Neural Representations of Speech in Human Auditory Cortex: Systems-Based Approaches

Jonathan Z. Simon

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

University of Maryland

Acknowledgements

Current Lab Members & Affiliates

Ross Baehr
Christian Brodbeck
Joshua Kulasingham
Sina Miran
David Nahmias
Krishna Puvvada
Peng Zan
Natalia Lapinskaya
Huan Luo
Mahshid Najafi
Alex Presacco
Jonas Vanthornhout
Ben Walsh
Yadong Wang
Juanjuan Xiang
Jiachen Zhuo

Past Lab Members & Affiliates

Nayef Ahmar
Sahar Akram
Murat Aytekin
Francisco Cervantes Constantino
Maria Chait
Marisel Villafane Delgado
Kim Drnec
Nai Ding
Victor Grau-Serrat
Julian Jenkins
Pirazh Khorramshahi

Collaborators

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

Past Undergraduate Students

Nicholas Asendorf
Anurupa Bhonsale
Sonja Bohr
Elizabeth Camenga
Julien Dagenais
Katya Dombrowski
Kevin Hogan
Andrea Shome
James Williams

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

Outline

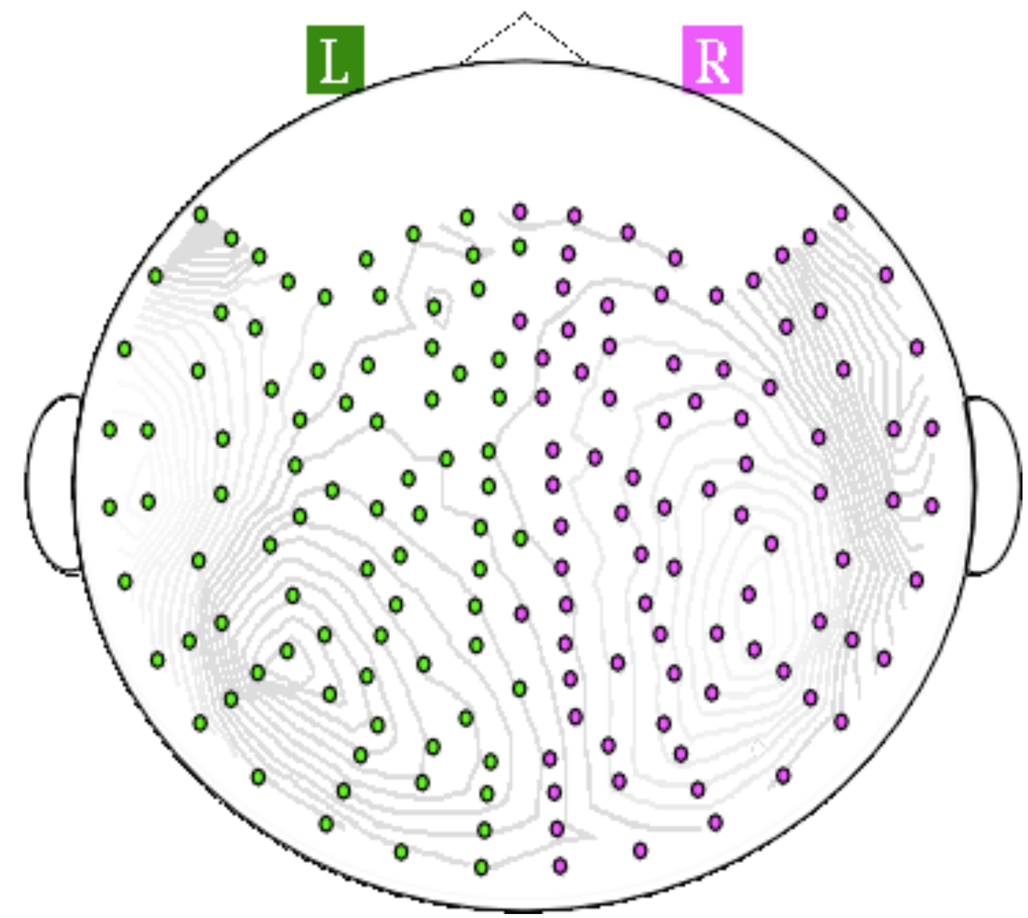
- Cortical Representations of Speech (via MEG)
  - Encoding vs. Decoding
- "Cocktail Party" Speech
- Recent Results
  - Attentional Dynamics
  - "Restoration" of Missing Speech
  - Speech Processing Across the Brain

Outline

- Cortical Representations of Speech (via MEG)
  - Encoding vs. Decoding
- "Cocktail Party" Speech
- Recent Results
  - Attentional Dynamics
  - "Restoration" of Missing Speech
  - Speech Processing Across the Brain

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

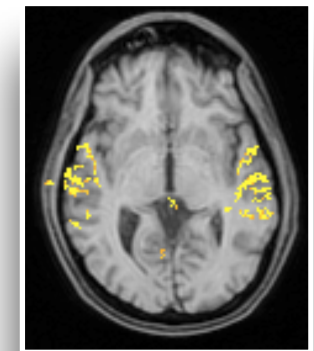


Functional Brain Imaging

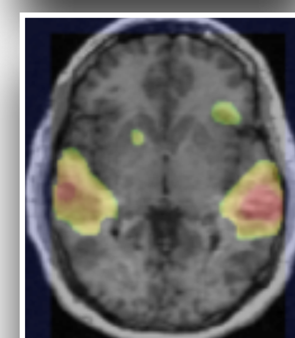
Functional Brain Imaging
= Non-invasive recording from human brain

Hemodynamic techniques

fMRI
functional magnetic resonance imaging



PET
positron emission tomography



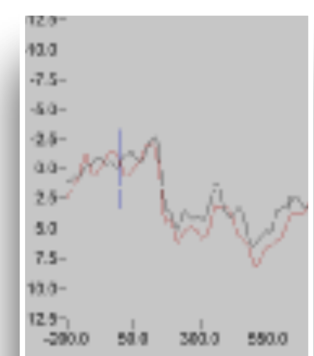
Excellent Spatial Resolution **(~1 mm)**

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

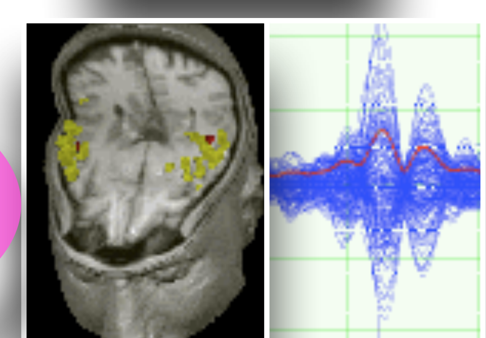
fMRI & MEG can capture effects in single subjects

Electromagnetic techniques

EEG
electroencephalography



MEG
magnetoencephalography



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

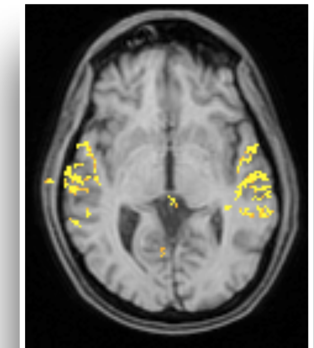
Excellent Temporal Resolution **(~1 ms)**

Functional Brain Imaging

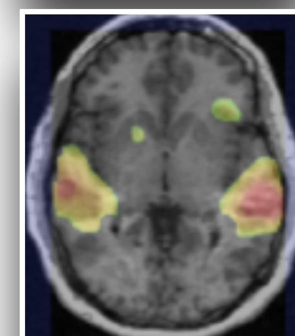
Functional Brain Imaging
= Non-invasive recording from human brain

Hemodynamic techniques

fMRI
functional magnetic resonance imaging



PET
positron emission tomography



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

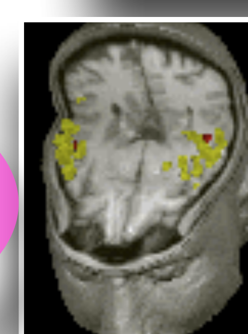
Poor Temporal Resolution **(~1 s)**

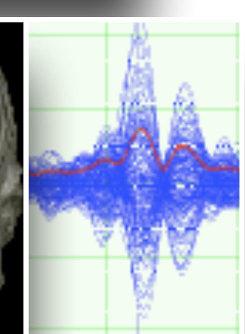
fMRI & MEG can capture effects in single subjects

Electromagnetic techniques

EEG
electroencephalography

MEG
magnetoencephalography



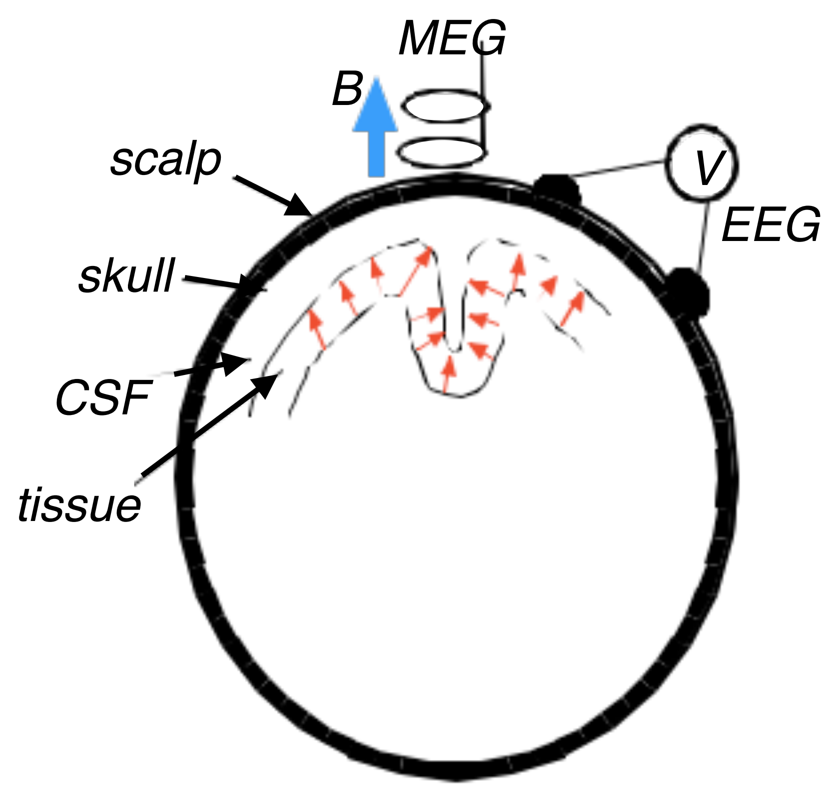


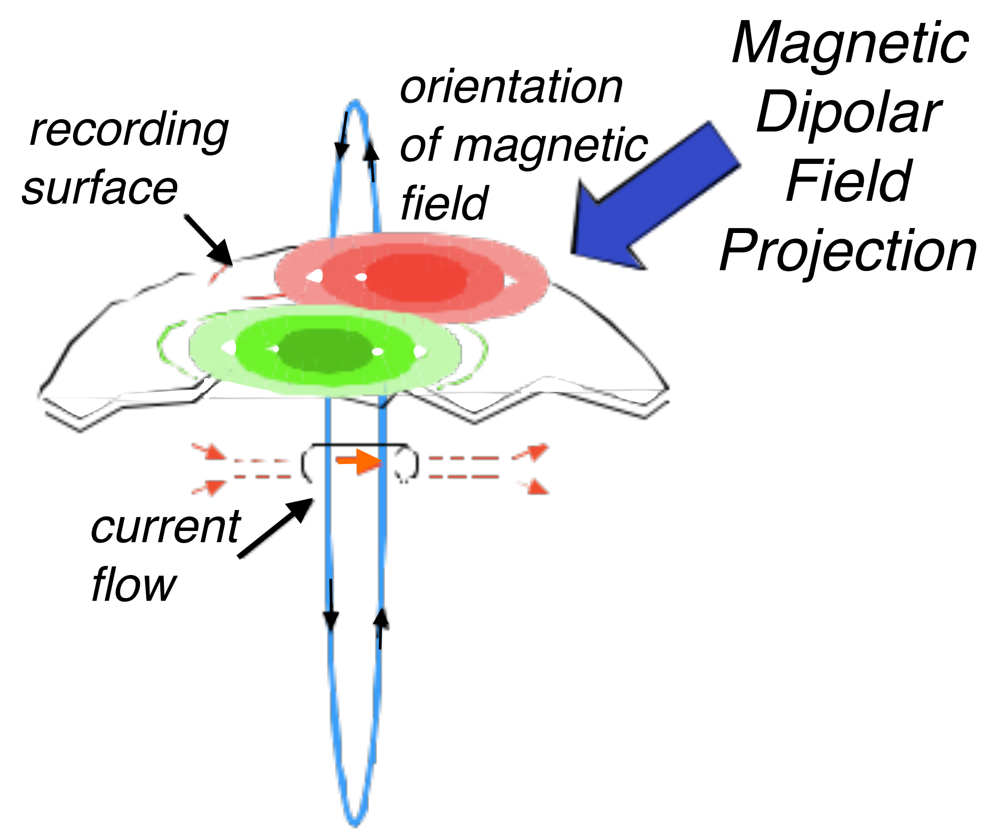
Poor Spatial Resolution **(~1 cm)**

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

Neural Signals & MEG

Photo by Fritz Goro

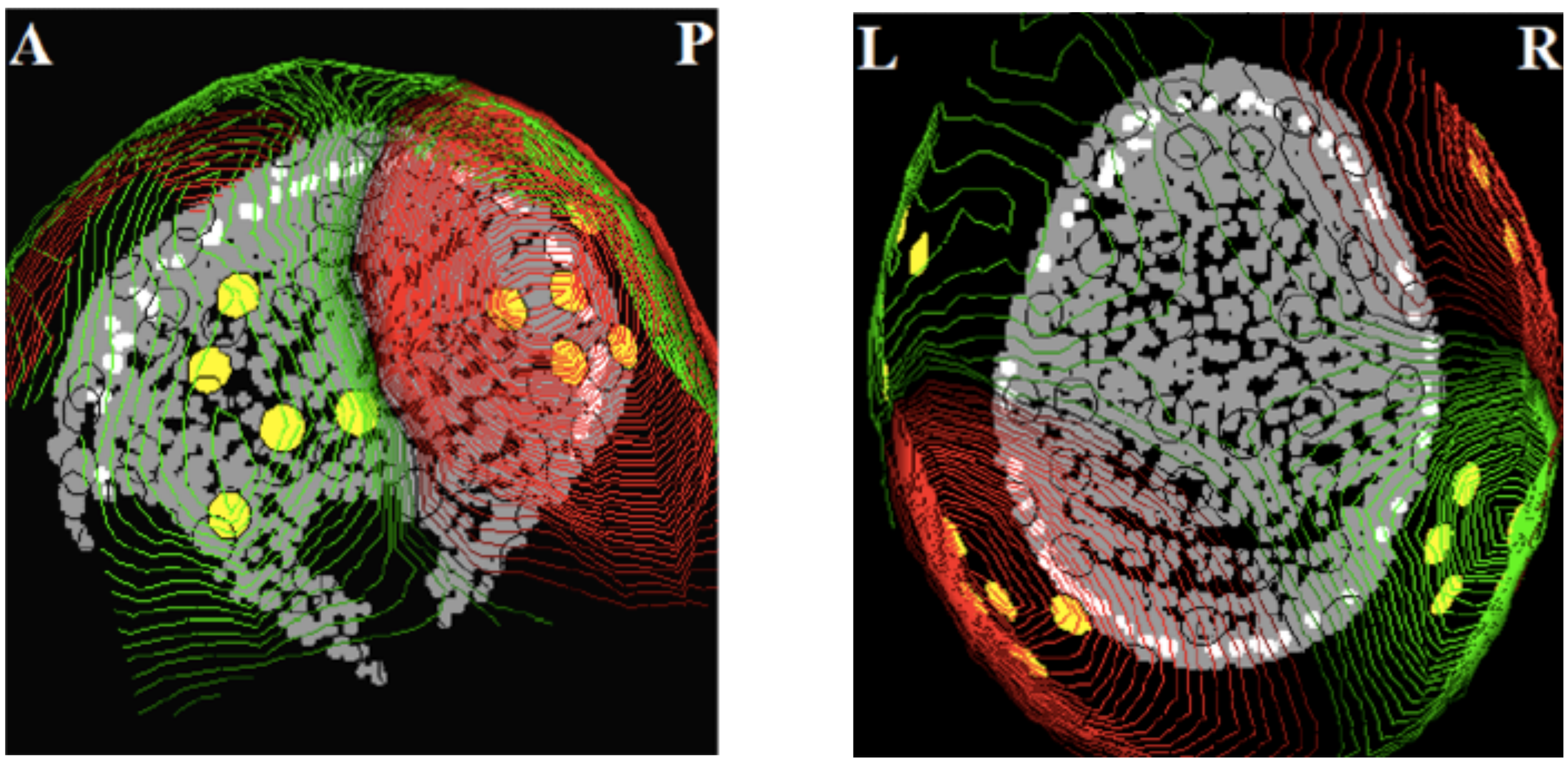




- Direct electrophysiological measurement
  - not hemodynamic
  - real-time
- No unique solution for distributed source

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

MEG Auditory Field

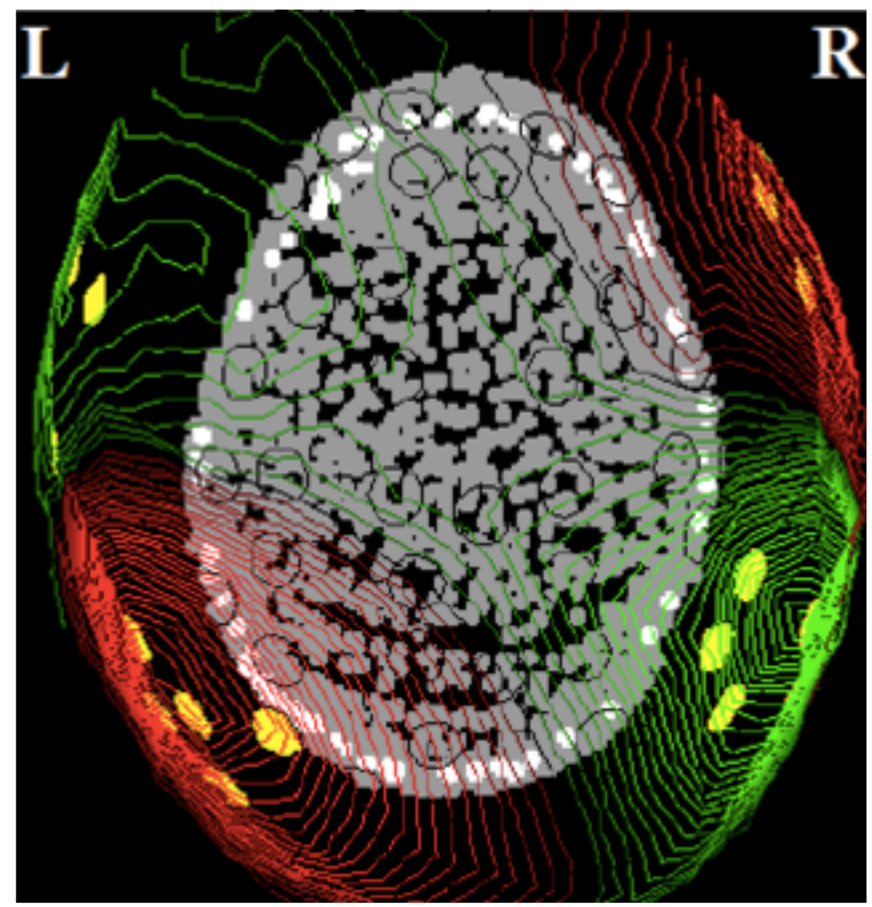


Sagittal View   Axial View

Strongly
Lateralized

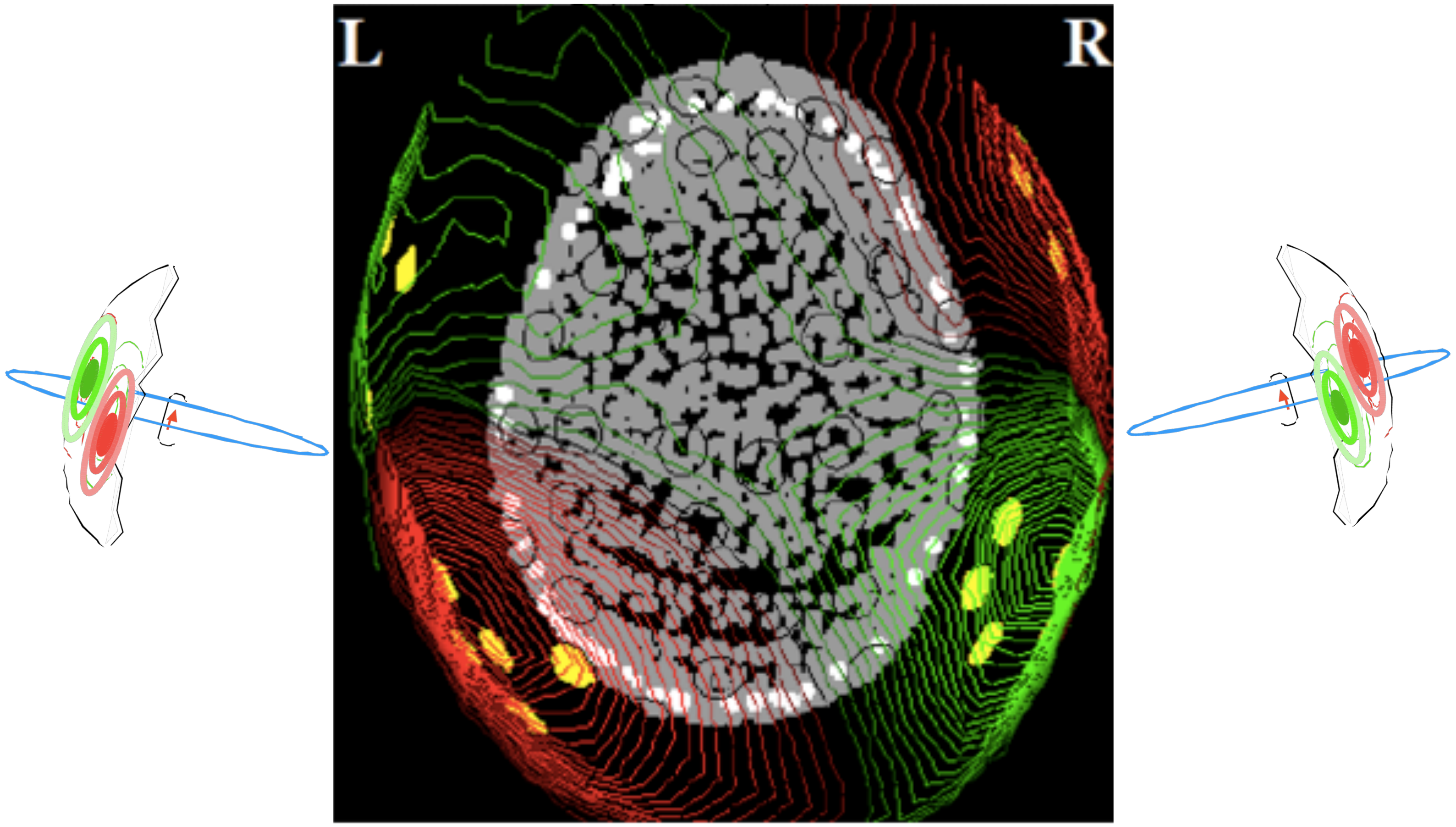
Chait, Poeppel and Simon, Cerebral Cortex (2006)

MEG Auditory Field



Chait et al., Cerebral Cortex (2006)

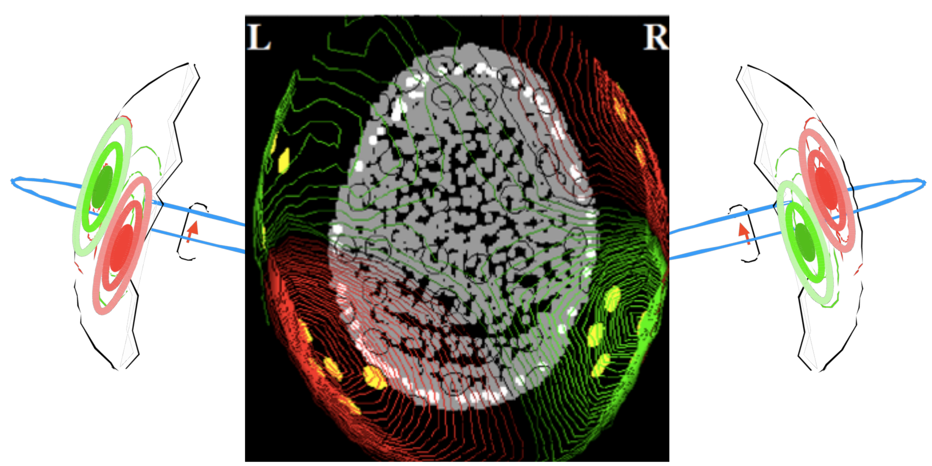
MEG Auditory Field

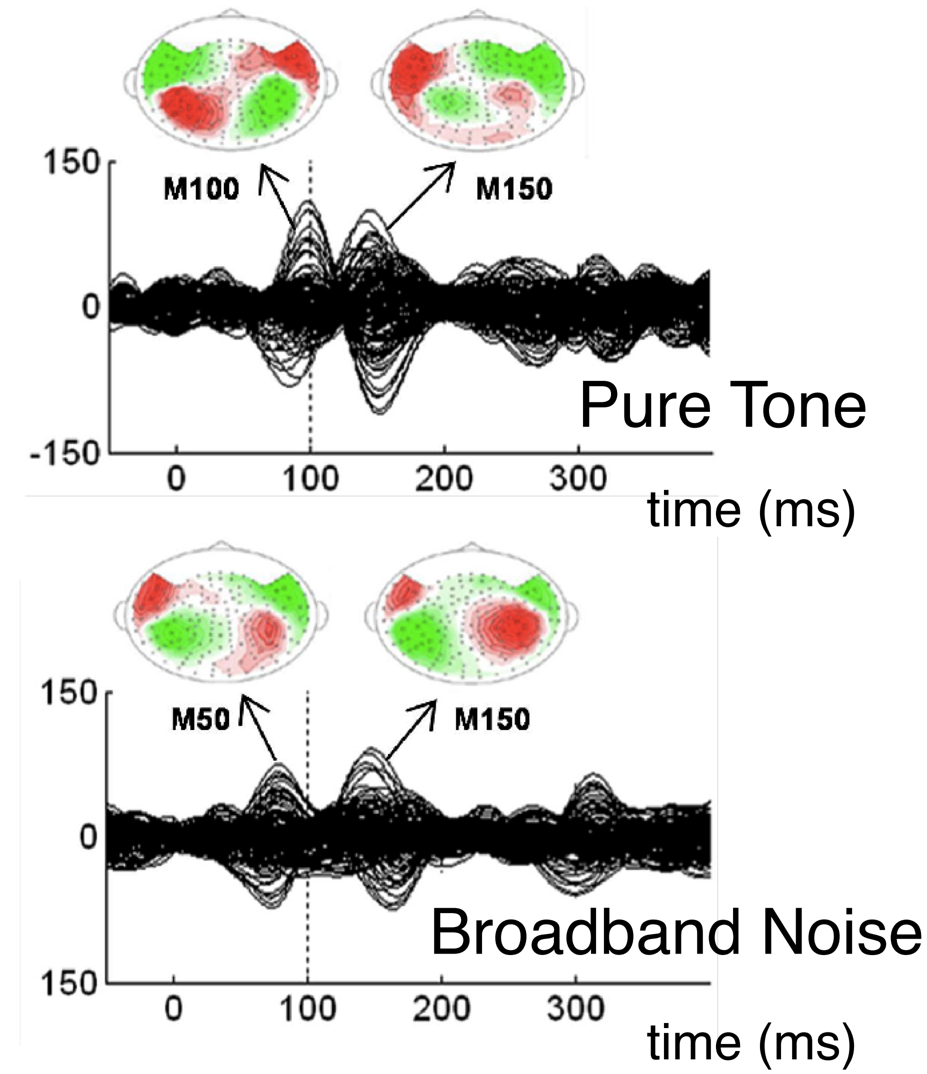


Chait et al., Cerebral Cortex (2006)

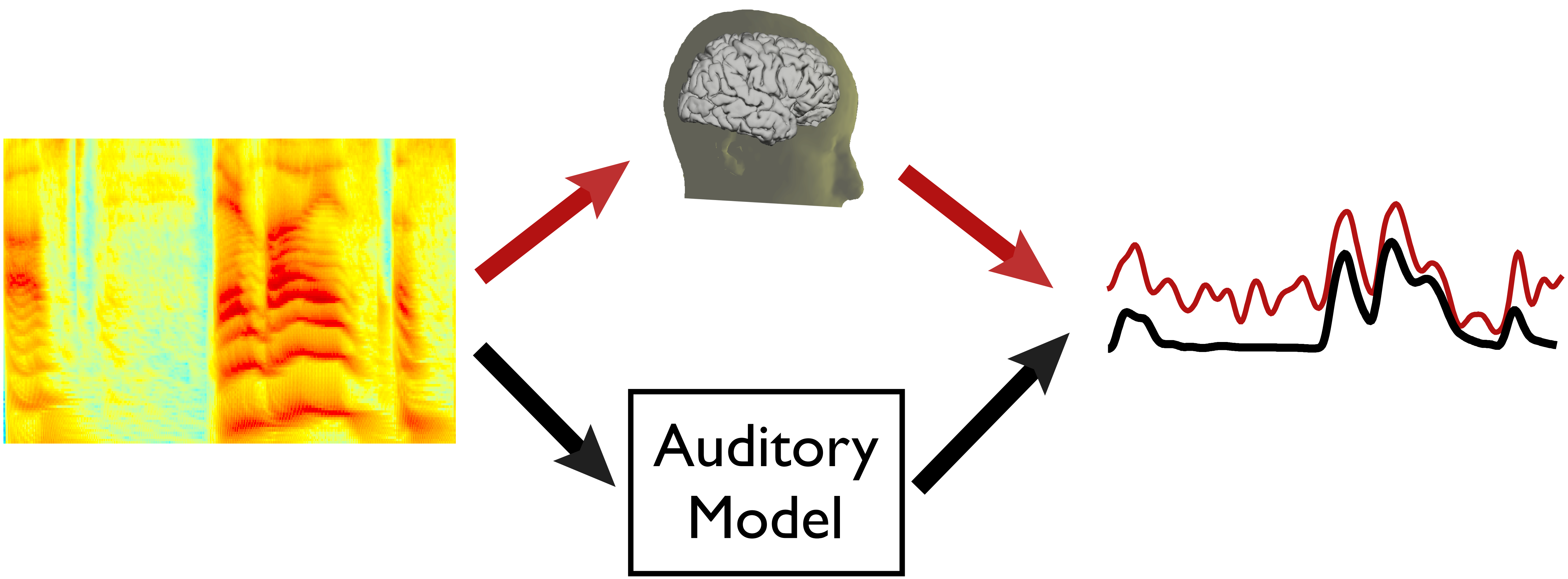
MEG & Auditory Cortex

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

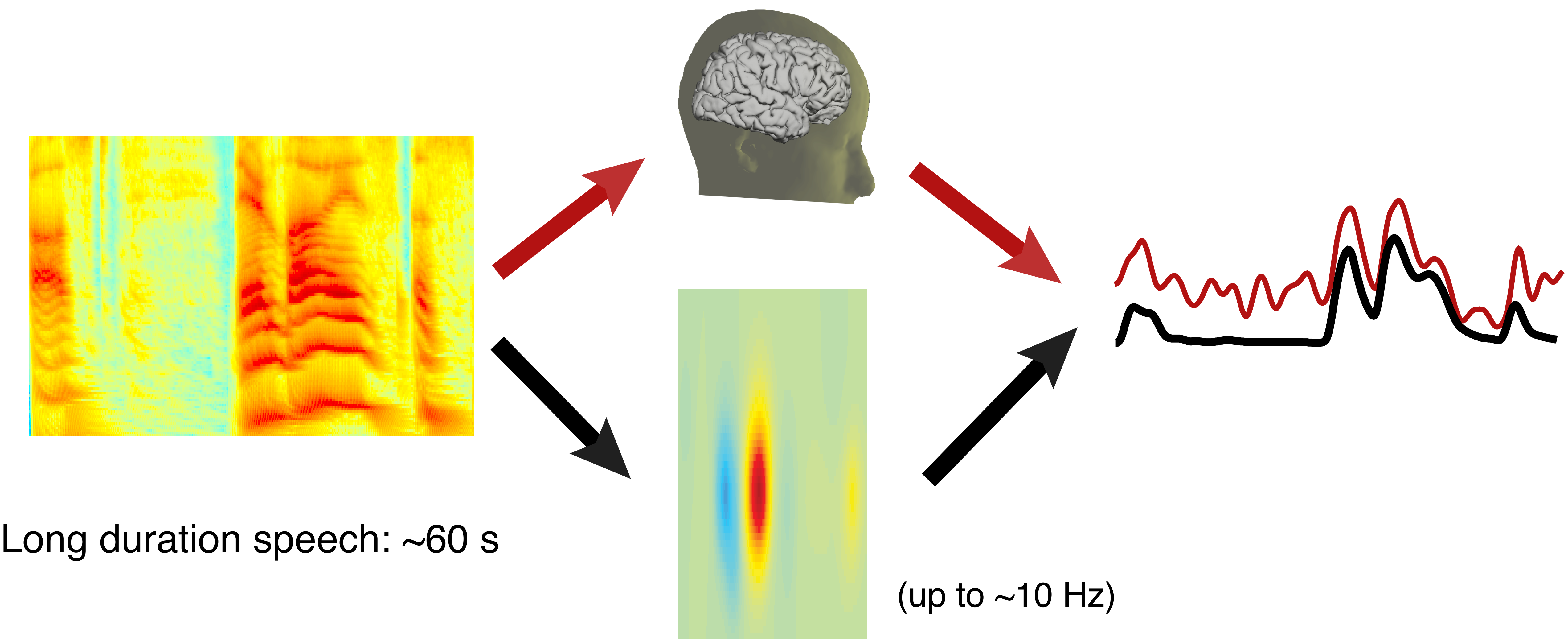




MEG Responses to Speech Modulations



MEG Responses Predicted by STRF Model

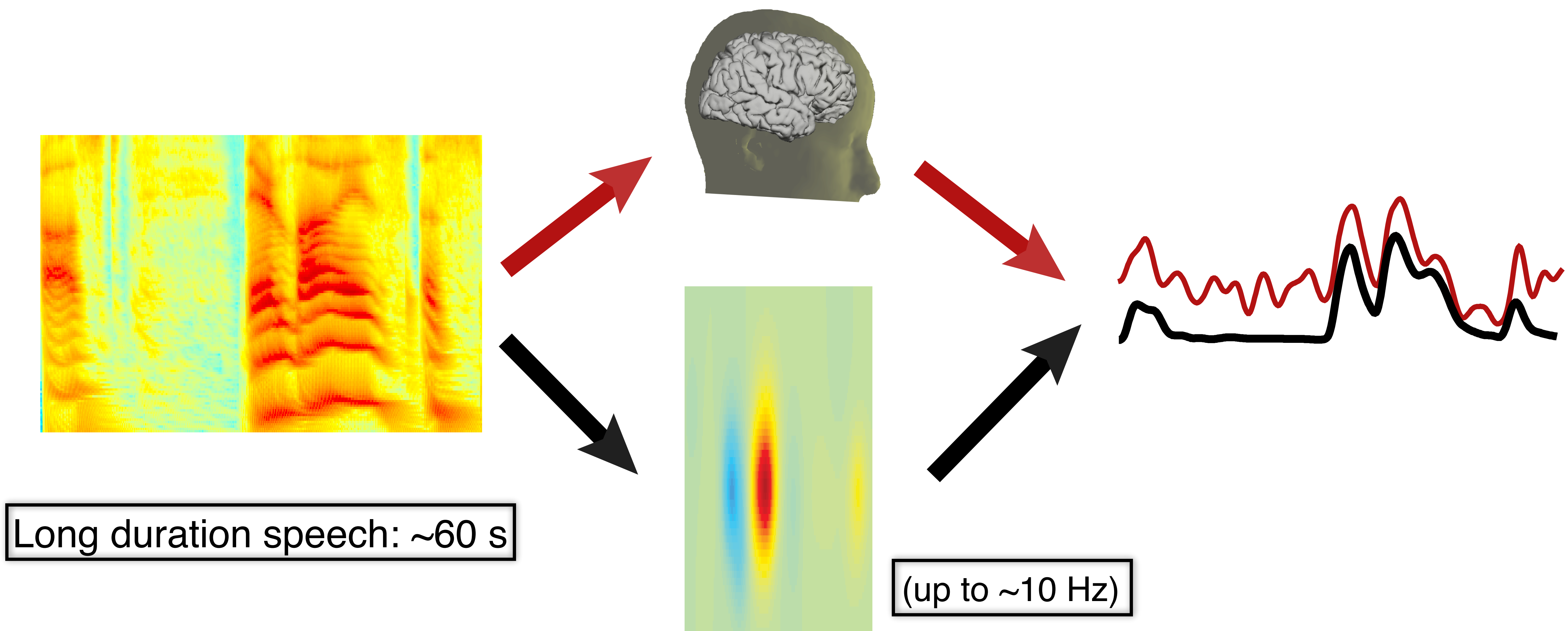


Linear Kernel = STRF

Ding & Simon, J Neurophysiol (2012)

“Spectro-Temporal Response Function”

MEG Responses Predicted by STRF Model

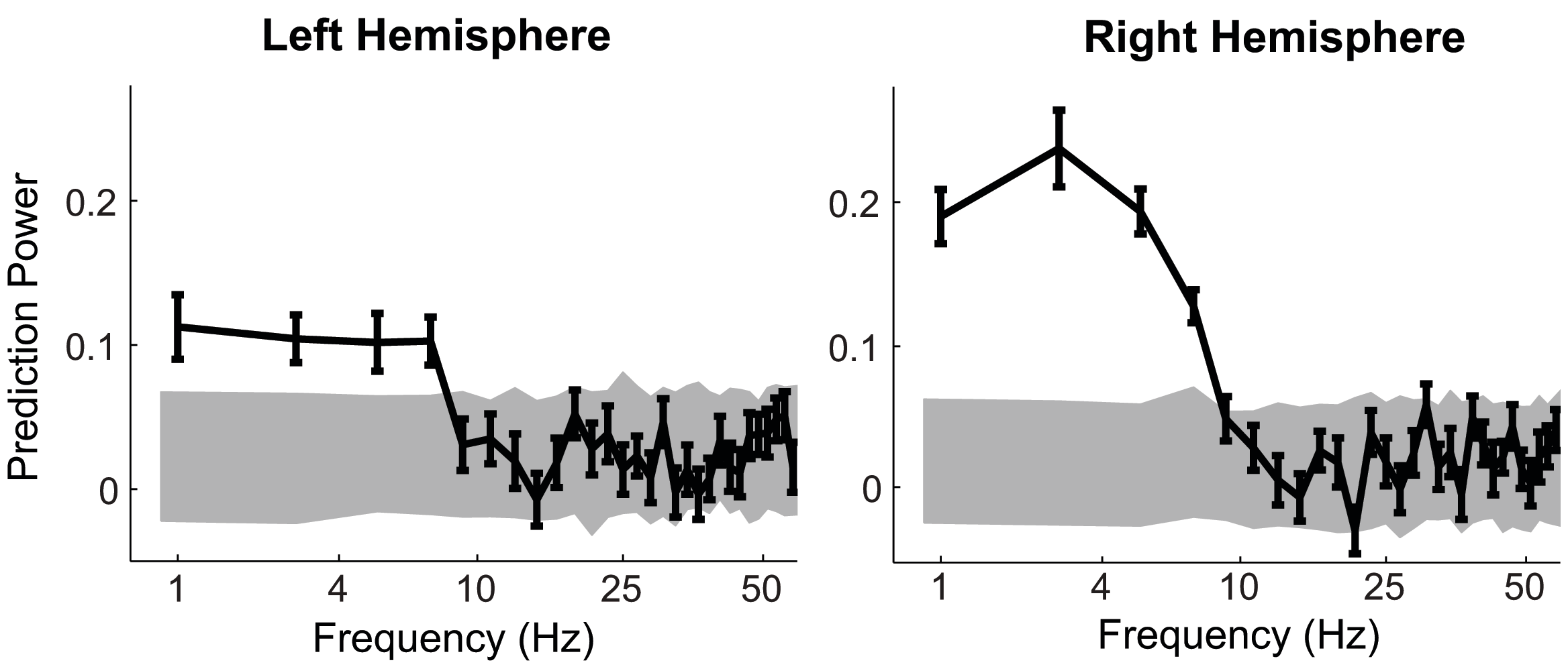


Linear Kernel = STRF

Ding & Simon, J Neurophysiol (2012)

