

Neural Representations of the Cocktail Party in Human Auditory Cortex

Jonathan Z. Simon

Department of Biology

Department of Electrical & Computer Engineering

Institute for Systems Research

University of Maryland

Acknowledgements

Current (Simon Lab & Affiliates)

Sahar Akram

Francisco Cervantes

Natalia Lapinskaya

Mahshid Najafi

Alex Presacco

Krishna Puvvada

Lisa Uible

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

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

Elliot Hong

Cindy Moss

David Poeppel

Shihab Shamma

Past Postdocs & Visitors

Aline Gesualdi Manhães

Dan Hertz

Yadong Wang

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

Ben Walsh

Funding **NIH**, USDA

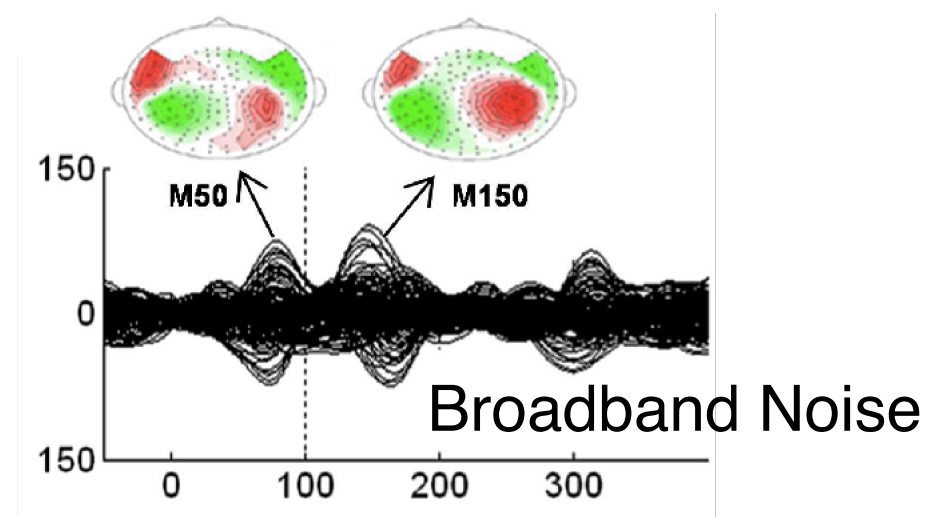
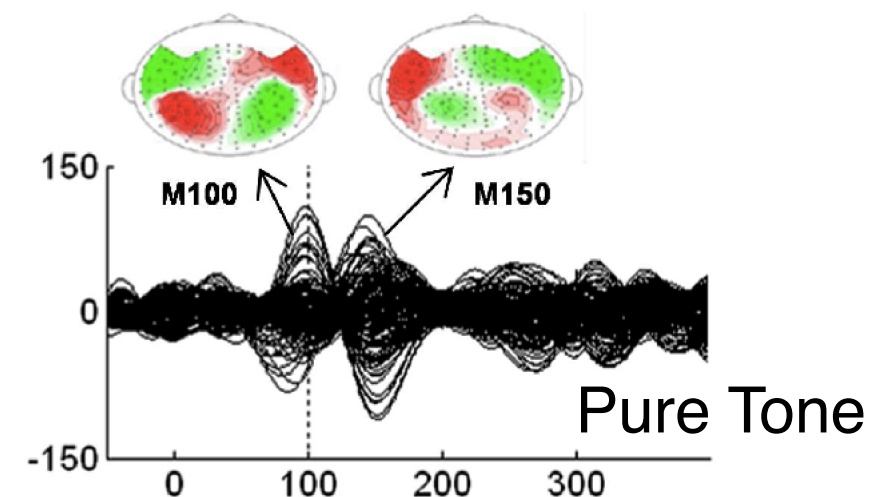
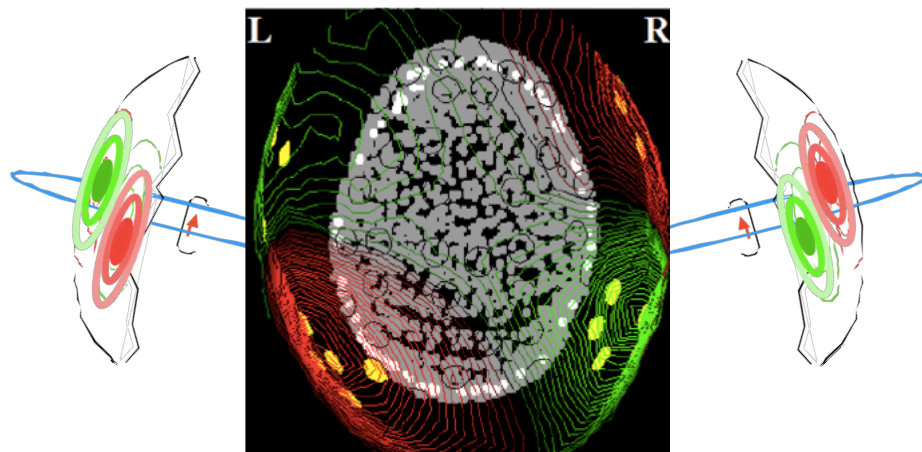
Outline

- Cortical Representations of Speech
(Encoding vs. Decoding)
- Attended vs. Unattended Speech
- New and Ongoing Studies:
 - ▶ Attentional Dynamics
 - ▶ Aging & Neural Representations of Speech
 - ▶ Neural Representations of the Background

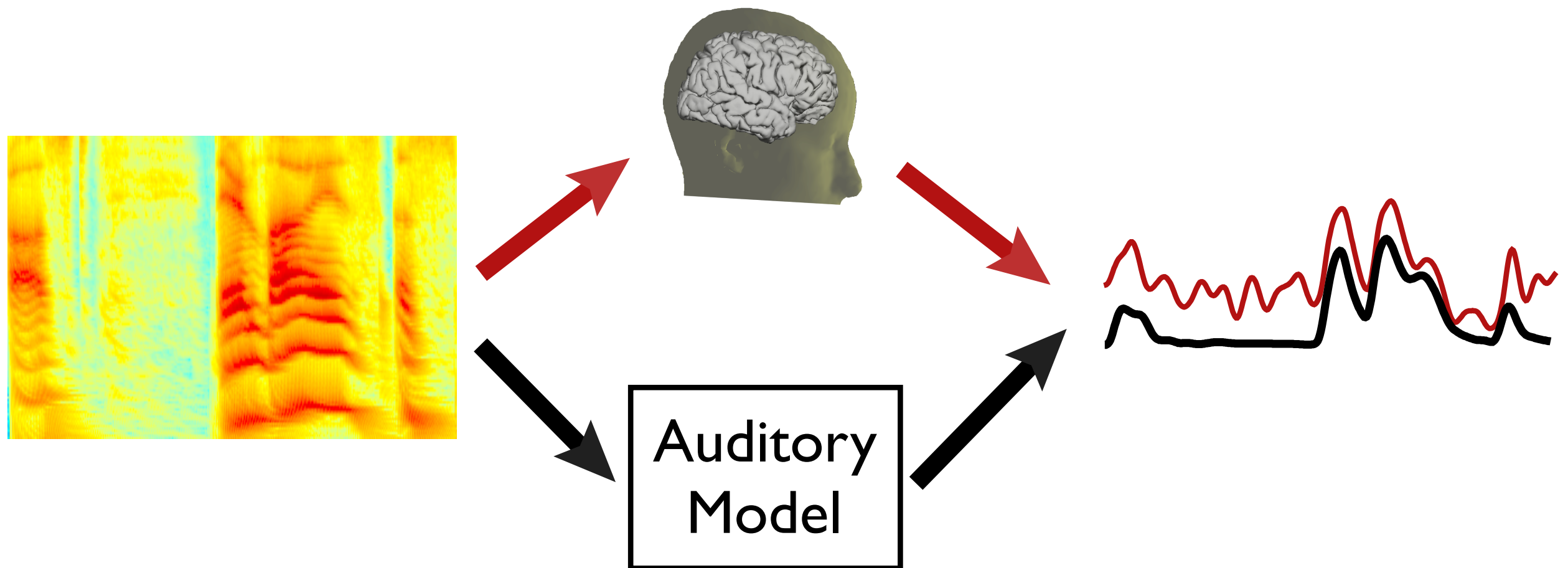
Time Course of MEG Responses

Auditory Evoked Responses

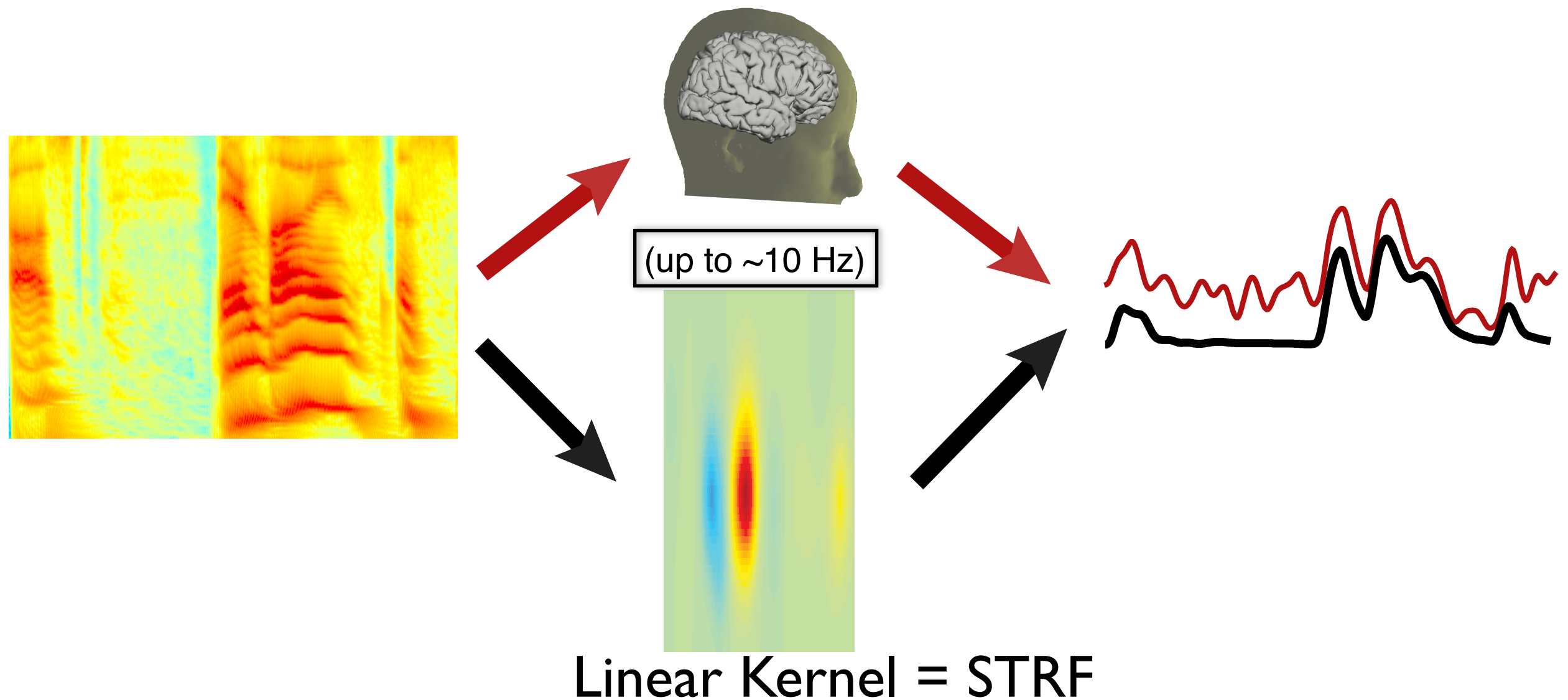
- Magnetoencephalography (MEG) Response Patterns Time-Locked to Stimulus Events
- Robust
- Strongly Lateralized



MEG Responses to Speech Modulations

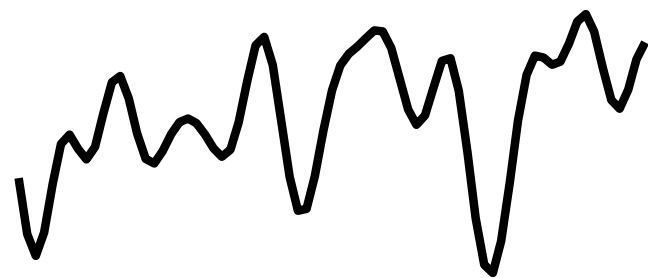


MEG Responses Predicted by STRF Model

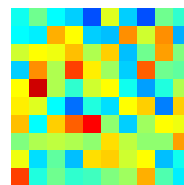


Neural Reconstruction of Speech Envelope

Speech Envelope

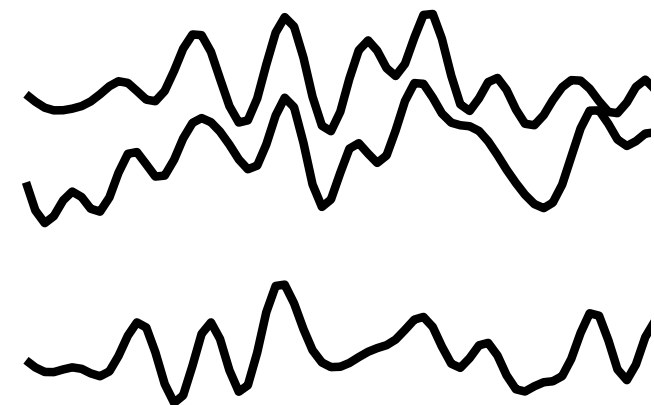


Decoder

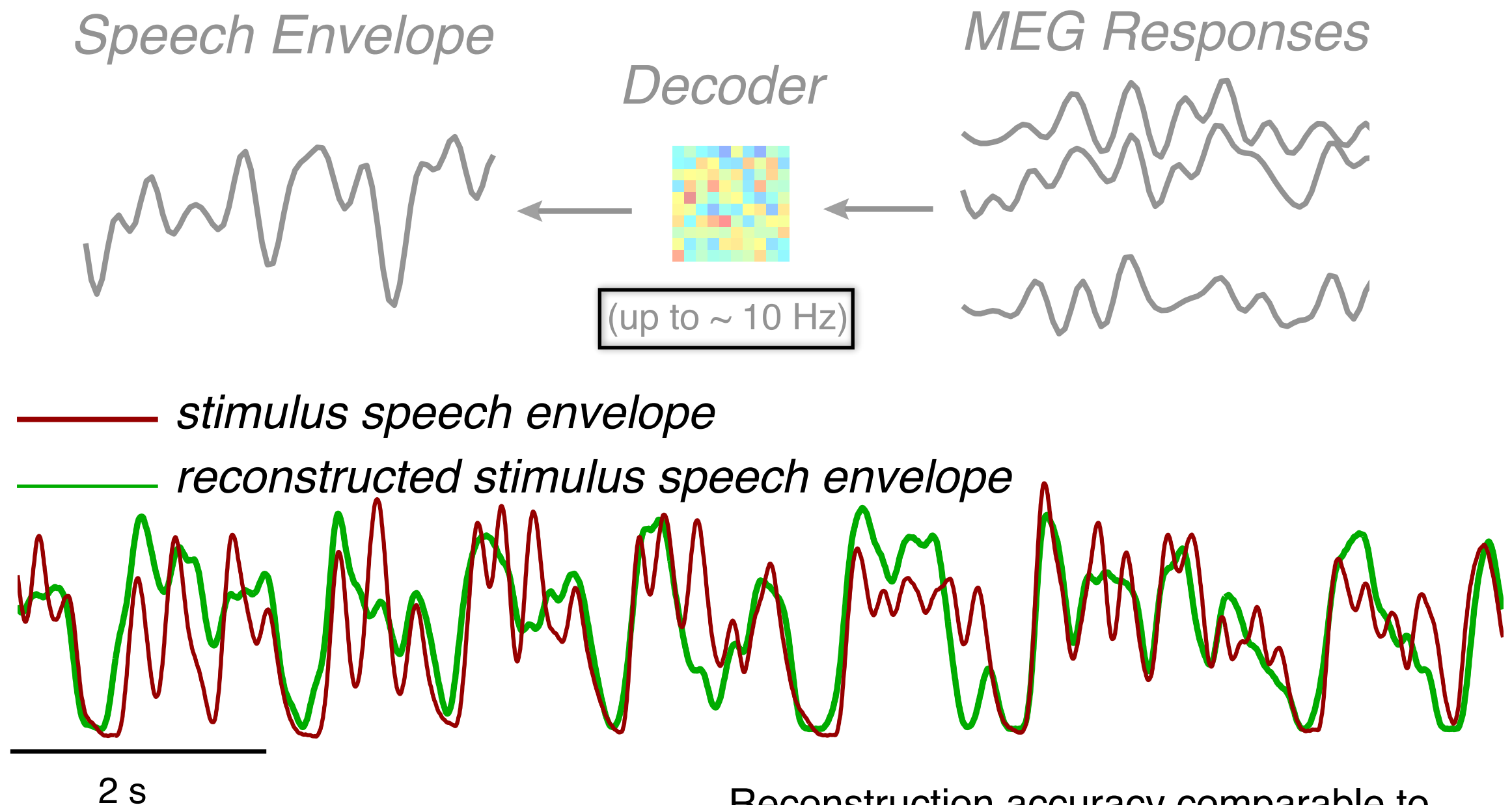


(up to ~ 10 Hz)

