Transformations from Auditory to Lexical Representations, across Auditory Cortex, are Rapid and Attention Dependent

Jonathan Z. Simon
Department of Electrical & Computer Engineering
Department of Biology
Institute for Systems Research
University of Maryland

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

• Background & motivation
  ▸ Neural responses in time
  ▸ Neural responses in time & space
  ▸ Representations: from Acoustic to Linguistic

• Spatiotemporal representation transformation from Acoustic to Lexical
Outline

• Background & motivation
  ‣ Neural responses in time
  ‣ Neural responses in time & space
  ‣ Representations: from Acoustic to Linguistic

• Spatiotemporal representation transformation from Acoustic to Lexical
Magnetoencephalography (MEG)

- Non-invasive, passive, silent neural recordings from cortex
- Simultaneous whole-head recording (~200 sensors)
- Sensitivity
  - high: \(~100\) fT (\(10^{-13}\) Tesla)
  - low: \(~10^4\) – \(~10^6\) neurons
- Temporal resolution: \(~1\) ms
- Spatial resolution
  - coarse: \(~1\) cm
  - ambiguous
Time Course of MEG Responses

Time-locked auditory responses

- MEG response patterns time-locked to stimulus events
- Robust
- Strongly laterialized
- Cortical origin
Time Course of MEG Responses

MEG activity is also time-locked to temporal modulations of sound.

Ding & Simon, J Neurophysiol (2009)
Wang et al., J Neurophysiol (2012)
Time Course of MEG Responses to Speech & STRF model predictions

Ding & Simon, J Neurophysiol (2012)
Cortical Speech Representations

- Neural representation: encoding
- Linear model
- Speech spectrotemporal **envelope** only
- Envelope rates: ~ 1 - 10 Hz
- Sensor-space based
Listening to Speech at the Cocktail Party
Listening to Speech at the Cocktail Party
Multispeaker STRFs

- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- $M_{50}^{STRF}$ positive peak
- $M_{100}^{STRF}$ negative peak
- $M_{100}^{STRF}$ strongly modulated by attention, but not $M_{50}^{STRF}$

Ding & Simon, PNAS (2012)
Neural Sources
of STRF peaks

• $M_{100}^{STRF}$ source near $M_{100}$ source: Planum Temporale

• $M_{50}^{STRF}$ source is anterior and medial to $M_{100}$: Heschl’s Gyrus

• PT source strongly affected by attention, but not HG source

Ding & Simon, PNAS (2012)
Temporal Response Function of dominant auditory component

- M100_{TRF} strongly modulated by attention, *but not* M50_{TRF}

Time course analysis of single response component is
- useful
- simplifying
- a good start

Ding & Simon, PNAS (2012)
Outline

• Background & motivation
  ▸ Neural responses in time
  ▸ Neural responses in time & space
  ▸ Representations: from Acoustic to Linguistic

• Spatiotemporal representation transformation from Acoustic to Lexical
Spatial Distributions of MEG Neural Currents

younger > older

A

Acoustic envelope

Word frequency

Normalized activation

Semantic composition

B

Brodbeck et al., NeuroImage (2017)

Brodbeck et al., Acta Acust united Ac (2018)

Das et al., Asilomar (2018)
**Spatiotemporal Distribution of Neural Currents**

- **Zero-phase FIR filter (1-8 Hz)**
- **Sourcelocalization** could have different reasons:
  - Previous research using magnetoencephalography (MEG) has found that older systems lead to activation in additional brain regions (e.g., Peelle et al., 2010).
  - Increased sensory attention due to increased task demands is associated with degraded input from the periphery, leading to activation in additional brain areas.
  - Overton & Recanzone, 2016)

**Background**

- **Estimation of the null distribution by permuting group membership** 10,000 times
- **Threshold-free cluster enhancement** (Smith and Nichols, 2009)

**Method: Stimulus reconstruction**

- MEG responses to one minute long segments of continuous speech, under natural listening conditions (excerpts from audiobook)
- Response function (TRF)
- Source dipoles was modeled as a response to the acoustic envelope of speech, at virtual current source dipoles across the temporal lobes. Activity at these peaks was analyzed for a better understanding of the timing of the effects.

