Study Report

Group 11

by Nathaniel, Florian, Alexandre, Livio and Timon

Teufel: Sense User’s Musical Mood

Study goal

To be able to “sense” the musical mood of a user is challenging and even with the right sensors and good machine learning algorithms it might take some time for the algorithms to propose to a user the songs he would expect based on his moods.
For this reason we thought of giving the user the possibility to overwrite the mood preset by the algorithm.
In our user study we want to find out what interface works best for our user group to get the most satisfaction out of a user. This study consists of two parts.
In a first step we want to find out what kind of attributes work best to identify a mood. Those attributes will then be used as a scale on the given interface. The chosen keywords need to be precise and accurate enough for a user to describe or recognize a given mood setting.
In a second step we want to analyze which attribute pairs when applied to our Navigator interface (a two-dimensional graph) work best for a precise determination of a song’s mood setting.

Specifically we have the following hypothesis:

H1: Which factors are most descriptive for a user’s given mood setting and help to describe a song?

H2: Which factor pairs are most accurate when used as the axis labeling to associate a song with a mood setting?

Limitations

After some thoughts and having taken a look at the results of the first study, we decided to focus on the navigator and discarded the sliders, even though when first asked the users prefered the sliders. However users might not know what interfaces is best for them without using them over time and compare variables (like speed, precision of the result etc).
Unfortunately we could not compare this over time with the same users in this course scope, but the expected results is that the casual music listener like our persona Peter would get tired of sliders.
In addition to that from the interaction we had with some potential user doing our study we found that the navigator is quite a lot more convenient and quicker to use than the sliders, although it might be slightly less precise since it only offers two dimensions for the attributes. The sliders seem to be too cumbersome for a user that only wants to quickly choose the kind of song he wants to listen to. What’s more with a navigator (in the form of a 2-dimensional graph) it is possible for a user to draw a curve from one mood (point on the graph) to another and passing through a third one as a slower transition between the current mood to the desired one. This would be a lot harder to do and visualize with sliders.
But we did not strike out the possibility of adding a level to our interface i.e. the default mode would be the navigator but an experienced user could have the ability to set his mood with sliders and he might be able to select which factor he wants to manually selects based on the one used to classify the musical mood.

Another limitation of our study is the fact that we questioned and interviewed mainly fellow students and parents (roughly 15 for the first study and 10 for the second).

Method

Finding the right Attributes

The attributes we chose to evaluate were: loudness, happiness, speechiness, tempo, energy and danceability. We chose those attributes based on spotify's Tuneable Track attributes[1]. This will allow an easy integration with Spotify API when it comes to finding the right track that match the chosen attribute values. This is the main reason why we did not let the user choose which words would better represent the mood for him because we then might not be able to classify the music based on this words (we are limited by the algorithms available).

We set up an online survey using Google Forms that presents three songs to the user and asks him/her to rate for each factor on a scale of 0 to 10 (this would mimic the behavior of a slider) how much it fits for the given song. Additionally we (optionally) asked the users about their age range, gender and occupation.

The songs we used are: Ed Sheeran - Shape of You, Kendrick Lamar - HUMBLE. and Yiruma - River Flows in You.
Those songs were picked based on having distinctive values for the given attributes, taken from the Spotify API. In other words, the songs should represent as different mood settings as possible.

Axis Labeling

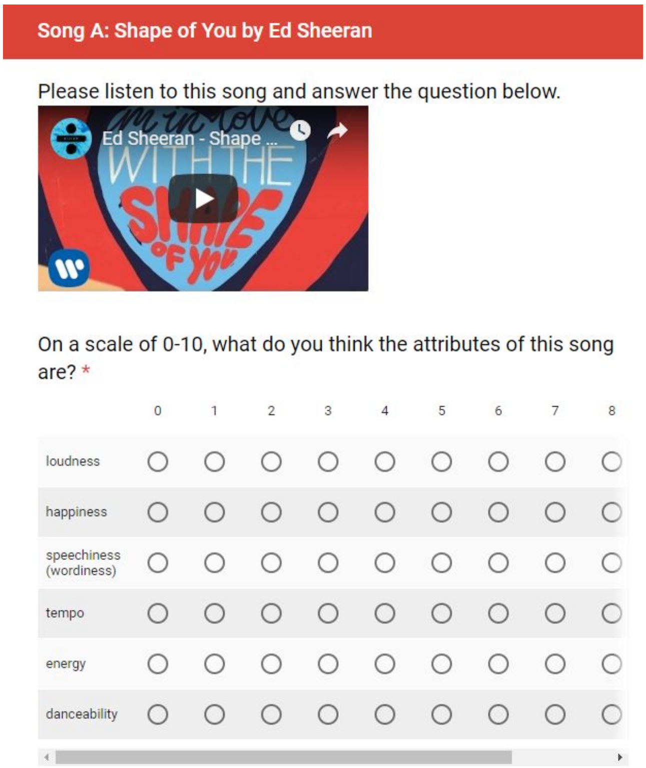
To find out how precise a user can associate two given attributes with a song we conducted moderated A/B tests. With the results from our first survey which gave us the answer to our first hypothesis - which attributes work best - we used those for our A/B test. From the four attributes we made four different variations for the axis labeling (one attribute per axis, two axis in total). For each of those four attribute pairs the user will be presented with a song that he has to describe using those two attributes on the axis. For each user the chosen song will be randomized such that we have different songs rated by different user for a given axis labeling.
On the second step we do another A/B testing where the user chooses a value for two given attributes on which we then do a test. This will allow us to have a large variety of different values chosen when conducted on a user group.

