Neural Representations of the Cocktail Party in Human Auditory Cortex

Jonathan Z. Simon

Department of Biology
Department of Electrical & Computer Engineering
Institute for Systems Research

University of Maryland

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

MPI Leipzig, 5 June 2014

Acknowledgements

Grad Students

- Francisco Cervantes
- Alex Presacco
- Krishna Puvvada

Past Grad Students

- Nayef Ahmar
- Claudia Bonin
- Maria Chait
- Marisel Villafane Delgado
- Kim Drnec
- 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 Hogan
- Kevin Kahn
- Andrea Shome
- Madeleine Varmer
- Ben Walsh

Collaborators' Students

- Murat Aytekin
- Julian Jenkins
- David Klein
- Huan Luo

Past Postdocs

- Dan Hertz
- Yadong Wang

Collaborators

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

Funding

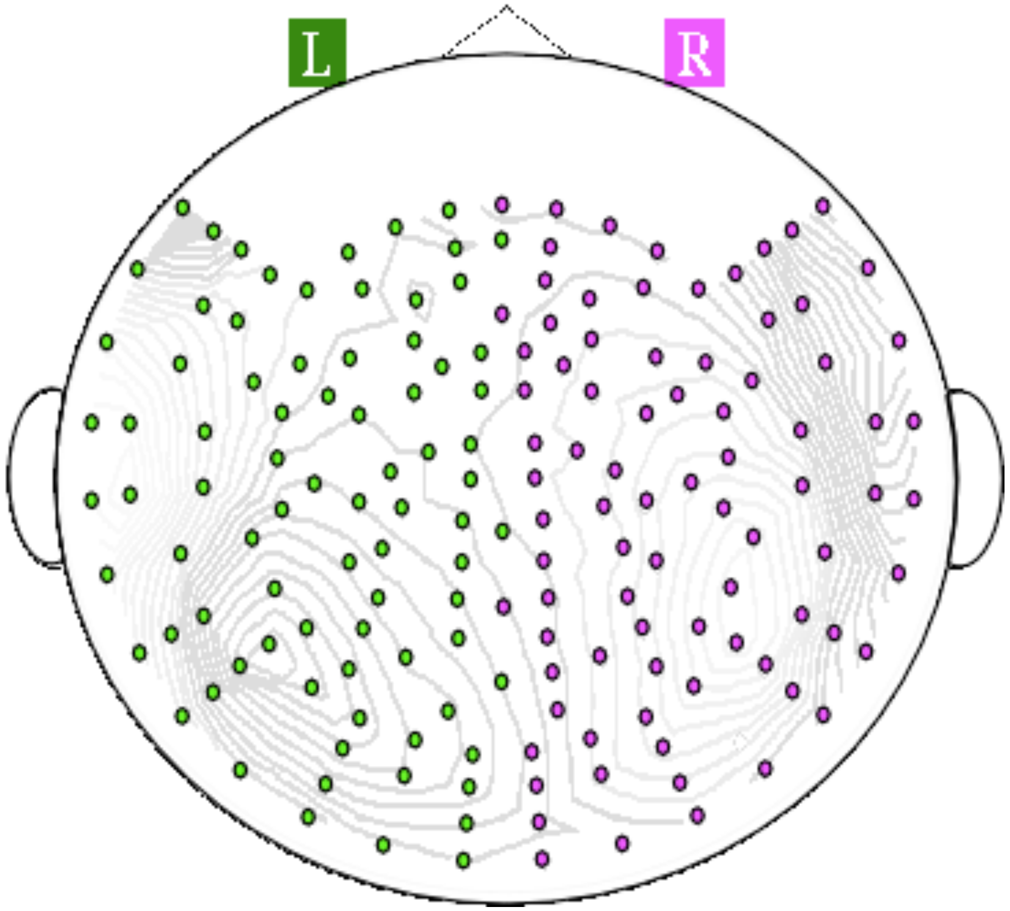
- NIH R01 DC 008342
- NIH R01 DC 007657
- NIH R01 DC 005660
- NIH R01 DC 000436
- NIH R01 AG 036424
- NIH R01 AG 027573
- NIH R01 EB 004750
- NIH R03 DC 004382
- USDA 20096512005791

Introduction

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

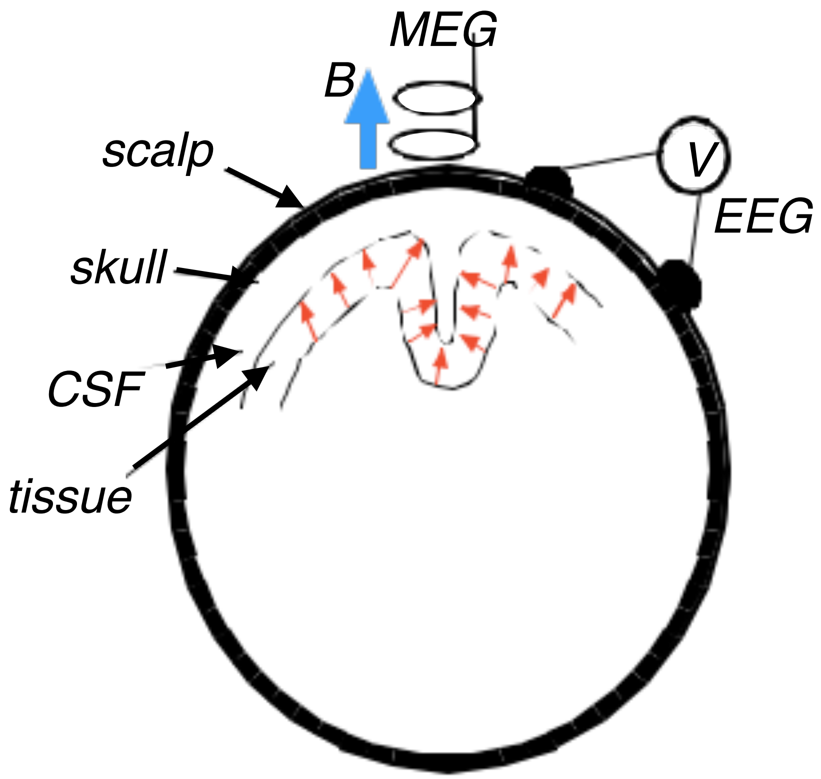
Magnetoencephalography

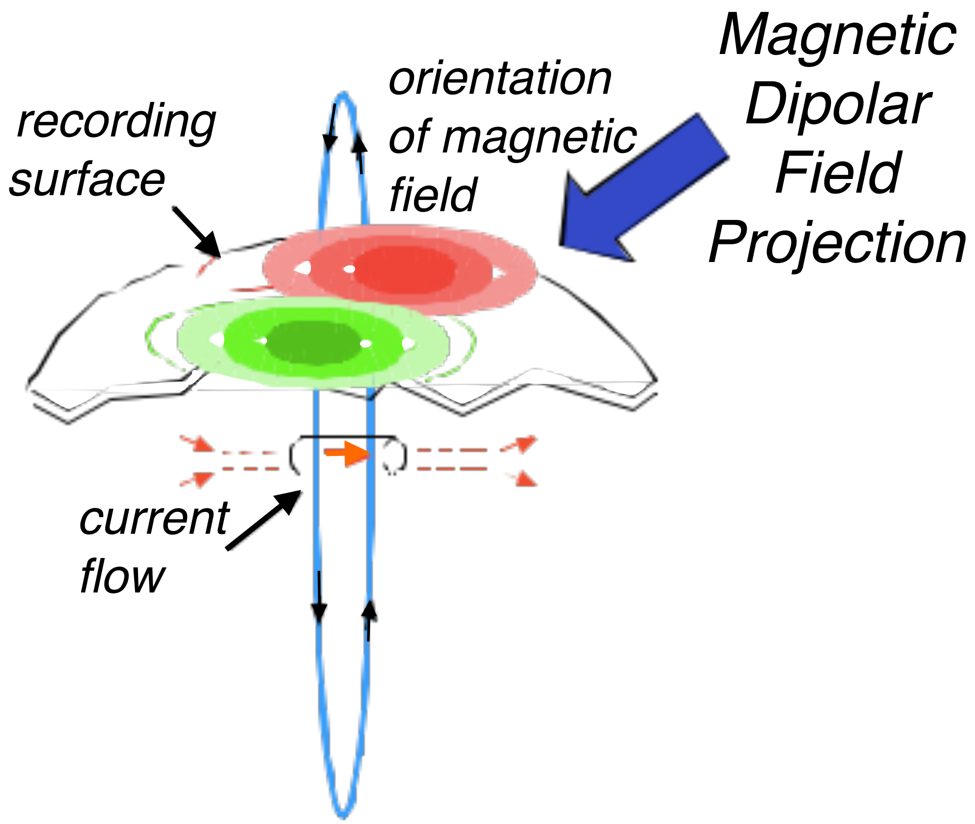
- Non-invasive, Passive, Silent Neural Recordings
- Simultaneous Whole-Head Recording (~200 sensors)
- Sensitivity
  - high: ~100 fT (10^{-13} Tesla)
  - low: ~10^4 – ~10^6 neurons
- Temporal Resolution: ~1 ms
- Spatial Resolution
  - coarse: ~1 cm
  - ambiguous



Neural Signals & MEG

Photo by Fritz Goro




recording surface
orientation of magnetic field
Magnetic Dipolar Field Projection
current flow

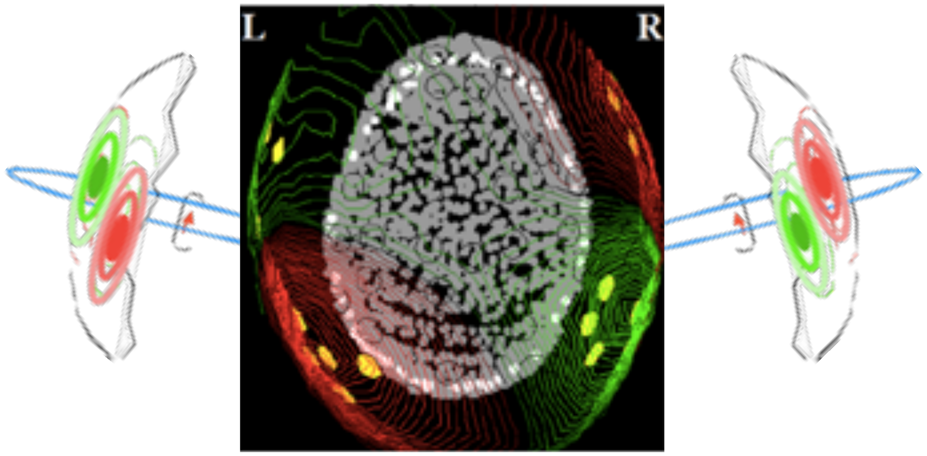
- Direct electrophysiological measurement
    - not hemodynamic
    - real-time
- No unique solution for distributed source

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

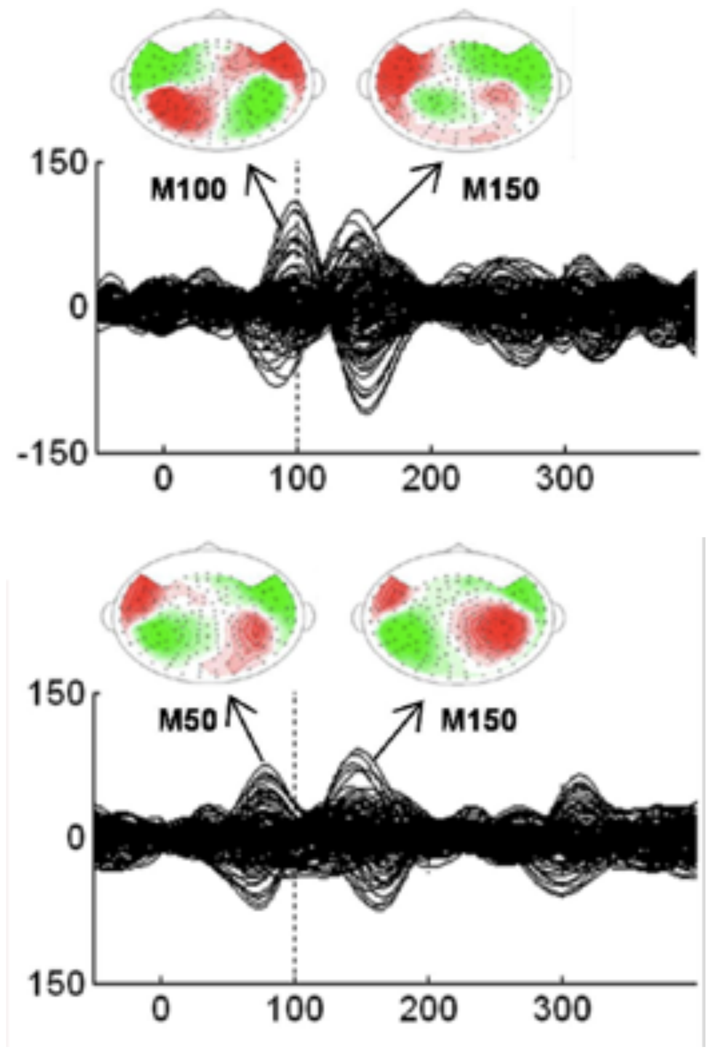
Time Course of MEG Responses

Auditory Evoked Responses

- MEG Response Patterns Time-Locked to Stimulus Events
- Robust
- Strongly Lateralized



Pure Tone



Broadband Noise

MEG Component Analysis

- Data driven spatial filtering:
  many available methods—ICA, PCA, DSS

- Generate spatial filters & their outputs ("components")

- DSS: Denoising Source Separation:
  Särelä & Valpola (2005)