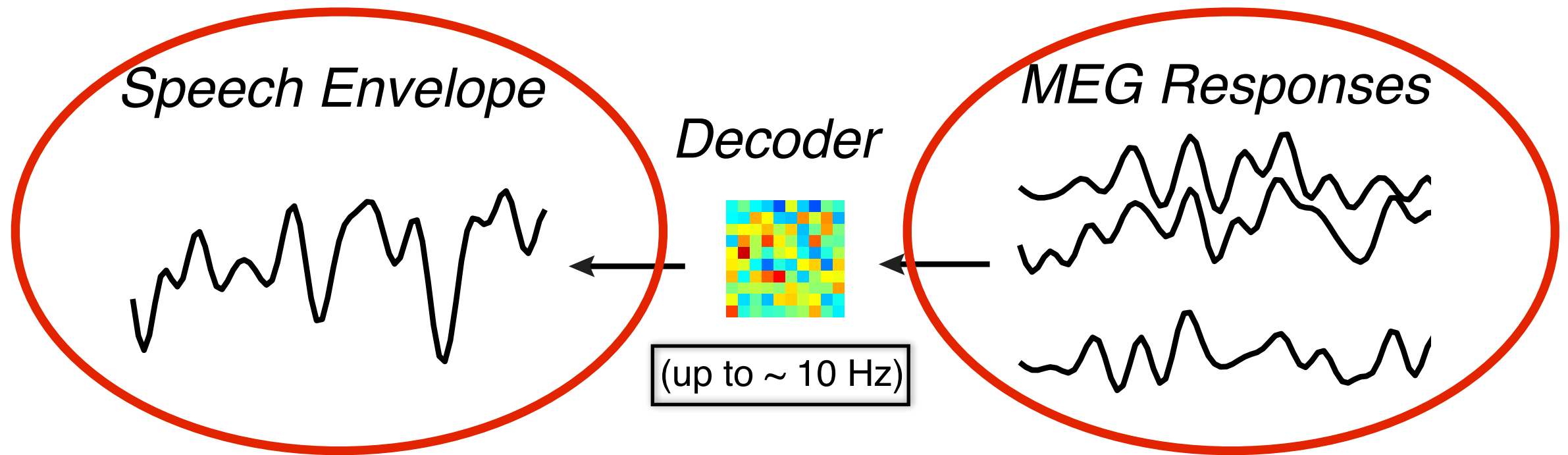
MEG Responses



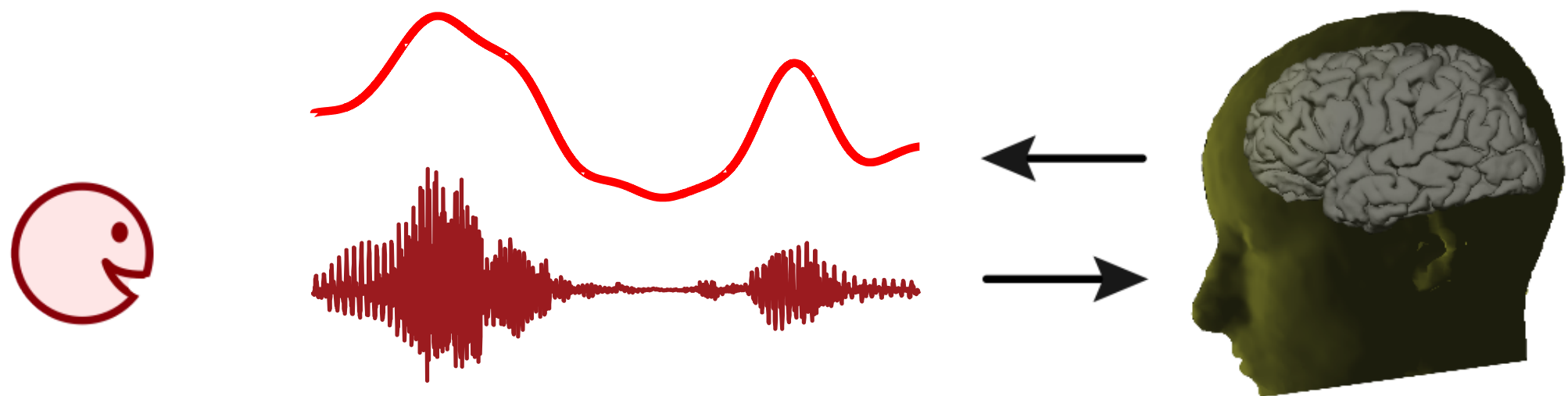
Neural Reconstruction of Speech Envelope



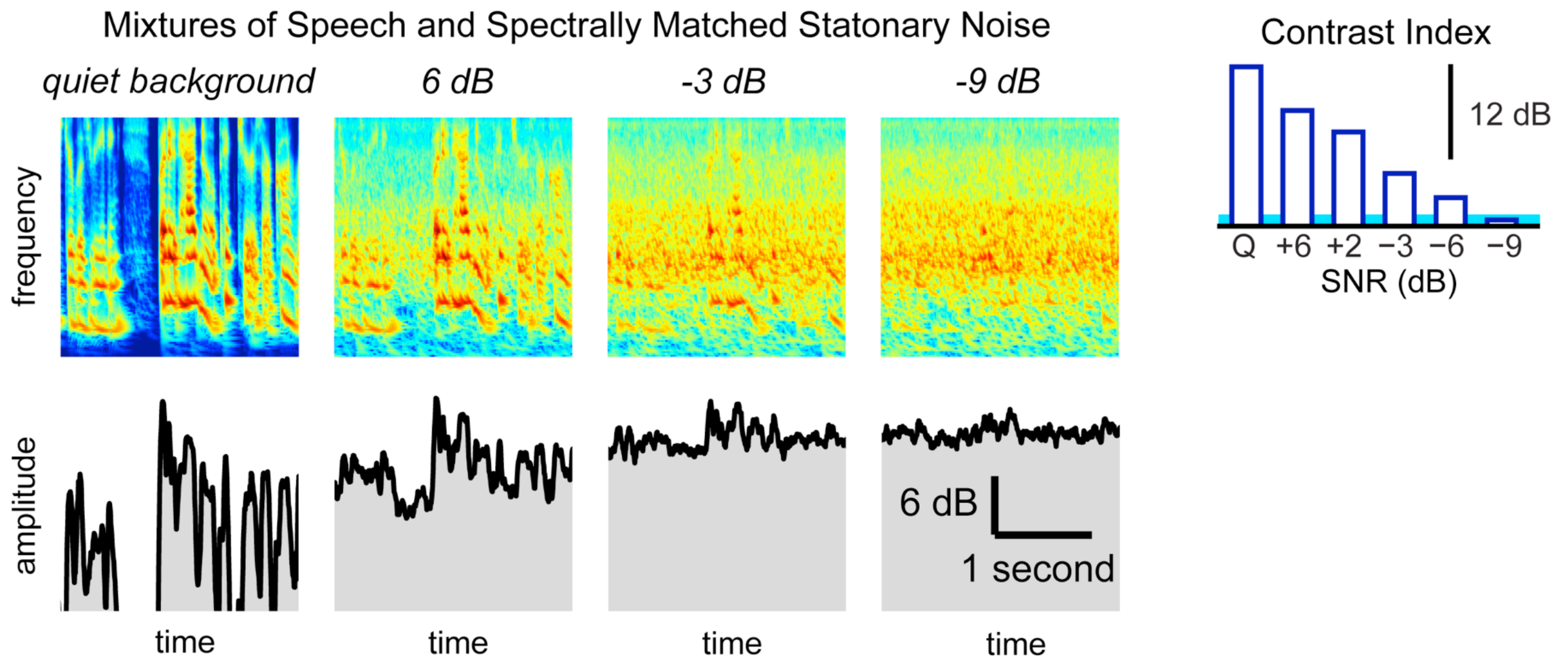
Reconstruction accuracy comparable to
single unit & ECoG recordings



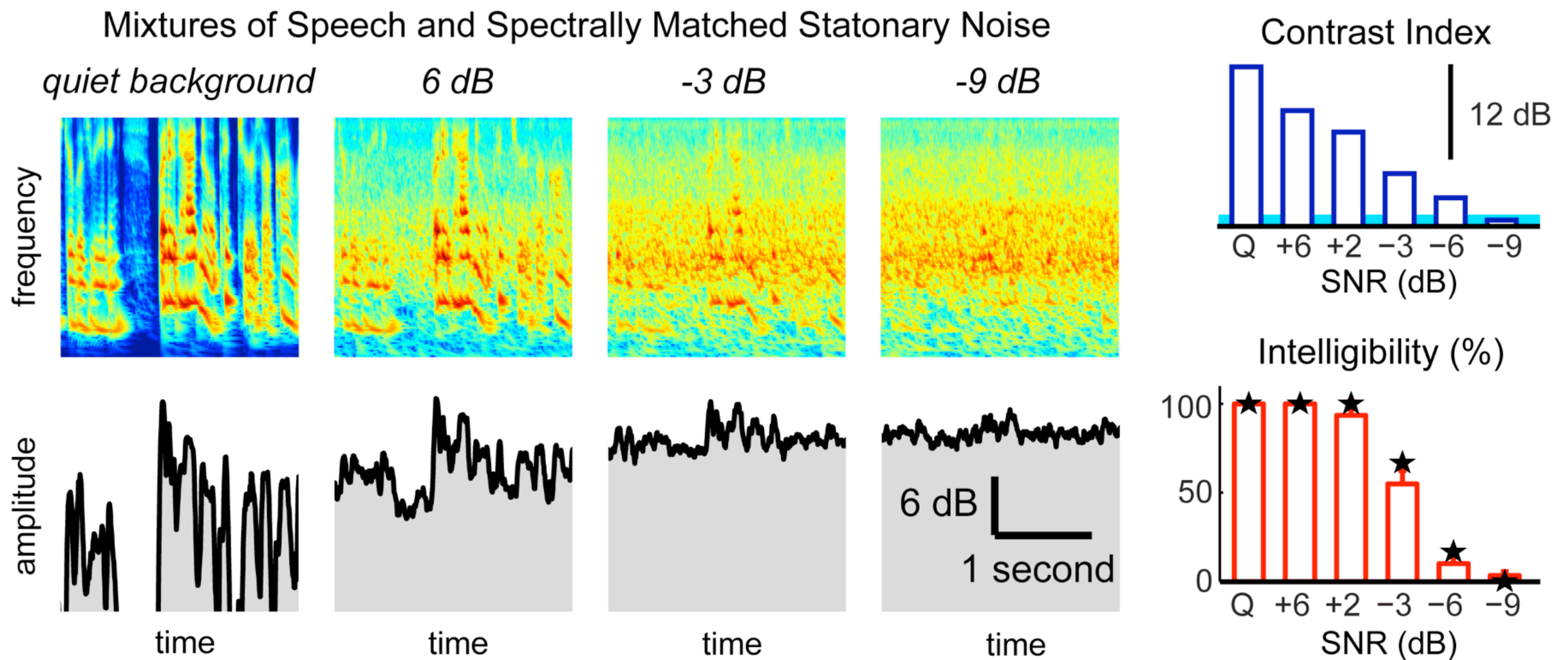
Neural Representation of Speech: Temporal



Speech in Noise

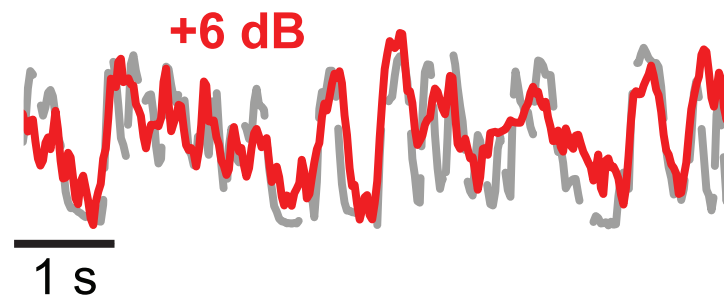


Speech in Noise



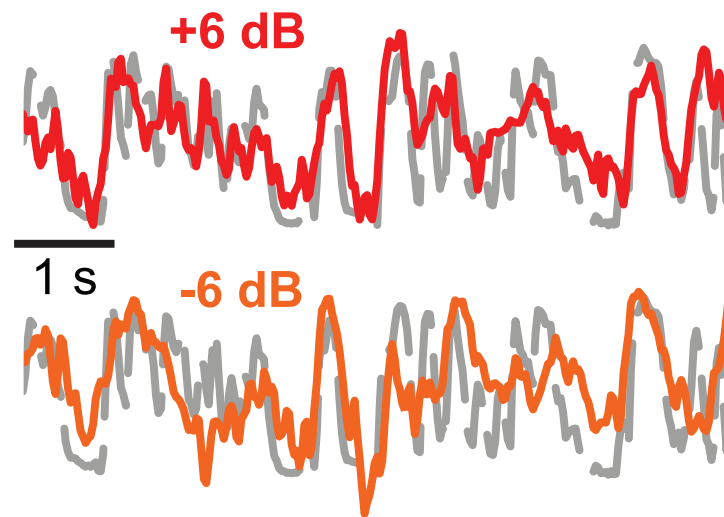
Speech in Noise: Results

Neural Reconstruction of
Underlying Speech Envelope



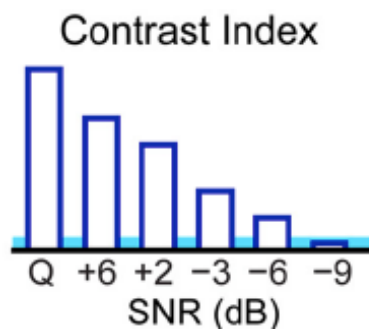
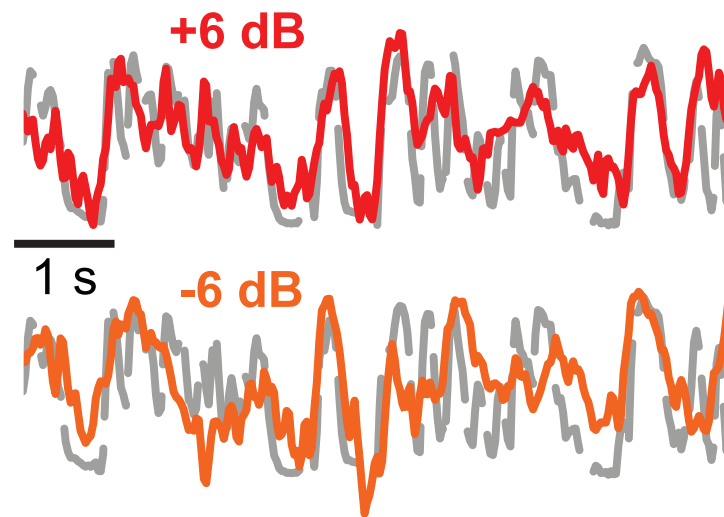
Speech in Noise: Results

