Comparing Human Perception Of Song Similarity With ML Models

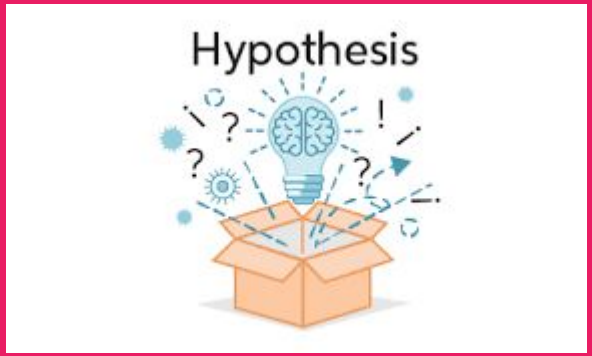
Presented By:
1. Pratyaksh Gautam (2020114002)
2. Harshit Gupta (2020114017)

OBJECTIVES

1. Identify and analyse the model's benefits and drawbacks in processing musical data, especially in comparison to human perception.

2. Examine any differences in how songs are perceived by musicians and non-musicians, particularly in terms of the timbre elements that they emphasize.

Hypothesis



We have no strong hypotheses because this is an exploratory study. We anticipate that music that the model deems similar will be perceived similarly by the majority of humans.

At the same time, we anticipate some flaws in the model, particularly when compared to listeners who are musicians.

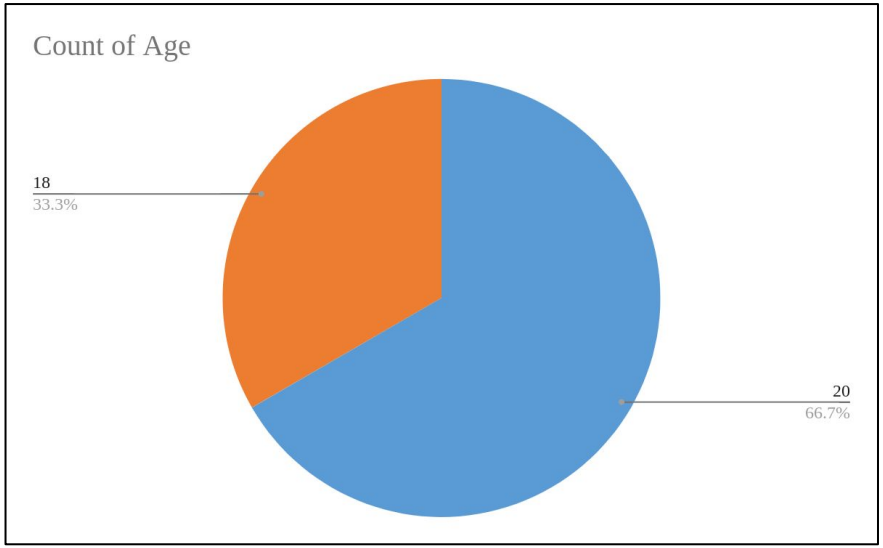
Methodology

1. Used an anonymous poll to gather information from participants.
2. To test the feasibility of the study, a pilot test of the form was administered to 2 individuals. **The final study did not involve the pilot testers.**
3. As a result of time constraints and the effects of listening fatigue, we reduced the number of song pairs from the original 12 to 6 during the pilot testing phase.
4. We have revised the explanations of important terminology to make them comprehensible.

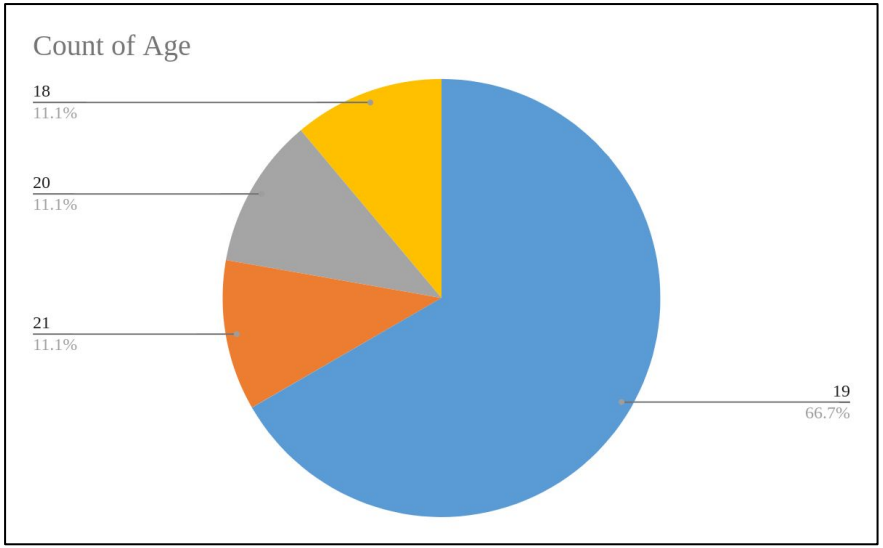
1. The dataset we retrieved the music pairs from for our investigation was **MagnaTagATune**. Six music pairs—two that are quite similar, two that are moderately similar, and two that are dissimilar—were kept in the sample at our discretion.
2. **Purposive sampling** (also convenience) was used to ensure a 50/50 balance between musicians and non-musicians, but we also gave participants the option to self-identify as either.
3. In order to compare the song similarity ratings between these pairings, we worked with the musical data using ML models from **MusiCNN**. The code file for the same is: https://colab.research.google.com/drive/1ldz20zTumkTYUokKLgNx6Q8ffLiRwg1J?usp=sharing. We extract the song features and compute the **cosine similarity** between them.
4. For our study, we obtained ordinal data ratings using a **5-point Likert Scale**. For the purpose of comparing them based on similarities, we classified these as unipolar and bipolar in terms of their emotions.

