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

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>

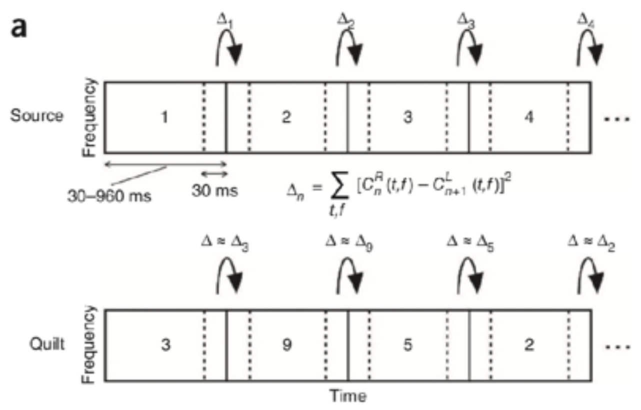

Fig 1. Quilting algorithm (from [1])

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

Finite Impulse Response (FIR) Modeling

- Responses: surface projected ('fsaverage6', 82k; FWHM = 6 mm) and GLMdenoise [4]
- 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 between full vs. reduced models at group-level (2^10 permutations)

Conclusions

- 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).
- Acoustic and linguistic contexts seem to modulate the encoding of phonemes in non-primary areas (e.g., STS and IFG) [1,2].
- Distinctive encoding patterns are found across phonemes being modulated by contexts.

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

Model Comparison

Effect of Envelope and Phonemes

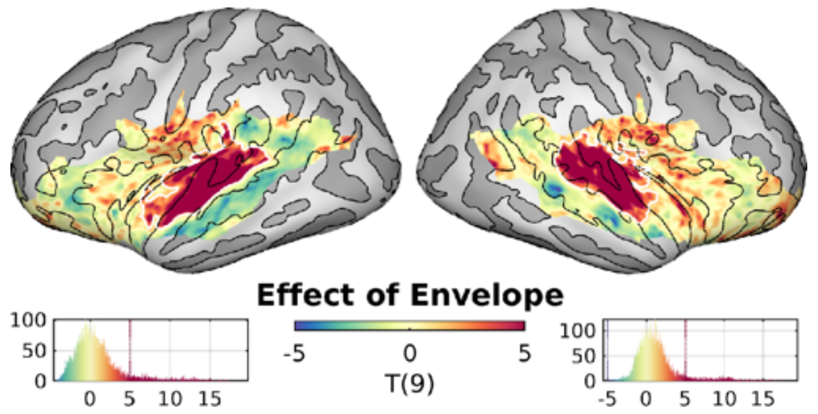

Fig 2. T-statistic map comparing prediction accuracies (r) from models with and without Envelope. White contours mark cluster-P < 0.05.

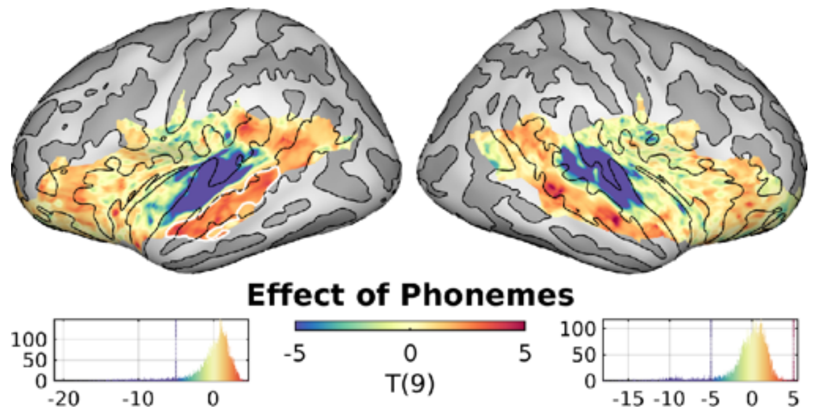

Fig 3. T-statistic map comparing prediction accuracies (r) from models with and without Phonemes. White contours mark cluster-P < 0.05.

