

# Activity in Human Auditory Cortex Represents Spatial Separation Between Concurrent Sounds

 Martha M. Shiell,  Lars Hausfeld, and  Elia Formisano

Department of Cognitive Neuroscience, Maastricht University, 6229 EV Maastricht, the Netherlands

The primary and posterior auditory cortex (AC) are known for their sensitivity to spatial information, but how this information is processed is not yet understood. AC that is sensitive to spatial manipulations is also modulated by the number of auditory streams present in a scene (Smith et al., 2010), suggesting that spatial and nonspatial cues are integrated for stream segregation. We reasoned that, if this is the case, then it is the distance between sounds rather than their absolute positions that is essential. To test this hypothesis, we measured human brain activity in response to spatially separated concurrent sounds with fMRI at 7 tesla in five men and five women. Stimuli were spatialized amplitude-modulated broadband noises recorded for each participant via in-ear microphones before scanning. Using a linear support vector machine classifier, we investigated whether sound location and/or location plus spatial separation between sounds could be decoded from the activity in Heschl's gyrus and the planum temporale. The classifier was successful only when comparing patterns associated with the conditions that had the largest difference in perceptual spatial separation. Our pattern of results suggests that the representation of spatial separation is not merely the combination of single locations, but rather is an independent feature of the auditory scene.

**Key words:** auditory cortex; auditory scene analysis; fMRI; multivariate pattern analysis; spatial cognition

## Significance Statement

Often, when we think of auditory spatial information, we think of where sounds are coming from—that is, the process of localization. However, this information can also be used in scene analysis, the process of grouping and segregating features of a soundwave into objects. Essentially, when sounds are further apart, they are more likely to be segregated into separate streams. Here, we provide evidence that activity in the human auditory cortex represents the spatial separation between sounds rather than their absolute locations, indicating that scene analysis and localization processes may be independent.

## Introduction

The primary and posterior auditory cortex (AC) are known for their sensitivity to spatial information (King and Middlebrooks, 2010), but how this information is processed is not yet understood. Unlike in the visual system, there is no known topographic map of auditory space in the mammalian cortex (Middlebrooks et al., 1998; Dery et al., 2016). Instead, in nonhuman animals, there are several auditory cortical fields that contain spatially selective neurons (Malhotra and Lomber, 2007), which tend to be

tuned broadly and contralaterally (Middlebrooks et al., 1998; Stecker et al., 2005; Woods et al., 2006). Single sound location can be decoded from neural activity based on distributed population activity (Furukawa et al., 2000; Stecker et al., 2003, 2005; Miller and Recanzone, 2009), but the details of these models are still being developed. The same principles appear to be true in the human AC, where brain imaging research has uncovered broad contralateral tuning at the level of neuronal populations (Salminen et al., 2009; Dery et al., 2016) and population-based decoding of single sound locations has been successful (Zhang et al., 2015; Dery et al., 2016; McLaughlin et al., 2016).

Neuropsychological research with lesion patients (Zatorre and Penhune, 2001) and deactivation research with transcranial magnetic stimulation (Ahveninen et al., 2013) demonstrate that activity in the posterior AC contributes to our ability to locate sounds in space consciously. However, spatial localization is not the only function supported in this region: cortex that is sensitive to spatial manipulations is also modulated by the number of auditory streams present in a scene, even when no spatial information is present (Smith et al., 2010). This nonspatial activity is

Received Nov. 22, 2017; revised March 5, 2018; accepted March 9, 2018.

Author contributions: M.M.S. and E.F. designed research; M.M.S. performed research; L.H. contributed unpublished reagents/analytic tools; M.M.S., L.H., and E.F. analyzed data; M.M.S., L.H., and E.F. wrote the paper.

This work was supported by the Netherlands Organization for Scientific Research (VENI Grant 451-17-033 to L.H. and VICI Grant 453-12-002 to E.F.). M.S. was supported by a Postdoctoral Training Fellowship from the Fonds de Recherche Santé Québec (Canada). We thank Íris Damião for assistance with preliminary analysis.

The authors declare no competing financial interests.

Correspondence should be addressed to M.M. Shiell, Department of Cognitive Neuroscience, Maastricht University, Oxfordlaan 55, 6229 EV Maastricht, the Netherlands. E-mail: marthashiell@gmail.com.

DOI:10.1523/JNEUROSCI.3323-17.2018

Copyright © 2018 the authors 0270-6474/18/384977-08\$15.00/0

thought to reflect the involvement of posterior AC in scene analysis (Smith et al., 2010); that is, the process for separating and grouping features from an auditory signal into different auditory streams (Bregman, 1994). Essentially, as the distance between sound sources grows, auditory streams from each source are perceived more clearly (Best et al., 2004; Middlebrooks and Onsan, 2012). This relationship is also reflected in fMRI of scene analysis tasks, where activation in the posterior AC increases with the spatial spread between concurrent sounds (Zatorre et al., 2002).

In the current study, we took a novel approach to investigating the representation of auditory spatial information from the perspective of how this information is used. We reasoned that, if spatial information is used for stream segregation, then it is the distance between sounds rather than their absolute positions that is essential. Therefore, we predicted that neural activity in the AC represents the separation between sounds independent of absolute location in space. To test our hypothesis, we used 7 tesla fMRI to measure human brain activity in response to spatially separated concurrent sounds. With a multivoxel pattern decoding approach, we investigated whether sound location and/or location plus separation between concurrent sounds could be decoded from the activity in Heschl's gyrus and the planum temporale. Our pattern of results suggests that the representation of spatial separation is not merely the combination of single locations, but rather is an independent feature of the auditory scene.

## Materials and Methods

### Study

The study was approved by the Ethical Review Committee for the Faculty of Psychology and Neuroscience at Maastricht University and all participants gave informed written consent. Participants were financially compensated for their time.

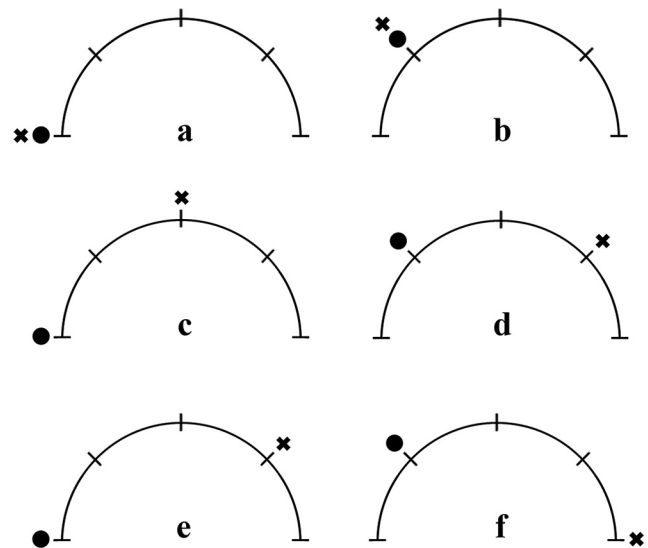
### Participants

Ten healthy participants who reported normal hearing were recruited from the Maastricht University community ( $n = 5$  males and  $n = 5$  females, mean age = 29.3 years, range = 27–34). Eight participants reported right-handedness and two reported left-handedness.

