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 sounds moving left, right, up, and down 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.

**Key words:** auditory motion; direction selectivity; fMRI; multivariate analyses; planum temporale; spatial hearing

**Significance Statement**

Compared with 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 light on how computing the location or direction of sounds is 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 middle-temporal cortex (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 (Newsome and Paré, 1988; Movshon and Newsome, 1996; Braddick et al., 2001), 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 (Searle et al., 1976; Blauert, 1982). In the auditory pathway, these cues are both processed and integrated in the brainstem, thalamus, and cortex to create an integrated neural representation of auditory space (Boudreau and Tsuchitani, 1968; Goldberg and Brown, 1969; Knudsen and Konishi, 1978; Ingham et al., 2001; for review, see Grothe et al., 2010).

Similar to the dual-stream model in vision (Ungerleider and Mishkin, 1982; Goodale and Milner, 1992), partially distinct ventral “what” and dorsal “where” auditory processing streams have been proposed in audition (Romanski et al., 1999; Rauschecker and Tian, 2000; Recanzone, 2000; Tian et al., 2001; Warren and Griffiths, 2003; Barrett and Hall, 2006; Lomber and Malhotra, 2008; Collignon et al., 2011). In particular, the dorsal stream is
thought to process sound source location and motion both in animals and humans (Rauschecker and Tian, 2000; Alain et al., 2001; Maeder et al., 2001; Tian et al., 2001; Arnott et al., 2004). 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 (Zatorre et al., 2002; Smith et al., 2010; Poirier et al., 2017).

One candidate region that might integrate spatial cues to compute motion and location information is the human planum temporale (hPT; Baumgart et al., 1999; Warren et al., 2002; Barrett and Hall, 2006). hPT is located in the superior temporal gyrus, posterior to Heschl’s gyrus, and is typically considered part of the dorsal auditory stream (Rauschecker and Tian, 2000; Warren et al., 2002; Derey et al., 2016; Poirier et al., 2017). Some authors have suggested that hPT engages in both the processing of moving sounds and the location of static sound sources (Zatorre et al., 2002; Smith et al., 2004, 2007, 2010; Krumbholz et al., 2005; Barrett and Hall, 2006; Derey et al., 2016). 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; Middlebrooks and Pettigrew, 1981; Imig et al., 1990; Rajan et al., 1990; Poirier et al., 1997; Doan et al., 1999). In contrast, other studies in animals (Poirier et al., 2017) and humans (Baumgart et al., 1999; Lewis et al., 2000; Bremmer et al., 2001; Pavani et al., 2002; Hall and Moore, 2003; Krumbholz et al., 2005; 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 human middle temporal cortex (hMT+/V5) in the visual system. hMT+/V5 is dedicated to process visual motion (Watson et al., 1993; Tootell et al., 1995; Movshon and Newsome, 1996) and displays a columnar organization tuned to axis of motion direction (Albright et al., 1984; Zimmermann et al., 2011). However, whether hPT discloses similar characteristic tuning properties remains unknown.

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

Materials and Methods

Participants

Eighteen participants with no reported auditory problems were recruited for the study. Two participants were excluded due to poor spatial hearing performance in the task, as it was lower by >2.5 SDs than the average of the participants. The final sample included 16 right-handed participants (8 females; age range, 20–42 years; mean ± SD age, 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 and the University of Trento. Experiments were undertaken with the understanding and written consent of each participant.

Auditory stimuli

Our limited knowledge of the auditory space processing in the human brain might be a consequence of the technical challenge of evoking a vivid perceptual experience of auditory space while using neuroimaging tools such as fMRI, EEG, or MEG. In this experiment, to create an externalized ecological sensation of sound location and motion inside the MRI scanner, we relied on individual in-ear stereo recordings that were recorded in a semianechoic room and on 30 loudspeakers on horizontal and vertical planes, mounted on two semicircular wooden structures with a radius of 1.1 m (Fig. 1A). Participants were seated in the center of the apparatus with their head on a chinrest, such that the speakers on the horizontal and vertical planes were equally distant from participants’ ears. Then, these recordings were replayed to the participants when they were inside the MRI scanner. By using such a sound system with in-ear recordings, auditory stimuli automatically convolved with each individual’s own pinna and head-related transfer function to produce a salient auditory perception in external space.

