Exploring the Relationship Between Music and Human Physiology Through Data

Elaine Chang (MIT)
Supervisor: Professor Elaine Chew | Music Theranostics Lab, King's College London

Introduction

Music affects more than emotion and cognition. It also changes physiological processes such as heart rate, respiration, and blood pressure, in listeners and in performers alike. Music cannot be treated as a single variable. Tempo, loudness, rhythm, structure, and a listener's own associations can all effect a person's response, so two pieces can move the body in different directions. Studying this quantitatively requires converting both the music and the physiological response into measurable features.

Research Question: How is raw observation of music and human physiology turned into structured data, and what can a model learn from the result?

My summer work spanned three stages of that pipeline: beat annotation, physiological data collection, and replication of a graph neural network (GNN) built by another researcher in the lab.

Methods

Beat tapping

A recorded live piano performance, which included six pieces and six etudes, needed beat annotations. I marked every beat by ear, listening to each recording repeatedly and at several playback speeds. Tempo and rhythmic features are derived from these marks and aligned against physiological signals.

Physiological data collection

HeartFM study: Participants wore a one- or three-lead ECG, a respiration belt, and a continuous blood pressure monitor, then listened to nine pieces drawn from a 30-piece library, some songs had altered tempos or loudness. Five silent minutes opened and closed each session as a baseline.

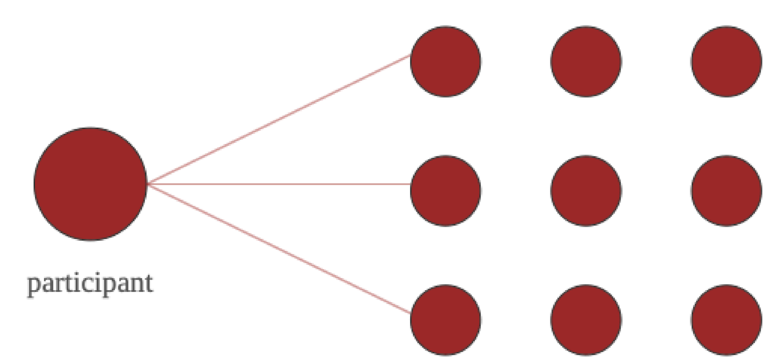
Trio performance study: A pianist, violinist, and cellist repeatedly performed the same pieces, roughly 50 performances over the summer, each wearing a three-lead ECG and respiration belt, with the same silent baselines. Repetition separates a consistent effect from variation between performances.

Graph Neural Network (GNN) Model

20 musical features were extracted from each of the 30 pieces in the HeartFM library, and 20 physiological features from each listening session, 10 from ECG, 10 from respiration, which were normalized before use. A GNN model was replicated from a previous lab project to try and predict hypertension in people listening to music.

Results

The replication used the HeartFM dataset to predict whether a participant was hypertensive. Each node is one participant–track pair holding only that session's physiological features and the edges carry the musical features and passes them in both directions. After pre-processing the data, there were 117 participants with 1,025 usable particpant-track pairs. The model used an 80-20% train-test split on the participant level, meaning that any participant in the training had all 9 nodes in the training to prevent any data leakage.



9 nodes, one per track in the playlist
20 physiological features per node
one hypertension label, shared by all 9

Fig 1: The label describes the person, not the session. The train/test split was made at the participant level, meaning no participant appeared in both training and test sets.

Five slightly different GNN models were made to try and optimize the results and accuracy. The model returned 5 different percentages for the testing data: accuracy, precision, recall, specificity, and F1.

