Neural Encoding of Phonemes Modulated by Linguistic Information

Seung-Goo Kim[1]*, Federico de Martino[2], Tobias Overath[1,3,4]

[1]Department of Psychology and Neuroscience, [3]Center for Cognitive Neuroscience, [4]Duke Institute for Brain Sciences, Duke University, Durham, NC, USA.

[2]Faculty of Psychology and Neuroscience, University of Maastricht, The Netherlands.

*seunggoo.kim@duke.edu

Introduction

Motivations

- Using a sound quilting algorithm [1] to control the temporal extent of speech structure, we showed that an acoustic analysis of temporal speech structure occurs in superior temporal sulcus (STS), while left inferior frontal gyrus (IFG) is engaged in mapping temporal speech structure to linguistic representations [2].
- Recently, linearized encoding analysis showed spectral and articulative features of natural speech in primary and non-primary auditory cortices [3].
- In the current study, we investigated how the phonetic information of natural speech encoded in the fMRI time-series is modulated by the temporal extent of speech structure and the linguistic context.

Methods

Participants

- Ten native English speakers (6 females) with normal hearing

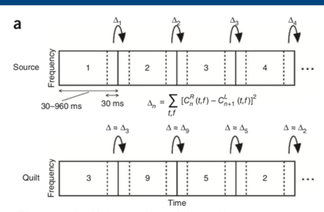

***Fig 1**. Quilting algorithm (from [1])*

Stimulus

- Stimulus: four bilingual female speakers reading books in English and Korean [2]
- Modified speech quilting algorithm [2]: quilting of phonemes (instead of fixed-duration segments [1])
- Design: 2x2 factorial; <English-or-Korean> x <Original-or-Phoneme Quilt>

Data Acquisition

- Multi-band echo-planar imaging: TR = 1.2 s, TE = 30 ms, MB = x3, vox = 2-mm-iso, 39 slices parallel to supratemporal planes over inferior frontal and temporal cortices
- Task: a button-press when the speaker changes from one to another (3-5 times during one trial)
- One trial = stimulus-time (33 s) + inter-trial-time (6 ± 4 s) + visual feedback (6 s)
- Multi-sessions: 8 runs x 3 sessions (except sub-01 and sub-08), total 205 min

Image Processing

- Volume-based: FSL5 (topup), SPM12 (slice timing correction, realignment)
- Surface-based: FreeSurfer ('fsaverage6', 82k; FWHM = 6 mm), GLMdenoise [4]

Finite Impulse Response (FIR) Modeling

- Features: overall cochleogram envelope and the durations of phoneme classes (manner of articulation), low-pass filtered and down-sampled to 1/TR (0.83 Hz)
- Lags: [0, 1, 2, .., 20] TRs (0, ..., 24 s)

Ridge Regression

- Optimization: ridge trace; the smallest λ such that all increments of β by an increment of λ to be smaller than 20% of initial λ; λ = from 10^0.5 to 10^11 [5]
- Cross-validation: 2-fold (odd/even-runs)
- Performance metric: Pearson correlation coefficient (not corrected for noise-ceiling)

Model Comparison

- Full model: phonemes and envelope in each condition separately (Phon+Env [L+Q+])
- Reduced models: models lacking some information (**Fig 3**).
- Interpretation: If r(A+B) >> r(A) in a voxel, B is markedly encoded in the voxel so that the prediction improves. If r(A+B) << r(A), B is not encoded in the voxel.
- Group-level inference based on permutations with replacement (n = 10; k = 2,000)

Model Performances

Collinearity of predictors

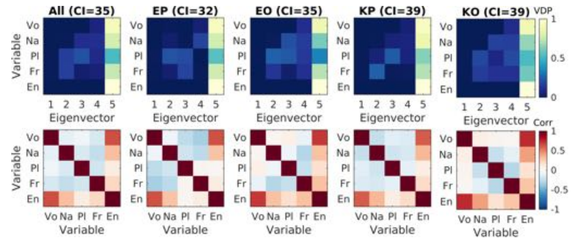

***Fig 2**. Variation Decomposition Proportion (upper) and Pearson correlation (lower)*

Models

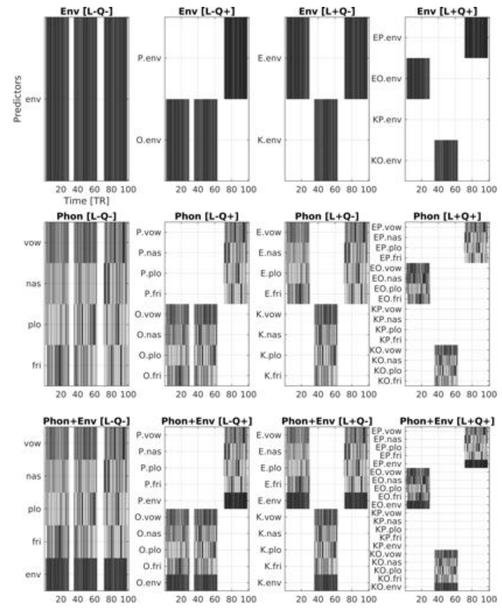

***Fig 3**. Examples of predictors. **Phon**emes (**vow**el, **nas**al, **plo**sive, **fri**cative) and **env**elope in different conditions are modeled either all together (**L-Q-**), regardless of language (**L-Q+**), regardless of quilting (**L+Q-**), or all separately (**L+Q+**)*

