Neural Representations of Speech in Human Auditory Cortex

Jonathan Z. Simon

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

University of Maryland

Acknowledgements

Current Lab Members & Affiliates

Ross Baehr
Christian Brodbeck
Alex Jiao
Joshua Kulasingham
Sina Miran
David Nahmias
Peng Zan

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
Natalia Lapinskaya
Huan Luo
Mahshid Najafi
Alex Presacco
Krishna Puvvada
Jonas Vanthornhout
Ben Walsh
Yadong Wang
Juanjuan Xiang
Jiachen Zhuo

Collaborators

Pamela Abshire
Samira Anderson
Behtash Babadi
Catherine Carr
Monita Chatterjee
Alain de Cheveigné
Stephen David
Didier Depireux
Mounya Elhilali
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

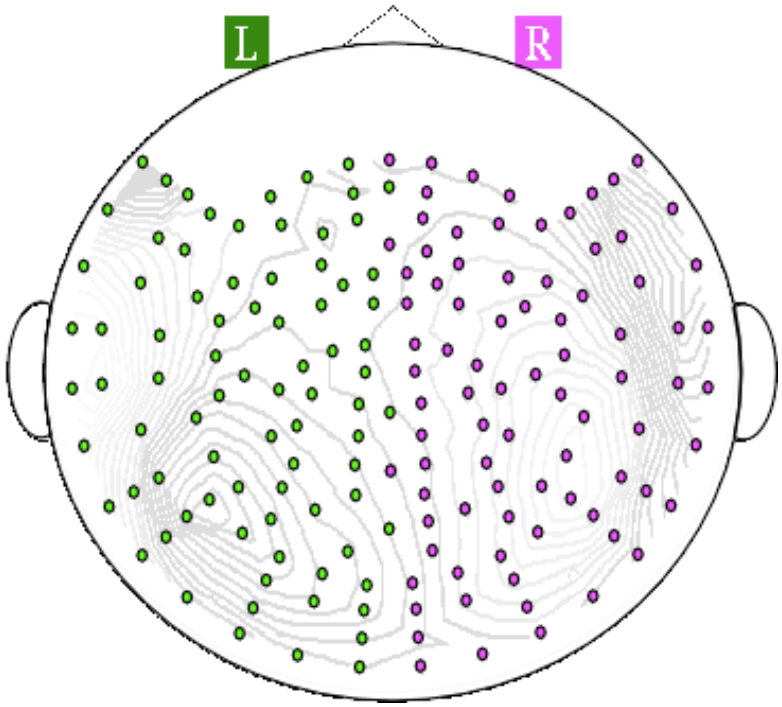
Funding NIH (*NIDCD*, *NIA*, *NIBIB*); NSF; DARPA; USDA; UMD

Outline

- Cortical Representations of Speech (via MEG)
  - *Encoding vs. Decoding*
- Cortical Representations of Speech in Noise
- Cortical Representations of "Cocktail Party" listening
- Recent Studies:
  - *Aging & Cortical Representations of Speech*
  - *Higher Level Interference*
  - *Representations of Internal Speech*

Magnetoencephalography (MEG)

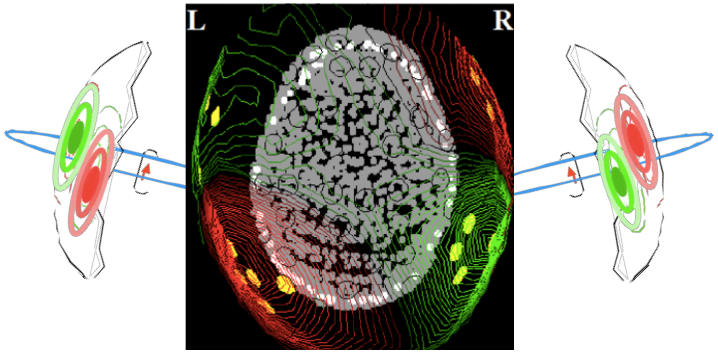
- Non-invasive, Passive, Silent Neural Recordings
- Simultaneous Whole-Head Recording (~200 sensors)
- Sensitivity
  - high: ~100 fT (10^{-13} Tesla)
  - low: ~10^4 – ~10^6 neurons
- Temporal Resolution: ~1 ms
- Spatial Resolution
  - coarse: ~1 cm
  - ambiguous

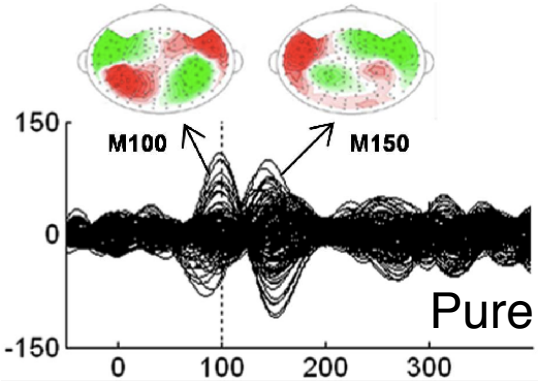


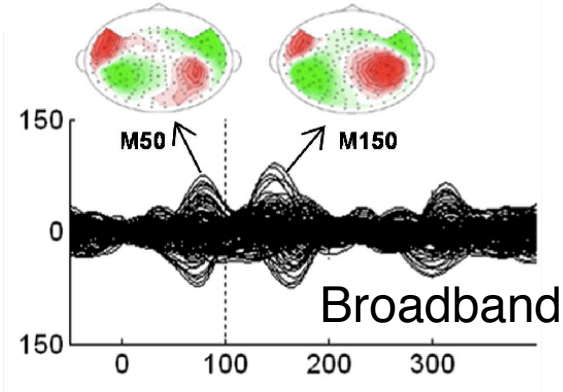
Time Course of MEG Responses

Time Locked Auditory Responses

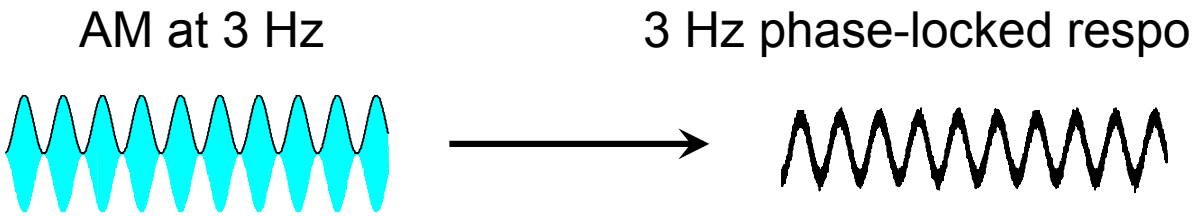
- MEG Response Patterns Time-Locked to Stimulus Events
- Robust
- Strongly Lateralized
- Cortical Origin Only



