Neural Representations of Speech in Human Auditory Cortex

Jonathan Z. Simon

Department of Electrical & Computer Engineering

Department of Biology

Institute for Systems Research

University of Maryland

Acknowledgements

Current Lab Members & Affiliates

Christian Brodbeck
Joshua Kulasingham
David Nahmias
Krishna Puvvada
Peng Zan

Past Lab Members & Affiliates

Nayef Ahmar
Sahar Akram
Murat Aytekin
Francisco Cervantes
Maria Chait
Marisel Villafane Delgado
Kim Drnec
Nai Ding
Victor Grau-Serrat
Julian Jenkins
Pirazh Khorramshahi
Natalia Lapinskaya
Huan Luo
Aline Gesualdi Manhães
Mahshid Najafi
Alex Presacco
Raul Rodriguez
Jonas Vanthornhout
Ben Walsh
Yadong Wang
Juanjuan Xiang
Jiachen Zhuo

Collaborators

Pamela Abshire
Samira Anderson
Behtash Babadi
Catherine Carr
Monita Chatterjee
Alain de Cheveigné
Stephen David
Didier Depireux
Mounya Elhilali
Jonathan Fritz
Stefanie Kuchinsky
Cindy Moss
David Poeppel
Shihab Shamma

Past Undergraduate Students

Abdulaziz Al-Turki
Nicholas Asendorf
Anurupa Bhonsale
Sonja Bohr
Elizabeth Camenga
Corinne Cameron
Julien Dagenais
Katya Dombrowski
Kevin Hogan
Kevin Kahn
Alexandria Miller
Andrea Shome
Madeleine Varmer
James Williams

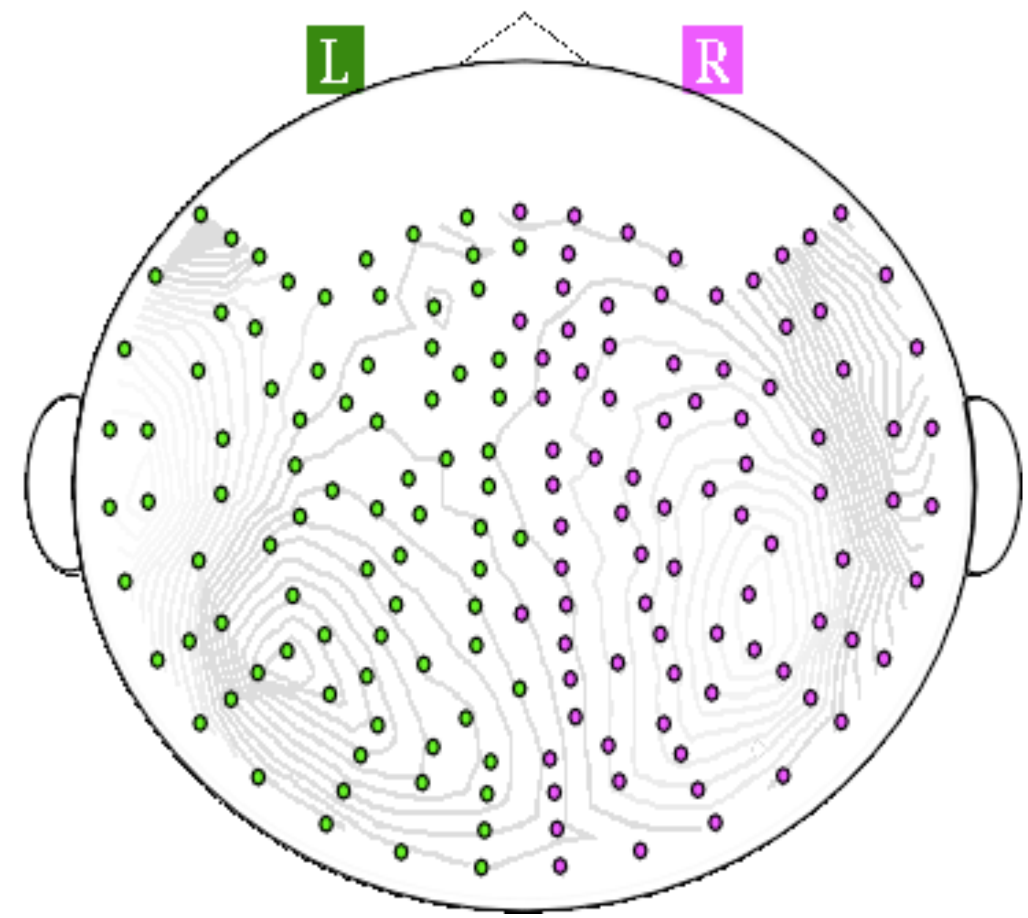
Funding NIH (*NIDCD*, *NIA*, *NIBIB*); NSF; DARPA; USDA; UMD

Outline

- Cortical Representations of Speech (via MEG)
  - Encoding vs. Decoding
- "Cocktail Party" Speech
- Recent Results
  - 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

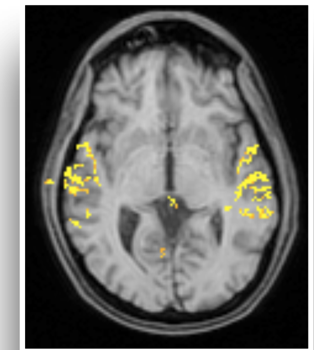


Functional Brain Imaging

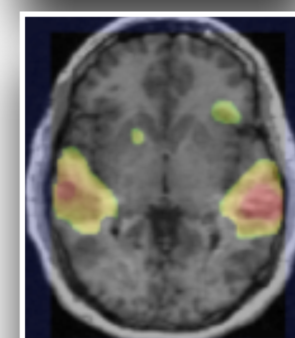
Functional Brain Imaging
= Non-invasive recording from human brain

Hemodynamic techniques

fMRI
functional magnetic resonance imaging



PET
positron emission tomography



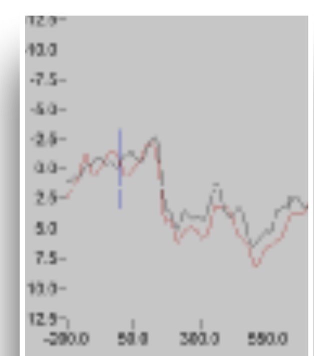
Excellent Spatial Resolution **(~1 mm)**

Poor Temporal Resolution **(~1 s)**

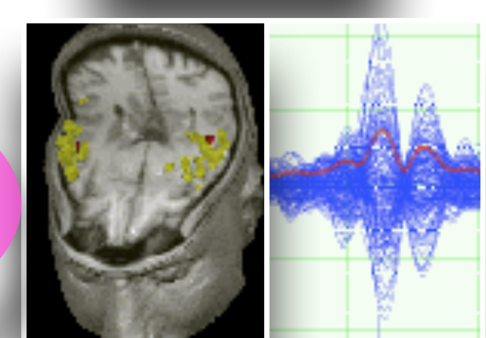
fMRI & MEG can capture effects in single subjects

Electromagnetic techniques

EEG
electroencephalography



MEG
magnetoencephalography



Poor Spatial Resolution **(~1 cm)**

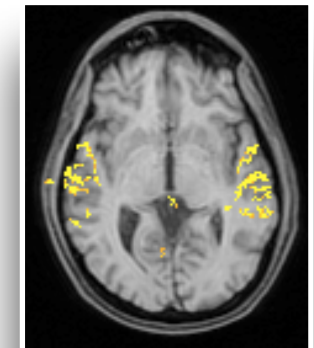
Excellent Temporal Resolution **(~1 ms)**

Functional Brain Imaging

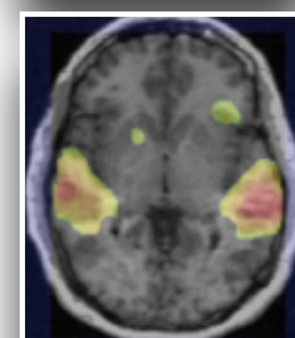
Functional Brain Imaging
= Non-invasive recording from human brain

Hemodynamic techniques

fMRI
functional magnetic resonance imaging



PET
positron emission tomography



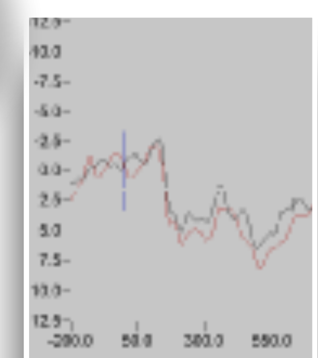
Excellent Spatial Resolution **(~1 mm)**

Poor Temporal Resolution **(~1 s)**

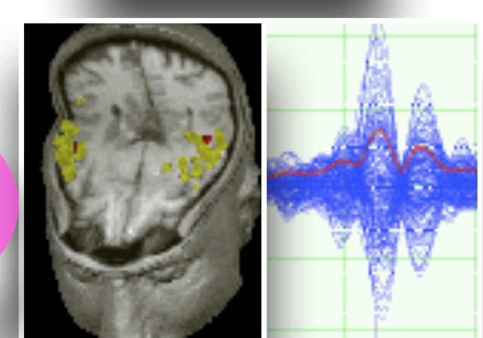
fMRI & MEG can capture effects in single subjects

Electromagnetic techniques

EEG
electroencephalography



MEG
magnetoencephalography



Poor Spatial Resolution **(~1 cm)**

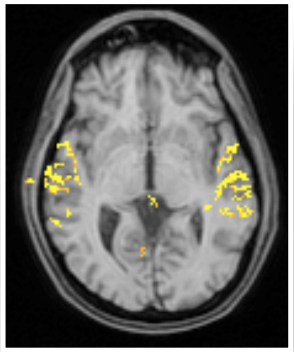
Excellent Temporal Resolution **(~1 ms)**

Functional Brain Imaging

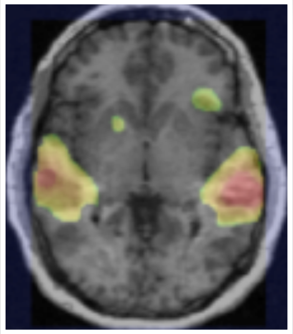
Functional Brain Imaging
= Non-invasive recording from human brain

Hemodynamic techniques

fMRI
functional magnetic resonance imaging



PET
positron emission tomography



Excellent Spatial Resolution **(~1 mm)**

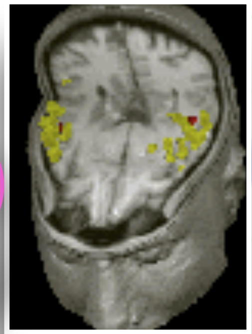
Poor Temporal Resolution **(~1 s)**

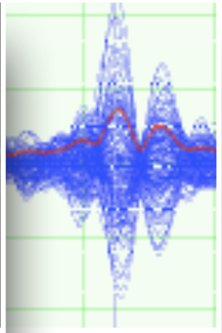
fMRI & MEG can capture effects in single subjects

Electromagnetic techniques

EEG
electroencephalography

MEG
magnetoencephalography





Poor Spatial Resolution **(~1 cm)**

Excellent Temporal Resolution **(~1 ms)**

Neural Signals & MEG

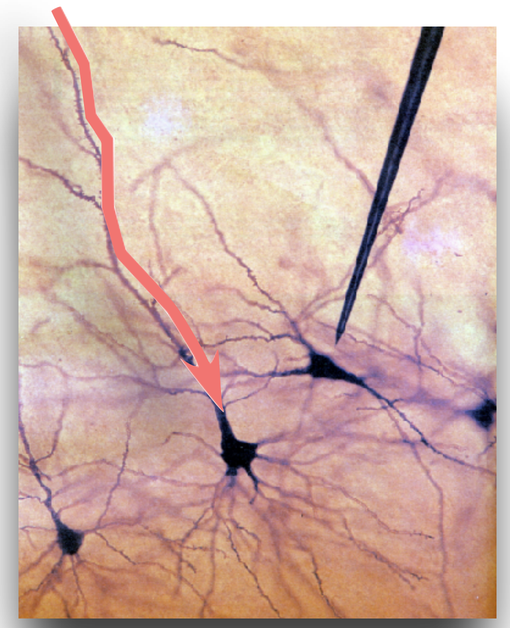
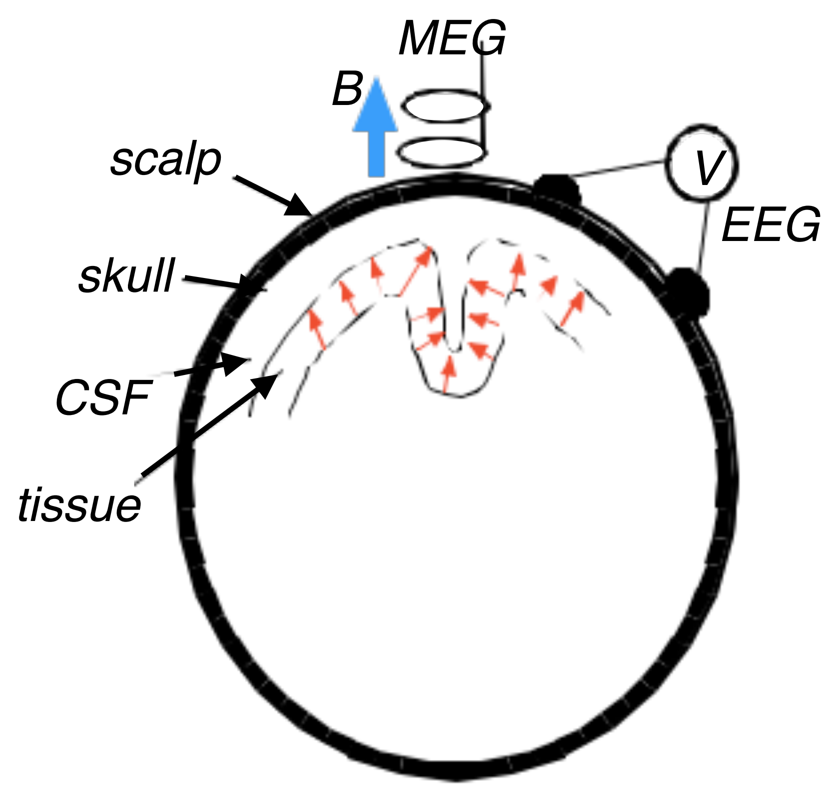
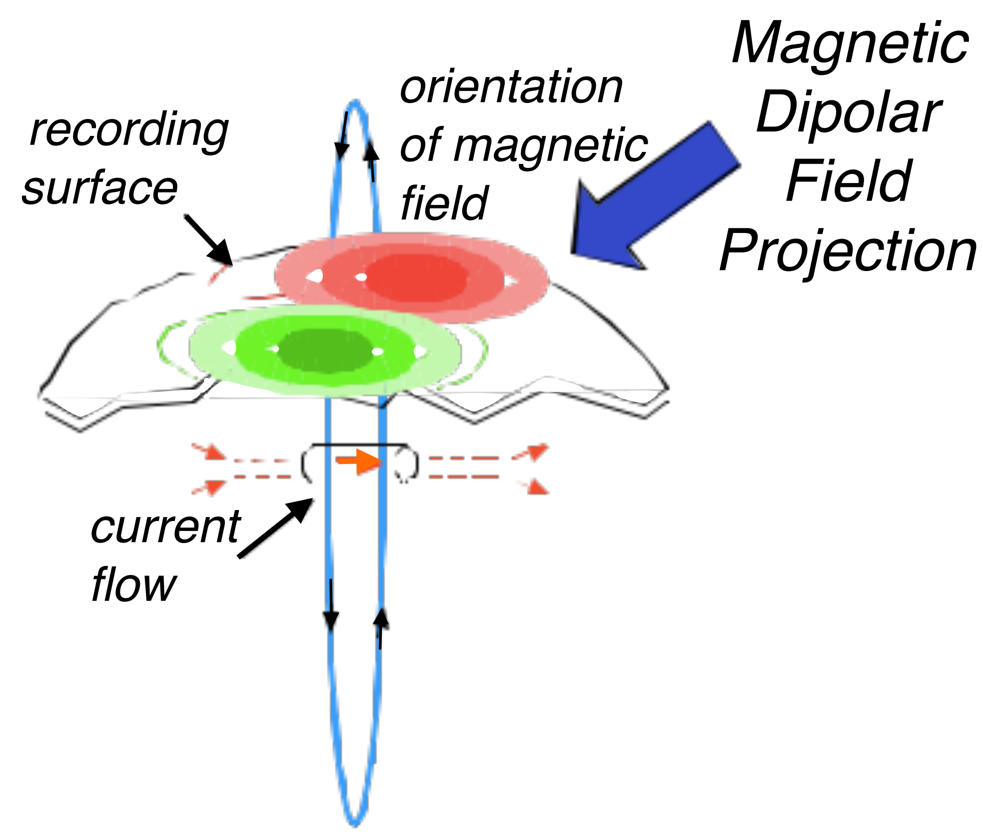


Photo by Fritz Goro



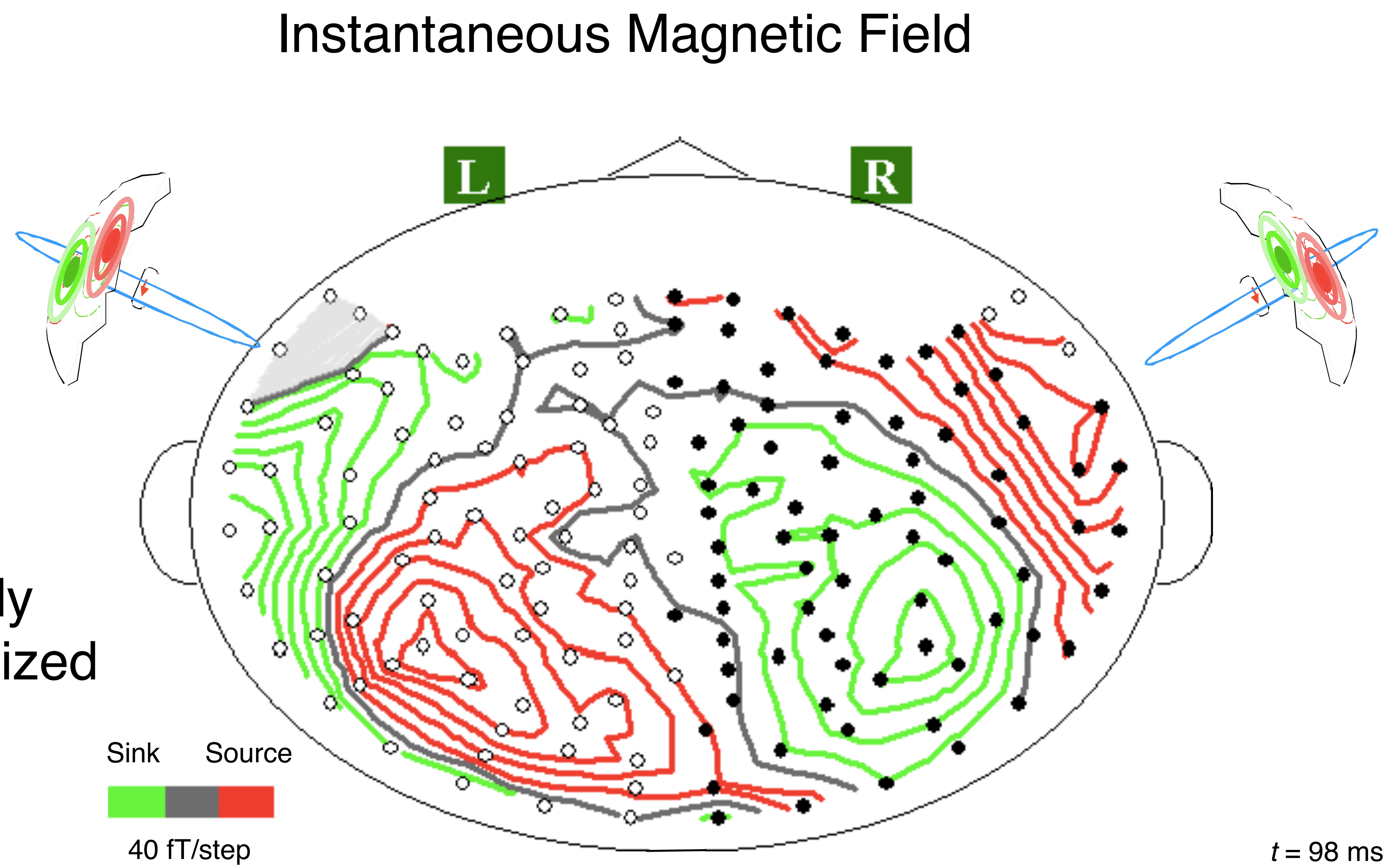


- Direct electrophysiological measurement
  - not hemodynamic
  - real-time
- No unique solution for distributed source