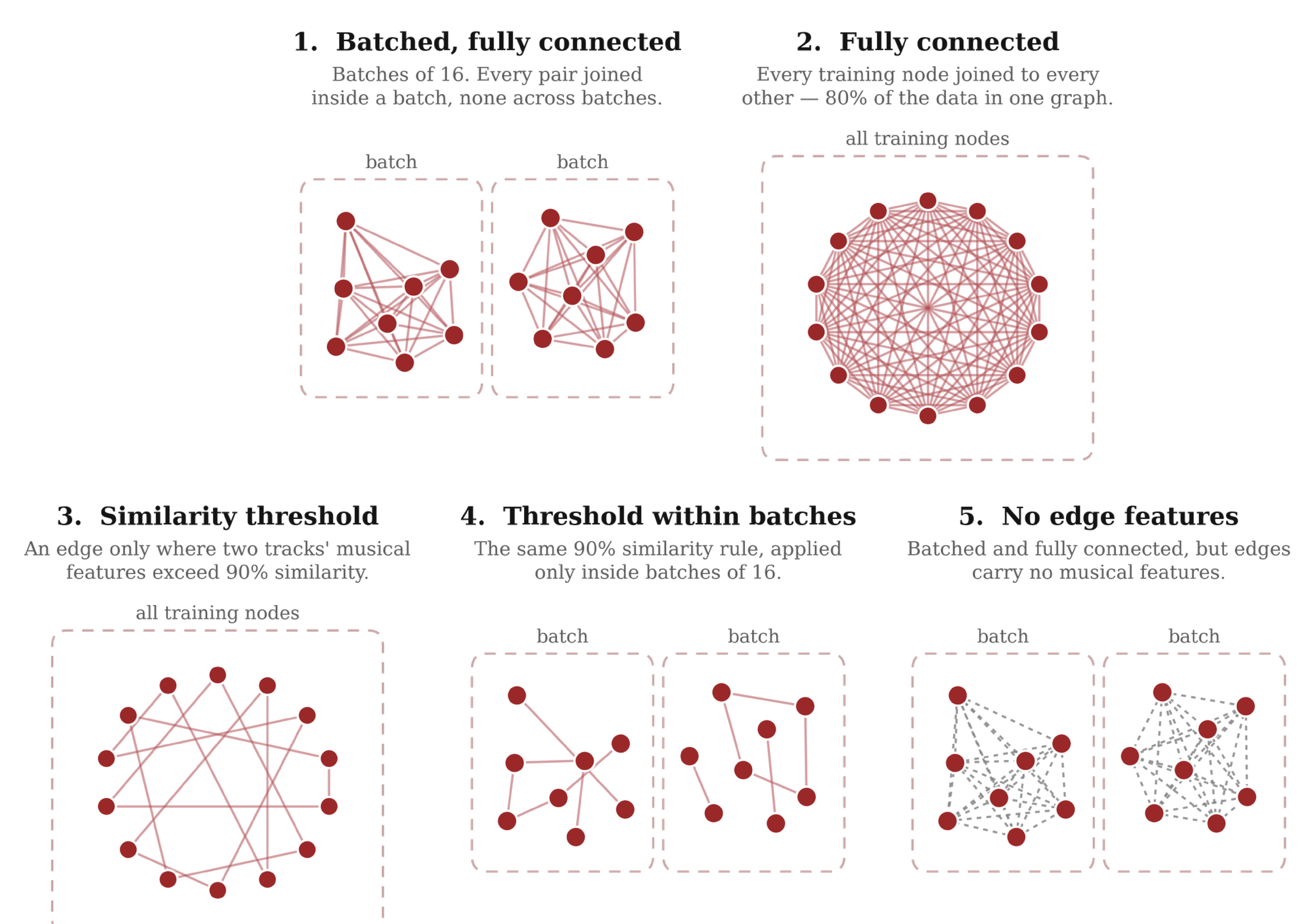


Fig 2: One construction reproduces the original project, the other four include minor differences that were meant to optimize results.

The replication did not reproduce the original result. None of the five graph constructions came close to the reported performance, and results varied substantially from one construction to the next.

Discussion

There are many explanations for why the results failed to meet the original that remain open, such as my pipeline differing from the original somewhere I did not identify. There were also many constraints that made the model difficult to use, a likely one being dataset size. Deep models generally require large quantities of training data, while HeartFM holds a limited number of participants. With little data, a complex model struggles to learn patterns that generalize to a new participant. Class imbalance is another. The dataset contains fewer hypertensive than non- hypertensive participants, so overall accuracy can stay high while the minority class is largely missed. A third issue is the premise of the graph itself. Edges were built on musical similarity, but musical similarity does not imply physiological similarity. Two pieces may share tempo and spectral characteristics while producing different responses in different people.

Conclusion

Three kinds of work fed one question. Beat tapping turned recorded performance into structured annotation, the HeartFM and trio sessions produced the ECG and respiration recordings any analysis depends on, and the GNN tested what could be learned when the two are combined.

The replication of that model did not reach the performance originally reported, and none of the five constructions closed the gap. The prediction task is left open rather than settled. The link between musical features and cardiovascular state is a topic that still needs to be investigated further.

Understanding the relationship between music and human physiology requires collaboration across disciplines and methods. Music has to be quantified without losing the complexity that makes it meaningful, physiological responses have to be measured carefully, and computational models have to be evaluated critically. A great deal of work remains before these relationships are well understood, but this summer gave me a foundation for continuing to explore how music, data, and human physiology connect.

1. Ellis, R. J., & Thayer, J. F. (2010). Music and autonomic nervous system (dys)function. Music Perception, 27, 317–326. https://doi.org/10.1525/mp.2010.27.4.317
2. Pal, P., Cotic, N., Soliński, M., Pope, V., Lambiase, P., & Chew, E. (2024). Music-based graph convolution neural network with ECG, respiration, pulse signal as a diagnostic tool for hypertension. Proceedings of the European Study Group in Cardiovascular Oscillations (ESGCO), Zaragoza, Spain. COSMOS source.
3. Soliński, M., Reed, C. N., & Chew, E. (2024). A framework for modeling performers' beat-to-beat heart intervals using music features and Interpretation Maps. Frontiers in Psychology, 15, 1403599. https://doi.org/10.3389/fpsyg.2024.1403599
4. COSMOS. (2025). Concert and Data at First European Hypertrophic Cardiomyopathy Patient Conference. COSMOS, King's College London.
5. Mejía-Mejía, E., Budidha, K., Abay, T. Y., May, J. M., & Kyriacou, P. A. (2020). Heart rate variability (HRV) and pulse rate variability (PRV) for the assessment of autonomic responses. Frontiers in Physiology, 11, 779. https://doi.org/10.3389/fphys.2020.00779