Decoding Music's Amplitude Envelope from Neural and Cardiovascular Signals

Emily Graber[1,2], Mateusz Solinski[3], Elaine Chew[3]

[1]STMS/CNRS/IRCAM, Sorbonne Université, Ministère de la Culture, Paris, France
[2]Department of Computer and Information Science, Allegheny College, Meadville, PA, USA
[3]Department of Cardiovascular Imaging, School of Biomedical Engineering & Imaging Sciences, Faculty of Life Sciences & Medicine and Department of Engineering, Faculty of Natural, Mathematical & Engineering Sciences, King's College London, London, UK

Poster Summary

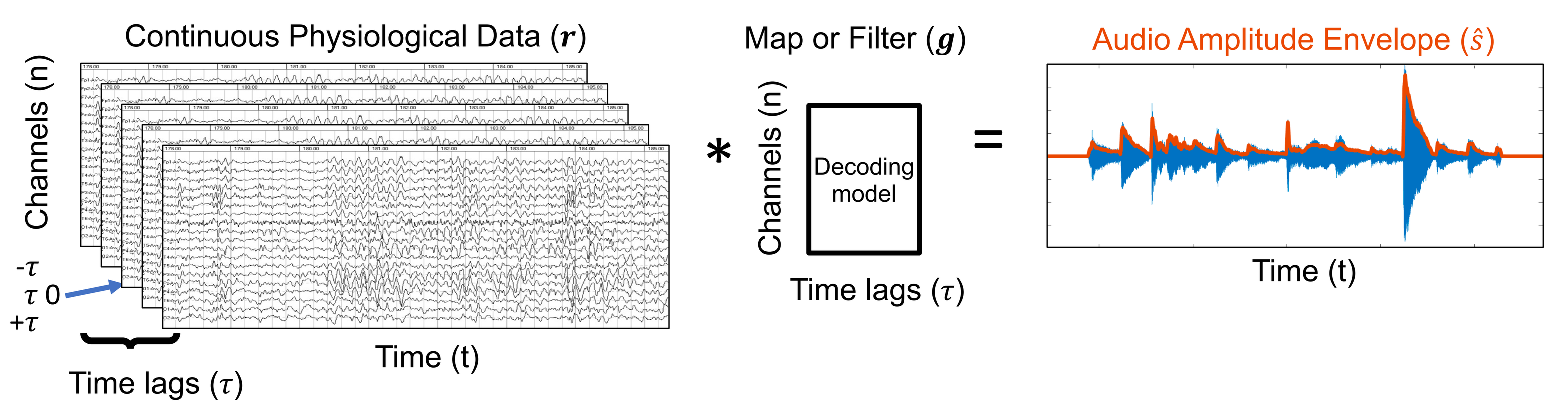
Cardiovascular signals during continuous music listening are analyzed with traditional electroencephalographic (EEG) decoding methods.

Amplitude envelope of music is more accurately reconstructed from cardiovascular signals with EEG than from EEG alone in live listening.

Background

Decoding

"Decoding" is a linear modeling method that maps continuous physiological data to the stimulus that generated it, typically audio amplitude envelope.



For EEG, the neural response data $\boldsymbol{r}$, with $\boldsymbol{t}$ time points and $\boldsymbol{n}$ channels is lagged, i.e. duplicated at delays $\boldsymbol{\tau}$, and weighted for each channel and delay, $\boldsymbol{g}(\boldsymbol{\tau},\boldsymbol{n})$, to optimally reconstruct a feature trace, $\hat{s}$, approximating the real feature trace, s, derived from the stimulus [1,2].

$$\hat{s}(t) = \sum_{n}\sum_{\tau} \boldsymbol{r}(\boldsymbol{t}+\boldsymbol{\tau},\boldsymbol{n})\boldsymbol{g}(\boldsymbol{\tau},\boldsymbol{n})$$

