Neural Representations of Speech, and Speech in Noise, 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

SPIRE 2016, 9 Jan 2016

Acknowledgements

Current (Simon Lab & Affiliates)

Francisco Cervantes
Natalia Lapinskaya
Mahshid Najafi
Alex Presacco
Krishna Puvvada
Lisa Uible
Peng Zan

Past (Simon Lab & Affiliate Labs)

Nayef Ahmar
Sahar Akram
Murat Aytekin
Claudia Bonin
Maria Chait
Marisel Villafane Delgado
Kim Drnec
Nai Ding
Victor Grau-Serrat
Julian Jenkins
David Klein
Ling Ma
Kai Sum Li
Huan Luo
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
Cindy Moss
David Poeppel
Shihab Shamma

Past Postdocs & Visitors

Aline Gesualdi Manhães
Dan Hertz
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
Isidora Ranovadovic
Andrea Shome
Madeleine Varmer
Ben Walsh

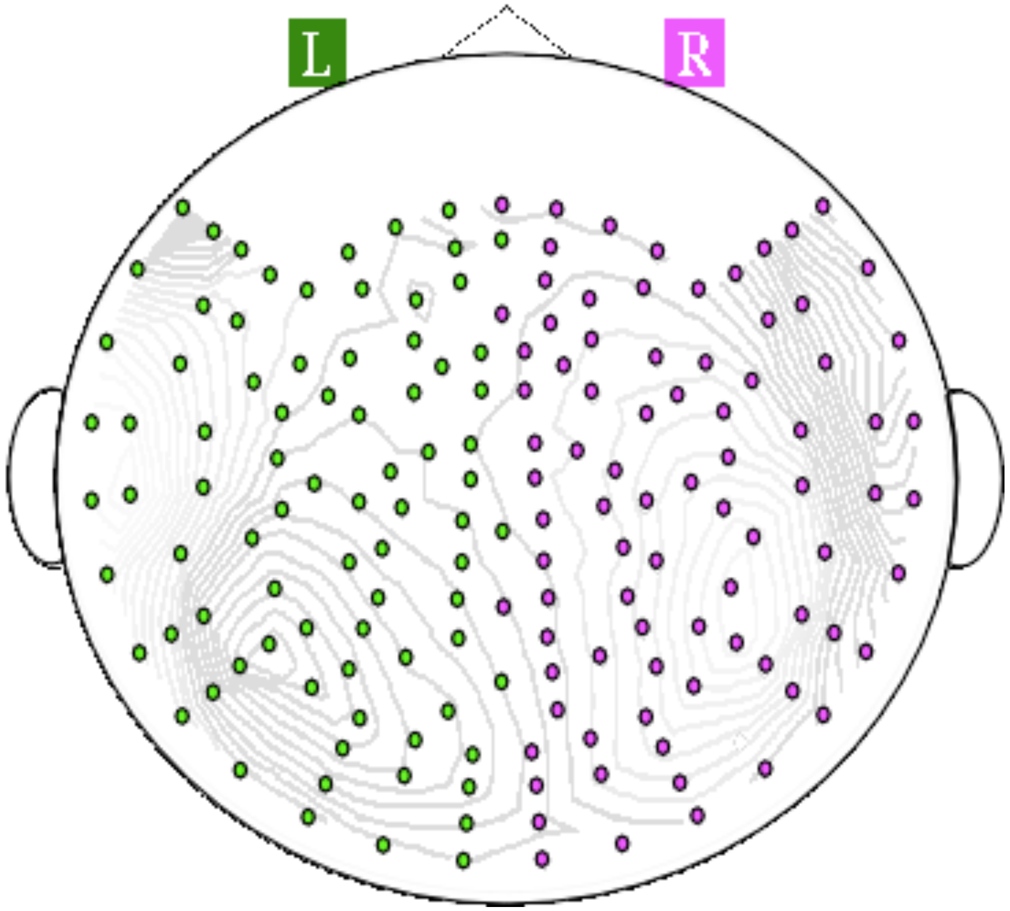
Funding NIH (*NIDCD*, *NIA*, *NIBIB*); USDA

Outline

- Cortical Representations of Speech (via MEG)
    - Encoding vs. Decoding
- Cortical Representations of Speech in Noise
- Recent Studies:
    - Attentional Dynamics
    - Aging & Cortical Representations of Speech
    - Higher Level Interference & Noise

Magnetoencephalography (MEG)

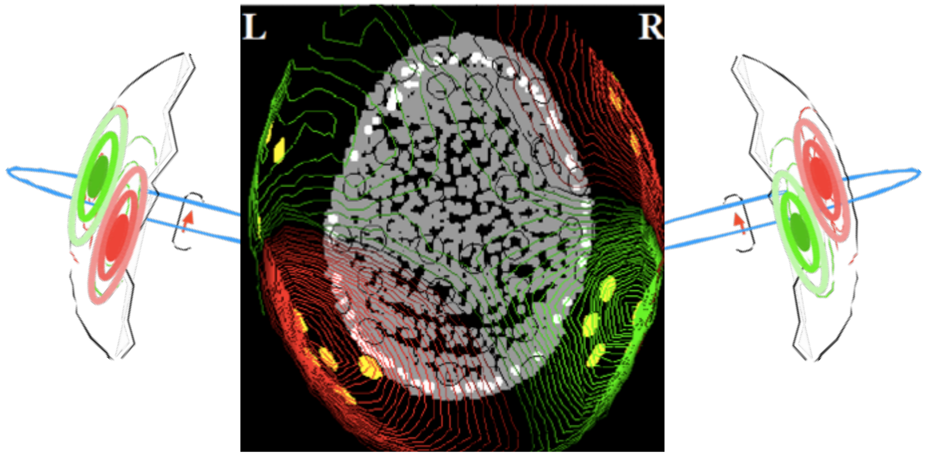
- Non-invasive, Passive, Silent Neural Recordings
- Simultaneous Whole-Head Recording (~200 sensors)
- Sensitivity
  - high: ~100 fT (10^{-13} Tesla)
  - low: ~10^4 – ~10^6 neurons
- Temporal Resolution: ~1 ms
- Spatial Resolution
  - coarse: ~1 cm
  - ambiguous

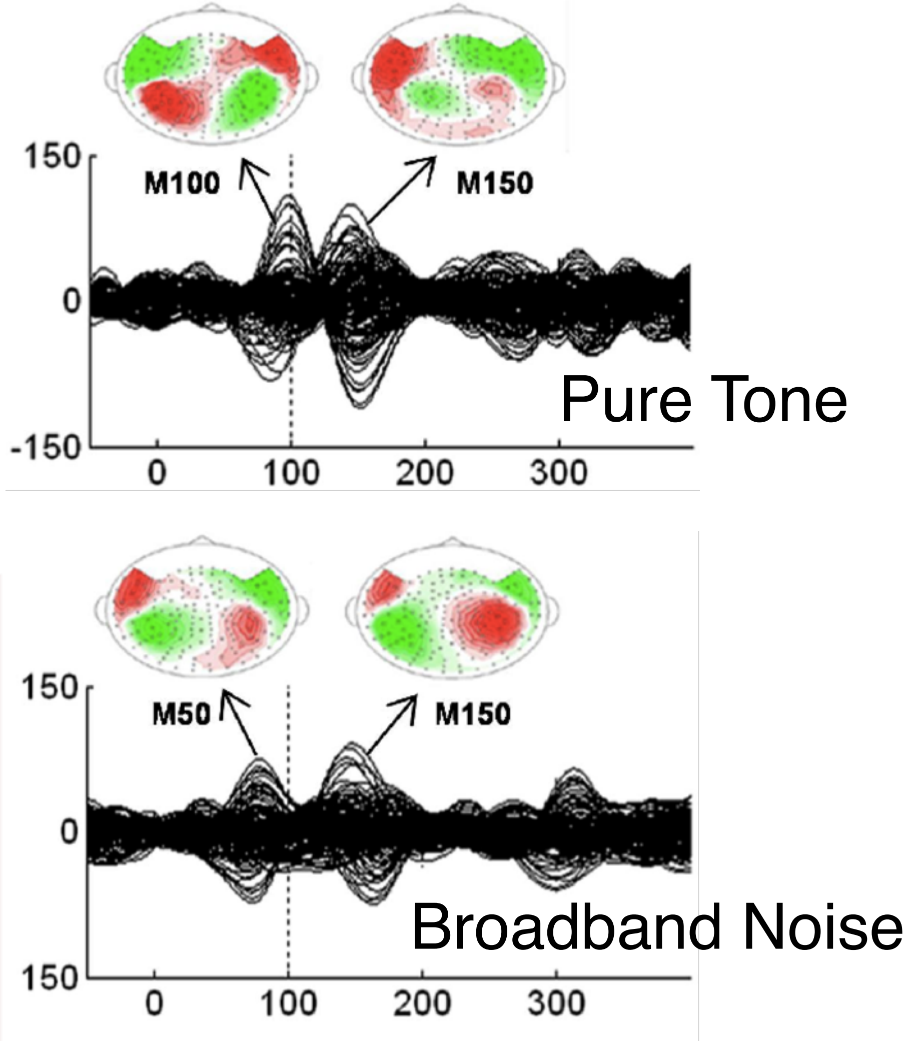


Time Course of MEG Responses

Time Locked Auditory Responses

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

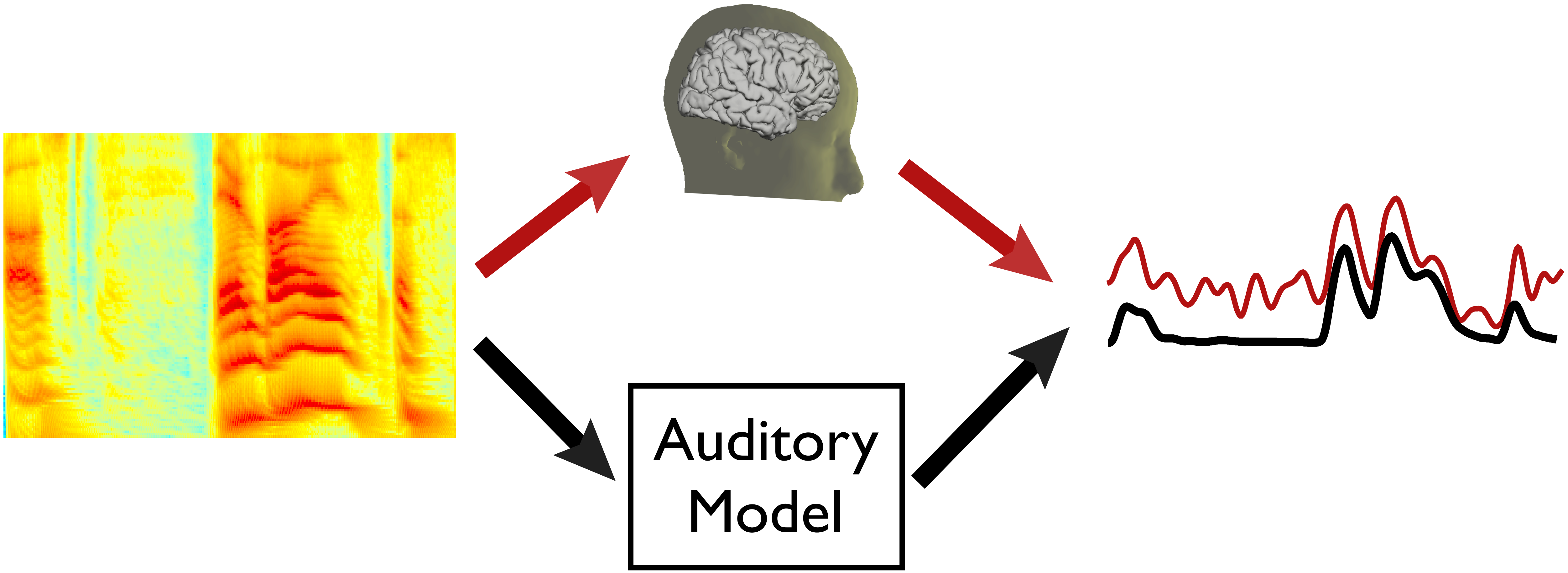




Pure Tone

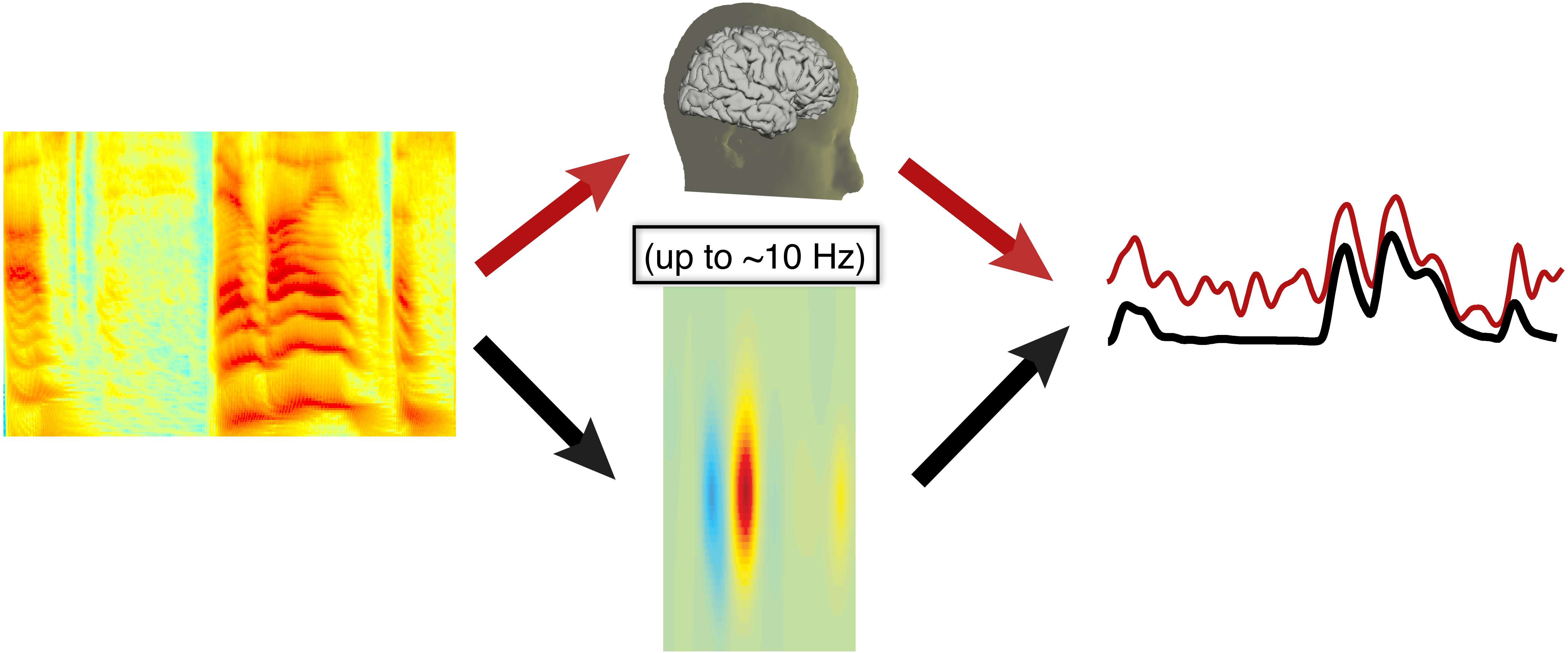
Broadband Noise

MEG Responses to Speech Modulations



Auditory Model

MEG Responses Predicted by STRF Model

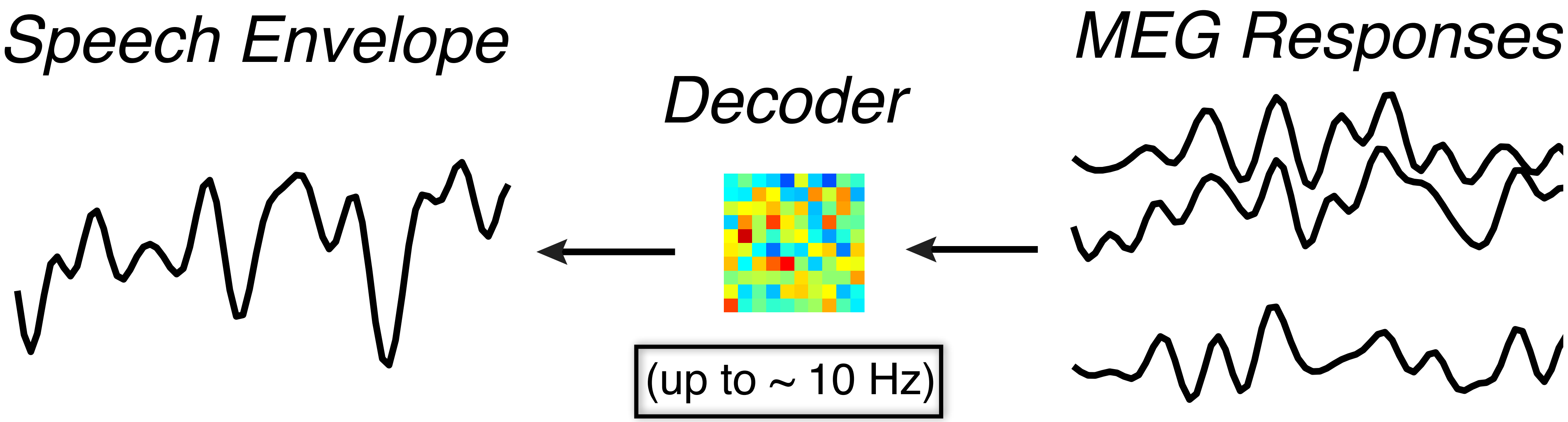


Linear Kernel = STRF

Ding & Simon, J Neurophysiol (2012)

"Spectro-Temporal Response Function"

