I J C T A, 9(7), 2016, pp. 3401-3413

Performance Analysis of Syllable Based Tamil Language Robust Speech Recognition System Using Modified Group Delay Function, Gammatone Cepstral Coefficients, Hidden Markov Model and Deep Neural Network

Sundarapandiyan S.*, Shanthi N. and Mohamed Yoonus M.*****

ABSTRACT

In the developing Natural Language Processing (NLP) technologies, Automatic Speech Recognition (ASR) plays a vital role in transmitting the speech data with efficiency and accuracy. The speech has been recognized by using the automatic speech recognition process which uses various languages while detecting the speech. ASR system is available for various languages, the proposed system uses the Tamil language based automatic recognized system with the help of the different models such as speech segmentation for extracting the syllables based on Modified Group Delay Function with Gammatone Cepstral Coefficients (GCC) related features from the audio signal, hybrid Deep Neural Network (DNN) and Hidden Markov Model (HMM) based Acoustic model and DNN based Language model in an effective manner. The segmented syllables are classified by using the syllable based acoustic model using DNN – HMM with syllable based language model using DNN. So, in this paper, the performance of the system is evaluated with the help of the two different databases such as Indian Institute of Information Technology – Hyderabad (IIIT-H) and Indic speech corpus and Linguistic Data Consortium for Indian Language (LDC-IL) compared with various metrics such as Signal to Noise Ratio, Root Mean Square Error, Entropy, accuracy of extracted features, detectability, execution time, sensitivity, specificity, and recognition accuracy and so on.

Keywords: Speech Recognition, Syllable segmentation, Acoustic model, Language model, Modified Group Delay Function with Gammatone Cepstral Coefficients, Deep Neural Network with the Hidden Markov Model process, IIIT-H Indic speech corpus speech database and LDC-IL speech database.

1. INTRODUCTION

Speech is the important communication medium for making the interaction between the people; also it is a sub-field of the computational linguistic which is used in various research areas such as multi-model interaction, electrical engineering, automatic translation [1]. Mostly the automatic speech recognition system uses various languages such as English, Sanskrit, Telugu, and Tamil and so on, for detecting the transferred speech data efficiently. Even though various languages present in the recognition system, Tamil language based recognition system placed a crucial role because it is the important Dravidian language. The only difference between the other languages and Tamil is with word arrangements and formation of the word. Also the Tamil language has various consonants, vowels and rhythm [2]. Based on the consonants, vowels, the Tamil text has been segmented into different syllable which helps to detect the spoken language in an

* Assistant Professor, Department of Computer Science and Engineering, Periyar Maniammai University, Vallam.

** Professor, Department of Computer Science & Engineering, Nandha Engineering College, Erode.

*** Senior Lecturer/Junior Research Officer, Linguistic Data Consortium for Indian Languages (LDC-IL), Central Institute of Indian Languages, Mysore.

தென்மேற்குப் பருவமழை இன்று தொடங்கும்

தென் + மேற் + குப் பரு + வம + ழை இன் + று தொடங் + கும்

effective manner. The sample Tamil spoken language text and the related syllable segmentation are shown as follows,

The above-spoken language has several consonants and vowels which are difficult to process and identify. So various speech recognition models such as syllable segmentation, acoustic model, language models are used to recognize the spoken text with high accuracy. So, this paper uses the three models for recognizing the Tamil spoken language by applying various methods. Initially the syllable is segmented with the help of the Modified Group Delay Function with Gammatone Cepstral Coefficients. The extracted features are used to create the acoustic model with Deep Neural Network with Hidden Markov Model [3] and the matching processing is enhanced by using the language model such as Deep Neural Network [4] which matches the testing features with trained features. Further this paper analyzes the efficiency of these proposed methods of using the IIIT-H Indic speech corpus [5] and LDC-IL [6] dataset in terms Signal to Noise Ratio, Root Mean Square Error, Entropy, accuracy of extracted features, detectability, execution time, sensitivity, specificity, and recognition accuracy and so on which is compared with various existing methods such as Mel Frequency Cepstral Coefficients with Hidden Markov Model, Mel Frequency Cepstral Coefficients with Deep Neural Networks, Gammatone Cepstral Coefficients with Hidden Neural Networks and Gammatone Cepstral Coefficients with Deep Neural Network

Apart from introducing this paper is organized as follows, Section 2 summarizes the related works for speech recognition system, Section 3 deals with the short descriptions about the proposed methodology and Section 4 discusses the performance analysis of the proposed methodologies and the Section 5 describes the conclusion.

2. RELATED WORKS

This section discusses various analyses about the speech recognition process.SankarBabu et al., [7] develop the automatic isolated speech recognition system by using the large-vocabulary continuous speech recognition process. The system analyzes the spoken language with isolated manner and encrypts the each text with a particular encryption key and stored in the database. The test features are decrypting the key and they match the spoken language by using the converter process. Then the performance of the system is implemented using the MATLAB tool and the efficiency is analyzed using the testing and training data set for both male and female voice in an effective manner.

Iswarya et al., [8] analyze the performance of the various Tamil speech feature extraction methods such as Linear predictive Cepstral coefficient and Mel-frequency Cepstral coefficient. These methods analyze the features with different directions and frames like 8, 12, 24 and extract the isolated features efficiently. The extracted features are classified by applying the probabilistic neural network which obtains the 97% recognition accuracy while recognizing Tamil speech characters.

Hema et al, [9] analyze the continuous speech and recognizing the speech by using the automatic segmentation process. The method analyzes the speech and segment text into the syllables and the similar syllables are clustered together. Based on the clusters, a separate model has been generated and labeled for each cluster to recognize the text based on the closed speaker value. The performance of the system is analyzed using the Tamil and Telugu database which achieve the 43.3% and 32.9% accuracy.

Alex Graves et al., [10] recognize the speech features by applying the deep neural network because it works well for sequential data. The network works based on the long short term memory process to analyze the interconnection between the speech features. The extracted features are classified in terms of the connectionist temporal classification process. The implemented system reduces the error rate up to 17.7% which is analyzed using the TIMIT phoneme recognition database effectively.

Mark Galeset al., [11] implement the large vocabulary continuous speech recognition system for improving the recognition rate. The author reduces the assumptions about the particular speech features which are classified by applying the Hidden Markov Model. This model uses various processes like feature projection, discriminative parameter estimation, covariance modeling, adaption, normalization, multipass and noise compensation process while detecting the speech feature.