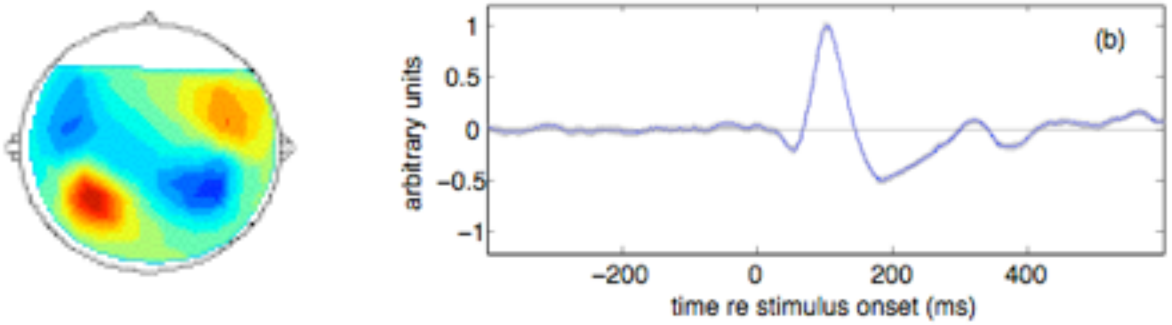
- DSS components ordered by reproducibility
  - 1st component "maximally reproducible" = most stimulus driven

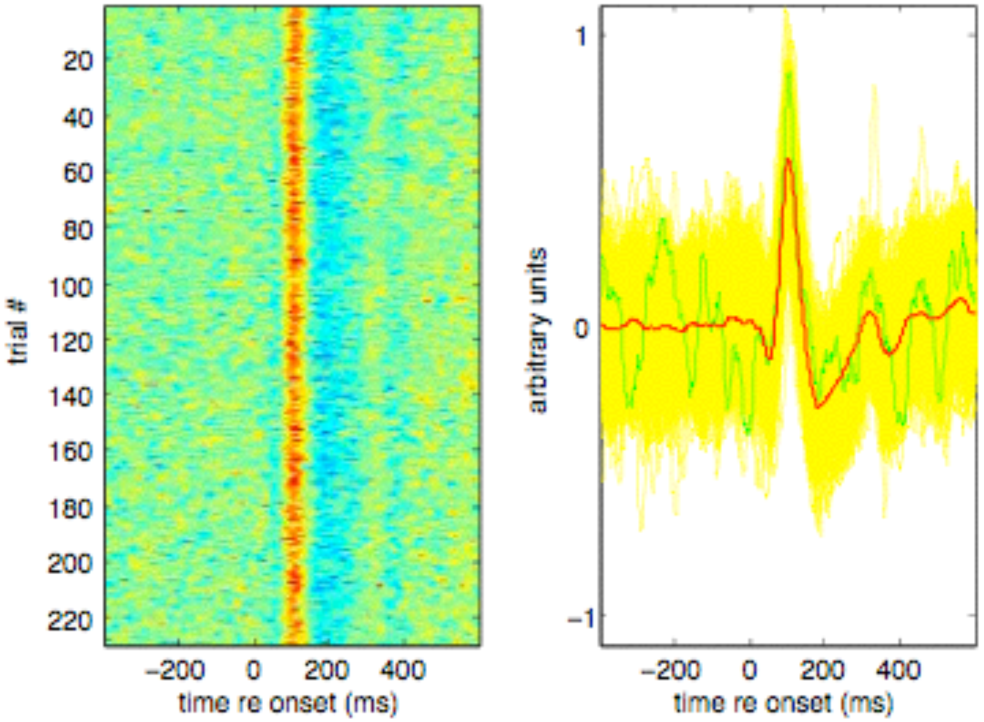
DSS Example

- Most reproducible filter & component

- Optimally filters out trial-to-trial-variable signal = neural noise

- Filter can be applied to other signals, e.g. single trials

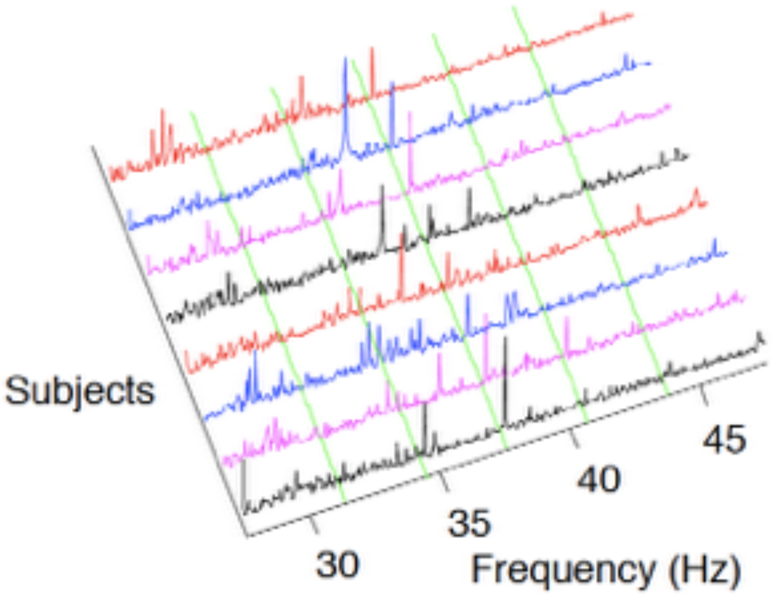




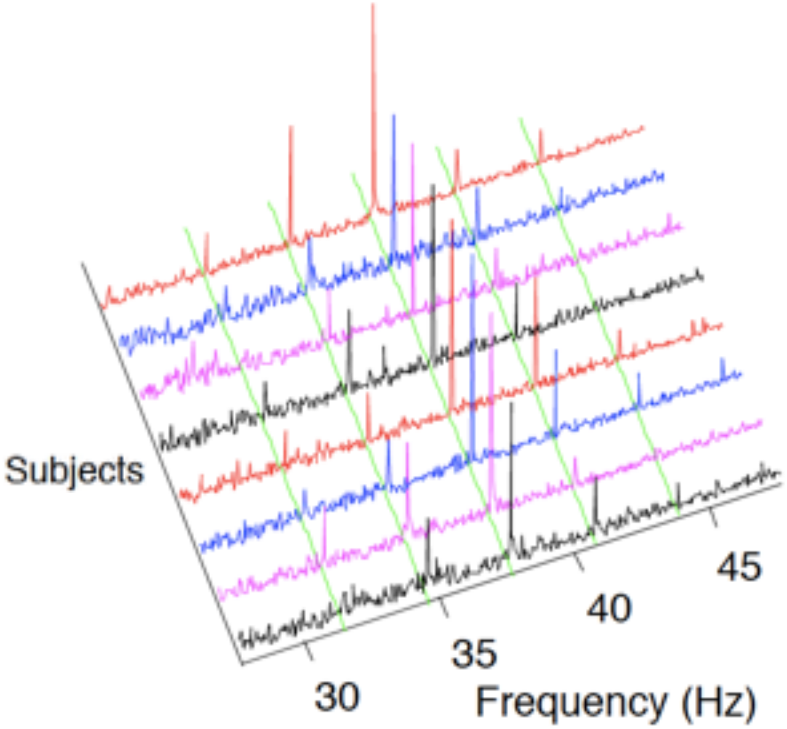
Särelä & Valpola (2005)
de Cheveigné & Simon, J. Neurosci. Methods (2008)

DSS Example: Spectral

Frequency Spectrum before DSS

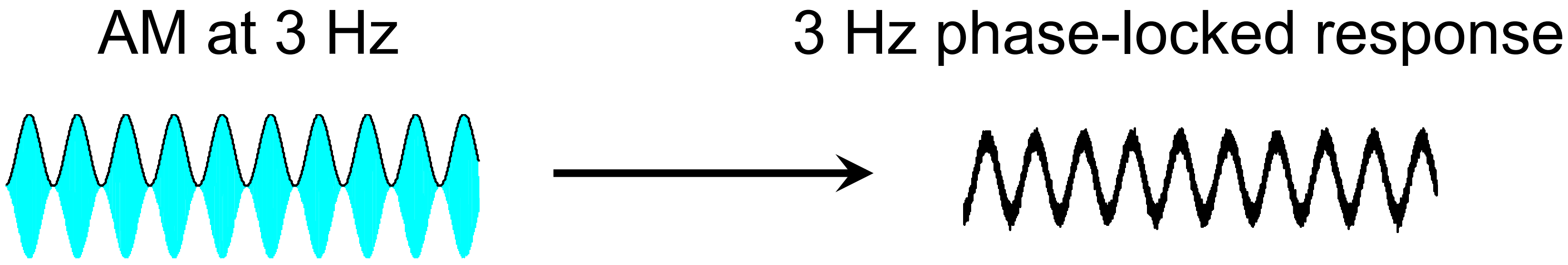


Frequency Spectrum after DSS



Ding & Simon, J. Neurophysiol (2009)

Phase-Locking in MEG to Slow Acoustic Modulations

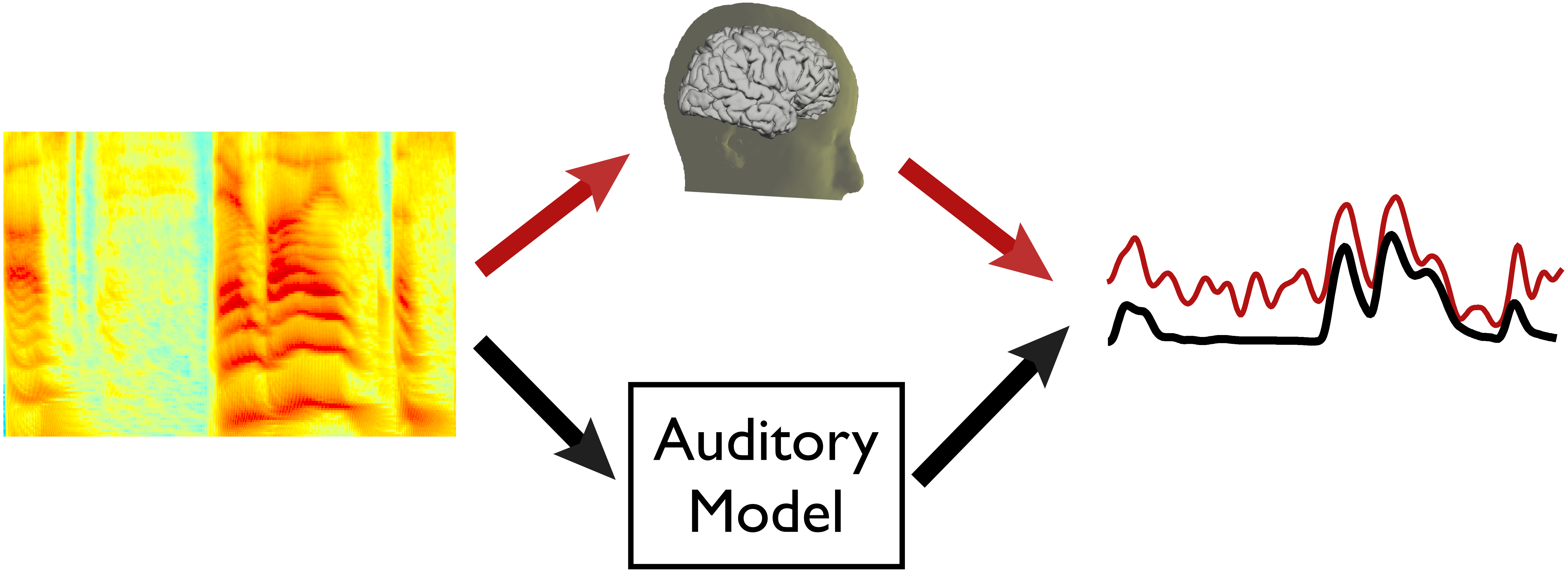


AM at 3 Hz

3 Hz phase-locked response

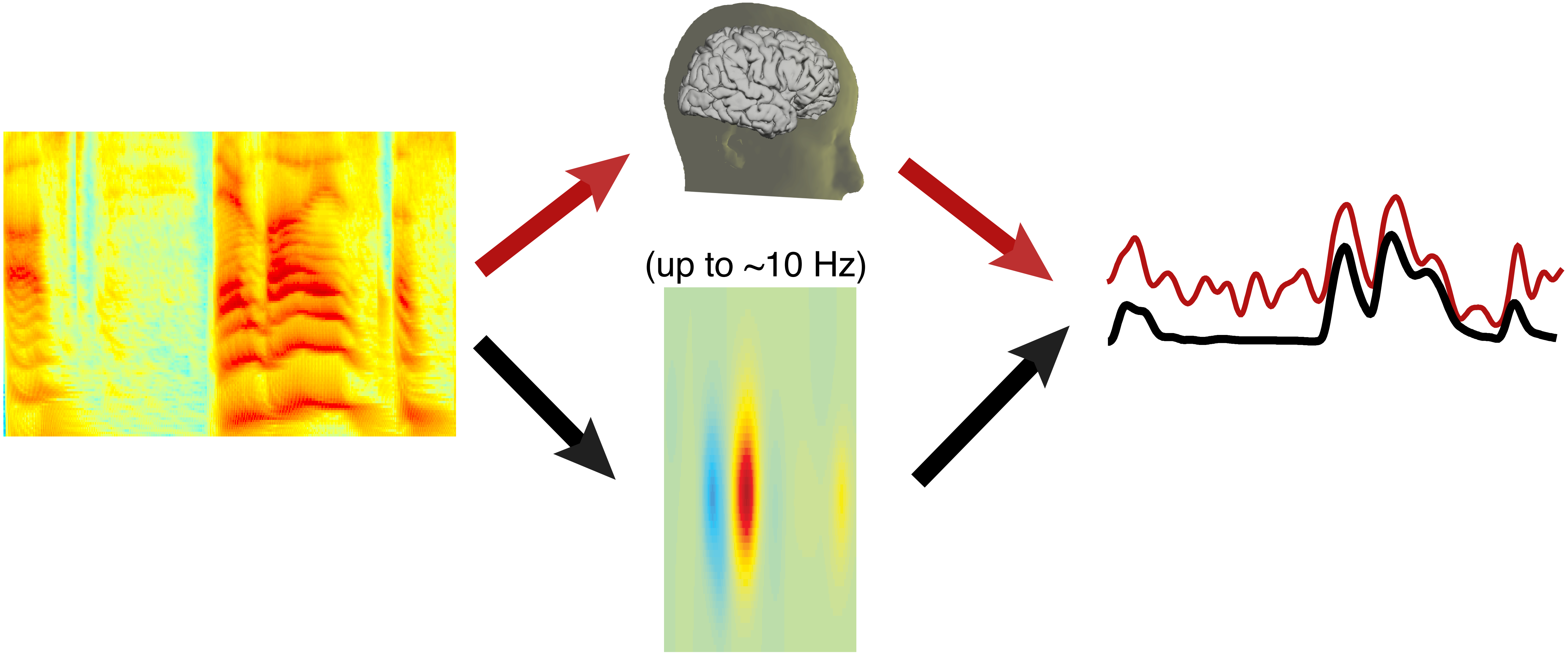
Ding & Simon, J Neurophysiol (2009)
Wang et al., J Neurophysiol (2012)

