Evaluation Methodology

Our experiment involves participants taking part in strength training exercises under three different music conditions: no music, music synchronized with their movements, and music out of sync with their movements. The goal is to assess how these conditions affect their motivation, enjoyment, exertion levels, and in-game performance.

We will be using Life Fitness G7 Cable Machine. Each participant is asked to perform 8 repetitions of double-arm biceps curls for 2 sets, with a 2-minute rest in between. The participants are allowed to choose their own weight based on their fitness levels. However the chosen weight must be within range (male 8-12kg, female 4-6kg). Participants will be performing under all three conditions with the same weight, and with sufficient rest between sessions to mitigate fatigue effects.

Dependent and Independent Variables

The independent variable is the condition of music while the participant is playing the game. The conditions are:

1. Sync: The rhythm is synchronised to the movement (4 beats up, 4 beats down).
2. Non-Sync: The rhythm is slightly ahead such that the movement is on the third beat.
3. No Music: No music is played during the game.

We have confined our game to only have 1 sound track to mitigate the effect of user preferences as much as possible. Our chosen sound track is *"DJ Got us Fallin in Love"* by Usher.

Experiment Design

We chose to use within-subject design for this experiment so each participant will experience all three conditions. To control for potential order effects, we will use Latin Square rotation, systematically varying the sequence in which participants encounter each condition. An example Latin Square for 6 participants is shown below, the numbers represent the order of conditions (the first participant will do Sync first, Non-Sync second, and No Music last).

Participant ID	Sync	Non-Sync	No Music
1	1	2	3
2	3	1	2
3	2	3	1
4	1	2	3
5	3	1	2
6	2	3	1

Pre and Post Questionnaires

Each participant will be asked to fill out a pre-questionnaire before the experiment and a post-questionnaire for every experiment condition, in total of three. The pre-questionnaire consists of the Physical Activity Readiness Questionnaire (PARQ), which assesses participants' ability to perform strength exercises, as well as a form collecting demographic and fitness background information (found under "*RawData/PARQ Forms"* and "*RawData/Pre Study"*). The post-questionnaire is adapted from the Intrinsic Motivation Inventory (IMI) and includes items measuring motivation, enjoyment, effort, and rate of perceived exertion (RPE), which can be found under *"RawData/Post Study"*.

In-Game Performance

During each gameplay session, a range of in-game performance metrics are logged to capture participants' interactions and outcomes, including time stamps, target and player positions (*targetY*, *playerY*), the status of the fishing dot (*dotStatus*), cumulative score, number of fish caught, whether beat is synchronised (*beatOffset*), and the weight used during the session. These data will be saved as CSV files for subsequent analysis, under the *"RawData/Game Logs"* directory.

Data Analysis

All data will be processed by the *Scripts/data_processing.ipynb* script. Processed data will be stored under the *ProcessedData* directory, under respective subfolders for each condition. Except for the pre-questionnaire data, since it is independent of conditions.

Because our questionnaires use Likert scales, which produce ordinal and non-parametric data, and our experiment uses a within-subjects design, we use the Friedman test to identify overall differences in motivation, enjoyment, exertion, and performance across the three music conditions. If the Friedman test indicates a significant effect, we then apply the Wilcoxon signed-rank test for post-hoc pairwise comparisons between conditions. The Friedman test allows us to determine whether any of the music conditions lead to statistically significant changes in participants' responses, while the Wilcoxon test helps pinpoint which specific pairs of conditions differ. Together, these tests enable us to assess both the presence and the source of any effects of music on participant outcomes. All tables and figures shown below are generated by the *"Scripts/data_analysis.ipynb"* script.

So far, we have 4 participants, all of them completed all 3 conditions with the order shown in the Latin Square. Their processed data is shown below:

Overall:

	player	condition	score	enjoyment	exertion	effort
0	charles	No Music	6312	11.0	9.0	7.0
1	eason	No Music	5360	16.0	12.0	13.0
2	gallon	No Music	6368	17.0	17.0	12.0
3	joshua	No Music	3576	-8.0	11.0	12.0
4	charles	Non Sync	6312	11.0	9.0	7.0
5	eason	Non Sync	4296	20.0	12.0	9.0
6	gallon	Non Sync	5376	13.0	17.0	12.0
7	joshua	Non Sync	3000	-4.0	13.0	14.0
8	charles	Sync	6464	11.0	10.0	7.0
9	eason	Sync	4472	20.0	15.0	9.0
10	gallon	Sync	6200	15.0	17.0	12.0
11	joshua	Sync	5272	-7.0	13.0	14.0