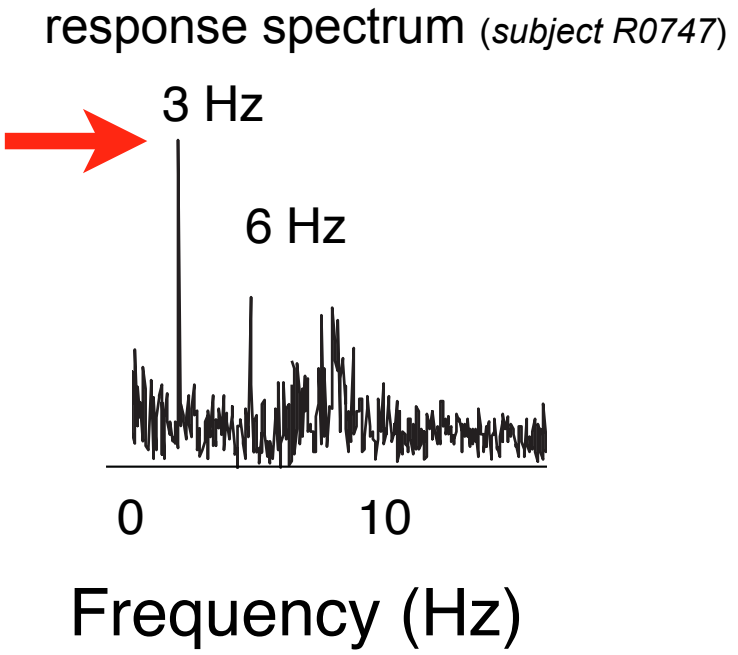




MEG Phase-Locked Responses to Slow Acoustic Modulations

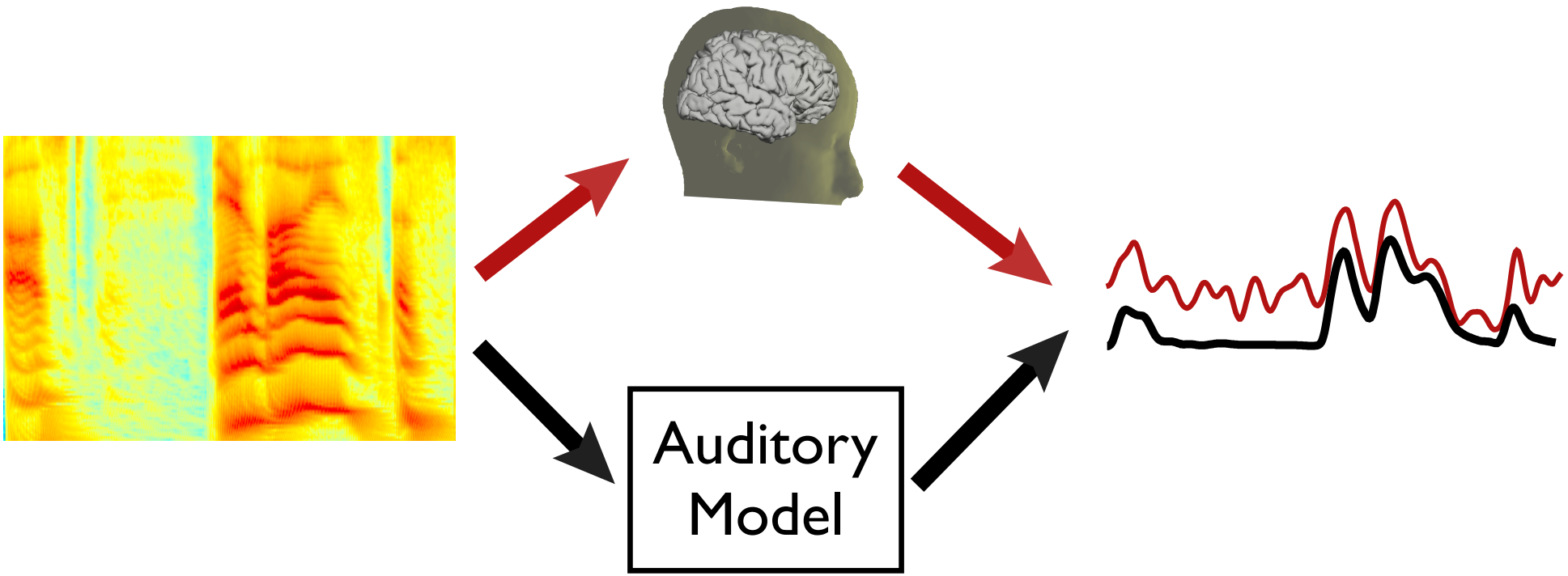


MEG activity is phase-locked to temporal modulations of sound

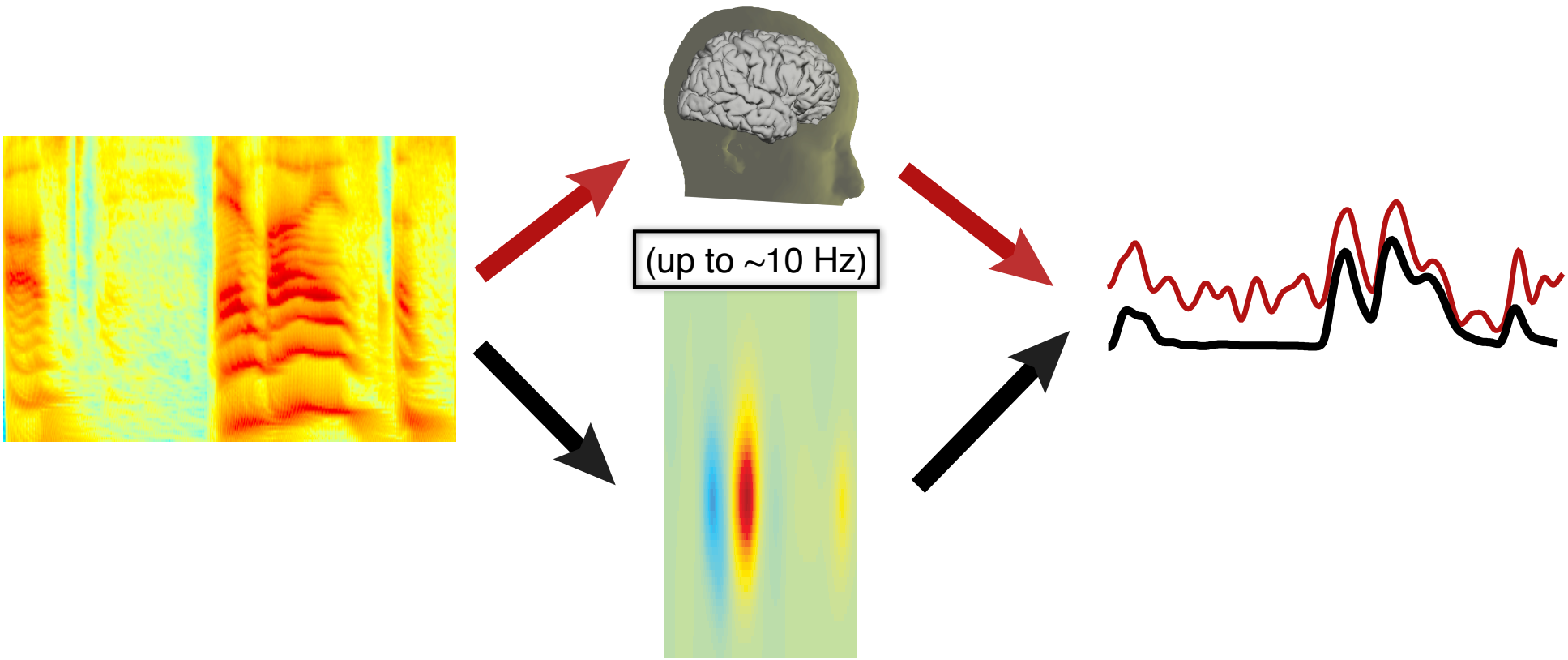


Ding & Simon, J Neurophysiol (2009)
Wang et al., J Neurophysiol (2012)

MEG Responses to Speech Modulations



MEG Responses Predicted by STRF Model

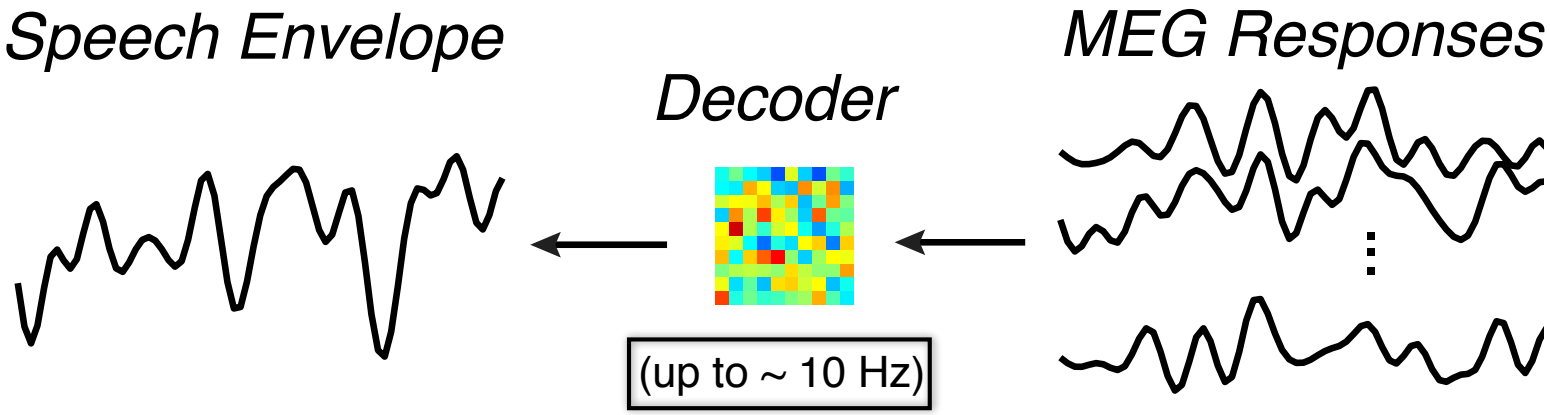


Linear Kernel = STRF

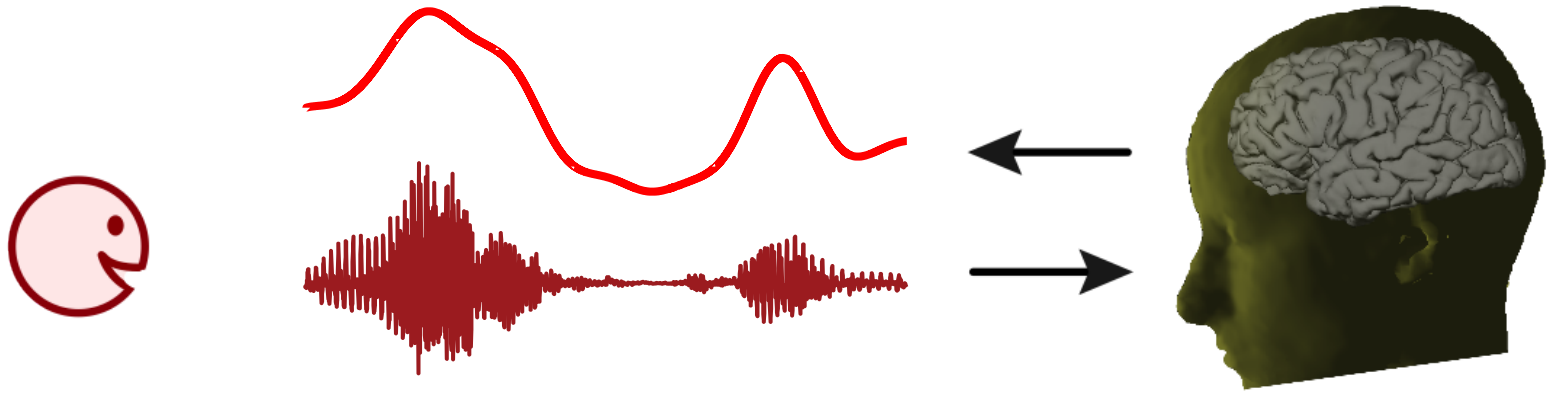
"Spectro-Temporal Response Function"

Ding & Simon, J Neurophysiol (2012)

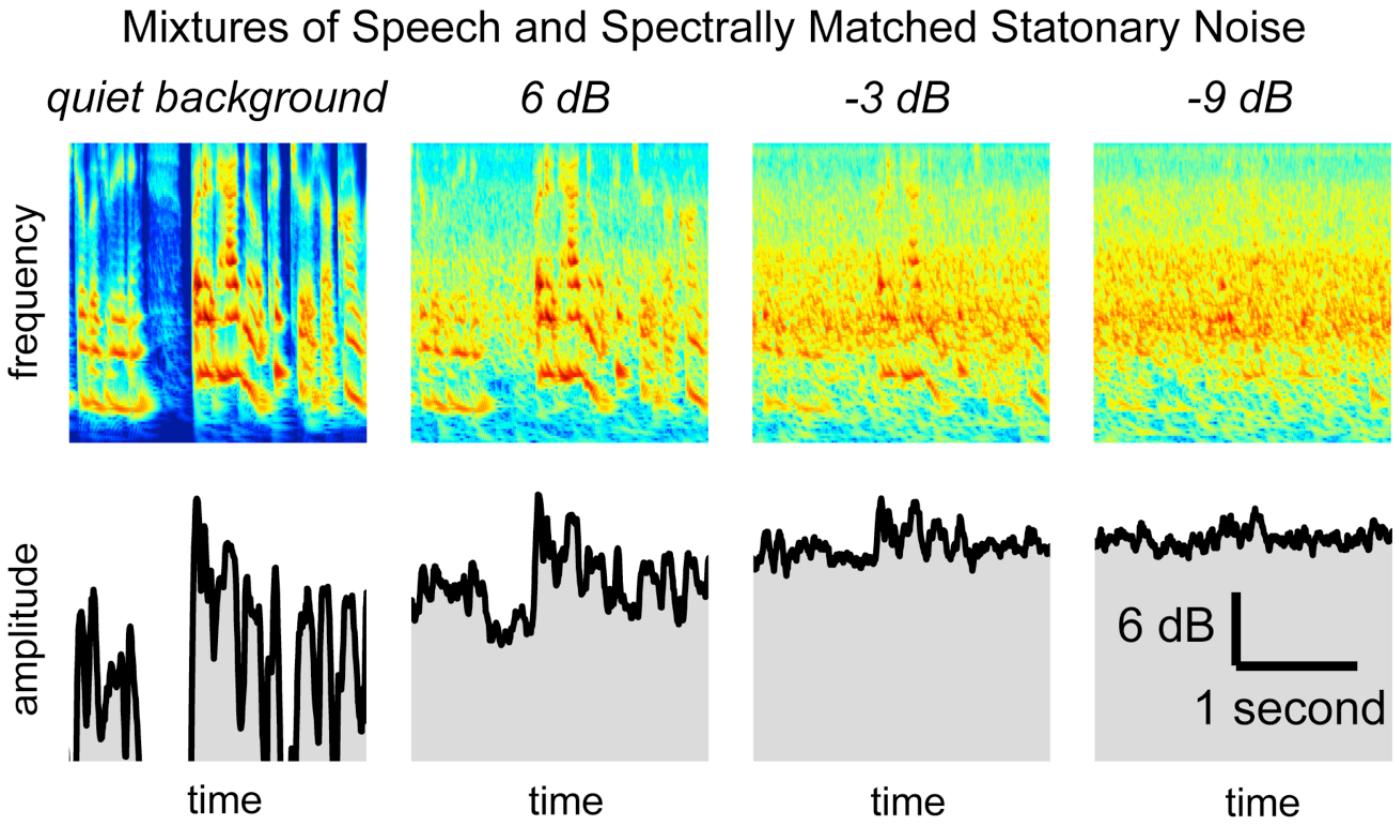
Neural Reconstruction of Speech Envelope

