Phonological contrasts are maintained despite neutralization: an intracranial EEG study

Anna Mai[1], Stéphanie Riès[2,3], Sharona Ben-Haim[4], Jerry Shih[5], Timothy Gentner[6,7,8]

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

The existence of language-specific abstract sound-structure units (such as the phoneme) is largely uncontroversial in phonology. However, whether the brain performs abstractions comparable to those assumed in phonology has been difficult to ascertain. Using intracranial electroencephalography (EEG) recorded during a passive listening task, this study takes advantage of an English neutralization pattern to provide evidence that the brain abstracts phonemic category identity in a structured, language-specific way from contextually conditioned acoustic variants of phonemes (allophones).

Intracranial EEG was recorded while ten participants listened to excerpts from the Buckeye Corpus (Pitt et al. 2007), a phonetically segmented and labeled corpus of American English conversational speech. Subsequently, epochs of high gamma power (HGP: z-scored analytic amplitude of 70-150Hz bandpass LFP) timelocked to the occurrence of voiceless coronal stops and coronal taps were excised from the neural recording and phonemically labeled as /d/ or /t/.

English /d/ and /t/ have many acoustically distinct allophones, but when either /d/ or /t/ occurs following a stressed syllable and between two vowels, their acoustic contrast is neutralized, and both are pronounced as a coronal tap (e.g. *writing, riding*). Given that auditory processing primarily proceeds in a feedforward manner from spectrotemporal features of the sensory input, it is anticipated that many speech selective sites will demonstrate an 'acoustic response', where the responses for all taps are more similar to one another than to the voiceless coronal stop allophone of /t/. However, if it is the case that phonemic identity is computed during language processing, even when the acoustic contrast between two phonemes is neutralized, then there also should exist sites demonstrating a 'phonemic response', where the neural response to underlyingly /t/ taps (**dx_t**; i.e., *writing*) is more similar to other allophones of /t/ (**t**; i.e., voiceless alveolar stops) than to underlyingly /d/ taps (**dx_d**; i.e., *riding*).

To test this, neural epochs were defined for all phones as the 500ms following phone onset with a 100ms baseline. Speech selective electrodes were then identified as those for which a sliding-window one-way ANOVA indicated that there was a significant difference in the electrode's mean response to a random selection of speech vs. silence epochs. Across the ten patients, 92 electrodes were selective for speech. Of those, a sliding-window one-way ANOVA comparing the HGP of **t**, **dx_t**, and **dx_d** epochs was significant for fifteen electrodes—four exhibiting a phonemic response, and eleven exhibiting an acoustic response.

To assess the likelihood of observing these numbers of acoustic and phonemic response sites by chance, a null distribution was generated by performing this statistical analysis for 3,500 arbitrary pairs of phones (i.e., A, B) with an arbitrary split of one phone (i.e., A, B_x, B_y). From this distribution, the probability of observing at least four phonemic sites and at least eleven acoustic sites was <1%. Figure 1 shows the null distribution of phonemic and acoustic sites, with a dotted white

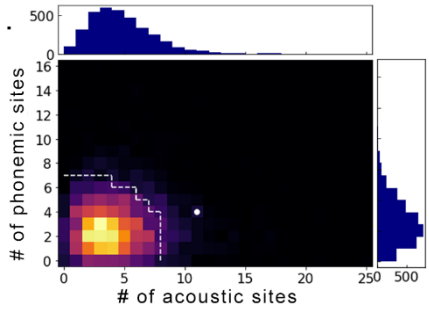

Figure 1

line marking the bound for 95% of the distribution's mass and a single white point indicating the observed number of phonemic and acoustic sites.

Sites of phonetic response tended to be temporal. In particular, one site in left posterior superior temporal gyrus (STG) exhibited a significant difference between alveolar stop /t/ and phonemically /t/ taps 150-250ms after phoneme onset and a lack of difference between taps (Figure 2a,c). This result suggests that this site tracks acoustic rather than phonemic identity, consistent with results reported in Mesgarani et al. (2014), Hullet et al. (2016), and elsewhere concerning STG's sensitivity to complex acoustic features during natural speech processing.

Sites of phonemic response tended to be frontal. In particular, one site in left cingulate cortex exhibited a significant difference between phonemically /d/ taps and both phonemically /t/ sounds 250-350ms following phone onset (Figure 2b,d). This activity suggests that phoneme-like abstraction occurs during passive listening to natural speech in frontal areas. Although lesion studies suggest that frontal areas are not the main substrate for speech perception (see Hickok & Poeppel 2007), this result may be consistent with work showing engagement of frontal areas during phonological processing (Burton 2001, Xie & Myers 2018). However, this result will require replication with MEG or intracranial patients with broader frontal coverage to fully substantiate.

From these results, we conclude that both acoustic and phonological contrasts are maintained during passive listening to natural speech. Most significantly, however, in showing that phonological contrasts are maintained in the absence of acoustic contrast, this work makes a substantial step towards understanding how units of analysis in phonological theory map onto the physiology of the brain. Future work with this dataset will assess the role of acoustic variability in the trial-by-trial neural response and will estimate spectrotemporal receptive fields for these speech-selective sites.

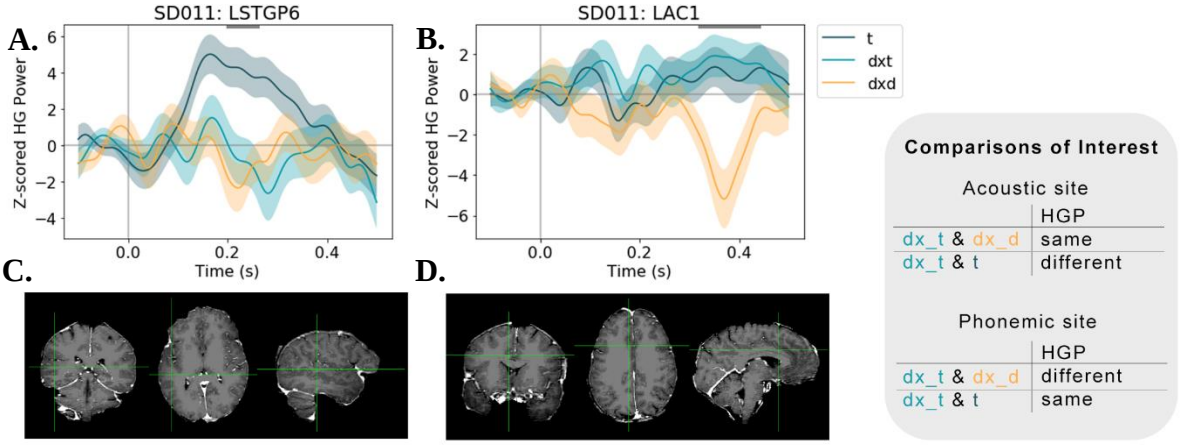

Figure 2

Works Cited

Burton, M. W. (2001). The role of inferior frontal cortex in phonological processing. *Cognitive Science*, 25(5), 695-709. **Hickok**, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature reviews neuroscience*, 8(5), 393-402. **Hullett**, P. W., Hamilton, L. S., Mesgarani, N., Schreiner, C. E., & Chang, E. F. (2016). Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli. *Journal of Neuroscience*, 36(6), 2014-2026. **Mesgarani**, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, 343(6174), 1006-1010. **Pitt**, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. and Fosler-Lussier, E. (2007) Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor). **Xie**, X., & Myers, E. (2018). Left inferior frontal gyrus sensitivity to phonetic competition in receptive language processing: A comparison of clear and conversational speech. *Journal of cognitive neuroscience*, 30(3), 267-280.