Cortical Encoding of Auditory Objects at the Cocktail Party

Jonathan Z. Simon

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

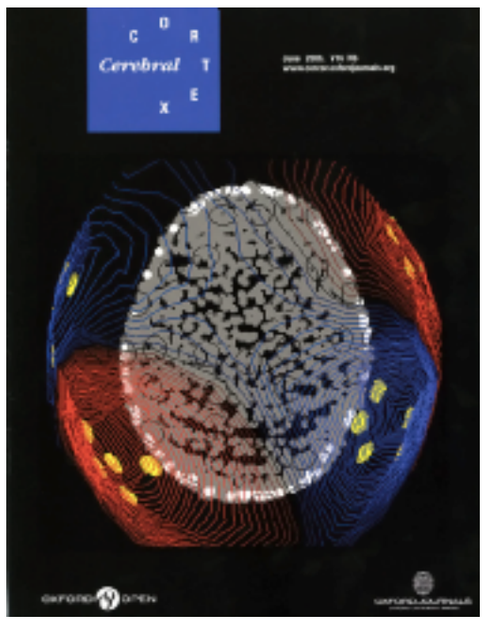
University of Maryland

ECE / ISR Joint Colloquium, 6 September 2013

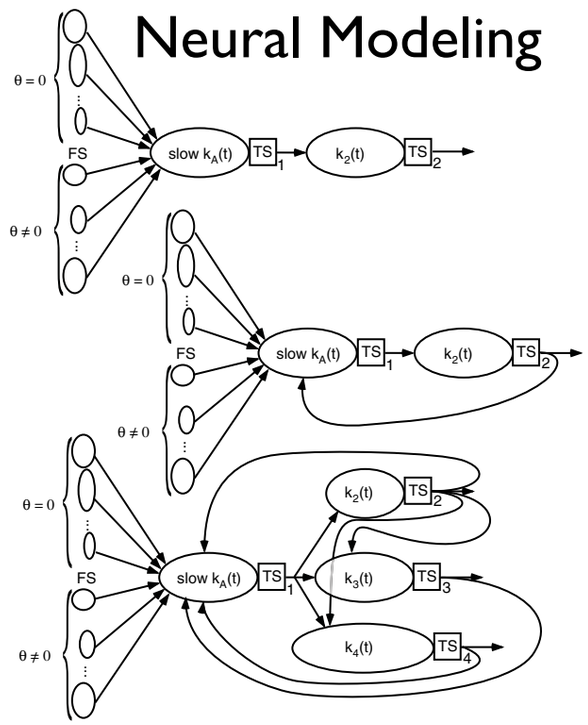
Jonathan Z. Simon

Neural processing of speech and complex auditory streams

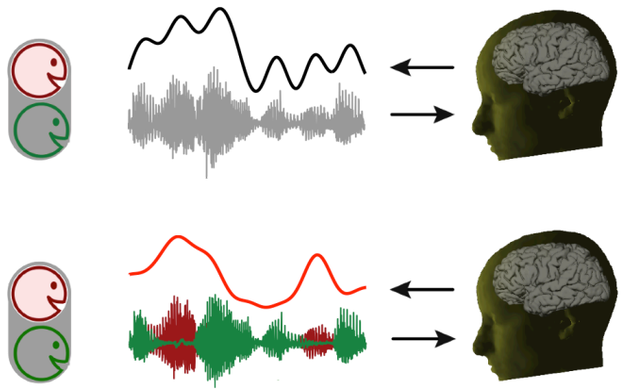
Magnetoencephalography



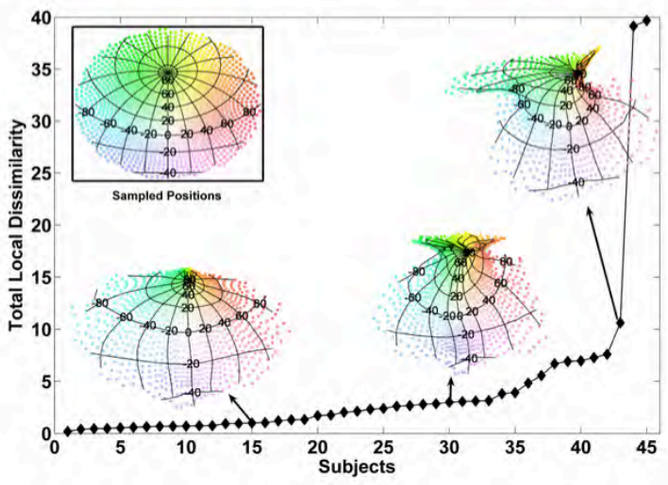
Neural Modeling



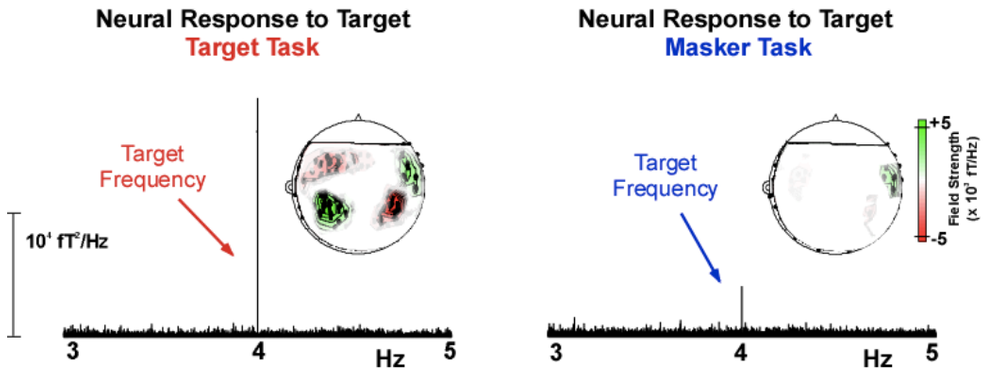
Neural Un-Mixing of Speech



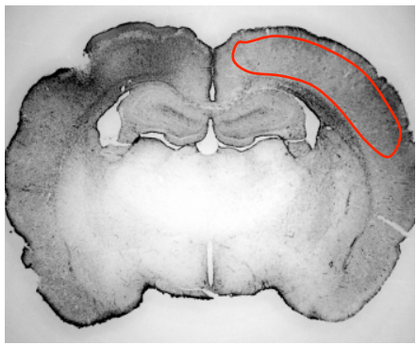
Neurally Inspired Algorithms



Neural Signal Processing



Advanced Neuroimaging



Acknowledgements

Grad Students

Francisco Cervantes
Alex Presacco
Krishna Puvvada

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

Collaborators

Catherine Carr
Alain de Cheveigné
Didier Depireux
Mounya Elhilali
Jonathan Fritz
Cindy Moss
David Poeppel
Shihab Shamma

Past Postdocs

Dan Hertz
Yadong Wang

Funding

NIH R01 DC 008342

Introduction

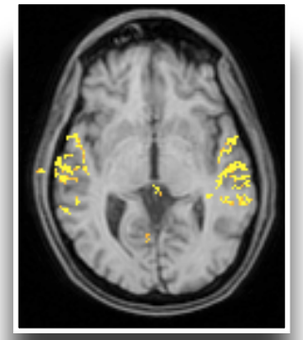
- Magnetoencephalography (MEG)
- Auditory Objects
- Neural Representations of Auditory Objects in Cortex: Decoding
- Neural Representations of Auditory Objects in Cortex: Encoding

Functional Brain Imaging

Functional Brain Imaging
= Non-invasive recording from human brain

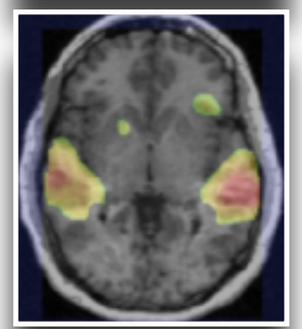
Hemodynamic techniques

fMRI
functional magnetic resonance imaging



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

PET
positron emission tomography

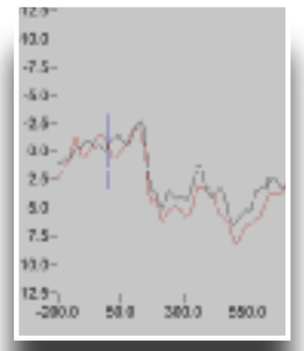


Poor Temporal Resolution **(~1 s)**

fMRI & MEG can capture effects in single subjects

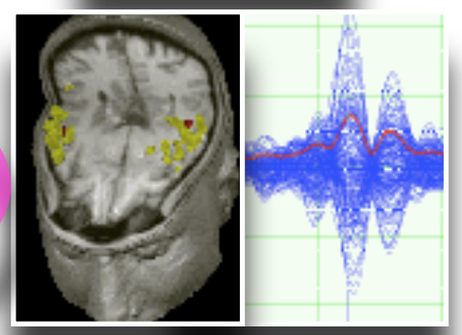
Electromagnetic techniques

EEG
electroencephalography



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

MEG
magnetoencephalography



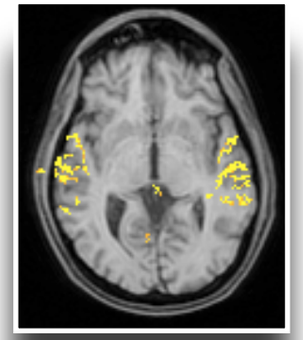
Excellent Temporal Resolution **(~1 ms)**

Functional Brain Imaging

Functional Brain Imaging
= Non-invasive recording from human brain

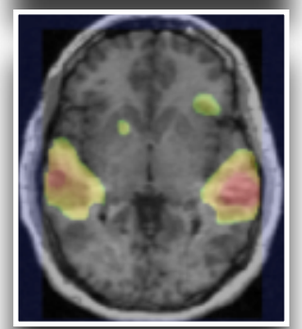
Hemodynamic techniques

fMRI
functional magnetic resonance imaging



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

PET
positron emission tomography

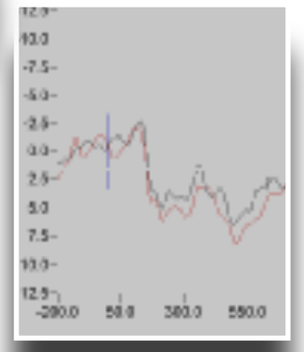


Poor Temporal Resolution **(~1 s)**

fMRI & MEG can capture effects in single subjects

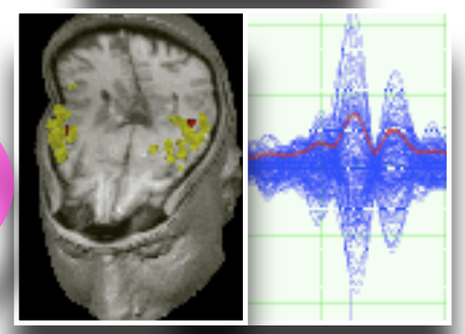
Electromagnetic techniques

EEG
electroencephalography



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

MEG
magnetoencephalography



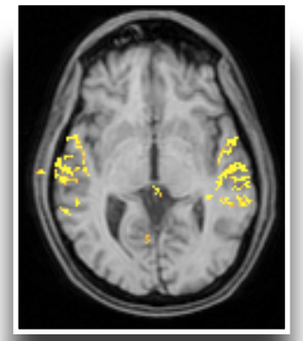
Excellent Temporal Resolution **(~1 ms)**

Functional Brain Imaging

Functional Brain Imaging
= Non-invasive recording from human brain

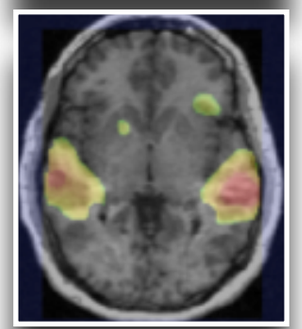
Hemodynamic techniques

fMRI
functional magnetic resonance imaging



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

PET
positron emission tomography

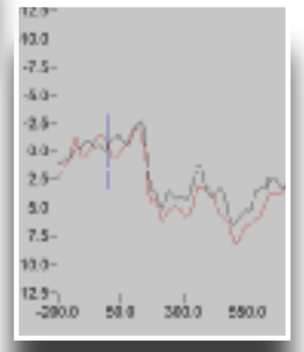


Poor Temporal Resolution **(~1 s)**

fMRI & MEG can capture effects in single subjects

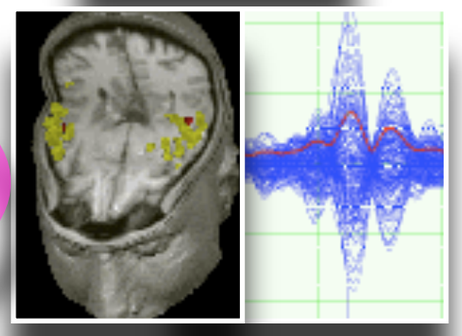
Electromagnetic techniques

EEG
electroencephalography



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

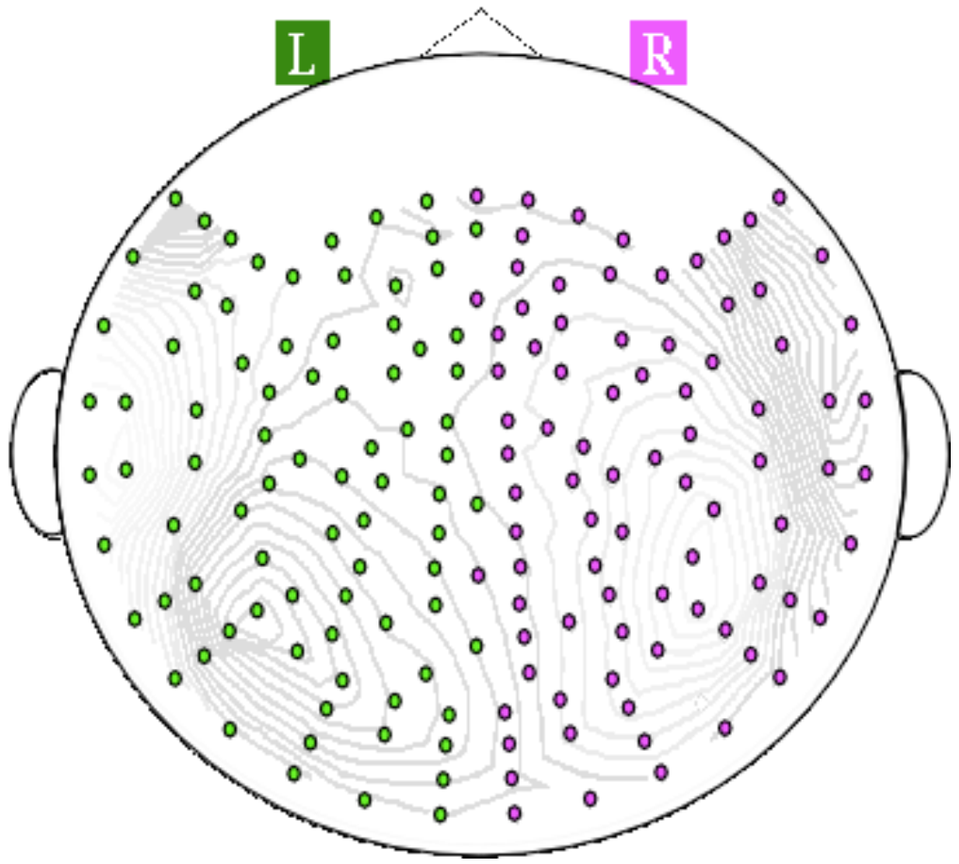
MEG
magnetoencephalography



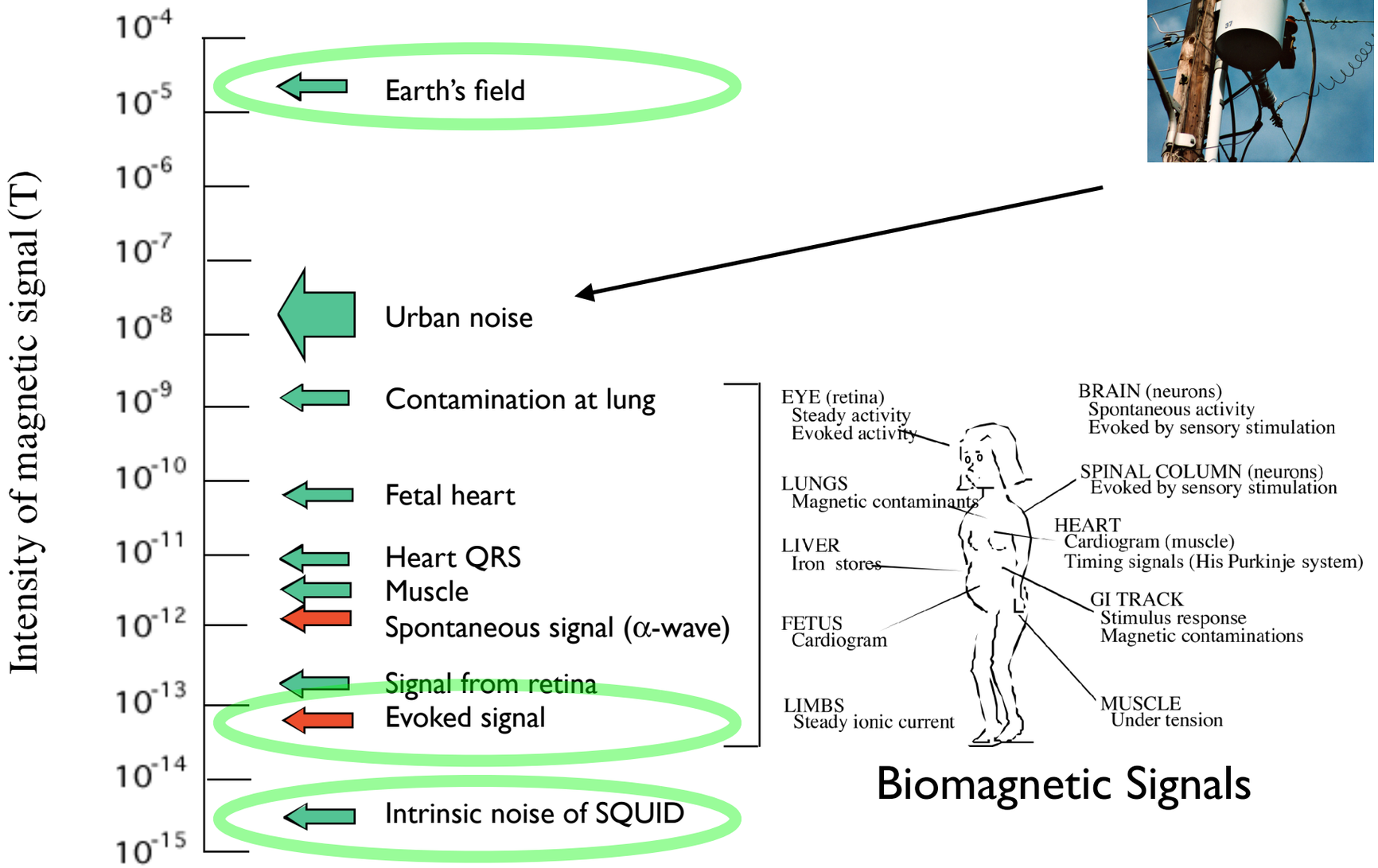
Excellent Temporal Resolution **(~1 ms)**

Magnetoencephalography

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



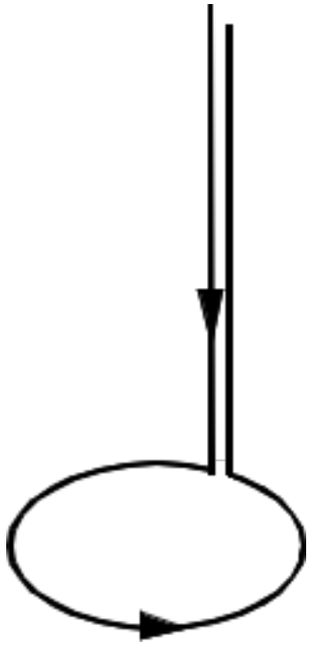
Magnetic Field Strengths



Biomagnetic Signals

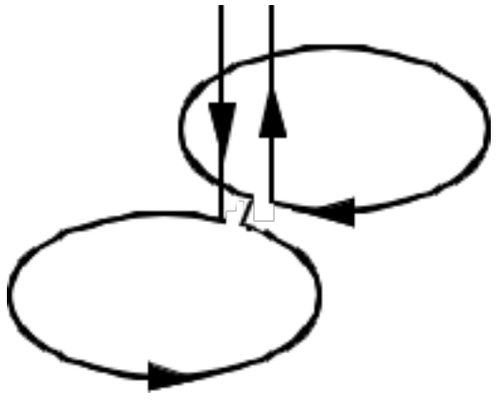
MEG SQUIDs

SQUID Magnetometer

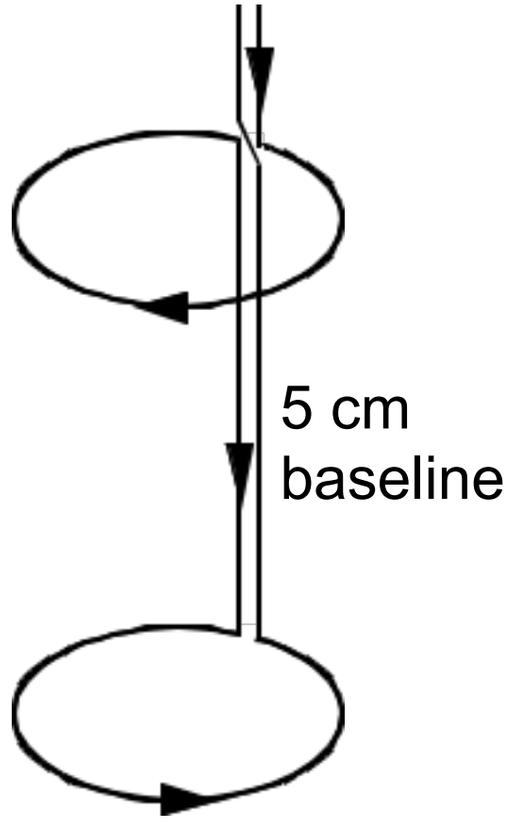


SQUID Gradiometers

Noise reduction from Differential measurement

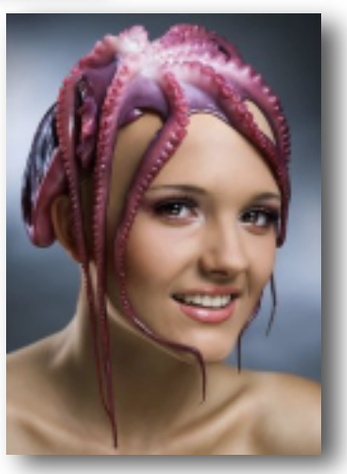


Planar Gradiometer



Axial Gradiometer

MEG = “Squid head”