MEG Responses to Speech Modulations



Auditory Model

MEG Responses Predicted by STRF Model

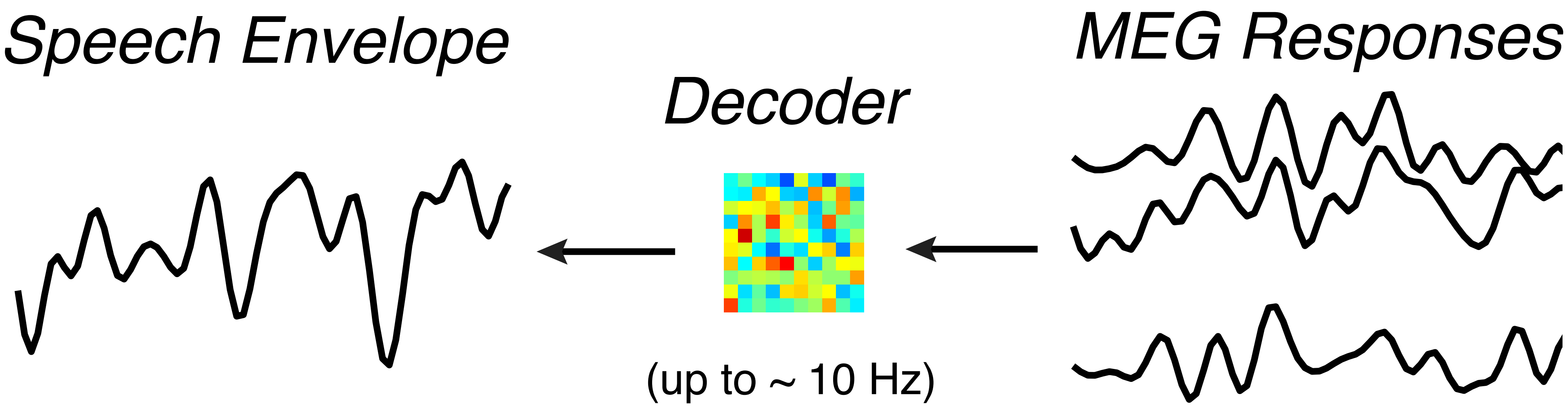


Linear Kernel = STRF

Ding & Simon, J Neurophysiol (2012)

“Spectro-Temporal Response Function”

Neural Reconstruction of Speech Envelope



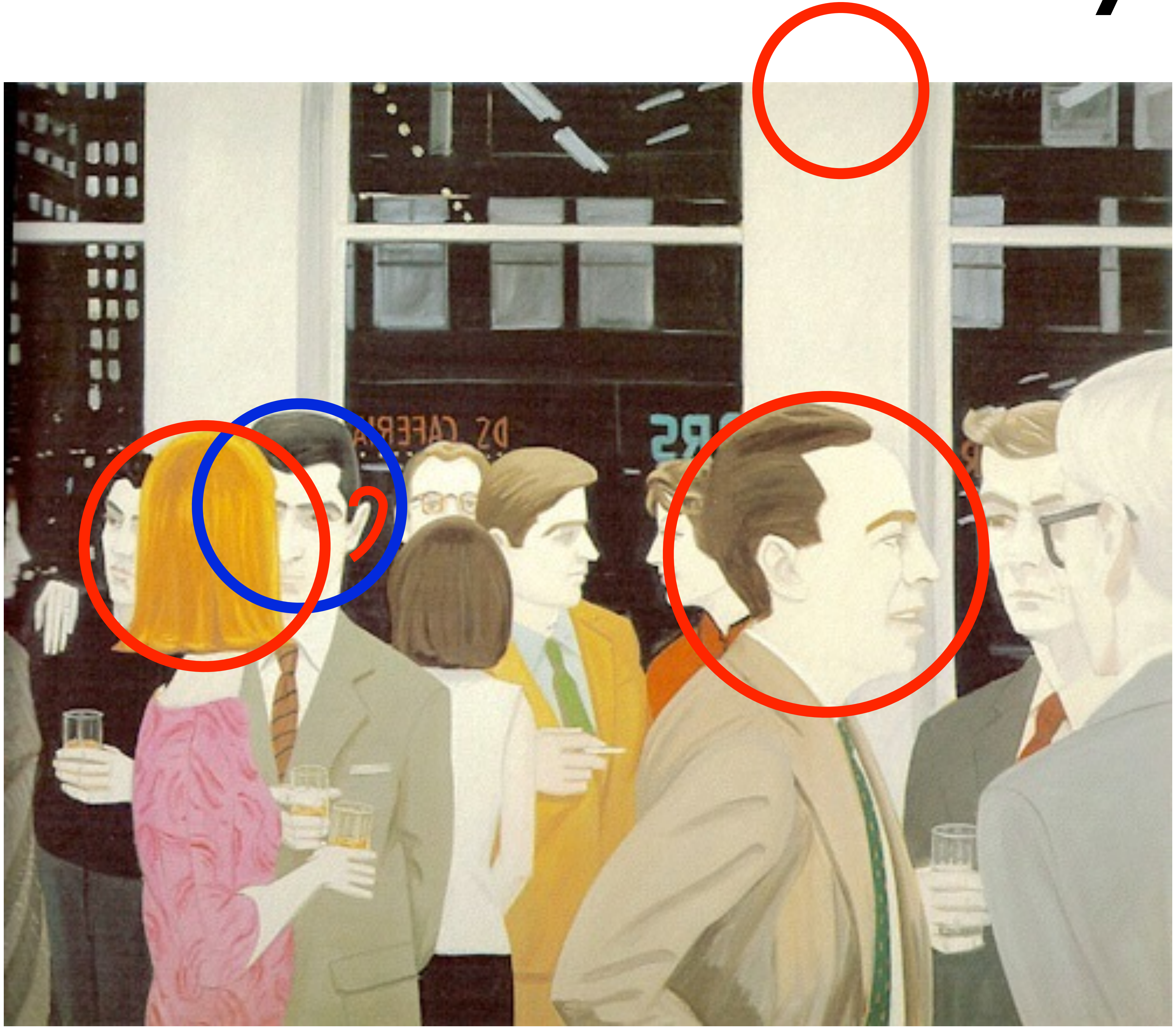
Auditory Objects

- What is an auditory object?
  - perceptual construct (not neural, not acoustic)
  - 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 at the Cocktail Party

