Attention modulates within-category differentiation of natural auditory objects in human auditory cortex

Ilkka Muukkonen[1,2], Onnipekka Varis[1], Patrik Wikman[1,3] (Corresponding Author)

[1]Department of Psychology, PO Box 21 FI-00014 University of Helsinki, Finland; [2]Department of Brain and Cognition, PO box 03711, KU Leuven, Belgium; [3]Advanced Magnetic Imaging Centre, PO Box 11000 FI-00076 Aalto University, Espoo, Finland

Abstract

Categorical representations of auditory objects arise in non-primary auditory cortex. However, it is unclear whether these regions support subcategorical differentiation (e.g., instrument types), and whether attention enhances this differentiation. fMRI data (n = 20) were acquired in two experiments: OA, where participants listened to sound objects (speech, instruments, animals); and 3OA, where they attended to designated objects in scenes containing one object per category. SVM decoders were trained on OA-data to distinguish subcategory objects (e.g., dog vs. bird). Decoders were tested: (1) on OA-data to identify regions with stimulus-dependent within-category differentiation; (2) on 3OA scenes to assess if attention boosts within-category differentiation (comparing attended vs. distractor object differentiation). Stimulus-dependent and attention-related speaker identity differentiation involved spectrally non-sensitive STG regions, whereas animal and instrument differentiation was confined to spectrally sensitive regions. Results suggest speaker identity differentiation involves higher-level object representations, while other naturalistic sounds are differentiated via lower-level acoustic features.

Keywords: auditory object, decoding, attention, scene, categoriy, within category differentiation

Introduction

Representations of natural auditory objects have been suggested to arise in the anteroventral auditory cortex (AC; (Boemio et al., 2005; Lewis et al., 2011; Petkov et al., 2009; Rauschecker & Scott, 2009; Theunissen & Elie, 2014). Most studies on the neural representation of natural auditory objects focus on differentiating categorical object representations (e.g., speech sounds vs. instrument sounds) from representations arising from sensitivity to low-level acoustic features (e.g., (Leaver & Rauschecker, 2010; Norman-Haignere et al., 2015; Wikman et al., 2025). However, object perception also involves differentiation of different individuals within a category. Such studies have focused on human voices, showing that several locations along the superior temporal gyrus (STG) show voice specificity (see (Belin et al., 2011). Yet, whether similar differentiation exists for other object types (e.g., instruments) remains unclear. Further, whether selective attention can operate on such object representation has not been previously studied.

We identified AC fields with within-category differentiation for speech, instruments, and animals using fMRI data (n = 20) from two experiments. In experiment OA, participants listened to single auditory objects (6 speakers, animals, and instruments). In experiment 3OA, participants attended to a designated object in scenes with one object from each category. Using OA-data, we determined spatial patterns that differentiate specific identities within categories (e.g., dog vs. other animals). For 3OA-data we determined where the *attended* identity could be decoded from the other within-category *attended* objects more reliably than *distractor* identities from other within-category *distractor* object identities (Figure 1). Consequently, we determined AC fields where selective attention boosts within-category differentiation when listening to scenes with multiple objects.

Methods

For each category, there were 6 different sound objects (4–7 s, speech: e.g., male, female; instruments: e.g., trumpet, guitar; animals: e.g., dog, bird), and for each object there were 8 different exemplars (144 stimuli in total). In 3OA, participants were to attend to one of the three sounds in the sound scene, and the experiment comprised several trials where each object was either the attended or a distractor in the scene. Thus, the trials were similar in their stimulus-level features, and only the focus of attention varied between trials. There were 4 runs in both OA and 3OA.

Preprocessed (fmriprep) and fsaverage-surface projected data (Fischl, 2012) were used for pairwise SVM-decoding (searchlight, 8 mm radius) to test: (1) *stimulus-level effects* (Figure 1, top) – i.e., whether and where neural response patterns can distinguish different auditory objects in OA; and (2) *attention-effects* (Figure 1, middle and bottom) – i.e., whether the attended object can be decoded from scenes with several objects. For both analyses, we decoded each possible auditory object pair within a category (e.g. trumpet vs. 12-string guitar, 15 comparisons per category), and took the mean of the pairwise decoding. For a description of the pipeline see Figure 1. All decoding was done within-subjects, and statistical inference calculated with permutation tests (cluster threshold: z=3.1, all FWER corrected p<.05).

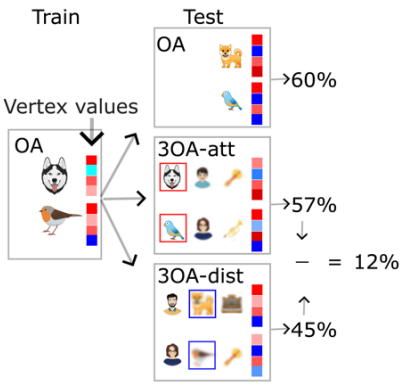

Figure 1: Schematic of pairwise SVM-decoding, showing decoding of dog vs. bird objects. In OA the SVM was trained using OA data and tested on OA-data using a leave-one-run-out method (note different exemplars in each run). In 3OA the SVM was trained using all OA-trials and tested on 3OA trials (all runs), separately for trials where the objects were attended (3OA-att) and trials where they were distractors (3OA-dist). Thereafter the difference of the as-attended and as-distractor decoding accuracies were calculated (relevant objects: red boxes; distractor: blue boxes).