### Stimuli

Stimuli were recorded for each participant individually with binaural microphones (OKM II Classic Microphone; sampling rate = 44.1 kHz) placed in the ear canals of each participant. Our goal was to recreate a natural perception of space, so we were not concerned with isolating individual spatial cues (i.e., interaural timing or level differences) or with the effect of echoes. The recordings were done in a room of 95 m<sup>3</sup>, with walls and ceiling made of gypsum board and a wooden floor covered by a thin carpet. Participants sat in the center of an array with eight speakers spaced at intervals of 30° from negative to positive 105° (0° denotes the midline), ~2.4 m distance from the participant. Participants mobilized their head in a chin rest and were instructed not to move for the duration of the recording session, during which each stimulus was presented once. Stimuli were 3-s-long amplitude-modulated (AM) broadband pink noise generated in MATLAB (The MathWorks). Broadband noise was used to mitigate potential effects of frequency information in the neural responses (Sollini et al., 2017). The AM rate was 6, 8, or 10 Hz. This amplitude modulation was either consistent for the duration of the stimulus or interrupted for 500 ms after 1500, 1750, or 2000 ms to form the “target.” Using Vizard Virtual Reality software (Worldviz, <http://www.worldviz.com/vizard-virtual-reality-software>), stimulus location was simulated via enhanced higher-order ambisonic spatialization to appear from five different locations:  $-90^\circ$ ,  $-45^\circ$ ,  $0^\circ$ ,  $+45^\circ$ , and  $+90^\circ$ . Therefore, there were a total of 60 stimuli recorded for each participant (3 AM rates  $\times$  4 target versions  $\times$  5 locations).

After the recording session, MATLAB was used to combine individual stimulus recordings into pairs to form the experiment conditions (Fig. 1). Each pair consisted of a 6 Hz AM rate stimulus, herein referred to as



**Figure 1.** Experimental conditions. *a–f*, Participants listened to 6 different pairs of concurrent sounds. Pairs varied in the location of sound 1 (either  $90^\circ$  or  $45^\circ$  from the midline, shown by the position of the black circle in the left and right columns, respectively) and in the spatial separation between sound 1 and sound 2 ( $0^\circ$ ,  $90^\circ$ , or  $135^\circ$  contralateral to sound 1, shown by the position of the black cross in the top, middle, and bottom rows, respectively). Half of the participants were tested with sound 1 in the left hemisphere, as shown, and half in the right.

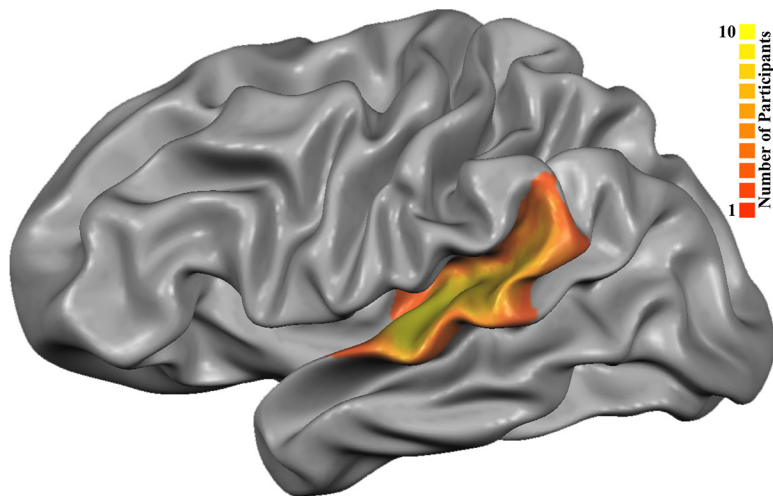
sound 1, presented with either an 8 or 10 Hz AM rate stimulus, herein referred to as sound 2. Sound 1 occurred at  $-90^\circ$ ,  $-45^\circ$ ,  $+45^\circ$ , or  $+90^\circ$  and sound 2 at  $0^\circ$ ,  $90^\circ$ , or  $135^\circ$  separation from sound 1 in the direction of the opposite hemifield. The spatial locations, separations, and AM rates of these stimuli were determined based on pilot testing showing that the sound locations could be discriminated easily and that the sounds were segregated easily when separated.

### Prescanning stimulus validation

On a separate day after the recording session, participants completed a behavioral testing session to ensure that the spatial properties of the stimuli were registered accurately in the recordings and that the participants could segregate the pairs of sounds based on spatial cues and perform the task that was required inside the scanner. The stimuli were presented via Psychopy (Peirce, 2007; <http://www.psychopy.org>) through the same model of earphones that were used inside the MRI scanner (model S14; Sensimetrics). The experimenter familiarized the participant with the spatialized stimuli and verified that the participant could distinguish the five locations. The participant completed four blocks of the behavioral task. The left and right hemifields were tested in separate alternating blocks. Each block consisted of 96 trials equally split across 2 AM rates for sound 2 and the 6 experiment conditions (2 locations of sound 1  $\times$  3 separations from sound 2). Trials were pseudorandomly ordered. In each trial, the target (the disruption in the AM rate, described above) occurred in either sound 1 or sound 2 at the 1500, 1750, or 2000 ms onset with equal probability. The participant's task was to indicate “yes” or “no” by button press if the target occurred in sound 1 and to ignore the presence or absence of the target in sound 2.

### MRI

On a separate day after the behavioral training session, participants underwent an MRI session on the Siemens 7 tesla MAGNETOM MRI scanner with a 32-channel Nova Medical head RF coil at the Scannexus facility in Maastricht, Netherlands ([www.scannexus.nl](http://www.scannexus.nl)). Participants completed 10 runs of fMRI to measure BOLD signal [T2\*-weighted gradient echo-planar imaging, volumes = 57, number of slices = 60, voxel size = 1.1 mm isotropic, matrix size =  $188 \times 188$ , TR = 7300 ms, TA = 1800 ms, silent gap = 5500 ms, generalized autocalibrating partially parallel acquisitions (GRAPPA) = 3], followed by two anatomical scans (0.7 mm isotropic voxels, matrix size =  $320 \times 320$ , number of slices = 256, flip angle =  $5^\circ$ , GRAPPA = 3): a T1-weighted image (MPRAGE se-



**Figure 2.** Region-of-interest (ROI). Shown is the group-aligned (Goebel et al., 2006) overlap of individual participant's ROIs ( $n = 10$ ) in the left hemisphere of a representative participant.

quence, TR = 3100 ms, time to inversion = 1500 ms, TE = 3.5 ms), and a proton density (PD) image (TR = 2160 ms, TE = 3.5 ms).