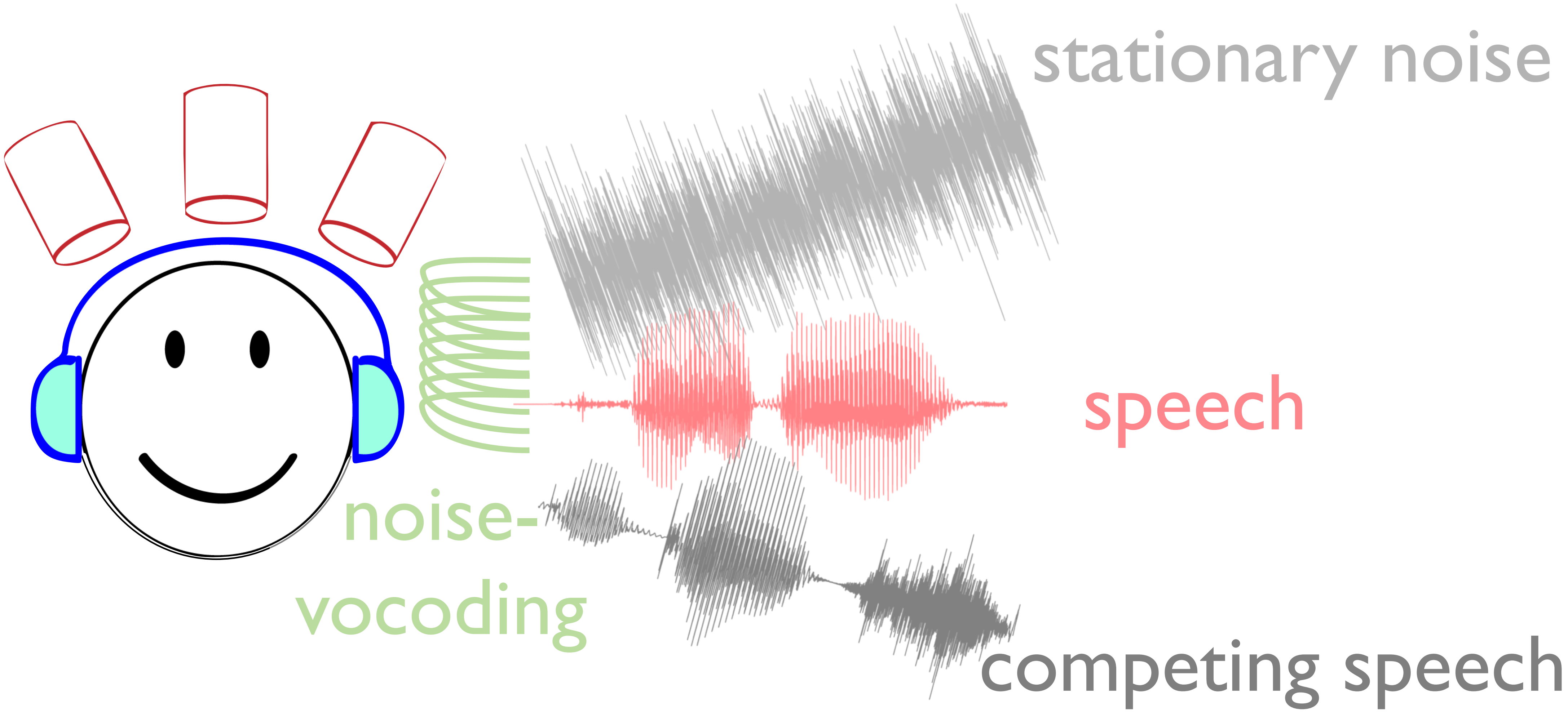


Alex Katz,
The Cocktail Party

Auditory Objects at the Cocktail Party

Alex Katz,
The Cocktail Party

Experiments



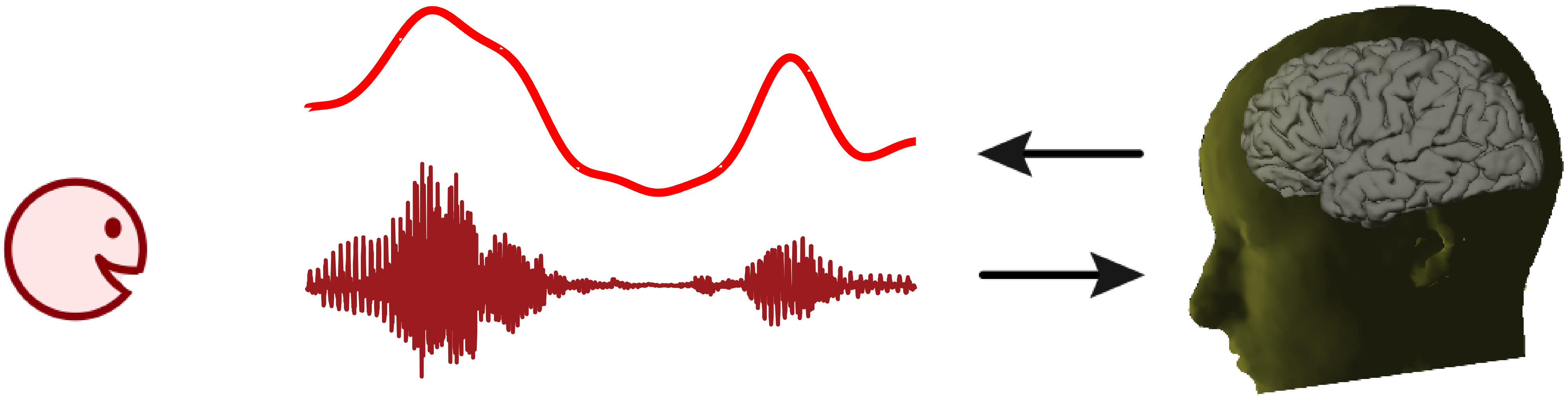
Speech Stream as an Auditory Object

- 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

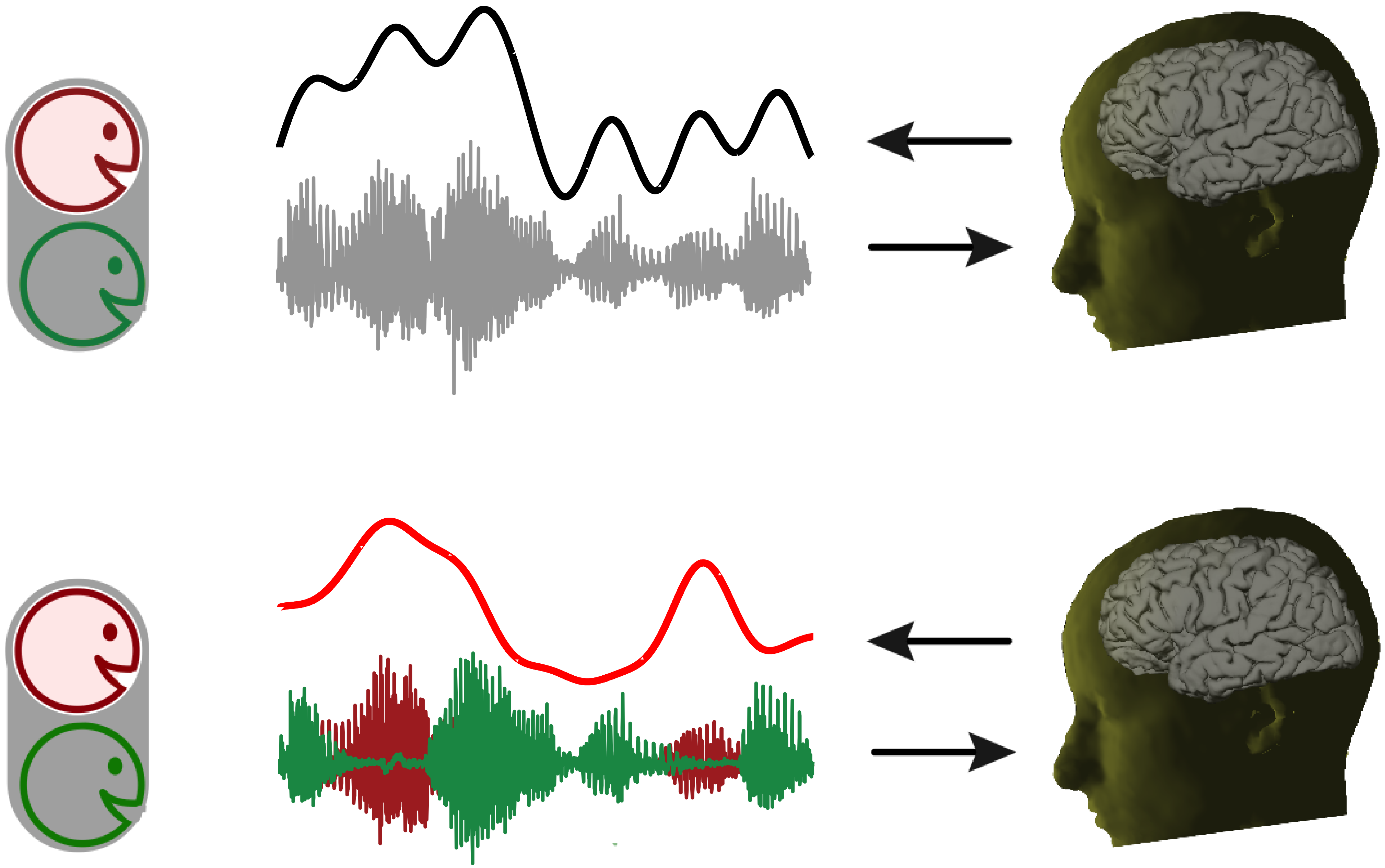
Neural Representation of an Auditory Object

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

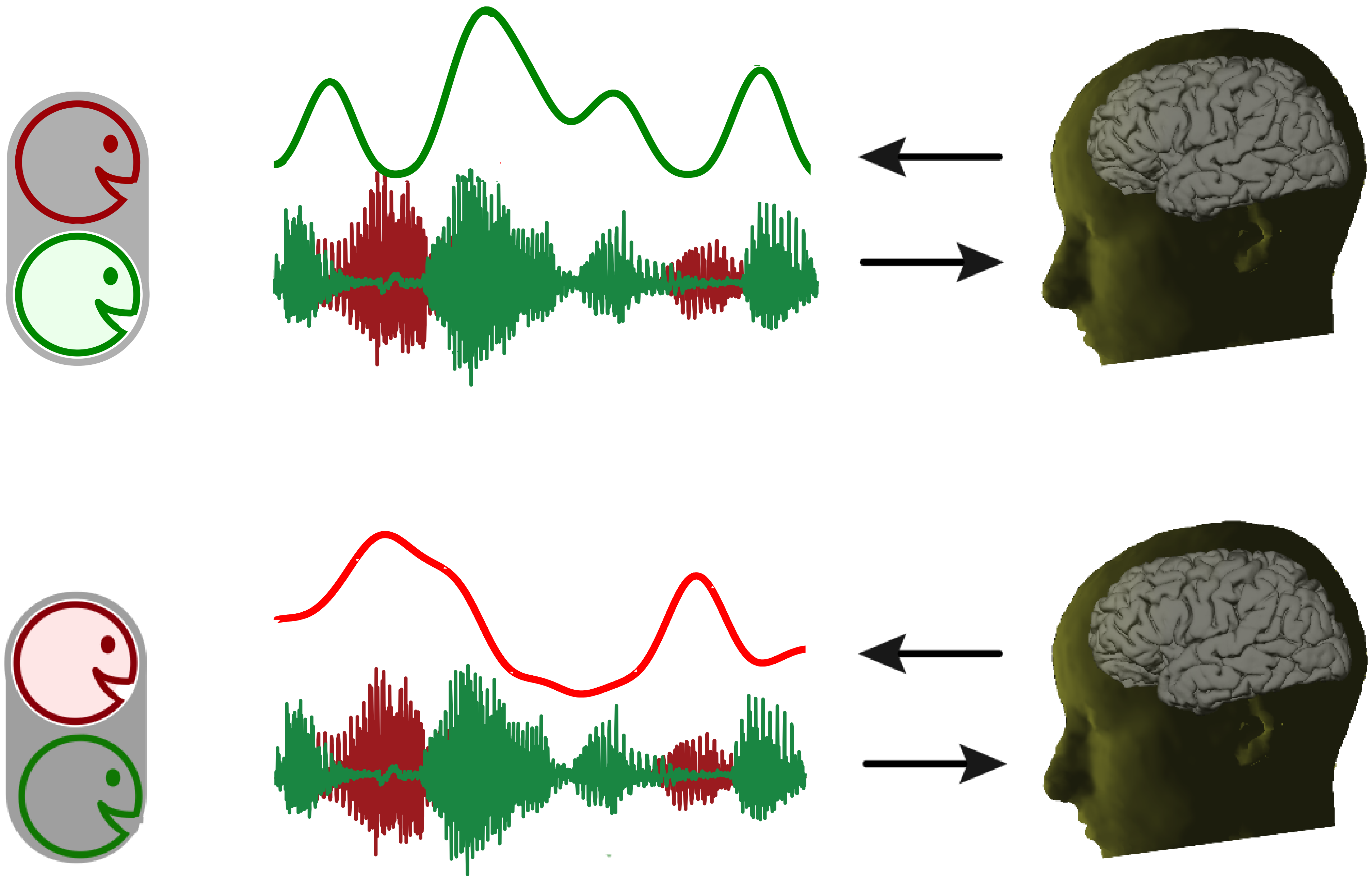
Selective Neural Encoding



Unselective vs. Selective Neural Encoding



Selective Neural Encoding

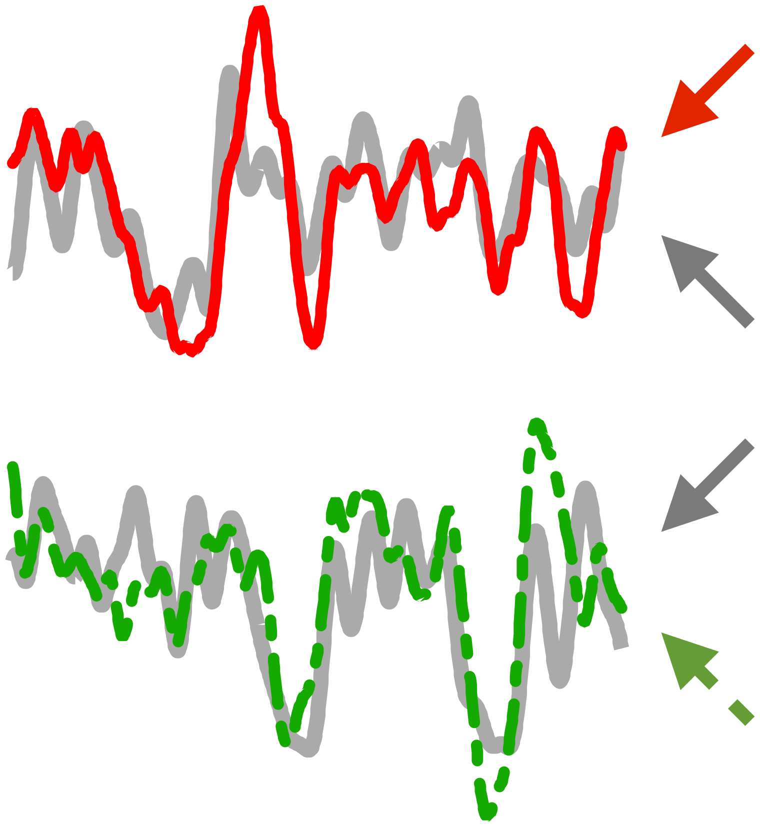


Stream-Specific Representation

grand average over subjects

attending to speaker 1

attending to speaker 2



reconstructed from MEG

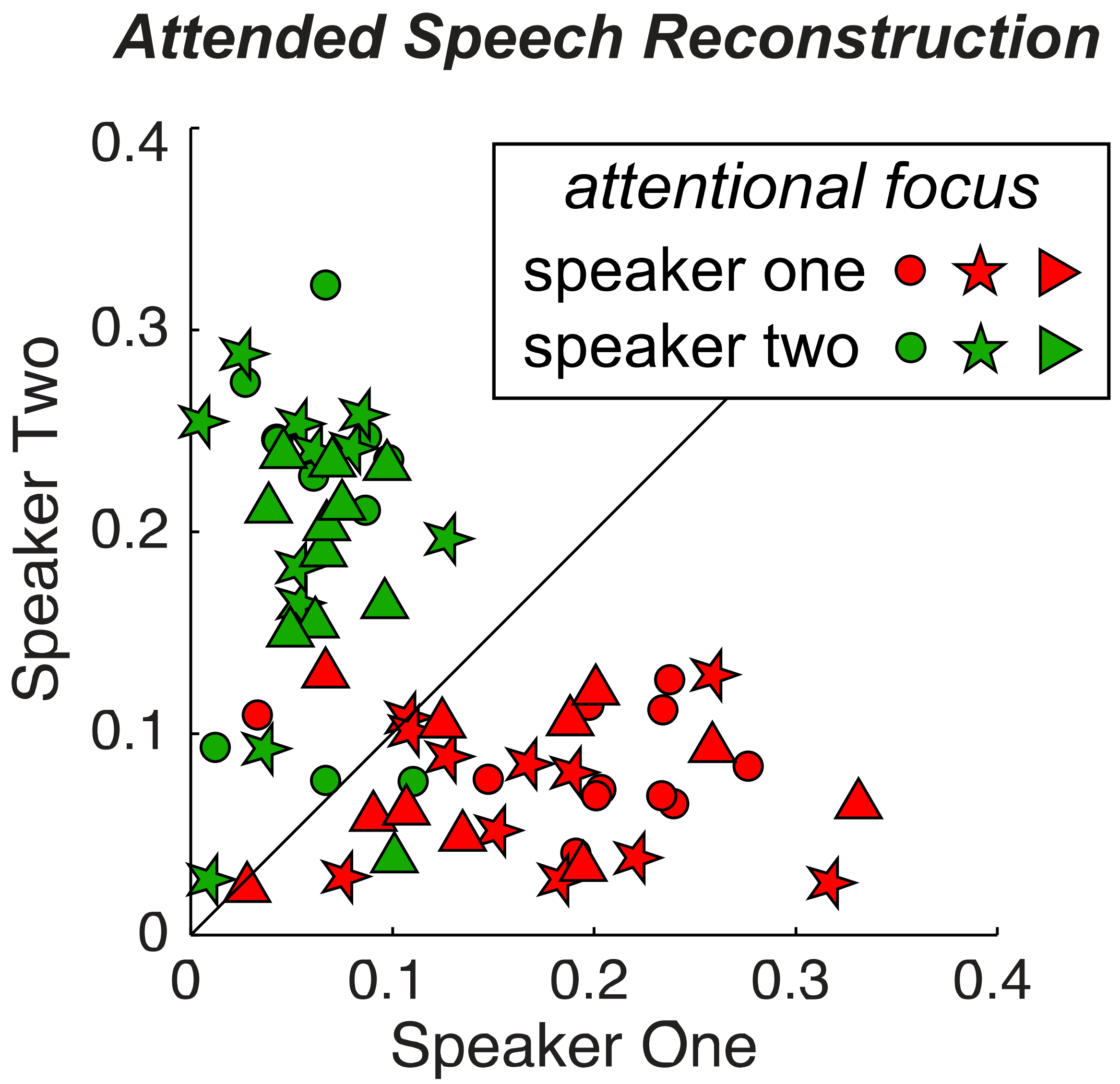
attended speech envelopes

reconstructed from MEG

Identical Stimuli!

