ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2020.0306

AUTOMATIC SPEECH RECOGNITION SYSTEM USING MFCC-BASED LPC APPROACH WITH BACK PROPAGATED ARTIFICIAL NEURAL NETWORKS

K. Pavan Raju[1], A. Sri Krishna[2] and M. Murali[3]

[1]*Department of Computer Science Engineering, Centurion University of Technology and Management, India*
[2]*Department of Information Technology, Shri Vishnu Engineering College for Women, India*
[3]*Department of Electronics and Communication Engineering, Centurion University of Technology and Management, India*

Abstract

Over the previous years, a marvelous quantity of study was performed by utilizing the artificial intelligence based deep learning approaches for the speech recognition applications. The automatic speech recognition (ASR) facing the problems in as preprocessing, feature extraction and classification stages mostly, thus solving these problems is mandatory to improve the classification accuracy of speech processing. To solve these issues, an advanced speech recognition methodology has developed by utilizing the Spectral Subtraction (SS) method of denoising with the combination of Mel-frequency Cepstral coefficients (MFCCs) and linear predictive coefficients (LPCs) feature extraction of speech signals. Then back propagated artificial neural networks (BP-ANN) is utilized for classifying the speech signals for the purpose of ASR, respectively. The simulation results show that the proposed approach gives the better classification accuracy compared to the state-of-ASR approaches..

Keywords:

Speech Processing, Automatic Speech Recognition, Mel-Frequency Cepstral Coefficients, Linear Predictive Coding, Artificial Neural Networks

1. INTRODUCTION

Speech is very crucial means of communication between humans. ASR means making the system smarts using computational algorithms which can recognizes the spoken word and converts it into text format or generates some control action [1]. ASR is a technique by which a device understands the human voice given as input and performs the assigned task. The primary task of such a system is the design of a model capable of recognizing the speech. But computers do not have the intelligence to interpret speech like human beings. Human beings can recognize the sound of interest from a large set of sounds which are concurrently audible. On the other hand, a computer-based system will consider the other sounds which are not of interest as noise. Hence to build a robust computer system, improvements are very much necessary. The units of speech that are popularly used for the recognition process are the phonemes and syllables. ASR systems are normally designed for recognizing three types of speech sound namely:

- *Recognition of Isolated Word*: This type of recognition is an easy process where the speaker needs to pause automatically between the words.
- *Recognition of Connected Word*: This type of recognizer can analyze limited amount of spoken words.
- *Recognition of Connected Speech or Continuous Speech*: This type of speech recognition allows the system to recognize normal form of conversational speech. Here, the system requires to be trained. Such systems are called speaker dependent system.

Systems which do require training is said to be a speaker independent system. The prime objective of speech synthesis/recognition system is the synthesis and recognition of continuous and natural speech as is used in a conversation. It is not easy to synthesize/recognize the speech in a natural conversation because there is negligible pause or sometimes no pauses between the words or phrase. Therefore, for conversional speech, which has a natural flow, the recognizer is required to apply the concept of "guessing". In statistical analysis this is done to generate appropriate speech units capable of producing correct sentences.

Speech pre-processing consists of speech encoding, segmentation, and noise removal in speech. Feature extraction step uses computational algorithm to find the discriminating features of the speech sample. In classification stage different supervised or unsupervised algorithm for the classification of speech. Depending upon type of ASR systems is classified into isolated word, connected word, continuous speech, and spontaneous SR system. Based on speaker mode ASR systems are classified into speaker dependent, speaker independent and speaker adaptive system [2]. In speech technology, features are the characteristics present in each sample that are measurable. The most important requirement in building a pattern recognition application is the extraction of a distinct feature set. A recognition system also requires the extraction of a discriminating feature set to identify the different speeches. As proposed work on Speech Recognition is exploiting increased complex feature space, some specific features used for different works may vary significantly. Based on vocabulary used, ASR systems categorized in to low, medium, and high dataset speech recognition systems. ASR systems have wide range of applications such as biometric speech recognition systems, speech to text conversion, audio conferencing, emotion recognition, robotics, synthetic speech recognition, education sector etc. Most of the ASR systems are trained using clean data, such a system gives unsatisfactory performance at real time speech recognition because of noise addition in the testing signal. Therefore, robust speech recognition under noisy condition is necessary. To deal with such a condition normally noise reduction or noise isolation techniques have been used. But when the noise cannot be isolated then noise reduction may degrade the signal content. Also pre-processing time which is used for the noise reduction is larger.

