

## **Phonemic category & higher-order acoustic features jointly drive neural response to speech**

Anna Mai<sup>1</sup>, Stéphanie Riès<sup>2,3</sup>, Sharon Ben-Haim<sup>4</sup>, Jerry Shih<sup>5</sup>, Timothy Gentner<sup>6,7,8</sup>

<sup>1</sup>UCSD, Linguistics, <sup>2</sup>SDSU, School of Speech, Language, and Hearing Sciences, <sup>3</sup>SDSU, Center for Clinical and Cognitive Neuroscience, <sup>4</sup>UCSD, Neurosurgery, <sup>5</sup>UCSD, Neurosciences, <sup>6</sup>UCSD, Psychology, <sup>7</sup>UCSD, Biological Sciences Division, Neurobiology Section, <sup>8</sup>UCSD, Kavli Institute for Brain and Mind

Using intracranial EEG recorded during a passive listening task, this study provides evidence that the brain abstracts phonemic category identity from an acoustically variable speech stream, and that it may do so in part using the covariance structure of the stimulus.

Intracranial EEG was recorded while ten participants listened to excerpts from the Buckeye Corpus (Pitt et al. 2007), a phonemically segmented and labeled corpus of American English conversational speech. Stimulus-timelocked broadband LFP (0.1-170Hz) and high gamma power (HGP: z-scored analytic amplitude of 70-150Hz bandpass LFP) were subsequently extracted from the recording.

Linear mixed-effects models were fit for each participant, modeling neural activity (HGP or LFP) with electrode channel and excerpt speaker as random effects. Fixed effects were either only spectrographic features, only phonemic labels, or both. Models were compared within response variable type (HGP or LFP) using the Akaike Information Criterion (AIC). All best-fit models carried 100% of the cumulative model weight and had an AIC score >200 lower than other models.

For broadband LFP, all participants' data were best fit by the model that included both spectrographic features and phonemic labels. These results demonstrate that broadband LFP contains phonemic category information that is not reducible to speech acoustics. For HGP, eight participants' data were best fit by the model that included only spectrographic features and two participants' data were best fit by the model that included both spectrographic features and phonemic labels. These results indicate that HGP is primarily driven by speech acoustics rather than phonemic category information. The variability in the best-fit model across participants may result from differential electrode coverage across participants and will take follow-up work to assess.

Based on these results, maximum noise entropy (MNE) models (Kaardal et al. 2017) were fit to assess what aspects of the stimulus drove the results. These MNE models are logistic functions of a linear combination of the first- and second-order features of the stimulus (the stimulus variance and covariance, respectively). First- and second-order models were fit to the HGP and broadband LFP of each participant for spectrographic stimuli that were either labeled or unlabeled for phonemic identity. For each channel, fit models were used to generate predicted neural responses. Pearson's r was calculated to assess the correlation of recorded vs. predicted responses and transformed using the Fisher Z-Transformation for comparison across conditions.

There was a statistically significant interaction between the effect of model order (first vs. second) and label status (labeled, unlabeled) on MNE model prediction quality, with  $p < 0.0001$  for all subjects, driven by second-order models fit using spectrograms labeled with phonemic identity. The fact that the inclusion of stimulus covariance structure improves model prediction for labeled data beyond the improvement for unlabeled data suggests that stimulus covariance structure and phonemic identity synergistically impact neural response, jointly providing some information not available in either feature alone.

Given previous work showing that the highest magnitude eigenvectors of speech covariance matrices are speaker-dependent (Zhou & Hansen 2005), this interaction may be driven by speaker-specific phonemic normalization. Future work will explore this possibility.