Using decoding error patterns to trace the neural signature of auditory scene perception

Yaelan Jung
Department of Psychology
University of Toronto
Toronto, ON, Canada
yaelan.jung@mail.utoronto.ca

Bart Larsen
Department of Psychology
University of Pittsburg
Pittsburgh, PA, USA
bsl18@pitt.edu

Dirk B. Walther
Department of Psychology
University of Toronto
Toronto, ON, Canada
bernhardt-walther@psych.utoronto.ca

***Abstract* — Multi-voxel pattern analysis has become a standard tool for analyzing neuroimaging data. In addition to the decoding accuracy, the particular pattern of decoding errors contains valuable information about the nature of neural code. We here use decoding errors in order to trace the processing of non-speech, non-music complex natural sounds in primary auditory cortex (A1) and its subdivisions, as well as across cortex. We use error patterns derived from an analysis of the physical stimulus properties and from a separate behavioral categorization experiment as references for this analysis. A linear mixed-effects model allows us to measure the similarity of the decoding error patterns to both of these references at the same time. Within A1 we find an interesting posterior-to-anterior trend, where the most posterior parts of A1 are linked more closely to properties of the stimuli, whereas more anterior parts of A1 are more closely aligned with human categorization behavior. In an exploratory searchlight analysis we find a similar trend across cerebral cortex. Decoding errors in occipital, posterior parietal and temporal cortex are more closely related to sound structure. Decoding errors in prefrontal cortex resemble behavioral errors. Our work illustrates the importance of decoding error analysis at the example of auditory processing in cortex and proposes linear mixed-effects modeling as a simple yet effective way for comparing decoding errors to reference error patterns obtained from auxiliary analyses or experiments.**

Keywords— fMRI, MVPA, Auditory cortex, mixed-effects model, error pattern analysis

I. Introduction

In the last decade, machine learning algorithms have given us access to information encoded in the human brain that was not accessible using univariate analysis methods. Inspection of the patterns of errors from decoding neural activity elicited by natural scene *images* has allowed us to compare the neural codes across different experimental conditions [1; 2; 3; 4]. Our group has previously introduced a framework for better quantification of multi-way error comparisons using linear mixed-effects models [5]. Specifically, we have developed a framework for linking error patterns in the neural code for natural scene categories elicited by specific, hypothesis-driven stimulus manipulations.

Here we demonstrate how the same modeling approach can be applied to determining the evolution of the neural codes involved the perception of complex, real-world *sounds*. When auditory information enters the brain, it is first processed in a series of sub-cortical nodes, including the superior olives in the brain stem, the inferior colliculus in the midbrain, and the medial geniculate nucleus of the thalamus. Throughout all of these processing steps, tonotopic mapping according to sound frequencies is preserved and carried all the way into primary auditory cortex, which is located on Heschl's gyrus (HG), in the Sylvian fissure, on the superior bank of the temporal lobe. Cortical processing of auditory information proceeds in the belt and parabelt areas just around A1, but little is known about further, higher-level processing of complex sounds that do not involve music or speech.

In the work presented in this article we were able to track auditory processing throughout cortex from codes that resemble the physical characteristics of the auditory stimuli to codes that strongly resemble categorization behavior by human observers. We used mixed-effects modeling of error patterns in order to compare the explanatory strengths between two fixed effects, one related to the physical structure of the stimuli and the other related to human behavior.

We found that the auditory cortex is largely dominated by stimulus-related properties, presumably due to its tonotopic organization. The middle part of Heschl's gyrus (TE1.0), however, more strongly represents auditory scene category properties related to human behavior. A whole-brain searchlight analysis revealed a prevalence of stimulus-related codes in occipital, posterior parietal and temporal cortex and a prevalence of behavior-related codes in frontal regions.

II. Methods

To assess the characteristics of neural activity patterns in comparison with the physical structure of sounds and human behavior, we first collected three types of error patterns:

978-1-5386-4291-7/18/$31.00 ©2018 IEEE

errors from a neural decoder, errors reflecting the physical structure of the stimuli, and errors from a behavioral categorization task. Then we used a linear mixed-effects (LME) model to evaluate how much variance in neural decoding errors can be explained by the error patterns of stimuli structure and patterns of behavioral errors.

A. Error patterns from neural classifier

In a neuroimaging experiment, we presented 13 participants (18-25 years old, 6 females) with sounds of real-world scenes, while their brain activity was recorded using functional magnetic resonance imaging (fMRI). Sixty-four sound clips of natural scene sounds were used in the experiment. Individual sound clips represented one of four scene categories (beach, city street, forest, and office). We presented four runs with eight blocks per run. During each block, a 15-second sound clip was played to the participant using Sensimetrics S14 MR-compatible in-ear earphones at a sound level of approximately 70-80 dB. Sound clips were equalized for perceptual loudness.