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

Stimuli recording

Sound stimuli consisted of 1250 ms pink noise (50 ms rise/fall time). In the motion condition, the pink noise was presented moving in the following four directions: leftward, rightward, upward, and downward. Moving stimuli covered 120° of the 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 pieces of equal length 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, 96.7°/s; fade in/out, 50 ms) containing the same sounds for all four directions. The choice of the movement speed of the motion stimuli aimed to create a listening experience relevant to conditions of everyday life. 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). In the static condition, the same pink noise was presented separately at one of the following four locations: left, right, up, and down. Static sounds were presented at the second-most outer speakers (−56° and +56° in the horizontal axis, and ±56° and ±56° in the vertical axis) to avoid possible reverberation differences at the outermost speakers. The static sounds were fixed at one location at a time instead of presented in multiple locations (Krumholz et al., 2005; Smith et al., 2004, 2010; Poirier et al., 2017). This strategy was purposely adopted for two main reasons. First, randomly presented static sounds can evoke a robust sensation of auditory apparent motion (Strybel and Neale, 1994; Lakatos and Shepard, 1997; for review, see Carlile and Leung, 2016). Second, and crucially for the purpose of the present experiment, presenting static sounds located in a given position and moving sounds directed toward the same position allowed us to investigate whether moving and static sounds share a common representational format using cross-condition classification (see below), which would have been impossible if the static sounds were randomly moving in space.

Before the recordings, the sound pressure levels (SPLs) were measured from the participant’s head position and ensured that each speaker conscience and ecological auditory space sensation (Pavani et al., 2002).

Auditory experiment

Auditory stimuli were presented via MR-compatible closed-ear head-phones (500–10 KHz frequency response; Serene Sound, Resonance Technology) that provided average ambient noise cancellation of −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 >80% accuracy.

Each run consisted of the eight auditory categories (four motion and four static) randomly presented using a block design. Each category of sound was presented for 15 s [12 repetitions of 1250 ms sound, no inter-stimulus interval (ISI)] and followed by 7 s gap for indicating the corresponding direction/location in space and 8 s of silence (total interblock interval, 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 pseudorandomized order. Participants were asked to press a button with their right index finger when the direction or location of the auditory block was matching with the auditory cue (Fig. 1B). The number of targets and the order (positions 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 the accuracy of response but not the reaction times.

Each session consisted of one block of each category, resulting in a total of eight blocks per run, with each run lasting 4 min and 10 s. Participants completed a total of 12 runs. The order of the blocks was randomized within each run and across participants.

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

fMRI data acquisition and analyses

Imaging parameters

Functional and structural data were acquired with a 4 T Bruker MedSpec Biospin MR scanner, equipped with an eight-channel head coil. Functional images were acquired with T2*-weighted gradient echoplanar sequence. Acquisition parameters were as follows: repetition time (TR), 2500 ms; echo time (TE), 26 ms; flip angle (FA), 73°; field of view, 192 mm; matrix size, 64 × 64; and voxel size, 3 × 3 × 3 mm3. A total of 39 slices was acquired in an ascending feet-to-head interleaved order with no gap. The three initial scans of each acquisition run were discarded to allow for steady-state magnetization. Before every two EPI runs, we performed an additional scan to measure the point-spread function of the acquired sequence, including fat saturation, which served for distortion correction that is expected with high-field imaging (Zeng and Constable, 2002).

A high-resolution anatomical scan was acquired for each participant using a T1-weighted 3D MP-RAGE sequence (176 sagittal slices; voxel size, 1 × 1 × 1 mm3; field of view, 256 × 224 mm; TR, 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 preprocessed and analyzed with SPM8 [Welcome Trust Centre for Neuroimaging London, UK (https://www.fil.ion.UCL.ac.uk/spm/software/spm8/)] 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 EPI template conforming to the Montreal Neurological Institute (MINI) space, and spatial smoothing (Gaussian kernel, 6 mm FWHM).

