Decoding single-trial frequency-following responses to speech stimuli using an animal model

Zhe-chen Guo¹, Jacie R. McHaney¹, Shengyue Xiong¹, Bharath Chandrasekaran¹, & Aravindkshan Parthasarathy²

¹Department of Communication Sciences and Disorders, Northwestern University
²Department of Communication Science and Disorders, University of Pittsburgh

1. Introduction

• Frequency-following responses to speech (speech FFRs) reflect not only stimulus features but also individual differences, providing a biomarker for listener identification [1–4].
• Affected by individual auditory experience.
• Less familiarity with the stimuli (e.g., Mandarin tones to English speakers) [5] and development of idiosyncratic FFR patterns over time [6] lead to better identification.
• It remains unclear how the decodability of individuals and the FFR features driving it are linked to neurophysiological changes, such as synaptopathy.
• Using an animal model (Mongolian gerbils) and a convolutional neural network (CNN) classifier, we examined:

1. Can individuals be decoded from single-trial FFRs?
2. What features are relevant and are they related to synapse count?
3. Do single-trial FFRs also differentiate age groups in a way that cannot be captured by traditional FFR metrics based on stimulus-relevant features?

2. Method

• Stimuli: 300-ms speech syllables (4 Mandarin tones x 4 talkers)
• 10 younger (15–40 weeks) and 9 middle-aged (75 weeks) gerbils
• Average immunostained cochlear synapse count < 8 kHz was calculated.
• FFRs were filtered between 70–1500 Hz, and converted to autocorrelograms (lags: 2–20 ms, window length: 40 ms, step size: 10 ms)
• ~640 trials per gerbil
• Trained a stack of 3 CNN layers to predict gerbil identity (or age group) from autocorrelograms of single-trial FFRs.
• Examined the features learned by the models using DeepLIFT [7] and correlated them with the synapse data.

3. Results

• Single-trial CNN models identified individual gerbils from single-trial FFRs with ~70% accuracy (baseline accuracy: 10–11%).
• Accuracy was significantly higher for middle-aged gerbils (79%) than for younger gerbils (68%) (permutted p = 0.033).
• Gerbils for which higher-frequency information was more relevant for identification tended to have a higher synapse count.

• The same models also accurately predicted age group (84% correct), with higher-frequency information being particularly relevant for the younger group.
• The age groups did not differ on any F0-based measures derived from average FFR and commonly used in the FFR literature [4, 8] (peak autocorrelation, F0 error, stimulus-response correlation).

• Single-trial speech FFRs contain robust biometric information for identifying individual subjects, consistent with previous research [1–4].
• The better identification of middle-aged gerbils may suggest less dedication to stimulus encoding and emergence of idiosyncratic FFR patterns over time [6].
• The findings further demonstrate that the patterns of relevant features in single-trial FFRs could reveal the underlying synaptopathy.
• The current decoding approach has the potential for clinical adoption.
• Tradition measures focusing on stimulus properties (e.g., F0 contours) are not sensitive to neurodevelopmental changes due to aging.
• Future work will examine the minimum number of trials required to reliably identify individuals and the applicability to human listeners.

4. Discussion


References