Neural encoding of lexical stress in human speech cortex

Ilina Bhaya-Grossman (ilina.bhaya-grossman@ucsf.edu)
University of California, San Francisco, USA

Yulia Oganian (yulia.oganian@uni-tuebingen.de)
University of Tuebingen, Germany

Emily Grabowski (emily_grabowski@berkeley.edu)
University of California, Berkeley, USA

Edward F. Chang (Edward.Chang@ucsf.edu)
University of California, San Francisco, USA

Abstract

Lexical stress – what distinguishes the noun "PRE-sent" from the verb "pre-SENT" – critically facilitates word recognition and comprehension in speech perception. To understand the neural mechanisms that enable the perception of lexical stress, we collected high-density intracranial electrocorticography recordings (ECoG) while ten English speaking participants performed two experiments: 1) passively listening to sentences with natural stress patterns and 2) actively identifying the stressed syllables in isolated spoken words. In Experiment 1, we identified neural populations that significantly encoded whether a syllable was stressed in natural speech. In Experiment 2, we found that stress-encoding neural populations were both sensitive to prior syllable contexts and categorical. Our findings characterize the distinct neural populations that process lexical stress, providing insight into the complex neural mechanisms that underlie this fundamental linguistic skill.

Keywords: speech perception; neural coding; electrocorticography (ECoG)

Introduction

Lexical stress—the perceived emphasis on certain syllables—plays a pivotal role in how we recognize words and respond to incoming speech (Cutler & Norris, 1988). In English, stressed syllables often contain longer vowels, have higher pitch, and have greater speech intensity *relative* to all other syllables in the word (Cutler, 2008; Fry, 1955; Mousikou et al., 2024). Therefore, accurately identifying which syllable in a word is stressed requires listeners to extract, normalize, and integrate multiple acoustic cues into a unified, categorical percept.

Neuroimaging studies have implicated the human temporal lobe, particularly the superior temporal gyrus (STG), in lexical stress identification tasks (Aleman et al., 2005; Domahs et al., 2013; Schwab et al., 2023). However, beyond identifying the cortical regions involved, it remains unclear which specific neural mechanisms within the STG enable listeners to extract and represent lexical stress as a multi-cue, relative, and categorical linguistic feature. Recent high-density ECoG studies have shown that neural populations in the STG separately encode the acoustic cues associated with lexical stress, including speech intensity (Oganian & Chang, 2019), vowel features (Oganian et al., 2023), and pitch (Tang et al., 2017). To our knowledge, no studies have yet leveraged ECoG to directly investigate how neural populations in the human temporal cortex transform representations of distinct acoustic cues into an integrated linguistic feature, lexical stress.

Results

To address this question, we conducted high-density ECoG recordings with ten participants undergoing clinical monitoring for epilepsy while they performed two lexical stress perception tasks. The millimeter and millisecond resolution of direct cortical surface ECoG enabled us to ask highly specific questions about how transient speech features relevant for lexical stress detection are encoded and transformed by the human temporal cortex.

Experiment 1. In Experiment 1, we tested whether neural activity recorded from single electrodes on the human temporal lobe encoded lexical stress information while participants passively listened to naturally-produced spoken sentences (Garofolo et al., 1993). Using temporal receptive field modelling (TRF) (Holdgraf et al., 2017), we identified electrodes primarily located in the human STG that significantly encoded the stress category of syllables (primary vs. unstressed), beyond the specific acoustic cues to stress such as speech intensity, pitch, and vowel duration or quality (example electrode shown in Fig. 1A). These STG electrodes did not exclusively encode lexical stress, but in some cases, also encoded the acoustic cues associated with stressed syllables more generally (Fig. 1B-D).

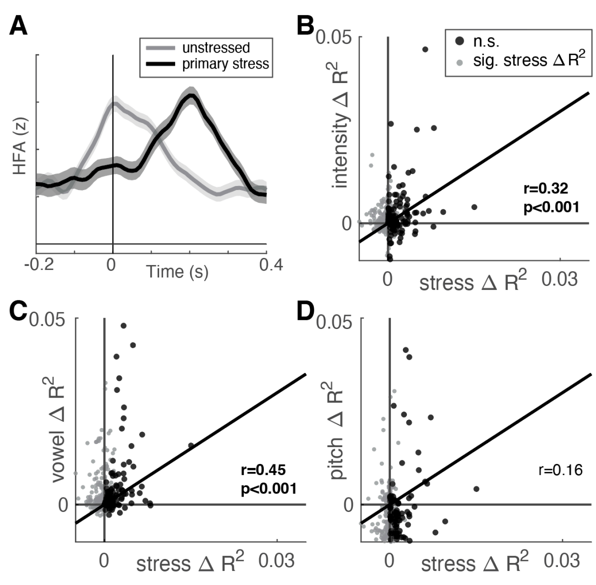

Fig. 1: A. Example electrode neural activity (z-scored High Frequency Activity [HFA]) shows highly differential response to unstressed and primary stressed syllables in natural speech. B. Comparing the contribution of stress features in a neural encoding model (TRF) to acoustic features such as intensity, C. vowel formants and duration, and D. pitch. Each axis in Fig. 1B-D corresponds to the unique variance, or the difference in model performance (R^2) between a model with all features and a model with the labeled feature (intensity, vowel, pitch) removed.