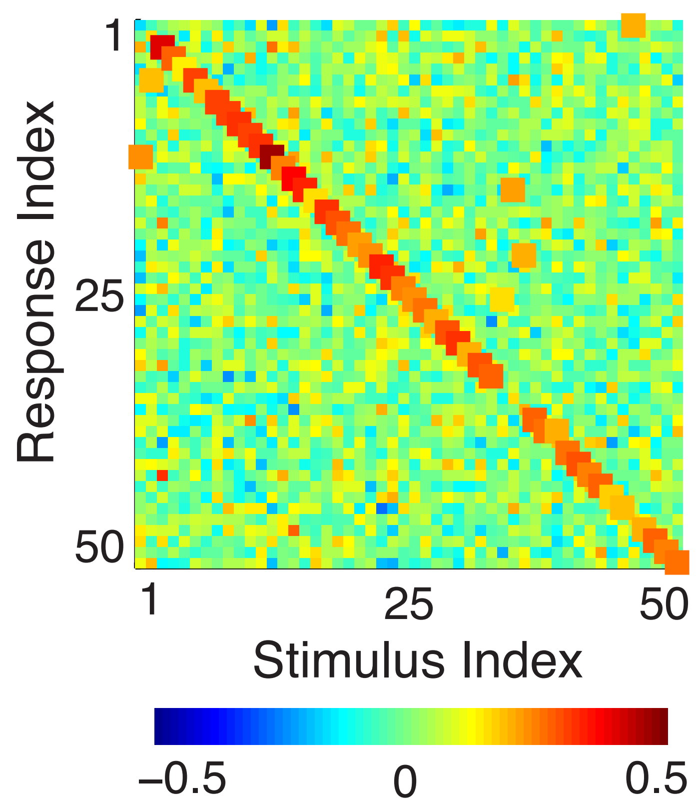
"Spectro-Temporal Response Function"

Frequency Dependence of STRF Predictability

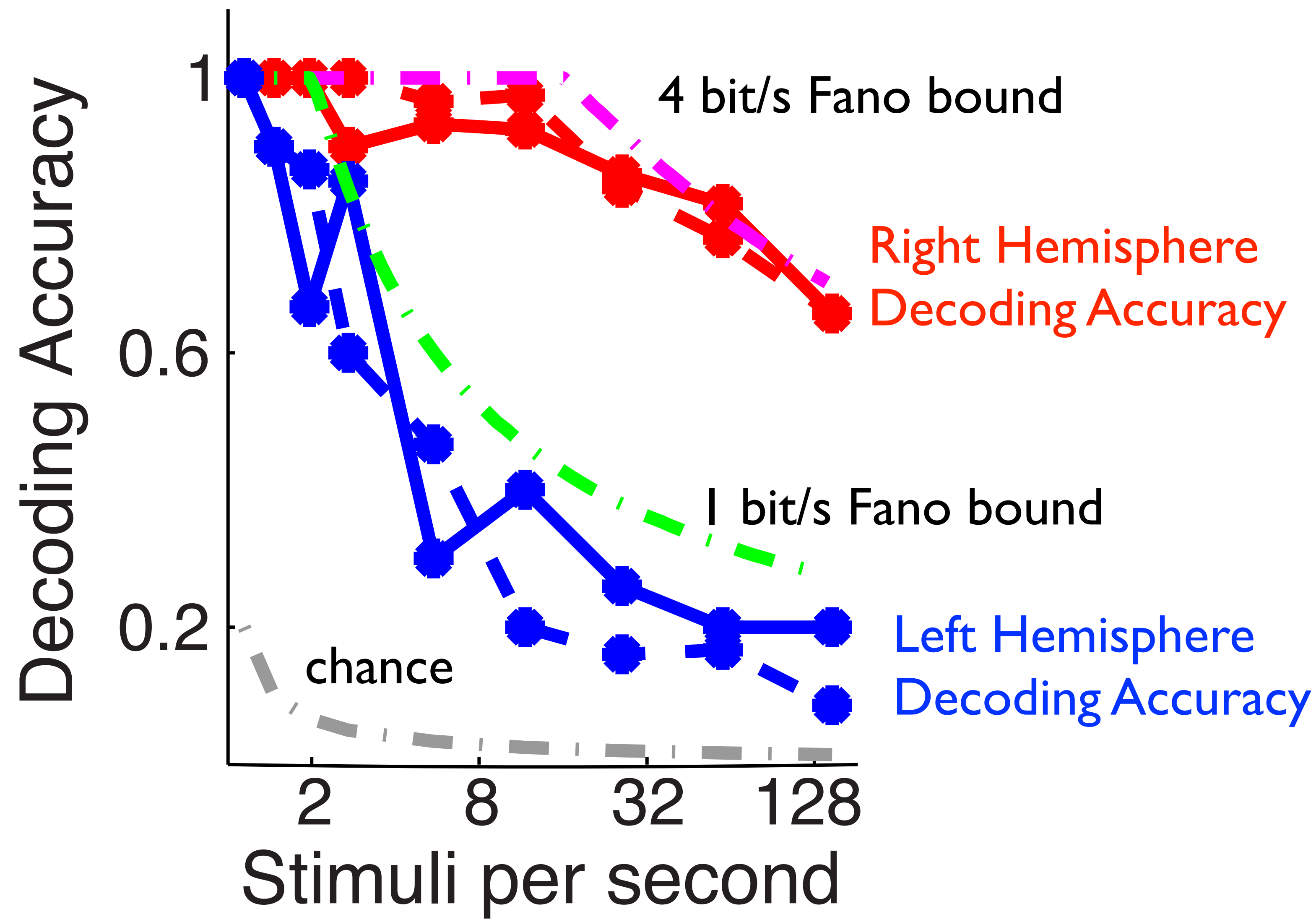


Ding & Simon, J Neurophysiol (2012)

Stimulus Information Encoded in Response

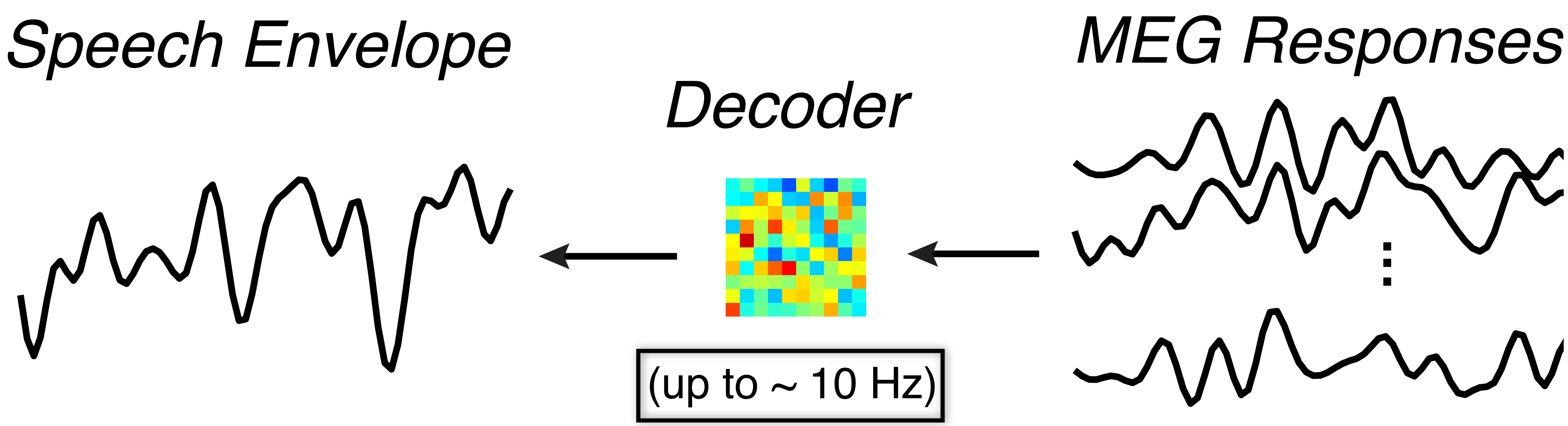


Correlation between stimulus envelope and reconstructed envelope (right hemisphere)

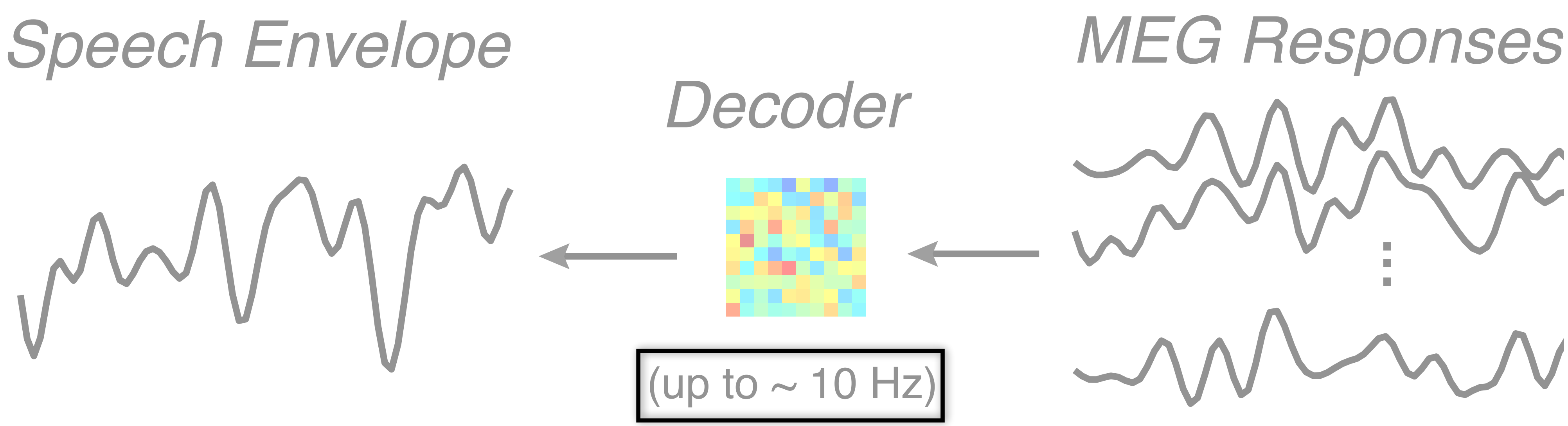


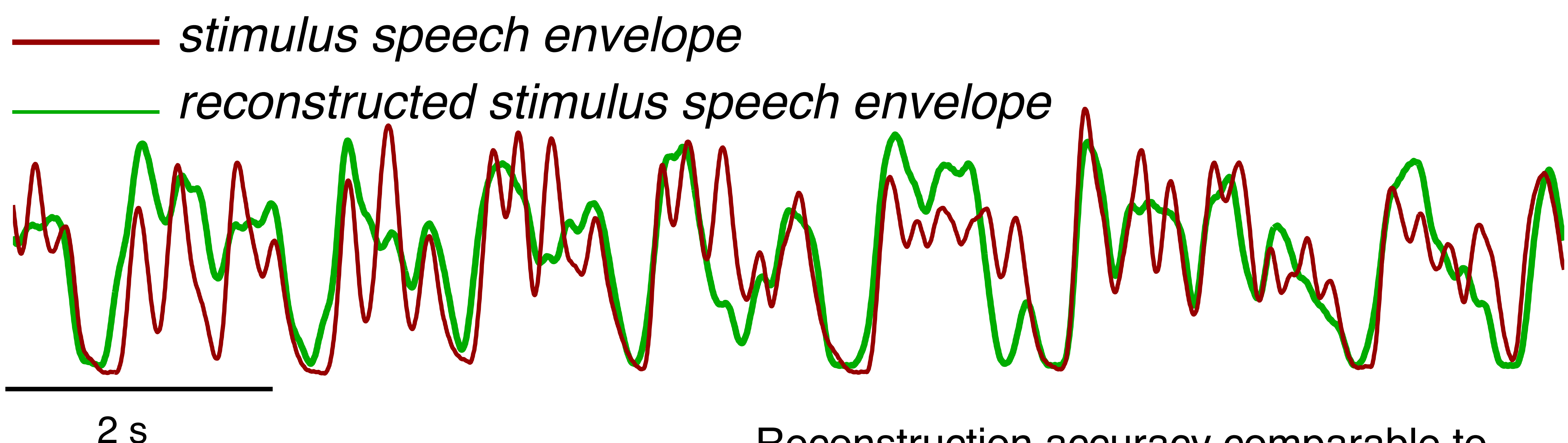
Ding & Simon, J Neurophysiol (2012)

Neural Reconstruction of Speech Envelope



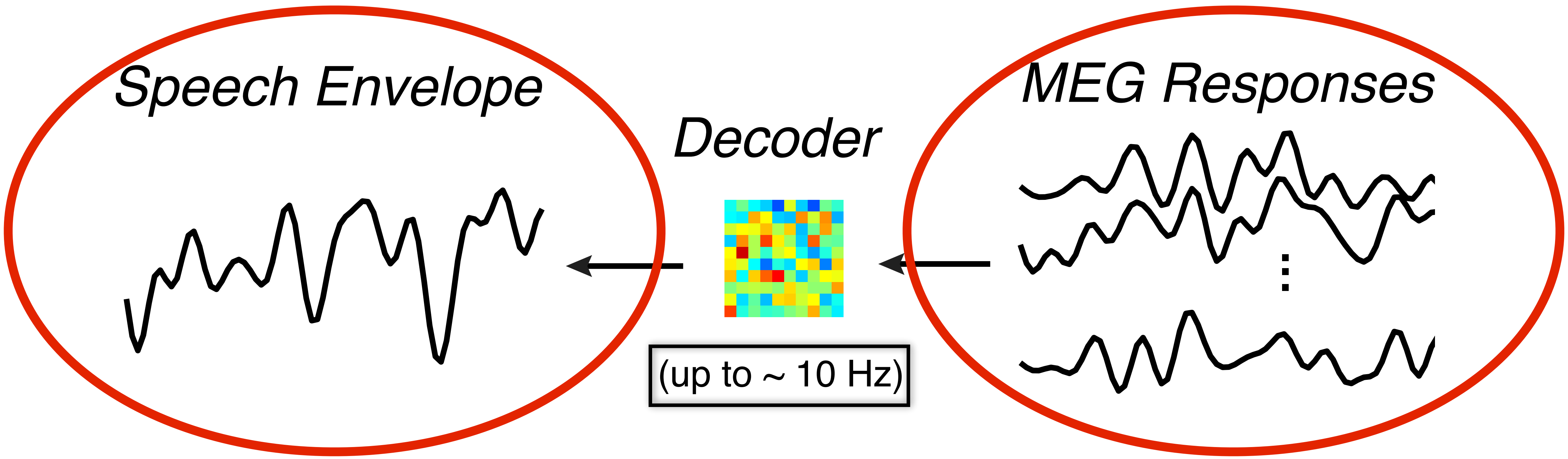
Neural Reconstruction of Speech Envelope



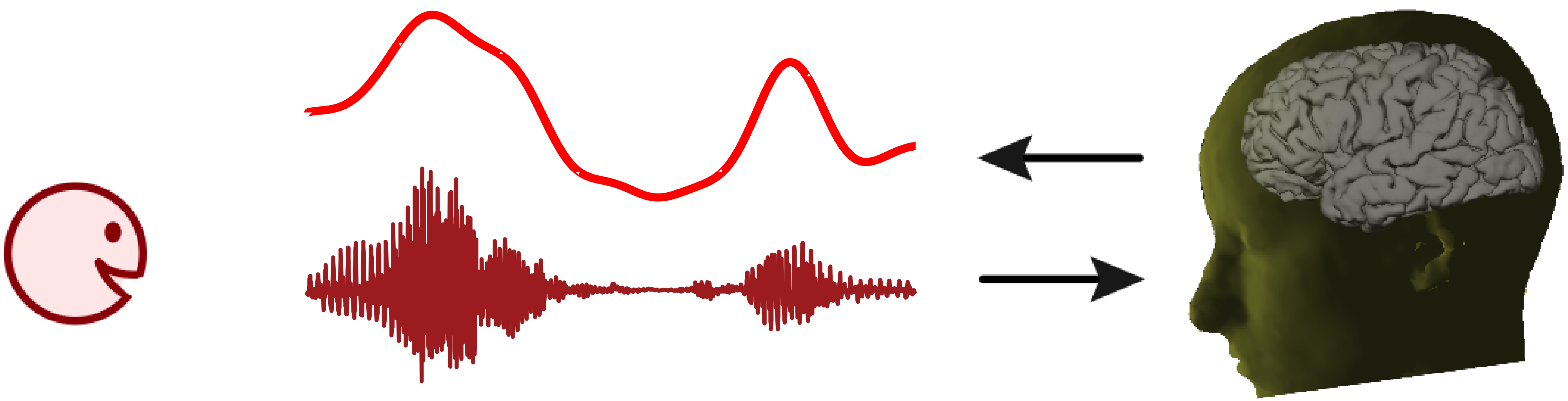


Reconstruction accuracy comparable to single unit & ECoG recordings

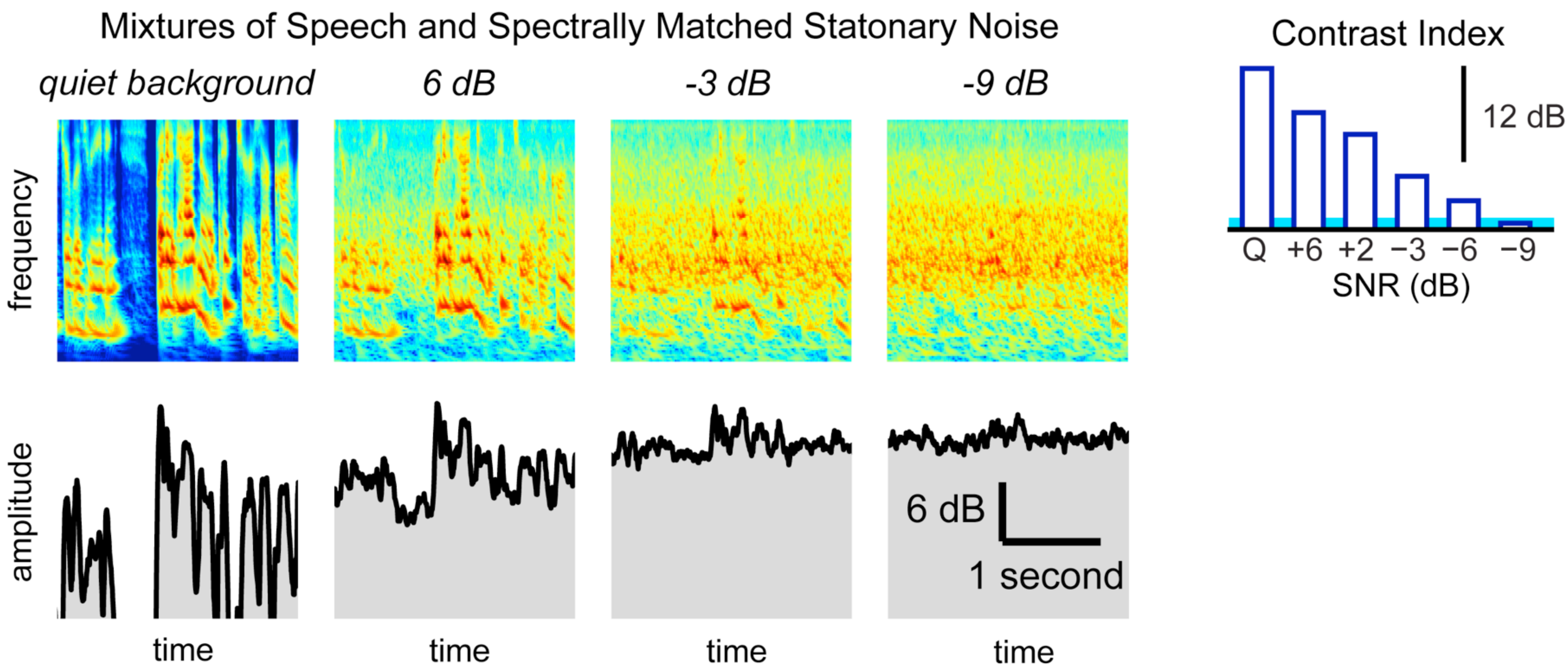
Ding & Simon, J Neurophysiol (2012)
Zion-Golumbic et al., Neuron (2013)


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

Neural Representation of Speech: Temporal

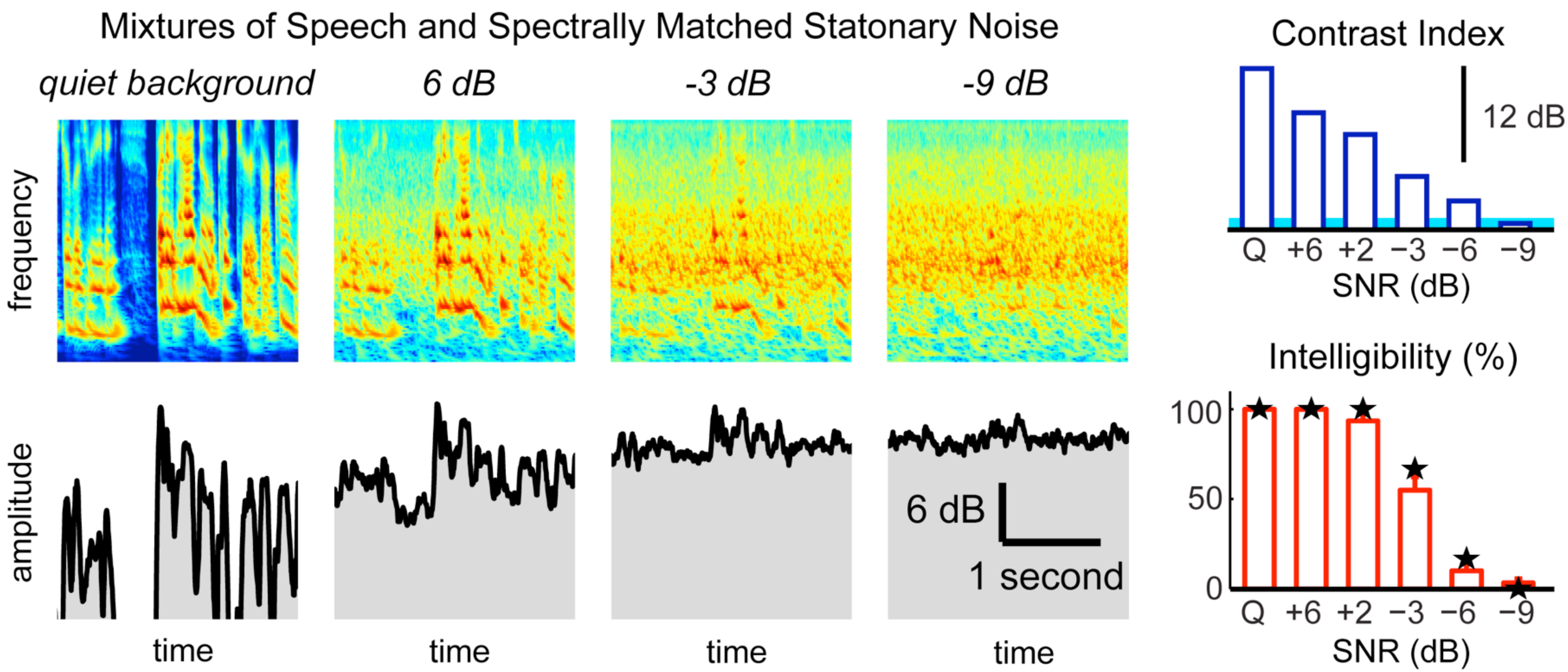


Speech in Stationary Noise



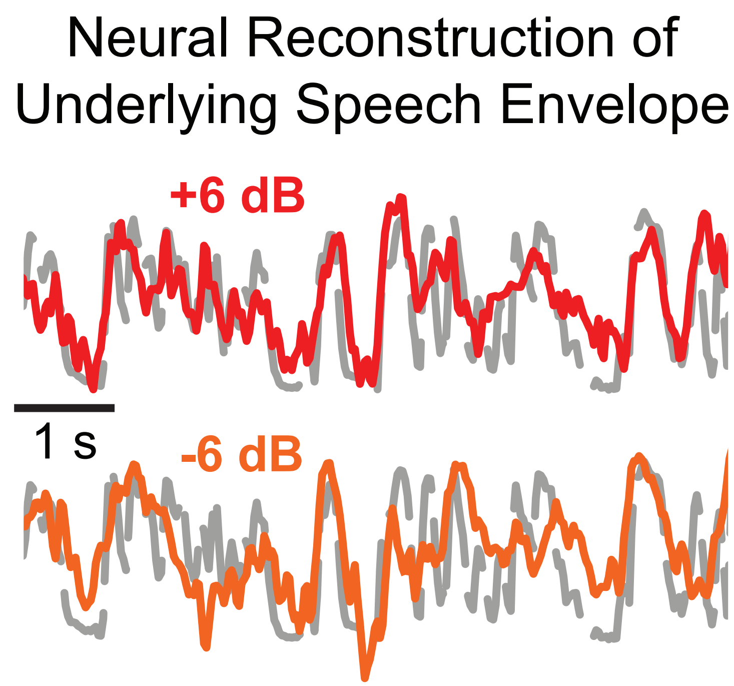
Ding & Simon, J Neuroscience (2013)

Speech in Stationary Noise



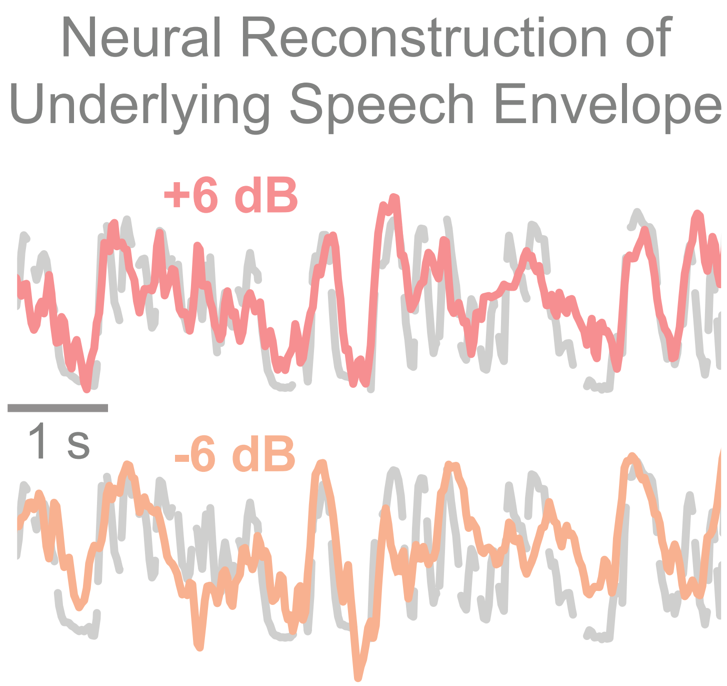
Ding & Simon, J Neuroscience (2013)

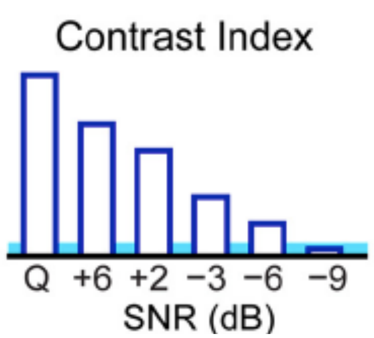
Speech in Noise: Results

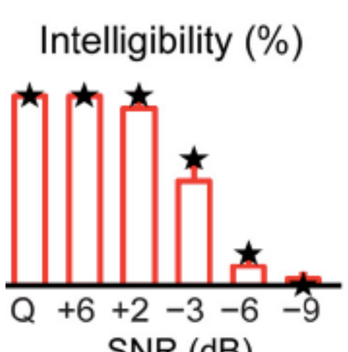


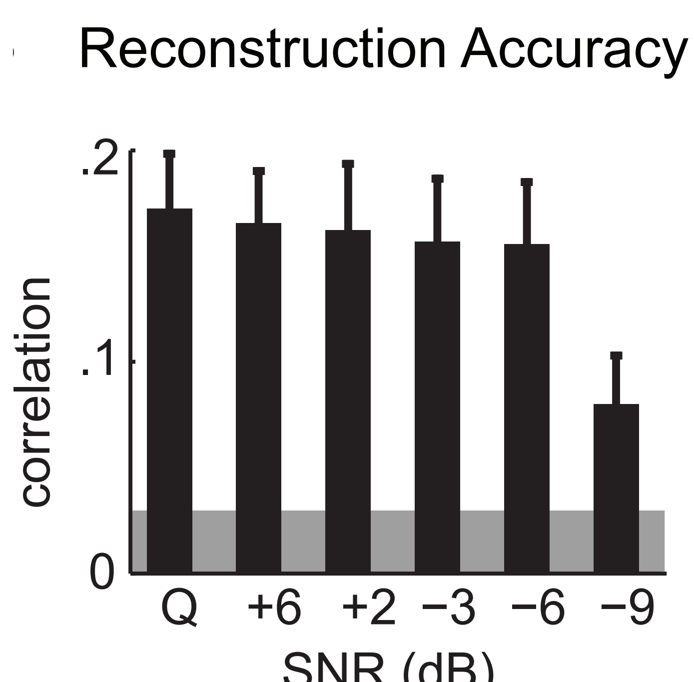
Ding & Simon, J Neuroscience (2013)

Speech in Noise: Results



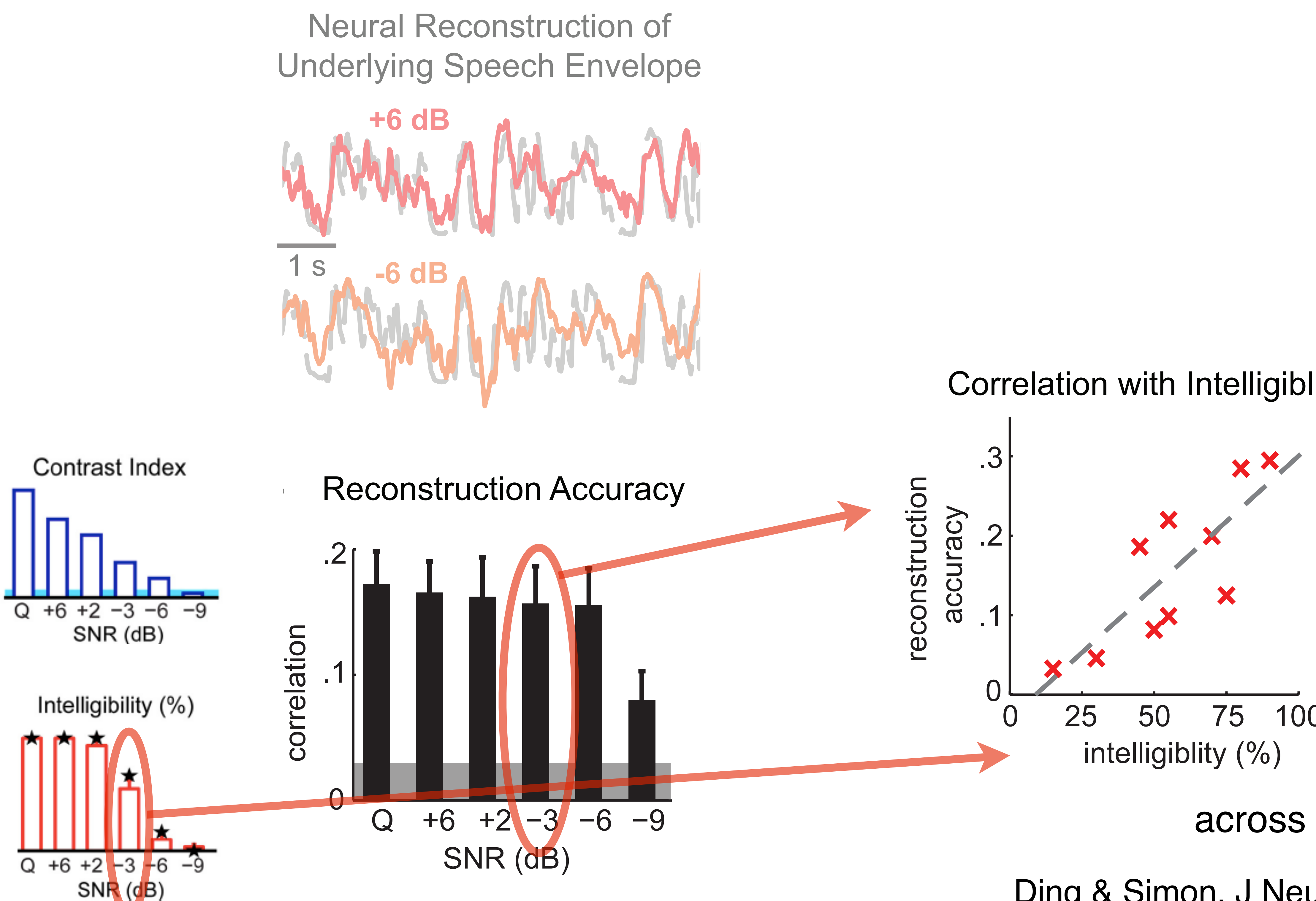






Ding & Simon, J Neuroscience (2013)

Speech in Noise: Results



across Subjects

Ding & Simon, J Neuroscience (2013)

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 Robustness of Speech Representation (Delta frequency band)

Outline

- Cortical Representations of Speech (via MEG)
  - Encoding vs. Decoding
- “Cocktail Party” Speech
- Recent Results
  - Attentional Dynamics
  - “Restoration” of Missing Speech
  - Speech Processing Across the Brain

Outline

- Cortical Representations of Speech (via MEG)
  - Encoding vs. Decoding
- “Cocktail Party” Speech
- Recent Results
  - Attentional Dynamics
  - “Restoration” of Missing Speech
  - Speech Processing Across the Brain

Listening to Speech at the Cocktail Party

Alex Katz,
The Cocktail Party

Listening to Speech at the Cocktail Party

Alex Katz,
The Cocktail Party

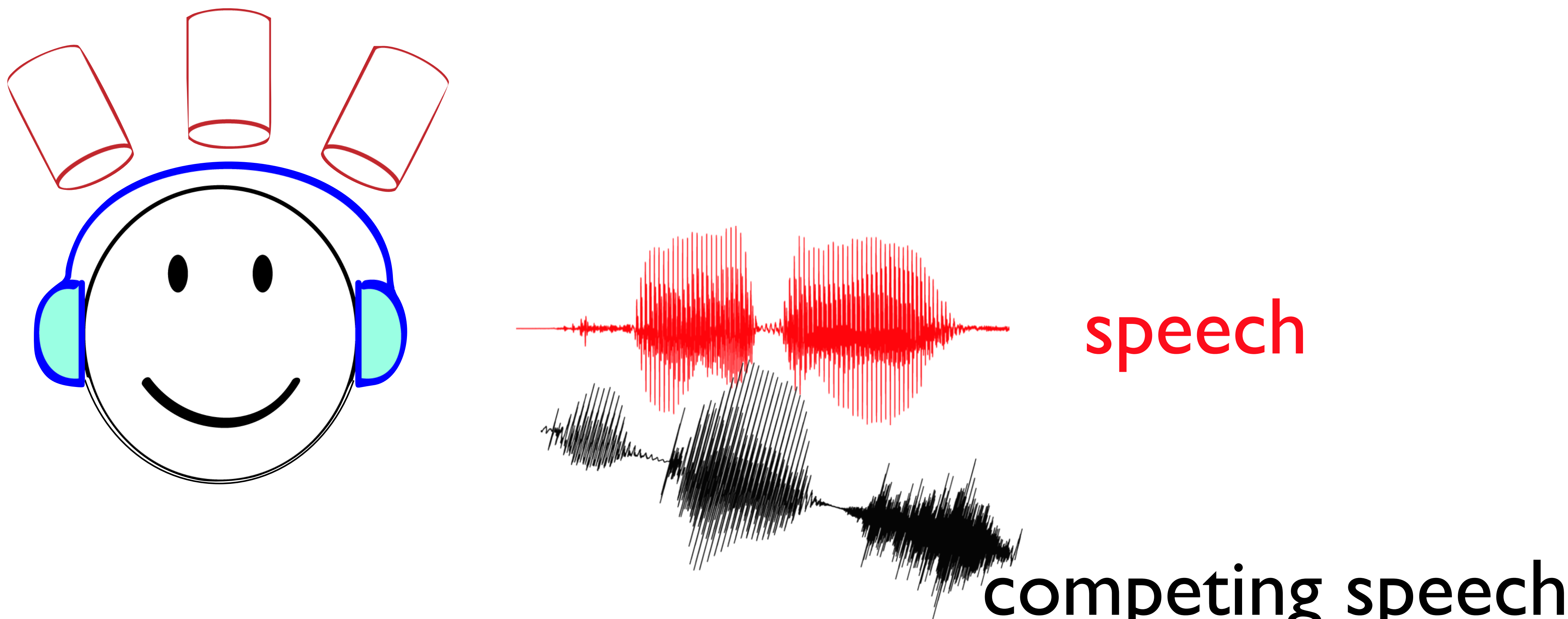
Listening to Speech at the Cocktail Party

Alex Katz,
The Cocktail Party

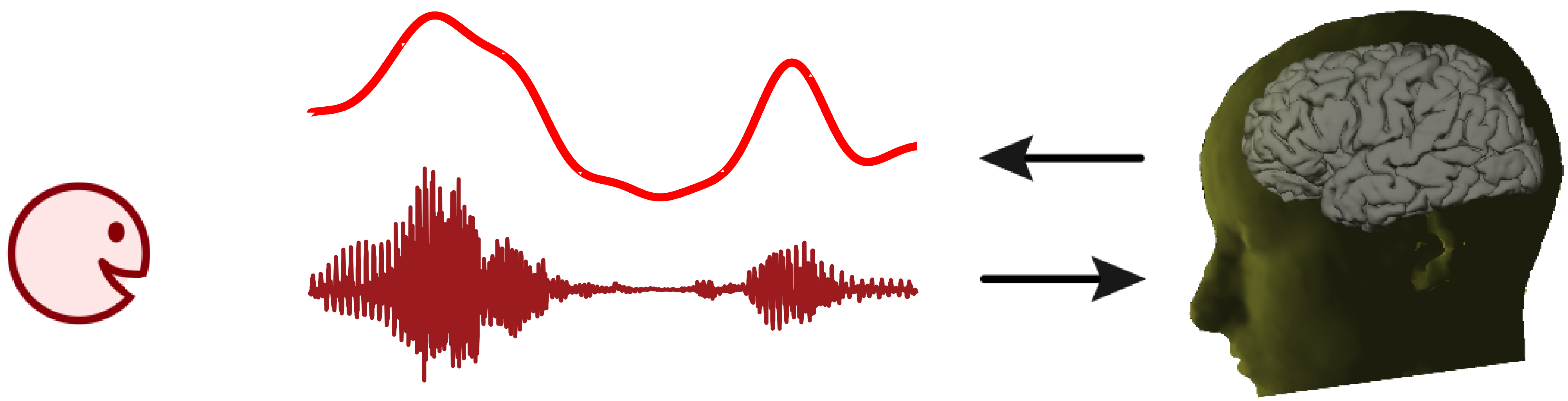
Listening to Speech at the Cocktail Party

