PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Spatial processing in the auditory cortex for stream segregation and localization

Martha M. SHIELL[1]; Elia FORMISANO[1]

[1] Maastricht University, Netherlands

ABSTRACT

The further apart two sounds are along the horizontal azimuth, the more likely they are to be heard as two separate sources. This spatial stream segregation (SSS) function may be dissociable from that of locating sound sources, but most neuroscience research on spatial hearing has focussed on the latter. Such research has uncovered spatial-sensitivity in the posterior auditory cortex, which has correspondingly been characterized as part of the "where" processing pathway for localization. However, other evidence indicates that this spatial sensitivity may be more relevant for SSS than for localization. We tested this hypothesis using 7 tesla functional magnetic resonance imaging. We measured brain activity in ten, normal-hearing adults while they listened to two concurrent auditory streams, spatially-separated along the horizontal azimuth. We applied multivoxel pattern classifiers to decode brain activity patterns associated with changes in location that resulted in either altered or constant SSS. We reasoned that if spatially-sensitive auditory cortex is optimized for SSS, then changes in spatial separation will be better decoded than mere changes in location. Our hypothesis was supported in the hemisphere contralateral to the sound locations. This result supports that spatial sensitivity in the auditory cortex is optimized for scene analysis rather than localization.

Keywords: Spatial stream segregation, localization, auditory cortex, auditory spatial processing

1. INTRODUCTION

The further apart two sounds are along the horizontal azimuth, the more likely they are to be heard as two separate sources. This phenomenon, known as spatial stream segregation (SSS) contributes to the essential function of auditory scene analysis. Typically, SSS is considered in tandem with localization processes: The brain combines input from each ear to calculate where the components of a sound wave originate from in space (localization), and simultaneously uses these spatial locations to help segregate these components into different auditory objects (SSS). However, this intuitive characterization is inconsistent with some evidence indicating that the two spatial functions are dissociable. For example, different weightings of interaural cues are used for a task that tests spatial relationships as compared to one that tests SSS (1). And, research on spatial hearing abilities after brain damage shows that while some patients are unable to locate the origin of sounds along the horizontal azimuth they can nonetheless benefit from spatial separation between target and masker in a sound identification task, and vice versa other patients are able to locate sounds but do not benefit from spatial separation in sound identification (2).

The neural substrates of SSS, dissociated from localization, remain unknown. While both primary and posterior mammalian auditory cortices (AC) show sensitivity to auditory spatial manipulations, this sensitivity has typically been considered as evidence for the involvement of these regions in localization as part of the "where" pathway of auditory processing (3). Thus, much research to understand processing in spatially-sensitive AC has centered on determining the code for localization of single sounds (for recent examples see 4,5). However, other evidence suggests that the spatial sensitivity of the AC may reflect the role of space in scene analysis: Activity levels here increase with spatial separation between concurrent sounds more than it does with changes in location of single sounds (6,7).

Here, we tested these conflicting interpretations of spatial-sensitivity in the AC by analyzing multivoxel patterns of brain activity measured by functional magnetic resonance imaging (fMRI) at 7 Tesla. In two different models, a classifier was trained to distinguish between brain activity

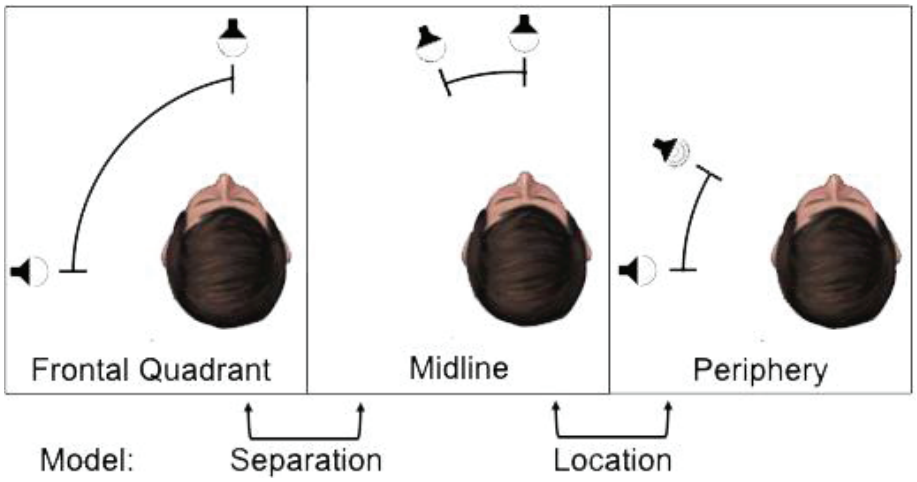

Figure 2 - The virtualized locations of two auditory streams in three different fMRI conditions (Frontal Quadrant, Midline, and Periphery), and the combination of these conditions into two models (Separation and Location).