To obtain blood oxygenation level-dependent activity related to auditory spatial processing, we computed single-subject statistical comparisons with a fixed-effect general linear model (GLM). In the GLM, we used eight regressors from each category (four motion direction, four sound source location). The canonical double-gamma hemodynamic response function implemented in SPM8 was convolved with a box-car function to model the above-mentioned regressors. Motion parameters derived from realignment of the functional volumes (three translational motion and three rotational motion parameters), button press, and the four auditory response cue events were modeled as regressors of no interest. During the model estimation, the data were high-pass filtered with a cutoff 128 s to remove the slow drifts/low-frequency fluctuations from the time series. To account for serial correlation due to noise in fMRI signal, autoregression [AR (1)] was used.
To obtain activity related to auditory processing in the whole brain, the contrasts tested the main effect of each category ([Left Motion], [Right Motion], [Up Motion], [Down Motion], [Left Static], [Right Static], [Up Static], [Down Static]). To find brain regions responding preferentially to the moving and static sounds, we combined all motion conditions [Motion] and all static categories [Static]. The contrasts tested the main effect of each condition ([Motion], [Static]), and comparison between the conditions ([Motion > Static], and [Static > Motion]). These linear contrasts generated statistical parametric maps (SPM[T]), which were further spatially smoothed (Gaussian kernel, 8 mm FWHM) and entered in a second-level analysis, corresponding to a random-effects model, accounting for intersubject variance. One-sample t tests were run to characterize the main effect of each condition ([Motion], [Static]) and the main effect of motion processing ([Motion > Static]) and static location processing ([Static > Motion]). Statistical inferences were performed at a threshold of \( p < 0.05 \) corrected for multiple comparisons (familywise error (FWE) corrected) either over the entire brain volume or after correction for multiple comparisons over small spherical volumes (12 mm radius) located in regions of interest (ROIs). 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).

ROI analyses

ROI definition. Due to the hypothesis-driven nature of our study, we defined hPT as an a priori ROI for statistical comparisons and 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 on-line 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 ROI analyses. The peak coordinate from the meta-analysis map was used to create 6 mm spheres (117 voxels) around the peak z-values of hPT [peak MNI coordinates (-56, -28, 8) and (60, -28, 8); left hPT (lhPT) and right hPT (rhPT) hereafter].

Additionally, hPT regions of interest were individually defined using anatomical parcellation with FreeSurfer (http://surfer.nmr.mgh.harvard.edu). The individual anatomical scan was used to perform cortical anatomical segmentation according to the atlas by Destrieux et al. (2010). We selected a planum temporale label defined as [G_temp_sup_Plan_tempo, 36] bilaterally. We equated the size of the ROI across participants and across hemispheres to 110 voxels (each voxel being 3 mm isotropic). For anatomically defined hPT ROIs, all further analyses were performed in subject space for enhanced anatomical precision and to avoid spatial normalization across participants. We replicated our pattern of results in anatomically defined parcels (lhPT and rhPT) obtained from the single-subject brain segmentation (for further analysis, see Battal et al., 2018).

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

ROI multivariate pattern analyses

Within-condition classification. Four-class and binary classification analyses were conducted within the hPT region to investigate the presence of auditory motion direction and sound source location information in this area. To perform multivoxel pattern classification, we used a univariate one-way ANOVA to select a subset of voxels (\( n = 110 \)) that are showing the most significant signal variation between the categories of stimuli (in our study, between orientations). This feature selection not only ensures a similar number of voxels within a given region across participants (dimensionality reduction), but, more importantly, identifies and selects voxels that are carrying the most relevant information across categories of stimuli (Cox and Savoy, 2003; De Martino et al., 2008), therefore minimizing the chance to include in our analyses voxels carrying noises unrelated to our categories of stimuli.

MVPA were performed in the lhPT and rhPT. Preprocessing steps were identical to the steps performed for univariate analyses, except for functional time series that were smoothed with a Gaussian kernel of 2 mm (FWHM). MVPA was performed using CoSMoMVPA (http://www.cosmomvpa.org; Oosterhof et al., 2016), implemented in MATLAB. Classification analyses were performed using support vector machine (SVM) classifiers as implemented in LIBSVM (https://www.csie.ntu.edu.tw/~cjlin/libsvm/; Chang and Lin, 2011). A general linear model was implemented in SPM8, where each block was defined as a

Figure 2. Univariate whole-brain results. A, Association test map was obtained from the on-line tool Neurosynth using the term Planum Temporale (FDR corrected, \( p < 0.05 \)). The black spheres are illustrations of a drawn mask (radius = 6 mm, 117 voxels) around the peak coordinate from Neurosynth (search term Planum Temporale, meta-analysis of 85 studies). B, Auditory motion processing (Motion > Static) thresholded at \( p < 0.05 \), whole-brain FWE corrected. C, Mean activity estimates (arbitrary units \( \pm \) SEM) associated with the perception of auditory motion direction (red) and sound source location (blue). ML, Motion left; MR, motion right; MD, motion down; MU, motion up; SL, static left; SR, static right; SD, static down; and SU, static up.

