An analysis of the performance evaluation of syllable based tamil speech recognition system

Article *in* ARPN Journal of Engineering and Applied Sciences · April 2017

2 authors, including:

Akila Ganesh
Vels Institute of science, Technology & Advanced Studies

VOL. 12, NO. 8, APRIL 2017 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

AN ANALYSIS OF THE PERFORMANCE EVALUATION OF SYLLABLE BASED TAMIL SPEECH RECOGNITION SYSTEM

A. Akila
Department of Computer Science, VELS University, Chennai, Tamil Nadu, India
E-Mail: akila.scs@velsuniv.ac.in

ABSTRACT

Automatic Speech Recognition has been a goal of research for many decades. Many research works have been developed successfully for automatic speech recognition (ASR) of English language. ASR for European languages has not reached their height as ASR in English language. In this work, an implementation of Tamil based automatic speech Recognition System is developed. The ASR has many phases to perform the recognition process. A novel Tamil speech recognition system has been proposed in this work which reduces the complexity and the vocabulary size of the recognition model by applying segmentation at different phases. The temporal features like short term energy, zero crossing rate and the feature vectors based techniques like Mel frequency Cepstral coefficient, linear predictive coding are used for the segmentation. The sound attributes such as Sound Intensity Level, Time Duration and Root Mean Square are used to enhance the effectiveness of the Tamil speech recognition system.

Keywords: temporal feature based segmentation, varied length maximum likelihood, speech outlay, normalization of speech, accuracy, computational time.

1. INTRODUCTION

The grapheme of Indian languages has originated from the ancient Brahmi script. The basic speech sound units and basic written form have one to one correspondence. An important characteristic of Indian language grapheme is their phonetic nature. The characters were the orthographic depiction of speech sounds. A character in Indian language scripts has been close to syllable and can be usually of the following form: C, V, CV, CCV and CVC, where C is a consonant and V is a vowel [1] [2]. The rules necessary to map the letters to sounds of Indian languages are almost in a straight line ahead. All Indian language scripts have common phonetic base [3]. Most of the Indian languages are syllable centric. It is intended to make a study of emergent of a speech recognition system at Syllable level.

Related work

The mode of communication like gesture, eye contact, sign language, written communication and vocal communication were being used by human beings for sharing thoughts and information to other people. The vocal communication has being one of the comfortable modes of communication [4], speech has been used for computer human interface. The automatic speech recognition has gained continuous attention in research field for more than five decades [5]. The Tamil speech recognition has its own challenge, because of the large character set, high grammatical rules [6], and varied accents [7]. Many research works have been carried out in Tamil speech recognition in the fields like varied recognition units [8][9][10], segmentation of the speech into subword units [11][12][13][14], designing the language model [15] and designing the decoder etc[16]. The Tamil speech recognition based research needs more enrichment as the English language.

Tamil speech recognition system

Tamil is one of the ancient languages which have been in practice for many centuries. The automated speech recognition system has been developed for Tamil language for many years. A glimpse of the Tamil based Speech Recognition is given below with the latest decade works, their main technology designed and the contributor of that work.

A segmentation procedure using a minimum phase group delay based approach was designed by T. Natarajan and his team to segment the spontaneous speech into syllable units [17]. The minimum phase signal was derived from the short term energy function, and then the group delay function finds the peaks of the spectrum.

An isolated word recognition system based on Neural Networks combined with the dynamic programming algorithm was designed by S. Saraswathi and T. V. Geetha [18] to increase the recognition rates.

A group delay based two level segmentation and rule based text segmentation approach was used in the Speech Recognizer which has the subword unit as syllable was designed by Lakshmi A and Hema A Murthy [12] to segment speech data and text. In the first level, the speech data was segmented into polysyllabic units using coarse grain window and at the second level, the segmentation was performed using fine grain window.

A spoken Tamil Character Recognition System was built by M. Chadrasekar and M. Ponnavaikko [16] which used Back Propagation Network (BPN) and the acoustic features of individual characters of Tamil language to perform the recognition process.

The word was decomposed into morpheme units by S. Saraswathi and T. V. Geetha using the language model [19]. The trigram language model and the katz backoff smoothing effect were used in the system to improve the performance of the system.

In Syllable based Continuous Speech Recognition for Tamil [13], R. Thangarajan and A. M.