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

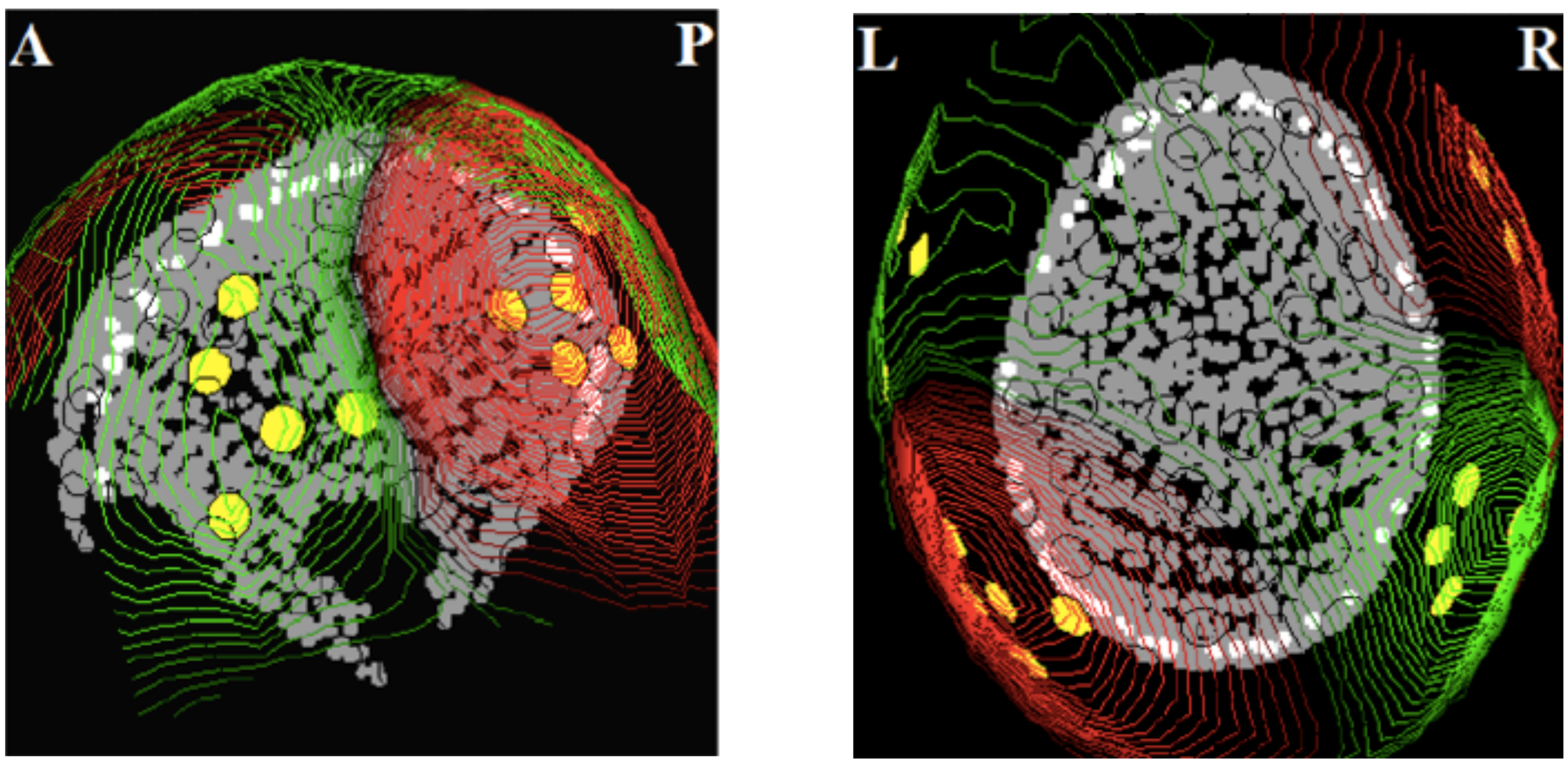
MEG Auditory Field

Flattened Isofield Contour Map



Strongly Lateralized

MEG Auditory Field

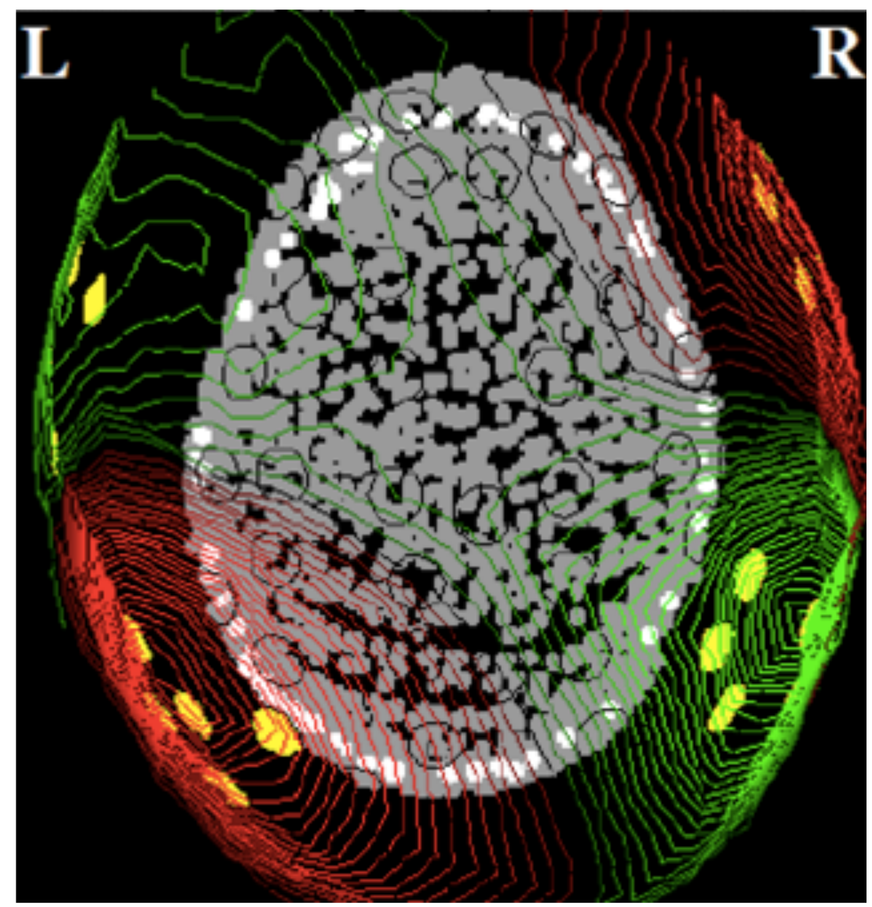


Sagittal View   Axial View

Strongly
Lateralized

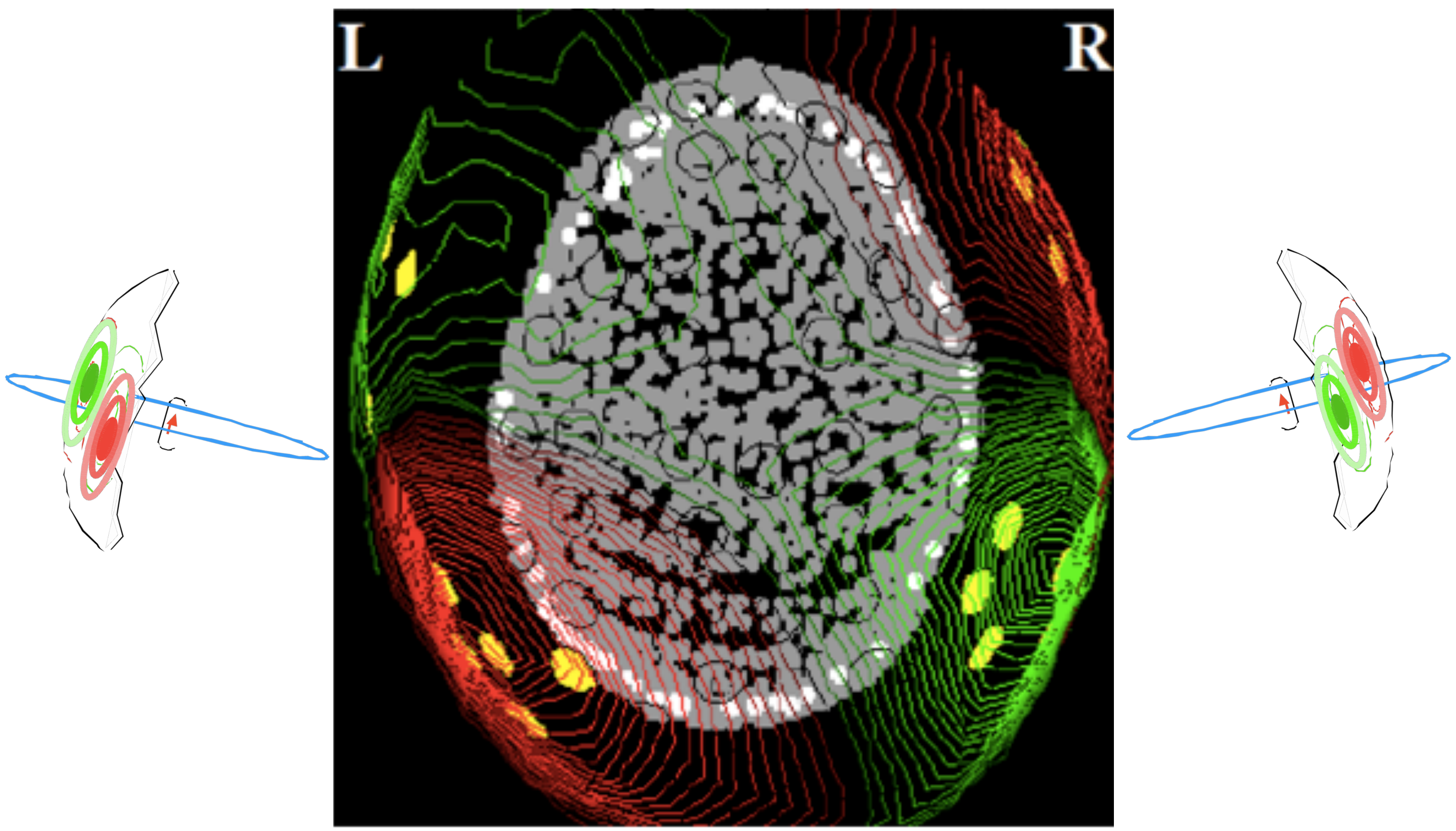
Chait, Poeppel and Simon, Cerebral Cortex (2006)

MEG Auditory Field



Chait et al., Cerebral Cortex (2006)

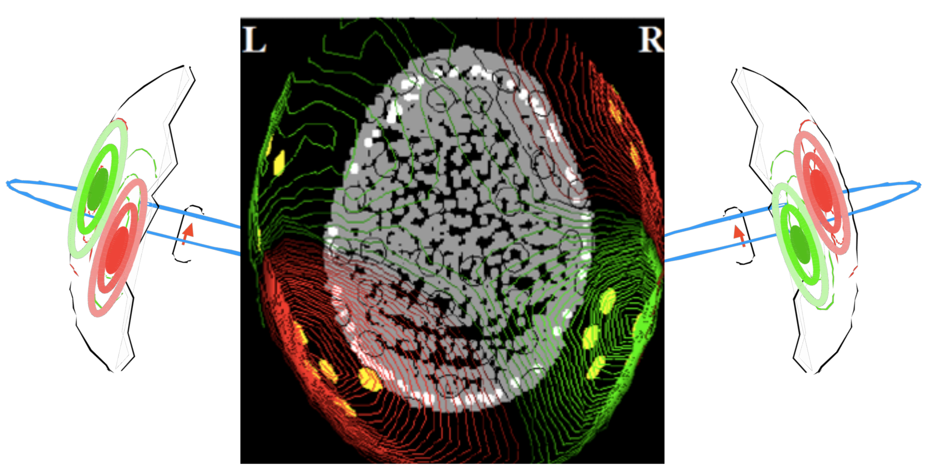
MEG Auditory Field

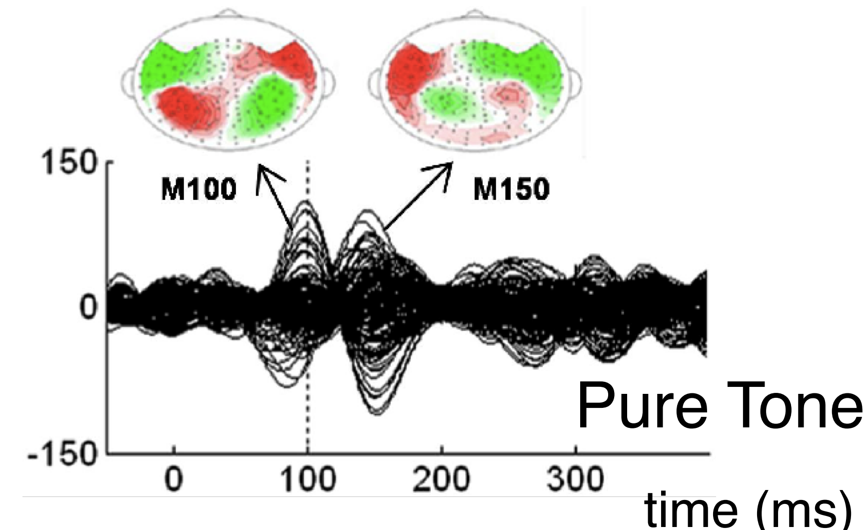


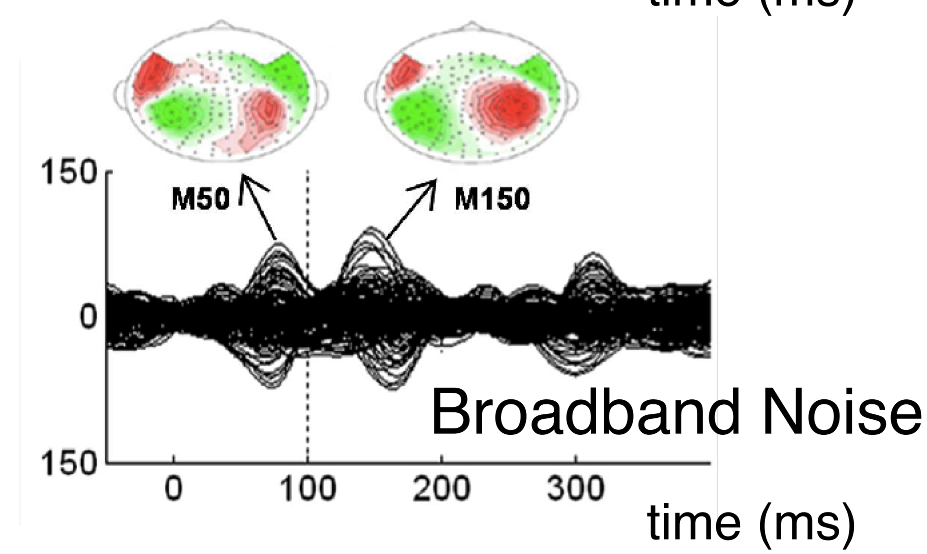
Chait et al., Cerebral Cortex (2006)

MEG & Auditory Cortex

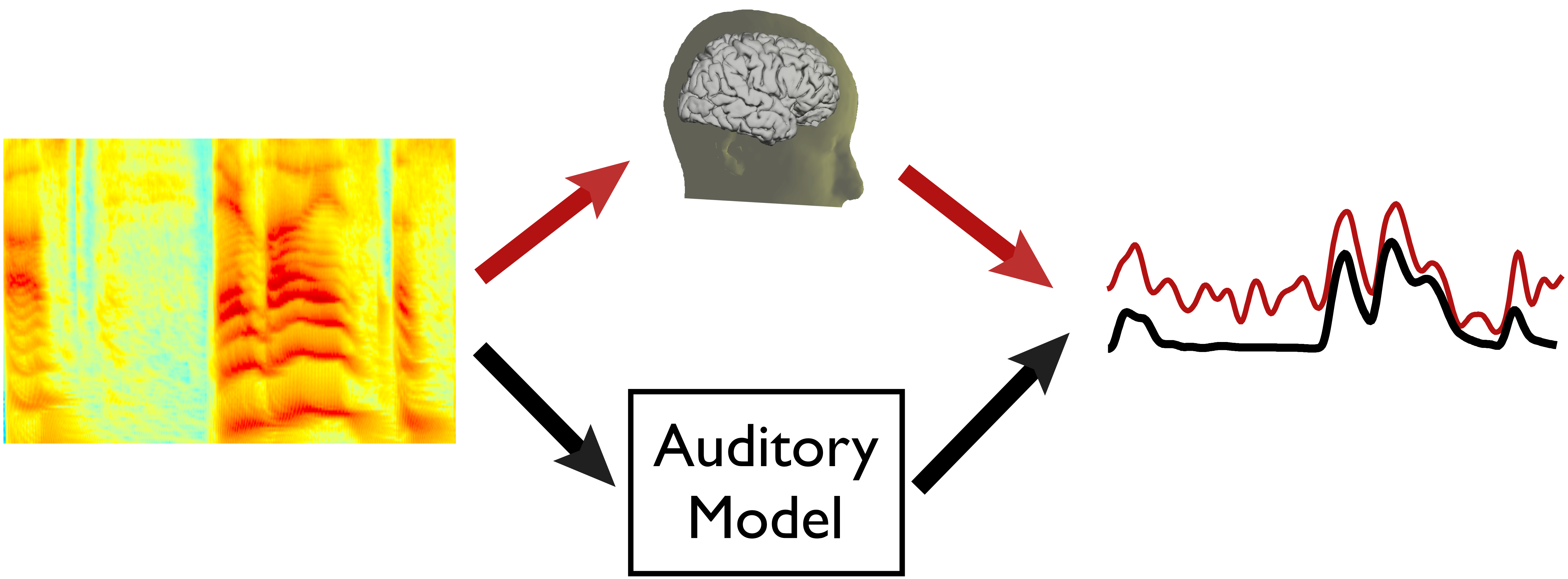
- Non-invasive, Passive, Silent Neural Recordings
- MEG Response Patterns Time-Locked to Stimulus Events
- Robust
- Strongly Lateralized
- Cortical Origin Only



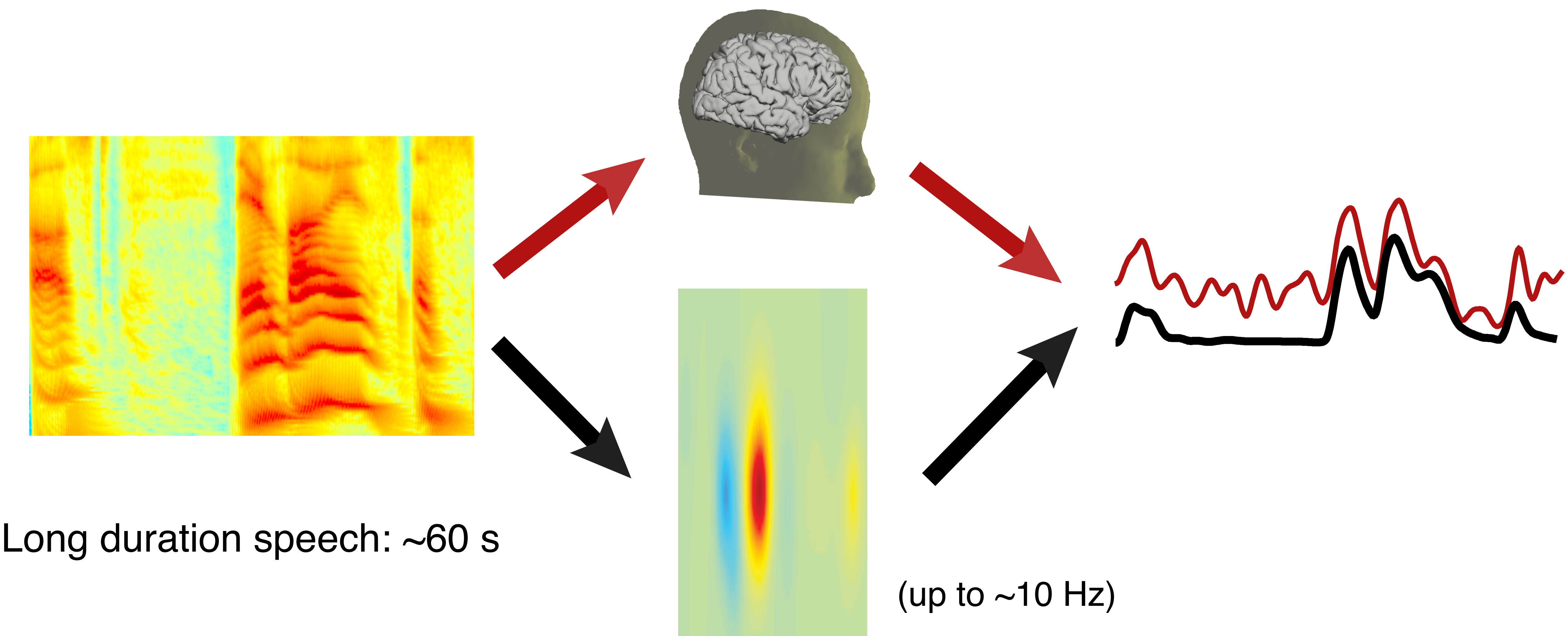




MEG Responses to Speech Modulations



MEG Responses Predicted by STRF Model

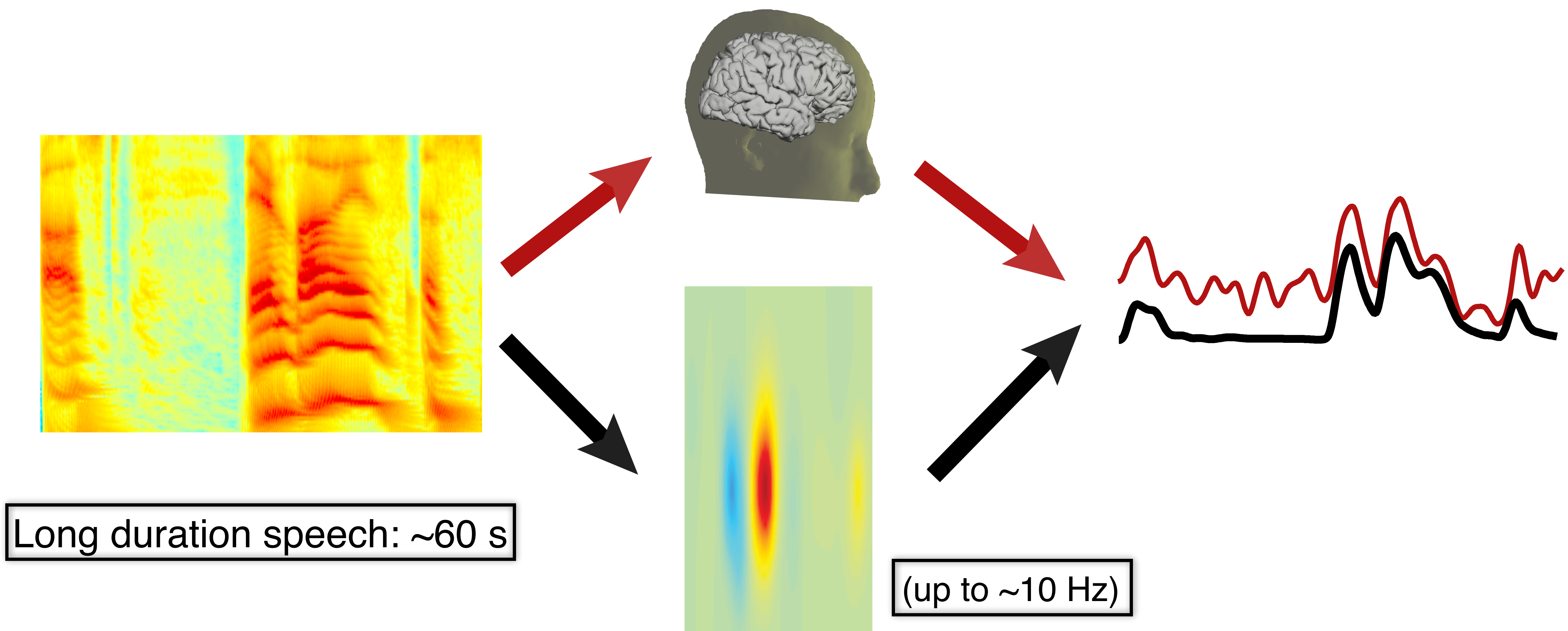


Linear Kernel = STRF

Ding & Simon, J Neurophysiol (2012)

"Spectro-Temporal Response Function"

MEG Responses Predicted by STRF Model

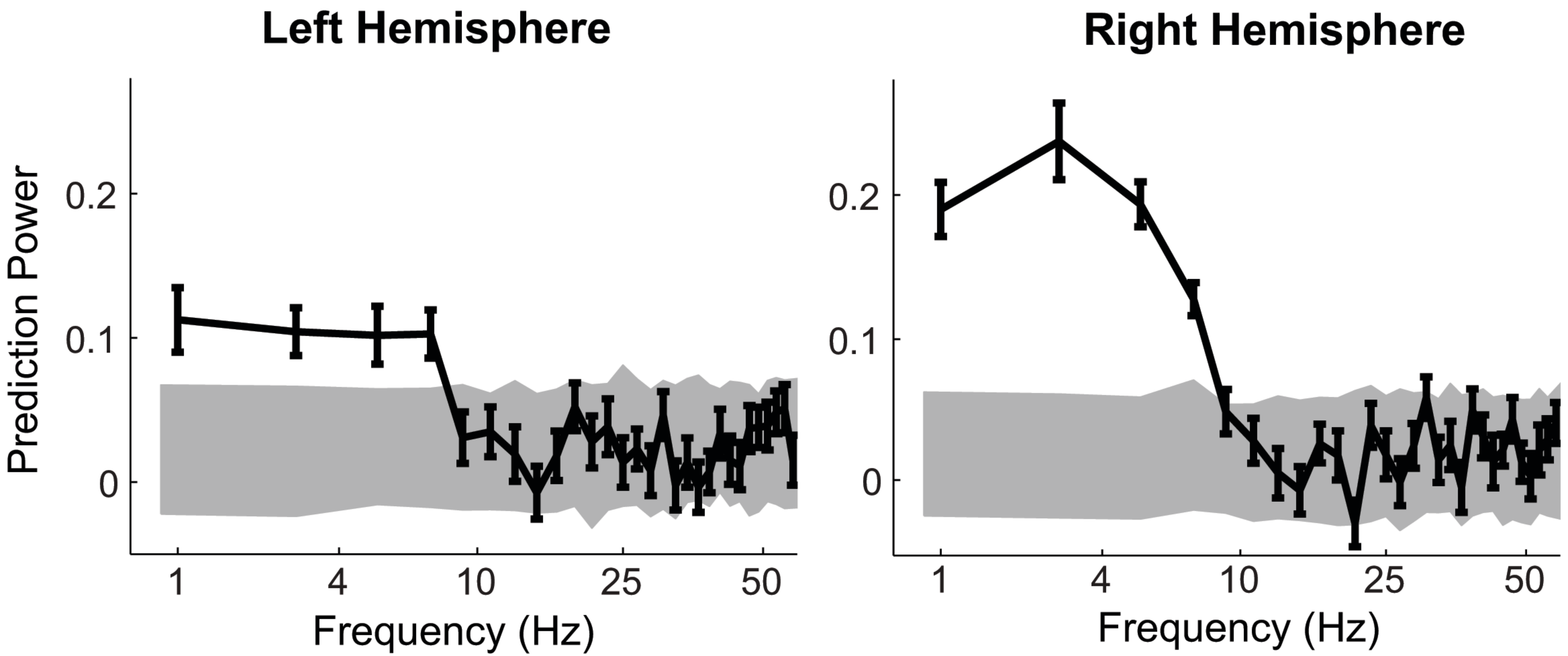


Linear Kernel = STRF

"Spectro-Temporal Response Function"

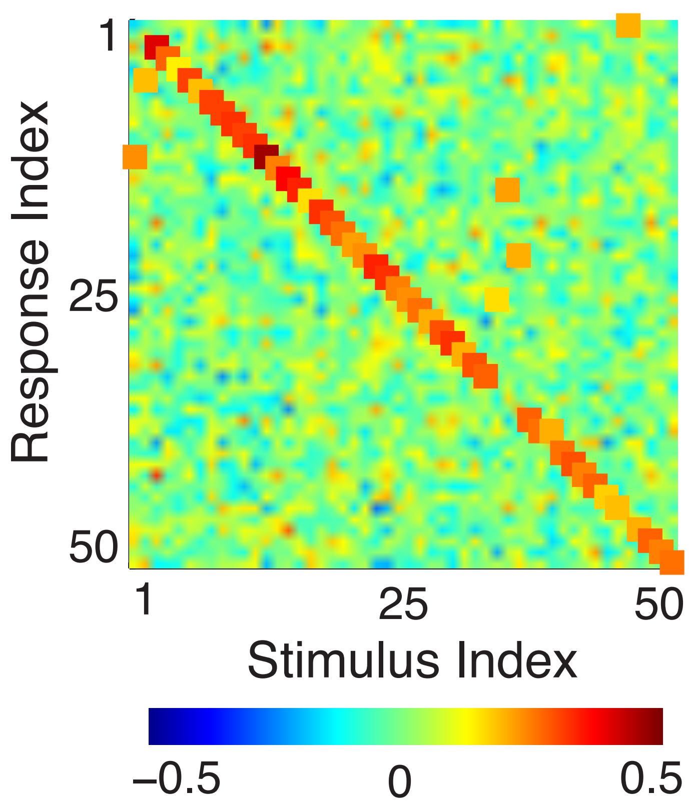
Ding & Simon, J Neurophysiol (2012)