[1] *"Get Recommendations Based on Seeds"*, Spotify. https://developer.spotify.com/documentation/web-api/reference/browse/get-recommendations/

Protocol

Finding the right Attributes

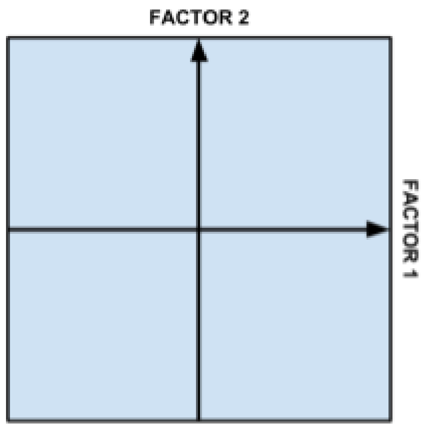
This survey is fully conducted online and without moderation. The user is given a link to the Google Form and performs the task individually. After gathering information for the user demographics the user in question is presented with the survey.
In a first step the user is asked to select 3-4 attributes which he/she would use to categorize music based on its mood.
In the second part the user is presented with three different songs and six potential attributes we consider for our future interface. The user then ranks each attributes with a value from 0 (extremely low, if not non-existent) to 10 (extremely high) that he thinks corresponds to the given song. This is repeated for each song, with the same attributes.



Axis Labeling

Our second study follows a classic protocol for User Study i.e. we invite the user, explain the tasks and what we try to find out. We then randomly choose a song to present to the user and let the user point on the Navigator where he considers this song fits in (based on the 2 given attributes). This procedure is repeated four times with different songs and different pairs of attributes.

In a second step the user selects a quadrant (among the four of the navigator) and we ask him to think about a song in his mind that would fit into this quadrant. Afterwards we present him with a song chosen by us as the Wizard of Oz (based on the chosen parameters) and



A simplified Navigator interface where we set Factor 1 & 2 to the attributes which we test the user on.

the user then rates the accuracy of the musical mood prediction. If the given song's mood setting matches that user's thought out song, we expect to get a high accuracy. If we provide a much different song than the user expected, he will rate it with a low accuracy.

Song Chosen? 1 *

- Call Me Maybe by Carly Rae Jepsen
- We are the Champions by Queen
- Bodak Yellow by Cardi B:
- Chariots of Fire by Vangelis

Song rating 1: Energy vs Happiness *

	Far Left/Bottom	Middle Left/Bottom	Slightly Left/Bottom	Middle	Slightly Right/Top	Middle Right/Top	Far Right/Top
Energy	○	○	○	○	○	○	○
Happiness	○	○	○	○	○	○	○

As the moderator we filled out this online survey with the data the user provided us in the moderated study. This allows us an easier evaluation after we conducted the study with multiple users.

Part B: Selecting mood

Think of POP SONGS ONLY!

How did participant select their Quadrant? *

	Top Left	Top Right	Bottom Left	Bottom Right
Pair A: Energy vs Happiness	○	○	○	○
Pair B: Speechiness vs Energy	○	○	○	○
Pair C: Tempo vs Happiness	○	○	○	○
Pair D: Happiness vs Speechiness	○	○	○	○

Song A Suggestions

Low Energy, Low Happiness: Shape of You Ed Sheeran
https://open.spotify.com/track/0FE9t6xYkqWXU2ahLh6D8X

Low Energy, High Happiness: Solo by Demi Lovato
https://open.spotify.com/track/6kPJZM97LwdG9QIsT7khp6

High Energy, Low Happiness: MIA by Bad Bunny feat Drake
https://open.spotify.com/track/116H0KvKr2Zl4RPuVBruDO

