J Inf Process Syst, Vol.13, No.5, pp.1135~1148, October 2017
https://doi.org/10.3745/JIPS.02.0033

ISSN 1976-913X (Print)
ISSN 2092-805X (Electronic)

JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Speech Query Recognition for Tamil Language Using Wavelet and Wavelet Packets

P. Iswarya* and V. Radha*

Abstract

Speech recognition is one of the fascinating fields in the area of Computer science. Accuracy of speech recognition system may reduce due to the presence of noise present in speech signal. Therefore noise removal is an essential step in Automatic Speech Recognition (ASR) system and this paper proposes a new technique called combined thresholding for noise removal. Feature extraction is process of converting acoustic signal into most valuable set of parameters. This paper also concentrates on improving Mel Frequency Cepstral Coefficients (MFCC) features by introducing Discrete Wavelet Packet Transform (DWPT) in the place of Discrete Fourier Transformation (DFT) block to provide an efficient signal analysis. The feature vector is varied in size, for choosing the correct length of feature vector Self Organizing Map (SOM) is used. As a single classifier does not provide enough accuracy, so this research proposes an Ensemble Support Vector Machine (ESVM) classifier where the fixed length feature vector from SOM is given as input, termed as ESVM_SOM. The experimental results showed that the proposed methods provide better results than the existing methods.

Keywords

De-noising, Feature Extraction, Speech Recognition, Support Vector Machine, Wavelet Packet

1. Introduction

Digital Speech is naturally most comfortable mode of communication in the field of Human Computer Interaction (HCI). Voice recognition is a task of decoding human speech into digitized form of speech that can be interpreted by device of a computer. Automatic Speech Recognition (ASR) system becomes complex due to variations of speaker, speaking style, environment, noise etc. Despite its limitations, speech recognition technology is valuable tool in many applications like live subtitling on television, dictation in Medical transcriptions, Command Control in Robotics, speech to text conversion for note making systems, replacement of keyboard and mouse for physically or visually challenged people. In noisy environment the accuracy level of ASR system will suffer greatly [1]. To recover original speech from noisy speech signal several speech de-noising methods are available. The wavelet based de-noising method is simple, and it analyze signal at different scales to remove noisy coefficients [2]. Wavelet based noise removal method removes corrupted noise using threshold

Manuscript received July 29, 2014; first revision November 4, 2014; second revision August 5, 2015; accepted August 5, 2015; onlinefirst December 17, 2015.
Corresponding Author: P. Iswarya (iswaryacbe333@gmail.com)
* Dept. of Computer Science, Avinashilingam Institute for Home science and Higher Education for Women, India ({iswaryacbe333, radhasrimai}@gmail.com)

algorithms for speech enhancement.

Generally hard thresholding and Soft thresholding performs better for external noise and internal noise respectively. But the proposed combined thresholding based noise removal technique provides better results for both internal and external noises.

Feature Extraction (FE) is one of the significant steps in ASR system which transforms original signal in to a form that is appropriate for the classification model. Traditionally most widely used feature extraction technique in the field of speech recognition is Mel Frequency Cepstral Coefficient (MFCC) and Linear Predictive Coding (LPC). However usage of MFCC has several issues and the experimental results showed that this technique does not work well in noisy speech environment [3]. Also to extract features, speech signal is divided into fixed size of frames and in these static size frames sharp transition that occur in signal cannot be analyzed [4,5]. Due to the limitations in MFCC and LPC, Wavelet Packet based Cepstral Coefficients (WPCC) is proposed. Support Vector Machine (SVM) is simple nonlinear classifier and has larger flexibility in handling classification task. Instead of tuning single model for better accuracy, combining the predictions of the homogeneous ensemble SVM model may improve performance further. The process of feature extraction results in variable length of feature vector for each isolated word. To convert variable size feature vector into fixed size feature vector Self Organizing Map (SOM) is used as an input to be fed into the ensemble classifier [6].

The paper structure as follows. Section 2 discusses about some of the related works in Tamil Speech recognition and Section 3 explains about the methodology of proposed system. In Section 4 wavelet based de-noising and silence removal methods are used for speech enhancement. Section 5describes briefly on LPC, MFCC and proposed wavelet packet based MFCC feature extraction. Section 6 gives detailed description on SVM, homogeneous ensemble SVM with SOM. Section 7 presents the recognition results of experiments with and without de-noising procedure. Finally conclusion is drawn in Section 8.

2. Related Works in Tamil Speech Recognition

Tamil is highly inflectional language than English. While building large vocabulary Tamil speech recognition system there exists a problem of excessive growth of vocabulary, caused by different forms of word derived from base word. To reduce the size of vocabulary, modified morpheme based language model for Tamil speech recognition was proposed which worked well for Tamil speech database than traditional N-gram language model [7]. In the paper [8], authors developed speech recognizer for written Tamil, because spoken Tamil varies from region to region and person to person. The word is recognized using new lexicon and top down parsing technique. They discussed about features of Tamil that affect the performance of speech recognition engine. In the paper [9], authors worked on isolated words for Tamil spoken language, here input signal are preprocessed using four types of filters, and from best filter output, LPCC feature extraction was done. The classification and recognition adopted using back propagation neural network, which has produced better results for limited vocabulary. In this paper [10], authors proposed the conventional approach, low frequency MFCC vectors are extracted and interrogated with frequency sub band decomposition. The implemented system shows better efficiency than existing MFCC method. In the paper [11], authors developed a speaker dependent

continuous speech recognition system for Tamil. The proposed method segments words from sentences and then character from words. The back propagation algorithm is used for training and testing a system. The system was tested for segmenting words from 9 spoken sentences and achieves accuracy of 80.95%.