Natarajan have given an algorithm which leverages linguistic rules of Tamil to identify poetic syllable in a Tamil word. The syllable in Tamil is built with combinations of short vowel (Kuril), long vowel (Nedil) and Consonant (Mei eluthu). The algorithm segments the given utterance into syllables using the syllabic rules of Tamil Language which consist of eight patterns. The algorithm showed a result of 80% accuracy.

P. G. Deivapalan and A. Hema Murthy [20] designed a syllable based isolated word recognizer which uses a group delay algorithm to segment the word into syllables. During training, syllable models were used and in testing the speech signal was segmented into syllable like units. An edit distance based approach was used to identify the word.

In automatic Speech Recognition, the phone has been the preferred subword unit [11]. Syllable a larger unit to phone has the problem of contextual variations. A group delay based two level segmentation was used to segment the speech into syllable units [21]. The Ngram language model was used as a post processing step to perform the syllable search. A Syllable modeling in Continuous Speech Recognition for Tamil language [22] was built by R. Thangarajan, A. M. Natarajan and M. Selvam, a syllable based model with prosodic syllable was proposed.

The language model was used at the segmentation phase, recognition phase and syllable, word level error correction phase. The ambiguity in the recognized phonemes were reduced by using inter and intra word based language models [15]. A strategy for recognizing oral words in Tamil language has been developed by A. N. Sigappi and S. Palanivel [23]. The strategy has been evolved using HMM and Auto Associative Neural Networks (AANN) models. The strategy was constructed using the Mel Frequency Cepstral Coefficients features obtained from the speech signal. In SRS built by Durai Murugan and his team [10] with medium size vocabulary, the system was trained with 21 speakers and the recognition accuracy was 84%. DTW based Speech Recognition System has a good accuracy for isolated Tamil Digits [8] [9]. The performance of the DTW is better for smaller corpus and is speaker dependent [24].

A phoneme based Context Dependent Acoustic model for Medium Vocabulary Continuous Speech Recognition in Tamil [25] was built by R. Thangarajan and his team. The vocabulary consists of 1700 unique words and a pronunciation dictionary with 44 phones was designed in this work. Identifying the language of the spoken utterance was done using the Discrete Hidden Markov Model (DHMM) by M. Sandhanam and his team [26]. The approach was designed to identify Indian languages like Telugu, Tamil, Hindi, Marathi, Malayalam and Kannada. A new chat system 'Tamil Vayadi' [27] which gets the voice input and converts it to text and transfer to the receiver in a serialized form by R. Sridhar and his team. The phoneme based unit was used as the recognition unit. A bigram language model was used for converting the speech to text.

The newspaper reading system was designed by J. Stephen and his team using the speech recognition, speech synthesis and web [28] to help the visually challenged and illiterates. The system has been designed to support the languages like Malayalam, Tamil, Hindi and English.

Need for syllable based Tamil speech recognition system

Syllable is a larger unit than a phone since it encompasses two or more phonemes [29]. The onset, the nucleus and the coda or rhyme makes a syllable [30]. The nucleus or the central part is not context dependent. The onset and coda may be dependent of the context. In Tamil language, a word is made of one or more syllables called Asai [7]. So when a syllable is used as a recognition unit for large vocabulary, it may lead to good recognition system with improved accuracy.

There are only few cases where a syllable may consist of a single phoneme only. Hence the problem of severe contextual effects present in phones is relatively reduced in syllables [31]. The pronunciation variation is more systematic in the level of syllable. The onset and nucleus of a syllable are not much contextual dependent.

Tamil Language is composed of greater grammatical rules in nature. The words in Tamil language can have same root with varied prefixes and suffixes. So models could not be effectively trained using a small set of training data. The problems in Tamil ASR due to inflectional morphology and OOV words could be solved using a morpheme based language models. Even though morpheme based language model performs better than trigram, bigram models, the accuracy of the morpheme model is the major issue. So a syllable based model is chosen in the proposed work.

2. PROPOSED MODEL – A NOVEL ADAPTIVE SPECTRAL FRAGMENTATION APPROACH FOR TAMIL SPEECH RECOGNITION

The main methodologies proposed in this work are Word Category Identification, Word Segmentation, Character Segmentation and Time Normalization whose detailed explanation of each methodology has been given in the successive sections.

