# Spatiotemporal tracking of phonetic content and probability in the human brain during continuous speech understanding

**William Turner (wft@stanford.edu)**
Department of Psychology, Stanford University

**Josef Parvizi (jparvizi@stanford.edu)**
Department of Neurology, Stanford University

**Laura Gwilliams (laura.gwilliams@stanford.edu)**
Department of Psychology, Stanford University

## Abstract

**To understand speech, the human brain integrates both the content and probability of what is being said. Whether content and probability are encoded in the same neural populations at the meso-scale, however, remains unknown. Here, we leverage the exceptional spatiotemporal precision of intracranial electroencephalography (iEEG), to track the neural encoding of phonetic features and phonetic surprisal (i.e. phoneme-level probabilities), during continuous speech processing. We identify neural populations that jointly encode phonetic information and phonetic surprisal in the superior temporal lobe. By contrast, we find that lexical surprisal (i.e. word-level probabilities) is encoded by adjacent but distinct populations. Overall, our findings have mechanistic implications for how content and probability are neurally integrated in the temporal lobe, to give rise to robust speech understanding.**

**Keywords:** speech processing; prediction; iEEG

## Introduction

Understanding spoken language typically feels effortless and automatic, despite the prevailing noise and ambiguity in the acoustic signal. Broadly, the human brain achieves this by integrating the *content* of speech input with the *probability* of that input occurring (Gwilliams & Davis, 2022; Jurafsky, 2003).

At the level of speech sounds, neural populations in the superior temporal gyrus (STG) have been associated with the encoding of phonetic feature *content* (Mesgarani et al., 2014; DiLiberto et al., 2015). Independent studies have also associated STG with the encoding of phoneme *probability* (Gagnepain, Henson, & Davis, 2012; Heilbron et al., 2022). Thus, at the macro-scale, as revealed by non-invasive methods such as fMRI and MEG, phonetic *content* and *probability* spatially overlap. What remains unknown, however, is whether at the "zoomed in" meso-scale, afforded by invasive electroencephalographic (iEEG) recordings, neural populations jointly encode phonetic content and phonetic probability.

Here, we leverage the exceptional spatio-temporal resolution of intracranial electro-encephalography (iEEG), to map populations responsive to phonetic information, as well as violations of probabilistic expectations ('prediction errors') in the brains of participants listening to continuous speech. We compute probabilistic expectations at two levels of abstraction: surprisal of phonemes in words ('phoneme surprisal'); surprisal of words in sentences ('lexical surprisal'). First, we identify neural populations sensitive to phonetic features (Mesgarani et al., 2014). Second, we identify neural populations sensitive to the contextual unexpectedness of phonemes and words. We examine spatial overlap and find evidence of joint encoding of phoneme content and phoneme surprisal, contrasting with lexical surprisal which was encoded in adjacent but partially dissociable populations.

## Method

Eleven patients with focal epilepsy were recruited during invasive electrophysiologic monitoring at the Stanford Medical Center (6 Female, Age: $M$ = 35.10, $SD$ = 8.58). Participants listened to 1hr of audiobook snippets. Stimuli were annotated for 14 binary phonetic features, in 3 feature families: voicing, manner of articulation, and place of articulation (Gwilliams et al. 2022). Logistic regression was used to predict binary encodings of each feature from the time course of activity at single electrodes (within stimulus, 5-fold cross validation).

We derived information theoretic measures of the unexpectedness (surprisal) of each phoneme and word. Phonetic surprisal was calculated from the contextual probability of each phoneme within words (Gwilliams et al., 2022). Lexical surprisal was calculated from GPT-2 derived word probability estimates (Heilbron et al., 2022). Ridge regression was used to decode surprisal estimates at each electrode.

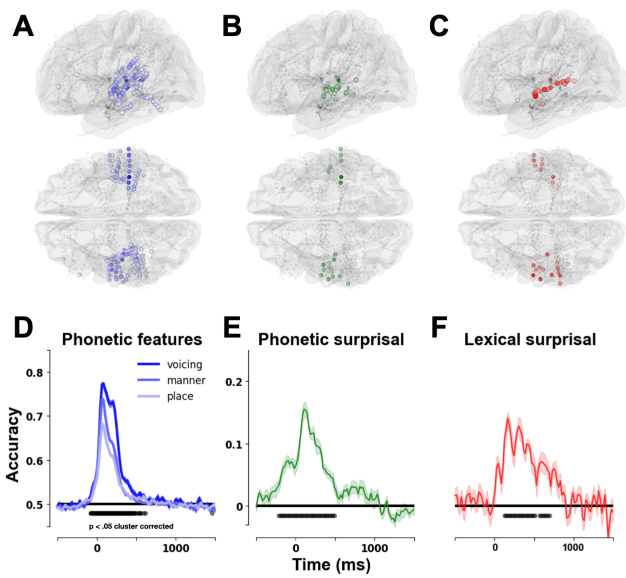


**Figure 1. Spatiotemporal mapping results.** Electrodes sensitive to **A)** phonetic information, **B)** phonetic surprisal, and **C)** lexical surprisal. Panels **D-F)** show decoding time courses for each electrode subset.

