Neural Representations of Speech at the “Cocktail Party” in Human Auditory Cortex

Jonathan Z. Simon

Department of Electrical & Computer Engineering
Department of Biology
Institute for Systems Research
University of Maryland
Acknowledgements

Current (Simon Lab & Affiliates)
Christian Brodbeck
Francisco Cervantes
David Nahmias
Mahshid Najafi
Krishna Puvvada
Peng Zan

Past (Simon Lab & Affiliate Labs)
Nayef Ahmar
Murat Aytekin
Claudia Bonin
Maria Chait
Marisel Villafane Delgado
Kim Drnec
Nai Ding
Victor Grau-Serrat
Julian Jenkins
Natalia Lapinskaya
Kai Sum Li
Huan Luo
Ling Ma
Alex Presacco

Raul Rodriguez
Ben Walsh
Juanjuan Xiang
Jiachen Zhuo

Collaborators
Pamela Abshire
Samira Anderson
Behtash Babadi
Catherine Carr
Monita Chatterjee
Alain de Cheveigné
Didier Depireux
Mounya Elhilali
Bernhard Englitz
Jonathan Fritz
Stefanie Kuchinsky
Cindy Moss
David Poeppel
Shihab Shamma

Past Postdocs & Visitors
Sahar Akram

Aline Gesualdi Manhães
Dan Hertz
Jonas Vanthornhout
Yadong Wang

Undergraduate Students
Abdulaziz Al-Turki
Nicholas Asendorf
Sonja Bohr
Elizabeth Camenga
Corinne Cameron
Julien Dagenais
Katya Dombrowski
Kevin Hogan
Kevin Kahn
Alexandria Miller
Andrea Shome
Sandra Soltz
Madeleine Varmer
James Williams

Funding NIH (NIDCD, NIA, NIBIB); USDA; UMD
Outline

• Cortical Representations of Speech (via MEG)
  ‣ Encoding vs. Decoding

• Cortical Representations of the “Cocktail Party”

• Recent Results
  ‣ Attentional Dynamics
  ‣ Aging & Cortical Representations of Speech
  ‣ Higher Level Interference & Noise
Functional Brain Imaging

= Non-invasive recording from human brain

**Hemodynamic techniques**

- **fMRI**
  - functional magnetic resonance imaging
  - Excellent Spatial Resolution (~1 mm)
  - Poor Temporal Resolution (~1 s)

- **PET**
  - positron emission tomography
  - Poor Spatial Resolution (~1 cm)
  - Excellent Temporal Resolution (~1 ms)

**Electromagnetic techniques**

- **EEG**
  - electroencephalography
  - Poor Spatial Resolution (~1 cm)

- **MEG**
  - magnetoencephalography
  - Excellent Temporal Resolution (~1 ms)

fMRI & MEG can capture effects in single subjects
MEG & Auditory Cortex

- Non-invasive, Passive, Silent Neural Recordings
- MEG Response Patterns Time-Locked to Stimulus Events
- Robust
- Strongly Lateralized
- Cortical Origin Only
MEG Responses to Speech Modulations
MEG Responses Predicted by STRF Model

Linear Kernel = STRF

“Spectro-Temporal Response Function”

Long duration speech: ~60 s

(up to ~10 Hz)

Ding & Simon, J Neurophysiol (2012)
MEG Responses Predicted by STRF Model

Linear Kernel = STRF

“Spectro-Temporal Response Function”

Long duration speech: ~60 s

(up to ~10 Hz)

Ding & Simon, J Neurophysiol (2012)
Neural Reconstruction of Speech Envelope

SpeechEnvelope

Decoder

MEGResponses

(up to ~ 10 Hz)
Neural Reconstruction of Speech Envelope

Decoder

Speech Envelope

MEG Responses

(Up to ~10 Hz)

stimulus speech envelope

reconstructed stimulus speech envelope

2 s

Reconstruction accuracy comparable to single unit & ECoG recordings

Ding & Simon, J Neurophysiol (2012)
Zion-Golumbic et al., Neuron (2013)
Neural Reconstruction of Speech Envelope

Reconstruction accuracy comparable to single unit & ECoG recordings (up to ~10 Hz)

Decoder

Speech Envelope

MEG Responses

(up to ~ 10 Hz)
Neural Representation of Speech: Temporal
Speech in Stationary Noise

Mixtures of Speech and Spectrally Matched Stationary Noise

- quiet background
- 6 dB
- -3 dB
- -9 dB

Contrast Index

Ding & Simon, J Neuroscience (2013)
Speech in Stationary Noise

Mixtures of Speech and Spectrally Matched Stationary Noise

quiet background  6 dB  -3 dB  -9 dB

Contrast Index

Intelligibility (%)
Speech in Noise: Results

Neural Reconstruction of Underlying Speech Envelope

+6 dB

-6 dB

1 s

Ding & Simon, J Neuroscience (2013)
Speech in Noise: Results

Neural Reconstruction of Underlying Speech Envelope

Reconstruction Accuracy

Contrast Index

Intelligibility (%)
Speech in Noise: Results

Neural Reconstruction of Underlying Speech Envelope

Reconstruction Accuracy

Correlation with Intelligibility

Ding & Simon, J Neuroscience (2013)
Noise-Vocoded Speech

Intelligibility Reflected only in Delta Band (1–4 Hz)

Ding, Chatterjee & Simon, NeuroImage (2014)
Multiple Cortical Speech Representations?