Most of the speech recognition systems have speech pre-processing for the speech enhancement as major step. Spectral subtraction is well-liked method for speech enhancement. In this method signal is enhanced by subtracting the average noise spectrum from the average signal spectrum. This method is not

applicable when noise profiling is not possible. Many filters [3] has been proposed with non-linear function have been used for speech enhancement. Active noise cancellation (ANC) [4] has been used for the real time noise robust speech recognition which uses additional hardware component at microphone to cancel the noise from the source. This method is better suited for low frequency noise reduction, but performance of such a system depends upon the location of cancellation source. Furthermore, cost of implementation is larger. Noise compensation has been used for noise minimization to develop robust SR system. Noise compensation is used to find missing component of speech but cannot guarantee.

Major limitation of this method is that it cannot be implemented for real time. Microphone array has been for speech enhancement. For this system DSP system [5] is needed. Because of additional hardware requirement, it is less feasible for low end devices. A biologically inspired artificial neuron network has been used for speech recognition in different languages in recent days with Malay language vowel classification etc. The machine learning based models are trained with clean speech data generally gives better performance when it is tested on clean test data but performs unsatisfactory when it is tested on noisy testing data. Thus, to improve the classification accuracy even in the noisy environment the deep learning models are suitable.

The remainder of the paper is structured as: Literature survey conducted is covered in section 2 with their advantages and drawbacks. The section 3 covers the detailed analysis of proposed speech recognition method while in section 4, the results obtained from simulations and comparisons with literatures are discussed. Finally, section 5 has the remarks that conclude the outcomes from the presented research work with possible future implementations.

2. LITERATURE SURVEY

Many hybrid models have been proposed by researchers, and the success of the models are assessed by the recognition accuracy. Author in [6] proposed a RASTA-LPC and DWT implementation approach for ASR. The speech recognition was finished with near assessment between customary Continuous-LPC and the new approach RASTA with DWT. This approach incorporated word reference format model for each word in the preparation period of the hybrid framework. This method proved an overall improvement in the performance. In [7] authors had intended to enhance the intelligibility of conversational speech produced by any speakers so that hearing-impaired people can easily understand the contents. It would be an assistive system for hearing-aid people in their speech communication with the help of electroencephalography (EEG) based features. They have also tried to enhance the intelligibility of speech produced speech impaired people with dysarthria, who have problems in articulation, for example, ataxic, Àaccid, and hyperkinetic in their paper, also using conversion techniques. In [8] authors proposed speaker-autonomous word recognition in light of numerous word formats utilizing the recurrent neural network (RNN) technique. This framework was tried on disconnected and associated models of speech. Phoneme-like layouts and different word formats are chosen consequently by framing.Nassifet al. [9] conducted the detailed survey on the speech recognition performances of a hybrid model with multiple front ends, neural network model with radial basis function which performed the frame labelling and the other. And presented a discrete HMM have been compared and both the front ends performed poorly with the deep learning based neural network front end producing a better result. Many techniques have been tried to suggest improving the accuracy of deep recognition. Moritzet al. [10] proposed speech recognition interpretation invariant back proliferation type network for separate word-based vocabulary framework. It performed superior to anything a complex nonstop acoustic parameter HMM show on a noisy environment. Speaker autonomous separated word recognition framework in light of HMM with 12 words vocabulary is proposed. A few LPC portrayals are tried as feature vectors and the outcome are compared with existing results.