Neural Reconstruction of Speech Envelope



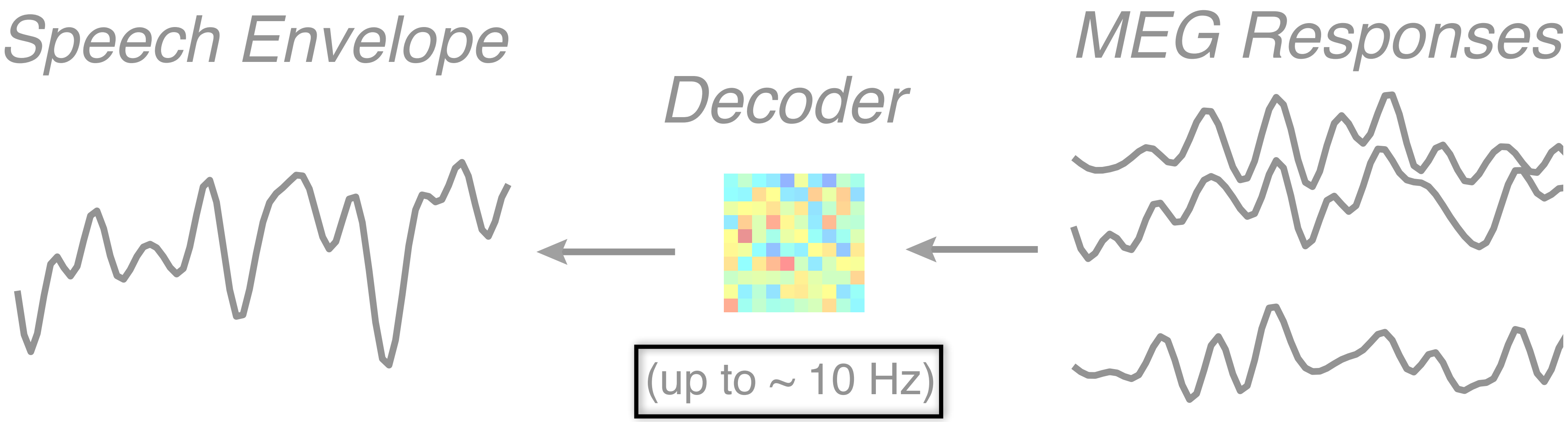
Speech Envelope

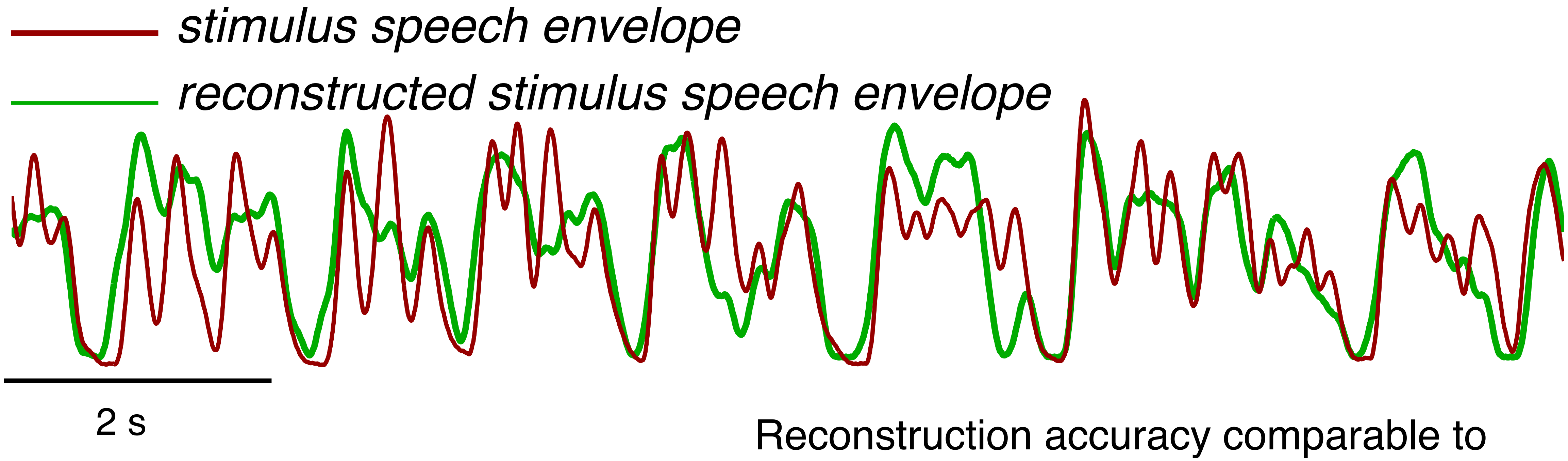
Decoder

MEG Responses

(up to ~ 10 Hz)

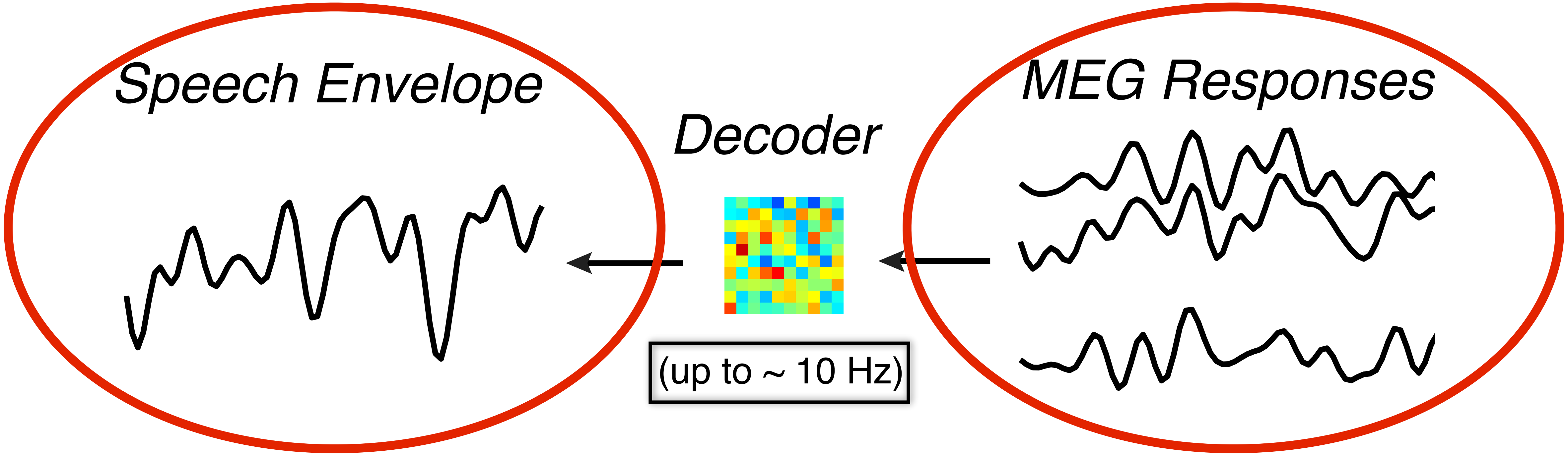
Neural Reconstruction of Speech Envelope



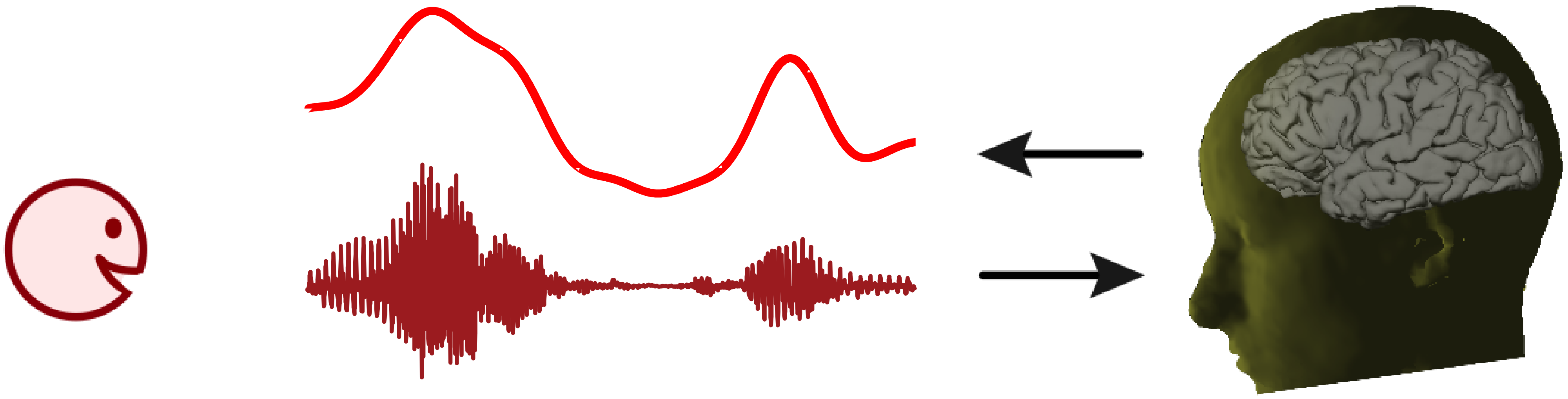


Reconstruction accuracy comparable to single unit & ECoG recordings

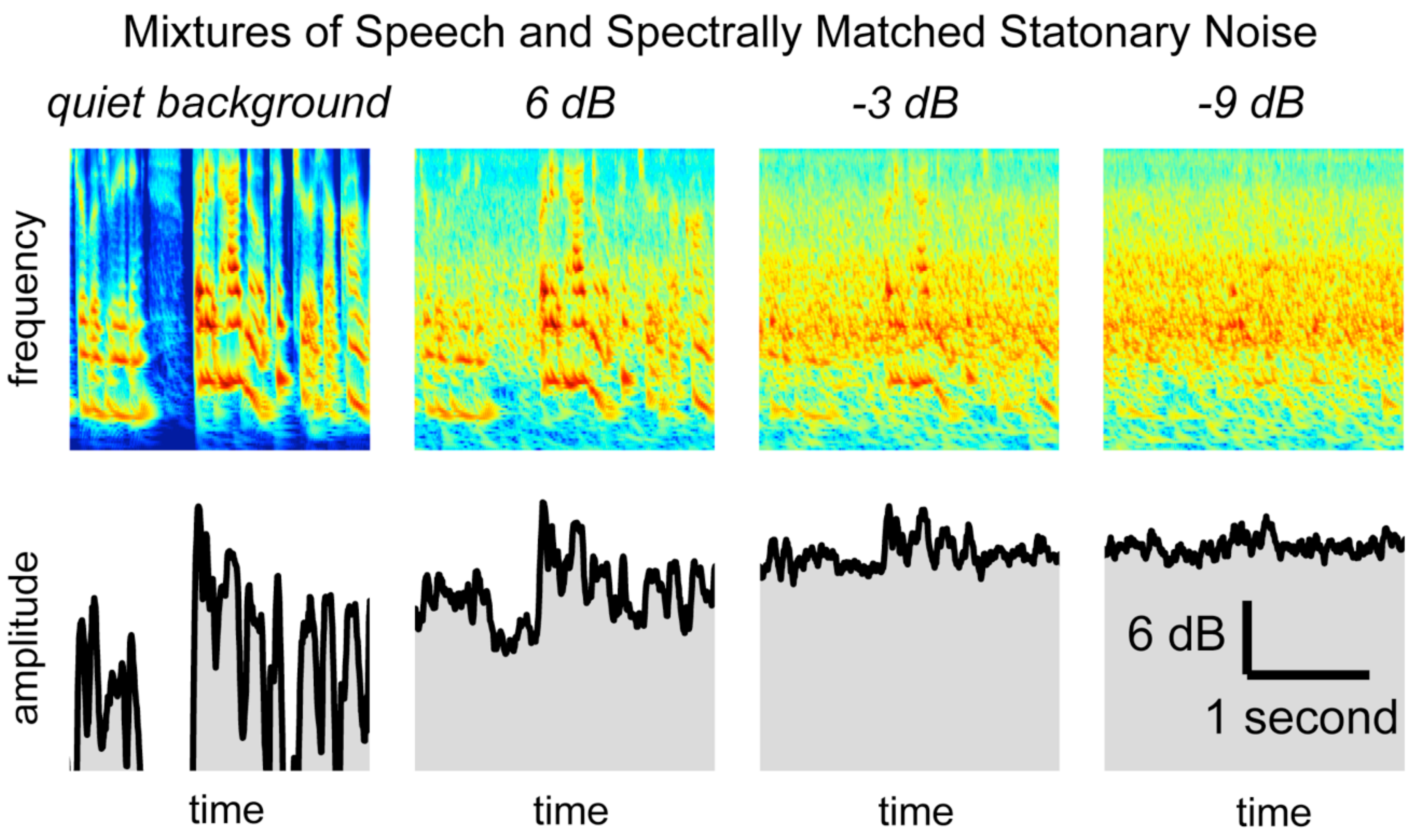
Ding & Simon, J Neurophysiol (2012)
Zion-Golumbic et al., Neuron (2013)


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

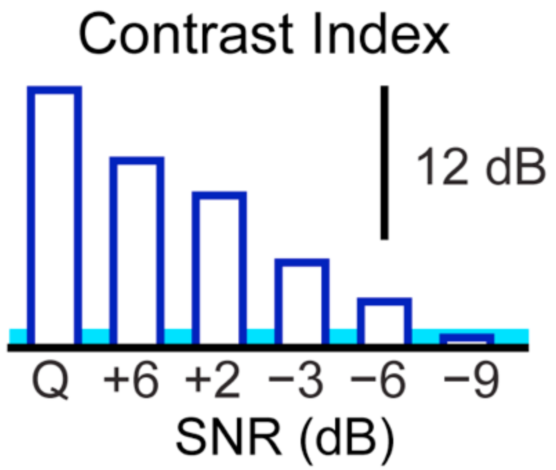
Neural Representation of Speech: Temporal



Speech in Stationary Noise



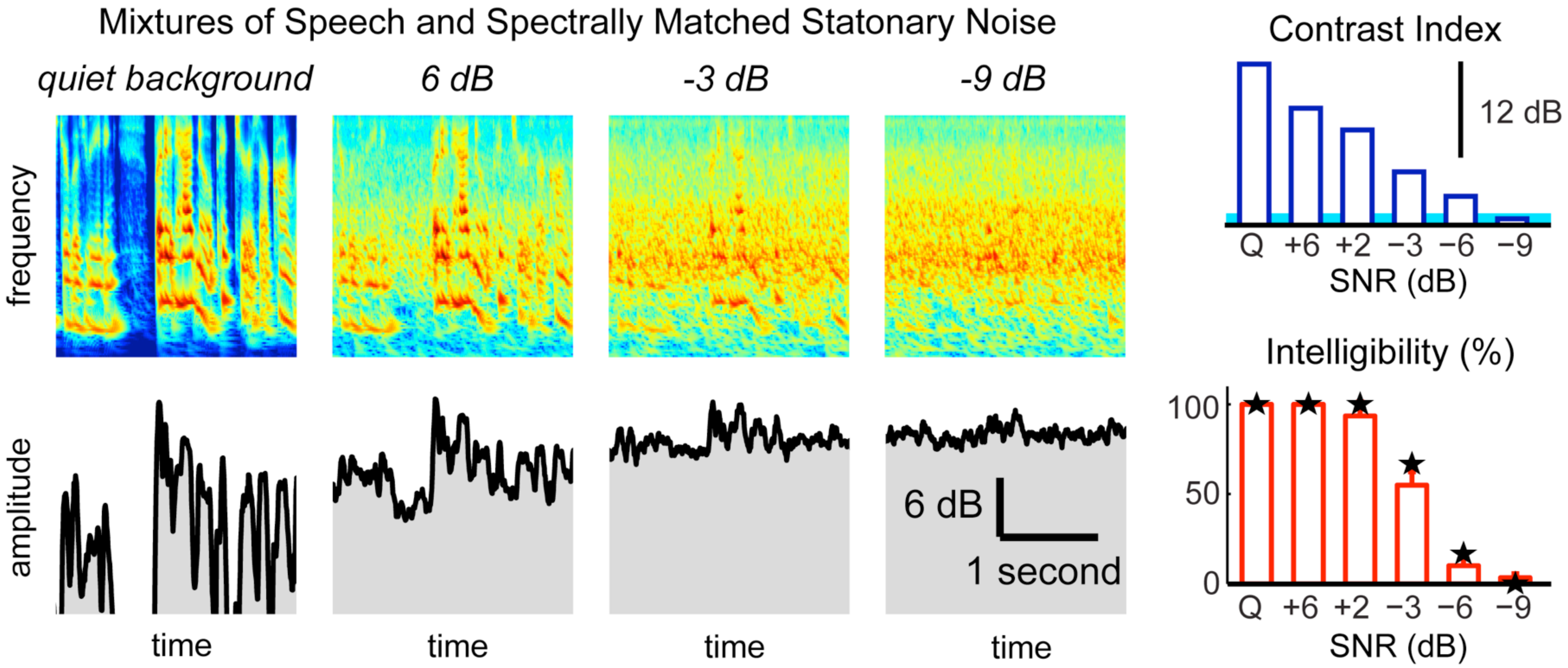
Mixtures of Speech and Spectrally Matched Statonary Noise



Contrast Index

Ding & Simon, J Neuroscience (2013)

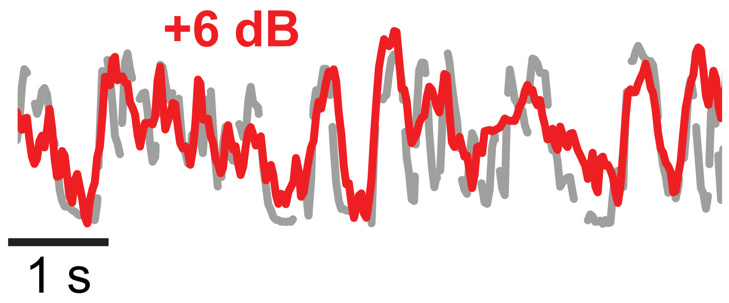
Speech in Stationary Noise



Ding & Simon, J Neuroscience (2013)

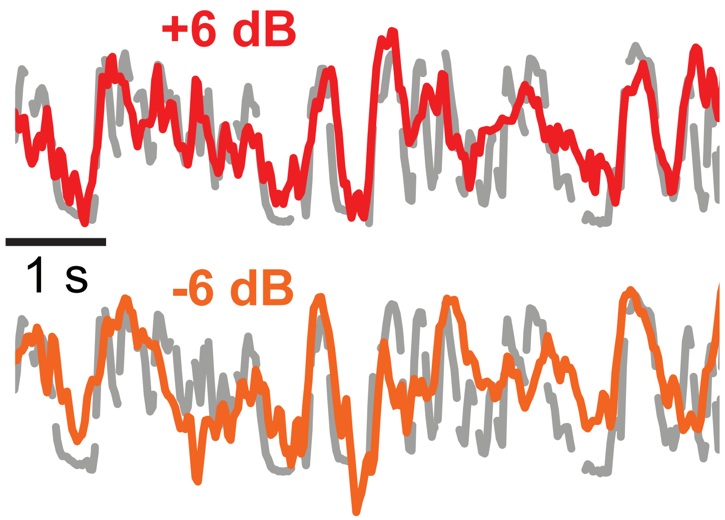
Speech in Noise: Results

Neural Reconstruction of
Underlying Speech Envelope



Ding & Simon, J Neuroscience (2013)

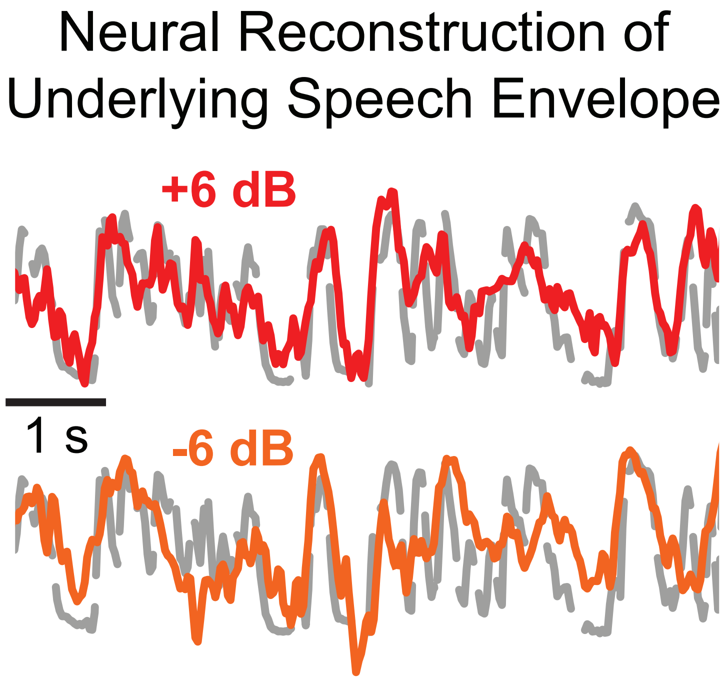
Speech in Noise: Results

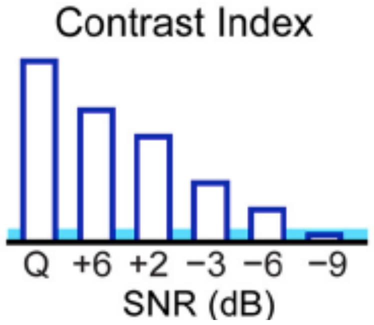


Neural Reconstruction of Underlying Speech Envelope

Ding & Simon, J Neuroscience (2013)

