Investigating Affective States with Music

Raakhee Sadhnani[1], Vijaya Laxmi Tripathi[2]
School of Computing
National University of Singapore
Singapore
{rsadhnani[1], vijaya.t[2]}@u.nus.edu

***Abstract*—Topics on the affective content of music can be observed through the lenses of a multi-modal concept that incorporates various facets such as audio, video, choreography and motion. We believes that listening to music is not a passive activity, rather it is a mirror to one's emotions. With this objective in mind, a research question was framed to conduct a study on the effect multi-cultural songs have on a group of participants by getting them to participate in our experiment and subsequently self-report their emotions. In undertaking this research, we focused on uni-modal feature level fusion, where we combined different types of features from the audio modality into a single representation and subsequently applied feature selection to obtain the most relevant audio features that were expected to have an impact on an individual's emotional state at a particular point in time. With this academic report, we present our work, investigated in an empirical way, through using data-driven approaches in understanding how music affects humans emotionally.**

***Index Terms*—music, emotional states, machine learning, modelling**

I. Introduction

Music is often referred to as the social language in expressing emotions and studies have suggested that the attraction of music comes from its "emotional powers" to alter emotions and communicate feelings. Throughout ages and across cultures, music has the ability to arouse and induce a range of intense and complex emotions, primarily through acoustic rendering. The emotional dimension of music cannot be under-stated. While it can make us smile and also bring us to tears, it also triggers a calming effect on us. It has been theorized that the detection and analysis of acoustical features such as tempo and loudness are associated with certain types of emotions. Notwithstanding the validity of this premise, we examine two common paradigm approaches in selecting categories for labelling music tracks. This first approach is categorical and it is primarily influenced by the basic emotions theoretical framework (Ekman, 1992) [10], where the focus is on music's ability to express a set of basic emotions characterized as, happiness, sadness, anger and fear in individuals. The alternative classification and its related taxonomy, as explored by Russell [1], is dimensional in nature as it investigates the phenomenon of music through representing emotions in a continuous dimensional space of arousal and valence evaluation.

This paper is organized as follows : Section I introduces our motivations and inspirations in undertaking this study. In Section II we provide a literature overview and touch on the different aspects of relevant and current research of how music listening evokes feelings and consider the underlying mechanics to it. We then move on to Section III describing the methodology adopted where we cover the music data-set creation, survey procedure and the overall process flow for our project. Section IV explains the emotion labeling and classification techniques while Section V describes how the survey experiment was conducted.Section VI presents the results and discussions. The next section VII gives a brief overview on the ethical impact of our study before we wrap up with conclusion and discussion VIII and finally possible future extensions in section IX.

A. Motivations and Goals

For a significant number of us, the presence of engaging in music as part of our everyday routine is a testament of our affective connections with it. Clearly, listening to music allows us to express our inner feelings and regulate our emotions. This reasoning was a necessary and sufficient condition to inspire us in seeking a deeper understanding on the relationship between music and emotions.

The next motivation was sparked by the fact that emotion recognition through music is a fairly nascent research area in music information retrieval, a field that focuses on developing computational systems. Music emotion recognition systems can be applied in many diverse areas such as music therapy and recommendation for the music streaming industry. The former is especially important in the ongoing pandemic period where the therapeutic properties of music are highly suggested in the clinical and psychological space for treating anxiety and mood disorders such as depression.

II. Literature Review

Studying the impact music has on an individual's emotional state is not novel. It is a well-researched topic over the last couple of years. Studies conducted previously have used psychological as well as affective computing methods to study this relationship. In this section we examine and review ongoing literature in this field.

Existing studies [3] conducted on Music Emotion Recognition (MER) tend to follow Thayer's dimensional model for classification of emotions in songs. Mogn, Argstatter and Wilker [13] conducted a study to identify whether individual's

can identify Ekman's six basic emotions in music pieces. They created 18 musical segments using solo instruments which were thought to be a representation of each of the six emotions in three different ways. Then they conducted a survey on 115 participants and asked them to select the emotion they thought was present in a music clip. The study was successful with the majority of the participants being able to correctly identify the emotions being conveyed via the musical excerpts. This study helps set the foundation for us to use Ekman's model of emotion classification in our study.

Most of the prior research done in this area, aimed to study the impact music has on individual's diagnosed with mood disorders like depression or patients receiving palliative care. Stewart, Garrido, Hense and McFerran [4] conducted a study to analyze the effect of music on 7 participants who have tendencies of depression. They found two patterns amongst the participants. Participants either opted to listen to songs which shifted their mood or ones which were reflective of their current mood. The authors found that participants had limited awareness of the impact music had on them before they participated in this study. They also found that for some of the participants, music intensified their depressed state. Since, the participants in their study suffered from mood disorders, this study is not conclusive for a general group of people. For our study, we aim to remove the bias of only considering participants suffering from a mood disorder or any other form of disease. Our participants will be from different selective ethnicities, age groups and gender.

Hu and Lee [5] in their study conducted a cross-cultural study of music mood perception between two cultures: American and Chinese. Their song dataset comprised of American songs and the participant groups were people studying in US universities. They concluded that culture does have an impact on how people perceive emotions in songs. Since all the participants listened to the same set of songs, with no cultural diversity accounted for in the dataset, the results were better for native Americans. In our study, participants will be made to listen to songs in a language which they use on a daily basis and our highly familiar with. This will help reduce any biases because of language barriers.

With the progress in artificial intelligence domain, a lot of prior work has been done to detect the emotions in the music audio or video tracks. Pandaya, Bhattari and Lee [6] proposed a multimodal approach, which uses audio, video and the facial expressions to classify the emotions in a music track. Our approach does not aim to use a multimodal method; it focuses on extracting features from the audio tracks and using machine learning models to classify the emotions across a selective cross-cultural dataset. Video and facial expressions across cultures might not always go in accordance with the audio emotions, hence we opt to focus on the audio features of a song.

Song, Dixon and Pearce [12] in their paper propose an approach for evaluating the musical features for emotion classification. They collect human annotated categorical data for this task. Using the social tagging feature provided by last.fm[1], they generate a ground truth with four classes; happy, sad, angry and relaxed. Then an experiment using SVM is run with different number of features to select the best features which play a role in emotion classification. Their model reports an accuracy of 51.9%. This suggests there is a lot of noise in the ground truth dataset of last.fm and hence for our approach using last.fm dataset can result in skewed results after the actual experiment is conducted. We propose to use a different form of social tagging made available in Spotify along with last.fm and choose the better alternative out of the two.

In the study by Essid and Richard [7] it was proposed that music should be perceived in other aspects besides acoustic rendering. Their article gives an in-depth overview of processing music content through various types of heterogeneous information, such as lyrics and user-generated metadata thus giving rise to multi-modal music analysis studies. In their paper, they have made a distinction between the 2 categories of multimodal techniques - cross modal processing and multi modal fusion. The former, examines the effort in characterizing the "relationships" between different modalities while the latter investigates how best to combine the information conveyed by the different modalities so as to achieve a more thorough analysis of content. Our team has explored the feasibility idea of integrating different types of features from different modalities into a single common feature representation for our project while going through this survey. The authors have recommended the principal component analysis approach as a feature transformation technique that only selects useful descriptors, that is a subset of the most relevant features. However, given its limitations, another idea revolves around decision-level fusion which combines the intermediate results from unimodal decisions through a model based on weighting system taken on a particular modality. However, these weights are either chosen heuristically or by a trial-and-error procedure that can be more formalized through Bayesian's framework, which is the reason we opt to not go with this model.