Neural Reconstruction of
Underlying Speech Envelope



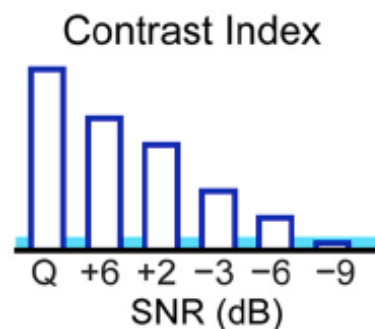
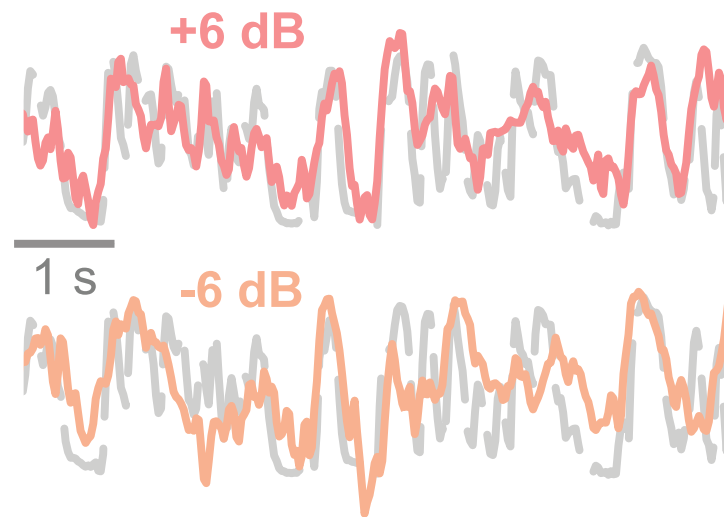
Speech in Noise: Results

Neural Reconstruction of
Underlying Speech Envelope

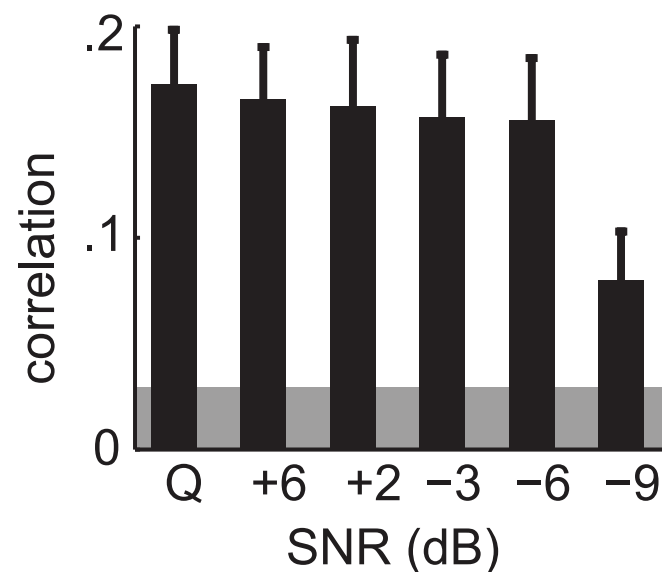


Speech in Noise: Results

Neural Reconstruction of
Underlying Speech Envelope

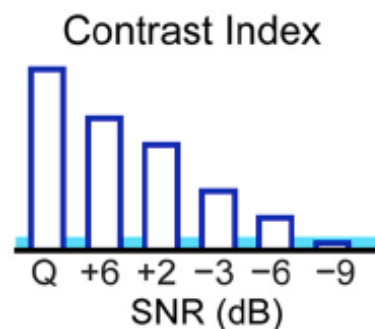
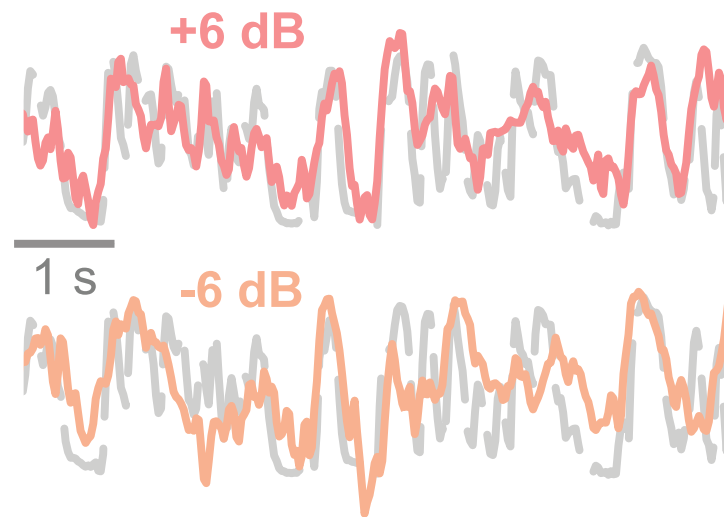


Reconstruction Accuracy

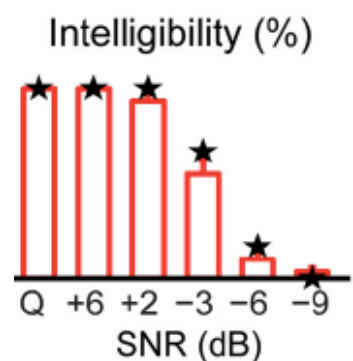
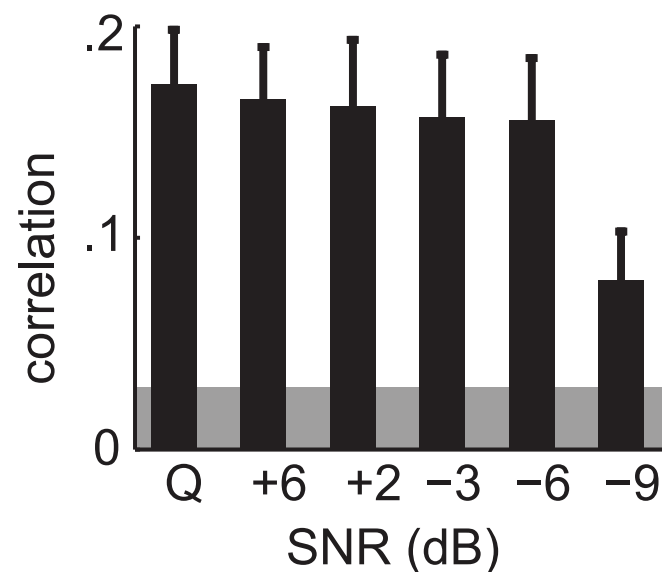


Speech in Noise: Results

Neural Reconstruction of Underlying Speech Envelope

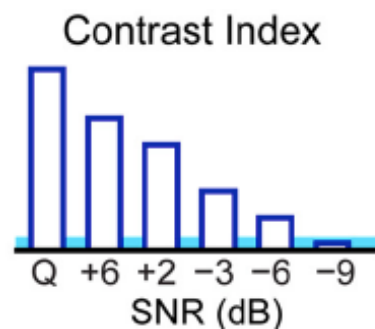
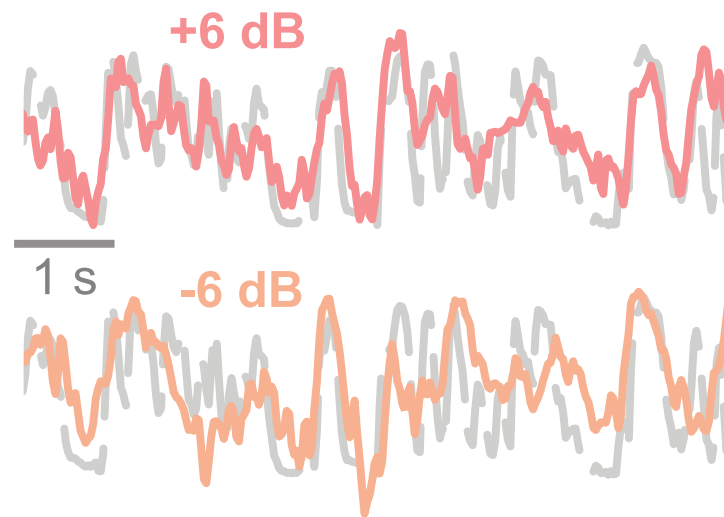


Reconstruction Accuracy

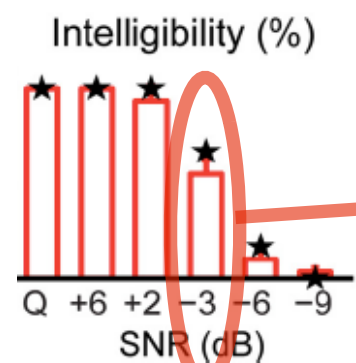
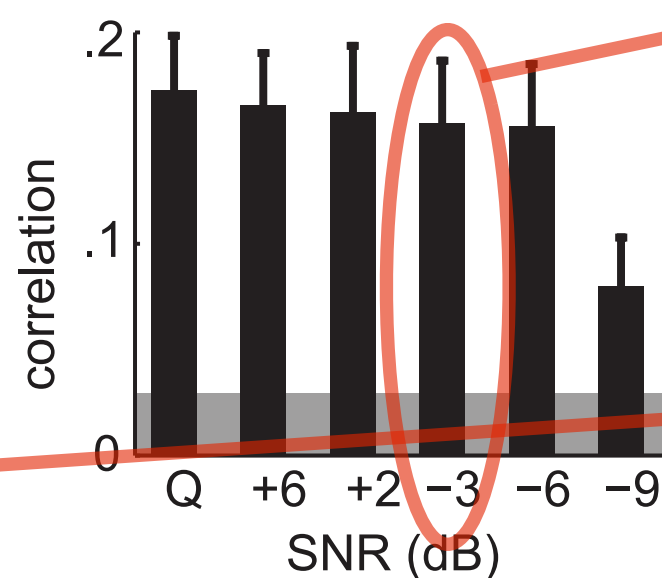


Speech in Noise: Results

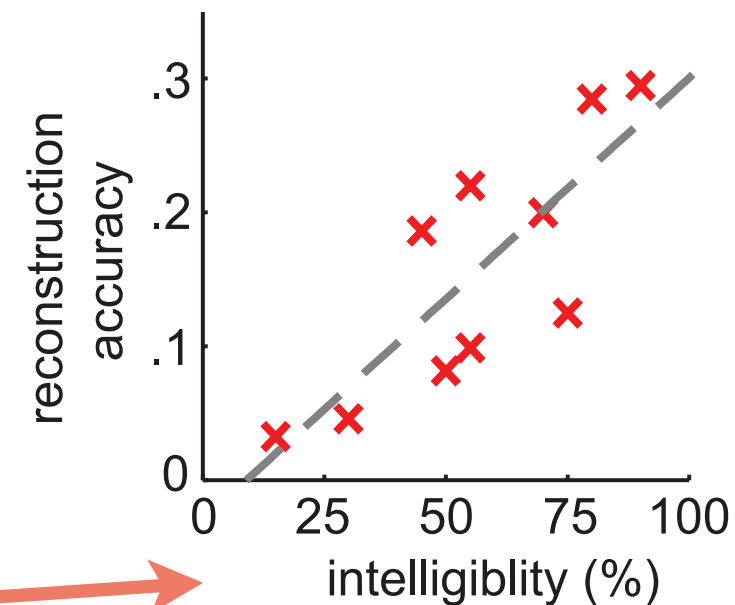
Neural Reconstruction of Underlying Speech Envelope



Reconstruction Accuracy



Correlation with Intelligibility



across Subjects

Cortical Speech Representations

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

Auditory Objects at the Cocktail Party



Alex Katz,
The Cocktail Party

Auditory Objects at the Cocktail Party



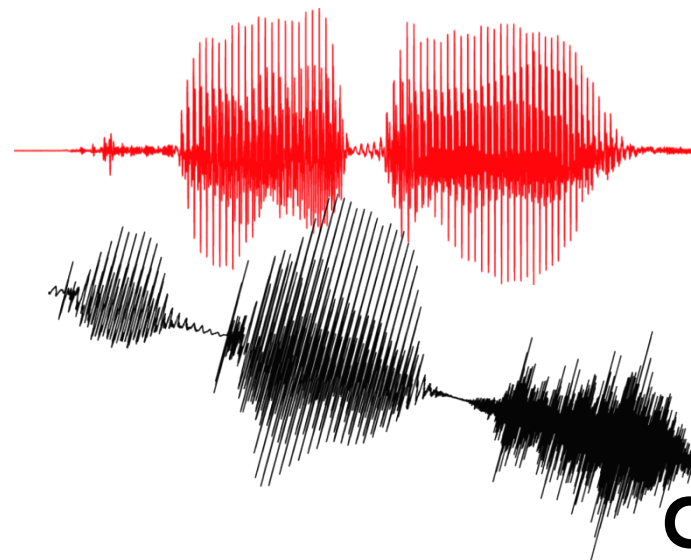
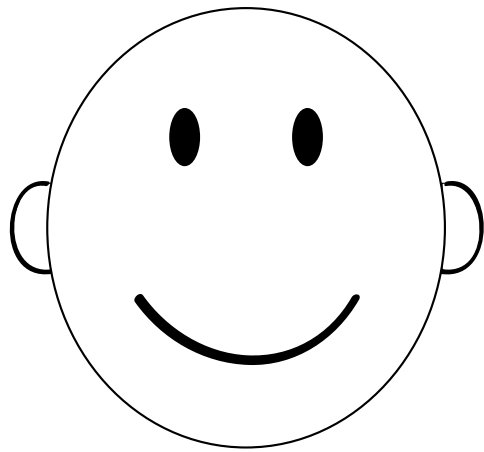
Alex Katz,
The Cocktail Party