The proposed model of Tamil Speech Recognition uses syllable as a recognition unit. The model has three major new methodologies proposed for the segmentation and normalization of speech respectively. The input utterance has to be preemphasised and then the endpoint detection using Mahalanobis Endpoint Detection algorithm. The algorithm uses the mean and variance to compute a mahalanobis distance [32]. The speech signal could be a connected speech which was divided into separate words using temporal characteristics. The isolated word was still fragmented into characters using the proposed algorithm Varied Length Maximum Likelihood [33][34][35]. The segmented characters were classified using the Gaussian Mixture Model and the syllables were formed using the prosodic syllabic rules[36][37][38][39].

The syllables were trained using the HTK tool. The same procedure was followed for the testing phase also.

3. SEGMENTATION OF CONNECTED WORD USING DYNAMIC THRESHOLDING SEGMENTATION (DTS) AND TEMPORAL FEATURE BASED SEGMENTATION(TFS)

Segmentation using dynamic thresholding

The connected word could be segmented using the techniques like K-mean clustering, fuzzy C-means clustering and Ostu thresholding method. The Ostu method has been implemented in image segmentation. The approaches like blocking black area and boundary detection were used to perceive the boundaries of the connected speech. The Dynamic thresholding segmentation (DTS) method was implemented for continuous speech [40][41][42]. The continuous speech was segmented into words as words have well defined acoustic representation. The technique was incorporated for connected speech in this work.

Temporal feature based segmentation

The speech input was segmented using the Short Term Energy (STE). The speech wave signal was split into many frames and for each frame the short term energy was calculated using equation 1 [40].

$$E_{sqr} = \sum_{n=1}^{N} s(n)^2 \tag{1}$$

where N is the length of the sampled signal, s(n) is the amplitude of the speech signal.

The Frame with Short Term Energy less than a threshold value was found and that frame was identified as the edge of each word in the connected word speech. The voice activity detection was done to identify the end points. The signal was divided into n frames of size m. The frames whose energy is down the threshold value were identified. The consecutive frames which have the short term energy less than threshold value were discarded and only the first frame of the consecutive frame was considered as the boundary. This process was used to split the connected speech into words.

4. CHARACTER BOUNDARY IDENTIFICATION USING MAXIMUM LIKELIHOOD (ML) AND VARIED LENGTH MAXIMUM LIKELIHOOD (VLML) ALGORITHM

Maximum likelihood (ML) algorithm

The maximum likelihood segmentation is a feature vector based segmentation approach. The uniqueness in spectral properties within a region is identified with the Maximum Likelihood algorithm. The feature extraction for each frame is calculated where each frame has the size as 25ms. The segment is made up of one or more frames [14]. To segment the frames, spectral distortion is used which is the deviation in the spectral properties. The spectral distortion within a segment should be small. The deviation is manipulated using intra segment distortion (δ) and generalized centroid (μ) [43]. The intra segment distortion (δ) was given by equation 2.

$$\delta(i,j) = \sum_{n=i}^{j} d(x_n, \mu) \tag{2}$$

where i and j are the first and last frames of the segment, d is the positive distortion measure. The d is the Euclidian distance between the n^{th} frame (x_n) and the centroid of each segment (μ). The equation 3 specifies the formula to calculate the centroid of each frame and μ′ represents the generalized centroid of all segments.

$$\mu = \arg\min_{\mu'} \frac{1}{j-i} \sum_{n=i+1}^{j} d(x_n, \mu') \tag{3}$$

Varied length maximum likelihood (VLML)

The maximum likelihood algorithm has the limitation that it imagines the size of each segment as same which is not practically feasible. The length of short vowel will be lower when compared to long vowel. So the necessity of an algorithm which considers each segment length to be different was more. The proposed method uses the formants to find the length of each segment to find the character boundary.

To overcome the drawback of the maximum likelihood algorithm, the proposed methodology uses a varied length segment. To identify the length of each segment, the formant values were used. Based on the first three formant values, the length of the each segment was fixed and the accurate length was identified using the baseline Maximum Likelihood algorithm.

5. IDENTIFICATION OF SPEECH OUTLAY AND NORMALIZATION OF SPEECH

Identification of speech outlay

The speech rate of the input signal was computed using sound attributes like sound intensity level, root mean square and the time duration. The range of these attributes was fixed by manipulating the attribute values for the training speech signals with various speech rates. The signals within the range of these attributes were found to be normal speech. The signals out of these ranges were treated as slow and fast speech signals. The range is given in the Table-1. The time duration of a slow signal would be obviously more because the duration of each phoneme would be more in the slow speech signal.

