Nonlinear EEG decomposition reveals distinct neural processes that track speech and music

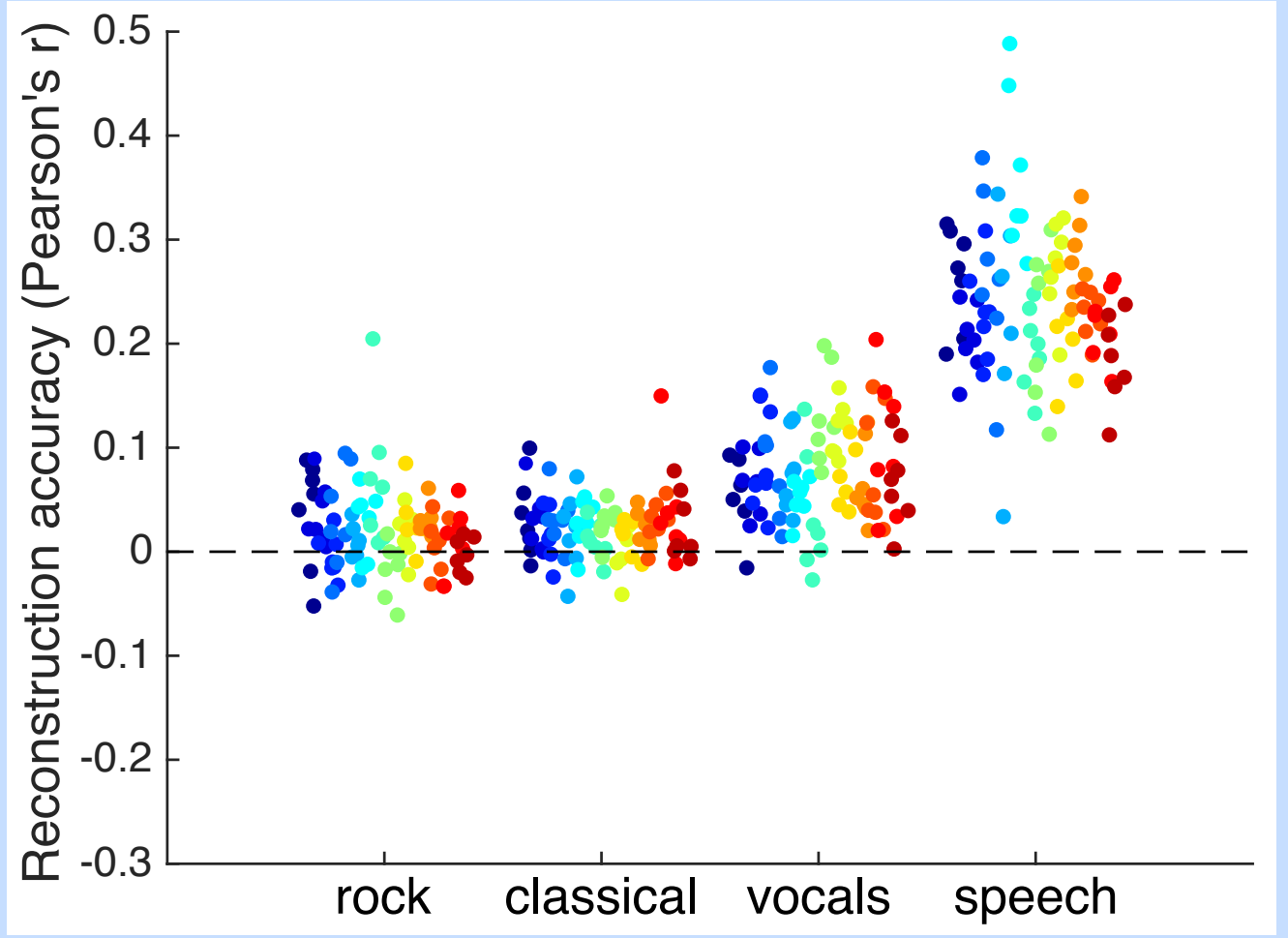
Nathaniel Zuk[1], Jeremy Murphy[1], Edmund Lalor[1,2,3]

1) Trinity College Institute of Neuroscience, Trinity College, Dublin, Ireland
2) Department of Biomedical Engineering, University of Rochester, Rochester, NY
3) Department of Neuroscience, University of Rochester Medical Center, Rochester, NY

Introduction

Previous research has shown that speech and music activate different neural circuits in the human brain [1], but it is unclear how these circuits differ in temporal activation. EEG-based reconstruction of the stimulus envelope has been used to study the neural processing of continuous speech [2,3]. However, we found that the reconstruction accuracy for music was markedly worse than the accuracy for speech (see figure on the right, each color is a different subject). The difference in accuracy could not be explained by differences in the statistics of the stimulus envelopes.

Here, we demonstrate a new technique for stimulus envelope reconstruction that utilizes the magnitude and phase of separate frequency bands of the EEG signal. Using this model, we show that temporal processing in the brain is distinct for speech and music.



