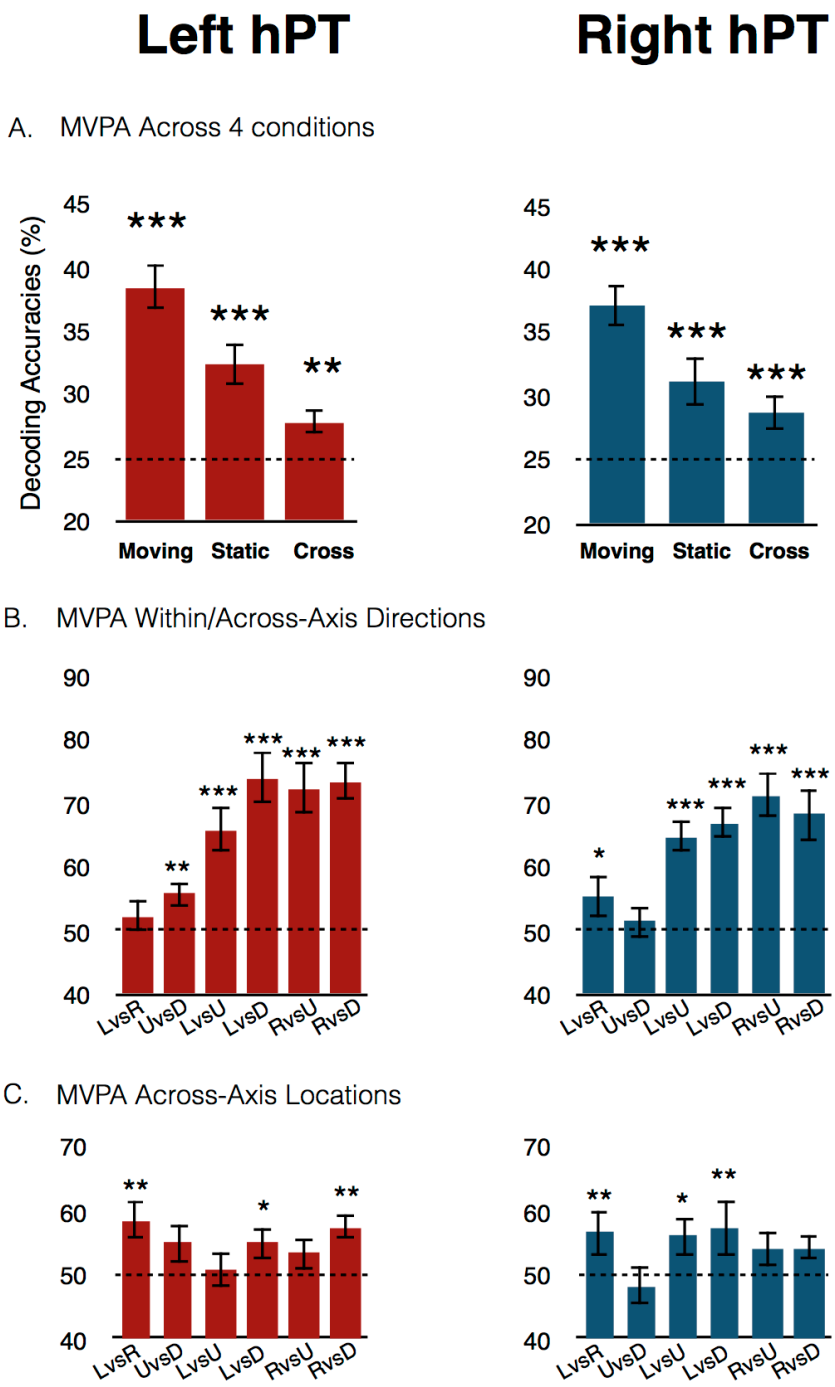

Figure 3. Within and cross-classification results. (A). Classification results for the 4 conditions in functionally defined hPT region. Within-condition and cross-condition classification results are shown in the same bar plots. Moving: four motion direction; Static: four sound source location; and Cross: cross-condition classification accuracies. **(B).** Classification results of within (left vs. right, up vs. down) and across axes (left vs. up, left vs. down, right vs. up, right vs. down) motion directions. **(C).** Classification results of within (left vs. right, up vs. down) and across axes (left vs. up, left vs. down, right vs. up, right vs. down) sound source locations. LvsR: Left vs. Right, UvsD: Up vs. Down, LvsU: Left vs. Up, LvsD: Left vs. Down, RvsU: Right vs. Up, RvsD: Right vs. Down classifications. FDR corrected p-values: (*) $p<0.05$, (**) $p<0.01$, (***) $p<0.001$ testing differences against chance level (dotted lines; see methods).

One may wonder whether the higher classification accuracy for across compared to within axes relates to the perceptual differences in discriminating sounds within the horizontal and vertical axes. Indeed, because we opted for an ecological design reflecting daily-life listening condition, we observed, as expected, that discriminating vertical directions was more difficult than discriminating horizontal ones (Middlebrooks and Green, 1991). In order to address this issue, we replicated our MVP-classification analyses after omitting the four subjects showing the lowest performance in discriminating the vertical motion direction, leading to comparable performance (at the group level) within and across axes. We replicated our pattern of results by showing preserved higher classification accuracies across-axes than within-axis (see Figure S4). Moreover, while accuracy differences between across- and within-axes classification was only observed in the motion condition, behavioral differences were observed in both static and motion conditions. To assess whether the higher across axes classification accuracies are due to difference in difficulties between horizontal and vertical sounds, we performed correlation analysis. For each participant, the behavioral performance difference between horizontal (left and right) and vertical (up and down) conditions was calculated. The difference between the classification accuracies in within axes (left vs. right and up vs. down) and across axes (left vs. up, left vs. down, right vs. up, and right vs. down) was calculated. Pearson's correlation between the (Δ) behavioral performance and (Δ) classification accuracies was not significant (see Figure S5). An absence of correlation suggests that there is no relation between the difference in classification accuracies and the perceptual difference. These results strengthen the notion that "axis of motion" organization observed in PT does not simply stem from behavioral performance differences.

These results strengthen the notion that the higher classification accuracies for axes of motion do not simply stem from behavioral performance differences.

MVPA – Cross-condition

To investigate if motion direction and sound source locations rely on shared representation in hPT, we trained the classifier to distinguish neural patterns from the motion directions (e.g. going to the left) and then tested on the patterns elicited by static conditions (e.g. being in the left), and vice versa.

Cross-condition classification revealed significant results on across 4 conditions (lhPT: mean ± SD = 27.8 ± 5.3, $p = 0.008$, rhPT: mean ± SD = 28.7 ± 3.8, $p<0.001$) and across axes (lhPT: mean ± SD = 57.6 ± 6.2, $p < 0.001$; mean ± SD = 58.8 ± 6.2, $p<0.001$). Within- axis categories did not reveal any significant cross-condition classification. These results suggest that a partial overlap between the neural patterns of moving and static stimuli in the hPT.

MVPA – Across-condition

Cross-condition classification results indicated a shared representation between motion directions and sound source locations. Previous studies argued that successful cross-condition classification reflects an abstract representation of stimuli conditions (Fairhall and Caramazza, 2013; Higgins et al., 2017; Hong et al., 2012). To test this hypothesis, patterns of the same orientation of motion and static conditions (e.g. leftward motion and left location) were involved in across-condition MVPA. The rational was that if the hPT region carries fully abstract representation of space, across-condition classification would give results in favor of the null hypothesis (no differences across conditions). In the across-condition classification analysis, accuracies from the four across-condition classification analyses were averaged and survived FDR corrections in bilateral hPT (lhPT: mean ± SD = 65.6 ± 5, $p < 0.001$, rhPT: mean ± SD = 61.9 ± 5.6, $p<0.001$), indicating that the neural patterns of motion direction can be reliably differentiated from sound-source location within hPT.

