Please dont stop the music!

Nikita Thomas

Suhani Jain

Vedika Agarwal

Evaluating the Influence of Music on Workplace Productivity using Multi-Modal Sensory Information

Track 1: Experimental Intervention

Problem Statement

The Problem

We aim to uncover how different music genres impact productivity by analyzing physiological data, emotions, and personal preferences.

Our goal is to identify which genre of music is best for each type of task and which modality it is driven by (the WHY?).

Motivation

Existing research on music's effect on productivity often relies on **subjective self-reports** and **lacks machine learning rigor** in analyzing the relationship between music and productivity, thus unable to pin-point the contribution of physiological responses, emotional states, and preferences to task performance. This study aims to overcome these limitations by using **objective, multimodal data-driven approaches**—analyzing PPG pulse data, emotional states and task performance metrics —**across a range of tasks** such as cognitive math, typing, and memory tasks.

Impact

Studies show that music can release dopamine, aid memory, increase motivation, de-stress and more, all leading to more positive work outcomes. This study could help:

1. **Employers:** Data-driven insights into designing customized, productive work environments.
2. **Individuals:** Personalized recommendations for choosing music that optimally enhances performance in specific tasks, based on their unique physiological and emotional profiles.
3. **Interdisciplinary Research:** A new, more precise framework for studying cognitive performance, blending insights from psychology, neuroscience, and human-computer interaction

Literature Survey

1. A State-space Investigation of Impact of Music on Cognitive Performance during a Working Memory Experiment by Md. Rafiul Amin,Maryam Tahir, and Rose T. Faghih

AIM

- Investigate the impact of music (**calming vs. vexing**) on working memory performance during an n-back task.
- Analyze both binary (correct/incorrect responses) and continuous (reaction time) measures.

INFERENCES

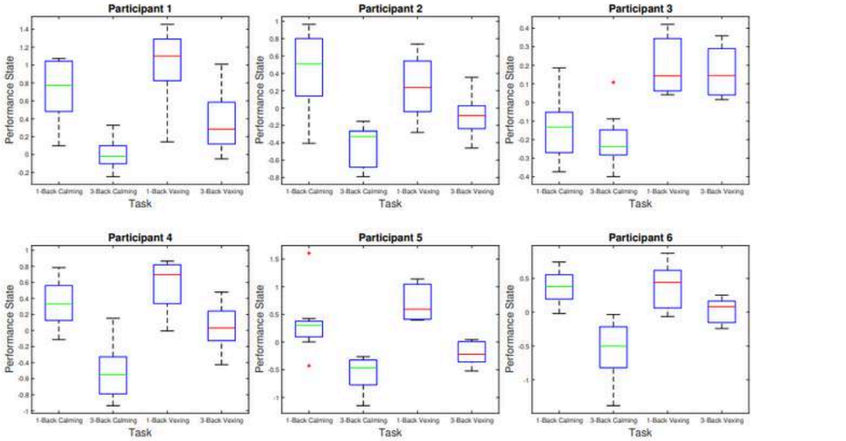


Fig. 3: **Distribution of Performance States During Different Tasks and Music.** Each sub-figure shows that the box-plot illustrating the distribution of the estimated performance states for each participant while performing two types of n-back tasks in presences calming and vexing music.

METHODOLOGY

- Participants: Six participants completed two sessions: one with calming music and another with vexing music (self-selected).
- Task: Participants performed 1-back (easier) and 3-back (harder) tasks to assess working memory.
- Data Collected:
  - Response data: Correct/incorrect responses.
  - Reaction times: Speed of responses.
  - Physiological measures: Heart rate, skin conductance, and temperature were recorded (but not analyzed here).
- High-Performance Index: HPI quantified high-performance periods.

IEEE Engineering in Medicine & Biology Society (EMBC)

2. **Enhancement of task performance aided by music** by B. Geethanjali, K. Adalarasu, M. Jagannath and R. Rajasekaran

AIM

- Examined the impact of Indian classical music and Indo jazz on task **performance** and **physiological** responses.
- Focused on how music affects reaction time, decision-making errors, pulse rate, and participants' moods and emotions.

INFERENCES

- Both kinds of music lowered pulse rates, led to better task performance, and enhanced overall mood and performance.
- They chose only 1 task and only 2 types of music.
- They analyze music impact on emotion and physical responses but fail to connect this to productivity

METHODOLOGY

Experimental Design:

- Participants were exposed to three conditions:
  - Listening to music without a task.
  - Performing a visual Go/No-go task with music.
  - Performing the task in silence.
- Physiological measures (pulse rate) and task performance (reaction time, omission, and commission errors) were recorded.
- Mood was assessed using PANAS, and emotions (valence and arousal) were measured with the Self-Assessment Manikin (SAM).
- Statistical Analysis: Non-parametric tests were used to compare pulse rate, reaction times, and errors across conditions, with significance set at $p < 0.05$.

3. **Task Appraisals, Emotions, and Performance Goal Orientation** by Cynthia D. Fisher, Amirali Minbashian, Nadin Beckmann, Robert E. Wood University of Melbourne

AIM/METHODOLOGY

The study aimed to explore how task appraisals—specifically, **task confidence** and **task importance**—affect real-time emotional fluctuations at work and how performance goal orientation (PGO) moderates these effects.

- Experience sampling methodology (ESM) was used with 135 middle managers from five multinational companies.
- Over three weeks, participants provided real-time reports on task appraisals and emotions.
- 4,172 task-related reports were analyzed using hierarchical linear modeling (HLM).

INFERENCES

- Higher task confidence was significantly associated with more positive emotions and fewer negative emotions.
- Higher task importance was positively related to both **positive and negative emotions**.

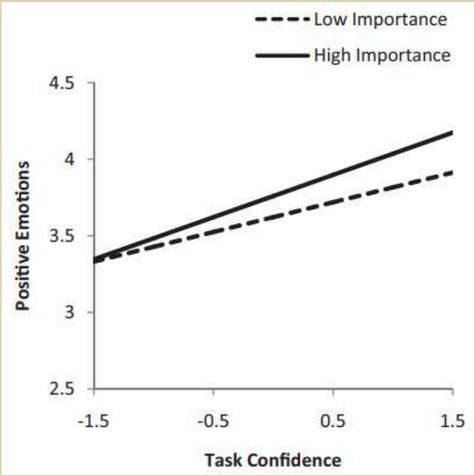


Figure 1. Level 1 interaction between task importance and confidence predicting positive emotions.

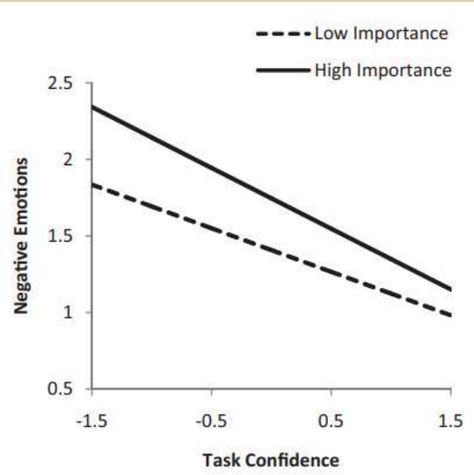


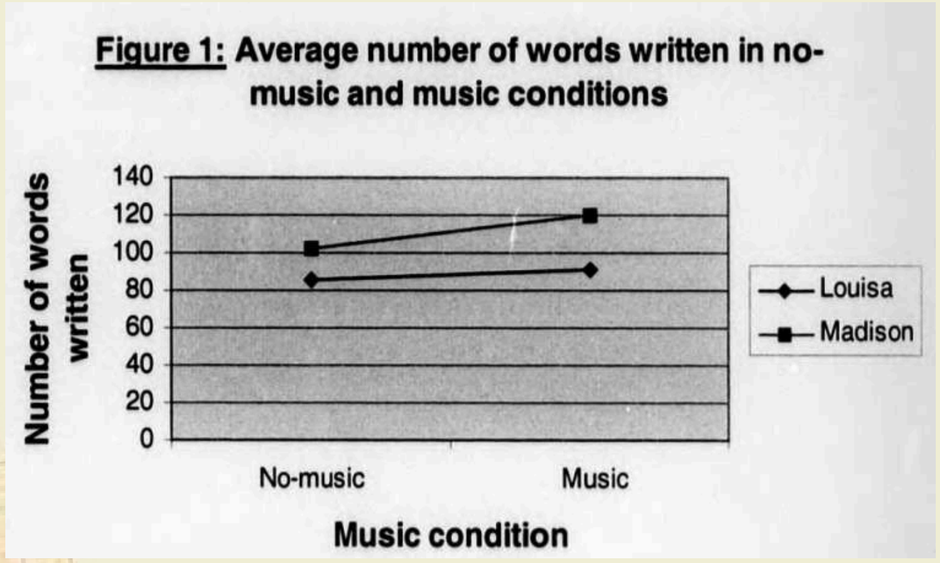
Figure 2. Level 1 interaction between task importance and confidence predicting negative emotions.

