The CSI Journal on
Computer Science and Engineering
Vol. 19, No. 2, 2025
Pages 11-18
**doi.org/10.22034/jcse.2025.182114**
**Regular Paper**

# Syllable based ASR system for isolated Tamil words using fuzzy based neural network

Kiruthika Krishnamoorthy[1] Rathinavelu Arumugam [1]

[1] Department of Computer Science and Engineering,
Dr. Mahalingam College of Engineering and Technology, Pollachi.

## Abstract

Speech recognition, also known as Automatic Speech recognition or computer speech recognition means understanding the voice and performing any required task or the ability to match a voice against a provided or acquired vocabulary. Automatic speech recognition (ASR) is an important topic of speech processing. Tamil words are analyzed and recognized with respect to syllables and features are to be extracted. In this research, we present a new approach to automatically segment the isolated Tamil words into syllable-like segments. The algorithm for syllable segmentation works by processing the short-term energy function of the continuous speech signal. The proposed approach for speech recognition uses the combined features of Mel-Frequency Cepstral Coefficients (MFCC) and Linear Predictive Coding (LPC). Recognition process uses Fuzzy based Neural Network is combined together with Feed Forward Neural Network to recognize the corresponding word.
**Keywords:** short term energy; mel-frequency cepstral coefficients; linear predictive coding (LPC); neural network.

## 1. Introduction

Speech is the vocal form of communication and it is the most primary mode among human beings to share information. People can also communicate using other ways like writing, physical action, lip reading etc., but speech is the most efficient way and this can be combined with other modes to enhance the efficiency of the communication in various environmental conditions. Each spoken word is the phonetic combination of a limited set of vowels and consonant speech sound units called phonemes. People can also communicate through various modalities such as vision, audition, smell and tactile sensation. Hence speech has become effective communication method while others are used only when speech is not possible.

Based on this significance, attempts have been made to develop Human Computer Interface (HCI), using speech technology. This can be achieved by building up an Automatic Speech Recognition (ASR) system. Automatic Speech Recognition is a technology that captures words spoken by the user using microphone and converting the speech signal to text. The system displays the text corresponding to the recognized speech.

Reference [1] provides segmentation algorithm based on the short-term energy function of the continuous speech signal using Mel Frequency Cepstral Coefficient and Articulatory features using Feed Forward Neural Network based on only two levels Segmentation of a word. A two stage recognition system for phonemes is proposed in [2] using articulatory and spectral features. An approach based on Segmentation of speech is employed in [3, 4 and 5] Ref. [6] provides methodology to syllable a speech using fuzzy based method. An approach based on articulatory features is employed in [7] to segment the speech into syllables.

A feed forward neural network model proposed in [8] uses random initial weights to take error from training model. An Efficient Method for isolated words in Speech Recognition can be implemented by extracting Mel Frequency Cepstral Coefficients (MFCC) and Linear Predictive Coefficients (LPC) using Neural Network Algorithm [9]. Ref. [10] provides a Grapheme segmentation of Tamil speech

signals with MFCC and LPCC features. Syllable based continuous speech recognition for Tamil language is implemented in Ref.
[11] Ref. [12] provides Recognition method using MFCC and DTW. Classification of a syllable using articulatory-acoustic feature was expressed in Ref. [13].

# Related work

A major challenge for automatic speech recognition relates to significant performance reduction in noisy environments. Speech recognition mainly follows below approaches.

- **Speech Preprocessing**
- **Syllable Segmentation**
- **Feature Extraction Techniques**
- **Speech Recognition**

### *A. Speech Preprocessing*

Preprocessing of speech signals, i.e. separating the voiced region from the silence/unvoiced portion of the captured signal. In speech signal extraction of voiced part by making the silence and unvoiced region leads to computational complexity. Speech preprocessing includes pre-emphasis, framing, windowing etc. Ref. [13].

- *Pre-emphasis*

Pre-emphasis of digitized speech will increase the magnitude of higher frequencies with respect to lower frequencies in order to improve the overall signal-to noise. To achieve this, the speech signal is passed through a first order low pass filter to spectrally flatten the signal and to make it less susceptible to finite precision effects later in the signal processing.

- *Sampling*

The process of converting a continuous-time signal into a discrete-time signal is called Sampling. An appropriate sampling rate has to be chosen for all speech signals. The sampling rate chosen for our study is 44100Hz.

- *Framing*

Speech signal is not a stationary signal. However, it can be assumed to be stationary over short time intervals. So, to extract spectral information from these intervals, a speech signal is divided into small frames. It is made stable for 10 ms to extract the features and this is called frame rate. So a 20-25 ms window is applied between the samples at 10 ms intervals so as to achieve the stationary behavior of the frame.