Frequency Dependence of STRF Predictability

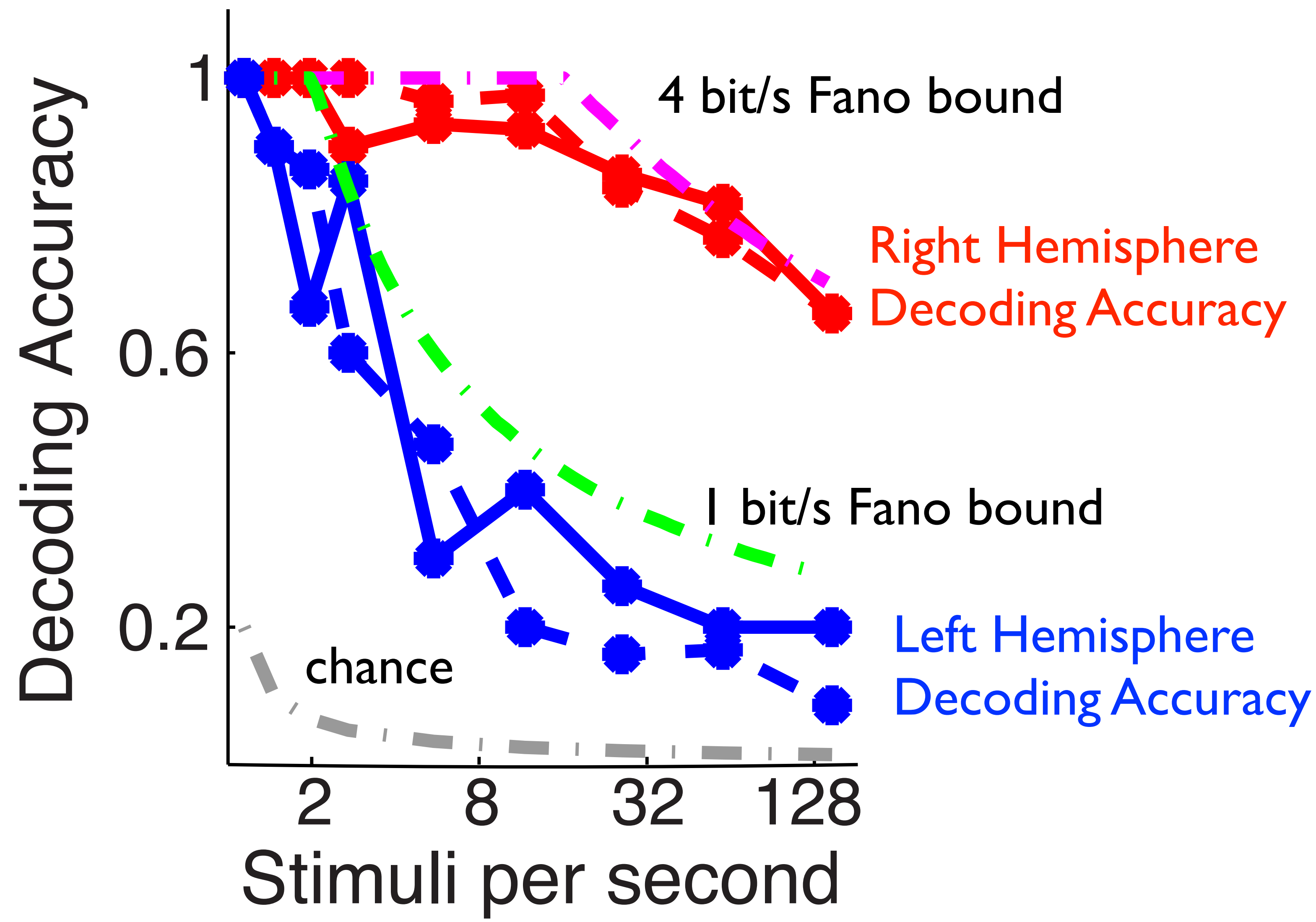


Ding & Simon, J Neurophysiol (2012)

Stimulus Information Encoded in Response

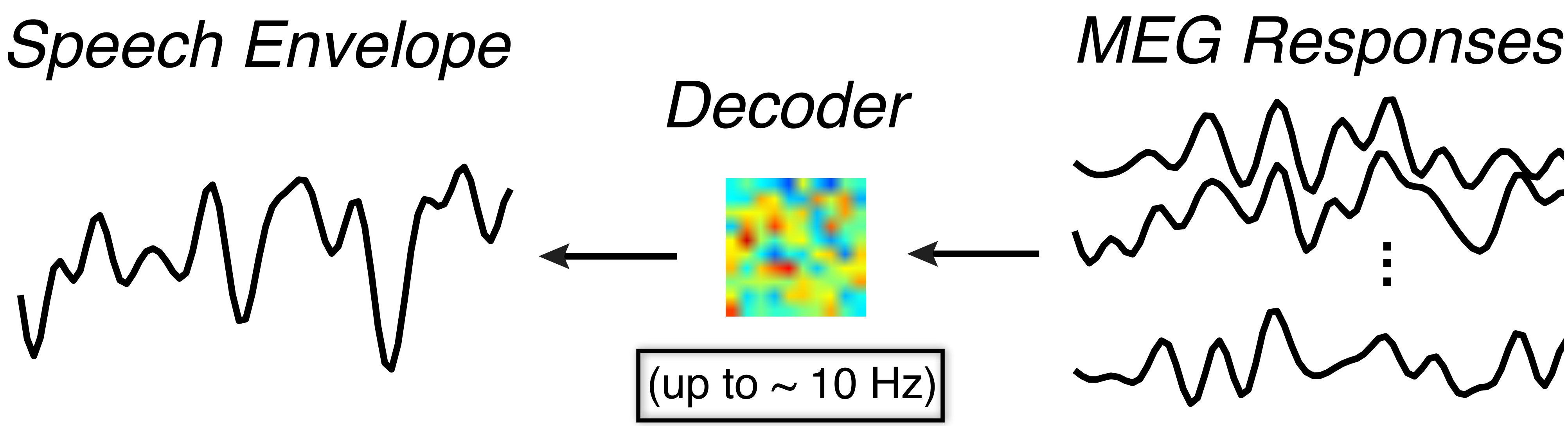


Correlation between stimulus envelope and reconstructed envelope (right hemisphere)

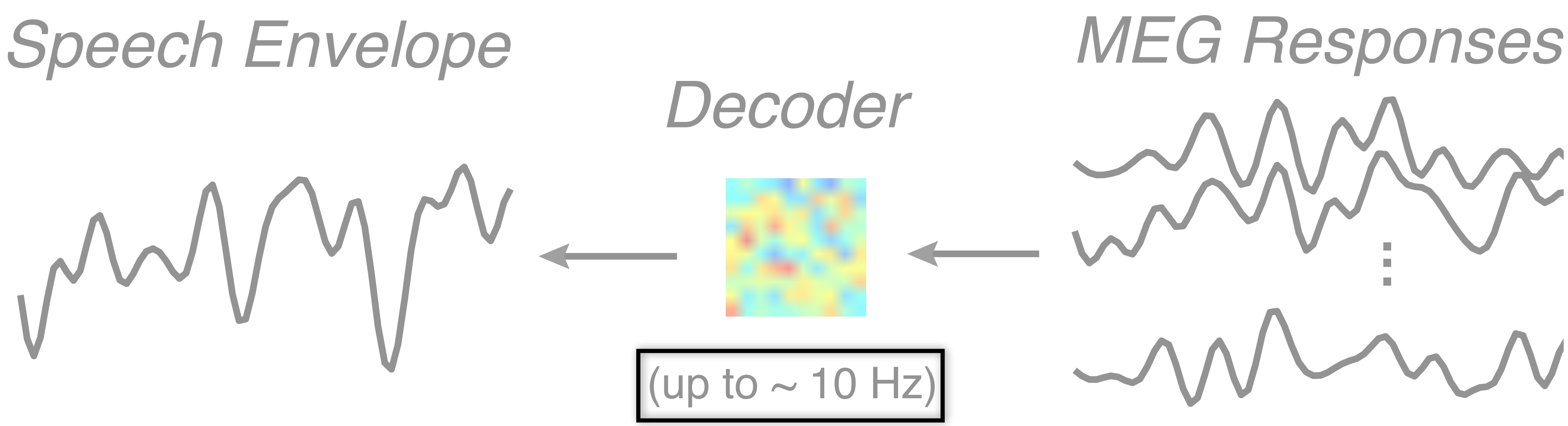


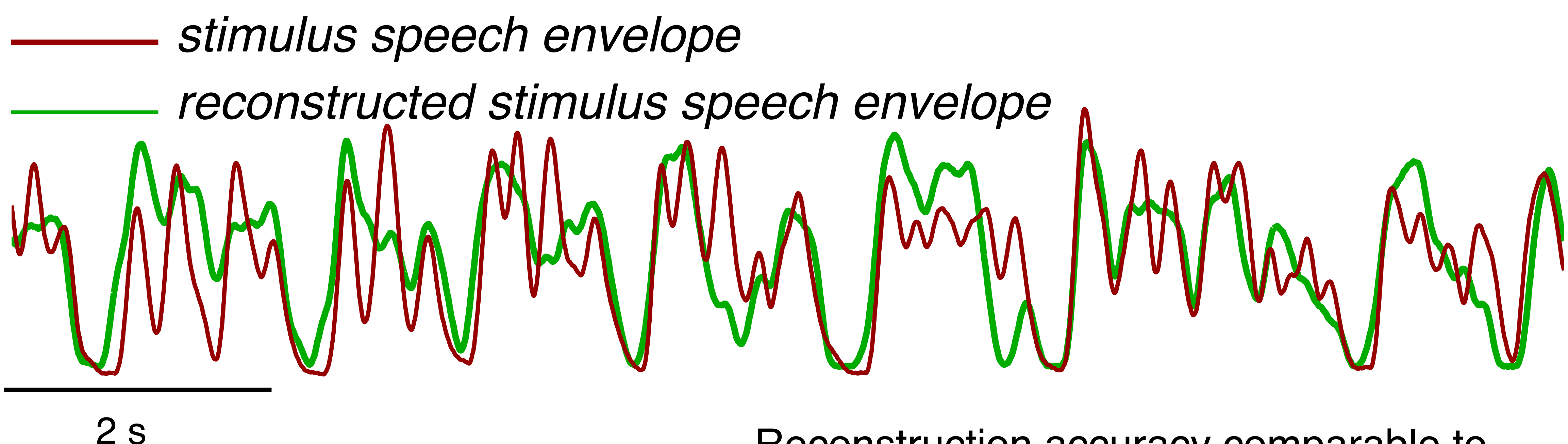
Ding & Simon, J Neurophysiol (2012)

Neural Reconstruction of Speech Envelope



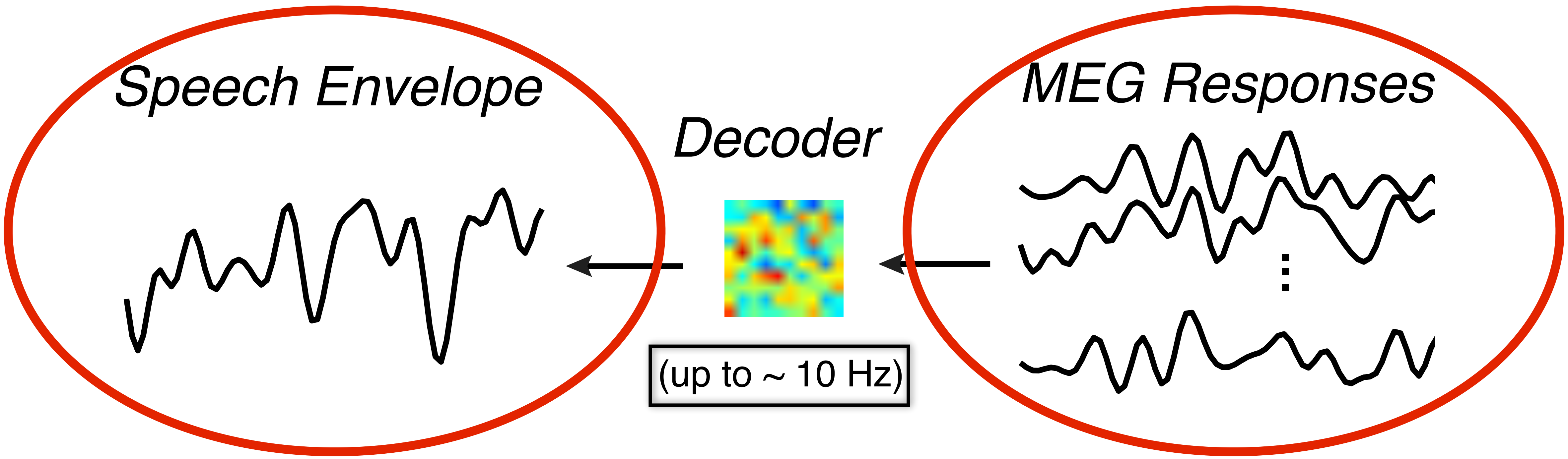
Neural Reconstruction of Speech Envelope



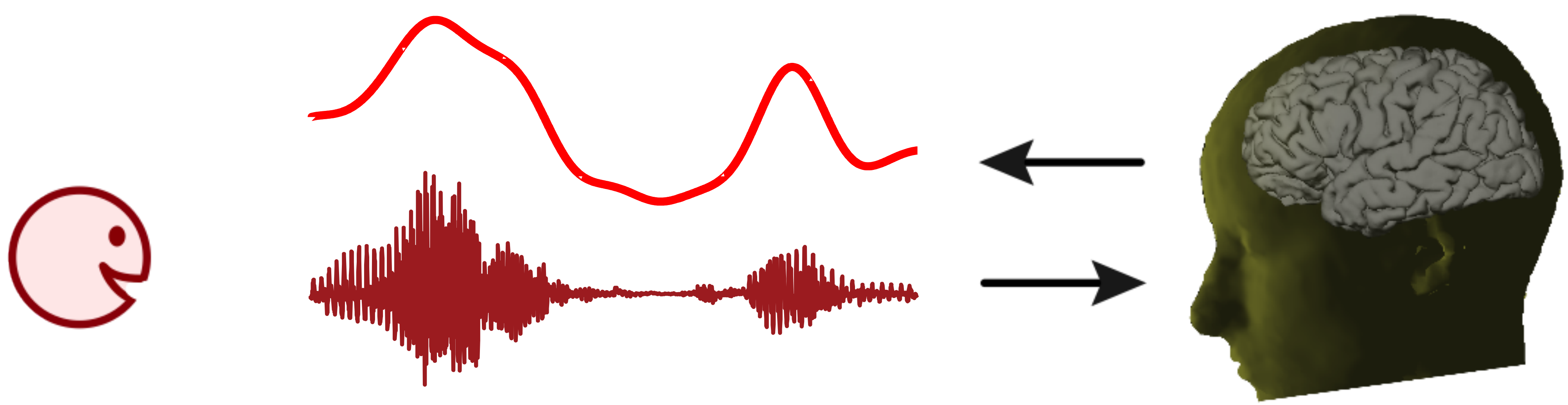


Reconstruction accuracy comparable to single unit & ECoG recordings

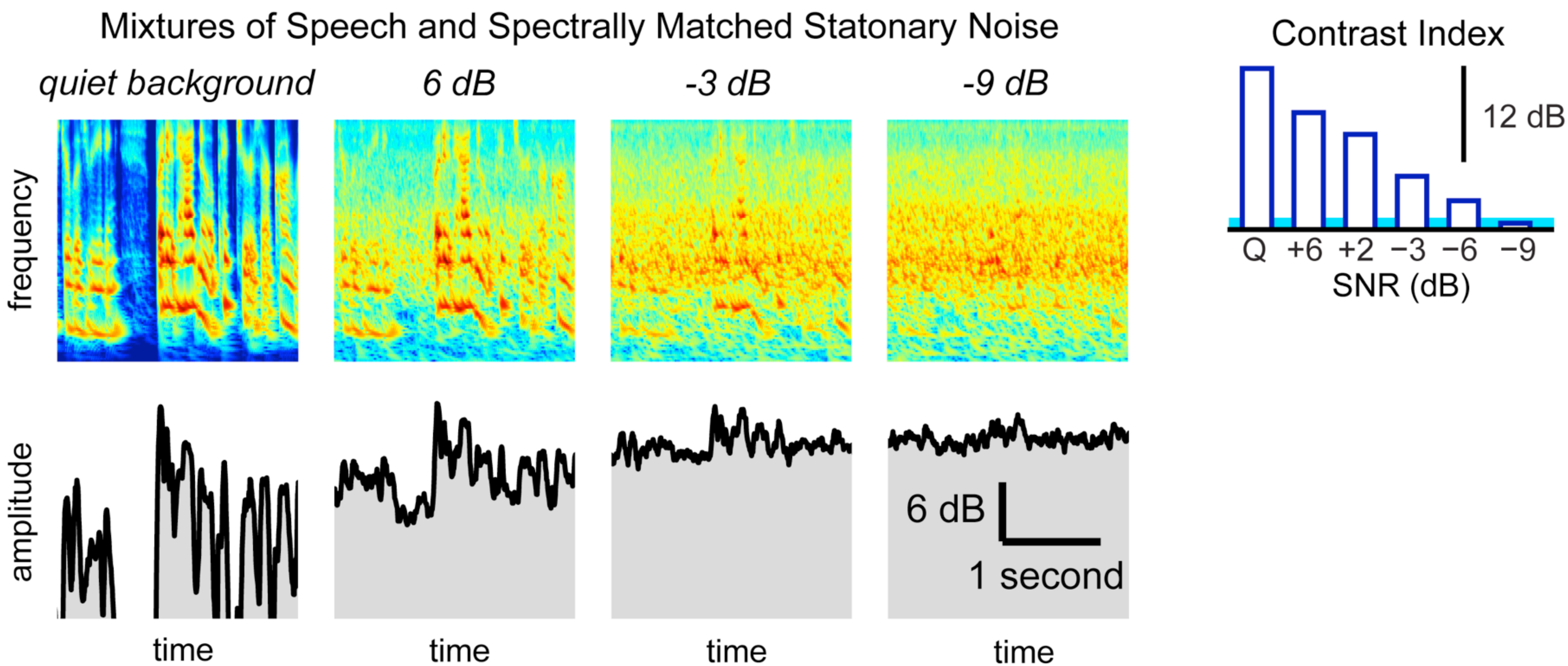
Ding & Simon, J Neurophysiol (2012)
Zion-Golumbic et al., Neuron (2013)


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

Neural Representation of Speech: Temporal

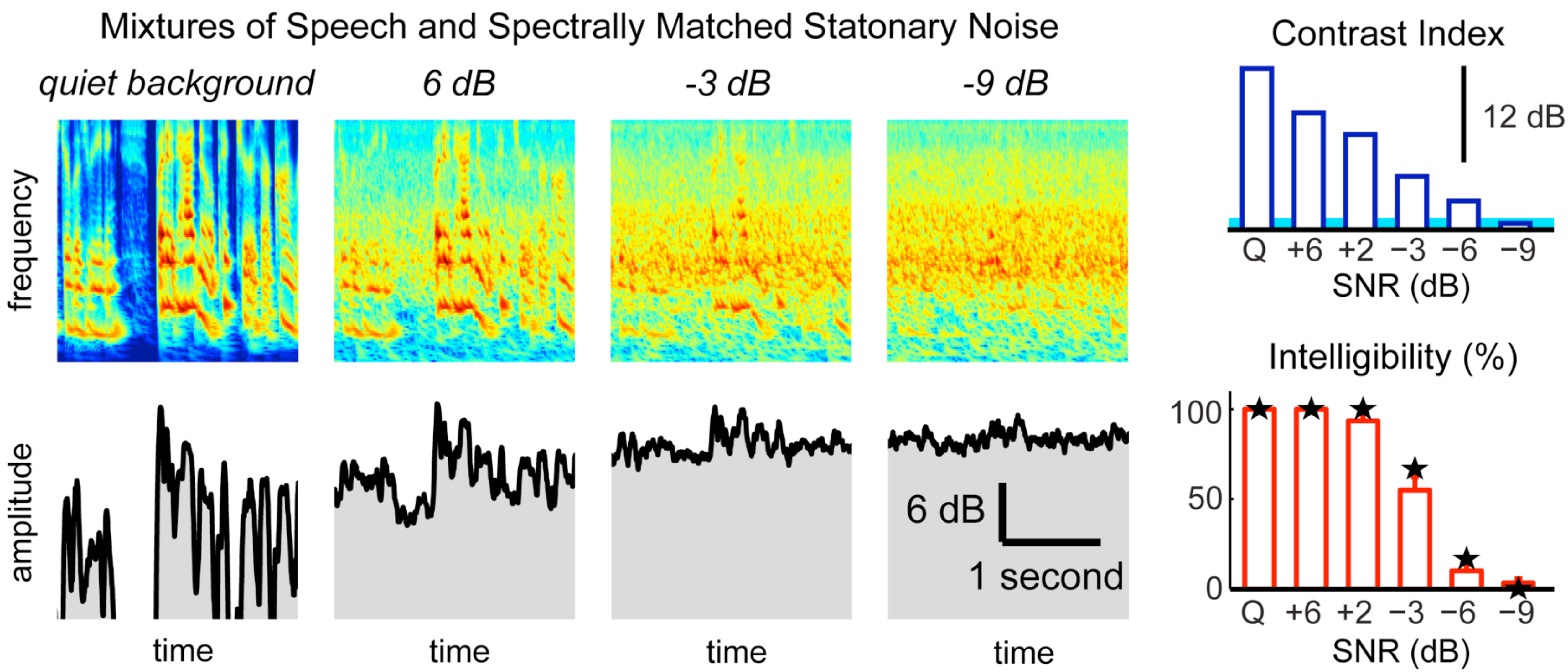


Speech in Stationary Noise



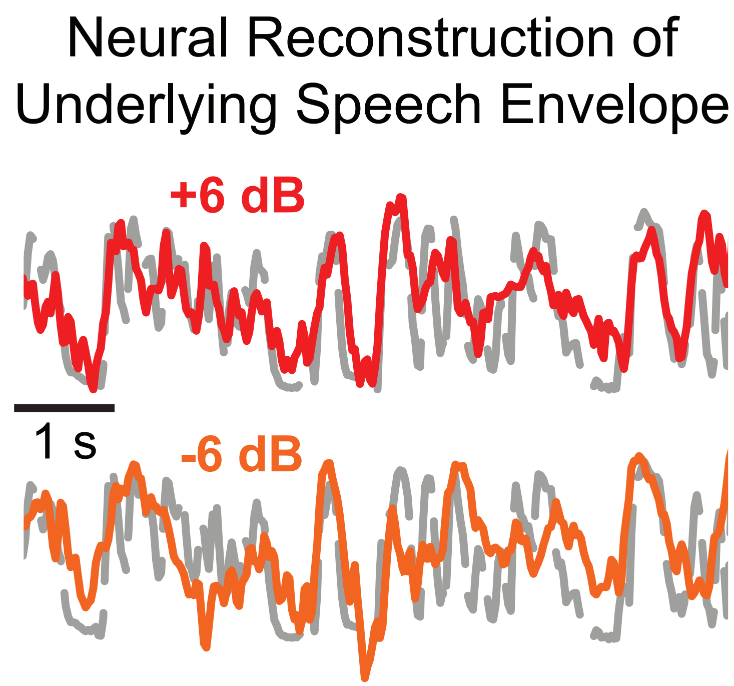
Ding & Simon, J Neuroscience (2013)

Speech in Stationary Noise



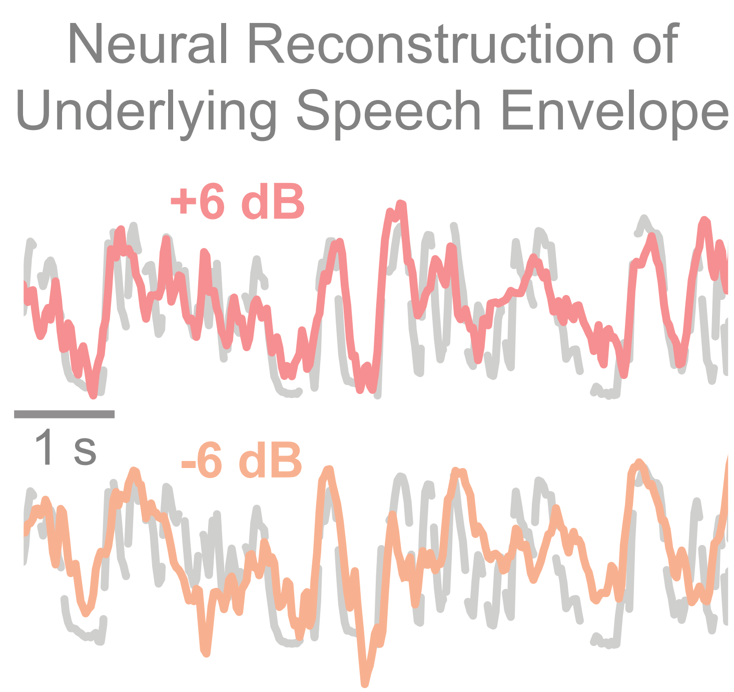
Ding & Simon, J Neuroscience (2013)

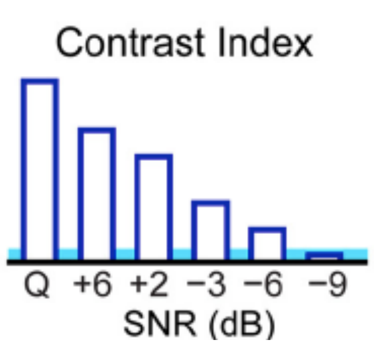
Speech in Noise: Results

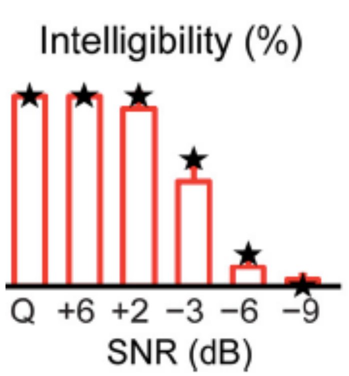


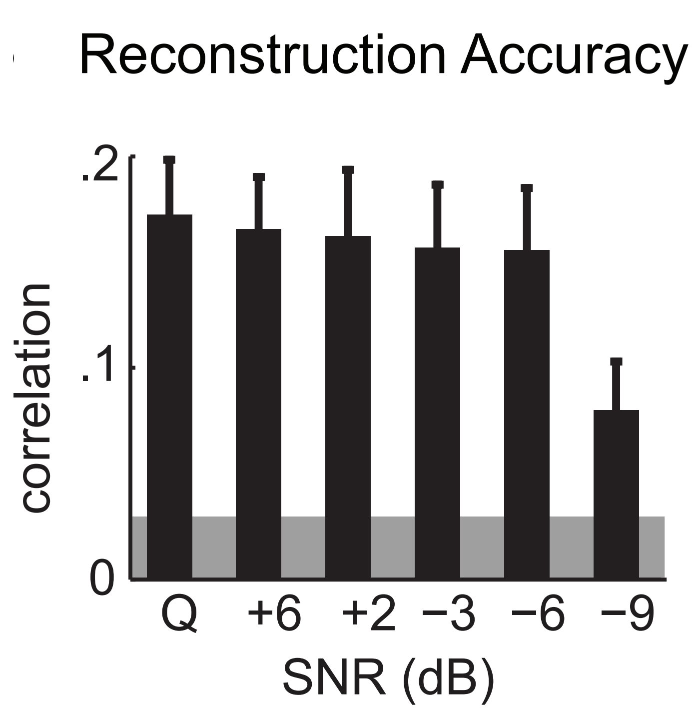
Ding & Simon, J Neuroscience (2013)

