Machine Learning Classifier Distinguishes between Imagery of Speech and Non-Speech Sounds

Team: Julia Leeman[1], Joseph Zhang[2], Pooja Kabber[3], Kristi Van Meter[3]
Leads: Evan Hare[1], Ricardo Morales-Torres[1], Tobias Overath[1,4,5]
[1]Department of Psychology and Neuroscience, Duke University; [2]Department of Biomedical Engineering, Duke University;
[3]Master in Interdisciplinary Data Science, Duke University; [4]Duke Institute for Brain Sciences, Duke University;
[5]Center for Cognitive Neuroscience, Duke University

Duke Institute for Brain Sciences

Introduction

- Imagery is defined as "representations and the accompanying experience of sensory information without a direct stimulus"[1].
- There are similarities in the neural correlates of imagining and perceiving stimuli [2,3,4,5,6].
- Auditory imagery has been shown to encode perceptual information, including timbre [7], loudness [8,9], pitch [9,10], and melody [11].
- Human vocal sounds are processed differently from non-vocal environmental sounds in perception [12].
- **We hypothesized that auditory imagery of speech is represented by different neural correlates than that of non-human sounds.**

Methods

Participants

- 25 English-speaking participants, (mean age = 22.26, range: 18 - 43, 13 females)

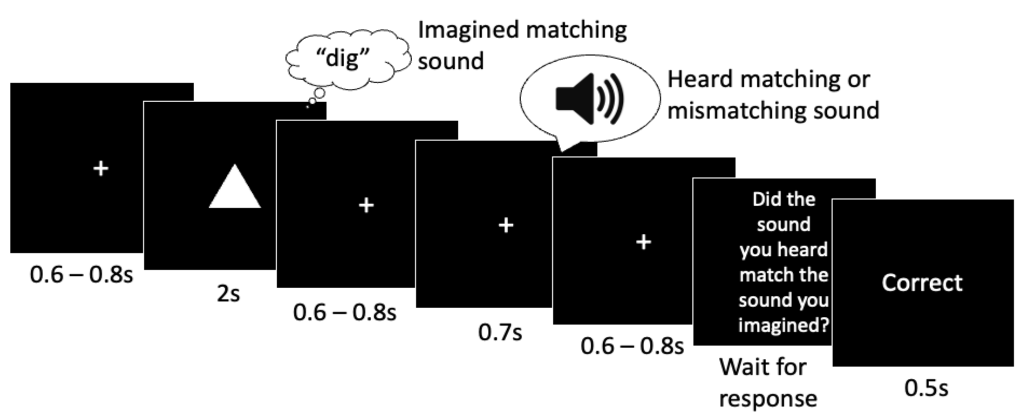
Stimuli

- Visual stimuli: square, circle, star, diamond, half-circle, triangle
- Speech sounds: English words "dig" and "cut"
- Artificial sounds: car horn, screenshot on an iPhone
- Animal vocalizations: chicken, frog

Design

- 3 Blocks of Training and Testing
- Participants learn to associate shapes and sounds. Then when presented with a shape, they imagine the associated sound.

Testing



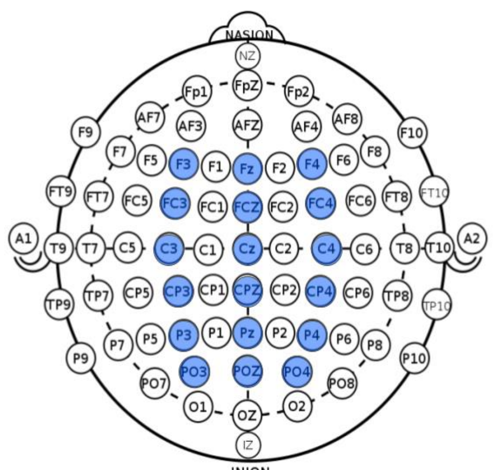
EEG Procedure

- Data was recorded during testing with a 64-channel BrainVision actiCAP EEG cap with a 10-20 montage at a sampling rate of 1,000 Hz
- Data were preprocessed and analyzed using custom MATLAB code and EEGLAB and FieldTrip toolboxes
- Preprocessing included re-referencing to the average of left and right mastoids, bandpass filtering from 0.1 to 50Hz, sparse interpolation of problem electrodes, independent component analysis, and epoching

Statistical Analysis

- ERP and time-frequency data were analyzed using FieldTrip statistics function using the Monte Carlo method and parametric statistical tests
- Data is also analyzed using ANOVA that uses category, stimuli, electrode laterality, and electrode anterior-posterior as factors, with subjects as random factor for our ROI
- Time-windows of N1 (50-150 ms), P200 (150-300 ms), LPC1 (350-500 ms), LPC2 (600-900 ms), and LPC3 (1200-1500 ms) are each tested

