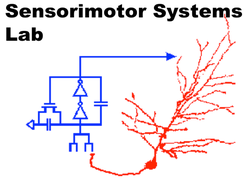
Neural Representation of Noisy Reverberant Speech in Human Auditory Cortex

Krishna C. Puvvada[1], Jonathan Z. Simon[1,2,3]

[1]Dept of Electrical & Computer Engineering, [2]Dept of Biology, [3]Institute for Systems Research

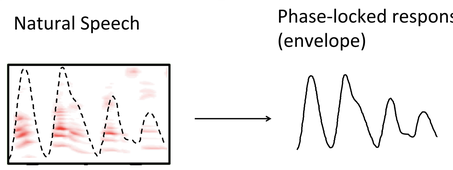
University of Maryland College Park

Computational Sensorimotor Systems Lab

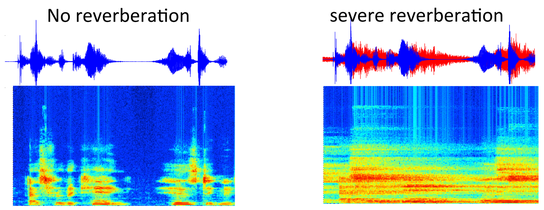


Background

- Cortical activity measured by magnetoencephalography (MEG) phase-locks to temporal modulations of stimuli.



- Natural speech corrupted with additive noise is represented in cortex as 'uncorrupted' speech and is quite robust to level of degradation. (Ding, N. & J.Z. Simon, 2013)
- Reverberation is another major source of speech degradation, causing both temporal and spectral smear.



- Is the neural representation of speech corrupted by reverberation (convolutive noise) an 'uncorrupted' version of speech ('clean' model) or reverberant version it self ('reverb' model) or both ('Mixed' model)?

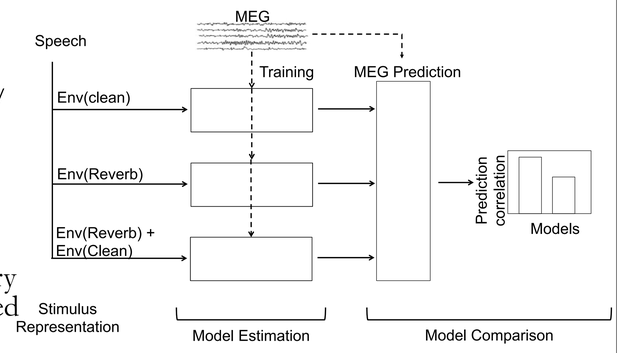
Experimental Design

- 4 reverberant conditions.
  - No reverb, mild, medium and severe reverberation.
- 3 different noise conditions.
  - No noise, +3 dB SNR, +6 dB SNR.
- 12 conditions in total.
- 60 second long story segments, 3 repetitions.
- $N = 12$ Subjects.
- To maintain attention, in each tiral the subject counts the number of times a keyword occurs in the story.
- MEG recording with 157 channels.
- 1kHz sampling , Time-shifted PCA based de-noising.
- Spatial filtering used to reduce 157 channels to 10, more reliable, virtual channels.

Methods

MEG prediction

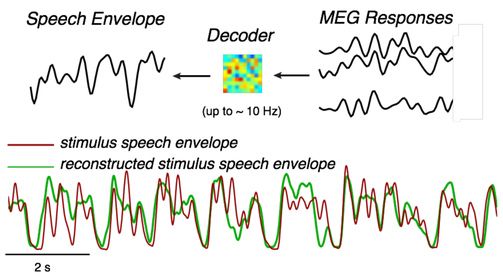
- MEG data predicted under different models/ stimulus representations, using optimum linear filters.
- First DSS auditory component is used as dominant auditory response.



Stimulus reconstruction

- Temporal envelope of stimulus is recon-structed from cortical responses using optimum linear filters.
- Reconstruction based on integrating neural responses over a temporal window.

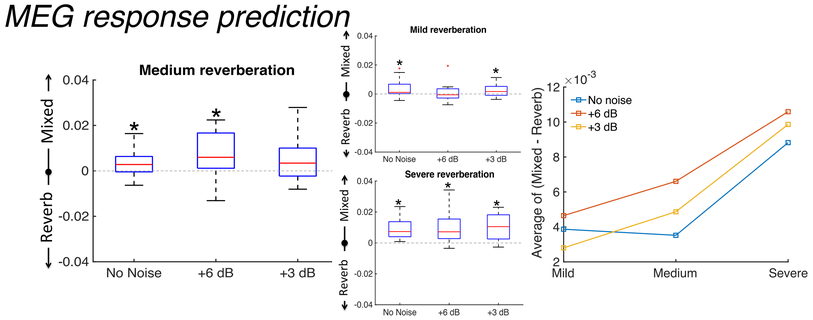
Stimulus envelope reconstruction from MEG response



- Optimum decoders are designed to reconstruct the speech envelope under each model (clean/reverb).
- Correlation between reconstructed and presumed model envelope is used as metric as to how faithfully the speech envelope is represented under presumed model.

Results

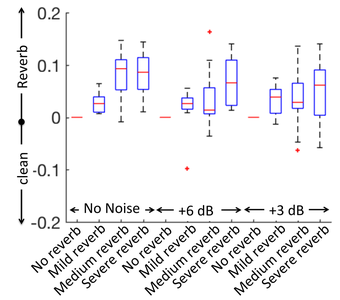
MEG response prediction



- All 3 models performed significantly above chance.
- Mixed model predicts MEG responses significantly better than both clean and reverb models (* $p<0.05$, permutation test).

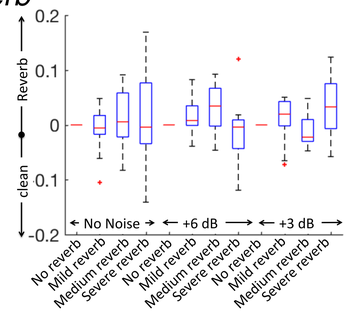
Stimulus reconstruction using Delta band (1-4 Hz) responses: clean vs. reverb

- Both models performed significantly above chance.
- Significant test shows that neural responses are more correlated with reverberant envelope than clean envelope.



Stimulus reconstruction using Theta band (4-8 Hz) responses: clean vs. reverb

- Both models performed significantly above chance.
- Neither model performed significantly better than the other.



Discussion

- Noisy reverberant speech is represented in cortex as *a mixture of both clean and reverberant versions of the speech*, possibly from different cortical areas.
- While Delta band neural responses (prosody level information in speech) are dominated by the reverberant representation of envelope, Theta band responses (syllabic information in speech) do not emphasize one representation over the other.
- Since the stimulus contrast is actually stronger in Theta than Delta, the shift away from the reverberance dominated model acts as evidence for *reverberance removal in theta band.*

References:

1) Ding N. and J. Z. Simon, (2013). Adaptive temporal encoding leads to a background insensitive cortical representation of speech. *Journal of Neuroscience.*

2) de Cheveigné, A., and J. Z. Simon (2008) Denoising Based on Spatial Filtering, *Journal of Neuroscience Methods.*

Acknowledgements: Funding from NIH R01 DC 014085