Alex Katz,
The Cocktail Party

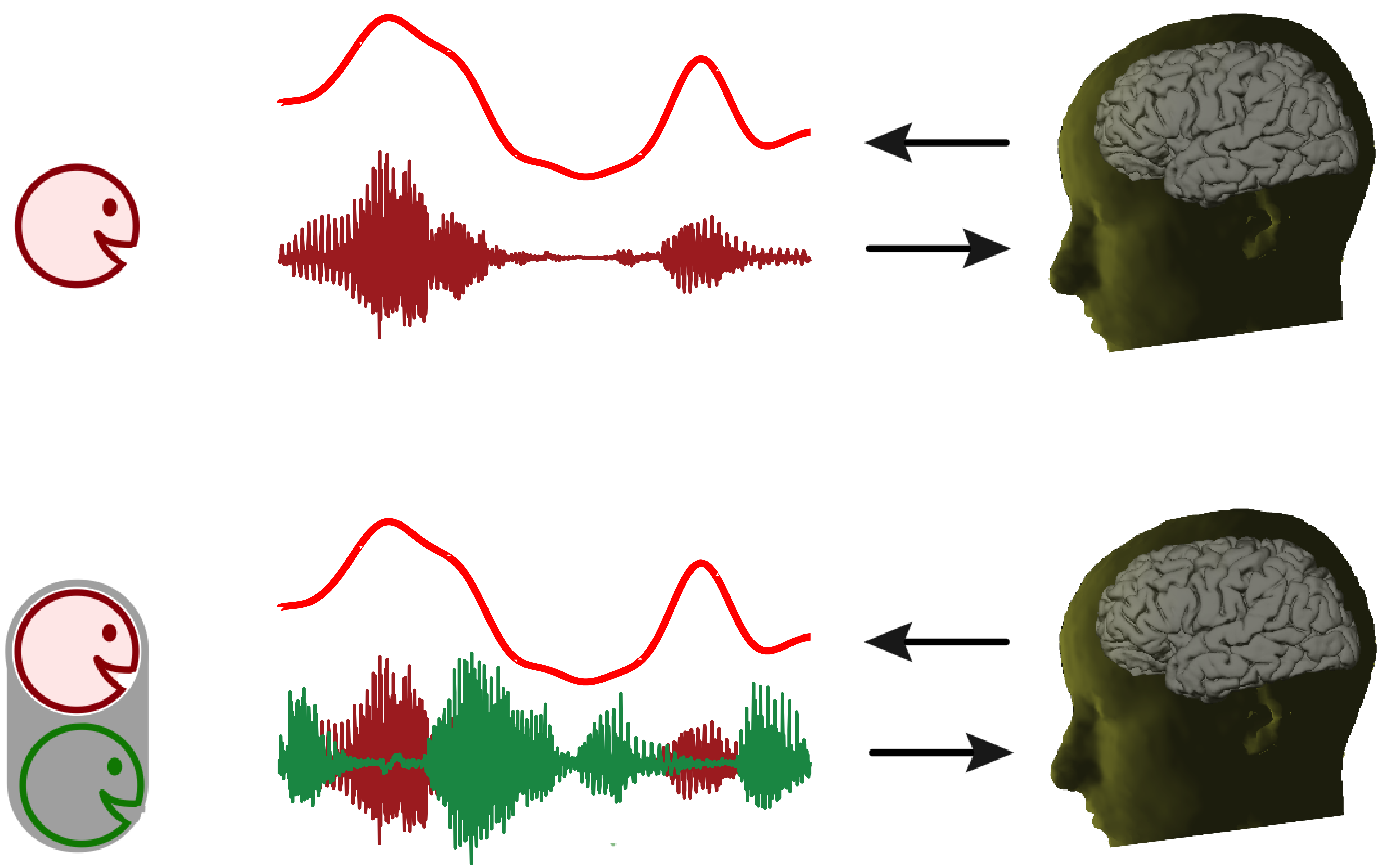
Competing Speech Streams



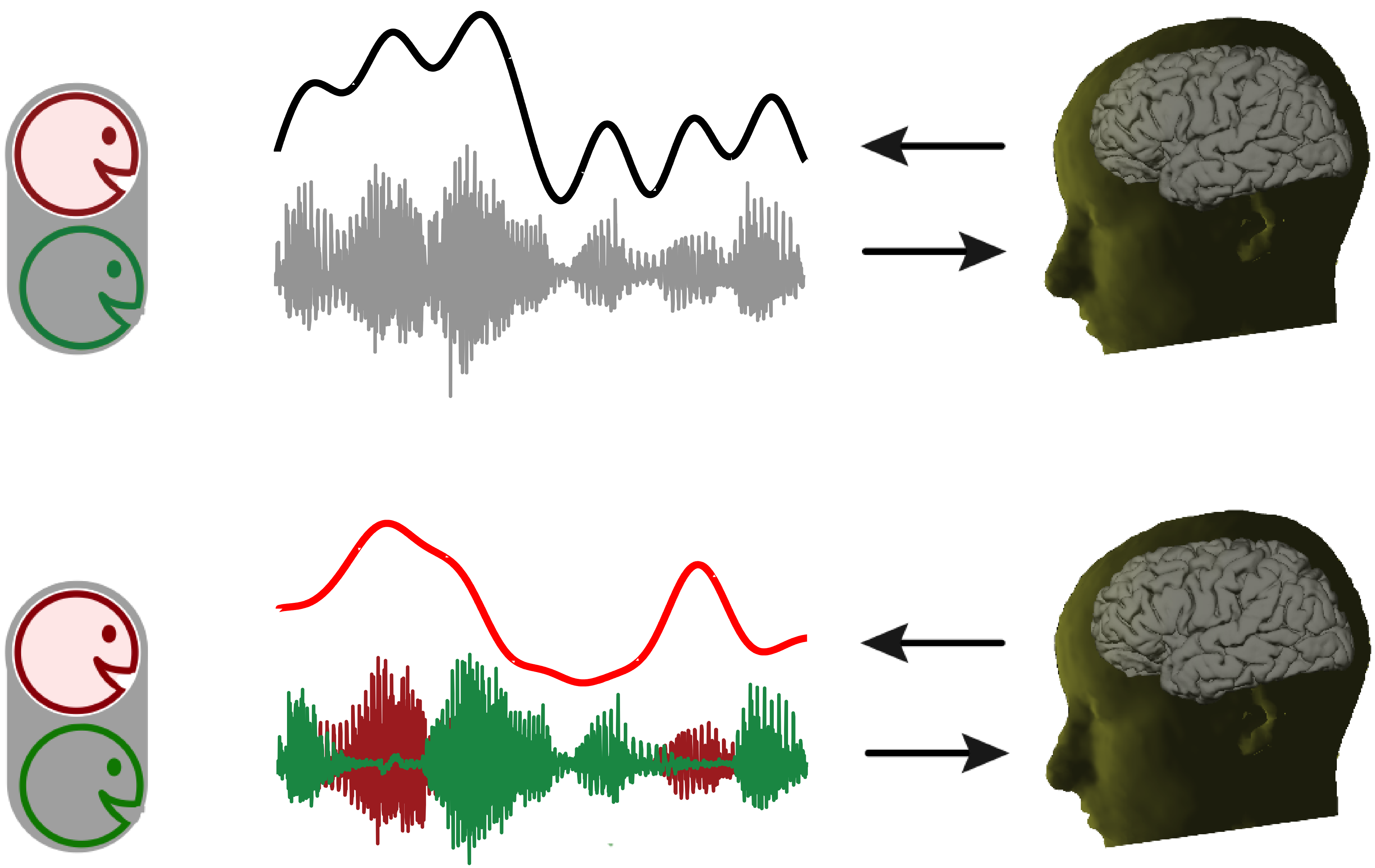
Selective Neural Encoding



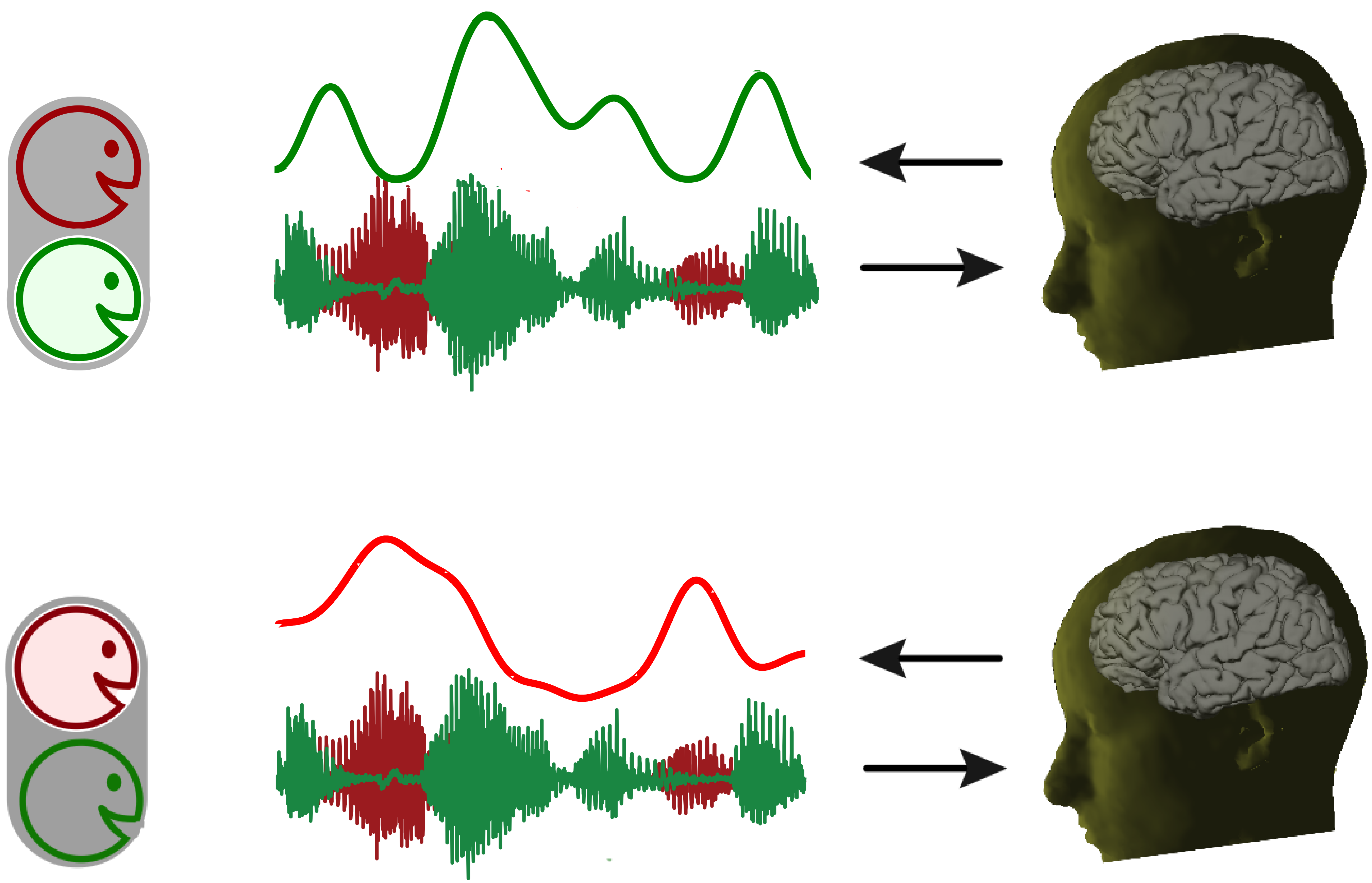
Selective Neural Encoding



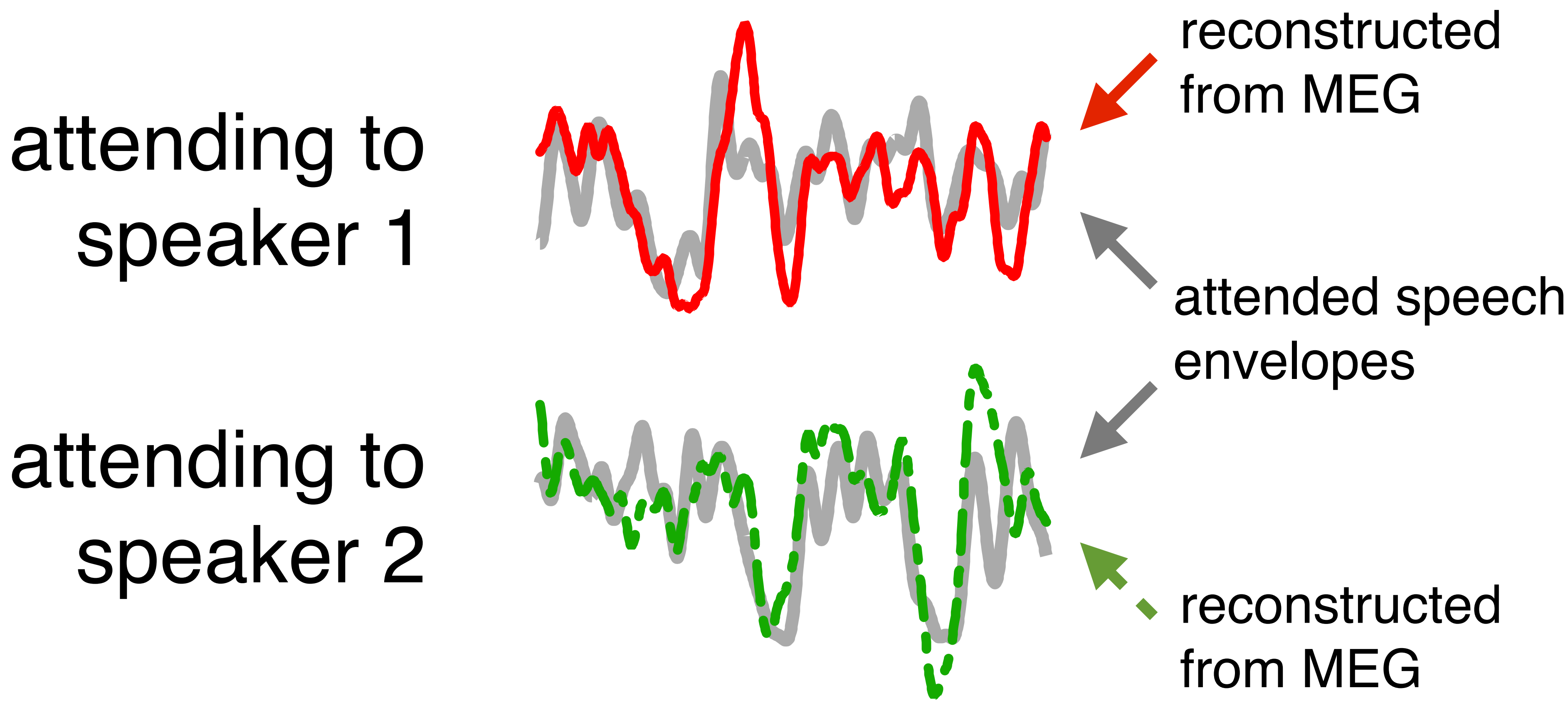
Unselective vs. Selective Neural Encoding



Selective Neural Encoding



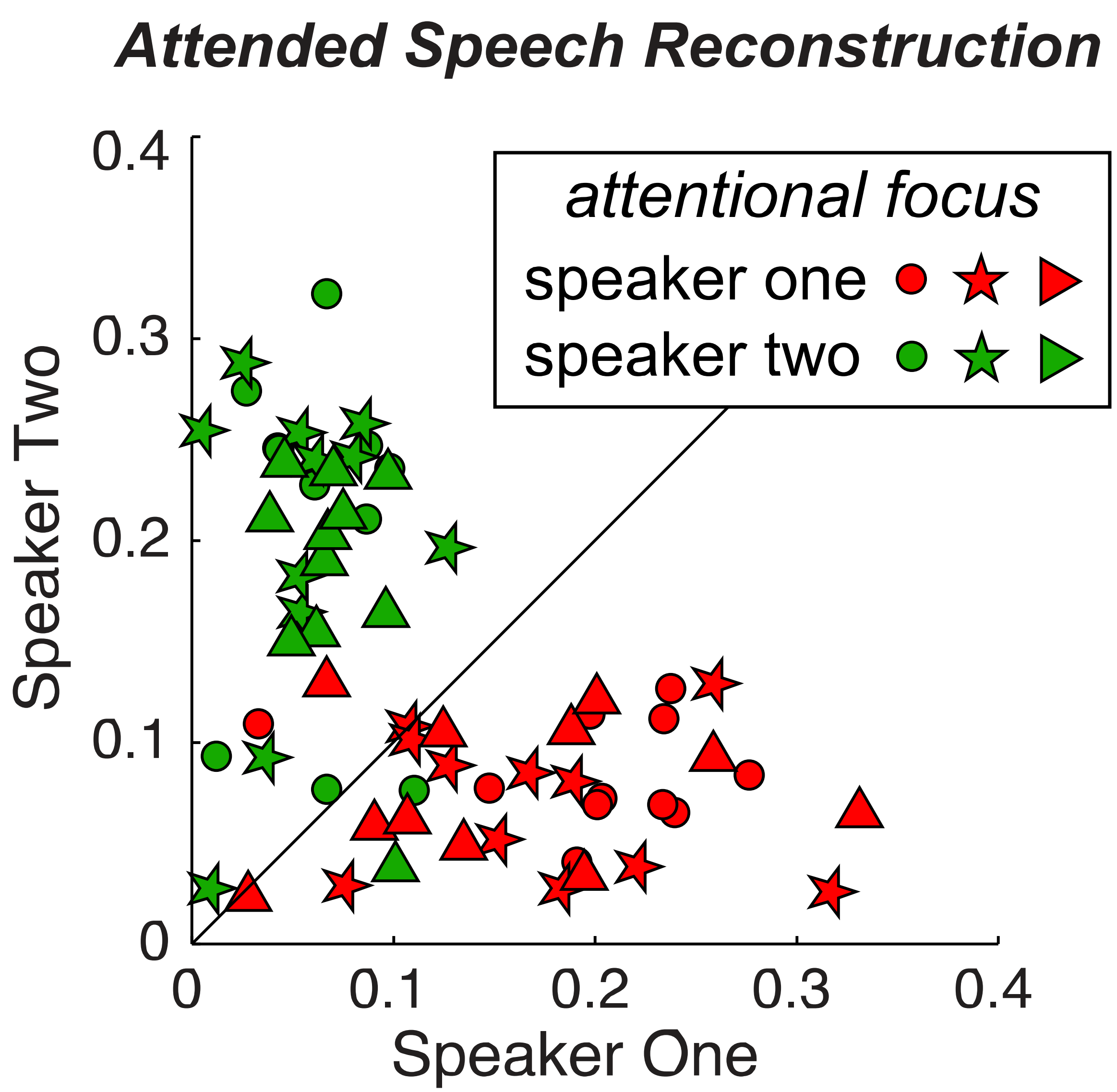
Selective Encoding: Results



Identical Stimuli!

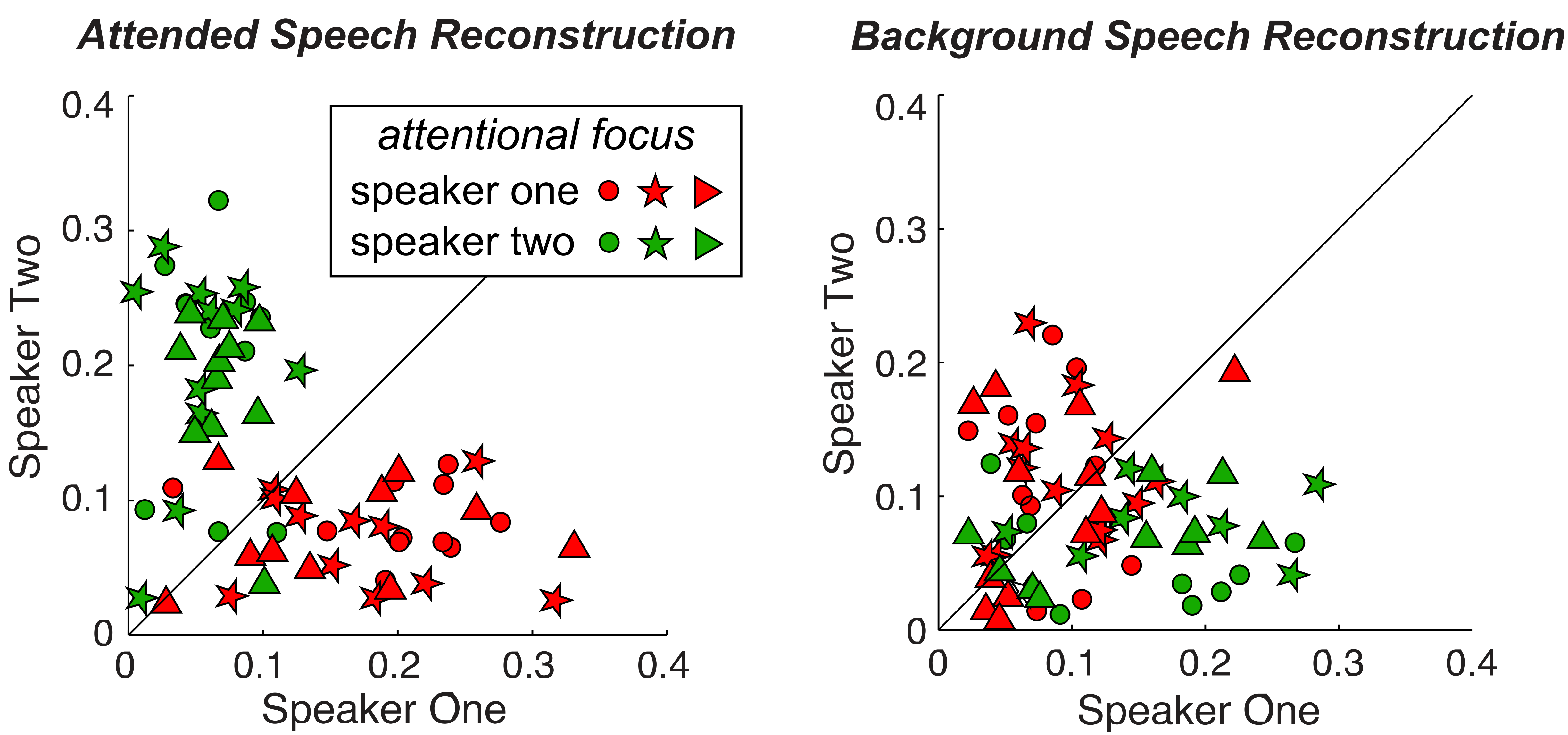
Ding & Simon, PNAS (2012)

Single Trial Speech Reconstruction



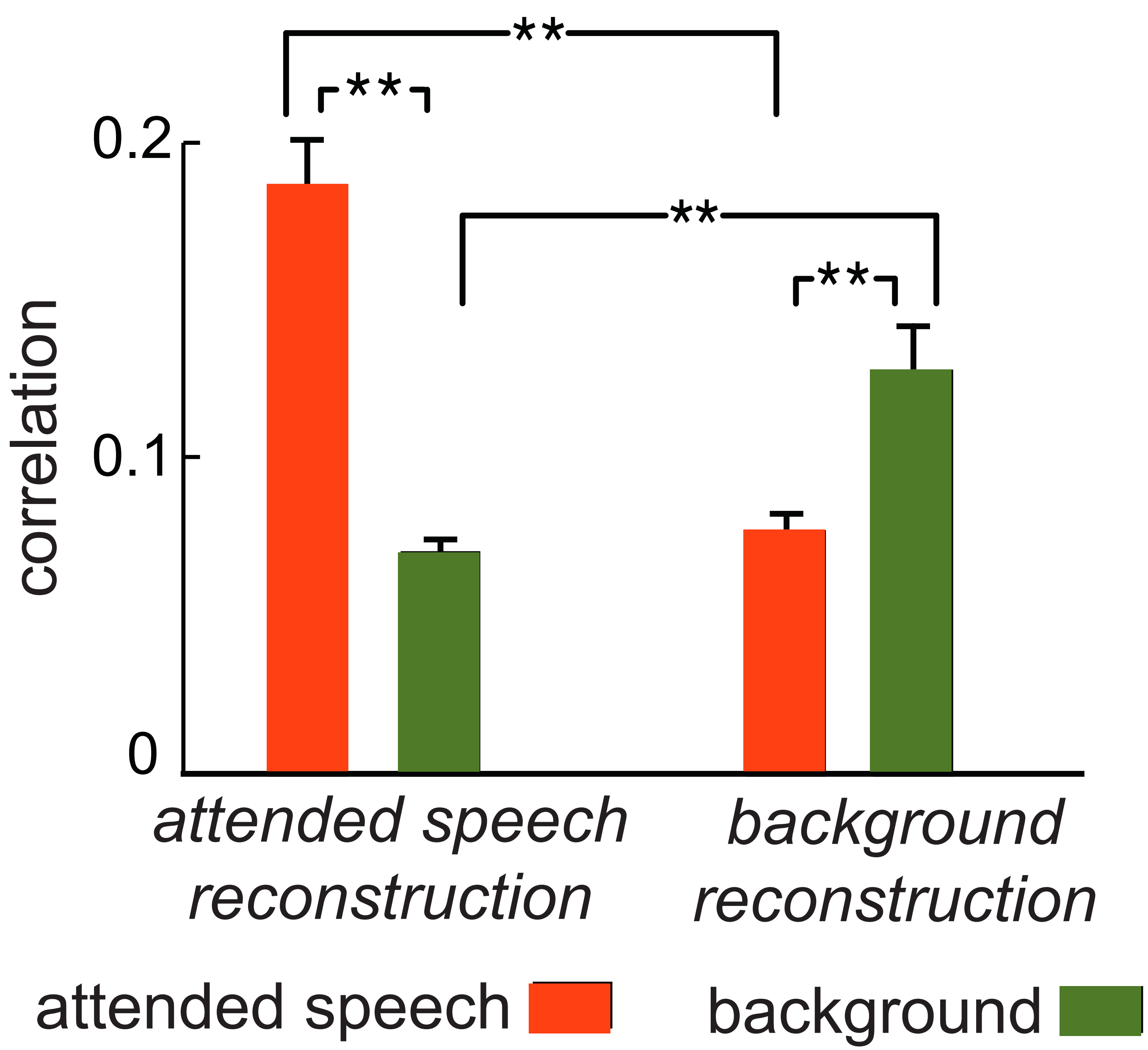
Ding & Simon, PNAS (2012)

Single Trial Speech Reconstruction



Ding & Simon, PNAS (2012)

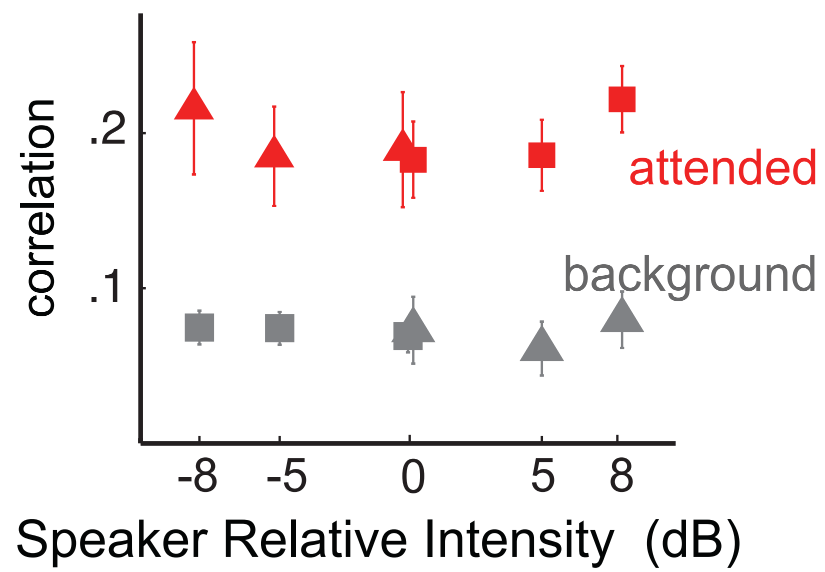
Overall Speech Reconstruction



Distinct neural representations for different speech streams

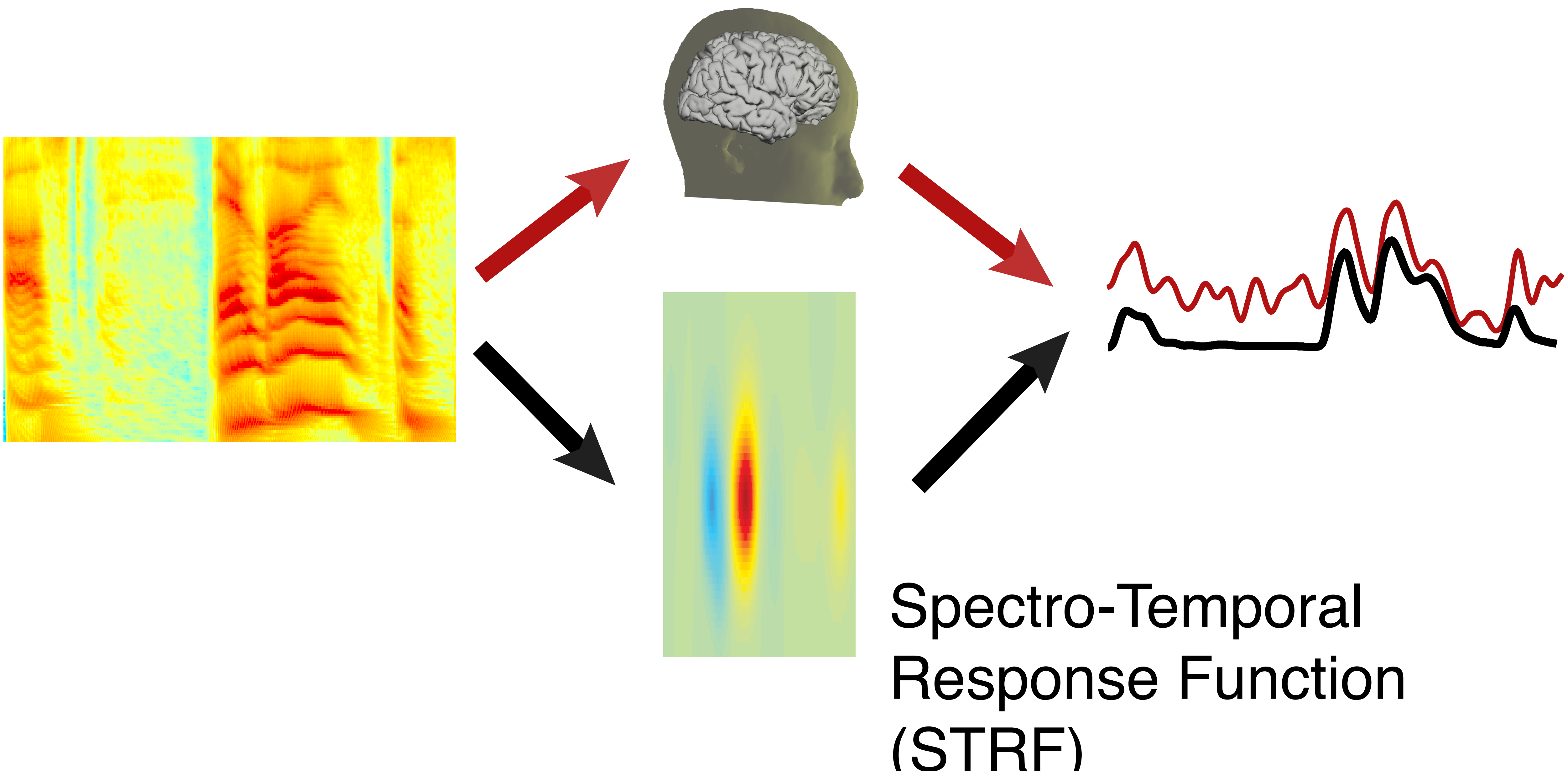
Invariance under Relative Loudness Change

Neural Results

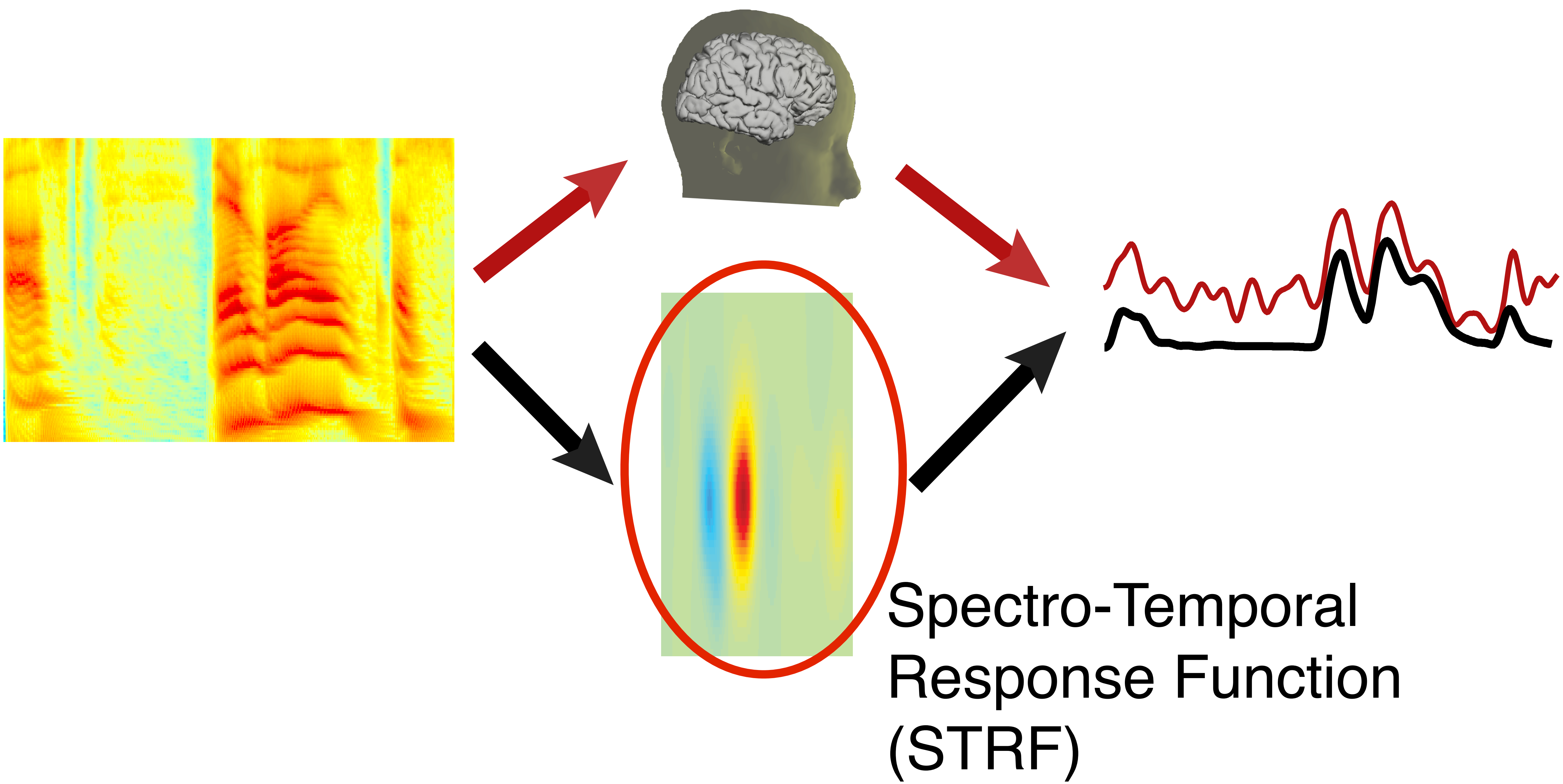


- Neural representation invariant to relative loudness change
- Stream-based Gain Control, not stimulus-based

