Neural Encoding of Speech in Auditory Cortex

Jonathan Z. Simon

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

University of Maryland

University College London, 22 June 2015

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

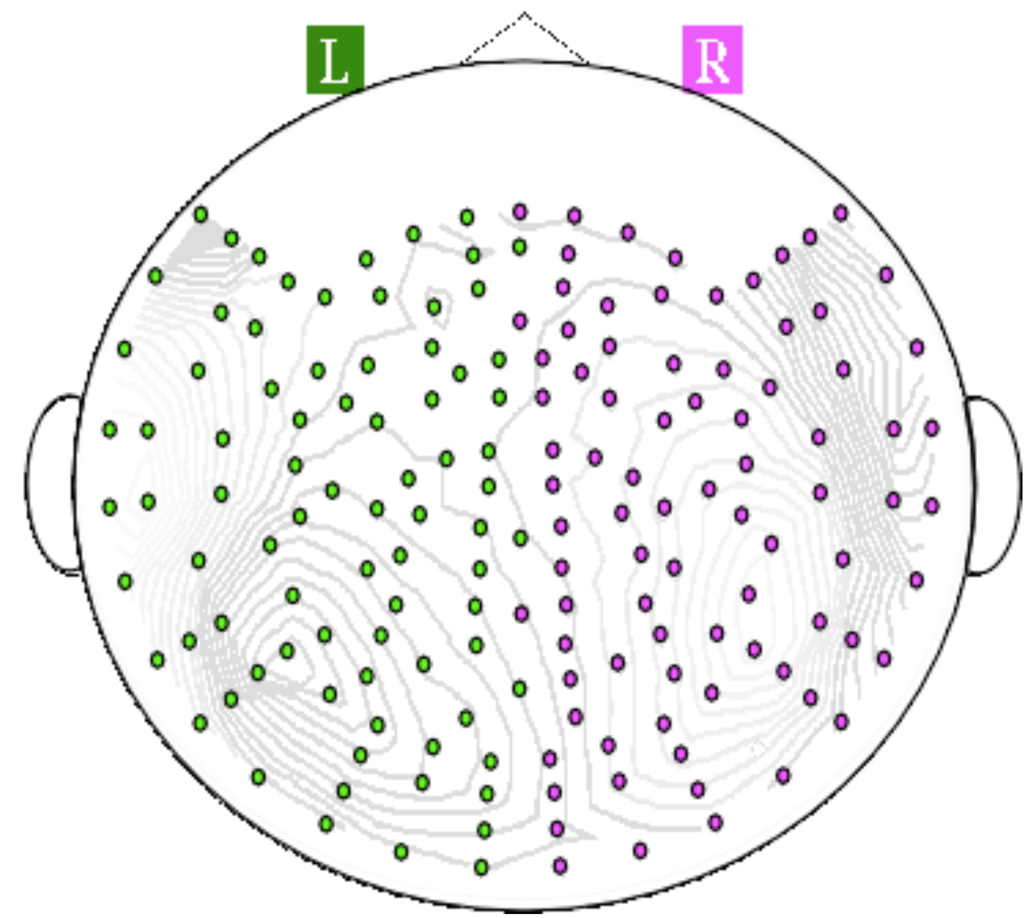
Funding NIH (*NIDCD*, *NIA*, *NIBIB*); USDA

Outline

- Magnetoencephalography (MEG)
- Cortical Representations of Speech
  - Encoding vs. Decoding
  - Attended vs. Unattended Speech
- Work in Progress
  - Attentional Dynamics
  - Aging and the Cocktail Party Problem
  - Foreground vs. Background

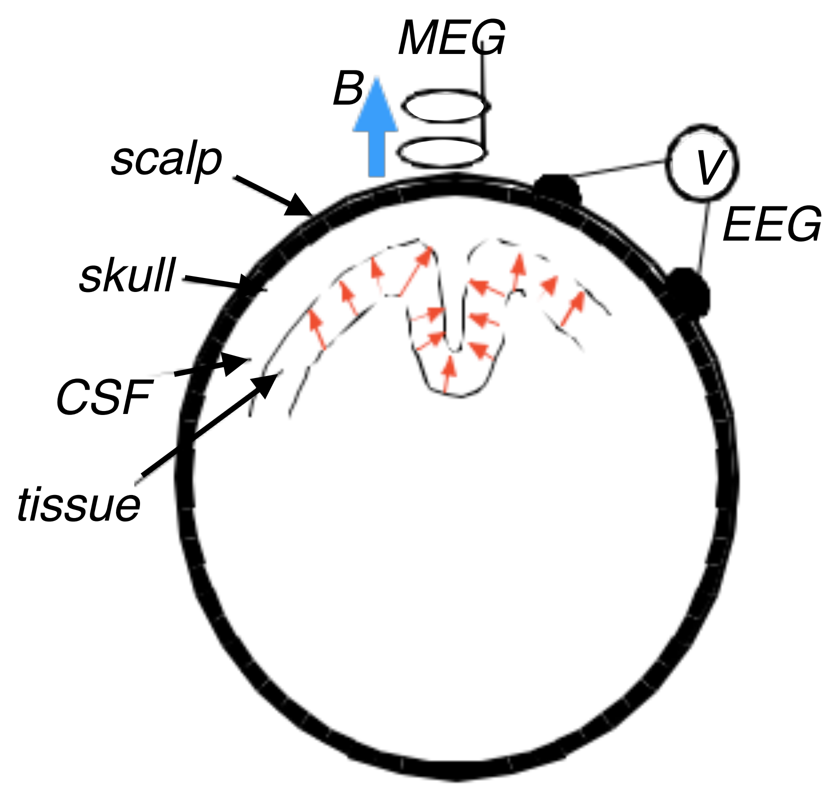
Magnetoencephalography

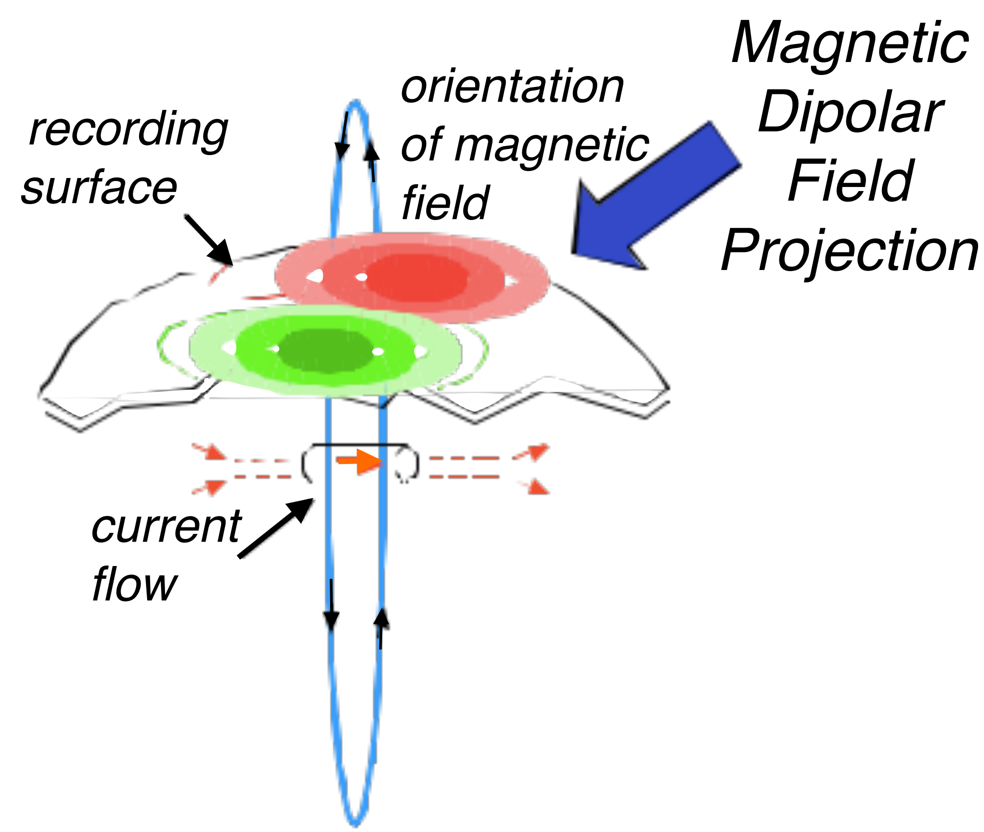
- Non-invasive, Passive, Silent Neural Recordings
- Simultaneous Whole-Head Recording (~200 sensors)
- Sensitivity
  - high: ~100 fT (10^{-13} Tesla)
  - low: ~10^4 – ~10^6 neurons
- Temporal Resolution: ~1 ms
- Spatial Resolution
  - coarse: ~1 cm
  - ambiguous



Neural Signals & MEG

Photo by Fritz Goro





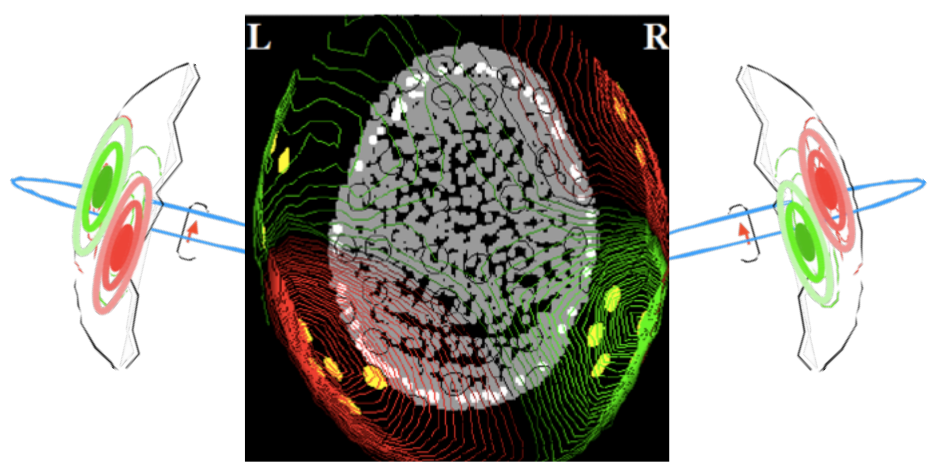
- Direct electrophysiological measurement
    - not hemodynamic
    - real-time
- No unique solution for distributed source

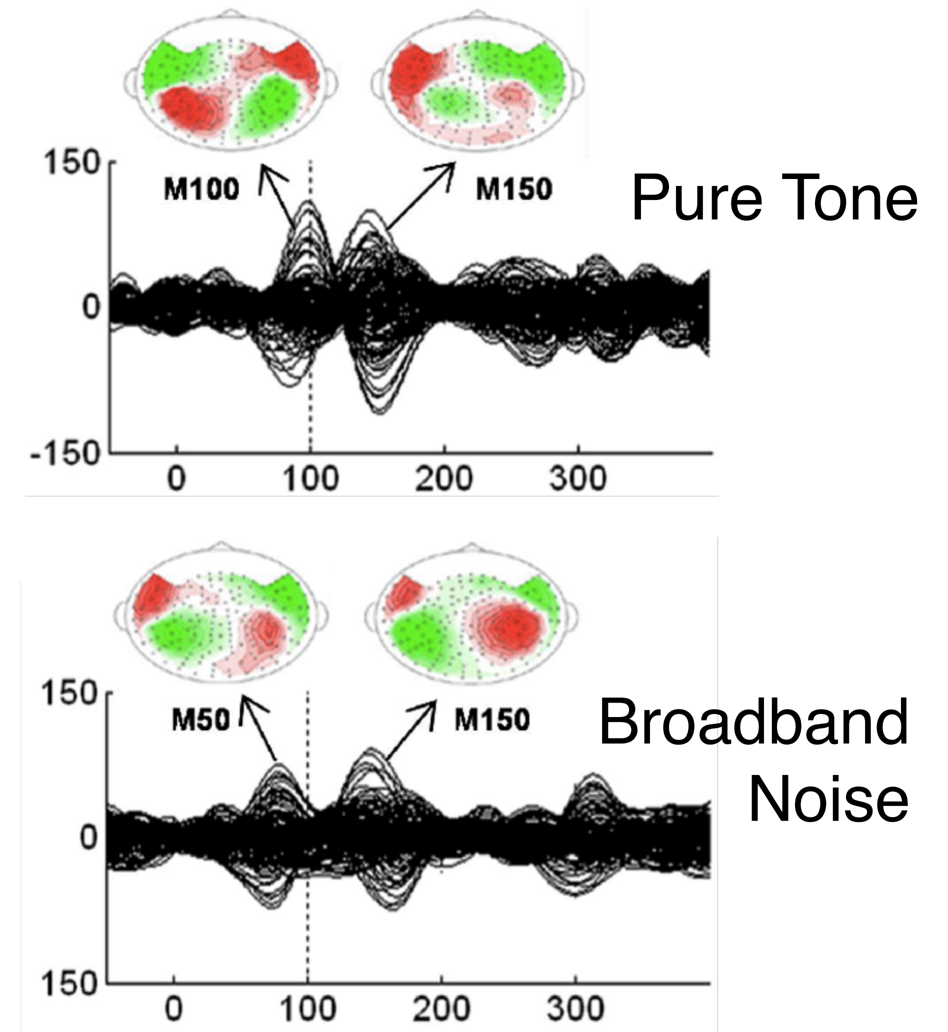
- Measures spatially synchronized cortical activity
- Fine temporal resolution (~ 1 ms)
- Moderate spatial resolution (~ 1 cm)

Time Course of MEG Responses

Auditory Evoked Responses

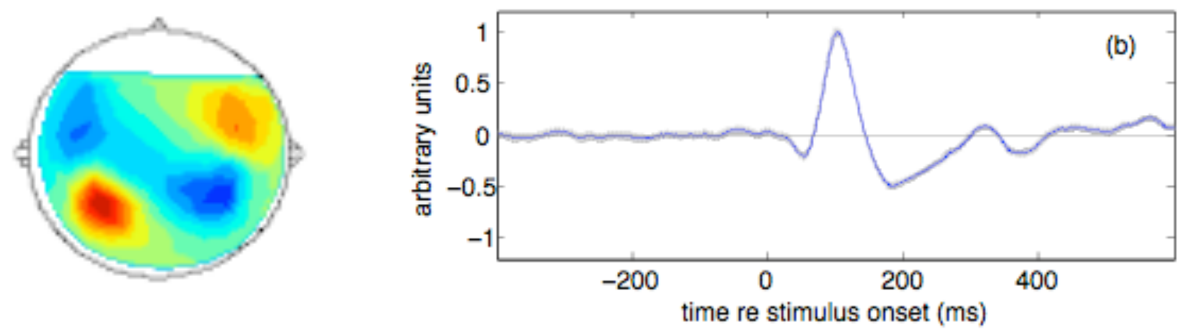
- MEG Response Patterns Time-Locked to Stimulus Events
- Robust
- Strongly Lateralized

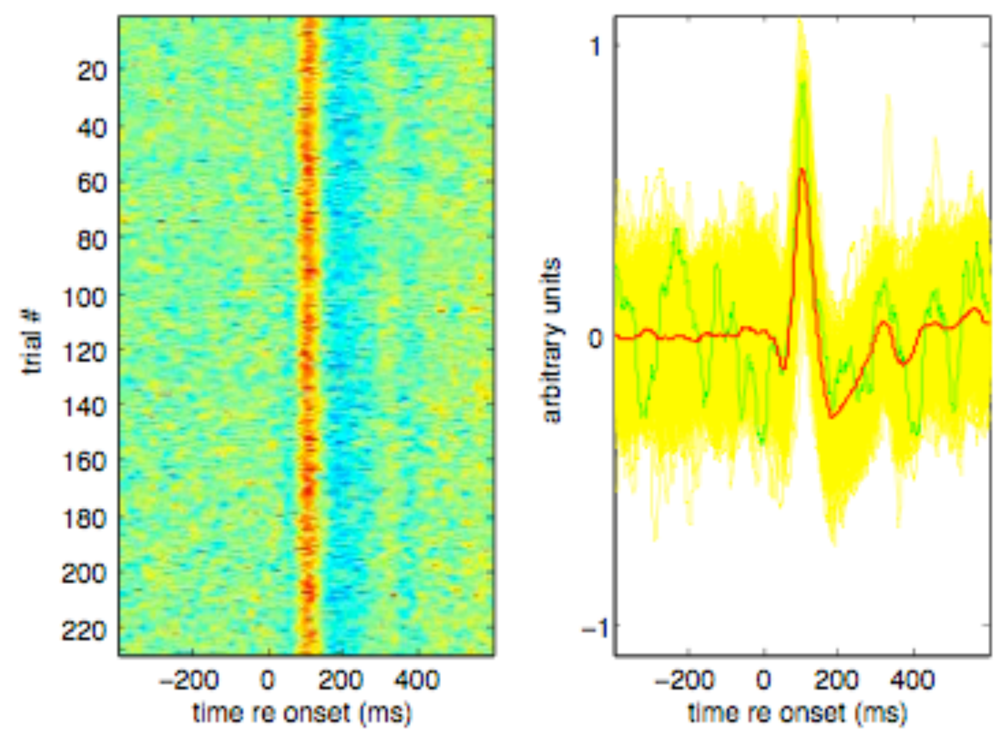




Component Analysis

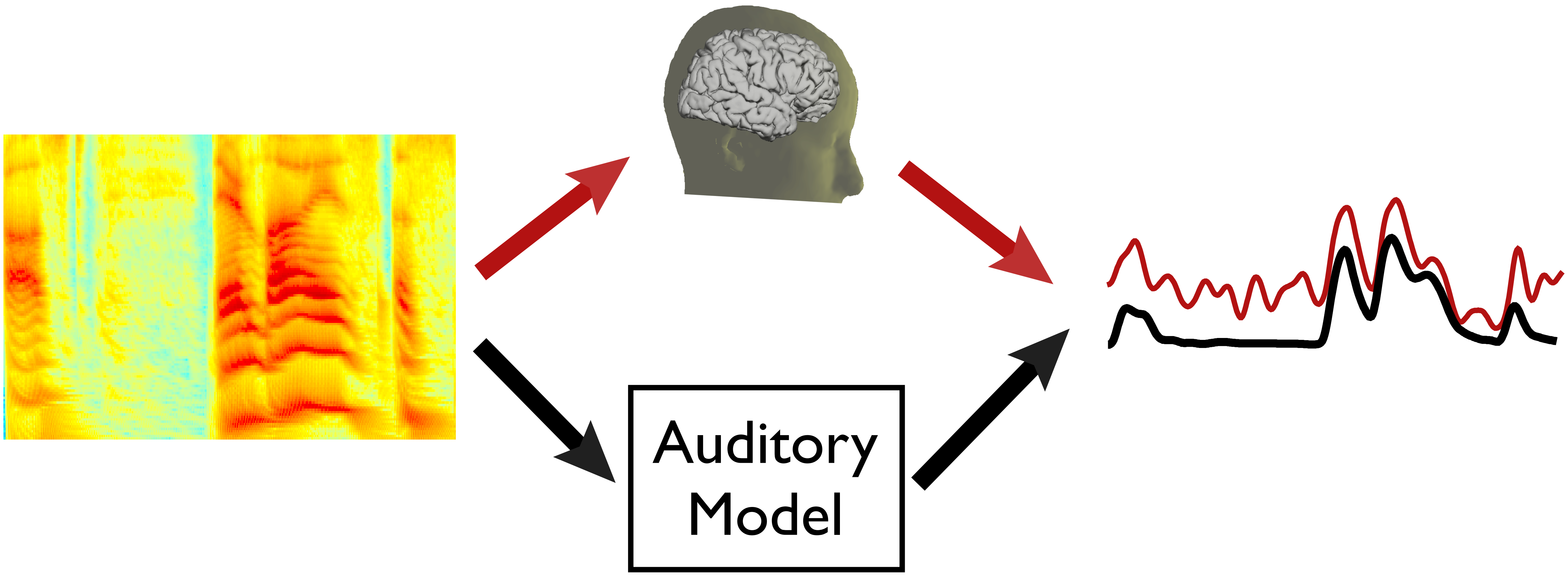
- Each component has both spatial and temporal profile
- Data driven, e.g., PCA, ICA, DSS
- DSS: ordered by trial-to-trial reproducibility
- ➔ Spatial Filter,
  e.g. for single trials
- **Can analyze temporal processing separately from anatomical origin**



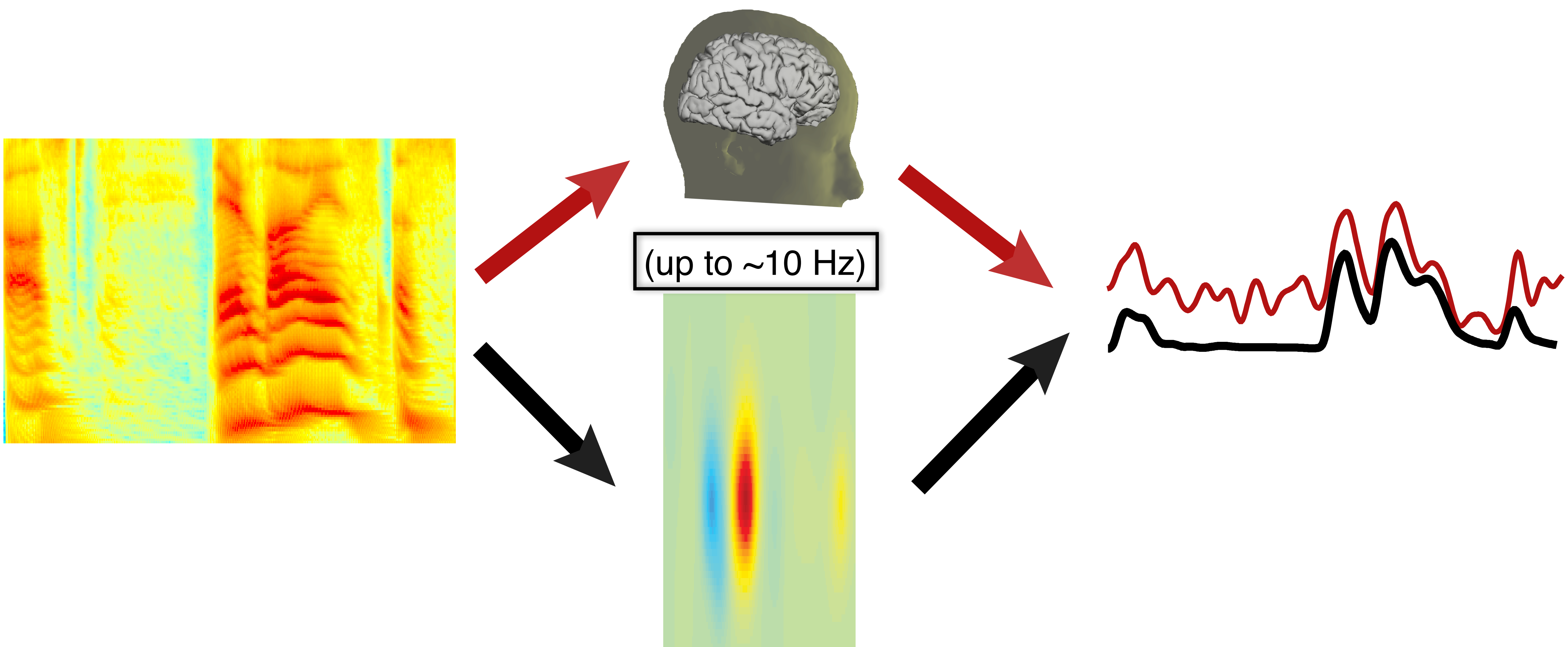


Särelä & Valpola (2005)
de Cheveigné & Simon, J. Neurosci. Methods (2008)

MEG Responses to Speech Modulations



MEG Responses Predicted by STRF Model

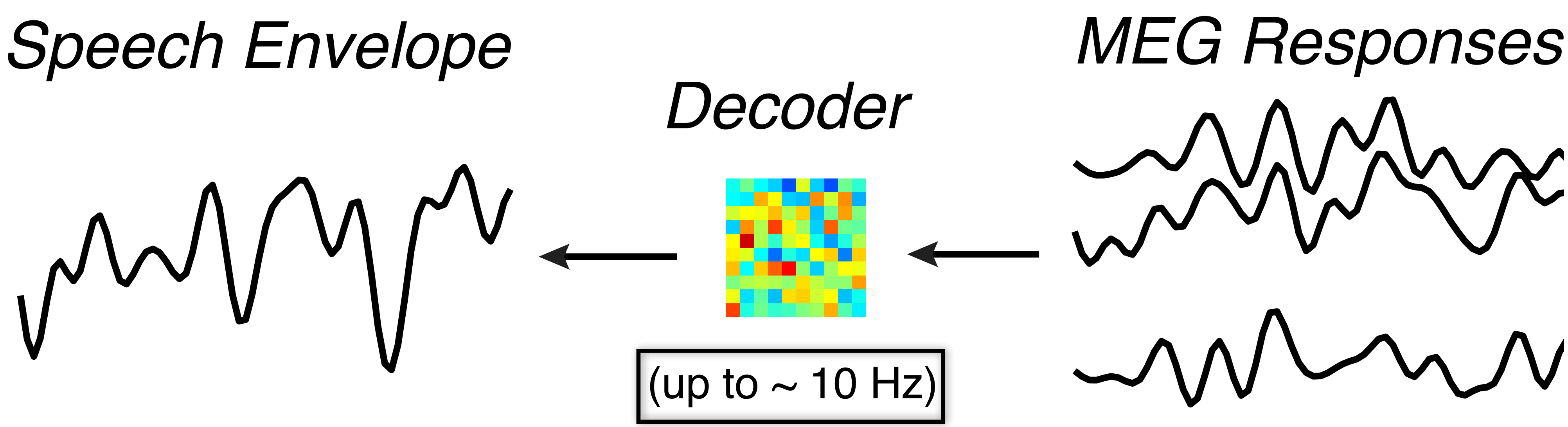


Linear Kernel = STRF

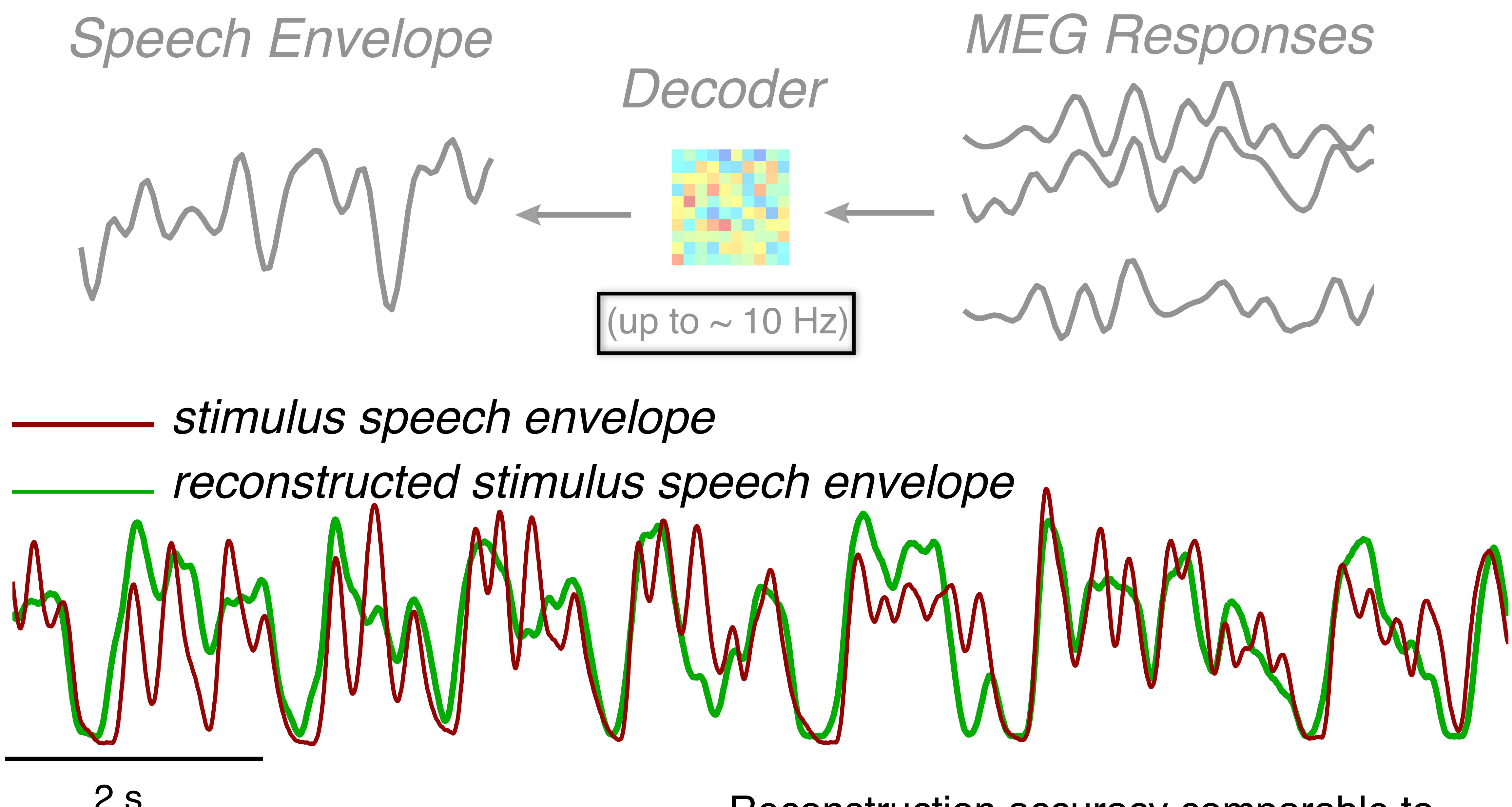
Ding & Simon, J Neurophysiol (2012)

"Spectro-Temporal Response Function"

Neural Reconstruction of Speech Envelope

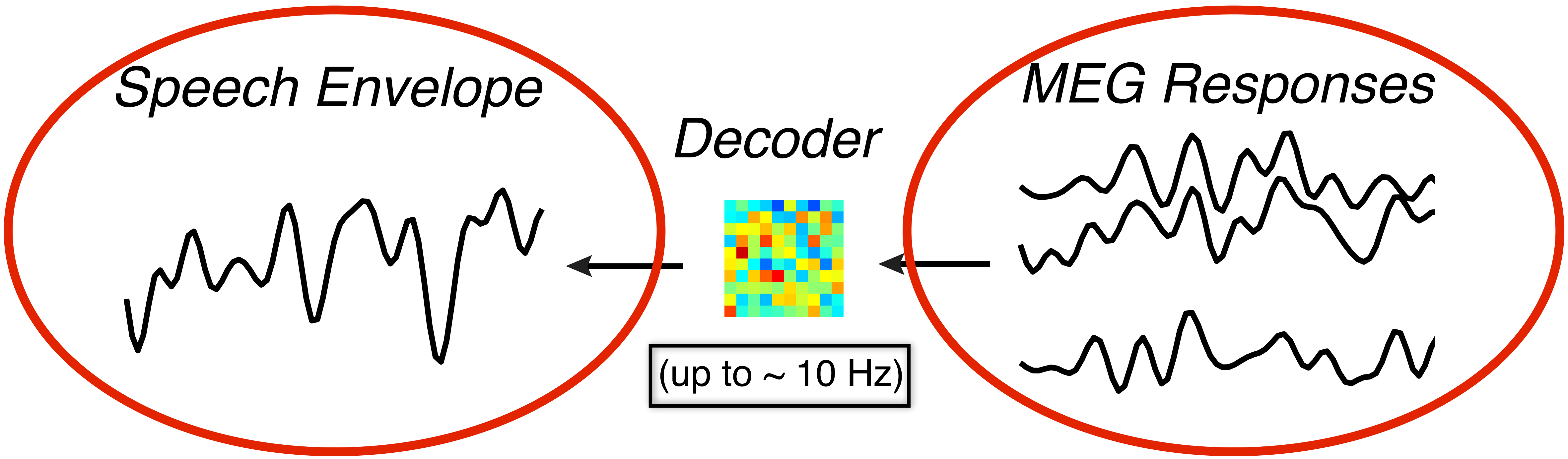


Neural Reconstruction of Speech Envelope



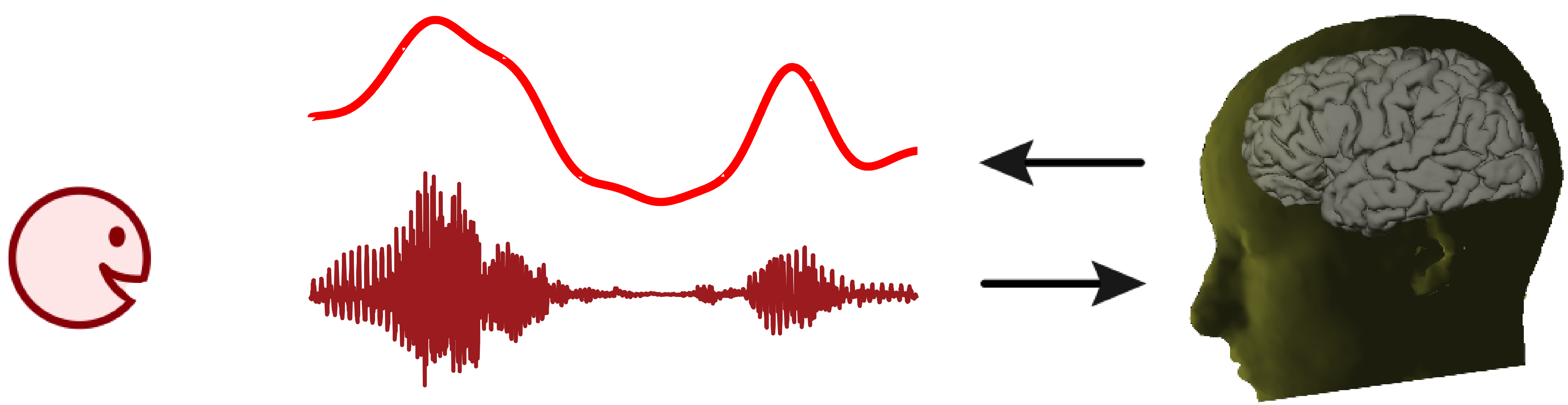
Reconstruction accuracy comparable to single unit & ECoG recordings

