

Neural Representations of Speech at 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

Current (Simon Lab & Affiliates)

Christian Brodbeck
Francisco Cervantes
David Nahmias
Mahshid Najafi
Krishna Puvvada
Peng Zan

Past (Simon Lab & Affiliate Labs)

Nayef Ahmar
Murat Aytekin
Claudia Bonin
Maria Chait
Marisel Villafane Delgado
Kim Drnec
Nai Ding
Victor Grau-Serrat
Julian Jenkins
Natalia Lapinskaya
Kai Sum Li
Huan Luo
Ling Ma
Alex Presacco

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
Stefanie Kuchinsky
Cindy Moss
David Poeppel
Shihab Shamma

Past Postdocs & Visitors

Sahar Akram

Aline Gesualdi Manhães
Dan Hertz
Jonas Vanthornhout
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
Andrea Shome
Sandra Soltz
Madeleine Varmer
James Williams

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

Outline

- Cortical Representations of Speech (via MEG)
 - ▶ Encoding vs. Decoding
- Cortical Representations of the “Cocktail Party”
- Recent Results
 - ▶ Attentional Dynamics
 - ▶ Aging & Cortical Representations of Speech
 - ▶ Higher Level Interference & Noise

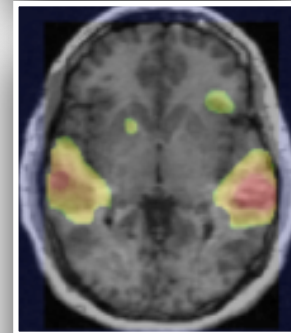
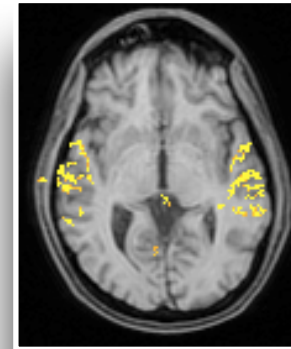
Functional Brain Imaging

Functional Brain Imaging
= Non-invasive recording from human brain

Hemodynamic techniques

fMRI
functional magnetic resonance imaging

PET
positron emission tomography



Excellent
Spatial
Resolution
(~1 mm)

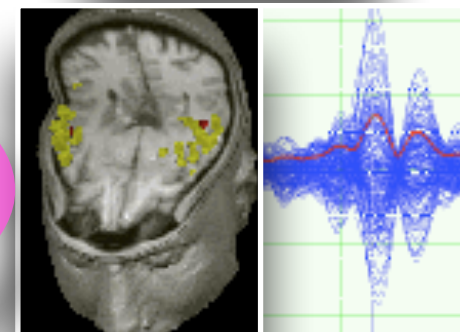
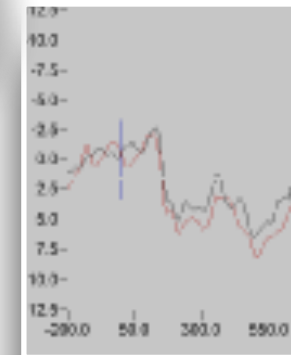
Poor
Temporal
Resolution
(~1 s)

fMRI & MEG can capture effects in single subjects

Electromagnetic techniques

EEG
electroencephalography

MEG
magnetoencephalography

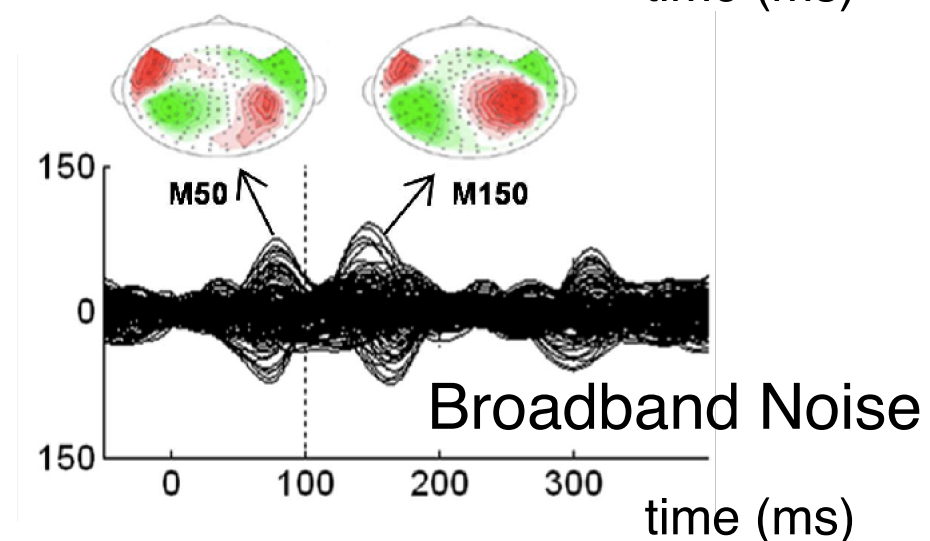
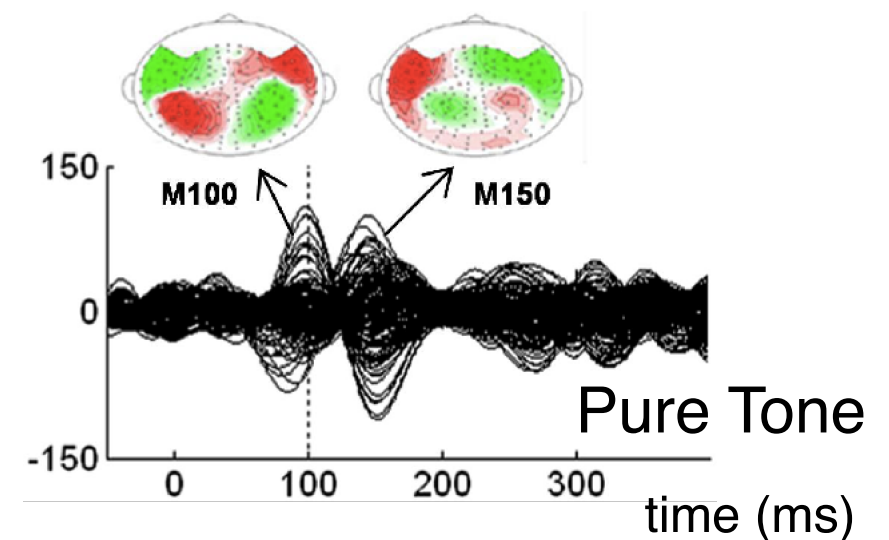
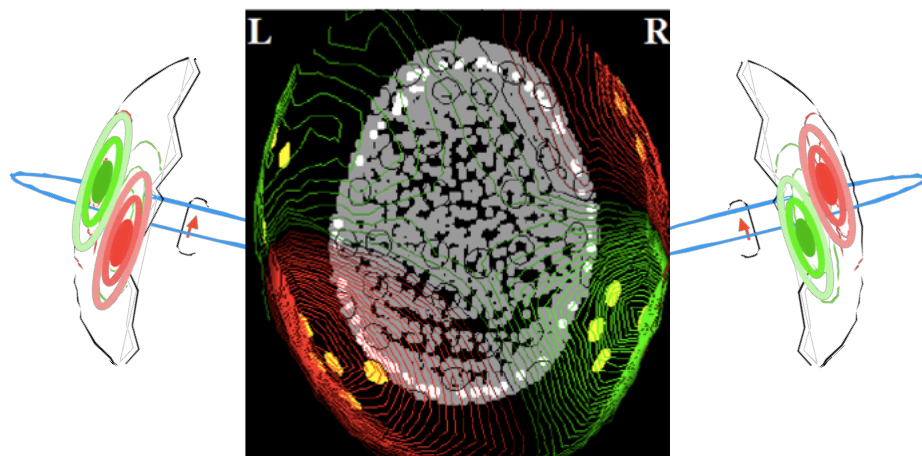


Poor
Spatial
Resolution
(~1 cm)

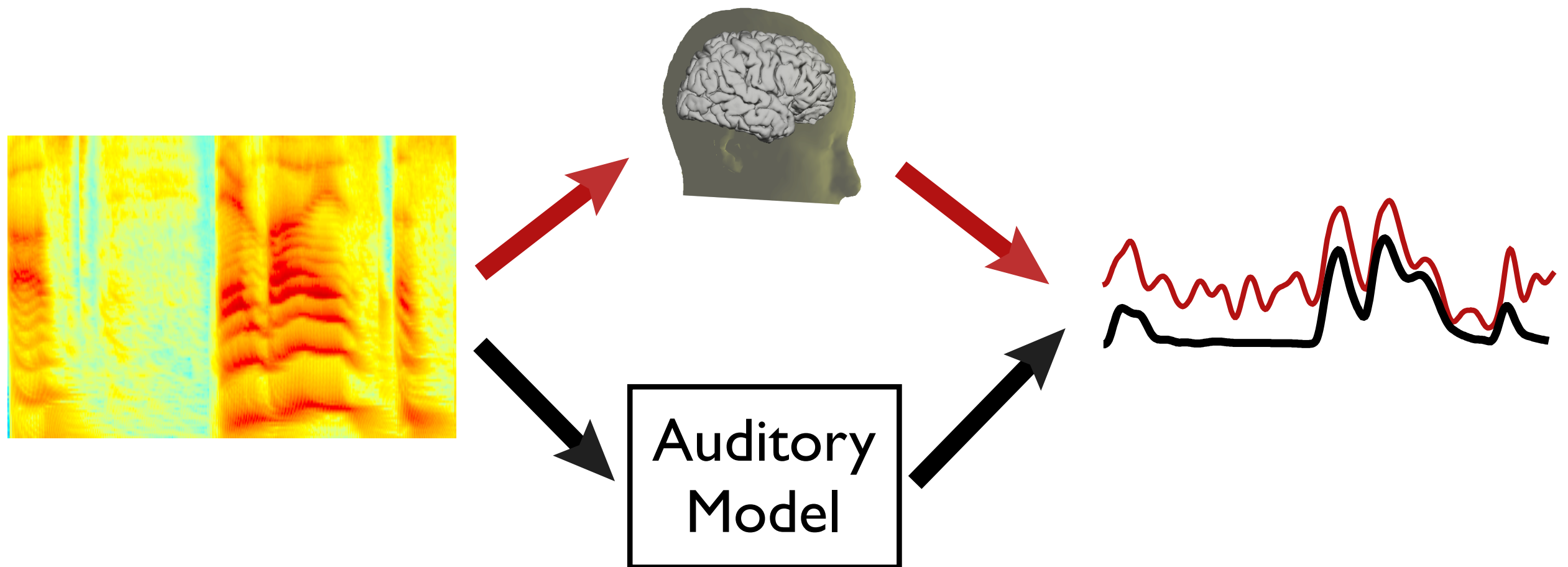
Excellent
Temporal
Resolution
(~1 ms)

MEG & Auditory Cortex

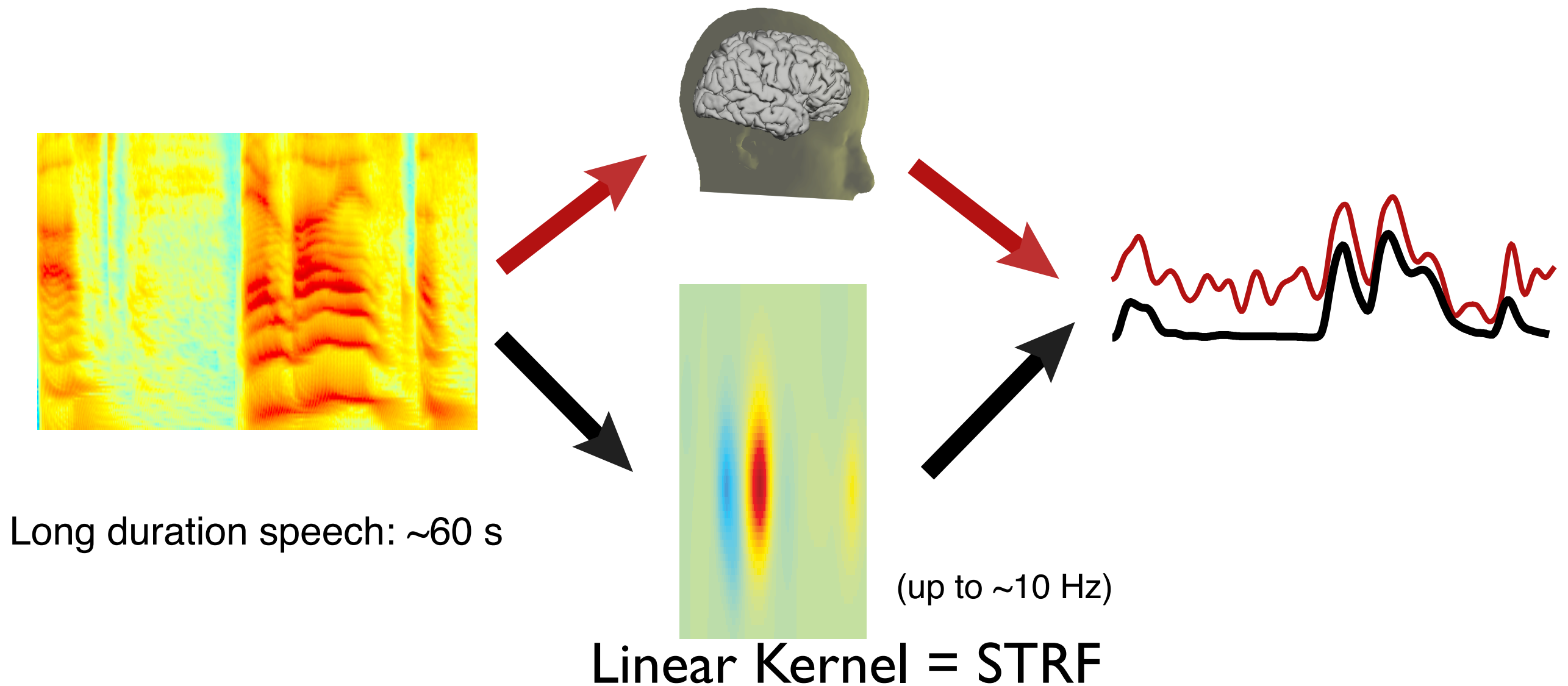
- Non-invasive, Passive, Silent Neural Recordings
- MEG Response Patterns Time-Locked to Stimulus Events
- Robust
- Strongly Lateralized
- Cortical Origin Only