- *Windowing*

The frames thus obtained are passed through a hamming filter to taper the signal at the edges of the frame and to prevent discontinuity within the frames and also at both the ends of each frame. Windowing is an important task in any signal processing applications. There are different types of windows like triangular, Blackman, hamming window etc. Among them hamming window best suited for speech signal processing.

### *B. Syllable Segmentation*

The sentences in a language are made up of a sequence of linguistic units which correspond to one or more sequences of acoustic units, namely, phoneme, syllable, word and sentences. Syllable structure was found to be an important factor in determining word errors, especially word deletions. Syllable seems to be an intuitive unit for representation as the variation observed is more systematic at the level of the syllable than at the level of the phoneme [1].

- Group Delay Based Segmentation
- **Short Term Energy Based Segmentation**
- Segmentation using Forced Viterbi Algorithm

### *C. Feature Extraction Techniques*

The spoken words can be recognized directly from the digitized waveform. It is a process extracting specific features of the preprocessed speech signal. Speech signals are non-stationary in nature and some form of statistical representations should be produced for reducing the speech signal variability. This can be achieved by performing feature extraction [9]. Various feature extraction techniques have been developed to extract spectral features from speech.

- **Mel Frequency Cepstral Coefficients (MFCC)**
- **Linear Predictive Coefficients (LPC)**
- Perceptual Linear Predictive (PLP) Coefficients
- Gammatone Frequency Cochleagram Coefficients (GFCC)

### *D. Speech Recognition*

Speech is most efficient way to exchange the information for human beings. Computer technology enables a device to recognize and understand spoken words by speaker. It is important that the machine can hear, understand, and act upon spoken information, and also speak to complete the information exchange.

Various speech recognition techniques have been developed and successfully used in many applications. They are divided into three broad categories,

- Template Matching
- Statistical Pattern Matching
- **Machine Learning**

#### *a. Machine Learning Approach*

Machine learning is an approach from artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning concerns on the development of computer programs that can access data and use it for themselves.
In Machine learning researchers have many different reasons for carrying out their work. Some are general principles of intelligent behavior, others are concerned with modeling human learning, and still others are oriented towards applications.
The main goal of machine learning is to program computers to use example data or past experience to solve a given problem. The approach combines the study of pattern recognition with the machines ability to analyze, learn and make a decision accordingly. Speech Classification process is for classifying the extracted features and relates the input sound to the best fitting sound from a database and represents them as an output. The commonly used techniques for Speech Classification are HMM (Hidden Markov Model), DTW (Dynamic Time Warping), VQ (Vector Quantization), ANN (Artificial Neural Network), etc. Several methods exist for this task such as,

- **Neural Networks**
- SVM(Support Vector Machine)
- Decision Trees and the combination of methods

### *I. Neural Network Recognition Method*

To perform particular function a mathematical model was developed called as neural network. It works like a human brain. Computer systems are trained by using machine learning algorithms to perform their task by their own. Neural Networks are a class of models within the general machine learning approach. Neural networks are a specific set of approach that has revolutionized the field of machine learning. It represents biological neural networks and the current algorithms so called deep neural networks have proven to work quite very well. Neural Networks are themselves general function approximations, that can be applied to literally almost any machine learning problem where the problem is about learning a complex mapping from the input to the output space. The diagrammatic representation of neural network is as shown below [13].

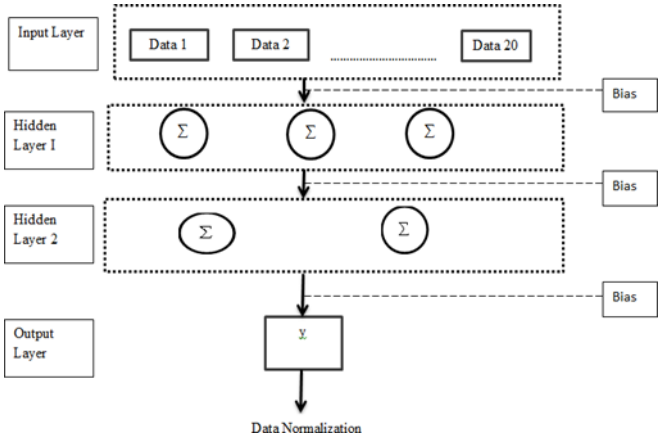


Fig. 1 structural representation of neuralnetwork

Neural networks are often used for statistical analysis and data modeling, in which their role is perceived as an alternative to standard nonlinear regression or cluster analysis techniques. A neural network is an artificial neural network where in connections between the nodes does not form a cycle.

The Neural network structure is compound 3 parts. The structure of the neural network for each word including input layer, two hidden layers and output layer. The number of nodes in each layer is 20, 3, 2 and 1 respectively. The input layer receives the feature vectors and forward to the next layer. All weights and bias in hidden layers and output layer are randomly initialized.

