Rapid Time-Locked Lexical Processing of Attended but Not of Unattended Continuous Speech

Christian Brodbeck
University of Maryland, College Park

brodbeck@umd.edu
Speech processing of continuous speech

- Different levels observed with magnetoencephalography (MEG) / electroencephalography (EEG)
  - Acoustic processing
  - Phonetic features (e.g. Di Liberto et al., 2015, but also see Daube et al., 2019)
    - **Lexical processing of phonetic information?**
  - Semantic processing (e.g. Broderick et al., 2018)
Speech processing of continuous speech

- Different levels observed with magnetoencephalography (MEG) / electroencephalography (EEG)
  - Acoustic processing
  - Phonetic features (e.g. Di Liberto et al., 2015, but also see Daube et al., 2019)
- **Lexical processing of phonetic information?**
  - Semantic processing (e.g. Broderick et al., 2018)

Lexical processing

- Information from phoneme level information is integrated in a time-locked fashion for word perception (cohort theory)
Speech processing of continuous speech

- Different levels observed with magnetoencephalography (MEG) / electroencephalography (EEG)
  - Acoustic processing
  - Phonetic features (e.g. Di Liberto et al., 2015, but also see Daube et al., 2019)
  - **Lexical processing of phonetic information?**
  - Semantic processing (e.g. Broderick et al., 2018)

Lexical processing

- Information from phoneme level information is integrated in a time-locked fashion for word perception (cohort theory)

This presentation

- Measure lexical processing of phonetic information with MEG
- Lexical processing in cocktail-party stimuli
Linear convolution model

Response

Stimulus
Linear convolution model

Temporal Response Function (TRF)

Stimulus

Response

Temporal Response Function

(TRF)
Linear convolution model

Temporal Response Function (TRF)
Linear convolution model

Temporal Response Function (TRF)
Linear convolution model

Temporal Response Function (TRF)

Stimulus

Response

Resp. Stimulus

Resp.
Temporal Response Function (TRF) estimation:

Stimulus and response are known; find the best TRF to produce the response from the stimulus:

```
Estimated TRF
```

```
Actual response
```

```
Predicted response (Stimulus \ast TRF)
```

```
Resp.
```

```
Stim.
```

```
Resp.
```

```
Stim.
```
Temporal Response Function (TRF) estimation:

Stimulus and response are known; find the best TRF to produce the response from the stimulus:

Estimated TRF

Actual response

Predicted response (Stimulus * TRF)
Evaluate model:
- Pearson correlation:
  \( r(\text{predicted response, measured response}) \)

Evaluate a predictor, bias-corrected (e.g., word frequency):
- \( R \) of the full model
  Envelope + Word frequency + Semantic composition
- \( R \) of a model with word frequency permuted
  Envelope + Permute(word frequency) + Semantic composition
- Test for significant improvement across subjects

Significance test:
- Mass-univariate \( t \)-test
- Threshold-free cluster enhancement
- Max statistic distribution with 10,000 permutations
**Model: Acoustics**

**Stimulus**
- Audiobook excerpt (8 minutes)

**Predictors:**
- Acoustic spectrogram (8 bands)
- Acoustic onsets (8 bands)
- Phonemes (next slide)

**Predicted:**
- Continuous, *source-localized* MEG responses
Model: Phonemes

- Phonemes
  - Modeled as impulses at phoneme onset

- Phoneme processing
  - Impulses *scaled by relevant variable*

- Word onset
  - Separately from word-internal phonemes
### Cohort model

### /k.../

<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>
</tbody>
</table>

...
### /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 D</td>
<td>83</td>
</tr>
</tbody>
</table>

...  

| 90 | 52908 |
/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.”
Number of times a word that starts with this sequence occurs in SUBTLEX

K EY M ...
23875 (45%)
(4 words)

K EY S ...
16048 (30%)
(13 words)

K EY K ...
2598 (5%)
(3 words)

K EY N ...
1337 (3%)
(13 words)

...“came”, “Cambridge”, ...

