The representation of low-level acoustic features in the human brain

Lucine Oganesian[1], Fatma Imamoglu[1,2], Jack L. Gallant[1]
University of California, Berkeley[1], International Computer Science Institute[2]

University of California Berkeley SUMMER UNDERGRADUATE RESEARCH FELLOWSHIPS

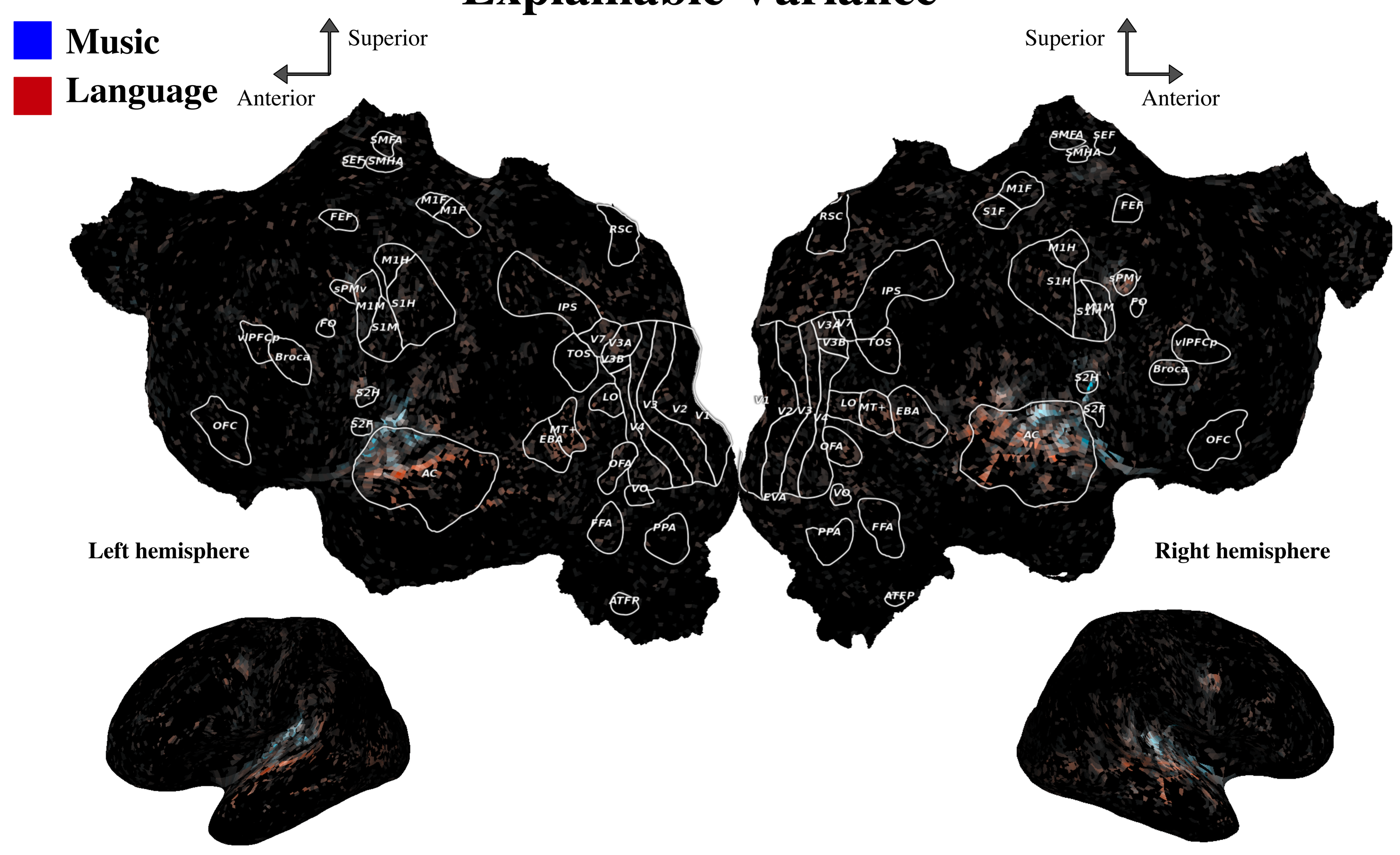
Introduction

Speech and music are two of the most important auditory signals for humans. However, little is known about how various acoustic features in speech and music are represented and processed in the human brain. A few previous studies have examined the representation of music in the human brain.[1], [2], [3], [4] However, these studies used cross-subject averaging or they merely probed music-speech contrasts. Thus, they were not sensitive enough to reveal details of tuning in individual human brains.

Here we asked whether acoustic features drawn from music information retrieval (MIR) can be used to better understand how speech and music are represented in the human brain. We also evaluated other acoustic features used in prior studies.[1] One human subject listened to broad speech and music stimuli while brain activity was measured using fMRI. Auditory features were extracted from the stimuli and a voxel-wise encoding model approach was used to estimate how each location in the cerebral cortex responded to these features. Our approach estimates a separate model for every voxel in every individual without loss of information due to averaging across voxels or subjects. Encoding models were verified by assessing prediction accuracy on a separate held-out data set (the validation set).