During the fMRI, auditory stimulus presentation was controlled by Psychopy (Peirce, 2007). Participants viewed a fixation cross presented via a projector and mirror. Each fMRI run contained 24 experiment trials and three catch trials. Both experiment and catch trials were preceded by a rest period (no stimulus or response) and catch trials were followed by a response period (explained below), making a total of 57 TRs per run. For the experiment/catch trials, the auditory stimulus was presented via MRI-compatible ear buds (S14; Sensimetrics) at a comfortable level during the 5500 ms silent gap between measurements. The order of the stimuli in experiment/catch trials was pseudorandom. The 24 experiment trials included 4 repetitions of each of the 6 conditions (2 locations for sound 1  $\times$  3 separations from sound 2) equally split across the 2 AM rates for sound 2 and with no target. Because we were not interested in hemispheric differences and our stimuli elicited no measurable differences in segregability between hemifields (see Results), half of the participants were tested with sound 1 in the left hemifield and half in the right hemifield. Catch trials always contained a target in either the attended or distractor sound and included an equal number from each condition over the 10 runs. During the response period after each catch trial, the fixation cross changed from black to red for 3000 ms. Participants were instructed to complete the same task as in the behavioral training session, but to only make a response when the fixation cross turned red. This random sampling of behavior allowed us to ensure that participants attended the stimuli and to avoid a potential confound by motor responses.

#### MRI preprocessing

MRI preprocessing was completed using automatic tools from Brain Voyager QX (Brain Innovation). T1-weighted images were divided by the PD images to minimize signal inhomogeneities from the receiver coil. The T1/PD was corrected for residual inhomogeneities, resampled to 1.0 mm isotropic resolution, and aligned to the AC–PC plane. Gray matter, white matter, and CSF were segmented automatically and the borders were edited manually as needed in the region of the primary auditory and posterior superior temporal cortex.

Preprocessing of the fMRI data consisted of slice scan-time correction (with sinc interpolation), 3D motion correction (trilinear/sinc interpolation to the first volume of the first run), and temporal high-pass filtering (five cycles per run with linear trend removal). Functional data were resampled to 1.0 mm isotropic space (sinc interpolation) and automatically registered to the participant's preprocessed anatomical image with manual corrections as needed (rigid-body transformation, six degrees of freedom).

#### Region-of-interest (ROI)

For each participant, each hemisphere of the preprocessed anatomical volume in native space was transformed into an inflated surface and an

ROI was drawn manually (Fig. 2). The ROI was defined liberally to include Heschl's gyrus, the planum temporale, and the surrounding cortex rather than approximating the border between these regions. This allowed us to avoid excluding potentially relevant information. Primary and posterior auditory regions were chosen based on evidence of their spatial sensitivity (Middlebrooks and Bremen, 2013; Dery et al., 2016). The ROI began at the medial end of the first transversal sulcus rostral to Heschl's gyrus and traveled caudally along the insular cortex following the angle of the sylvian fissure and its ascending limb. At the most dorsal point of the ascending limb of the sylvian fissure, the ROI sliced horizontally to the midpoint of the ascending superior temporal gyrus and followed the line of this gyrus' lateral edge ventrally and rostrally. The rostral border of the ROI was delineated by a vertical slice at the point where Heschl's sulcus meets the superior temporal gyrus.

The ROI was transformed into volume space and the resultant voxels were used to extract fMRI signal for the multivoxel pattern analysis (MVPA).

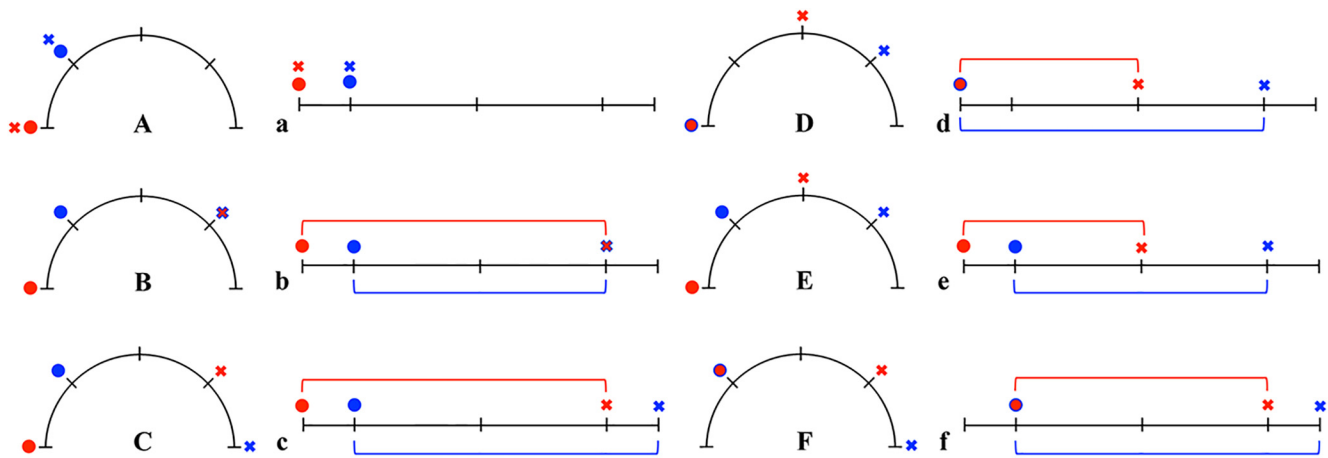
#### MVPA

Our hypothesis was that activity in the AC represents the separation between concurrent sounds. To test this hypothesis and to interpret our results, we developed six different models, which are explained further below. Each model consisted of two conditions and a classifier was trained to distinguish the multivoxel patterns elicited by each condition (De Martino et al., 2008; Formisano et al., 2008). We predicted that the performance of the classifier would be modulated by how strongly the conditions in each model differed in separation, with greater separation associated with higher decoding accuracy. This prediction was based on evidence that similarity in perceptual information correlates with similarity of elicited multivoxel activity patterns (O'Toole et al., 2005).