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

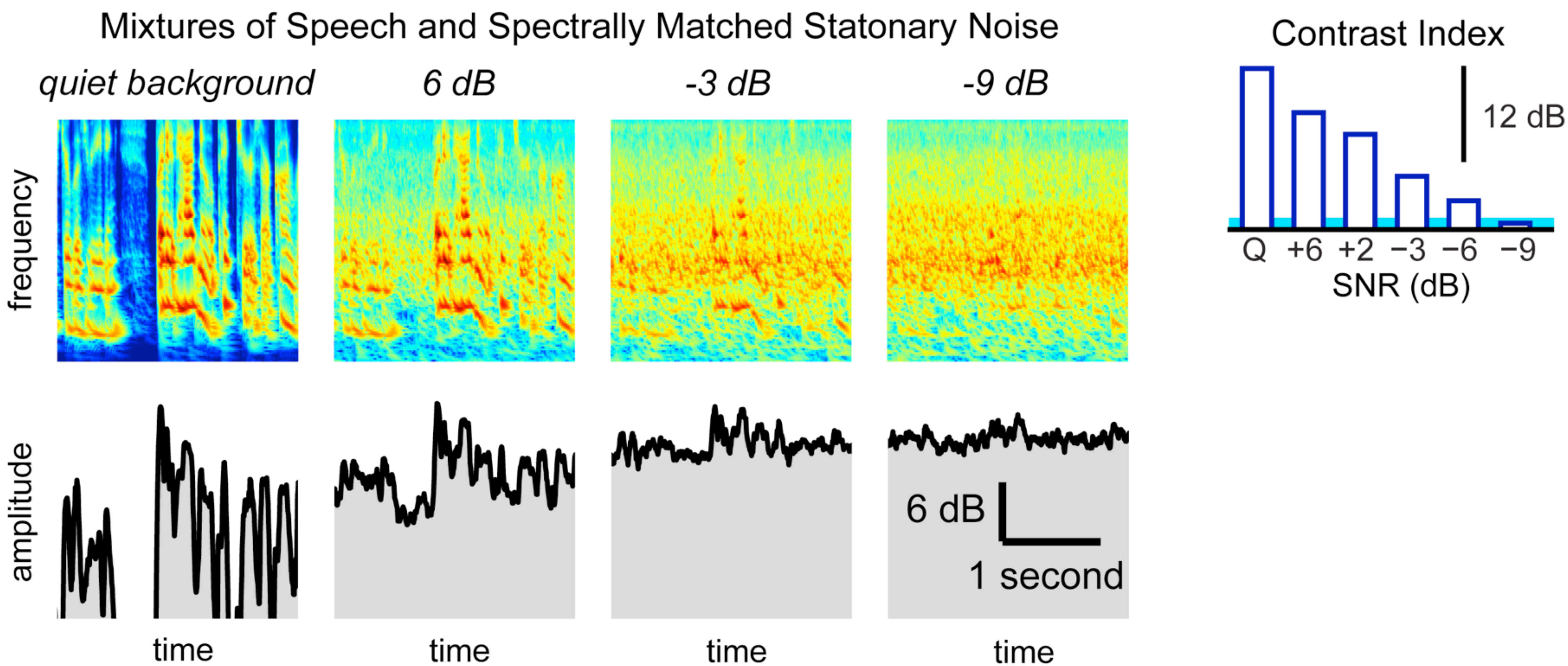


Speech Envelope
Decoder
(up to ~ 10 Hz)
MEG Responses

Neural Representation of Speech: Temporal

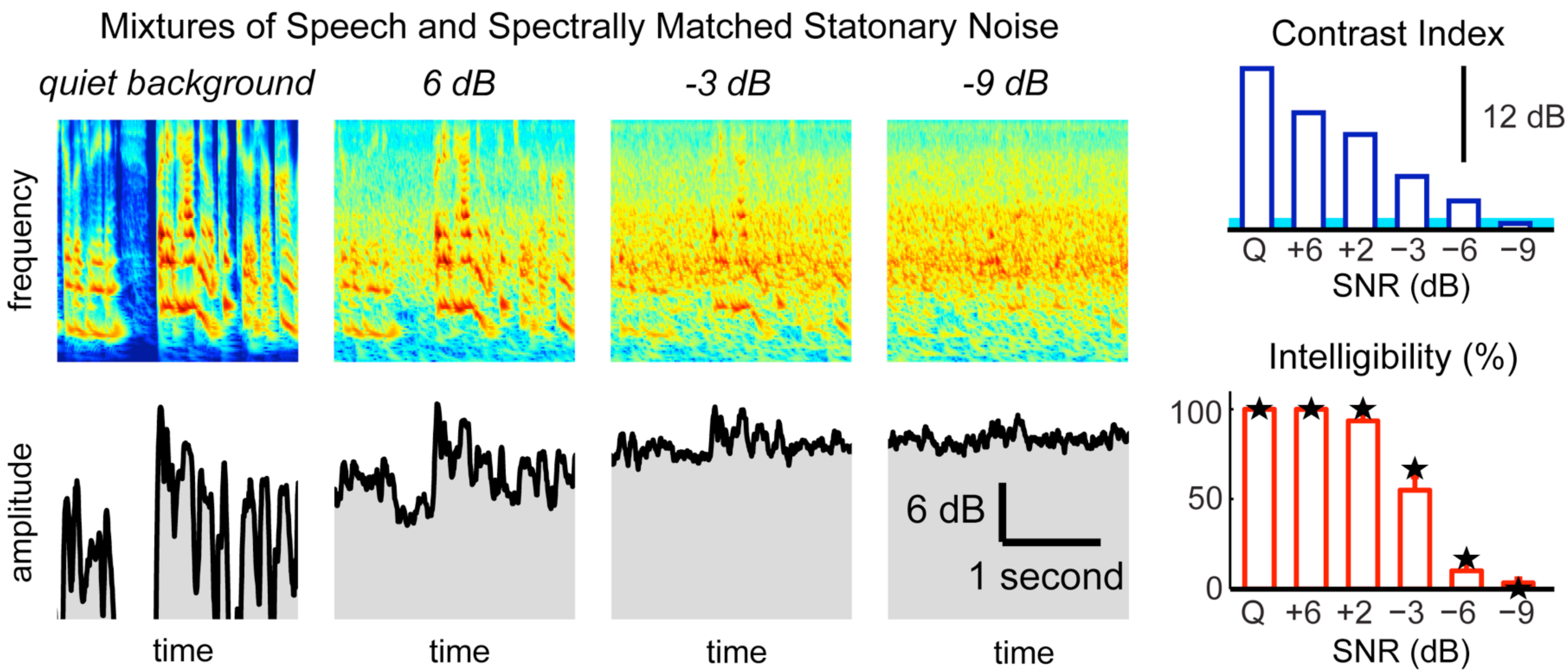


Speech in Noise



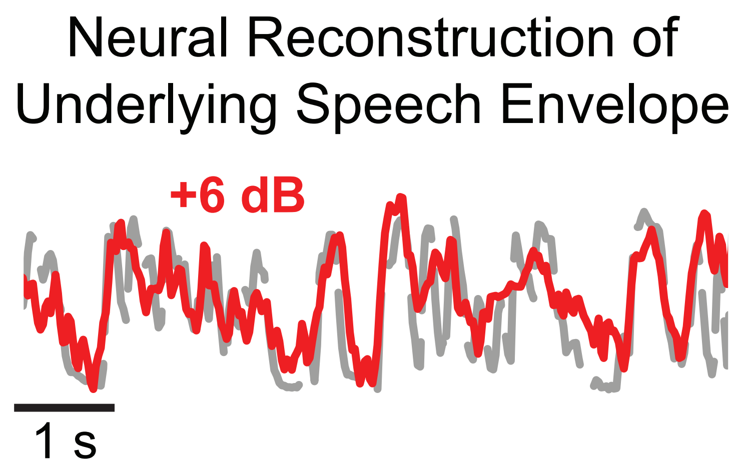
Ding & Simon, J Neuroscience (2013)

Speech in Noise



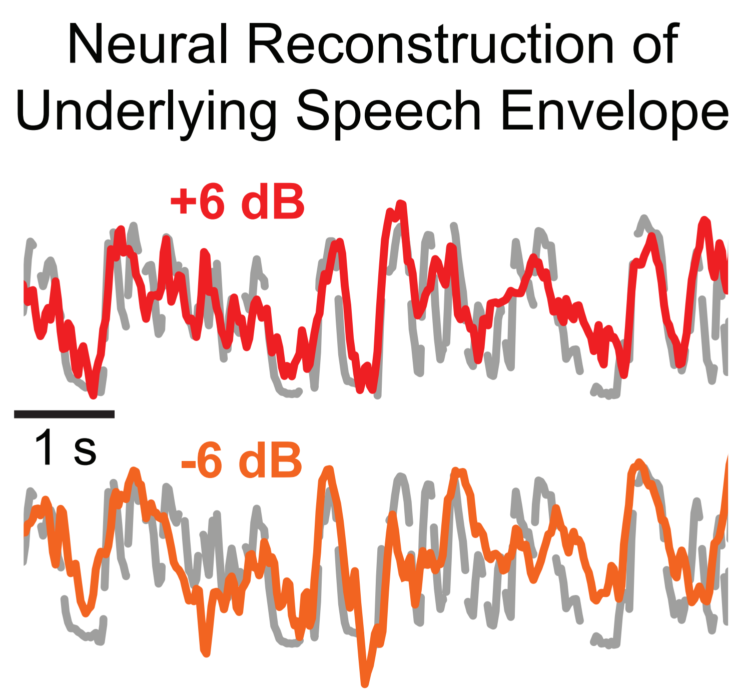
Ding & Simon, J Neuroscience (2013)

Speech in Noise: Results



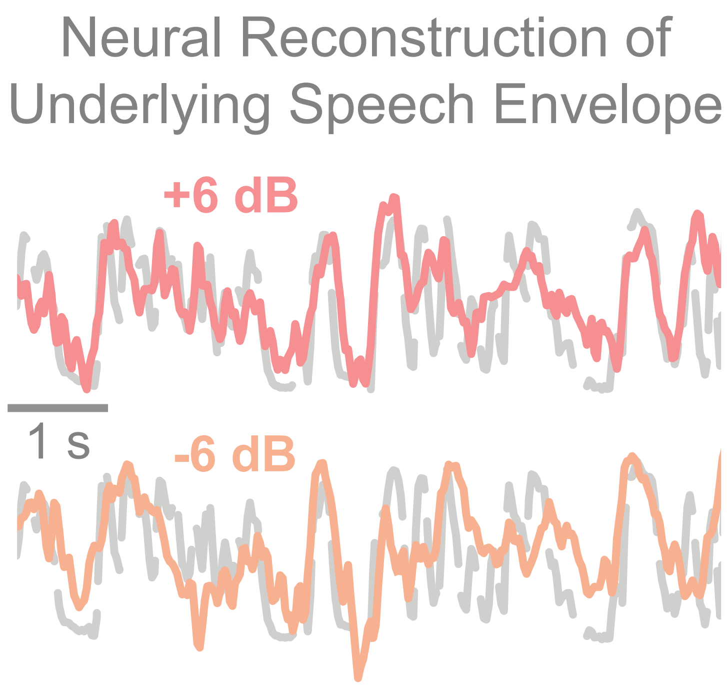
Ding & Simon, J Neuroscience (2013)

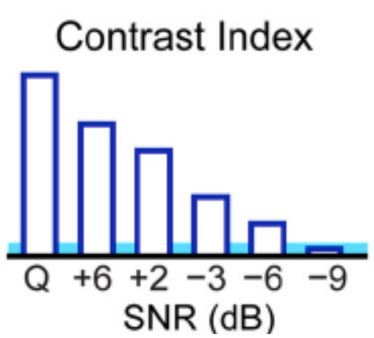
Speech in Noise: Results

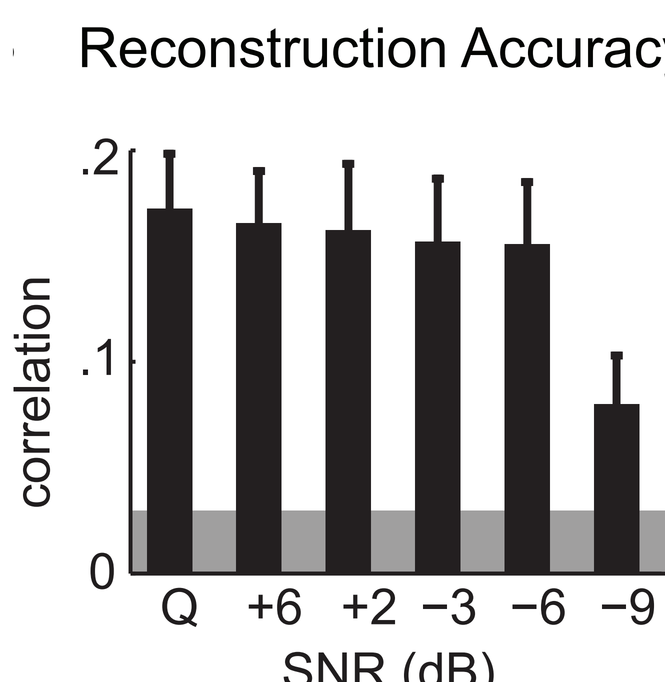


Ding & Simon, J Neuroscience (2013)

Speech in Noise: Results

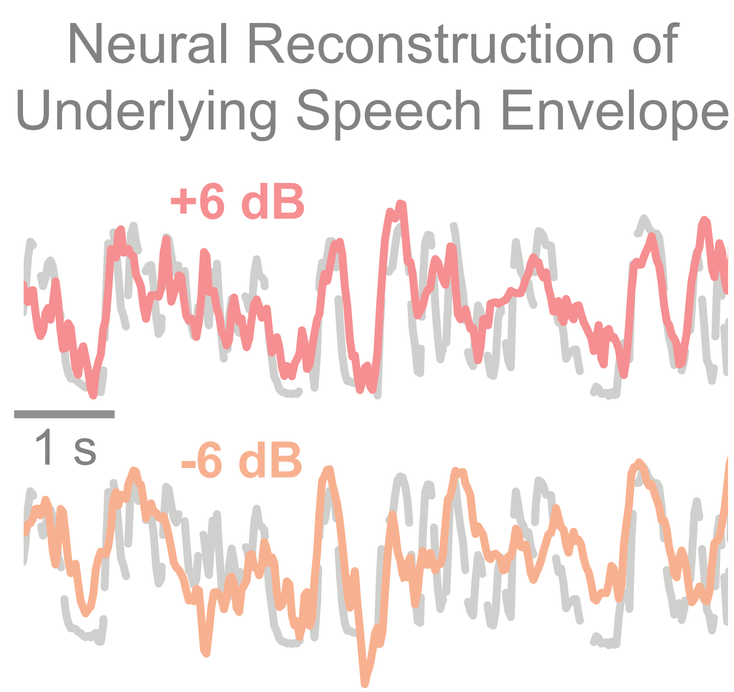


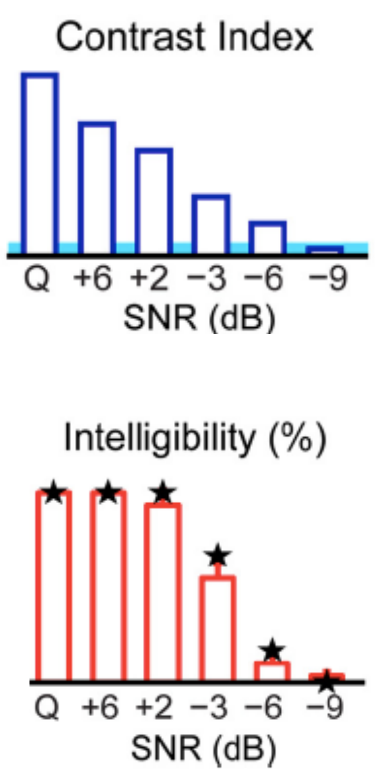


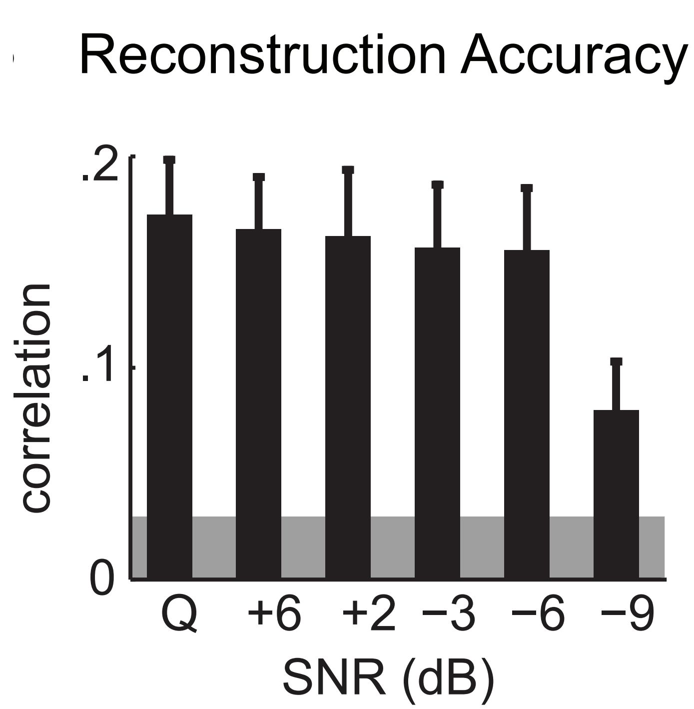


Ding & Simon, J Neuroscience (2013)

Speech in Noise: Results

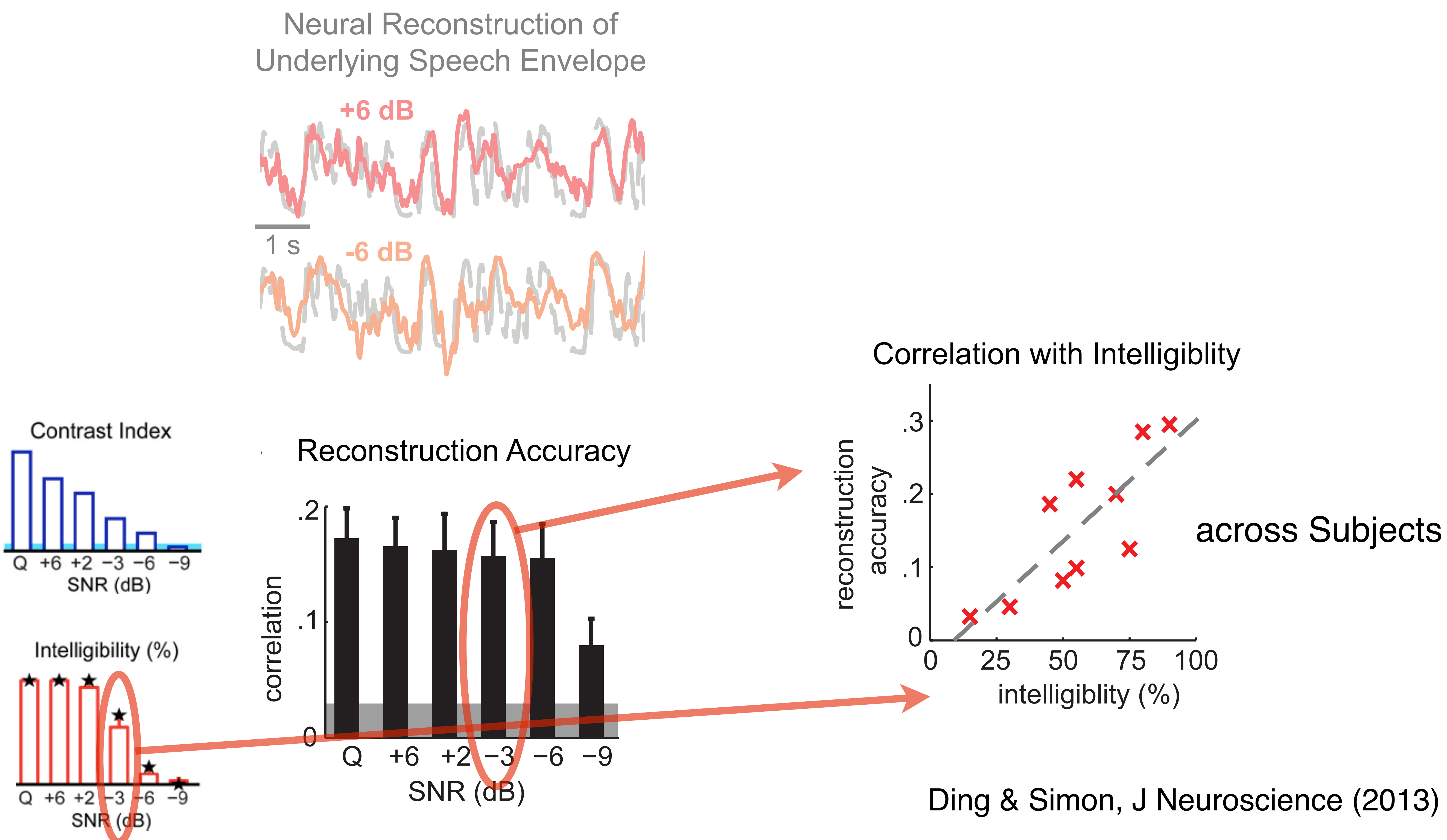






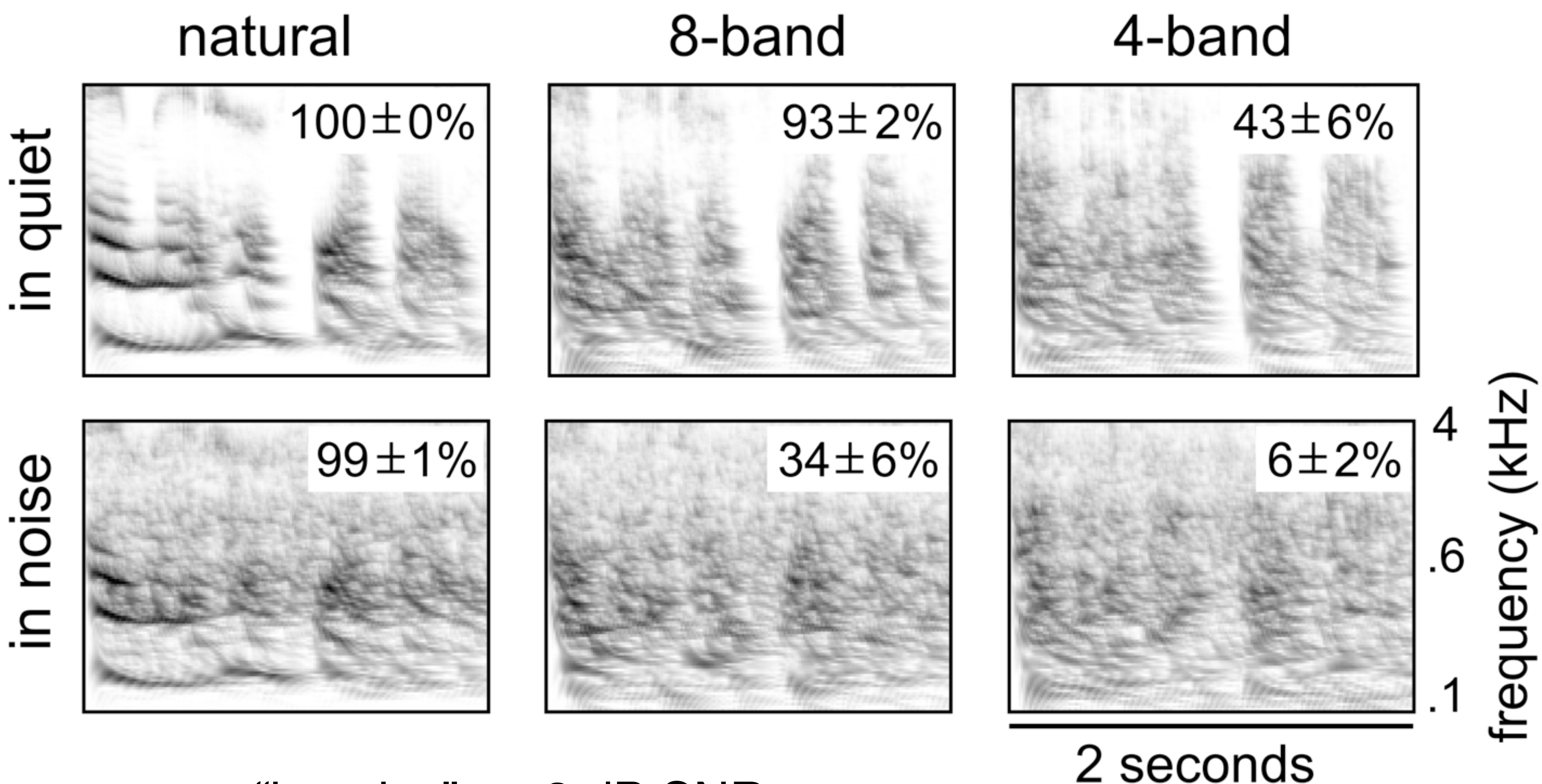
Ding & Simon, J Neuroscience (2013)

Speech in Noise: Results



Ding & Simon, J Neuroscience (2013)

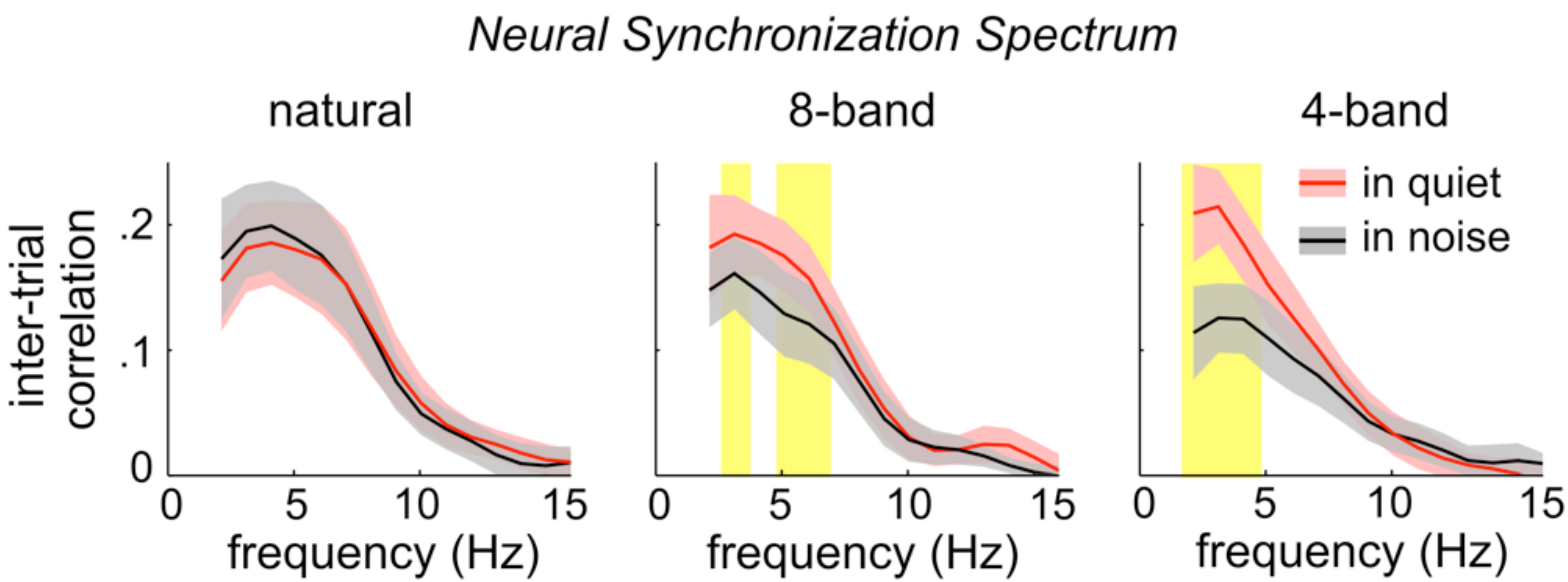
Noise-Vocoded Speech



"in noise" = +3 dB SNR

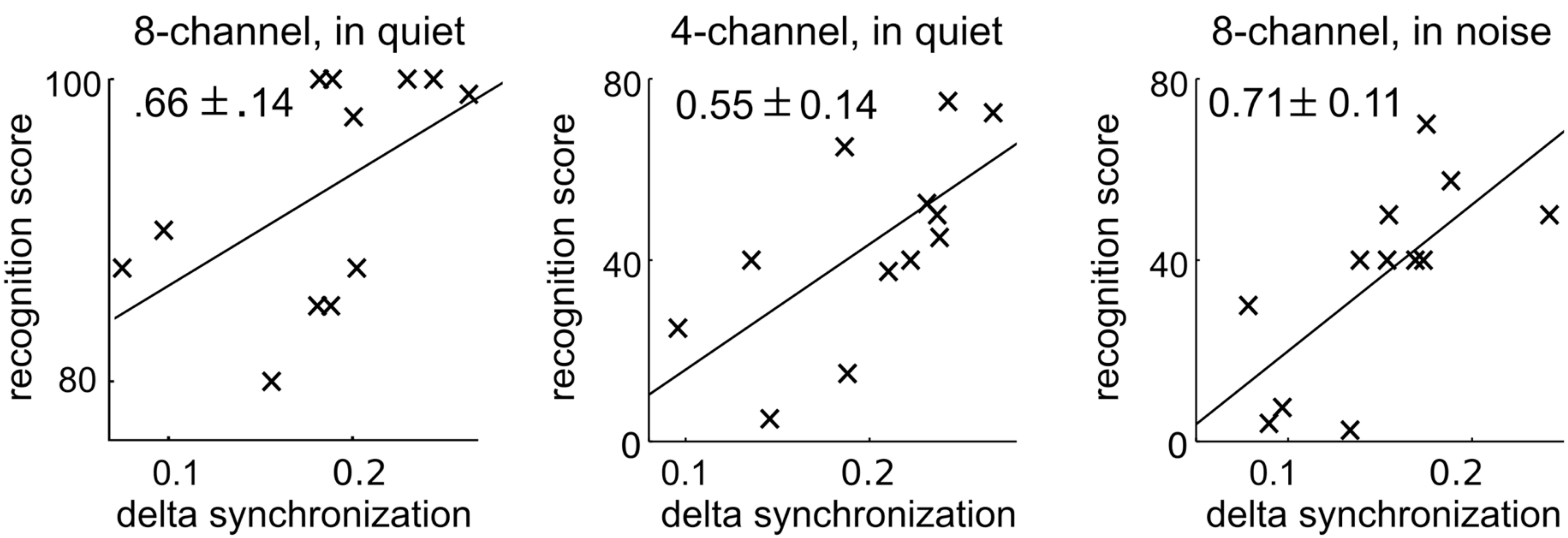
Ding, Chatterjee & Simon, NeuroImage (2014)

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



Cortical Speech Representations

- Neural Representations: Encoding & Decoding
- Linear models: Useful & Robust
- Speech **Envelope** only (as seen by MEG)
- Envelope Rates: ~ 1 - 10 Hz

