Exploring the Relationship Between Music and Human Physiology Through Data

Elaine Chang
Music Theranostics Lab at King's College London
Supervisor: Professor Elaine Chew

Abstract

Music produces measurable changes in heart rate, respiration, and blood pressure, but quantifying that relationship is difficult. This paper describes summer research at the Music Theranostics Laboratory at King's College London, including beat annotation of a recorded performance, physiological data collection for two studies, and an attempted replication of a graph neural network predicting hypertension from musical and physiological features. The replication did not reproduce the original results under any of the five graph constructions. Further work is needed to establish whether the gap reflects a pipeline difference or not.

1. Introduction

Music is a constant presence in everyday life. Whether it is a song during a commute, an instrument being practiced, or the soundtrack of a film, people interact with music in countless ways. Because it is so prevalent, music has long been studied for its effects on human emotion, cognition, and behavior. But music influences more than how people think and feel. It also produces changes in physiological processes such as heart rate, respiration, and blood pressure. This raises a broader question: how does music interact with the human body, and how can those interactions be measured and understood?

I explored this question during my summer at the Music Theranostics Laboratory at King's College London led by Professor Elaine Chew. The lab investigates the relationship between music and physiology with the goal of understanding how music can be used in therapeutic and clinical settings. My work consisted of three interconnected components: beat tapping, physiological data collection, and an attempt to replicate an existing graph neural network (GNN) that another researcher in the lab had built from music and physiological data.

Although these activities look very different from one another, each represents a different stage of a single research process. Beat tapping required careful listening to recorded performances and annotating every beat. Data collection required working directly with participants and physiological equipment. Building the GNN required extracting musical and physiological features and applying machine learning to them. Each stage depended on the kind of work done in the others, and the GNN attempted to combine these different types of information into one computational model. The beats I tapped and the data collection sessions I helped run did not feed directly into my own model, which drew on data the lab had collected previously, but the tasks were the same tasks that produce a dataset like that one.

I did not have a single goal for the summer. Instead, I focused on understanding how raw observations of music and human physiology can be transformed into structured data, and ultimately into evidence about the relationship between music and health. This paper describes my work across all three areas, the methods involved, the challenges I encountered, including a replication that did not reproduce the results it was aiming at, and what the experience taught me about conducting interdisciplinary research.

2. Background

The relationship between music and human physiology has been studied for decades. Music influences arousal and emotional state, which in turn affect physiological processes. Changes in musical characteristics such as tempo, loudness, and rhythm can produce different responses in both listeners and performers. Physiological measures such as electrocardiography (ECG), respiration, and blood pressure provide ways of quantifying those responses.

One of the best known examples of research connecting music and human functioning is the "Mozart effect." Early studies suggested that listening to Mozart could temporarily improve performance on certain spatial reasoning tasks. Later research challenged the strength and generality of those claims, but the phenomenon contributed to broader interest in how music influences cognition and physiological arousal. More recent work has taken a more nuanced view, considering how the characteristics of the music, the listener, and the experimental setting interact with one another.

Music therefore cannot be treated as a single variable. Two pieces can produce very different physiological responses depending on their tempo, loudness, rhythm, and structure, and on the emotional associations a listener brings to them. This creates a methodological challenge: how can the relationship between music and physiology be investigated quantitatively?

Answering that question requires converting both music and physiological response into measurable variables. ECG recordings are continuous measurements of cardiac activity, and respiration recordings capture changes in breathing over time. These raw signals contain a great deal of information, but they cannot be used directly in a machine learning model. They must first be cleaned, processed, and normalized into meaningful features. My summer research let me work across all three of the disciplines: music, physiology, and computer science.

3. Beat Tapping

One of the first parts of my summer research was beat tapping, the most direct connection to the musical side of the project. Beat tapping involves listening to a recording and marking the location of every beat. These annotations can then be used to analyze characteristics of the performance and to extract other musical features. Identifying a beat feels intuitive while listening, but marking each one precisely requires sustained attention, especially in passages with changes in rhythm and tempo.

Before I joined the lab, a live performance by a pianist had been recorded and needed beat annotations. Over the summer I tapped the beats for that performance, which included six pieces and six etudes. For each recording I listened multiple times and at several playback speeds in order to place each beat as accurately as I could.

