Automatic Music Generation Based on Emotions

Bingcheng Li, Caiyi (Giselle) Deng, Divit Karmiani

# **Idea**

## Goals and objectives

We want to develop an automatic music generation pipeline based on emotions. The pipeline would take in videos with a person’s facial expression and audio as input, detect that person’s emotions, and generate brand-new music pieces based on the emotions.

## **Data source**

We will use open-source datasets for emotion detection and music generation.

The eNTERFACE' 05 Audio-Visual Emotion Database[[1]](#footnote-0) provides 1166 video sequences presented by 42 subjects eliciting six particular emotional states: anger, disgust, fear, happiness, sadness, and surprise, which we will use to train the emotion detection model.

Lakh Piano Dataset[[2]](#footnote-1) provides us with 20,000 MIDI samples of different genres, which will be used for music generation.

## Application

There are two potential applications for the pipeline:

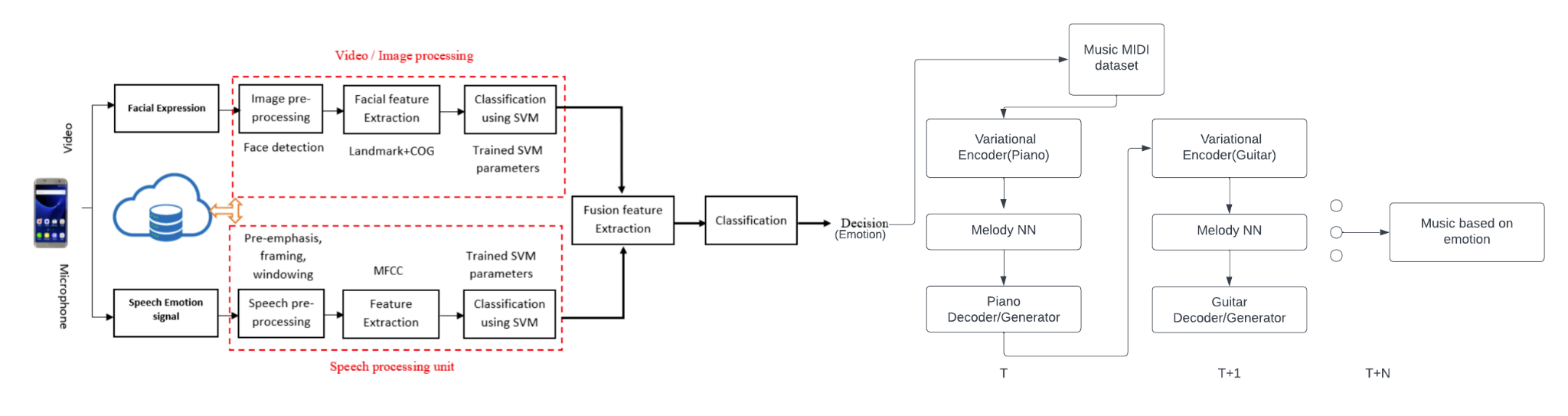
* A mental support system that induces an emotion using music basics on a person’s mood. For example, if a person is angry, we can generate soft and smooth music to relax them.
* Background music generator for a video clip. For a silent video with a person’s facial expression and audio, we can add background music based on context to make the video more vivid.

# **Significance**

There are two gaps that we are trying to fill. The first one is that emotion recognition based on facial expressions or speech alone may have potential issues. For instance, one can easily hide their emotion by changing their facial expression or speech tone. We want to combine two modalities (facial expression and speech) to train a model that returns more accurate emotional classification.

The second gap is that there is no existing model to generate music based on facial expressions and sound. However, there are machine learning models to implement emotion detection based on facial or speech recognition and automatic music generators based on emotion. We plan to combine two models together to build a machine learning pipeline that can detect emotion based on facial expression and speech, and then generate music that matches the detected mood.

**Model Architecture**

****

# **Plan**

| **Timeline** | **Objective** |
| --- | --- |
| 11/3/2022 | Research about different papers and online resources |
| 11/10/2022 | Building and training facial expression based model |
| 11/17/2022 | Building and training Speech based emotion recognition model |
| 11/24/2022 | Building Automated Music generated model |
| 12/1/2022 | Build end-to-end pipeline to connect the 2 models |
| 12/8/2022 | Get ready for demo |

# 

# **Team members**

Bingcheng Li: I don’t have much machine learning experience, but I write clean codes and am willing to learn. I believe my research experience would also help me contribute more to researching previous work and various resources.

Caiyi (Giselle) Deng: I’m familiar with machine learning algorithms and architecture. I don’t have too much experience on the application side of machine learning, but I believe my background will help us understand the problem conceptually and build up the framework. Additionally, I have solid coding skills and have worked with many python frameworks. I believe my experience will be helpful in parsing input and building demos.

Divit Karmiani: I have some experience in Data Science and Machine Learning. Hence, I am excited to put this knowledge into use and steer our team toward a successful end-to-end project. I have never worked on Deep Learning and Neural Networks before and also never built an end-to-end pipeline requiring sensors. This will be a great learning experience for me.

1. <https://ieeexplore.ieee.org/document/1623803?reload=true> (A copy of the dataset can be found in the “results” section of the website <http://www.enterface.net/enterface05> ) [↑](#footnote-ref-0)
2. <https://towardsdatascience.com/generating-music-using-deep-learning-cb5843a9d55e> [↑](#footnote-ref-1)