Krishna et al. [11] presented an ASR system for people who became hearing impaired post lingual with database of 34. Segregated words and words in sentences were tried at three levels of understandability. Subsequently, for each subject, three evaluations of k factors were gotten. Furthermore, sound-related, visual, and sound related visual sentence recognition was assessed utilizing regular method of sentence production. Author in [12] developed a Deep Neural Networks (DNNs) based separate word speech recognition system. Vector quantization is utilized to refer layouts for recognition framework. LPC parameters were utilized as features for recognition. In [13], authors proposed hybridized LSTM and MFCC approach for ASR. The trial was finished with near assessment between customary continuous-HMM (GHMM) and the new approach LSTM and MFCC. This approach incorporated word reference format model for each word in the preparation period of the Hybrid framework. This method proved an overall improvement in the performance. Li et al. [14] had tested the ASR performance for especially cerebral palsy in the task of controlling home electronics by utilizing the RNN models. In their later work, they have proposed a robust feature extraction method for ASR systems of speaker-impaired people. Jiang et al. developed a speaker independent system for controlling the movements of a robot using voice commands [15]. Back propagation unsupervised deep learning algorithm was used to train neural network.

Author in [16] introduced the means required in the outline of a speaker-autonomous speech recognition framework using RNN transducer. This examination work principally centered on the pre-handling stage that concentrates remarkable features of a speech signal and thinks about that feature utilizing RNN models. These procedures are connected for recognition of disengaged and in addition associated spoken word. Kim et al. [17] proposed confined speech recognition framework for CTC. In this framework confined word is perceived from their shape utilizing n monotonic chunk-wise attention (MoChA). Individuals with CTC can make utilization of this MoChA. Shan et al. used a linear spectral fusion(LSF) along with a maximum mutual information criterion (MMI) because MMI does not require evenly distributed training data and could increase the probability of training data, MMI is combined with shallow fusion and cold fusion in order to achieve rapid adaptation using one word of adaptation data [18]. The MMI-LSF algorithms have been found to provide a relative decrease in the error rate of about 11.1% using only 0.25s of adaptation data. A study for integrating voice command for movement of wheelchair for obstacles avoidance is addressed in [19]. The hybrid method by utilizing the HMM based CNN is

obtained by mixing zero crossing, extremes, dynamic time wrapping, energy and fundamental frequency. This ASR is mainly used for handicapped person by simple voice message. The use of extended HMM based CNN Law to acquire target continuously and keep on updating. The acquired data helped to find the obstacles in acquiring targets with automatic speech analysis. Author in [20] proposed a method which used LPC coefficients decomposed for audio-to-byte (A2B) and byte-to-audio (B2A) conversions. These LPC coefficients provided better representation. The proposed method provided good recognition using B2A it is evaluated using isolated digits in language amidst the presence of white Gaussian noise.

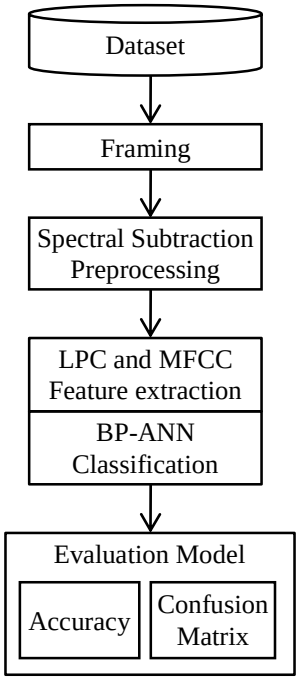

Fig.1. Proposed methodology of ASR system.

3. PROPOSED METHOD

The proposed speech recognition process consisting of two major processes namely training and testing. In the training phase different sources of speech signals are trained using the MFCC and LPC features with the help of back propagated ANN network and all the trained is stored into the dataset, respectively. During the testing phase, random speech signal is applied for the purpose speech recognition system as shown in Fig.1. The test signal is initially passed through the framing step; here the different frequencies of speech signal will be extracted. Thus, it is easy to remove the noise frequency from the incoming signal. For removal of noise spectral subtraction method is utilized effectively in preprocessing stage. Then, the combinations of LPC and MFCC features are extracted from the noise free speech signal and these features are applied to back propagate ANN respectively. Finally, various types of quality evaluation have been performed on the system to measure the accuracy of recognition. The detailed operation of each stage as follows:

3.1 SPECTRAL SUBTRACTION PREPROCESSING

Noise is a general term used for any unwanted or unknown information attached with the signal when recorded, transmitted, processed, or stored. Noise also means some random signal which does not have any specific information attached to it. Noise reduction should be done to remove or reduce this noise from the original signal in order to produce effective result. De-noising is done not only to remove the noises but also to improve the quality of speech signal by separating any independent signals attached to the original signal. De-noising algorithms can be classified as two types based on their domains as Spectral Subtraction and Filtering algorithm.