Pouyanfar and Sameti [8] collected 280 popular songs from the All-Music Guide (AMG) and classified them into four basic emotion categories – happy, angry, sad and relaxed, based on Thayer's 2D emotion model. They pre-process the songs and extract the best features from them. Then, they created two categories of data. One category had all four emotion categories and another category just had half of the sad and relaxed songs. They used SVM to train this model. The features used in this paper were limited. Our study aims to use Spotify API[2] to extract more features for training the models. We will also be conducting a survey to validate the results with how emotions are actually perceived by individuals.

Brata and Darmawan [14] proposed an approach to classify the moods in Balinese songs using Spotify API audio features and K-means clustering algorithm. Similar to [12] they use a limited number of features for cluster generation. They use

[1]https://www.last.fm/home

[2]https://developer.spotify.com/

valence and energy for classification of songs into 4 different clusters. They generate a ground truth by asking 10 students to manually rate the emotions in a song on a scale of 1-5 and then compute the accuracy of the K-means algorithm based on this. They report an accuracy of 32% for the four classes of emotions. We believe including audio features like energy, tempo, danceability etc. will help in achieving better accuracy results in the emotion classification model and propose to do in this paper.

Riberio, Santos, Albuquerque and Silva [9] aimed to study the emotional state generated after an Emotional Inductions through Music (EIM) paradigm with a 6-minute recovery phase. They monitored the effect by using a valence and arousal self-report measure and physiological assessment. They concluded that self-report measures tended to change to neutral after 2 minutes in the recovery phase, but skin conductance levels data suggested a longer lasting arousal for both positive and negative emotional states. In accordance with this study, we will record the participant's emotional state immediately after listening to a song. The participants will have a very brief recovery phase before they listen to the next song excerpt, which would allow their emotional state to not become neutral.

In summarizing this section, we described studies relating to classification of emotions in songs using various machine learning and deep learning methods along with the inclusion of cultural and clinical perspective that the work was conducted on. We also examined the different emotion state models in which ongoing research has been carried. The review also examined alternative ways in measuring the affective content of music. In the following section we will discuss our framework proposed.

III. Methodology

Our methodology constituted of the following sub-areas, namely the music data-set creation and its associated audio features, the survey procedure, process framework and the tools used.

A. Dataset Creation and Features

Given that there was already a multitude of comprehensive music databases made available in the open source community, our team nevertheless opted to source playlists mostly from Spotify music domain. This was because research and investigations showed that there was a lack of consensus from the former method as to what categories ought be used in classifying music based on both audio and non-audio features. Hence, going with a pre-compiled music data-set meant that we were incorporating inherent subjectivity and bias from the original gathering method that could have influenced emotion distribution on the resulting output.

Our rationale for choosing Spotify was due to the following two reasons. Firstly, it is the most popular audio streaming platform for gaining access to massive amount of data of close to 70 million over tracks and popular user playlists. Secondly, it has evolved to fuse information from acquisition of Echo Nest, a music intelligent service that provides automatic data extraction from songs by web crawling and applying digital signal processing technique on the audio itself [2]. We also mined user generated playlists from the social music platform Last.fm using its rest api as an equivalent alternative.

To build our data-set from Spotify, we used its library and obtained authorization credentials to connect to its web API for retrieving all the songs within a playlist. The selection criteria applied in this extraction relied on the concept of social tagging wherein a playlist was only marked if it had more than a thousand like votes in Spotify. Within the research community, this form of tagging is considered to be a reliable way of capturing contextual knowledge such as genre, mood, instrumentation about music [15]. Further explanation on this concept is provided under Section IV-B.

Our resulting output was a dictionary composed of close to 2000 songs containing metadata information on the artist, album and the audio features that represented a fair distribution across the languages. Example of some key features derived include, but were not limited to the following:

1) Danceability : This explains how suitable a track is for dancing through key elements of tempo, rhythm and beat stability.
2) Tempo: This refers to the overall estimated tempo in beats per minutes. (BPM)
3) Valence: The musical positiveness conveyed by a track.

B. Survey Procedure and Participants

A survey design that could be translated into measurable factors was constructed as a critical component contributing to the objectives of our study. Prior to survey execution, subjects were solicited for our experiment. They were given a general idea on the goal of the experiment but they were not informed about the specific aims so as to eliminate any form of biased responses. We gathered participants (n= 120) from different cultures and across age ranges to understand what feelings were evoked from different music styles. The reason behind including participants from various age cohorts was to get a cross-generational perspective on the results obtained, post the experiment. They were asked to fill out a questionnaire, designed to collect data on their demographics (age, gender), background and information on music exposure such as familiarity with musical instruments. The survey form is described and a snapshot of it is referenced as Appendix C.

C. Process Flow

Our process flow is summarized in the following diagram, referenced as Figure 1 (see Figure 1). It clearly provides a visual representation of the sequential tasks involved in this study. The specific set of steps implemented and executed are described in the respective sections of this paper.

D. Tools

A broad range of open source tools and its various libraries were used in this study. We have tabled them against their