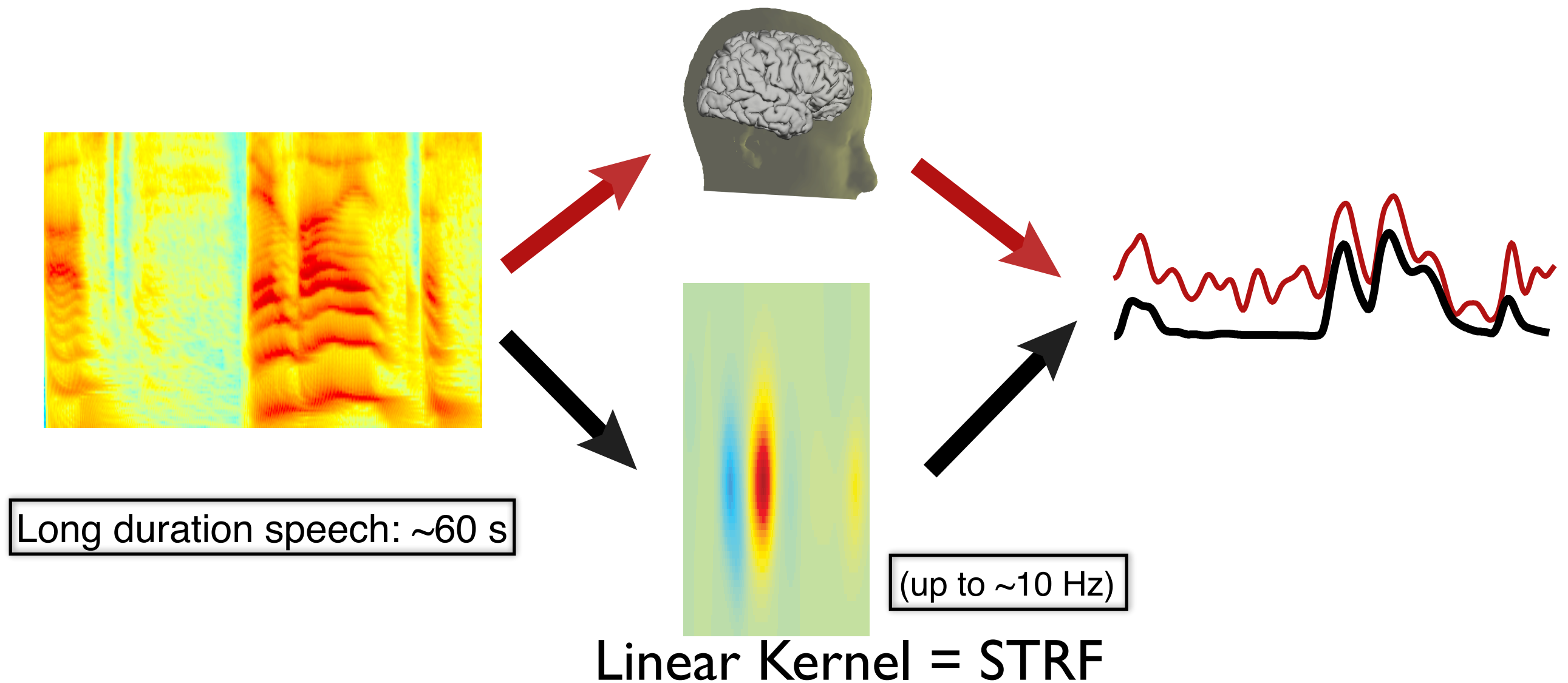
MEG Responses to Speech Modulations



MEG Responses Predicted by STRF Model

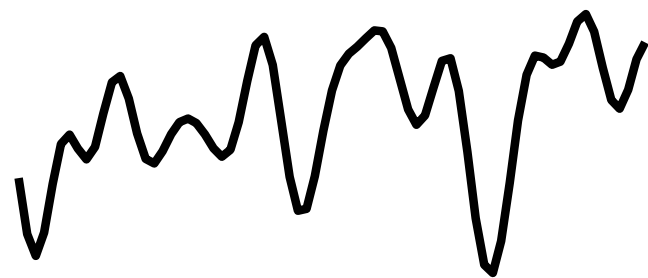


MEG Responses Predicted by STRF Model

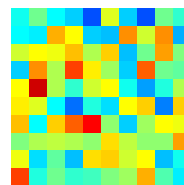


Neural Reconstruction of Speech Envelope

Speech Envelope

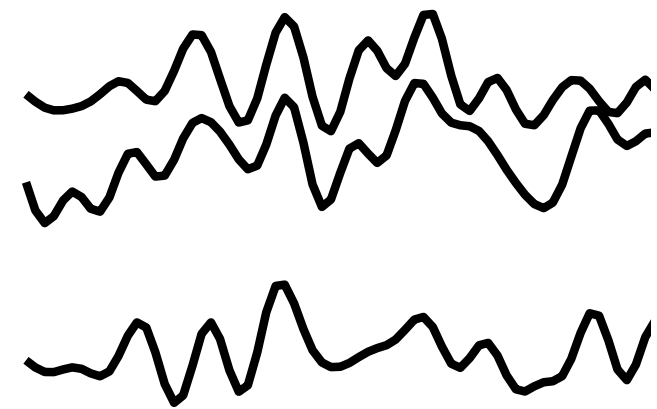


Decoder

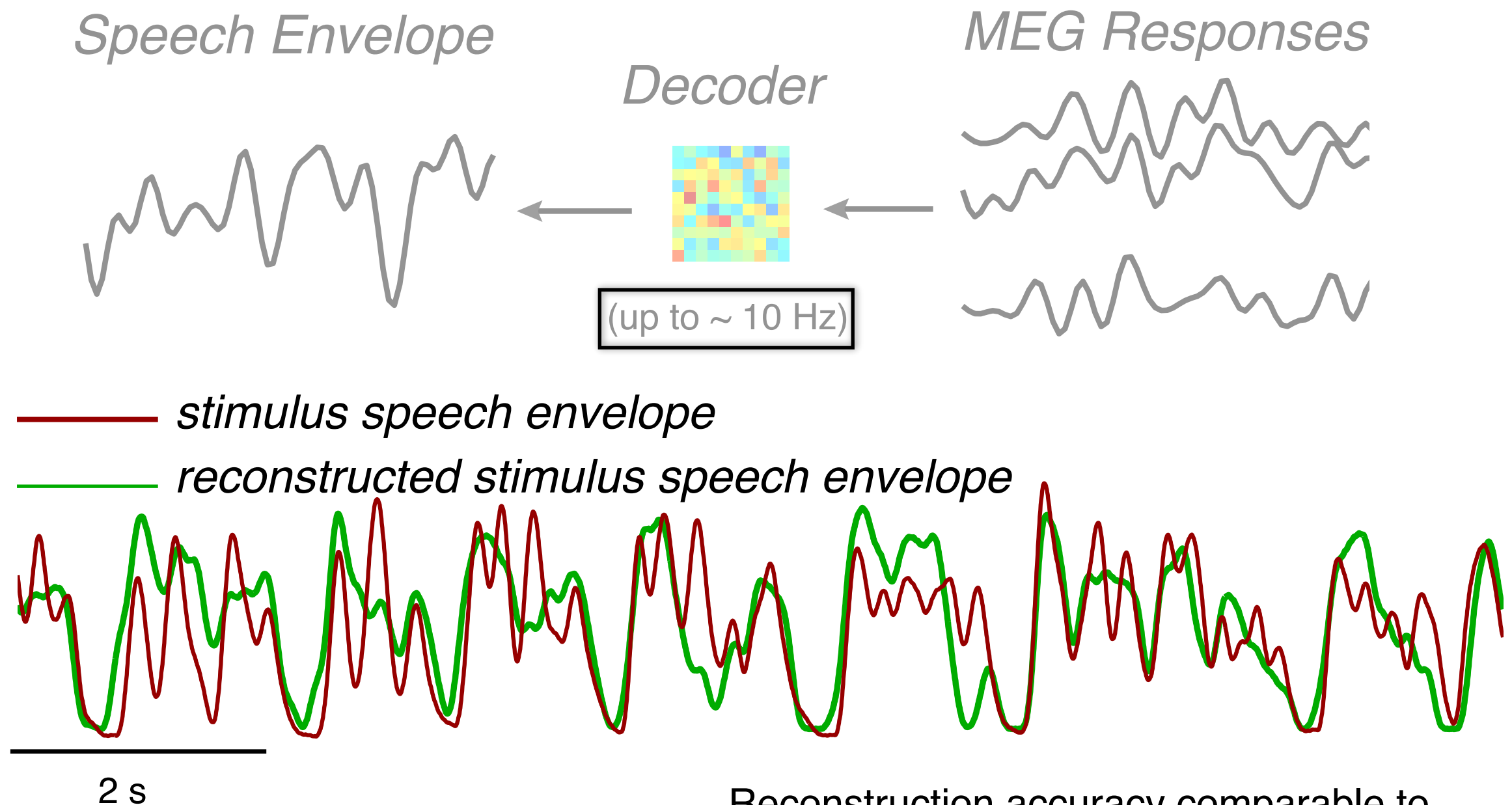


(up to ~ 10 Hz)

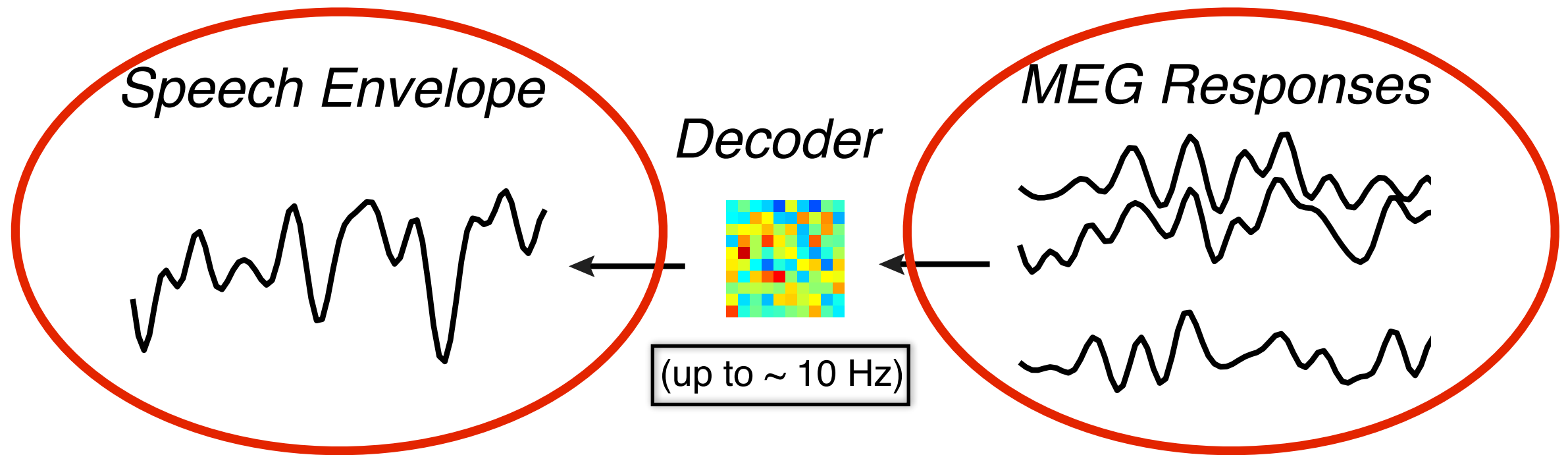
MEG Responses



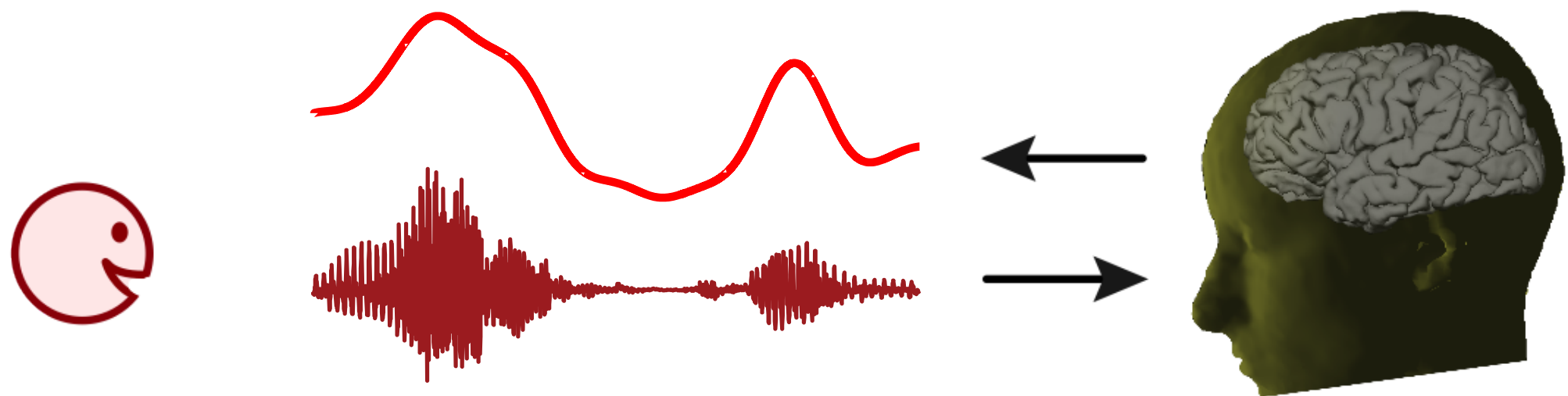
Neural Reconstruction of Speech Envelope



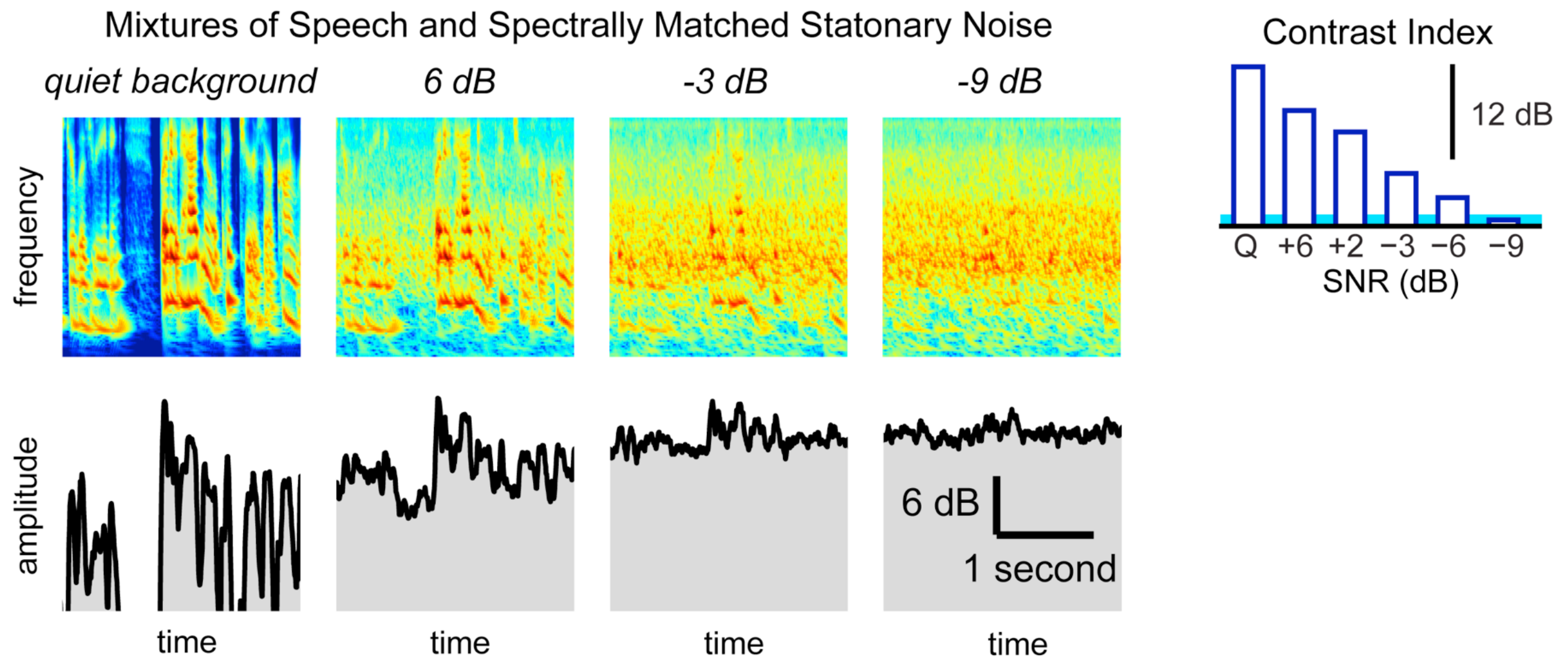
Reconstruction accuracy comparable to
single unit & ECoG recordings



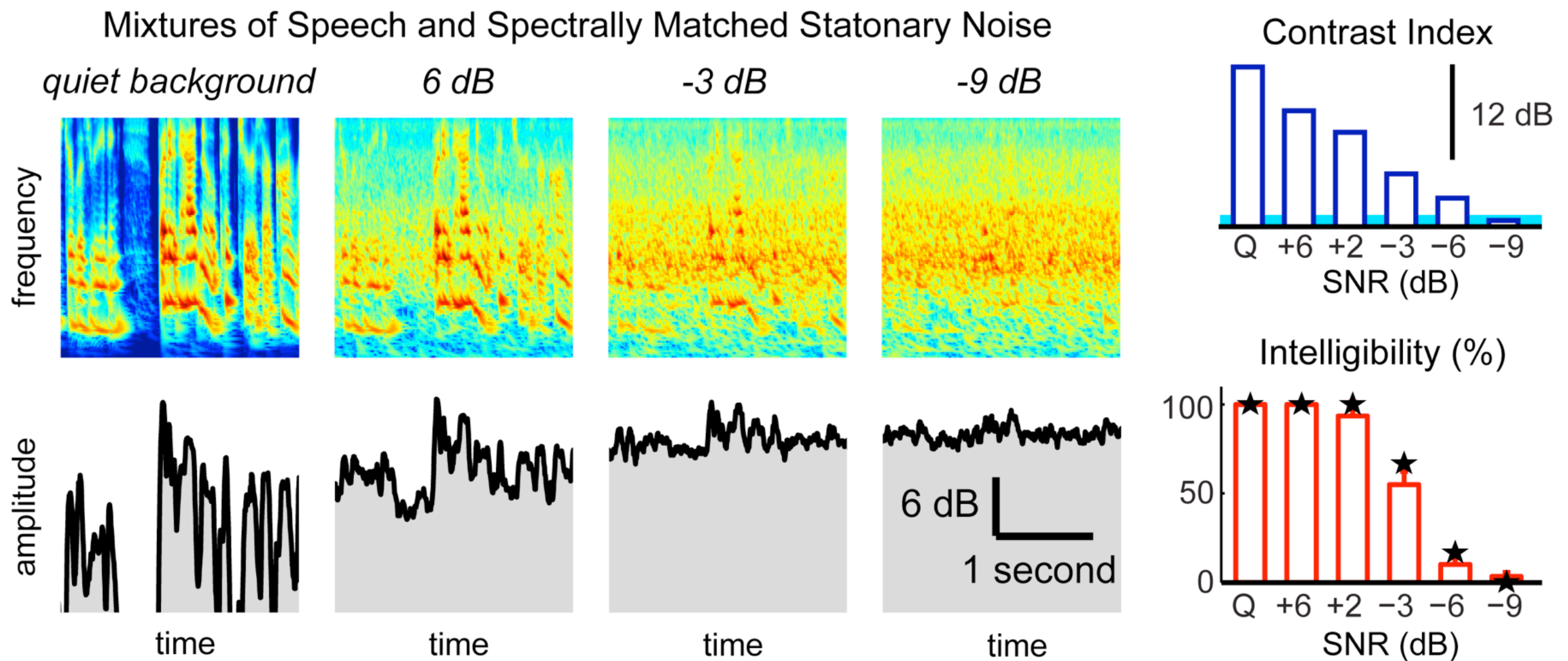
Neural Representation of Speech: Temporal



Speech in Stationary Noise

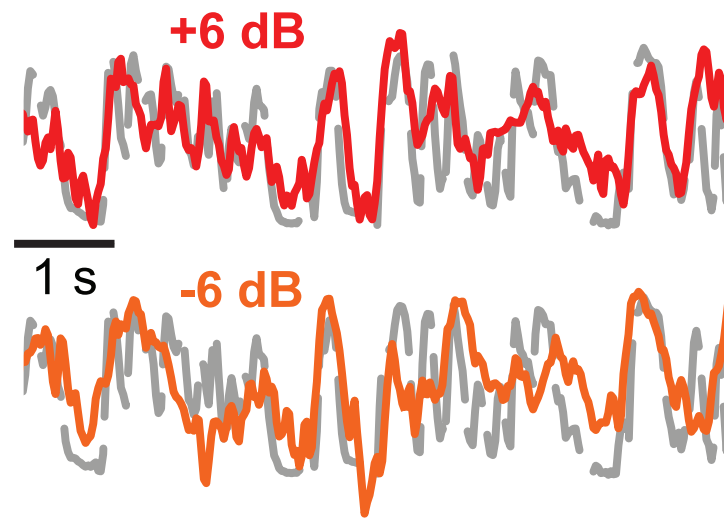


Speech in Stationary Noise



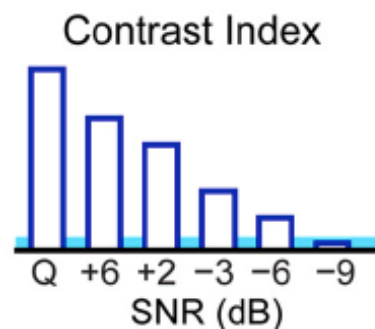
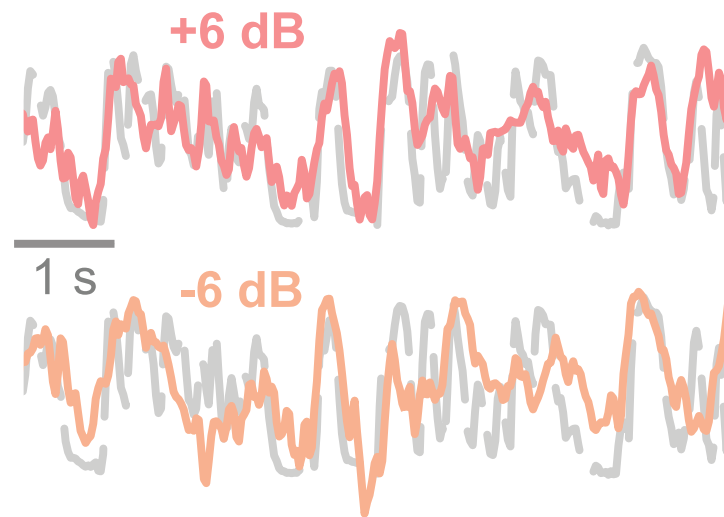
Speech in Noise: Results