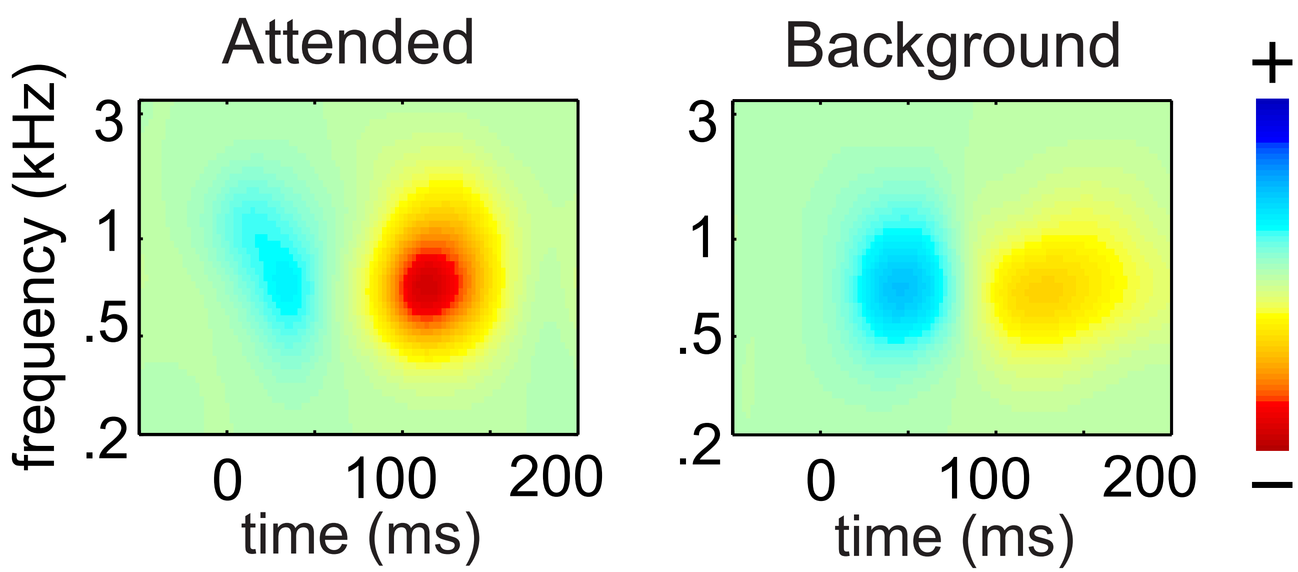
Forward STRF Model



Forward STRF Model

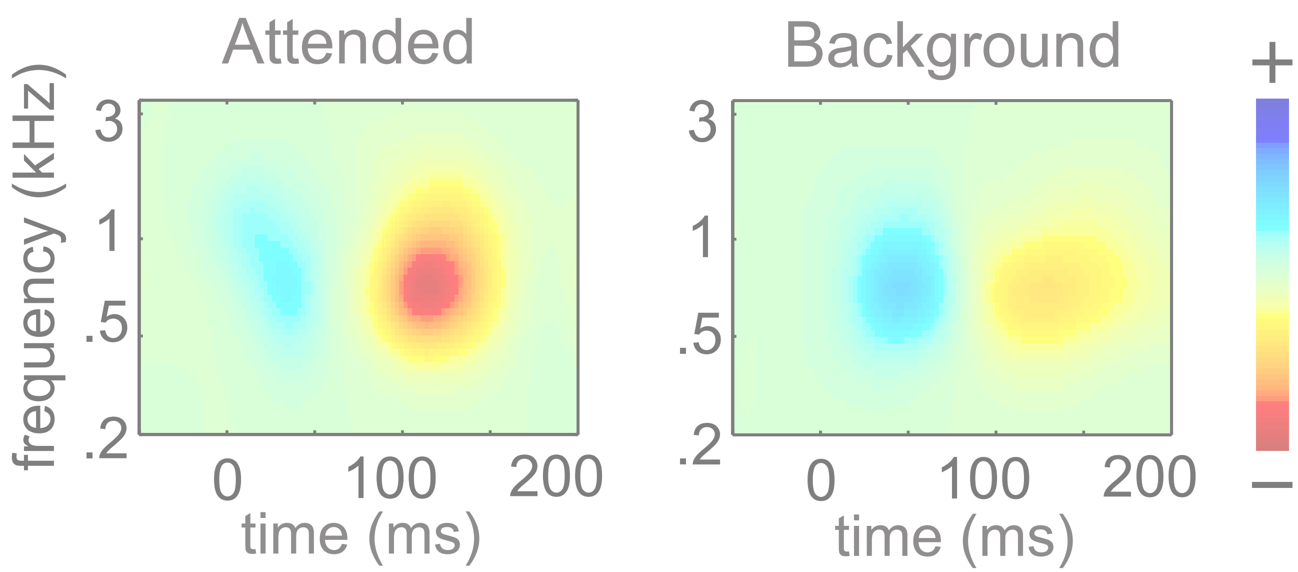


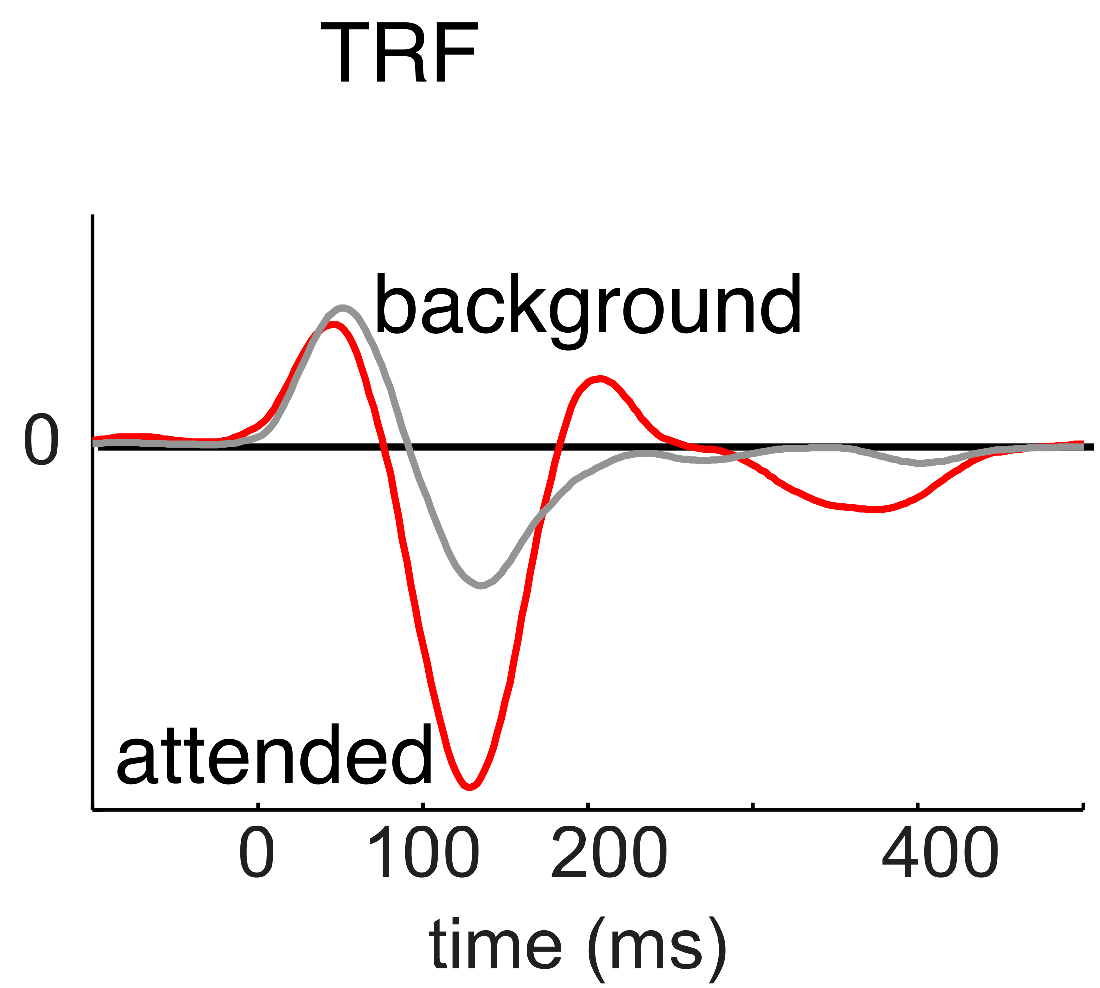
STRF Results



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

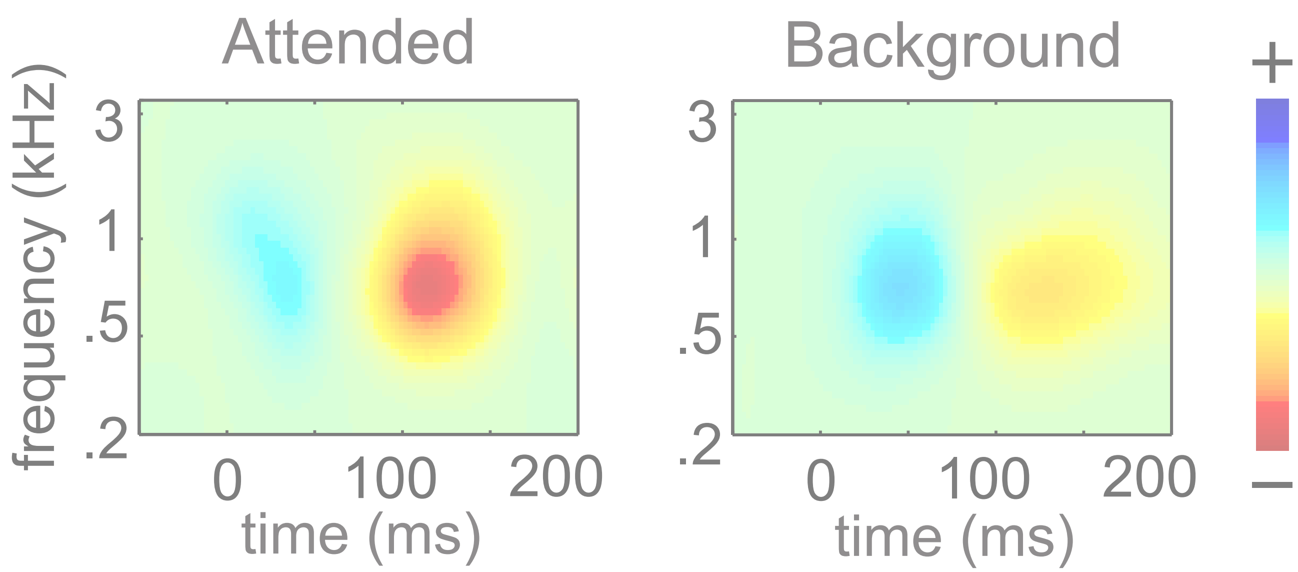
STRF Results

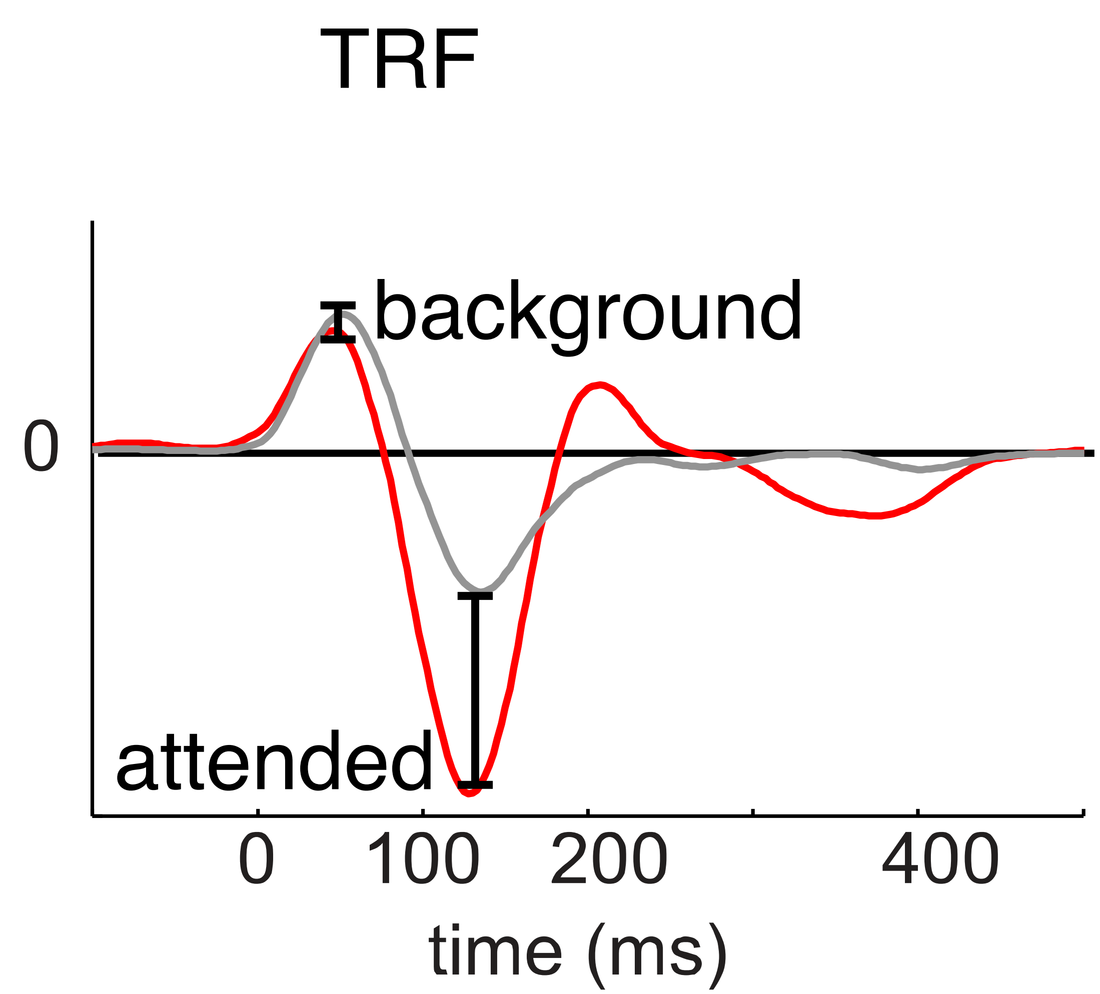




- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- $M50_{STRF}$ positive peak
- $M100_{STRF}$ negative peak

STRF Results





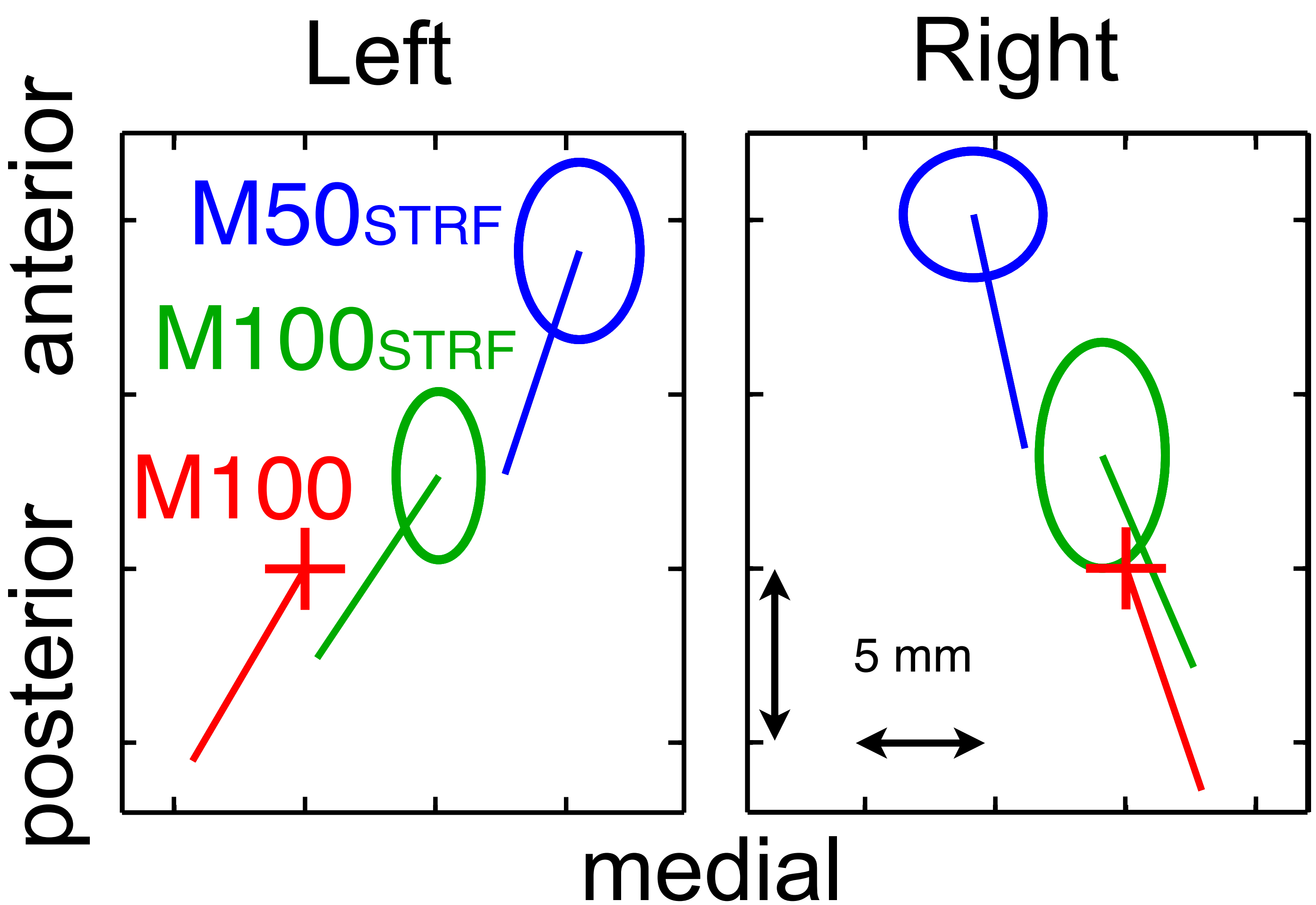
- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- $M50_{STRF}$ positive peak
- $M100_{STRF}$ negative peak
- **$M100_{STRF}$ strongly modulated by attention, *but not $M50_{STRF}$***

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***



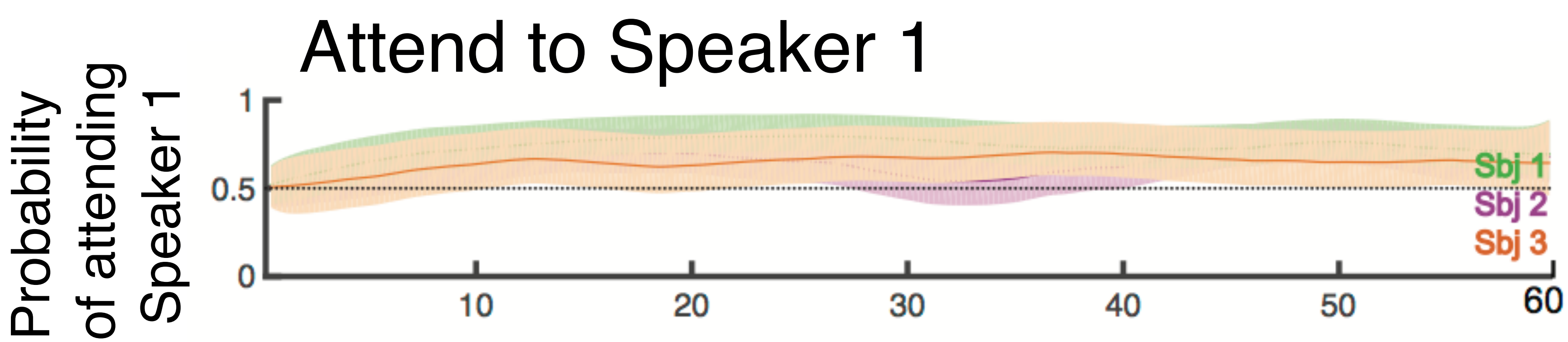
Outline

- Cortical Representations of Speech (via MEG)
  - Encoding vs. Decoding
- “Cocktail Party” Speech
- Recent Results
  - Attentional Dynamics
  - “Restoration” of Missing Speech
  - Speech Processing Across the Brain

Outline

- Cortical Representations of Speech (via MEG)
  - Encoding vs. Decoding
- “Cocktail Party” Speech
- Recent Results
  - Attentional Dynamics
  - “Restoration” of Missing Speech
  - Speech Processing Across the Brain

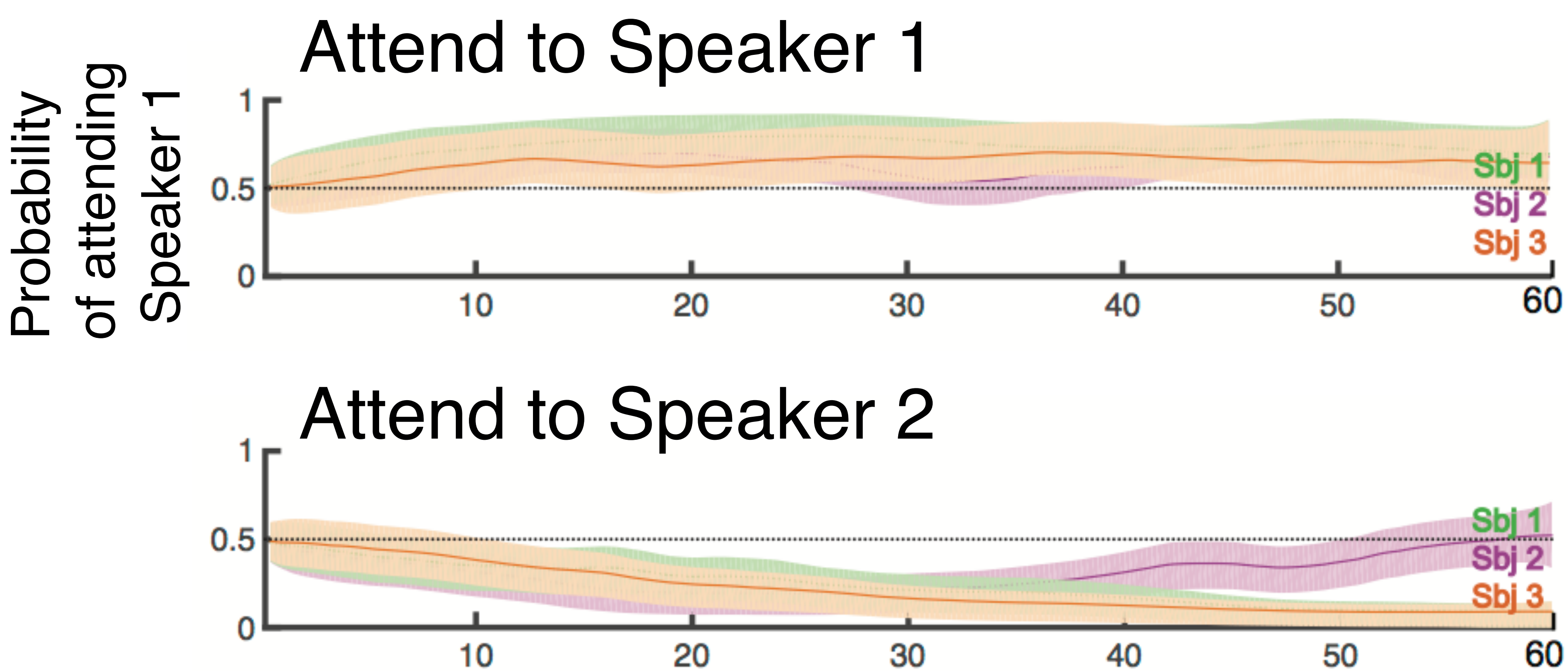
Attentional Dynamics



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

Akram et al. NeuroImage (2016)

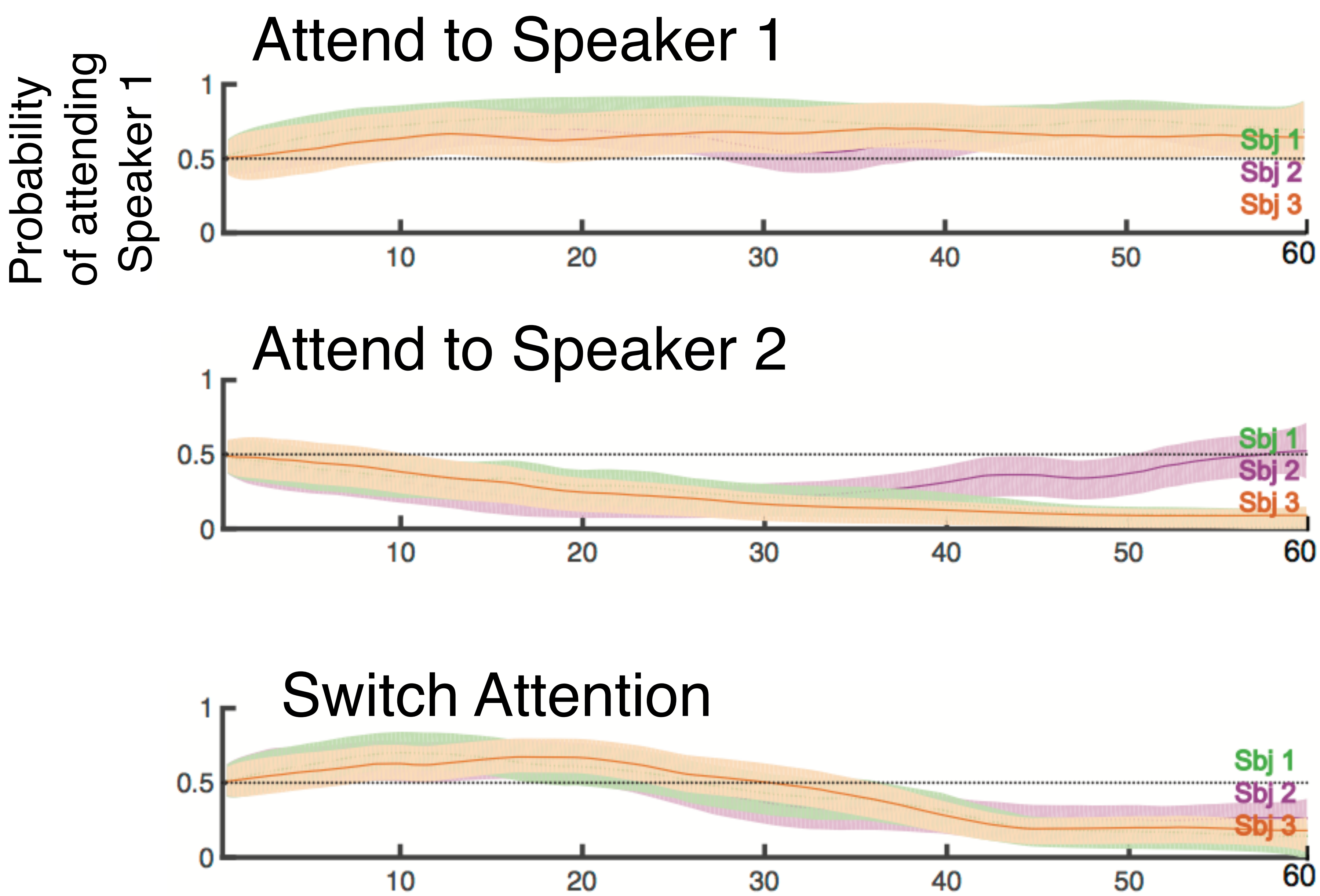
Attentional Dynamics



- 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)

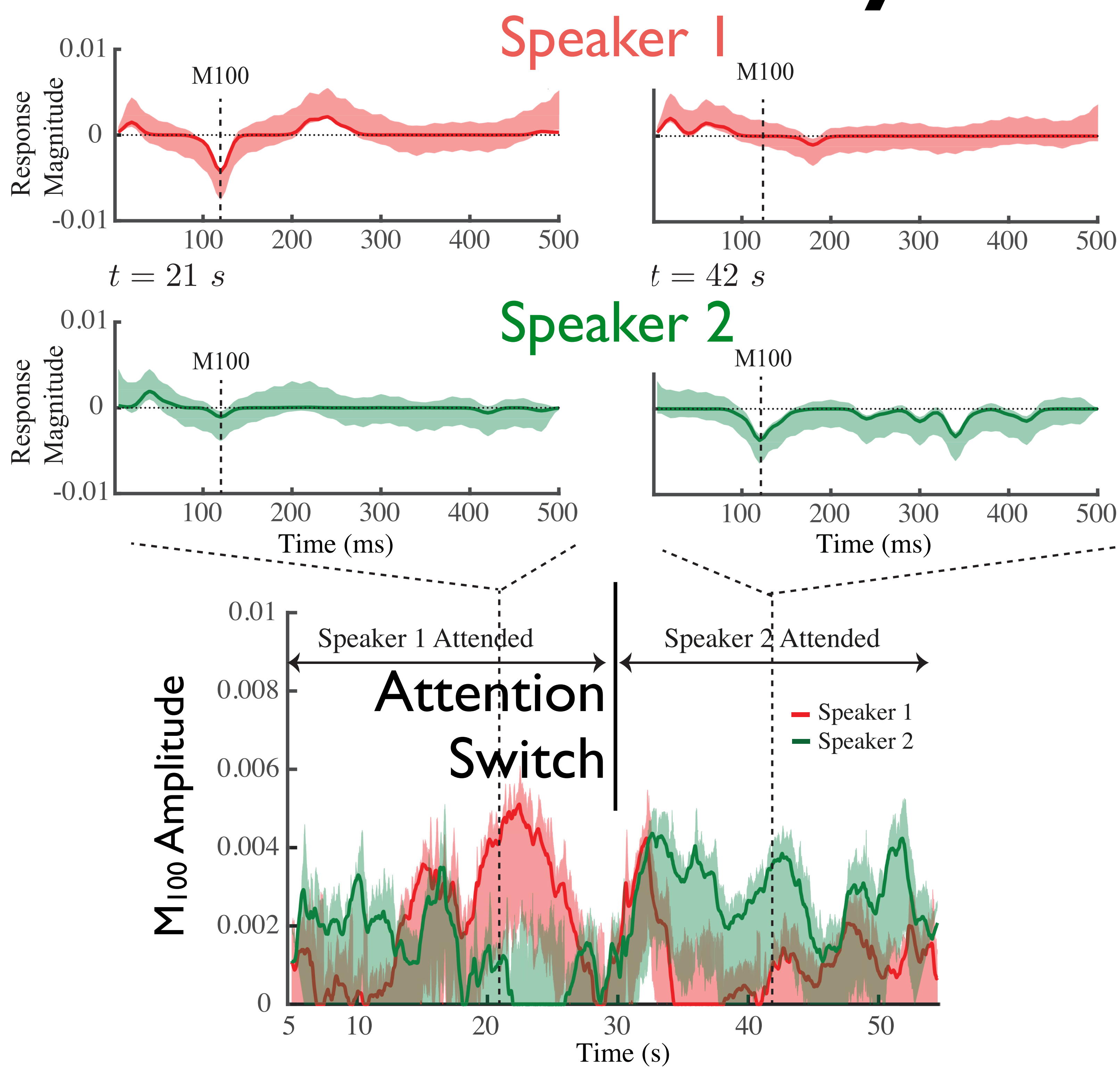
Attentional Dynamics



- 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
- Uses SPARLS algorithm developed by Babadi

Akram et al., IEEE Tr Biomed Eng (2017)

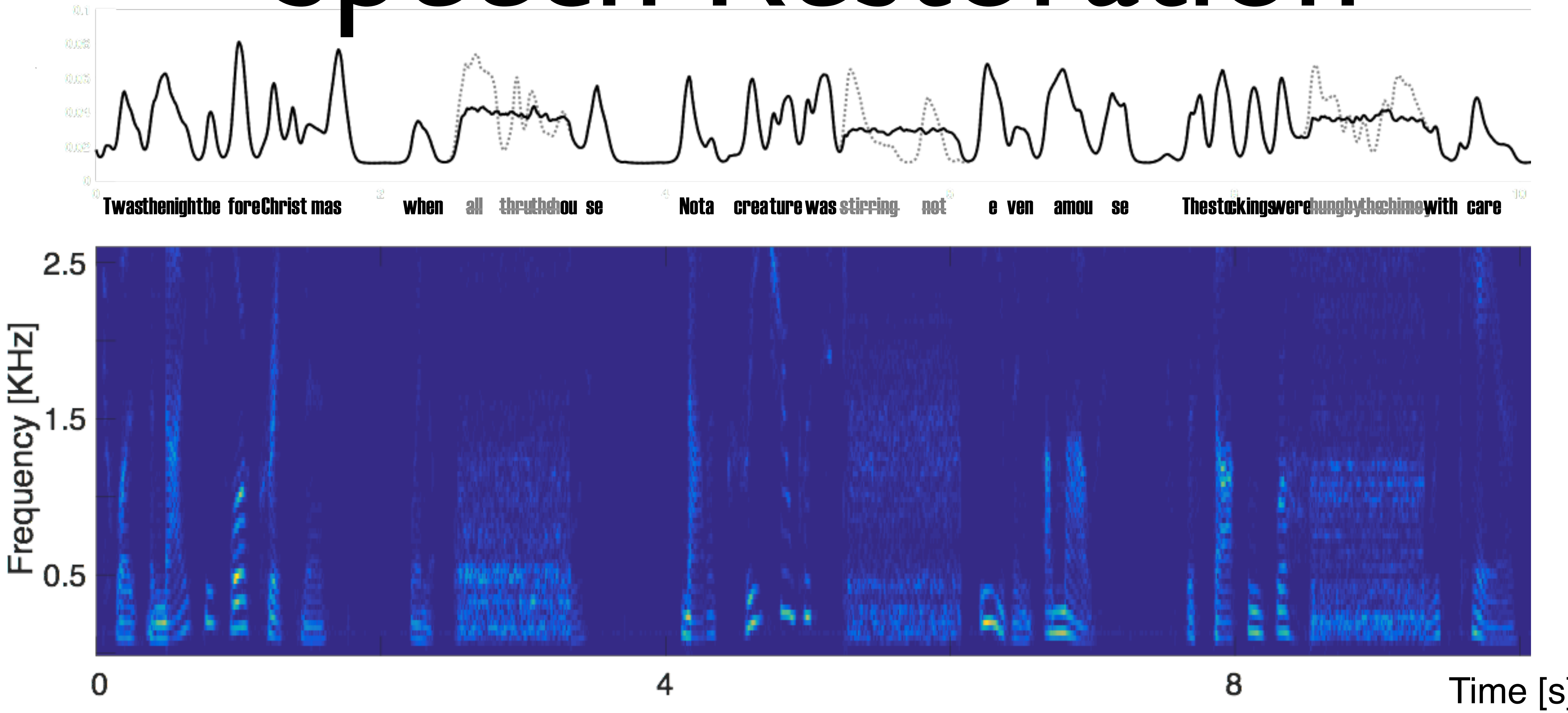
Outline

- Cortical Representations of Speech (via MEG)
  - Encoding vs. Decoding
- “Cocktail Party” Speech
- Recent Results
  - Attentional Dynamics
  - “Restoration” of Missing Speech
  - Speech Processing Across the Brain

Outline

- Cortical Representations of Speech (via MEG)
  - Encoding vs. Decoding
- "Cocktail Party" Speech
- Recent Results
  - Attentional Dynamics
  - "Restoration" of Missing Speech
  - Speech Processing Across the Brain

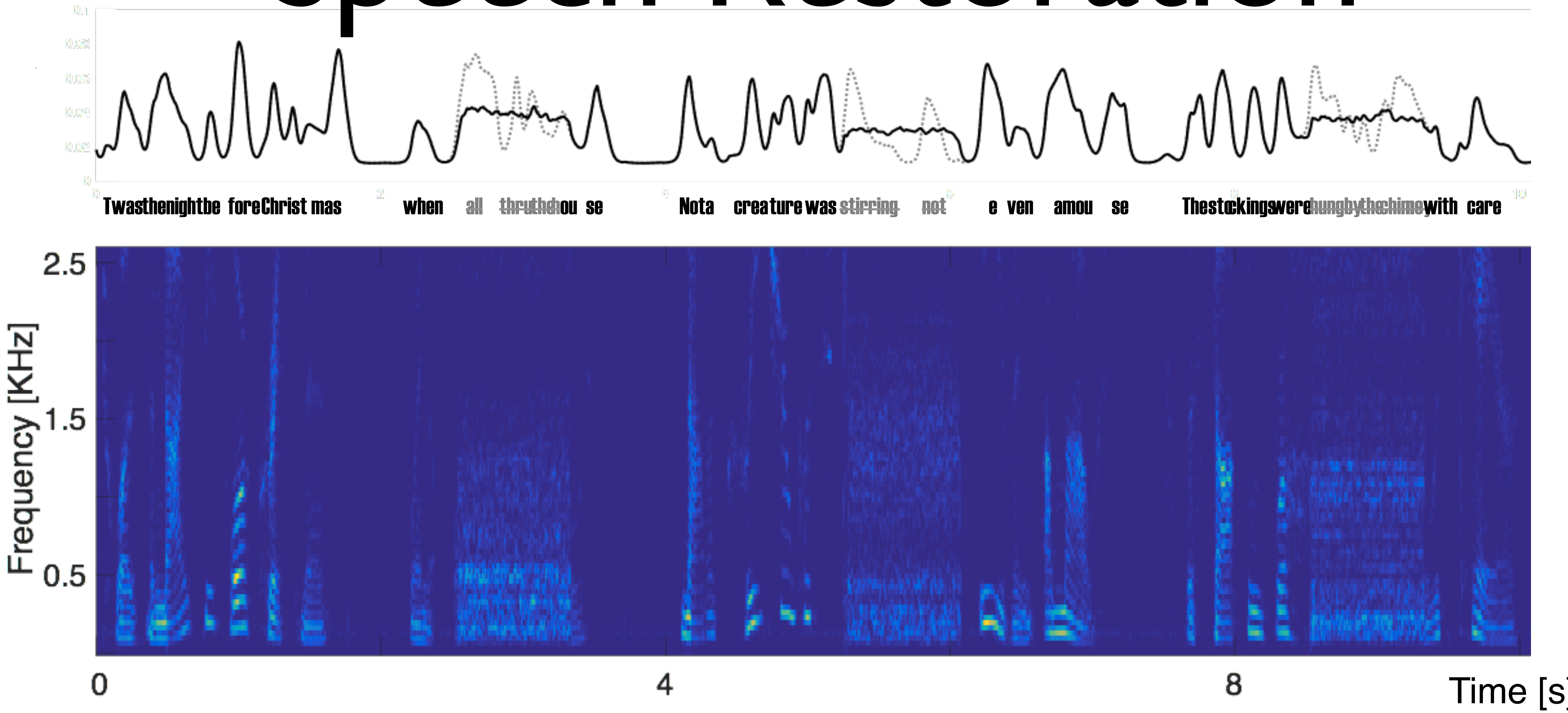
Speech Restoration



- Can sustained, non-stationary, speech be restored?
  - Might be aided by contextual knowledge/familiarity
  - Might be aided by strong rhythmicity

Cervantes Constantino & Simon, in Preparation

Speech Restoration



- Can sustained, non-stationary, speech be restored?
  - Might be aided by contextual knowledge/familiarity
  - Might be aided by strong rhythmicity

Cervantes Constantino & Simon, in Preparation

Speech Restoration

Twas the night before Christmas, when all through the house
not a creature was stirring, not even a mouse.
The stockings were hung by the chimney with care,
in hopes that St. Nicholas soon would be there.

The children were nestled all snug in their beds,
while visions of sugar plums danced in their heads.
And Mama in her 'kerchief, and I in my cap,
had just settled our brains for a long winter's nap.

When out on the lawn there arose such a clatter,
I sprang from my bed to see what was the matter.
Away to the window I flew like a flash,
tore open the shutter, and threw up the sash.

The moon on the breast of the new-fallen snow
gave the lustre of midday to objects below,
when, what to my wondering eyes should appear,
but a miniature sleigh and eight tiny reindeer.

With a little old driver, so lively and quick,
I knew in a moment it must be St. Nick.
More rapid than eagles, his coursers they came,
and he whistled and shouted and called them by name.

"Now Dasher! Now Dancer! Now, Prancer and Vixen!
On, Comet! On, Cupid! On, Donner and Blitzen!
To the top of the porch! To the top of the wall!
Now dash away! Dash away! Dash away all!"

As dry leaves that before the wild hurricane fly,
when they meet with an obstacle, mount to the sky
so up to the house-top the coursers they flew,
with the sleigh full of toys, and St. Nicholas too.

And then, in a twinkling, I heard on the roof
the prancing and pawing of each little hoof.
As I drew in my head and was turning around,
down the chimney St. Nicholas came with a bound.

He was dressed all in fur, from his head to his foot,
and his clothes were all tarnished with ashes and soot.
A bundle of toys he had flung on his back,
and he looked like a peddler just opening his pack.

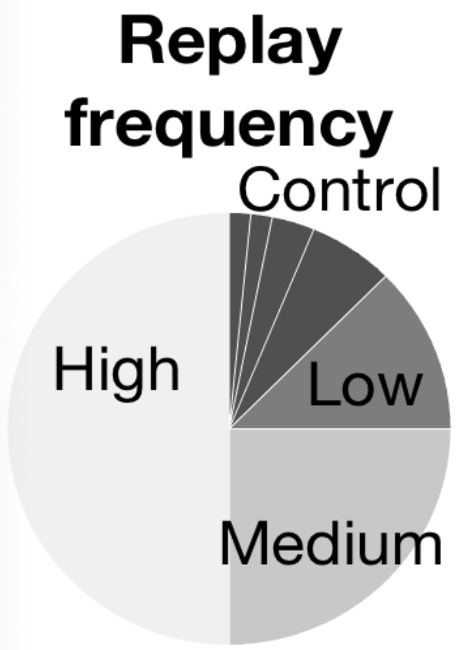
His eyes--how they twinkled! His dimples, how merry!
His cheeks were like roses, his nose like a cherry!
His droll little mouth was drawn up like a bow,
and the beard on his chin was as white as the snow.

The stump of a pipe he held tight in his teeth,
and the smoke it encircled his head like a wreath.
He had a broad face and a little round belly,
that shook when he laughed, like a bowl full of jelly.

He was chubby and plump, a right jolly old elf,
and I laughed when I saw him, in spite of myself.
A wink of his eye and a twist of his head
soon gave me to know I had nothing to dread.

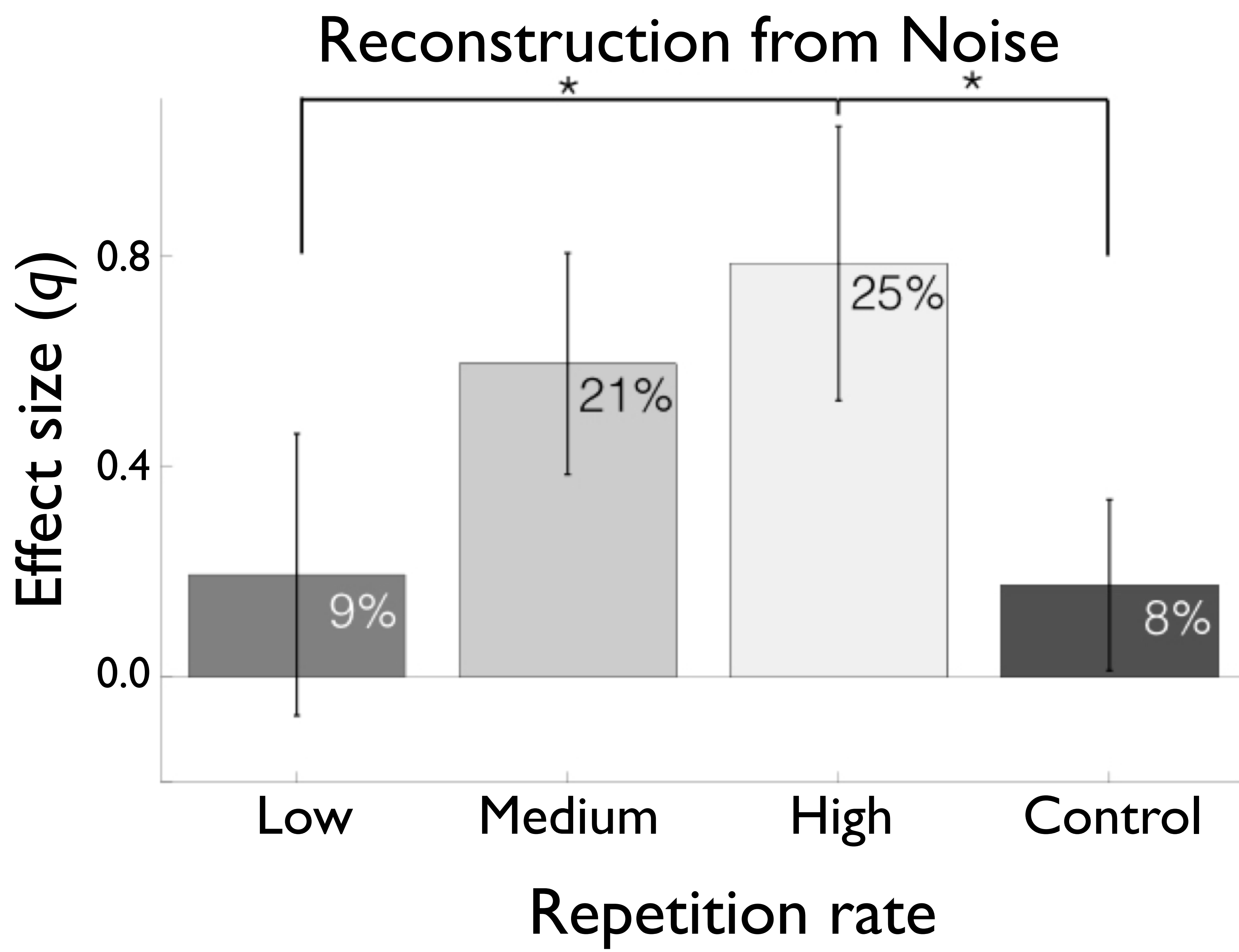
He spoke not a word, but went straight to his work,
and filled all the stockings, then turned with a jerk.
And laying his finger aside of his nose,
and giving a nod, up the chimney he rose.

