Cortical Encoding of Auditory Objects at the Cocktail Party

Jonathan Z. Simon
University of Maryland

ARO Presidential Symposium, February 2013

Introduction

- Auditory Objects
- Magnetoencephalography (MEG)
- Neural Representations of Auditory Objects in Cortex: Decoding
- Neural Representations of Auditory Objects in Cortex: Encoding

Auditory Objects

- What is an auditory object?
  - perceptual (not neural) construct
  - commonalities with visual objects
  - several potential formal definitions

Auditory Object Definition

- Griffiths & Warren definition:
  - corresponds with *something* in the sensory world
  - object information *separate from* information of rest of sensory world
  - abstracted: object information *generalized over particular* sensory experiences

Auditory Objects

- a few more details…
  - punctate or streaming, serial or parallel
  - not quite "*what*" in "*what vs. where*"

Auditory Objects at the Cocktail Party

Alex Katz,
The Cocktail Party

Auditory Objects at the Cocktail Party

Alex Katz,
The Cocktail Party

Auditory Objects at the Cocktail Party

Alex Katz,
The Cocktail Party

Auditory Objects at the Cocktail Party

Alex Katz,
The Cocktail Party

Auditory Objects at the Cocktail Party

Alex Katz,
The Cocktail Party

Auditory Objects at the Cocktail Party

Alex Katz,
The Cocktail Party

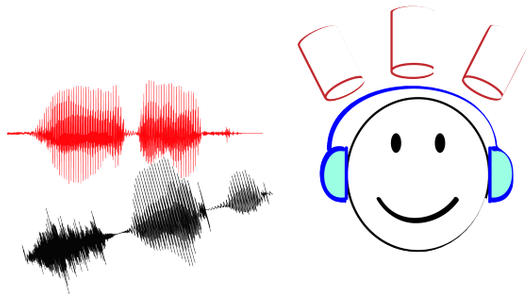
Auditory Objects at the Cocktail Party

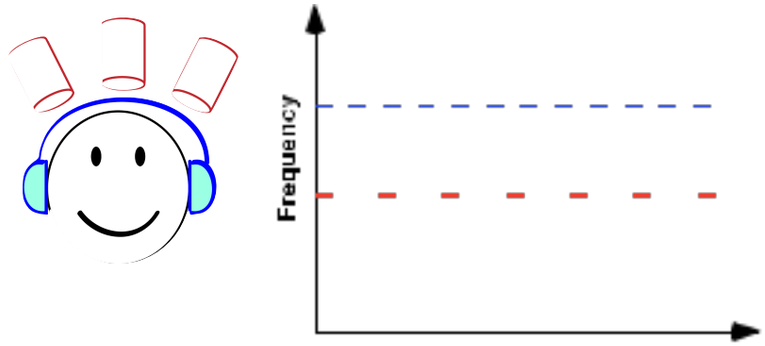
Alex Katz,
The Cocktail Party

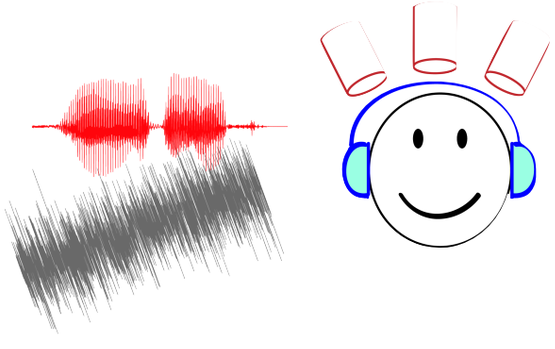
Auditory Objects at the Cocktail Party

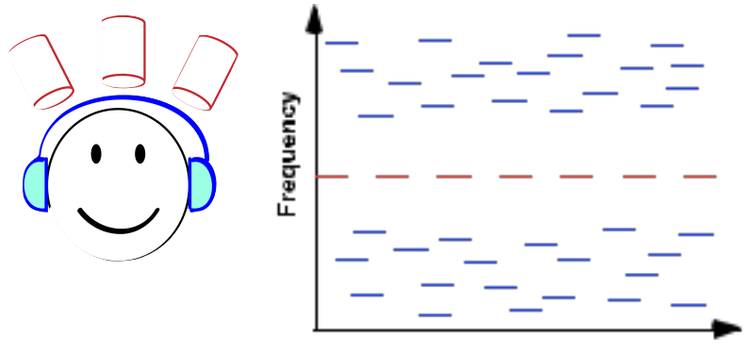
Ding & Simon, PNAS (2012)

Experiments









Magnetoencephalography (MEG)

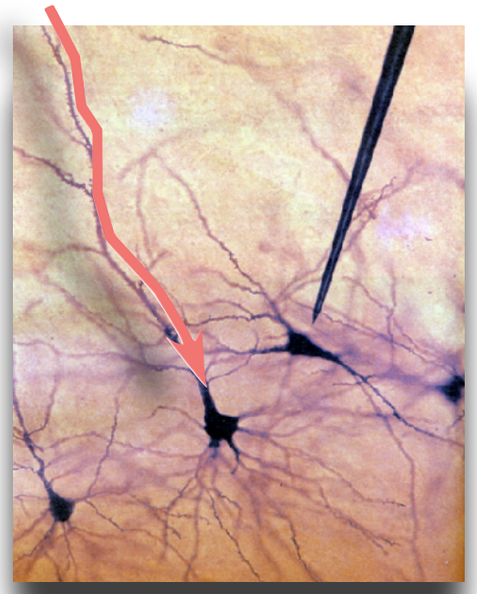
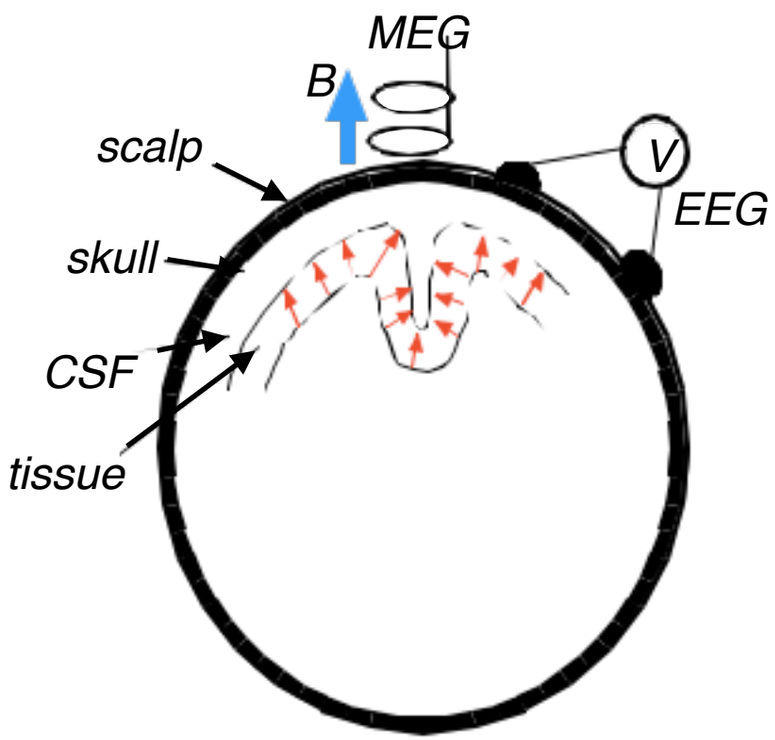
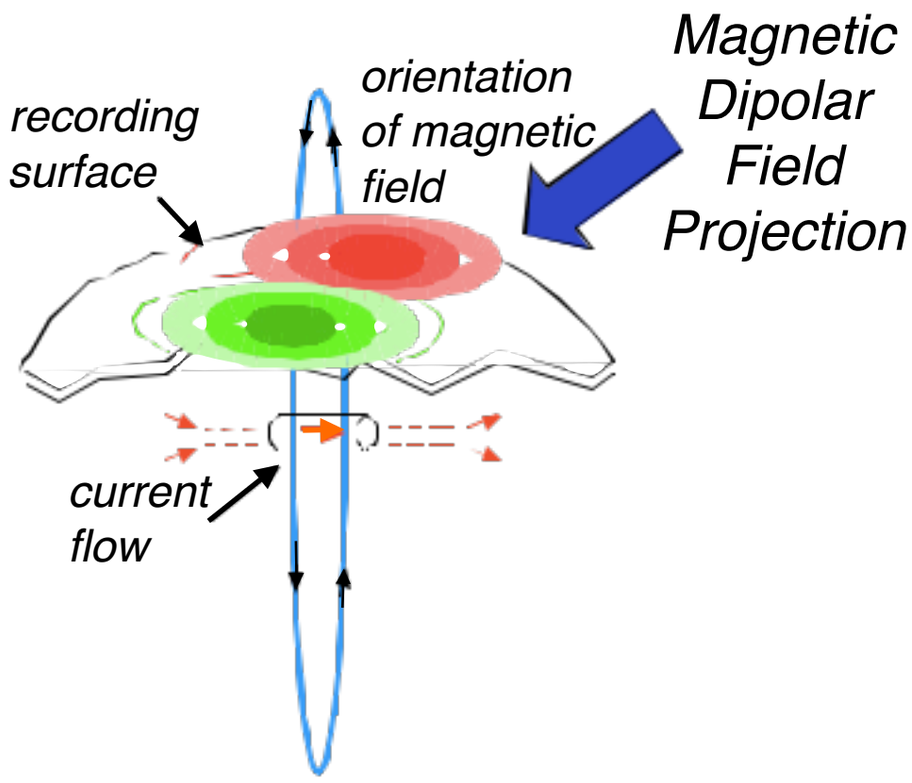


Photo by Fritz Goro





- 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)

Magnetoencephalography (MEG)

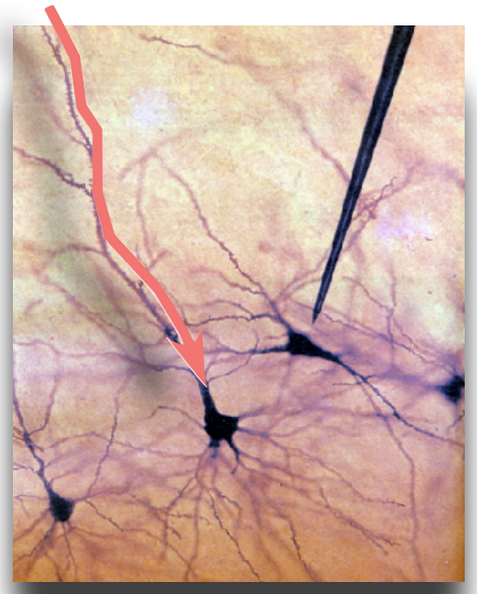
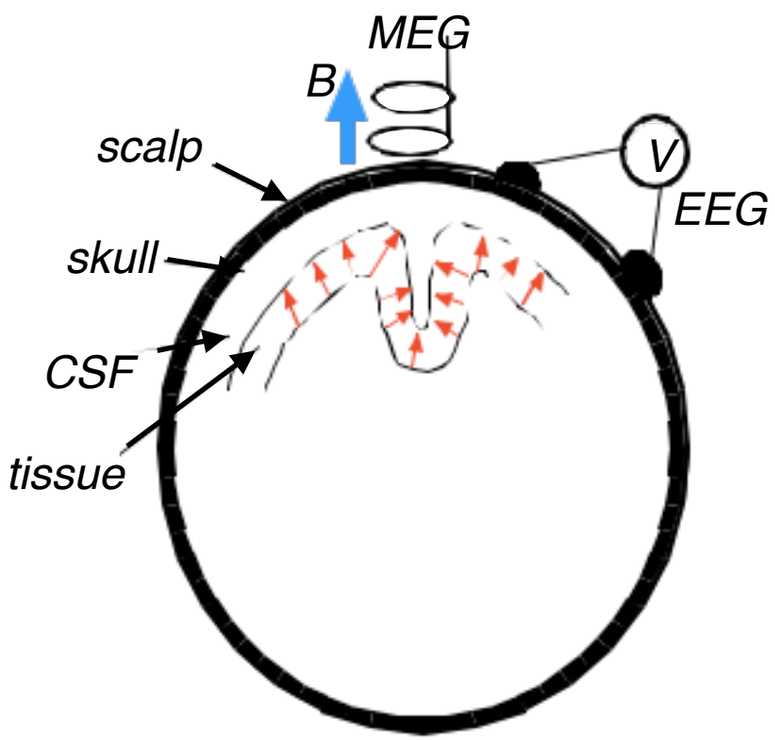
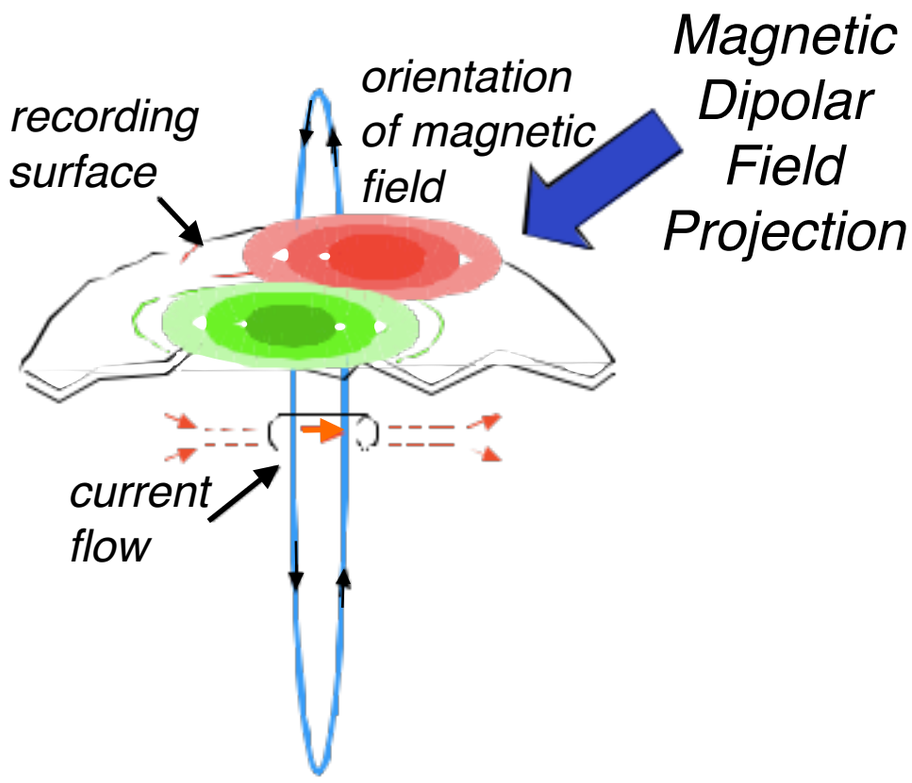


Photo by Fritz Goro

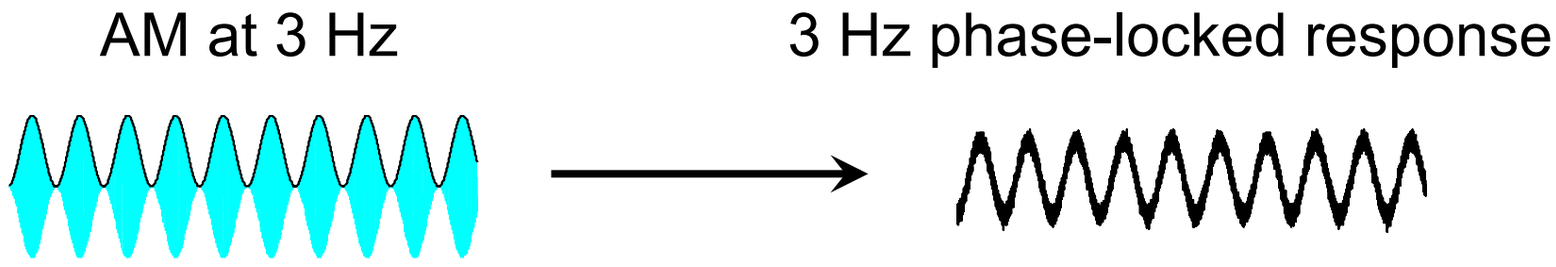




- Direct electrophysiological measurement
  - not hemodynamic
  - real-time
- No unique solution for distributed source

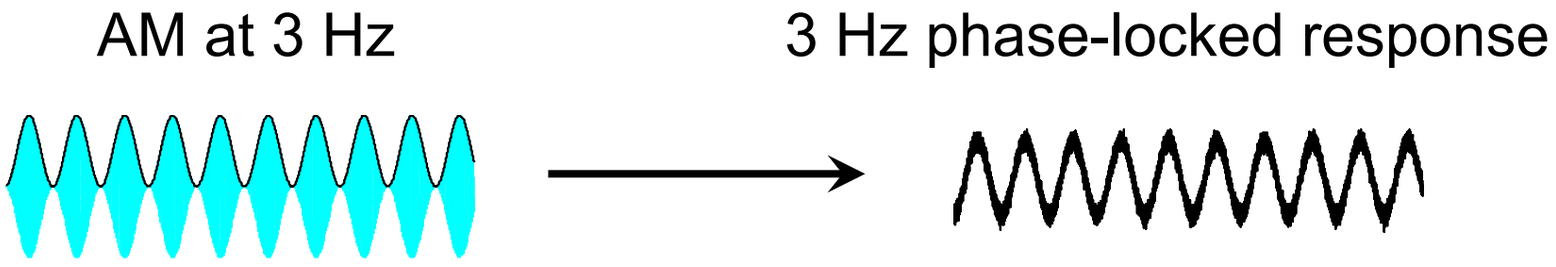
- Measures spatially synchronized cortical activity
- Fine temporal resolution (~ 1 ms)
- Moderate spatial resolution (~ 1 cm)

Phase-Locking in MEG to Slow Temporal Modulations

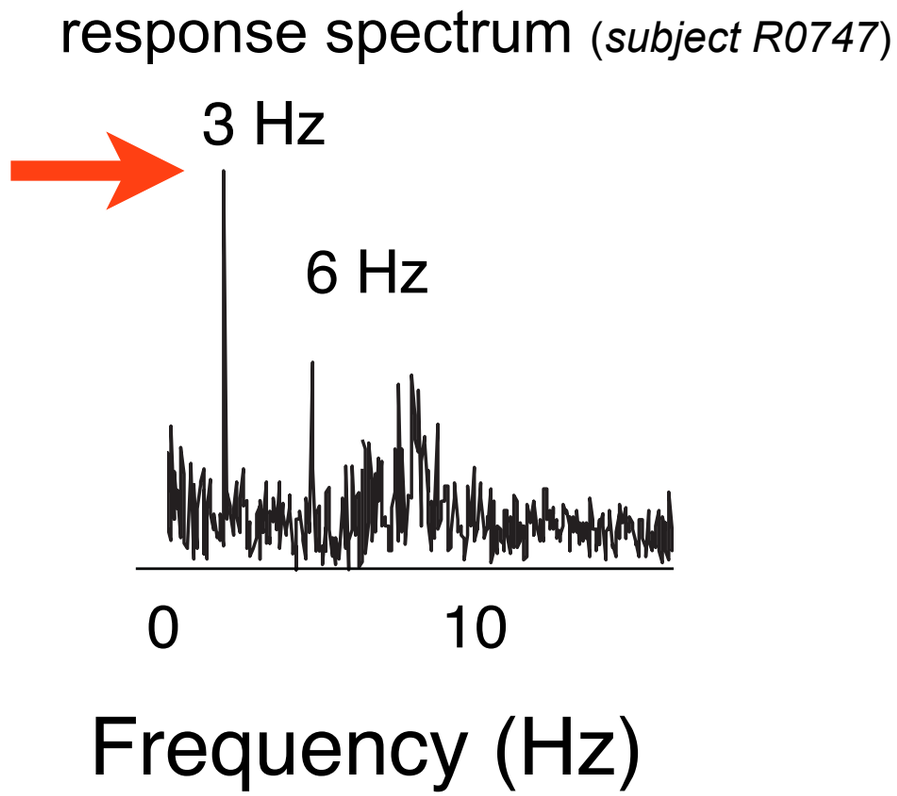


Ding & Simon, J Neurophysiol (2009)
Wang et al., J Neurophysiol (2012)

Phase-Locking in MEG to Slow Temporal Modulations

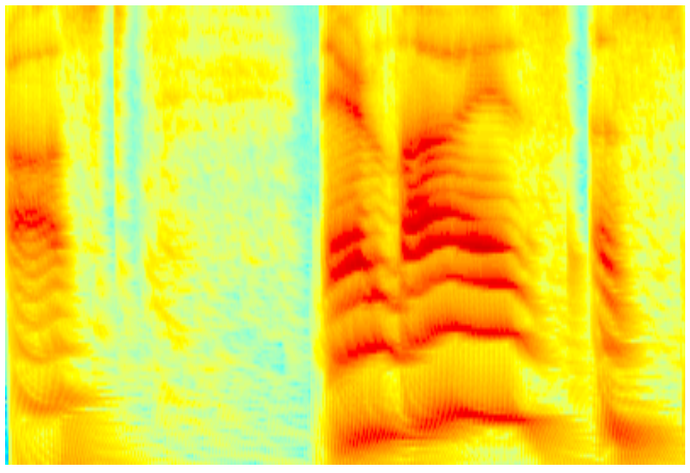


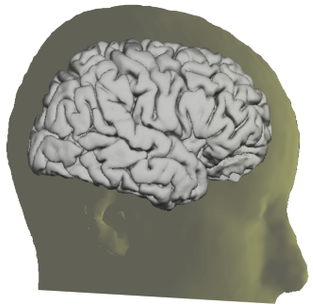
MEG activity is precisely phase-locked to temporal modulations of sound

