Neural Representations of 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

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

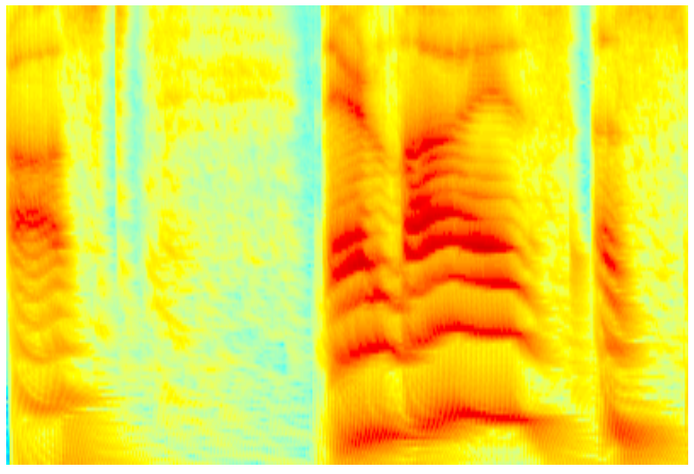
Funding

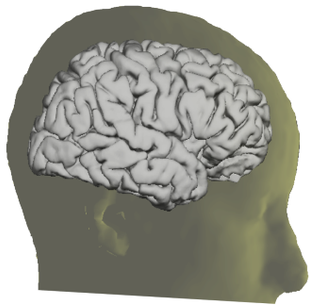
NIH R01 DC 008342
NIH R01 DC 007657
NIH R01 DC 005660
NIH R01 DC 000436
NIH R01 AG 036424
NIH R01 AG 027573
NIH R01 EB 004750
NIH R03 DC 004382
USDA 20096512005791

Introduction

- Magnetoencephalography (MEG) & Speech
- Speech as example of Auditory Object
- Neural Representations of Auditory Objects (e.g., speech) in Auditory Cortex

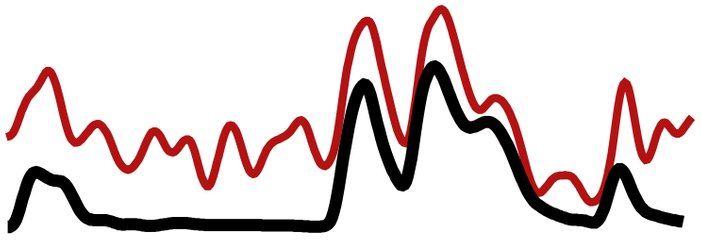
MEG Responses to Speech



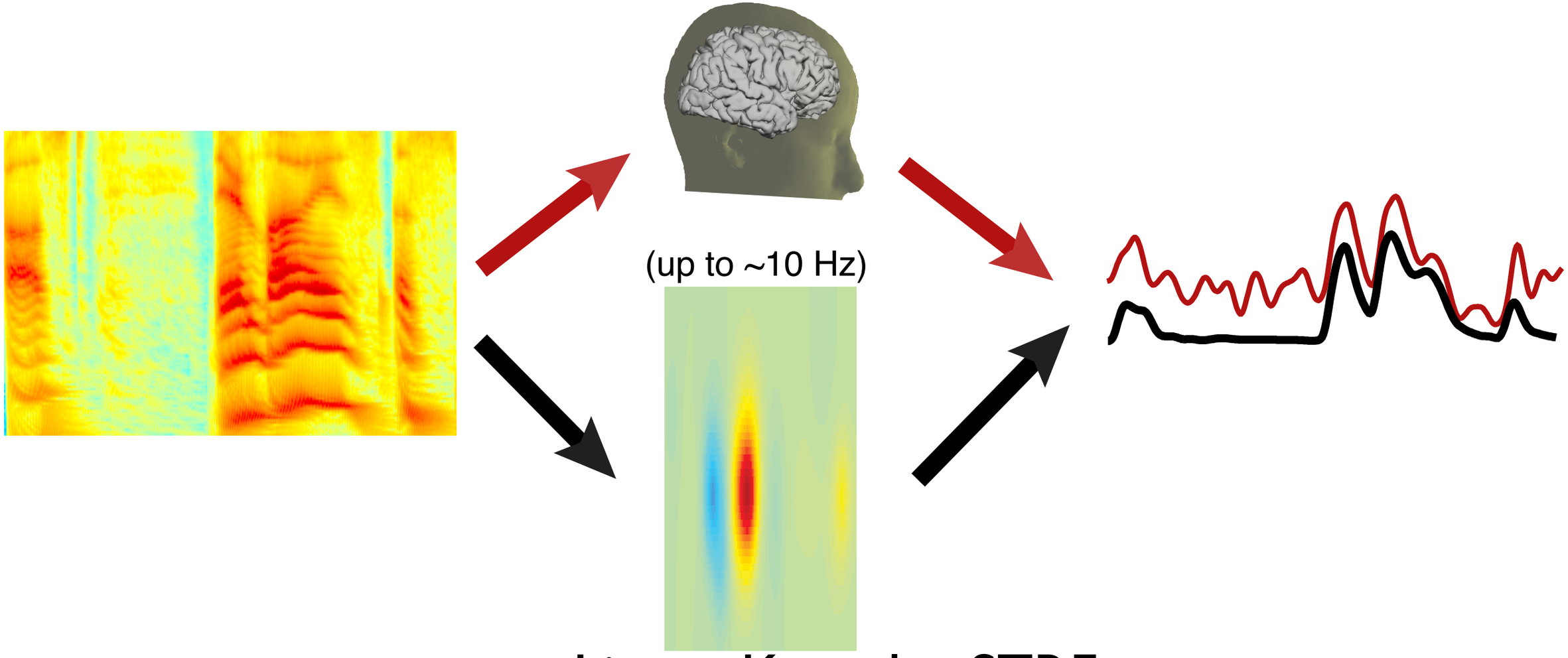


(up to ~10 Hz)

Auditory Model



MEG Responses Predicted by STRF Model

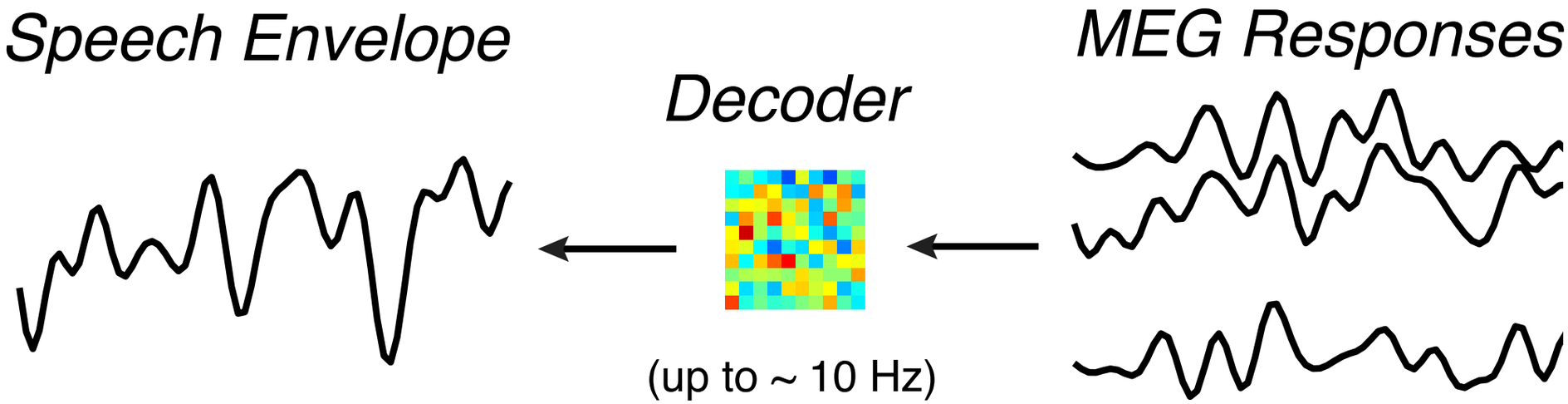


Linear Kernel = STRF

"Spectro-Temporal Response Function"

