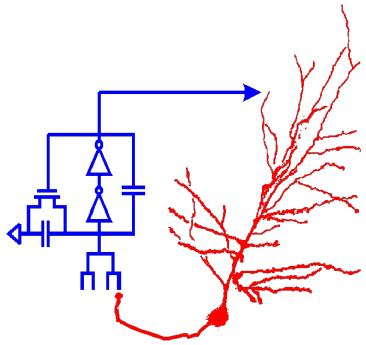
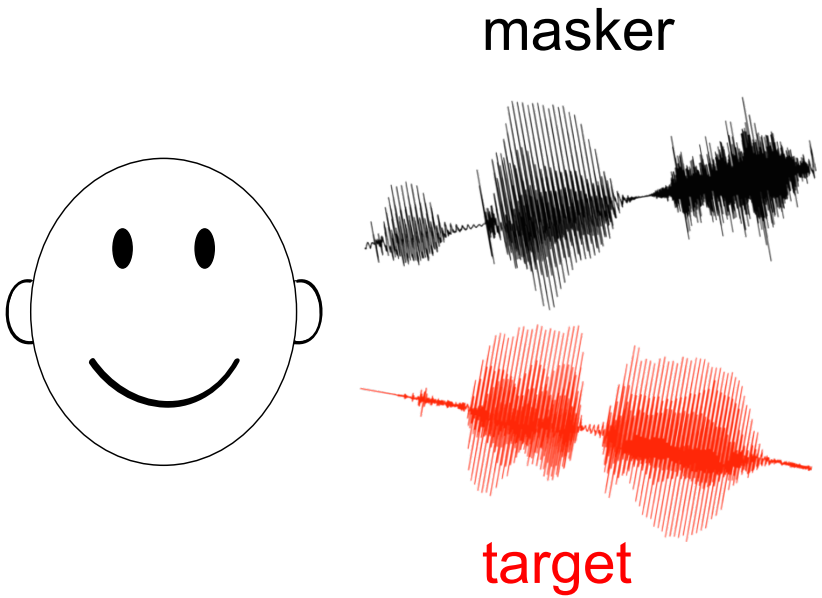
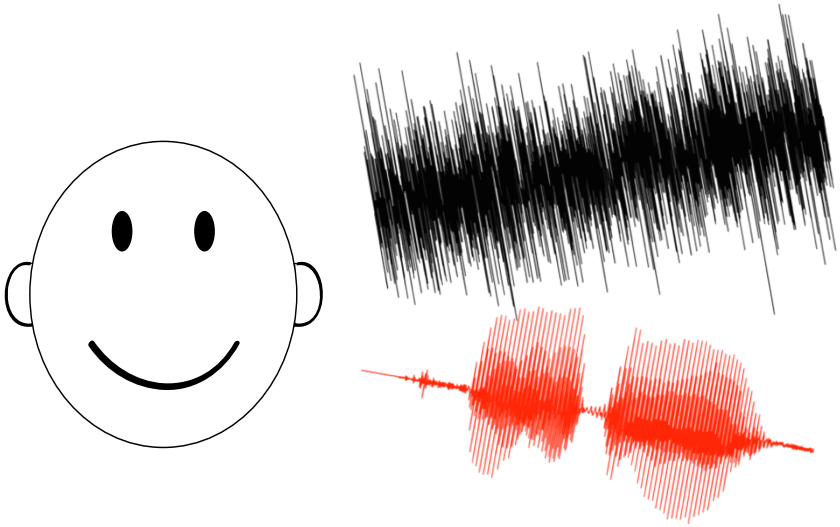


Robust Cortical Encoding of Slow Temporal Modulations of Speech

Nai Ding, Jonathan Z. Simon
University of Maryland, College Park

Speech Listening in Challenging Listening Environments

Speech Masker	Noise Masker
	
strong informational masking (**top-down selection** needed)	strong energetic masking (**bottom-up saliency** reduced)

Experiments

Speech Masker

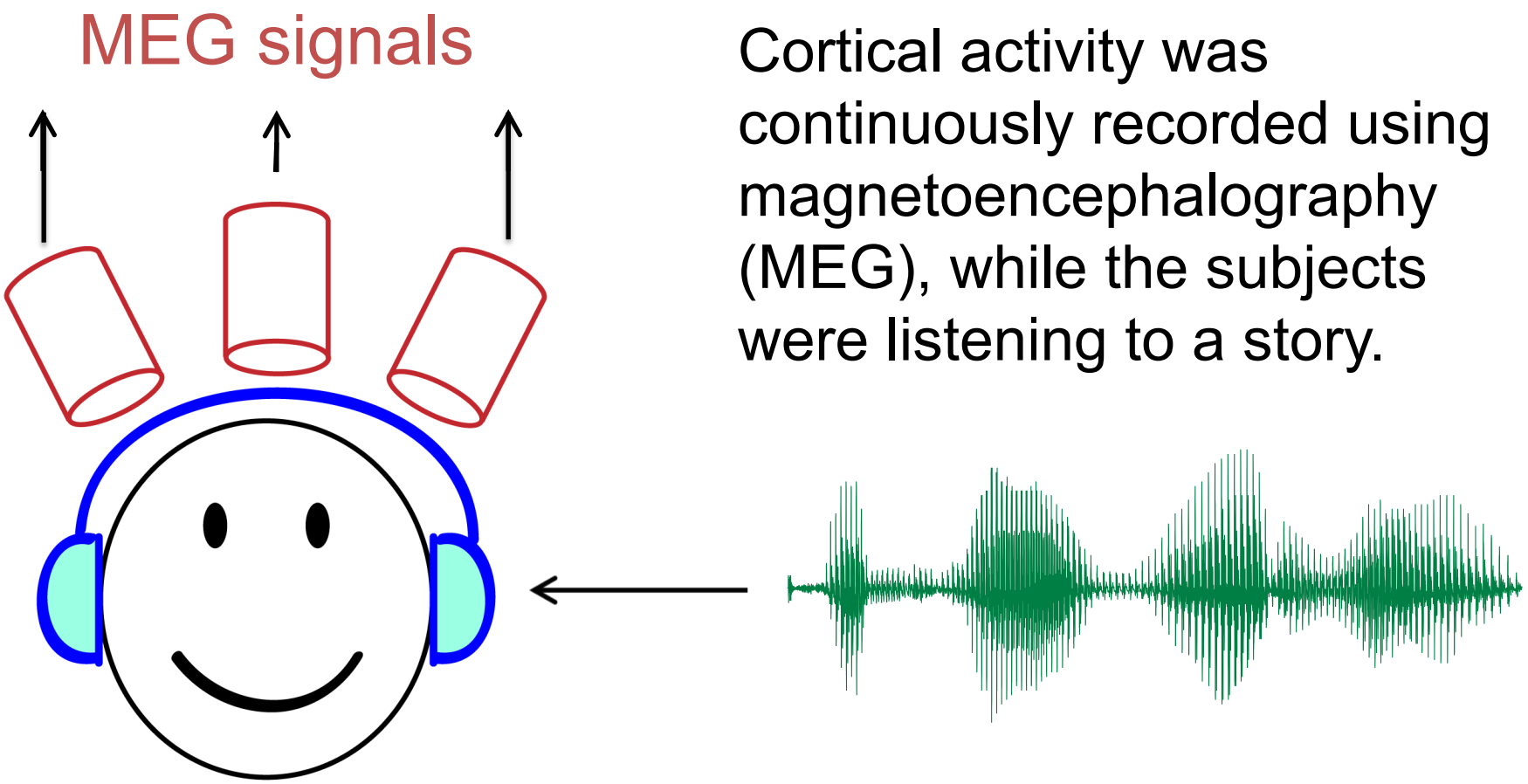
Stimulus:
- a male speaker
- a female speaker
- mixed into a single channel
- The listeners were asked to attend to one speaker.