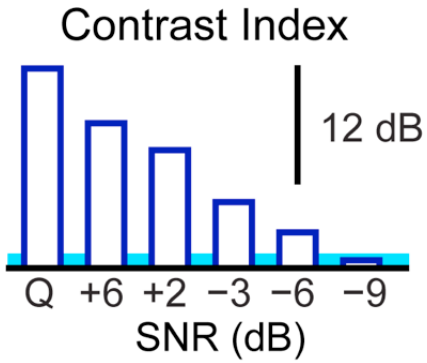


Neural Representation of Speech: Temporal



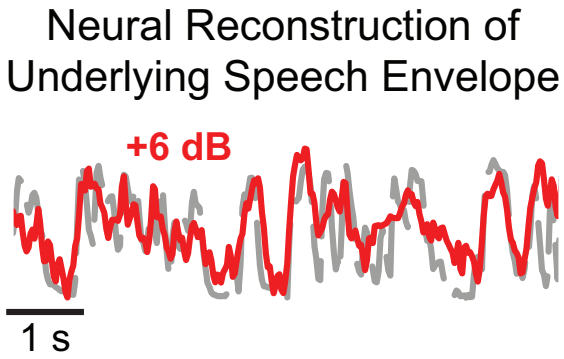
Speech in Stationary Noise

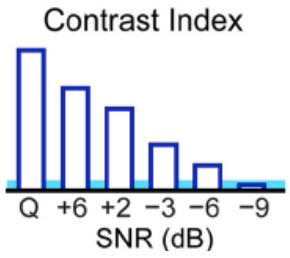


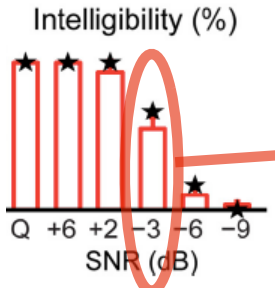


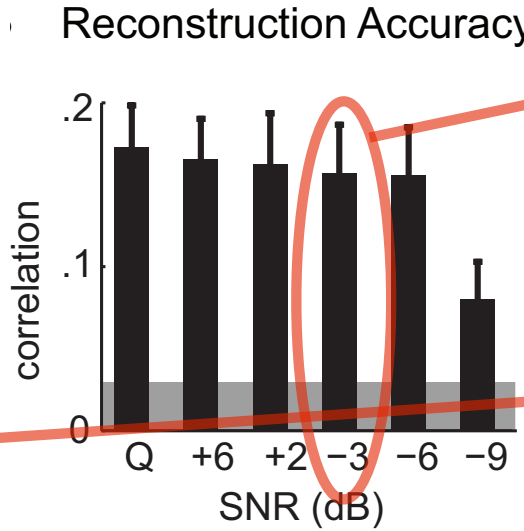
Ding & Simon, J Neuroscience (2013)

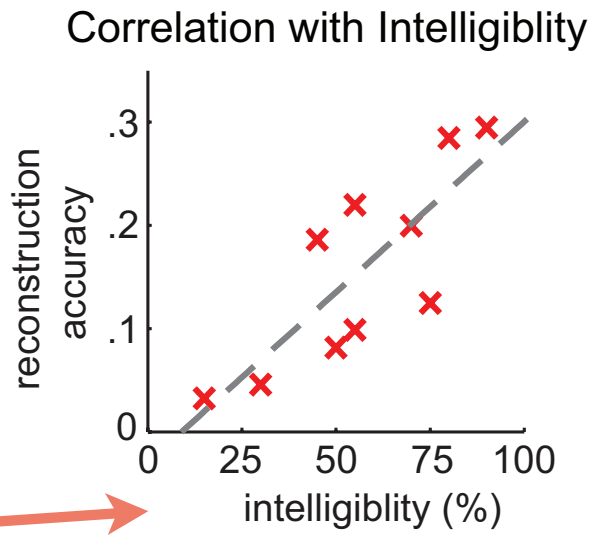
Speech in Noise: Results







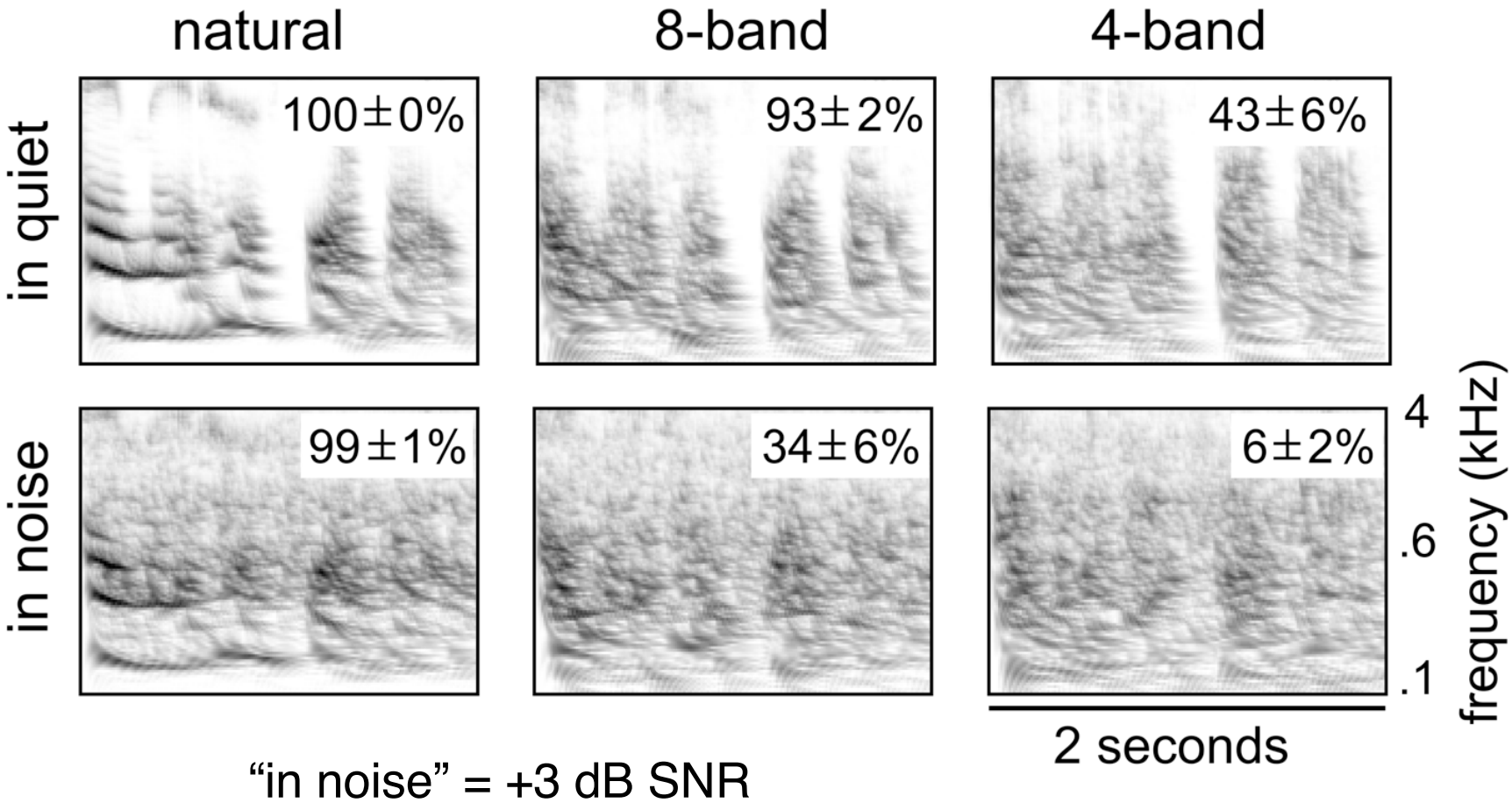




across Subjects

Ding & Simon, J Neuroscience (2013)

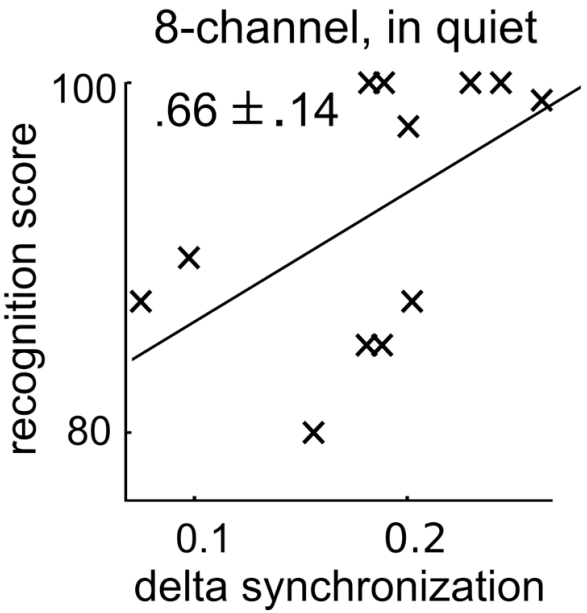
Noise-Vocoded Speech

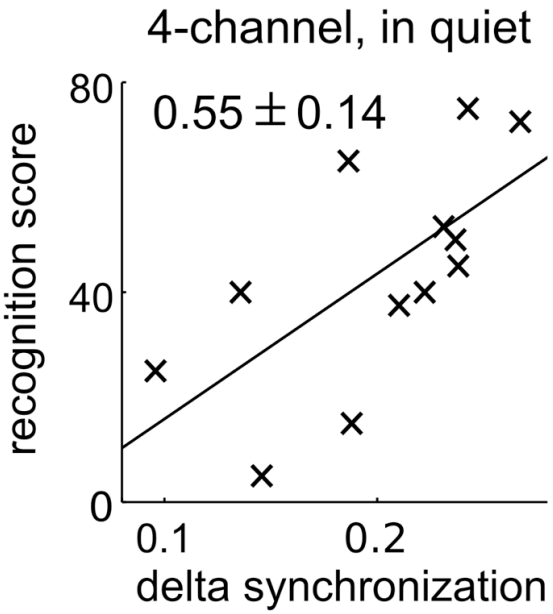


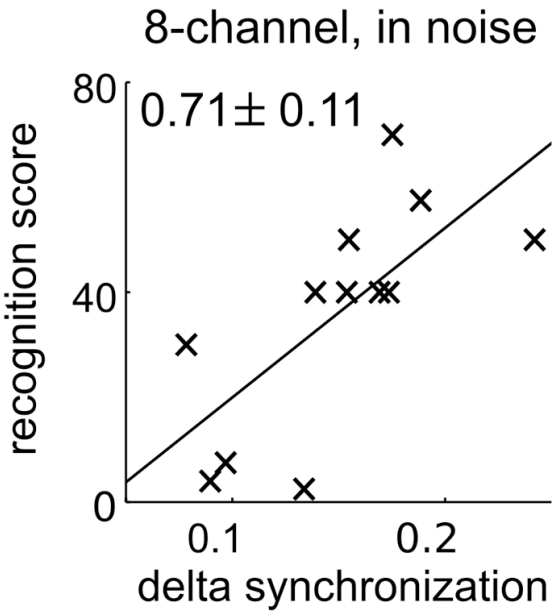
"in noise" = +3 dB SNR

Ding, Chatterjee & Simon, NeuroImage (2014)

