Neural Representations of Continuous Speech in Auditory Cortex

Jonathan Z. Simon

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

University of Maryland

Acknowledgements

Grad Students

Francisco Cervantes
Mahshid Najafi
Alex Presacco
Krishna Puvvada
Ben Walsh

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

Acknowledgements

Grad Students

Francisco Cervantes
Mahshid Najafi
Alex Presacco
Krishna Puvvada
Ben Walsh

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)
- Cortical Representations of Speech
  - Encoding vs. Decoding
  - Attended vs. Unattended Speech
  - Foreground vs. Background

Neural Signals & MEG

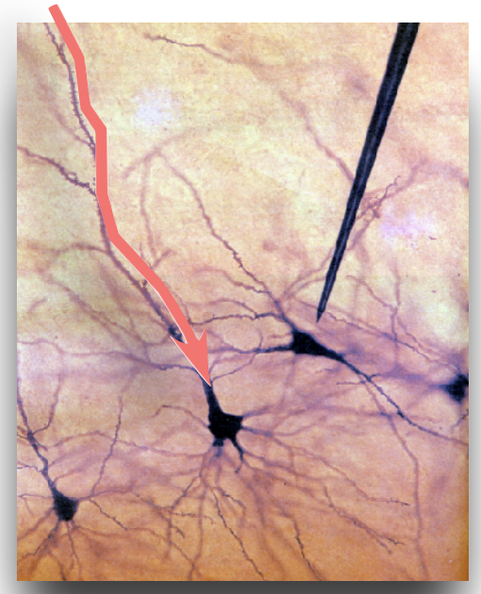
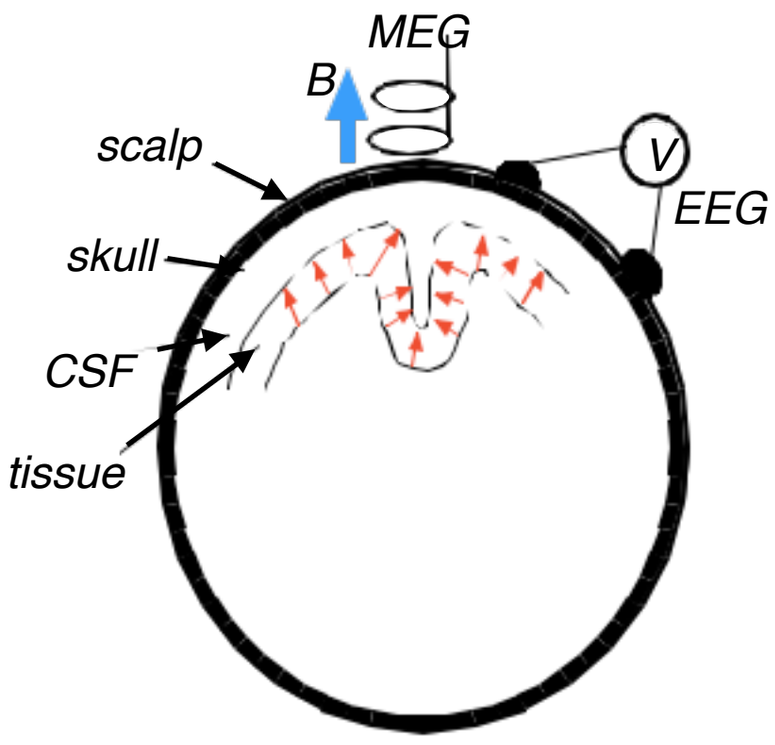
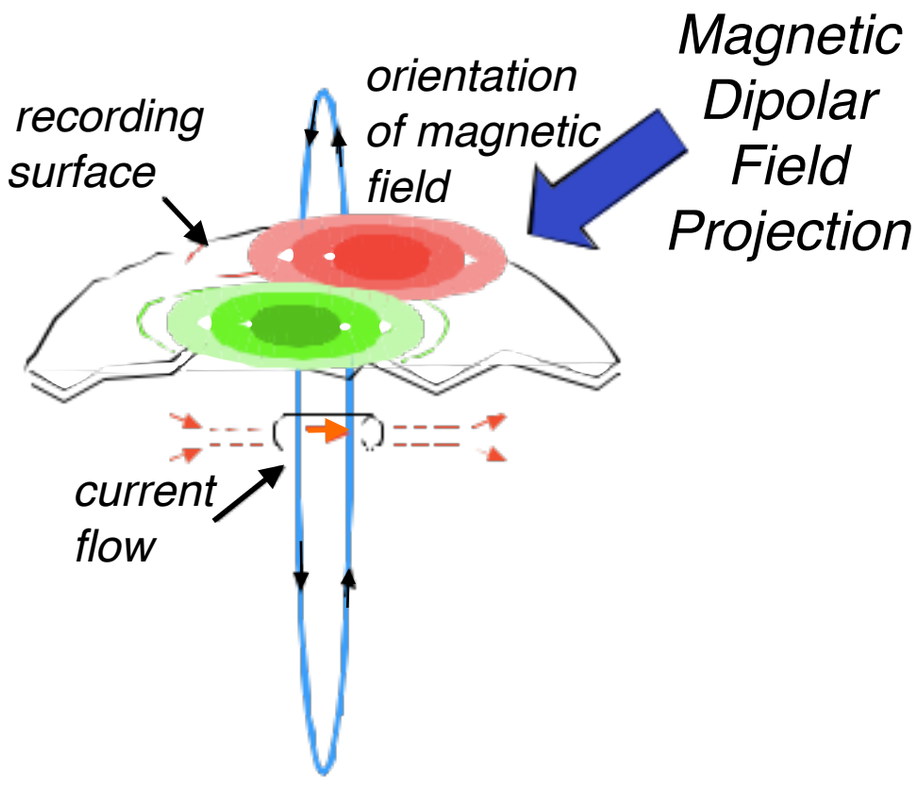


Photo by Fritz Goro

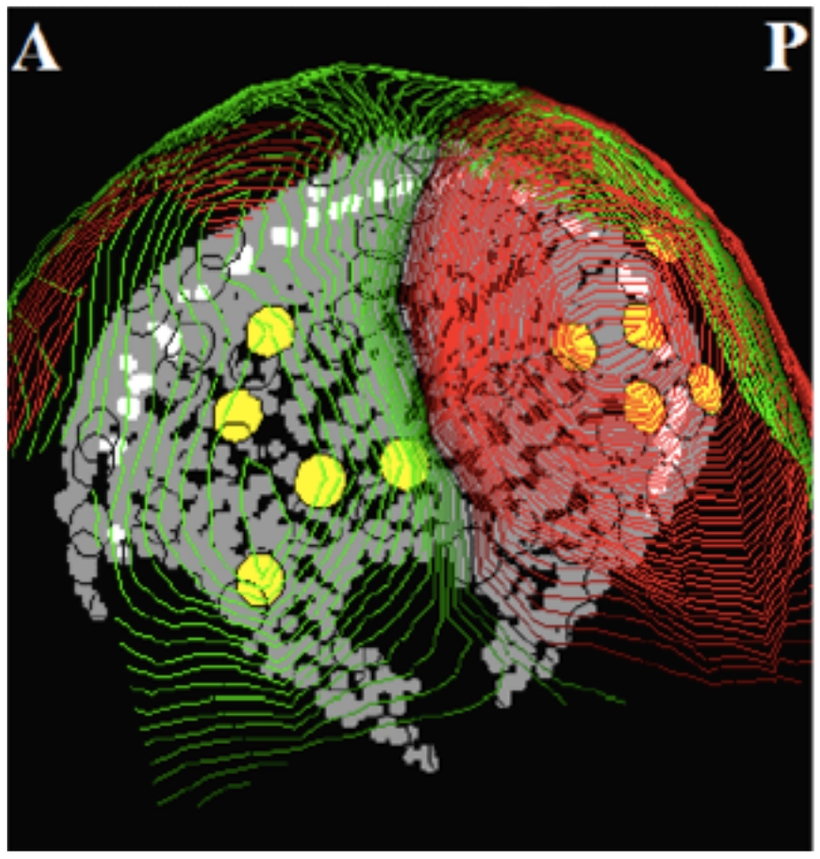




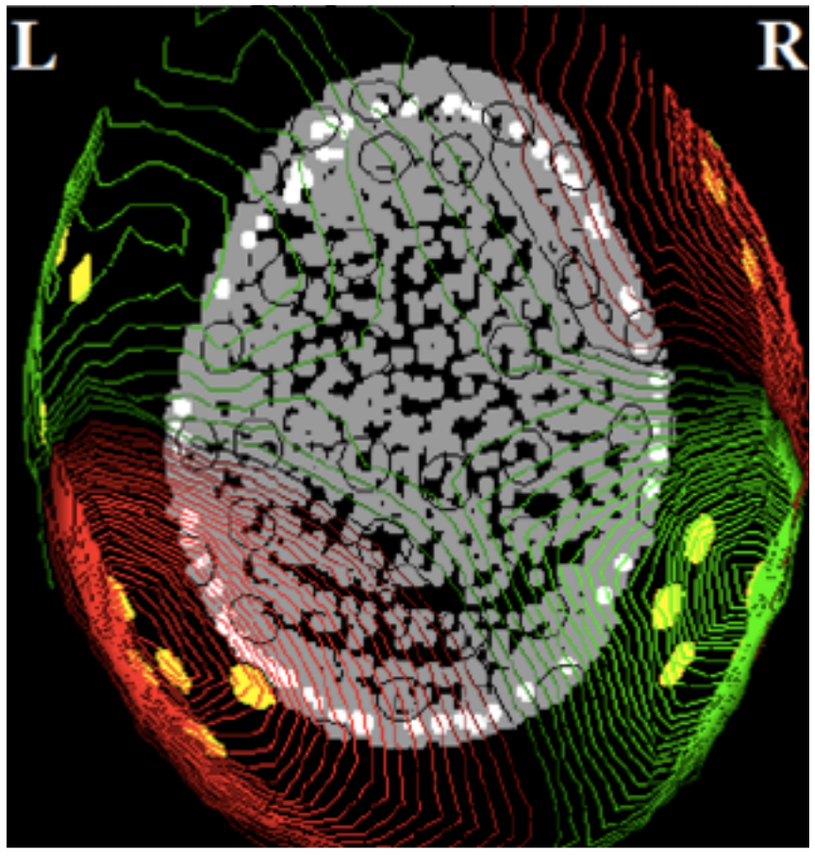
- Direct electrophysiological measurement
  - not hemodynamic
  - real-time
- No unique solution for distributed source

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

MEG Auditory Field



Sagittal View

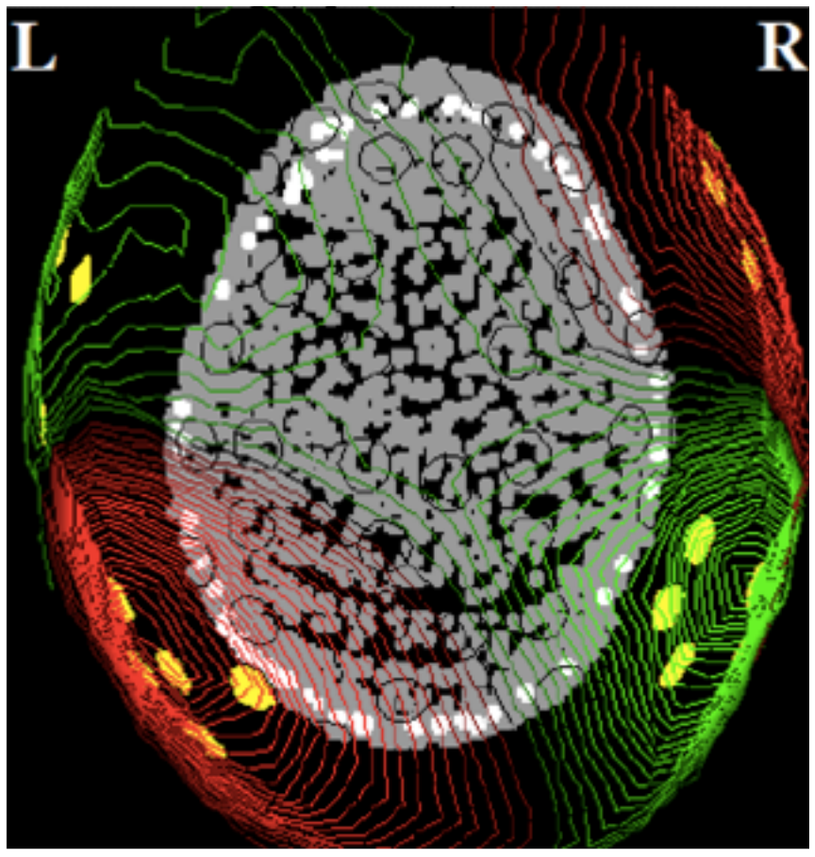


Axial View

Strongly
Lateralized

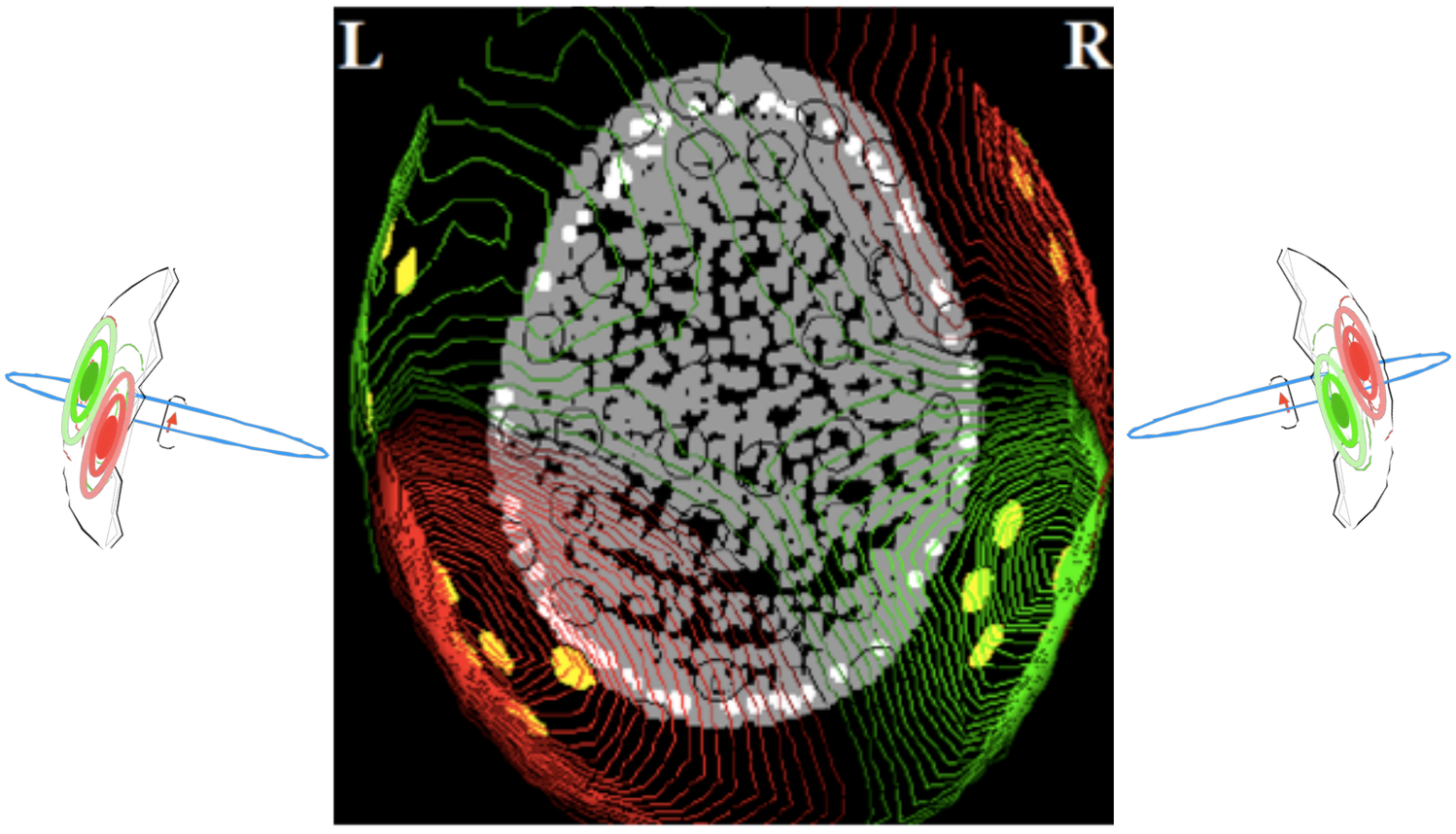
Chait, Poeppel and Simon, Cerebral Cortex (2006)

MEG Auditory Field



Chait et al., Cerebral Cortex (2006)

MEG Auditory Field

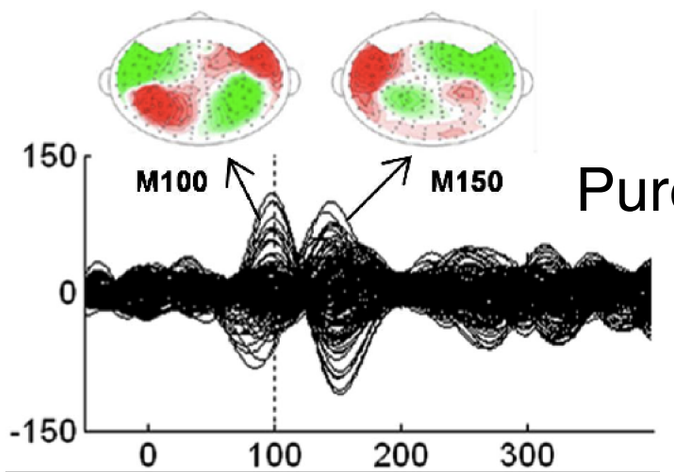


Chait et al., Cerebral Cortex (2006)

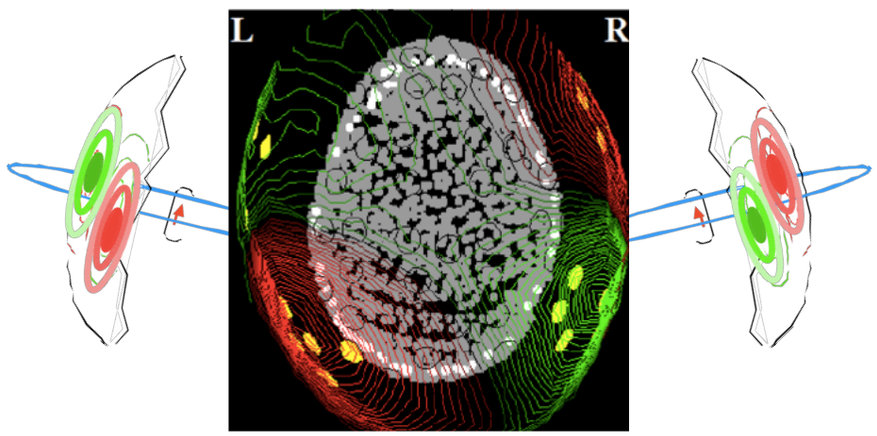
Time Course of MEG Responses

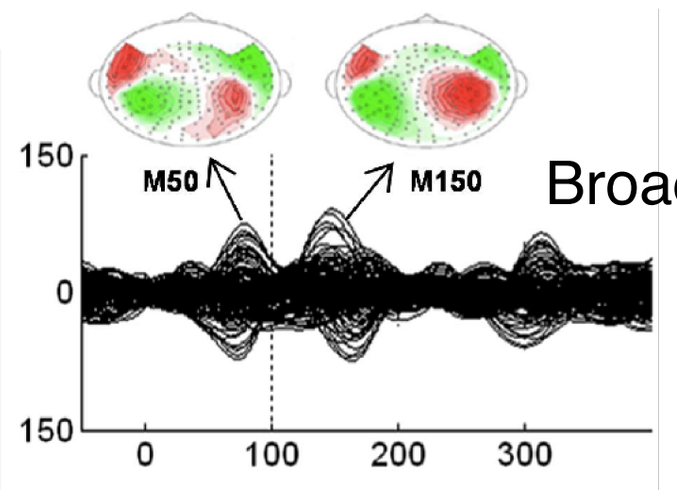
Auditory Evoked Responses

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



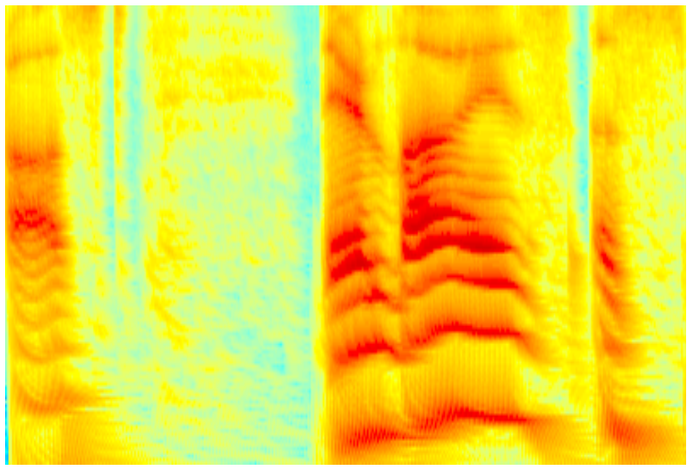
Pure Tone

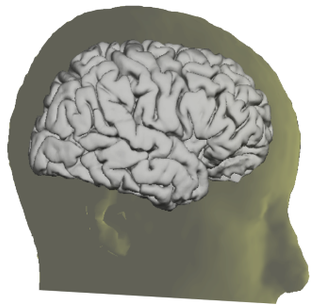




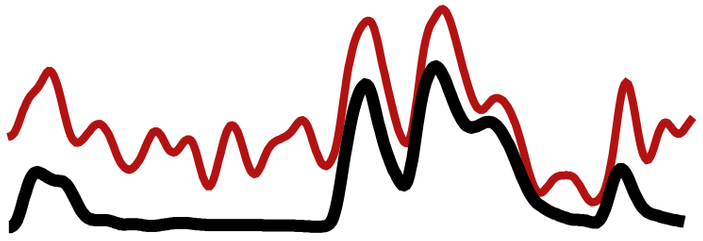
Broadband Noise

MEG Responses to Speech Modulations

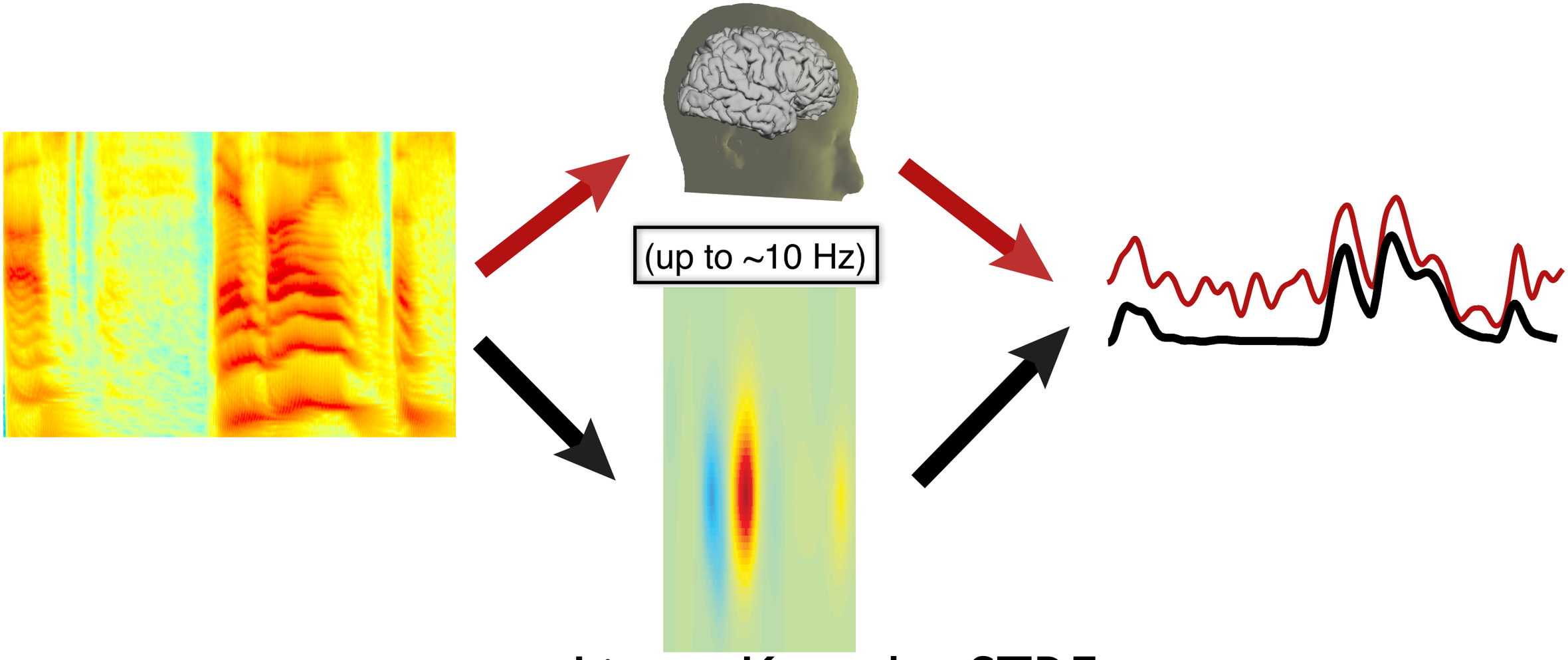




Auditory Model



MEG Responses Predicted by STRF Model

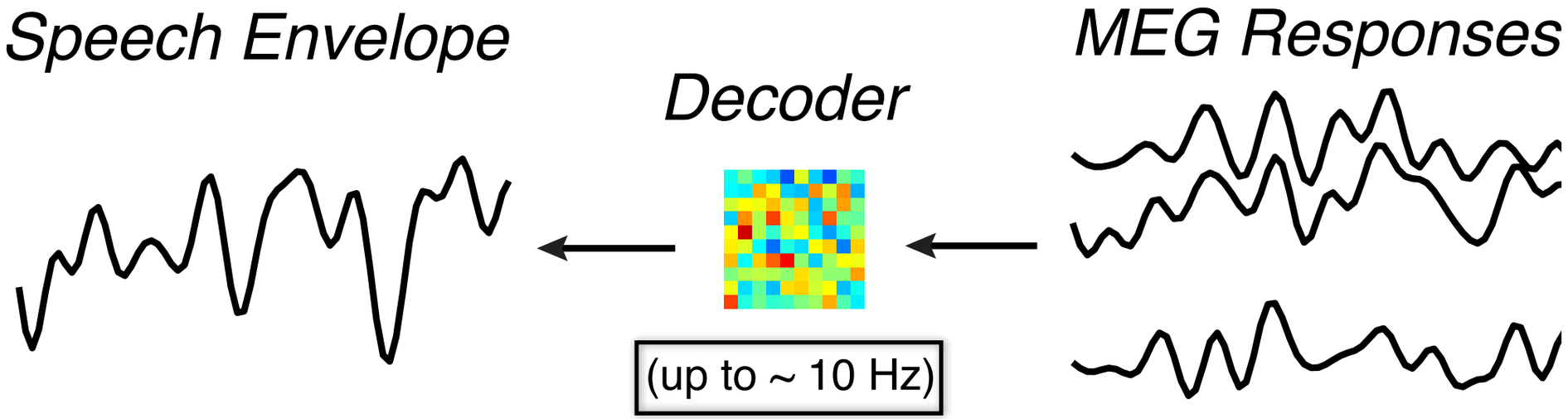


Linear Kernel = STRF

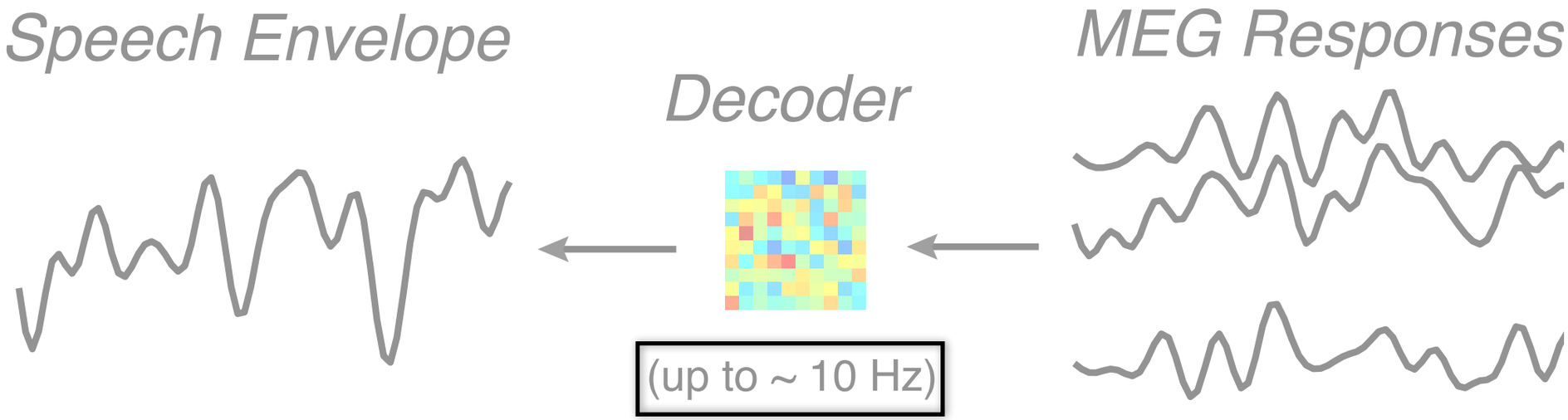
"Spectro-Temporal Response Function"

