1. Conclusion

In summary, while most visible speech information is, unsurprisingly, seen around the lips and it is also distributed over the entire face. Whenever the component is removed for example face or teeth, the intelligibility falls. Visual speech analysis should take care when discarding some of the many possible facial features which are available.

Chapter 1 is an introductory chapter. It discusses the basic concept involved in automatic speech recognition and migration from automatic speech recognition technology to audio-visual speech recognition, by highlighting the challenges in the area of audio-visual speech recognition.

Chapter 2 provides the review of the literature and word carried out by the researchers, methods used in the literature are primarily based on Contour Coding, Dynamic Programming, Multi-State Time Delayed Neural Network, Active Shape Model, Hidden Markov Model, Active Contour Model, Discrete Cosine Transform, and Principal Component Analysis etc. The Principal Component Analysis is widely chosen for dimension reduction and act as basis for building shape models. This chapter also provides information about speech production, speech sounds and features.

Chapter 3 deals with the methodology used in order to analyze patterns generated by processing the utterances of known dataset. It also describes about the available databases and the database related to isolated city names formed in this research work, which was made up of 10 speakers with 10 utterances of 10 isolated city names. The methodology clearly describes method of acquiring samples, converting samples into discrete set of image frame vectors. The preprocessing of frame vector was discussed prior to feature extraction using subjective and objective quality metrics as well as use
of image enhancement operations for improving the quality of image for further processing. There might be chances of having duplicate frame or the image frame with statistical characteristics and needs to be identified in order to minimize overhead on frame analysis for the utterance. This was done by implementing Principal component analysis based feature extraction of frame and detecting duplicate frame from image vector. The visual feature extraction was carried out using incremental difference procedure, preparing and representing contours for analyzing of shape of the mouth. Audio-visual integration is an area for research. This chapter focused on an informative approach towards integration of audio and visual integration using MFCC components based recognitions.

Chapter 4 devotes the result outcome from the methodology discussed in chapter 03. The patterns are analyzed visually as well as in the form speech signal. The method used of the visual analysis is low level analysis. The visual frame vector was enhanced and the statistical quality measures were recorded and the results were presented in the forms of graphs. This was required because the video input was sampled into audio as well as visual frame vector. The alignment of visual frame vector was not applicable because the data/samples acquired from the subjects have been taken under constraint environments. The frame vector was checked for redundant existence of frame under frame vector using Principal component analysis and it was observed that out of the 1254 frames of subject ‘Purva’, 107 frames was identified as redundant and the average percentage of redundant frame was 8.53% and 91.47% frames are found to be distinct. Similarly for subject 2, out of 1220 frames, 40 frames are found to be redundant and redundant frame percentage was recorded as 3.38 and distinct frame percentage was recorded as 96.72%. The frame vector was highly depending upon the environment used for acquisition of data, lighting condition, distance etc. still this was important in order minimize the overheads on the systems.

The visual feature extraction from the frame vector corresponding to the isolated city name was done through the incremental difference procedure for finding midpoint for HRL and VRL. This midpoint difference feature was computed with respective to every frame of frame vector corresponding to the word and mean difference corresponding to the word was computed. This was done for all words uttered by the speaker. The distance matrix was prepared separately with respective to each word.
uttered by the speaker. The speech signal was analyzed using MFCC and MFCC feature vector corresponding to each speech signal of utterances were recorded. The speaker dependent distance matrix was prepared; similarly word specific distance matrix was also computed. The speaker dependent confusion matrix and speaker independent confusion matrix was prepared and it was observed that for speaker 1 the speaker dependant recognition of word was found to be 88.89%, for speaker 2 it was recorded as 92.22% and speaker independent recognition of word was found to be 87.77%. The work was illustrated with two subjects, but similar type of results was recorded for other speakers too. The audio-visual fusion was out of the scope of this research work as this area was itself a subject of research. The objective of extracting of visual features of speechreading was found to be satisfactory as visual feature vector is clearly separable in space and can be used for enhancing speech recognitions systems

2. Future work

The research discussed in this thesis demonstrated the uses of visual features of lip reading are found to be an added parameter for the automatic speech recognition system. The work was carried out under constrained environment and it is very obvious that the base experiments must be extended to broader examples. In future audio-visual database will be extended towards multi-lingual audio-visual database with participation of subjects of difference age groups and instead of acquiring full face frontal video the side poses of subjects will be recorded. In next coming years, one expects to see lot of research activity in the area of audio-visual speech recognition.
APPENDIX – A

LIST OF PUBLICATIONS

1. Incremental Difference as Feature for Lipreading, International Conference in Advances in Computer Science, (ACS 2010), 21st-22nd Dec 2010, Trivandrum, Kerala, India. (Accepted For Oral Presentation)


3. Criminal Information Analysis using Facial Expression, International Journal of Computer Science and Application (IJCSA ISSN: 0974-0767) (Communicated)


5. Identification of Criminal using Facial Expression, AVISHKAR 2009, Organized by Solapur University, Solapur, 11-13 Jan 2010 (Recipient of First Prize under Teacher Category of Engineering and Technology)