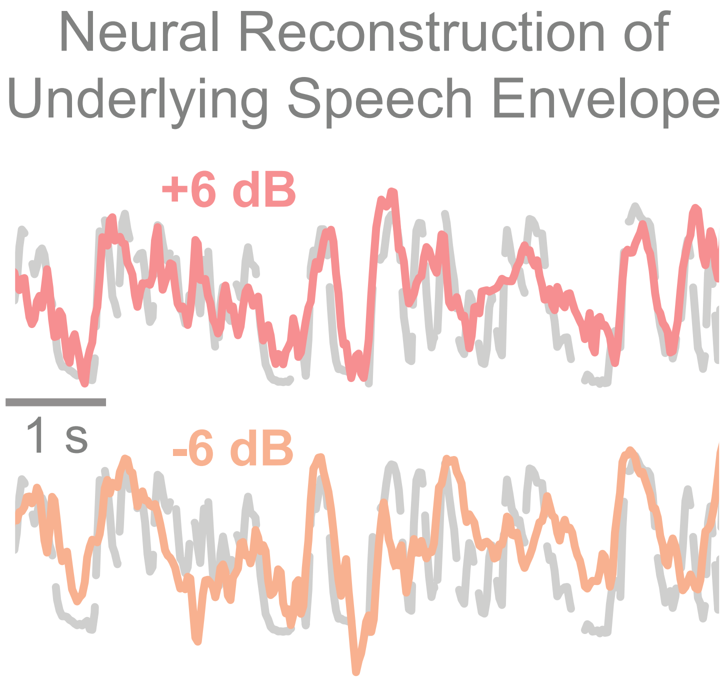
Speech in Noise: Results

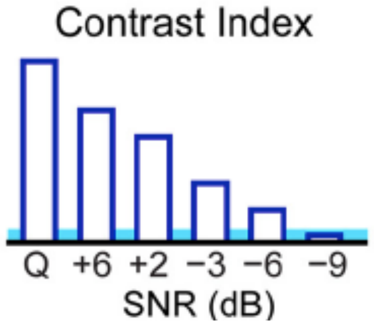


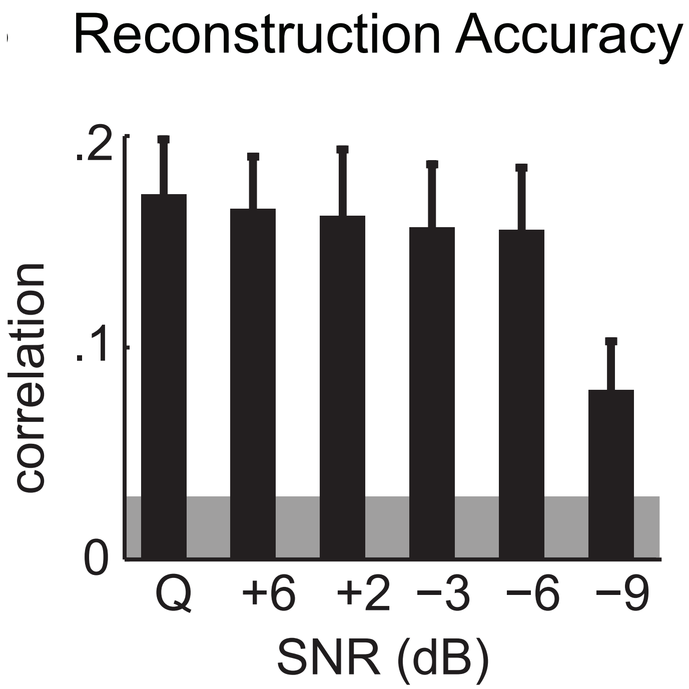


Ding & Simon, J Neuroscience (2013)

Speech in Noise: Results

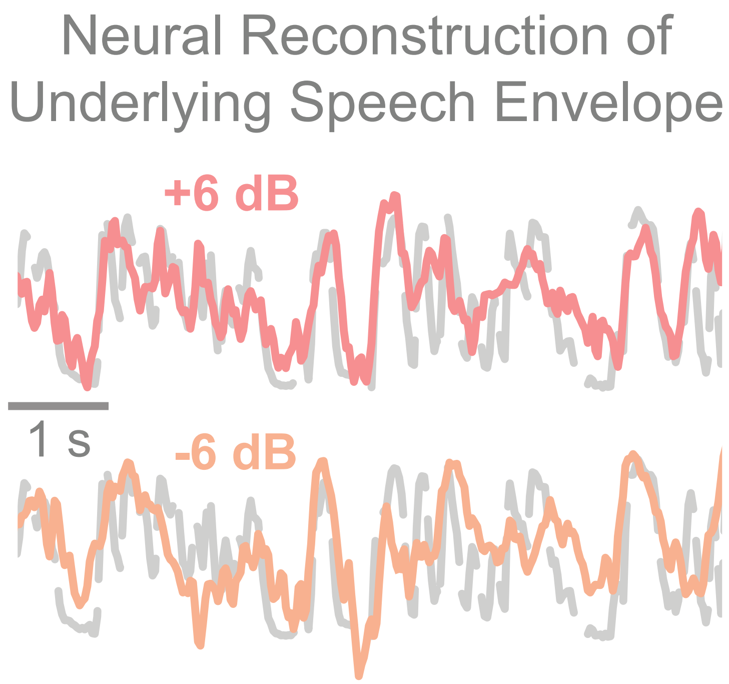

Neural Reconstruction of Underlying Speech Envelope

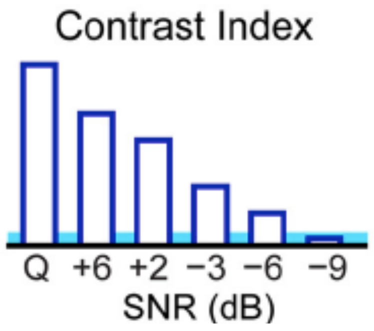

Contrast Index

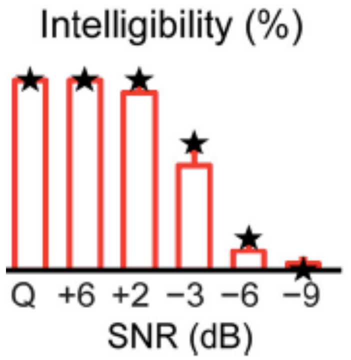

Reconstruction Accuracy

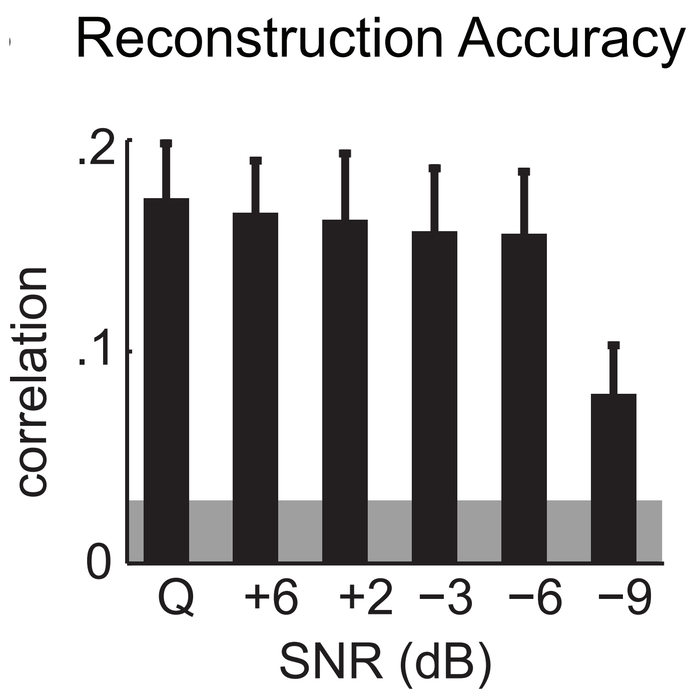
Ding & Simon, J Neuroscience (2013)

Speech in Noise: Results



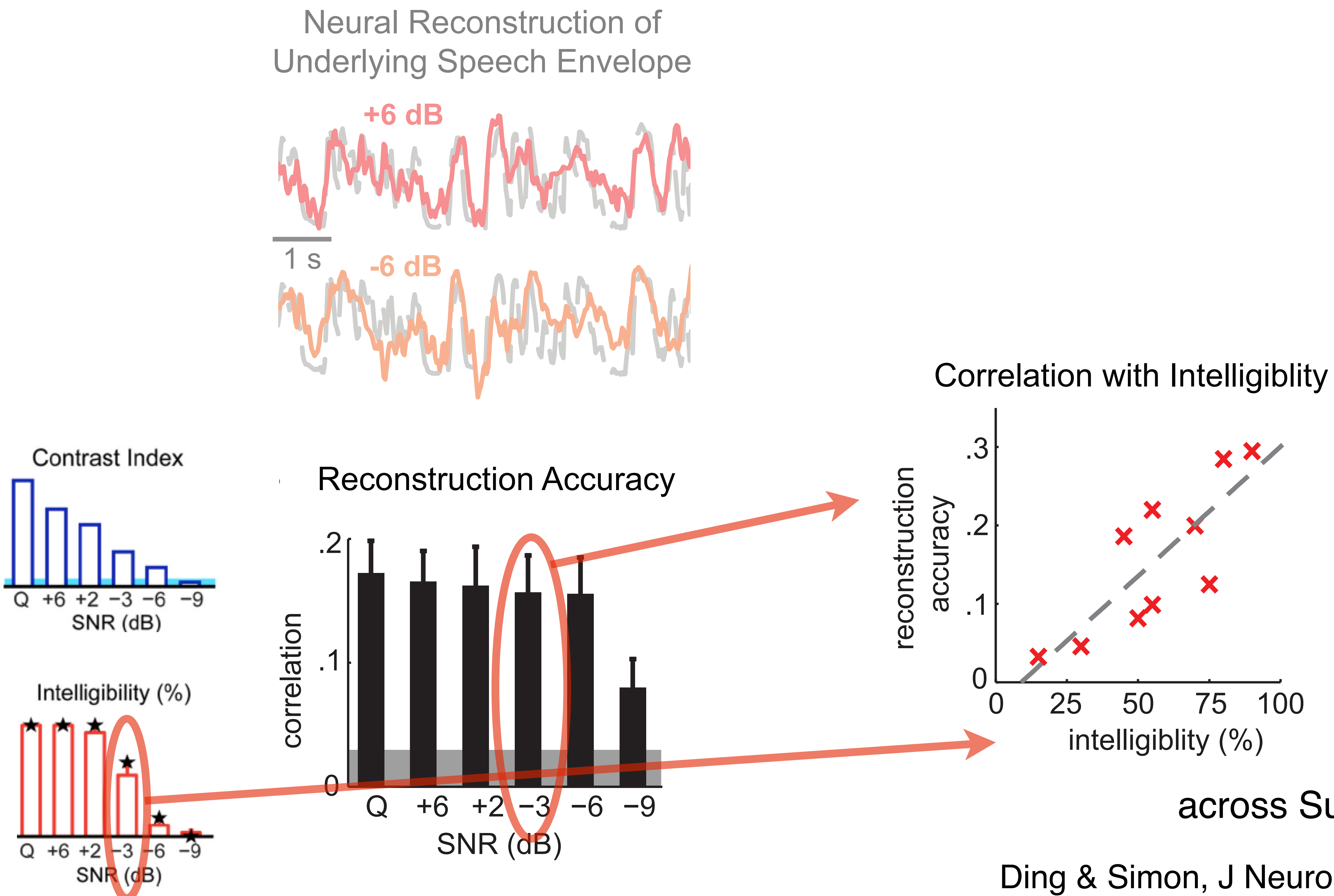






Ding & Simon, J Neuroscience (2013)

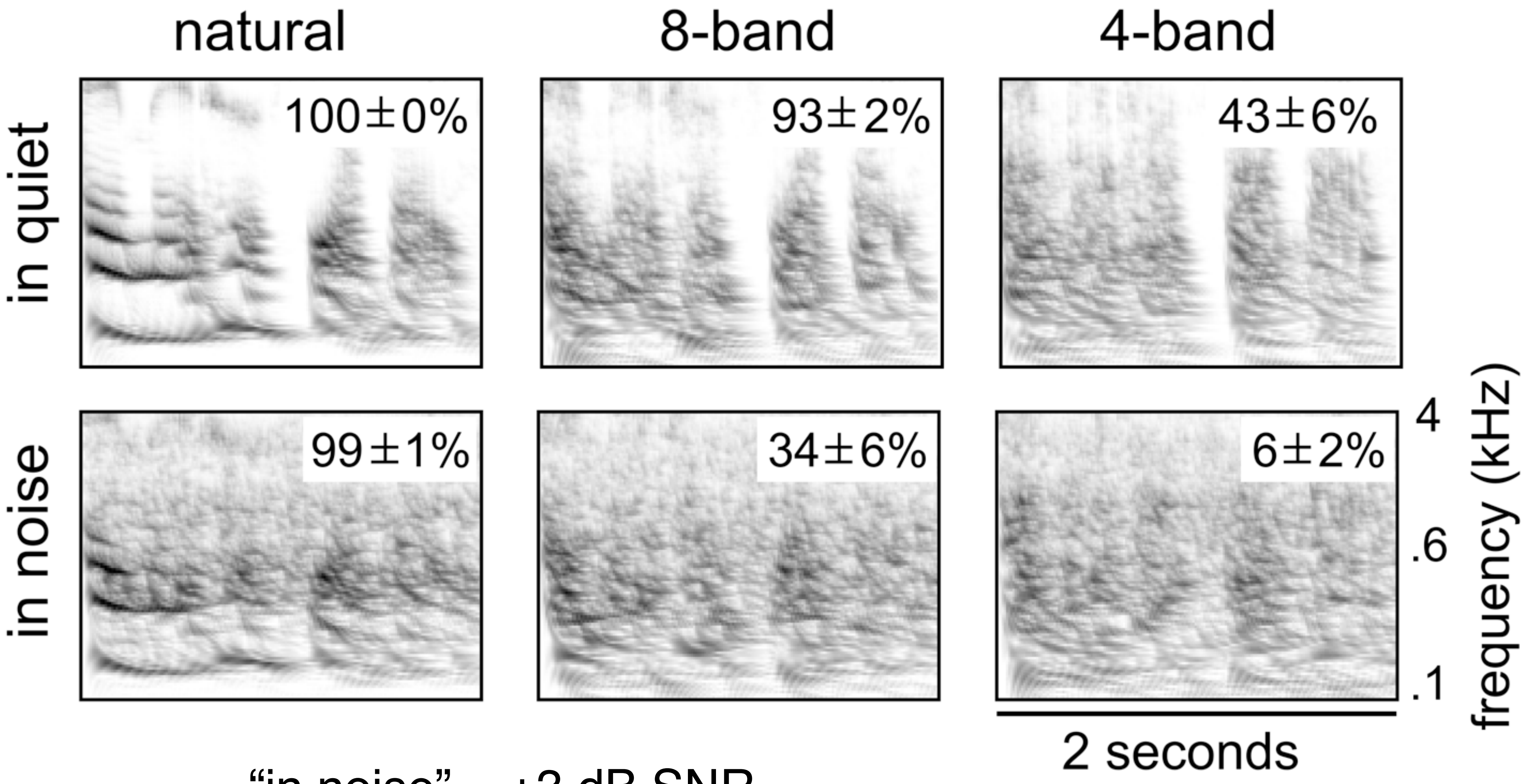
Speech in Noise: Results



across Subjects

Ding & Simon, J Neuroscience (2013)

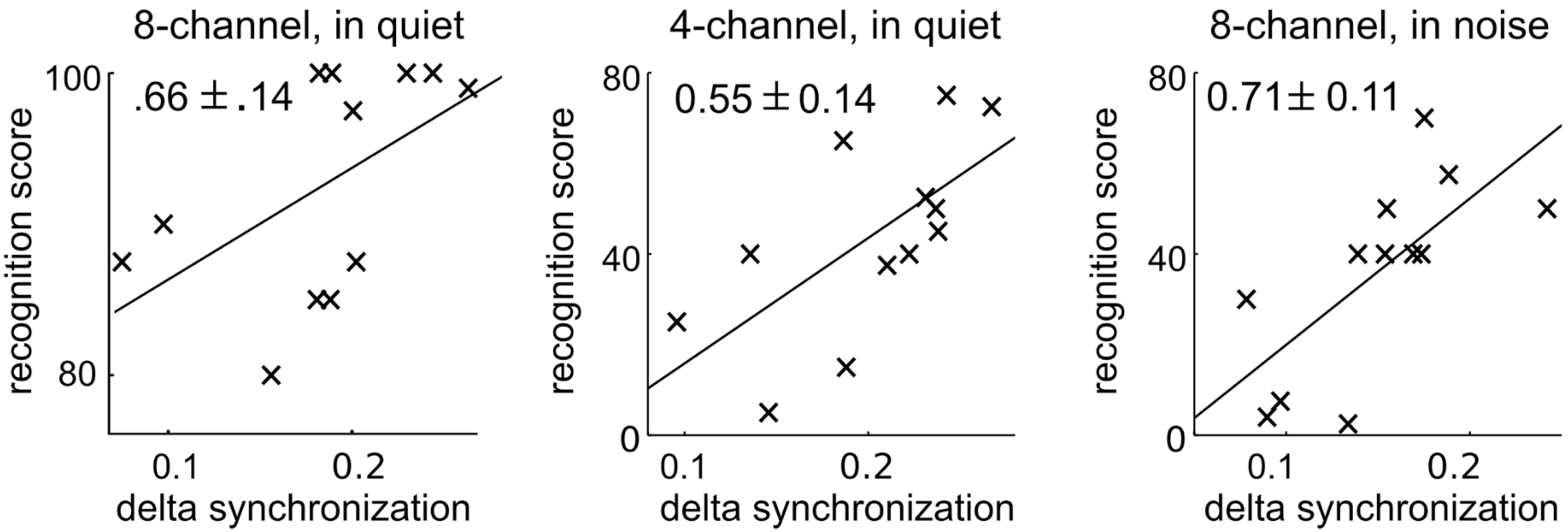
Noise-Vocoded Speech



“in noise” = +3 dB SNR

Ding, Chatterjee & Simon, NeuroImage (2014)

Noise-Vocoded Speech: Results



Multiple 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)

