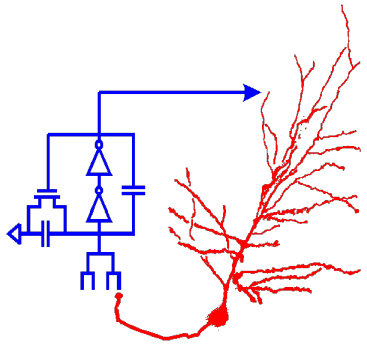


Cortical Representation of Continuous Speech in Complex Auditory Scenes

Nai Ding, Jonathan Z. Simon

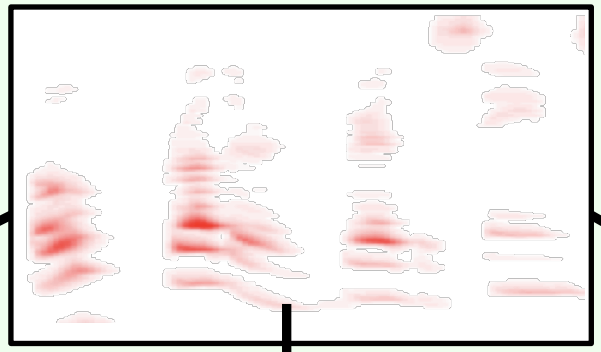
University of Maryland, College Park

supported by NIH R01 DC 008342

Speech Perception Is Largely Invariant to the Auditory Background

variable acoustics → invariant perception

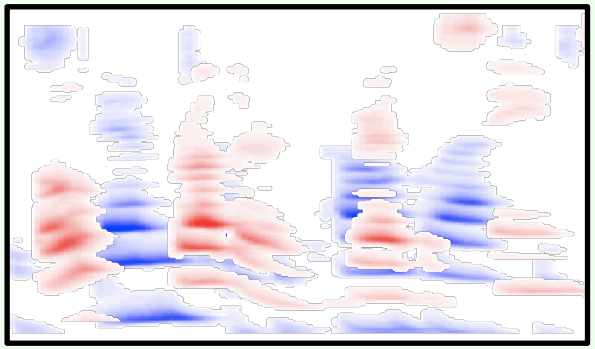
Important Examples of Acoustic Variations of Speech



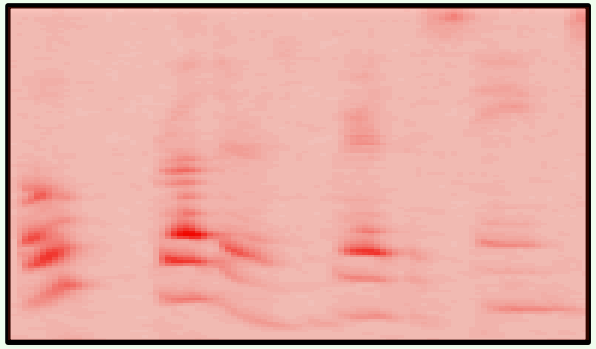
with a competing speech stream

embedded in noise

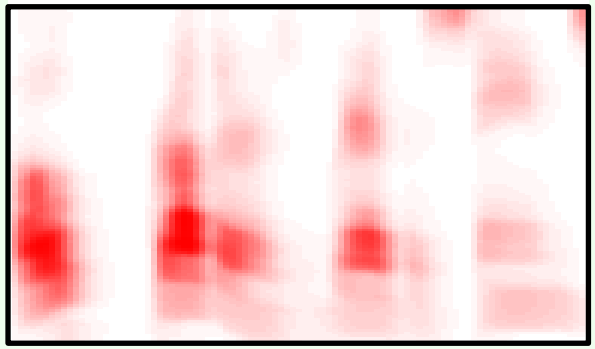
reduced spectral resolution



top down selection

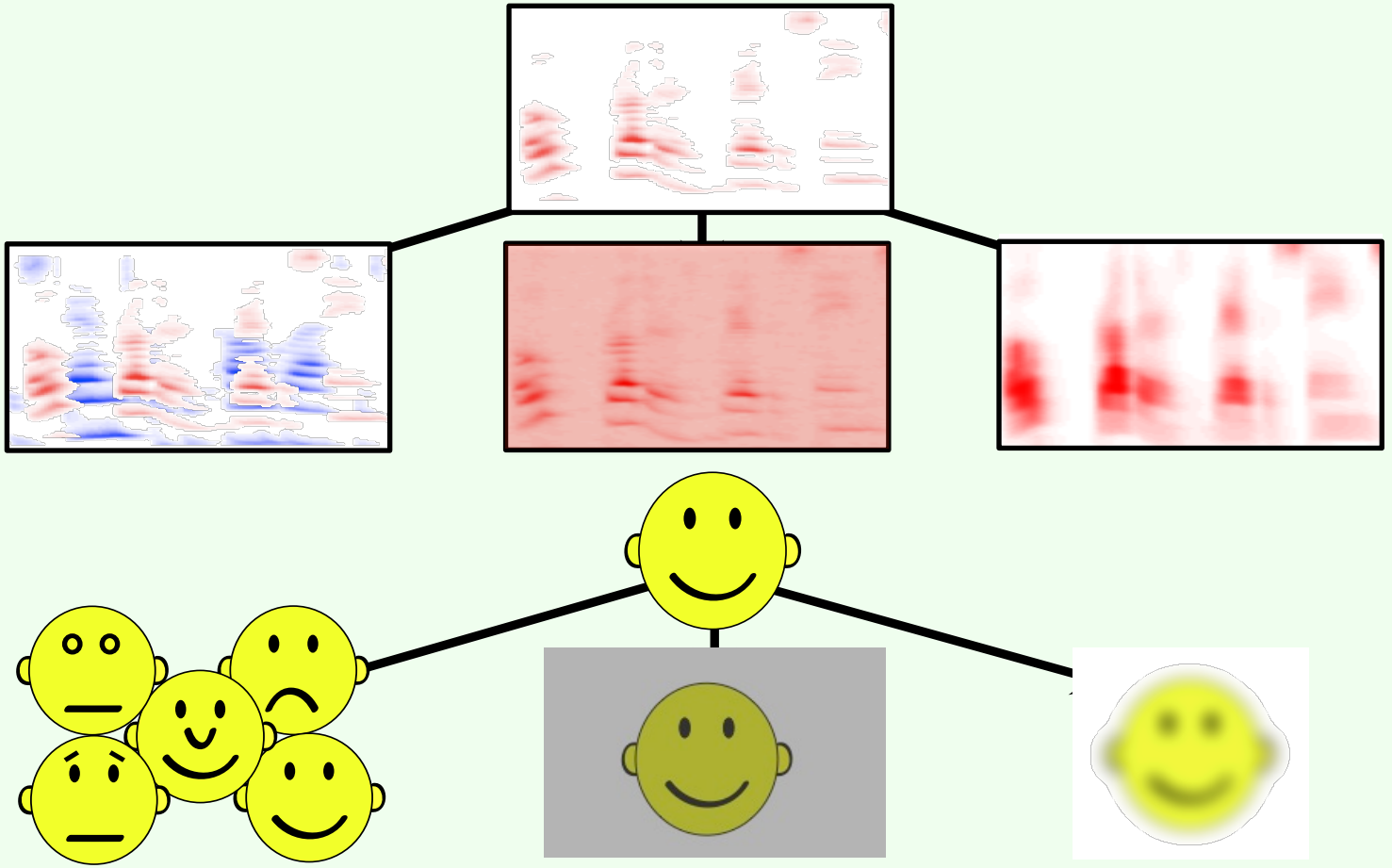


bottom up extraction



recognition

Acoustic Variations of Speech & Visual Analogies

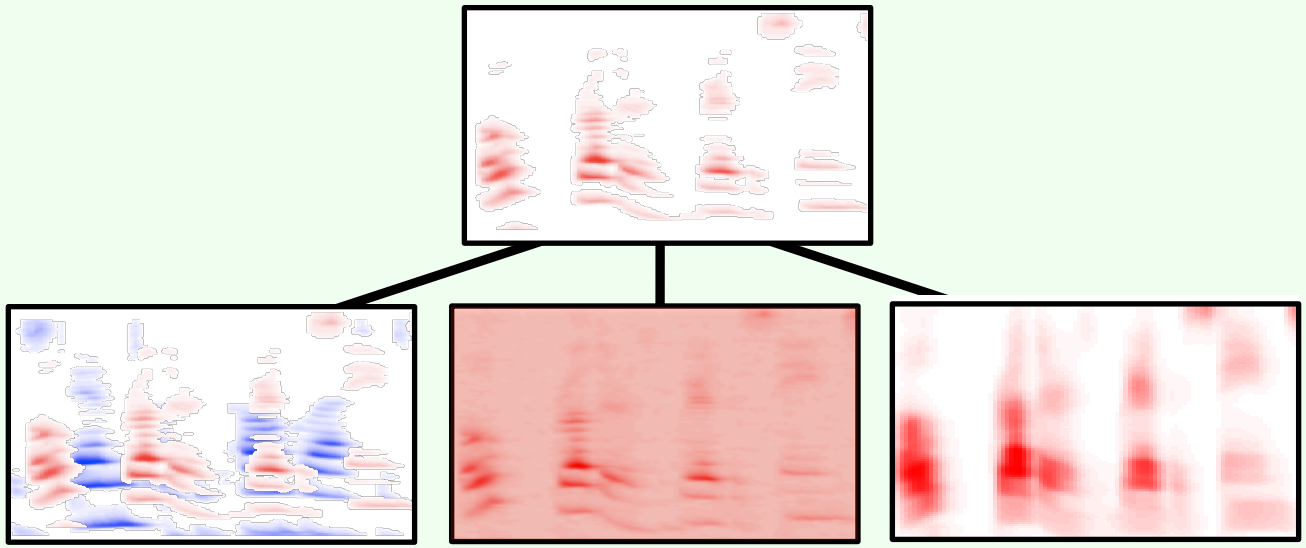


competing objects
(informational masking)

shadow
(energetic masking)

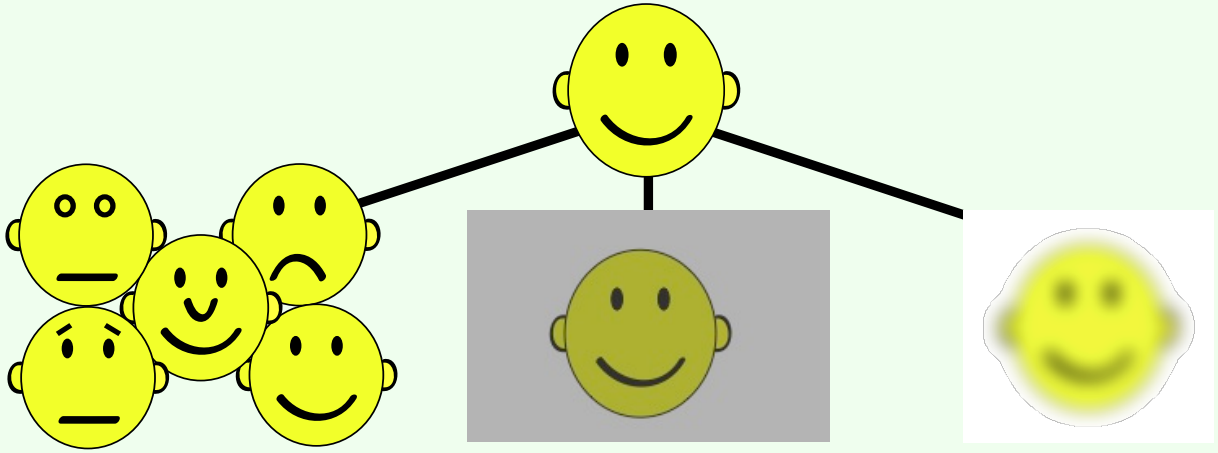
blurring
(recognition)