Journal of Applied Psychology

4. **The Effect of Student Music Choice on Writing Productivity** by Renee Donohoe And Teresa McNeely

AIM

- Aimed to investigate the impact of **student-selected** background music on writing productivity in elementary school students.
- Specifically understanding whether allowing students to choose their own music would increase the number of words written during writing sessions.



METHODOLOGY

Participants:

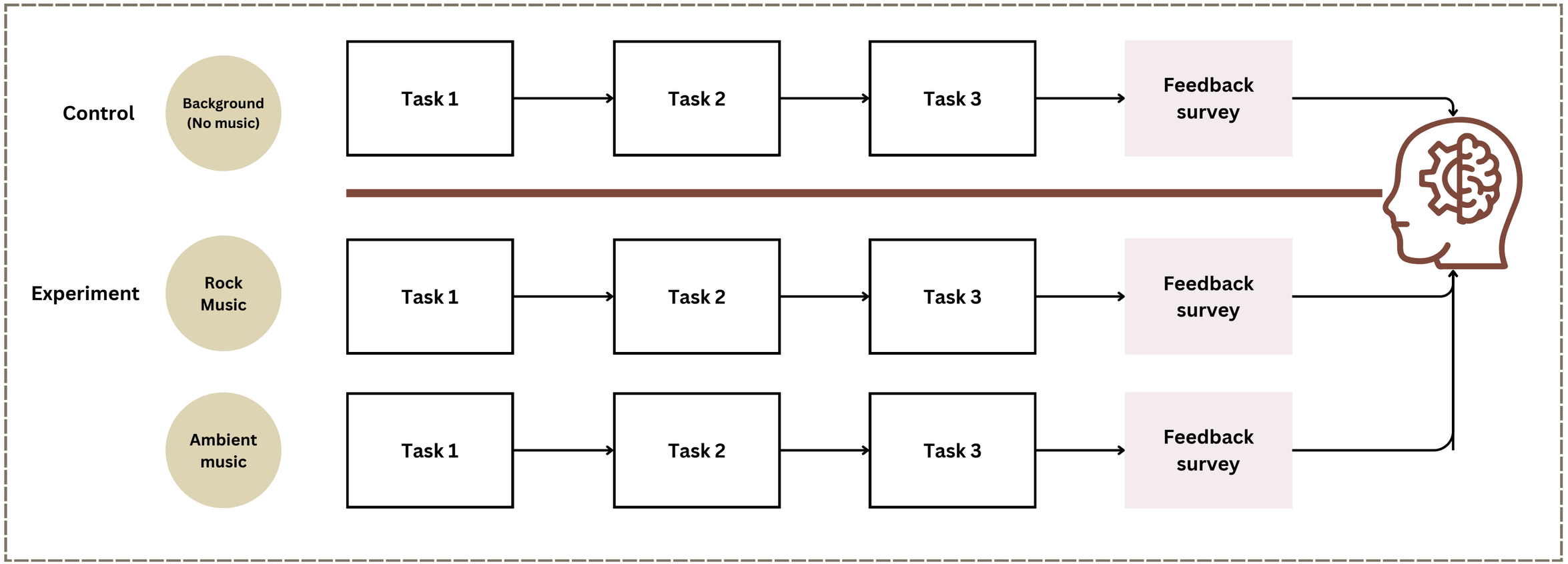
- Two fourth-grade classes, one from Madison County and one from Louisa County.
- Madison County had no prior exposure to music in the classroom, while Louisa County regularly incorporated it.

Design:

- A quasi-experimental, repeated measures design was used, with students completing four writing sessions: one baseline session without music and three with music, each lasting 15 minutes.
- A multi-section survey assessed students' music preferences, and a mixtape of songs from the top two genres was created for the sessions.

Experiment Protocol

Experiment Protocol



Recorded data from 17 participants

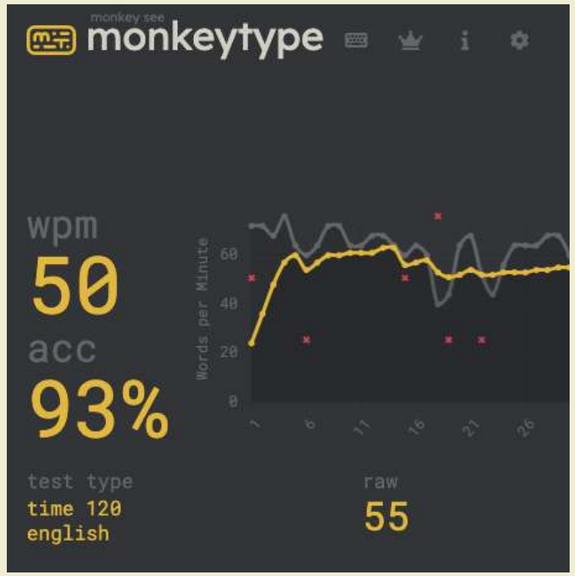
Tasks: Each Participant was asked to perform 3 tasks in a randomised order while listening to audio:
The user was asked to fill a short survey after each song

Stimuli: Each participant listened to an audio in the following categories. The order will be randomised to account for stimuli presentation

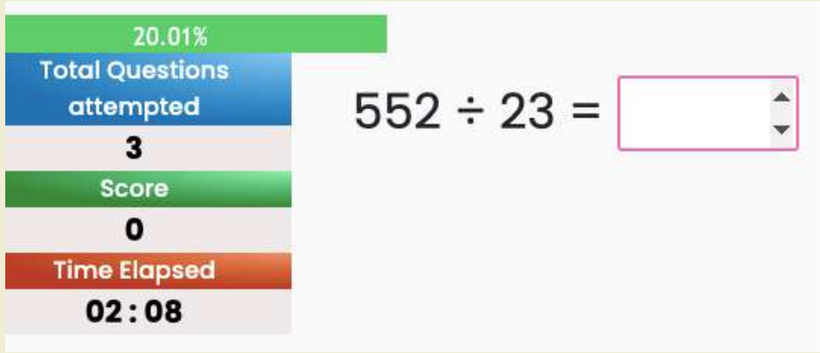
- **Background Noise** (Control)
- **Ambient Productive Music** (Instrumental)
- **Agitating Rock Music**

The audio will be played at the same volume for everyone in earphones. At the end of the experiment each participant will be asked to rate their preference

Experiment



Time: 2 minutes
Metrics:
WordsPerMinute
Accuracy



15 questions (Mixed)
Metrics:
Time Taken
Accuracy

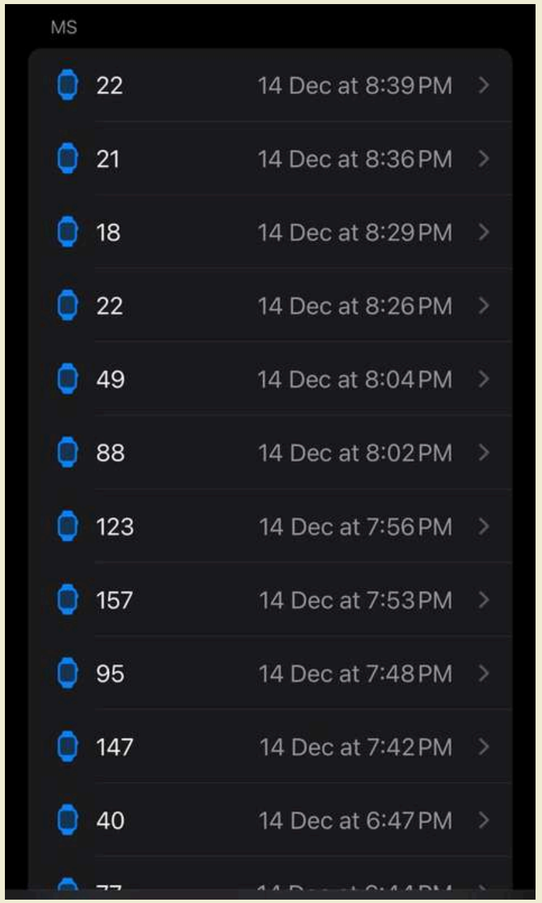
Time: 2 Minutes
Metrics:
Recall Accuracy
Average Response time