Sampled Demographics

1. Age



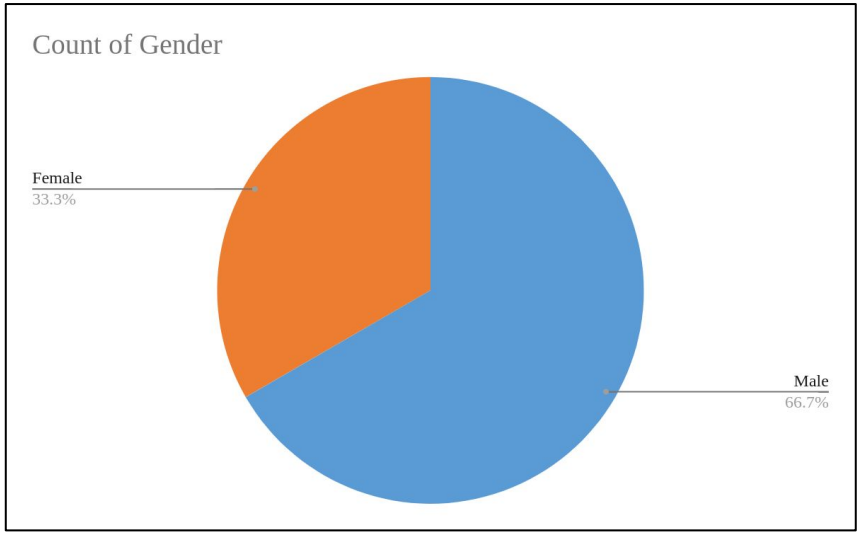
Form 1

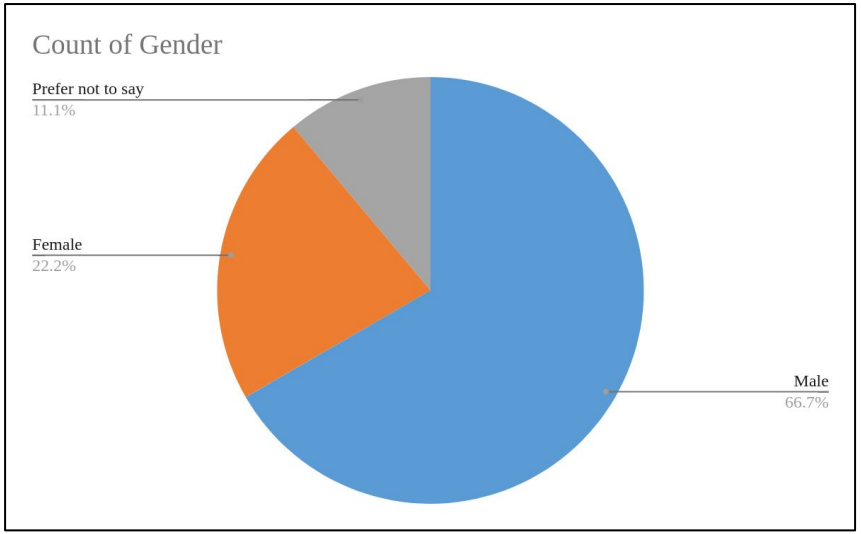


Form 2

Sampled Demographics

2. Gender



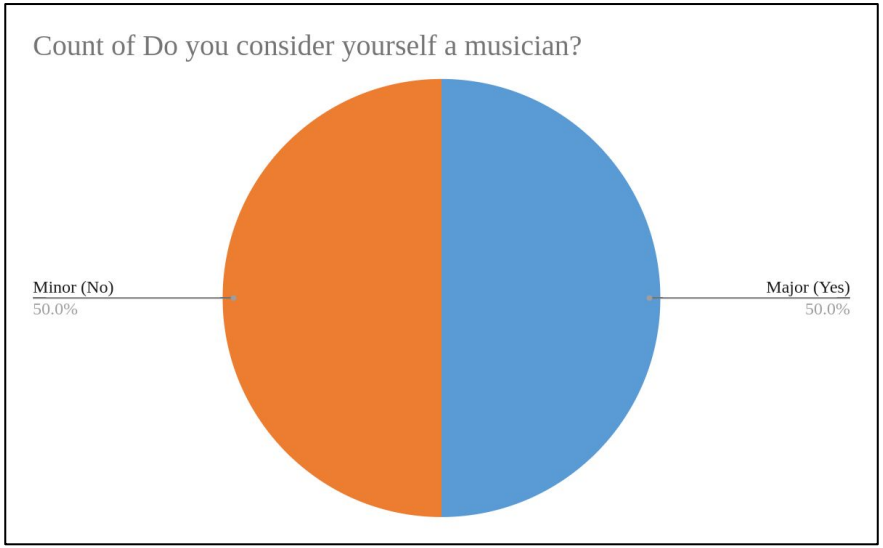


Form 1

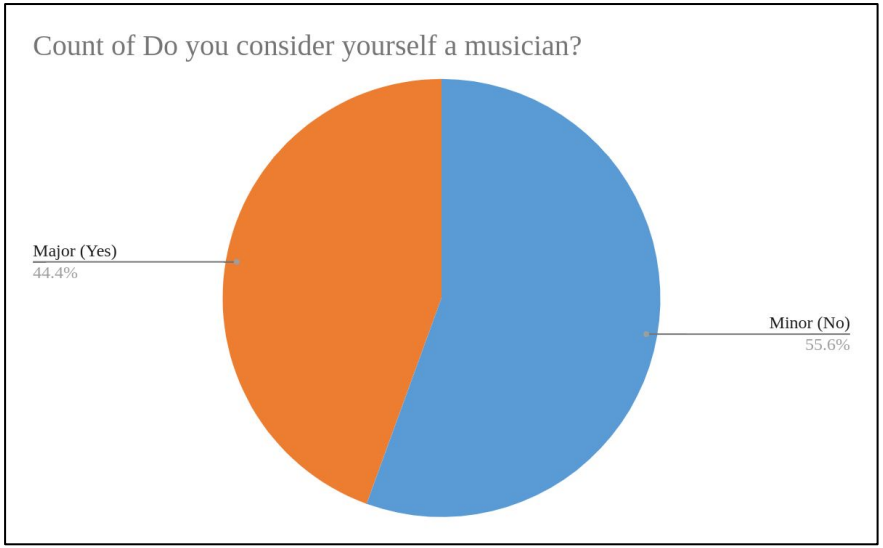
Form 2

Sampled Demographics

3. Do you consider yourself a musician?



Form 1



Form 2

Experiment Design

Link of the Form: https://forms.office.com/r/upgM26EaFz

Comparing Human Perception Of Song Similarity Against Deep Learning Models

The completion of this experiment takes 25 to 30 minutes. The following guidelines must be followed:
1. Find a peaceful spot to sit.
2. Before beginning the experiment, put on some headphones or earphones.
3. You can listen to the sound clip as many times as you like.