In the paper [12], authors presented novel speech word recognition system for Tamil which consists of three stages. The first input speech signal is preprocessed using Gaussian filtering technique. From noiseless signal MFCC feature vectors are extracted from training dataset and test dataset. Then Feed Forward Backpropagation Neural Network (FFBNN) undergoes training and testing with their respective datasets. The performance of proposed technique gives better recognition results than existing HMM and Associative ANN technique. In the paper [13], authors formulated a speaker dependent medium sized vocabulary Tamil Speech recognition mechanism. Here the system was trained and tested with HMM and Auto Associative Neural Networks (AANN) using 8,000 and 2,000 samples respectively. The MFCC feature extraction techniques are applied to input speech samples to extract feature vectors. The performance states that HMM with 5 states and 4 mixtures yields high recognition performance than AANN. In the paper [14], authors proposed the speaker independent isolated Tamil words recognition system using Discrete Wavelet Transform (DWT), and multilayer perceptron network trained with back propagation training algorithm. The db4 type of wavelet used for wavelet based feature extraction. Then the speech samples in database successively undergo an eight level decomposition to obtain approximation and detail coefficients. Here 70% of data used for training, 15% for validation, 15% for testing and finally it achieves overall recognition accuracy of 90%. In the paper [15], authors developed the speaker independent Triphone based medium vocabulary continuous speech recognizer for Tamil language. The implementation of the system is done with Sphinx-4 framework of HMM model with 3 emitting and 1 non-emitting states with continuous density of 8 Gaussian per state was used. They constructed a phoneme based context dependent acoustic model for 1,700 unique words, then pronunciation dictionary with 44 base phones and triphone based statistical language model. The system results in good word accuracy and same word error rate for train and test utterances. In the paper [16], authors were improved the accuracy of Tamil speech system by designing language models at various level such as segmentation phase, recognition phase and syllable and word level error correction phase. They improved the recognition accuracy at each phase and finally 87.1% accuracy was obtained. In the paper [17], authors developed speaker independent isolated Tamil digits recognition used and achieved overall recognition accuracy of 91.8%. From input speech signals MFCC feature vectors are extracted and trained using Vector Quantization (VQ) approach. The codebook for each digit is generated using Linde-Buzo-Gray (LBG) VQ training algorithm. In the paper [18], authors designed the framework for Tamil speech based query processing architecture to retrieval English textual documents. They integrated speech recognition and cross language text retrieval system.

From several related works of Tamil Speech recognition, it is found that many of the research were carried out using MFCC, LPC, and wavelet based feature extraction techniques. Also for recognition purpose Hidden markov model and neural networks were used by many authors. Then few papers made use of noise filtering techniques for noise removal. This paper concentrates on enhancing the methods at various stages such as preprocessing, feature extraction and classification which provides better results than conventional techniques.

3. Proposed Methodology

The first step in speech recognition is preprocessing speech signals which reduce noises based on Combined Thresholding wavelet noise removal algorithm. The next task is to eliminate silence and to identify only the presence of spoken word from beginning to end of the sentence. In second step signal is divided into short chunks called as frames. Then each frame is decomposed using discrete wavelet packet transform, instead of doing Discrete Fourier Transform (DFT) in MFCC based feature extraction. To avoid distortions in speech signal cepstral mean normalization technique is applied [6]. In order to make fixed length trajectory model input to ensemble SVM classifier, SOM is applied on feature vectors. The proposed speech recognition system is shown in Fig. 1.

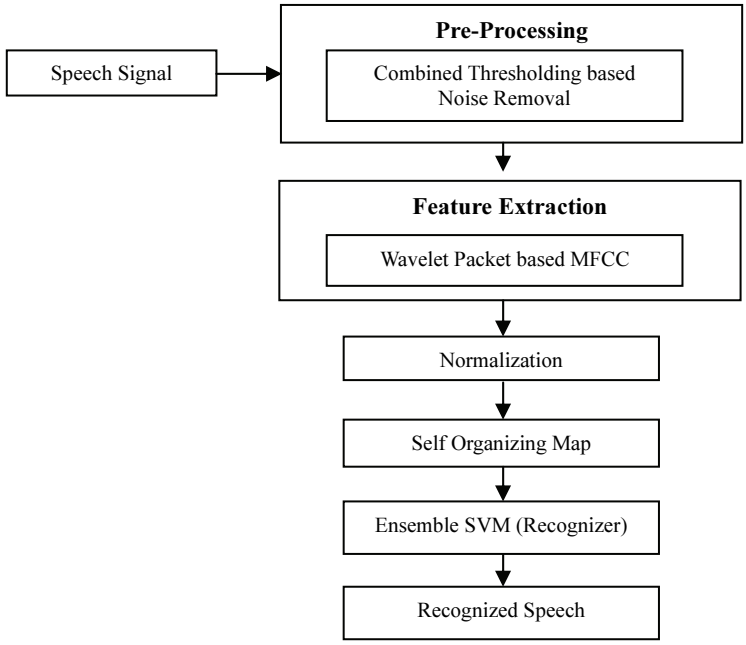


Fig. 1. Proposed speech recognition system.