Experiment and EEG preprocessing

14 subjects passively listened to:

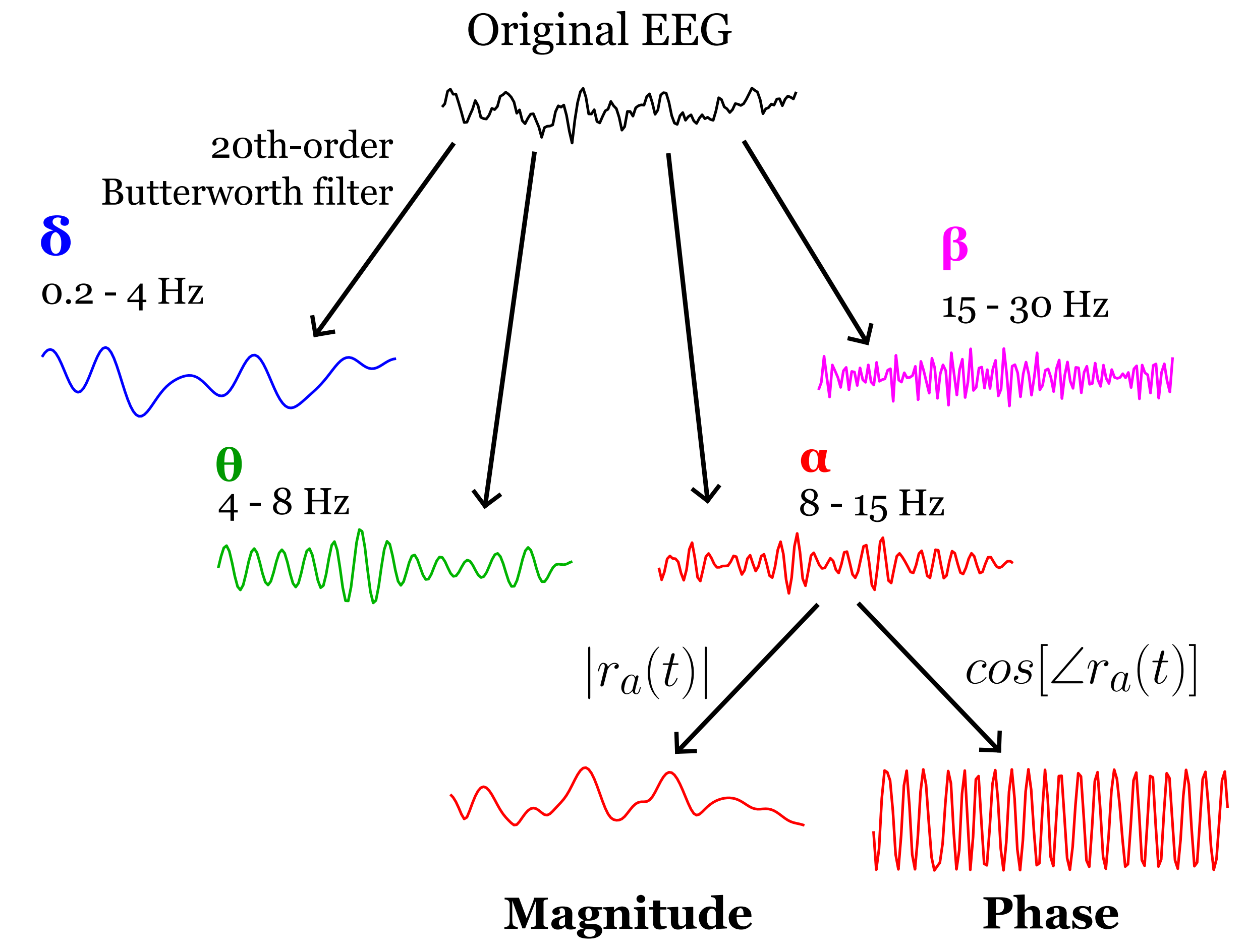
- **Rock** (rock songs, including vocals)
- **Classical** (excerpts from classical music)
- **Vocals** (from the rock songs, no other instruments)
- **Speech** (audiobook excerpts)

Subjects listened to 6-7/10 stimuli of each type

Stimulus envelope was computed averaging the gammatonegram with 16 channels, 125 - 8000 Hz

Recorded 128-channel EEG, filtered from 0.2 - 32 Hz with zero-phase chebyshev type-II filter, artifacts removed with ICA

Analytic EEG representation

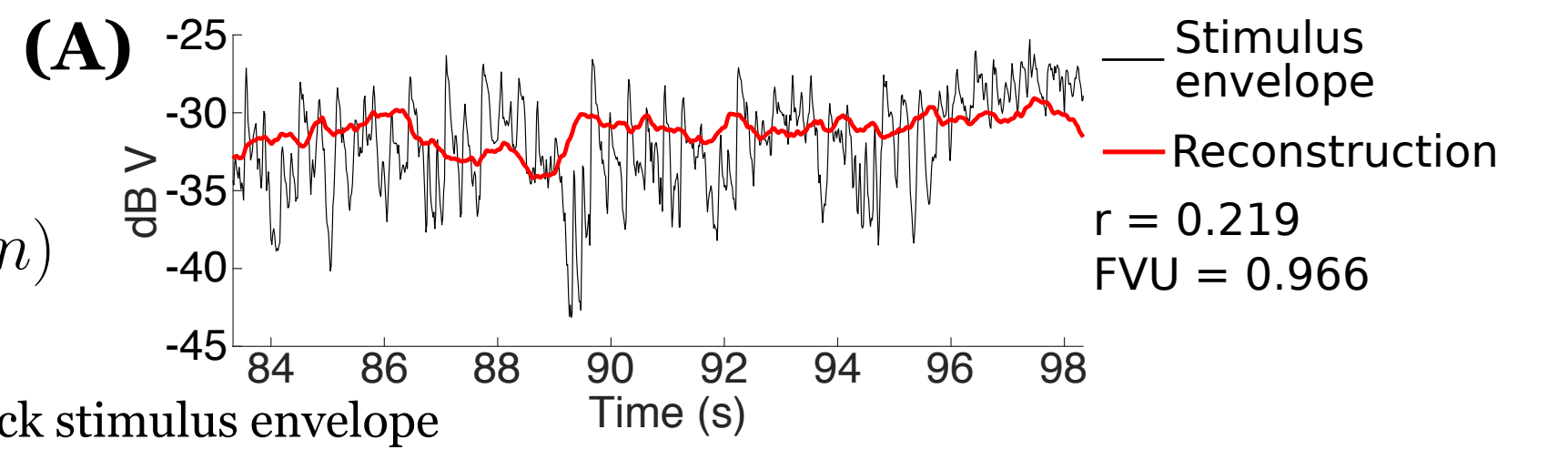


Used PCA components of the 128-channel EEG, 6-16 principal components were retained for each subject, based on the minimization of cross-validation error.