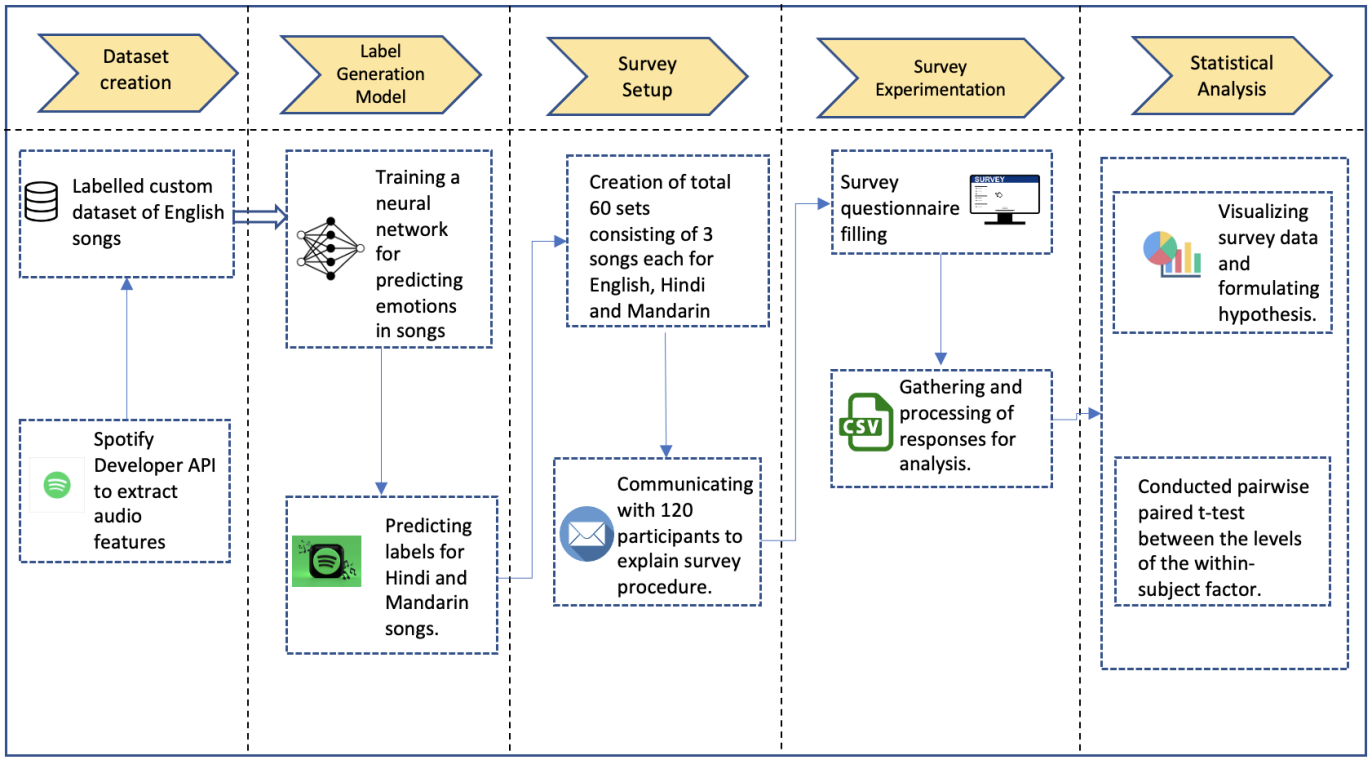

Fig. 1. Process Flow.

purpose for fit as stated in Table I. Python was the default language used for our machine learning and deep learning tasks along with the editing of the audio clips. Our survey was a react-app that was hosted on the Firebase platform which was preferred for free hosting and storage. It must be mentioned that we chose to develop the survey with React over the direct Google Forms because moving forward our idea is to integrate it with a front-end real-time dashboard.

TABLE I
TOOLS USED

S. No.	Tool	Task
1.	Python	Dataset creation and ML models
2.	R	Statistical Modelling and Inference
3.	Tableau	Data Visualization
4.	ReactJS and Firebase Console	Survey
5.	HTML and CSS	Survey Layout

IV. IMPLEMENTATION

Our implementation steps revolved mostly around the annotation strategy and its generated data-set that was subsequently used for executing the survey experiment.

A. Annotation Strategy

Determining the emotional category to which a song belongs can be quite challenging and this one was of the key areas of our study. In this subsection, we describe the labeling process before moving on to explain the machine learning models explored for classifying the emotions in songs. To reiterate what was mentioned in the introduction, annotating the music excerpts followed the categorical discrete approach and it was limited to categories expressed as happy, sad, calm and energetic. It is thought that this set of basic emotions are readily expressed and perceived equally across cultures. They also forms the base from which other views of emotional expressivity are derived from.

B. Data Generation Post Labelling Process

Data labelling or rather data annotation was a necessary design consideration in evaluating the auditory stimuli (music pieces) as they transmitted varying emotional sensations in different combinations to participants participating in our experimental study. Our algorithms could successfully comprehend the inputs only if this step was carried out prior.

From Spotify, the first dataset was un-annotated while the second was annotated with four emotion labels characterized as happy, sad, energetic and calm. The third song dataset, from Last.fm was labelled as well, similar to the second. To further explain this idea, we filtered tags based on a corpus of affect-related terms in the title of the playlist. For example, to search for Happy songs in the English language, the search contained the word ”Happy”. English songs were collected and aggregated in this manner for parameters containing the search terms of sad, happy, energetic and calm. Appendix B gives a

snapshot view of this final dataset output. For the un-annotated dataset, the search criteria for playlists was generated not by using tagging of keywords, rather it was sourced from using top playlists such as Top 100 Chinese Songs, Top US 100 songs and Top 100 Hindi Songs across the three languages.

We also adopted a music labelling transfer task where the model, trained to classify labels from the English song language playlist was transferred over to be labels for the target language, Hindi and Mandarin playlist for the annotated data-sets. The reason triggering this was that both Hindi and Mandarin did not have popular user playlists for all the emotions, so to avoid sourcing noisy data, which might have a potential to skew our results later on, the similar annotation process was propogated through.

C. Machine Learning Models

Machine learning algorithms were experimented with so as to learn that given certain audio features, which one best contributes towards them being classified into a certain emotional category. In this section we go through both the unsupervised as well as supervised models.

1) Unsupervised Techniques: Using unsupervised machine learning techniques, we clustered songs into different buckets of emotion classes on the basis of similarity in audio features. Clustering was performed on each language's dataset independently. We used the un-annotated dataset generated in IV-B.

K-means, mini-batch K-means and BIRCH were tried for clustering the songs. Only audio features which were distinct across songs and numerical in nature were chosen. Others like key, track name, duration_ms etc. were dropped with the rationale of contributing very little towards classification of emotions in songs. Prior to performing clustering, the features were normalized using Min_Max_Scaler. Due to lack of a ground truth, calculation of an exact accuracy was not possible. Upon visualizing the clusters formed it was noticed that the clusters overlapped with each other in case of K-Means and mini-batch K-means. In case of BIRCH, 3 distinct clusters were not obtained. On manually checking the generated clusters, a lot of discrepancy was found in the clusters created. Several songs which were considered to be *sad* were classified as *happy*. The manual verification process was performed by 4 people for English songs, 3 for Hindi and 1 for Mandarin.

Figure 2 shows the clusters formed with k equal to 3 for 300+ English songs. It can be observed there is an overlap between the clusters. Similar results were observed across the other two languages. Due to a lack of good and distinct clusters, we decided to not pursue this approach further.

2) Supervised Techniques: Due to a lack of good results from the unsupervised algorithms, we shifted our focus to supervised machine learning methods. For the purpose of training supervised models, we used the second and third datasets generated in IV-B. Same set of experiments were run on both since there was no conclusive evidence proving Spotify to be better than last.fm or vice versa.

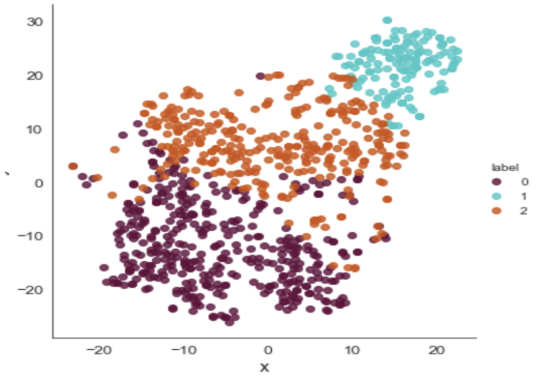

Fig. 2. K-Means clustering visualization with k=3

TABLE II
MODEL ACCURACY

Dataset	Model Accuracy
Last.fm	48.34
Last.fm + Spotify	58.76
Spotify	74.54

A multi-layer multi-class neural network model was used for the purpose of training the model[3]. The model was trained to predict a label out of *Happy,Sad,Energetic* and *Calm*. Prior to training the models, we used Min_Max_Scaler to normalize all the features in our Spotify and last.fm datasets. We also categorically encoded the output labels. The annotated English songs dataset was split into 80-20 train-test sets.