4. Pre-processing in Speech Signal

Speech de-noising is a process of eliminating corrupted noise from original speech signal to improve recognition accuracy. Noise may available in different form depending upon the environmental conditions. In this paper experiments are conducted in class room environment, so there may be presence of white noise, babble noise and external noise such as pen click sound, knocking table noise, were analyzed at -10 dB, -5 dB, 0 dB, 5 dB, 10 dB SNR levels.

Wavelets are successfully applied in the field of speech recognition [2] using wavelet transform analysis. DWT uses multi-resolution technique to analyze signal at different frequency bands. DWT is applied on original speech signal and order the coefficients in increasing frequency. Selection of threshold is important issue in de-noising; choosing a too high level threshold value will remove some of the original content and to choose low threshold value will not remove noise properly. Donoho and

Johnstone [19] developed a minimax thresholding method for calculating threshold value which is simple, proper and more conservative. In this paper, hard thresholding, soft thresholding and combined thresholding are used to estimate wavelet coefficients in wavelet threshold de-noising. To obtain reconstructed noise removed signal inverse discrete wavelet transform is performed. De-noising procedure is represented in Fig. 2. Trial and error method is used to select the threshold value, T. The value of T in Eq. (1), Eq. (2) and Eq. (3) is 0.01, 0.1 and 0.3 respectively.

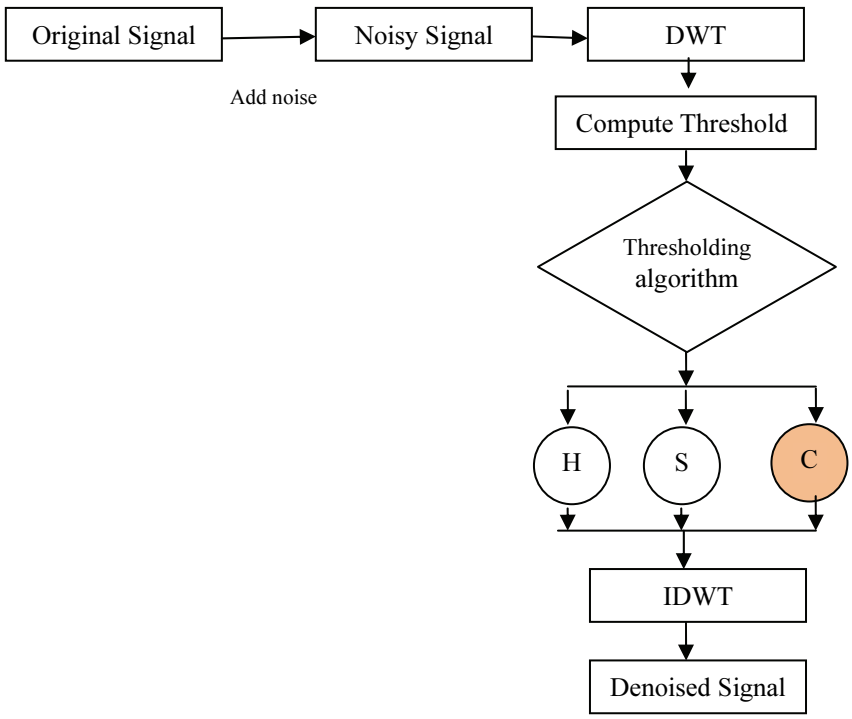


Fig. 2. Speech de-noising procedure (H=hard, S=soft, C=combined).

4.1 Hard Thresholding

Hard thresholding is process of setting the selected coefficient values to zero, if their absolute values are lesser than threshold value T, otherwise it keep the coefficients without alter.

$$T_{Hard}(coeff) = \begin{cases} if\ coeff(i) < T & then\ coeff(i) = 0 \\ else\ coeff(i) = coeff(i) \end{cases} \quad (1)$$

4.2 Soft Thresholding

Soft thresholding sets some of signal coefficients to zero elements whose coefficient values are lesser than threshold value, and if absolute values are greater than threshold, it shrinks the coefficients to values closer to zero, using threshold value.

$$T_{Soft}(coeff) = \begin{cases} if\ coeff(i) < T & then\ coeff(i) = 0 \\ else\ Sign(coeff(i)) * (coeff(i) - T) \end{cases} \quad (2)$$

4.3 Proposed Combined Thresholding

The proposed method of combined thresholding is combination of both hard thresholding and soft thresholding. If coefficient values are below threshold, then it set selected coefficients to zero, and if absolute value exceeds the threshold level, it applies combined thresholding.

$$T_{Combined}(coeff) = \begin{cases} if\ coeff(i) < T & then\ coeff(i) = 0 \\ else\ Sign(coeff(i)) * (coeff(i) - T) + coeff(i) \end{cases} \quad (3)$$

5. Feature Extraction

The most dominant approaches in the field of speech recognition are Mel frequency cepstral coefficients and linear predictive coding [5]. Usually in speech signal, lower frequency has greater magnitude where as higher frequency is lesser in magnitude. Therefore to emphasis high frequency content, pre-emphasis filter is used for compensation, and it is carried out using Eq. (4). The speech signal of Tamil word 'amma' before and after pre-emphasis is presented in Fig. 3.

$$F_{preem}(z) = 1 - a_{preem}(z^{-1}) \quad (4)$$

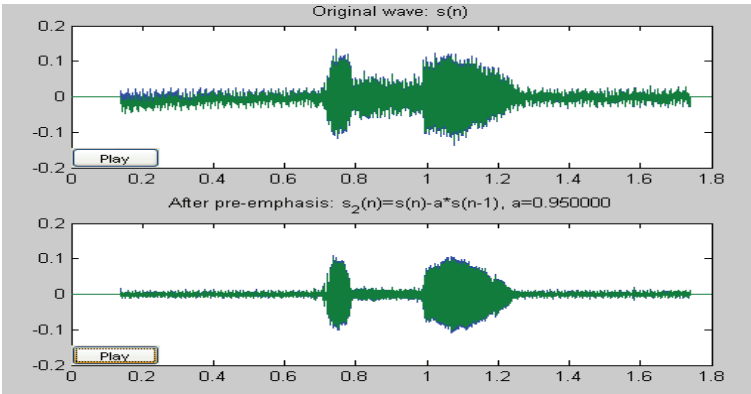


Fig. 3. Speech waveform of Tamil word "amma" before and after pre-emphasis filter.