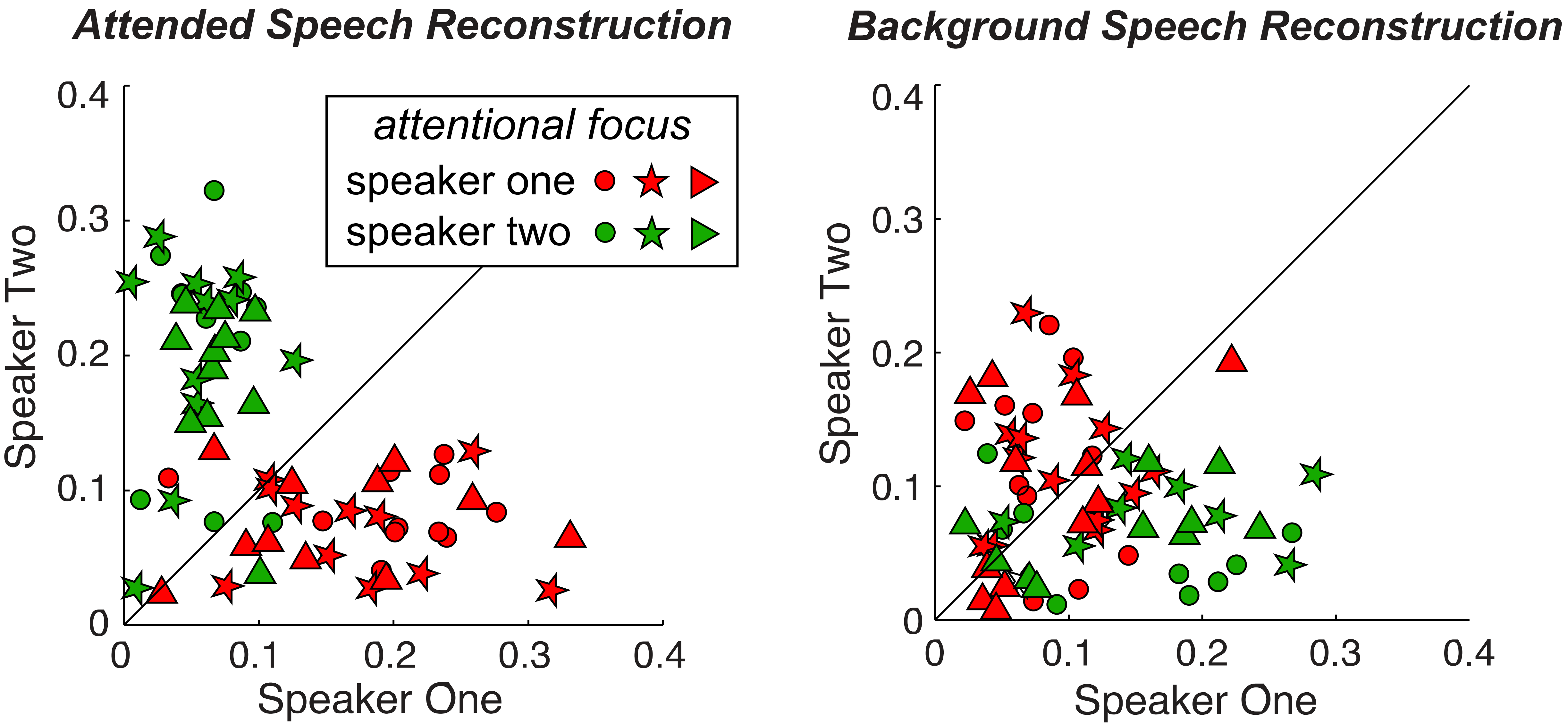
Ding & Simon, PNAS (2013)

Single Trial Speech Reconstruction



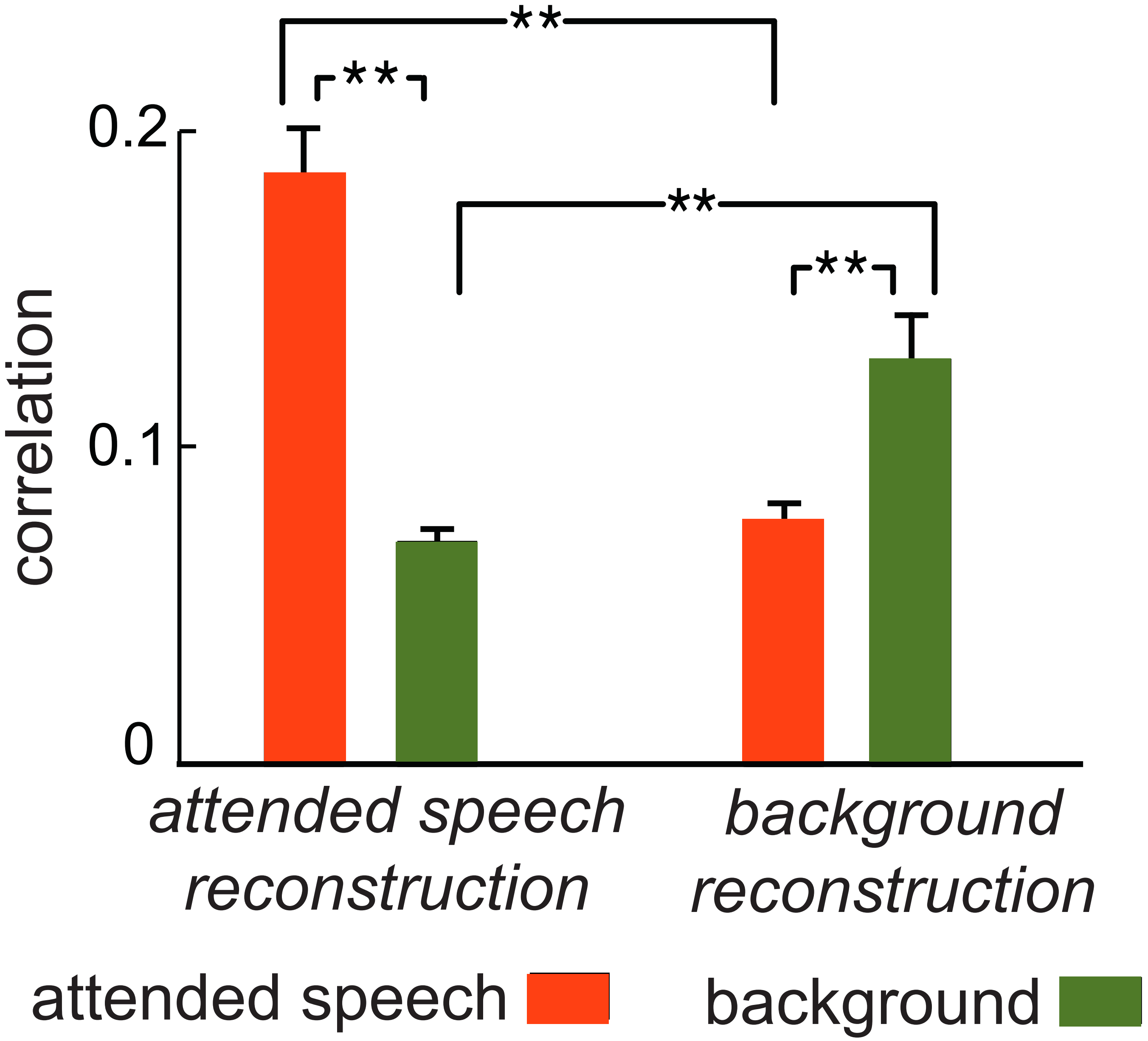
Ding & Simon, PNAS (2013)

Single Trial Speech Reconstruction



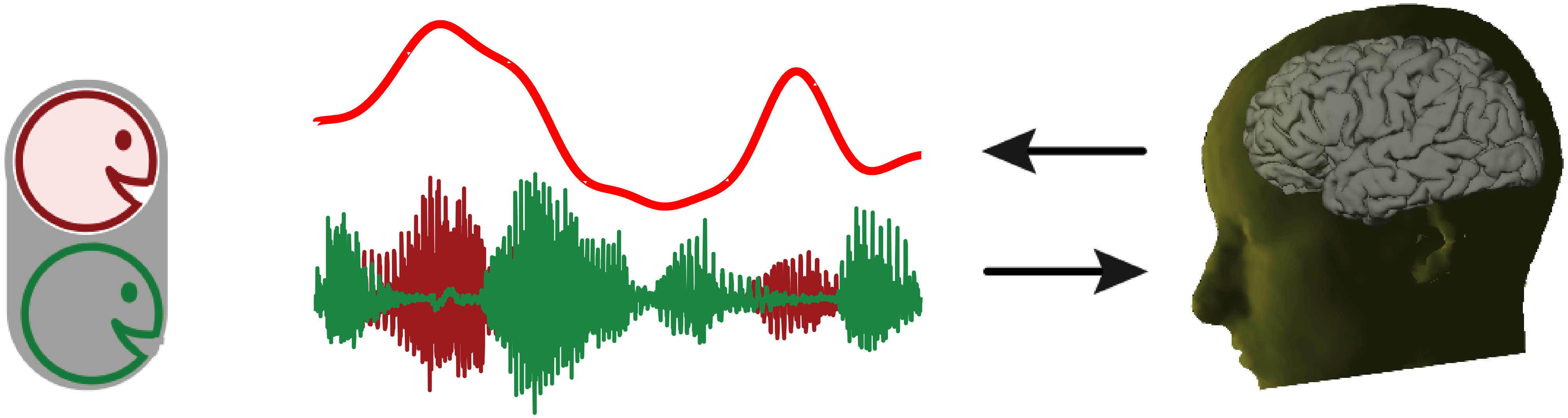
Ding & Simon, PNAS (2013)