Neural Signals & MEG

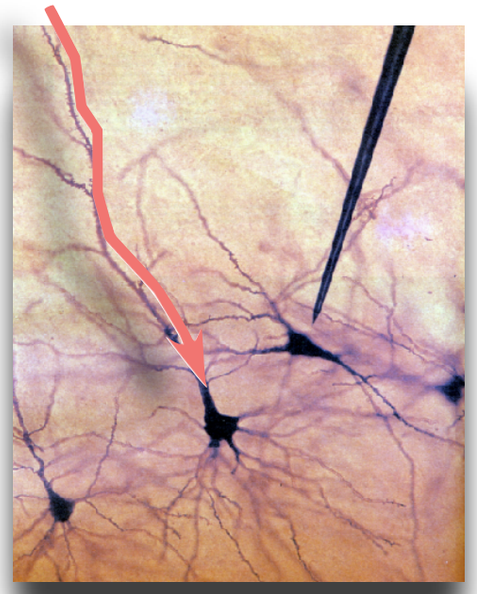
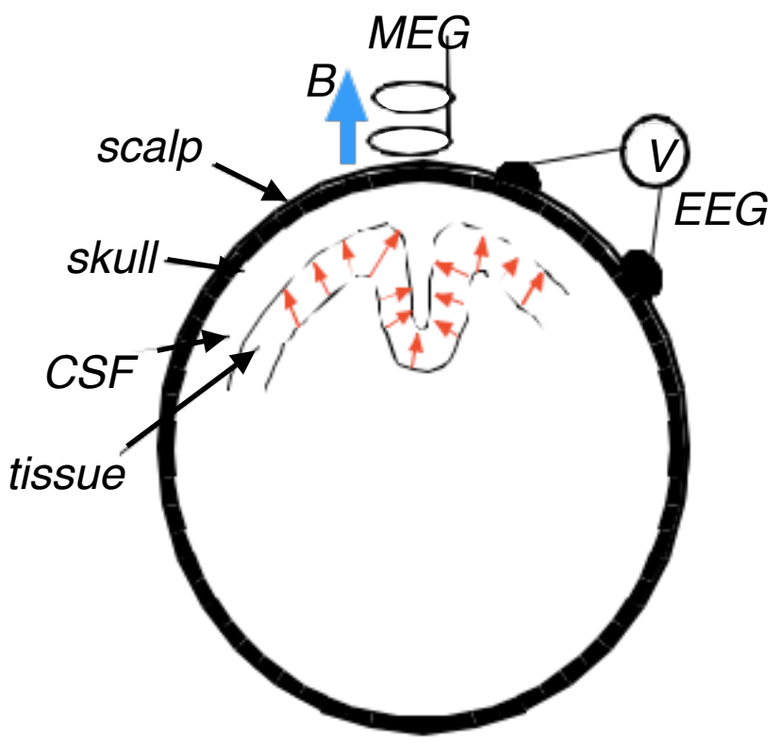
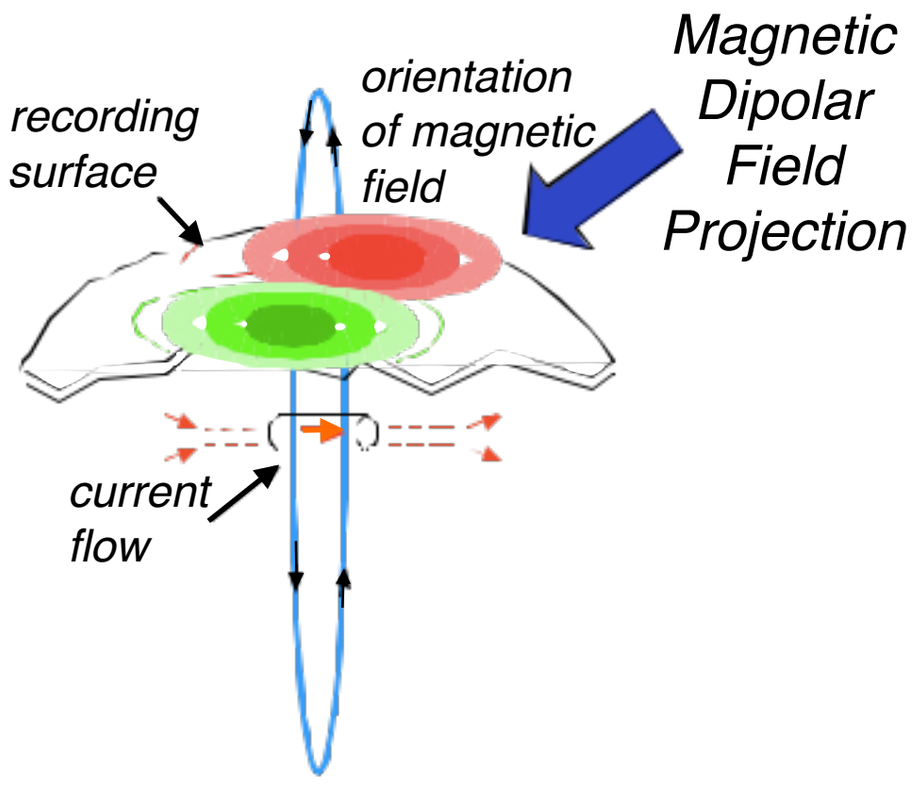


Photo by Fritz Goro



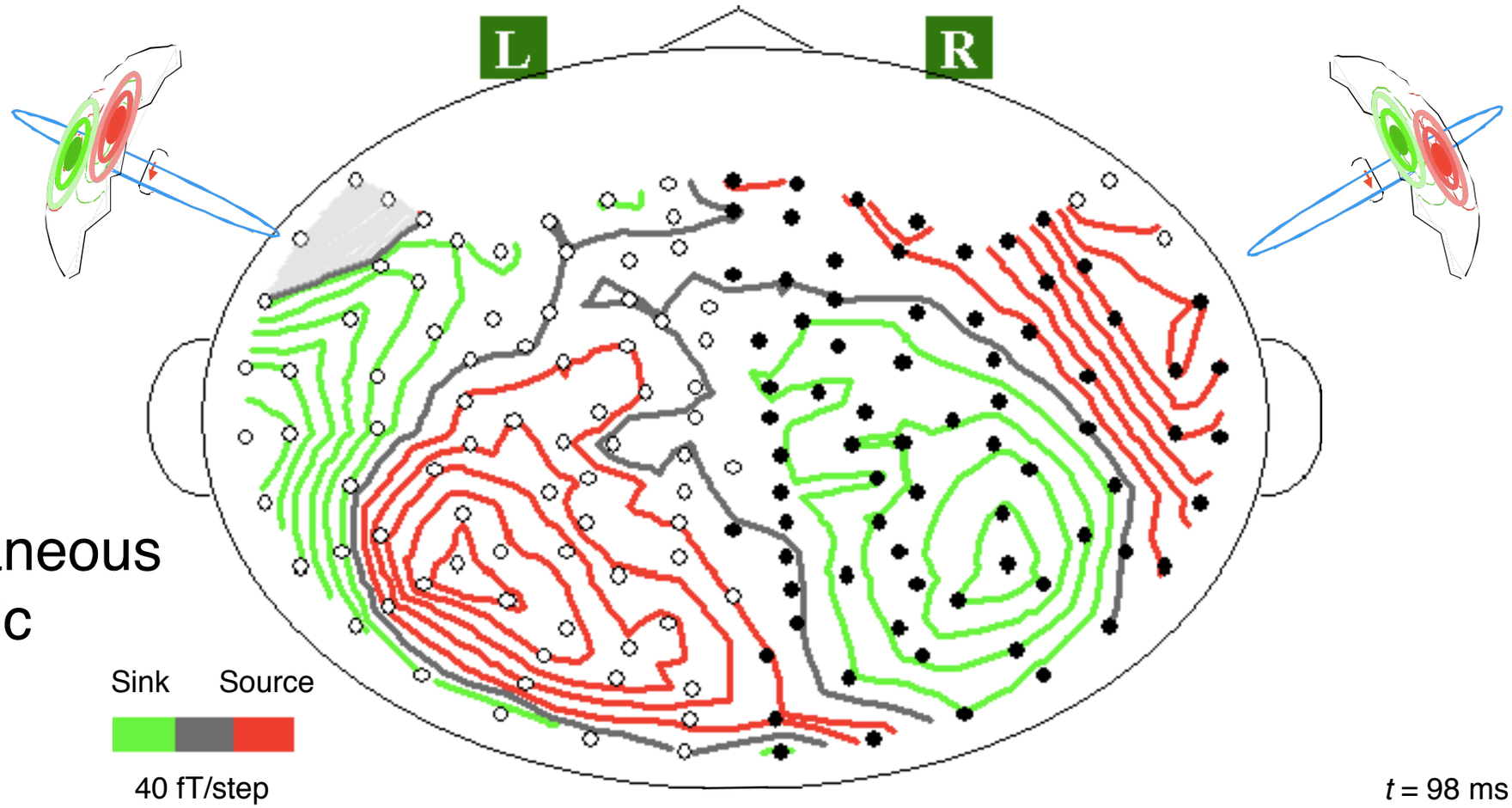


- Direct electrophysiological measurement
  - not hemodynamic
  - real-time
- No unique solution for distributed source

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

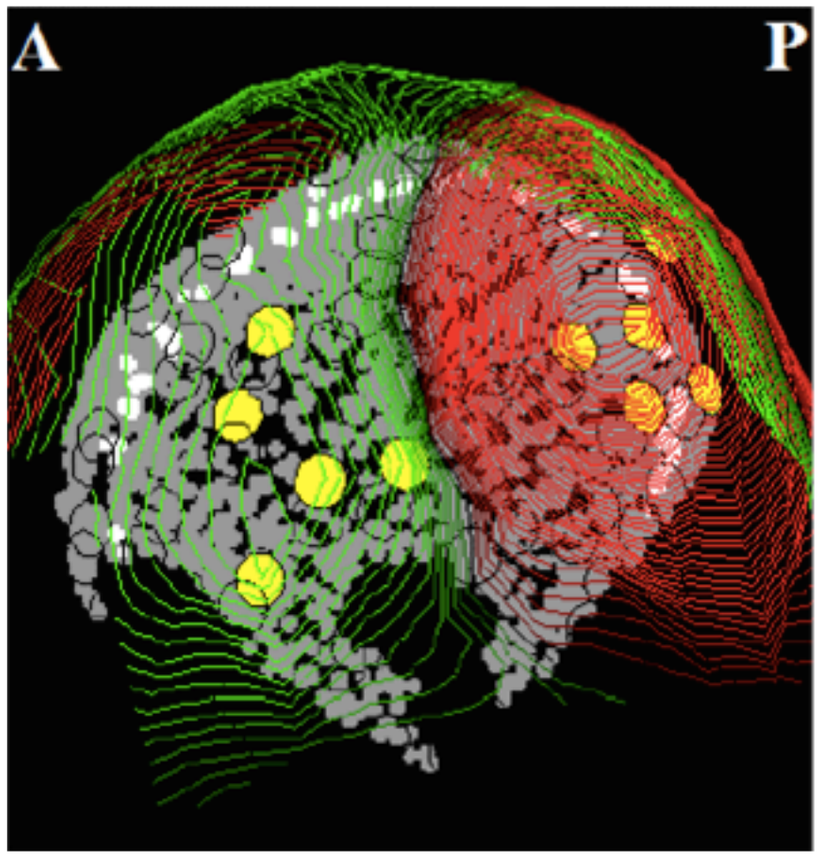
MEG Auditory Field

Flattened Isofield Contour Map

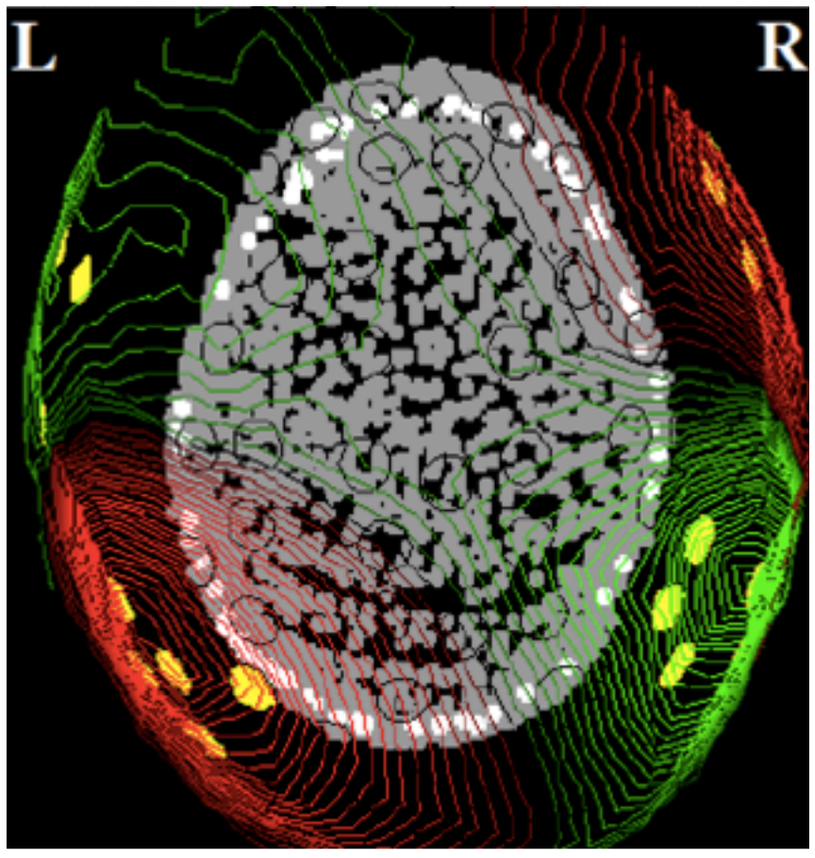


MEG Auditory Field

3-D Isofield Contour Map



Sagittal View



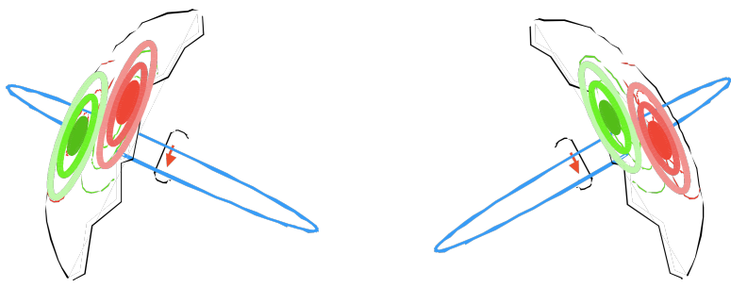
Axial View

Chait, Poeppel and Simon, Cerebral Cortex (2006)

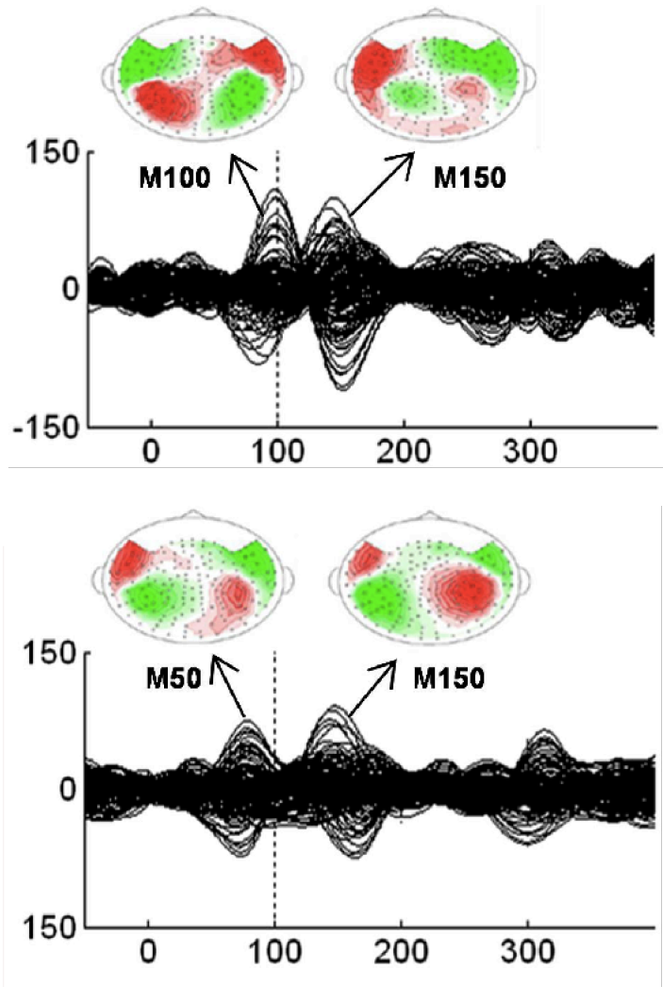
Time Course of MEG Responses

Auditory Evoked Responses

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

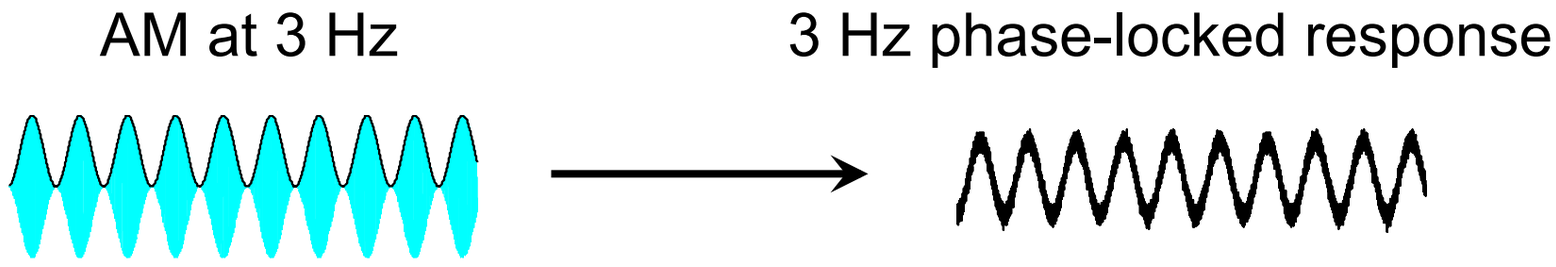


Pure Tone



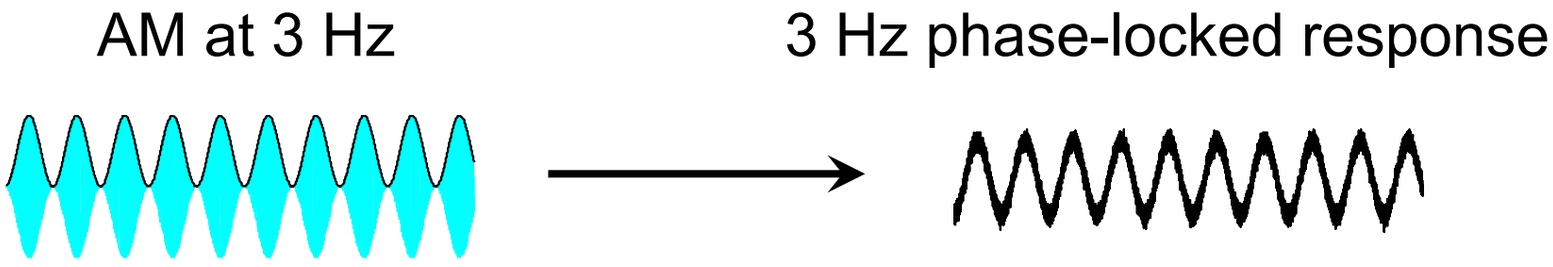
Broadband Noise

Phase-Locking in MEG to Slow Acoustic Modulations

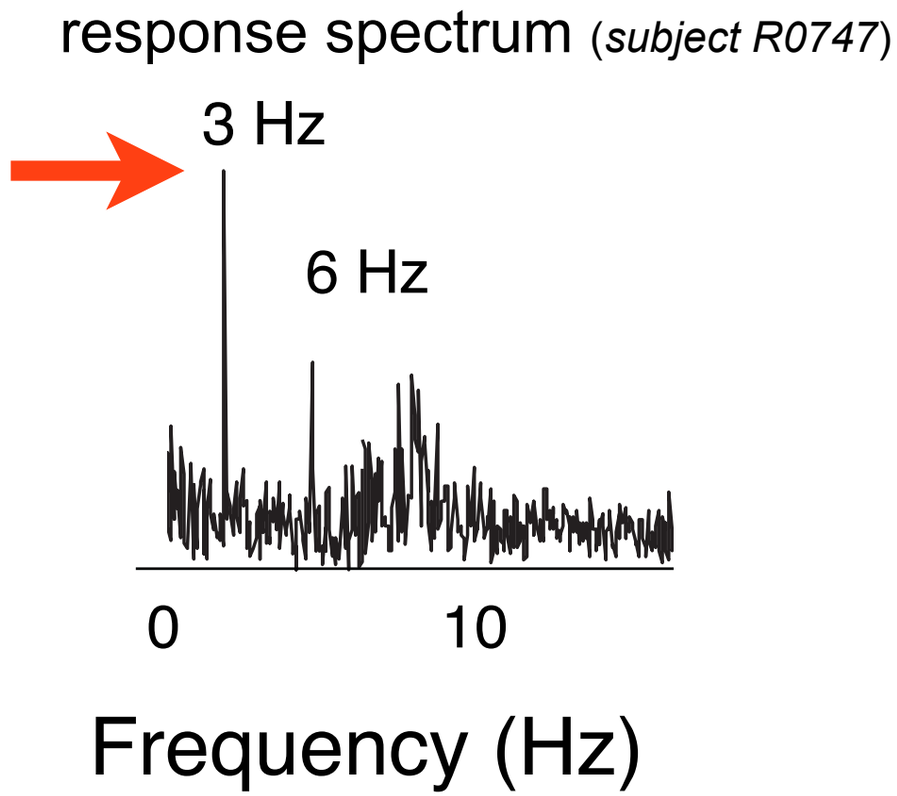


Ding & Simon, J Neurophysiol (2009)
Wang et al., J Neurophysiol (2012)

