Comparing Methods for Deriving the Auditory Brainstem Response to Continuous Speech in Human Listeners

Tong Shan[1], Ross K Maddox[1,2]

Departments of [1]Biomedical Engineering, and [2]Neuroscience, University of Rochester, Rochester, NY

maddoxlab.urmc.edu

Hearing & Balance Research Collective

Introduction

- Much remains unknown about how the brain encodes speech subcortically.
- Auditory brainstem responses (ABR) are used to characterize subcortical activity in human beings, but traditionally rely on transient stimuli.
- We have developed deconvolution methods to investigate natural speech encoding at subcortical level using different features as regressors [1,2,3]

GOAL: Quantitatively compare the deconvolved responses using three different regressors to help guide decisions on what approach to choose when deriving ABRs from natural speech and other natural sounds.

Methods

Stimulus with two conditions

- Original speech source: 40 excerpts of 64 s-long English audiobooks (male narrator) [4]
- **Conditions**
  1) Unaltered original speech
  2) "Peaky" speech [2]: re-synthesized from the original stimuli by aligning the phase of the harmonics at glottal pulses and making the speech impulse-like but maintaining the intelligibility. (Fig.3)

Deconvolution

- Encoding Model

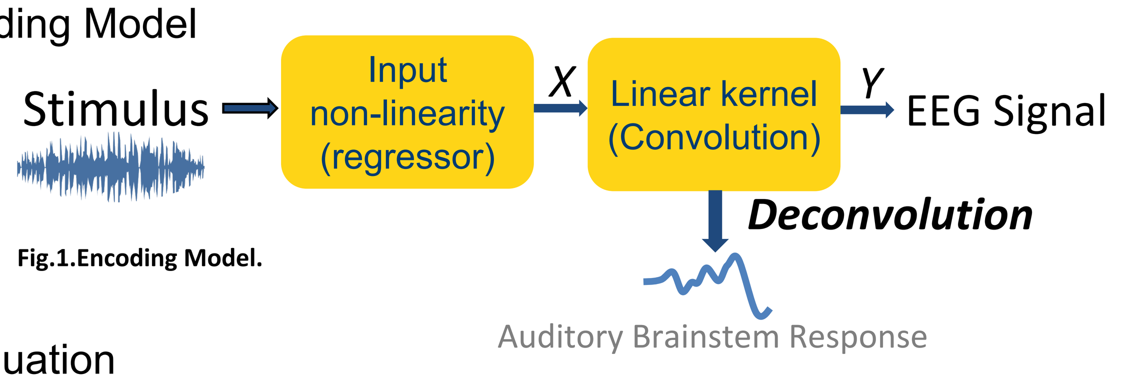

Fig.1.Encoding Model.

- Equation

$$response = \mathcal{F}^{-1}\left\{\frac{\sum_n b_n X_n^* Y_n}{\frac{1}{N}\sum_n X_n^* X_n}\right\},$$

where $response$ denotes the derived impulse response (ABR), X the FFT of the stimulus with the non-linearity applied (i.e., regressor), Y the FFT of EEG signal, * the complex conjugate, $\mathcal{F}^{-1}$ the inverse FFT, b_n the averaging weight of the n^{th} trial proportional to the inverse of the trial variance, N the total number of trials, n the index of n^{th} trial.

Subject Data from Polonenko & Maddox [4]

- 22 adults (aged 18-40) with normal hearing thresholds.

Three regressors

- **Halfwave-Rectified stimulus waveforms (HWR)** (Maddox & Lee, 2018)

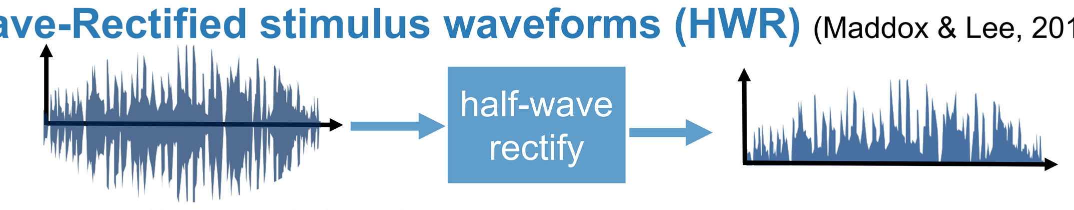

Fig.2. Half-wave rectified stimulus generation

- **Glottal pulse train (Pulse)** (Polonenko & Maddox, 2021)