Region of Interest (ROI)



Machine Learning

- Classification for time series data using a subset of subjects (n = 22)
- Preprocessing – Downsampling (from 1000 Hz to 150 Hz) and extraction of ROI electrodes to dimensions N x 18 x t
- Compared two models (LSTM, EEGNet)
- Used grid search hyperparameter tuning
- Used k-folds (10-folds) cross validation to choose EEGNet as final model (also for limited sample size)

Results

Differences in ERPs for Perception Attenuated in Imagery

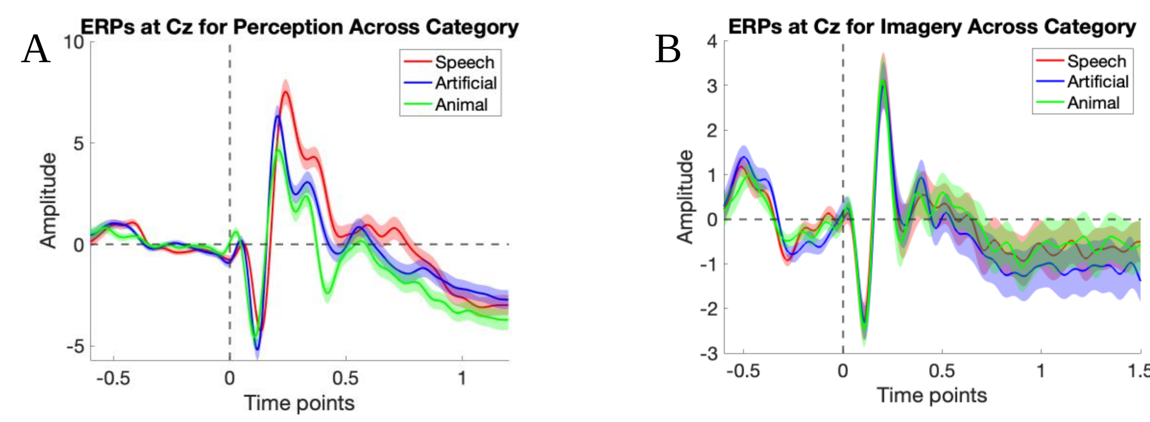


Figure 1. Grand average ERPs for each category at electrode Cz. **A,** ERPs for perception. ANOVA on perception: At P200, speech vs. animal and artificial vs. animal are significantly different. At LPC1 and LPC2, all pairs are significantly different. **B,** ERPs for imagery. ANOVA on perception: At P200, artificial vs. animal are significantly different. At LPC1, speech vs. animal and artificial vs. animal are significantly different.

Significant Clusters in Across Category ERP Comparison

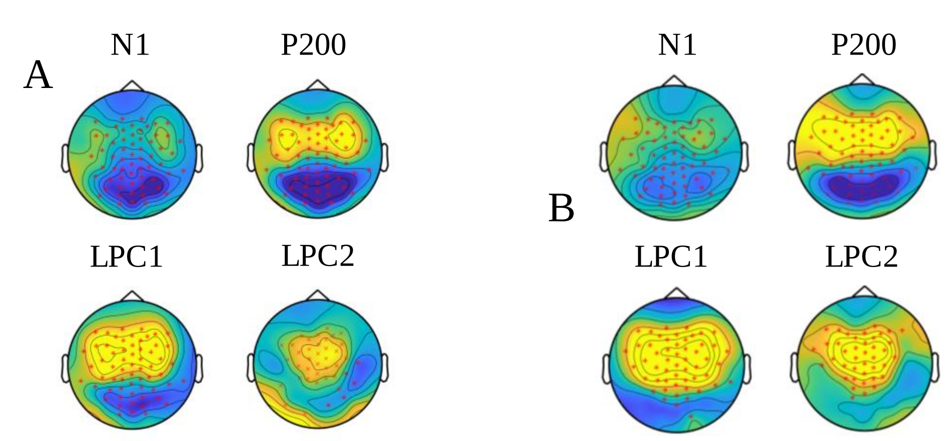


Figure 2. Cluster statistics for the difference between the grand average ERPs of the speech, artificial, and animal sounds for perception. Taken at time windows N1, P200, LPC1, and LPC2. **A,** Speech vs. artificial are significantly different for perception. **B,** Speech vs. animal are significantly different for perception. Same significant clusters when compared to ANOVA results. Legend: * for $p < 0.01$ and x for $p < 0.05$.