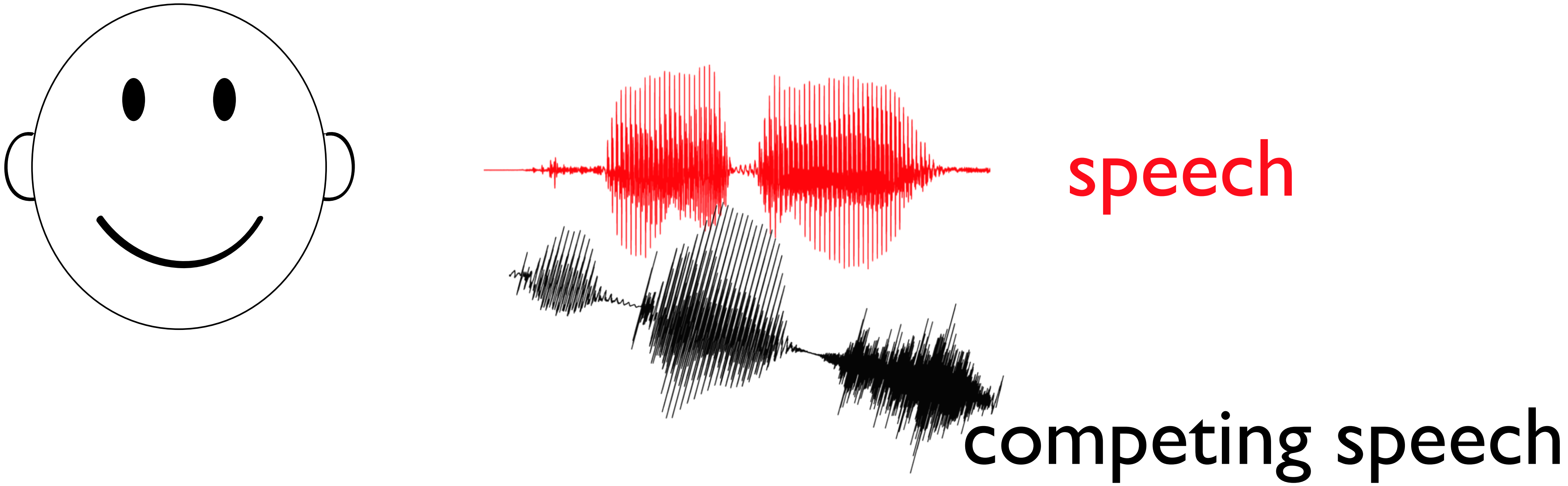
Listening to Speech at the Cocktail Party

Alex Katz,
The Cocktail Party

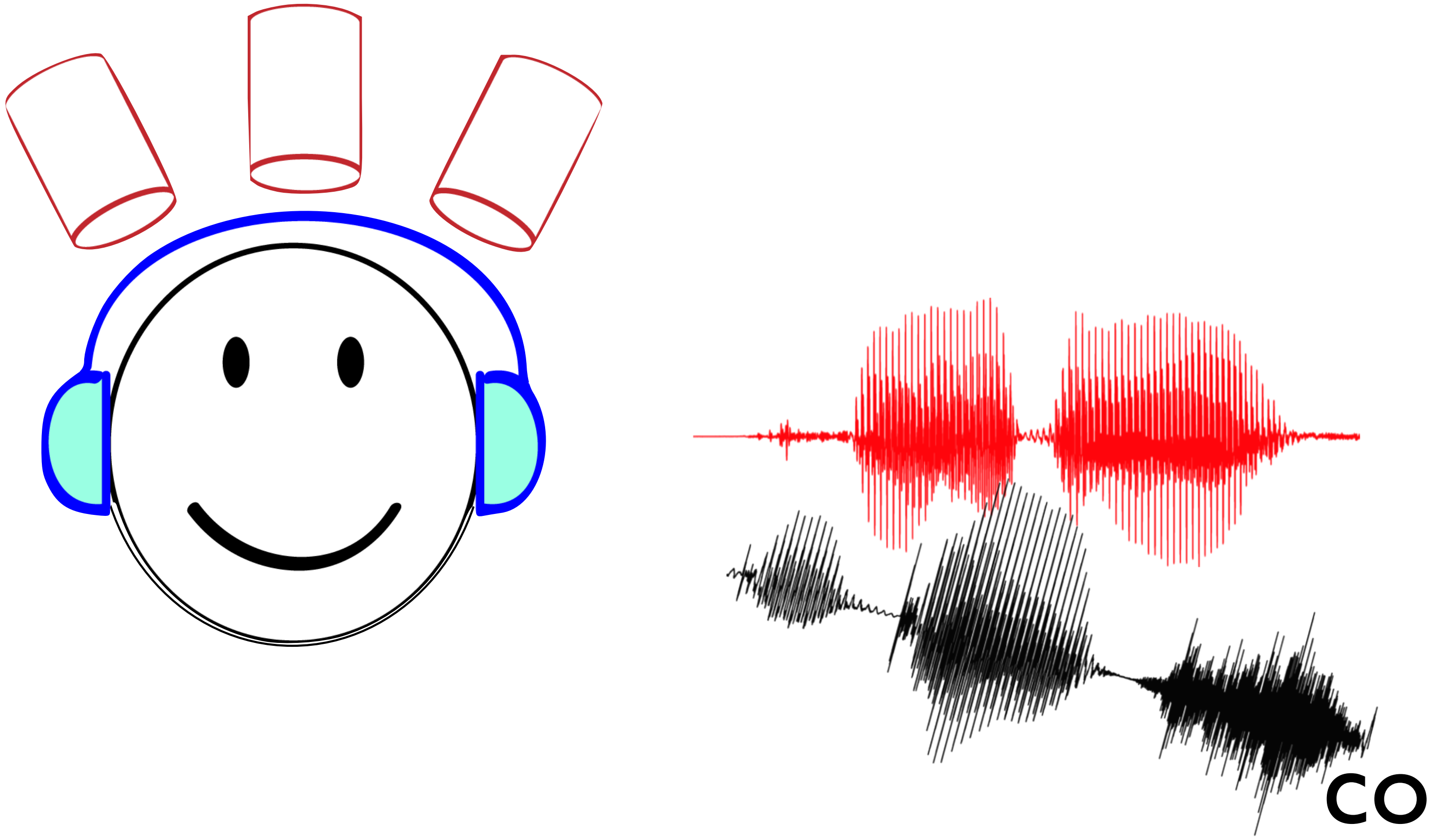
Listening to Speech at the Cocktail Party