He sprang to his sleigh, to his team gave a whistle,
And away they all flew like the down of a thistle.
But I heard him exclaim, 'ere he drove out of sight,
"Happy Christmas to all, and to all a good night!"

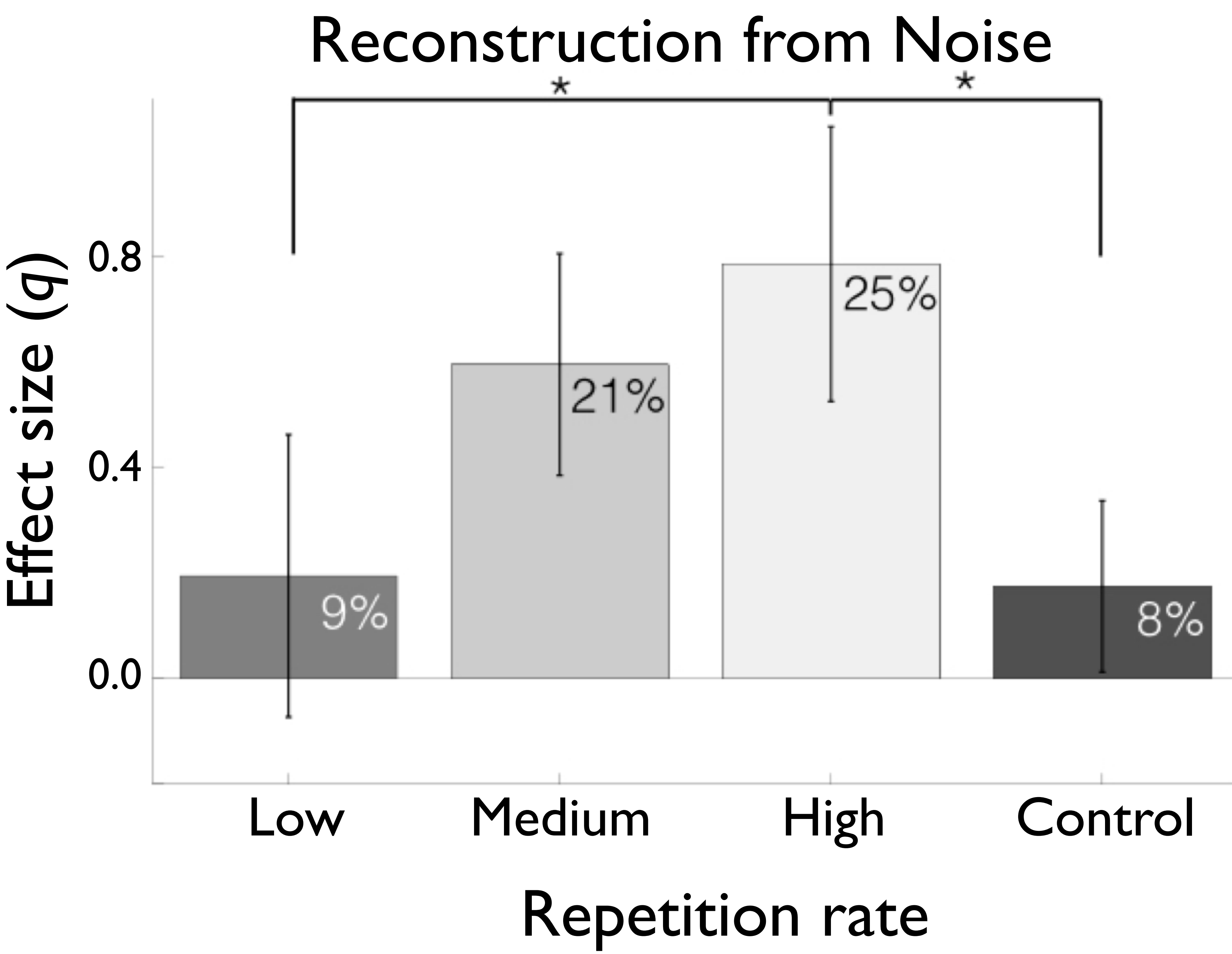


- Hypothesis: contextual knowledge of missing speech can be controlled by exposure to the speech

Speech Restoration

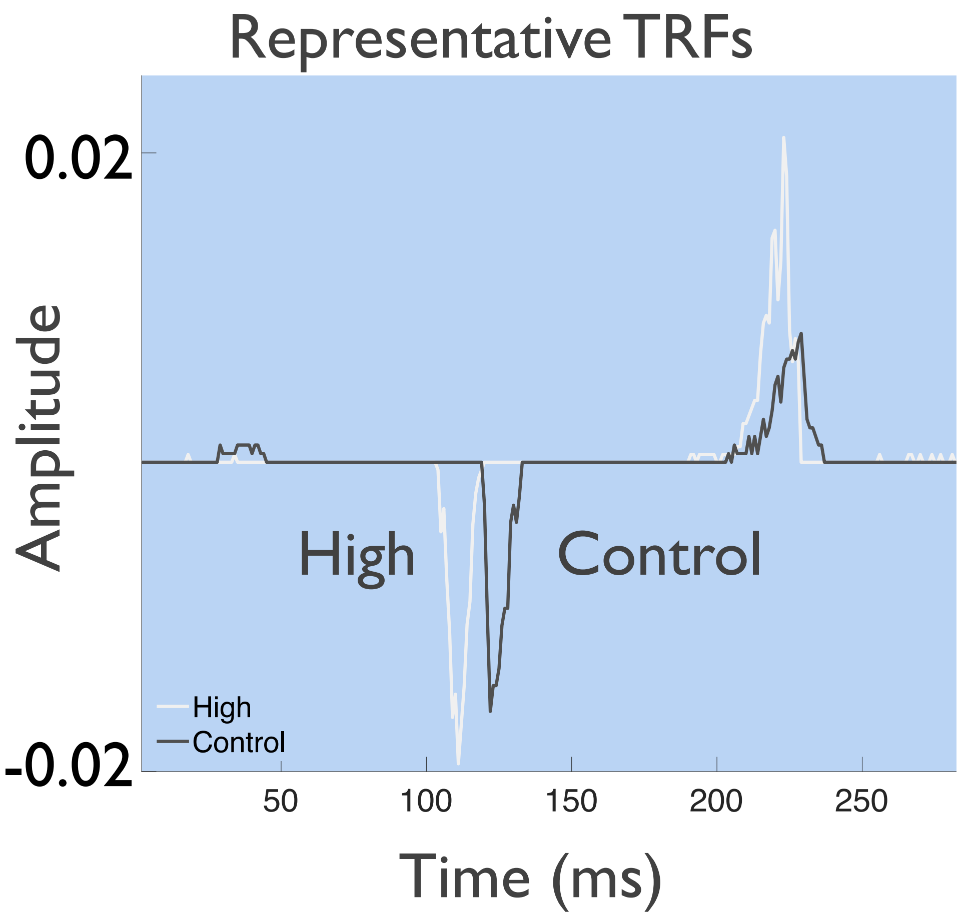


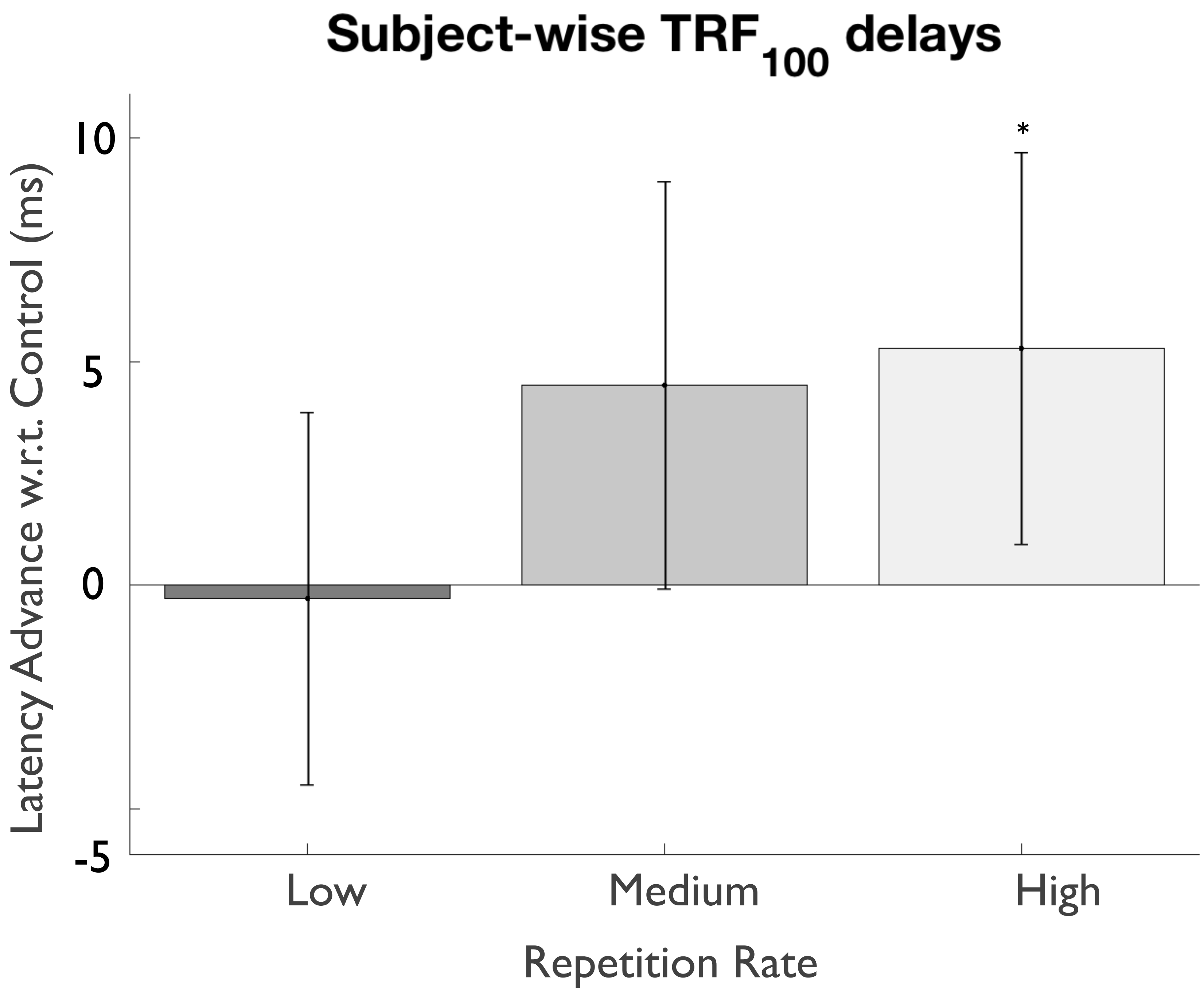
Speech Restoration



- Decoding of the ***missing*** speech token improves with prior experience
- Performance is a considerable fraction of that for clean speech

Speech Anticipation





- Prior experience speeds subsequent responses

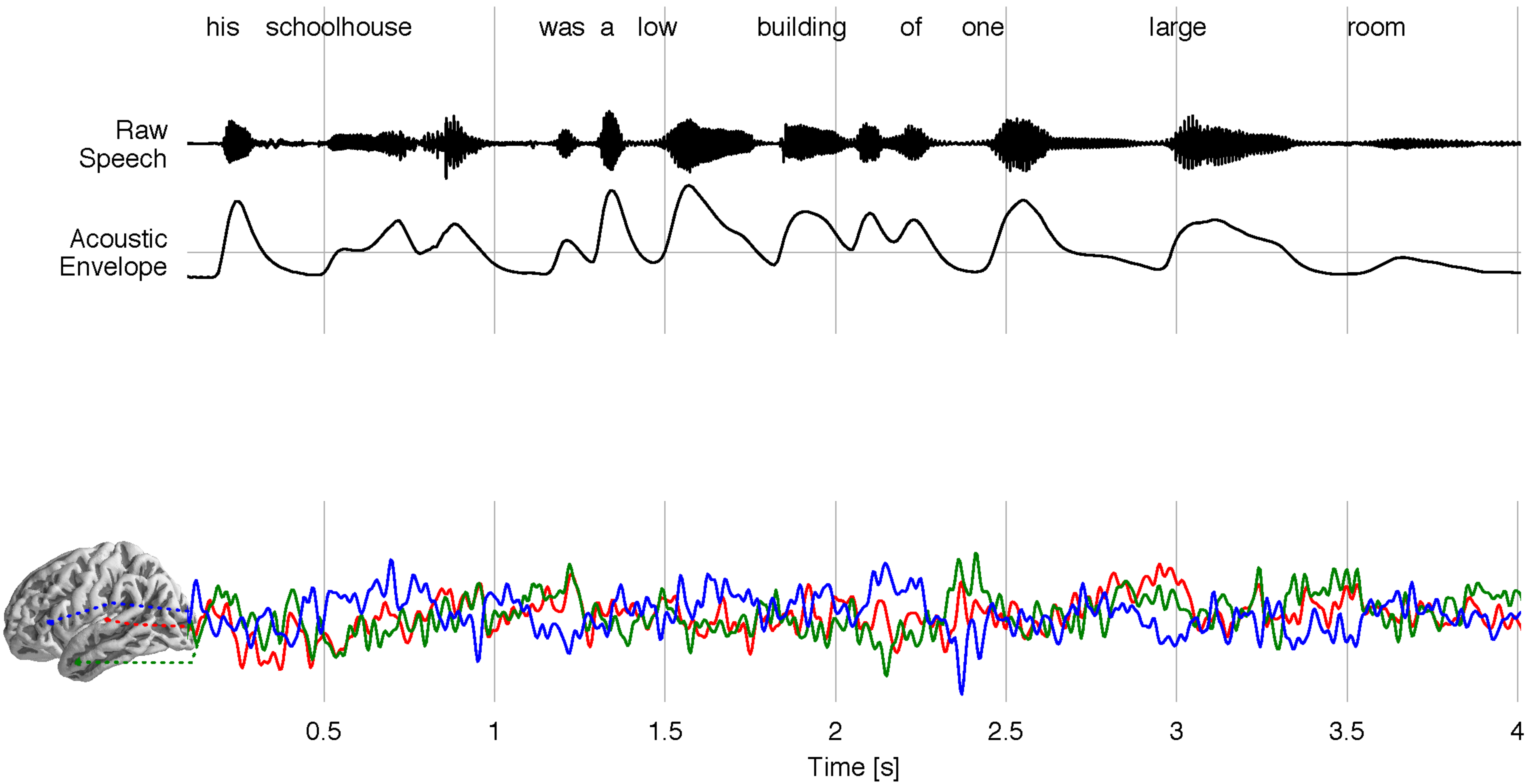
Outline

- Cortical Representations of Speech (via MEG)
  - Encoding vs. Decoding
- “Cocktail Party” Speech
- Recent Results
  - Attentional Dynamics
  - “Restoration” of Missing Speech
  - Speech Processing Across the Brain

Outline

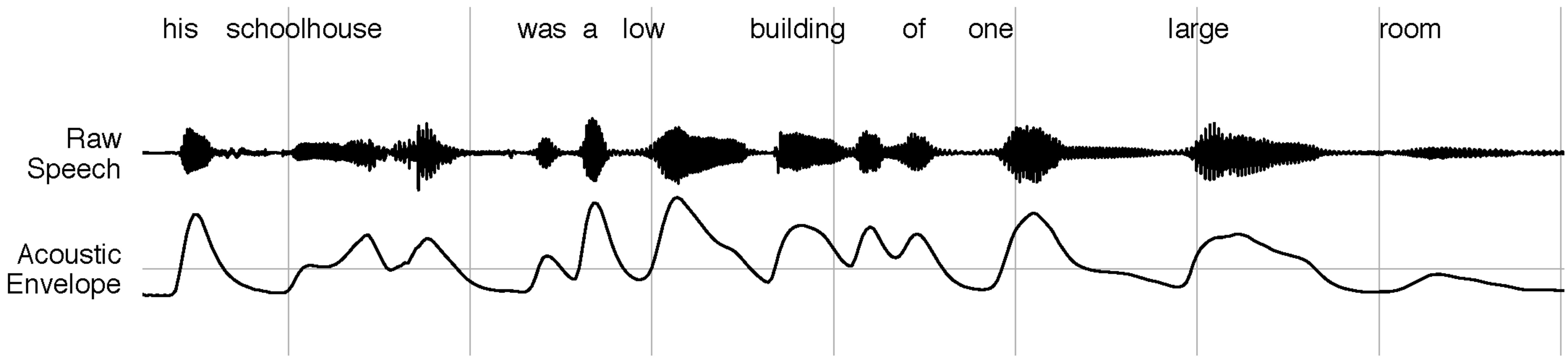
- Cortical Representations of Speech (via MEG)
  - Encoding vs. Decoding
- “Cocktail Party” Speech
- Recent Results
  - Attentional Dynamics
  - “Restoration” of Missing Speech
  - Speech Processing Across the Brain

Localizing Speech Processing



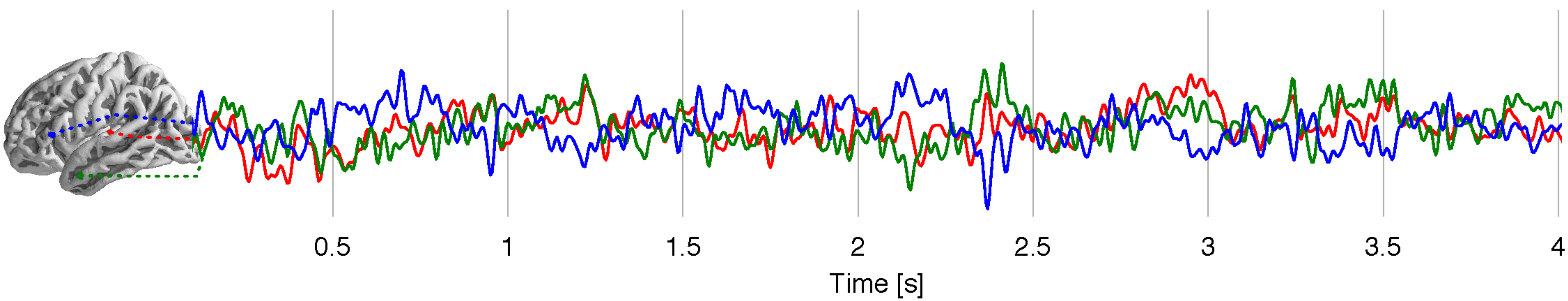
Brodbeck et al., bioRxiv 2017

Localizing Speech Processing



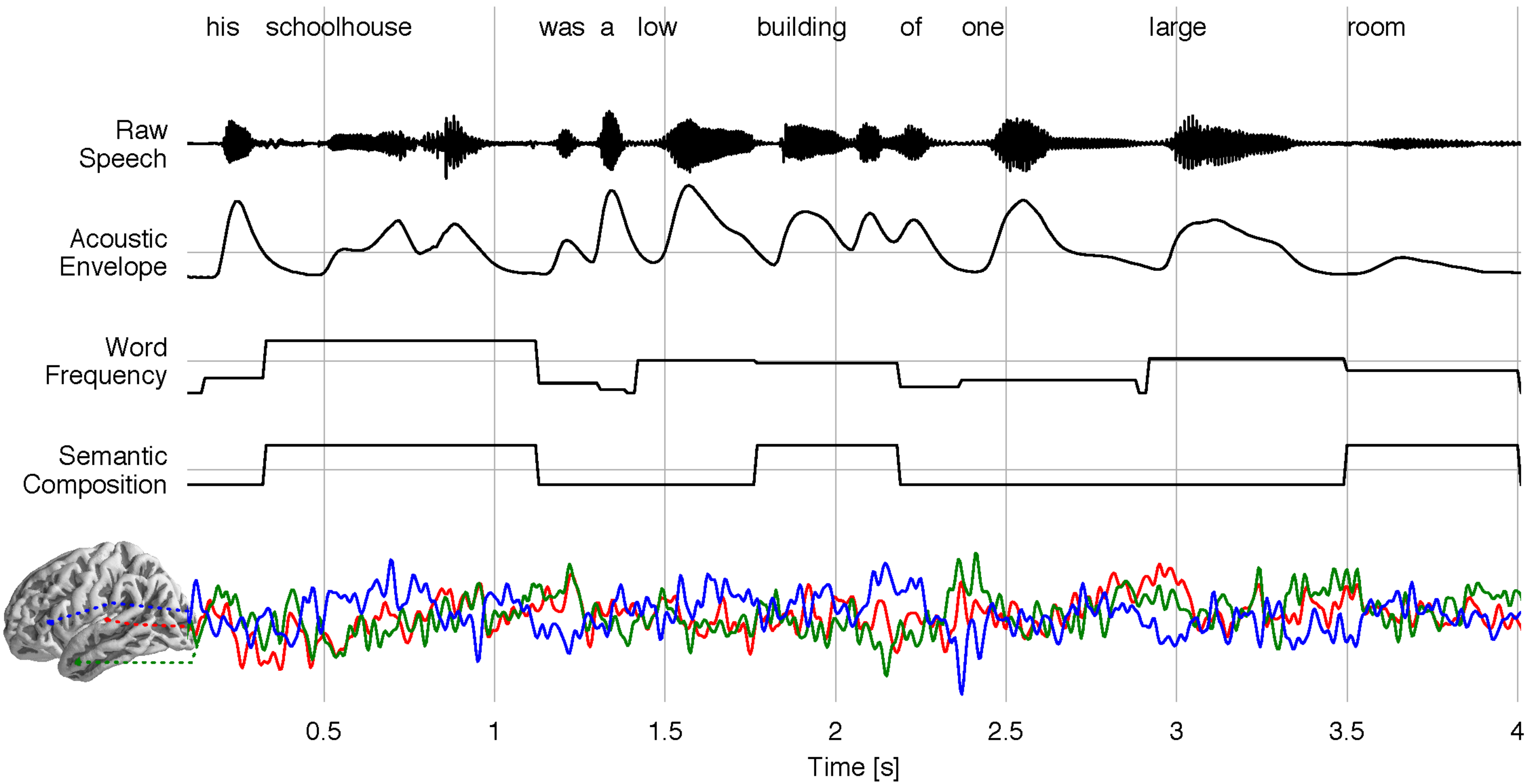
Forward model = source-to-sensor matrix (L)

Each neural source is linear superposition of sensor responses

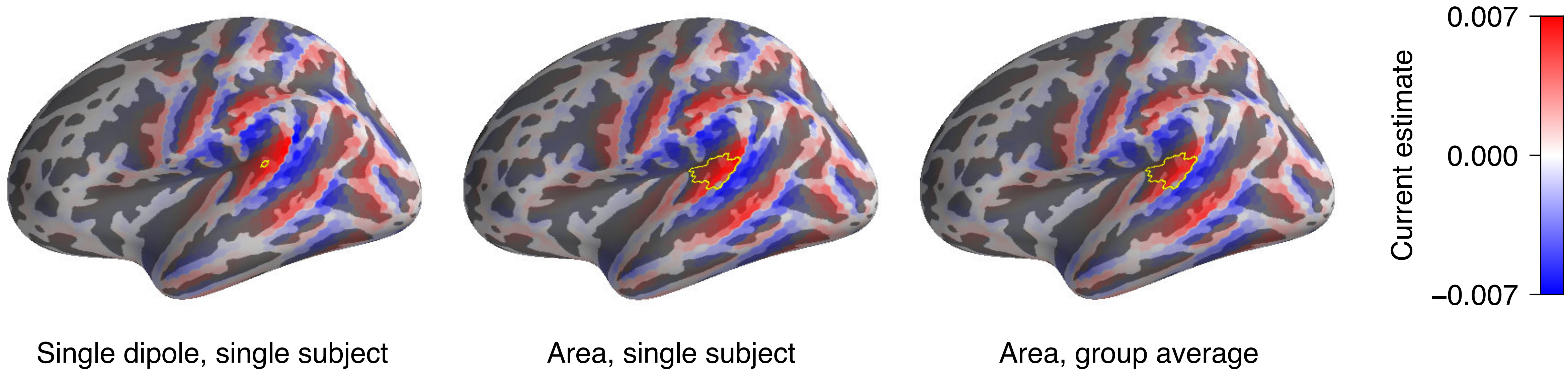


Brodbeck et al., bioRxiv 2017

Localizing Speech Processing



Point Spread Function



Source estimate for hypothetical point source

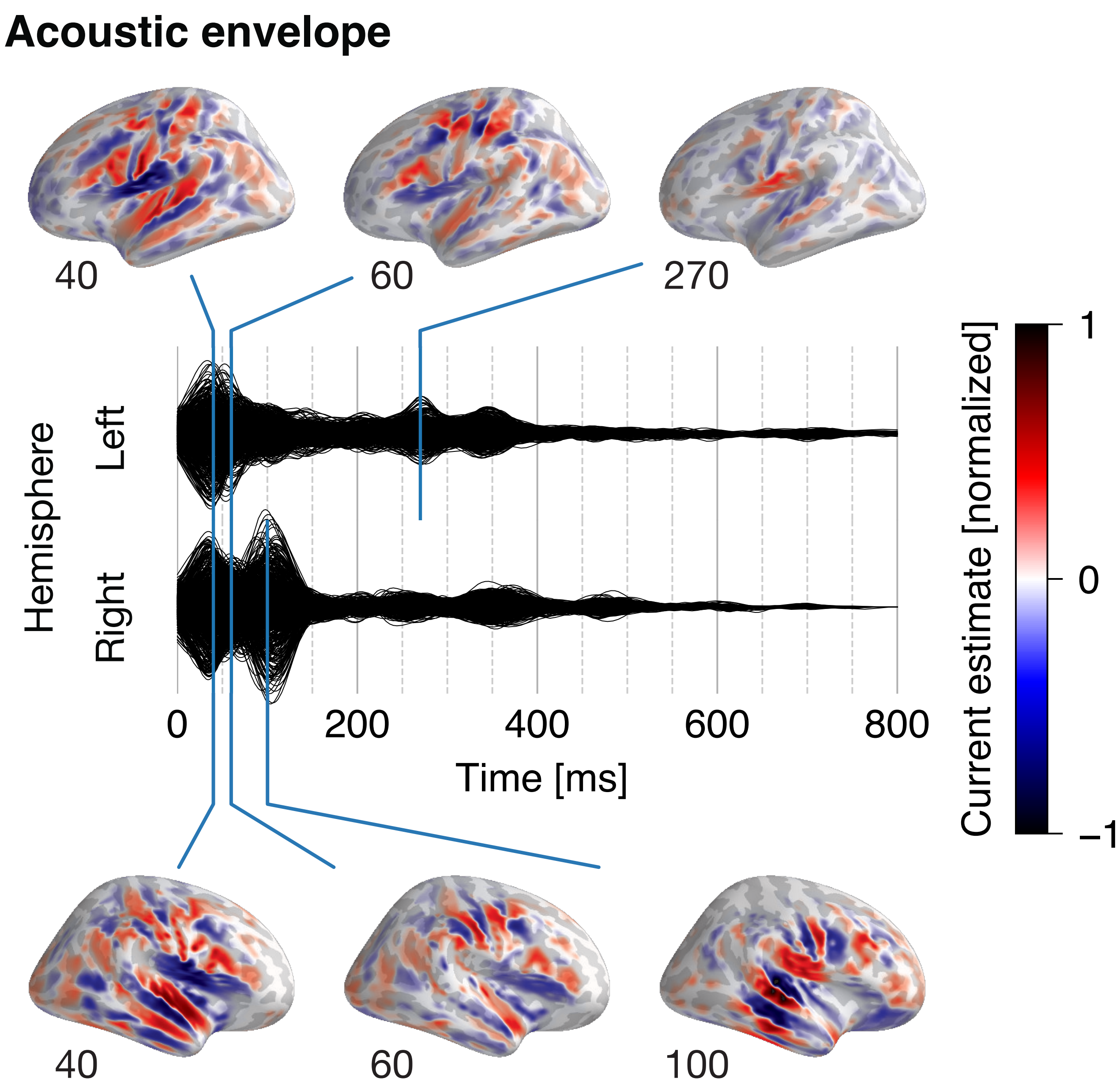
Forward model = source-to-sensor matrix L

Inverse model = sensor-to-source matrix G

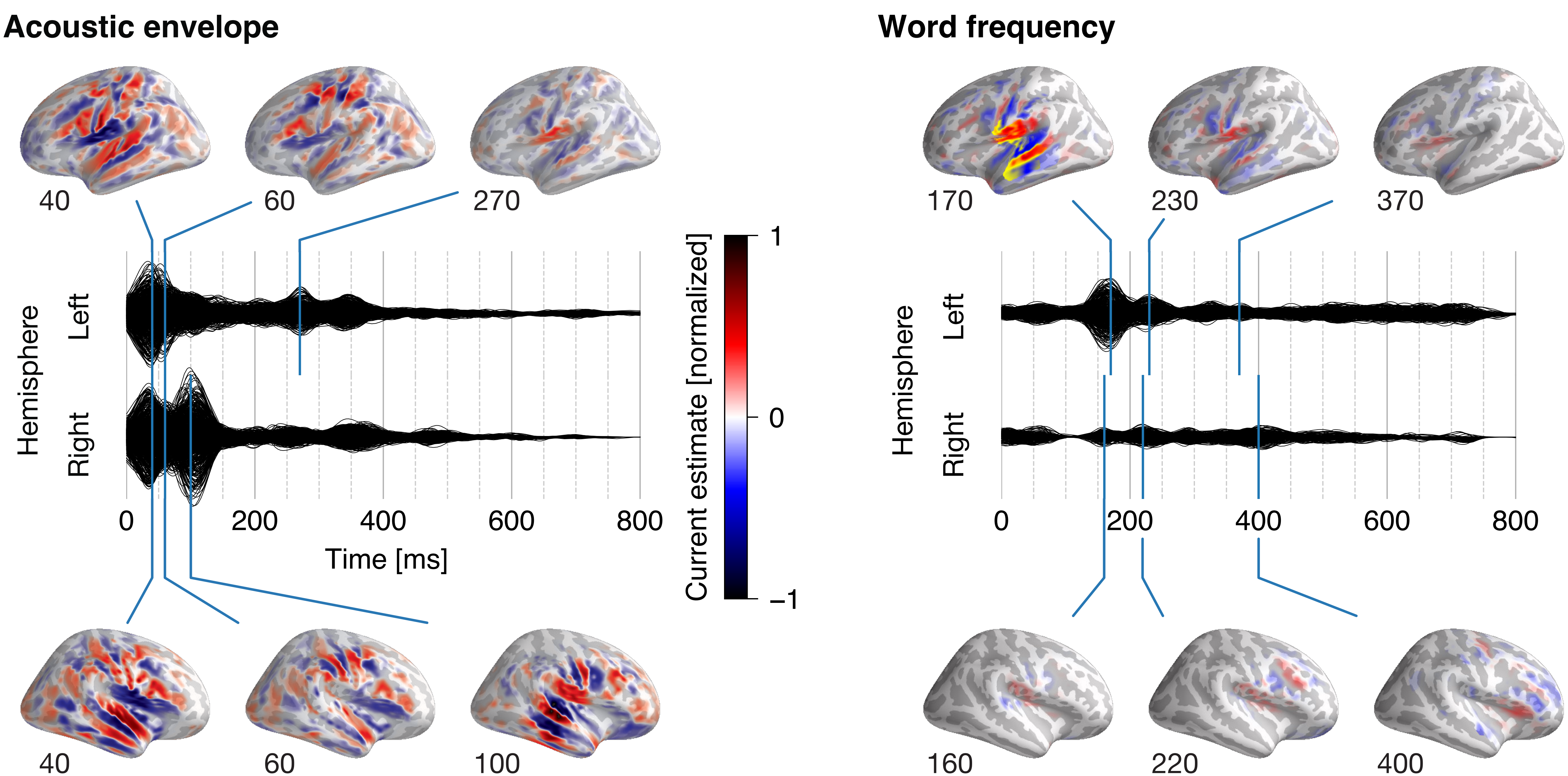
➡ Source *estimate* of a *hypothetical* source j: G•L•j.

= ***Point Spread Function***

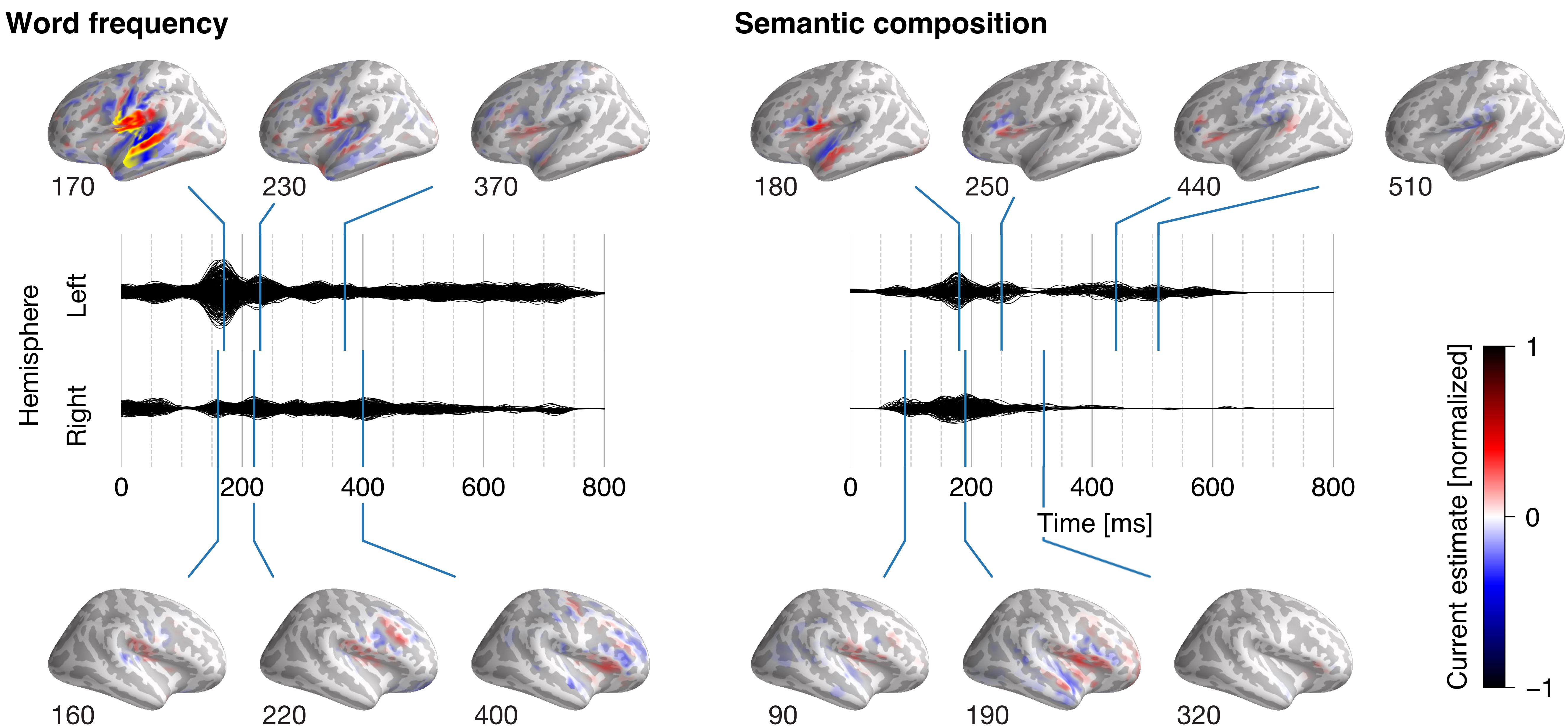
Localized TRFs



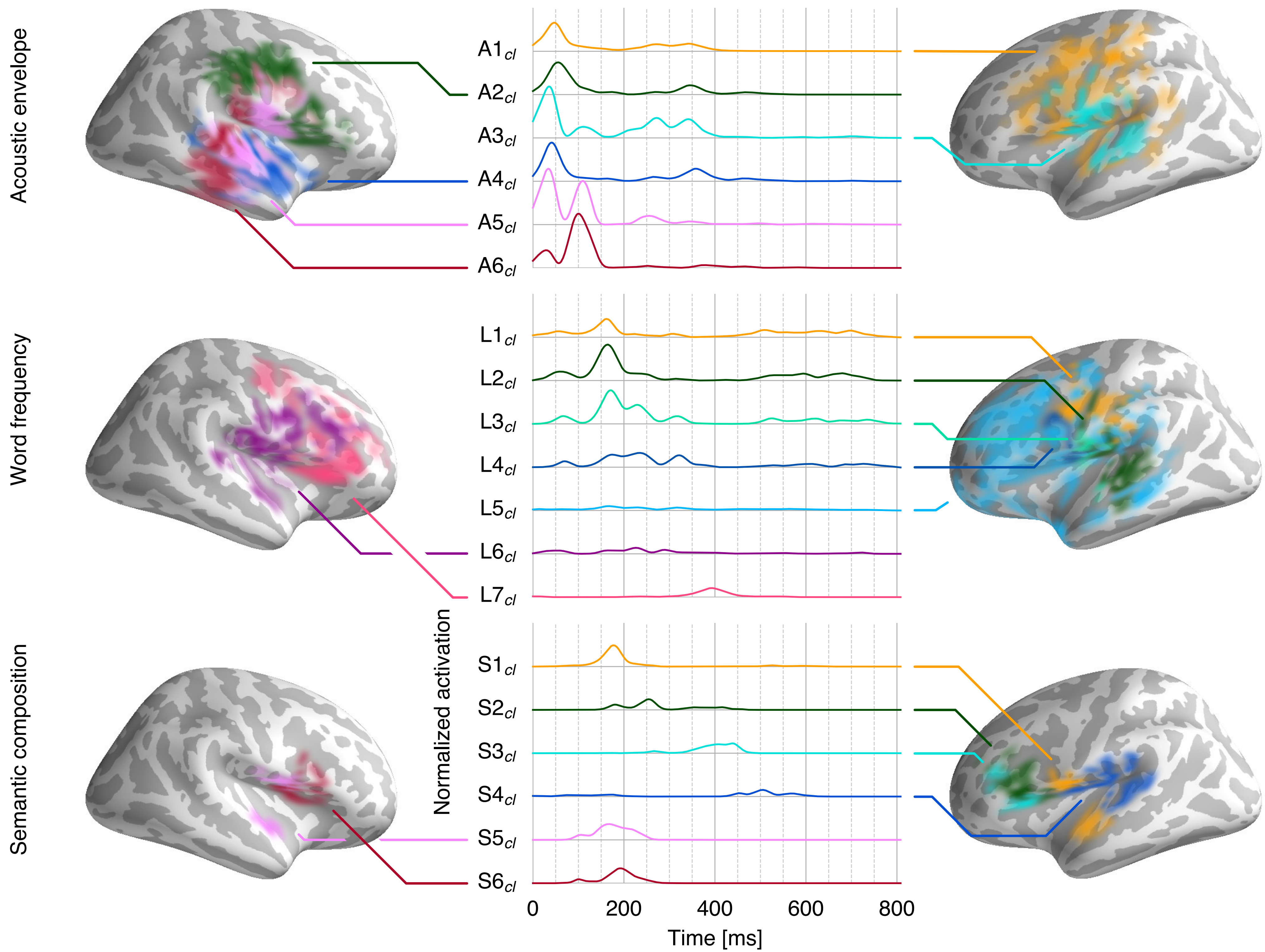
Localized TRFs



Localized TRFs



Clustered Localized TRFs



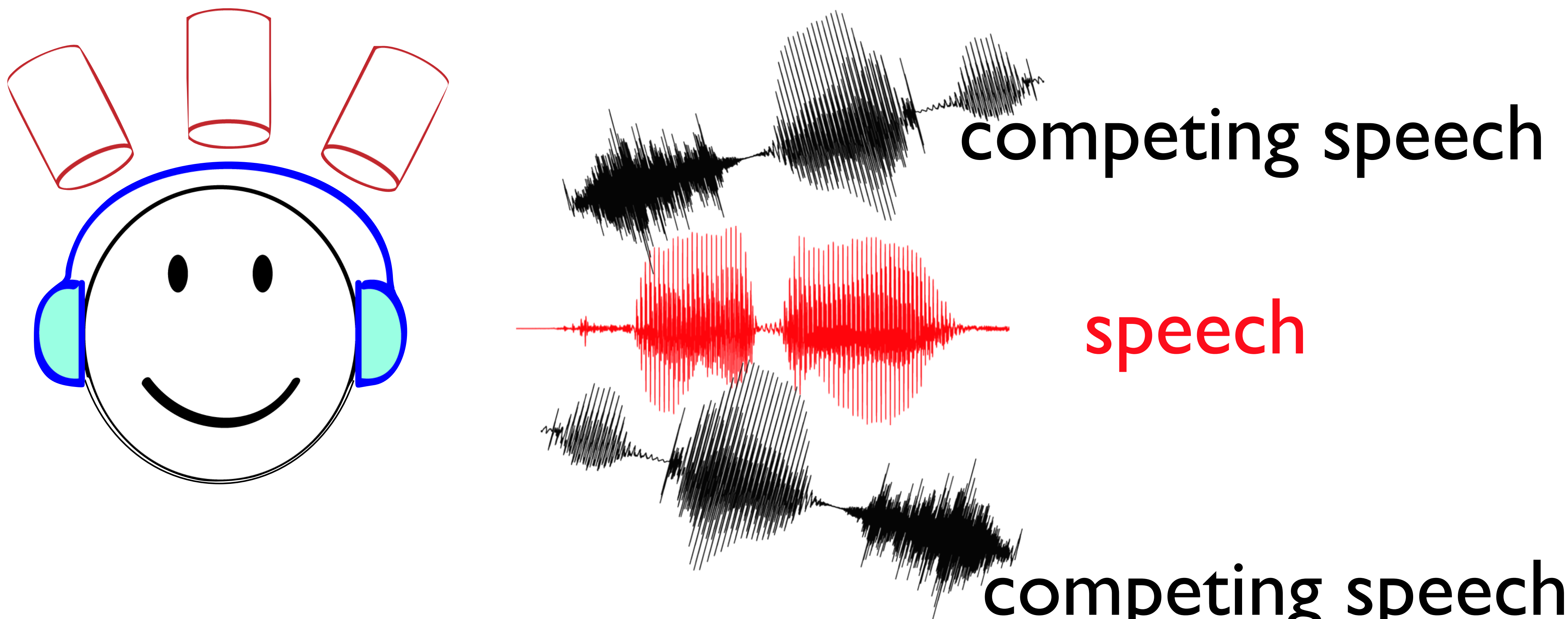
Clustering allows "beating the point spread"