FMRI data were recorded on a Siemens Tim Trio MRI scanner at 3 Tesla, using an echo-planar sequence with TE = 28 ms and TR = 2.5 s. 48 axial slices with 3 mm thickness were recorded without gap, resulting in isotropic voxels of 3 x 3 x 3 mm. These data were then motion-corrected, spatially smoothed with a 2mm Gaussian kernel, temporally smoothed with a high-pass filter at 1/400 Hz, and normalized to the mean of each run's initial fixation period.

We used multi-voxel pattern analysis to decode the scene categories of the sounds from the fMRI data. Specifically, we trained a linear support vector machine (SVM) to predict auditory scene categories in a leave-one-run-out cross validation procedure. The classification analysis was performed in anatomically defined ROIs, the auditory cortex (ACX) as well as its anatomical sub-divisions: the planum temporale (PT), posteromedial HG (TE1.1), middle HG (TE1.0), anterolateral HG (TE1.2), and the planum polare (PP). To decrease the dimensionality of the neural data, we used voxel selection based on the ANOVA on each voxel's activity with the main effect of scene category in a nested leave-one-out cross validation.

The performance of the neural decoder was recorded in a confusion matrix, whose rows indicate the true label of each category and whose columns indicate the categories predicted by the neural decoder. Diagonal elements represent correct predictions, off-diagonal elements decoding errors. We here mainly focus on comparing the decoding errors.

B. Error patterns corresponding to sound stimuli

Two types of the reference errors were computed, errors related to physical structure of stimuli and those related to behavior. First, we computed the auditory features of the sounds using the cochleagram, which analyzes the frequency content of the sounds by modeling the biomechanics of the ear [6]. The cochleagrams of the individual sound stimuli was computed at 128 frequency bands and integrated over the duration of the sounds. The power distribution of the sounds within these frequency bands was used as input to a linear SVM, which predicted scene categories of the sounds in a 16-fold cross validation. The off-diagonal elements of the resulting confusion matrix contain the error patterns reflecting the physical structure of the sounds (Fig 1).

C. Error patterns corresponding to behavior

To obtain error patterns related to human behavior, we conducted a separate behavioral experiment. Twenty-five participants (18 to 21 years old, 16 females) were separately recruited for this experiment. Participants listened to individual sound clips and responded with the perceived scene category of the sound clip. To ensure that participants produced enough errors, we degraded the quality of the sound stimuli by overlaying them with masking sounds. Masking sounds were consisting of 30 ms snippets of pure tones, whose frequency was randomly chosen between 50 and 2000 Hz. Performance of individual participants was recorded in a confusion matrix. We averaged the confusion matrices across all the participants, and extracted off-diagonal elements as the behavioral error pattern (Fig 1).

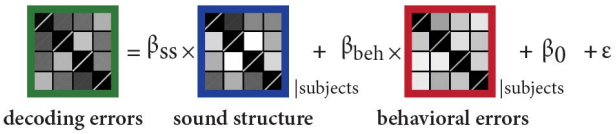

Fig 1. Formula of the LME model for analyzing error patterns.

D. Mixed-effect models

To understand the relationship between the neural decoding errors and the two types of reference errors, we set up a linear mixed-effects model (LME) with the reference errors as the fixed effects and subjects of the neuroimaging experiment as the random effects.

A LME describes the relationship between the observed data (here, decoding errors) and the experiment design (here, the reference errors). In this model, we set the reference errors as the fixed-effects coefficients: error patterns of sound structure (blue confusion matrix in Fig 1) and error patterns of behavior (red confusion matrix in Fig 1) to determine how much the respective coefficients contribute to explaining the variance in decoding errors (green confusion matrix in Fig1).

Here, β_{SS} are the fixed-effects coefficients for the error patterns based on the physical structure of the sounds, β_{beh} are the fixed-effects coefficients for the behavioral errors, β_0 is the intercept, and ε are the residuals. The slope of the random effects for each fixed effect was estimated separately to control for the variance at the level of subjects.

The estimates of the fixed-effects coefficients are interpreted as the correspondence of the structure of the neural code with either physical sound structure or human categorization behavior. To evaluate which type of reference error pattern

explains the variance in decoding errors best, we compared the two fixed-effect coefficients using an F-test. Significance in this F-test indicates that the two fixed-effects coefficients are significantly different in their ability to explain the variance in the decoding errors.

We first examined this relationship between decoding errors and the reference errors in the auditory cortex and its anatomical sub-divisions. Using the LME model of the decoding error patterns in each ROI we determined whether that ROI is more closely related to the physical structure of the sounds or to categorization behavior. In addition, we performed an exploratory searchlight analysis, which used a linear SVM classifier in a 9.076 cm^3 (7×7×7 voxels) cubic searchlight. In the searchlight analysis, we corrected for multiple comparisons at the cluster level. To this end we used AFNI tools 3dFWHMx and 3dClustSim (AFNI version 18.024), to perform cluster simulation, which resulted in a minimum cluster size of 13 voxels.