Analytic EEG model improves reconstruction accuracy for rock music and isolated vocals

Stimulus envelope reconstruction:

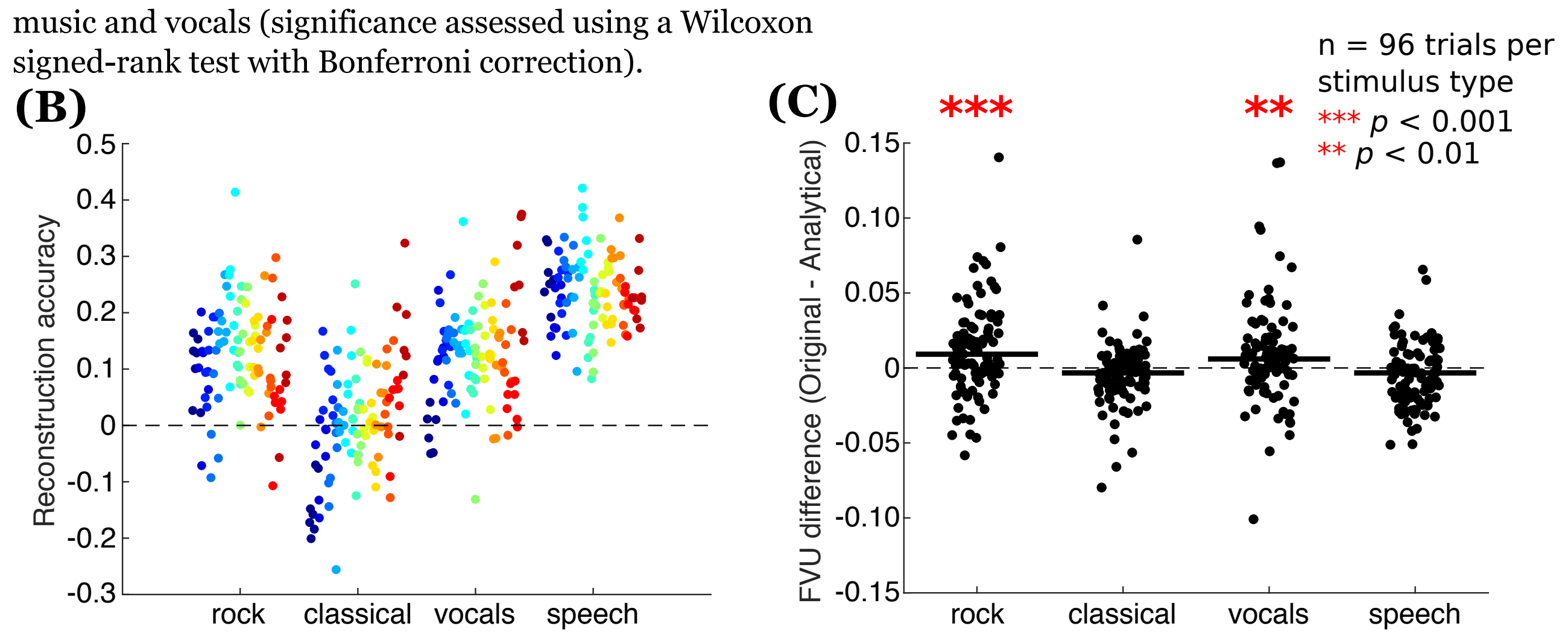
$$\widehat{s}(t) = \sum_n \sum_t r(t+\tau)g(\tau,n)$$