3. PROPOSED METHODOLOGY

In this paper, the Tamil language based speech recognition system is implemented by utilizing the IIIT-H Indic speech corpus and LDC-IL dataset. It has three different stages such as syllable segmentation, feature extraction and recognition stage [12]. During the recognition process, the system uses both acoustic and language model for improving the recognition rate. The proposed system block diagram is shown in the following figure 1. The first stage of the process is syllable segmentation in which the Fourier transform has been computed of $x(n)$ and $nx(n)$ and choose $X(k)$ and $Y(k)$. From the Fourier transform, the cepstrally smoothen is estimated to $|X(k)|$ and $S(\omega)^2$. Then the group delay function is calculated as follows,

$$\tau(\omega) = \left(\frac{X_R(\omega) Y_R(\omega) + Y_I(\omega) X_I(\omega)}{S(\omega)^{2\gamma}} \right) \tag{1}$$

Finally the tune the parameters γ, α depending on the speech environment. Based on the environment, the speech text syllables are extracted, then Gammatone wavelet coefficients are applied to extract the features as follows.

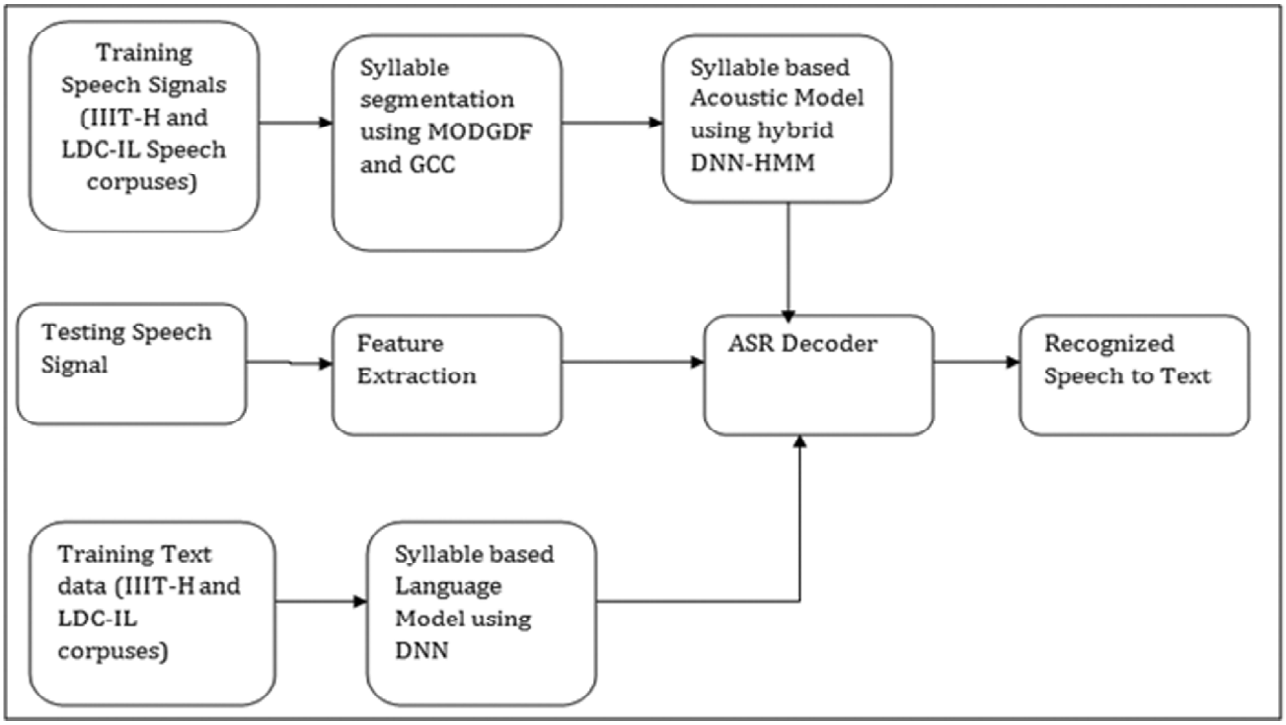
Training Speech Signals (IIIT-H and LDC-IL Speech corpuses)
Syllable segmentation using MODGDF and GCC
Syllable based Acoustic Model using hybrid DNN-HMM
Testing Speech Signal
Feature Extraction
ASR Decoder
Recognized Speech to Text
Training Text data (IIIT-H and LDC-IL corpuses)
Syllable based Language Model using DNN

Figure 1: Proposed System Block Diagram

From the extracted syllable windowing process is applied upto 10 to 50 ms.The windowing process is done with the help of the Hamming window which is represented as follows,

$$Y(n) = X(n)*W(n) \tag{2}$$

Where *Y*(*n*) is the output of the windowing signal, *X*(*n*) is the input of the audio signal, n is the number of samples present in each frame and *W* is the Hamming window. Then the FFT method is applied to the windowing signal to convert each frame of *N* samples to the frequency domain.

$$Y(w) = FFT\left[h(t) * X(t)\right] = H(w) * X(w) \tag{3}$$

The output of the FFT value is fed into the Gammatone filterbank which has the Equivalent rectangular bandwidth (ERB) that it used to extract the meaningful sound frequencies. The ERB is calculated as follows,

$$ERB(f) = \frac{f}{9.26} + 24.7 \tag{4}$$

After that the center frequency of the channel is computed as follows,

$$f_c(k) = -C + e\frac{k\log\left(\frac{f\ \min + C}{f\ \max + c}\right)}{K}.(f\ \max + C) \tag{5}$$

Where, fmin and fmax is the lowest and highest cutoff frequency of the filter bank. The estimated center frequency value is applied to Gammatone wavelet for obtaining the various features from the audio signal. The computed filter output is applied to the log energy function and discrete cosine transform to obtain the human voice based features in an effective manner. From the extracted feature the acoustic model has been created for recognizing the speech by matching the testing and training features.

