COMP5425 Final Report

Author SID: 520110054, 540648861, 540526237, 540465176
Author Unikey: jche0758, ylia0668, zliu0097, zhwu0482

1. Abstract

In today's fast-paced digital era, individuals are frequently exposed to information overload. It heightens the stress and abrupt mood fluctuations. Despite that, mainstream music platforms rely on history-based recommenders and often overlook the listener's emotional state. This study explores whether real-time biometric signals, beats-per-minute (BPM), can be used to infer emotional states and deliver music that matches the emotion. We develop a lightweight music recommendation system which converts heart-rate variability (HRV) into emotion-aligned music suggestions. Drawing on ten-minute chest-ECG epochs for Calm, Sad and Happy states from six participants in the WESAD dataset (recorded under the TSST paradigm)[1], we extract four HRV features (RMSSD, SDNN, LF/HF ratio, BPM) and train a 100-tree Random-Forest classifier that attains 83 % accuracy on a held-out test set. We map BPM values directly to emotional states using a static dictionary (predictions.json). Then the system retrieves music tracks aligned with the predicted affective state. Overall, the project demonstrates effective interdisciplinary collaboration, integrating physiological signals, data science, and music theory to build a BPM-driven, emotion-aware recommendation system.

2. Introduction

In the era of information overload, individuals are increasingly overwhelmed with digital content. While modern technology provides unprecedented convenience for accessing information, fostering social connections, and delivering entertainment, it simultaneously contributes to heightened anxiety, emotional fatigue, and psychological strain. As a result, there is growing recognition of the need to cultivate self-awareness and emotional resilience. According to Shaffer and Ginsberg's study, "among various physiological signals, heart rate and heart rate variability (HRV) have been proven to be effective indicators for detecting emotional states due to their close association with autonomic nervous system activity" [2]. "Heart rate variability has been used to assess activation of the sympathetic and parasympathetic nervous systems" [3].

Meanwhile, music has long been recognised as a powerful medium for expressing and externalising emotion. As demonstrate by Pring et al. [4], "humans perceive a range of fundamental emotional connotations from music", and listeners' choices can often reflect their momentary psychological states. However, the recommendation engines that power mainstream platforms are fundamentally retrospective and generalised. They rely on historical listening behaviour, such as play counts, likes, and popularity metrics, while disregarding the listener's immediate physiological or emotional state. As a result, these static, history-based algorithms struggle to provide music that aligns with users' present emotional needs, limiting their capacity to support real-time emotion regulation and immersive affective engagement.

These limitations motivated us to create a new platform. The core research question of this project is how might we develop a lightweight, BPM-based emotion inference model that delivers responsive and emotionally resonant music recommendations. Our goal is to explore how a minimal, fast-response design can provide meaningful emotional alignment through music. To address this question, the project set out to achieve four major objectives. First, we aimed to develop an emotion-aware music recommendation system that maps user-provided heart rate (BPM) to one of three emotional states, including Calm, Sad, or Happy, by using the trained model data and rule-based logic as a fallback. Second, the system is a lightweight, web-based interface that enables users to interact with the system in real time without requiring external sensors. Third, we want to create an immersive listening experience by incorporating visual feedback. Finally, we integrated a feedback mechanism that allows users to evaluate whether the suggested track aligns with their emotional experience. These objectives collectively support the overarching goal of delivering a simple and emotionally intelligent music recommendation experience grounded in Zen-inspired interaction design.

This project achieves the following aspects:

1. Physiology-driven music recommendation: We demonstrate that four simple HRV features—RMSSD, SDNN, LF/HF ratio, and mean BPM—can achieve 83% emotion classification accuracy on held-out data, confirming the feasibility of affect recognition from a single physiological signal.
2. Millisecond-level lookup: By exporting model outputs to a static lookup dictionary (predictions.json), Zen Music Recommender (ZMR) enables browser-side recommendations with an end-to-end latency of ≤150 milliseconds, eliminating the need for server-side computation.
3. End-to-end pipeline: We provide a complete and reproducible workflow covering data preprocessing, feature extraction, emotion classification, dictionary mapping, music retrieval, and front-end design, supporting future development and academic replication.
4. Emotion-reflective front-end: The user interface conveys emotional states through dynamic gradients, rotating album art, and a pulse indicator with synchronised breathing animations, enhancing affective immersion.
5. Feedback-driven refinement: A post-playback pop-up invites users to indicate whether the recommended music matched their emotional state (Yes/No), enabling future refinement of accuracy and paving the way for

adopting finer-grained, multidimensional emotion labels as the dataset scales.