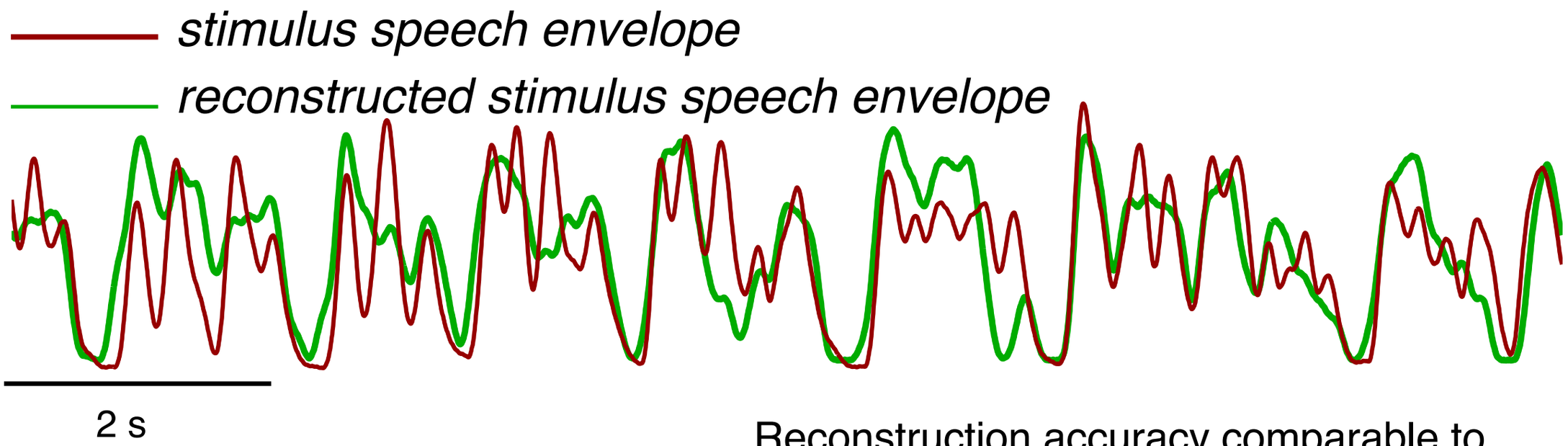
Ding & Simon, J Neurophysiol (2012)

Neural Reconstruction of Speech Envelope



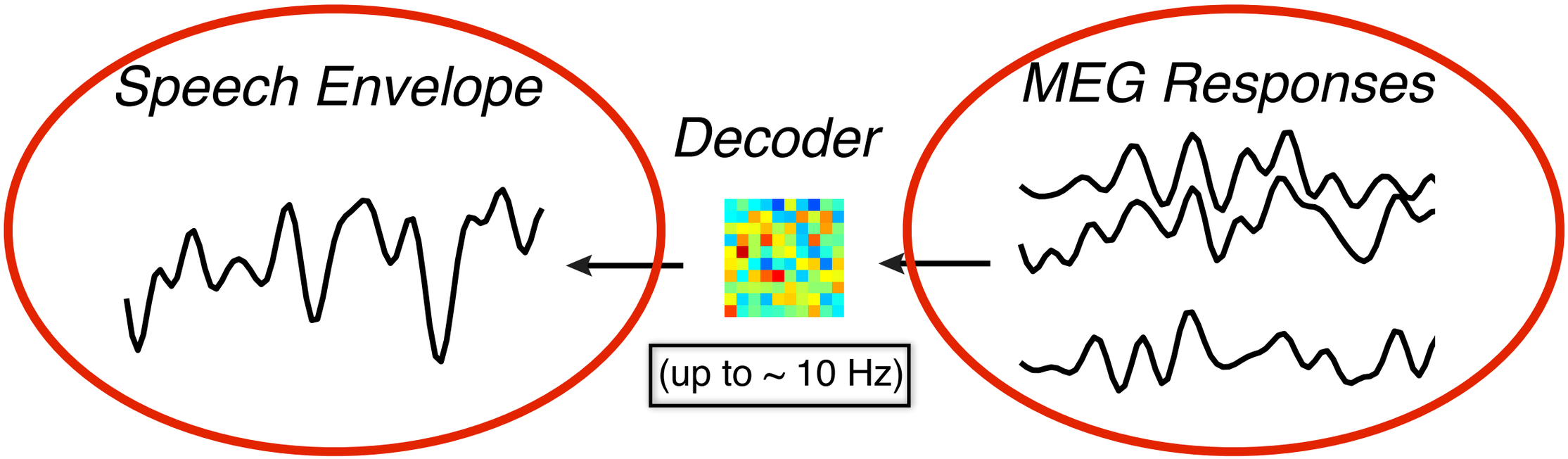
Neural Reconstruction of Speech Envelope



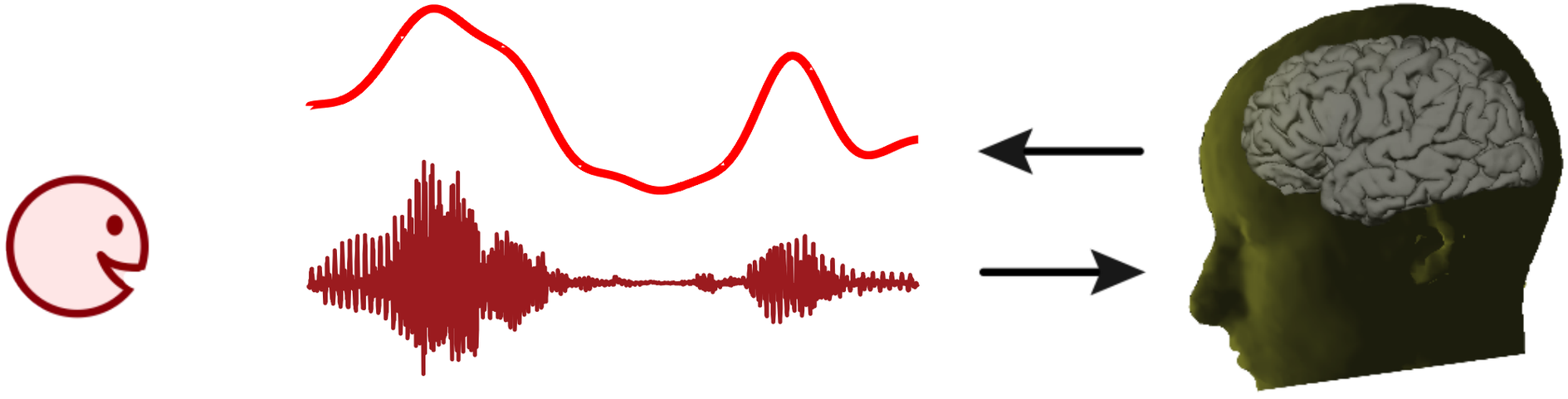


Reconstruction accuracy comparable to single unit & ECoG recordings

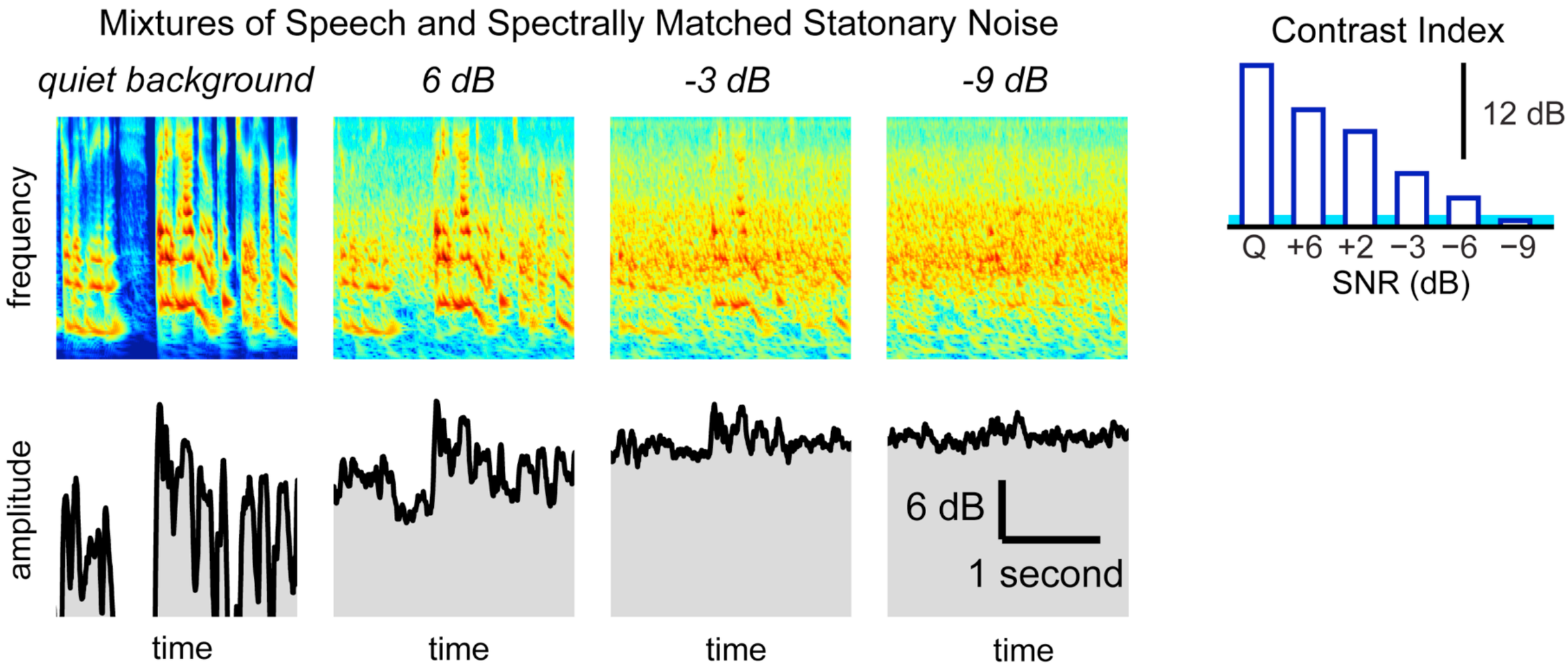
Ding & Simon, J Neurophysiol (2012)
Zion-Golumbic et al., Neuron (2013)


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

Neural Encoding of Speech: Temporal

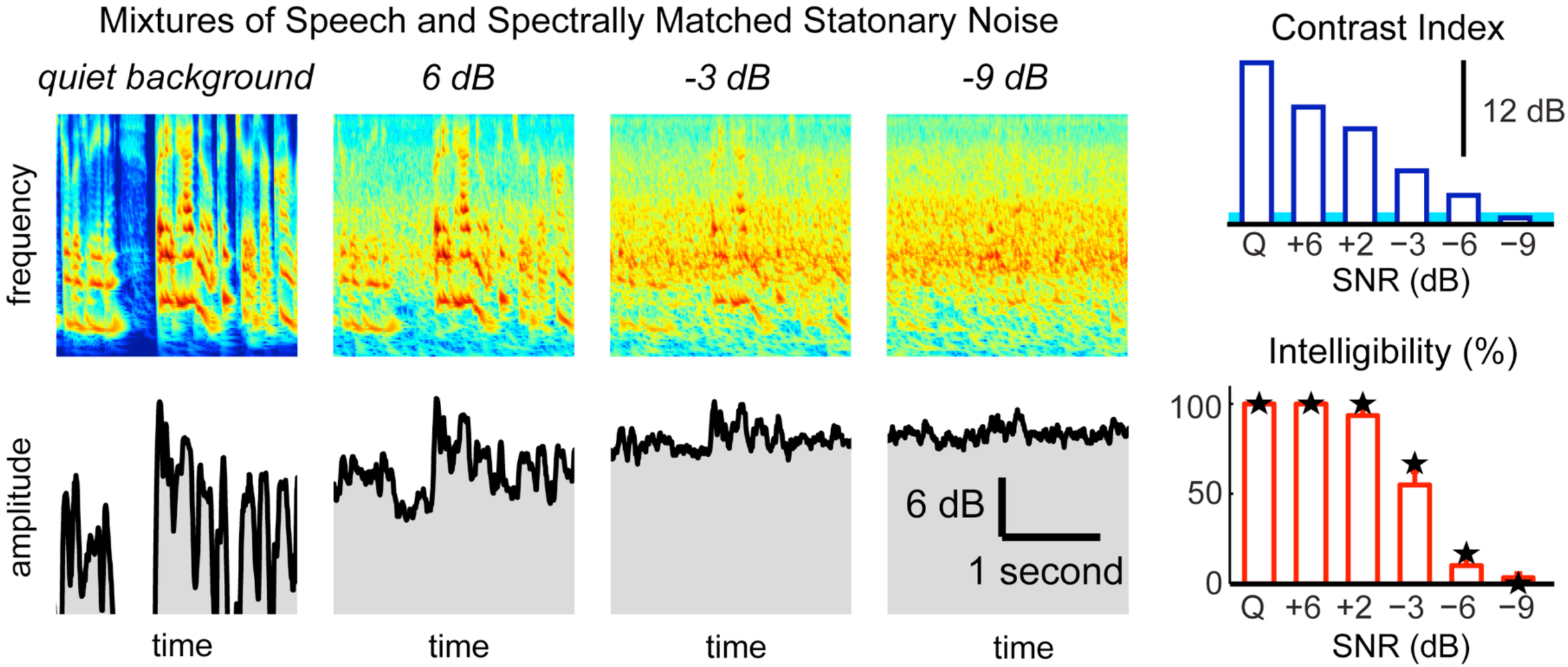


Speech in Noise



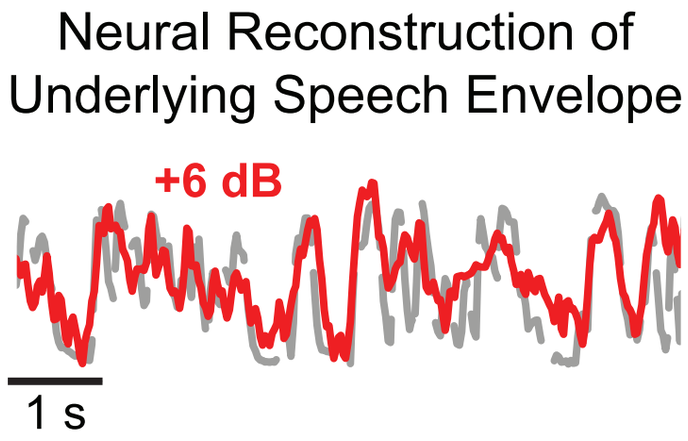
Ding & Simon, J Neuroscience (2013)

Speech in Noise

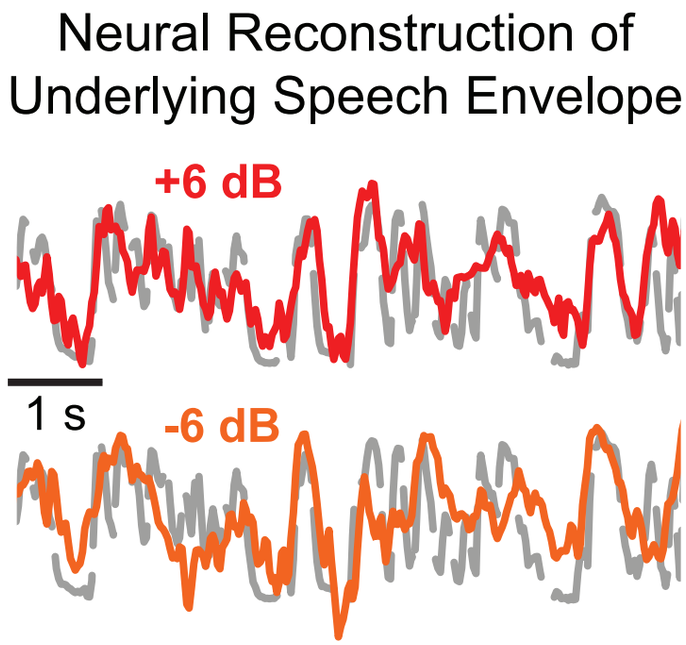


Ding & Simon, J Neuroscience (2013)