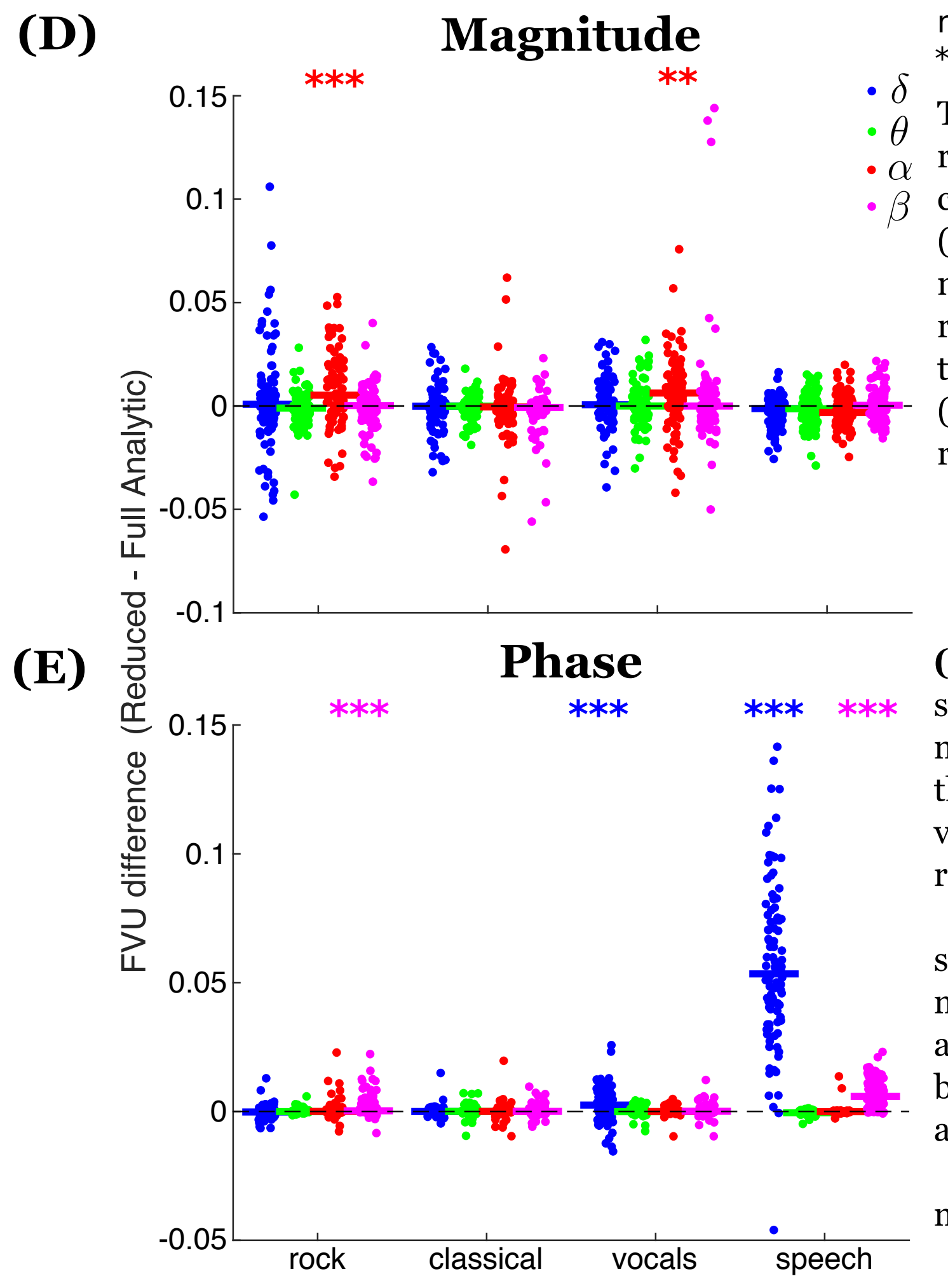
(A) Example reconstruction for a rock stimulus envelope (τ = 0 to 800 ms). **(B)** Across all subjects there was an increase in r values for the rock music and vocals using the analytic EEG model. **(C)** We used the difference in FVU to quantify the significance of the improvement in the reconstruction using the analytic EEG model. The analytic EEG model performance is significantly better for rock music and vocals (significance assessed using a Wilcoxon signed-rank test with Bonferroni correction).

Fraction of variance unexplained:

$$FVU = \frac{\sigma^2_{err}}{\sigma^2_{stim}}$$



EEG alpha magnitude and delta phase encode rock music and speech respectively



n = 96 trials per stimulus type
*** $p < 0.001$

To assess which components are most useful for reconstruction, we used the FVU difference to compare a model with one component missing ("reduced") to the full analytic model. **(D)** Alpha magnitude significantly improved the reconstruction accuracy for rock music and for the isolated vocals from the rock songs (significance assessed using a Wilcoxon signed-rank test with Bonferroni correction).

(E) Including delta phase significantly improved speech reconstruction relative to the reduced model. Delta phase also significantly improved the reconstructions for vocals, suggesting that vocals may activate both music- and speech-related neural processes.

Additionally, beta phase also contributed significantly to the reconstructions for rock and music. However, these improvements were small, and we failed to find any specific contribution of beta phase to the reconstructions in our later analyses (not shown).

Our focus here will primarily be on alpha magnitude and delta phase.