...

* Required

Demographics

Your personal information will only be collected and used for training and analysis. Before moving on to the next stage, anonymity will be guaranteed for all participant-sensitive data.

Key Instructions

The key terminology involved and their definitions are mentioned below. Please open this link in a new tab if you would like to refer to the terms repeatedly while participating in the experiment.

https://docs.google.com/document/d/1vrqqUaPzIE9357pEHR3ZYJapunlPvezJaIc_Wv6p5Eo/edit?usp=sharing

Definitions:

- **Melody: The main tune of the song. It is the memorable part of the song, which you can hum along to and remember afterwards.**
- **Rhythm: The underlying beat of the music. It refers to both the speed, or tempo, of the music, as well as how the beats are organized: for eg. in repeating groups of 3 or 4.**
- **Harmony: The other tunes in the song that support the melody. This includes things like the notes and chords played on a guitar while a vocalist sings a melody.**
- **Timbre: The element of the songs other than melody, rhythm and harmony. The 'sound color'; it is what allows you to tell when the same note is played on a saxophone vs when it is played on a violin.**

Back Next Page 2 of 9

Comparing Human Perception Of Song Similarity Against Deep Learning Models ...

* Required

Music Pair 1

You must evaluate the two musical compositions' similarity based on the many criteria listed in the link. **Please DO NOT pay attention to the lyrics in these clips.**

YOU CAN REPLAY THE CLIPS AS MANY TIMES AS YOU LIKE :)

Music Clip 1
https://drive.google.com/file/d/1p9HVDrEG6rMW6dW_IoE7iN1jXRmv01K6/view?usp=share_link
Music Clip 2
https://drive.google.com/file/d/1imgS4Imn8I_56aQJv7cBIG8vCmSAsnSL/view?usp=share_link

5. How much are the two musical compositions alike? *

	Not Similar		Neutral		Similar
Melody	○	○	○	○	○
Rhythm	○	○	○	○	○
Harmony	○	○	○	○	○
Timbre	○	○	○	○	○
Overall (Total)	○	○	○	○	○

6. Look for abstract emotions in the **first musical clip. Now,** evaluate the abstract emotions in the first musical clip onto the scale. *

	Softness/ Sadness/ Soothing		Neutral		Hardness/ Happiness/ Exciting
1. Softness-Hardness	○	○	○	○	○
2. Sadness-Happiness	○	○	○	○	○
3. Soothing-Exciting	○	○	○	○	○

7. Look for abstract emotions in the **second musical clip. Now,** evaluate the abstract emotions in the first musical clip onto the scale. *

	Softness/ Sadness/ Soothing		Neutral		Hardness/ Happiness/ Exciting
1. Softness-Hardness	○	○	○	○	○
2. Sadness-Happiness	○	○	○	○	○
3. Soothing-Exciting	○	○	○	○	○

8. How else do you think the two musical pieces compare or contrast? Why do you think they are similar or dissimilar?

Enter your answer

Back | Next | Page 3 of 9

Section 9

Thank You For Participating!

We wish you a great day ahead :)