Ding & Simon, J Neurophysiol (2012)

Neural Reconstruction of Speech Envelope



Auditory Objects

- What is an auditory object?
  - perceptual (not neural, not acoustic)
  - commonalities with visual objects
  - example: speech stream (“voice”)
  - several 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

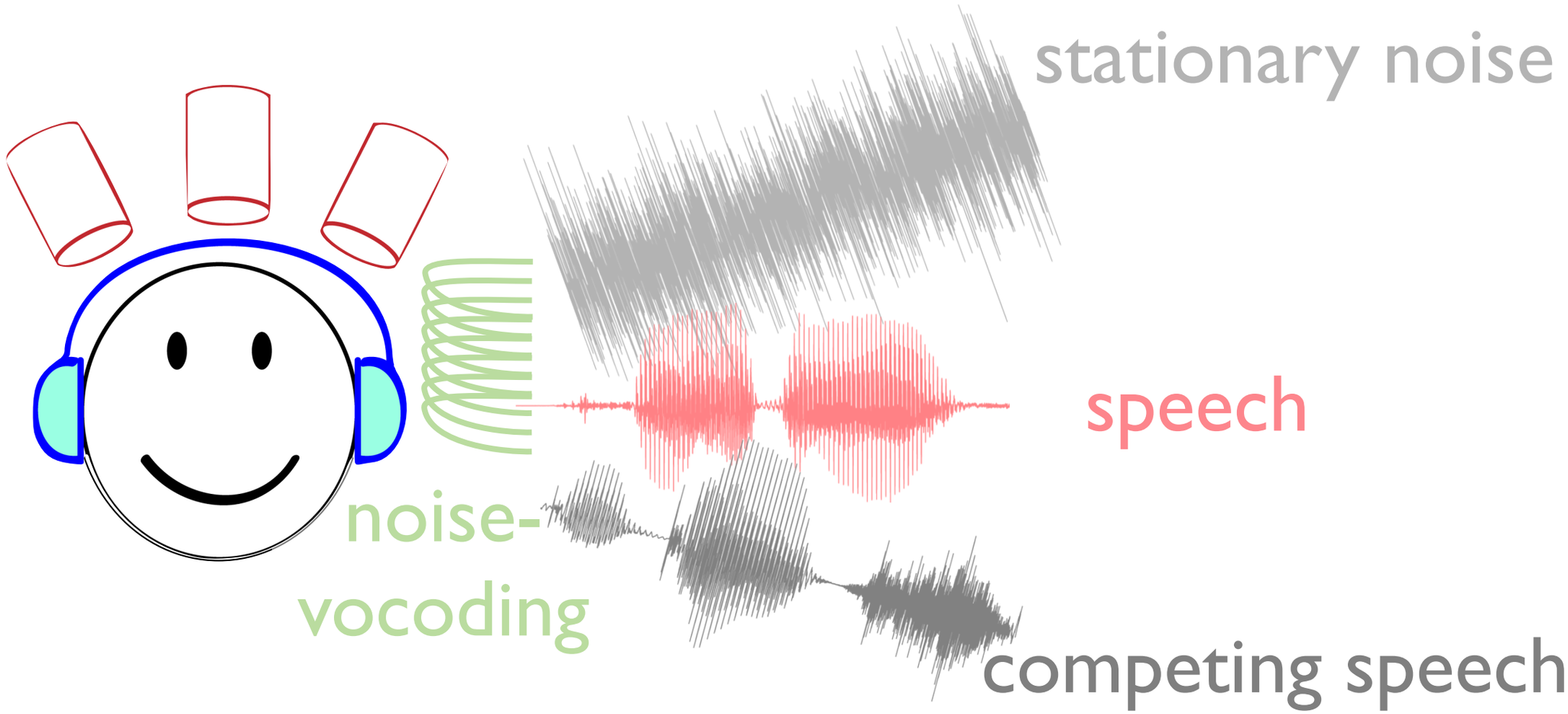
Auditory Objects at the Cocktail Party

Alex Katz,
The Cocktail Party

Auditory Objects at the Cocktail Party

Alex Katz,
The Cocktail Party

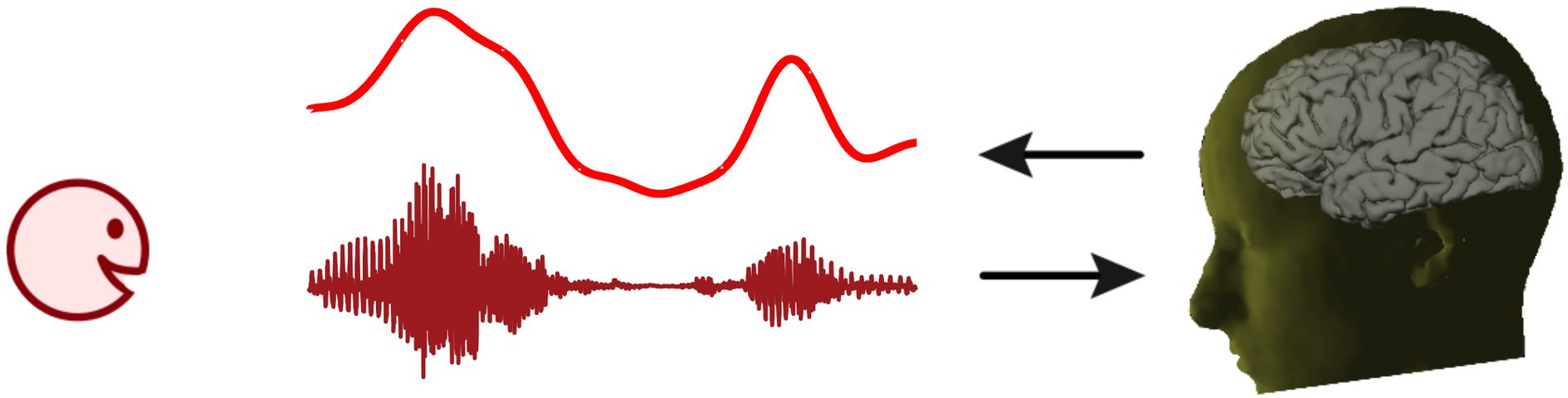
Experiments



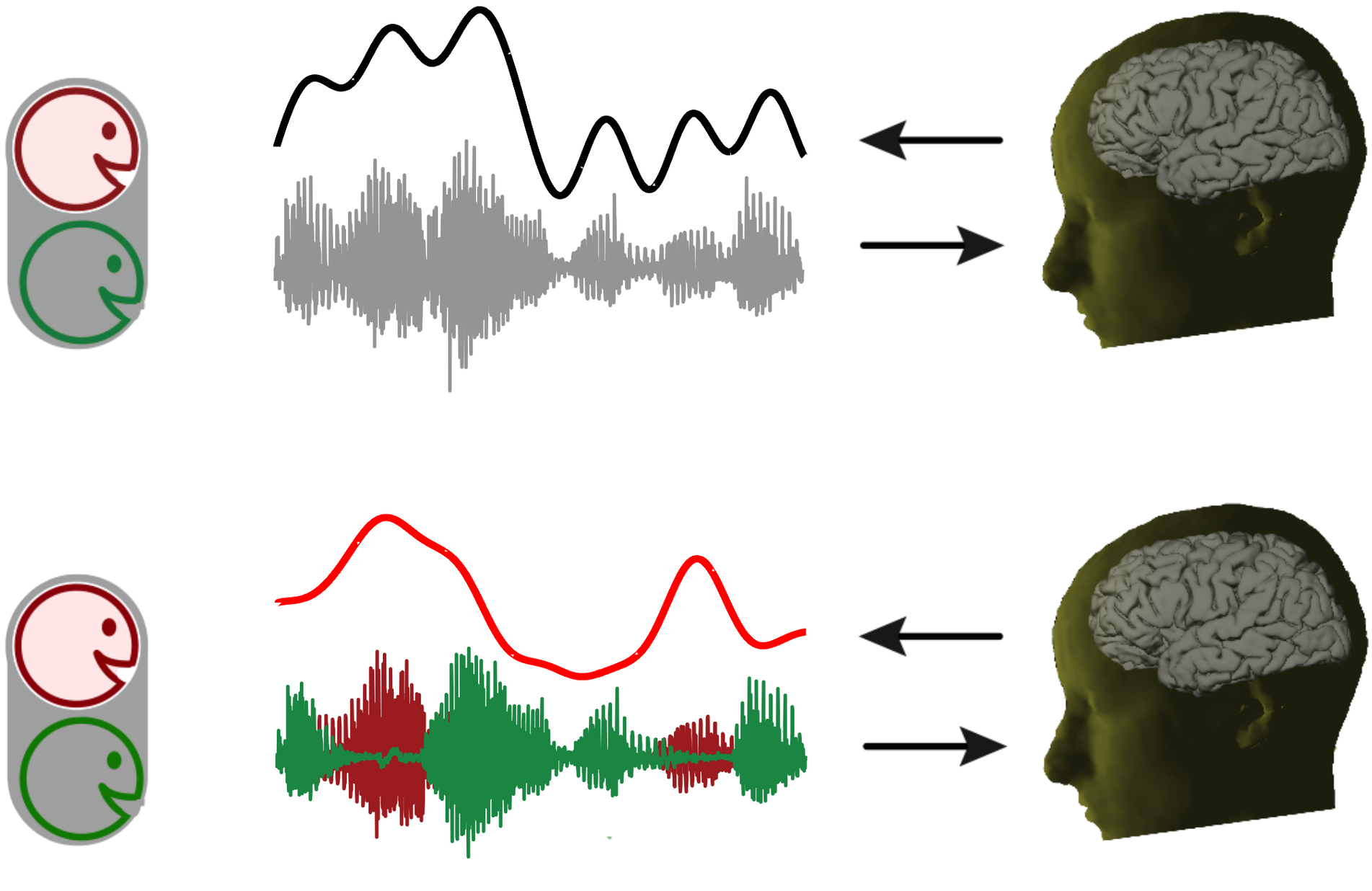
Neural Representation of an Auditory Object

