ISSN:0975-9646

Tamil Speech Recognition Using Hybrid Technique of EWTLBO and HMM

Dr.E.Chandra M.Sc., M.phil., PhD[1], S.Sujiya M.C.A., MSc(Psyc)[2]

[1.] *Director, Department of Computer Science, Dr.SNS Rajalakshmi College of Arts & Science, Coimbatore-32, India*

[2.] *Assistant Professor, Avinashilingam Institute for Home Science and Higher Education for Women , Coimbatore, India*

***Abstract*-Speech Recognition technology is one of the fast growing engineering technologies. The task of speech recognition is to convert speech into a sequence of words by a computer program. As the most natural communication modality for humans, the ultimate dream of speech recognition is to enable people to communicate more naturally and effectively. The accuracy of automatic speech recognition remains one of the most important research challenges after years of research and development. There are a number of well-known factors that determine the accuracy of a speech recognition system. Several speech recognition algorithms were adapted for different language. In this research continuous speech recognition is concentrated over Tamil language. This paper divides the proposed framework into three stages namely preprocessing, feature extraction and classification. Feature extraction is the process of retaining useful information of the signal while discarding redundant and unwanted information. In feature extraction stage, the proposed framework uses MFCC for transforming the signal into a form appropriate for classification. Classification is done by combination of EWTLBO and HMM. Teaching–Learning-Based Optimization (TLBO) is recently being used as a reliable, accurate and robust optimization technique scheme for global optimization over continuous spaces. This research presents an improved variant of TLBO algorithm, called Enhanced Weighted Teaching–Learning-Based Optimization (EWTLBO). A performance comparison of the proposed method is provided against the original TLBO and some other algorithms. An additional parameter "weight" is introduced to the existing TLBO algorithm to increase convergence rate. To enhance the search space of the algorithm an elitist concept is introduced to improve the performance of existing TLBO algorithm. EWTLBO is used to find the optimization value which is passed through the HMM to get the recognized output.**

Keywords: Preprocessing, Feature Extraction, Enhanced Weighted Teaching Learning based optimization (EWTLBO), Hidden Markov Model

I. INTRODUCTION:

Speech recognition and speech classification is one of the most important task in signal processing. It is also known as automatic speaker recognition[1]. The idea behind the speaker recognition is to convert the speech signals into sequence of words. Speech is one of the natural form of techniques for human communication and it is one of the recent methods in signal processing.

- Vocabulary size and confusability
- Speaker dependence vs. independence
- Isolated, discontinuous, or continuous speech
- Task and language constraints
- Read vs. spontaneous speech
- Adverse conditions.

These are some condition for efficiency system. Speech recognition has fallen into three categories they are

- Template-based approaches
- Knowledge-based approaches
- Statistical-based approaches.

Several approaches are handled for speech recognition, which is based on speech, speaker and vocabularies. Automatic speech recognition is one of the techniques, in which the concern person speaks through the microphone or telephone, these spoken words or written words are recognized or identified by the computer[2] [3].

Speech recognition is achieved for many languages around the world, researchers focus on developing the speech recognition through some techniques. Nowadays many research and researchers are focused to develop the speech recognition for Indian languages [4].

Tamil belongs to Dravidian language and it is mostly used in Tamilnadu, one of the state in India. Tamil is one of the official languages used in Tamilnadu and Sri Lanka and other country such as Singapore, Malaysia. Tamil is one of the popular languages used widely by 80 million people in the world. [5].There are lot of research works for recognizing the English speech recognition [6, 7 and 8]. This paper focuses on increasing the Tamil speech recognition rate. Section II reveals about the related work about speech recognition, Section III describes about the methodology of the proposed framework which is divided into three parts namely preprocessing, feature extraction and classification. Section IV reveals about the experimental results of the proposed framework and highlights the performance of the proposed work. Section V reveals about the conclusion of the proposed system.

II. RELATED WORK

Many researchers are focused and concentrated on sign language recognition and classification that are come from the countries like American, Australian, Korean and Japanese [9, 10, 11, 12 and 13]. Speech recognition system for children is developed and identified using a novel based fuzzy-based discriminative feature representation [14]. Artificial neural network is well known for pattern recognition and later 1980 ANN is proved to be suitable for speech recognition and speech applications [15].

