Progression of acoustic, phonemic, lexical and sentential neural features emerge during speech listening

I.M Dushyanthi Karunathilake[1], Christian Brodbeck[2], Shohini Bhattasali[1], Philip Resnik[1], Jonathan Z. Simon[1]

[1]University of Maryland, College Park, [2]University of Connecticut

INTRODUCTION

Understanding speech requires analyzing the acoustic waveform via intermediate abstract representations including phonemes, words and ultimately meaning along with other cognitive operations. While recent neurophysiological studies have reported that the brain tracks acoustic and linguistically meaningful units, the impact of different kinds of speech information and how these feature responses are modulated by top-down mechanisms is not well understood.

Motivation

- How are different speech features driven by bottom-up and top-down mechanisms (and when)?
- Investigate the progression and representation of different speech features along the speech and language hierarchy.
- How the speech features emerge for different speech conditions?

METHODS

30 younger adults (18-30 years), Native English speakers

Neural Recording - Magnetoencephalography (MEG)

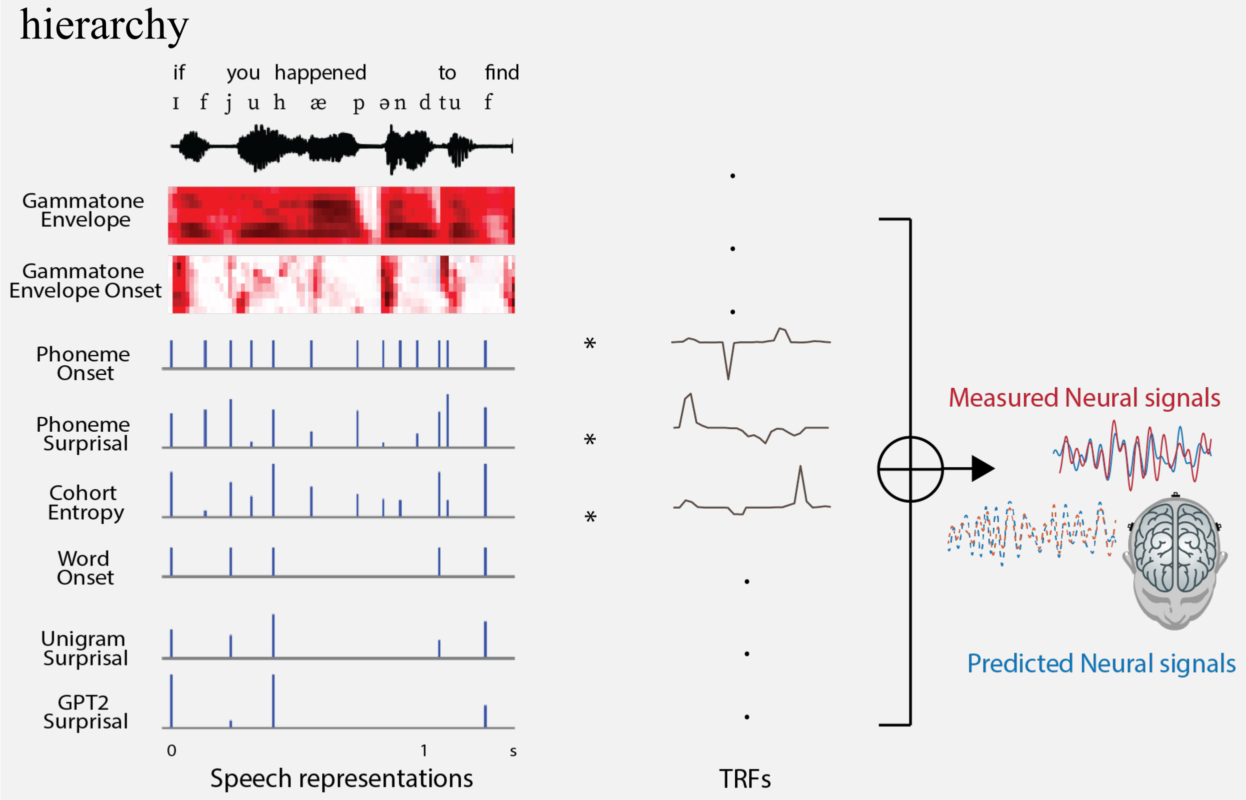
Task - Listening to 1-min-long continuous speech, 4 passage types