Speech as an Auditory Object



auditory object

auditory scene

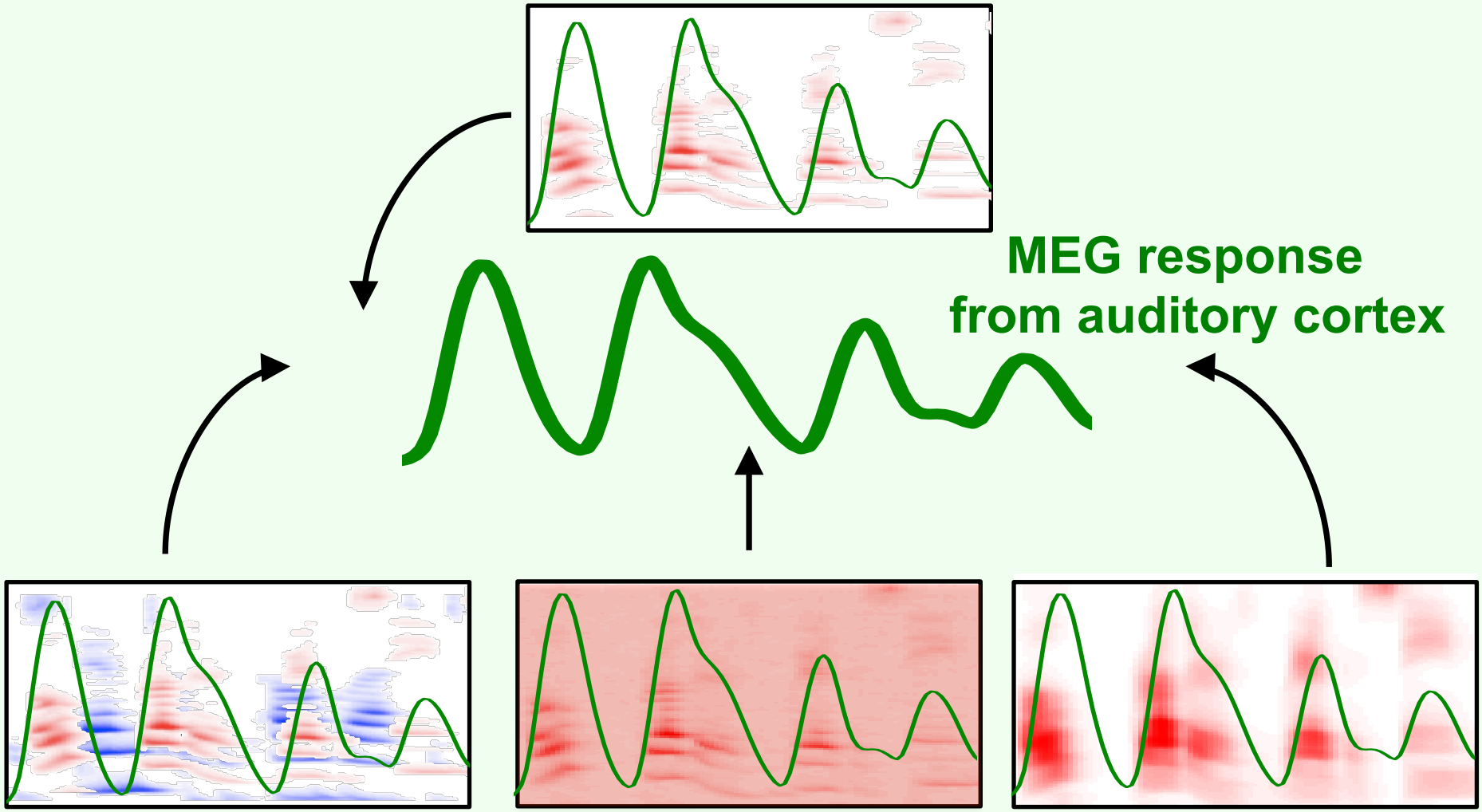


visual object

visual scene

Griffiths & Warren, 2004
Shamma, Elhilali & Micheyl, 2011

Auditory-object Based Neural Representation in Auditory Cortex?



The Questions to Address

1. Neural entrainment to speech as a correlate of invariant perception of auditory objects?

   (Poeppel & Giraud, 2012; Shamma et al. 2011; Schroeder and Lakatos, 2009)

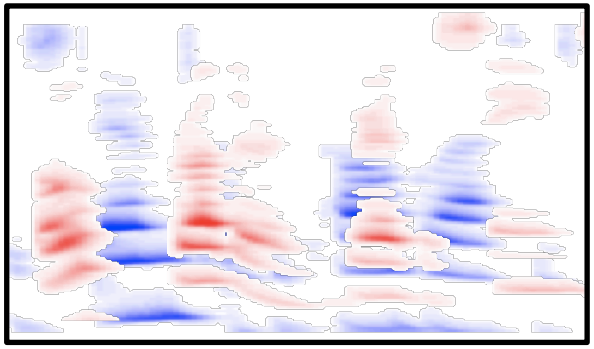
2. Which aspects of neural responses are influenced by acoustic degradation?

Experimental Paradigm

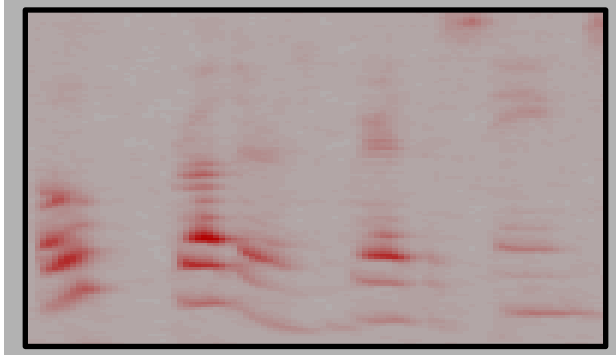
- Cortical activity was continuously recorded using MEG, while the subjects were listening to a story.
- Each stimulus is 1 minute in duration, and the subjects answer comprehension questions after each section.

Neural Encoding of Degraded Speech: Competing Speech Streams

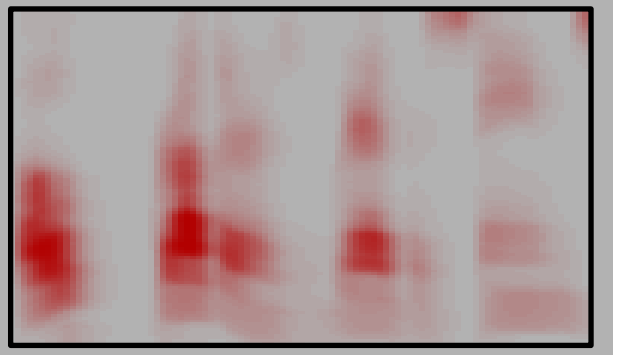
speech presented with a competing speech stream



speech embedded in noise



speech with reduced spectral resolution



Listen to One Speaker while Another Speaker Talking in the Background

Experimental Details

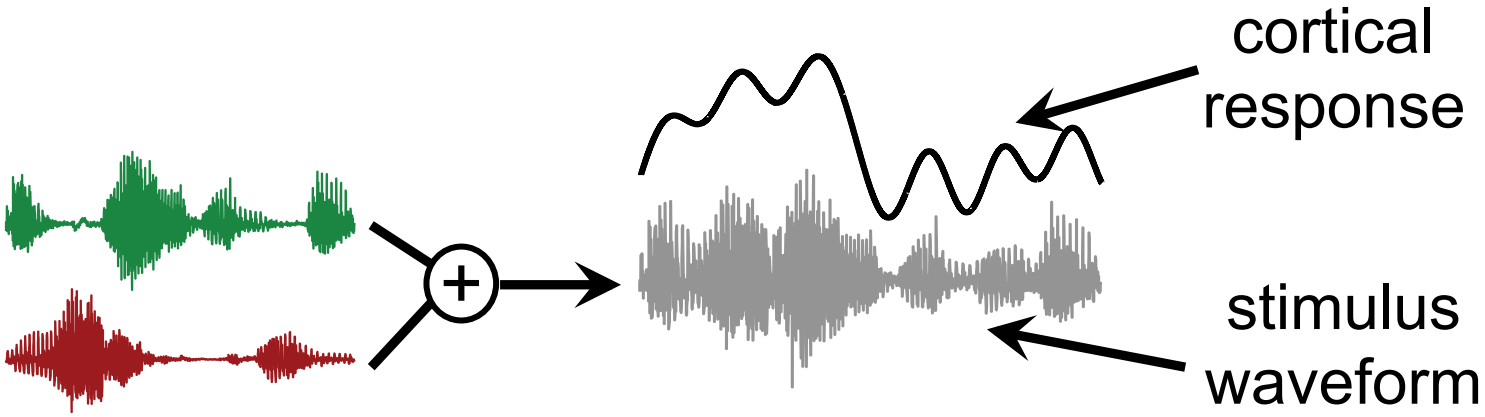
- a female speaker + a male speaker, each narrating a story.

- The subjects attended to one speaker, and answer comprehension questions.

- The intensity ratio between the two speakers vary between -8 and 8 dB, and speech intelligibility remains above 50%.

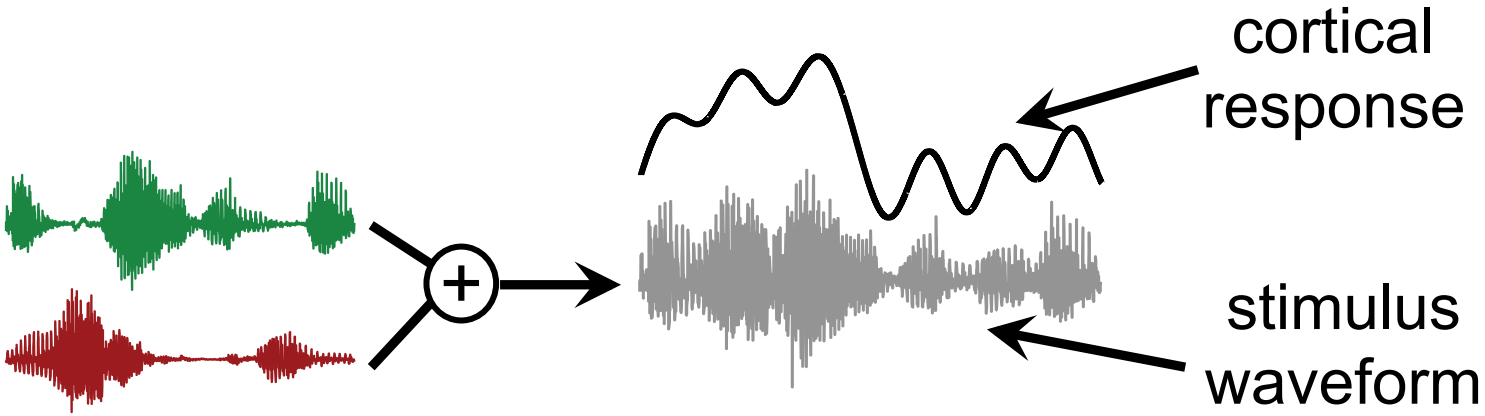
Stimulus-based Cortical Encoding?

