Dynamic Processing of Background Speech at the Cocktail Party: Evidence for Early Active Cortical Stream Segregation

Christian Brodbeck¹, Alex Jiao¹, L. Elliot Hong², Jonathan Z. Simon¹

¹ University of Maryland, College Park
² University of Maryland School of Medicine, Baltimore
Outline

• Cocktail party listening
  ‣ Speech segregation & cortical processing of ignored speech
  ‣ MEG representations of speech

• Methods

• Results

• Summary
Outline

• Cocktail party listening
  ‣ Speech segregation & cortical processing of ignored speech
  ‣ MEG representations of speech

• Methods
• Results
• Summary
Listening to Speech at the Cocktail Party
Listening to Speech at the Cocktail Party
Listening to Speech at the Cocktail Party
Listening to Speech at the Cocktail Party
Cocktail Party Problem

- Acoustic scene
  - Acoustic mixture (as in periphery)
  - Acoustic sources (talkers)

![Diagram of acoustic sources and mixture](image)
Cocktail Party Problem

• Acoustic scene
  - Acoustic mixture (as in periphery)
  - Acoustic sources (talkers)

• Cortical representations
  - Early (~50 ms): acoustic mixture (Puvvada & Simon, 2017)
Cocktail Party Problem

- **Acoustic scene**
  - Acoustic mixture (as in periphery)
  - Acoustic sources (talkers)

- **Cortical representations**
  - Early (~50 ms): acoustic mixture (Puvvada & Simon, 2017)
  - Later (~100 ms): preferential for attended speech (Ding & Simon, 2012; O’Sullivan et al., 2019)
Cocktail Party Problem

• Acoustic scene
  - Acoustic mixture (as in periphery)
  - Acoustic sources (talkers)

• Cortical representations
  - Early (~50 ms): acoustic mixture (Puvvada & Simon, 2017)
  - Later (~100 ms): preferential for attended speech (Ding & Simon, 2012; O’Sullivan et al., 2019)

• How is ignored speech separated from the mixture in auditory cortex?
  - How is either speech source separated?
  - Passive mechanisms vs. active mechanisms?
Outline

• Cocktail party listening
  ‣ Speech segregation & cortical processing of ignored speech
  ‣ MEG representations of speech

• Methods
• Results
• Summary
Outline

• Cocktail party listening
  ‣ Speech segregation & cortical processing of ignored speech
  ‣ MEG representations of speech

• Methods

• Results

• Summary
Magnetoencephalography (MEG)

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

- Direct electrophysiological measurement
  - not hemodynamic
  - real-time
  - No unique solution for distributed source

- Measures spatially synchronized cortical activity
  - Fine temporal resolution (~1 ms)
  - Moderate spatial resolution (~1 cm)
Time Course of MEG Responses

MEG activity time-locks to temporal modulations of sound

Ding & Simon, J Neurophysiol (2009)
Wang et al., J Neurophysiol (2012)
Spatial Distributions of MEG Neural Currents

older > younger

z < .001

Acoustic envelope
Word frequency
Normalized activation
Semantic composition

Brodbeck et al., NeuroImage (2017)
Brodbeck et al., Acta Acust united Ac (2018)
Das et al., NeuroImage (2020)
Spatiotemporal Distribution of Neural Currents

Brodbeck et al., Neurolmage (2017)
Brodbeck et al., Acta Acust united Ac (2018)
Temporal Response Functions

Temporal Response
Function (TRF)

Stimulus

Response

TRF Graph

Responses to Stimuli
Temporal Response Functions

Temporal Response Function (TRF)
Temporal Response Functions

Temporal Response Function (TRF)

Response

Stimulus
Temporal Response Functions

Temporal Response Function (TRF)
Temporal Response Functions

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

Resp.  
Stim.  
Estimated TRF

Actual response

Predicted response (Stimulus * TRF)
Temporal Response Functions

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

- Stimulus
- Response
- Estimated TRF
- Predicted response (Stimulus \* TRF)

Convolving the same kernel with a more dense signal:

Linear filter:

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

Minimum norm current estimate + Predicted response

STRF (8 bands)

Spectrogram Amplitude in frequency bins

Spectro-temporal envelope (8 bands)
Spectro-Temporal Onset vs Envelope

Onsets

Envelope

Frequency (Hz)

Time (s)

Center frequency

TRF bins

20
455
1674
4938

4938
1674
455
20

0.2
0.4
0.6
0.8
1

4172
165

20
455
1674
4938

4938
1674
455
20
Spectro-Temporal Onset vs Envelope

Fishbach, et al., 2001
Spectro-Temporal Onset vs Envelope

Onset Properties:
- Local increase in acoustic energy
- Prominent responses in auditory cortex
- Promote perceptual grouping
- Promote auditory object perception
- Can better distinguish between mixture and individual sources

Cervantes Constantino et al., 2017
Hamilton et al., 2018
Daube et al., 2019
Bregman et al., 1994
Spectro-Temporal Response Functions