Each hemisphere was analyzed separately, following from evidence of successful unilateral decoding of spatial stream segregation in the cat auditory cortex (Middlebrooks and Bremen, 2013). Note that a control analysis with bilateral decoding did not improve classifier accuracy for any of the models tested. Analysis was completed with custom MATLAB scripts. In each voxel within the ROI, we calculated the percentage signal change for each experimental trial relative to the preceding rest trial. Data from catch trials and response trials were excluded. The participant's 10 runs were split into training and testing data using a leave-run-out scheme, leading to 10 combinations of 36 training (9 runs) and 4 testing trials (1 run) for each model. For each training set, we excluded voxels with outliers (absolute  $Z$ -transformed percentage signal change  $> 5.0$ ) under the assumption that such values are likely driven by noise. To reduce the dimensionality of the data, we selected the half of the voxels with the highest absolute percentage signal change. These exclusion thresholds were chosen *a priori*, but *post hoc* replications of the analysis with different thresholds (excluding  $Z$ -transformed percentage signal change  $> 4.0$  or  $6.0$ ; selecting either 25% or 75% of the voxels with the highest absolute percentage signal change) produced the same pattern of significant results as reported below.

For each model, the multivoxel patterns of percentage signal change were analyzed by a linear support vector machine (SVM) (Formisano et al., 2008). Voxels were selected iteratively using a recursive feature elimination (RFE) procedure (De Martino et al., 2008) consisting of 15 levels. In each level of the RFE, the SVM training and testing was repeated four times with a random sampling of 90% of the trials. Each voxel was labeled with a weight that represented the contribution of that voxel to the classification's success, averaged across the four repetitions, and the lowest weighted voxels were discarded. The number of discarded voxels at each level was adjusted for each hemisphere's ROI such that  $\sim 250$  voxels remained at the 15th level of RFE. Decoding accuracy at each level was





**Figure 3.** Experiment models. *A–F*, Each model consisted of two conditions shown separately in red and blue. A blue outline indicates a sound location common to both conditions. Circles and crosses represent locations for sounds 1 and 2, respectively, with locations at  $-90^\circ$ ,  $-45^\circ$ ,  $0^\circ$ ,  $+45^\circ$ , and  $+90^\circ$  (black markers from left to right on each semicircle/horizontal line). Lowercase letters (*a–f*) illustrate models in a diagram (unscaled) representation of perceptual space, where acuity is amplified around the midline. Red and blue lines indicate perceptual separation between concurrent sounds in each condition.

calculated as the average proportion of correctly classified testing trials across the 10 splits and two classes. The maximum accuracy across RFE levels was selected for each hemisphere and each model.

Although the theoretical chance level for decoding accuracy is 50%, the final step of the RFE procedure inflates this value. Therefore, we calculated chance level empirically for each of the 20 tested hemispheres individually by permuting the condition labels for each model and repeating the full RFE procedure 600 times (100 times per model). Across models and 20 hemispheres, the median empirical chance level was 53.9%, with an interquartile range of 0.004%. To account for this variability, for each hemisphere, we subtracted its empirical chance level from each model's classification accuracy. This gave the measure of decoding accuracy above chance (DAC). Because half of the participants were tested with sound 1 in the left hemifield and half in the right, results for the left and right hemispheres were classified as either ipsilateral or contralateral according to their relation to sound 1.

#### Model design

Our models (Fig. 3*A–F*) were designed *a priori* to test our main hypothesis and to control for confounding interpretations. Model A consisted of concurrent sounds with no separation, which, based on behavioral performance in the prescanning stimulus validation (see Results), participants were unable to segregate into separate streams. The inclusion of model A allowed us to verify the sensitivity of the MVPA procedure. In case our subsequent decoding of concurrent sounds was unsuccessful, the success or failure of decoding model A was intended to help us to determine whether our failure with concurrent sounds was derived from a lack of sensitivity. Models B–E were included to test our main prediction, that the performance of the classifier will be modulated by the model's change in separation. Note that, in our design, it was not possible to isolate separation from location: a change in separation necessarily required a change in location. Therefore, our predictions rested on the assumption that, if separation was an important feature, then it would add information beyond mere location and thereby improve DAC. Our models were designed in terms of absolute auditory space, where  $45^\circ$  separation in the periphery is considered equal to the same separation at the midline. From this perspective, models B, D, and F include changes in separation between the two decoded conditions, whereas models C and E maintain a constant separation (Fig. 3*B–F*). Based on our prediction, we expected that model B would have higher DAC compared with model C and, likewise, model D compared with model E. Together, these two comparisons allowed us to control for the confound interpretation that differences in DAC may be attributed to exclusion of either the  $90^\circ$  or  $135^\circ$  separation conditions (which were excluded from models C and E, respectively). Because participants were instructed to attend sound 1 and the location of sound 1 does not

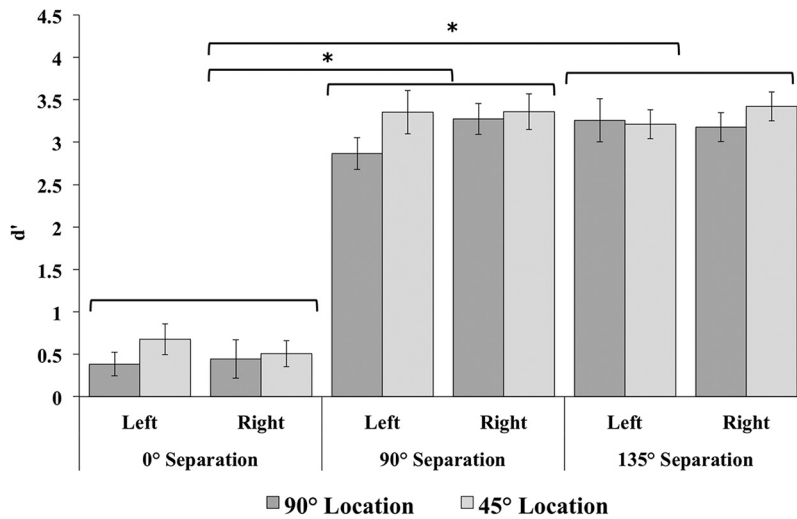
change in model D, model F was included for comparison with model B as a control for the possibility of attentional effects in the comparison of models D and E. Because only model D was successfully decoded above chance (see Results), only the planned comparison between models D and E was completed (see “Experimental design and statistical analysis” section).