3.1. Acoustic Model and Language Model

The acoustic model is created by using the deep neural network with Hidden Markov Model which works based on the IR based and ONC based model [13]. The model uses the lexicons that convert the word-level transcriptions into the mapped phones and sequences. The sequence has the independent states that help to retrieve probability value of the feature and the features are trained using the deep neural network. The network analyzes the non-linear relationship between each feature state which matches the complex speech data patterns. During the training process, the network uses the learning rate and creates the efficient acoustic model. The created acoustic model helps while matching the test features into the trained (database) features. During the matching process, the system uses the deep neural network to recognize the speech character with minimum time. Initially, the words are treated as the N-dimensional sparse vector that has the 1 and 0 index value. Based on the index value the input words are mapped using linear projections which create the lookup table that includes the list of words, vocabularies. When the new features are entered into the system, the neural network mapping the new entries with the look table in terms of the ith row in the feature space. The sequence feature vector is stored in the history which is concatenated with the projection layers that is used to retrieve the speech with minimum time. The collected information is summed in the output layer for achieving the target in an effective manner. During the target speech recognition process, the error rate has been reduced by updating the weights and bias value which is defined as follows,

$$d = \tan\ h\left(\sum_{l=1}^{(n-1)*P} M_{jlcl} + b_j\right)\forall_j = 1...H \tag{6}$$

$$o_i = \sum_{j=1}^{H} V_{ij} d_j + k_i \forall_i = 1,\ ...\ N \tag{7}$$

$$p_i = \frac{\exp(o_i)}{\sum_{r=1}^{N} \exp(o_r)} = P\left(w_j = \left\langle i \middle| h_j \right\rangle\right) \tag{8}$$

Where,

M_{jlcl} *represented as the weight of the projection and hidden layer in the network,*

V_{ij} *represented as the weight of the hidden layers and output layer in the network*

b_j *is the bias of the network,*

$P\left(w_j = \left\langle i \middle| h_j \right\rangle\right)$ *represented as the output layer posterior probability*

Based on the mapping process, the testing features are mapped with the trained acoustic features which efficiently retrieve the Tamil speech text with minimum computation time also it eliminates the errors present in the recognition process. Then the performance of the proposed system is analyzed using the experimental results and discussions which are explained as follows.

4. PERFORMANCE EVALUATION OF THE PROPOSED SYLLABLE-BASED TAMIL SPEECH RECOGNITION SYSTEM

This section discusses that the proposed syllable based Tamil speech recognition system by utilizing the IIIT-H Indic speech corpus and LDC-IL speech corpus effectively. The detailed description of the both dataset details is explained as follows.

4.1. IIIT-H Indic speech corpus

IIIT-H is the International Institute of Information Technology Hyderabad speech and vision laboratory database for making the research about the speech recognition. The dataset consists of various languages such as Bengali, Hindi, Kannada, Malayalam, Marathi, Tamil and Telugu related texts and speech data. While capturing this language data set, 1000 sentences were chosen for each language, it contains the moreover than 5000 most frequently used words in every language. In our proposed system, Tamil language is used to analyze the speech recognition process and the sample Tamil language related statistics is shown in table 1.

4.2. LDC-IL dataset

LDC-IL is the Linguistic Data Consortium for Indian Languages [14]. It is used to do the research in the fields of speech recognition, synthesis, character recognition and corpora creation process. The dataset contains the two Indian and English language related corpora speech records which were captured at the time of microphone, telephone both mobile and landline conversation. These data have been collected from150 males and 150 females which have more than 10,000 words. Also the dataset consists of several languages such as Bengali, Hindi, Konkani, Oriya, Malayalam, Punjabi and Tamil related speech records.

Table 1
I Indic speech corpus-based Tamil Speech text Corpus

S. No	*Text Corpus*	*No of Sentence*	*No of words*		*No of syllables*		*No of phones*	
			Total	*Unique*	*Total*	*Unique*	*Total*	*Unique*
1.	Wikipedia	99650	1888462	857850	3193292	10525	5688710	35
2	Optimal	1000	7045	2182	232284	930	42134	35

In our proposed system Tamil speech recognition system is used, in which the Tamil speech corpus is based on the monolingual language. The data were captured from the 450 different native speakers from Tamilnadu with various gender, regions and dialectical environment. Each of the recorded data has a proper noun, frequent word, phonetically balanced vocabulary, form, control words and command.

These captured speech corpora are processed by applying the syllable segmentation performed with the help of the Modified Group Delay Function with Gammatone Cepstral Coefficients and the speech is recognized by Deep Neural Network with Hidden Markov Model. Then the performance of the proposed system is evaluated with the help of various performance metrics which are discussed as follows.

4.3. Performance of the Syllable Segmentation Process

This section evaluates the performance of the syllable segmentation process. The method extracts the syllable from the each sentence which is used to extract the speech-related features. Then the segmented syllable efficiency is evaluated using various metrics such as Peak Signal to Noise Ratio (PSNR), Root Mean Square Error Value (RMSE), Mean Absolute Error (MAE), and Precision and Recall.

RMSE is the measure which is used to estimate that, whether, the proposed system efficiently segments the perfect syllable from the sequence of a sentence which is measured as follows.

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n} e_i^2} \tag{9}$$

Where e_i^2 is defined as sum-of-squares error for samples of model errors. Then the square error is calculated as follows.

$$MSE = \frac{1}{mn}\Sigma_0^{m-1}\Sigma_0^{n-1}\left\|p(i,j)-q(i,j)\right\|^2 \tag{10}$$

In addition, the efficiency of the segmented syllable has been evaluated using the PSNR measure which is defined as follows,

$$PSNR = 20\log_{10}\left[\frac{MAXPIX}{MSE}\right] \tag{11}$$

Where, m is represented as the number of rows of segmented syllable image pixel, i is the index of that row, n is defined as the number of columns of segmented syllable image pixels and j is defined as the index of that column. *MAXPIX* is defined as maximum signal value. If the PSNR value is high, the image quality is better otherwise image quality is low.

$$MAE = Max(abs\ (original\ pixel - degraded\ pixel)) \tag{12}$$

MAE is defined as the maximum absolute edge value, the difference between *original pixel – degraded pixel*. Finally, the entropy of the segmented syllable region is defined as follows,

$$Entropy = -k\ \Sigma_i p_i \ln p_i \tag{13}$$

Then the related graph representation of the error value and quality of the image is shown in figure 2 and 3.

The above figure 2 and 3 clearly show that the proposed method segments the image with minimum error rate (IIIT-H-0.32 pps, LDC-IL-0.27 pps) and high entropy value (IIIT-H-1.29 J, LDC-IL-1.4922 J) which means segments the image with high quality which is evaluated using the PSNR (IIIT-H-80.01 decibel, LDC-IL-82.01 decibel) metrics on both IIIT-H Indic speech corpus and LDC-IL database. Moreover, the segmented syllable accuracy and efficiency is evaluated using the precision and recall metrics. Precision and Recall are defined as follows

$$precision = \frac{True\ Positive}{True\ Positive\ +\ False\ Positive} \tag{14}$$

$$Recall = \frac{True\ Positive}{True\ Positive\ +\ False\ Positive} \tag{15}$$

These precision and recall indicate that how the proposed system segments the exact syllable which is shown in figure 4.