+ Add new

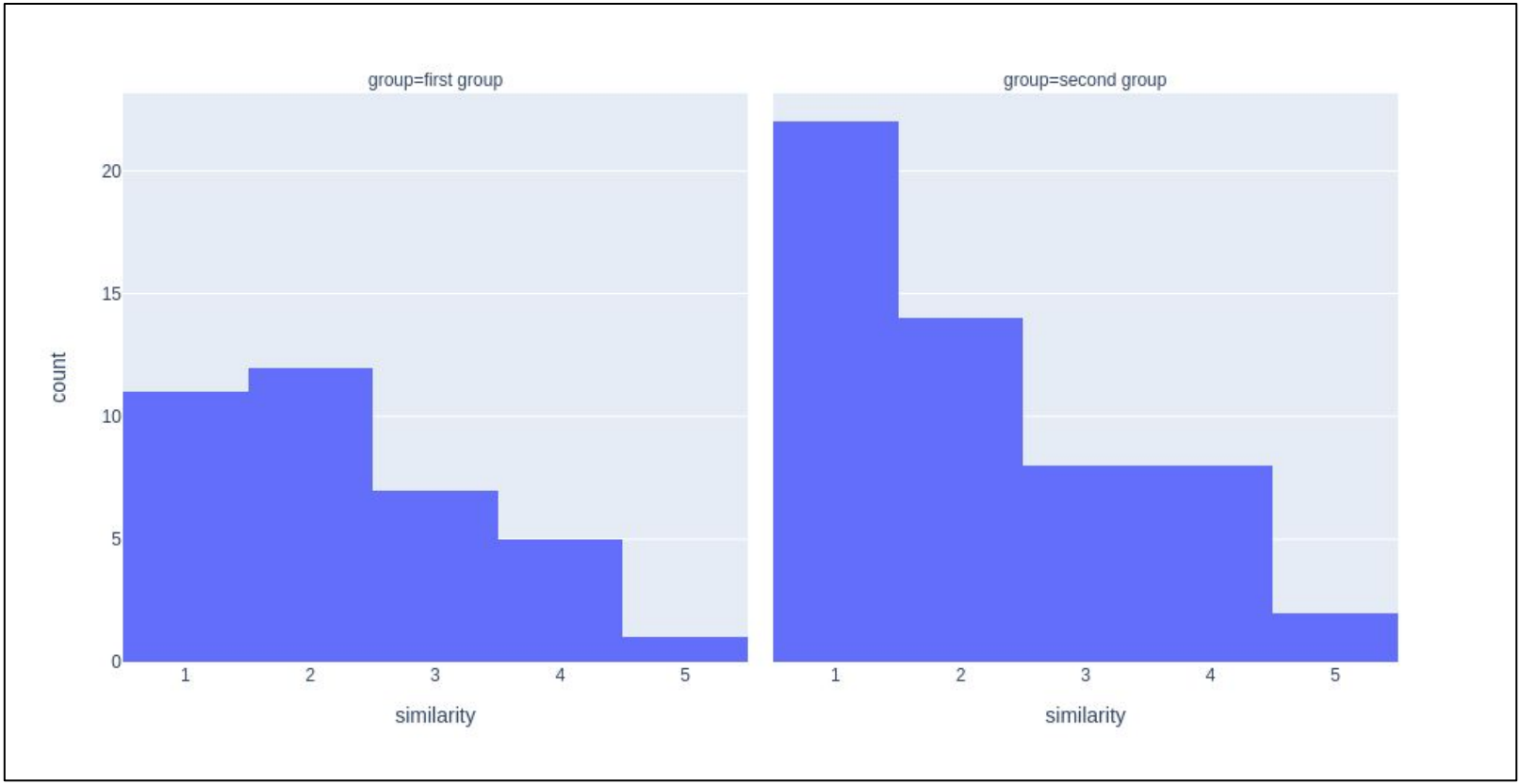
Variables

Independent Variable:
Whether the rating received is by a model or a person and musician or non-musician in case of a person.

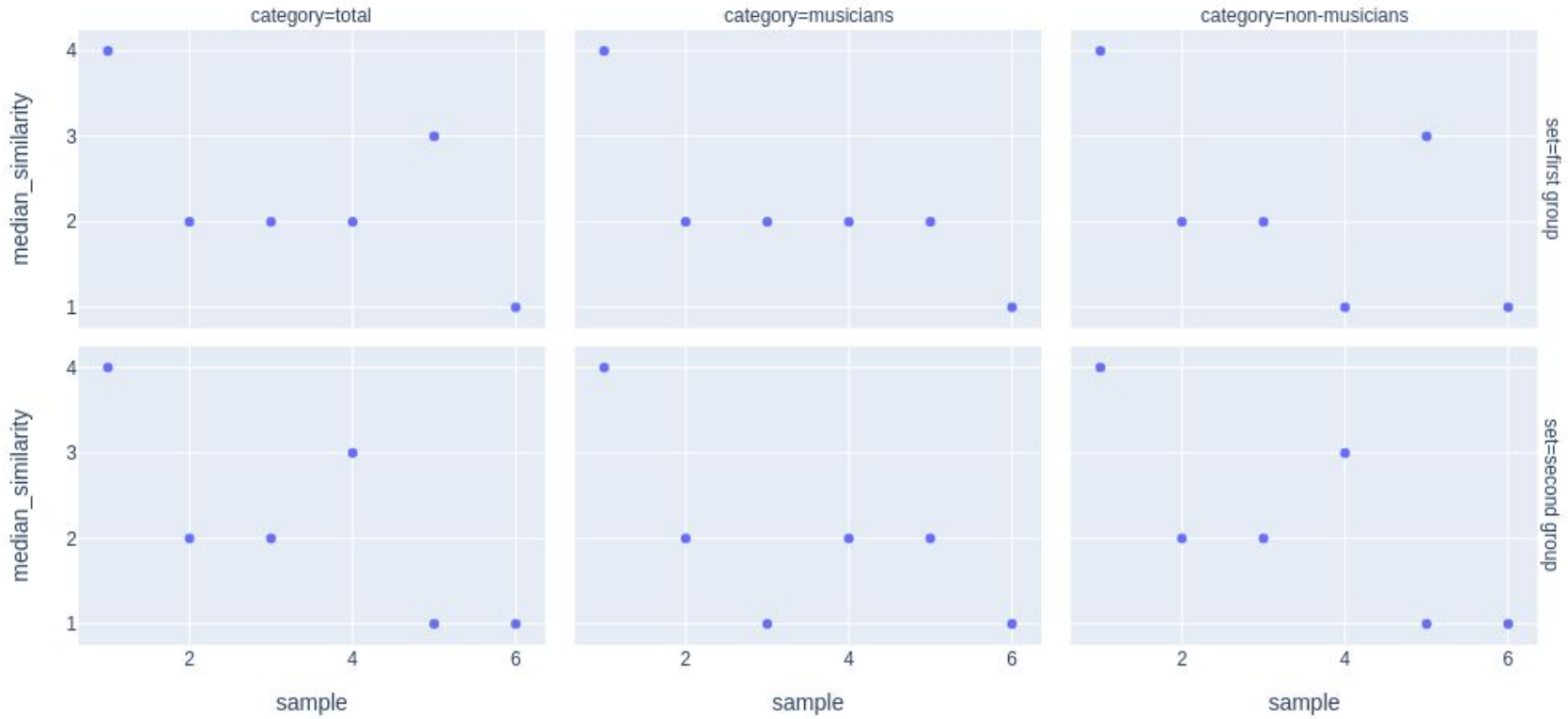
Extraneous Variables:
1. Listening habits of the subjects.
2. Listening apparatus used by the participant.
3. Genre and other abstract features of the musical piece.
4. Demographics of the participants.
5. Environment surrounding the participant.

Dependent Variables:
How similarly do they rate two songs.