encoding the physical stimulus

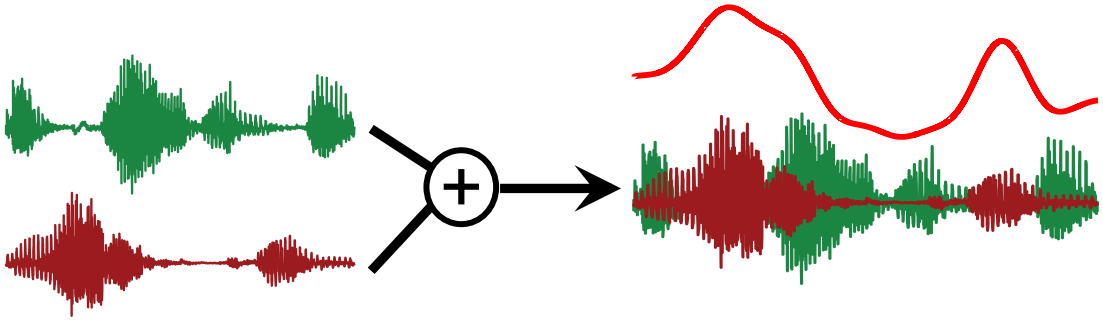


Object-based Cortical Encoding?
(A) Invariant to the Background

encoding the physical stimulus

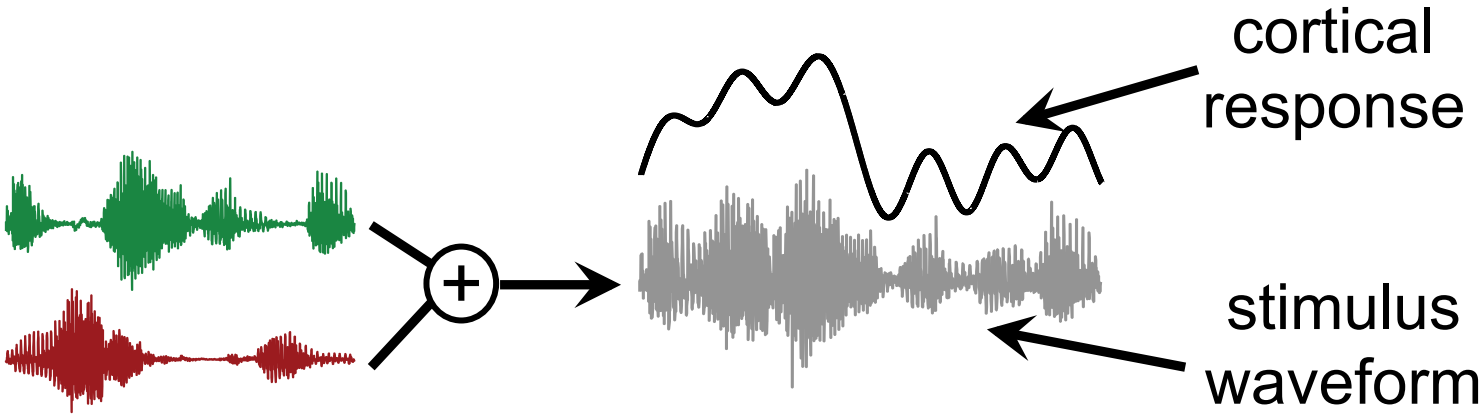


encoding an auditory object

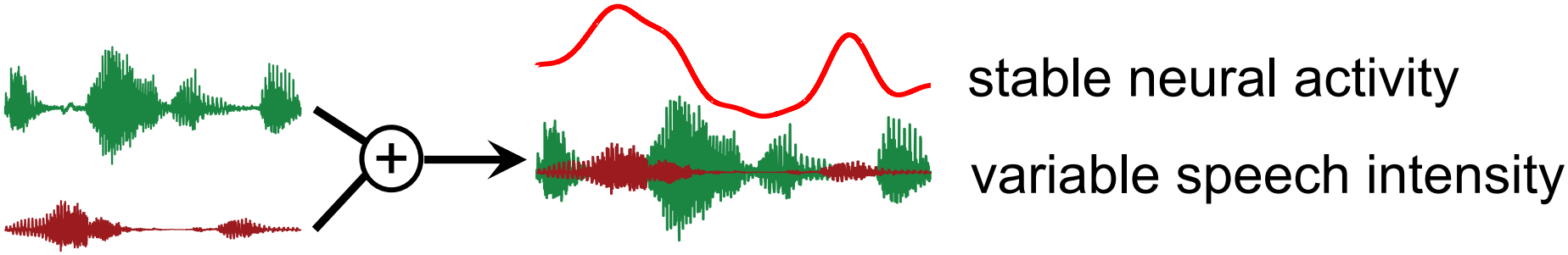


Object-based Cortical Encoding?
(B) Invariant to Basic Acoustic Variations

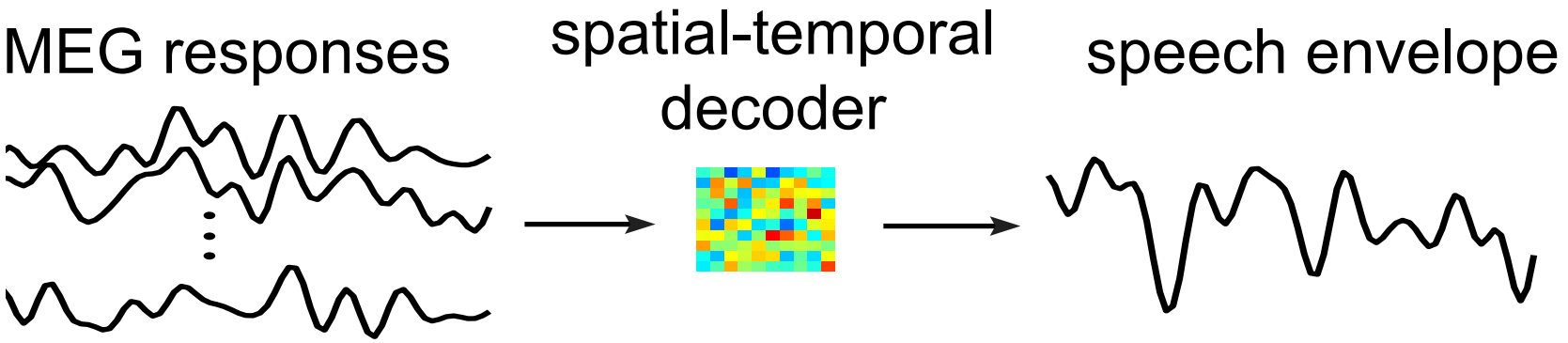
encoding the physical stimulus

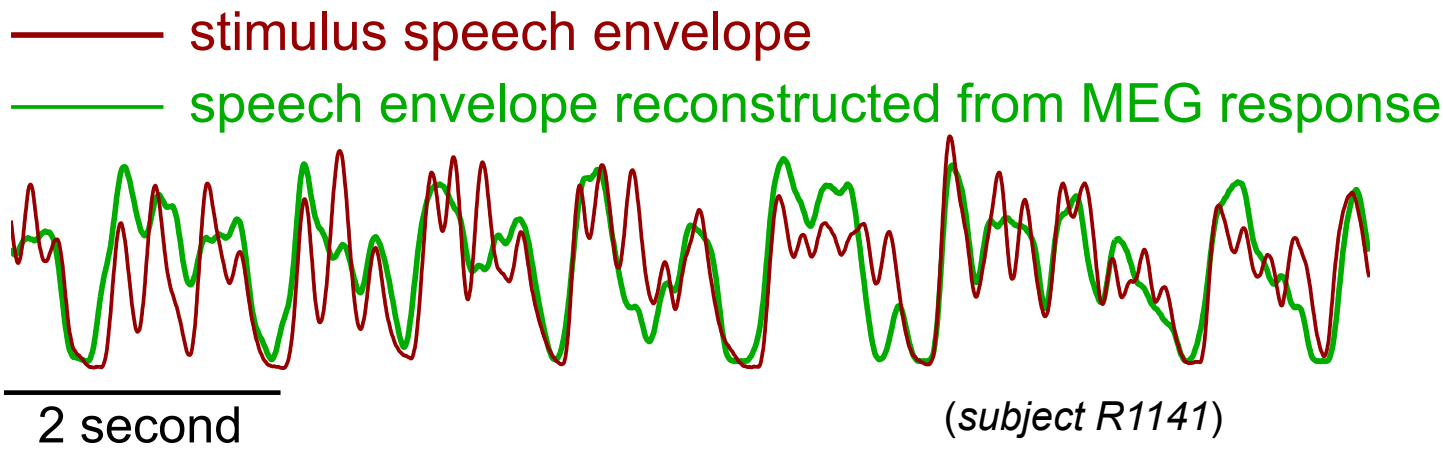


encoding an auditory object



Data Analysis: a Decoding Approach

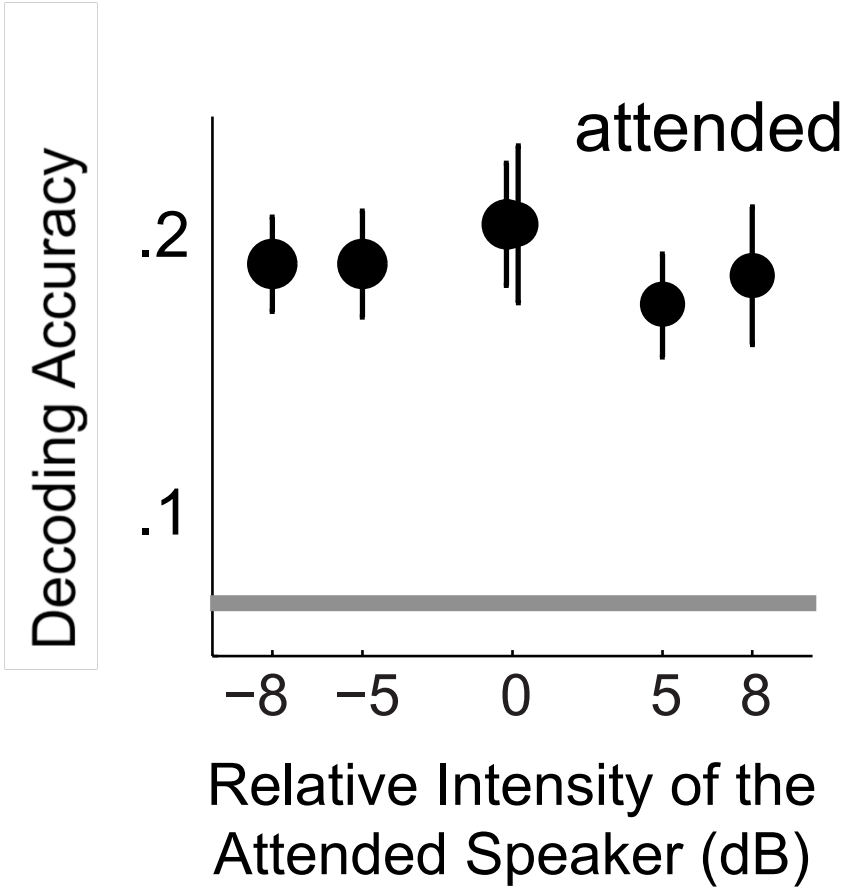




The correlation between stimulus and reconstructed envelope is a measure of neural encoding accuracy.

Ding & Simon, J Neurophys (2012)

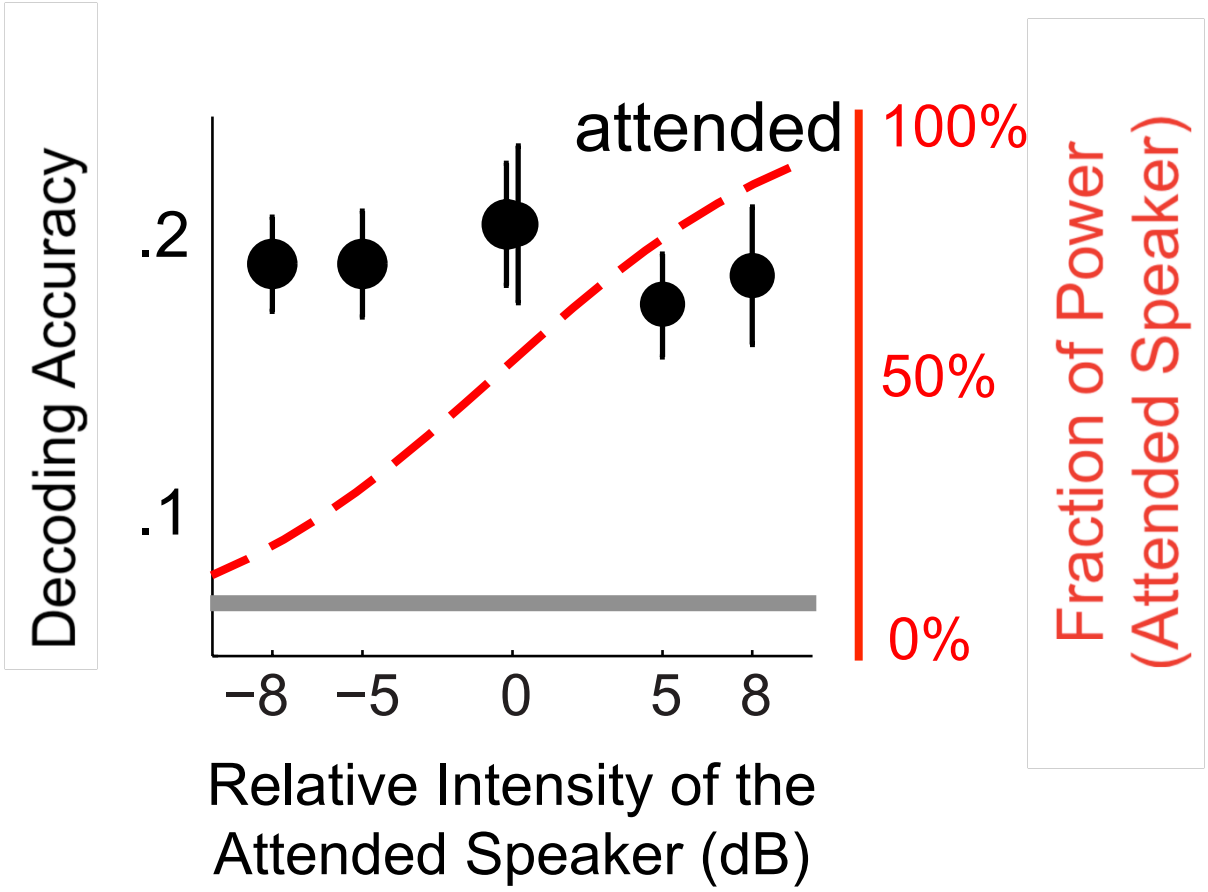
Decoding Speech Envelopes from the MEG Responses



Cortical representation of the attended speech is invariant to the relative intensity of the speaker.

(Ding & Simon, PNAS 2012)

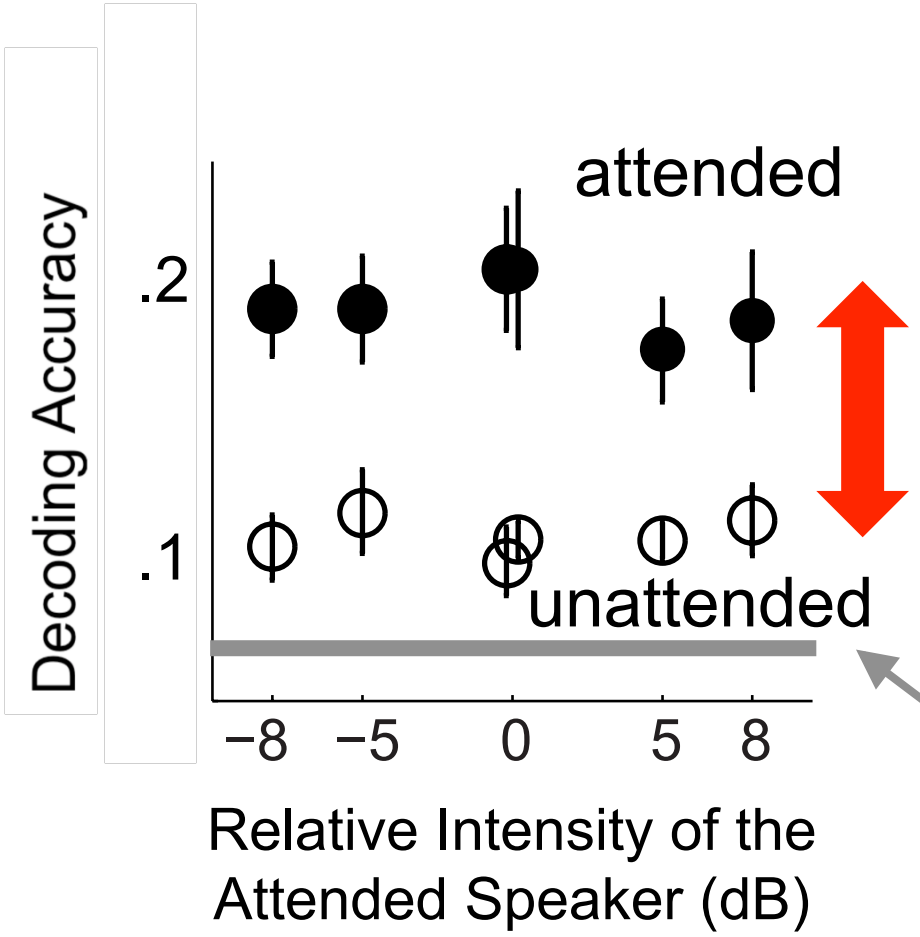
Decoding Speech Envelopes from the MEG Responses



Intensity of the speaker is compensated by the neural system.