Phase-Locking in MEG to Slow Acoustic Modulations

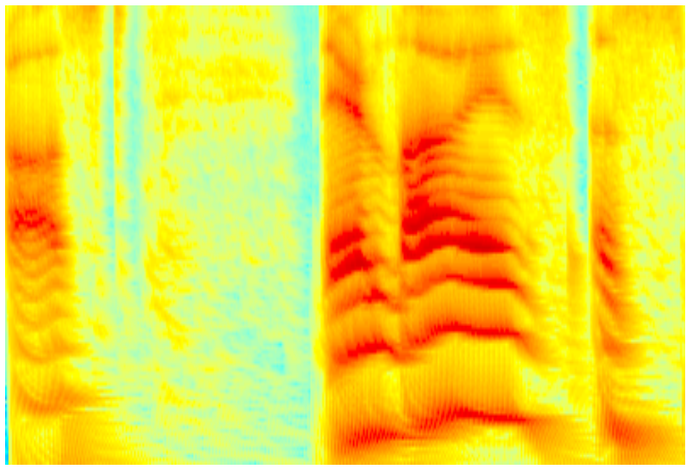


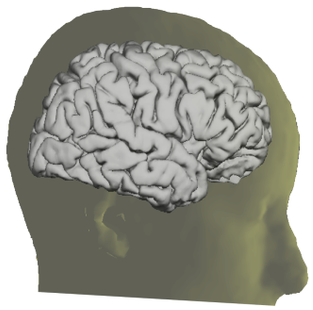
MEG activity is precisely phase-locked to temporal modulations of sound



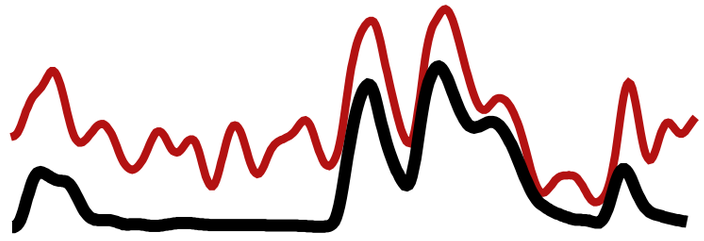
Ding & Simon, J Neurophysiol (2009)
Wang et al., J Neurophysiol (2012)

MEG Responses to Speech Modulations

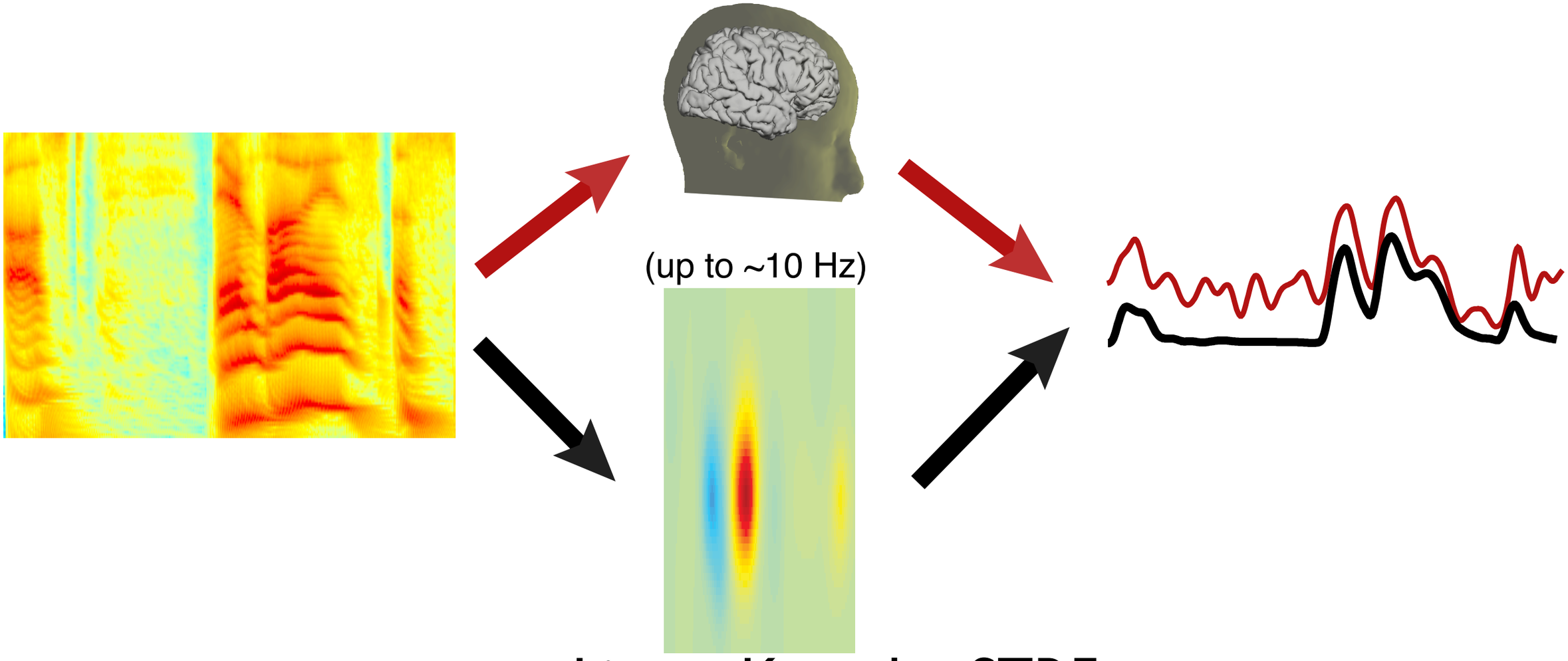




Auditory Model



MEG Responses Predicted by STRF Model

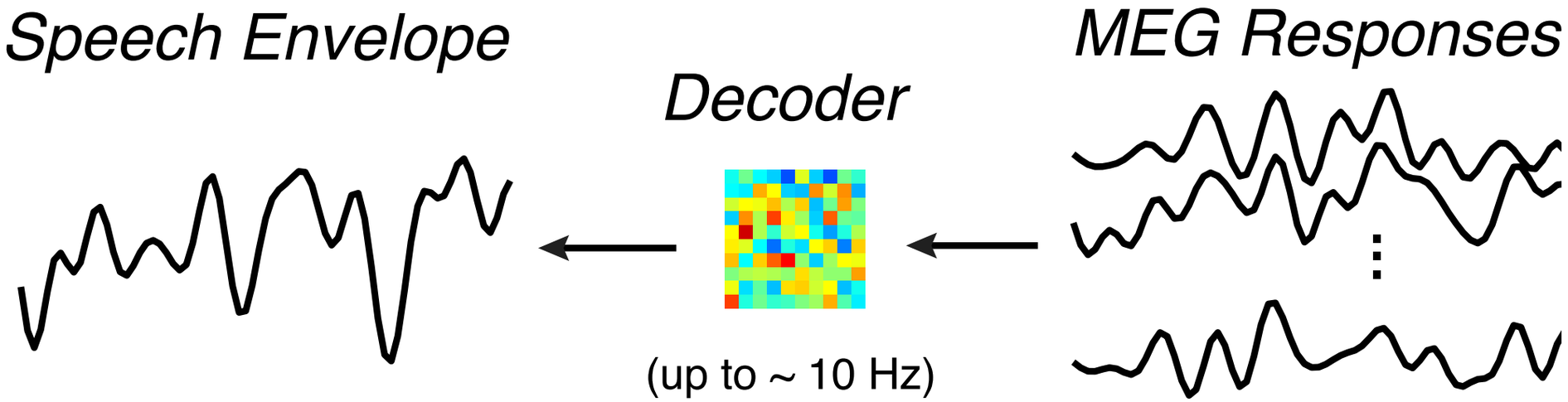


Linear Kernel = STRF

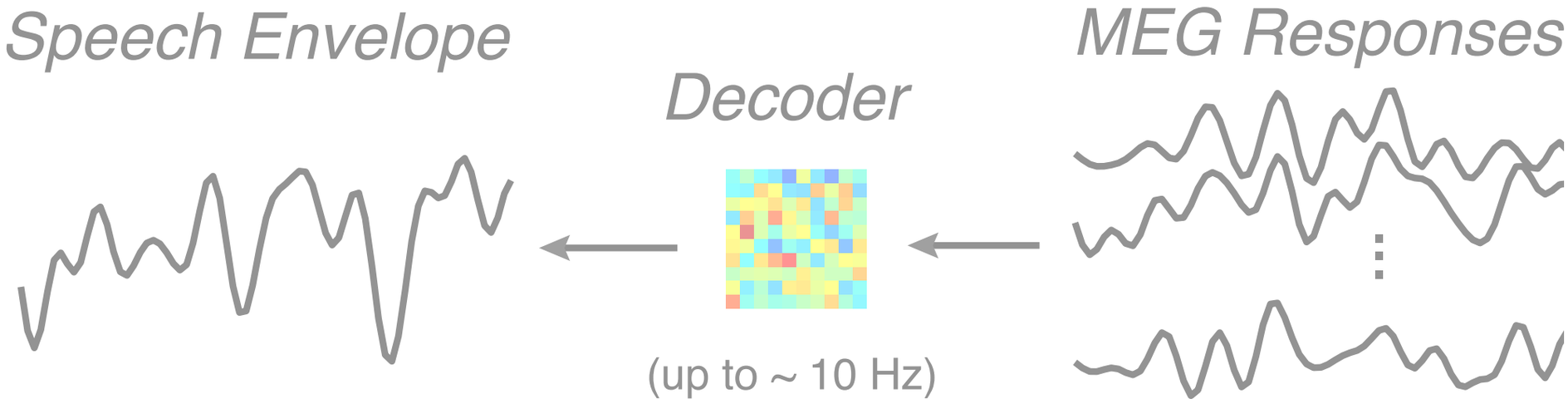
"Spectro-Temporal Response Function"

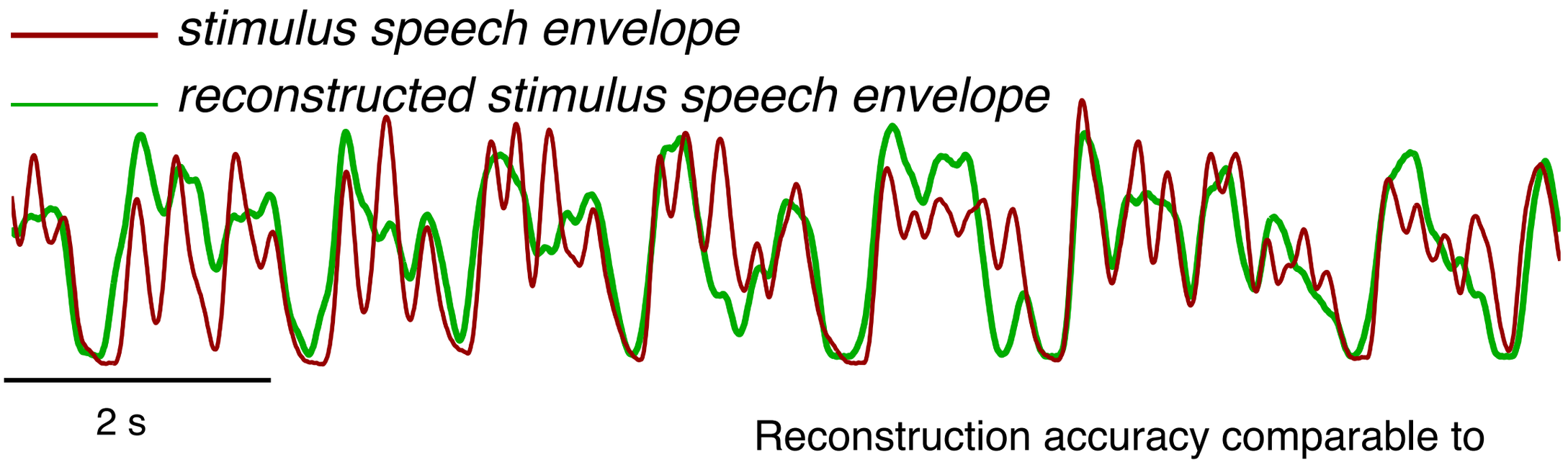
Ding & Simon, J Neurophysiol (2012)

Neural Reconstruction of Speech Envelope



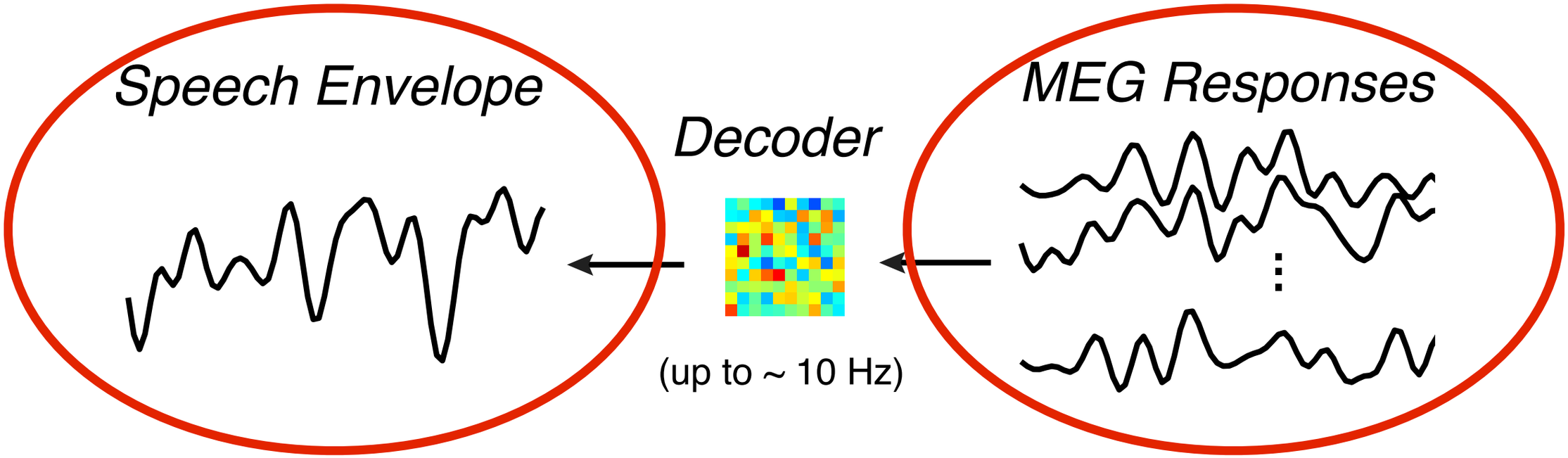
Neural Reconstruction of Speech Envelope





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)

Auditory Objects

- What is an auditory object?
  - perceptual construct (not neural, not acoustic)
  - commonalities with visual objects
  - several potential formal definitions

Auditory Object Definition

- Griffiths & Warren definition:
  - corresponds with *something* in the sensory world
  - object information *separate from* information of rest of sensory world
  - abstracted: object information *generalized over particular* sensory experiences

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

Auditory Objects at the Cocktail Party

Alex Katz,
The Cocktail Party

Auditory Objects at the Cocktail Party

Alex Katz,
The Cocktail Party

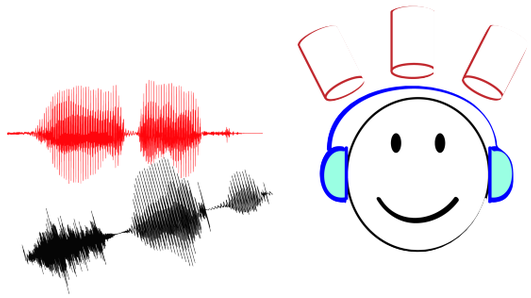
Auditory Objects at the Cocktail Party

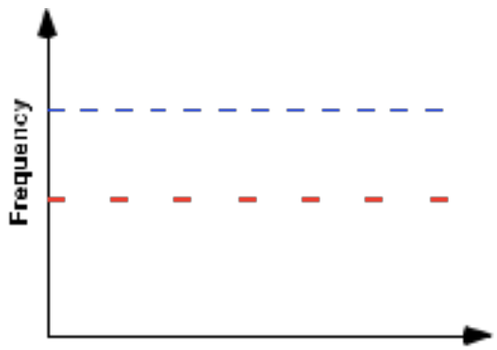
Alex Katz,
The Cocktail Party

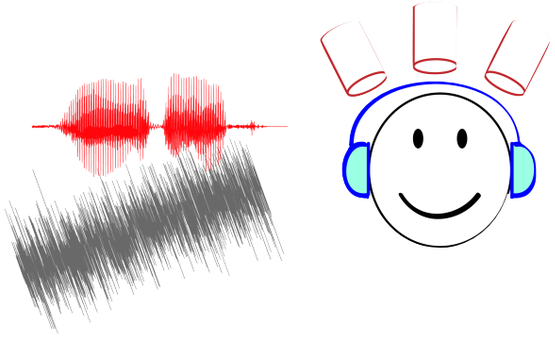
Auditory Objects at the Cocktail Party

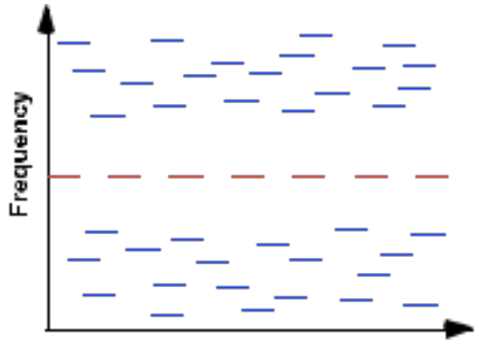
Alex Katz,
The Cocktail Party

Experiments

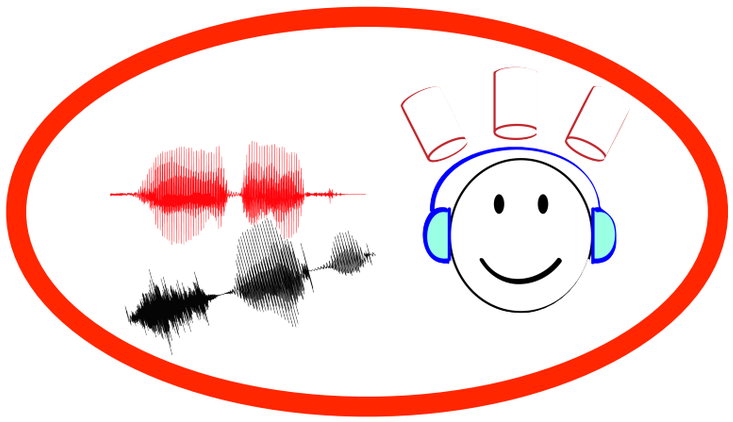


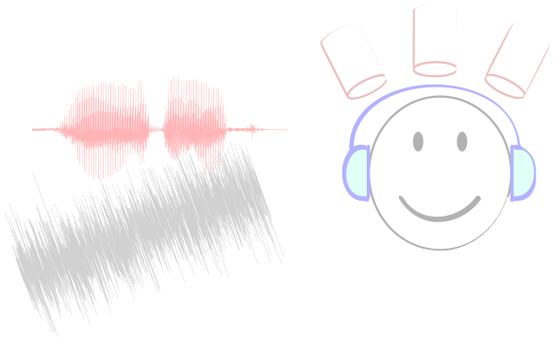


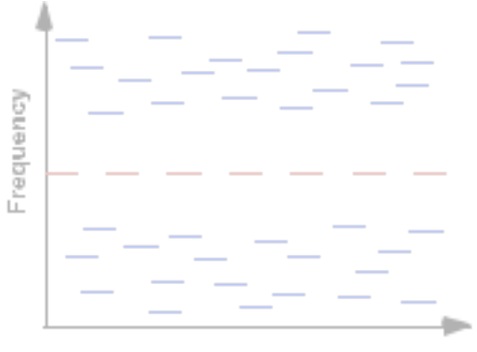




Experiments







Ding & Simon, PNAS (2012)

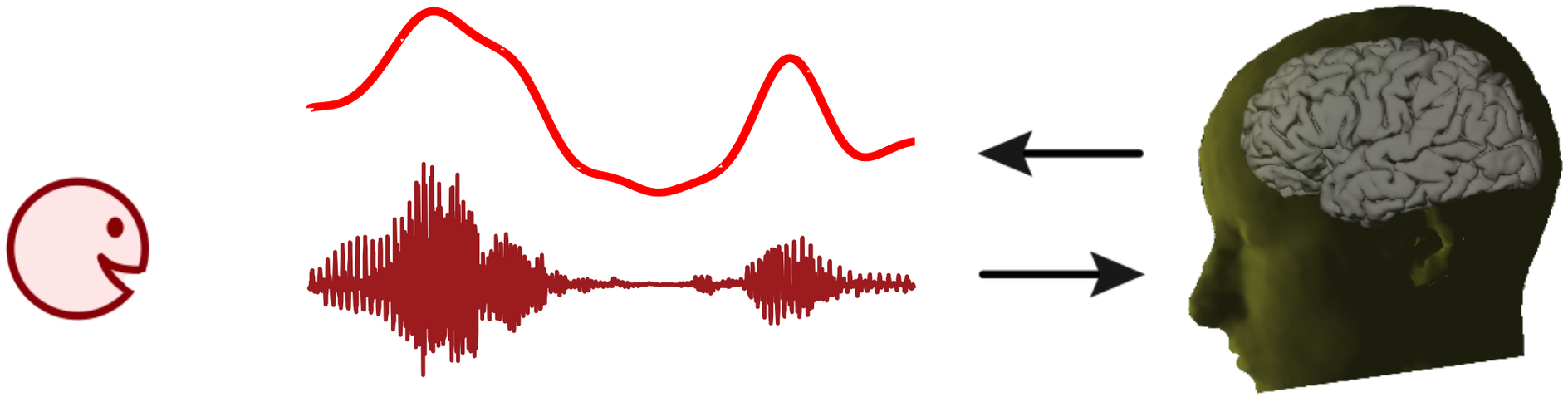
Speech Stream as an Auditory Object

- corresponds with something in the sensory world
- information *separate from* information of rest of sensory world
  e.g. other speech streams or noise
- abstracted: object information *generalized over particular* sensory experiences
  e.g. different sound mixtures