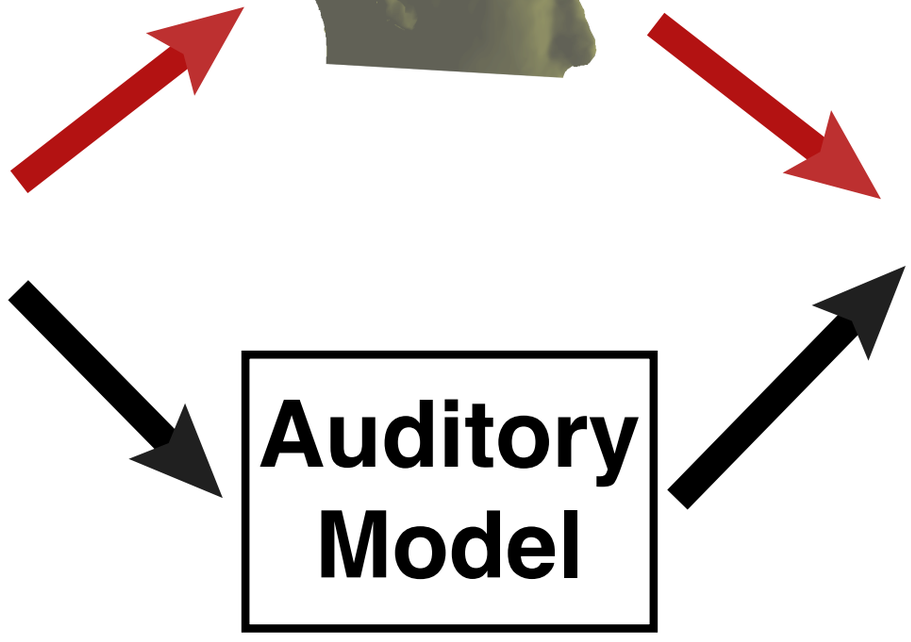


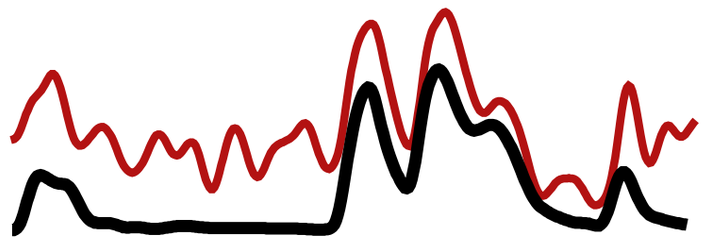
Ding & Simon, J Neurophysiol (2009)
Wang et al., J Neurophysiol (2012)

MEG Responses to Speech Modulations

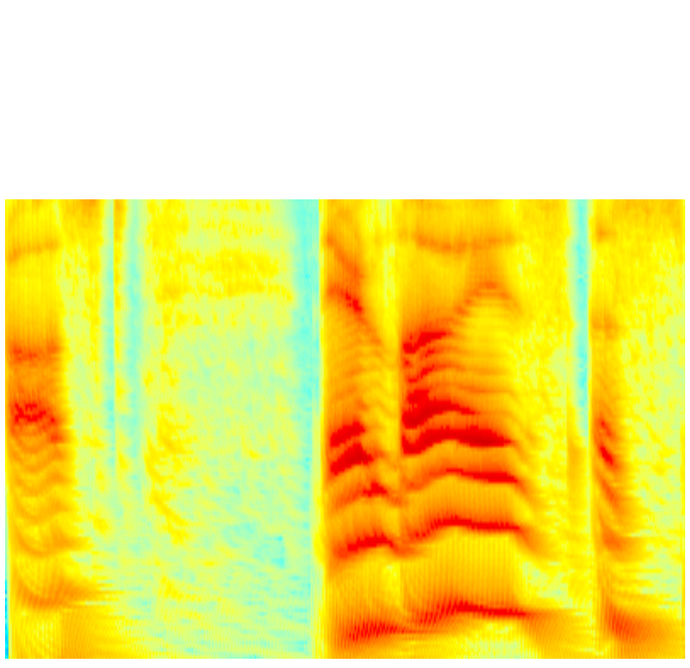


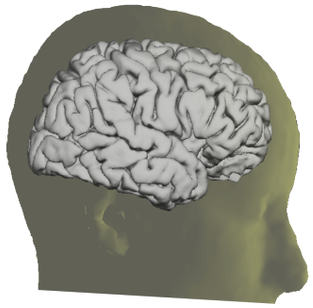


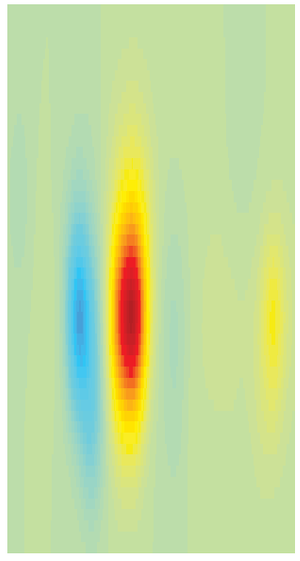




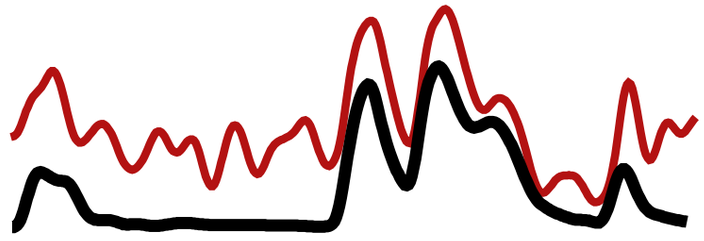
MEG Responses Predicted by STRF Model







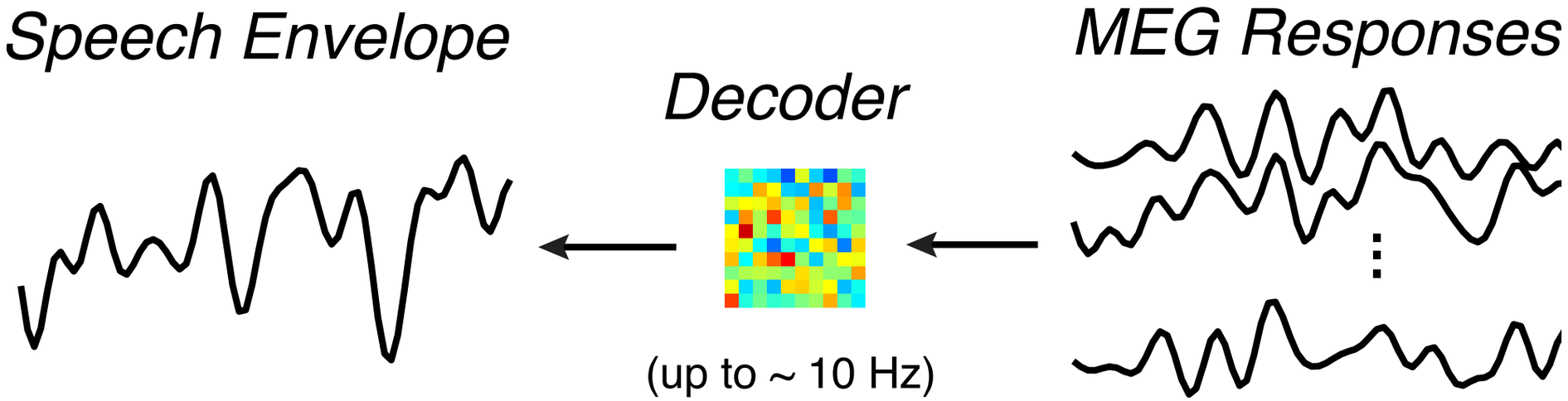
(up to ~10 Hz)



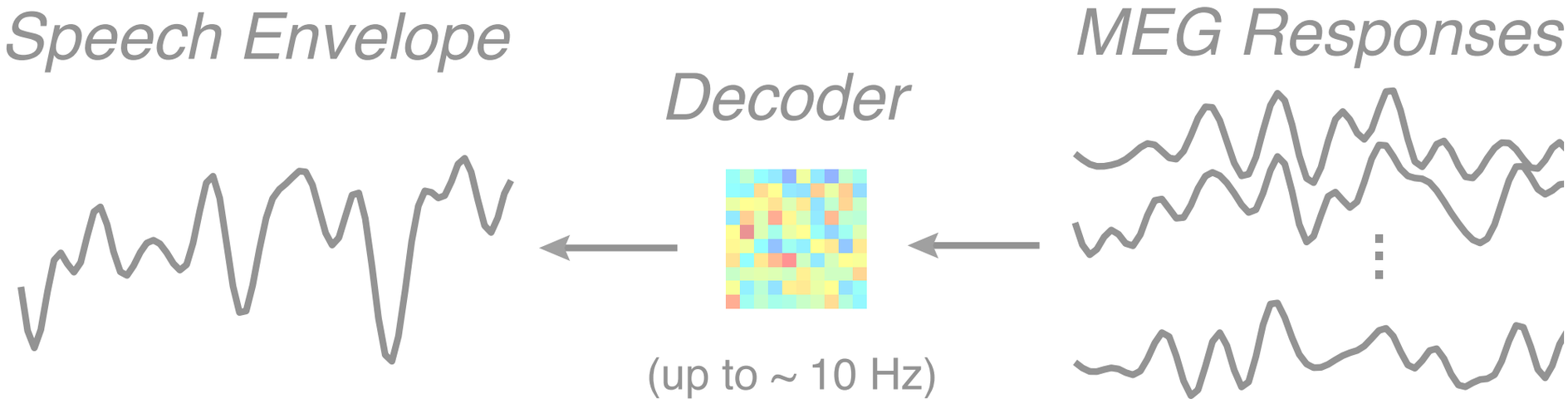
Spectro-Temporal Response Function (STRF)

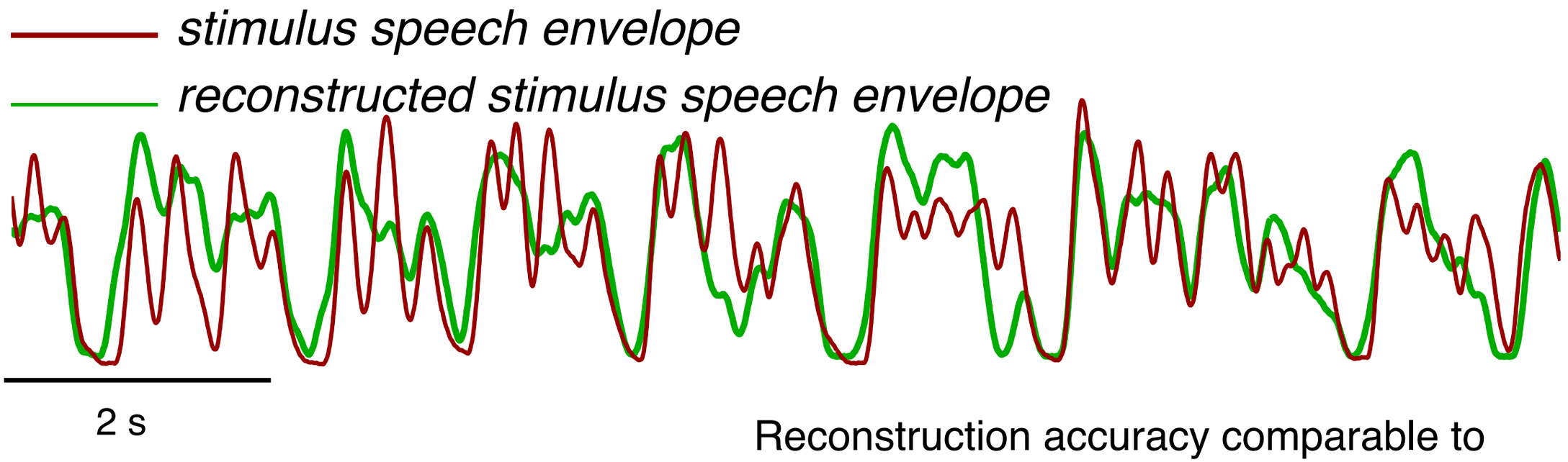
Ding & Simon, J Neurophysiol (2012)

Neural Reconstruction of Speech Envelope



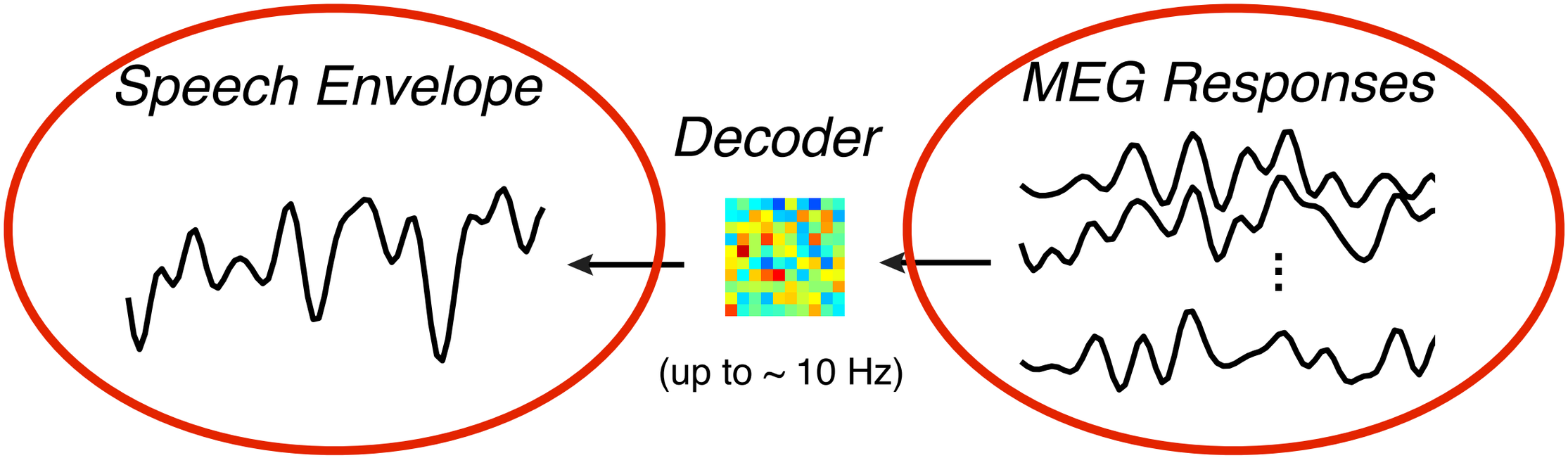
Neural Reconstruction of Speech Envelope





Reconstruction accuracy comparable to single unit & ECoG recordings

Ding & Simon, J Neurophysiol (2012)
Zion-Golumbic et al., Neuron (in press)


Speech Envelope
Decoder
MEG Responses
(up to ~ 10 Hz)

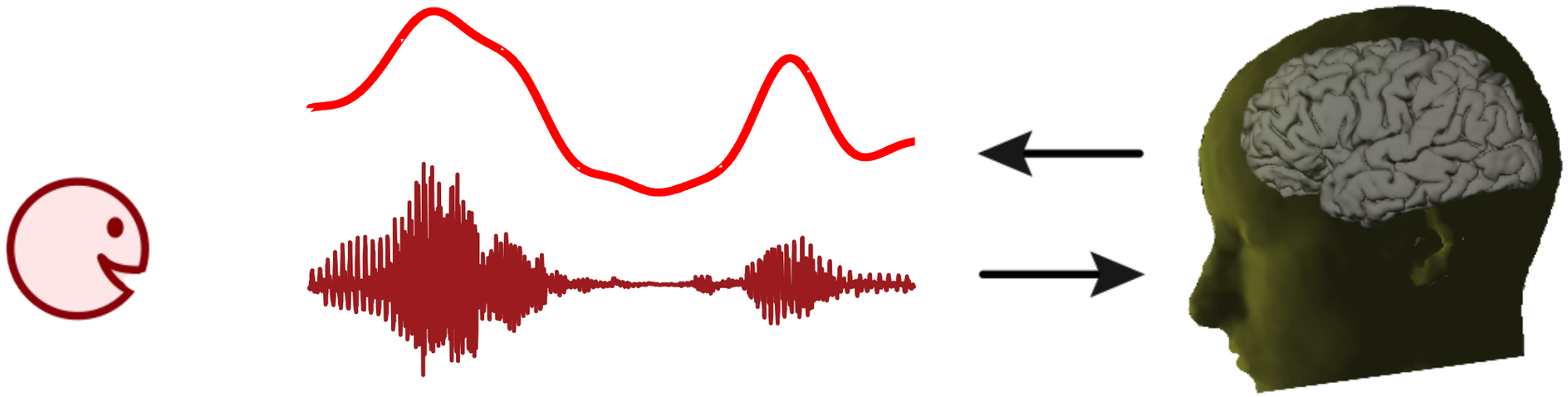
Speech Stream as an Auditory Object

- corresponds with something in the sensory world
- information *separate from* information of rest of sensory world
  e.g. other speech streams or noise
- abstracted: object information *generalized over particular* sensory experiences
  e.g. different sound mixtures

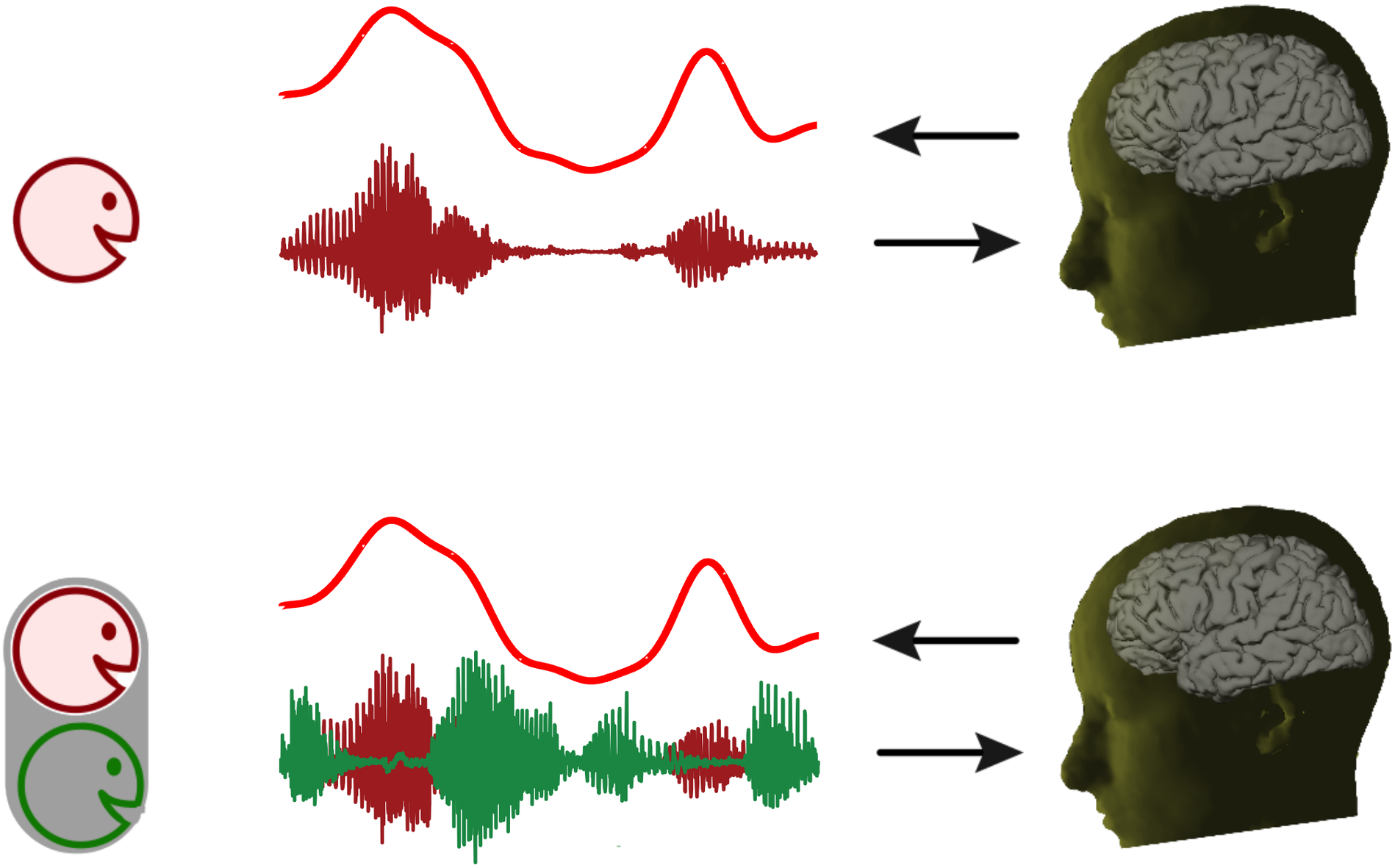
Neural Representation of an Auditory Object

- neural representation is of something in sensory world
- when other sounds mixed in,
  neural representation is of auditory object,
  not entire acoustic scene
- neural representation invariant
  under broad changes in specific acoustics