Alex Katz,
The Cocktail Party

Experiments



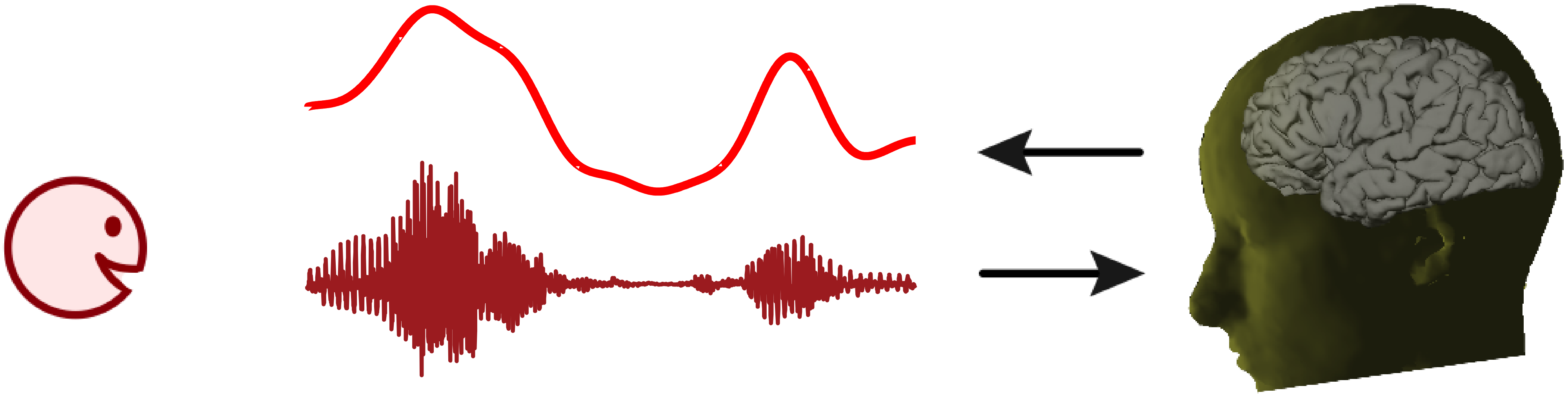
Experiments



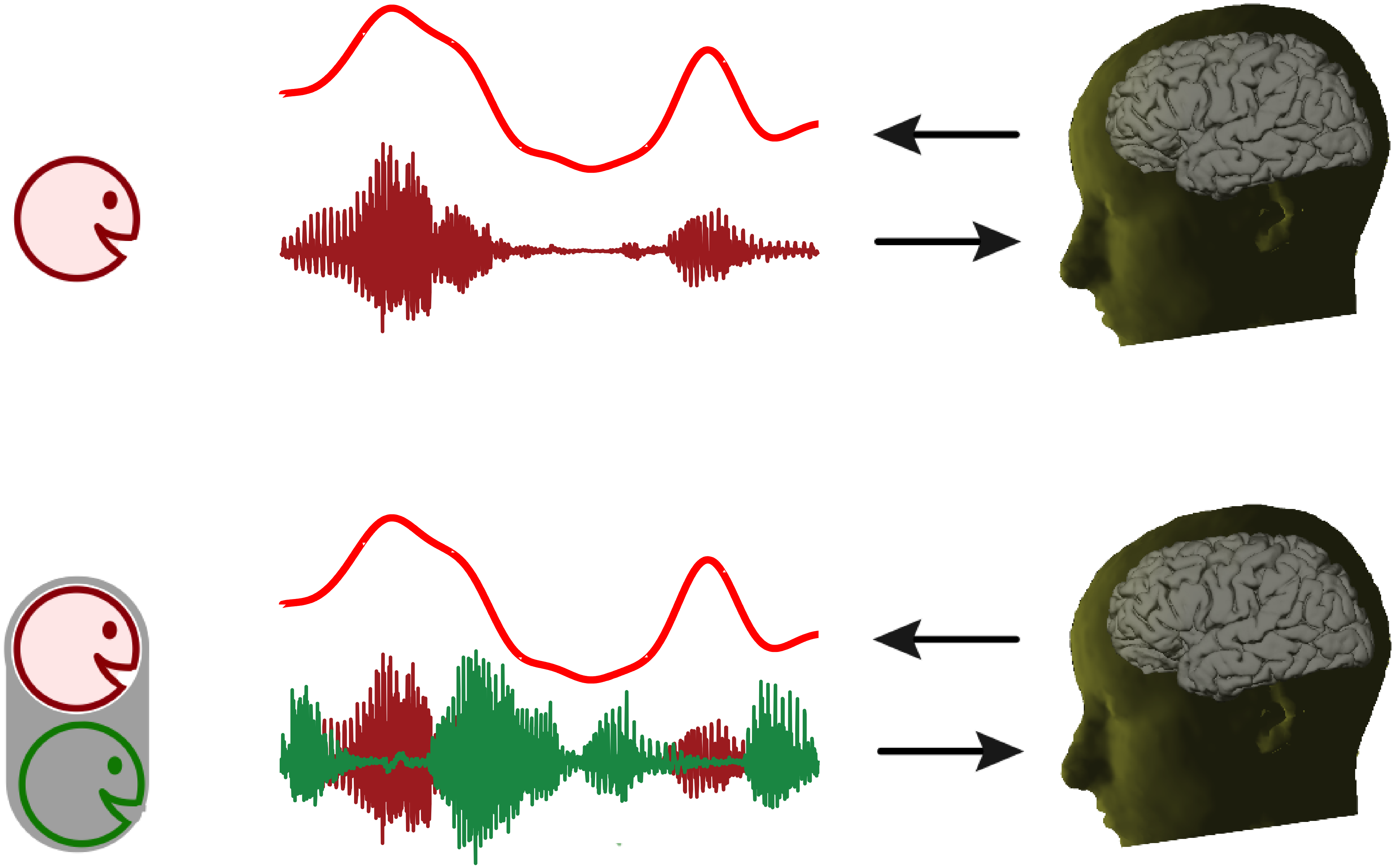
speech

competing speech

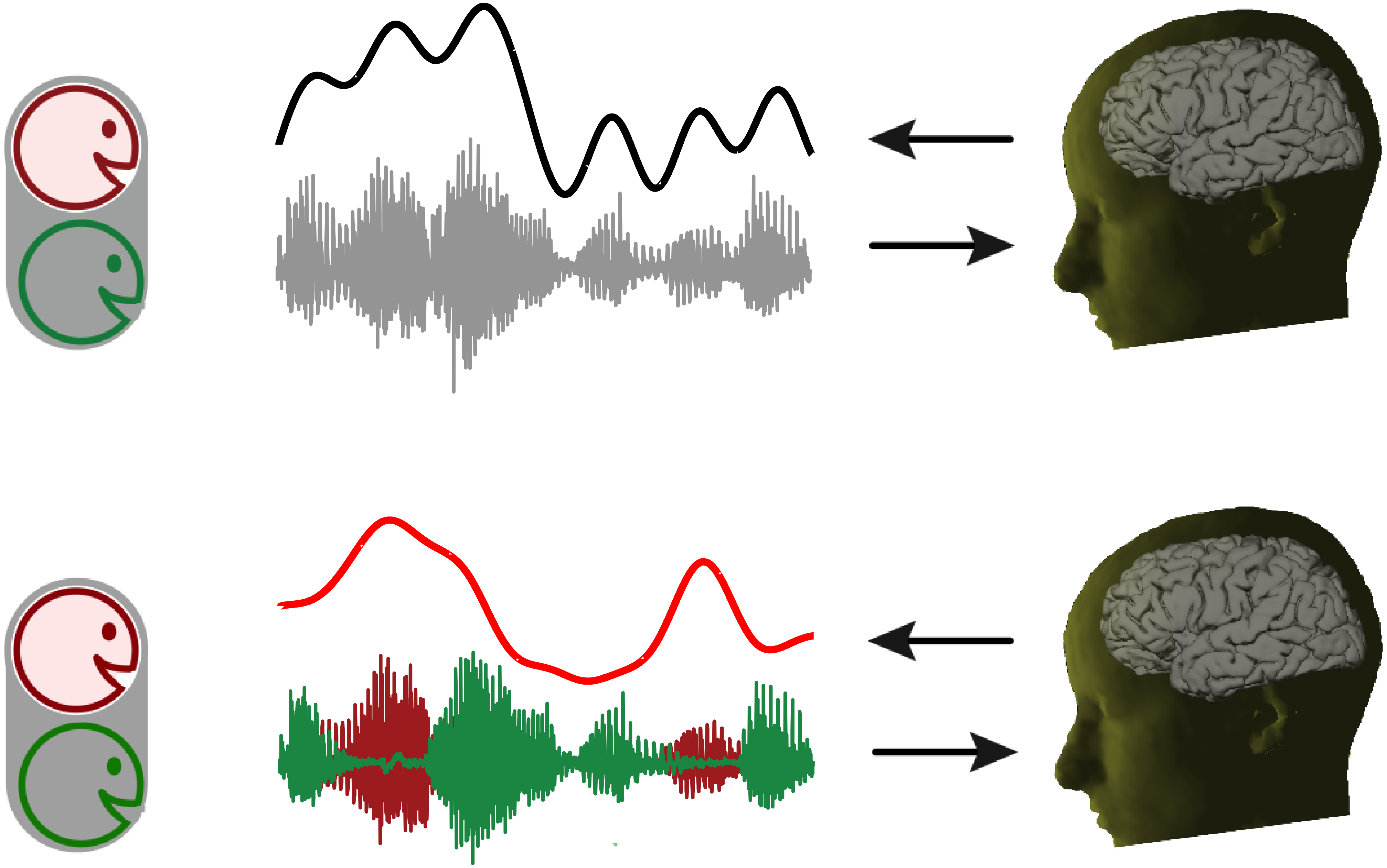
Selective Neural Encoding



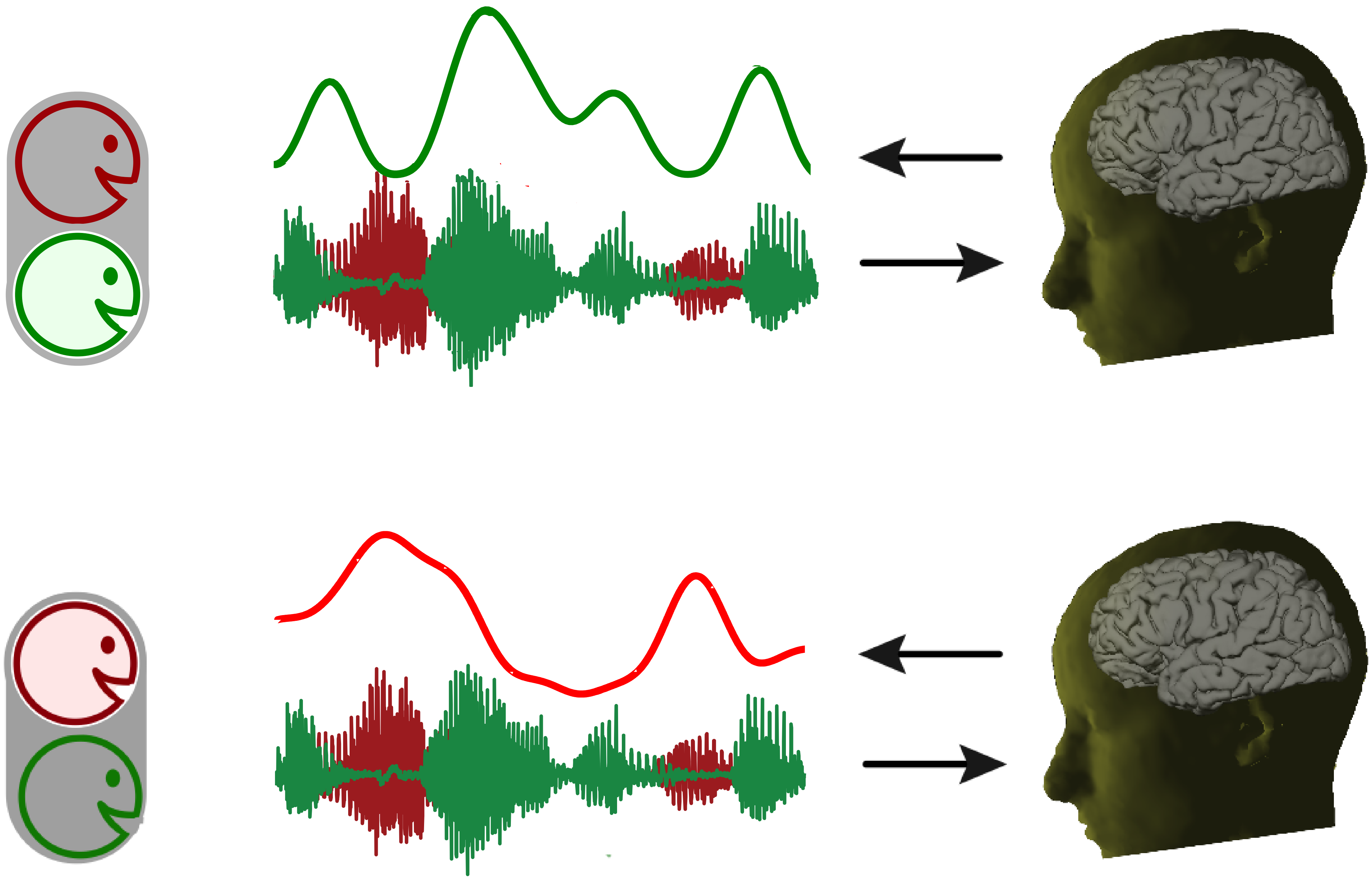
Selective Neural Encoding



Unselective vs. Selective Neural Encoding



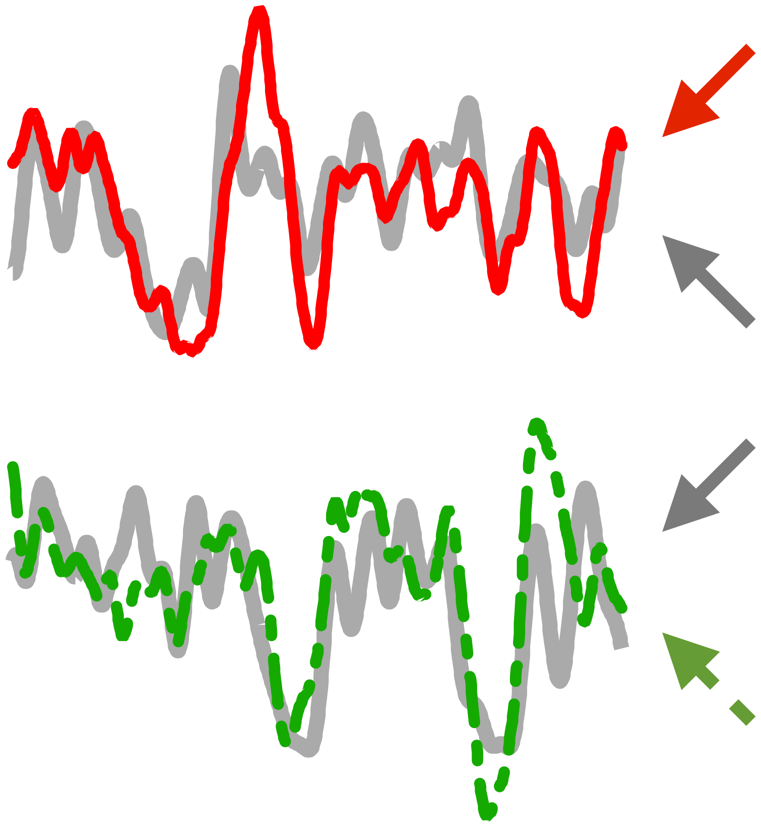
Selective Neural Encoding



Stream-Specific Representation

attending to
speaker 1

attending to
speaker 2



reconstructed
from MEG

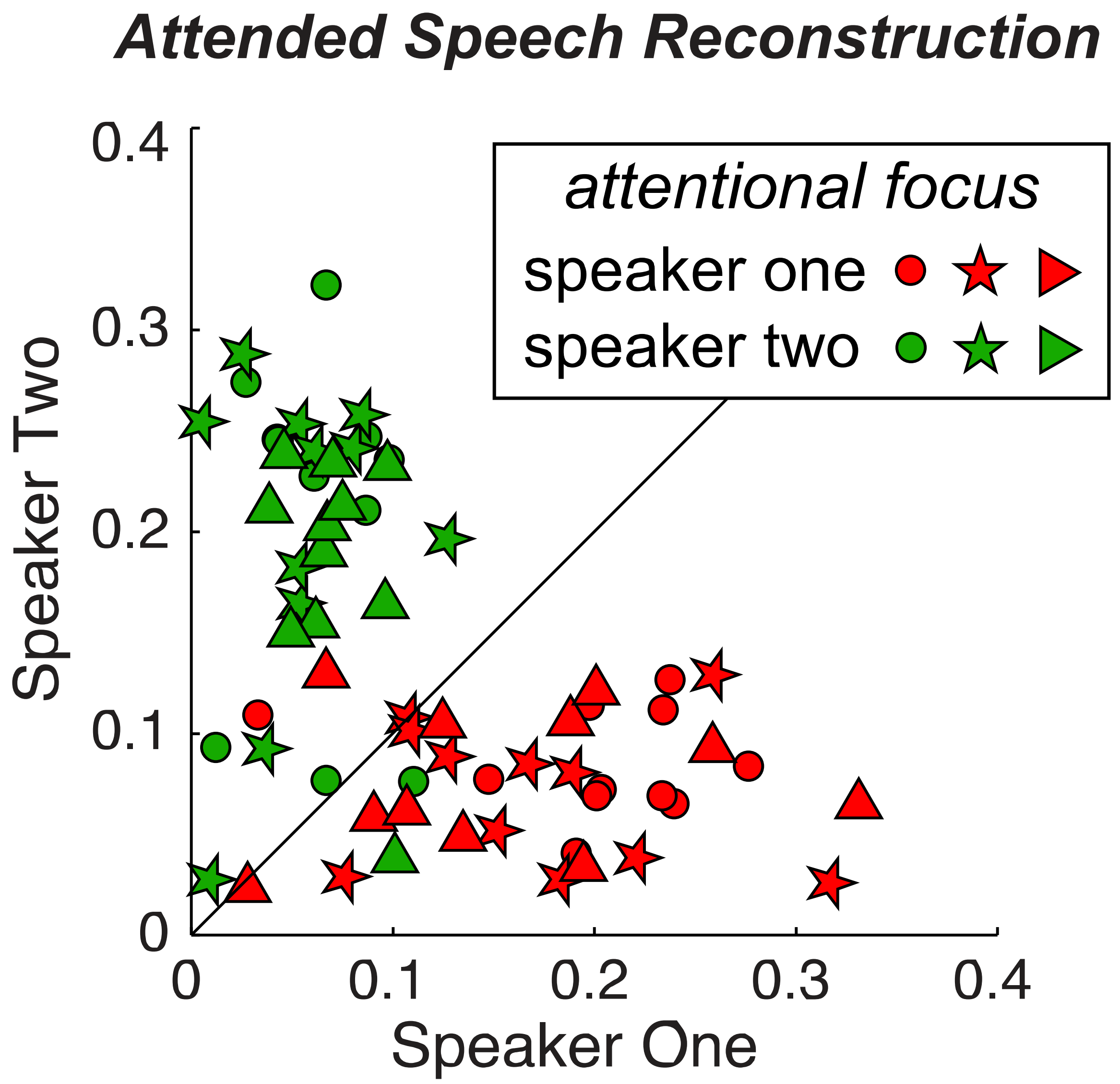
attended speech
envelopes

reconstructed
from MEG

Identical Stimuli!

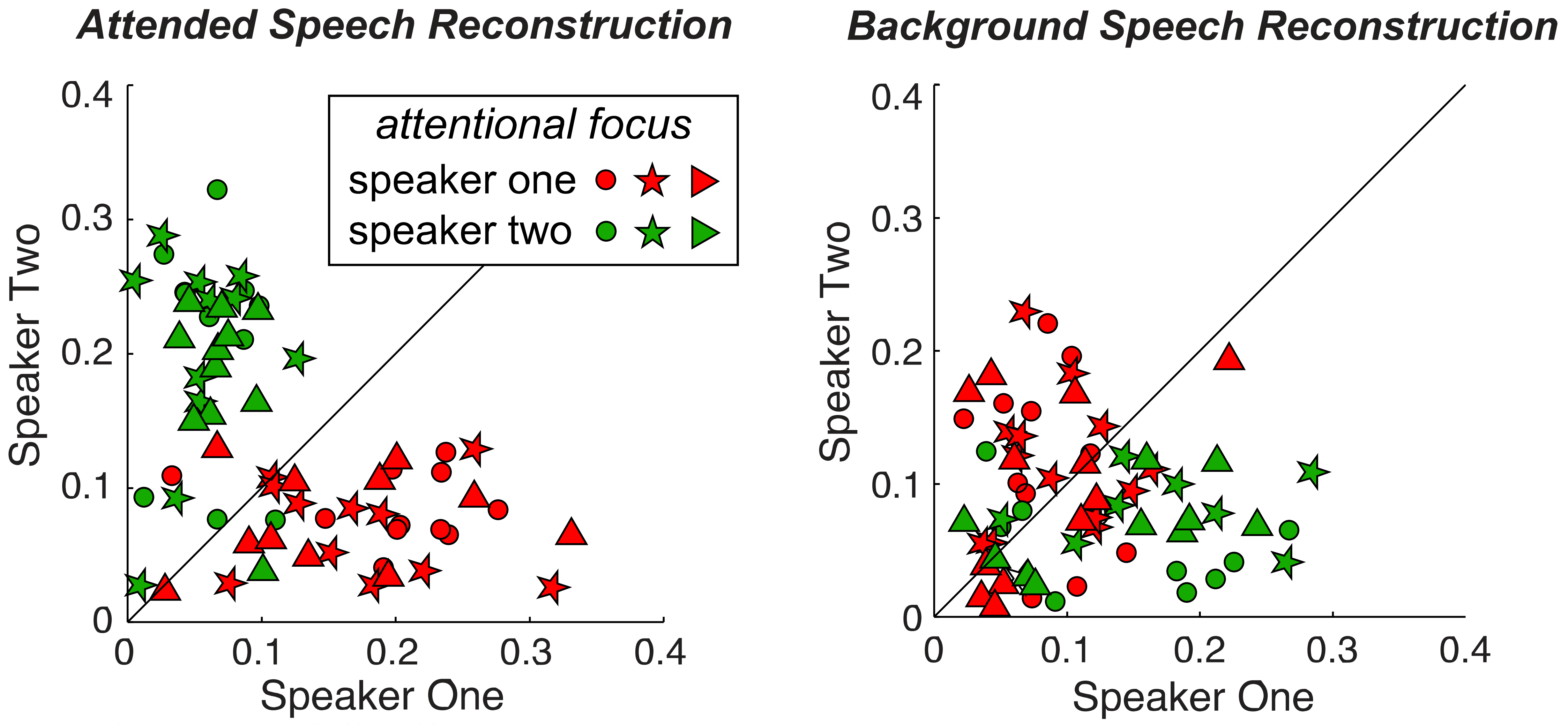
Ding & Simon, PNAS (2012)

Single Trial Speech Reconstruction



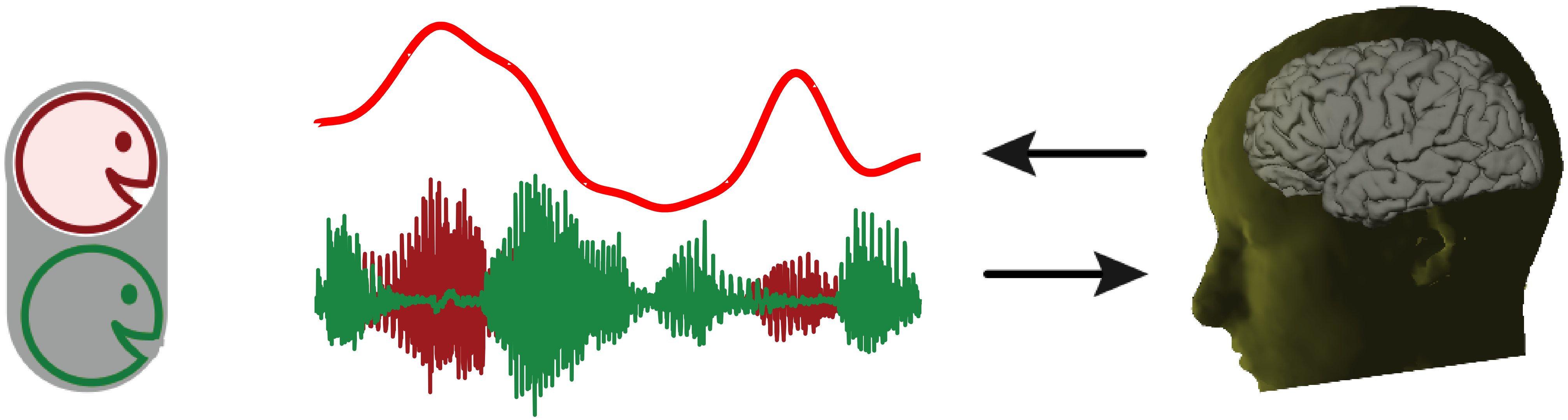
Ding & Simon, PNAS (2012)

Single Trial Speech Reconstruction

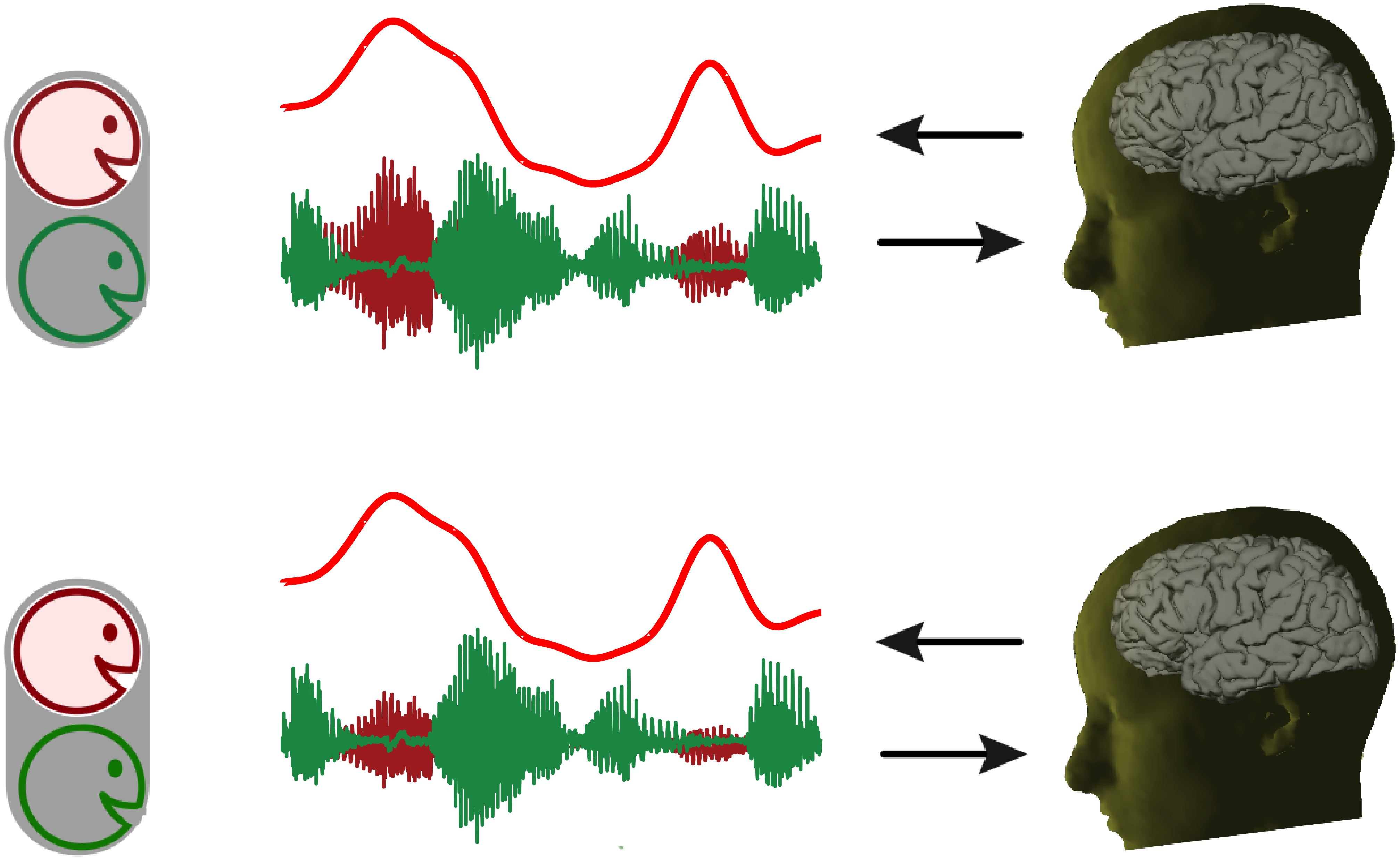


Ding & Simon, PNAS (2012)

Invariance Under Relative Loudness Change?

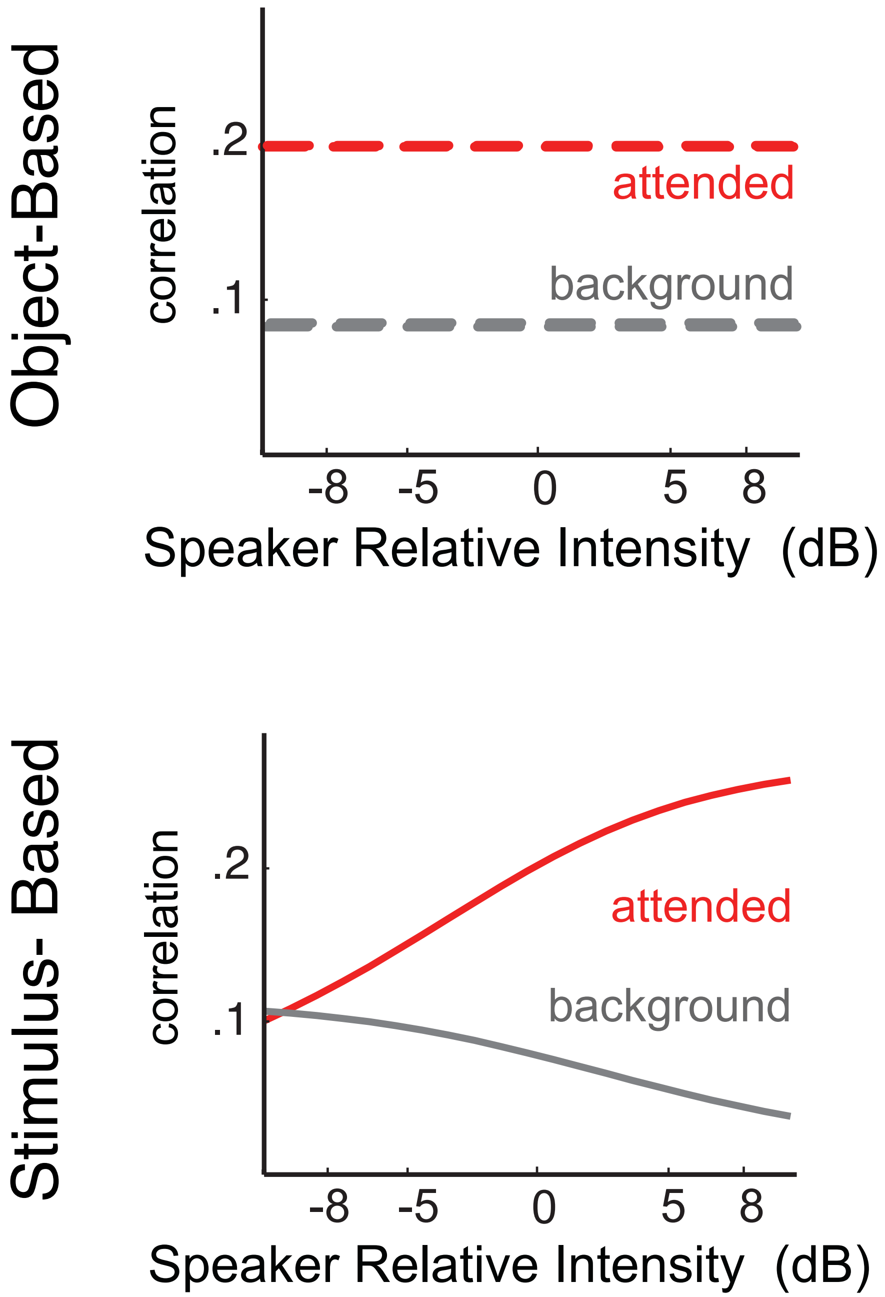


Invariance Under Relative Loudness Change?



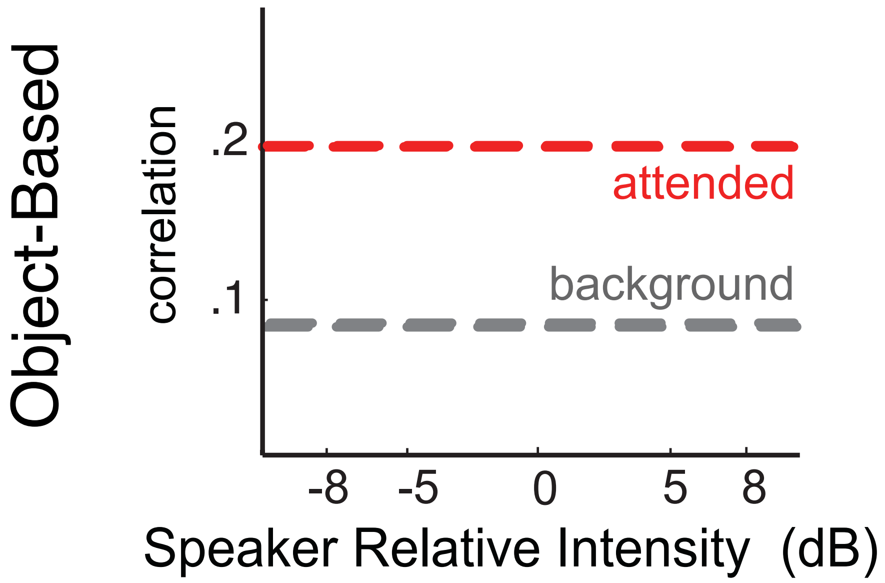
Stream-Based Gain Control?