The feed forward neural network was the first and simplest type of artificial neural network devised. In this network, the information moves in only one direction. Neural network is one of the good choices among all. The term feed-forward means that the information is only allowed to "travel" in one direction.

It means that the output of one layer becomes the input of the next layer, and so forward. Feed-forward networks are advantageous as they have the fastest models to execute.

The multi-layer feed forward is used in back propagation neural network as shown in Figure 4 with total number of features as number of input neurons in input layer for LPC and MFCC parameters respectively. Neural Network model consists of input layer, hidden layer and output layer. Variable number of hidden layer input part can be tested for best results. We train network for different combinations of epochs with goal as minimum error rate.

To add the second step in our Neural-Fuzzy, using fuzzy logic because it allows us to avoid the retraining the neural networks. Fuzzy logic provides human capabilities to capture uncertainties that cannot be described by precise mathematical models. And fuzzy logic can able to the reasoning with some particular form of knowledge.

In many Speech recognition systems, many techniques are implemented and work in a cooperative relationship. Neural Networks perform very well at learning phoneme probability from highly parallel audio input, while neural Models can use the phoneme observation probabilities that Neural Networks provide to produce the likeliest phoneme sequence or word.

Neural networks offer exciting advantages such as adaptive machine learning, parallelism, fault tolerance, and generalization.

The comparisons between different techniques that have been used in multiple references are shown in Table 1.

Table 1. Comparisons between different techniques

| *S.no* | *Paper* | *Technique* | *Result* | *Issues* |
|---|---|---|---|---|
| 1 | "Syllable based Continuous Speech Recognition for Tamil Language", C. Sivaranjini and B. Bharathi. | For segmentation: Viterbi algorithm Features extraction : MFCC Recognition : Hidden Markov model | Input signal is split into syllables and performed using Hidden Markov model | Needs manually segmented data |
| 2 | "Speech Recognition using Neuro Fuzzy Network", Maitri Shah, Yash Sharma. | PSD and PCA features are extracted and signals are recognized using fuzzy neuro network | Quiet robust and it will improve higher accuracy in recognition | It needs elaborate feasibility for this approach |
| 3 | "Segmentati on Of Speech Into Syllable Units Using Fuzzy Smoothed Short Term Energy Contour", Ghazaal Sheikhi, Farshad Almasganj. | Feature Extraction: Combine MFCC and LPC Recognition: Neural Network | During syllable segmentatio n duration will be reduced | For each sample, previous 7 samples are to be considered for smoothing |

| | | | | |
|---|---|---|---|---|
| 4 | "Isolated word Recognition System for Tamil Spoken Language using Back Propagation Neural Network Based on LPC Features", Dr.V.Radha, Vimala.C, M.Krishnave ni. | Back propagation Neural Network technique based on LPC features | It uses multilayer feed forward neural network for speech recognition | Needs adapted network with specified parameter |
| 5 | "Syllable based Speech Recognition System for Tamil Language using Acoustic and Articulatory" " Gayathri S, Rathinavelu A, Jayashree C. | Combination of MFCC and articulatory features with FFNN for recognition | Disyllable segmentatio n can be performed | Not more than two level segmentatio n of a word can be performed |

# 3. Existing framework

In the existing ASR system for Tamil language, syllable based Automatic Speech recognition system for isolated words perform two levels syllable segmentation. Feature extractions are extracted using the combination of MFCC with articulatory features and Feed Forward Neural Network technique is used for recognition. The existing system for isolated Tamil words containing the following process.

### *A. Syllabe Segmentation- Short Term Energy*

Several algorithms have been proposed for speech signal endpoint detection. The most popular algorithm is the one based on energy contour because of simple computing and which helps to indicate the present of speech signal. Each of energy algorithms provides different energy contours. The most important of syllable segmentation is the parameter value and the threshold setting of the detection technique.

### *B. Articulatory Feature Extraction*

Articulation is the movement of the tongue, lips, jaw, and other speech organs in ways that make speech sounds. Spectral features mainly represent the gross shape of the vocal tract, but not the information related to the excitation source or the positioning and movements of various articulators. But, the production of each sound unit is characterized by articulatory and excitation features in addition to vocal tract features. A unique combination of articulators in the vocal tract and specific source of excitation results in the production of a particular sound unit. Articulatory phonetics focused on transformation of aerodynamic energy into acoustic energy [1]. The Articulatory Features are derived from the spectral features using FFNNs**.**

### *C. Recognition*

Template matching approach like DTW can be used for recognition which measures the similarity between two sequences which may vary in time or speed. The sequences are warped non-linearly in the time dimension to measure their similarity. The feature vectors for all the words are extracted and a reference template for each word in the corpus is prepared. The DTW distance between the feature vectors are calculated and the minimum distance is found out. The production of sound unit is characterized by articulatory features in addition to vocal tract features [12].