(Ding & Simon, PNAS 2012)

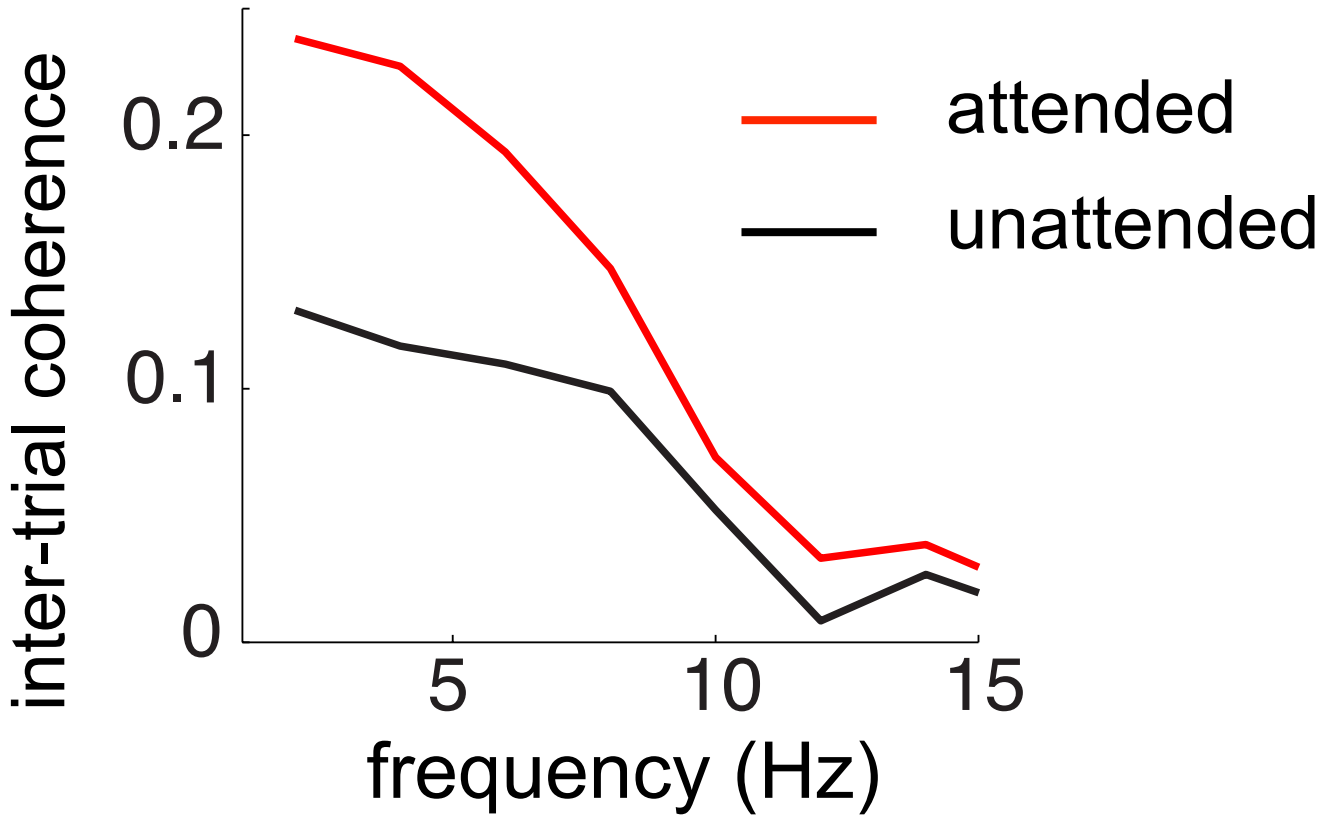
Decoding Speech Envelopes from the MEG Responses



The neural response carries more information about the attended speaker than the unattended speaker.

(Ding & Simon, PNAS 2012)

Phase Coherence Spectrum



Delta and theta band responses are both modulated by top-down attention.

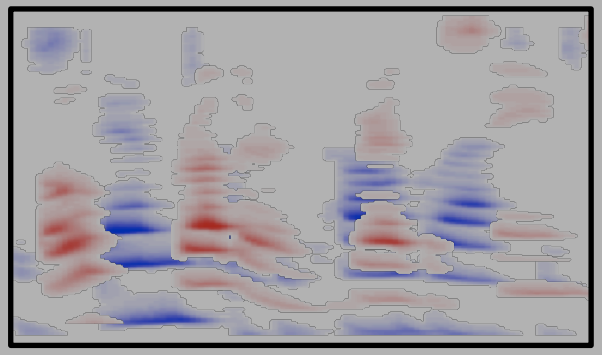
Cortical Entrainment Is Strongly Modulated by Top-down Attention

- Cortical activity is predominantly entrained to the attended speech stream, even when the competing stream is much louder.

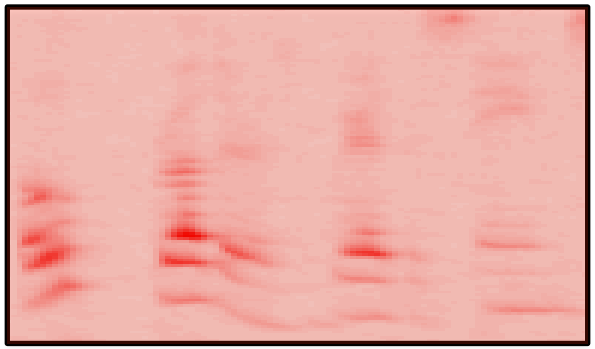
- Both delta and theta band responses are strongly modulated by attention.

Neural Encoding of Degraded Speech: Speech in Noise

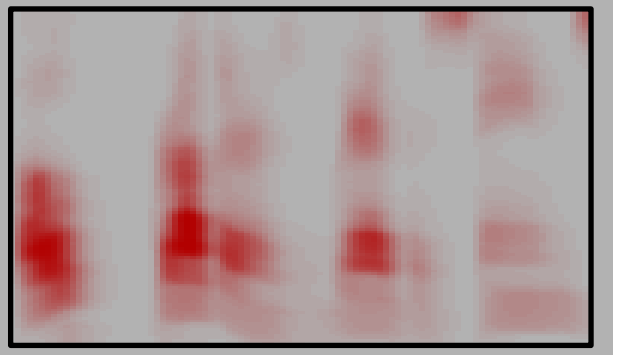
speech presented with a competing speech stream