Auditory Objects at the Cocktail Party



Alex Katz,
The Cocktail Party

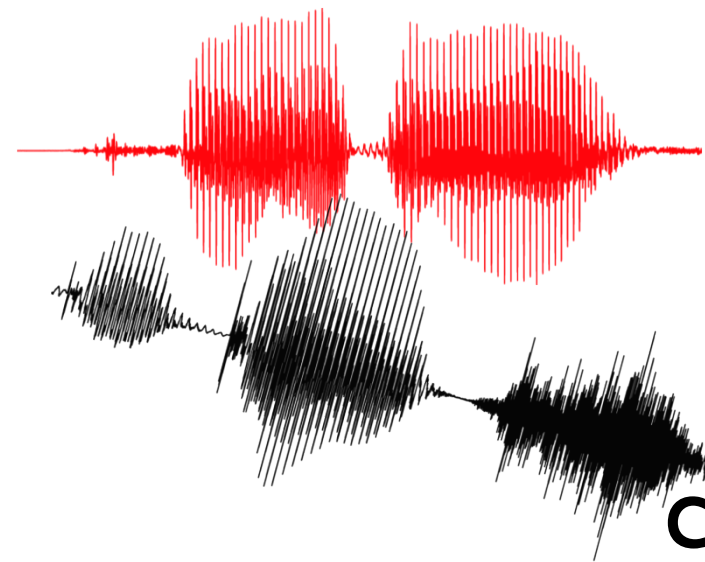
“Classic” Experiment



speech

competing speech

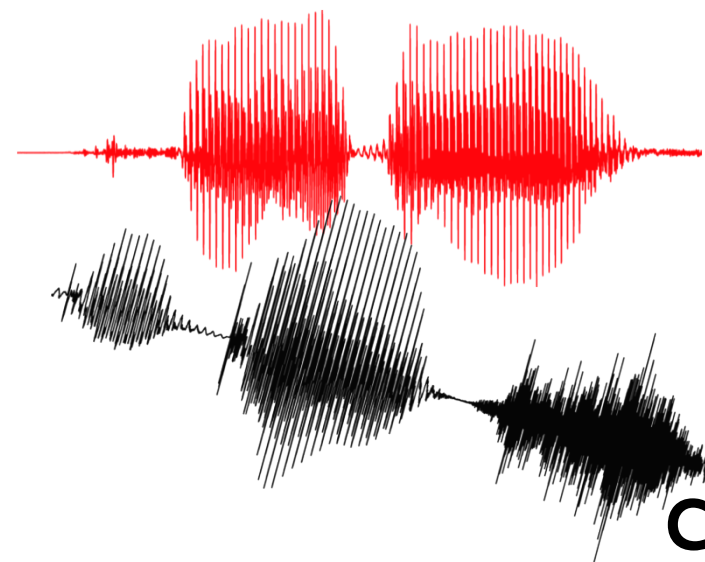
“Classic” Experiment



speech

competing speech

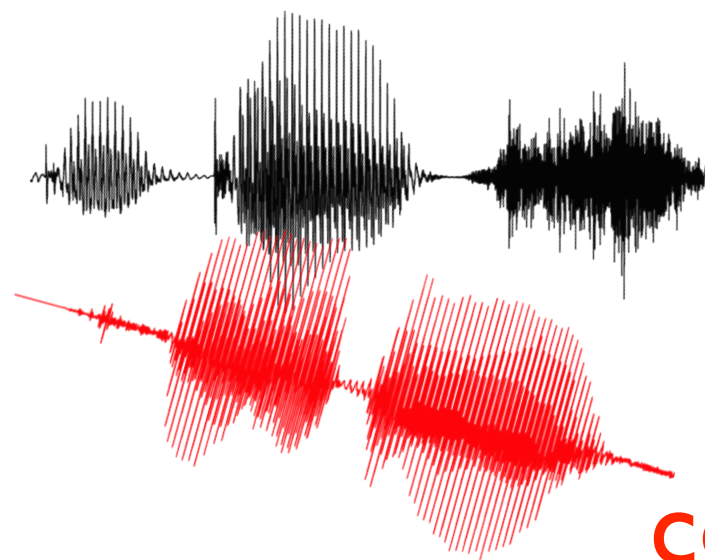
Recent Experiment



speech

competing speech

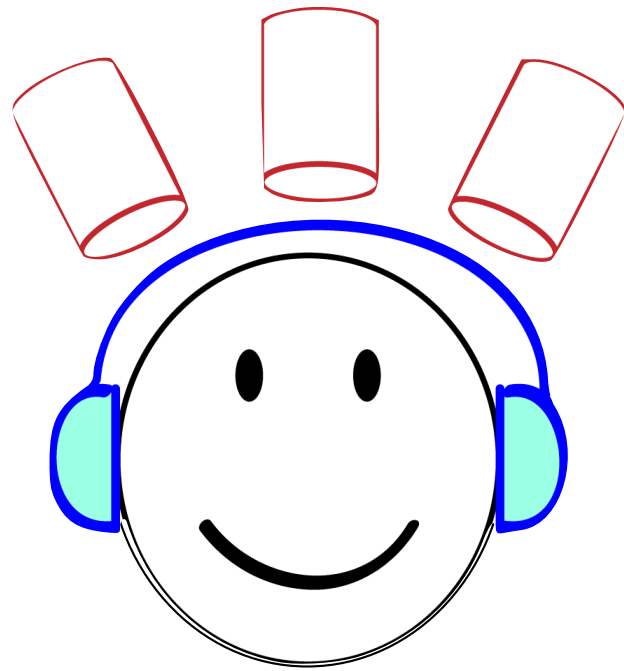
Recent Experiment



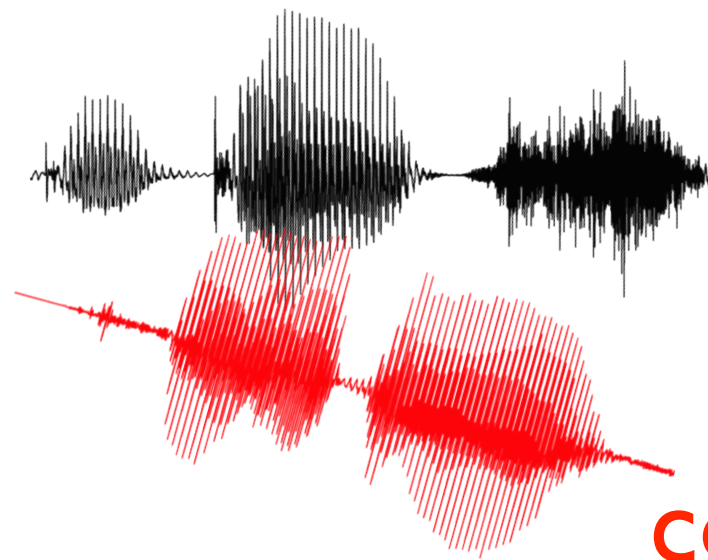
speech

competing speech

Recent Experiment



Attentional
Switch



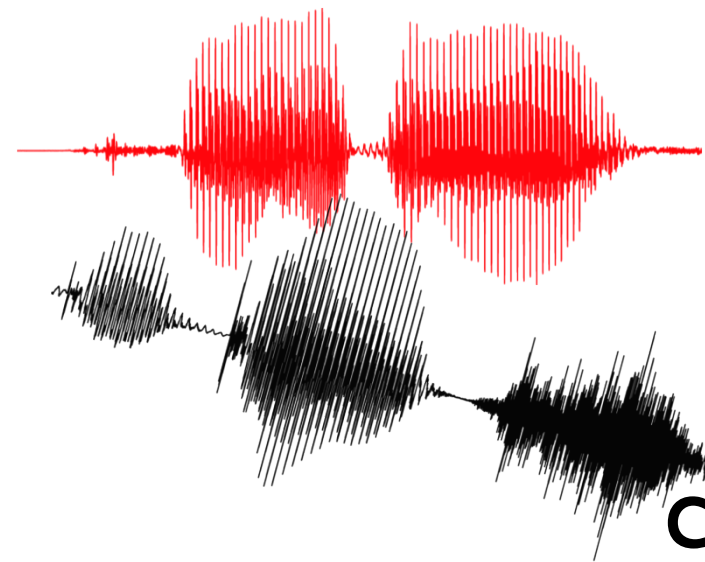
speech

competing speech

Recent Experiment