Noise Masker

Stimulus:
- a female speaker
- spectrally matched stationary noise
- mixed into a single channel

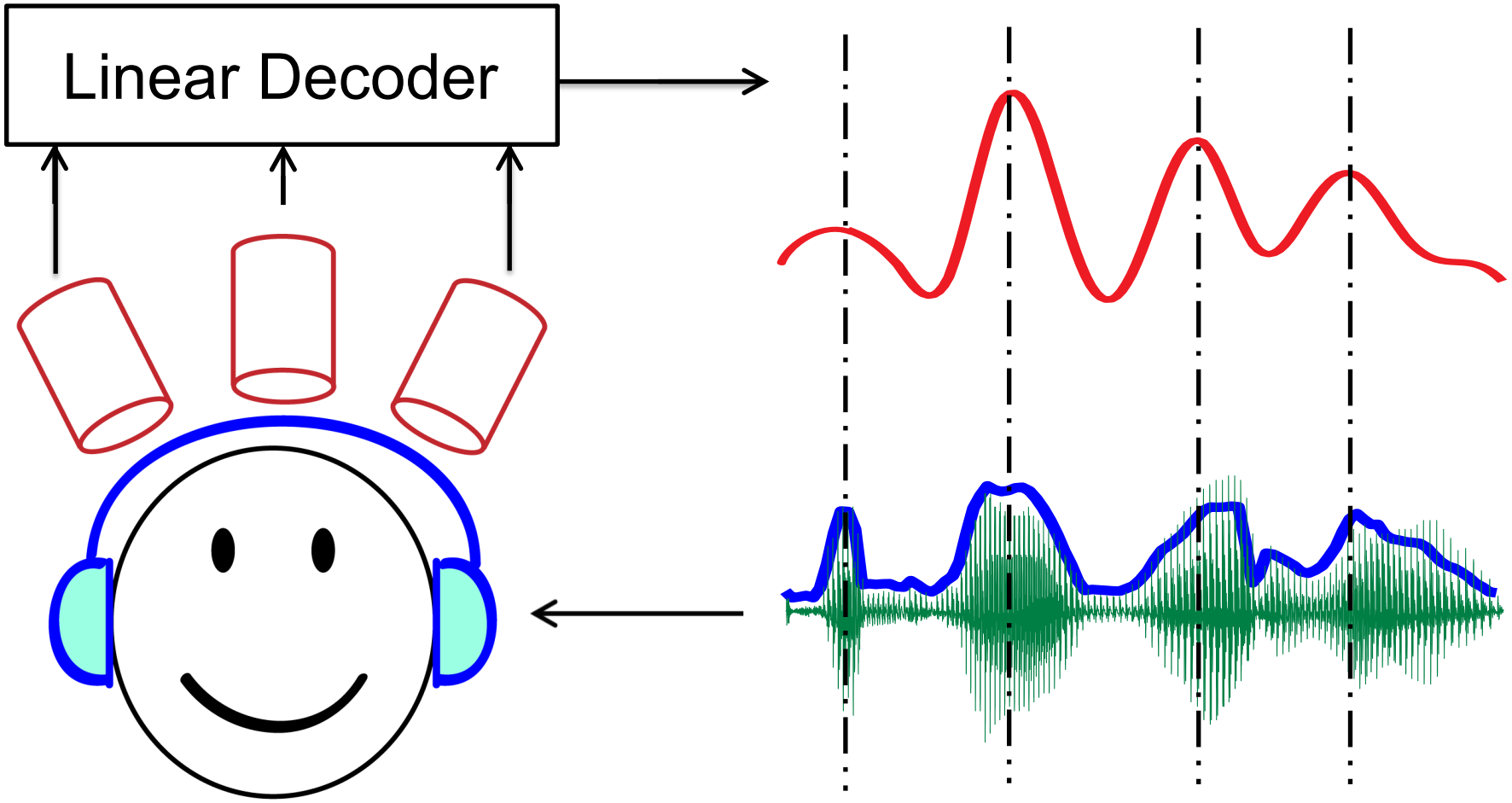
Procedure: The subjects listened to a story in 1-minute sections. They answered comprehension questions, and rated speech intelligibility, after each section.

Neural Recording



Cortical activity was continuously recorded using magnetoencephalography (MEG), while the subjects were listening to a story.