Data Collection

Our data collected consists of :

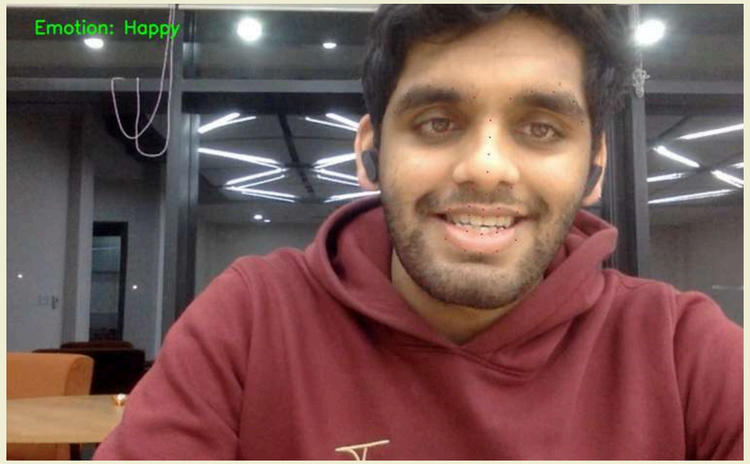
PPG Data	During the experiment. Recorded using **Apple watch health data**
Feedback Survey	Participants' subjective feedback on music preference and perceived productivity
AU scores	AU (Action Unit) analysis via facial expression recognition from **frames extracted from laptop camera** of participant performing task to extract emotion data.
Task performance	Task performance will be evaluated based on accuracy and time taken to complete task

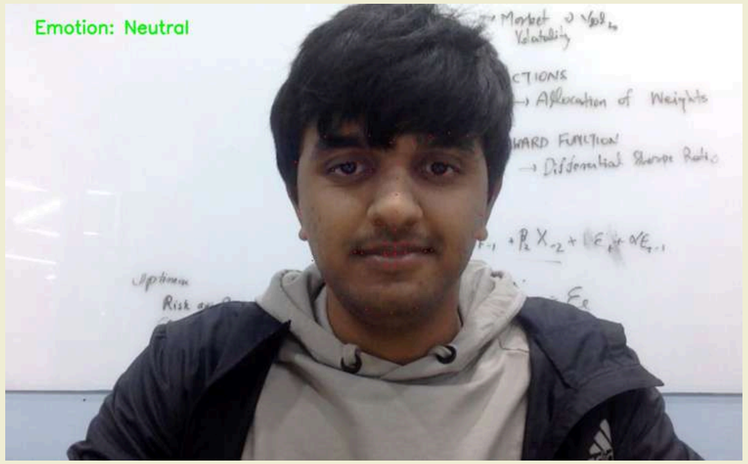
HRV



Collected using Breathe sessions on Apple Watch

Emotion Recognition



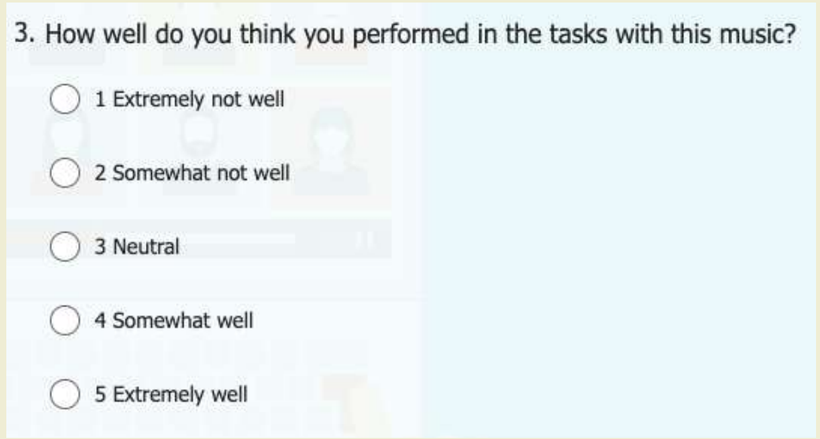


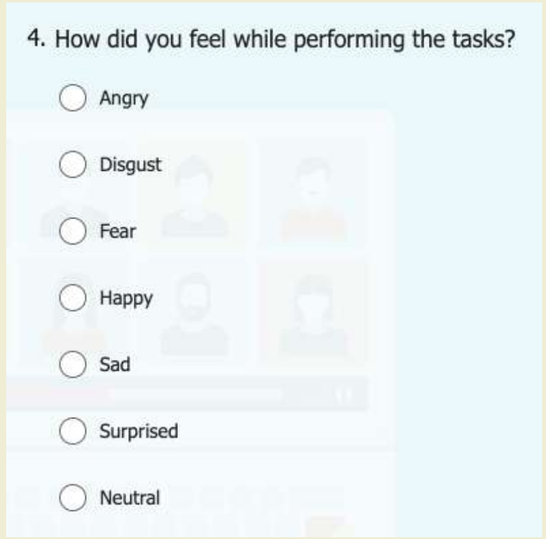
Took 2-3 Pictures of participant performing task randomly and classified emotions in the following categories using OpenFace FACS

- sad
- anger
- fear
- disgust
- happy
- surprise
- neutral

Survey

After each song, rating perceived productivity across all 3 tasks





- Rate Music Preference
- Report emotion while performing the task (Ground truth)
- Rate your performance/productivity while listening to music (Scale 1-5)

Dataset

Typing

WPM	Accuracy	FACS	Music	Perceived Performance	Perceived Emotion	Music Rating
63	95	Neutral	No Music	4	Neutral	3
60	93	Happy	Ambient	4	Happy	5
66	94	Happy	Rock	5	Happy	5

Memory

Recall Speed	Accuracy	HRV	FACS	Music	Perceived Performance	Perceived Emotion	Music Rating
1.01	80	85	Neutral	No Music	4	Neutral	3
1.01	76	88	Happy	Ambient	4	Happy	5
0.98	84	88	Neutral	Rock	5	Happy	5

Math

Time Taken(Seconds)	Accuracy	HRV	FACS	Music	Perceived Performance	Perceived Emotion	Music Rating
97	73	63	Neutral	No Music	4	Neutral	4
61	100	75	Happy	Ambient	4	Happy	5
129	86	74	Happy	Rock	5	Happy	5