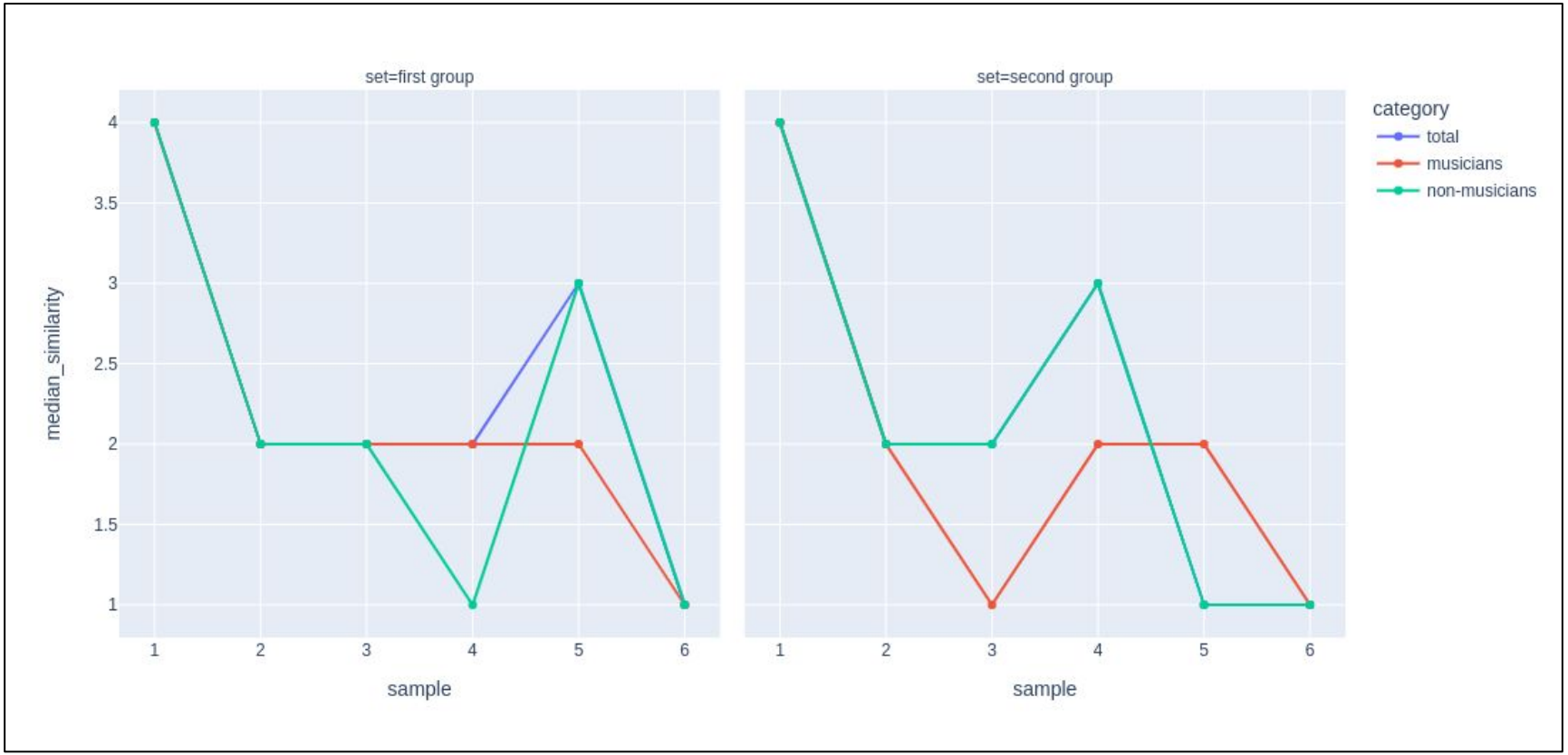
Results and Analysis



Since the data is ordinal, we do not expect it to be normally distributed. Thus, we cannot apply parametric tests to it. We note that the second group's data is a bit more skewed to dissimilarity.

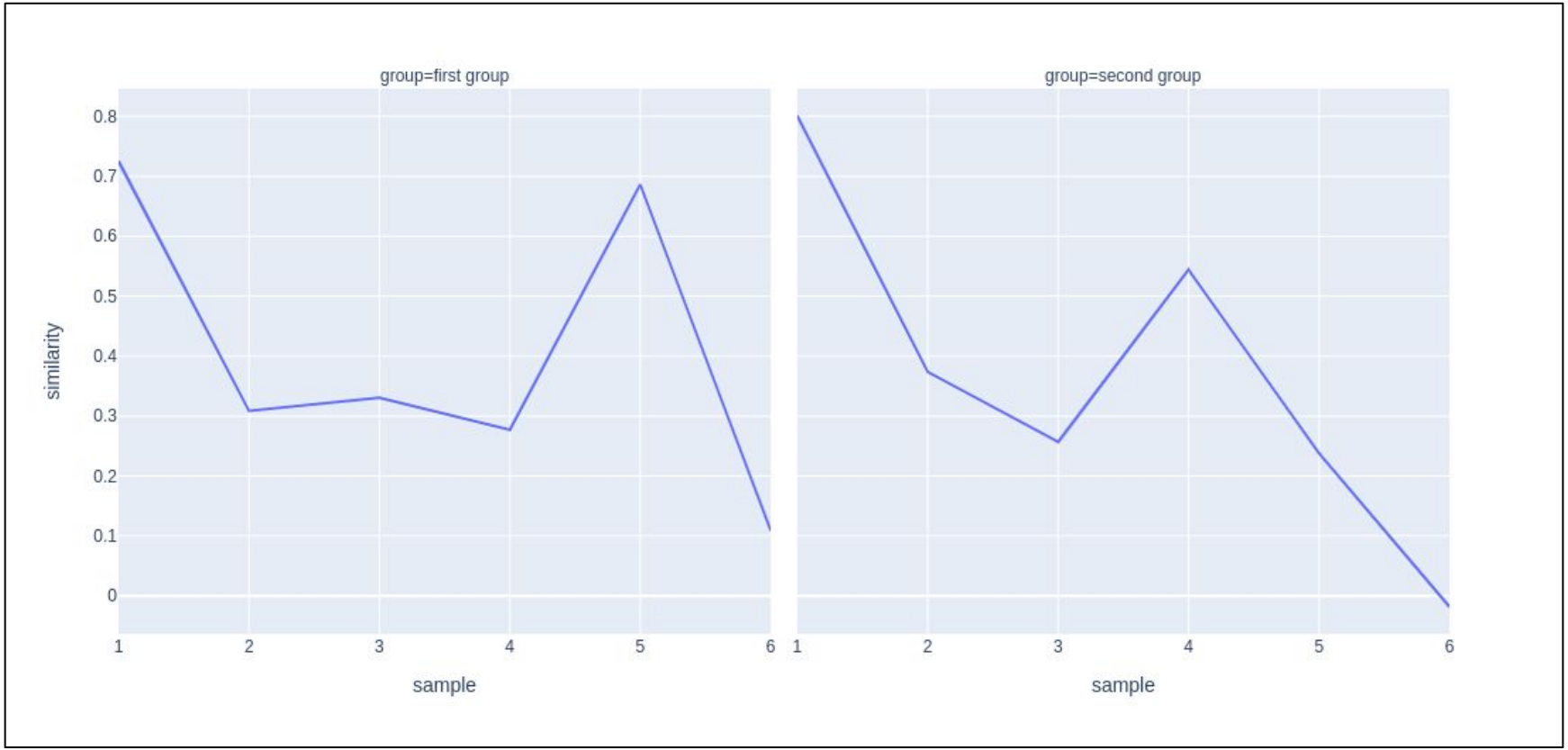


Keep in mind that the first and second group rated on different samples entirely, and thus should not be compared directly.