Narrative	*"If you happened to find yourself on the banks of the Ohio River on a particular afternoon in the⋯"*	Speech-like prosody and rhythm
Scrambled Words	*"A liquid is only speak, second even for good reach the attack us. Living fact, which it's was plants⋯"*	
Non-words	*"Sustument eviless, joservil edfolke provericant zin tahovasibed bi conson sketting pitablion⋯"*	
Speech-Modulated Noise		

MEG data was band passed 1-10 Hz

Source localization using MNE, Temporal Lobe

Analysis - Temporal Response Functions (TRFs) including different speech representations along the speech and linguistic hierarchy



Gammatone envelope - Acoustic power in logarithmically spaced 8 bands

Gammatone envelope onset - Rising slope of acoustic power in the same bands

Phoneme surprisal - How surprising the current phoneme is given the previous phoneme sequence

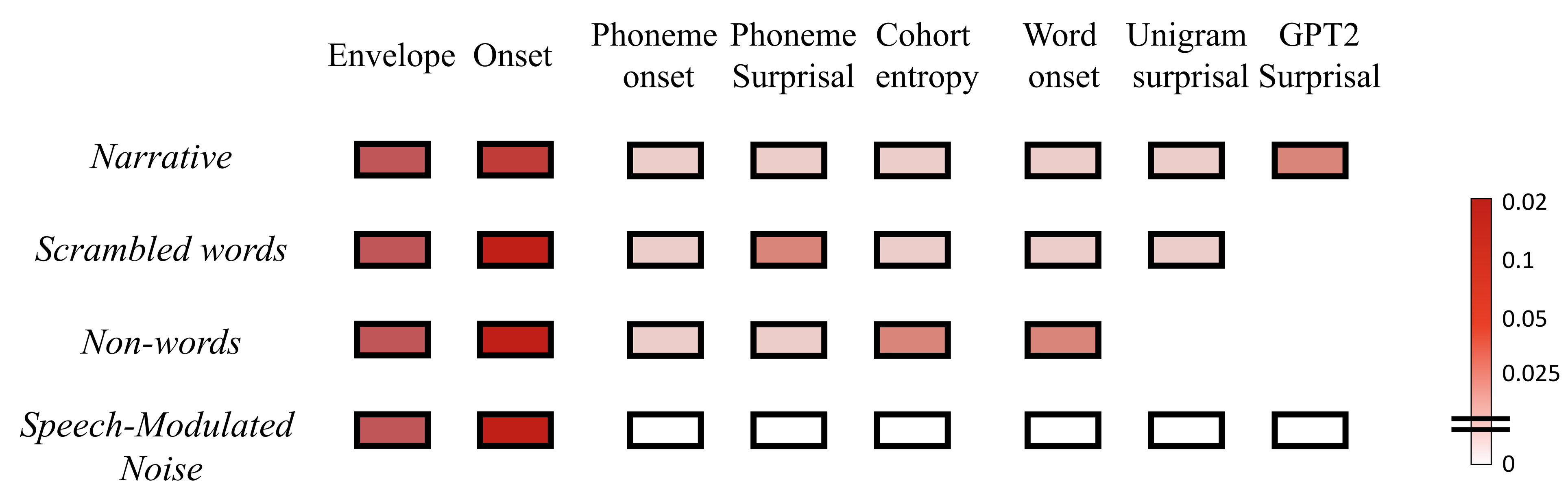
Cohort entropy - Lexical competition among words that are compatible with the phoneme sequence

Unigram surprisal - Context independent word surprisal calculated using SUBTLEX database