Significant Clusters in Across Category Time-Frequency Comparison for Perception and Imagery

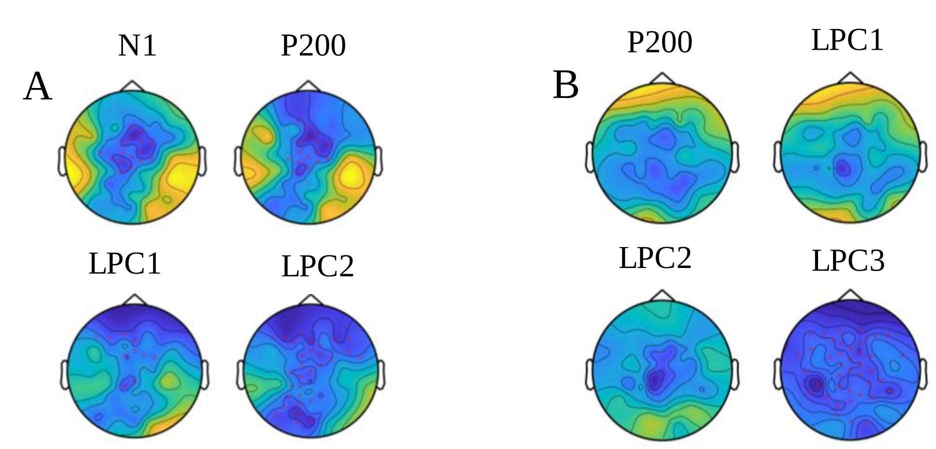


Figure 3. Cluster statistics for the difference between alpha power of the speech, artificial, and animal sounds for both perception and imagery, and at time windows N1, P200, LPC1, LPC2, and LPC3. **A,** Speech vs. animal are significantly different for perception. **B,** Speech vs. artificial are significantly different for imagery. Legend: * for $p < 0.01$ and x for $p < 0.05$.

Classification Using Machine Learning

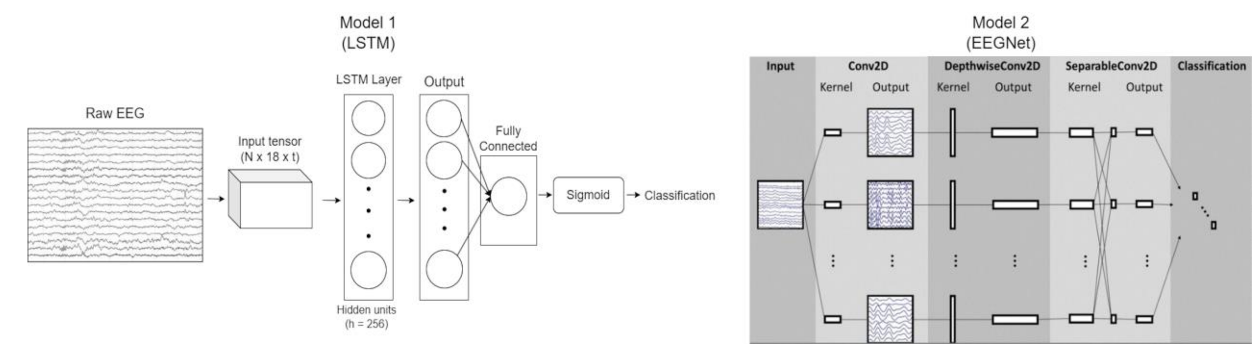


Figure 4. Model architectures for the LSTM-based model and EEGNet

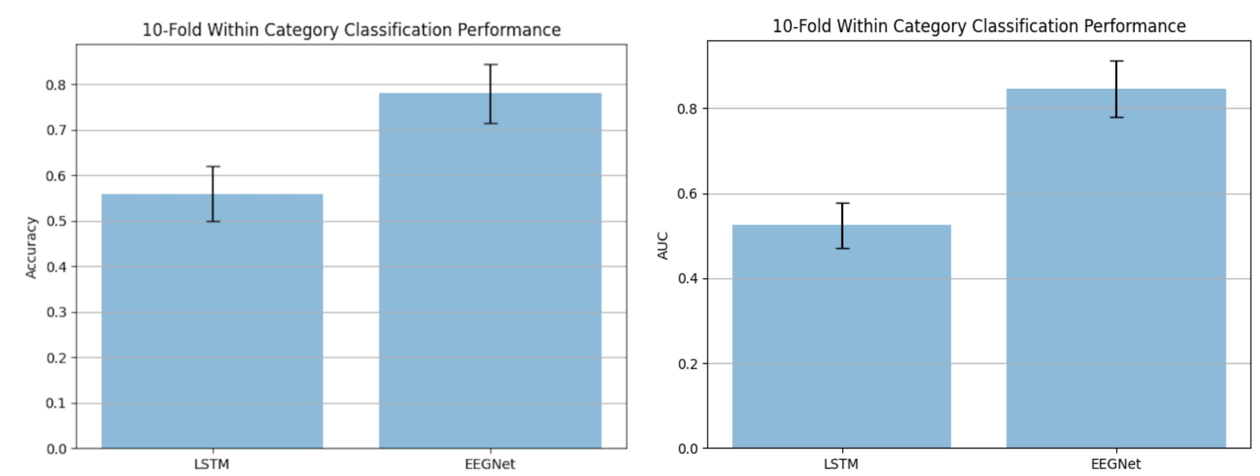


Figure 5. Performance metrics (Accuracy and AUC score) for all models using 10-folds measured within the categories (Speech, Animal, Artificial) with error bars for variance of scores across categories.

