Face the Music

Amelia Simonoff

DD/MM/2022

A thesis presented for the degree of

Bachelor of Arts in Music

At the

THE UNIVERSITY OF CHICAGO | COLLEGE

1 Abstract

Since time immemorial, music has both elicited and strengthened emotions.

Current facial recognition technology has progressed to the point that it can be done in nearly real-time. In this project, I use the Python library DeepFace to read peoples' emotions in real-time, and then the Spotify Web API and Web Playback SDK to play music related to those emotions.

I set up this project on a Raspberry Pi, continually running a Python script that reads data from the camera and then pulls data from Spotify. I then set up the Raspberry Pi in the Media Arts, Data, and Design (MADD) Center at the University of Chicago.

2 Background

Music and emotions have been indelibly linked throughout human history. From the first chants to pop songs, music has evoked emotions in its listeners since time immemorial. Here, I was interested in the other direction: what if we create music based on an emotion, instead of the other way around?

To answer that question, then, I needed a way to read the emotions of users in near real-time, and then a way to either synthesize music of a certain emotion or pull from an already-existing database. I knew that facial recognition software could accomplish the first step, and so I chose to use a pre-trained model.

Facial recognition software has been available for the past DECADE (citation), and its implementations in the modern day are not without controversy. However, much research has been done on facial recognition, and current implementations are incredibly lightweight, especially using pre-trained models.

Unfortunately, real-time music synthesis is not yet available. While artificial-intelligence based musical synthesis has already passed the Turing Test (citation), the models are not quite versatile enough yet to produce a continuous stream of music with changing input parameters (i.e., the emotion of the user). As such, I decided to use a corpus of music that is already labeled, which is found in many streaming services, such as Apple Music or Spotify.

Spotify, the most-used music streaming service in the world (citation), has many playlists with emotional labels, such as "Happy Beats" or "Sad Piano", etc.. However, Spotify does not label their music based on emotion outright, and instead classifies music by a large variety of parameters (see Methods).

CONCLUSION

3 Significance Statement

Music is cool facial recognition is dope something about fronto-temporal dementia

4 Methods

This project is running on a Python backend with a Node.js environment communicating to a React application to output the music to a browser. The general overview of the project is outlined in Figure 1 (INSERT FLOWCHART).

I use a Python environment running OpenCV to pull video data from the webcam. Each frame is passed to DeepFace, which is a wrapper for multiple state-of-the-art emotion recognition models. I rewrote the code for real-time analysis of incoming video frames.

Each frame is classified by the probability that it is one of the following emotions: `angry, disgusted, scared, happy, sad, surprised`, and `neutral` (citation). Then, the most likely emotion is chosen as the `dominant emotion`, which is then sent for downstream processing.

Each video frame is also rendered, and a user's emotion, if available, is displayed on the frame. Pulling and rendering video frames takes significantly less time than analyzing an emotion in a frame, so both processes (known as workers for multiprocessing) run asynchronously and communicate using a Queue datastructure.

Parsed emotions are also passed to a third worker, which communicates with Spotify through the Spotify Web API. Upon initialization, the worker authenticates the user using the Spotify OAuth service, and then pulls the user's top songs and artists. The worker then makes a list of the genres of the user's top songs, and compares it with the available genres for recommendation, which is used as a mask.

After initialization, the worker loads the parameters for the recommendation request for each emotion. Every ten seconds, the worker queries Spotify for a list of recommendations given a random number of user top sonsg, artists, and genres as seeds, as well as the parameters given by the current emotion. The worker then plays a randomly chosen song from the returned list of suggestions, starting between one-sixth and one-third of the way through the song.[1]

The tunable parameters that I chose for each emotion for the recommendations query from Spotify are as follows. Each parameter has a maximum, minimum, and target value (e.g. `min_acousticness`, `max_danceability`, `target_valence`): `acousticness`, `danceability`, `energy`, `instrumentalness`, `liveness`, `loudness`, `popularity`, `speechiness`, `tempo`, and `valence`. See the supplemental material for the tuned parameters for each emotion.

Concurrently, a Node.js process supporting a React application is running to display the Spotify Web Playback SDK in a browser. The backend once again authorizes the user, and the browser functions as a device for Spotify playback, which is used as the default playback destination for the Spotify worker.

5 Code Availability

The code is available on Github under the GNU General Public License, version 3.

6 Observations

hehe look at all those ~~monkeys~~ people controlling this with their face hehe

7 Discussion

8 Conclusion

1. I chose these values because they are close to the start of the song but are past any lead-ins or introns, which might not accurately represent the emotion codified by the song.