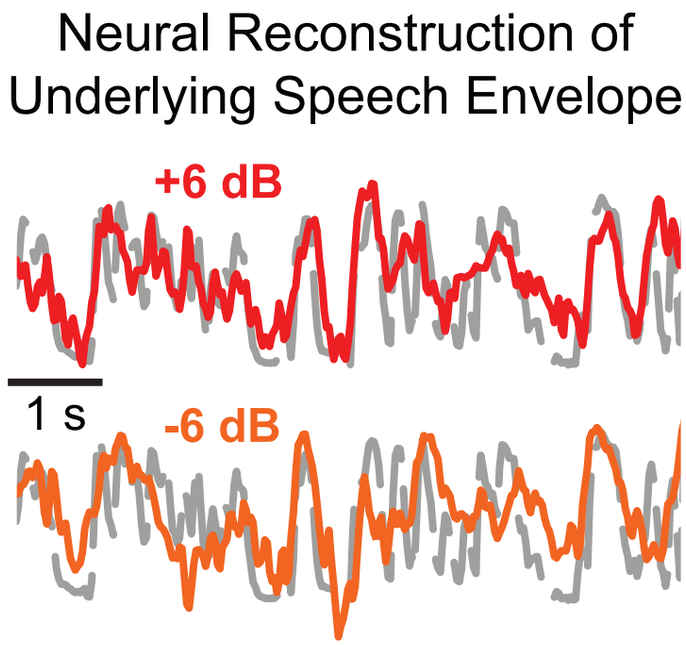
Speech in Noise: Results

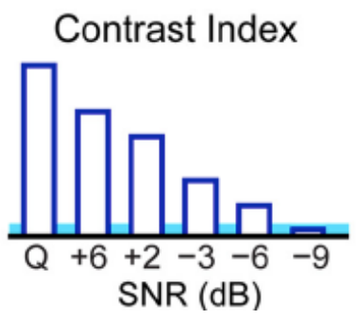


Speech in Noise: Results

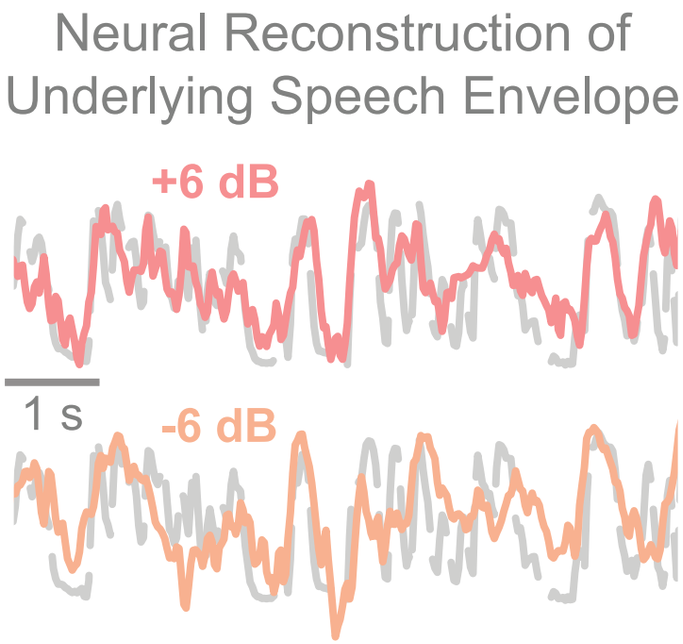


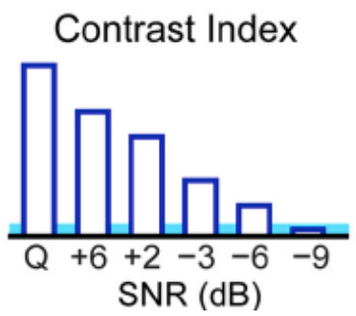
Speech in Noise: Results

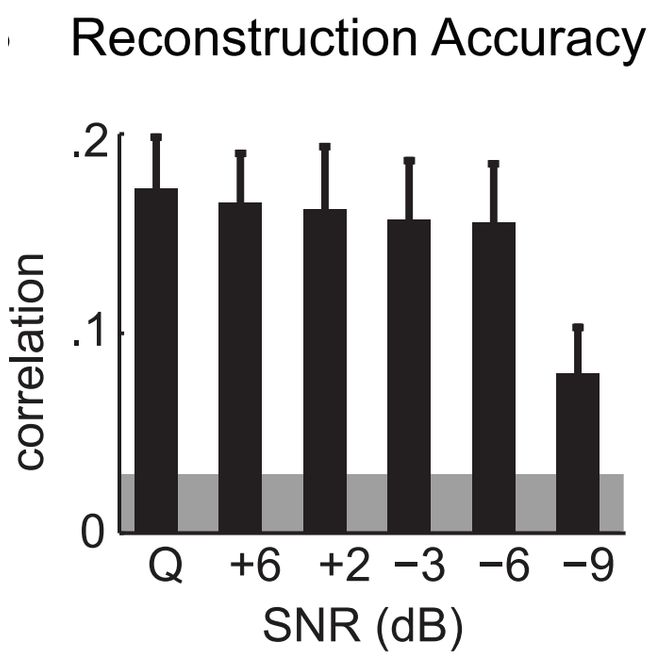




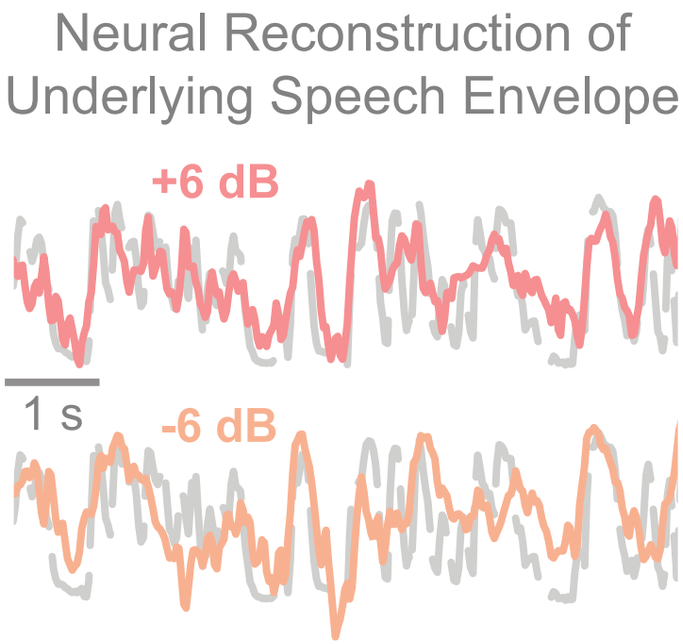
Speech in Noise: Results

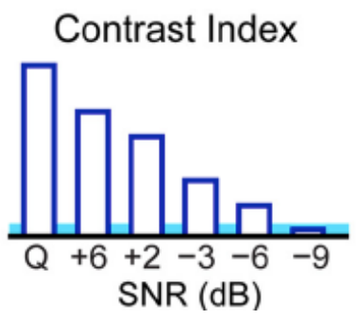


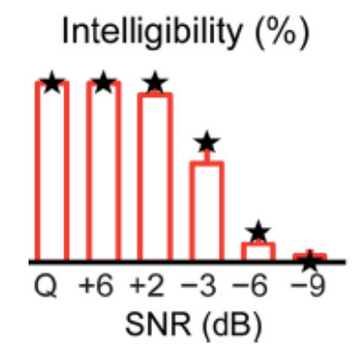


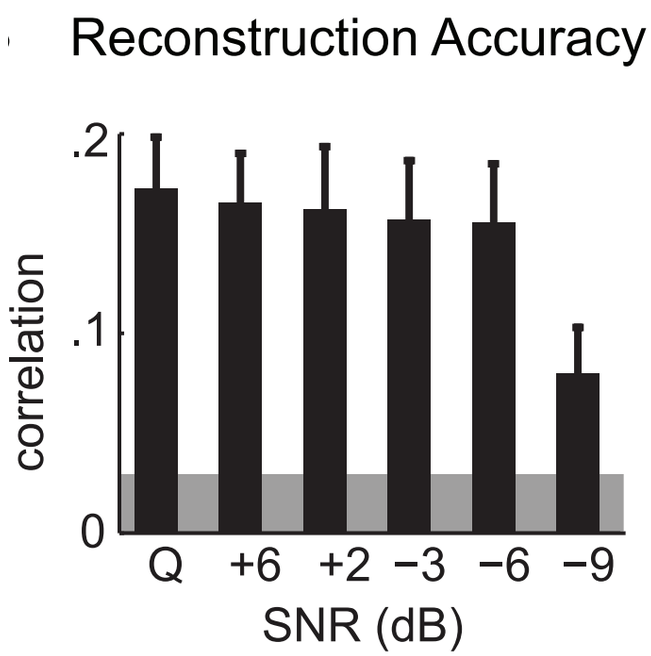


Speech in Noise: Results

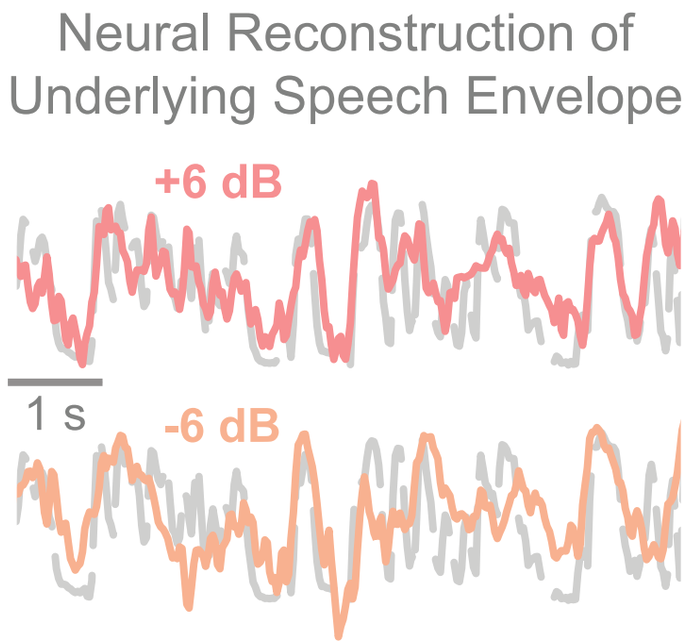


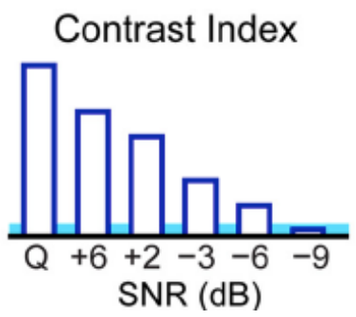


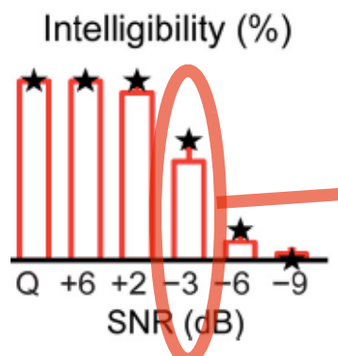


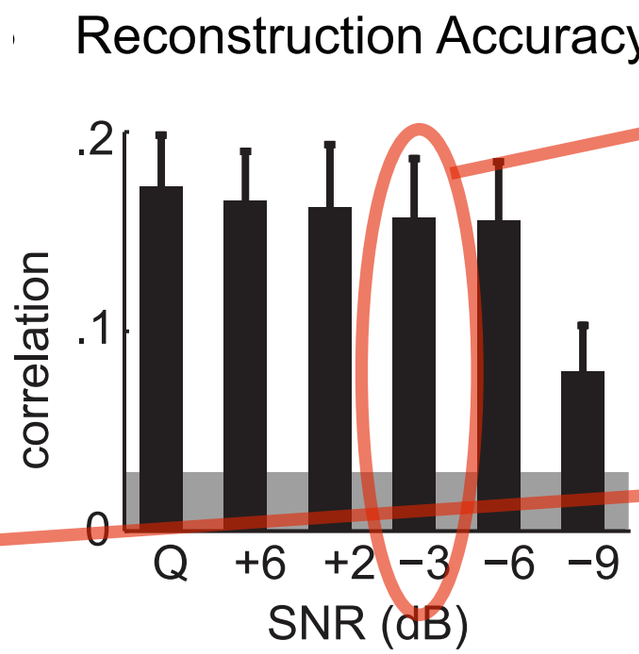


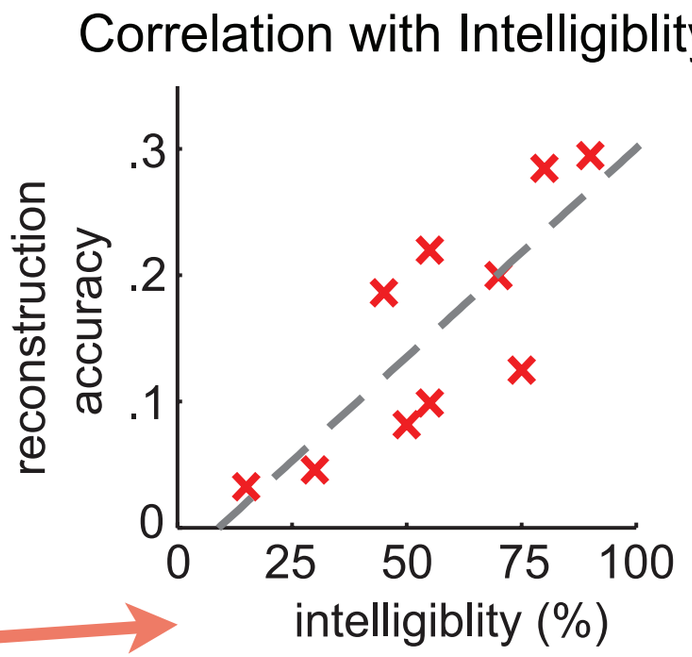
Speech in Noise: Results











across Subjects

The Cocktail Party

Alex Katz,
The Cocktail Party

The Cocktail Party

Alex Katz,
The Cocktail Party

The Cocktail Party

Alex Katz,
The Cocktail Party

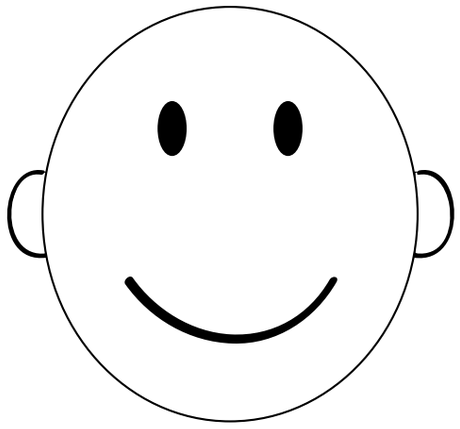
The Cocktail Party

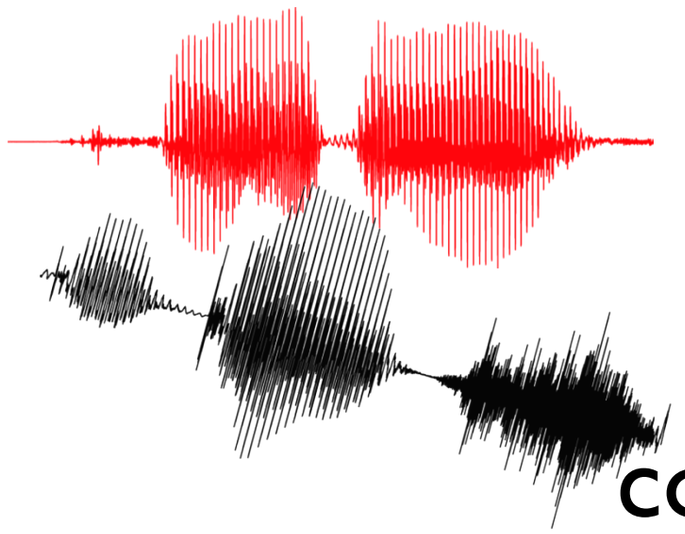
Alex Katz,
The Cocktail Party

The Cocktail Party

Alex Katz,
The Cocktail Party

Experiments

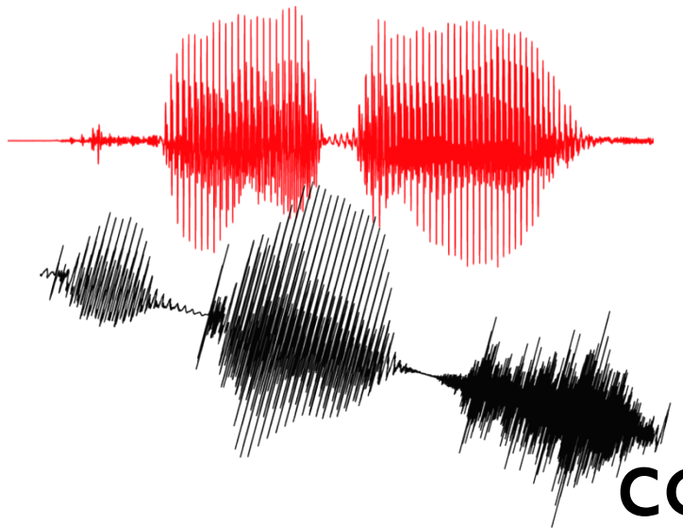




speech

competing speech

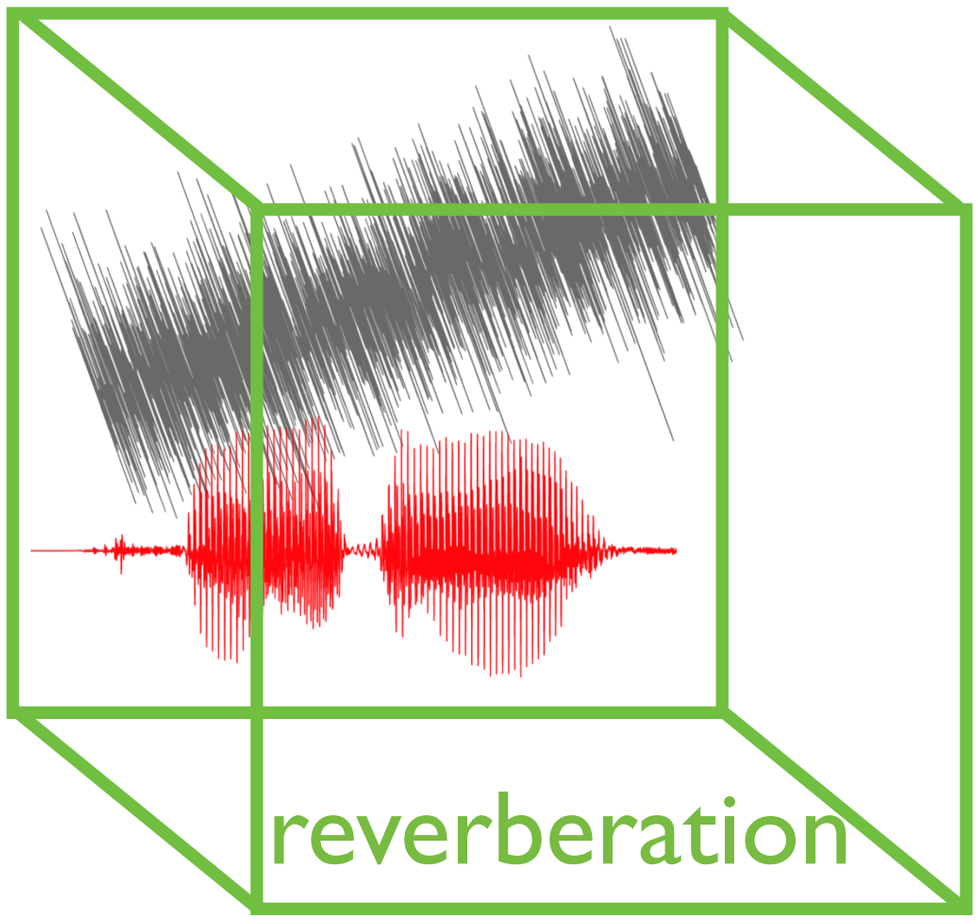
Experiments



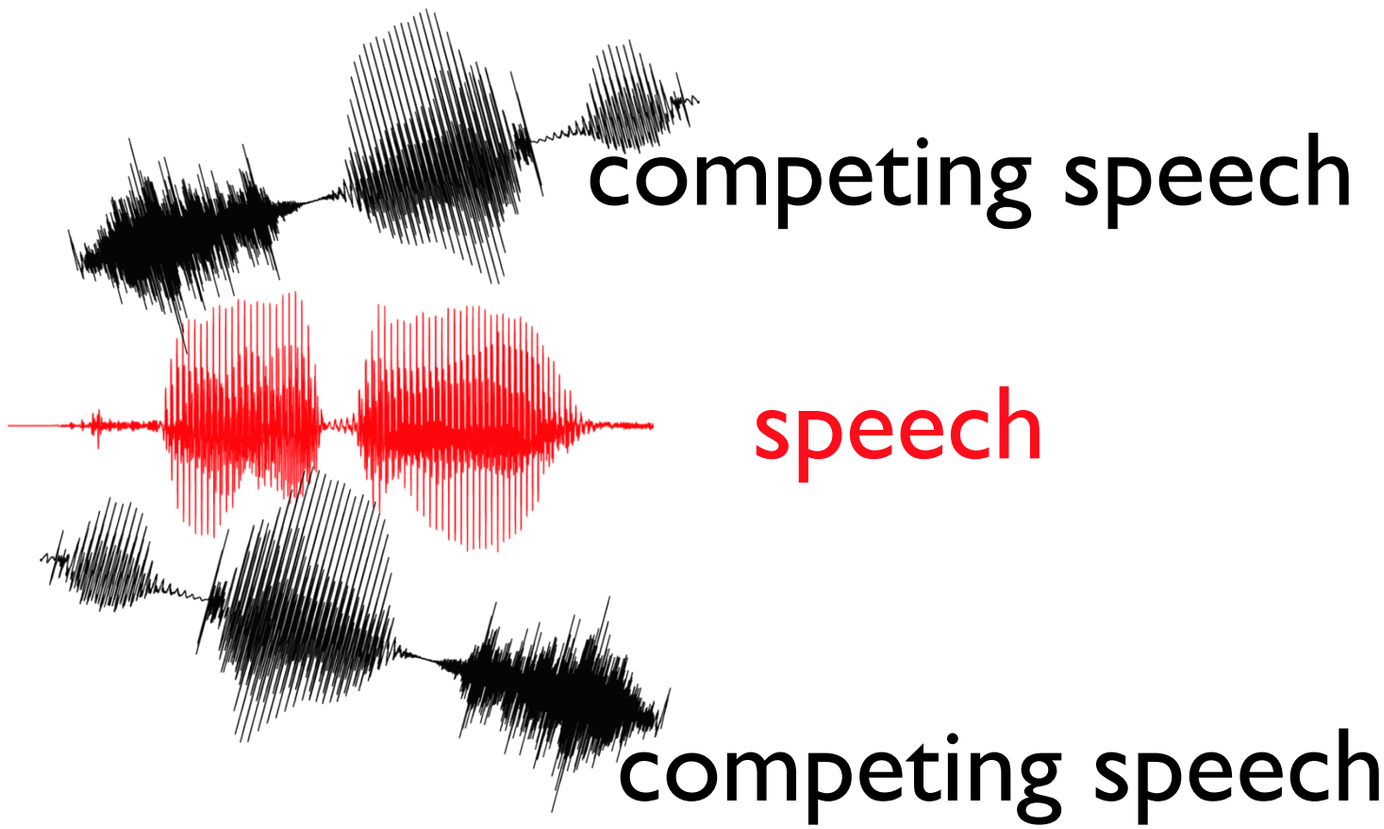
speech

competing speech