We use a Sequential model with a plain stack of 2 layers. The first layer of the neural network takes as input the 10 audio features and passes that to a second Dense layer. The final layer in the network is the output layer which has four output classes possible (Happy, Sad, Energetic and Calm). We used the KerasClassifier as the estimator and passed the base model defined above to it, with the batch size equal to 200 and epochs equal to 1000.

The estimator is evaluated using K-Fold Cross (k=10) validation using the training data. This validation is done to check whether the model defined above is a good fit for the problem we are trying to solve. This accuracy is calculated by taking an average of the accuracy of each fold. After evaluation of the model, we train the model. The model is tested on the split test set to check its soundness.

Post the training of the model, we used it to predict the labels for Hindi and Mandarin songs. The accuracy for the models tested on last.fm dataset, last.fm combined with Spotify dataset and Spotify dataset alone is shared in Table II. For the purpose of conducting the survey experiment, we proceed with using data-set which gives the best results.

[3]https://towardsdatascience.com/predicting-the-music-mood-of-a-song-with-deep-learning-c3ac2b45229e

TABLE III
STATISTICS OF ANNOTATED MUSIC DATASET

Language	Emotion				Total
	Calm	Happy	Sad	Energetic	
Chinese	52	59	193	50	354
English	234	207	293	640	1374
Hindi	1	25	229	174	429

D. Annotated Dataset

We proceeded to utilize the annotated Spotify music dataset from the supervised approach for all three languages, as it provided the most acceptable results in terms of accuracy. We have been careful to position the algorithm choice against the following requirements that we consider important for it to be useful in progressive affective computing research, namely reproducibility, reliability and interpretability.

Table III presents the total number of songs across each emotion for all the three languages. Since there was only one *calm* labelled song generated for the Hindi dataset, we decided to drop the label *calm* for the purpose of conducting the experiment. We only selected songs from Happy, Sad and Energetic categories for providing audio simulation to the participants.

In readying the data for the survey experiment, 20 songs were randomly sampled from each of the three categories and 3 languages. Hence, each language had 20 sets which contained three clips ordered and labelled in the following manner of sad, happy and energetic (SHE).

V. EXPERIMENT

The idea of the experiment was to have each of the participants listen to excerpts of 3 audio tracks, lasting over 60 seconds per clip in their identified first language and rate the emotions that they most accurately experienced or expressed after each song on a continuous 5 point scale with 1 being the least affected and 5 being the most impacted. It should be noted that the length of each music excerpt was edited using our programmed python script from the original length to precisely 1 minute as this was gauged to be long enough to trigger an emotional response in participants.

A. Survey Execution

The choice of our survey medium and the social context surrounding in which the music listening experiment was conducted was essentially online. This was deemed to be the best method for achieving a high response rate with negligible material cost. The duration in which the entire process was framed is referenced using a timeline as detailed in Appendix D. Survey links were circulated along with the instructions and audio sets assigned to each participant via email. They were advised to attempt the survey in an isolated and undisturbed environment. The web link to the survey is : https://emotionmusicsurvey-58649.web.app/

B. Preliminary Evaluation

From the survey conducted, we observed a drop-out rate estimated at around 8 percent, meaning 109 respondents responded effectively in our bucket of analyzed data out of the 120 participants. Appendix A examines the association between participants demographics and genre preferences for example. Clearly participants in the $<$ age range of 30 group formed the bulk of our subjects and they were most oriented to listening to English songs in no specific ordered period but mostly daily across all kinds of music genres, generally feeling calm and relaxed after music activity.

VI. RESULTS

We present here our findings on the curve of emotional state from our experimental study in a two-step approach. First, we analyzed summary statistics through chart plotting using histograms and box plots. Next, we went on to infer our results by formulating the hypothesis, checking the assumptions around our dataset distribution and computing the t-test.

A. Descriptive Statistics

To understand the results gathered from participants post the survey, we performed an exercise using histograms and box-plots for exploratory data analysis.

1) Histograms: For histograms, we subtracted the pre and post emotion scores for each song clip to get the absolute change in each participant's ratings for the four emotions. A positive change score indicates an increase of that emotion in an individual after listening to the music excerpt while a negative score indicates a decrease of that emotion in an individual. For example, if a participant gave a happy rating of 4, before listening to the intended song and a happy rating of 3 post-listening to the same clip, then that was constituted as a decrease of 1 in the individual's rating for happy emotion.

Figure 3 shows the change in the ratings observed in the participants after they listened to a sad song. 45 participants reported an increase in their sadness level and 40 reported a decrease in their happiness level.

Figure 4 shows the change in the ratings observed in the participants after listening to a happy Song. 58 participants reported a decrease in their sadness level and 56 reported an increase in their happiness level.

Figure 5 shows the change in the ratings observed in the participants after they listened to a sad song. After listening to Energetic songs, the variance across all four emotions was close to nil, with most of the participants reporting a difference of 0. This is in accordance with Thayer's model where energetic and happy emotions lie in the same quadrant.

2) Box Plots: Box plots were another graphical technique that we used to examine the data for important features such as symmetry and to detect pockets of observations that were removed from the overall data. Generally the box plots demonstrated that the distribution was somewhat symmetrical and it formed a first point of check for us to proceed with the paired two sample t-Test.