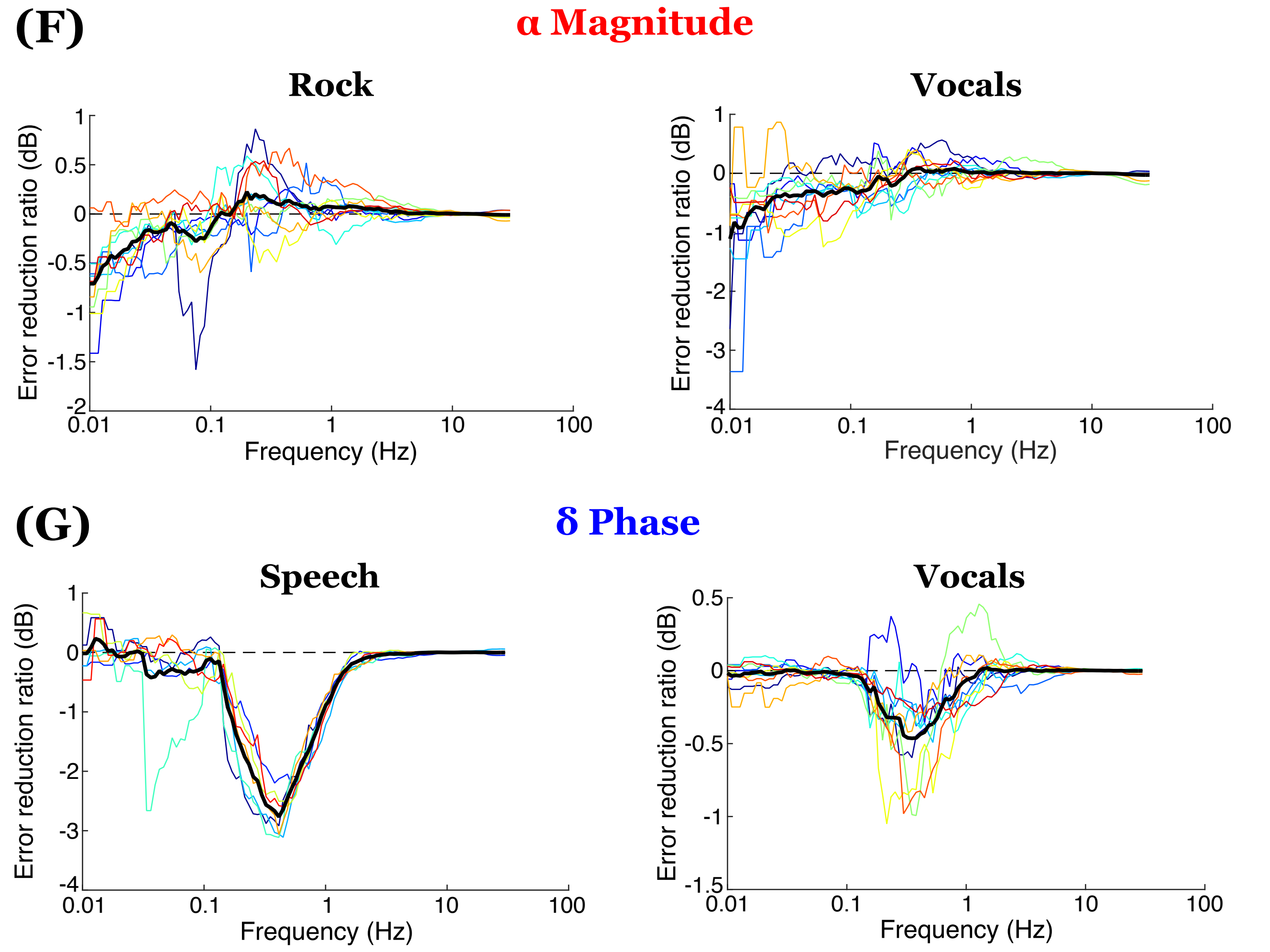
Conclusions and future work

- Distinct components of the EEG signal are involved when reconstructing speech and rock music envelopes, suggesting that two different temporal processes encode these stimuli.
- Alpha envelope reconstructs a buildup signal that correlates with rock onset ramps. This may be some general attention signal [4], or it may relate to the strength of temporal patterns in the music [5].
- Delta fine-structure encodes speech and may include separate activity from both the auditory cortex and parietal area.
- Beta phase is clearly relevant for rock and speech rconstructions, but no clear results were observed using the analysis described here. How it relates to rock music and speech requires further study.
- This analysis could be used in the future to study the neural processing of continuous musical recordings instead of synthetic stimuli as is typically done for music perception research.

Delta phase and alpha magnitude reconstruct very low-frequency fluctuations in speech and music envelopes respectively

$$s(t) \xrightarrow{Fourier} S(f)$$

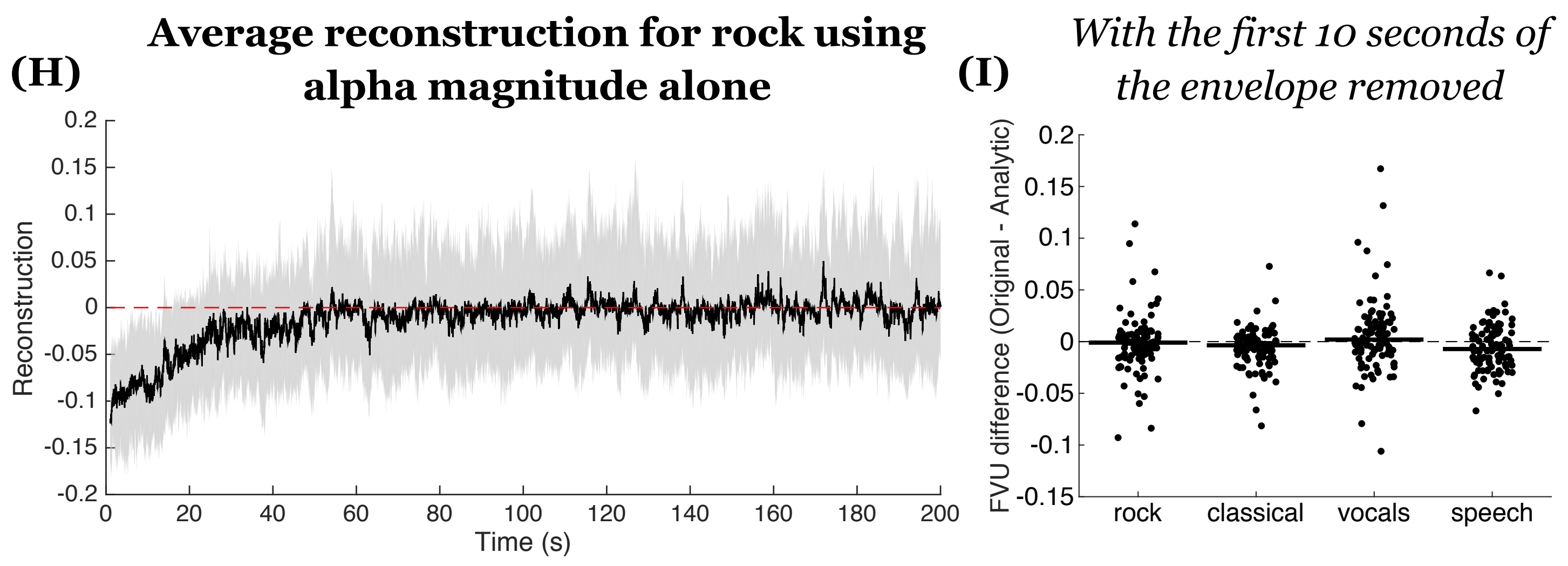
Error reduction ratio: $$= \frac{\sum_f [S(f) - \hat{S}(f)]^2}{\sum_f S(f)^2}$$



In order to understand the specific contribution of the EEG component to the reconstruction, all reconstructions in this analysis were created using the individual component alone. Error reduction ratios were computed with an octave bandwidth surrounding each frequency. In all plots, colored lines are the results for an individual stimulus, averaged across subjects, and the thick black lines is the average across all stimuli.

