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@uclouvain.be; 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 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 male and female 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.

SIGNIFICANCE STATEMENT

In comparison to what we know about visual motion, little is known about how the brain implements spatial hearing. Our study reveals that motion directions and sound source locations can be reliably decoded in the human Planum Temporale

(hPT) and that they rely on partially shared pattern geometries. Our study therefore sheds important new lights on how computing the location or direction of sounds are implemented in the human auditory cortex by showing that those two computations rely on partially shared neural codes. Furthermore, our results show that the neural representation of moving sounds in hPT follows a "preferred axis of motion" organization, reminiscent of the coding mechanisms typically observed in the occipital hMT+/V5 region for computing visual motion.

INTRODUCTION

While the brain mechanisms underlying the processing of visual localization and visual motion have received considerable attention (Braddick et al., 2001; 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 brainstem, thalamus, and cortex in order to create an integrated neural representation of auditory space (Boudreau and Tsuchitani, 1968; Goldberg and Brown, 1969; Knudsen & Konishi 1978; Ingham et al., 2001; see Grothe et al., 2010 for review).

Similar to the dual-stream processing model in vision (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; Alain et al., 2001; 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 substrates (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 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). In contrast, other studies in animals (Poirier et al., 2017) and humans (Baumgart and Gaschler-Markefski, 1999; Bremmer et al., 2001; 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 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, as it was lower by more than 2.5 standard deviations than the average of the participants. 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.

127 **Auditory stimuli**

128 Our limited knowledge of the auditory space processing in the human brain might be
129 a consequence of the technical challenge of evoking vivid perceptual experience of
130 auditory space while using neuroimaging tools such as fMRI, EEG, or MEG. In this
131 experiment, to create an externalized ecological sensation of sound location and
132 motion inside the MRI scanner, we relied on individual in-ear stereo recordings that
133 were recorded in a semi-anechoic room and from 30 loudspeakers on horizontal and
134 vertical planes, mounted on two semi-circular wooden structures with a radius of 1.1m
135 (see Figure 1A). Participants were seated in the center of the apparatus with their head
136 on a chin-rest, such that the speakers on the horizontal and vertical planes were
137 equally distant from participants' ears. Then, these recordings were re-played to the
138 participants when they were inside the MRI scanner. By using such sound system with
139 in-ear recordings, auditory stimuli automatically convolved with each individuals' own
140 pinna and head related transfer function to produce a salient auditory perception in
141 external space.

--- Figure 1 about here ---

142 The auditory stimuli were prepared using custom MATLAB scripts (r2013b; Matworks).
143 Auditory stimuli were recorded using binaural in-ear omni-directional microphones
144 (Sound Professionals-TFB-2; 'flat' frequency range 20–20,000 Hz) connected to a
145 portable Zoom H4n digital wave recorder (16-bit, stereo, 44.1 kHz sampling rate).
146 Microphones were positioned at the opening of participant's left and right auditory
147 ear canals. While auditory stimuli were played, participants were listening without

performing any task with their 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 participant's report that our stimuli were perceived as smoothly moving (no perception of successive snapshots).

171 In the static condition, the same pink noise was presented separately at one of 4
172 locations: left, right, up, and down. Static sounds were presented at the second most
173 outer speakers (-56° and +56° in the horizontal axis, and +56° and -56° in the vertical
174 axis) in order to avoid possible reverberation difference at the outermost speakers.
175 The static sounds were fixed at one location at a time instead of presented in multiple
176 locations (Smith et al., 2008; 2010; Poirier et al., 2017; Krumbholz et al., 2005). This
177 strategy was purposely adopted for two main reasons. First, randomly presented static
178 sounds can evoke robust sensation of auditory apparent motion (Strybel & Neale
179 1994; Lakatos et al., 1997; see Carlile 2016 for review). Second, and crucially for the
180 purpose of the present experiment, presenting static sounds located in a given
181 position and moving sounds directed toward the same position allowed us to
182 investigate whether moving and static sounds share a common representational
183 format using cross-condition decoding (see below), which would have been
184 impossible if the static sounds where randomly moving in space.

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

191

192 **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 auditory categories (4 motion and 4 static) randomly presented using a block-design. Each category of sound 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 (50 ms fade in/out) 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 category. This procedure was adopted to ensure that the participants gave their response using equal motor command for each category and to ensure the response is produced after the end of the stimulation period for each category. Participants were instructed to emphasize accuracy of response but not reaction times.

214 Each scan consisted of one block of each category, resulting in a total of 8
215 blocks per run, with each run lasting 4 m 10 s. Participants completed a total of 12
216 runs. The order of the blocks was randomized within each run, and across participants.
217 Based on pilot experiments, we decided to not rely on a sparse-sampling
218 design as sometimes done in the auditory literature in order to present the sounds
219 without the scanner background noise (Hall et al., 1999). These pilot experiments
220 showed that the increase in the signal to noise ratio potentially provided by sparse
221 sampling did not compensate for the loss in the number of volume acquisitions.
222 Indeed, pilot recordings on participants not included in the current sample showed
223 that, given a similar acquisition time between sparse-sampling designs (several
224 options tested) and continuous acquisition, the activity maps elicited by our spatial
225 sounds contained higher and more reliable beta values using continuous acquisition.
226

227 **fMRI data acquisition and analyses**

228 *Imaging parameters*

229 Functional and structural data were acquired with a 4T Bruker MedSpec
230 Biospin MR scanner, equipped with an 8-channel head coil. Functional images were
231 acquired with T2*-weighted gradient echo-planar sequence. Acquisition parameters
232 were: repetition time of 2500 ms, echo time of 26 ms, flip angle of 73°, a field of view
233 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
234 slices were acquired in ascending feet-to-head interleaved order with no gap. The
235 three initial scans of each acquisition run were discarded to allow for steady-state
236 magnetization. Before every two EPI runs, 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 participant 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

260 category (four motion direction, four sound source location). The canonical double-
261 gamma hemodynamic response function implemented in SPM8 was convolved with
262 a box-car function to model the above mentioned regressors. Motion parameters
263 derived from realignment of the functional volumes (3 translational motion and 3
264 rotational motion parameters), button press, and the four auditory response cue
265 events were modeled as regressors of no interest. During the model estimation, the
266 data were high-pass filtered with cut-off 128s to remove the slow drifts/ low-frequency
267 fluctuations from the time series. To account for serial correlation due to noise in fMRI
268 signal, autoregressive (AR (1)) was used.

269 In order to obtain activity related to auditory processing in the whole brain,
270 the contrasts tested the main effect of each category ([Left Motion], [Right Motion],
271 [Up Motion], [Down Motion], [Left Static], [Right Static], [Up Static], [Down Static]). To
272 find brain regions responding preferentially to the moving and static sounds, we
273 combined all motion conditions [Motion] and all static categories [Static]. The
274 contrasts tested the main effect of each condition ([Motion], [Static]), and comparison
275 between the conditions ([Motion > Static], and [Static > Motion]). These linear
276 contrasts generated statistical parametric maps (SPM[T]) which were further spatially
277 smoothed (Gaussian kernel 8 mm FWHM) and entered in a second-level analysis,
278 corresponding to a random effects model, accounting for inter-subject variance. One-
279 sample t-tests were run to characterize the main effect of each condition ([Motion],
280 [Static]), and the main effect of motion processing ([Motion > Static]) and static
281 location processing ([Static > Motion]). Statistical inferences were performed at a
282 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. 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 association test, 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 (uniformity test; P (activation | term)), this method identifies regions whose report in a neuroimaging study is diagnostic of a certain term being dominant in the study (association test; 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