Based on the above context, it performs only disyllable segmentation for isolated Tamil words and also in recognition FFNN calculates only randomized values in the hidden layer.

# 4. Proposed framework

The main objective of the system is to develop a syllable based ASR system for isolated Tamil words using Fuzzy based Neural Network which is quiet robust to noise and it can yield higher recognition results compared to other popular algorithms. Short Term based syllabification technique for isolated Tamil words can segment a word with any number of syllables. The main objective of the ASR system is to improve performance. So MFCC and LPC are combined together to extract the features. This can be achieved by identifying the limits in the existing system and rectifying it. Existing ASR system performs disyllable segmentation for isolated Tamil words and in recognition hidden part, only randomized values are calculated. This technique has many disadvantages and a solution for these problems have been discussed in this section.

The Proposed ASR system comprises modules for Speech Corpus Collection, Pre-processing the input signal, Syllable Segmentation based on short term energy calculation, Features are extracted using Mel Frequency Cepstral Coefficients and Linear Predictive Coefficients, Fuzzy and Feed forward neural network are used to recognize the input speech.

Figure 2 illustrates the overall architectures of proposed ASR system.

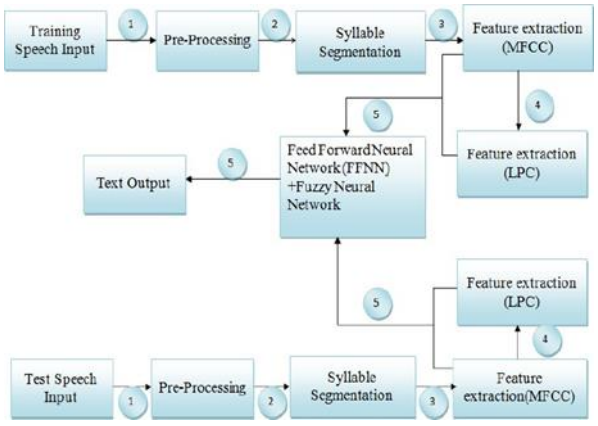


Fig.2 syllable based asr sytem

Syllables have better representational and durational stability relative to the phonemes. For syllable segmentation short term energy of an input signal is calculated to detect the boundaries, and speech signal is smoothened. Most of the recognition systems are developed using some features using MFCC or LPC.

Mel-Frequency Cepstral Coefficients (MFCC) is a representation of the real Cepstral of a windowed short- time

signal [16]. The LPC renders a robust, reliable and accurate method for estimating the parameters that represent the vocal tract system [9]. Fuzzy neural network algorithm can be used to adapt for human sound because this algorithm has high accuracy for classifying when compare with other popular algorithm in classification.

The Neural network structure is composed of 3 parts. The structure of the neural network includes input layer, hidden layers and output layer [14]. The input layer receives the feature vectors and forward to the next layer. All weights and bias in hidden layers and output layer are initialized based on membership function. In each round, data passes all layers. This process is called feed forward. The second process is the data normalization.

From these, speech corpus developed in .wav format by 1 female speaker using the tool Wavosaur.1.3.0.0. The corpus includes 25 words.

# 5. Methodolgy

Among the few attempts on Tamil ASR, most of the experiments are carried out using some of the feature extractions. And most of the works are based on machine learning approaches. As an alternative, this work focuses on developing a syllable based ASR system using fuzzy neuro method, to achieve good recognition rate.

### *A. Dataset Preparation*

For the purpose of speech recognition, a speech corpus has to be prepared. The training set including different categories with different combination of words has been developed. Tamil words include 12 vowels, 18 consonants. Figure 3 shows the waveform of the word **"*வாழ் த்துக்கள்*".**

Fig.3 Waveform of the word"*வாழ்த்துக்கள்* .wav"

*a. Speech Acquisition*

Initial phase of any speech recognizer requires speech data. Speech corpus has to be prepared for both training and testing the system which includes all set of words. During speech acquisition the acoustic waves emitted by the vocal tract system are captured by a microphone. In human speech production system, the articulation produces sound waves and ear convey sit to the brain for processing.