(F) Reconstructions based on alpha magnitude have reduced errors below 0.1 Hz. The minimum error occurs around 0.01 Hz, which refers to a period that is roughly half the duration of the stimuli. **(G)** In contrast, delta phase reconstructs a range of frequencies with a minimum at 0.4 Hz, whose period is roughly the duration between silences in the speech, as well as the duration of phrases.

Alpha magnitude primarily captures the onset ramp of rock music envelopes



(H) By averaging the reconstructions using alpha magnitude across trials and subjects (black line is the median, gray region is the interquartile range), it's clear that the reconstruction captures an intiail buildup signal that is correlated with the onset ramp of the rock envelopes. These onsets in the rock envelopes last about 10 seconds. **(I)** When the first 10 seconds of the envelopes are removed and the reconstructions are analyzed, the improvement found in (C) disappears (Wilcoxon signed-rank test, p > 0.05 for all stimuli). This validates that the improved reconstructions for rock music and vocals observed earlier is due to alpha magnitude's reconstruction of the a buildup that is correlated with the onset ramp.

References

[1] Norman-Haignere et al (2015), *Neuron*, 88(6):1281-1296.
[2] O'Sullivan et al (2015), *Cerebral Cortex*, 25(7):1697-1706.
[3] Ding & Simon (2012), *J Neurophysiol*, 107(1):78-89.
[4] Wöstmann et al (2016), *Proc Natl Acad Sci*, 113(14)3873-3878.
[5] Barascud et al (2016), Proc Natl Acad Sci, 113(5):E616-E625.

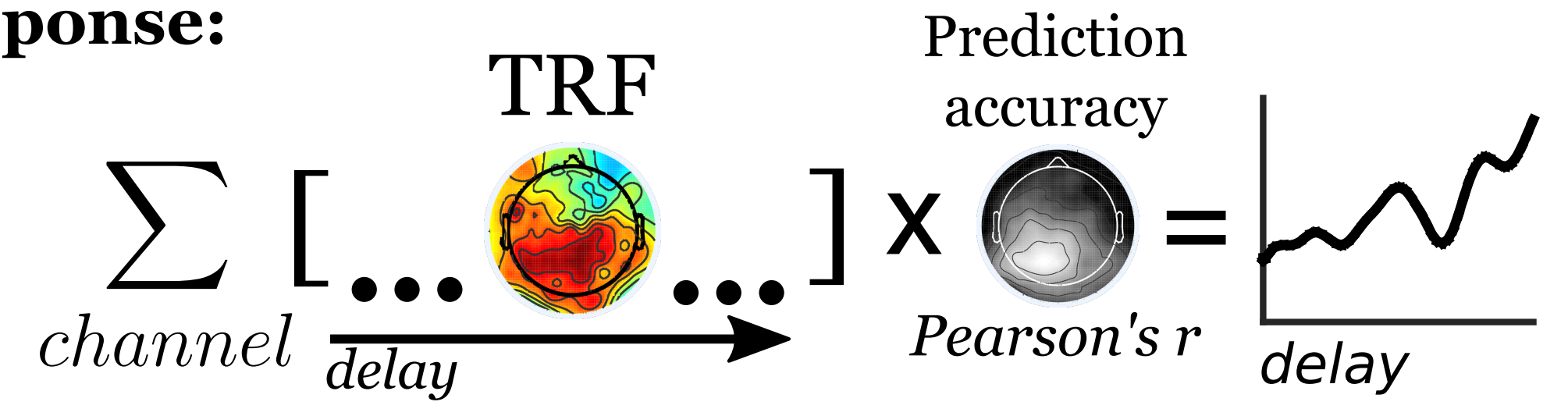
Acknowledgements

This work was funded by an RPG2013-1 grant through the Irish Research Council, a career development award (CDA/15/3316) from Science Fountation Ireland, and by startup funds through the Biomedical Engineering Department at the University of Rochester.