ARPN Journal of Engineering and Applied Sciences

Table-1. Limits of the speech outlay for a normal speech signal.

Speech Outlay	Lower Bound	Upper Bound
Sound Intensity Level	116 dB	121 dB
Time Duration	0.9 ms	2.0 ms
Root Mean Square	0.04	0.07

Normalization of speech

To make the speech a normal one, the time normalization is carried out which uses the scaling factors to perform time scale modification [34]. The rate of speech is used to identify the variation in speech rate. The speech signal was classified as fast or slow or normal speech. The fast speech signal was stretched and the slow speech signal is compressed by applying the scaling factor [44]. No change was made for a normal speech signal. The process of altering the speech signal either by expansion or compression using the scaling factor is called as Time Scale Modification. If ROS computed with the equation 4 is larger than the normal range, it is treated as fast signal and if smaller than the normal range, then the speech signal as slow. The threshold is found by manipulating the rate of speech for a normal signal which is characterized by using the PRAAT tool.

$$ROS = \frac{\sum_{i=1}^{n} MD_i}{n} \quad (4)$$

6. RESULT AND DISCUSSION

The evaluation of the proposed models was manipulated using the computational time and the accuracy.

Accuracy

The accuracy could be measured using the measures like Word Error Rate (WER), Correctness Percentage, False Detection Percentage, Match Error Rate (MER) etc. The accuracy of segmentation algorithms could be measured using the Missing Detection Percentage (MDP) and False Detection Percentage (FDP) [45]. The MDP and the deviation from the manual method were the measures used in this work to measure the accuracy. The accuracy of the procedure was manipulated by finding the boundary frame number of each segment using the proposed model and the segmentation boundary found manually. The deviation was calculated by the difference of the manual method and the proposed segmentation procedure. The Missing Detection Percentage is measured by finding the percentage of the number of frames missed from boundary detection divided by the actual number of frames which is given in equation 5 [45]. The false detection percentage is the percentage of number of segments detected incorrect by actual number of segments. The FDP was computed using equation 6.

$$MDP = \frac{Missed\ Detection}{Actual\ points} X\ 100 \quad (5)$$

$$FDP = \frac{False\ Detections}{Total\ Number\ of\ Segments} \quad (6)$$

Accuracy of DTS and TFS

The edge between each word in the connected speech was found using the existing method dynamic thresholding segmentation and the proposed method temporal feature based segmentation. The performance accuracy was computed by identifying the deviation from the boundary position found manually from the connected speech signals with the help of audacity tool which is shown in Table-2. It could be noticed from the Figure.2 that the deviation was less in TFS compared to the DTS algorithm.

Table-2. Edge points of each word in the connected speech manipulated using DTS and TFS.

Signals		CT1	CT2	CT3
Manual	B1 (in frames)	28	36	38
	B2 (in frames)	70	70	85
	B3 (in frames)	113	101	135
DTS	B1 (in frames)	20	34	29
	B2 (in frames)	78	66	68
	B3 (in frames)	116	94	112
TFS	B1 (in frames)	21	30	34
	B2 (in frames)	69	64	79
	B3 (in frames)	104	93	122

The deviation was less by 0.69% in the proposed TFS algorithm. The missed detection percentage of the proposed TFS algorithm was lesser than the DTS algorithm as shown in Table-3. The MDP in temporal feature segmentation was 8.9% less than the existing dynamic thresholding segmentation algorithm.

Table-3. Missed detection percentage of the DTS and TFS algorithms.

Algorithm	Missed Detection Percentage
DTS	23.82%
TFS	14.94%

Accuracy of ML and VLML

The boundary points manipulated for sample words using the VLML and ML algorithms were compared with the boundary points identified manually using the Audacity. It could be noticed from the table that

VOL. 12, NO. 8, APRIL 2017 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

the segment length of each character was more or less same of 21 frames. When the number of samples was less than the segment size, the remaining samples were added to the last segment, so the size of the last segment was more than 21 frames. When the input word contains a long vowel, then the length of the long vowel was misdetected with a lesser length in Maximum Likelihood Algorithm. The Varied Length Maximum Likelihood algorithm fixes the length of the segment based on the category of the character and finds better boundary position compared to the Maximum Likelihood algorithm.