The Cocktail Party

Alex Katz,
The Cocktail Party

The Cocktail Party

Alex Katz,
The Cocktail Party

The Cocktail Party

Alex Katz,
The Cocktail Party

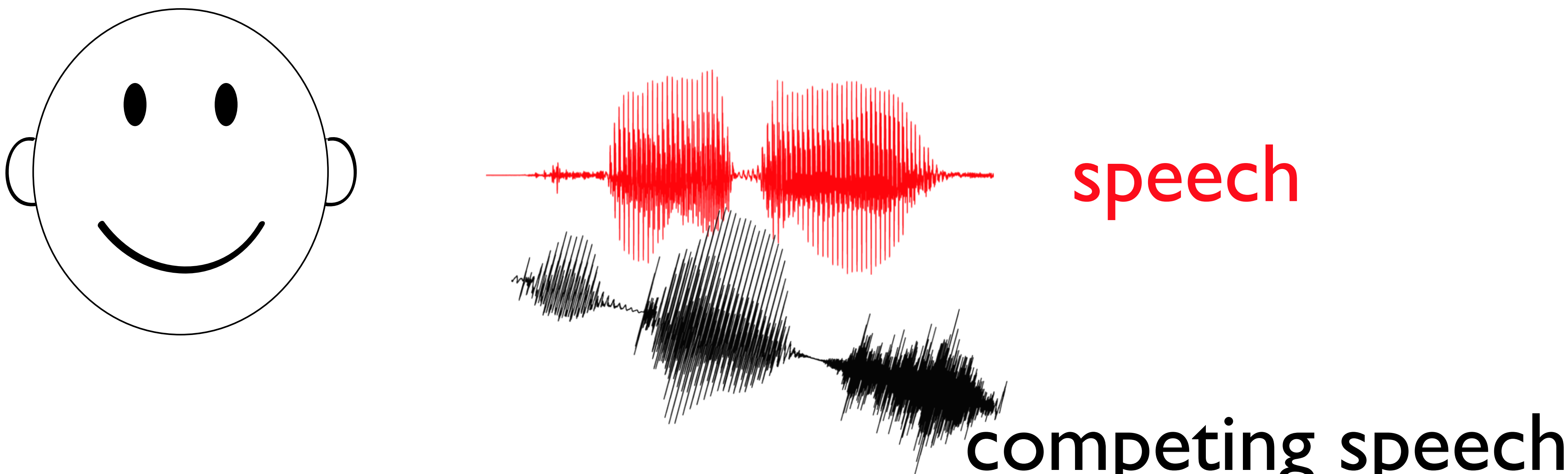
The Cocktail Party

Alex Katz,
The Cocktail Party

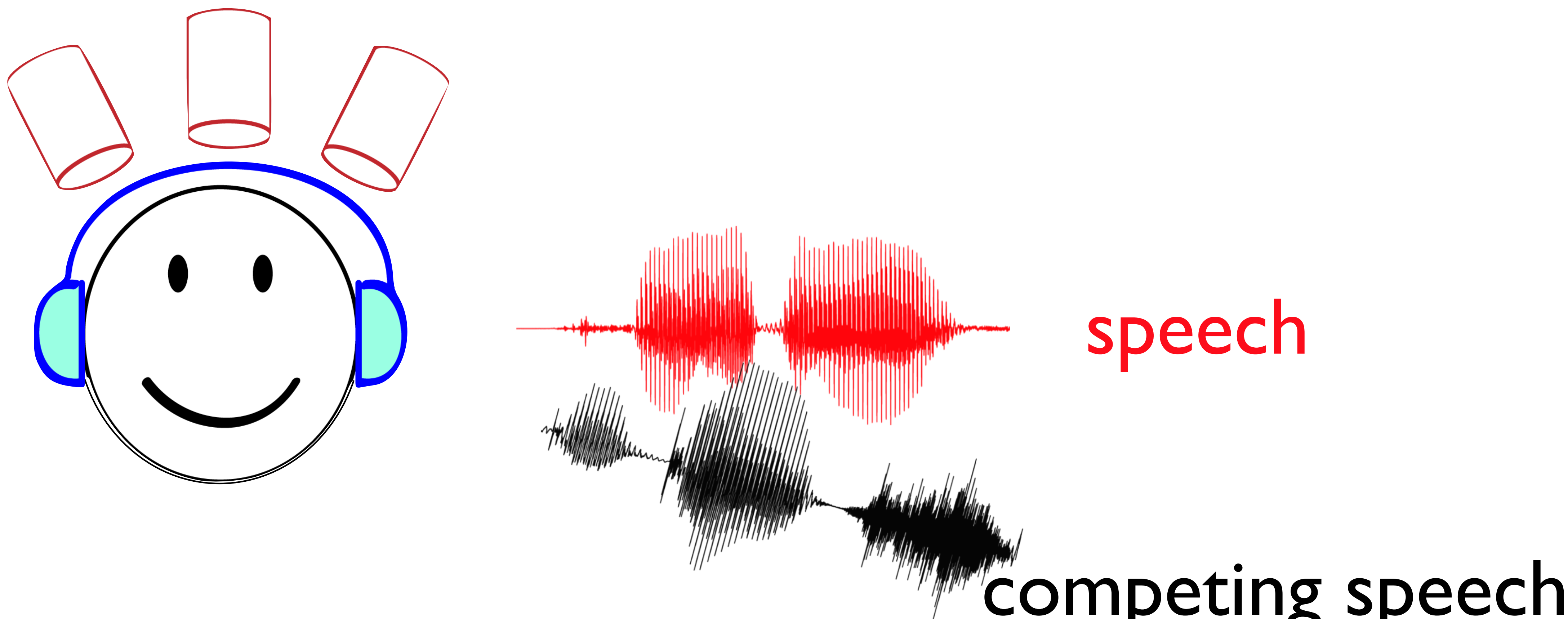
The Cocktail Party

Alex Katz,
The Cocktail Party

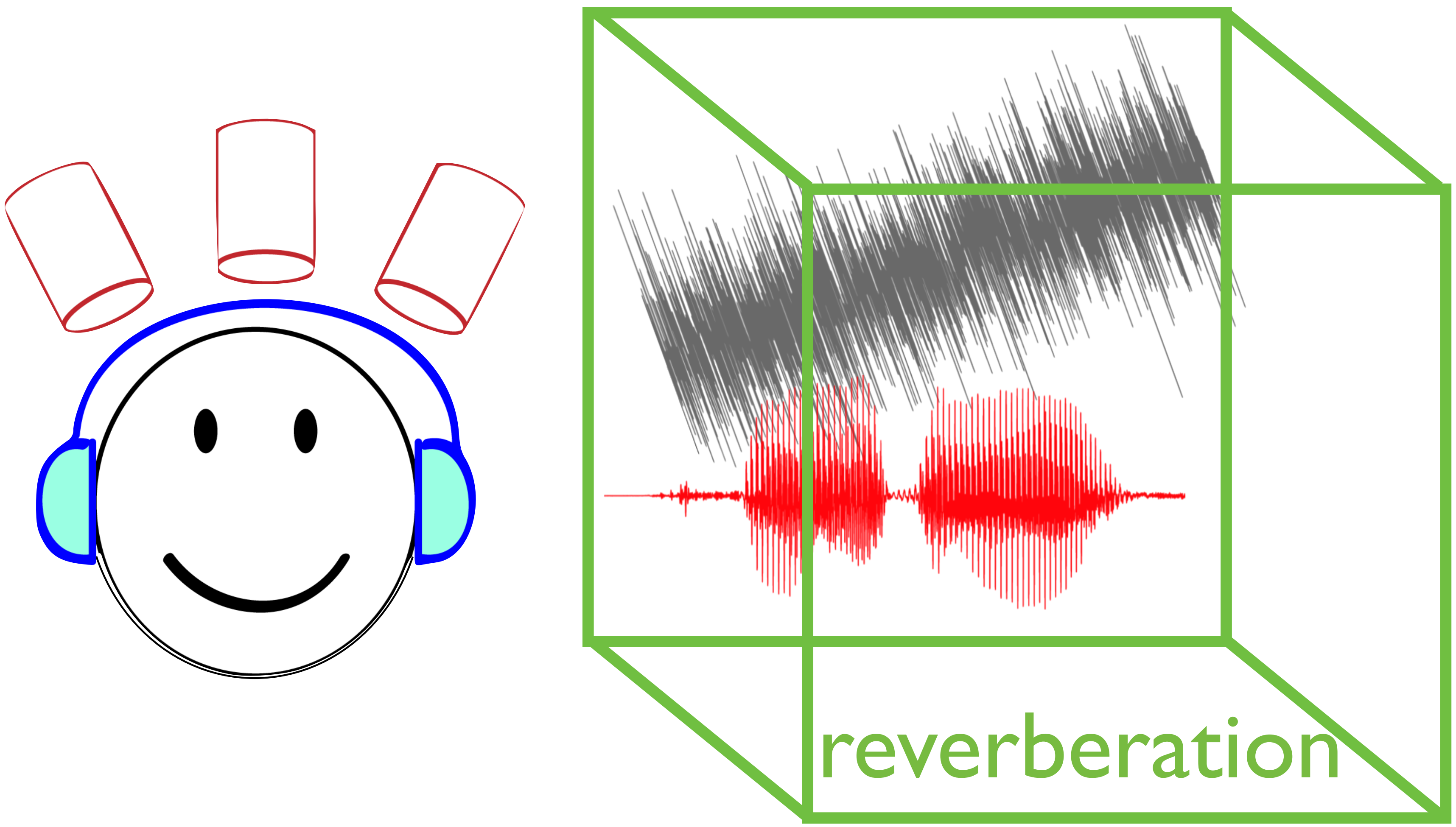
Experiments



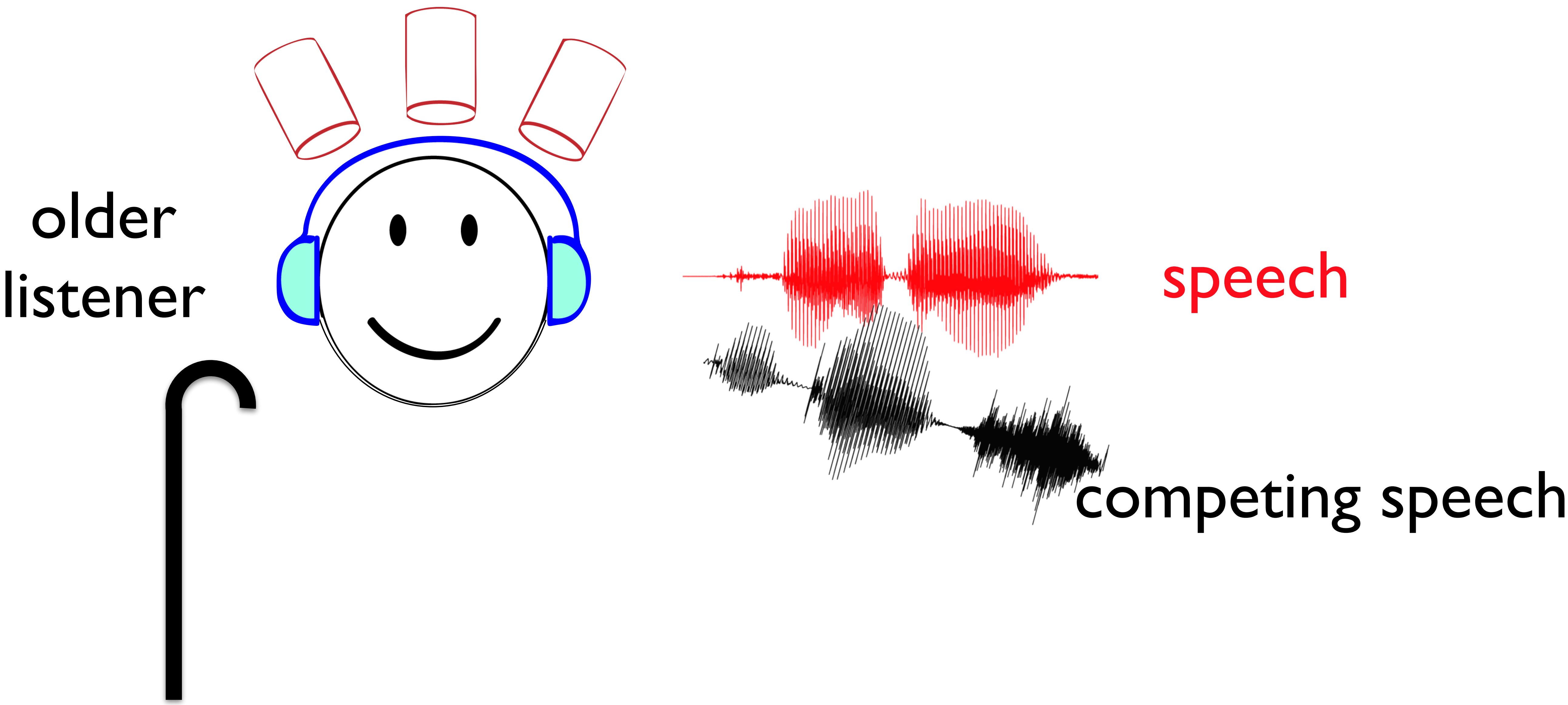
Experiments



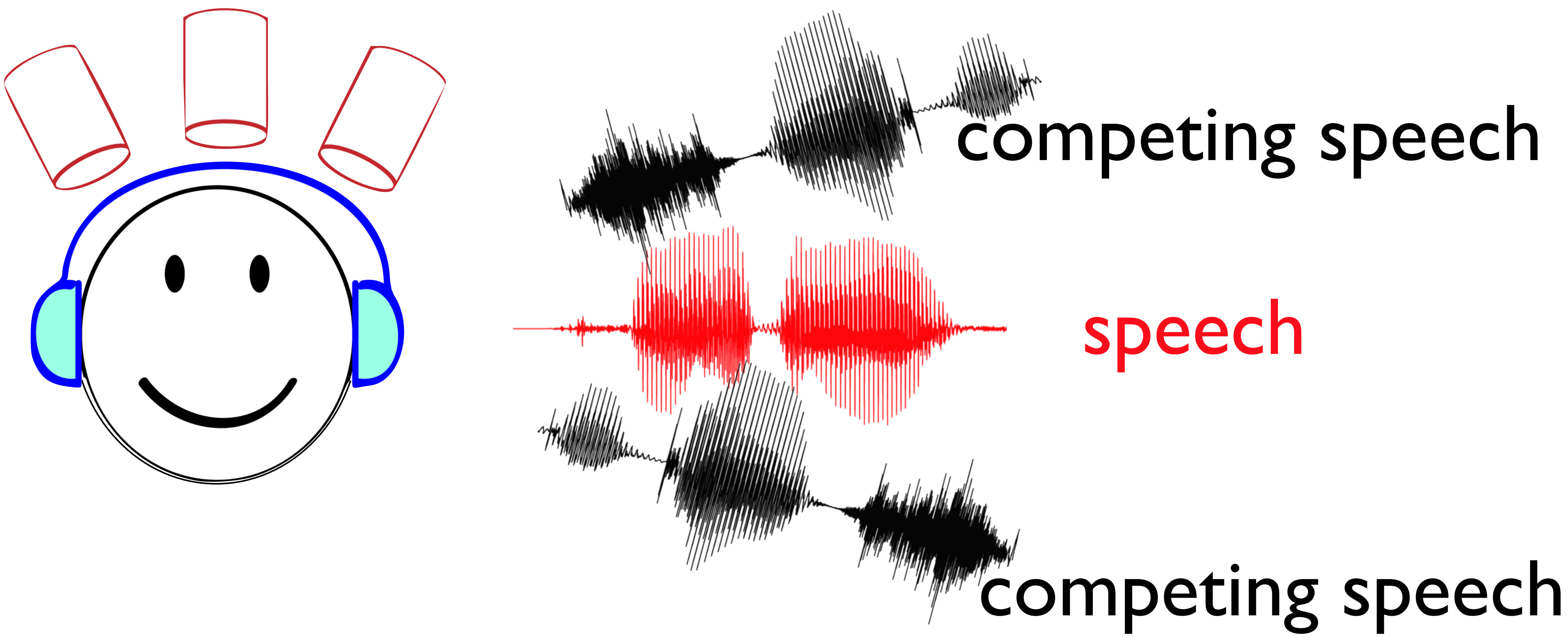
Experiments in Progress



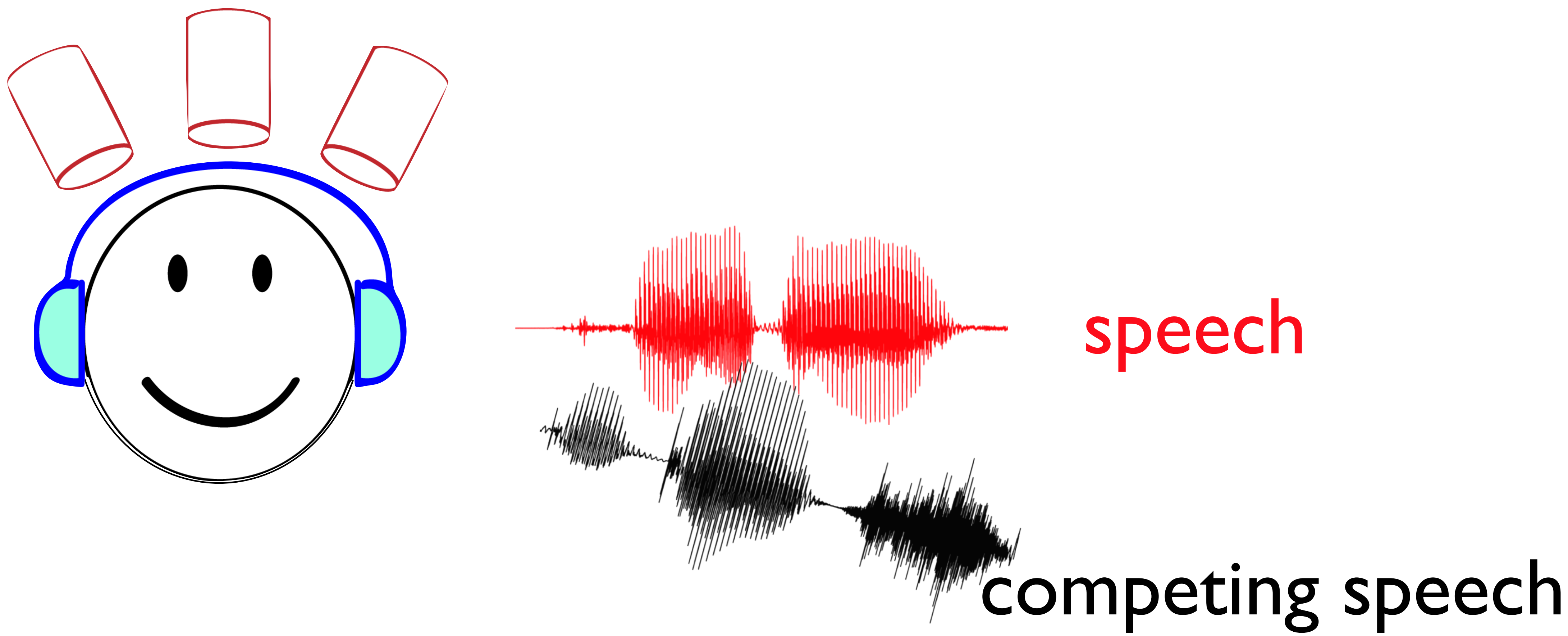
Experiments in Progress



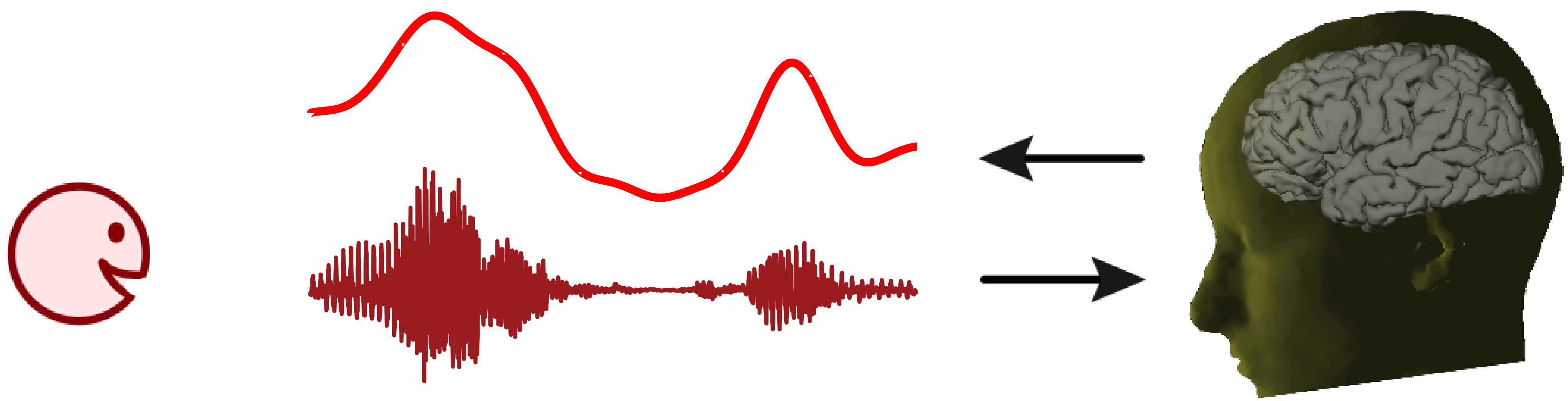
Experiments in Progress



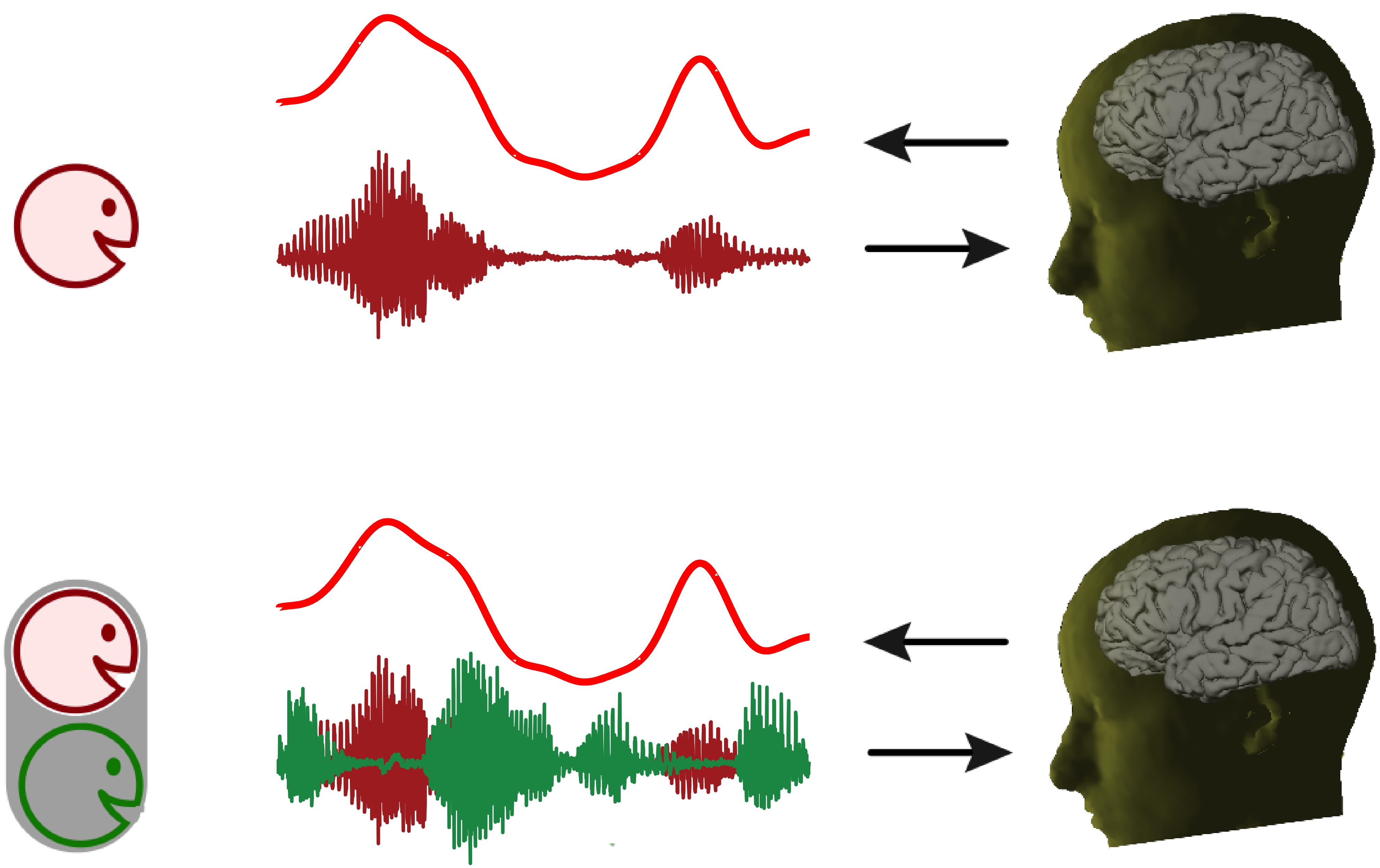
Two Competing Speakers



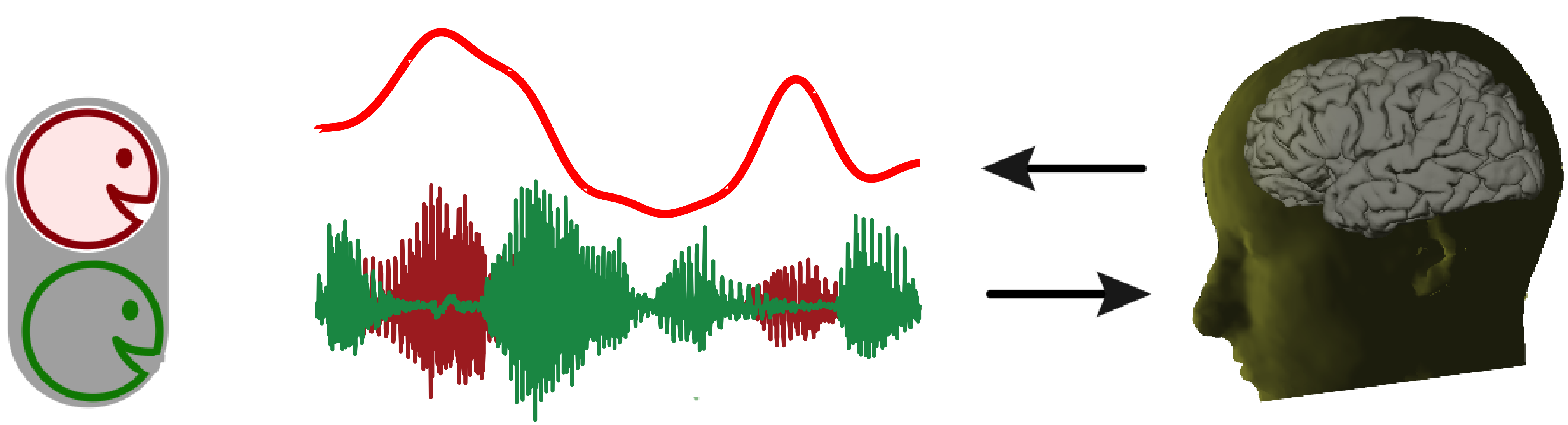
Selective Neural Encoding



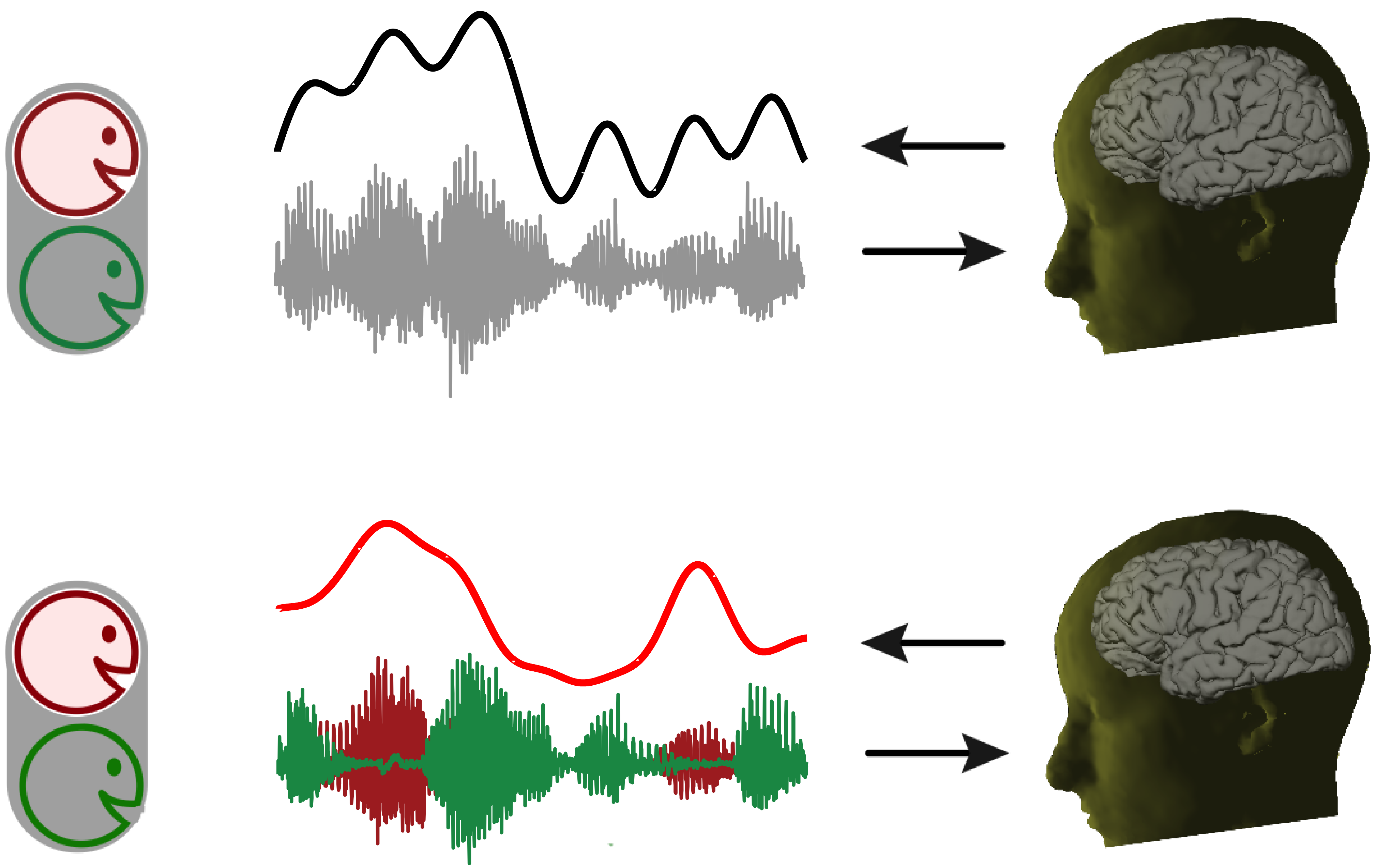
Selective Neural Encoding



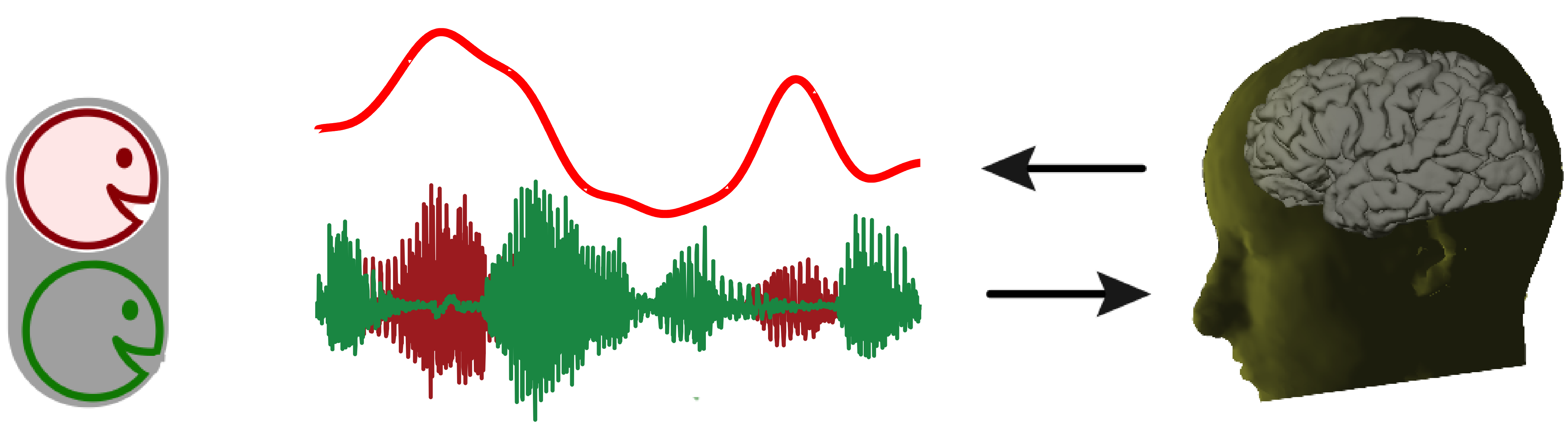
Selective Neural Encoding



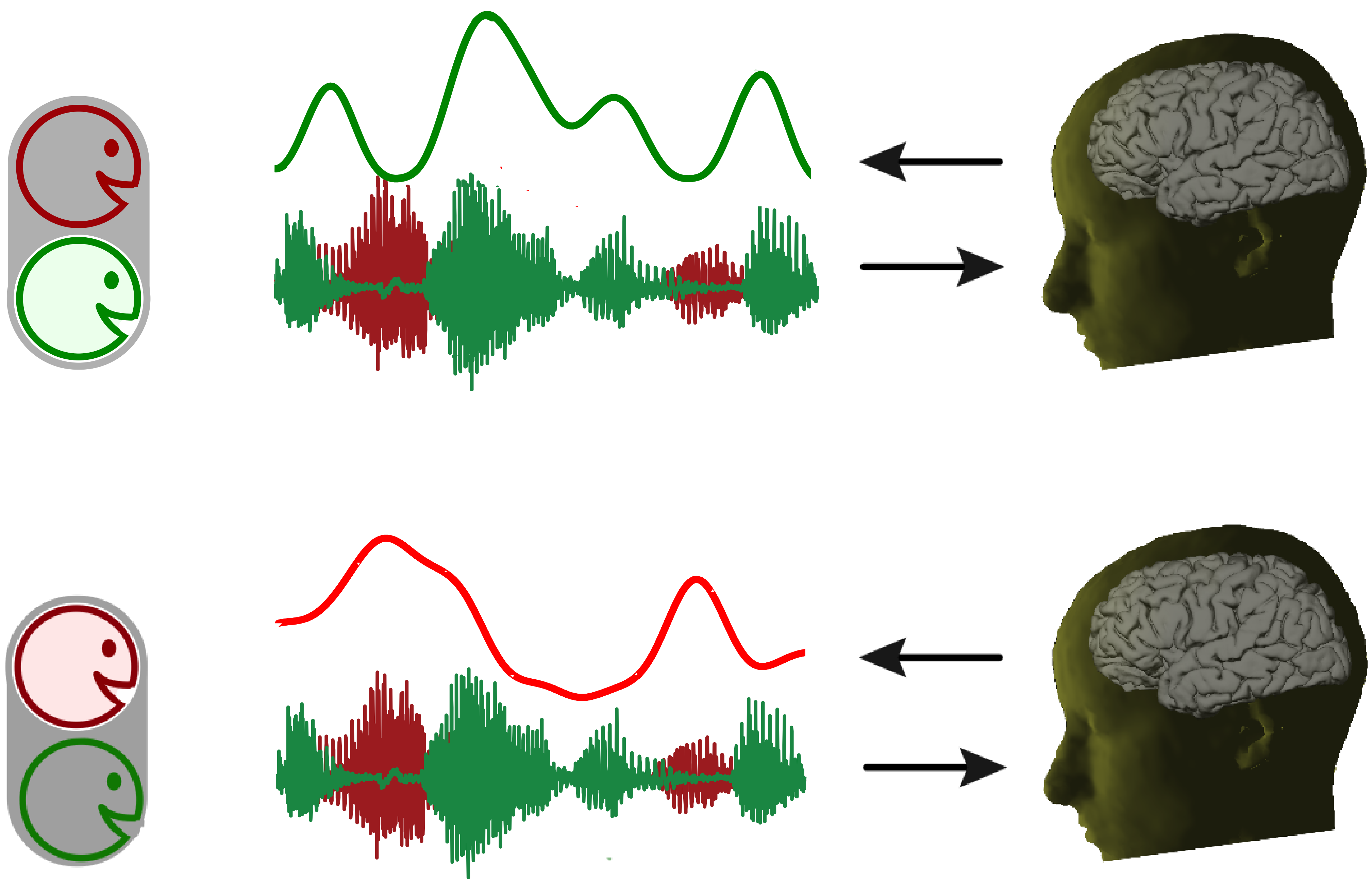
Unselective vs. Selective Neural Encoding