Spectral Subtraction (SS) is the process of subtracting spectrum noise from noisy signal spectrum. Spectral subtraction can be applied in applications where noise is accessed in separate channel. The main advantage of this method is its less complexity nature. Consider the following signal model

$$Y(n) = X(n) + N(n) \quad (1)$$

where $Y(n)$ will be the signal, $X(n)$ the additive noise and $N(n)$ noisy signal. Discrete time index is represented as n. Taking Fourier transform on the Eq.(1) gives,

$$Y(f) = X(f) + N(f) \quad (2)$$

It is the frequency domain of the Eq.(2), where f is the frequency variable. In SS, input signal is buffered and divided equally into segments of length N. The segments are windowed using Hamming window. Discrete Fourier Transform (DFT) is used to transform the signal to N spectral sample.

$$Y_w(n) = w(n)*y(n)$$
$$Y_w(n) = w(n)*[x(n)+n(n)]$$
$$Y_w(n) = x_w(n) + n_w(n) \quad (3)$$

where frequency domain of windowing is,

$$Y_w(f) = W(f) * Y(f) \quad (4)$$

where '*' is convolution. Spectral subtraction block diagram is shown in Fig.2 and spectral subtraction equation can be represented as:

$$\left|\hat{X}(f)\right|^b = \left|Y(f)\right|^b - \alpha\left|\bar{N}(\bar{f})\right|^{\bar{b}} \quad (5)$$

where $\left|\left|\hat{X}(f)\right|^b\right.$ will be original signal spectrum estimate. Here noise is considered a stationary random process, or it varies slowly.

Discrete Fourier transformation (DFT) is used for converting time domain into frequency domain followed by Magnitude operator. To reduce noise variant distortions, a low pass filter is used. Spectral subtraction induces distractions; to remove it post processing is done. Initially DFT is used for converting time to frequency domain, an Inverse DFT is used to convert the signal back to time domain.

3.2 FEATURE EXTRACTION

In any type of deep learning, extracting features is considered an important process; due to this reason any features which are obtained from these processes directly affects the efficiency of any classification process. Moreover, feature extraction is the major stage of any intelligent system, which likely removes redundant data and only intrinsic value of the actual original data is present. Thus, performing feature extraction affirms the significant information. In a speech signal, spectral feature set characterizes the properties of the signal in the frequency domain.

The DFT is applied to the signal to obtain this feature set. This is because DFT of a signal gives a high dimensional representation of a speech signal with very distinct spectral details. The DFT spectrum which is generated is then transformed into a more compact feature set which represent the speech signal and is used in speech related tasks. These representations may be in the form of MFCCs, LPCs etc., which are used in the synthesis recognition process. It helps provide additional information to the prosodic features which proves to be especially useful.

3.2.1 LPC:

LPC is a very powerful and popular method for feature extraction in speech technology. Here, the basic technique is to predict the present speech sample based on the linear combination of the past speech samples. It has widespread use in the area of speech research because it is simple to implement, computationally fast and mathematically precise. It is used for the feature extraction of speech relating to vowels, consonants, syllables, phonemes, isolated words etc. It is a digital method in which encoding is done for an analog signal where the particular value of the signal is predicted from the previous value with the help of a linear function. Linear prediction is a method which has wide application in other areas too. The LPC model is mainly used because at any time, say n, a speech sample, $S(n)$, can be obtained by linearly combining the previous samples of speech, i.e.