References

Clynes Manfred and Nettheim Nigel. Music, Mind, and Brain: The Neuropsychology of Music. Edited by Manfred Clynes. Chap. IV: The Living Quality of Music: Neurobiologic Patterns of Communicating Feeling. New South Wales State Conservatorium of Music, 1983.

Emotions can be communicated through physical expression, such as a change in finger pressure over time. These can then be converted to essentic forms with recognizeable emotional characteristics. Essentic forms (i.e., musical pieces with specific dynamic expressive communications) can then be created and synthesized by computers.

Cohen Annabelle J. Music and Emotion: Theory and Research. Edited by Patrik N. Juslin and John N. Sloboda. Chap. 11: Music as a Source of Emotion in Film. Oxford University Press, 2001.

Music adds to the diegetic functions of the film while being non-diegetic itself. Music, among other things, induces the mood of the scene in film. Emotions of characters can be communicated through the musical scores. Music also controls emotional responses, "directs attention to an object and ascribes its meaning to that object", and commands interest. Music represents emotion in the abstract, while the screen represents the object of emotional direction.

Gabrielsson Alf. Music and Emotion: Theory and Research. Edited by Patrik N. Juslin and John N. Sloboda. Chap. 19: Emotions in Strong Experiences with Music. Oxford University Press, 2001.

Lists reports of intense, positive, negative, and mixed reactions to music. Reactions are all mostly for basic emotions, with some complex emotions. It is difficult to distinguish between emotions expressed in and aroused by music.

Harrer G. and Harrer H. Music and the Brain: Studies in the Neurology of Music. Edited by MacDonald Critchley and R. A. Henson. Chap. 12: Music, Emotion and Autonomic Function. William Heinemann Medical Books Limited, 1977.

Emotional reactions to music depend on the attitude of the listener towards the music, their current mood, etc.. Different music can produce cardiovascular or motor responses. Responses greater in performers than in listeners. Pulse can synchronize to the beat of the music (or some subbeat). Muscular movement can be recorded throughout a musical piece (EMG). Low dosages of tranquilizers suppress muscle movements during musical listening, but not emotional experiences. Large doses suppress both.

Henson R. A. Music and the Brain: Studies in the Neurology of Music. Edited by MacDonald Critchley and R. A. Henson. Chap. 14: The Language of Music. William Heinemann Medical Books Limited, 1977.

Music is a language that communicates musical ideas well and succinctly, which speech is unable to. Music cannot be described in words. There may sometimes be ”meanings beyond musical ideas”, but that is up to the listener to decide. Some of these ideas evoke emotions: ”...music is then the image of these emotions...”.

Huron David and Margulis Elizabeth Hellmuth. Handbook of Msuic and Emotion: Theory, Research, Applications. Edited by Patrk N. Juslin and John A. Sloboda. Chap. 21: Musical Expectancy and Thrills. Oxford University Press, 2011.

Music is often sought out to change one’s mood, second only to a conversation with a close friend. Prediction events in music elicit dopaminergic reward responses similarly to any other expected event. Musical frisson is elicited through mechanisms similar to fear.

Juslin Patrik N. Music and Emotion: Theory and Research. Edited by Patrik N. Juslin and John N. Sloboda. Chap. 14: Communicating Emotion in Music Performance: A Review and Theoretical Framework. Oxford University Press, 2001.

Performers can communicate emotions to listeners with high accuracy through cues. Complex emotions are a combination of basic emotions, which can vary in time and depend on note ordering. Timing patterns alone communicate some emotions, but less effective than tempo or dynamics. Emotions are communicated in music similar to using prosodic cues in the voice. Muic might stimulate emotions through an "emotional contagion" or through violation of musical expectations.

———. Musical Emotions Explained: Unlocking the secrets of musical affect. Chap. 15: Does Music Arouse Emotions? How Do We Know? Oxford University Press, 2019.

This book is the easiest to read so far. Music arousing emotions can be measured through the following mechanisms: self-reporting, psychophysiological measurements, specific patterns of neural activation (most likely a network of the amygdala, hippocampus, para-hippocampus, temporal poles, and maybe the pre-genual cingulate cortex), or action tendencies.

———. Musical Emotions Explained: Unlocking the secrets of musical affect. Chap. 16: The Prevalence of Emotional Reactions. Oxford University Press, 2019.

Music is incredibly prevalent but doesn't always elicit emotions. Music usually arouses positive emotions: joy, interest, contentment, and love; but may also elicit complex emotions. Mixed emotional responses can also occur to music. Music doesn't induce unique emotions. Interestingly, music can also arouse emotions at the individual but also interpersonal and intergroup levels.