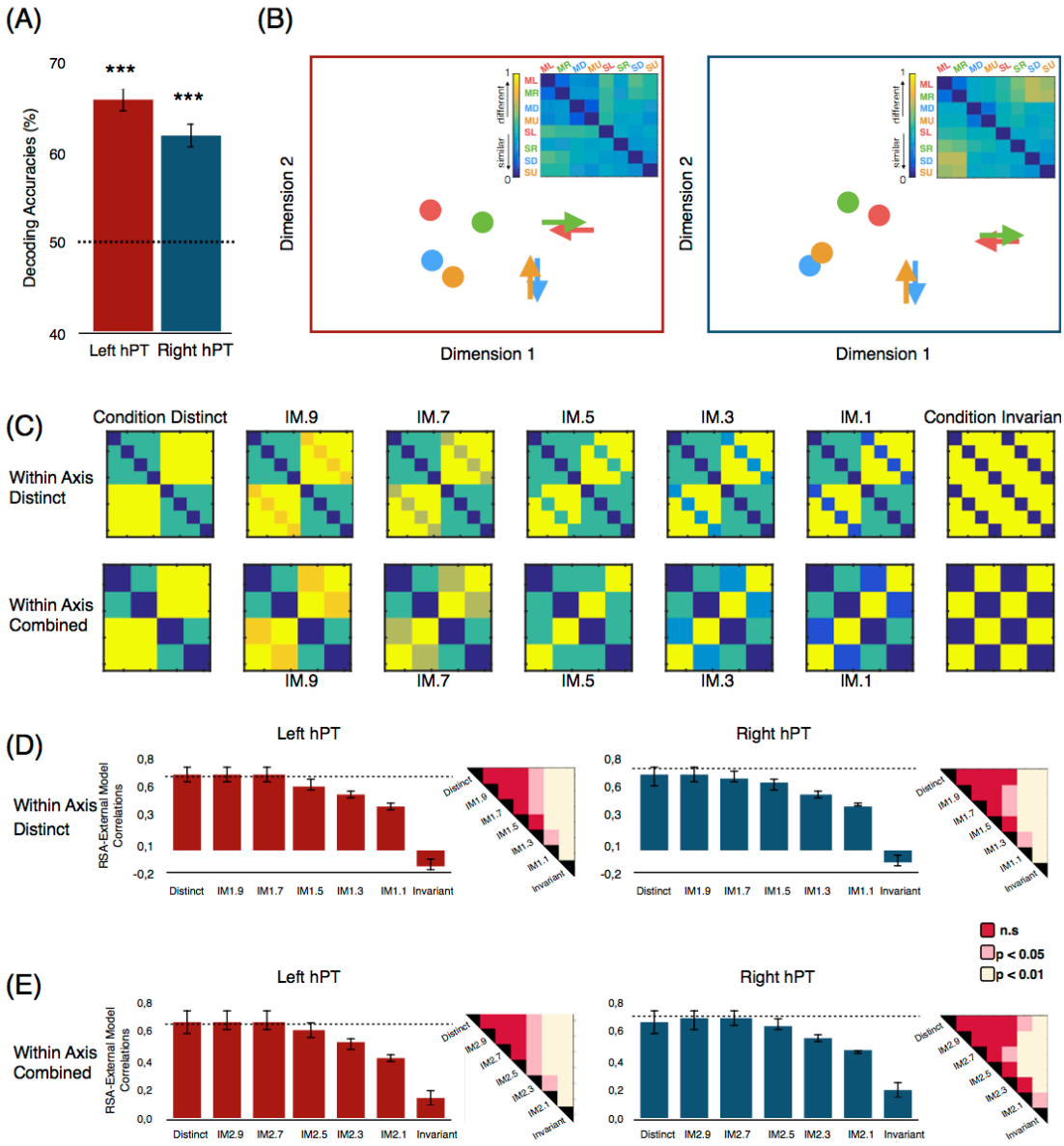

Figure 4. Pattern dissimilarity between motion directions and sound source locations. (A). Across-condition classification results of across 4 conditions are represented in each ROI (lhPT and rhPT). 4 binary classifications [leftward motion vs. left location], [rightward motion vs. right location], [upward motion vs. up location], and [downward motion vs. down location] were computed and averaged to produce one accuracy score per ROI. FDR corrected p-values: (***) p<0.001. Dotted lines represent chance level. **(B).** The embedded top panel shows neural RDMs extracted from left and right hPT, and multi-dimensional scaling (MDS) plot visualizes the similarities of the neural pattern elicited by 4 motion directions (arrows) and 4 sound source locations (dots). Color codes for arrow/dots are as follows: green indicates left direction/location, red indicates right direction/location, orange indicates up direction/location, and blue indicates down direction/location. ML: motion left, MR: motion right, MD: motion down, MU: motion up, SL: static left, SR: static right, SD: static down, and SU: static up. **(C-D).** The results of representational similarity analysis (RSA) in hPT are represented. **(C).** RDMs of the computational models that assume different similarities of the neural pattern based on auditory motion and static conditions. **(D).** RSA results for every model and each ROI. For each ROI, the dotted lines represent the reliability of the data considering the signal-to-noise ratio (see Materials and Methods), which provides an estimate of the highest correlation we can expect in a given ROI when correlating computational models and neural RDMs. Error bars indicate SEM. IM1: Intermediate models with within-axis conditions distinct, IM2: Intermediate model with within-axis conditions combined. Each right up corner of the bar plots show visualization of significant

differences for each class of models and hemispheres separately (Mann-Whitney-Wilcoxon rank-sum test, FDR corrected).

RSA

Multi-dimensional Scaling

Visualization of the representational distance between the neural patterns evoked by motion directions and sound source locations further supported that within-axis directions show similar geometry compared to the across-axes directions, therefore, it is more difficult to differentiate the neural patterns of opposite directions in MVP-classification. MDS also showed that in both lhPT and rhPT, motion directions and sound source locations are separated into 2 clusters (Figure 4B).

RSA with external models

The correlation between model predictions and neural RDMs for the lhPT and rhPT is shown in Figure 4D. The cross-condition classification results indicated a shared representation within the neural patterns of hPT for motion and static sounds. We examined the correspondence between the response pattern dissimilarities elicited by our stimuli with 14 different model RDMs that included a fully condition distinct, fully condition-invariant models, and intermediate models with different degrees of shared representation.

First class of computational RDMs were modeled with the assumption that the neural patterns of within-axis sounds are fully distinct. The analysis revealed a negative correlation with the fully condition-invariant model in the bilateral hPT (lhPT: mean r ± SD = -0.12 ± 0.18, rhPT: mean r ± SD = -0.01 ± 0.2) that increased gradually as the models progressed in the condition-distinct direction. The model that best fit the data was the M9 model in the bilateral hPT (lhPT: mean r ± SD = 0.66± 0.3, rhPT: mean r ± SD = 0,65 ± 0.3). A similar trend was observed for the second class of models that have the assumption of within-axis sounds evoke similar neural patterns. Condition-invariant model provided the least explanation of the

data (lhPT: mean r ± SD = 0.14 ± 0.25, rhPT: mean r ± SD = 0.2 ± 0.29), and correlations gradually increased as the models progressed in the condition-distinct direction. The winner models in this class were the models M9 in lhPT and M7 in the rhPT (lhPT: mean r ± SD = 0.67 ± 0.22, rhPT: mean r ± SD = 0.69 ± 0.15).

In addition, we assessed differences between correlation coefficients for computational models using Mann-Whitney-Wilcoxon rank-sum test for each classes of models and hemispheres separately (Figure 4E). In Within-Axis Distinct class in lhPT, pairwise comparisons of correlation coefficients did not show significant difference for [Condition Distinct vs. IM.9; p = 0.8], [Condition Distinct vs. IM7; p = 0.6; rhPT], [Condition Distinct vs. IM.5; p = 0.09], however as the models progressed in further condition invariant direction, the difference between correlation coefficients for models reached significance ([Condition Distinct vs. IM.3], p = 0.012; [Condition Distinct vs. IM.1], p = 0.007; [Condition Distinct vs. Condition Invariant], p <0.001), indicating Condition Distinct model provided stronger correlation compared to the models representing conditions similarly. In rhPT, the rank-sum tests between each pairs revealed no significant difference for [Condition Distinct vs. IM.9; p = 0.9], [Condition Distinct vs. IM7; p = 0.7], [Condition Distinct vs. IM.5; p = 0.3], and also [Condition Distinct vs. IM.3; p = 0.06], however as the models progressed in further condition invariant direction, the difference between correlation coefficients for models reached significance ([Condition Distinct vs. IM.1], p =0.006; [Condition Distinct vs. Condition Invariant], p <0.001).

The two classes of models were used to deepen our understanding of characteristic tuning of hPT for separate direction/location or axis of motion/location. To reveal the differences between Within-Axis Distinct and Within-Axis Combine classes of models, we relied on 2-sided signed-rank test. Within-Axis Combined models explained our stimuli space better than Within-Axis Distinct models supporting similar pattern representation within planes (p = 0.0023).

