

Prakhar K.(13485) and Soumya Gayen(13708)

Mentor: Prof. Piyush Rai

Emotional state detection via speech in spoken Hindi

So much work has been done on Speech-Recognition that it is considered a more or less solved problem. The interpretation of emotional state of the speaker, is a relatively newer topic with scope for improvement. Presently there are methods with high accuracy(around 90%) for emotional state detection from speech input, using a combination of a Gaussian Mixture Model with an SVM classifier (an anchor model) applied on MFCC vectors, generated by preprocessing the audio corpus[1]. We intend to experiment this and other methods on an artificial audio-corpus generated by curating speech performed by actors in sophisticated environments[3]. We will be using anchor model. Then we'll extend our model to perform classification on natural audio samples, which are filled with a lot of white and colored noise. As suggested by another model[2], we may use the efficient GMM-SVM with an RNN recognition model to improve upon the accuracy. To speed up the neural net, we can implement an estimate based on an SGD method(see [here](#)). Using the vector quantization method in [1] and the RNN classifier in [2] we intend to study performance vs accuracy results and determine the best combination. Since most people in our country are familiar with spoken Hindi, our emotion detection system will be trained on the database in same language. If time permits, we will try real-time emotional state detection on an audio-track.

Database:

1. Artificial Database: We will build our own speech corpus by extracting audio from the bollywood movies.
2. Natural Database: Will record numerous audio clips ourselves in natural settings(i.e. with mobile microphone and not in any studio) to build the speech corpus.

Tentative Timeline:

- week-1 to 2: Reading, building the speech corpus, and preprocessing of data
- week-3 to 4: anchor model implementation
- week-5 to 6: adding i-vector features and RNN (or other Neural Networks method)
- week-7: Final analysis and compilation of results

Relevant links are:

1. [Emotion recognition from speech using Gaussian Mixture Model and vector quantization](#)
2. [Speech emotion recognition with i-vector feature and RNN model](#)
3. [Emotion recognition from speech: a review](#)
4. [RECOGNIZING EMOTION IN SPEECH USING NEURAL NETWORKS](#)

These two are our broad areas of interests.