GPT2 surprisal - Context based word surprisal measured using GPT2 language model

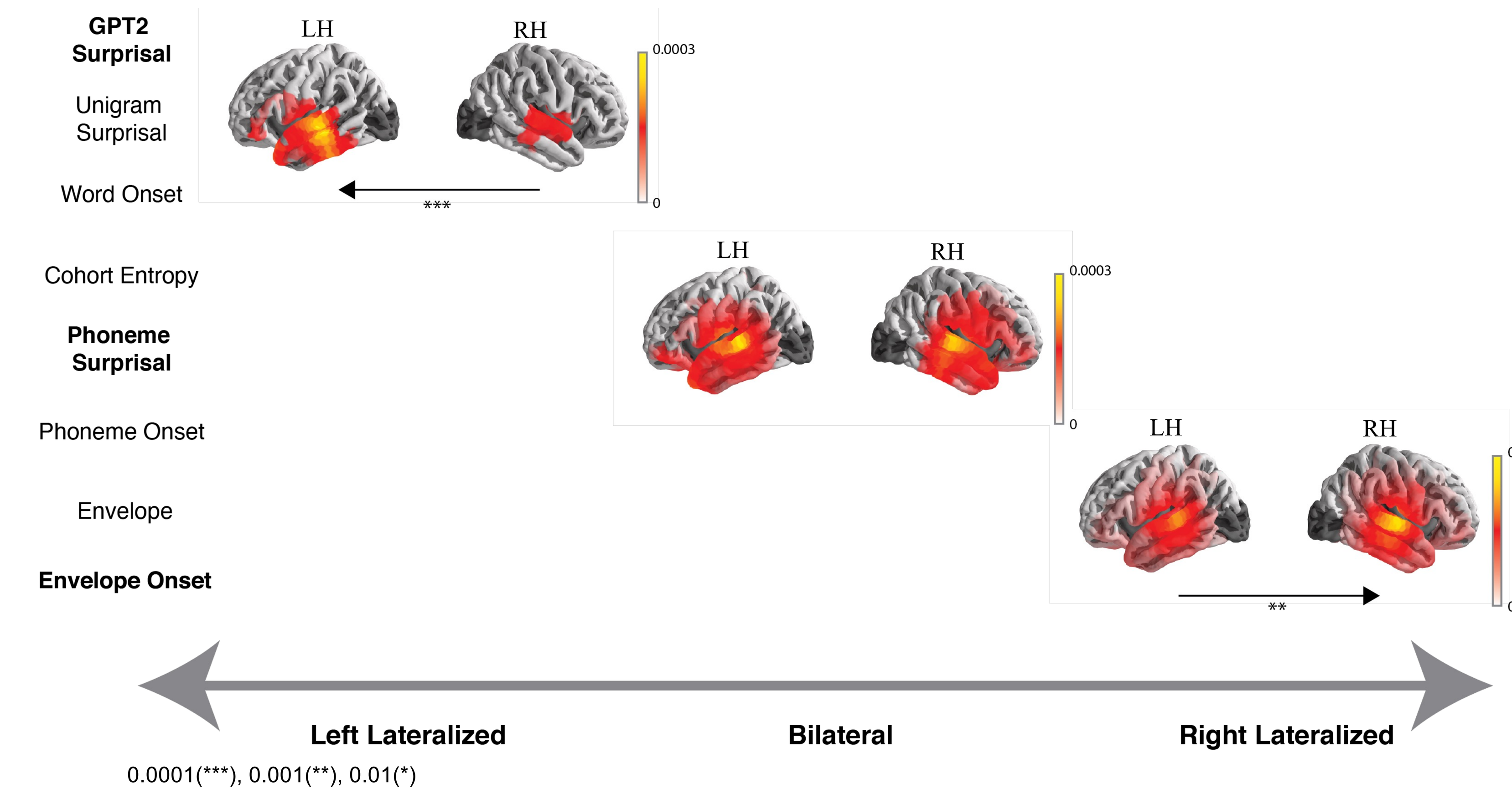
RESULTS

Emergence of neural features as the incremental processing occur



- Acoustic features are encoded for both non-speech and speech stimuli
- (Sub)-lexical features are encoded only when (sub)-lexical boundaries are intelligible
- Context based word surprisal emerges for narrative passage
- When context supports, context based surprisal is better tracked compared to unigram surprisal
- Unigram surprisal and GPT2 surprisal were not defined for non-words
- When there is no context, GPT2 surprisal converges to unigram surprisal. Therefore, GPT2 surprisal was not included in scrambled word passage TRF modelling