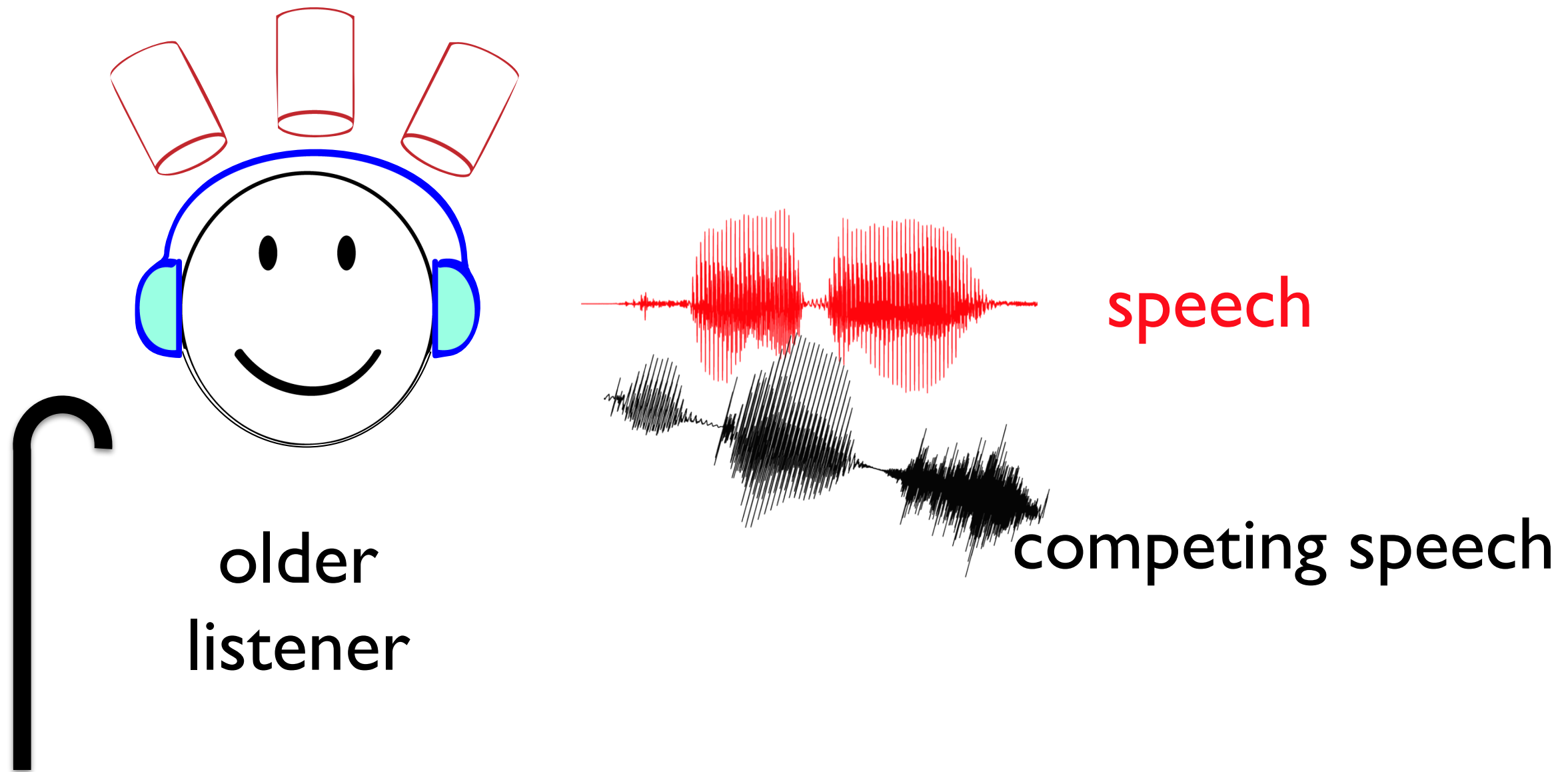
Attentional
Switch



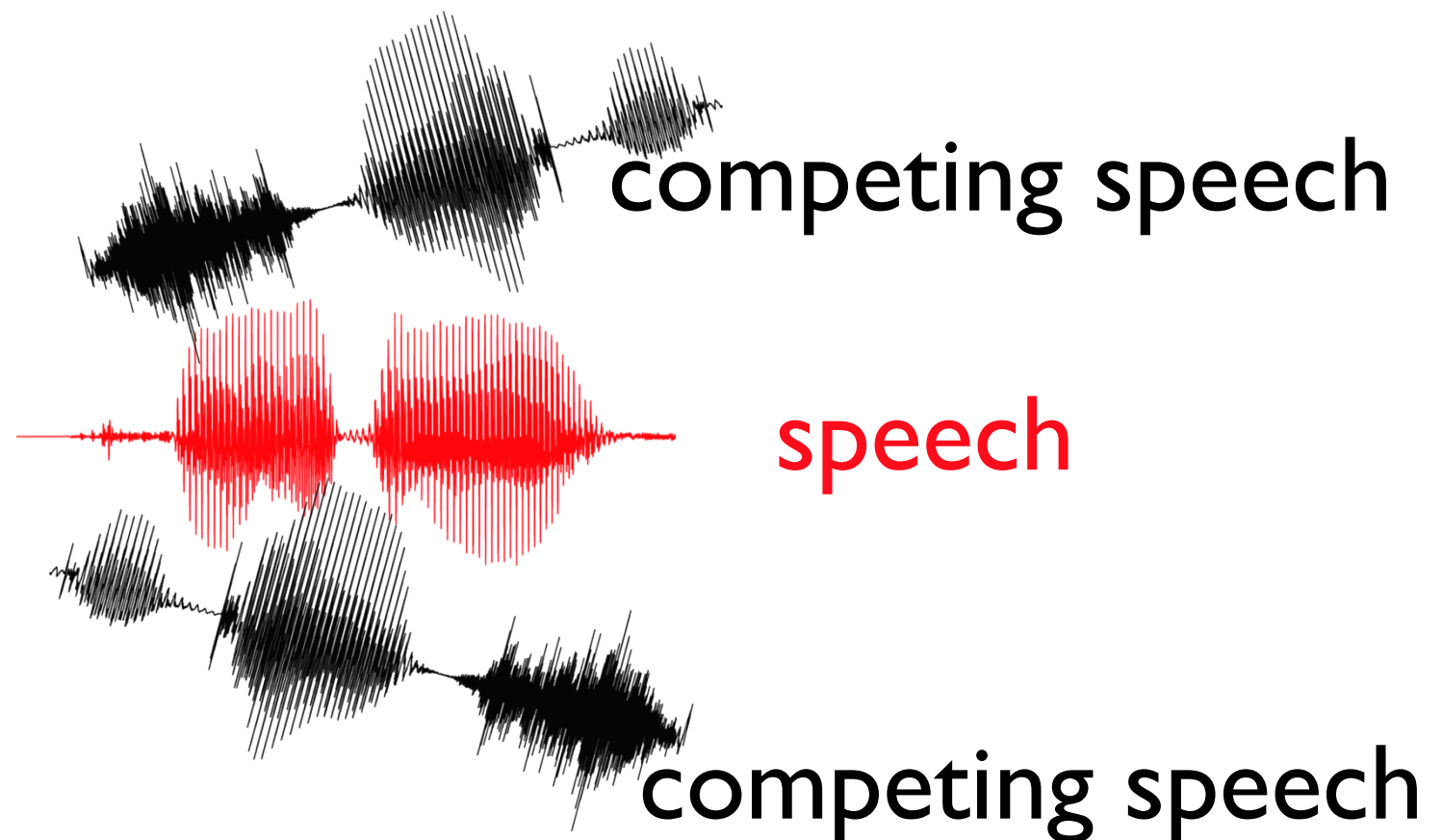
speech

competing speech

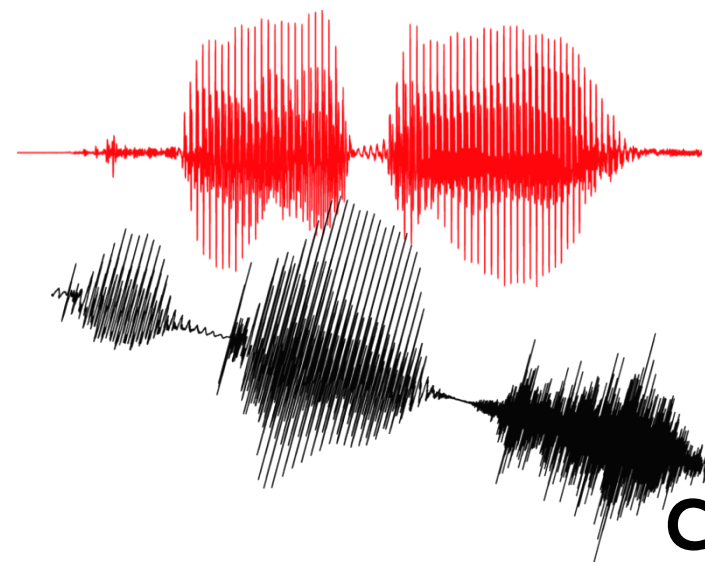
Experiments in Progress



Experiments in Progress



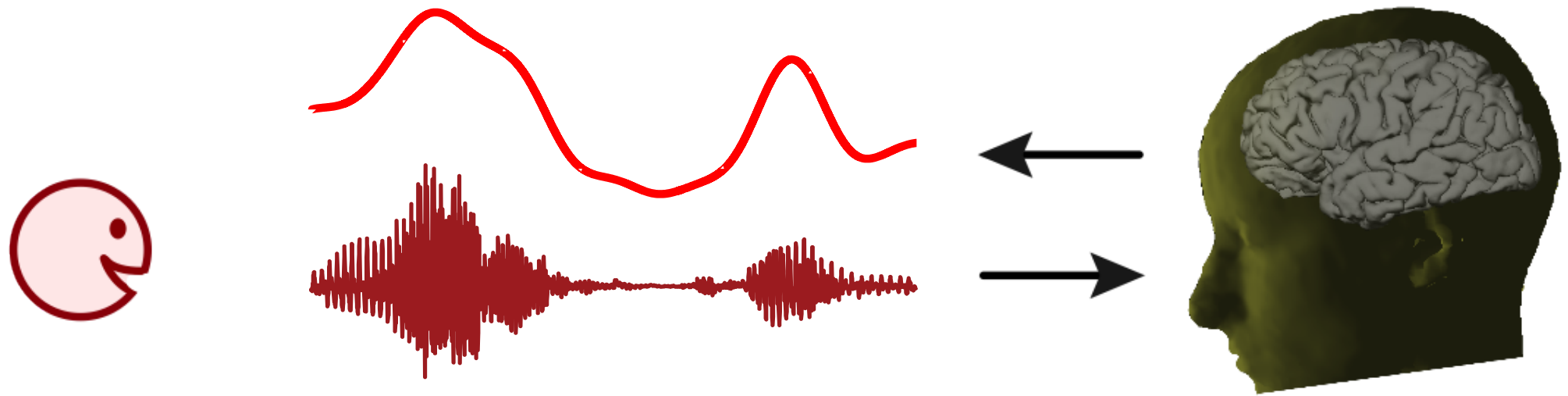
Two Competing Speakers



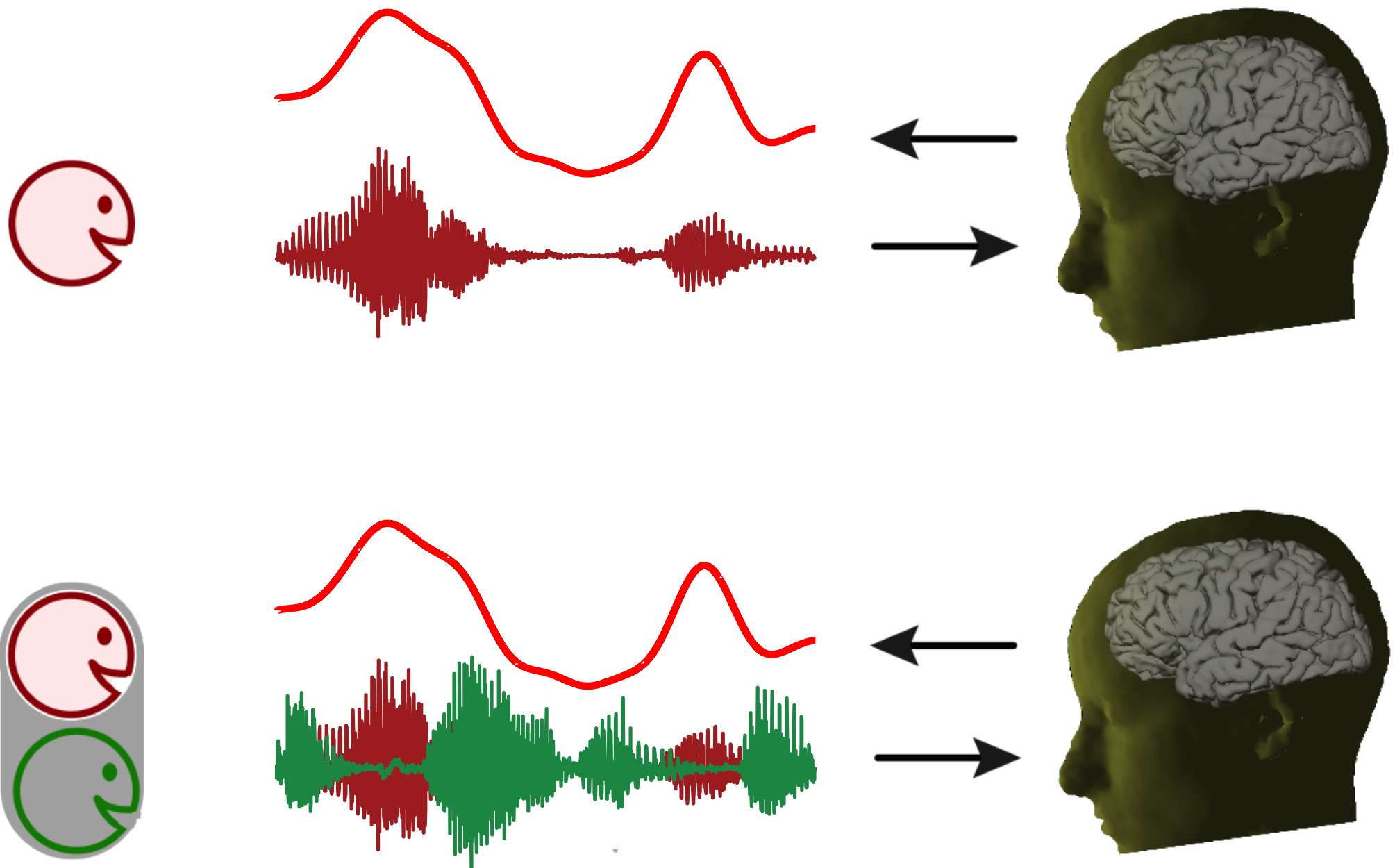
speech

competing speech

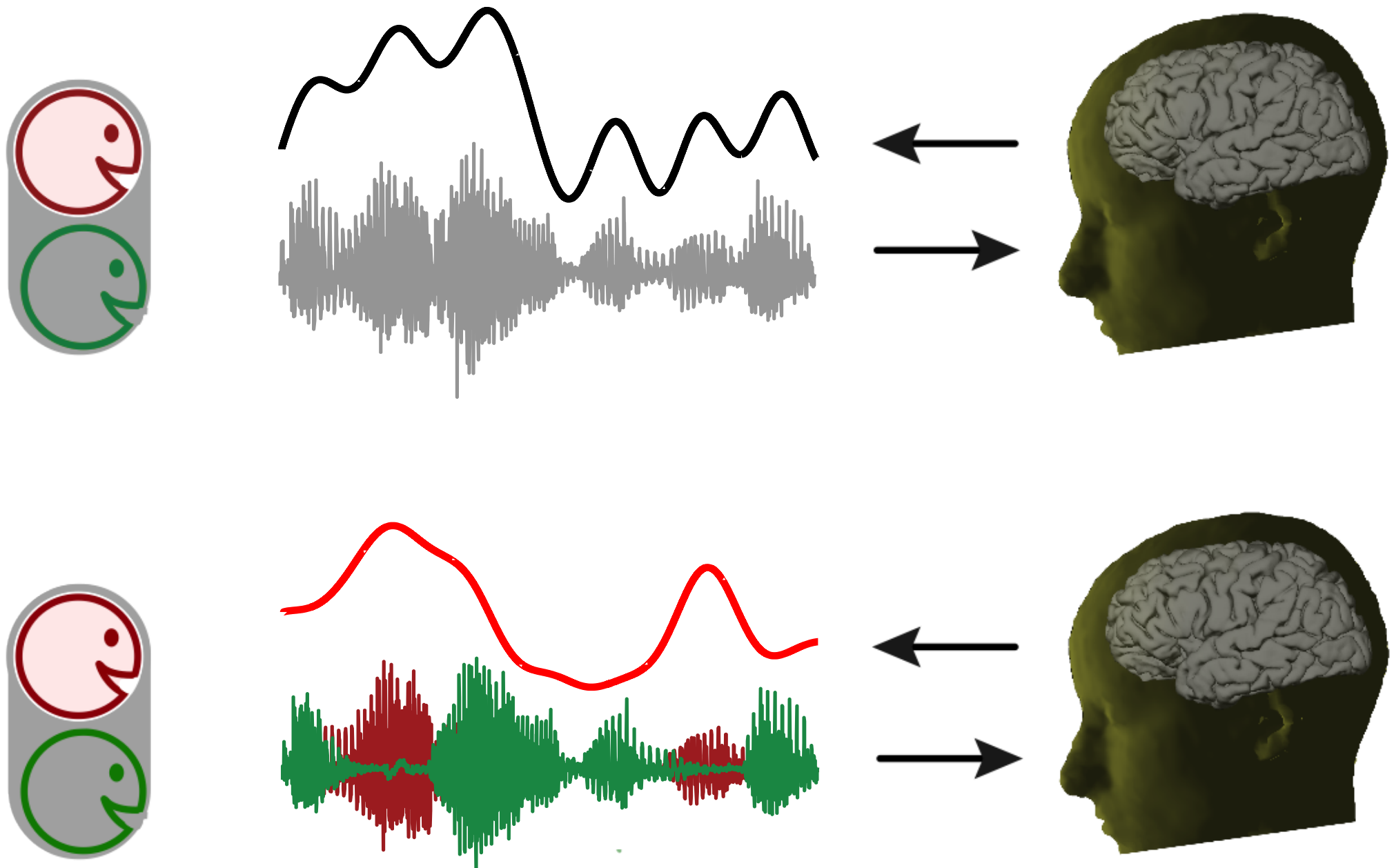
Selective Neural Encoding



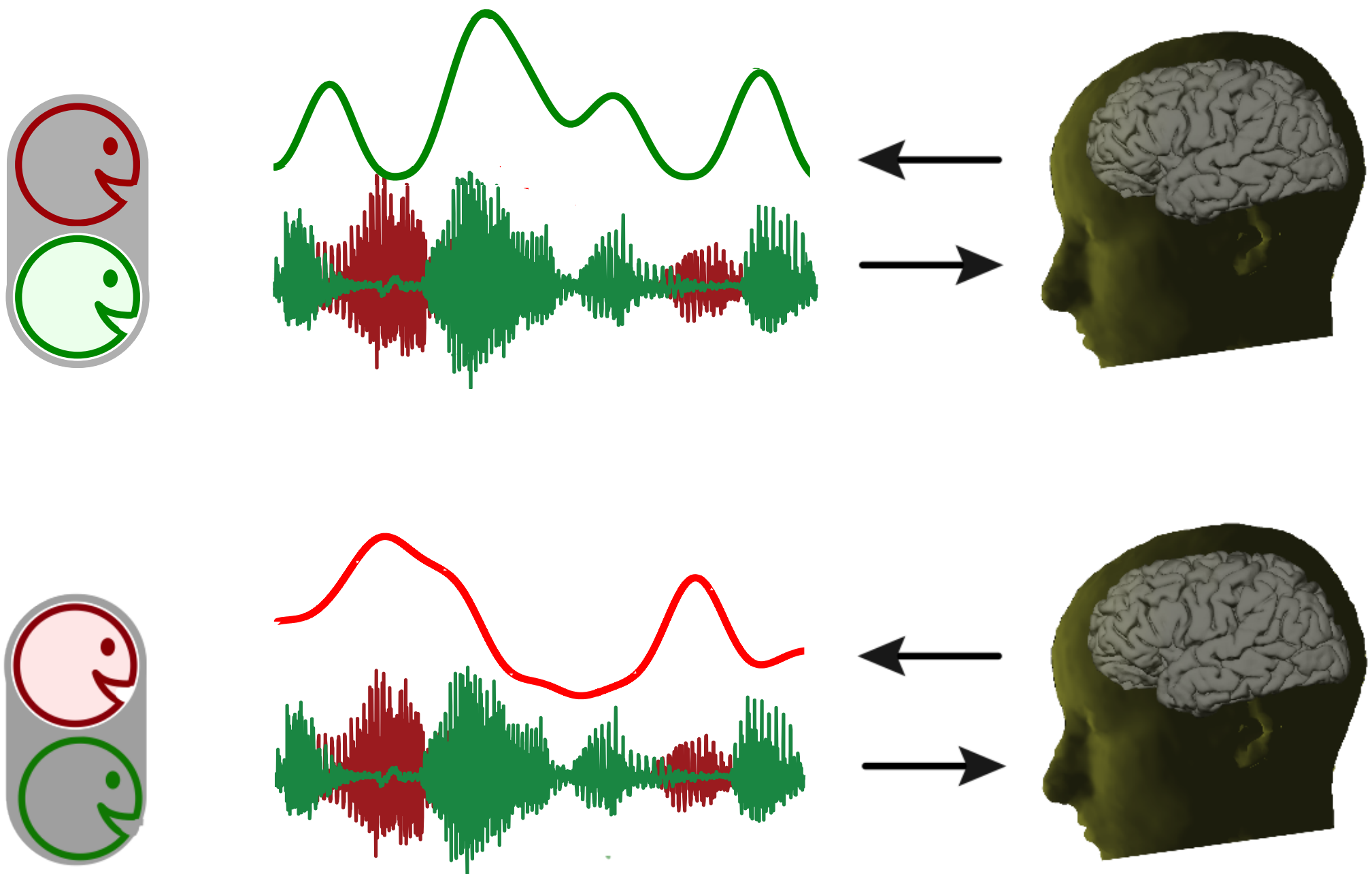
Selective Neural Encoding



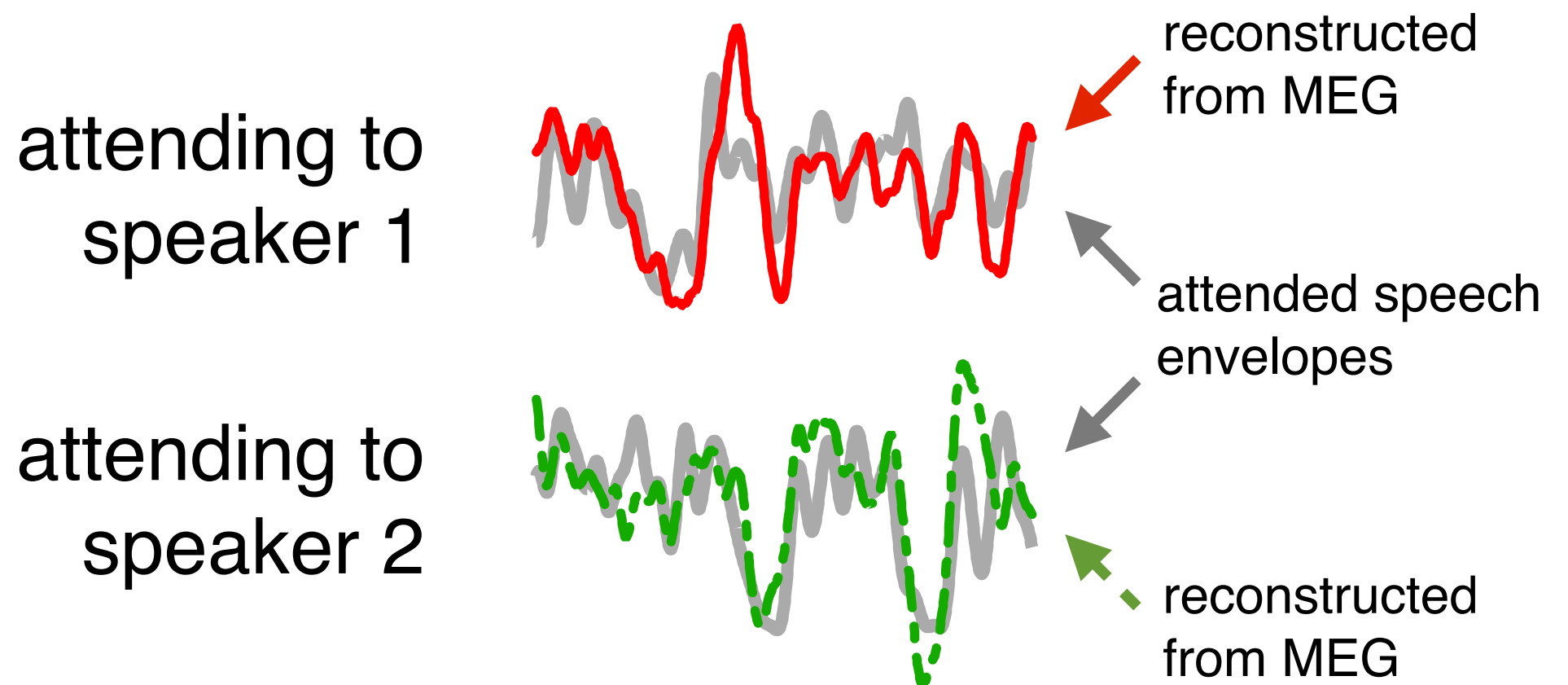
Unselective vs. Selective Neural Encoding