Speech in Noise: Results



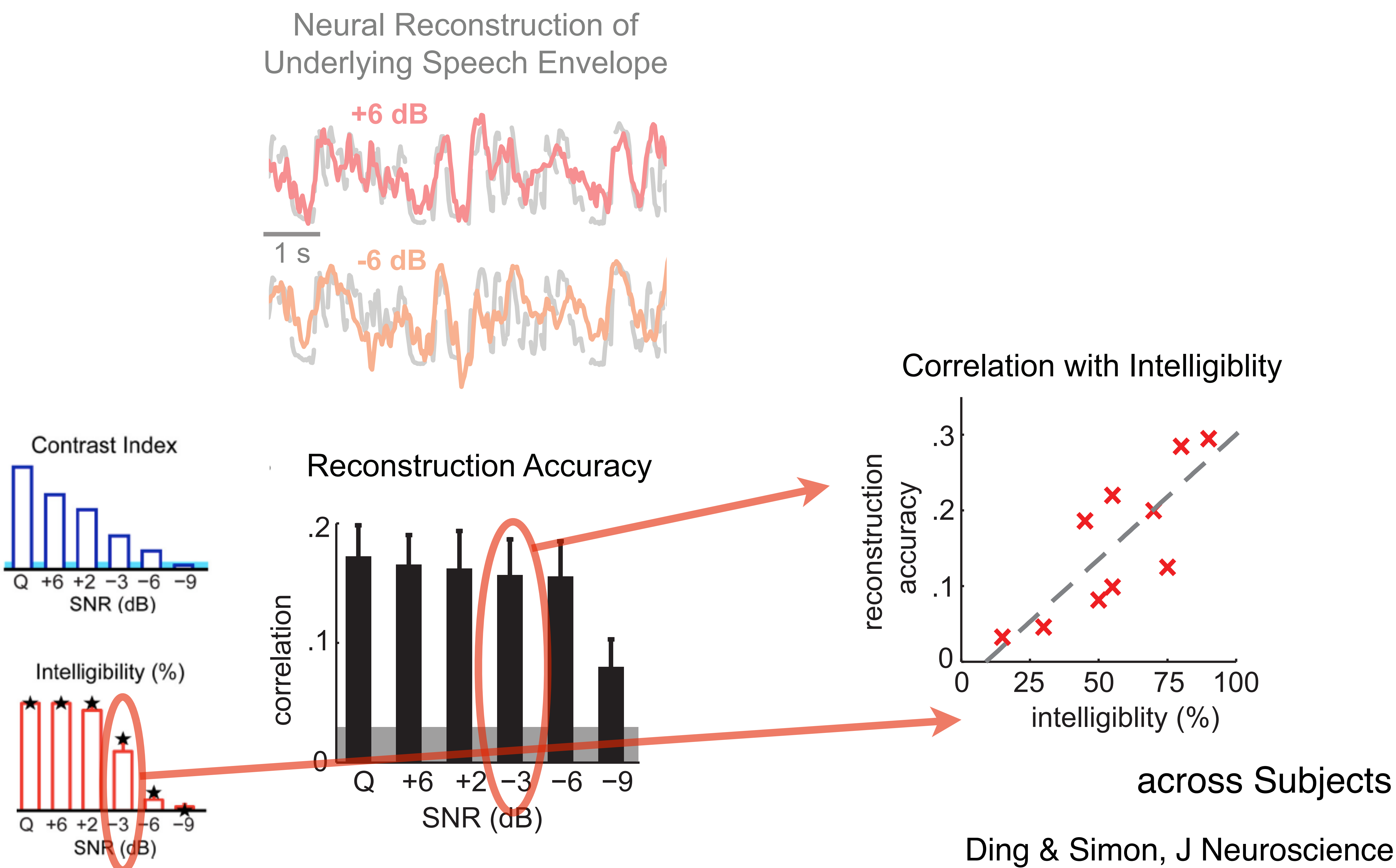






Ding & Simon, J Neuroscience (2013)

Speech in Noise: Results



Ding & Simon, J Neuroscience (2013)

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: ~ 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

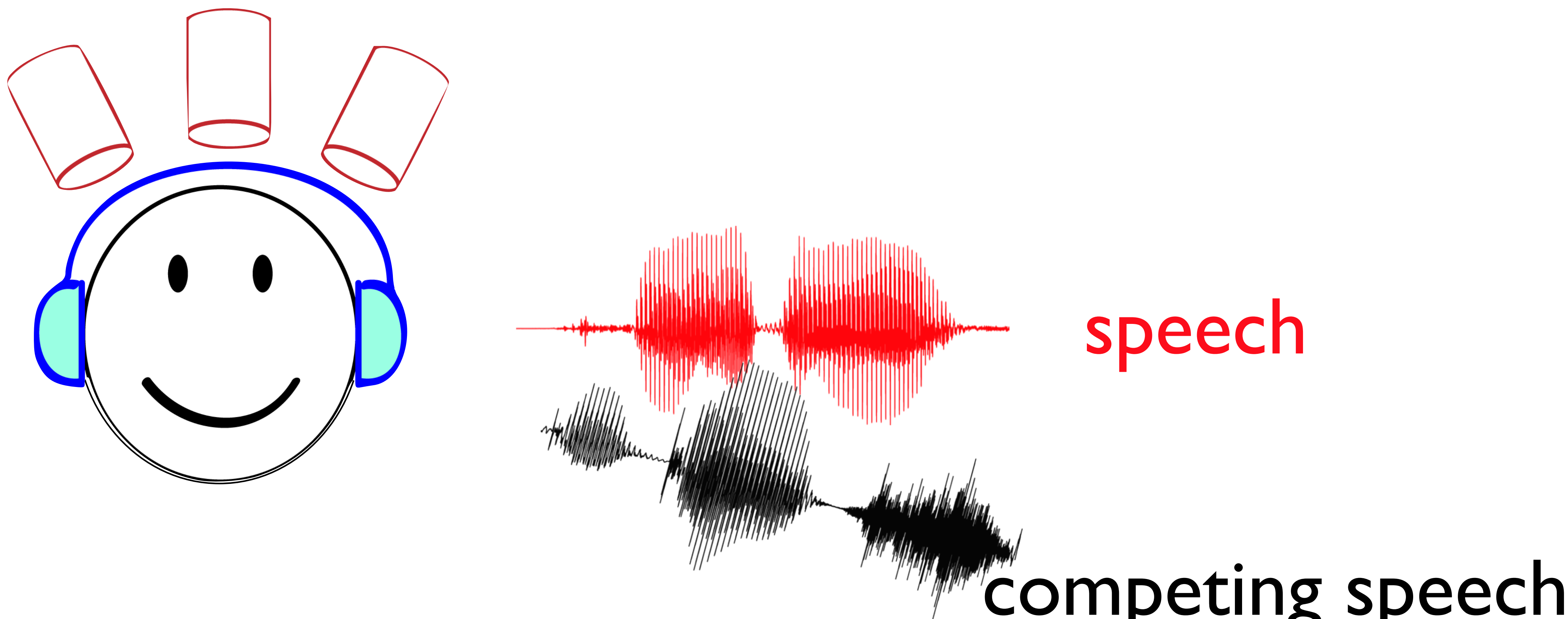
Listening to Speech at the Cocktail Party

Alex Katz,
The Cocktail Party

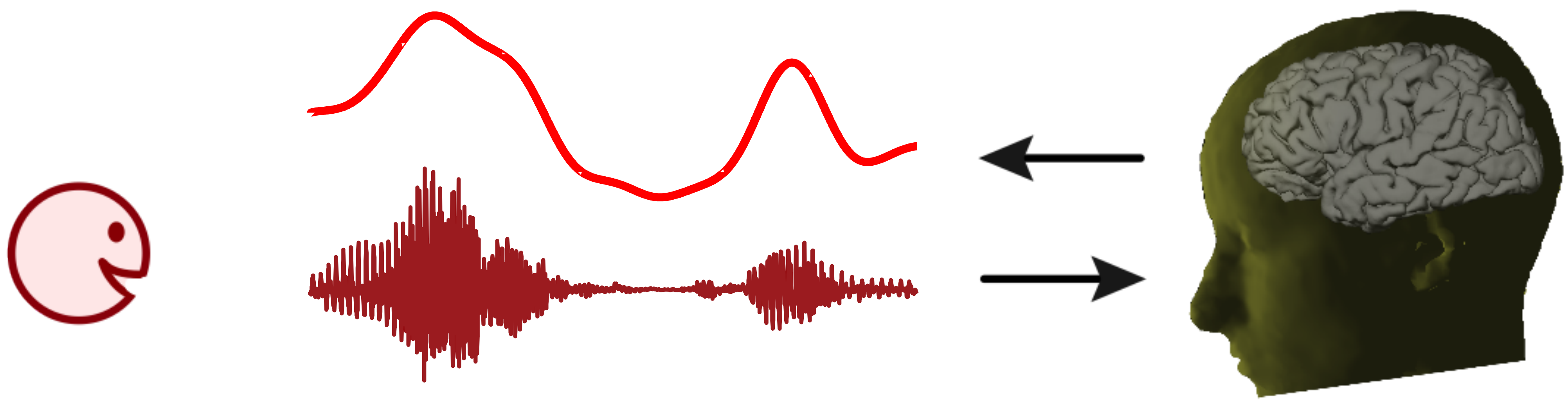
Listening to Speech at the Cocktail Party

Alex Katz,
The Cocktail Party

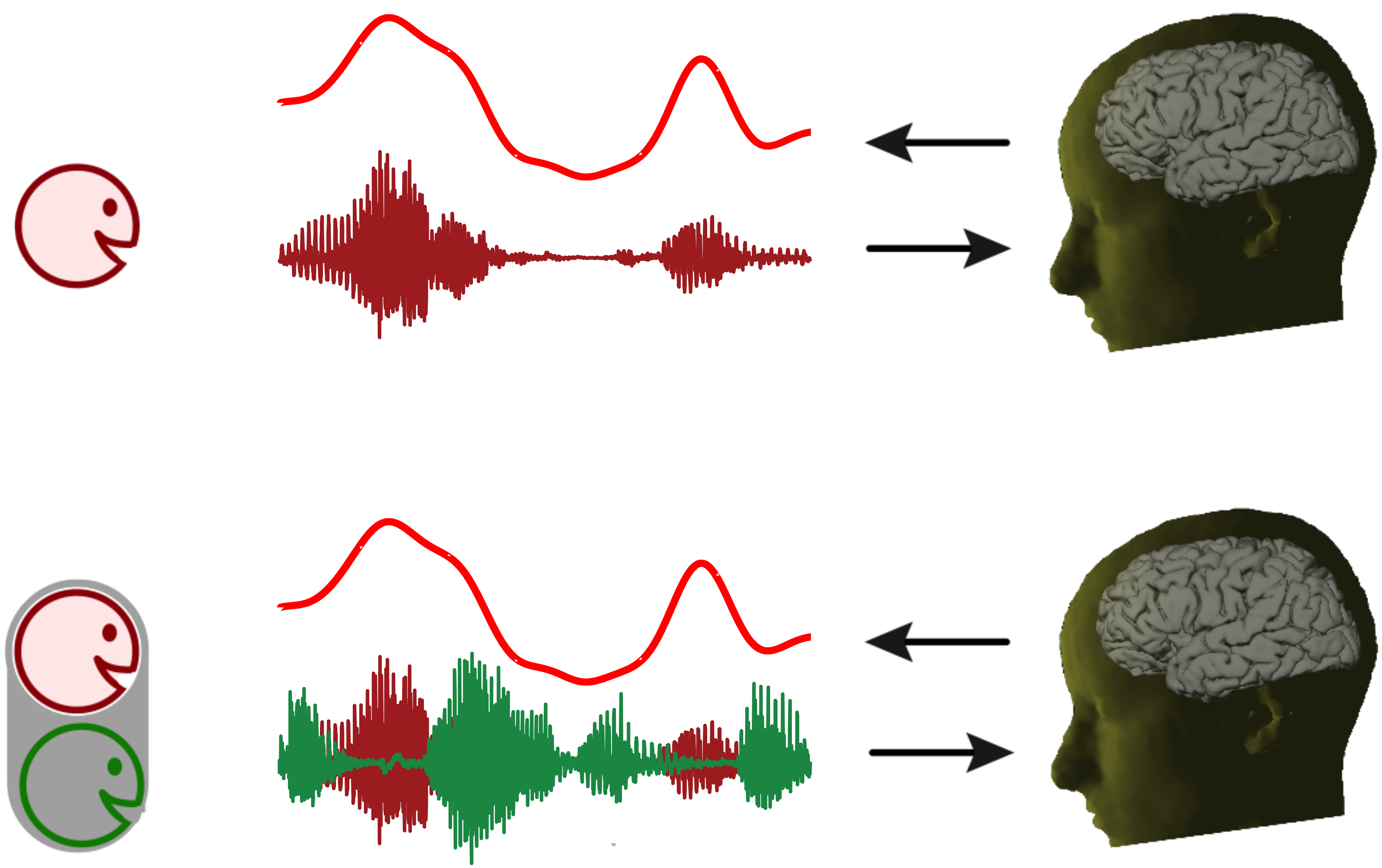
Competing Speech Streams



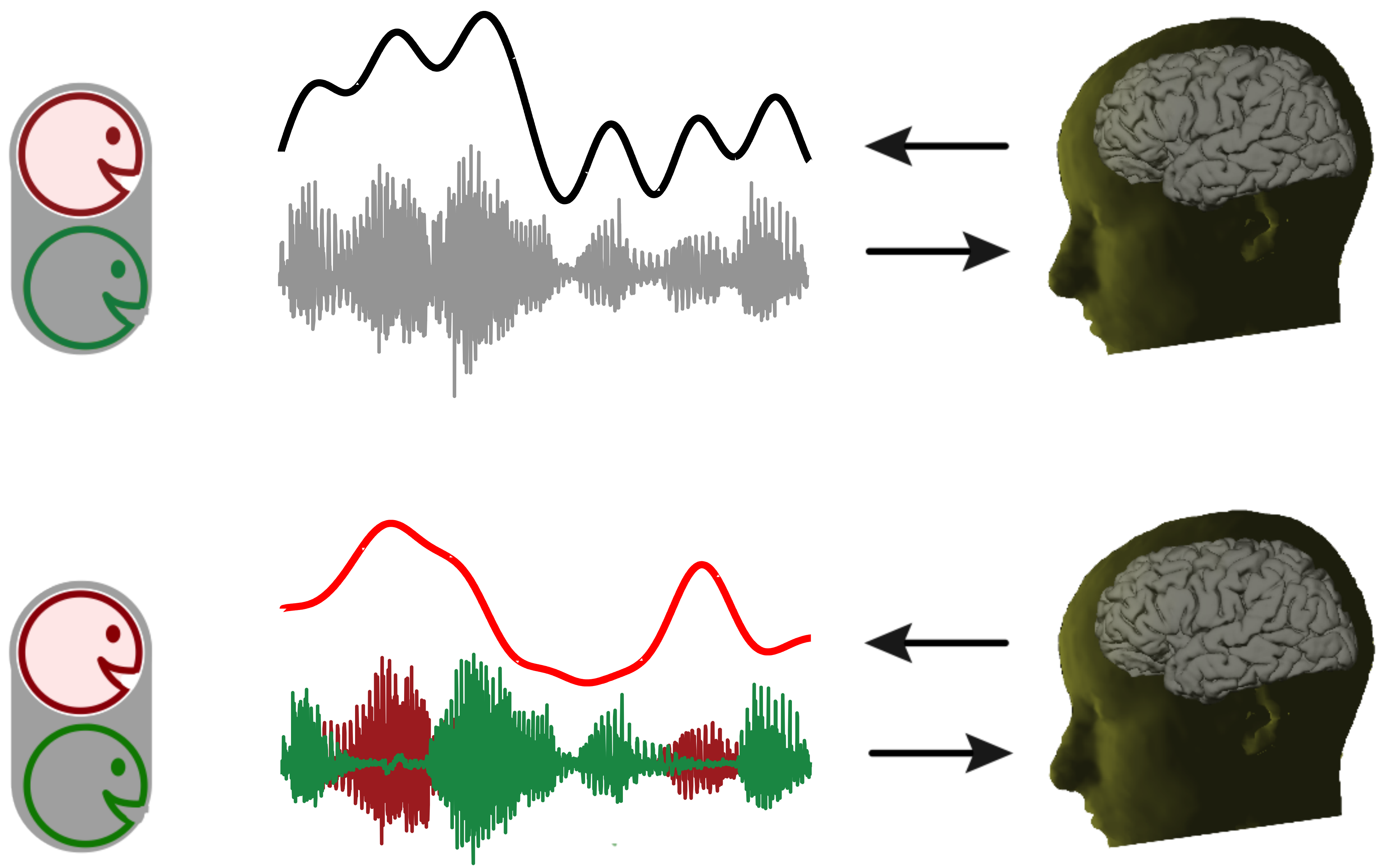
Selective Neural Encoding



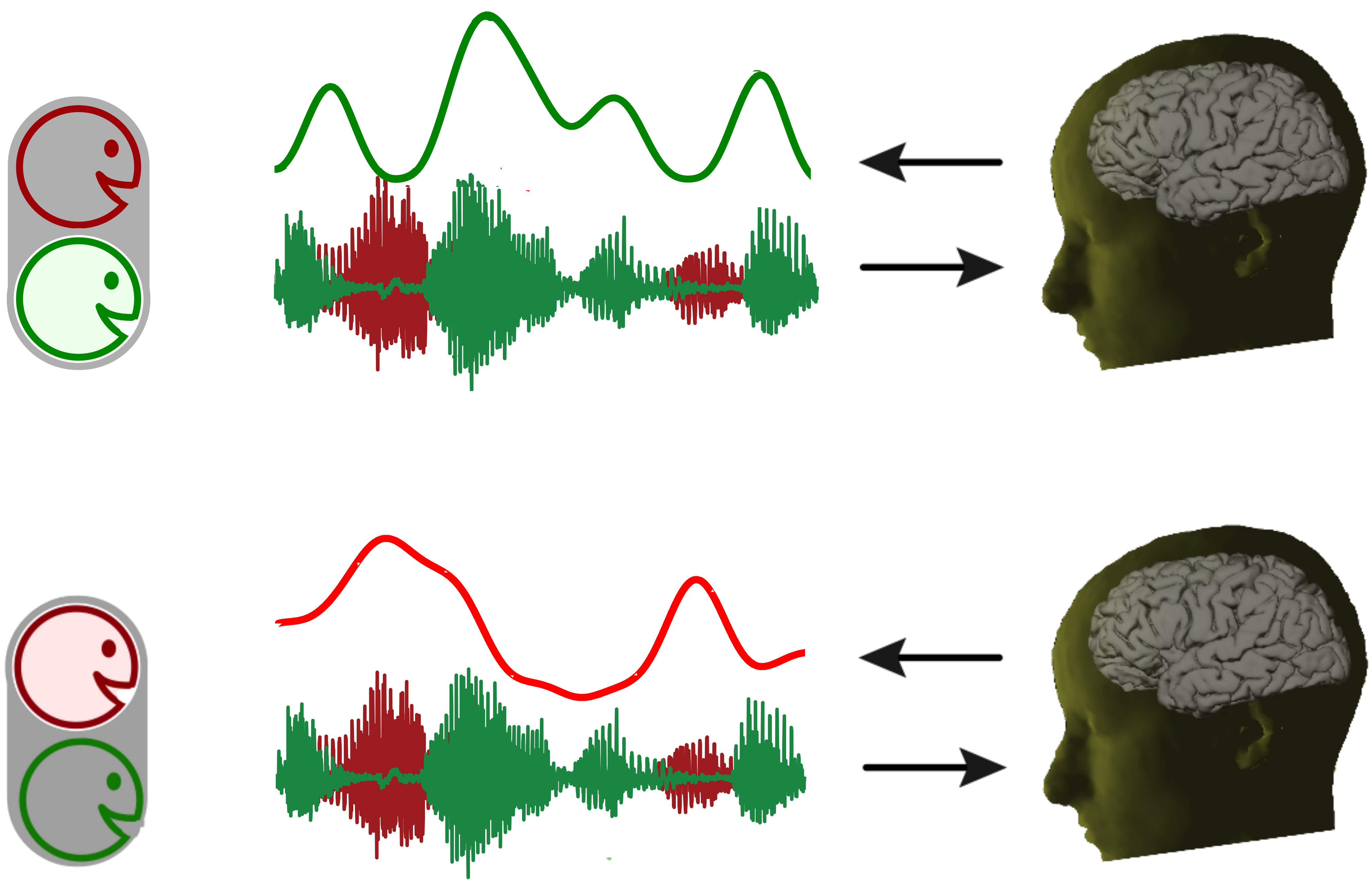
Selective Neural Encoding



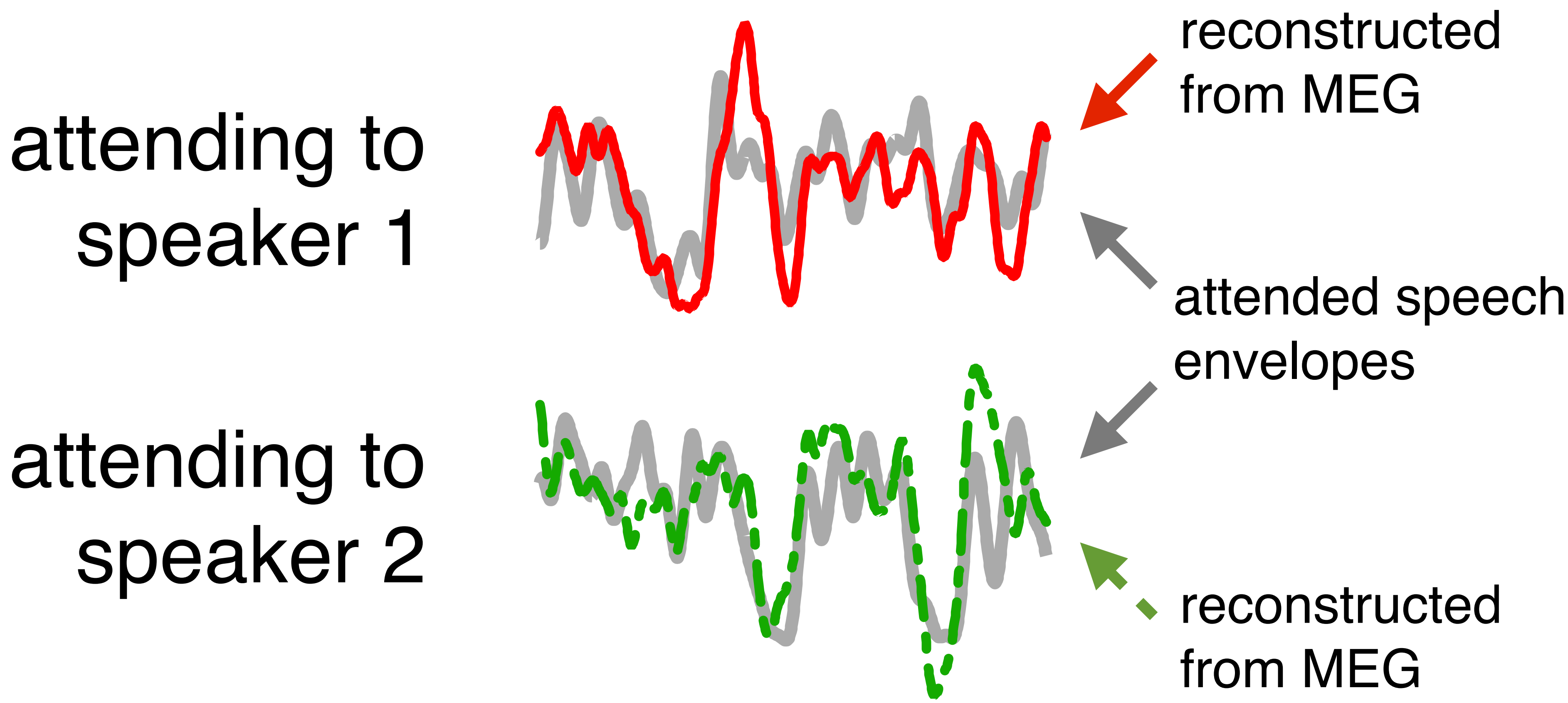
Unselective vs. Selective Neural Encoding



Selective Neural Encoding



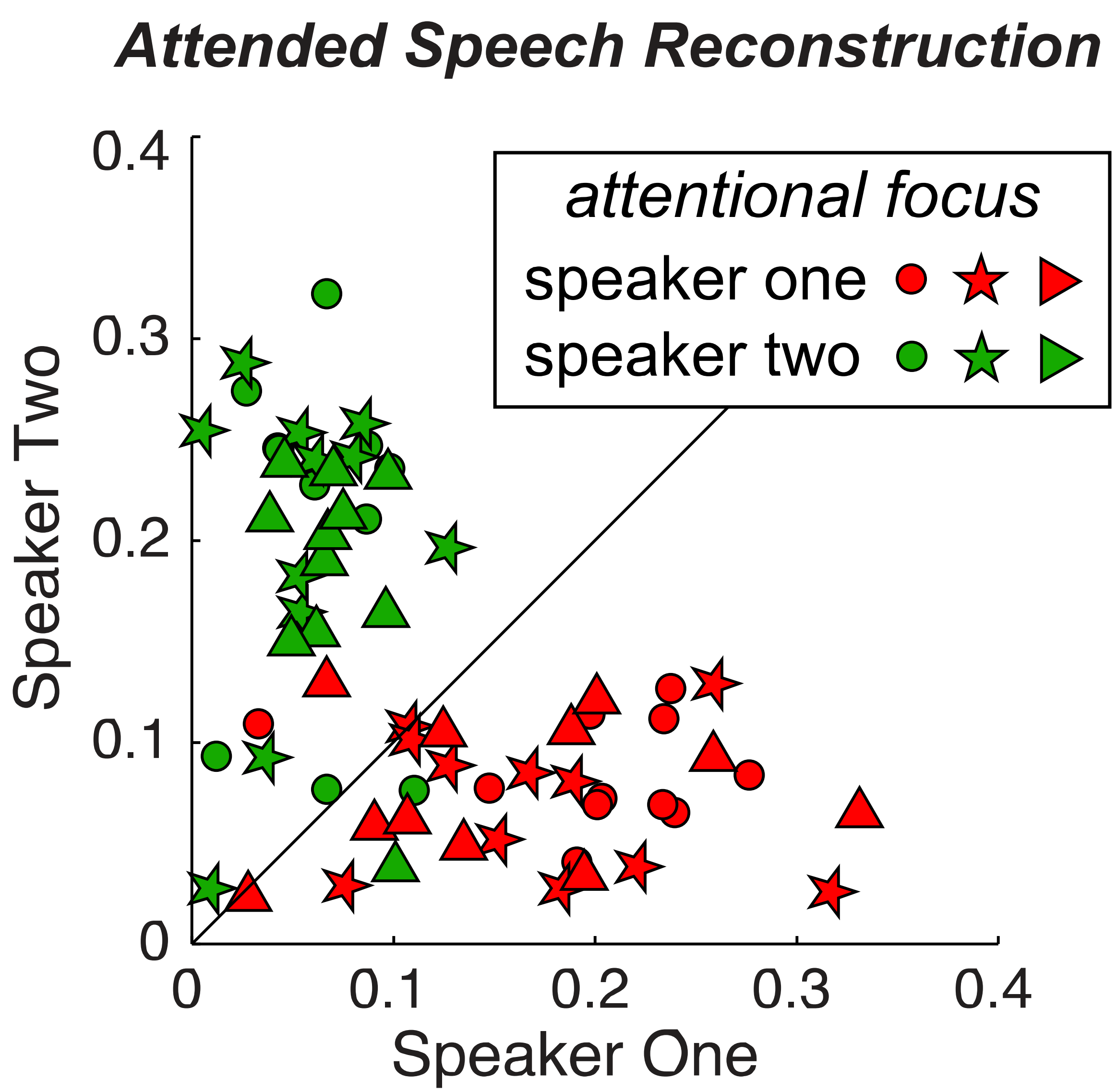
Selective Encoding: Results



Identical Stimuli!