patterns associated with changes in location that resulted in either altered or constant SSS. We followed the assumption that decoding accuracy between two conditions will increase with the degree of their perceptual difference. This assumption is based on evidence that similarity in perceptual information correlates with similarity of elicited multivoxel activity patterns (O'Toole et al., 2005). Thus, we reasoned that if spatially-sensitive AC is optimized for SSS, then the classifier will have higher accuracy when decoding location changes associated with SSS as compared to those associated with location alone.

2. METHODS

2.1 Participants

The study was approved by the Ethical Review Committee for the Faculty of Psychology and Neuroscience at Maastricht University. Participants were recruited from the Maastricht University undergraduate psychology program. All gave informed written consent and were compensated for their time either by course credits or gift vouchers. Ten right-handed healthy adults with self-reported normal hearing participated in the experiment (4 males and 6 females, age in years old: mean = 21.5, range = 18 - 31). The behavioural data from these participants is also reported on in our companion paper (8).

2.2 Stimuli

The stimuli are described in our companion paper (8), and were adapted from Middlebrooks and Onsan (1) rhythmic-masking release paradigm. In short, the stimuli consisted of two spatially-separated concurrent auditory streams. Each stream consisted of repeated 20 ms noise bursts, arranged to form one of two rhythms, termed rhythms A and B. The rhythms were such that the two streams created no energetic acoustic masking for one another. As such, when the streams were co-located, the participant heard a single stream of noise bursts, and only when the streams were sufficiently spatially-separated could the rhythm (A or B) be identified by the participant. The streams were spatialized with non-individualized head-related transfer functions from the CIPIC library (9) which were selected based on the participants' anthropometry and subjective choice (10).

The streams were arranged in three spatial conditions, shown in Figure 1. In the frontal quadrant condition, the locations of the two streams were fixed at -80 and -10 degrees (where the midline is 0 degrees). This spatial separation between the streams was sufficiently large so that they were easily segregated, and the identity of the streams' rhythm was clear to the participant. In the periphery and midline conditions, only one stream had a fixed location across participants: at -80 degrees for the periphery condition, and at -10 degrees for the midline condition. Here, the location of the second stream was determined for each participant individually, such that the spatial separation between the streams allowed the participant to correctly identify the rhythm with 81% accuracy. This was

calculated by a staircase procedure that was completed over two sessions of behavioural testing prior to the fMRI session. The details of this procedure are described in our companion paper (8).

2.3 fMRI Acquisition

Participants were scanned 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). Participants completed 10 runs of fMRI to measure BOLD signal [T2*-weighted gradient echo-planar imaging, volumes = 60, number of slices = 60, voxel size = 1.1 mm isotropic, matrix size = 188 x 188, TR = 10080 ms, TA = 1780 ms, silent gap for stimulus presentation = 8300 ms, generalized autocalibrating partially parallel acquisitions (GRAPPA) = 3]. Following this, we completed a T1-weighted scan (MPRAGE sequence, TR = 3100 ms, time to inversion = 1500 ms, TE = 3.5 ms, 0.7 mm isotropic voxels, matrix size = 320 x 320, number of slices = 256, flip angle = 5°, GRAPPA = 3), and a proton density scan TR = 2160 ms, TE = 3.5 ms, 0.7 mm isotropic voxels, matrix size = 320 x 320, number of slices = 256, flip angle = 5°, GRAPPA = 3).

Auditory stimulus presentation during the fMRI was controlled by Psychopy (11), and was accompanied by a visual fixation cross presented via a projector and mirror. Each fMRI run contained 30 measurements of experiment trials interleaved with 30 measurements of rest. For experiment trials, the auditory stimulus was presented via MRI-compatible ear buds (S14; Sensimetrics) at a comfortable level during the 8300 ms silent gap between measurements. The order of the stimuli was pseudorandom. The 30 trials included 3 repetitions of each of the 2 rhythms and 3 conditions. During the subsequent rest trial, the fixation cross changed from black to red for 2000 ms. Participants were instructed to respond whether the previous target stream was rhythm A or rhythm B with this cue.

