Neural Entrainment to Speech in Noise: Analysis in the Time and Frequency Domain

Nai Ding[1,3], Jonathan Z. Simon[1,2]

[1]Department of Electrical & Computer Engineering, [2]Department of Biology, University of Maryland College Park;
[3]Department of Psychology, New York University

Computational Sensorimotor Systems Lab

Introduction

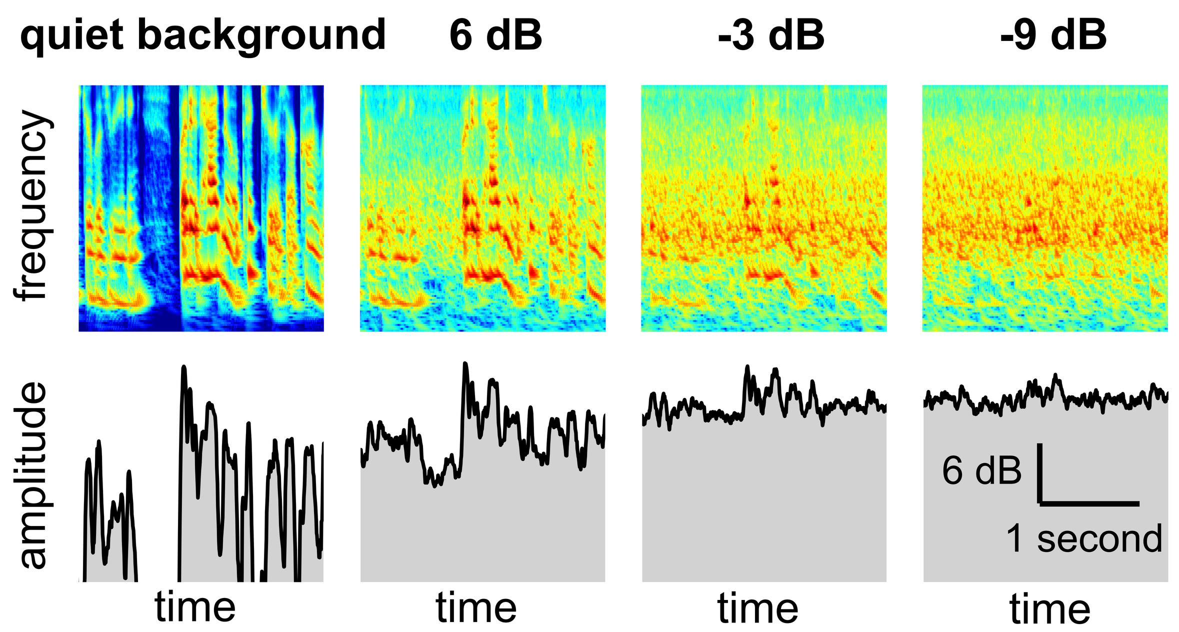
First, we investigate how the cortical representation of speech is affected by noise. We record from human subjects listening to a narrated story using magnetoencephalography (MEG). The narrated story is presented in spectrally matched stationary noise at different signal-to-noise ratios (SNR). We find that the low frequency cortical activity that follows the speech envelope is robust to noise until about -9 dB SNR and is related to individual intelligibility score.

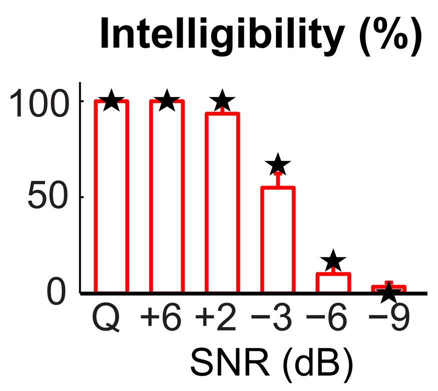
Second, we compare the time and frequency domain analysis of the MEG response synchronized to the speech envelope.

Stimuli & Data Analysis

Stimuli

The speech materials were selected from a narrated story, and cut into 50 second duration segments. A spectrally matched stationary noise was mixed into speech with one of six SNRs, i.e. quiet (no noise added in), +6 dB, +2 dB, -3 dB, -6 dB, and -9 dB. All the sections were presented sequentially and then repeated twice (3 trials in total). The subjects had to answer a comprehension question after each section, and rate speech intelligibility during the first presentation of each section. The background noise reduces the dynamic range of the stimulus (as evident from the stimulus envelope), and distorts the spectro-temporal features of speech.





subjective speech intelligibility (bars) & questions correctly answered (stars).