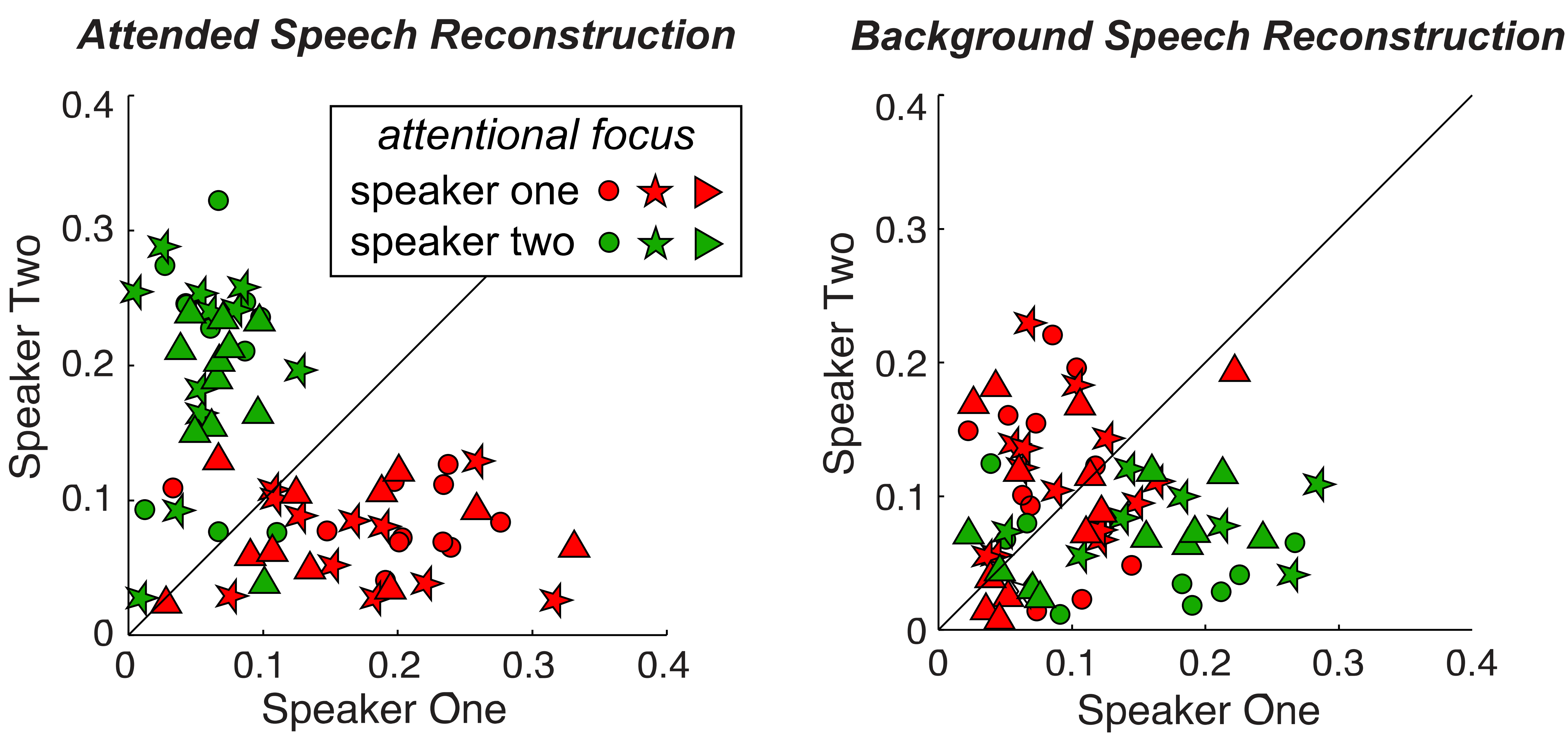
Ding & Simon, PNAS (2012)

Single Trial Speech Reconstruction



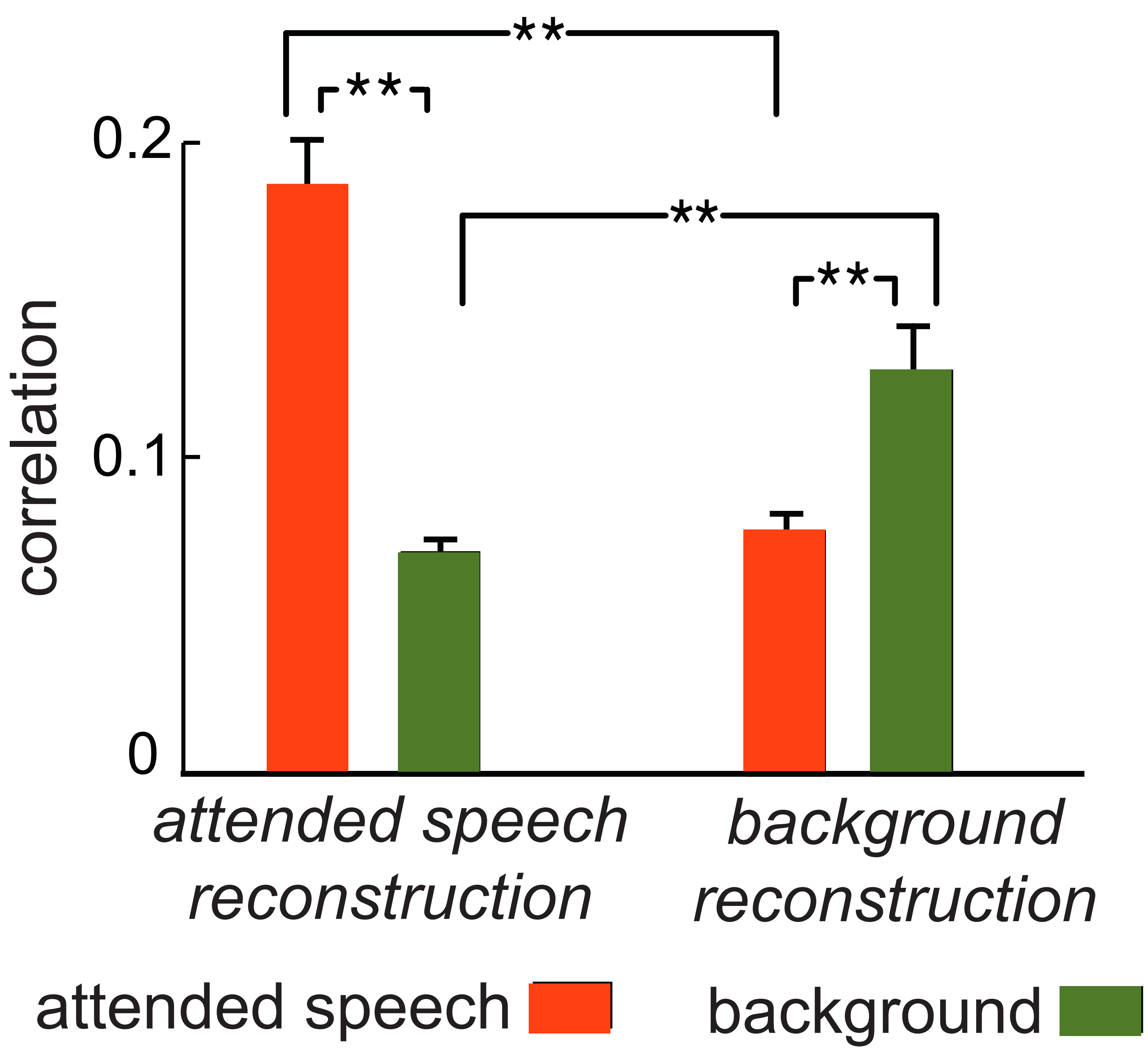
Ding & Simon, PNAS (2012)

Single Trial Speech Reconstruction



Ding & Simon, PNAS (2012)

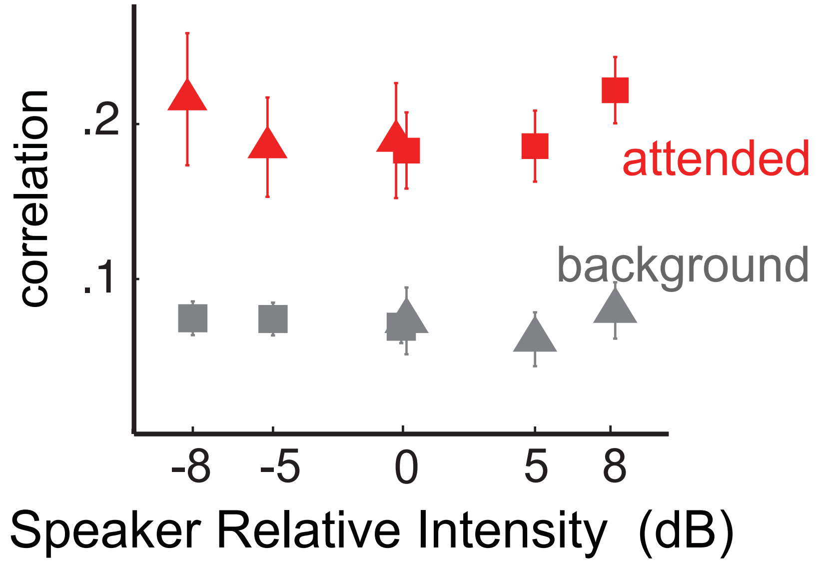
Overall Speech Reconstruction



Distinct neural representations for different speech streams

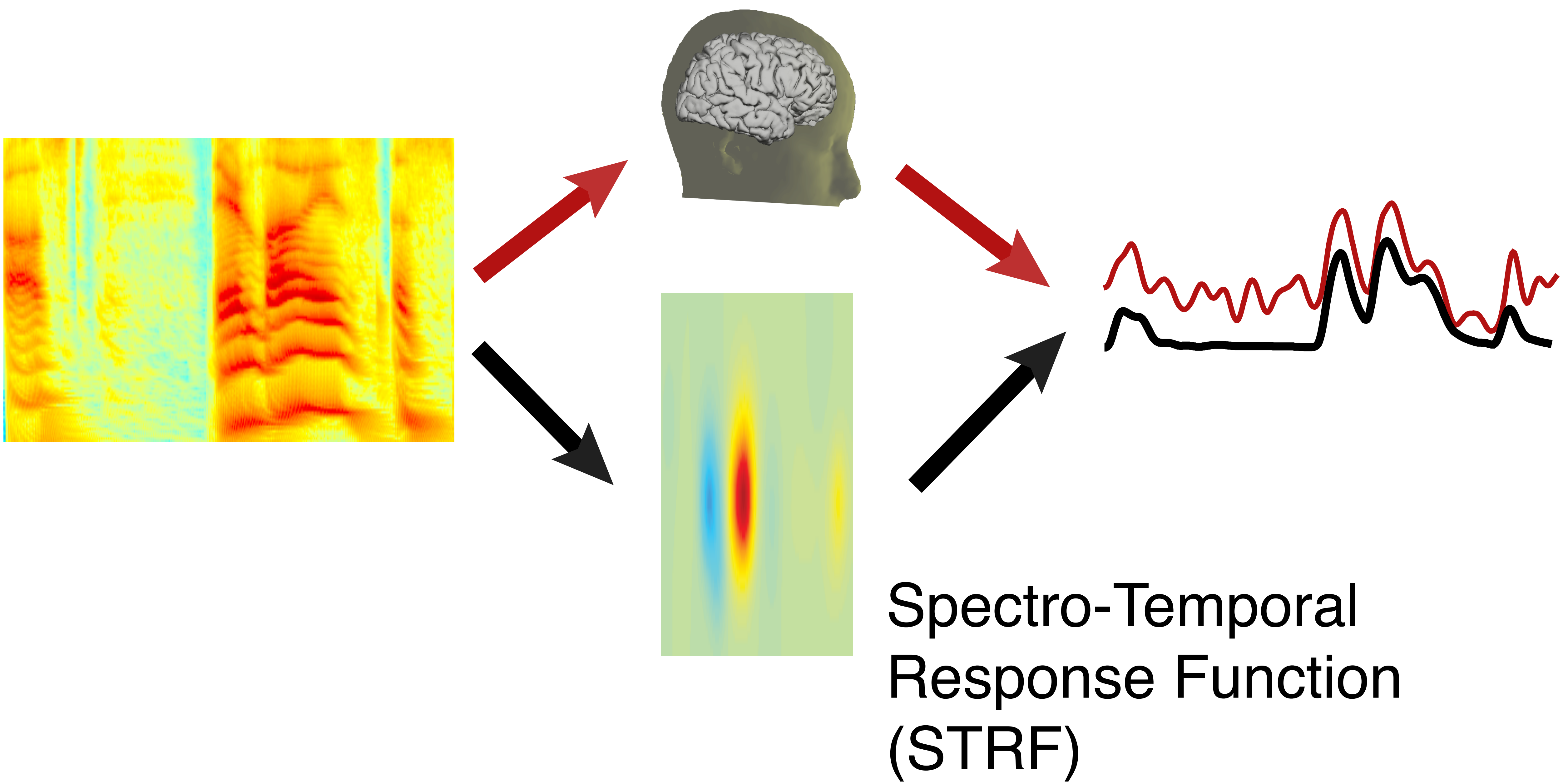
Invariance under Relative Loudness Change

Neural Results

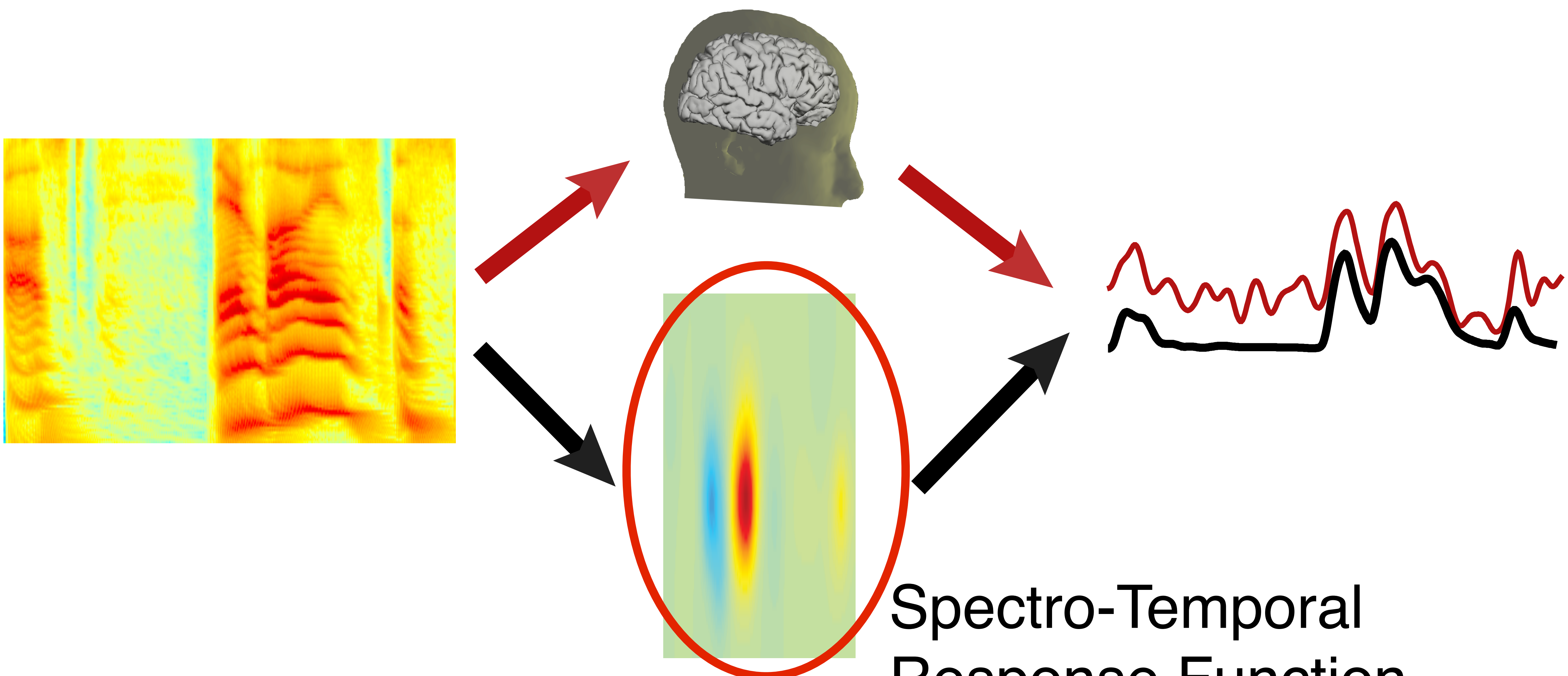


- Neural representation invariant to relative loudness change
- Stream-based Gain Control, not stimulus-based

Forward STRF Model

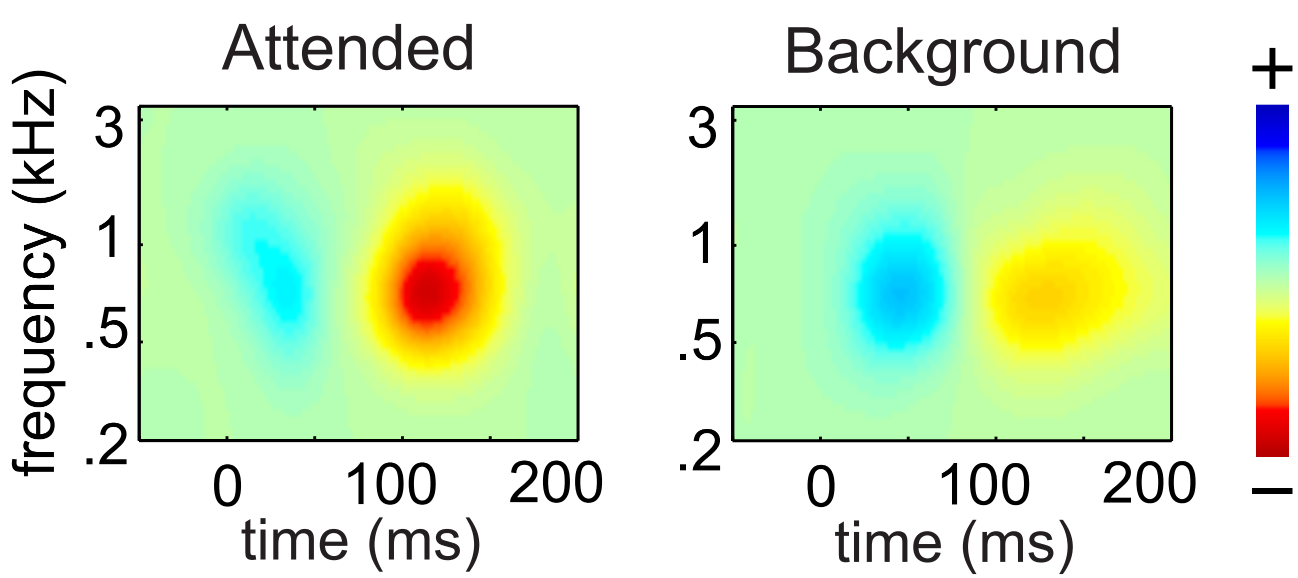


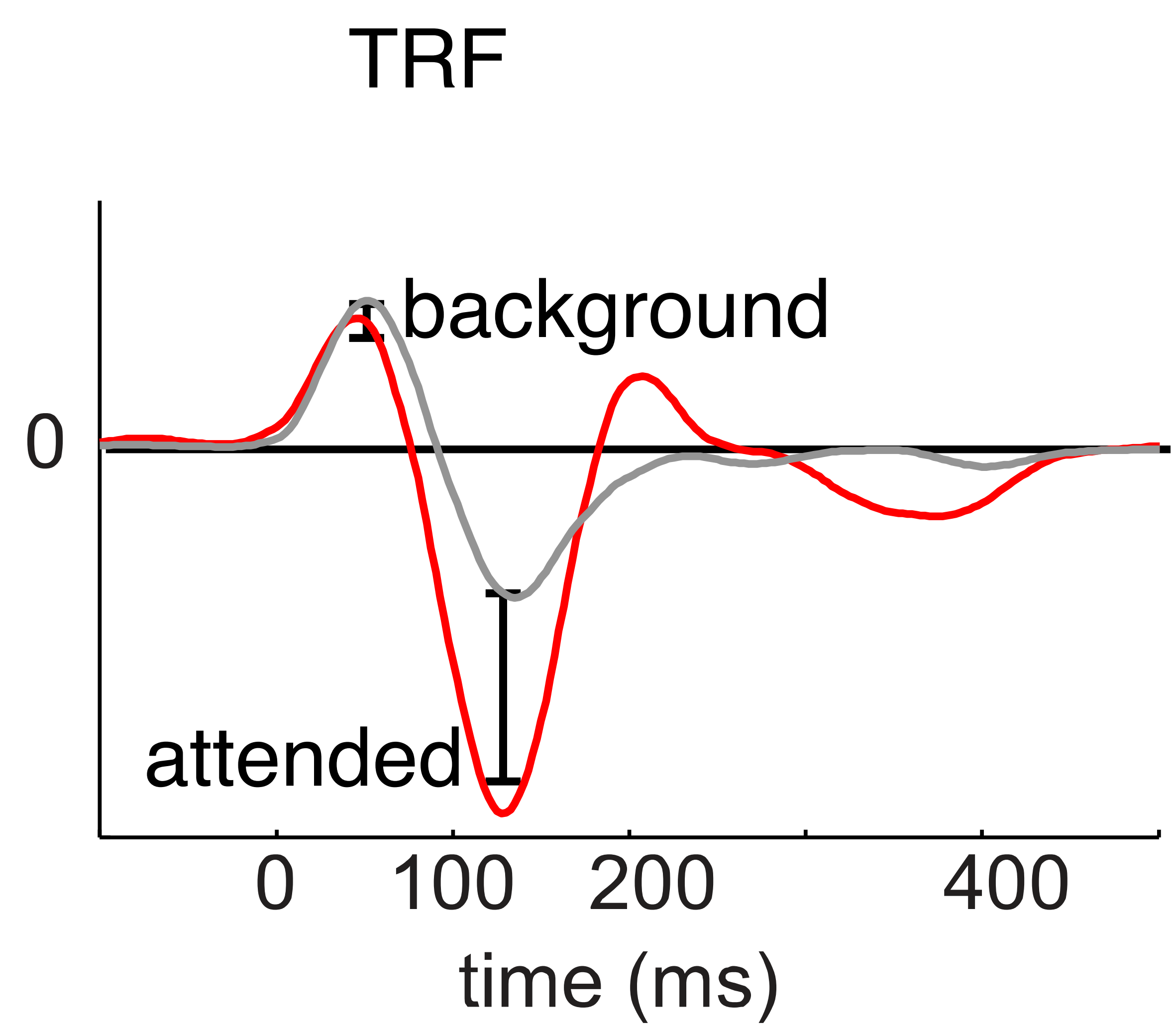
Forward STRF Model



Spectro-Temporal
Response Function
(STRF)