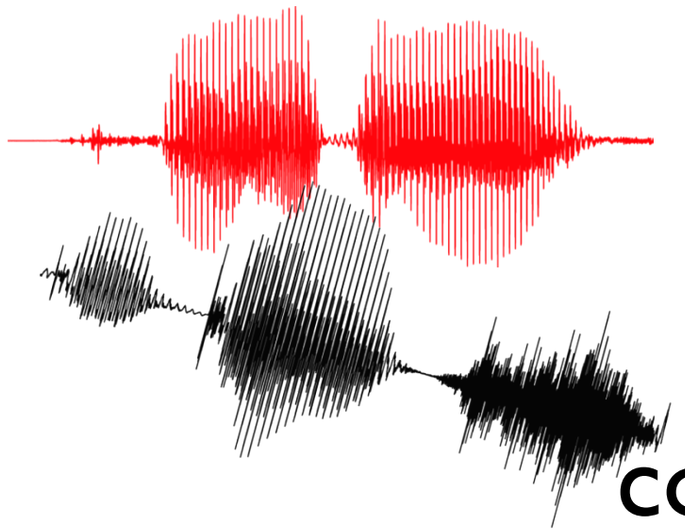
Experiments in Progress



Experiments in Progress



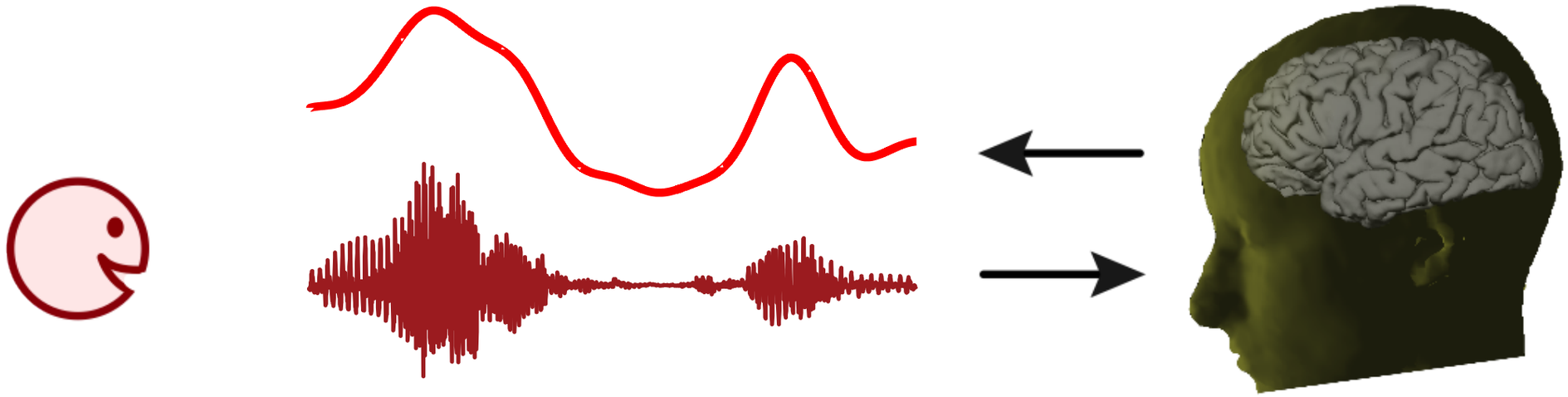
Two Competing Speakers



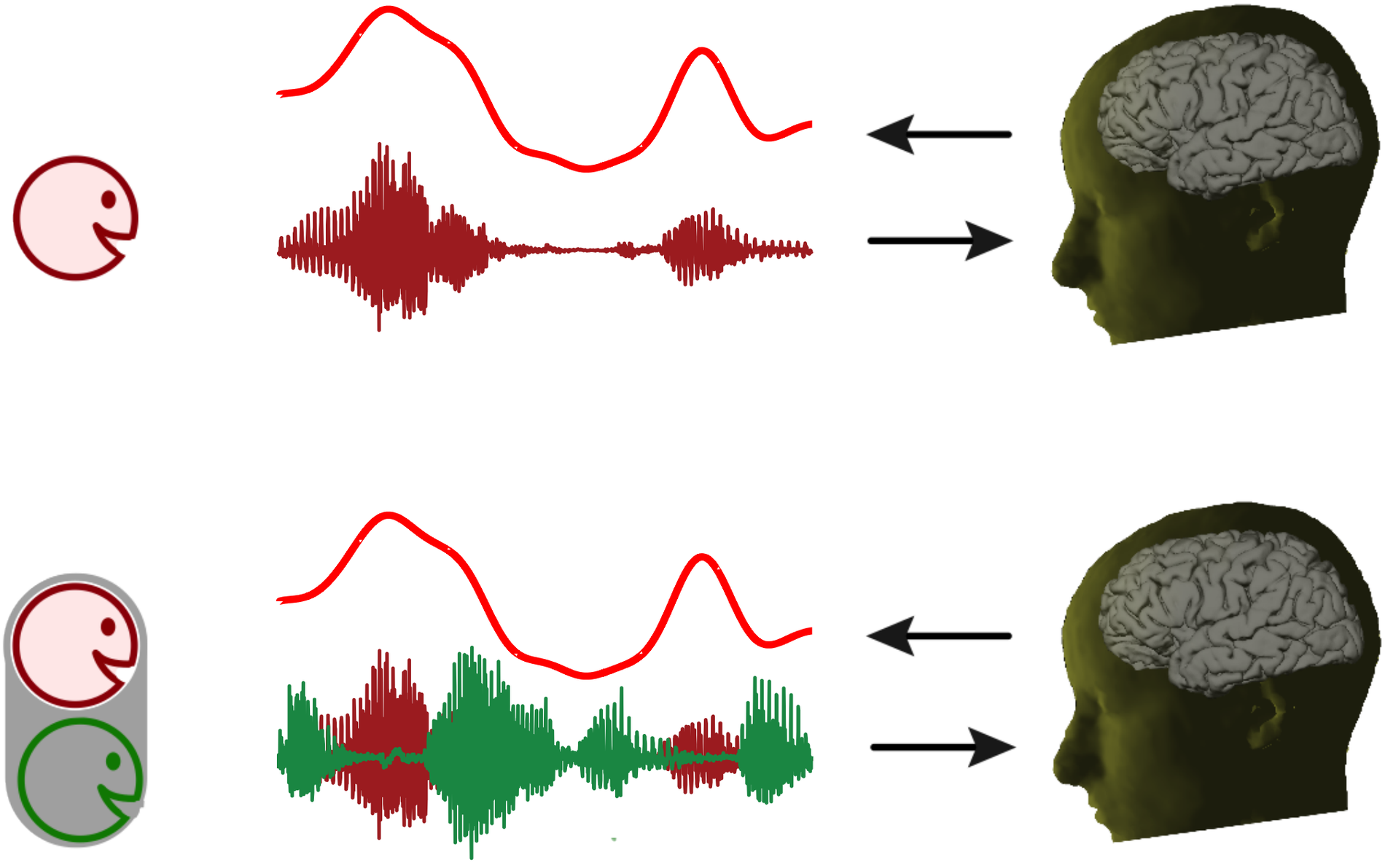
speech

competing speech

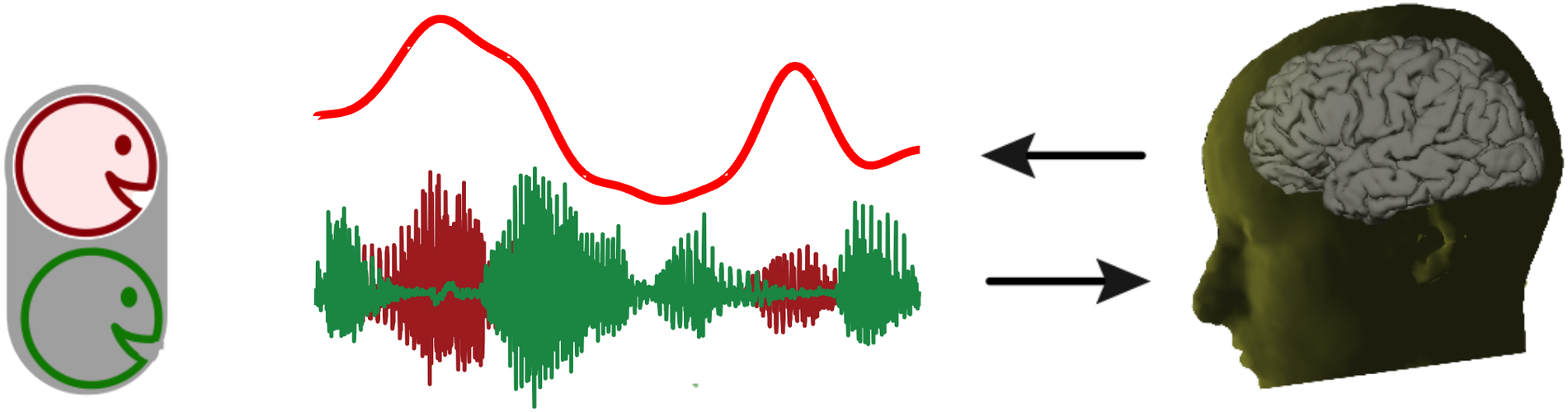
Selective Neural Encoding



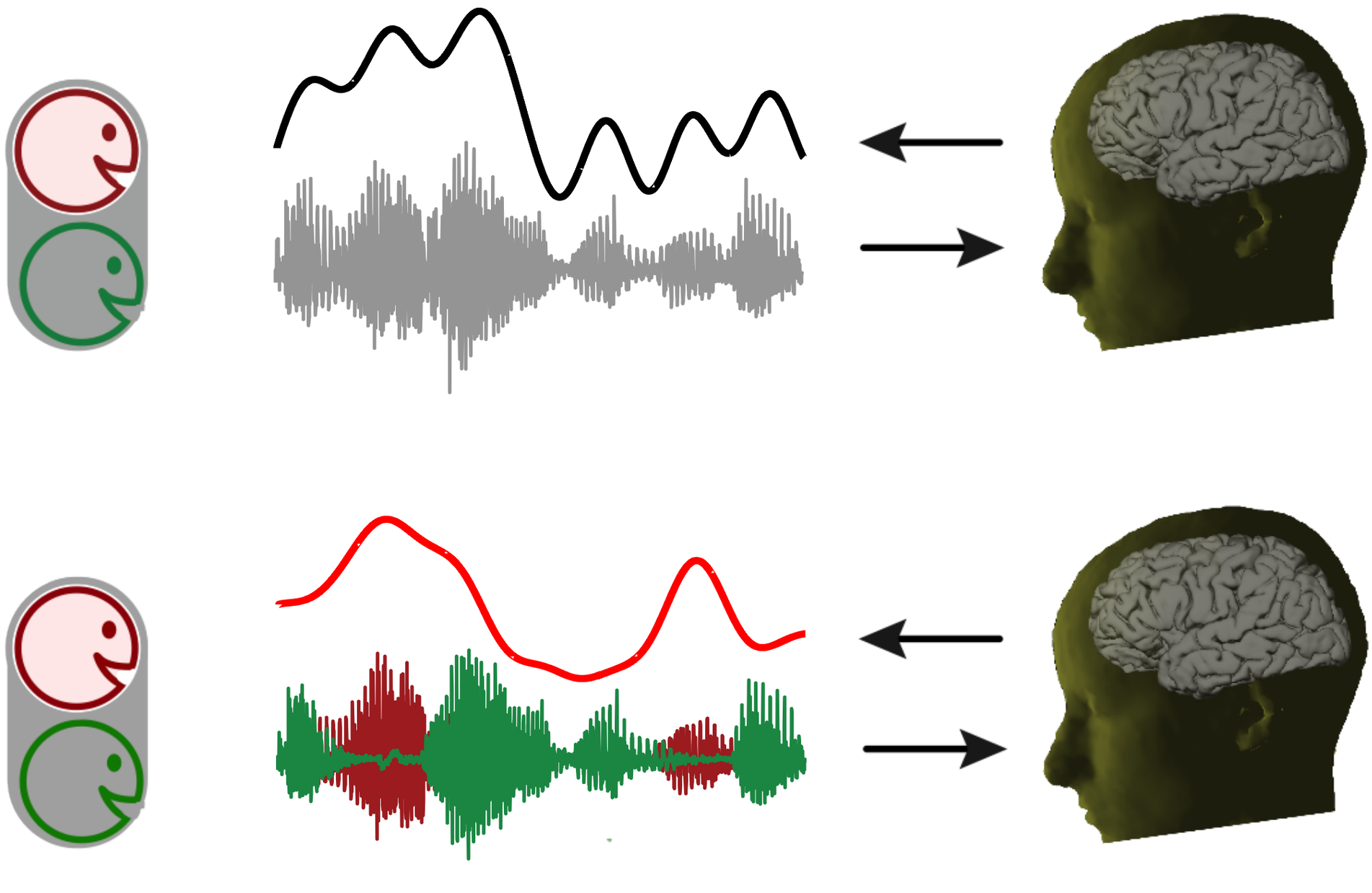
Selective Neural Encoding



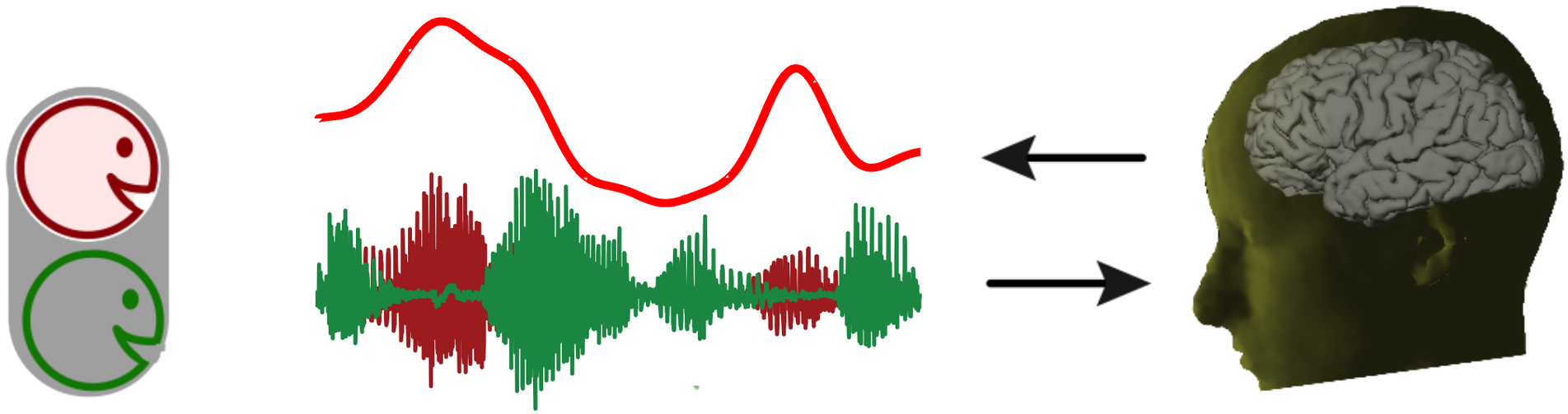
Selective Neural Encoding



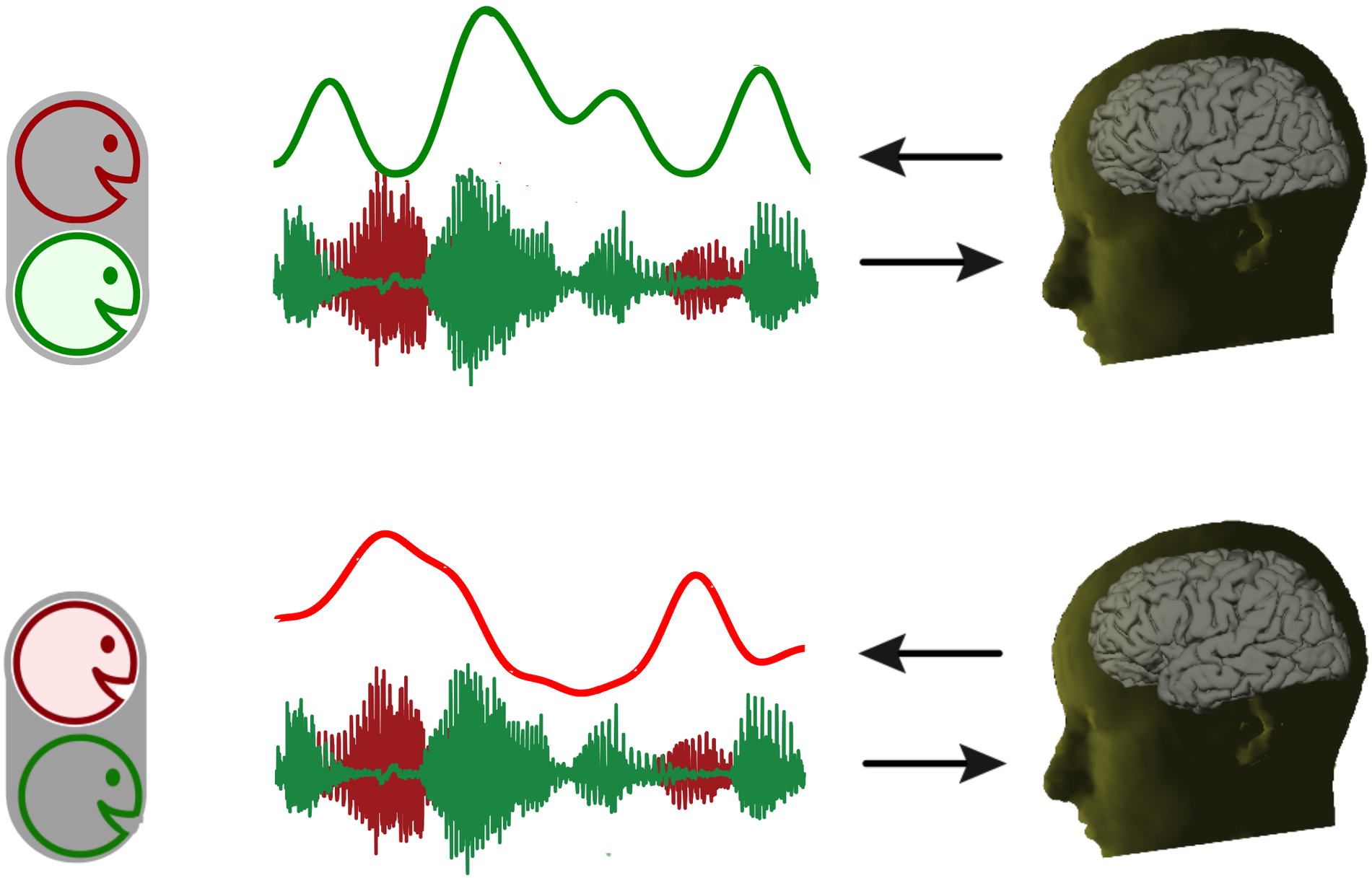
Unselective vs. Selective Neural Encoding



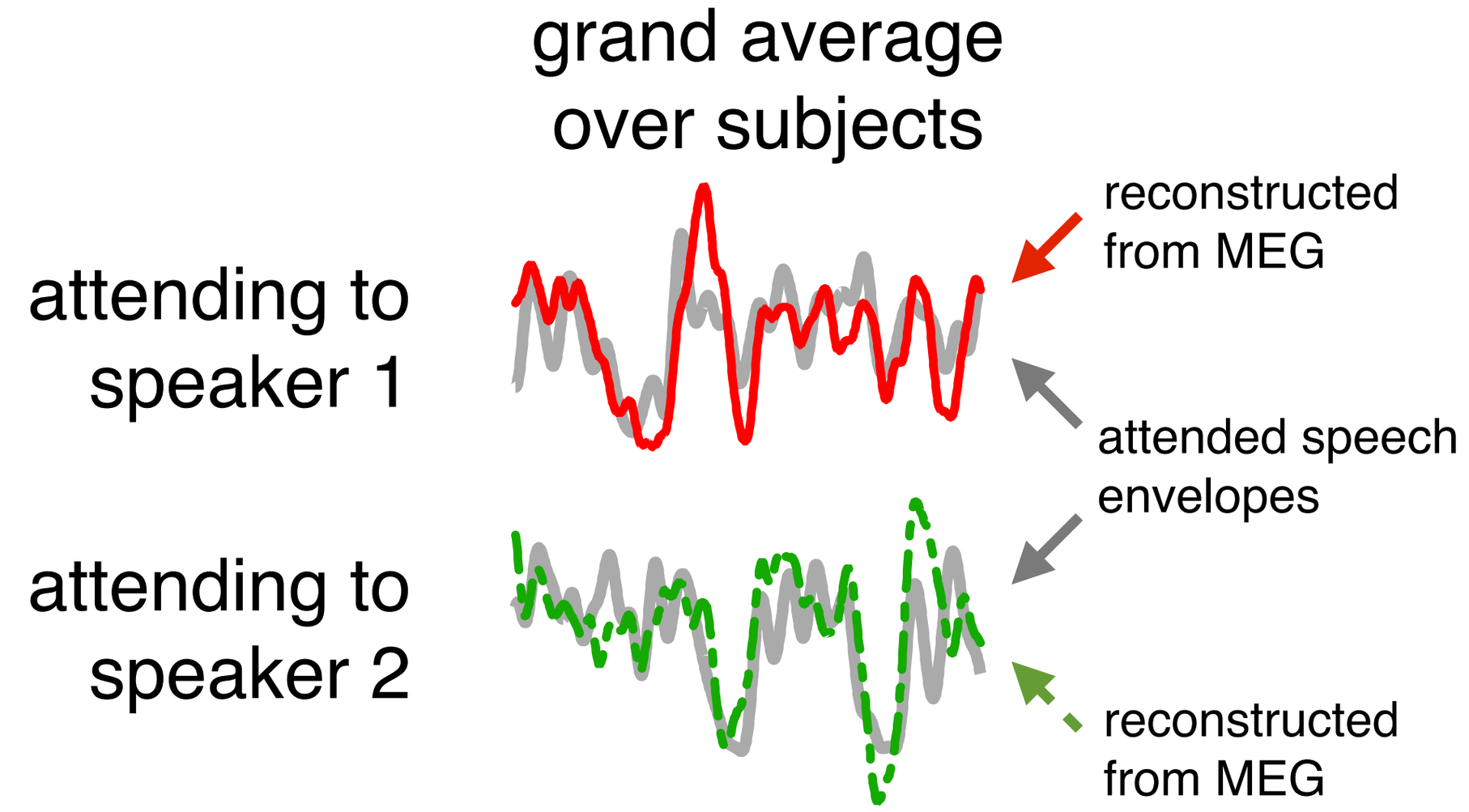
Unselective vs. Selective Neural Encoding



Selective Neural Encoding



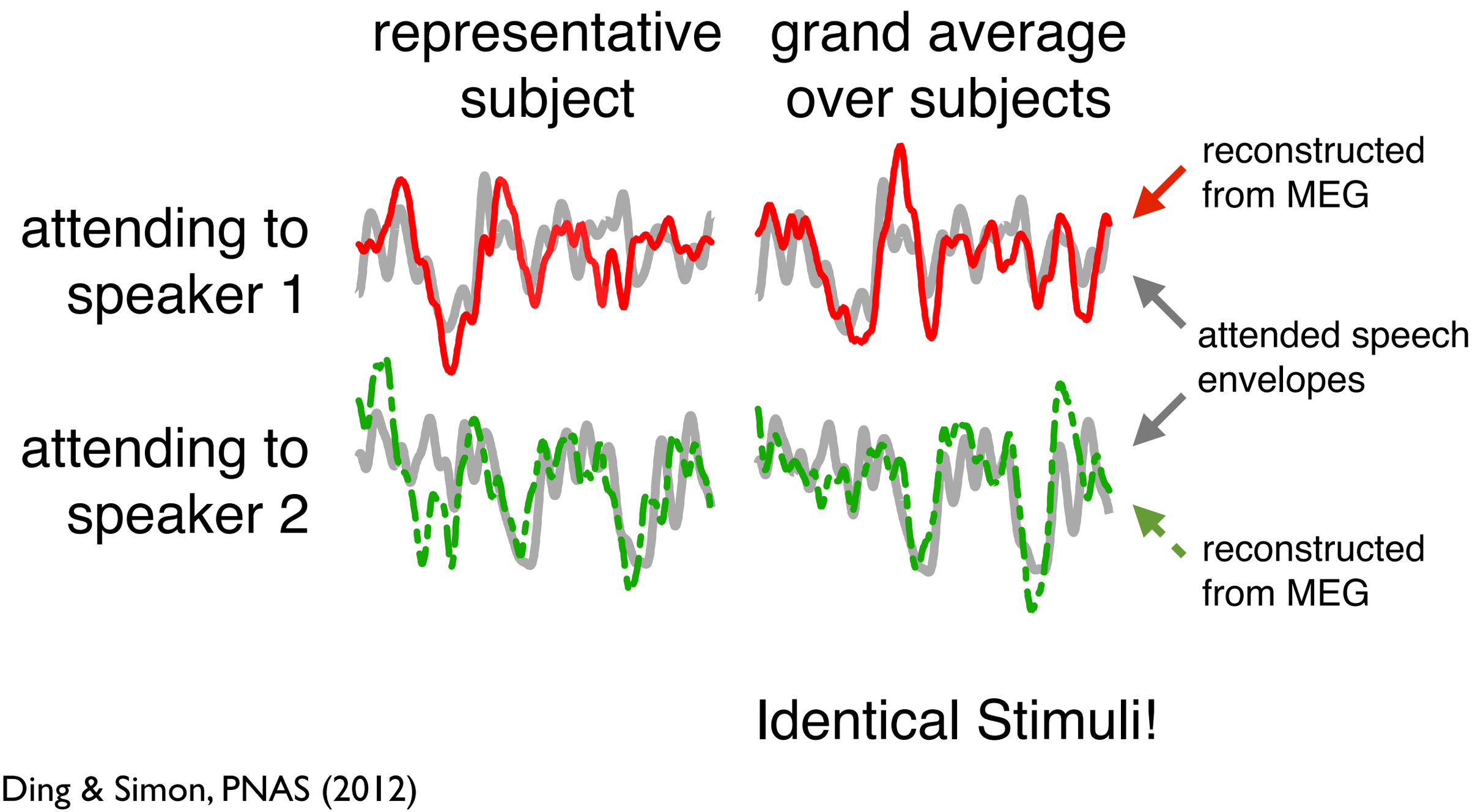
Stream-Specific Representation



Identical Stimuli!

Ding & Simon, PNAS (2012)

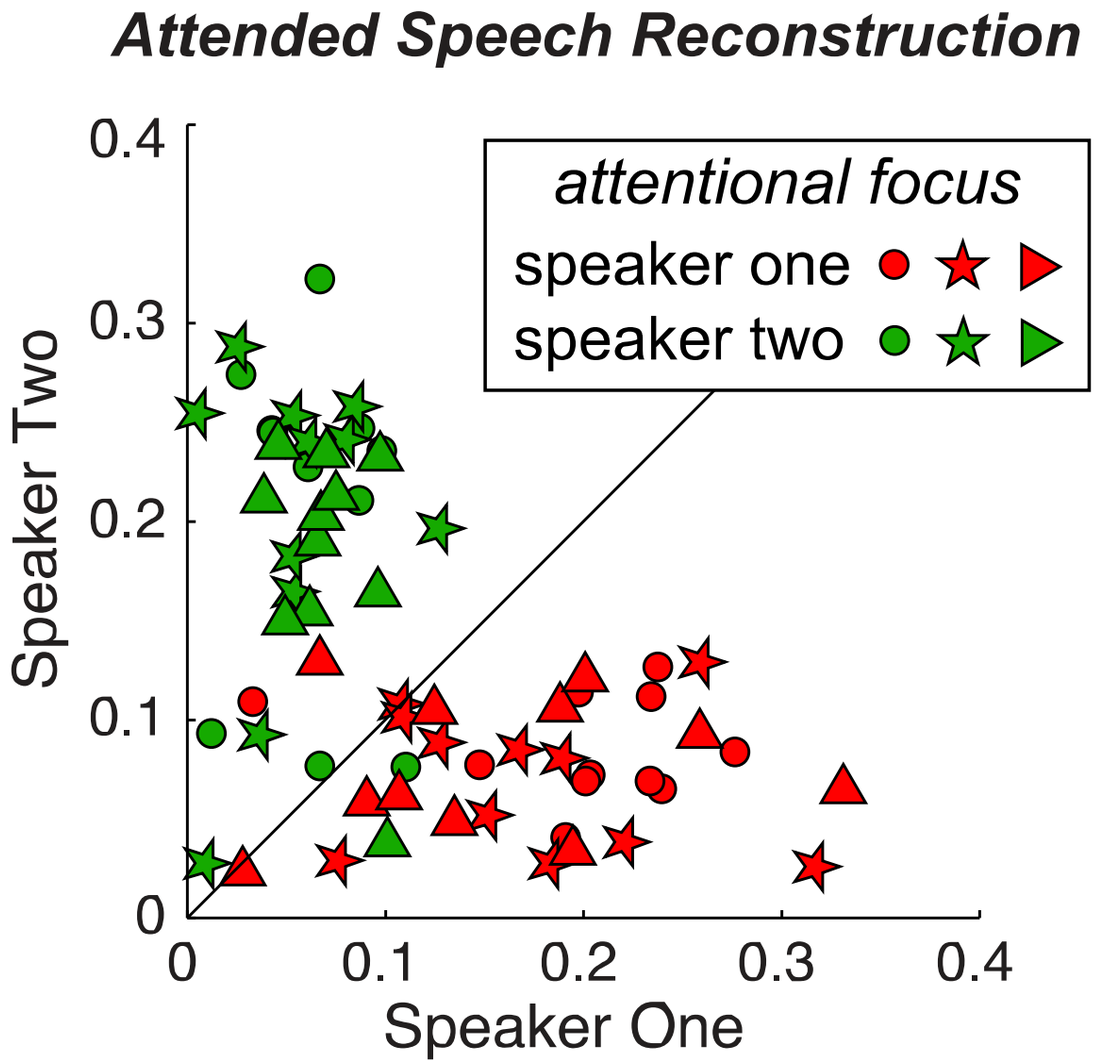
Stream-Specific Representation



Identical Stimuli!

Ding & Simon, PNAS (2012)

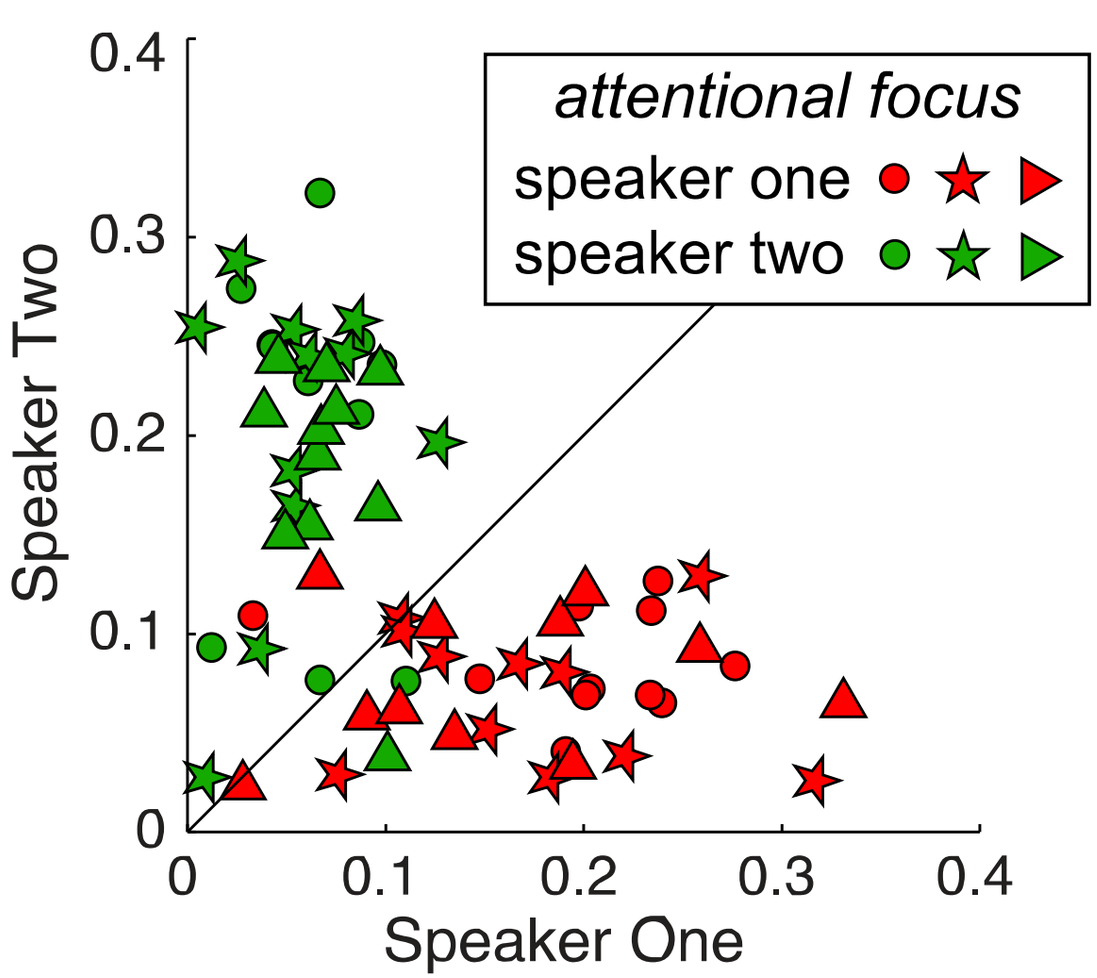
Single Trial Speech Reconstruction



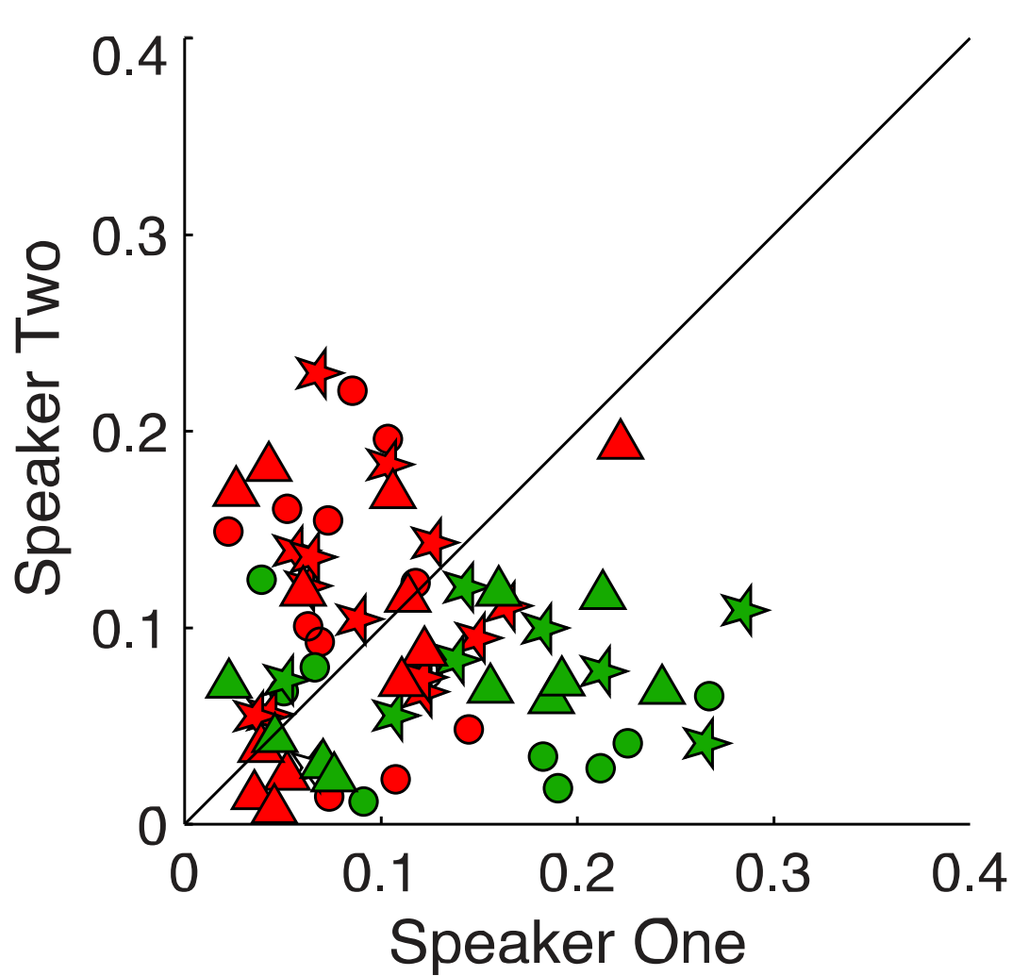
Ding & Simon, PNAS (2012)

Single Trial Speech Reconstruction

Attended Speech Reconstruction

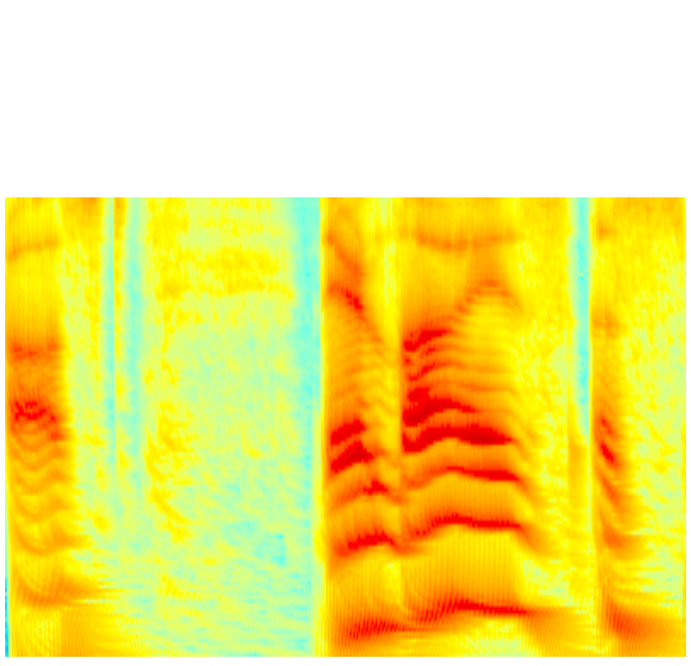


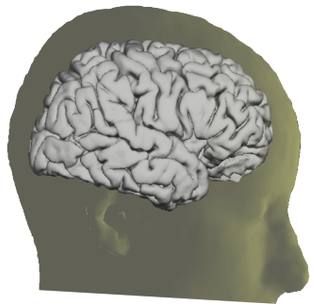
Background Speech Reconstruction

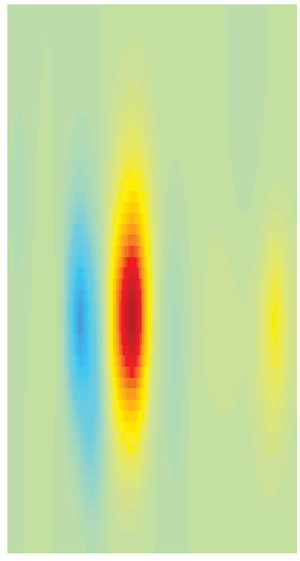


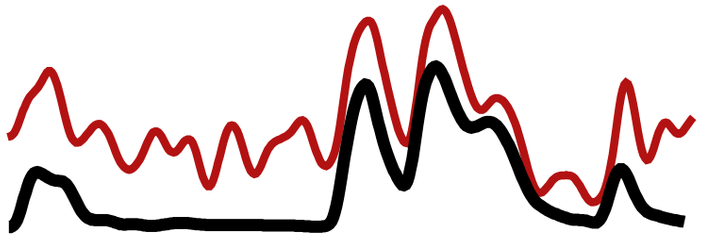
Ding & Simon, PNAS (2013)