We separate them into different tables for better visualization:

Enjoyment Table:

condition	No Music	Non Sync	Sync
player			
charles	11.0	11.0	11.0
eason	16.0	20.0	20.0
gallon	17.0	13.0	15.0
joshua	-8.0	-4.0	-7.0

Exertion Table:

condition	No Music	Non Sync	Sync
player			
charles	9.0	9.0	10.0
eason	12.0	12.0	15.0
gallon	17.0	17.0	17.0
joshua	11.0	13.0	13.0

Score Table:

condition	No Music	Non Sync	Sync
player			
charles	6312	6312	6464
eason	5360	4296	4472
gallon	6368	5376	6200
joshua	3576	3000	5272

Effort Table:

condition	No Music	Non Sync	Sync
player			
charles	7.0	7.0	7.0
eason	13.0	9.0	9.0
gallon	12.0	12.0	12.0
joshua	12.0	14.0	14.0

Friedman Test

Description

The Friedman test is applied to compare motivation and performance scores across the three music conditions: no music, music, and music out of sync with movement. This test does not assume normality and is good for analysing non-parametric questionnaire and discrete data.

Null Hypothesis (H_0)

There are no differences in the distributions of motivation or performance scores across the three music conditions. Music does not affect these outcomes.

Alternative Hypothesis (H_1)

At least one music condition leads to different motivation or performance scores compared to the others.

Results

```
Friedman Test:
    Enjoyment: statistic=0.545, p-value=0.7613
    Exertion: statistic=4.667, p-value=0.0970
    Score: statistic=4.933, p-value=0.0849
    Effort: statistic=0.000, p-value=1.0000
```

The results of the Friedman test indicate that there are no statistically significant differences in enjoyment, exertion, score, or effort across the three music conditions, as all p-values are greater than the conventional threshold of 0.05. This means we fail to reject the null hypothesis. Thus, we say that the type of music played during the strength training sessions did not lead to measurable changes in participants' motivation or performance.

Wilcoxon Signed-Rank Test

Description

We use Wilcoxon Signed-Rank Test for post-hoc pairwise comparisons between music conditions. Although the Friedman test currently indicates no significant effect, our small sample size of 4 may limit the test. Analysis code is still done for Wilcoxon test for the future when we have more participants. We are aiming for about 15-20 participants for our user study.

Null Hypothesis (H_0)

There is no difference in motivation or performance scores between the two music conditions being compared.

Alternative Hypothesis (H_1)

There is a difference in motivation or performance scores between the two music conditions being compared.

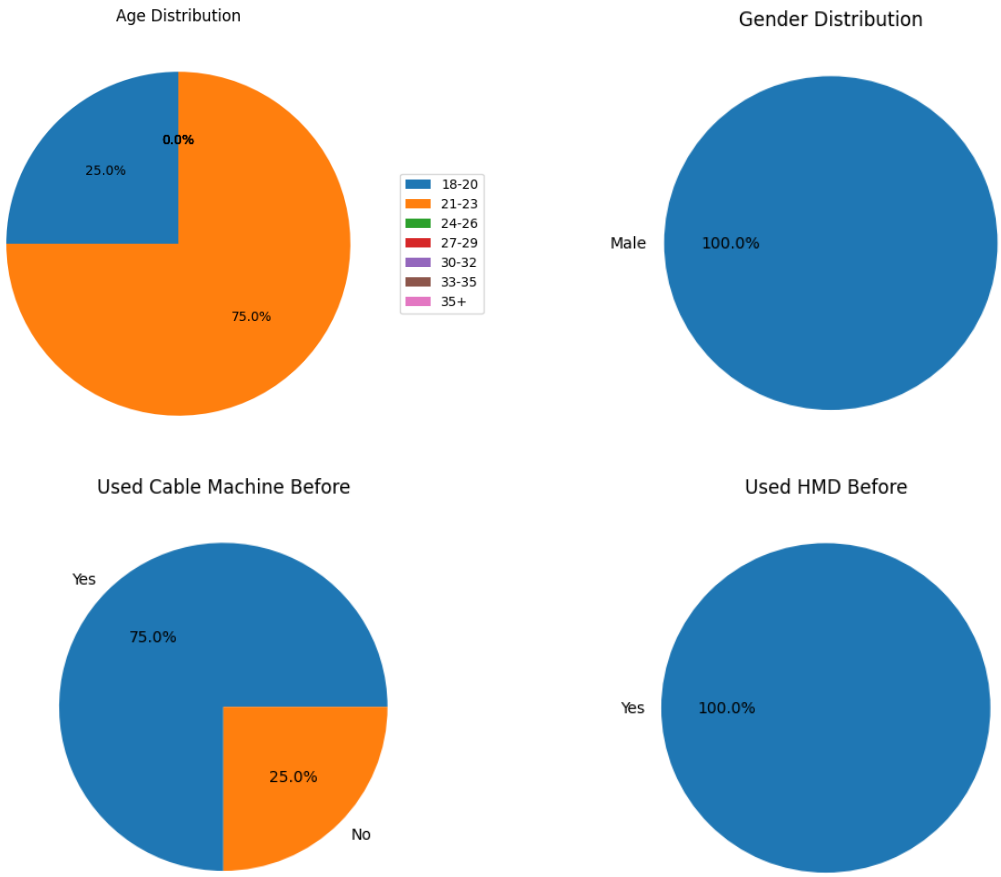
Results

```
Wilcoxon Signed-rank Test:
    (No Music vs Sync)
    Enjoyment: statistic=2.000, p-value=0.5930
    Exertion: statistic=0.000, p-value=0.1088
    Score: statistic=5.000, p-value=1.0000
    Effort: statistic=1.000, p-value=0.6547

    (No Music vs Non Sync)
    Enjoyment: statistic=2.000, p-value=0.5637
    Exertion: statistic=0.000, p-value=0.3173
    Score: statistic=0.000, p-value=0.1088
    Effort: statistic=1.000, p-value=0.6547

    (Non Sync vs Sync)
    Enjoyment: statistic=1.000, p-value=0.6547
    Exertion: statistic=0.000, p-value=0.1797
    Score: statistic=0.000, p-value=0.1250
    Wilcoxon test not applicable for Effort
```