High Energy, High Happiness: Rise by Jonas Jack & Jack
https://open.spotify.com/track/3u1S1OmAUhx5DRlLrXqyp3

Participant's reaction to accuracy of chosen Pair A Song *

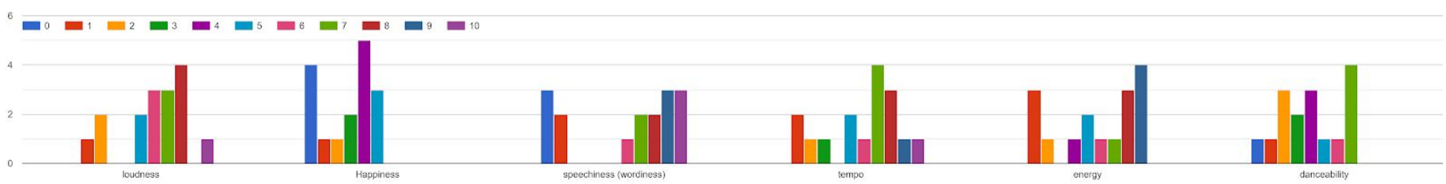
	1	2	3	4	5	
Not accurate at all	○	○	○	○	○	Very accurate

Results

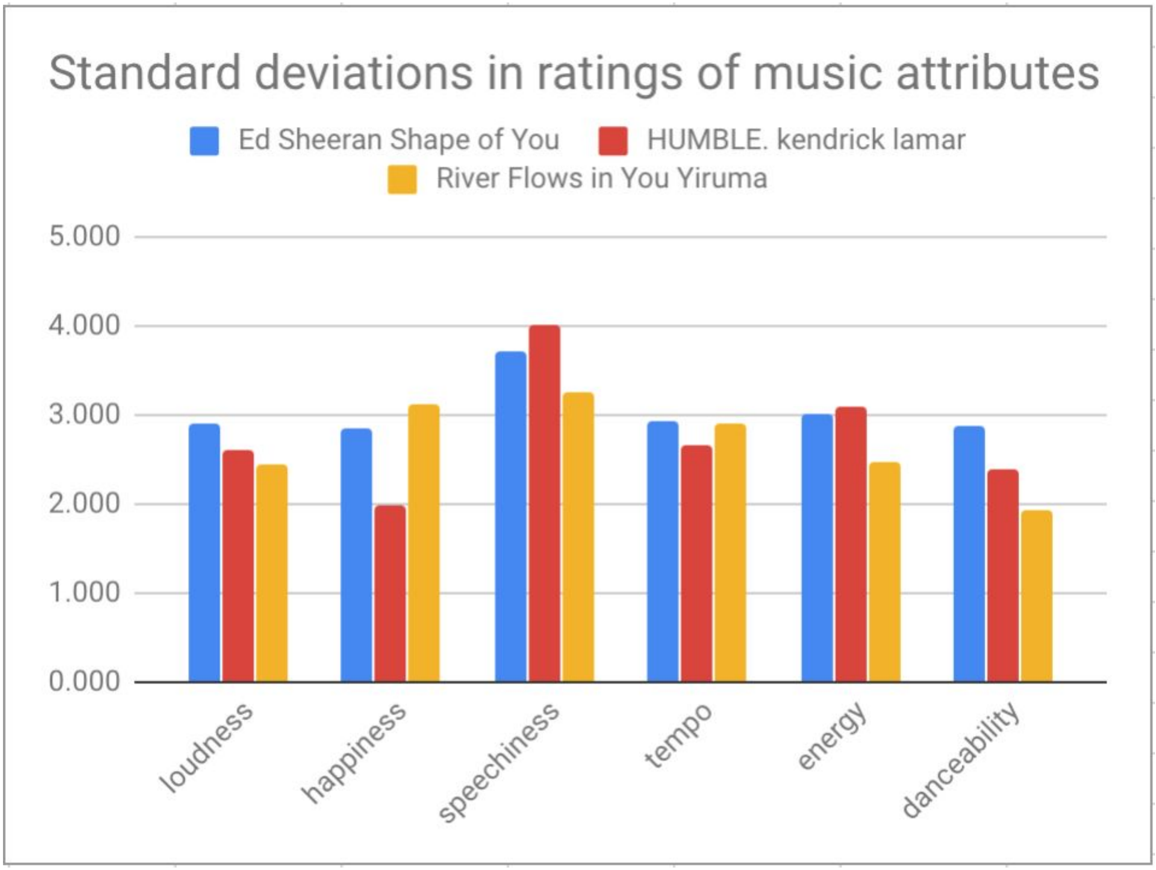
Finding the right Attributes

Based on the results of the initial survey, it was not clear if there was a subset of distinguishing characteristics that users would likely be able to agree upon what mood a music depicts. There was a wide spread in opinion amongst respondents.
For example for the song HUMBLE by Kendrick Lamar this were the results of the users rating (the scale of 0-10 could mimic the behavior of a one-dimensional slider)