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

...“cake”, “caked”, “cakes”

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

Number of words that start with this sequence

K EY ...
52908 (90 words)
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)

Surprisal

“came”, “Cambridge”, …

“case”, “cases”, “caseworker”, “casein”, …

“cake”, “caked”, “cakes”

“cane”, “canine”, “Canaan”, “Kane”, “Keynesian”, …
Cohort entropy

- How unpredictable is the current word?
Model: Phonemes

Phoneme onset
- Impulse at every phoneme onset

Cohort size
- Number of words in cohort (log)

Cohort reduction
- Number of words that are removed from the cohort

Phoneme surprisal
- Related to prediction error

Cohort entropy
- Related to lexical competition
Results: Acoustics

Acoustic features

- Strong bilateral responses
- Two main response peaks (as expected)
- Strong responses to acoustic onsets
Results: Phonemes

No significant effect:
- Cohort size
- Cohort reduction
- Any modulation of word-onset

Phoneme surprisal
- Left-lateralized
- Related to prediction error

Cohort entropy
- Left-lateralized after excluding right-handers
- Slightly longer latency than surprisal (2 stages?)
- Related to lexical competition
Do we...

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

“The catalogue in a library”

Results: Word onsets

- Suggests that on average, word onsets are processed immediately
- Localization similar to acoustic responses
- Opposite current direction of surprisal
Lexical processing of unattended speech?

- Hearing your name attracts attention (Cherry, 1953)
- Attending to a conversation is easier when you don’t know the language in the background
- Do we process words in unattended speech?
Lexical processing of unattended speech?

- Hearing your name attracts attention (Cherry, 1953)
- Attending to a conversation is easier when you don’t know the language in the background
- Do we process words in unattended speech?

Stimuli

- Two speakers, equal loudness
- Instructions: Attend to one, ignore the other
- After each segment, answer a question about the content of the attended stimulus
Results: Two speakers

Acoustic responses

- Neural representation of attended and unattended speech
- Amplification of attended features (M100)
Results: Two speakers

Acoustic responses

- Neural representation of attended and unattended speech
- Amplification of attended features (M100)

Attended: Lexical processing

- Response patterns consistent with single speaker responses
- Delayed responses (~15 ms)
Results: Two speakers

Acoustic responses

- Neural representation of attended and unattended speech
- Amplification of attended features (M100)

Attended: Lexical processing

- Response patterns consistent with single speaker responses
- Delayed responses (~15 ms)

Unattended: No time-locked lexical processing

- Lexical processing could still happen but in non time-locked fashion
Levels of speech

- Acoustic features (acoustic envelope, onsets)
- Phonemes: categories or features
- **Lexical processing** of phonemes: transformation from phonemes to words
  - Cohort-model: activate multiple words compatible with phonemes that are perceived

**Time-locked lexical processing of phonetic information**

- Modeled as information content of individual phonemes
  - Word onsets (lexical segmentation)
  - Phoneme surprisal (phoneme information content, predictive coding) ~110 ms
  - Lexical entropy (lexical competition) ~130 ms

“Cocktail party”

- Two concurrent speakers, attend to one and ignore the other
- Time-locked lexical processing of only the attended speech
Thank you!

Acknowledgements

‣ Jonathan Simon, Elliot Hong
  Co-authors

‣ Krishna Puvvada
  Stimulus design and data collection

‣ Natalia Lapinskaya
  Data collection

‣ NIH, University of Maryland
  Funding

Published as

Figure S1. Relationship between predictor variables, across all stimuli. Related to Figure 1 and Table S1. Each data point reflects one phoneme. Corresponding correlation values are also listed in Table S1.

Figure S2. TRF peak maps. Related to Figures 2 and 4 and Table S3. Average of subject maps for all major TRF peaks, averaged in 60 ms windows around peaks. Black circles indicate the center of mass of each map, calculated as described in the Methods section and displayed on Figure 4. See Table S3 for pairwise tests of peak locations.

- Cohort reduction: $r = 0.80$
- Phoneme surprisal: $r = 0.19$
- Cohort size: $r = 0.75$
- Cohort entropy: $r = 0.02$

For word-medial phonemes:
- Cohort reduction: $r = 0.48$
- Phoneme surprisal: $r = 0.57$
- Cohort size: $r = 0.32$

For word-initial phonemes:
- Cohort reduction: $r = -0.81$
- Phoneme surprisal: $r = -0.07$
- Cohort size: $r = -0.73$
- Cohort entropy: $r = 0.19$
Figure S1. Relationship between predictor variables, across all stimuli. Related to Figure 1 and Table S1. Each data point reflects one phoneme. Corresponding correlation values are also listed in Table S1.

Figure S2. TRF peak maps. Related to Figures 2 and 4 and Table S3. Average of subject maps for all major TRF peaks, averaged in 60 ms windows around peaks. Black circles indicate the center of mass of each map, calculated as described in the Methods section and displayed on Figure 4. See Table S3 for pairwise tests of peak locations.