When humans speak, air passes from the lungs through the oral cavity and nasal cavity and produces sound. These sounds forms vowels and consonants usually called phonemes. A phoneme is a unit or individual of sounds and which it contains different group of sounds. The phonemes are combined together to form words and words are connected together to form sentences. Before acquiring the speech, there are some specifications to be followed and is given in Table 2.

| SPECIFICATIONS | DESCRIPTIONS |
|---|---|
| Input file format | wav |
| Sampling rate | 44100 Hz |
| Sampling format | 15 bits |
| Input Channel | Mono |
| Recording Software | Wavosaur |

### *B. Speech Preprocessing*

The signal obtained from the wavosaur is first preprocessed in order to make it more compatible, noise- free and suitable for feature extraction. Speech preprocessing steps include,

- End Point Detection (EPD)
- Windowing

*a. End Point Detection (EPD)*

While recording sound waves, there is always a possibility that the spoken word is preceded and succeeded by silence. However we are in demarcating the words spoken, there is always a limit to visual editing and so there exists a consequent scope for improvement. Speech consists of voiced and unvoiced parts where the major portion is voiced. Voiced speech is periodic in nature, can be identified and extracted and is the primary ingredient of pre-processing, whereas unvoiced speech is non-periodic and random. So, separating the voiced and unvoiced speech has become a subject of interest and is one of the key pre-processing steps in the speech recognition process.

*b. Windowing*

Each sample is multiplied by an N sample window w (n) where the window chosen is the hamming window and also frames are obtained through hamming filter. It reduces the discontinuities of the speech signal at the edges of each frame which in turn minimizes the adverse effects of chopping N samples and also it is used for minimizing the spectral distortion [15]. The plot illustrates before preprocessing and after preprocessing of a word**"*வாழ் த்துக்கள்* .wav"**.

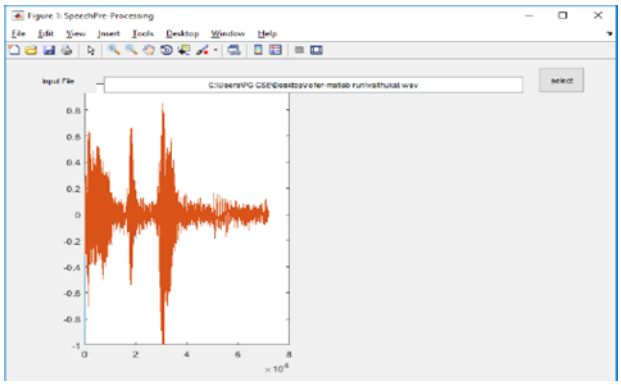

Fig.4 before preprocessing

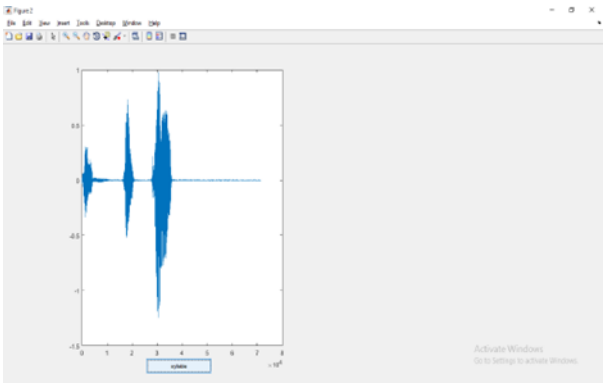

Fig.5 after preprocessing

### *C. Syllable Segmentation*

The preprocessed input signal is segmented into syllables which will be used for speech recognition process. For the input signal, short term energy is calculated. Absolute energy is selected for the process [1]. The calculated short term energy is smoothened using sgolay filter since the original STE signal contains many local maxima and local minima which will alter the boundary detection.

From the smoothened STE signal, boundary detection algorithm is executed to find the syllable boundaries. Approximate syllable duration will be 300ms and this value can be used to improve the segmentation accuracy. The valleys in the STE signal represent the syllable boundaries and the peaks represent syllable centers, since the syllable center will be vowel and the vowels have high energy corresponding to the peaks in the STE signal. Figure 6 shows the short term energy of the word **"*வாழ் த்துக்கள்*.wav".** Figure 8, 9, 10 shows the segmented syllables "***வாழ் த்***", "***துக்***" **and** "***கள்***".

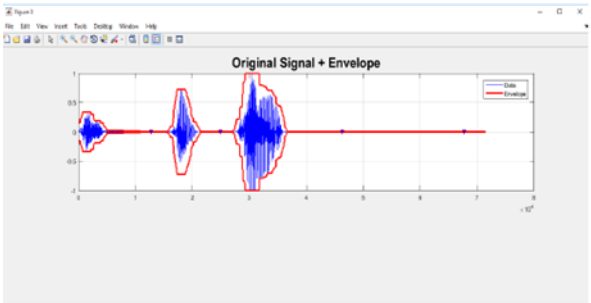


Fig.6 PLOT OF THE WORD"*வாழ்த்துக்கள்* .WAV"