Forward STRF Model



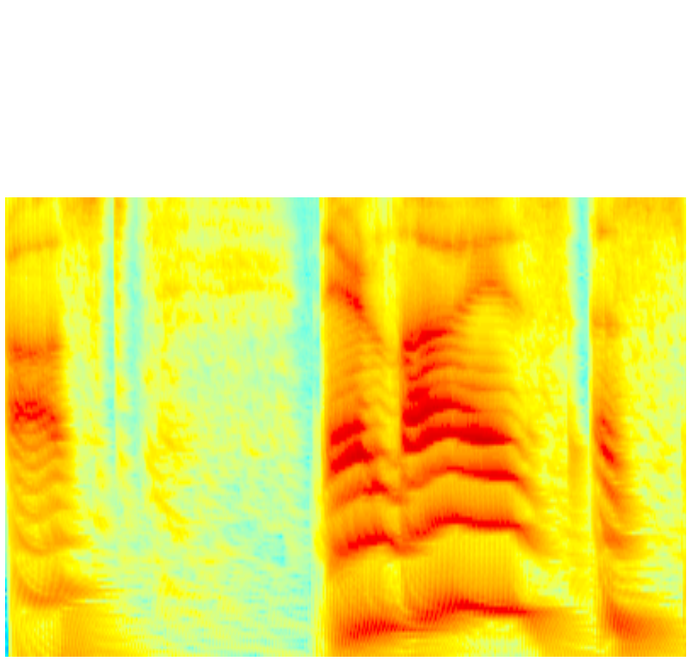


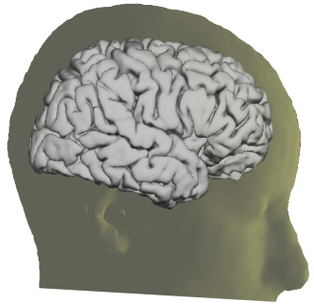


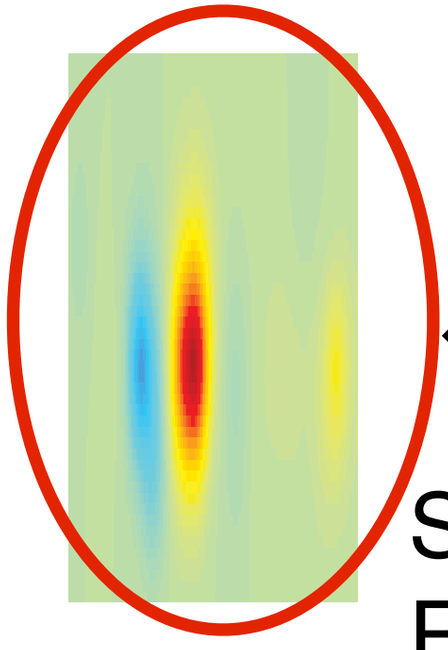


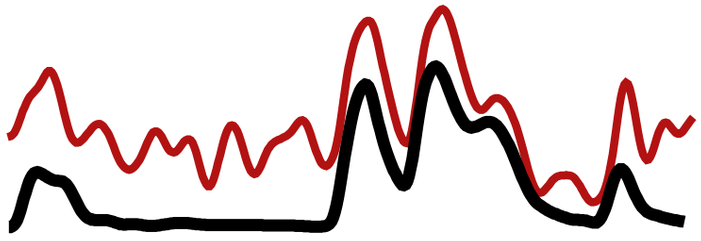
Spectro-Temporal Response Function (STRF)

Forward STRF Model



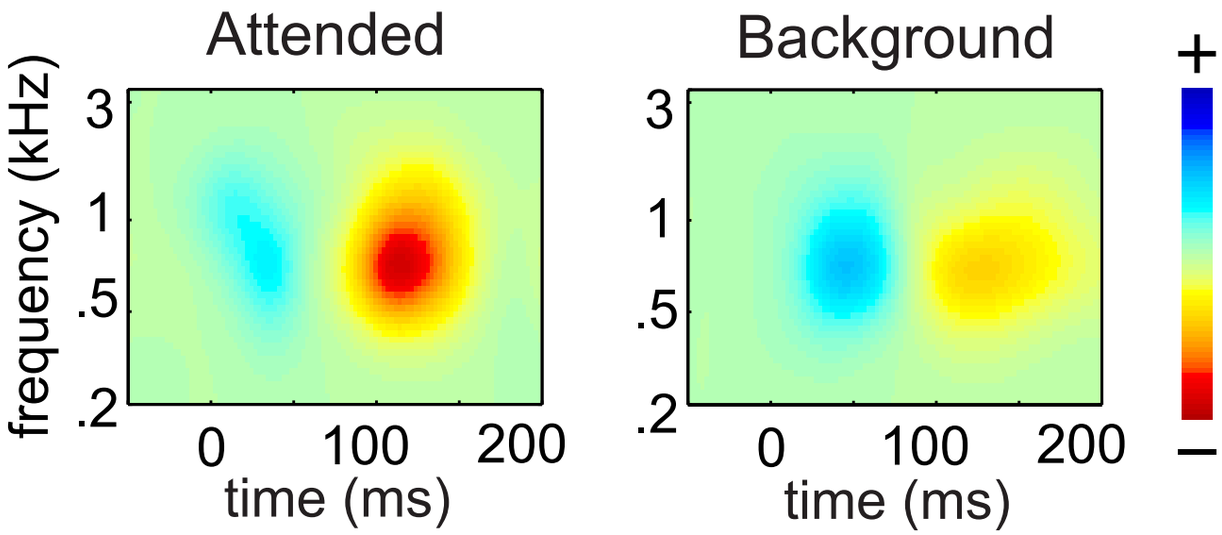






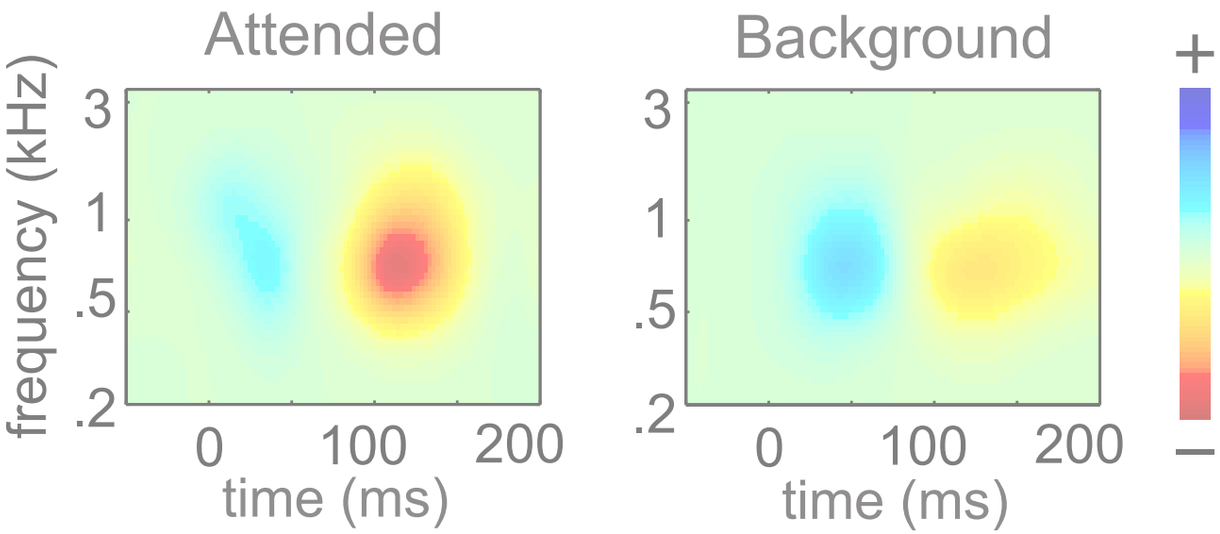
Spectro-Temporal
Response Function
(STRF)

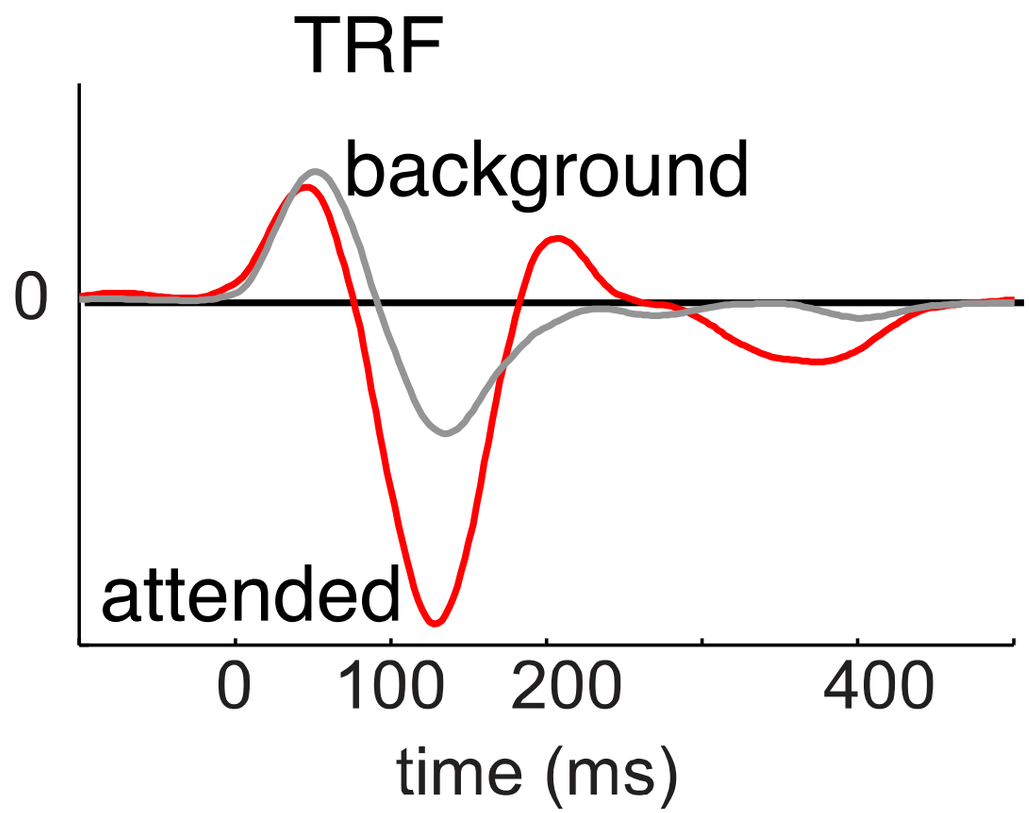
STRF Results



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

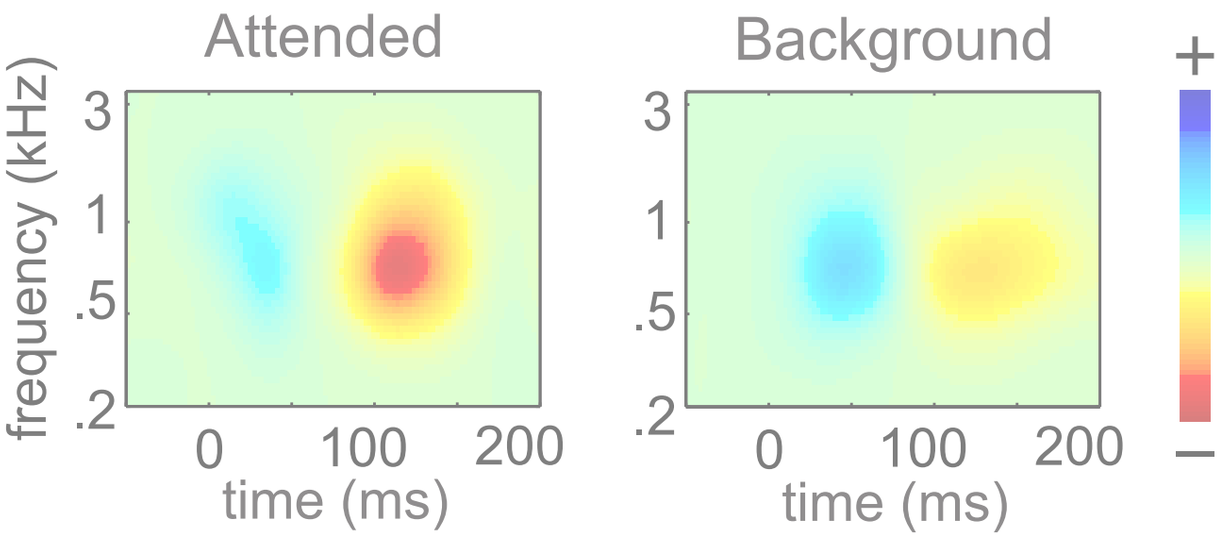
STRF Results

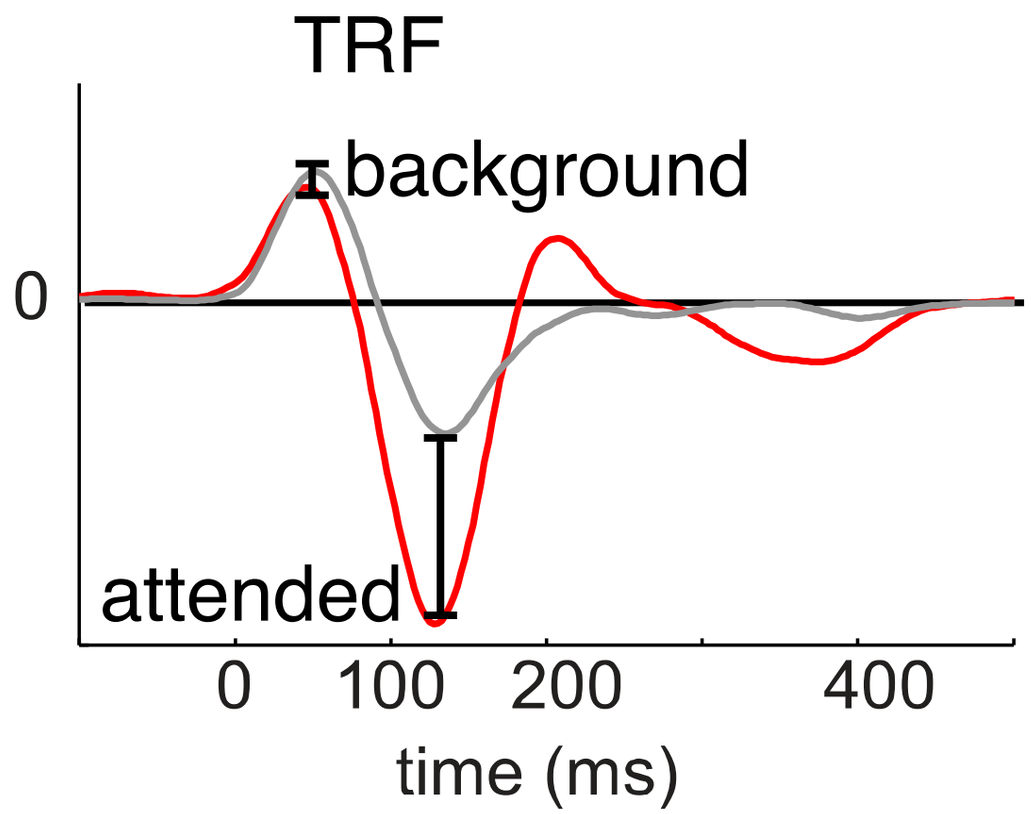




- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- $M50_{STRF}$ positive peak
- $M100_{STRF}$ negative peak

STRF Results





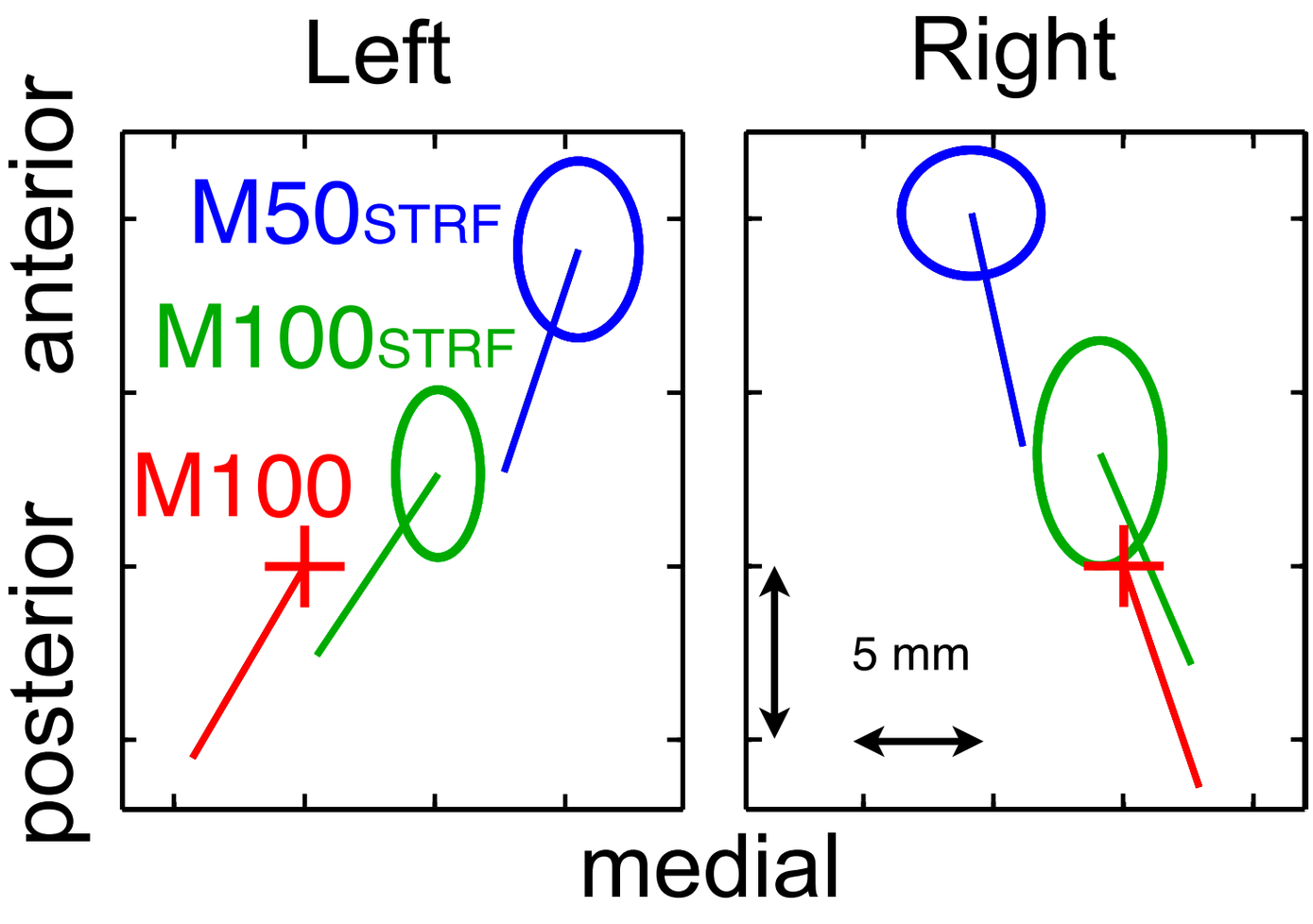
- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- $M50_{STRF}$ positive peak
- $M100_{STRF}$ negative peak
- **$M100_{STRF}$ strongly modulated by attention, *but not $M50_{STRF}$***

Neural Sources

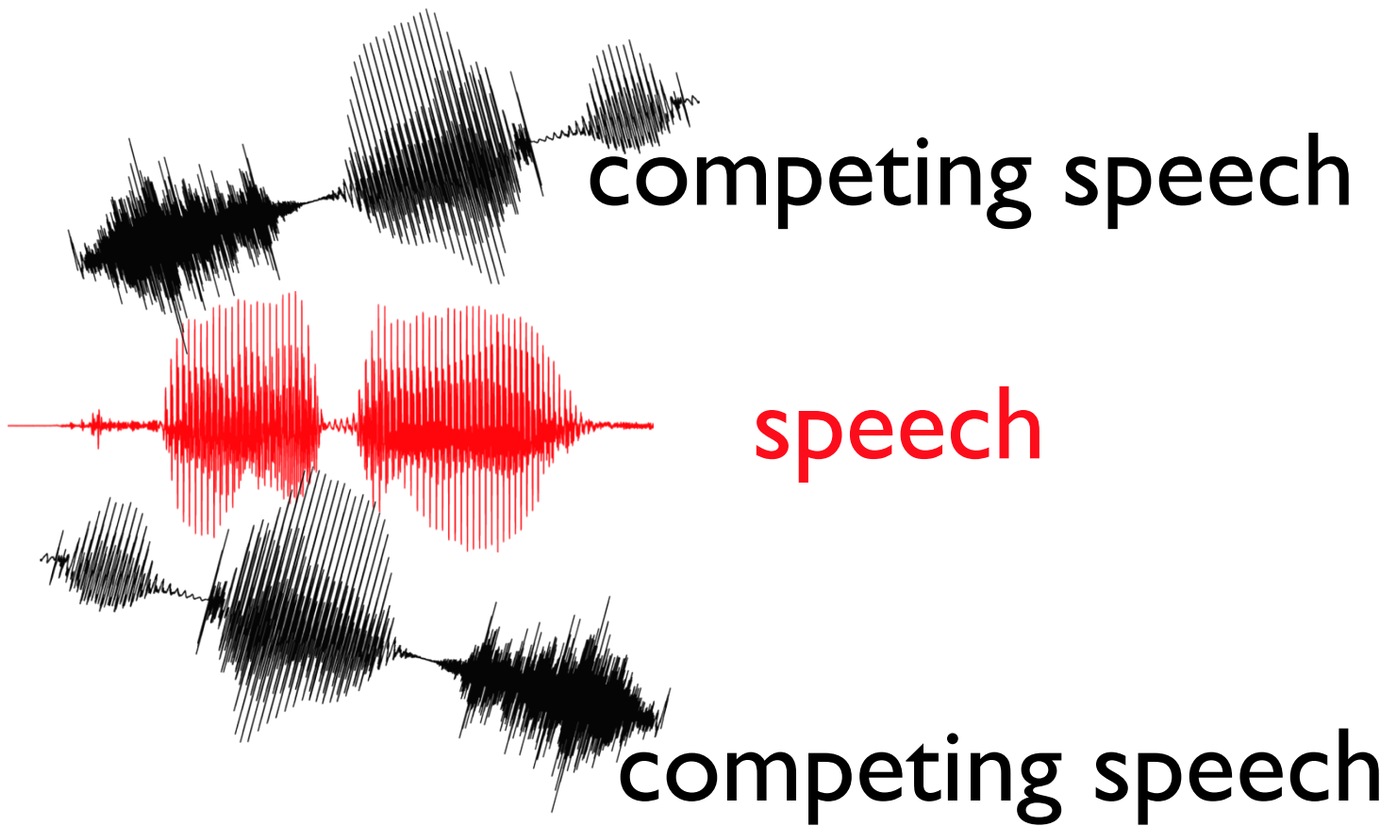
- M100$_{STRF}$ source near (same as?) M100 source: Planum Temporale

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

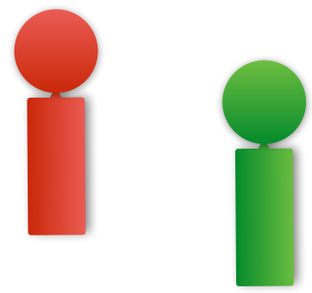
- **PT strongly modulated by attention, *but not HG***