## Results

Averaging over phonetic features, we identified 163 electrodes sensitive to phonetic information (one sample t-tests on accuracy across stimuli, $p < .05$ fdr-corrected). These were predominantly located in superior temporal lobe, with a left hemispheric bias (Figure 1A). We identified 25 electrodes sensitive to phonetic surprisal, and 27 electrodes sensitive to lexical surprisal, again predominantly in superior temporal lobe (Figure 1B-C).

Decoding time-courses locked to phoneme onset (Figure 1D-F) revealed sustained representation of each linguistic feature for >500ms (early decoding in D-E due to coarticulation). Examining population overlap (Figure 2), we find that electrodes sensitive to phonetic surprisal are also sensitive to phonetic features. However, electrodes sensitive to lexical surprisal appear qualitatively more dissociable.

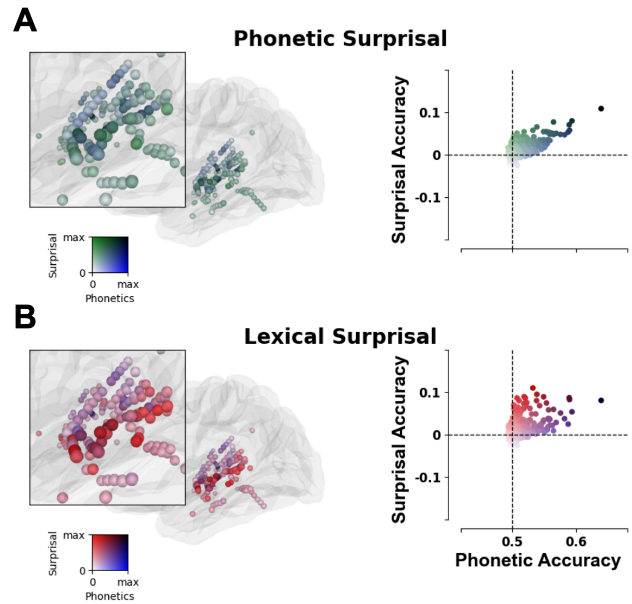


**Figure 2. Probing population overlap. A)** Electrodes sensitive to phonetic surprisal are sensitive to phonetics (each point shows joint accuracy at single electrodes). **B)** Qualitative dissociations between lexical surprisal and phonetics, with a less clustered joint accuracy profile.

## Discussion

We have examined the neural encoding of the content and probability of phonetic inputs during continuous speech processing. We found that neural populations are jointly sensitive to phonetic features and phonetic surprisal, in superior temporal lobe. In contrast, we found evidence that lexical surprisal, which tracks probabilities at the word level, is encoded in adjacent but partly dissociable populations. Our findings suggest that both the content and probability of phonetic input are co-localised at the meso-scale, with lexical probabilities showing distinct encoding. Given the view that high-level linguistic predictions are broadly distributed (Heilbron et al., 2022), it is also striking that we find no evidence of populations sensitive to lexical surprisal in frontal regions. This is despite adopting a relatively lenient criterion for classifying electrodes as surprisal responsive, as potential acoustic correlates have yet to be controlled. Instead, we find that lexical surprisal is encoded predominantly in temporal cortex, consistent with the view that local recurrent computations may facilitate more complex forms of linguistic computation than traditionally thought (Yi, Leonard, & Chang, 2019).

## References


Di Liberto, G. M., O'sullivan, J. A., & Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, *25*(19), 2457-2465.

Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Current Biology*, *22*(7), 615-621.

Gwilliams, L., & Davis, M. H. (2022). Extracting language content from speech sounds: the information theoretic approach. In *Speech perception* (pp. 113-139). Cham: Springer International Publishing.

Gwilliams, L., King, J. R., Marantz, A., & Poeppel, D. (2022). Neural dynamics of phoneme sequences reveal position-invariant code for content and order. *Nature Communications*, 13(1), 6606.

Heilbron, M., Armeni, K., Schoffelen, J. M., Hagoort, P., & De Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National Academy of Sciences*, 119(32), e2201968119.

Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. *Probabilistic linguistics*, *21*, 1-30.

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, 343(6174), 1006-1010

Yi, H. G., Leonard, M. K., & Chang, E. F. (2019). The encoding of speech sounds in the superior temporal gyrus. *Neuron*, 102(6), 1096-1110.