Overall Speech Reconstruction

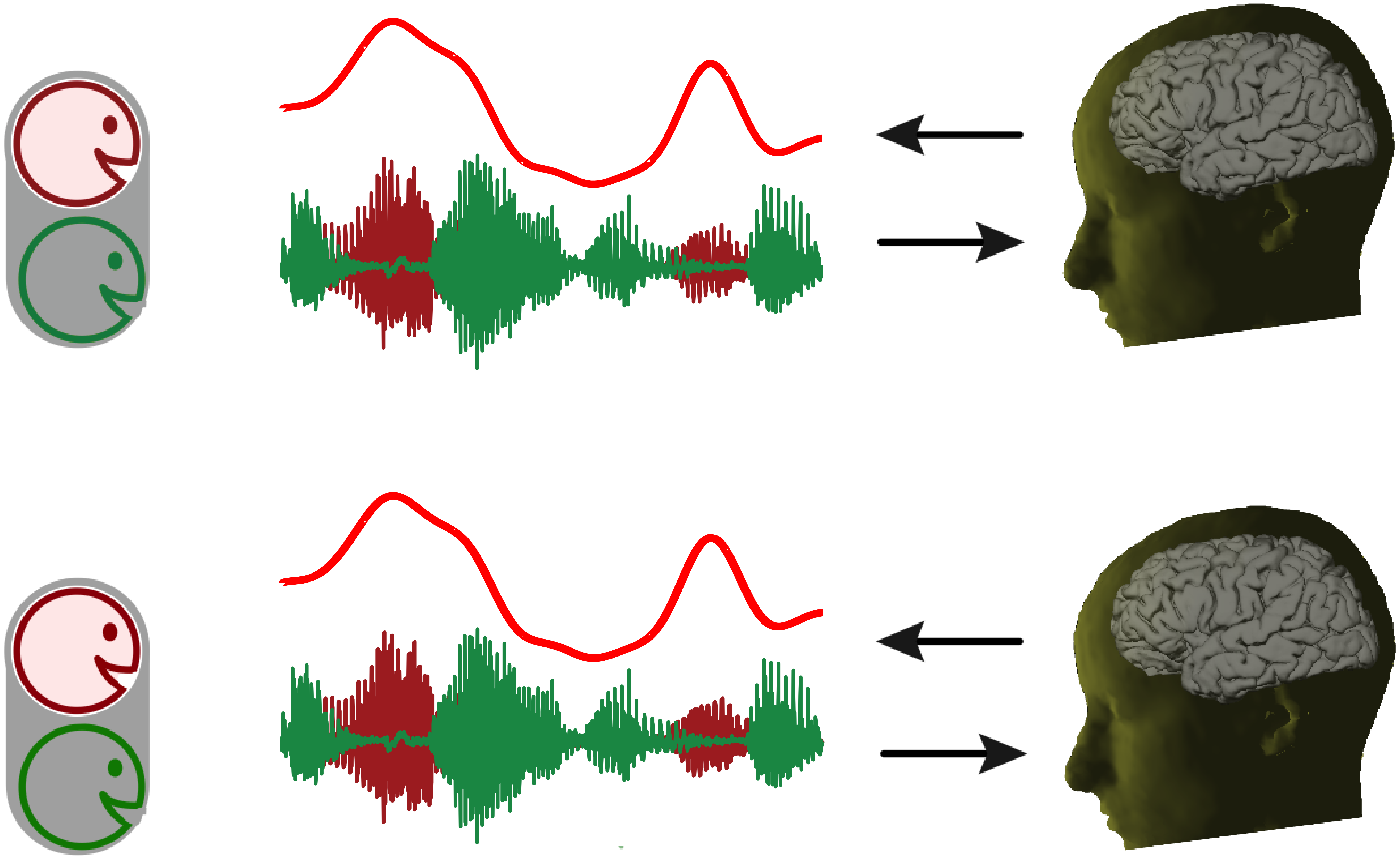


Distinct neural representations for different speech streams

Invariance Under Acoustic Changes

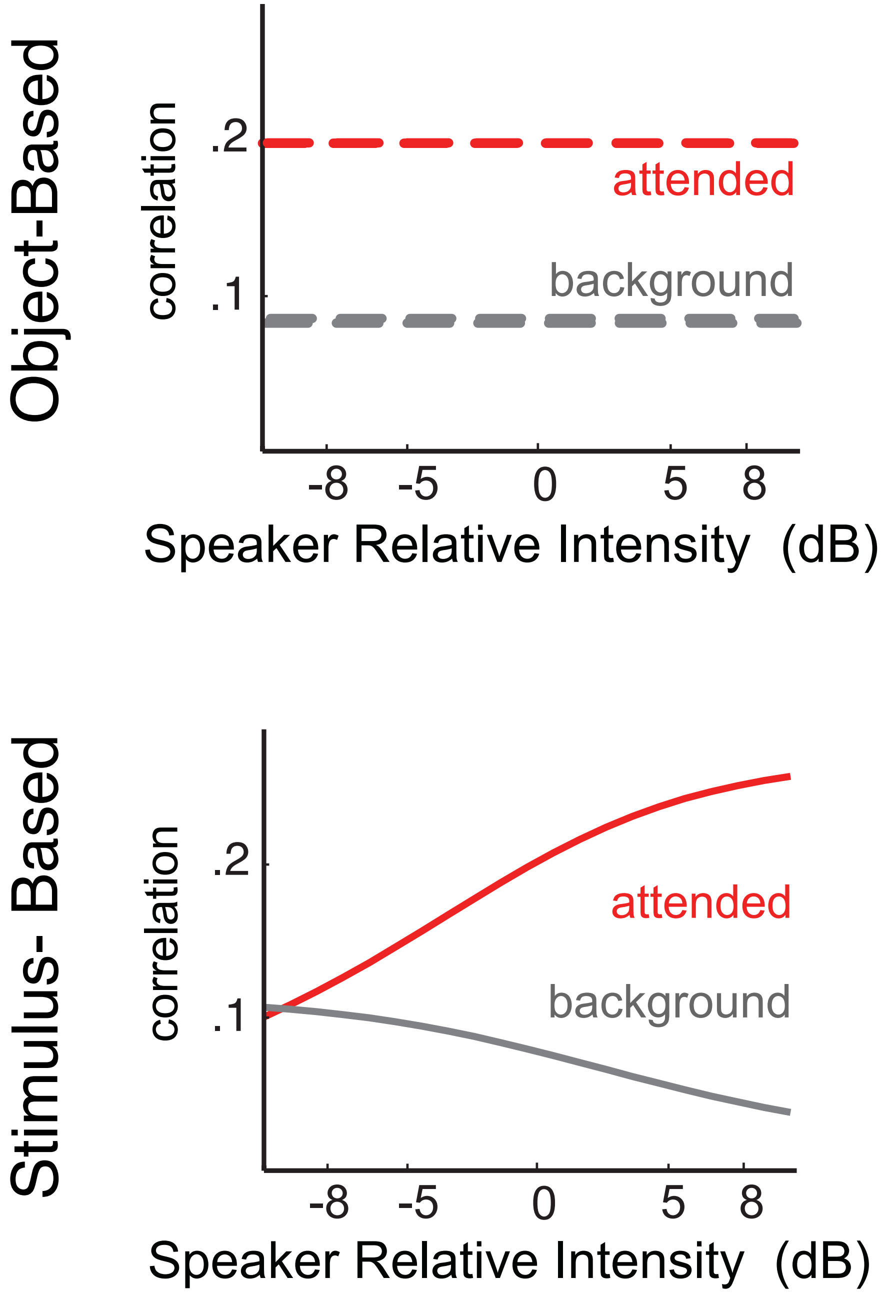


Invariance Under Acoustic Changes

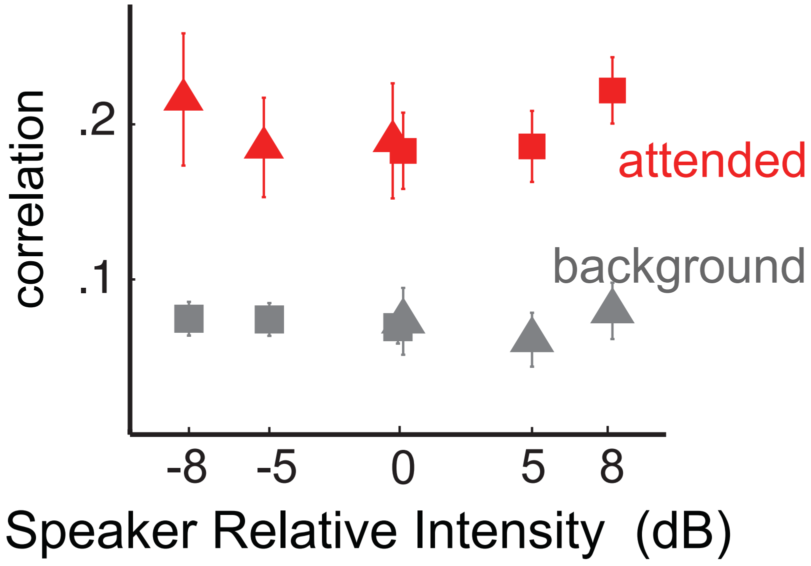


Stream-Based Gain Control?

Gain-Control Models



Neural Results

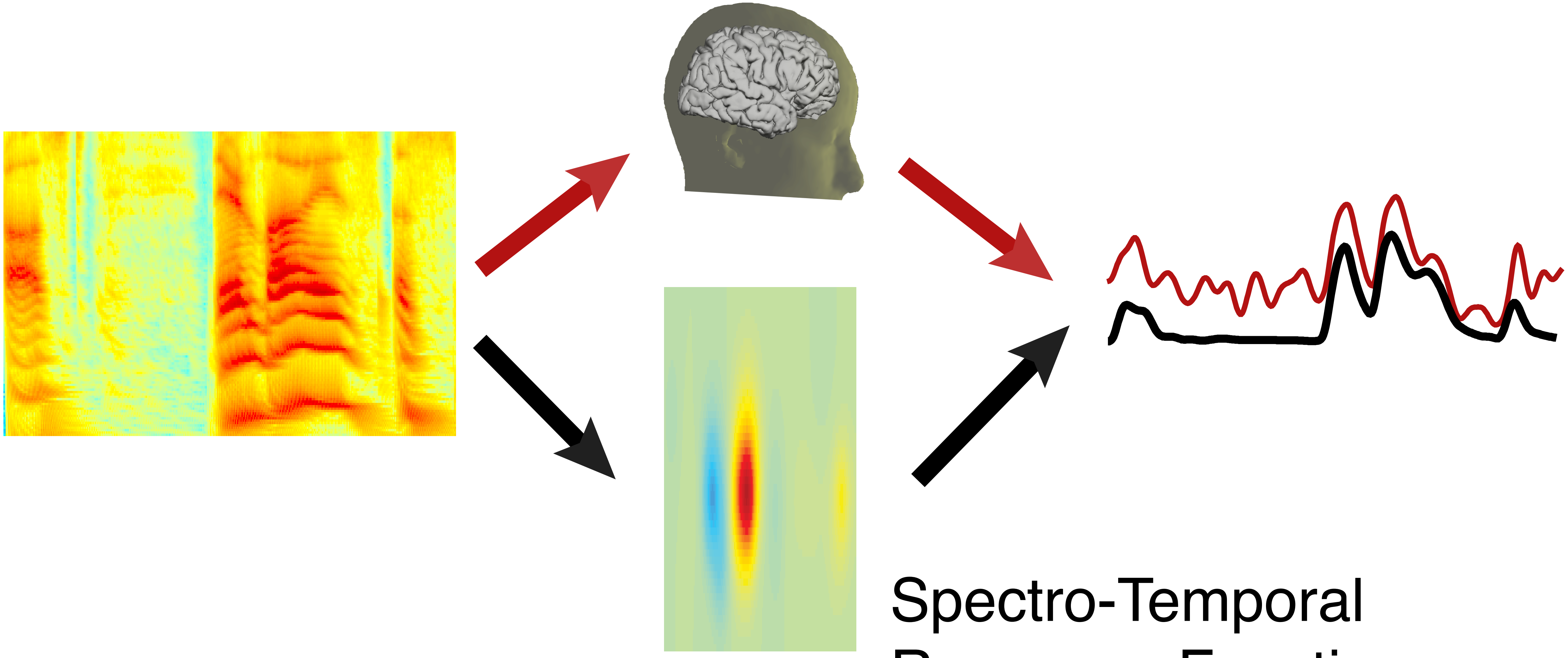


- Stream-based not stimulus-based
- Neural representation is invariant to acoustic changes.

Neural Representation of an Auditory Object

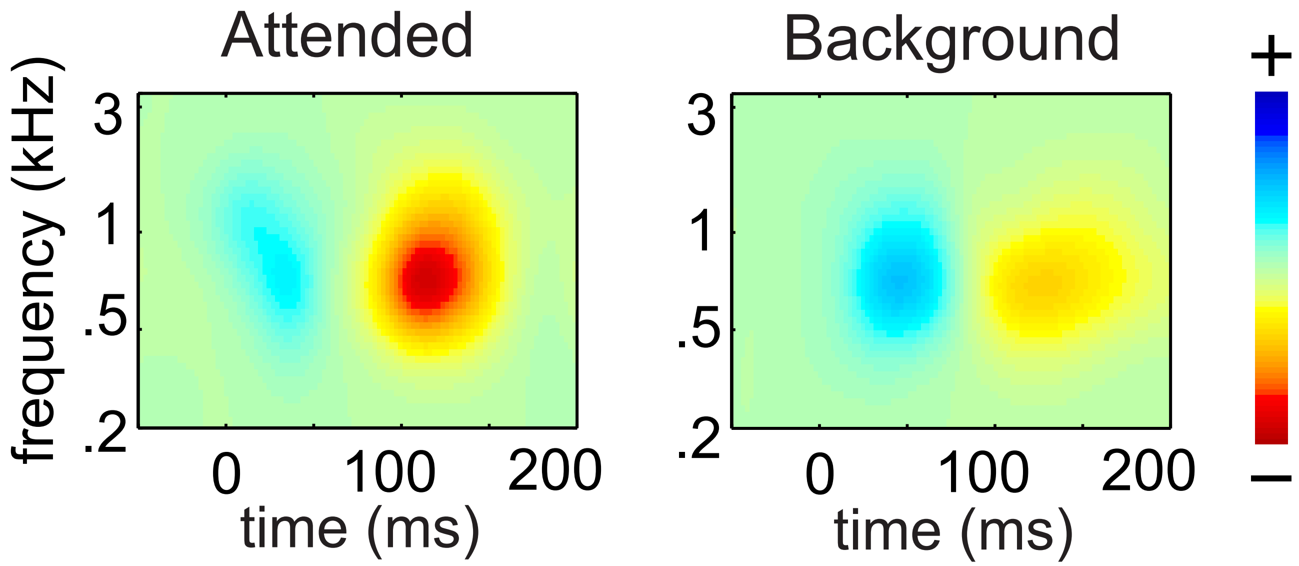
- ✓ neural representation is of something in sensory world
- ✓ when other sounds mixed in,
  neural representation is of auditory object,
  not entire acoustic scene
- ✓ neural representation invariant
  under broad changes in specific acoustics

Forward STRF Model



Spectro-Temporal
Response Function
(STRF)

STRF Results

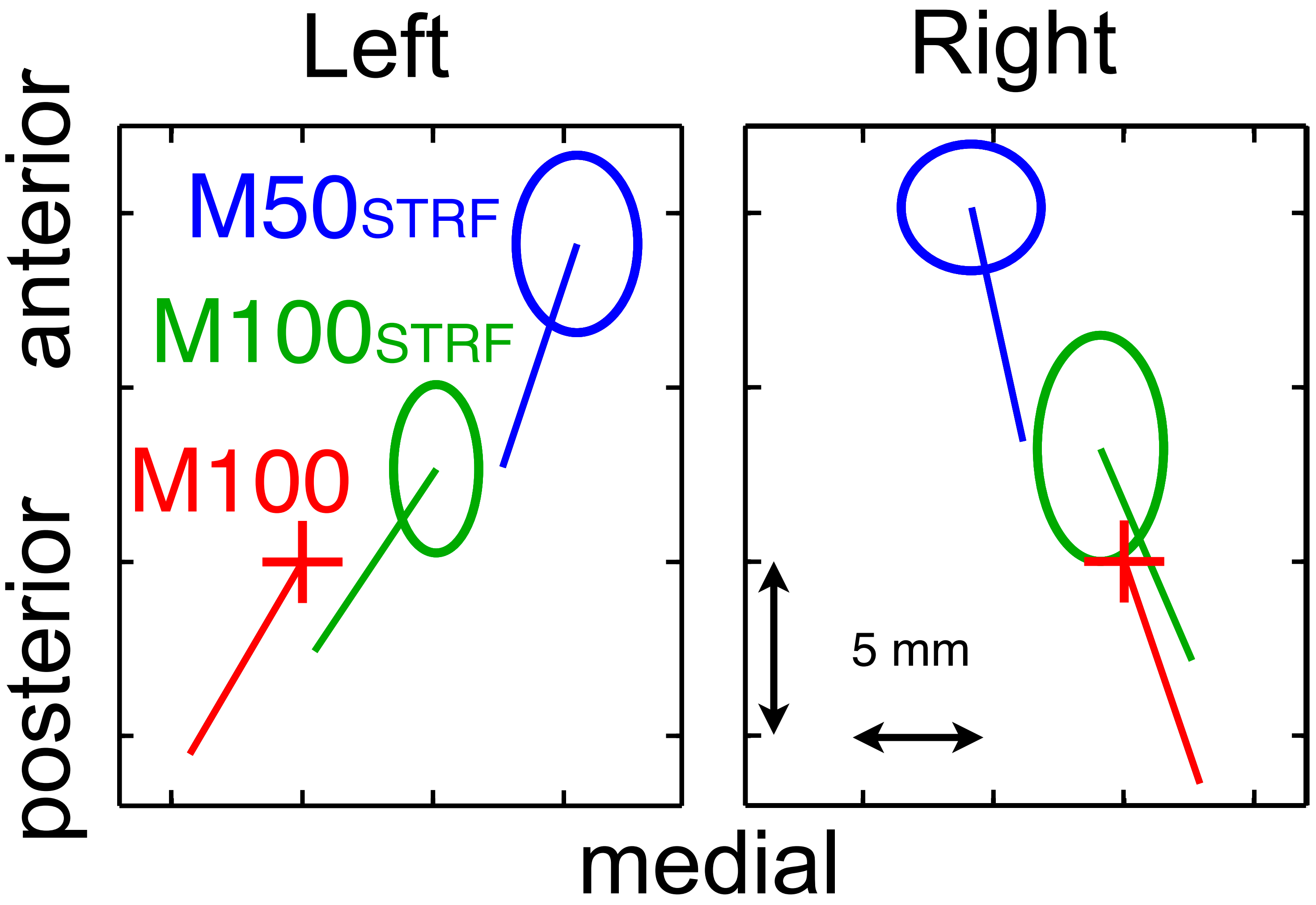


- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- M50$_{\text{STRF}}$ positive peak
- M100$_{\text{STRF}}$ negative peak