Neural Reconstruction of
Underlying Speech Envelope

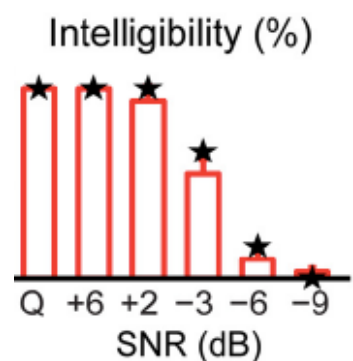
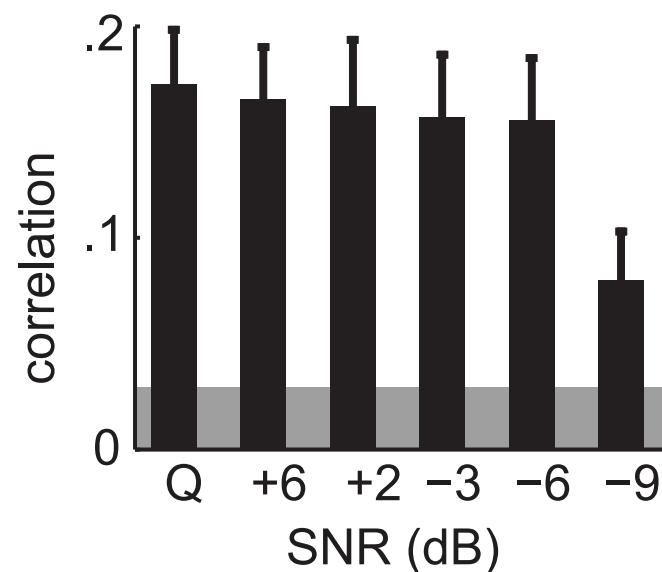


Speech in Noise: Results

Neural Reconstruction of Underlying Speech Envelope

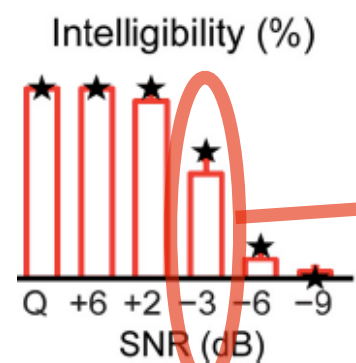
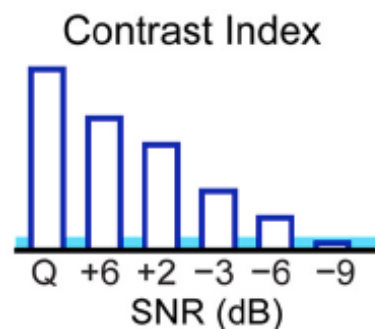
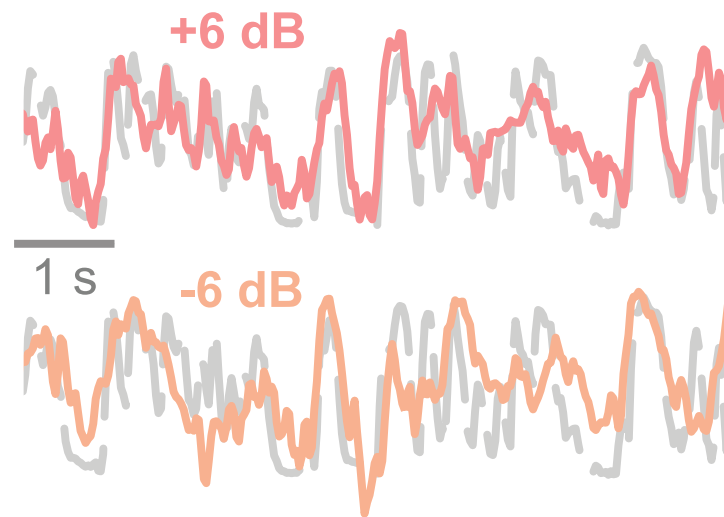


Reconstruction Accuracy

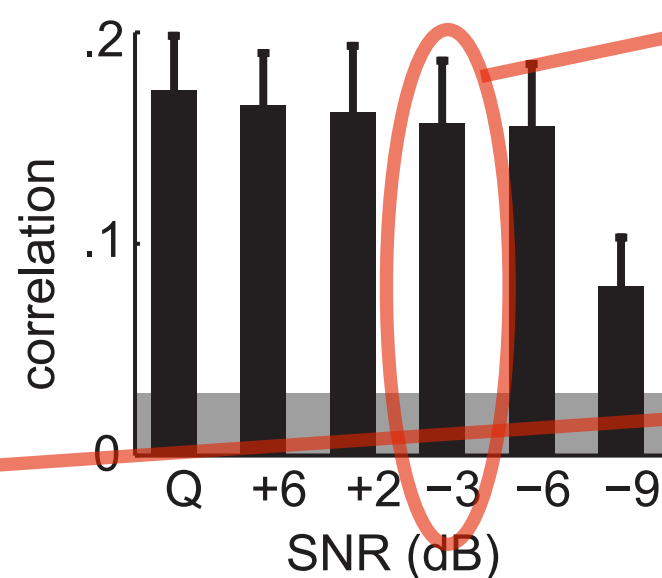


Speech in Noise: Results

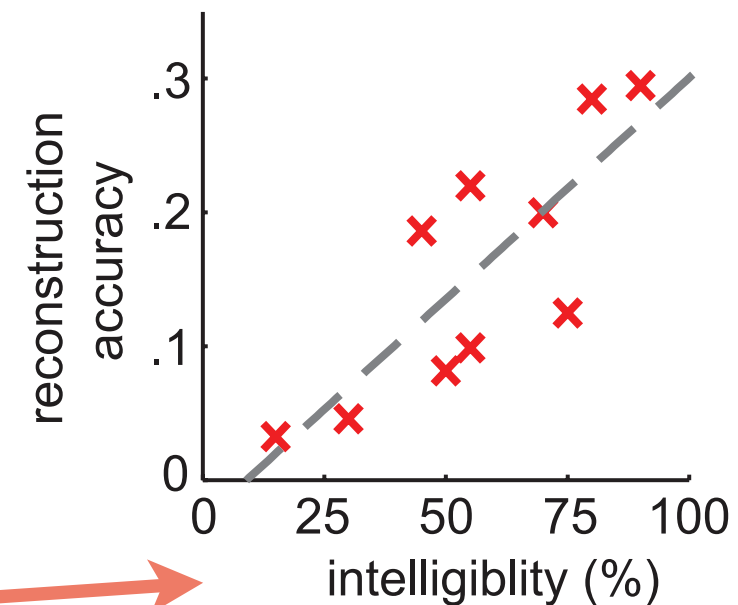
Neural Reconstruction of Underlying Speech Envelope



Reconstruction Accuracy

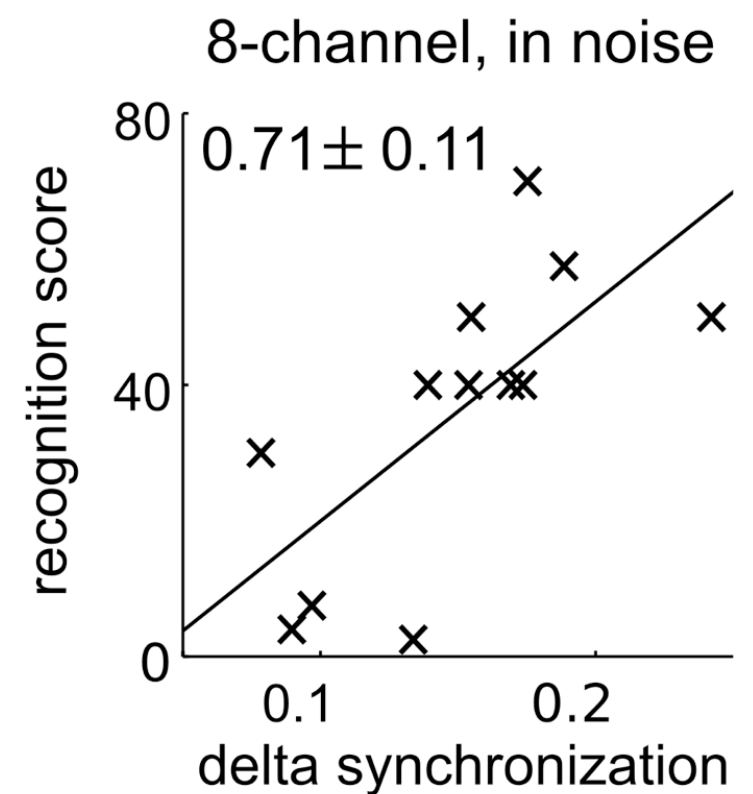
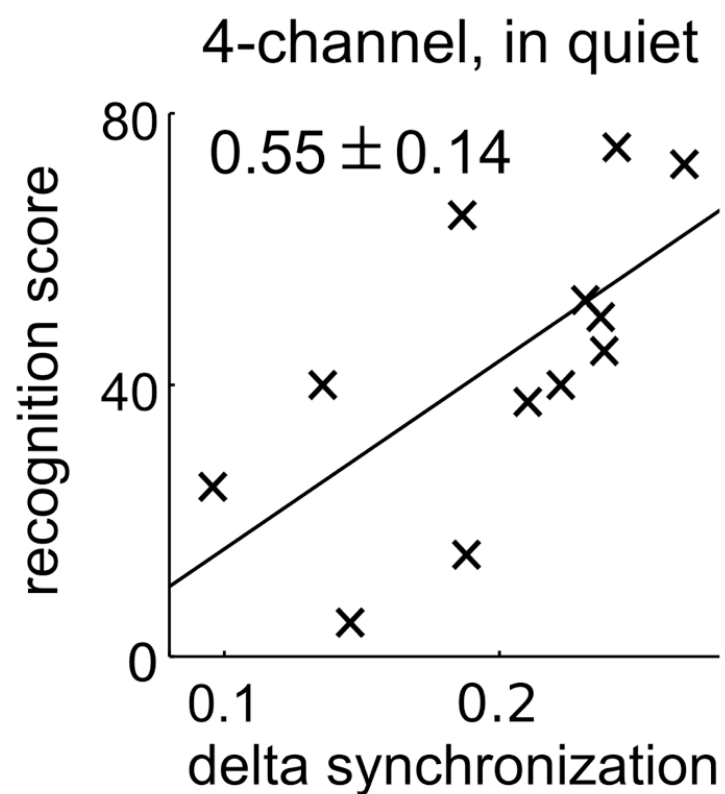
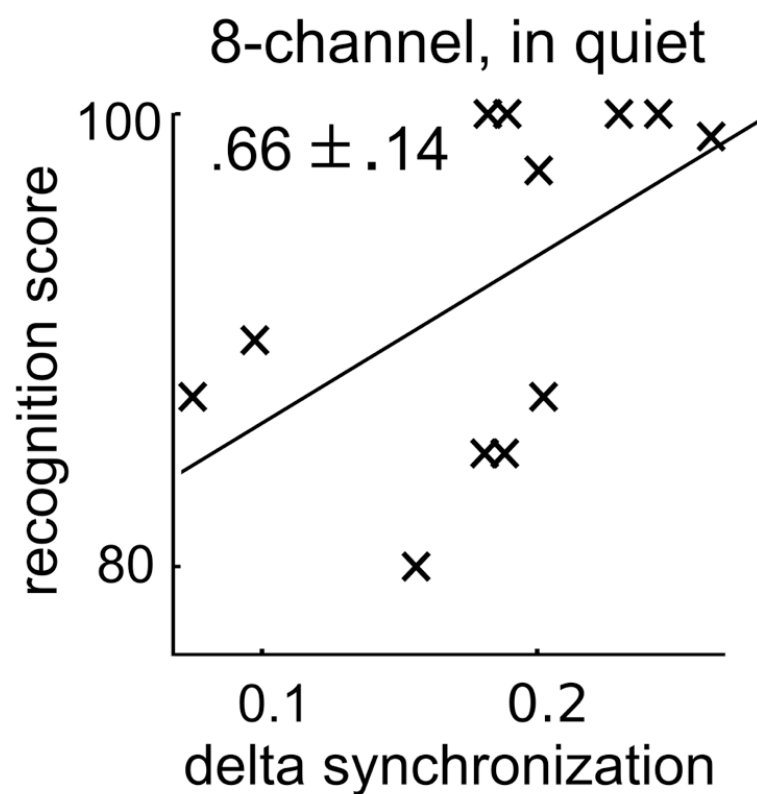


Correlation with Intelligibility



across Subjects

Noise-Vocoded Speech



Intelligibility Reflected only in Delta Band (1–4 Hz)

Multiple Cortical Speech 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: $\sim 1 - 10$ 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