Although we chose to design our models in terms of absolute auditory space, the perception of auditory space is not equal along the horizontal azimuth. Specifically, spatial acuity for broad and narrow band-pass-filtered noises tends to decrease from the midline to the periphery (Makous and Middlebrooks, 1990; Perrott et al., 1993; Best et al., 2004), meaning that a change in location from  $45^\circ$  to  $90^\circ$  is perceptually smaller than a change from the midline to  $45^\circ$ . We chose not to prioritize this perceptual perspective in our design because, without already completing the experiment, it was unclear whether it was critical for the representation of space in the auditory cortex and properly controlling for it would have required a full behavioral experiment with the psychoacoustical conditions of the scanner and our specific stimuli. It should be noted that even if differences in spatial acuity across the horizontal azimuth are taken into account for our models (Fig. 3*a–f*), the comparisons outlined above still test our hypothesis. From this perceptual perspective, model E has a change in separation (Fig. 3*e*), but this change in separation remains smaller than that which occurs in model D, which has the largest change in separation of all models (see difference between blue and red line lengths, Fig. 3*a–f*).

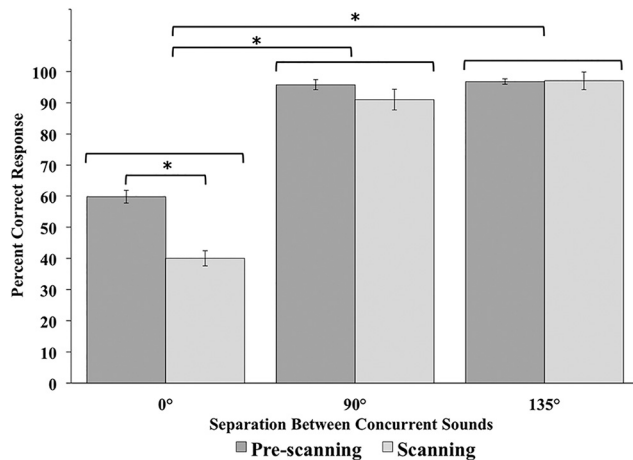
#### Experimental design and statistical analysis

**Prescanning stimulus validation.** For each condition in the prescanning stimulus validation session,  $d'$  was calculated as the Z-transformed rates for hits minus false alarms of the target in sound 1. To validate the effectiveness of our stimuli, we performed a  $2 \times 2 \times 3$  repeated-measures ANOVA ( $n = 10$ ) with the within-subject factors location ( $45^\circ$  or  $90^\circ$ ), hemifield (left or right), and separation ( $0^\circ$ ,  $90^\circ$ , or  $135^\circ$ ), and two-tailed *post hoc t* test comparisons with Bonferroni correction (number of comparisons = 3).

**Behavior inside the scanner.** To assess participant attention inside the scanner, we calculated the rate of responses and the percentage correct in the response trials. To ensure that stimuli elicited effective stream segregation inside the scanner, we compared percentage correct responses for catch trials inside the scanner with the prescanning stimulus validation in a  $3 \times 2$  repeated-measures ANOVA ( $n = 10$ ) with the within-subject factors separation ( $0^\circ$ ,  $90^\circ$ , or  $135^\circ$ ) and testing environment (prescanning or scanning). We completed two-tailed *post hoc t* test comparisons with Bonferroni correction (number of comparisons = 6) to interpret the ANOVA results.



**Figure 4.** Prescanning stimulus validation. Shown is the group average target detection during prescan stimulus validation with SE bars ( $n = 10$ ).  $*p < 0.0001$ .



**Figure 5.** Behavior inside the scanner. Shown is the group average performance on the behavioral task with SE bars ( $n = 10$ ).  $*p < 0.001$ .

### DAC

Decoding results for each model in ipsilateral and contralateral groups of hemispheres were tested for significant DAC with a one-sample ( $n = 10$ ) Wilcoxon signed-rank test corrected by false discovery rate (Benjamin and Hochberg, 1995) (number of comparisons = 12) with  $q < 0.05$ .

### A priori models comparison

We predicted *a priori* that DAC would be higher in model D than model E, in model B than C, and B than F. Because only model D showed DAC significantly above chance (see Results section), only the prediction for models D and E was completed. This prediction was tested at the group hemisphere level ( $n = 10$ ) with one-tailed paired-sample Wilcoxon signed-rank tests corrected by false discovery rate (Benjamin and Hochberg, 1995) (number of comparisons = 2) with  $q < 0.05$ .

### Post hoc comparison of Euclidean distance

Model D was the only model that did not include a change from 90° to 45°, which may have contributed to its success. Potentially, the representations of the 90° and 45° locations may have been too noisy to be discriminated. In this case, model D succeeded over model E due to fewer instances of 90° and 45° locations (two vs three), leading to more consistent activity patterns overall. To test this possibility, we selected voxels in each hemisphere that were included in the most accurate RFE level of both models D and E, and common to at least five splits in each. From

these voxels, we calculated the average pairwise Euclidean distance of multivoxel patterns across trials separately in two conditions: sound 1 at  $-90^\circ$  and sound 2 at  $+45^\circ$ , versus sound 1 at  $-45^\circ$  and sound 2 at  $+45^\circ$  (Fig. 1*d,e*), which were chosen because they differed between models D and E (blue/outlined shapes in Fig. 3*d,e*). Averages were compared separately in ipsilateral and contralateral groups of hemispheres with a two-tailed two-sample Wilcoxon rank-sum test ( $n = 10$ ) with Bonferroni correction (number of comparisons = 2).

## Results

### Prescanning stimulus validation

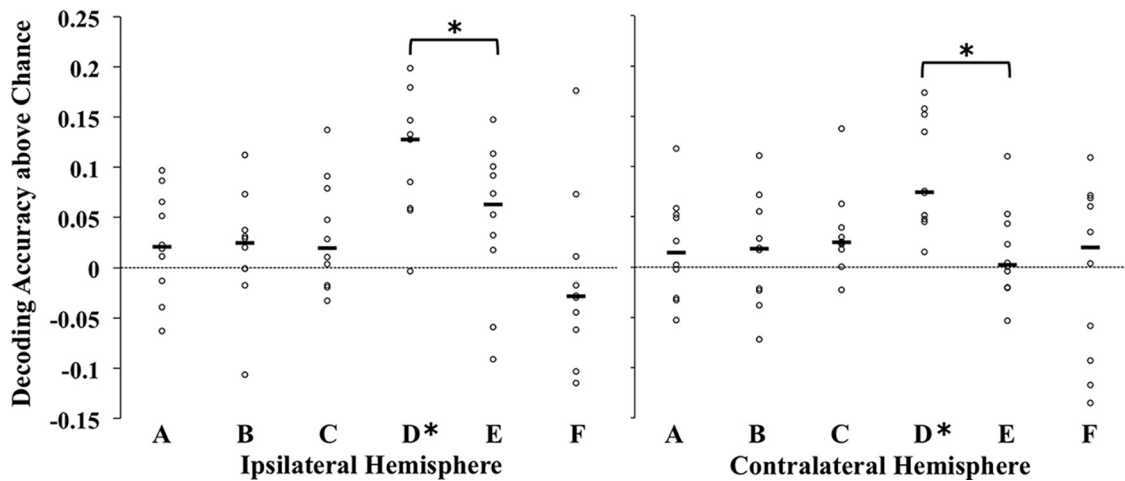
To validate our stimuli, we compared participant performance in a two-alternative forced-choice target detection task with a 2 (location)  $\times$  2 (hemifield)  $\times$  3 (separation) repeated-measures ANOVA (Fig. 4). We found a main effect of separation ( $n = 10$ ,  $F_{(1,25,11,28)} = 153.4$ ,  $p < 0.001$ , Greenhouse–Geisser corrected). Two-tailed *post hoc t* test comparisons ( $n = 10$ ) with Bonferroni correction