Selective Neural Encoding

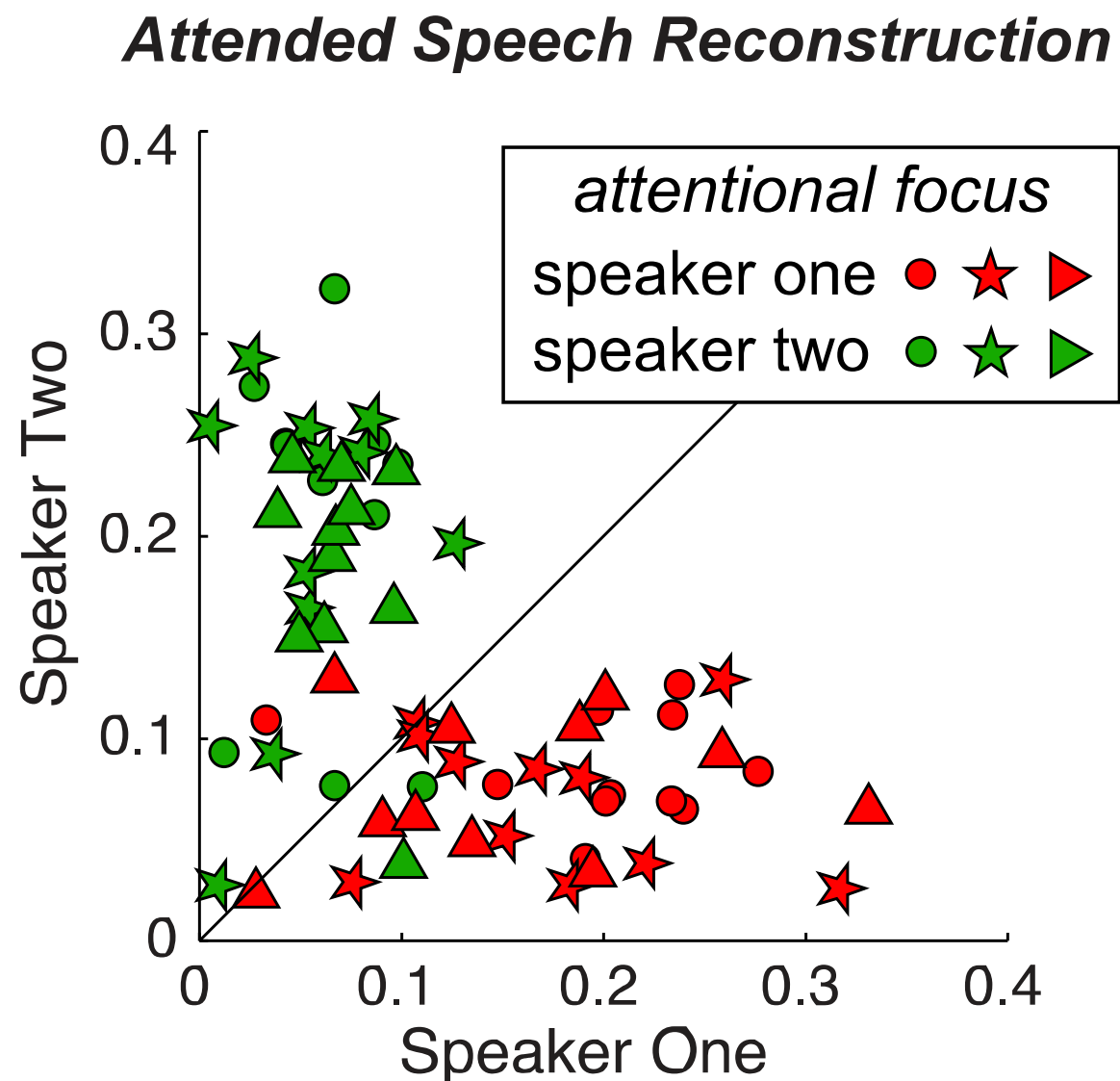


Stream-Specific Representation



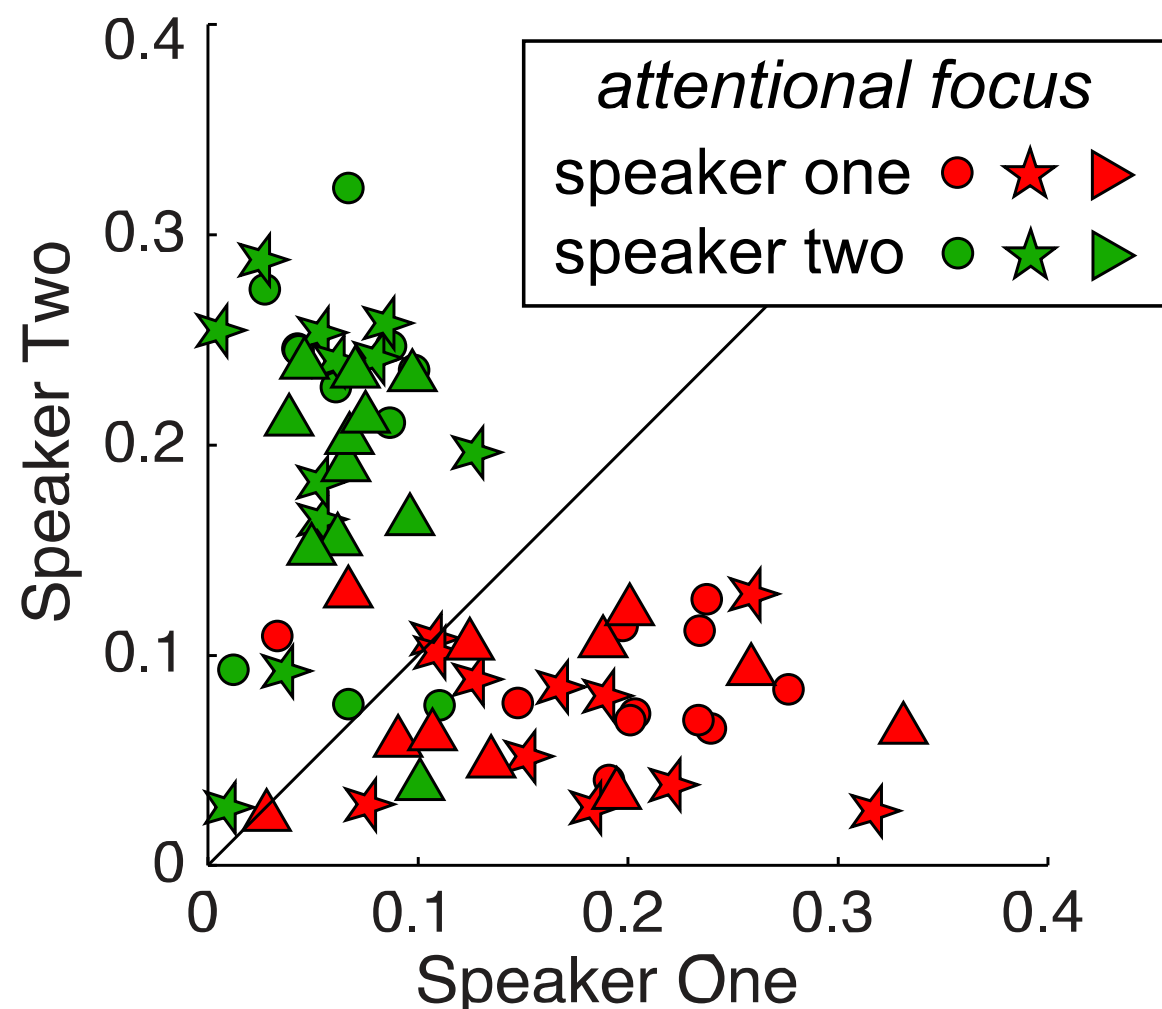
Identical Stimuli!