305 was used to create a 6 mm spheres (117 voxels) around the peak z-values of hPT (peak
306 MNI coordinates [-56 -28 8] and [60 -28 8]; lhPT and rhPT hereafter, respectively).
307 Additionally, hPT regions of interest were individually defined using
308 anatomical parcellation with FreeSurfer (http://surfer.nmr.mgh.harvard.edu). The
309 individual anatomical scan was used to perform cortical anatomical segmentation
310 according to the Destrieux (Destrieux et al., 2010) atlas. We selected planum
311 temporale label defined as [G_temp_sup-Plan_tempo, 36] bilaterally. We equated the
312 size of ROI across participants and across hemispheres to 110 voxels (each voxel being
313 3mm isotropic). For anatomically defined hPT ROIs, all further analyses were carried
314 out in subject space for enhanced anatomical precision and to avoid spatial
315 normalization across participants. We replicated our pattern of results in anatomically
316 defined parcels (left and right hPT) obtained from the single subject brain
317 segmentation (for further analysis see Battal et al., 2018).
318
319 *ROI Analyses*
320 *Univariate*
321 The beta parameter estimates of the 4 motion directions and 4 sound source locations
322 were extracted from lhPT and rhPT regions (Fig 2C). In order to investigate the
323 presence of motion directions/sound source locations selectivity and condition effect
324 in hPT regions, we performed a 2 Conditions (hereafter refers to the stimuli type:
325 motion, static) x 4 Orientations (hereafter refers to either direction or location of the
326 stimuli: left, right, down, and up) repeated measures ANOVA in each hemisphere
327 separately on these beta parameter estimates. Statistical results were then corrected

328 for multiple comparisons (number of ROIs x number of tests) using the false discovery
329 rate (FDR) method (Benjamini and Yekutieli, 2001). A Greenhouse–Geisser correction
330 was applied to the degrees of freedom and significance levels whenever an
331 assumption of sphericity was violated.
332
333 *ROI - Multivariate pattern analyses*
334 *Within Condition Classification*
335 Four-class and binary classification analyses were conducted within the hPT
336 region in order to investigate the presence of auditory motion direction and sound
337 source location information in this area. To perform multi-voxel pattern classification,
338 we used a univariate one-way ANOVA to select a subset of voxels (n=110) that are
339 showing the most significant signal variation between the categories of stimuli (in our
340 study, between orientations). This feature-selection not only ensures a similar number
341 of voxels within a given region across participants (dimensionality reduction), but,
342 more importantly, identifies and selects voxels that are carrying the most relevant
343 information across categories of stimuli (Cox and Savoy 2003; De Martino et al., 2008),
344 therefore minimizing the chance to include in our analyses voxels carrying noises
345 unrelated to our categories of stimuli.
346 Multivariate pattern analyses (MVPA) were performed in the lhPT and rhPT.
347 Preprocessing steps were identical to the steps performed for univariate analyses,
348 except for functional time series that were smoothed with a Gaussian kernel of 2 mm
349 (FWHM). MVPA was performed using CoSMoMVPA (http://www.cosmomvpa.org/;
350 (Oosterhof et al., 2016), implemented in MATLAB. Classification analyses were

351 performed using support vector machine (SVM) classifiers as implemented in LIBSVM
352 (http://www.csie.ntu.edu.tw/~cjlin/libsvm; Chang and Lin, 2011). A general linear
353 model was implemented in SPM8, where each block was defined as a regressor of
354 interest. A beta map was calculated for each block separately. Two multi-class and six
355 binary linear SVM classifiers with a linear kernel with a fixed regularization parameter
356 of C = 1 were trained and tested for each participant separately. The two multi-class
357 classifiers were trained and tested to discriminate between the response patterns of
358 the 4 auditory motion directions and locations, respectively. Four binary classifiers
359 were used to discriminate brain activity patterns for motion and location within axes
360 (left vs. right motion, left vs. right static, up vs. down motion, up vs. down static,
361 hereafter within axis classification). We used 8 additional classifiers to discriminate
362 across axes (Left vs. Up, Left vs. Down, Right vs. Up, and Right vs. Down motion
363 directions, Left vs. Up, Left vs. Down, Right vs. Up, and Right vs. Down sound source
364 locations, hereafter across axes classification).

365 For each participant, the classifier was trained using a cross-validation leave-
366 one-out procedure where training was performed with n-1 runs and testing was then
367 applied to the remaining one run. In each cross-validation fold, the beta maps in the
368 training set were normalized (z-scored) across conditions, and the estimated
369 parameters were applied to the test set. To evaluate the performance of the classifier
370 and its generalization across all the data, the previous step was repeated 12 times
371 where in each fold a different run was used as the testing data and the classifier was
372 trained on the other 11 runs. For each region per participant, a single classification
373 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).

Within-orientation 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 in which the classifiers were trained and tested on the same orientation pairs: leftward motion vs. left static, rightward motion vs. right static, upward motion vs. up static, and downward motion vs. down static. If the same orientation (leftward and left location) across conditions (motion and static) generates similar patterns of activity in hPT

region, the classifier would not be able to differentiate leftward motion direction from left sound location. However, significant within-orientation classification would indicate that the evoked patterns within hPT contain differential information for motion direction and sound source location in the same orientation (e.g. left).

The mean of the four binary classifications was computed to produce one accuracy score per ROI. Prior to performing the within-orientation and cross-condition MVPA, each individual pattern was normalised separately across voxels so that any cross or within-orientation classification could not be due to global univariate activation differences across the conditions.

Statistical analysis: MVPA

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 individual's null distribution one value was randomly chosen and averaged across all the participants. This step was repeated 100,000 times resulting in a group level null distribution of 100,000 values. The classification accuracies across participants was considered as significant if $p<0.05$ after corrections for multiple comparisons (number of ROIs x number of tests) using the FDR method (Benjamini and Yekutieli, 2001).

419 Similar approach was adopted to assess statistical difference between the
420 classification accuracies of two auditory categories (e.g. four motion direction vs. four
421 sound source location, left motion vs. left static, left motion vs. up motion, etc.). We
422 performed additional permutation tests (100,000 iterations) by building a null
423 distribution for t-stats after randomly shuffling the classification accuracy values across
424 two auditory categories and re-calculating the two-tail t-test between the classification
425 accuracies of these two categories. All p-values were corrected for multiple
426 comparisons (number of ROIs x number of tests) using the FDR method (Benjamini
427 and Yekutieli, 2001).
428
429 *Representation Similarity analysis*
430 *Neural Dissimilarity matrices*
431 To further explore the differences in the representational format between
432 sound source locations and motion directions in hPT region, we relied on
433 representation similarity analysis (RSA; Kriegeskorte et al., 2008). More specifically,
434 we tested the correlation between the representational dissimilarity matrix (RDM) of
435 right and left hPT in each participant with different computational models that
436 included condition-invariant models assuming orientation invariance across
437 conditions (motion, static), condition-distinct models assuming that sound source
438 location and motion direction sounds elicit highly dissimilar activity patterns, and a
439 series of intermediate graded models between them. The RSA was performed using
440 CosmoMVPA toolbox (Oosterhof et al., 2016) implemented in MATLAB. To perform
441 this analysis, we first extracted in each participant 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 auditory category 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 element of the RDM contains the dissimilarity index between the patterns of activity generated by two categories, in other words the RDM represents how different is the neural representation of each category from the neural representations of all the other categories in the selected ROI. The 16 neural RDMs (1 per participant) 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) (Figure 4C).