$$s(n) \approx a_1 s(n\text{-}1)+a_2 s(n\text{-}2)+\ldots+ a_n s(n\text{-}p) \tag{6}$$

where the coefficients are $a_1, a_2, a_3,\ldots,a_n$ are assumed to remain constant through the speech analysis frame. Thus, by adding an excitation term $G_u(n)$, the above equation can be converted into an equality as follows:

$$s(n) = Gu(n) + \sum_{i=1}^{p} a_i s(n-1) \tag{7}$$

By expressing Eq.(7) in Z domain, the relation given below can be derived:

$$S(n) = Gu + \sum_{i=1}^{p} a_i z^{-1} s(z) \tag{8}$$

The relation can be converted to a transfer function as shown below:

$$H(z) = \frac{s(z)}{Gu(z)} = \frac{1}{1-\sum_{i=1}^{p} a_i z^{-1}} = \frac{1}{A(z)} \tag{9}$$

In this work, feature vectors based on LPC along with other important parameters containing phonetic information are used. Feature vector having Cepstral weights are extracted by processing each frame of a signal having continuous speech. Then by using an analog-to-digital converter, sampling and quantization is carried out. The signal is flattened spectrally by pre-emphasizing the speech signal. This is done by using a 1st order digital filter. The transfer function of the digital filter is given by:

$$H(z) = 1 - az^{-1},\ for\ 0 \le a \le 1 \tag{10}$$

Here, signal having consecutive speech is considered as a single frame. To overcome the drawbacks of Gibbs phenomenon, a windows function called the Hamming window is used. The frames are multiplied by this function. It is computed as:

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi m}{N-1}\right),\ for 0 \le n \le N \tag{11}$$

Here, the total number of samples in the block is denoted by N. All the frames in the signal are now auto correlated to yield:

$$r_f(m) = \sum_{n=0}^{N-m} \tilde{x}_f(n)\tilde{x}_f(n+m), m = 0,1,2...,p \tag{12}$$

Here, the biggest auto-correlated value denotes the order of LPC analysis

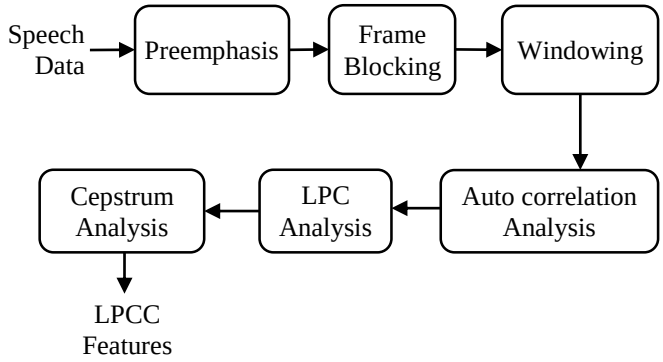

Fig.3. LPC operation diagram

3.2.2 MFCC:

Pre-emphasis is the initial stage of extraction. It is the process of boosting the energy in high frequency. It is done because the spectrum for voice segments has more energy at lower frequencies than higher frequencies. This is called spectral tilt which is caused by the nature of the glottal pulse. Boosting high-frequency energy gives more info to Acoustic Model which improves phone recognition performance. MFCC can be extracted by following method.

Step 1: The given speech signal is divided into frames (~20ms). The length of time between successive frames is typically 5-10ms.

Step 2: Hamming window is used to multiply the above frames to maintain the continuity of the signal. Application of hamming window avoids Gibbs phenomenon. Hamming window is multiplied to every frame of the signal to maintain the continuity in the start and stop point of frame and to avoid hasty changes at end point. Further, hamming window is applied to each frame to collect the closest frequency component together.

Step 3: Mel spectrum is obtained by applying Mel-scale filter bank on DFT power spectrum. Mel-filter concentrates more on the significant part of the spectrum to get data values. Mel-filter bank is a series of triangular band pass filters similar to the human auditory system. The filter bank consists of overlapping filters. Each filter output is the sum of the energy of certain frequency bands. Higher sensitivity of the human ear to lower frequencies is modeled with this procedure. The energy within the frame is also an important feature to be obtained. Compute the logarithm of the square magnitude of the output of Mel-filter bank. Human response to signal level is logarithm. Humans are less sensitive to small changes in energy at high energy than small changes at

low energy. Logarithm compresses dynamic range of values.