To analyze speech, the pre-emphasized signal is divided into short intervals called frames. Each frame consists of M samples which are represented in form of feature vector. At the beginning and end of the frame, signal discontinuities may occur to minimize distortions hamming windowing function is applied using Eq. (5) [20].

$$w(n) = 0.54 - 0.46\cos(\frac{2n\pi}{N-1}) \quad 0 \le n \le N\text{-}1 \quad (5)$$

5.1 Linear Predictive Coding

Initial step in LPC is to pre-process a speech signal that includes framing and windowing. In LPC unknown sample parameters can be approximated using linear combination of past speech samples. Estimate LPC coefficients [20], by assuming orthogonality principle of Yule walkers equation and further it is solved using auto correlation method. The resulting equation is solved by Levinson Durbin's recursive procedure. The procedure works recursively and finally, all the prediction coefficients were obtained in a descending order of p. To compute LPC it is necessary to minimize sum of squared difference between actual ones and estimated samples of linear ones.

5.2 Mel Frequency Cepstral Coefficients

The first step in MFCC feature extraction is to undergo pre-emphasis, framing and windowing. Then to convert time domain samples to frequency domain samples DFT is applied. The frequency spectrum of the signal does not follow a linear scale, and also pitch in frequency measured using Mel scale. The set of triangular filters in Mel scale is called Mel scale filter bank, which is linear below 1,000 Hz, therefore logarithmic frequency spacing is done above 1,000 Hz. The Mel scale is defined in below equation as

$$Mel(f) = 2595 * log_{10}(1 + {}^{f}/_{700}) \tag{6}$$

Finally MFCCs are obtained, by converting frequency domain samples back into time domain samples using Discrete Cosine Transformation (DCT) which is defined in Eq. (7). The double delta MFCC of 39 dimensions per frame are extracted and used for experiments.

$$c_i = \sqrt{\frac{2}{N}} \sum_{k=1}^{N} m_k \cos\left(\frac{\pi_i}{N}(k\text{-}0.5)\right) \tag{7}$$

5.3 Proposed Discrete Wavelet Packet based MFCC

The Wavelet Packet (WP) Transform provides multi-resolution property over fixed time frequency resolution of Fourier transform. An use of non stationary signal in wavelet transform analysis provide better results than using Fourier transform, because WPT able to analyze unvoiced and nasal sound in the signal, which is not done in DFT based MFCC [2]. In proposed feature extraction technique part of DFT in MFCC is replaced with Wavelet packet based transform that the model is termed as WPMFCC and it is shown in Fig. 4.

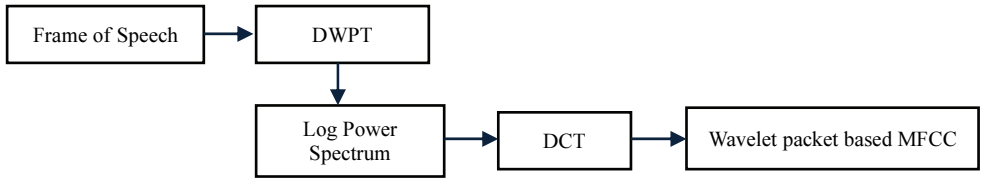


Fig. 4. Wavelet packet based MFCC.

The WP is computationally efficient alternative with sufficient frequency resolution. The WP decomposes both approximation and detail coefficients, then two orthogonal bases at a parent node (depth, j; number of subspaces, p) defined by

$$\varphi_{j+_1}^{2p}(k) = \sum_{n-\infty}^{\infty} h[n]\varphi_j^P\,(k - 2^j n) \tag{8}$$

$$\varphi_{j+_1}^{2p}(k) = \sum_{n-\infty}^{\infty} g[n]\varphi_j^P\,(k - 2^j n) \tag{9}$$

$h[n]$ and $g[n]$ correspond to low-pass and high-pass filters, respectively, given by following equation.

$$h[n] = \langle \varphi_{j+1}^{2p}\,(u), \varphi_j^P\,(u - 2^j n)\rangle \tag{10}$$

$$g[n] = \langle \varphi_{j+1}^{2p+1}\,(u), \varphi_j^P\,(u - 2^j n)\rangle \tag{11}$$

Thus by utilizing j level wavelet packet decomposition fully a symmetric binary tree structure is formed having more than 2^{2j-1} orthogonal bases. The proposed WP based MFCC feature extraction able to derive efficient, noise robust features that can be extracted from frequency sub-bands of wavelet packet.