(See Sections 4.1 and 4.2 for implementation details)

3. Related Work

Emotion-aware music recommendation systems aim to enhance user experience by aligning music suggestions with the listener's emotional state. Various approaches have been explored so far, ranging from real-time physiological sensing to multimodal emotion recognition techniques. Several studies have investigated using physiological signals, such as heart rate, to predict emotional states for music recommendation. For instance, Ayata, Yaslan, and Kamasak [5] proposed a system that utilises wearable devices to capture ECG signals, feeding this data into a machine learning algorithm to predict emotions and recommend music accordingly. Beyond physiological signals, some systems incorporate multiple methods for emotion detection. For example, certain applications analyse facial expressions using computer vision techniques alongside heart rate data to determine the user's emotional state and suggest suitable music [6]. These multimodal systems aim to improve emotion recognition accuracy by combining various data sources.

While these systems demonstrate the potential of integrating emotion recognition into music recommendation, they often rely on continuous real-time data collection, specialised hardware and complex processing pipelines. This requirement can pose challenges regarding accessibility, user privacy, and system responsiveness. In contrast, our system emphasises simplicity and user autonomy by allowing manual input of heart rate (BPM) values. By utilising a precomputed prediction dictionary derived from a trained machine learning model, we achieve fast-response emotion inference without the need for real-time physiological sensing. This approach reduces system complexity, enhances reliability, and ensures a consistent user experience across devices.

4. Solution

4.1 Emotion Classification Model

To achieve emotion recognition based on physiological signals, this project constructed an emotion classification system consisting of four parts: data preprocessing, feature extraction, classification model training and generation of mapping dictionaries. We used the ECG signals collected through chest-worn devices in the Wearable Stress and Affect Detection (WESAD) dataset and selected data from six subjects, namely S2, S3, S5, S8, S11, and S17, including three females and three males. According to the WESAD data definition, the original tags 1 (Baseline), 2 (Stress), and 3 (Amusement) are respectively mapped to Calm, Sad, and Happy to align with the emotion tags of the music recommendation system. The original ECG signal was continuously collected at a sampling rate of 700 Hz. We intercepted 10 minutes of continuous signals for each emotional stage (a total of 420,000 sampling points) and automatically eliminated the segments where the R peak detection failed to ensure the validity of the input data.

The data preprocessing is divided into four steps: signal cleaning, feature calculation, standardisation processing and label coding.

Firstly, the denoising of ECG signals, the detection of R peaks and the extraction of heart rate sequences are performed using the neurokit2.ecg_process() function. In the feature calculation stage, during the training stage, the nk.hrv() function is called to extract the HRV time-domain features (RMSSD, SDNN) and frequency-domain features (LF/HF ratio) at one time. While in the prediction stage, to optimise the calculation efficiency, it is split the calculation into calls to hrv_time() and hrv_frequency() to extract the time-domain and frequency-domain features, respectively. All features are standardised based on the training set statistics (mean/standard deviation) (StandardScaler), and the text labels are converted into numerical codes (Calm→0, Sad→1, Happy→2) through the LabelEncoder. The final constructed feature matrix contains four feature indicators: time-domain features RMSSD (reflecting the variability between adjacent heartbeats) and SDNN (characterising the overall heart rate variability), frequency-domain features LF/HF (indicating the balance between sympathetic and parasympathetic activity), and BPM (mean heart rate).

The emotion classification model employs a stratified sampling strategy (stratify=y) to divide the training set (70%) and the test set (30%), ensuring a balanced distribution of emotion categories. We choose RandomForest (RandomForestClassifier) as the core classifier. Its advantage lies in reducing the risk of overfitting through the integration of multiple decision trees and naturally supporting the assessment of feature importance. It is particularly suitable for scenarios with low-dimensional HRV features and high noise sensitivity. We set the model hyperparameters to 100 decision trees (n_estimators=100) and fixed the random seed (random_state=42). As a result, a classification accuracy of 83% was achieved on the test set, verifying the effectiveness of feature engineering.

