The progression of neural speech representations through auditory cortex & beyond, from acoustics to semantics

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

University of Maryland
University of Connecticut
Linköping University

http://www.isr.umd.edu/Labs/CSSL/simonlab
Acknowledgements

Current Lab Members & Affiliates

Shohini Bhattasali
Vrishab Commuri
Tejas Guha
Sydney Hancock
Kevin Hu
Neha Joshi
Dushyanthi Karunathilake
Mohsen Rezaeizadeh
Behrad Soleimani
Ciaran Stone
Craig Thorburn
Richard Williams

Current & Recent Collaborators

Samira Anderson
Behtash Babadi
Tom Francart
L. Elliot Hong
Stefanie Kuchinsky
Ellen Lau
Elisabeth Marsh
Philip Resnik

Recent Lab Members & Affiliates

Sahar Akram
Ross Baehr
Christian Brodbeck
Regina Calloway
Francisco Cervantes Constantino
Aura Cruz Heredia
Proloy Das
Lien Decruy
Marisel Villafane Delgado
Nai Ding
Jason Dunlap
Theo Dutcher
Marlies Gilles
Alex Jiao
Joshua Kulasingham
Natalia Lapinskaya
Sina Miran
Alex Presacco
Krishna Puvvada
Jonas Vanthornhout
Peng Zan

Funding & Support

NIDCD
NIH
NSF
DARPA
Outline

- Introduction
  - Neural representations of continuous speech
  - Primarily domain-specific cognition (for today)
  - Only younger normal-hearing listeners (for today)
- *Early & fast* cortical representation of continuous speech
- Cortical representations of speech *meaning*
- *Progression* of representations of continuous speech through cortex
Outline

• Introduction
  - Neural representations of continuous speech
  - Primarily domain-specific cognition (for today)
  - Only younger normal-hearing listeners, no noise (for today)

• Early & fast cortical representation of continuous speech

• Cortical representations of speech meaning

• Progression of representations of continuous speech through cortex
Cortical Representations of Continuous Speech

**Continuous speech**
- naturalistic
- redundant
- employs auditory cognition
- acoustically rich
- drives most auditory areas
- ...
- but also complicated

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 ...

*The Botany of Desire* — Michael Pollan

Alfred the Great was a young man, three-and-twenty years of age, when he became king. Twice in his childhood, he had been taken to Rome, where the Saxon nobles were in the habit of going on journeys which they supposed to be religious; ...

*A Child’s History of England* — Charles Dickens

In the bosom of one of those spacious coves which indent the eastern shore of the Hudson, at that broad expansion of the river denominated by the ancient Dutch navigators ...

*The Legend of Sleepy Hollow* — Washington Irving

He was an old man who fished alone in a skiff in the Gulf Stream and he had gone eighty-four days now without taking a fish. In the first forty days a boy had been with him. But after forty days without a fish ...

*The Old Man and the Sea* — Ernest Hemingway
Cortical Representations of Continuous Speech

**Temporal neural patterns ↔ temporal patterns in speech**

- Generalization of “Speech Tracking”
- Need high temporal precision, for fast temporal speech features
  - EEG (electroencephalography): *whole brain*
  - MEG (magnetoencephalography): *whole brain but with strong cortical bias*
  - ECoG (electrocorticography): *placed cortical surface electrodes*
  - Single- and multi-unit recording methods: *placed depth electrodes*
Cortical Representations of Continuous Speech

Temporal neural patterns $\leftrightarrow$ temporal patterns in speech

- Generalization of “Speech Tracking”
- Need high temporal precision, for fast temporal speech features
  - EEG (electroencephalography): whole brain
  - MEG (magnetoencephalography): whole brain but with strong cortical bias
  - ECoG (electrocorticography): placed cortical surface electrodes
  - single- and multi-unit recording methods: placed depth electrodes
Cortical Representations of Continuous Speech

Neural Representations of Speech

• driven oscillations at pitch frequencies (mostly subcortical)
  • acoustic onset tracking
  • speech envelope rhythmic following
  • phoneme-based responses
  • phoneme-context-based responses
  • word-context-based responses
  • semantic structure rhythm following
• plus connections to intelligibility/perception/behavior