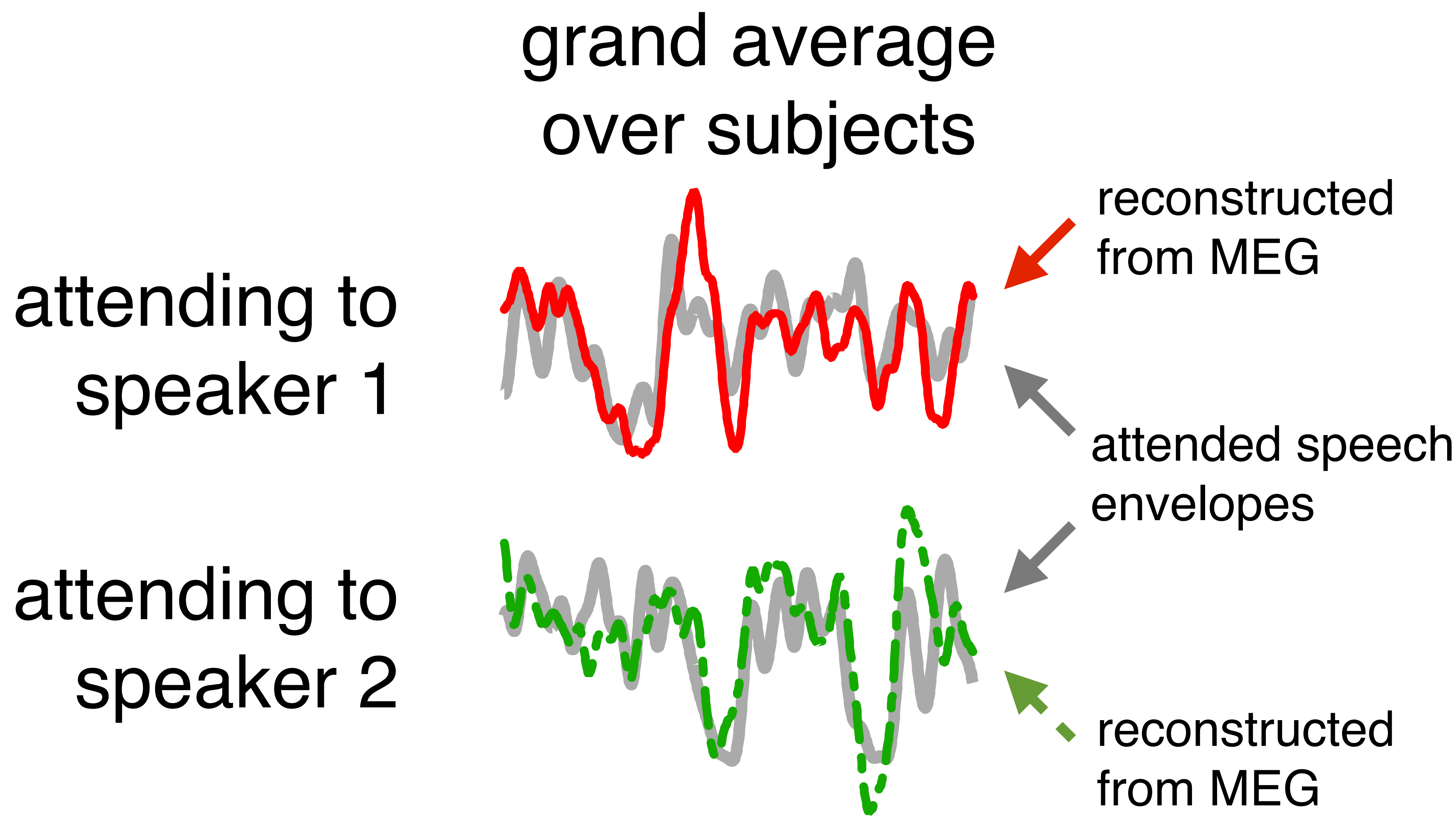
Unselective vs. Selective Neural Encoding



Selective Neural Encoding



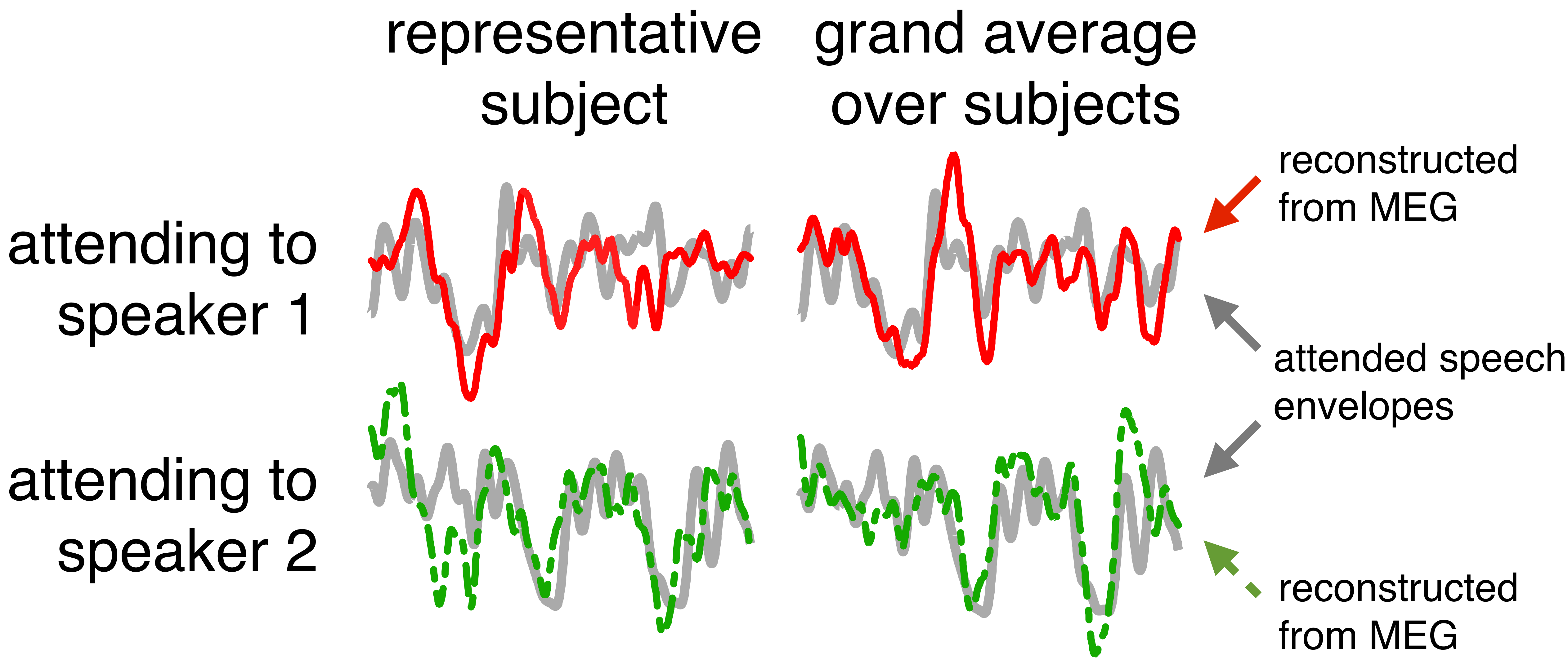
Stream-Specific Representation



Identical Stimuli!

Ding & Simon, PNAS (2012)

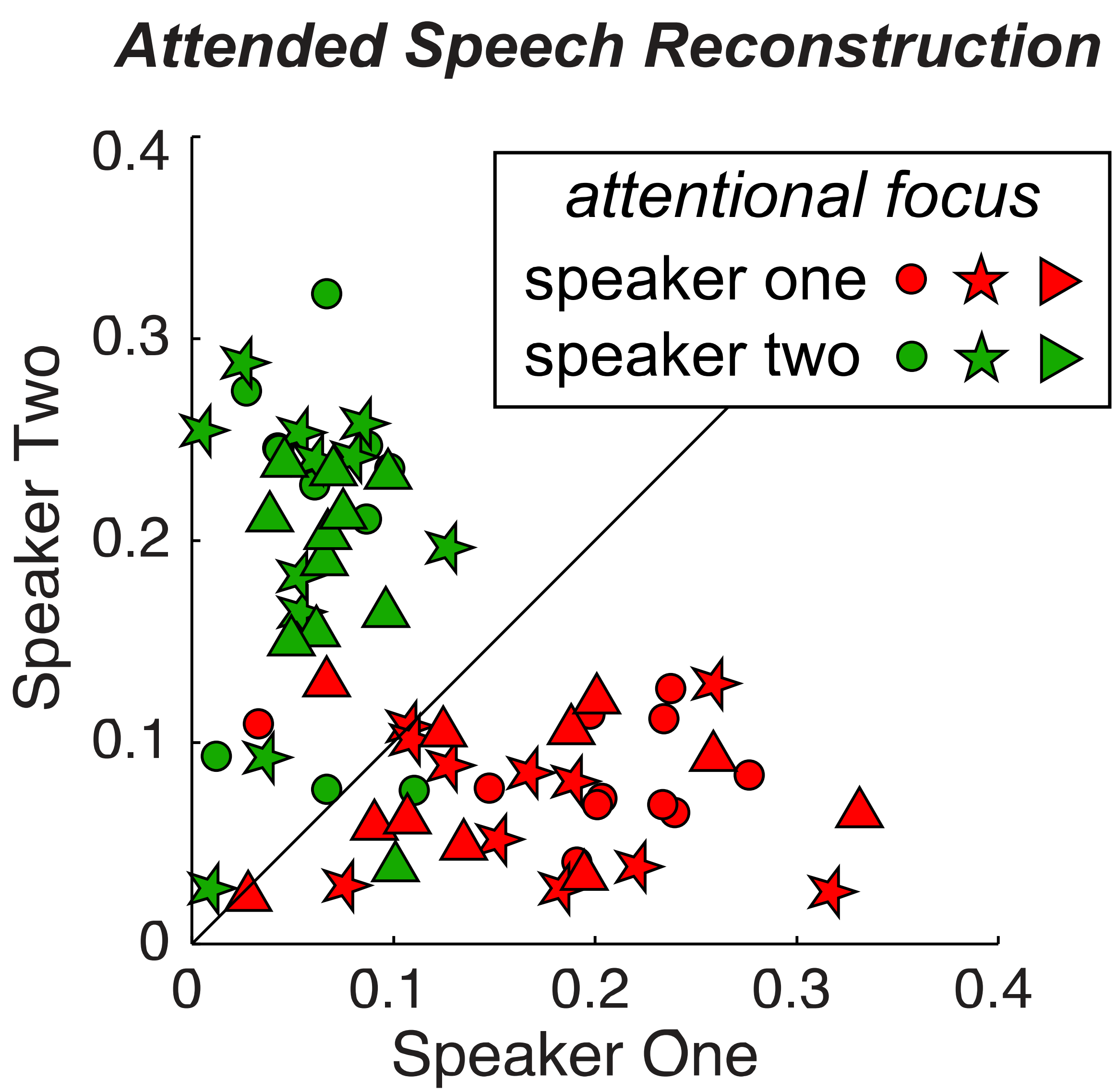
Stream-Specific Representation



Identical Stimuli!

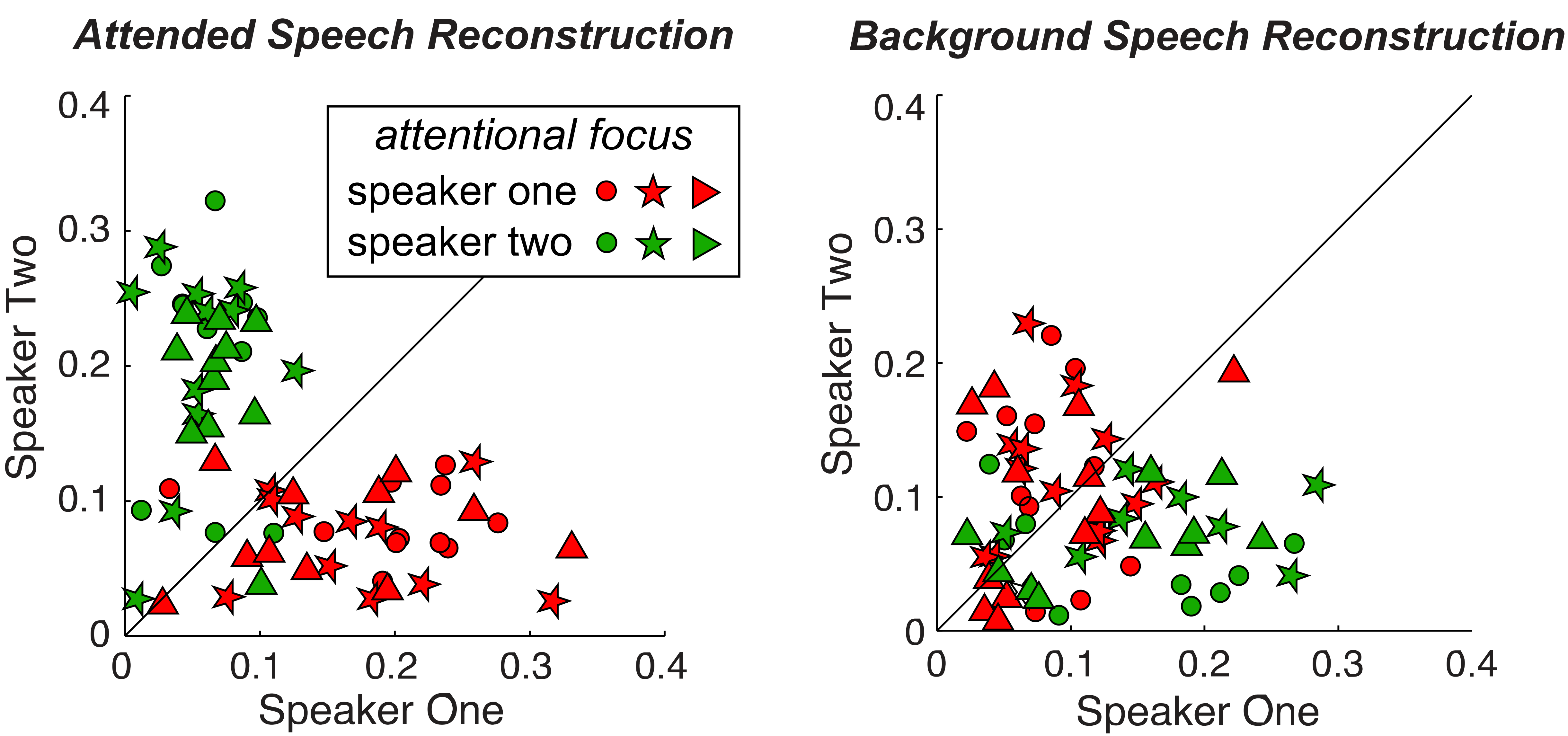
Ding & Simon, PNAS (2012)

Single Trial Speech Reconstruction



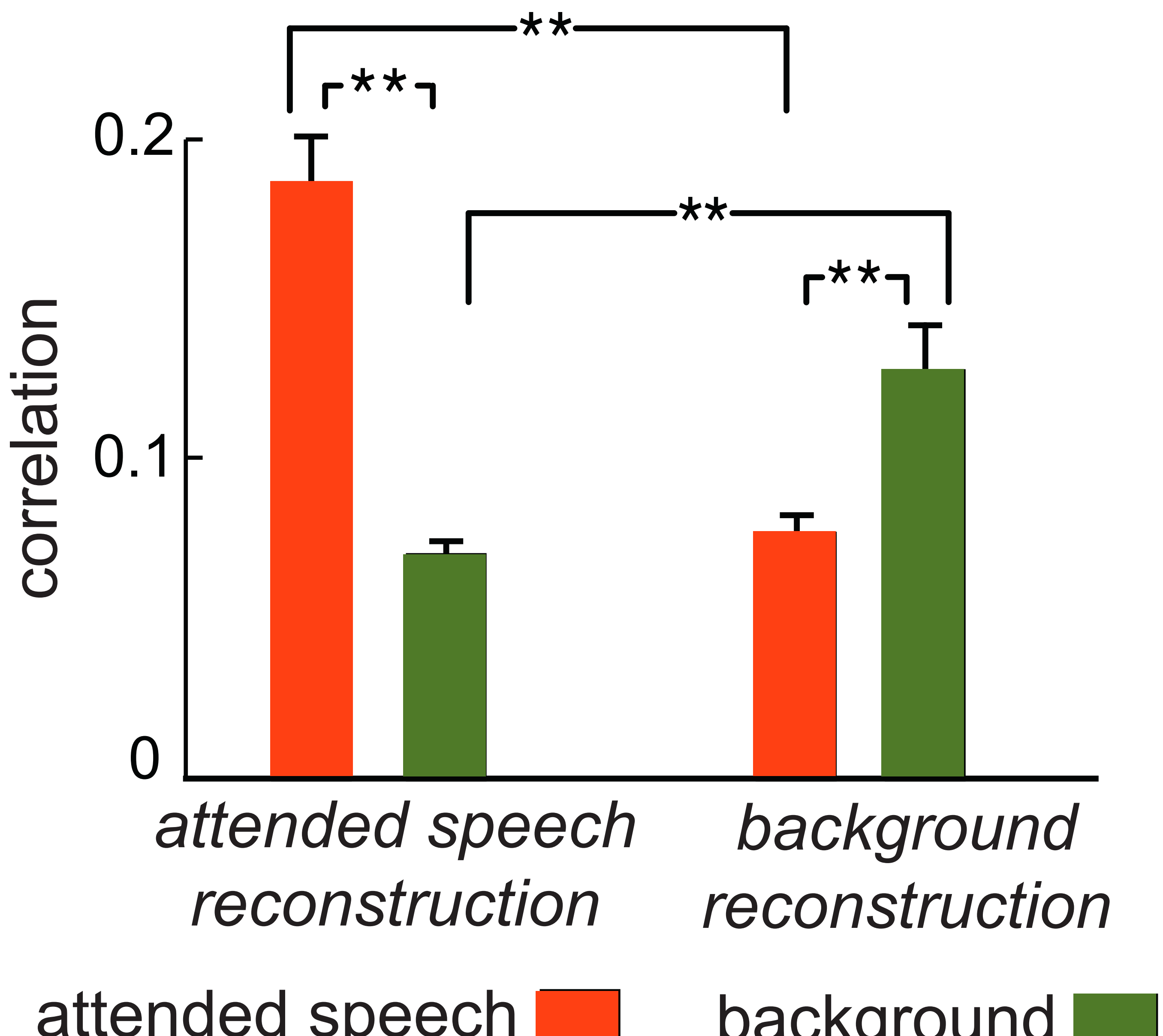
Ding & Simon, PNAS (2012)

Single Trial Speech Reconstruction



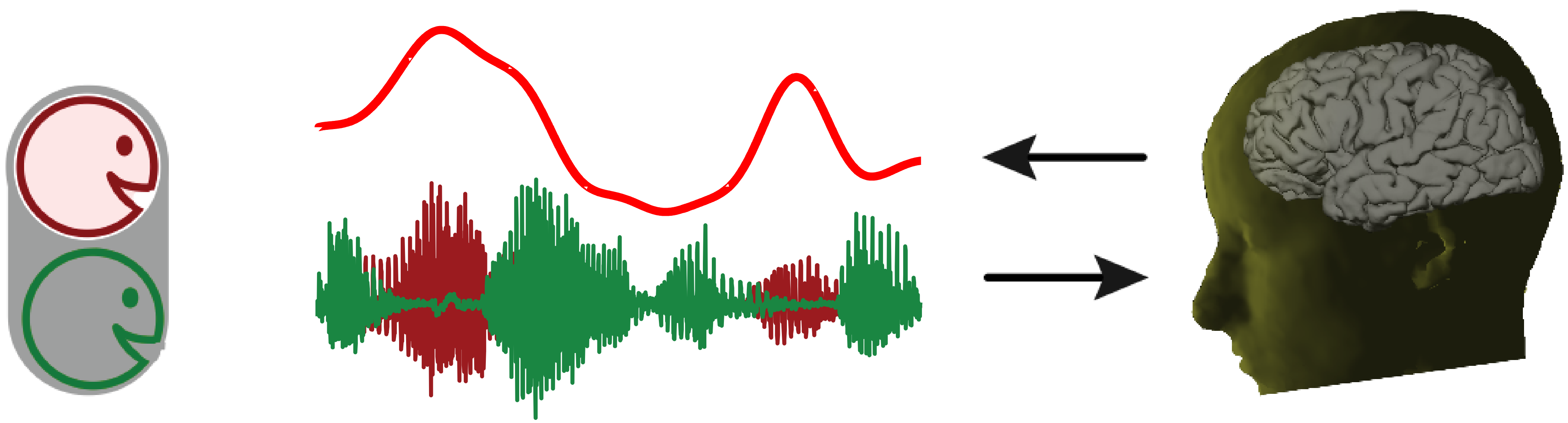
Ding & Simon, PNAS (2013)

Overall Speech Reconstruction

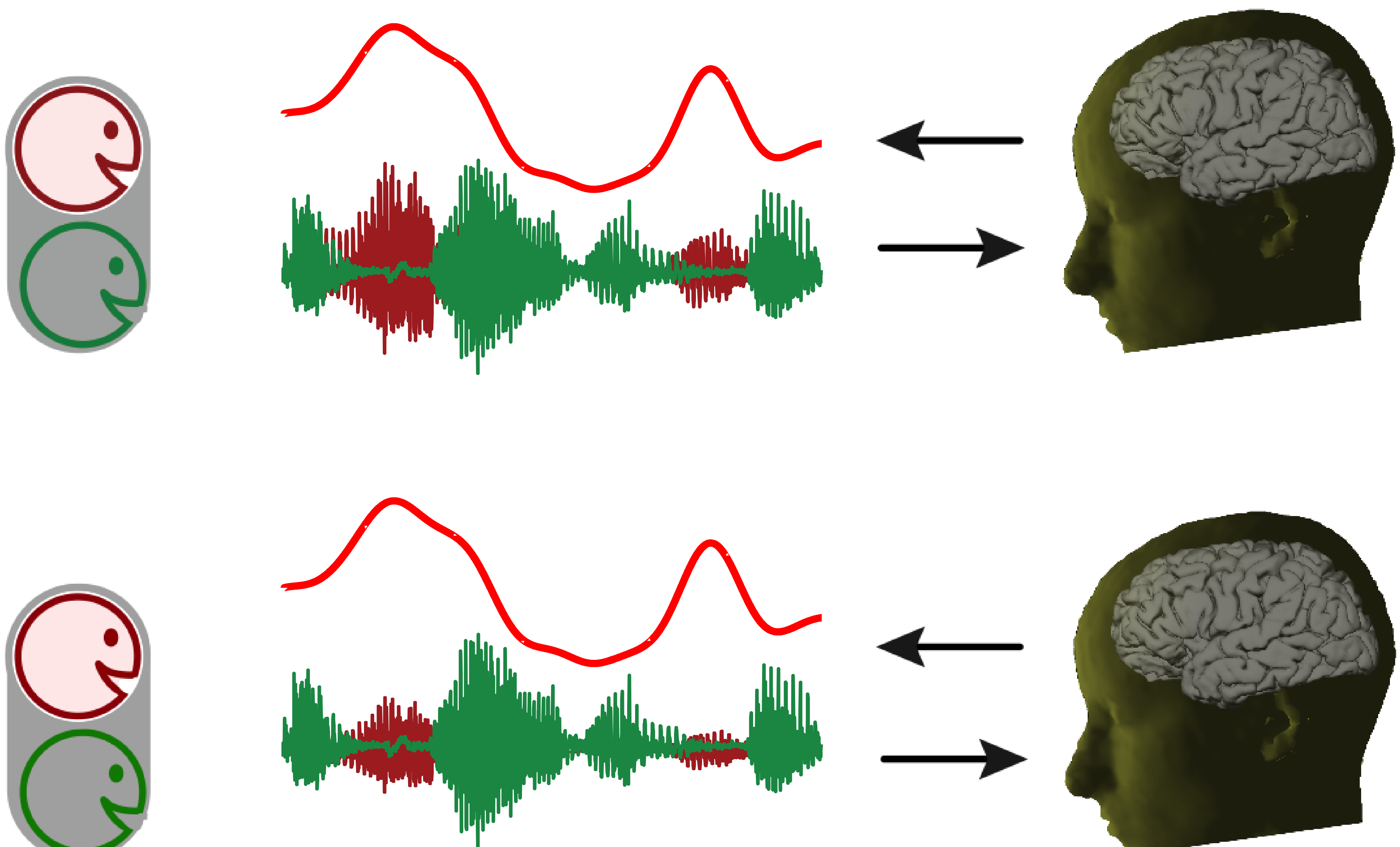


Distinct neural representations for different speech streams

Invariance Under Relative Loudness Change?

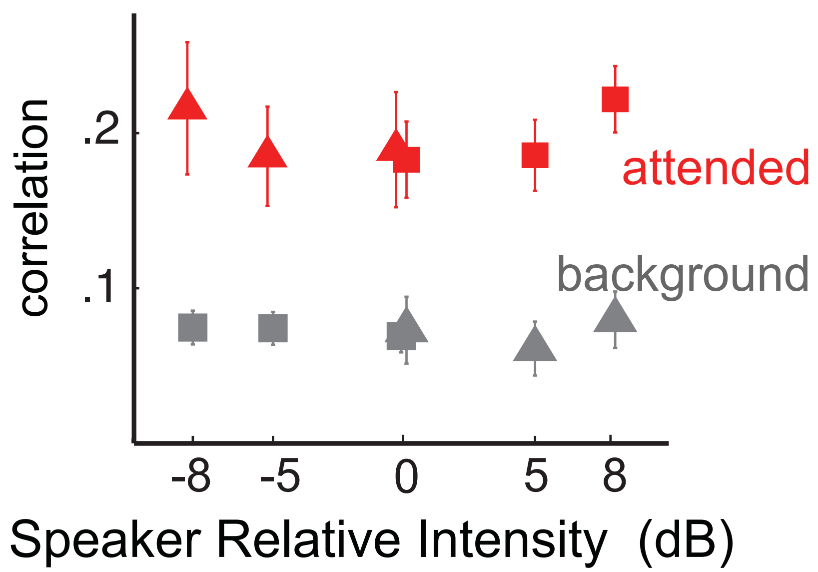


Invariance Under Relative Loudness Change?



