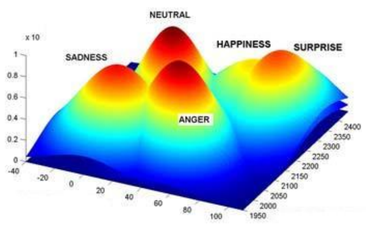

Emotion Orientation Extraction from Speech Signal

{ Kalpesh Patil, Meet Shah, Navjot Singh, Yash Bhalgat }

Method

1. **Preprocessing**: To boost the amount of energy in the high frequencies preemphasis is done
2. **Framing**: small frames with the time length in the range of 20-40 ms. It enables the non stationary speech signal to be segmented into quasi-stationary frames. Hamming Windowing technique is used.
3. **Mel Filterbank and DCT**: 12 MFCC coefficients are extracted
4. **1st and 2nd order difference**: Dynamics coefficients are then obtained by combining the first and second differences of filter energies

Intricacies and Issues

Emotion recognition is a complex task due to the following reasons:

1. Differentiating between various emotions and deciding which particular speech features are more useful is not clear. (i.e separation of linguistic content from background noise and further extraction of phonemes representing emotion).
2. Various factors like gender, accent (thus language), loudness, etc need to be considered for successfully deciding upon the emotion the speech presents.
3. The methods and features used in the system are not language independent. Hence, the method is not directly optimal for cross-lingual emotion recognition.
4. Emotional states do not have clear-cut boundaries and often differ from person to person, which imposes subjectivity in the problem.
5. As far as databases are concerned, the complexity of the problem along with cross cultural diversities makes the development of a complete common database an extremely difficult task.

References

[1] Shaikh Nilofer R.A., R R Deshmukh and V B Waghmare. GraphTrack: Comparative Analysis of Pitch and Formant for Recognizing Emotions of Isolated Marathi Speech. In *IJCA '15*

[2] Shashidhar G. Koolagudi, K. Sreenivasa Rao. Emotion recognition from speech: a review. In *Int J Speech Technol '12*

[3] Aastha J. Rajneet K. Astudy of Speech Emotion Recognition Methods In *IJCSMC, '13*

