Direct Brain Recordings Show Strengthening Prioritization of Speech Over Music

Rajvi Agravat[1], Maansi Desai[2], Alyssa M. Field[2], Gabrielle Foox[5], Sandra Georges[5], Jacob Leisawitz[5], Saman Asghar[5], Anne E. Anderson[5], Dave Clarke[4], Elizabeth C. Tyler-Kabara[4], Andrew J. Watrous[5], Howard L. Weiner[5], Liberty S. Hamilton[2,3,6,7,8]

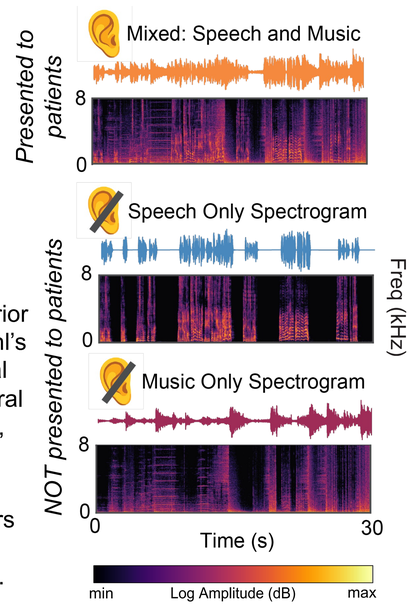
[1]Interdisciplinary Neuroscience Program, The University of Texas at Austin; [2]Department of Speech, Language, and Hearing Sciences, The University of Texas at Austin; [3]Department of Neurology, Dell Medical School, The University of Texas at Austin; [4]Dell Children's Medical Center; [5]Texas Children's Hospital, Baylor College of Medicine; [6]Department of Neuroscience, University of California, Berkeley; [7]Department of Statistics, University of California, Berkeley; [8]Department of Neurosurgery, University of California, San Francisco

Introduction

- In everyday life, our brains constantly separate overlapping sounds, such as speech and music, to focus on important information, using a process called ***auditory streaming***.
- This is essential for understanding conversations and distinguishing different sounds in natural environments.
- Previous studies have examined explicit attention to a target with a distracting sound, but real-world auditory processing often occurs ***without directed focus***.
- This study reveals **how different temporal lobe regions encode speech and music streams in a developing brain, without explicit attention to either modality.**

sEEG Data Collection & Methods

- **Participants:** 47 patients (age 4 to 22; 25M/22F) undergoing sEEG monitoring.
- **Neural recordings:** High-gamma activity (70-150 Hz) from 1,053 temporal lobe electrodes.
- **ROIs analyzed**: superior temporal gyrus, Heschl's gyrus, middle temporal gyrus, superior temporal sulcus, planum polare, planum temporale.
- **Stimuli:** Naturalistic children's movie trailers containing mixed speech-music content.

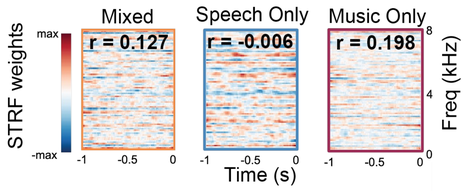


- **Audio separation for model fitting:** ***Post data collection, the original mixed audio was separated into speech-only and music-only streams*** using a DNN-based source separation algorithm (Moises).
  This yielded three versions of each stimulus:
  (1) **original mixed audio** (30.8% of all timepoints)
  (2) **isolated speech** (9.9% of all timepoints)
  (3) **isolated music** (41.1% of all timepoints).
- **Analysis:** Fit 3 separate Spectro-Temporal Receptive Field (STRF) models to the original, speech-only, and music-only stimuli to understand stream encoding.

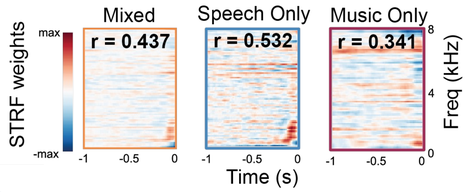
$$sEEG(t,n) = \sum_f \sum_\tau w(f,\tau,n)\, s(f,t-\tau) + \varepsilon(t,n)$$

predicted electrode activity | receptive field | stimulus representation | residuals

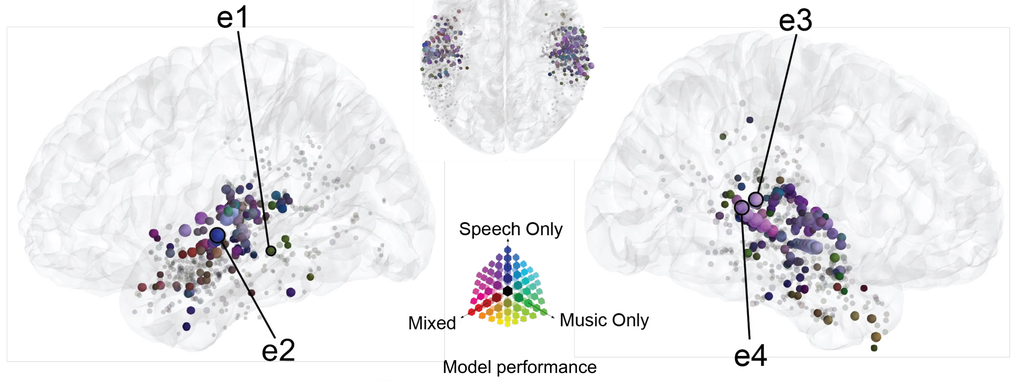
Auditory cortex preferentially tracks speech over music without explicit attention