STRF Results





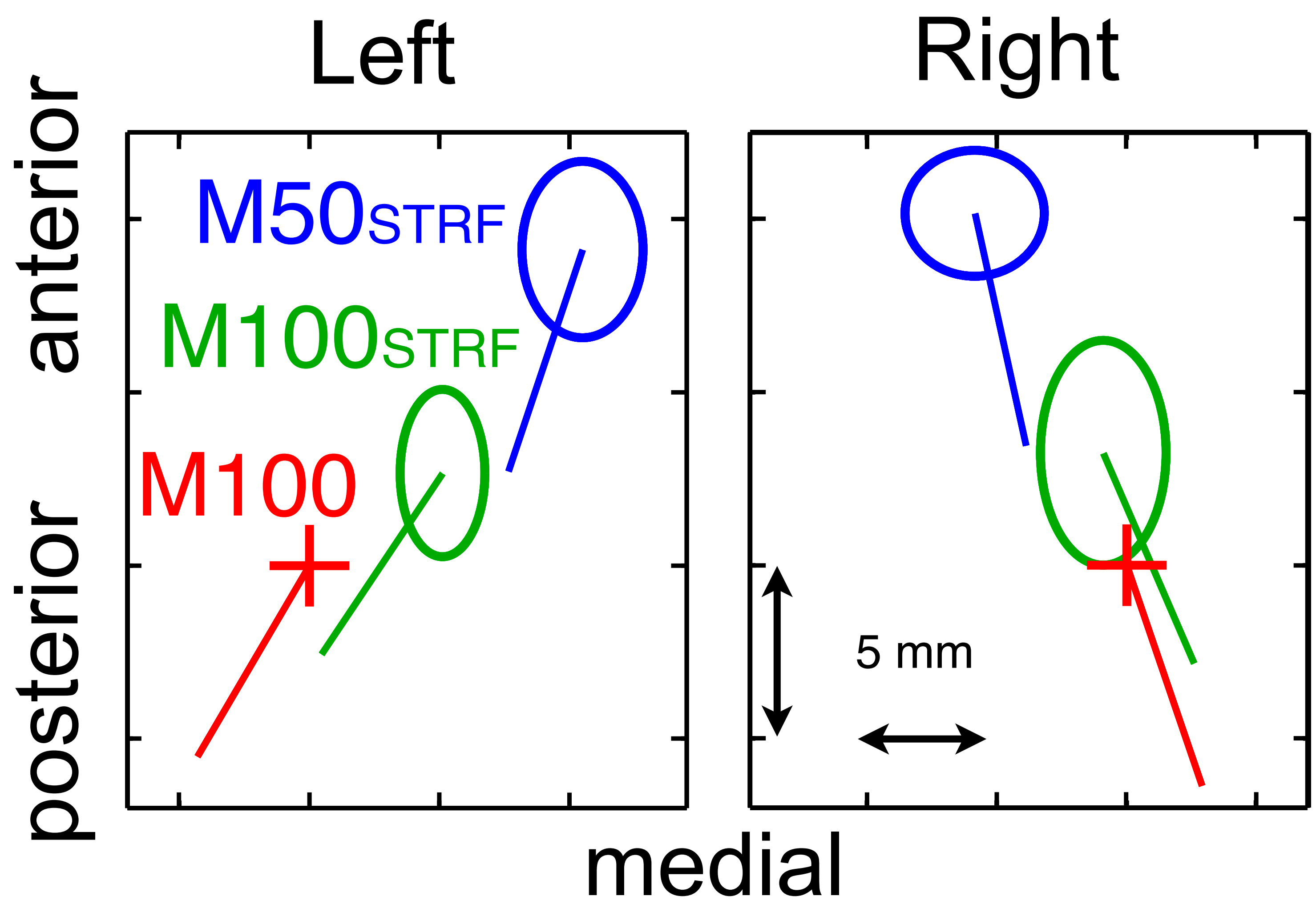
- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- $M50_{STRF}$ positive peak
- $M100_{STRF}$ negative peak
- **$M100_{STRF}$ strongly modulated by attention, *but not $M50_{STRF}$***

Neural Sources

- $M100_{STRF}$ source near (same as?) M100 source:
  Planum Temporale

- $M50_{STRF}$ source is anterior and medial to M100 (same as M50?):
  Heschl's Gyrus

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



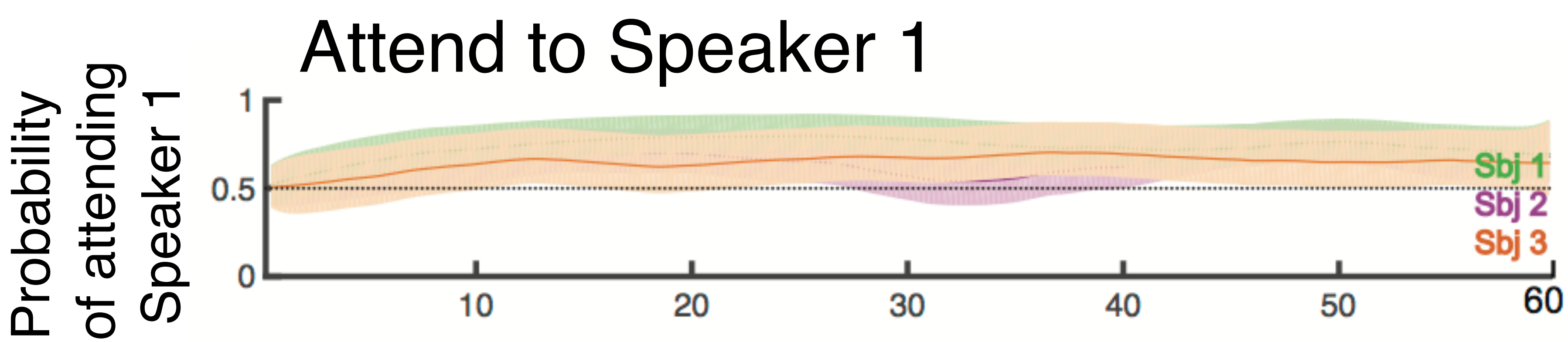
Recent Results

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

Recent Results

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

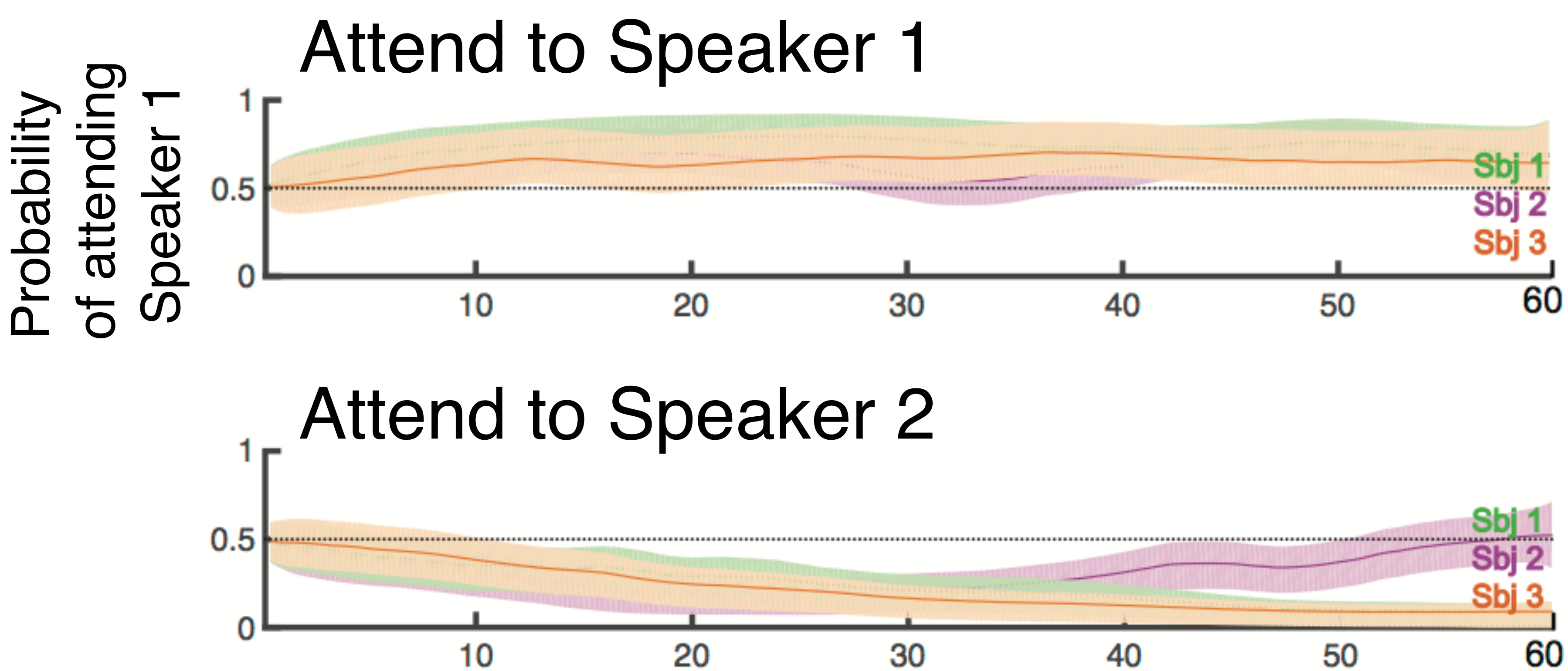
Attentional Dynamics



- Simple *dynamical* model of neural correlate of attentional direction
- Time resolution ~5 s (not, e.g., 60 s)

Akram et al. NeuroImage (2016)

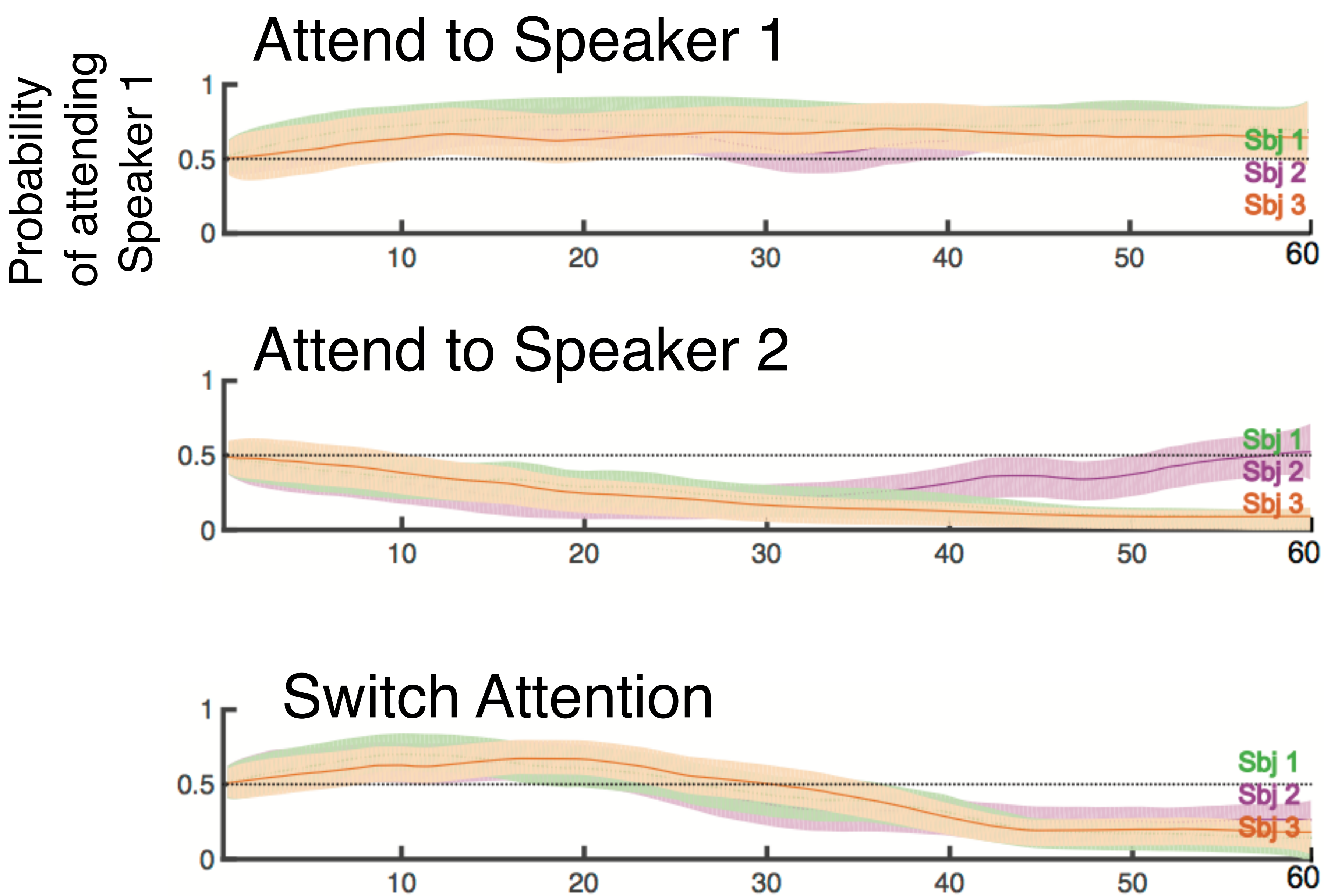
Attentional Dynamics



- Simple *dynamical* model of neural correlate of attentional direction
- Time resolution ~5 s (not, e.g., 60 s)
- Less conservative in assumptions regarding actual subject behavior

Akram et al. NeuroImage (2016)

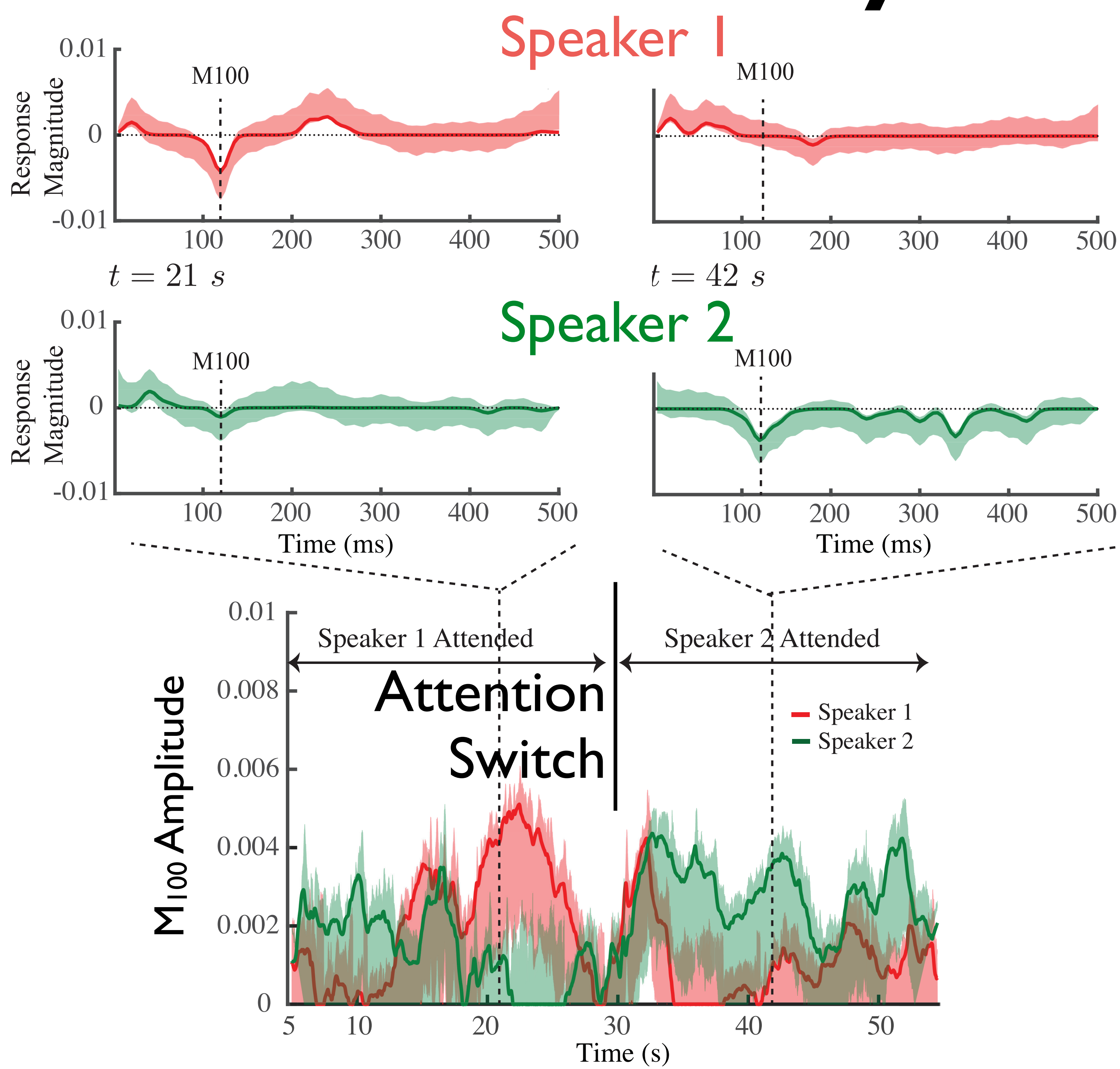
Attentional Dynamics



- Simple *dynamical* model of neural correlate of attentional direction
- Time resolution ~5 s (not, e.g., 60 s)
- Less conservative in assumptions regarding actual subject behavior
- Observable attentional (neural) dynamics

Akram et al. NeuroImage (2016)

TRF Dynamics



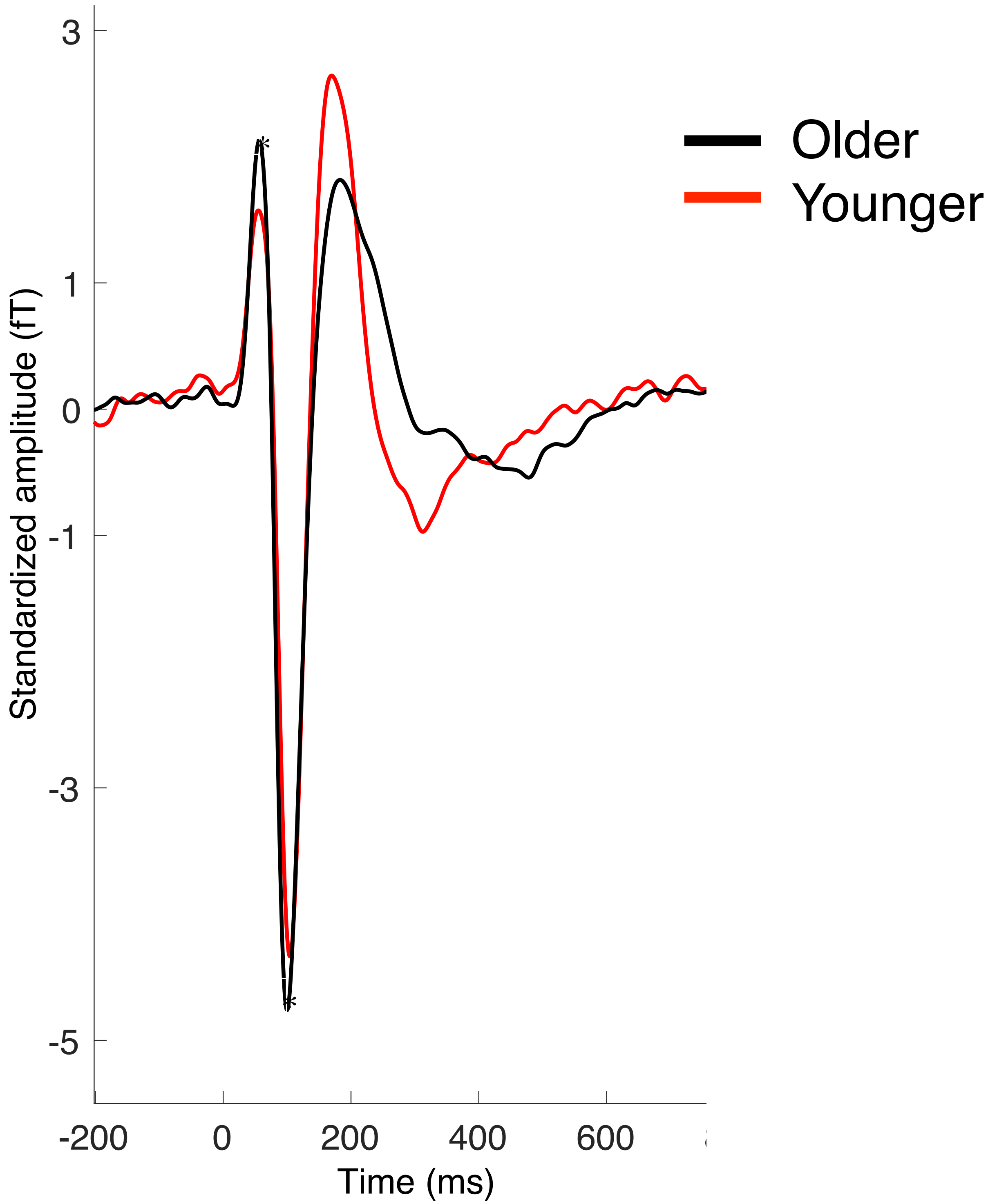
- Dynamical model entire TRF, including attentional modulation
- Time resolution still ~5 s

Akram et al. IEEE Tr Biomed Eng (2017)

Recent Results

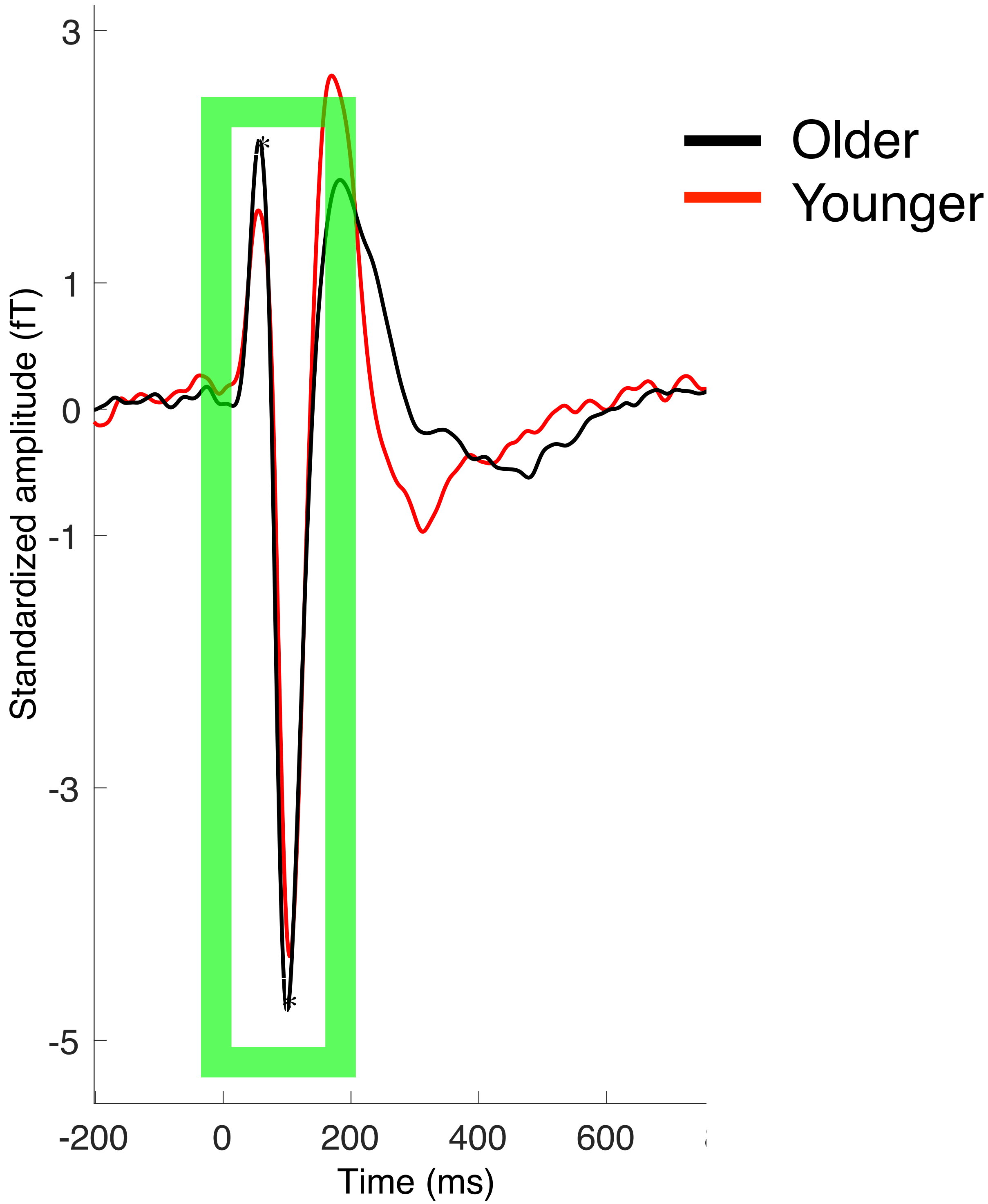
- Attentional Dynamics
- Aging & Cortical Representations of Speech
- High Level Interference & Noise