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

465 would be no similarity between any motion directions with any sound source location.
466 The dissimilarity values in the diagonal were set to 0, simply reflecting that neural
467 responses for the same direction/location are identical to themselves.
468
469 *Condition-Invariant model*
470 The condition-invariant models assume a fully shared representation for
471 specific/corresponding static and motion conditions. For example, the models
472 consider the neural representation for the left sound source location and the left
473 motion direction highly similar. All within-condition (left, right, up and down)
474 comparisons are set to 0 (i.e. highly similar) regardless of their auditory condition. The
475 dissimilarities between different directions/locations are set to 1 meaning that each
476 within condition sound (motion or static) is different from all the other within
477 conditions.
478
479 *Intermediate models*
480 To detect the degree of similarity/shared representation between motion
481 direction and sound source location patterns, we additionally tested 2 classes of 5
482 different intermediate models. The two classes were used to deepen the
483 understanding of characteristic tuning of hPT for separate direction/location or axis of
484 motion/location. The two model classes represent 2 different possibilities. The first
485 scenario was labeled as Within-Axis Distinct, and these models assume that each of
486 the 4 directions/locations (i.e. left, right, up, down) would generate a distinctive neural
487 representation different from all of the other within-condition sounds (e.g. the

488 patterns of activity produced by the left category are highly different from the patterns
489 produced by right, up and down categories) (see Figure 4C, upper panel). To
490 foreshadow our results, we observed preference for axis of motion in MVP-
491 classification, therefore we created another class of models to further investigate
492 neural representations of within-axis and across-axes of auditory motion/space. The
493 second scenario was labeled with Within-Axis Combined, and these models assume
494 that opposite direction/locations within the same axis would generate similar patterns
495 of activity (e.g. the pattern of activity of horizontal (left and right) categories are
496 different from the patterns of activity of vertical categories (up and down) (see Figure
497 4C, lower panel).

498 In all intermediate models, the values corresponding to the dissimilarity
499 between same auditory spaces (e.g. left motion and left location) were gradually
500 modified from 0.9 (motion and static conditions are mostly distinct) to 0.1 (motion and
501 static conditions mostly similar). These models were labeled IM9, 7, 5, 3, and 1
502 respectively.

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

506

507 *Performing RSA*

508 We computed Pearson's correlation to compare neural RDMs and
509 computational model RDMs. The resulting correlation captures which computational
510 model better explains the neural dissimilarity patterns between motion direction and

511 sound source location conditions. To compare computational models, we performed
512 Mann-Whitney-Wilcoxon rank-sum test between for every pair of models. All p-values
513 were then corrected for multiple comparisons using the FDR method (Benjamini and
514 Yekutieli, 2001). To test differences between two classes of models (Within-Axis
515 Combined vs. Within-Axis Distinct), within each class, the correlation coefficient values
516 were averaged across hemispheres and across models. Next, we performed
517 permutation tests (100,000 iterations) by building a null distribution for differences
518 between classes of models after randomly shuffling the correlation coefficient values
519 across two classes, and re-calculating the subtraction between the correlation
520 coefficients of Within-Axis Combined and Within-Axis Distinct classes.

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

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

531

532 **RESULTS**

533 *Behavioral results*

534 During the experiment, we collected target direction/location discrimination
535 responses (see Figure 1C). The overall accuracy scores were entered into 2 x 4
536 (Condition, Orientation) repeated measures ANOVA. No main effect of Condition
537 ($F_{1,15}$ = 2.22; p = 0.16) was observed, indicating that the overall accuracy while
538 detecting direction of motion or sound source location did not differ. There was a
539 significant main effect of orientation ($F_{1.6,23.7}$ = 11.688; p = 0.001), caused by greater
540 accuracy in the horizontal orientations (left and right) as compared to the vertical
541 orientations (up and down). Post-hoc two-tailed t-tests (Bonferroni corrected) revealed
542 that accuracies did not reveal significant difference within horizontal orientations (left
543 vs right; t_{15} = -0.15, p=1), and within vertical orientations (up vs down; t_{15} = 0.89, p =
544 1). However, left orientation accuracy was greater as compared to down (t_{15} = 3.613,
545 p = 0.005), and up (t_{15} = 4.51, p < 0.001) orientations and right orientation accuracy
546 was greater as compared to the down (t_{15} = 3.76, p = 0.003) and up (t_{15} = 4.66, p <
547 0.001) orientation accuracies. No interaction between Condition x Orientation was
548 observed, pointing out that differences between orientations in terms of performance
549 expresses both for static and motion. A Greenhouse–Geisser correction was applied
550 to the degrees of freedom and significance levels whenever an assumption of
551 sphericity was violated.
552
553 *fMRI results – whole-brain univariate analyses*
554 To identify brain regions that are preferentially recruited for auditory motion
555 processing, we performed a univariate RFX- GLM contrast [Motion > Static] (Figure
556 2B). Consistent with previous studies (Dormal et al., 2016; Getzmann and Lewald,

557 2012; Pavani et al., 2002; Poirier et al., 2005; Warren et al., 2002), whole-brain
558 univariate analysis revealed activation in the superior temporal gyri, bilateral hPT,
559 precentral gyri, and anterior portion of middle temporal gyrus in both hemispheres
560 (Figure 2B, Table 1). The most robust activation (resisting whole brain FWE correction,
561 $p<0.05$) was observed in the bilateral hPT (peak MNI coordinates [-46 -32 10] and [60
562 -32 12]). We also observed significant activation in occipito-temporal regions (in the
563 vicinity of anterior hMT+/V5) as suggested by previous studies (Dormal et al., 2016;
564 Poirier et al., 2005; Warren et al., 2002).
565
566 *fMRI results – ROI univariate analyses*
567 Beta parameter estimates were extracted from the pre-defined ROIs (see
568 methods) for the four motion directions and four sound source locations from the
569 auditory experiment (Figure 2C). We investigated the condition effect and the
570 presence of direction/location selectivity in lhPT and rhPT regions separately by
571 performing 2 x 4 (Conditions, Orientations) repeated measures of ANOVA with beta
572 parameter estimates. In lhPT, main effect of Conditions was significant ($F_{1,15} = 37.28$,
573 $p < 0.001$), indicating that auditory motion evoked higher response compared to
574 static sounds. There was no significant main effect of Orientations ($F_{1.5,22.5} = 0.77$, $p =$
575 0.44), and no interaction ($F_{3,45} = 2.21$, $p = 0.11$). Similarly, in rhPT, only main effect of
576 Conditions was significant ($F_{1,15} = 37.02$, $p < 0.001$). No main effect of Orientation
577 ($F_{1.5,23.2} = 1.43$, $p = 0.26$) or interaction ($F_{3,45} = 1.74$, $p = 0.19$) was observed. Overall,
578 brain activity in the hPT as measured with beta parameter estimate extracted from

579 univariate analysis did not provide evidence of motion direction or sound source
580 location selectivity.

--- Figure 2 about here ---

581 *fMRI results – ROI multivariate pattern analyses*

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

586 *MVPA – Within Condition*

587 Multi-class across four conditions classification accuracy in the hPT was significantly
588 above chance (chance level = 25%) in both hemispheres for motion direction (lhPT:
589 mean ± SD = 38.4 ±7, $p < 0.001$; rhPT: mean ± SD = 37.1 ± 6.5, $p<0.001$), and sound
590 source location (lhPT: mean ± SD = 32.4 ± 6.7, $p < 0.001$; rhPT: mean ± SD = 31.2 ±
591 7.5, $p < 0.001$) (Figure 3A). In addition, we assessed the differences between
592 classification accuracies for motion and static stimuli by using permutation tests in
593 lhPT ($p = 0.024$) and rhPT ($p = 0.024$), indicating greater accuracies for classifying
594 motion direction than sound source location across all regions.

595 *MVPA – Binary Within-Axis*

596 Binary horizontal (left vs. right) within-axis classification showed significant results in
597 both lhPT and rhPT for static sounds (lhPT: mean ± SD = 58.6 ± 14.5, $p<0.001$; rhPT:
598 mean ± SD = 56.5 ± 11.9, $p = 0.008$) (Figure 3C), while motion classification was

599 significant only in the rhPT (mean ± SD = 55.5 ± 13.9, p = 0.018) (Figure 3B).
600 Moreover, binary vertical (up vs down) within-axis classification was significant only in
601 the lhPT for both motion (mean ± SD = 55.7 ± 7.1, p = 0.01), and static (mean ± SD
602 = 54.9 ± 11.9, p = 0.03) conditions (Figure 3B-C).