6. Recognition Using Machine Learning Approach

6.1 Support Vector Machine

The machine learning models are simple and easily adaptable, especially supervised learning model in machine learning approach used in many applications like text categorization, Information filtering, Speech recognition, parsing, etc. [21]. The supervised learning models learn from known input data and known response data to build a predictor model that will generate response for new data. SVM is also a supervised learning model which is simple and performs better than Artificial Neural Network (ANN) in many cases [22]. The author Vapnik introduced a SVM classifier, and this initially developed to work for linear problems. Later linear SVM classifier is extended to handle non-linear data, and also binary SVM to multiclass SVM. The SVM learning for linear classifier defined as

$$F(x) = W^T X + b \tag{12}$$

The general category of SVM is kernel function, for linear data, dot product is used as kernel algorithm. For nonlinear type of data, kernel function varies depending upon the application data [23]. The speech recognition system has several words and each word represent a class so that multiclass SVM with polynomial kernel function is implemented. SVM parameters are set as given below.

SVM_Params={'-*b* 0 -*c* 100 -*d* 3 -*g* 1 -*t* 3 '}

Here, -b indicates the probability estimates and takes the values 0 or 1 to indicate whether to train a SVC (Support Vector Classification) or SVR (Support Vector Regression) model for probability estimates, 0 or 1. For our method we used SVC. To allow some flexibility in separating the categories, SVM models have a cost parameter, C, that controls the tradeoff between allowing training errors and forcing rigid margins. It creates a soft margin that permits some misclassifications. Increasing the value of C increases the cost of misclassifying points and forces the creation of a more accurate model that may not generalize well. The cost parameter is set to 100. The next parameter, d, is degree in kernel function which is assigned a value 3, while g (gamma in kernel function) is assigned a value of 1. The last parameter '-t' is used identify the kernel function to be used. The kernel function, $K(x_i, xj) \equiv (x_i)^T \phi(x_j)$, can belong to any one of the following four types available.

1. Linear Kernel : $K(x, x_j) = x_i^T x_j$
2. Polynomial Kernel : $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d$, $\gamma > 0$.
3. Radial Basis Function (RBF) : $K(x_i, x_j) = exp(-\gamma||x_i - x_j||^2)$, $\gamma > 0$.
4. Sigmoid Kernel : $K(x_i , x_j) = tanh(x_i^T x_j + r)$.

Here, γ, r and d are kernel parameters. For our method we used the polynomial kernel function.

6.2 Proposed Ensemble Support Vector Machine with Self Organizing Map (ESVM_SOM)

The use of multiclass in ensemble SVM shows better recognition performance than using single SVM [24]. In this paper ensemble SVM works based on boosting technique and it does training to each SVM with different training samples that will be chosen according to the manner in which the probability of distribution is updated. Then the individual SVM predicted results are unified using majority voting based approach. The boosting algorithm used in ensemble is adaboost algorithm which selects different training samples for learning. Adaboost algorithm [24] is to build a strong classifier by linear combination of weak classifier $h_t(x)$.

$$f(x) = \sum_{t=1}^{T} \alpha_t \; h_t(x) \tag{13}$$

Simple majority voting based approach is used for combining the result of several SVM. Then the final class decision is assigned to a sample by predicting which will be voted for the most by majority of classifiers.

The process feature extraction results in variable length of feature vectors where SOM is a neural network that convert varying size into fixed size of feature vectors that will fed into the classifier as inputs. Then the use of SOM with Ensemble SVM improved the recognition accuracy and minimizes the training time further. SOM is unsupervised learning method that works based on competitive leaning strategy. The SOM algorithm [25] uses as input the variable length feature vector and maps it to a constant size of six clusters while preserving the input size. The algorithm consists of three tasks, namely, Competitive task, Cooperative task and Adaptation task. This section presents details of the conventional SVM and ensemble classifiers.

7. Experimental Results and Discussion

The experiments were conducted using Tamil queries taken from Forum for Information Retrieval and Evaluation (FIRE) dataset 2011. Fifty short Tamil title topic queries uttered by 20 persons with 3 repetitions total of 3,000 sentences used for training, and 10 persons with 2 repetitions total of 1,000 sentences used for testing. The silence removal and segmentation of speech are done by computing signal energy and spectral centroid features.

The first metric used during evaluation of preprocessing algorithms is Signal to Noise Ration (SNR). SNR is used to quantify how much a signal has been corrupted by noise. It is defined as the ratio of signal power to the noise power corrupting the signal. The SNR is calculated in two ways one is Pre SNR and other is Post SNR which are obtained before and after applying the preprocessing operation. De-noising is successful if Post SNR is higher than Pre SNR. Eq. (14) presents the formula used to estimate SNR.

$$SNR_{db} = 10log_{10}\left(\frac{P_{signal,dB}}{P_{noise,dB}}\right) = P_{Signal,dB} - P_{noise,dB} \tag{14}$$

Mean Square Error (MSE) is and is used to quantify the difference between values implied and the true being estimated. The input MSE is defined using Eq. (15) and output MSE is defined using Eq. (16), where x_i is the original signal, y_i is the noisy signal and $\bar{x}_i$ is estimated xi (noisy signal y passed through de-noising algorithm). Lower MSE indicates a closer match between the two signals.

$$Input\ MSE = \frac{1}{N}\sum_i (X_i - Y_i)^2 \tag{15}$$

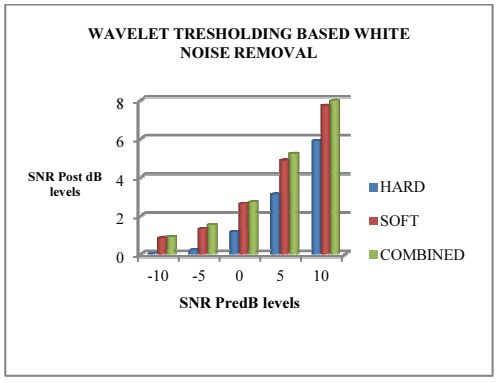
$$Output\ MSE = \frac{1}{N}\sum_i (X_i - \bar{X}_i)^2 \tag{16}$$

The next performance metric, accuracy is estimated using Eq. (17). A high accuracy value indicates maximized speech recognition performance.