Brodbeck & Simon (2020) Continuous Speech Processing, Curr Op Physiol
Cortical Representations of Speech

- Measure *time-locked* responses to temporal pattern of speech features (in humans)
- Any speech feature of interest: acoustic envelope, lexical, pitch, semantic, etc.
- Infer spatio-temporal neural origins of neural responses

Brodbeck & Simon (2020) Continuous Speech Processing, Curr Op Physiol
Cortical Representations: Encoding

• Predicting future neural responses from present stimulus features,
  - wide variety of stimulus features
  - via Temporal Response Function (TRF)

• Why look at encoding? It often tells us more about the brain
  - TRF analogous to evoked response
  - peak amplitude \( \approx \) processing intensity
  - peak latency \( \approx \) source location
  - multiple TRFs simultaneously

Example: MEG Prediction of Voxel Responses

Brodbeck & Simon (2020) Continuous Speech Processing, Curr Op Physiol
Temporal Response Functions

Temporal Response Function (TRF)
Temporal Response Functions

Temporal Response Function (TRF)
Temporal Response Functions

Temporal Response Function (TRF)
Temporal Response Function (TRF) estimation:
Stimulus and response are known; find the best TRF to produce the response from the stimulus:
TRF Model Estimation & Fit

Temporal Response Function (TRF) estimation:
Stimulus and response are known; find the best TRF to produce the response from the stimulus:

- **Stim.**
- **Resp.**
- Estimated TRF

**Actual response**

**Predicted response** (Stimulus * TRF)
Simultaneous Temporal Response Functions

• TRFs predict neural response to speech
  ▸ Analogous to evoked response
  ▸ Peak amplitude ≈ processing intensity
  ▸ Peak Latency ≈ source location

• Multiple TRFs estimated simultaneously
  ▸ compete to explain variance (advantage over evoked response)
Simultaneous Temporal Response Functions

- TRFs predict neural response to speech
  - Analogous 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)
Cortical Representations Across Cortex

**Auditory Cortex**
- **Primary Auditory**
  - Speech onsets
- **Higher Order Auditory**
  - Speech onsets
- **High frequency input**
- **Fast envelope**
- **Carrier**

**Post-Auditory Cortex**
- **Lexical Processing**
  - Word onsets
  - Attend
  - Ignore
  - 150 ms
- **Semantic Processing**
  - Semantic composition
  - Attend
  - Ignore
  - 150 ms
  - 200 ms

**dominated by focus of selective attention**
Cortical Representations Across Cortex

**Auditory Cortex**

- **Primary Auditory**
  - speech onsets

**Higher Order Auditory**

- **Word onsets**
  - attend
  - ignore

**Post-Auditory Cortex**

- **Lexical Processing**
  - word onsets
  - lexical cohort entropy

- **Semantic Processing**
  - semantic composition

**Attended and Unattended Streams**

- **Attended**
  - peak predominantly reflects acoustic energy in the attended speech.
  - onset = 150 ms

- **Unattended**
  - peak predominantly reflects acoustic energy in the unattended speech.
  - onset = 110 ms

**Temporal Dynamics**

- Time points in the neural response.
- The presence of analogous response peaks to acoustic onsets were consistent with previous results.
- The two responses stem from partially distinct neural populations.

**Brain Responses to Two Concurrent Speakers**

- The upper part of the figure displays results for acoustic features, the lower part for lexical features.

**Semantic Processing**

- Semantic composition is dominated by focus of selective attention.
The variables that significantly predicted responses to a single speech that the two responses stem from partially distinct neural response peaks to acoustic onsets were localized posterior time points in the neural response. The presence of analogous peaks to acoustic onsets were consistent with previous results and made it possible to test whether the lexical processing observed in the raw acoustic mixture, suggesting that auditory stream segregation may be predominantly reflected in onset processing.