Noise-Vocoded Speech: Results





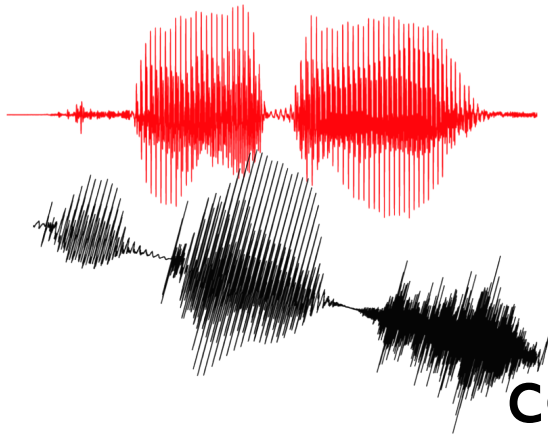


Cortical Speech Representations

- Neural Representations: Encoding & Decoding
- Linear models: Useful & Robust
- Speech **Envelope** only (as seen in MEG)
- Envelope Rates: ~ 1 - 10 Hz
- Intelligibility linked to lower range of frequencies (Delta)

Listening to Speech at the Cocktail Party

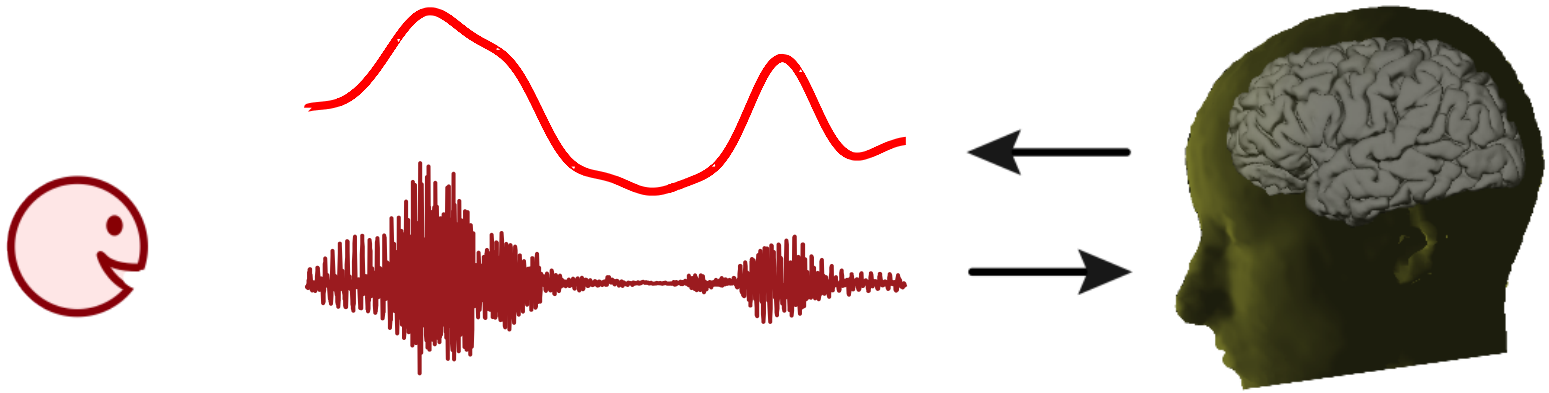
Experiments



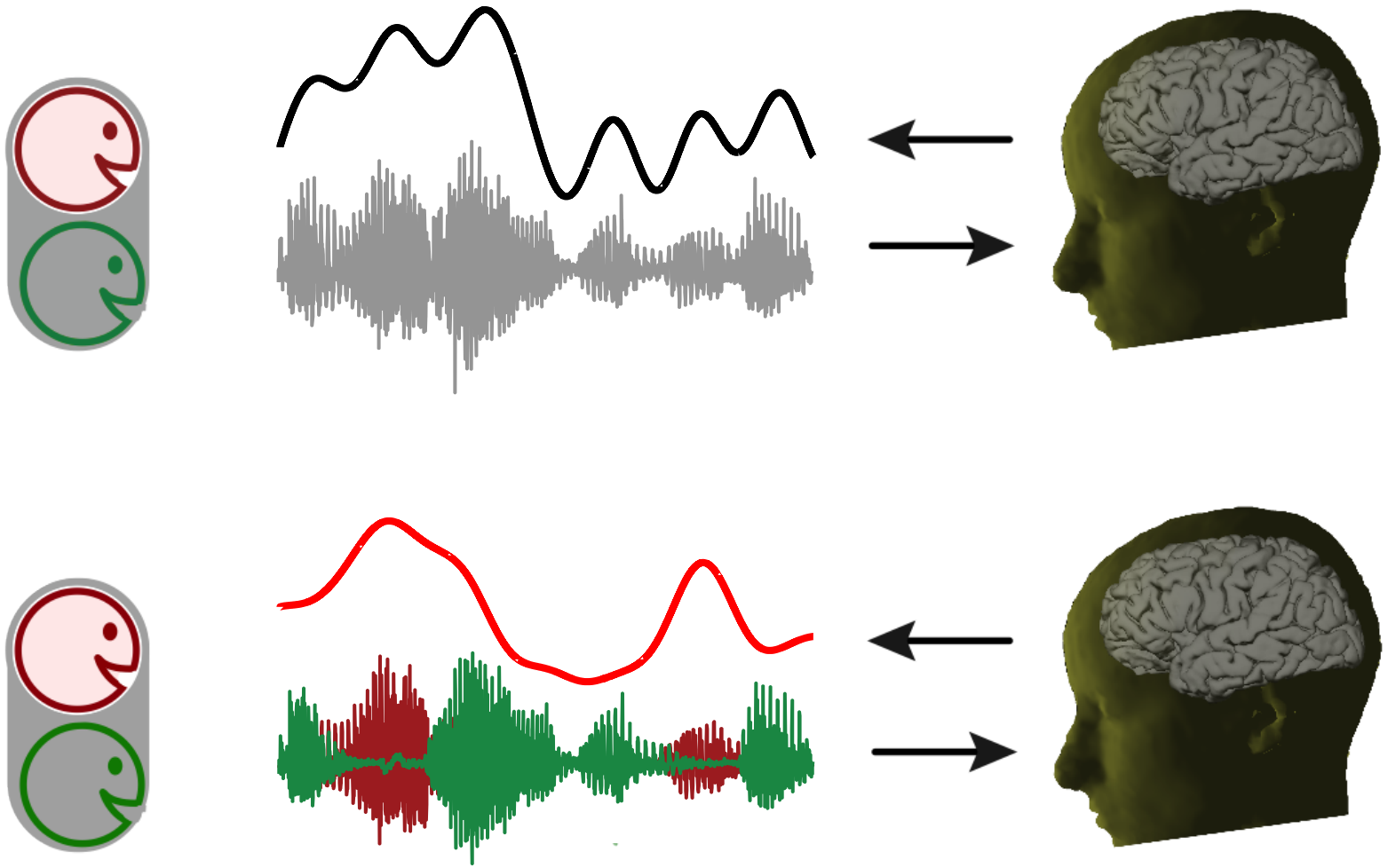
speech