In earlier speech recognition is recognized by using machines in the year of 1950, later 1952 a system was built for single speaker recognition which was isolated digit recognition [16]. In [17] Russian studies the speech recognition system came with a ideas through pattern recognition, and Japanese illustrates dynamic programming for speech recognition and linear predicting coding idea was given by itakura's researchers.

III. METHODOLOGY

The proposed system concentrates on continuous speech recognition on Tamil language. The proposed framework is divided into three steps namely preprocessing, feature extraction and classification. The overview of proposed methodology is shown in figure 1

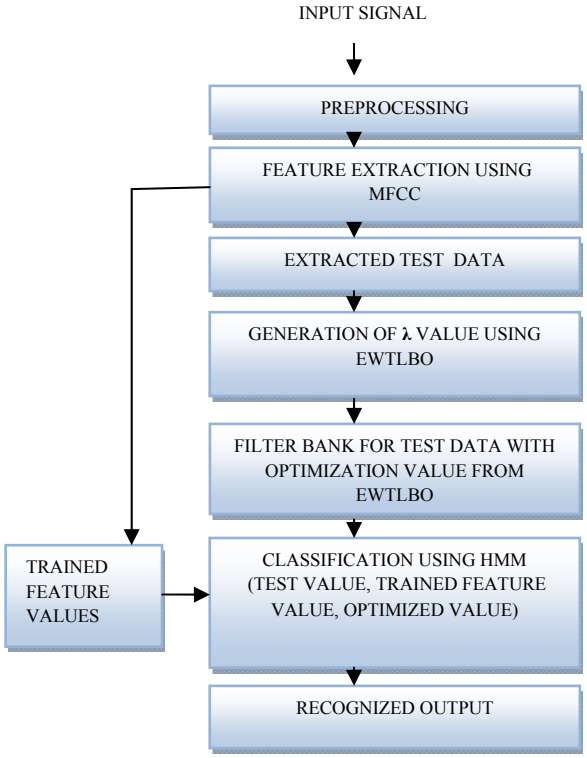


Figure .1 Proposed Methodology Overview

A. Preprocessing

In speech recognition, first phase is preprocessing which deals with a input speech signal which is an analog signal at the recording time, which varies with time. To process the signal by digital means, it is necessary to sample the continuous-time signal into a discrete-time discrete- valued (digital) signal. The purpose of preprocessing is to derive a set of parameters to represent the input speech signals in a form which is convenient for subsequent processing.

B. Speech Feature Extraction

Feature extraction is one of the most techniques used in signal processing, for getting the efficient system. The goal of this work is to sufficiently represent the characteristics of the speech signal with reduced redundancy. Features are extracted based on frames (windows). In the proposed system MFCC is used to extract the appropriate features for further classification

i. Mel Frequency Cepstral Coefficients (MFCC)

Mel Frequency Cepstral Coefficients (MFCC) are derived from the Fast Fourier Transform (FFT) of the audio clip. The basic difference between the FFT and the MFCC is that in the MFCC, the frequency bands are positioned logarithmically which approximates the human auditory system's response more closely than the linearly spaced frequency bands of FFT. This allows for better processing of data. The objective of the MFCC processor is to mimic the behavior of the human ears.

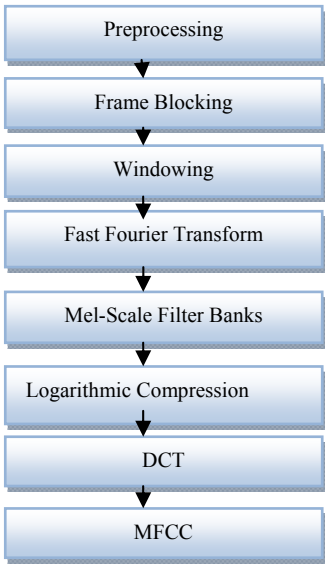


Figure 2.Block diagram of Mel-Frequency

The first step is preprocessing the speech signal, in which the amplitude which has zero is minimized or eliminated. Continuous speech signal is divided into frames of N samples in the frame blocking step. Adjacent frames are being separated by M (M<N). The values used are M = 128 and N = 256. The next step in the processing is to window each individual frame so as to minimize the signal discontinuities at the beginning and end of each frame. Preprocessing is used to minimize the spectral distortion by using the window to taper the signal to zero at the beginning and end of each frame. If we define the window as w (n), 0 ≤ n ≤ N −1, where N is the number of samples in each frame, then the result of windowing is the signal.

$$y(n) = x(n)w(n), 0 \leq n \leq N - \quad (1)$$

Typically the Hamming window is used, which has the form:

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N \quad (2)$$

Use of speech spectrum for modifying work domain on signals from time to frequency is made possible using Fourier coefficients. At such applications the rapid and practical way of estimating the spectrum is use of rapid Fourier changes.

$$x_k = \sum_{n=0}^{N-1} x_L e^{-j2\pi kn/N}, where\ k = 0,1,2,\ldots,N - \quad (3)$$

The relationship between frequency in Mel and frequency f

in Hz is specified as in

$$Mel(f) = 2925\ log_{10}\left(1 + \frac{f}{70}\right. \quad (4)$$

C. CLASSIFICATION

In this paper Tamil speech is classified and recognized by hybrid approach, a combination of Enhanced Weighted Teaching Learning Based Optimization (EWTLBO) and

Hidden Markov Model (HMM). As compared to the existing system the HMM method has good accuracy when compared with neural networks and other optimization techniques. In neural network to get the desired accuracy more number of parameters has to be set. This is not possible if the dataset set is changed and it involves different concepts. Sometimes the iteration fails to converge when the goal is attained. To overcome this, HMM was designed with the optimization concept. Where the coefficients in HMM are found by the EWTLBO process, EWTLBO is an optimization technique enhanced from the existing TLBO which is used to find out the significant parameters for HMM

ENHANCED WEIGHTED TEACHING–LEARNING-BASED OPTIMIZATION (EWTLBO)

Teaching-Learning-Based Optimization (TLBO) is recently being used as a steady, accurate and strong optimization technique scheme for global optimization over continuous spaces. This paper presents an enhanced version of TLBO algorithm, called the Enhanced Weighted Teaching-Learning-Based Optimization (WTLBO) .Two level of enhancement is done to the original TLBO algorithm First, a parameter 'weight' is introduced to the teacher's phase to find the best teacher based on student understanding. This parameter is introduced to increase the convergence rate of the speech. Secondly, an Elitist concept is introduced for improvisation of the TLBO algorithm. This concept is used to identify the exploration and exploitation of the TLBO algorithm.

This optimization method is based on the effect of the knowledge of a teacher based on the output of learners in a class. It is a population based method and unlike other population based methods it uses a population of solutions to proceed to the universal solution. A group of learners form the population in TLBO. In any optimization algorithms there are a number of different design variables. These different design variables in TLBO are analogous to different subjects offered to learners and the learner's result is analogous to the 'fitness', as in other population-based optimization techniques. As the teacher is considered the most learned person in the society, the best solution so far is analogous to Teacher in TLBO. The process of TLBO is divided into two parts.

- Teacher phase
- Learner phase

"Teacher phase" means learning from the teacher.

"Learner phase" means learning through the interaction between learners.

i. Initialization

Here some notations are given below and it is used to describe the TLBO.

N: number of learners in class i.e. "class size"

D: number of courses offered to the learners

MAXIT: maximum number of allowable iterations

The population X is randomly initialized by a search space bounded by matrix of N rows and D columns. The j^{th} parameter of the i^{th} learner is assigned values randomly using the equation

$$x_{(i,j)}^{0} = x_j^{min} + rand \times \left(x_j^{max} - x_j^{nd}\right) \quad (5)$$

where rand represents a uniformly distributed random variable within the range (0, 1), $x_j^{max} - x_j^{n}$ represent the minimum and maximum value for jth parameter. The parameters of ith learner for the generation g are given by

$$X_{(i)}^{g} = \left[x_{(i,1)}^{g}, x_{(i,2)}^{g}, x_{(i,3)}^{g}, \ldots, x_{(i,j)}^{g}, \ldots, x_{(i,\ldots}^{g}\right. \quad (6)$$

ii. Teacher phase

The mean parameter Mg of each subject of the learners in the class at generation g is given as

$$M^{g} = \left[m_1^{g}, m_2^{g}, \ldots, m_j^{g}, \ldots, m\right. \quad (7)$$

The learner with the minimum objective function value is considered as the teacher Xg Teacher for corresponding iteration. The Teacher phase continues the algorithm by shifting the mean of the learners towards its teacher. To obtain a new set of better learners a random weighted differential vector is formed from the current mean and the desired mean parameters and it is added to the existing population of learners.