Music selective electrode in STG

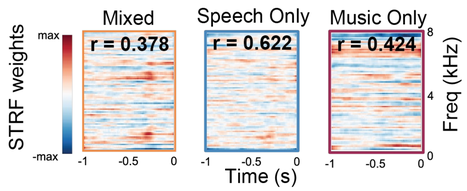


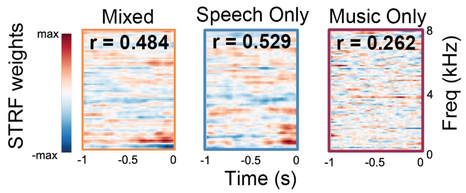
Speech selective electrode in STG



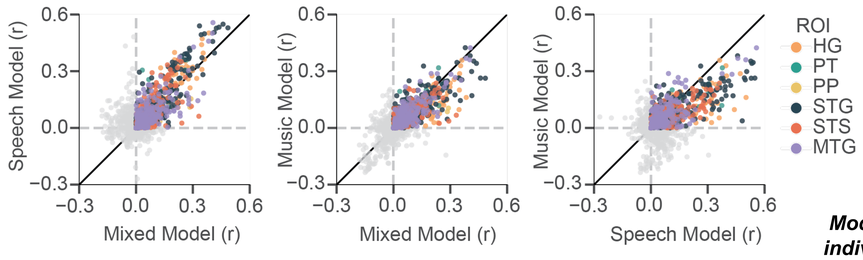
Speech selective electrode in STG

Speech selective electrode in STG

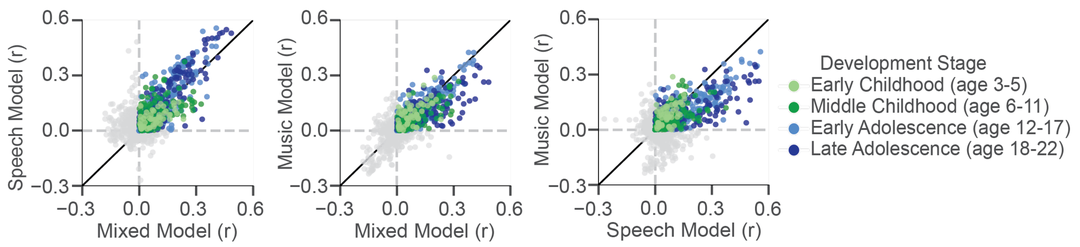




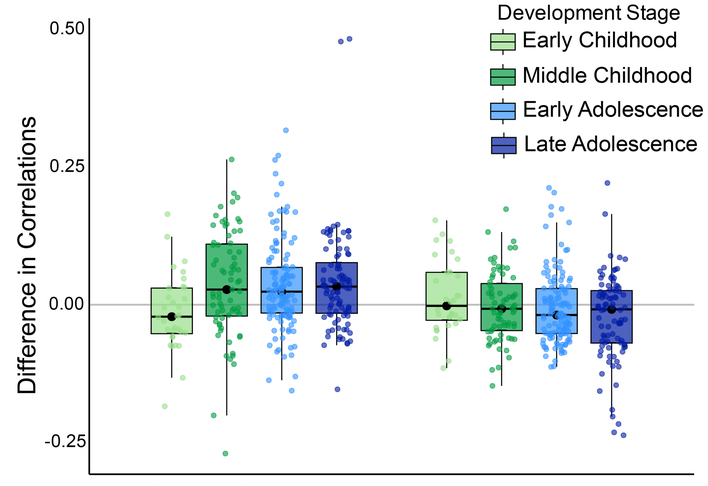
Key Finding: Despite participants passively listening to mixed speech+music stimuli, spectro-temporal receptive field models trained on DNN-separated speech components significantly outperformed mixed-audio and music-only models in predicting neural responses across higher-order auditory regions.



Model performance for individual electrodes by ROI or age



Speech selectivity increases with age in STG (321 electrodes)



Speech Selectivity $r_{speech} - r_{mixed}$

Music Selectivity $r_{music} - r_{mixed}$

Conclusion

- Although participants heard a combination of speech and music during the task, higher-order auditory cortex activity was **better modeled by the speech-separated audio**, indicating **preferential encoding of speech.**
- Statistical analysis revealed significantly **greater speech selectivity across multiple temporal lobe regions**, including STG, STS, and MTG (all $p < 0.0001$), as well as PT ($p < 0.03$). HG showed significant selectivity for both speech-only and music-only conditions (both $p < 0.0001$), indicating a distinct response pattern.
- Speech selectivity **strengthened from childhood to adolescence** in STG, indicating continued maturation of auditory streaming.

References & Funding

- Desai, M et al. (2021) J Neuro. 41(43): 8946-8962
- Holdgraf et al. (2017) Frontiers in Systems Neuroscience 11,61.
- The Musician's App. https://moises.ai/

This study is funded under a grant by the National Institutes of Health to LSH (1R01DC018579, NIDCD).