- 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

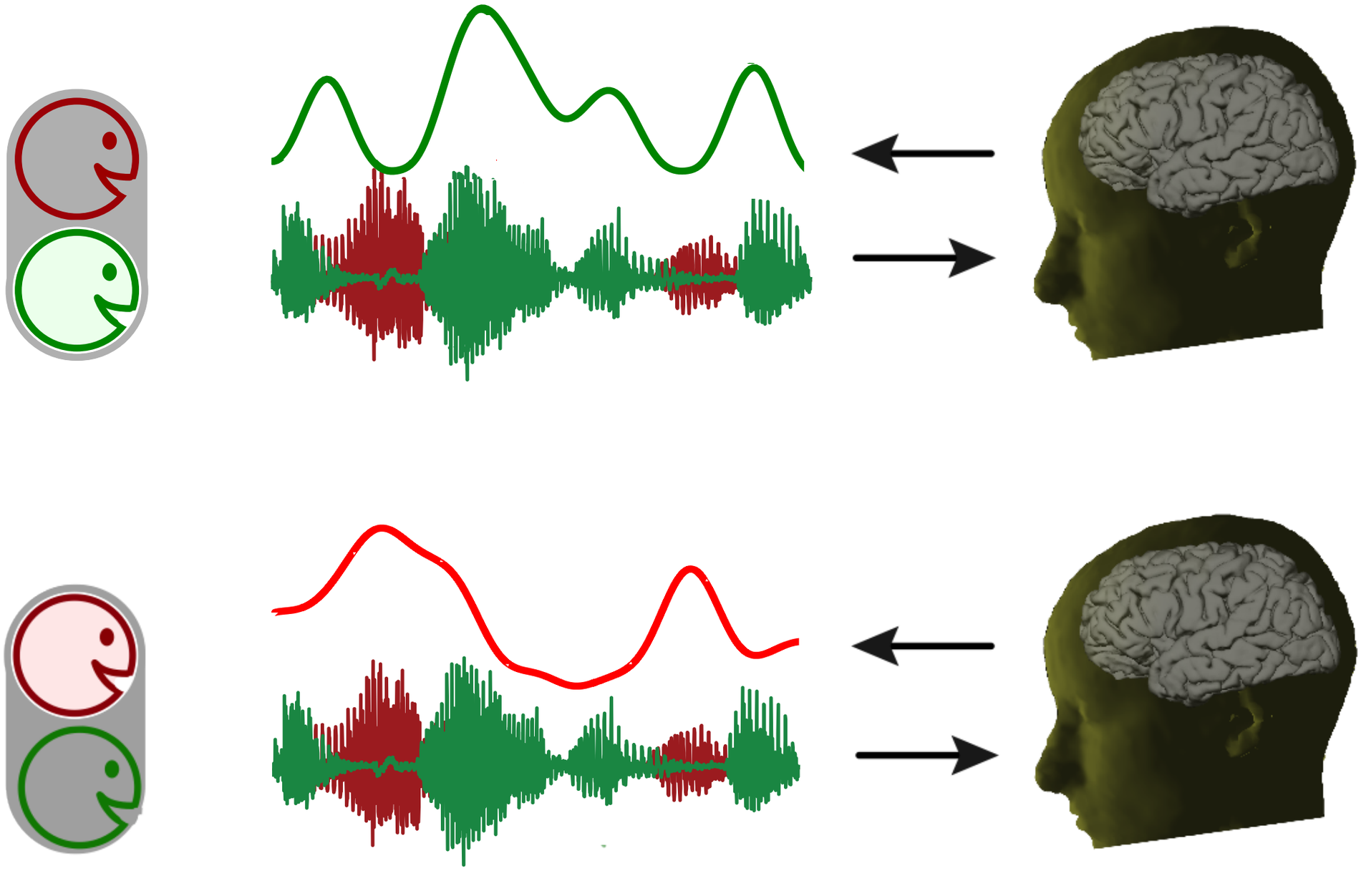
Selective Neural Encoding



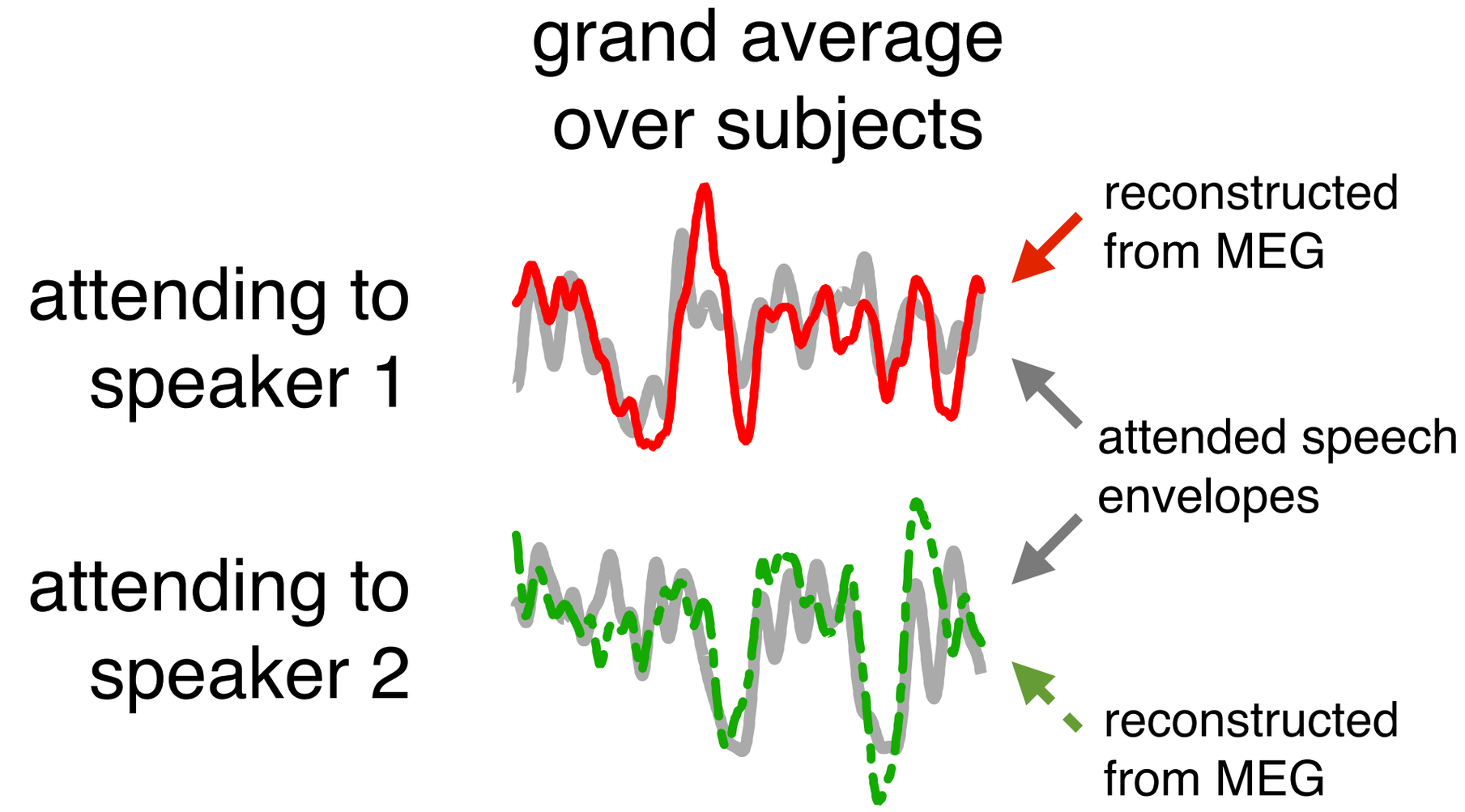
Unselective vs. Selective Neural Encoding