competing speech

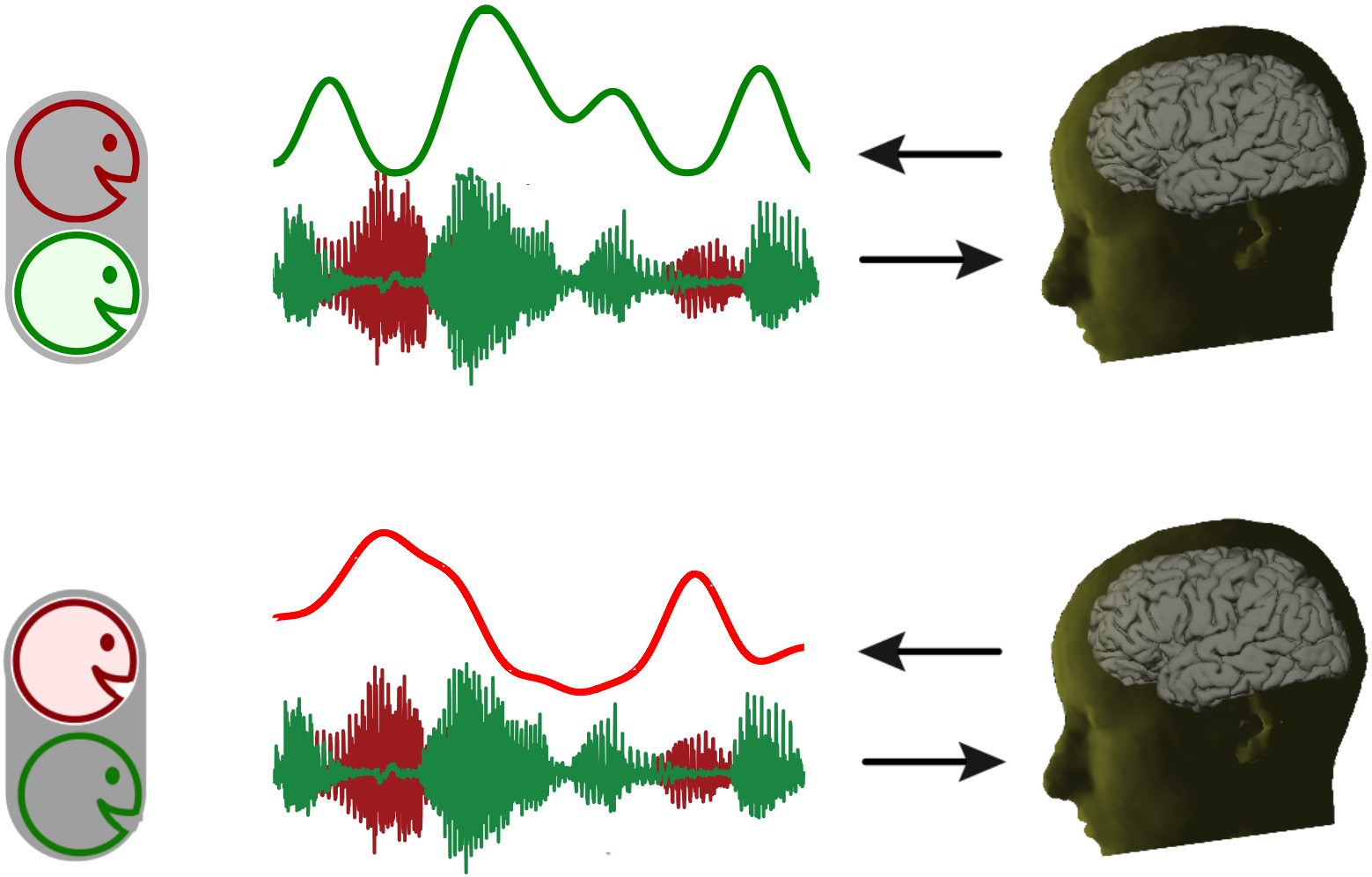
Selective Neural Encoding



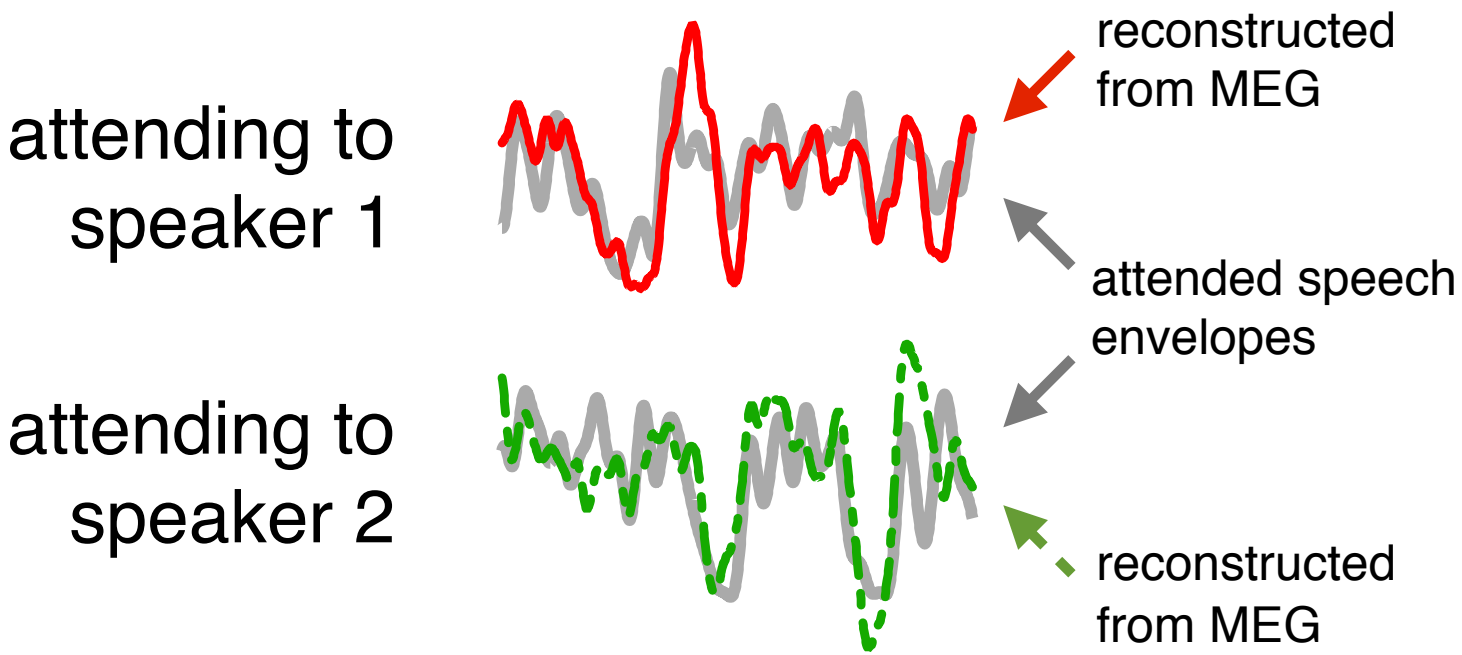
Unselective vs. Selective Neural Encoding



Selective Neural Encoding



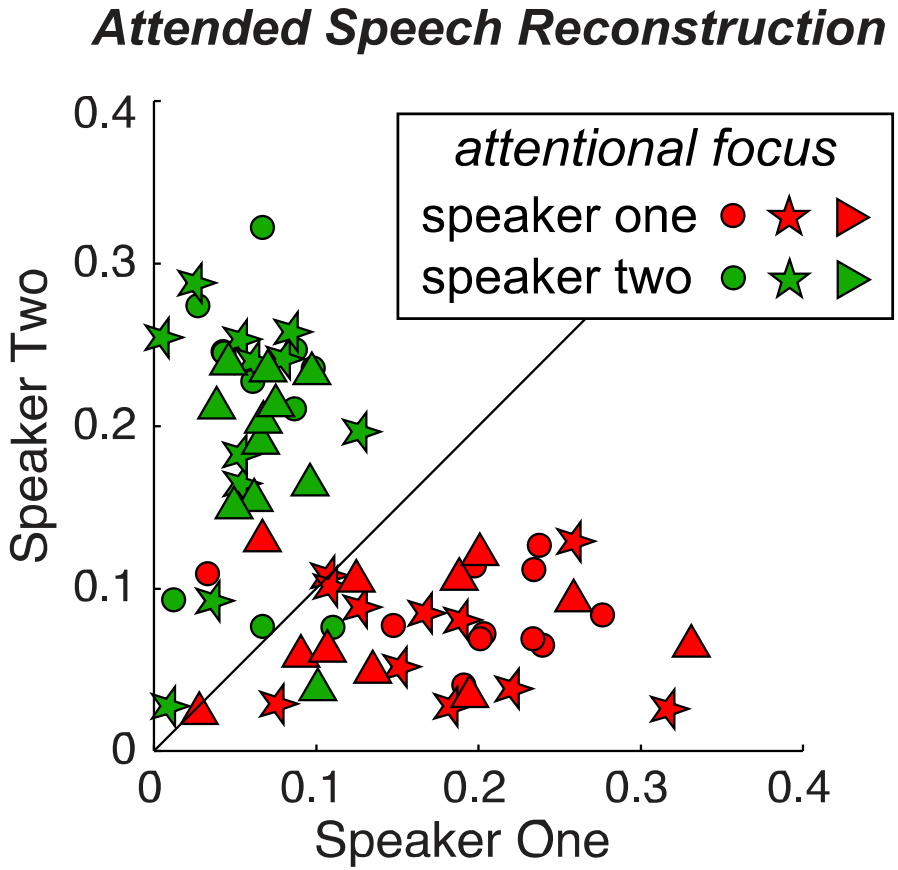
Stream-Specific Representation



Identical Stimuli!

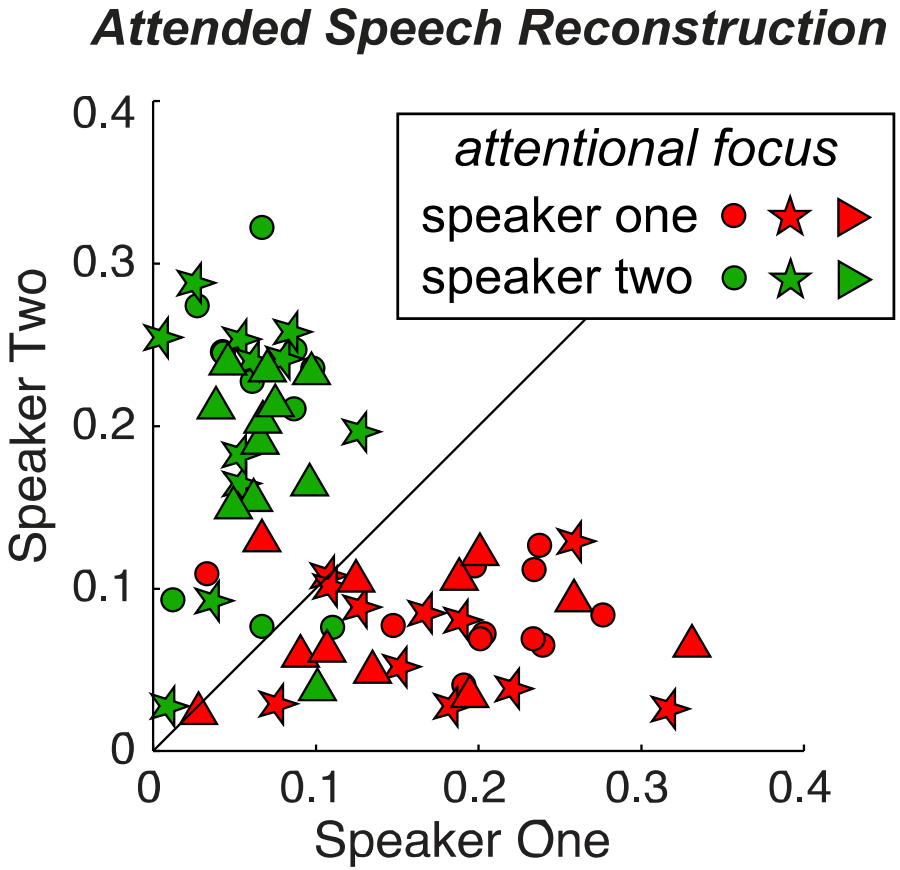
Ding & Simon, PNAS (2012)

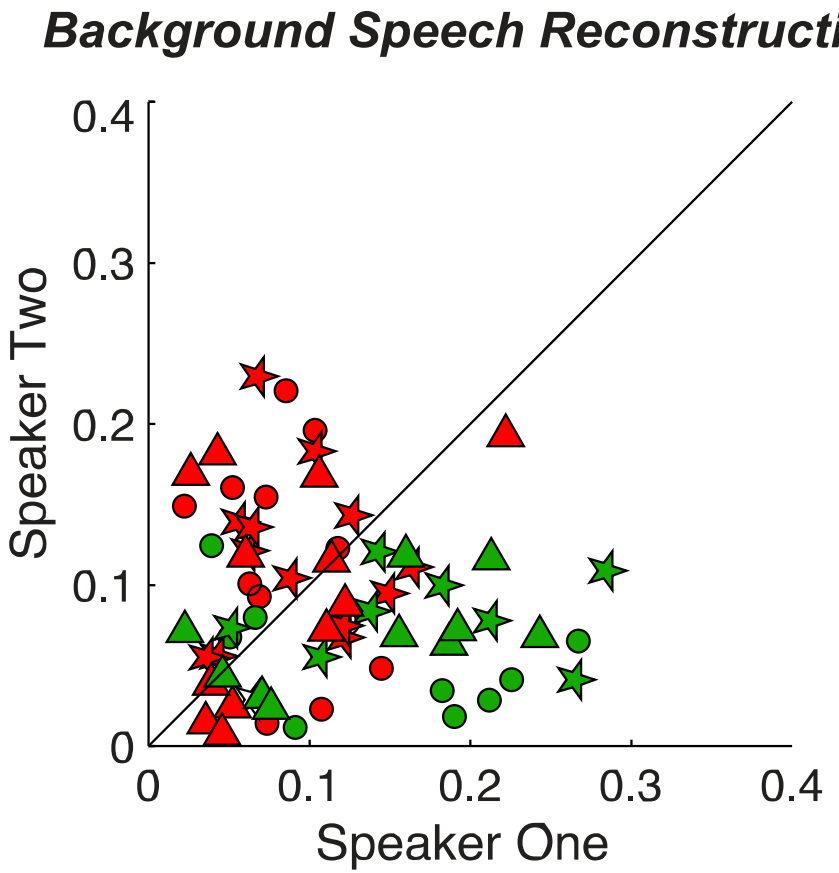
Single Trial Speech Reconstruction



Ding & Simon, PNAS (2012)