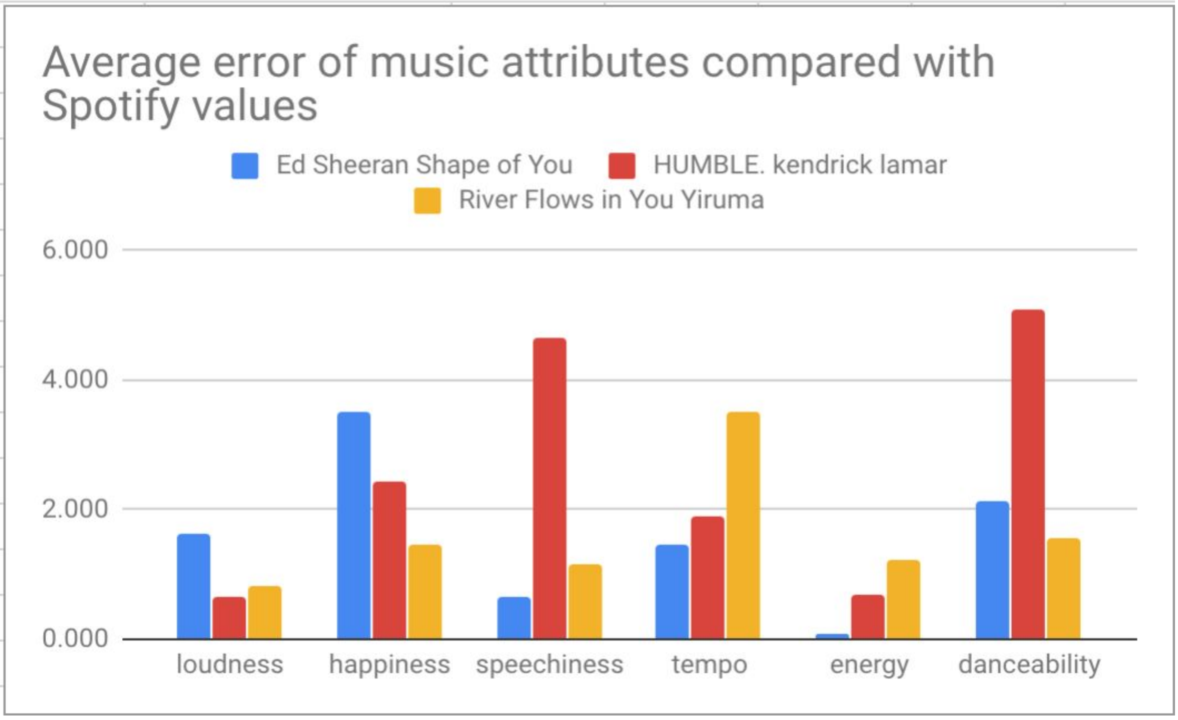
On a scale of 0-10, what do you think the attributes of this song are?



It was similar for the two other proposed songs of the study so in order to get more insight about the data we compared the standard deviations in the ratings of the music attributes :



When compared to the values we get from the Spotify API we get the following average absolute errors:



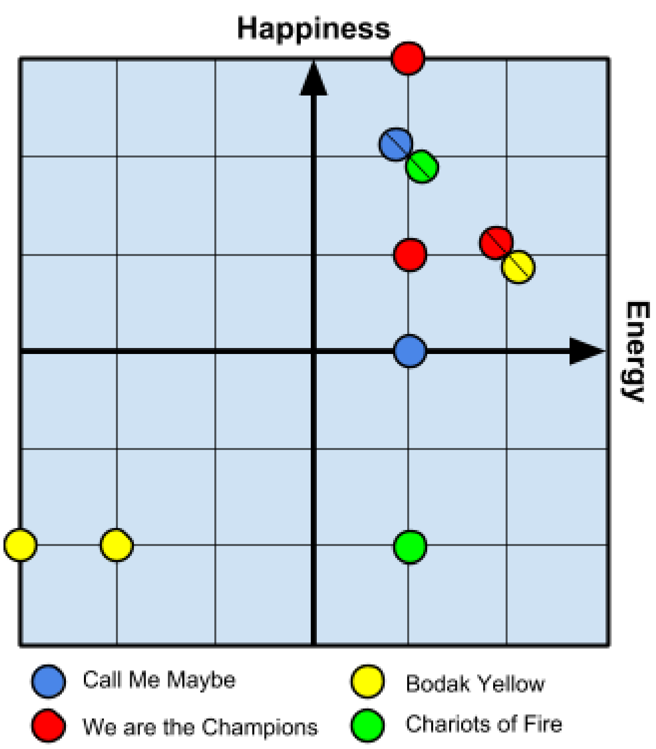
In terms of accuracy (compared with the Spotify-generated feature values, which we took as reference even though Spotify could be wrong but we have not found other ways to judge), loudness, energy and happiness topped the list in being consistent compared to user's understandings of the attribute words.