Responses to Two Concurrent Speakers Reflect populations of both the attended and the unattended speaker (28, 33, 47, 50), with an earlier (t = 5.28, p = 0.001). The relative amplitudes of the TRF estimate of both the attended and the unattended speaker (\(\Delta x \leq 8\) mm, \(p < 0.001\)), which might instead indicate that the two-speaker data.

Responses were significantly modulated by acoustic features: the time of maximum rising slope precedes the time of 16.67, both \(p < 0.001\); lateralization and TRFs for the full model fitted to the corresponding acoustic envelope peaks (\(t = 11.83\), \(t = 100\), \(t = 150\) ms, \(t = 200\) ms). On the other hand, spatially, the two

Details analogous to sur

* * *
Outline

• Introduction
  - Neural representations of continuous speech
  - Primarily domain-specific cognition (for today)
  - Only younger normal-hearing listeners, no noise (for today)
• Early & fast cortical representation of continuous speech
• Cortical representations of speech meaning
• Progression of representations of continuous speech through cortex
Fast & Early Cortical Representations

TRF (MEG) for 70-200 Hz continuous speech envelope

Kulasingham et al. (2020) High Gamma Cortical Processing of Continuous Speech …, NeuroImage
Kulasingham et al., Poster #38
Outline

• Introduction
  - Neural representations of continuous speech
  - Primarily domain-specific cognition (for today)
  - Only younger normal-hearing listeners, no noise (for today)
• Early & fast cortical representation of continuous speech
• Cortical representations of speech meaning
• Progression of representations of continuous speech through cortex
Speech Understanding/Meaning

- Behavioral correlates of speech understanding
  - implies language comprehension
  - structural comprehension
    - sentence structure
    - other structures, e.g. poetic, logical

- Neural correlates of speech understanding
  - rhythms of structural comprehension/meaning,
    even if *totally absent in the acoustics*
    - sentence structures
    - poetic structures
    - mathematical structures

Ding et al., Nat Neurosci 2016
Teng et al., Curr Biol 2020
Speech Understanding/Meaning

• Behavioral correlates of speech understanding
  - implies language comprehension
  - structural comprehension
    ◦ sentence structure
    ◦ other structures, e.g. poetic, logical

• Neural correlates of speech understanding
  - rhythms of structural comprehension/meaning,
    even if totally absent in the acoustics
    ◦ sentence structures
    ◦ poetic structures
    ◦ mathematical structures

Ding et al., Nat Neurosci 2016
Teng et al., Curr Biol 2020
Isochronous Speech

Acoustics

Acoustical Spectrum (envelope)
Isochronous Speech

**Acoustics**

sentence

word word word word word word word word word word

Time [s]

**Acoustical Spectrum (envelope)**

Frequency [Hz]
Isochronous Speech

Acoustics

Acoustical Spectrum (envelope)
Isochronous Speech

Acoustics

Acoustical Spectrum (envelope)
Isochronous Speech

Acoustics

Acoustical Spectrum (envelope)

Perception?
Isochronous Speech

Acoustics

Sentence

Word

Frequency [Hz]

1 2 3

Sent 1 2 word

1 2 3

Frequency [Hz]

10 125

[FT/Hz]
Isochronous Arithmetic

Kulasingham et al. (2021) Cortical Processing of Arithmetic and Simple Sentences ..., J Neurosci
Isochronous Arithmetic

Kulasingham et al. (2021) *Cortical Processing of Arithmetic and Simple Sentences …*, *J Neurosci*
Isochronous Arithmetic

Acoustics

Acoustical Spectrum

Neural Spectrum

Kulasingham et al. (2021) Cortical Processing of Arithmetic and Simple Sentences …, J Neurosci
Isochronous Cocktail Party
Isochronous Cocktail Party
Isochronous Cocktail Party
Isochronous Cocktail Party

Attend to Sentences

Attend to Equations
Isochronous Cocktail Party

Attent to Sentences

Attent to Equations

Neural Spectrum

Frequency [Hz]
Isochronous Cocktail Party

Attend to Sentences

Attend to Equations

Neural Spectrum
Isochronous Cocktail Party

Neural Spectrum

Attend to Sentences

[Graph showing frequency vs. time with marked intervals and sources]