DISCUSSION

In line with previous studies, our univariate results demonstrate higher activity for moving than for static sounds in the superior temporal gyri, bilateral hPT, precentral gyri, and anterior portion of middle temporal gyrus in both hemispheres (Baumgart and Gaschler-Markefski, 1999; Krumbholz et al., 2005; Pavani et al., 2002; Poirier et al., 2005; Warren et al., 2002). The most robust cluster of activity was observed in the bilateral hPT (Figure 2B, Table 1). Moreover, activity estimates extracted from independently defined hPT also revealed higher activity for moving relative to static sounds in this region. Both whole-brain and ROI analyses therefore clearly indicated a functional preference (expressed here as higher activity level estimates) for motion processing over sound-source location in bilateral hPT regions (Figure 2).

Does hPT contain information about specific motion directions and sound source locations? At the univariate level, our four (left, right, up and down) motion directions and sound source locations did not evoke differential univariate activity in hPT region (see Fig. 2C). We then carried out multivariate pattern classification in order to investigate whether information related to motion directions and sound-source locations could be retrieved from the distributed activity reliably elicited by each separate condition across voxels of the hPT.

We observed that bilateral hPT contains reliable information about the four auditory motion directions (Figure 3). Our results therefore demonstrate that despite no univariate differences, area hPT contains reliable distributed information about separate directions of motions (Alink et al., 2012; Dormal et al., 2016; Jiang et al., 2014, 2016). Our results are therefore similar to the observations made with fMRI in the human visual motion area hMT+/V5 showing reliable direction-selective information despite comparable voxel-wise univariate activity levels across directions (Kamitani and Tong, 2006). Within-axis MVP-classification results revealed that even if both horizontal (left versus right), and vertical (up versus down) motion directions can be classified in the hPT region (see Fig. 3C-D across axes (e.g. left versus up) direction classifications revealed much higher decoding accuracies compared to within-axis classifications. These results indicate that classifying motion direction information is much more reliable across axis of motion, rather than

separate directions within horizontal (left versus right) or vertical (up versus down) axes. Such enhanced classification accuracy across axes versus within axis is reminiscent of observations made in MT+/V5 where a large-scale axis of motion selective organization was observed in non-human primates (Albright et al., 1984), and in human area MT+/V5 (Zimmermann et al., 2011). Further examination with RSA provided additional evidence that within-axis combined models (aggregating the opposite directions/location) explain better the neural representational space of hPT by showing higher correlations values compared to within-axis distinct models. These results strengthen the idea of representation of opposite directions/locations are relatively similar in the neural patterns of hPT (see Figure 4D-E).

The resemblance between our findings with the conclusions reached in the hMT+/V5 for visual motion (Zimmermann et al., 2011) suggests that the topographic organization principle of hMT+/V5 and hPT shows similarities in representing motion directions. The functional organization of the middle occipito-temporal region hMT+/V5 is characterized by columns containing neurons that react specifically to a certain visual motion direction (Albright et al., 1984). Those columns vary smoothly for certain motion direction but are also found running side by side with their respective opposing motion direction counterparts (Albright et al., 1984; Born and Bradley, 2005; Diogo et al., 2003; Geesaman et al., 1997; Zimmermann et al., 2011). By aggregating opposing motion directions, larger axis of motion features can be constructed that are more easily detectable with fMRI than individual direction selective columns (Zimmermann et al., 2011). Moreover, neural responses to opposite directions were suggested to play a role in encoding visual motion direction by triggering excitatory/inhibitory mechanism within hMT+/V5 (Heeger et al., 1999). Due to this topographic organization principle of area hMT+/V5, and probably in combination with excitatory/inhibitory activity features of the opposing motion directions, it has been suggested that the representation of preferred axis of motion is more systematic from the pattern of fMRI activity when compared to the opposite direction of motion (Zimmermann et al., 2011; but see below for alternative accounts). The observed motion opponent mechanism in visual motion area could also exist in hPT region and influence the

reliable axis of motion classification. A number of electrophysiological studies have indeed demonstrated the existence of motion direction sensitive neurons in the auditory cortex of mammals (Ahissar et al., 1992; Doan et al., 1999; Poirier et al., 1997) and showed higher spatial selectivity (sharper spatial tuning) in the caudal fields (homologue to area hPT) (Woods et al., 2006; Zhou and Wang, 2012).

Even if it has been proposed that successful classification may potentially stem from the spatial biases within each voxel that relates to the underlying cortical columnar organization or other types of direction selective signals (Bartels et al., 2008; Haynes and Rees, 2006; Kamitani and Tong, 2005), alternative explanations have also been provided. Indeed, if fMRI signal within a voxel would exclusively reflect a sampling of cortical columns, smoothing of the data would substantially decrease the classification accuracies due to averaging out the random biases in the neighboring voxels (Kamitani and Sawahata, 2010). Contrary to that, evidence points to no influence of smoothing (Op de Beeck, 2010). Studies conducted on early visual cortex proposed that classifying orientation preference reflects much larger scale (e.g. retinotopy) rather than columnar organization (Op de Beeck, 2010; Freeman et al., 2011, 2013; Beckett et al., 2012). Interestingly, high-field fMRI studies showed that the signal carries information related to both large- and fine-scale (columnar level) biases (Gardumi et al., 2016; Pratte et al., 2016; Sengupta et al., 2017). A study that investigated the effect of spatial resolution and smoothing on the classification accuracies on two different auditory tasks, concluded that the influence of large- and fine-scale spatial biases depends on the specific task of interest (Gardumi et al., 2016). These studies support the notion that MVP-classification results could reflect the combination of both large- and fine-scale organization. The present study sheds important new lights on the coding mechanism of motion direction within the hPT and demonstrates that fMRI signal in the hPT contains direction specific information and point toward an "axis of motion" organization. This result highlights intriguing similarities between the visual and auditory systems, arguing for common neural coding strategies of motion directions across the senses. However, further studies are needed to test the similarities between the coding mechanisms implemented in visual and auditory

motion selective regions, and more particularly, what drives the directional information captured with fMRI in the auditory cortex.

Supporting univariate motion selectivity results in bilateral hPT, MVPA revealed higher classification accuracies for moving than for static sounds (Fig. 3A-B). However, despite minimal univariate activity elicited by sound-source location in hPT, and the absence of reliable univariate differences in the activity elicited by each position (see Fig. 2C), within-axis MVP-classification results showed that beside the vertical axis (up versus down), sound source location information can be reliably decoded bilaterally in hPT (Figure 3C). Our results are in line with previous studies showing that posterior regions in auditory cortex exhibit location sensitivity both in animals (Recanzone, 2000; Stecker et al., 2005; Tian et al., 2001) and humans (Ahveninen et al., 2006, 2013; Brunetti et al., 2005; Deouell et al., 2007; Derey et al., 2016; Krumbholz et al., 2005; Warren and Griffiths, 2003; Zatorre et al., 2002). In contrast to what was observed for motion direction however, within-axis classifications did not reveal lower decoding accuracies when compared to across-axis classifications in hPT for static sounds. This indicates that auditory sound source locations might not follow similar topographic representations than the one observed for motion directions.

To which extend the neural representation of motion directions and sound source locations overlaps has long been debated (Grantham, 1986; Kaas et al., 1999; Poirier et al., 2017; Romanski et al., 2000; Smith et al., 2004, 2007; Zatorre and Belin, 2001). Despite the fact that hPT preferentially represents directional motion (observed in our study by higher univariate responses and higher within-condition classification accuracies), the cross-condition classification results revealed that auditory motion (e.g. going to the left) and sound-source location (being on the left) share partial neural representations in hPT (Figure 3A). The idea of cross-condition classification between motion direction and sound-source location necessarily relies on whether there is a shared computation between sounds located on a given position and sounds moving towards this position. Low-level features of these two types of auditory stimuli vary in many ways and produce large difference at the univariate level in the cortex (see Figure 2B). However, perceiving,

for instance, a sound going toward the left side or located on the left side evoke a sensation of location/direction in the external space that is common across conditions. Our significant cross-condition classification may therefore relate to the evoked sensation/perception of an object being/going to a common external spatial location. Electrophysiological studies in animals demonstrated that motion-selective neurons in the auditory cortex displayed similar response profile to sounds located or moving toward the same position in external space, suggesting that the processing of sound-source locations may contribute to the perception of moving sounds (Ahissar et al., 1992; Doan et al., 1999; Poirier et al., 1997). Results from human psychophysiological and auditory evoked potential studies also strengthen the notion that sound source location contributes to motion perception (Getzmann and Lewald, 2011; Strybel and Neale, 1994). Our cross-condition MVPA results therefore extend the notion that motion directions and sound source locations might have common features that are partially shared for encoding spatial sounds.