The MEG Response Is Phase Locked to the Speech Envelope

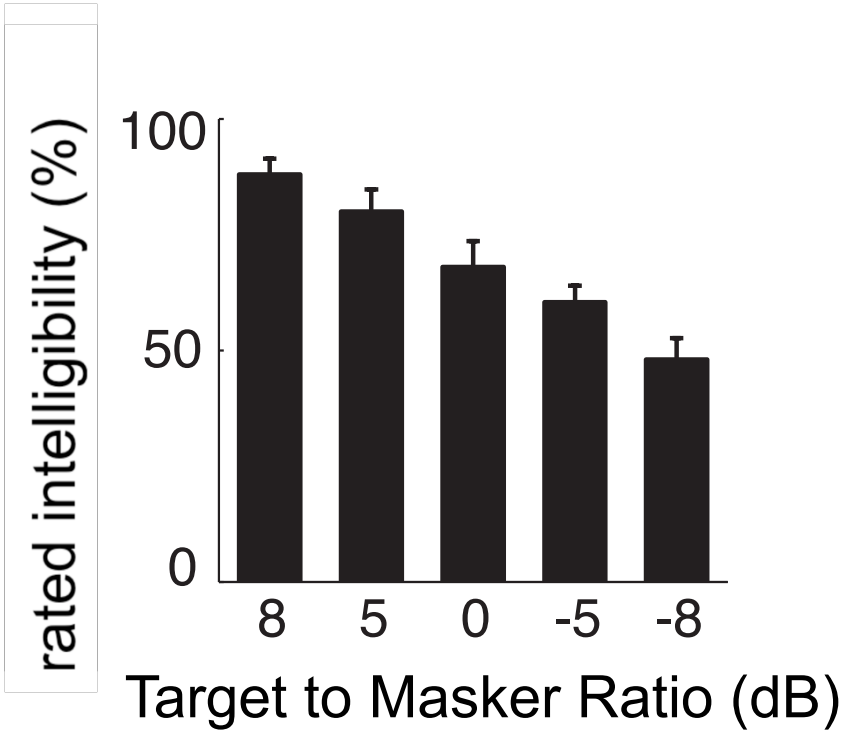


phase locking to slow temporal modulations < 10 Hz

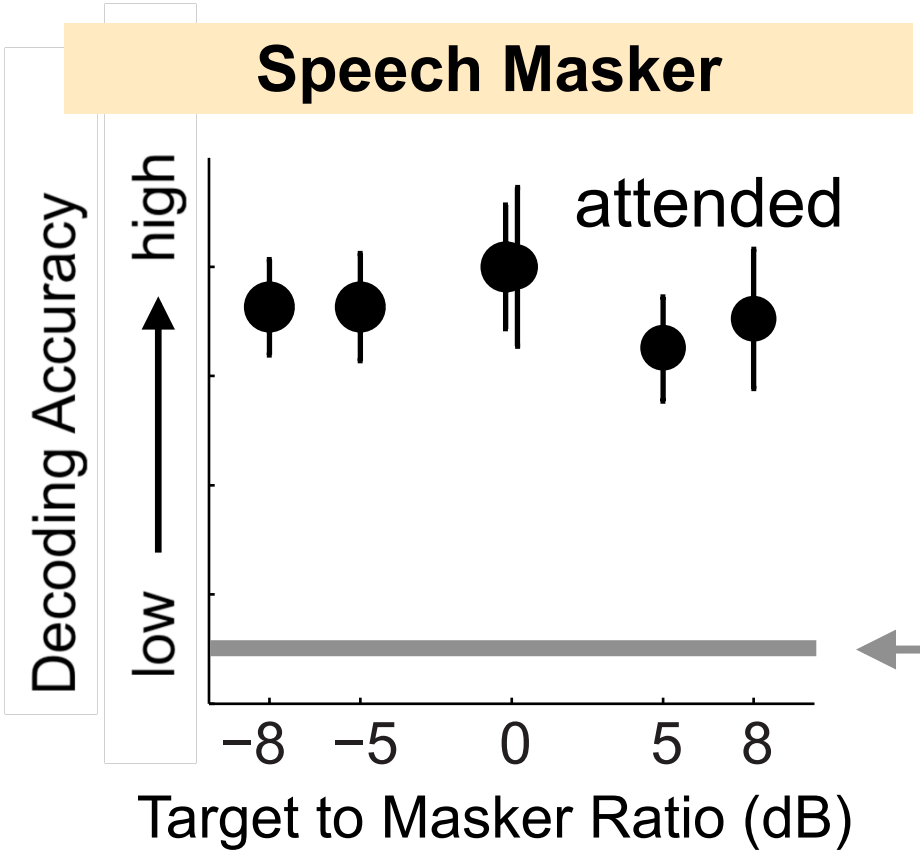
(Ding & Simon, J Neurophysiol. 2012)

Rated Intelligibility

Speech Masker



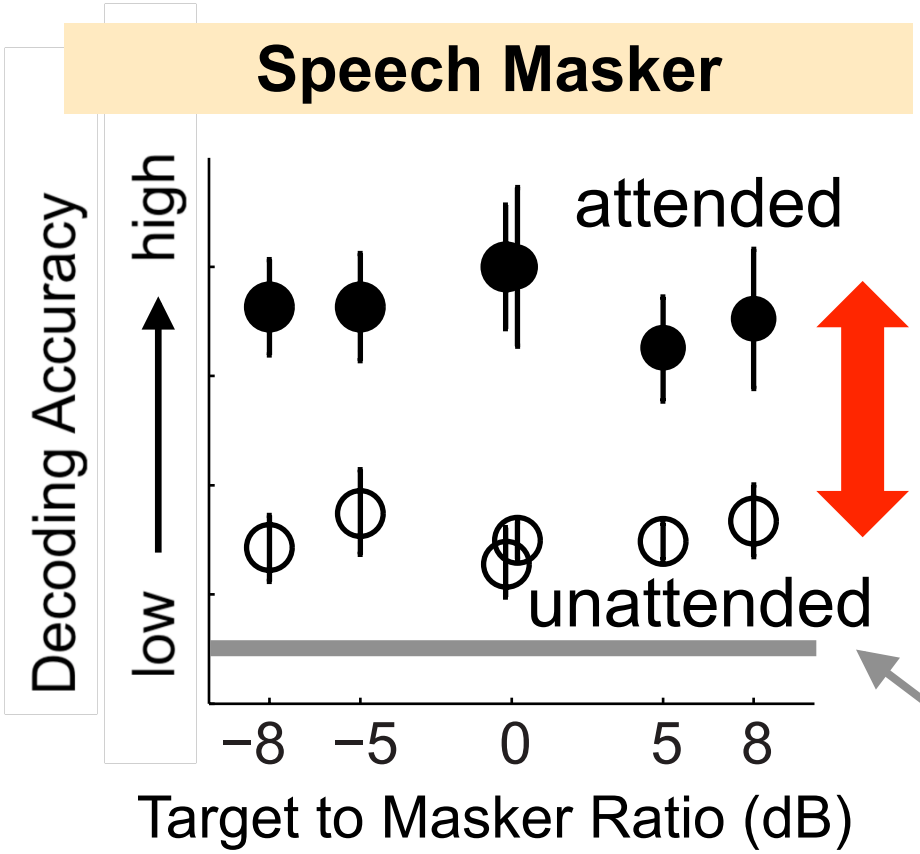
Decoding Speech Envelopes from the MEG Responses



The neural representation is unaffected by TMR.