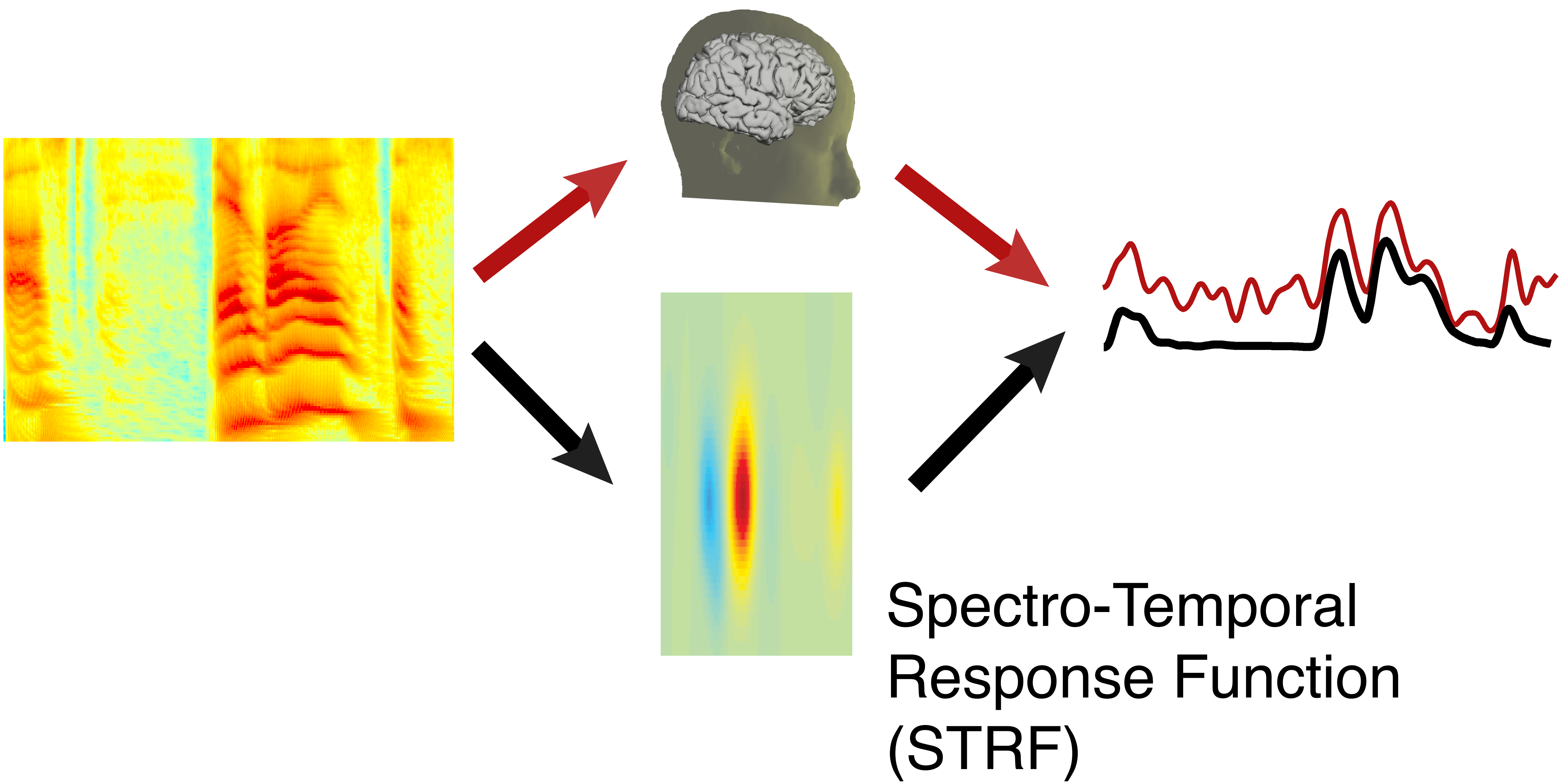
Invariance under Relative Loudness Change

Neural Results

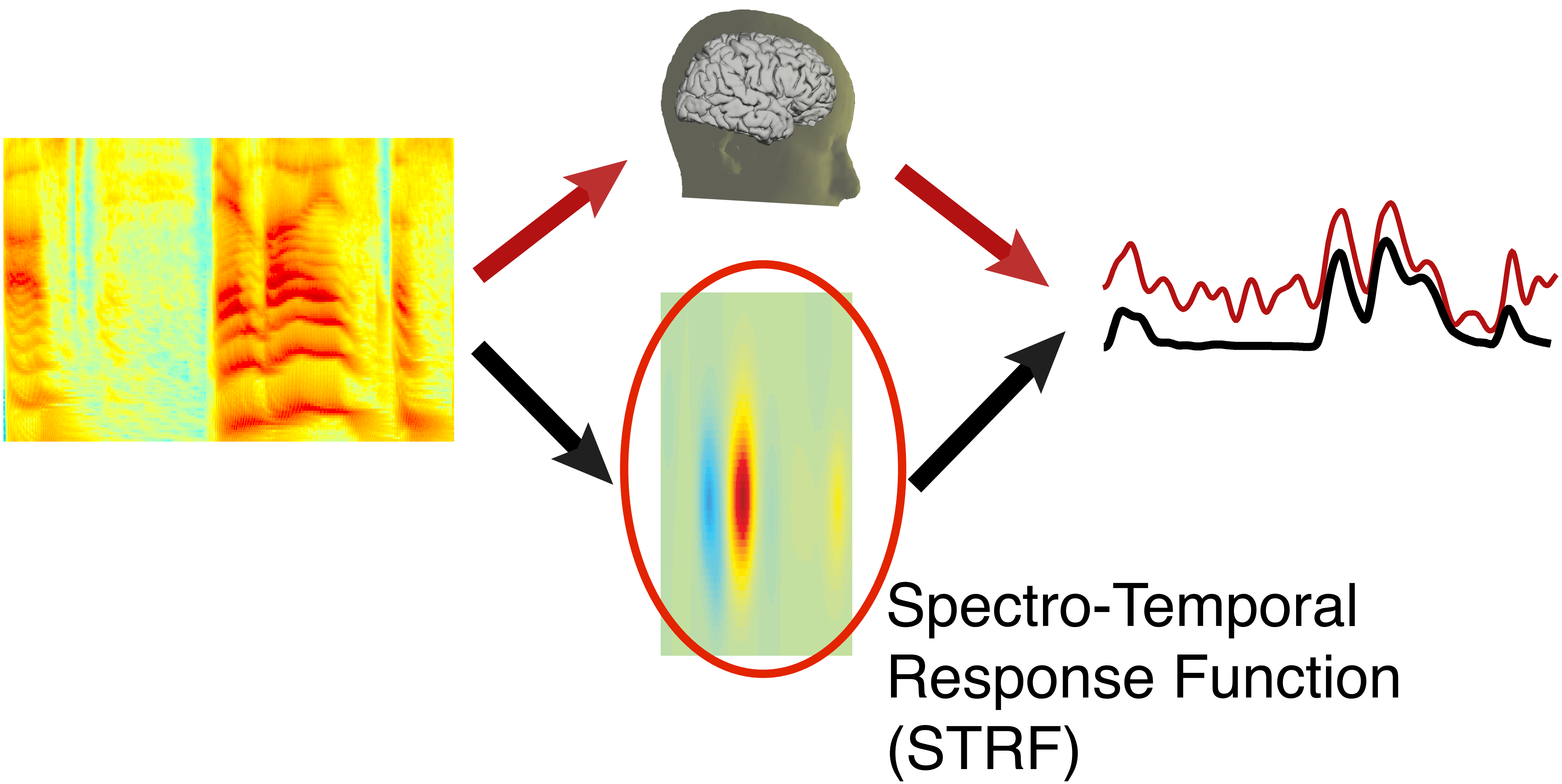


- Neural representation invariant to relative loudness change
- Stream-based Gain Control, not stimulus-based

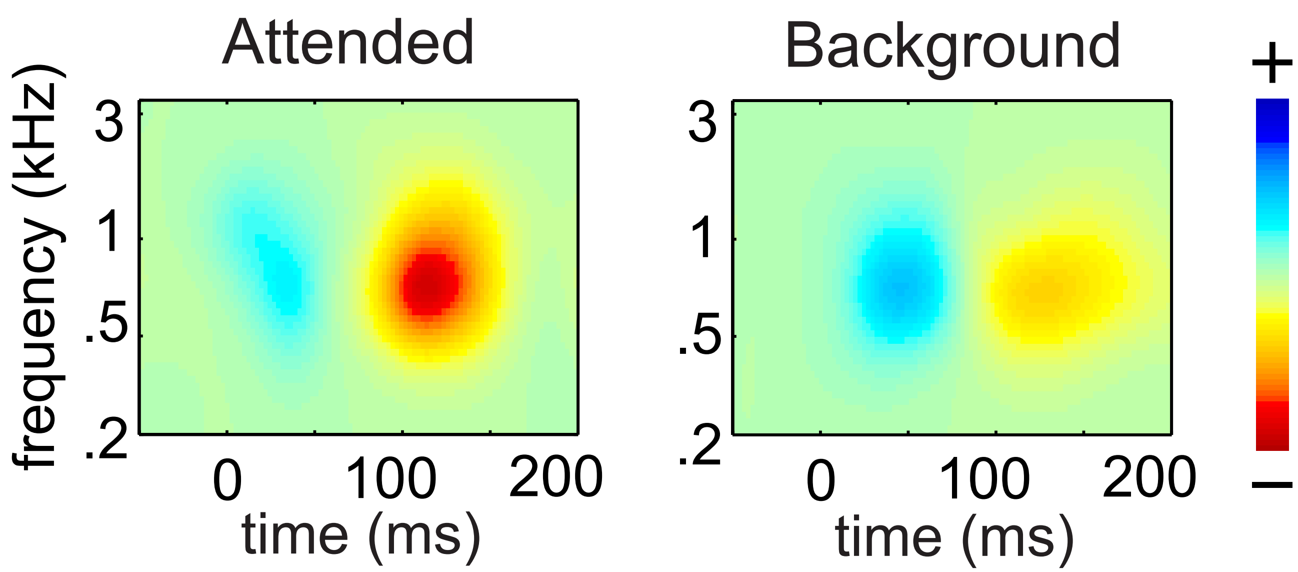
Forward STRF Model



Forward STRF Model

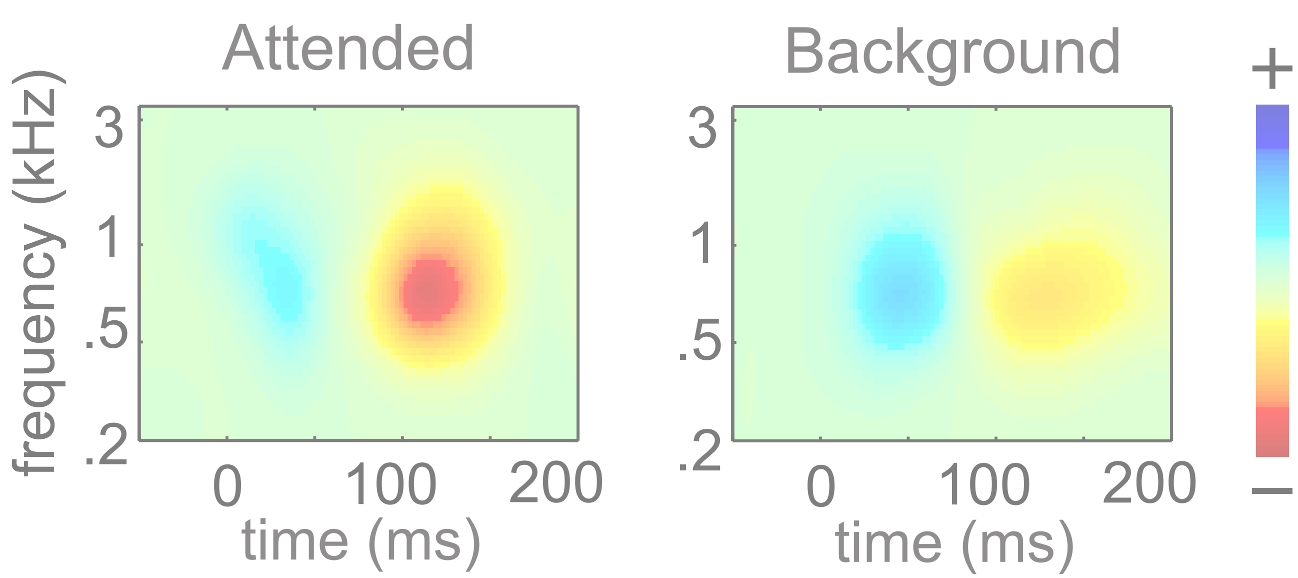


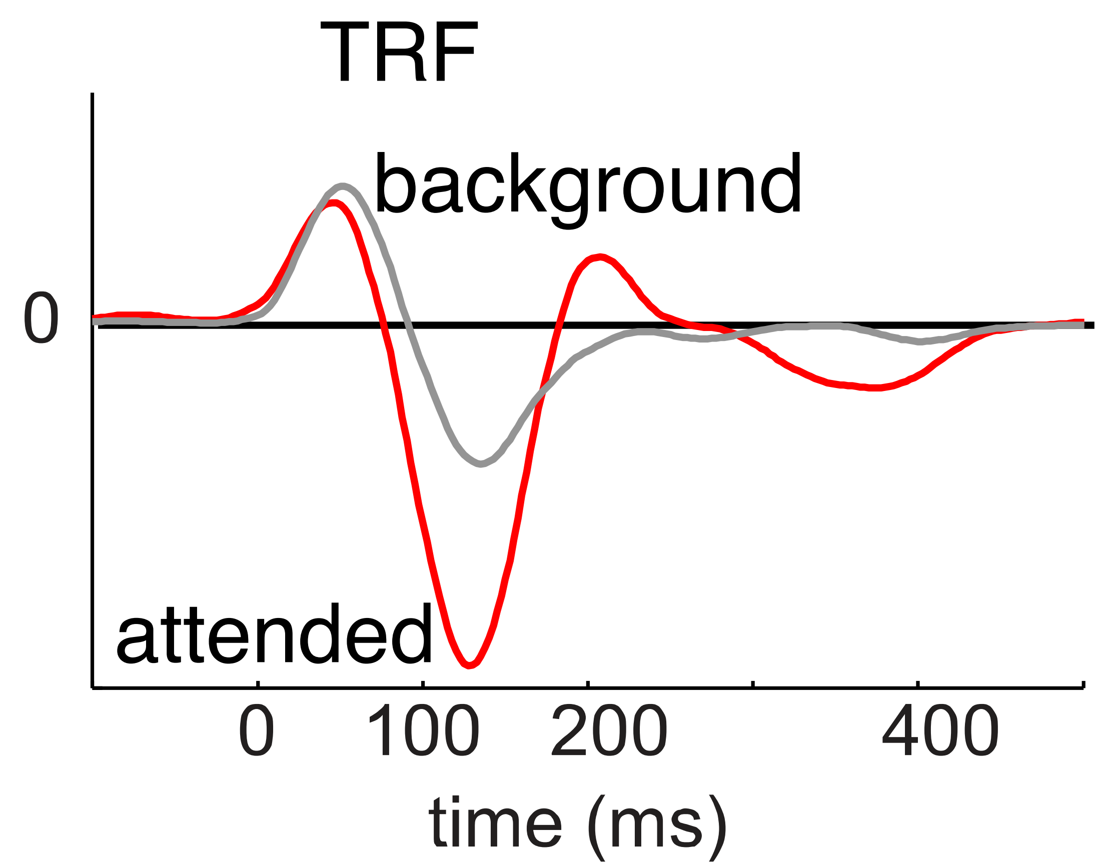
STRF Results



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

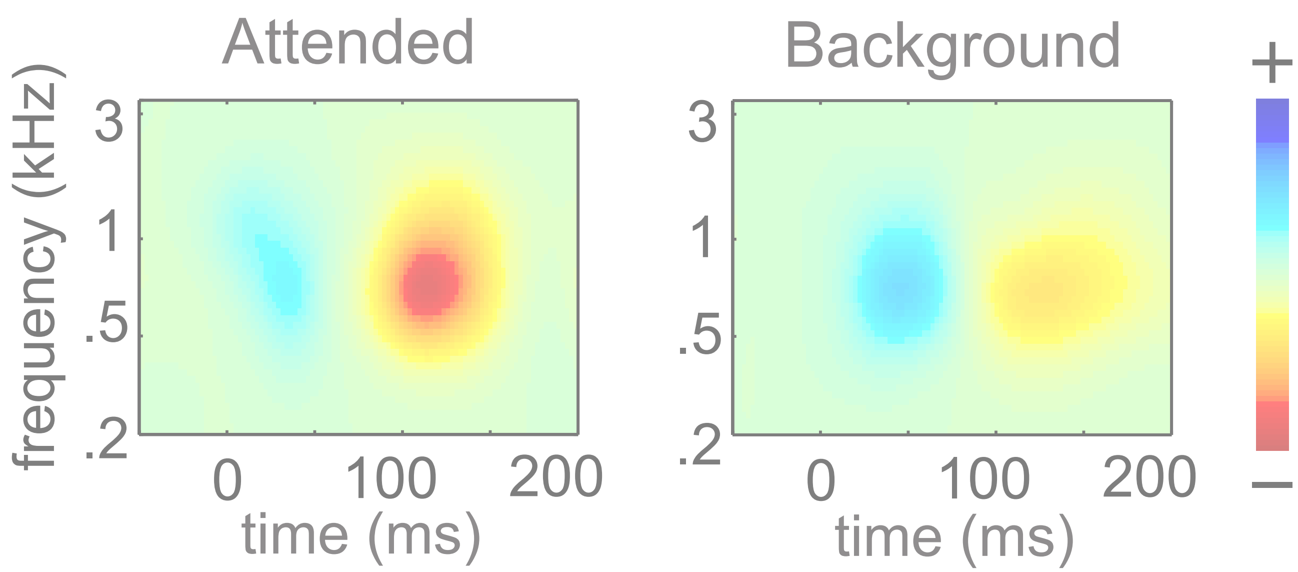
STRF Results

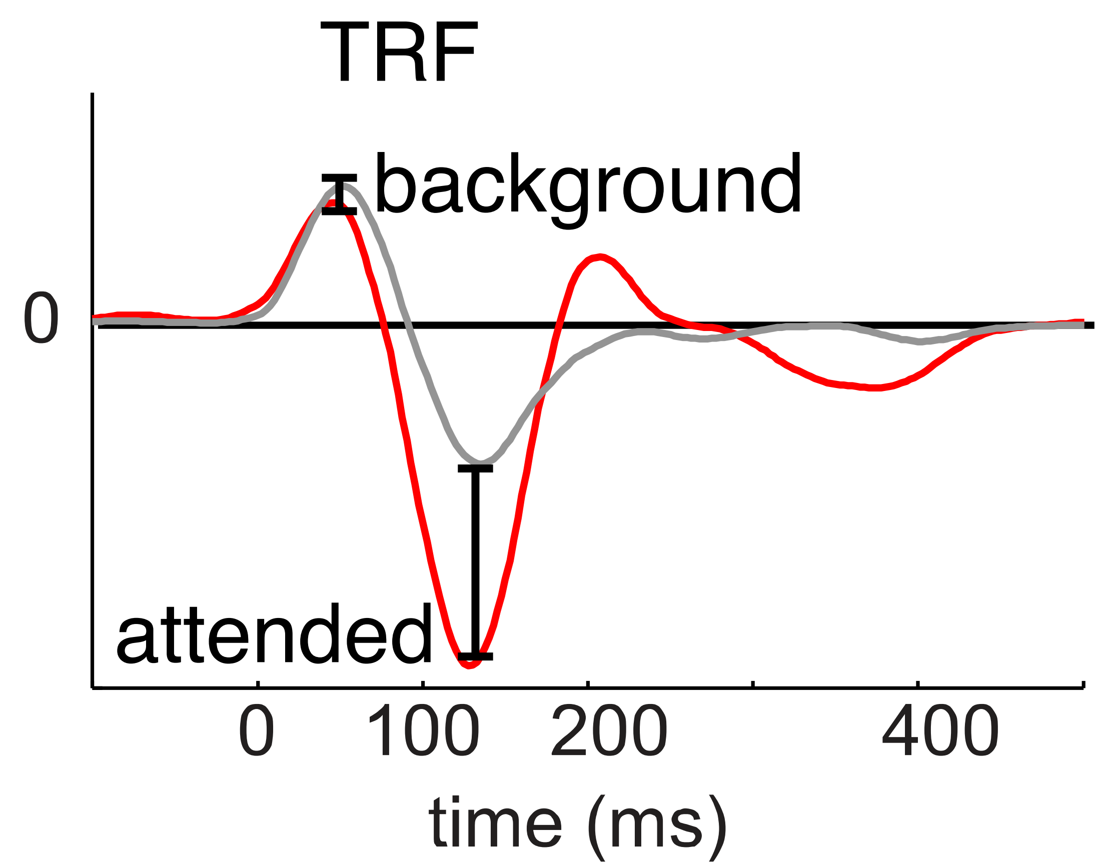




- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- $M50_{STRF}$ positive peak
- $M100_{STRF}$ negative peak

STRF Results

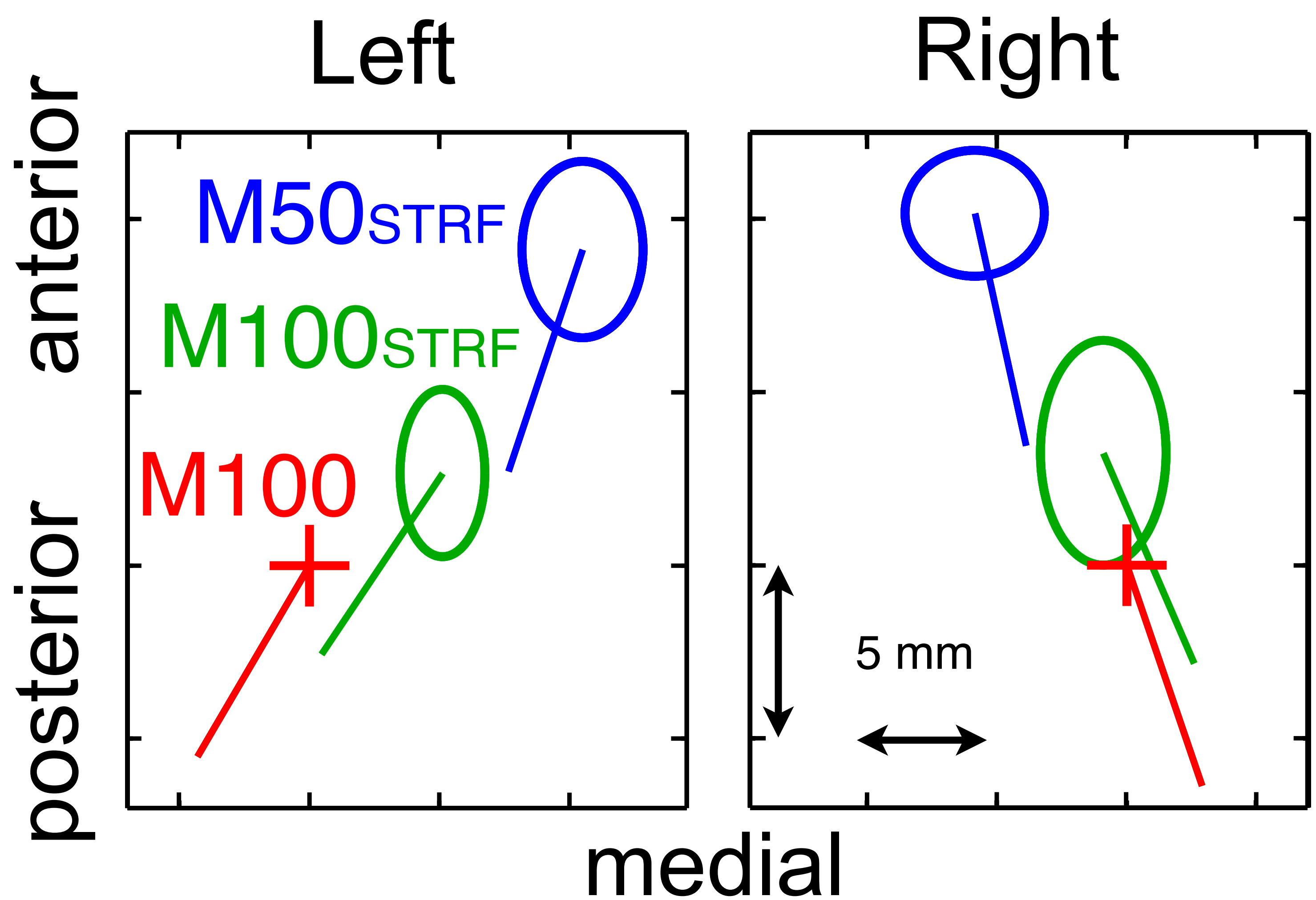




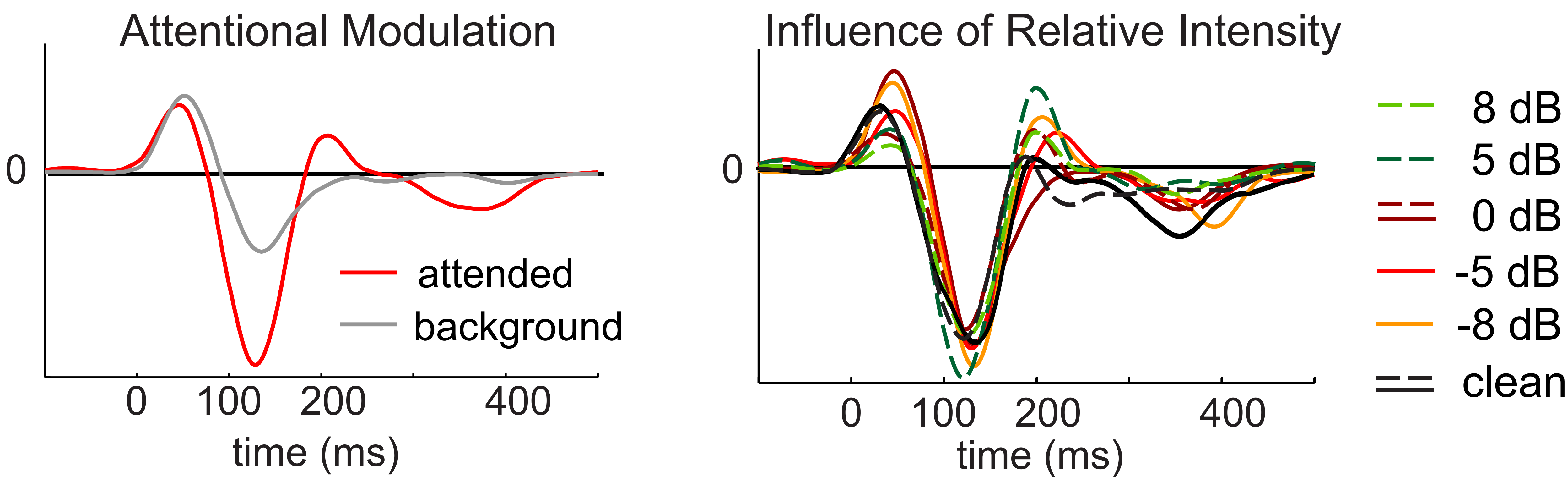
- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- $M50_{STRF}$ positive peak
- $M100_{STRF}$ negative peak
- **$M100_{STRF}$ strongly modulated by attention, *but not $M50_{STRF}$***

Neural Sources

- M100$_{STRF}$ source near (same as?) M100 source:
  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***



Cortical Object-Processing Hierarchy



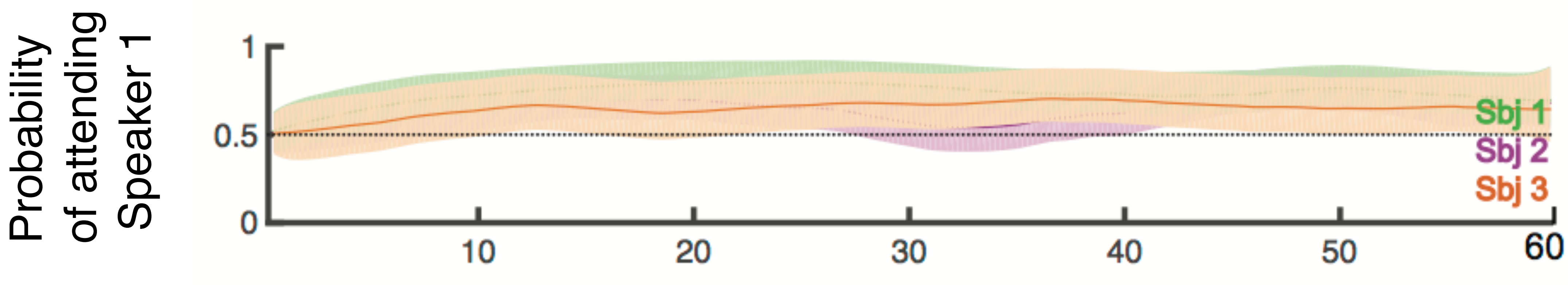
- $M100_{STRF}$ strongly modulated by attention, but not $M50_{STRF}$.
- $M100_{STRF}$ invariant against acoustic changes.
- Objects well-neurally represented at 100 ms, but not 50 ms.

Studies In Progress

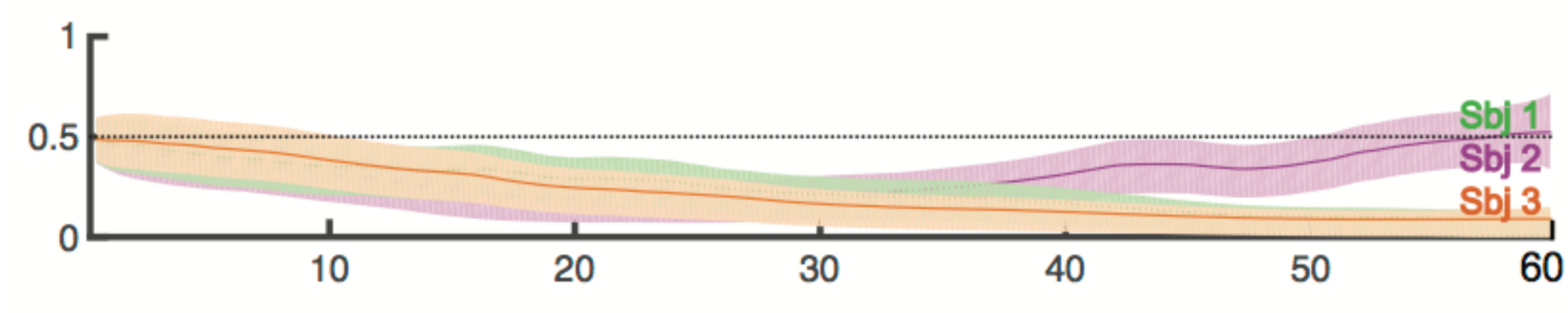
- Attentional Dynamics
- Aging & Neural Representations of Speech
- Neural Representations of the Background