Step 4: Mel-scaling and smoothing (pull to right). Mel scale is approximately linear below 1 kHz and logarithmic above 1 kHz.

Step 5: Compute the logarithm of the square magnitude of the output of Mel filter bank.

Step 6: DCT is further stage in MFCC which converts the frequency domain signal into time domain and minimizes the redundancy in data which may neglect the smaller temporal variations in the signal. Mel-cepstrum is obtained by applying DCT on the logarithm of the mel-spectrum. DCT is used to reduce the number of feature dimensions. It reduces spectral correlation between filter bank coefficients. Low dimensionality and 17 uncorrelated features are desirable for any statistical classifier. The cepstral coefficients do not capture the energy. So, it is necessary to add energy feature. Thus twelve (12) Mel Frequency Cepstral Coefficients plus one (1) energy coefficient are extracted. These thirteen (13) features are generally known as base features.

Step 7: Obtain MFCC features.

The MFCC i.e. frequency transformed to the cepstral coefficients and the cepstral coefficients transformed to the MFCC by using the equation.

$$mel(f) = 2595 \times \log_{10}\left(1 + \frac{f}{100}\right) \quad (13)$$

where *f* denotes the frequency in Hz The step followed to compute MFCC. The MFCC features are estimated by using the following equation.

$$C_n = \sum_{n=1}^{K} (\log S_k)\left[n(K-0.5)\frac{\pi}{K}\right], n = 1, 2..., K \quad (14)$$

Here, *K* represents the number of Mel cepstral coefficient, C_0 is left out of the DCT because it represents the mean value of the input speech signal which contains no significant speech related information. For each of the frames (approx. 20ms) of speech that has overlapped, an acoustic vector consisting of MFCC is computed. This set of coefficients represents as well as recognize the characteristics of the speech.

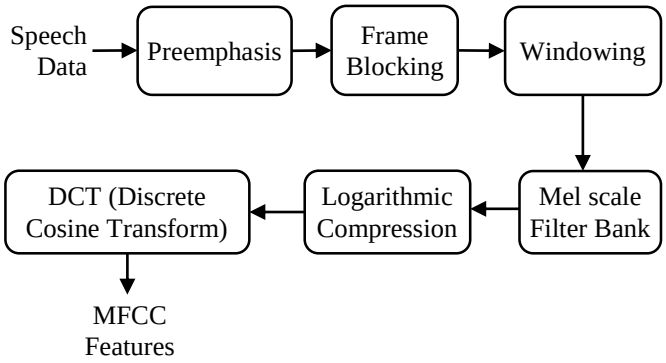

Fig.4. MFCC operation diagram

3.3 BP-ANN CLASSIFICATION

Neural networks have been effectively applied across a range of problem domains like audio, medicine, engineering, geology, physics, and biology. From a statistical viewpoint, neural networks are interesting because of their potential use in prediction and classification problems. BP-ANN is a method developed using emulation of birth neural scheme. The neurons are connected in the predefined architecture for effectively performing the classification operation. Depending on the LPC and MFCC features, the weights of the neurons are created. Then, the relationships between weights are identified using its characteristic features. The quantity of weights decides the levels of layers for the proposed network. BP-ANN basically consists of two stages for classification such as training and testing. The process of training will be performed based on the layer-based architecture. The input layer is used to perform the mapping operation on the input dataset; the features of this dataset are categorized into weight distributions.

The BP-ANN architecture has eight layers with weights. It contains the sequence of three alternating Convolutional2D layer and MaxPooling2D layer and three fully connected layers. The first convolutional2D layer of the net takes in 224*224*3 samples of speech signals and applies 96 11×11 filters at stride 4 samples, followed by a ReLU activation layer and cross channel normalization layer. The second layer (MaxPooling) contains 3*3 filters applied at stride 2 samples and zero paddings. Next convolutional 2Dlayer applies 5 256*256 sample filters at stride 4 samples, followed by max pooling2D layer which contains 3×3 samples filters applied at stride 2 samples and zero paddings. The third convolutional 2D layer of the net takes applies 384 3×3 filters at stride 1 sample and one padding. The last dense layer of the BP-ANN contains three fully connected layers with ReLU activation and 50% dropout to give 60 million parameters. Then the classification operation was implemented in the two levels of hidden layer as shown in Fig.5. The two levels of hidden layer hold individually normality and abnormalities of the ASR characteristic information. Based on the features criteria, it is categorized as normal and abnormal classification stage. These two levels are mapped as labels in output layer. When the test speech signal is applied, its LPC and MFCC features are applied for testing purpose in the classification stage. Based on the maximum feature matching criteria utilizing Euclidean distance manner it will function. If the feature match occurred with hidden layer 1 label, then it is classified as recognized speech signal from the database, respectively.