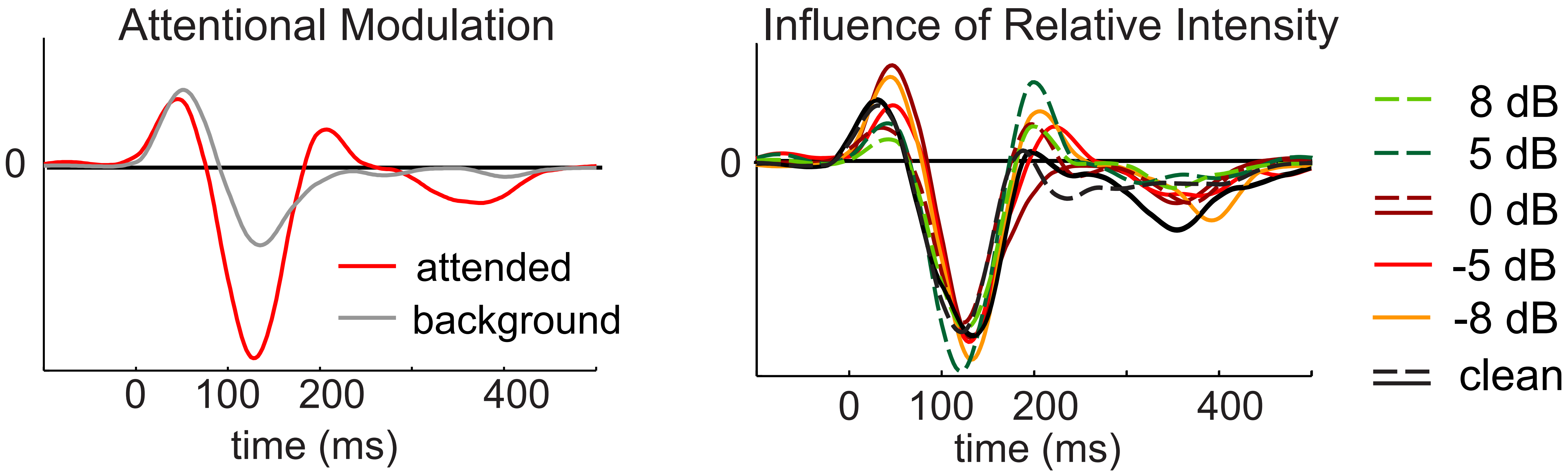
Neural Sources

- M100STRF source near (same as?) M100 source:
  Planum Temporale

- M50STRF source is anterior and medial to M100 (same as M50?):
  Heschl's Gyrus

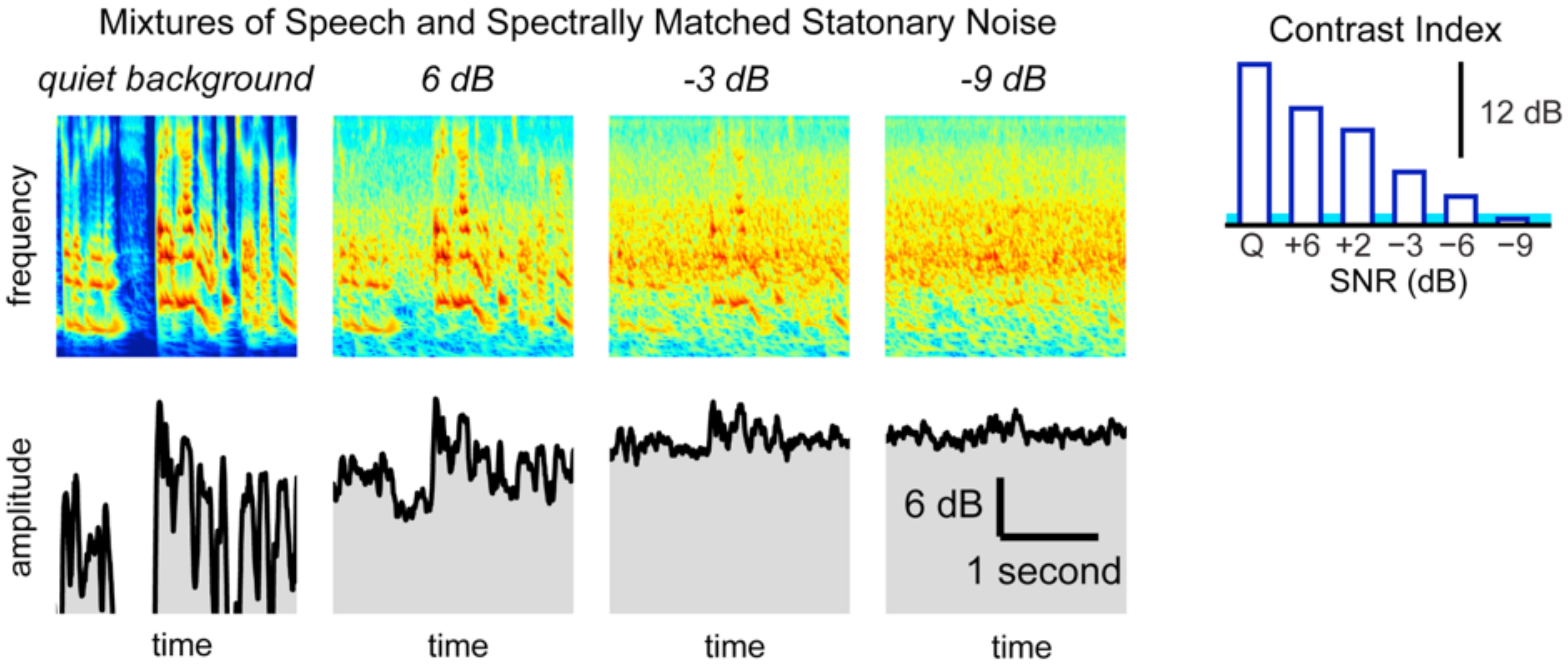


Cortical Object-Processing Hierarchy



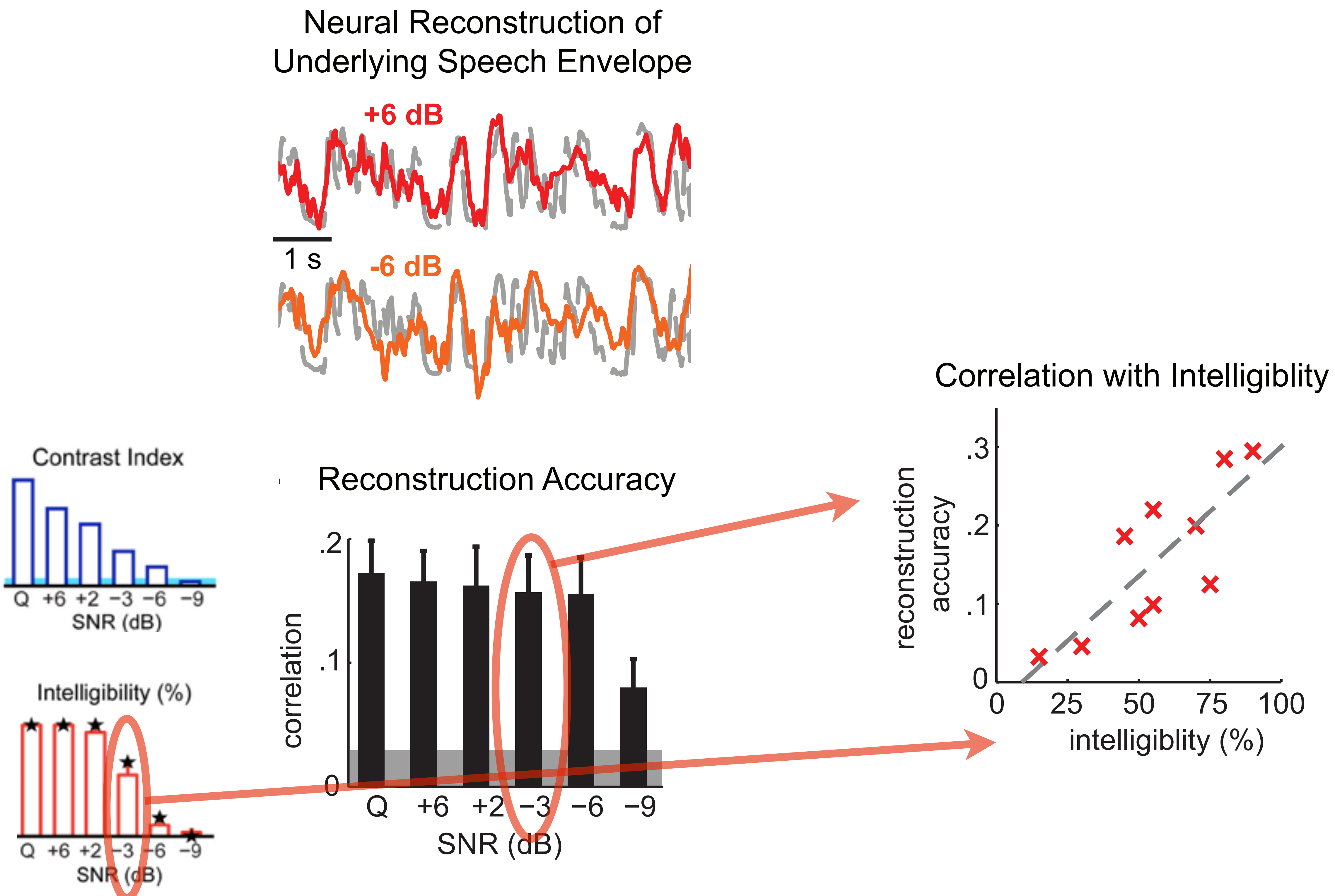
- M100$_{\text{STRF}}$ strongly modulated by attention, but not M50$_{\text{STRF}}$.
- M100$_{\text{STRF}}$ invariant against acoustic changes.
- Objects well-neurally represented at 100 ms, but not 50 ms.

