This thesis primarily addresses the problem of automatic speech recognition for recognising isolated words in the Malayalam language. Since the performance of a speech recognition system relies on the pre-processing steps, feature extraction techniques adopted, post processing methods applied on the feature vector set obtained and the pattern classifiers used, the main objective of this work is to build a speech recognition system with maximum recognition accuracy. To achieve this goal, new algorithms and improvements are necessary at each stage of the speech recognition process. Different performance assessment measures are employed for evaluating the performance of the speech recognition systems developed. These representation models are proved to be effective in improving the recognition rate. So this research work proposes new enhanced algorithms and improved techniques for building an efficient speech recognition system.

9.1 Summary of the Research Work

In spite of the advances in technology and hardware during the last few decades, Automatic Speech Recognition is still a challenging and difficult task when it comes to real world applications. Now-a-days, speech recognition has wide applications in almost all fields of life. Due to this wide variety of applications, the requirements for each application are different. Researchers
are therefore trying to explore effective ways to build efficient speech recognition systems for each application. This work intends to enhance the performance of the already existing methods to improve the recognition rate.

From the literature survey conducted, it was not possible to rule out a specific technique, which always produces the best results. So in this research work, first a study is carried out to find a suitable combination of techniques for the efficient recognition of the speech samples in the databases created for the Malayalam language. The present work can be considered to have two phases. In the first phase, sixteen different speech recognition systems are developed to find the best feature extraction technique and pattern classifier combination which produced the best recognition rate for the speech databases created in Malayalam. During the second phase, new enhanced algorithms and improvements are proposed, designed and developed to further improve the recognition rate. The major highlights of both phases are given below.

The main highlights of the activities in the first phase of the research work are:

- Creation of three databases in Malayalam for vowels, digits and isolated words with 100 speakers, 200 speakers and 1000 speakers respectively.
- Exploitation of End Point Detection algorithm and Pre-emphasis Filters for noise reduction and speech enhancement.
- Development of a speech recognition system using Linear Predictive Coding (LPC) parameters and its implementation using four different classifiers like Artificial Neural Networks (ANN), Support Vector Machines (SVM), Hidden Markov Models (HMM) and Naive Bayes Classifiers.
Future Directions and Conclusion

- Evaluation of the consistency of the features based on recognition accuracy depending on the number of speakers.
- Implementation of Mel Frequency Cepstral Coefficients (MFCC) features for the recognition of Malayalam speech databases using different classifiers.
- Exploitation of wavelet denoising algorithms based on Soft Thresholding for denoising of signals.
- Experiments done for selecting the best wavelet family and mother wavelet for the databases created.
- Design and implementation of a speech recognition system for the three databases created using wavelet based Discrete Wavelet Transforms (DWT) feature extraction technique using various classifiers.
- Evaluating the performance of the speech recognition system developed using Wavelet Packet Decomposition (WPD) feature vector set and the above given classifiers.
- Comparison of the performance of these feature extraction techniques and classifiers to select the combination of feature extraction technique and classifier which performed best for Malayalam in terms of recognition accuracy.

The main highlights of the activities in the second phase of the research work are:

- Introduction of a new improved algorithm for feature extraction called Discrete Wavelet Packet Decomposition (DWPD) for the better performance of the speech recognition system. The characteristics of
different feature extraction techniques are combined to create a new hybrid algorithm which can produce better results based on recognition accuracy.

- Exploitation of Principal Component Analysis (PCA) technique for effective reduction of the dimension of the feature vector set without much affecting the recognition accuracy. PCA provides the dominant traits on which greater emphasis is laid.

- Devising of a new algorithm for smoothing the speech signals before pre-processing for reducing the amount of noise present in the signals. Noise is omnipotent and an algorithm that can weed out the sudden spikes which are only mere disturbances helps in enhancing the quality of the original signal. This proposed Adaptive Smoothing algorithm along with the wavelet denoising method based on Soft Thresholding called Adaptive Smoothing Soft Thresholding (ASST) helps in the enhancement of the signal by reducing the Signal-to-Noise Ratio and the error produced.

- Formulation of a new method for post processing based on the statistical technique of Three Sigma Control Limits which utilises the features of Mean. Statistical methods provide better uniformity among the data values in the feature vector set and also ensure reliable identification.

- Introduction of a new method for limiting the range of data using statistical technique based on Quartiles which uses the characteristics of Median. The performance of the feature vectors varies with the selection of the range. Choosing a range that can even out the variations in the data set helps in better analysis and classification.
Future Directions and Conclusion

- Devised a new method that can be applied during the post processing stage based on the statistical mode calculation.

- Investigation of ensemble classifiers based on three ensemble learning methods namely Bagging, Boosting and Stacking by combining the various classifiers for the better performance of the system. This hybrid architecture of combining different classifiers using different ensemble learning methods overcomes the limitations of using a single classifier. Ensemble learning utilises the best classifier combinations based on two popular schemes namely Voting and Stacking.

- Achievement of encouraging and improved recognition rates for all the four stages of development of a speech recognition system using the proposed new algorithms and methods applied during the pre-processing, feature extraction, post processing and classification stages.

9.2 Future Directions

In this research work, we have designed a speech recognition system with a fair degree of accuracy. Our main emphasis was on speech recognition for isolated words which finds applications in industry and man-machine interfaces. However, there is scope for further research in this field and some of the future prospects are listed below.