This work provided an interesting contrast to the computational work I did later. Beat tapping let me engage with the music as a listener; when I was building the GNN, the same kind of material appeared only as numerical features. That contrast highlighted an important challenge: many characteristics of music that seem obvious to a human listener still have to be explicitly defined and measured before they can enter a computational analysis.

Beat tapping also showed how small pieces of manually generated data become useful in larger research pipelines. Beat annotations can be used to extract tempo and rhythmic features, which can then be combined with physiological measurements. In this way, a task grounded in human perception of music becomes the basis for quantitative analysis.

4. Physiological Data Collection

I also assisted with physiological data collection for two studies: the HeartFM study and a trio performance study. Working on both gave me a clearer understanding of everything that happens before data ever reaches a machine learning model.

4.1 The HeartFM Study

The HeartFM study investigates physiological responses to music. During a recording session, participants are connected to several monitoring devices. This includes either a one-lead or three-lead ECG, a respiration belt, and a continuous blood pressure monitor.

Participants listened to a playlist of nine classical piano pieces drawn from a larger library used across the study. The pieces could be presented under different conditions, such as altered tempo or loudness. These manipulations allow researchers to investigate whether changes in specific musical characteristics produce corresponding changes in physiological response.

Each session also included periods without music. Participants completed a five-minute silent baseline before listening and another five-minute silent baseline afterward. The baselines provide a reference point against which responses during music can be compared.

My role involved assisting with the collection process: helping prepare participants, confirming that the recording equipment was working properly, and monitoring the session as it ran.

4.2 The Trio Performance Study

The second study involved a live performance by a pianist, violinist, and cellist. Rather than studying people listening to recorded music, it examined physiological responses during performance itself.

Each performer was connected to a three-lead ECG and a respiration belt. As in the HeartFM study, each recording began with a five-minute silent baseline, followed by the performance, and concluded with another five-minute silent baseline.

The performers played the same piece multiple times. Repetition matters here because it creates the opportunity to compare physiological responses across performances and to distinguish consistent effects from variation between one performance and the next. Over the course of the summer, roughly fifty performances were recorded.

Assisting with this study gave me experience with a very different kind of physiological dataset. The live setting introduces a different relationship between performer and music, rather than passively receiving a recording, the musician is actively producing the music while simultaneously experiencing it.

Working on both studies demonstrated the importance of experimental design. The equipment and measurements overlapped substantially, but the research questions and contexts did not. HeartFM allowed researchers to manipulate the characteristics of recorded music, while the trio study focused on repeated live performance. Those differences shape how the resulting physiological data should be interpreted.

5. Replicating a Graph Neural Network

The most computationally intensive part of my summer was a replication. Another researcher in the lab had built a graph neural network that used HeartFM data to predict whether a participant was hypertensive from a graph-based representation of music and physiological features, and had reported strong performance. My task was to rebuild the pipeline from the same data and see whether I arrived at the same place, and then to test alternative graph constructions of my own on top of it.

The first step was to represent the music numerically. Reproducing the feature extraction from an earlier project in the lab, I extracted twenty musical features from each of the thirty pieces in the HeartFM library, from which every participant's nine-track playlist was drawn. These features were chosen to capture different characteristics of the music so that pieces could be compared quantitatively.

The physiological data required a similar transformation. Rather than feeding raw ECG and respiration signals into the model, I extracted twenty physiological features, ten from the ECG data and ten from the respiration data, and normalized them before use.

The next step was to construct a graph holding both the participants' physiological responses and the musical features. Each node represented a participant–track pair, so a participant with a complete session contributed nine nodes, one for each track in their playlist. A node contained only the physiological features recorded while that participant listened to that track. Each participant also carried a binary label indicating whether they were hypertensive. Because that label describes the person rather than any single listening session, all of a participant's nodes share the same label. A complete dataset of 117 participants and the graph contained 1,025 nodes.

The edges represented relationships between nodes and carried the musical features. If node A corresponded to one track and node B to another, node A passed its track's features to node B and node B passed its own features back.

I built five versions of the graph. One reproduced the construction used in the original project; the other four were alternatives of my own, designed to test how sensitive performance was to the way edges were defined. The first used batches of sixteen nodes, with each batch fully connected. The second was fully connected across all training nodes, which made up eighty percent of the data. The third connected nodes only when their musical features exceeded a ninety percent similarity threshold. The fourth applied that same threshold within batches of sixteen. The fifth used batches of sixteen with no musical features on the edges at all.

