Audio Engineering Society

Late Breaking Demo Paper

Presented at the AES International Conference on
Artificial Intelligence and Machine Learning for Audio
2025 September 8–10, London, UK

Decoding Melodic Acoustic Features from Neural Data

Zorka Bozilovic[1] and Iran R. Roman[1]

[1] *Queen Mary University of London*

Correspondence should be addressed to Zorka Bozilovic (`zorka.bozilovic@gmail.com`)

ABSTRACT

We decoded acoustic features of music from EEG using machine learning, focusing on envelopes and onsets and comparing musicians and non-musicians. Results showed higher accuracy when decoding from musicians, especially in the theta band. These findings highlight measurable differences in neural representation of music and suggest potential applications in neural audio engineering, assistive technologies, and rehabilitation.

1 Introduction

This late-breaking paper investigates how auditory stimuli relate to brain activity by decoding acoustic features of music from EEG. This deepens understanding of how music transforms into neural signals and suggests directions for neural audio engineering (e.g., reconstructing imagined sounds) which may benefit assistive technologies and neurological rehabilitation [1]. Brain encoding and decoding have both been studied with audio [2]: encoding predicts EEG, while decoding reconstructs stimuli. A recent study examined musical feature encoding across EEG frequency bands [3], but did not explore decoding. We applied machine learning to a public EEG dataset recorded as participants listened to melodies. **Our paper's contributions are:** (1) Decoded melodic envelope and onsets from EEG; (2) Evaluated decoding across musicians vs. non-musicians and EEG bands; (3) Open-sourced the code.[1]

[1] `github/ZorkaBozilovic/audio-brain-decoding`

2 Methods

Dataset. We used the dataset from [4] (analyzed in [3]), which includes EEG (64ch) from 10 non-musicians and 10 pianists. Each individual passively listened to 10 melodies while EEG was recorded. This was repeated three times, yielding 30 trials per participant. EEG was re-referenced to the mean of the two mastoid channels.

Preprocessing. EEG was filtered into delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), and beta (12–30 Hz) bands[2]. We extracted the envelopes and onset features from the dataset's melodies.

Decoding acoustic features from neural data. We used the temporal response function (TRF, a linear model mapping neural responses to stimulus features via convolution [2]). When decoding, EEG is the input and musical features are the output. One model was fit

[2] Gamma was excluded due to the sampling rate in the pipeline [2].

per participant, with separate models for envelope and onsets. We used `eelbrain`'s boosting function with a 300ms window and evaluated decoding accuracy using Pearson's r between predicted and actual features.

3 Results

Figure 1 shows decoding performance across frequency bands. Correlation coefficients were significantly higher for musicians than non-musicians, with the theta band (4–8 Hz) yielding best decoding overall. Envelope decoding generally outperformed onsets decoding.

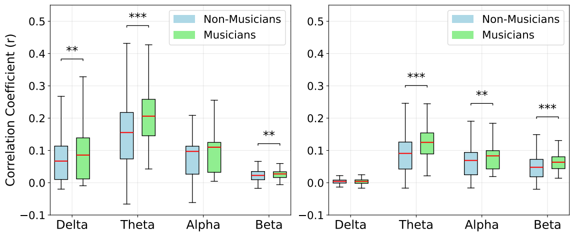

Fig. 1: Decoding correlation coefficients across EEG bands for envelopes (left) & onsets (right). The mean is shown in red. T-tests show significant differences across non-musicians & musicians (***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$).

Figure 2 shows that musicians had stronger filter amplitudes and stronger voltage field, particularly over right-lateralized channels. This aligns with prior findings on hemispheric lateralization for musicians [5].

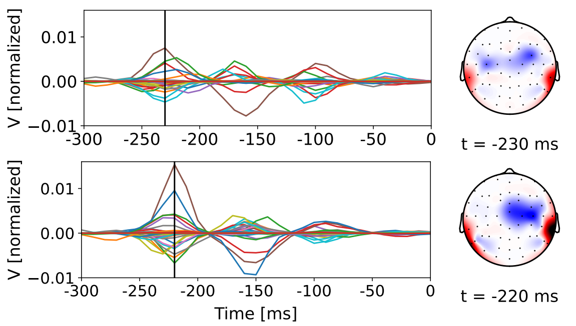

Fig. 2: Mean filters learned to decode musical envelopes from theta-band. Left: butterfly plots show the filter for each EEG channel. Right: topographic maps at peak latencies show spatial distribution of voltage field strength. Top: non-musicians (N=10); bottom: musicians (N=10).

Figure 3 shows three representative decoding examples. The top (who is a musician) shows closer alignment in shape and timing between decoded and true envelope, particularly toward the end. In contrast, non-musicians (middle and bottom) show reduced synchrony.

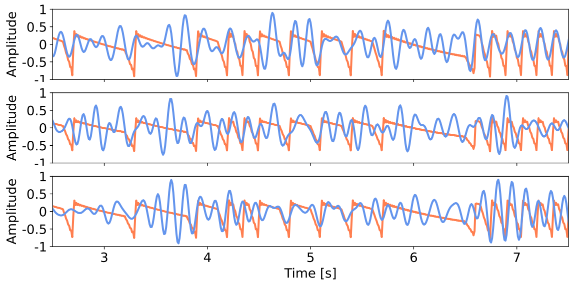

Fig. 3: Examples of decoded (blue) vs. true (orange) envelopes in the theta band. Top: high correlation ($r = 0.43$, musician); middle: average performance ($r = 0.18$, non-musician); bottom: poor decoding ($r = -0.02$, non-musician).

4 Conclusion

Musical expertise enables more accurate decoding of musical features from EEG. Performance was highest in the theta band. Our findings suggest that musical training yields measurable differences in brain responses to music, offering valuable insights for future neural audio engineering interfaces.

References

[1] Moses, D. A. et al., "Neuroprosthesis for decoding speech in a paralyzed person with anarthria," *New England Journal of Medicine*, 385(3), 2021.

[2] Brodbeck, C. et al., "Eelbrain, a Python toolkit for time-continuous analysis with temporal response functions," *eLife*, 12, 2023.

[3] Galeano-Otálvaro, J.-D. et al., "Neural encoding of melodic expectations in music across EEG frequency bands," *European Journal of Neuroscience*, 60(11), pp. 6734–6749, 2024.

[4] Di Liberto, G. M. et al., "Cortical encoding of melodic expectations in human temporal cortex," *Elife*, 9, p. e51784, 2020.

[5] Koelsch, S. et al., "Effects of musical expertise on the early right anterior negativity: an event-related brain potential study," *Psychophysiology*, 2002.