Listening to Speech at the Cocktail Party



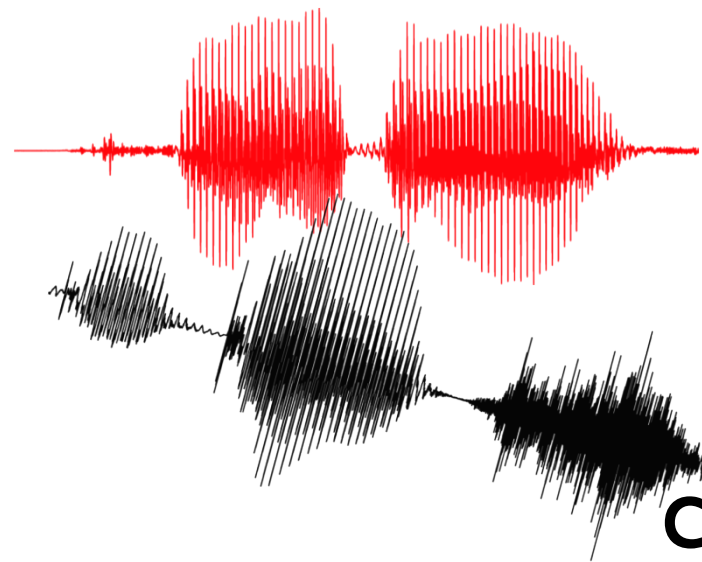
Alex Katz,
The Cocktail Party

Listening to Speech at the Cocktail Party



Alex Katz,
The Cocktail Party

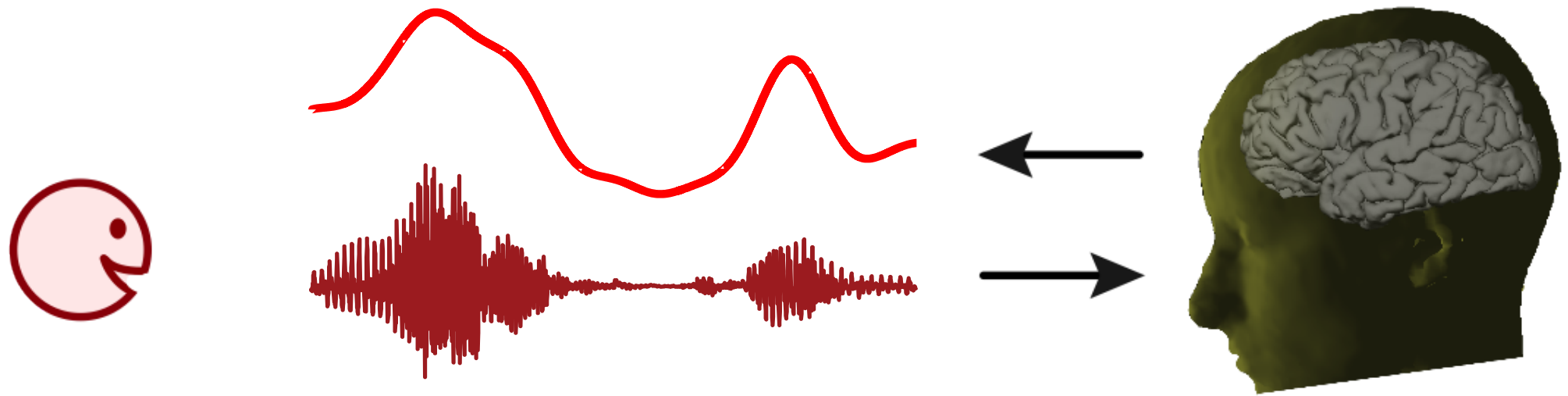
Competing Speech Streams



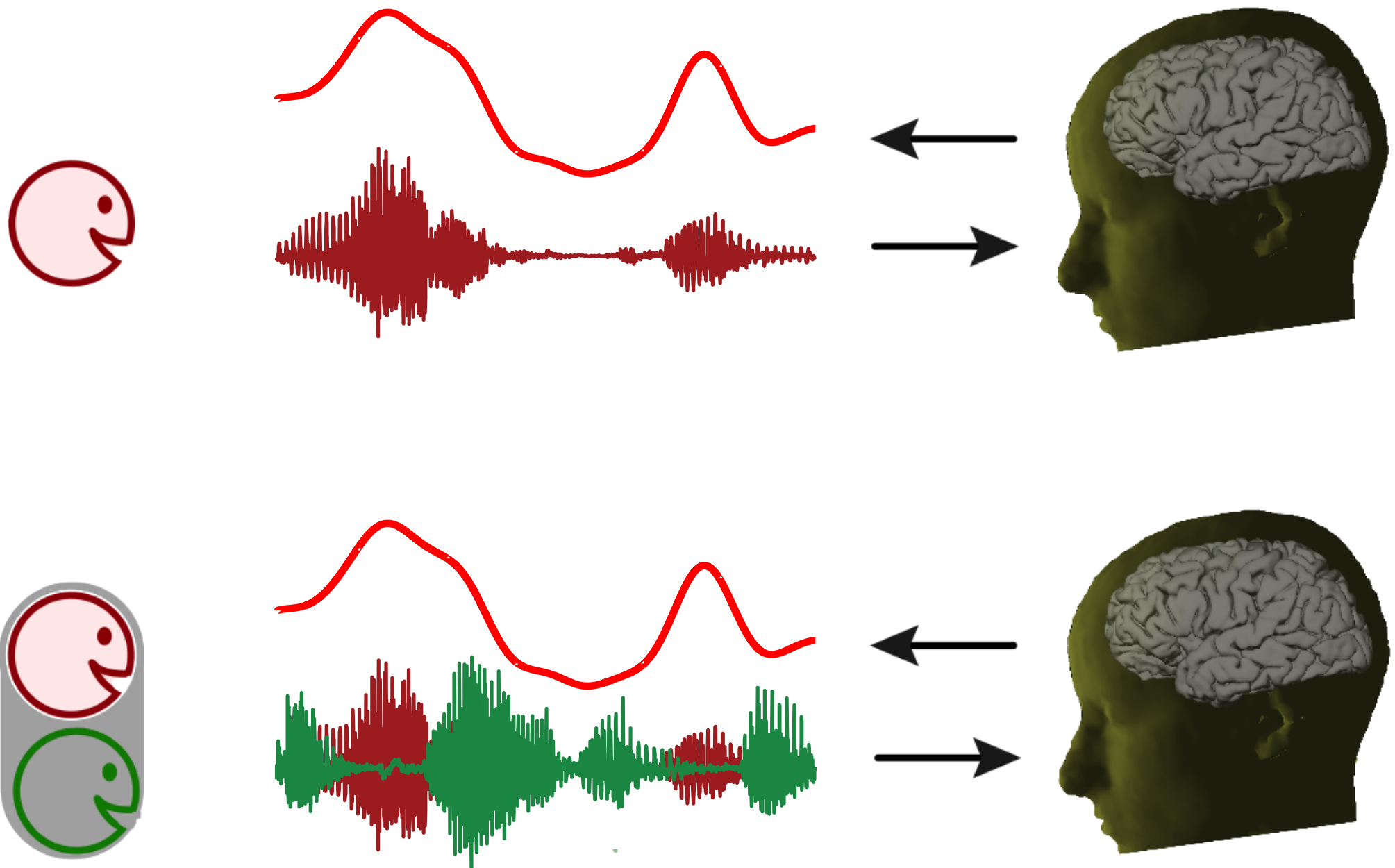
speech

competing speech

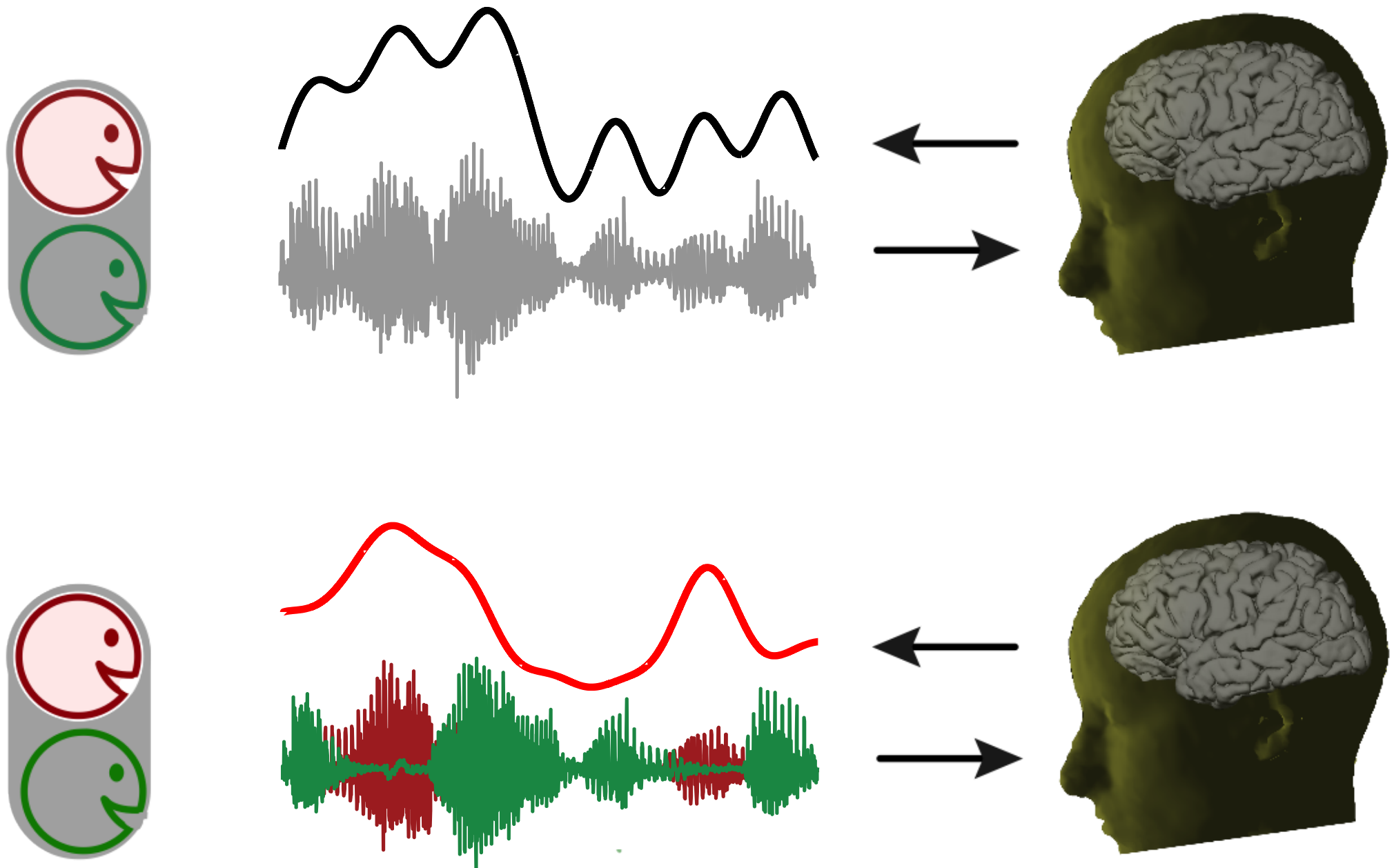
Selective Neural Encoding



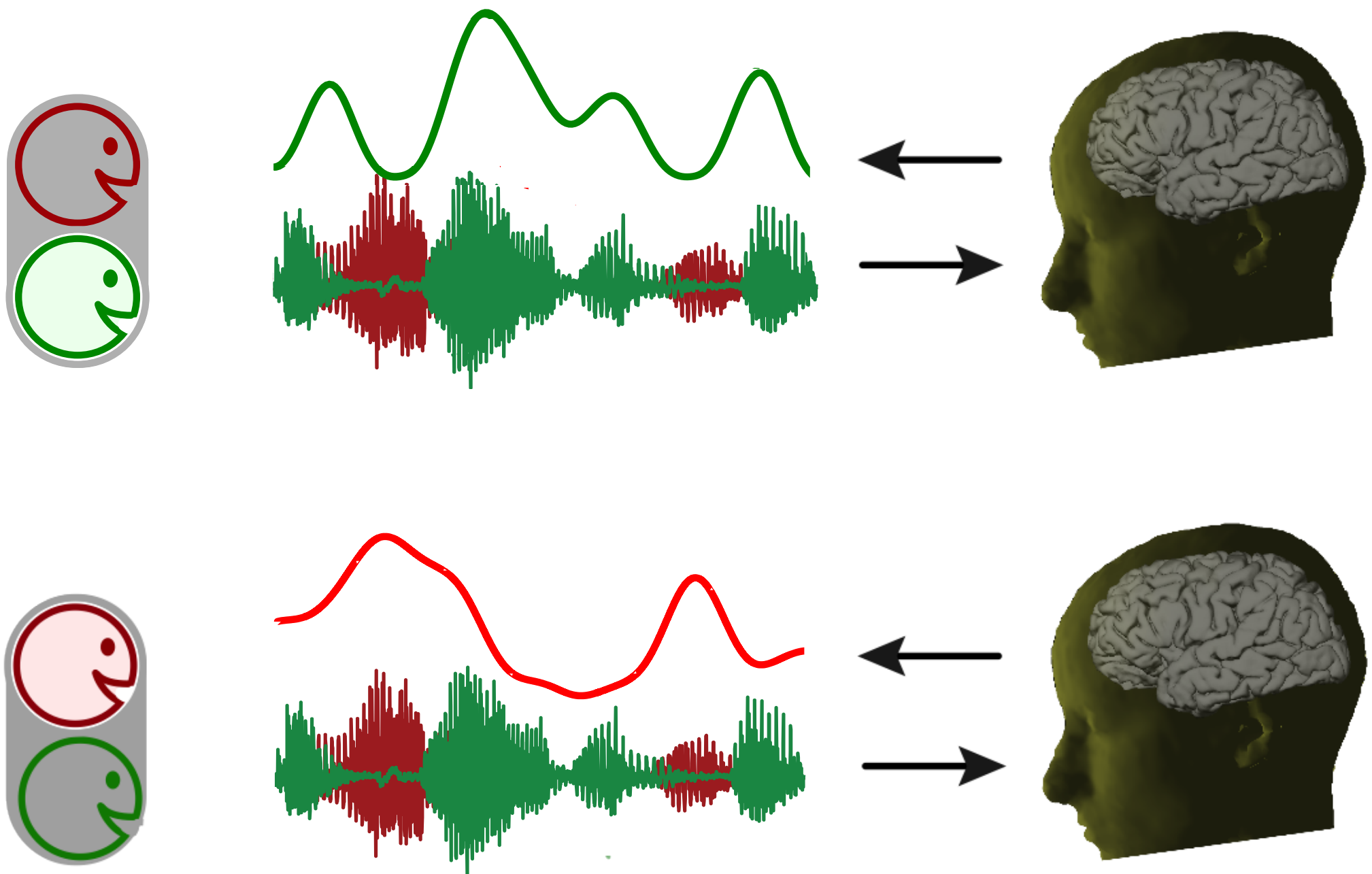
Selective Neural Encoding



Unselective vs. Selective Neural Encoding



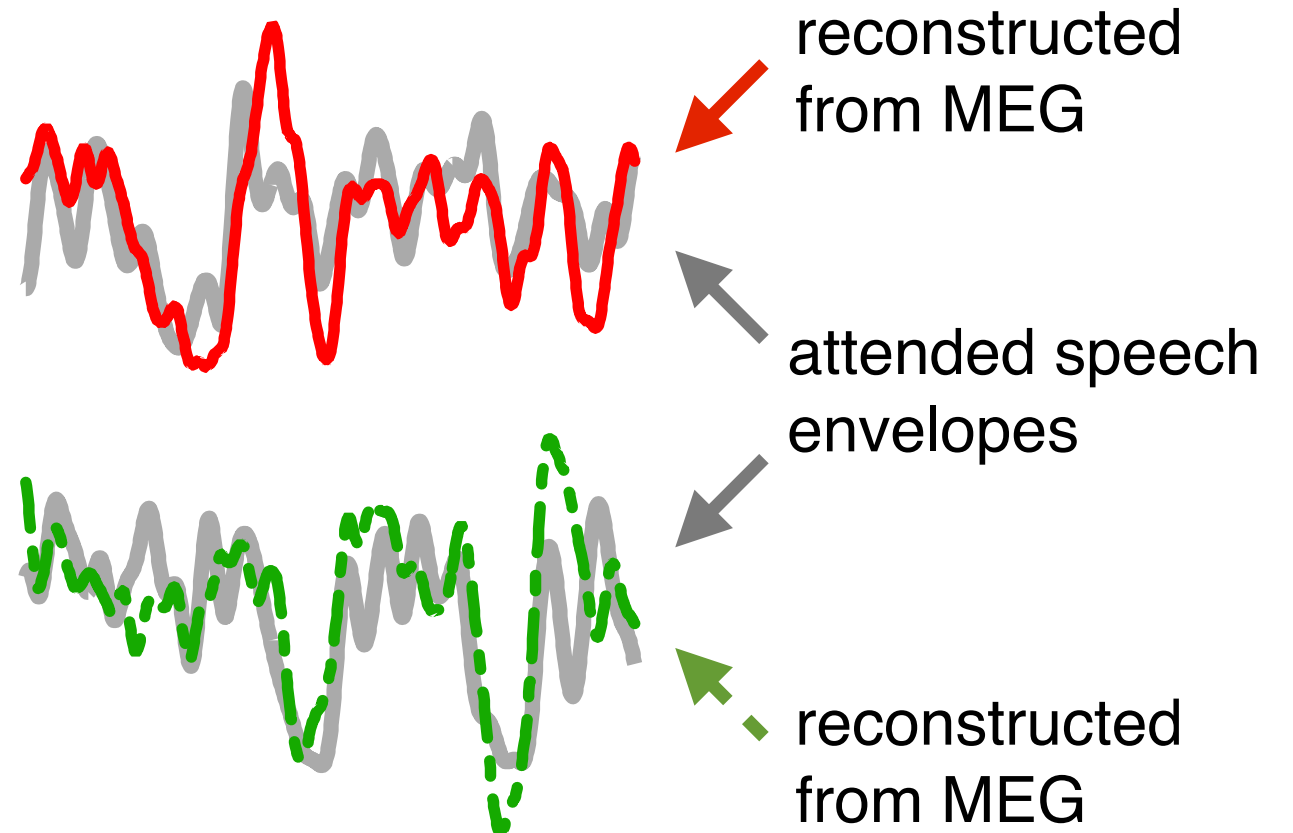
Selective Neural Encoding



Selective Encoding: Results

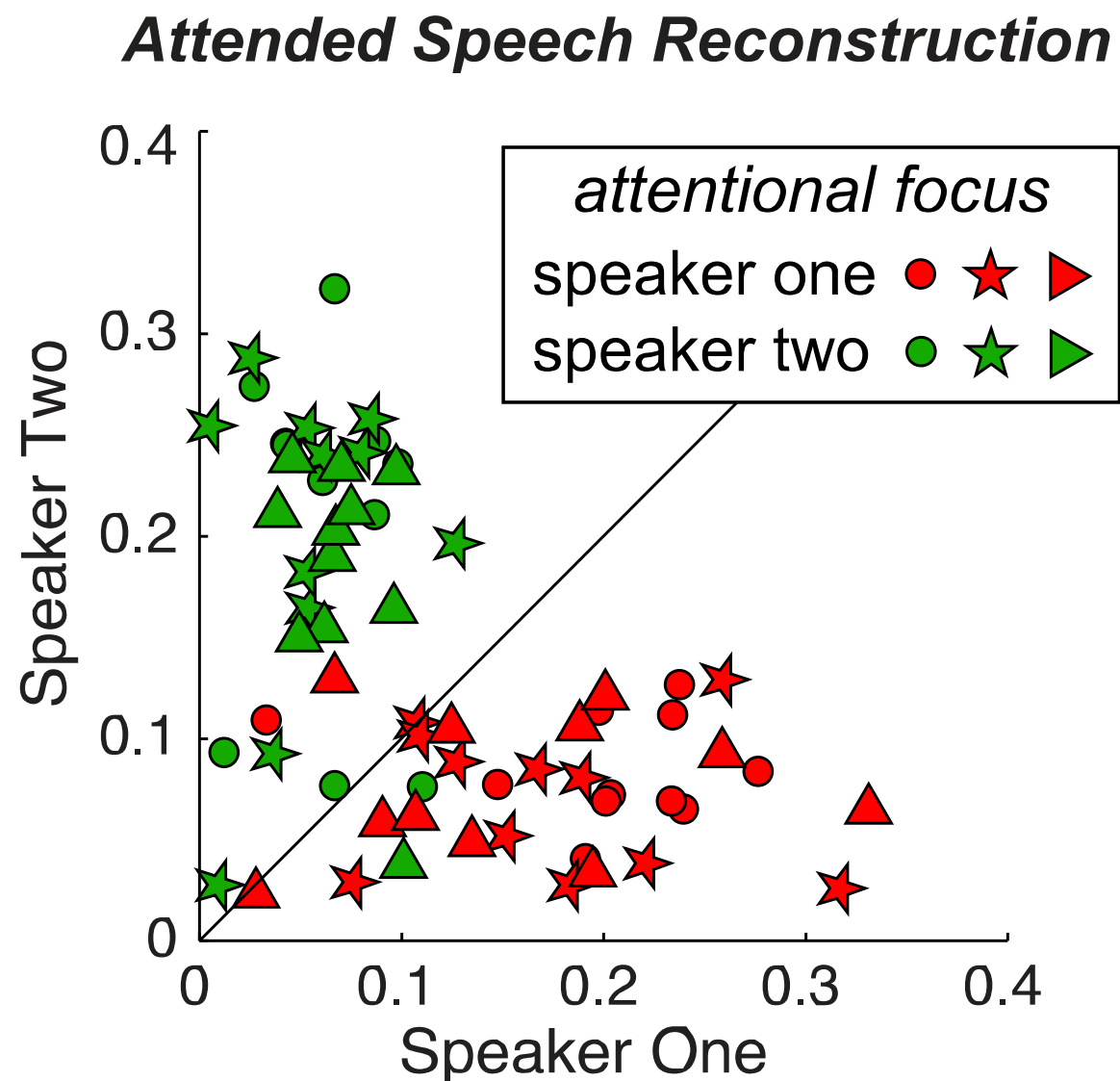
attending to
speaker 1

attending to
speaker 2



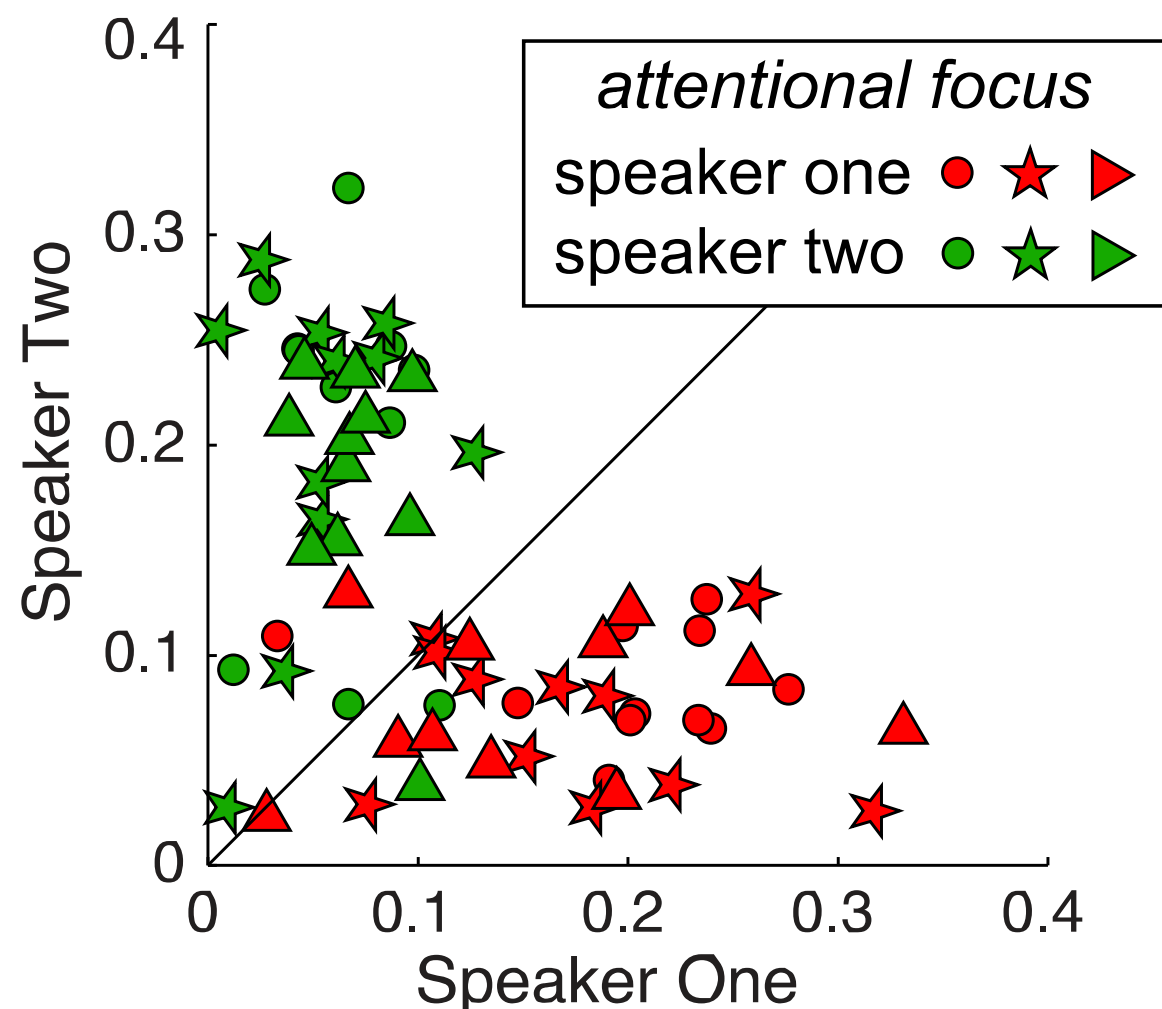
Identical Stimuli!