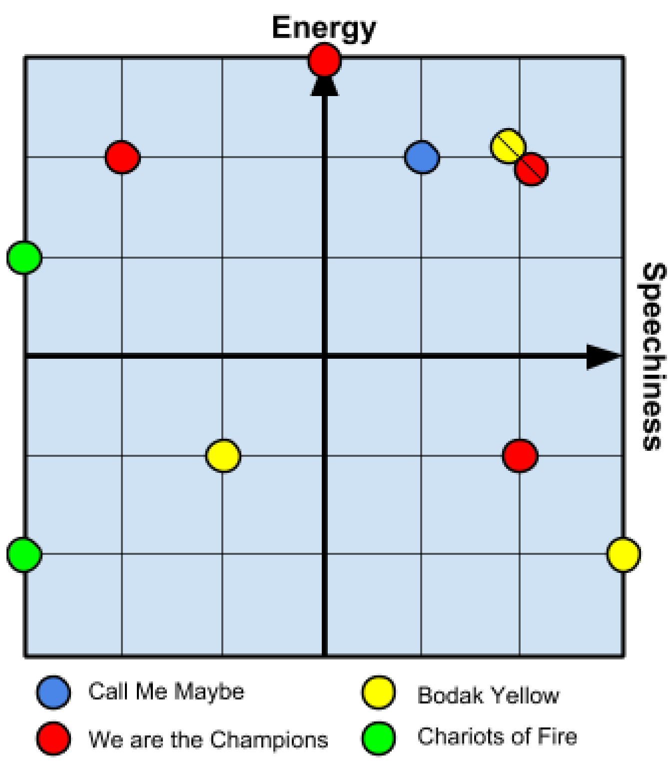
Considering the result of our studies and the finding from other research[2] we therefore considered the following four attributes as potential axis labels:
Energy, Happiness, Speechiness and Tempo.