Selective Neural Encoding



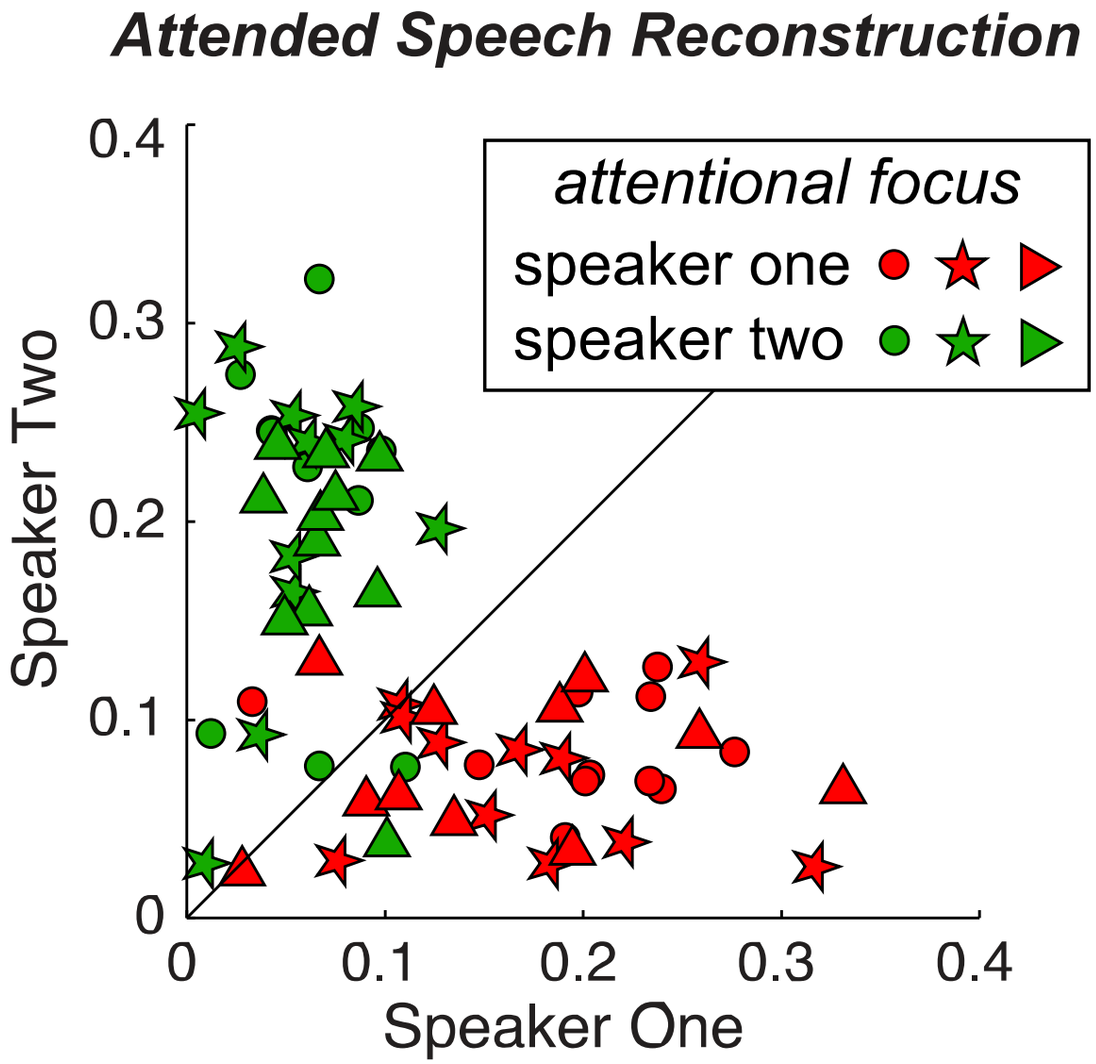
Stream-Specific Representation



Identical Stimuli!

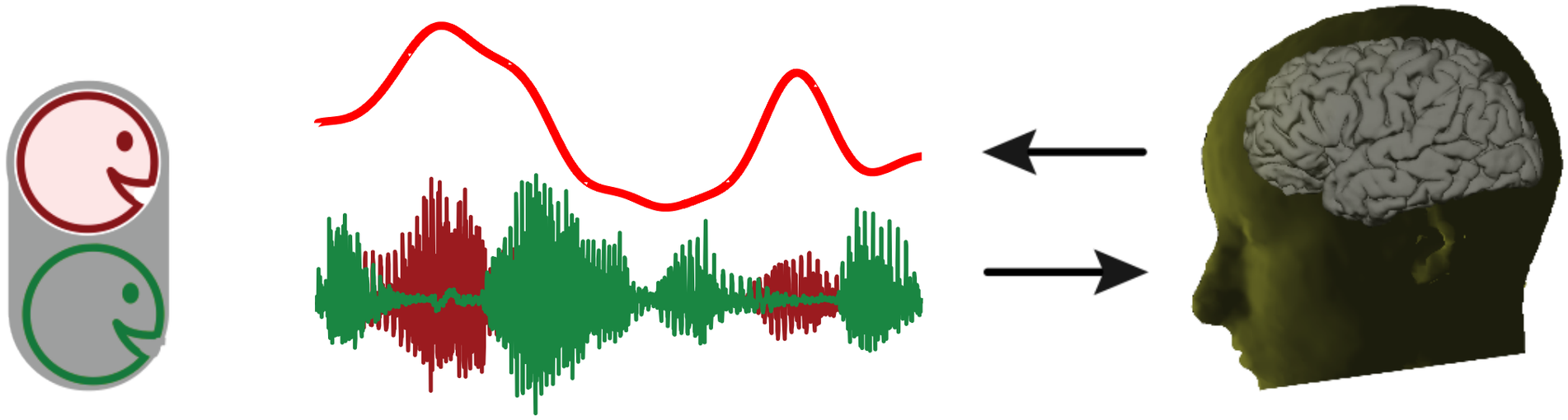
Ding & Simon, PNAS (2013)