A

Neurosynth PT

B

M > S

C

Left hPT [-56 -28 8]

Right hPT [60 -28 8]
regressor of interest. A \( \beta \) map was calculated for each block separately. Two multiclass and six binary linear SVM classifiers with a linear kernel with a fixed regularization parameter of \( C = 1 \) were trained and tested for each participant separately. The two multiclass classifiers were trained and tested to discriminate among the response patterns of the four auditory motion directions and locations, respectively. Four binary classifiers were used to discriminate brain activity patterns for motion and location within axes (left vs right motion, left vs right static, up vs down motion, up vs down static; hereafter called within-axis classification). We used eight additional classifiers to discriminate across axes (left vs up, left vs down, right vs up, and right vs down motion directions; left vs up, left vs down, right vs up, and right vs down sound source locations; hereafter called across-axes classification).

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

Cross-condition classification. To test whether motion directions and sound source locations share a similar neural representation in hPT region, we performed cross-condition classification. We performed 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 applied 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 performed).

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 four binary classifications in which the classifiers were trained and tested on the same orientation pairs: leftward motion versus left static, rightward motion versus right static, upward motion versus up static, and downward motion versus 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. Before performing the within-orientation and cross-condition MVPA, each individual pattern was normalized separately across voxels so that any cross-orientation 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 nonparametric tests permuting condition labels and bootstrapping (Stelzer et al., 2013). Each permutation step included shuffling of the condition labels and rerunning the classification 100 times on the single-subject level. Next, we applied a bootstrapping procedure to obtain a group-level null distribution that is representative of the whole group. From each individual’s null distribution, one value was randomly chosen and averaged across all of 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 \( \times \) number of tests) using the FDR method (Benjamini and Yekutieli, 2001).

A similar approach was adopted to assess statistical difference between the classification accuracies of two auditory categories (e.g., four motion direction vs four sound source location, left motion vs left static, left sound source location vs left motion, etc.). We performed additional permutation tests (100,000 iterations) by building a null distribution for \( t \) statistics after randomly shuffling the classification accuracy values across two auditory categories and recalculating the two-tailed \( t \) test between the classification accuracies of these two categories. All \( p \) values were corrected for multiple comparisons (number of ROIs \( \times \) number of tests) using the FDR method (Benjamini and Yekutieli, 2001).

Representation similarity analysis

Neural dissimilarity matrices

To further explore the differences in the representational format between sound source locations and motion directions in hPT region, we relied on representation similarity analysis (RSA; Kriegeskorte et al., 2008). More specifically, we tested the correlation between the representational dissimilarity matrix (RDM) of rhPT and lhPT in each participant with different computational models that included condition-invariant models assuming orientation invariance across conditions (motion, static), condition-distinct models assuming that sound source location and motion direction sounds elicit highly dissimilar activity patterns, and a series of intermediate graded models between them. The RSA was performed using the CosmoMVPA toolbox (Oosterhof et al., 2016) implemented in MATLAB. To perform 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) 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 a \( 8 \times 8 \) correlation matrix for each participant that was then converted into an RDM 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 the neural representation of each category is from the neural representations of all the other categories in the selected ROI. The 16 neural RDMs (1 per participant) for each of the two 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., 2017; see Fig. 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 would be no similarity between any motion directions with any sound source location. The dissimilarity values in the diagonal were set to 0, simply reflecting that neural responses for the same direction/location are identical to themselves.

Condition-invariant model. The condition-invariant models assume a fully shared representation for specific/corresponding static and motion conditions. For example, the models consider the neural representation for the left sound source location and the left motion direction highly similar. All within-condition (left, right, up, and down) comparisons are set to 0 (i.e., highly similar), regardless of their auditory condition. The dissimilarities between different directions/locations are set to 1, meaning that each within-condition sound (motion or static) is different from all the other within-condition sounds.

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

In all intermediate models, the values corresponding to the dissimilarity between same auditory spaces (e.g., left motion and left location) were gradually modified from 0.9 (motion and static conditions are mostly distinct) to 0.1 (motion and static conditions mostly similar). These models were labeled IM9, IM7, IM5, IM3, and IM1, respectively.