[4] MFCC Tutorial: http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs

Images

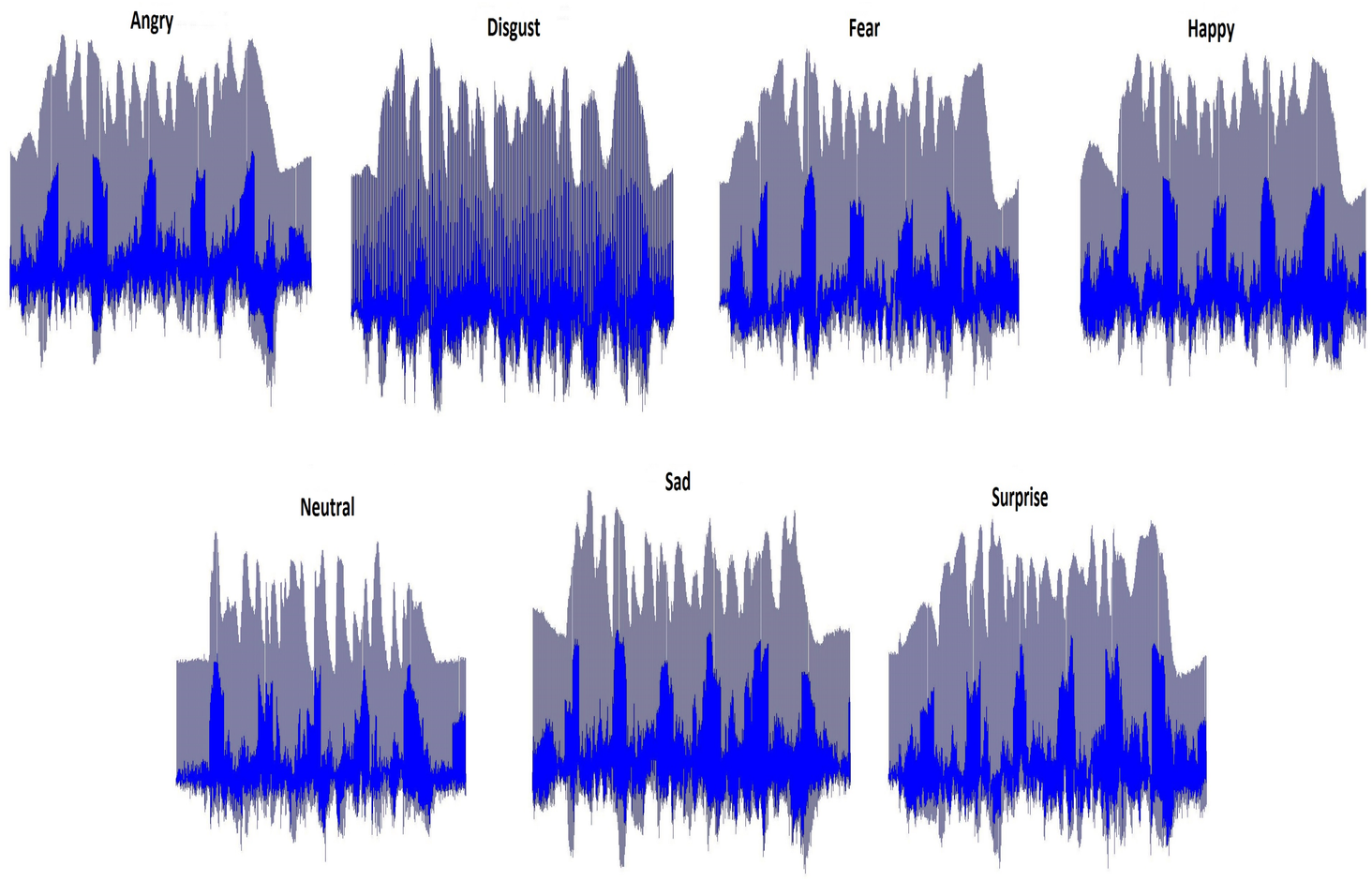

Fig :Feature vectors etracted using MFCC for various emotions

Classification Models

Once we obtain the vectors for the speech signal, we use classification techniques like SVMs, k-means, HMMs and Convolutional Neural Networks (state-of-the-art). If information of no. of different emotion classes is available, K-Means clustering or use K-PCA (for non linear mapping) and then clustering/classifiers can be used. Results from recent proceedings:

In HMM, the state transition probabilities and the output symbol probabilities are uniformly initialized. Number of states are chosen to be close to the no. of classes.

We summarized the results of standard SVM kernels - linear, polynomial and radial-basis - and also **stage SVM** for an extensive comparison with the HMM and K-NN based methods.

Among the various methods of supervised statistical pattern recognition, the Nearest Neighbor is the most traditional one, it does not consider a priori assumptions about the distributions from which the training examples are drawn. It involves a training set of all cases.

Features & Classifiers Used	Accuracy (%)
MFCC & HMM [17]	59
MFCC & K-NN [31]	67
MFCC & FFNN [25]	55
Energy + pitch + duration & SVM [13]	48
MFCC & SVM using linear kernel [5]	65
MFCC & SVM using polynomial kernel [18]	60
MFCC & SVM using radial basis function kernel [2]	55.4
MFCC & 3 – Stage SVM [proposed method]	68

Applications

1. **Call-center systems:** Recognize voices, detect anger in speech and prioritize angry calls! In real time applications such as call analysis in the emergency services like ambulance and fire brigade, verification of emotions to analyze genuineness of requests is important
2. **Medicine:** Can be used in rehabilitation, help monitoring and counselling patients/clients.
   Can be used in therapy of Autism, for people who struggle to express/interpret emotions
3. **E-learning:** Detect the state of the learner and adjust presentation style of the tutor based on the feedback
4. **Law:** Deeper discovery of depositions and can be used as a non-invasive lie detector test

Future Work

1. Higher accuracy can be obtained using the combination of more features and classification models. Hybrid approach may produce better results
2. Sparse coding could yield better performance. But this can only be concluded after verifying on some different database.
3. Expression of emotions is an universal phenomenon, which may be independent of speaker, gender and language. Cross lingual emotion recognition study may be another interesting work for further research.