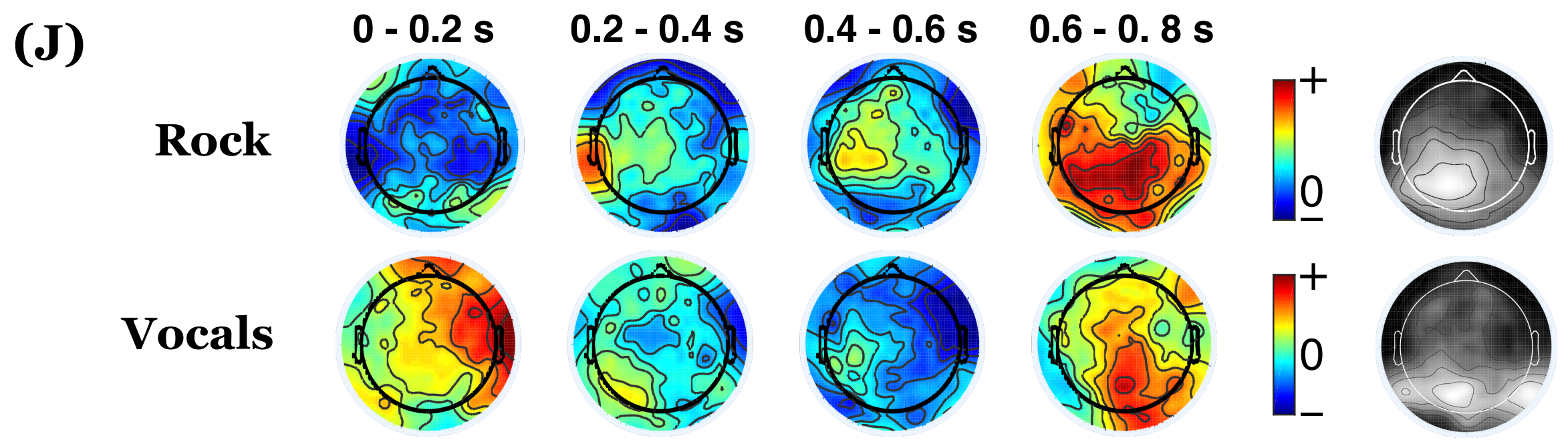
Alpha envelope and delta fine-structure have distinct spatial and temporal activation patterns

EEG prediction: $r(t,n) = \sum_t s(t-\tau)\underbrace{w(\tau,n)}_{TRF}$

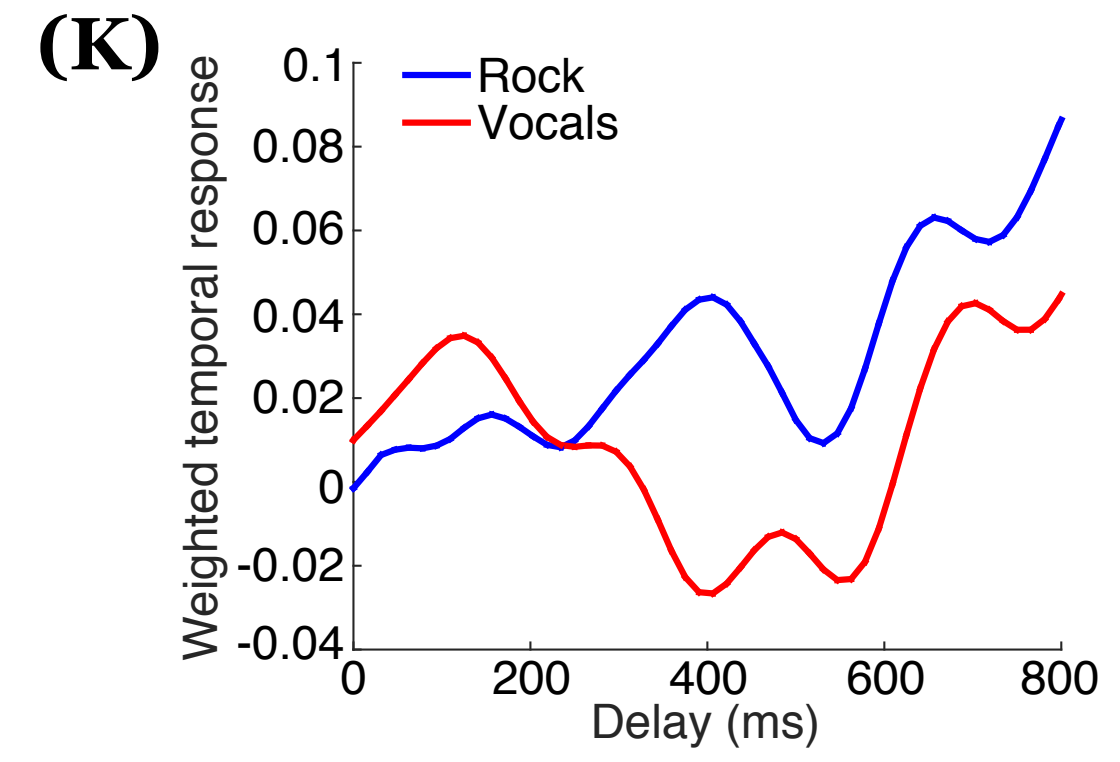
Weighted temporal response:



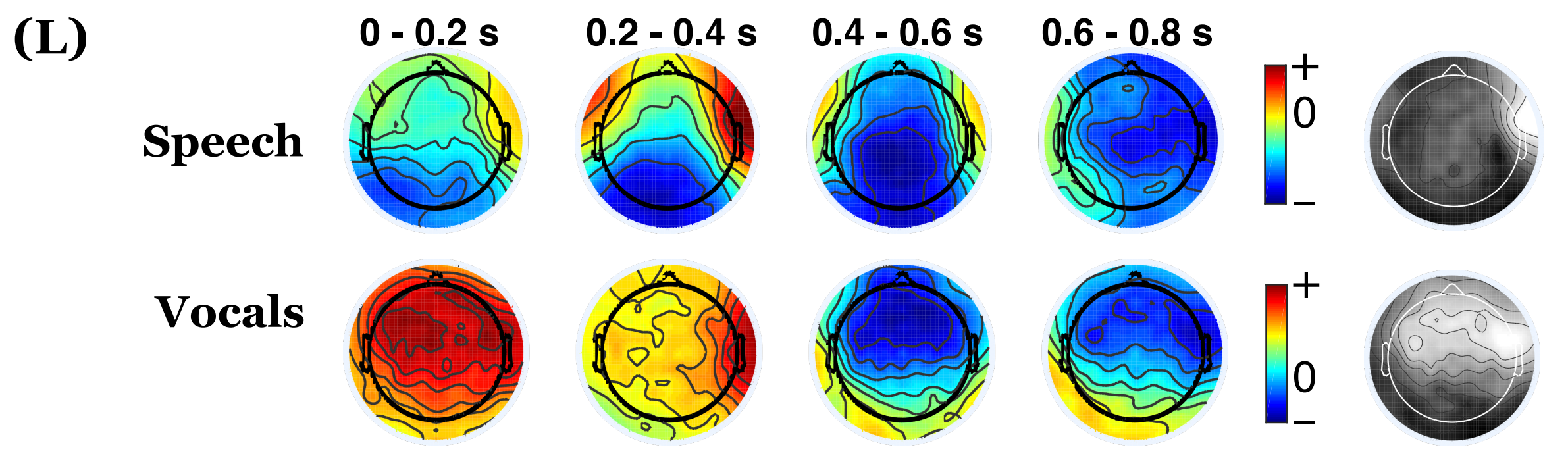
Alpha magnitude activation primarily focused in parietal cortex