$$Xnew_{(i)}^{g} = X_{(i)}^{g} + rand \times \left(X_{Teacher}^{g} - T_F M\right. \quad (8)$$

is the teaching factor that decides the value of mean to be changed. Value of can be either 1 or 2. The value of is decided randomly with equal probability as,

$$T_{FFFF} = round\left[1 + rand\,(0,1)\{2 - :\right. \quad (9)$$

Where is not a parameter of the TLBO algorithm. The value of is not given as an input to the algorithm and its value is randomly decided by the algorithm using Eq. (9). After conducting a number of experiments on many benchmark functions, it is concluded that the algorithm performs better if the value of is between 1 and 2. However, the algorithm starts to perform much better if the value of is either 1 or 2 and hence to simplify the algorithm, the teaching factor is recommended to take either 1 or 2 depending on the rounding up criteria.

If $Xnew$ is found to be a greater learner than X in generation g, then it replaces inferior learner X in the matrix.

iii. Learner phase

In this phase the interaction of learners with each other takes place. The process of mutual interaction tends to increase the knowledge of the learner. This random interaction among learners improves his/her knowledge. For a given learner X, another learner X is randomly selected (i ≠ r). The ith parameter of the matrix Xnew in the learner phase is given as

$$Xnew_{(i)}^{g} = \begin{cases} X_{(i)}^{g} + rand \times \left(X_{(i)}^{g} - X_{(r)}^{g}\right) \text{ if } f\left(X_{(i)}^{g}\right) < f\left(X_{(r)}^{g}\right) \\ X_{(i)}^{g} + rand \times \left(X_{(r)}^{g} - X_{(i)}^{g}\right) \text{ otherwise} \end{cases} \quad (10)$$

To modify the equation 8 and 10 through weight a modification is given to TLBO by setting the minimum and maximum value for the weight through above equation.

$$w = w_{max} - \left(\frac{w_{max} - w_{min}}{\text{max iteration}}\right) \quad (11)$$

In the above equation $w_{max} - w_{}$ are the minimum and maximum values for the weight factor 'w','i' is the current iteration number.

Hence the teacher phase can be improved as

$$Xnew^{r}_{(i)} = w * X^{g}_{(i)} + rand * \left(X^{g}_{Teacher} - T_F M\right) \quad (12)$$

and an improved learner can be written as

$$Xnew^{g}_{(i)} = \begin{cases} w * X^{g}_{(i)} + rand \times \left(X^{g}_{(i)} - X^{g}_{(r)}\right) \\ \quad if\ f\left(X^{g}_{(i)}\right) < f\left(X^{g}_{(r)}\right) \\ w * X^{g}_{(i)} + rand \times \left(X^{g}_{(r)} - X^{g}_{(i)}\right)\ otherwise \end{cases} \quad (13)$$

$$X_{new} = X_{old} + r\left(X_{teacher} - (T_r)Mean\right) \quad (14)$$

If the modification is best than the previous solution then randomly select the solution and modify them comparing by each other, modify the duplicate solution based on the elite solution and finally termination is fulfilled