Significant cross-condition classification based on multivariate fMRI signal has typically been considered as a demonstration that the region implements a partly common and abstracted representation of the tested conditions (Fairhall and Caramazza, 2013; Higgins et al., 2017; Hong et al., 2012). For instance, a recent study elegantly demonstrated that the human auditory cortex at least partly integrates interaural time and level differences (ITD and ILD) into a higher-order representation of auditory space based on significant cross-cue classification (training on ITD and classifying ILD, and reversely). In the present study, we argue that even if cross-condition MVP-classification indeed demonstrate the presence of partially shared information across sound source location and direction; successful cross-MVPA results should however not be taken as evidence that the region implements a purely abstract representation of auditory space. Our successful across-condition classification (see Figure 4A) demonstrated that, even though there are shared representations for moving and static sounds within hPT, classifiers are able to easily distinguish motion directions from sound source locations (e.g. leftward versus left location). RSA analyses further supported the idea that moving and static sounds elicit distinct patterns in hPT (see Figure 4B-D). Altogether, our

results suggest that hPT contains both motion direction and sound-source location information but that the neural patterns related to these two conditions are only partially overlapping. Our observation of significant cross-condition classification based on highly distinct pattern of activity between static and moving sounds may support the notion that even if location information could serve as a substrate for movement detection, motion encoding does not solely rely on location information (Ducommun et al., 2002; Getzmann, 2011; Poirier et al., 2017).

REFERENCES

Ahissar, M., Ahissar, E., Bergman, H., and Vaadia, E. (1992). Encoding of sound-source location and movement: activity of single neurons and interactions between adjacent neurons in the monkey auditory cortex. J. Neurophysiol. *67*, 203–215.

Ahveninen, J., Jaaskelainen, I.P., Raij, T., Bonmassar, G., Devore, S., Hamalainen, M., et al. (2006). Task-modulated "what" and "where" pathways in human auditory cortex. Proc. Natl. Acad. Sci. U. S. A. *103*, 14608–14613.

Ahveninen, J., Huang, S., Nummenmaa, A., Belliveau, J.W., Hung, A.-Y., Jääskeläinen, et al. (2013). Evidence for distinct human auditory cortex regions for sound location versus identity processing. Nat. Commun. *4*, 871–882.

Albright, T.D., Desimone, R., and Gross, C.G. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. J. Neurophysiol. *51*, 16–31.

Alink, A., Euler, F., Kriegeskorte, N., Singer, W., and Kohler, A. (2012). Auditory motion direction encoding in auditory cortex and high-level visual cortex. Hum. Brain Mapp. *33*, 969–978.

Altman, J.A. (1968). Are there neurons detecting direction of sound source motion? Exp. Neurol. *22*, 13–25.

Altman, J.A. (1994). Processing of Information Concerning Moving Sound Sources in the Auditory Centers and its Utilization by Brain Integrative Structures. Sens. Syst. *8*, 255–261.

Arnott, S. R., Binns, M. A., Grady, C. L., and Alain, C. (2004). Assessing the

auditory dual-pathway model in humans. Neuroimage *22*, 401–408.

Barrett, D.J.K., and Hall, D.A. (2006). Response preferences for "what" and "where" in human non-primary auditory cortex. Neuroimage *32*, 968–977.

Bartels, A., Logothetis, N.K., and Moutoussis, K. (2008). fMRI and its interpretations: an illustration on directional selectivity in area V5 / MT. Trends in Neuro. *31*, 444-453.

Baumgart, F., and Gaschler-Markefski, B. (1999). A movement-sensitive area in auditory cortex. Nature *400*, 1997–1999.

Beckett, A., Peirce, J.W., Sanchez-Panchuelo, R.M., Francis, S., and Schluppeck, D. (2012). Contribution of large scale biases in decoding of direction-of-motion from high-resolution fMRI data in human early visual cortex. Neuroimage *63*, 1623–1632.

Op de Beeck, H.P. (2010). Against hyperacuity in brain reading: Spatial smoothing does not hurt multivariate fMRI analyses? Neuroimage *49*, 1943–1948.

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Ann. Stat. *29*, 1165–1188.

Benson, D.A., Hienz, R.D., and Goldstein, M.H. (1981). Single-unit activity in the auditory cortex of monkeys actively localizing sound sources: Spatial tuning and behavioral dependency. Brain Res. *219*, 249–267.

Blauert, J. (1982). Binaural localization. Scand.Audiol.Suppl *15*, 7–26.

Born, R.T., and Bradley, D.C. (2005). Structure and Function of Visual Area MT. Ann. Review of Neuro. *28*, 157–189.

Boudreau, J.C. & Tsuchitani, C., (1968). Binaural interaction in the cat superior olive S segment. *Journal of neurophysiology*, 31(3), pp.442–454

Braddick, O.J., O'Brien, J.M.D., Wattam-Bell, J., Atkinson, J., Hartley, T., and Turner, R. (2001). Brain areas sensitive to coherent visual motion. Perception *30*, 61–72.

Bremmer, F., Schlack, A., Shah, N.J., Zafiris, O., Kubischik, M., Hoffmann, K., Zilles, K., and Fink, G.R. (2001). Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. Neuron *29*, 287–296.

Britten, K.H., Newsome, W.T., Shadlen, M.N., Celebrini, S., and Movshon, J.A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. Vis. Neurosci. *13*, 87–100.

Brunetti, M., Belardinelli, P., Caulo, M., Del Gratta, C., Della Penna, S., Ferretti, et al. (2005). Human brain activation during passive listening to sounds from different locations: an fMRI and MEG study. Hum. Brain Mapp. *26*, 251–261.

Carlile, S., and Leung, J. (2016). The Perception of Auditory Motion. Trends Hear. *20*, 1–19.

Collignon, O., Vandewalle, G., Voss, P., Albouy, G., Charbonneau, G., Lassonde, M., et al. (2011). Functional specialization for auditory-spatial processing in the occipital cortex of congenitally blind humans. Proc. Natl. Acad. Sci. U. S. A. *108*, 4435–4440.

Cox D.D., and Savoy R.L. (2003) Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage *19*, 261–270.

Deouell, L.Y., Heller, A.S., Malach, R., Esposito, M.D., and Knight, R.T. (2007). Cerebral Responses to Change in Spatial Location of Unattended Sounds. Neuron *55*, 985–996.

Derey, K., Valente, G., De Gelder, B., and Formisano, E. (2016). Opponent Coding of Sound Location (Azimuth) in Planum Temporale is Robust to Sound-Level Variations. Cereb. Cortex *26*, 450–464.

Diogo, A.C.M., Soares, J.G.M., Koulakov, A., Albright, T.D., and Gattass, R. (2003). Electrophysiological imaging of functional architecture in the cortical middle temporal visual area of Cebus apella monkey. J. Neurosci. *23*, 3881–3898.

Doan, D.E., Saunders, J.C., Field, F., Ingham, N.J., Hart, H.C., and McAlpine, D. (1999). Sensitivity to Simulated Directional Sound Motion in the Rat Primary Auditory Cortex. J. Neurophysi. 81, 2075–2087.

Dormal, G., Rezk, M., Yakobov, E., Lepore, F., and Collignon, O. (2016). Auditory motion in the sighted and blind: Early visual deprivation triggers a large-scale imbalance between auditory and "visual" brain regions. Neuroimage *134*, 630–644.

Ducommun, C.Y., Murray, M.M., Thut, G., Bellmann, A., Viaud-Delmon, I., Clarke, S., et al. (2002). Segregated processing of auditory motion and auditory location: an ERP mapping study. Neuroimage *16*, 76–88.

Edelman, S., Grill-Spector, K., Kushnir, T., and Malach, R. (1998). Toward direct visualization of the internal shape represetation space by fMRI. Psychobiology *26*, 309–321.

Eickhoff, S.B., Paus, T., Caspers, S., Grosbras, M.H., Evans, A.C., Zilles, K., et al. (2007). Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. Neuroimage *36*, 511–521.

Fairhall, S.L., and Caramazza, A. (2013). Brain Regions That Represent Amodal Conceptual Knowledge. J. Neurosci. *33*, 10552–10558.