A total of 156 rows of data

EDA

Mean and Std

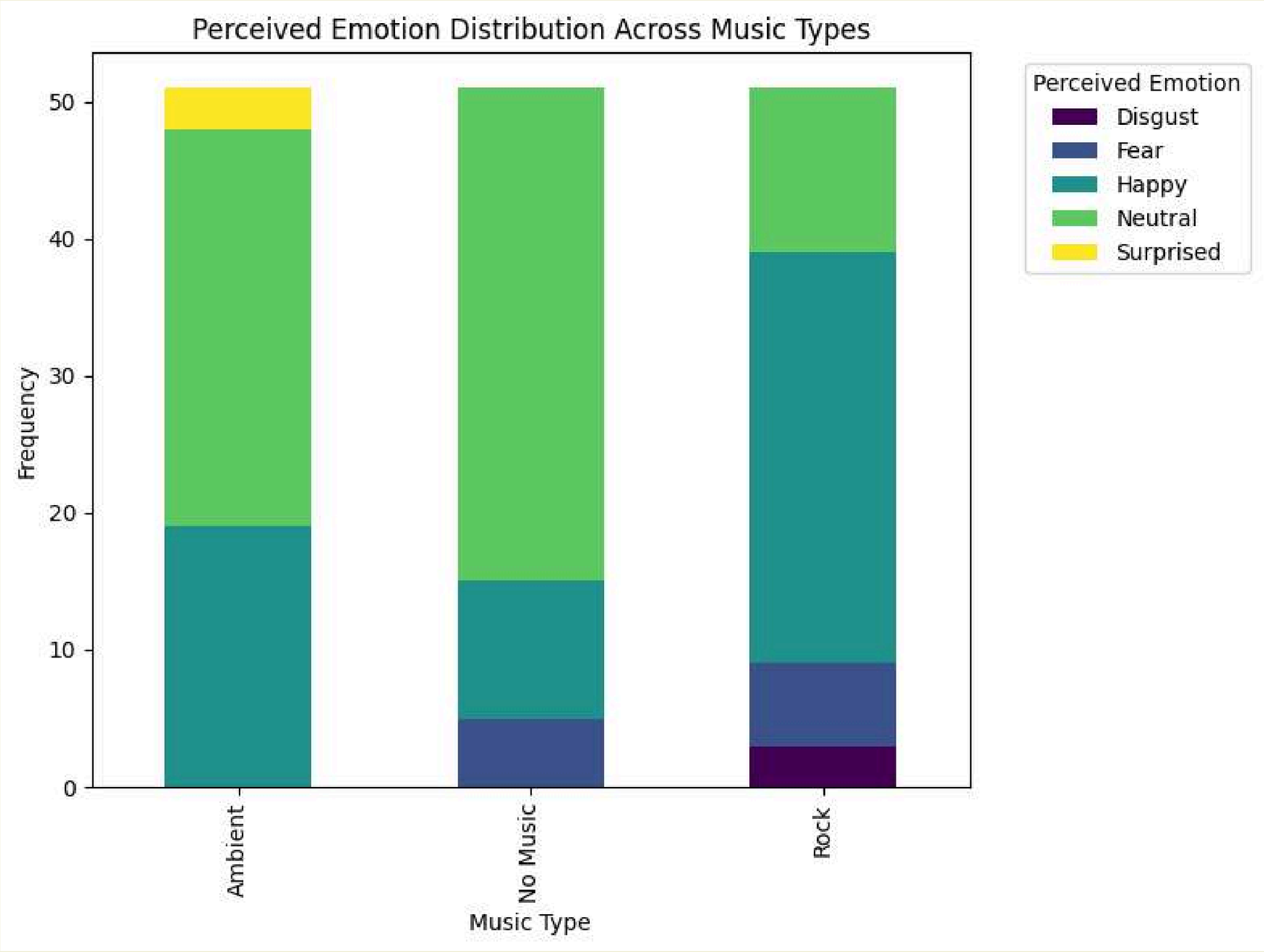
Recall Speed (Memory Task Only)

Music Condition	Mean Recall Speed	SD Recall Speed	Key Insights
Ambient	0.8911	0.2734	Recall speed improves slightly under Ambient but is moderately variable.
No Music	0.9215	0.2756	Recall speed improves further, showing similar variability as Ambient.
Rock	0.9351	0.3074	Highest recall speed, but with increased variability.

Task	Music Condition	Mean Accuracy	SD Accuracy	Key Insights
Math	Ambient	86.00	11.30	Highest accuracy for Math tasks under Ambient with moderate variability.
	No Music	85.88	10.99	Slightly lower accuracy with similar variability.
	Rock	82.88	13.06	Lowest accuracy for Math, with higher variability.
Memory	Ambient	84.06	10.59	High accuracy with moderate variability, similar to No Music condition.
	No Music	84.24	8.06	Comparable to Ambient, with lower variability.
	Rock	81.24	15.30	Slightly lower accuracy with much higher variability, indicating inconsistency.
Typing	Ambient	94.88	1.83	Typing accuracy remains consistently high under Ambient with very low variability.
	No Music	95.00	3.43	Highest typing accuracy under No Music, but with slightly more variability.
	Rock	94.11	3.99	Typing accuracy remains high under Rock, with slightly higher variability compared to others.

HRV (Math and Memory Tasks)

Task	Music Condition	Mean HRV	SD HRV	Key Insights
Math	Ambient	52.41	29.18	Highest HRV for Math under Ambient, with high variability indicating mixed relaxation levels.
	No Music	45.65	18.08	Lower HRV, suggesting less relaxation under No Music.
	Rock	50.35	32.89	Comparable HRV to Ambient, but with the highest variability across participants.
Memory	Ambient	48.12	22.99	Moderate HRV for Memory under Ambient, with slightly lower variability.
	No Music	45.59	20.47	Slightly lower HRV compared to Ambient, indicating less relaxation.
	Rock	44.24	25.90	Lowest HRV for Memory under Rock, with highest variability.


Perceived Emotion Distribution Across Music Types
Frequency
0
10
20
30
40
50
Ambient
No Music
Rock
Music Type
Perceived Emotion
Disgust
Fear
Happy
Neutral
Surprised

WPM (Typing Task Only)

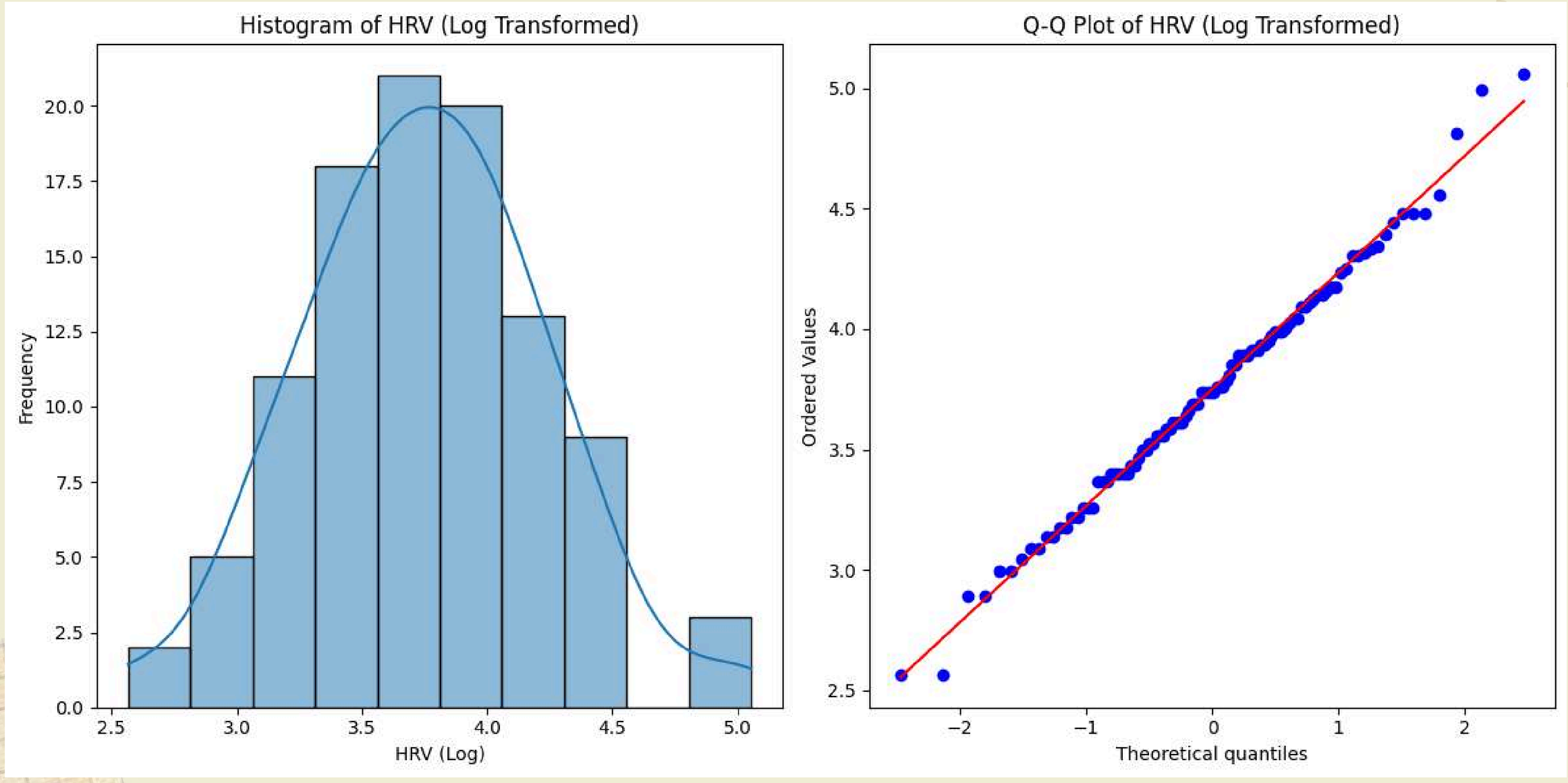
Music Condition	Mean WPM	SD WPM	Key Insights
Ambient	61.18	14.18	Typing speed is highest under Ambient with moderate variability.
No Music	58.18	14.54	Typing speed slightly decreases under No Music, but variability remains similar.
Rock	56.53	17.11	Lowest typing speed under Rock, with the highest variability across participants.