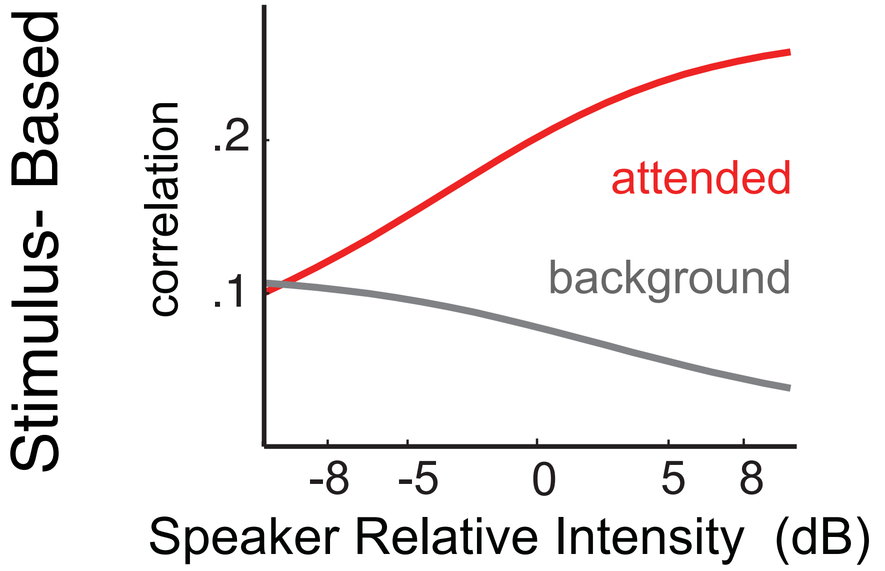
Gain-Control Models



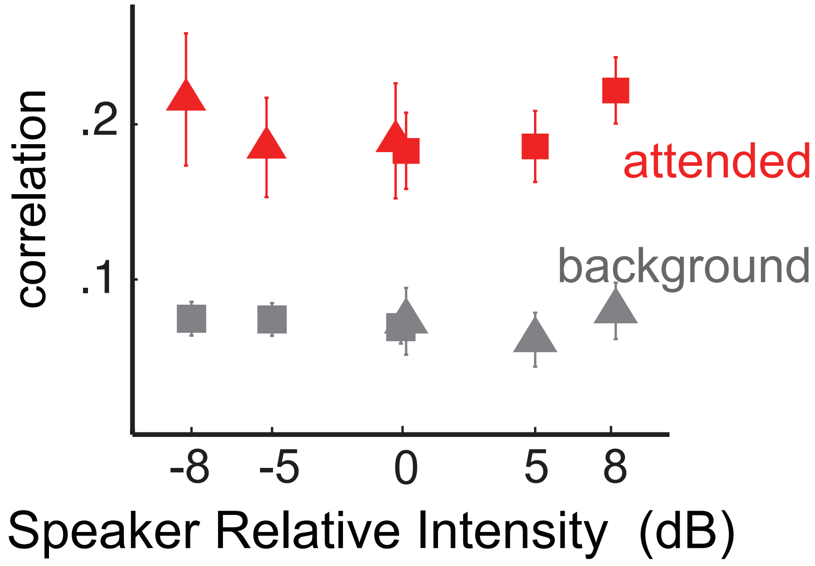
Stream-Based Gain Control?

Gain-Control Models



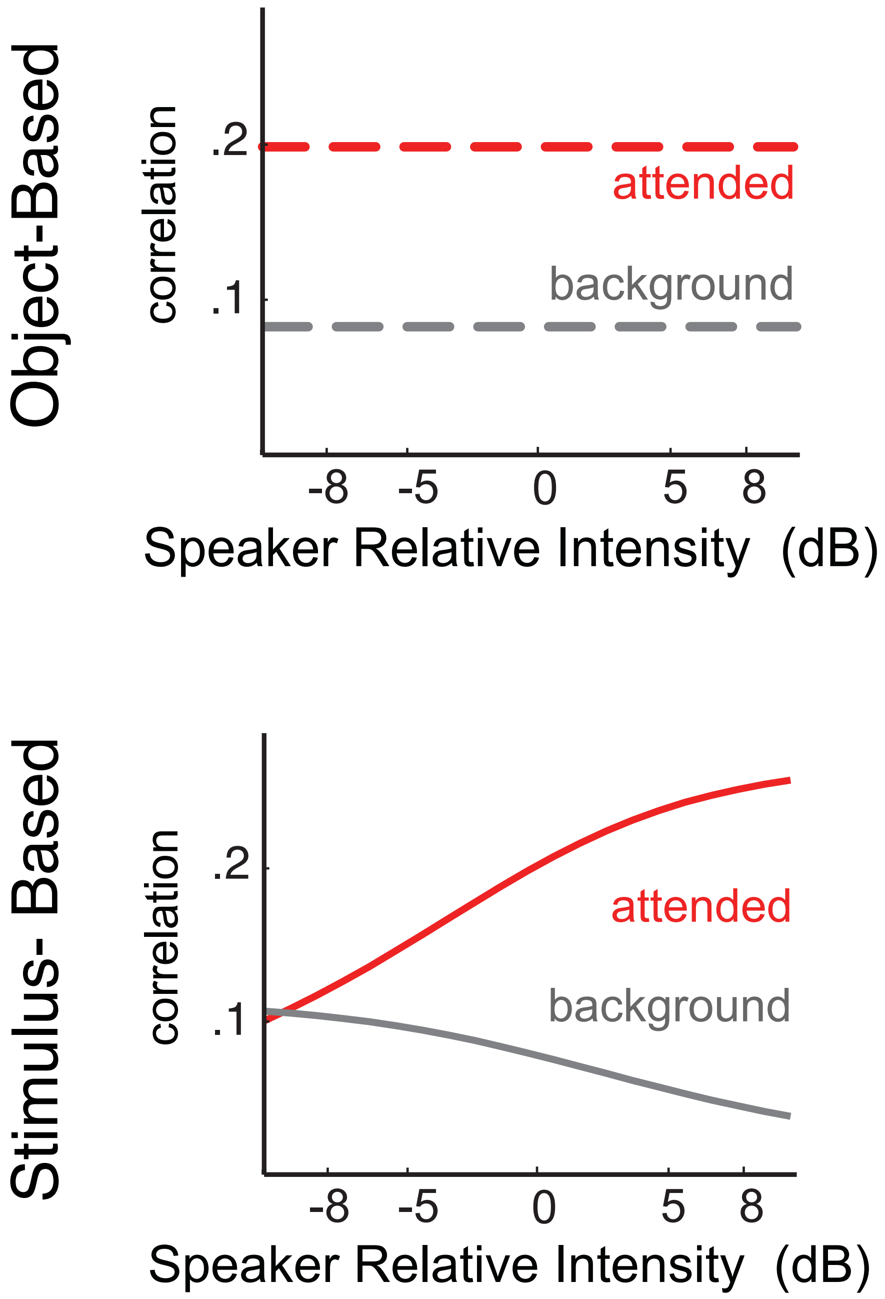


Neural Results

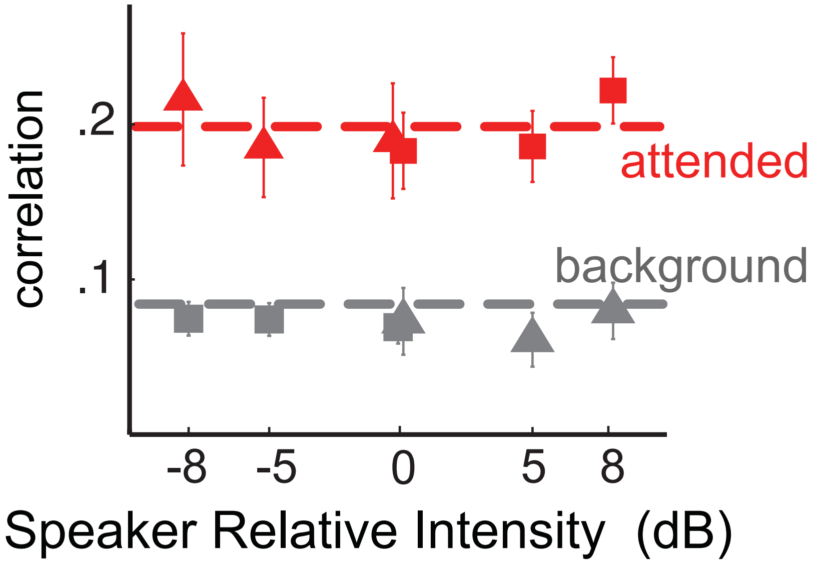


Stream-Based Gain Control?

Gain-Control Models

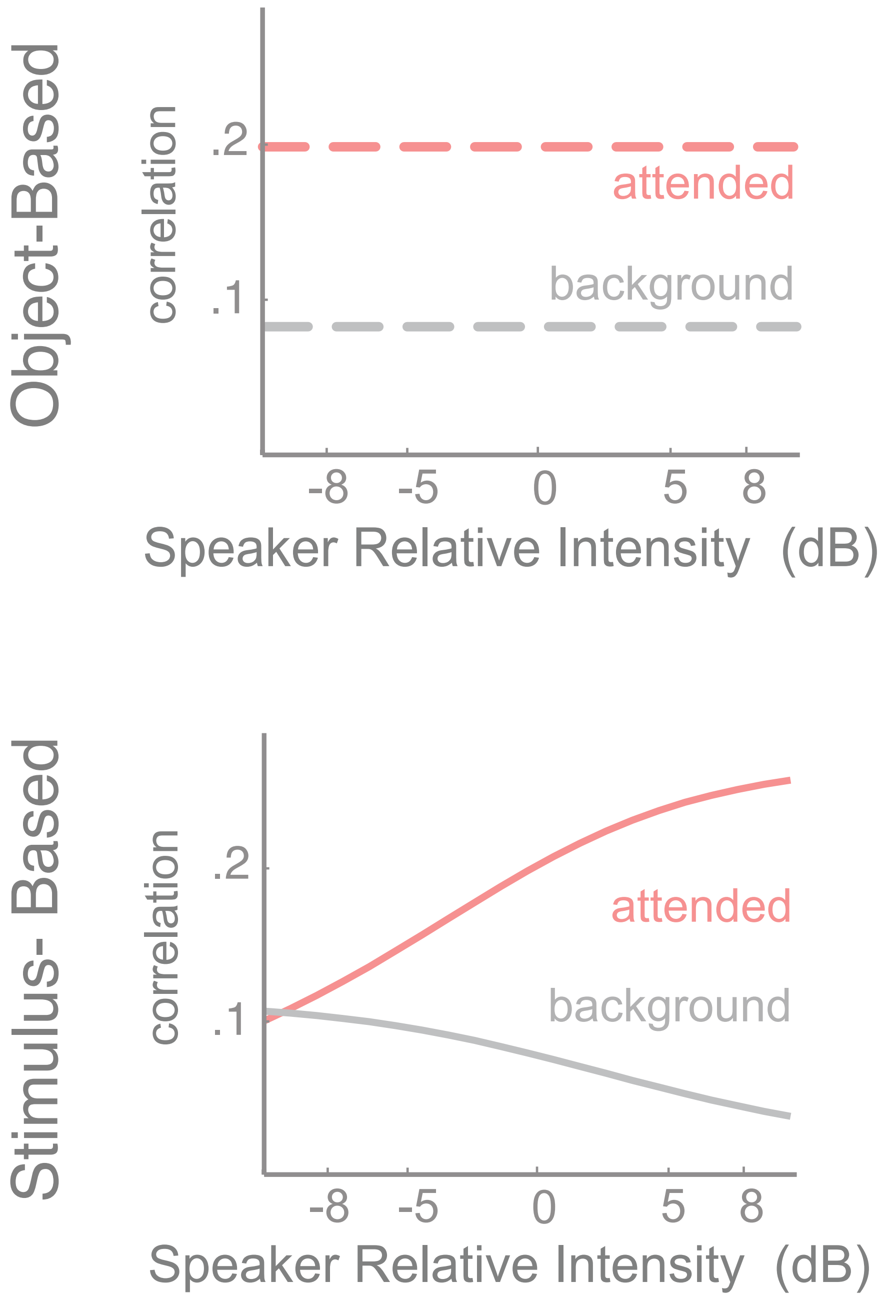


Neural Results

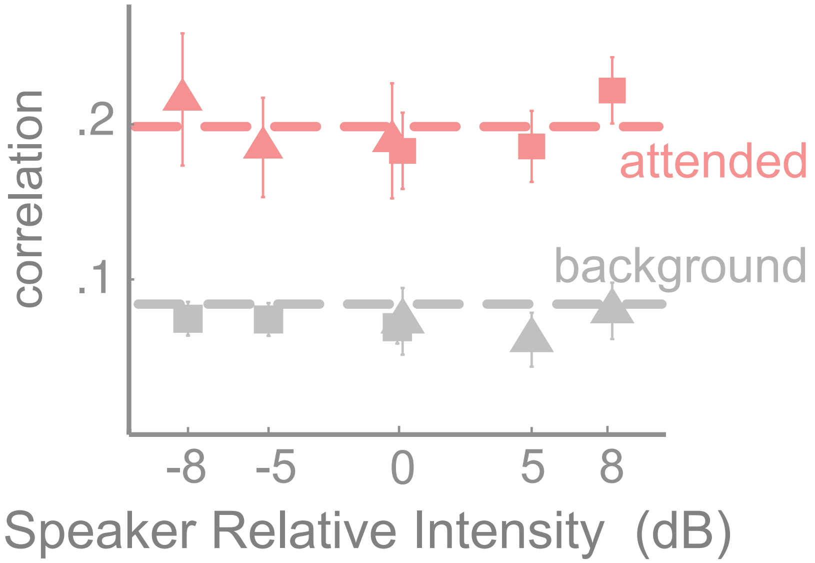


Stream-Based Gain Control?

Gain-Control Models

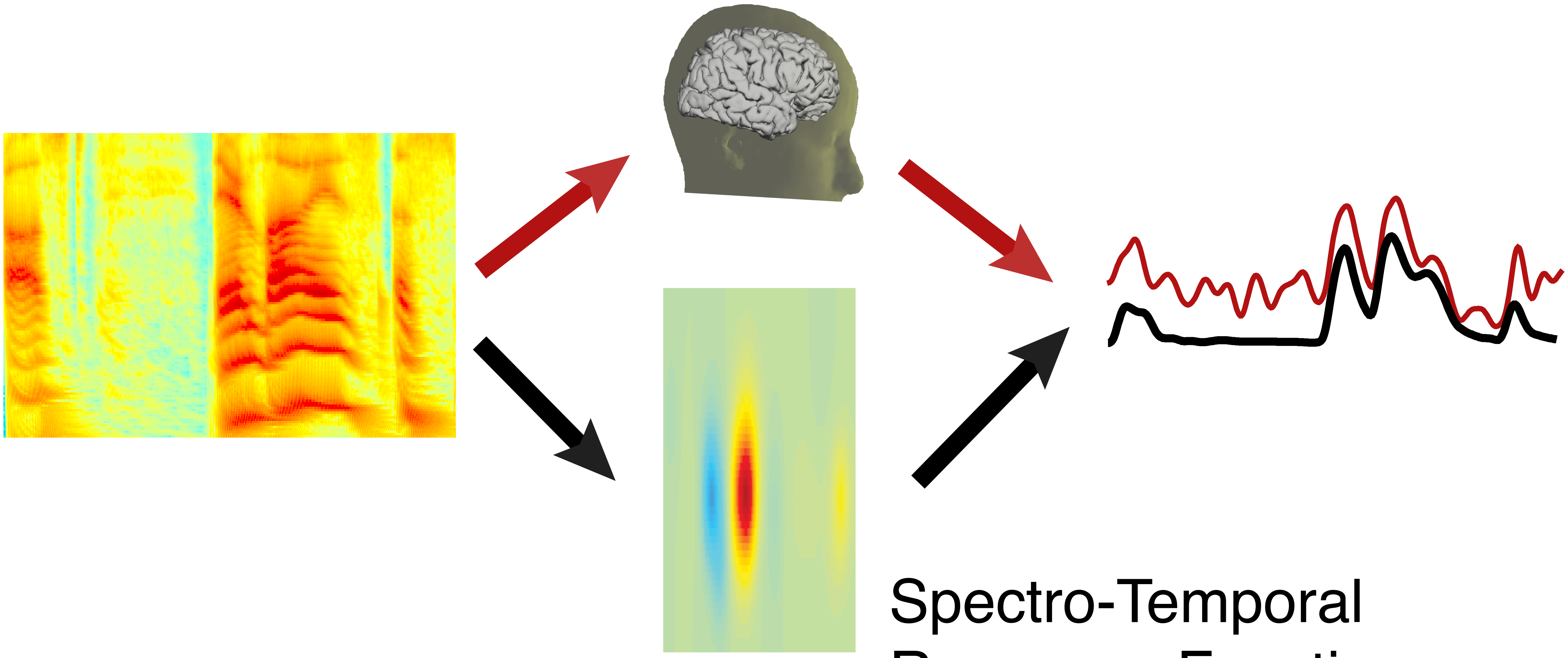


Neural Results



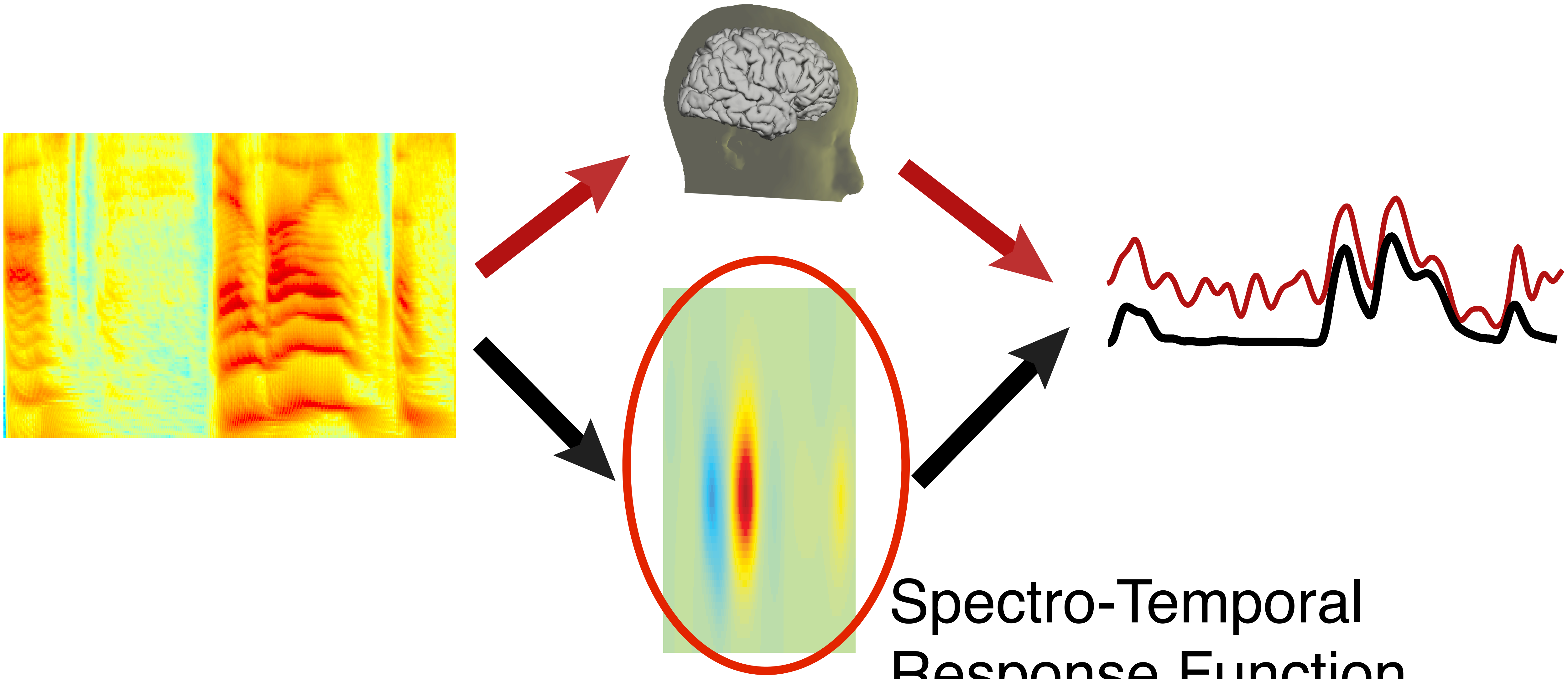
- Stream-based not stimulus-based
- Neural representation is invariant to acoustic changes.

