

CS490: Read the Room

Jordan Smith (JAS449)

Advisor: Scott Petersen

2/5/2020

Inspiration:

With the rise of home assistants like Amazon Alexa and Google Home I got to thinking about stretching the capabilities of the voice-enabled home assistant. I thought if a home assistant could recognize how a group of people is feeling it would be able to guide them in the best direction of music choice to complement a given mood. For instance, if there were a room with two people speaking with low energy and giving off an impression of a relaxed environment, the system would recommend a genre of music like classical or jazz to adhere to the ambiance of the room. On the other hand, if there were a large group of people bustling with energy, speaking rapidly and loudly, the system may recommend a genre more akin to electronic dance music or pop music to complement the mood of the crowd.

Abstract:

This project aims to create a home-assistant system that guides the user in music genre selection based on the mood of the room. The mood will be determined by an analysis of the soundscape of the room. More specifically, by four different major components: Speaker diarization (the number of different unique speakers), sound wave pitch analysis, sound wave amplitude analysis, and speech-to-text natural language processing. I'll create a program using Convolutional Neural Networks that will take in a waveform file and analyze it for these key components. I'll use existing emotional sound data to train the neural network to classify the emotional content of the soundscape. In combination with knowing how many people are in the room and the words that are being said, I'll be able to make a statement about the general mood of the room. From there I will have the system match a genre of music to the space, and effectively "Read the Room."

Primary Goals:

1. Speaker Diarization:

Determining how many people are in a room will help the system determine a base-line energy level for the room.

2. Python Audio Analysis

I'll use machine learning and Mel-Frequency Cepstral Components(MFCC) to analyze the amplitude and frequency data of the soundwaves for emotional content.

3. Natural Language Processing

I'll use speech-to-text and search for key phrases within audio snippets that could help the system determine the mood of the room.

4. Translating Mood to Music Genre

I'll create a program that interprets all the audio analysis and produces a statement about the mood of the room. The program will then match the mood to a music genre and communicate to the user what it read.

5. Integrating Home Assistant

Ideally, I'll compartmentalize all of the major components into the home assistant Amazon Alexa, by developing my own Alexa skill. However, it is more likely that I will create an audio collection and analysis system on a computer and have the home assistant communicate with that computer.

Stretch Goals:

1. User Approval

I'll implement a way for the user to give positive or negative feedback to the systems genre recommendation. I'd then tune the system such that the user's previous reactions to previous music recommendations would affect what it suggests next. This way the system could not only read the room but also get to know the user himself.

Deliverables:

- Abstract: 250-500 word introduction to my work.
- Written Summary (2-4 pages): A thorough explanation and walkthrough of my project
- Video Demo: a demonstration of the system working in different situations with different numbers of voices and different emotional profiles
- Source Code: all source files, code, and any other byproducts of my work.

Timeline:

Weeks 4-6: Creating an Alexa developer account and learning to navigate creating Alexa skills. Also conducting more research on Convolutional Neural Networks and emotion recognition through sound.

Weeks 6 - 8: Beginning the creation of a system by focusing first on speaker diarization and music genre recommendation.

Weeks 8 - 10: Looking to integrate speech emotion recognition and speech-to-text language processing into the system.

Weeks 10 - 12: Focusing on having the entire system work together and connecting it with a home assistant.

Weeks 12 - 14: Should time permit, I'll begin working on my stretch goal of integrating user feedback into making the home assistant better at reading the room.

Resources:

Chu, Reza. "Speech Emotion Recognition with Convolution Neural Network." *Medium*, Towards Data Science, 1 June 2019, towardsdatascience.com/speech-emotion-recognition-with-convolution-neural-network-1e6bb7130ce3.

Dasgupta, Poorna. "Detection and Analysis of Human Emotions through Voice and Speech Pattern Processing." *ArXiv.org*, 2017, arxiv.org/pdf/1710.10198.pdf.

Fujita, Yusuke, et al. "End-to-End Neural Speaker Diarization with Permutation-Free Objectives." *ArXiv.org*, 12 Sept. 2019, arxiv.org/abs/1909.05952.

Gubin, M V. "Using Convolutional Neural Networks to Classify Audio Signal in Noisy Sound Scenes." *Using Convolutional Neural Networks to Classify Audio Signal in Noisy Sound Scenes - IEEE Conference Publication*, 2018, ieeexplore.ieee.org/abstract/document/8570117.

Kerkeni, Leila, et al. "Automatic Speech Emotion Recognition Using Machine Learning." *IntechOpen*, IntechOpen, 25 Mar. 2019, www.intechopen.com/online-first/automatic-speech-emotion-recognition-using-machine-learning.

Muhr, Alex. "Building a Vocal Emotion Sensor with Deep Learning." *Medium*, Towards Data Science, 14 Aug. 2019, towardsdatascience.com/building-a-vocal-emotion-sensor-with-deep-learning-bedd3de8a4a9.

Smales, Mike. "Sound Classification Using Deep Learning." *Medium*, Medium, 27 Feb. 2019, medium.com/@mikesmales/sound-classification-using-deep-learning-8bc2aa1990b7.

Tarunika, Krishna, et al. "Applying Machine Learning Techniques for Speech Emotion Recognition." *Applying Machine Learning Techniques for Speech Emotion Recognition - IEEE Conference Publication*, 2018, ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8494104.

Zhang, Aonan, et al. "Fully Supervised Speaker Diarization." *ArXiv.org*, 19 Feb. 2019, arxiv.org/abs/1810.04719.