603 *MVPA – Binary Across-Axis*

604 We used 8 additional binary classifiers to discriminate across axes moving and static
605 sounds. Binary across-axes (Left vs. Up, Left vs. Down, Right vs. Up, and Right vs.
606 Down) classifications were significantly above chance level in the bilateral hPT for
607 motion directions (Left vs. Up lhPT: mean ± SD = 65.8 ±14.8, p<0.001; rhPT: mean ±
608 SD = 64.8 ± 9.4, p<0.001; Left vs. Down lhPT: mean ± SD = 74 ±15.9, p<0.001; rhPT:
609 mean ± SD = 66.9 ± 10.4, p<0.001; Right vs. Up lhPT: mean ± SD = 72.4 ±15.8,
610 p<0.001; rhPT: mean ± SD = 71.4 ± 13.3, p<0.001; Right vs. Down lhPT: mean ± SD
611 = 73.4 ±11.8, p<0.001; rhPT: mean ± SD = 68.2 ± 15.9, p<0.001) (Figure 3B).
612 However, sound source locations across-axes classifications were significant only in
613 certain conditions (Left vs. Up rhPT: mean ± SD = 56 ± 12.7, p=0.018; Left vs. Down
614 lhPT: mean ± SD = 55.2 ±11.2, p=0.024; rhPT: mean ± SD = 57.3 ± 17.2, p=0.003;
615 Right vs. Down lhPT: mean ± SD = 57.6 ±8.4, p=0.005) (Figure 3C).

616

617 *MVPA – "Axis of Motion" Preference*

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

One may wonder whether the higher classification accuracy for across compared to within axes relates to the perceptual differences (e.g. difficulty in localizing) 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 participants 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. 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 within axes (left vs. right and up vs. down) and across axes (left vs. up, left vs. down, right vs. up, and right

665 vs. down) was calculated. Spearman correlation between the (Δ) behavioral
666 performance and (Δ) classification accuracies was not significant in the bilateral hPT
667 (lhPT r = 0.18, p = 0.49; rhPT r = 0.4, p = 0.12). An absence of correlation suggests
668 that there is no relation between the difference in classification accuracies and the
669 perceptual difference. These results strengthen the notion that "axis of motion"
670 organization observed in PT does not simply stem from behavioral performance
671 differences.
672
673 *MVPA – Cross-condition*
674 To investigate if motion direction and sound source locations rely on shared
675 representation in hPT, we trained the classifier to distinguish neural patterns from the
676 motion directions (e.g. going to the left) and then tested on the patterns elicited by
677 static conditions (e.g. being in the left), and vice versa.
678 Cross-condition classification revealed significant results across 4
679 direction/location (lhPT: mean ± SD = 27.8 ± 5.3, p = 0.008, rhPT: mean ± SD = 28.7
680 ± 3.8, p<0.001) (Figure 3A). Within- axis categories did not reveal any significant
681 cross-condition classification. These results suggest that a partial overlap between the
682 neural patterns of moving and static stimuli in the hPT.
683
684 *MVPA – Within-orientation*
685 Cross-condition classification results indicated a shared representation
686 between motion directions and sound source locations. Previous studies argued that
687 successful cross-condition classification reflects an abstract representation of stimuli

688 conditions (Fairhall and Caramazza, 2013; Higgins et al., 2017; Hong et al., 2012). To
689 test this hypothesis, patterns of the same orientation of motion and static conditions
690 (e.g. leftward motion and left location) were involved in within-orientation MVPA. The
691 rational was that if the hPT region carries fully abstract representation of space, within-
692 orientation classification would give results in favor of the null hypothesis (no
693 differences within the same orientation). In the within-orientation classification
694 analysis, accuracies from the four within-orientation classification analyses were
695 averaged and survived FDR corrections in bilateral hPT (lhPT: mean ± SD = 65.6 ± 5,
696 $p < 0.001$, rhPT: mean ± SD = 61.9 ± 5.6, $p<0.001$), indicating that the neural patterns
697 of motion direction can be reliably differentiated from sound-source location within
698 hPT (Figure 4A).

--- Figure 4 about here ---

699 *RSA*

700 *Multi-dimensional Scaling*

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

707

708 *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 class of models and hemispheres separately (Figure 4D-E). In Within-Axis Distinct class in

732 lhPT, pairwise comparisons of correlation coefficients did not show significant
733 difference for [Condition Distinct vs. IM.9; p = 0.8], [Condition Distinct vs. IM7; p =
734 0.6; rhPT], [Condition Distinct vs. IM.5; p = 0.09], however as the models progressed
735 in further condition invariant direction, the difference between correlation coefficients
736 for models reached significance ([Condition Distinct vs. IM.3], p = 0.012; [Condition
737 Distinct vs. IM.1], p = 0.007; [Condition Distinct vs. Condition Invariant], p <0.001),
738 indicating Condition Distinct model provided stronger correlation compared to the
739 models representing conditions similarly. In rhPT, the rank-sum tests between each
740 pairs revealed no significant difference for [Condition Distinct vs. IM.9; p = 0.9],
741 [Condition Distinct vs. IM7; p = 0.7], [Condition Distinct vs. IM.5; p = 0.3], and also
742 [Condition Distinct vs. IM.3; p = 0.06], however as the models progressed in further
743 condition invariant direction, the difference between correlation coefficients for
744 models reached significance ([Condition Distinct vs. IM.1], p =0.006; [Condition
745 Distinct vs. Condition Invariant], p <0.001).

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

752

753 **DISCUSSION**

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

765 Does hPT contain information about specific motion directions and sound
766 source locations? At the univariate level, our four (left, right, up and down) motion
767 directions and sound source locations did not evoke differential univariate activity in
768 hPT region (see Figure 2C). We then carried out multivariate pattern classification and
769 observed that bilateral hPT contains reliable distributed information about the four
770 auditory motion directions (Figure 3). Our results are therefore similar to the
771 observations made with fMRI in the human visual motion area hMT+/V5 showing
772 reliable direction-selective information despite comparable voxel-wise univariate
773 activity levels across directions (Kamitani and Tong, 2006; but see Zimmermann et al.,
774 2011). Within-axis MVP-classification results revealed that even if both horizontal (left
775 versus right), and vertical (up versus down) motion directions can be classified in the
776 hPT region (see Figure 3C-D), across axes (e.g. left versus up) direction classifications

777 revealed much higher decoding accuracies compared to within-axis classifications.
778 Such enhanced classification accuracy across axes versus within axis is reminiscent of
779 observations made in MT+/V5 where a large-scale axis of motion selective
780 organization was observed in non-human primates (Albright et al., 1984), and in
781 human area MT+/V5 (Zimmermann et al., 2011). Further examination with RSA
782 provided additional evidence that within-axis combined models (aggregating the
783 opposite directions/location) explain better the neural representational space of hPT
784 by showing higher correlations values compared to within-axis distinct models (see
785 Figure 4D-E). Our findings suggest that the topographic organization principle of
786 hMT+/V5 and hPT shows similarities in representing motion directions. Animal studies
787 have shown that hMT+/V5 contains motion direction selective columns with specific
788 directions organized side by side with their respective opposing motion direction
789 counterparts (Albright et al., 1984; Born and Bradley, 2005; Diogo et al., 2003;
790 Geesaman et al., 1997), an organization also probed using fMRI in humans
791 (Zimmermann et al., 2011; but see below for alternative accounts). The observed
792 difference for within versus between axis direction decoding may potentially stem
793 from such underlying cortical columnar organization (Bartels et al., 2008; Haynes and
794 Rees, 2006; Kamitani and Tong, 2005). Alternatively, it has been suggested that
795 classifying orientation preference reflects much larger scale (e.g. retinotopy) rather
796 than columnar organization (Op de Beeck, 2010; Freeman et al., 2011, 2013).
797 Interestingly, high-field fMRI studies showed that the fMRI signal carries information
798 related to both large- and fine-scale (columnar level) biases (Gardumi et al., 2016;
799 Pratte et al., 2016; Sengupta et al., 2017). 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 (Figure 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 Figure 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; Alain et al., 2001). 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

822 source locations might not follow similar topographic representations than the one
823 observed for motion directions.