Forward STRF Model



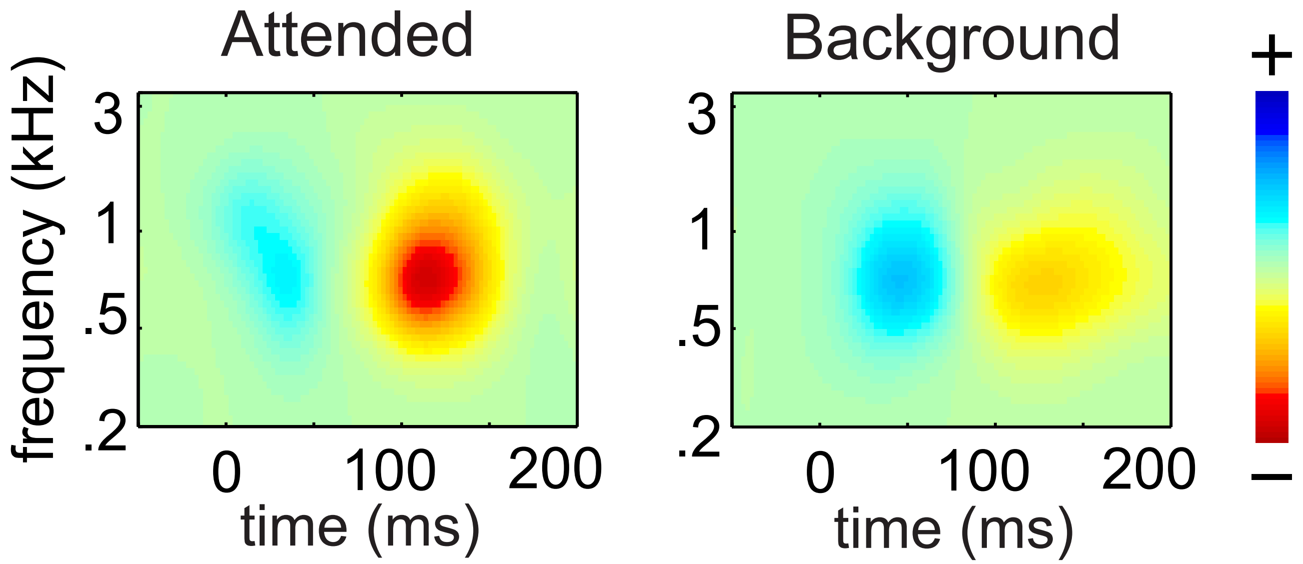
Spectro-Temporal
Response Function
(STRF)

Forward STRF Model



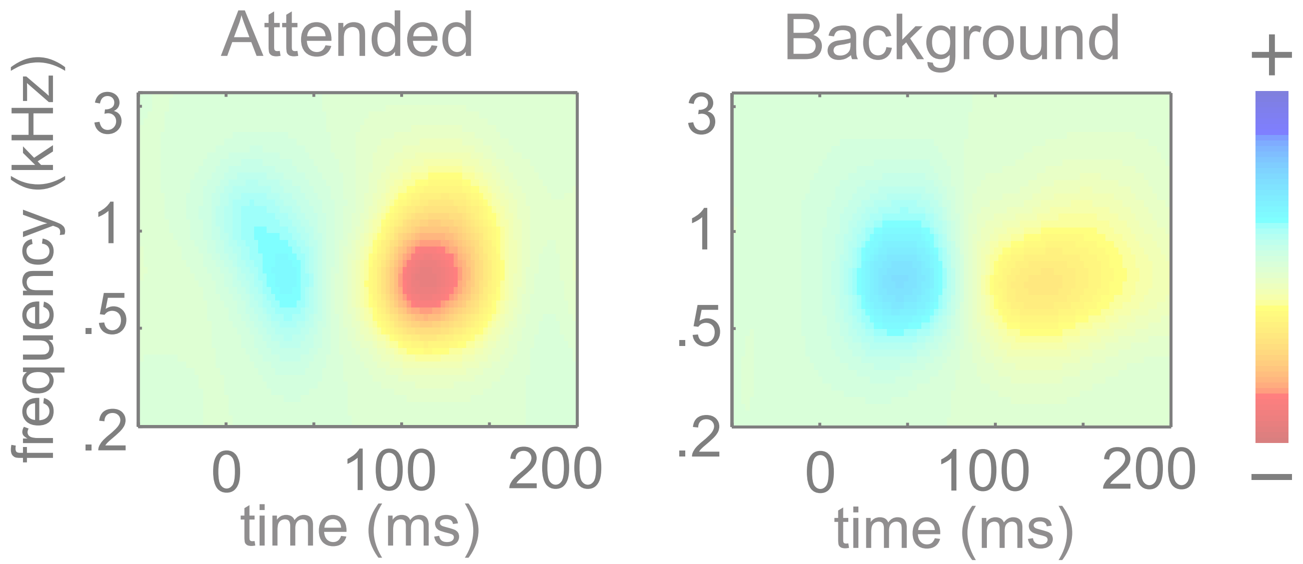
Spectro-Temporal Response Function (STRF)

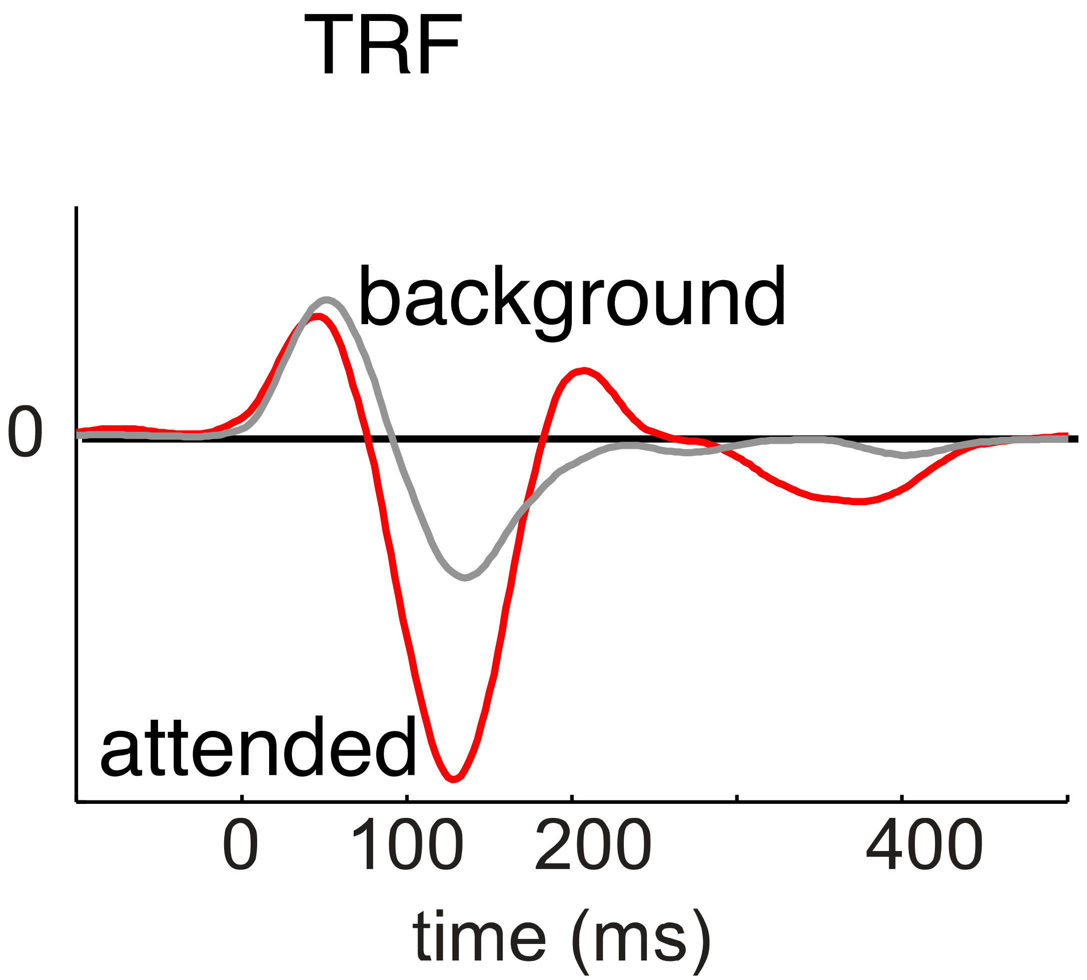
STRF Results



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

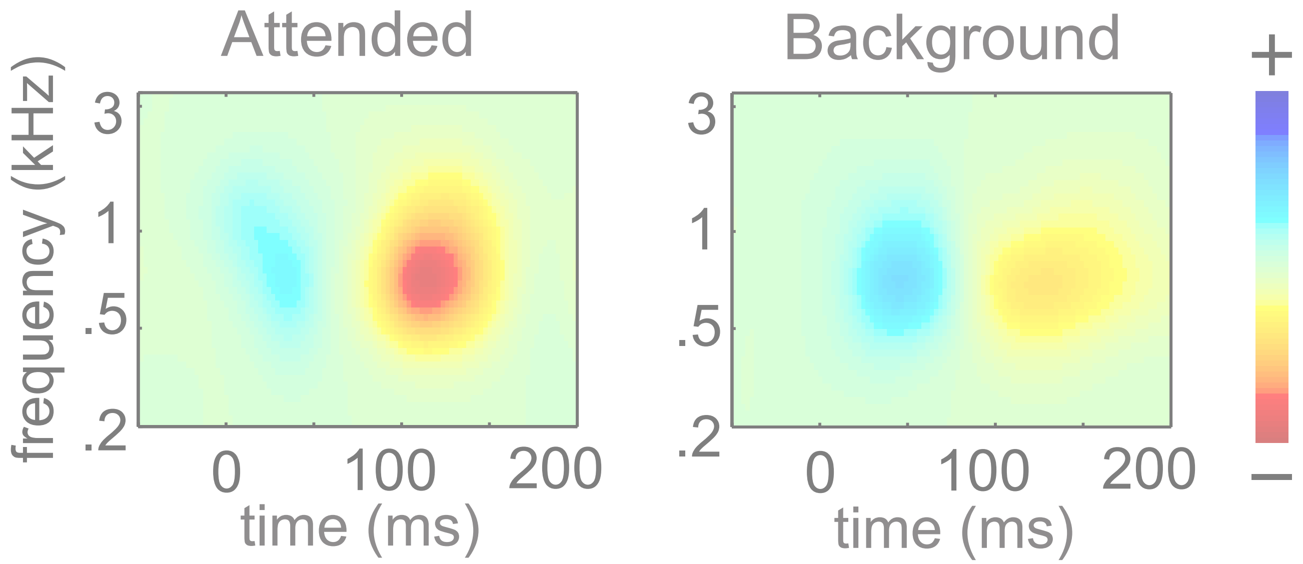
STRF Results

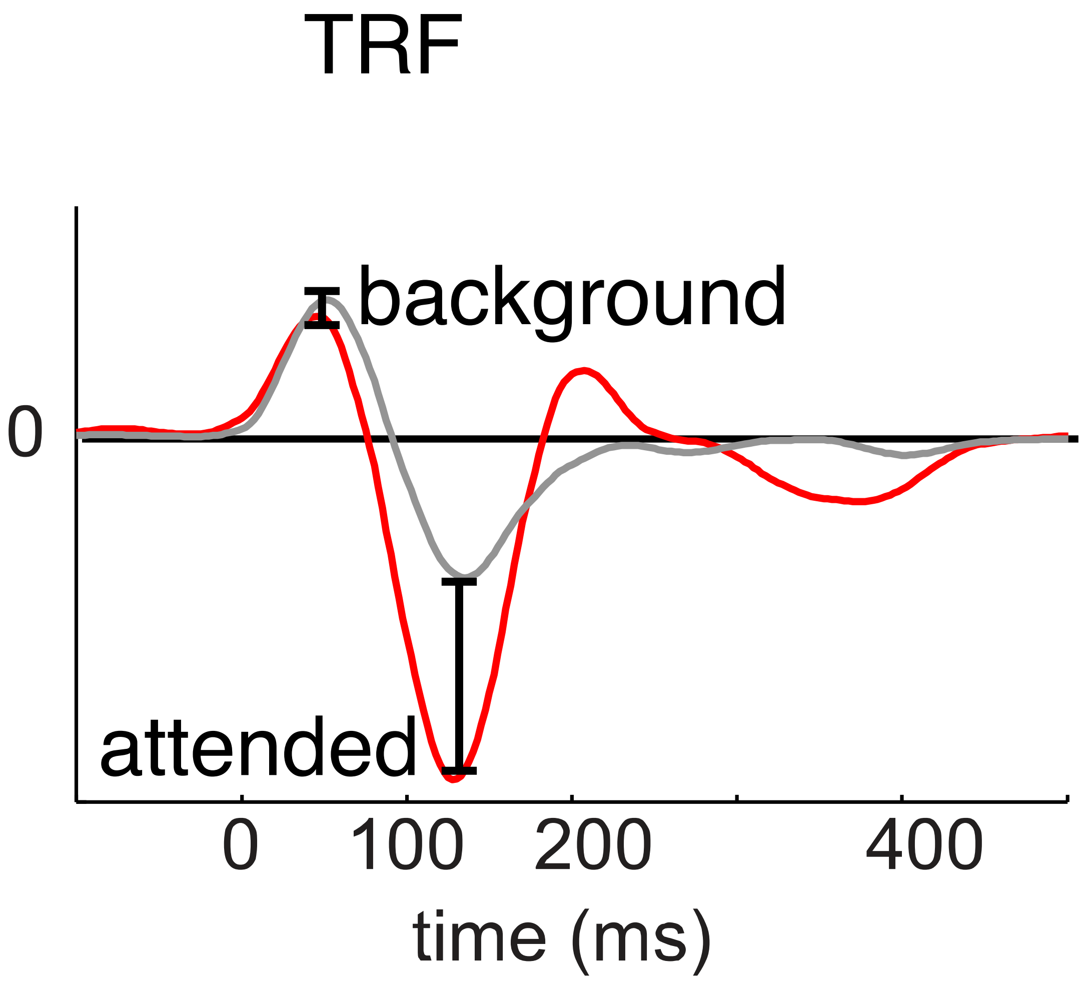




- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- M50_{STRF} positive peak
- $\text{M100}_{\text{STRF}}$ negative peak

STRF Results





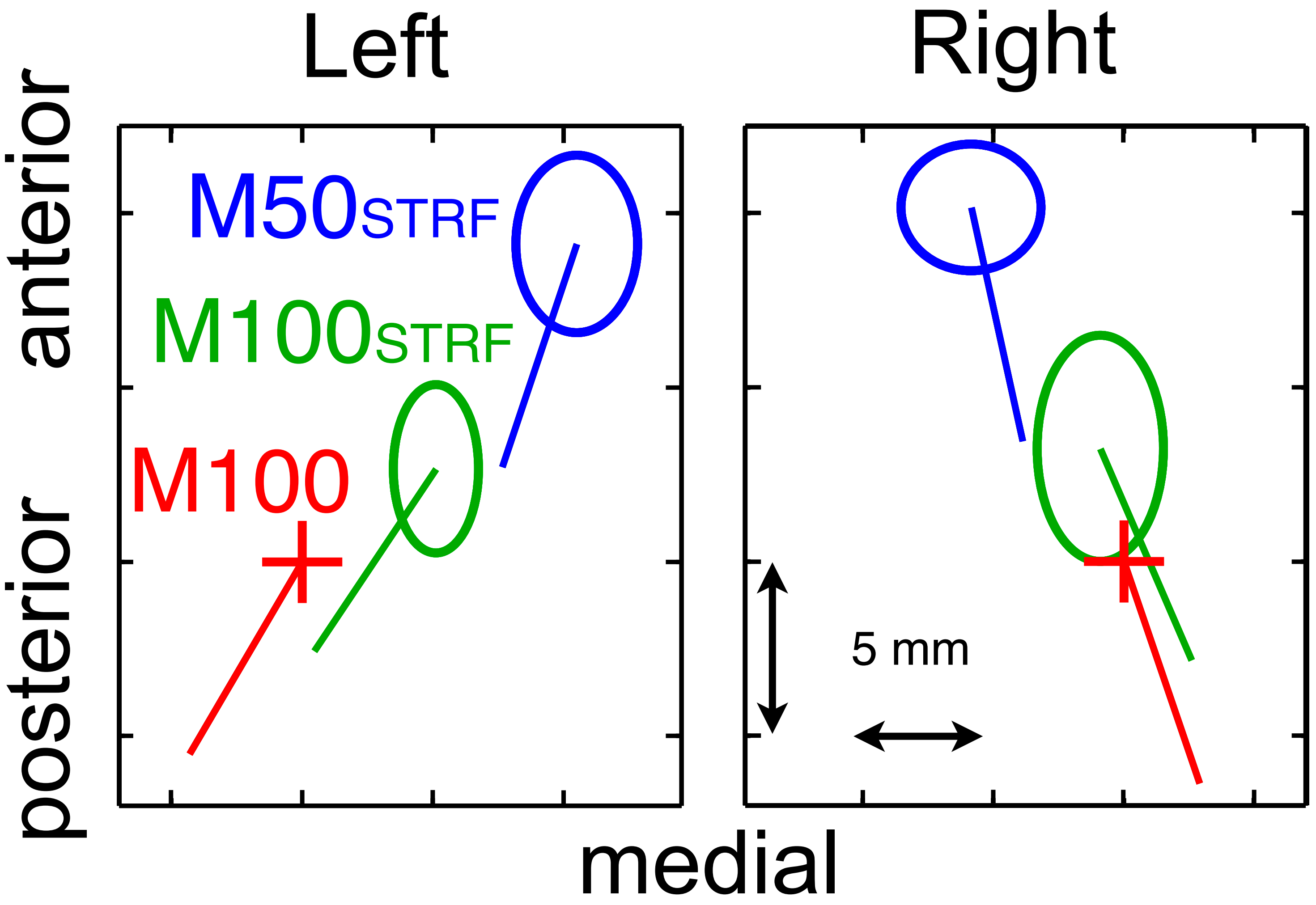
- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- M50$_{\text{STRF}}$ positive peak
- M100$_{\text{STRF}}$ negative peak
- **M100$_{\text{STRF}}$ strongly modulated by attention, *but not M50$_{\text{STRF}}$***