(Ding & Simon, PNAS 2012)

Decoding Speech Envelopes from the MEG Responses

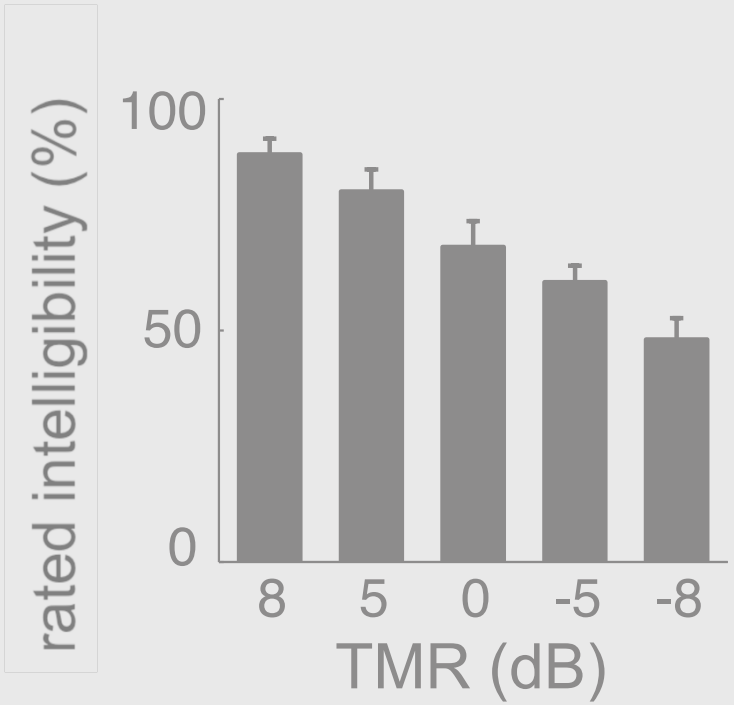


The neural response carries more information about the attended speaker than the unattended speaker.

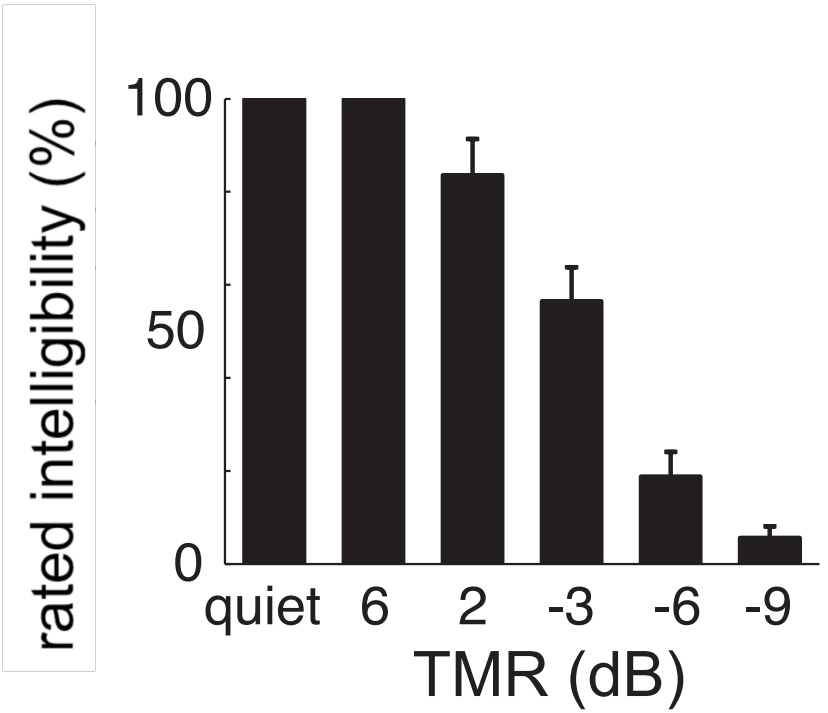
(Ding & Simon, PNAS 2012)