Total aligns perfectly with non-musicians in the second group.

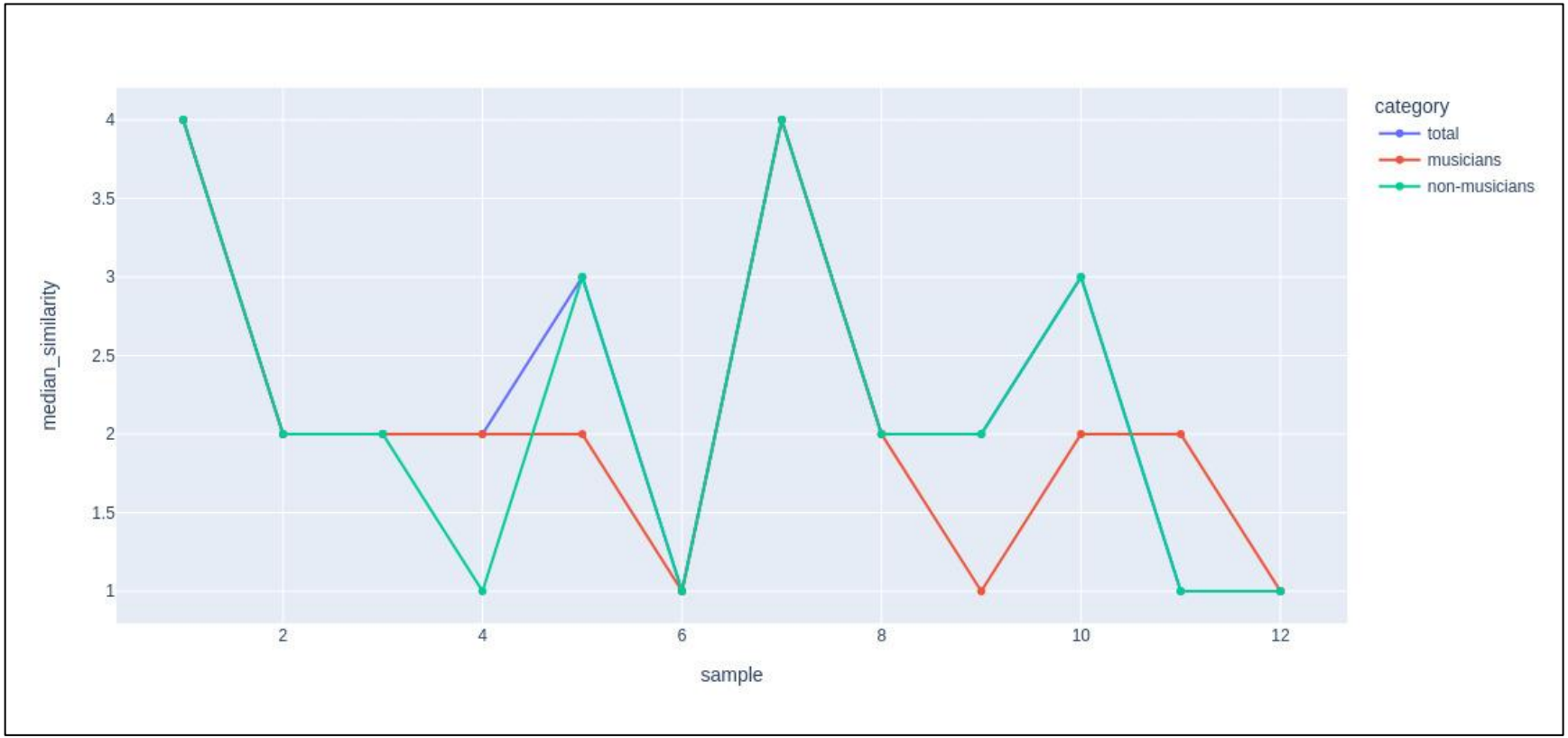
The musicians all agreed that Sample 3 in Form 2 had a timbre similarity of 1.



We conclude that the model gives similar similarity scores just as how the human participants do for the song pairs.
Link to plots: https://colab.research.google.com/drive/1RURL6U4GgF9WY9FJSZaXCfH_NvzNRLEh?usp=sharing