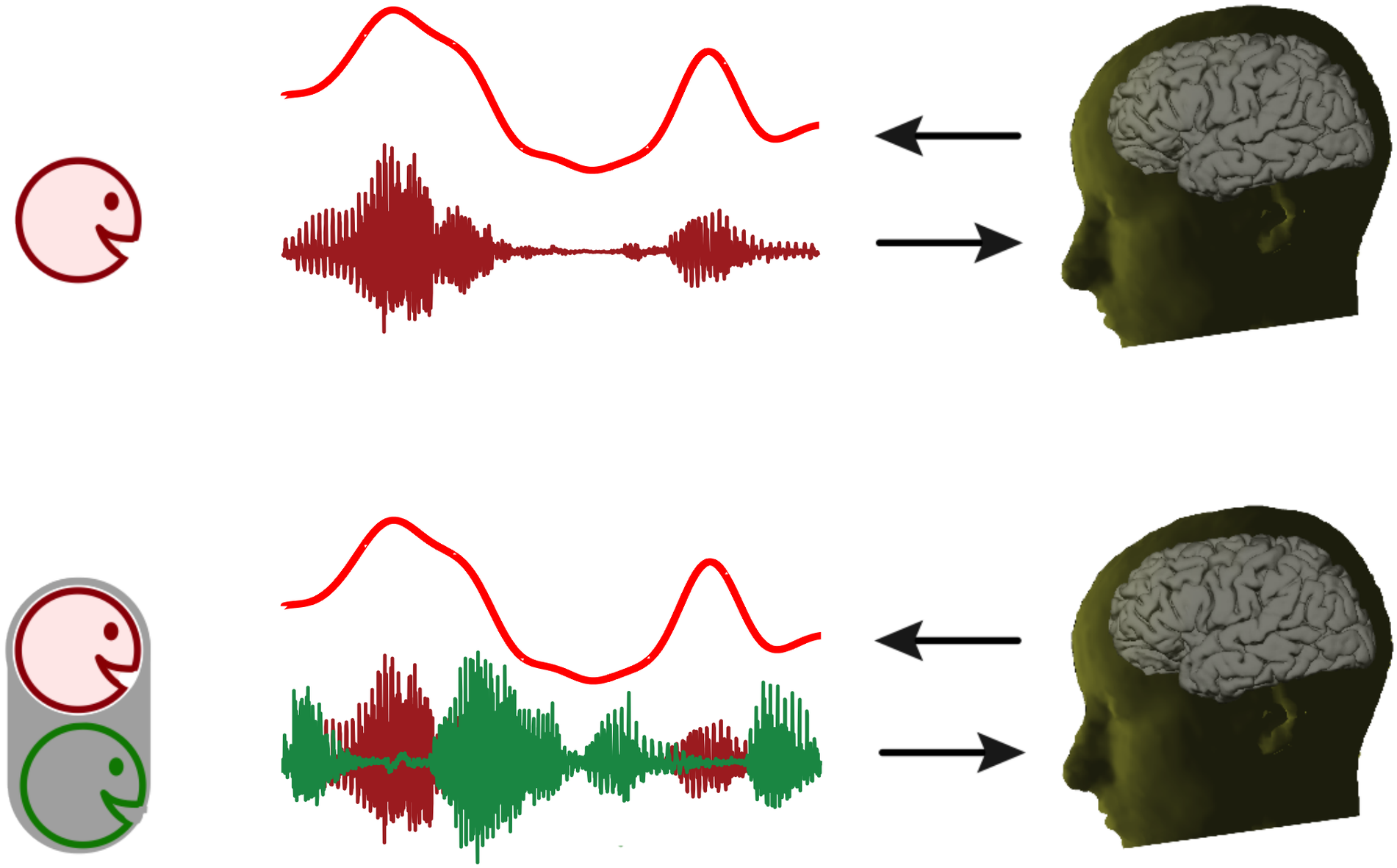
Neural Representation of an Auditory Object

- neural representation is of something in sensory world
- when other sounds mixed in,
  neural representation is of auditory object,
  not entire acoustic scene
- neural representation invariant
  under broad changes in specific acoustics

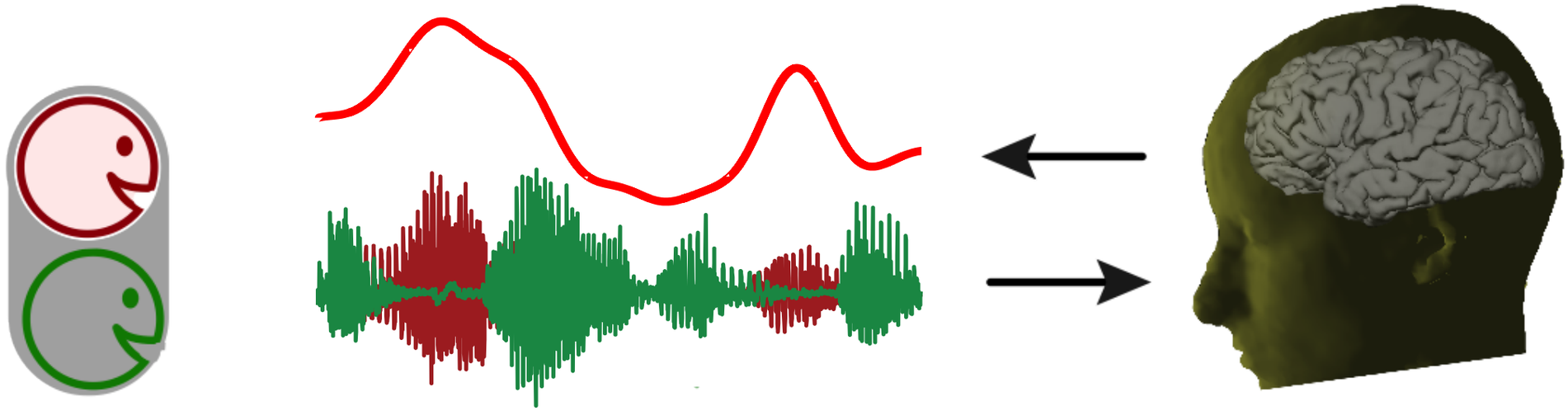
Selective Neural Encoding



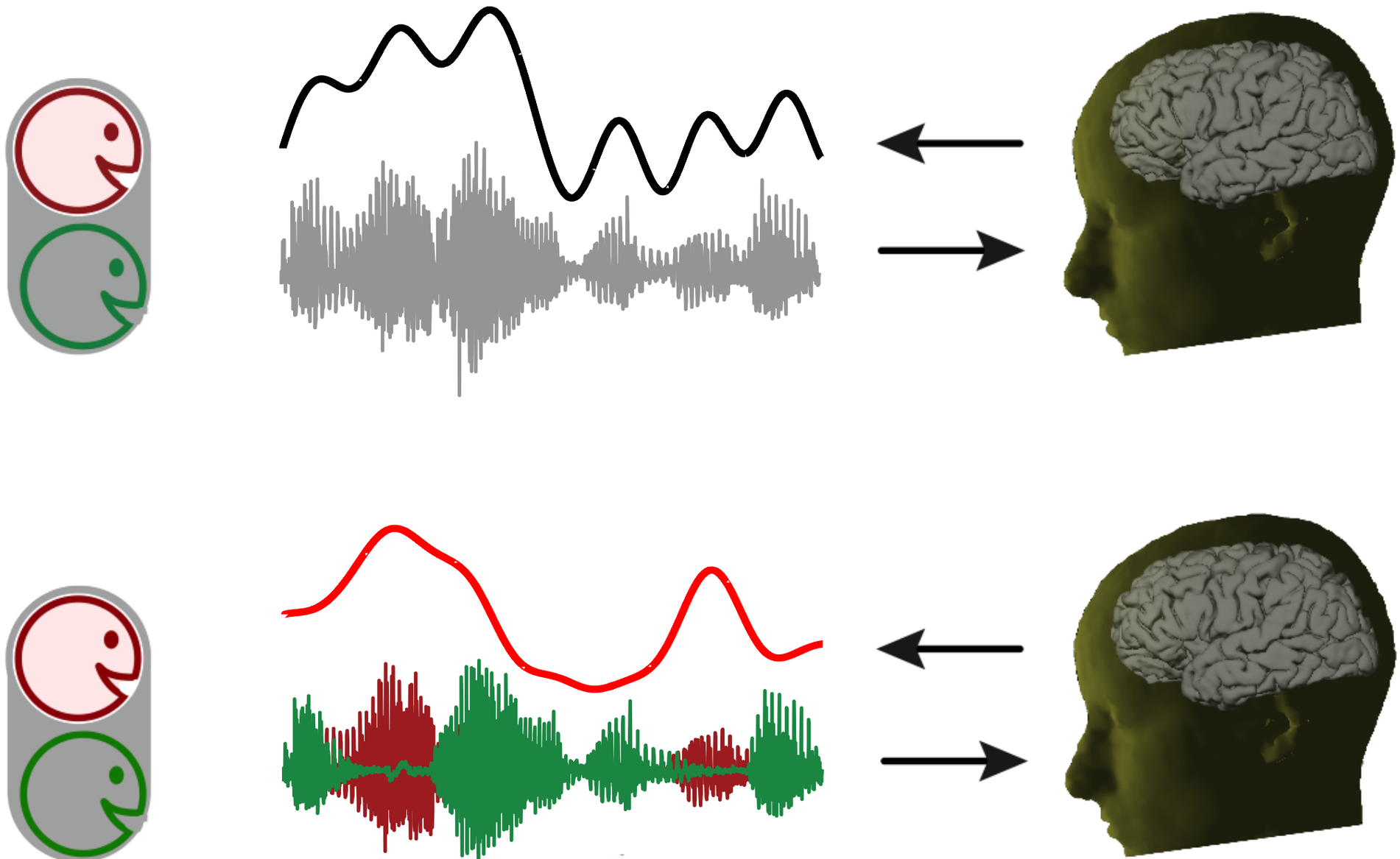
Selective Neural Encoding



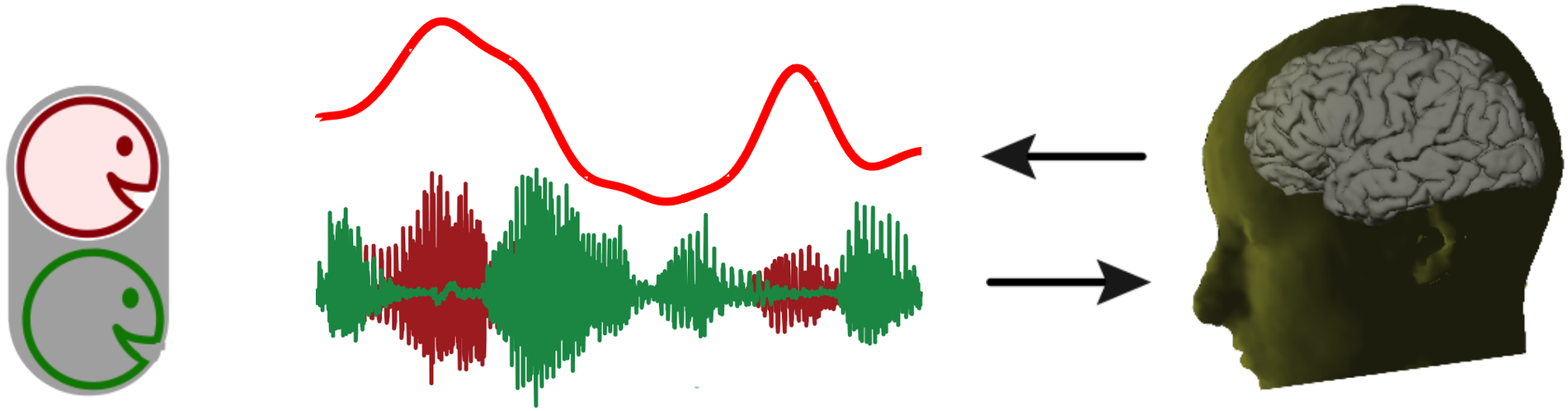
Selective Neural Encoding



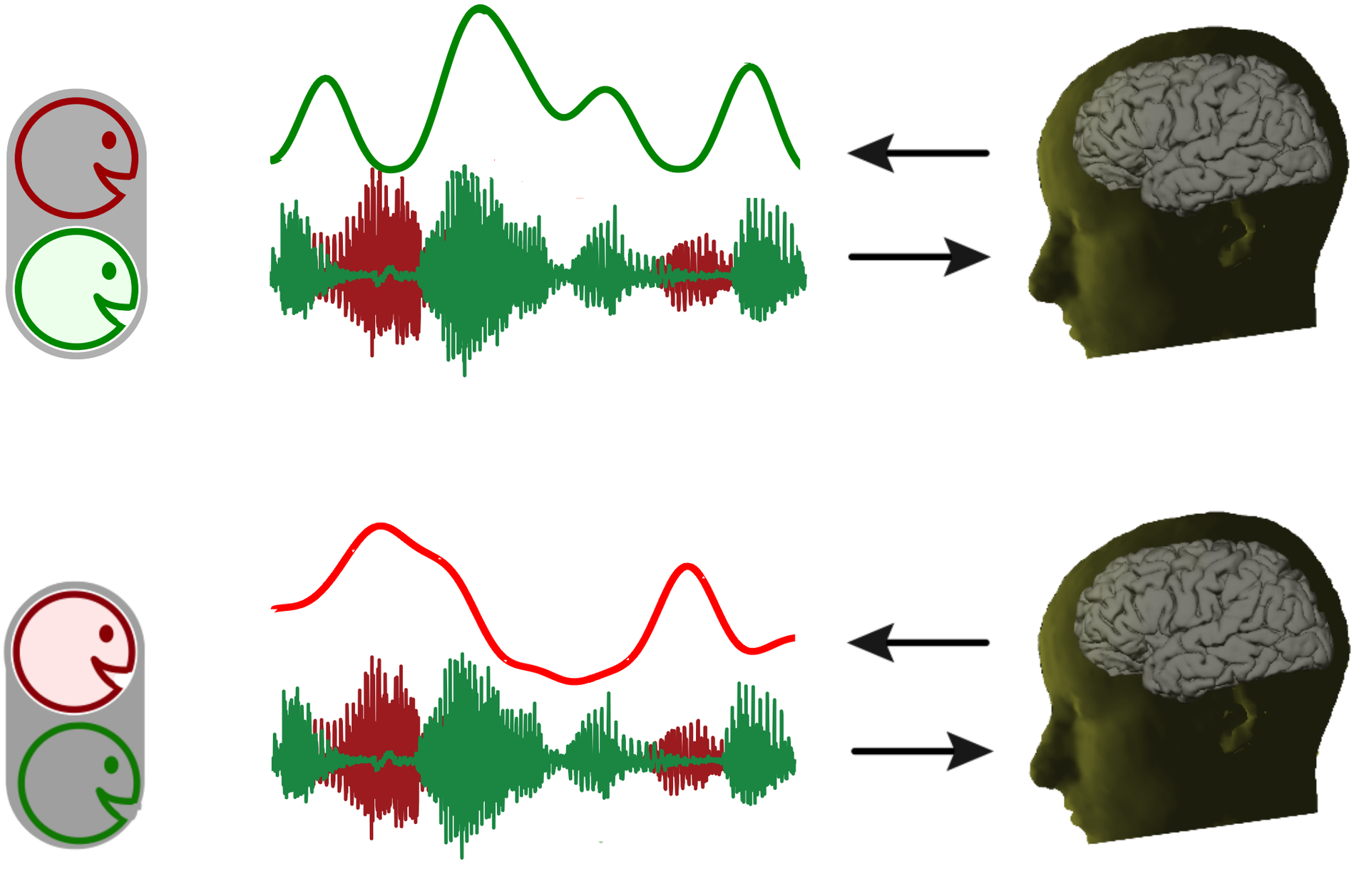
Unselective vs. Selective Neural Encoding



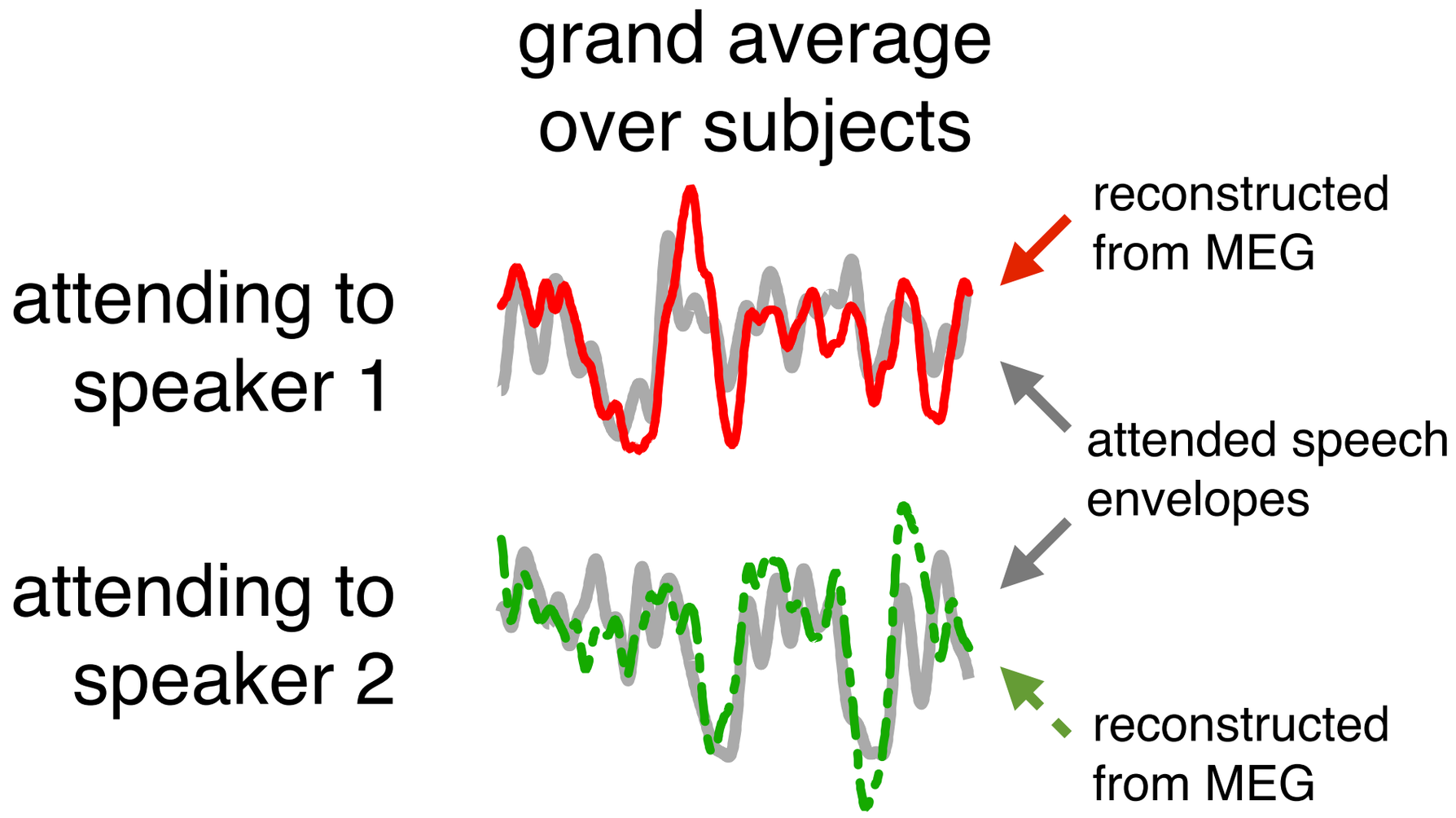
Unselective vs. Selective Neural Encoding



