**SPEECH ENHANCEMENT AND MOOD DETECTION**

Methodology and Techniques

**NIMISH AGRAWAL, 140905596**

**KRITIK MATHUR, 140905586**

**Stage 1: SEPARATING HUMAN SPEECH FROM NOISE**

DATASET USED: British National Corpus http://ota.ox.ac.uk/desc/2554

The data-set is a collection of hours of spoken British English (public speaking and interviews) with information relating to start time and duration of phenomes in XML format.

When an audio is received, Fast Fourier Transform is performed to get a frequency domain amplitude graph. Depending on the loudness of the frequencies present in the audio, it is determined if a frequency pertaining to a phenome of the English language is present. Frequencies that don't correspond to any phenomes will be silenced out and Inverse Fourier Transform is performed on the newly obtained frequency domain distribution.

The audio thus obtained is generally free of noise.

**Stage 2:** **PREDICTING MOOD OF THE SPEAKER FROM THE FILTERED SPEECH**

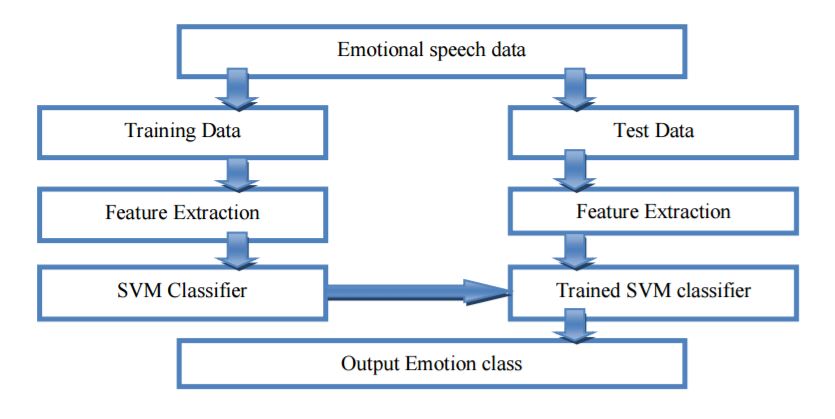
Once, we have noise-free speech input from stage-1, we can proceed to the emotion recognition stage. We use Support Vector Machines (SVMs) as a classifier to classify emotions such as Anger, Happiness, Sadness, etc.

There are a variety of temporal and spectral features that can be extracted from human speech. We use statistics relating to the pitch, Mel Frequency Cepstral Coefficients (MFCCs) and formants of speech as inputs to classification algorithms.

Broadly, different sub-stages for emotion recognition will be:

1. Feature Extraction – from stage-1 output
2. Feature Labelling – to train the SVM
3. SVM Training – to generate a model for mood prediction
4. Feature Extraction – for test data
5. SVM classification – for testing using the generated model

This process along with training and testing phases can be represented by the below flowchart:



**FLOWCHART FOR EMOTION RECOGNITION PROCESS**