The developmental trajectory and neuroanatomical correlates of speech comprehension

Jill Kries (jill.kries@stanford.edu)
Stanford University, Stanford, CA, USA

Brian Ervin (Brian.Ervin@cchmc.org)
Cincinnati Children's Hospital Medical Center, Cincinnati, OH, US

Jason Buroker (Jason.Buroker@cchmc.org)
Cincinnati Children's Hospital Medical Center, Cincinnati, OH, US

Gregory Hickok (gshickok@uci.edu)
University of California, Irvine, CA, USA

Ravindra Arya (Ravindra.Arya@cchmc.org)
Cincinnati Children's Hospital Medical Center, Cincinnati, OH, US

Laura Gwilliams (laura.gwilliams@stanford.edu)
Stanford University, Stanford, CA, USA

Abstract

Language acquisition spans early childhood into young adulthood. Neural mechanisms underlying this prolonged development, especially within naturalistic speech comprehension settings, remain understudied. Here, we used intracranial EEG recordings from 42 children with epilepsy while they listened to natural speech. We reconstructed acoustic and symbolic features from the neural data and found significant above-chance decoding in 77% of children. An ascending pattern of across-age feature decoding indicates later development of more symbolic features. The features seem to be co-encoded in temporal regions. These findings provide first insights into the language hierarchy development and the neuroanatomical correlates of speech comprehension throughout childhood.

Keywords: language development; speech comprehension; modelling; intracranial EEG

Introduction

Brain function and structure mature throughout childhood and up to young adulthood, with language development progressing from infant babbling to complex sentence processing by age 3 and beyond. In adults, natural speech comprehension involves parallel encoding of hierarchical representations, from acoustic to symbolic features (Brodbeck et al., 2022; Gwilliams et al., 2024). Developmental studies have shown that infants encode acoustic features, such as the speech envelope (Kalashnikova et al., 2018; Ríos-López et al., 2020) and phonetic features (Di Liberto et al., 2018). However, research on the development of higher-level speech features related to information-theoretic cues and prediction remains limited. Moreover, these studies used scalp EEG, with limited spatial precision. Here, we therefore (1) evaluate whether the feature hierarchy can be decoded above chance in children and adolescents using intracranial EEG recordings, (2) characterize the developmental trajectory for each level of the language hierarchy, and (3) identify where in the brain each feature is encoded.

Methods

42 participants aged between 5 and 20 years of age listened to 2 to 3.5 minutes of natural speech. Participants were patients with drug-resistant epilepsy undergoing standard-of-care stereo-EEG monitoring. Electrode locations were individually chosen clinically, however, sampling of peri-sylvian language areas is very common. We excluded participants (1) if brain anatomy atlas labels indicated no contact in speech and language areas (n=2), and (2) if they had less than 2 speech-responsive electrodes (n=2).

We annotated the speech stimuli with a feature hierarchy ranging from sensory to symbolic features, comprising the speech amplitude envelope, phonetic features, phoneme entropy, phoneme surprisal (Gwilliams & Davis, 2022), and GPT2 surprisal (i.e., context-dependent). The neural data underwent preprocessing, which included band-pass filtering (1-15Hz for envelope decoding and 70-150Hz for the other four features), common-average referencing, downsampling, and normalization.

Our investigation focused on three main aspects: 1) assessing above-chance decoding performance across childhood and adolescence for the feature hierarchy (above-chance decoding: true performance larger than 95% of random permutations), 2) examining the decoding performance of each feature across age to determine the developmental trajectory for each level of the language hierarchy, and 3) characterizing the anatomical location associated with each feature.

For the speech envelope decoding analysis, we used continuous neural data (for spatio-temporal receptive field (STRF) decoding), while we epoched the data around phoneme/token onset for pulse-based speech features. We used ridge regression to decode speech envelope, phoneme entropy, phoneme surprisal, and word surprisal, and logistic regression for phonetic feature decoding from the neural recordings.