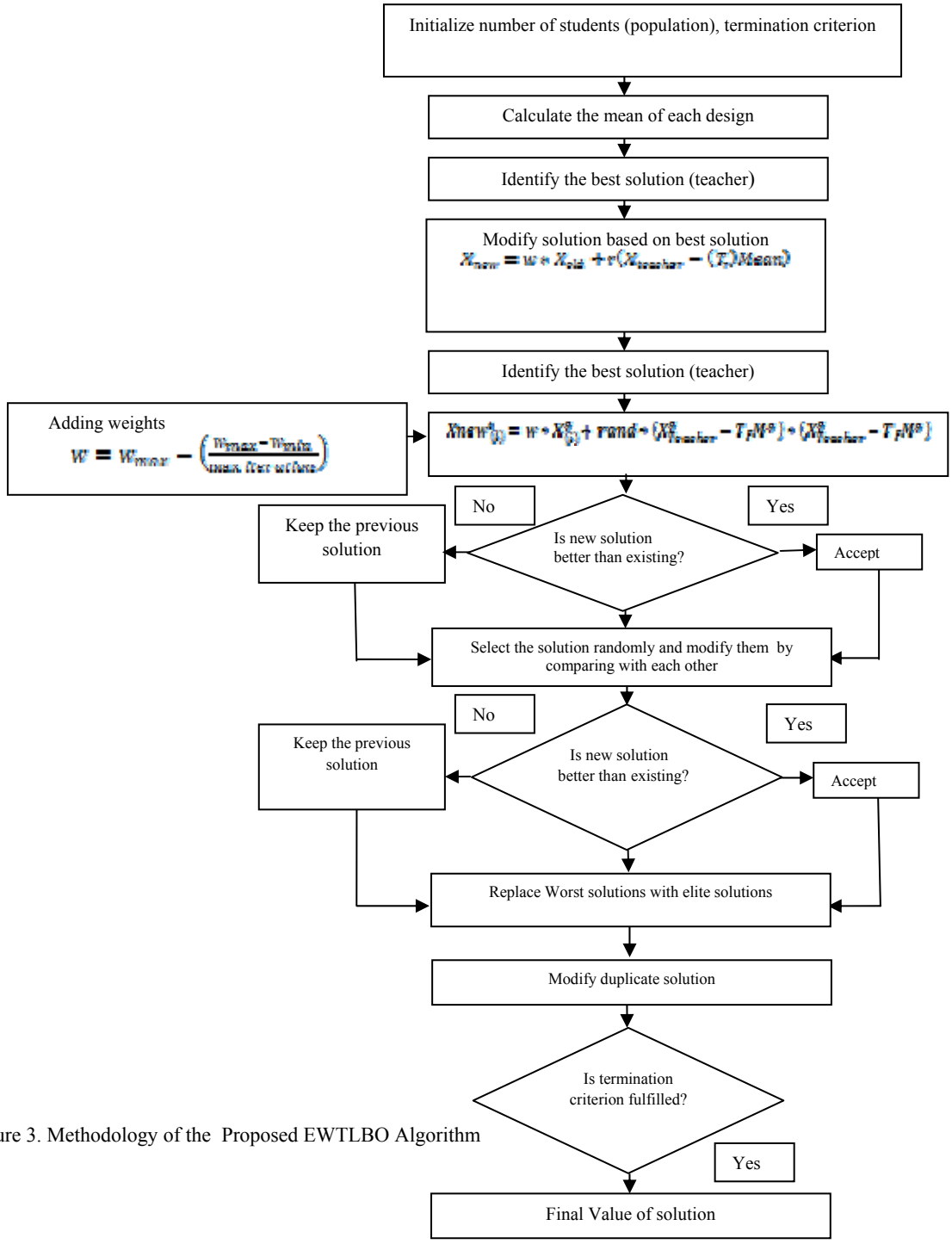


Figure 3. Methodology of the Proposed EWTLBO Algorithm

HIDDEN MARKOV MODEL

Hidden Markov Models (HMM) [1, 2] have a rich history in sequence data modeling (in speech recognition and bioinformatics applications) for the purpose of classification, segmentation, and clustering. HMM's success is based on the convenience of their simplifying assumptions. The space of probable sequences is constrained by assuming only pair-wise dependencies over hidden states. Pair-wise dependencies also allow for a class of efficient inference algorithms whose critical steps build on the Forward- Backward algorithm [1].

In this paper, three parameters are used in Hidden Markov model for classification of speech signal namely training, testing, features and optimization value generated by EWTLBO. Recognized Tamil speech output is generated by Viterbi algorithm.

i. Viterbi algorithm

The forward algorithm tells us how likely it is that a sequence of observations is generated by a model λ. However, in some cases the individual states of a Markov model may have some meaning, for example the phonemes in a word, and we might be interested in the sequence of states that is most likely to have generated the observation sequence $O = o_1 o_2 \ldots$. This boils down to maximizing $P(q|O,$ or equivalently maximizing $P(q|O,$ as:

$$\max_q P(q|O,\lambda) = \max_q \frac{P(q|O,\lambda)}{P(O)} = \max_q P(q|O,\lambda)$$

[15]

V. EXPERIMENTAL RESULTS

The proposed system is implemented using MATLAB 7.14 version. The experiments are carried out using Laptop with Intel CORE 2 DUO processor configuration in Microsoft Windows 2007 Environment. It is commonly used simulator tool. Each speaker recorded a news text through recorder having an inbuilt microphone. The recordings are in stereo recording and the extracted channels are also included in the specific files. It includes audio file, text file, NIST files which were saved as .ZIP Files. All the speech data are transcribed and labeled at the sentence level. All utterances were collected with headset microphone.