Rated Intelligibility

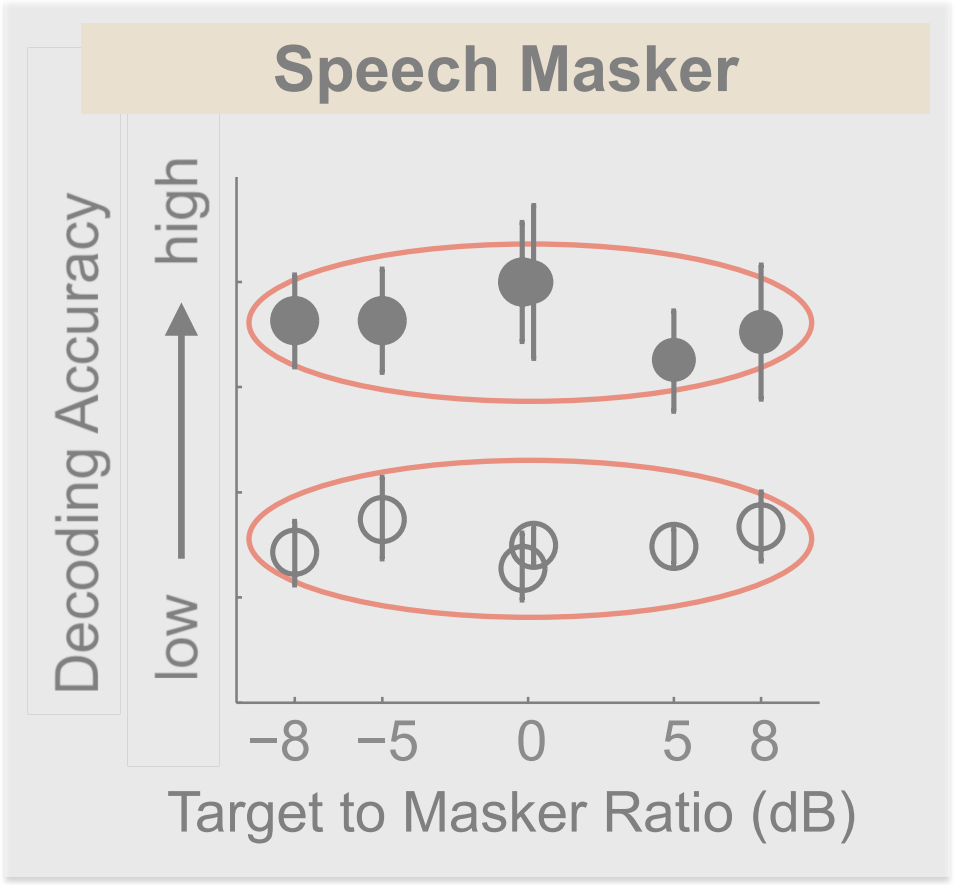
Speech Masker

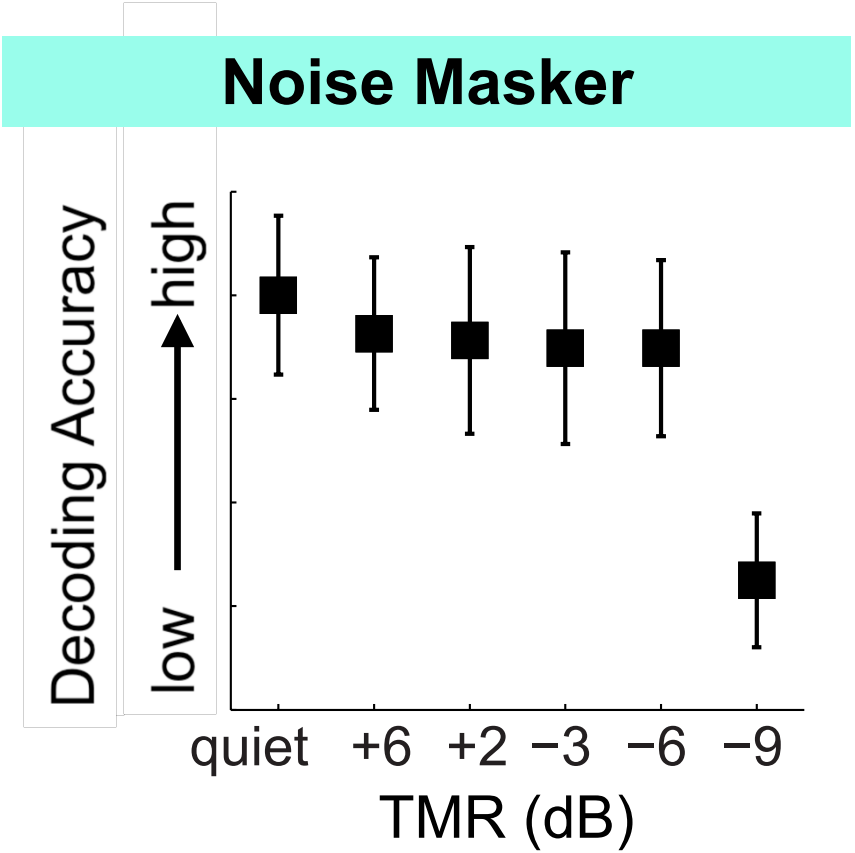


Noise Masker



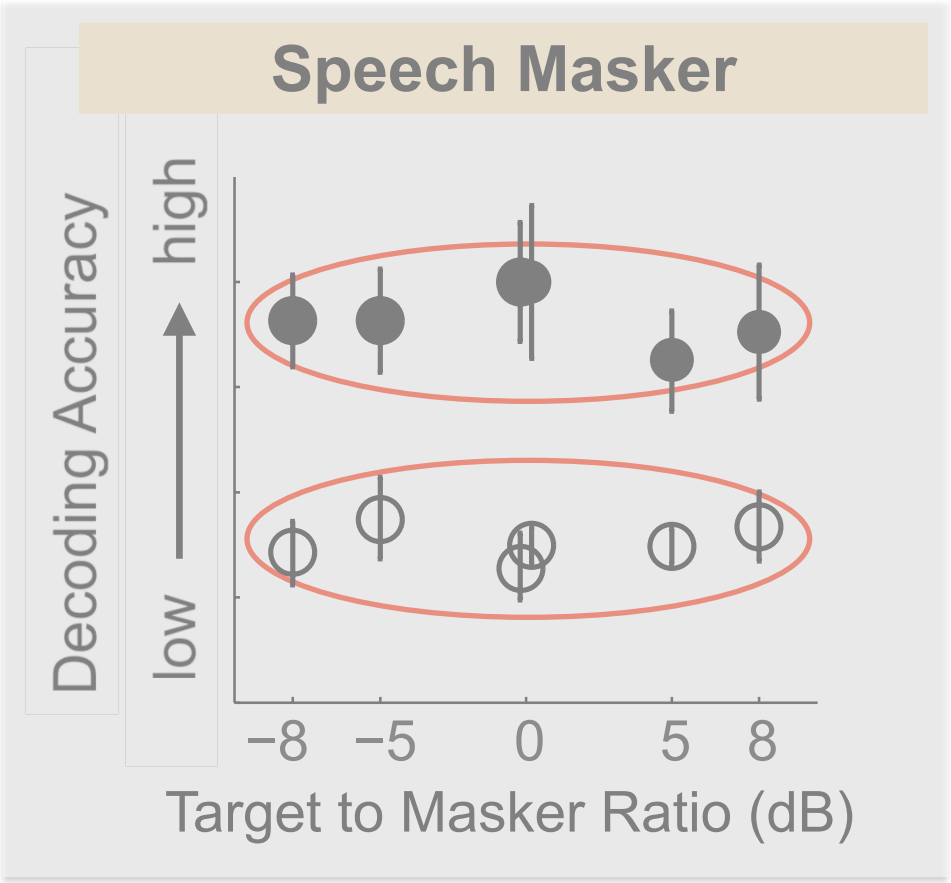
Neural Reconstruction

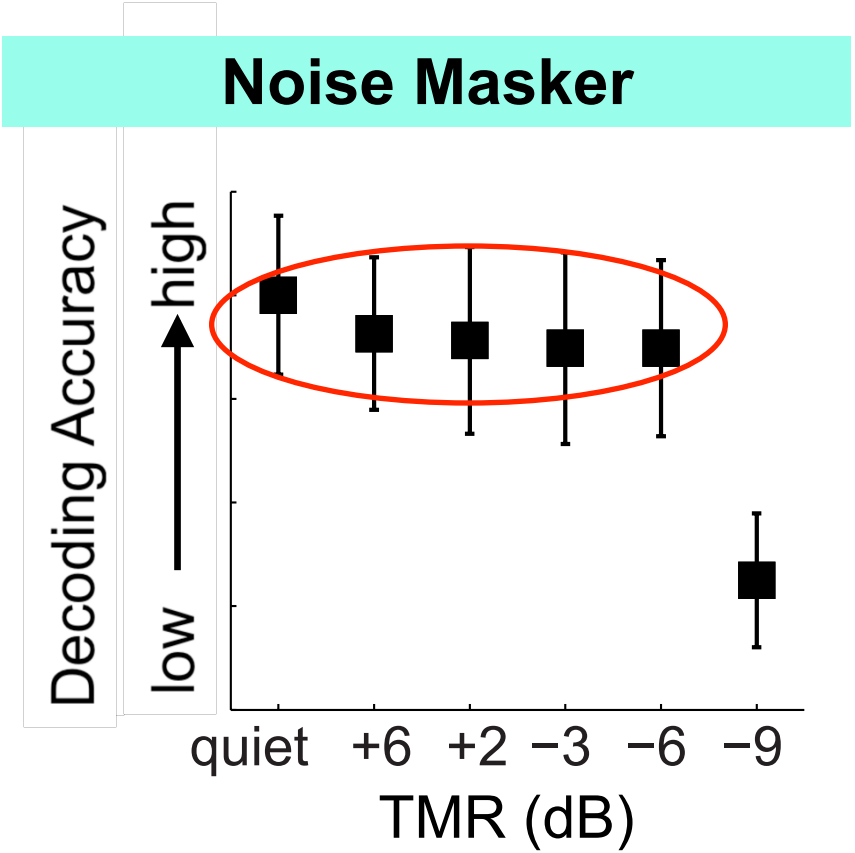




The neural response is not significantly affected by the TMR, for either speech or noise masker.