Effects of Linguistic and Acoustic Contexts

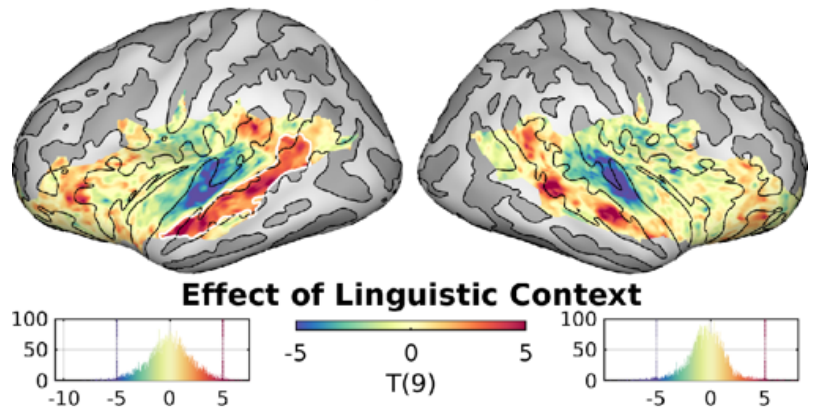

Fig 4. T-statistic map comparing models encoding all conditions and all but language.

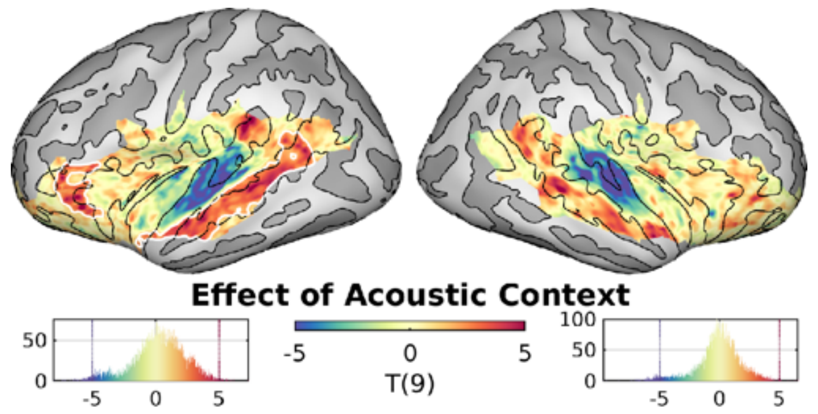

Fig 5. T-statistic map comparing models encoding all conditions and all but quilts.

Interaction Between Linguistic and Acoustic Contexts

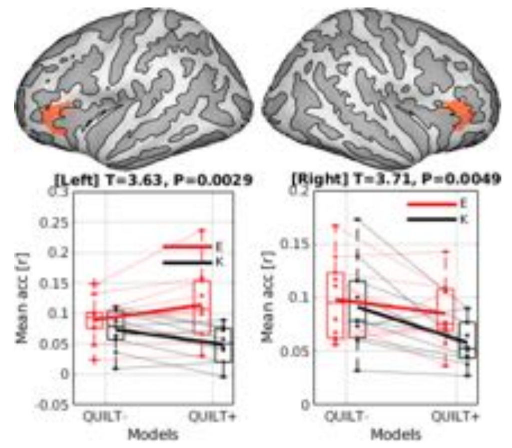

Fig 6. Region-averaged interaction in the inferior frontal gyrus (pars triangularis and pars opercularis)

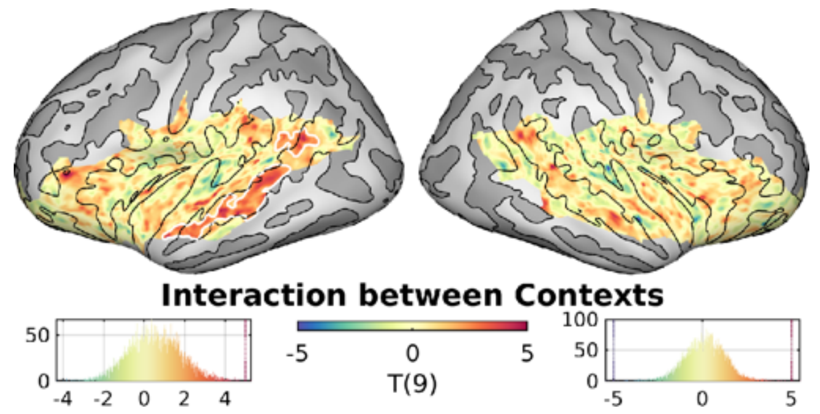

Fig 7. T-statistic map comparing prediction accuracy increases by incorporating quilts [e.g., Fig 5] in English conditions than in Korean conditions (English > Korean).