Speech feature processing **hemispheric lateralization**



- Low-level feature processing are right lateralized and Higher level features processing are left lateralized
- Non-words processing mostly bi-lateral (Lateralization may be task dependent)

Temporal Response Functions

Early word processing ~100 ms

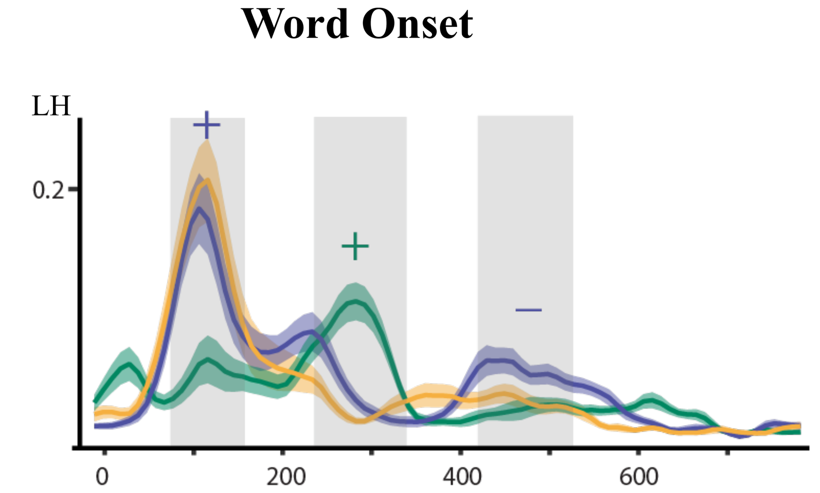
Scrambled ≈ Narrative

Different neural mechanisms for non-words ~300 ms

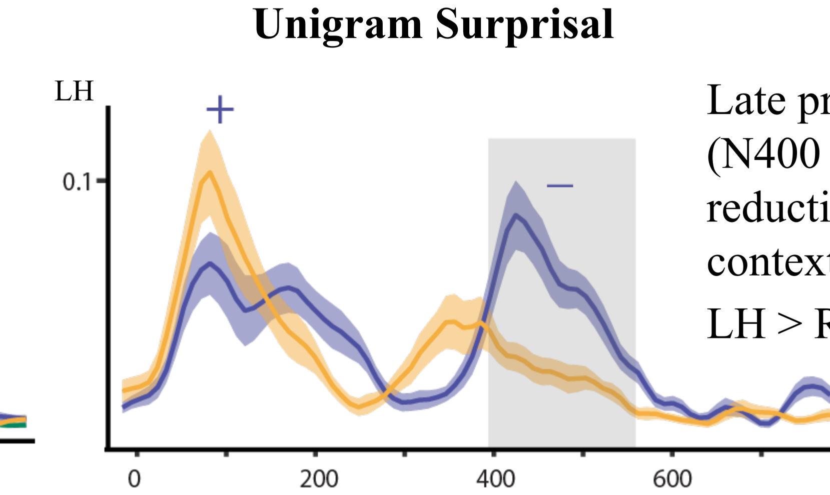
Words LH > RH

Non-words LH ≈ RH

Word Onset



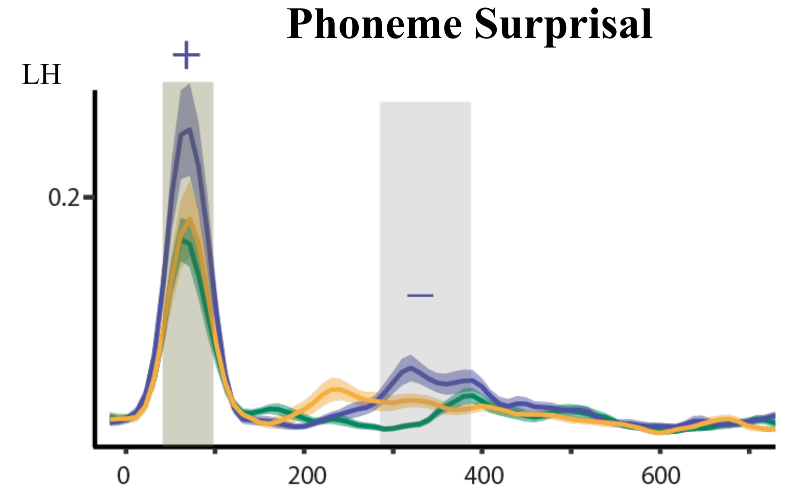
Unigram Surprisal



Late processing ~450 ms (N400 like response, reduction in surprisal due to context)