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

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

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

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

Results
Behavioral results
During the experiment, we collected target direction/location discrimination responses (Fig. 1C). The overall accuracy scores were entered into 2 × 4 (Condition, Orientation) repeated-measures ANOVA. No main effect of Condition (F₁,₁₁₅ = 2.22; p = 0.16) was observed, indicating that the overall accuracy while detecting the direction of motion or sound source location did not differ. There was a significant main effect of orientation (F₁,₁₁₅ = 11.688; p = 0.001), caused by greater accuracy in the horizontal orientations (left and right) compared with the vertical orientations (up and down). Post hoc two-tailed t tests (Bonferroni corrected) revealed that accuracies did not reveal significant differences within horizontal orientations (left vs right; t₁₁₅ = −0.15, p = 1), and within vertical orientations (up vs down; t₁₁₅ = 0.89, p = 1). However, left orientation accuracy was greater compared with down (t₁₁₅ = 3.613, p = 0.005) and up (t₁₁₅ = 4.51, p < 0.001) orientation accuracies, and right orientation accuracy was greater compared with the down (t₁₁₅ = 3.76, p = 0.003) and up (t₁₁₅ = 4.66, p < 0.001) orientation accuracies. No interaction between Condition and Orientation was observed, pointing out that differences between orientations in terms of performance expresses both for static and motion. A Greenhouse–Geisser correction was applied to the degrees of freedom and significance levels whenever an assumption of sphericity was violated.

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

ROI univariate analyses
Beta parameter estimates were extracted from the predefined ROIs (see Materials and Methods) for the four motion directions and four sound source locations from the auditory experiment (Fig. 2C). We investigated the condition effect and the presence of direction/location selectivity in lhPT and rhPT regions separately by performing 2 × 4 (Conditions, Orientations) repeated measures of ANOVA with β parameter estimates. In lhPT, the

<table>
<thead>
<tr>
<th>Area</th>
<th>MNI coordinates (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Motion &gt; Static</td>
<td></td>
</tr>
<tr>
<td>L planum temporale</td>
<td>10,506</td>
</tr>
<tr>
<td>L Middle temporal G</td>
<td>−56</td>
</tr>
<tr>
<td>L Precentral G</td>
<td>−46</td>
</tr>
<tr>
<td>L Putamen</td>
<td>−22</td>
</tr>
<tr>
<td>L Middle temporal G</td>
<td>43</td>
</tr>
<tr>
<td>R Superior temporal G</td>
<td>7074</td>
</tr>
<tr>
<td>R Superior temporal G</td>
<td>62</td>
</tr>
<tr>
<td>R Superior temporal G</td>
<td>52</td>
</tr>
<tr>
<td>R Precentral G</td>
<td>50</td>
</tr>
<tr>
<td>R Superior frontal S</td>
<td>159</td>
</tr>
<tr>
<td>R Middle temporal G</td>
<td>136</td>
</tr>
<tr>
<td>R Middle occipital G</td>
<td>24</td>
</tr>
</tbody>
</table>

Coordinates reported in this table are significant (p < 0.05, FWE corrected) after correction over small spherical volumes (10 mm radius) of interest (ROIs) or the whole brain (whole brain). The coordinates used for correction over small SUVs are as follows (in x, y, z, in MNI space): left middle temporal gyrus (MT+V5) [−42, −64, 6] (Dormal et al., 2016); right middle temporal gyrus (MT+V5) [42, −60, 4] (Dormal et al., 2016); right superior frontal sulcus [32, 6, 48] (Collignon et al., 2011); right middle occipital gyrus [46, −76, 6] (Collignon et al., 2011). K, Number of voxels when displayed at p(uncorrected) < 0.001; L, left; R, right; G, gyrus; S, sulcus.
main effect of Conditions was significant \((F_{1,15} = 37.28, p < 0.001)\), indicating that auditory motion evoked a higher response compared with static sounds. There was no significant main effect of Orientations \((F_{1,5,22,5} = 0.77, p = 0.44)\), and no interaction \((F_{1,3,45} = 2.21, p = 0.11)\). Similarly, in rhPT, only the main effect of Conditions was significant \((F_{1,15} = 37.02, p < 0.001)\). No main effect of Orientation \((F_{1,5,23,2} = 1.43, p = 0.26)\) or interaction \((F_{1,3,45} = 1.74, p = 0.19)\) was observed. Overall, brain activity in the hPT, as measured with a \(b\) parameter estimate extracted from the univariate analysis, did not provide evidence of motion direction or sound source location selectivity.