(number of comparisons = 3) indicated that participants performed worse for 0° separation than 90° ( $t_{(9)} = -11.6$ ,  $p < 0.0001$ ) or 135° separation ( $t_{(9)} = -15.2$ ,  $p < 0.0001$ ), confirming that the sounds were better segregated when separated. There was no difference between separations of 90° and 135° ( $t_{(9)} = 0.54$ ,  $p = 0.6$ ) or main effects of the locations of sound 1 (left vs right hemifield:  $F_{(1,9)} = 1.53$ ,  $p = 0.25$ ; 90° vs 45° locations:  $F_{(1,9)} = 4.39$ ,  $p = 0.067$ ), which mitigates the potential effects of differences in cognitive load between the different separation (excluding 0° separation) and location conditions.

### Behavior inside the scanner

All participants responded to all 30 catch trials except for two participants who responded to 28/30 catch trials (each with missed responses in two different runs) and one participant who responded to 29/30 catch trials. This nearly perfect response rate indicates that participants mostly attended to the stimuli during the fMRI scanning session, but that there may have been lapses of attention. Such lapses are not surprising given that stimuli were presented only intermittently approximately every 14 s and that fMRI scanning lasted  $\sim 90$  min. However, we do not anticipate that the participants' state of attention affected our results based on evidence showing that spatial stream segregation is present in AC responses even under conditions of anesthesia (Middlebrooks and Bremen, 2013).

To validate the effectiveness of our stimuli inside the scanner, we compared percentage correct responses for catch trials inside the scanner with the prescan stimulus validation in a 3 (separation)  $\times$  2 (testing environment) repeated-measures ANOVA (Fig. 5). We found a main effect for separation ( $F_{(2,18)} = 344.3$ ,  $p < 0.0001$ ) and testing environment ( $F_{(1,9)} = 12.0$ ,  $p = 0.007$ ) and an interaction between the two ( $F_{(2,18)} = 11.3$ ,  $p = 0.0006$ ). Two-tailed *post hoc t* tests with Bonferroni correction (number of comparisons = 6) showed that, across testing environments, performance was worse for 0° separation than 90° ( $t_{(9)} = 24.4$ ,  $p < 0.0001$ ) and 135° ( $t_{(9)} = 21.4$ ,  $p < 0.0001$ ), but not different between the 90° and 135° separations ( $t_{(9)} = 1.76$ ,  $p = 0.67$ ). Both the main effect of testing environment and the interaction between testing environment and separation appeared to have been driven by decreased performance in the 0° separation condition



**Figure 6.** DAC. Shown are individual (circles) and group median (bars,  $n = 10$ ) DAC in each hemisphere for six models (Fig. 3A–F). Chance level, indicated by the dashed line, was calculated separately for each participant and hemisphere and was combined across models. Hemispheres were coded ipsilaterally or contralaterally relative to the location of sound 1. \* $q < 0.05$ .

in the scanner compared with prescanning ( $t_{(9)} = 6.2, p < 0.001$ ). Importantly, there was no difference in behavior between prescanning and scanning behavior for the  $90^\circ$  ( $t_{(9)} = 1.2, n.s.$ ) and  $135^\circ$  ( $t_{(9)} = 0.06, n.s.$ ) separation conditions, confirming that participants segregated sounds 1 and 2 inside the scanner.

**DAC**

DAC for six models in each hemisphere is shown in Figure 6. Only model D produced significant DAC, which occurred in both hemispheres (ipsilateral:  $W = 54, q = 0.024$ ; contralateral:  $W = 55, q = 0.024$ ).

**A priori model comparison**

Consistent with our prediction, DAC for model D was higher than that for model E in both hemispheres (ipsilateral:  $W = 49, q = 0.016$ ; contralateral:  $W = 52, q = 0.008$ ).

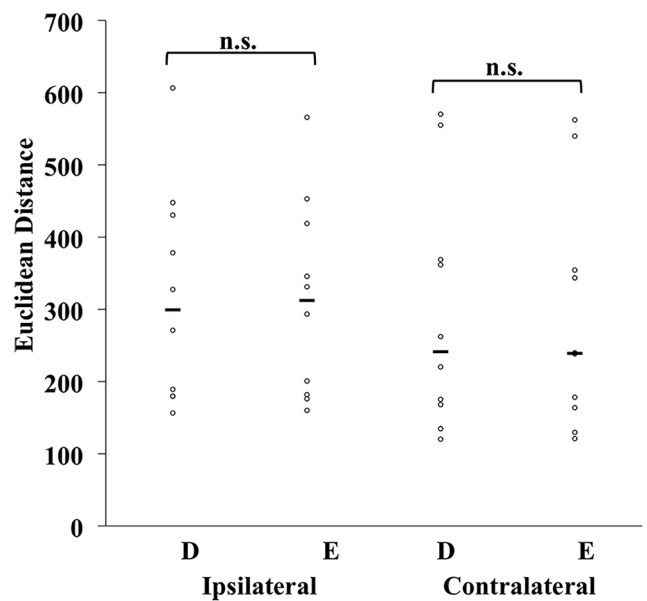
**Post hoc comparison of Euclidean distance**

To partially control for the possibility that model D’s success was due to its exclusion of a change in location from  $90^\circ$  to  $45^\circ$ , we compared average pairwise Euclidean distance across trial repetitions in the two conditions that were different in these models (Fig. 1*d,e*) with a two-tailed two-sample Wilcoxon rank-sum test with Bonferroni correction (number of comparisons = 2). There were no differences between conditions (Fig. 7) in either the ipsilateral ( $W = 31, p = 0.76$ ) or contralateral ( $W = 5, n.s.$ ) hemispheres, confirming that these models did not differ in the consistency of their patterns over trial repetitions.