LH > RH

Phoneme Surprisal

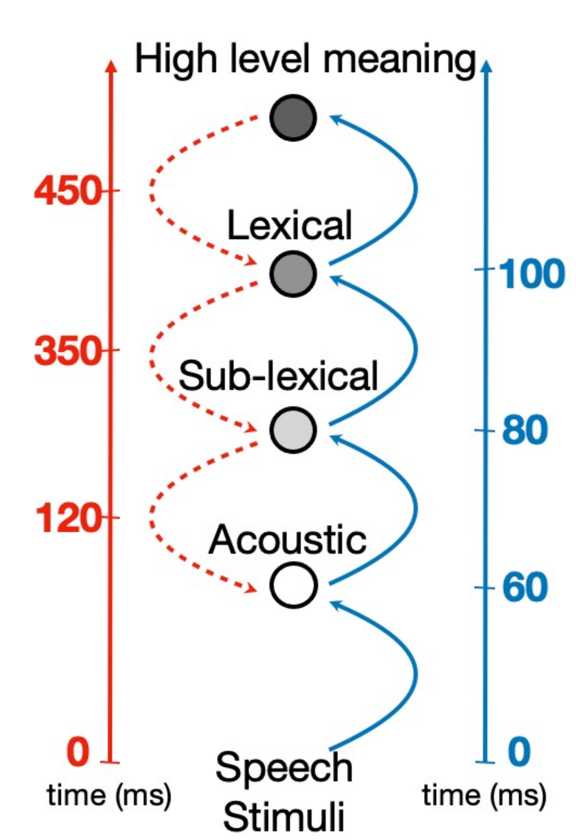


Early phoneme onset processing ~80 ms

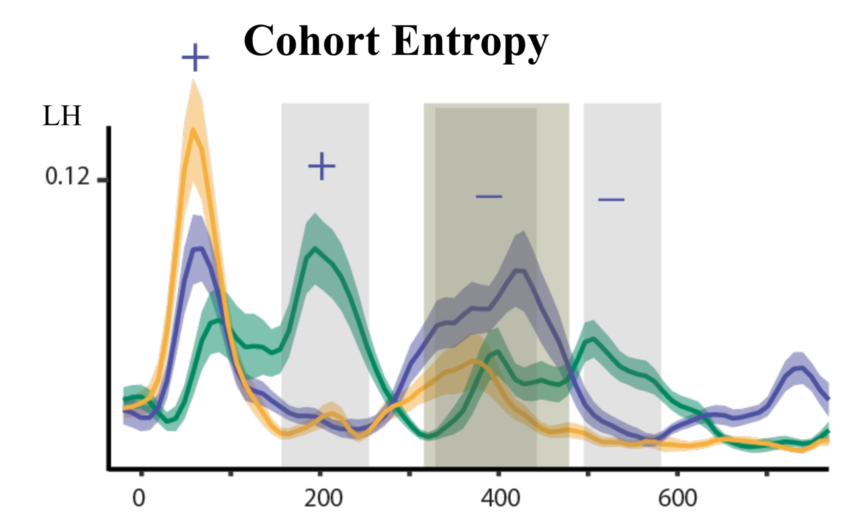
Scrambled ≈ Non-word

Non-words > Narrative

Noise lacks second peak ~350 ms



Cohort Entropy



Late context processing ~400 ms (N400 like response, reduced for narrative)