Neural Reconstruction



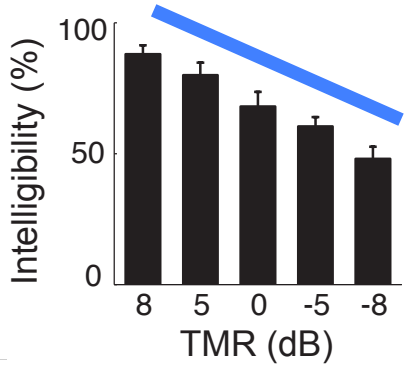


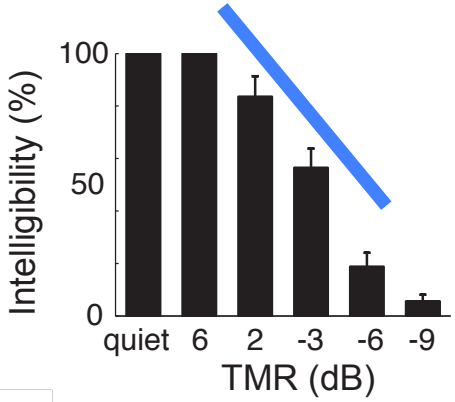
The neural response is not significantly affected by the TMR, for either speech or noise masker.

Neural Encoding Accuracy & Speech Intelligibility

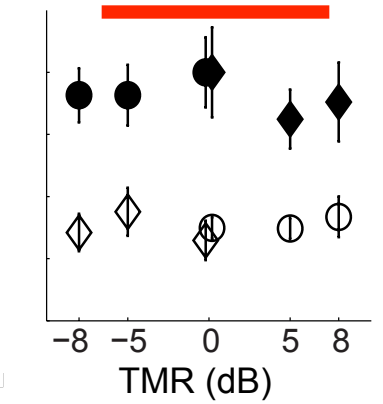
Speech Masker | **Noise Masker**

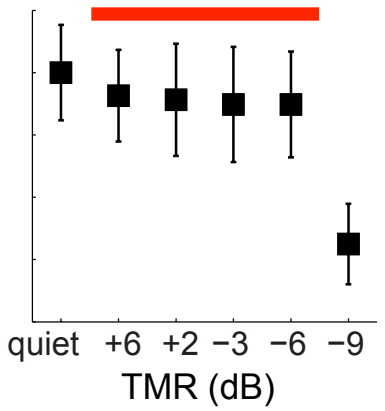
Rated Intelligibility





Neural Encoding Accuracy

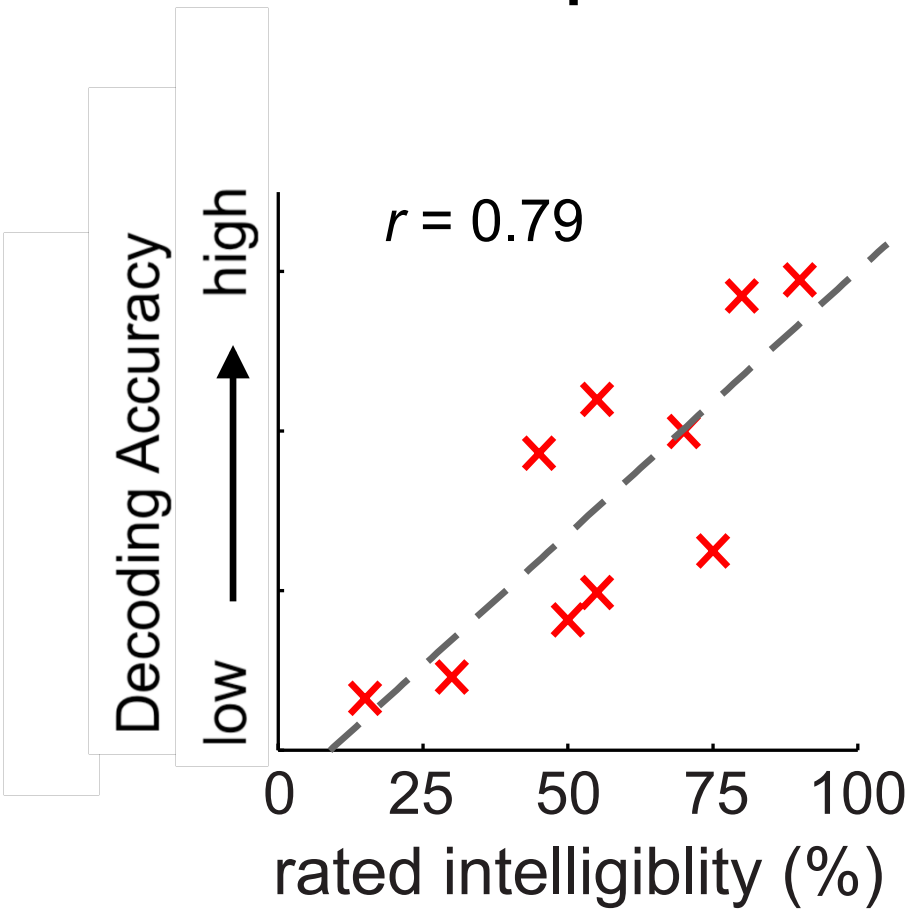




The neural synchronization to speech envelope is more robust to acoustic interference than speech intelligibility.

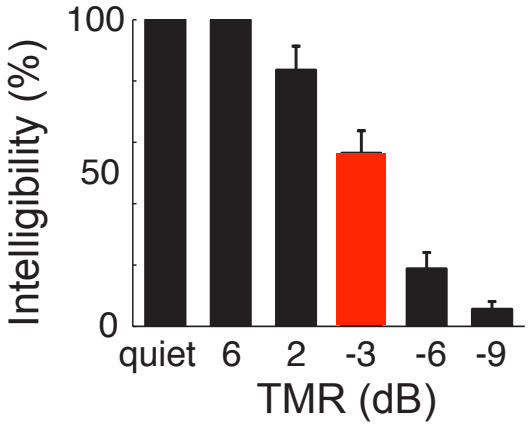
It may reflect the detection of syllabic structure rather than phonetic categories.

Neural Encoding Accuracy & Speech Intelligibility



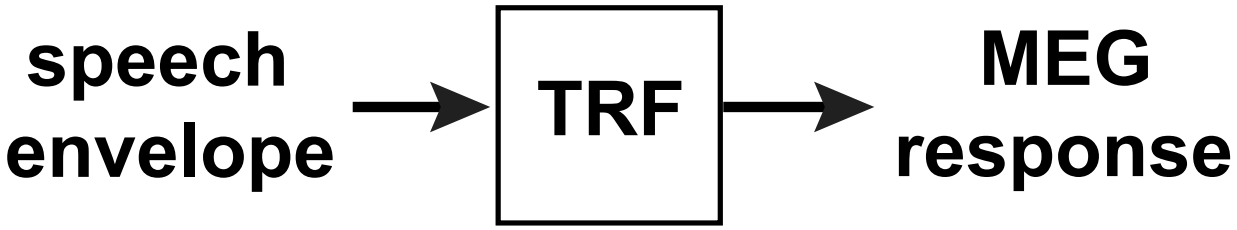
For the noise masker,

At -3 dB TMR, neural reconstruction accuracy predicts individual subjects' speech rating.