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

**Within condition classification.** Multiclass across four conditions, the classification accuracy in the hPT was significantly above chance (chance level, 25%) in both hemispheres for motion direction (lhPT: mean ± SD = 38.4 ± 7, \(p < 0.001\); rhPT: mean ± SD = 37.1 ± 6.5, \(p < 0.001\)) and sound source location (lhPT: mean ± SD = 32.4 ± 6.7, \(p < 0.001\); rhPT: mean ± SD = 31.2 ± 7.5, \(p < 0.001\); Fig. 3A). In addition, we assessed the differences between classification accuracies for motion and static stimuli by using permutation tests in lhPT \((p = 0.024)\) and rhPT \((p = 0.024)\), indicating greater accuracies for classifying motion direction than sound source location across all regions.

**Binary within-axis classification.** Binary horizontal (left vs right) within-axis classification showed significant results in both lhPT and rhPT for static sounds (lhPT: mean ± SD = 58.6 ± 14.5, \(p < 0.001\); rhPT: mean ± SD = 56.5 ± 11.9, \(p = 0.008\); Fig. 3C), while motion classification was significant only in the rhPT (mean ± SD = 55.5 ± 13.9, \(p = 0.018\); Fig. 3B). Moreover, binary vertical (up vs down) within-axis classification was significant only in the lhPT for both motion (mean ± SD = 55.7 ± 7.1, \(p = 0.01)\) and static (mean ± SD = 54.9 ± 11.9, \(p = 0.03)\) conditions (Fig. 3B, C).

**Binary across-axis classification.** We used eight additional binary classifiers to discriminate across axes-moving and static sounds. Binary across-axes (Left vs Up, Left vs Down, Right vs Up, and Right vs Down) classifications were significantly above chance level in the bilateral hPT for motion directions (Left vs Up: lhPT: mean ± SD = 65.8 ± 14.8, \(p < 0.001\); rhPT: mean ± SD = 64.8 ± 9.4, \(p < 0.001\); Left vs Down: lhPT: mean ± SD = 74 ± 15.9, \(p < 0.001\); rhPT: mean ± SD = 66.9 ± 10.4, \(p < 0.001\); Right vs Up: lhPT: mean ± SD = 72.4 ± 15.8, \(p < 0.001\); rhPT: mean ± SD = 71.4 ± 13.3, \(p < 0.001\); Right vs Down: lhPT: mean ± SD = 73.4 ± 11.8, \(p < 0.001\); rhPT: mean ± SD = 68.2 ± 15.9, \(p < 0.001\); Fig. 3B). However, sound source location across-axes classifications were significant only in certain conditions (Left vs Up rhPT: mean ± SD = 56 ± 12.7, \(p = 0.018\); Left vs Down: lhPT: mean ± SD = 55.2 ± 11.2, \(p = 0.024\); rhPT: mean ± SD = 57.3 ± 17.2, \(p = 0.003); Right vs Down lhPT: mean ± SD = 57.6 ± 8.4, \(p = 0.005)\; Fig. 3C).

"Axis of motion" preference. To test whether neural patterns within hPT contain information about opposite directions/locations within an axis, we performed four binary within-axis classifications. Similar multivariate analyses were performed to investigate the presence of information about across-axes directions/locations. The classification accuracies were plotted in Figure 3, B and 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 that were 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 up] vs [left vs right]: \(p < 0.001\); [right vs down] 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 up] vs [left vs right]: \(p = 0.02)\) and [right vs down] vs [left vs right]: \(p = 0.003)\) were significantly higher compared with the horizontal-within-axis classification accuracies. Similarly, across-axes classification accuracies were significantly higher when compared with vertical-within-axis classification accuracies in lhPT.
Fig. 3

Figure 4

terns from the motion directions (e.g., going to the left) and sound source locations rely on shared represen-
tation in hPT, we trained the classifier to distinguish neural patterns of moving and static stimuli in the
bilateral hPT. These results strengthen the notion that the axis of motion organization observed in PT does not simply stem from behavioral
performance 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). To address this issue, we replicated our classification MVPAs 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-axes and within-axes classifications was observed only 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 differences 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 vs down) was calculated. Spearman correlation between the (Δ) behavioral performance and (Δ) classification accuracies was not significant in the bilateral hPT (lhPT: \( r = 0.18, p = 0.49 \); rhPT: \( r = 0.4, p = 0.12 \)). An absence of correlation suggests that there is no relation between the difference in classification accuracies and the perceptual difference. These results strengthen the notion that the axis of motion organization observed in PT does not simply stem from behavioral performance differences.