Freeman, J., Brouwer, G.J., Heeger, D.J., and Merriam, E.P. (2011). Orientation decoding depends on maps, not columns. J. Neurosci. *31*, 4792–4804.

Freeman, J., Heeger, D.J., and Merriam, E.P. (2013). Coarse-Scale Biases for Spirals and Orientation in Human Visual Cortex. J. Neurosci. *33*, 19695–19703.

Gardumi, A., Ivanov, D., Hausfeld, L., Valente, G., Formisano, E., and Uludağ, K. (2016). The effect of spatial resolution on decoding accuracy in fMRI multivariate pattern analysis. Neuroimage *132*, 32–42.

Geesaman, B.J., Born, R.T., Andersen, R.A., and Tootell, R.B. (1997). Maps of complex motion selectivity in the superior temporal cortex of the alert macaque monkey: a double-label 2-deoxyglucose study. Cereb Cortex *7*, 749–757.

Getzmann, S. (2011). Auditory motion perception: onset position and motion direction are encoded in discrete processing stages. Eur. J. Neurosci. *33*, 1339–1350.

Getzmann, S., and Lewald, J. (2011). The effect of spatial adaptation on auditory motion processing. Hear. Res. *272*, 21–29.

Getzmann, S., and Lewald, J. (2012). Cortical processing of change in sound location: Smooth motion versus discontinuous displacement. Brain Res. *1466*, 119–127.

Grantham, D.W. (1986). Detection and discrimination of simulated motion of auditory targets in the horizontal plane. J. Acoust. Soc. Am. *79*, 1939–1949.

Griffiths, T.D., Büchel, C., Frackowiak, R.S., and Patterson, R.D. (1998). Analysis of temporal structure in sound by the human brain. Nat. Neurosci. *1*, 422–427.

Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, et al. (1999). "Sparse" temporal sampling in auditory fMRI. Hum. Brain Mapp. *7*, 213–223.

Hall, D.A., and Moore, D.R. (2003). Auditory Neuroscience : The Salience of Looming Sounds. Curr. Biology *13*, 91–93.

Haxby, J. V, Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science *293*, 2425–2430.

Haynes, J.D., and Rees, G. (2006). Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. *7*, 523–534.

Heeger, D.J., Boynton, G.M., Demb, J.B., Seidemann, E., and Newsome, W.T. (1999). Motion opponency in visual cortex. J. Neurosci. *19*, 7162–7174.

van der Heijden, K., Rauschecker, J.P., Formisano, E., Valente, G., and de Gelder, B. (2018). Active sound localization sharpens spatial tuning in human primary auditory cortex. J.Neurosci. 0587-18.

Higgins, N.C., McLaughlin, S.A., Rinne, T., and Stecker, G.C. (2017). Evidence for cue-independent spatial representation in the human auditory cortex during active listening. Proc. Natl. Acad. Sci. *114*, E7602–E7611.

Imig, T.J., Irons, W. a, and Samson, F.R. (1990). Single-unit selectivity to azimuthal direction and sound pressure level of noise bursts in cat high-frequency primary auditory cortex. J. Neurophysiol. *63*, 1448–1466.

Imig, T.J., Bibikov, N.G., Poirier, P., and Samson, F.K. (2000). Directionality derived from pinna-cue spectral notches in cat dorsal cochlear nucleus. J. Neurophysiol. *83*, 907–925.

Ingham, N.J., Hart, H.C., and McAlpine, D. (2001). Spatial receptive fields of inferior colliculus neurons to auditory apparent motion in free field. J. Neurophysiol. *85*, 23–33.

Jiang, F., Stecker, G.C., and Fine, I. (2014). Auditory motion processing after

early blindness. J. Vis. *14*, 4–4.

Jiang, F., Stecker, G.C., Boynton, G.M., and Fine, I. (2016). Early Blindness Results in Developmental Plasticity for Auditory Motion Processing within Auditory and Occipital Cortex. Front. Hum. Neurosci. *10*, 324.

Kaas, J.H., Hackett, T.A., and Tramo, M.J. (1999). Auditory processing in primate cerebral cortex. Curr. Opin. Neurobiol. *9*, 164–170.

Kamitani, Y., and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *8*, 679–685.

Kamitani, Y., and Tong, F. (2006). Decoding Seen and Attended Motion Directions from Activity in the Human Visual Cortex. Curr. Biol. *16*, 1096–1102.

Kamitani, Y., and Sawahata, Y. (2010). Spatial smoothing hurts localization but not information: Pitfalls for brain mappers. Neuroimage *49*, 1949–1952.

Kriegeskorte N, Mur M, Bandettini P. (2008). Representational similarity analysis—connecting the branches of systems neuroscience. Front. Syst. Neurosci. *2*, 1-28.

Kriegeskorte, N., Formisano, E., Sorger, B., and Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. Proc. Natl. Acad. Sci. *104*, 20600–20605.

Krumbholz, K., Schönwiesner, M., Rübsamen, R., Zilles, K., Fink, G.R., and von Cramon, D.Y. (2005). Hierarchical processing of sound location and motion in the human brainstem and planum temporale. Eur. J. Neurosci. *21*, 230–238.

Lewis, J.W., Beauchamp, M.S., and DeYoe, E.A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. Cereb.Cortex *10*, 873–888.

Lomber, S.G., and Malhotra, S. (2008). Double dissociation of "what" and "where" processing in auditory cortex. Nat. Neurosci. *11*, 609–616.

Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., et al. (2001). Distinct pathways involved in sound recognition and localization: a human fMRI study. Neuroimage *14*, 802–816.

Middlebrooks, J., and Pettigrew, J. (1981). Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location. J

Neurosci *1*, 107–120.

Middlebrooks, J.C., and Green, D.M. (1991). Sound localization by human listeners. Annu. Rev. Psychol. *42*, 135–159.

Middlebrooks, J.C. (2002). Auditory space processing: Here, there or everywhere? Nat. Neurosci. *5*, 824–826.

Middlebrooks, J.C., and Bremen, P. (2013). Spatial Stream Segregation by Auditory Cortical Neurons. J. Neurosci. *33*, 10986–11001.

Movshon, J.A., and Newsome, W.T. (1996). Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. J. Neurosci. *16*, 7733–7741.

Newsome, W.T., and Park, E.B. (1988). A Selective Impairment of Motion Perception Following Lesions of the Middle Temporal Visual Area (MT). J. Neurosci. *8*, 2201–2211.

Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci *10*, 424–430.

Oosterhof, N.N., Connolly, A.C., and Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. Front. Neuroinform. *10*, 1–27.

Ortiz-Rios, M., Azevedo, F.A.C., Kuśmierek, P., Balla, D.Z., Munk, M.H., Keliris, G.A., et al. (2017). Widespread and Opponent fMRI Signals Represent Sound Location in Macaque Auditory Cortex. Neuron *93*, 971–983.e4.

Pavani, F., Macaluso, E., Warren, J.D., Driver, J., and Griffiths, T.D. (2002). A common cortical substrate activated by horizontal and vertical sound movement in the human brain. Curr. Biol. *12*, 1584–1590.

Poirier, C., Collignon, O., Devolder, A.G., Renier, L., Vanlierde, A., Tranduy, D., et al. (2005). Specific activation of the V5 brain area by auditory motion processing: an fMRI study. Brain Res.Cogn Brain Res. *25*, 650–658.

Poirier, C., Baumann, S., Dheerendra, P., Joly, O., Hunter, D., Balezeau, et al. (2017). Auditory motion-specific mechanisms in the primate brain. PLoS Biol. *15*, 1–24.

Poirier, P., Jiang, H., Lepore, F., and Guillemot, J.P. (1997). Positional,

directional and speed selectivities in the primary auditory cortex of the cat. Hear. Res. *113*, 1–13.

Pratte, M.S., Sy, J.L., Swisher, J.D., and Tong, F. (2016). Radial bias is not necessary for orientation decoding. Neuroimage *127*, 23–33.

Rajan, R., Aitkin, L.M., and Irvine, D.R. (1990). Azimuthal sensitivity of neurons in primary auditory cortex of cats. II. Organization along frequency-band strips. J. Neurophysiol. *64*, 888–902.

Rauschecker, J.P., and Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc.Natl.Acad.Sci.U.S.A. *97*, 11800–11806.

Recanzone, G.H. (2000). Spatial processing in the auditory cortex of the macaque monkey. Proc. Natl. Acad. Sci. U. S. A. *97*, 11829–11835.