**Discussion**

We measured human brain activity in response to spatially separated concurrent sounds with fMRI at 7 tesla. Using a MVPA approach, we investigated whether sound location and/or location plus separation between sounds could be decoded from the activity in Heschl’s gyrus and the planum temporale. As discussed further below, our pattern of results supports our hypothesis that activity in the AC represents the spatial separation between concurrent sounds, as per the role of spatial information in scene analysis, but only when perceptual spatial acuity across the horizontal azimuth is taken into account. To our knowledge, this work marks the first direct evidence for this hypothesis.

Models B, D, and F each included a change in absolute separation (Fig. 3), but the classifier was only successful at decoding



**Figure 7.** Euclidean distance across trials. Shown are individual (circles) and group median (horizontal bars) average pairwise Euclidean distance across trial repetitions for the two conditions shown in Figure 1, *d* and *e* ( $n = 10$ ).

the brain activity associated with the two conditions in model D. Model D differed from the other two models in three ways: (1) it included a location change from the midline to  $45^\circ$ , (2) it excluded a location change from  $90^\circ$  to  $45^\circ$ , and (3) it contained the largest change in perceptual separation (see difference between red and blue line lengths, Fig. 3*b–f*). With respect to the first listed feature, model E similarly included a location change from the midline to  $45^\circ$ . If this feature were responsible for model D’s success, then model E should have also produced significant DAC, which was not the case. The second listed feature may have contributed to model D’s success if the neural representations of the  $90^\circ$  and  $45^\circ$  locations were either not sufficiently different from one another or were too noisy to be decoded in models B and F. In the former possibility, model E would have been equivalent to model D and thus similarly discriminable. To the contrary, DAC for model D was higher than that for model E. In the latter possibility, model E would have had less consistent multi-

voxel patterns across trials than model D. Our comparison of Euclidean distance (Fig. 7; see Materials and Methods and Results) indicates that this was not the case. Therefore, by process of elimination, we reason that model D's success is best explained by the third listed feature: its inclusion of the largest change in perceptual separation of all tested models. Given that multivoxel pattern discriminability correlates with perceptual discriminability (O'Toole et al., 2005), we propose that the change in perceptual separation in model D was sufficient to elicit distinct differences in multivoxel patterns, whereas the smaller perceptual changes in models B and F were not.

Our interpretation implies that the representation of space in the AC mirrors the relationship between auditory spatial acuity and azimuth, where acuity decreases with eccentricity. With broad and narrow band-pass-filtered noise, this relationship has been demonstrated with measures of localization accuracy (Makous and Middlebrooks, 1990), the minimum audible angle between two consecutive sounds (Perrott et al., 1993), and the minimum audible angle between two concurrent sounds (Best et al., 2004). Because the rate of change in acuity differs depending on the experimental paradigm, it is not possible to infer from these previous works precisely how acuity can be equated across the azimuth. In the study most similar to the paradigm of our experiment, Best et al. (2004) asked participants to indicate whether they heard one or two distinct sound sources in response to concurrent spatially separated broadband noises. For a sound located at 90°, 45° of separation was required for perfect perception of two distinct sources, whereas for a sound located at the midline, only 12° of separation was required. This indicates that, for every degree of separation from the midline, 3.75° are required for the equivalent perception of separation from 90°. This is approximately consistent with our diagram representation of perceptual space in Figure 3, *a–f*.

In our experiment, participants were instructed to attend to one of the concurrent streams, which was cued both by the AM rate and as being at the more eccentric location. Therefore, models D and E differed in the location of attention, which remained stationary at 90° in model D and moved from 45° to 90° locations in model E. Our intuition was that a change in the location of the attended sound in model E should have made the associated brain activity patterns more discriminable, thereby decreasing the difference between models D and E. Due to the failure of the classifier for models B and F, we were unable to test this intuition directly, but the fact that the DAC for model D was higher than that for model E suggests that this potential attentional effect was not sufficiently strong enough to overcome the difference in separation change between the two models. In our current design, we cannot control for the less intuitive possibility, that decoding was biased to the location change of the unattended sound, but we do not believe that such an explanation could explain our results completely given that spatial stream segregation is present in AC responses even under conditions of anesthesia (Middlebrooks and Bremen, 2013).

The comparison between models D and E is particularly interesting when considered with the null result for model A. We were unable to decode activity patterns associated with 90° and 45° locations when the concurrent sounds were colocated in model A (Fig. 3A), which is consistent with the decreased spatial acuity for the locations of noises in the periphery (Makous and Middlebrooks, 1990; Perrott et al., 1993). Despite the fact that these activity patterns were not discriminable in model A, the location change from 90° to 45° was sufficient to offset the change in separation in model E such that DAC was lower for model E

than D. This discrepancy may indicate that separation is not merely the combination of single sound locations, but rather is an independent quality of the auditory scene. Note that this interpretation does not imply that location is not processed in the auditory cortex, but rather suggests that location processing occurs independently of separation.

The proposed dissociation between location and separation processing is consistent with evidence that spatial processing for stream segregation invokes different neural mechanisms than that for localization. Specifically, Duffour-Nikolov et al. (2012) tested neuropsychological lesion patients on both auditory spatial localization and the ability to benefit from spatial separation between a target and masker sounds, known as spatial release from masking. Five patients showed a deficit in localization but preserved spatial release from masking and one showed the opposite pattern, indicating that the two processes are affected independently by brain damage (Duffour-Nikolov et al., 2012). Support for this dissociation also comes from research on thresholds for minimum audible angles and spatial stream segregation in healthy adults. Here, the two measures varied unsystematically across conditions with different types of stimuli, implying that they relied on different mechanisms (Middlebrooks and Onsan, 2012).

Previous research on auditory spatial processing in the cortex has demonstrated that the location of a single sound can be decoded by population-based tuning (Furukawa et al., 2000; Stecker et al., 2003, 2005; Miller and Recanzone, 2009; Zhang et al., 2015; Derey et al., 2016; McLaughlin et al., 2016), but it remains unclear how these representations translate into representations of scenes with sounds from multiple locations (Middlebrooks and Bremen, 2013; Day and Delgutte, 2013). Our proposal that auditory cortex activity represents separation between concurrent sounds implies that modeling single locations will be insufficient to fully understand auditory spatial processing. Research on spatial stream segregation in the primary auditory cortex of cats is consistent with this idea (Middlebrooks and Bremen, 2013). Here, neuronal activity was modulated by the degree of separation between concurrent auditory streams and spatial tuning changed depending on the spatial configuration of the sounds (Middlebrooks and Bremen, 2013).