Single Trial Speech Reconstruction

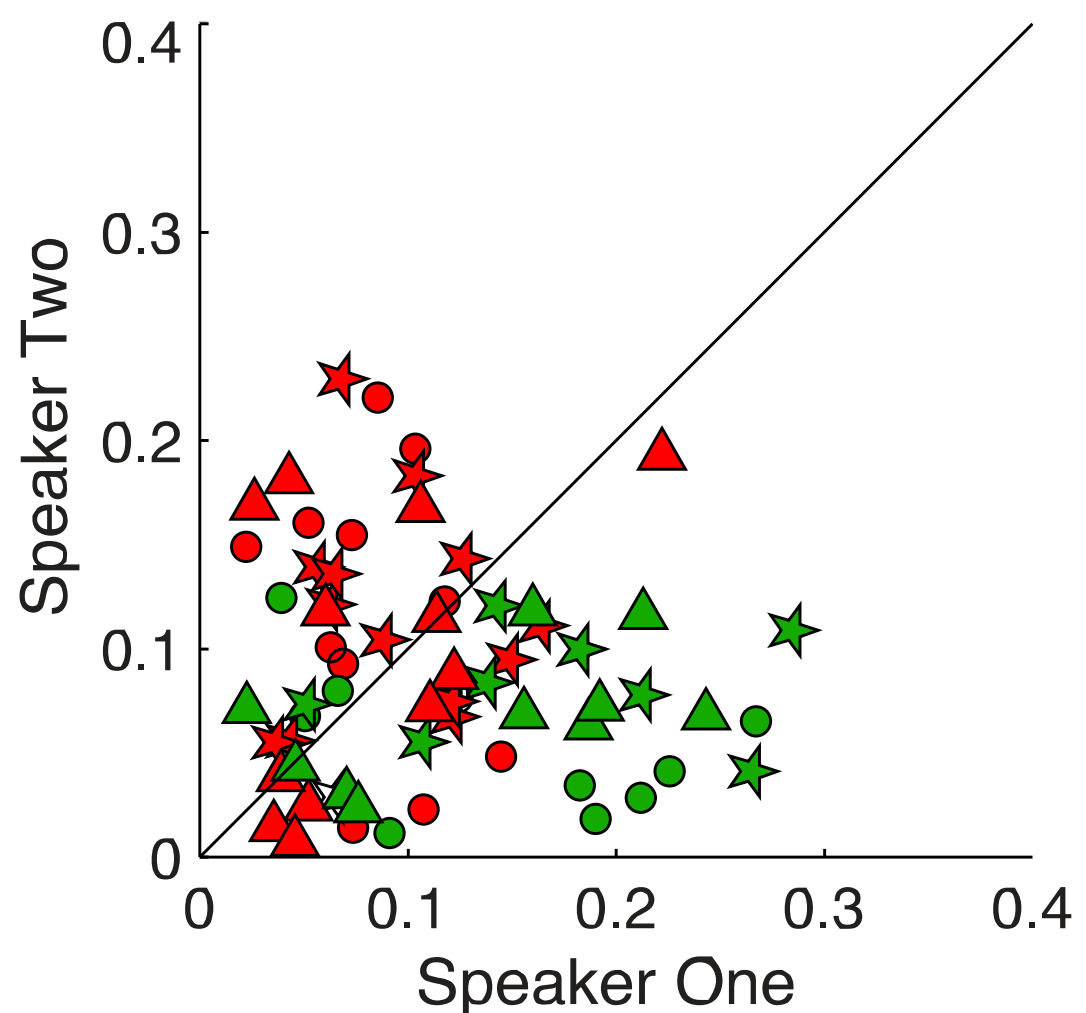


Single Trial Speech Reconstruction

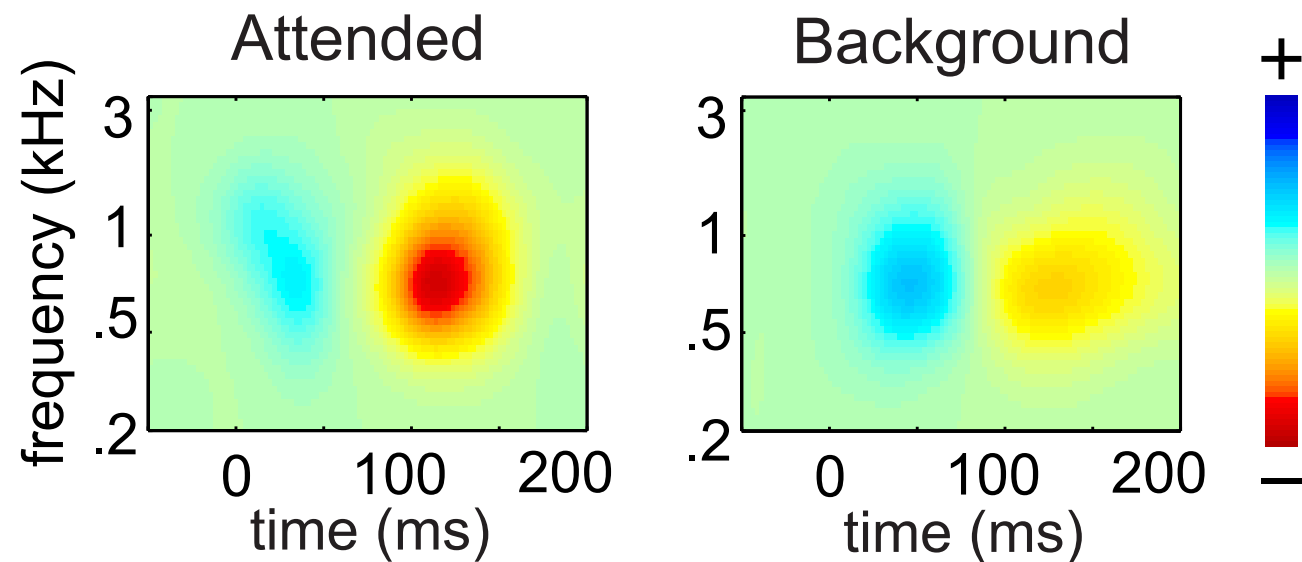
Attended Speech Reconstruction



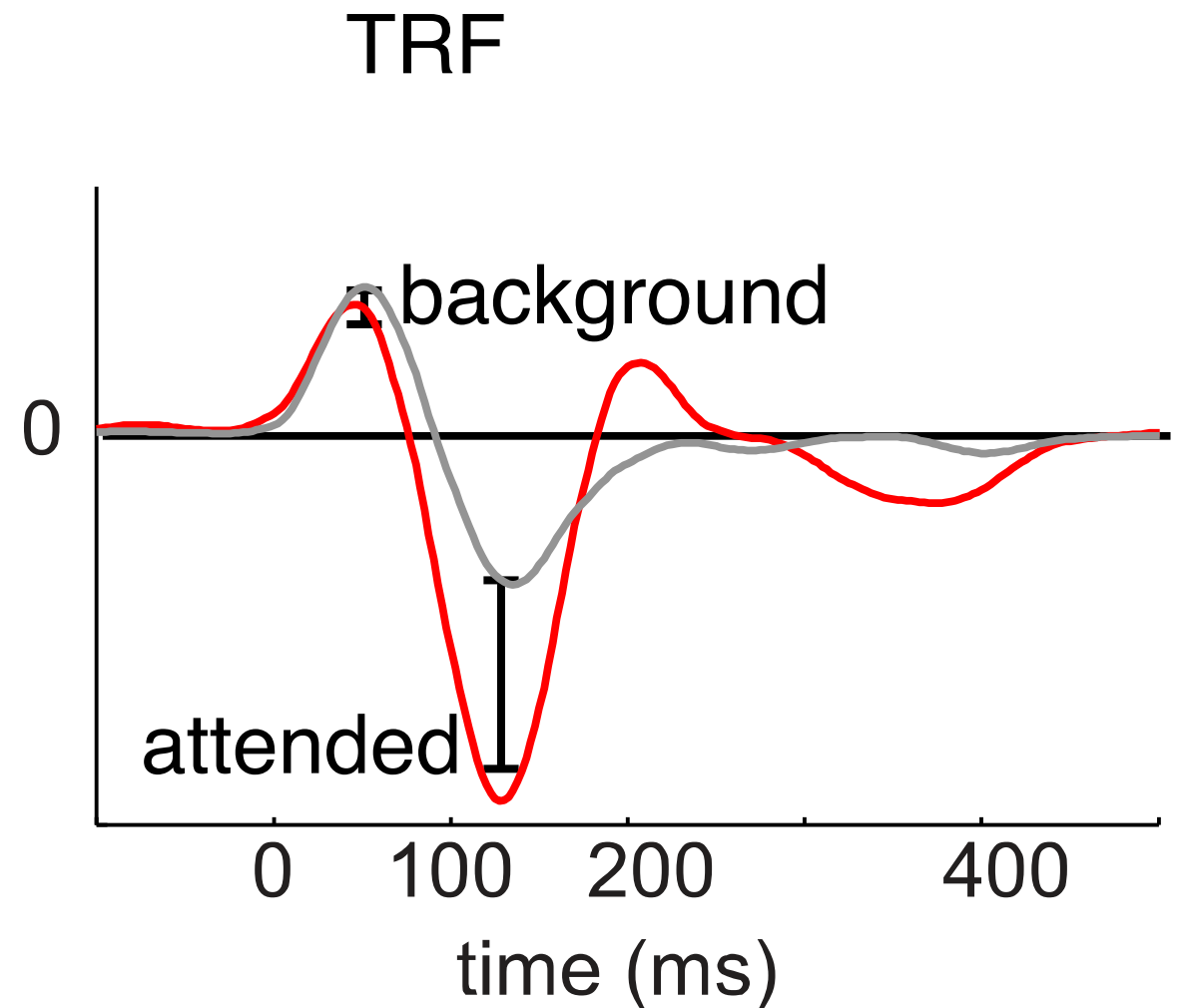
Background Speech Reconstruction



STRF Results

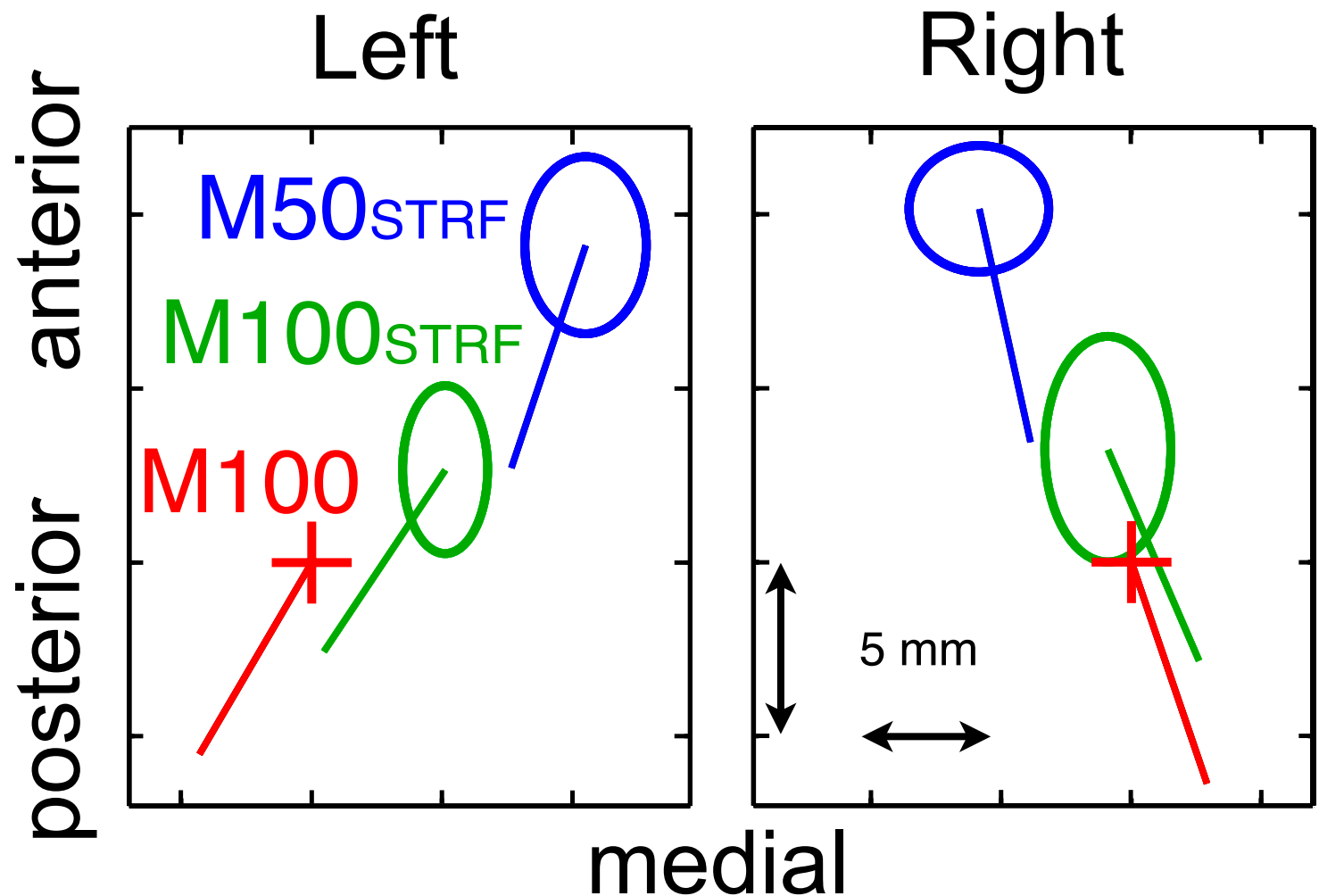


- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- M50_{STRF} positive peak
- M100_{STRF} negative peak
- **M100_{STRF} strongly modulated by attention, *but not* M50_{STRF}**