824 To which extend the neural representation of motion directions and sound
825 source locations overlaps has long been debated (Grantham, 1986; Kaas et al., 1999;
826 Poirier et al., 2017; Romanski et al., 2000; Smith et al., 2004, 2007; Zatorre and Belin,
827 2001). A number of neurophysiological studies have reported direction-selective
828 neurons along the auditory pathway (Ahissar et al. 1992; Spitzer and Semple 1993;
829 Doan et al. 1999; Stumpf et al. 1992; Toronchuk et al. 1992; McAlpine et al. 2000).
830 However, whether these neurons exhibit selectivity to motion direction specifically, or
831 also to spatial location remains debated (Poirier et al. 1997; McAlpine et al. 2000;
832 Ingham et al. 2001; Oliver 2003). In our study, we revisit this question by relying on
833 cross-condition classification that revealed that auditory motion (e.g. going to the left)
834 and sound-source location (being on the left) share partial neural representations in
835 hPT (Figure 3A). This observation suggests that there is a shared level of computation
836 between sounds located on a given position and sounds moving towards this
837 position. Low-level features of these two types of auditory stimuli vary in many ways
838 and produce large difference at the univariate level in the cortex (see Figure 2B).
839 However, perceiving, for instance, a sound going toward the left side or located on
840 the left side evoke a sensation of location/direction in the external space that is
841 common across conditions. Our significant cross-condition classification may
842 therefore relate to the evoked sensation/perception of an object being/going to a
843 common external spatial location. Electrophysiological studies in animals
844 demonstrated that motion-selective neurons in the auditory cortex displayed similar

845 response profile to sounds located or moving toward the same position in external
846 space, suggesting that the processing of sound-source locations may contribute to
847 the perception of moving sounds (Ahissar et al., 1992; Doan et al., 1999; Poirier et al.,
848 1997). Results from human psychophysiological and auditory evoked potential studies
849 also strengthen the notion that sound source location contributes to motion
850 perception (Getzmann and Lewald, 2011; Strybel and Neale, 1994). Our cross-
851 condition MVPA results therefore extend the notion that motion directions and sound
852 source locations might have common features that are partially shared for encoding
853 spatial sounds.

854 Significant cross-condition classification based on multivariate fMRI signal has
855 typically been considered as a demonstration that the region implements a partly
856 common and abstracted representation of the tested conditions (Fairhall and
857 Caramazza, 2013; Higgins et al., 2017; Hong et al., 2012). For instance, a recent study
858 elegantly demonstrated that the human auditory cortex at least partly integrates
859 interaural time and level differences (ITD and ILD) into a higher-order representation
860 of auditory space based on significant cross-cue classification (training on ITD and
861 classifying ILD, and reversely) (Higgins et al., 2017). In the present study, we argue
862 that even if cross-condition MVP-classification indeed demonstrate the presence of
863 partially shared information across sound source location and direction; successful
864 cross-MVPA results should however not be taken as evidence that the region
865 implements a purely abstract representation of auditory space. Our successful across-
866 condition classification (see Figure 4A) demonstrated that, even though there are
867 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.

Alain, C., Arnott, S.R., Hevenor, S.J., Graham, S., Grady, C.L., 2001. ''What'' and ''where'' in the human auditory system. Proc. Natl. Acad. Sci. U.S.A. *98*, 12301–12306.

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.

Battal, C., Rezk, M., Mattioni, S., Vadlamudi, J., Collignon, O. (2018). Representation of auditory motion directions and sound source locations in the human planum temporale. http://biorxiv.org/cgi/content/short/302497v2.

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.

Chang, C., and Lin, C., (2011). LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology *2*, 27:1-27:27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

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.

De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., and Formisano, E. (2008). Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. NeuroImage *43*, 44–58.

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.

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

963 Diogo, A.C.M., Soares, J.G.M., Koulakov, A., Albright, T.D., and Gattass, R.
964 (2003). Electrophysiological imaging of functional architecture in the cortical middle
965 temporal visual area of Cebus apella monkey. J. Neurosci. *23*, 3881–3898.

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

969 Dormal, G., Rezk, M., Yakobov, E., Lepore, F., and Collignon, O. (2016).
970 Auditory motion in the sighted and blind: Early visual deprivation triggers a large-
971 scale imbalance between auditory and "visual" brain regions. Neuroimage *134*, 630–
972 644.

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

976 Edelman, S., Grill-Spector, K., Kushnir, T., and Malach, R. (1998). Toward direct
977 visualization of the internal shape represetation space by fMRI. Psychobiology *26*,
978 309–321.

979 Eickhoff, S.B., Paus, T., Caspers, S., Grosbras, M.H., Evans, A.C., Zilles, K., et al.
980 (2007). Assignment of functional activations to probabilistic cytoarchitectonic areas
981 revisited. Neuroimage *36*, 511–521.

982 Fairhall, S.L., and Caramazza, A. (2013). Brain Regions That Represent Amodal

983 Conceptual Knowledge. J. Neurosci. *33*, 10552–10558.

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

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

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

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

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

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

998 Getzmann, S., and Lewald, J. (2012). Cortical processing of change in sound
999 location: Smooth motion versus discontinuous displacement. Brain Res. *1466*, 119–
1000 127.

1001 Grantham, D.W. (1986). Detection and discrimination of simulated motion of
1002 auditory targets in the horizontal plane. J. Acoust. Soc. Am. *79*, 1939–1949.

1003 Griffiths, T.D., Büchel, C., Frackowiak, R.S., and Patterson, R.D. (1998). Analysis
1004 of temporal structure in sound by the human brain. Nat. Neurosci. *1*, 422–427.

Grothe B., Pecka M., McAlpine D. (2010). Mechanisms of sound localization in mammals. *Physiol Rev* **90**:983–1012.

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.

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.

1028 Ingham, N.J., Hart, H.C., and McAlpine, D. (2001). Spatial receptive fields of
1029 inferior colliculus neurons to auditory apparent motion in free field. J. Neurophysiol.
1030 *85*, 23–33.

1031 Jiang, F., Stecker, G.C., and Fine, I. (2014). Auditory motion processing after
1032 early blindness. J. Vis. *14*, 4–4.

1033 Jiang, F., Stecker, G.C., Boynton, G.M., and Fine, I. (2016). Early Blindness
1034 Results in Developmental Plasticity for Auditory Motion Processing within Auditory
1035 and Occipital Cortex. Front. Hum. Neurosci. *10*, 324.

1036 Kaas, J.H., Hackett, T.A., and Tramo, M.J. (1999). Auditory processing in
1037 primate cerebral cortex. Curr. Opin. Neurobiol. *9*, 164–170.

1038 Kamitani, Y., and Tong, F. (2005). Decoding the visual and subjective contents
1039 of the human brain. *8*, 679–685.

1040 Kamitani, Y., and Tong, F. (2006). Decoding Seen and Attended Motion
1041 Directions from Activity in the Human Visual Cortex. Curr. Biol. *16*, 1096–1102.

1042 Kamitani, Y., and Sawahata, Y. (2010). Spatial smoothing hurts localization but
1043 not information: Pitfalls for brain mappers. Neuroimage *49*, 1949–1952.