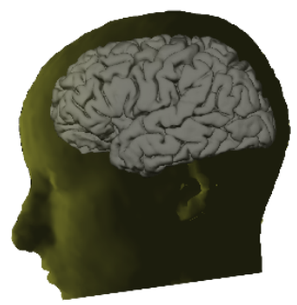


Three Competing Speakers

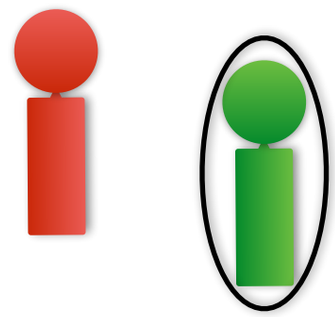


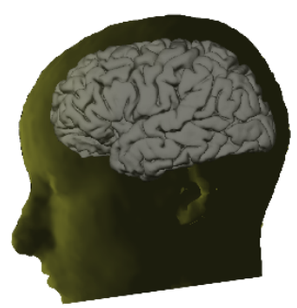
Foreground vs. Background



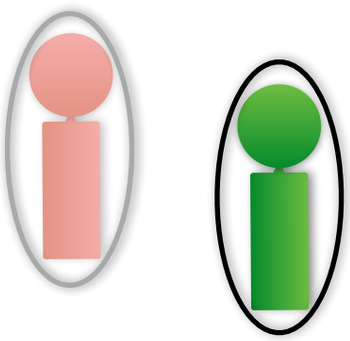


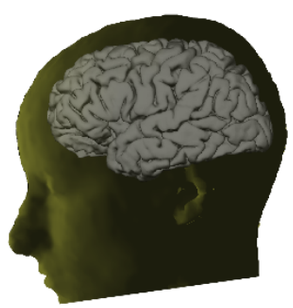
Foreground vs. Background



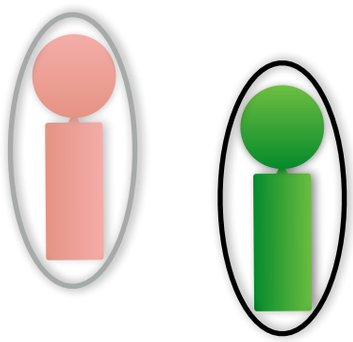


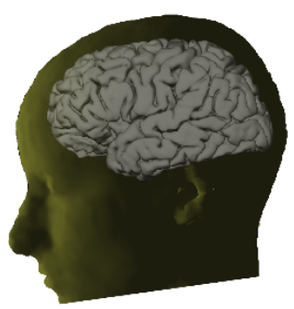
Foreground vs. Background

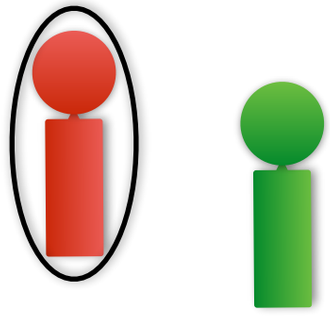




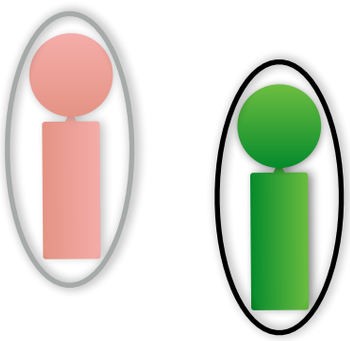
Foreground vs. Background

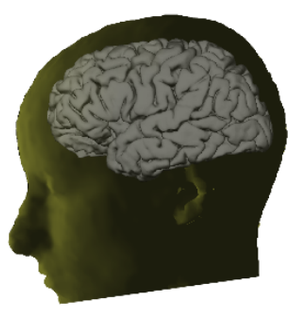


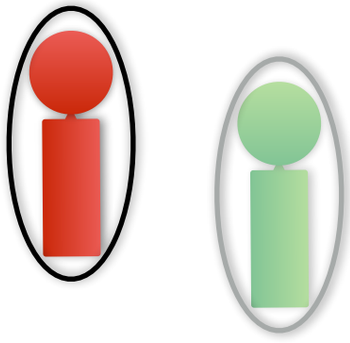




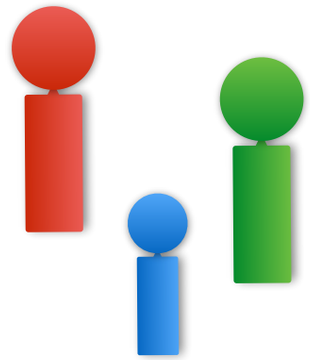
Foreground vs. Background



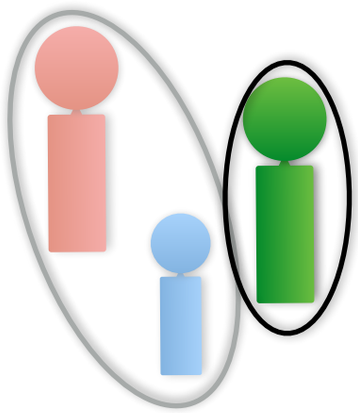


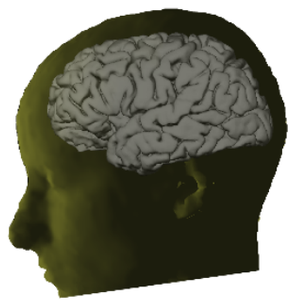


Foreground vs. Background

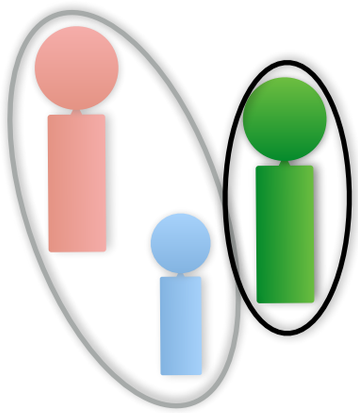


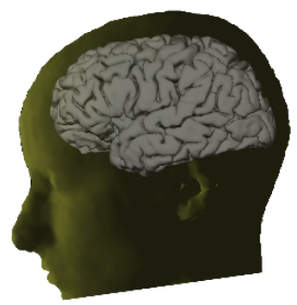
Foreground vs. Background

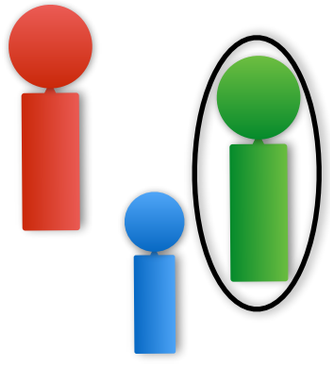




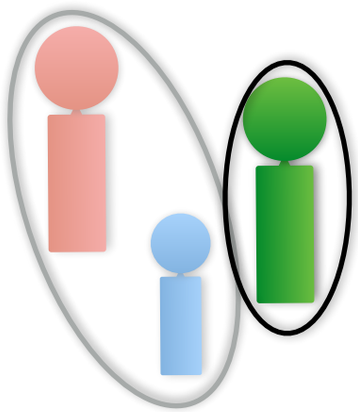
Foreground vs. Background

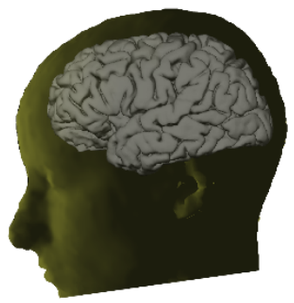


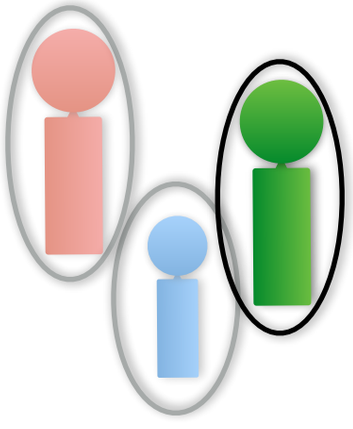




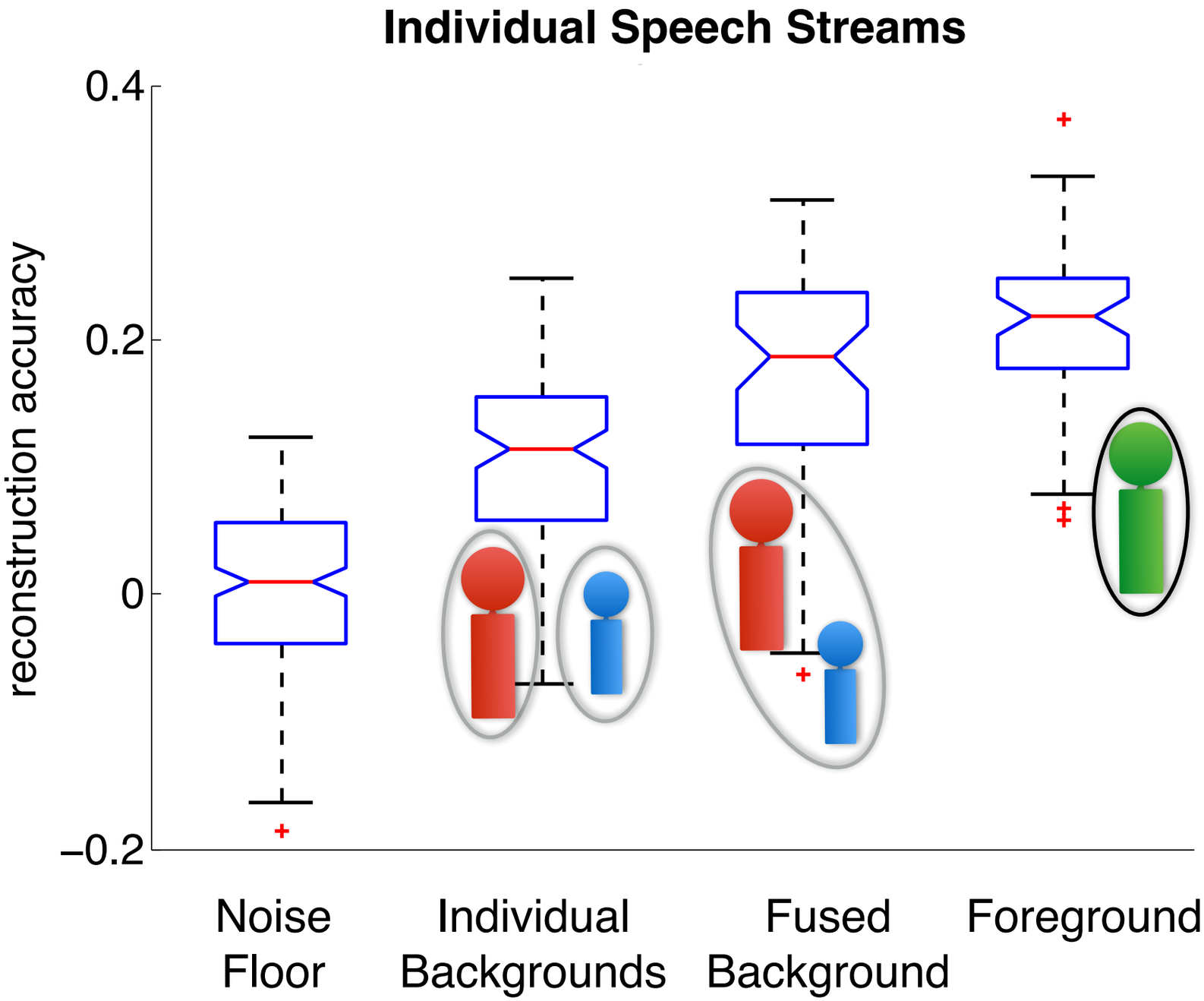
Foreground vs. Background





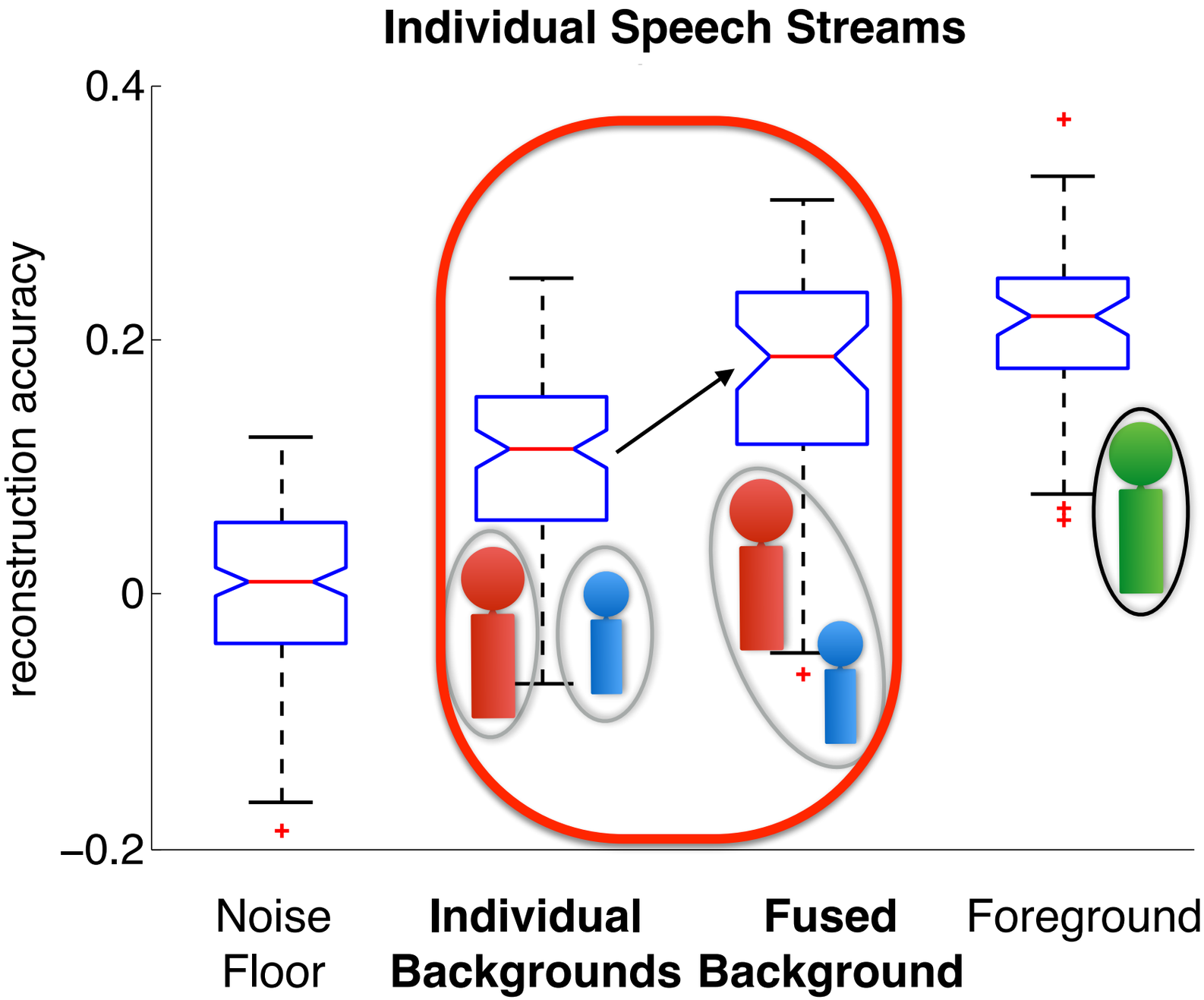


Backgrounds vs. Background



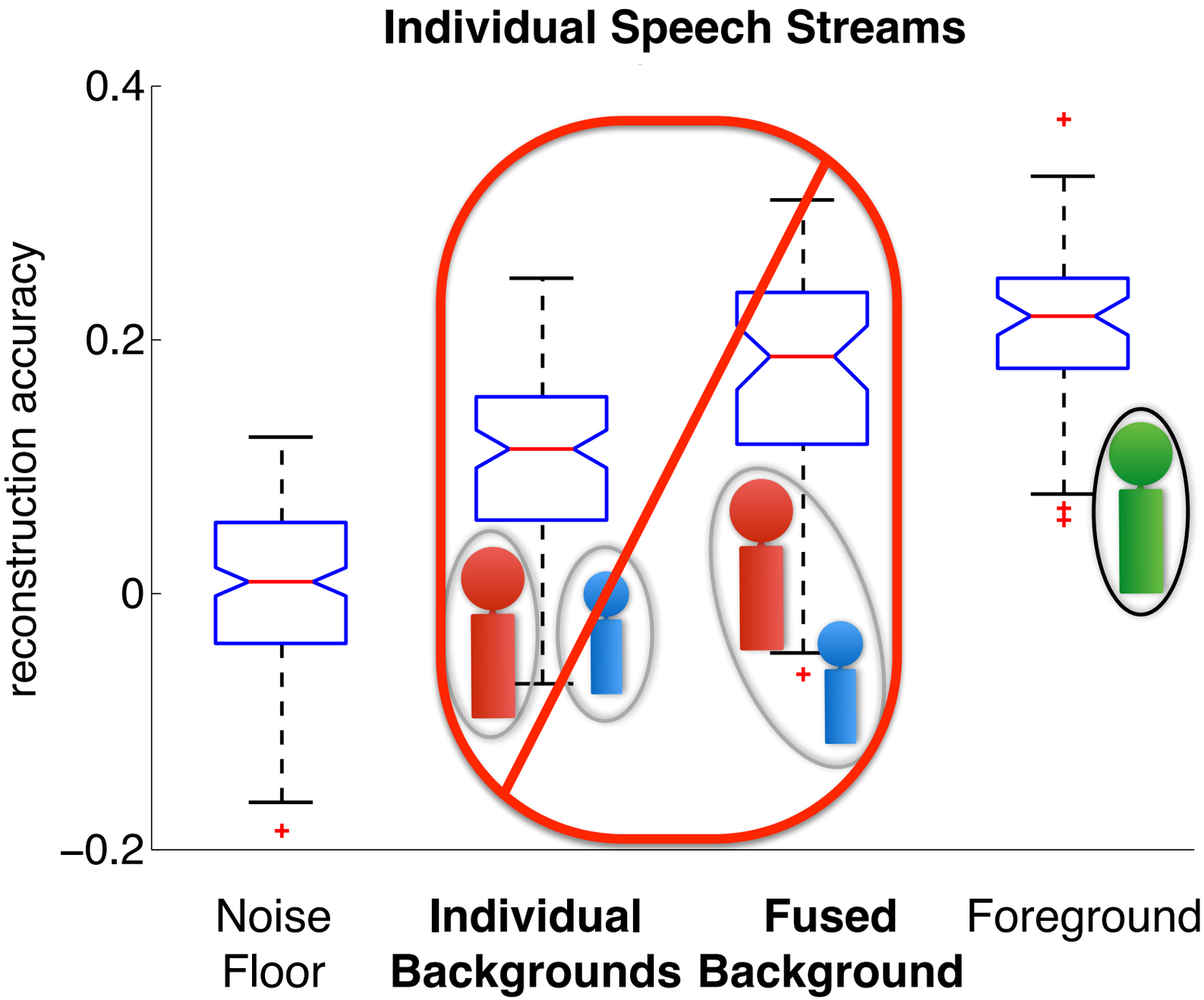
Integration Window over Late Times Only

Backgrounds vs. Background



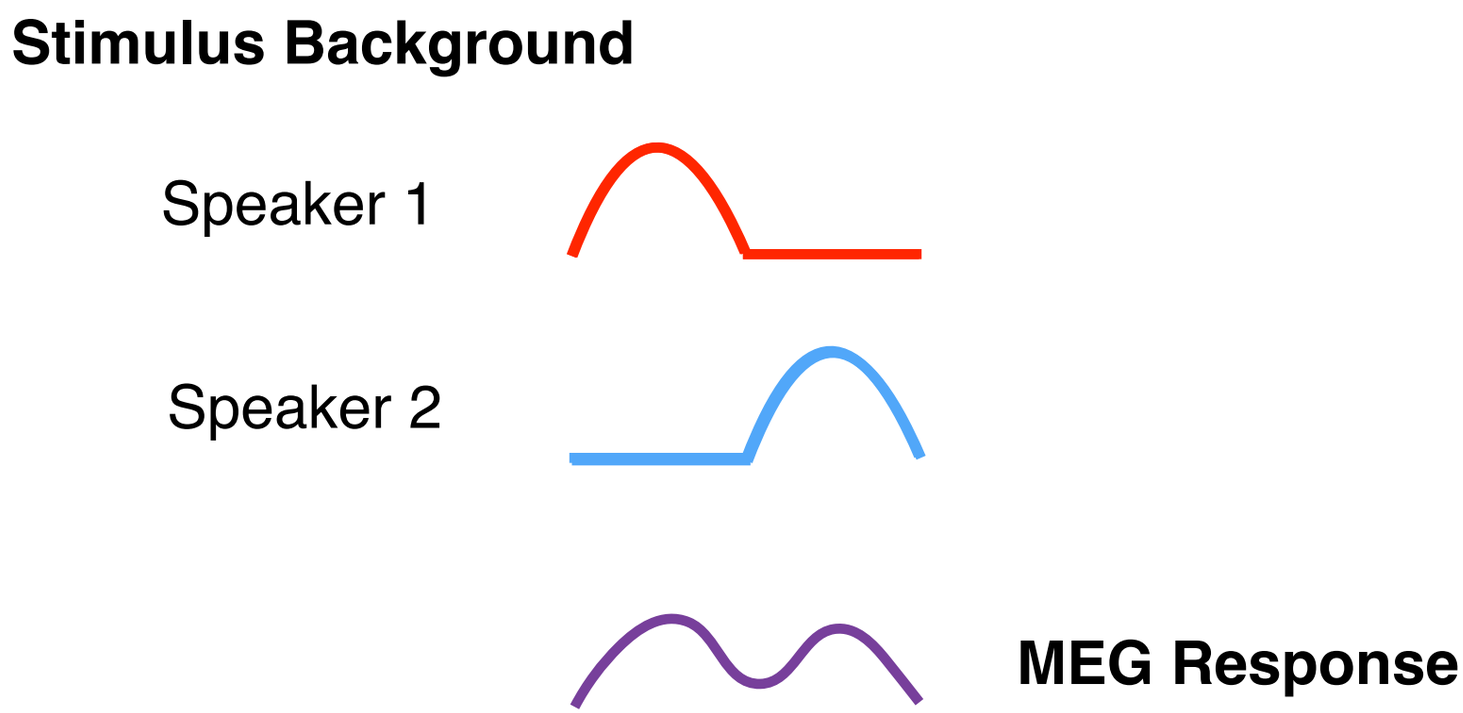
Integration Window over Late Times Only

Backgrounds vs. Background

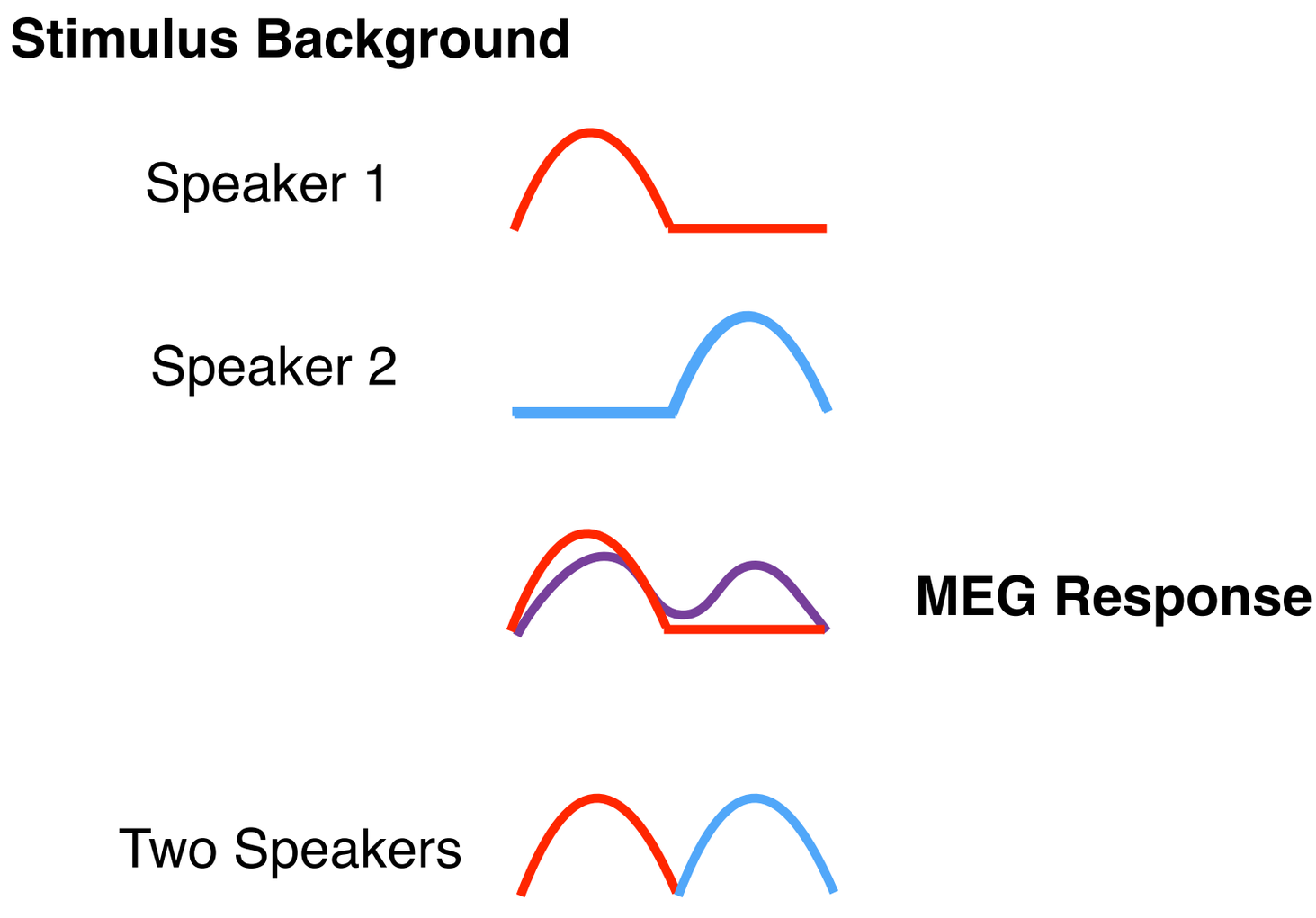


Integration Window over Late Times Only

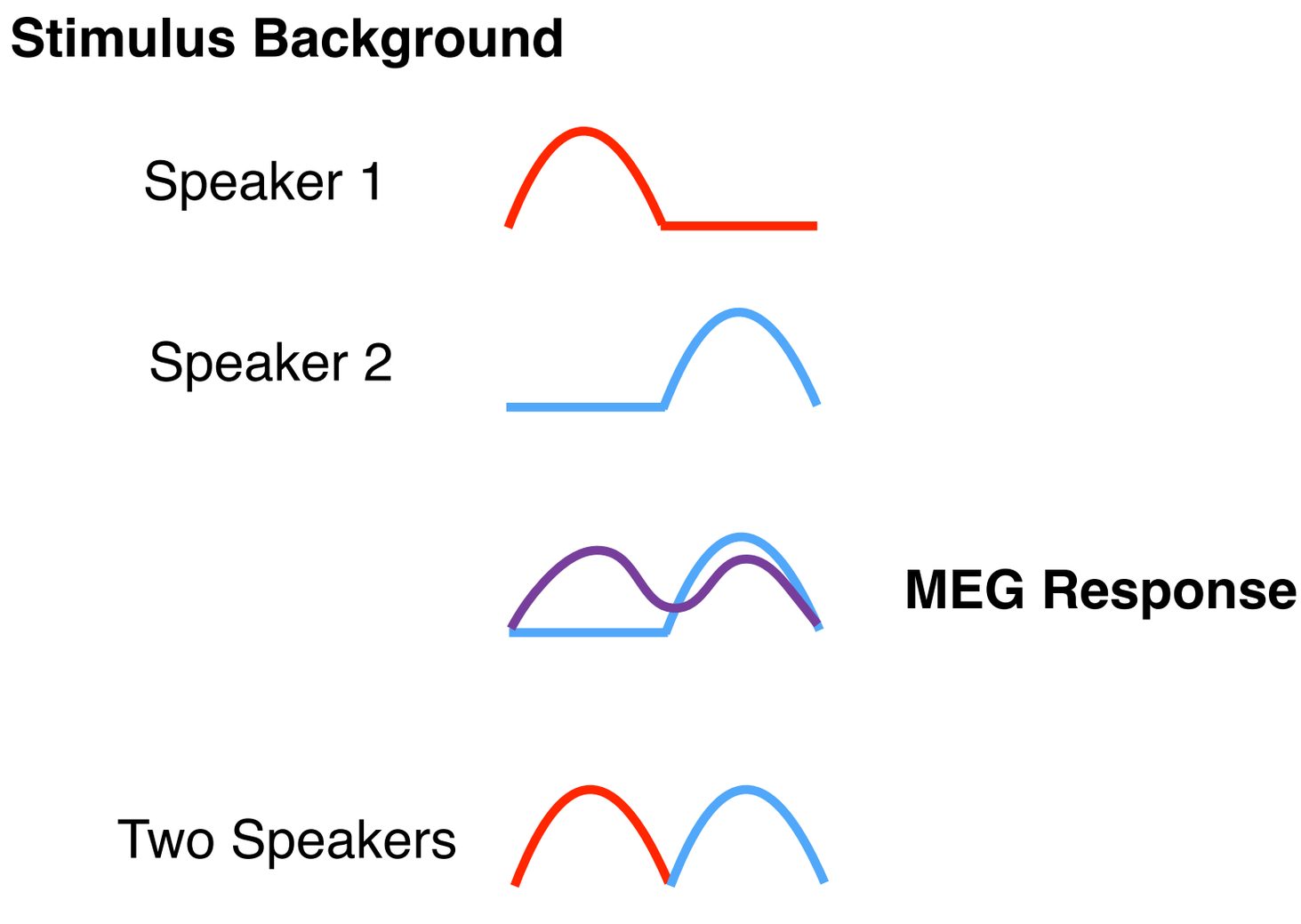
Backgrounds vs. Background
Why not?