Attentional Dynamics

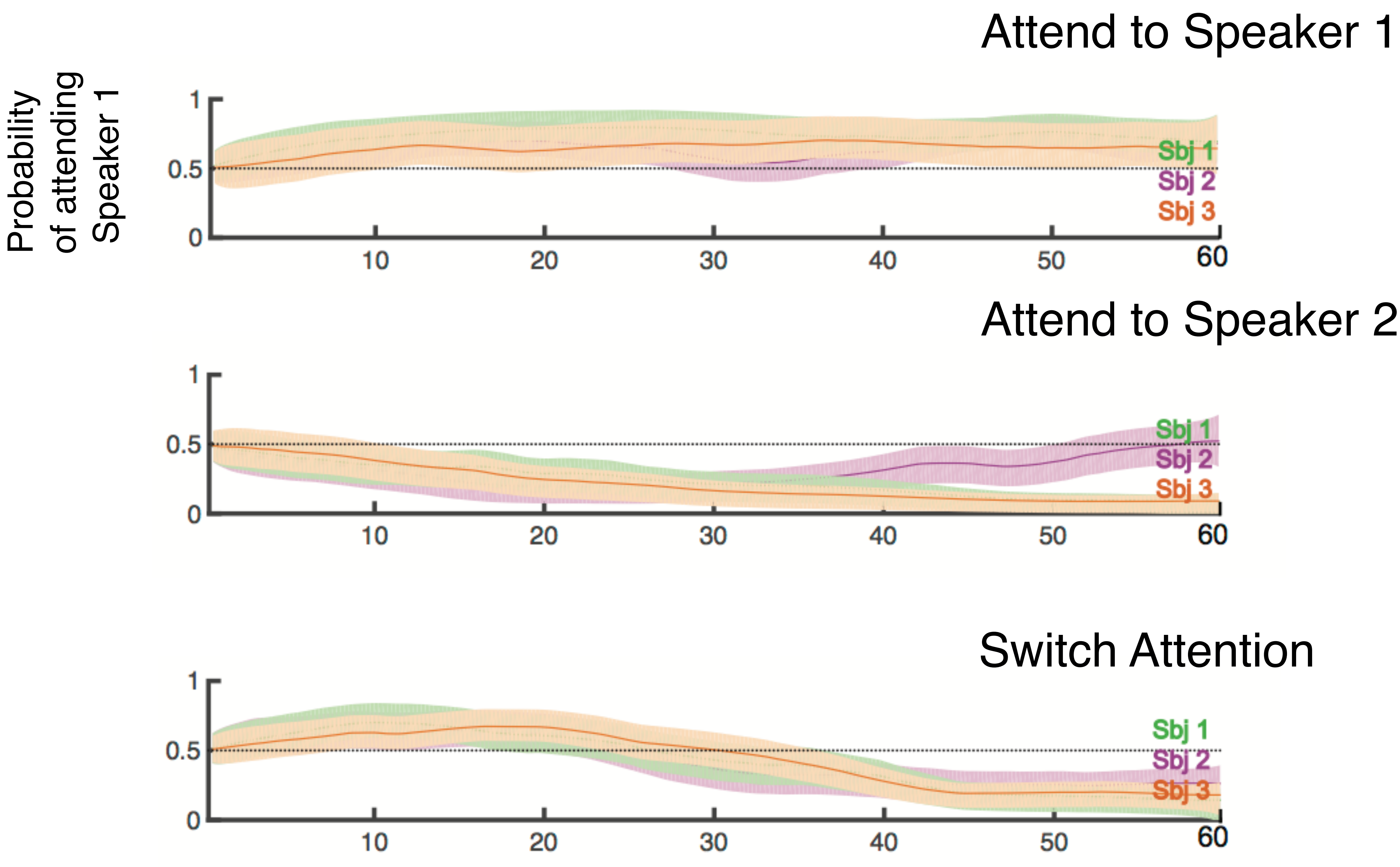
Attend to Speaker 1



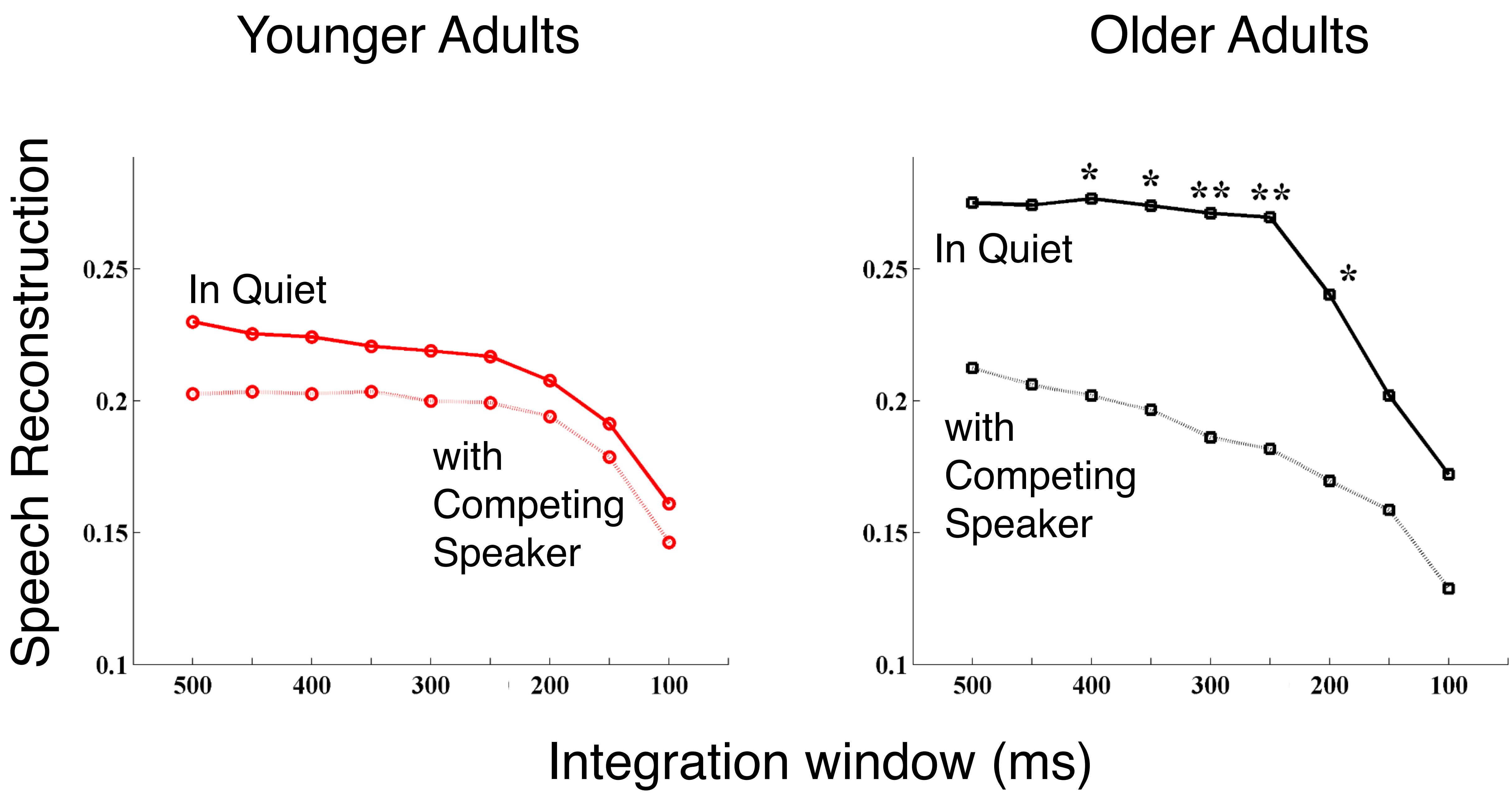
Attend to Speaker 2



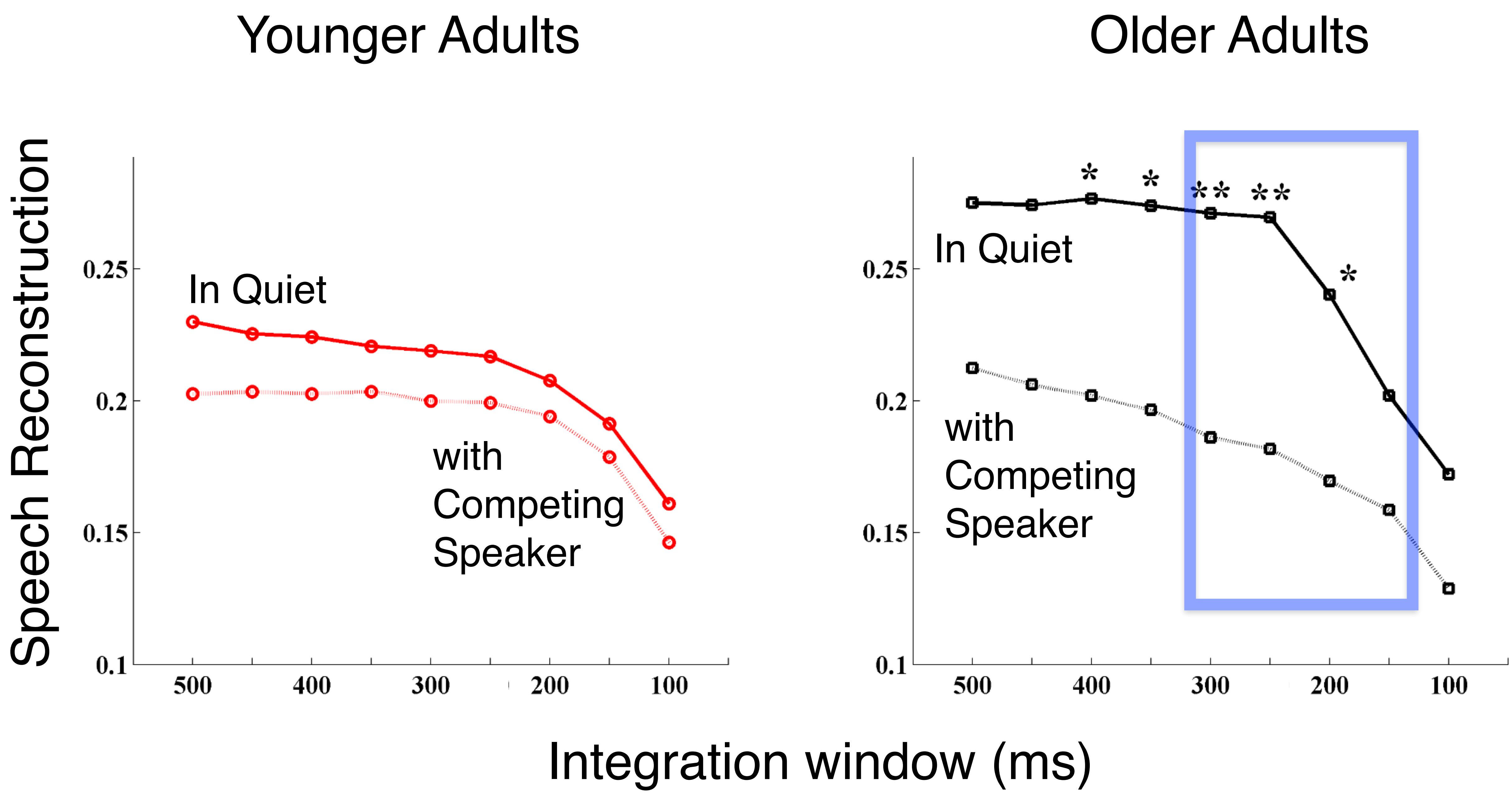
Attentional Dynamics



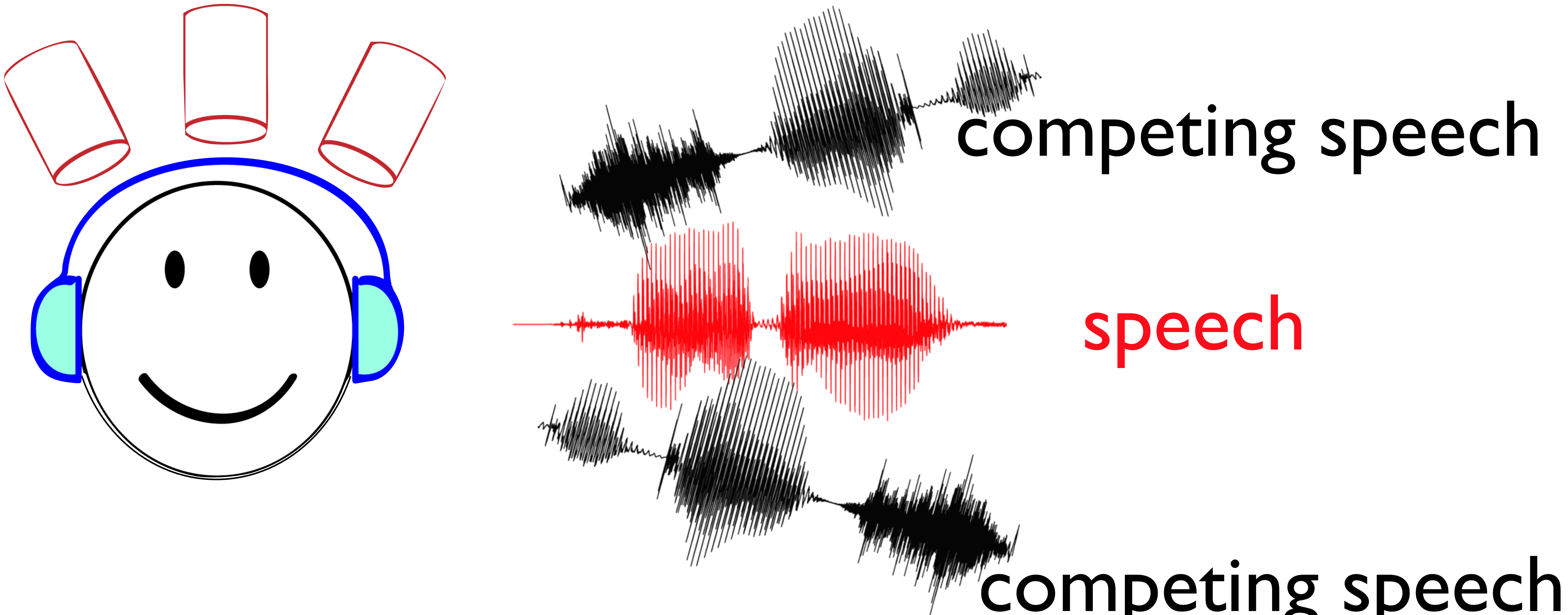
Younger vs. Older Listeners



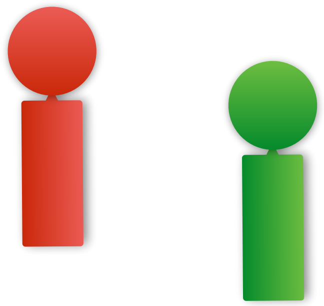
Younger vs. Older Listeners

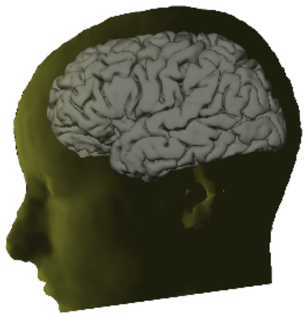


Three Competing Speakers

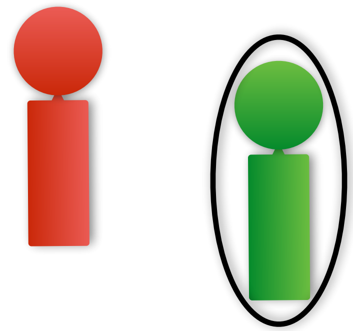


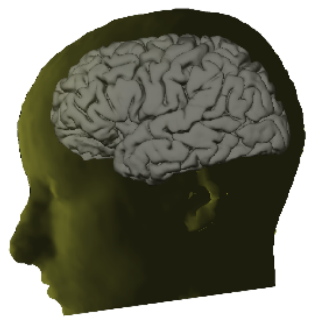
Foreground vs. Background



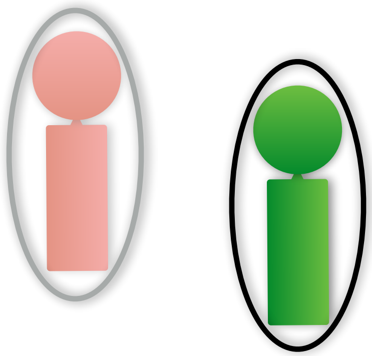


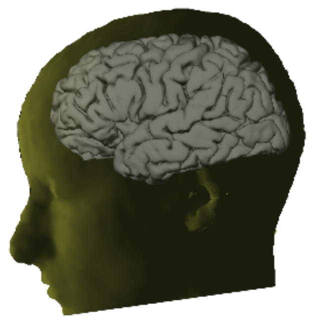
Foreground vs. Background



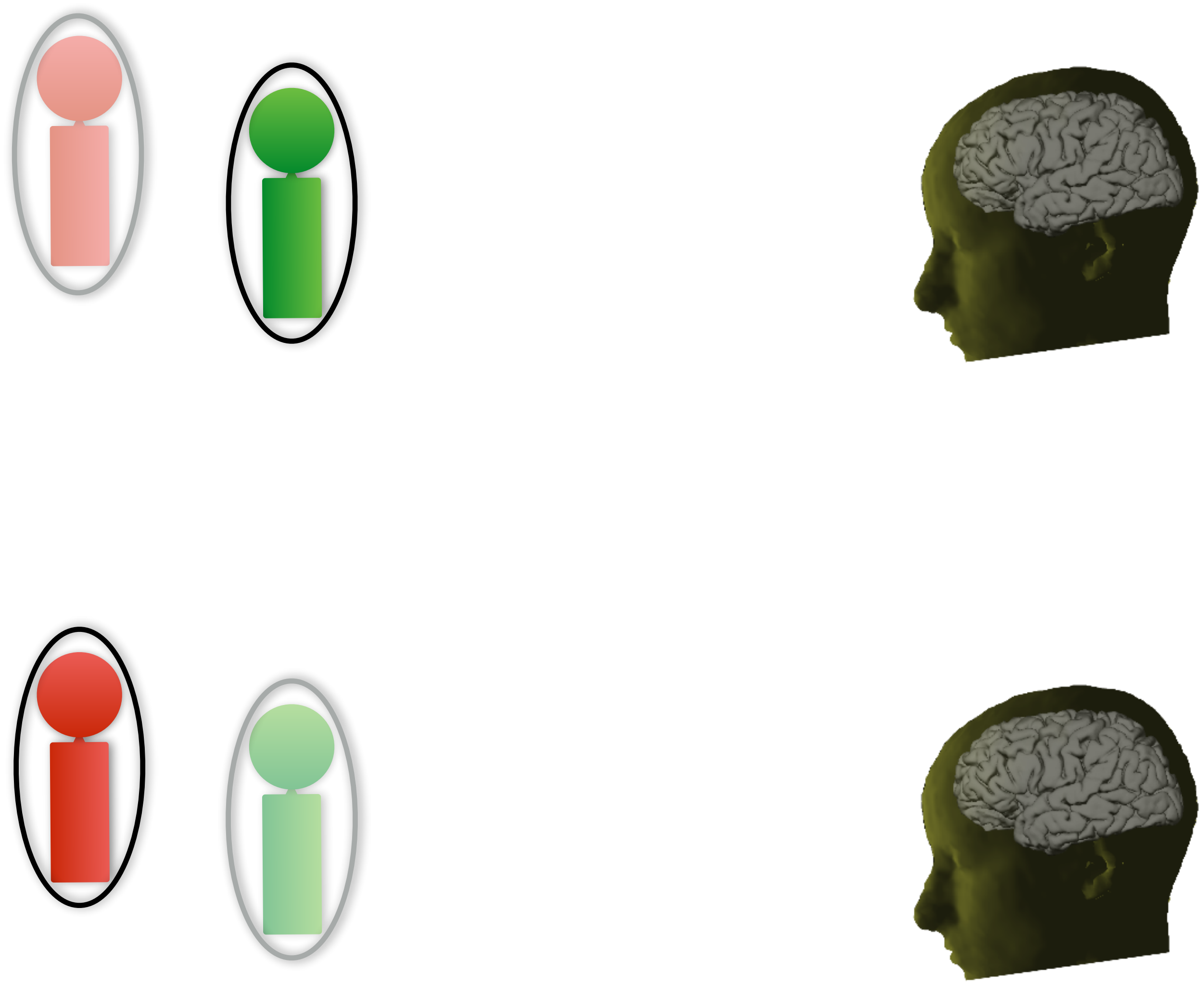


Foreground vs. Background

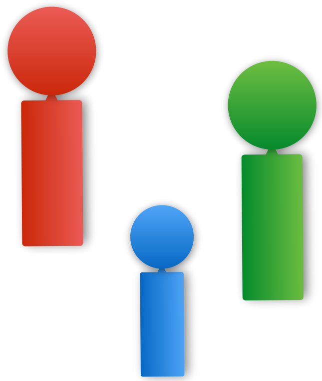


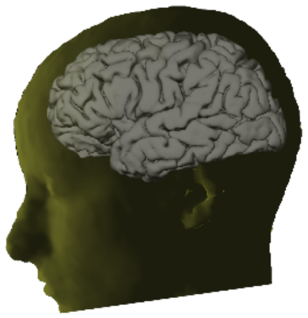


Foreground vs. Background

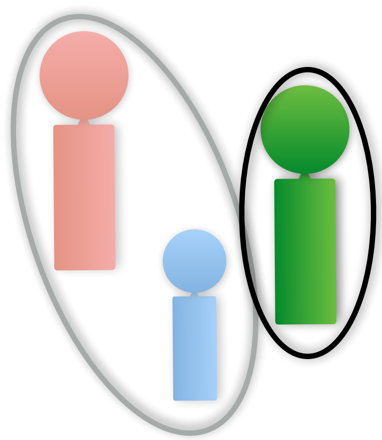