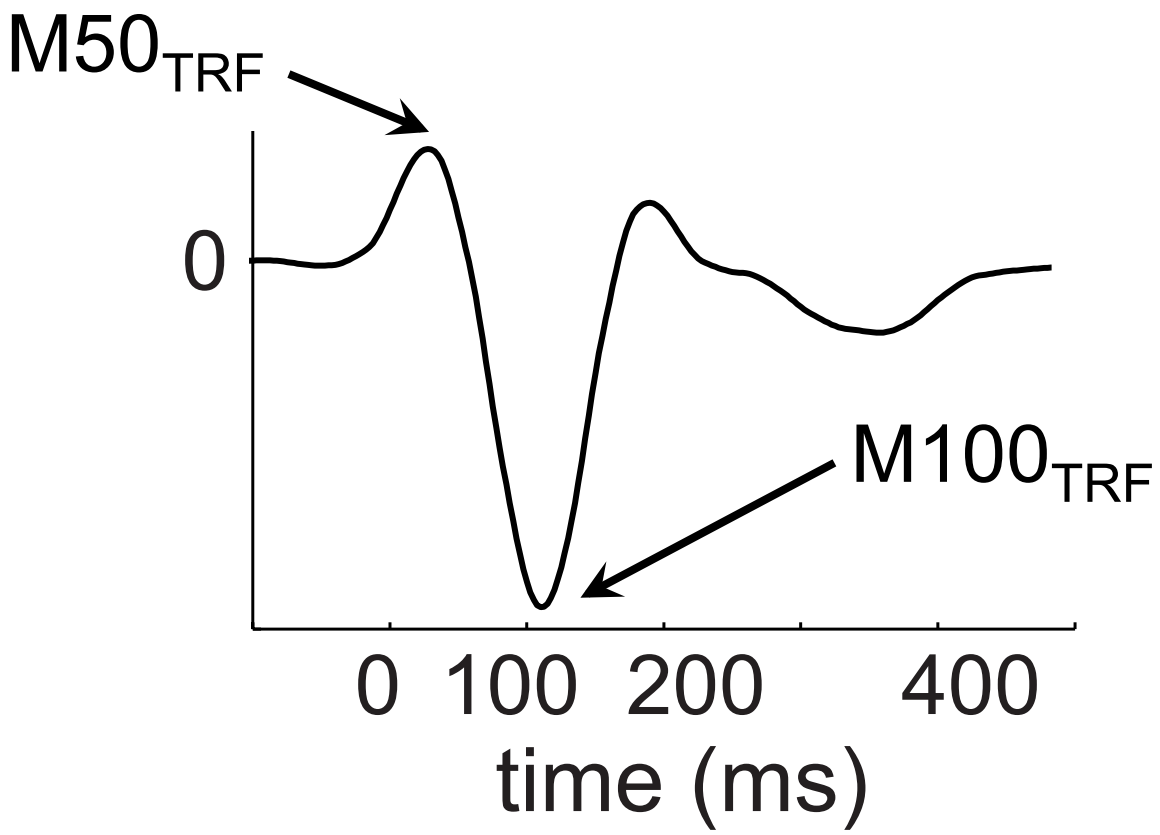
Spatial-temporal Details of the MEG Response to Speech

The neural response from each MEG sensor can be modeled by a temporal response function (TRF).



The TRF is the impulse response that transforms the speech envelope (stimulus) to the MEG recording (response).

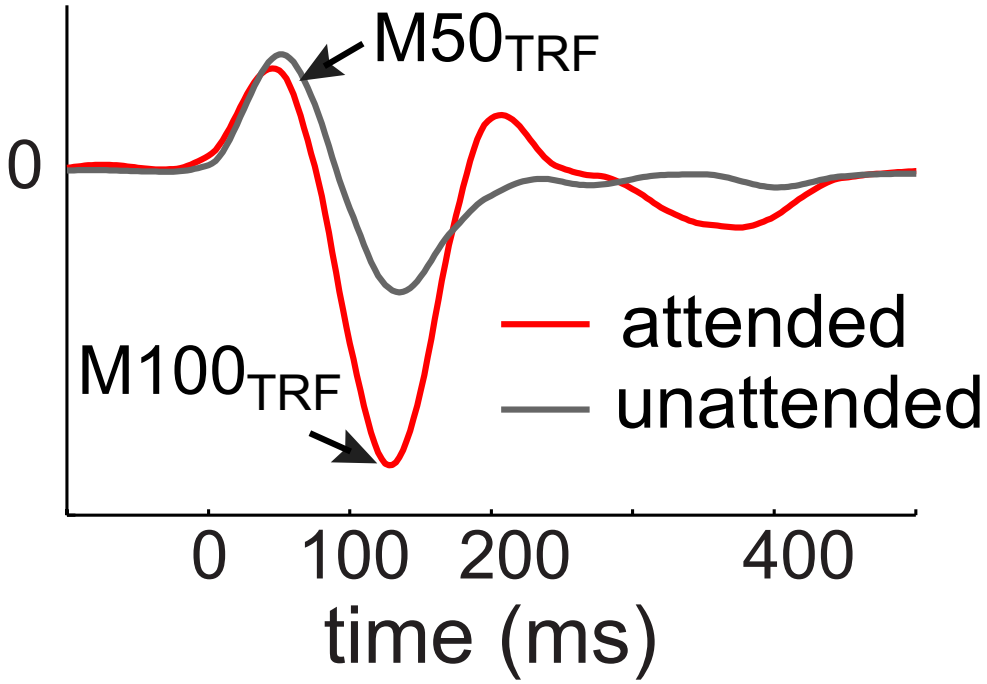
Temporal Response Function (TRF)



The source of $M100_{TRF}$ is from posterior association auditory cortex, while the source of $M50_{TRF}$ is more close to core auditory cortex.

TRF (Attention Effect)

Speech Masker

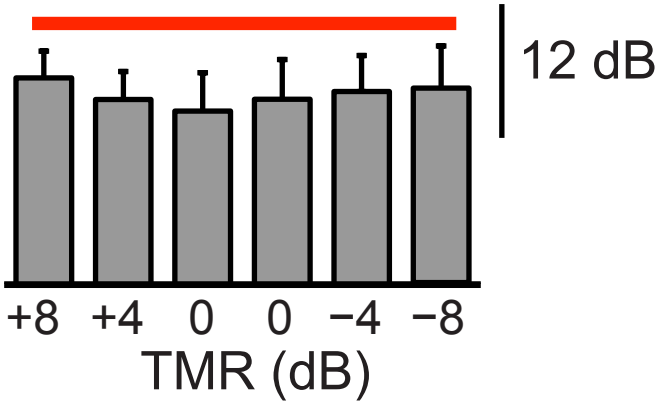


The $M100_{TRF}$ is strongly modulated by attention while the $M50_{TRF}$ is not.