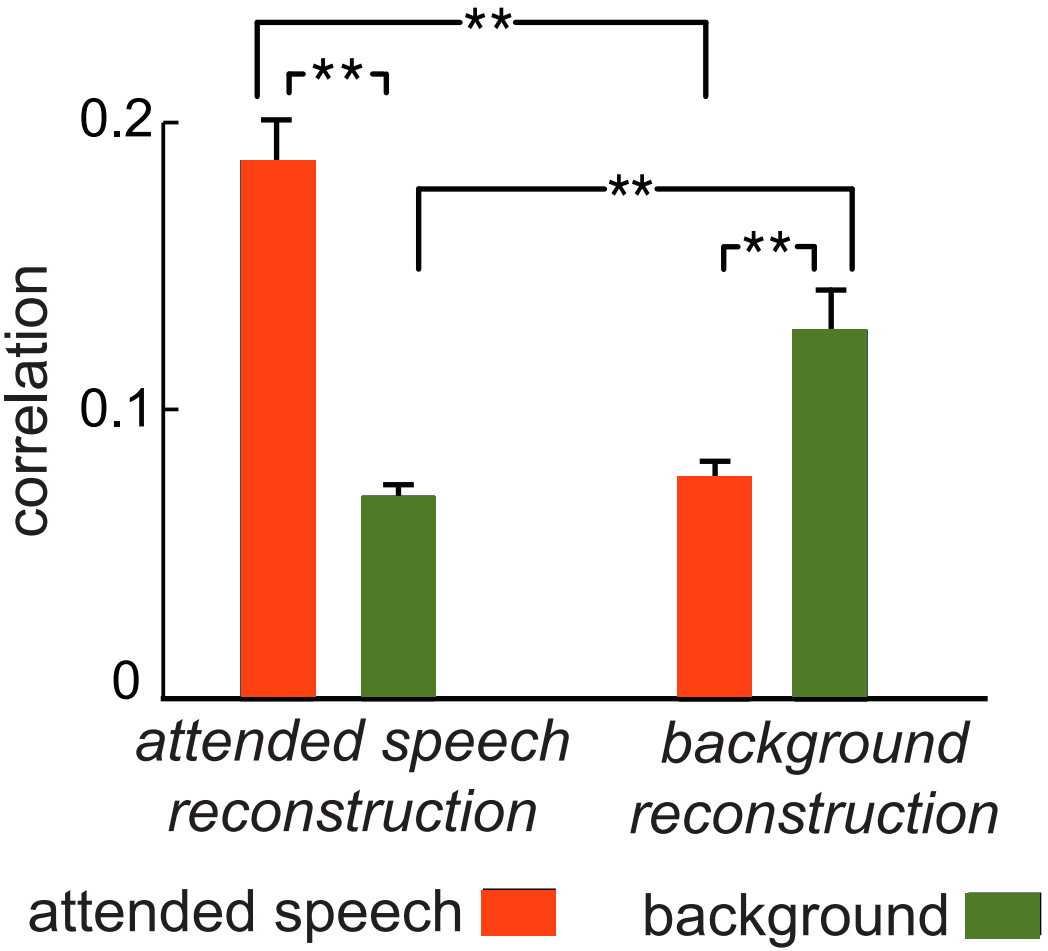
Single Trial Speech Reconstruction





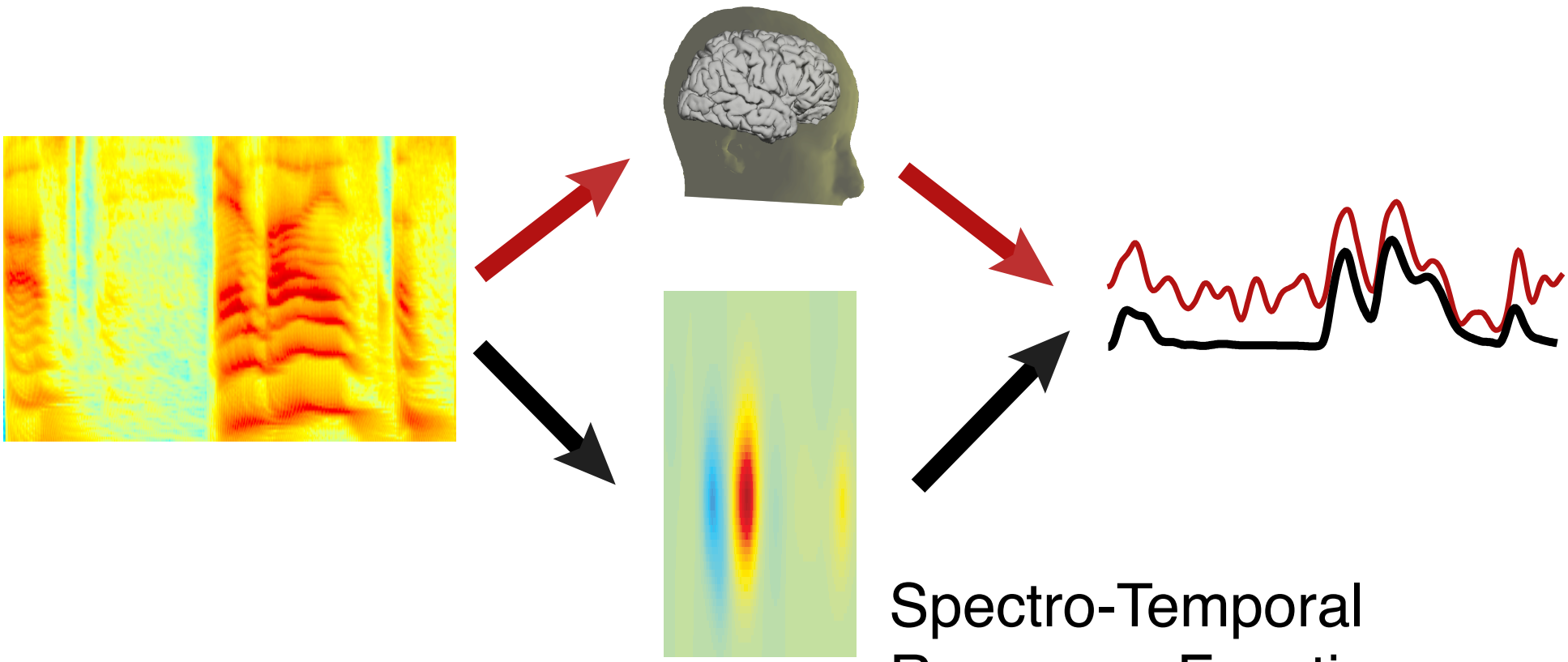
Ding & Simon, PNAS (2012)

Overall Speech Reconstruction



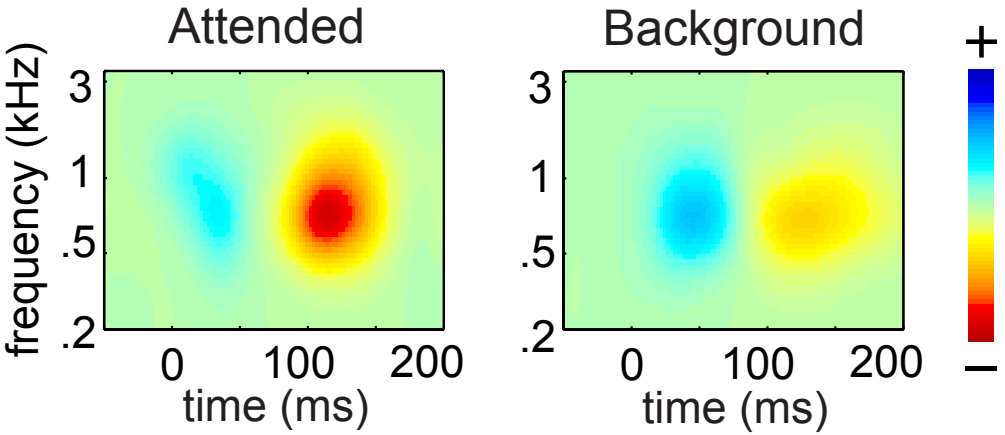
Distinct neural representations for different speech streams

Forward STRF Model



Spectro-Temporal
Response Function
(STRF)

STRF Results



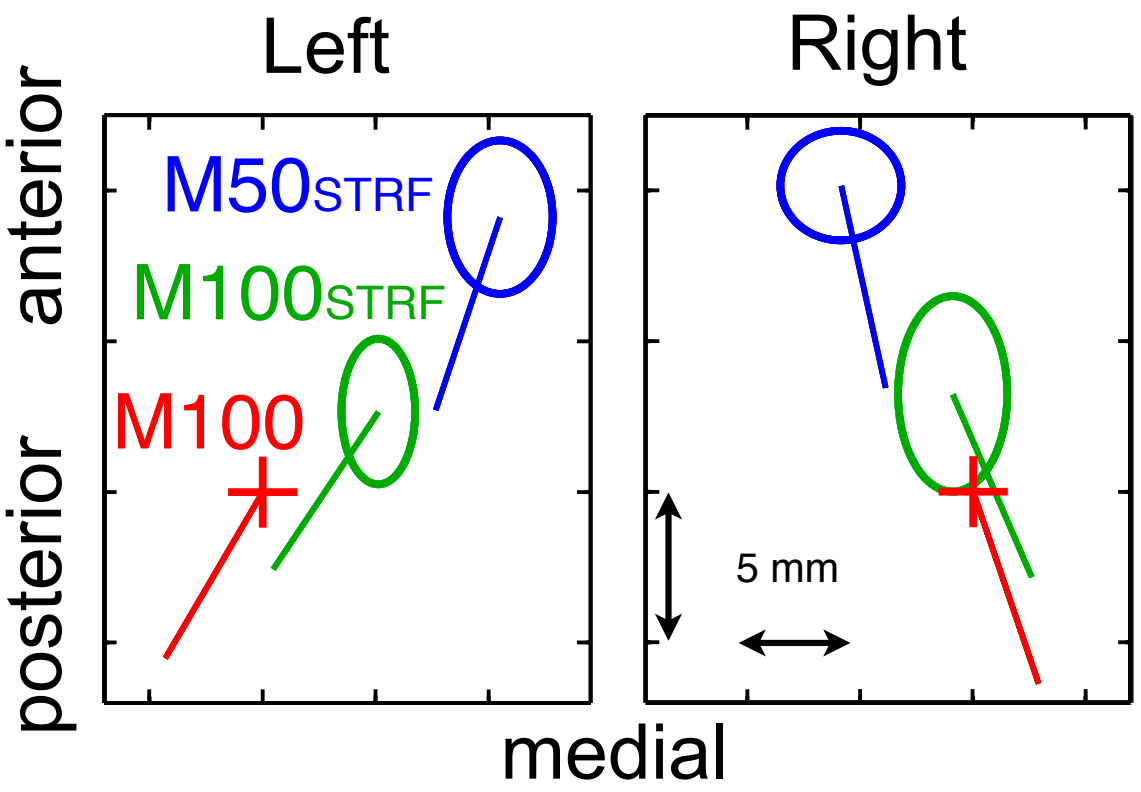
- STRF separable (time, frequency)
- 300 Hz - 2 kHz dominant carriers
- $M50_{STRF}$ positive peak
- $M100_{STRF}$ negative peak