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 monitoring of attention
- "Restoration" of speech at brain level
  - neural processing tracks behavior
- Systems Approach works at neural source level
  - with higher order aspects of speech

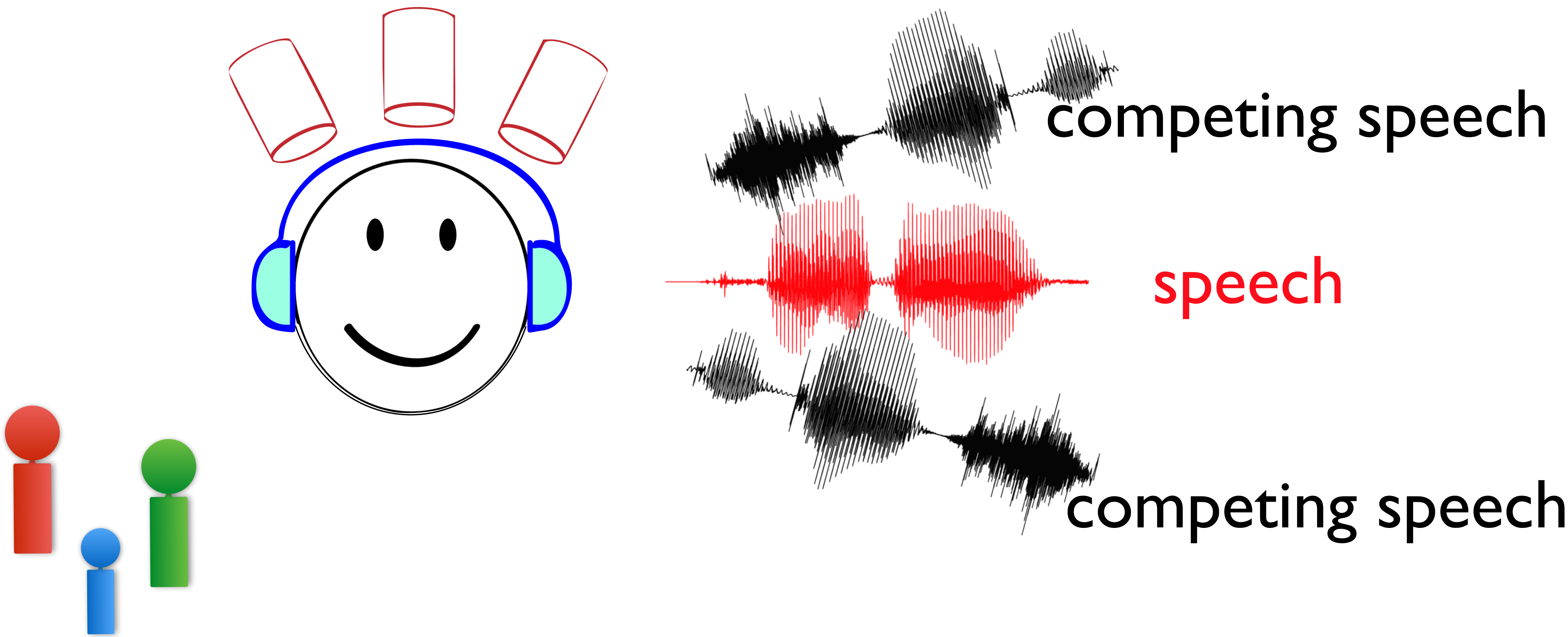
Thank You

Three Competing Speakers



Puvvada & Simon, bioRχiv (2017)

Three Competing Speakers



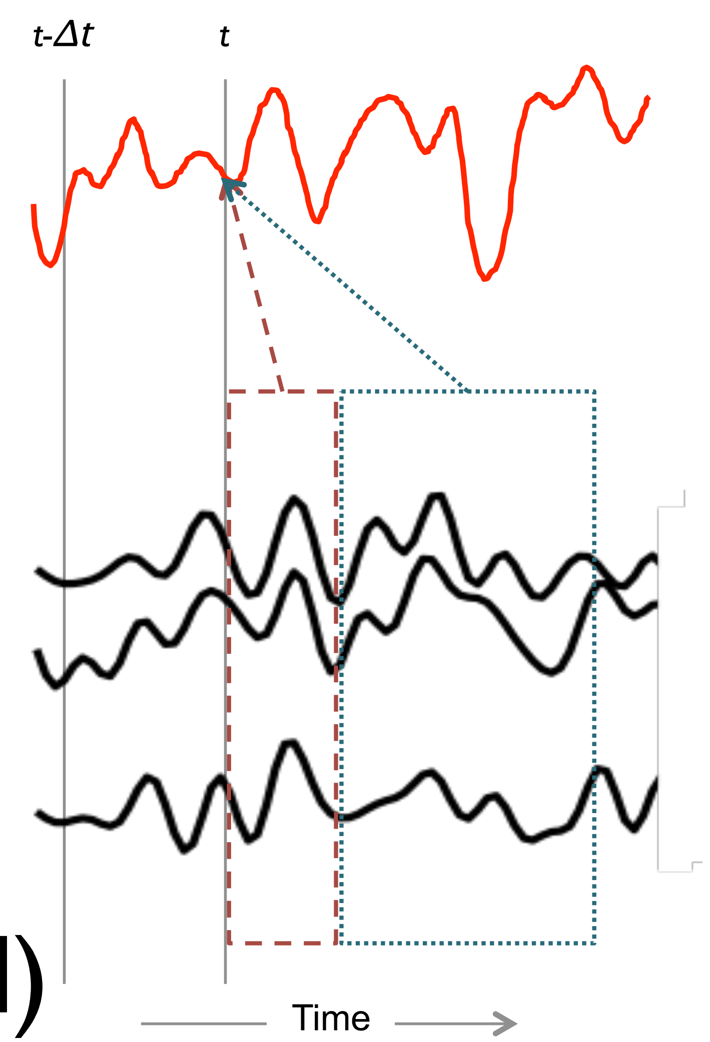
Puvvada & Simon, bioRχiv (2017)

Idea

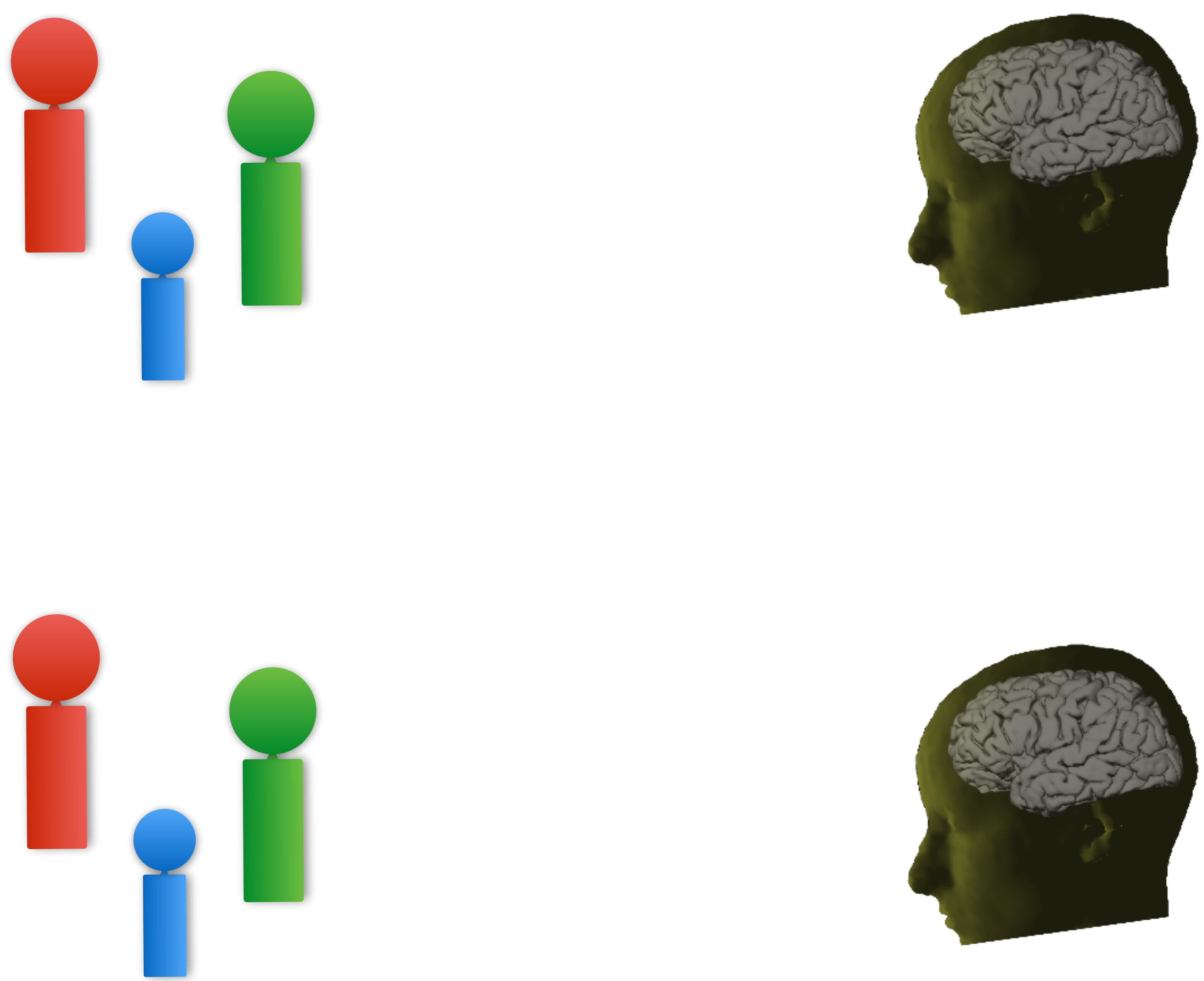
- Latency as Proxy for Cortical Area(s)
  - Earlier Latency Responses from Heschl's Gyrus
  - Later Latency Responses from Planum Temporale (and beyond)
- Not just for Response but also Reconstruction
  - Earlier Integration Window Reconstructs from HG
  - Later Integration Window Reconstructs from PT (and beyond)

Idea

- Latency as Proxy for Cortical Area(s)
  - Earlier Latency Responses from Heschl's Gyrus
  - Later Latency Responses from Planum Temporale (and beyond)
- Not just for Response but also Reconstruction
  - Earlier Integration Window Reconstructs from HG
  - Later Integration Window Reconstructs from PT (and beyond)

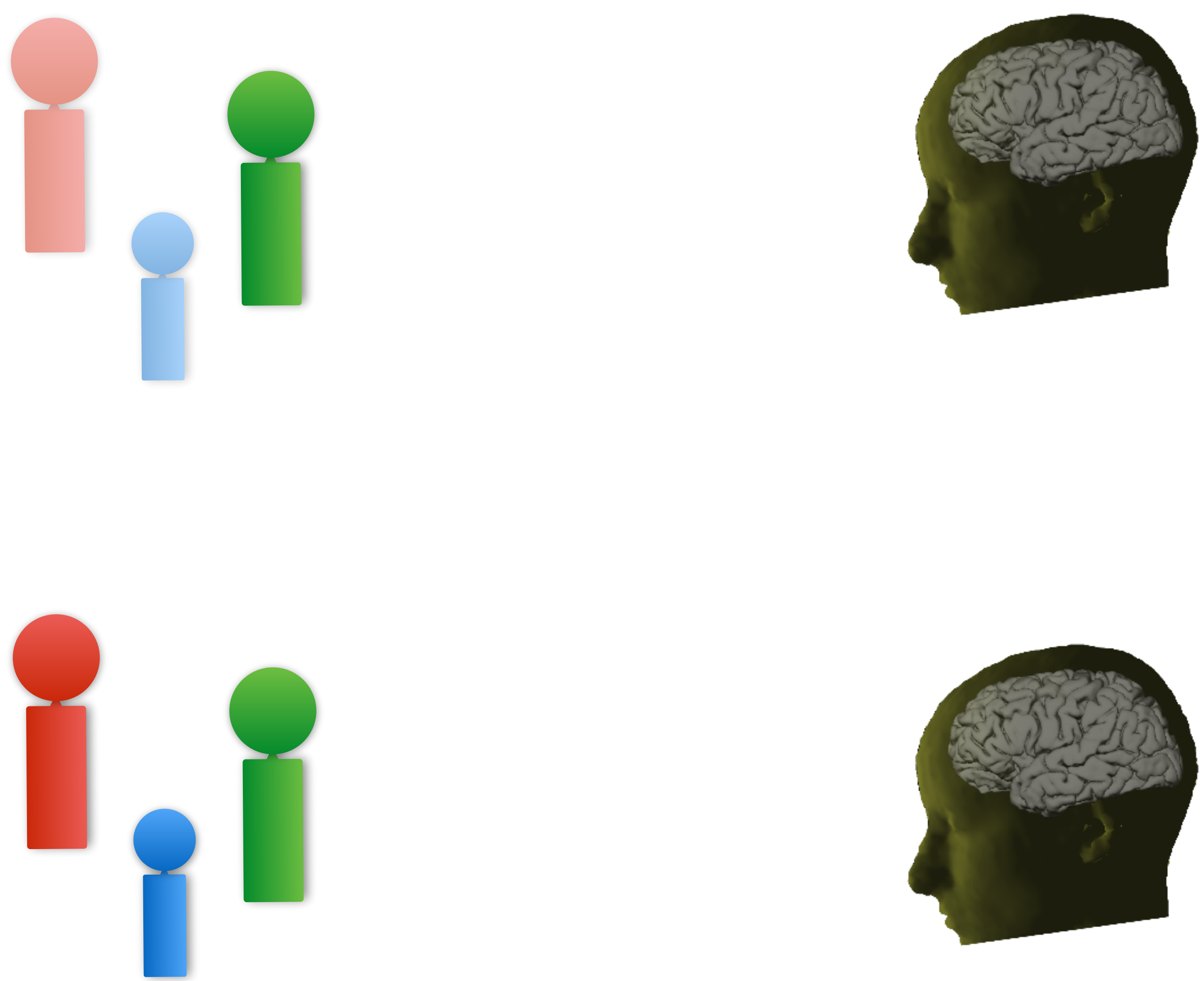


Where in Cortex is there a Segregated Foreground?



Puvvada & Simon, bioRχiv (2017)

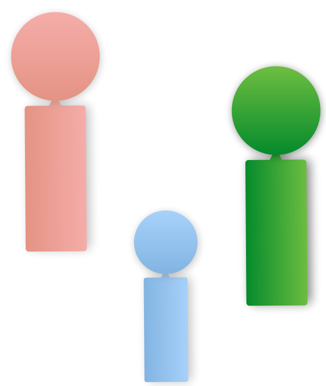
Where in Cortex is there a Segregated Foreground?

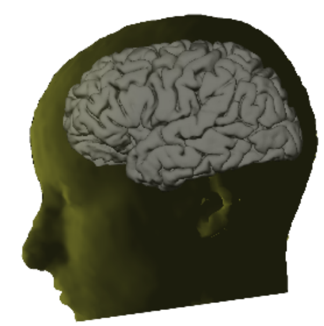


Puvvada & Simon, bioRχiv (2017)

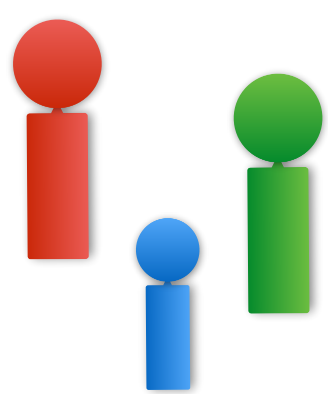
Where <u>in Cortex</u> is there a Segregated Foreground?

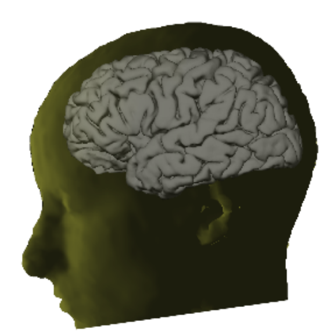
Late?





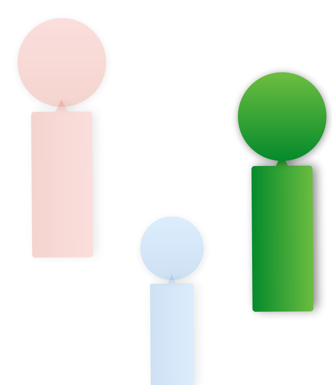
Early?



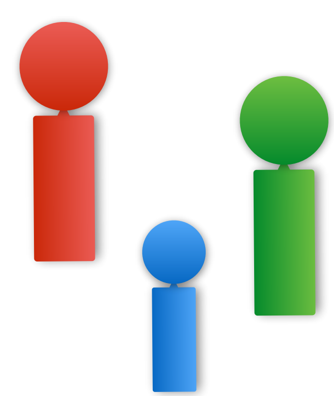


Puvvada & Simon, bioRχiv (2017)

Late Cortical Reconstruction



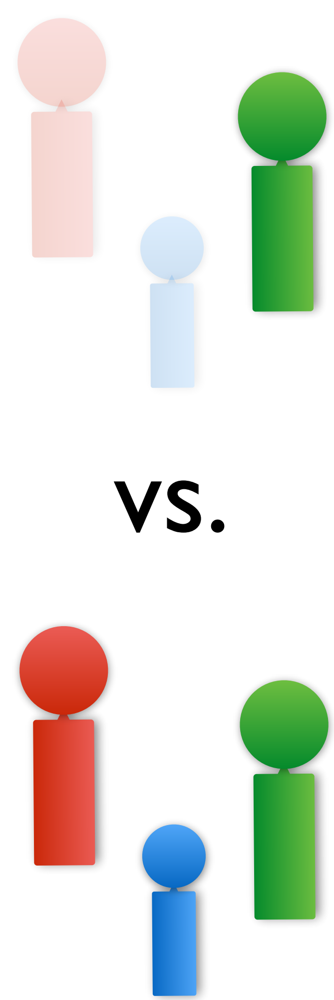
vs.

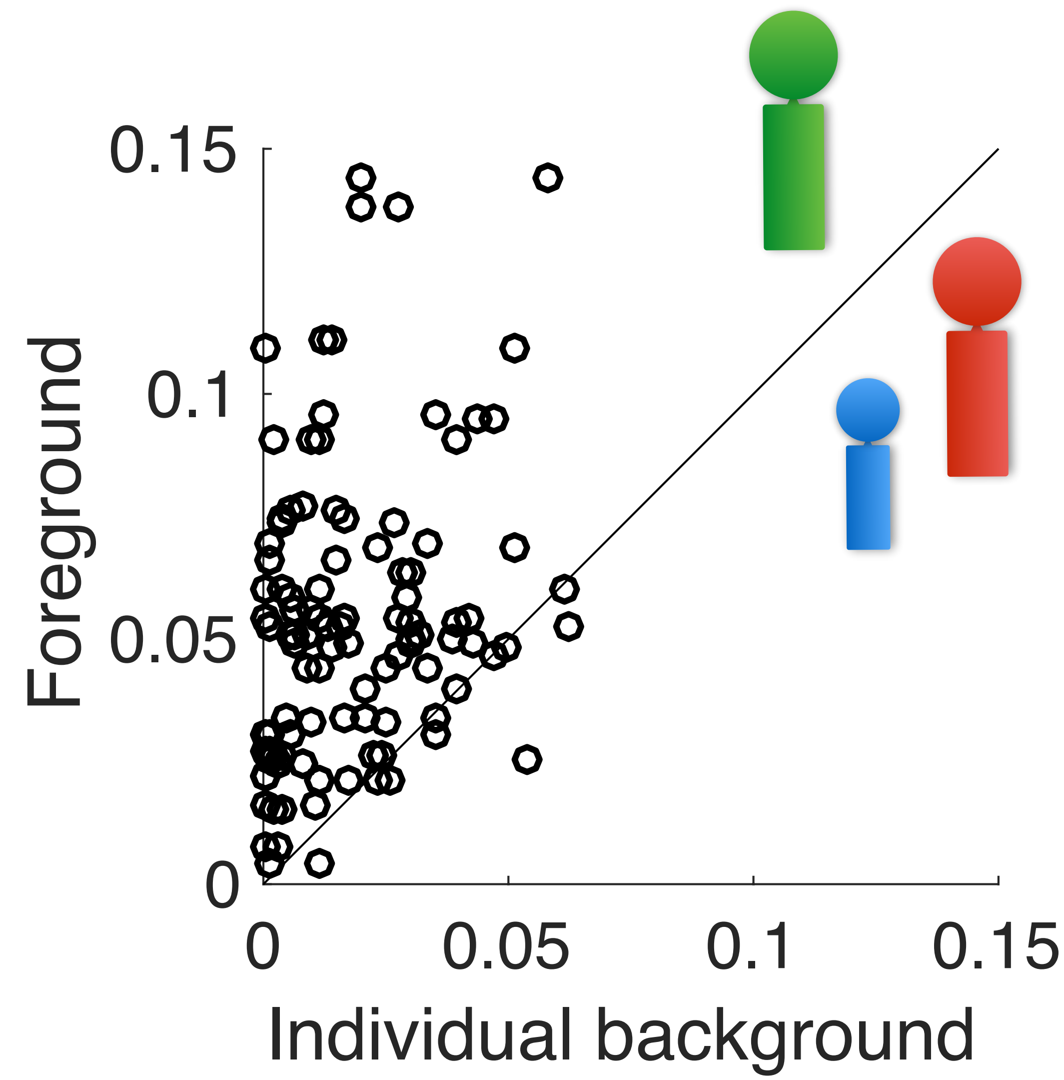


Puvvada & Simon, bioRχiv (2017)

Integration Window over Late Times Only

Late Cortical Reconstruction



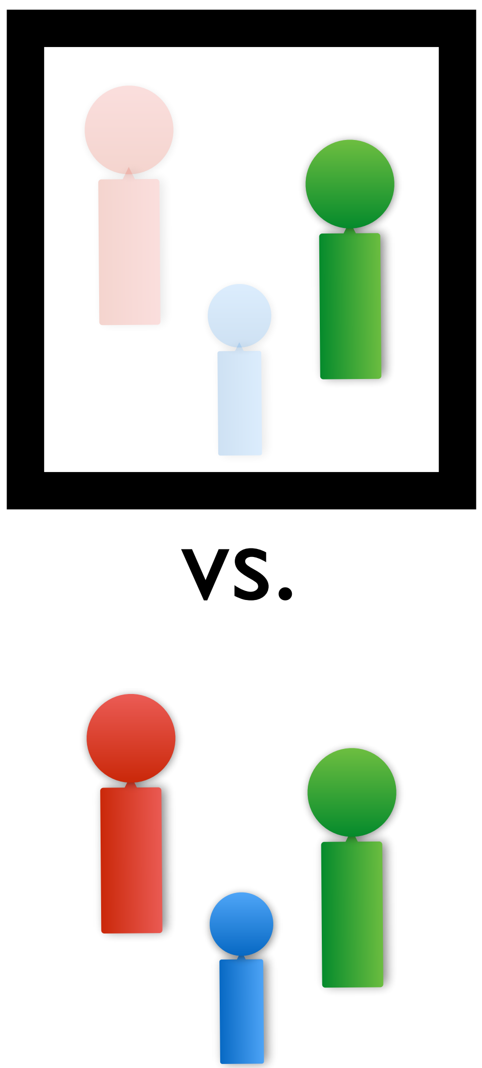


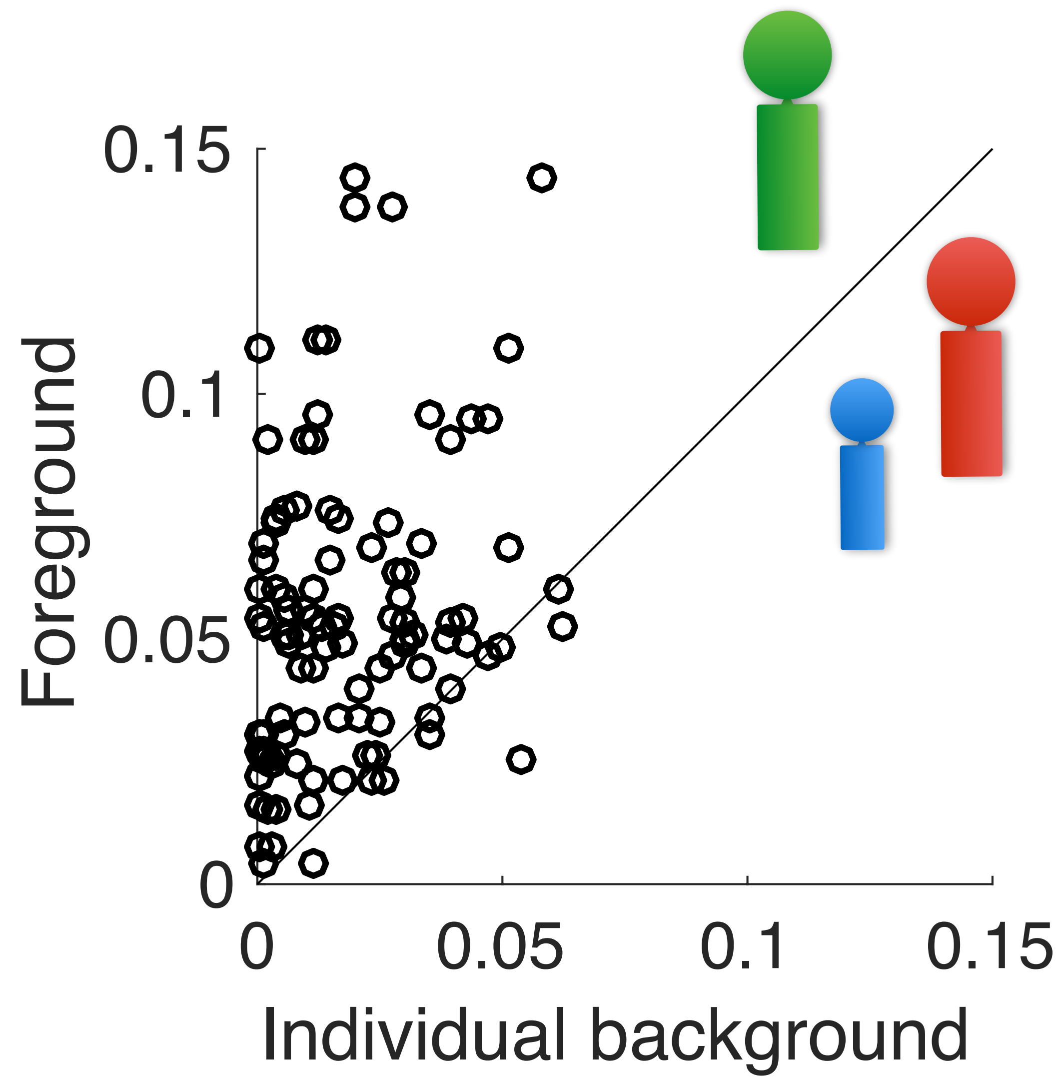
Note: r^2 Scatterplot (not r)

Puvvada & Simon, bioRχiv (2017)

Integration Window over Late Times Only

Late Cortical Reconstruction



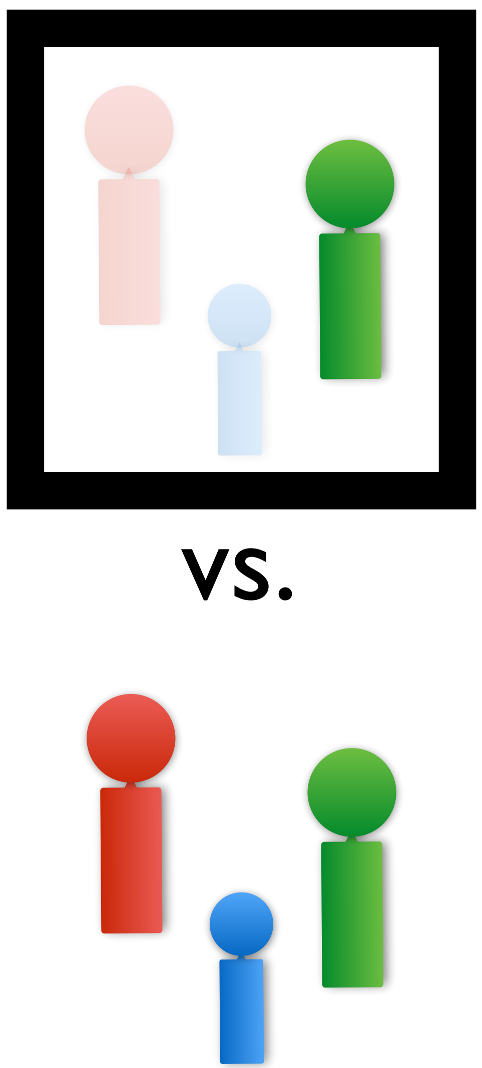


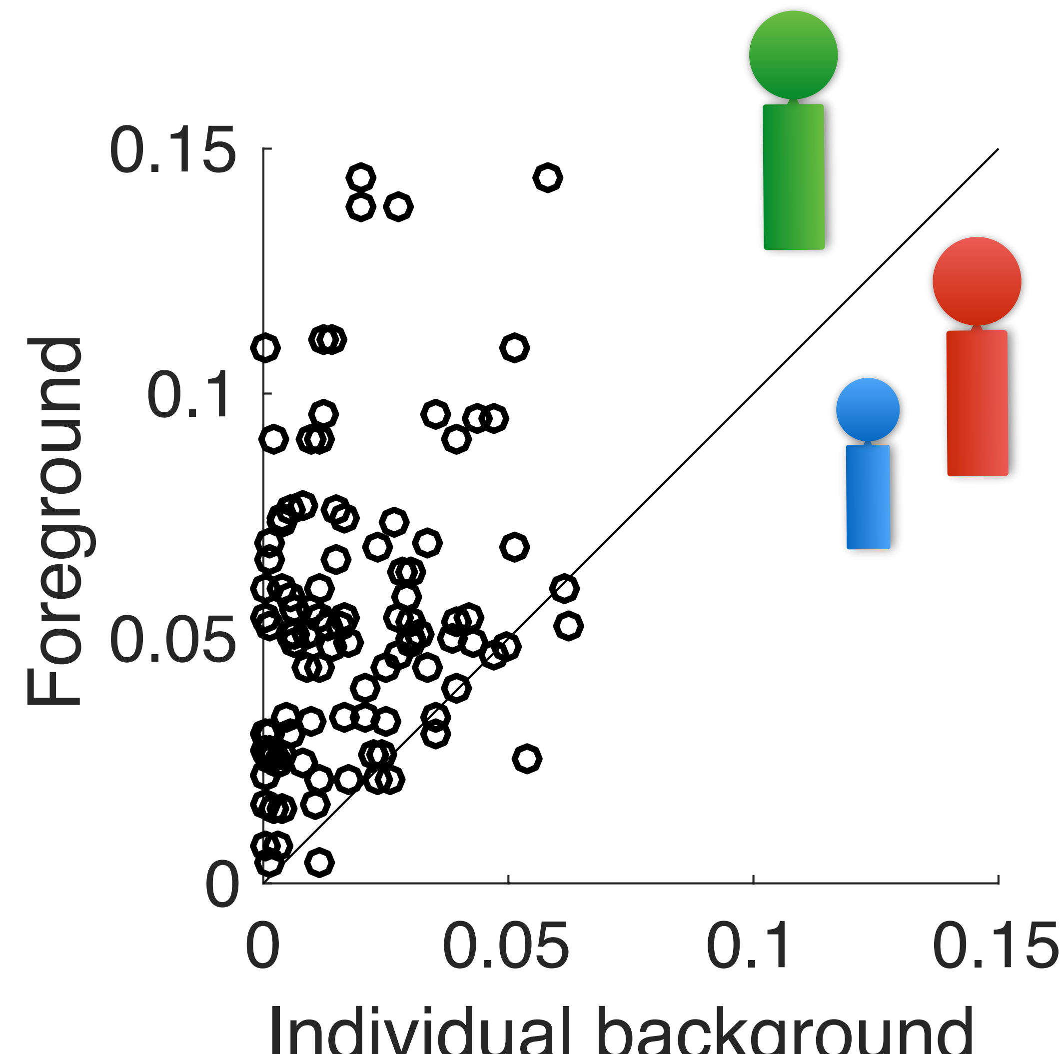
Note: r^2 Scatterplot (not r)

Puvvada & Simon, bioRχiv (2017)

Integration Window over Late Times Only

Late Cortical Reconstruction

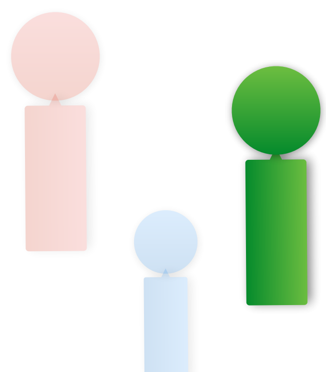




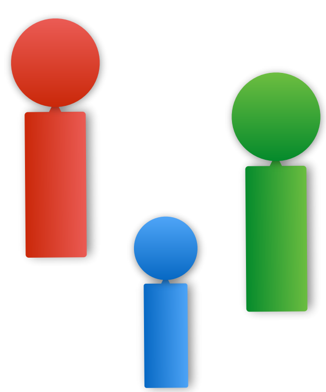
PT represents attended speech with much greater fidelity than unattended

Puvvada & Simon, bioRχiv (2017)

Early Cortical Reconstruction



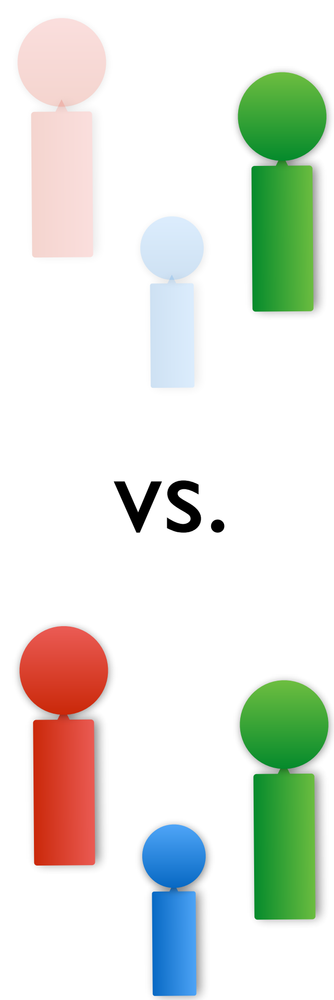
vs.