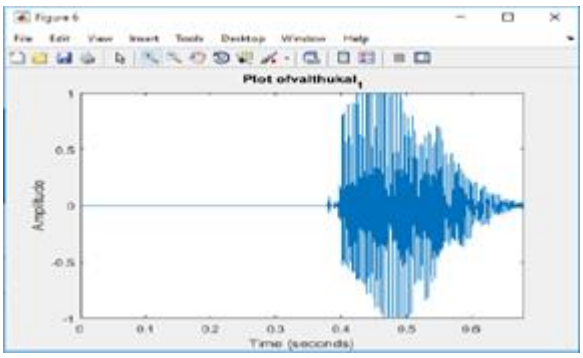


Fig.7 plot of the syllable "*வாழ்த்*.wav"

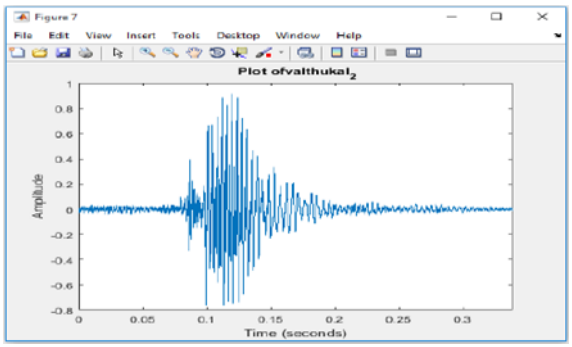


Fig.8 plot of the syllable"*துக்*.wav"

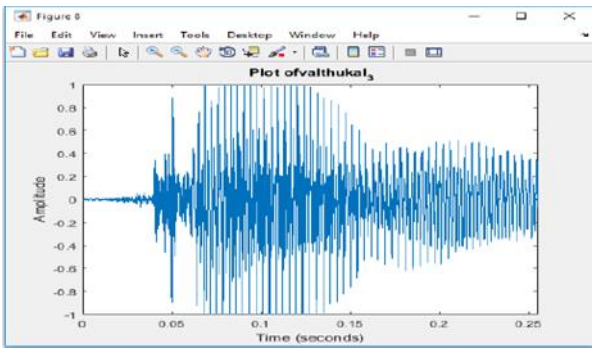


Fig.9 plot of the syllable "*கள்* .wav"

Peaks in the smoothened short term energy curve indicate syllable center and valley indicate the syllable boundaries. Based on the peaks in the input signal and high and low frequency of the signal it can segment a word and syllable it.

### *D.Feature Extraction*

Feature extraction is a basic and fundamental preprocessing step for recognition and machine learning problem. It's a special form of dimensionality reduction technique which is used to reduce the data which is very large to be processed by an algorithm. In feature extraction, input data is transformed into a set of features which provides the relevant information for performing a desired task without the need of the full size data but using the reduced set. The speech recognition techniques have a background of DSP i.e. Digital signal processing. The techniques used in this research work are the combination of Mel Frequency Cepstral Coefficients and Linear Predictive Coding [16].

#### *a. MFCC Features*

The most commonly used feature extraction method in automatic speech recognition (ASR) is Mel-Frequency

Cepstral Coefficients (MFCC) and it is one of the most powerful speech feature extraction technique and works on human auditory perception system[1]. The MFCC features are based on the known variation of the human ear"s critical bandwidth frequencies with filters spaced linearly at low frequencies and logarithmically at high frequencies used to capture the important characteristics of speech. The signal is divided into overlapping frames to compute MFCC coefficients. Let each frame consist of N samples and let adjacent frames be separated by M samples where M<N. Each frame is multiplied by a Hamming window [16]. Then the signal is converted from time domain to frequency domain by subjecting it to Fourier Transform.

In the next step the frequency domain signal is converted to Mel frequency scale, which is more appropriate for humans. This is done by a set of triangular filters that are used to compute a weighted sum of spectral components so that the output of the process approximates a Mel scale. Then the log Mel scale spectrum is converted to time domain using Discrete Cosine Transform (DCT) [1]. MFCCs are commonly derived as follows,

- Take the Fourier transform of (a windowed excerpt of) an input signal.
- Map the power values of the spectrum obtained above onto the Mel scale, using triangular overlapping windows.
- Take the log values of the powers at each of Mel frequencies.
- After take the discrete cosine transform for the list of Mel log powers, as if it were a signal.
- The MFCCs are the amplitudes of the resulting spectrum.

The result of the conversion in the input is called Mel Frequency Cestrum Coefficient. The set of coefficients is called acoustic vectors. Therefore, each input signal is transformed into a sequence of acoustic vectors. A block diagram of the MFCC processes is shown in the Figure 10.