The deviation between the boundary point manipulated VLML and manual method is computed and similarly between the ML and manual is also computed. It is observed that the deviation is less in VLML algorithm when compared with ML algorithm. The Table-4 shows the boundary deviation of the ML and VLML algorithms. It also specifies that the boundary of a phoneme could be identified by ML algorithm but it could not be used for other subword units like character, syllable etc. The VLML algorithm uses the formant values to find the category of the character. It could be used to categories other subword units also by identifying the individual characters and then concatenating the characters to form syllables or morphemes etc.

Table-4. Performance comparison of ML and VLML algorithm.

Properties	Maximum Likelihood	Varied Length Maximum Likelihood
Boundary Deviation Percentage	2.353571	2.010714
Segmentation	All segments are equally segmented	Segmented depending on Character Category
Character Categorization	Not Possible	Does categorization using Formants

Accuracy of Tamil based speech recognition model with and without speech rate normalization using sound attributes (SRNSA)

The efficiency of the Tamil Speech Recognition Model with and without Speech Rate Normalization using Sound Attributes (SRNSA) was compared by accuracy with a dataset from the datasets. The Sound Intensity Level, Time Duration and Root Mean Square are computed with the entire dataset and the range of Sound Intensity Level, Time Duration and Root Mean Square for the normal speech signal is set. It was experienced that the Sound Intensity Level, the Time Duration and the Root Mean Square of a slow speech signal exceeds the normal signal range. So it was clear that when values of the three features for a given input signal exceeds the upper bound of Table-1, the signal has been identified as a slow speech signal. When the values of the three features for a given input signal were less the lower bound of the Table-1, the signal has been identified as a fast speech signal.

The performance of the Tamil Speech Recognition Model with SRNSA and without SRNSA in identifying the count of the character segments is compared. The proposed model after doing the time normalization, computes the counting of character segments correctly for all the speech rate category signals as shown in Table 5. The model without SRNSA computes correctly the count for the normal speech signal but does not identify for the slow and fast speech signal which is also shown in Table-5 when performed for the speech corpus. Proposed model without SRNSA gives an accuracy of 70% and the proposed model with SRNSA gives an accuracy of 74%.

Table-5. Performance comparison of the proposed model with and without SRNSA.

Speech Signal	Segments Count Computed		
	Actual Segment Count	Proposed Model with SRNSA	Proposed Model without SRNSA
S1	3	3	5
S2	3	3	3
S3	3	3	1

The accuracy was compared with the proposed model with SRNSA and proposed model without SRNSA. The accuracy was more with the model with SRNSA because the data corpus used for training and testing the system consists of slow, fast and normal speech rate signals. When the speech signal without SRNSA was used to recognize the speech, the variation in speech rate affected the performance. The variation was shown in the character fragmentation algorithm. The character with slow speech rate was having duration more than the normal character length. So when the character fragmentation was done, the misdetection of boundary was more. This leads to the decrease in accuracy percentage of the proposed model without SRNSA. As the time normalization were done with slow and fast speech signals in the model with SRNSA, the accuracy was better as the boundary of the characters were identified with less misdetection.

The false detection percentage is the percentage of number of segments detected incorrect by actual number of segments. The FDP was computed using equation 6. The false detection was more in the proposed model without SRNSA as the duration of the slow and fast speech signals were more without normalization. This leads to more number of misdetections. The false detection percentage of the proposed model with and without SRNSA was shown Table-6.

Table-6. False detection percentage of the proposed model with and without SRNSA.

Algorithm	False Detection Percentage
Proposed Model with SRNSA	8.3
Proposed Model without SRNSA	45.8

Computation time of the proposed algorithms

Computation time which is also called running time is the length of time necessary to perform a computational process. The computation time could be measured using CPU Time and Elapsed clock time.

Elapsed Clock Time is the moment in time taken from start of computer program to the end. Elapsed time includes I/O time and all other types of wait. Elapsed time is the time measured by an ordinary clock [46]. The CPU time and elapsed time was manipulated using equation 7 and equation 8.

$$\text{CPU Time} = \text{PST} - \text{PET} \tag{7}$$

Where PST is the process start time and PET is the process end time.

$$\text{Elapsed } Time = CST - CET \tag{8}$$

Where CST is the clock start time and CET is the clock end time.