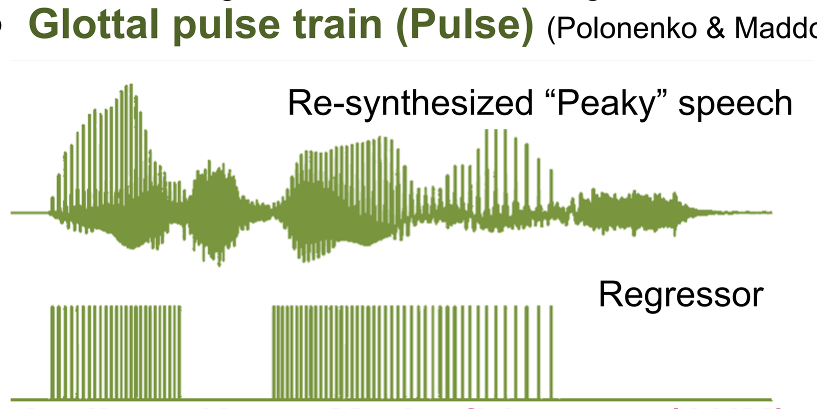

Fig.3. Glottal pulse train, figure adapted from Polonenko & Maddox (2021)

- **Auditory Nerve Model firing rate (ANM)** (Shan et al., 2022)
  - Generated from Zilany et al (2014) [5, 6], which models the detailed transformation from acoustic signals to the AN representation of the stimulus (show in fig.4)
  - The response will take the peripheral nonlinear effects of the acoustical differences into account