Figure 4. Feature Extraction output

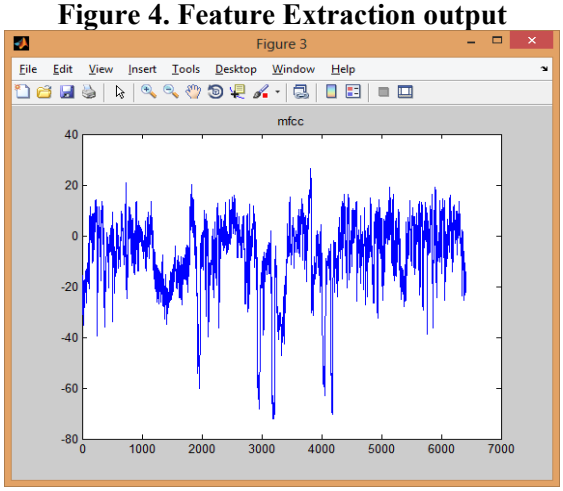
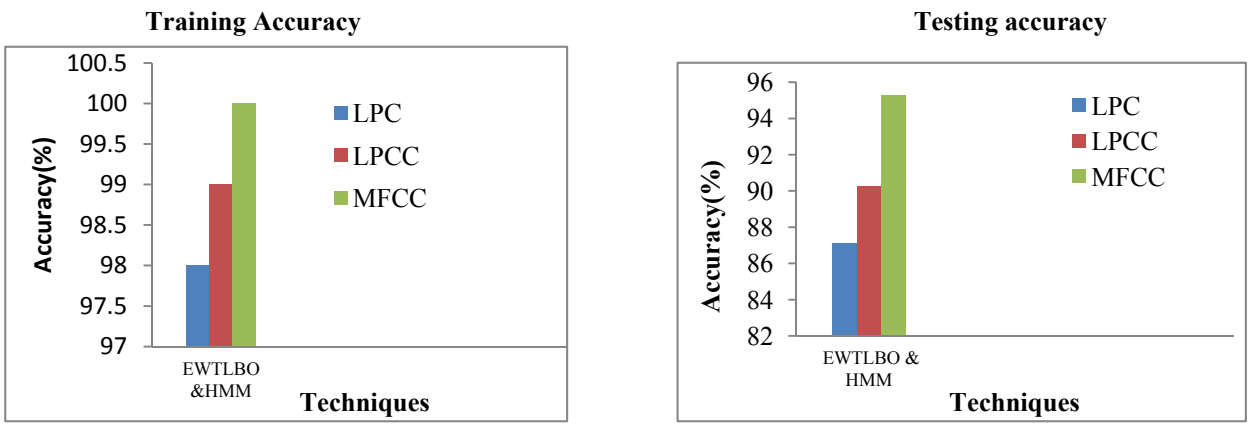


Figure 5. Comparison of Speech Accuracy



From the above table it is clearly represented that the training time of MFCC produces 100% accuracy and the testing time produces 95.26% accuracy .Hence the feature are extracted by MFCC are best for recognition

Performance Measure of the proposed system

Speech verification performance will be reported using the false acceptance rate (FAR), the false rejection rate (FRR).

$$FAR = \frac{\#\ accepted\ imposter\ claims}{\#\ imposter\ accesses} \times 100\%$$

I.

Table –1 : Comparison of FAR and FRR

Techniques	FAR (%)	FRR (%)
MFCC with EWTLBO & HMM	1.65	1.2

The above Table-1 provides the False Acceptance Rate and False Rejection Rate for proposed system . From the table it is clearly observed that the FAR and FRR is very low for the proposed system.

Table 2. Speech Recognition of the Proposed system

Techniques	Speech Recognition rate
Hybrid EWTLBO With HMM	**95.26%**

The overall prediction results of the proposed system is depicted and from the simulation results it is clearly observed that the proposed system gives the better accuracy rate of 95.26 % in recognizing Tamil Speech.

CONCLUSION & FUTURE WORK

Speech is a way of communication for human beings who can interact with each other. It is a difficult task to make a computer understand spoken commands. The proposed system is developed for recognizing Tamil spoken words. To employ this work, feature extraction is done after required preprocessing techniques. The most popular speech feature extraction and classification techniques were implemented and analyzed. Totally, three feature extraction algorithms namely MFCC, LPC, LPCC are compared with the proposed classification techniques. MFCC extracts the best feature for classification in the proposed method. The classification algorithms that are used for Tamil speech recognition are Enhanced Weighted Teaching Learning based Optimization (EWTLBO) with HMM. This method provides 96.26 % recognition rate, low false acceptance rate and false rejection rate.