- **Extending the work to Continuous speech recognition:** This works concentrates only on the recognition of isolated words. The new proposed algorithms can also be tried on continuous speech since it also includes pre-processing, feature extraction, post processing and classification modules.
• **Expanding the work to large vocabulary systems:** This research work was carried out for only medium number of data such as 12 vowels, 10 digits and 20 isolated words. In future, this can be expanded to large vocabulary systems.

• **Extending the work to speaker recognition:** This work focuses only on recognising speech. Another area to which this research work can be extended is of speaker recognition. Speaker recognition deals with identifying the speaker instead of recognising what he says.

• **Extending the work to language independent speech recognition system:** This work is meant to design efficient speech recognition for Malayalam language. This work can also be extended to different languages since the architecture of the speech recognition system is the same.

**9.3 Conclusion**

Speech recognition is a complicated task and state-of-the-art recognition systems show that its performance depends on many factors like the number of speakers, the database used and the different techniques adopted during the different stages of development of the system. The main intention of this research work is to build a speech recognition system for recognising speaker independent isolated words in Malayalam with utmost recognition accuracy. So databases are created in Malayalam and experiments are performed therein. From the literature study, it was not possible to select a specific combination of the feature extraction method and a classifier which always generated good results. Hence sixteen different experiments were carried out for selecting the best combination with the best recognition rate using 4 feature extraction techniques and 4 pattern classifiers. Among these techniques, DWT and MLP combinations were found to produce the best
results. Among the different wavelet families available, better performance was obtained using the Daubechies family of wavelets with order 4 (db4).

In this research work, new algorithms and improvements were proposed, designed and developed during the four stages of development of the speech recognition system. The proposed hybrid algorithm DWPD which was developed during the feature extraction stage by combining the features of both DWT and WPD produced improvements in the degree of recognition accuracy. The newly proposed adaptive smoothing technique which was applied during the pre-processing stage played a significant role in removing the sudden spikes due to noise, thus improving the SNR value. These smoothed signals when applied to wavelet denoising using Soft Thresholding, yielded better recognition accuracy. All the three statistical thresholding techniques proposed during the post processing stage based on Three Sigma Limits, Quartiles and Confidence Interval Mode were found to be efficient in selecting the feature vectors for pattern recognition. These techniques were employed to bring the feature vectors to a particular predefined range. Among these three methods, the results obtained using Quartiles were proved to be superior. It was observed that the ensemble learning methods based on Bagging, Boosting and Stacking which were applied during the classification stage also generated better results. Thus the newly proposed algorithms and improved techniques performed well and produced better results for the speech recognition system developed for Malayalam.

The thesis findings can be used for different purposes with variations in language, databases etc. The consequent applications derived from the thesis findings will be on the rise.
The thesis findings can be used to develop similar applications in foreign languages and it can be used in Automated Teller Machines (ATMs) for dispensing cash and other financial transactions across all languages. This will allow foreigners as well as citizens of the country, a large degree of freedom from language-barriers in conducting financial transactions. The spoken digits recognition system has great relevance in this field.

An efficient speaker independent isolated words recognition system has a number of applications in different fields. It has applications across the Internet in helping farmers and other less literate people to access international markets for information on commodities and future pricing positions. It has also a wide range of applications in robotics where actions and tasks can be executed/cancelled using voice commands in native languages. The need for high-level technical expertise in accessing state-of-the-art technologies can be transgressed and brought down within the reach of the common man. The automotive industry which uses robots can enhance their productivity by including voice detection. Strategically placed in airports and other places of public interest, it can be used to locate criminals whose voice data is already available. It can help investigating agencies in tracking down criminals. It can be used to supplement security measures by bringing in additional check measures to establish authenticity. The spoken words recognition system is of great relevance in this context.

The findings can also be used in our quest to derive unspoken words from an existing dataset. A database of consonants and vowels can be developed and maintained to create speech signals by combining the consonants and vowels. This can mimic an undelivered speech and identify potential speakers for a particular theme or event. A database of all consonants
and vowels can create a dictionary of all words that can be spoken by an individual. This can be compared with actual spoken words to verify identity.

Presently all dictionaries are language specific. English-English, English – Malayalam etc. With a voice recognition system in place, we can develop dictionaries that can transcend language barriers. Suppose we have an English – Malayalam dictionary and Malayalam – Hindi dictionary, an English – Hindi dictionary can be only a few seconds away in the hands of a software personnel. In a similar way we can develop dictionaries against all languages that will ultimately remove barriers of language. The works of speech translators can also be made easy. In the recent past we have seen translators working hard during visit by foreign dignitaries. A speaker and a hearing mechanism can covert alien language to ones own mother tongue without waiting for a translator.

The music industry also finds varied uses. The ragas developed in ancient ages are prone to dilution in the hands of inexperienced artists. Music competitions are won based on the judgements by the judges who may or may not be right. Presently the ragas and other intricate music systems can be verified using voice recognition systems. The feature vectors corresponding to the original raga can be stored and compared with the performances by artists.

Finally, this research work has been a comprehensive approach for the development of a speech recognition system with emphasis on all the different aspects namely, the pre-processing steps, feature extraction techniques, post processing methods and the classification techniques. No one technique is perfect in itself and we have therefore adopted a hybrid architectural approach.