The emotion classification model we established can be called in real time on the server side for emotion prediction. However, to balance the flexibility and efficiency of system deployment, we generated a static mapping dictionary (predictions.json) for the chest ECG signals in the WESAD dataset based on the statistical relationship between BPM and emotion labels. The rounded BPM value is directly mapped to the emotion label. This scheme sacrifices a small amount of accuracy (BPM rounding) in exchange for millisecond-level response speed. It is suitable for low computing power environments such as embedded devices, and achieves cross-platform compatibility through JSON files, facilitating integration into subsequent music recommendation systems.

4.2 Music Recommendation Platform

4.2.1 System Design and Implementation

The music recommendation platform follows a modular architecture consisting of a Flask-based backend, a responsive HTML/CSS frontend with integrated JavaScript logic, and structured data files (music.json, predictions.json) for emotion-tagged content retrieval. This design separates user interaction, emotion inference, and content recommendation into clearly defined components, facilitating maintainability and extensibility. The system maps users' heart rate values to emotional states and recommends music. Unlike traditional affective computing systems, our solution emphasises simplicity and immersive user experience by allowing manual input to support fast-response emotion inference and reduce sensing complexity.

The system is composed of three main components:

(1) **heart rate input interface** allows manual entry of their current heart rate (BPM),
(2) The **emotion inference layer** maps BPM values to emotional categories based on a machine learning model trained by the project team, which achieved approximately 80% accuracy.
(3) **Music recommendation engine**: Suggests music tracks corresponding to the inferred emotional state, emphasising calmness, clarity, and aesthetic harmony.

4.2.2 Emotion Mapping Logic

The system adopts a combined approach to emotion inference, including rule-based logic with data-driven model predictions. The default fallback rule maps BPM values below 65 to Sad, above 100 to Happy, and values in between to Calm. When available, a precomputed prediction dictionary (predictions.json), which was generated by the previously trained machine learning model, overrides the rule-based mapping. This dictionary maps selected BPM values to predicted emotional labels with improved accuracy. Rather than integrating the model directly into the runtime environment, we export its predictions as a static dictionary. This decision was based on both technical and contextual considerations. The current version of the model was trained and evaluated on a set of static physiological samples, each representing a single snapshot of user state. It was not designed for continuous, real-time inference or streaming input. Embedding such a model into the live Flask application would require additional processing pipelines for feature extraction and device-level sensing integration.

Given that the emotion classification task is based mainly on BPM, we found that the precomputed mapping retained sufficient accuracy while simplifying deployment. This approach eliminated the need for server-side ML runtimes or heavy dependencies, enabling a faster, more responsive frontend experience. It also improved system reliability, ensuring consistent behaviour across devices. Furthermore, this design remains extensible: future iterations of the model can be retrained simply by updating or replacing the dictionary, without altering system logic or architecture. Once the emotion is inferred, the system filters a local music library (music.json) for tracks labelled with the corresponding emotion. One track is randomly selected from the filtered list. Its metadata (title, artist, file path, album cover) is stored in Flask's session object to persist across page reloads or redirects. This ensures consistent playback and enables the system to maintain state without requiring resubmission of the input.

4.2.3 Frontend Design

To support a seamless and responsive user experience, the system employs a session-based architecture with a Post-Redirect-Get (PRG) workflow. Upon submitting the BPM input, the emotion is mapped, and the system selects an appropriate music track. The system redirects the user to a dedicated result page, avoiding browser warnings about resubmitting form data. The selected track's metadata is retrieved from the session and used to render the playback interface.

The frontend interface is designed to visually convey the user's inferred emotional state. Background gradients dynamically shift based on the emotion, the album artwork rotates, and a pulse dot animates with synchronised "breathing" effects. These animations are scaled in real time according to the user's BPM input, visually mirroring the rhythm of their physiological state. A notable feature is the feedback popup, which is triggered upon completion of the track. This floating dialogue prompts users to assess the emotional relevance of the recommended music using a binary choice (Yes / No). The popup follows a glassmorphism design style, with blurred backgrounds, rounded edges, and gradient highlights. All interactive elements—including "Start," "Stop," and feedback buttons—share a unified visual language, promoting design consistency and reducing cognitive load.