Axis Labeling

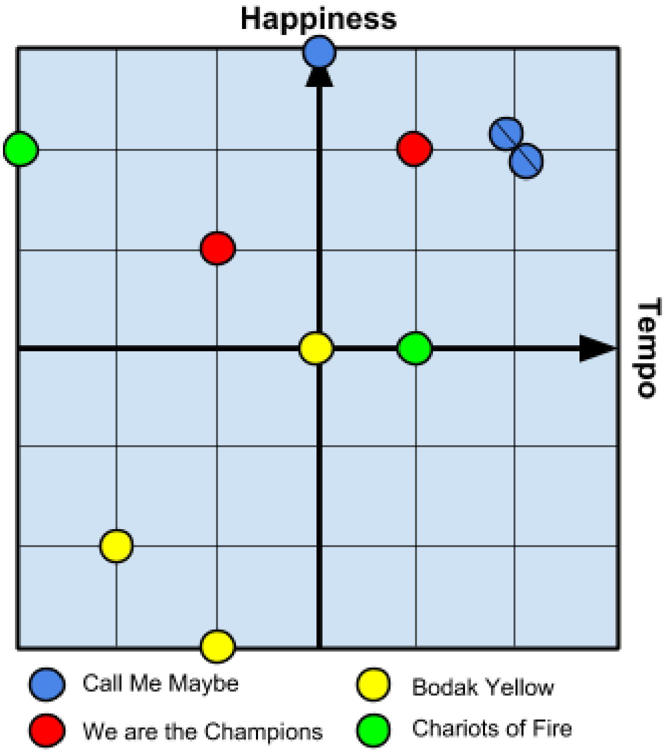
First Part

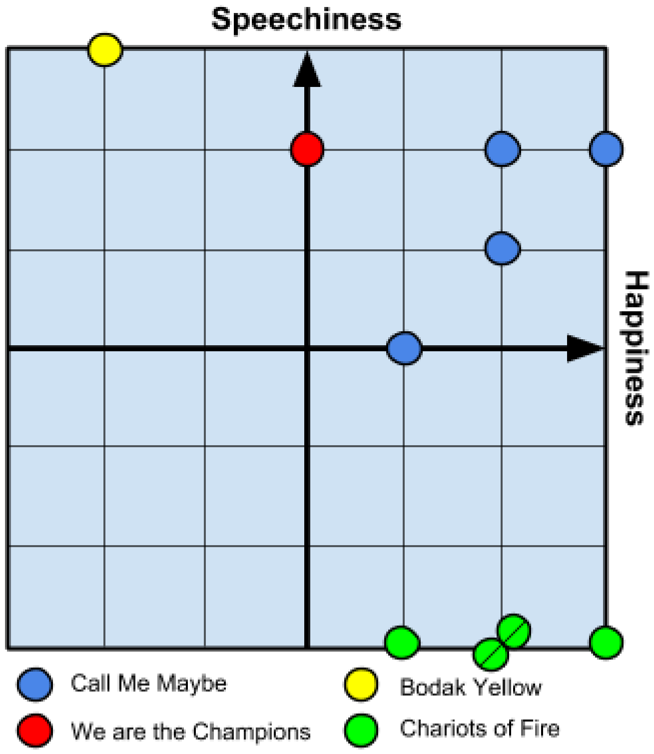
We made users listened to four songs in a random order and made them rate the songs on a two dimensional graph with randomly varying pair of axis.
The results we got are summarized in the 4 graphs below :



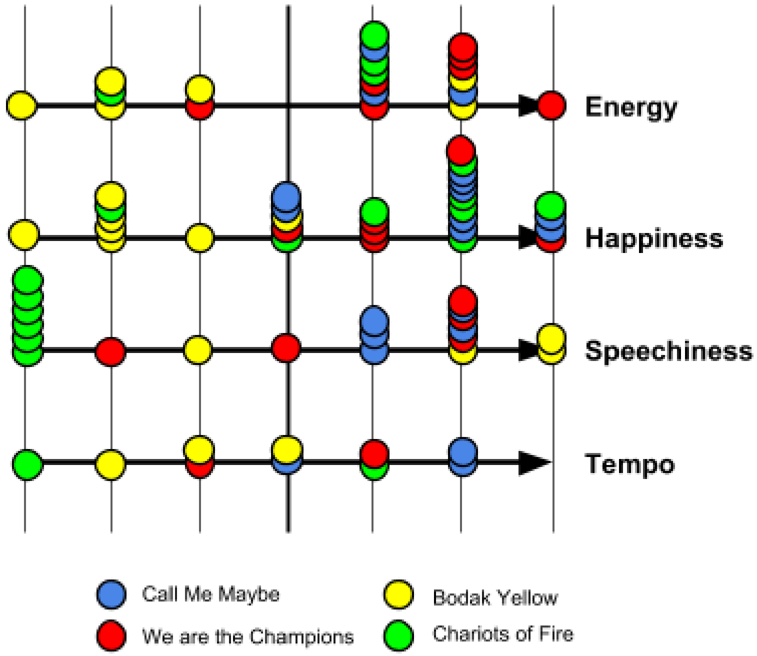


[2] Nuzzolo, Michael. *Music Mood Classification.* https://sites.tufts.edu/eeseniordesignhandbook/2015/music-mood-classification/