1044 Kriegeskorte N, Mur M, Bandettini P. (2008). Representational similarity
1045 analysis—connecting the branches of systems neuroscience. Front. Syst. Neurosci. *2,*
1046 1-28.

1047 Kriegeskorte, N., Formisano, E., Sorger, B., and Goebel, R. (2007). Individual
1048 faces elicit distinct response patterns in human anterior temporal cortex. Proc. Natl.
1049 Acad. Sci. *104*, 20600–20605.

1050 Krumbholz, K., Schönwiesner, M., Rübsamen, R., Zilles, K., Fink, G.R., and von

1051 Cramon, D.Y. (2005). Hierarchical processing of sound location and motion in the
1052 human brainstem and planum temporale. Eur. J. Neurosci. *21*, 230–238.

1053 Lewis, J.W., Beauchamp, M.S., and DeYoe, E.A. (2000). A comparison of visual
1054 and auditory motion processing in human cerebral cortex. Cereb.Cortex *10*, 873–888.

1055 Lomber, S.G., and Malhotra, S. (2008). Double dissociation of "what" and
1056 "where" processing in auditory cortex. Nat. Neurosci. *11*, 609–616.

1057 Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P.,
1058 et al. (2001). Distinct pathways involved in sound recognition and localization: a
1059 human fMRI study. Neuroimage *14*, 802–816.

1060 McAlpine, D., Jiang, D., Shackleton, T.M., and Palmer, A.R. (2000). Responses
1061 of neurons in the inferior colliculus to dynamic interaural phase cues: evidence for a
1062 mechanism of binaural adaptation. J Neurophysiol *83*, 1356– 1365.

1063 Middlebrooks, J., and Pettigrew, J. (1981). Functional classes of neurons in
1064 primary auditory cortex of the cat distinguished by sensitivity to sound location. J
1065 Neurosci *1*, 107–120.

1066 Middlebrooks, J.C., and Green, D.M. (1991). Sound localization by human
1067 listeners. Annu. Rev. Psychol. *42*, 135–159.

1068 Middlebrooks, J.C. (2002). Auditory space processing: Here, there or
1069 everywhere? Nat. Neurosci. *5*, 824–826.

1070 Middlebrooks, J.C., and Bremen, P. (2013). Spatial Stream Segregation by
1071 Auditory Cortical Neurons. J. Neurosci. *33*, 10986–11001.

1072 Movshon, J.A., and Newsome, W.T. (1996). Visual response properties of striate
1073 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.

Oliver, D. L., Beckius, G. E., Bishop, D. C., Loftus, W. C., & Batra, R. (2003). Topography of interaural temporal disparity coding in projections of medial superior olive to inferior colliculus. *The Journal of neuroscience 23*, 7438-7449.

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.

Spitzer, M.W., and Semple, M.N. (1993). Responses of inferior colliculus neurons to time-varying interaural phase disparity: effects of shifting the locus of virtual motion. J Neurophysiol *69* 1245–1263.

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.

Stumpf, E., Toronchuk, J.M., Cynader, M.S., (1992). Neurons in cat primary auditory cortex sensitive to correlates of auditory motion in three-dimensional space. Exp. Brain Res. *88*, 158–168.

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.

Toronchuk, J.M., Stumpf, E., Cynader, M.S., (1992). Auditory cortex neurons sensitive to correlates of auditory motion: underlying mechanisms. Exp. Brain Res. *88*, 169–180.

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.

1166 (2011). Large-scale automated synthesis of human functional neuroimaging data. Nat.
1167 Methods *8*, 665–670.

1168 Young, E.D., Spirou, G.A., Rice, J.J., Voigt, H.F., and Rees, A. (1992). Neural
1169 Organization and Responses to Complex Stimuli in the Dorsal Cochlear Nucleus.
1170 Philos. Trans. R. Soc. B Biol. Sci. *336*, 407–413.

1171 Zabicki, A., De Haas, B., Zentgraf, K., Stark, R., Munzert, J., and Krüger, B.
1172 (2016). Imagined and Executed Actions in the Human Motor System: Testing Neural
1173 Similarity Between Execution and Imagery of Actions with a Multivariate Approach.
1174 Cereb. Cortex *27*, 1–14.

1175 Zatorre, R.J., and Belin, P. (2001). Spectral and Temporal Processing in Human
1176 Auditory Cortex. Cereb. Cortex *11*, 946–953.

1177 Zatorre, R.J., Bouffard, M., Ahad, P., and Belin, P. (2002). Where is "where" in
1178 the human auditory cortex? Nat. Neurosci. *5*, 905–909.

1179 Zeng, H., and Constable, R.T. (2002). Image distortion correction in EPI:
1180 Comparison of field mapping with point spread function mapping. Magn. Reson.
1181 Med. *48*, 137–146.

1182 Zhou, Y., and Wang, X. (2012). Level dependence of spatial processing in the
1183 primate auditory cortex. J. Neurophysiol. *108*, 810–826.

1184 Zimmermann, J., Goebel, R., de Martino, F., van de Moortele, P.F., Feinberg,
1185 D., Adriany, G., et al. (2011). Mapping the Organization of Axis of Motion Selective
1186 Features in Human Area MT Using High-Field fMRI. PLoS One *6*, 1–10.

1187

1188 **FIGURE LEGENDS**

1189 **Figure 1. Stimuli and Experimental Design. (A).** The acoustic apparatus used to
1190 present auditory moving and static sounds while binaural recordings were carried out
1191 for each participant before the fMRI session. **(B).** Auditory stimuli presented inside the
1192 MRI scanner consisted of 8 conditions: leftward, rightward, downward and upward
1193 moving stimuli and left, right, down and up static stimuli. Each condition was
1194 presented for 15 s (12 repetition of 1250 ms sound, no ISI) and followed by 7 s gap
1195 for indicating the corresponding direction/location in space and 8 s of silence (total
1196 inter-block interval was 15 s). Sound presentation and response button press were
1197 pseudo-randomized. Participants were asked to respond as accurately as possible
1198 during the gap period. **(C).** The behavioural performance inside the scanner.

1199 **Figure 2. Univariate whole brain results. (A).** Association test map was obtained from
1200 the online tool Neurosynth using the term "Planum Temporale" (FDR corrected p
1201 <0.05). The black spheres are illustration of drawn mask (radius = 6mm, 117 voxels)
1202 around the peak coordinate from Neurosynth (search term "planum temporale",
1203 meta-analysis of 85 studies). **(B).** Auditory motion processing [motion > static]
1204 thresholded at p <0.05 FWE-corrected. **(C).** Mean activity estimates (arbitrary units ±
1205 SEM) associated with the perception of auditory motion direction (red) and sound-
1206 source location (blue). ML: motion left, MR: motion right, MD: motion down, MU:
1207 motion up, SL: static left, SR: static right, SD: static down, and SU: static up.

1208 **Figure 3. Within and cross-classification results. (A).** Classification results for the 4
1209 conditions in functionally defined hPT region. Within-condition and cross-condition
1210 classification results are shown in the same bar plots. Moving: four motion direction;

1211 Static: four sound source location; and Cross: cross-condition classification accuracies.
1212 **(B).** Classification results of within (left vs. right, up vs. down) and across axes (left vs.
1213 up, left vs. down, right vs. up, right vs. down) motion directions. **(C).** Classification
1214 results of within (left vs. right, up vs. down) and across axes (left vs. up, left vs. down,
1215 right vs. up, right vs. down) sound source locations. LvsR: Left vs. Right, UvsD: Up vs.
1216 Down, LvsU: Left vs. Up, LvsD: Left vs. Down, RvsU: Right vs. Up, RvsD: Right vs. Down
1217 classifications. FDR corrected p-values: (*) $p<0.05$, (**) $p<0.01$, (***) $p<0.001$ testing
1218 differences against chance level (dotted lines; see methods).

