**LITERATURE REVIEW**

Review of literature is based on the previous work on Affective Algorithmic Music Composition, more specifically to works that process music in symbolic form in order to generate music with a given emotion. Generating emotionally relevant music for videos and books is a problem that people have attempted to solve in different ways. A common approach for this problem consists of designing a rule-based system to map musical features to a given emotion in a categorical or dimensional space [1]. Steve Rubin and Maneesh Agrawala [2] proposed a system that would resequence music files in order to create a soundtrack that conveys the emotion perceived from an audio book. Their method requires premade music files for different emotions and worked best when the transcript of the audio book was labelled by humans.

**1.1 Sentiment analysis of the audio**

Sentiment analysis is a topic that has been worked on extensively over the past few decades. Most of the work in it focuses on textual data for analysis. This method of analysis may not always accurately recognize the context of the situation described in the text (sarcasm, joy, sadness, etc). Sentiment analysis on audio introduces additional features such as tone, pitch, timbre. But audio sentiment analysis is still in a nascent stage in the research community. Using an hybrid approach comprising of audio analysis as well as text analysis would greatly improve the accuracy with which the system can extract emotions.

Navas, Eva, Inma Hernez, and Iker Luengo [4] have shown that these prosodic features and acoustic features such as power and pitch contribute to the sentiment variation. By incorporating these audio features of the accompanying speech, the capability of the model to recognize the aforementioned context can be increased.

Arpit Shah and Shivani Firodiya [3] used acoustic features such as MFCC, STFT, Contrast, Mel Spectrum, Chroma and Tonnetz extracted from the audio clips of the RAVDESS dataset to train their fully connected DNN model for sentiment analysis on audio.

Classification of sentiments or emotions can be achieved through different measures. A popular approach is Russell's circumplex model [5] which classifies emotions based on their valence and arousal values. The valence dimension indicates whether an emotion is positive or negative, whereas arousal indicates the intensity of the emotion.

**1.2 Generation of Music with sentiments**

There have been multiple studies in the domain of music generation. However, generating music that conveys a certain sentiment has been a major challenge. The goal of Affective Music Composition (AMC) is to automatically generate music that is perceived to have specific emotions. Sentiment is perceived in music due to several features such as melody, harmony, tempo, timbre, etc. Quieter volume, slower tempo and lower pitch is found to convey the emotion of sadness whereas louder volume, faster tempo and higher pitch is perceived as happier. Ferreira and Whitehead [1] have presented a generative mLSTM model that can be controlled to generate symbolic music with a given sentiment. Their model had been trained on a corpus of MIDI files.

They treated the music composition problem as a language modeling problem and hence represented music pieces as a sequence of words and punctuation marks from a vocabulary that represents events retrieved from the MIDI file. Their model used a Genetic Algorithm (GA) to optimize the weights of their neurons in order to lead the mLSTM to generate only positive or negative pieces. Two independent executions of this GA was performed, one to optimize the mLSTM for generating positive pieces of music and another was for negative pieces. This model could also be used to perform sentiment analysis of symbolic music.

**REFERENCES**

[1] Lucas N. Ferreira, Jim Whitehead. “Learning to Generate Music With Sentiment”, *20th International Society for Music Information Retrieval Conference, Delft, The Netherlands,* 2019.

[2] Steve Rubin and Maneesh Agrawala. 2014. “Generating emotionally relevant musical scores for audio stories”. In Proceedings of the 27th *annual ACM symposium on User interface software and technology (UIST '14).* Association for Computing Machinery, New York, NY, USA, 439–448.

[3] Arpit Shah, Shivani Firodiya, “Audio Sentiment Analysis after a Single-Channel Multiple Source Separation.” Indiana University Bloomington.

[4] Navas, Eva, Inma Hernez, and Iker Luengo. “An objective and subjective study of the role of semantics and prosodic features in building corpora for emotional TTS.” IEEE transactions on audio, speech, and language processing 14.4 (2006).

[5] Russell, J. A. A circumplex model of affect. Journal of personality and social psychology 39, 6 (1980), 1161.