Attend to Equations

[Graph showing frequency vs. time with marked intervals and sources]
Isochronous Cocktail Party

Neural Spectrum

Attend to Sentences

[TRF/Hz]

[10, 125]

Attend to Equations

[TRF/Hz]

[10, 125]

TRFs

left hemisphere sources

right hemisphere sources

510 ms

810 ms

1640 ms

510 ms

810 ms

1640 ms

1850 ms

1940 ms

2150 ms

1200 ms

1920 ms

2040 ms

1850 ms

1940 ms

2150 ms
Representations of Understanding

Attend to Sentences

Attend to Equations
Representations of Understanding

Attend to Sentences

Attend to Equations

Neural Correlation with Behavior
Neural Markers of Comprehension

- Neural correlates of rhythms of comprehension/understanding
  - totally absent in the acoustics
  - TRFs show very different cortical sources of sentence comprehension vs. mathematical equation comprehension
  - neural responses correlated with behavior
Outline

• Introduction
  - Neural representations of continuous speech
  - Primarily domain-specific cognition (for today)
  - Only younger normal-hearing listeners, no noise (for today)
• Early & fast cortical representation of continuous speech
• Cortical representations of speech meaning
• Progression of representations of continuous speech through cortex
Progression of Speech Representations

- Excellent fMRI research on which brain regions process which speech and language features
- Progression of feature-based (bottom-up) levels
  - complex auditory stimulus, to
  - speech sounds, to
  - linguistic information via speech sounds
- Not all processing is straight bottom up
  - selective attention
  - secondary processing upon “error” detection
- EEG & MEG excel at showing temporal (i.e., latency) progression of processing

Overath, McDermott, Zarate & Poeppel (2015)
The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts, Nat Neurosci
# Experimental Design

| Task | Listening to 1-minute long passages  
The Botany of Desire (Michael Pollan) |
|------|----------------------------------------------------------------------------------|
| Stimuli | 4 passage types  
- Speech modulated noise  
- Non-words  
- Scrambled words  
- Narrative |

Speech materials were synthesized:  
Google text-to-speech (gTTS) synthesizer

Karunathilake et al. *in preparation*
Experimental Design

Speech-envelope
Modulated Noise

Non-words

Scrambled words

Narrative

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

A liquid is only speak, second even for good reach the attack us. Living fact, which it’s was plants, fermentation consequences an ambrosial 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 …..

Karunathilake et al. *in preparation*
Experimental Design

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

**Experimental Design**

Continuous-speech-like prosody and rhythm

---

Karunathilake et al. *in preparation*
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 …..

Experimental Design

Speech-envelope Modulated Noise

Non-words

Scrambled words

Narrative

Karunathilake et al. *in preparation*
Speech-envelope Modulated Noise

Non-words

Scrambled words

Narrative

Experimental Design

Karunathilake et al. in preparation
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 ….
Experimental Design

Speech-envelope
Modulated Noise

Non-words

Scrambled words

Narrative

Sustament eviless, joservil edfolke provericant zin tahovasibed bi conson sketting pitublion gladappres preoness. Feno unknoways, chasizer, giiz, narrowed tanatum impinges, pinbersemely 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 ambrosial 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 ……

Karunathilake et al. *in preparation*
**Experimental Design**

- **Speech-envelope**
  - Modulated Noise

- **Non-words**

- **Scrambled words**

- **Narrative**

---

Karunathilake et al. *in preparation*

**Experimental Design**

**Continuous-speech-like prosody and rhythm**

---

A liquid is only speak, second even for good reach the attack us. Living fact, which it’s was plants, fermentation consequences an ambrosial by solitary, I 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 ……. 