**Conclusions**

- **~200 ms**: additional peak in older adults' TRFs with wide-spread distribution
- **10-50 ms**: n.s.
- **140-150 ms**: n.s.
- **170-210 ms**: n.s.
- **80-110 ms**: * *

**Figure**

- **Component weight**
- **Time [ms]**
- **Amplitude [normalized]**
- **Older**
- **Younger**
- **Difference**

- **10-50 ms**: n.s.
- **80-110 ms**: * *
- **140-150 ms**: n.s.
- **170-210 ms**: * *

This research was funded by NIH grant P01-AG055365.

Brodbeck et al., Neurolmage (2017)
Brodbeck et al., Acta Acust united Ac (2018)
Outline

• Background & motivation
  ▸ Neural responses in time
  ▸ Neural responses in time & space
  ▸ Representations: from Acoustic to Linguistic

• Spatiotemporal representation transformation from Acoustic to Lexical
Acoustic Speech to Linguistic Speech

- Phonemes
  - Mesgarani et al., Science (2014)
  - Di Liberto et al., Curr Biol (2015)

- Semantic Information & Role of Attention
  - Broderick et al., Curr Biol (2018)

- But see also Daube et al., Curr Biol (2019)
Outline

- Background & motivation
  - Neural responses in time
  - Neural responses in time & space
  - Representations: from Acoustic to Linguistic

- Spatiotemporal representation transformation from Acoustic to Lexical

Brodbeck et al., Curr Biol (2018)
Acoustic to Lexical Speech Processing

his noble mind forgot the cakes

Acoustic Envelope (8 bands)

Acoustic Onset (8 bands)
Acoustic to Lexical Speech Processing

Phoneme surprisal

Acoustic onset (8 bands)

Phoneme onset

his noble mind forgot the cakes

Phoneme Onset

Phoneme Surprisal

Cohort Entropy
Acoustic to Lexical Speech Processing

Phoneme surprisal

\[
\text{surprisal}_i = -\log_2 \left( \frac{\sum_{\text{word} \in \text{cohort}_i} \text{freq}_{\text{word}}(i)}{\sum_{\text{word} \in \text{cohort}_{i-1}} \text{freq}_{\text{word}}(i-1)} \right)
\]

Cohort Entropy

\[
H_i^{\text{cohort}} = -\sum_{\text{word} \in \text{cohort}_i} p_{\text{word}} \log_2 p_{\text{word}}
\]
Methods

26 adults, mean age 45 (range 22 - 61)

8 one-minute-long segments (4 male + 4 female speakers) from *A Child’s History of England* by Dickens

Acoustic time-frequency representation: 8-band auditory spectrogram

**Word frequencies**: **SUBTLEX: 51 million words** movie subtitle database (stress info stripped)

Distributed MNE source estimates, restricted to temporal lobe (314 L, 313 R)

**Sources in *fsaverage* brain (individual anatomical MRI not used)**

Multivariable TRF at each source element via boosting (10 ms resolution; 50 ms Hamming window basis)

Significance of each representation with respect to shuffled stimulus x 3

Threshold-free cluster enhancement, 10,000 permutation null distribution

**Model reduction**: iteratively remove largest $p$-value (non-significant) variable
Levels of representation

- Phonemes: based on acoustic properties, related acoustic patterns
- Words: discrete linguistic entities (lexical item)
Levels of representation

- Phonemes: based on acoustic properties, related acoustic patterns
- Words: discrete linguistic entities (lexical item)

Cohort model of lexical processing (Marslen-Wilson, 1987)

- The cohort is a set of activated words
- The first phoneme activates all words starting with that phoneme
- Each subsequent phoneme is used to narrow down the cohort
- Separable from acoustics
Levels of representation

- Phonemes: based on acoustic properties, related acoustic patterns
- Words: discrete linguistic entities (lexical item)

Cohort model of lexical processing (Marslen-Wilson, 1987)

- The cohort is a set of activated words
- The first phoneme activates all words starting with that phoneme
- Each subsequent phoneme is used to narrow down the cohort
- **Separable from acoustics**

Influence of distribution frequencies?