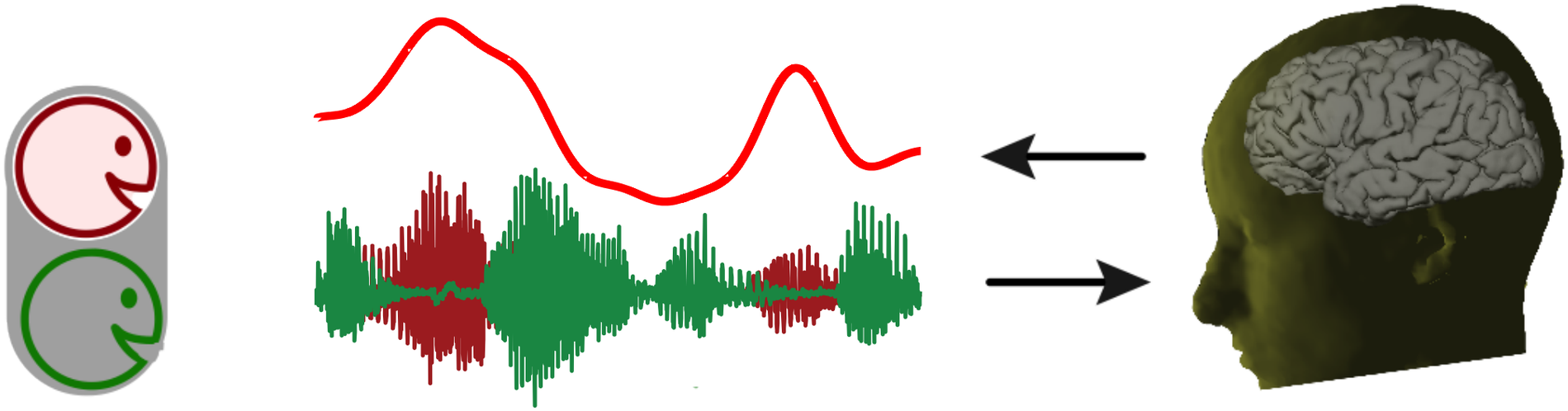
Selective Neural Encoding



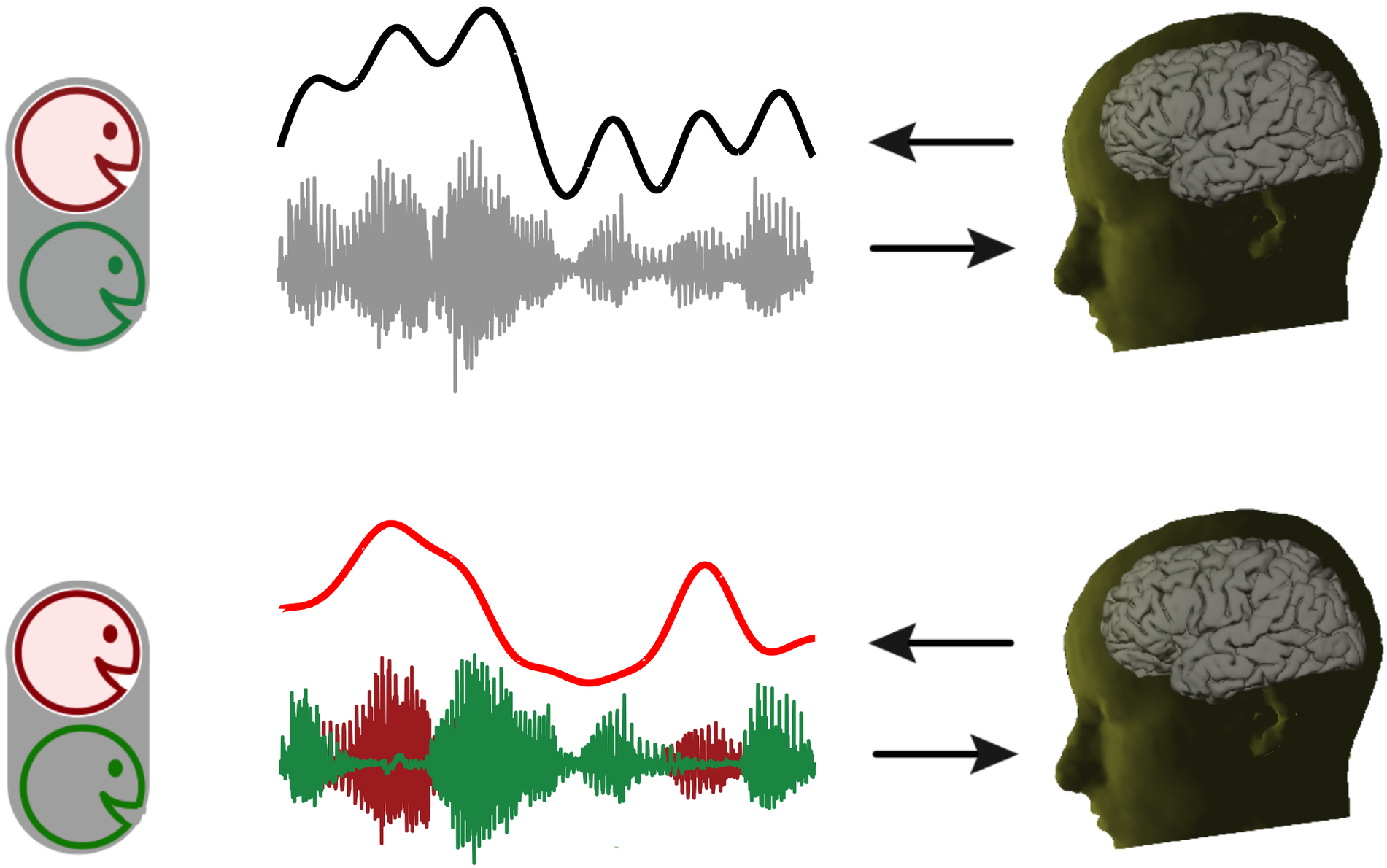
Selective Neural Encoding



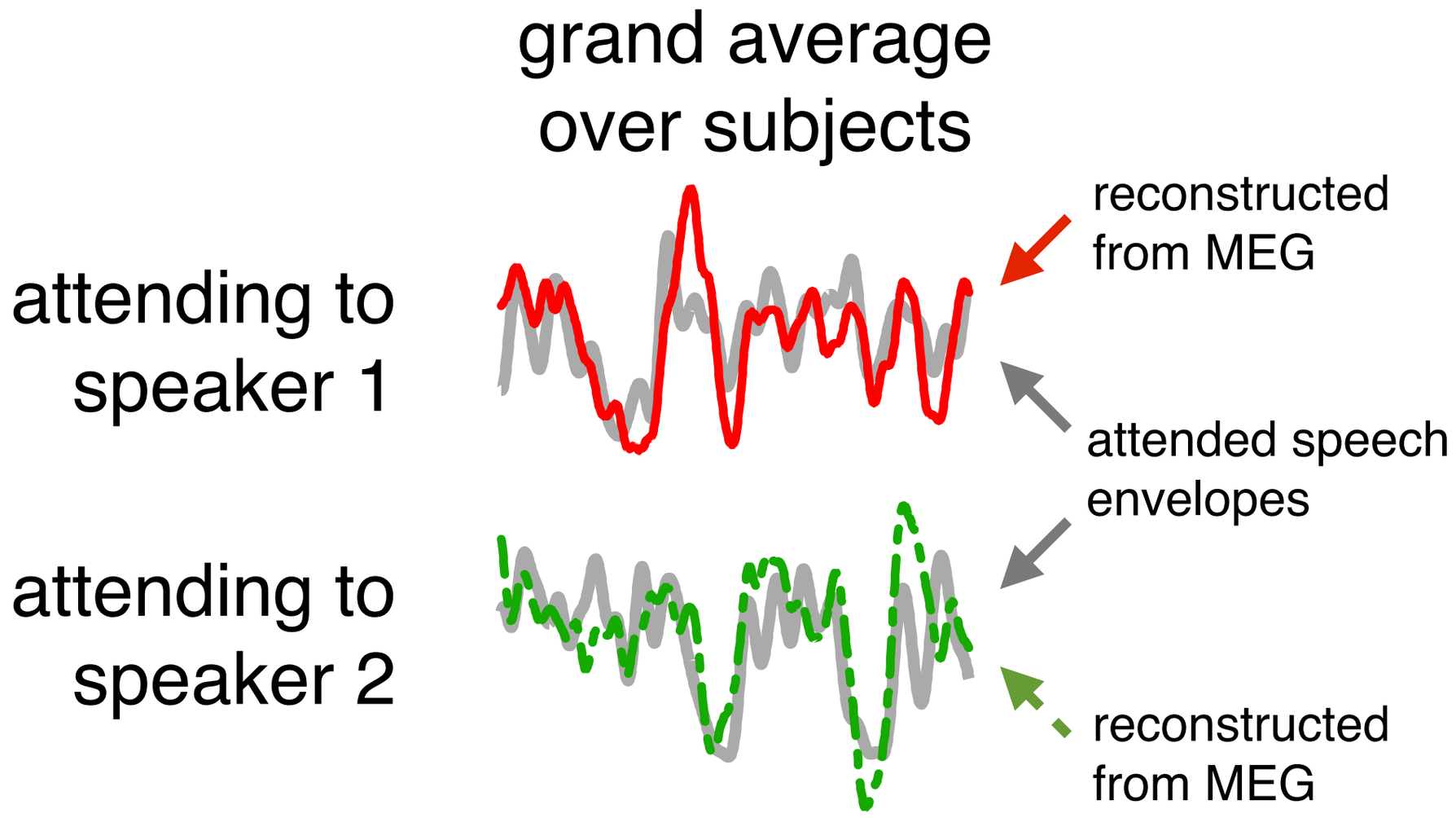
Selective Neural Encoding



Unselective vs. Selective Neural Encoding

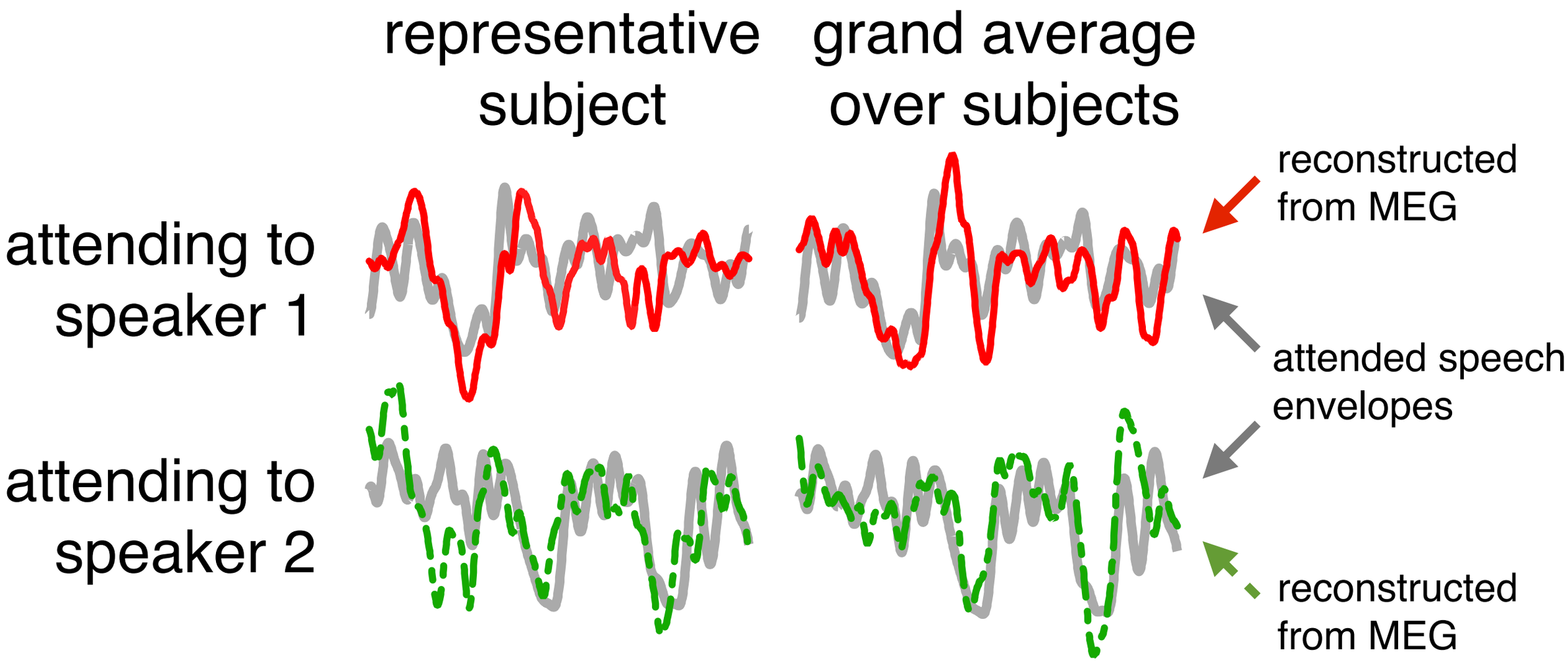


Stream-Specific Representation



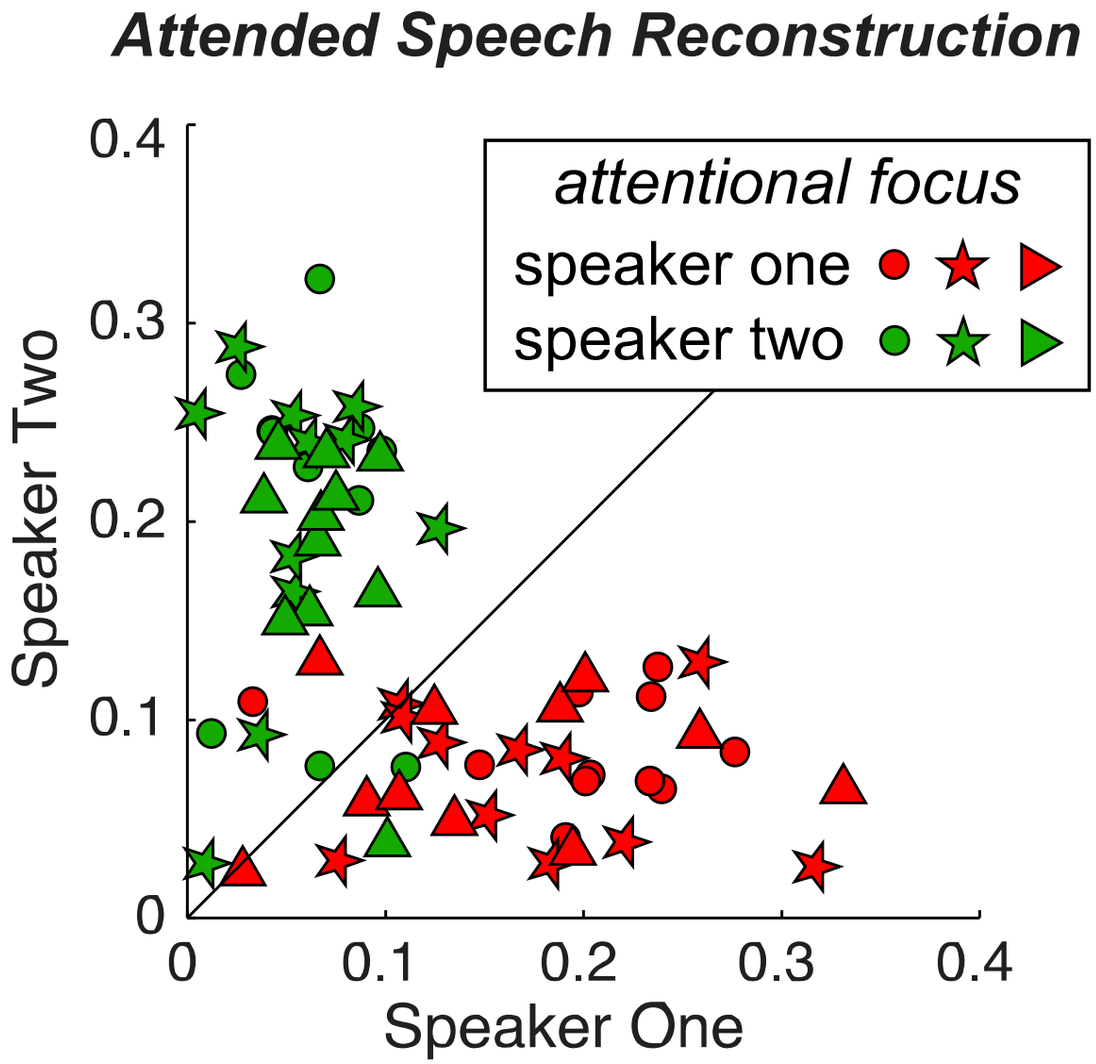
Identical Stimuli!

Stream-Specific Representation



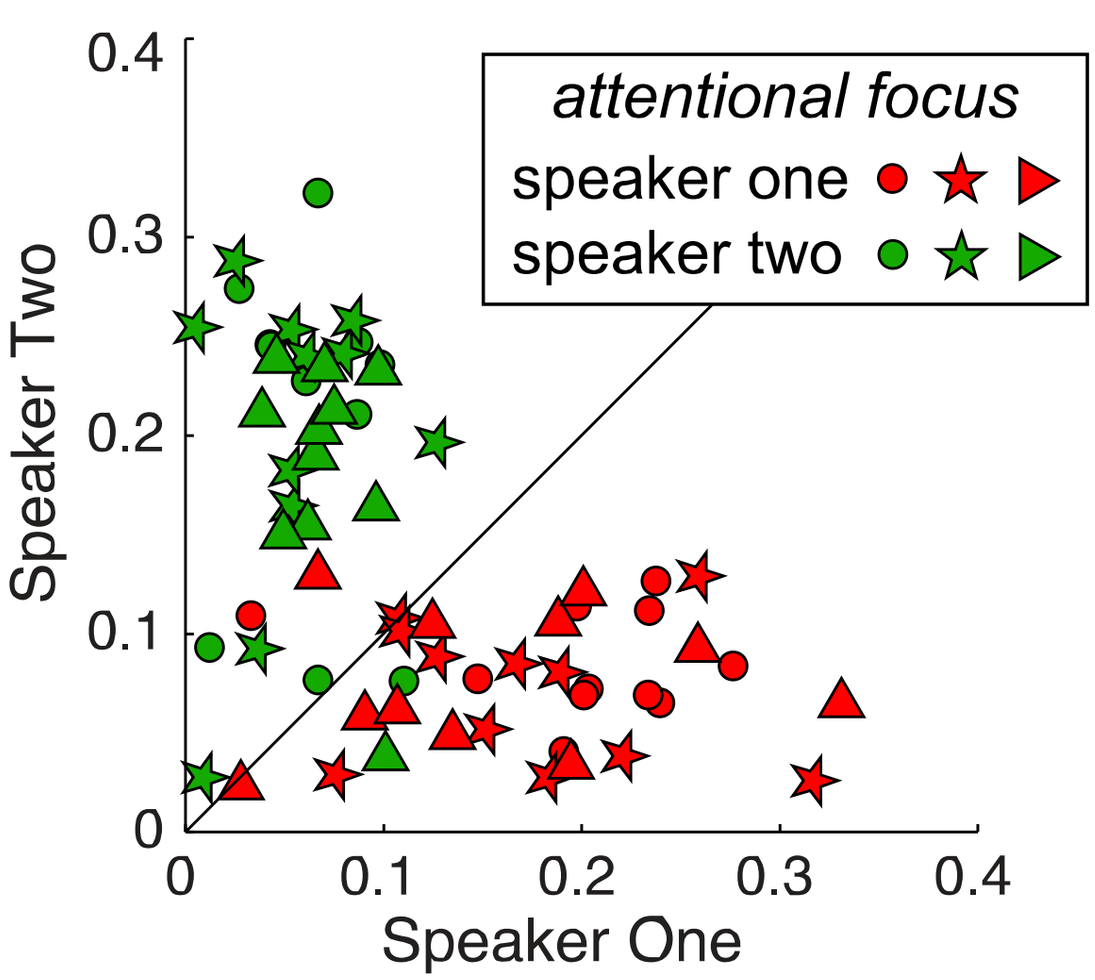
Identical Stimuli!