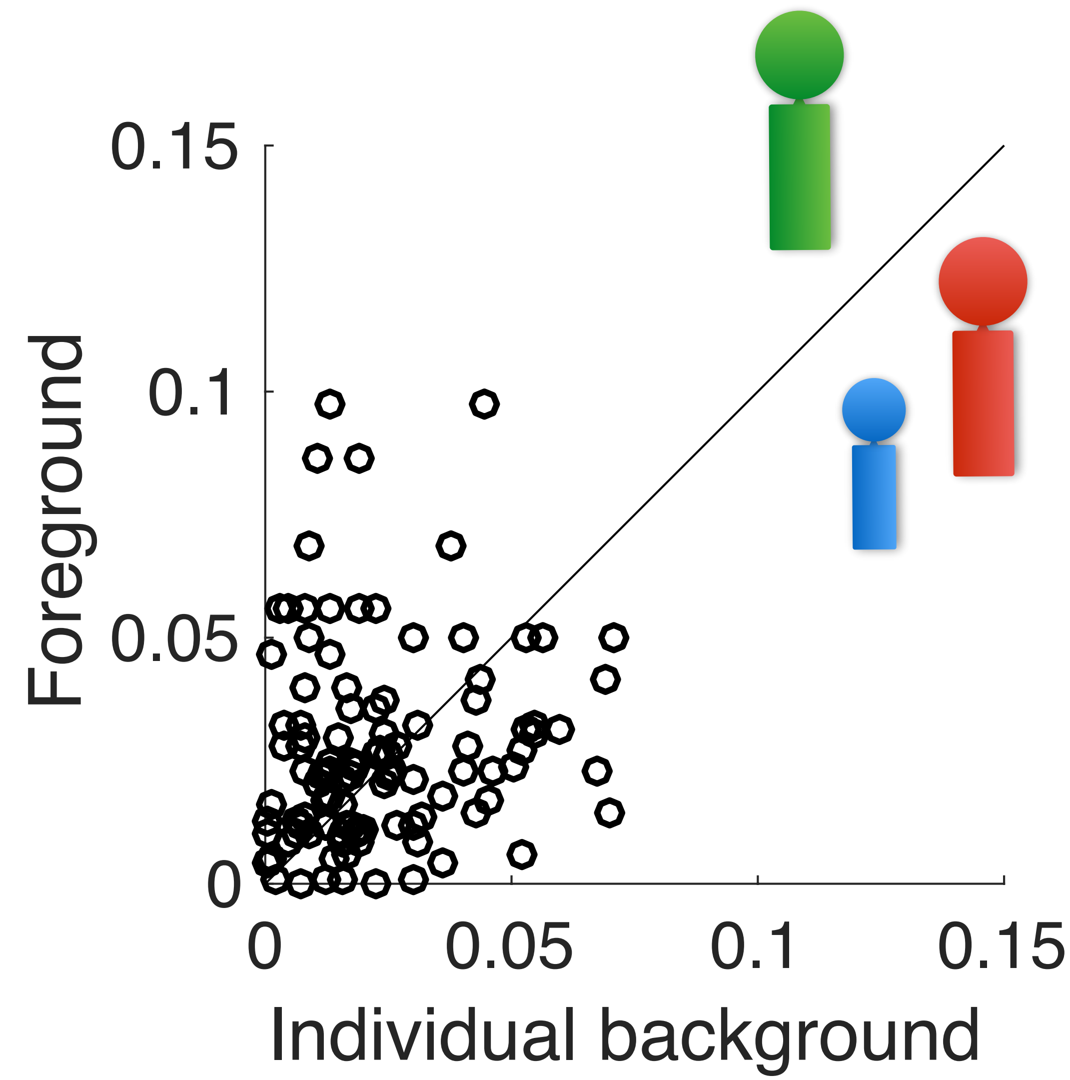


Puvvada & Simon, bioRχiv (2017)

Integration Window over Early Times Only

Early Cortical Reconstruction

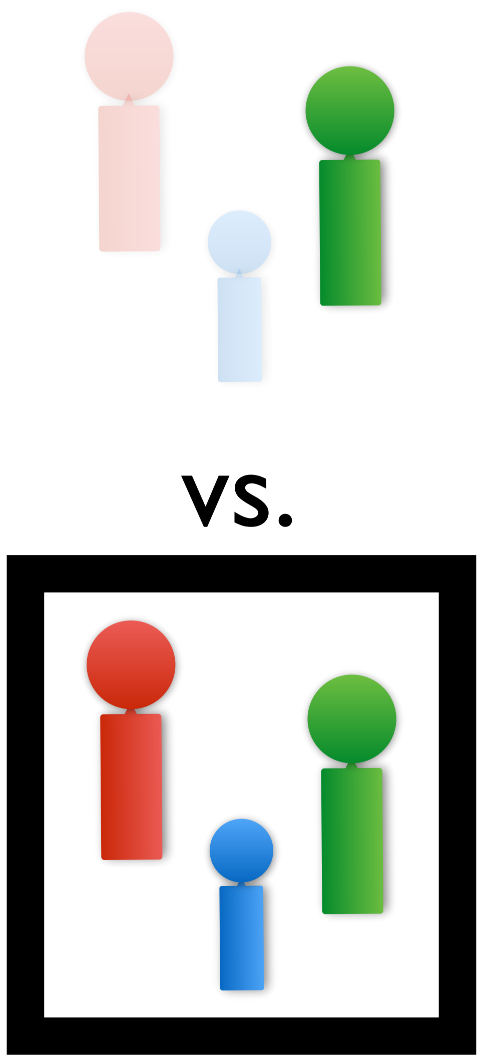


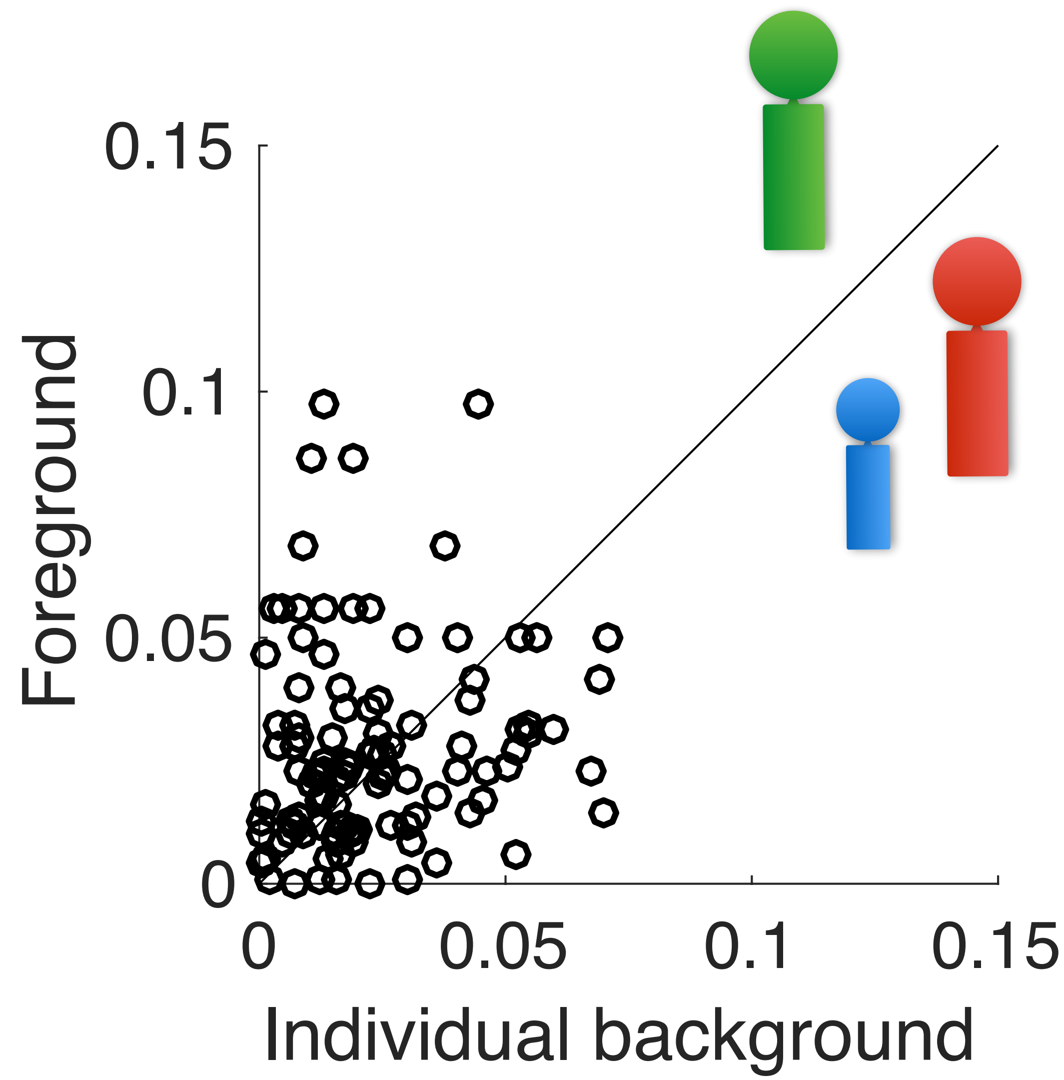


Puvvada & Simon, bioRχiv (2017)

Integration Window over Early Times Only

Early Cortical Reconstruction

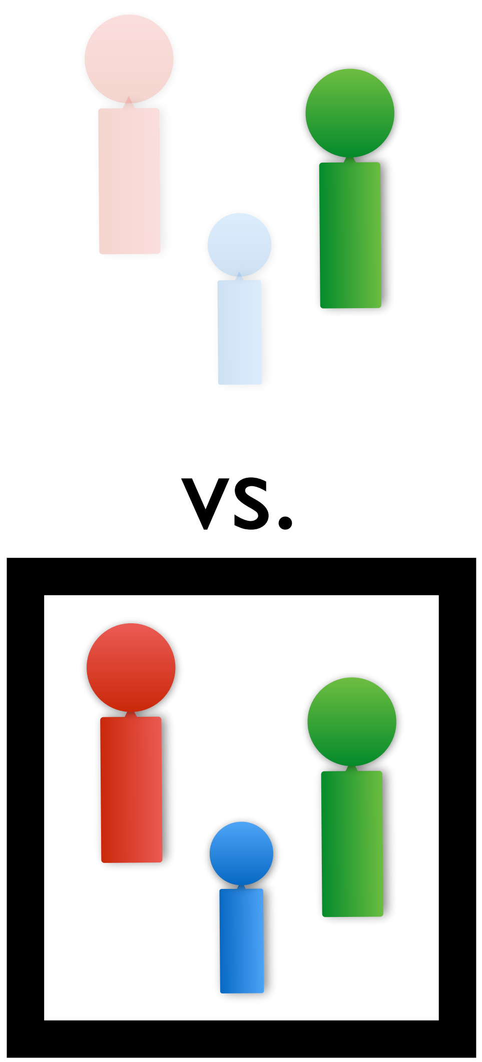


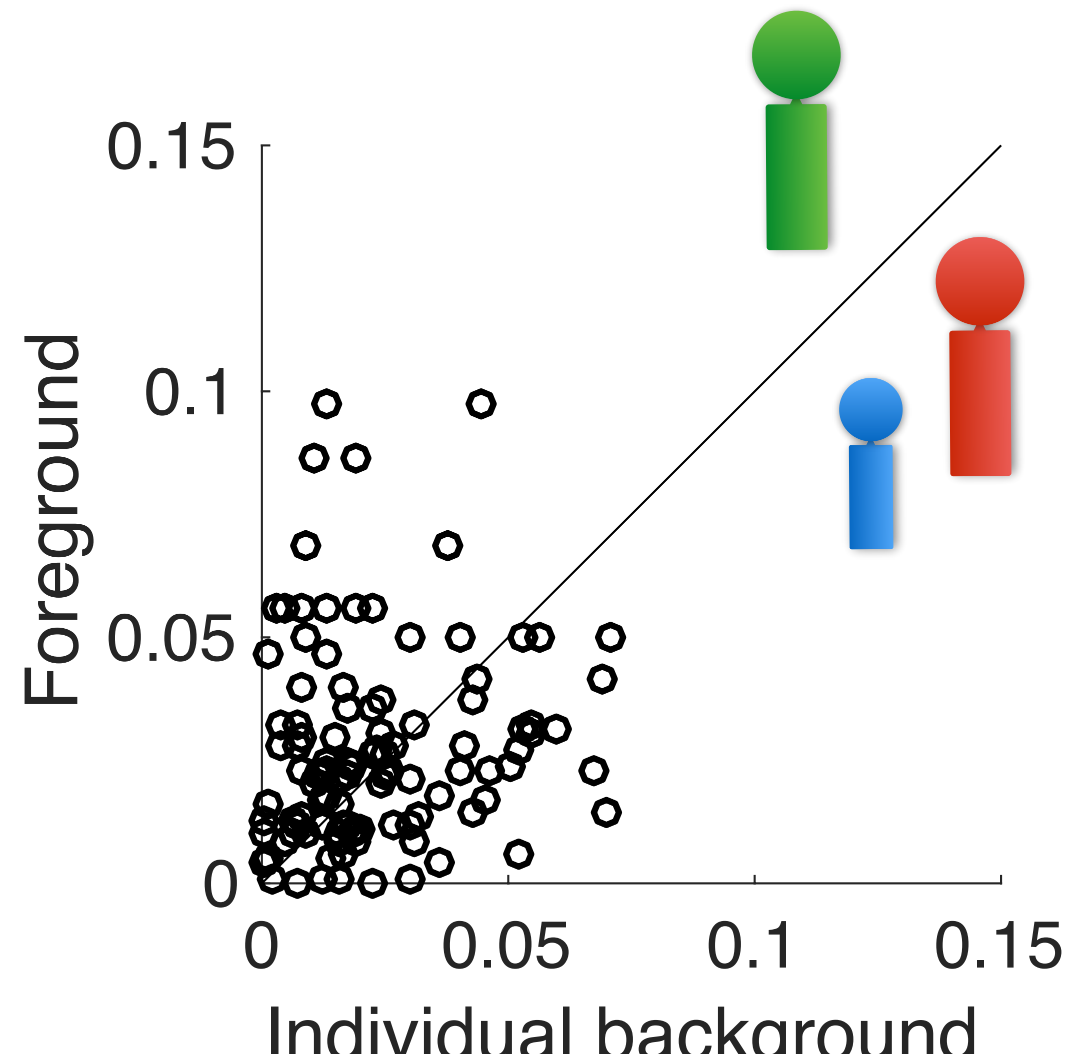


Puvvada & Simon, bioRχiv (2017)

Integration Window over Early Times Only

Early Cortical Reconstruction



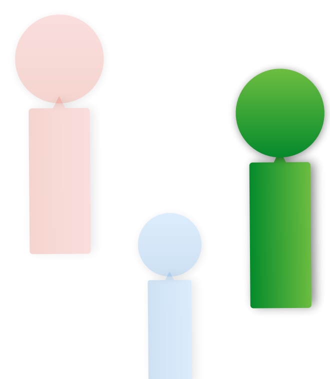


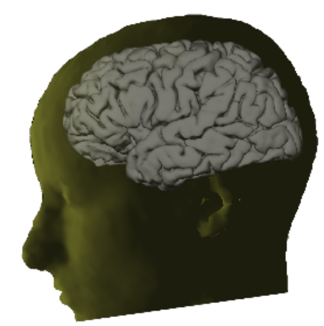
HG represents attended and unattended speech with almost equal fidelity

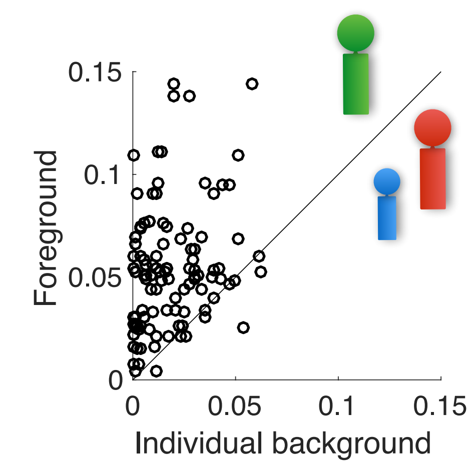
Puvvada & Simon, bioRχiv (2017)

Where in Cortex is there a Segregated Foreground?

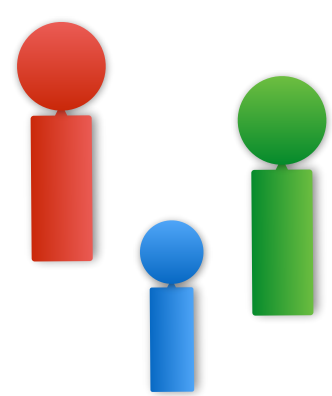
Planum Temporale

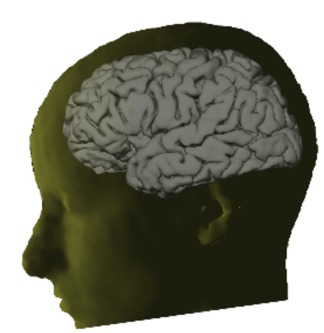


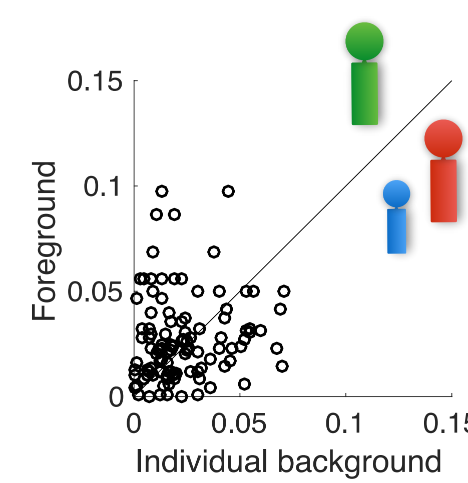




but not Heschl's Gyrus

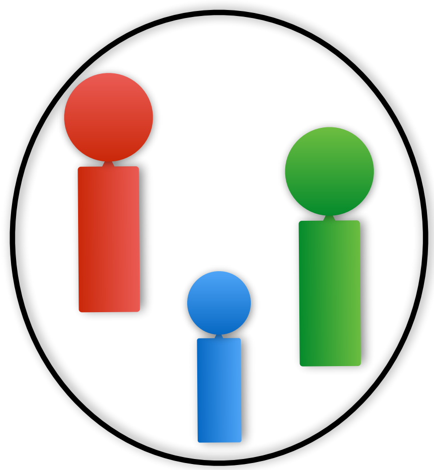


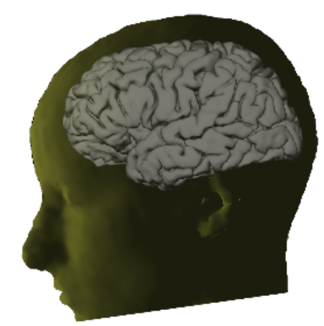




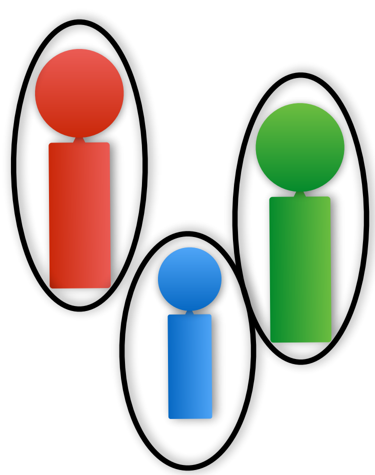
Puvvada & Simon, bioRχiv (2017)

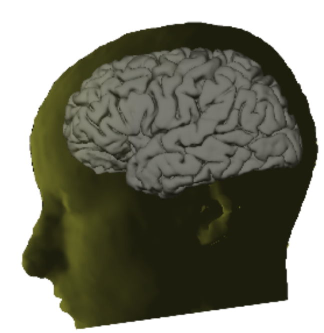
Early Entire Acoustic Scene vs. Individual Streams





vs.



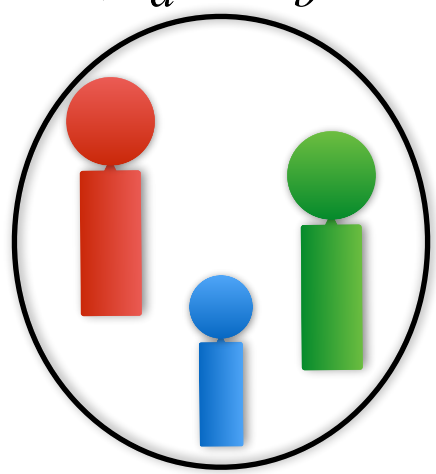


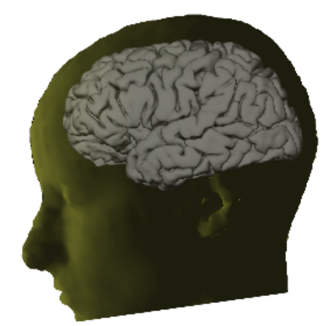
Puvvada & Simon, bioRχiv (2017)

Integration Window over Early Times Only

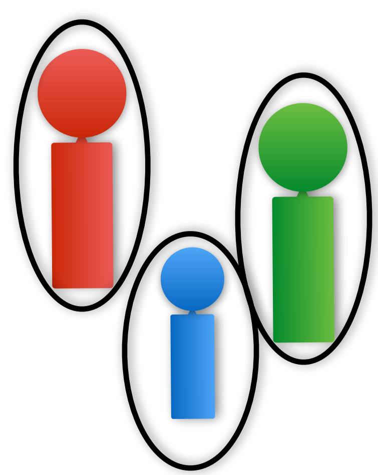
Early Entire Acoustic Scene vs. Individual Streams

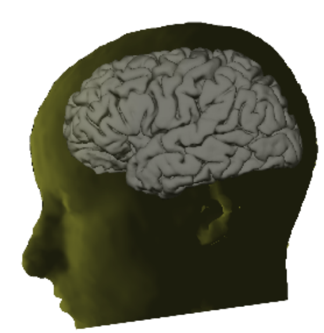
$Env(S_a + S_b + S_c)$





vs.



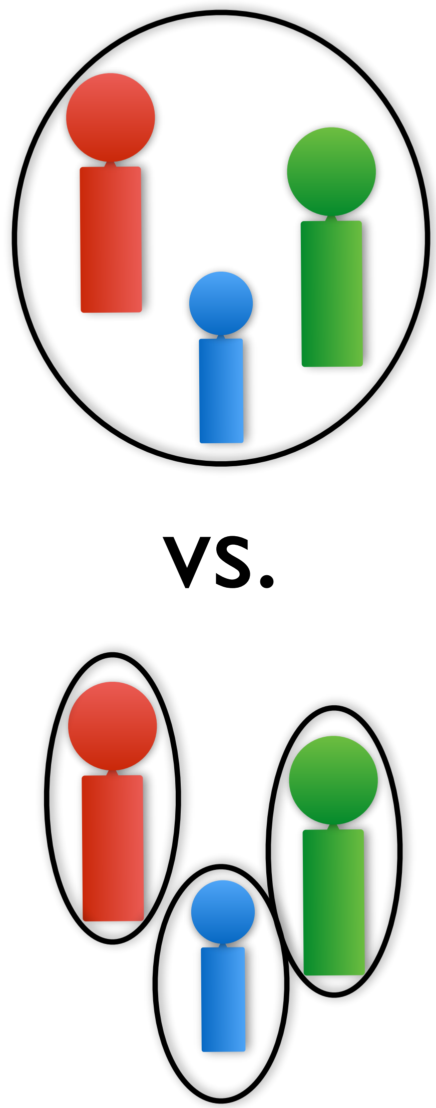


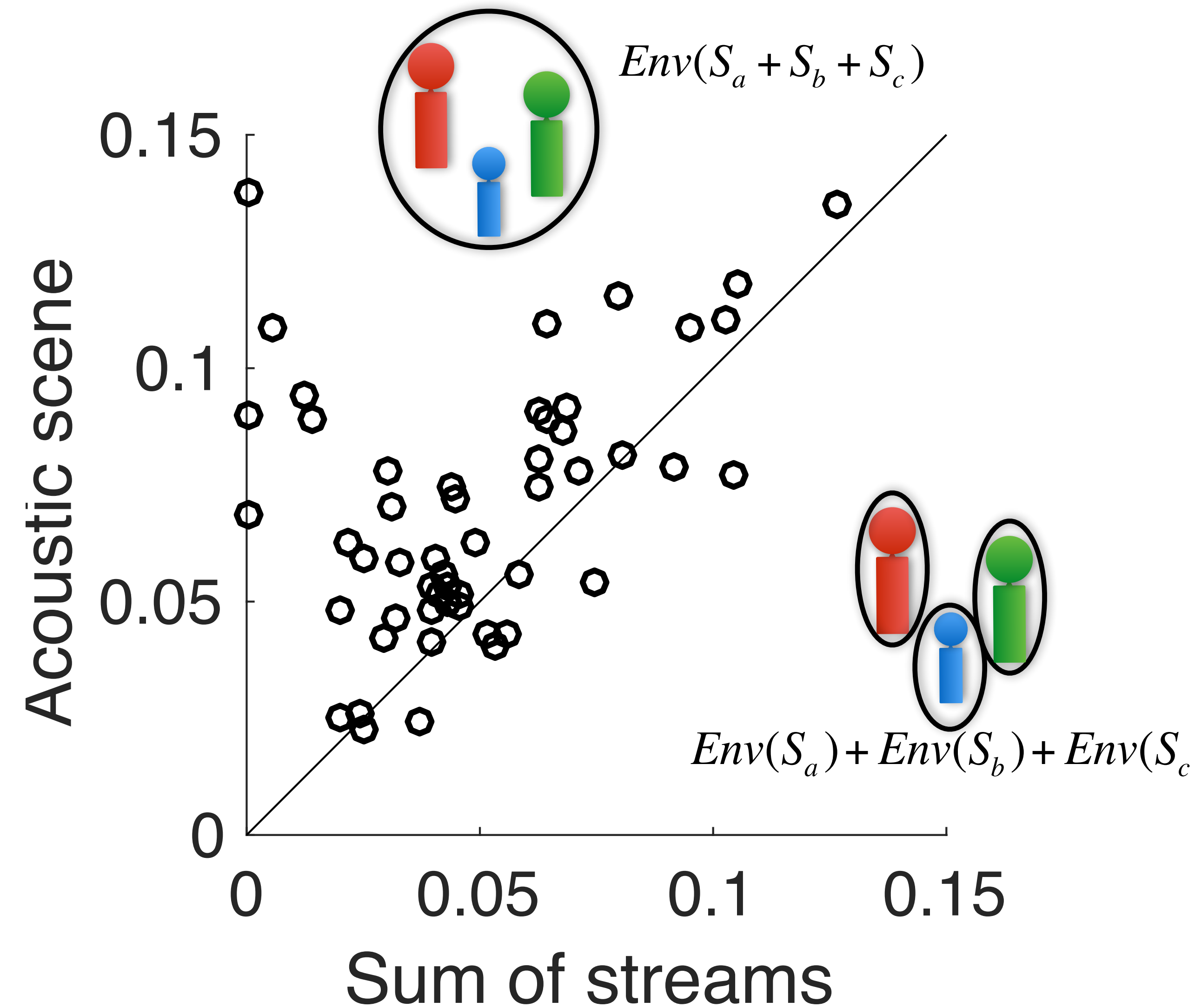
$Env(S_a) + Env(S_b) + Env(S_c)$

Puvvada & Simon, bioRχiv (2017)

Integration Window over Early Times Only

Early Entire Acoustic Scene vs. Individual Streams

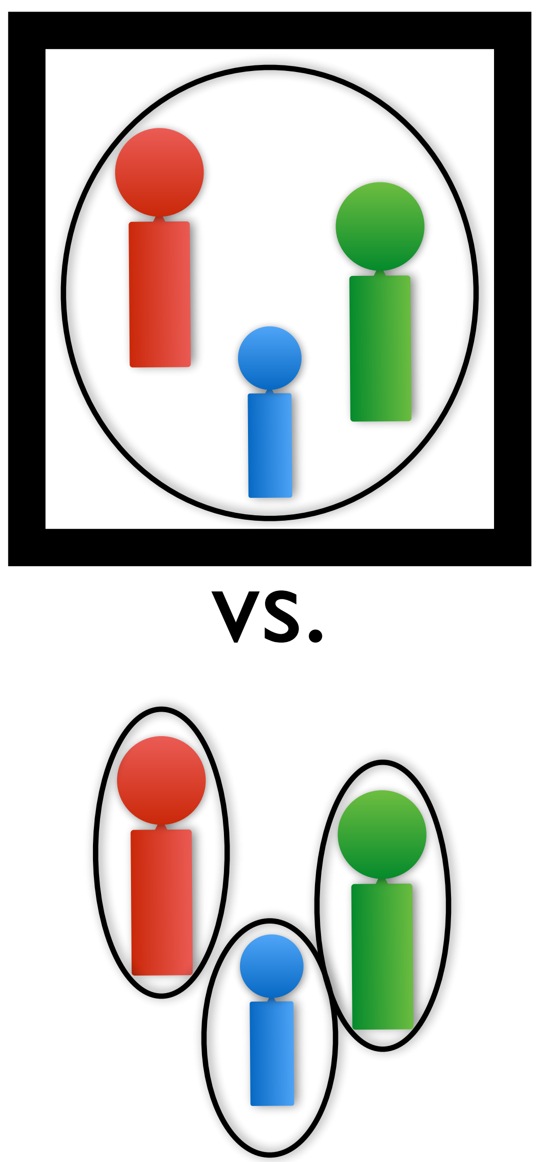


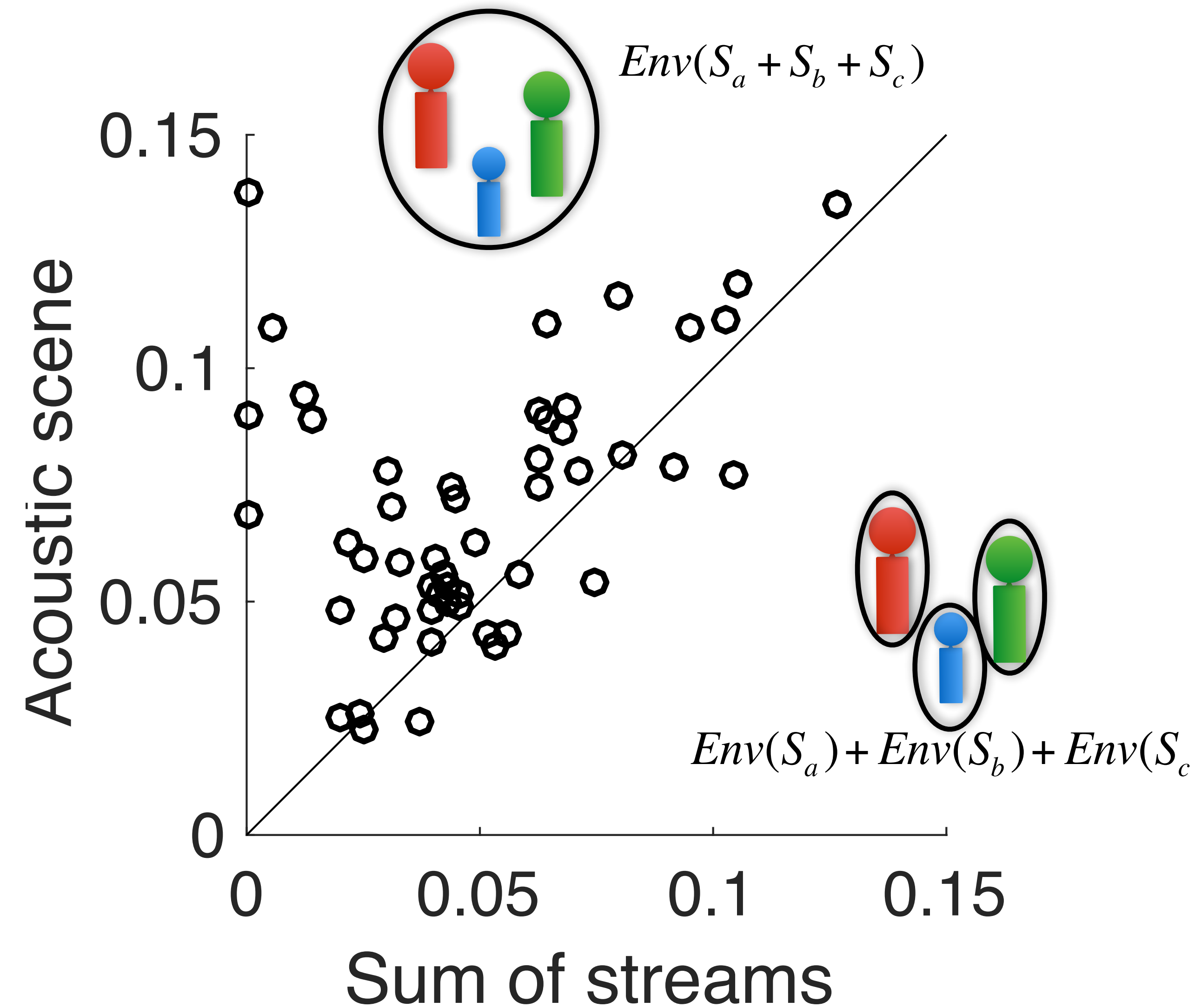


Puvvada & Simon, bioRχiv (2017)

Integration Window over Early Times Only

Early Entire Acoustic Scene vs. Individual Streams

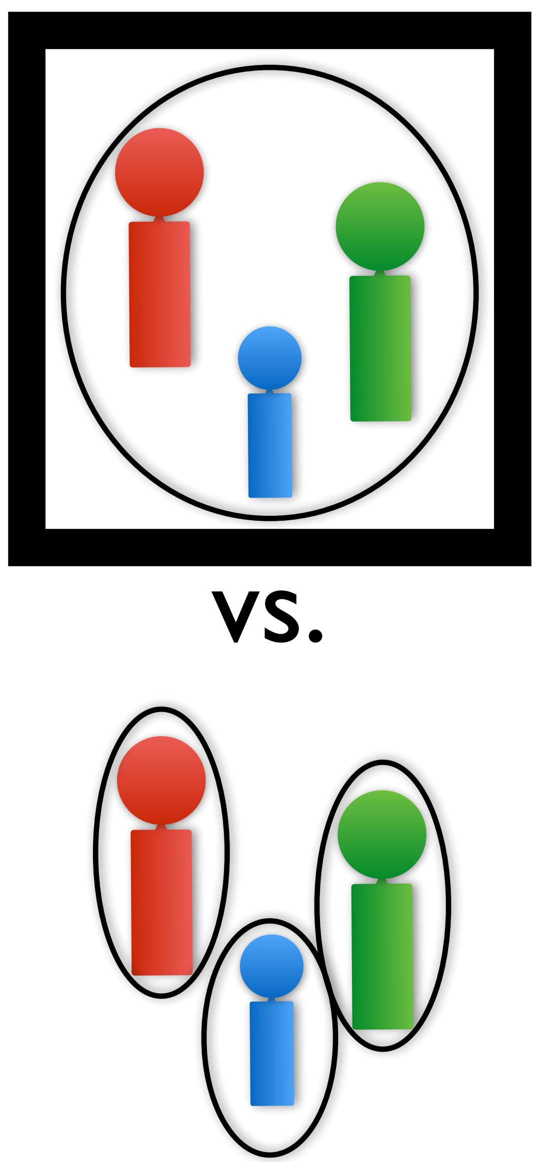


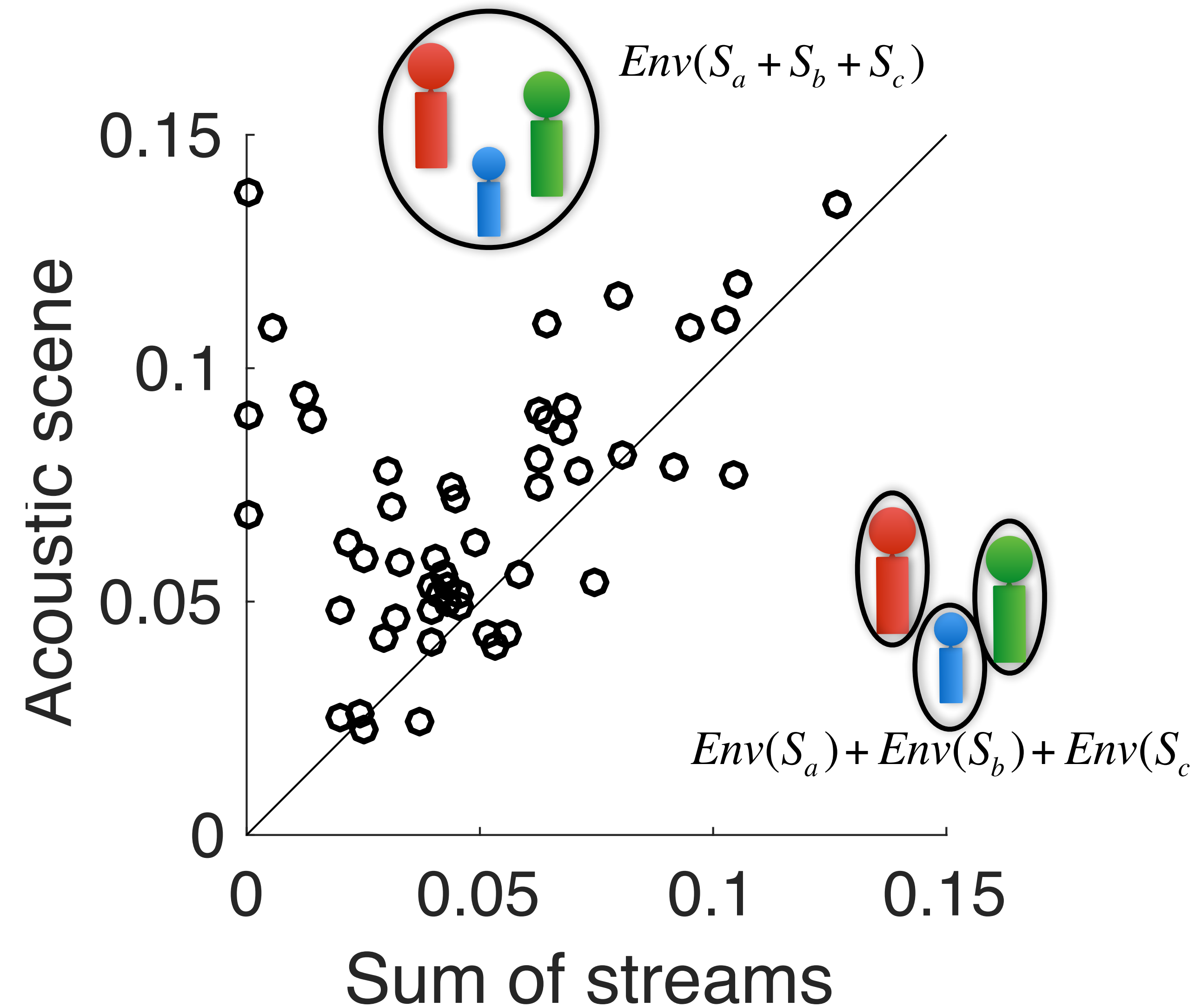


Puvvada & Simon, bioRχiv (2017)

Integration Window over Early Times Only

Early Entire Acoustic Scene vs. Individual Streams





HG Represents the (holistic) Acoustic Scene, not Individual Streams

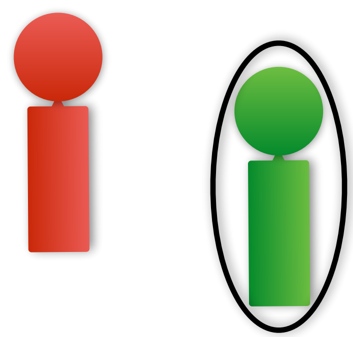
Puvvada & Simon, bioRχiv (2017)

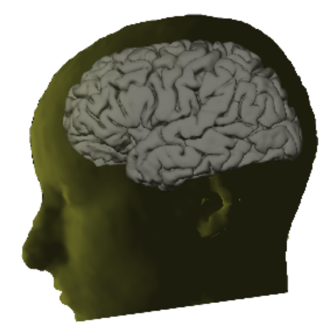
Foreground vs. Background

Puvvada & Simon, bioRχiv (2017)

Integration Window over Late Times Only

Foreground vs. Background





Puvvada & Simon, bioRχiv (2017)

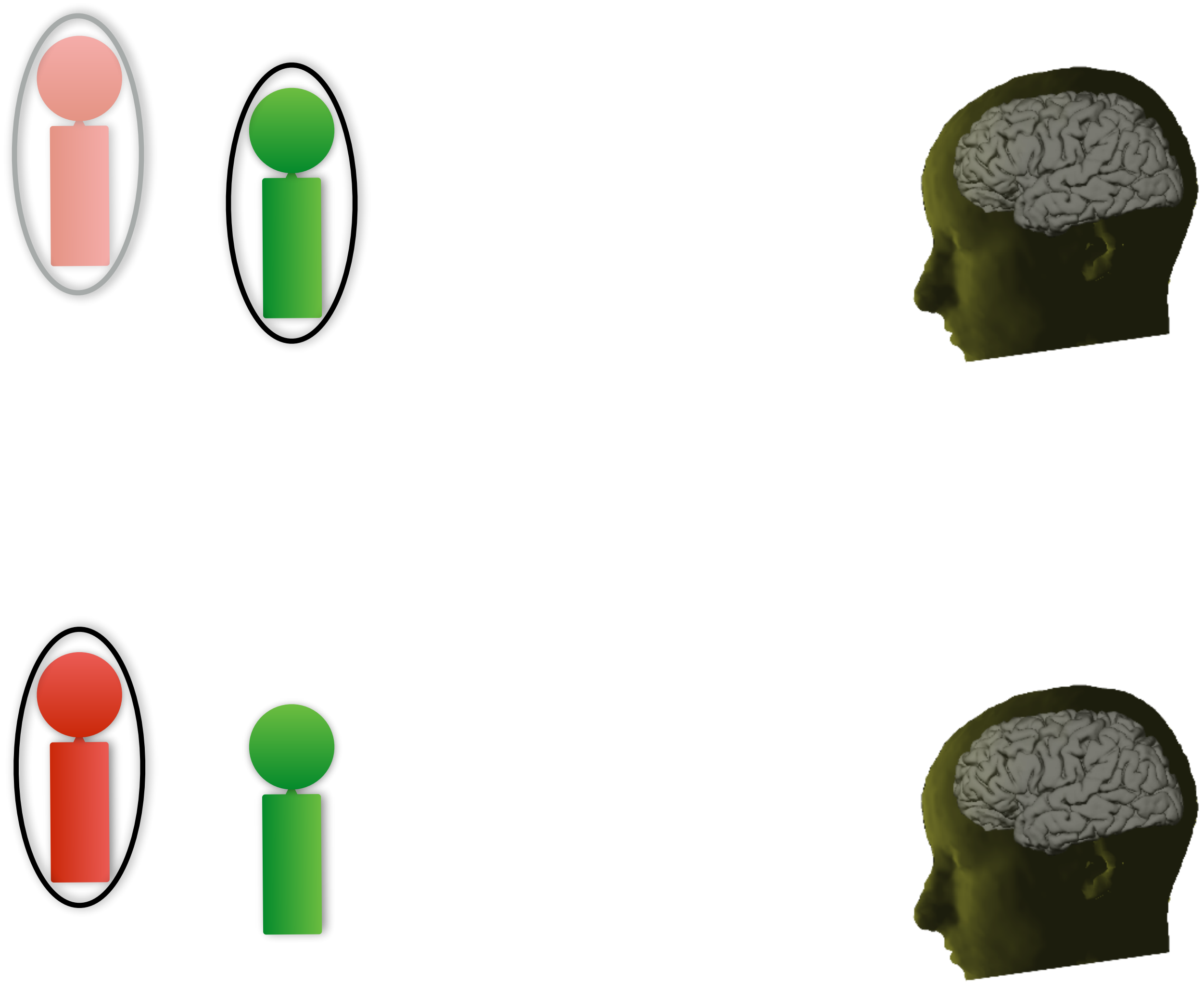
Integration Window over Late Times Only

Foreground vs. Background

Puvvada & Simon, bioRχiv (2017)