To address autoplay restrictions imposed by browsers, the audio player is initialised with muted autoplay and subsequently triggered using JavaScript upon page load. Client-side validation is also implemented to ensure that BPM inputs fall within a realistic physiological range (40–180 BPM). When invalid input is detected, the system provides animated visual cues and error messages to gently prompt correction. Collectively, these design and implementation choices aim to foster a calm, meditative user experience consistent with the system's Zen philosophy, while preserving reliability, responsiveness, and emotional relevance.

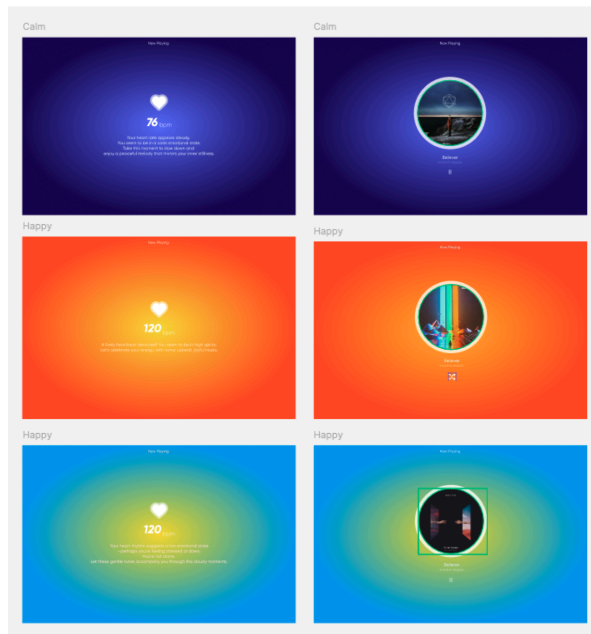

Figure 1: Front-end Interface Design, Calm, Happy, & Sad

5. Limitations and Opportunities for Future Development

5.1 Limitations of the emotion Classification Model

Although the emotion classification model has a relatively systematic design in terms of data preprocessing, feature extraction, model training and system deployment, there are still several limitations. Firstly, the limited scale of the dataset and insufficient subject representativeness can easily lead to overfitting of the model and weak generalisation ability. Secondly, the process of emotional label mapping has semantic simplification, ignoring the complexity and diversity of emotional states. Furthermore, the feature dimension lacks nonlinear HRV features, which limits the model's ability to distinguish complex emotions. Although the static BPM mapping dictionary strategy improves the deployment efficiency, it results in severe information loss and is difficult to support dynamic scenarios.

To overcome the above limitations, future work can be improved in the following aspects. Firstly, the scale of the data set should be further expanded, and more subjects with different ages, genders and health statuses should be introduced to improve the generalisation ability and adaptability of the model. Secondly, we should consider a more fine-grained and multi-dimensional emotional labelling system and combine it with subjective questionnaires or multimodal data (such as the fusion of facial expressions, speech, and physiological signals) to depict emotional states more realistically. In terms of feature extraction, nonlinear HRV indicators and deep learning methods can be introduced to automatically extract time-series features, to enhance the recognition ability of the model for complex emotional states. In addition, future research can explore cross-subject adaptation algorithms to reduce the impact of individual differences on model performance and achieve wider population adaptation.

5.2 Limitations for Music Recommendation Platform

Emotion recognition and mapping were implemented using a hybrid approach. The original plan proposed training a machine learning model on physiological features derived from the WESAD dataset, targeting three emotional states: *calm*, *sad*, and *happy*. This objective was fulfilled, and the team-trained model achieved an accuracy of approximately 80%. In the final system, the model predictions are stored in a dictionary (predictions.json) for lightweight integration. While real-time inference was initially considered, this was pragmatically replaced with static mapping to maintain architectural simplicity. A fallback rule-based logic was also implemented to ensure robustness when model predictions are unavailable. The music retrieval component operates as planned by filtering a local music library (music.json) based on emotion tags. Although the proposal suggested a top-N retrieval strategy using additional musical features (e.g., valence, tempo), the final system opts for emotion-level filtering followed by a random selection of one matching track. This simplification is reasonable given the scale of the music dataset and the project's emphasis on responsiveness and interaction minimalism.