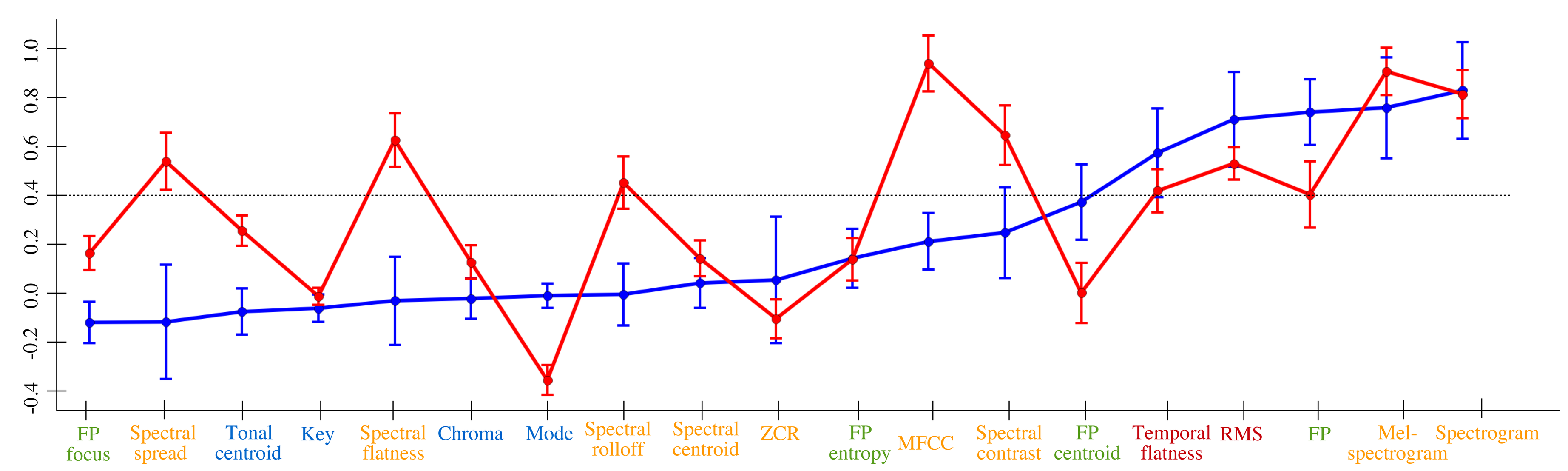
Results

Explainable Variance

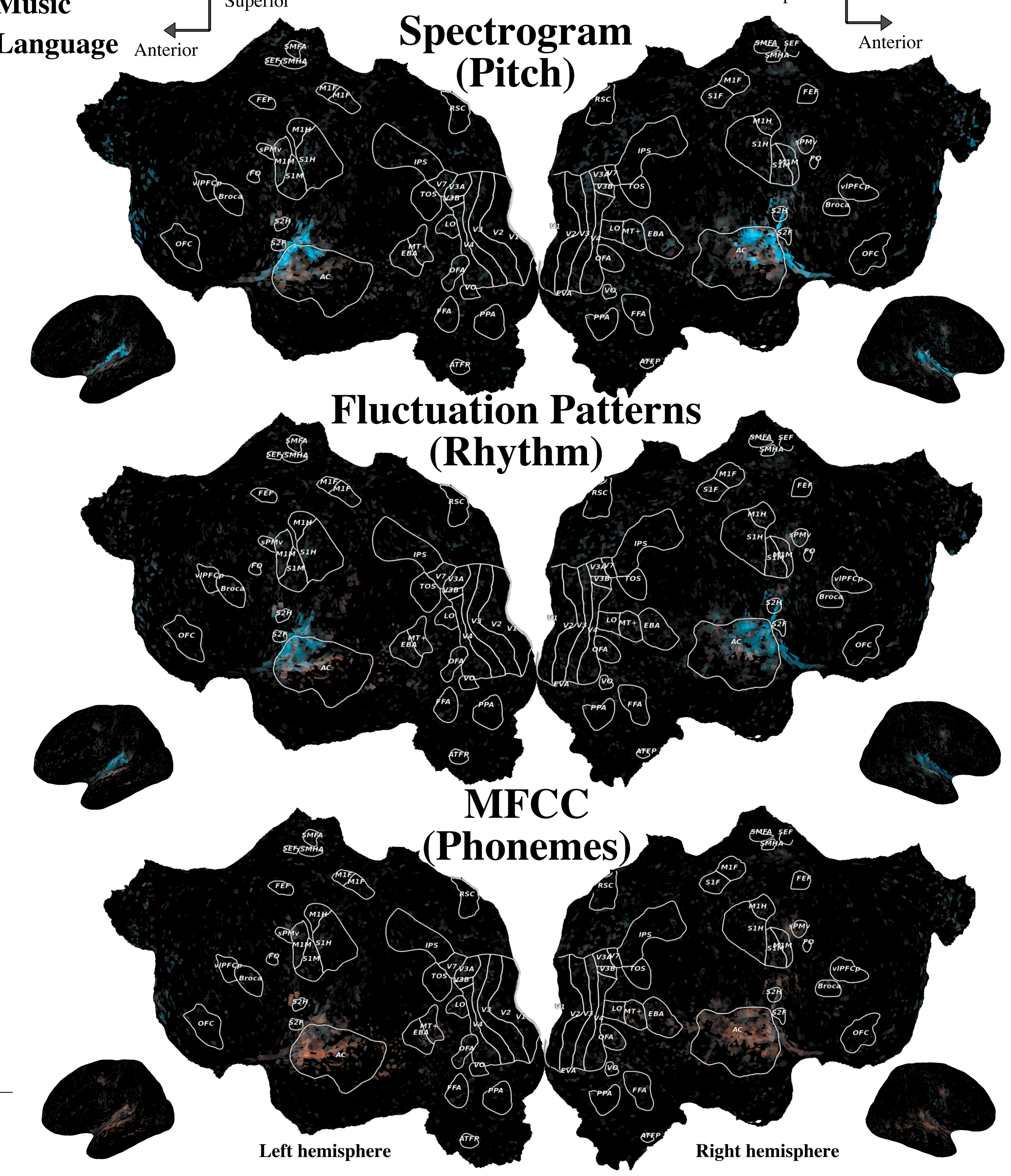


Average Prediction Correlation Value Per Feature

(as a fraction out of the full model's average prediction correlation value)