FUTURE WORK

The current research work is extended for Continuous speech recognition of Tamil language. The potent advantage of Optimization Technique with HMM approach along with MFCC features is more suitable for these requirements and offers good recognition result. These techniques will enable us to create increasingly powerful systems, deployable on a worldwide basis in future.

REFERENCES

[1] Rabiner and Juang, 1986] L. Rabiner and B. Juang. An introduction to hidden Markov models. IEEE ASSp Magazine, 3(1 Part 1):4–16, 1986.

[2] S. Eddy. Profile hidden markov models.Bioinformatics, 14(9):755–763, 1998.

[3] Mr. R. Arun Thilak & Mrs. R. Madharaci, “Speech Recognizer for Tamil Language”, Tamil Internet 2004, Singapore.

[4] M. Chandrasekar, M. Ponnavaikko, “Spoken TAMIL Character Recognition”, in Electronic Journal Technical Acoustics (EJTA), ISSN 1819-2408, 2007.

[5] M. Chandrasekar, and M. Ponnavaikko, “Tamil speech recognition: a complete model”, Electronic Journal Technical Acoustics (EJTA) ,ISSN 1819-2408, 2008.

[6] Katagiri.S, Lee. C. H. A new hybrid algorithm for speech recognition based on HMM segmentation and learning vector quantization. IEEE Trans. on Speeech and Audio Processing, vol. 1, no. 4, 421–430, 1993.

[7] 63. C. Dugast, L. Devillers, X. Aubert. Combining TDNN and HMM in a hybrid system for improved continuous speech-recognition system. IEEE Trans. on Speeech and Audio Processing, vol. 3, no. 1, 1994.

[8] G. Zavaliagkos, Y. Zhao, R. Schwart, J. Makhoul. A hybrid segmental neural net/hidden Markov model system for continuous speech recognition. IEEE Trans. on Speech and Audio Processing, vol. 2, no. 1, 151–160, 1994.

[9] T. Starner, J. Weaver, and A. Pentland, ―Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video,‖ IEEE Trans. Pattern Analysis Machine Intelligence, Dec. 1998, vol.20, no. 12, pp. 1371-1375.

[10] M.W. Kadous, ―Machine recognition of Australian signs using power gloves: Toward large-lexicon recognition of sign language,‖ Proc. Workshop Integration Gesture Language Speech, 1996,pp. 165–174,.

[11] J. S. Kim,W. Jang, and Z. Bien, ―A dynamic gesture recognition system for the Korean sign language (KSL),” IEEE Trans. Syst., Man, Cybern. B, Apr. 1996,vol. 26, pp. 354–359,.

[12] H. Matsuo, S. Igi, S. Lu, Y. Nagashima, Y. Takata, and T. Teshima, ―The recognition algorithm with noncontact for Japanese sign language using morphological analysis,‖ Proc. Int. Gesture Workshop, 1997, pp. 273–284.

[13] Aleem Khalid Alvi, M. Yousuf Bin Azhar, Mehmood Usman, Suleman Mumtaz, Sameer Rafiq, Razi Ur Rehman, Israr Ahmed T , ―Pakistan Sign Language Recognition Using Statistical Template Matching,‖ World Academy of Science, Engineering and Technology, 2005.

[14] Seyed Mostafa Mirhassani# ˚ı ,#Hua-Nong Ting, “Fuzzy-based discriminative feature representation for children’s speech recognition”, Digital Signal Processing 31 (2014) 102–114.

[15] Sabato Marco Siniscalchi, Torbjørn Svendsenc and Chin-Hui Lee, “An artificial neural network approach to automatic speech processing”, Neurocomputing 140 (2014) 326–338.

[16] Rabiner, L. R., Levinson. S. E.,Rosenberg A>E and Wilpon J. G, “ Speaker Independent Recognition of Isolated Words using Clustering Techniques, IEEE. Trans. Acoustics, Speech, Signal Proc., ASSP-27:336-349,1979.

[17] Davis K.H., Bidulph R and Balashek S, “ Automatic Recognition of Spoken Digits”, J. Acoust. Soc. Am., 24 (6):637-642, 1952.

[18] Rabiner, L. and Juang, B., Fundamentals of speech recognition. Prentice Hall, Inc., Upper Saddle River, New Jersey, 1993.

[19] Rabiner, L. and Schafer, R., Digital Processing of Speech Signals. Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1978.