Selective Neural Encoding

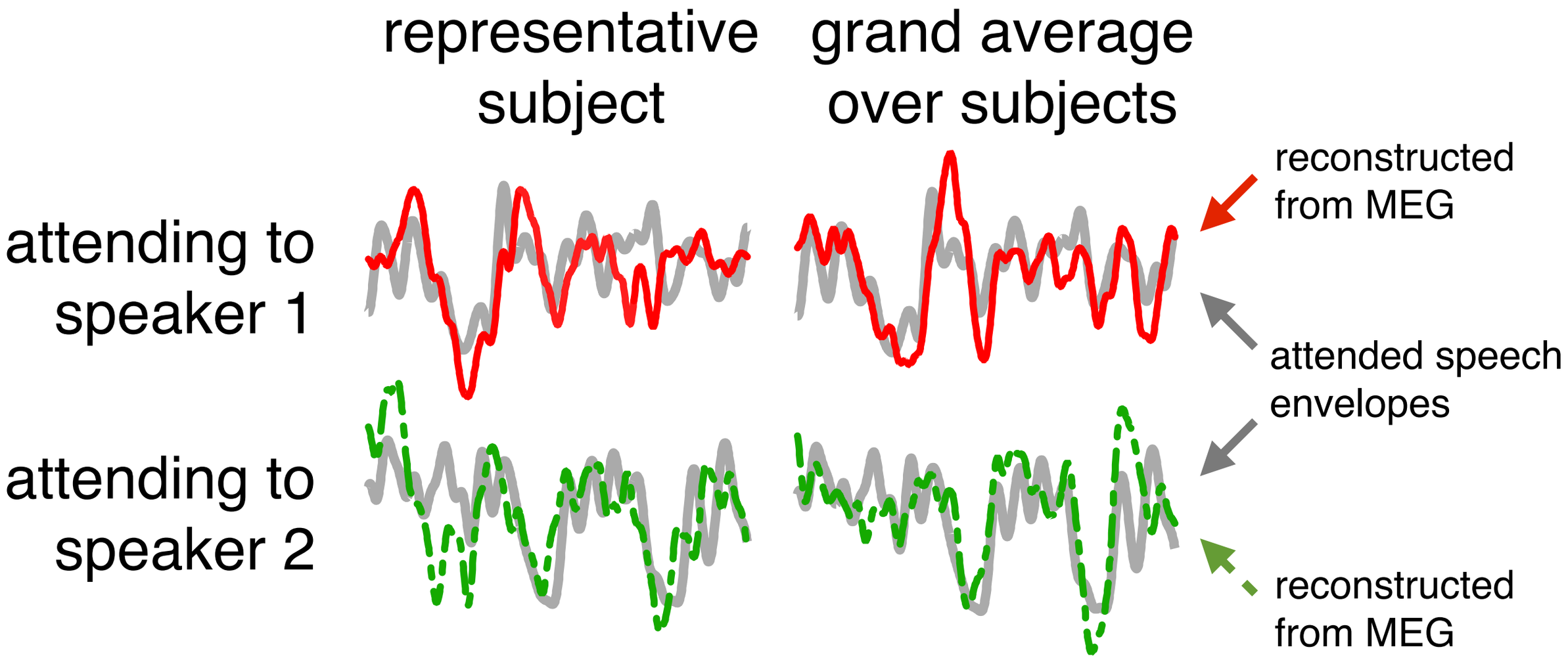


Stream-Specific Representation



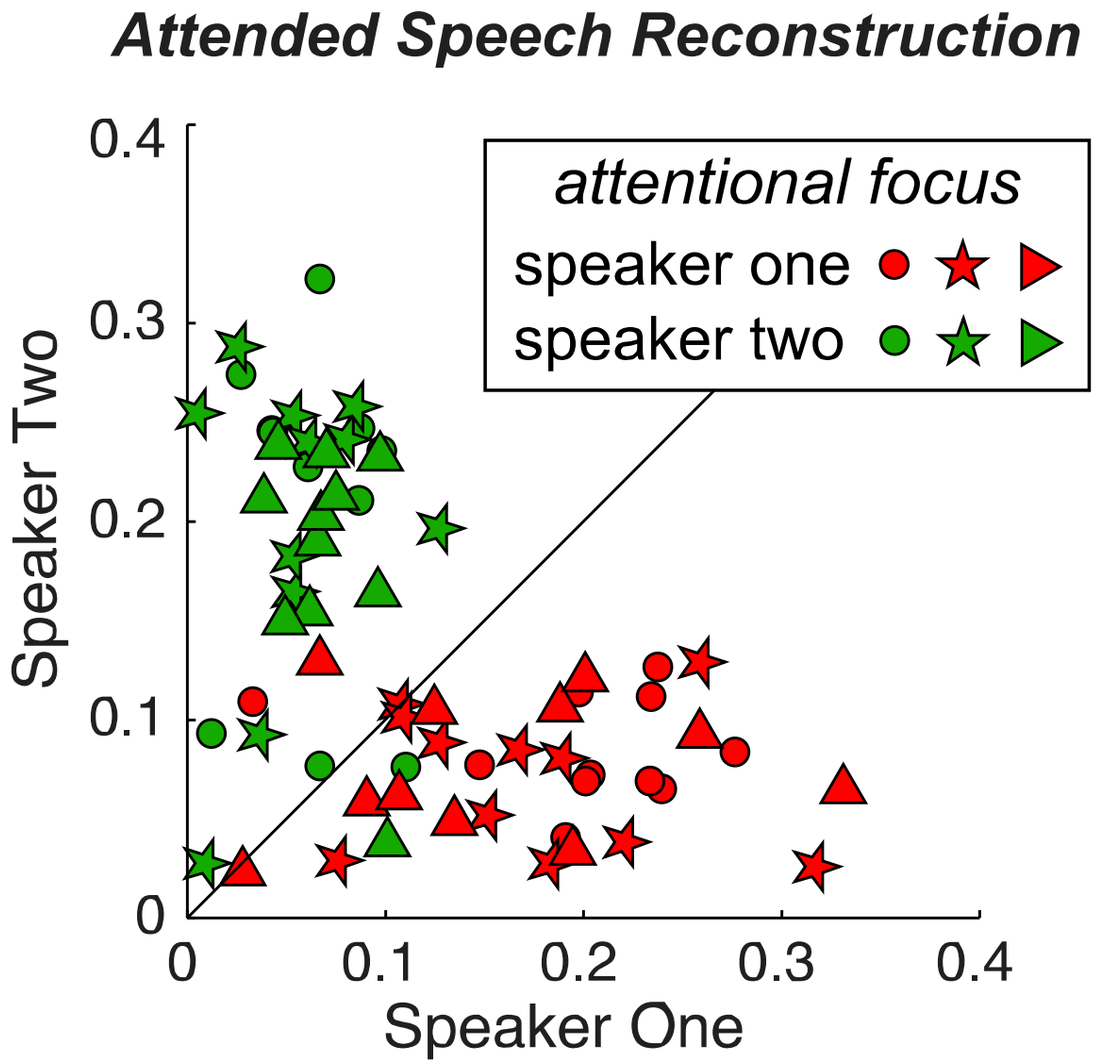
Identical Stimuli!

Stream-Specific Representation



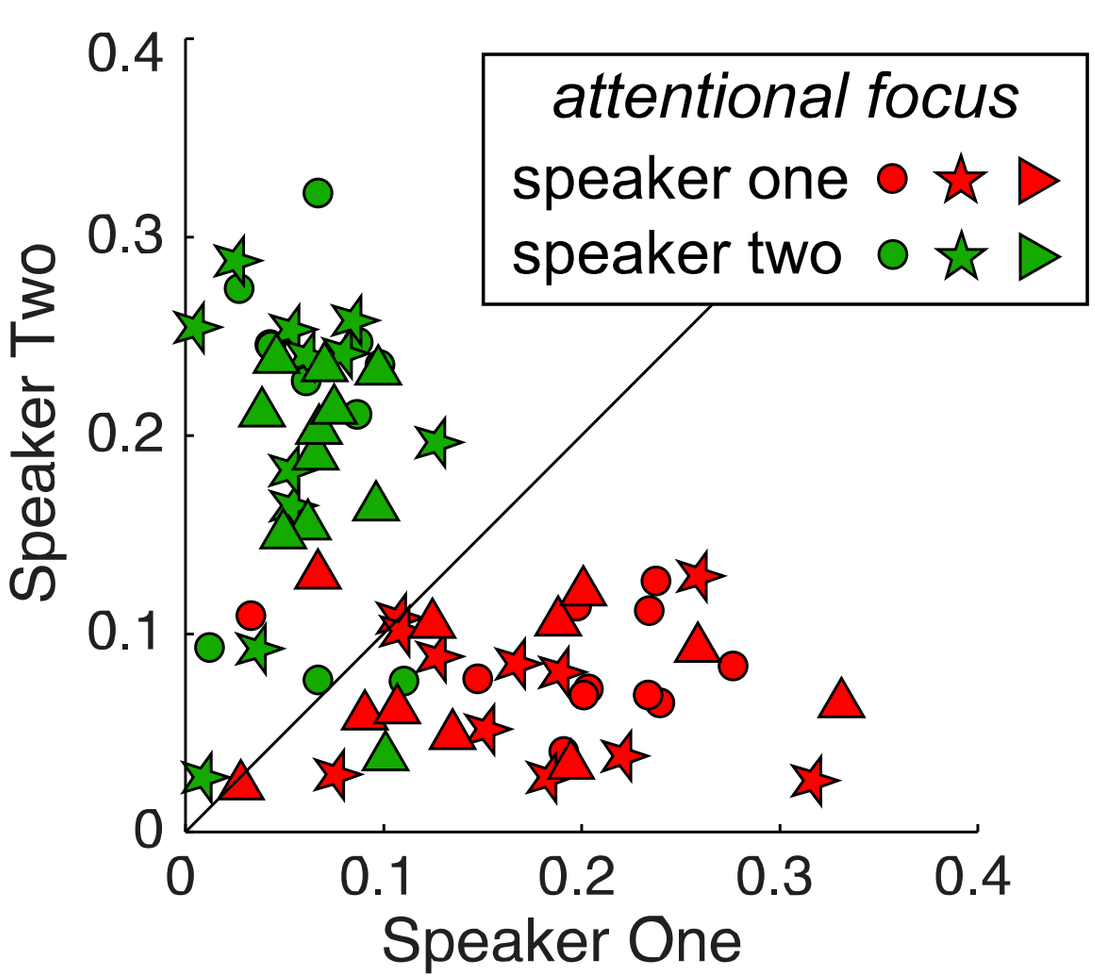
Identical Stimuli!