Backgrounds vs. Background
Why not?



Backgrounds vs. Background
Why not?

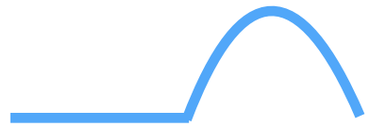


Backgrounds vs. Background
Why not?

Stimulus Background

Speaker 1

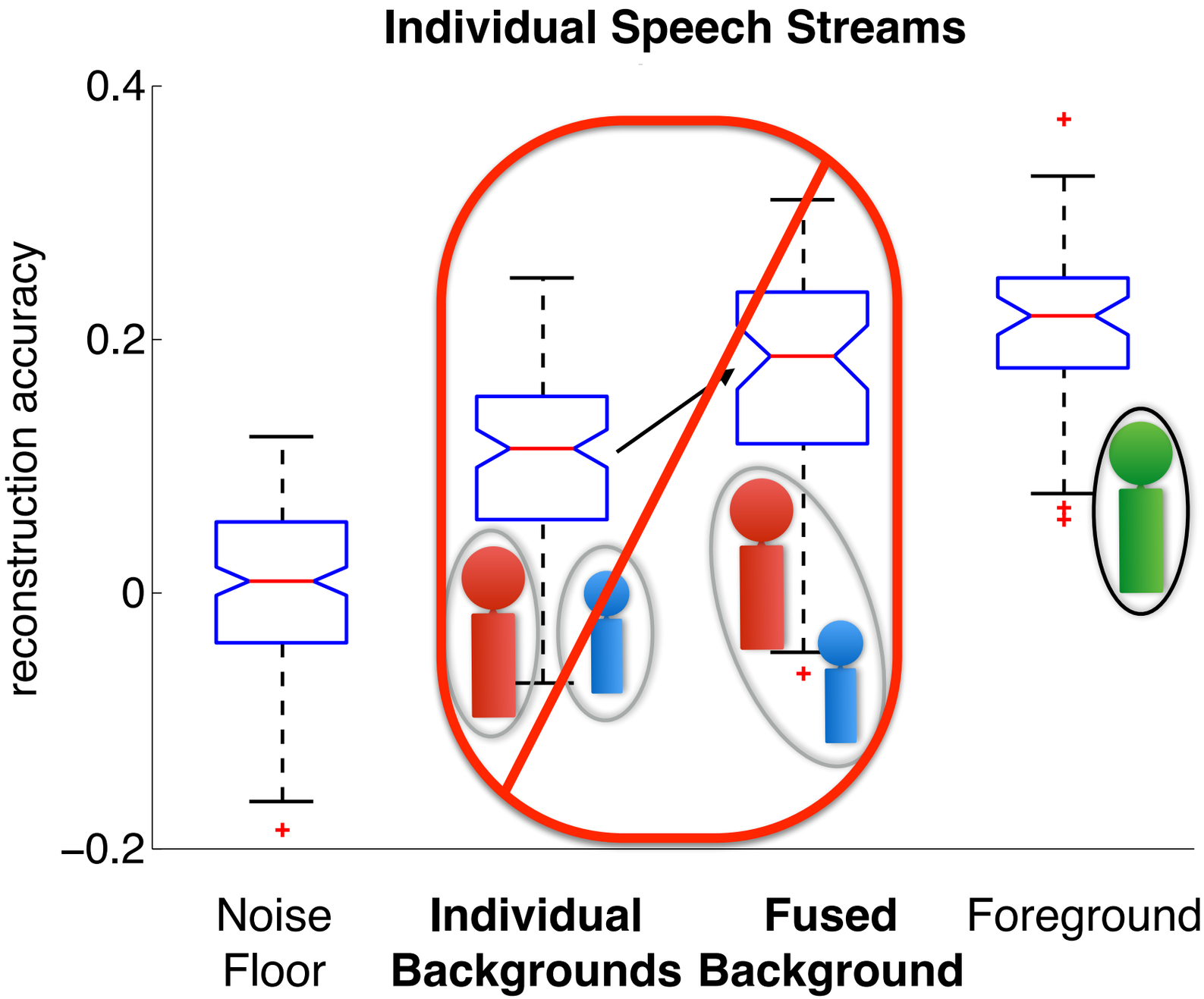
Speaker 2



MEG Response

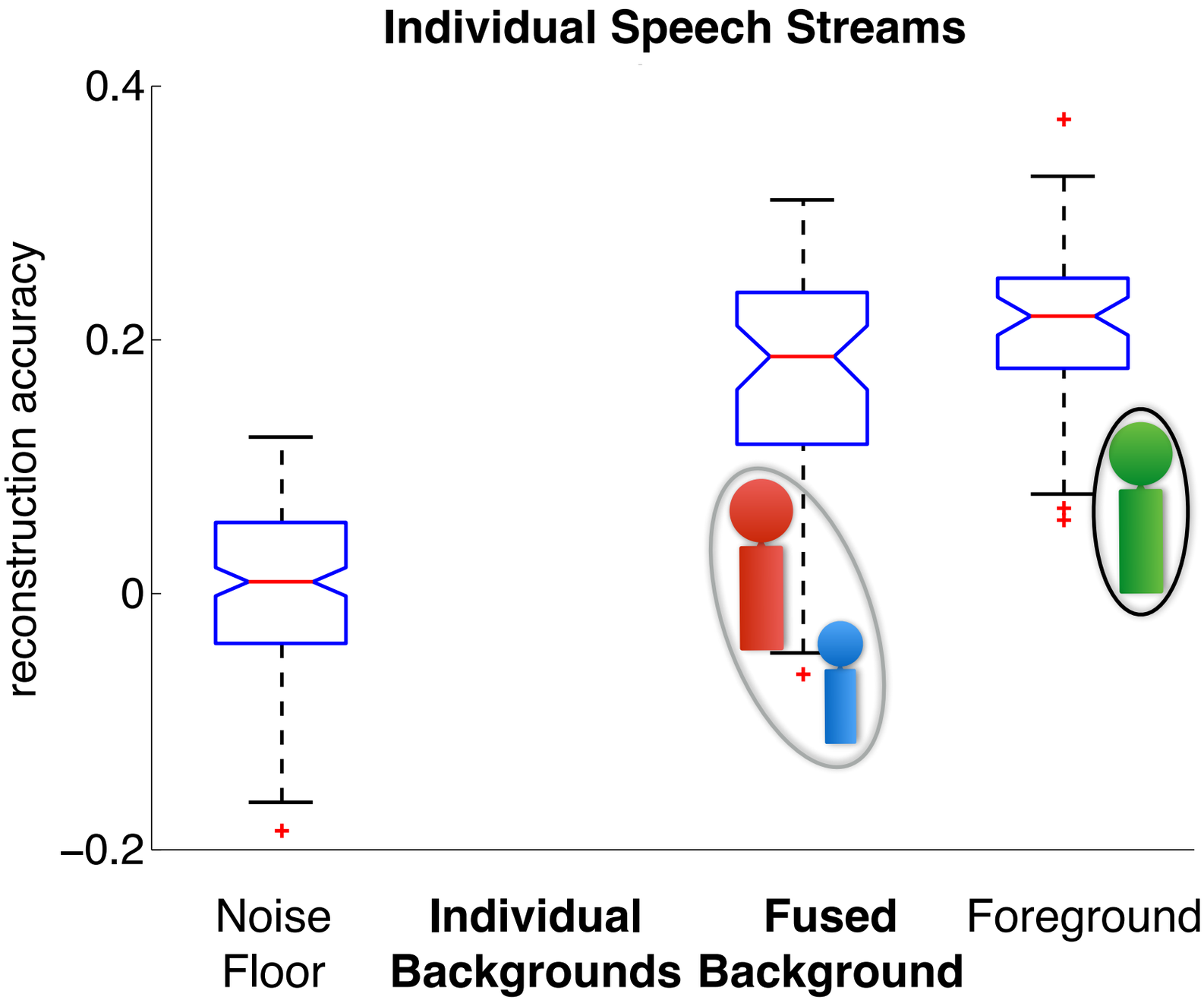
Two Speakers

Backgrounds vs. Background



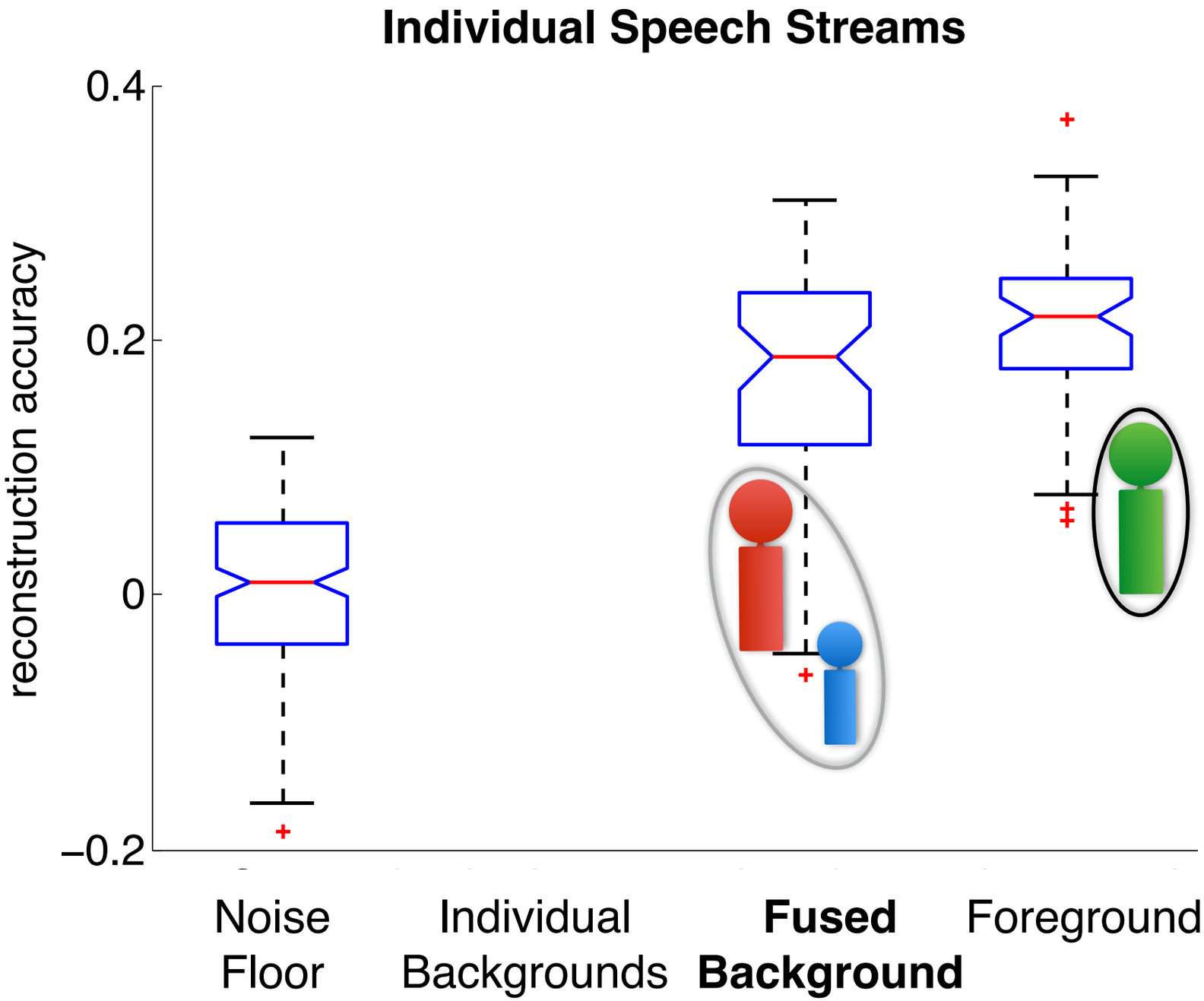
Integration Window over Late Times Only

Backgrounds vs. Background



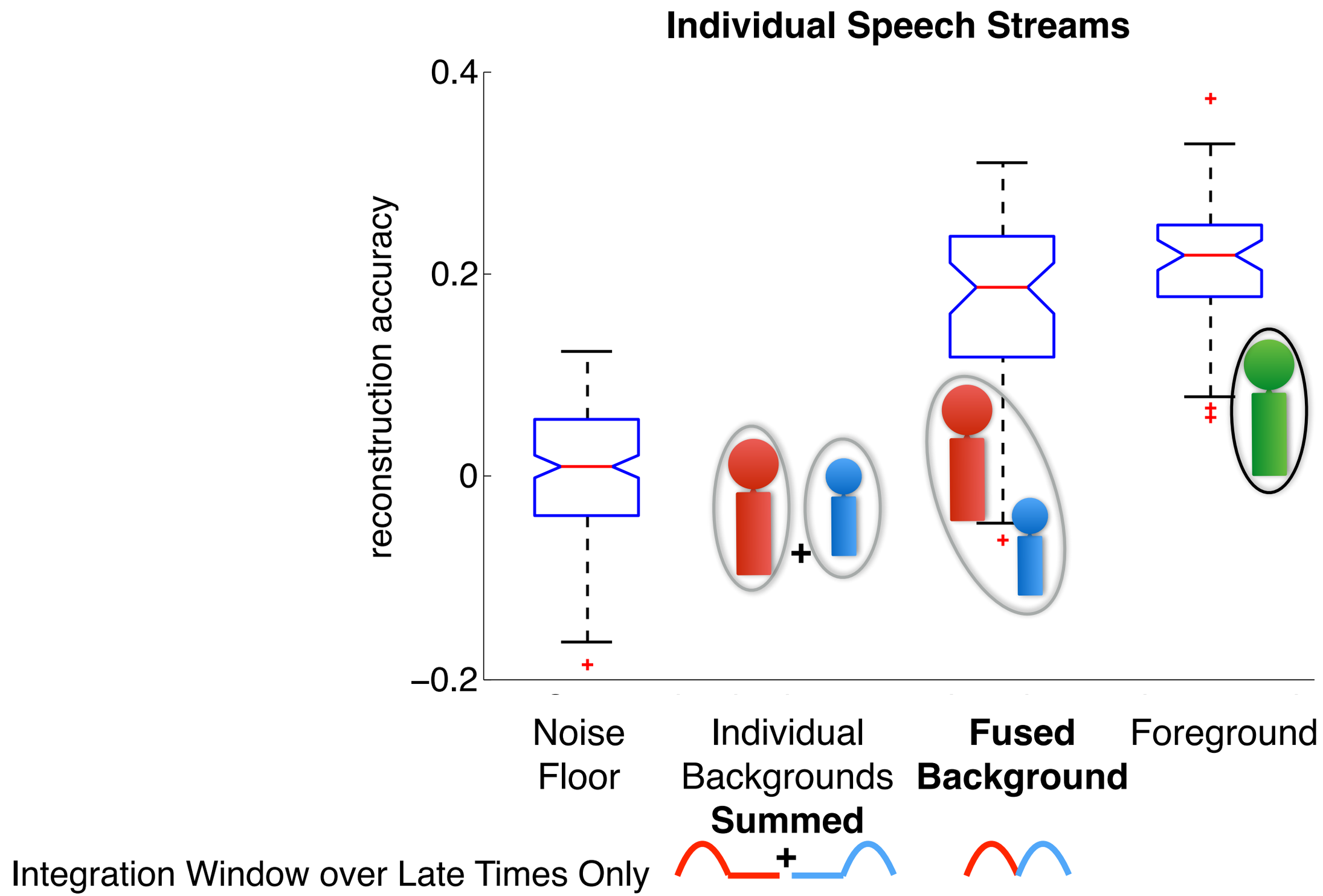
Integration Window over Late Times Only

Backgrounds vs. Background

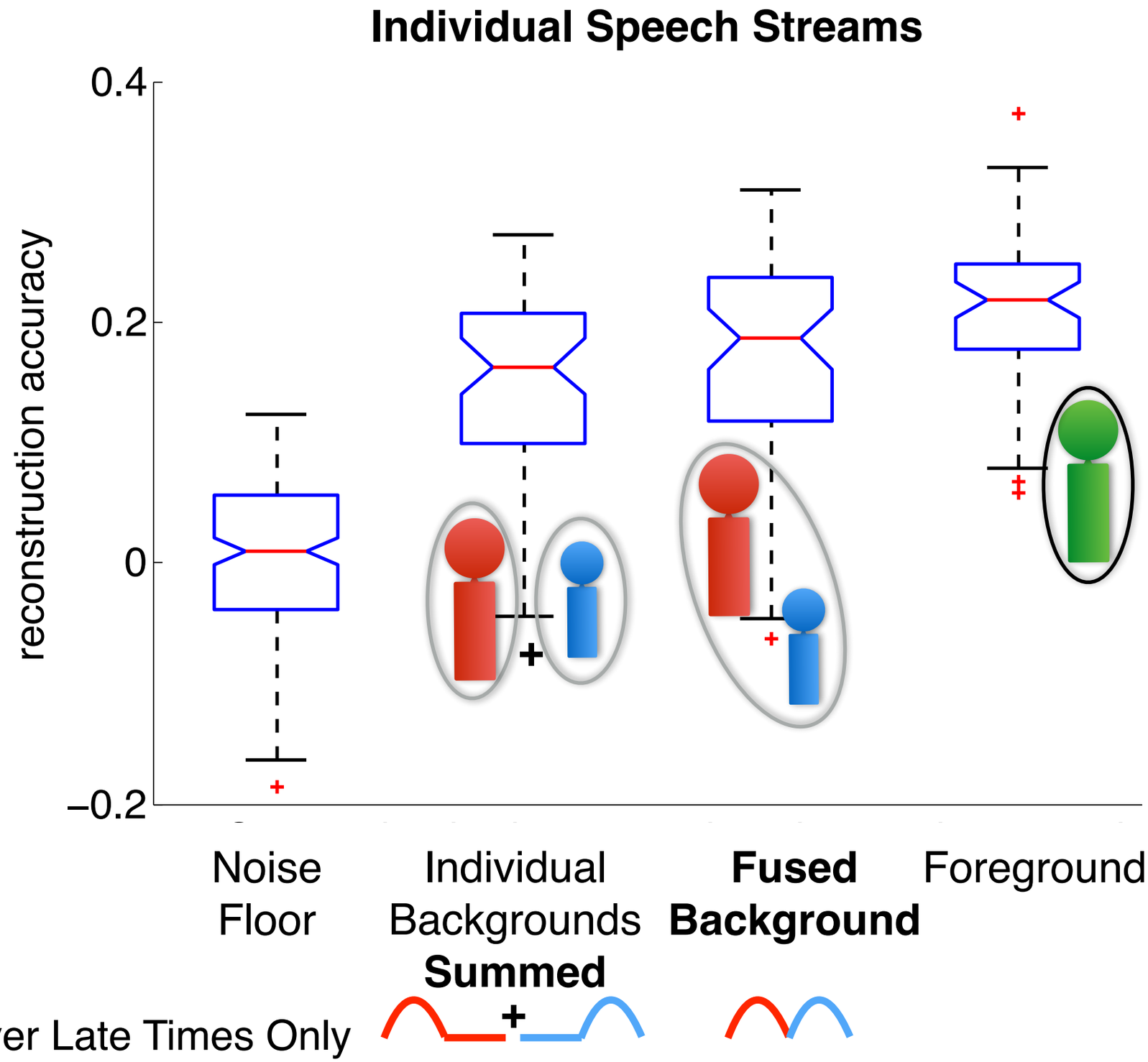


Integration Window over Late Times Only

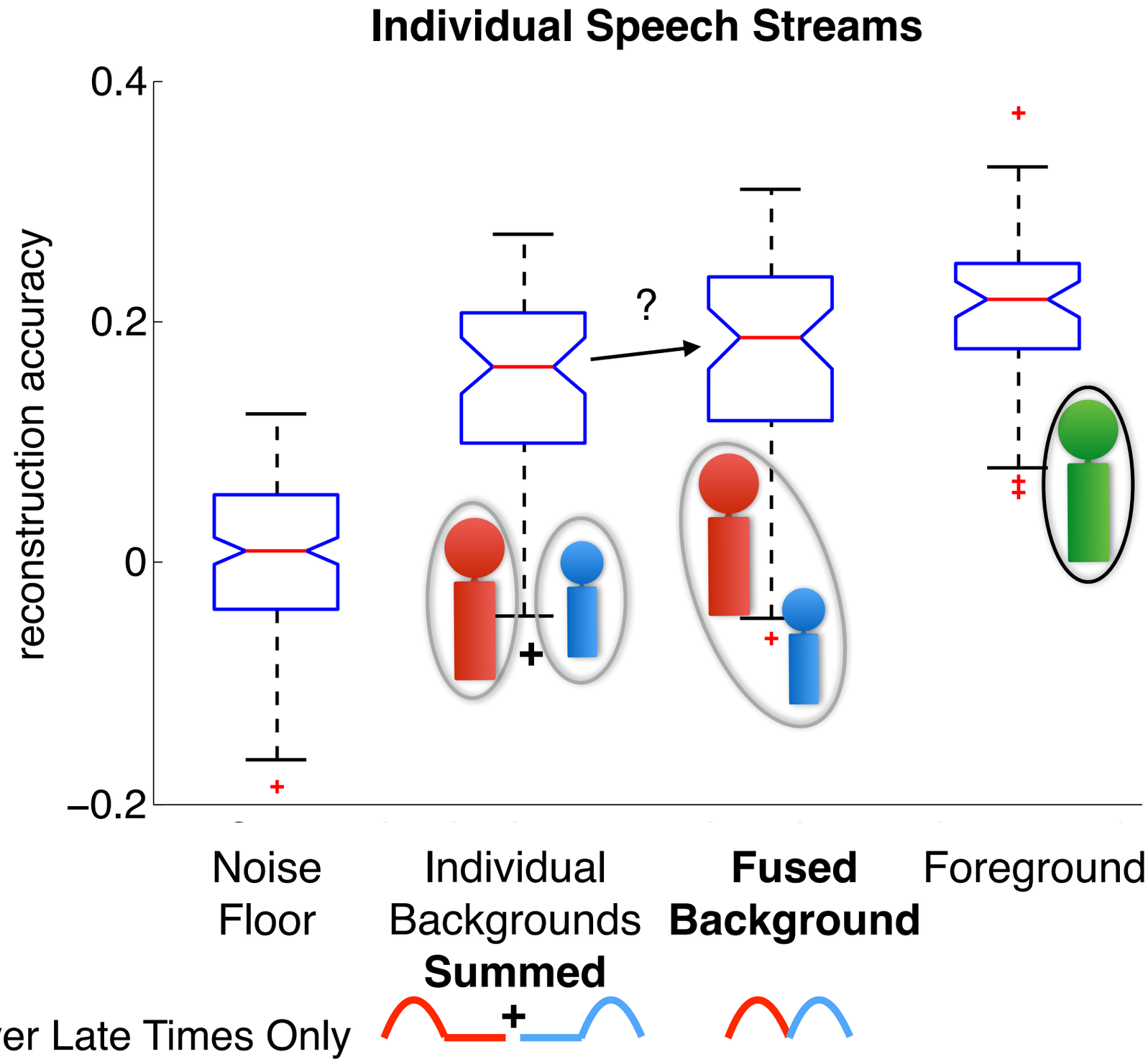
Backgrounds vs. Background



Backgrounds vs. Background

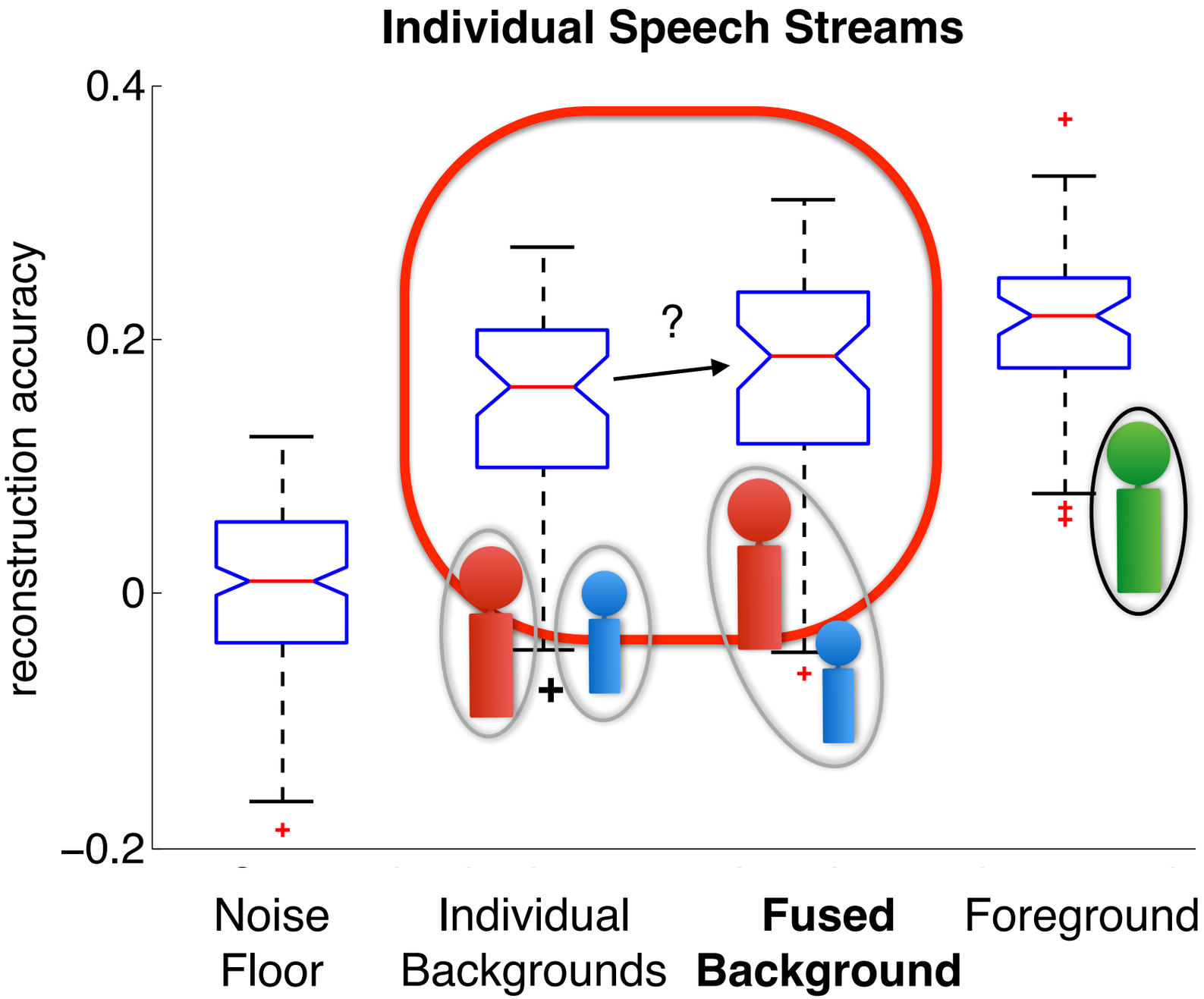


Backgrounds vs. Background



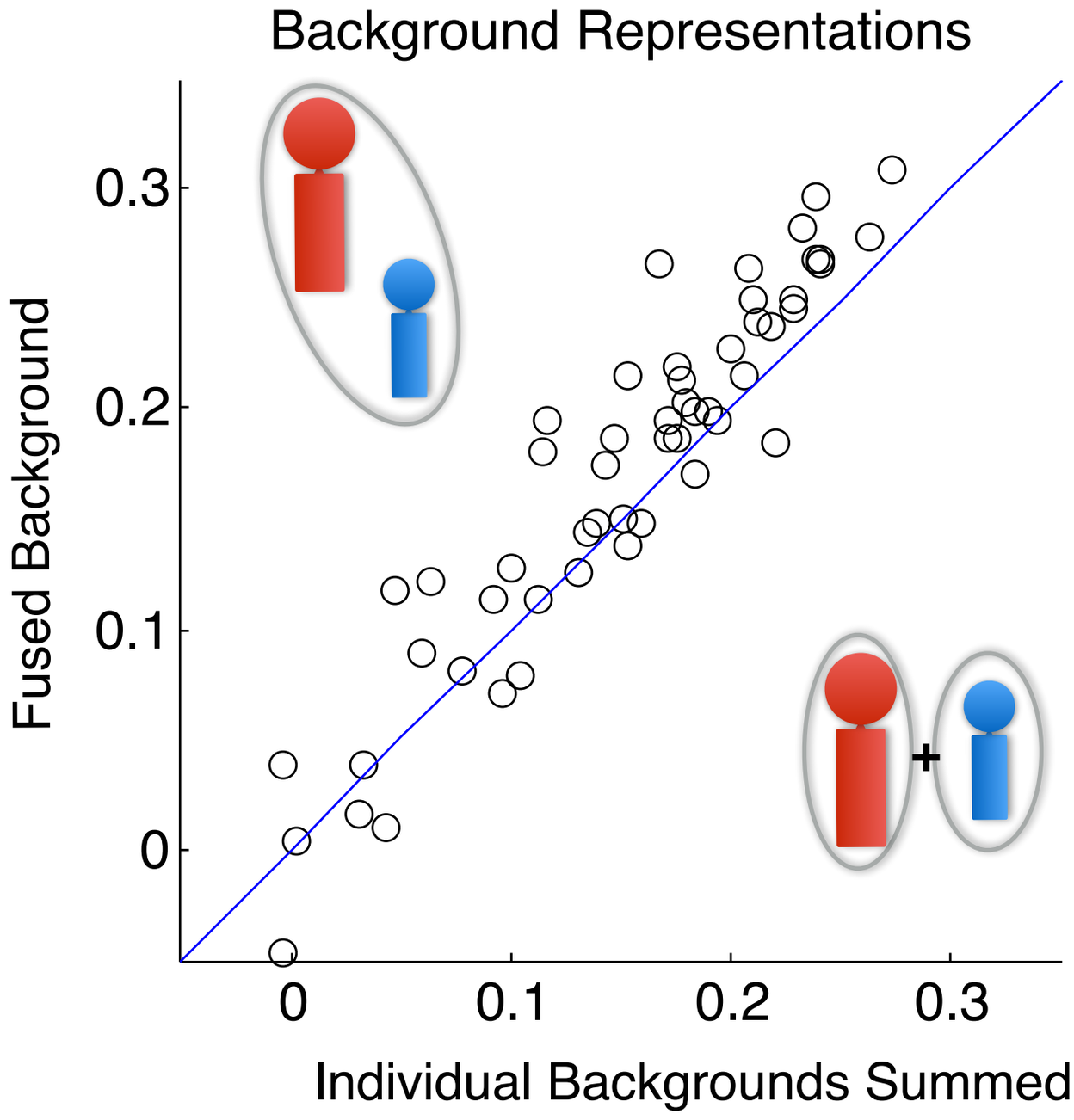
Integration Window over Late Times Only

Backgrounds vs. Background



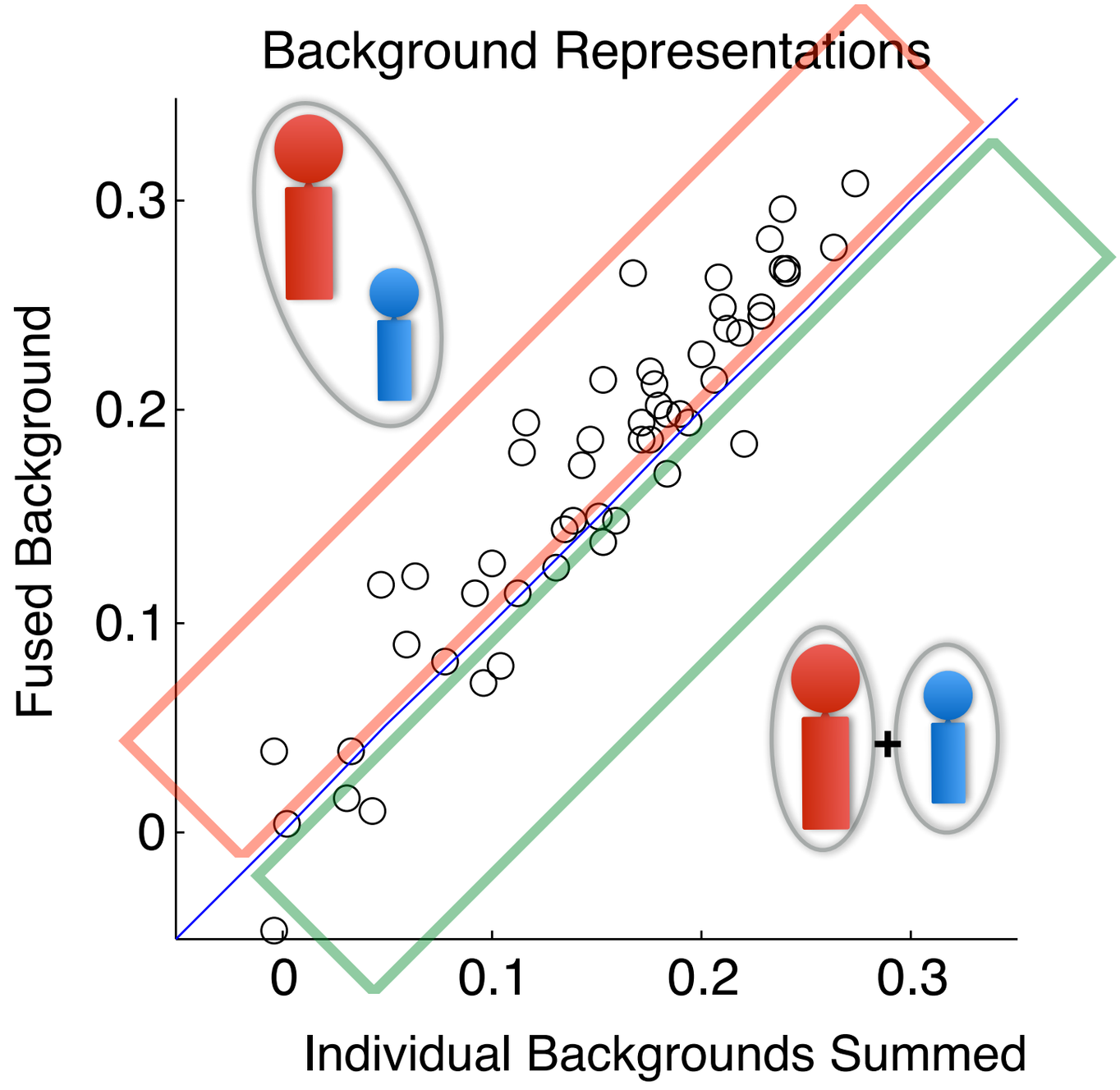
Integration Window over Late Times Only