Minimum norm current estimate

Predicted response

Spectrogram

Amplitude in frequency bins
Spectro-Temporal Response Functions

2 STRFs
Envelope STRF + Onset STRF (8 bands each)
Outline

• Cocktail party listening
  ‣ Speech segregation & cortical processing of ignored speech
  ‣ MEG representations of speech

• Methods

• Results

• Summary
Outline

• Cocktail party listening
  ‣ Speech segregation & cortical processing of ignored speech
  ‣ MEG representations of speech

• Methods

• Results

• Summary
Methods

Data set already used and described in Brodbeck, 2019

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

16 one-minute-long segments constructed from the same passages with two competing speakers, male + female, equal loudness

- Subjects’ instructions: Attend to one, ignore the other (counter-balanced)
- After each segment, answer question about content of the attended stimulus

Distributed MNE source estimates, restricted to Region of Interest (below)

- 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
Outline

• Cocktail party listening
  ‣ Speech segregation & cortical processing of ignored speech
  ‣ MEG representations of speech

• Methods

• Results

• Summary
Outline

• Cocktail party listening
  ▸ Speech segregation & cortical processing of ignored speech
  ▸ MEG representations of speech

• Methods

• Results

• Summary
Successful Response Prediction

Source Localization consistent with auditory cortex (Heschl’s gyrus, superior temporal Gyrus)
Successful Response Prediction

Significant prediction

Onsets

Envelope

Source Localization consistent with auditory cortex
(Heschl’s gyrus, superior temporal Gyrus)
Onset & Envelope STRFs

Onset STRF
• Typical response pattern
  + early peak
  – later peak

Envelope STRF
• Less well-defined

Onset STRF

- Typical response pattern
  + early peak
  – later peak

Envelope STRF
• Less well-defined
Cocktail Party Listening

<table>
<thead>
<tr>
<th>Acoustic mixture</th>
<th>Attended source</th>
<th>Ignored source</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Acoustic mixture" /></td>
<td><img src="image2.png" alt="Attended source" /></td>
<td><img src="image3.png" alt="Ignored source" /></td>
</tr>
</tbody>
</table>

Potential representations

- Acoustic mixture (input)
- Recovered source signals
  - Attended source
  - Ignored source (?)

Significant responses

- Significant response to onsets in the ignored source
- After accounting for mixture and attended source

Brodbeck et al., bioRxiv 2019

Attended vs. mixture onsets
Ignored vs. mixture onsets

<table>
<thead>
<tr>
<th>Time (ms)</th>
<th>Onset response comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>Source current (normalised) in mixture</td>
</tr>
<tr>
<td>40+4</td>
<td>Mixture onset (ms)</td>
</tr>
<tr>
<td>40*2</td>
<td>Source current variance in mixture</td>
</tr>
<tr>
<td>40+4</td>
<td>Attended onsets</td>
</tr>
<tr>
<td>40*2</td>
<td>Ignored onsets</td>
</tr>
</tbody>
</table>

Attended - Ignored
Ignored - Mixture

<table>
<thead>
<tr>
<th>Time (ms)</th>
<th>Onset response comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>Source current (normalised) in mixture</td>
</tr>
<tr>
<td>40+4</td>
<td>Mixture onset (ms)</td>
</tr>
<tr>
<td>40*2</td>
<td>Source current variance in mixture</td>
</tr>
<tr>
<td>40+4</td>
<td>Attended onsets</td>
</tr>
<tr>
<td>40*2</td>
<td>Ignored onsets</td>
</tr>
</tbody>
</table>

Attended - Ignored
Ignored - Mixture
Cocktail Party Listening

Potential representations
- Acoustic mixture (input)
- Recovered source signals
  - Attended source
  - Ignored source (?)

6 STRFs(!)
Envelope STRF + Onset STRF for each of
Mixture, Attended, Ignored
(8 bands each)
Cocktail Party STRFs

Onset STRFs

Mixture onset STRF

Source onset STRFs

Attended source

Ignored source

Acoustic mixture

Attended source

Ignored source

Source current (normalized)

Center frequency

.p

.p

.p

.p

.p

.p

.p

.p

.p

.p
Onset STRFs

Source onset STRFs

Mixture onset STRF

Source current (normalized)

Attended source

Ignored source

Attended - Ignored

Acoustic mixture
Cocktail Party STRFs

Onset STRFs

Mixture onset STRF

Source onset STRFs

Attended source

Ignored source

Envelope STRFs

Mixture onset STRF

Source onset STRFs

Attended source

Ignored source

Envelope STRFs

Center frequency
Cocktail Party Onset STRFs

STRF for mixture
- + large peak (72 ms)
- – smaller peak (126 ms)

STRF for attended
- + peak (81 ms)
- – peak (150 ms)

STRF for ignored
- + peak (88 ms)
- – peak absent
Cocktail Party Onset STRFs

STRF for mixture
• + large peak (72 ms)
• – smaller peak (126 ms)