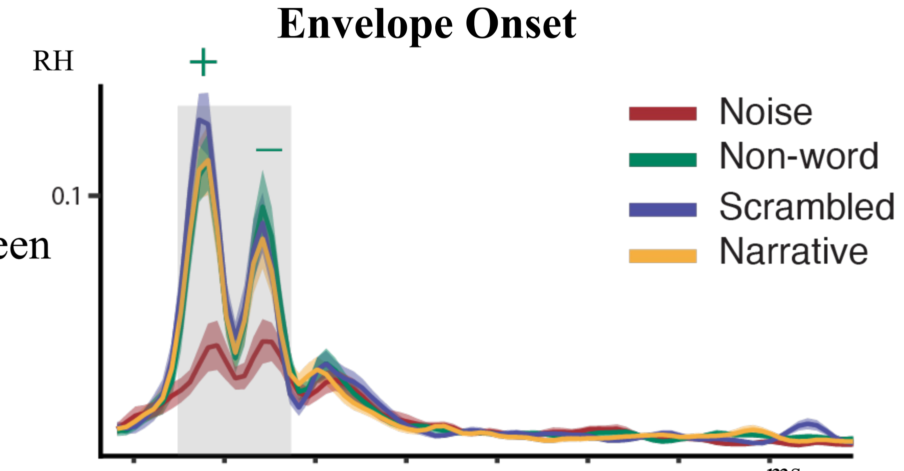
Additional/delayed peaks in non-words (difference in stimulus distributions)

Envelope Onset

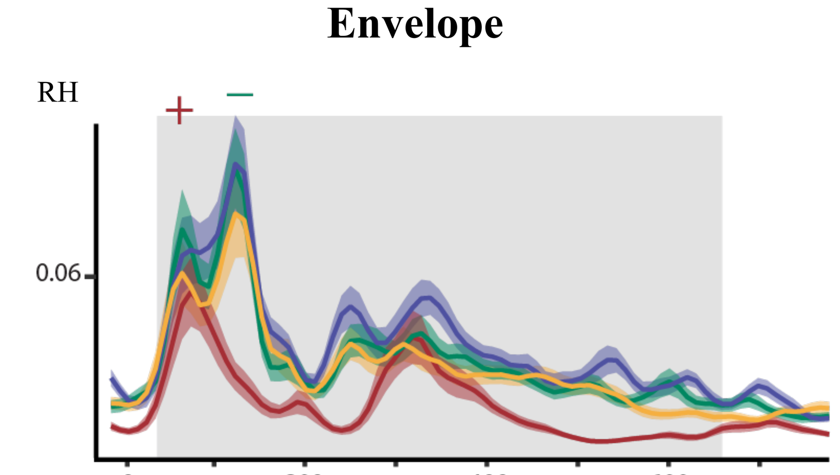
Speech > Noise

No difference between speech conditions

Acoustic driven



Envelope



Speech > Noise

Scrambled ≈ Non-word

Non-words > Narrative

Early processing ~60 ms

Noise lacks second peak ~120 ms

CONCLUSION

- Cortical response time-locks to emergent features from acoustics to context as incremental steps in the processing of speech input occur
- Lower-level acoustic feature responses are right lateralized whereas, context based responses are left lateralized
- Linguistic features are processed when the linguistic boundaries are intelligible
- Higher level processing/top-down mechanisms in addition to lower level processing/ bottom up mechanisms

Acknowledgements

This work was supported by the National Institutes of Health grants R01-DC019394 and National Science Foundation SMA 1734892.

References

[1]. Xu, J. *et al.* (2005) 'Language in Context: Emergent Features of Word, Sentence, and Narrative Comprehension'. *NeuroImage*, 25(3), pp. 1002–1015.

[2]. Brodbeck, C., Hong, L.E. and Simon, J.Z. (2018) 'Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech'. *Current Biology*, 28(24), pp. 3976-3983.e5.

[3]. Gillis, M. *et al.* (2021) 'Neural Markers of Speech Comprehension: Measuring EEG Tracking of Linguistic Speech Representations, Controlling the Speech Acoustics'. *The Journal of Neuroscience*, 41(50), pp. 10316–10329.

[4]. Brodbeck, C. et al. (2022) 'Parallel Processing in Speech Perception with Local and Global Representations of Linguistic Context'. ELife, 11, p. e72056.

[5]. Schrimpf, M. et al. (2021) 'The Neural Architecture of Language: Integrative Modeling Converges on Predictive Processing'. Proceedings of the National Academy of Sciences, 118(45), p. e2105646118.

For poster and sample audio clips:

Contact- dushk@umd.edu