Development of speech in the brain using intracranial recordings

Liberty S. HAMILTON[1,2]

1. Department of Neuroscience and Department of Statistics, University of California, Berkeley, CA, USA. 2. Department of Speech, Language, and Hearing Sciences and Department of Neurology, The University of Texas at Austin, TX, USA.

E-mail: *liberty.hamilton@berkeley.edu*

Abstract: We recorded intracranially in patients aged 4 to 21 to understand neural development of speech. We fit encoding models predicting brain responses from spectrograms or phonological features and found phoneme invariance emerges during adolescence. These results have implications for understanding typical and atypical language development.

Keywords: Intracranial recordings, Speech, Language, Development

Introduction

Typically, infants learn the phonemes of their native language as early as 6 months to one year of age [1]. Still, many speech-related tasks, including understanding speech in noise, do not mature until adolescence [2]. In children with epilepsy, language networks may be affected by ongoing seizures and other comorbidities. Understanding how such networks develop in the brain is important for identifying the basis of typical and atypical language development, and for guiding interventions in children with communication disorders. Here, we answer this question using intracranial recordings during speech tasks from children, adolescents, and young adults with drug resistant epilepsy.

Methods

Participants: We acquired intracranial recordings from depth or grid electrodes in 50 patient participants aged 4–21 undergoing surgical monitoring for drug resistant epilepsy at Dell Children's Medical Center in Austin TX or Texas Children's Hospital in Houston, TX. Recordings included bilateral coverage of speech-sensitive peri-Sylvian cortex including Heschl's gyrus (HG), planum temporale (PT), planum polare (PP), superior temporal gyrus (STG), superior temporal sulcus (STS), and middle temporal gyrus (MTG) (Figure 1B). Patients provided written informed consent (if aged 18+) or assent and parental permission for those under 18. All procedures were approved by the University of Texas at Austin Institutional Review Board.

Task: Participants watched audiovisual movie clips taken from well-known movie studios such as Pixar, Dreamworks, and Disney. These movies were transcribed for the presence of phoneme and word level information using FAVE-align and boundaries were manually corrected [3], [4] (Figure 1A).

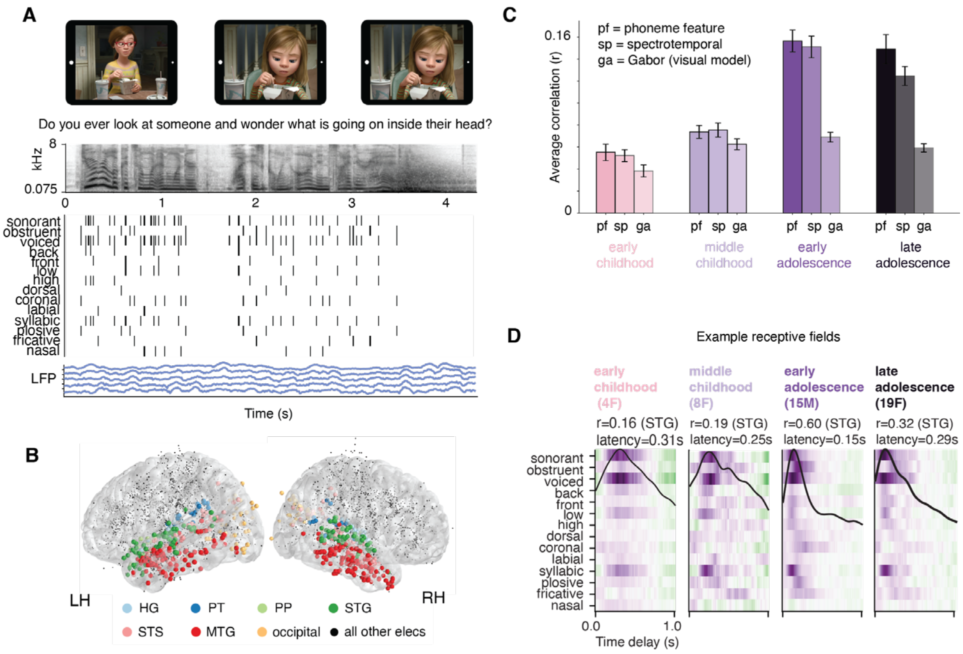

Figure 1: (A) Task and features used in the encoding model. Patients watched audiovisual movies presented on an iPad (top). The spectrogram and phonological features were used as stimulus features to predict the high gamma local field potential (LFP) recordings from each electrode. (B) Electrode coverage across all 50 participants, with temporal lobe and occipital regions of interest highlighted. (C) Average encoding model performance across phonological feature model (pf), spectrogram model (sp), and Gabor wavelet model (ga) for each age group. (D) Example receptive fields showing decrease in latency across age.