Experiment 2. Experiment 1 did not control for the multiple, co-varying acoustic cues to stress, so we were unable to determine whether the neural encoding of lexical stress on the STG was dependent on the relationships between syllables (relative) for particular acoustic cues, like intensity. Therefore, in Experiment 2, we asked participants to actively indicate which syllable they perceived as stressed in synthesized two-syllable pseudowords (e.g., “hu-ka”, “ma-lu”). We systematically varied the intensity of the first syllable and fixed the intensity of the second syllable to experimentally test whether neural responses to the second syllable depended on the intensity of the first (Fig. 2A).

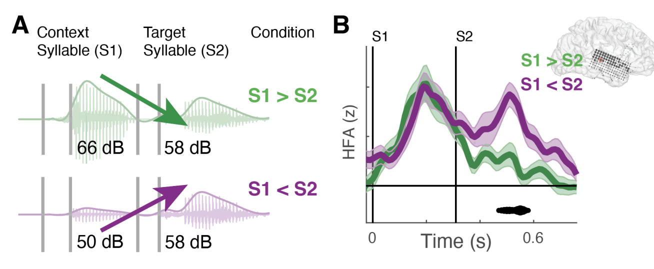

Fig. 2: A. Experiment 2 example stimuli, where the first syllable has variable intensity and the second syllable is fixed. The task condition refers to whether the second syllable (S2) is greater or less than the first (S1). B. Example electrode shows a differential neural response (indicated by thick black lines) aligned to S2 for the S1<S2 condition even though the acoustic content of S2 across conditions is identical.

We found electrodes in the STG that responded stronger to the second syllable when it was greater in intensity than the first, demonstrating a relative sensitivity to intensity difference between consecutive syllables. Furthermore, electrodes with relative sensitivity to the intensity difference between syllables showed non-linear, categorical responses, serving as possible neural foundations for the categorical perception of lexical stress. Our results suggest that electrodes identified in Experiment 1 and Experiment 2 perform highly non-linear transformations of acoustic information critical for detecting stressed syllables in words.

Conclusions

Our findings confirm the fundamental role of the STG in lexical stress perception and further characterize distinct neural populations that encode lexical stress content and related acoustic cues. Specifically, we demonstrate that these neural populations categorically encode relative intensity differences across syllables. This study provides insight into the complex and nonlinear neural computations that transform acoustic speech signals into linguistic representations.

Acknowledgments

We thank Matthew Leonard, Keith Johnson, and members of the Chang lab for their feedback on this work. This work was supported by the National Institutes of Health (NIDCD R01DC012379 to E.F.C).

References

Aleman, A., Formisano, E., Koppenhagen, H., Hagoort, P., de Haan, E. H. F., & Kahn, R. S. (2005). The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. *Cerebral Cortex (New York, N.Y.: 1991)*, *15*(2), 221–228.

Cutler, A. (2008). Lexical Stress. In *The Handbook of Speech Perception* (pp. 264–289). Blackwell Publishing Ltd.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 157–177.

Domahs, U., Klein, E., Huber, W., & Domahs, F. (2013). Good, bad and ugly word stress--fMRI evidence for foot structure driven processing of prosodic violations. *Brain and Language*, *125*(3), 272–282.

Fry, D. B. (1955). Duration and Intensity as Physical Correlates of Linguistic Stress. *The Journal of the Acoustical Society of America*, *27*(4), 765–768.

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., & Pallett, D. S. (1993). *DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. NIST speech disc 1-1.1* (NASA STI/Recon Technical Report N, Vol. 93, p. 27403). adsabs.harvard.edu. https://ui.adsabs.harvard.edu/abs/1993STIN...9327403G

Holdgraf, C. R., Rieger, J. W., Micheli, C., Martin, S., Knight, R. T., & Theunissen, F. E. (2017). Encoding and Decoding Models in Cognitive Electrophysiology. *Frontiers in Systems Neuroscience*, *11*, 61.

Mousikou, P., Strycharczuk, P., & Rastle, K. (2024). Acoustic correlates of stress in speech perception. *Journal of Memory and Language*, *136*(104509), 104509.

Oganian, Y., Bhaya-Grossman, I., Johnson, K., & Chang, E. F. (2023). Vowel and formant representation in the human auditory speech cortex. *Neuron*, *111*(13), 2105–2118.e4.

Oganian, Y., & Chang, E. F. (2019). A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Science Advances*, *5*(11), eaay6279.

Schwab, S., Mouthon, M., Jost, L. B., Salvadori, J., Stefanos-Yakoub, I., da Silva, E. F., Giroud, N., Perriard, B., & Annoni, J.-M. (2023). Neural correlates of lexical stress processing in a foreign free-stress language. *Brain and Behavior*, *13*(1), e2854.

Tang, C., Hamilton, L. S., & Chang, E. F. (2017). Intonational speech prosody encoding in the human auditory cortex. *Science*, *357*(6353), 797–801.