Effects of Individual Phoneme Classes

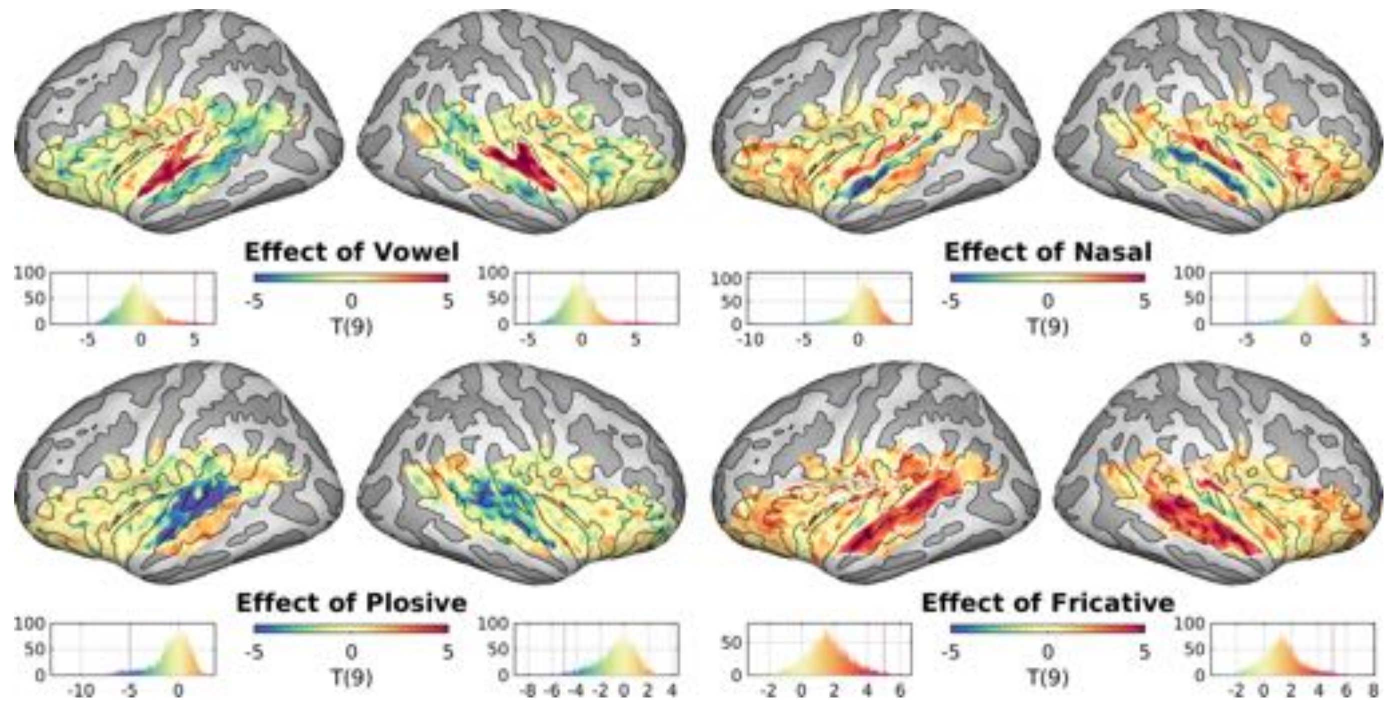

Fig 8. T-statistic maps for uniquely explained variance by individual phoneme classes.

Phoneme-Class Encoding

Encoding of Individual Phoneme Classes

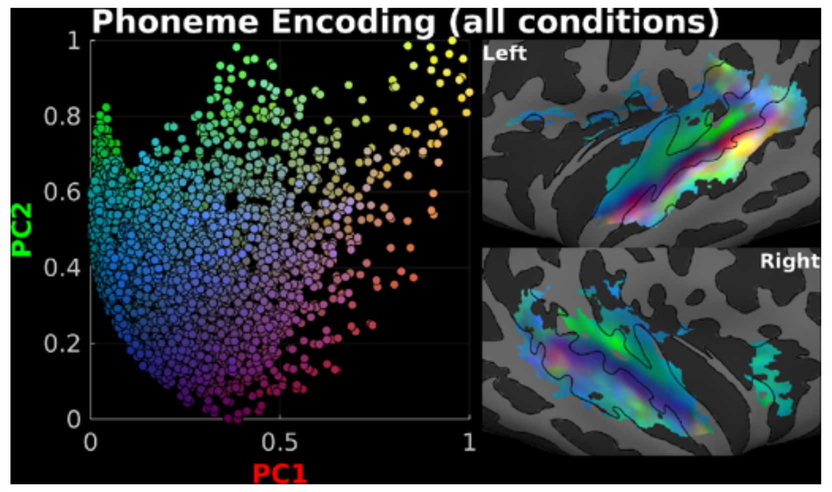

Fig 9. Projection of phoneme encoding patterns [V,N,P,F] to the first three principal components and its RGB visualization shown in the PC1-PC2 plane (left) and the anatomical space (right).