Romanski, Tian, Fritz, Mishkin, Goldman-Rakic, and Rauschecker (2000). Reply to "What', "where" and "how" in auditory cortex'. Nat. Neurosci. *3*, 966.

Romanski, L.M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P.S., and Rauschecker, J.P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat.Neurosci. *2*, 1131–1136.

Searle, C.L., Braida, L.D., Davis, M.F., and Colburn, H.S. (1976). Model for auditory localization. J. Acoust. Soc. Am. *60*, 1164–1175.

Sengupta, A., Yakupov, R., Speck, O., Pollmann, S., and Hanke, M. (2017). The effect of acquisition resolution on orientation decoding from V1 BOLD fMRI at 7 T. Neuroimage *148*, 64–76.

Smith, K.R., Okada, K., Saberi, K., and Hickok, G. (2004). Human cortical auditory motion areas are not motion selective. Neuroreport *15*, 1523–1526.

Smith, K.R., Saberi, K., and Hickok, G. (2007). An event-related fMRI study of auditory motion perception: No evidence for a specialized cortical system. Brain Res. *1150*, 94–99.

Smith, K.R., Hsieh, I.H., Saberi, K., and Hickok, G. (2010). Auditory spatial and object processing in the human planum temporale: no evidence for selectivity. J. Cogn. Neurosci. *22*, 632–639.

Stecker, G.C., Harrington, I.A., and Middlebrooks, J.C. (2005). Location

coding by opponent neural populations in the auditory cortex. PLoS Biol. *3*, 0520–0528.

Stelzer, J., Chen, Y., and 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.

Strybel, T.Z., and Neale, W. (1994). The effect of burst duration, interstimulus onset interval, and loudspeaker arrangement on auditory apparent motion in the free field. J. Acoust. Soc. Am. *96*, 3463–3475.

Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J.P. (2001). Functional Specialization in Rhesus Monkey Auditory Cortex. Science *292*, 290–293.

Tootell, R.B.H., Reppas, J.B., Dale, A.M., Look, R.B., Sereno, M.I., Malach, R., et al. (1995). Visual motion aftereffect in human cortical area MT revealed by functional magnetic resonance imaging. Nature *375*, 139–141.

Wang, H.X., Merriam, E.P., Freeman, J., and Heeger, D.J. (2014). Motion Direction Biases and Decoding in Human Visual Cortex. J. Neurosci. *34*, 12601–12615.

Warren, J., Zielinski, B., and Green, G. (2002). Perception of sound-source motion by the human brain. Neuron *34*, 139–148.

Warren, J.D., and Griffiths, T.D. (2003). Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. J.Neurosci. *23*, 5799–5804.

Woods, T.M., Lopez, S.E., Long, J.H., Rahman, J.E., and Recanzone, G.H. (2006). Effects of Stimulus Azimuth and Intensity on the Single-Neuron Activity in the Auditory Cortex of the Alert Macaque Monkey. J. Neurophysiol. *96*, 3323–3337.

Hong, S., Tong, F., and Seiffert, A.E. (2012). Direction-selective patterns of activity in human visual cortex suggest common neural substrates for different types of motion. Neuropsychologia *50*, 514–521.

Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., and Wager, T.D. (2011). Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods *8*, 665–670.

Young, E.D., Spirou, G.A., Rice, J.J., Voigt, H.F., and Rees, A. (1992). Neural

Organization and Responses to Complex Stimuli in the Dorsal Cochlear Nucleus. Philos. Trans. R. Soc. B Biol. Sci. *336*, 407–413.

Zabicki, A., De Haas, B., Zentgraf, K., Stark, R., Munzert, J., and Krüger, B. (2016). Imagined and Executed Actions in the Human Motor System: Testing Neural Similarity Between Execution and Imagery of Actions with a Multivariate Approach. Cereb. Cortex *27*, 1–14.

Zatorre, R.J., and Belin, P. (2001). Spectral and Temporal Processing in Human Auditory Cortex. Cereb. Cortex *11*, 946–953.

Zatorre, R.J., Bouffard, M., Ahad, P., and Belin, P. (2002). Where is "where" in the human auditory cortex? Nat. Neurosci. *5*, 905–909.

Zeng, H., and Constable, R.T. (2002). Image distortion correction in EPI: Comparison of field mapping with point spread function mapping. Magn. Reson. Med. *48*, 137–146.

Zhou, Y., and Wang, X. (2012). Level dependence of spatial processing in the primate auditory cortex. J. Neurophysiol. *108*, 810–826.

Zimmermann, J., Goebel, R., de Martino, F., van de Moortele, P.F., Feinberg, D., Adriany, G., et al. (2011). Mapping the Organization of Axis of Motion Selective Features in Human Area MT Using High-Field fMRI. PLoS One *6*, 1–10.

Supplemental Information

To confirm our results and the location of functionally and independently defined hPT, we performed similar univariate and multi-variate analyses on anatomically defined hPT. By using FreeSurfer (http://surfer.nmr.mgh.harvard.edu) anatomical parcellation, hPT region of interest were individually defined in each subject. The individual anatomical scan was used to perform cortical anatomical segmentation according to the Destrieux (Destrieux et al., 2010) atlas. We selected planum temporale label defined as [G_temp_sup-Plan_tempo, 36] bilaterally. We equated the size of ROI across participants and across hemispheres to 110 voxels (each voxel being 3mm isotropic).

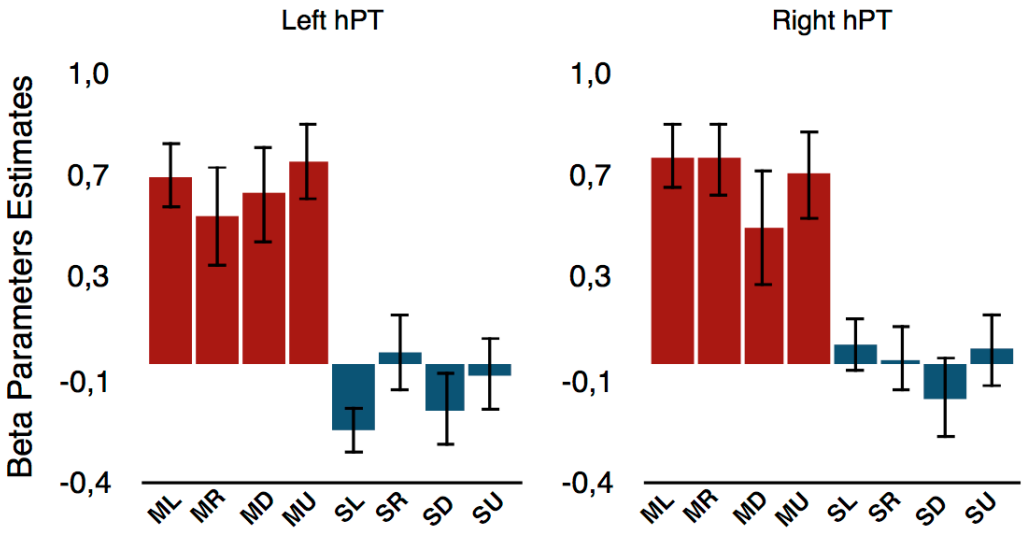

Figure S1. Univariate results of anatomically defined hPT. (A). Mean activity estimates (arbitrary units ± SEM) associated with the perception of auditory motion direction (red) and sound-source location (blue). ML: motion left, MR: motion right, MD: motion down, MU: motion up, SL: static left, SR: static right, SD: static down, and SU: static up.

