CHAPTER VI

CONCLUSION & FUTURE SCOPE

6.1. SUMMARY OF WORK

An introduction to the speech processing is followed by a study on speaker information in speech. A discussion on speech production, its characteristics, feature extraction from a speech signal is presented. The importance of feature selection is given. Datasets used for the experimentation as mentioned in a detailed fashion. Then a detailed study on three baseline systems is provided along with the results.

Several optimization methods were investigated in order to identify means to improve the system’s robustness and flexibility to handle speech patterns. It was found that a MOHABC model based on 39-Mel frequency cepstral coefficients extracted at 25 msec frame rates was able to provide higher degree of recognition accuracy than other methods pursued. Further, it was found that high Gaussian value did not contribute any significant acoustic cues and/or additional discriminant information to the system.

Then we presented the way to deal with sparse data in case of speaker verification. Here we considered speech signals of very less duration like 2 ~ 6 sec for training as well as testing. For mapping we used MAP adaptation and the benchmark dataset considered is TIMIT corpus. Using the phonemes from the cohort speech signal we can fill the holes present in the sparse speech signal. To achieve this task we have come up with a new procedure where we index the phonemes of speech signals in an indexing table. When the proposed feature selection algorithm was implanted as a part of this verification process, the results surpassed the efficiency of conventional methods.
6.2. CONCLUSIONS

Chapter -1 gives the presentation, foundation of highlight choice and inspiration of the issue. The explanation behind selecting the issue and its significance is likewise clarified. A brief discussion on past and present research work in several areas like speaker verification, sparse data, feature selection, optimization techniques and multi objective optimization is presented. The advantage of multi objective optimization is studied in Chapter -2. Some of the systems have their drawbacks with reference to the classifier used. ANNs need a longer time to execution. Decision trees suffer from limited complexity on the constructed boundary, the tree structure may not be global optimal. SVM is sensitive to noise, considers only two classes. This method of classification shows poor performance when number of samples if less than the number of features. It uses an expensive method of five-fold cross validation for probability estimation. One of the reasons for their poor performance is the number of features they have considered for classification.

Our thesis is divided into two parts. In the first part, we have proposed two feature set optimization algorithms. First algorithm is HABC which used the advantages of ant colony and bee colony to develop a hybrid model for feature set optimization. When the proposed algorithm is used for text independent speaker verification it optimized the feature set to 85%. This shows an immense gain in the accuracy rate of recognition.

Second algorithm is MOHABC which is an enhancement of the first algorithm. Multi objective optimization concept is embedded into HABC algorithm. When this optimization algorithm is applied for text independent speaker verification we obtained 87% of feature set optimization. There by it improved the accuracy rate of verification to 97.5%.

For experimentation of the above mentioned two systems we used BELIN dataset and telephone conversation dataset as explained in Chapter-3. We realized that in real time scenario the speech signal will not be pure i.e. it will contain a lot of noise in it. So we added 5db and 10db of noise in telephone conversation dataset and evaluated the system. In both the cases the proposed algorithms improved their accuracy rates in terms of EER and DCF as shown in Chapter -4.
In the second part, we proposed a system for text independent speaker verification system in the case where test and train data is sparse in nature. In real time we may encounter the cases where the speech data we get for training as well as testing is very less in amplitude. In the proposed system rather than relying on the text a speaker utters, we relied on the acoustic phonemes which we gathered from cohort speakers to fill the holes in sparse speech signal. Using a benchmark dataset and an indexing method we succeed in filling the acoustic holes. After obtaining a speech signal sufficient for verification, we used the second proposed optimization algorithm for speaker verification. The results of the resultant system surpassed the efficiency of conventional models for speaker verification using sparse data. The efficiency of proposed system is evaluated in terms of EER and DCF as shown in Chapter -5.

6.3. FUTURE SCOPE

The presented work in this thesis has shown a considerable amount of improvement in the area of speaker verification. Future scope is discussed as follows:

A. A variation in emotional state, stress, speaking style of a speaker before and after training degrades systems performance. This can be a point of investigation in future.

B. An increase in the number of cohort speakers increases the false acceptance rate. So in future an improvement in the methods to optimize number of cohort speakers will definitely improve the systems performance.

C. An extensive study of speaker mapping methods can enhance the efficiency of the system.