During graph convolution, information from neighboring nodes was passed to each node. Because the edges carried musical information, the model could take into account both the physiological characteristics of neighboring observations and the musical relationships connecting them.

The models were evaluated on accuracy, precision, recall, specificity, and F1 score. Because every node belonging to a participant carries that participant's label, a split made randomly over nodes would place some of a person's nodes in training and the rest in testing. The model could then score well simply by recognizing a participant it had already seen, rather than by learning anything that generalizes. The split was therefore made at the participant level, meaning every node belonging to a given participant appeared either in the training set or in the test set, never both, with roughly eighty percent of participants assigned to training.

6. Results and Discussion

The replication was not successful. None of the five graph constructions came close to the performance reported in the original project, which remained well out of reach across every version I tried. Results also varied substantially from one construction to the next. The fully connected graph, for example, reached an accuracy of 72.11 percent, precision of 55 percent, recall of 55.93 percent, and specificity of 79.39 percent.

Dataset size was one likely constraint. Deep learning models generally benefit from large quantities of training data, while the HeartFM dataset contains a relatively limited number of participants and observations. With limited data, a complex model can struggle to learn patterns that generalize to new participants.

Class imbalance was another. The dataset contained fewer hypertensive than non-hypertensive participants, so a model could achieve fairly high overall accuracy while performing poorly on the minority class. The gap between accuracy, recall, and specificity reflects exactly that.

A third challenge involved the graph construction itself. The graph was built on musical similarity, but musical similarity does not necessarily imply physiological similarity. Two pieces may share tempo or spectral characteristics while producing different physiological responses in different people, and two very different pieces may produce similar effects. The assumptions behind the graph may therefore not reflect the relationships that actually matter for this prediction task.

7. Conclusion

My summer at the Music Theranostics Laboratory let me approach the relationship between music and physiology from several directions. Physiological data collection gave me firsthand experience with the experimental work required to generate research data. Beat tapping showed me how musical characteristics are manually identified and converted into a form suitable for computational analysis. The GNN project let me rebuild another researcher's model from the same data and then test variations of my own on top of it.

My replication did not reproduce the original model's results. That outcome was still a valuable part of the experience. Rather than showing that the research had failed, the results made clear how difficult it is to predict human physiological outcomes from music, and how much small datasets, class imbalance, feature selection, participant variability, and assumptions about relationships between observations can constrain what a model can learn.

Most importantly, the summer changed my understanding of what research is. It is not simply building a model and obtaining a high accuracy score. It involves asking questions, collecting data, making methodological decisions, encountering unexpected results, and working out what those results can teach you. Moving between data collection, musical annotation, and machine learning let me see that process from several vantage points at once.

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.

References

1. Ellis, R. J. & Thayer, J. F. Music and autonomic nervous system (dys)function. Music Percept. 27,
317–326, DOI: https://doi.org/10.1525/mp.2010.27.4.317 (2010).

2. Pal, P., Cotic, N., Soliński, M., Pope, V., Lambiase, P. & Chew, E. Music-based graph convolution
neural network with ECG, respiration, pulse signal as a diagnostic tool for hypertension. In Proceedings of the European Study Group in Cardiovascular Oscillations (ESGCO), Zaragoza, Spain, 23–25 October 2024 (2024). COSMOS source.
3. Soliński, M., Reed, C. N. & Chew, E. A framework for modeling performers' beat-to-beat heart
intervals using music features and Interpretation Maps. Front. Psychol. 15, 1403599, DOI: https://doi.org/10.3389/fpsyg.2024.1403599 (2024). COSMOS source.
4. COSMOS. Concert and Data at First European Hypertrophic Cardiomyopathy Patient Conference.
COSMOS, King's College London (2025). COSMOS source.
5. Mejía-Mejía, E., Budidha, K., Abay, T. Y., May, J. M. & Kyriacou, P. A. Heart rate variability (HRV)
and pulse rate variability (PRV) for the assessment of autonomic responses. Front. Physiol. 11, 779, DOI: https://doi.org/10.3389/fphys.2020.00779 (2020).