**Project Title: Speech Based Summarization and Emotion Analysis**

**Members:**

Pradeep Thalasta 4579117326 thalasta@usc.edu

Paridhi Verma 2139220711 vermap@usc.edu

Raksha Nagesh Bysani 3449867668 bysani@usc.edu

Suraj Hebbar Shankar 4376950843 surajhes@usc.edu

**1. Introduction**

Public speaking skills help you communicate important messages inside and outside of the organization. Public speaking is so important that it could be the deciding factor in many things such as your career development, your business growth and even in the relationships you have with your friends and family. We plan to build a tool based on analysis of the recorded speech that can help an individual improve by classifying the feedback of his speech. This is done by highlighting the key points mentioned in the speech and analyzing the over-all emotion. This problem is interesting because it helps the person to improvise the content and the emotion associated with it.

Existing approaches to solve this problem are done in two phases. First by analyzing the sentiment of the speech. Second by summarizing the speech data. These two phases are performed for a corpus based on English Language. In our approach, we try to combine both the phases by analyzing the emotion of the speech as well as summarize the content by using machine learning algorithms, followed by emotion analysis and summarization for German Corpus.

**2. Method**

**Materials:**

We used publicly available emoDB which is a Berlin database for emotional speeches. A database of 700 emotional utterances spoken by actors was recorded as part of a DFG funded research project SE462/3-1 in 1997-1999. Every utterance is named according to the same scheme:

* Positions 1-2: number of speaker
* Positions 3-5: code for text sample
* Position 6: emotion (letter stands for German emotion word)
* Position 7: if there are more than two versions these are numbered a, b, c ....

Example: 03a01Fa.wav is the audio file from Speaker 03 speaking text a01 with the emotion "Freude" (Happiness).

The database has samples for the following emotions: Anger (W), Boredom (L), Disgust (E), Anxiety/Fear (A), Happiness (F), Sadness (T) and Neutral version(N).

**Procedure:**

1. Corpus Building:

Each audio sample was tokenized sentence-wise by taking the audio file and the time intervals spoken for each sentence. (PRADEEP **EXPLAIN THE PROCEDURE AND ADD TOOLS USED)** Tools:

2. Extracting Model Parameters:

(PRADEEP **EXPLAIN THE PROCEDURE AND ADD TOOLS USED)** Tools:

3. Emotion Analysis:

We predicted the emotion behind a test speech. The following procedure was followed for this:

* ***Naive Bayes****:*

*Training –*

We used Gaussian Naïve Baye’s to train the model. This is because the features that were extracted were continuous real-valued Gaussian distributions. The model was trained for each of the following features – F0, spectral centroid, MFCC, energy levels, chroma, spectral flux , spectral spread, spectral entropy, ZCR, loudness, energy entropy, chroma deviation and spectral roll-of.

*Classification –*

The test data was predicted the emotion for each feature. The emotion class with the highest frequency amongst all the features was assigned to the sample data.

Tools:

Pandas, SKLEARN

4. Speech Summarization:

The objective of this step was to provide an extractive summarization of the content of the speech. In this step, we used Text Rank algorithm on the transcript of the speech.

[PRADEEP – **EXPLAIN BAG OF WORDS MODEL USAGE AND ADD TOOLS]**

After this step, we had the representative sentences for each sample speech along with their emotion.

5. Extracting Test Data:

**OPTIONAL**

[PARIDHI – **WRITE ABOUT YOUR CRAWLER IF YOU CAN MAKE IT WORK TO OBTAIN TEST DATA FOR ALL 7 EMOTIONS]**

**Evaluation:**

1. Corpus:

2. Emotion detection:

3. Speech Summarization:

[AS DECIDED, WE CAN MANUALLY GENERATE THE SUMMARY FOR TEXT USING VARIOUS SUMMARIZATION TOOLS ONLINE AND ASSIGN A SIMILARITY SCORE LIKE F1 SCORE]

**3. Results** 

**4. Discussion**

[WRITE ABOUT THE PRECISSION RATE OF TOKENIZING AUDIO TO FETCH THE FEATURE VALUES FOR SAMPLE DATA]

**Future research**:

**5. References**

1. [European Language Resources Association (ELRA) 2016] Mathieu Chollet, Torsten Wortwein, Louis-Philippe Morency, Stefan Scherer, “**A Multimodal Corpus for the Assessment of Public Speaking Ability and Anxiety**”.
2. [IEEE Transactions on Speech and Audio Processing ( Volume: 12, Issue: 4, July 2004 )] S. Furui, T. Kikuchi, Y. Shinnaka. “**Speech-to-text and speech-to-speech summarization of spontaneous speech**”.
3. Felix Burkhardt, Astrid Paeschke, Miriam Rolfes, Walter Sendlmeier und Benjamin Weiss.  
   “**A Database of German Emotional Speech**” [dataset].

**6. Division of Labor between teammates**

**Emotion Analysis:** Raksha

**Speech Summarization:** Pradeep

**Corpus Building:** All

**Summary generation for Evaluation:** Pradeep

**Evaluation**:

**7. Word Count for the document:** 735 words