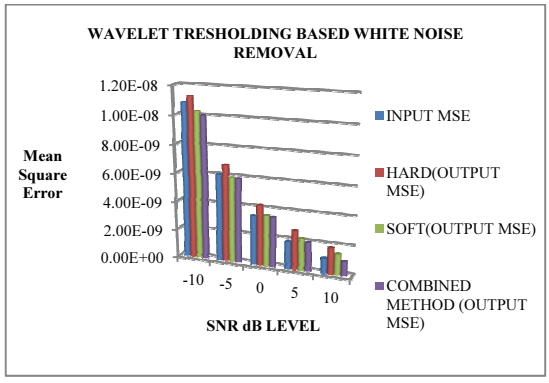
$$Accuracy\ (\%) = \frac{No.of\ words\ correctly\ recognized}{Total\ No.of.Words} \times 100 \tag{17}$$

The "connected component" Speech consist of two or more words which are segmented. The first stage in speech recognition is de-noising a speech signal using hard thresholding, soft thresholding and combined thresholding methods. The signal is analyzed at different SNR dB levels by implementing three wavelet thresholding algorithms and corresponding results are shown in Fig. 5(a)-(f).

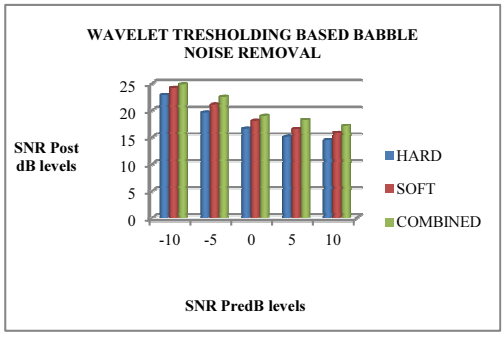
The result shows that, compared to hard thresholding, the soft thresholding attains high PSNR and low MSE values for white noise and babble noise removal, but for external noise, hard thresholding performs better. The proposed combined thresholding technique removes all noises effectively than other two de-noising methods at different dB levels. Each noise has its own desired characteristics, and the white noise of non-stationary signal makes large or rapid changes in its spectrum over time. Thus makes the Thresholding algorithm cannot remove high noise dB levels efficiently, where as other noise does not make rapid changes over time.



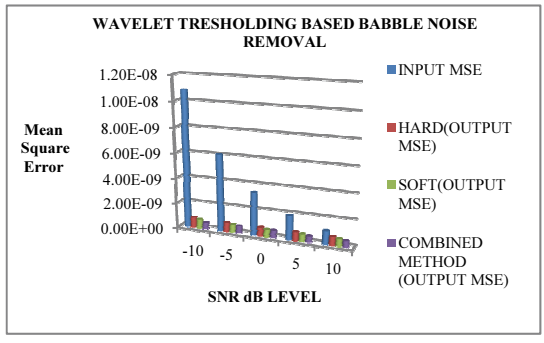
(a)



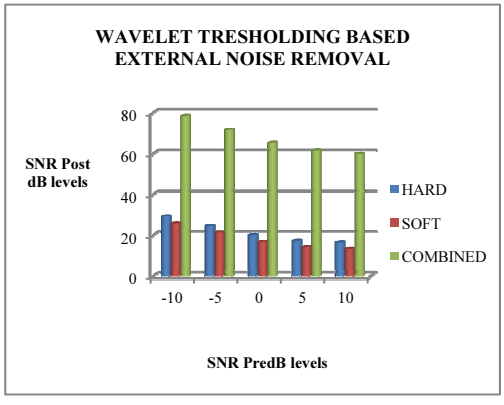
(d)



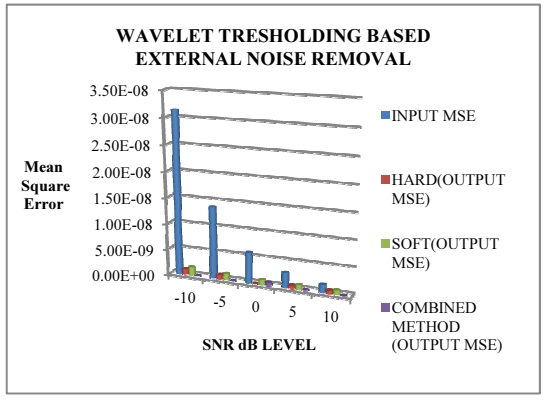
(b)



(e)



(c)



(f)

Fig. 5. (a-c). represents SNR pre dB and SNR post dB levels, and **(d-f)** represents input MSE and output MSE values of hard, soft, combined thresholding methods respectively.

To analyze the effectiveness of LPC, MFCC and proposed WPMFCC feature extraction techniques, their corresponding feature vectors are given as input to the Support Vector machine classifier. Each feature extraction technique performance is analyzed, by applying de-noising, and without de-noising technique. The feature extraction results are shown in Fig. 6.