Prediction Accuracies

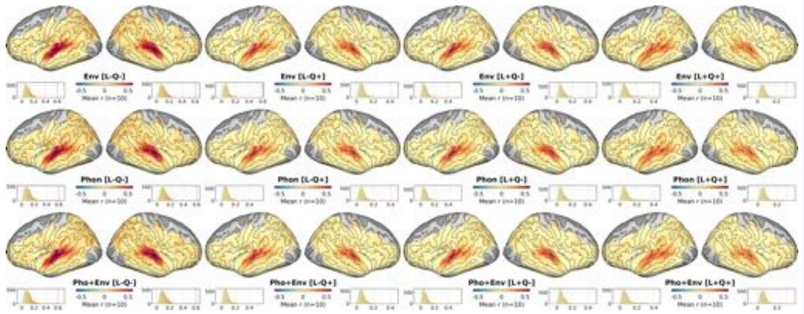

***Fig 4**. Group-averaged Pearson correlation for each model.*

Model Comparisons

Effect of Envelope

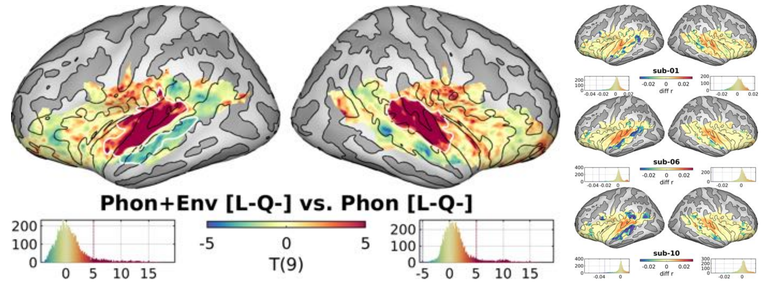

***Fig 5.** T-statistic map comparing prediction accuracies (r) from models with and without Envelope. White contours mark cluster-level p < 0.05. Positive clusters (red) indicate areas where the addition of Envelope in the model improved the prediction of fMRI time-series whereas negative clusters (blue) indicate areas where the addition of Envelope worsen the prediction. Individual differences in prediction accuracy (r) in selected subjects are also shown (right).*

Effect of Phoneme

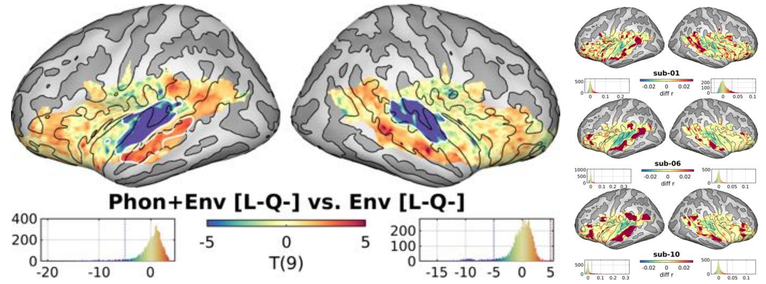

***Fig 6.** T-statistic map comparing models with and without Phonemes. White contours mark cluster-level p < 0.05. Individual r differences in selected subjects are shown (right).*

Effects of Linguistic and Temporal Contexts

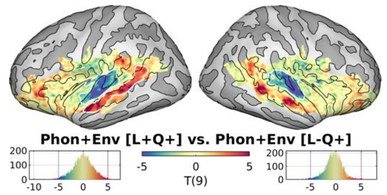

***Fig 7.** T-statistic map comparing models encoding all conditions and all but language*

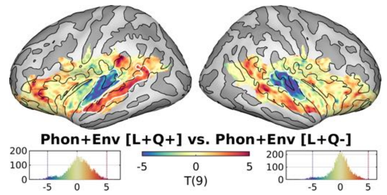

***Fig 8.** T-statistic map comparing models encoding all conditions and all but quilts*

Conclusions

- As shown previously [3], acoustic features (e.g. envelope) are encoded in primary auditory areas (e.g., HG, PT), whereas phonetic features are encoded in non-primary auditory areas (e.g., STS).
- The temporal extent of speech structure and linguistic context seem to modulate the encoding of these features in non-primary areas (e.g., STS) [1,2].

References

[1] Overath et al. (2015). Nature Neurosci. [2] Overath & Paik. (under review). [3] de Heer, Huth et al. (2017). J. Neurosci. [4] Kay et al. (2013). Front. Neurosci. [5] Santoro et al., 2014, PLoS Comput Biol