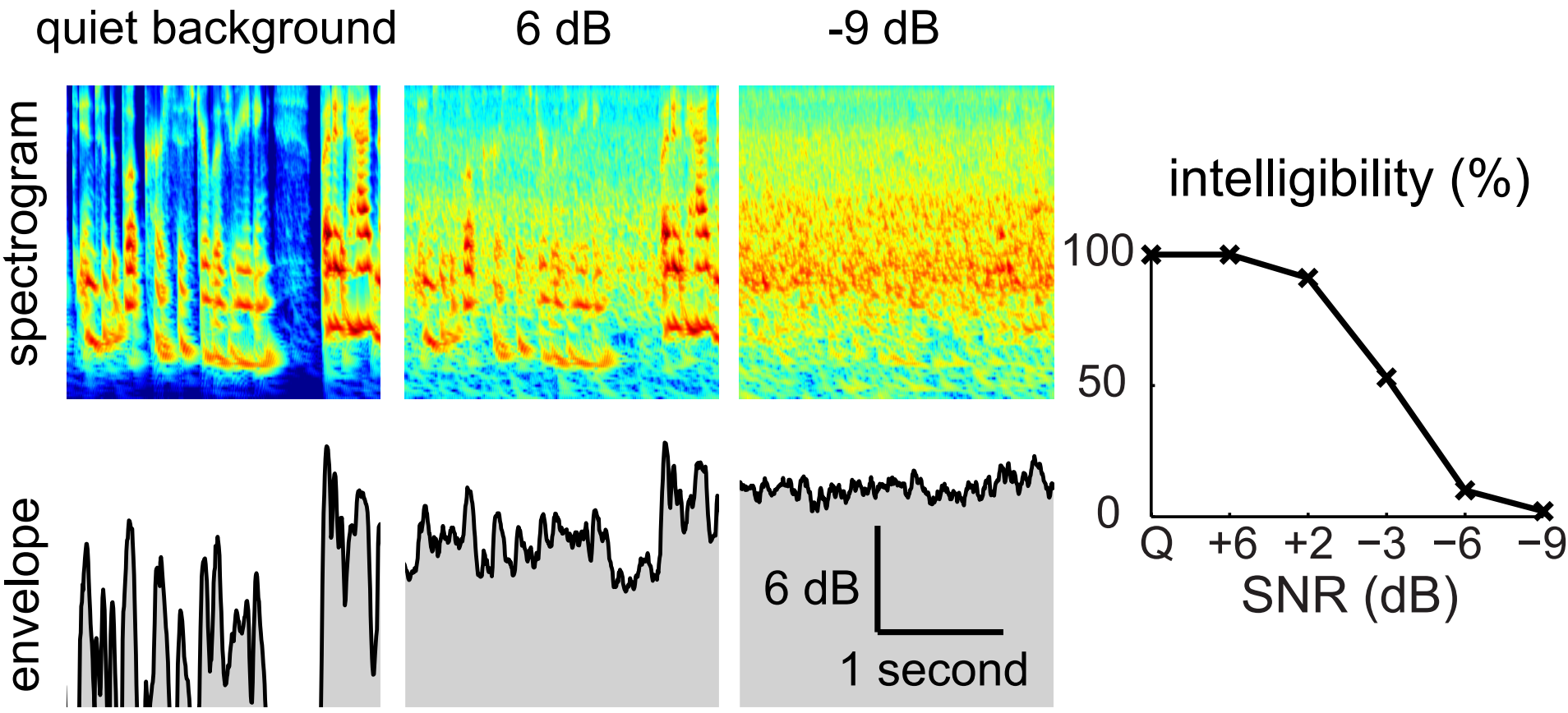
speech embedded in noise



speech with reduced spectral resolution

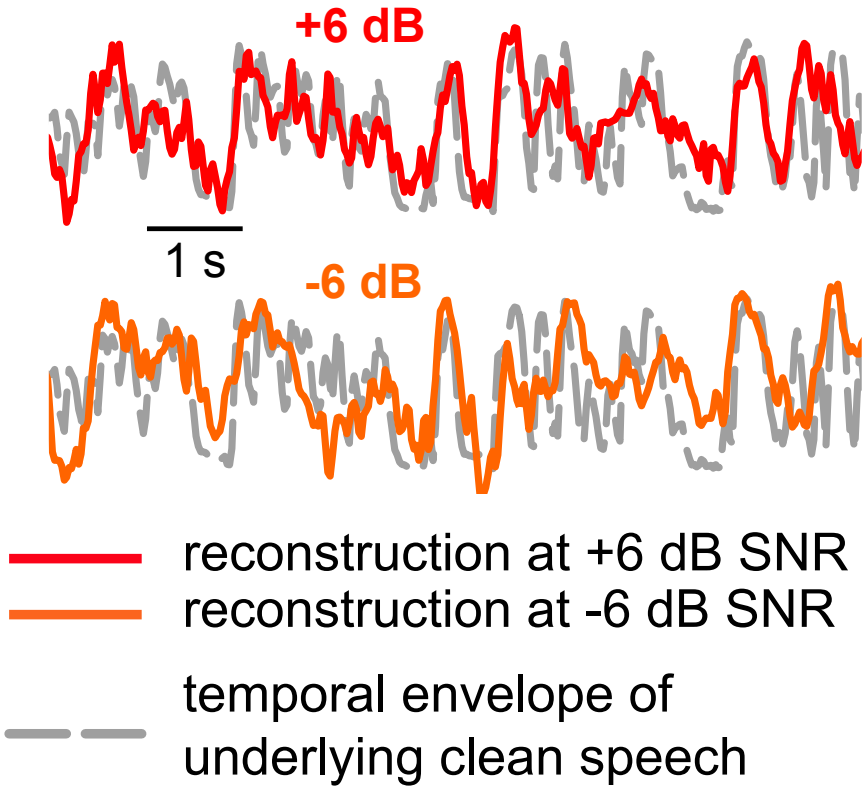


Stationary Noise Strongly Interferes with Speech

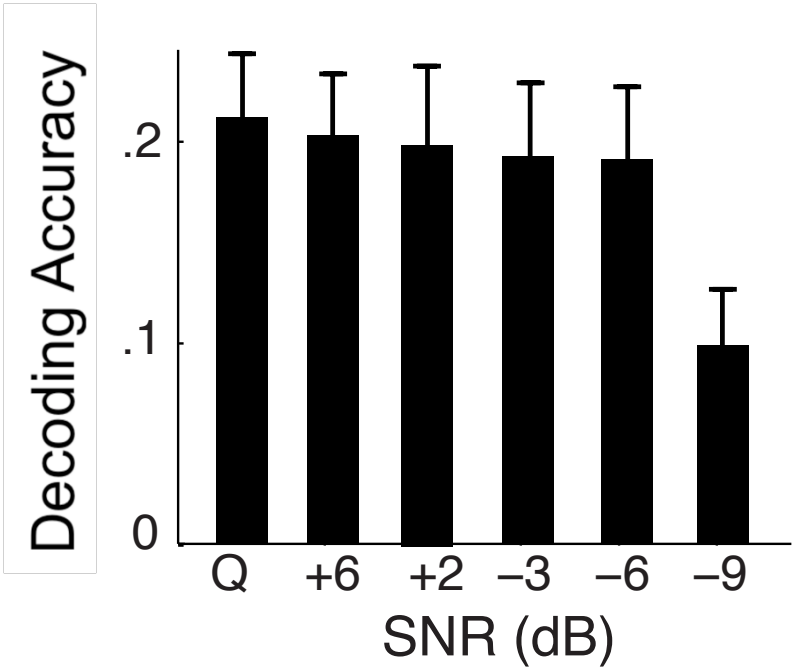


Neural Entrainment to Speech Is Robust to Noise

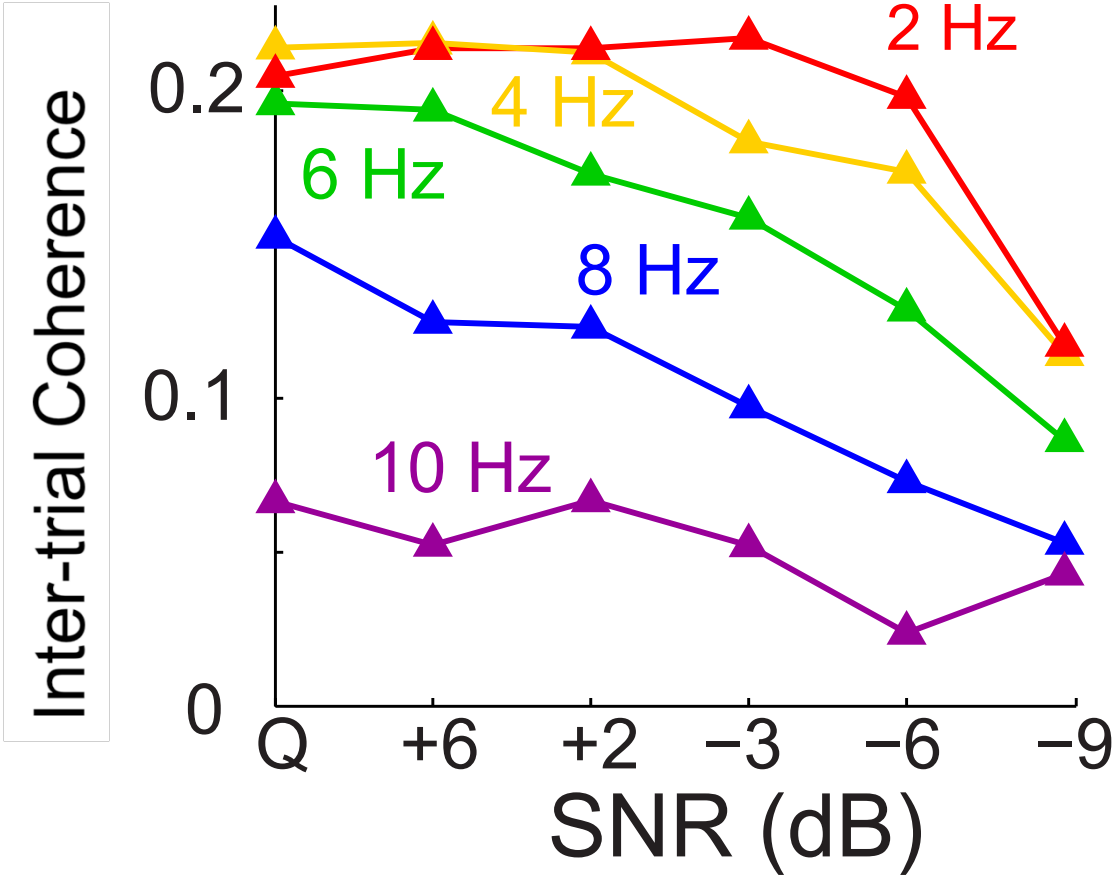
Examples of Neural Reconstructions



Reconstruction Accuracy

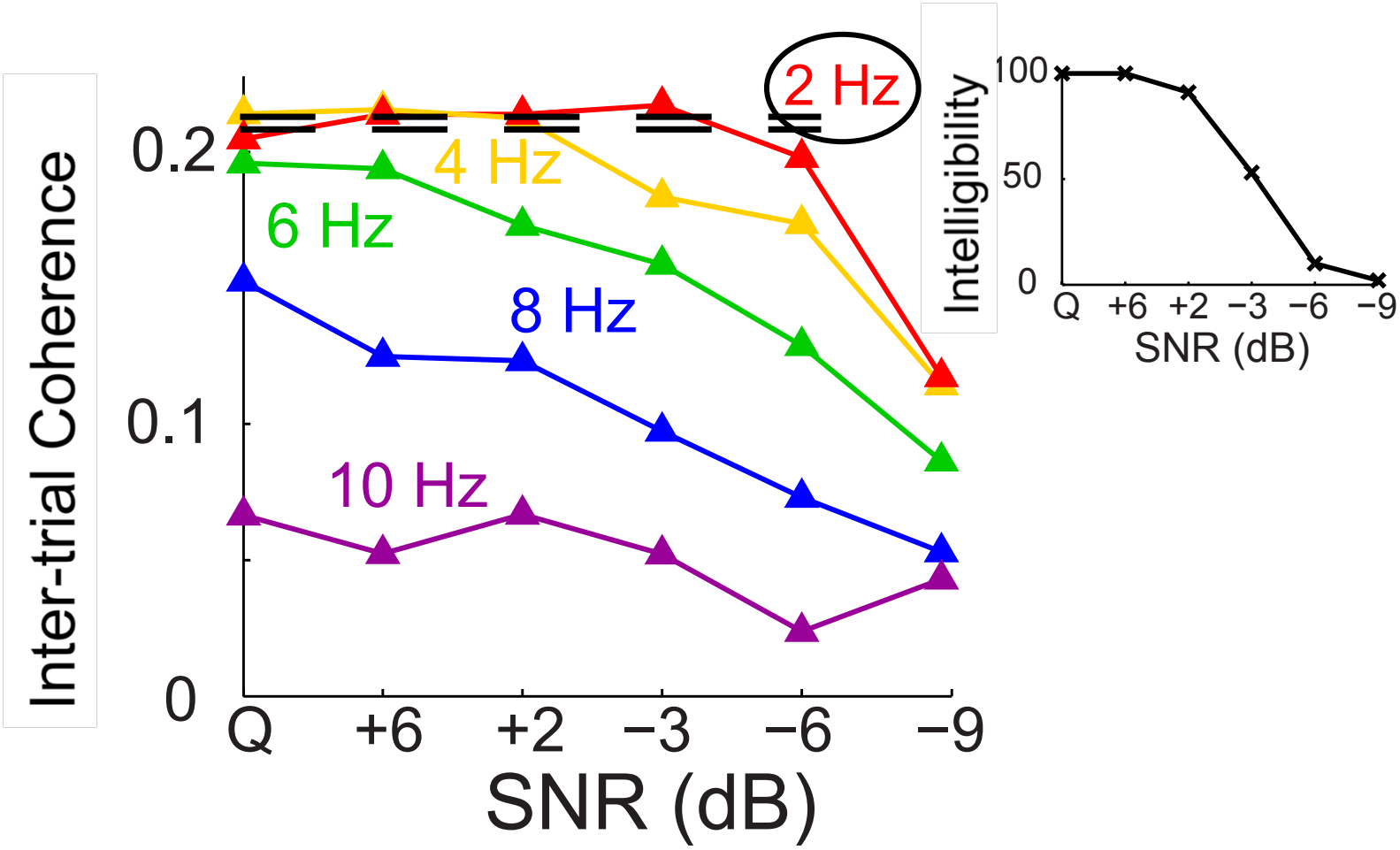


Frequency Dependent Noise Susceptibility



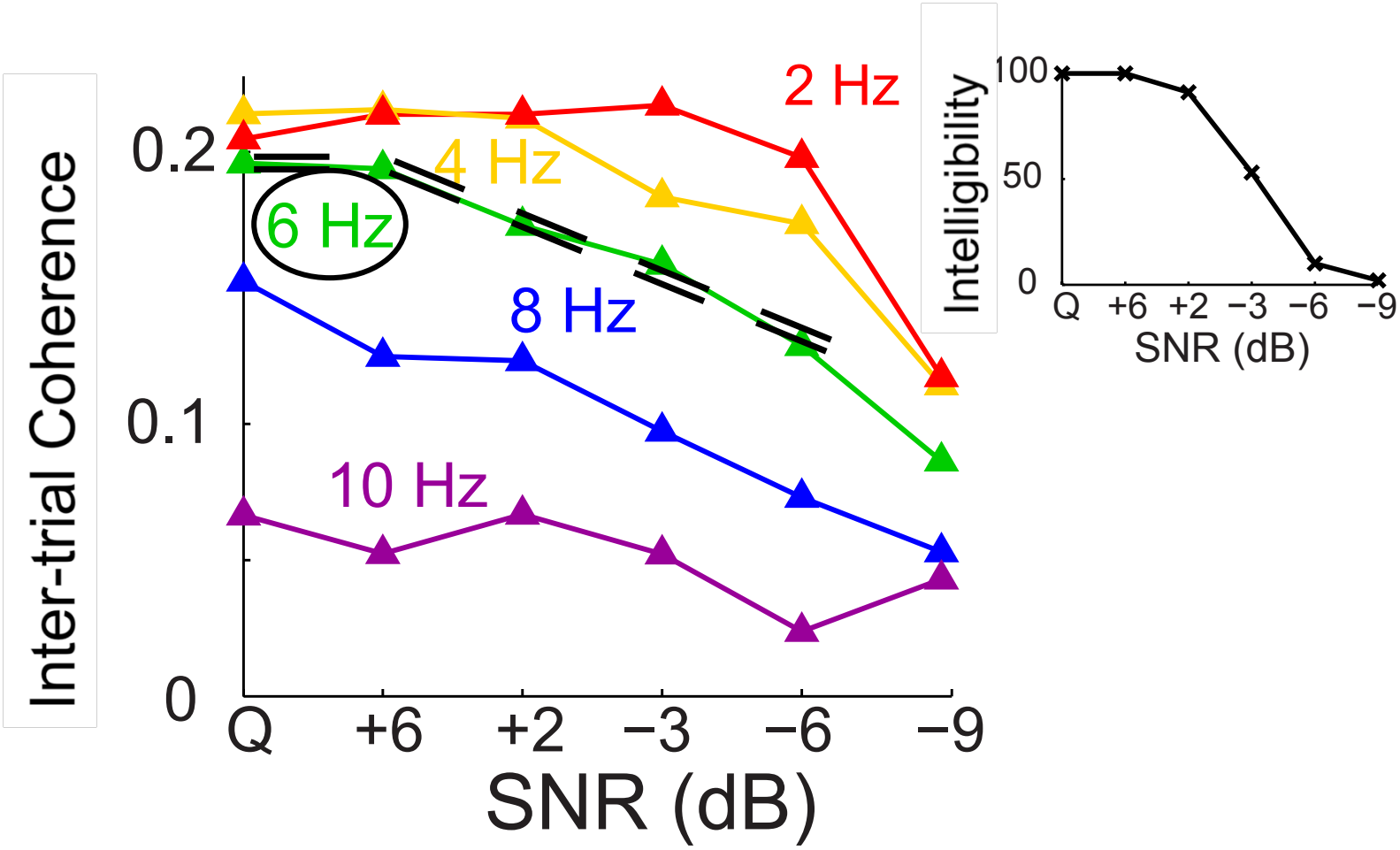
Frequency Dependency of Response Stability

Delta band activity is robust to noise.