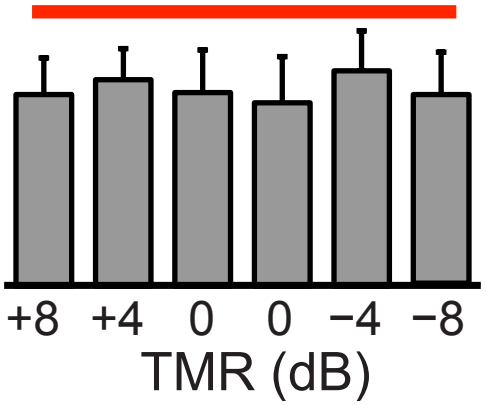
TRF (TMR Effect)

Speech Masker

$M50_{TRF}$ amplitude

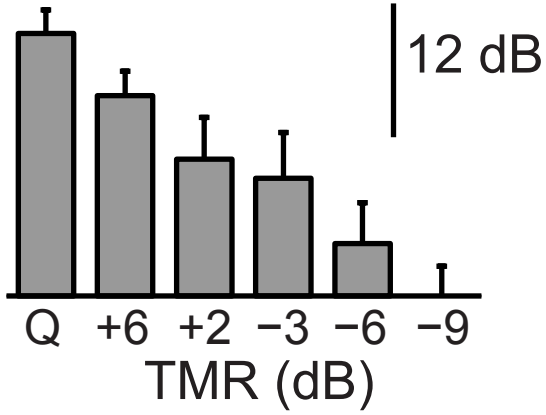


$M100_{TRF}$ amplitude

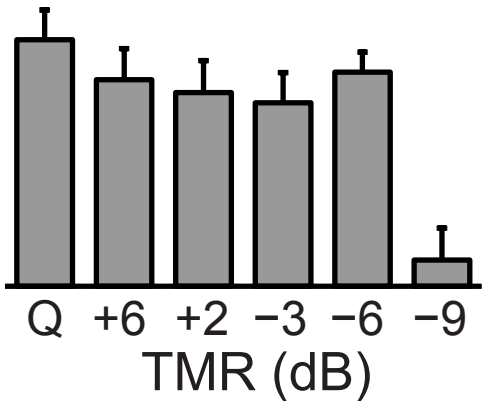


Noise Masker

$M50_{TRF}$ amplitude



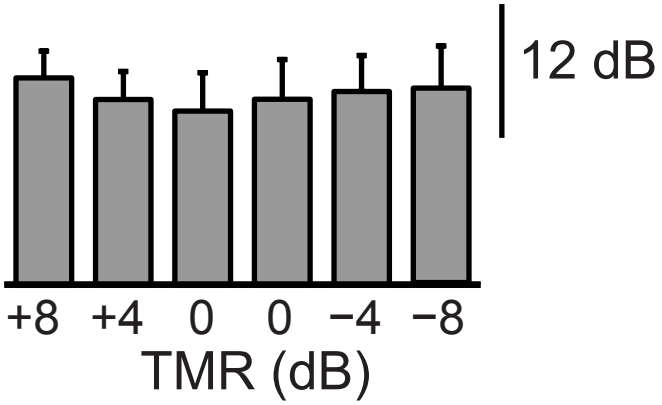
$M100_{TRF}$ amplitude



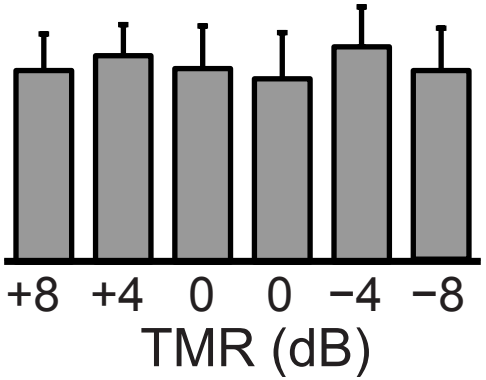
TRF (TMR Effect, $M100_{TRF}$)

Speech Masker

$M50_{TRF}$ amplitude

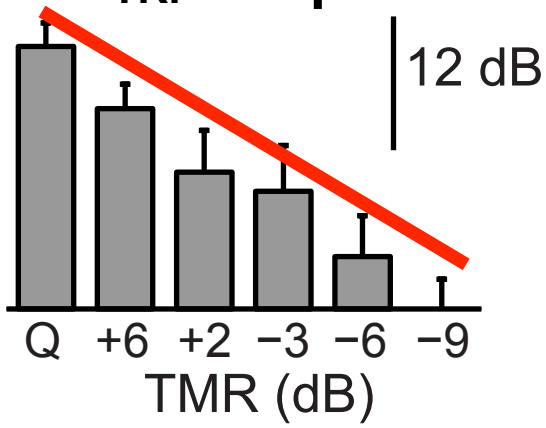


$M100_{TRF}$ amplitude

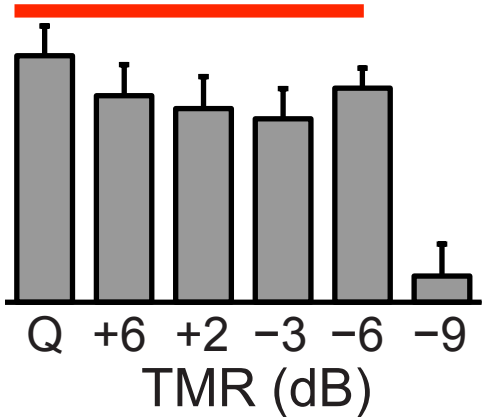


Noise Masker

$M50_{TRF}$ amplitude



$M100_{TRF}$ amplitude



TRF

- The $M50_{TRF}$ is
  - not modulated by attention.
  - affected by energetic but not informational maskers.

  It may reflect the audibility of the speech stream.

- The $M100_{TRF}$ is
  - modulated by attention.
  - robust to both energetic and informational maskers.

  It may reflect the creation of a neural representation specific to the attended auditory stream.

Summary

- Long-latency (~100 ms) cortical activity encodes the temporal modulations of:
  - the target speech stream, but
  - not the physical acoustics

- This auditory stream-specific neural synchronization predicts an individual's speech score in noise, and is likely to contribute to the perception of the syllabic structure of speech.

Acknowledgments: Supported by NIH R01 DC 008342

Thanks!

Acknowledgments: Supported by NIH R01 DC 008342