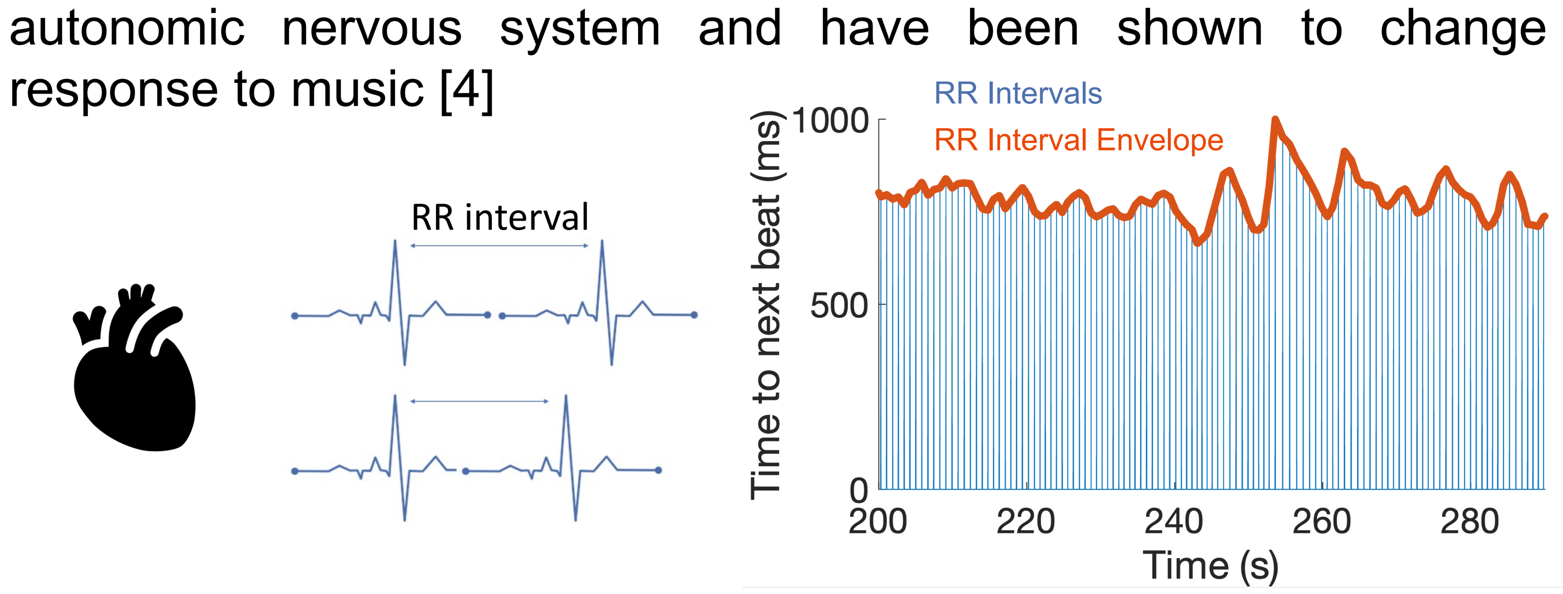
The dimensionality of the response data $\boldsymbol{r}$, increases with each lag added. -Lagged data introduce *future* data into the present prediction, +lagged data introduce *past* data into the present prediction

Reconstructed feature traces, $\hat{s}$, can be compared by correlation to the real feature traces, s, to evaluate decoding effectiveness and to determine the utility of the feature trace selected or the listener's state of focus [3].

Cardiovascular Signals

Electrocardiography (ECG) measures continuous electrical signals from the heart.

Inter-heartbeat (RR) intervals extracted from ECG are mediated by the autonomic nervous system and have been shown to change in response to music [4]



Question

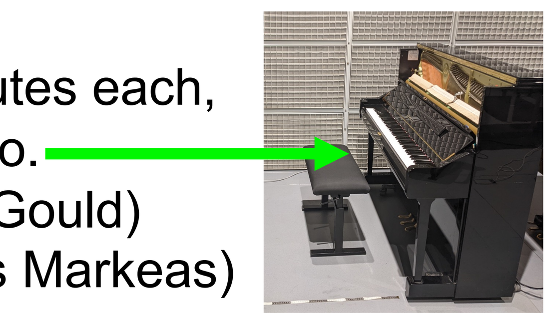
Can ECG RR interval envelope be used to reconstruct the amplitude envelope of music and reveal listener engagement?

contact: egraber@allegheny.edu

Methods

Stimuli

Two pre-recorded solo piano excerpts, 15 minutes each, re-rendered acoustically on a reproducing piano.