Aging & Auditory Cortex



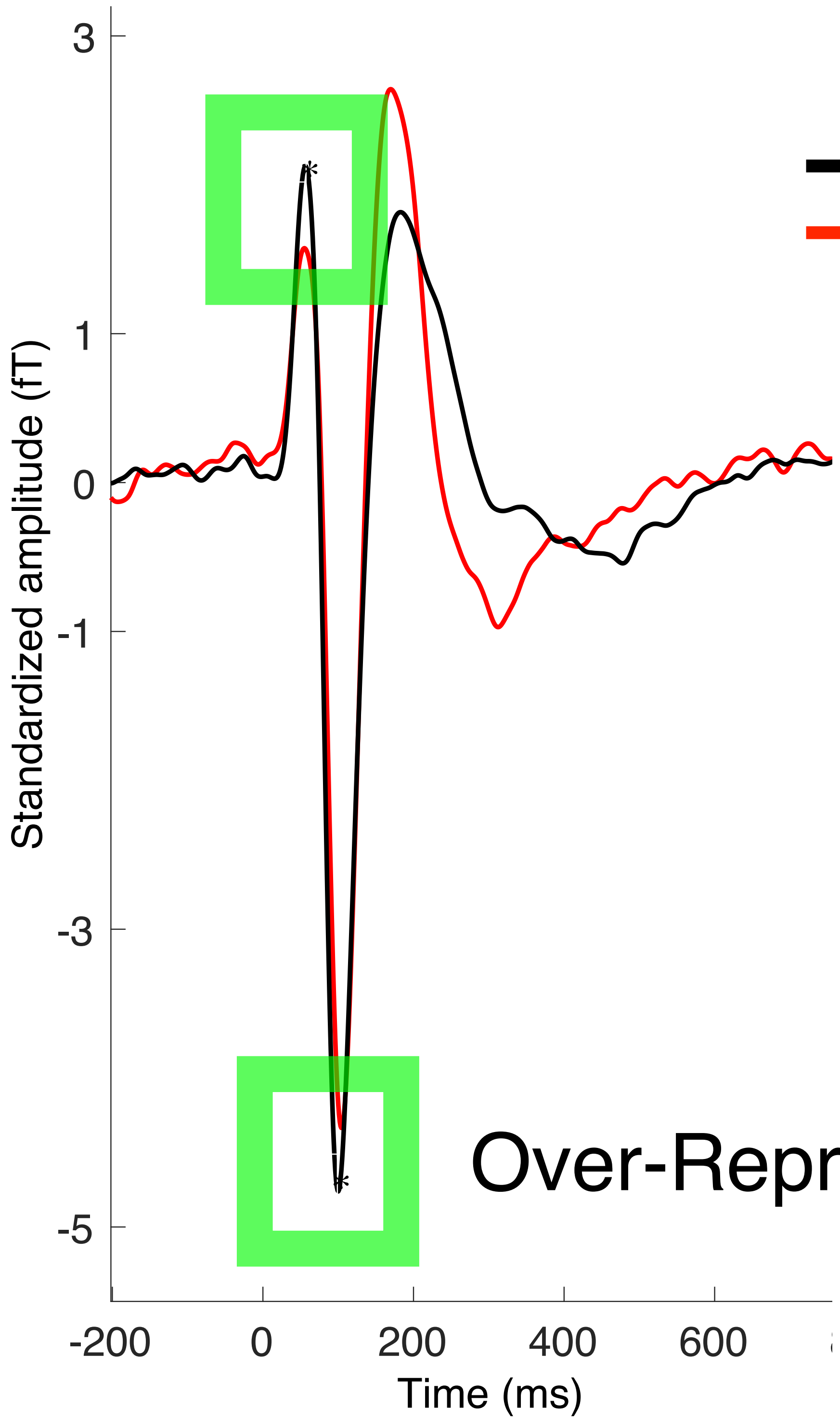
Average Responses to Pure Tone

Aging & Auditory Cortex



Average Responses to Pure Tone

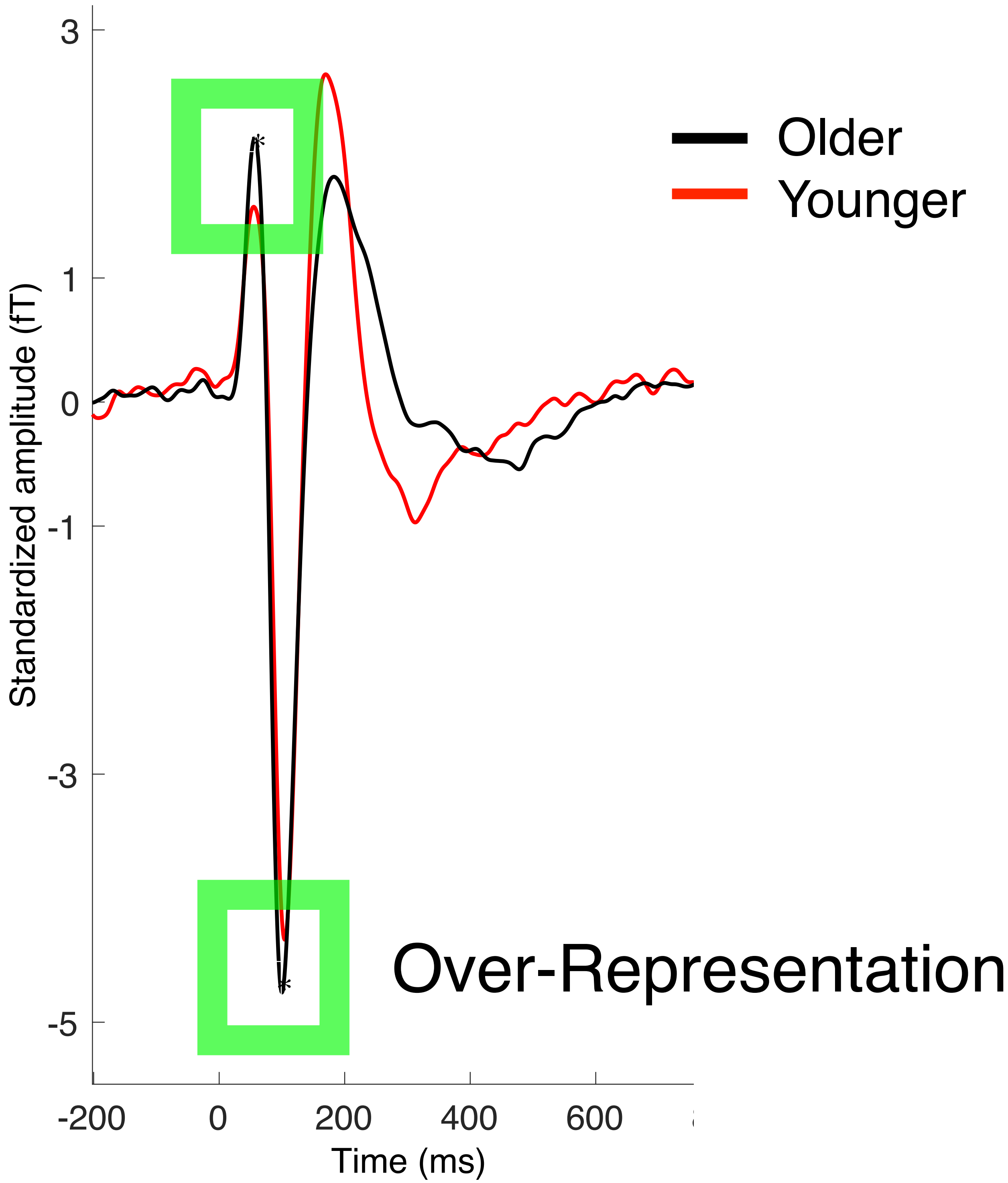
Aging & Auditory Cortex



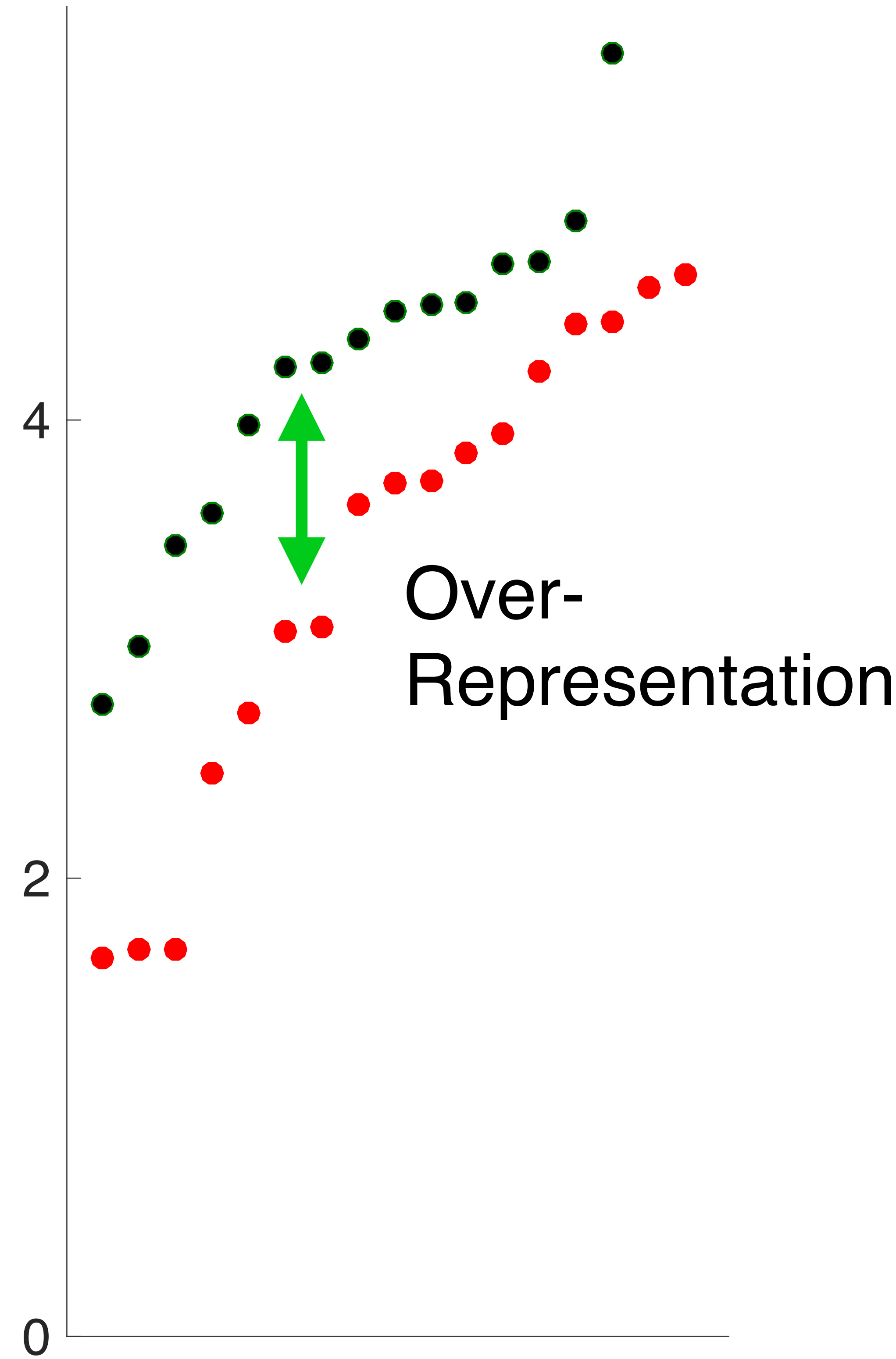
Over-Representation

Average Responses to Pure Tone

Aging & Auditory Cortex

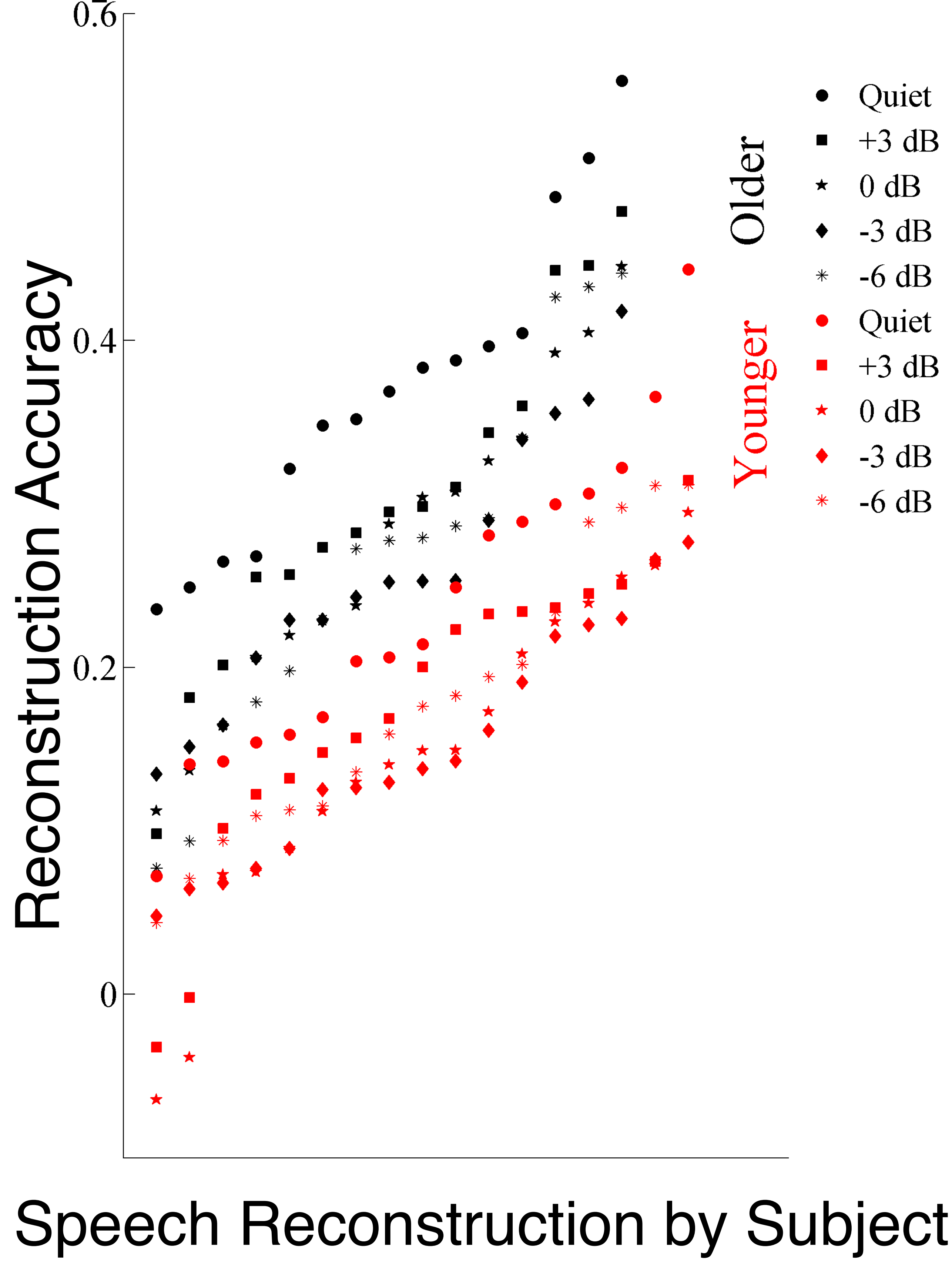


Average Responses to Pure Tone



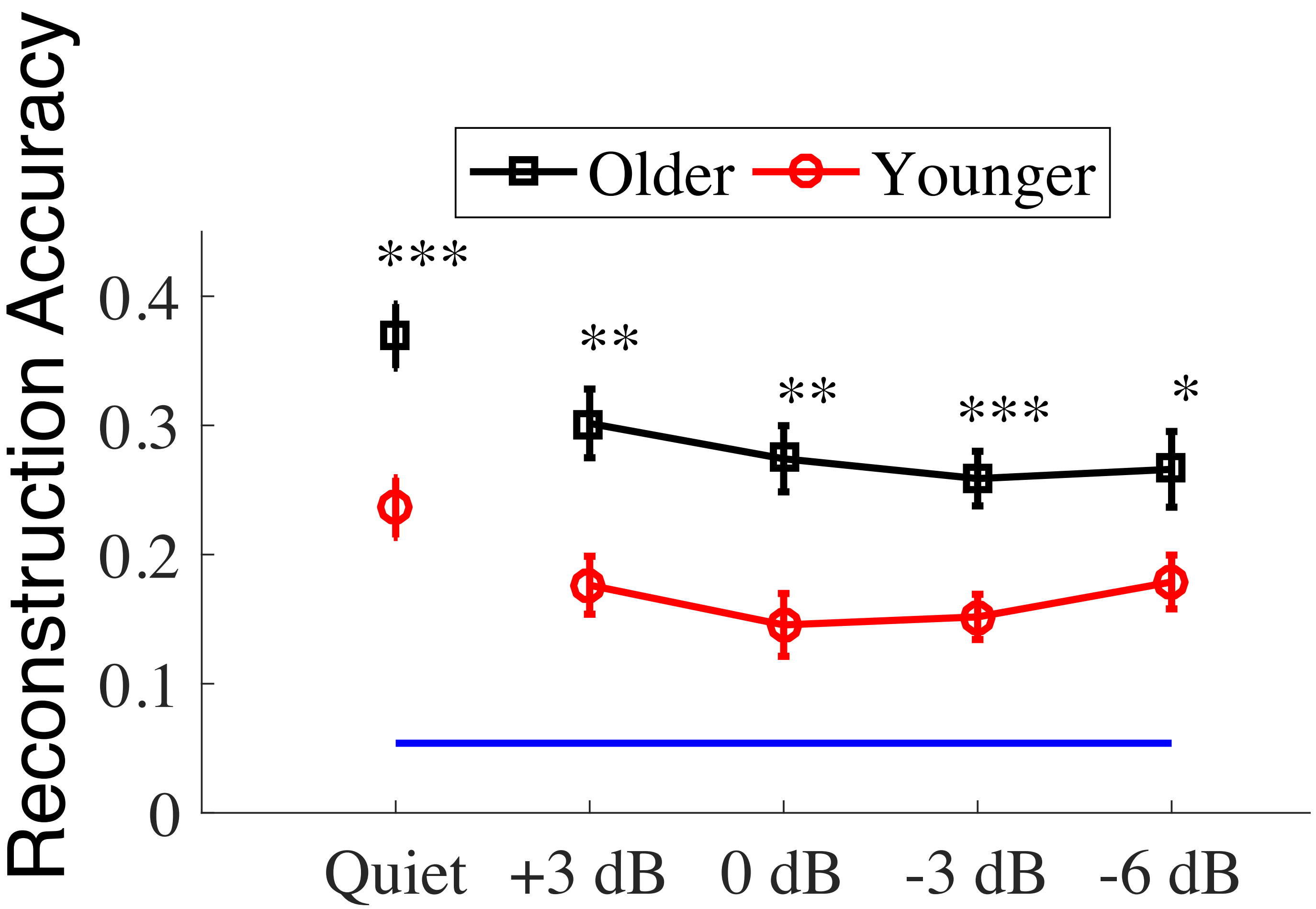
M100 Power by Subject

Speech Over-Representation

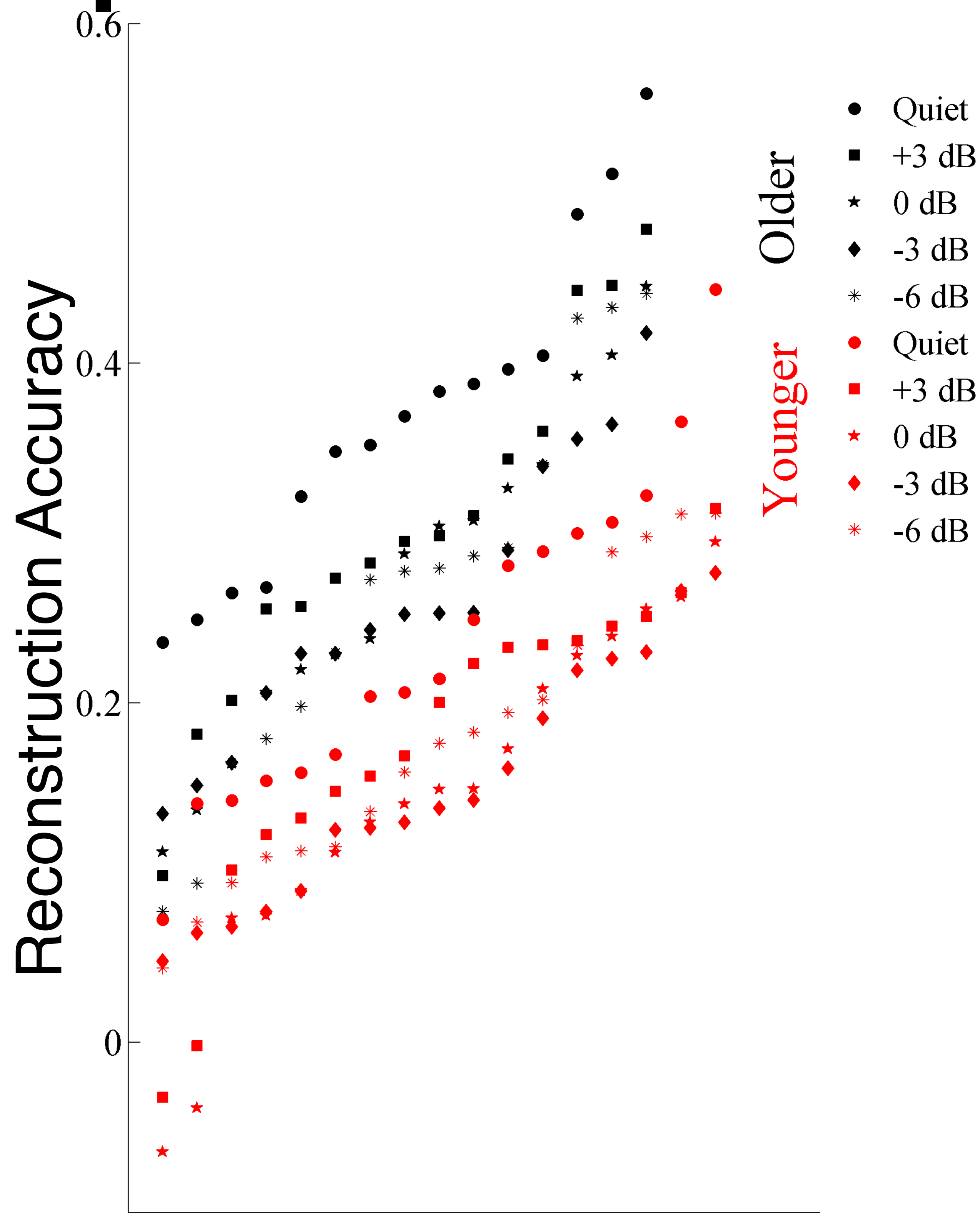


Presacco et al., J Neurophysiol (2016a)

Speech Over-Representation



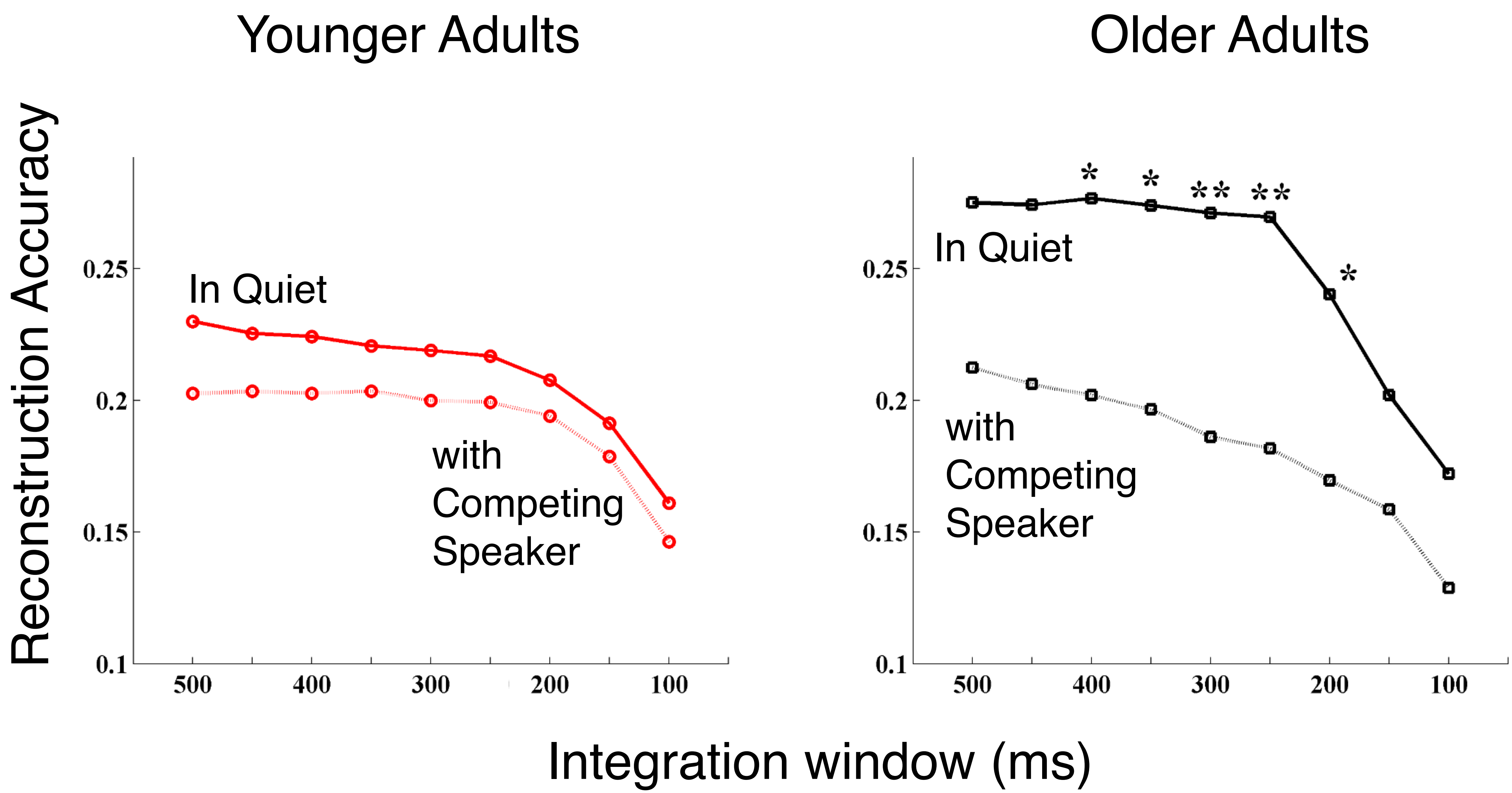
Speech Reconstruction by SNR



Speech Reconstruction by Subject

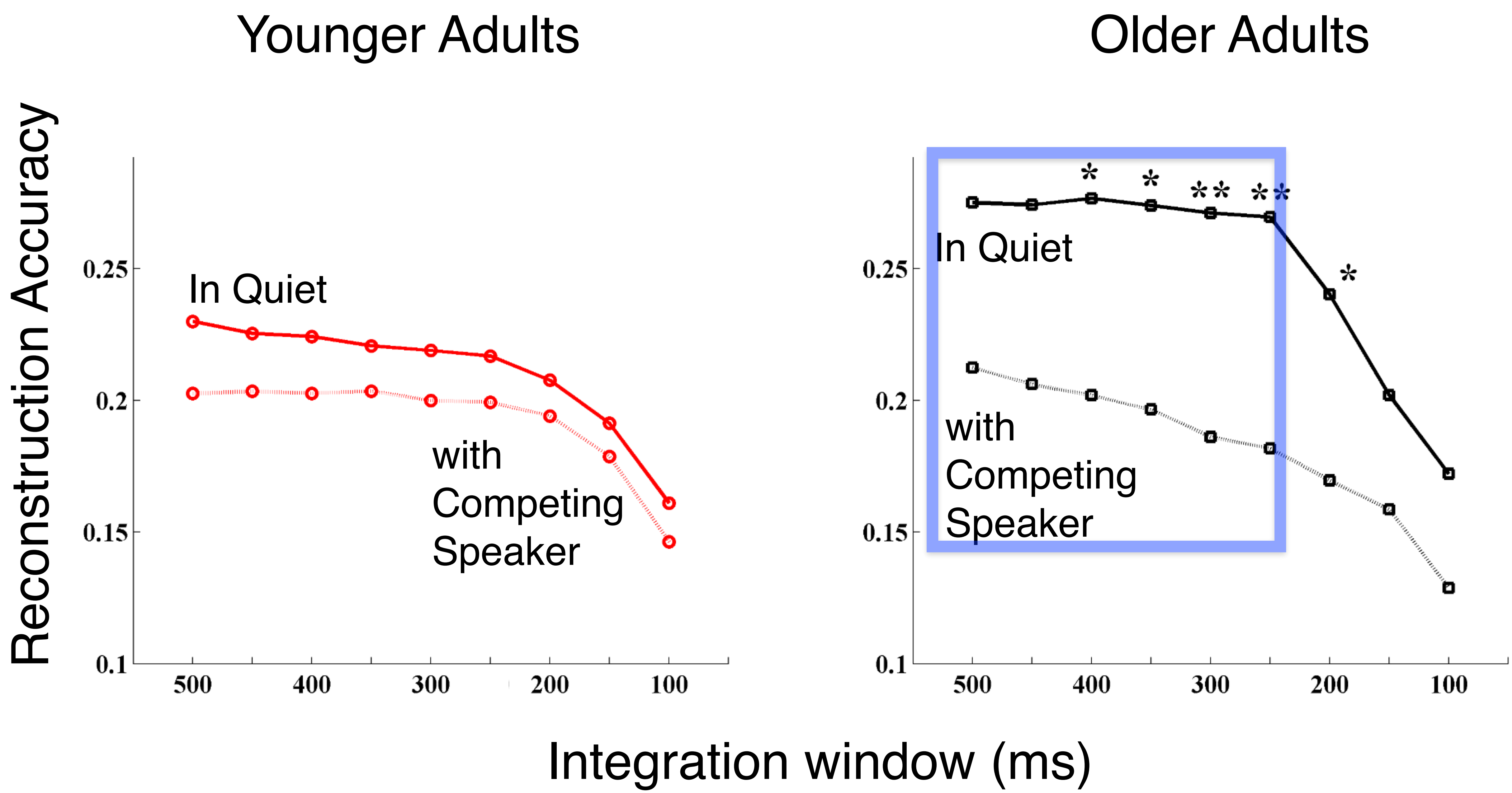
Presacco et al., J Neurophysiol (2016a)