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***



Recent Results

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

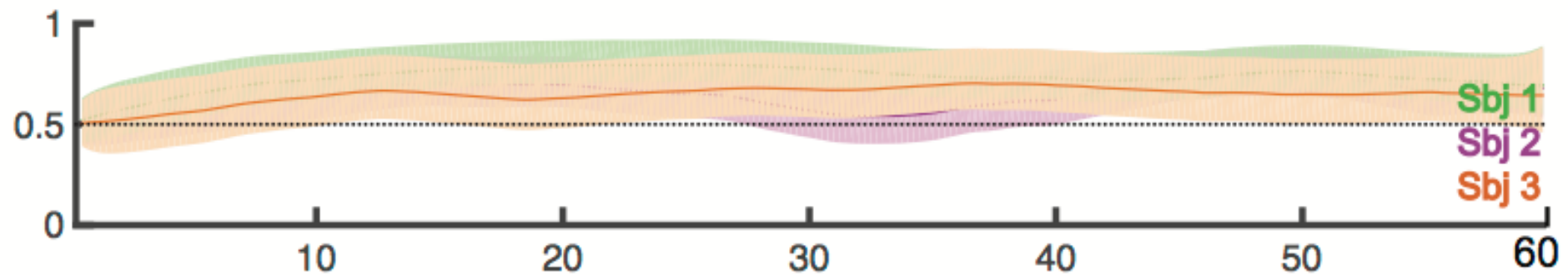
Recent Results

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

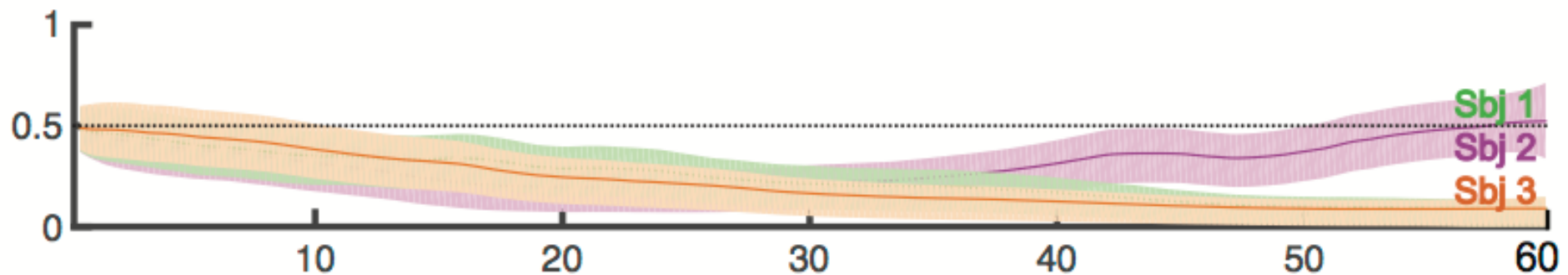
Attentional Dynamics

Attend to Speaker 1

Probability
of attending
Speaker 1



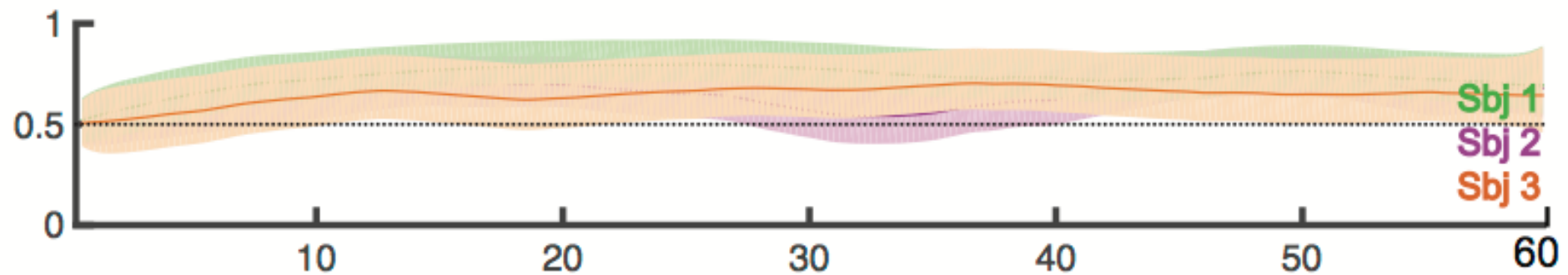
Attend to Speaker 2



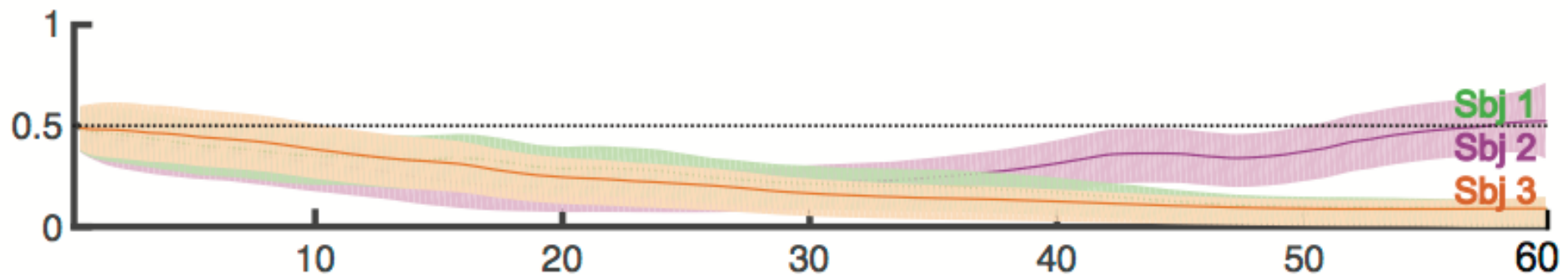
Attentional Dynamics

Attend to Speaker 1

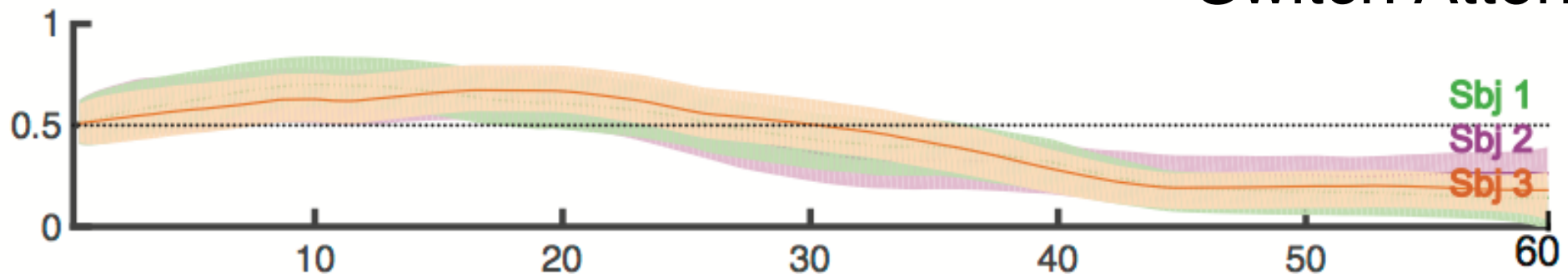
Probability
of attending
Speaker 1



Attend to Speaker 2



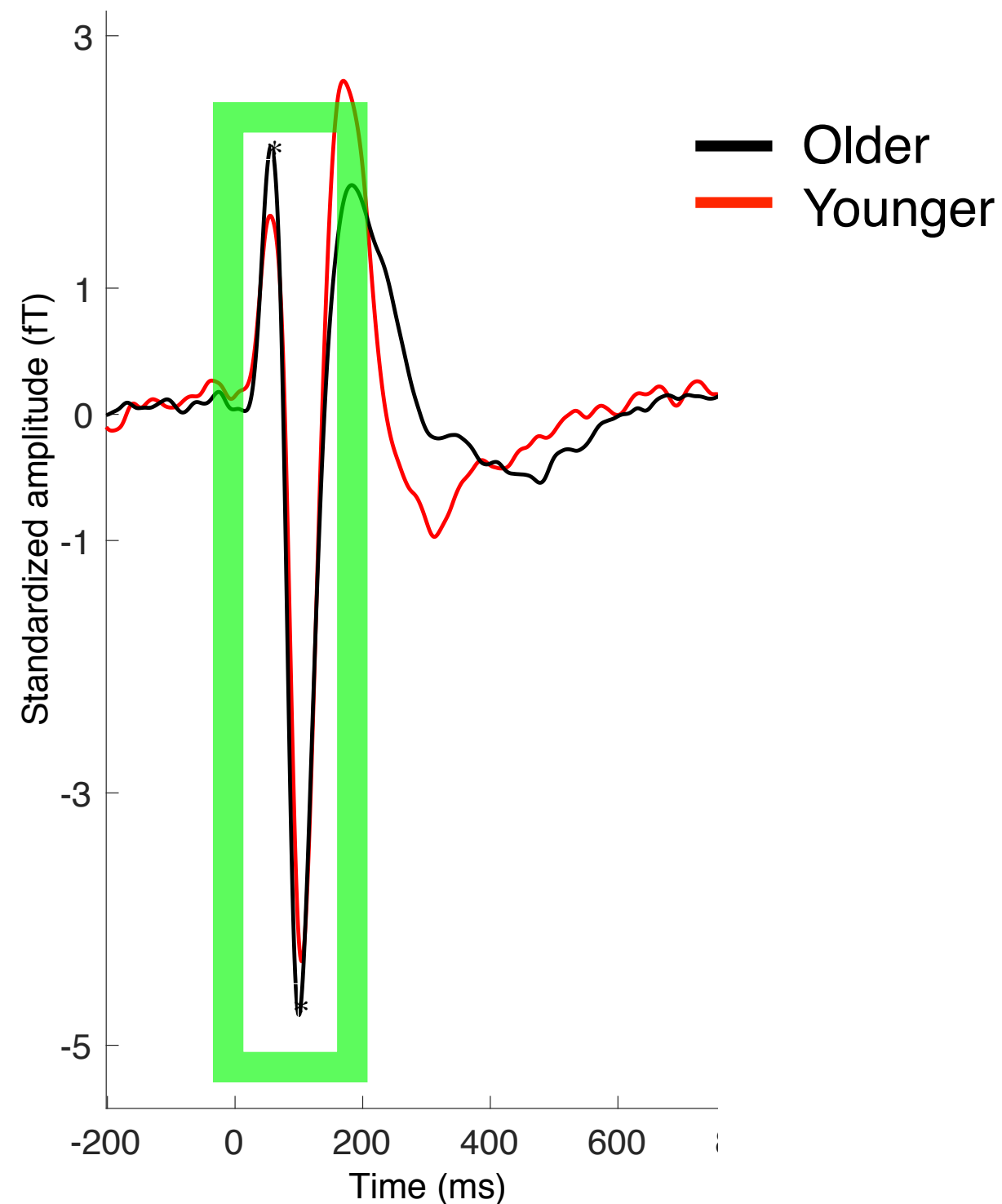
Switch Attention



Recent Results

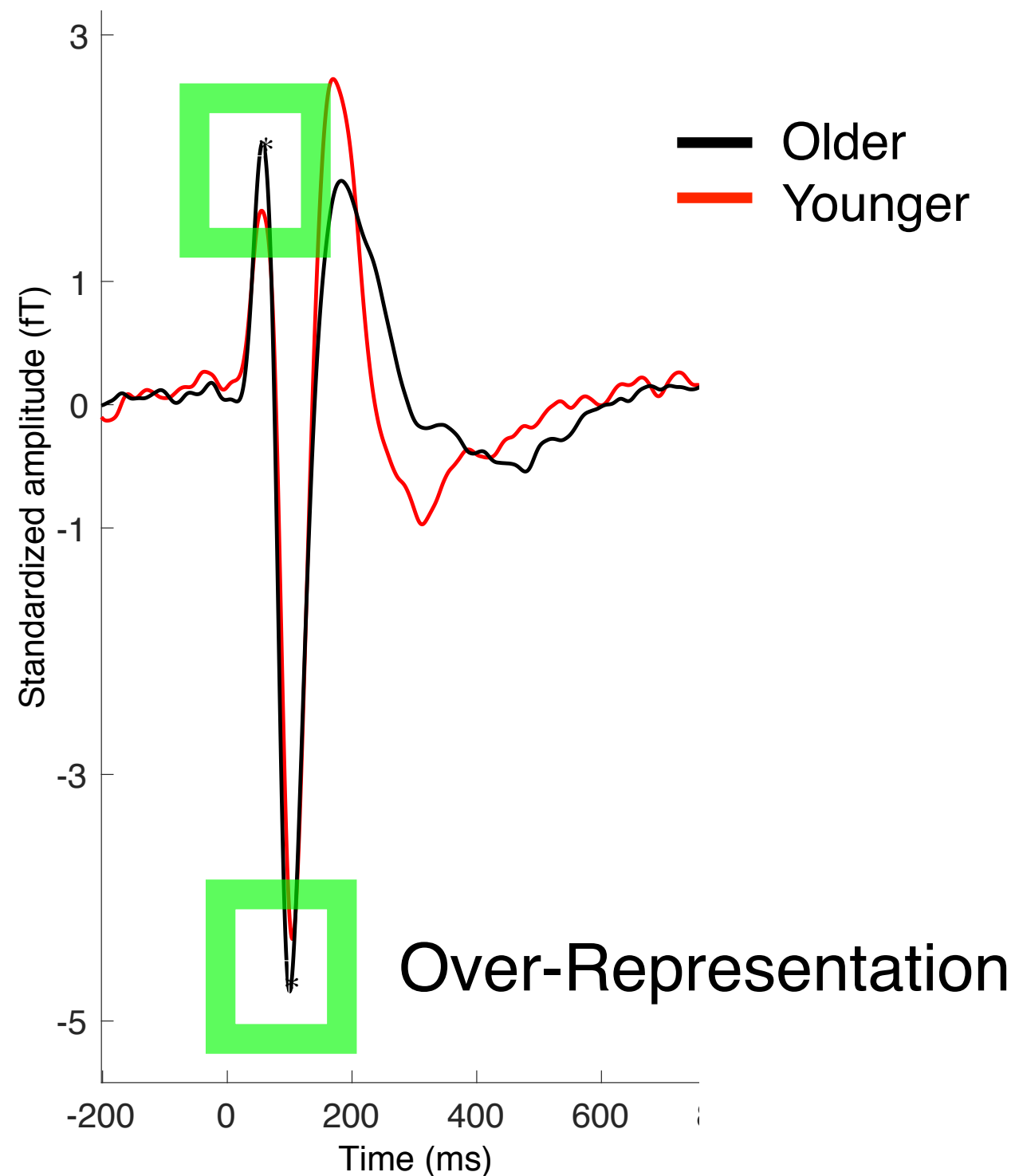
- Attentional Dynamics
- Aging & Cortical Representations of Speech
- High Level Interference & Noise