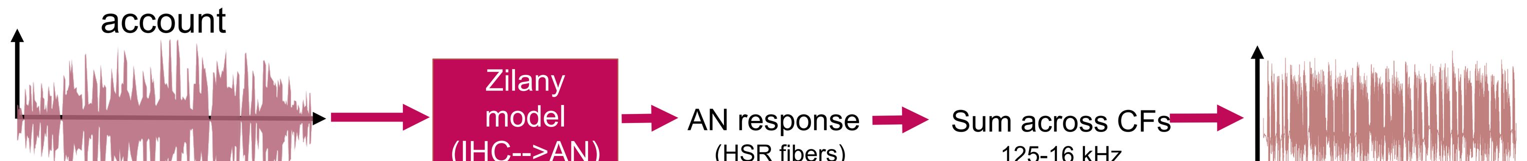

Fig.4. ANM regressor generation

Results

Deconvolution Derived Waveforms

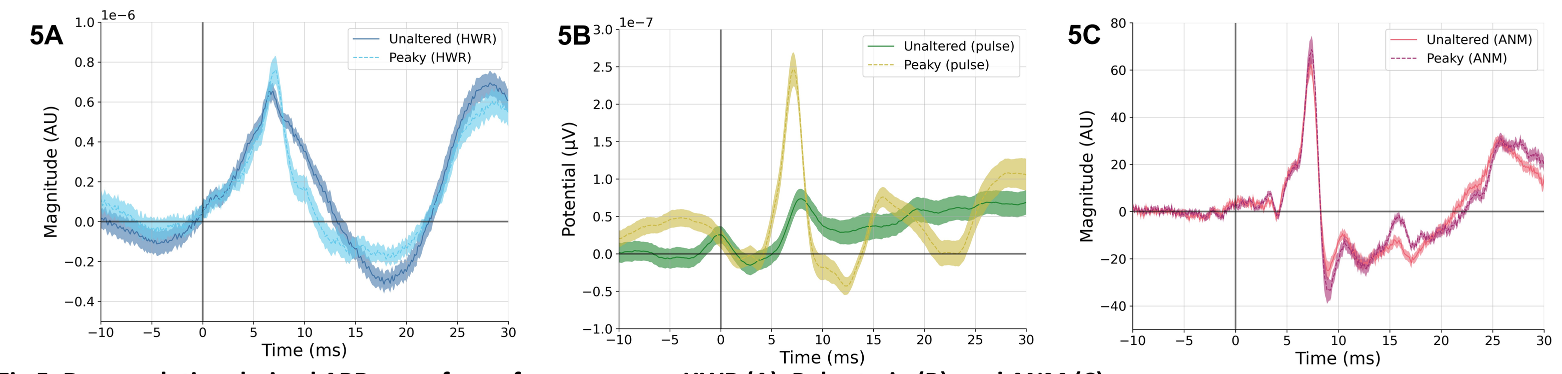

Fig.5. Deconvolution derived ABR wave forms from regressor HWR (A), Pulse train (B), and ANM (C).

Comparison of Signal-to-Noise Ratio (SNRs)

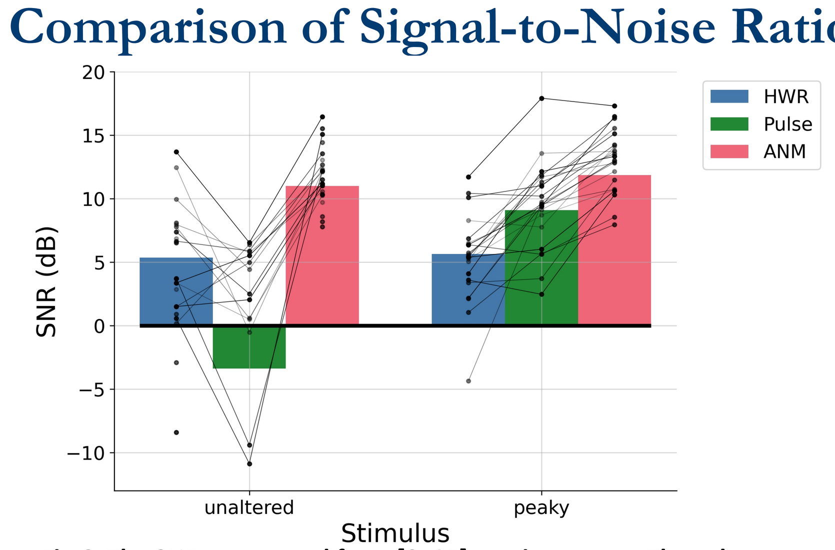

Fig.6. The SNRs computed from [0, 15] ms time range. The color-coded bars are the corrected SNR of pooled responses, and the think black lines are the SNRs from each subject.

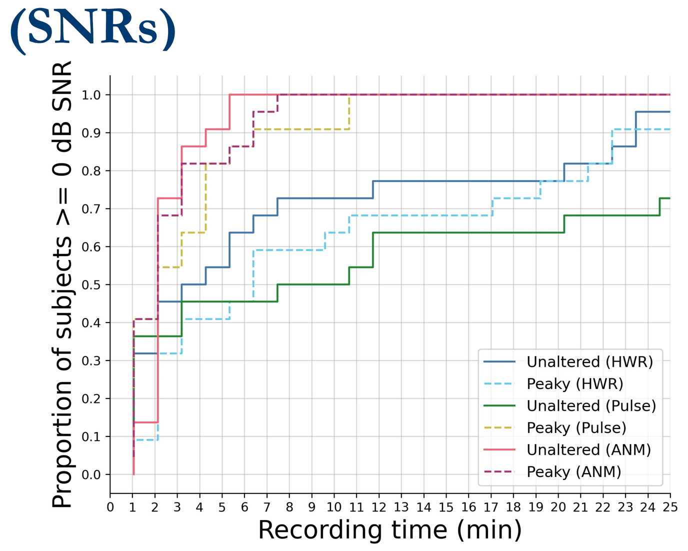

Fig.7. The cumulative proportion of subjects that has ABR SNR >= 0 dB versus recording time needed.