Data Analysis

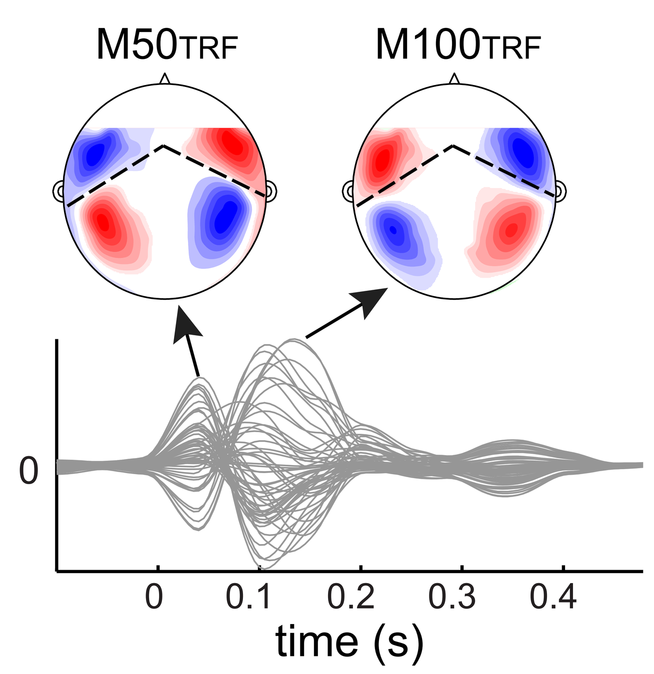
MEG: 157-channel, whole-head MEG. 1 kHz sampling rate, resampled to 40 Hz. The neural source of MEG activity is localized using a equivalent current dipole model, one per hemisphere. 10 subjects participated in the experiment.

Temporal Response Function:
If the MEG response is modeled as the speech envelope processed by a linear system, the temporal response function (TRF) is the impulse response of the linear system. In other words, the neural response from a MEG sensor is modeled by the stimulus envelope convolved with a TRF. The TRF can be interpreted as the neural response evoked by a unit power increase of the stimulus and is estimated using boosting with 10-fold cross validation (David et al. 2007).

The TRF is analogous to the spectro-temporal receptive field (STRF) model, except that it contains only the temporal dimension. Nevertheless, the TRF models the response from large neural populations rather than e.g. a single neuron. The TRF may reflect the impulse response from a single neural population or the compound impulse responses from multiple neural populations.

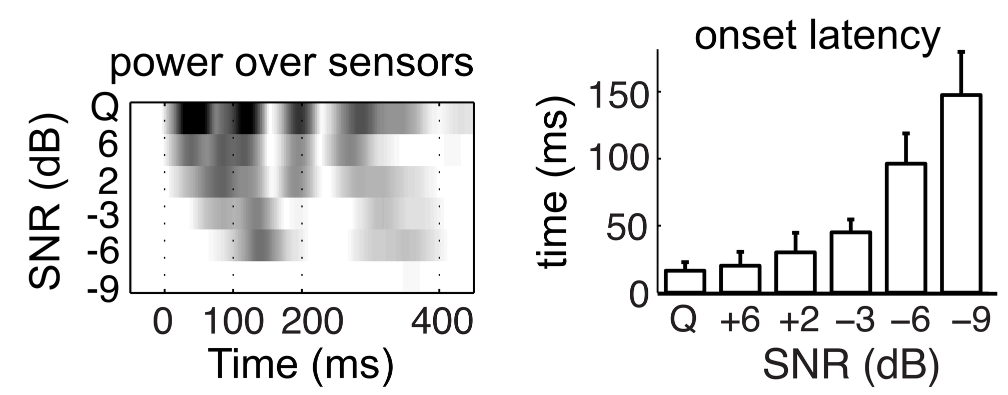
Coherence Spectrum:
The coherence spectrum refers to the inter-trial correlation of the MEG response filtered into narrow frequency bands. It reflects how repeatable the MEG response is when the same stimulus is played multiple times, and is a measure of the degree of phase-locking of the MEG response. It is a non-parametric characterization of the neural response and does not explicitly model the relationship between the stimulus and response.