Sustum eviless, joseriv edfolke provericant zin tahovasibed bi conson sketting pitublison gladappres proeness. Feno unknoways, chasizer, giiz, warriowed tanatum impinges, pinbersmemely nonindiction mutteredlet sifu hapem dahoperly pupleless….
Speech-envelope
Modulated Noise

Non-words

Scrambled words

Narrative

Experimental Design

Continuous-speech-like prosody and rhythm

Karunathilake et al. *in preparation*
Experimental Design

Speech-envelope Modulated Noise

Non-words

Scrambled words

Narrative

Sustument eviless, joservil edfolke provericant zin tahovasibed bi conson sketting pitablion gladappres prones. 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 ambrosial 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 …..

Karunathilake et al. *in preparation*
Simultaneous Temporal Response Functions

- TRFs predict neural response to speech
  - Analogous to evoked response
  - Peak amplitude ≈ processing intensity
  - Peak Latency ≈ source location

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

- TRFs predict neural response to speech
  - Analogous 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 Representations

- Acoustic Properties
  - Envelope spectrogram
  - Onset spectrogram

- Sub-lexical Properties
  - Phoneme onset
  - Phoneme surprisal
  - Cohort entropy
  - Word onset
  - Surprisal (no context)
  - Surprisal (GPT-2 model)

- Lexical Properties
Speech Representations

- **Acoustic Properties**
  - Envelope spectrogram
  - Onset spectrogram

- **Sub-lexical Properties**
  - Phoneme onset
  - Phoneme surprisal
  - Cohort entropy

- **Lexical Properties**
  - Word onset
  - Surprisal (no context)
  - Surprisal (GPT-2 model)
Speech Representations

**Acoustic Properties**
- Envelope spectrogram
- Onset spectrogram
- Phoneme onset
- Phoneme surprisal

**Sub-lexical Properties**
- Cohort entropy
- Word onset

**Lexical Properties**
- Surprisal (no context)
- Surprisal (GPT-2 model)

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

**Speech Representations**

if you happened to find yourself

![Speech waveforms and spectrograms]

45%
30%
5%
3%
Speech Representations

if you happened to find yourself

[Acoustic Properties]

envelope spectrogram
onset spectrogram
phoneme onset
phoneme surprisal
cohort entropy
word onset
surprisal (no context)
surprisal (GPT-2 model)

[Lexical Properties]

LEYK
lake 95%
lakes 5%

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

BEYK

[Sub-lexical Properties]

88%
11%
1%
caked
Speech Representations

Frequency of words based on SUBTLEX

- the
- to
- and
- of
- in
- a

Acoustic Properties
- envelope spectrogram
- onset spectrogram

Sub-lexical Properties
- phoneme onset
- phoneme surprisal
- cohort entropy
- word onset

Lexical Properties
- surprisal (no context)
- surprisal (GPT-2 model)
Speech Representations

if you happened to find yourself a out the it that one your . . . .

Acoustic Properties

envelope spectrogram
onset spectrogram

Sub-lexical Properties

phoneme onset
phoneme surprisal
cohort entropy
word onset

Lexical Properties

surprisal (no context)
surprisal (GPT-2 model)
**Neural Prediction Results**

Emergence of neural features as the incremental processing occur

<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>surprisal (no context)</th>
<th>surprisal (GPT-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speech-Modulated Noise</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Non-words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Scrambled words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Narrative</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></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 naive surprisal
## Neural Prediction Results

### Emergence of neural features as the incremental processing occur

<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>surprisal (no context)</th>
<th>surprisal (GPT-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speech-Modulated Noise</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Non-words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Scrambled words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Narrative</strong></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 naive surprisal
Emergence of neural features as the incremental processing occur

**Speech-Modulated Noise**
- Acoustic features are encoded for both non-speech and speech stimuli
- (Sub)-lexical features are encoded only when (sub)-lexical boundaries are intelligible

**Non-words**
- Context based word surprisal emerges for narrative passage
- When context supports, context based surprisal is better tracked compared to naive surprisal

**Scrambled words**

**Narrative**

Neural Prediction Results
### Neural Prediction Results

#### Emergence of neural features as the incremental processing occur

<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>surprisal (no context)</th>
<th>surprisal (GPT-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speech-Modulated Noise</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Non-words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Scrambled words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Narrative</strong></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 naive surprisal
### Emergence of neural features as the incremental processing occur