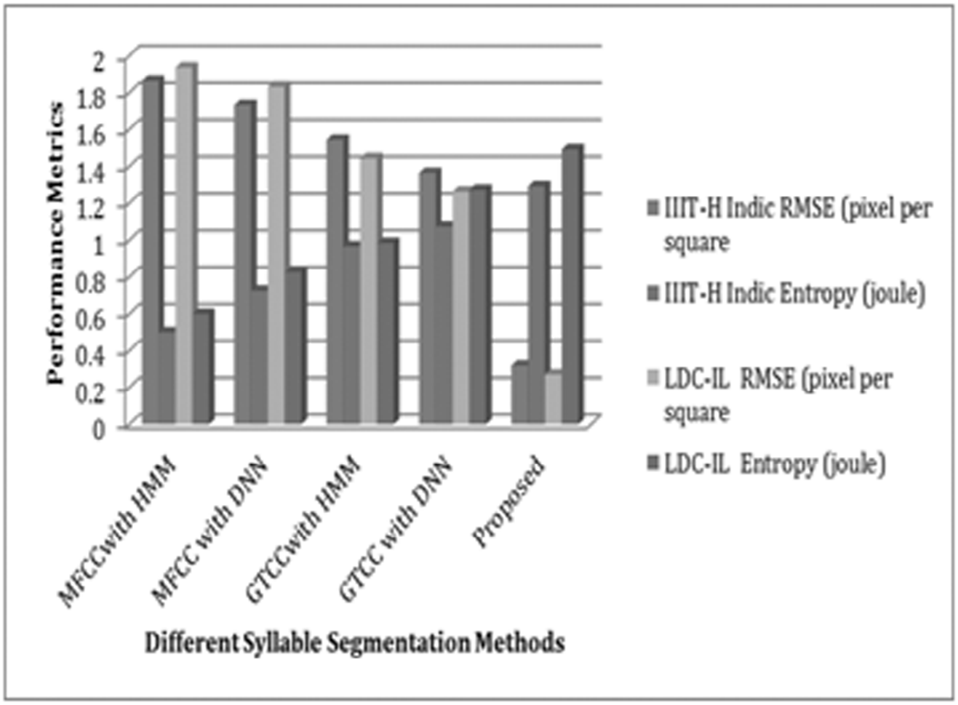

Figure 2: RMSE and Entropy

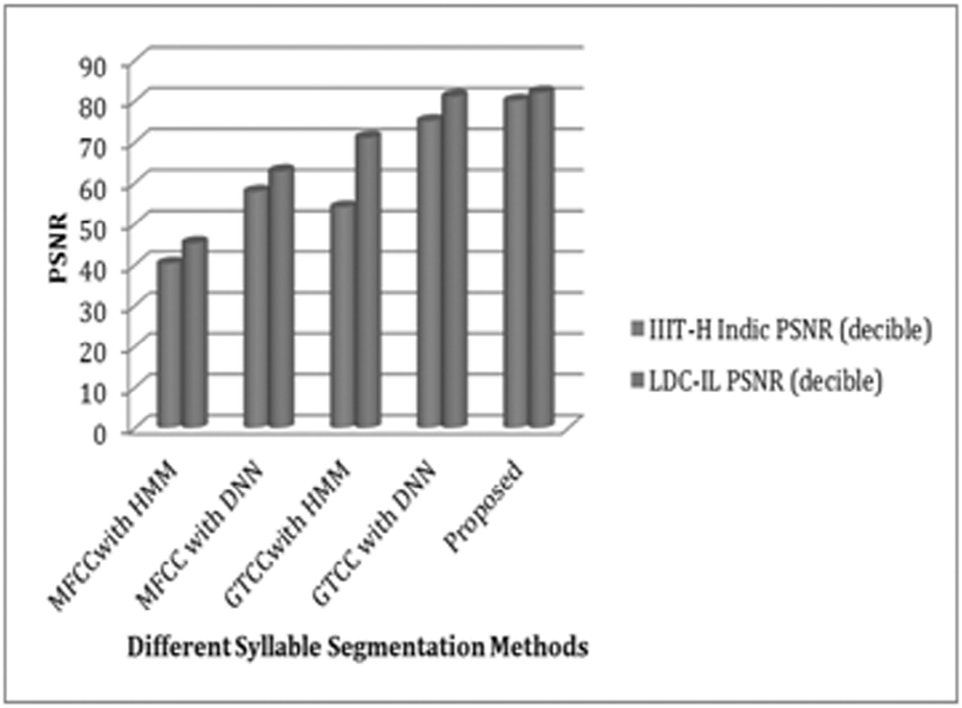

Figure 3: PSNR

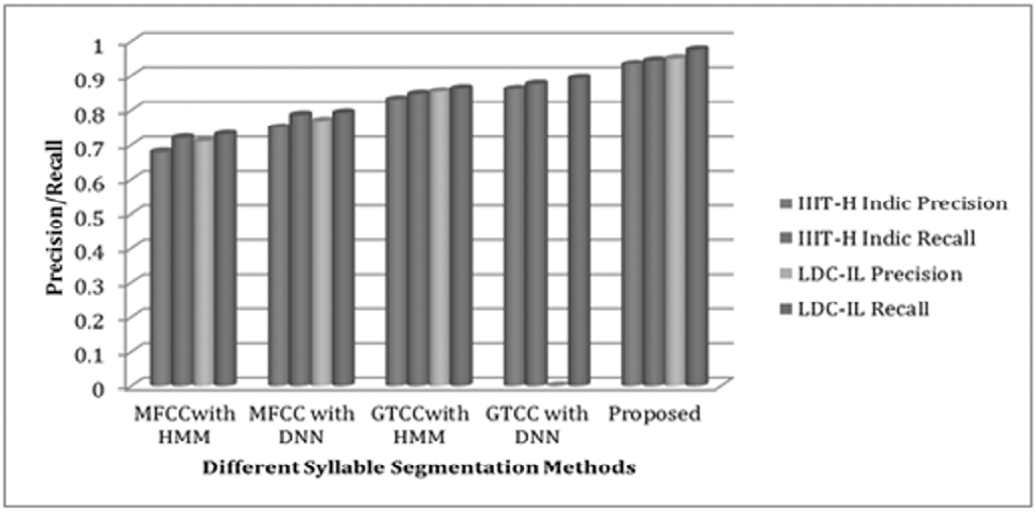

Figure 4: Precision and Recall

From the above figure 4,it is understood that the proposed method efficiently segment the exact syllable with highest values on both IIIT-H Indic speech corpus and LDC-IL database such as (IIIT-H-0.935, LDC-IL-0.952) precision value and (IIIT-H-0.946, LDC-IL-0.976) recall value. The segmented syllables are fed into the next feature extraction phase which is discussed as follows.