STRF for attended
• + peak (81 ms)
• – peak (150 ms)

STRF for ignored
• + peak (88 ms)
• – peak absent

Onset STRFs

Early response to mixture greater & earlier than sources
Early response to sources not distinguishable
Later response only to attended source, not ignored
Not all onsets are the same

• Source onsets can be masked by other source
  - “Masked onset”
  - Typically occurs when other (masking) source sustained
  - ➔ No onset apparent in mixture despite source onset

• Source onsets may not be masked by other source
  - “Overt onset”
  - Onset apparent in both mixture and source
  - Other source does not interfere

• Overt onsets allow segregation via filtering

• Covert onsets more difficult to unmix
Not all onsets are the same

- Onsets in one source can be masked by other
  - Typically occurs when masking source sustained
  - No onset apparent in mixture despite source onset
Not all onsets are the same

- Onsets in one source can be masked by other
  - Typically occurs when masking source sustained
  - ➔ No onset apparent in mixture despite source onset

Overt onset
Not all onsets are the same

- Onsets in one source can be masked by other
  - Typically occurs when masking source sustained
  - No onset apparent in mixture despite source onset
Not all onsets are the same

• Onsets in one source can be masked by other
  - Typically occurs when masking source sustained
  - \(\rightarrow\) No onset apparent in mixture despite source onset
Masked onsets and active vs. passive segregation

- Can we distinguish active segregation mechanisms from passive?
  - Masked onsets cannot be processed passively
Masked onsets and active vs. passive segregation

• Can we distinguish active segregation mechanisms from passive?
  - Masked onsets cannot be processed passively

Masked & Overt onset streams generate separate STRFs
Auditory Cortex and Overt vs. Masked Onsets

**Overt Onset Responses**
- Similar to onset responses above
- Early response to sources not distinguishable
- Later response only to attended source, not ignored
Auditory Cortex and Overt vs. Masked Onsets

**Overt Onset Responses**
- Similar to onset responses above
- Early response to sources not distinguishable
- Later response only to attended source, not ignored

![Graph showing Overt Onset Responses](image)
Auditory Cortex and Overt vs. Masked Onsets

**Overt Onset Responses**
- Similar to onset responses above
- Early response to sources not distinguishable
- Later response only to attended source, not ignored

**Masked Onset Responses**
- Smaller peaks (≠0)
- Early peak shows effect of attention
- …
Masked onsets engage extended cortical processing

- Masked onset peaks delayed relative to overt onset peaks
  - early masked peaks delayed ~20 ms
  - later (attended) masked peak delayed ~45 ms
- More time spent processing masked peaks
- Evidence for early active processing in segregation
Outline

• Cocktail party listening
  ‣ Speech segregation & cortical processing of ignored speech
  ‣ MEG representations of speech

• Methods

• Results

• Summary
Outline

• Cocktail party listening
  ‣ Speech segregation & cortical processing of ignored speech
  ‣ MEG representations of speech

• Methods

• Results

• Summary
Summary I

• Spectrotemporal acoustic onsets robustly represented in auditory cortex
  - Onsets explain more variance when onsets and envelopes are allowed to compete

• Onsets in both attended and ignored speech represented in auditory cortex, in addition to onsets in acoustic mixture

• Early onset processing does not distinguish between attended and ignored speech
  - except masked onsets (attended > ignored): early effects of selective attention
Summary II

- Auditory cortex “un-masks” masked onsets
  - Related to neural filling-in?
- Unmasking requires additional processing time
  - SNR-dependent delays well known, but here shown to be dynamic
- Scene segregation not merely passive spectrotemporal filtering
  - Scene segregation employs active processing
Summary II

- Auditory cortex “un-masks” masked onsets
  - Related to neural filling-in?
- Unmasking requires additional processing time
  - SNR-dependent delays well known, but here shown to be dynamic
- Scene segregation not merely passive spectrotemporal filtering
  - Scene segregation employs active processing
Thank You
Acknowledgements

Current Lab Members & Affiliates

Christian Brodbeck
Alex Presacco
Proloy Das
Lien Decruy
Jason Dunlap
Theo Dutcher
Kevin Hu
Dushyanthi Karunathilake
Joshua Kulasingham
David Nahmias
Peng Zan

Julian Jenkins
Pirazh Khorramshahi
Natalia Lapinskaya
Huan Luo
Sina Miran
Mahshid Najafi
Krishna Puvvada
Jonas Vanthornhout
Ben Walsh
Yadong Wang
Juanjuan Xiang
Jiachen Zhuo

Didier Depireux
Mounya Elhilali
Tom Francart
Jonathan Fritz
Michael Fu
L. Elliot Hong
Stefanie Kuchinsky
Steven Marcus
Cindy Moss
David Poeppel
Shihab Shamma

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

Past Undergraduate Students

Nicholas Asendorf
Ross Baehr
Anurupa Bhonsale
Elizabeth Camenga
Katya Dombrowski
Kevin Hogan
Alex Jiao
James Williams

Funding

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