———. Musical Emotions Explained: Unlocking the secrets of musical affect. Chap. 17: How Does Music Arouse Emotions? Oxford University Press, 2019.

Music is, indeed, a stimulus which then, through a complex mechanism, arouses emotions, just as any other stimulus. Emotions can be triggered by a "musical event," containing music, listener, and

context. Mechanisms for emotion effectations are the BRECVEMA model: Brainstem reflex, Rhythmic entrainment, Evaluative conditioning, Contagion, Visual imagery, Episodic memory, Musical expectancy, and Aesthetic judgement. Could also be cognitive goal appraisal. He goes deeper in depth in each of these in the subsequent chapters.

Juslin Patrik N. Musical Emotions Explained: Unlocking the secrets of musical affect. Chap. 25: Predictions, Implications, Complications. Oxford University Press, 2019.

He has a table of each of the BRECVEMA aspects with, among other things, induced affect, induction speed, degree of volitional influence, availability to consciousness, and dependendence on musical structure. B: arousal, surprise; R: arousal increase/decrease; E: basic emotions; C: basic emotions; V: all possible emotions; E: all possible emotions (esp nostalgia and longing); M: anticipation, anxiety, surprise, releif, tension. All are medium dependent on musical structure except E, E (low) and M (high). Musical emotions involve A1, cingulate, PFC, PAG, emotional processing. Real-world musical events often multiply mediated, not just one-to-one emotion:music. BRECVEMA response framework explains individual differences between listeners, and also explains no response (if no relevant information present). Personal musical induction mechanisms still influenced by context. Musical emotion evocation is moderately universal across cultures: low-level mechanisms conserved, responses not.

———. Musical Emotions Explained: Unlocking the secrets of musical affect. Chap. 29: Aesthetic Criteria: Meet the Usual Suspects. Oxford University Press, 2019.

This chapter and this part all deal with the Aesthetic judgement part of the BRECVEMA model. To be aesthetically valuable, music should be heard as expressive. Not all good music is beautiful but all good music is expressive. Novel music is interesting. People prefer music that arouses emotions, but emotional arousal does not require aesthetic judgements. Preferences for specific styles can either help or hinder appreciation of the music at hand.

Juslin Patrik N. Musical Emotions Explained: Unlocking the secrets of musical affect. Chap. 30: A Novel Approach Towards Aesthetic Judgement. Oxford University Press, 2019.

Music judged as aesthetically pleasing, greater than some aesthetic threshold, induces positive emotions, while displeasing emotions induce negative emotions. Not all art strives to evoke positive emotions.

———. Musical Emotions Explained: Unlocking the secrets of musical affect. Chap. 32: The Last Chorus: Putting It All Together. Oxford University Press, 2019.

Music can effect pleasurable sad (or otherwise negative) emotions. Listeners can prepare to feel the strongest emotions from a certain musical listening.

Juslin Patrik N. and Sloboda John N. Music and Emotion: Theory and Research. Edited by Patrik N. Juslin and John N. Sloboda. Chap. 4: Psychological Perspective on Music and Emotion. Oxford University Press, 2001.

Emotions are hard to define, but can be measured using self-reporting, expressive behaviors, or physiological measures. Emotions are eliciting through proximally eliciting stimuli in an unconscious manner. Sources of emotions in music can be either intrinsic to the music or extrinsic, with complex interactions.

Scherer Klaus R. and Zentner Marcel R. Music and Emotion: Theory and Research. Edited by Patrik N. Juslin and John N. Sloboda. Chap. 16: Emotional Effects of Music: Production Rules. Oxford University Press, 2001.

Music can evoke emotions through: A. Central route production: 1. appraisal theory, which evaluates the con/de-structiveness of a musical stimulus and its compatibility with internal and external standards; 2. by bringing up memories with an emotional valence; or 3. empathetic responses to the performers. B. Peripheral route production: 1. proprioceptive feedback; or 2. by facilitating pre-existing emotions.

Schubert Emery. Music and Emotion: Theory and Research. Edited by Patrik N. Juslin and John N. Sloboda. Chap. 17: Continuous Measurement of Self-Report Emotional Response to Music. Oxford University Press, 2001.

Emotions can only be induced over time.

Sloboda John A. and O'Neill Susan A. Music and Emotion: Theory and Research. Edited by Patrik N. Juslin and John N. Sloboda. Chap. 18: Emotions in Everyday Listening to Music. Oxford University Press, 2001.