Single Trial Speech Reconstruction

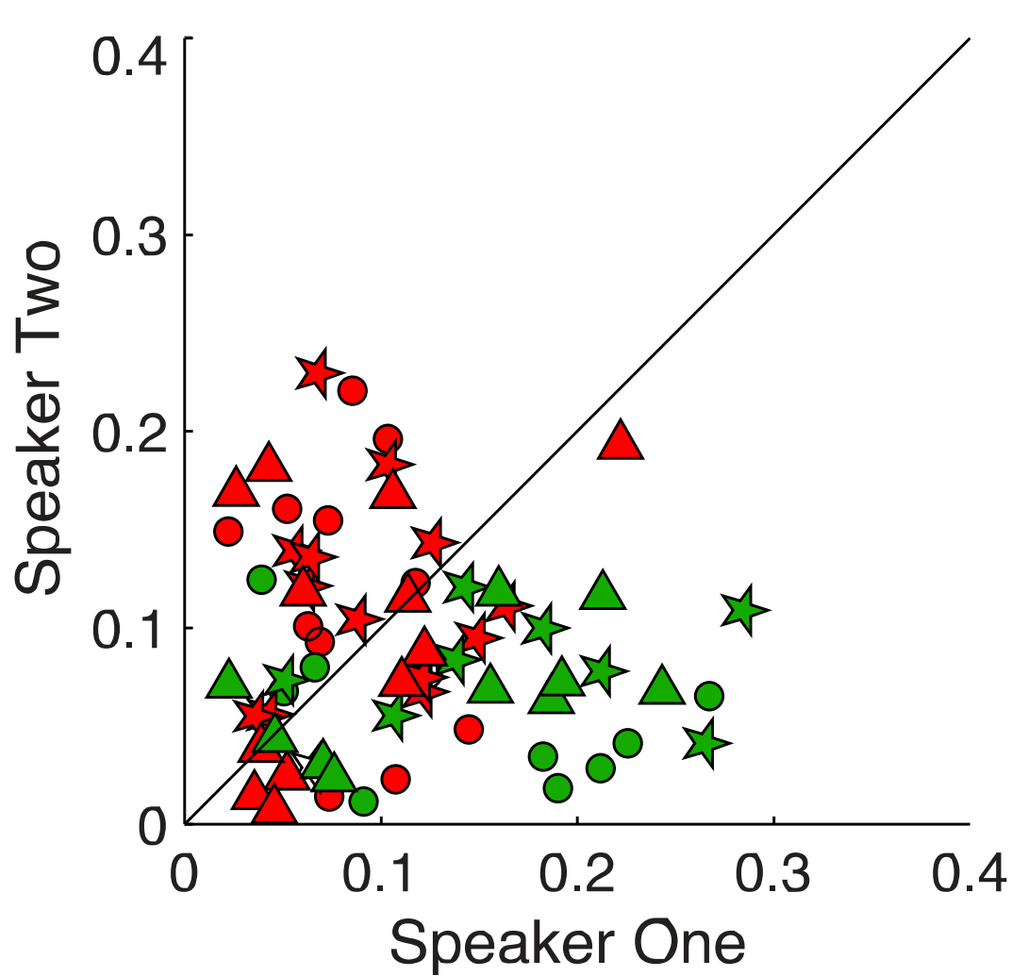


Single Trial Speech Reconstruction

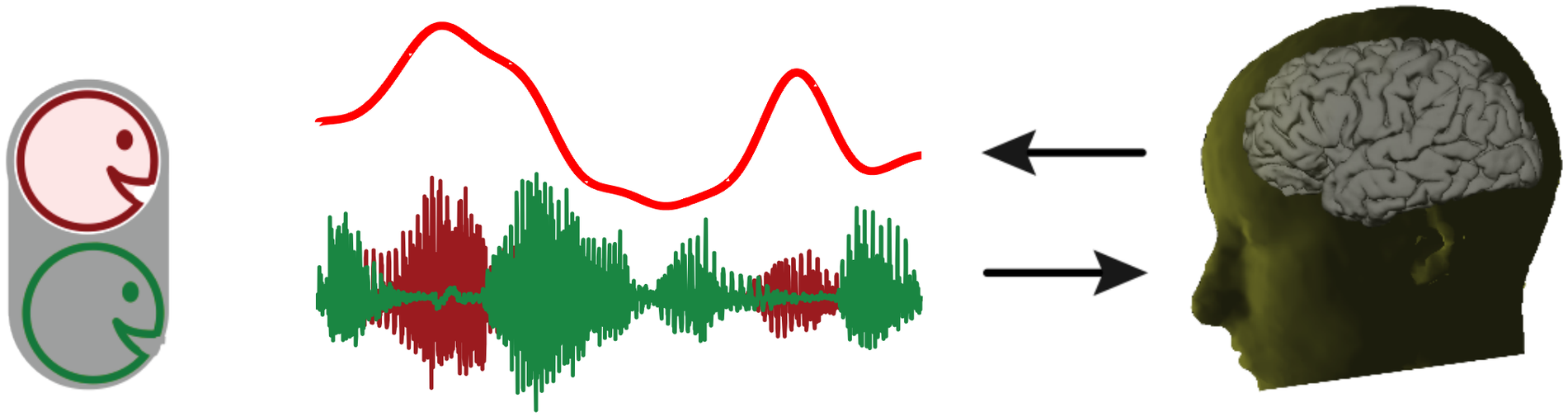
Attended Speech Reconstruction



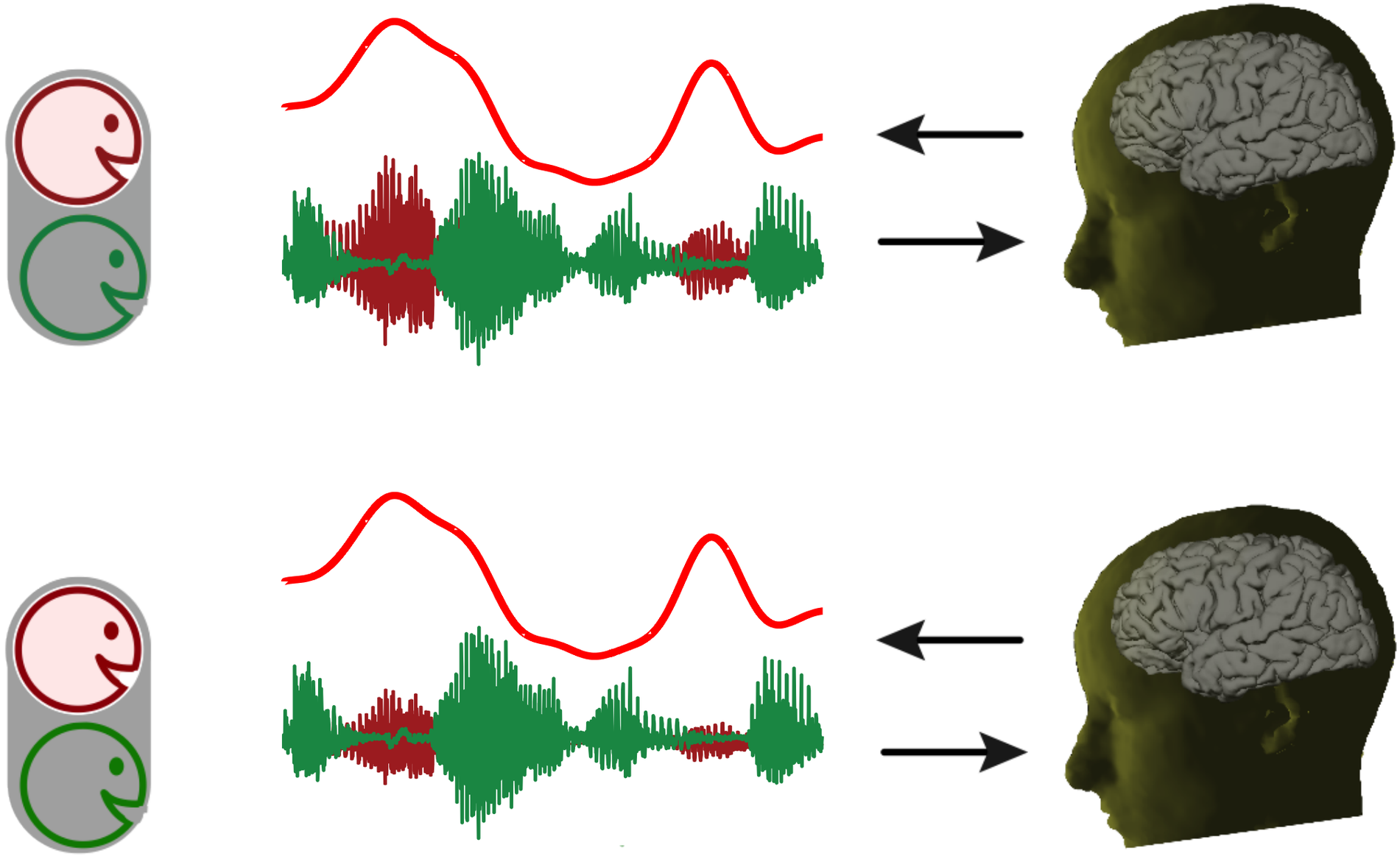
Background Speech Reconstruction



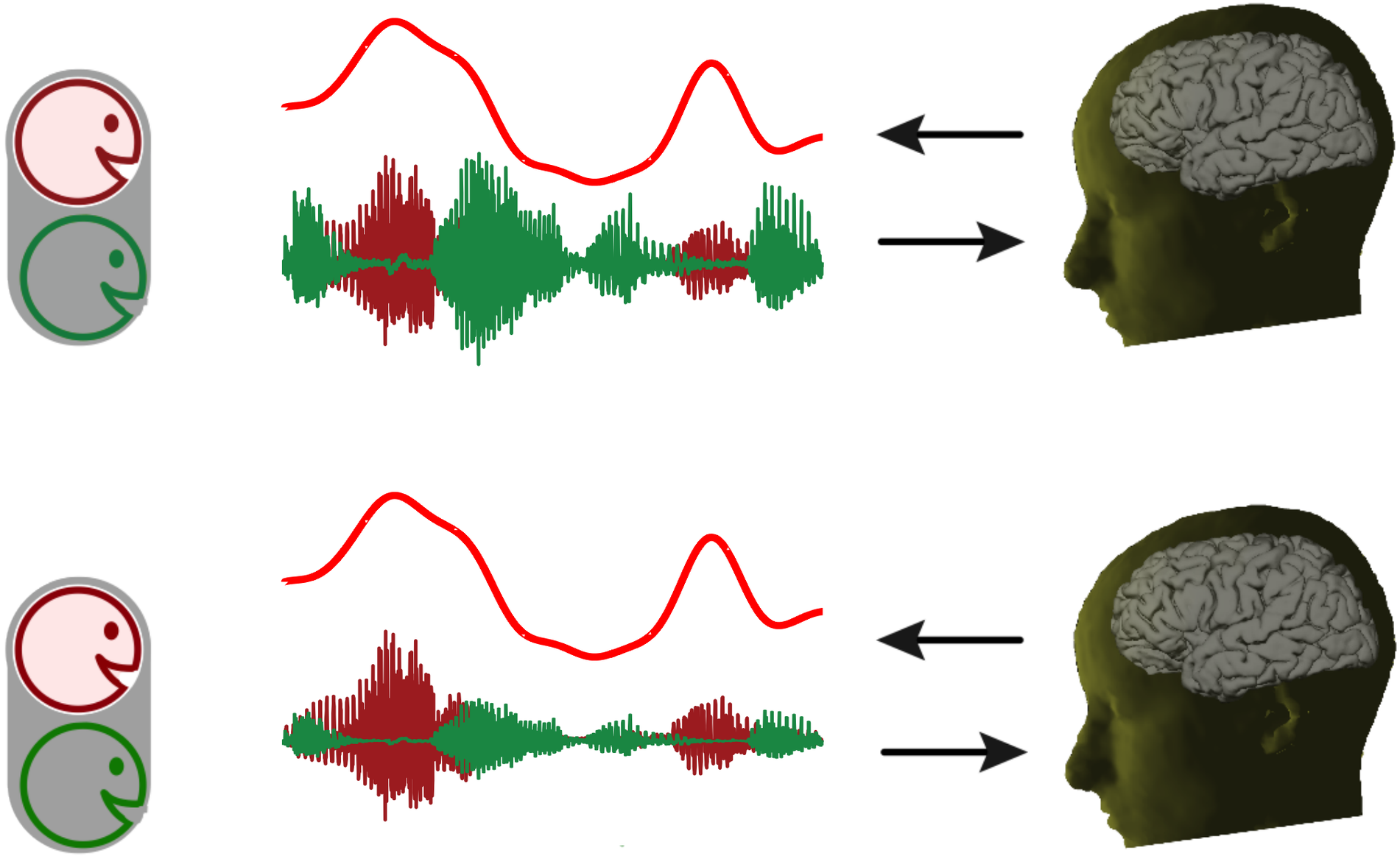
Invariance Under Acoustic Changes



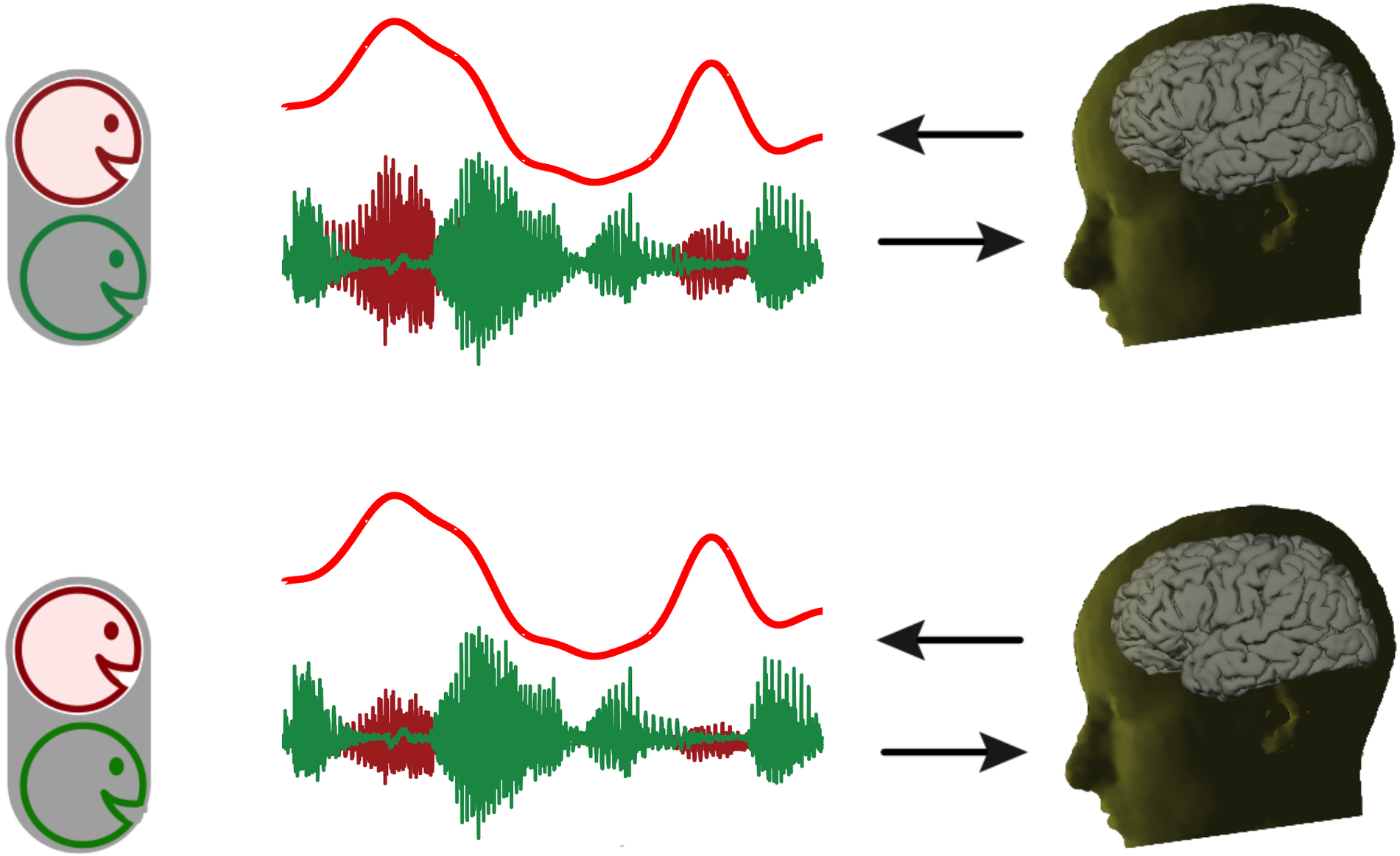
Invariance Under Acoustic Changes



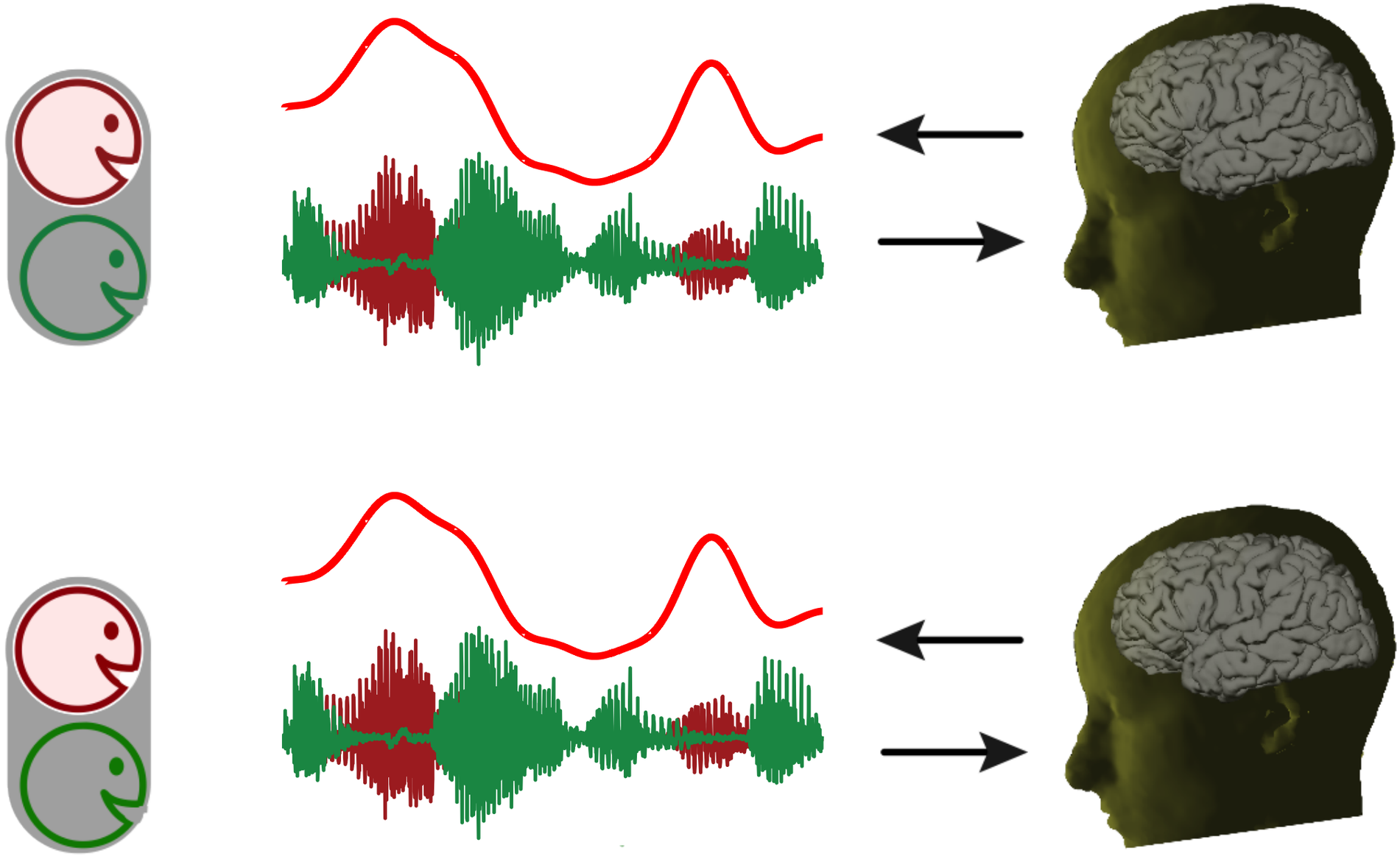
Invariance Under Acoustic Changes



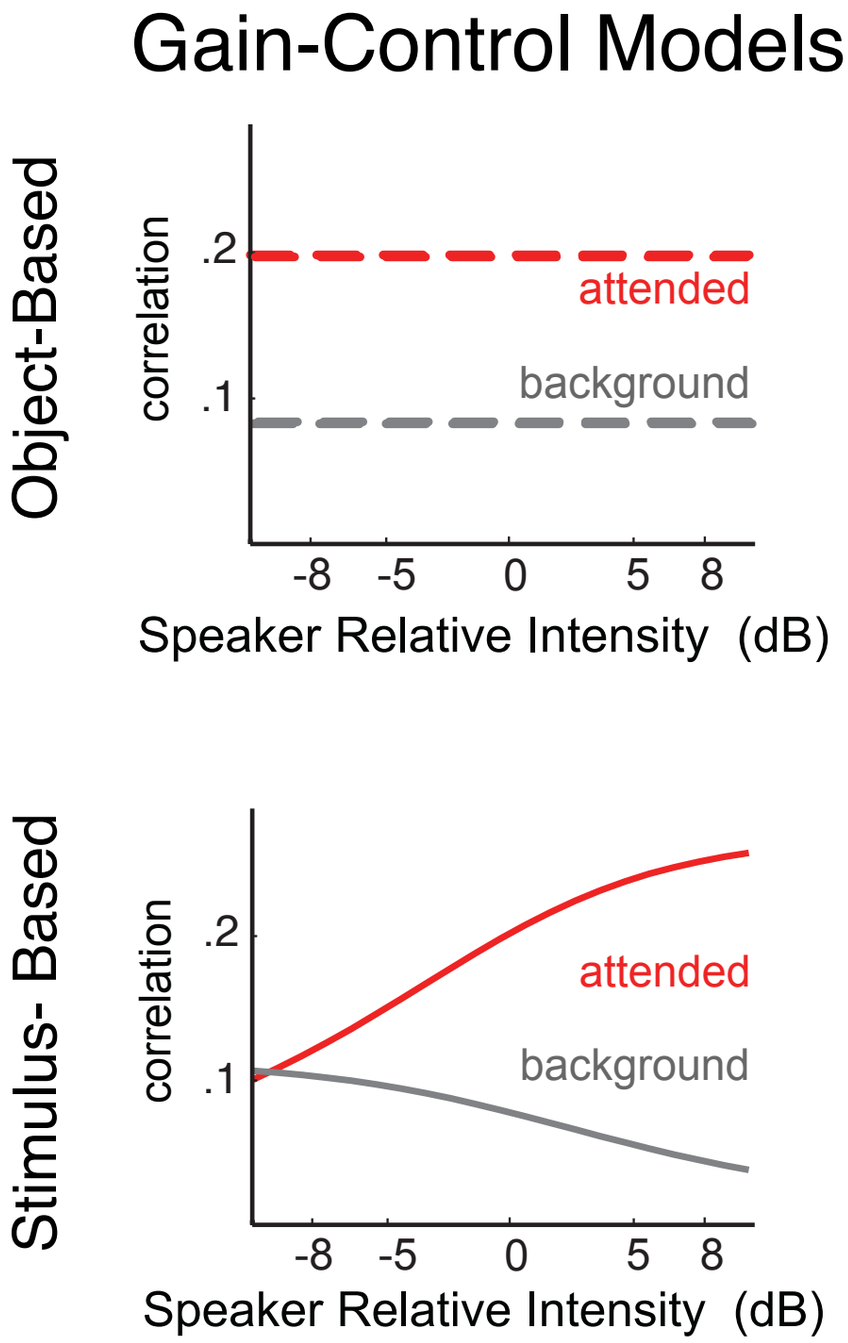
Invariance Under Acoustic Changes



Invariance Under Acoustic Changes



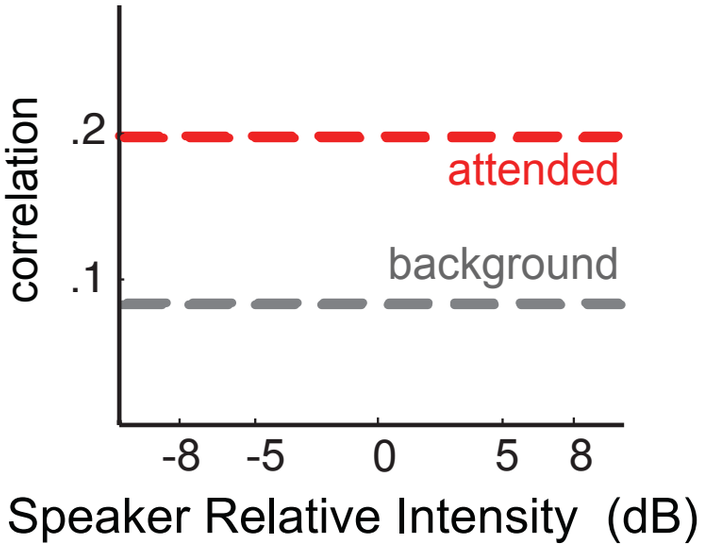
Stream-Based Gain Control?



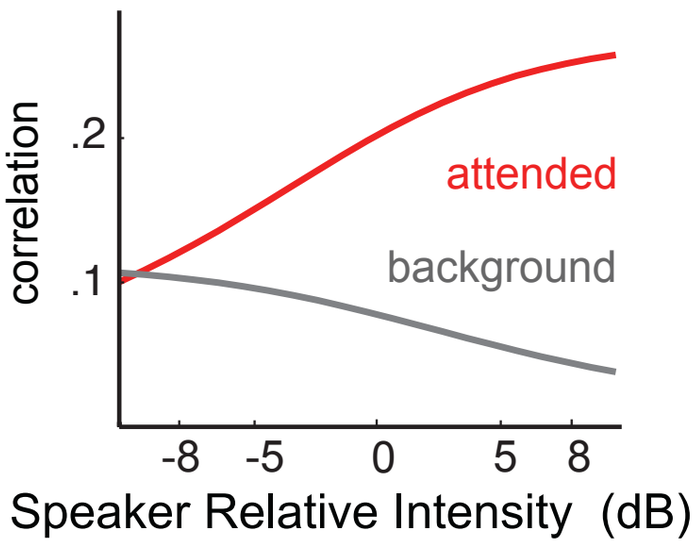
Stream-Based Gain Control?

Gain-Control Models

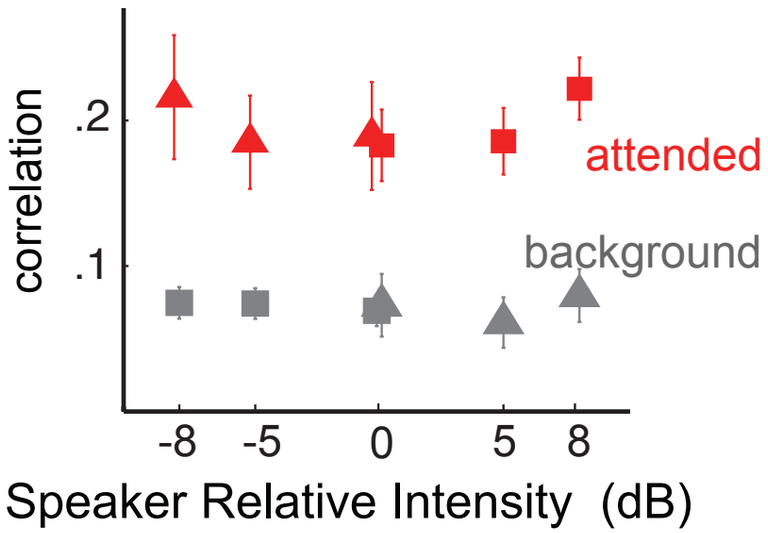
Object-Based



Stimulus- Based



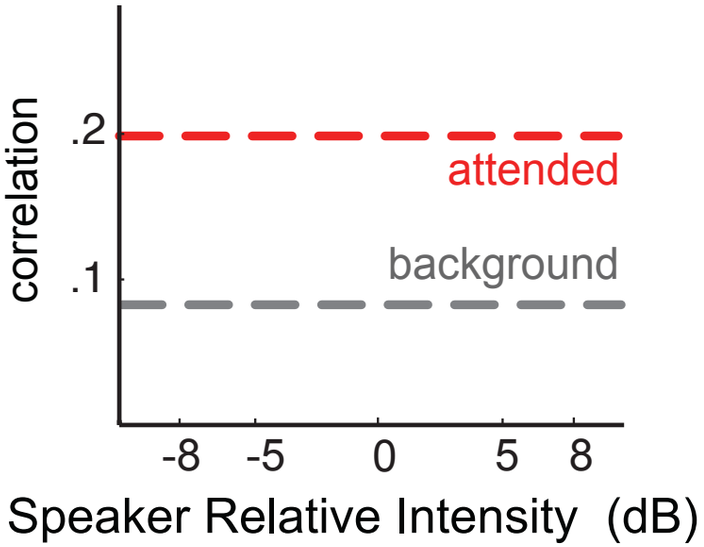
Neural Results



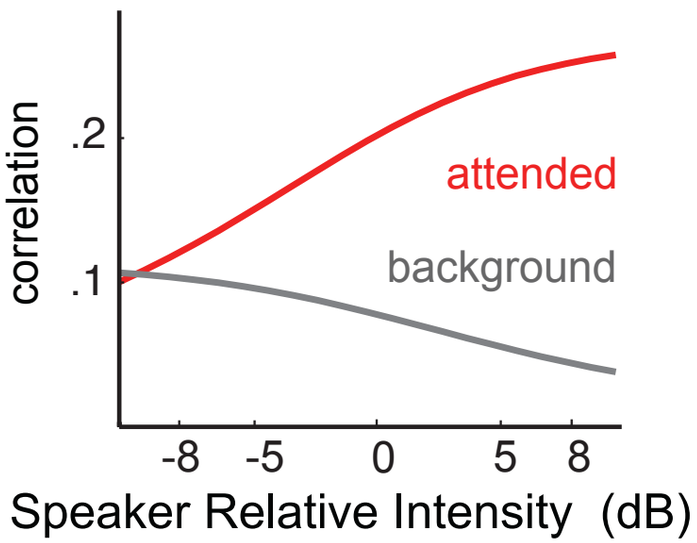
Stream-Based Gain Control?

Gain-Control Models

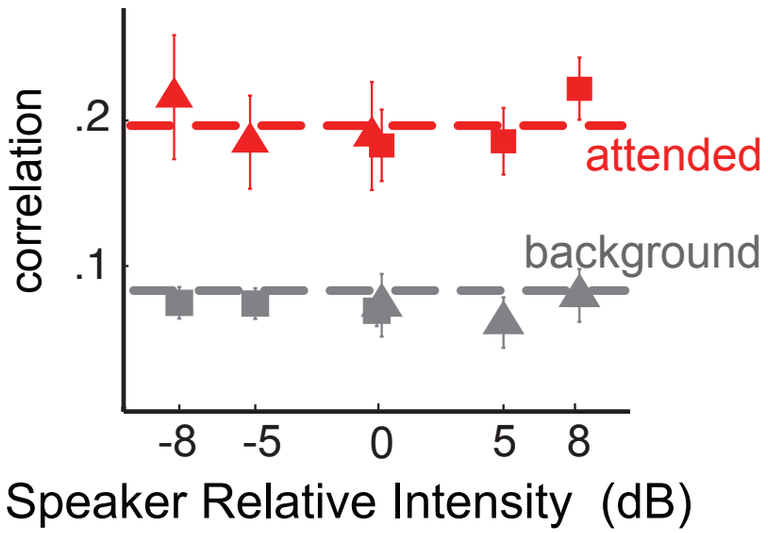
Object-Based



Stimulus- Based



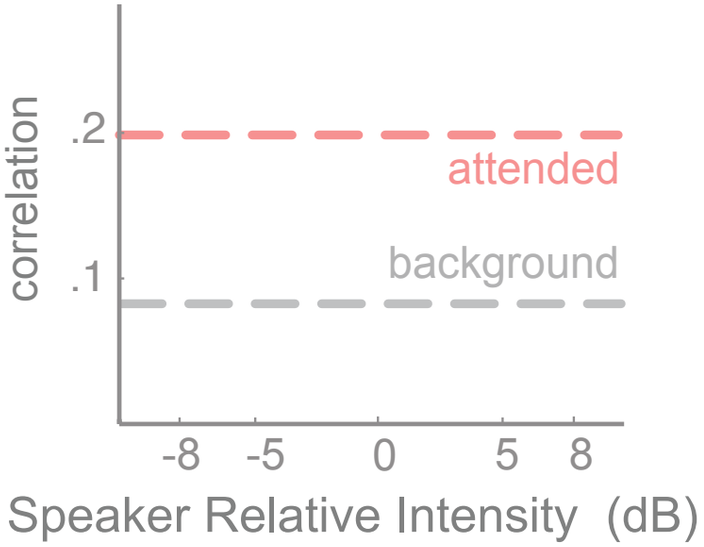
Neural Results



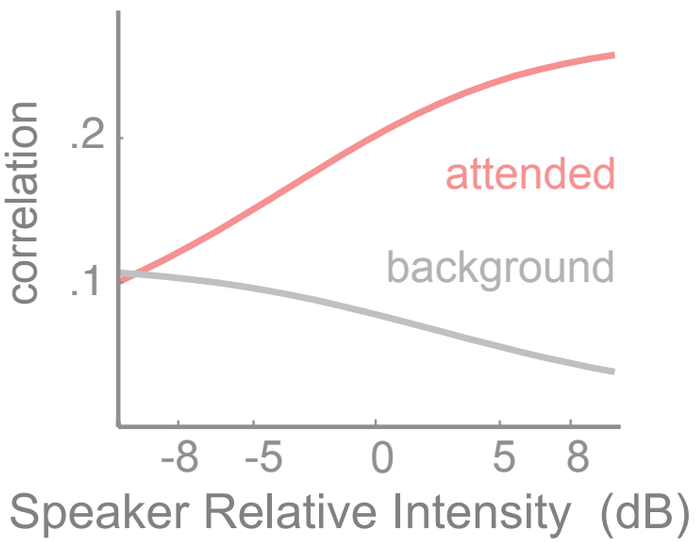
Stream-Based Gain Control?