4.4. Performance of the Feature Extraction Process

This section discusses how the proposed system efficiently retrieves various speech-related features such as phone type, vowel length, lip rounding, vowel height, vowel frontness, consonant voicing, cluster, aspirations and place of articulation are extracted. The efficiency of the extracted feature performance is analyzed using the detectability of the extracted features and accuracy of the extracted features which are discussed as follows.

4.4.1. Accuracy

Accuracy is the very crucial measure which is used to identify elaborately how the proposed system efficiently retrieves the speech features. Then the resultant feature accuracy is compared with the existing methods such as MFCC with HMM [15], MFCC with DNN [16], GTCC with HMM [17] and GTCC with DNN[18] which is shown in figure 5. The accuracy of the feature is computed on both IIIT-H Indic speech corpus and LDC-IL database as follows,

$$Accuracy\ of\ \ the\ Feature = \frac{Number\ of\ \ feature\ extracted}{Total\ Number\ of\ \ feature} * 100 \tag{16}$$

From the above figure 5, it clearly known that the proposed MODGDF-GWCC with DNN-HMM method extract the speech relate features with 98.64% accuracy of IIIT-H dataset and 99.23% accuracy of LDC-IL dataset when compared to other existing methods such as MFCC with HMM(IIIT-H-77.36%, LDC-IL-78.43%), MFCC with DNN (IIIT-H-83.14%, LDC-IL-85.31%),GTCC with HMM (IIIT-H-85.47%, LDC-IL-87.67%) and GTCC with DNN (IIIT-H-88.67%, LDC-IL-99.23%). In addition the efficiency of the extracted feature is analyzed using the detectability metrics.

4.4.2. Detectability

Detectability is the measure which is used to analyze whether the extracted speech features are used to recognize the speech while matching the features. Then the performance of the proposed system detectability value is shown in the following figure 6.