Temporal Response Function



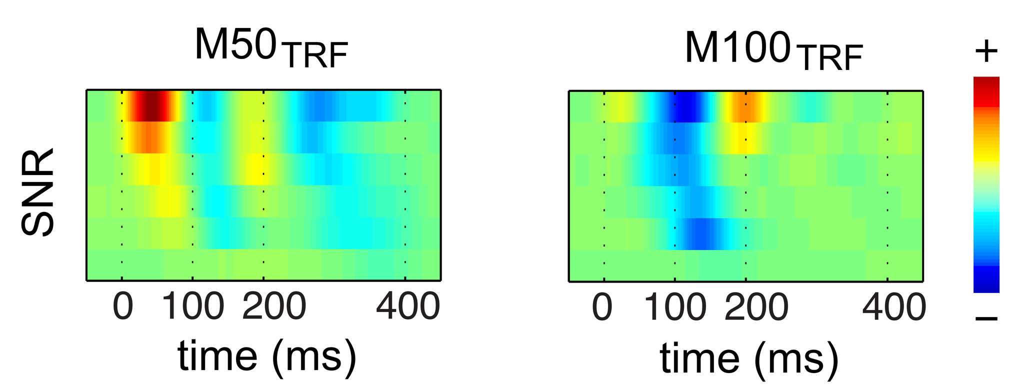
A temporal response function (TRF) is estimated for each MEG sensor. It represents the neural response evoked by a unit power increase of the stimulus. The TRF has two salient peaks, the M50TRF and M100TRF, which have opposite polarity. The source of the M100TRF is consistent with the source of the M100 evoked by a tone pip, which is in posterior association auditory cortex. The source of the M50TRF is more anterior than the source of the M100TRF, and is more close to core auditory cortex.

Power of the TRF over all MEG Sensors



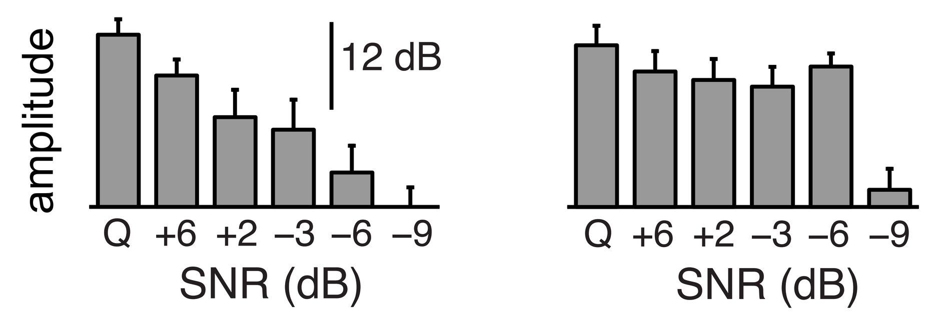
The onset latency of the TRF is elongated as the SNR decreases. The amplitude, however, is relatively stable between -6 and 6 dB SNR.

TRF at the Neural Source Locations of the M50TRF & M100TRF



The TRF projected to the source locations of the M50TRF and M100TRF. The polarity of the M100TRF is consistent with the polarity of the M100 and is defined as being negative.

Amplitude of the M50TRF & M100TRF



The amplitude of the M50TRF (left) continuously decreases with SNR while the amplitude of the M100TRF (right) is stable above -9 dB.

Summary

Longer-latency (~100 ms) responses from posterior auditory cortex are robust to noise, but not shorter-latency (~50 ms) responses from areas more close to core auditory cortex.