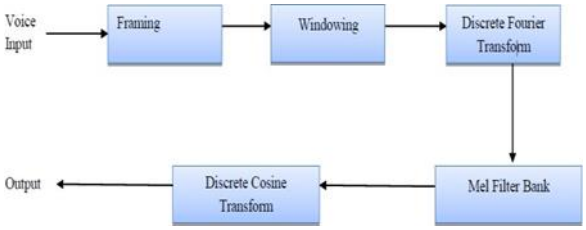


Fig.10 Mfcc Feature Extraction

## b. *LPC Features*

It is desirable to compress the input signal for efficient transmission and storage. The Digital signal is compressed before transmission for efficient utilization of channels into wireless media. LPC is most widely used for medium or low bit rate coder. Then LPC calculates a power spectrum for each input signal. It is used for formant analysis [16]. LPC is one of the most powerful speech analysis techniques and also used for formant estimation technique.

While passing the input speech signal from speech analysis filter to remove the redundancy in signal, and it generated a residual error to the output. It can be quantized by smaller number of bits are compared to original signal. So here, instead of transferring entire signal we can transfer this residual error and speech parameters to generate the original signal.

A parametric model which is computed based on least mean squared error theory, and this technique called as linear prediction (LP). By this method, the speech signal is approximated as a linear combination of its previous samples. In this technique, the obtained LPC coefficients describe the formants. The frequencies which occur in resonant peaks are called as formant frequencies.
In this method, the locations of the formants in the input speech signal are estimated by computing the linear predictive coefficients over a sliding window and finding the peaks in the spectrum of the resulting LP filter. Then excluded 0th coefficient and used next ten LPC Coefficients.
In speech synthesis, during vowel sound vocal cords vibrate harmonically and so quasi periodic signals are produced. Vocal tract are responsible for speech response at the time of filter. Biological phenomenon of speech generation can be easily converted in to equivalent mechanical model. Periodic impulse train and random noise can be considered as excitation source and digital filter as vocal track [16].

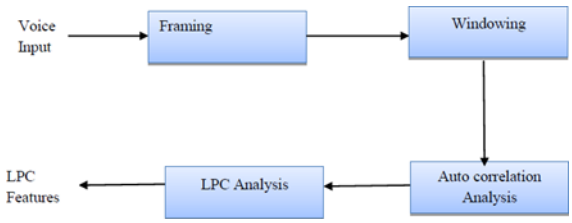


Fig.11 Lpc Feature Extraction Technique

## E. *Recognition*

The features are extracted so far the speech signal, MFCC and LPC features are combined together to form any number of syllable recognition system. In the first stage features are predicted for the segmented syllables and in the second stage syllables are recognized based on the previously predicted MFCC features using Feed Forward Neural Network and Fuzzy Neuro Network. Number of nodes in the input layer is the combination of input and output nodes for the first stage and their features which correspond to the syllables are used for training the FFNN.

Neuro fuzzy system is trained by the means of a data driven learning method derived from neural network theory. It takes into account local information to cause local changes in fundamental fuzzy system. It is the system that is initialized with or without prior knowledge about the data. It approximates an n-dimensional unknown function which is partly represented by training examples. Fuzzy rules are interpreted as vague prototypes of the training data.

Neuro fuzzy is represented as three layered feed forward neural network where first layer comprises of input variables, the second symbolizes fuzzy rules and the third

layer gives us the output layer. Here the fuzzy sets are converted to connection weights.

This work uses the Neuro-Fuzzy algorithm from to adapt for human sound because this algorithm has high accuracy for classifying infant sound when compare with other popular algorithm in classification. The training processes, including neural network, data normalization and fuzzy logic [14].

# 6. Performance Evaluation

### *A. Syllable Segmentation Accuracy*

Syllable Segmentation accuracy based on the actual duration of the syllable is calculated using below equation „1".

$$\text{relativeError} = \frac{|\text{actualDuration} - \text{estimatedDuration}|}{\text{actualDuration}} \quad (1)$$

It was noted that for the selected twenty five different categories of words. And the words have relative error that shows the actual syllable duration in the word and the duration estimated using proposed algorithm.

### *B. Recognition Accuracy*

The performance of a speech recognition system is measured using Recognition rate as given in „2".

$$RecognitionAccuracy = \frac{\text{No. of times words recognized}}{\text{Total no. of trials}} \times 100 \quad (2)$$

Table 3 shows the recognition accuracy of fuzzy neuro recognition process compared with the two-stage recognition process.

Table 3. Recognition Accuracy Rate

| SPEECH CORPUS | BASELINE ACCURACY | FUZZY NEURO ACCURACY |
|---|---|---|
| 51 Tamil Syllables | 76.9% | 86.66% |

# 7. Summary and conclusion

In this work, any number of syllable segmentation with recognition system is developed using feed forward neural network with fuzzy neuro network and it can yield higher recognition result compared to other popular algorithms. For the coarticulated spoken words, syllables are identified and for each syllable MFCC with LPC are determined. These two features are combined together to provide a better recognition system with improved accuracy.

# 8. Acknowledgement

This work is performed at Dr.Mahalingam College of Engineering and Technology as a part of project work titled "Syllable based ASR System for Isolated Tamil Words using Fuzzy Based Neural Network" supported by Department of Computer Science and Engineering.