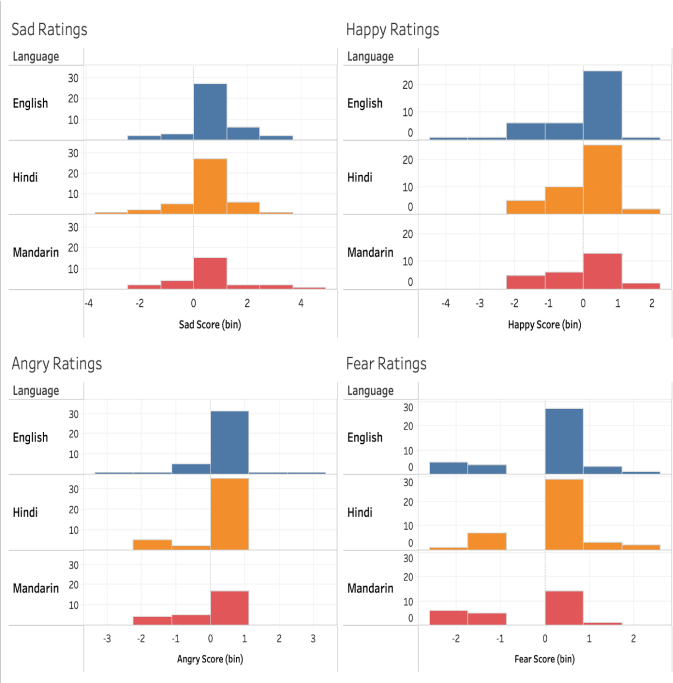

Fig. 3. Emotion rating post listening to Sad songs

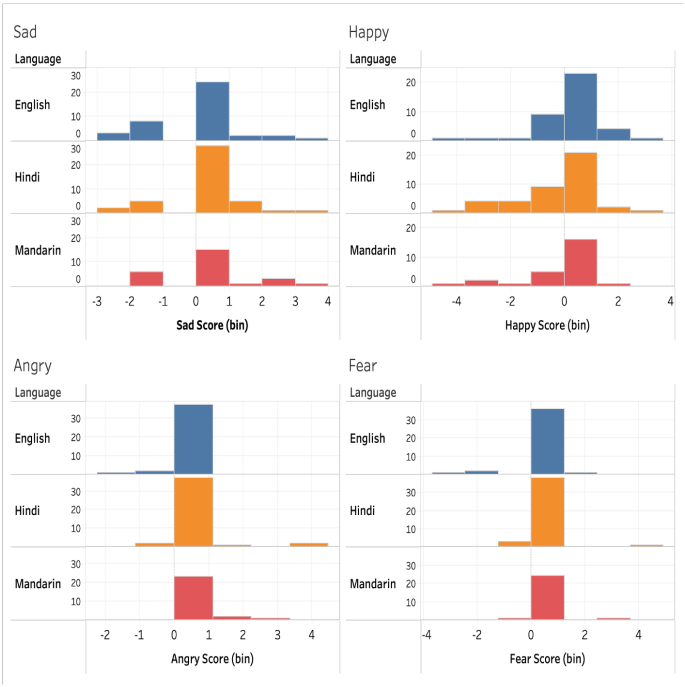

Fig. 5. Emotion rating post listening to Energetic songs

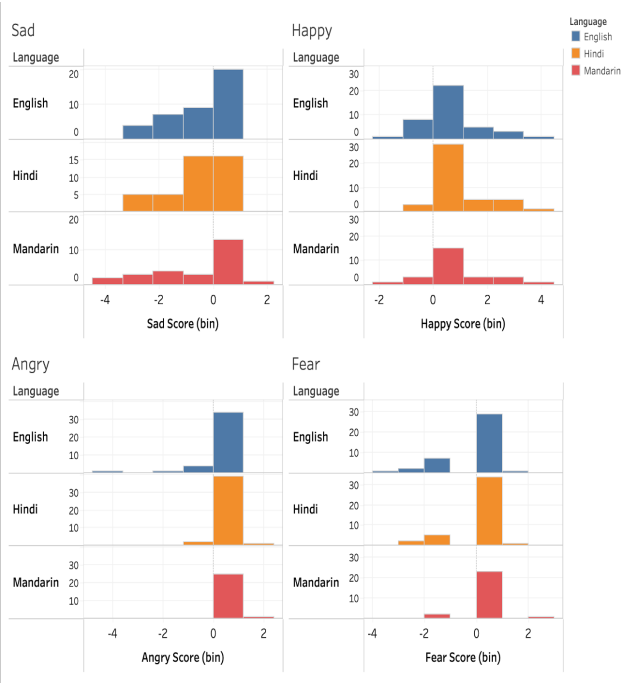

Fig. 4. Emotion rating post listening to Happy songs

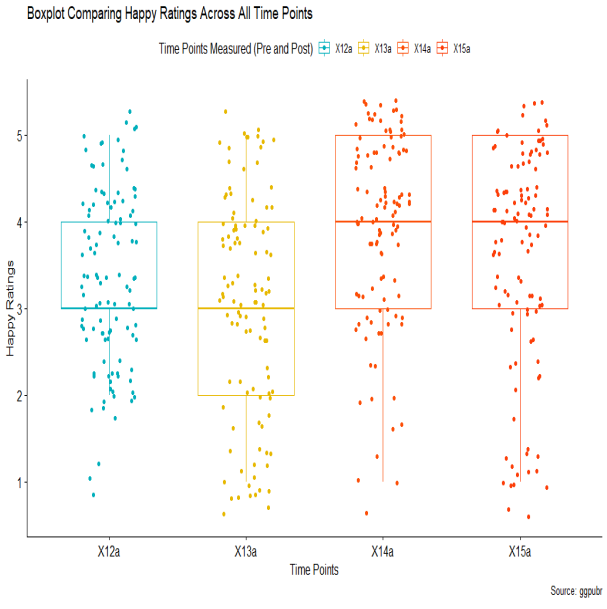

Fig. 6. Happy ratings compared across all time points

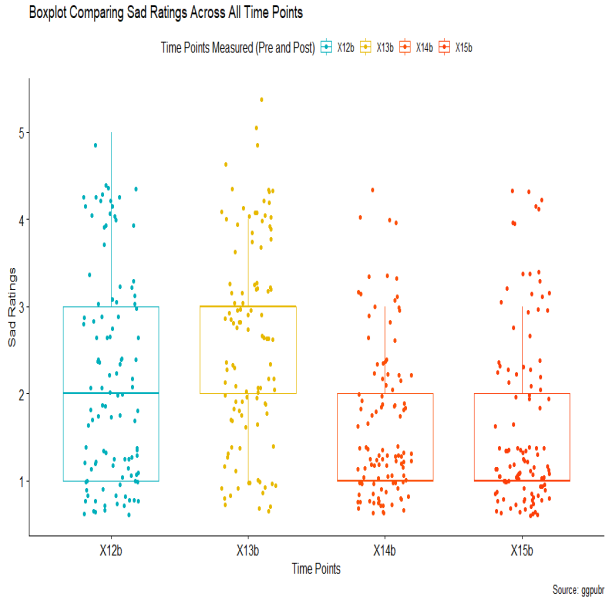

Fig. 7. Sad ratings compared across all time points

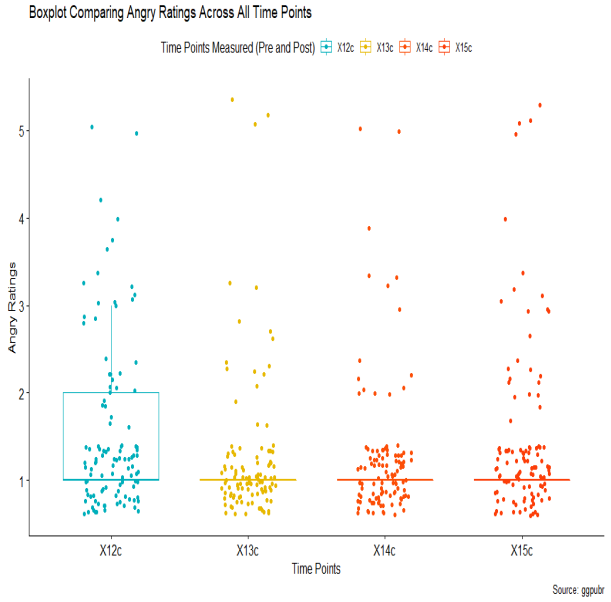

Fig. 8. Angry Ratings compared across all time points

The lowest or minimum rating score before a happy track was played to all participants was 2 while the lower quartile was three and the upper quartile was 4. The maximum score excluding outliers was 5. The mean was also quite consistent except that it took a slight dip of 7 percent after listening to a sad excerpt.