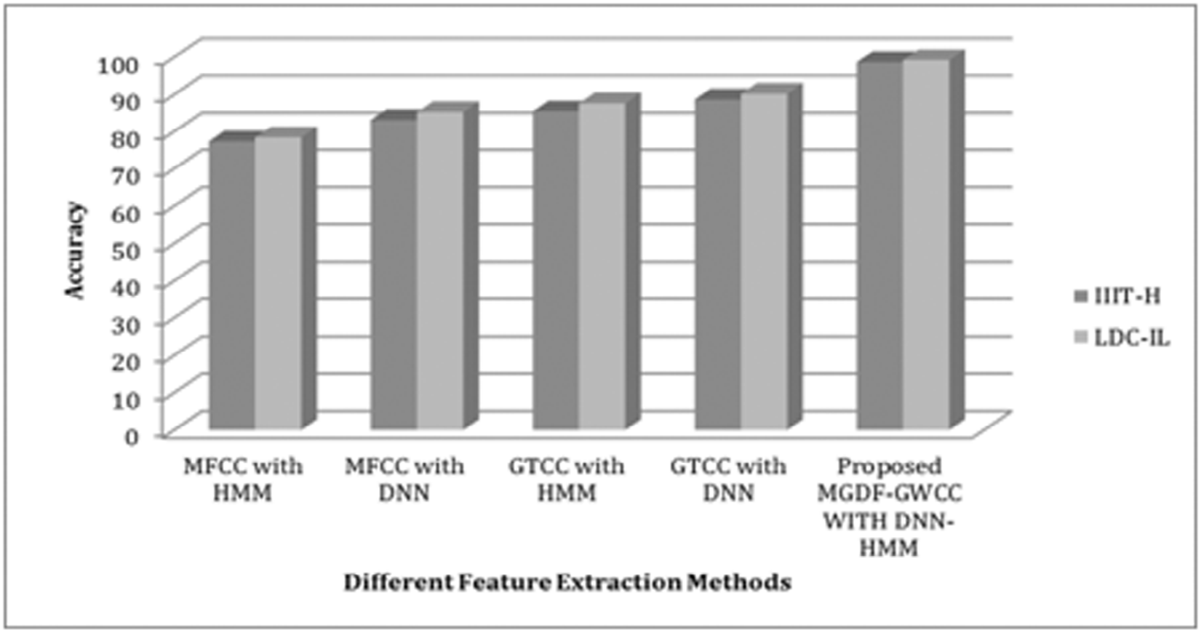

Figure 5: Accuracy of Extracted Features

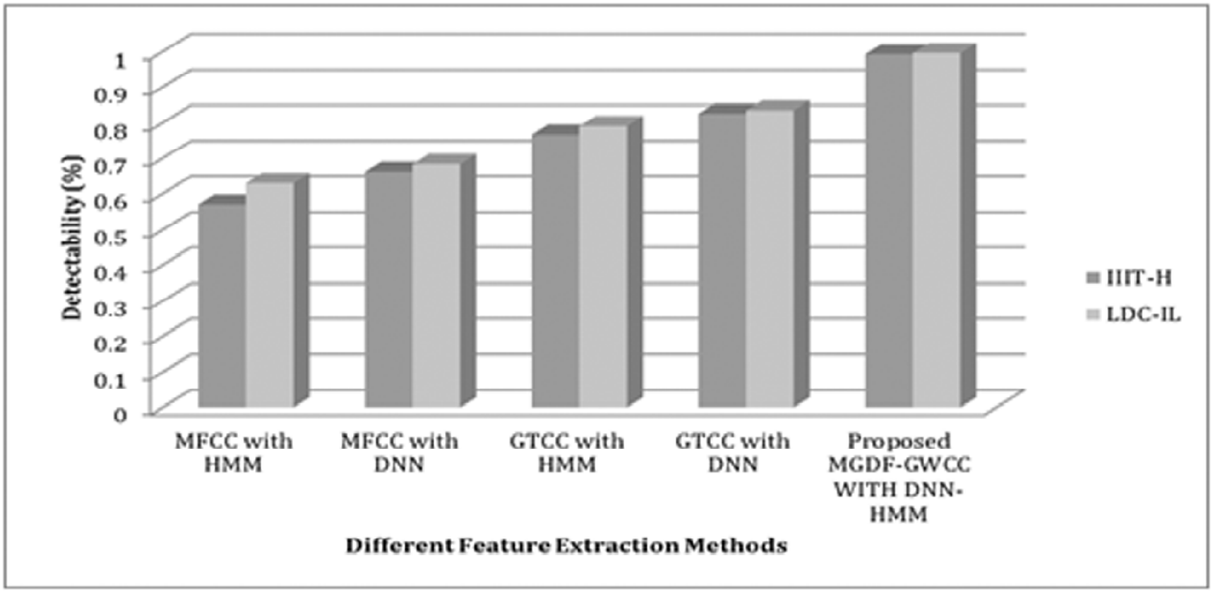

Figure 6: Detectability of Extracted Features

From the above figure 6, it could be inferred that the proposed MODGDF-GWCC with DNN-HMM method extract the speech features which is widely used during the speech recognition process and the detectability rate on both IIIT-H Indic speech corpus and LDC-IL databases such as 0.993% on IIIT-H and LDC-IL on 0.996% accuracy when compared to other existing methods such as MFCC with HMM(IIIT-H-0.57%, LDC-IL-0.63%), MFCC with DNN (IIIT-H-0.662%, LDC-IL-0.684%),GTCC with HMM (IIIT-H-0.767%,LDC-IL-0.791%) and GTCC with DNN (IIIT-H-0.826%,LDC-IL-0.835%). In addition the efficiency of the extracted feature is analyzed using the detectability metrics. Thus the extracted features are fed into speech recognition stage for classifying the features with minimum execution time which is discussed as follows.

4.5. Performance of the Speech Recognition Process

This section discusses that the efficiency of the speech recognition process in terms of the Sensitivity, Specificity, execution time and Accuracy is discussed. The extracted features are fed into the language model such as the deep neural network which consumes efficient sensitivity and specificity value. Then the sensitivity and specificity value are calculated as follows,

$$\text{Sensitivity} = TP/((TP + FN)) \tag{17}$$

$$\text{Specificity} = TN/((TN + FP)) \tag{18}$$

Where, TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

From the above equations, the sensitivity and specificity value is calculated which is shown in figure 7.

The following figure 7 shows that the proposed method achieves the high sensitivity (98.23%) and specificity (93.45%) value when compared to the other recognition methods such as MFCC with HMM (sen-76.34%, spec-71.23%), MFCC with DNN (sens-79.03%, spec-80.23%), CTCC with HMM (sen-82.45%, spec-83.42%),GTCC with DNN (sen-85.3%, spec-83.56%) on IIIT-H Indic speech corpus dataset. In addition the proposed method achieves the high value on LDC-IL database such as sensitivity (98.76%) and specificity (95.4%) value when compared to the other recognition methods such as MFCC with HMM (sen-78.25%, spec-73.19%), MFCC with DNN (sens-80.13%, spec-82.89%), CTCC with HMM (sen-84.6%, spec-86%),GTCC with DNN (sen-87.9%, spec-89.13%). The sensitivity and specificity value the speech recognition accuracy which is shown in figure 8.

The following above figure 8, shows that the proposed method achieves the high accuracy 98.3% value when compared to the other recognition methods such as MFCC with HMM (85%), MFCC with DNN (82.2%), GTCC with HMM (85.6%),GTCC with DNN (88.32%) on IIIT-H dataset. Moreover the proposed method also achieves the higher accuracy on LDC-IL dataset such as 99.23% high accuracy when compared to the MFCC with HMM (86.3%), MFCC with DNN (84.5%), GTCC with HMM (87.54%), GTCC with DNN (89.04%) In addition, the proposed system retrieves the speech from the database with minimum execution time which is listed in the following table 2.