Time Taken (Math Task Only)

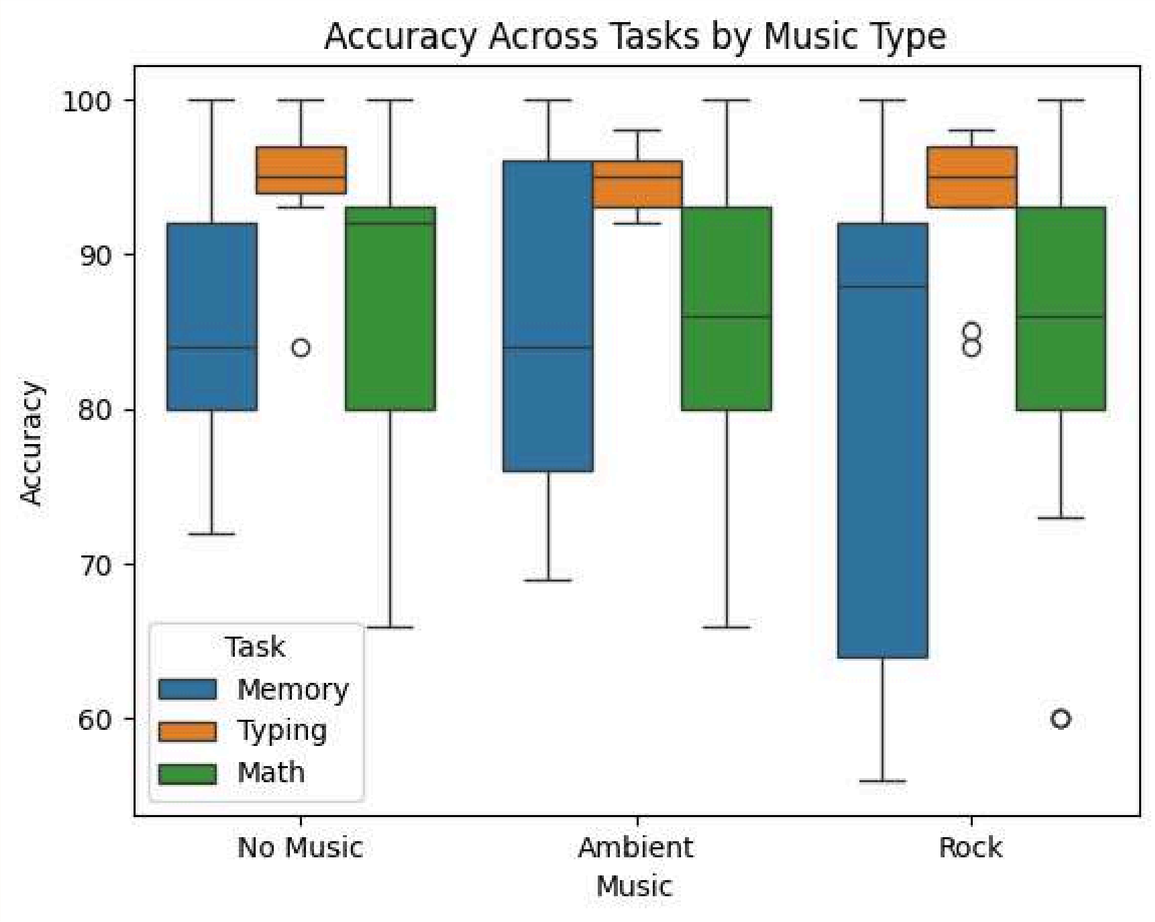
Music Condition	Mean Time Taken (Seconds)	SD Time Taken	Key Insights
Ambient	137.29	61.75	Participants take the least time under Ambient, though variability is moderate.
No Music	138.82	57.34	Slightly higher time than Ambient, but with slightly lower variability.
Rock	141.59	71.50	Participants take the longest time under Rock, with the highest variability.

Table 2: Normality (After Transformation)

Metric	Music Condition	Transformation	Shapiro-Wilk p-value	Normality
Accuracy	No Music	Box-Cox	0.00721	Not Normal
	Ambient	Box-Cox	0.00130	Not Normal
	Rock	Box-Cox	0.00168	Not Normal
Music Rating	No Music	Box-Cox	3.1e-11	Not Normal
	Ambient	Box-Cox	2.8e-06	Not Normal
	Rock	Box-Cox	1.9e-07	Not Normal
HRV	No Music	Log	0.30404	Normal
	Ambient	Log	0.88269	Normal
	Rock	Log	0.60793	Normal


Histogram of HRV (Log Transformed)
Frequency
HRV (Log)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
2.5
3.0
3.5
4.0
4.5
5.0
Q-Q Plot of HRV (Log Transformed)
Ordered Values
Theoretical quantiles
2.5
3.0
3.5
4.0
4.5
5.0
−2
−1
0
1
2

Visualisation



Memory Task

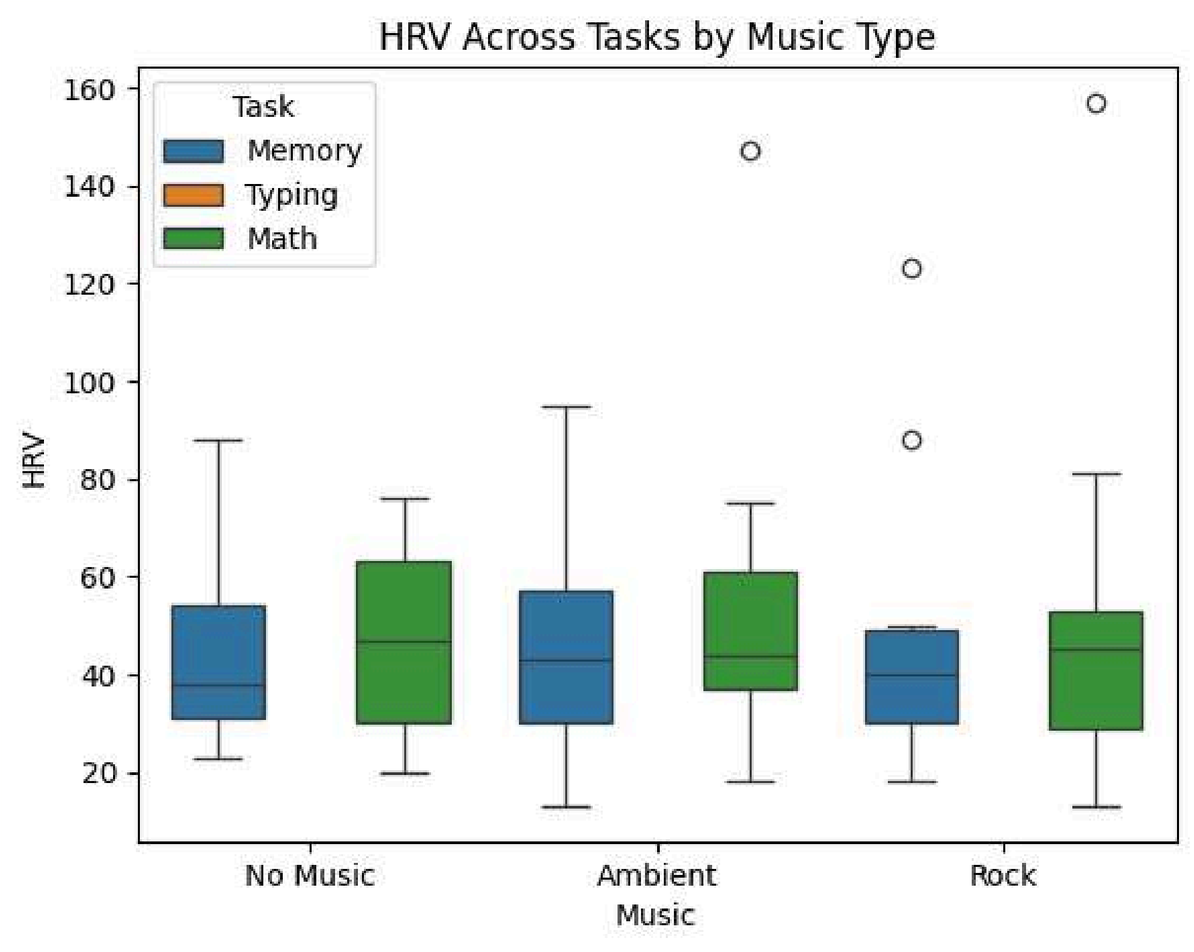
- Rock music negatively impacts performance, showing the largest variability and lower accuracy.
- Ambient music shows slightly better results, similar to No Music.

