Conclusion and Future Enhancement
6. CONCLUSION AND FUTURE ENHANCEMENT

Speaker and speech recognition systems have become so advanced that business and health care professionals are turning to speech recognition solutions for everything, from providing telephone support to writing medical reports. Technological advances have made speech recognition software and devices more functional and user friendly, with most contemporary products performing tasks with over 90 percent accuracy.

The continuous improvement in speech recognition algorithms and the more recent development of personal computers have made it possible for nearly anyone to set up an ASR system in his/her home. Currently there are several speech recognition packages that can be purchased for use on the home computer. The primary purpose of these is to reduce the need for the keyboard by turning speech into text.

To satisfy the needs of the consumers and businesses by simplifying customer interaction, increasing efficiency, and reducing operating costs, speech recognizers are used in a wide range of applications. Recent advances in speech recognition software are creating a dynamic environment, since this technology appeals to anyone who needs or wants a hands-free approach to computing tasks. As the merger of large vocabularies and continuous recognition continues, more and more companies move towards speech recognition with the industry taking up the role of a leader in the technology sector.

6.1. CONCLUSION

In this research work, methods and techniques have been developed for doctor (speaker) and speech recognition for healthcare industry from a remote environment. The proposed work performs a speaker dependent, text-dependent
doctor identification, as an authentication process for the small vocabulary continuous speech recognition system.

In doctor identification, K-means clustering algorithm is used to prune or reduce search space of feature vectors during matching the input signal with the speech reference template from the database. The main disadvantage of identifying the optimal ‘K’ value in the traditional K-means algorithm is solved by an automated process that uses EM algorithm to find the optimal K. After identifying the most probable cluster that matches with the input signal, a multi-step approach with 4 tests, namely, cross correlation test, frequency multiplication test, frequency cross correlation test and peak signal comparison test, is used. The test results were combined using Neyman-Pearson Likelihood ratio test which produced a true result if two or more of the 4 tests produced a true result. A true result indicates that a speaker has been identified.

After successful doctor identification, the next step is continuous speech recognition. Two hybrid models using HMM and MLP were developed. The research work uses a novel architecture called Feed Forward Neural Network as the basic structure of the proposed hybrid systems. The first model combines FFNN and HMM and the second model combines Multi-Dimensional Labeling features with HMM Model.

While conducting experiments, it is clear that the performance order of the tests starting from the best result were cross correlation test, frequency multiplication test, frequency cross correlation test and peak signal comparison test. Cross correlation test and frequency multiplication test had the best successful detection capacity without high false alarm rate (wrongly identifying or not identifying speech signal) and therefore these tests were given the highest weights when compared to all other tests. Combined frequency cross correlation test also produced similar result, but not upto the mark when compared with cross
correlation test and frequency multiplication test. Peak signal comparison test was
the least successful of the tests. It failed in the process of detecting non-speech
signals.

It was also found that the average time taken for identification is directly
proportional to the number of candidate clusters selected for comparison. As the
number of clusters increases, the time for matching process also increases, which
in turn increases the execution time.

Further, the effect of using clustering to reduce search space was also
analysed. From the experimental results, it can be concluded that the accuracy of
recognition, time taken for successful recognition and thereby the execution time
were greatly enhanced by the use of clustering.

The performance of the system was determined by rigorous and objective
testing; it was found that the system performed reasonably successfully when the
test audience was completely enrolled in the database. However, the system fell
short of expectation when subjects in the test audience were not enrolled in the
database, and yielded higher error rates in these cases. During the course of
testing, it was also found that the system was, to a large extent, resistant to slight
variations in the rate of speech; however, it required at least about two seconds of
continuous speech (with pauses removed) to establish a successful match.

From the experimental results conducted during continuous speech
recognition, it was found that both the systems perform moderately well,
producing good recognition rates. The results were compared using a standard
VQ/HMM hybrid system. In all the experiments, the proposed hybrid model
outperformed the standard model. The success rate was 97 per cent and 99 per
cent on an average for FFNN/HMM model and FFNN Multi-Dimensional
Labeling/HMM model respectively. Further, by examining the recognition speed,
it was observed that the performance of FFNN/HMM model with respect to time
in seconds was better than that of Multi dimensional Labeling FFNN/HMM model. The reason might be due to the heavy computations that have to be performed in the formation of codebooks and labels.

The system was further tested in a clinic with five doctors and 80 reports. The real time clinic experiments also supported the simulated environment experiments indicating that multi dimensional labeling FFNN/HMM system is the top most in terms of recognition rate. Further, the findings also revealed that the report preparation speed increased with the implementation of CSR. The best performance was given by FFNN/HMM model, followed by multi-dimensional labeling FFNN/HMM model.

6.2. SUGGESTIONS FOR FUTURE DEVELOPMENT

- A hybrid variety of EM and K-means clustering is used as a reducing agent in search space. Other combinations like Learning Vector Quantization (LVQ) and HMM, can be applied and its performance can be studied.

- The speed of speech recognition can further be improved by analyzing other types of hybrid systems that apply fuzzy concepts in labeling and codebook generation process.

- The system assumes that all the speech data captured during identification and recognition are from clean noise-free environment. Performance degradation occurs if the data has noise. Noise reduction and enhancement techniques can be applied as pre-processing techniques to improve the performance of the system.