Aging & Integration Time



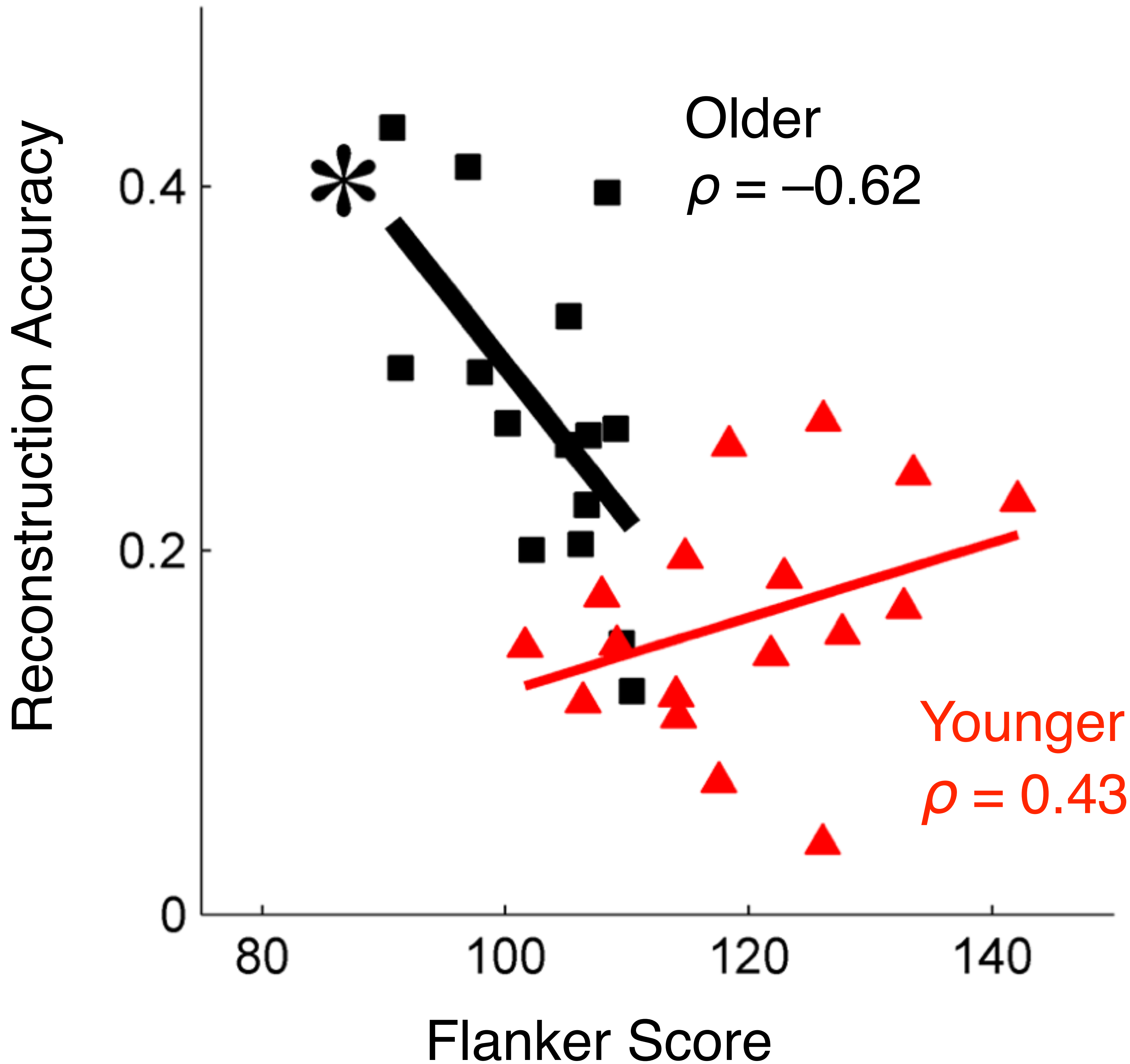
Presacco et al., J Neurophysiol (2016a)

Aging & Integration Time



Presacco et al., J Neurophysiol (2016a)

Neural vs Inhibitory Control

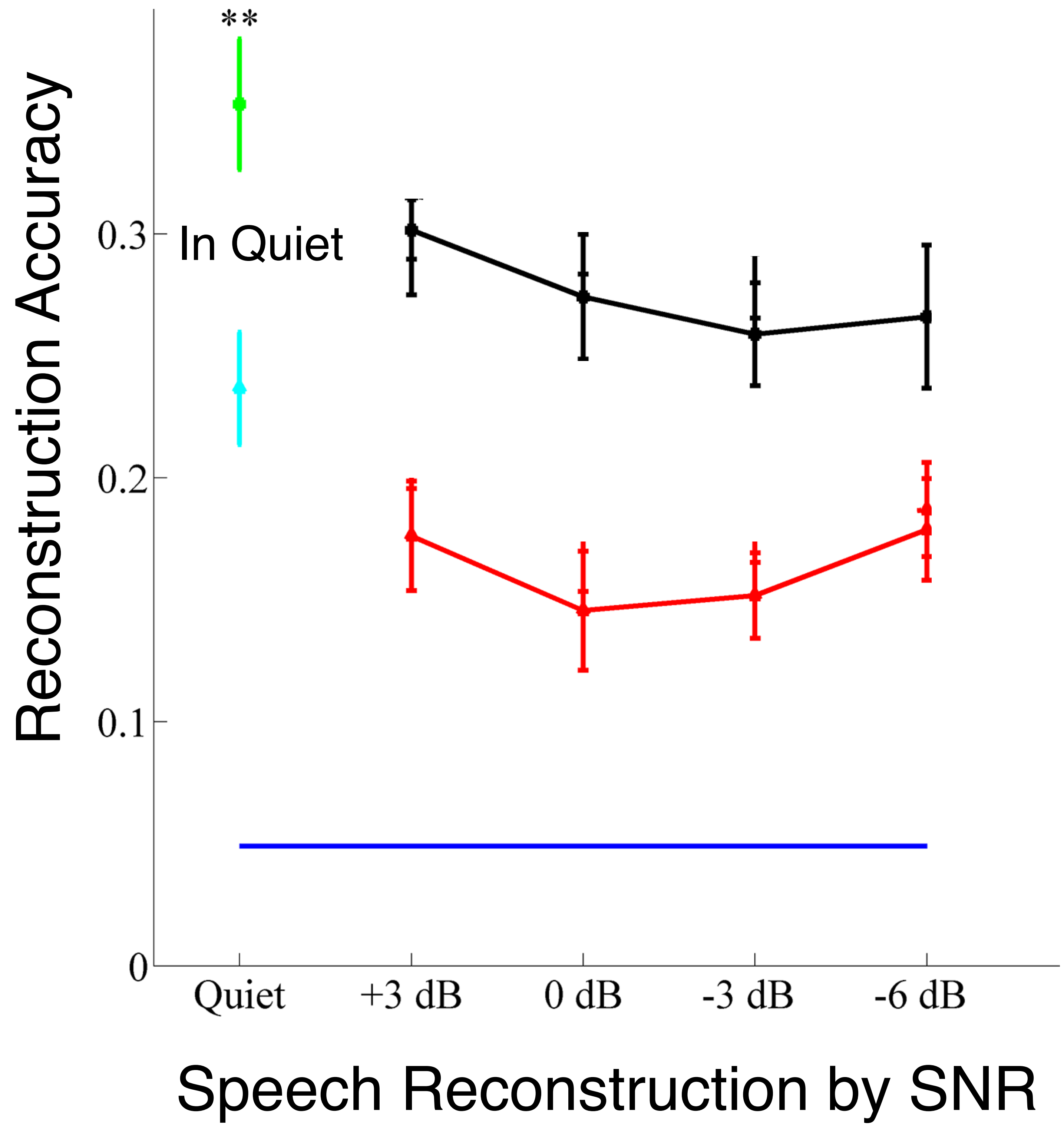


Presacco et al., J Neurophysiol (2016b)

Recent Results

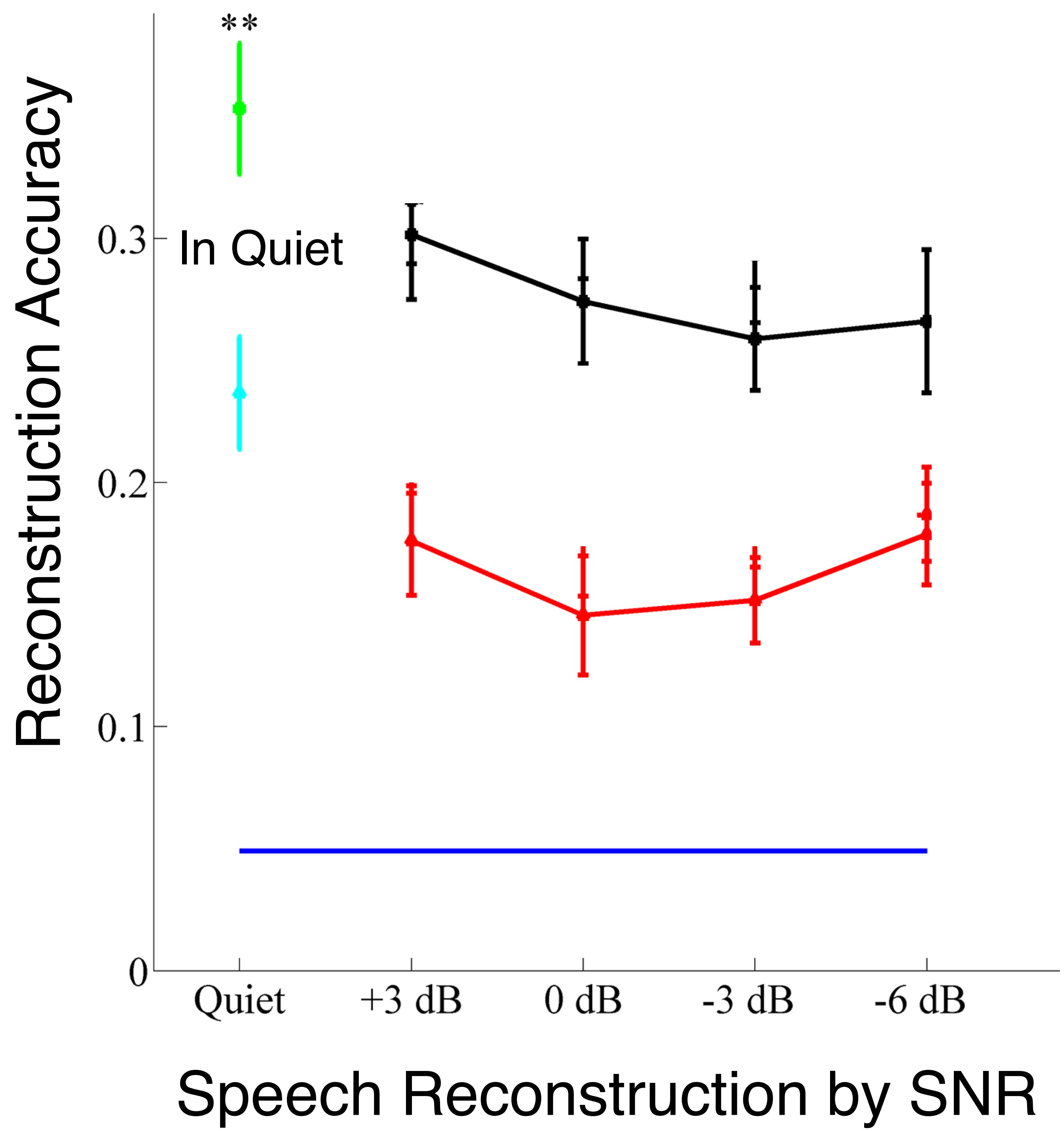
- Attentional Dynamics
- Aging & Cortical Representations of Speech
- High Level Interference & Noise

High Level Interference Effects



Presacco et al., J Neurophysiol (2016b)

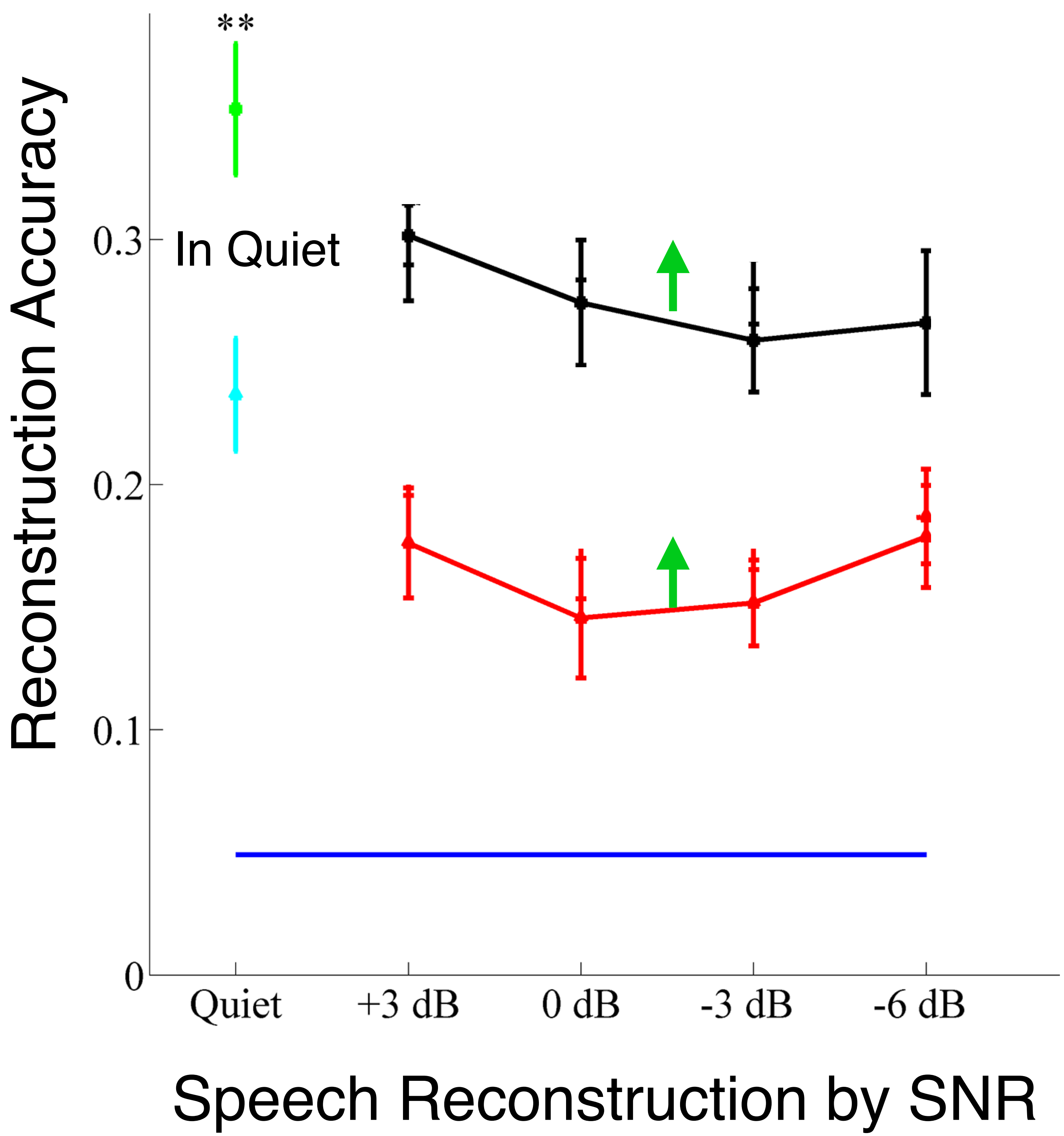
High Level Interference Effects



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

Presacco et al., J Neurophysiol (2016b)

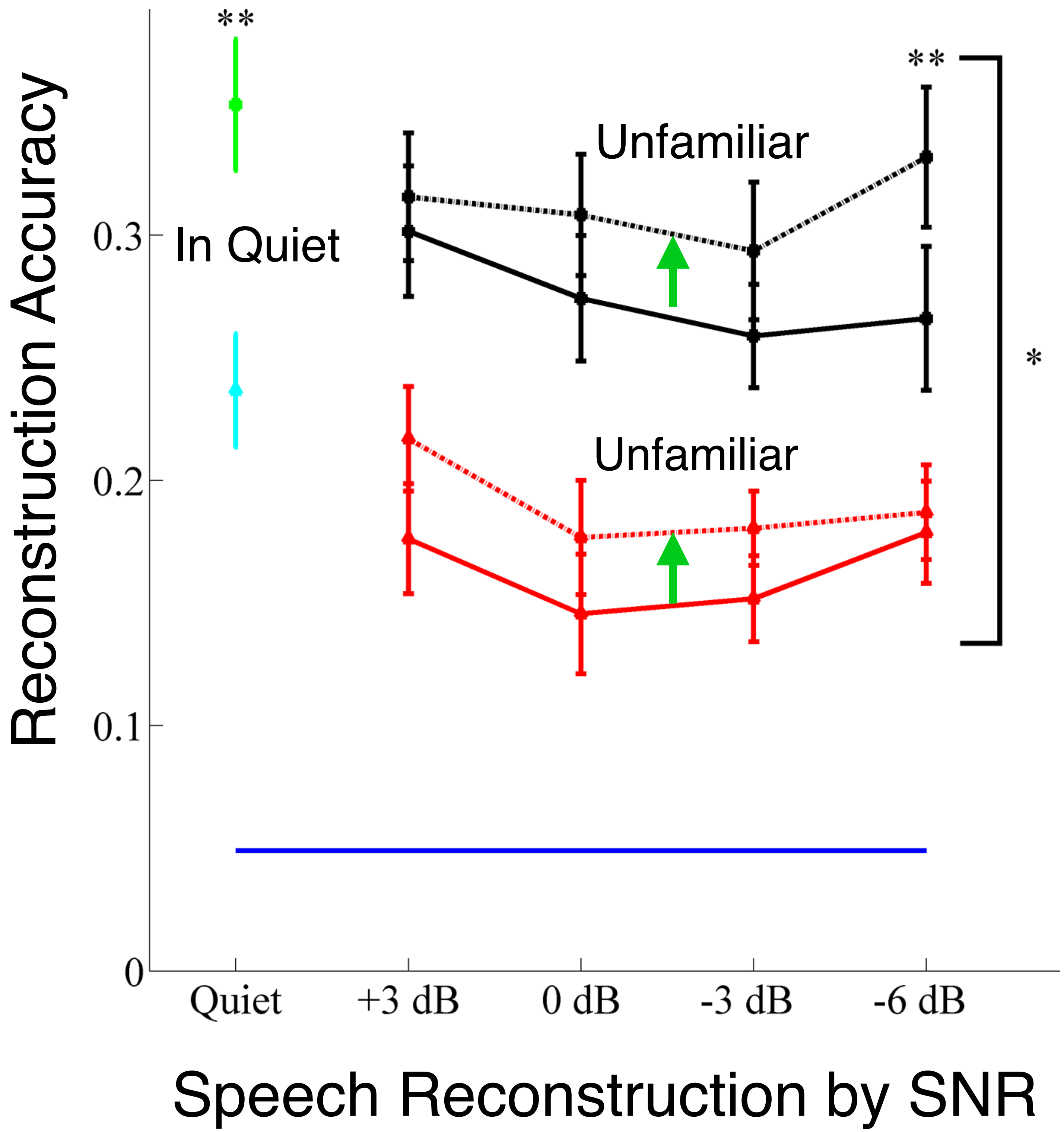
High Level Interference Effects



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

Presacco et al., J Neurophysiol (2016b)

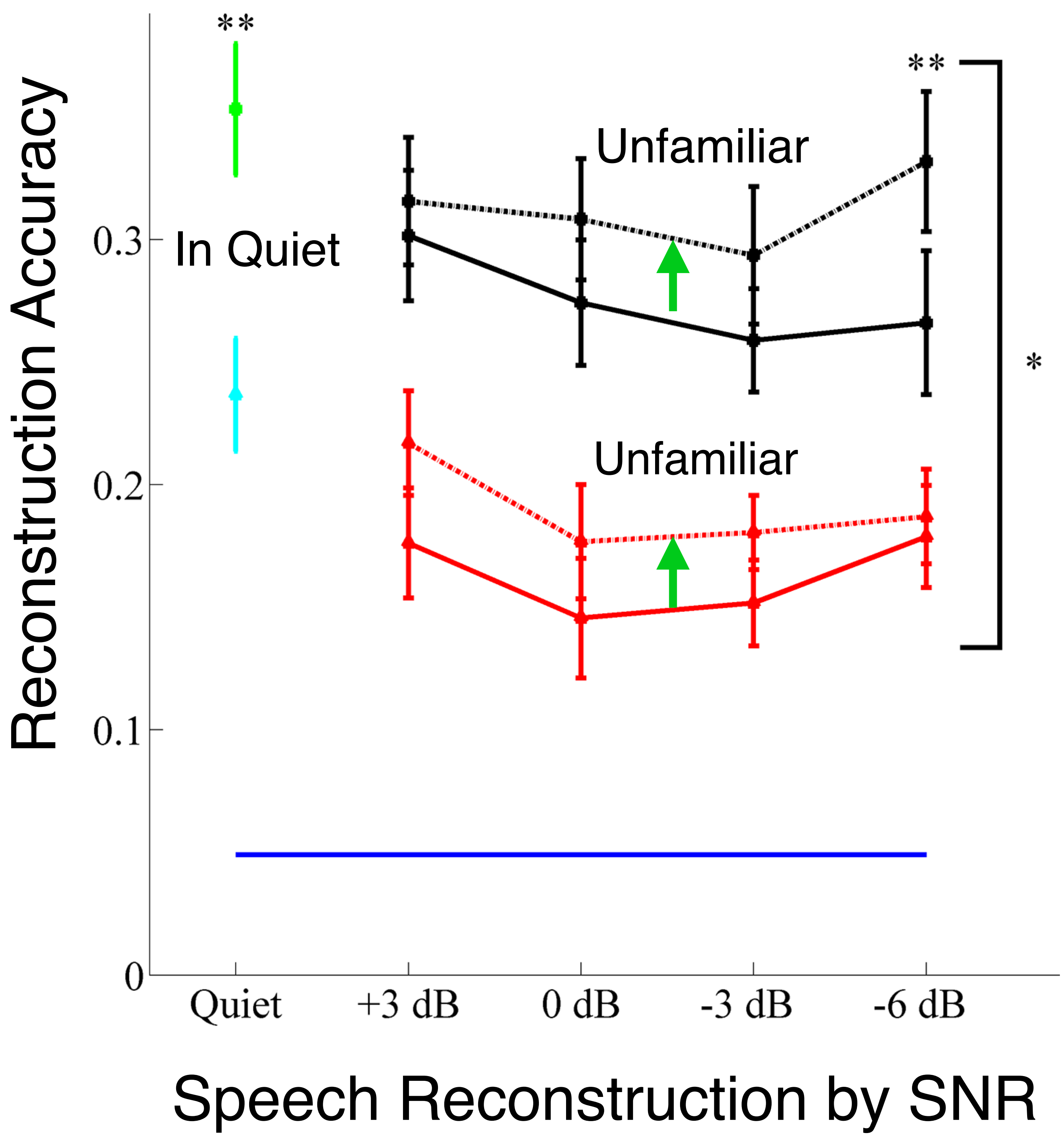
High Level Interference Effects



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

Presacco et al., J Neurophysiol (2016b)

High Level Interference Effects



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

Presacco et al., J Neurophysiol (2016b)

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