Aging & Auditory Cortex



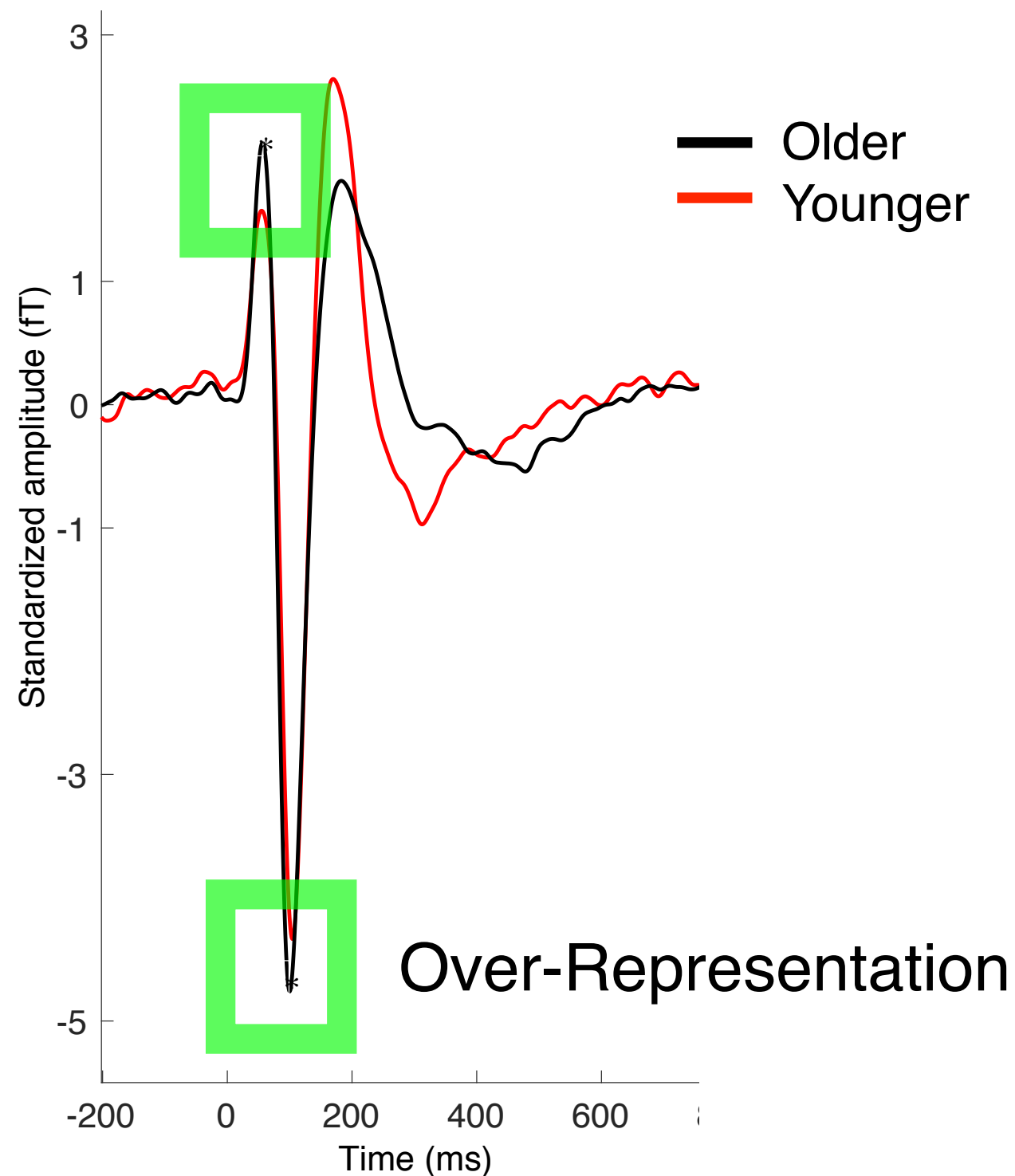
Average Responses to Pure Tone

Aging & Auditory Cortex

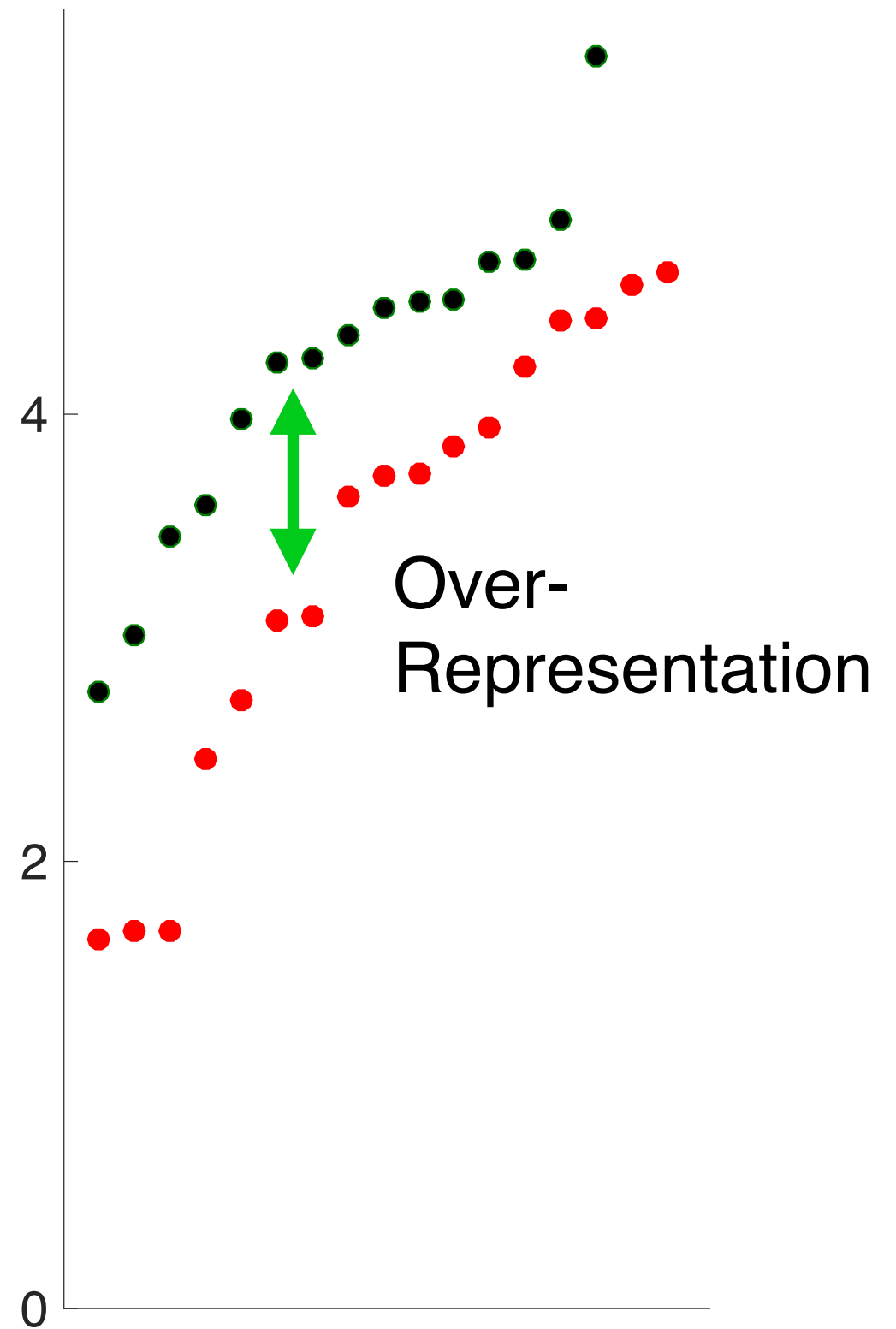


Average Responses to Pure Tone

Aging & Auditory Cortex

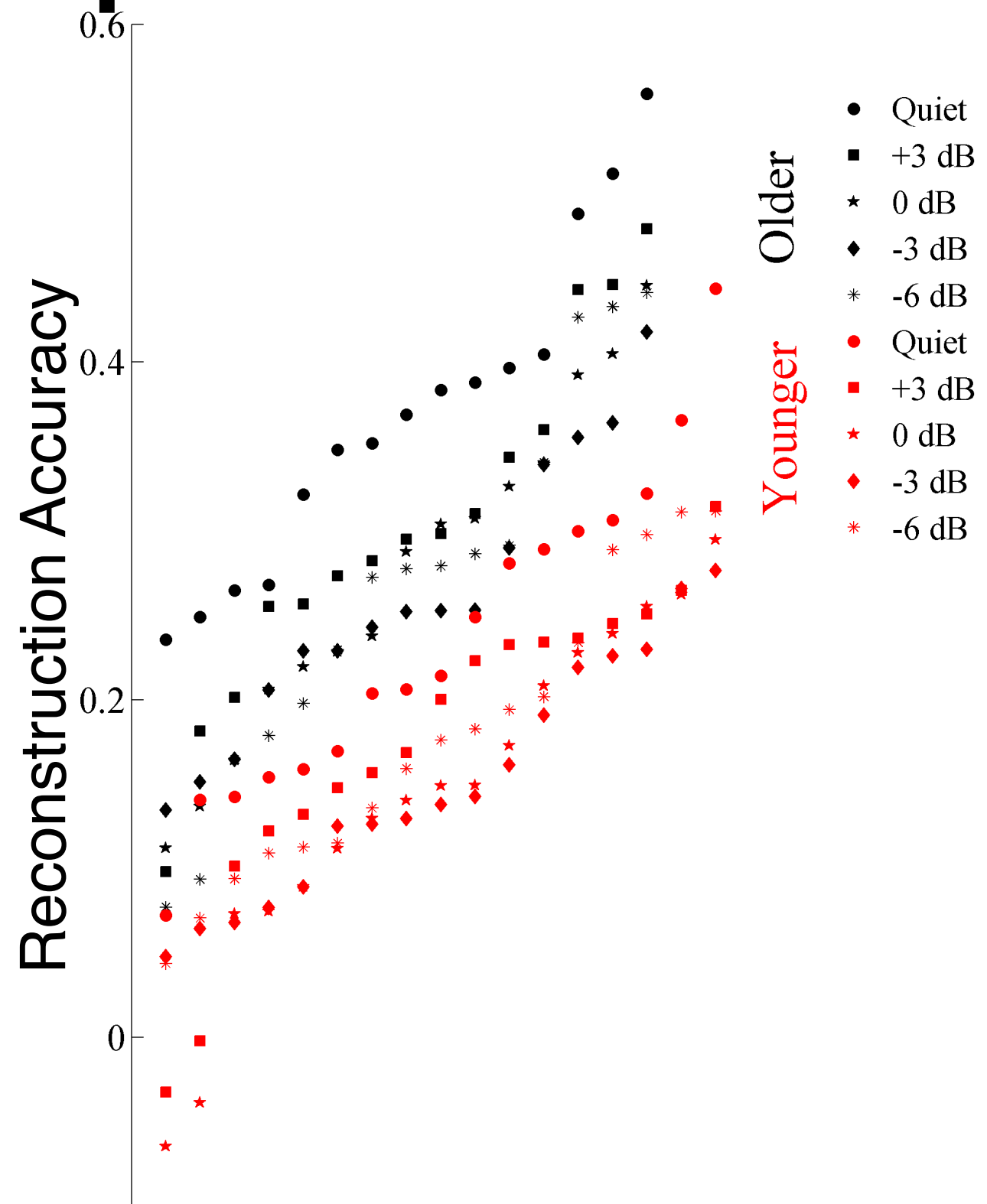


Average Responses to Pure Tone

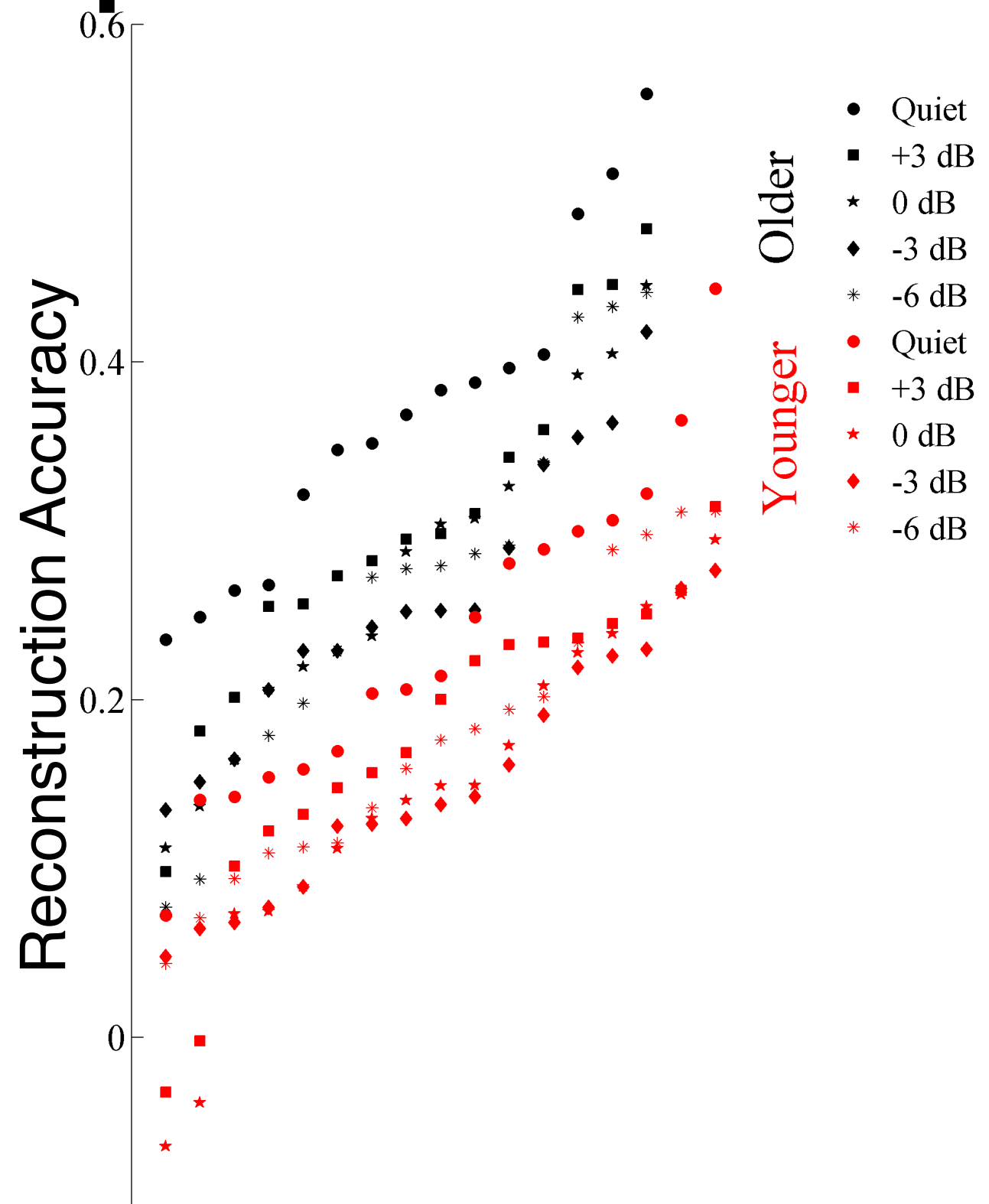
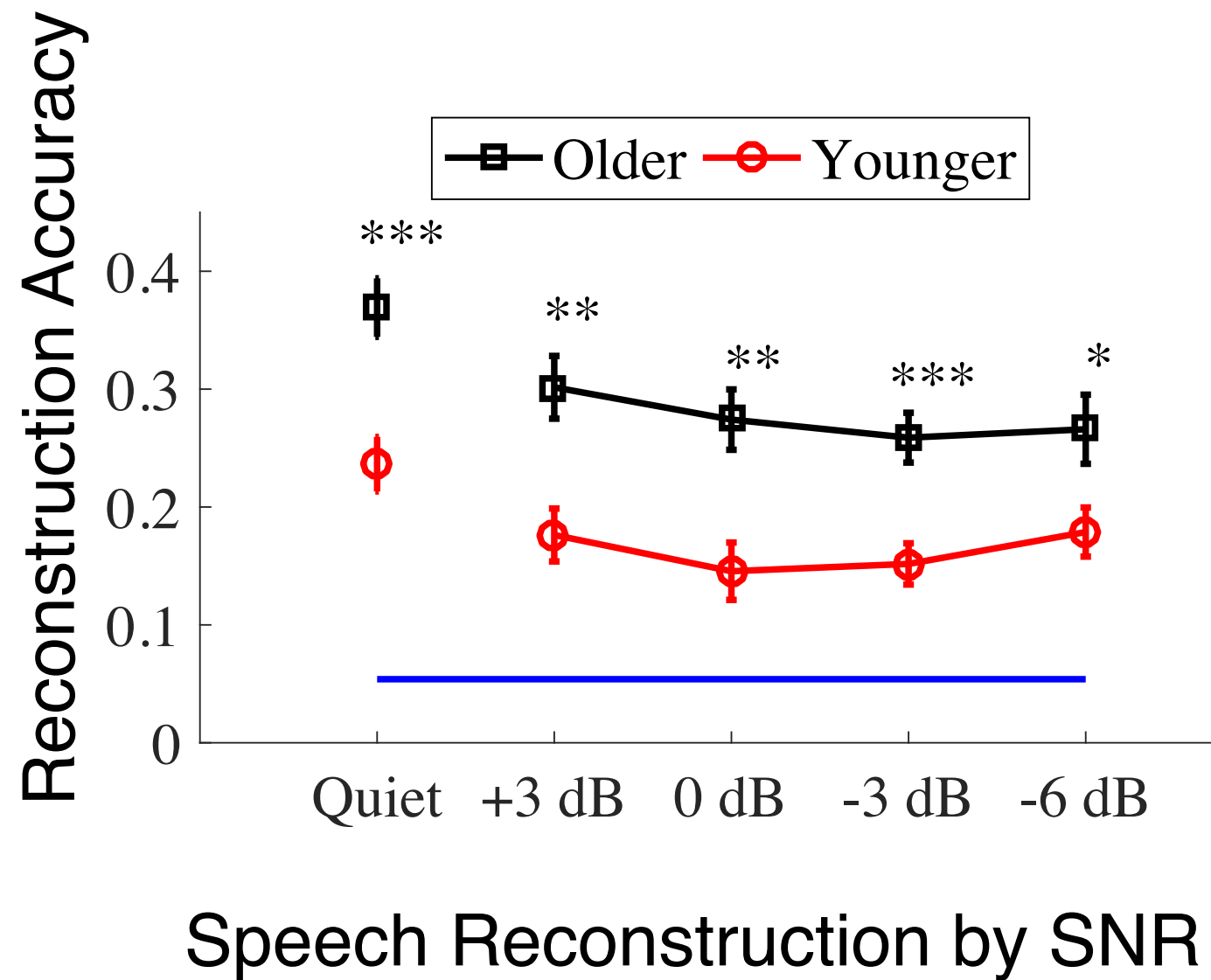