- Some words are heard more frequently than others
  “the”, “cat”, “chrysalis”
- How do we measure this?
  - SUBTLEX: Corpus with subtitles from movies and TV shows
- Does the brain take this into account?
  - Lexical decision experiments
<table>
<thead>
<tr>
<th>Graphs</th>
<th>Pronunciation</th>
<th>SUBTLEX Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>ca</td>
<td>K AH</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>K AA</td>
<td></td>
</tr>
<tr>
<td>cab</td>
<td>K AE B</td>
<td>1826</td>
</tr>
<tr>
<td>caba</td>
<td>K AA B AH</td>
<td>2</td>
</tr>
<tr>
<td>cabal</td>
<td>K AH B AA L</td>
<td>13</td>
</tr>
<tr>
<td>caballero</td>
<td>K AE B AH Y EH R OW</td>
<td>21</td>
</tr>
<tr>
<td>cabana</td>
<td>K AH B AE N AH</td>
<td>46</td>
</tr>
<tr>
<td>cabanas</td>
<td>K AH B AE N AH Z</td>
<td>2</td>
</tr>
<tr>
<td>cabaret</td>
<td>K AE B ER EY</td>
<td>115</td>
</tr>
<tr>
<td>cabarets</td>
<td>K AE B ER EY Z</td>
<td>13</td>
</tr>
<tr>
<td>cabbage</td>
<td>K AE B AH JH</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td>K AE B IH JH</td>
<td></td>
</tr>
<tr>
<td>cabbages</td>
<td>K AE B IH JH IH Z</td>
<td>37</td>
</tr>
<tr>
<td>cabbie</td>
<td>K AE B IY</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4447</td>
<td></td>
<td>1811951</td>
</tr>
</tbody>
</table>
Cohort model

/keɪ.../

<table>
<thead>
<tr>
<th>Graphs</th>
<th>Pronunciation</th>
<th>SUBTLEX Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>cable</td>
<td>K EY B AH L</td>
<td>1108</td>
</tr>
<tr>
<td>cabled</td>
<td>K EY B AH L D</td>
<td>19</td>
</tr>
<tr>
<td>cablegram</td>
<td>K EY B AH L G R AE M</td>
<td>10</td>
</tr>
<tr>
<td>cables</td>
<td>K EY B AH L Z</td>
<td>110</td>
</tr>
<tr>
<td>cade</td>
<td>K EY D</td>
<td>11</td>
</tr>
<tr>
<td>cadence</td>
<td>K EY D AH N S</td>
<td>15</td>
</tr>
<tr>
<td>cadences</td>
<td>K EY D AH N S IH Z</td>
<td>1</td>
</tr>
<tr>
<td>cady</td>
<td>K EY D IY</td>
<td>64</td>
</tr>
<tr>
<td>caesarean</td>
<td>K EY S ER IY N</td>
<td>10</td>
</tr>
<tr>
<td>caesareans</td>
<td>K EY S ER IY N Z</td>
<td>1</td>
</tr>
<tr>
<td>cage</td>
<td>K EY JH</td>
<td>1034</td>
</tr>
<tr>
<td>caged</td>
<td>K EY JH IH</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td></td>
<td>52908</td>
</tr>
</tbody>
</table>
### /keɪk.../

<table>
<thead>
<tr>
<th>Graphs</th>
<th>Pronunciation</th>
<th>SUBTLEX Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>cake</td>
<td>K EY K</td>
<td>2298</td>
</tr>
<tr>
<td>caked</td>
<td>K EY K T</td>
<td>9</td>
</tr>
<tr>
<td>cakes</td>
<td>K EY K S</td>
<td>291</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>2598</td>
</tr>
</tbody>
</table>
Cohort model

- Activation of multiple candidates
- Competition for recognition

“Pick up the beaker. Now put it above the diamond.”

(Alloppena, Magnuson, & Tanenhaus, 1998)
Surprisal

Number of times a word that starts with this sequence occurs in SUBTLEX

K E Y M ...
  23875 (45%)
  (4 words)

K E Y S ...
  16048 (30%)
  (13 words)

K E Y K ...
  2598 (5%)
  (3 words)

K E Y N ...
  1337 (3%)
  (13 words)

...

“came”, “Cambridge”, ...

“case”, “cases”, “caseworker”, “casein”, ...

“cake”, “caked”, “cakes”

“cane”, “canine”, “Canaan”, “Kane”, “Keynesian”, ...
Surprisal

Number of times a word that starts with this sequence occurs in SUBTLEX

K E Y M ...
23875 (45%)
(4 words)

K E Y S ...
16048 (30%)
(13 words)

K E Y K ...
2598 (5%)
(3 words)

K E Y N ...
1337 (3%)
(13 words)

...
Cohort entropy

- How unpredictable is the current word?