- Among the broadband regressors, ANM performed the best for the waveforms, the SNR and the acquisition time. But the differences of the regressors' power spectra may lead to different spectra of the responses.

Regressors Have Different Spectra

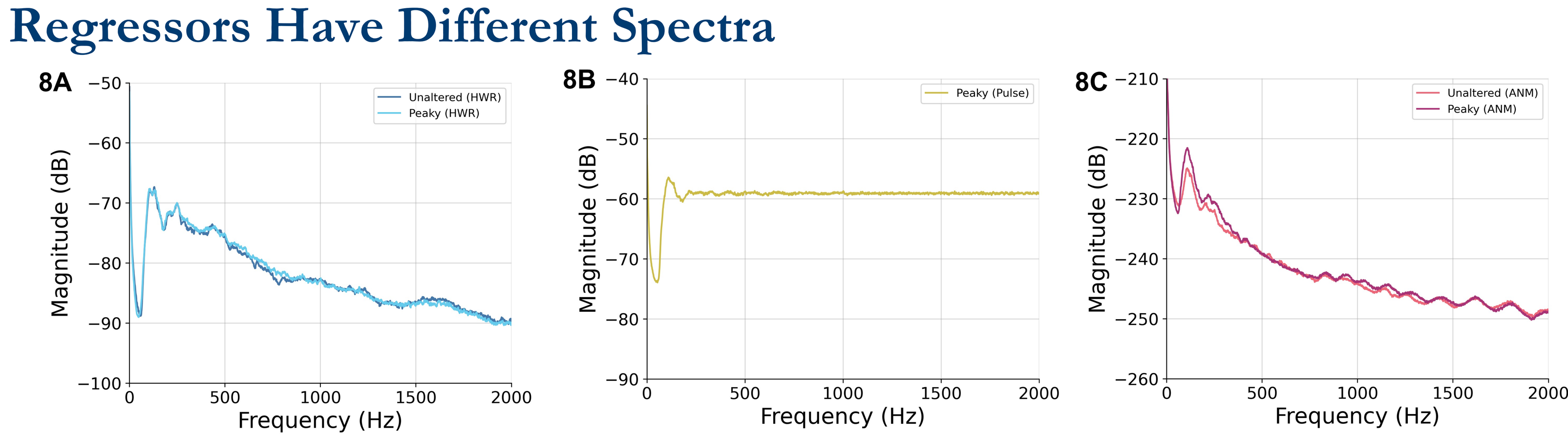

Fig.8. Averaged power density spectrum with welch estimate for HWR (A), Pulse train (B), and ANM (C). They have different spectra.

Spectral Coherence Analysis

- Compare the power of the 3 regressors in predicting EEG across frequency band
- Less sensitive to spectrum of predicted response

$$C_{xy}(f) = \frac{E[X_i^*(f)\,Y_i(f)]}{\sqrt{E[X_i^*(f)\,X_i(f)]\,E[Y_i^*(f)Y_i(f)]}}$$

Where $C_{xy}(f)$ denotes the coherence between signal x and y at frequency band f, $E[\,]$ the expected value across slices, * the complex conjugation. X_i the Fourier transform for x slice i, and Y_i the Fourier transform for y slice i.

- Our analysis computes the coherence of the predicted EEG (*x*) by the ABR kernels generated by the 3 regressors (averaged across 22 subjects) and the true EEG signal (*y*)
  - Predicted EEG is computed by convolving the averaged ABR kernel with the regressors for each stimulus