1219 **Figure 4. Pattern dissimilarity between motion directions and sound source locations.**
1220 **(A).** Across-condition classification results of across 4 conditions are represented in
1221 each ROI (lhPT and rhPT). 4 binary classifications [leftward motion vs. left location],
1222 [rightward motion vs. right location], [upward motion vs. up location], and [downward
1223 motion vs. down location] were computed and averaged to produce one accuracy
1224 score per ROI. FDR corrected p-values: (***) $p<0.001$. Dotted lines represent chance
1225 level. **(B).** The embedded top panel shows neural RDMs extracted from left and right
1226 hPT, and multi-dimensional scaling (MDS) plot visualizes the similarities of the neural
1227 pattern elicited by 4 motion directions (arrows) and 4 sound source locations (dots).
1228 Color codes for arrow/dots are as follows: green indicates left direction/location, red
1229 indicates right direction/location, orange indicates up direction/location, and blue
1230 indicates down direction/location. ML: motion left, MR: motion right, MD: motion
1231 down, MU: motion up, SL: static left, SR: static right, SD: static down, and SU: static
1232 up. **(C-E).** The results of representational similarity analysis (RSA) in hPT are
1233 represented. **(C).** RDMs of the computational models that assume different similarities

of the neural pattern based on auditory motion and static conditions. **(D-E).** 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).

1244 **TABLES**

1245 **Table 1.** Results of the univariate analyses for the main effect of auditory motion
1246 processing [motion > static], and auditory localization processing [static > motion].
1247 Coordinates reported in this table are significant ($p < 0.05$ FWE) after correction over
1248 small spherical volumes (SVC, 12 mm radius) of interest (#) or over the whole brain (*).
1249 Coordinates used for correction over small spherical volumes are as follows (x, y, z, in
1250 MNI space): left middle temporal gyrus (hMT+/V5) [-42 -64 4] (Dormal et al., 2016),
1251 right middle temporal gyrus (hMT +/V5) [42 – 60 4] (Dormal et al., 2016), right superior
1252 frontal sulcus [32 0 48] (Collignon et al., 2011), right middle occipital gyrus [48 -76 6]
1253 (Collignon et al., 2011). K represents the number of voxels when displayed at p(unc)
1254 < 0.001. L: left, R: right, G: gyrus, S: sulcus.

1255

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#

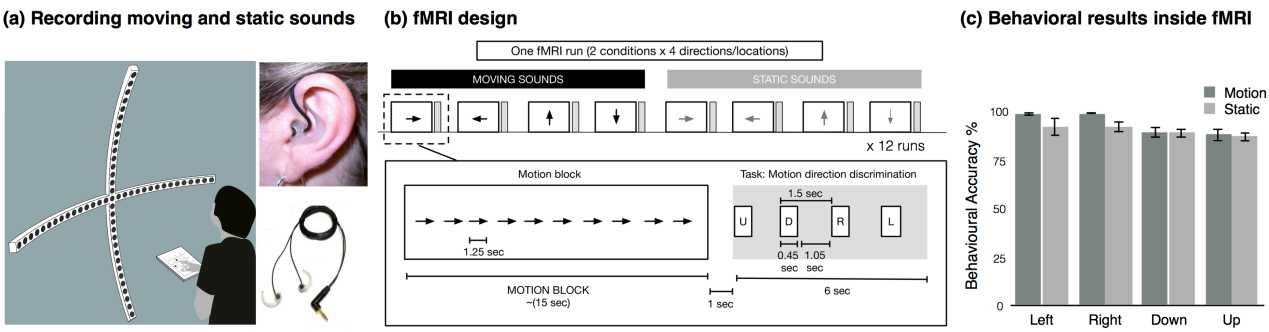
(a) Recording moving and static sounds
(b) fMRI design
(c) Behavioral results inside fMRI
One fMRI run (2 conditions x 4 directions/locations)
MOVING SOUNDS
STATIC SOUNDS
x 12 runs
Motion block
1.25 sec
MOTION BLOCK
~(15 sec)
1 sec
Task: Motion direction discrimination
1.5 sec
U
D
R
L
0.45 sec
1.05 sec
6 sec
Motion
Static
Behavioural Accuracy %
100
75
50
25
0
Left
Right
Down
Up

Figure 1

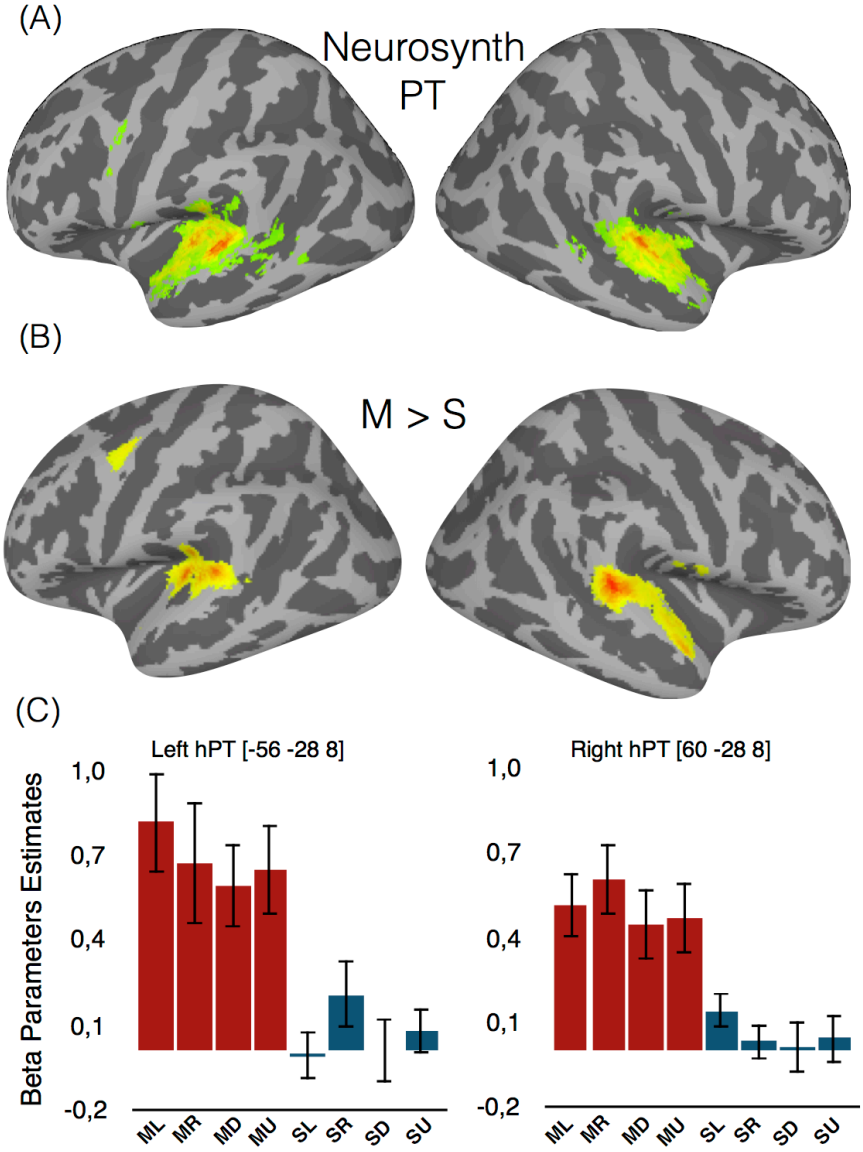
(A)
Neurosynth
PT
(B)
M > S
(C)
Left hPT [-56 -28 8]
Right hPT [60 -28 8]
Beta Parameters Estimates
1,0
0,7
0,4
0,1
-0,2
ML
MR
MD
MU
SL
SR
SD
SU