Here we observe that the maximum score rated actually dropped when a happy and energetic track was played right after the subjects reported experiencing negative feelings in response to a music track that conveyed sadness. It is unsure however, if these feelings were perceived as genuine feelings or otherwise. This was not a surprising outcome as it proved that listening to music that was not classified as sad could actually infleunce and uplift the sad emotions participants felt previously.

We observe that music had very little impact in calming participants down when the emotion experienced was anger. This could be explained by the fact that most people do not come into an experiment study feeling angry so music had very little to enforce an impact on for this classification of emotion.

The same reasoning applies as the case for angry emotions.

In summary, across the 4 ratings, we observed outliers in the data which constituted a paradox and could be interesting to pursue for a further explanation from a psychological standpoint. For example, while happiness was conceptualized as a positive emotion, there were a handful of participants who reported feeling sad for a music intended to convey happiness is worthy of empirical investigation. There were also a small number of participants who were happy after listening to songs perceived as sad. One explanation could be that it triggered a specific joy-related memory in the listener.

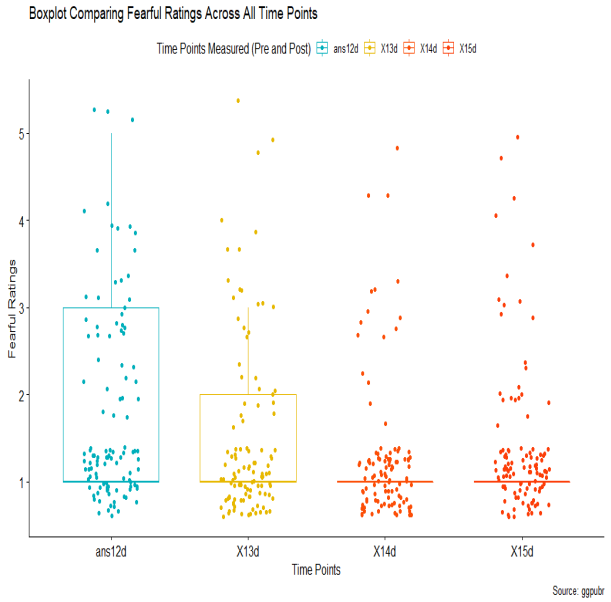

Fig. 9. Fear Ratings compared across all time points

B. *Statistical Inference*

The purpose of undertaking statistical inference for the survey results was essentially to factor uncertainty into account when drawing conclusions for our study. We achieved this by first formalizing a hypothesis and questioned the probability of observing a difference of means across time periods of the audio listening exercise. The details are as follows :

1) Hypothesis Formalized: A separate hypothesis was formulated for each emotion category that we detailed out in the survey form from question 12 through to 15. These emotions were happy, sad, anger and fear. The repeated measures here

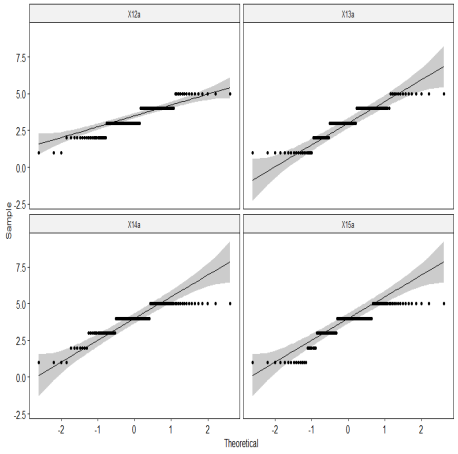

Fig. 10. QQ Plot Happy

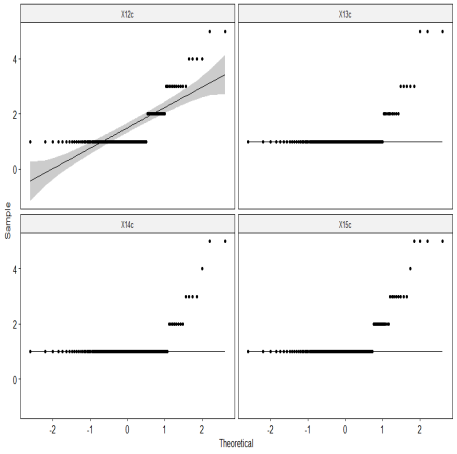

Fig. 12. QQ Plot Angry

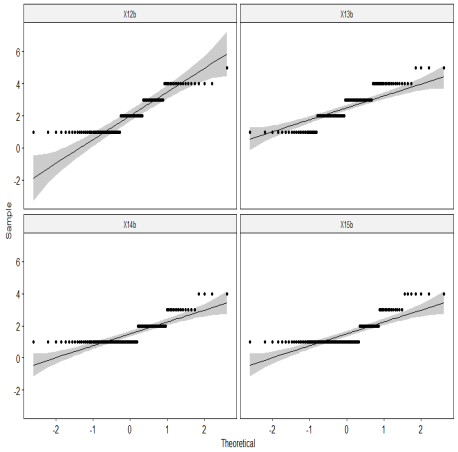

Fig. 11. QQ Plot Sad

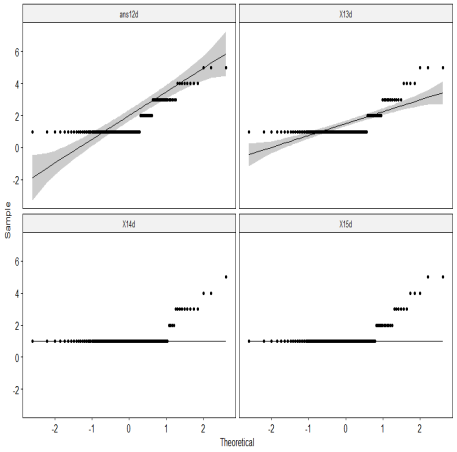

Fig. 13. QQ Plot Fear

were the time points post listening to each clip and we wanted to test whether the means for all time points were equal. While the null hypothesis (H0) is stated as no effect and no difference, the alternate hypothesis (H1) is the exact opposite to the motivation of our study. An example of the hypothesis framed is described below as :

H0 : Mean ratings when emotions are rated as happy across all time points are the same. H1 : Mean ratings when emotions are rated as happy for at least two time points are different from the others.

H0 : Mean ratings when emotions are rated as sad across all time points are the same. H1 : Mean ratings when emotions are rated as sad for at least two time points are different from the others.

2) Normality Assumption: We tested for normality for each time point to check if the data was normally distributed and the graphical results are shown below. In general, we observe a bit of deviation but in general it was minimal to skew the overall results.

3) t-Tests: To illustrate if our music-emotion induction experiment was successful, we conducted a two group by four time point repeated measure of analysis to examine if indeed the ratings score changed significantly between the pre and post test condition in the two groups. Basically, we conducted one within subject test for each emotion. The withing subject factors here are the time points and the dependent variables are the emotion ratings. An adjusted p-value was used to determine the significance level of our statistical tests.

The pairwise differences between time points were statistically significant as the p-value was less than 0.05. Hence, there was strong evidence to reject the null hypothesis for happy and sad emotions. However, there was only partial evidence to reject the null hypothesis that mean ratings across all time points were the same for emotions tagged as anger and fear. The details of these results are attached in the Appendix C.