III. Results

A. ROI-based analysis

We were able to decode auditory scene categories from the auditory cortex and its sub-divisions significantly above chance (details reported elsewhere).

We analyzed patterns of decoding errors using the LME model approach described above. In the entire auditory cortex, we did not see any difference in the fixed-effects coefficients (see Fig 2), $F(2, 153) = 0.004$, $p = .951$, suggesting that decoding errors were equally well explained by sound structure errors and by behavior errors.

However, when we examined the individual sub-divisions of the auditory cortex, we observed an interesting trend. We saw slight domination of sound structure in the most posterior part of the auditory cortex, the planum temporale (although not significantly, $F(2, 153) = .06$, $p = .806$). In the middle HG (TE1.0), decoding errors were mostly explained by behavior

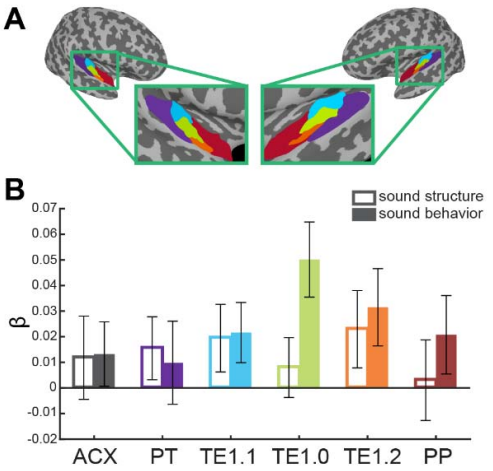

Fig 2. A) Anatomically defined sub-regions of the auditory cortex. B) The estimates of the fixed effects from an LME model in the auditory cortex (ACX) and its sub-regions.

but not by sound structure, $F(2, 153) = 3.835$, $p = .025$. Numerically, this pattern persists in the anterolateral HG (TE1.2), even though the difference in two fixed-effect coefficients did not reach statistical significance, $F(2, 153) = .01$, $p = .0916$. These findings in the LME model show that although the entire auditory cortex did not show a strong domination by sounds structure versus behavior, sub-regions of the auditory cortex exhibit a posterior-to-anterior trend: anterior regions relate more to stimulus structure, and posterior regions more to behavior.

B. Searchlight analysis

To further explore how neural representation in different brain regions can be explained by properties of physical structure and behavior, we performed an exploratory searchlight analysis throughout the whole brain. The same leave-one-run-out cross validation analysis as for the ROIs was performed at each searchlight location. Decoding accuracy and error patterns were recorded at each location. We modeled the contributions to the decoding errors with the same LME model as for the pre-defined ROIs. To illustrate whether each searchlight location is more related to behavior versus sound structure, we show the contrast between two coefficients by subtracting the estimates of the sound structure coefficients from the estimates of behavior coefficients $(\hat{\beta}_{beh} - \hat{\beta}_{SS})$. Thus, the searchlight locations with positive contrast (warm colors in Fig 3) indicate that decoding errors are explained better by behavior, and the locations with negative contrast (cold colors in Fig 3) indicate that decoding errors are better explained by sound structure.

The searchlight results were fairly consistent with our ROI-based findings. First, we found two big significant clusters at the location of the auditory cortex (417 voxels in the right hemisphere, 151 voxels in the left hemisphere). These two clusters had intermixed blobs of positive and negative contrasts between the two coefficients; in the central part of each cluster, the contrast between the two coefficients was positive, i.e., the estimates of the behavioral coefficients were bigger than those for the sound structure. In the posterior part of the cluster, the contrasts were negative, showing estimates for sound structure errors exceeding the estimates for the behavioral errors. We also found clusters with positive contrasts in the frontal lobe, specifically in the right superior/medial frontal gyrus (157 voxels) and the left middle/superior frontal gyrus (296 voxels). We also found clusters with positive contrasts in the left precentral gyrus (44 voxels) as well as in the right angular gyrus, spilling over to the right middle temporal gyrus (355 voxels). In comparison, we found a cluster with negative contrasts in primary visual cortex (left middle occipital gyrus; 43 voxels).