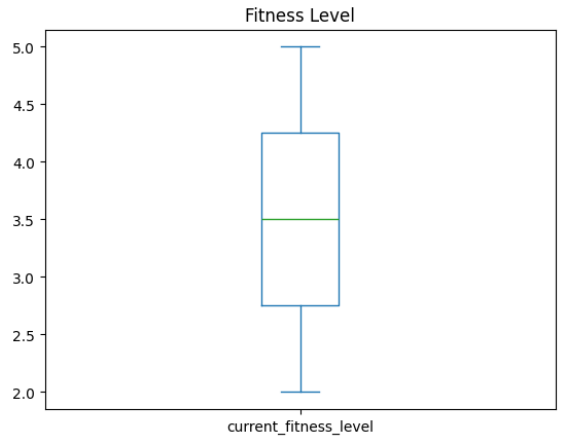
The Wilcoxon Signed-Rank Test results currently show that none of the pairwise comparisons between music conditions yielded statistically significant differences in enjoyment, exertion, score, or effort, as all p-values are above 0.05. This means we fail to reject the null hypothesis for each comparison, indicating that, based on the current data, there is no evidence that any specific music condition leads to different motivation or performance outcomes compared to the others. These findings are consistent with the Friedman test results.

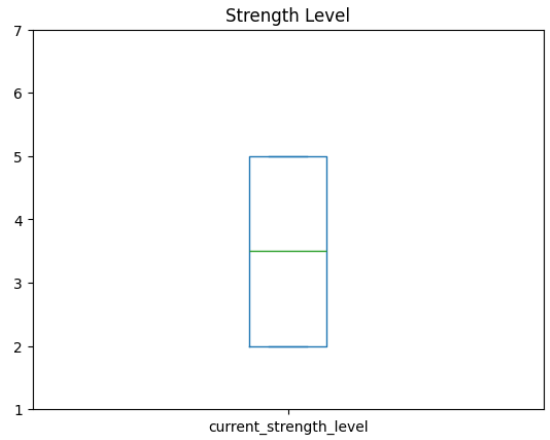
Demographics

The *"Scripts/data_analysis.ipynb"* script also plots for several demographics data based on our pre-questionnaire. Our 4 participants are mostly age between 21-23, all of which are male, all have used head-mounted display (HMD) before and all except for one have used cable machine before.



The fitness and strength levels of our participants is shown through a box plot (scale 1-7). Most people report strength levels to be between 2 and 5, with the median being 3.5. There is moderate spread but no extreme low/high values. The strength plot shows similar median but more people report themselves to be at either extreme.





From the post-questionnaire, the Sync group shows significantly higher exertion compared to No Music and Non-Sync, the enjoyment of the three groups is similar, the score of Non-Sync group is the lowest, and the groups with music show significantly lower effort compared to the group without music. However, a larger sample size is needed for more conclusive results.