Neural Sources

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

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

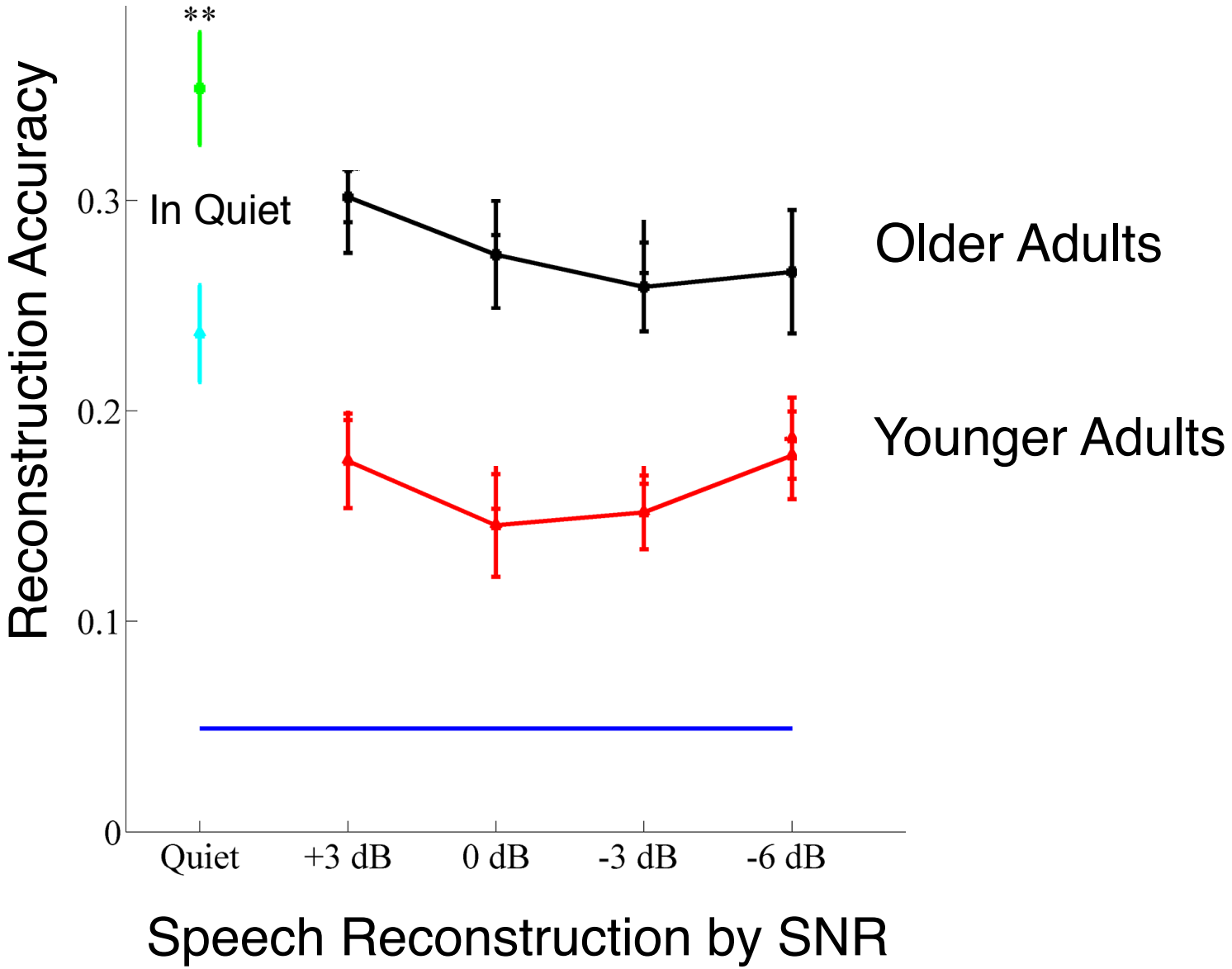
- **PT strongly modulated by attention, *but not HG***



Recent Studies

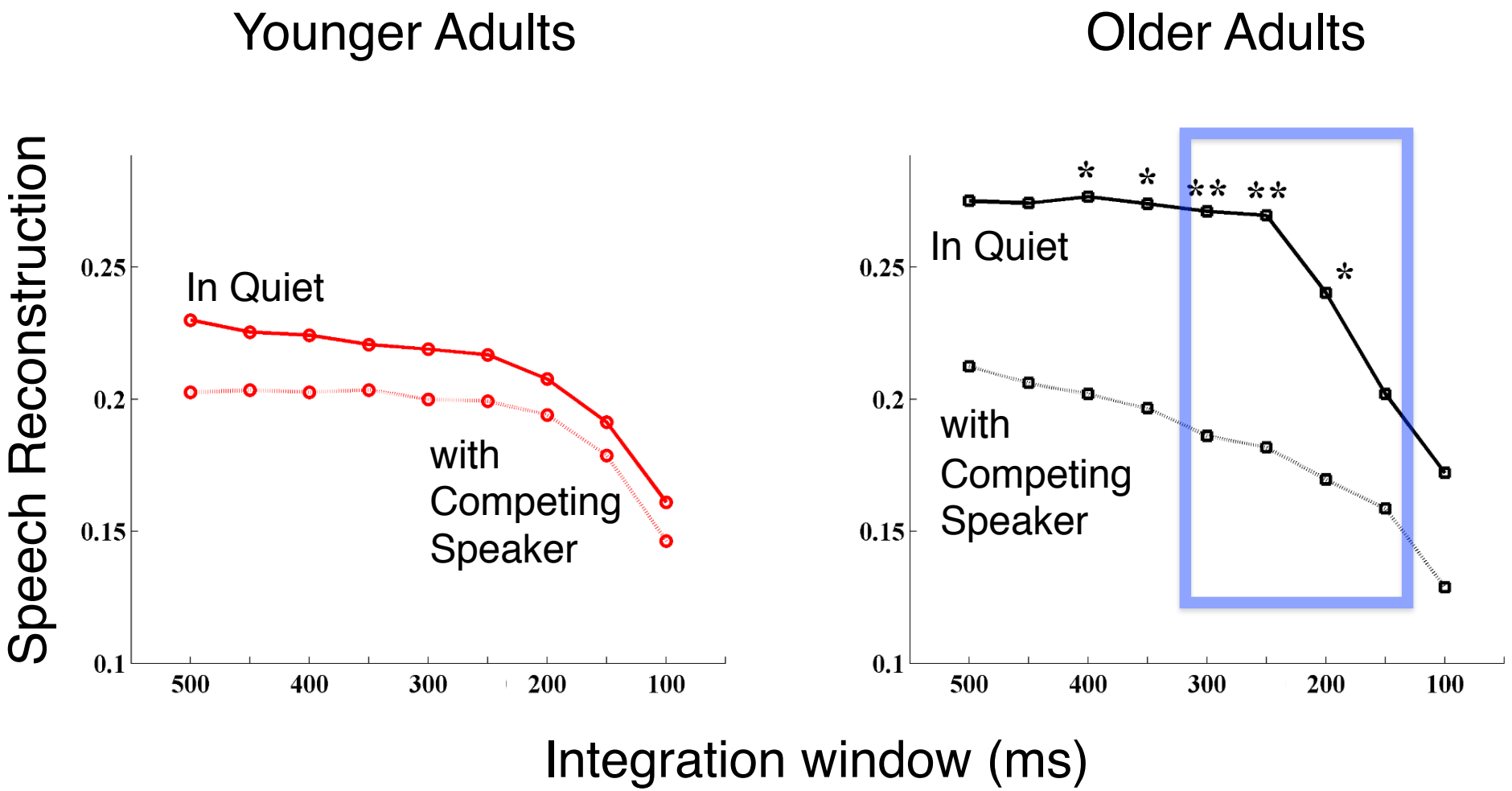
- Aging & Cortical Representations of Speech
- High Level Interference & Noise
- Representations of Internal Speech

Younger vs. Older Listeners



Presacco et al., J Neurophysiol (2016a)