Single Trial Speech Reconstruction



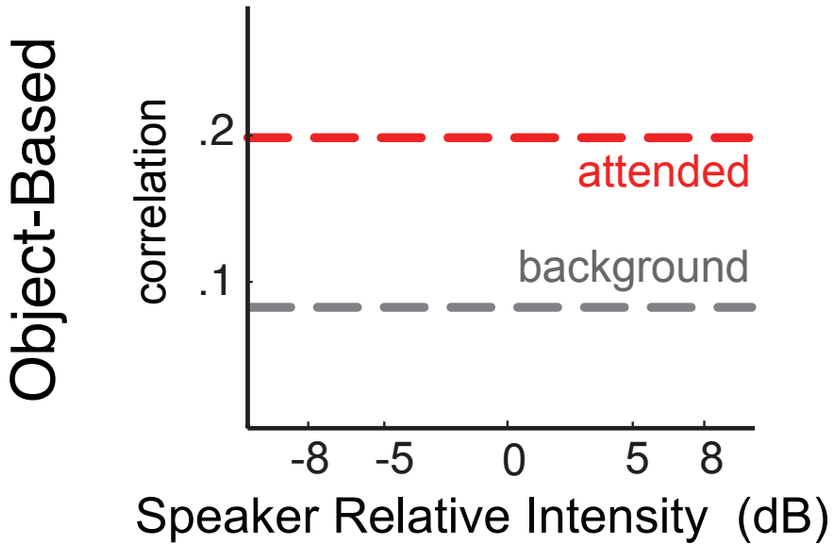
Ding & Simon, PNAS (2013)

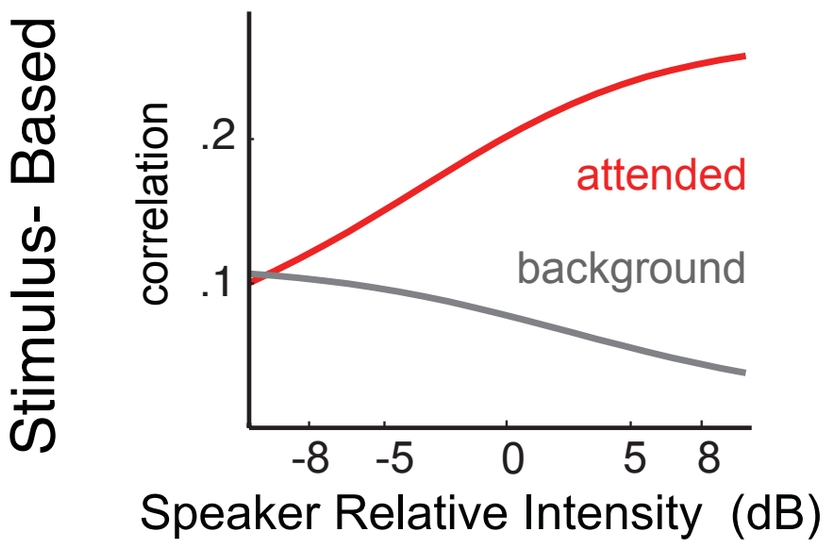
Invariance Under Acoustic Changes



Independence from Acoustic Specifics

Neural Response vs. Acoustics





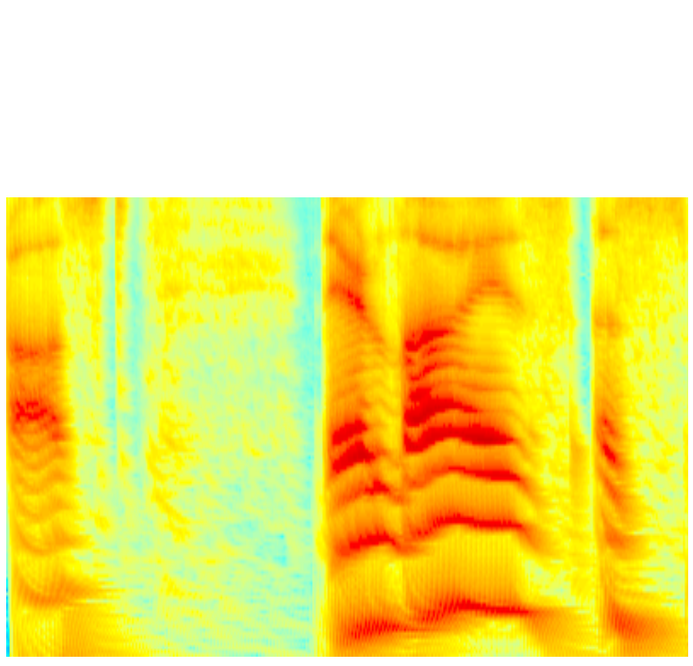
Neural Representation of an Auditory Object

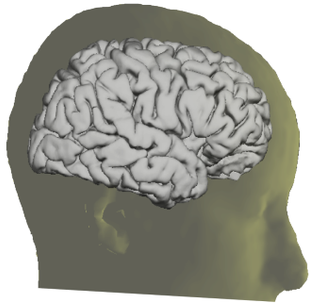
✓ neural representation is of something in sensory world

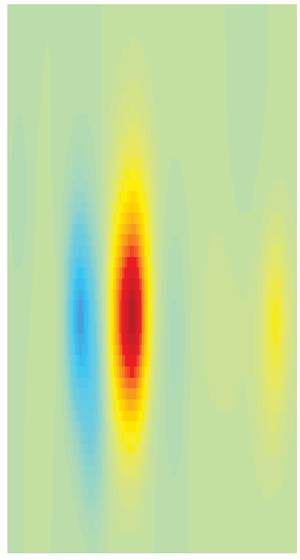
✓ when other sounds mixed in,
neural representation is of auditory object,
not entire acoustic scene

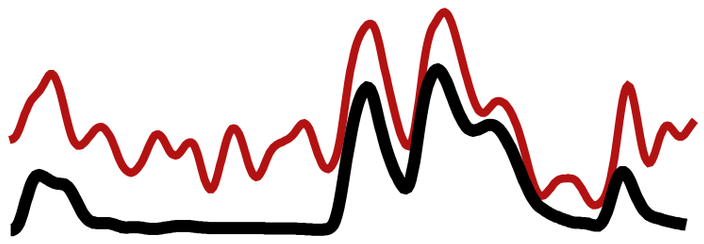
✓ neural representation invariant
under broad changes in specific acoustics