Prediction correlation flat maps for three acoustic features



Methods

Experimental Design

One human subject

Stimuli
~134 minutes of classical piano music
84 minutes of spoken stories

fMRI Parameters
TR: 2.0045 seconds
Voxel size: 2.24 x 2.24 x 3.5 mm^3

Acoustic Features

Loudness
Root mean square energy
Temporal flatness

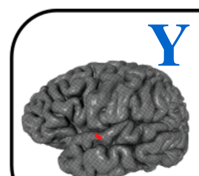
Tonal
Chromagram
Tonal centroid
Key and Mode

Rhythm
Fluctuation pattern
Fluctuation pattern centroid
Fluctuation pattern focus
Fluctuation pattern entropy

Spectral
Spectral centroid
Spectral spread
Spectral flatness
Spectral contrast
Spectral rolloff
MFCC
Melspectrogram
Spectrogram
Zero crossing rate

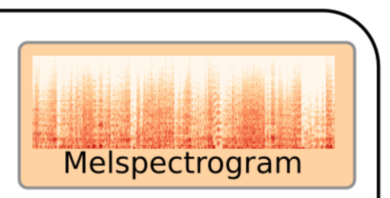
Training

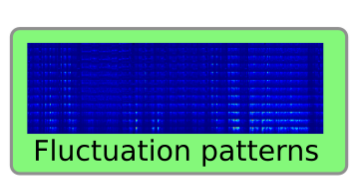
Validation



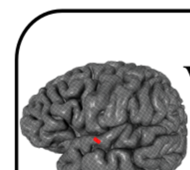
Y Training data

Stimulus feature spaces





F ⋮



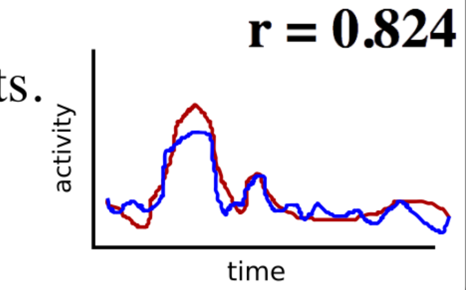
Validation data

W Estimate model weights

Ridge Regression Model Fit

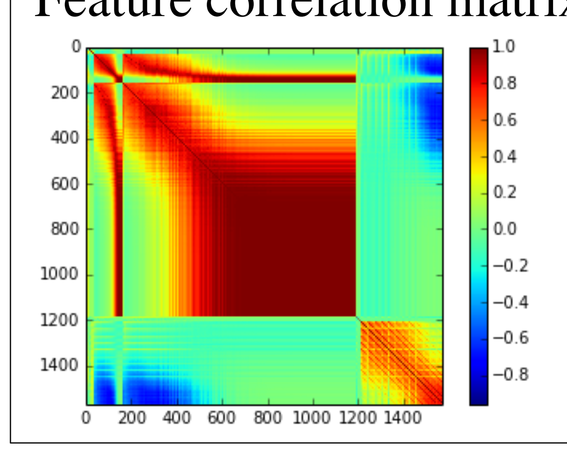
$$W = (F^T F + \lambda I)^{-1} F^T Y$$

Predict validation responses per voxel using the estimated model weights. Correlate with actual responses to verify models.



Future Directions

Feature correlation matrix



Future analysis will need to account for the covariance between features.

Different genres of music will need to be used as stimuli to validate these results.

References Acknowledgements

[1] Alluri et al., 2012. Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. NeuroImage 59.
[2] Alluri et al., 2013. From Vivaldi to Beatles and back: Predicting lateralized brain responses to music. NeuroImage 83.
[3] Toiviainen et al., 2014. Capturing the musical brain with Lasso: Dynamic decoding of musical features from fMRI data. NeuroImage 88.
[4] Leaver and Rauschecker., 2010. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J Neurosci. 30(22): 7604–7612

- SURF (funding)
- Searland Music Data (music stimuli)
- The Moth Radio Hour (speech stimuli)
- MIRToolbox and Librosa (open source MIR software used for benchmarking in-house software and for feature extraction)