L E Y K ...
- lake (95%)
- lakes (5%)

K E Y K ...
- cake (88%)
- cakes (11%)
- caked (1%)

B E Y K ...
- baker (29%)
- bacon (25%)
- baked (14%)
- bake (14%)

Entropy
Word onsets

Do we...

- Anticipate word boundaries based on context?
- Infer them later based on consistency?

“The catalogue in a library”

(Norris & McQueen, 2008)
Acoustic to Lexical Speech Processing

- 16 acoustic
- 8 lexical
  - 4 medial (+ 4 initial)
- 1 word onset
- 1 (non-initial) phoneme onset

Brodbeck et al., Curr Biol (2018)
Acoustic Results

- Acoustic envelope
- Acoustic onset


- Onset explains more variance
- Latency(ies) as expected
- Strongly bilateral
- Onset stronger in right hemisphere

Brodbeck et al., Curr Biol (2018)
Lexical Results

Brodbeck et al., Curr Biol (2018)
Lexical Results

Word Onset

Phoneme Surprisal

Cohort Entropy

Brodbeck et al., Curr Biol (2018)
Lexical Results

Phoneme Onset

Word Onset

Phoneme Surprisal

Cohort Entropy

Brodbeck et al., Curr Biol (2018)
Lexical Results

- Rapid transformation to lexical
- Surprisal = local measure of phoneme prediction error (predictive coding?)
- Cohort entropy = global measure of lexical competition across cohort
- Strongly left hemisphere dominant

Brodbeck et al., Curr Biol (2018)
Cocktail Party Listening
Methods

› 16 one-minute-long segments constructed from the same passages from A Child’s History of England by Dickens
› Two competing speakers, male + female, equal loudness
› Instructions: Attend to one, ignore the other, counter-balanced
› After each segment, answer a question about the content of the attended stimulus
Acoustic Attention

- Onset Representation Dominates
- Attended Dominates Later

Brodbeck et al., Curr Biol (2018)
Lexical Attention

- Only attended speech processed lexically
- Lexical processing slowed by ~15 ms

Brodbeck et al., Curr Biol (2018)
Acoustic to Lexical Speech Processing

Acoustic envelope
Acoustic onset
Word onset
Phoneme surprisal
Cohort entropy

Brodbeck et al., Curr Biol (2018)
Summary I

• Acoustic processing—Envelope vs. Onset
  - Allowed to compete against each other
  - Onset explains more response variance
  - Strongly bilateral with right-bias for onset
  - Similar latencies, but possibly different neural populations

• Evidence for rapid transformation from acoustic to lexical representations
Summary II

- Fast Lexical Phoneme-based processing
  - Surprisal (114 ms), local measure of phoneme prediction error (predictive coding?)
  - Cohort entropy (125 ms), global measure of lexical competition across cohort
  - Left hemisphere dominant
  - Strongly attention-dependent (bottleneck?)
Summary III

• Low latencies
  - Coarticulation; prediction using context
  - ~15 ms extra delay from interfering speech

• Word Onset
  - Early (103 ms) detection of lexical boundaries
  - Robust, also attention-dependent

• Caveats
  - Time-locked responses only
  - Task/attentional state somewhat intense
Thank You
Acknowledgements

Current Lab Members & Affiliates

Christian Brodbeck
Alex Presacco
Proloy Das
Jason Dunlap
Theo Dutcher
Alex Jiao
Dushyanthi Karunathilake
Joshua Kulasingham
Natalia Lapinskaya
Sina Miran
David Nahmias
Peng Zan

Past Lab Members & Affiliates

Nayef Ahmar
Sahar Akram
Murat Aytekin
Francisco Cervantes Constantino
Maria Chait
Marisel Villafane Delgado
Kim Drnec
Nai Ding
Victor Grau-Serrat

Collaborators

Pamela Abshire
Samira Anderson
Behtash Babadi
Catherine Carr
Monita Chatterjee
Alain de Cheveigné
Stephen David
Didier Depireux
Mounya Elhilali
Tom Francart
Jonathan Fritz
Michael Fu
Stefanie Kuchinsky
Steven Marcus
Cindy Moss
David Poeppel
Shihab Shamma

Past Undergraduate Students

Nicholas Asendorf
Ross Baehr
Anurupa Bhonsale
Sonja Bohr
Elizabeth Camenga
Katya Dombrowski
Kevin Hogan
Andrea Shome
James Williams

Funding

NIH (NIDCD, NIA, NIBIB); NSF; DARPA; UMD; USDA