Computational performance of the DTS and TFS

The CPU Time and the Elapsed Time of the some sample connected speech signal used in the proposed work were manipulated using equation 7 and equation 8. It could be noticed from the Table-7 that the CPU Time is always smaller than Elapsed Time as the Elapsed Time includes the waiting Time for I/O Operations. The CPU and the elapsed time were manipulated for the two segmentation algorithms for connected speech with Matlab as the segmentation algorithms were implemented with Matlab.

Table-7. Computation time of DTS and TFS algorithms.

Computation Time	DTS (in sec)	TFS (in sec)
Elapsed Time	4.4453	1.5665
CPU Time	4.7011	0.3828

It could be observed from the Table 7 that the Elapsed Time of the Dynamic Threshold Segmentation (DTS) algorithm is larger when compared to the proposed Temporal Feature based Segmentation (TFS) algorithm. Similarly the CPU time of the DTS algorithm is larger when compared to TFS algorithm which is shown in Table-7.

Computational performance of the ML and VLML

The CPU Time and the elapsed time was more in the VLML algorithm as shown in Table-8 because VLML performs the manipulation for identifying the length of each segment which was not performed in ML algorithm.

Table-8. Computational time of ML and VLML algorithms.

Algorithm	CPU Time(in Sec)	Elapsed Time (in Sec)
ML	0.453	0.344
VLML	0.593	0.515

The ML algorithm performs the boundary adjustment using the Intra Segment Distortion. The proposed methodology VLML considers the length of each segment to be different. The segment length has been computed using the formant values and then uses the baseline ML algorithm to adjust the boundary. So the computational time of the proposed methodology was more than the baseline ML algorithm.

Computational performance of the Tamil based speech recognition model with and without SRNSA

The computational time of the proposed model with and without SRNSA was computed using the measures CPU time and the elapsed time. The CPU time for the proposed model with SRNSA was less than the proposed model without SRNSA. As the elapsed time includes the I/O time, waiting time etc, it was more for the proposed model with SRNSA when compared to proposed model without SRNSA which is shown in Table-9.

Table-9. Computational time of the proposed model with and without SRNSA.

Algorithm/Computational Time Parameters	Proposed Model with SRNSA	Proposed Model without SRNSA
CPU Time (in sec)	134.44	145.60
Elapsed Time (in sec)	235.52	195.90

7. CONCLUSIONS

The proposed method TFS has used the short term energy to perform segmentation. The performance of the proposed methodology and the dynamic thresholding segmentation was measured using the accuracy and computation Time. The accuracy of boundary identification was increased in the proposed methodology by using short term energy which is a temporal feature. The accuracy was measured in terms of the frame number. The computational time of the proposed TFS algorithm is lesser than the DTS algorithm.

The proposed methodology Varied Length Maximum Likelihood has lesser deviation in identifying the boundary of each character when compared to the

ARPN Journal of Engineering and Applied Sciences

baseline maximum likelihood algorithm. The segmentation was done dynamically depending on the character category in VLML which leads way to use the same segmentation algorithm for different subword units. The computation time was more in the VLML than the ML algorithm because in VLML the length of each segment was computed prior to the boundary identification using the maximum likelihood algorithm.

The performance of the proposed model Adaptive Spectral Fragmentation Approach for Tamil Speech Recognition with the Speech Rate Normalization using Sound Attributes and without the Speech Rate Normalization using Sound Attributes was compared using the measures Accuracy and the False Detection Percentage. The proposed model with SRNSA has shown better performance that model without SRNSA. The computation time of the model with SRNSA was more when compared to the model without SRNSA because it has more operations such as identification of speech rate and normalization of speech rate.

REFERENCES

[1] Biljana Prica and Sinisa Ilic, Recognition of Vowels in Continuous Speech by using Formants, Facta University-series: Electronics and Energetics, Vol.23, No. 3, 2010, pp. 379-393.

[2] Kumara. K.H, Dias. N.G.J and Sirisena. H, Automatic Segmentation of given set of Sinhala Text into Syllables for Speech Synthesis, Journal of Science of the University of Kelaniya 3, 2007, pp. 53-62.

[3] Kalyani.N and Sunitha. K.V.N, Syllable Analysis to Build a Dictation System in Telugu Language, IJCSIS,Vol.6, No.3, 2009, pp.171-176.

[4] Chandra. E and A. Akila, An Overview of Speech Recognition and Speech Synthesis Algorithms, International Journal of Computer Technology & Applications, Vol.3, No.4, 2012, pp 1426-1430.