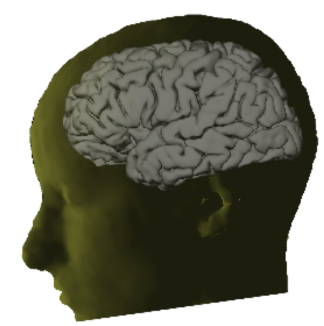
Foreground vs. Background



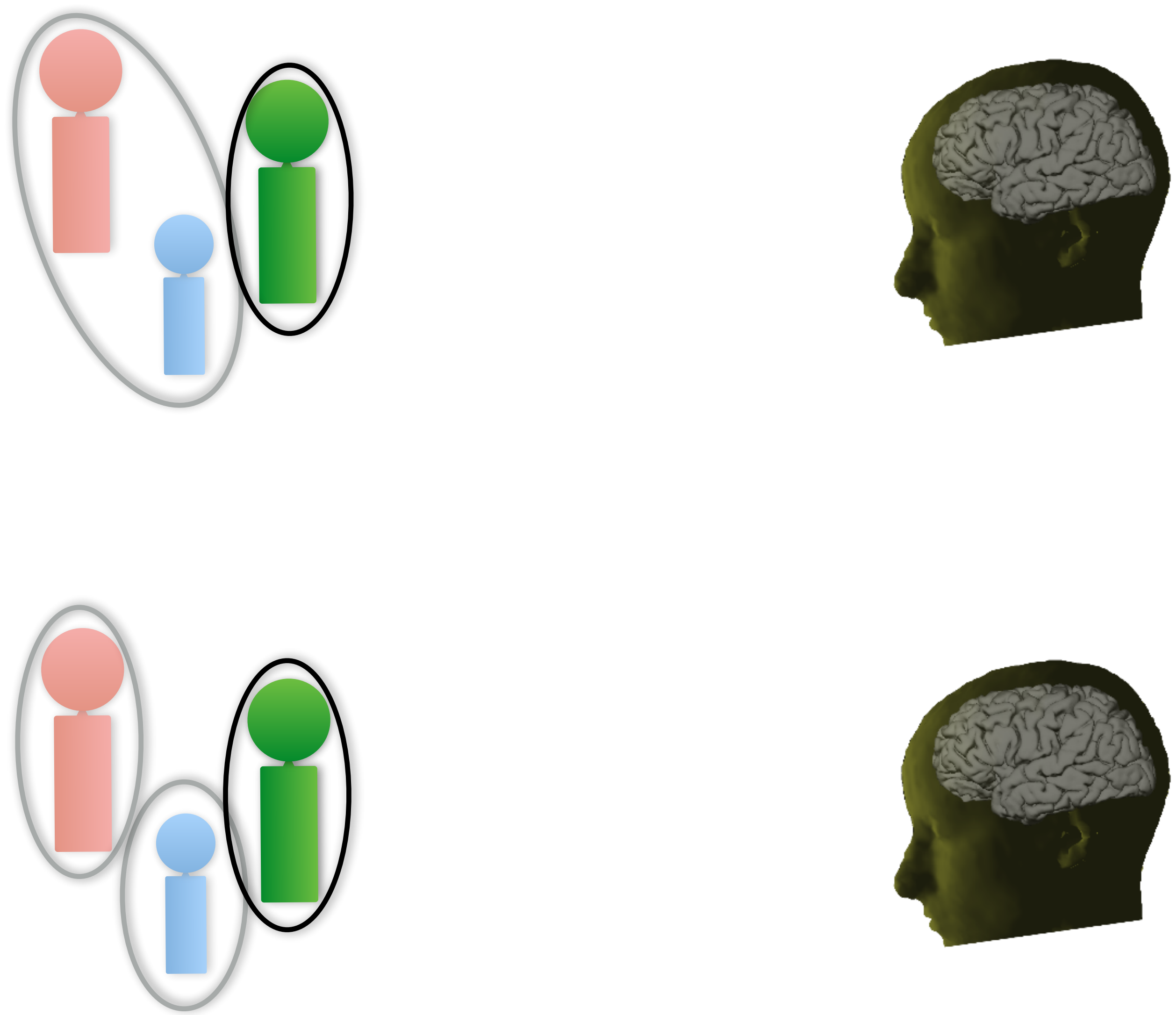


Foreground vs. Background

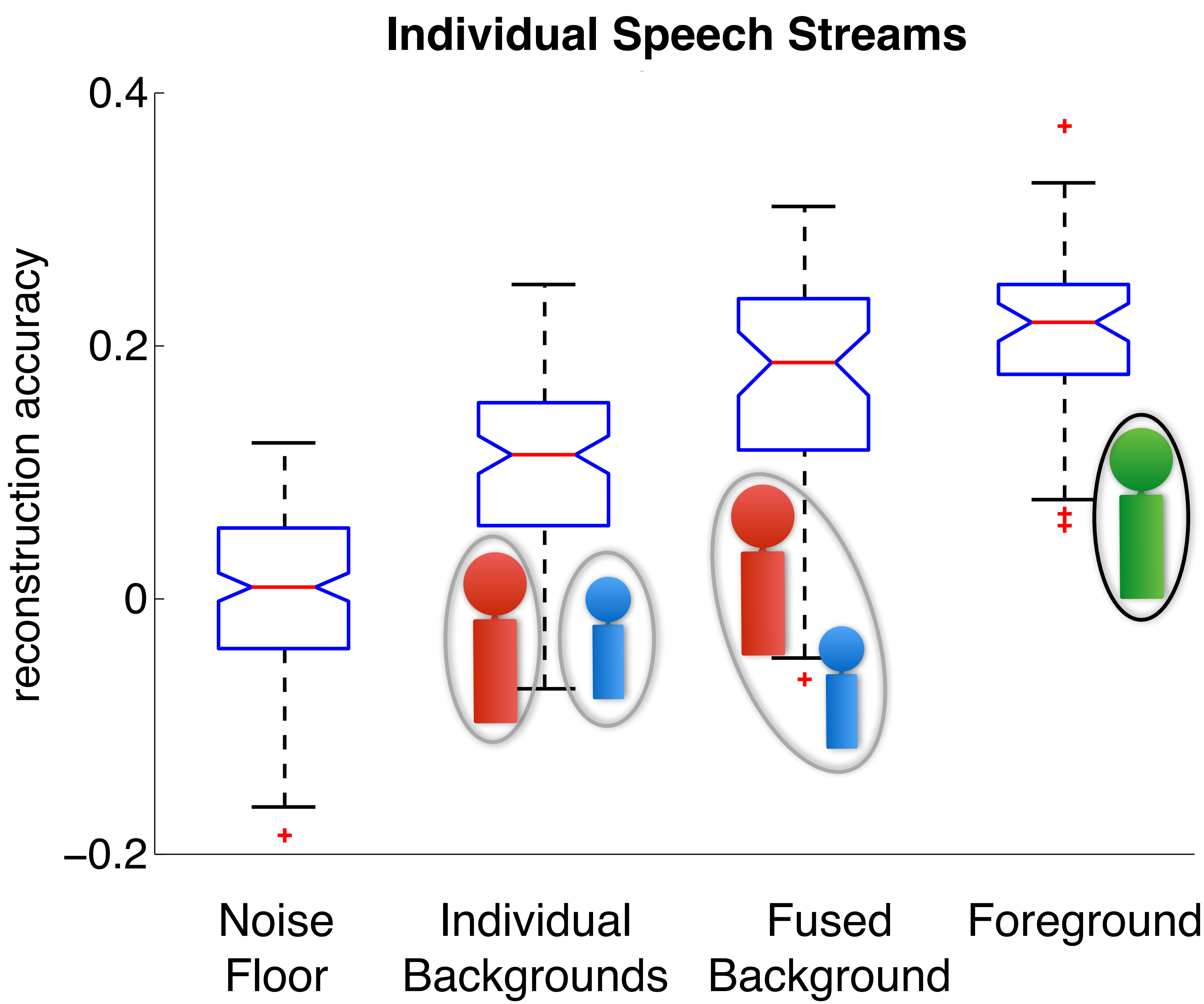




Foreground vs. Background

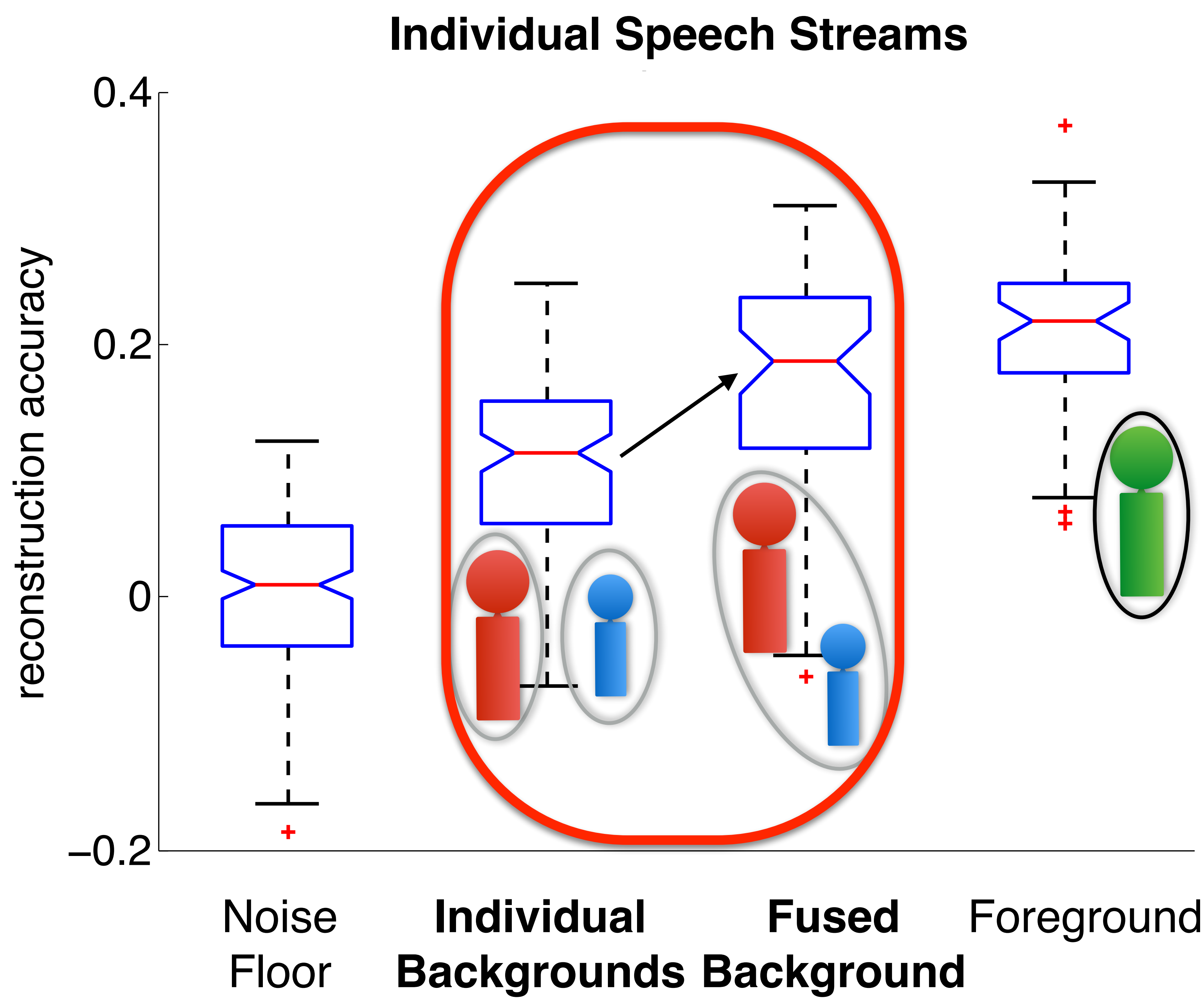


Backgrounds vs. Background



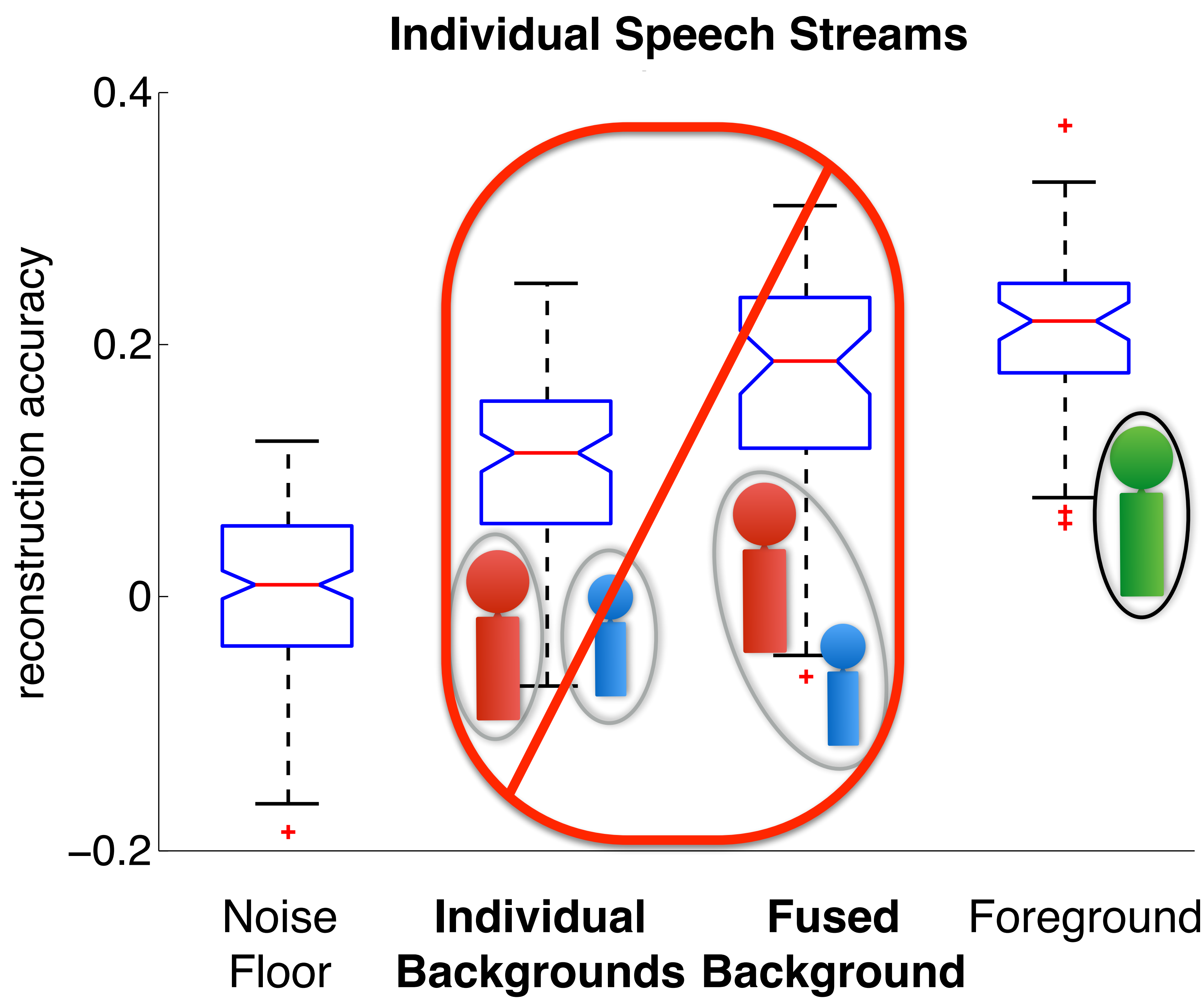
Integration Window over Late Times Only

Backgrounds vs. Background



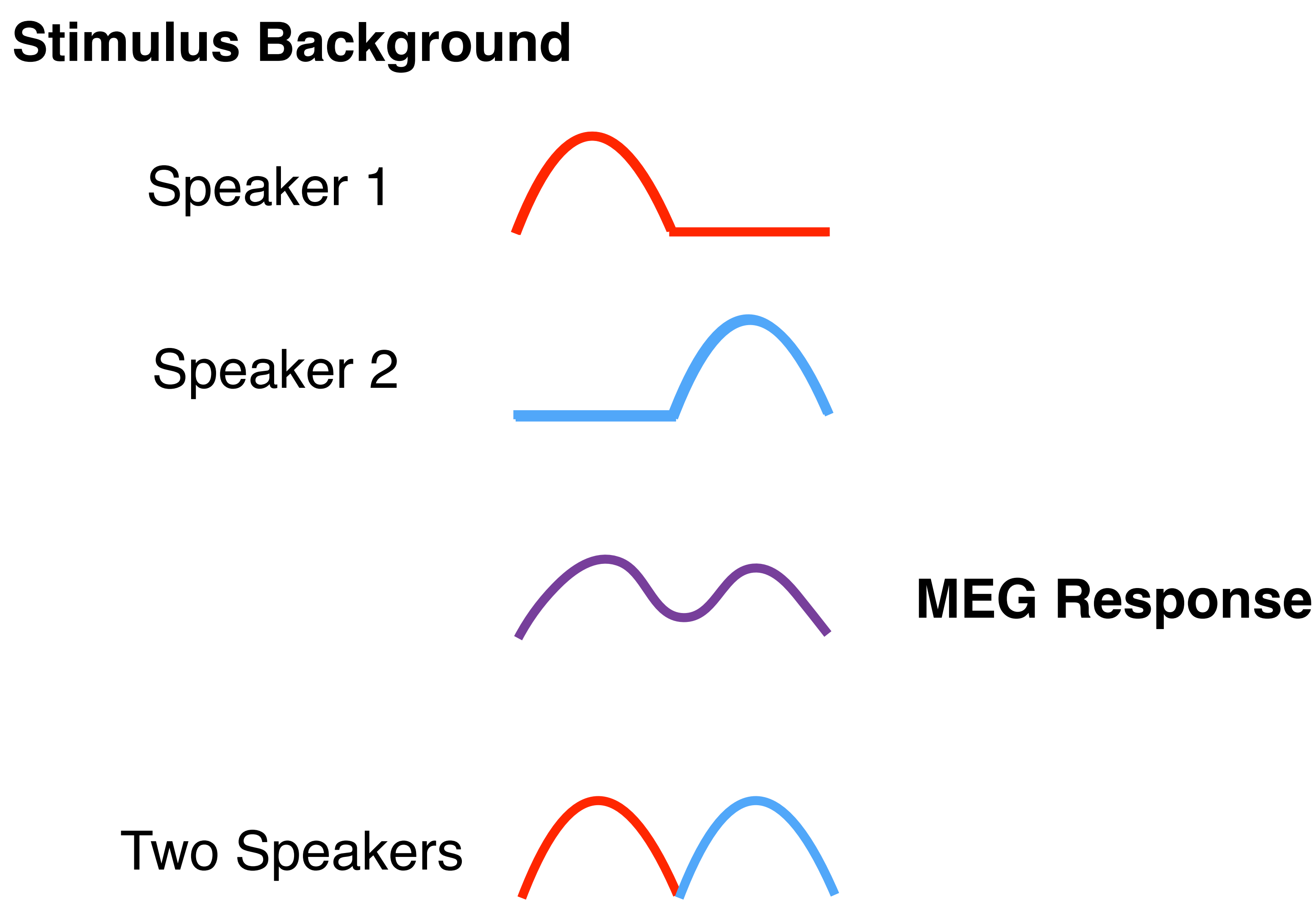
Integration Window over Late Times Only

Backgrounds vs. Background

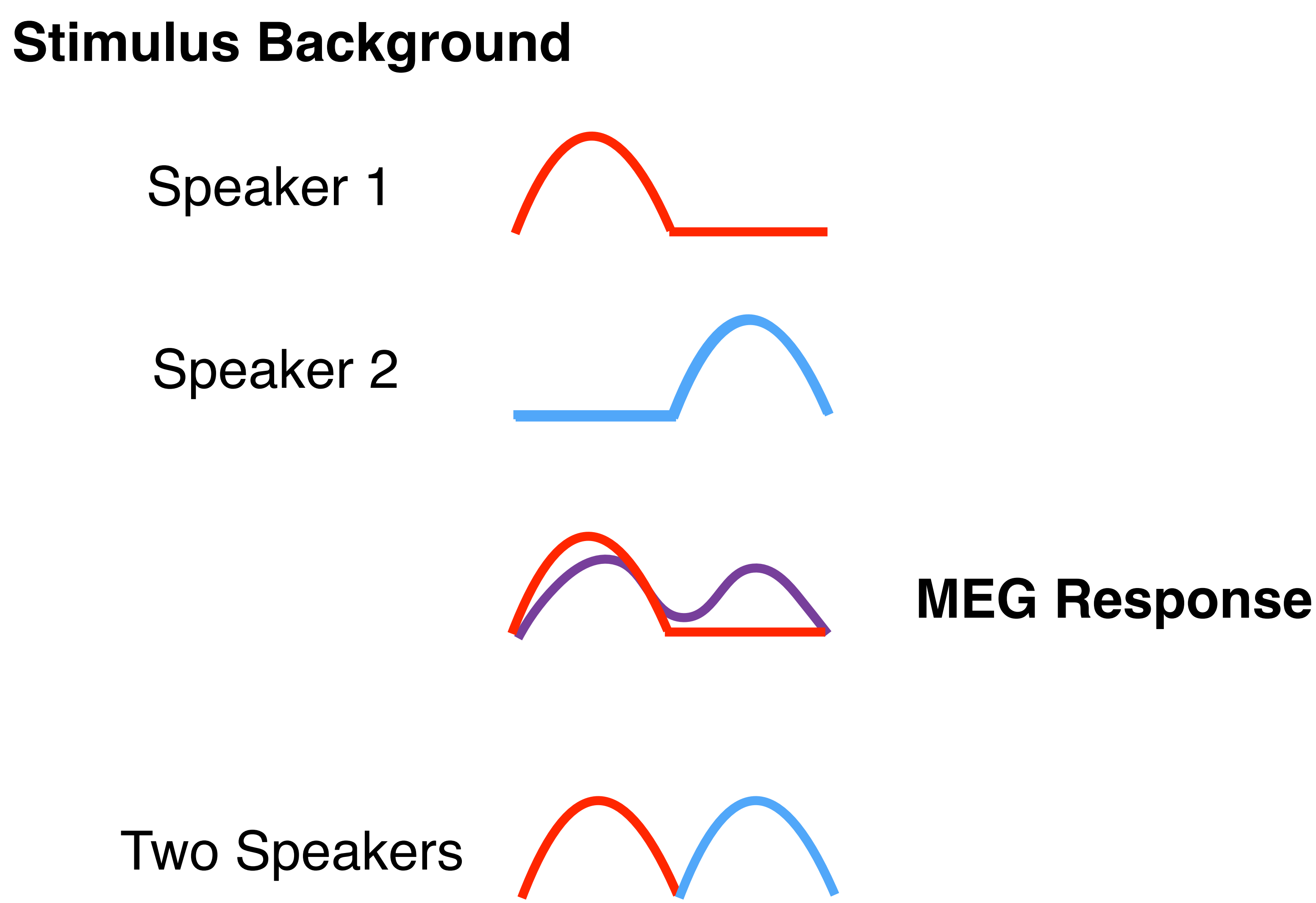


Integration Window over Late Times Only

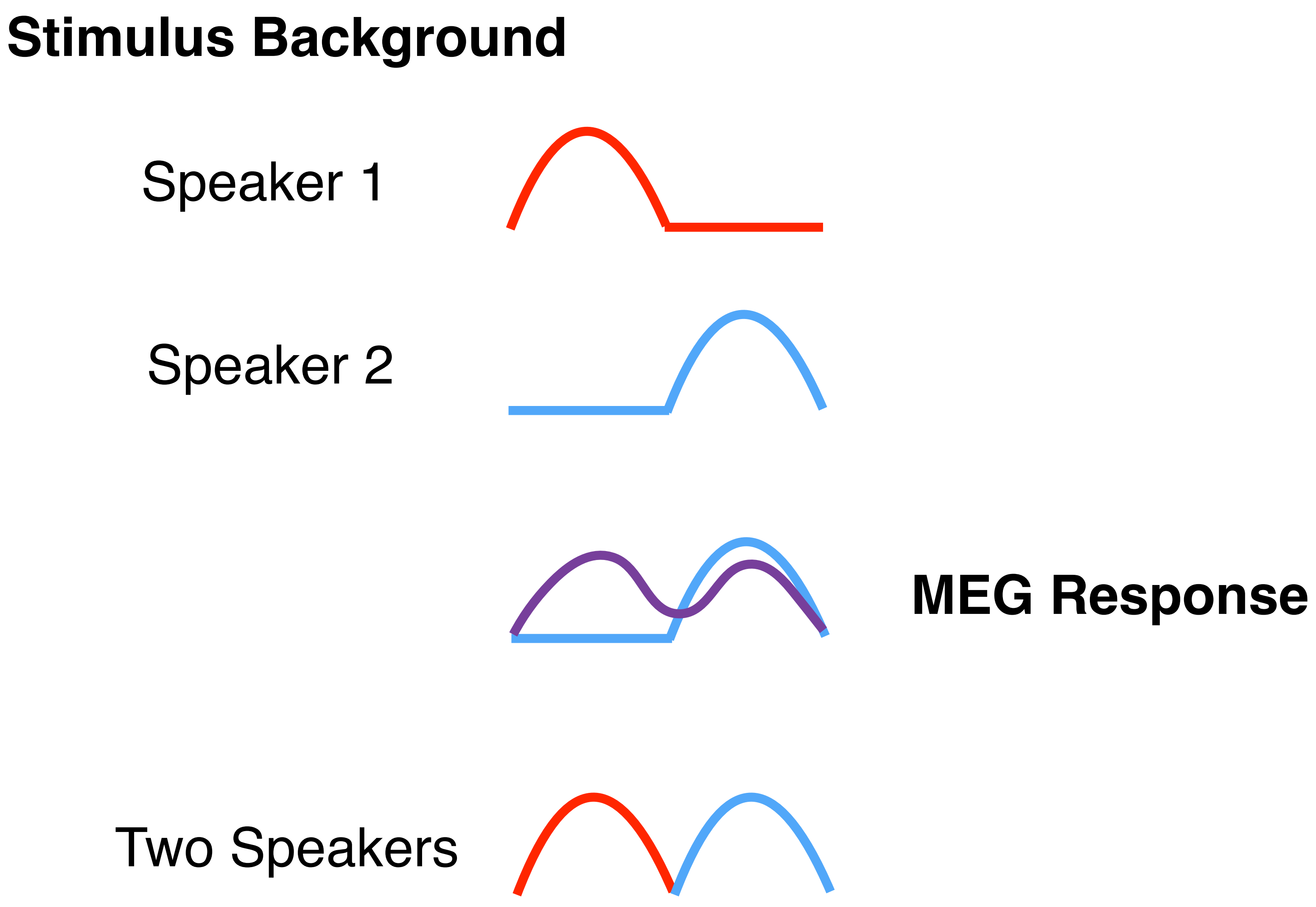
Backgrounds vs. Background
Why not?



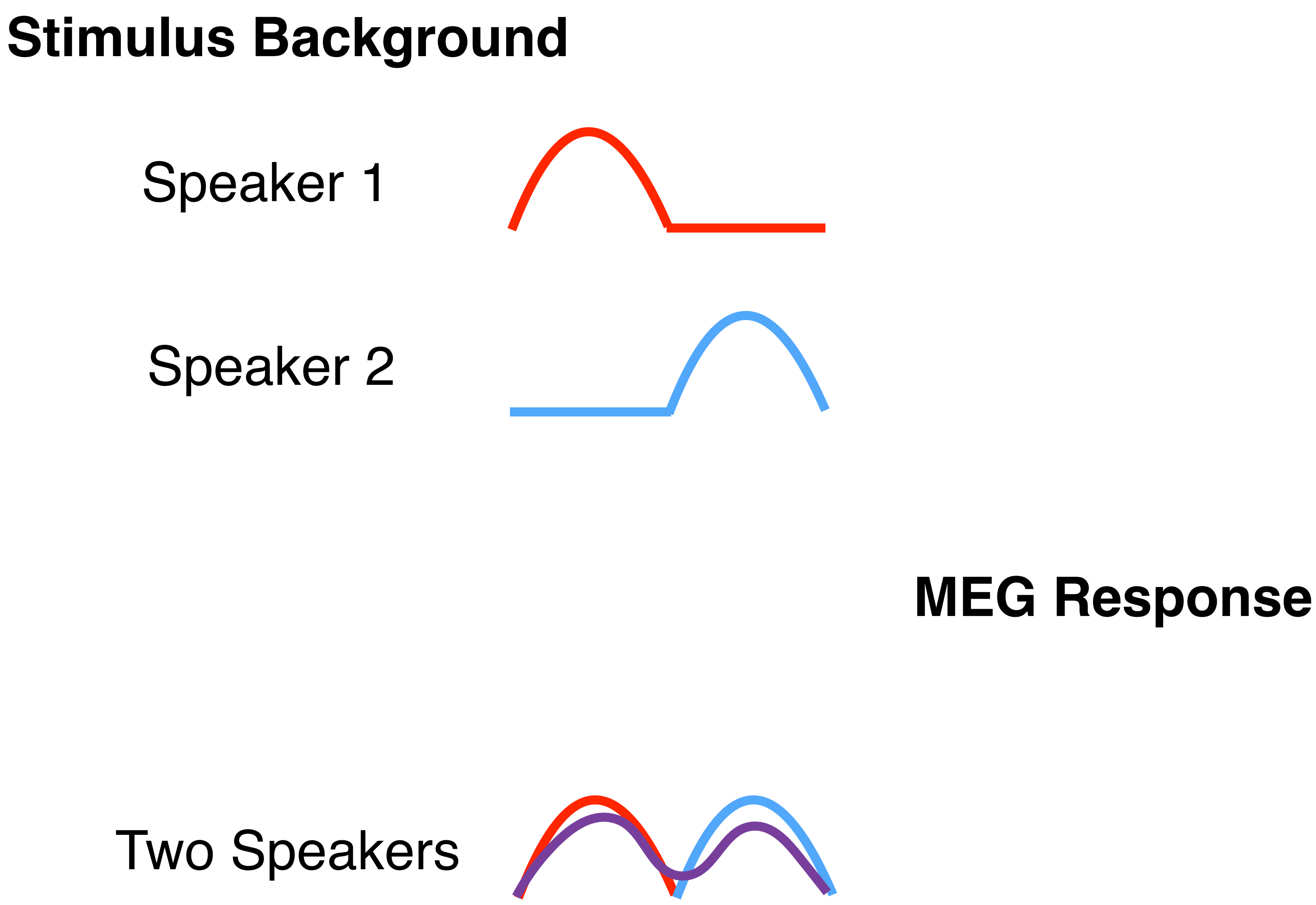
Backgrounds vs. Background
Why not?



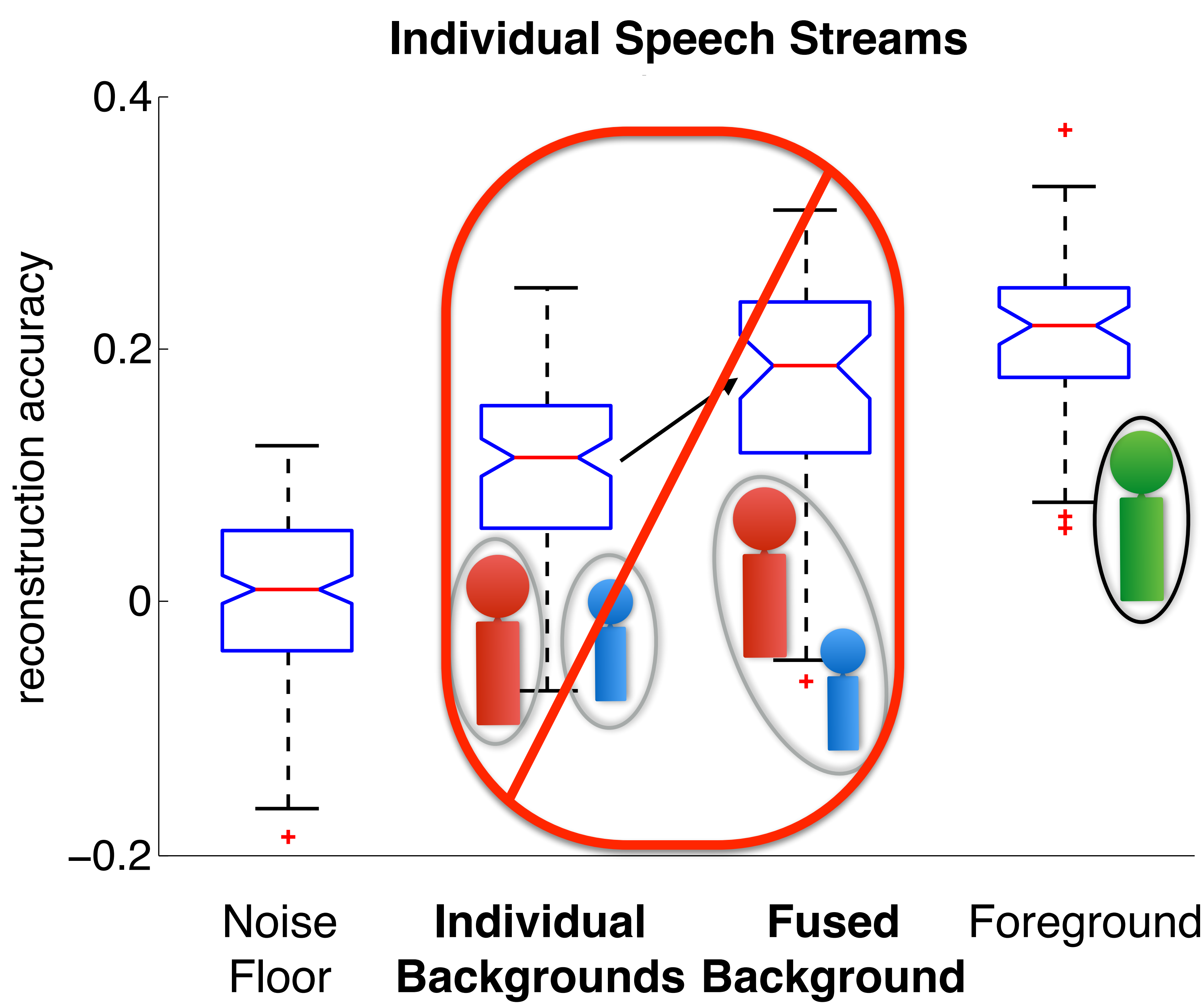
Backgrounds vs. Background
Why not?



Backgrounds vs. Background
Why not?