The Fig. 6 results indicate that the proposed WPMFCC with level 4 decomposition attain good results when compared with standard MFCC and LPC technique. The speech query recognition accuracy was

tested with and without de-noising. MFCC does not work better under noisy conditions, and their percentage of accuracy is almost similar to accuracy of LPC, but it achieves more recognition rate after de-noising.

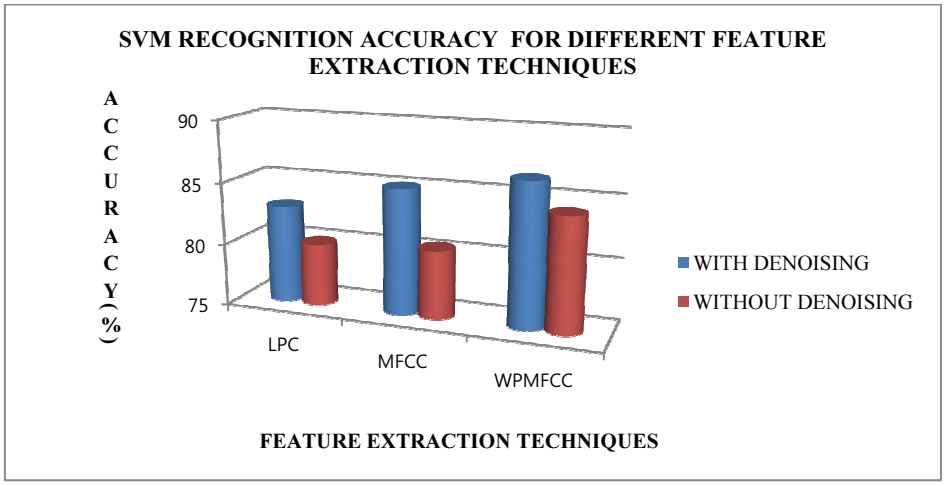


Fig. 6. Average speech recognition accuracy of different feature extraction techniques with Support Vector Machine (SVM) classifier on entire test dataset (1,000 sentences).

After finding best feature extraction technique, it is necessary to improve the effectiveness of classifier performance. The proposed WPMFCC feature extraction technique fed as input to the SVM, ESVM and ESVM_SOM classifiers. The different classifier results with WPMFCC feature extraction technique was presented in Fig. 7.

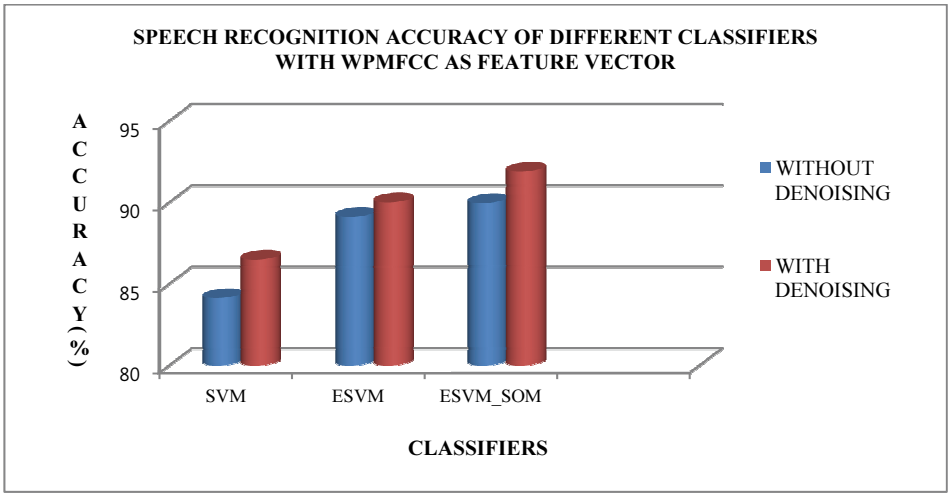


Fig. 7. Average speech recognition accuracy of different classifiers on test dataset (1,000 sentences).

The results show that Ensemble classifier provides more accuracy than single SVM and also fixed feature vector with ensemble SVM gives even better results. The experimental analysis shows that WPMFCC with ESVM_SOM gives average better recognition accuracy of 92%.

8. Conclusion

The proposed methods of combined thresholding, WPMFCC and ESVM_SOM outperform other existing methods. The combined thresholding based noise removal works out well for both external and internal noises. The discrete Fourier transformation has fixed time resolution property over speech signal, that does not provide satisfactory results, but wavelet packet transform provides multi resolution analysis. Also use of fixed length trajectory vector as input to the ensemble SVM increased the recognition accuracy further, than using varied size feature vector in ESVM. Further improvements can be done on proposed method with varying wavelet structure and wavelet family to attain maximum accuracy.

References

[1] A. G. Chitu, L. J. Rothkrantz, P. Wiggers, and J. C. Wojdel, "Comparison between different feature extraction techniques for audio-visual speech recognition," *Journal on Multimodal User Interfaces*, vol. 1, no. 1, pp. 7-20, 2007.

[2] R. Aggarwal, J. K. Singh, V. K. Gupta, S. Rathore, M. Tiwari, and A. Khare, "Noise reduction of speech signal using wavelet transform with modified universal threshold," *International Journal of Computer Applications*, vol. 20, no. 5, pp. 14-19, 2011.

[3] R. Sarikaya, B. L. Pellom, and J. H. Hansen, "Wavelet packet transform features with application to speaker identification," in *Proceedings of 3rd IEEE Nordic Signal Processing Symposium*, Vigso, Denmark, 1998, pp. 81-84.