Math Task

- Accuracy is highest with No Music.
- Both Ambient and Rock music show a slight drop, with Rock introducing more variability.

Typing task is least affected

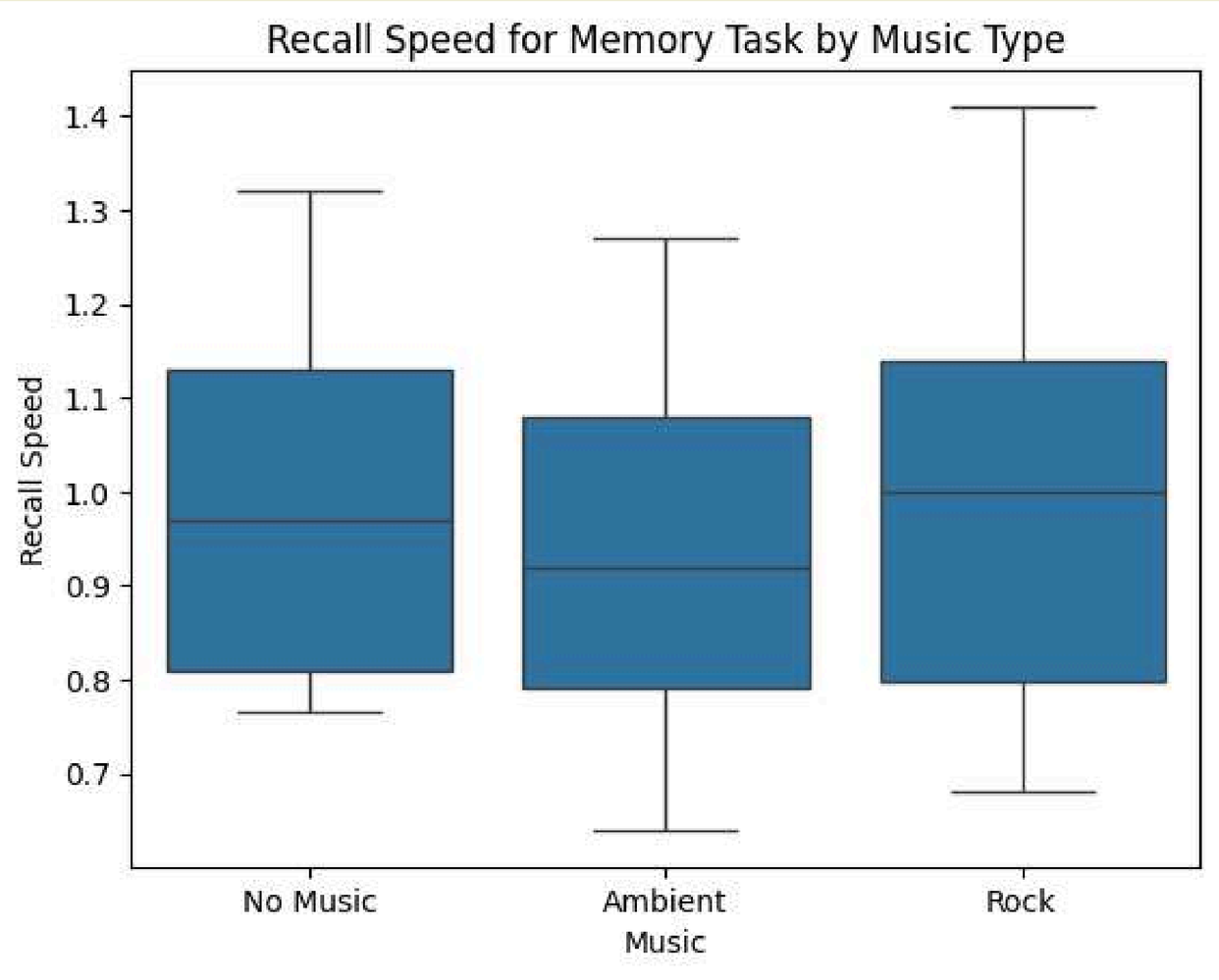
Visualisation



Heart Rate Variability (HRV)

- Rock Music introduces higher variability (outliers appear), especially for Memory task.
- indicating it may induce more physiological stress or arousal.
- HRV impacted Memory tasks more compared to Math task

Visualisation



Recall Speed for Memory Task

- No Music and Ambient Music show similar recall speeds, with medians around 0.8–0.9.
- Rock Music has a slightly higher median (~1.0) showing that the task took longer time for completion, but it also shows a wider spread of values.

Outliers

- All music types show low recall speed outliers close to 0, indicating occasional poor performance.

General Trend

- Rock Music slightly improves median

Checking Assumptions

Table 1: Normality and Homogeneity (Before Transformation)

Metric	Music Condition	Shapiro-Wilk p-value	Levene's Test p-value	Normality	Homogeneity
Accuracy	No Music	0.00031	0.41346	Not Normal	Homogeneous
	Ambient	0.00015		Not Normal	
	Rock	0.000002		Not Normal	
HRV	No Music	0.01403	0.77217	Not Normal	Homogeneous
	Ambient	0.00169		Not Normal	
	Rock	0.000012		Not Normal	
Recall Speed	No Music	0.01406	0.85898	Not Normal	Homogeneous
	Ambient	0.05821		Normal	
	Rock	0.20141		Normal	
Music Rating	No Music	2.6e-11	1.7e-07	Not Normal	Not Homogeneous
	Ambient	3.5e-06		Not Normal	
	Rock	2.4e-07		Not Normal	
WPM	All Conditions	> 0.48	0.71352	Normal	Homogeneous
Time Taken	Rock	0.01638	0.69723	Not Normal	Homogeneous

Checking Assumptions

Table 1: Normality and Homogeneity (Before Transformation)

Metric	Music Condition	Shapiro-Wilk p-value	Levene's Test p-value	Normality	Homogeneity
Accuracy	No Music	0.00031	0.41346	Not Normal	Homogeneous
	Ambient	0.00015		Not Normal	
	Rock	0.000002		Not Normal	
HRV	No Music	0.01403	0.77217	Not Normal	Homogeneous
	Ambient	0.00169		Not Normal	
	Rock	0.000012		Not Normal	
Recall Speed	No Music	0.01406	0.85898	Not Normal	Homogeneous
	Ambient	0.05821		Normal	
	Rock	0.20141		Normal	
Music Rating	No Music	2.6e-11	1.7e-07	Not Normal	Not Homogeneous
	Ambient	3.5e-06		Not Normal	
	Rock	2.4e-07		Not Normal	
WPM	All Conditions	> 0.48	0.71352	Normal	Homogeneous
Time Taken	Rock	0.01638	0.69723	Not Normal	Homogeneous

ANOVA

Metric	Effect	p-value (Music)	p-value (Task)	p-value (Music:Task Interaction)
Accuracy (Box-Cox)	No effect of Music, Task has significant effect	0.6961	0.0000	0.9659
HRV (Log Transformed)	No effect of Music or Task	0.5061	0.2698	0.5788
Music Rating (Box-Cox)	Music affects rating	0.0287	0.6198	0.5400

Chi-squared test

Test	Chi-Square Value (Chi^2)	p-value	Interpretation
Chi-Square Test (Mood & Music)	39.71	3.63e-06	Significant relationship between mood and music type