[5] Akila. A and Chandra. E, Isolated Tamil Word Speech Recognition System using HTK, International Journal of Computer Science Research and Application, Vol.3, No.2, 2013, pp 30-38.

[6] Henry Charles. A.P and Devaraj. G, Alaigal-A Tamil Speech Recognition, Tamil Internet, 2004, pp.125-131.

[7] South Asia Language Resource Center, www.southasia.sas.upenn.edu/ Tamil/index.html

[8] Dharun. V. S, Karnan. M, Voice and Speech Recognition for Tamil Words and Numerals, IJMER, Vol. 2, No. 5. Sep-Oct 2012, pp.3406-3414.

[9] Karpagavalli. S, Usha Rani. K, Deepika. R and Kokila. P, Isolated Tamil Digits Speech Recognition using Vector Quantization, IJERT, Vol.1, No.4, June 2012, pp.1-9.

[10] Durai Murugan. L, Rajasekhar. S and Eliayaraja. E, Automated Transcription System for Tamil Language, IJCE, Vol.1, No.3, March 2012, pp 7-14.

[11] Thagarajan. R, Natarajan. A. M and Selvam. M , Word and Triphone based Approaches in Continuous speech Recognition for Tamil Language, WSEAS Transactions on Signal Processing, Vol.4, No.3, March 2008, pp.76-85.

[12] Lakshmi. A and Hema. A. Murthy, A Syllable Based Continuous Speech Recognition for Tamil, Interspeech 2006-ICSLP, September 17-21, Pittsburgh, Pennsylvania, 2006, pp.1878-1881.

[13] Thangarajan. R and Natarajan. A.M, Syllable based Continuous Speech Recognition for Tamil, South Asian Language Review, Vol. XVIII, No.1, January 2008, pp.71-85.

[14] Anindya Sarkar and Sreenivas. T.V, Automatic Speech Segmentation using Average Level Crossing Rate Information, Proc. of ICASSP - March 2005, Philadelphia, pp-I-397-I-400.

[15] Saraswathi. S and Geetha. T.V, Design of Language Models at Various Phase of Tamil Speech Recognition System, International Journal of Engineering Science and Technology, 2010, Vol.2, No.5, pp:244-257

[16] Chadrasekar. M and Ponnavaikko. M, Spoken Tamil Character Recognition, Electronic Journal of Technical Acoustics, 2007, Vol.22, pp.1-11.

[17] Nagarajan. T, Hema. A. Murthy and Rajesh. M .Hegde, Segmentation of Speech into Syllable-like Units, Energy 1, No. 1, 2003, pp.2893-2896.

[18] Saraswathi. S and Geetha. T.V, Implementation of Tamil Speech Recognition System using Neural Networks, Springer Berlin Heidelberg in Applied Computing, 2004, pp.169-176.

[19] Selvarajan Saraswathi and Geetha. T.V, Morpheme based Language Model for Tamil Speech Recognition System, The International Arab Journal of Information Technology 4, No.3, 2007, pp. 214-219.

[20] Deivapalan. P.G and Hema. A .Murthy, A Syllable-based IWR for Tamil Handling OOV words, National Conference on Communications, IIT Bombay, 2008, pp. 267-271.

[21] Lakshmi. A and Hema. A. Murthy, A New Approach to Continuous Speech Recognition in Indian Languages, Proceedings of NCC, 2008, pp.277-281.

[22] Thangarajan. R, Natarajan. A. M and Selvam. M, Syllable Modeling in Continuous Speech Recognition for Tamil Language, International Journal of Speech Technology, Vol.12, No. 2009, pp.47-57.

[23] Sigappi. A. N and Palanivel. S, Spoken Word Recognition Strategy for Tamil Language, International Journal of Computer Science Issues, Vol. 9, Issue 1, No 3, January 2012, pp.227-233.

[24] Anusuya. M. A and Katti. S. K, Classification Techniques used in Speech Recognition Applications: A Review, International Journal of Computer Technology Applications, Vol.2, No.4, 2011, pp. 910-954.

[25] Thangarajan. R, Natarajan. A. M and Selvam. M, Phoneme based Approach in Medium Vocabulary Continuous Speech Recognition in Tamil, International Journal of Computer processing of Oriental Languages, pp.1-17

[26] Sadanandam. M, Kamakshi Prasad. V, Janaki. V and Nagesh. A, Text Independent Language Recognition System using DHMM with New Features, IEEE 11th International Conference on Signal Processing, Vol.1, 2012, pp.511-514.