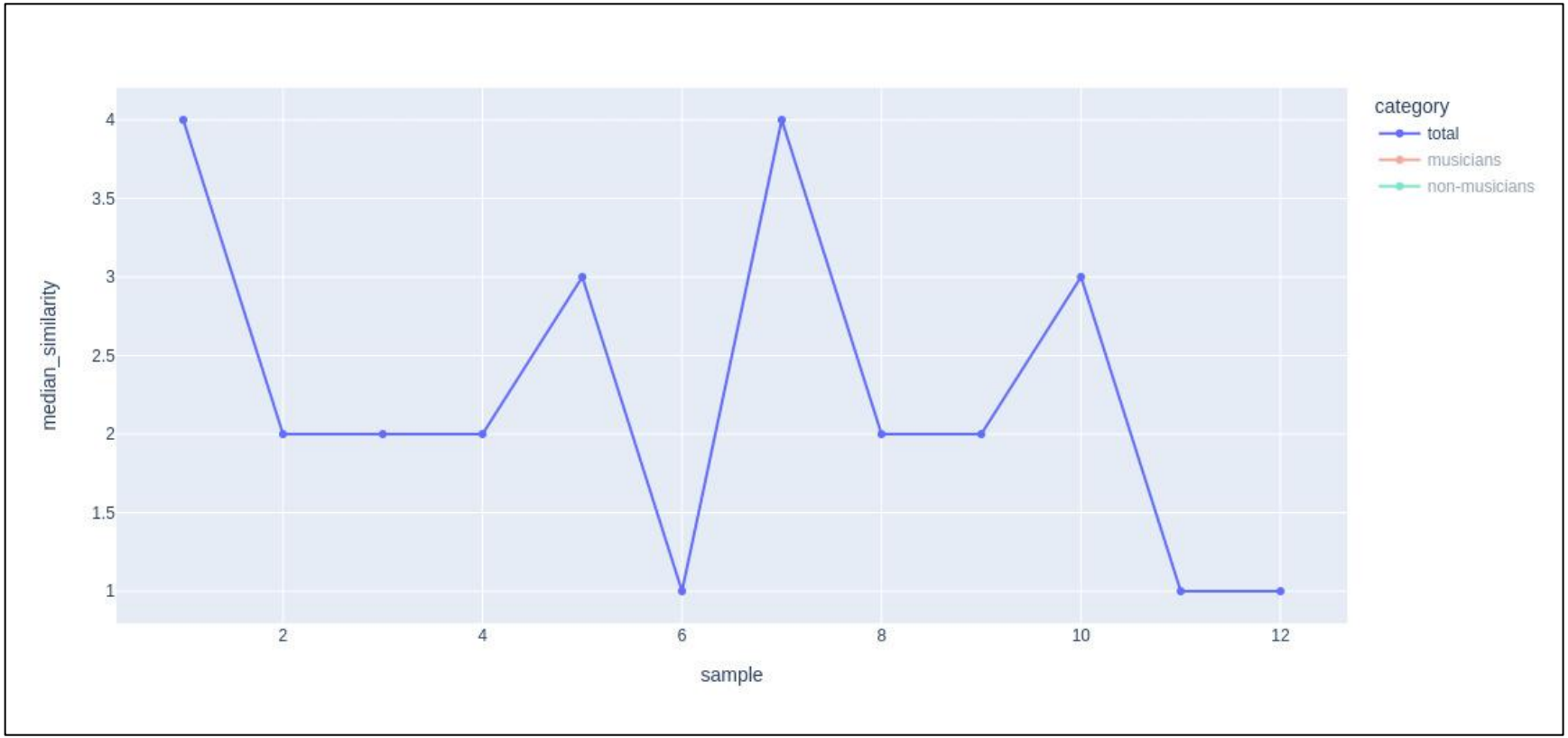
ADDITIONAL SLIDES

Median Similarity Scores (Musicians, Non-Musicians, Overall)



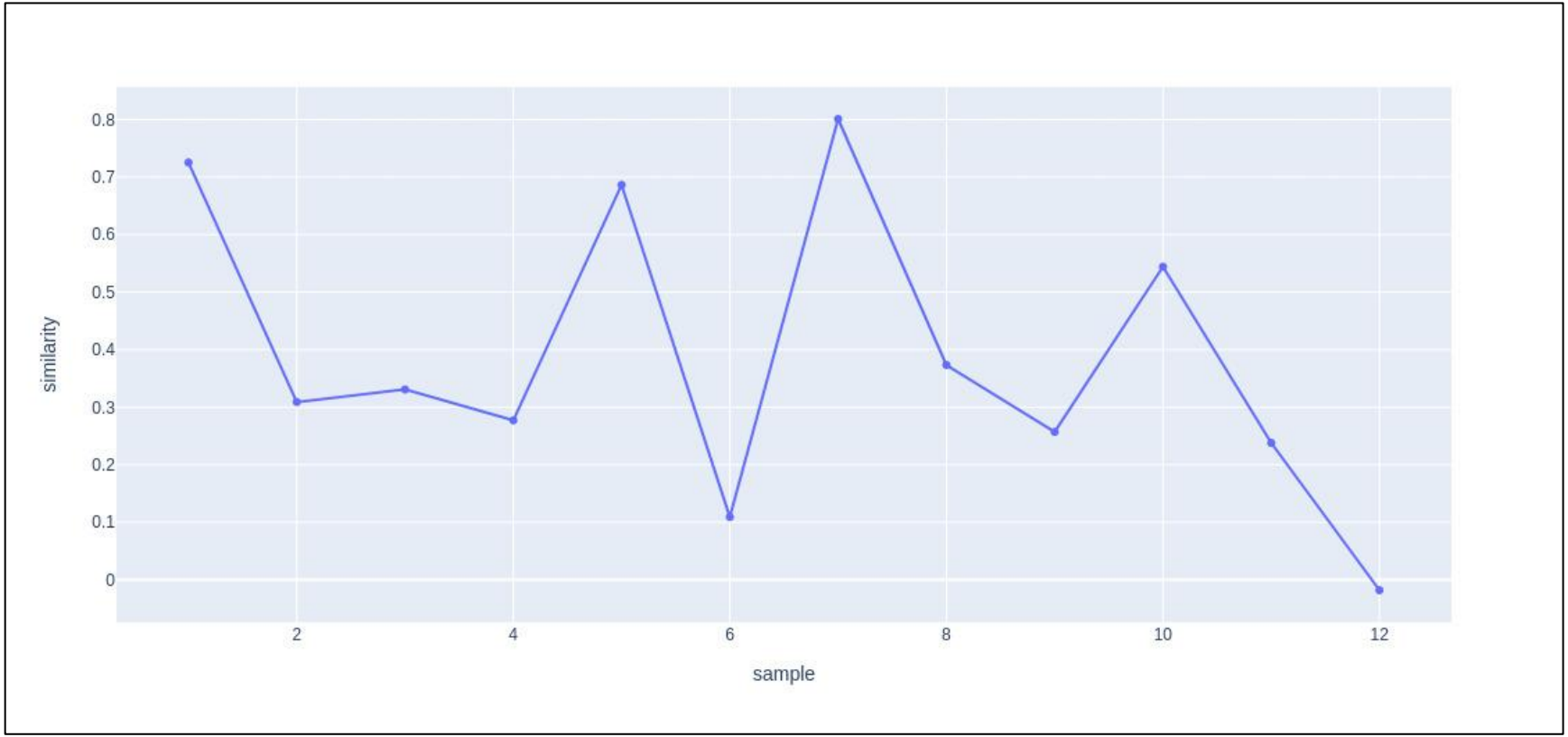
We observe that compared to musicians, the non-musicians curve is inline with the overall scores curve. Indicates that certain criteria which musicians judge the samples is not being accommodate for by the model.