M100 Power by Subject

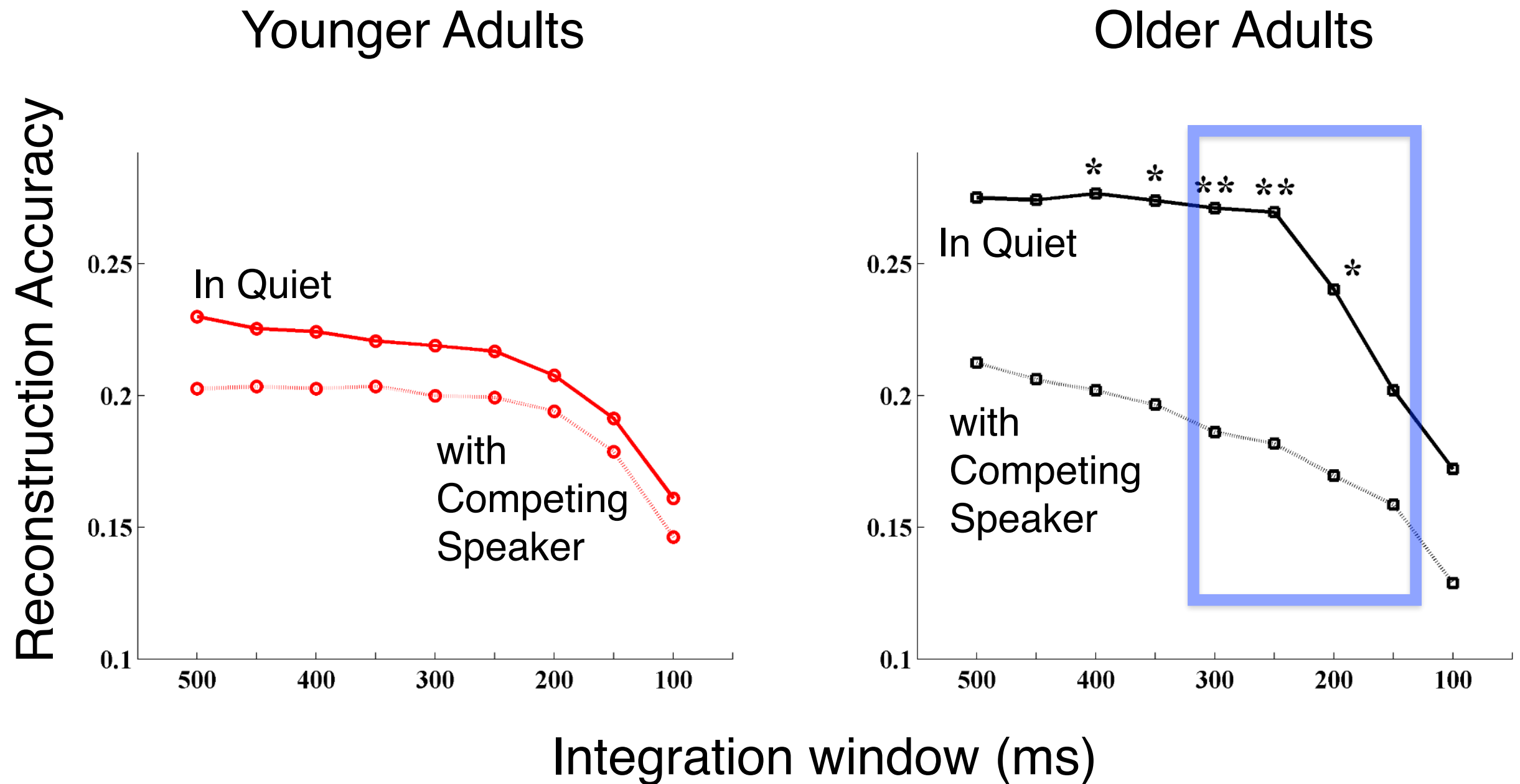
Speech Over-Representation



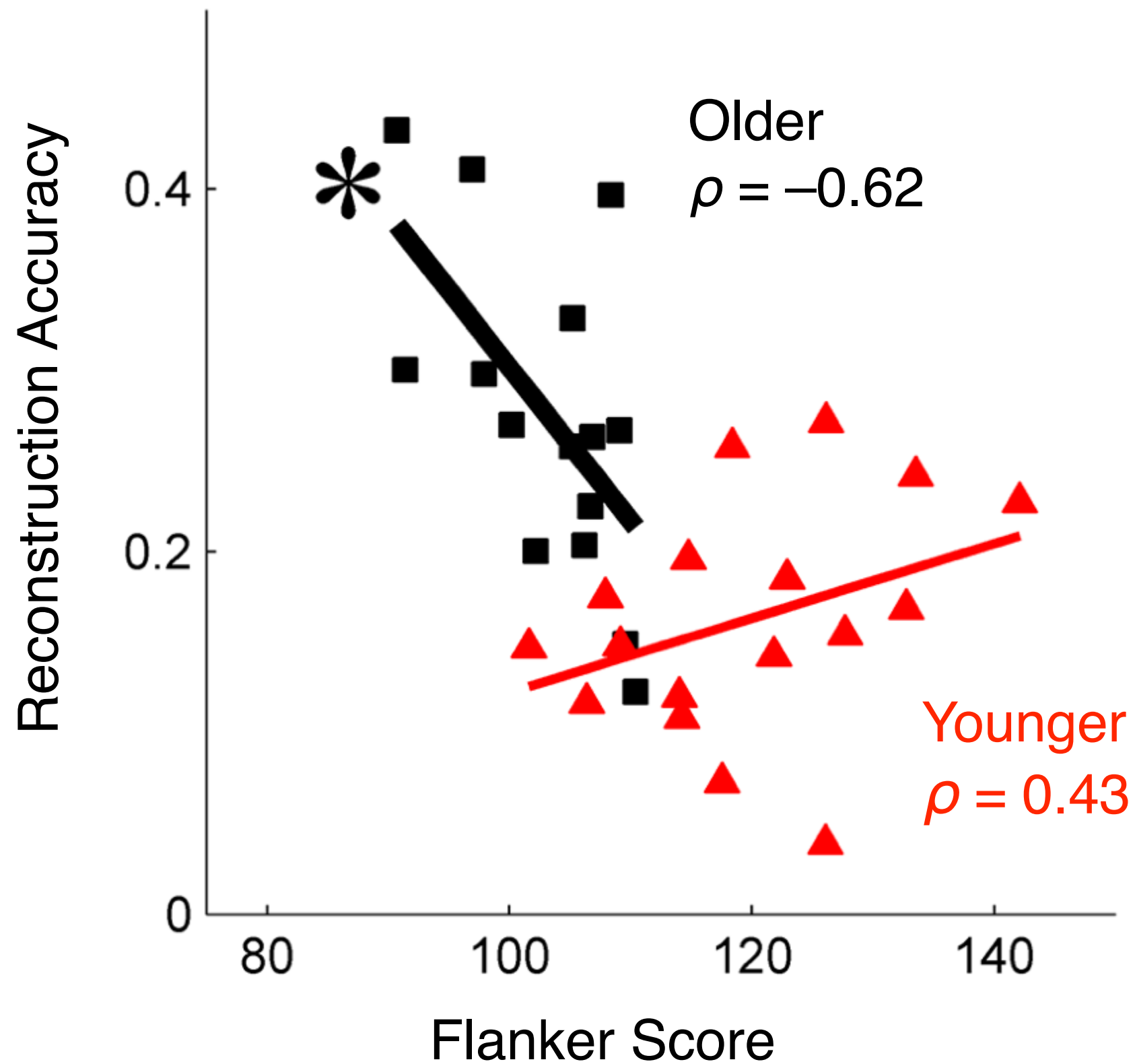
Speech Over-Representation



Aging & Integration Time



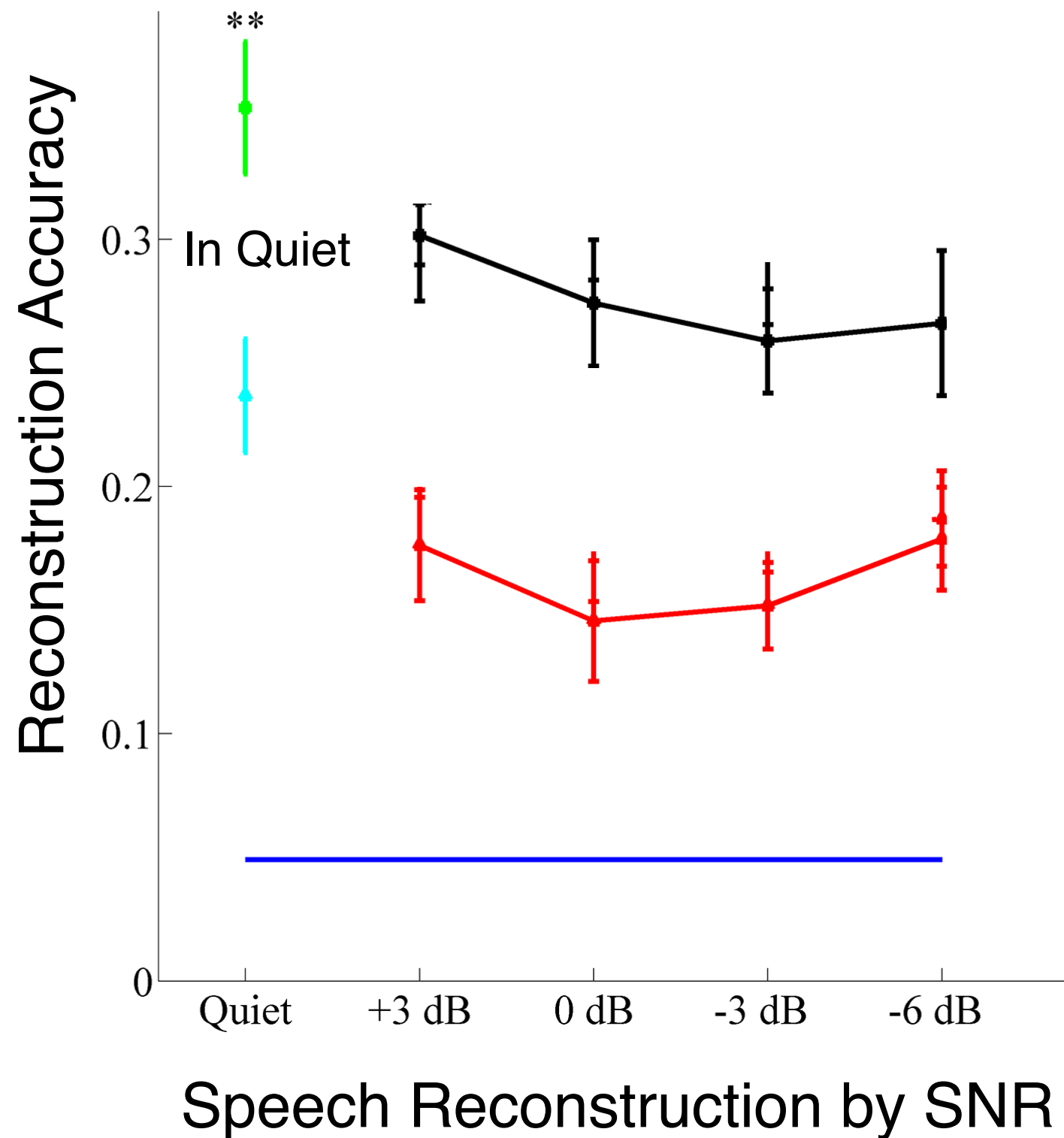
Neural vs Inhibitory Control



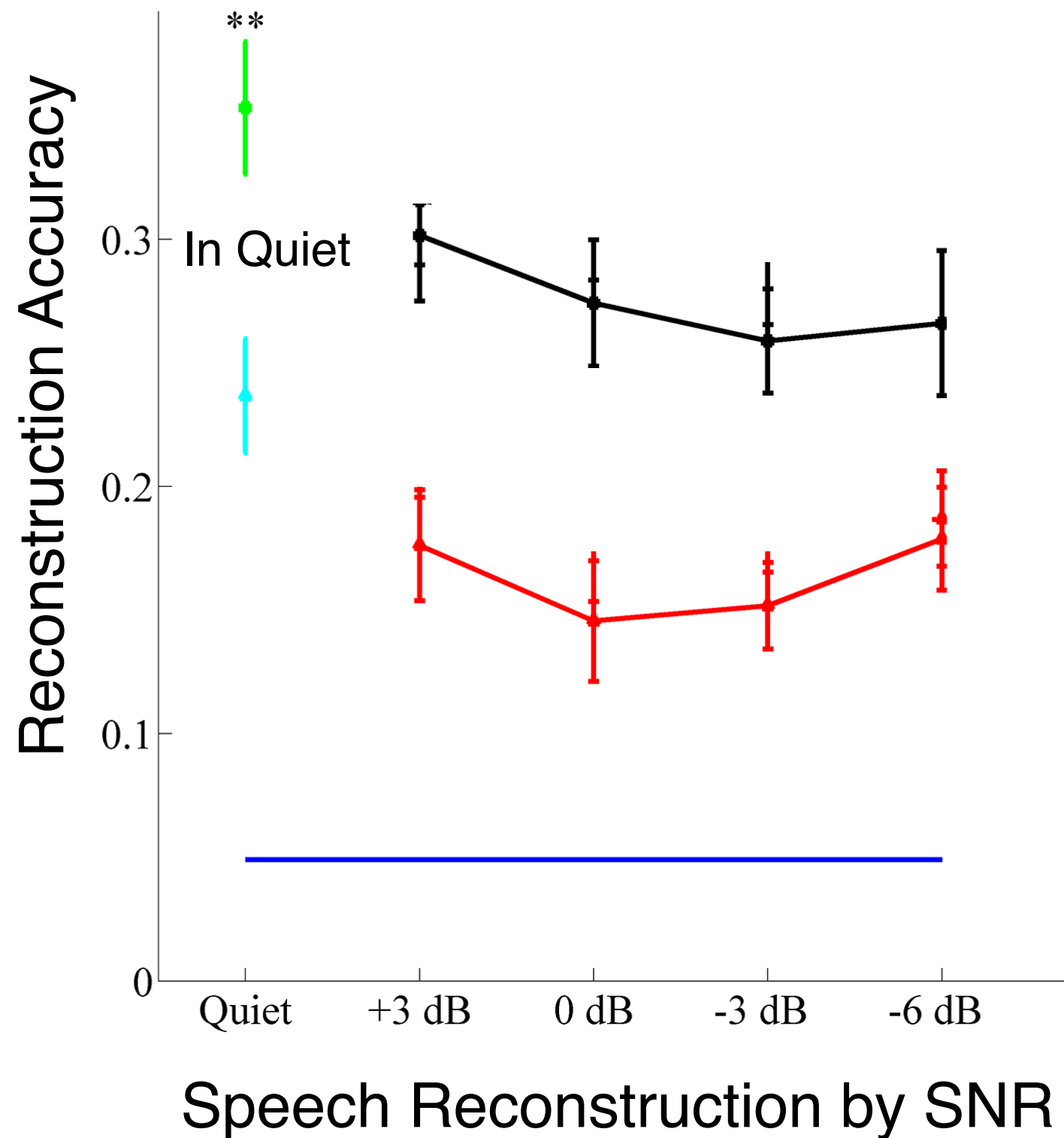
Recent Results

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

High Level Interference Effects

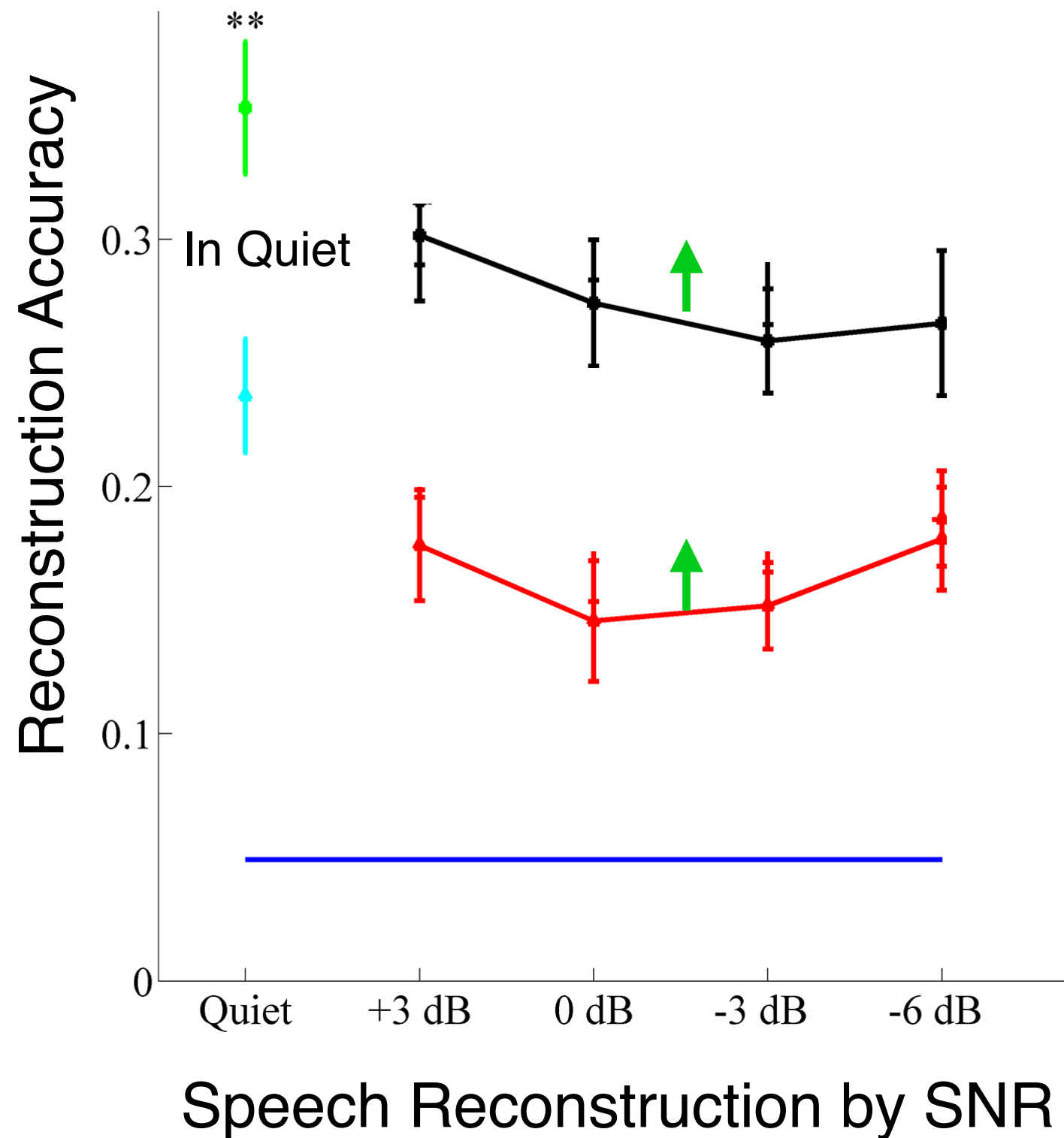


High Level Interference Effects



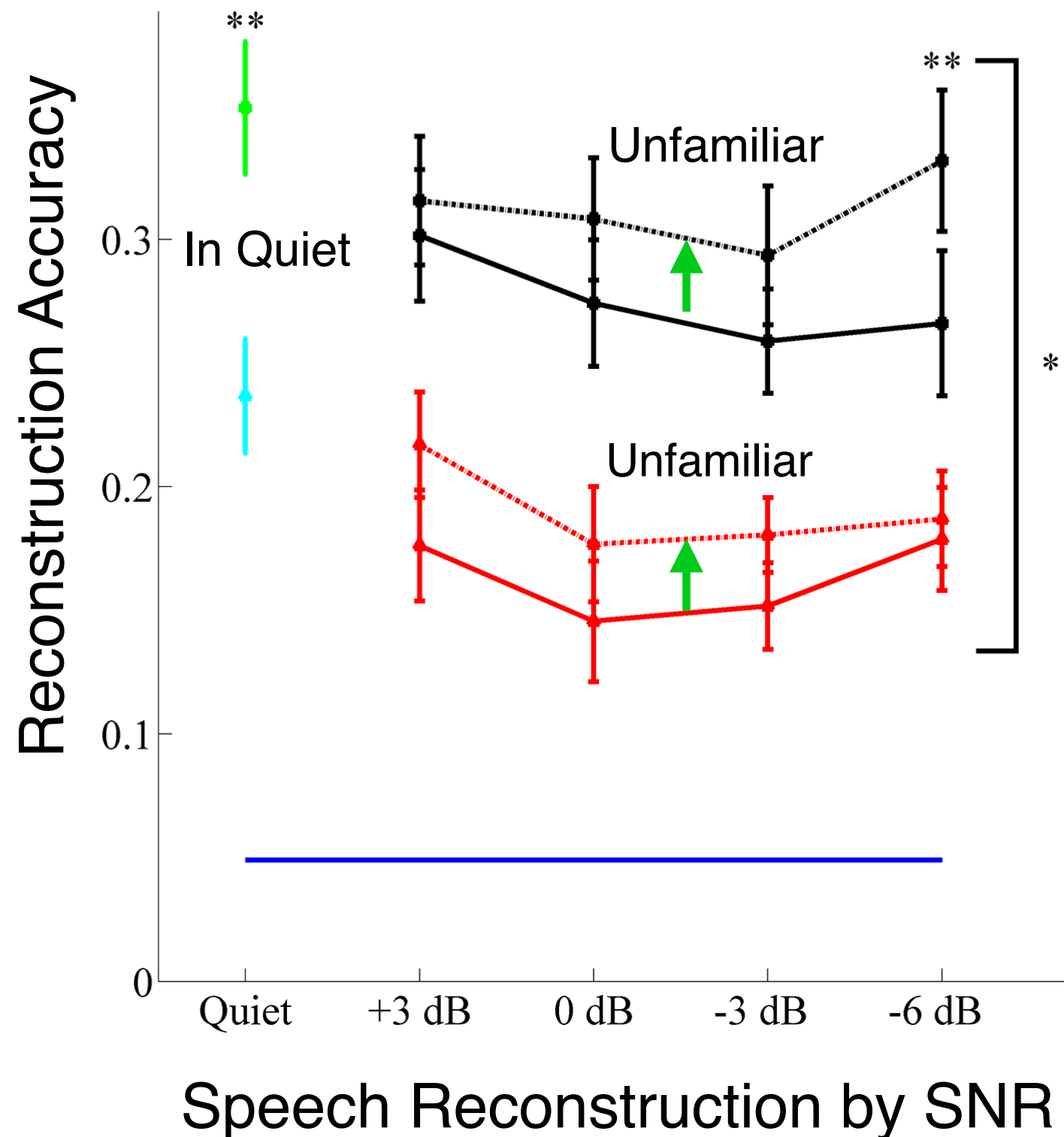
- **Unfamiliarity of Background**
 - Boosts Intelligibility of Attended Speech

High Level Interference Effects



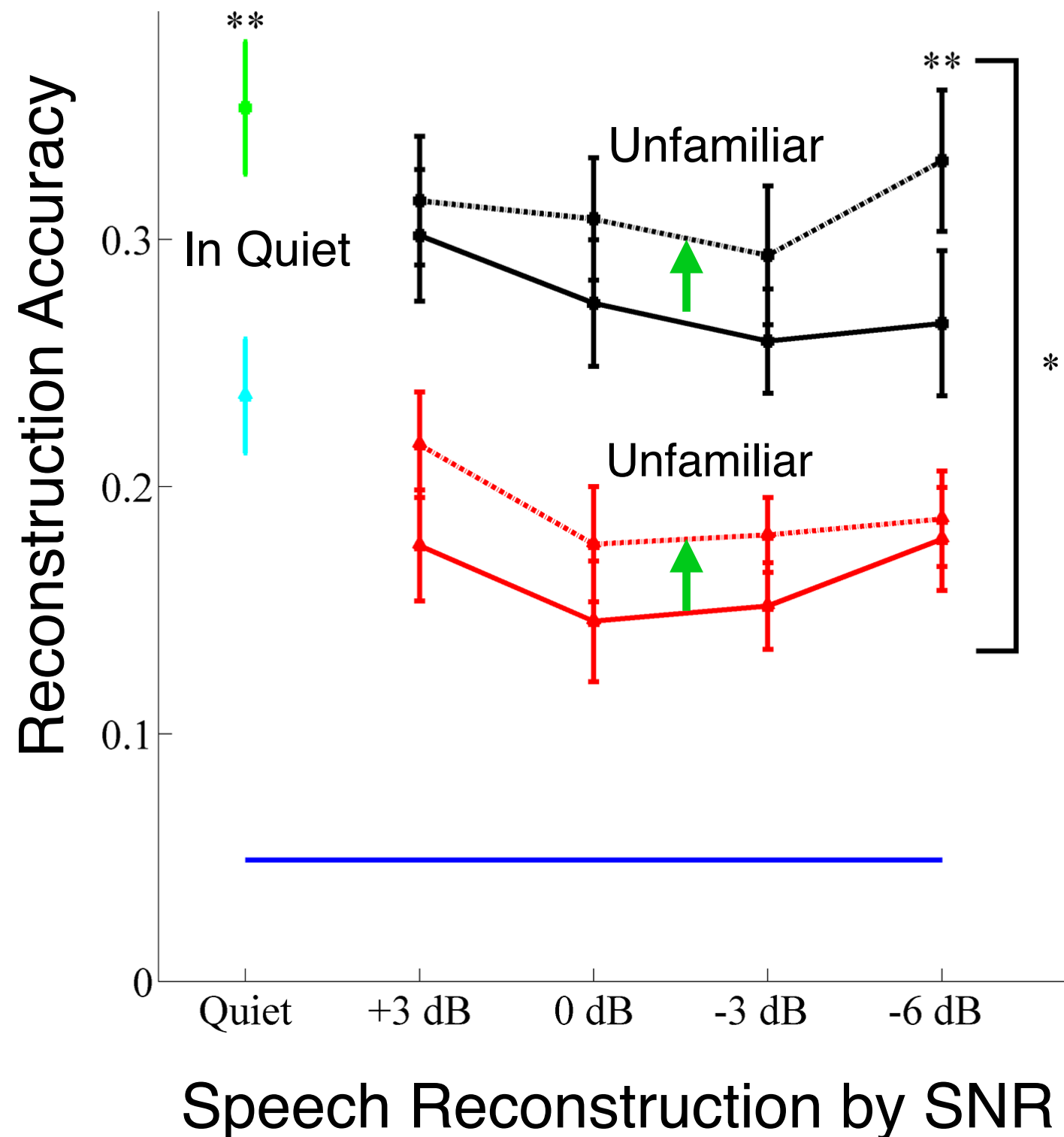
- **Unfamiliarity of Background**
 - Boosts Intelligibility of Attended Speech

High Level Interference Effects



- **Unfamiliarity of Background**
 - Boosts Intelligibility of Attended Speech

High Level Interference Effects



- **Unfamiliarity of Background**
 - Boosts Intelligibility of Attended Speech
 - *Also Boosts Cortical Reconstruction of Attended Speech*

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)
- Robust Dynamical Foreground Monitoring
- Over-Representation with Aging
 - Reconstruction depends on integration time
 - Over-Representation tracks inhibitory control
- Background familiarity: neural tracks behavior

Thank You