Gain-Control Models

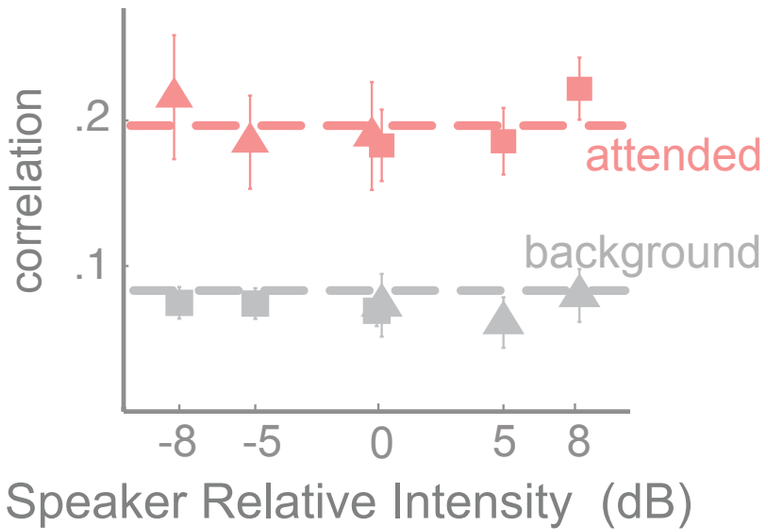
Object-Based



Stimulus- Based



Neural Results



- Stream-based not stimulus-based
- Neural representation is invariant to acoustic changes.

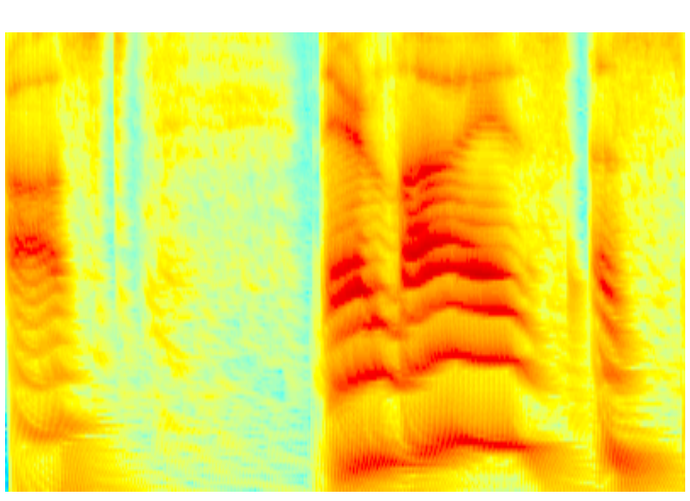
Neural Representation of an Auditory Object

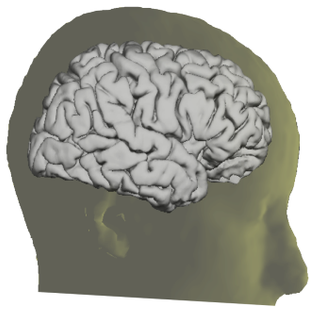
✓ neural representation is of something in sensory world

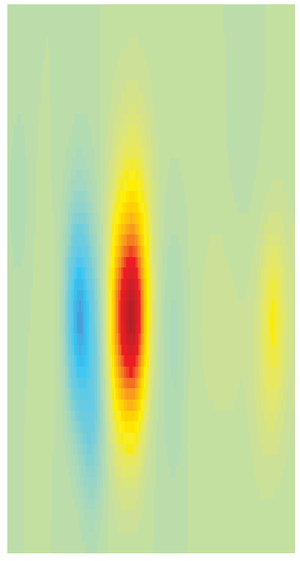
✓ when other sounds mixed in,
neural representation is of auditory object,
not entire acoustic scene

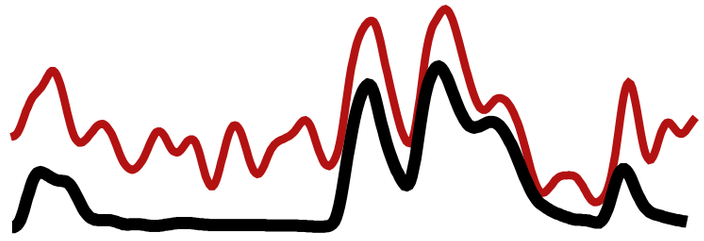
✓ neural representation invariant
under broad changes in specific acoustics

Forward STRF Model



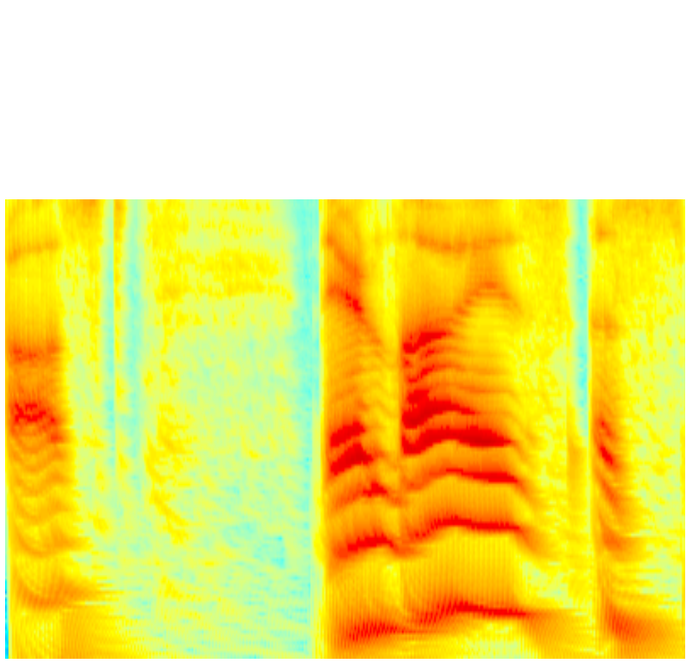


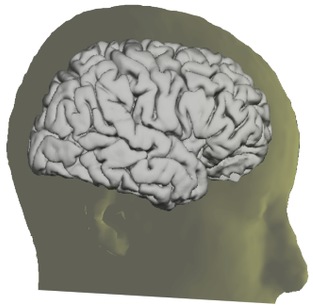


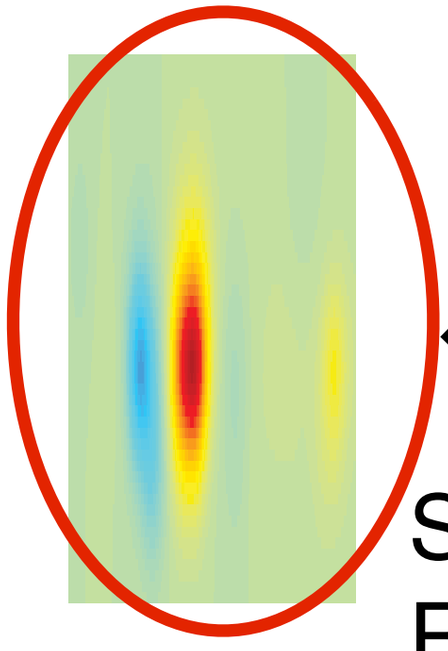


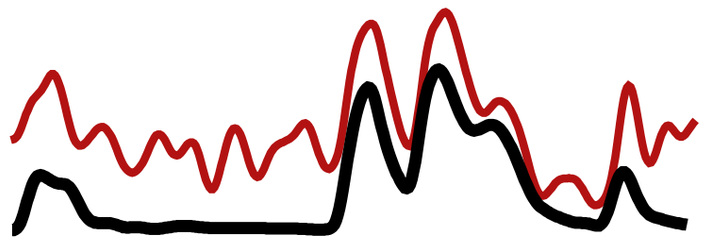
Spectro-Temporal Response Function (STRF)

Forward STRF Model



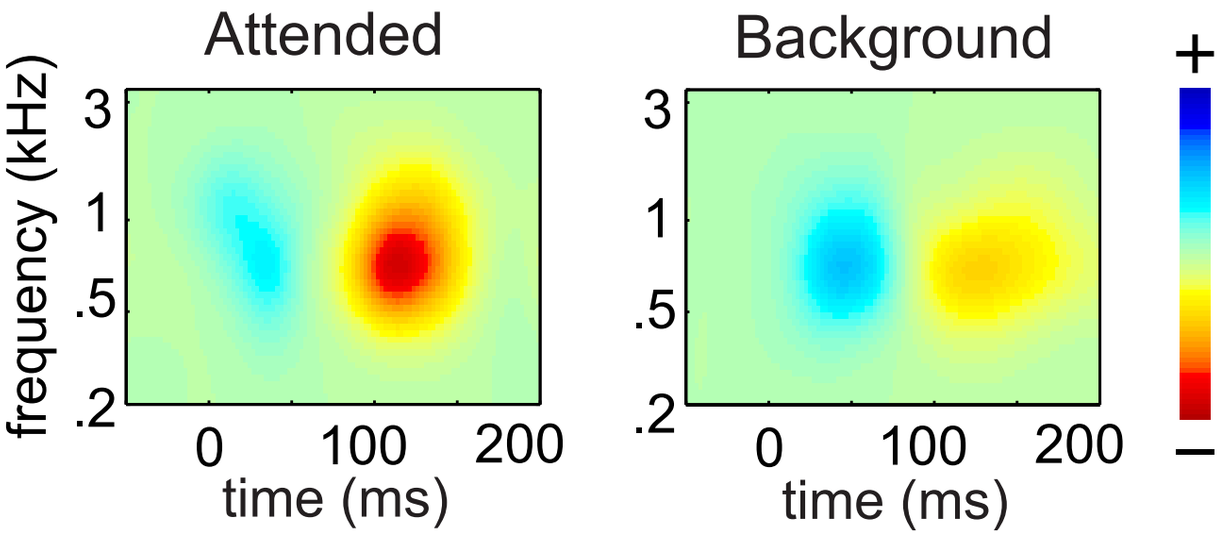






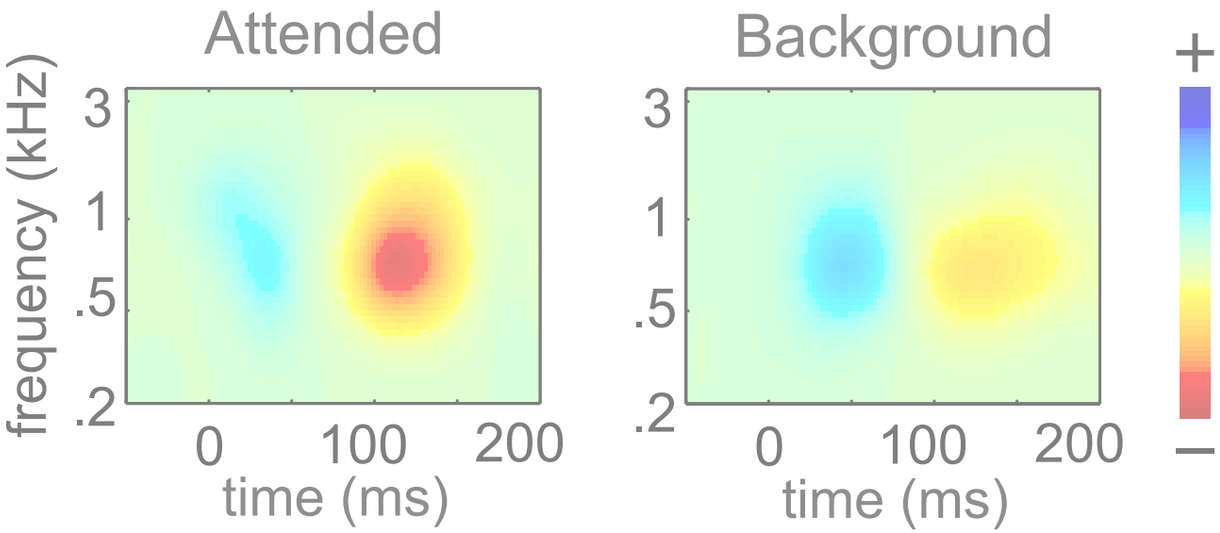
Spectro-Temporal
Response Function
(STRF)

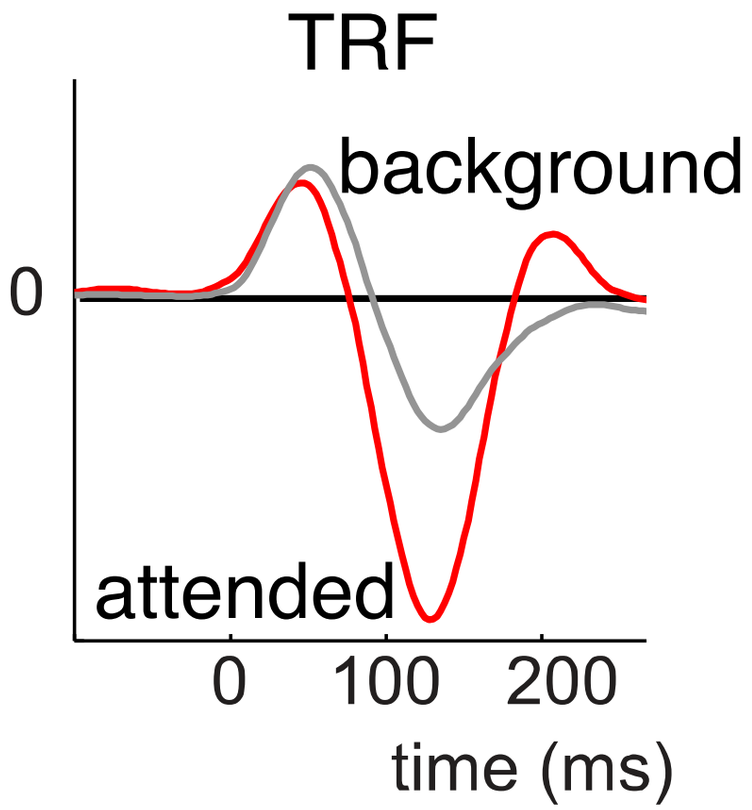
STRF Results



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

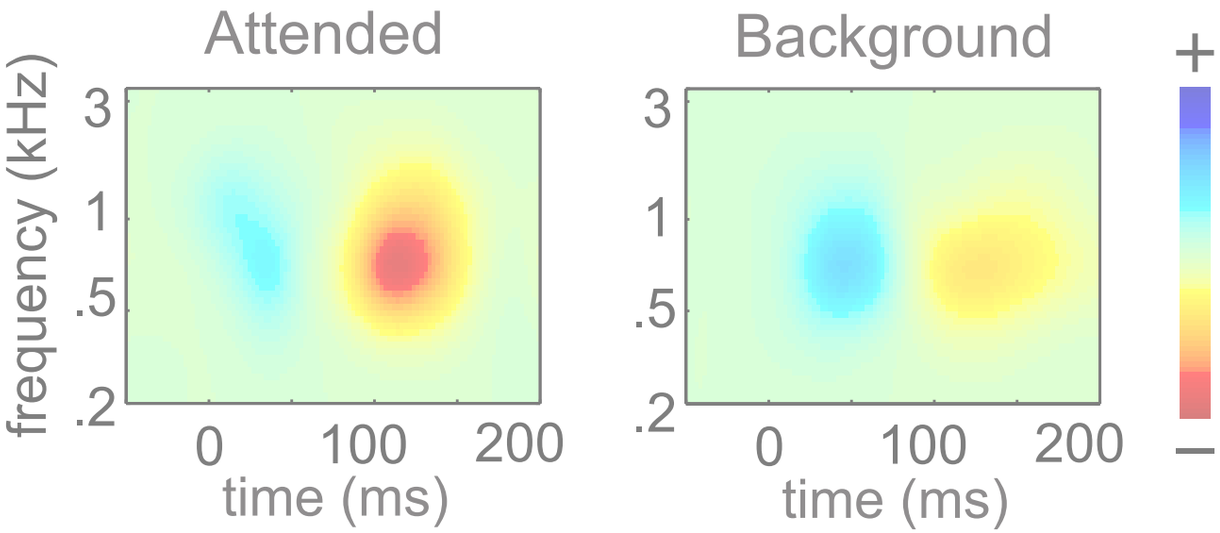
STRF Results

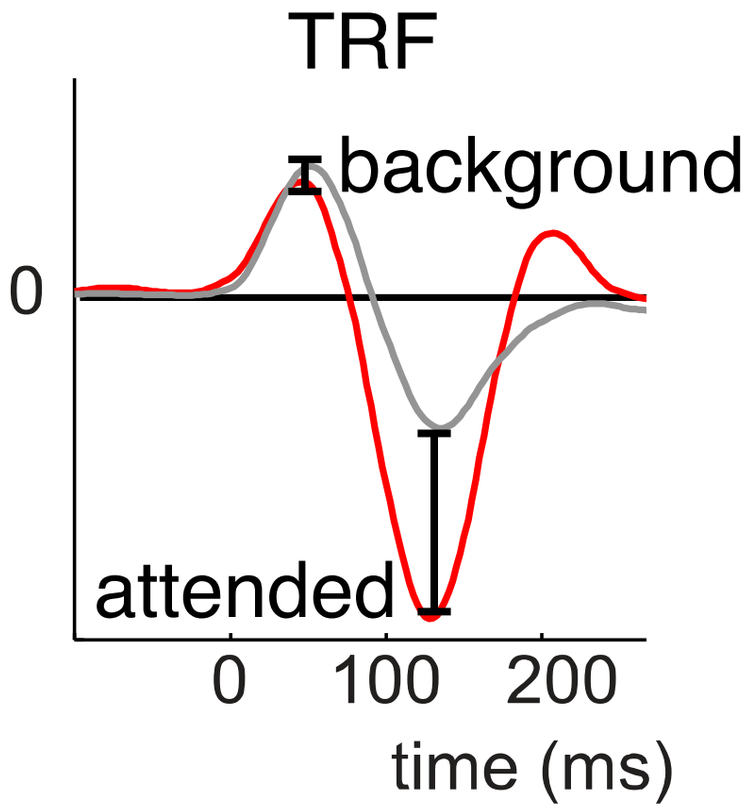




- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- $M50_{STRF}$ positive peak
- $M100_{STRF}$ negative peak

STRF Results



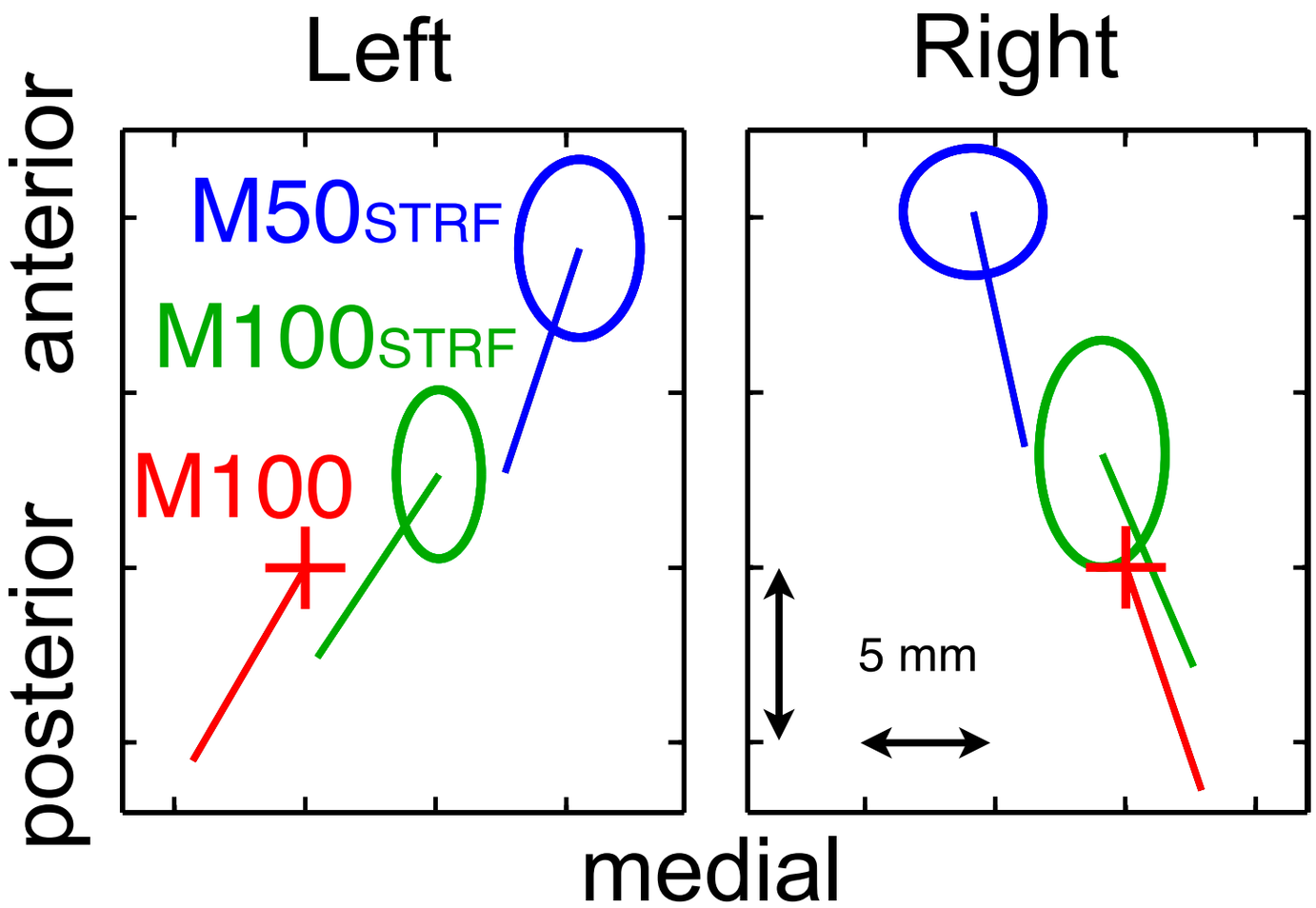


- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- $M50_{STRF}$ positive peak
- $M100_{STRF}$ negative peak
- **$M100_{STRF}$ strongly modulated by attention, *but not $M50_{STRF}$***