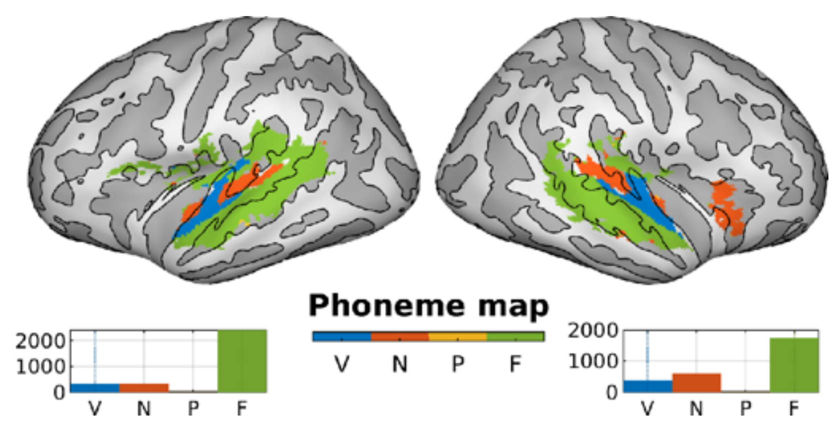

Fig 10. Winner-take-all map of phoneme encoding. cluster-P < 0.05 in any phoneme class. V, vowel; N, nasal; P, plosive; F, fricative.