Discussion

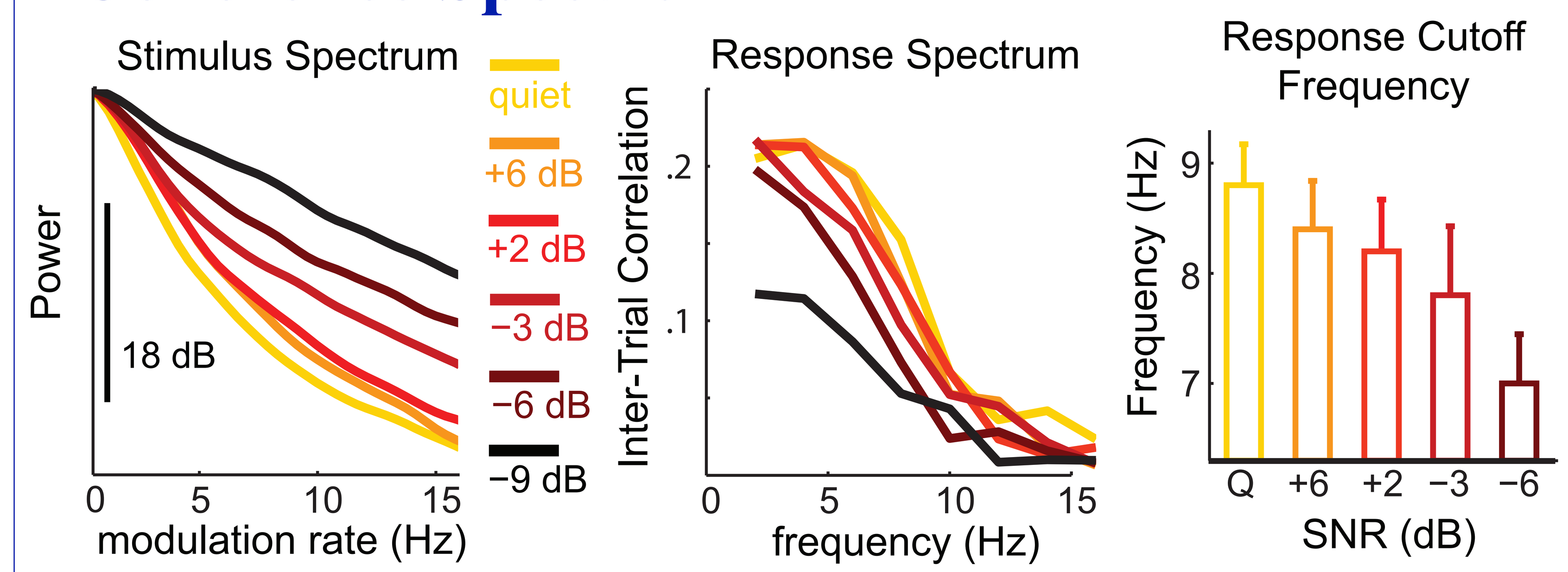
A previous study (Ding & Simon, 2012) shows that, when a listener selectively listens to one of two competing speakers, the M100TRF is modulated by attention while the M50TRF is not. Combining these observations, we hypothesize that the M50TRF mainly reflects the bottom-up saliency of an auditory stream and the M100TRF reflects the perceptual dominance.

The Pros and Cons of TRF Analysis

***Pros*: 1. It reveals the timing of response components. 2. Easy to interpret under the framework of linear system theory.**