Single Trial Speech Reconstruction

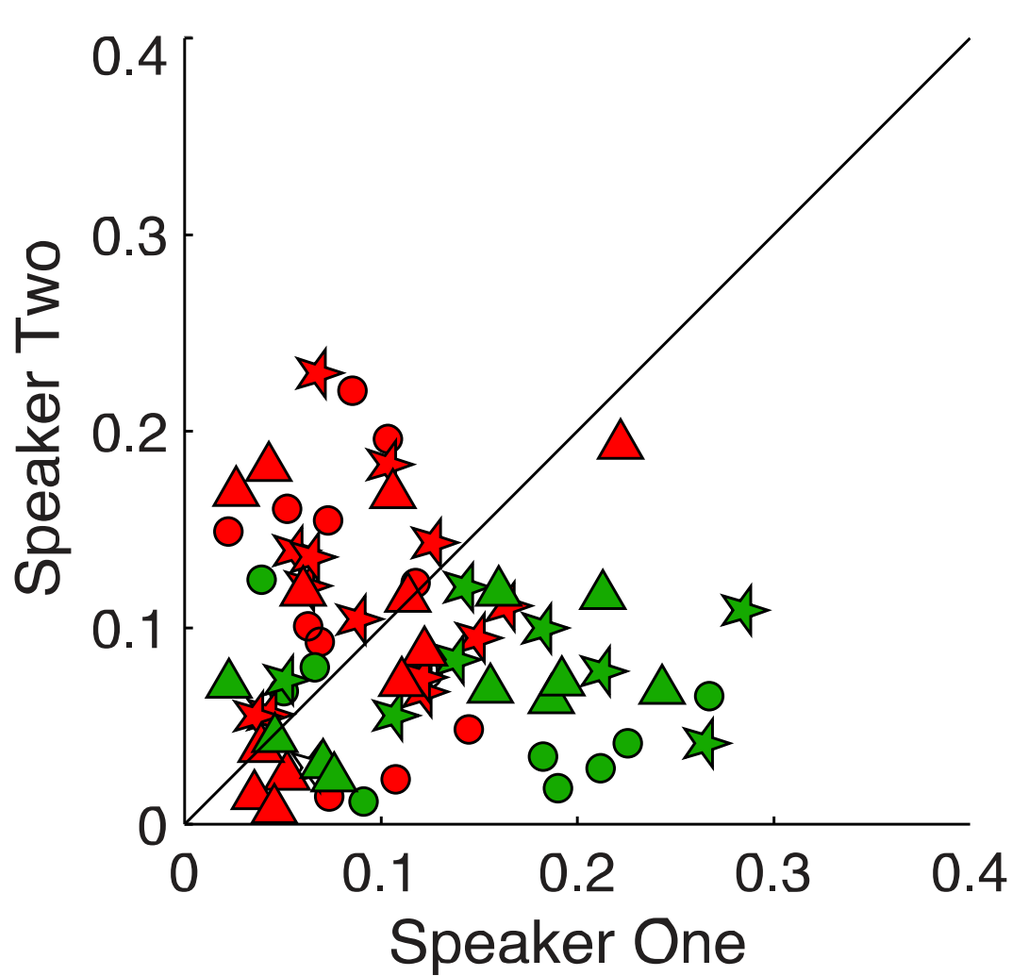


Single Trial Speech Reconstruction

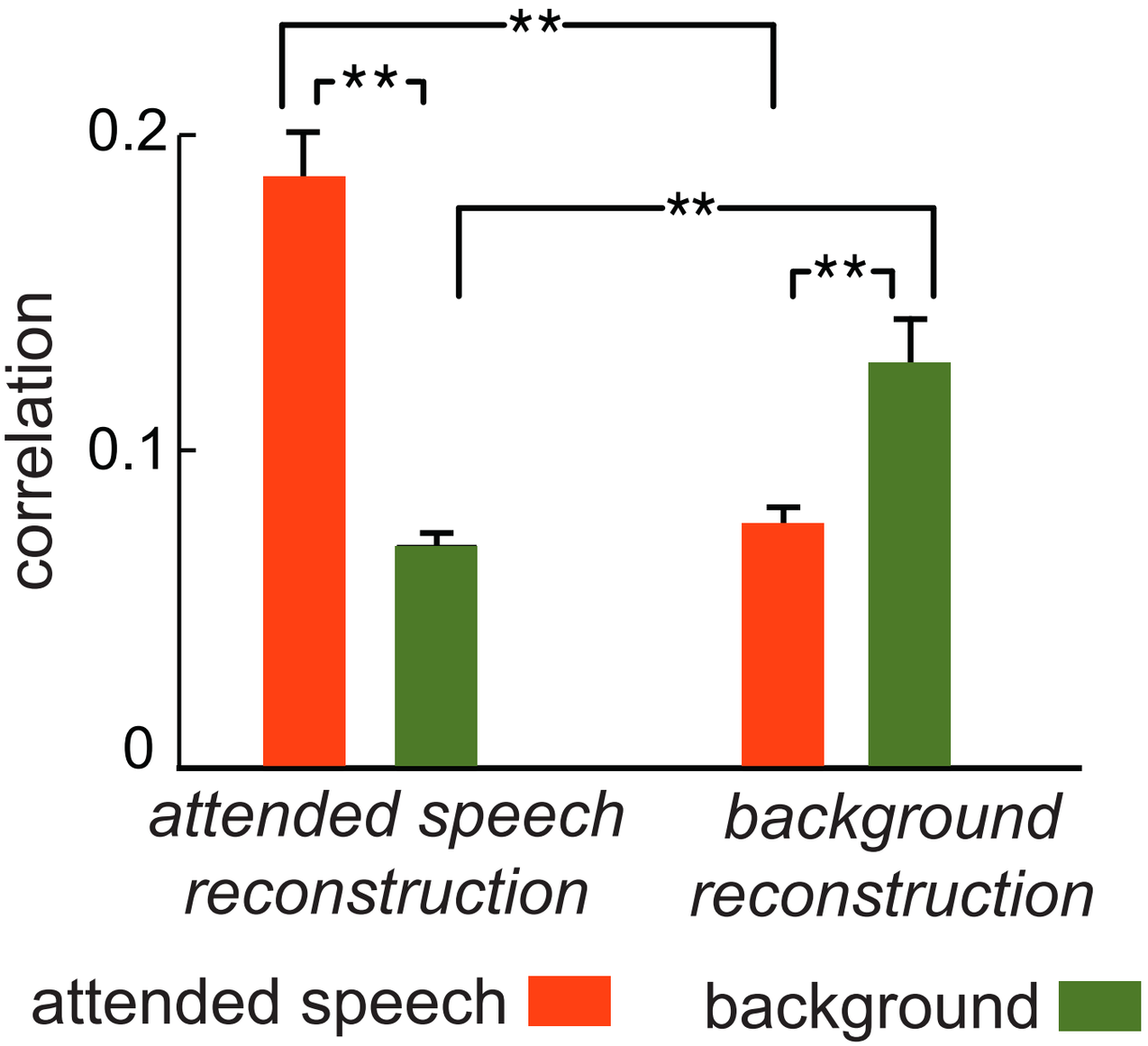
Attended Speech Reconstruction



Background Speech Reconstruction

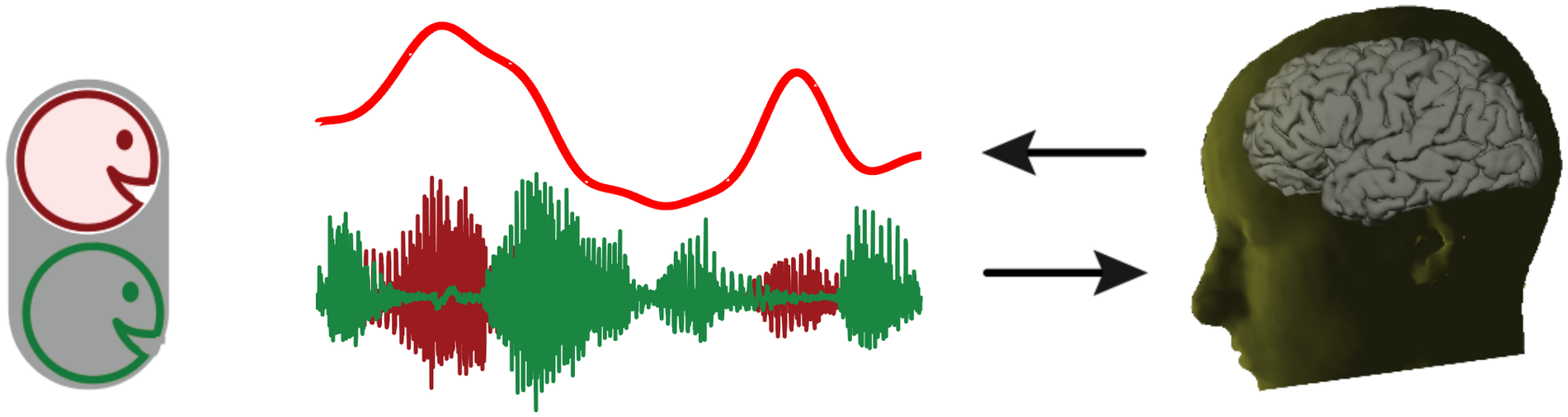


Overall Speech Reconstruction

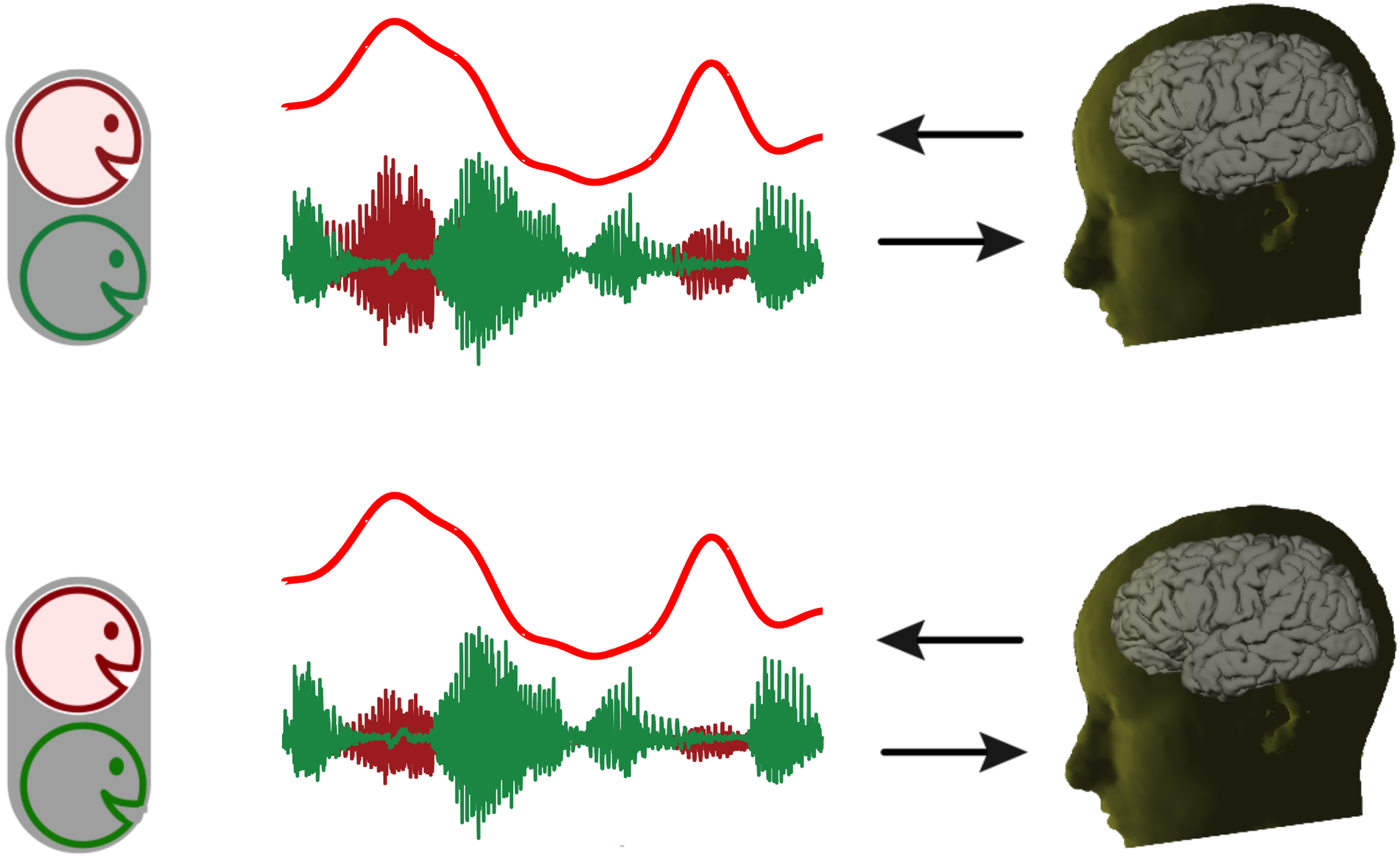


Distinct neural representations for different speech streams

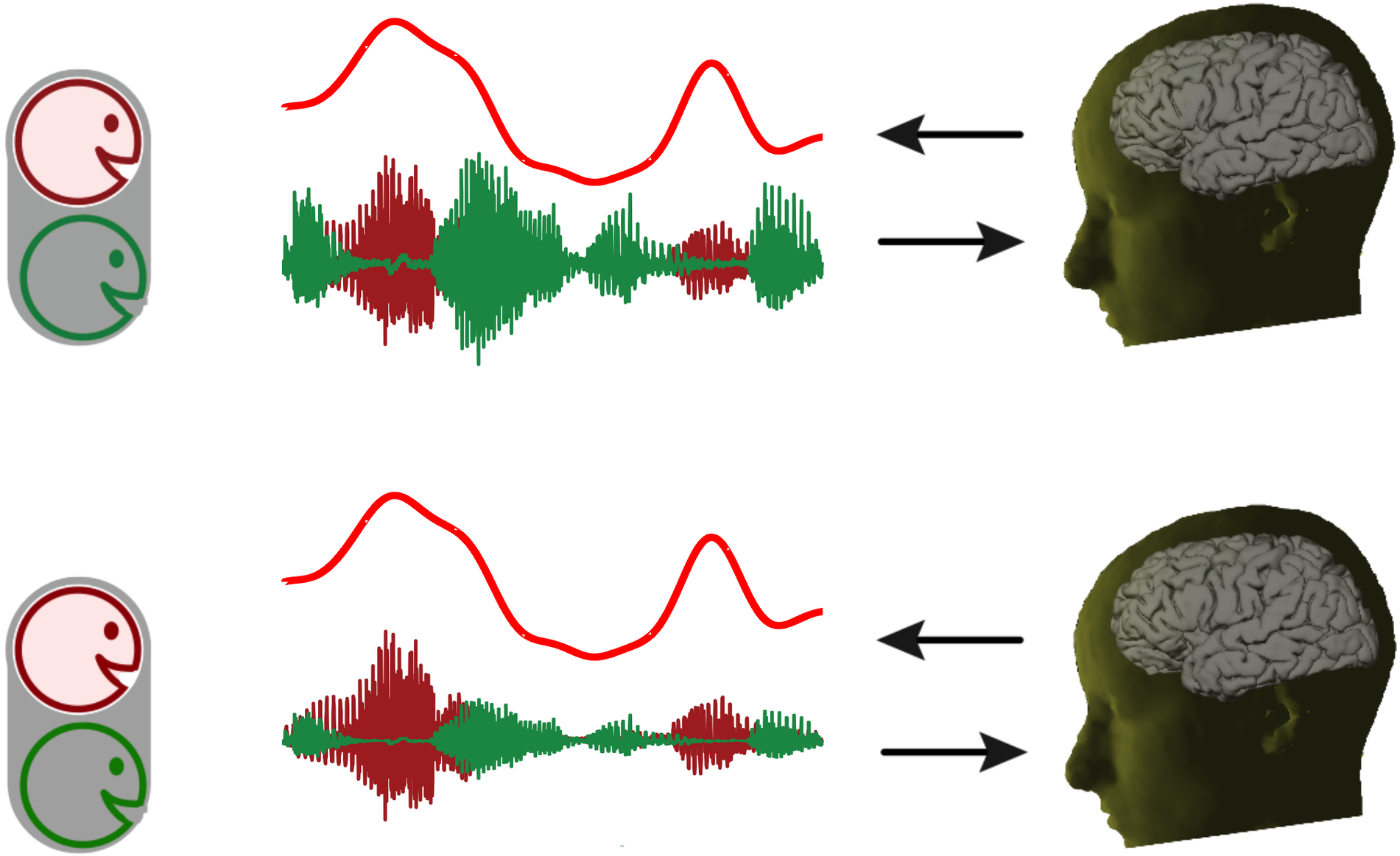
Invariance Under Acoustic Changes



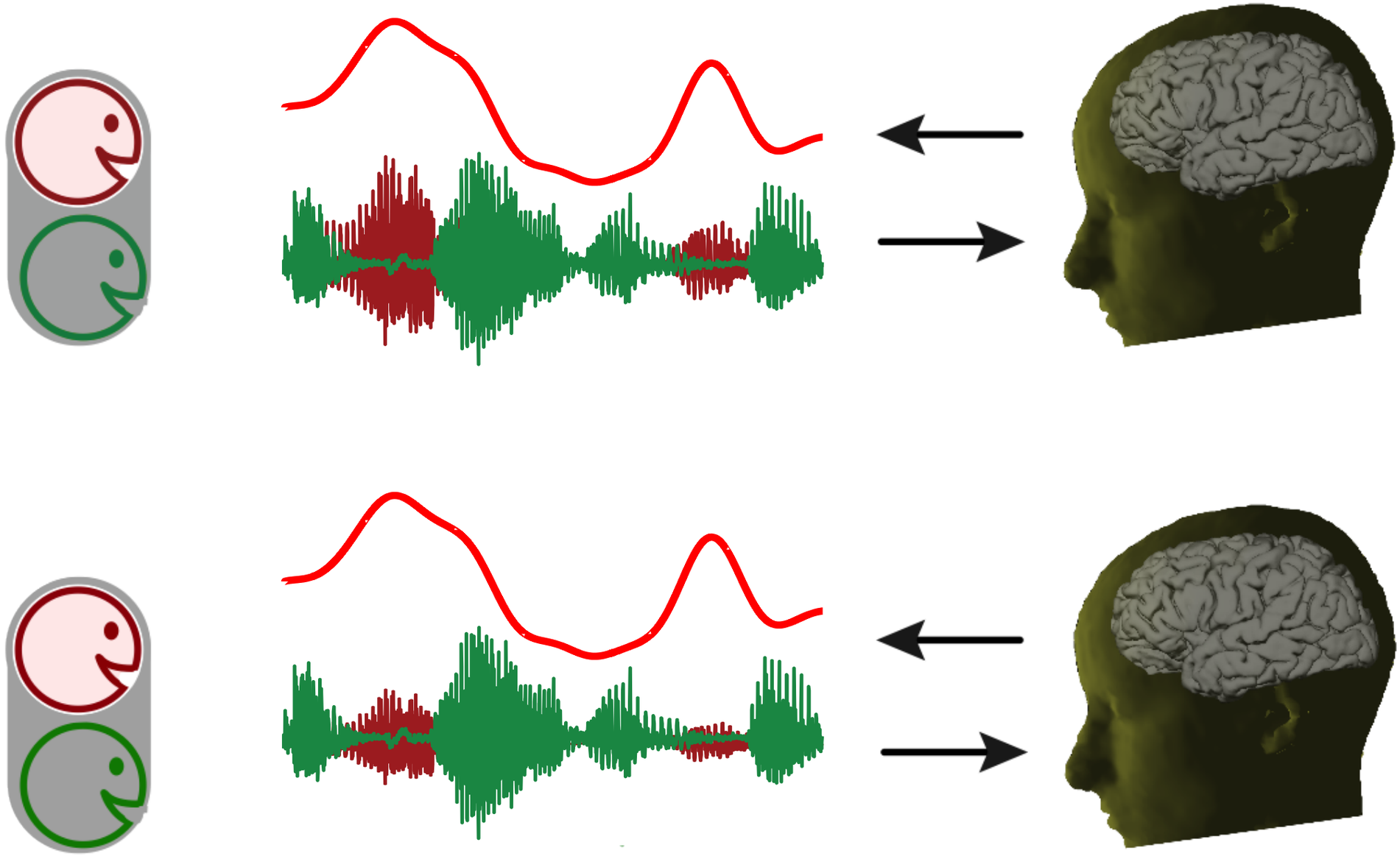
Invariance Under Acoustic Changes



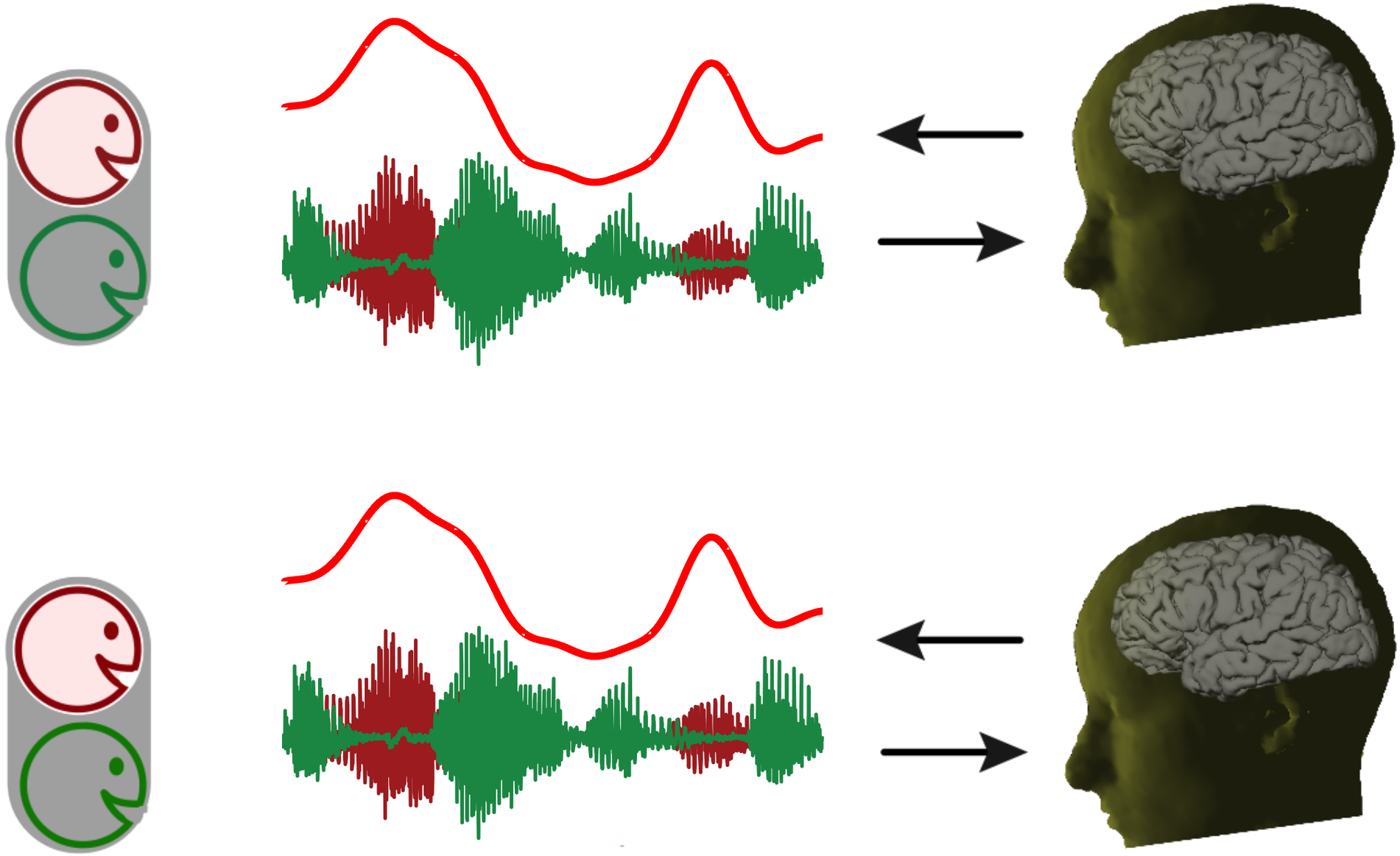
Invariance Under Acoustic Changes



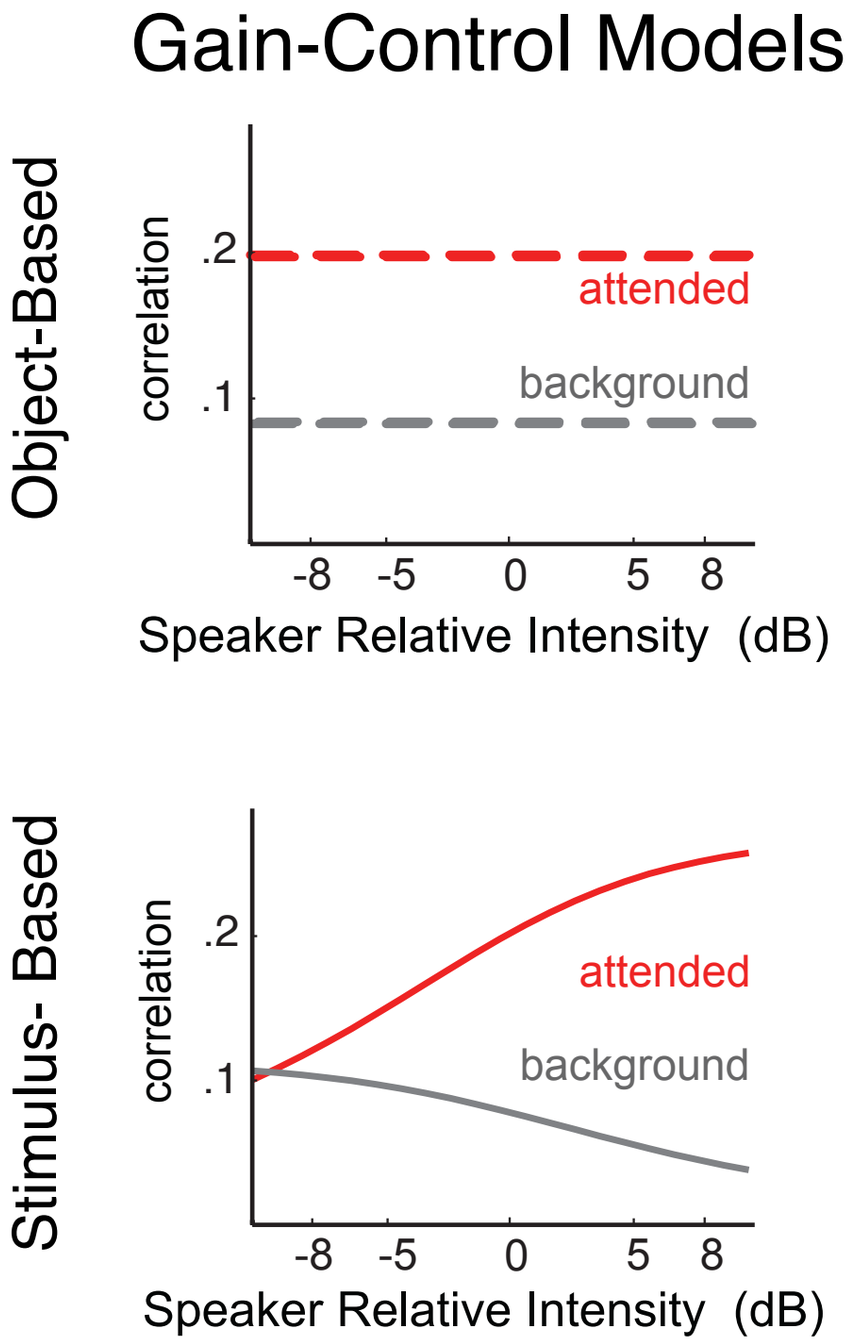
Invariance Under Acoustic Changes



Invariance Under Acoustic Changes



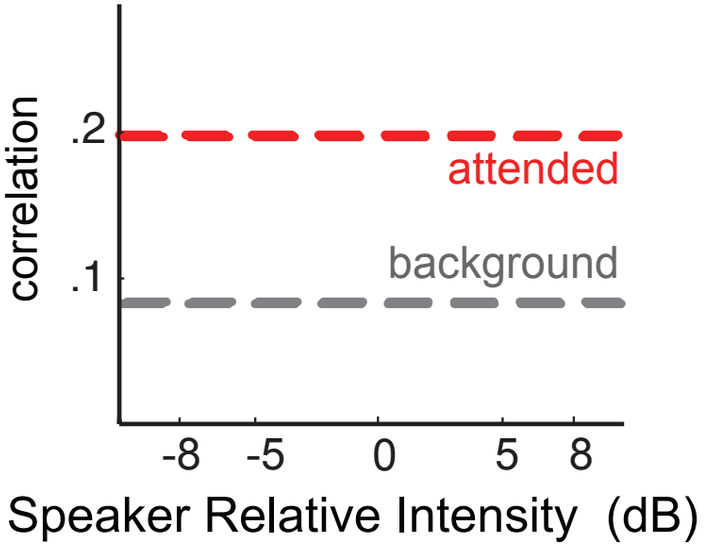
Stream-Based Gain Control?



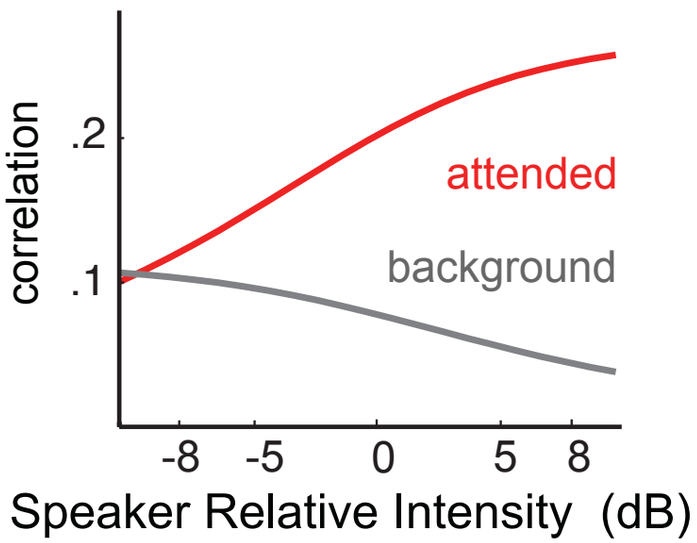
Stream-Based Gain Control?

Gain-Control Models

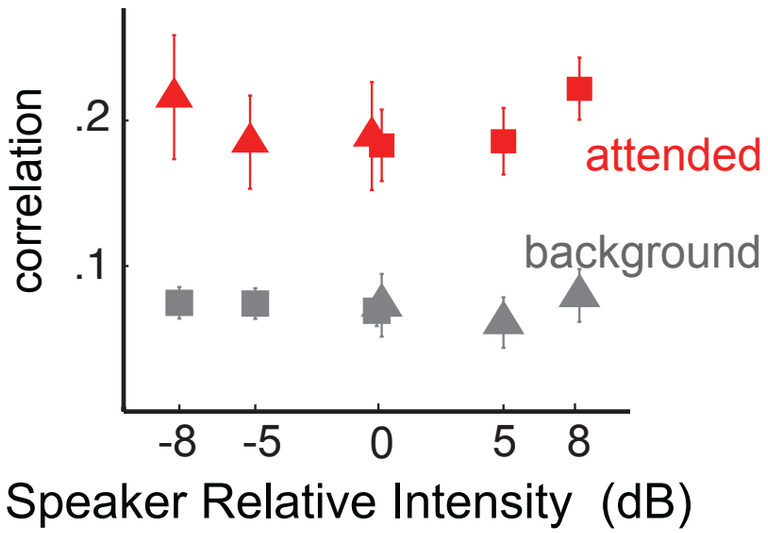
Object-Based



Stimulus- Based