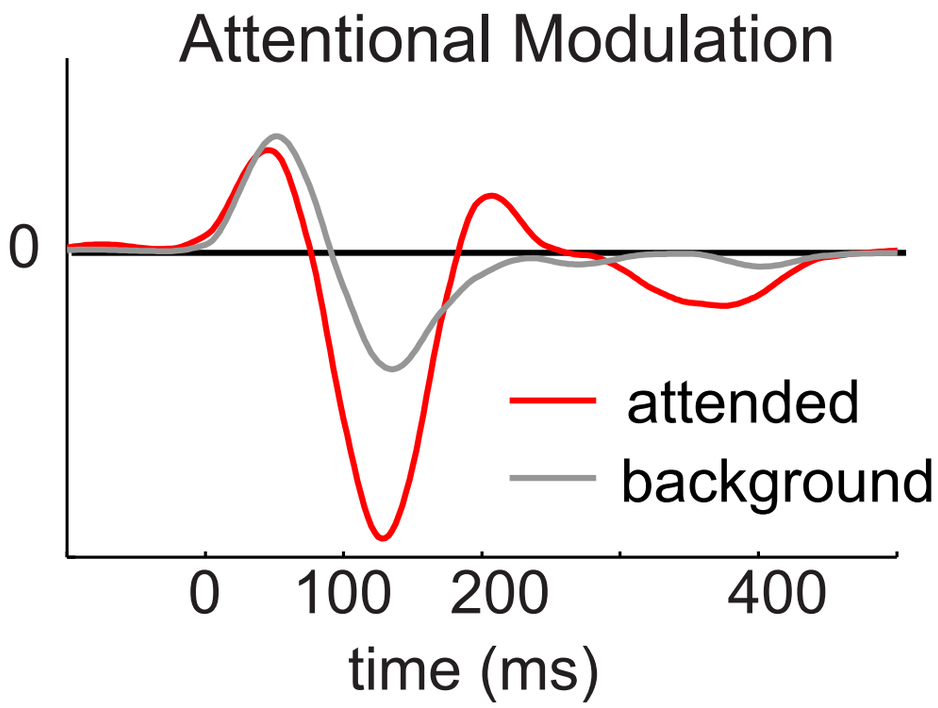
Neural Sources

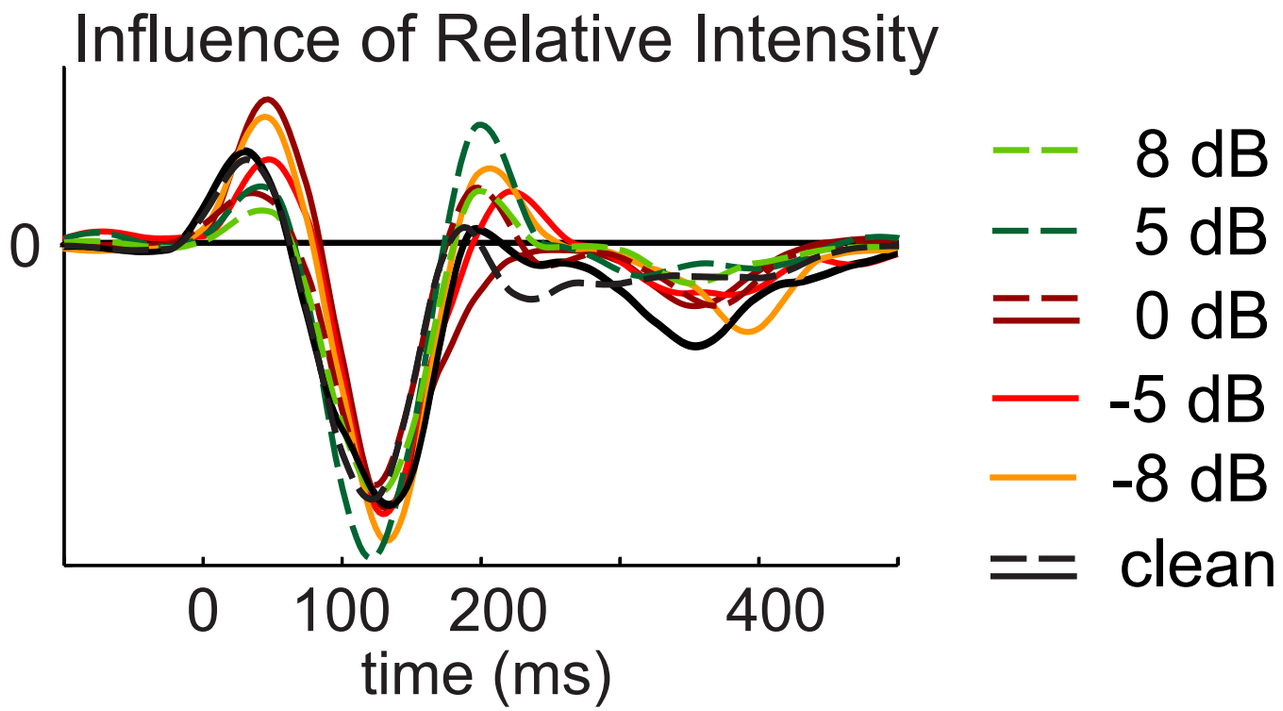
- M100$_{STRF}$ source near (same as?) M100 source:
  Planum Temporale

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



Cortical Object-Processing Hierarchy





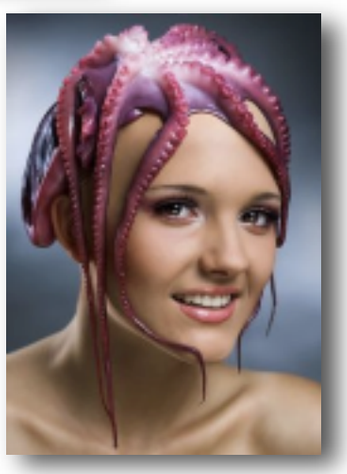
- $M100_{STRF}$ strongly modulated by attention, but not $M50_{STRF}$.
- $M100_{STRF}$ invariant against acoustic changes.
- Objects well-neurally represented at 100 ms, but not 50 ms.

Summary

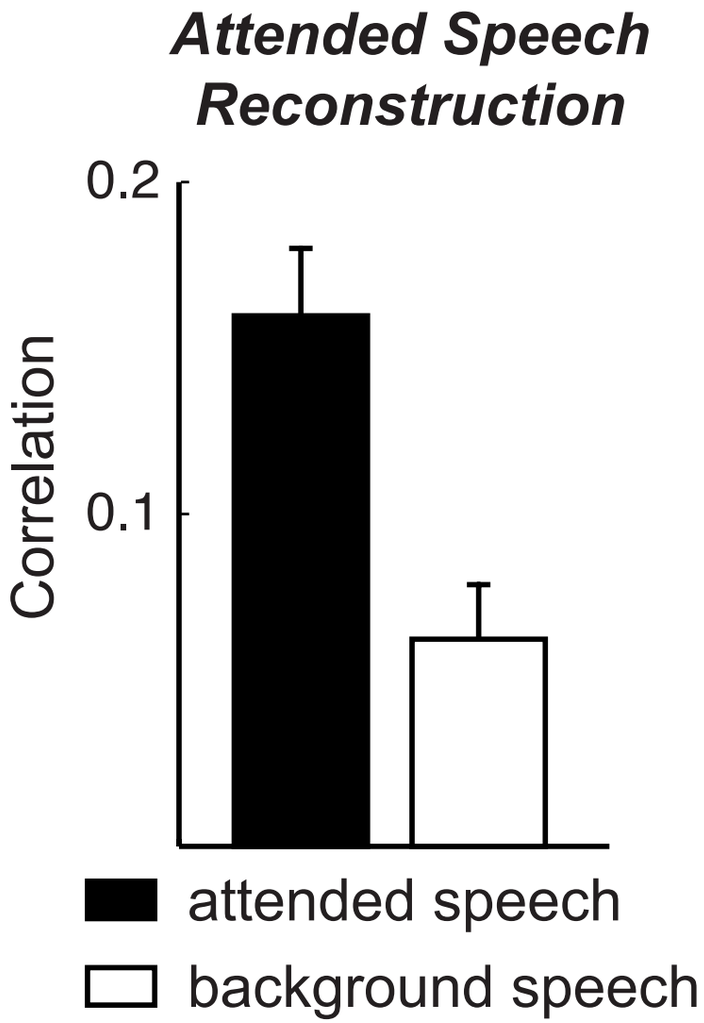
- Cortical representations of speech found here:
  - ✓ consistent with being *neural* representations of auditory *(perceptual)* objects
    - ✓ meet 3 formal criteria for auditory objects
- Object representation fully formed by 100 ms latency (PT), but not by 50 ms (HG)
- Not special to speech

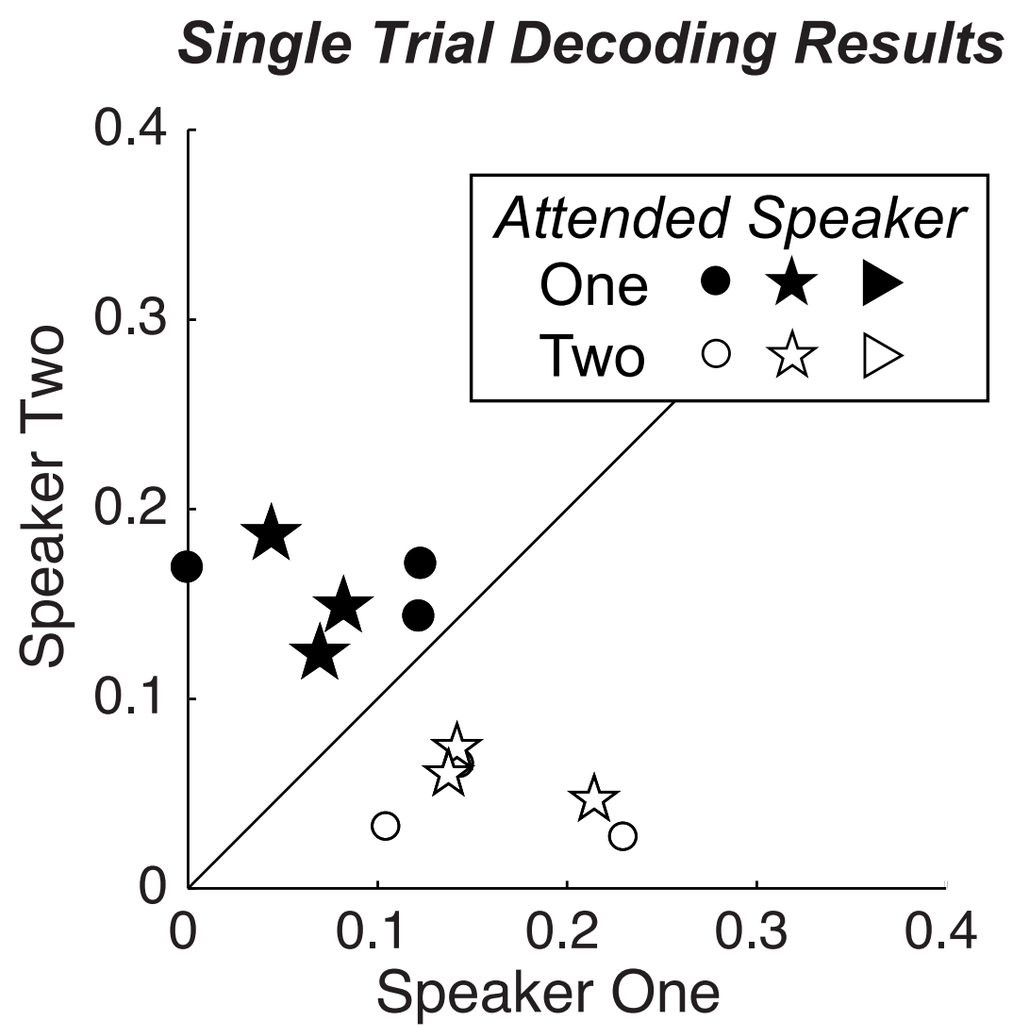
Thank You


HELLO NEW MEMBERS
NO HATS
At the monthly meeting of
Squidheads Anonymous

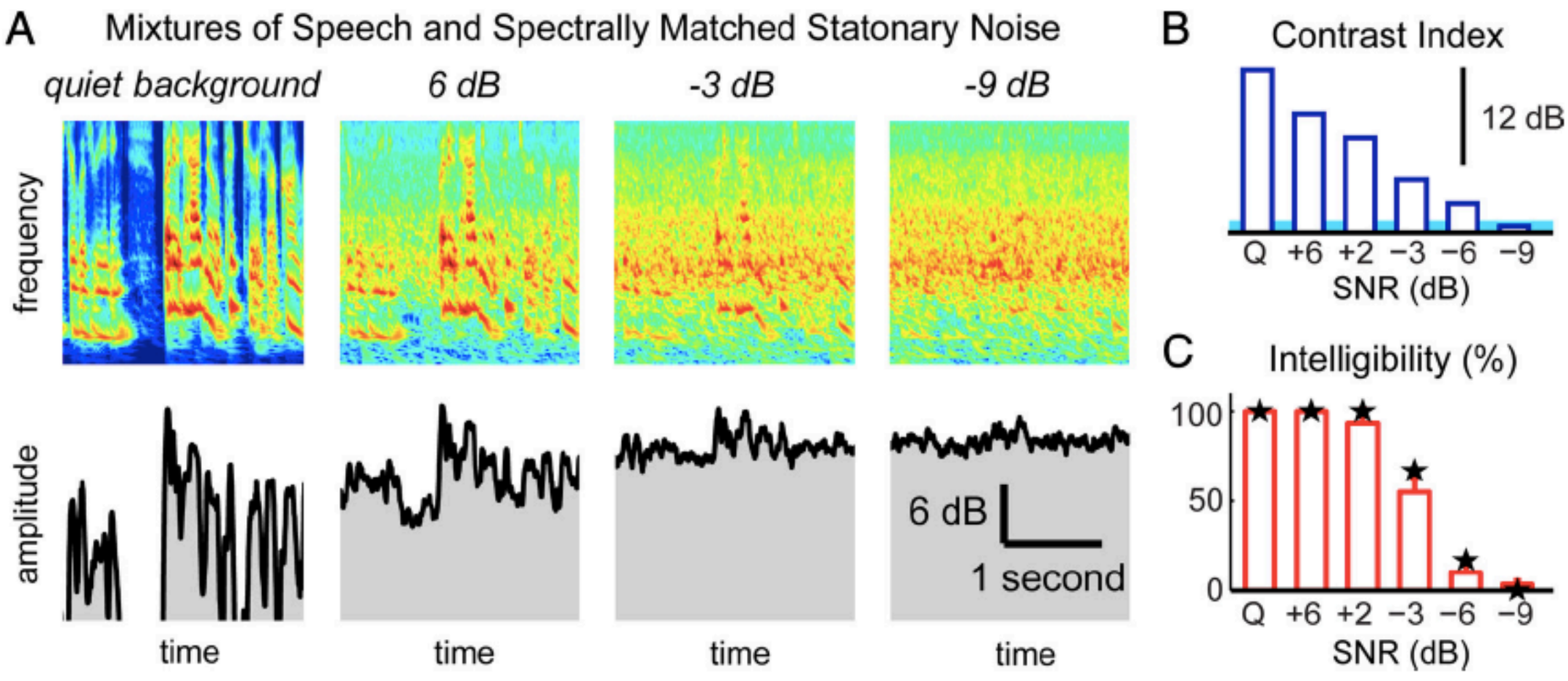


Reconstruction of Same-Sex Speech

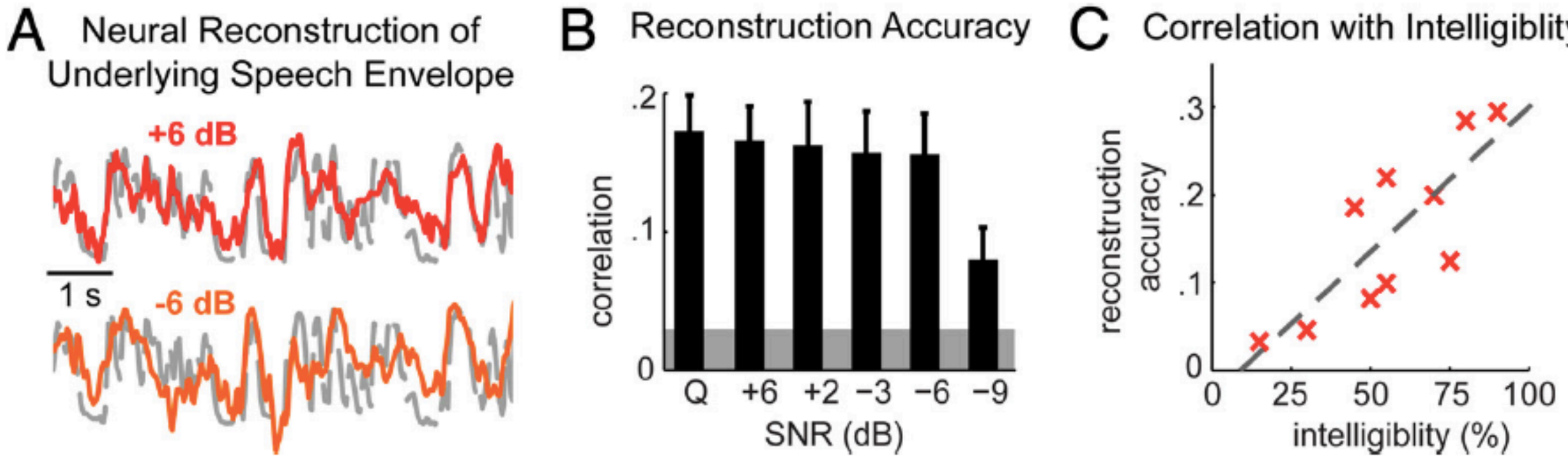




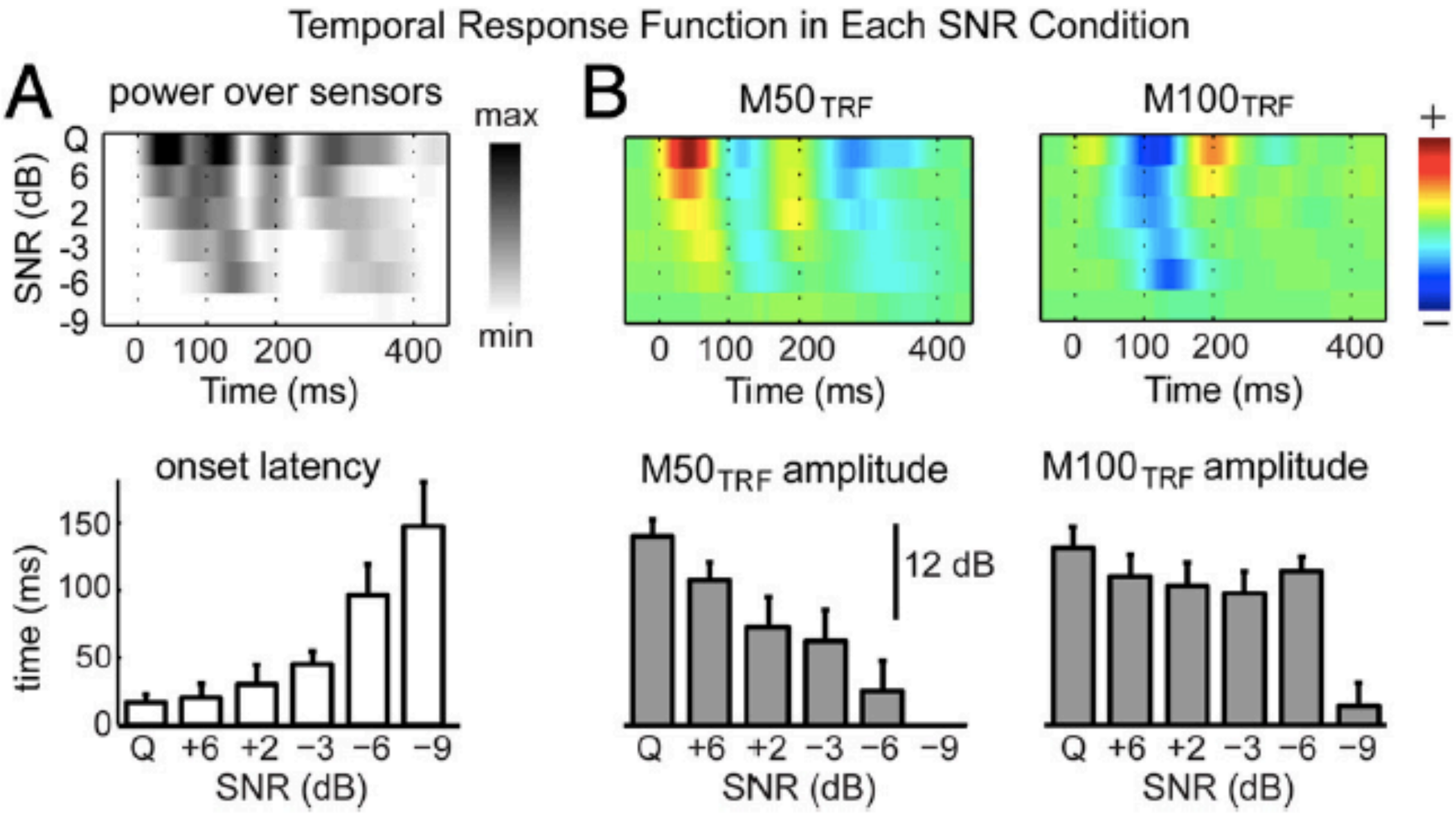
Speech in Noise: Stimuli



Speech in Noise: Results



Speech in Noise: Results



Speech in Noise: Results