Single Trial Speech Reconstruction

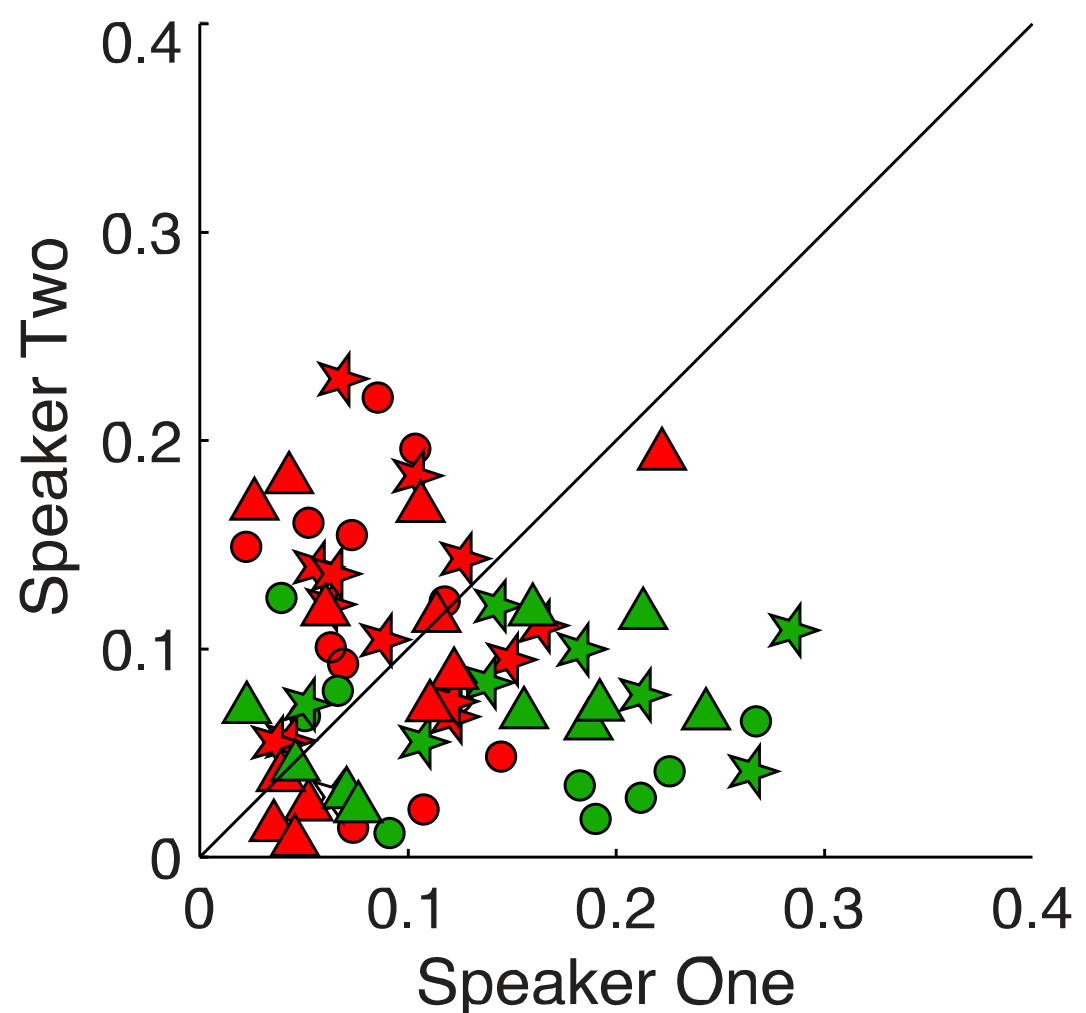


Single Trial Speech Reconstruction

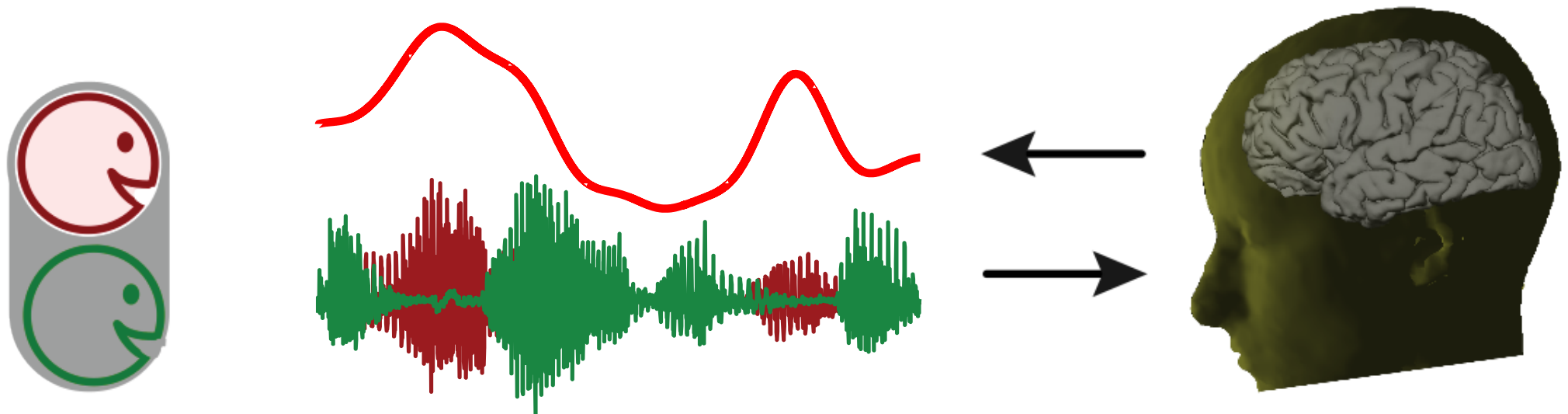
Attended Speech Reconstruction



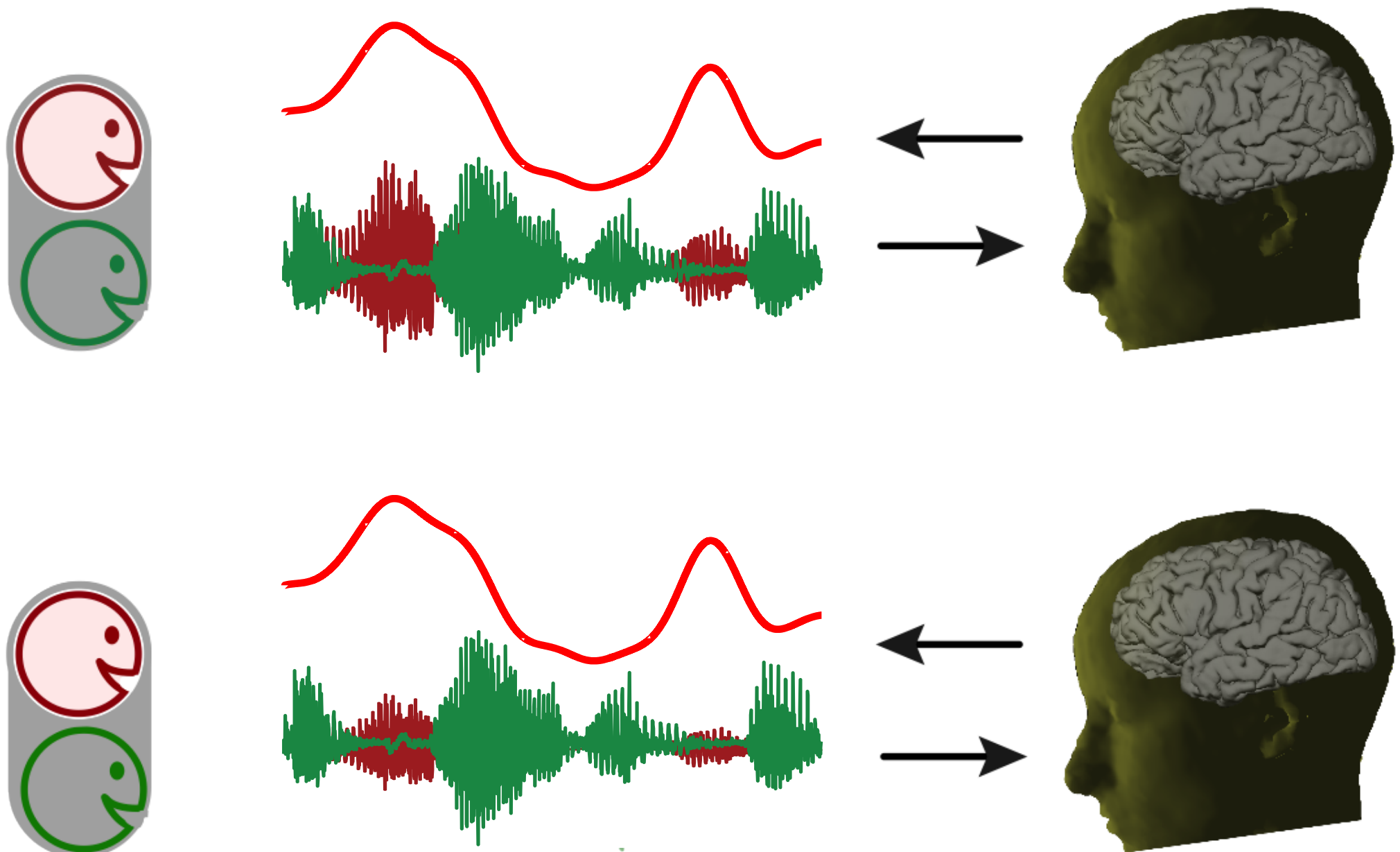
Background Speech Reconstruction



Invariance Under Relative Loudness Change?

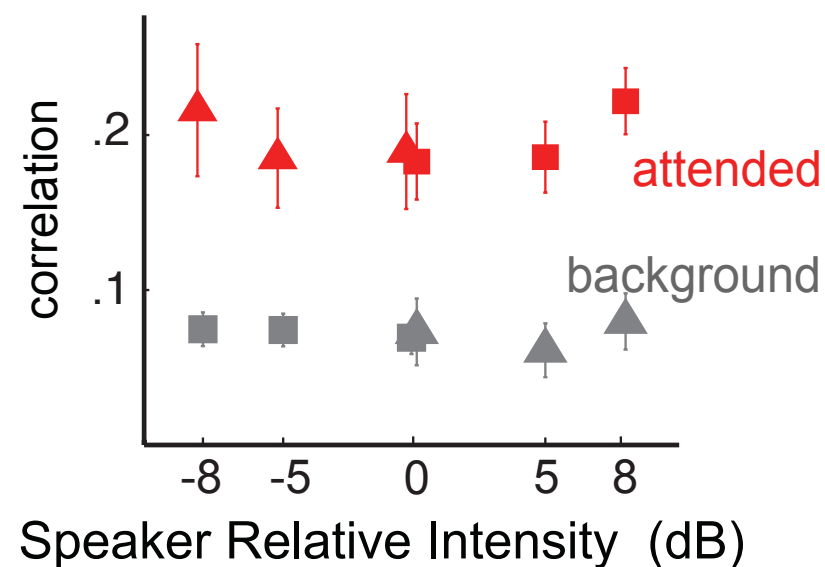


Invariance Under Relative Loudness Change?



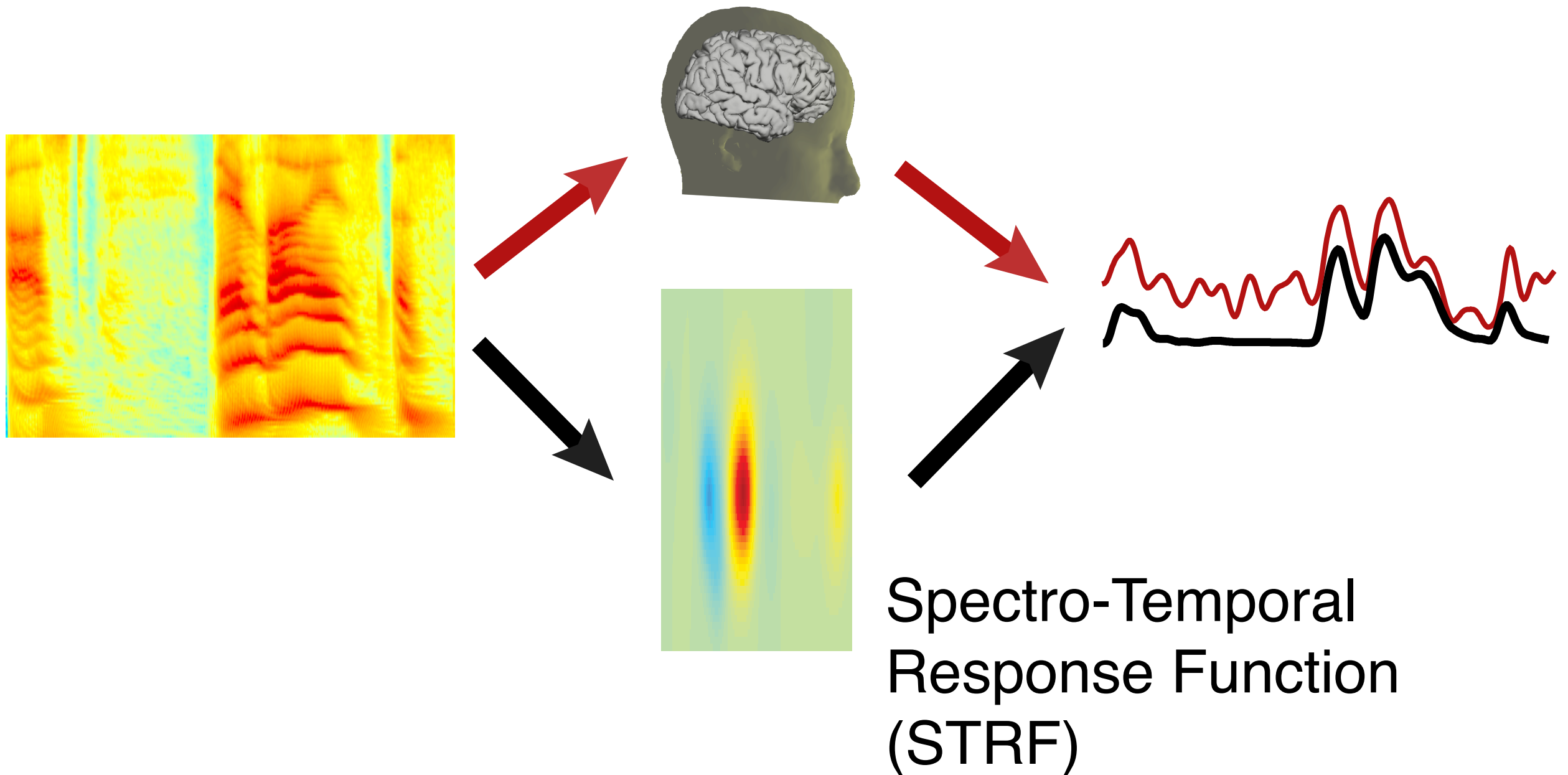
Invariance under Relative Loudness Change

Neural Results

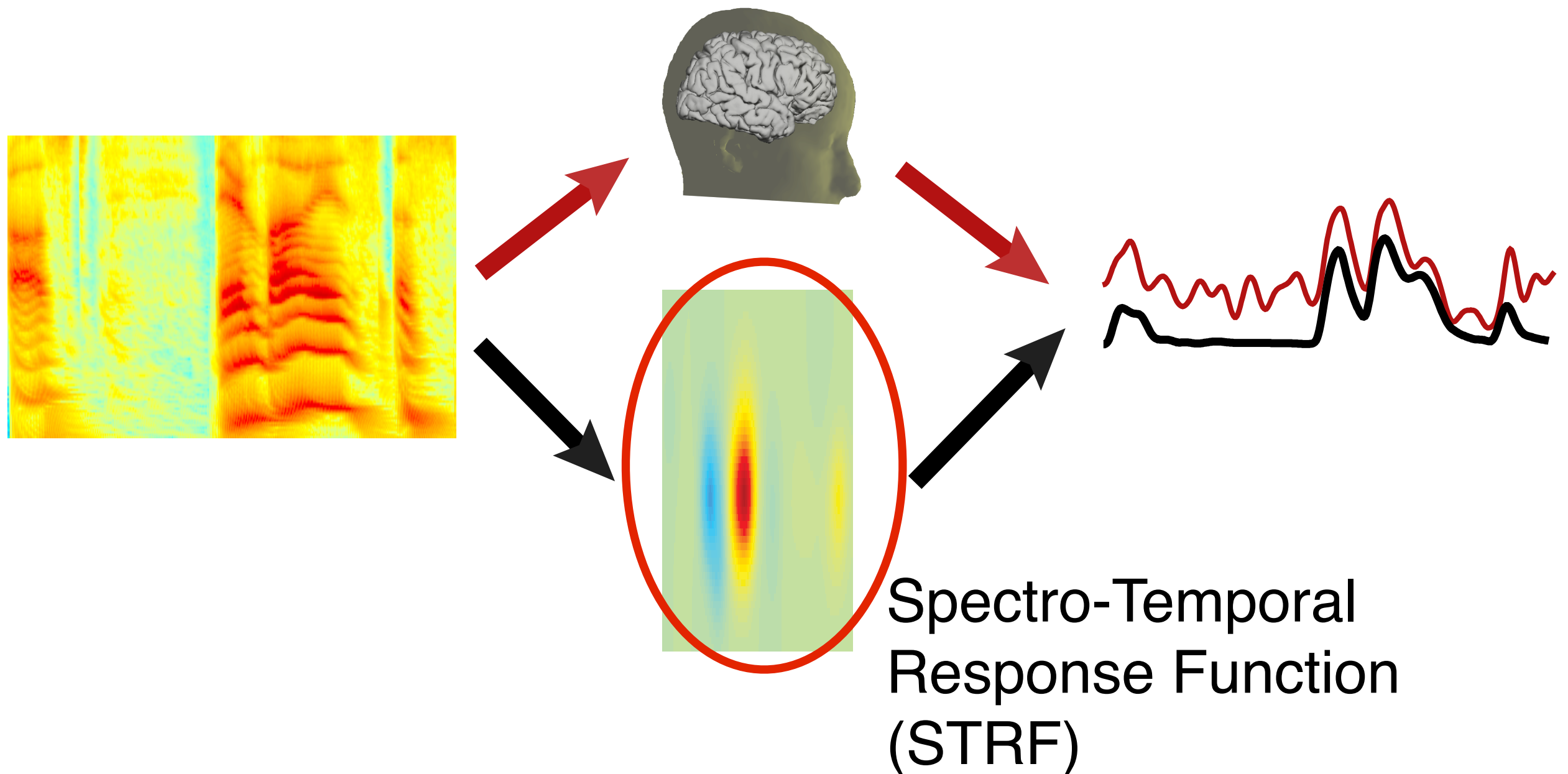


- Neural representation invariant to relative loudness change
- Stream-based Gain Control, not stimulus-based