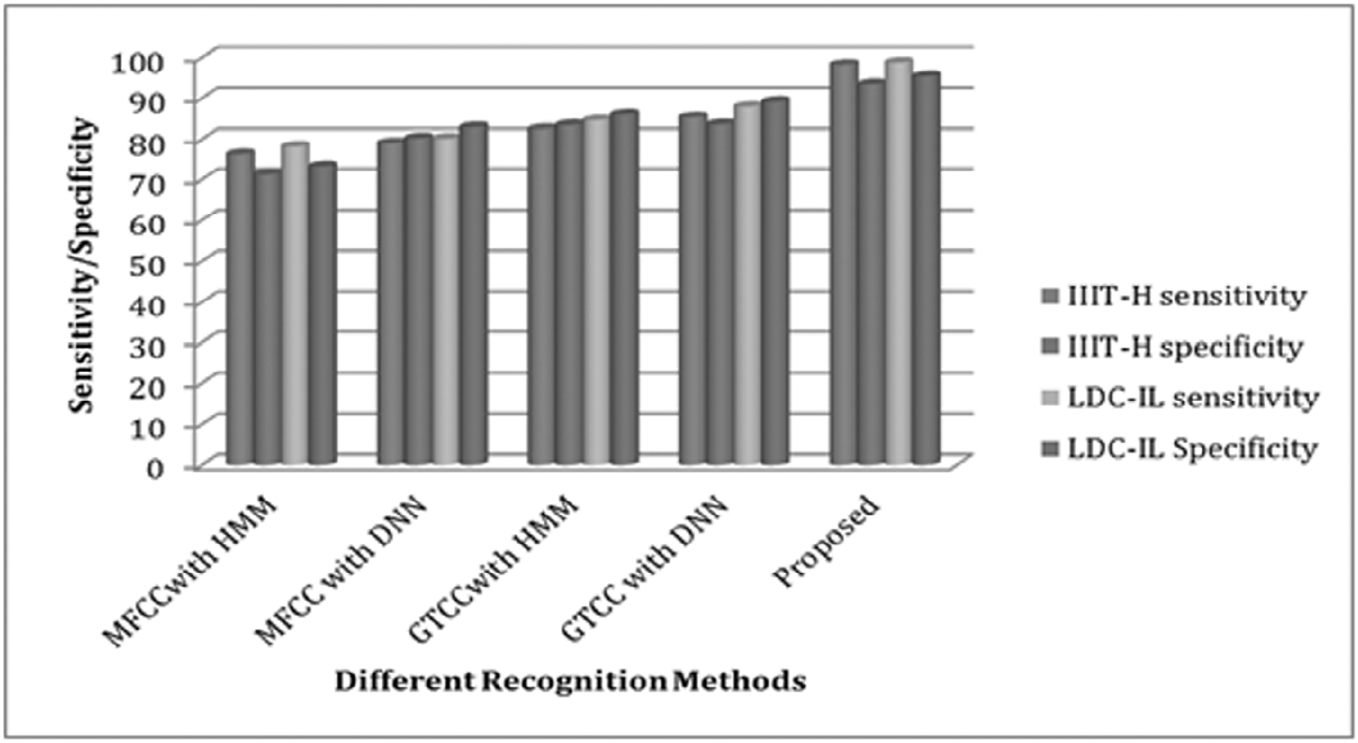

Figure 7: Sensitivity and Specificity

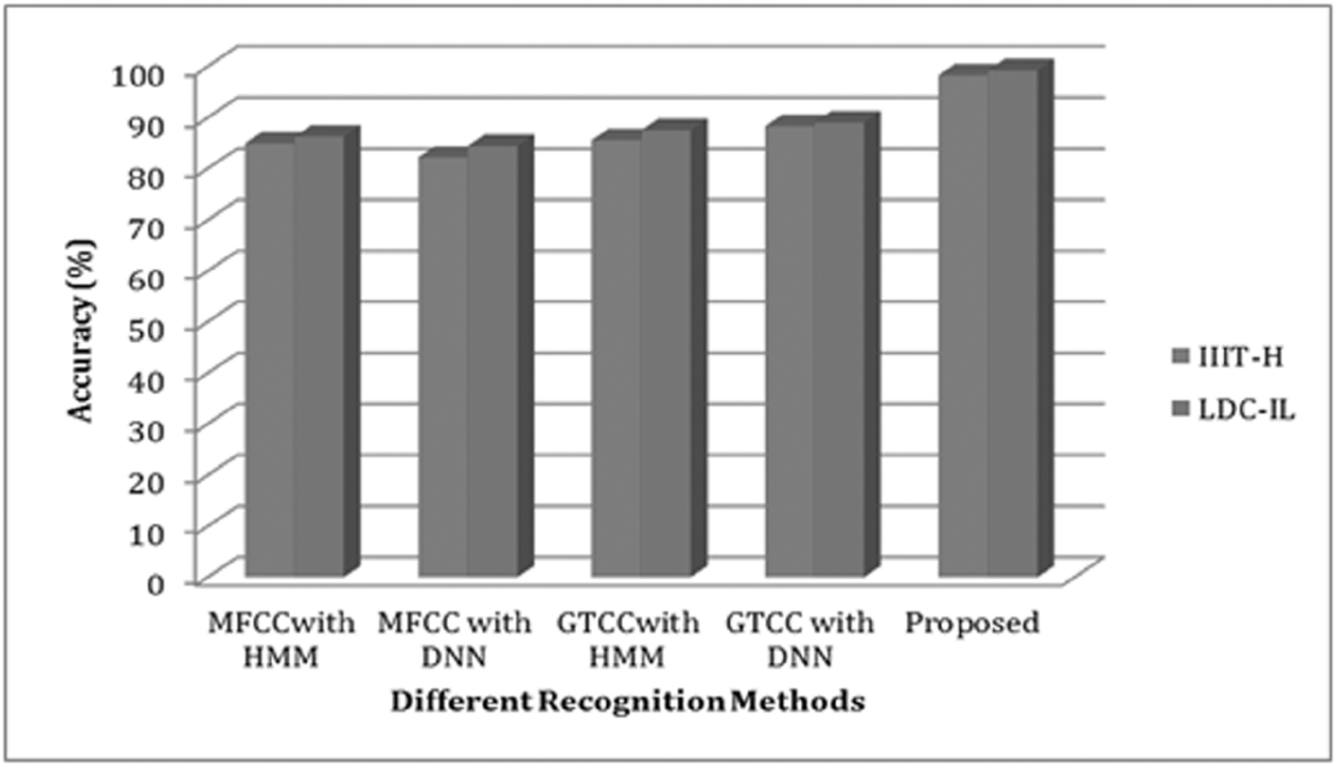

Figure 8: Accuracy

Table 2
Execution Time

S. No	*Methods*	*Execution Time (s)*	
		IIIT-H Indic speech corpus	*LDC-IL*
1	MFCC with HMM	26.869	32
2	MFCC with DNN	16.07	25
3	GTCC with HMM	14.10	17
4	GTCC with DNN	13.514	13
5	Proposed MODGDF-GWCC WITH DNN-HMM	9.64	6

From the above table 2, it clearly can be understood that the proposed method MODGDF-GWCC WITH DNN-HMM retrieves the speech-related features from the database with minimum execution time (IIIT-H-9.64s, LDC-IL-6s) when compared to the other methods such as MFCC with HMM(IIIT-H26.869s, LDC-IL32s), MFCC with DNN(IIIT-H-16.07s, LDC-IL-25s), GTCC with HMM(IIIT-H-14.10s, LDC-IL-17s) and GTCC with DNN(IIIT-H13.514s, LDC-IL-13s).

4.6. Overall Results and Discussions

The above discussions clearly show prove that the proposed syllable segmentation, feature extraction and recognition process achieve the efficient results of both datasets are shown in the following table 3.

The above Table 3, clearly shows that the proposed system achieves the efficient results while matching the speech-related features with both the IIIT-H Indic speech corpus and LDC-IL database with minimum execution time. Then the graphical representation of the overall efficiency is shown in the following figure 9.

Table 3
Overall Performance of the Proposed System

S. No	*Methods*	*IIIT-H Indic speech corpus Database*					*LDC-IL Database*				
		P	*R*	*AoF*	*Dec*	*RA*	*P*	*R*	*Aof*	*Dect*	*RA*
1	MFCC - HMM	68	72	77.36	57	85	71	73	78.43	63	86.3
2	MFCC- DNN	75	78	83.14	66	82.2	77	79.2	85.31	68	84.5
3	GTCC-HMM	83	84	85.47	76	85.6	85.4	86.3	87.67	79	87.54
4	GTCC- DNN	86	87	88.67	82	88.3	88	89.3	90.21	83	89.04
5	Proposed	93	94	98.64	99	98.3	95.2	97.6	99.23	99	99.23

P-Precision, R-Recall, AoF- Accuracy of extracted Features, Dec-Detectability,
RA- Recognition Accuracy

