Progression of acoustic, phonemic, lexical and sentential neural features emerge during speech listening
LM Dushyanthi Karanathilake1, Christian Brodbeck2, Shohini Bhattacharl2, Philip Resnik1, Jonathan Z. Simon1
1University of Maryland, College Park, 2University 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 do 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

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

RESULTS
Emergence of neural features as the incremental processing occur
<table>
<thead>
<tr>
<th>Envelope Onset</th>
<th>Phoneme onset</th>
<th>Phoneme Surprisal</th>
<th>Cohort entropy</th>
<th>Word onset</th>
<th>Unigram GPT2 Surprisal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scrambled words</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-words</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech-Modulated Noise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

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

Emergent Neural responses for non-speech & speech conditions

Speech > Noise
- Early processing ~60 ms
- Noise lacks second peak ~120 ms

Phoneme Surprisal
- Early phoneme onset processing ~80 ms
- Scrambled ≈ Non-word
- Non-words > Narrative
- Noise lacks second peak ~350 ms

Lexical
- Late context processing ~400 ms (N400 like response, reduction in surprisal due to context) LH > RH
- Additional/delayed peaks in non-words (difference in stimulus distributions)

High level meaning
- Late processing ~450 ms (N400 like response, reduction in surprisal due to context) LH > RH

Speech Stimuli

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

Speech like prosody and rhythm

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

References