Total Median Similarity Scores



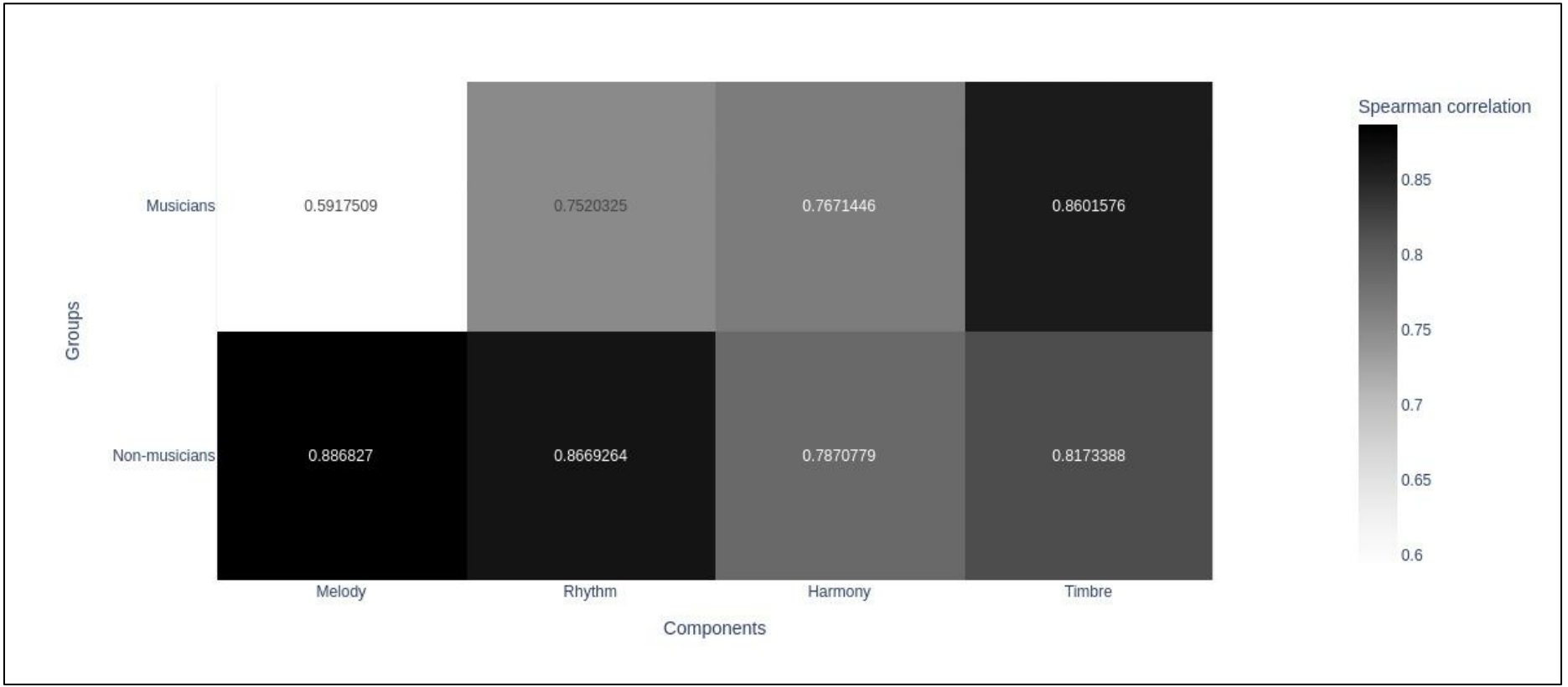
Total median similarity scores accommodates for both musicians and non-musicians. Link to plots: https://colab.research.google.com/drive/1RURL6U4GgF9WY9FJSZaXCfH_NvzNRLEh?usp=sharing

Similarity Scores As Given By The MusiCNN Model



The curve of both the graphs is similar! We conclude that the MusiCNN model gives similar similarity scores just as how the human participants do for the song pairs.

Correlation between overall perceived similarity and a similarity of a specific component in the song pairs, by Musicians and Non-Musicians



Analysis Of The Above Correlation Plot

For both of the data collection forms, we plot the correlation scores shown above. These results were calculated using the median ratings for each track. We used **Spearman Rank Correlation** which is a non-parametric test that is used to measure the degree of association between two variables.

1. The relationship between melody similarity and perceived overall similarity was strongest among non-musicians.
2. At the same time, musicians tend to have relatively little association between melodic similarity and overall similarity.
3. Musicians tend to have strongest association between timbral similarity and overall similarity.

Conclusion

The MusiCNN similarity ratings align very well with the median perceived similarity of the song-pairs by human participants. This holds across a diverse set of genres and listeners who are musicians and nonmusicians.

Drawbacks

1. We initially split our sample into 2 forms to allow us to go through a greater variety of songs. But, it ended up having a negative effect of greatly reducing our sample size.
2. Song similarity as annotated by human raters is an ordinal measure. However, that from MusiCNN is an ratio measure. It is therefore difficult to interpret their alignment outside of a qualitative assessment.
3. Low generalizability due to concentrated age range of sample and convenience sampling.

Some Interesting Comments

Musicians

“Pitch of the percussion instrument in second clip is lower, and there are gaps in the percussion instrument.”

“Rhythm emphasis is very similar - stress on the 1st and 4th beat was the same
But objectively the speed is different, so ranks low on the rhythm similarity. The ethos is also difference - if not for the vocal portion, then the vibe would be extremely different.”

“The first has a fixed andante rhythm, the second had variations. The first was completely major with no accidentals, the second had a few.”

Non-Musicians

“The first one felt mysterious and second one felt like a bgm of a villain.”

“First one was a peaceful happy type, reminding of snowy areas. The second was exciting happy type ”

“The emotions are very alike yet different. First is joy whereas the second is kinda lovey.”

“First clip felt a bit dystopian , regret and sad. The second one felt bittersweet i guess, felt peaceful sad n happy, not sure but was calming. ”

Thank You! :)
Any Questions?