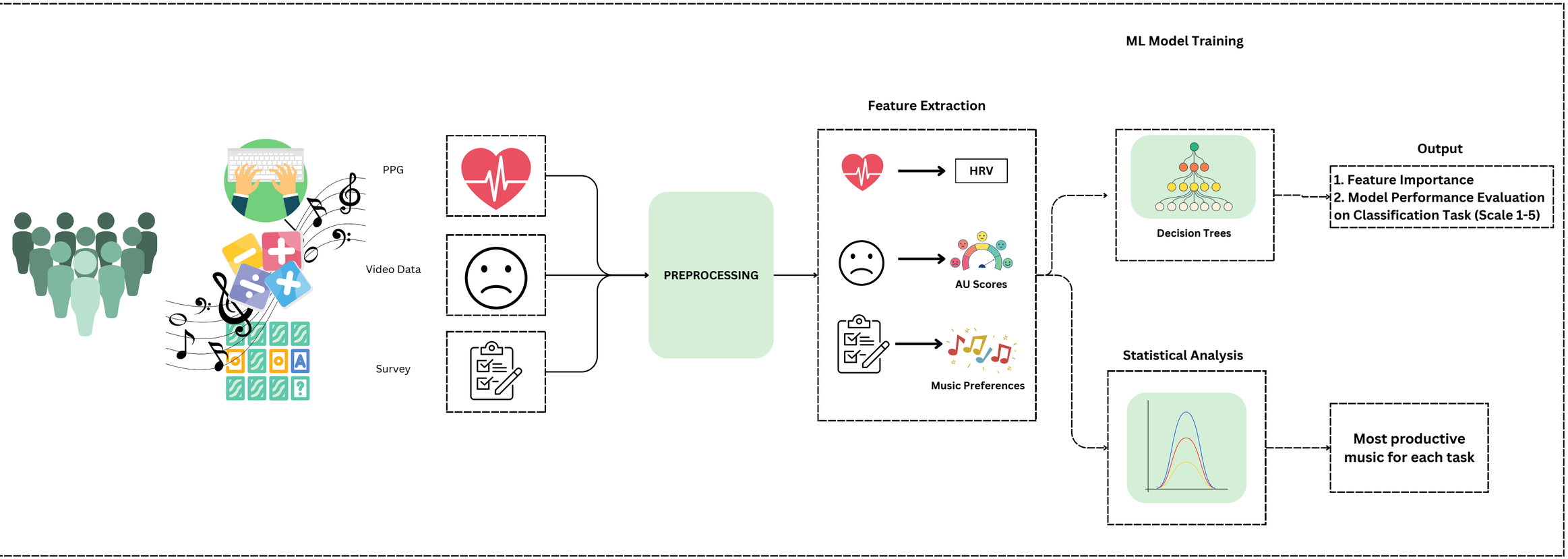
Interpretation

Why Results are Not Significant

1. **Variability Within Groups:**
   - High within-group variability can obscure differences between groups. For example, if participants perform differently under the same music condition (e.g., some perform much better, others much worse), the large variability may prevent the ANOVA from detecting significant effects.
2. **Small Sample Size:**
   - A small sample size reduces the power of the test, making it harder to detect significant differences even if they exist. With fewer data points, the chance of detecting a true effect decreases.
3. **Effect Size vs. Statistical Significance:**
   - While large differences in averages may suggest a meaningful effect, statistical significance requires more confidence that the differences are not due to chance. High variability or small sample size can obscure significant results.
4. **Data Distribution:**
   - ANOVA assumes that data within each group are normally distributed. Skewed data or outliers can distort the results, making it harder for the test to detect true differences.

ML

OVERVIEW


ML Model Training
Feature Extraction
PPG
Video Data
Survey
PREPROCESSING
HRV
AU Scores
Music Preferences
Decision Trees
Output
1. Feature Importance
2. Model Performance Evaluation on Classification Task (Scale 1-5)
Statistical Analysis
Most productive music for each task

Metrics

Memory Task	**Accuracy - k*Recall Time**
Math Task	**Accuracy/Time Taken**
Typing Task	**Accuracy * WPM**