Within orientation classification. Cross-condition classification results indicated a shared representation between motion directions and sound source locations. Previous studies have argued that successful cross-condition classification reflects an abstract representation of stimuli conditions (Hong et al., 2012; Fairhall and Caramazza, 2013; Higgins et al., 2017). To test this hypothesis, patterns of the same orientation of motion and static conditions (e.g., leftward motion and left location) were involved in within-orientation MVPAs. The rational was that if the hPT region carries a fully abstract representation of space, within-orientation classification would give results in favor of the null hypothesis (no differences within the same orientation). In the within-orientation classification analysis, accuracies from the four within-orientation classification analyses were averaged and survived FDR corrections in bilateral hPT (lhPT: \( \text{mean} \pm \text{SD} = 65.6 \pm 5, p < 0.001 \); rhPT: \( \text{mean} \pm \text{SD} = 61.9 \pm 5.6, p < 0.001 \)), indicating that the neural patterns of motion direction can be reliably differentiated from sound source location within hPT (Fig. 4A). RSA

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

Correlation with computational 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.

The first class of computational RDMs was 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, \( r \pm \text{SD} = -0.12 \pm 0.18 \); rhPT, \( r \pm \text{SD} = -0.01 \pm 0.2 \)) 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 \( r \pm \text{SD} = 0.66 \pm 0.3 \); rhPT, mean \( r \pm \text{SD} = 0.65 \pm 0.3 \)). A similar trend was observed for the second class of models that have the assumption of within-axis sounds that evoke similar neural patterns. The condition-invariant model provided the least explanation of the data (lhPT, mean \( r \pm \text{SD} = 0.14 \pm 0.25 \); rhPT, mean \( r \pm \text{SD} = 0.2 \pm 0.29 \)), 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 \( r \pm \text{SD} = 0.67 \pm 0.22 \); rhPT, mean \( r \pm \text{SD} = 0.69 \pm 0.15 \)).

In addition, we assessed differences between correlation coefficients for computational models using the Mann–Whitney–Wilcoxon rank-sum test for each class of models and hemispheres separately (Fig. 4D,E). In the Within-Axis Distinct class in lhPT, pairwise comparisons of correlation coefficients did not show a
significant difference for [Condition Distinct versus IM.9 (p = 0.8)], [Condition Distinct versus IM7 (p = 0.6); rhPT], and [Condition Distinct versus IM.5 (p = 0.09)]; however, as the models progressed further in the condition-invariant direction, the difference between correlation coefficients for models reached significance ([Condition Distinct vs IM.3], p = 0.012; [Condition Distinct vs IM.1], p = 0.007; [Condition Distinct vs Condition Invariant], p < 0.001), indicating that the Condition-
Distinct model provided a stronger correlation compared with the models representing conditions similarly. In rhPT, the rank-sum tests between each pairs revealed no significant difference for [Condition Distinct versus IM.9 ($p = 0.9$)], [Condition Distinct versus IM7 ($p = 0.7$)], [Condition Distinct versus IM5 ($p = 0.3$)], and also [Condition Distinct versus IM.3 ($p = 0.06$)]; however, as the models progressed further in the condition-invariant direction, the difference between correlation coefficients for models reached significance ([Condition Distinct vs IM.1], $p = 0.006$; [Condition Distinct vs Condition Invariant], $p < 0.001$).