Neural data preprocessing: Neural data were preprocessed by manually rejecting epileptiform activity, applying a common average reference, and extracting high gamma band power (70-150 Hz) by calculating the analytic amplitude of the Hilbert transform. We fit linear encoding models to predict neural activity from acoustic or phonetic information in the stimulus. Acoustic information was operationalized as an 80-band mel spectrogram, while phonetic information was a binary matrix at the onset of each phoneme, with features for place and manner of articulation [5] (Figure 1A). Model performance was evaluated by calculating the correlation between the neural response predicted by each model and the actual neural response to held out data. To measure neural processing speed, we calculated the peak latency of the encoding model weights and analyzed trends across age groups (Figure 1D) – early childhood (age 4-5), middle childhood (age 6-12), early adolescence (age 13-17) and late adolescence (age 18-21). We also correlated neural selectivity with behavioral measures of verbal comprehension, attention, and reading as assessed through preoperative neuropsychological testing.

Results

We found robust responses to acoustic information in early auditory areas such as HG across the entire age range. However, robust responses to phoneme categories did not emerge until early adolescence (Figure 1C). Response latencies in speech-selective cortex (STG, STS, MTG) became faster and more precise with age (Figure 1D), while auditory responses in HG were relatively more stable. Measures of neural selectivity were most strongly correlated with actual patient age, rather than developmental measures of language ability.

Discussion

Our results suggest that robust cortical representations of phonetic information emerge relatively late, concordant with behavioral reports of later crystallization of speech in noise processing. Our results also suggest that phoneme processing develops along a biological timeline, relatively independent of language and attention measures. One limitation of this study is that intracranial recordings can only be performed in patients with drug resistant epilepsy, which can affect language and development. However, we found that age of onset of epilepsy and duration of epilepsy was not correlated with our neural measures.

Conclusions

By incorporating movie watching tasks, we were able to assess speech and language development across a wide age range with a child-friendly research paradigm. Our results may have implications for language-related disorders including dyslexia and auditory processing disorder.

Acknowledgements

We would like to thank the patients and their families for their participation in this research. The author would like to thank Maansi Desai, Alyssa Field, Sandra Georges, Jacob Leisawitz, Gabrielle Foox, Saman Asghar, Dave Clarke, Nancy Nussbaum, Rosario DeLeon, Elizabeth Tyler-Kabara, Howard Weiner, Andrew Watrous, and Anne Anderson for contributing to this project. This work was funded by R01DC018579 (PI: Hamilton) from the National Institutes of Health National Institute on Deafness and Other Communication Disorders and the Texas Speech-Language-Hearing Foundation.

References

[1]P. Iverson and P. K. Kuhl, "Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling," *J. Acoust. Soc. Am.*, vol. 97, no. 1, pp. 553–562, 1995.
[2]D. H. Sanes and S. M. N. Woolley, "A behavioral framework to guide research on central auditory development and plasticity," *Neuron*, vol. 72, no. 6, pp. 912–929, Dec. 2011.
[3]I. Rosenfelder, J. Fruehwald, and K. A. J. Y. Evanini, "FAVE (Forced Alignment and Vowel Extraction) Program Suite," 2011. [Online]. Available: http://fave.ling.upenn.edu.
[4]M. Desai, J. Holder, C. Villarreal, N. Clark, B. Hoang, and L. S. Hamilton, "Generalizable EEG Encoding Models with Naturalistic Audiovisual Stimuli," *J. Neurosci.*, vol. 41, no. 43, pp. 8946–8962, Oct. 2021.
[5]N. Mesgarani, C. Cheung, K. Johnson, and E. F. Chang, "Phonetic feature encoding in human superior temporal gyrus," *Science*, vol. 343, no. 6174, pp. 1006–1010, Feb. 2014.