4. SIMULATION RESULTS

This section focuses on how best to build a syllable speech recognition system that is based. Many attempts have been attracted by the outstanding properties of earning syllable as the essential acoustic modeling components. The database contains 2 words per speaker which are same across all the 50 speakers. A total of 200 unique syllables is present in the training data and 50 unique syllables in the core test set. Test syllables which are not in the training data are replaced with corresponding phonemes.

Table.1. Training times required by the various methods

Training method	Pre-training time	Fine-tuning training time
SVM [7]	3.4 hours	1.35 hours
RNN [8]	1.8 hours	1.24 hours
BP-ANN	52 minutes	48 minutes

Table 2. Performance of the ASR system using different features

Training method	Recognition Accuracy (%)
LBP	59.41
LPC	73.27
MFCC	85.15
LPC+ MFCC	98.33

From the Table.1, it is observed that the proposed BP-ANN deep learning based training model has faster training performance compared to the state of art approaches SVM [7] and RNN [8], as the proposed method has consisting of multi layers with error propagation mechanism. From the table 2, it is observed that the proposed combination of LPC+ MFCC features haves higher Recognition Accuracy compared to the individual features. The performance metrics used to evaluate the proposed methods are accuracy (*AC*), recall (*RE*), and specificity (*SP*). Let *TP*, *TN*, *FP* and *FN* be the count of true positive, true negative, false positive, and false negative, respectively. Then the equations are shown in following equations:

Accuracy: It is defined as the number of data points predicted correctly to the total sum of all data points.

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

Recall: It tells the proportion of the speech signal is recognized and tested positive.

$$RE = \frac{TP}{TP + FN} \tag{16}$$

Precision: It tells the proportion of the speech signal is recognized more precisely.

$$PR = \frac{TP}{TP + FP} \tag{17}$$

F1-Score: it is calculated using precision and recall as follows:

$$F1 = 2\frac{PR \times RE}{PR + RE} \tag{18}$$

Table.3. Performance comparison obtained quality metrics of proposed ASR system with existing ASR systems.

Method	Accuracy	Recall	Precision	F1-score
SVM [7]	87	92	83	87.26
HMM [9]	91	93	89.5	91.47
RASTA–LPC-DWT [6]	89.5	57.3	83.4	67.58
RNN [8]	97.49	94.3	95.6	94.40
Proposed	98.33	98.93	97.73	97.49

From the qualitative evaluation, it is observed that the proposed method can effectively shows the better performance of speech recognition compared to the conventional approaches SVM [7], HMM [9], and RNN [8] as the proposed methodology utilities the hybrid LPC and MFCC features respectively.

5. CONCLUSION

This article majorly focusing on the development of speech recognition by utilizing the hybrid features such LPC and MFCC features, respectively. By using the spectral subtraction method in preprocessing stage, they were effectively removed the noise from the speech with the capable of effective extraction of source speech from noisy environment. The feature extraction has performed by LPC and MFCC method very accurately with all the types of features including echo-based phase variations. This work can be effectively extended to implement the recognition of variety of emotions from the speech signal.

REFERENCES

[1] H. Liu and H. Motoda, "Feature Selection for Knowledge Discovery and Data Mining", *Kluwer Academic Publishers*, 2012.

[2] X. Tang, Y. Dai and Y. Xiang, "Feature Selection based on Feature Interactions with Application to Text Categorization", *Expert Systems with Applications*, Vol. 120, pp. 207-216, 2019.