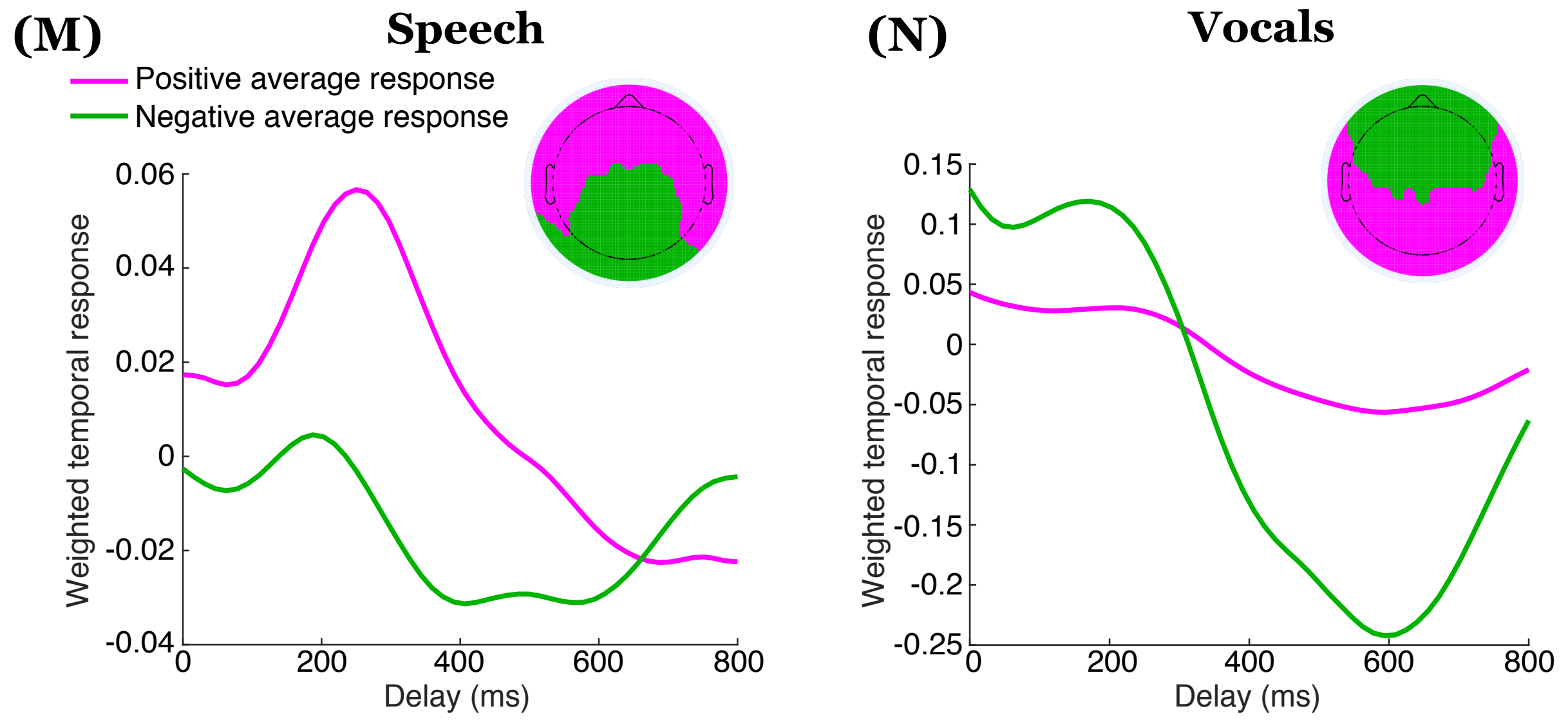


TRF models were created using delays between -200 to 1000 ms. Only the range between 0 - 800 ms is shown. **(J)** EEG predctions of alpha magnitude were best around parietal regions for both rock music and vocals. **(K)** A weak negativity can be found between 400-600 ms. The temporal response for rock music instead appears to modulate with a periodicity of about 200 ms, but this is less clear.



Delta phase activation consists of two spatially separated temporal components





(L) The temporal response for delta phase appeared to contain two different components, one mor lateral that contained a positivity, and a second more central that contained a negativity. Prediction accuracies for speech tended to be best laterally on the scalp (r ~ 0.11), with a weaker but still notable accuracy centrally (r ~ 0.06). The positive (purple) and negative (green) components were analyzed separately by grouping channels based on their average TRF parameters relative to the median across channels. **(M)** For speech, the weighted temporal responses for channels with a positive response is clearly different thant the temporal function for channels with negative weights. **(N)** For vocals, this distinction is less clear, suggesting a neural process using delta phase may be active during speech but not vocals.