ML Models

Task-Wise Impact of Music on Performance to assess Productivity

Random Forest Classifier

Linear Regression

Random Forest Regressor

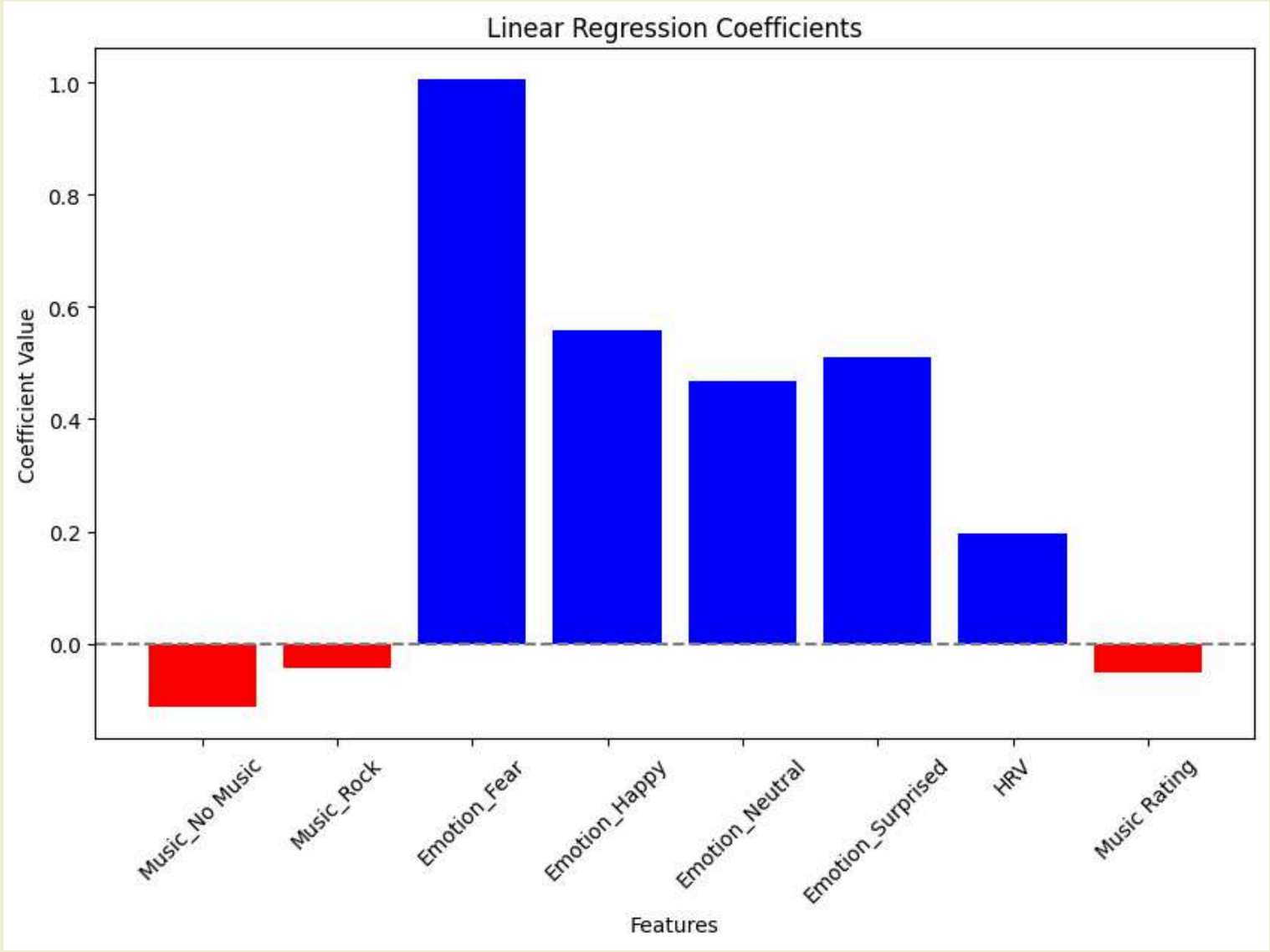
Gradient Boosting

SMOTE for Class Imbalance

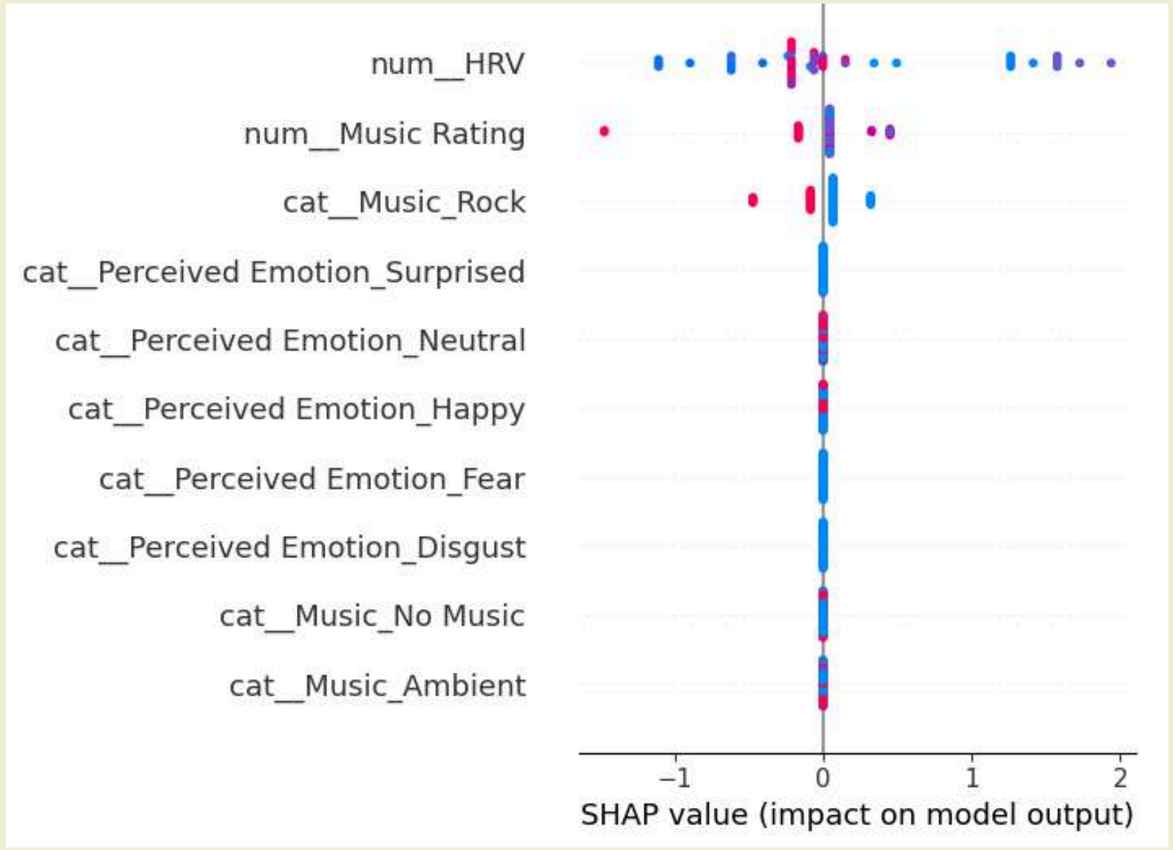
Results

Linear Regressor for Math Task

```
Mean Squared Error: 0.1084
R2 Score: -0.026979278804
```

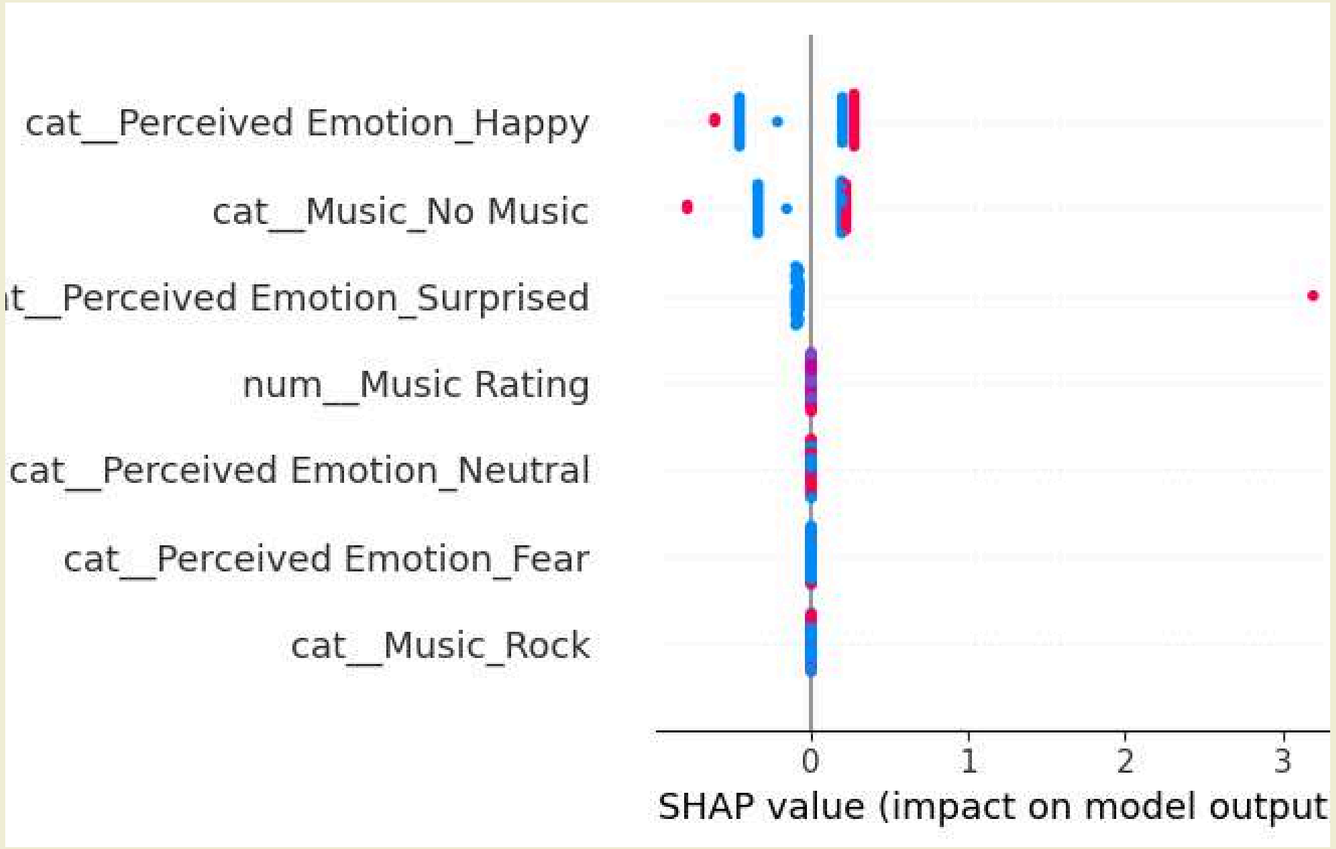


Gradient Boosting Regressor for Memory Task




```
Mean Squared Error: 143.839866
R2 Score: -0.23461311973634014
```

Gradient Boosting Regressor for Typing Task




```
Mean Squared Error: 472.562268
R2 Score: -0.7432191651495996
```

Impact and Deployability

Impact

Task-Wise Impact of Music on Performance to assess Productivity

A comprehensive analysis assessing the impact of music genre on performance across various productivity tasks such as **typing, cognitive tasks, and memory**. This analysis examines which features contribute most to performance changes, such as:

- **Emotional responses** evoked by music (captured through facial action units)
- **Physiological changes** like heart rate variations (HRV data)
- **Individual preferences** for music genres

This analysis can help inform which music genre impacts each task but also **why.** Individuals or employers can thus derive more nuanced recommendations to boost productivity based on their work settings.

Challenges

Data Quality	Ensuring clean, accurate data from the Apple Watch PPG sensors and facial emotion recognition (face stays in frame with optimal lighting). Also, PPG data captures internal emotional arousal which may not directly correspond with emotion data.
Self-Reporting Bias	Participants' subjective feedback on music preference and perceived productivity did not always correspond to actuals.
Data Samples	Collection of sufficient number of samples to derive reliable results.
Generalization	Results from controlled tasks may not generalize well to real-world environments due to limited participant data, resulting in high individual variability.

Deployment

Data Collection

Collect more data from diverse participants to ensure higher accuracies in ML Model and statistical significance in hypothesis testing

Work Settings

Based on tasks required for the work setting, choose and suggest music types to employees. Issues could include data collection for multimodal sensors.

Real Time Data

Synchronization suggestions for real time personalized feedback on music playlists to optimize tasks performance while collecting multimodal data

Thank You!