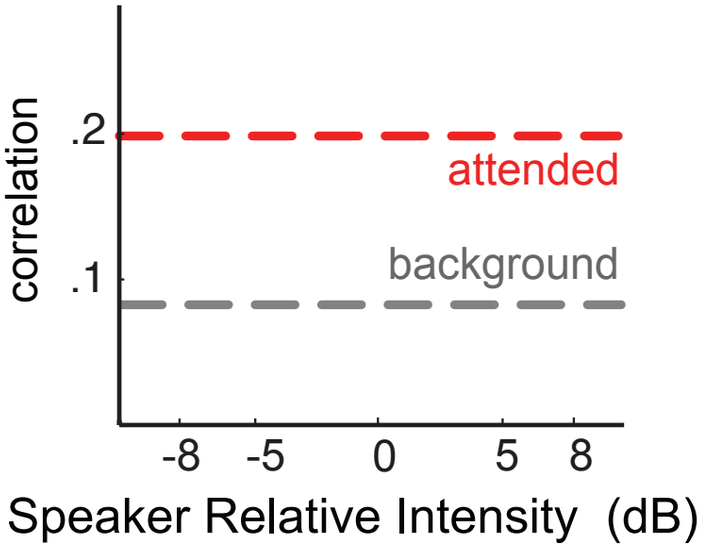
Neural Results



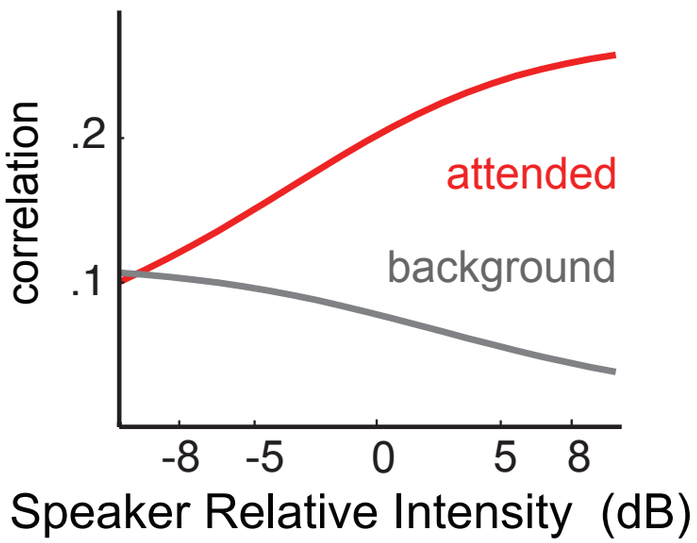
Stream-Based Gain Control?

Gain-Control Models

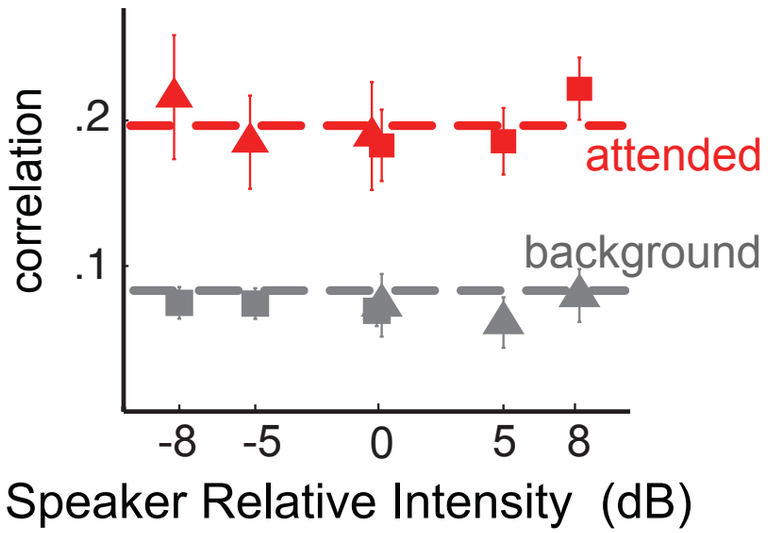
Object-Based



Stimulus- Based



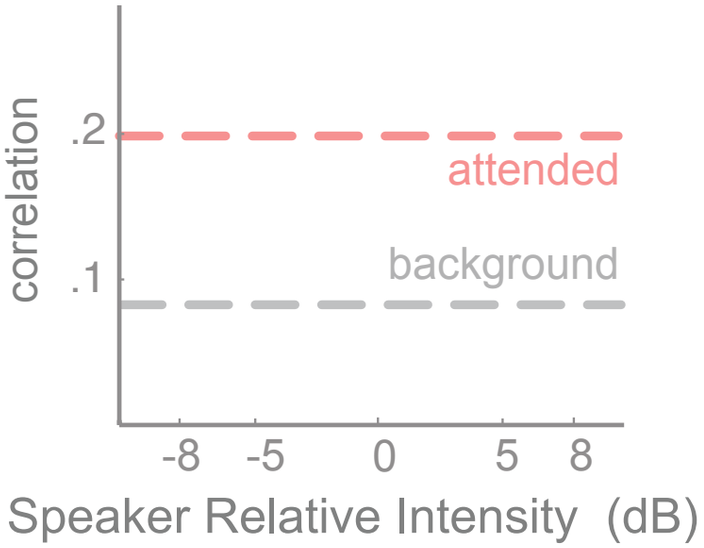
Neural Results



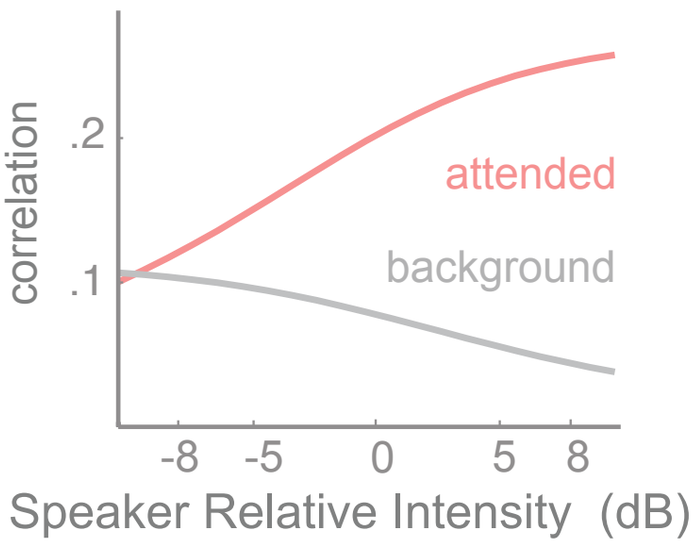
Stream-Based Gain Control?

Gain-Control Models

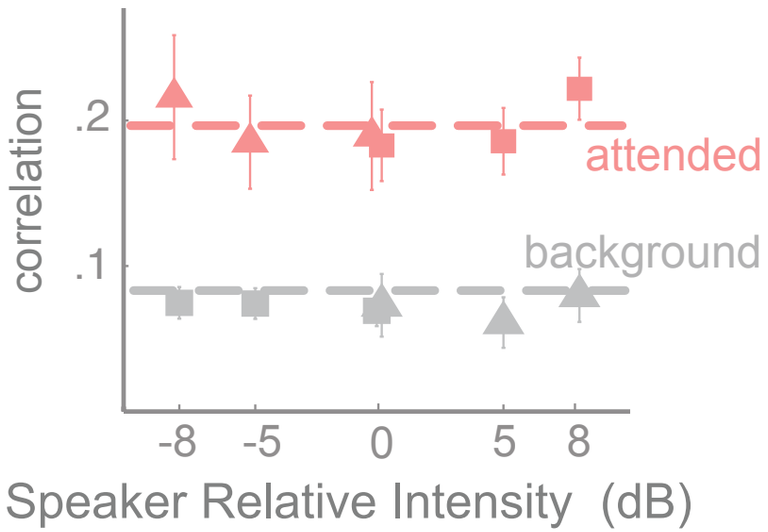
Object-Based



Stimulus- Based



Neural Results



- Stream-based not stimulus-based
- Neural representation is invariant to acoustic changes.

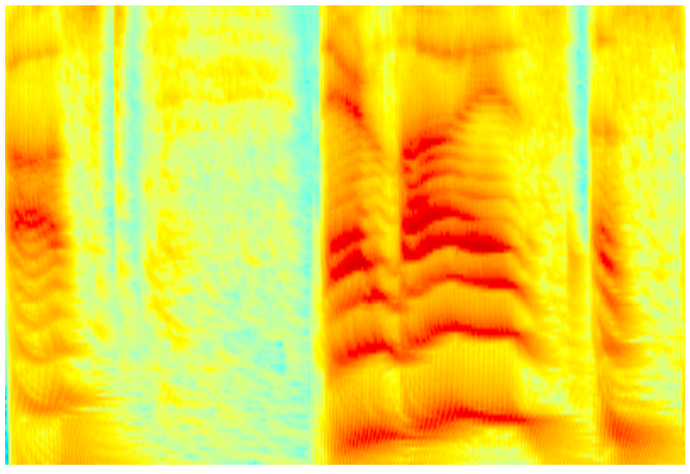
Neural Representation of an Auditory Object

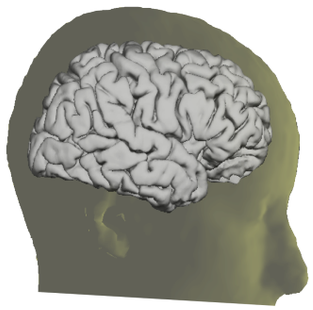
✓ neural representation is of something in sensory world

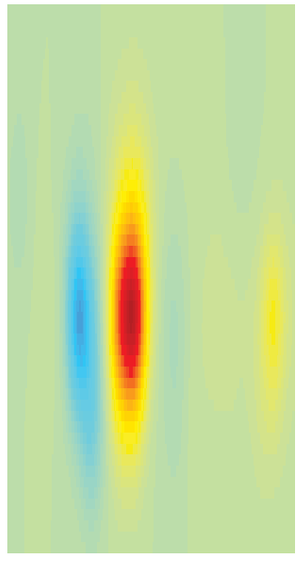
✓ when other sounds mixed in,
neural representation is of auditory object,
not entire acoustic scene

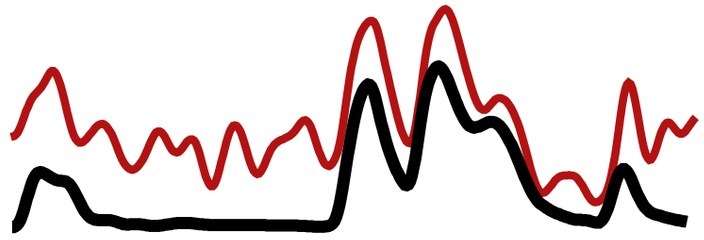
✓ neural representation invariant
under broad changes in specific acoustics