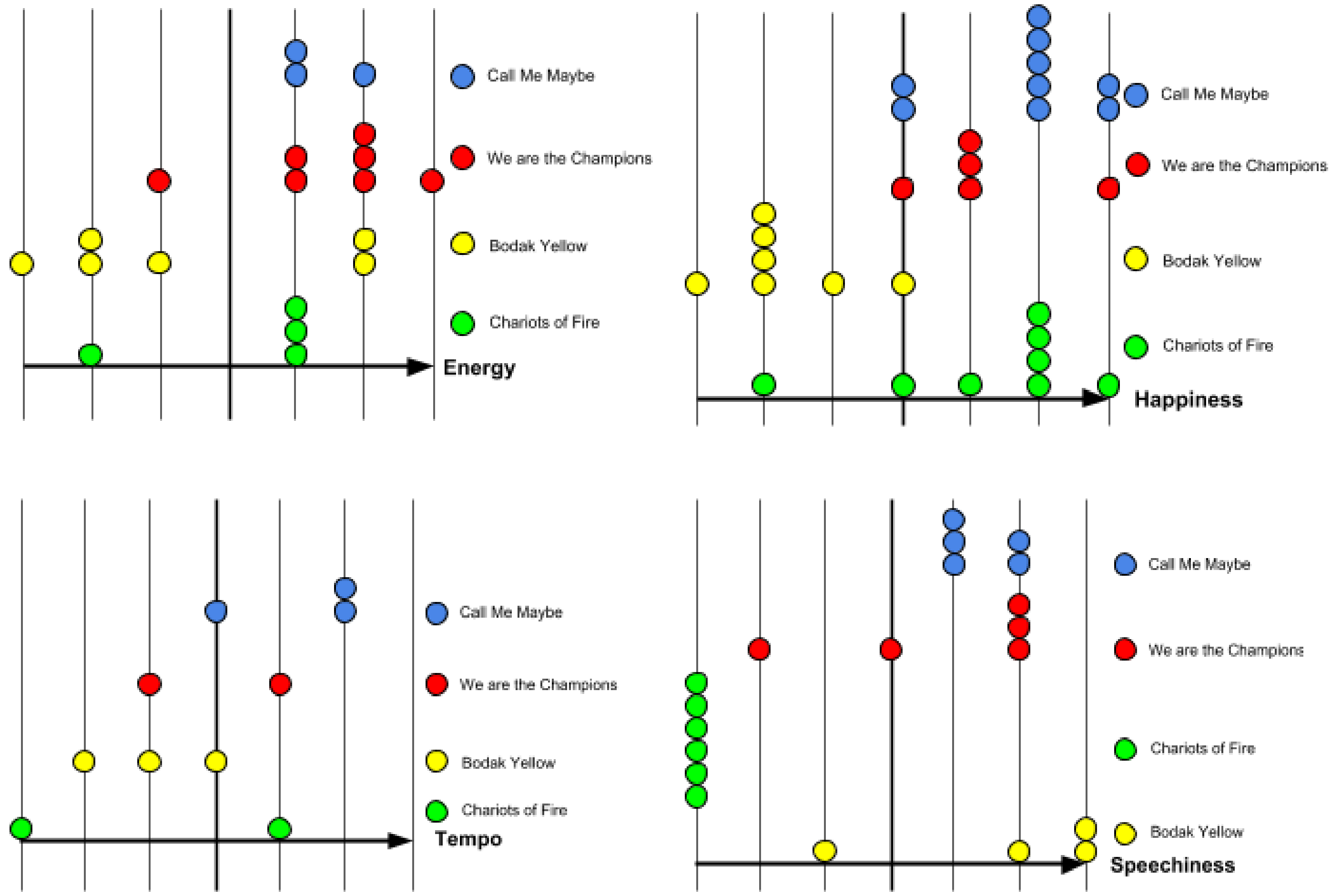




Depending on the axis pairs and the song we see that some rating coordinates are closer to each other than more spread out others.
We then asked ourselves how each individual attribute performed, we thus examined each attribute when we collapse the data into 1-dimensional scale :



If we focus on individual attributes we get :



With those graphs we can see that, where we have enough data, users tend to agree more on where to rate a song according to a given attribute. We have less dispersion here (when we used a 2 dimensional graph as the rating technic) than in the first study where the rating technic mimicked a 1-dimensional slider.

Second Part

We asked the users to think of a pop song and then pick the quadrant where they think it resides on each of the four axis pairs

- **Happiness vs Energy (Pair A)**

Spotify Algorithm Suggestions
Low Energy, Low Happiness: Shape of You Ed Sheeran
Low Energy, High Happiness: Solo by Demi Lovato
High Energy, Low Happiness: MIA by Bad Bunny feat Drake
High Energy, High Happiness: Rise by Jonas Jack & Jack
Average similarity score: 3.1/5.0

- **Energy vs Speechiness (Pair B)**

Spotify Algorithm Suggestions

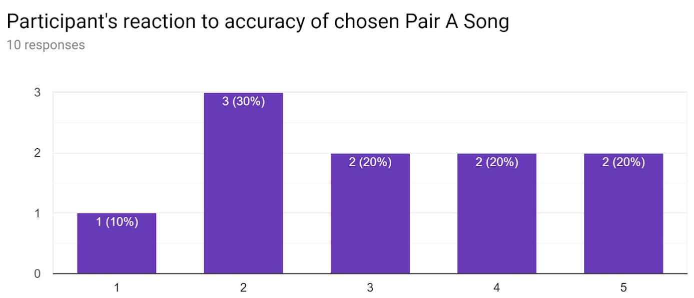
Low Speechiness, Low Energy: Without Me by Halsey
Low Speechiness, High Energy: Happier by Marshmello, Bastille
High Speechiness, Low Energy: One I want by Majid Jordan
High Speechiness, High Energy: Panda by Desiigner
Average similarity score: 3.4/5.0