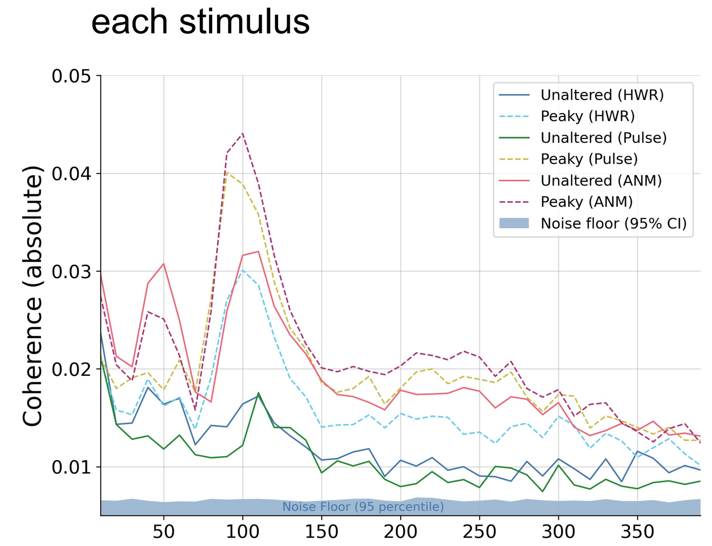

Fig.9. Coherence in frequency range [0, 400] Hz using a window size of 0.1s. At most frequency range, the ANM regressor is best and then the pulse regressor for "peaky" speech.

Using Phase-only Regressors

- To fairly compare time-domain response waveforms in a way that does not depend on the regressor magnitude spectrum, we performed deconvolution using a phase-only regressor.
- The equation to make phase-only regressor is as follows:

$$X(f)_{phase-only} = \frac{X(f)}{|X(f)|}$$

- Where $X(f)$ is the FFT of the regressor and $|X(f)|$ is the magnitude of the regressor in frequency domain, and $X(f)_{phase-only}$ is the phase-only regressor we will use in the following deconvolution

Deconvolution Derived Waveforms from Phase-only Regressors

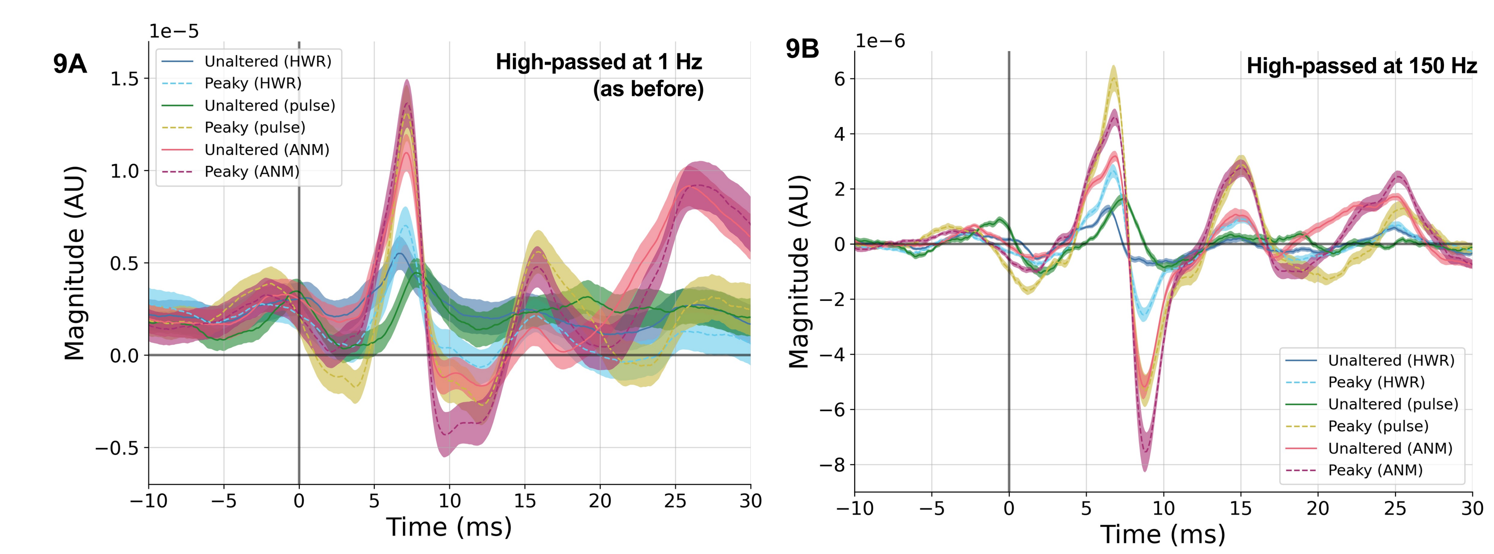

Fig.10A. The waveforms derived from phase-only regressors.