2.4 fMRI Preprocessing

Image 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 cortices.

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).

For each hemisphere of the preprocessed anatomical volume in native space, a region-of-interest (ROI) was manually drawn. The ROI was defined liberally to include Heschl's gyrus, the upper half of the posterior superior temporal gyrus, the planum temporale, and the ascending limb of the Sylvian fissure. This broad definition allowed us to avoid excluding potentially relevant information. The ROI was drawn on an inflated reconstruction of the cortical surface and projected into the co-registered volume space; the resultant voxels were used to extract fMRI signal for the multivoxel pattern analysis (MVPA).

2.5 MVPA

The MVPA was implemented with a custom MATLAB script. We designed two models (Figure 1), termed the separation and location models. Each model consisted of a combination of two conditions: the frontal quadrant and midline conditions for the separation model, and the midline and periphery conditions for the location model. Note that in both models, the most peripheral stream changes location across the two conditions. In the separation model this change in location is associated with a change in SSS, whereas in the location model SSS it is held constant. For each model, trials of rhythm A and B were trained and tested separately.

In each voxel within the ROI, we calculated the percentage signal change for each experimental trial relative to its subsequent rest trial. The participant's 10 runs were split into training and testing data using a leave-run-out scheme with 8 runs for training and 2 for testing, and 5 cross-validations.

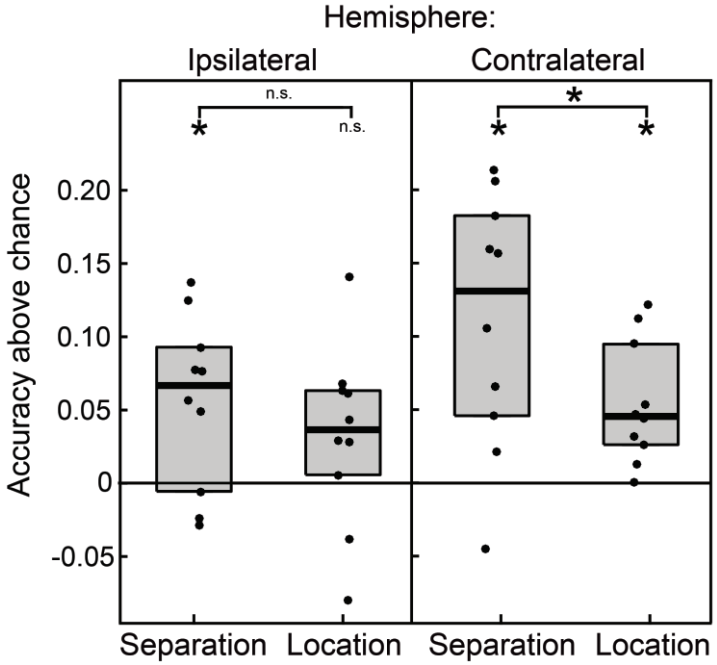

Figure 1 - Decoding accuracy from the multivoxel pattern analysis for two hemispheres and two models. Chance (y = 0) was calculated for each hemisphere separately through random permutations of training and testing labels. Thick bars indicate group median values, boxes indicate group quartiles, and circles indicate individual participants.

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.

For each model, we applied a linear support vector machine (SVM) to learn to distinguish between the multivoxel patterns of activity associated with each condition in the model (12). Voxels within each hemisphere's ROI were reduced iteratively using a recursive feature elimination (RFE) procedure (13) 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 approximately 250 voxels remained at the 15th level of RFE. Decoding accuracy at each level was calculated as the average proportion of correctly classified testing trials across the two classes. The maximum accuracy across RFE levels was selected for each split and then these values were averaged. Finally, the accuracies across rhythm conditions were averaged to give a single accuracy value for each hemisphere and model.

Although the theoretical chance level for decoding accuracy is 50%, the RFE procedure inflates this value. Therefore, we calculated chance level empirically for each of the 20 tested hemispheres individually. This was done by permuting the condition labels for each model and repeating the full RFE procedure 1200 times (200 times per model and rhythm). For each hemisphere, we subtracted its empirical chance level from each model's classification accuracy to give decoding accuracy above chance (DAC). Wilcoxon signed-rank tests were used to compare DAC to chance and between models in each hemisphere with a threshold of $p < 0.05$, adjusted for six multiple comparisons by false-discovery rate (14).