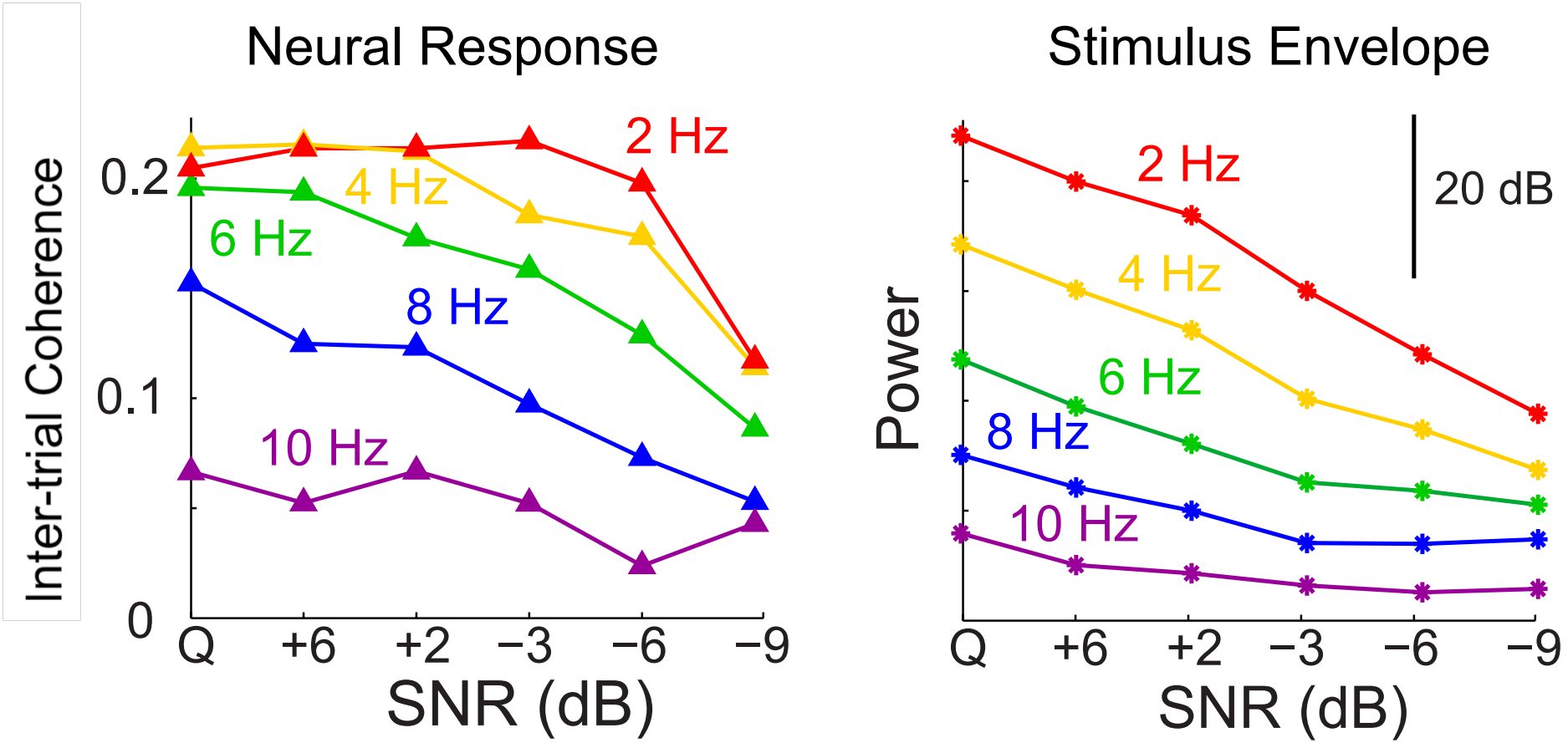


Frequency Dependency of Response Stability

Theta band activity is sensitive to noise.



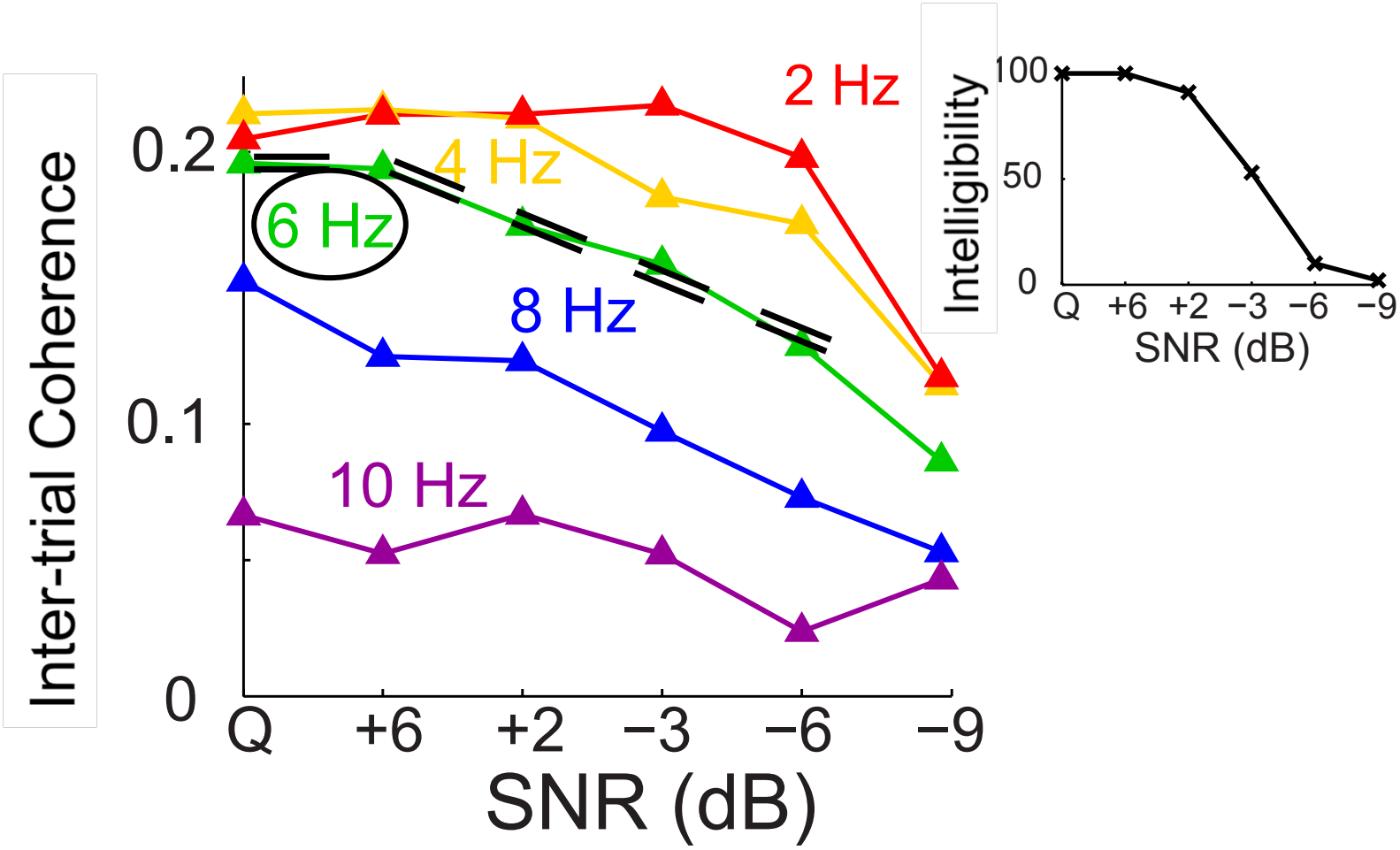
Response Coherence vs. Stimulus Power



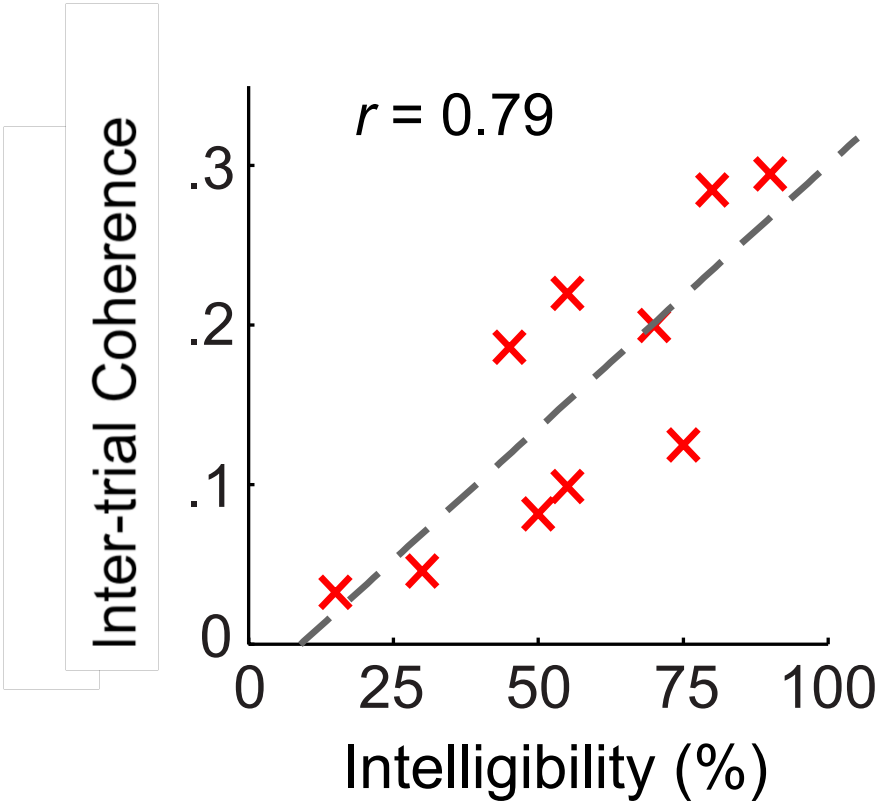
The frequency-dependent robustness of the neural response is acoustic properties of the stimulus.

Frequency Dependency of Response Stability

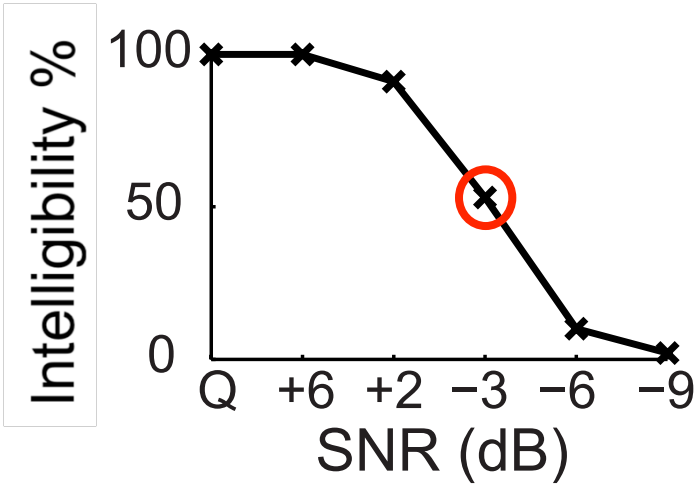
Theta band activity is sensitive to noise.



Delta Activity Predicts Individual Variance of Speech Score



In noisy environments, individual subjects' speech score is predicted by low-frequency (<4 Hz) neural entrainment to speech.



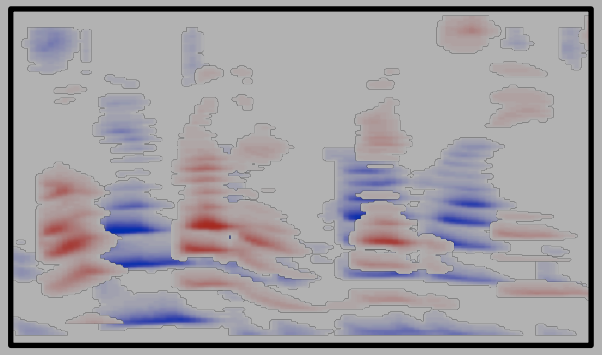
Cortical Entrainment to Speech Is Robust to Energetic Masking

- Delta band activity is largely invariant to background noise and predicts individual differences in speech perception.

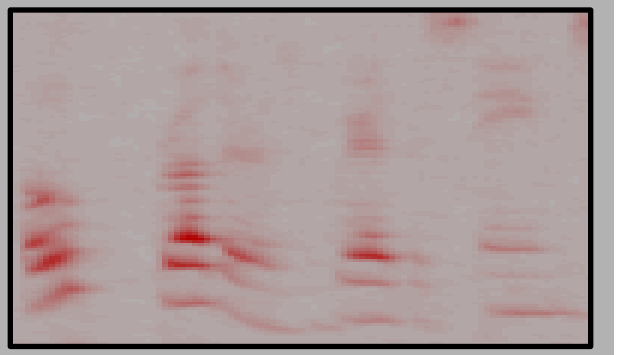
- Theta band activity reflects how speech intelligibility is affected by noise level.