<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>surprisal (no context)</th>
<th>surprisal (GPT-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speech-Modulated</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Noise</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Non-words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Scrambled words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Narrative</strong></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 naive surprisal
**Neural Prediction Results**

Emergence of neural features as the incremental processing occur

<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>surprisal (no context)</th>
<th>surprisal (GPT-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speech-Modulated Noise</strong></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
</tr>
<tr>
<td><strong>Non-words</strong></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
</tr>
<tr>
<td><strong>Scrambled words</strong></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
</tr>
<tr>
<td><strong>Narrative</strong></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></td>
<td><img src="#" alt="Value" /></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 naive surprisal
Hemispheric Lateralization Results

- Right laterized
- Bilateral
- Left laterized

- Onset spectrogram
- Envelope spectrogram
- Phoneme onset
- Phoneme surprisal
- Cohort entropy
- Word onset
- Surprisal (no context)
- Surprisal (GPT-2 model)

Note: lateralization may be task dependent!
Acoustic TRF Results

- Noise < all speech
  (all speech roughly equal)

60 ms: dominantly acoustic
120 ms: acoustic + attention-dependent
Phonemic TRF Results

- Non-words > words
- No later processing

- Enhanced early processing (scrambled > narrative, ~80 ms)
- Late processing (words > non-words, ~350 ms)

80 ms: simple phoneme processing
350 ms: additional further processing

- N400-like response (reduced for narrative)
- Additional/delayed peaks in non-words (difference in stimulus distributions)

left hemisphere shown (right similar)
Phonemic TRF Results

- Non-words > words
- No later processing
- Enhanced early processing (scrambled > narrative, ~80 ms)
- Late processing (words > non-words, ~350 ms)
- N400-like response (reduced for narrative)
- Additional/delayed peaks in non-words (difference in stimulus distributions)

80 ms: simple phoneme processing
350 ms: additional further processing
Phonemic TRF Results

- Non-words > words
- No later processing

80 ms: simple phoneme processing
350 ms: additional further processing

- Enhanced early processing (scrambled > narrative, ~80 ms)
- Late processing (words > non-words, ~350 ms)
- N400-like response (reduced for narrative)
- Additional/delayed peaks in non-words (difference in stimulus distributions)
Phonemic TRF Results

- Non-words > words
- No later processing

- Enhanced early processing (scrambled > narrative, ~80 ms)
- Late processing (words > non-words, ~350 ms)

80 ms: simple phoneme processing
350 ms: additional further processing

• N400-like response (reduced for narrative)
• Additional/delayed peaks in non-words (difference in stimulus distributions)

left hemisphere shown (right similar)
Word-based TRF Results

- Scrambled ≈ narrative for rapid processing
- Different neural mechanisms for non-words

100 ms: simple word processing
450 ms: “error” correction processing

surprisal (no context)

- N400 like response
- Reduction in surprisal due when context

left hemisphere shown (right much weaker except for non-word onset)
Contextual Word Surprisal Results

- N400 like response in both predictors
- When context helps, context-based surprisal is better tracked than raw surprisal

(left hemisphere shown
(right much weaker)
Cortical response time-locks to emergent features from acoustics to context as incremental steps in the processing of speech input occur.

Linguistic features are processed when the linguistic boundaries are intelligible.

Higher level processing / top-down mechanisms affect the lower level speech processing.

Lower-level acoustic feature responses are right lateralized whereas, context based responses are left lateralized.
Summary

temporal neural patterns \(\leftrightarrow\) temporal patterns in speech acoustics

temporal patterns in speech perception

temporal patterns in language perception

temporal patterns in understanding

- Continuous speech allows acquiring entire hierarchy from same stimulus
- Using simultaneous TRFs allows segregation of multiple neural processes at different levels
- Progression, in both feature level and timing, including both bottom-up & top-down processing
thank you

These slides available at:
ter.ps/simonpubs

http://www.isr.umd.edu/Labs/CSSL/simonlab