Di Liberto, et al. (2015) Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing

Kayser et al. (2015) Irregular Speech Rate Dissociates Auditory Cortical Entrainment, Evoked Responses, and Frontal Alpha

Ding et al. (2015) Cortical tracking of hierarchical linguistic structures in connected speech
Cortical Speech Representations

• Neural Representations: Encoding & Decoding
• Linear models: Useful & Robust
• Speech **Envelope** only (as seen in MEG)
• Envelope Rates: ~ 1 - 10 Hz
• Intelligibility linked to lower range of frequencies (Delta)
Listening to Speech at the Cocktail Party

Alex Katz,
The Cocktail Party
Listening to Speech at the Cocktail Party
Listening to Speech at the Cocktail Party
Listening to Speech at the Cocktail Party
Competing Speech Streams
Selective Neural Encoding
Selective Neural Encoding
Unselective vs. Selective Neural Encoding
Selective Neural Encoding
Selective Encoding: Results

Identical Stimuli!

Ding & Simon, PNAS (2012)
Single Trial Speech Reconstruction

Ding & Simon, PNAS (2012)
Single Trial Speech Reconstruction

Attended Speech Reconstruction

Background Speech Reconstruction

attentional focus
speaker one
speaker two

Ding & Simon, PNAS (2012)
STRF Results

- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- M50_{STRF} positive peak
- M100_{STRF} negative peak
- M100_{STRF} strongly modulated by attention, but not M50_{STRF}
Neural Sources

- M100_{STRF} source near (same as?) M100 source: Planum Temporale
- M50_{STRF} source is anterior and medial to M100 (same as M50?): Heschl’s Gyrus
- PT strongly modulated by attention, *but not HG*
Recent Results

- Attentional Dynamics
- Aging & Cortical Representations of Speech
- High Level Interference & Noise
Recent Results

• Attentional Dynamics
• Aging & Cortical Representations of Speech
• High Level Interference & Noise
Attentional Dynamics

Akram et al., NeuroImage (2016)
Attentional Dynamics

Probability of attending Speaker 1

Attend to Speaker 1

Attend to Speaker 2

Switch Attention

Akram et al., NeuroImage (2016)
Recent Results

- Attentional Dynamics
- Aging & Cortical Representations of Speech
- High Level Interference & Noise
Average Responses to Pure Tone
Aging & Auditory Cortex

Average Responses to Pure Tone

Over-Representation

Evoked response to the 500 Hz tone burst

Younger

Older

X: 0.7861
Y: 0.1801

Aging & Auditory Cortex

Average Responses to Pure Tone
Aging & Auditory Cortex

Average Responses to Pure Tone

M100 Power by Subject

Evoked response to the 500 Hz tone burst

Younger vs. Older

Over-Representation

Subject

z-score

Over-Representation
Speech Over-Representation

Presacco et al., J Neurophysiol (2016a)
Speech Over-Representation

Speech Reconstruction by SNR

Presacco et al., J Neurophysiol (2016a)
Aging & Integration Time

Younger Adults

In Quiet

Reconstruction Accuracy

Integration window (ms)

Older Adults

In Quiet

with Competing Speaker

Presacco et al., J Neurophysiol (2016a)
Neural vs Inhibitory Control

Presacco et al., J Neurophysiol (2016b)
Recent Results

- Attentional Dynamics
- Aging & Cortical Representations of Speech
- High Level Interference & Noise
High Level Interference Effects

Reconstruction Accuracy

In Quiet

Speech Reconstruction by SNR

Presacco et al., J Neurophysiol (2016b)
High Level Interference Effects

- Unfamiliarity of Background - Boosts Intelligibility of Attended Speech

Presacco et al., J Neurophysiol (2016b)
High Level Interference Effects

• Unfamiliarity of Background
  - Boosts Intelligibility of Attended Speech

Presacco et al., J Neurophysiol (2016b)
High Level Interference Effects

- Unfamiliarity of Background
  - Boosts Intelligibility of Attended Speech

Presacco et al., J Neurophysiol (2016b)
High Level Interference Effects

- Unfamiliarity of Background
  - Boosts Intelligibility of Attended Speech
  - Also Boosts Cortical Reconstruction of Attended Speech

Presacco et al., J Neurophysiol (2016b)
Summary

• Cortical representations of speech
  - representation of envelope (up to ~10 Hz)
  - robust against a variety of noise types
  - neural representation of perceptual object

• Object-based representation at 100 ms latency (PT), but not by 50 ms (HG)

• Robust Dynamical Foreground Monitoring

• Over-Representation with Aging
  - Reconstruction depends on integration time
  - Over-Representation tracks inhibitory control

• Background familiarity: neural tracks behavior
Thank You