Neural Encoding of Degraded Speech: Reduced Spectral Resolution

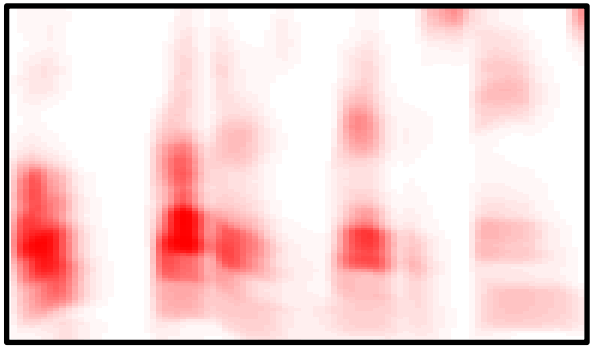
speech presented with a competing speech stream



speech embedded in noise

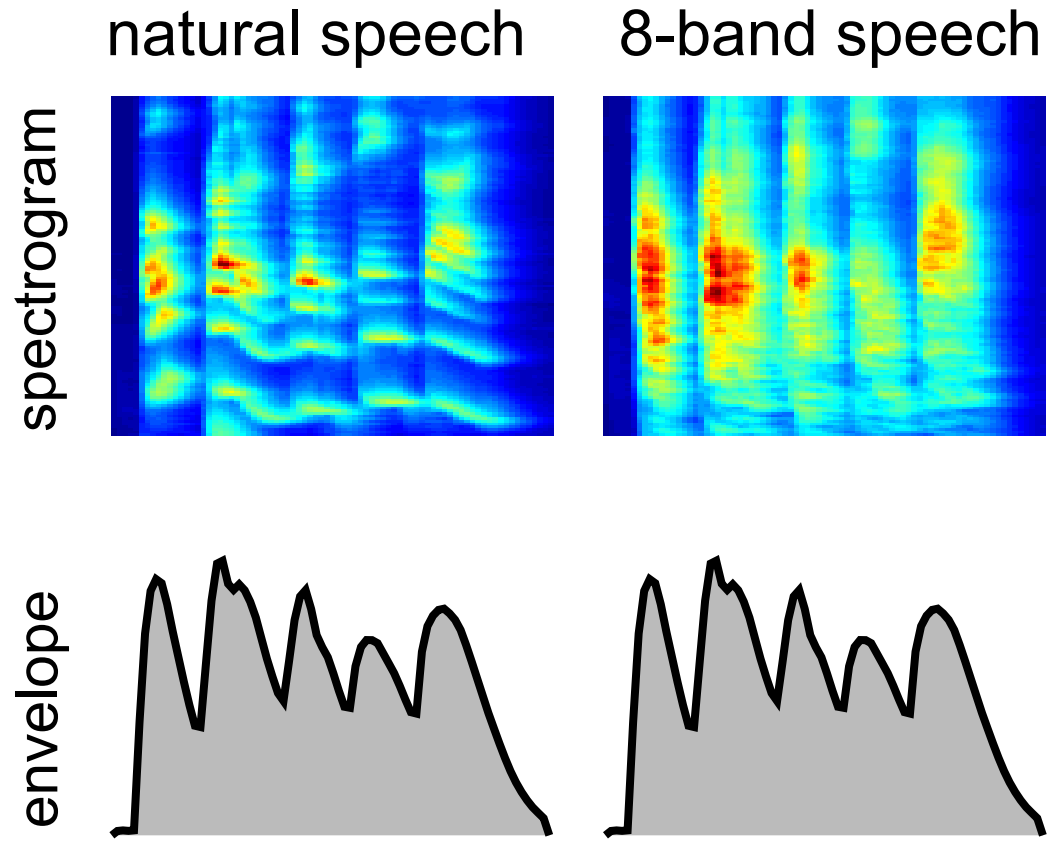


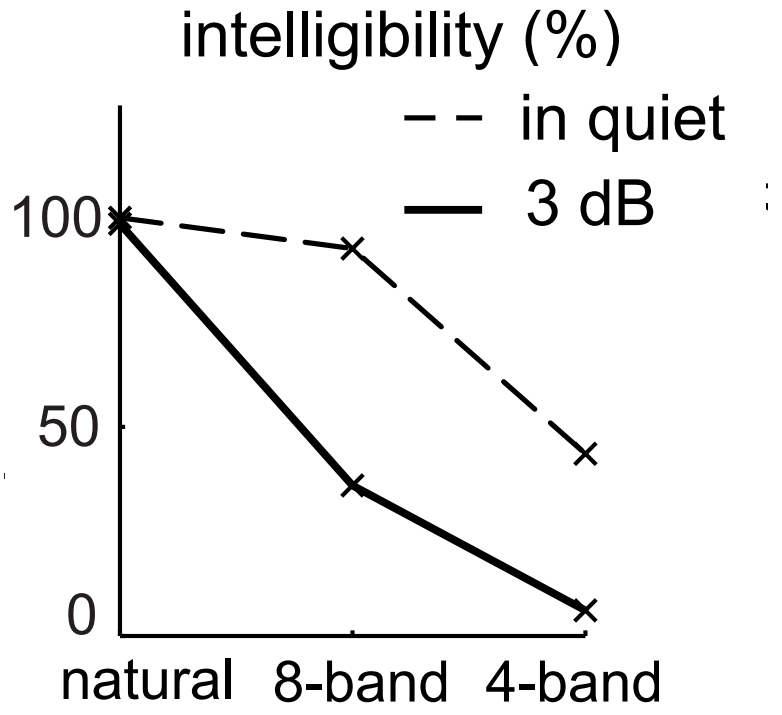
speech with reduced spectral resolution



Ding, Chatterjee & Simon, in prep

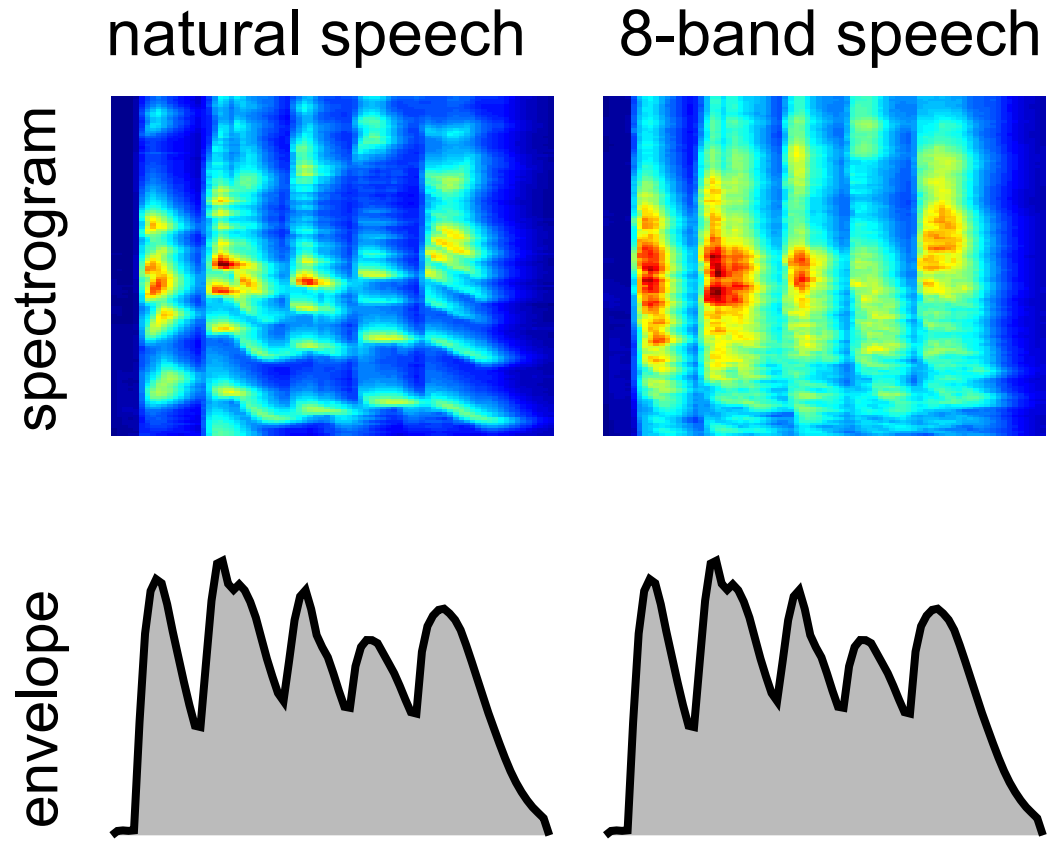
Reduce Spectral Resolution and Preserve Temporal Information

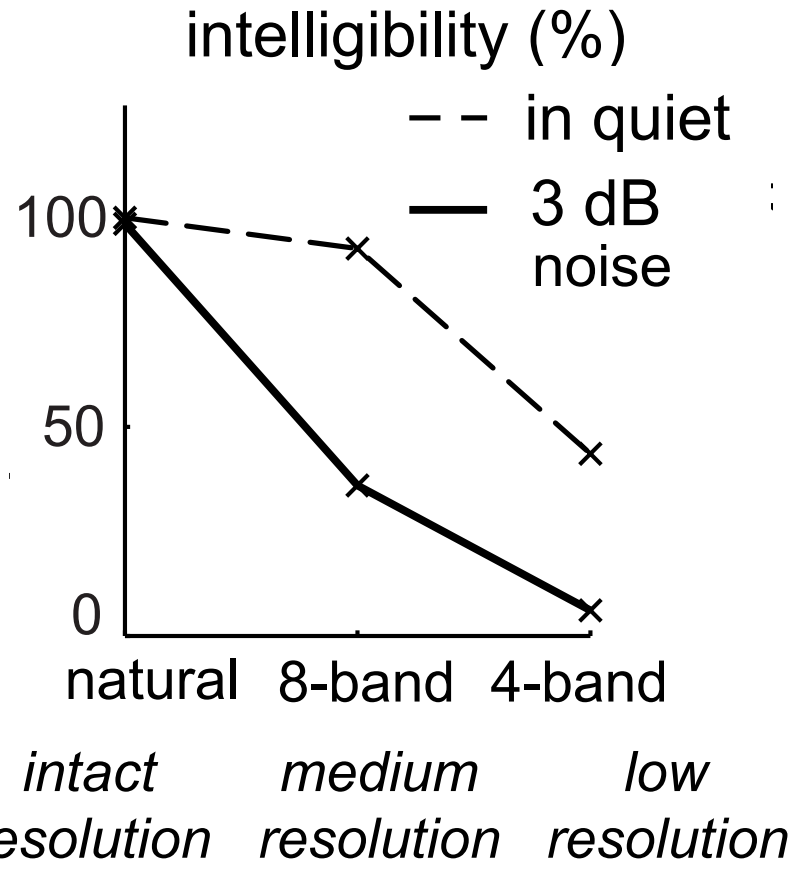




The spectral resolution of speech is reduced by noise band vocoding.

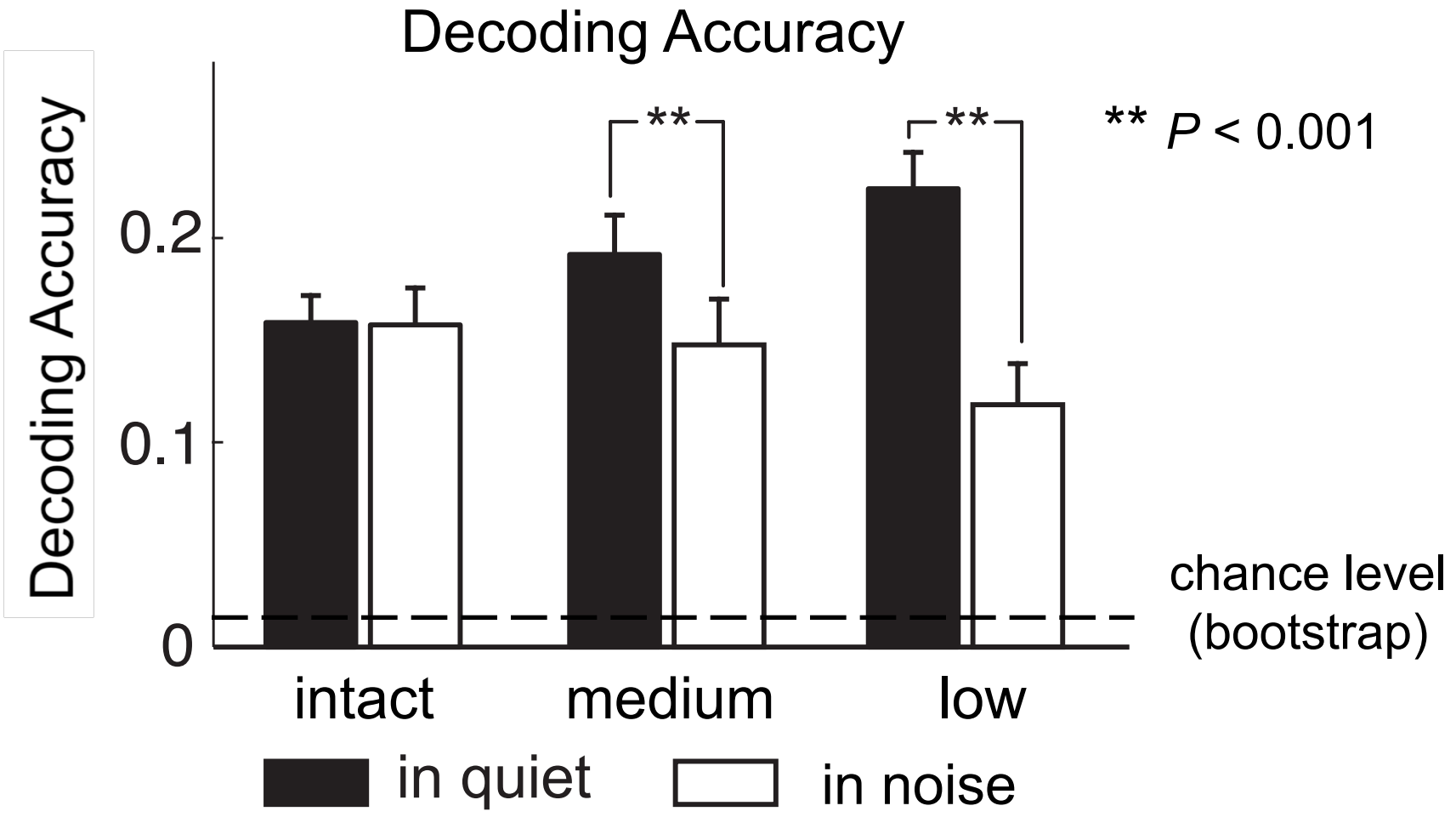
Reduce Spectral Resolution and Preserve Temporal Information