# References

[1] Gayathri S, Rathinavelu A and Jayashree C, (2018), "Syllable based Speech Recognition System for Tamil Language using Acoustic and Articulatory", International Conference on Frontiers in Engineering, Applied Sciences and Technology, NIT, Trichy.

[2] Manjunath K E, K. Sreenivasa Rao and Gurunath Reddy M (2015),"Two-Stage Phone Recognition System using Articulatory and Spectral Features", Spaces-2015.

[3] V. Kamakshi Prasad, T. Nagarajan, Hema A. Murthy, (2004), "Automatic Segmentation Of Continuous Speech Using Minimum Phase Group Delay Functions", Speech Communication, pp.429–446.

[4] Nutthacha Jittiwarangkul, Somchai Jitapunkul, Sudaporn Luksaneeyanawin, Visarut Ahkuputra, Chai Wutiwiwatchai (1998), "Thai Syllable Segmentation For Connected Speech Based On Energy", IEEE Asia-Pacific Conference on Circuits and Systems.

[5] Ghazaal Sheikh, Farshad Almasganj, (2011), "Segmentation Of Speech Into Syllable Units Using Fuzzy Smoothed Short Term Energy Contour", Biomedical Engineering (ICBME).

[6] Suyanto & Agfianto Eko Putra, (2014), "Automatic Segmentation of Indonesian Speech into Syllables using Fuzzy Smoothed Energy Contour with Local Normalization, Splitting, and Assimilation", Journal of ICT and Research Applications.

[7] K. Sreenivasa Rao, Manjunath K E, (2017), "Speech Recognition Using Articulatory and Excitation Source Features", Springer Briefs in Electrical and Computer Engineering, pp.17 to 44.

[8] Salam Hamdan, Adnan Shaout, (2016), "Hybrid Arabic Speech Recognition System Using FFT, Fuzzy Logic and Neural Network", IRACST - International Journal of Computer Science and Information Technology & Security, vol. 6, issue. 4.

[9] Mayur R Gamit, Kinnal Dhameli (2015), "Isolated words using MFCC, LPC and Neural Network", International Journal of Research in Engineering and Technology, vol. 4, issue .6.

[10] Geetha K, Dr. R.Vadivel (2017), "Grapheme Segmentation of Tamil Speech Signals using Excitation Information with MFCC and LPCC Features", International Journal of Computer Science and Information Technologies, vol. 8.

[11] C.Sivaranjani, B.Bharathi (2016), "Syllable Based Continuous Speech Recognition for Tamil Language", International of Advanced Engineering Technology vol. 8, issue. 1, pp.01-04.

[12] Dalmiya C, Dr. Dharun V, Rajesh K, (2013), "An Efficient Method For Tamil Speech Recognition Using MFCC And DTW", IEEE Conference on Information and Communication Technologies (ICT), pp. 1263-1268.

[13] Nitin Washani, Sandeep Sharma, (2015), "Speech Recognition System: A Review", International Journal of Computer Applications, vol. 115, issue. 18.

[14] Maitri Shah, Yash Sharma (2017), "Speech Recognition using Neuro Fuzzy Network", IJARIIE- ISSN", vol. 3, issue. 2.

[15] Dr.V.Radha, Vimala.C, M.Krishnaveni (2017), "Isolated word Recognition System for Tamil Spoken Language using Back Propagation Neural Network Based on LPC Features", An International Journal (CSEIJ),vol. 1,isuse. 4.
[16] Namrata Dave(2013), "Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition", International journal for advance research in engineering and technology, vol. 1,issue. 6.
[17] Lakshmi A, Hema A Murthy (2006), "A syllable based Continuous Speech Recognizer for Tamil", interspeech.
[18] K. Geetha and E. Chandra (2015), "Monosyllable Isolated Word Recognition for Tamil language using Continuous Density Hidden Markov Model", Ieee international conference on electrical, computer and communication technologies (icecct).
[19] Moirangthem Tiken Singh, Abdur Razzaq Fayjie, Biswajeet Kachari (2015), "A Survey Report on Speech Recognition System", International Journal of Computer Applications, vol. 121, issue. 11.
[20] Prerana Das, Kakali Acharjee, Pranab Das and Vijay Prasad (2016), "Voice recognition system: speech-to-text" Journal of Applied and Fundamental Sciences.
[21] Preeti Saini, Parneet Kaur (2013), "Automatic Speech Recognition: A Review", International Journal of Engineering Trends and Technology, vol. 4, issue. 2.
[22] Prachi Khilari, Bhope V. P, (2015), "A review on speech to text conversion methods" International Journal of Advanced Research in Computer Engineering & Technology, vol. 4, issue. 7.