Using Audio FX to alter music emotion

Stelios Katsis [1]

[1]Dept. of Electrical and Computer Engineering NTUA

Introduction

Music Emotion Recognition (MER) utilizes machine learning to analyze musical features like melody and rhythm, offering insights into the emotional impact of music on listeners. While MER has advanced significantly, the influence of audio effects (FX) such as reverb and distortion on perceived emotions remains under-explored. This study investigates how audio FX dynamically shape emotional perception in music, addressing a critical gap in Music Information Retrieval (MIR).

Emotion theory

Psychological theories of emotion are widely discussed and play a key role in this article. The most common emotion theories adopted by scientists are two:

- **Discrete Emotion Theory** According to this theory, the emotions can be grouped in a number of discrete categories (e.g. happy, sad, calm, excited etc). This theory is simpler, as there is a clear distinction between emotions, but it is also less accurate, because it cannot depict the emotional complexity and diversity.
- **Continuous Emotion Theory** According to this theory, the emotions are characterized based on two values called valence (depicting the joy and happiness of a song) and arousal (depicting the excitement and intensity). These values are placed in a 2-dimensional plane and are between [-1,1]. This theory, on the other hand, is more complex, yet more expressive.

Methodology

The way we are going to proceed with our research is to use 2 different models with different recorded performance on two different datasets, each with its own advantages and difficulties

1. **Models** As shown in Fig.1, MERT is a relatively new state-of-the-art model that shows remarkable performance in the MER field. We will try two different variations of the MERT model, the **MERT-v1-330M** model and the **MERT-v0-public**, a simple version of MERT trained with publicly available audio
2. **Datasets**. We will use the dataset **EMOPIA**, with four discrete emotional values, and the **DEAM** dataset with continuous values in the VA plane.

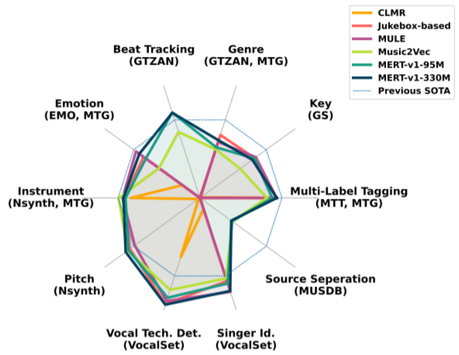

Figure 1. Comparison of various music audio self-supervised models as reported through the MARBLE.

Model Evaluation

In order to study the emotional effect of audio effects, we first have to construct a model, capable of identifying accurately the emotion of a given audio. Both MERT models produce a vector of 1024 elements, describing the audio. That's why we have to build two adapters for the MERT models, one for regression and one for classification problems.

- **Regression** In the DEAM dataset, the audio is presented alongside valence and arousal metrics, so in order to address the initial emotion recognition problem, we must use regression. In our case, we will use a 3-layer neural network. The third layer of the NN has two neurons (one for valence and one for arousal) and a tanh activation function
- **Classification** In the EMOPIA dataset we only have the labels Q1-Q4, corresponding to the four quadrants of the VA plane, meaning it is a classification problem. In this case we will use the XTBoost model, famous for its accuracy and precision.

Model	Valence			Arousal		
	MAE	RMSE	R^2	MAE	RMSE	R^2
MERT-v1-330M	0.1320	0.1681	0.4308	0.1081	0.1427	0.6643
MERT-v0-public	**0.1258**	**0.1587**	**0.4926**	**0.1044**	**0.1399**	**0.7045**

Table 1. Performance metrics for regression predictions across MERT models.

Model	Precision	Recall	F1-score	Accuracy
MERT-v1-330M	0.67	0.67	0.67	0.6646
MERT-v0-public	0.66	0.65	0.66	0.6584

Table 2. Performance metrics for classification predictions across MERT models.

Audio FX Integration

In this step we will use a python library called **pedalboard**, in order to insert 5 basic effects in the music. In order to incorporate the intensity of each audio effect and its impact on emotion, we select 50 songs from each dataset and apply 5 different levels for each one of the effects, having a total of 1250 tracks from each dataset. We will use 5 different audio effects:

Reverb Distortion Delay EQ Pitch Shift

Results Analysis

In the last phase, we will analyze the influence of audio effects on the emotional perception of music by visualizing the previously obtained results. To achieve this, diagrams will be generated for each category and dataset, providing a comprehensive comparison and aiding in understanding the effect of various audio manipulations on the detected emotions.

In the poster, only a few selected plots are presented to provide an overview of the results. However, the full analysis, including detailed insights and visualizations, is thoroughly documented in the accompanying article and Jupyter notebook on GitHub.

While emotional perception is inherently subjective and unique to each individual, this study successfully uncovered some primitive yet meaningful patterns about how audio effects influence emotions. For instance, effects like distortion and EQ show significant correlations with heightened arousal and specific emotional labels, whereas others like reverb and pitch shift often evoke more subtle or context-dependent emotional shifts.

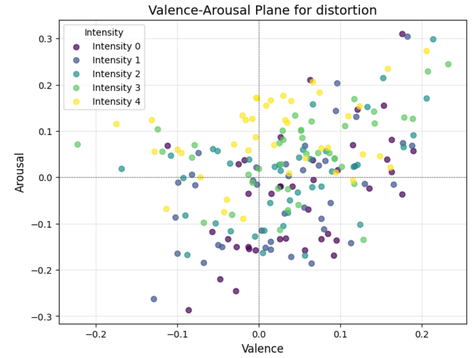

Figure 2. Correlation between Distortion and Arousal

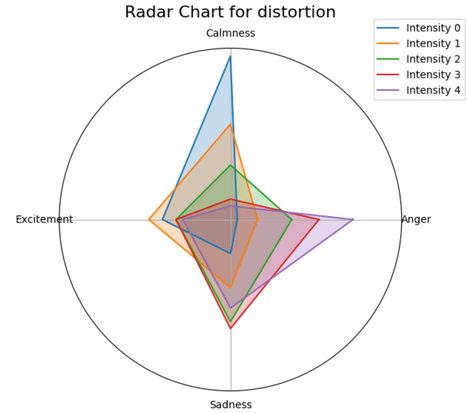

Figure 3. Emotion transfer from Calmness to Anger

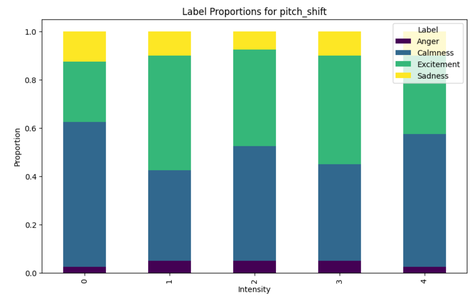

Figure 4. Pitch shift increases calmness