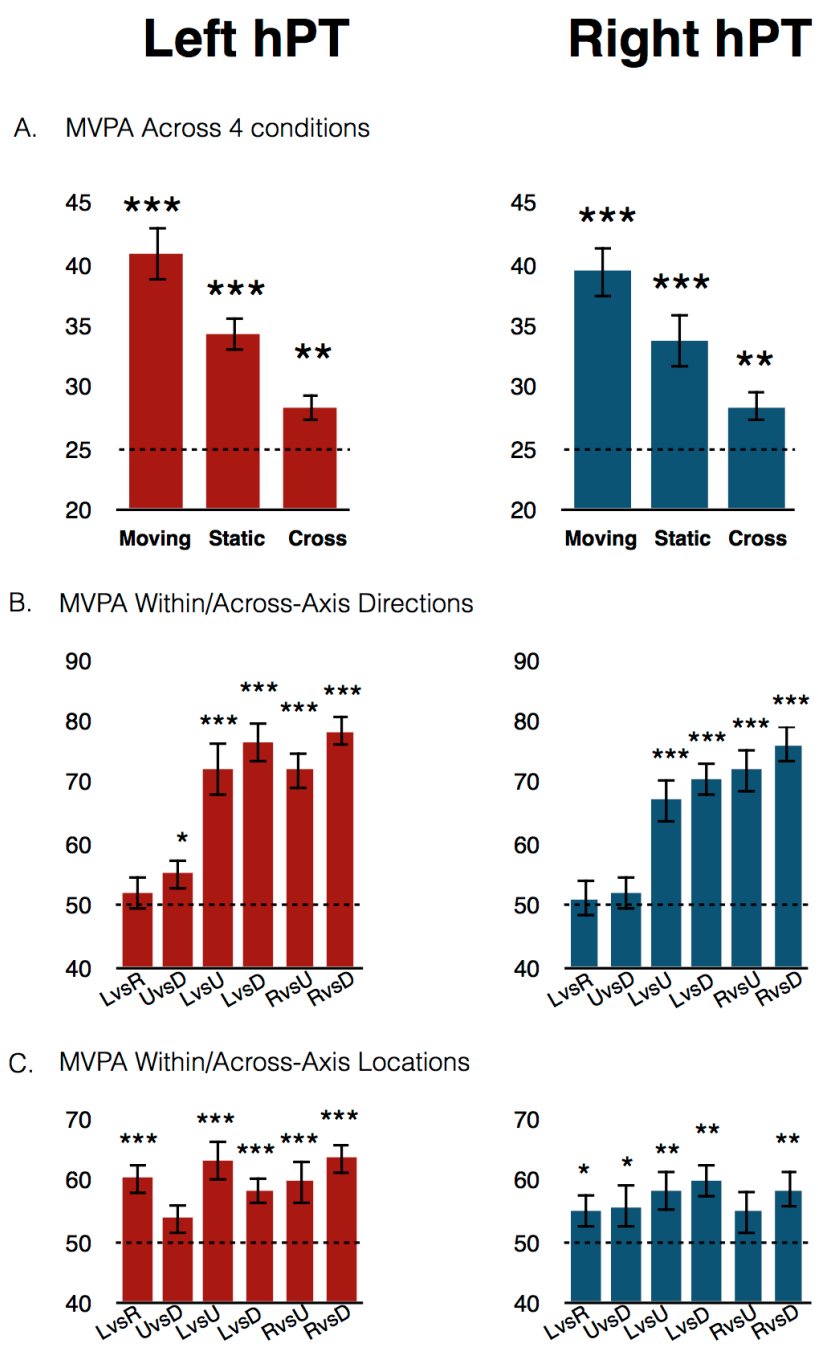

Figure S2. Within and cross-classification results of anatomically defined hPT. (A). Classification results of across 4 conditions. Within-condition and cross-condition classification results are shown in the same bar plots. **(B).** Classification results of within (left vs. right, up vs. down) and across axes (left vs. up, left vs. down, right vs. up, right vs. down) motion directions. **(C).** Classification results of within (left vs. right, up vs. down) and across axes (left vs. up, left vs. down, right vs. up, right vs. down) sound source locations. LvsR: Left vs. Right, UvsD: Up vs. Down, LvsU: Left vs. Up, LvsD: Left vs. Down, RvsU: Right vs. Up, RvsD: Right vs. Down classifications. FDR corrected p-values: (*) is $p<0.05$, (**) $p<0.01$, (***) $p<0.001$ testing differences against chance level (dotted lines; see methods).

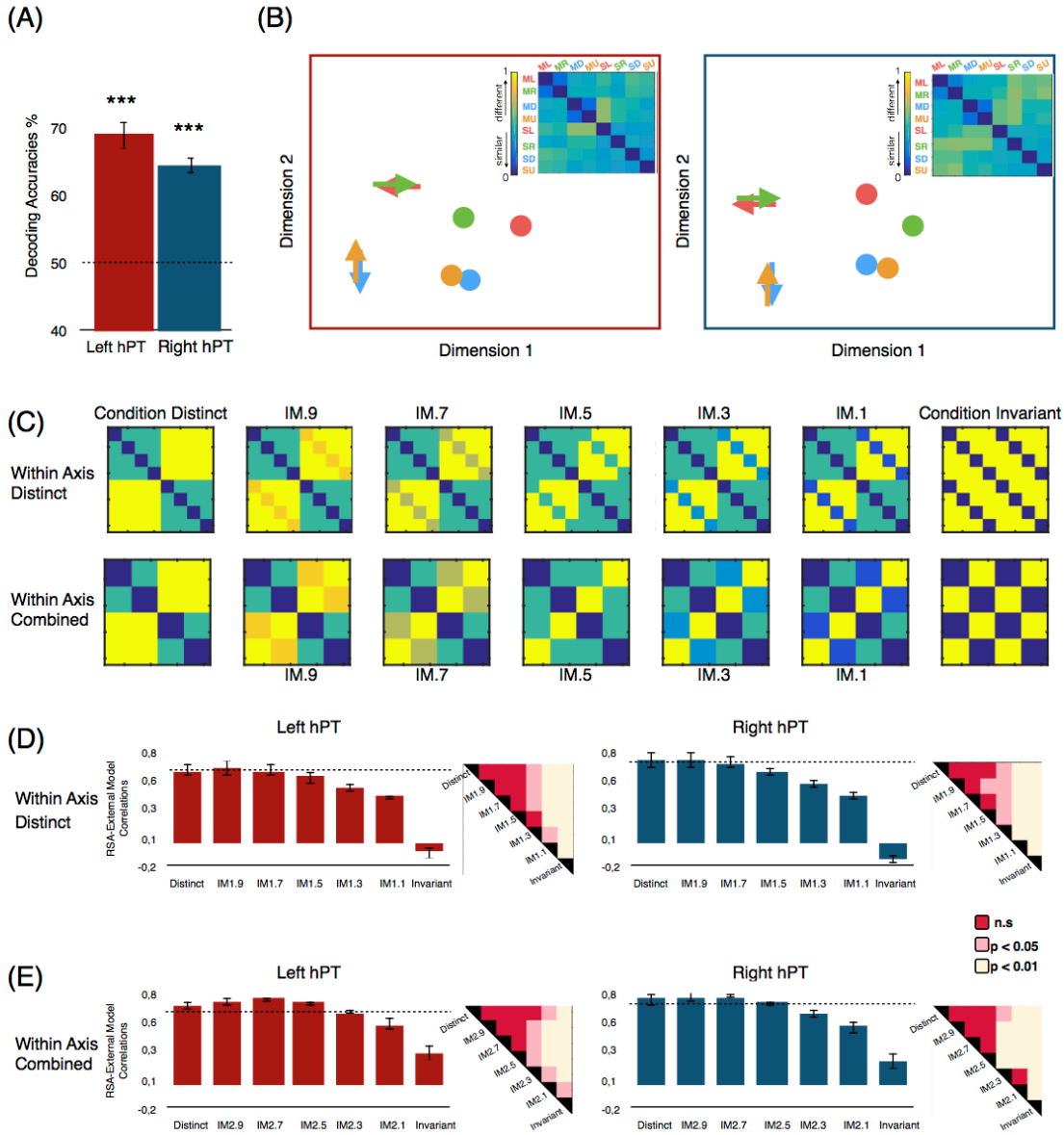

Figure S3. Pattern dissimilarity between motion directions and sound source locations in anatomically defined hPT. (A). Across-condition classification results of across 4 conditions are represented in each ROI (lhPT and rhPT). 4 binary classifications [leftward motion vs. left location], [rightward motion vs. right location], [upward motion vs. up location], and [downward motion vs. down location] were computed and averaged to produce one accuracy score per ROI. FDR corrected p-values: (***) p<0.001. Dotted lines represent chance level. **(B).** The embedded top panel shows neural RDMs extracted from left and right hPT, and multi-dimensional scaling (MDS) plot visualizes the similarities of the neural pattern elicited by 4 motion directions (arrows) and 4 sound source locations (dots). Color codes for arrow/dots are as follows: green indicates left direction/location, red indicates right direction/location, orange indicates up direction/location, and blue indicates down direction/location. ML: motion left, MR: motion right, MD: motion down, MU: motion up, SL: static left, SR: static right, SD: static down, and

SU: static up. **(C-D).** The results of representational similarity analysis (RSA) in hPT are represented. **(C).** RDMs of the computational models that assume different similarities of the neural pattern based on auditory motion and static conditions. **(D).** RSA results for every model and each ROI. For each ROI, the dotted lines represent the reliability of the data considering the signal-to-noise ratio (see Materials and Methods), which provides an estimate of the highest correlation we can expect in a given ROI when correlating computational models and neural RDMs. Error bars indicate SEM. IM1: Intermediate models with within-axis conditions distinct, IM2: Intermediate model with within-axis conditions combined. Each right up corner of the bar plots show visualization of significant differences for each class of models and hemispheres separately (Mann-Whitney-Wilcoxon rank-sum test, FDR corrected).