Results

First, we tested whether speech features are decoded significantly above chance-level across age. On average across speech features, we found that 77% of participants showed significant decoding (envelope: 77%; phonetic features: 76%; phoneme surprisal: 78%; phoneme entropy: 78%; word surprisal: 78%). Second, we explored the developmental trajectory of the decoding performance for each feature. We found that a cubic fit is better than a linear or quadratic fit for each feature. However, the cubic models did not significantly explain the data for any of the features (envelope: R2=0.03, p=.5; phonetic features: R2=0.05, p=.6; phoneme surprisal: R2=0.024, p=.8; phoneme entropy: R2=0.06, p=.4; word surprisal: R2=0.1, p=.8). We show the decoding performance data and non-significant model fits in figure 1 (arrows show max. decoding performance of model fits).

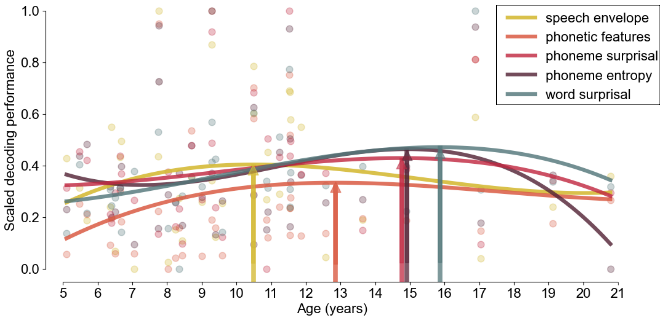

Figure 1: Developmental trajectory of the speech processing hierarchy. Decoding performance was min-max scaled for each feature separately, such that scores are between 0 & 1.

Third, we characterized the anatomical location of speech feature decoding across participants (fig. 2). We tested, for each feature, which regions show the highest ratio of significant electrodes to total number of electrodes (table 1). With this data-driven approach, we found that all features show the highest ratio of significant electrodes in left planum temporale. The 2nd to 5th highest ranking areas differ between features (table 1).

Sensory ⟷ Symbolic

	Envelope	Phonetic features	Phoneme surprisal	Phoneme entropy	Word surprisal
Area with the highest ratio	Left planum temporale **(100%)**	Left planum temporale **(58.3%)**	Left planum temporale **(33.3%)**	Left planum temporale **(33.33%)**	Left planum temporale **(33.3%)**
Area with 2nd highest ratio	Right superior temporal gyrus **(90.9%)**	Right pars opercularis **(41.6%)**	Right lateral ventricle **(25%)**	Right central operculum **(24.4%)**	Right superior temporal gyrus **(33.3%)**
Area with 3rd highest ratio	Left superior temporal gyrus **(65.7%)**	Right superior temporal gyrus **(36.3%)**	Right central operculum **(22.2%)**	Right anterior insula **(21.4%)**	Left amygdala **(18.7%)**
Area with 4th highest ratio	Right frontal operculum **(63.1%)**	Right lateral ventricle **(33.3%)**	Right parietal operculum **(15.3%)**	Right superior frontal gyrus **(17.6%)**	Left posterior insula **(17.6%)**
Area with 5th highest ratio	Left lateral ventricle **(58.3%)**	Right precentral gyrus **(33.3%)**	Right superior temporal gyrus **(15.1%)**	Left inferior temporal gyrus **(15.7%)**	Right pars triangularis **(16.6%)**

Table 1: Neuroanatomical areas with the highest ratio between number of significant electrodes and number of total electrodes, per feature. We only included anatomical labels with at least 10 electrodes.