Backgrounds vs. Background



Integration Window over Late Times Only

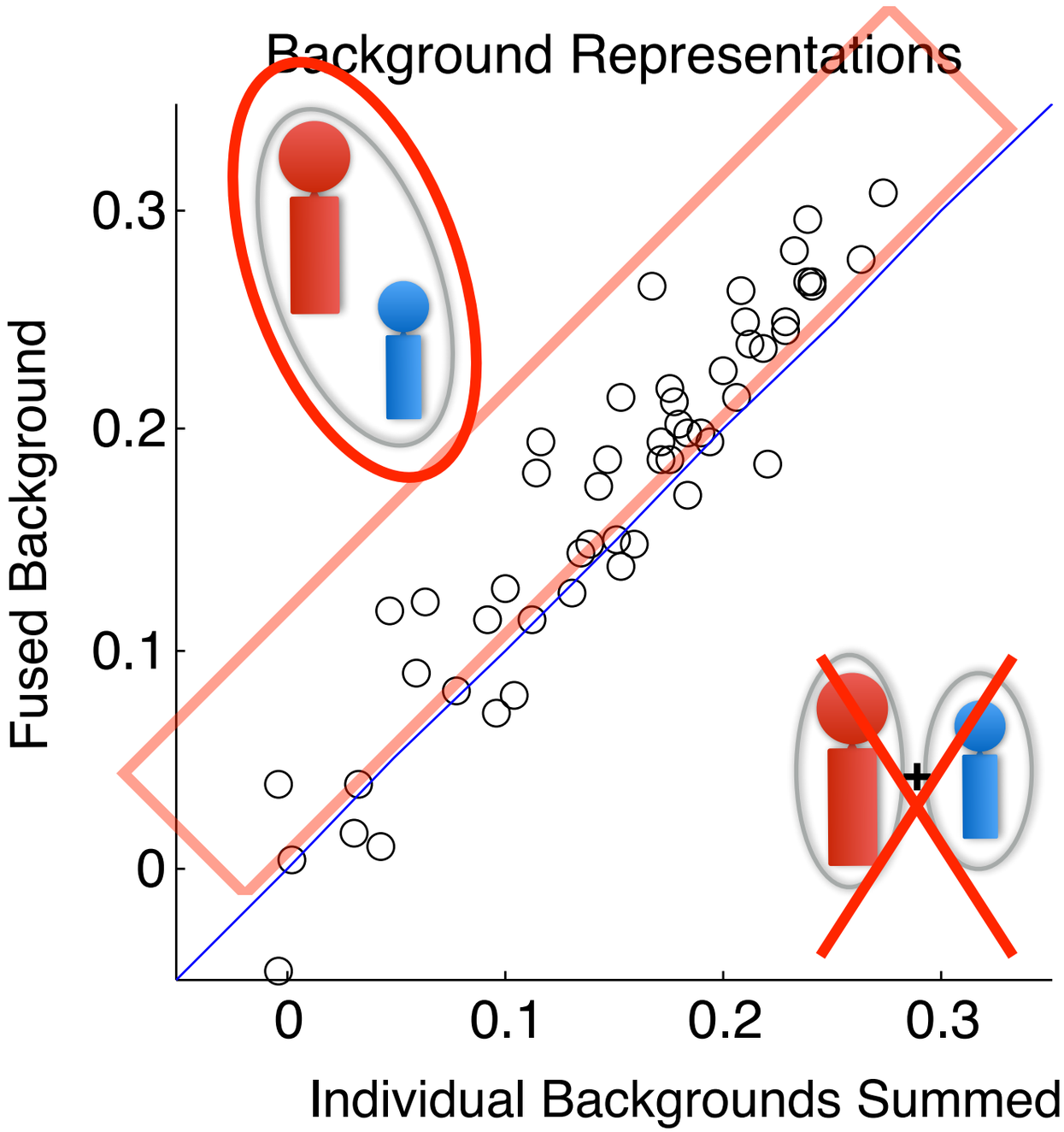
Backgrounds vs. Background



Integration Window over Late Times Only

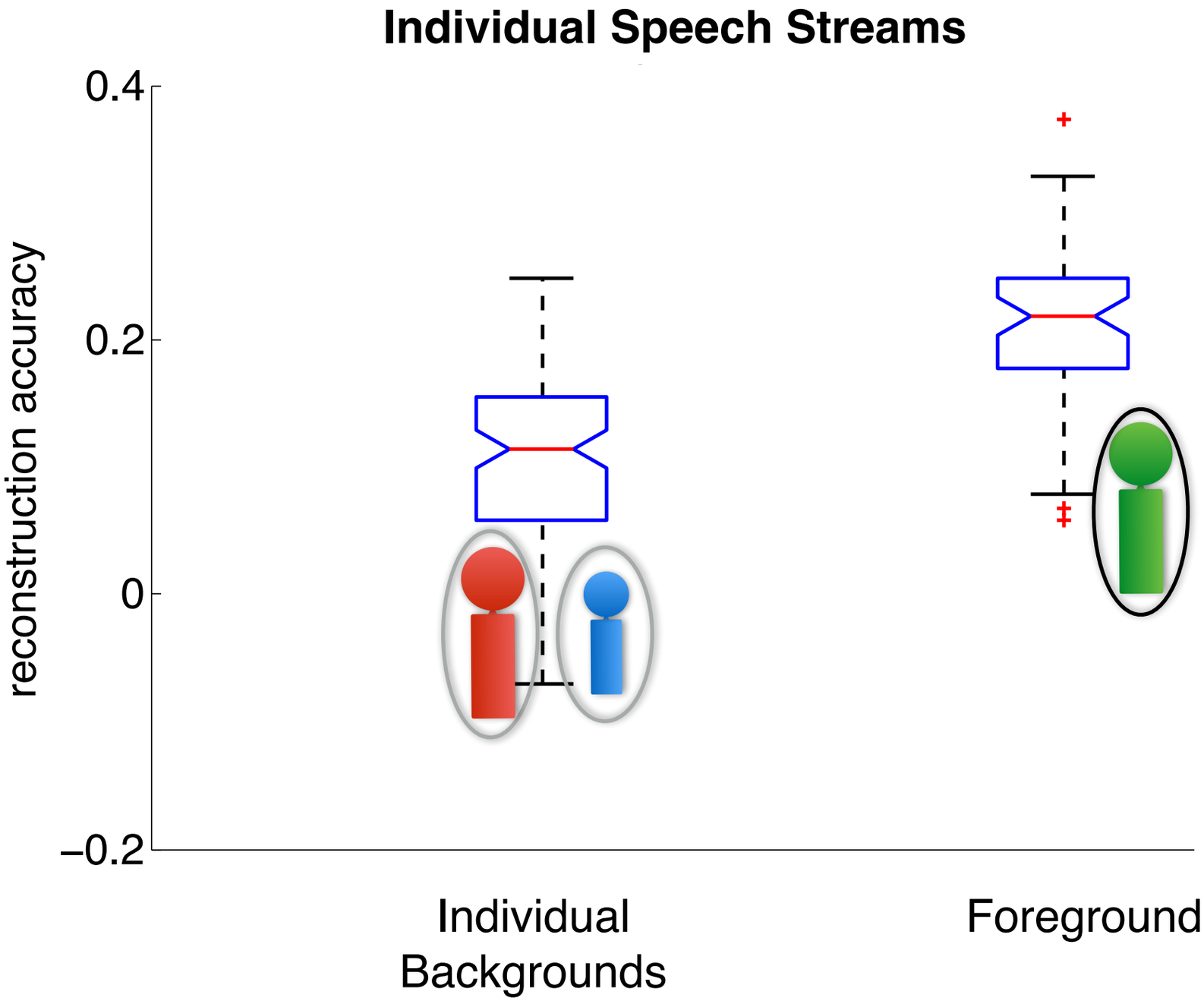
Backgrounds vs. Background

High latency areas (PT) represent ***fused*** background with better fidelity than ***individual*** backgrounds (p = 1.3E-05)



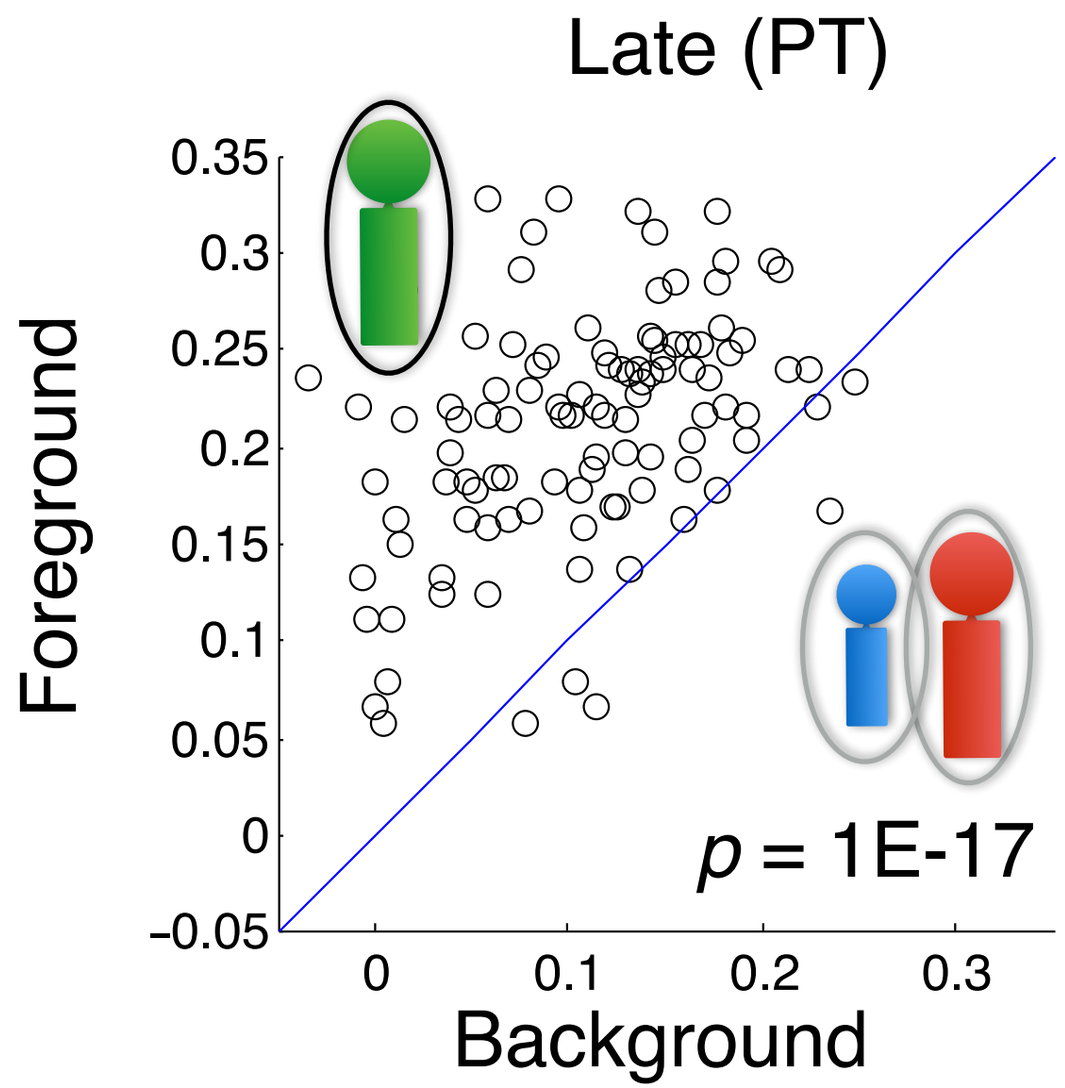
Integration Window over Late Times Only

Foreground vs. Background

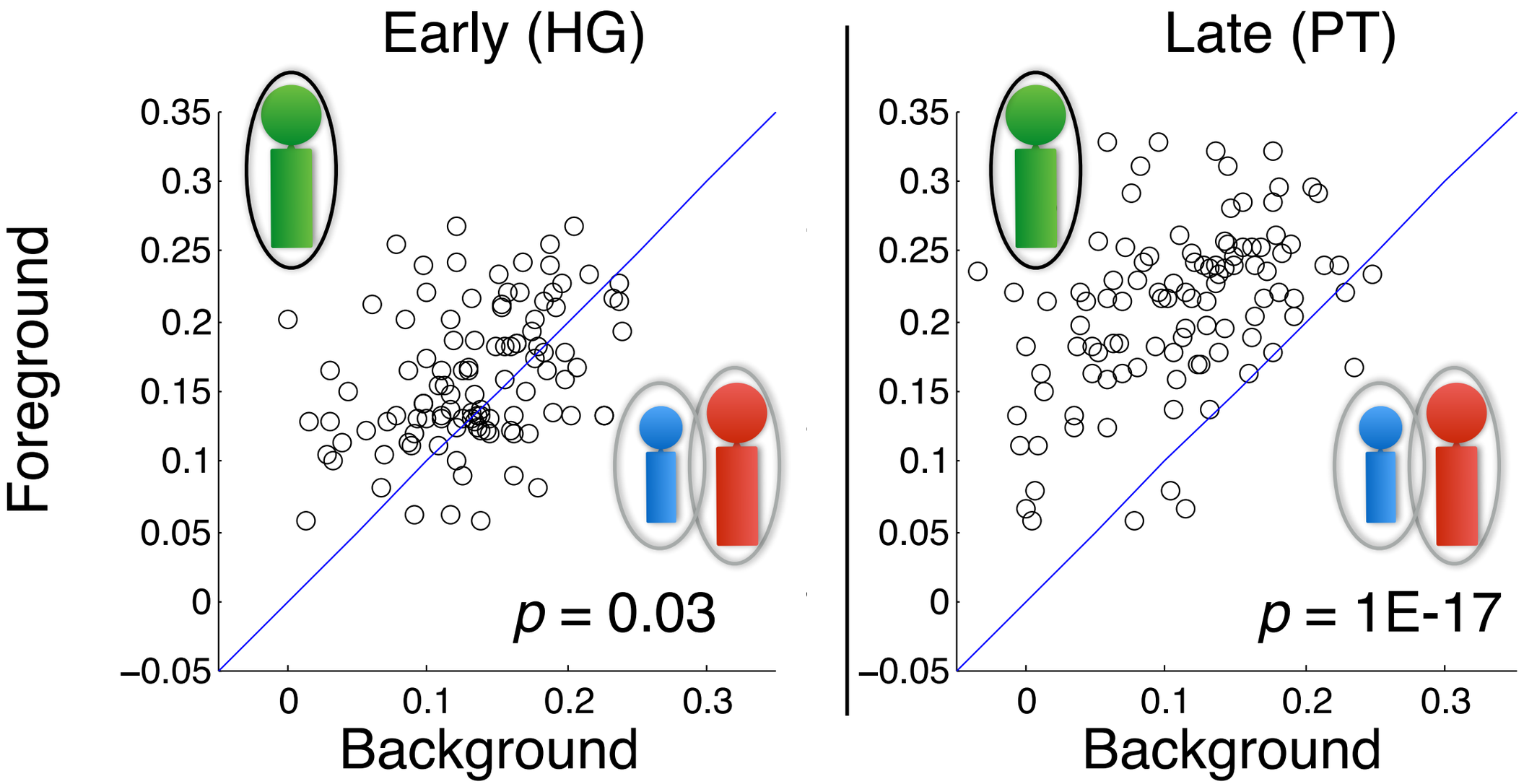


Integration Window over Late Times Only

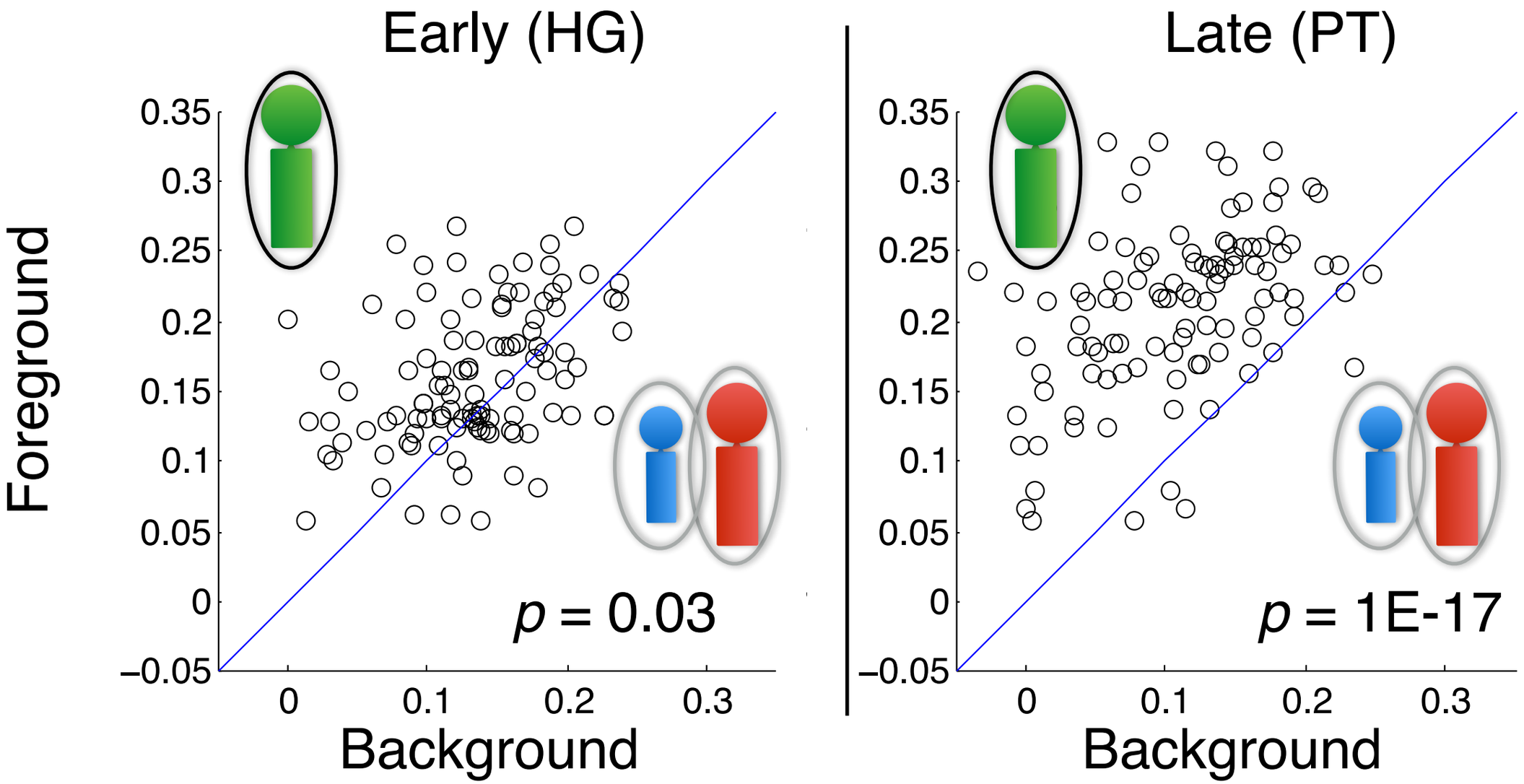
Foreground vs. Background
Early vs. Late



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)
- Object representation at 100 ms latency (PT), but not by 50 ms (HG)
- Consistent with being neural representations of auditory perceptual object
- Preliminary evidence for
  - ✓ PT: additional fused background representation
  - ✓ HG: almost equal representations

Thank You