Comparing Human Perception Of Song Similarity With ML Models

Harshit Gupta (2020114017)
Pratyaksh Gautam (2020114002)

Background

What do we mean by song similarity?

Two pieces of music can be considered similar on the basis of various factors such as:

1. Melody
2. Rhythm
3. Harmony
4. Timbre

And more abstract factors as well such as:

1. Energy
2. Repetitiveness
3. Warmth
4. Anger
5. Softness

Music Models

MusiCNN, which is pronounced "musician," is a collection of pre-trained deep convolutional neural networks for music audio tagging.

We can use the encoder's feature vectors and a measure like cosine similarity to obtain numerical similarity ratings between songs.

Motivation

We wish to discover the model's weaknesses in order to increase our understanding of music perception and design better structures for it. It also allows us to gain insight into the parameters that must be considered while building these models rather than simply employing them blindly.

Objectives

We aim to identify and analyse the model's strengths and weaknesses in understanding and dealing with musical data, particularly when compared to human perception.

We'd also like to discover if there are any differences in song perception between musicians and non-musicians, particularly in terms of which components of timbre they emphasise.

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.

Experiment Design

Sampling

We intend to use purposive sampling to get a sample of ten musicians and ten non-musicians.

We will also sample music data across genres, time periods, languages, tempo, instruments played, and so on.

We use simple random sampling for the music groups.

Methodology

We want to conduct the research as a survey using a thorough web form. The form will include song pairs and questions asking the subject to rate the similarities of the two songs on several scales such as energy, emotion, harmony, and so on. We will also want their broad reasons for the same.

We begin the experiment with 2-3 volunteers in a controlled quiet environment. This will allow us to foresee and address any design flaws before releasing the survey to a bigger audience.

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.

Thank You!

Questions?