Neural Sources

- $\mathrm{M100}_{\mathrm{STRF}}$ source near (same as?) M100 source:
  Planum Temporale

- $\mathrm{M50}_{\mathrm{STRF}}$ source is anterior and medial to M100 (same as M50?):
  Heschl's Gyrus

- **PT strongly modulated by attention, *but not HG***



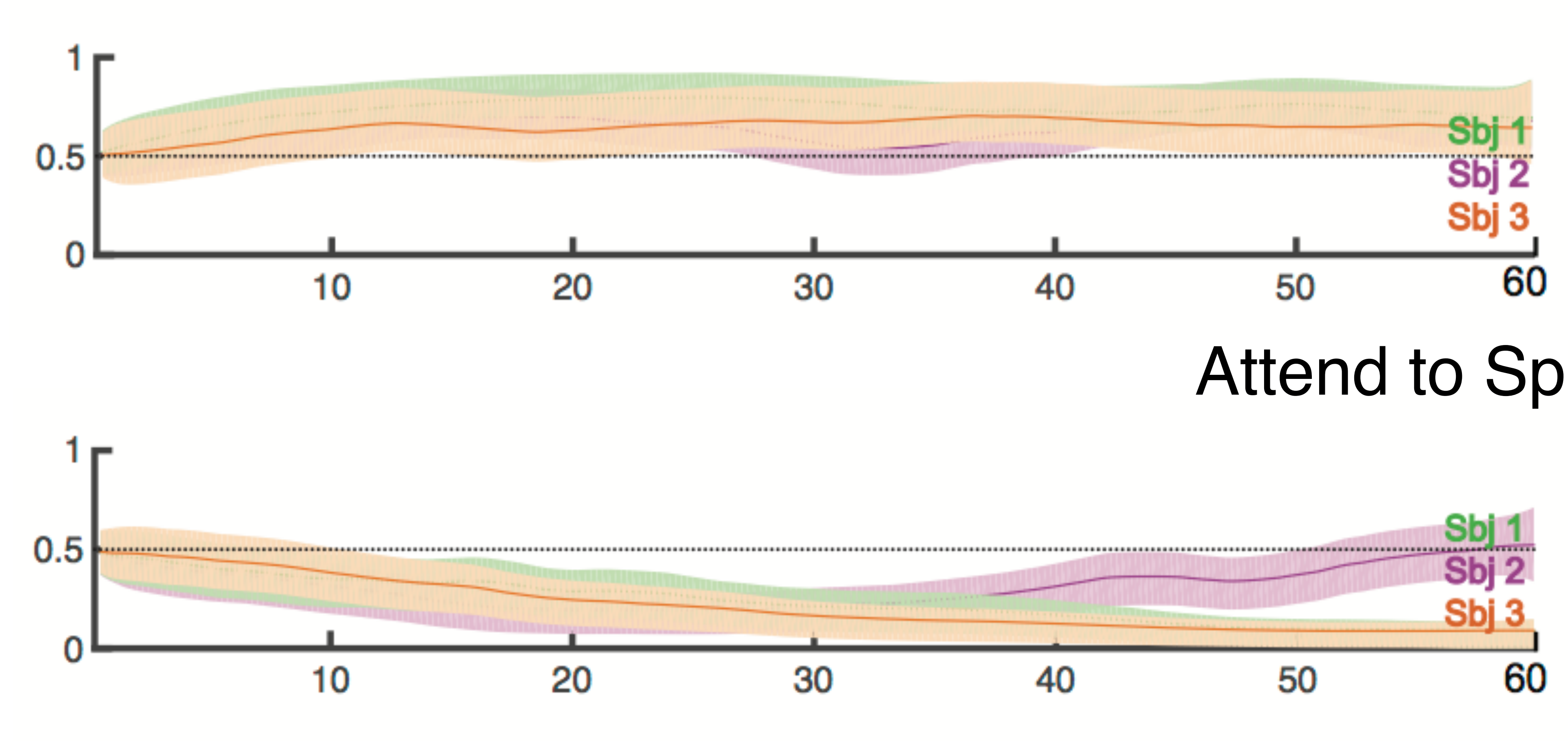
Recent Studies

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

Attentional Dynamics

Attend to Speaker 1

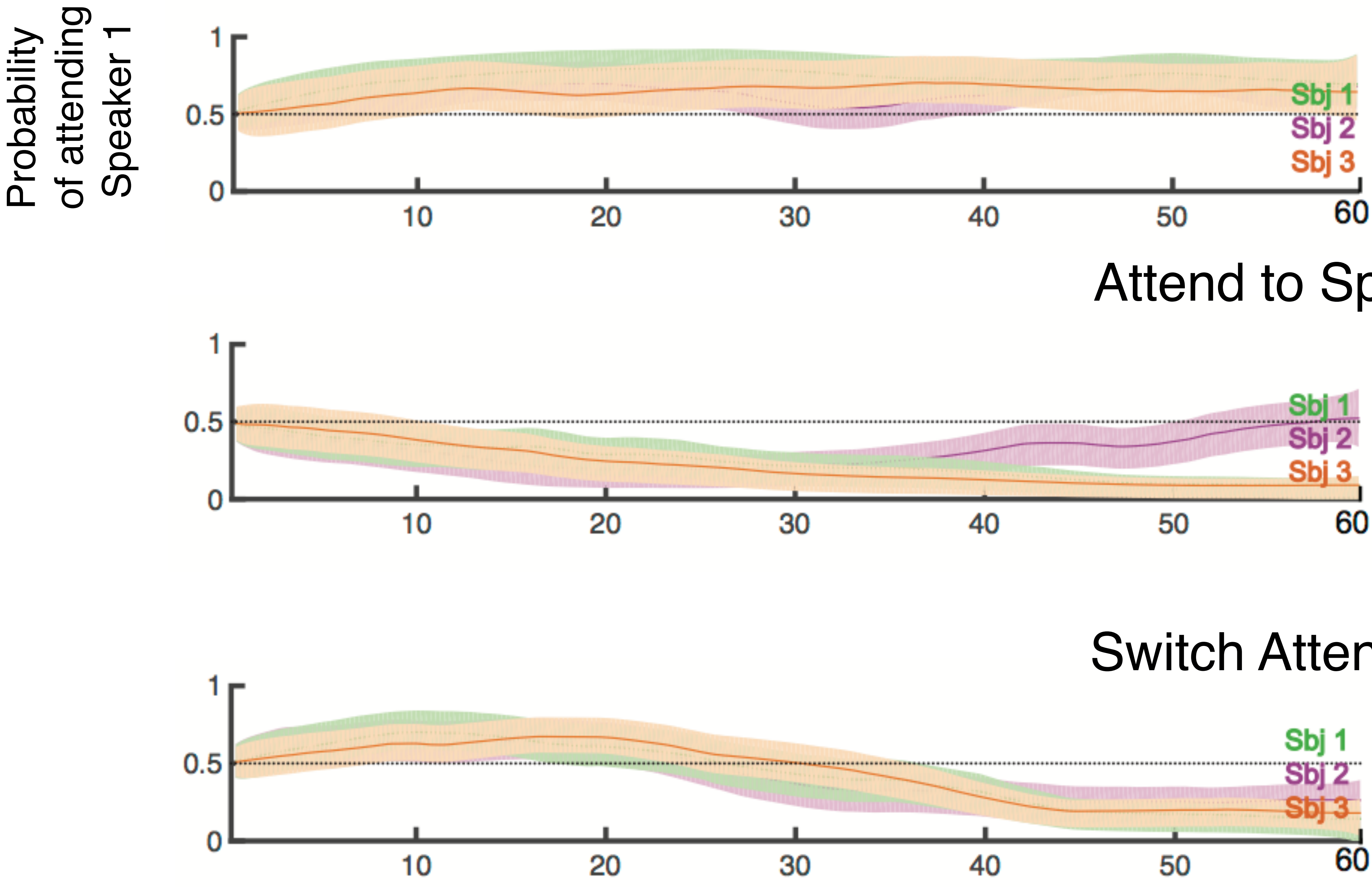
Probability of attending Speaker 1



Attend to Speaker 2

Akram et al., NeuroImage (2016)

Attentional Dynamics

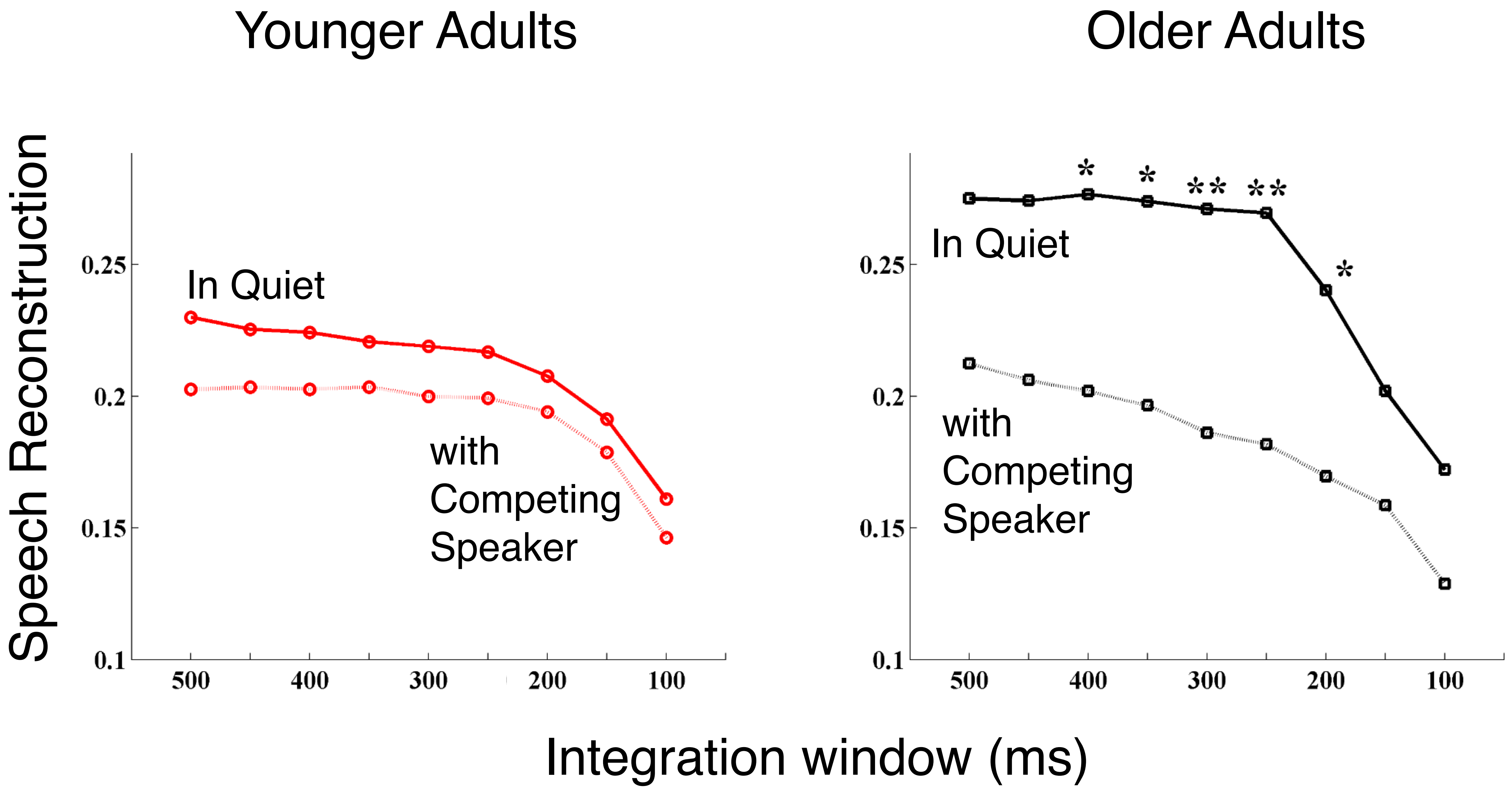

Attend to Speaker 1

Attend to Speaker 2

Switch Attention

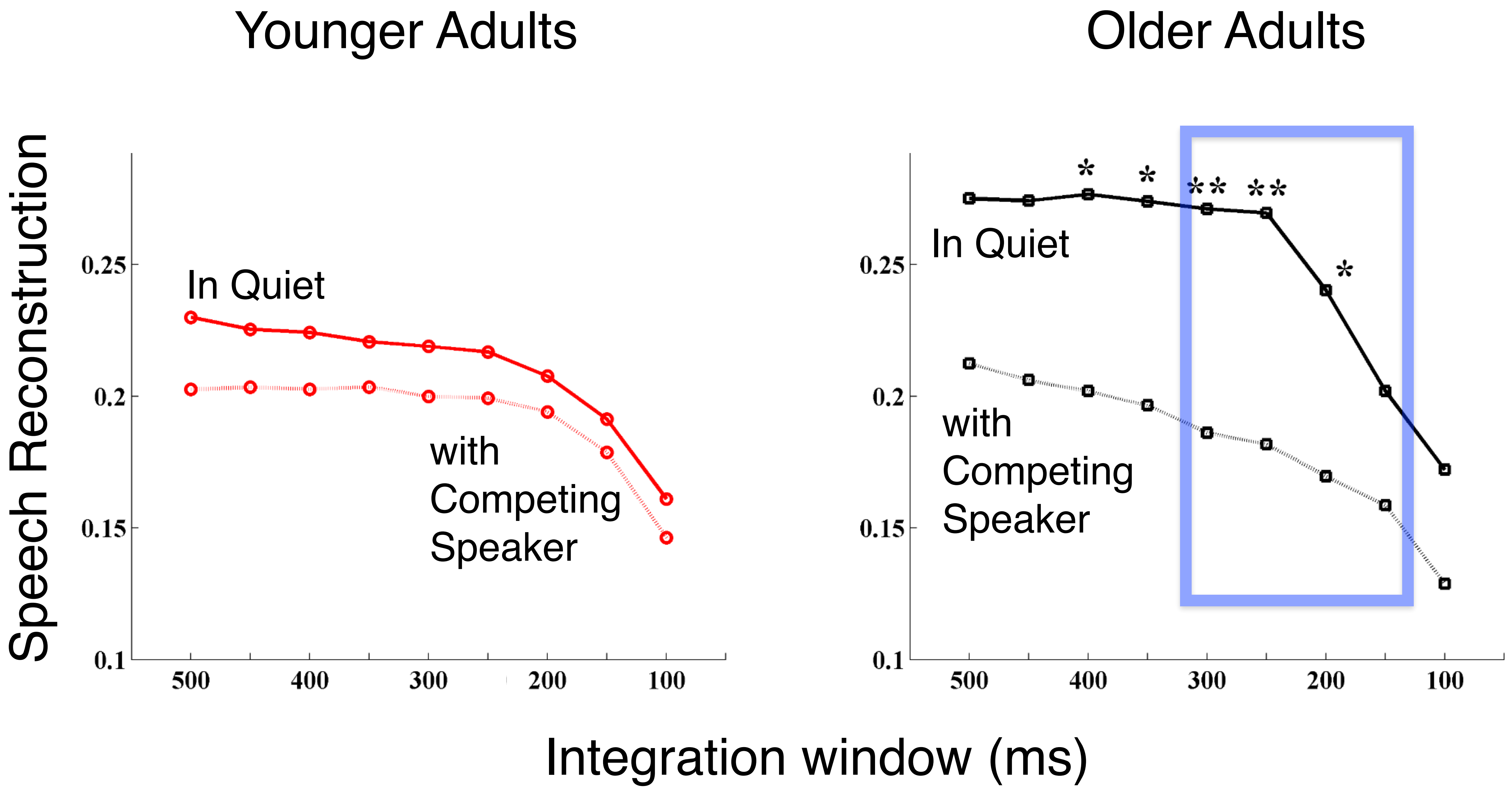
Akram et al., NeuroImage (2016)

Younger vs. Older Listeners



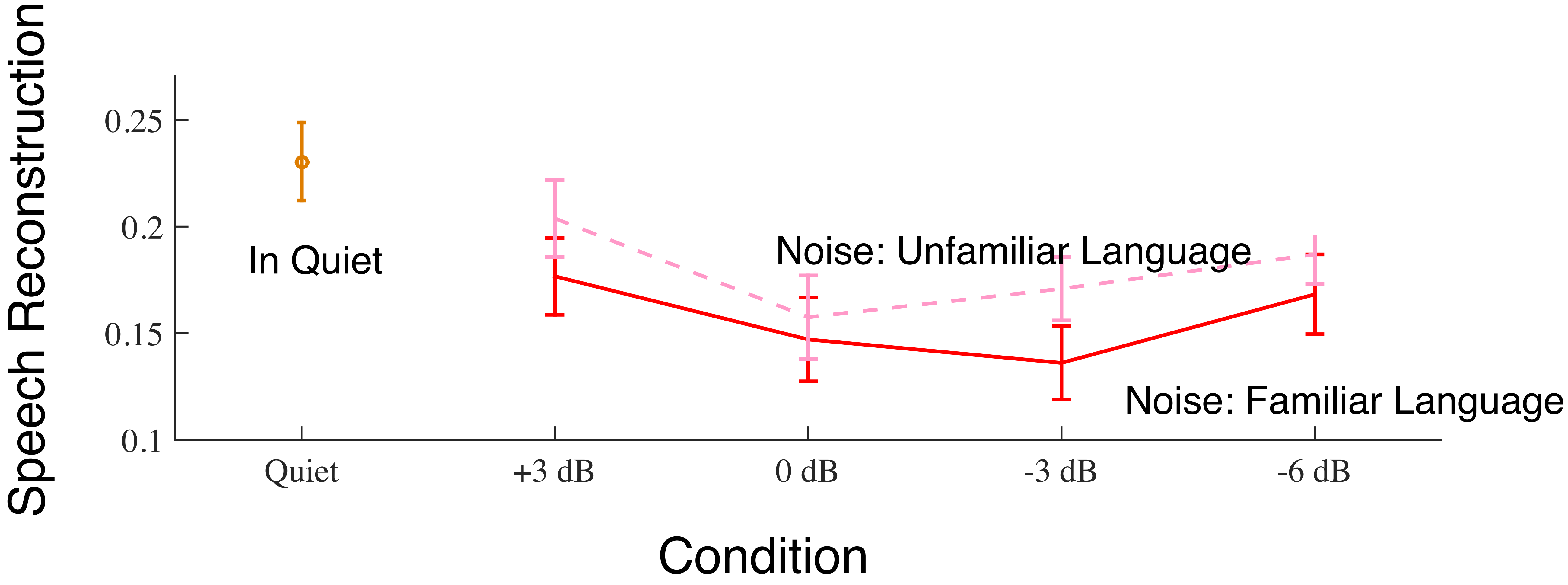
Effect absent in Midbrain (FFR Response)

Younger vs. Older Listeners



Effect absent in Midbrain (FFR Response)

High Level Interference



Effect absent in Midbrain (FFR Response)

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)
- At least 2 different object-based representations, e.g., delta vs. theta; effect of language; phoneme acoustics vs. perception

Thank You