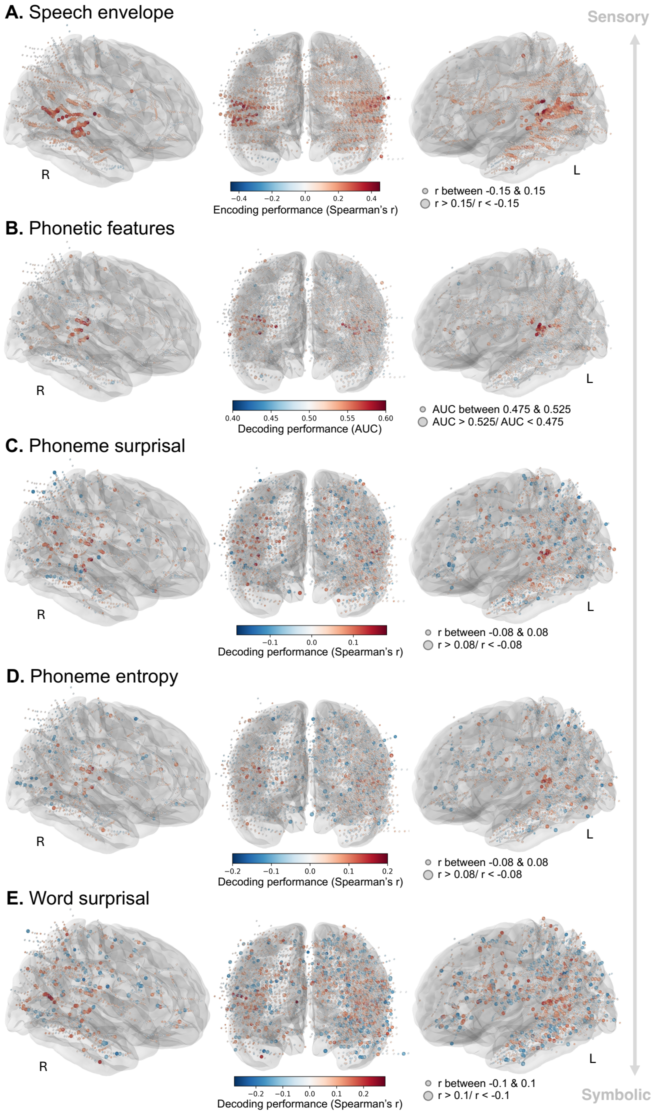

Figure 2: Decoding model performance for each feature and electrode. Electrodes with higher encoding strength were plotted larger (specific legend can be found at each panel).

Discussion & Conclusion

In this study, we reconstructed acoustic and symbolic features from the neural data. First, we found significant above-chance decoding across features in 77% of children. Second, the spatial analysis shows that features might be co-encoded in temporal regions. However, in future analyses, the covariance between features should be taken into account, which may result in more diverging feature localization. Third, our developmental trajectories within-feature were not significant, suggesting that a larger cross-sectional sample and more trials per participant will be necessary to determine the developmental acquisition of acoustic and symbolic features. Nonetheless, these findings hold promise for future extensions of this research question.

Acknowledgments

We would like to thank all participants of this study and their families.

References

Brodbeck, C., Bhattasali, S., Cruz Heredia, A. A. L., Resnik, P., Simon, J. Z., & Lau, E. (2022). Parallel processing in speech perception with local and global representations of linguistic context. *ELife*, *11*. https://doi.org/10.7554/eLife.72056

Di Liberto, G. M., Peter, V., Kalashnikova, M., Goswami, U., Burnham, D., & Lalor, E. C. (2018). Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia. *NeuroImage*, *175*, 70–79. https://doi.org/10.1016/j.neuroimage.2018.03.072

Gwilliams, L., Davis, M.H. (2022). Extracting Language Content from Speech Sounds: The Information Theoretic Approach. In: Holt, L.L., Peelle, J.E., Coffin, A.B., Popper, A.N., Fay, R.R. (eds) Speech Perception. Springer Handbook of Auditory Research, vol 74. Springer, Cham. https://doi.org/10.1007/978-3-030-81542-45

Gwilliams, L., Marantz, A., Poeppel, D., & King, J.-R. (2024). *Hierarchical dynamic coding coordinates speech comprehension in the human brain*. https://doi.org/10.1101/2024.04.19.590280

Kalashnikova, M., Peter, V., di Liberto, G. M., Lalor, E. C., & Burnham, D. (2018). Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-32150-6

Ríos-López, P., Molinaro, N., Bourguignon, M., & Lallier, M. (2020). Development of neural oscillatory activity in response to speech in children from 4 to 6 years old. *Developmental Science*, *23*