- Baroque, Bach Goldberg Variations (Glenn Gould)
- Contemporary, Markeas Improv (Alexandros Markeas)

Participants

24 subjects. 22 analyzed: 11 female, mean age 30.52 (SD = 7.5), mean years musical training 14.5 (SD = 10.2), normal hearing

Task

Listen to piano excerpts and rate personal liking and perceived public liking of the music on a scale from -10 (hate) to 10 (love)

ECG Belt

EEG Cap

EEG Recording

7 channels (Fpz, Fz, Cz, Pz, Oz, FC1 and FC2), Enobio Neuroelectric system with USB connection to MacBook Pro laptop. 500 Hz sampling rate, initial reference left mastoid.

ECG Recording

Bluetooth Polar H10 chests strap at 130Hz with one channel located centrally just below the sternum. An additional EEG channel was used to record ECG at 500Hz from a second location near the 4th intercostal space along the left margin of the sternum.

Data Alignment + Acquisition

EEG, ECG, audio onset/offset triggers, and rendered audio were synchronized and recorded with *Lab Streaming Layer* [5].

Preprocessing

EEG were cleaned in MATLAB with EEGLAB [6]. Data were band-passed 0.1– 40 Hz, downsampled to 100 Hz, decomposed with ICA to remove ocular artifacts, and re-referenced to the common average.

R peaks were identified and extracted from ECG. RR envelopes were created and interpolated to 100 Hz.

Decoding

Linear decoding was carried out using the mTRF Toolbox v2.3 [1,2]. EEG was z-scored collectively for each participant. RR envelopes were z-scored collectively for each participant. Audio RMS envelope and EEG low pass filtered to 10Hz. Lags of -100 – 400 ms were used.

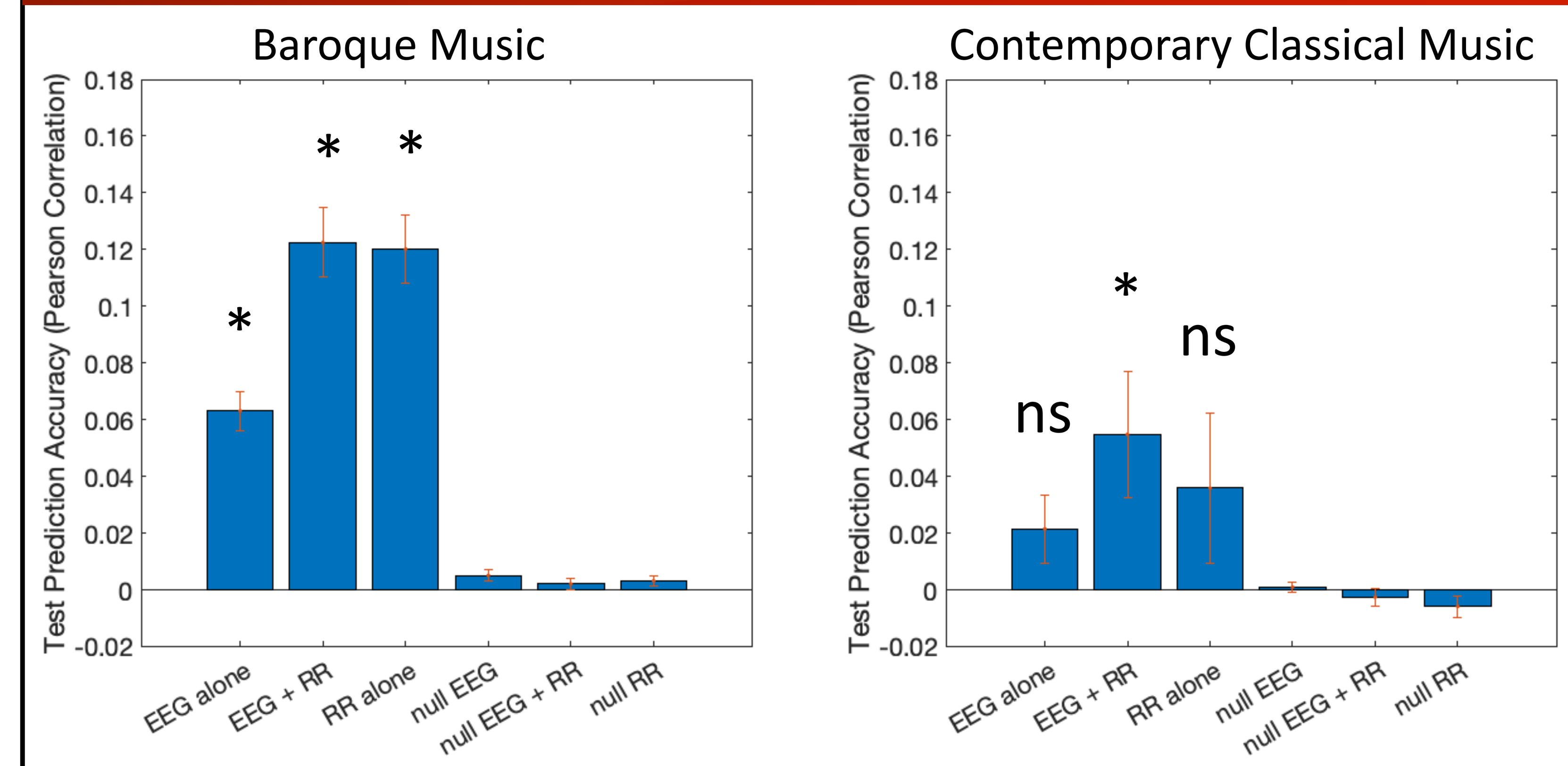
Three Model Types: EEG alone | EEG + RR env | RR env alone