On the interface and interaction side, the final implementation substantially exceeds initial expectations. The proposal briefly mentioned a basic UI for BPM input and music playback. In contrast, the final system includes a well-styled frontend featuring emotion-synchronised background gradients, animated breathing effects on the album cover, and a pulse indicator that visually reflects the user’s BPM. The user interface adopts a Zen-inspired aesthetic, leveraging soft gradients, glassmorphism components, and unified visual styling for all interactive buttons. Input validation and animated error feedback (e.g., input shake) further enhance the user experience.

A particularly notable enhancement is the inclusion of a **user feedback mechanism**, which was not originally specified in the proposal. Upon music playback completion, a floating pop-up invites the user to evaluate whether the recommended music matched their emotional state. It lays a strong foundation for future evaluation of recommendation accuracy and user satisfaction. The system also adheres to sound architectural practices, adopting a **Post-Redirect-Get (PRG)** pattern to prevent duplicate form submissions and improve session stability. Autoplay constraints imposed by modern browsers were addressed through muted preload strategies and JavaScript-triggered playback initiation.

In summary, the implemented system meets all essential functional goals described in the proposal, while also introducing several additional features—including consistent visual design, session handling, and a structured feedback loop—that contribute to a refined and emotionally engaging user experience. Limitations, such as the absence of real-time physiological sensing and a more granular music retrieval strategy, are acknowledged and positioned as future work.

6. Results

The Music Recommender was successfully implemented and evaluated across key dimensions of functionality, responsiveness, and user experience. The final system fulfils its intended purpose of providing lightweight, BPM-driven music recommendations with a seamless and aesthetically coherent interface. The application demonstrates consistent performance across major browsers and platforms. The use of a Post-Redirect-Get (PRG) workflow and session-based state management allows users to revisit the result page without data loss or unintended form resubmissions. Audio playback is reliably triggered using JavaScript-based autoplay fallback, and animations (breathing cover, pulse indicator) are dynamically adjusted to reflect the input BPM. Although the deployed system does not perform real-time model inference, it uses a static dictionary derived from a trained machine learning model with an estimated accuracy of 80%. This mapping enables sufficiently accurate emotion inference from BPM values. For BPM inputs included in the prediction dictionary, the system retrieves the model's emotion output; otherwise, it uses a fallback rule. This hybrid approach ensures complete coverage while retaining model fidelity where possible.

The local music library is categorised into three emotional states (*Calm*, *Sad*, *Happy*). Upon emotion inference, one track is randomly selected from the corresponding emotion group. In all tested cases, the system returned valid and thematically appropriate audio files along with the associated metadata (title, artist, album image). Album artwork and track name are displayed in the interface and animated synchronously. A feedback popup window is triggered after track playback, allowing users to indicate whether the recommended music aligned with their emotional state. Preliminary informal testing with student users showed that the majority found the system intuitive and the feedback interface non-intrusive. The final interface incorporates Zen-inspired visual design, including glassmorphism buttons, emotion-reflective background gradients, and smooth micro-interactions (e.g., hover effects, animated error feedback). Input validation is enforced with soft animations (shake + message fade) to maintain user flow. The design promotes emotional resonance and clarity, contributing to the overall effectiveness of the system.

References

[1]Schmidt, P, Reiss, A, Duerichen, R, Marberger, C & Van Laerhoven, K 2018, 'Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection', in Proceedings of the 20th ACM International Conference on Multimodal Interaction, ACM, New York, NY, USA, pp. 400–408.

[2] Shaffer, F, & Ginsberg, JP 2017, 'An Overview of Heart Rate Variability Metrics and Norms', Frontiers in Public Health, vol. 5, pp. 258–258.

[3] Iwanaga, M, Kobayashi, A. & Kawasaki, C 2005, 'Heart rate variability with repetitive exposure to music', Biological Psychology, vol. 70, no. 1, pp. 61–66.

[4] Pring, EX, Olsen, KN, Mobbs, AED & Thompson, WF 2024, 'Music communicates social emotions: Evidence from 750 music excerpts', Scientific Reports, vol. 14, no. 1, pp. 27766–14.

[5] Ayata, D., Yaslan, Y., & Kamasak, M. E. (2020). Emotion recognition from EEG signals by using multivariate empirical mode decomposition. *Pattern Analysis and Applications*, 23(2), 305–319. https://doi.org/10.1007/s10044-019-00834-4

[6] Soleymani, M., Lichtenauer, J., Pun, T., & Pantic, M. (2012). A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3(1), 42–55. https://doi.org/10.1109/T-AFFC.2011.25