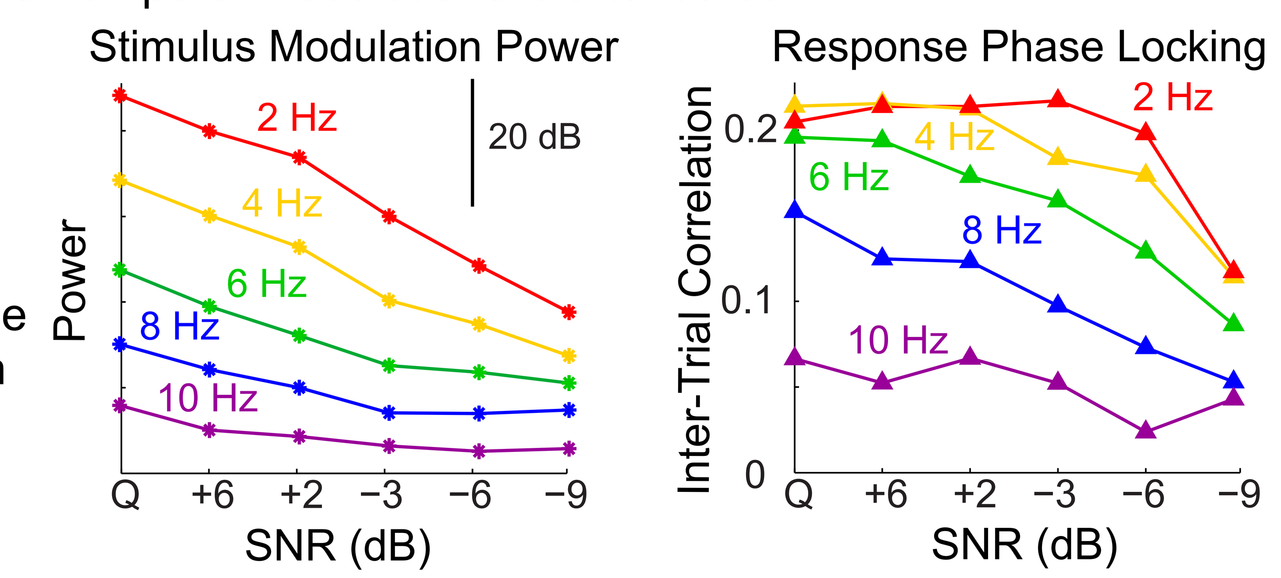
***Cons*: 1. The shape of TRF is affected by the estimation procedure. 2. The shape of the TRF shows some variability across subject, making group analysis harder. 3. Linear model.**

Coherence Spectrum



Left: The modulation spectrum of the stimulus, normalized based on its power density at 0.1 Hz. Noise introduces more high-frequency temporal modulations. *Middle*: The coherence spectrum of neural response is consistently low-pass in shape but with a cutoff frequency that decreases with poorer SNR (*Right*). Therefore, even though noise introduces more faster spectral modulations, the response to faster temporal modulations is attenuated.

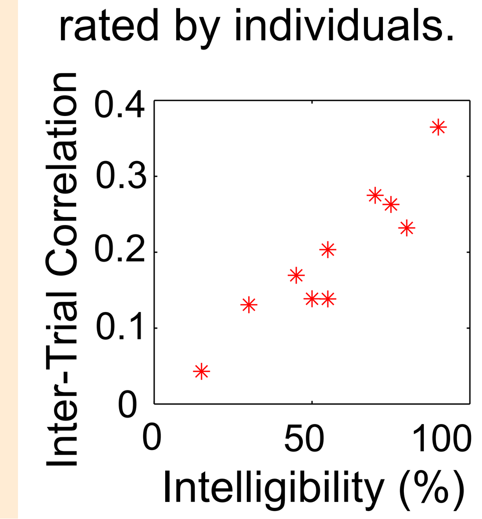
Left: Stimulus modulation spectrum again (unnormalized) *Right*: Response synchronization by SNR.



Summary

Delta band (< 4 Hz) but not theta band (>4 Hz) neural activity is robust to noise.

At -3 dB SNR, the precision of delta-band neural phase locking predicts the intelligibility score rated by individuals.



Discussion

A previous study (Ding & Simon, 2012) shows that, when a listener selectively listens to one of two competing speakers, both delta and theta band activity is modulated by attention. Therefore, the delta and theta band activity are both influenced by top-down modulation but only theta band activity is influenced by bottom-up saliency of the sound target.

The Pros and Cons of Coherence Spectrum Analysis

***Pros*: 1. It captures any response that is repeatable over trials, rather than just the linear component. 2. The inter-trial correlation is a robust non-parametric measure. 3. The shape of the coherence spectrum is highly consistent across subjects. 4. It reflects the frequency composition of the response.**

***Cons*: 1. It is influenced by both the stimulus and the response. 2. It is sensitive to any phase-locked response, which complicates its interpretation. 3. It requires repetitions of the same stimulus. 4. It is not a method widely used in e.g. single unit recording or fMRI.**

References: *Competing speech experiment*: Ding & Simon, PNAS, 2012;
STRF estimation: David, Mesgarani & Shamma, Network, 2007.
STRF/TRF for MEG response to speech: Ding & Simon, J Neurophys, 2012;
Acknowledgement: work supported by NIH R01 DC-008342.