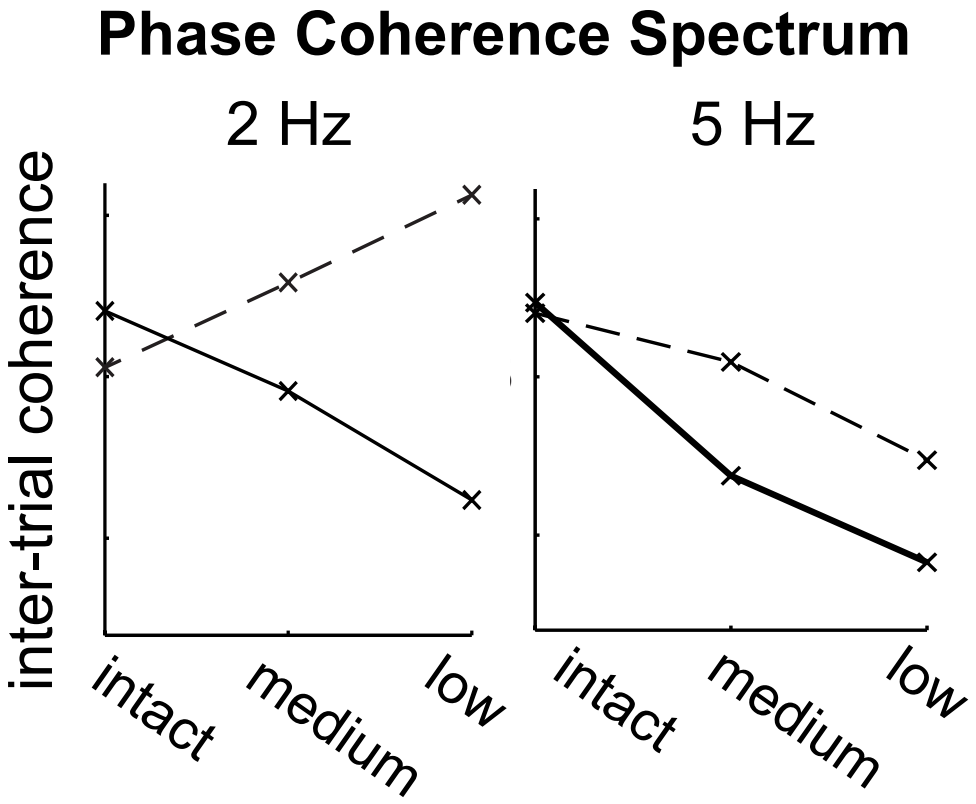


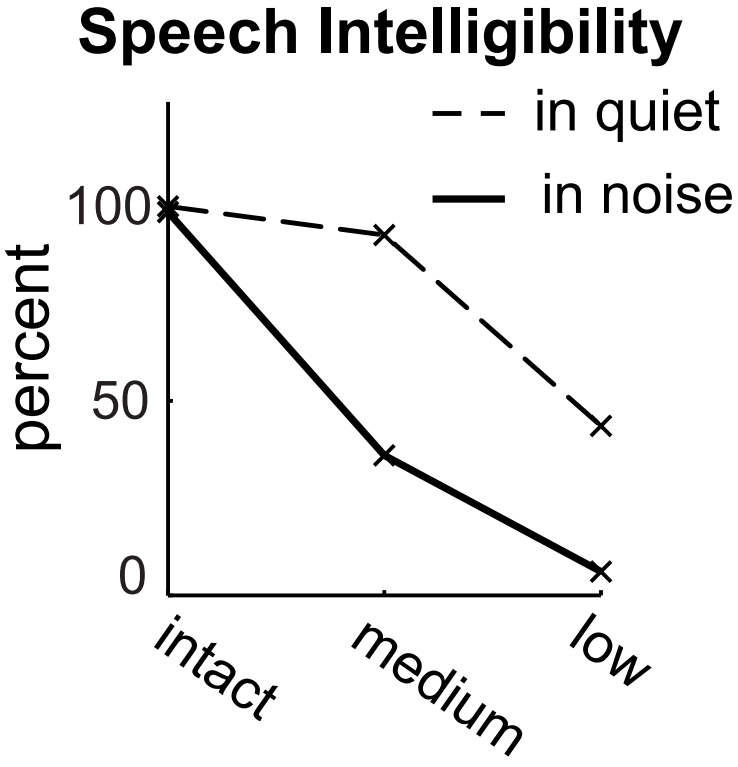
The spectral resolution of speech is reduced by noise band vocoding.

Noise-robust Cortical Entrainment Requires Fine Spectral Resolution

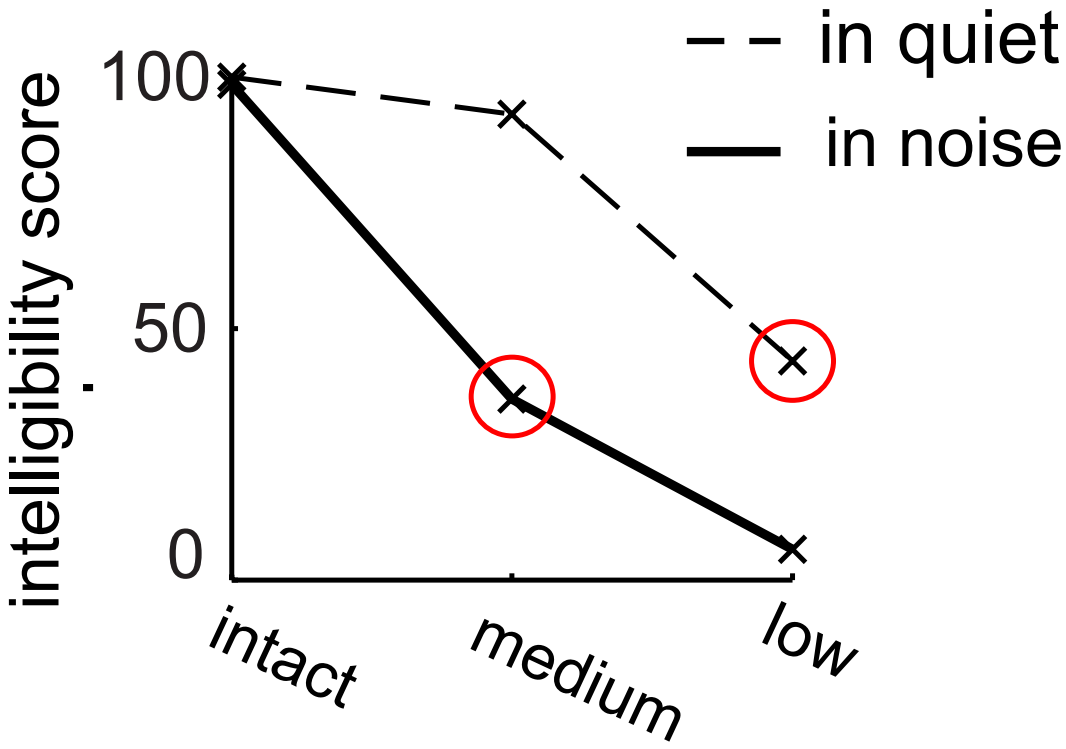


Theta but not Delta Band Activity Matches Speech Score

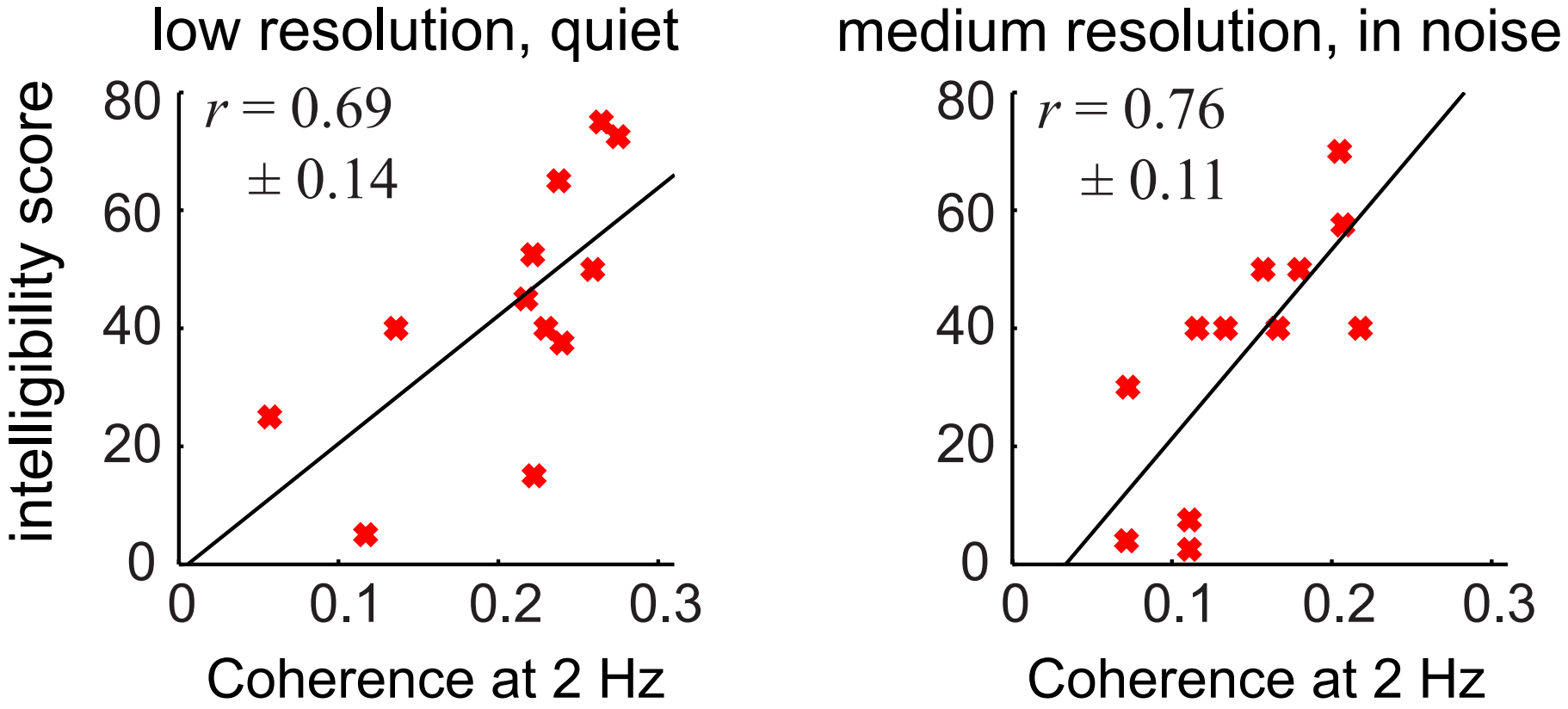




Predictors for Individual Differences?



Synchronization in Delta Band Predicts Individual's Intelligibility Score



Subjects showing better neural synchronization to speech tend to understand speech better.

Spectral Resolution of Speech Influences Cortical Synchronization

- Robust cortical synchronization to speech requires an interplay between spectral and temporal processing.
- Delta band activity reflects individual's speech recognition ability.
- Theta band activity reflects how speech intelligibility is affected by acoustic degradation.

Summary (I)
Object-based Neural Representation

- In auditory cortex, there is a neural representation of the perceived auditory object, which is distinguishable from the physically presented auditory scene.

Summary (II)
A Hierarchy of Cortical Processing

- Delta band activity is very robust to acoustic degradation, and predicts individual differences in speech perception.

  (the perception/detection of an auditory object)

- Theta band activity reflects how speech intelligibility is affected by acoustic degradation.

  (the amount of decodable speech information)

- Thanks!

- Simon & de Chevinge