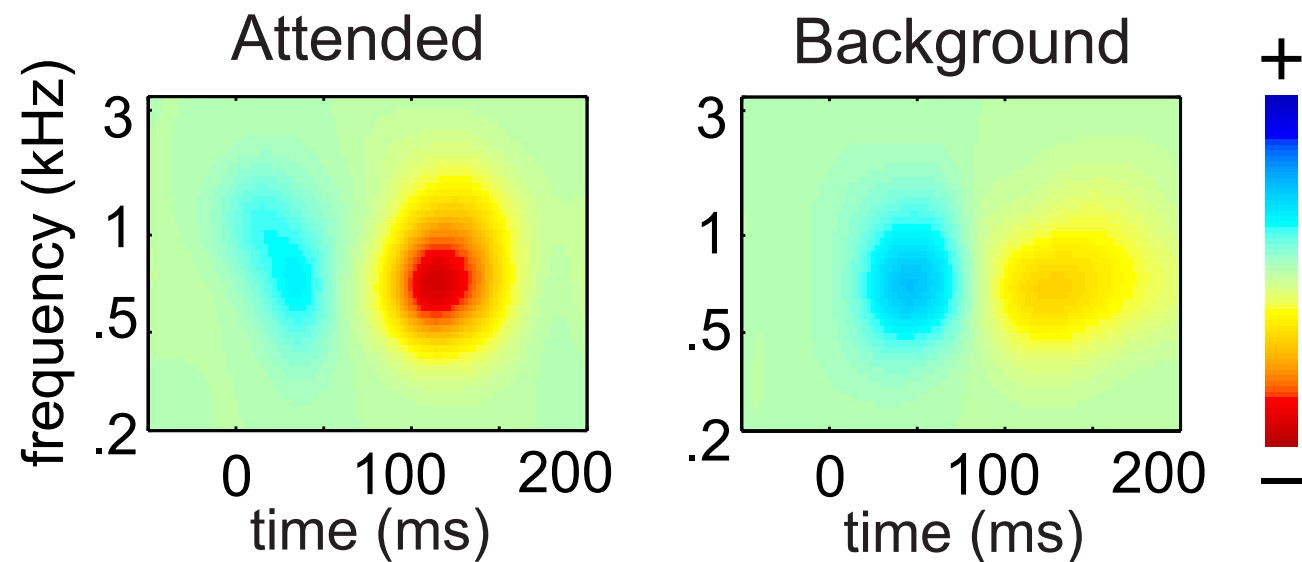
Forward STRF Model



Forward STRF Model

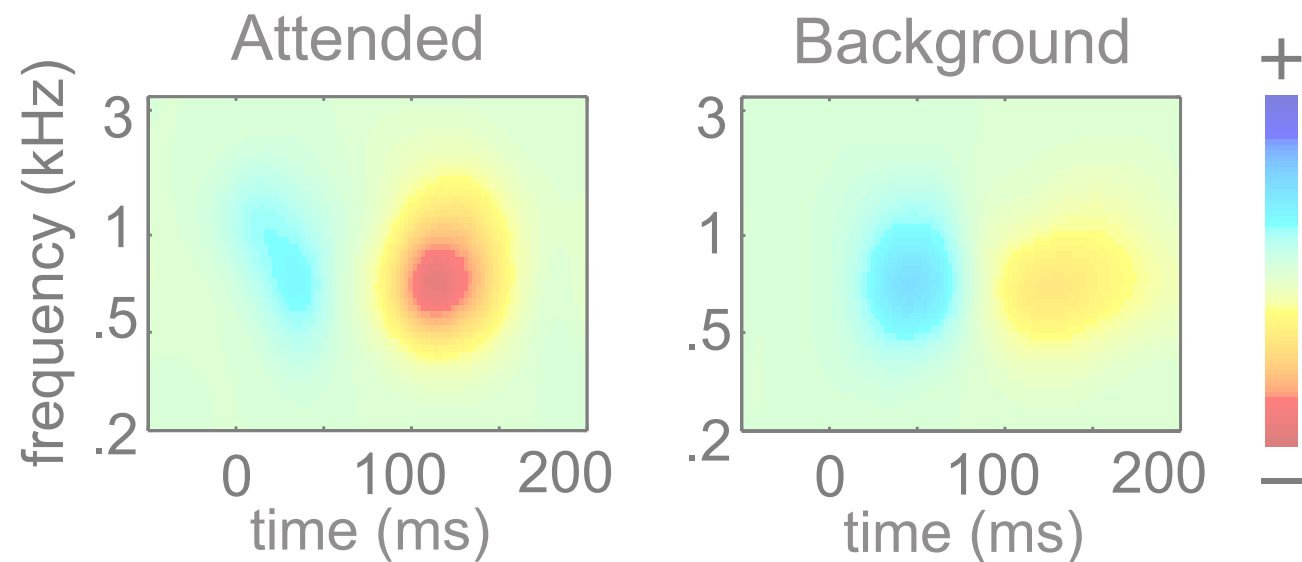


STRF Results

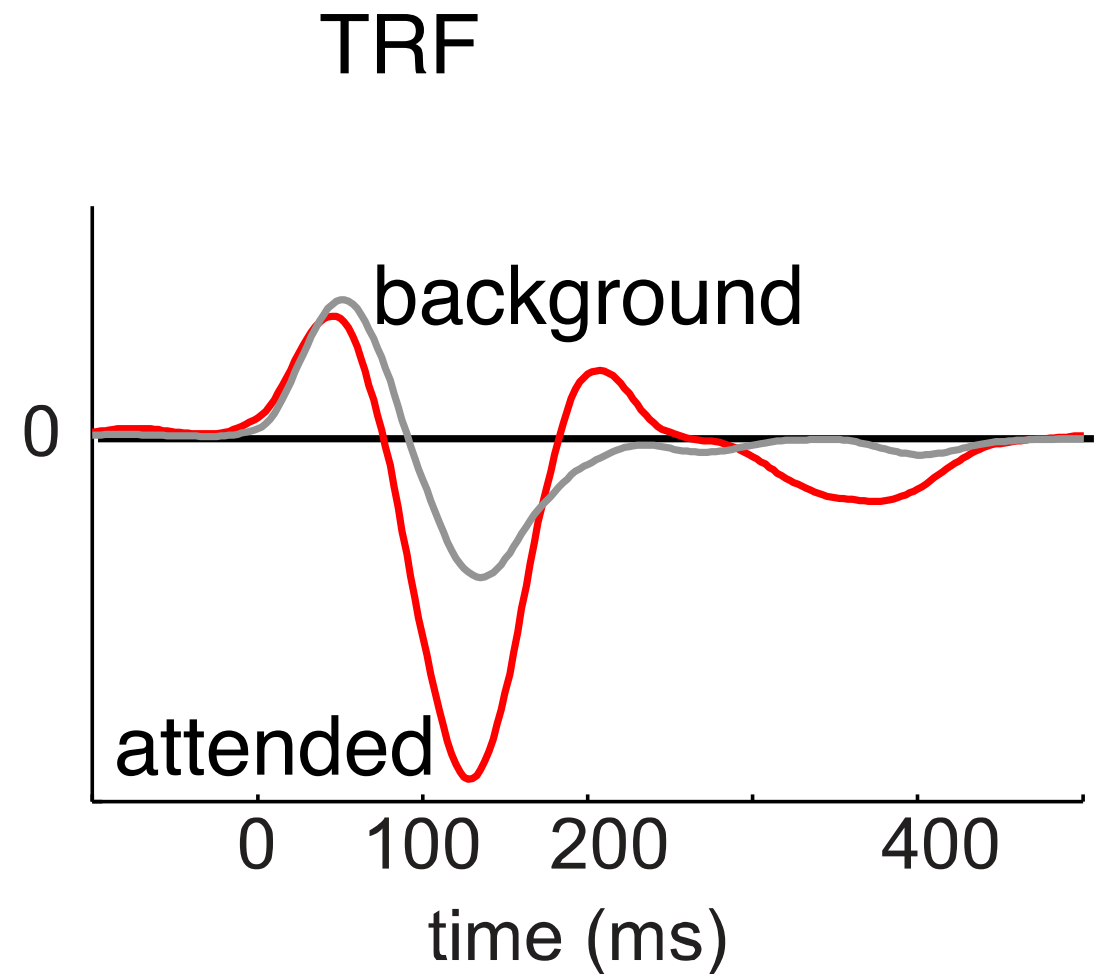


- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- M50_{STRF} positive peak
- M100_{STRF} negative peak

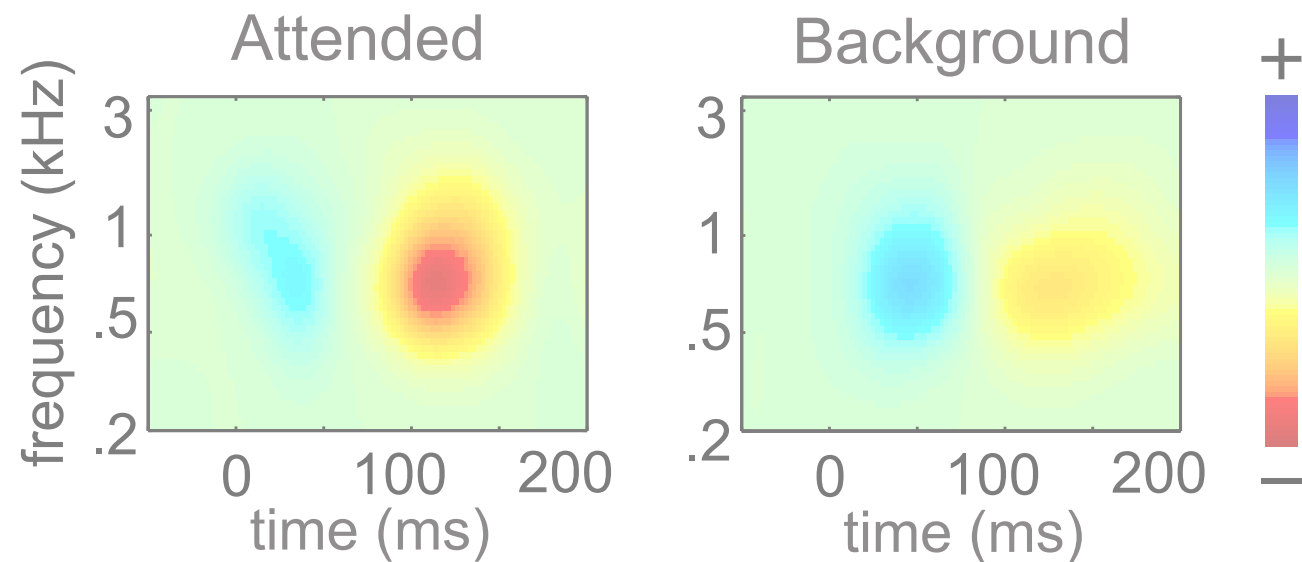
STRF Results



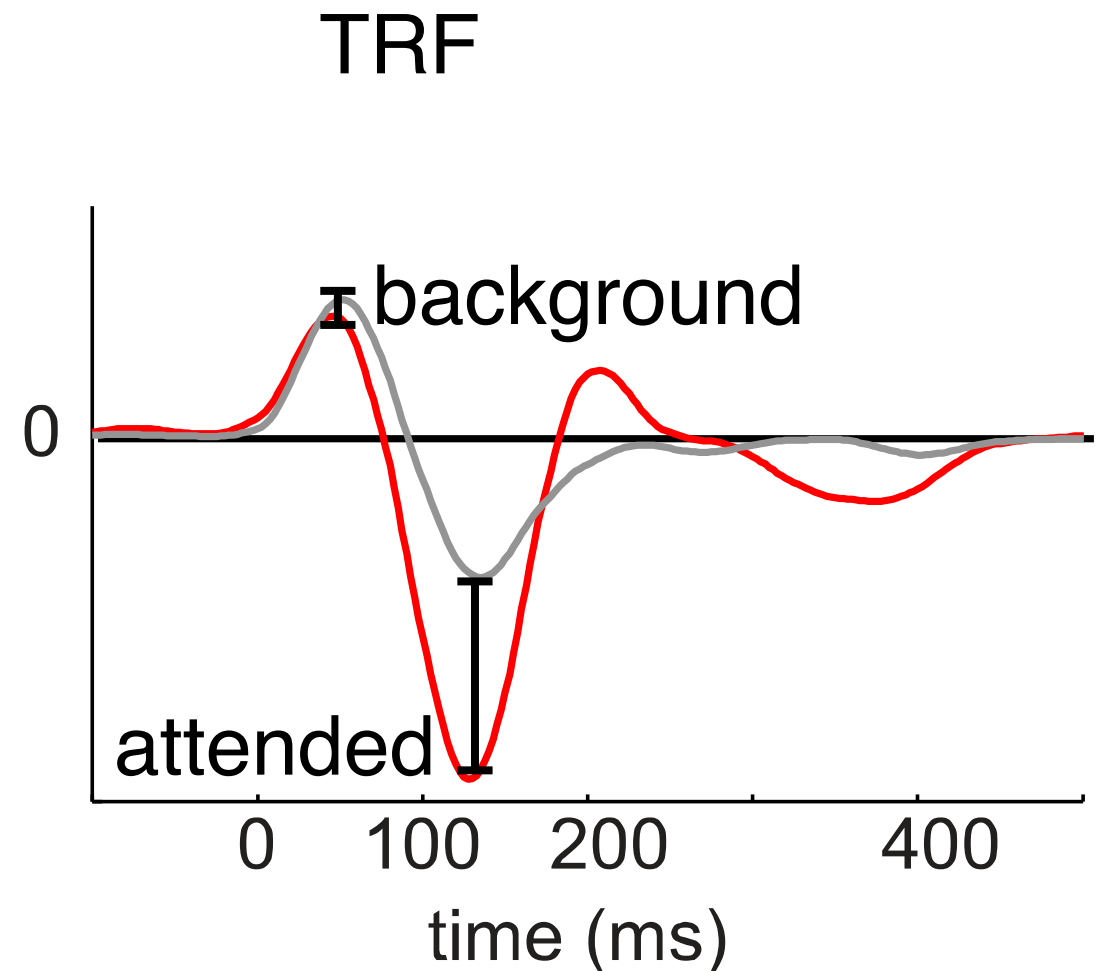
- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- M50_{STRF} positive peak
- M100_{STRF} negative peak



STRF Results

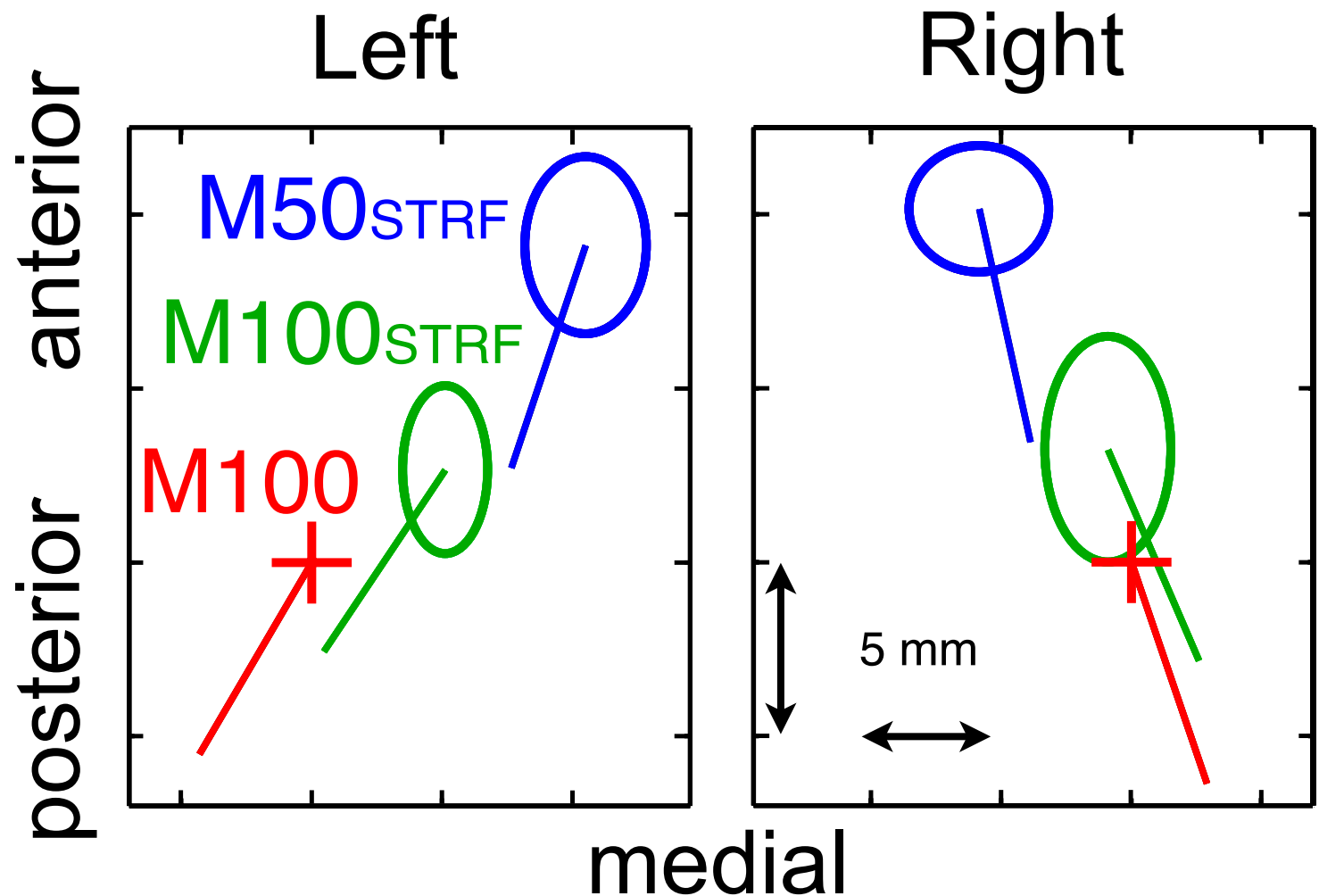


- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- $M50_{STRF}$ positive peak
- $M100_{STRF}$ negative peak
- **$M100_{STRF}$ strongly modulated by attention, *but not* $M50_{STRF}$**



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***



Studies In Progress

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

Summary

- Temporal Speech Envelope reconstructable from temporal neural responses (up to ~ 10 Hz)
- Cortical Processing Hierarchy: Representation consistent with being neural representation of auditory perceptual object
- Object representation at 100 ms latency (PT), but not by 50 ms (HG)
- Latency / Integration window matters
- Object of attention dynamically trackable
- Care needed with diverse subject pool

Thank You