Figure 2

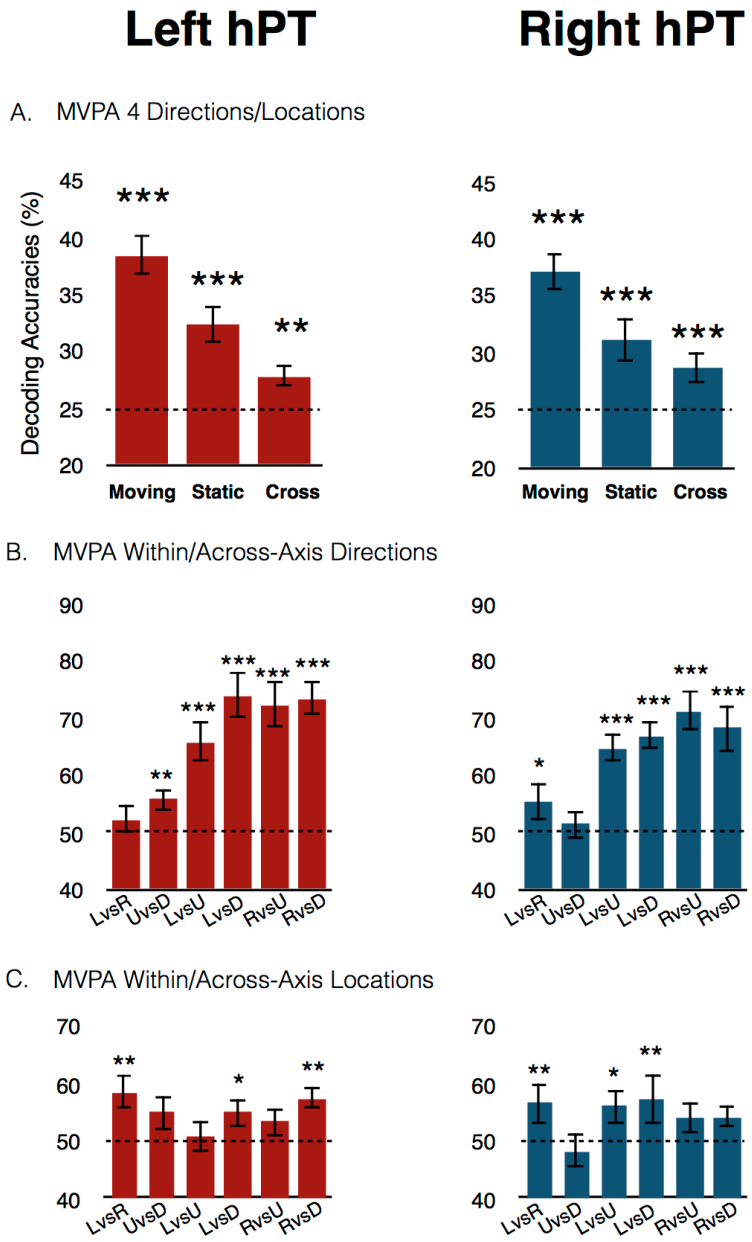
Left hPT
Right hPT
A. MVPA 4 Directions/Locations
Decoding Accuracies (%)
Moving
Static
Cross
B. MVPA Within/Across-Axis Directions
LvsR
UvsD
LvsU
LvsD
RvsU
RvsD
C. MVPA Within/Across-Axis Locations

Figure 3

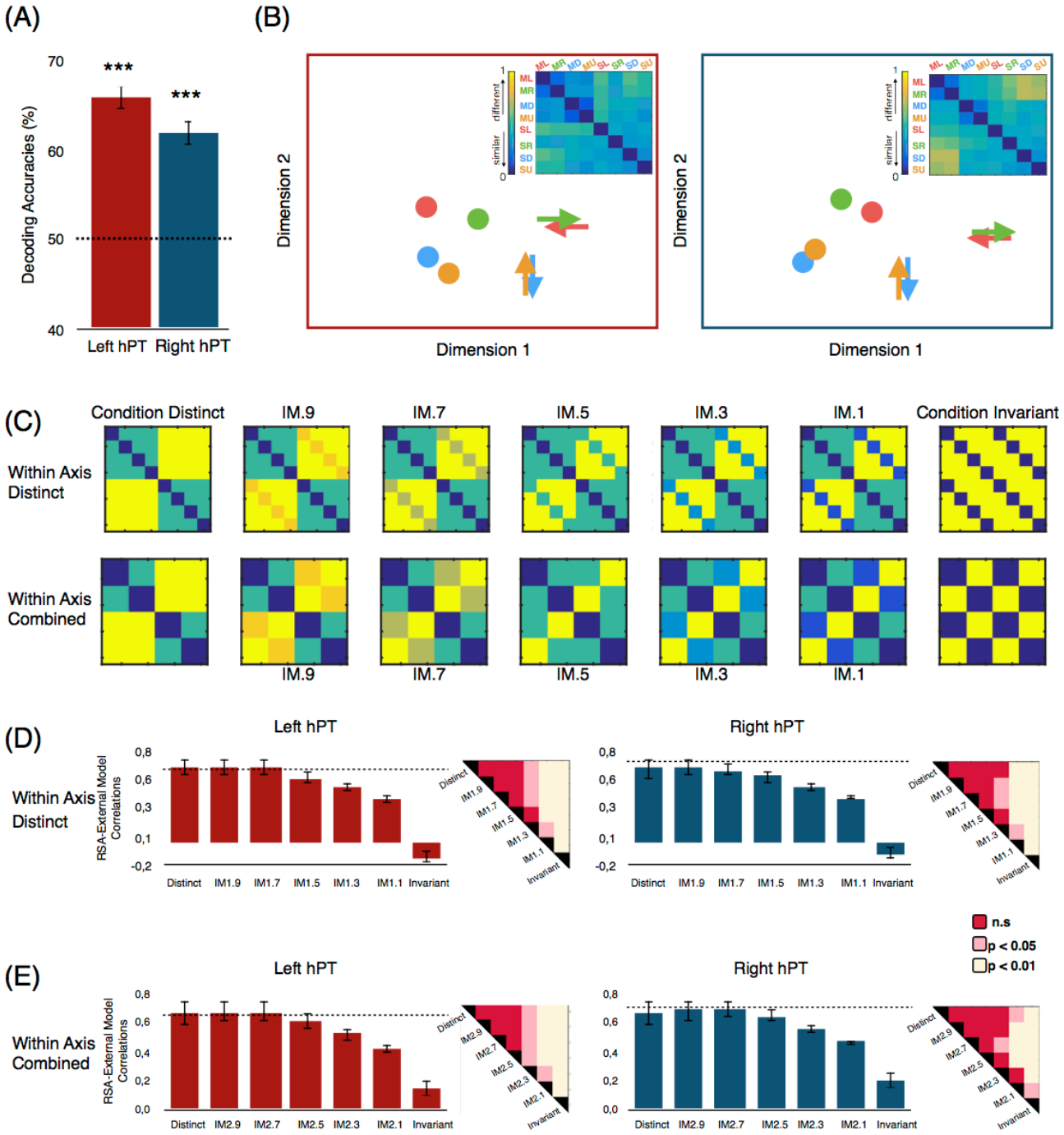
(A)
(B)
Decoding Accuracies (%)
70
60
50
40
***
***
Left hPT
Right hPT
Dimension 2
Dimension 1
ML
MR
MD
MU
SL
SR
SD
SU
different
similar
(C)
Condition Distinct
IM.9
IM.7
IM.5
IM.3
IM.1
Condition Invariant
Within Axis Distinct
Within Axis Combined
(D)
Left hPT
Right hPT
Within Axis Distinct
RSA-External Model Correlations
Distinct
IM1.9
IM1.7
IM1.5
IM1.3
IM1.1
Invariant
n.s
p < 0.05
p < 0.01
(E)
Within Axis Combined
IM2.9
IM2.7
IM2.5
IM2.3
IM2.1

Figure 4