Altogether, these searchlight results show a posterior-to-anterior trend: in the posterior part of the brain, the neural representations are more related to the physical structure of the stimuli, and in the anterior part of the brain, the neural representations are more related to human behavior. We

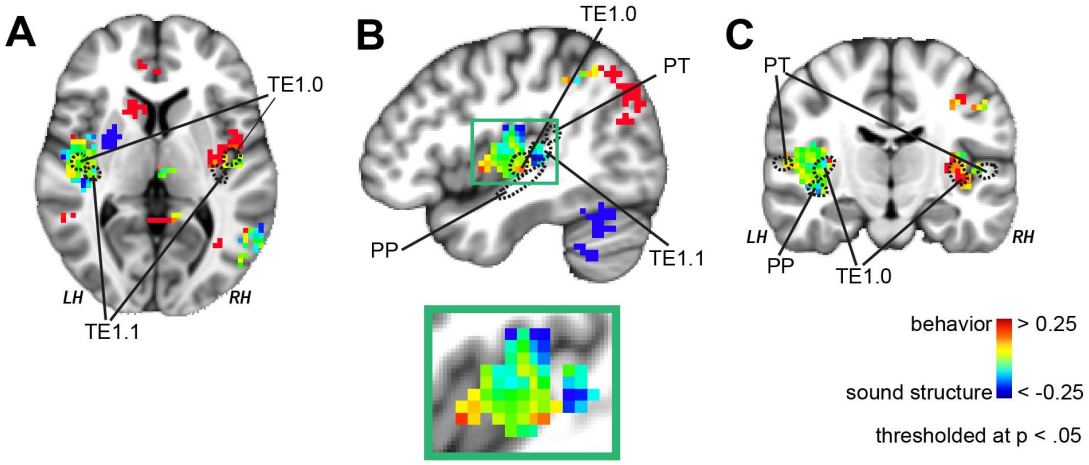

Fig 3. Searchlights with the difference in regression coefficients between behavior and sound structure, thresholded using the F-test comparing the effects of two coefficients at $p < 0.5$. Warm colors indicate the searchlight locations where the behavioral errors can explain the decoding errors significantly better than stimuli structure errors, cold colors regions where sound structure outweighs behavior. The predefined regions of interests are outlined with dotted lines

observe this trend within primary auditory cortex (see Fig 3B) as well as throughout the entire brain.

IV. Discussions

We have demonstrated here how a simple LME model can be used effectively in order to characterize neural codes of complex auditory scenes as being more closely related to stimulus structure versus behavior. We have applied this procedure to the perception of real-world complex auditory scenes. Using an exploratory searchlight analysis we found regions beyond the auditory cortex and its vicinity in the occipital, parietal and frontal lobes that are involved in processing complex sounds. The linear mixed-effects modeling approach allowed us to characterize these regions as being more related to sound structure (mostly posterior regions) or more related to sound categorization behavior (mostly anterior regions). Our approach of analyzing classification errors is similar in spirit to some forms of representational similarity analysis [8]. Importantly, we here analyze decoding errors rather than correlations of activity vectors. This allows us to more directly compare to auxiliary analyses and experiments, which can easily be cast as classification analysis.

To summarize, analysis of decoding errors allows us to more comprehensively characterize neural codes than using mere decoding accuracy. For perception, the physical structure of the stimuli and the behavioral errors of human observers categorizing the stimuli are useful guideposts for following the path of processing the information through the brain, as we have shown here with the example of auditory perception. However, we believe that this approach of characterizing decoding errors is applicable more broadly to other domains of cognitive neuroscience as well, such as research on memory or action planning.

References

[1] Walther, D. B., Chai, B., Caddigan, E., Beck, D. M., & Fei-Fei, L. (2011). Simple line drawings suffice for functional MRI decoding of natural scene categories. Proceedings of the National Academy of Sciences, 108(23), 9661-9666

[2] Walther, D. B. (2013, June). Using confusion matrices to estimate mutual information between two categorical measurements. In Pattern Recognition in Neuroimaging (PRNI), 2013 International Workshop on (pp. 220-224). IEEE.

[3] Olivetti, E., & Walther, D. B. (2015, June). A Bayesian Test for Comparing Classifier Errors. In Pattern Recognition in NeuroImaging (PRNI), 2015 International Workshop on (pp. 69-72). IEEE.

[4] Choo, H., & Walther, D. B. (2016). Contour junctions underlie neural representations of scene categories in high-level human visual cortex. NeuroImage, 135, 32-44.

[5] Choo, H., & Walther, D. B. (2017, June). Modeling the effect of stimulus perturbations on error correlations between brain and behavior. In Pattern Recognition in Neuroimaging (PRNI), 2017 International Workshop on (pp. 1-4). IEEE

[6] Wang, D., & Brown, G. J. (2006). Computational auditory scene analysis: Principles, algorithms, and applications. Wiley-IEEE Press.

[7] Norman-Haignere, S., Kanwisher, N., & McDermott, J. H. (2013). Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. Journal of Neuroscience, 33(50), 19451-19469.

[8] Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. Frontiers in systems neuroscience, 2, 4.