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

**Discussion**

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

Does hPT contain information about specific motion directions and sound source locations? At the univariate level, our four (left, right, up, and down) motion directions and sound source locations did not evoke differential univariate activity in hPT region (Fig. 2C). We then performed multivariate pattern classification and observed that bilateral hPT contains reliable distributed information about the four auditory motion directions (Fig. 3). Our results are therefore similar to the observations made with fMRI in the human visual motion area hMT+/V5 showing reliable direction-selective multivariate information despite comparable voxelwise univariate activity levels across directions (Kamitani and Tong, 2006; but see Zimmermann et al., 2011). Within-axis MVP classification results revealed that even if both horizontal (left vs right), and vertical (up vs down) motion directions can be classified in the hPT region (Fig. 3C,D), across-axes (e.g., left vs up) direction classifications revealed higher decoding accuracies compared with within-axis classifications. Such enhanced classification accuracy across axes versus within axis is reminiscent of observations made in MT+/V5, where a large-scale axis of motion-selective organization was observed in non-human primates (Albright et al., 1984) and in area hMT+/V5 in humans (Zimmermann et al., 2011). Further examination with RSA provided additional evidence that within-axis combined models (aggregating the opposite directions/location) better explain the neural representational space of hPT by showing higher correlation values compared with within-axis distinct models (Fig. 4D,E). Our findings suggest that the topographic organization principle of hMT+/V5 and hPT shows similarities in representing motion directions. Animal studies have shown that hMT+/V5 contains motion direction-selective columns with specific directions organized side by side with their respective opposing motion direction counterparts (Albright et al., 1984; Geesaman et al., 1997; Diogo et al., 2003; Born and Bradley, 2005), an organization also probed using fMRI in humans (Zimmermann et al., 2011; but see below for alternative accounts). The observed difference for within-axis versus between-axis direction classification may potentially stem from such underlying cortical columnar organization (Kamitani and Tong, 2005; Haynes and Rees, 2006; Bartels et al., 2008). Alternatively, it has been suggested that classifying orientation preference reflects a much larger-scale (e.g., retinotopy) rather than columnar organization (Op de Beek, 2010; Freeman et al., 2011, 2013). Interestingly, high-field fMRI studies showed that the fMRI signal carries information related to both large-scale and fine-scale (columnar level) biases (Garduni et al., 2016; Pratte et al., 2016; Sengupta et al., 2017). The present study sheds important new light on the coding mechanism of motion direction within the hPT and demonstrates that the fMRI signal in the hPT contains direction-specific information and points 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, MVP revealed higher classification accuracies for moving sounds than for static sounds (Fig. 3A,B). However, despite minimal univariate activity elicited by sound source location in hPT, and the absence of reliable univariate differences in the activity elicited by each position (Fig. 2C), within-axis MVP classification results showed that sound source location information can be reliably decoded bilaterally in hPT (Fig. 3C). Our results are in line with those of previous studies showing that posterior regions in auditory cortex exhibit location sensitivity both in animals (Recanzone, 2000; Tian et al., 2001; Stecker et al., 2005) and humans (Alain et al., 2001; Zatorre et al., 2002; Warren and Griffiths, 2003; Brunetti et al., 2005; Krumbholz et al., 2005; Ahveninen et al., 2006, 2013; Deouell et al., 2007; Derey et al., 2016; Ortiz-Rios et al., 2017). In contrast to what was observed for motion direction, however, within-axis and across-axis classifications did not differ in hPT for static sounds. This indicates that sound source locations might not follow topographic representations that are similar to the one observed for auditory motion directions.

The extent to which the neural representations of motion directions and sound source locations overlap has long been debated (Grantham, 1986; Kaas et al., 1999; Romanski et al., 2000; Zatorre and Belin, 2001; Smith et al., 2004, 2007; Poirier et al., 2017). A number of neurophysiological studies have reported direction-selective neurons along the auditory pathway (Ahissar et al., 1992; Stumpf et al., 1992; Toronchuk et al., 1992; Spitzer and Semple, 1993; Doan et al., 1999; McAlpine et al., 2000). However, whether these neurons code similarly for auditory motion and sound source location remains debated (Poirier et al., 1997; McAlpine et al., 2000; Ingham et al., 2001; Oliver et al., 2003). In our study, cross-condition classification revealed that auditory motion (e.g., going to the left) and sound source location (being on the left) share partial neural representations in hPT (Fig. 3A). This observation suggests that there is a shared level of computation between sounds located on a given position and sounds.
moving toward this position. Low-level features of these two types of auditory stimuli vary in many ways and produce large differences at the univariate level in the cortex (Fig. 2B). However, perceiving, for instance, a sound going toward the left side or located on the left side evokes a sensation of location/direction in the external space that is common across conditions. Our significant cross-condition classification may therefore relate to the evoked sensation/perception of an object being going to a common external spatial location. Electrophysiological studies in animals demonstrated that motion-selective neurons in the auditory cortex display a response profile that is similar for sounds located or moving toward the same position in external space (Ahissar et al., 1992; Poirier et al., 1997; Doan et al., 1999).

Results from human psychophysiological and auditory evoked potential studies also strengthen the notion that sound source location contributes to motion perception (Strybel and Neale, 1994; Getzmann and Lewald, 2011). Our cross-condition MVP results therefore extend the notion that motion directions and sound source locations might have common features that are partially shared for spatial hearing.

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