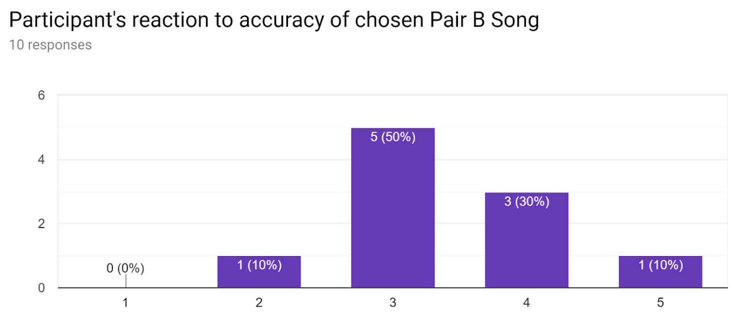
- **Happiness vs Tempo (Pair C)**

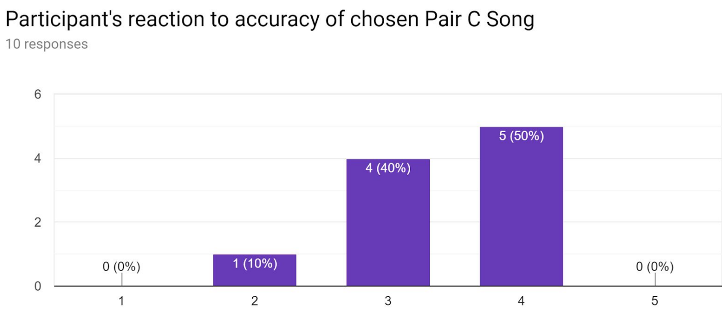
Spotify Algorithm Suggestions
Low Tempo, Low Happiness: Lucid Dreams by Juice WRLD
Low Tempo, High Happiness: Sunflower - Spiderman
High Tempo, Low Happiness: Rockstar by Post Malone
High Tempo, High Happiness: Mala Mia by Maluma
Average similarity score: 3.4/5.0

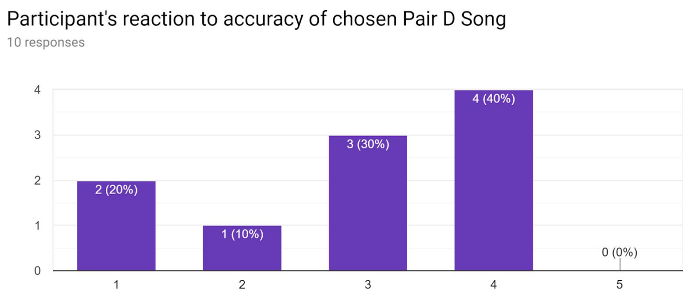
- **Speechiness vs Happiness (Pair D)**

Spotify Algorithm Suggestions
Low Happiness, Low Speechiness: When the party's over by Billie Ellish
Low Happiness, High Speechiness: Youngblood by 5 Seconds of Summer
High Happiness, Low Speechiness: Your Song By Ellie Goulding
High Happiness, High Speechiness: Back on Road by Gucci Mane
Average similarity score: 2.9/5.0









The average similarity scores are all around 3 but the pairs Energy vs Speechiness and Happiness vs Tempo performed better with 3.4

Implications

Based on the results of our first studies and findings from other research we concluded that the factors Energy, Happiness, Speechiness and Tempo might be the most descriptive of a music's mood when a user rate songs. And with this in mind we wanted to find out which factor pairs is the most accurate when used as axis labeling to associate a song with a mood setting. Considering the result of both our studies we find the pair Energy and Happiness as the most promising. Nonetheless we think mood related words alone might not be the best solution for a user to find the music he wants based on the mood settings he inputs.
Therefore in order to further improve our design and restrict the user frustration of not getting the songs he would expect based on the mood he sets we thought about giving him instant preview of the type of songs to expect given the mood he inputs or maybe ask for more user inputs like a popover that would propose 3 genres for a given mood setting (i.e. for low happiness, low energy it could propose instrumental, chill or gothic).
Therefore we will use our results to set up our Navigator interface with a Happiness and Energy axis but we will also consider further options. The most relevant option will be to directly influence the genre. Since the genres themselves are very different and we haven't studied in depth songs from different genres in our studies, we will consider adding an additional option to set the genre.
Using direct previews or song suggestions on the interface we will be able to solve potential misunderstanding or a different interpretation of the attributes. Therefore the user is directly confronted with a song and can better associate the chosen attribute values with the song he/she will get. This allows a more precise choice for future songs.
We can even add the Slider interface as an optional interface that allows experienced users to precisely set the mood and genre of the songs they are looking for.