Forward STRF Model



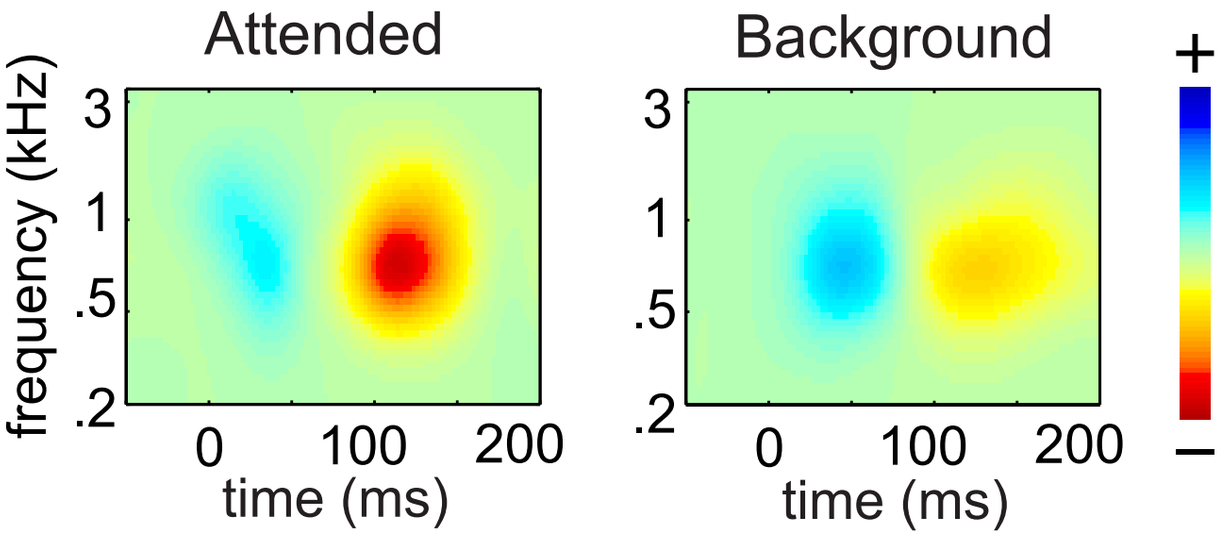






Spectro-Temporal Response Function (STRF)

STRF Results



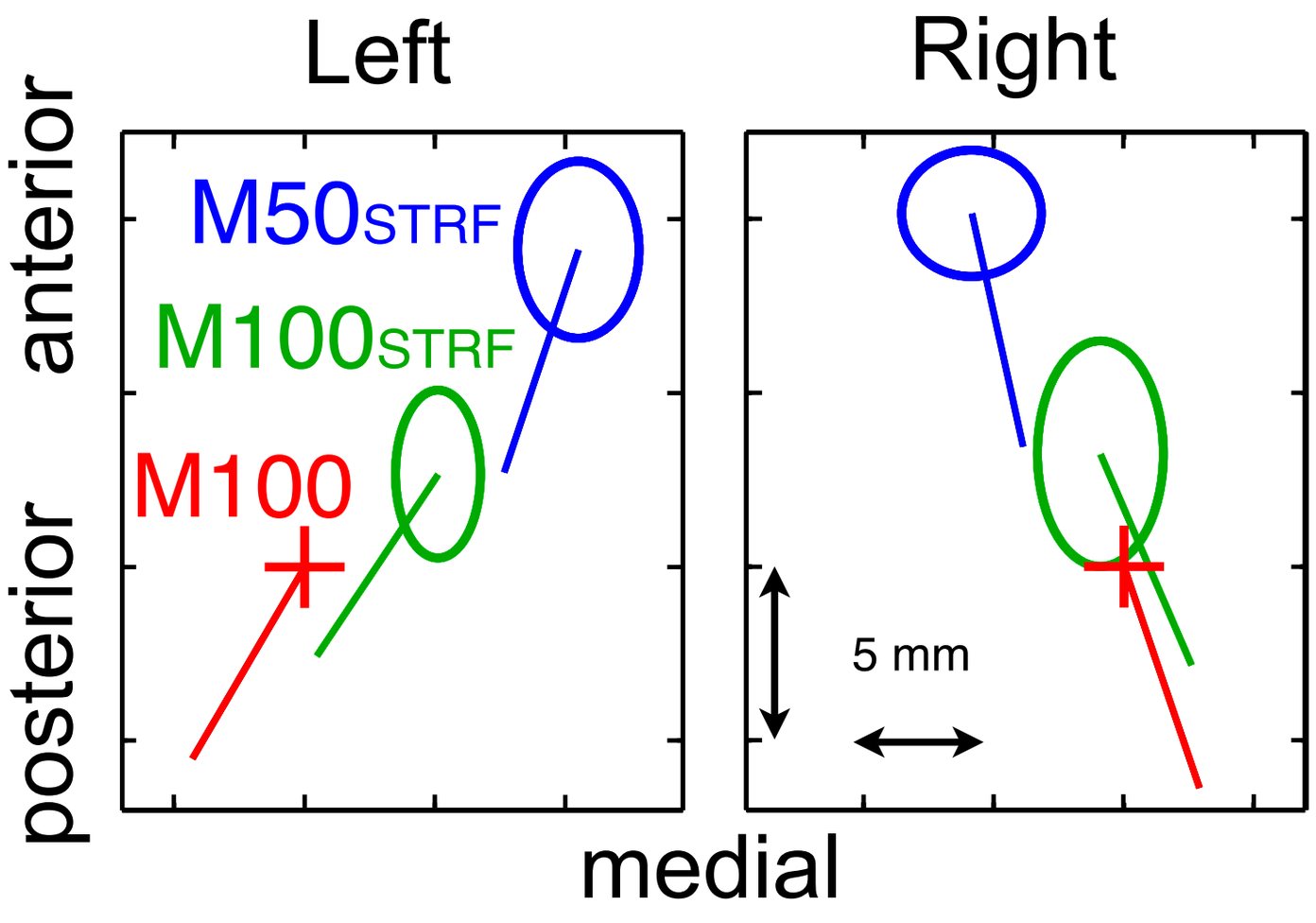
- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- $M50_{STRF}$ positive peak
- $M100_{STRF}$ negative peak