Music is ubiquitous and intrinsically socially and culturally linked, which people often forget. People listen to music to fill functional niches, and can influence mood. People like music moe when they can control it (i.e., in private spaces, or busking). Music is directly tied to personality and identity, and can be used daily as self-therapy.

Stravinsky Igor. An Autobiography. Simon and Schuster, 1936.

Quote: "Most people like music because it gives them certain emotions, such as joy, grief, sadness, an image of nature, a subject for daydreams, or – still better – oblivion from "everyday life." They want a drug – "dope." It matters little whether this way of thinking of music is expressed directly or is wrapped up in a veil of artificial circumlocutions. Music would not be worth much if it were reduced to such an end. When people have learned to love music for itself, when they listen with other ears, their enjoyment will be of a far higher and more potent order, and they will be able to judge it on a higher plane and realize its intrinsic value.".

Sundberg Johan. Music, Mind, and Brain: The Neuropsychology of Music. Edited by Manfred Clynes. Chap. VII: Speech, Song, and Emotions. New South Wales State Conservatorium of Music, 1983.

Sundberg discusses vocally generated emotions, focusing on phonation frequencies. He also discusses singing with an emotional expression, with the following observed criteria, increasing: 1. Tempo: sorrow, neutral, joy, anger, fear. 2: Pauses between syllables: neutral, joy, anger, sorrow, fear. 3: Voice amplitude:

fear, neutral, sorrow, joy, fear. 4: Tone onsets/offsets: anger/fear, joy, neutral, sorrow. Joy requires more than just signal characteristics (eg pitch, loudness, outset, etc.).