Speech in Noise



Ding & Simon, J Neuroscience (2013)

Speech in Noise: Results



Neural Reconstruction of Underlying Speech Envelope

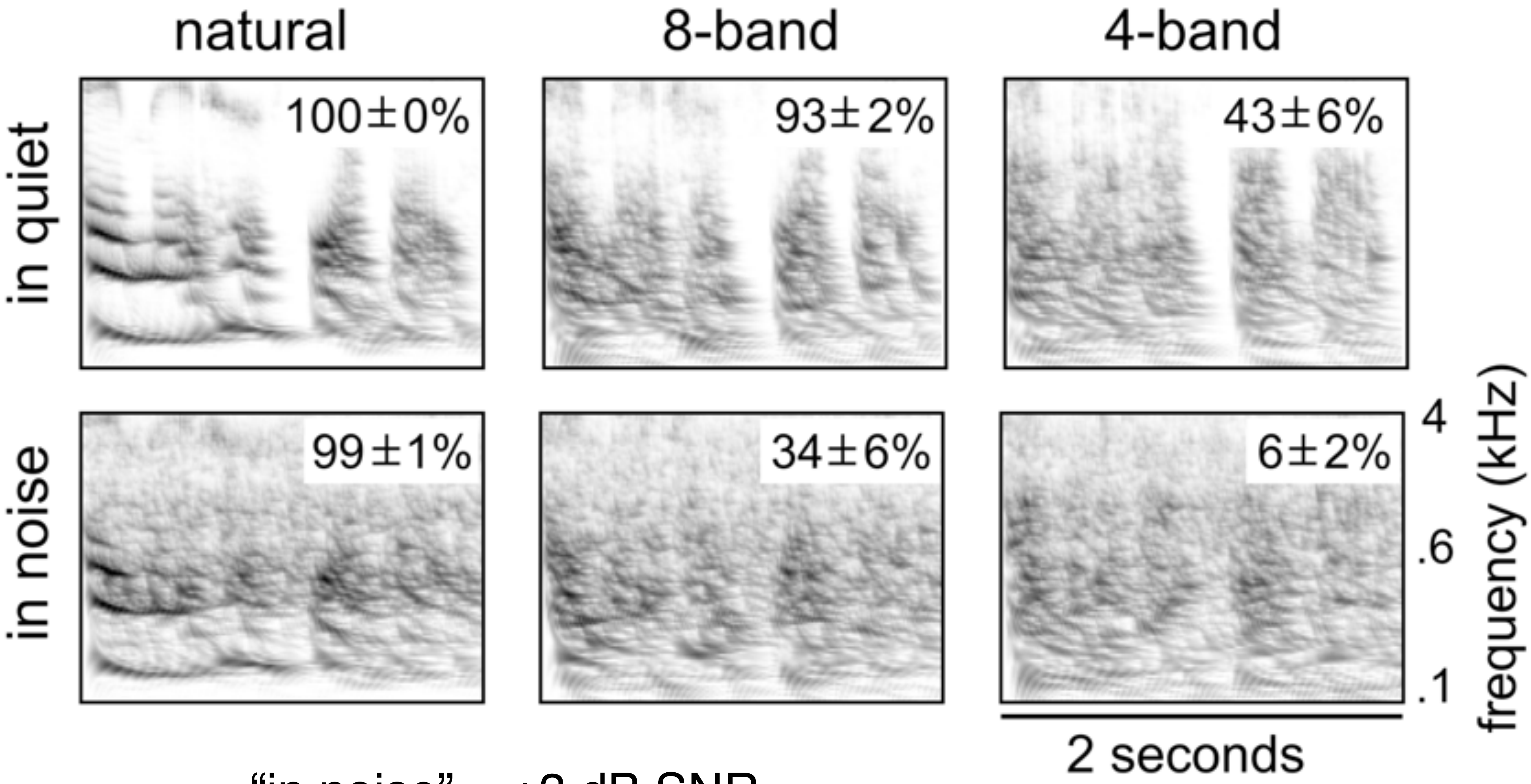
Contrast Index

Intelligibility (%)

Reconstruction Accuracy

Correlation with Intelligiblity

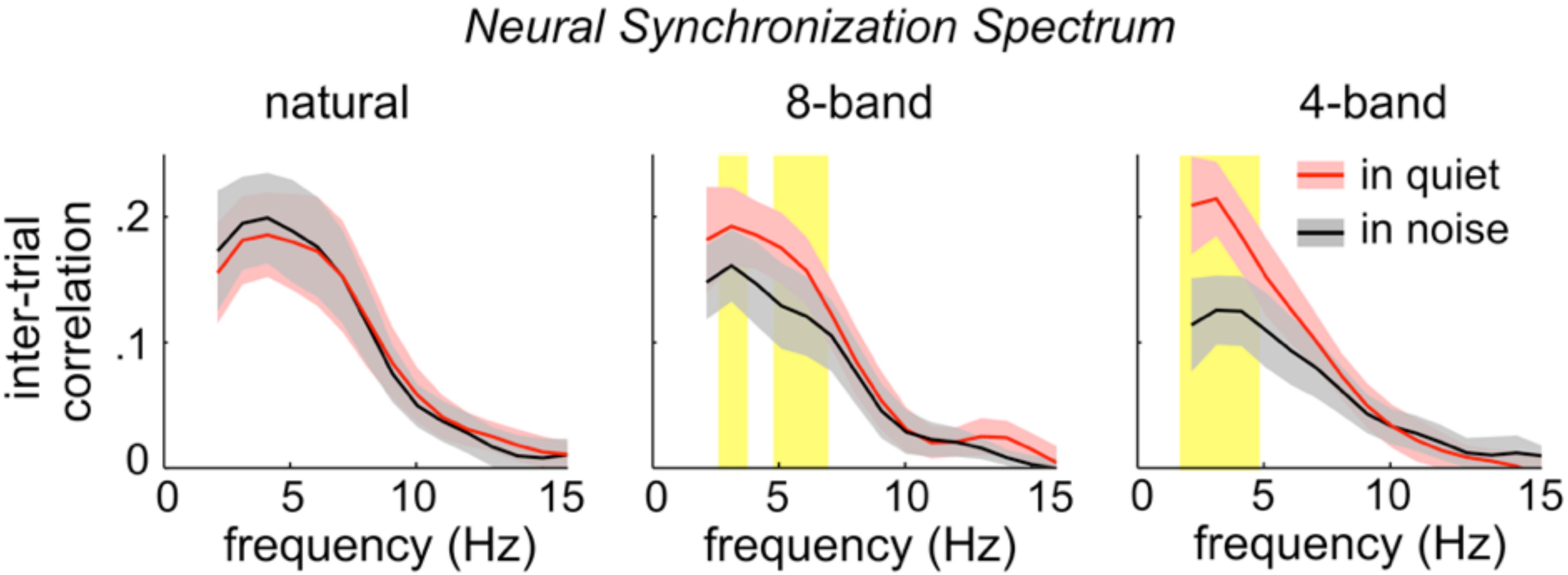
Noise-Vocoded Speech



“in noise” = +3 dB SNR

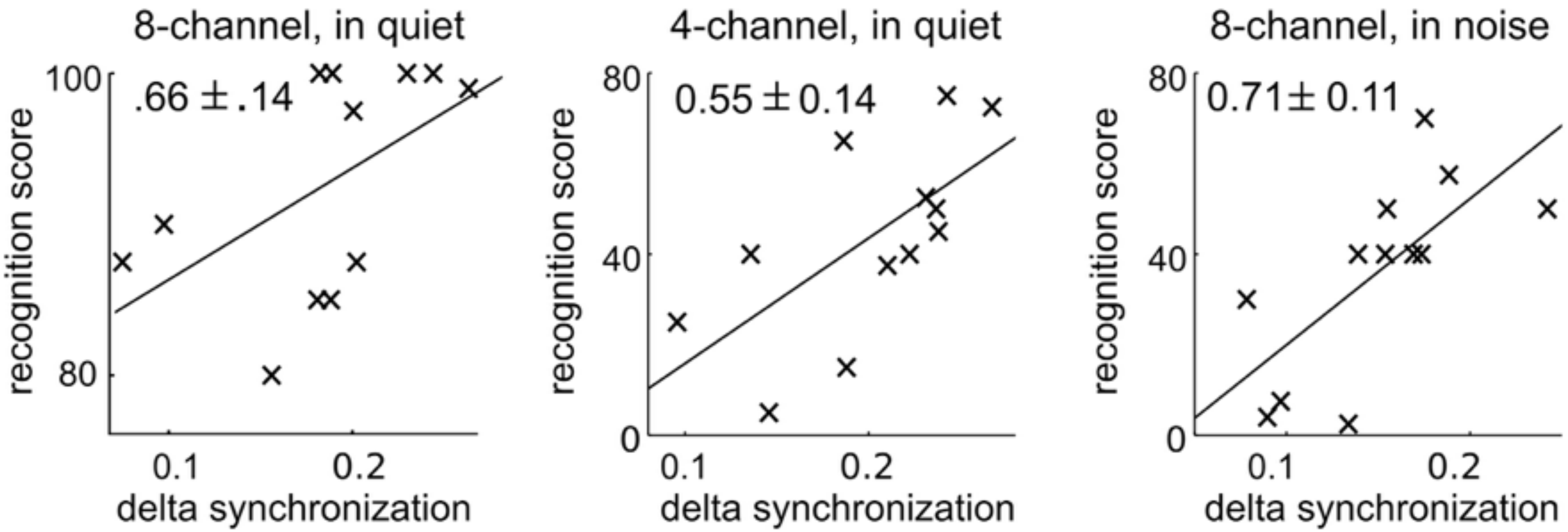
Ding, Chatterjee & Simon, NeuroImage (2014)

Noise-Vocoded Speech: Results



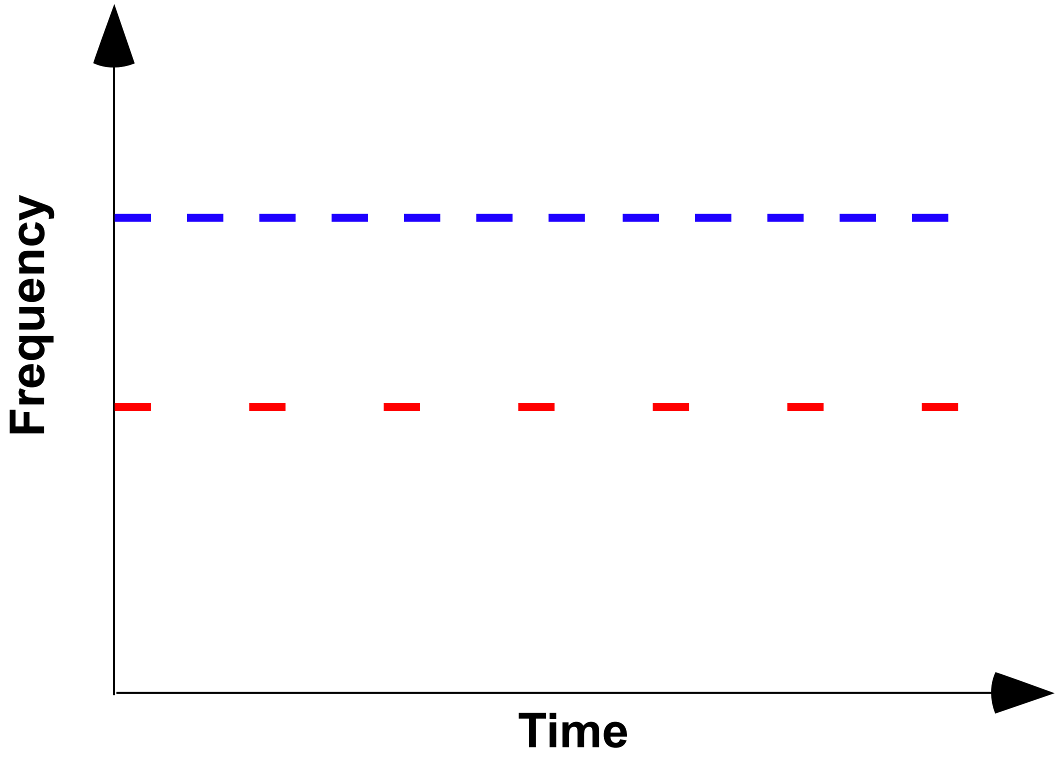
- Cortical entrainment to natural speech robust to noise
- Cortical entrainment to vocoded speech is not
- Not explainable by passive envelope tracking mechanisms
  - noise vocoding does not directly affect the stimulus envelope

Noise-Vocoded Speech: Results

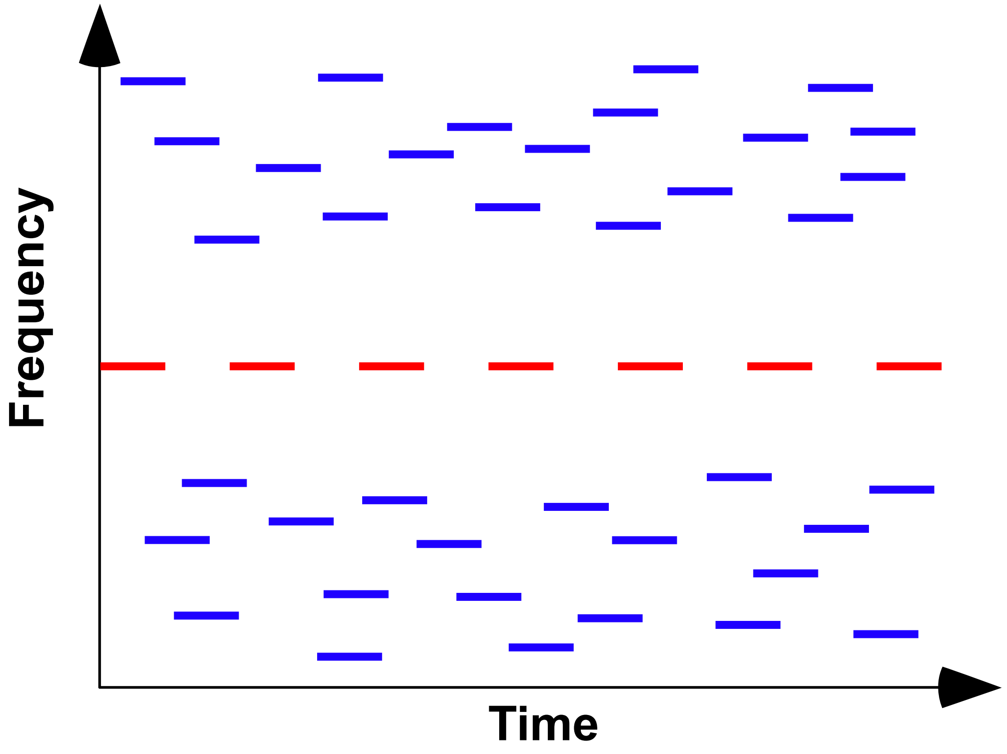


Not Just Speech

Competing Tone Streams



Tone Stream in Masker Cloud



Xiang et al., J Neuroscience (2010)

Elhilali et al., PLoS Biology (2009)

Summary

- Cortical representations of speech found here:
  - ✓ consistent with being *neural* representations of auditory *perceptual* objects
  - ✓ very robust to noise (~intelligibility)
  - ✓ relies on *spectro*-temporal fine structure
  - ✓ explicitly temporal representation
- Object representation at 100 ms latency (PT), but not by 50 ms (HG)

Thank You