Interaction of Contexts on Phoneme Encoding

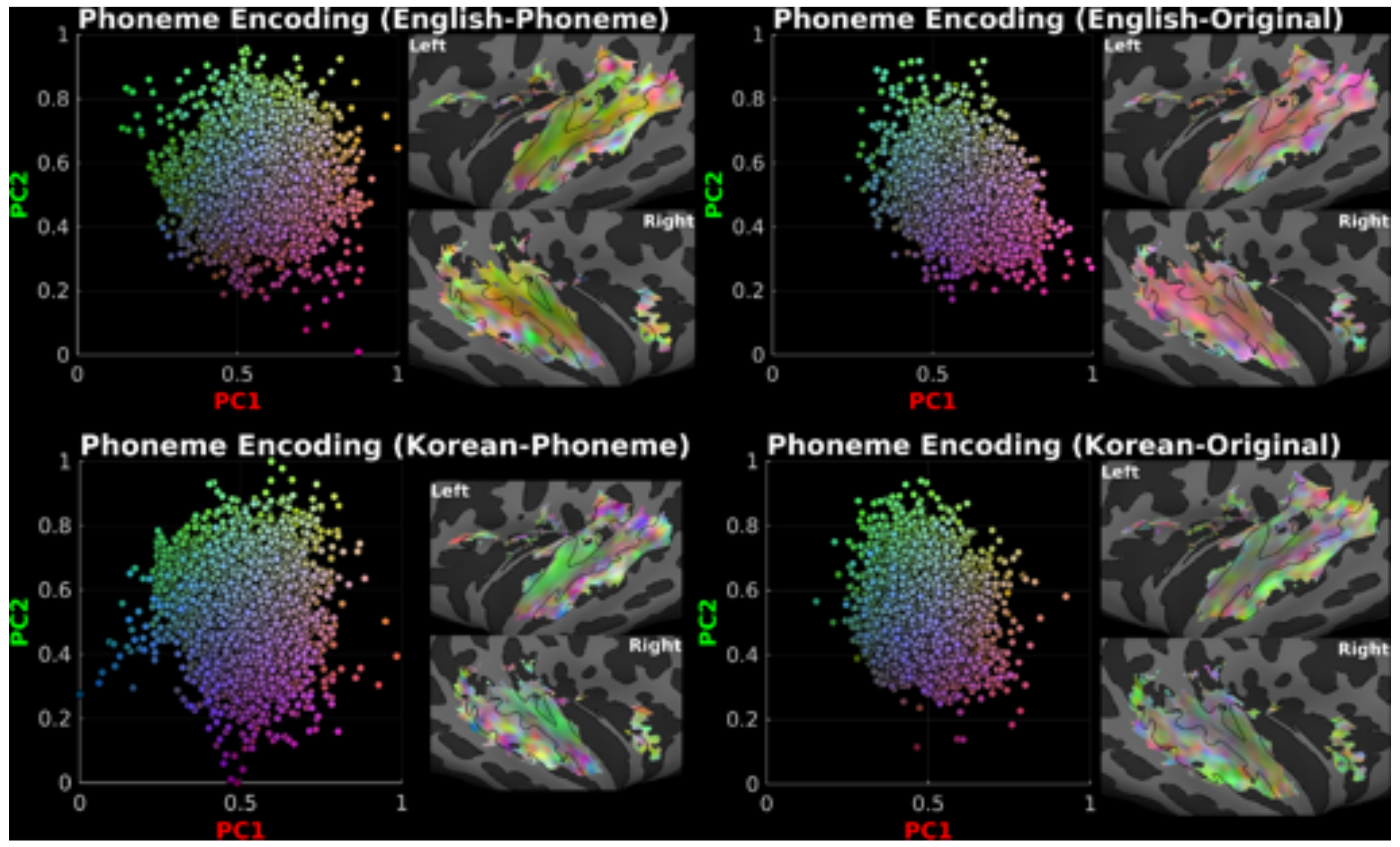

Fig 11. Projection of phoneme encoding patterns [V,N,P,F] to the first three principal components and its RGB visualization shown in the PC1-PC2 plane and the anatomical space in four conditions.

Multivariate Distance of Phoneme Encoding Between Languages

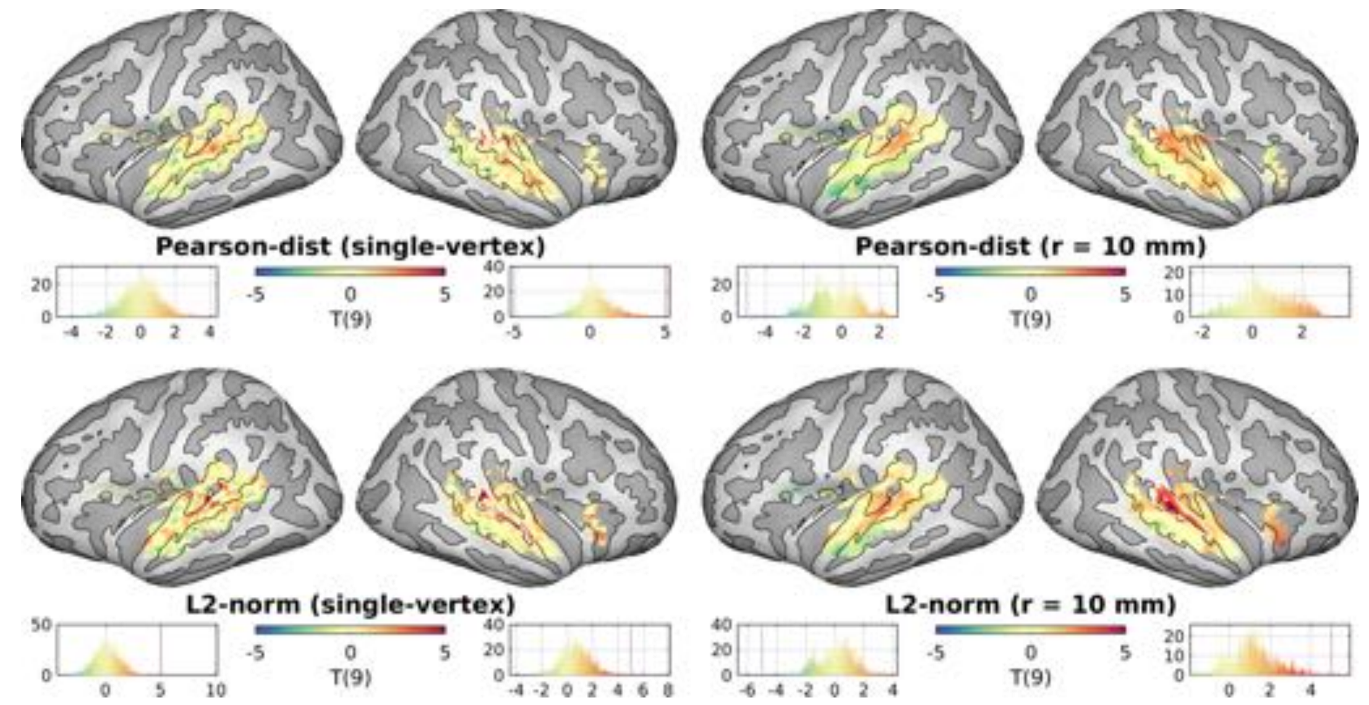

Fig 12. T-statistic maps testing a distance being greater in English than Korean.