[27] Rajeswari Sridhar, Madhavan. K. V, Nagarajan. S and Nishanth. S, Incremental Language Model and Dynamic Decoder for Tamil Chat System, IEEE International Conference on Advances in Computing, Communications and Informatics, 2013, pp. 2057-2062.

[28] Jose Stephen, Anjali. M and Bhadran. V. K, Voice Enabled Multilingual Newspaper Reading System, IEEE Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS), 2013, pp.317-320.

[29] Rajesh Janakiraman, Chaitanya Kumar. J, Hema. A. Murthy , Robust Syllable Segmentation and its Application to Syllable-Centric Continuous Speech Recognition, IEEE National Conference on In Communications (NCC), 2010, pp.1-5.

[30] Todd Stephenso, Speech Recognition using Phonetically Featured Syllables, Master's Thesis, Centre for Cognitive Science, University of Edinburgh, 1998, pp.3.

[31] Aravind Ganapathiraju, Jonathan Hamaker, Joseph Picone, Mark Ordowski, and George. R. Doddington, Syllable-based Large Vocabulary Continuous Speech Recognition, IEEE Transactions on Speech and Audio Processing, Vol.9, no. 4 , 2001,pp.358-366.

[32] Matthias Wolfel and Hazim Kemal Ekenel, Feature Weighted Mahalanobis Distance: Improved Robustness for Gaussian Classifiers, 13th European signal processing conference, 2005, pp.1-4.

[33] Ravichandran. T and Durai Samy. K, Performance Evaluation and Comparison of Voice Activity Detection Algorithms, Medwell Journals, International Journal of Soft Computing, 2007, pp: 257-261.

[34] Jin Ah Kang and Seung Ho Choi, A Speaking Rate Normalization Method using Time Scale Modification for Speech Recognition, ICCA 2013, Vol.24, pp.95-97, 2013.

[35] Hema. A. Murthy and Yegnanarayana. B, Formant Extraction from Group Delay Function, Speech Communication 10, No.3, 1991, pp.209-221.

[36] Matthew Nicholas Stuttle, A Gaussian Mixture Model Spectral Representation for Speech Recognition, Hughes Hall and Cambridge University Engineering Department, July 2003, pp.30-73.

[37] Akila. A. Ganesh and Chandra Ravichandran, Syllable Based Continuous Speech Recognizer with Varied Length Maximum Likelihood Character Segmentation, IEEE conference Proceedings of ICACCI 2013, August 2013, pp. 935-940.

[38] Akila. A. Ganesh and Chandra Ravichandran, Grapheme Gaussian Model and Prosodic Syllable Based Tamil Speech Recognition System, IEEE Conference Proceedings of ICSC 2013, December 2013, pp. 424-429.

[39] Steve Young *et al*, The HTK Book. Version 3.4, Cambridge University, March 2006, pp.23-46.

[40] Rabiner L.R. and Juang B.H, Fundamentals of Speech Recognition. Prentice-Hall, 1993, pp.10-422.

[41] Md Mijanur Rahman and Md Al-Amin Bhuiyan, Dynamic Thresholding on Speech Segmentation, International Journal of Research in Engineering and Technology, Vol.2, Issue.9, September 2013, pp.404-411.

[42] Miael Nilsson and Marcus EJnarsson, Speech Recognition using Hidden Markov Model, Department of Telecommunications and speech Processing, Biekinge Institute of Technology, 2002, pp.17-39.

[43] Ramasubramanian. V and Sreenivas. T .V, Automatically Derived Units for Segment Vocoders, Proc. of ICASSP-2004, Vol. 1, pp.473-476.

[44] Benoit Sylvestre, Time-Scale Modification of Speech: A Time Frequency Approach, ME Thesis, McGill University, April 1991, pp.77-91.

[45] Behrouz Abdolali and Hossein Sameti, A Novel Method for Speech Segmentation Based on Speakers Characteristics, Signal & Image Processing: International Journal (SIPIJ),Vol.3, No.2, April 2012, pp.65-78.

[46] Computational Time Calculation, http://www.mathworks.in/matlabcentral/ answers/19069-computation-time-calculation.