- **$M100_{STRF}$ strongly modulated by attention, *but not $M50_{STRF}$***

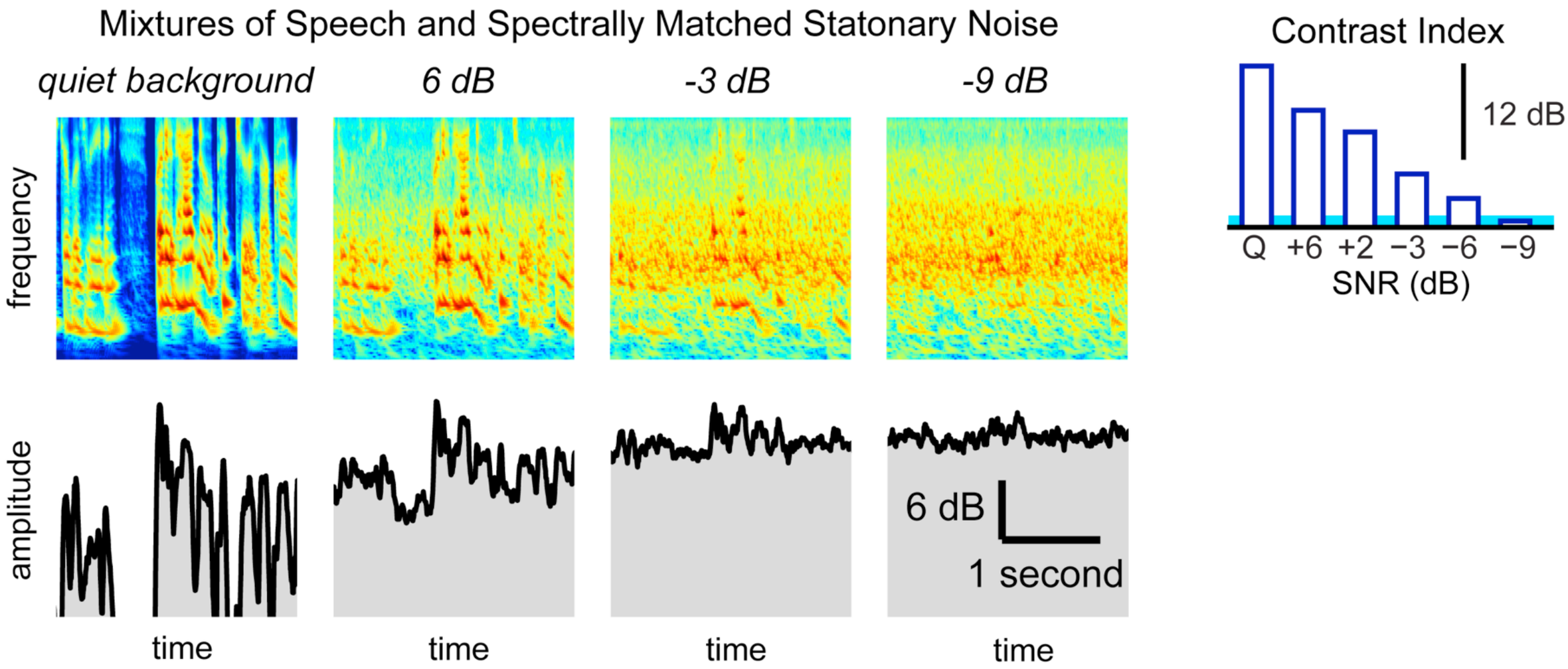
Neural Sources

- 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

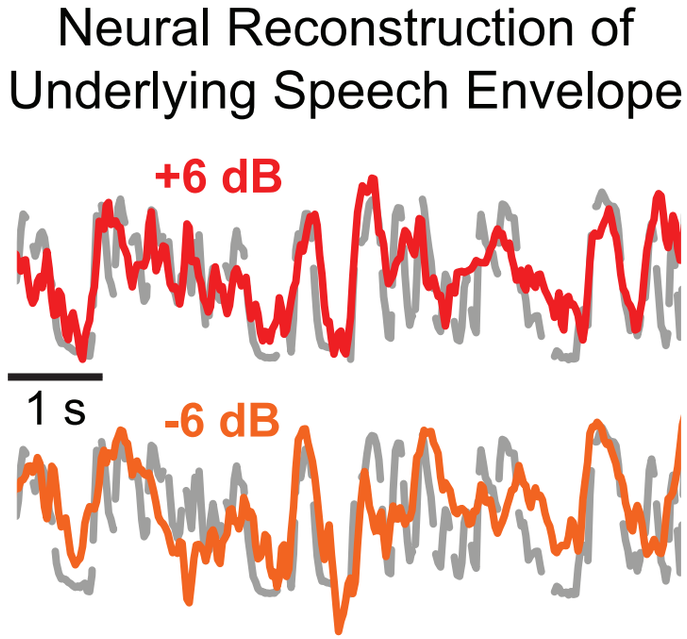


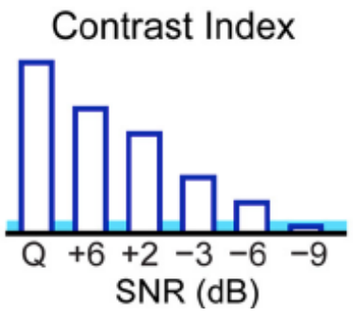
Speech in Noise

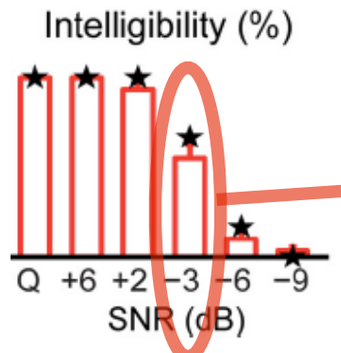


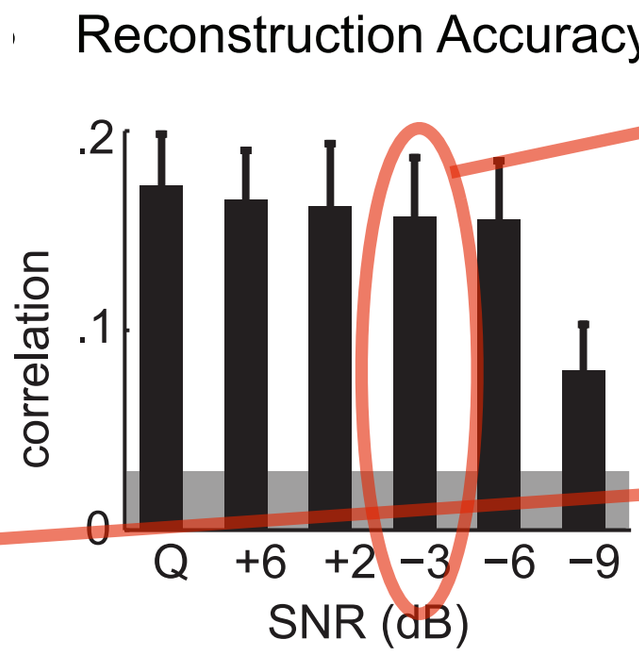
Ding & Simon, J Neuroscience (2013)

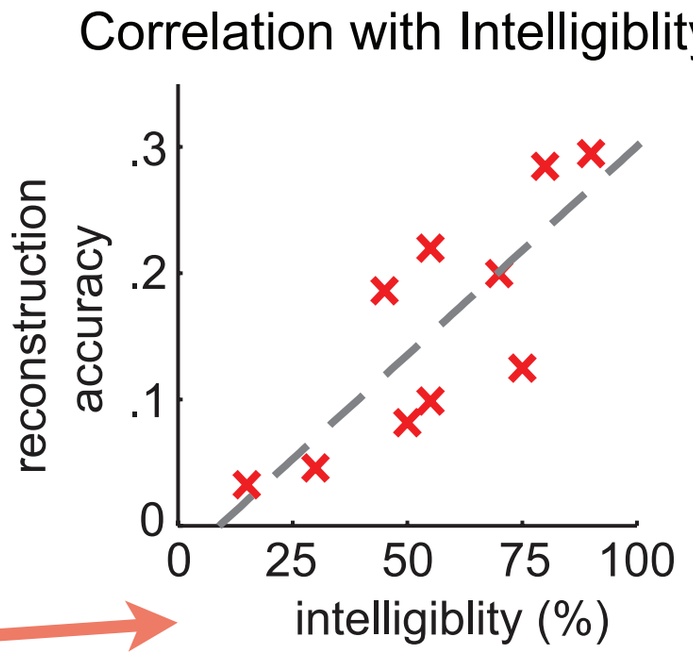
Cortical Representation of Speech in Noise