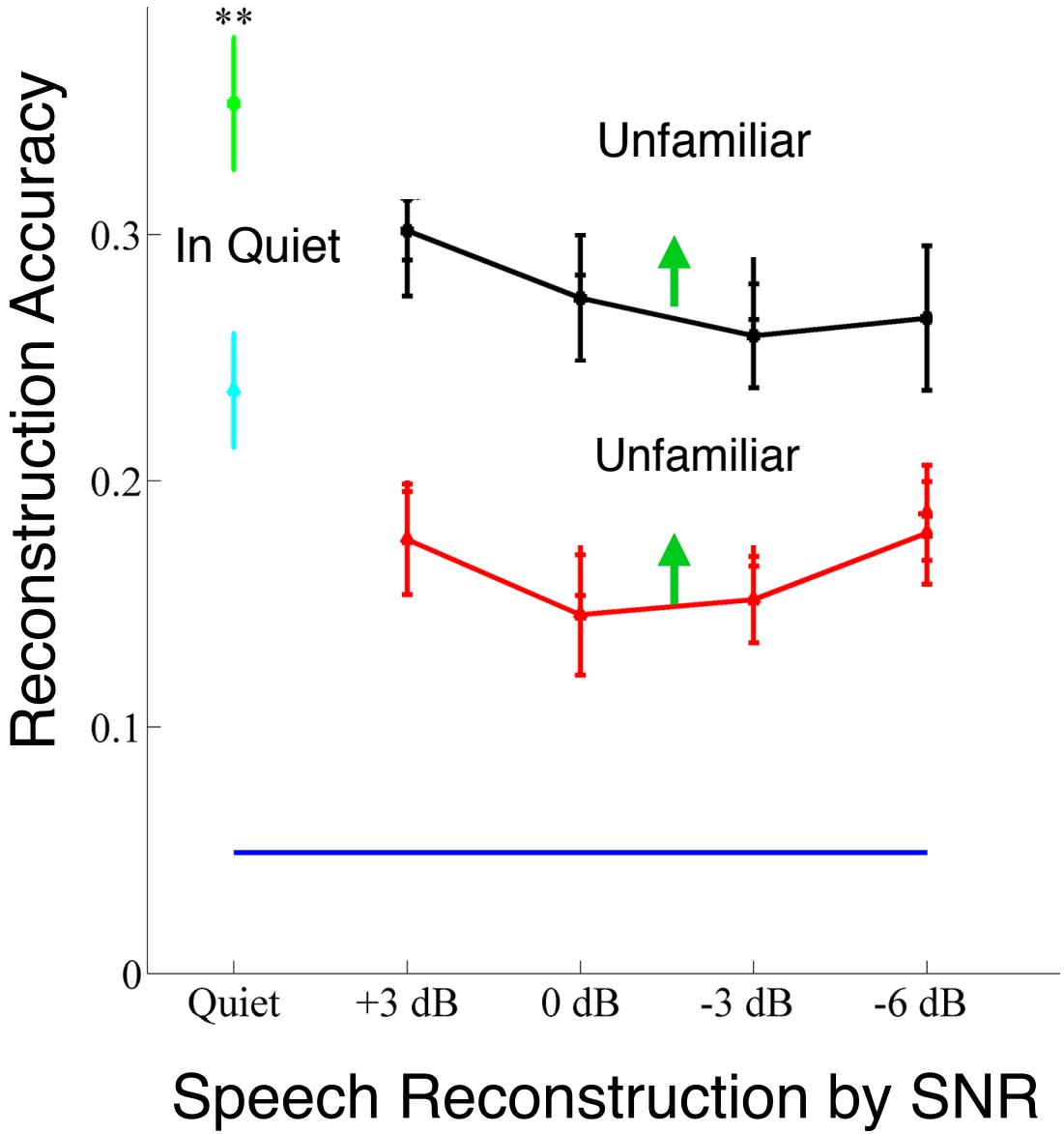
Younger vs. Older Listeners



Presacco et al., J Neurophysiol (2016a)

Effect absent in Midbrain (FFR Response)

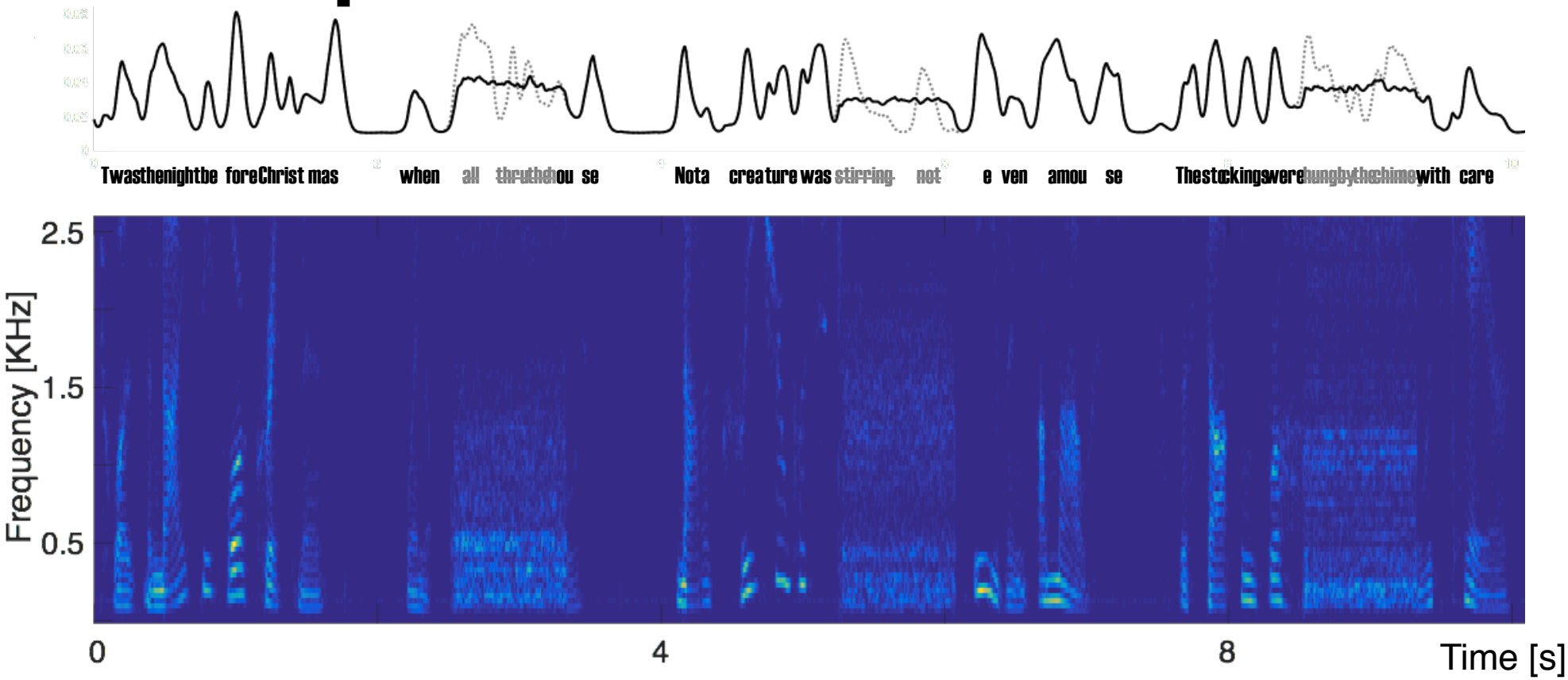
High Level Interference Effects



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

Presacco et al., J Neurophysiol (2016b)

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, bioRxiv 251793

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.

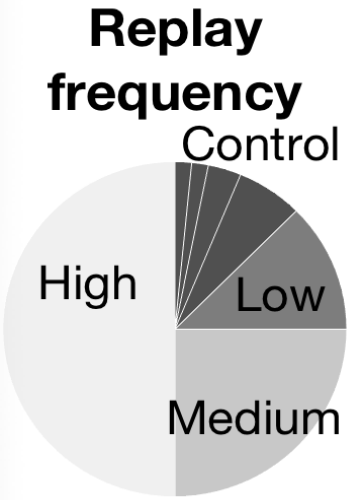
His eyes--how they twinkled! His dimples, how merry!
His cheeks were like roses, his nose like a cherry!
His droll little mouth was drawn up like a bow,
and the beard on his chin was as white as the snow.

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.

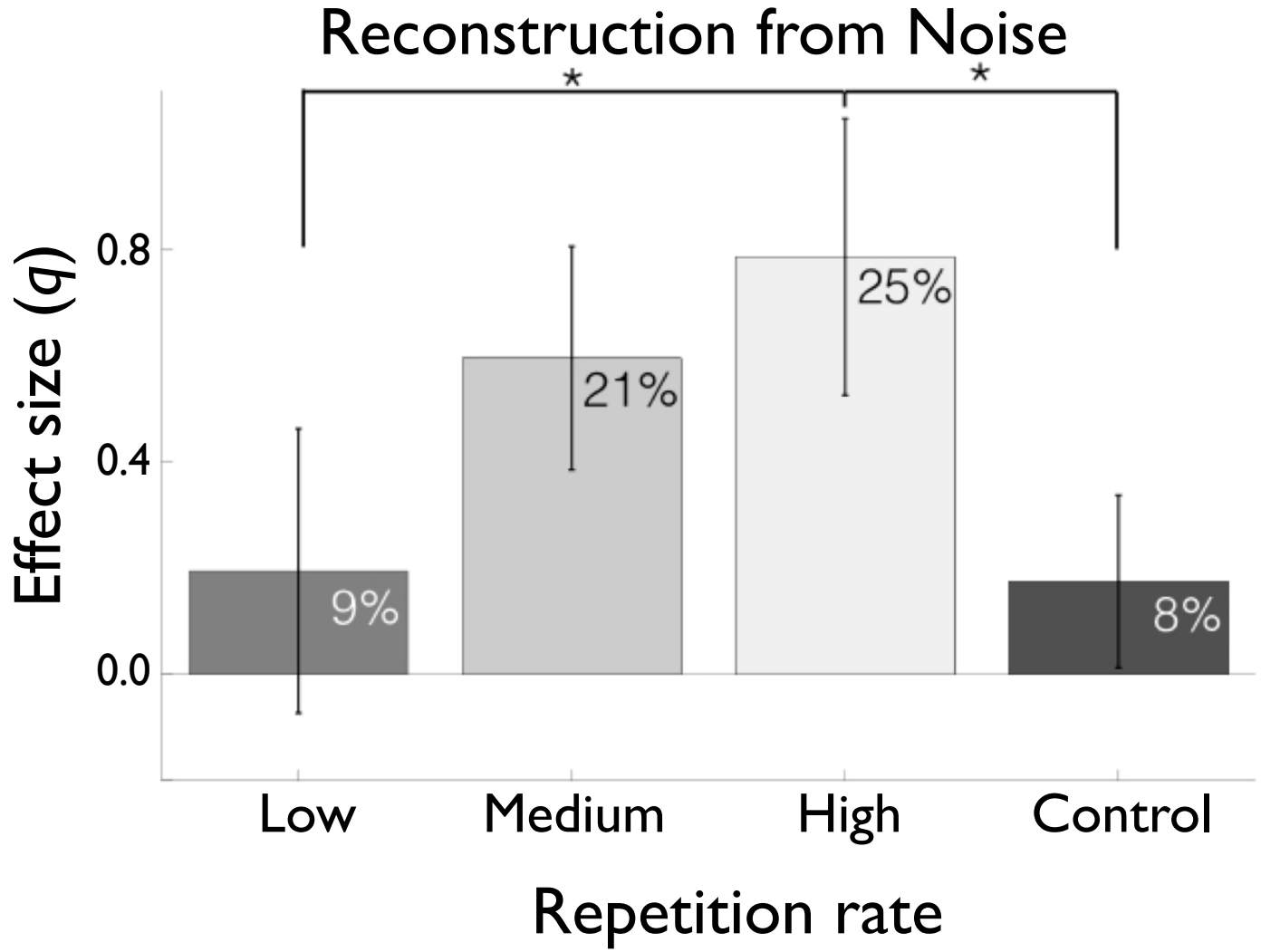
He spoke not a word, but went straight to his work,
and filled all the stockings, then turned with a jerk.
And laying his finger aside of his nose,
and giving a nod, up the chimney he rose.

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



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

Summary

- Cortical representations of speech
  - representation of envelope (up to ~10 Hz)
  - robust against a variety of noise types
  - neural representation of perceptual object
- Object-based representation at 100 ms latency (PT), but not by 50 ms (HG)
- Aging shows over-representation (and time integration deficits)
- Applies to acoustically missing internal speech

Thank You