Forward STRF Model



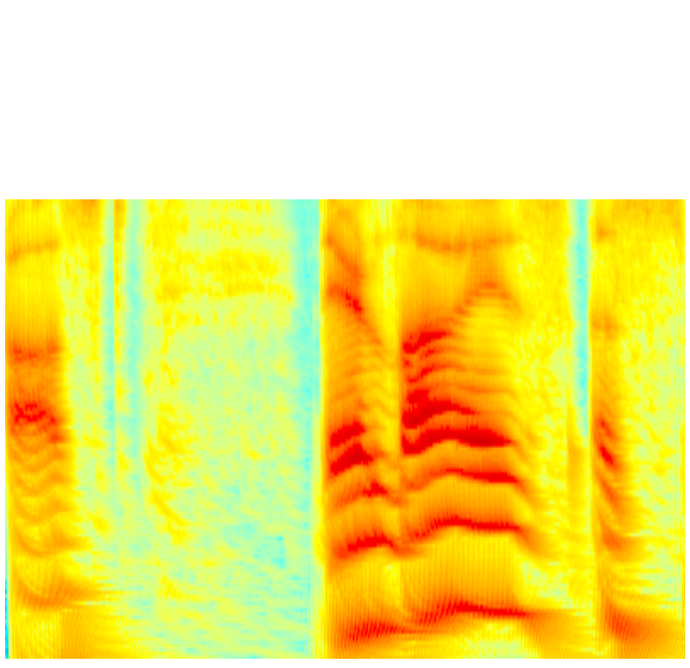


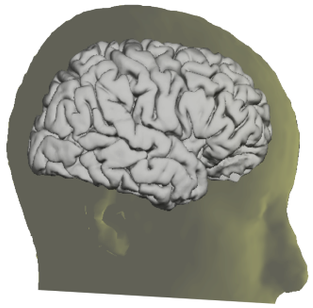


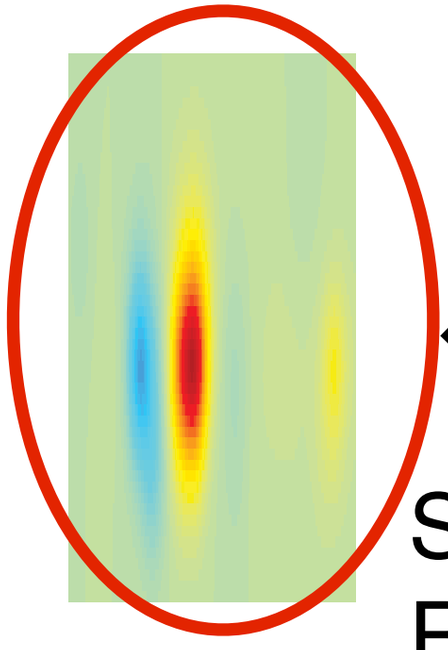


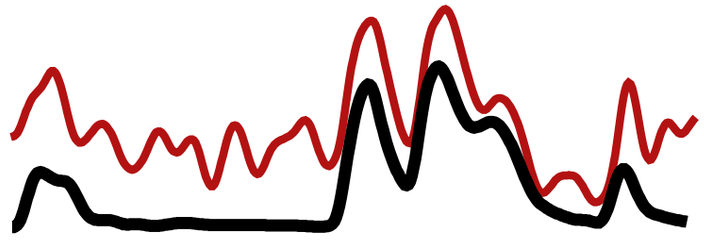
Spectro-Temporal Response Function (STRF)

Forward STRF Model



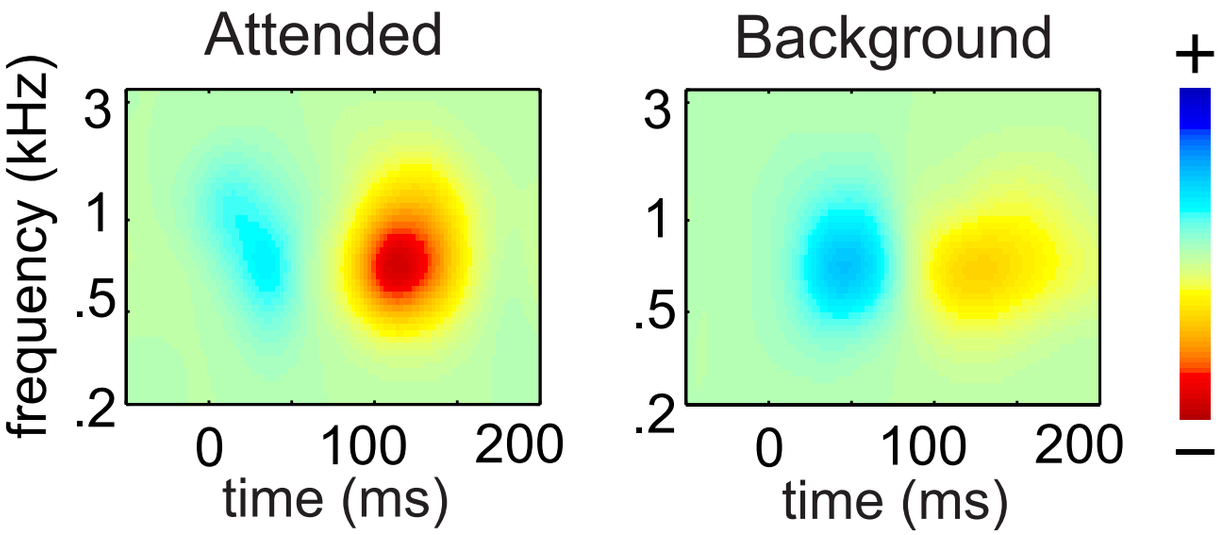






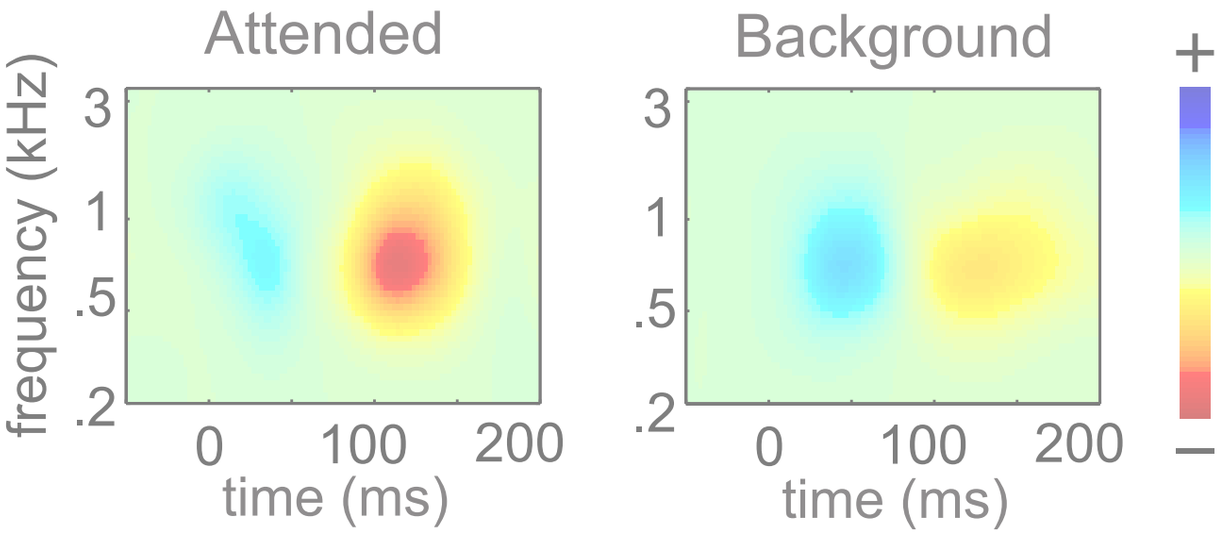
Spectro-Temporal Response Function (STRF)

STRF Results

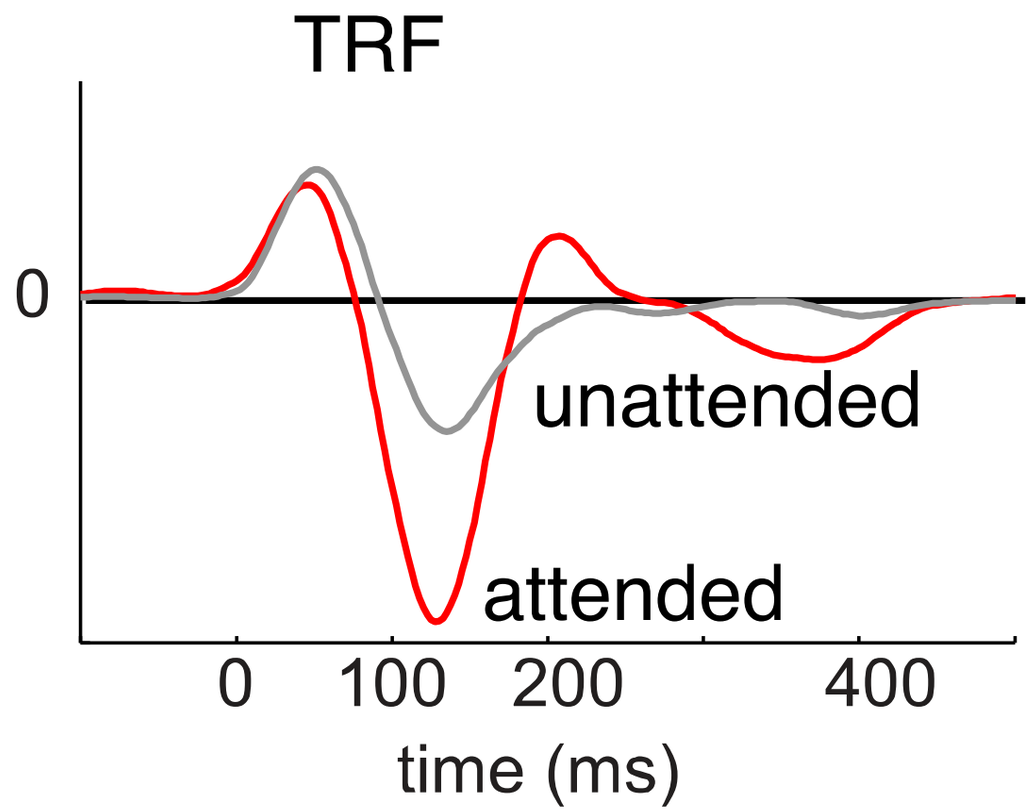


- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- $M50_{STRF}$ positive peak
- $M100_{STRF}$ negative peak

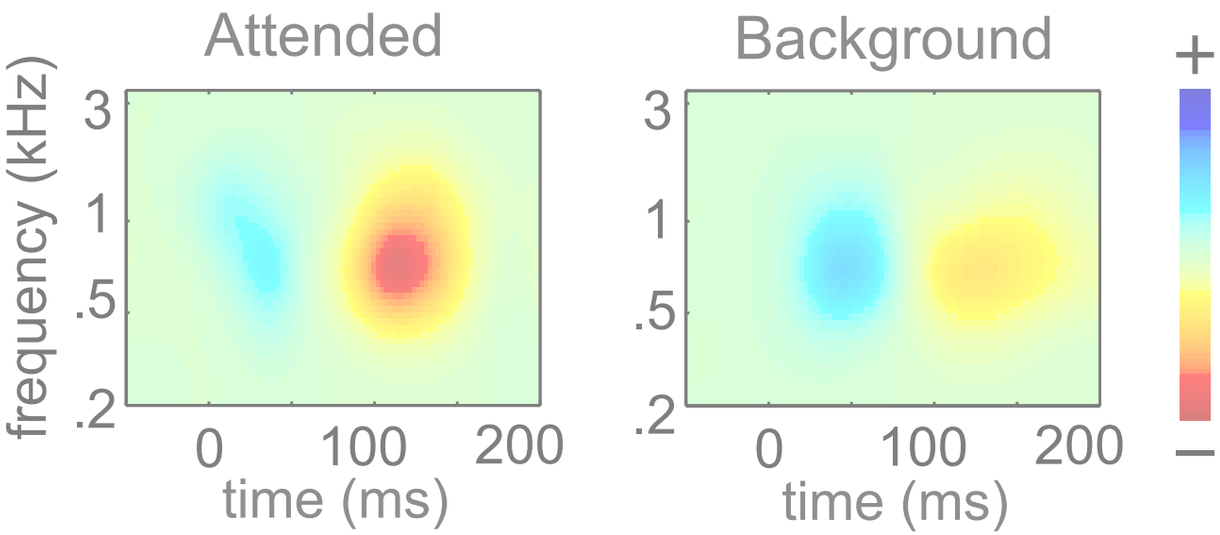
STRF Results

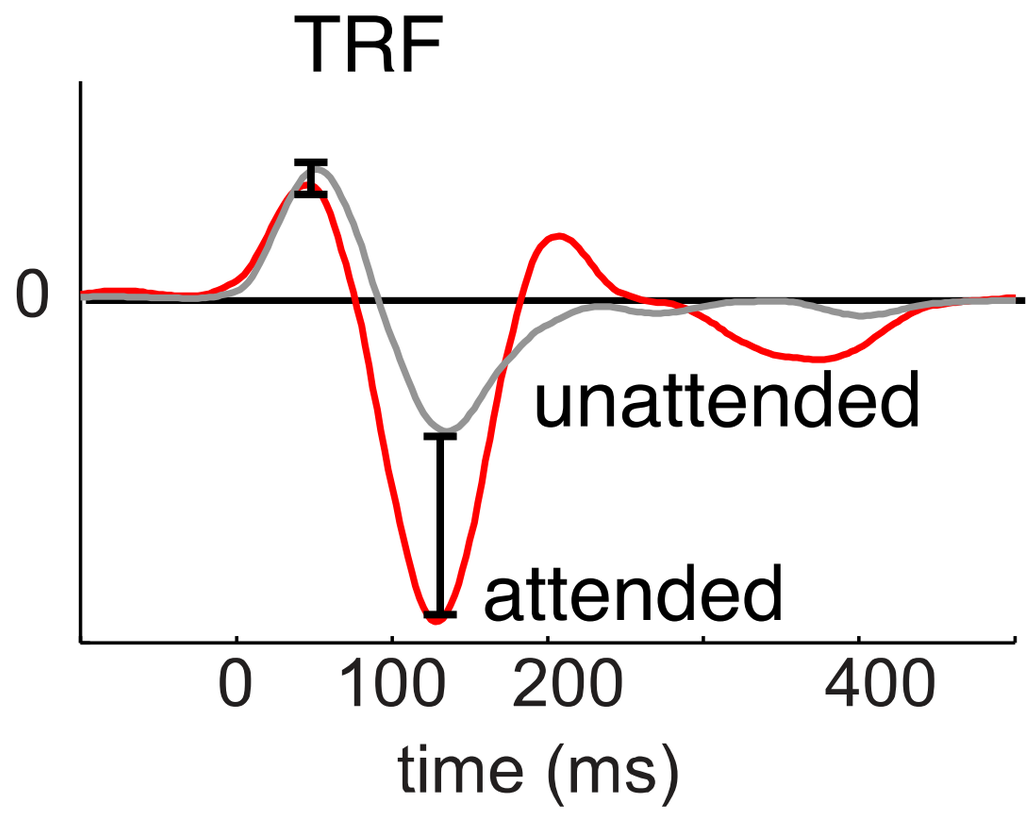


- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- $M50_{STRF}$ positive peak
- $M100_{STRF}$ negative peak



STRF Results



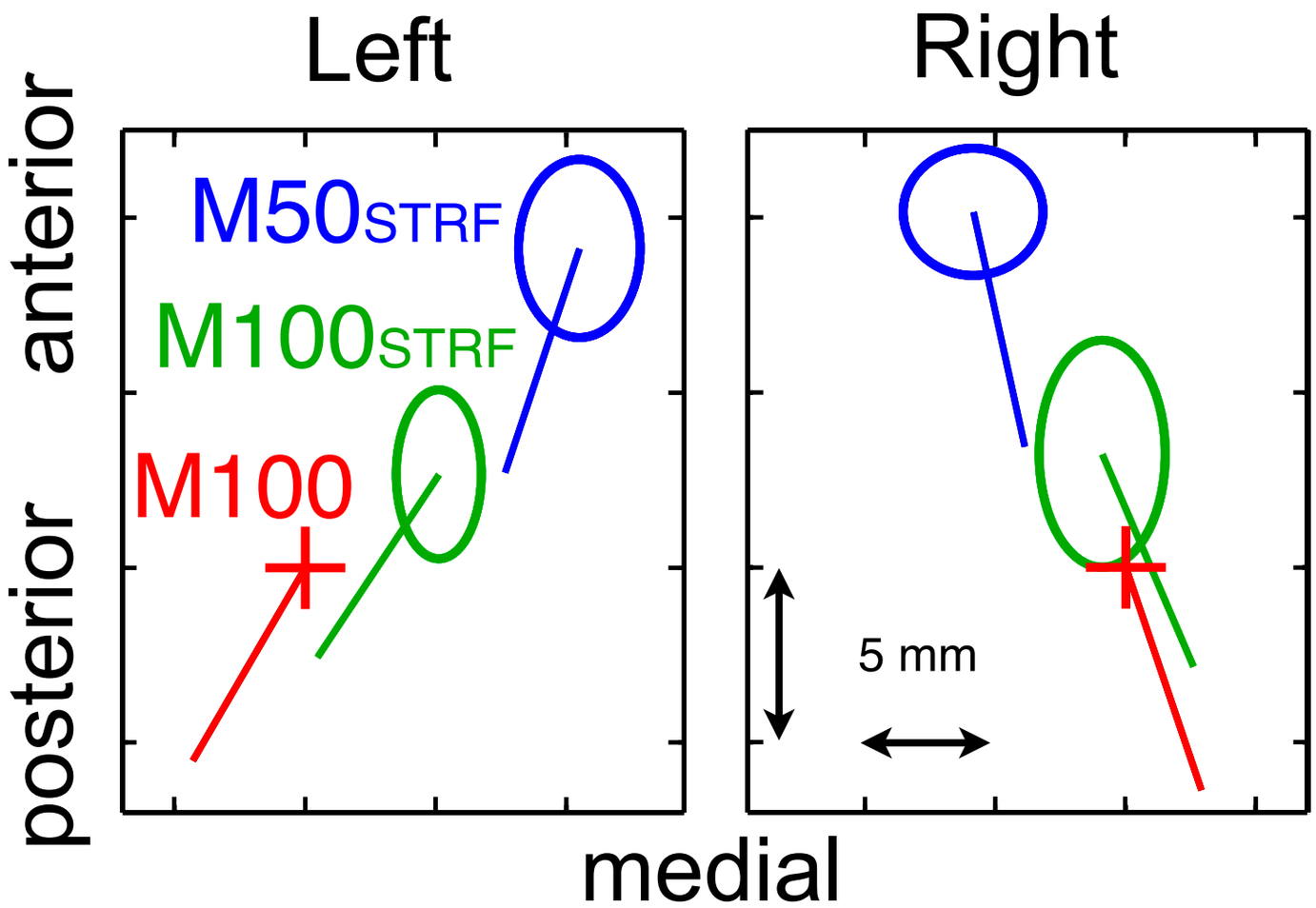


- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- $M50_{STRF}$ positive peak
- $M100_{STRF}$ negative peak
- **$M100_{STRF}$ strongly modulated by attention, *but not $M50_{STRF}$***

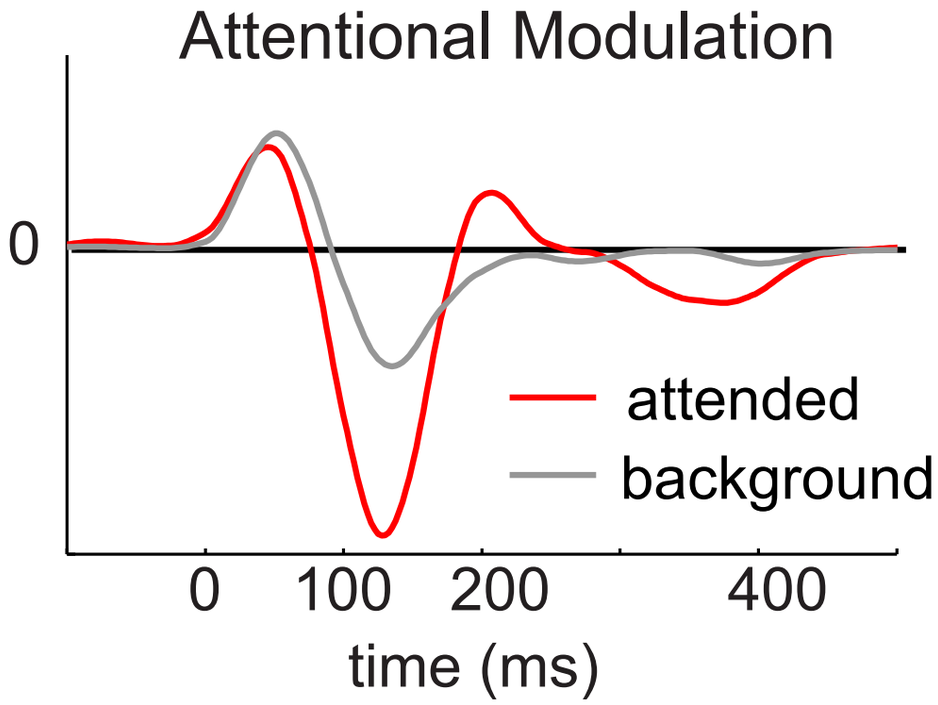
Neural Sources

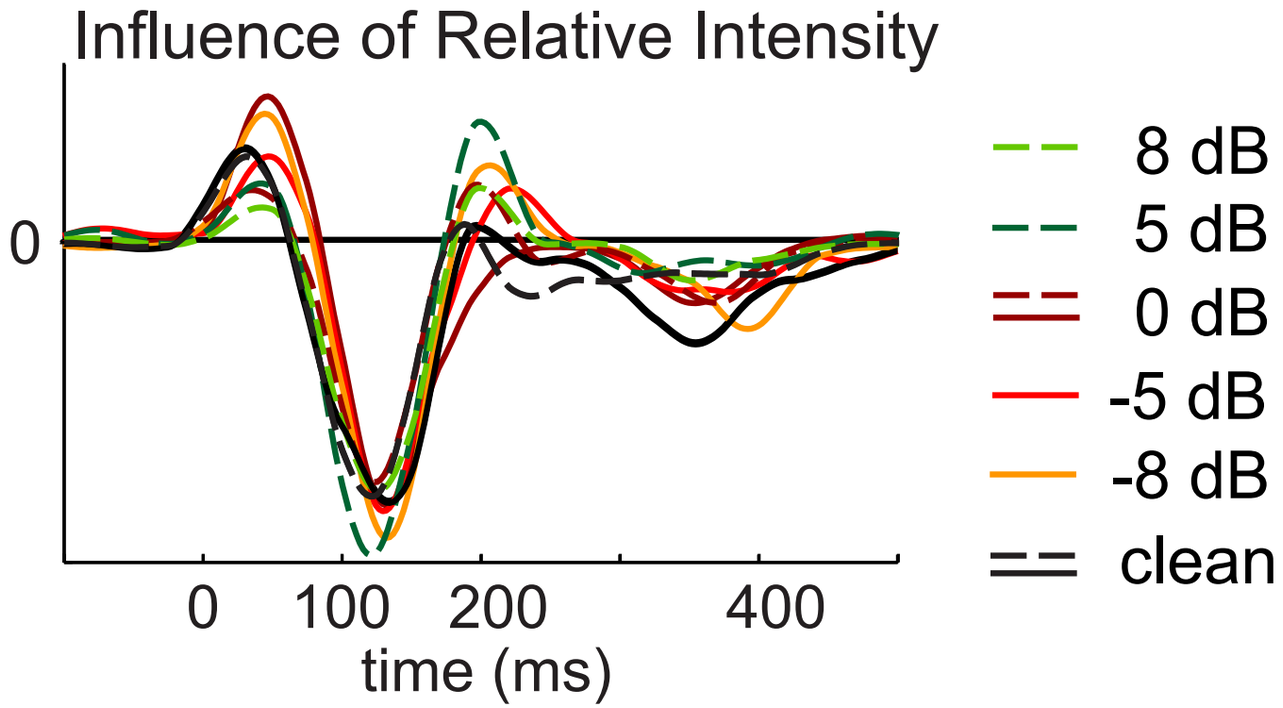
- $M100_{STRF}$ source near (same as?) M100 source: PT