Integration Window over Late Times Only

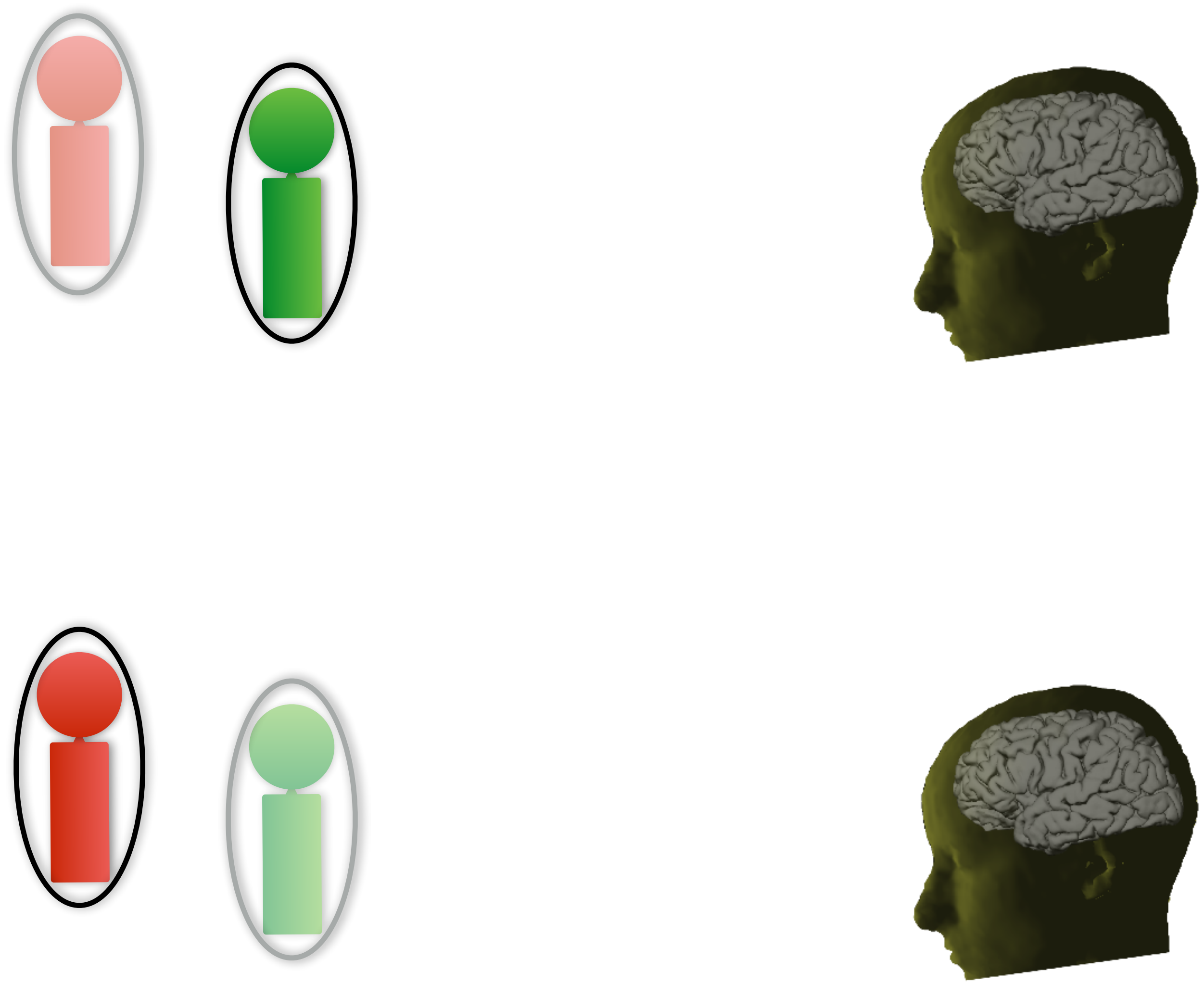
Foreground vs. Background



Puvvada & Simon, bioRχiv (2017)

Integration Window over Late Times Only

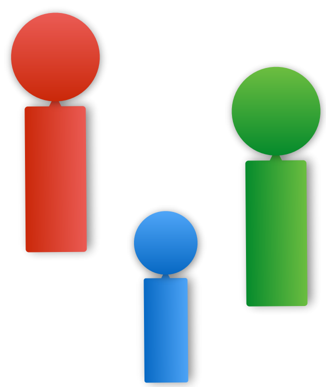
Foreground vs. Background

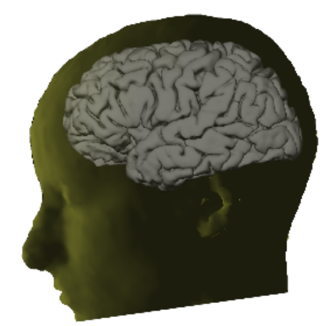


Puvvada & Simon, bioRχiv (2017)

Integration Window over Late Times Only

Foreground vs. Background

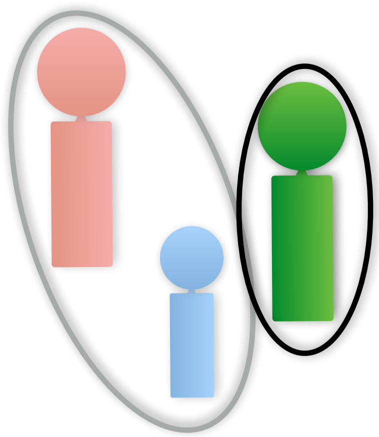


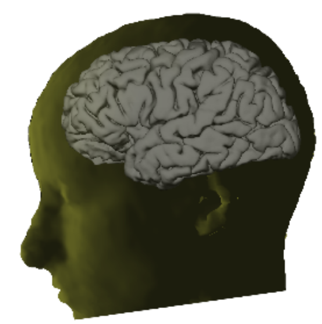


Puvvada & Simon, bioRχiv (2017)

Integration Window over Late Times Only

Foreground vs. Background





Puvvada & Simon, bioRχiv (2017)

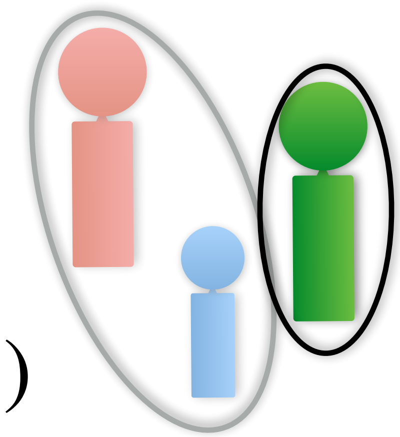
Integration Window over Late Times Only

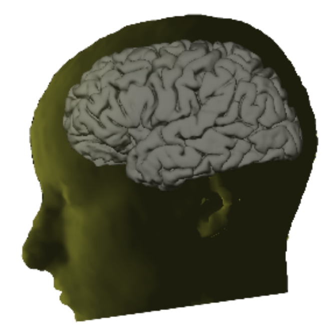
Foreground vs. Background

Puvvada & Simon, bioRχiv (2017)

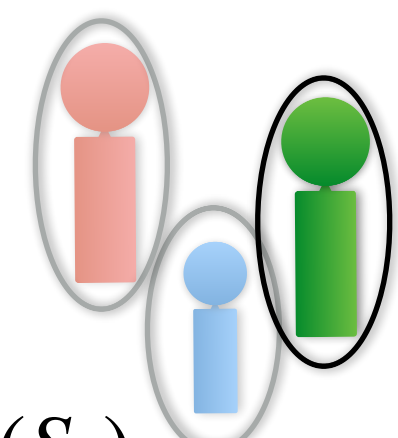
Integration Window over Late Times Only

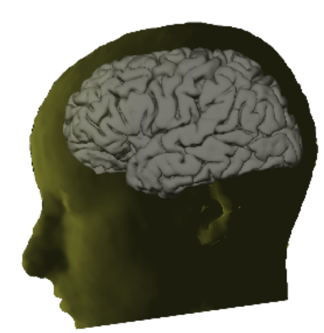
Foreground vs. Background





$Env(S_b + S_c)$



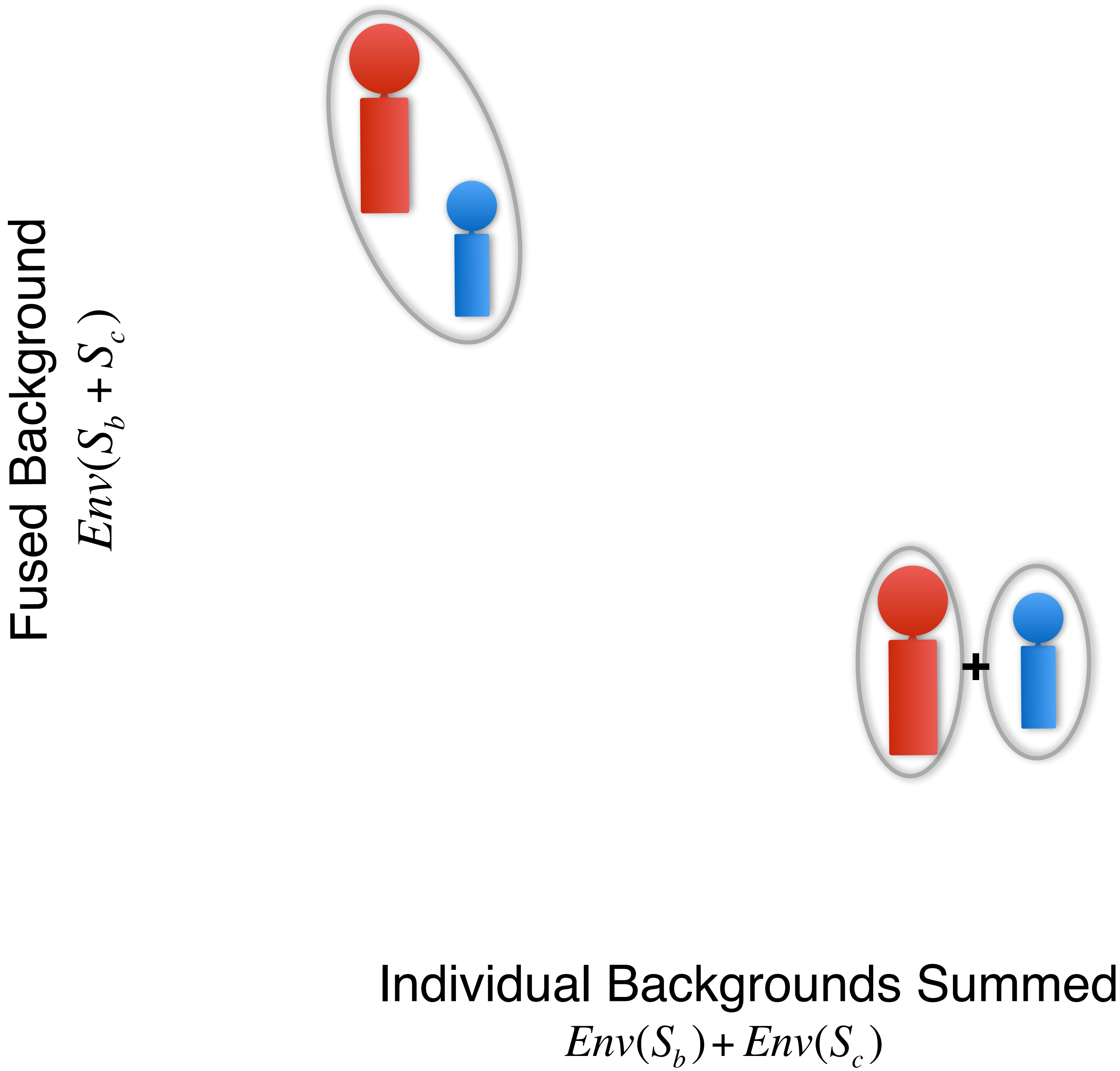


$Env(S_b) + Env(S_c)$

Puvvada & Simon, bioRχiv (2017)

Integration Window over Late Times Only

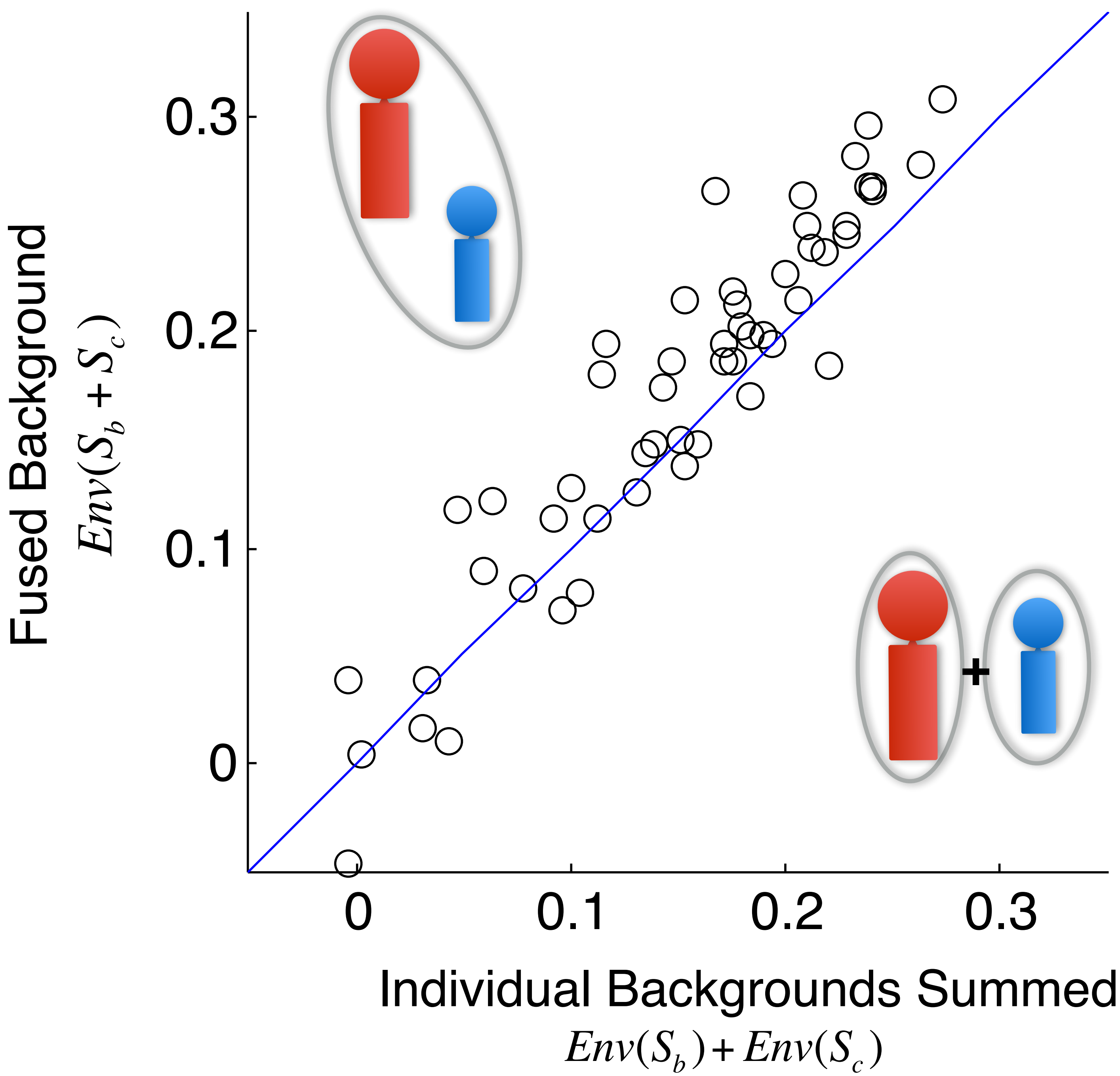
Background vs. Backgrounds



Puvvada & Simon, bioRχiv (2017)

Integration Window over Late Times Only

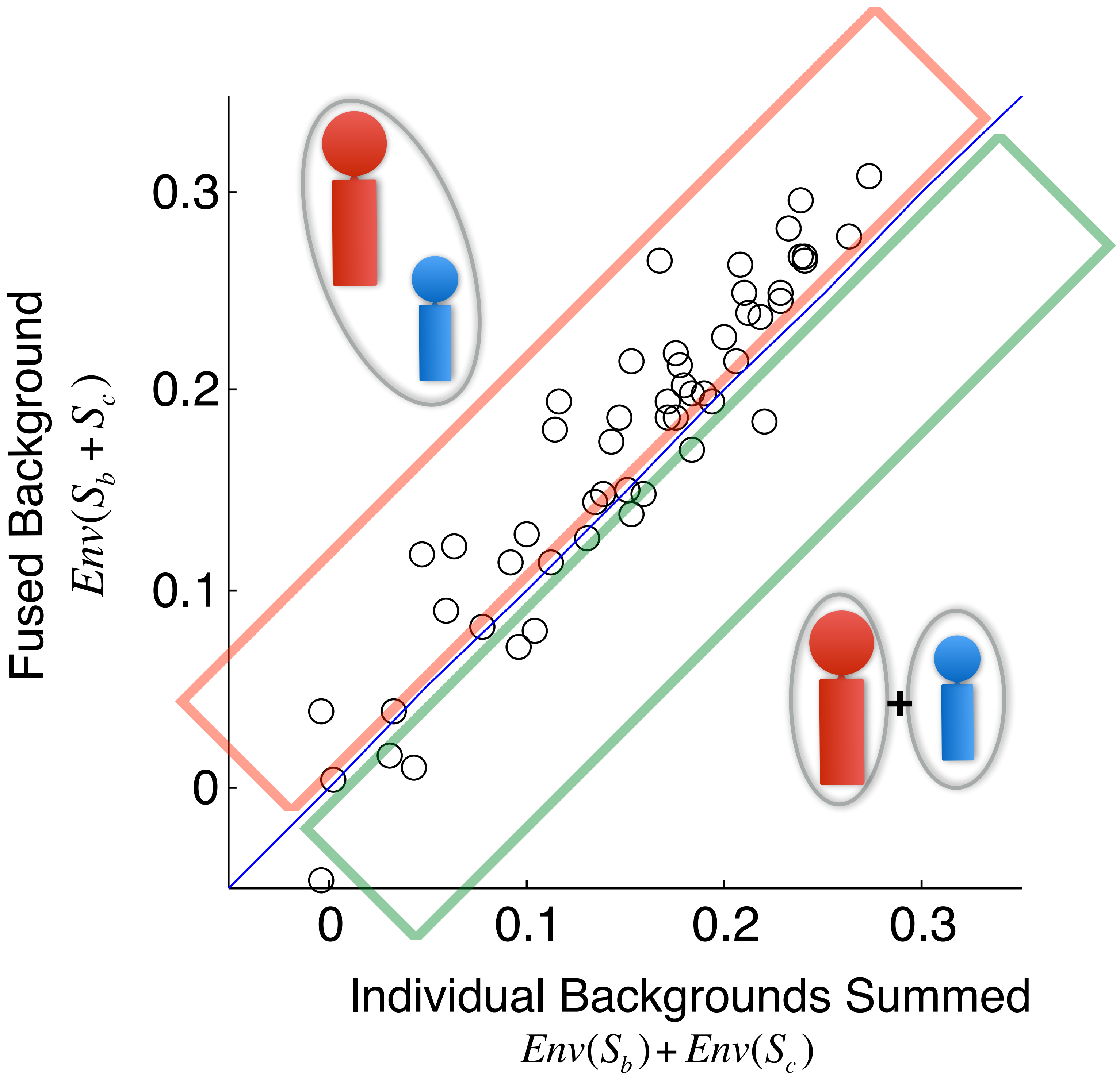
Background vs. Backgrounds



Puvvada & Simon, bioRχiv (2017)

Integration Window over Late Times Only

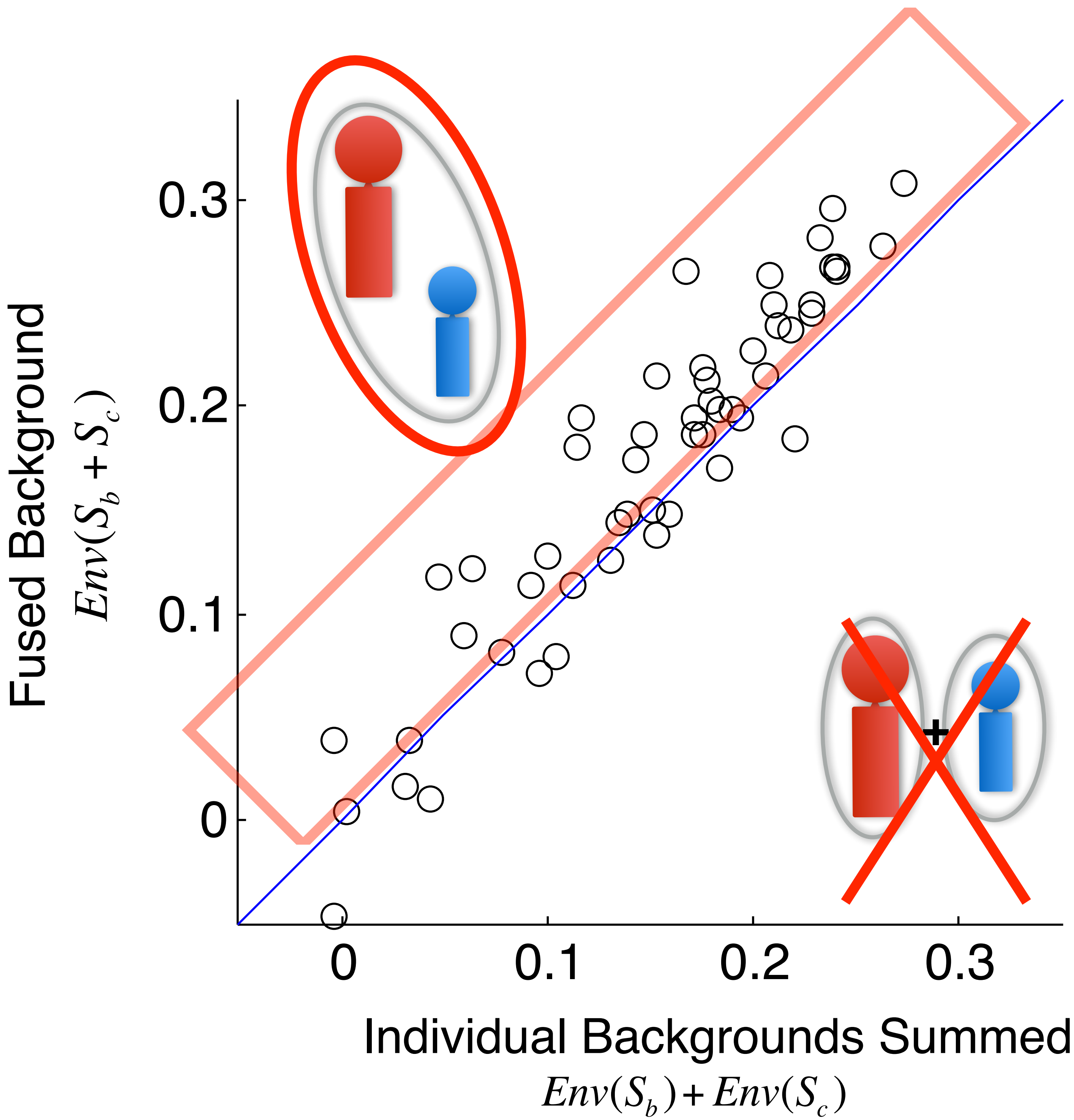
Background vs. Backgrounds



Puvvada & Simon, bioRχiv (2017)

Integration Window over Late Times Only

Background vs. Backgrounds



PT represents a fused background with much better fidelity than individual backgrounds

Puvvada & Simon, bioRχiv (2017)

Forward Model?

Current Competing Speaker TRF model:

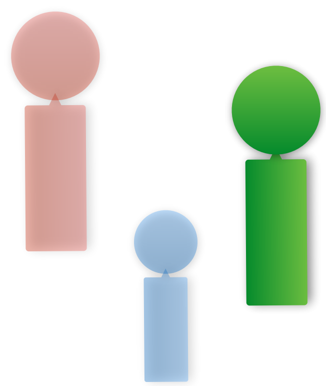
$$r(t) = \sum_{\tau} TRF_a(t-\tau)S_a(\tau) + \sum_{\tau} TRF_b(t-\tau)S_b(\tau) + \sum_{\tau} TRF_c(t-\tau)S_c(\tau) + \varepsilon(t)$$

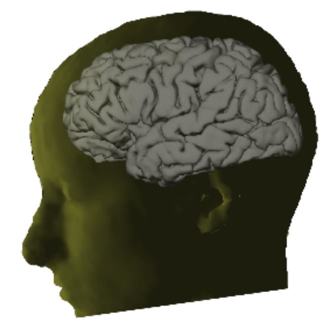
Puvvada & Simon, bioRχiv (2017)

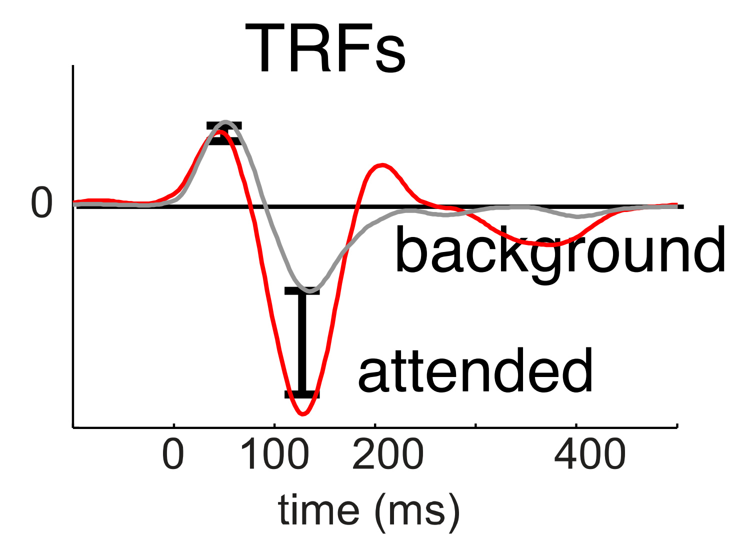
Forward Model?

Current Competing Speaker TRF model:

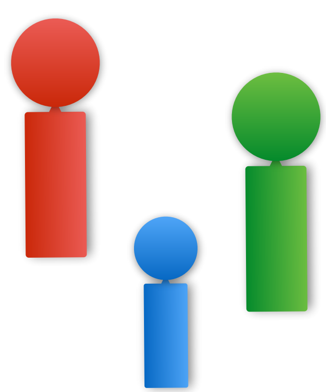
$$r(t) = \sum_{\tau} TRF_a(t-\tau)S_a(\tau) + \sum_{\tau} TRF_b(t-\tau)S_b(\tau) + \sum_{\tau} TRF_c(t-\tau)S_c(\tau) + \varepsilon(t)$$

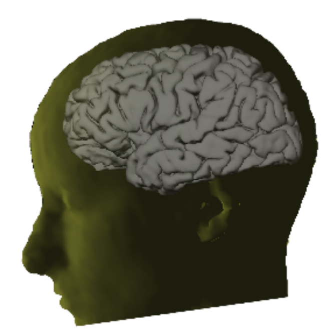






or





or…

Puvvada & Simon, bioRχiv (2017)

Better Forward Model?

$$r(t) = \sum_{\tau=0}^{\tau=\tau_1} TRF_{Scene}(t-\tau) S_{Scene}(\tau) +$$

$$+ \sum_{\tau=\tau_1}^{\tau=\tau_2} TRF_{Foregound}(t-\tau) S_{Foregound}(\tau)$$

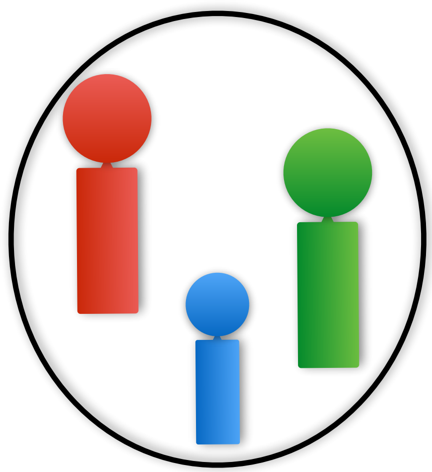
$$+ \sum_{\tau=\tau_1}^{\tau=\tau_2} TRF_{Background}(t-\tau) S_{Background}(\tau)$$

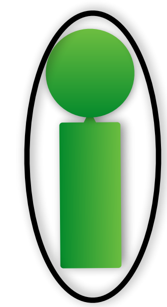
$$+\varepsilon(t)$$

Puvvada & Simon, bioRχiv (2017)

Better Forward Model?

$$r(t) = \sum_{\tau=0}^{\tau=\tau_1} TRF_{Scene}(t-\tau)S_{Scene}(\tau) +$$

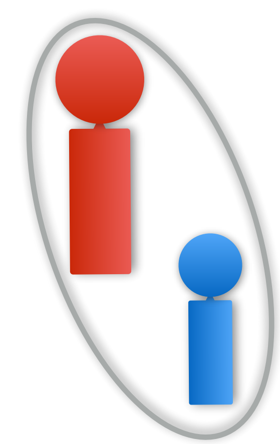




$$+\sum_{\tau=\tau_1}^{\tau=\tau_2} TRF_{Foregound}(t-\tau)S_{Foregound}(\tau)$$

$$+\sum_{\tau=\tau_1}^{\tau=\tau_2} TRF_{Background}(t-\tau)S_{Background}(\tau)$$

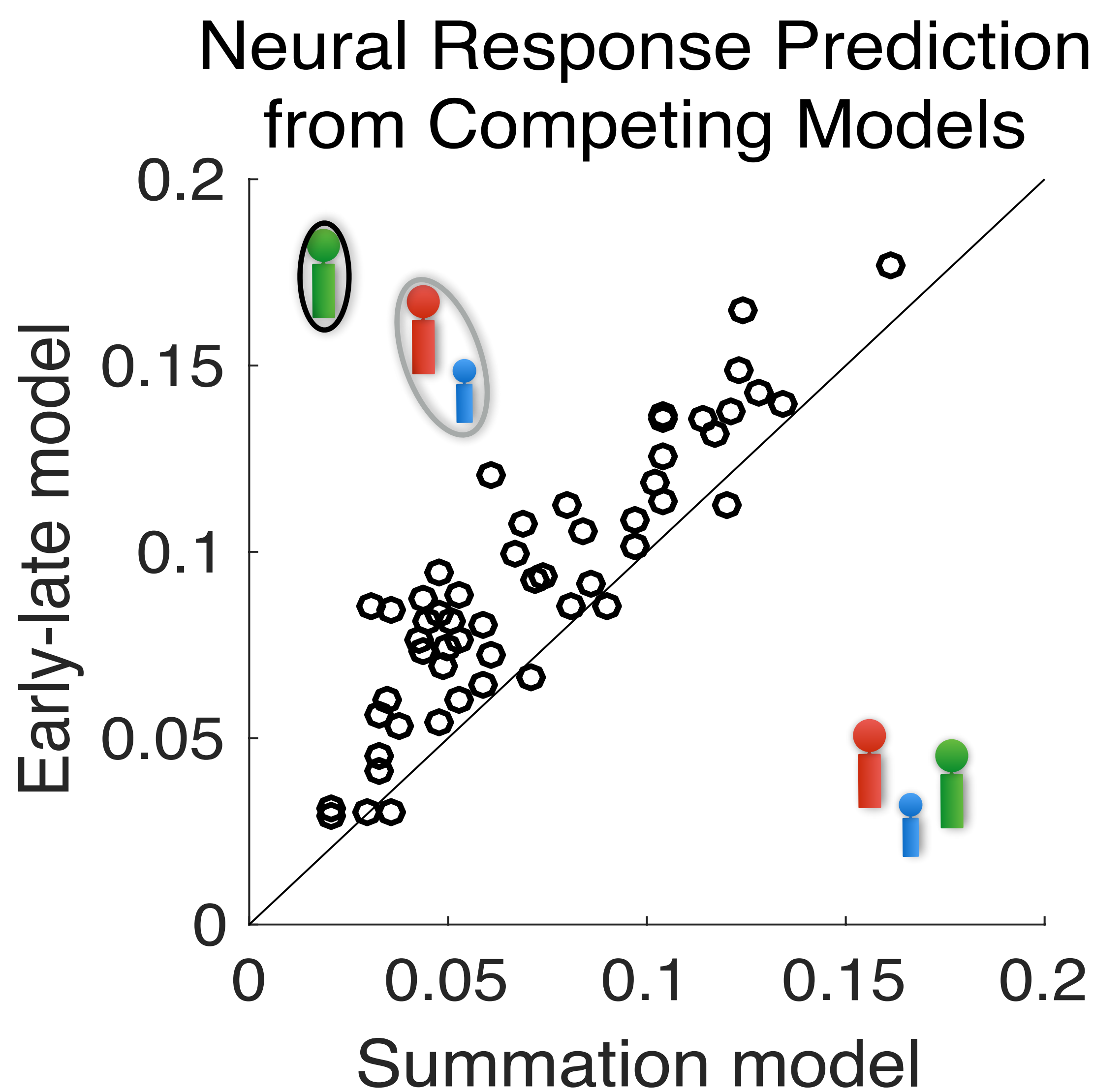
$$+\varepsilon(t)$$



Puvvada & Simon, bioRχiv (2017)

Forward Models Compared

$$r(t) = \sum_{\tau=0}^{\tau=\tau_1} TRF_{Scene}(t-\tau)S_{Scene}(\tau) + \sum_{\tau=\tau_1}^{\tau=\tau_2} TRF_{Foregound}(t-\tau)S_{Foregound}(\tau) + \sum_{\tau=\tau_1}^{\tau=\tau_2} TRF_{Background}(t-\tau)S_{Background}(\tau) + \varepsilon(t)$$

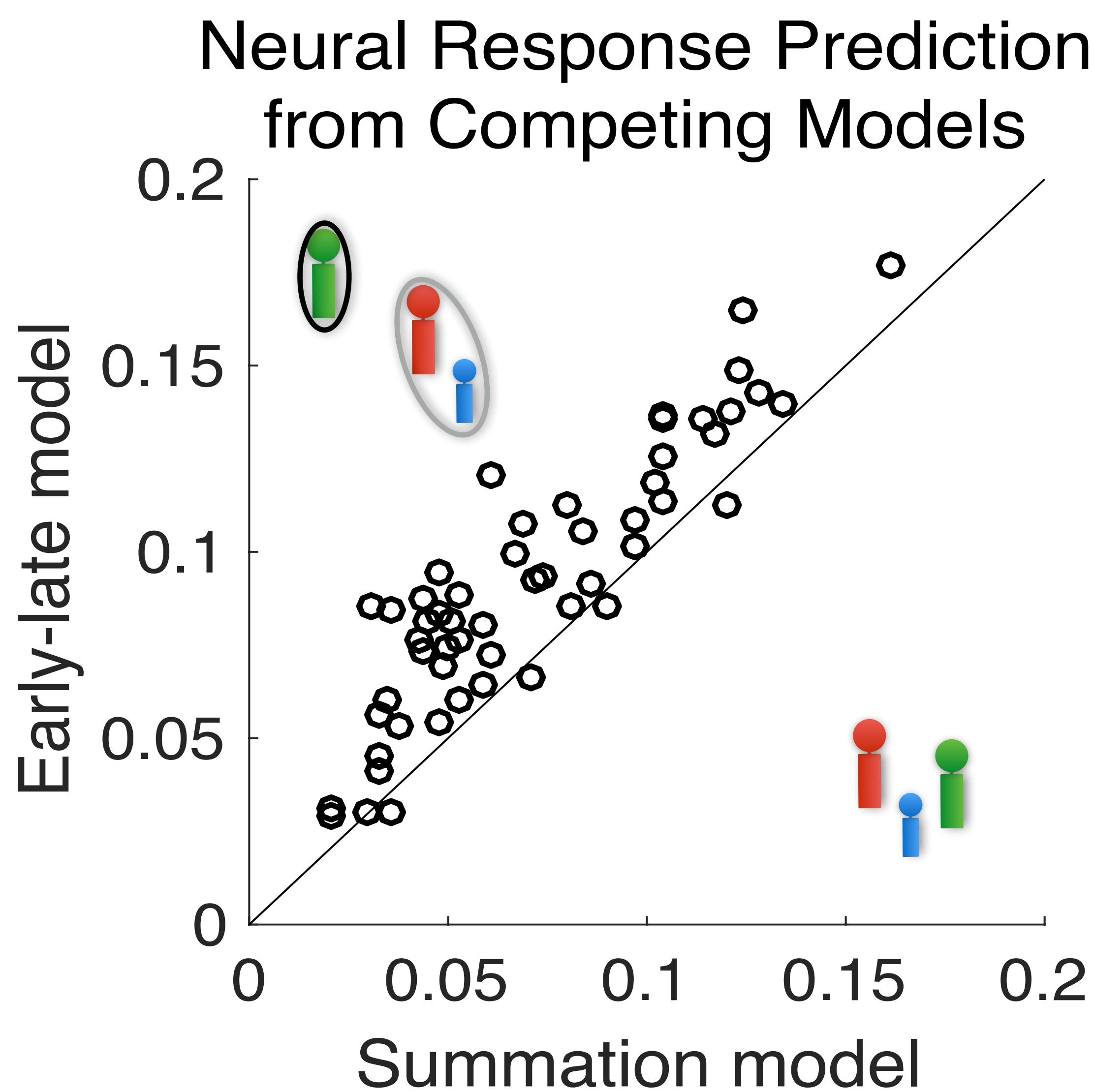


$$r(t) = \sum_{\tau} TRF_a(t-\tau)S_a(\tau) + \sum_{\tau} TRF_b(t-\tau)S_b(\tau) + \sum_{\tau} TRF_c(t-\tau)S_c(\tau) + \varepsilon(t)$$

Puvvada & Simon, bioRχiv (2017)

Forward Models Compared

$$r(t) = \sum_{\tau=0}^{\tau=\tau_1} TRF_{Scene}(t-\tau)S_{Scene}(\tau) + \sum_{\tau=\tau_1}^{\tau=\tau_2} TRF_{Foregound}(t-\tau)S_{Foregound}(\tau) + \sum_{\tau=\tau_1}^{\tau=\tau_2} TRF_{Background}(t-\tau)S_{Background}(\tau) + \varepsilon(t)$$



$$r(t) = \sum_{\tau} TRF_a(t-\tau)S_a(\tau) + \sum_{\tau} TRF_b(t-\tau)S_b(\tau) + \sum_{\tau} TRF_c(t-\tau)S_c(\tau) + \varepsilon(t)$$

Early-late model outperforms naive model

Puvvada & Simon, bioRχiv (2017)

Latencies as Proxy for Cortical Areas

- Using biologically defined integration windows to reconstruct stimulus can distinguish between different representations
  - Early areas (HG) are best at reconstructing the entire acoustic sound scene
  - Later areas (PT) are best at reconstructing the foreground stream, with an integrated background
- Modified TRF model performs better than naive