C. Feedback Mined

A word cloud generator was also visualized through simple text mining on the open ended text based question of our survey form. As expected, the information conveyed though the frequency and proportional text size of jumbled and disparate words matched up to our statistical findings.

VII. Ethical Impact

The impact and usage of artificial intelligence is increasing with each passing year and it is penetrating every aspect of humans' life. This growth requires us to stop and consider the ethical impact our research in this area has.

The current approach proposed in this paper, of users self-reporting their emotional states, is a non-intrusive way of gathering information about user's emotional states. Down the line though, this might not be feasible as it can get cumbersome to enter data before being allowed to listen to music. The new methodology of tracking the emotional state might shift focus to using the sensory data from consumer wearable or capturing facial expressions to predict the current emotional state of a user. The use of sensors or a front camera can be intrusive, and many can consider it to be an invasion of privacy.

Deng, Leung, Milani and Chen [11] conducted a study to study the effect of music on an individual's emotional state using self-reporting measures and by accessing a participant's previous song history. They use the song history in one sitting to build a recommendation system which will suggest songs to listener's based on the inferred emotional state in that sitting. Apart from an invasion of privacy, suggesting songs based on current state can further worsen the condition for some listeners [4].

Beside the impact it can have on an individual, classification of emotions in songs might lead to a consumer bias. Certain individuals might choose to avoid specific songs which have been tagged as depressing, sad or anxiety inducing. This can lead to a reduction in the listener-ship of these songs and hence a loss to the stakeholders involved.

VIII. Conclusion And Discussion

Based on the stimuli used in our study and the results generated, music has been shown to successfully induce either positive or negative emotions and influence mood alterations on an individual's emotional state. While intended emotions to achieve this can at times be recognized from the music piece itself, it is noteworthy to call out that this observation is not as strong when bench-marking against emotions demonstrated such as anger and fear. Our alternative hypothesis is supported stating that the mean ratings when emotions are rated as either happy and sad for a minimum of two time points differ. While our findings definitely enrich and provide empirical support to music-emotions relationship, much work is yet to be investigated upon and incorporated in our study through an adequate concept around the theory of emotion. Nevertheless, this is a foundation that we can look back to and build upon.

IX. Future Works

We believe that progressing from a restrictive uni-modal model to a multi-modal decision fusion concept which combines features from different modalities into a single feature representation matrix would be the next step for elevating the work done on this project. It is also hoped that we can extend the research in studying the relationship between emotions and music genres. One idea to progress with this is to use principal component analysis to determine music genres which are orthogonal in nature and use it to build a predictive model against emotions expressed. Related to this, is to incorporate a combination of factors around user context such as an individuals' personality and association to music in determining how emotions are affected by music in totality.

Acknowledgment

Our work was conducted with the support of Professor Desmond Ong, Varsha Suresh and Gerard Christopher Yeo Zheng Jie. We would like to thank them for their guidance and feedback on our project. We also wish to acknowledge all the participants who helped us in conducting our study by providing their responses to our survey questionnaire.

References

[1] Russell, J. A. (1980). A circumplex model of affect. J. Pers. Soc. Psychol. 39, 1161–1178.

[2] Renato PANDA, Hugo REDINHO, Carolina GONÇALVES, Ricardo MALHEIRO and Rui Pedro PAIVA. 2015. How Does The Spotify API Compare To The Music Emotion Recognition State-Of-The-Art?, Proceedings of the 18th Sound and Music Computing Conference, June 29th – July 1st 2021.

[3] Kim, Junghyun & Lee, Seungjae & Kim, SungMin & Yoo, Won. (2011). Music mood classification model based on arousal-valence values. International Conference on Advanced Communication Technology, ICACT.

[4] Stewart J, Garrido S, Hense C, McFerran K. Music Use for Mood Regulation: Self-Awareness and Conscious Listening Choices in Young People With Tendencies to Depression. Front Psychol. 2019;10:1199. Published 2019 May 24.

[5] Hu, Xiao & Lee, J.H.. (2012). A Cross-cultural study of music mood perception between American and Chinese listeners. Proceedings of the 13th International Society for Music Information Retrieval Conference, ISMIR 2012. 535-540.

[6] Pandeya, Y.R.; Bhattarai, B.; Lee, J. Deep-Learning-Based Multimodal Emotion Classification for Music Videos. Sensors 2021, 21, 4927.

[7] Essid, Slim & Richard, Gaël. (2012). Fusion of Multimodal Information in Music Content Analysis.

[8] S. Pouyanfar and H. Sameti, "Music emotion recognition using two level classification," 2014 Iranian Conference on Intelligent Systems (ICIS), 2014, pp. 1-6

[9] Ribeiro, F. S., Santos, F. H., Albuquerque, P. B., & Oliveira-Silva, P. (2019). Emotional Induction Through Music: Measuring Cardiac and Electrodermal Responses of Emotional States and Their Persistence. Frontiers in Psychology, 10.

[10] Ekman, P., and Friesen, W. V. (1984). Emotion Facial Action Coding System (EM-FACS). San Francisco, CA: University of California Press.

[11] James J. Deng, Clement H. C. Leung, Alfredo Milani, and Li Chen. 2015. Emotional States Associated with Music: Classification, Prediction of Changes, and Consideration in Recommendation. ACM Trans. Interact. Intell. Syst. 5, 1, Article 4 (March 2015), 36 pages.

[12] Song, Yading & Dixon, Simon & Pearce, Marcus. (2012). Evaluation of Musical Features for Emotion Classification. Proceedings of the 13th International Society for Music Information Retrieval Conference, ISMIR 2012.

[13] Mohn, Christine & Argstatter, Heike & Wilker, Friedrich-Wilhelm. (2010). Perception of six basic emotions in music. Psychology of Music. 39. 10.1177/0305735610378183.

[14] Brata, I & Darmawan, I. (2021). Mood Classification of Balinese Songs with the K-Means Clustering Method Based on the Audio-Content Feature. JELIKU (Jurnal Elektronik Ilmu Komputer Udayana). 9. 331. 10.24843/JLK.2021.v09.i03.p03.

[15] Lamere, Paul & Pampalk, Elias. (2008). Social Tags and Music Information Retrieval.. Journal of New Music Research. 37. 24. 10.1080/09298210802479284.