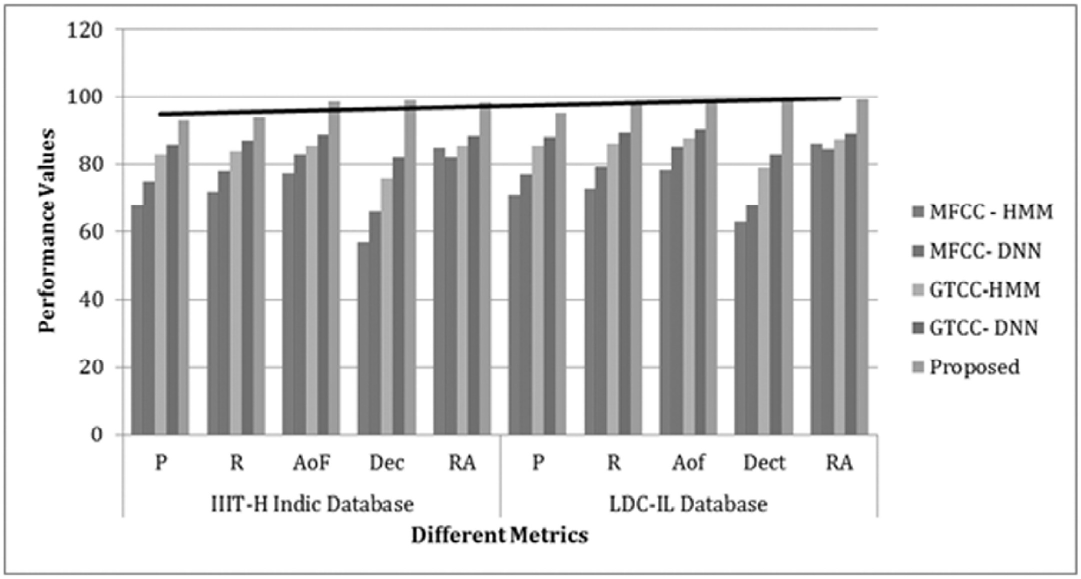

Figure 9: Overall Efficiency

From the above discussions, it is clearly known that the proposed system efficiently recognizes the Tamil speech with minimum time. Thus the proposed modified global delay function with Gammatone wavelet coefficient based features are efficiently trained and matched with language model which improves the overall recognition accuracy.

5. CONCLUSION

Thus the paper developing the Tamil speech recognition system by utilizing the three different models such as syllable segmentation, acoustic model creation and language model on the IIIT-H Indic speech corpus speech and LDC-IL database. The implemented system performance is analyzed using various performance metrics such as PSNR, MAE, entropy, precision, recall, the accuracy of extracted features, detectability of the features, recognition accuracy, sensitivity, specificity and execution time. From the analysis process, the proposed system achieves the highest recognition rate when compared to the other existing methods. Thus the performance of the system is analyzed using the experimental results and the proposed system recognizes the character with minimum complexity, time and high accuracy.

REFERENCE

[1] V. Kamakshi Prasad., T. Nagarajan., Hema A. Murthy., “Automatic segmentation of continuous speech using minimum phase group delay functions”, Speech Communications, Vol. 42, pp. 429-446, 2004.

[2] Bharathi et al., “A Neural Network based speech based recognition system for isolated Tamil words” SSN college of engineering.

[3] Mark Gales and Steve Young, “The Application of Hidden Markov Models in Speech Recognition”, Foundations and TrendsinSignal Processing, Vol. 1, No. 3, 2007.

[4] Dutta, K; Sarma, K.K., Dept. of ECE., “Multiple feature extraction for RNN-based Assamese speech recognition for speech to text conversion application” IEEE, Page(s):600–603, Date: 28-29 Dec. 2012.

[5] Kishore Prahallad, E. Naresh Kumar, Venkatesh Keri, S. Rajendran, Alan W. Black, “The IIIT-H Indic speech corpus Speech Databases”.

[6] http://www.ldcil.org/.

[7] Sankar Babu, Dr. M. Anto Bennet, R. Kaushik Krishna, S. Jaya Prakash, B.S. Jayavignesh, “Performance & Analysis of Encrypted Approach on Speech to Text Converter”, International Journal of Emerging Technologies in Computational and Applied Sciences.

[8] Iswarya, Radha, “Comparative analysis of feature extraction techniques for Tamil speech recognition”, available at., http://searchdl.org/public/book_series/elsevierst/1/117.pdf.

[9] Hema A. Murthy, T. Nagarajan and N. Hemalatha, “Automatic Segmentation and Labeling of Continuous Speech without Bootstrapping”, available at., http://www.iitm.ac.in/donlab/website_files/publications/2004/eusipco.pdf.

[10] Alex Graves, Abdel-rahman Mohamed and Geoffrey Hinton, “Speech Recognition With Deep Recurrent Neural Networks”, http://www.cs.toronto.edu/~fritz/absps/RNN13.pdf.

[11] Mark Gales and Steve Young, “The Application of Hidden Markov Models in Speech Recognition”, Foundations and Trendsin Signal Processing, Vol. 1, No. 3, 2007.

[12] Prasad, V. K., Nagarajan, T., and Murthy, H. A., “Automatic segmentation of continuous speech using minimum phase group delay functions,in Speech Communication, vol. 42, Apr. 2004, pp. 1883–1886.

[13] Lakshmi, A, “A Syllable-based Continuous Speech Recognizer for IndianLanguages,” MS Thesis, Indian Institute of Technology, Department of Computer Science and Engg., Madras, India, August 2007.

[14] http://www.ldcil.org/areasOfWorkSpeech.aspx.

[15] Dalmiya C, Dr. Dharun V, Rajesh K, “An EfficientMethod for Tamil Speech Recognition using MFCCand DTW”, IEEE Conference on Information andCommunication Technologies (ICT), pp 1263-1268,2013.

[16] V. Kamakshi Prasad, T. Nagarajan and Hema A. Murthy, “Continuous speech recognition using automatically segmented data at syllabic units,” in Proceedings of the Sixth International Conference on Signal Processing, ICSP, pp. 235-238, August 2002.

[17] Q. Li and Y. Huang, “An auditory-based feature extraction algorithm for robust speaker identification under mismatched conditions,” IEEETransactions on Audio, Speech, and Language Processing, vol. 19, no. 6, pp. 1791–1801, 2011.

[18] R. Schluter, L. Bezrukov, H. Wagner, and H. Ney, “Gammatonefeaturesand feature combination for large vocabulary speech recognition,” in IEEEInternational Conference on Acoustics, Speech and Signal Processing, 2007.ICASSP 2007., vol. 4, April 2007, pp. IV–649 –IV–652.