- $M50_{STRF}$ source is anterior and medial to M100 (same as M50?): HG



Cortical Object-Processing Hierarchy

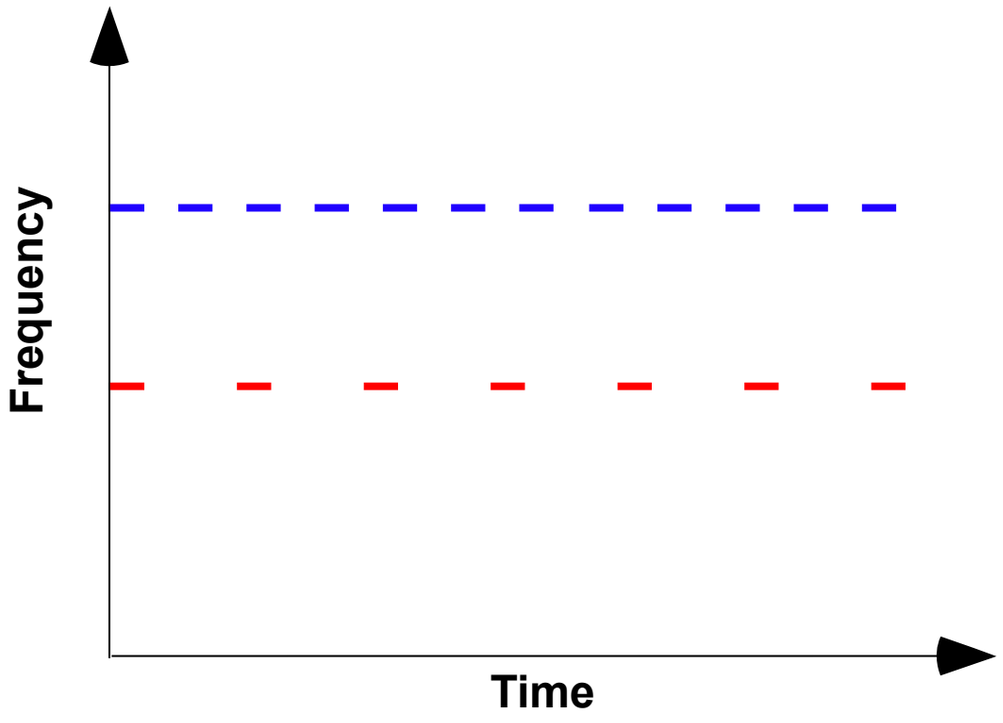




- $M100_{STRF}$ strongly modulated by attention, but not $M50_{STRF}$.
- $M100_{STRF}$ invariant against acoustic changes.
- Objects well-neurally represented at 100 ms, but not 50 ms.

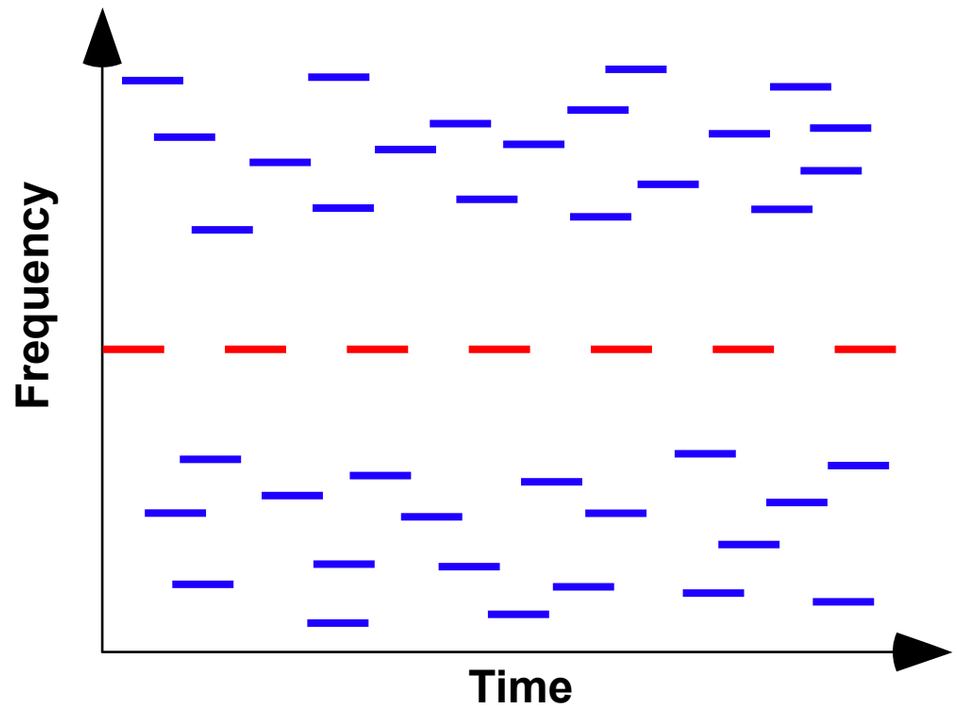
Not Just Speech

Competing Tone Streams



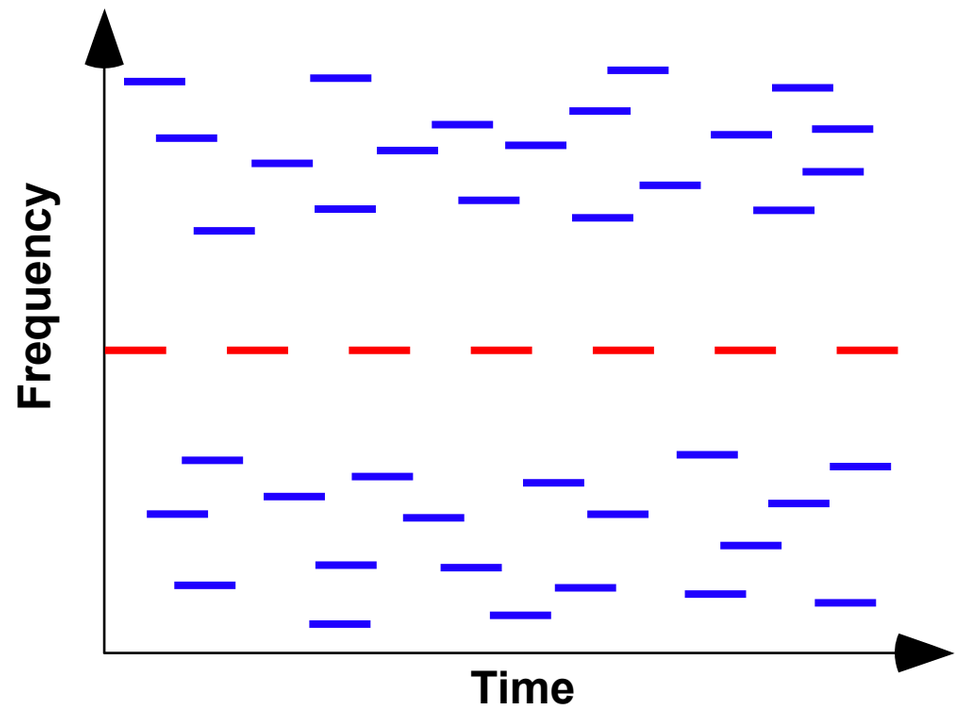
Xiang et al., J Neuroscience (2010)

Tone Stream in Masker Cloud

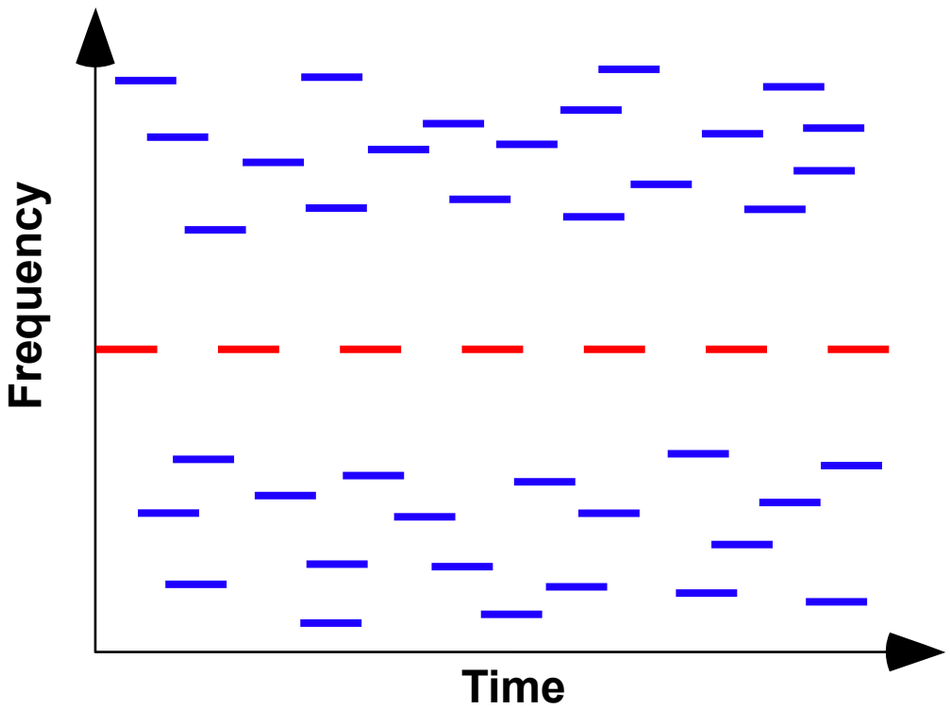


Elhilali et al., PLoS Biology (2009)