Fig.10B. The waveforms derived from phase-only regressors and were high-passed at 150 Hz.

Comparison of SNRs of phase-only responses

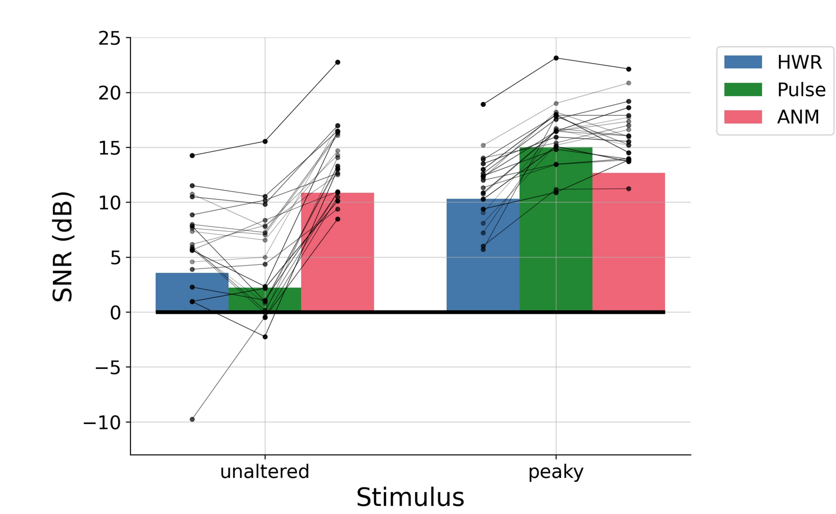

Fig.11. The SNRs computed from [0, 15] ms time range for the ABR derived from phase-only regressors. The color-coded bars are the corrected SNR of pooled responses, and the think black lines are the SNRs from each subject.

Summary

- Both the ANM and glottal pulse regressor (the latter only for peaky speech) provided comparable high-quality ABRs and quick acquisition and substantially outperformed the HWR regressor of Maddox and Lee [1].
- The glottal pulse regressor has the disadvantage of being applicable only to re-synthesized peaky speech, but the advantage of providing ABRs in meaningful physical units (i.e., microvolts).
- The ANM regressor has the advantage of being applicable to both natural and peaky speech, and in principle other natural sounds such as music.
- Because regressors have different magnitude spectra, we developed the phase-only regressor to more fairly compare the quality of waveforms deconvolved using different regressors.

References

[1] Maddox, R. K. & Lee, A. K. C. Auditory Brainstem Responses to Continuous Natural Speech in Human Listeners. eNeuro 5, doi:10.1523/ENEURO.0441-17.2018 (2018).
[2] Melissa J Polonenko, Ross K Maddox (2021) Exposing distinct subcortical components of the auditory brainstem response evoked by continuous naturalistic speech eLife 10:e62329. https://doi.org/10.7554/eLife.62329
[3] Tong Shan, Madeline S. Cappelloni, Ross K. Maddox. Music and Speech Elicit Similar Subcortical Responses in Human Listeners. bioRxiv 2022.10.14.512309; doi: https://doi.org/10.1101/2022.10.14.512309
[4] Polonenko, Melissa J; Maddox, Ross K (2021), Exposing distinct subcortical components of the auditory brainstem response evoked by continuous naturalistic speech, Dryad, Dataset, https://doi.org/10.5061/dryad.12jm63xwd
[5] Zilany, M. S., Bruce, I. C. & Carney, L. H. Updated parameters and expanded simulation options for a model of the auditory periphery. The Journal of the Acoustical Society of America 135, 283-286 (2014).
[6] Zilany, M. S., Bruce, I. C., Nelson, P. C. & Carney, L. H. A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics. The Journal of the Acoustical Society of America 126, 2390-2412 (2009).

Research supported by NIH grant R01DC017962, NSF CAREER grant 2142612 and a pilot grant from the Schmitt Program in Neuroscience