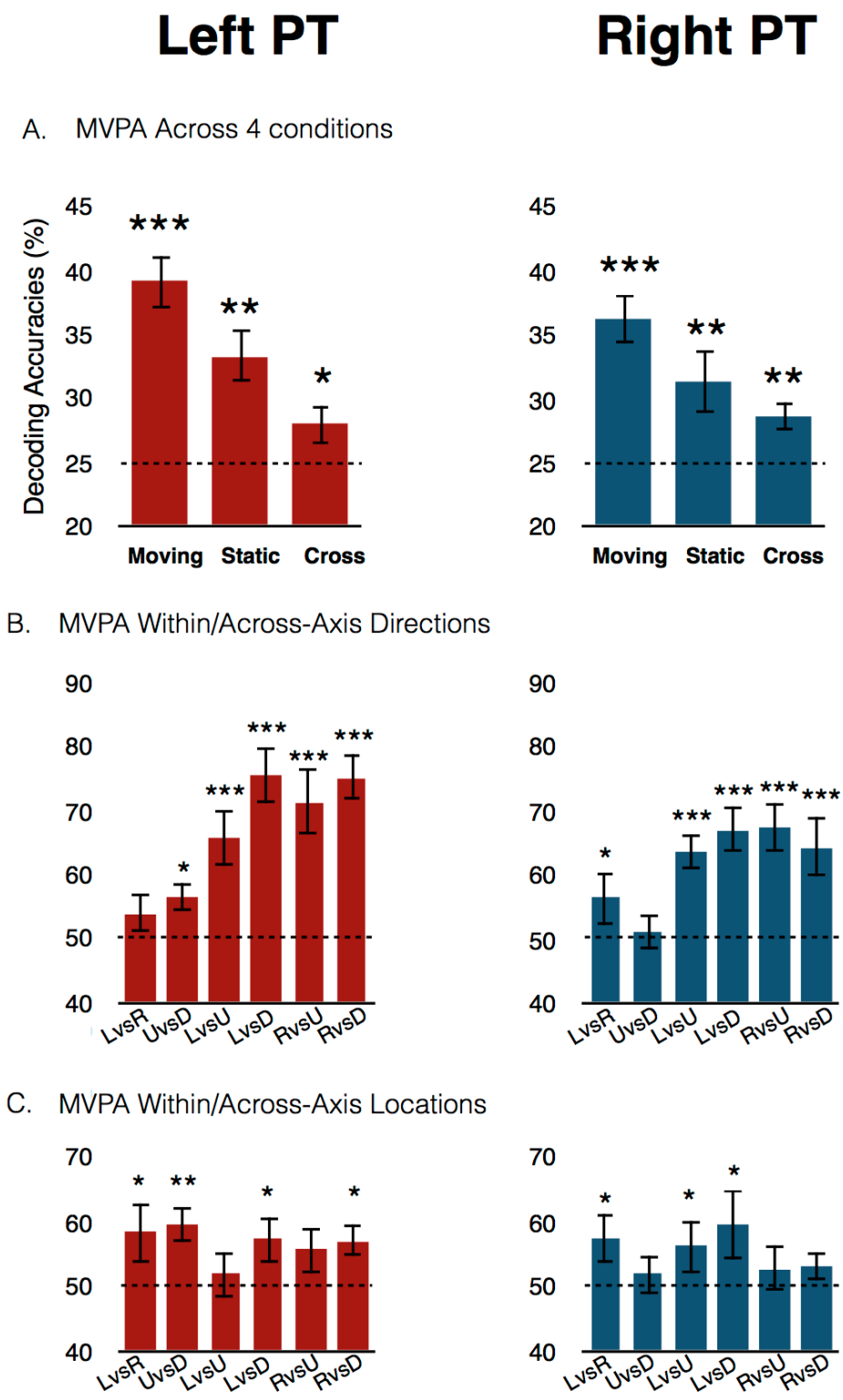

Figure S4. Within and cross-classification results after omitting the 4 worst subjects in the vertical discrimination task (N=12). The four subjects were omitted due to their lower performance in detection of vertical motion (behavioral accuracy between 65 - 77 %). We re-ran the MVPA with 12 subjects (behavioral accuracy > 87%). **(A).** Classification results of across 4 conditions. Within-condition and cross-condition classification results are shown in

the same bar plots. **(B).** Classification results of within (left vs. right, up vs. down) and across axes (left vs. up, left vs. down, right vs. up, right vs. down) motion directions. **(C).** Classification results of within (left vs. right, up vs. down) and across axes (left vs. up, left vs. down, right vs. up, right vs. down) sound source locations. LvsR: Left vs. Right, UvsD: Up vs. Down, LvsU: Left vs. Up, LvsD: Left vs. Down, RvsU: Right vs. Up, RvsD: Right vs. Down classifications. FDR corrected p-values: (*) is p<0.05, (**) p<0.01, (***) p<0.001 testing differences against chance level (dotted lines; see methods).

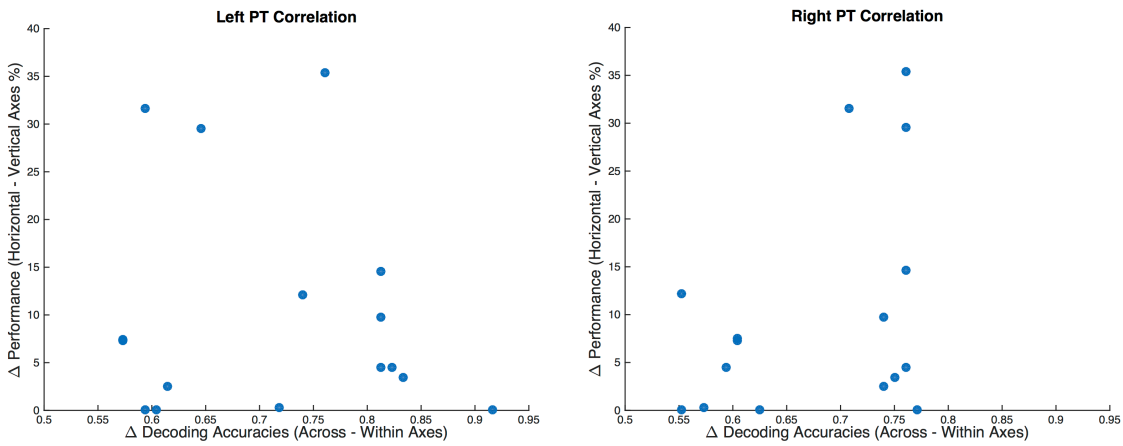

Figure S5. Results of correlation between decoding accuracies and behavioral performance in the orthogonal axes.

Supplemental Analysis

The beta parameter estimates of the 4 motion directions and 4 sound source locations were extracted from lhPT and rhPT anatomically defined parcels (Fig S1). In order to investigate the presence of motion directions/sound source locations selectivity and condition effect in hPT regions, we performed a 2 Conditions (motion, static) x 4 Orientations (left, right, down, and up) repeated measures ANOVA in each hemisphere separately on these beta parameter estimates. A Greenhouse–Geisser correction was applied to the degrees of freedom and significance levels whenever an assumption of sphericity was violated. In lhPT, main effect of Conditions was significant ($F_{1,15}$ = 102.42, $p < 0.001$), indicating that auditory motion evoked higher response compared to static sounds. There was no significant main effect of Orientations ($F_{1.5,22.5}$ = 0.49, p = 0.5), and no interaction

($F_{3,45}$ = 1.88, p = 0.14). Similarly, in rhPT, only main effect of Conditions was significant ($F_{1,15}$ = 54.63, $p < 0.001$). No main effect of Orientation ($F_{1.5,23.2}$ = 1.78, p = 0.2) or interaction ($F_{3,45}$ = 1.5, p = 0.9) was observed. Overall, brain activity in the hPT as measured with beta parameter estimate extracted from univariate analysis did not provide evidence of motion direction or sound source location selectivity.