[3] K. Scarfone aand P. Mell, "Guide to Intrusion Detection and Prevention Systems (IDPS)", Technical Report, National Institute of Standards and Technology, pp. 1-78, 2012.

[4] S. Mohammadi, H. Mirvaziri, M. Ghazizadeh Ahsaee and H. Karimipour, "Cyber Intrusion Detection by Combined Feature Selection Algorithm", *Journal of Information Security and Applications*, Vol. 44, No. 2, pp. 80-88, 2019.

[5] S. Zaman and F. Karray, "Features selection for intrusion detection systems based on support vector machines", *Proceedings of 6th IEEE International Conference on Consumer Communications and Networking*, pp. 1-8, 2009.

[6] S. Maza and M. Touahria, "Feature Selection Algorithms in Intrusion Detection System: A Survey", *KSII Transactions on Internet and Information Systems*, Vol. 12, No. 10, pp. 1-14, 2018.

[7] K. Chen, F.Y. Zhou and X.F. Yuan, "Hybrid Particle Swarm Optimization with Spiral-Shaped Mechanism for Feature Selection", *Expert Systems with Applications*, Vol. 128, pp. 140-156, 2019.

[8] M. Keshtgary and N. Rikhtegar, N., "Intrusion Detection Based on a Novel Hybrid Learning Approach", *Journal of AI and Data Mining*, Vol. 6, No. 1, pp. 157-162, 2018.

[9] N. Acharya and S. Singh, "An IWD-Based Feature Selection Method for Intrusion Detection System.", *Soft Computing*, Vol. 22, No. 13, pp. 407-416, 2018.

[10] A.S. Eesa, Z. Orman and A.M.A. Brifcani, "A New Feature Selection Model Based on ID3 and Bees Algorithm for Intrusion Detection System", *Turkish Journal of Electrical Engineering and Computer Sciences*, Vol. 23, No. 2, pp. 615-622, 2015.

[11] E. Zorarpaci and S.A. Ozel, "A Hybrid Approach of Differential Evolution and Artificial Bee Colony for Feature

Selection", *Expert Systems with Applications*, Vol. 62, pp. 91-103, 2016.

[12] Barnali Sahu,Satchidananda Dehuri and Alok Jagadev, "A Study on the Relevance of Feature Selection Methods in Microarray Data", *The Open Bioinformatics Journal*, Vol. 11, No. 2, pp. 117-139, 2018.

[13] Swagatam Das, Arijit Biswas, Sambarta Dasgupta and Ajith Abraham, "*Bacterial Foraging Optimization Algorithm: Theoretical Foundations, Analysis, and Applications*", Foundations of Computational Intelligence, Vol. 3, pp. 23-55, 2009.

[14] L.P. Dias and J.J.F. Cerqueira, "Using Artificial Neural Network in Intrusion Detection Systems to Computer Networks", *Proceedings of 9th IEEE International Conference on Computer Science and Electronic Engineering*, pp. 1-8, 2017.

[15] J. Martens and I. Sutskever, "Learning Recurrent Neural Networks with Hessian-Free Optimization", *Proceedings of 28th IEEE International Conference on Machine Learning*, pp. 1-6, 2011.

[16] Xuegong Zhang, Xin Lu and Qian Shi, "Recursive SVM Feature Selection and Sample Classification for Mass-Spectrometry and Microarray Data", *BMC Bioinformatics*, Vol. 7, No. 19, pp. 1-18, 2006.

[17] V.R. Shewale and H.D. Patil, "Performance Evaluation of Attack Detection Algorithms using Improved Hybrid IDS with Online Captured Data", *International Journal of Computer Applications*, Vol. 146, No. 8, pp. 1-12, 2016.

[18] S. Kalaivani and Gopinath Ganapath, "Bio-Inspired Modified Bees Colony Feature Selection based Intrusion Detection System for Cloud Computing Application", *International Journal of Advanced Science and Technology*, Vol. 29, No. 3, pp. 1-12, 2020.

[19] S. Kalaivani and Gopinath Ganapath, "Bacterial Foraging Optimization for Enhancing the Security in Intrusion Detection System", *International Journal of Scientific and Technology Research*, Vol. 9, No. 2, pp. 1-8, 2020.