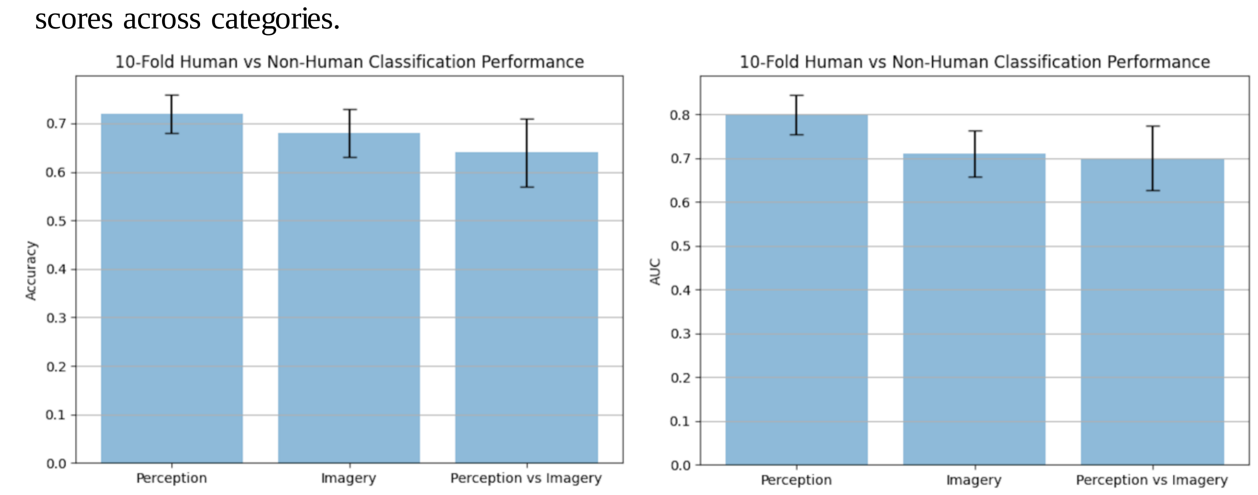


Figure 6. Performance metrics (Accuracy and AUC score) for EEGNet using 10-folds measured for human vs. non-human sounds for perception, imagery and across the two (train model on perception and test on imagery) with error bars for variance of scores across the 10 folds.

Conclusion

- Speech vs nonspeech ERP component differences are generally significant in auditory perception but less so in imagery.
- Time-frequency analysis shows some significant categorical alpha band differences in late time windows for both perception and imagery
- EEGNet machine learning classifier sorts individual ERPs evoked by speech vs nonspeech that are both perceived and imagined with accuracy significantly above baseline
- Categorical differences in imagined auditory perception are likely present in neural responses captured by EEG, but require a sensitive data-driven approach to examine

Future Directions

- Use DeepExplain package to extract sections of the neural time
- Explore categorical differences such as lexically meaningful speech vs spoken nonsense words
- Compare neural measures between participants with different levels of reported auditory imagery ability

References: [1] Pearson et al. (2015) *Trends in Cognitive Sciences*, **19**(10), 590–602. [2] Dijkstra et al. (2017) *Journal of Neuroscience*, **37**(5), 1367–1373. [3] Dijkstra et al. (2019) *Trends in Cognitive Sciences*, **23**(5), 423-434. [4] Gelding et al. (2019) *Scientific Reports*, **9**(1), 1-13. [5] McNorgan (2012) *Frontiers in Human Neuroscience*, 6. [6] Regev et al. (2021) *Cerebral Cortex*, **31**(8), 3622-3640. [7] Tuźnik et al. (2018) *International Journal of Psychophysiology*, **129**, 9-17. [8] Tian et al. (2018) *Nature Human Behaviour*, **2**(3), 225-234. [9] Wu et al. (2011) *Psychophysiology*, **48**(3), 415-419. [10] Halpern (2015) *Psychomusicology: Music, Mind, and Brain*, **25**(1), 37-47. [11] Zatorre et al. (2010) *Journal of Cognitive Neuroscience*, **22**(4), 775-789. [12] Belin et al. (2000) *Nature*, **403**(6767), 309-312.