9 Supplemental Material

9.1 Tuned Parameters

```
# emotion_labels = ['angry', 'disgusted', 'scared', \
#     'happy', 'sad', 'surprised', 'neutral']

angry_params = {
    'min_acousticness': 0,
    'min_danceability': 0.2,
    'min_energy': 0.4,
    'min_instrumentalness': 0,
    'min_liveness': 0,
    'min_loudness': -60,
    'min_popularity': 0,
#     'min_speechiness': 0,
    'min_tempo': 80,
    'min_valence': 0,

    'max_acousticness': 0.8,
    'max_danceability': 0.65,
    'max_energy': 1,
    'max_instrumentalness': 1,
    'max_liveness': 0.4,
    'max_loudness': 0,
    'max_popularity': 100,
```

```
    'max_speechiness': 1,
#       'max_tempo': 120,
    'max_valence': 0.6,

    'target_acousticness': 0.3,
    'target_danceability': 0.4,
    'target_energy': 0.65,
    'target_instrumentalness': 0.5,
    'target_liveness': 0.5,
#       'target_loudness': -60,
#       'target_popularity': 0.5,
    'target_speechiness': 0.2,
    'target_tempo': 140,
    'target_valence': 0.2
}

disgusted_params = {
    'min_acousticness': 0,
    'min_danceability': 0.2,
    'min_energy': 0.4,
    'min_instrumentalness': 0,
    'min_liveness': 0,
    'min_loudness': -60,
    'min_popularity': 0,
    'min_speechiness': 0,
    'min_tempo': 0,
```

```
    'min_valence': 0,

    'max_acousticness': 1,
    'max_danceability': 0.6,
    'max_energy': 0.7,
    'max_instrumentalness': 1,
    'max_liveness': 0.4,
    'max_loudness': 0,
#   'max_popularity': 100,
    'max_speechiness': 1,
    'max_tempo': 120,
    'max_valence': 0.5,

#   'target_acousticness': 0.5,
    'target_danceability': 0.4,
    'target_energy': 0.55,
    'target_instrumentalness': 0.5,
    'target_liveness': 0.5,
    'target_loudness': -60,
#   'target_popularity': 0.5,
    'target_speechiness': 0.5,
#   'target_tempo': 0.5,
    'target_valence': 0.2
}

scared_params = {
```

```
    'min_acousticness': 0,
    'min_danceability': 0,
    'min_energy': 0,
    'min_instrumentalness': 0,
    'min_liveness': 0,
    'min_loudness': -60,
    'min_popularity': 0,
    'min_speechiness': 0,
    'min_tempo': 0,
    'min_valence': 0,

    'max_acousticness': 1,
    'max_danceability': 0.4,
    'max_energy': 0.5,
    'max_instrumentalness': 1,
    'max_liveness': 0.4,
#     'max_loudness': 0,
#     'max_popularity': 100,
    'max_speechiness': 0.5,
    'max_tempo': 120,
    'max_valence': 0.6,

    'target_acousticness': 0.5,
    'target_danceability': 0.2,
    'target_energy': 0.3,
    'target_instrumentalness': 0.5,
```

```
    'target_liveness': 0.5,
    'target_loudness': -60,
#       'target_popularity': 0.5,
    'target_speechiness': 0.5,
#       'target_tempo': 0.5,
    'target_valence': 0.35
}

happy_params = {
    'min_acousticness': 0,
    'min_danceability': 0.5,
    'min_energy': 0.5,
    'min_instrumentalness': 0,
    'min_liveness': 0,
    'min_loudness': -60,
#       'min_popularity': 0,
    'min_speechiness': 0,
    'min_tempo': 60,
    'min_valence': 0.4,

    'max_acousticness': 1,
    'max_danceability': 1,
    'max_energy': 1,
    'max_instrumentalness': 1,
    'max_liveness': 0.4,
    'max_loudness': 0,
```

```
    'max_popularity': 100,
    'max_speechiness': 1,
    'max_tempo': 120,
    'max_valence': 1,

#    'target_acousticness': 0.5,
    'target_danceability': 0.8,
    'target_energy': 0.8,
#    'target_instrumentalness': 0.5,
    'target_liveness': 0.3,
#    'target_loudness': -60,
#    'target_popularity': 0.5,
#    'target_speechiness': 0.5,
#    'target_tempo': 0.5,
    'target_valence': 0.8
}

sad_params = {
    'min_acousticness': 0,
    'min_danceability': 0,
    'min_energy': 0,
    'min_instrumentalness': 0,
    'min_liveness': 0,
    'min_loudness': -60,
    'min_popularity': 0,
#    'min_speechiness': 0,
```

```
    'min_tempo': 0,
    'min_valence': 0,

    'max_acousticness': 1,
    'max_danceability': 0.4,
    'max_energy': 0.5,
    'max_instrumentalness': 1,
    'max_liveness': 0.4,
    'max_loudness': 0,
#     'max_popularity': 100,
    'max_speechiness': 1,
    'max_tempo': 120,
    'max_valence': 0.4,

    'target_acousticness': 0.6,
    'target_danceability': 0.2,
    'target_energy': 0.2,
    'target_instrumentalness': 0.8,
    'target_liveness': 0.3,
#     'target_loudness': -60,
#     'target_popularity': 0.5,
    'target_speechiness': 0.5,
    'target_tempo': 80,
    'target_valence': 0.2
}
```

```
surprised_params = {
    'min_acousticness': 0.3,
    'min_danceability': 0.4,
    'min_energy': 0.3,
#    'min_instrumentalness': 0,
    'min_liveness': 0,
    'min_loudness': -60,
    'min_popularity': 0,
    'min_speechiness': 0,
    'min_tempo': 0,
    'min_valence': 0.2,

    'max_acousticness': 1,
    'max_danceability': 0.6,
    'max_energy': 0.8,
    'max_instrumentalness': 1,
    'max_liveness': 0.4,
    'max_loudness': 0,
    'max_popularity': 100,
    'max_speechiness': 1,
#    'max_tempo': 120,
    'max_valence': 1,

    'target_acousticness': 0.5,
    'target_danceability': 0.5,
    'target_energy': 0.5,
```

```
    'target_instrumentalness': 0.5,
    'target_liveness': 0.5,
    'target_loudness': -60,
#     'target_popularity': 0.5,
    'target_speechiness': 0.5,
    'target_tempo': 108,
    'target_valence': 0.5
}

neutral_params = {
    'min_acousticness': 0,
    'min_danceability': 0,
    'min_energy': 0.3,
    'min_instrumentalness': 0,
    'min_liveness': 0,
    'min_loudness': -60,
    'min_popularity': 0,
    'min_speechiness': 0,
    'min_tempo': 0,
    'min_valence': 0.3,

    'max_acousticness': 0.8,
    'max_danceability': 0.5,
    'max_energy': 0.7,
    'max_instrumentalness': 1,
    'max_liveness': 0.2,
```

```
    'max_loudness': 0,
    'max_popularity': 100,
    'max_speechiness': 0.6,
    'max_tempo': 140,
    'max_valence': 0.7,

    'target_acousticness': 0.5,
    'target_danceability': 0.4,
    'target_energy': 0.5,
    'target_instrumentalness': 0.5,
    'target_liveness': 0.5,
    'target_loudness': -60,
#   'target_popularity': 0.5,
    'target_speechiness': 0.5,
#   'target_tempo': 0.5,
    'target_valence': 0.45
}
```