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

I.M Dushyanthi Karunathilake, Christian Brodbeck, Shohini Bhattasali, Philip Resnik, Jonathan Z. Simon

University of Maryland, College Park
University of Connecticut
How do our brains perceive and construct meaning from continuous sounds?

• Acoustic inputs are turned into a meaning via intermediate stages.
  - Intermediate stages can perform
    ▶ Acoustic analysis (acoustic envelope)
    ▶ Phonological analysis (phonemes)
    ▶ Lexical processing of word (words)
    ▶ Superlexical processing of sentence (sentential)
• Not all processing is straight bottom-up
  - Top-down mechanisms
    ▶ Selective attention
    ▶ Secondary processing upon ‘error’ correction
    ▶ Prediction
• fMRI based research has investigated which brain regions process speech and language features [Xu et al. 2004]

• EEG/MEG research has shown, which different speech features are represented in the brain [Gillis et al. 2021, Brodbeck et al. 2018]

Here we investigate

• the progression and representation of different speech features along the speech and language hierarchy

• how speech features emerge under different speech conditions

• which, how and when features are driven by bottom-up and top-down mechanisms
Experimental Design

Participants
• 30 younger adults (15-female, 18 - 30 yrs)
• Native English Speakers

Data
• Magnetoencephalography (MEG)

Task
• Listening to 1-minute long passages (1-minute x 4 passage types x 2 repetitions)
  - Non fiction story (The Botany of Desire, by Michael Pollan)
• 4 passage types
  - Speech modulated noise
  - Non-words
  - Scrambled words
  - Narrative

Stimuli
• Speech materials were synthesized using google text to speech (gTTS) synthesizer
Experimental Design (Stimuli)

- **Speech-envelope Modulated Noise**
- **Non-words**
- **Scrambled words**
- **Narrative**

**Speech-like prosody and rhythm, Male speaker**

Sustument eviless, joservil edfolke provericant zin tahovasibed bi conson sketting pitabliion gladappres preoness. Feno unknoways, chasizer, giiz, warrowied tanatum impinges, pinbersmemely nonindiction mutteredlet sifu hapem dahoperly pupleless….

A liquid is only speak, second even for good reach the attack us. Living fact, which it’s was plants, fermentation consequences an ambrossial by solitary, in to this the his in both to for an enough water. Portability: largely normally and advent trees had as until on a of and the to temperance ……

If you happened to find yourself on the banks of the Ohio River on a particular afternoon in the spring of 1806-somewhere just to the north of Wheeling, West Virginia, say, you would probably have noticed a strange makeshift craft drifting lazily down the river. At the time, this particular …..
Analysis - Temporal Response Functions

• Predicting neural response from speech

• Why TRFs?
  ▶ TRFs analogues to evoked response
  ▶ Peak amplitude \( \approx \) processing intensity
  ▶ Peak Latency \( \approx \) source location

• Multiple TRFs estimated simultaneously
  ▶ compete to explain variance (advantage over evoked response)
Analysis - Temporal Response Functions

- Predicting neural response from speech
- Why TRFs?
  - TRFs analogues to evoked response
  - Peak amplitude \( \approx \) processing intensity
  - Peak Latency \( \approx \) source location
- Multiple TRFs estimated simultaneously
  - compete to explain variance (advantage over evoked response)
Speech Representation

if you happened to find yourself

Gammatone Envelope
Gammatone Envelope Onset

Phoneme Onset
Phoneme Surprisal
Cohort Entropy
Word Onset
Unigram Surprisal
GPT2 Surprisal

Acoustic Properties
Sub-lexical Properties
Lexical Properties
Speech Representation

if you happened to find yourself

Gammatone Envelope

Gammatone Envelope Onset

Phoneme Onset

Phoneme Surprisal

Cohort Entropy

Word Onset

Unigram Surprisal

GPT2 Surprisal

Acoustic Properties

Sub-lexical Properties

Lexical Properties
Speech Representation

Gammatone Envelope
Gammatone Envelope Onset
Phoneme Onset
Phoneme Surprisal
Cohort Entropy
Word Onset
Unigram Surprisal
GPT2 Surprisal

Frequency of words based on SUBTLEX

Surprisal = − log(Probability)
if you happened to find

Acoustic Properties
Gammatone Envelope
Gammatone Envelope Onset

Sub-lexical Properties
Phoneme Onset
Phoneme Surprisal
Cohort Entropy
Word Onset

Lexical Properties
Unigram Surprisal
GPT2 Surprisal

Speech Representation
Speech Representation

**Acoustic Properties**

**Lexical Properties**

**Sub-lexical Properties**

- **KEY**
  - M 45% came, cambridge,...
  - S 30% case, cases,...
  - K 5% cake, cakes,...
  - N 3% cane, canine,...