3. RESULTS

DAC for each hemisphere and each model are shown in Figure 2, with statistical values in Table 1. For the location model, decoding was above chance only in the contralateral hemisphere. For the separation model, both hemispheres were above chance. Consistent with our prediction, DAC was

higher for the separation model than the location model, but only in the contralateral hemisphere.

Table 1 – Statistical test values for comparisons of the measure of decoding accuracy above chance. Statistical significance is indicated by bold.

Hemisphere	Test	*W*	*p* (FDR-corrected)
Ipsilateral (left)	Location > Chance	49	**0.041**
	Separation > Chance	42	0.192
	Separation > Location	36	0.224
Contralateral (right)	Location > Chance	53	**0.018**
	Separation > Chance	55	**0.012**
	Separation > Location	49	**0.027**

4. SUMMARY

Using MVPA, we found that decoding brain activity patterns that were associated with changes in SSS was better than that associated with changes in location only. This finding supports the hypothesis that spatially-sensitive AC is optimized for SSS. Interestingly, the contralateral hemifield also represented location information, which suggests that localization processing is carried out in spatially-sensitive AC in addition to SSS. The current dataset may help dissociate localization and SSS further through future examination of the distribution of each voxel's contribution to the classification success. In any case, our result demonstrates that further understanding of auditory spatial processes cannot be accomplished by examining responses to single sound locations only, but will require concurrent spatially-separated sounds. This implication is relevant not only for future neuroscience research, but also for applied research in hearing technology development.

ACKNOWLEDGEMENTS

This research was supported by grant NWO VICI 453-12-002 to EF. MS was supported by a postdoctoral fellowship from the Canadian Institutes of Health Research.

REFERENCES

1. Middlebrooks JC, Onsan ZA. Stream segregation with high spatial acuity. J Acoust Soc Am. Acoustical Society of America; 2012;132(6):3896–911.
2. Duffour-Nikolov C, Tardif E, Maeder P, Thiran AB, Bloch J, Frischknecht R, et al. Auditory spatial deficits following hemispheric lesions: Dissociation of explicit and implicit processing. Neuropsychol Rehabil. Psychology Press; 2012 Sep 20;22(5):674–96.
3. Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neuro. 2009 May 26;12(6):718–24.
4. Derey K, Valente G, de Gelder B, Formisano E. Opponent Coding of Sound Location (Azimuth) in Planum Temporale is Robust to Sound-Level Variations. Cereb Cortex. 2015 Nov 5;:1–47.
5. Ortiz-Rios M, Azevedo FAC, Kuśmierek P, Balla DZ, Munk MH, Keliris GA, et al. Widespread and Opponent fMRI Signals Represent Sound Location in Macaque Auditory Cortex. Neuron. 2017 Feb;93(4):971–4.
6. Zatorre RJ, Bouffard M, Ahad PA, Belin P. Where is “where” in the human auditory cortex? Nat Neuro. 2002 Aug 19;5(9):905–9.
7. Smith KR, Hsieh I-H, Saberi K, Hickok G. Auditory spatial and object processing in the human planum temporale: no evidence for selectivity. J Cogn Neuro. 2009 Apr;22(4):632–9.
8. Shiell MM, Formisano E. Acuity of spatial stream segregation along the horizontal azimuth with non-individualized head-related transfer functions. Aachen; 2019. pp. 1–6.
9. Algazi VR, Duda RO, the DTWO, 2001. The CIPIC HRTF database - IEEE Conference Publication. New Paltz, USA; 2001.
10. Seeber BU, Fastl H. Subjective selection of non-individual head-related transfer functions. Georgia

Institute of Technology; 2003 Jul 1.
11. Peirce JW. PsychoPy—Psychophysics software in Python. J Neurosci Methods. 2007 May;162(1-2):8–13.
12. Formisano E, De Martino F, Bonte M, Goebel R. "Who" is saying “what?” Brain-based decoding of human voice and speech. Science. 2008 Nov 7;322(5903):970–3.
13. De Martino F, Valente G, Staeren N, Ashburner J, Goebel R, Formisano E. Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. NeuroImage. 2008 Oct 15;43(1):44–58.
14. Benjamin Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc Series B Methodol. 1995;57(1):289–300.