APPENDIX A
SURVEY DEMOGRAPHICS

Characteristic	N	Emotional States				p-value[2]
		<29	>60	30-44	45-59	
		N = 91 (83%)[1]	N = 1 (0.9%)[1]	N = 13 (12%)[1]	N = 4 (3.7%)[1]	
Gender	109					0.3
female		46 (51%)	1 (100%)	7 (54%)	0 (0%)	
male		44 (48%)	0 (0%)	6 (46%)	4 (100%)	
preferNotToSay		1 (1.1%)	0 (0%)	0 (0%)	0 (0%)	
Preferred_Language	109					0.2
chinese		12 (13%)	1 (100%)	3 (23%)	2 (50%)	
english		45 (49%)	0 (0%)	5 (38%)	1 (25%)	
hindi		34 (37%)	0 (0%)	5 (38%)	1 (25%)	
Listening_Period	109					0.2
free		23 (25%)	1 (100%)	4 (31%)	3 (75%)	
notSpecific		62 (68%)	0 (0%)	9 (69%)	1 (25%)	
work		6 (6.6%)	0 (0%)	0 (0%)	0 (0%)	
Listening_Frequency	109					0.7
3-6days		25 (27%)	0 (0%)	3 (23%)	0 (0%)	
3days		17 (19%)	0 (0%)	3 (23%)	2 (50%)	
daily		49 (54%)	1 (100%)	7 (54%)	2 (50%)	
Music_Genre	109					0.2
acoustic		4 (4.4%)	0 (0%)	1 (7.7%)	0 (0%)	
all		51 (56%)	0 (0%)	7 (54%)	0 (0%)	
Bollywood		1 (1.1%)	0 (0%)	0 (0%)	0 (0%)	
Classic Rock		1 (1.1%)	0 (0%)	0 (0%)	0 (0%)	
classical		4 (4.4%)	1 (100%)	1 (7.7%)	1 (25%)	
hiphop		4 (4.4%)	0 (0%)	0 (0%)	0 (0%)	
indie		10 (11%)	0 (0%)	2 (15%)	0 (0%)	
Light indian/Hindi music		0 (0%)	0 (0%)	0 (0%)	1 (25%)	
mando pop		1 (1.1%)	0 (0%)	0 (0%)	0 (0%)	
Minimal Techno		1 (1.1%)	0 (0%)	0 (0%)	0 (0%)	
pop		11 (12%)	0 (0%)	2 (15%)	2 (50%)	
Pop and R&B		1 (1.1%)	0 (0%)	0 (0%)	0 (0%)	
Qawwali		1 (1.1%)	0 (0%)	0 (0%)	0 (0%)	
Rock, Metal		1 (1.1%)	0 (0%)	0 (0%)	0 (0%)	
Post_Music_Emotion	109					0.8
calm		48 (53%)	0 (0%)	6 (46%)	2 (50%)	
energetic		18 (20%)	0 (0%)	3 (23%)	0 (0%)	
happy		23 (25%)	1 (100%)	4 (31%)	2 (50%)	
sad		2 (2.2%)	0 (0%)	0 (0%)	0 (0%)	
Music.Training	109	15 (16%)	0 (0%)	2 (15%)	0 (0%)	>0.9

[1] n (%)

APPENDIX B
ENGLISH SONGS ANNOTATED DATASET

name	album	artist	length	danceability	acousticness	energy	instrumentalness	liveness	valence	loudness	speechiness	tempo	key	time_signature	mood
10000 Hours	10000 Hours	Thomas Vee	168960	0.66	0.0428	0.611	0	0.0942	0.336	-5.863	0.0354	89.895	3	4	Calm
11 Minutes (with Halsey feat.	11 Minutes (with Halsey feat. Travis Barker)	YUNGBLUD	239507	0.464	0.0116	0.852	0	0.108	0.233	-3.804	0.067	160.075	11	4	Energetic
16 Shots	16 Shots	Stefflon Don	224727	0.684	0.0271	0.79	0.000371	0.0854	0.549	-4.871	0.0735	95.083	3	4	Energetic
20 All-Time Greatest Hits!	Get Up Offa That Thing	James Brown	250200	0.883	0.225	0.664	2.30E-06	0.941	0.8	-10.395	0.411	118.104	4	4	Happy
432 Water Crystals	Just Look at You	369	187385	0.558	0.985	0.249	0.925	0.103	0.111	-14.715	0.0303	94.991	7	4	Calm
5 Day Mischon	Day 5: For Carol	Tom Misch	394862	0.485	0.765	0.497	0.886	0.132	0.065	-10.022	0.0308	100.857	2	4	Calm
7 rings	thank u, next	Ariana Grande	178626	0.778	0.592	0.317	0	0.0881	0.327	-10.732	0.334	140.048	1	4	Energetic
99 Luftballons	99 Luftballons	Nena	233000	0.466	0.089	0.438	5.62E-06	0.113	0.587	-12.858	0.0608	193.1	4	4	Happy
A Beautiful Lie + 30 Seconds T	ATTACK	Thirty Seconds To Mars	189200	0.331	0.00344	0.876	0.000835	0.732	0.299	-1.894	0.0603	175.009	5	4	Energetic
A Brief Inquiry Into Online Rel	Be My Mistake	The 1975	256688	0.572	0.835	0.155	0.000137	0.0906	0.0949	-14.405	0.0344	109.923	6	3	Sad
A Burden to Bear	A Burden to Bear	Emmanuelle Rimbaud	129410	0.394	0.995	0.0475	0.955	0.105	0.172	-26.432	0.072	71.241	6	5	Calm
A Collection	Crystal Baller - 2006 Remaster	Third Eye Blind	255760	0.379	0.0142	0.695	1.13E-05	0.196	0.432	-5.922	0.0705	193.672	0	4	Energetic
A Deeper Understanding	Pain	The War On Drugs	330760	0.523	0.00387	0.833	0.000339	0.0841	0.482	-4.374	0.0405	115.975	7	4	Sad
A Different Way (with Lauv)	A Different Way (with Lauv)	DJ Snake	198285	0.784	0.495	0.757	1.18E-06	0.142	0.587	-3.912	0.0384	104.996	8	4	Energetic
A La Plage	A La Plage	Ron Adelaar	141888	0.504	0.994	0.0584	0.956	0.115	0.553	-20.461	0.0516	134.209	5	4	Calm
A Place In The Sun	My Own Worst Enemy	Lit	169026	0.494	0.00129	0.946	0	0.398	0.741	-2.757	0.0637	103.408	4	4	Energetic
A Sentimental Education	Fire in Cairo	Luna	209052	0.587	0.379	0.713	0.439	0.332	0.816	-8.858	0.0268	130.05	1	4	Sad
A Song For Every Moon	Easily	Bruno Major	210240	0.772	0.491	0.256	0.00612	0.144	0.357	-8.545	0.0481	118.902	7	3	Sad

APPENDIX C
SURVEY

Emotional Responses to Music

(Analyzing changes to an individual's emotional state arising from listening to a given music track)

Welcome aswe@gm to our survey

Let's Start !

1. What age range best describes you?

- ○ Less than 29 years
- ○ Between 30 to 44 years
- ○ Between 45 to 59 years
- ○ More than 60 years

2. Which gender do you identify by?

- ○ Male
- ○ Female
- ○ Prefer not to say
- ○ Others

3. Which of the following language do you most prefer to listen to songs in?

○ English ○ Hindi ○ Chinese

4. When do you listen to music the most?

- ○ Free time
- ○ Working hours
- ○ No specific order of time

5. How often do you listen to your preferred music genre?

- ○ Daily
- ○ 3 days a week
- ○ 4-6 days a week

6. What specific genre of music do you listen to the most?

- ○ Pop
- ○ Acoustic
- ○ Hip-Hop
- ○ Indie/Indie-pop
- ○ Classical
- ○ Every kind of genre
- ○ Others

APPENDIX D
SURVEY TIMELINE