Results

Stimulus-level differences within animal and instrument categories (Figure 2, middle and bottom, yellowish and white) were present in primary and belt AC. For animal sounds, stimulus effects spanned the whole AC, while attention effects (Figure 2, blueish and white) overlapped in the posterior AC. For instrument sounds, stimulus-effects were more localized in superior temporal plane, and attention modulated responses in anterior parts. In contrast, for separating different speakers from each other, only higher-level auditory areas, around area A4 (Glasser et al., 2016), showed significant results. Further, attention effects were mostly non-overlapping along STG, though close to fields displaying stimulus-effects.

Discussion

We found both stimulus and attention effects for all three sound categories within AC. The attention effects were mostly overlapping with the stimulus effects, showing that attention may use similar neural resources to differentiate different objects within a category. Notably, the regions showing speech-sound effects corresponded to regions that have previously been observed to process speech object-level information (Norman-Haignere et al., 2015; Norman-Haignere & McDermott, 2018). Similar regions have also been suggested to differentiate human voices from each other (Bonte et al., 2014). Thus, selective attention may operate on such voice representations.

In contrast, stimulus-dependent and attention-related subcategorical differentiation for animal and instrument objects was observed in parts of AC known to process acoustic features. The instrument cluster was centered on fields associated with fine spectral and slow temporal modulation and the animal cluster on fields associated with fast temporal and coarse-scale spectral modulation (Norman-Haignere et al., 2015). This highlights that for these categories low-level acoustic features, or combinations of features, can be used for within-category differentiation and selective attention can boost sensitivity to this variation in multi-object scenes.

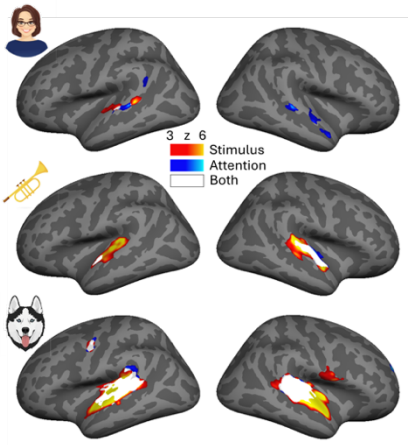

Figure 2: Mean pairwise SVM-decoding within speech (top), instrument (middle), and animal sound categories (bottom), for objects played alone (OA, red, yellow); and for the difference between attended and distractor sounds (3OA, blue). Overlap between both experiments is depicted in white.

Acknowledgements

This work was supported by The Research Council of Finland (grant number #1348353). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We want to thank M. Salmikivi, W. Vikatmaa, S. Rossow, and L. Lehtimäki for helping with data acquisition, as well as Ville Laaksonen and Jaakko Kauramäki for help with the study design.

References

Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature neuroscience*, *8*(3), 389-395.

Bonte, M., Hausfeld, L., Scharke, W., Valente, G., & Formisano, E. (2014). Task-Dependent Decoding of Speaker and Vowel Identity from Auditory Cortical Response Patterns. *Journal of Neuroscience*, *34*(13), 4548-4557. https://doi.org/10.1523/Jneurosci.4339-13.2014

Fischl, B. (2012). FreeSurfer. *Neuroimage*, *62*(2), 774-781.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., & Jenkinson, M. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171-178.

Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *The Journal of neuroscience*, *30*(22), 7604.

Lewis, J. W., Talkington, W. J., Puce, A., Engel, L. R., & Frum, C. (2011). Cortical networks representing object categories and high-level attributes of familiar real-world action sounds. *Journal of cognitive neuroscience*, *23*(8), 2079-2101.

Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron*, *88*(6), 1281-1296.

Norman-Haignere, S. V., & McDermott, J. H. (2018). Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex. *Plos Biology*, *16*(12), e2005127.

Petkov, C. I., Logothetis, N. K., & Obleser, J. (2009). Where are the human speech and voice regions, and do other animals have anything like them? *The Neuroscientist*, *15*(5), 419-429.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature neuroscience*, *12*(6), 718-724.

Theunissen, F. E., & Elie, J. E. (2014). Neural processing of natural sounds. *Nature Reviews Neuroscience*, *15*(6), 355-366.

Wikman, P., Muukkonen, I., Kauramäki, J., Laaksonen, V., Varis, O., Petkov, C., & Rauschecker, J. (2025). Selective attention network in naturalistic auditory scenes is object and scene specific. *bioRxiv*, 2025.2001.2003.631190. https://doi.org/10.1101/2025.01.03.631190