Project Title: Speech Based Summarisation and Sentiment Analysis

Members:

Pradeep Thalasta	4579117326	thalasta@usc.edu
Paridhi Verma	5333348251	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 organisation. 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 analysing the over-all sentiment. This problem is interesting because it helps the person to improvise the content and the sentiment associated with it.

Existing approaches to solve this problem are done in two phases. First by analysing the sentiment of the speech. Second by summarising the speech data. Both of these are performed for a corpus based on English Language. In our approach, we try to combine both the phases by analysing the sentiment of the speech as well as summarise the content by using machine learning algorithms, followed by sentiment analysis and summarisation for French Corpus.

2. Method:

Materials:

The SIWIS French Speech Synthesis Database includes high quality French speech recordings and associated text files, aimed at building TTS(Text To Speech) systems, investigate multiple styles, and emphasis. A total of 9750 utterances from various sources such as parliament debates and novels were uttered by a professional French voice talent. A subset of the database contains emphasised words in many different contexts. The database includes more than ten hours of speech data and is freely available.

Tools: gensim, NLTK, pyAudioAnalysis

Procedure:

1. **Sentiment Analysis:** We try to tag each word with its modulation taken from the input audio data and classify each audio sample as Good, Bad, Anger, Sad, Optimistic and Anxiety.
   A. Naive Bayes: We will use the audio feature along with text from the training corpus to develop a Naive Bayes model and use this model to assign the class to the given input.
   B. Support Vector Machines: We use the transcript of the speech to convert the given speech to mathematical model using Word to Vector. We then use the generated vectors to classify the input based on SVM.

2. **Summarisation**: We build the summary by implementing Text Rank algorithm on the transcript of the speech.

Evaluation:

1. **Sentiment Analysis:**
   - We use F1 score to evaluate the module that is used to detect sentiment from the given audio sample.
   - We will also perform manual evaluation of the labels generated by the models
2. **Summarisation**: We will manually evaluate the summary generated based on the survey taken by the users.

3. 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 summarisation of spontaneous speech**".

3. [Yamagishi, Junichi, Honnet, Pierre-Edouard, Garner, Philip, Lazaridis, Alexandros] **The SIWIS French Speech Synthesis Database**, 2016 [dataset].

4. Division of Labour:

Sentiment Analysis: All, **Summarisation:** All

5. Word Count: 459