In closing, our results emphasize the importance of considering concurrent locations in the study of auditory spatial processing. We believe this approach is relevant not only for understanding auditory processing, but also for understanding spatial cognition more broadly. Our interpretation supports that spatial cognition may be customized to the particular goals of a system. If the auditory system aims to analyze scenes, then the most relevant feature for this goal is the separation between concurrent sounds. Continued research on this relationship between spatial processing and scene analysis may be clinically relevant for rehabilitation after hearing loss (Francart et al., 2014) and in aging (Gallun et al., 2013).

## References

- Ahveninen J, Huang S, Nummenmaa A, Belliveau JW, Hung AY, Jääskeläinen IP, Rauschecker JP, Rossi S, Tiitinen H, Raij T (2013) Evidence for distinct human auditory cortex regions for sound location versus identity processing. *Nat Commun* 4:2585. [CrossRef Medline](#)
- Benjamin Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Methodol* 57:289–300.
- Best V, van Schaik A, Carlile S (2004) Separation of concurrent broadband sound sources by human listeners. *J Acoust Soc Am* 115:324–336. [CrossRef Medline](#)
- Bregman AS (1994) *Auditory scene analysis*. Cambridge, MA: MIT.



- Day ML, Delgutte B (2013) Decoding sound source location and separation using neural population activity patterns. *J Neurosci* 33:15837–15847. [CrossRef Medline](#)
- De Martino F, Valente G, Staeren N, Ashburner J, Goebel R, Formisano E (2008) Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage* 43:44–58. [CrossRef Medline](#)
- Derey K, Valente G, de Gelder B, Formisano E (2016) Opponent coding of sound location (azimuth) in planum temporale is robust to sound-level variations. *Cereb Cortex* 26:450–464. [CrossRef Medline](#)
- Duffour-Nikolov C, Tardif E, Maeder P, Thiran AB, Bloch J, Frischknecht R, Clarke S (2012) Auditory spatial deficits following hemispheric lesions: dissociation of explicit and implicit processing. *Neuropsychol Rehabil* 22:674–696. [CrossRef Medline](#)
- Formisano E, De Martino F, Bonte M, Goebel R (2008) “Who” is saying “what?” Brain-based decoding of human voice and speech. *Science* 322:970–973. [CrossRef Medline](#)
- Francart T, Lenssen A, Wouters J (2014) Modulation enhancement in the electrical signal improves perception of interaural time differences with bimodal stimulation. *J Assoc Res Otolaryngol* 15:633–647. [CrossRef Medline](#)
- Furukawa S, Xu L, Middlebrooks JC (2000) Coding of sound-source location by ensembles of cortical neurons. *J Neurosci* 20:1216–1228. [CrossRef Medline](#)
- Gallun FJ, Diedesch AC, Kampel SD, Jakien KM (2013) Independent impacts of age and hearing loss on spatial release in a complex auditory environment. *Front Neurosci* 7:252. [CrossRef Medline](#)
- Goebel R, Esposito F, Formisano E (2006) Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum Brain Mapp* 27:392–401. [CrossRef Medline](#)
- King AJ, Middlebrooks JC (2010) Cortical representation of auditory space. In: *The auditory cortex* (Winer JA, Schreiner CE, eds), pp 329–341. Boston: Springer.
- Makous JC, Middlebrooks JC (1990) Two-dimensional sound localization by human listeners. *J Acoust Soc Am* 87:2188–2200. [CrossRef Medline](#)
- Malhotra S, Lomber SG (2007) Sound localization during homotopic and heterotopic bilateral cooling deactivation of primary and nonprimary auditory cortical areas in the cat. *J Neurophysiol* 97:26–43. [CrossRef Medline](#)
- McLaughlin SA, Higgins NC, Stecker GC (2016) Tuning to binaural cues in human auditory cortex. *J Assoc Res Otolaryngol* 17:37–53. [CrossRef Medline](#)
- Middlebrooks JC, Bremen P (2013) Spatial stream segregation by auditory cortical neurons. *J Neurosci* 33:10986–11001. [CrossRef Medline](#)
- Middlebrooks JC, Onsan ZA (2012) Stream segregation with high spatial acuity. *J Acoust Soc Am* 132:3896–3911. [CrossRef Medline](#)
- Middlebrooks JC, Xu L, Eddins AC, Green DM (1998) Codes for sound-source location in nontopographic auditory cortex. *J Neurophysiol* 80:863–881. [CrossRef Medline](#)
- Miller LM, Recanzone GH (2009) Populations of auditory cortical neurons can accurately encode acoustic space across stimulus intensity. *Proc Natl Acad Sci U S A* 106:5931–5935. [CrossRef Medline](#)
- O’Toole AJ, Jiang F, Abdi H, Haxby JV (2005) Partially distributed representations of objects and faces in ventral temporal cortex. *J Cogn Neurosci* 17:580–590. [CrossRef Medline](#)
- Peirce JW (2007) PsychoPy: psychophysics software in Python. *J Neurosci Methods* 162:8–13. [CrossRef Medline](#)
- Perrott DR, Costantino B, Cisneros J (1993) Auditory and visual localization performance in a sequential discrimination task. *J Acoust Soc Am* 93:2134–2138. [CrossRef Medline](#)
- Salminen NH, May PJ, Alku P, Tiitinen H (2009) A population rate code of auditory space in the human cortex. *PLoS One* 4:e7600. [CrossRef Medline](#)
- Smith KR, Hsieh IH, Saberi K, Hickok G (2010) Auditory spatial and object processing in the human planum temporale: no evidence for selectivity. *J Cogn Neurosci* 22:632–639. [CrossRef Medline](#)
- Sollini J, Mill R, Sumner CJ (2017) Spatial processing is frequency-specific in auditory cortex but not in the midbrain. *J Neurosci* 37:6588–6599. [CrossRef Medline](#)
- Stecker GC, Mickey BJ, Macpherson EA, Middlebrooks JC (2003) Spatial sensitivity in field PAF of cat auditory cortex. *J Neurophysiol* 89:2889–2903. [CrossRef Medline](#)
- Stecker GC, Harrington IA, Middlebrooks JC (2005) Location coding by opponent neural populations in the auditory cortex. *PLoS Biol* 3:e78. [CrossRef Medline](#)
- Woods TM, Lopez SE, Long JH, Rahman JE, Recanzone GH (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. [CrossRef Medline](#)
- Zatorre RJ, Penhune VB (2001) Spatial localization after excision of human auditory cortex. *J Neurosci* 21:6321–6328. [CrossRef Medline](#)
- Zatorre RJ, Bouffard M, Ahad P, Belin P (2002) Where is “where” in the human auditory cortex? *Nat Neurosci* 5:905–909. [CrossRef Medline](#)
- Zhang X, Zhang Q, Hu X, Zhang B (2015) Neural representation of three-dimensional acoustic space in the human temporal lobe. *Front Hum Neurosci* 9:203. [CrossRef Medline](#)