Evaluation

10-fold leave-one out cross validation was used to estimate model performances. Null distributions were created by repeatedly dephasing the audio and physiology data. Additional modeling was run with simulated physiology data with properties similar to the RR envelopes.

Pearson correlation computed between predicted and actual amplitude envelopes for all model types.

Results



For the Baroque Music

Music envelope could be decoded in (1) the EEG data alone, (2) a combination of the EEG and RR interval envelopes, and (3) from the RR interval envelopes alone

RR interval envelopes always improved the modeling accuracy.

For the Contemporary Music

Only EEG + RR data produced a model that could predict the music amplitude envelope above the noise floor.

Perspectives

Preliminary results suggest that added ECG can significantly improve EEG decoding accuracy, particularly for the Contemporary stimuli. However, as only seven EEG channels were recorded, the contributions that other channels could have made are unknown.

Modeling success may depend on the specific stimuli or the consistency of the physiological responses to the stimuli. Responses to Contemporary music could be variable, or subjects may not engage favorably with Contemporary music.

Low frequency components of EEG produced during speech and music listening have been found to track speech more strongly than music [7], thus RR envelopes, which are on the order of 1 Hz, require further investigation and validation.

Future work could investigate if ECG decoding can illuminate subjective evaluations of the music, musicianship [8], or the processing of high-level musical features like musical phrasing.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłowdoska Curie grant agreement No. 101028532 and the European Research Commission Advanced Grant No. 788960.

References

[1] Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. Frontiers in Human Neuroscience, 10(November), 1–14.
[2] Crosse, M. J., Zuk, N. J., Di Liberto, G. M., Nidiffer, A. R., Molholm, S., & Lalor, E. C. (2021). Linear Modeling of Neurophysiological Responses to Speech and Other Continuous Stimuli: Methodological Considerations for Applied Research. In Frontiers in Neuroscience (Vol. 15).
[3] O'Sullivan, Power, Mesgarani, Rajaram, Foxe, Shinn-Cunningham, Slaney, Shamma & Lalor, (2014). Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. Cerebral Cortex, 25(7), 1697–1706.
[4] Koelsch, S., & Jancke, L. (2015). Music and the heart. European Heart Journal, 36(44), 3043–3048.
[5] Lab Streaming Layer [Computer Software], https://github.com/sccn/labstreaminglayer.
[6] Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods, 134, 9–21.
[7] Zuk, N. J., Murphy, J. W., Reilly, R. B., & Lalor, E. C. (2021). Envelope reconstruction of speech and music highlights stronger tracking of speech at low frequencies. PLoS Computational Biology, 17(9).
[8] Saari, Burunat, Brattico & Toiviainen, (2018). Decoding Musical Training from Dynamic Processing of Musical Features in the Brain. Scientific Reports, 8, 708(1-12).

This poster was presented virtually at the 2023 International Conference for Music Perception and Cognition, Tokyo, Japan