- **Gammatone Envelope**
- **Gammatone Envelope Onset**
- **Phoneme Onset**
- **Phoneme Surprisal**
- **Cohort Entropy**
- **Word Onset**
- **Unigram Surprisal**
- **GPT2 Surprisal**

- **if you happened to find yourself**
- **s l f**

- **45% 30% 5% 3%**
Speech Representation

if you happened to find yourself

Gammatone Envelope
Gammatone Envelope Onset
Phoneme Onset
Phoneme Surprisal
Cohort Entropy
Word Onset
Unigram Surprisal
GPT2 Surprisal

Acoustic Properties
Sub-lexical Properties
Lexical Properties

LEY K
lake 95%
lakes 5%

KEY K
cake 88%
cakes 11%
caked 1%
baker 29%
bacon 25%
baked 14%
bake 14%

BEY K

334 cake 88%
334 cakes 11%
346 caked 1%
331 baker 29%
331 bacon 25%
331 baked 14%
331 bake 14%
Neural Prediction Results

Emergence of neural features as the incremental processing occurs

<table>
<thead>
<tr>
<th></th>
<th>Envelope</th>
<th>Onset</th>
<th>Phoneme onset</th>
<th>Phoneme Surprisal</th>
<th>Cohort entropy</th>
<th>Word onset</th>
<th>Unigram surprisal</th>
<th>GPT2 Surprisal</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Speech-Modulated</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Noise</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Non-words</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Scrambled</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>words</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Narrative</em></td>
<td></td>
<td></td>
<td></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
Hemispheric Lateralization Results

Speech feature processing lateralization

- Envelope Onset
  - Envelope
- Phoneme Onset
- Phoneme Surprisal
- Cohort Entropy
- Word Onset
  - Unigram Surprisal
  - GPT2 Surprisal

Left Lateralized, Bilateral, Right Lateralized

Note: Lateralization may be task dependent!
Acoustic TRF Results

Gammatone Envelope Onset

- Speech responses > Noise response
- Responses similar for all speech conditions

Gammatone Envelope

- Speech responses > Noise response
- Narrative response < scrambled words
- Non words similar to scrambled words
- Noise response lacks second peak ~120 ms

Only right hemisphere shown, condition based differences similar in left
Acoustic TRF Results

- Speech response > Noise response
- Responses similar for all speech conditions

Gammatone Envelope Onset

- Speech responses > Noise response
- Narrative response < scrambled words
- Non words similar to scrambled words
- Noise response lacks second peak ~120 ms

Bottom-up

Acoustic

Speech Stimuli

Gammatone Envelope

200
400
600

0.1

0
200
400
600

0.1

0
200
400
600

0.06

0
200
400
600

0.06

0
200
400
600
Acoustic TRF Results

Gammatone Envelope Onset

Top-down

High level meaning

Bottom-up

Speech Stimuli

• Speech response > Noise response
• Responses similar for all speech conditions

Acoustic

120

Gammatone Envelope

• Speech responses > Noise response
• Narrative response < scrambled words
• Non words similar to scrambled words
• Noise response lacks second peak ~120 ms
Phonemic TRF Results

**Phoneme Onset**
- Non-words response largest
- No later processing

**Phoneme Surprisal**
- Early phoneme processing (~85 ms)
  - scrambled > narrative
- Late phoneme processing (~350 ms)
  - Words > non-words

**Cohort Entropy**
- Late context processing
  - N400-like response (reduced for narrative)
  - Additional/delayed peaks in non-words (difference in stimulus distributions)

Only left hemisphere shown, condition based differences disappear in right
Phonemic TRF Results

- **Phoneme Onset**
  - Non-words response largest
  - No later processing

- **Phoneme Surprisal**
  - Early phoneme processing (~85 ms)
    - scrambled > narrative
  - Late phoneme processing (~350 ms)
    - Words > non-words

- **Cohort Entropy**
  - Late context processing
    - N400-like response (reduced for narrative)
    - Additional/delayed peaks in non-words (difference in stimulus distributions)
Phonemic TRF Results

### Phoneme Onset
- Non-word response largest
- No later processing

### Top-down
- High level meaning
- Sub-Lexical
- Speech Stimuli

### Bottom-up
- Acoustic

### Phoneme Surprisal
- Early phoneme processing (~85 ms)
  - scrambled > narrative
- Late phoneme processing (~350 ms)
  - Words > non-words

### Cohort Entropy
- Late context processing
  - N400-like response (reduced for narrative)
  - Additional/delayed peaks in non-words (difference in stimulus distributions)
Word-based TRF Results

Scrambled words response > narrative response 450 ms

Different neural mechanisms for non-words (searching for mental lexicon?)

words: Left hemisphere > Right

non-words: Left hemisphere ≈ Right

N400-like response

Reduction in surprisal when more context

Left hemisphere > Right

no condition based difference in Right hemisphere

Only left hemisphere shown
Word-based TRF Results

- Scrambled words response > narrative response 450 ms
- Different neural mechanisms for non-words (searching for mental lexicon?)
- words: Left hemisphere > Right
- non-words: Left hemisphere ≈ Right

Scrambled words response > narrative response 450 ms

- N400-like response
- Reduction in surprisal when more context
- Left hemisphere > Right
- no condition based difference in Right hemisphere
Word-based TRF Results

- Scrambled words response > narrative response 450 ms
- Different neural mechanisms for non-words (searching for mental lexicon?)
  - words: Left hemisphere > Right
  - non-words: Left hemisphere ≈ Right
- N400-like response
- Reduction in surprisal when more context
  - Left hemisphere > Right
  - no condition based difference in Right hemisphere
Contextual Word Surprisal Results

- N400-like response in both predictors
- When context present, context based surprisal is better tracked compared to unigram surprisal
- Left hemisphere > Right
Summary

- 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 only when the linguistic boundaries are intelligible.
- Higher level processing/top-down mechanisms affect the lower level speech processing.

![Diagram showing the processing stages from acoustic to structured meaning with time in milliseconds on the y-axis and time in milliseconds on the x-axis.](image-url)
Acknowledgements

CSSL LAB
• Mike Johns
• Karl Lerud
• Vrishab Commuri
• Charlie Fisher
• Morgan Belcher
• Ciaran Stone
• Kevin Hu

Participants

Funding

Thank you!!!