SPEECH EMOTION AND DRUNKENNESS DETECTION USING A CONVOLUTIONAL NEURAL NETWORK

JOSHUA MILLER, JILLIAN DONAHUE, BEN SCHMITZ
University of Rochester, Department of Audio and Music Engineering

ABSTRACT
One problem with Artificial intelligence (AI) is that it lacks emotional or situational knowledge about the human with which it interacts. This project attempts to propose a solution to this problem by detecting emotion or drunkenness through speech input. Using convolutional neural networks, models for four states were created: happy, sad, angry, and intoxicated. Our network aims to classify these four states with accuracy above 80% by building upon previous research in emotion detection.

OBJECTIVES
- Create a model able to classify the following states using input voice data: happy, sad, angry, and intoxicated
- Source enough good speech data to thoroughly train a model for the desired 80% accuracy figure

BACKGROUND
Speech Emotion Detection Problem
- Speech emotion detection has been a challenging and complex problem to tackle
- An accurate representation of the classified emotions needs to be created
- Complex networks and large amounts of data are typically required for an accurate model

IMPLEMENTATION
Neural Network Methods
- Methods for speech emotion detection based on neural network technology are in vogue currently
- Neural nets have the ability to learn complex patterns like those that indicate emotion in speech without much guidance from programmers

Preprocessing
- Audio files were downsampled from 44.1 kHz to 16 kHz to limit spectral data to the frequencies most relevant for speech features
- Augmented each audio file with several times its original length of white noise at 15 dB SNR to avoid overfitting
- Took wide band spectrograms with an 80 sample window, 70 sample overlap, and a DFT size of 512, removed all frequency bands below 0 Hz and above 4 kHz, then rescaled using bicubic image resizing to optimize the data for quick training

CONCLUSIONS
- Network parameters not yet fully optimized
- Shortage of drunk data might be contributing to low classification rate
- Adding more convolutional layers may also give better results

FUTURE WORK
- Allow user input for classification in real time
- Discover the influence of language by using multiple databases
- Implementing curriculum learning for higher accuracy

SOURCES