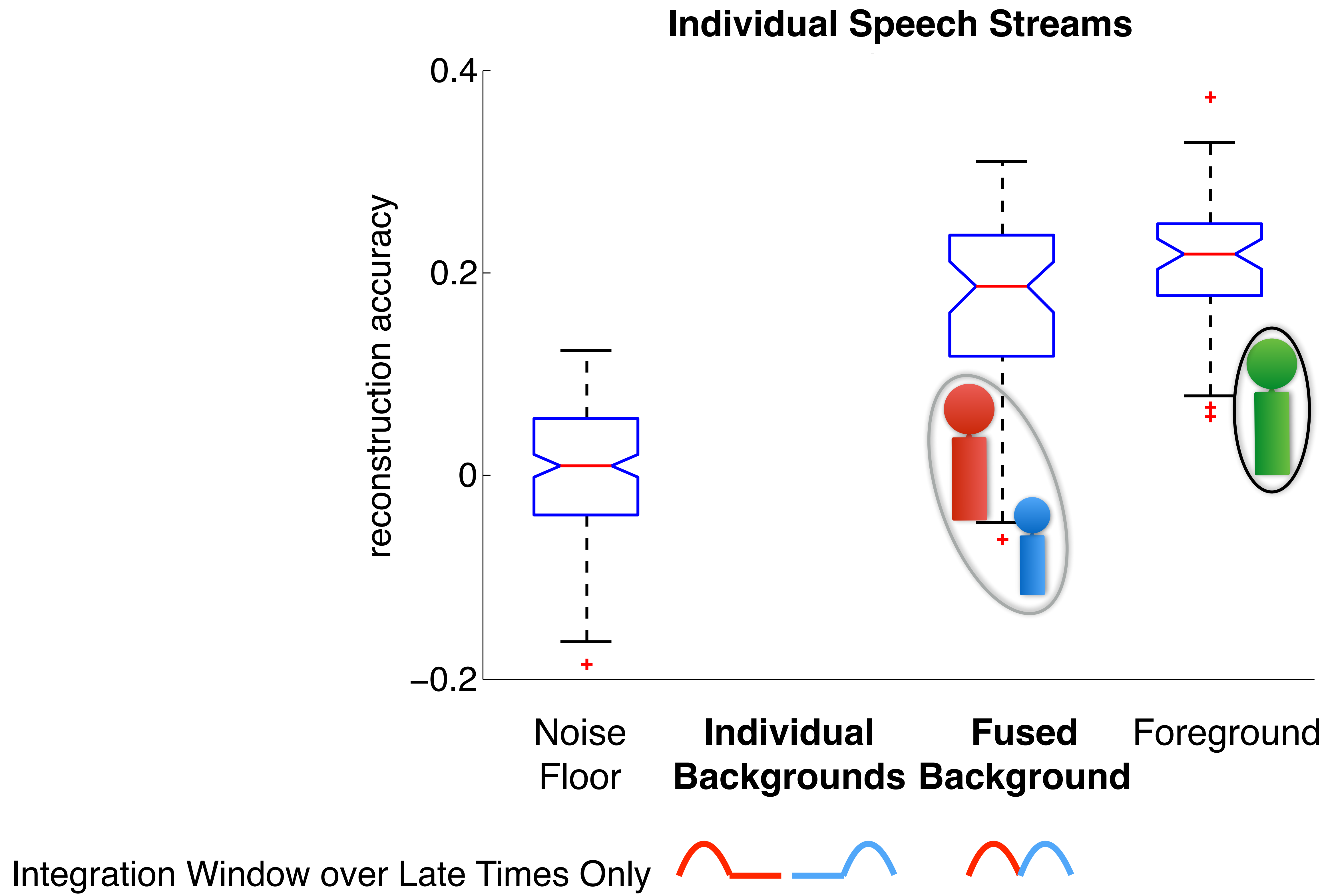


Backgrounds vs. Background

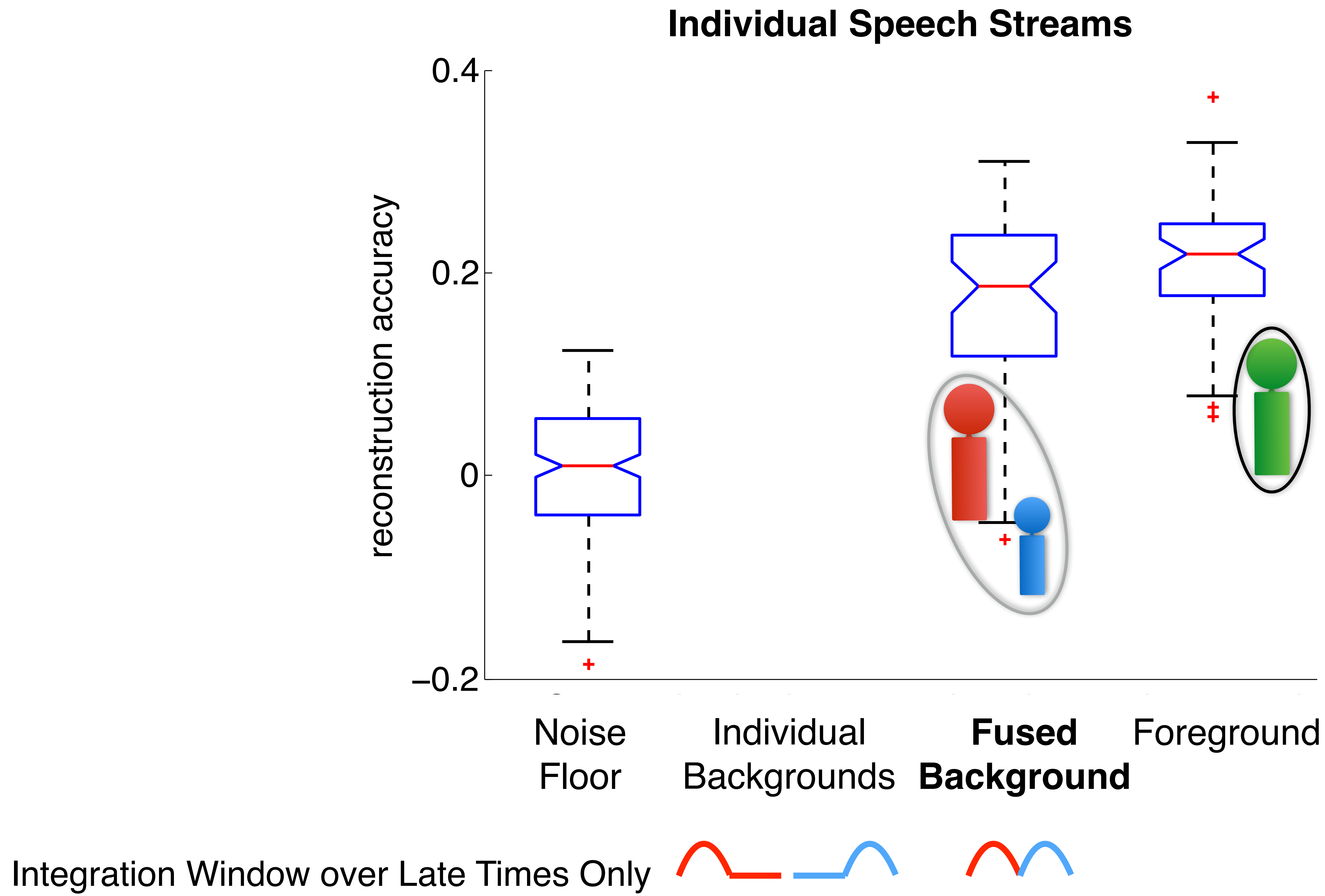


Integration Window over Late Times Only

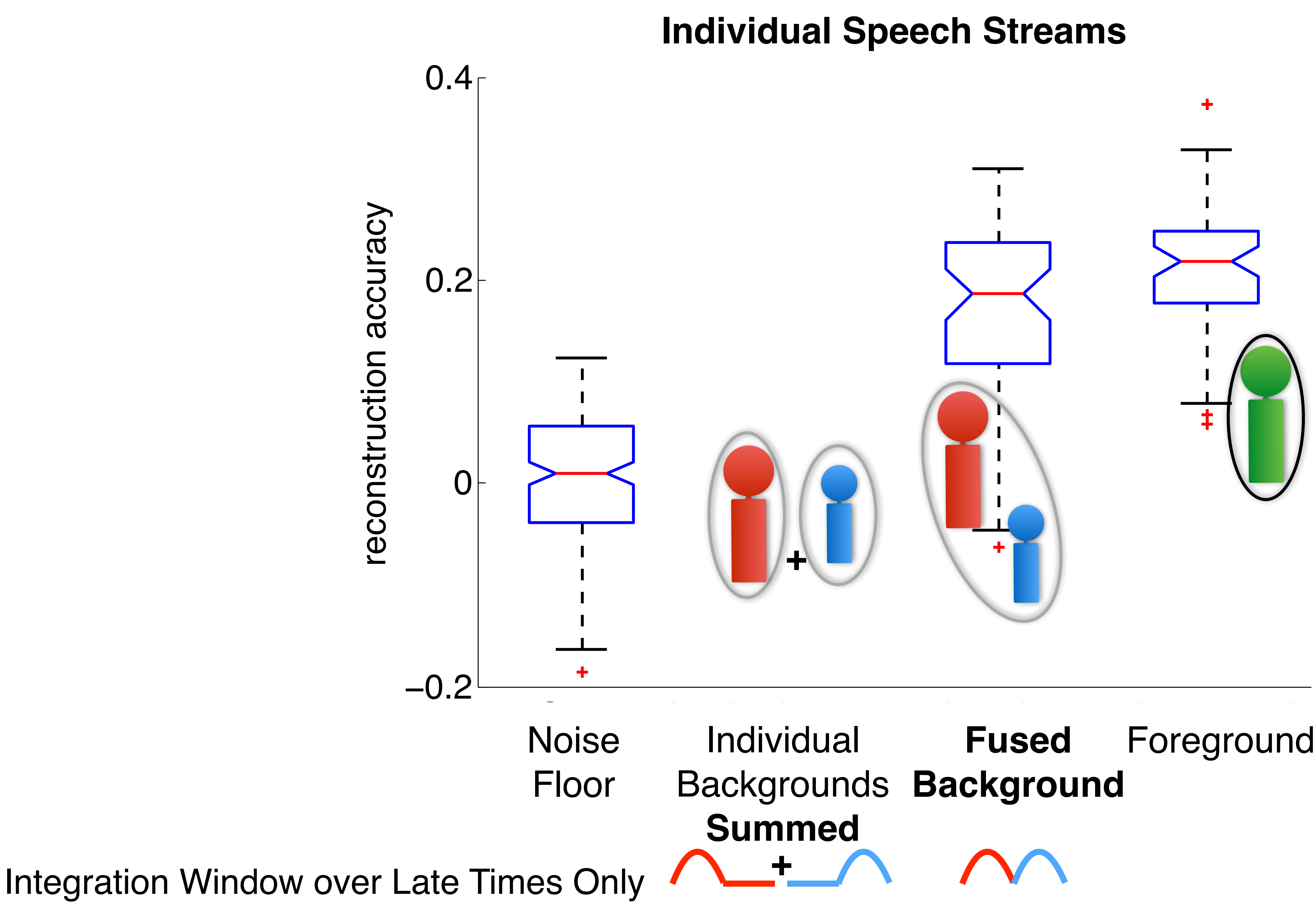
Backgrounds vs. Background



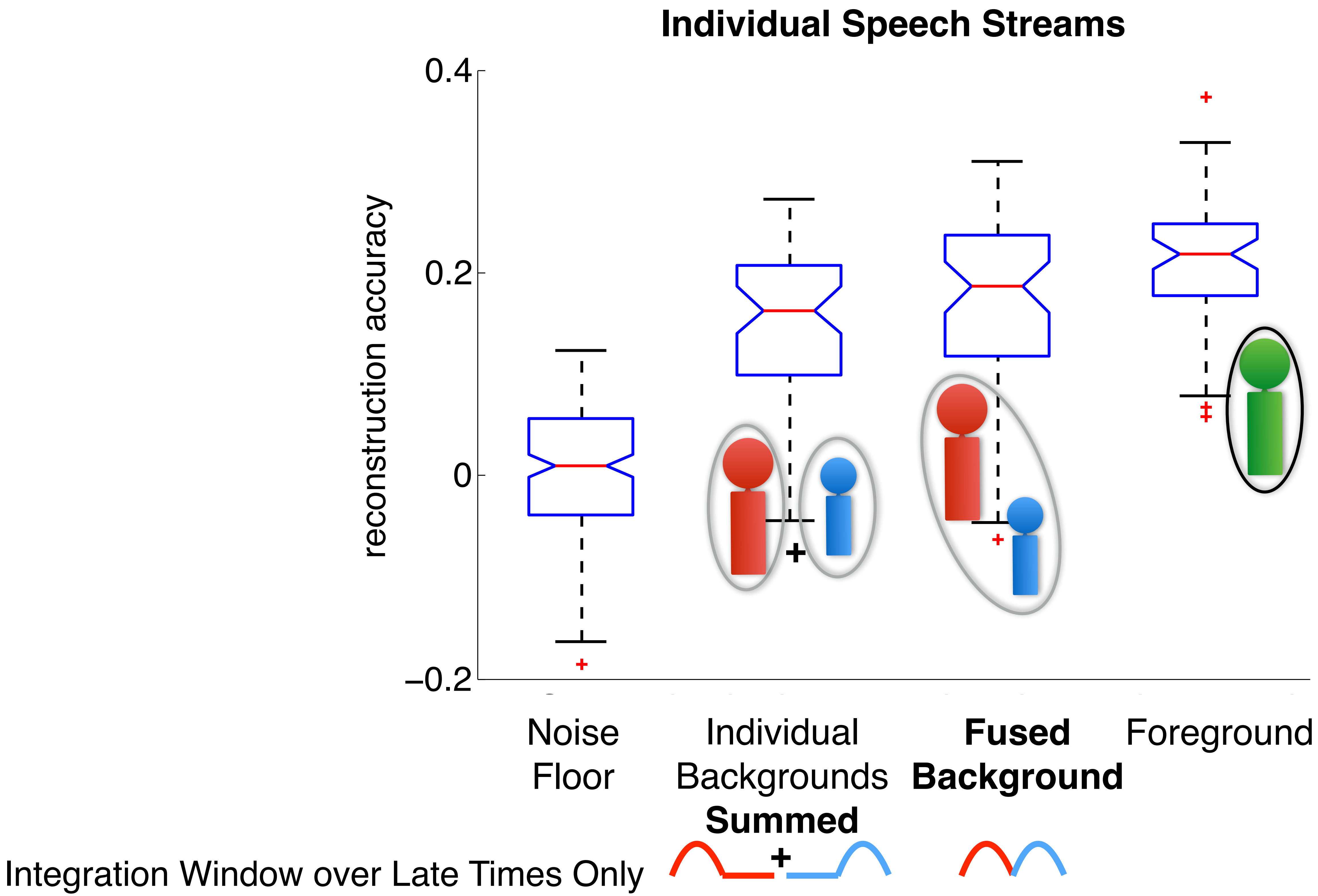
Backgrounds vs. Background



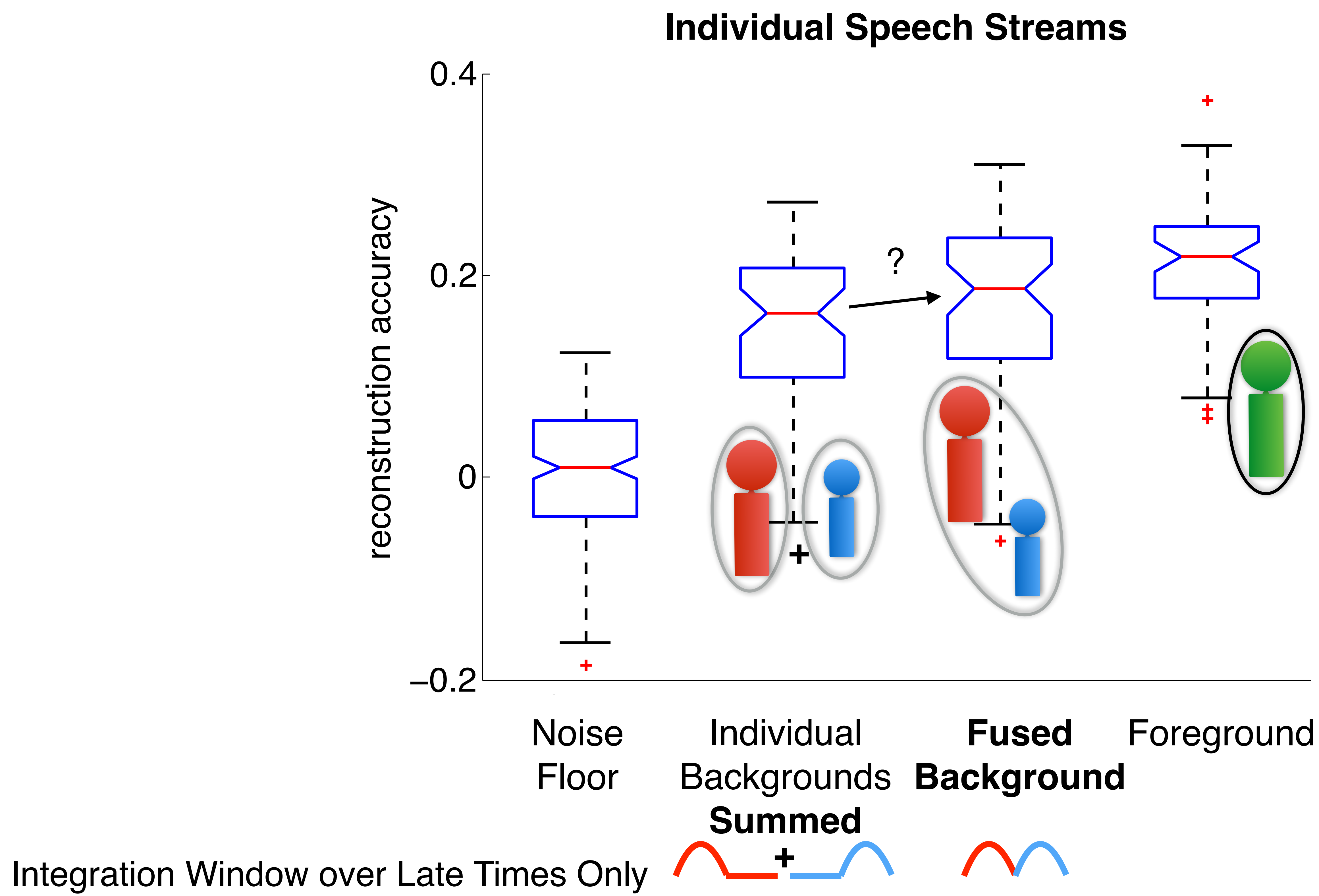
Backgrounds vs. Background



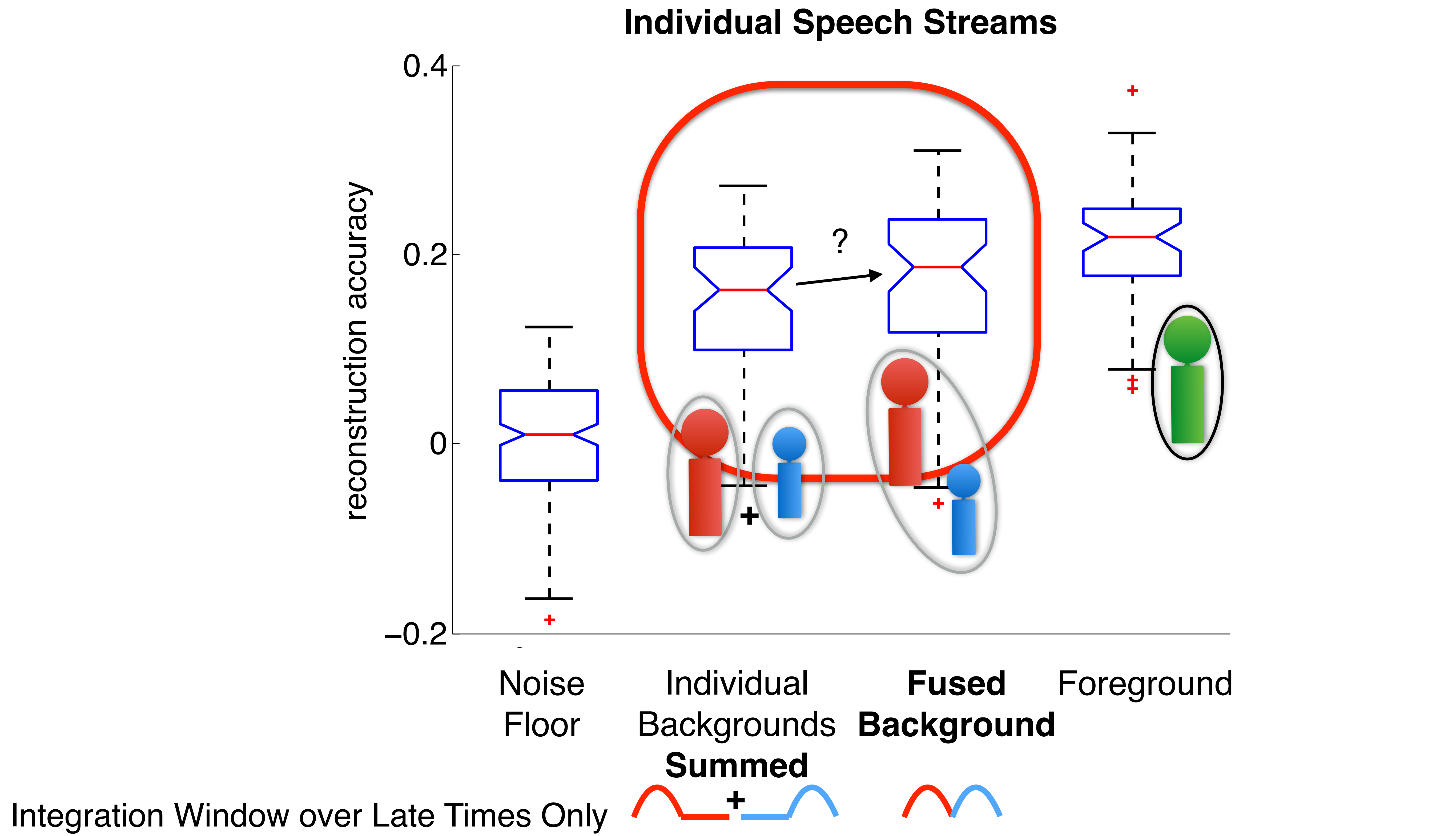
Backgrounds vs. Background



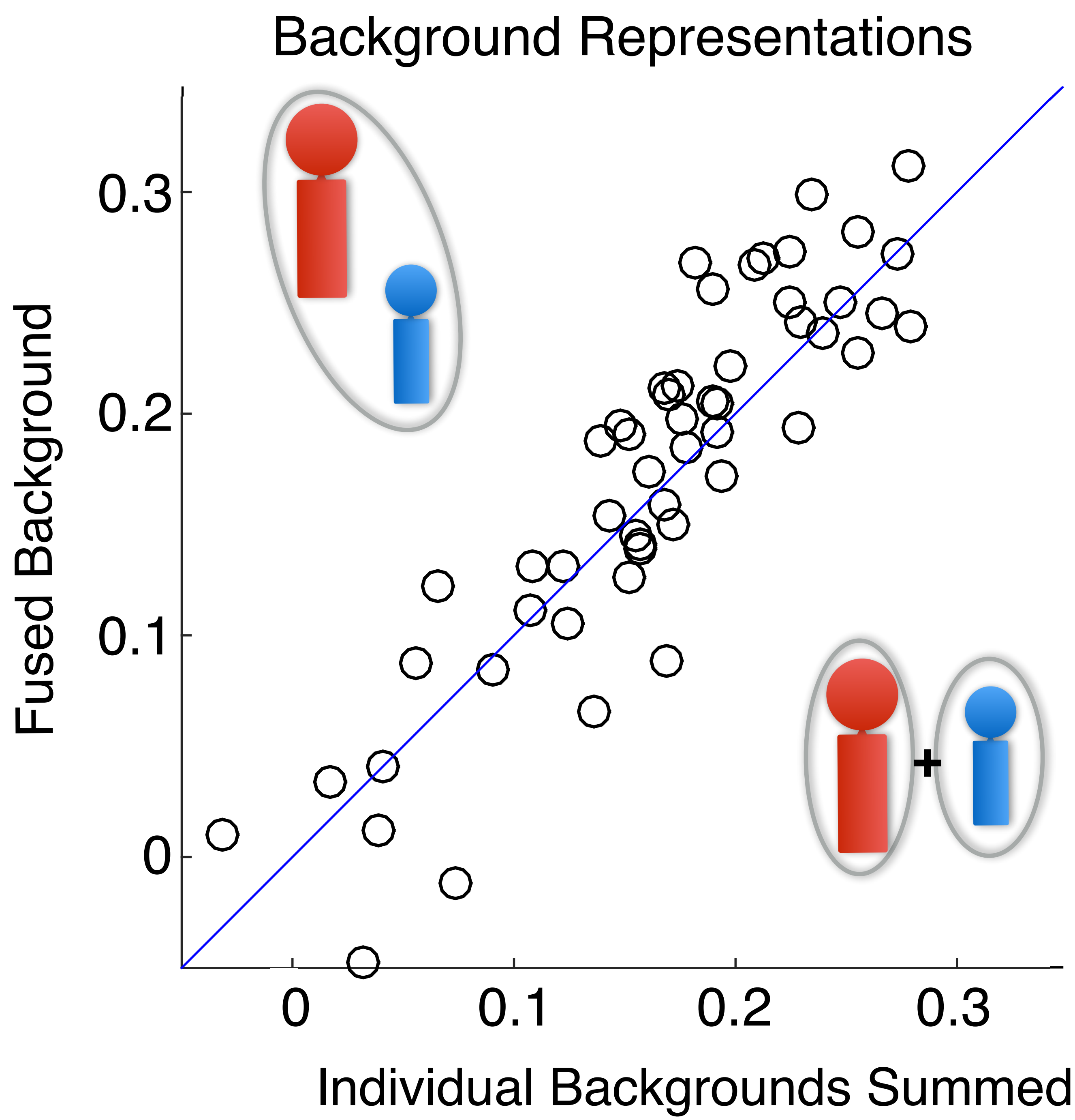
Backgrounds vs. Background



Backgrounds vs. Background

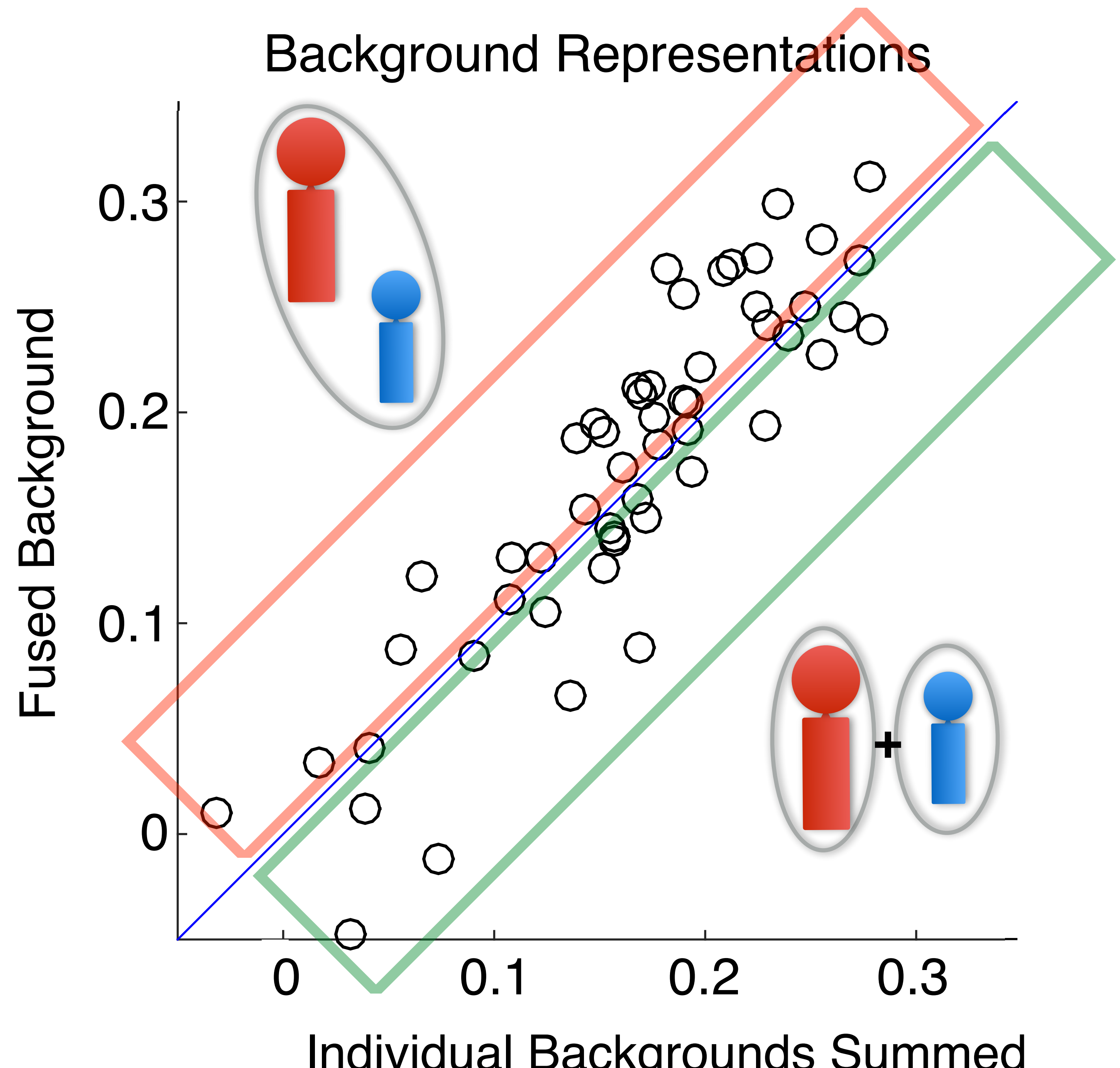


Backgrounds vs. Background



Integration Window over Late Times Only

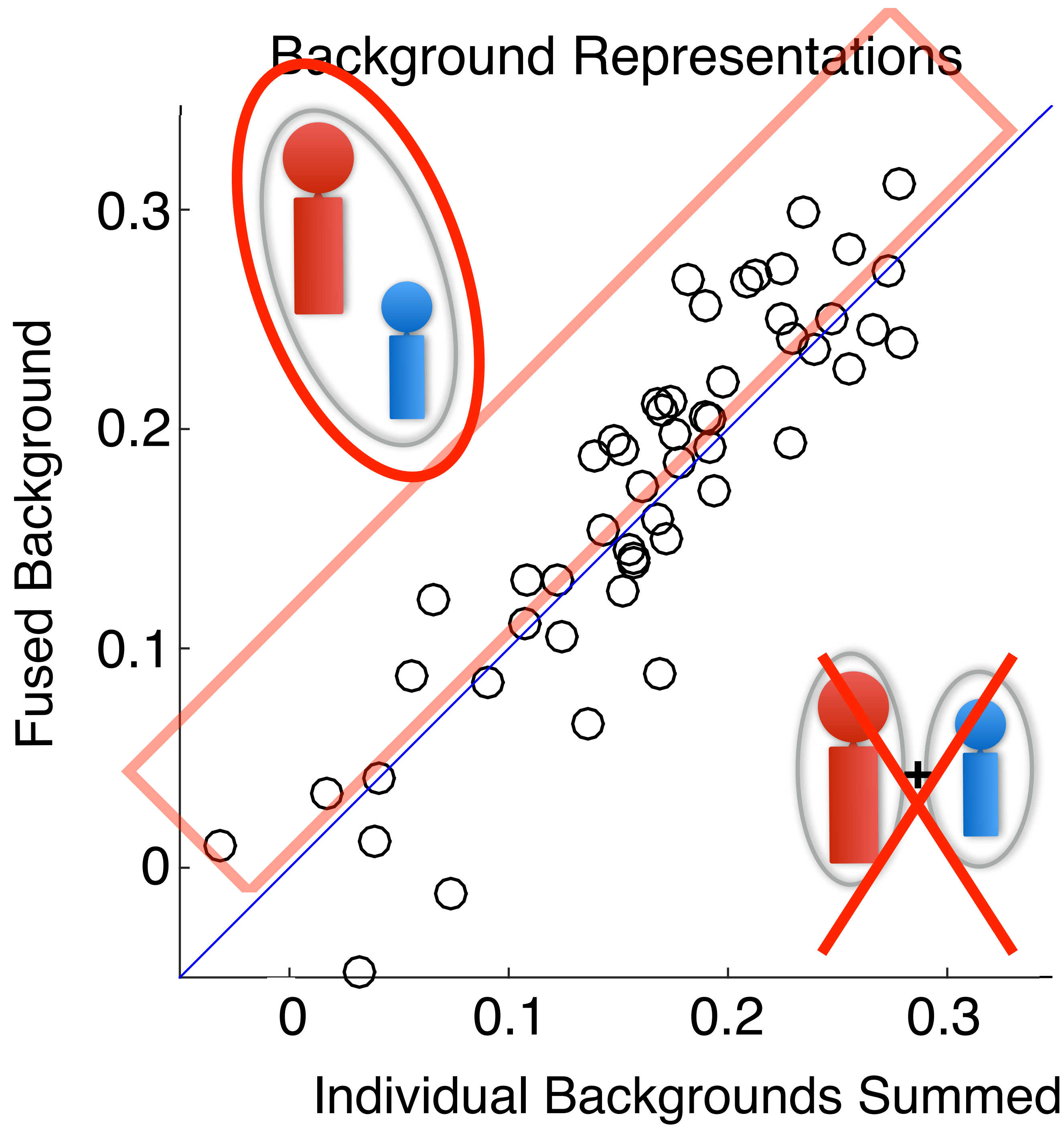
Backgrounds vs. Background



Integration Window over Late Times Only

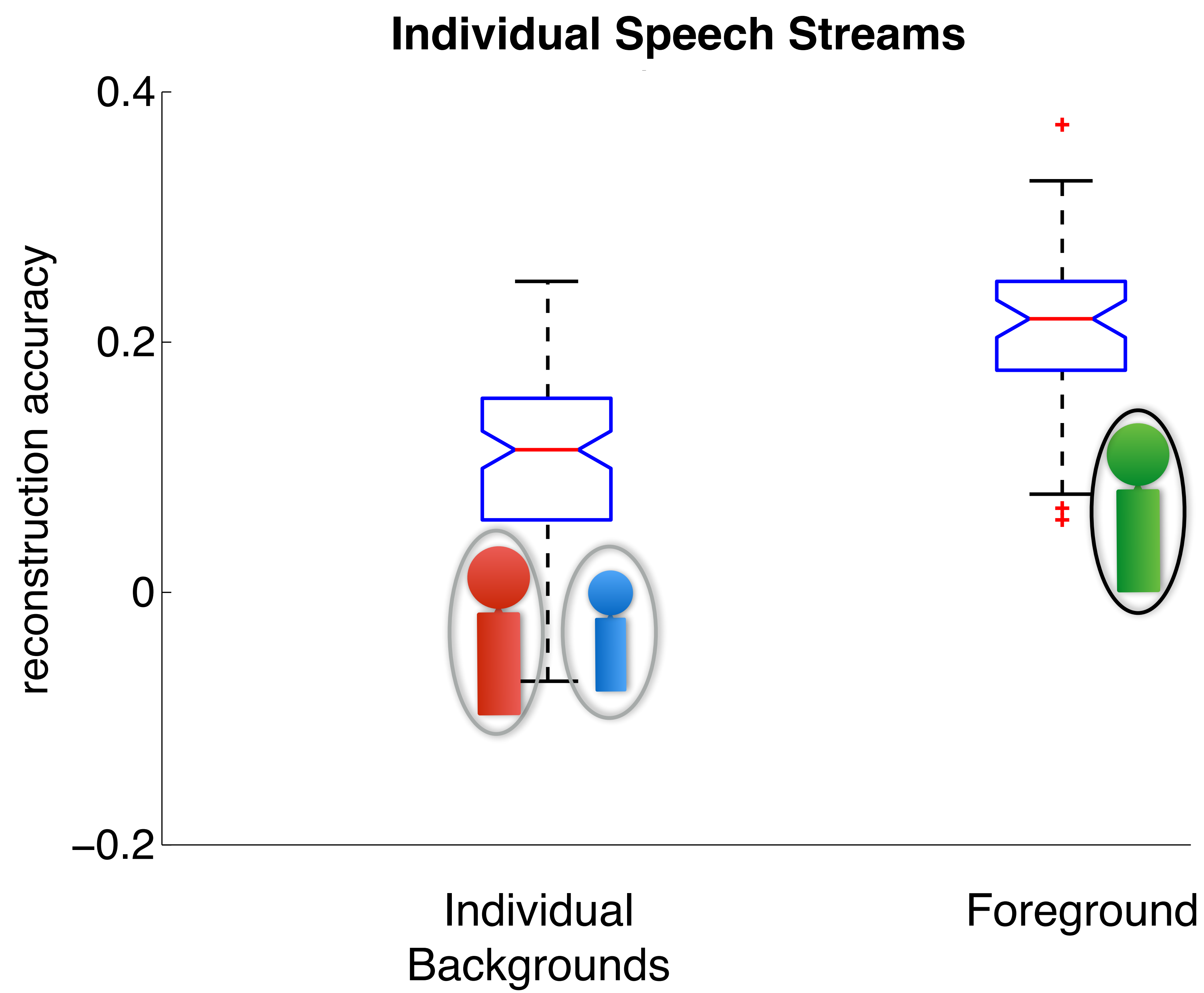
Backgrounds vs. Background

High latency areas (PT) represent ***fused*** background with better fidelity than ***individual*** backgrounds ($p = 0.012$)



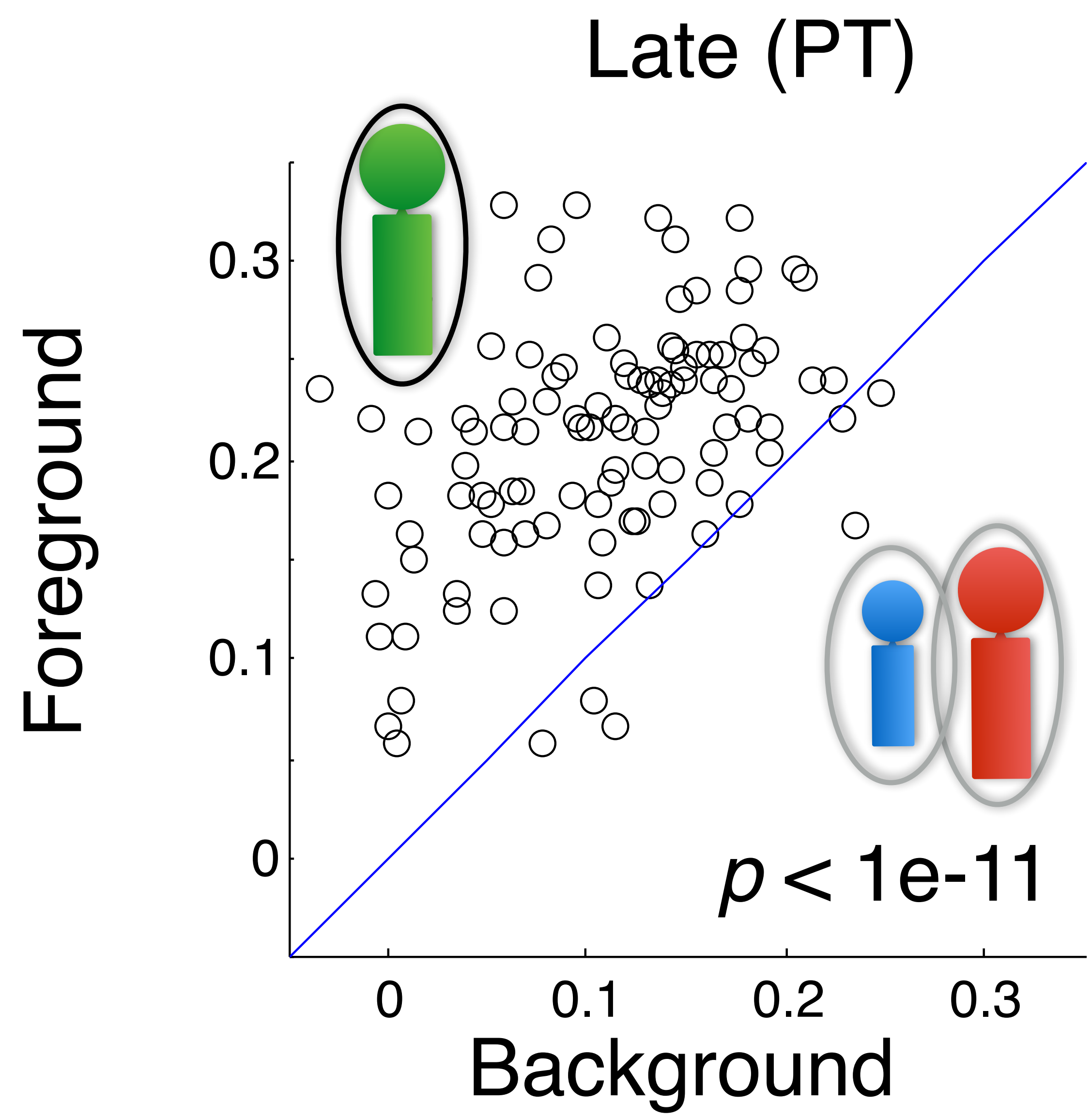
Integration Window over Late Times Only

Foreground vs. Background

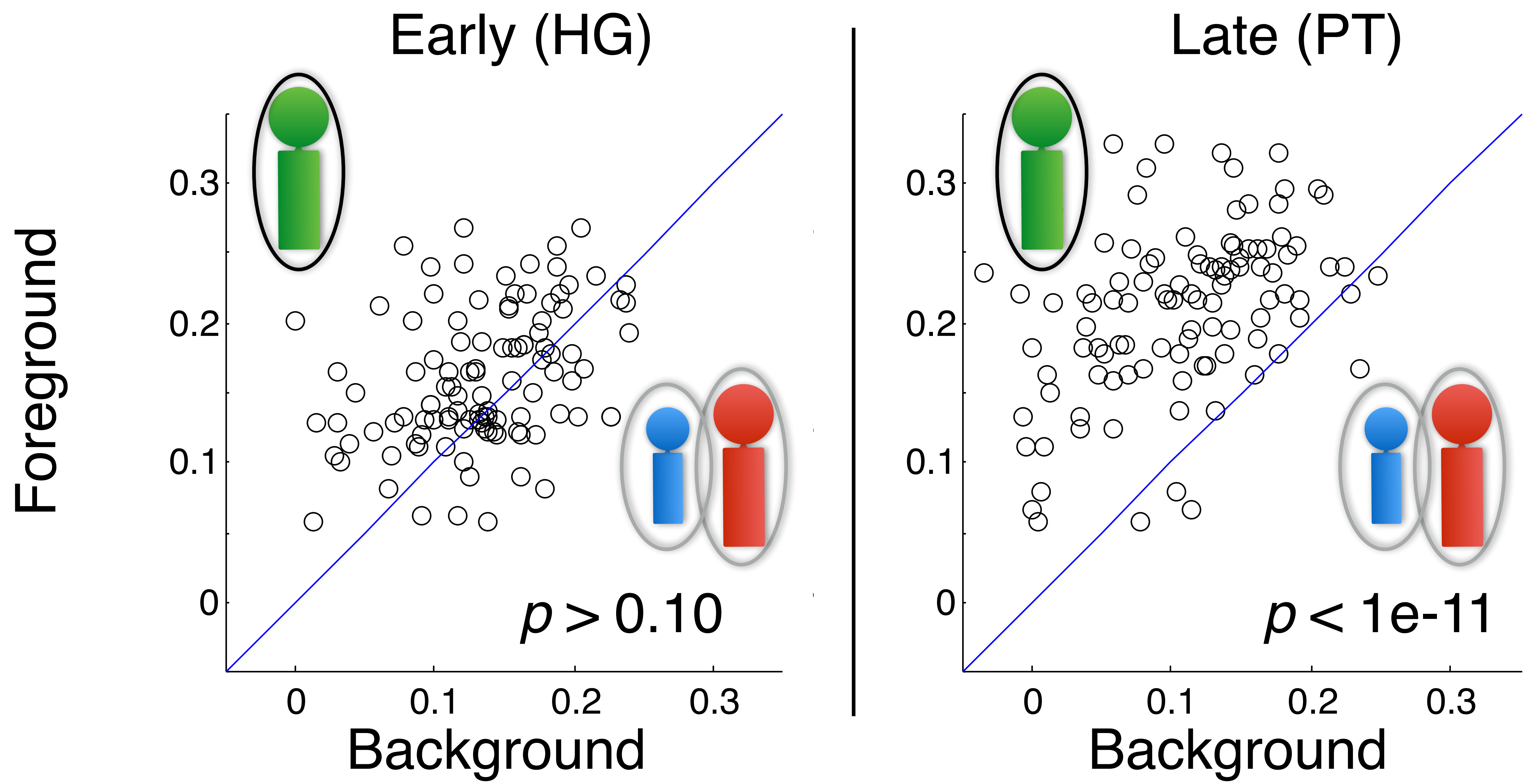


Integration Window over Late Times Only

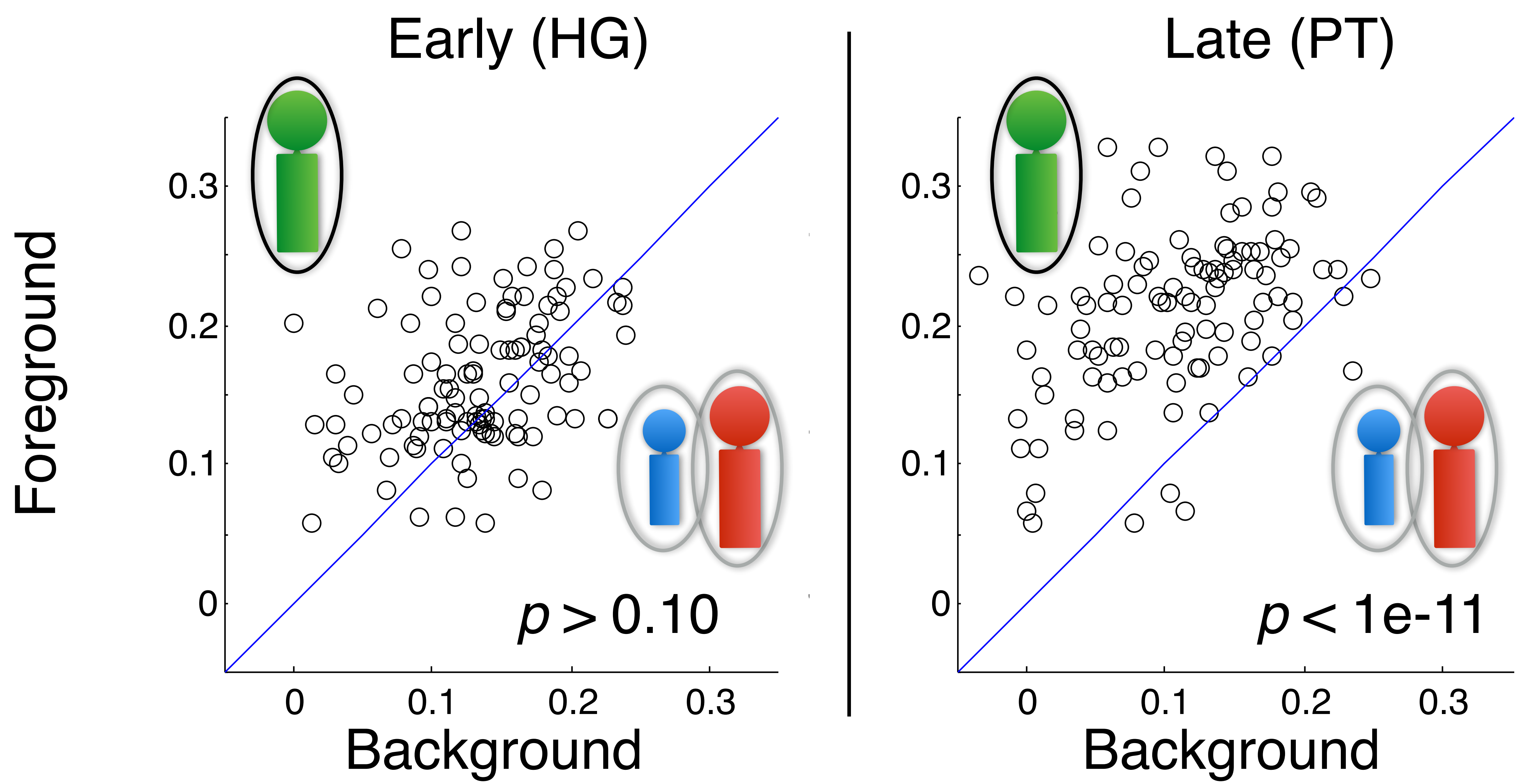
Foreground vs. Background
Early vs. Late



Foreground vs. Background
Early vs. Late



Foreground vs. Background
Early vs. Late



HG represents attended and unattended speech with *almost* equal fidelity

Summary

- Cortical representations of speech
  - representation of envelope (up to ~10 Hz)
- Cortical Processing Hierarchy: Consistent with being neural representation of auditory perceptual object
- Object representation at 100 ms latency (PT), but not by 50 ms (HG)
- Preliminary evidence for
  - PT: additional fused background representation
  - HG: almost equal representations

Thank You