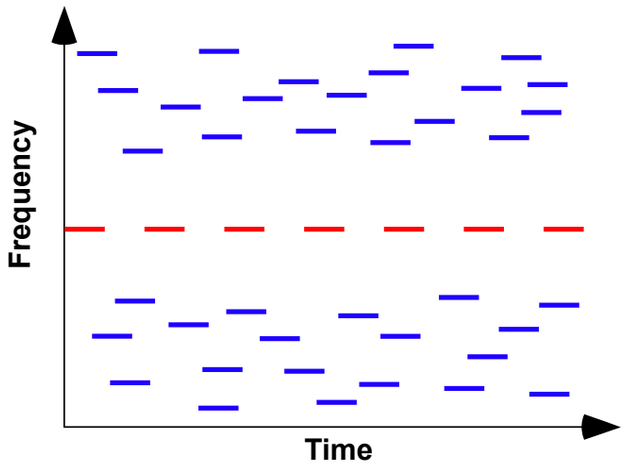
Tone Stream in Masker Cloud

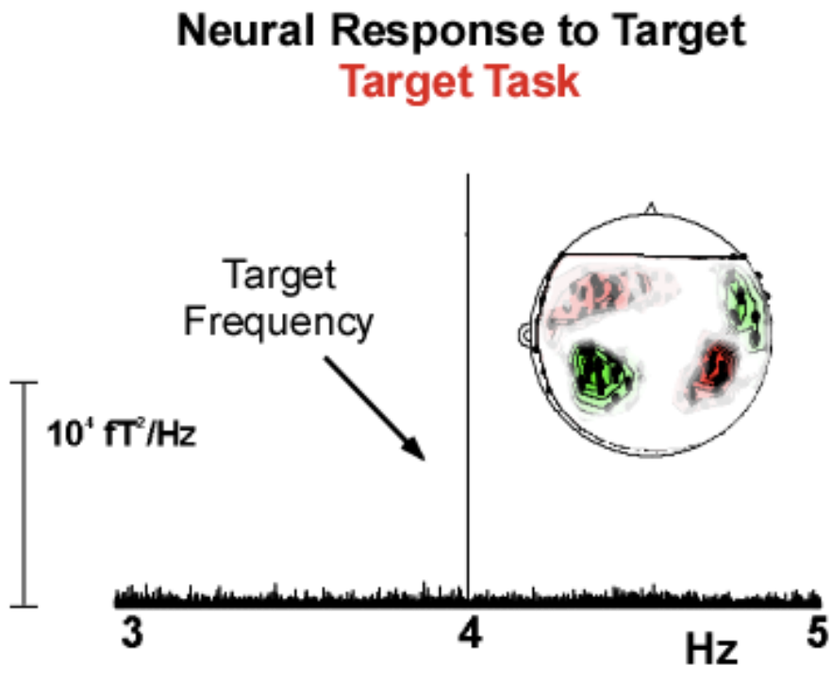


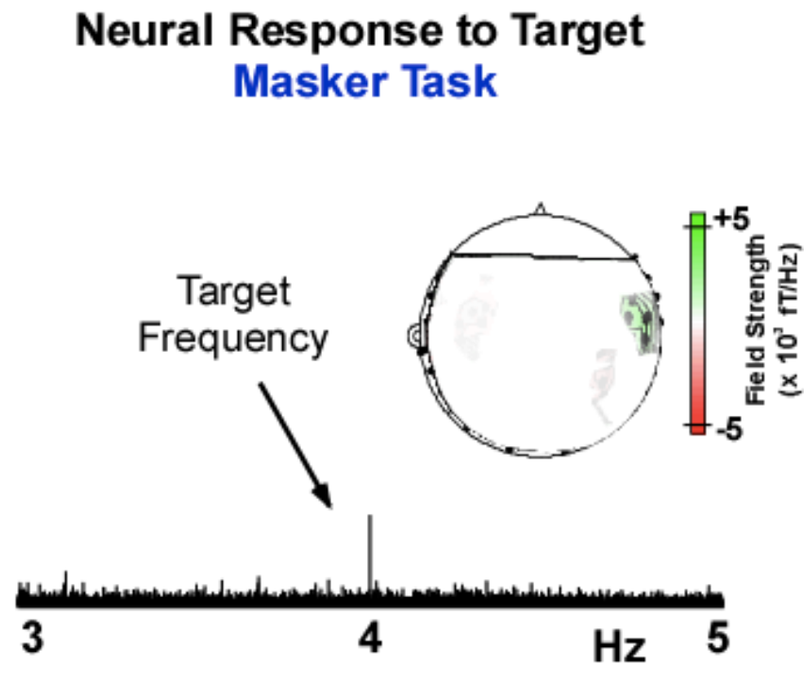
Tone Stream in Masker Cloud



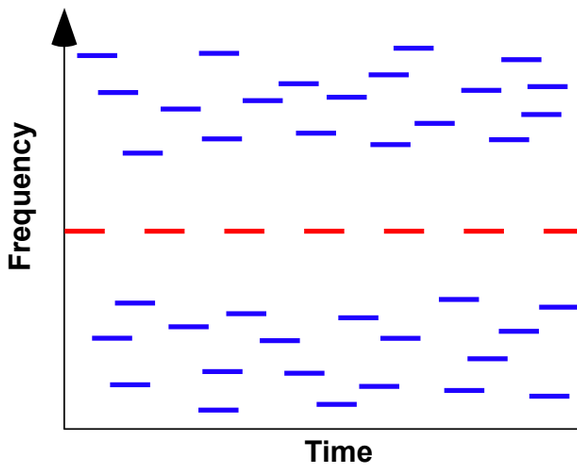
Tone Stream in Masker Cloud

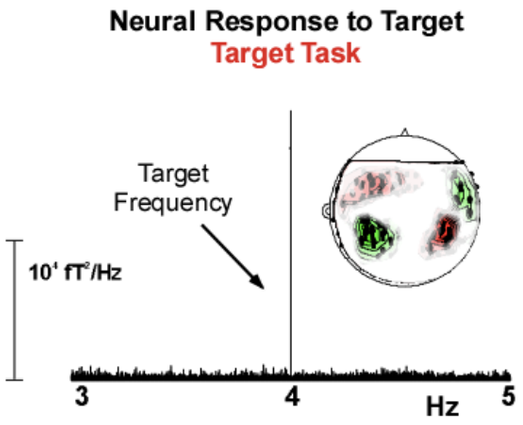


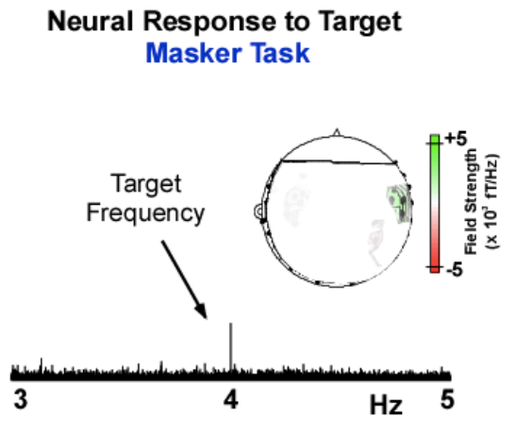


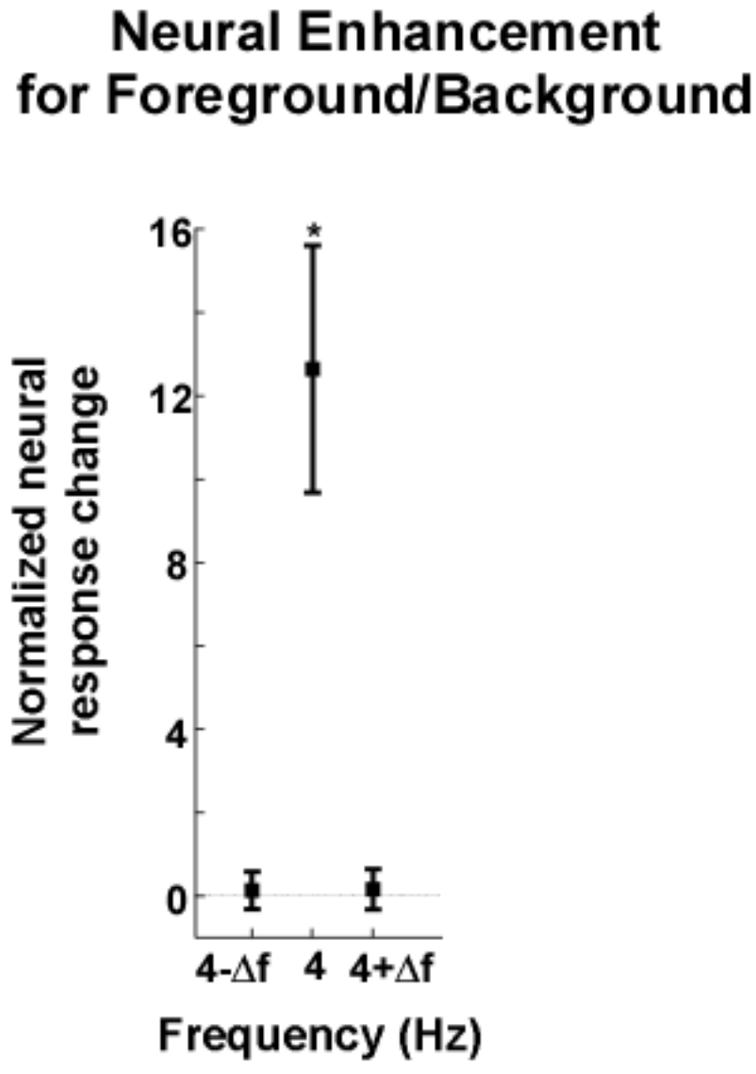


Tone Stream in Masker Cloud

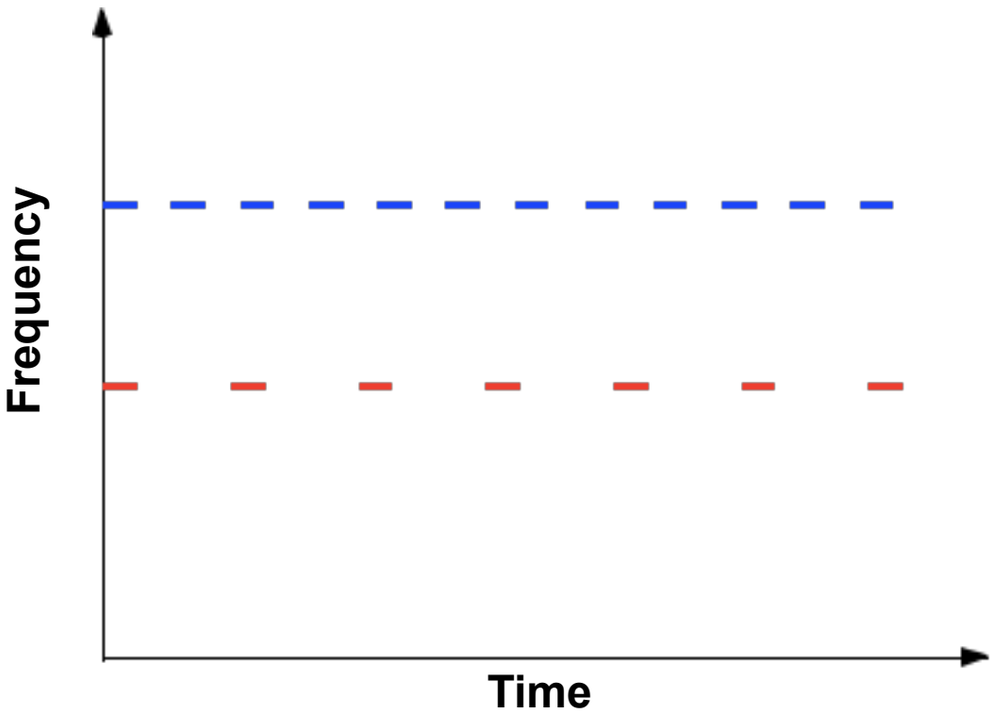




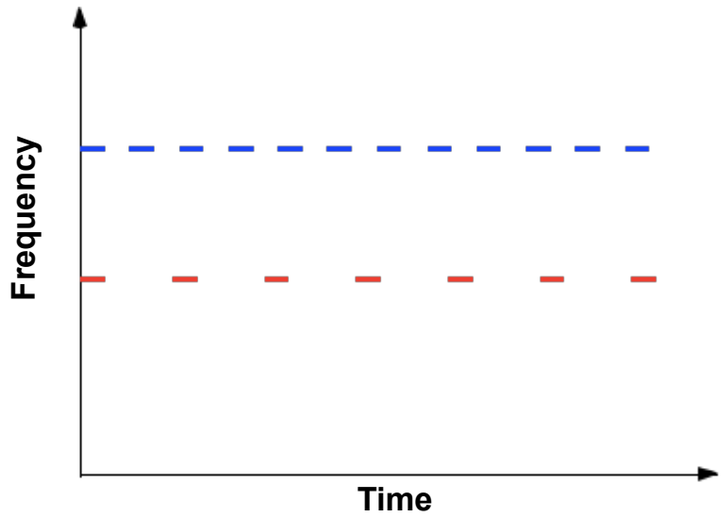


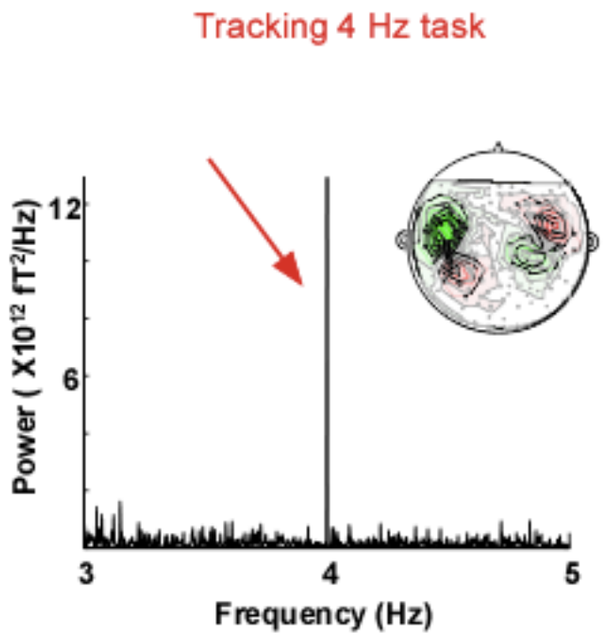


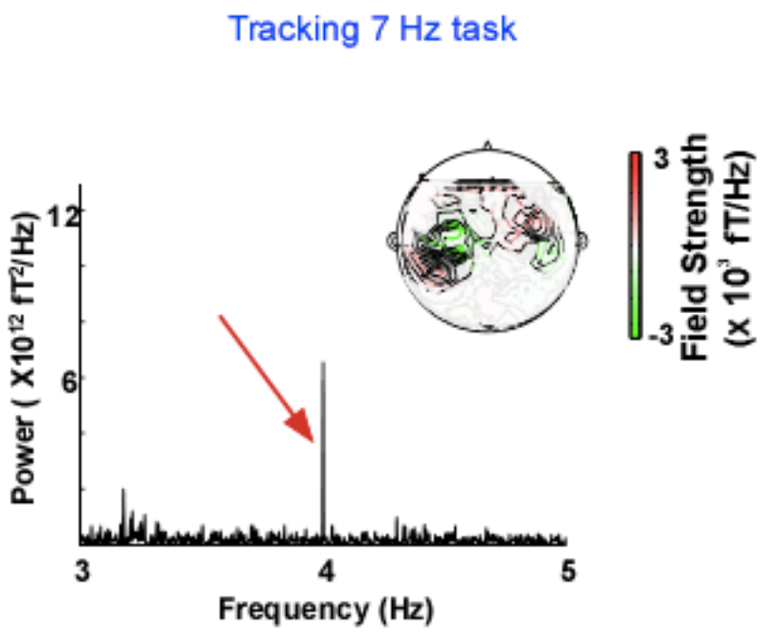
Competing Tone Streams

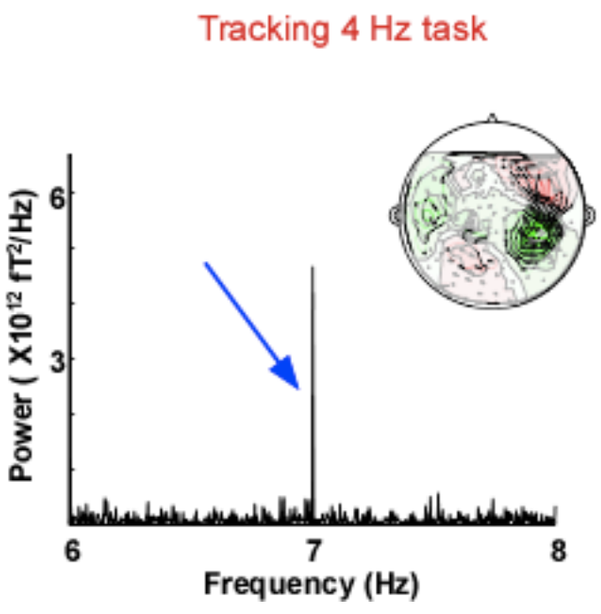


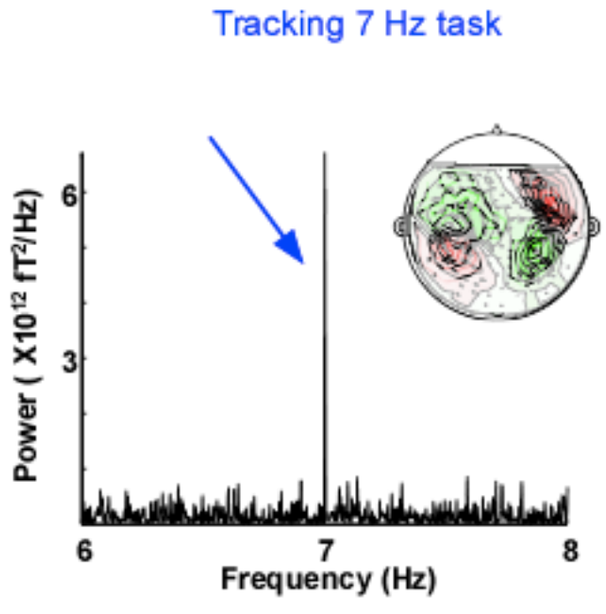
Competing Tone Streams



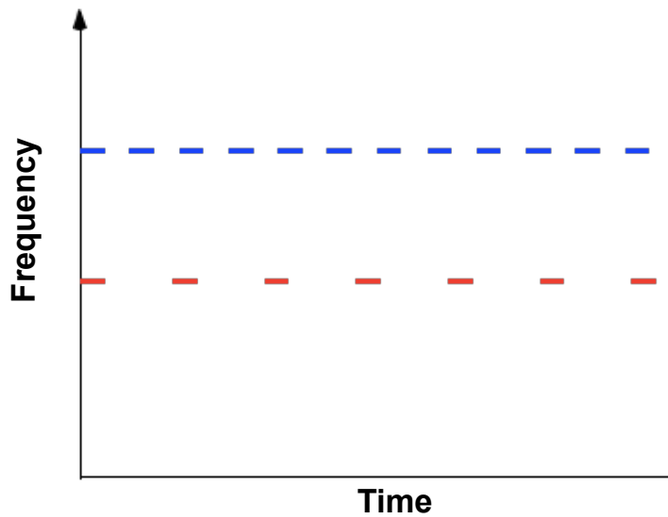


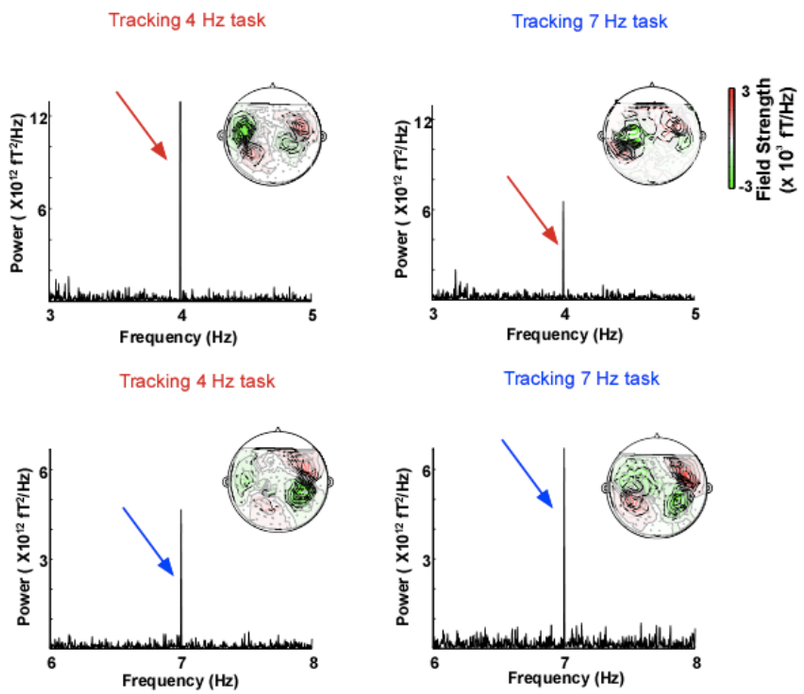


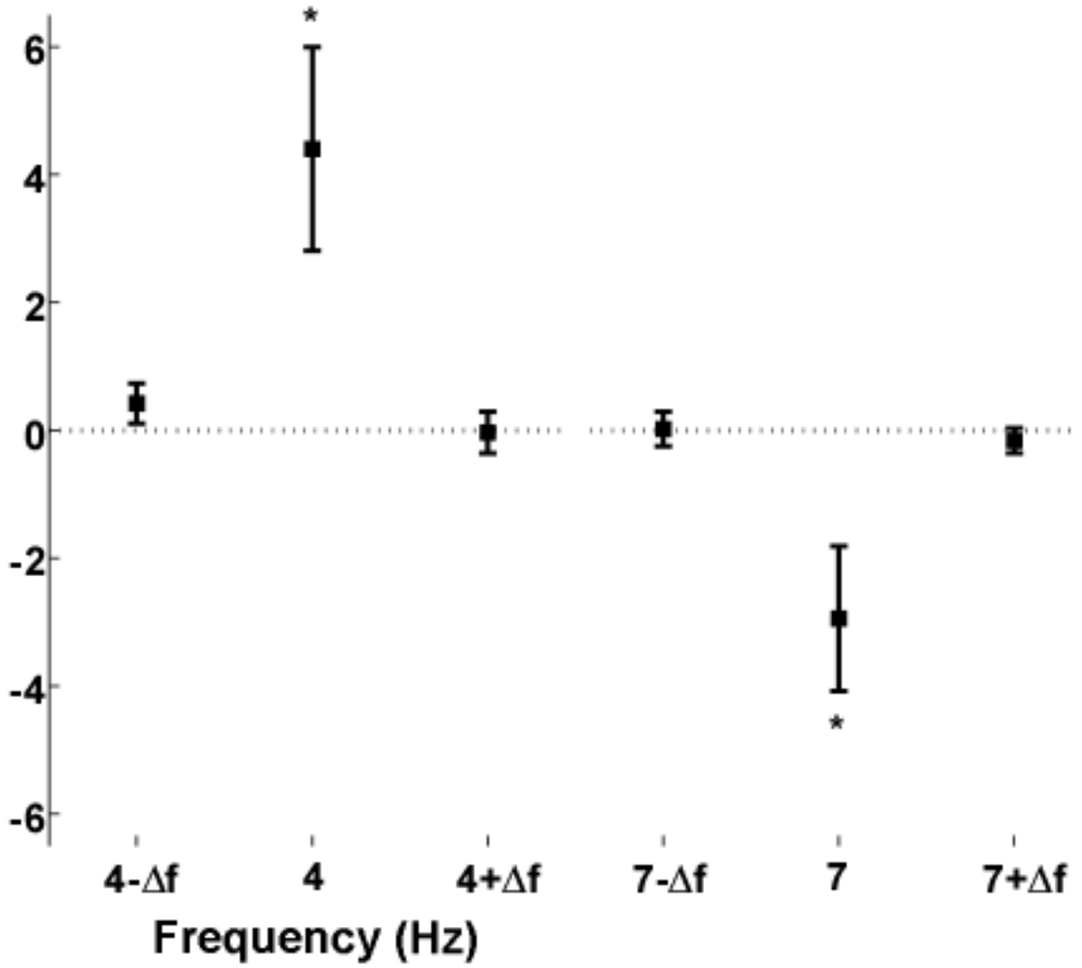




Competing Tone Streams







Summary

- Cortical representations of speech found here:
  - ✓ consistent with being *neural* representations of auditory *(perceptual)* objects
    - ✓ meet 3 formal criteria for auditory objects
- Object representation fully formed by 100 ms latency (PT), but not by 50 ms (HG)
- Not special to speech

Acknowledgements

Grad Students

Francisco Cervantes
Marisel Villafane Delgado
Kim Drnec
Krishna Puvvada

Past Grad Students

Nayef Ahmar
Claudia Bonin
Maria Chait
Nai Ding
Victor Grau-Serrat
Ling Ma
Raul Rodriguez
Juanjuan Xiang
Kai Sum Li
Jiachen Zhuo

Undergraduate Students

Abdulaziz Al-Turki
Nicholas Asendorf
Sonja Bohr
Elizabeth Camenga
Corinne Cameron
Julien Dagenais
Katya Dombrowski
Kevin Kahn
Andrea Shome
Ben Walsh

Collaborators' Students

Murat Aytekin
Julian Jenkins
David Klein
Huan Luo

Collaborators

Catherine Carr
Alain de Cheveigné
Didier Depireux
Mounya Elhilali
Jonathan Fritz
Cindy Moss
David Poeppel
Shihab Shamma

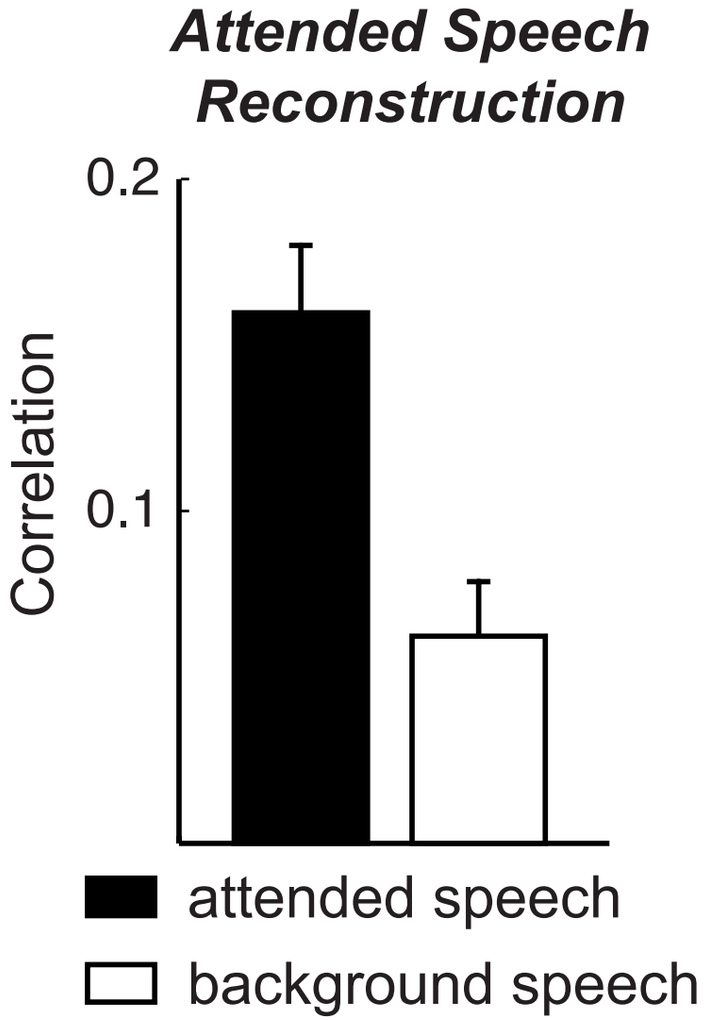
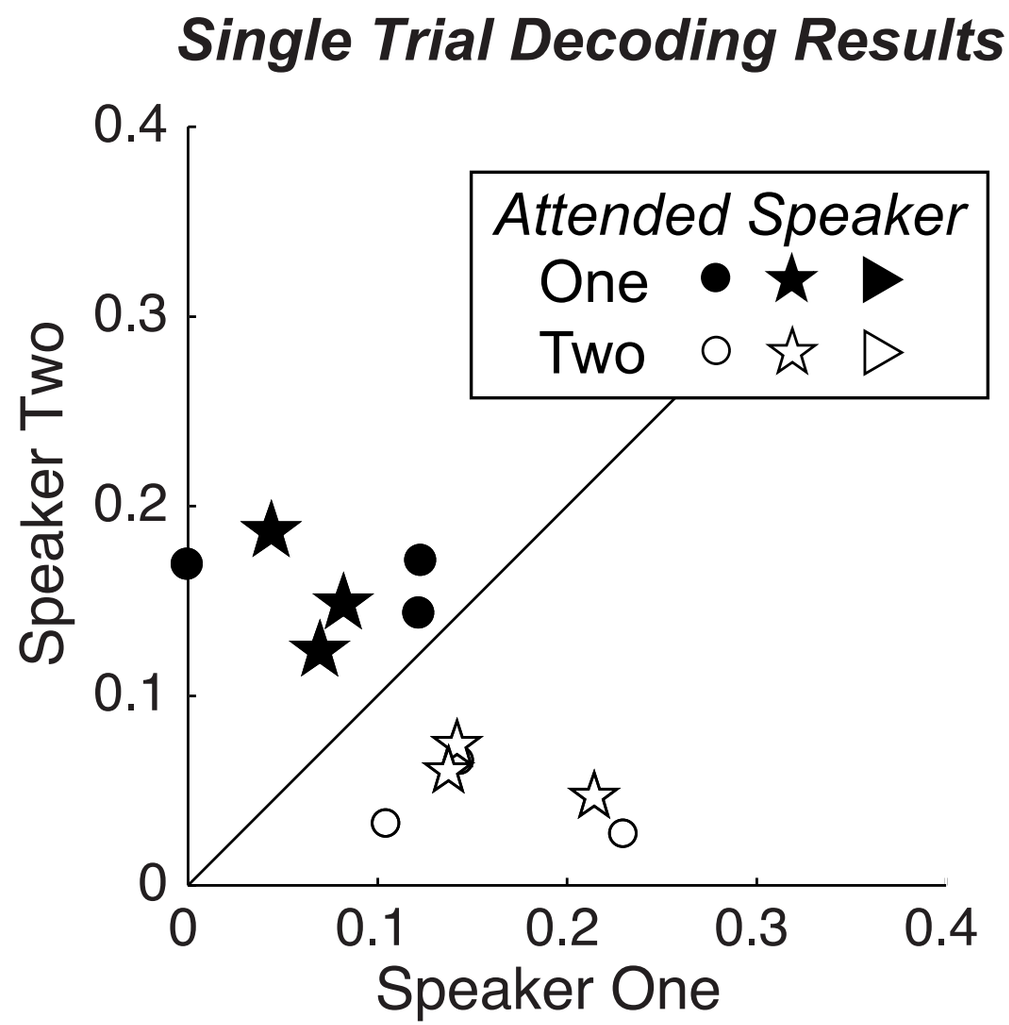
Past Postdocs

Dan Hertz
Yadong Wang

Funding

NIH R01 DC 008342

#373 Cortical Representations of Music in Human Listeners
#889 Cortical Encoding of Speech in Challenging Listening Environments
#961 Neural Entrainment to Speech, a Matter of Time or Frequency?

Thank You

Reconstruction of Same-Sex Speech