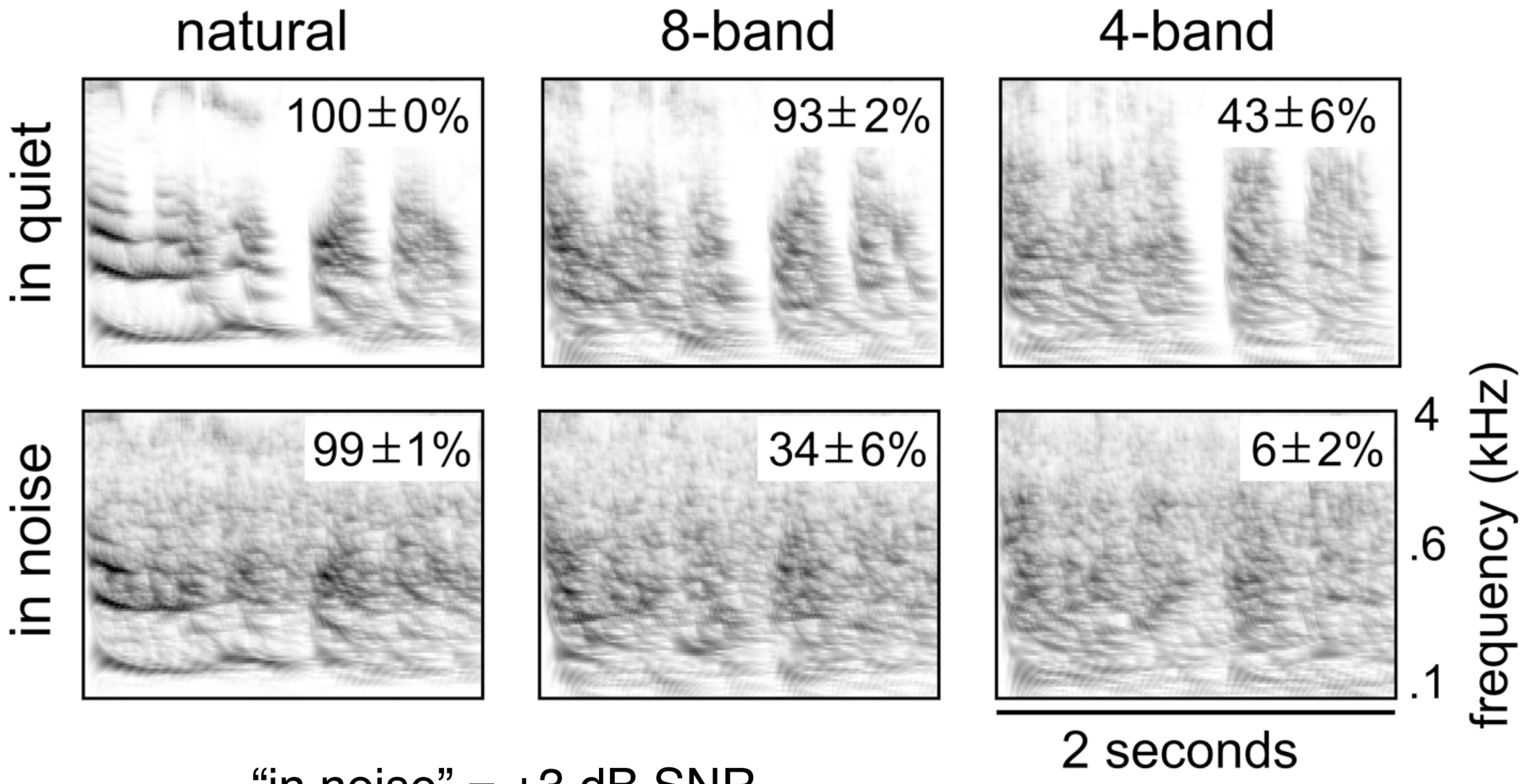








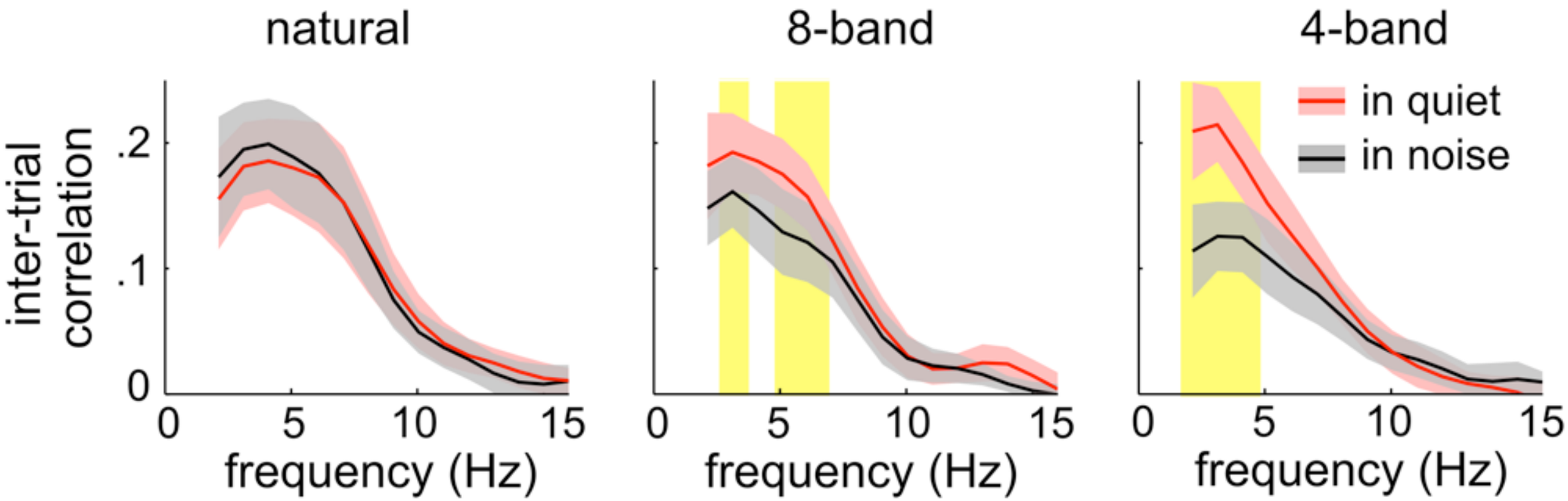
Noise-Vocoded Speech



"in noise" = +3 dB SNR

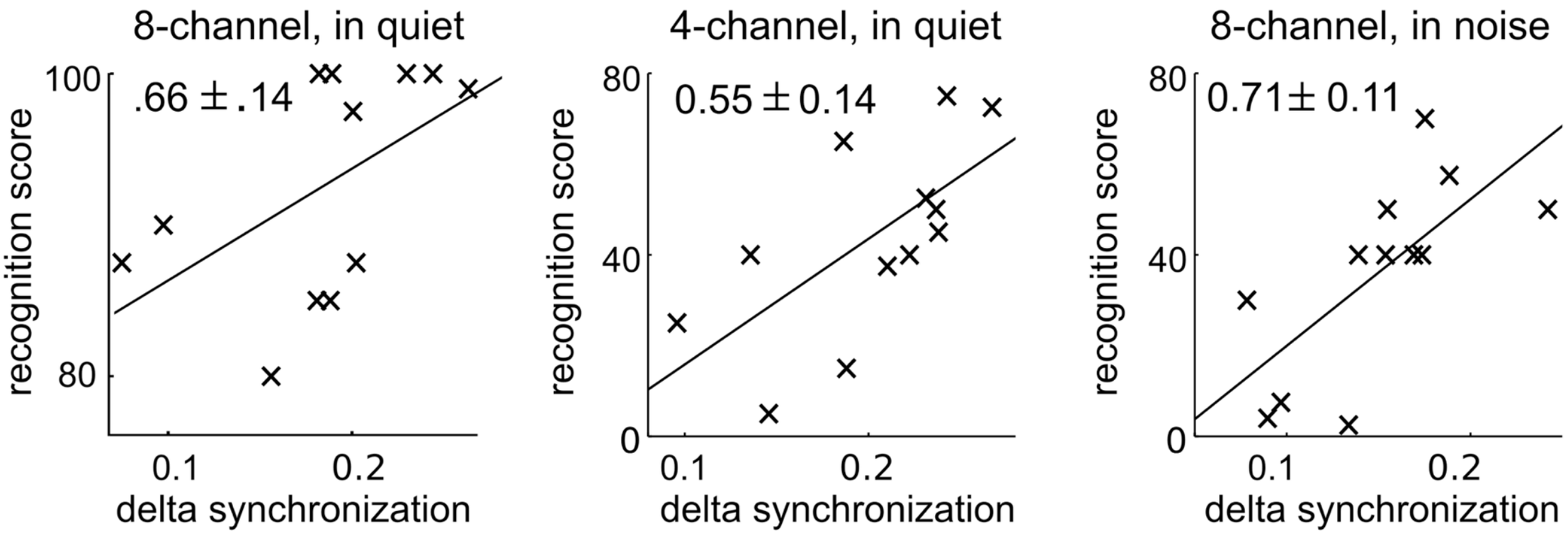
Ding, Chatterjee & Simon, NeuroImage (2014)

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



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)