[4] J. N. Gowdy and Z. Tufekci, "Mel-scaled discrete wavelet coefficients for speech recognition," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Istanbul, Turkey, 2000, pp. 1351-1354.

[5] N. S. Nehe and R. S. Holambe, "DWT and LPC based feature extraction methods for isolated word recognition," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2012, no. 1, pp. 1-7, 2012.

[6] B. Bharathi, V. Deepalakshmi, and I. Nelson, "A neural network based speech recognition system for isolated Tamil words," in *Proceedings of International Conference on Neural Networks and Artificial Intelligence,* Brest, Belarus, 2006.

[7] S. Saraswathi and T. Geetha, "Morpheme based language model for Tamil speech recognition system," *International Arab Journal of Information Technology*, vol. 4, no. 3, pp. 214-219, 2007.

[8] R. Arun Thilak and R. Madharaci, "Speech recognizer for Tamil language," in *Proceedings of Tamil Internet Conference*, Singapore, 2004, pp. 1-7.

[9] V. Radha, C. Vimala, and M. Krishnaveni, "Isolated word recognition system for Tamil spoken language using back propagation neural network based on LPCC features," *Computer Science & Engineering*, vol. 1, no. 4, pp. 1-11, 2011.

[10] I. Patel and Y. S. Rao, "Speech recognition using HMM with MFCC: an analysis using frequency spectral decomposition technique," *Signal & Image Processing: An International Journal (SIPIJ)*, vol. 1, no. 2, pp. 101-110, 2010.

[11] M. Chandrasekar and M. Ponnavaikko, "Tamil speech recognition: a complete model," *Electronic Journal Technical Acoustics*, article no. 20, 2008. http://www.ejta.org/en/chandrasekar2.

[12] S. Rojathai and M. Venkatesulu, "A novel speech recognition system for Tamil word recognition based on MFCC and FFBNN," *European Journal of Scientific Research,* vol. 85, no. 4, pp. 578-590, 2012.

[13] A. N. Sigappi and S. Palanivel, "Spoken word recognition strategy for Tamil language," *International Journal of Computer Science Issues*, vol. 9, no. 1, pp. 1694-0814, 2012.

[14] P. Sivaraj and M. Rama, "Recognition of isolated spoken words using DWT," *International Journal of Engineering & Science Research*, vol. 2, no. 9, pp. 1187-1196, 2012.
[15] R. Thangarajan, A. M. Natarajan, and M. Selvam, "Word and triphone based approaches in continuous speech recognition for Tamil language," *WSEAS Transactions on Signal Processing*, vol. 4, no. 3, pp. 76-86, 2008.
[16] S. Saraswathi and T. V. Geetha, "Design of language models at various phases of Tamil speech recognition system," *International Journal of Engineering, Science and Technology*, vol. 2, no. 5, pp. 244-257, 2010.
[17] S. Karpagavalli, K. U. Rani, R. Deepika, and P. Kokila, "Isolated Tamil digits speech recognition using vector quantization," *International Journal of Engineering Research and Technology*, vol. 1, no. 4, pp. 1-12, 2012.
[18] P. Iswarya and V. Radha, "Speech based query processing architecture for Tamil–English in cross language text retrieval system," *International Journal of Emerging Trends in Engineering and Development*, vol. 7, no. 2, pp.437-442, 2012.
[19] D. L. Donoho and I. M. Johnstone, "Minimax estimation via wavelet shrinkage," *Annals of Statistics*, vol. 26, no. 3, pp. 879-921, 1998.
[20] P. Iswarya and V. Radha, "Comparative analysis of feature extraction techniques for Tamil speech recognition," in *Proceedings of International Conference on Emerging Research in Computing, Information, Communication and Application*, Yelahanka, India, 2013, pp. 755-761.
[21] A. Ekbal and S. Saha, "Simulated annealing based classifier ensemble techniques: Application to part of speech tagging," *Information Fusion*, vol. 14, no. 3, pp. 288-300, 2013.
[22] A. R. Ahmad, M. Khalid, and R. Yusof, "Machine learning using support vector machines," Centre for Artificial Intelligence and Robotics, Kuala Lumpur, Malaysia, 2002.
[23] A. Ben-Hur and J. Weston, "A user's guide to support vector machines," 2007; http://pyml.sourceforge.net/doc/howto.pdf.
[24] H. C. Kim, S. Pang, H. M. Je, D. Kim, and S. Y. Bang, "Constructing support vector machine ensemble," *Pattern Recognition*, vol. 36, no. 12, pp. 2757-2767, 2003.
[25] K. M. P. Sampath, P. W. D. C. Jayathilake, R. Ramanan, S. Fernando, and S. Chatura De Silva, "Speech recognition using neural networks," 2003; http://docslide.us/documents/speech-recognition-using-neural-network.html.

P. Iswarya

She received Master of computer science degree in Avinashilingam Institute for Home Science and Higher Education for Women in 2011. Currently she is a Ph.D. scholar and his area of interest includes speech recognition, natural language processing and information retrieval. She has 8 publications in International Journals and Conferences.

V. Radha

She is Professor in Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, India. She has more than 21 years of teaching experience and 7 years of Research Experience. Her area of specialization includes Image Processing, Optimization Techniques, Voice Recognition and Synthesis, Speech and signal processing and RDBMS. She has more than 45 publications at national and International level journals and conferences. She has one Major Research Project funded by UGC.