TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 116 - 124

An improved Sub – Word Model based Continuous Robust Automatic Speech Recognition System for Tamil Language

[1]Sundara Pandiyan S [2]Julian Benadit P& [3]Michael Moses T

[1,2&3]Assistant Professor,Department of Computer Science and Engineering,School of Engineering and Technology,CHRIST (Deemed to be University),Bangalore,India.

Article Info
Volume 82
Page Number: 16585 - 16591
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 28 February 2020*

I. Abstract:

Automatic Speech Recognition (ASR), translating spoken words into text is the challenging task due to various language characteristics such as phoneme, syllable and nature of language chosen for this study, different properties of speech signals such as gender, frequency, environment noise, and different age groups.

Keywords: Share Automatic Speech Recognition, MODGDF, LDC-IL, Properties of Group Delay

I. INTRODUCTION

The objective of this study explores the syllable based continuous speech recognition for the Tamil language rather than the conventional speech recognition based on phoneme and word. Automatic Speech Recognition (ASR), translating spoken words into text is the challenging task due to various language characteristics such as phoneme, syllable and nature of language chosen for this study, differentproperties of speech signalssuch as gender, frequency, environment noise, and different age groups.

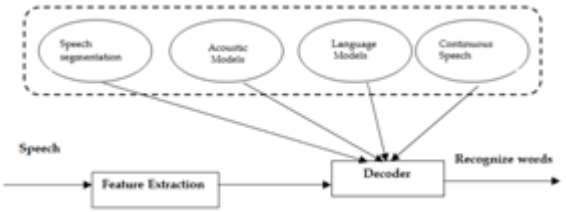


Figure 1 Automatic Speech Recognition System

ASR is defined as follows

$$\hat{w} = \underset{w}{\operatorname{argmax}} \{P(W|Y)\} \quad (1)$$

Applying Baye's rule in Equation (1)

$$\hat{w} = \underset{w}{\operatorname{argmax}} \{P(Y|W)P(W)\} \quad (2)$$

Where P(Y|W)is determined by an Acoustic Model and the P (W) is determined by aLanguage model.Language sub-word units such as phoneme and syllable areinfluenced in the accuracy of ASR. Syllable is the core unit of ASRsystem, and it gives better representation and duration stability about the phoneme (Nagarajanet al. 2003) Figure 1.Shows the various functional components of ASR system.

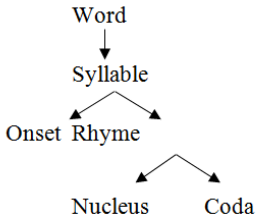


Figure 2 Parts of Syllable

Figure2 shows the parts of syllable, and the syllable is defined as the combination of vowels and consonants. Vowel is the nucleus of syllable and consonant in the syllable may be placed on either side of vowel and in general the format of any syllable is c^*vc^*. Where c is a consonant and v is a vowel. Syllable has many prosodic properties such

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 116 - 124

as pitch, accent and stress; allthese are linked with human speech production and perception. Tamil is a syllable centric language(Timothy et al. 2016) and it has rich linguistic rules to extract the syllable segments from a text. The ASR decoder recognizes continuous speech into words using Baye's rule on Hidden Markov Model (HMM).

II. LITERATURE SURVEY

The syllable was proposed as a unit of ASR as early as 1975(Matti Karjalainen 1987)due to its strong links with human speech production and perception. Syllable based ASR system needs segmentation of speech signal into syllabic units. A minimum phase Group Delay Function is applied to segment spontaneous speech into syllable-like units. The construction of the filterbank of the auditory system can be based on frequency response of the basilar membrane. Many models have been proposed to mimic the basilar membrane in the inner ear. The Gammatone functions provide a good fit to the auditory response (Aniruddha Adiga et al. 2013).

Joint features derived from Modified Group Delay Function (MODGDF) and Mel FrequencyCepstral Coefficients (MFCC) gives a 10.6 % improvement for Tamil Language ASR (Rajesh M.hegde et al. 2014).Feature Extraction for continuous ASR is a complex and important task. The Gaussian Mixture Model (GMM) based Hidden Markov Model (HMM) has been applied to generate the Acoustical Model of ASR. There are many efforts that have been taken to apply syllable that is the basic unit of ASR for Tamil Language (Lakshmi et al. 2006). This motivation comes from the fact that Tamil is syllabic Language. The Tamil Language has Low Resourcespeech corpus for training required to build ASR. However the speechsparsity and uneven problems in speech corpus are alleviated using Deep Neural Network (DNN) based ASR framework. The DNN based Acoustical Modeling method is the special Artificial Neural Network (ANN) and hybrid with HMM approach. Neural Network Language Model (NNLM)(Ankur Gandhe et al. 2006)is superior to standard *n-gram* technique that excepting the fact that it has high computational training complexity.

III. SPEECH CORPUS

The LDC-IL Speech database has both telephone speech and microphone speech.

3.1 IIIT-H Indic Speech Database

The IIIT-H Indic (kishore Prahallad et al.) and LDC-IL (Pukhraj p. Shrishrimal et al. 2012) Speech Databases are used in our ASR Experiments. IIIT-H Speech Database consists of seven Indian Languages and the details of IIIT-H Speech Database are shown in Table.1

Table 1 Statistics of the Optimal Speech and Text Selection.

Languages	No. of sentences	No. of words		No. of Syllables		Average Words per line	Duration (hh:mm)	Avg. Dur of each Utterance (sec)
		Total	Unique	Total	Unique			
Bengali	**1000**	**7877**	**2285**	**25757**	**866**	**7**	**1:39**	**5.94**
Hindi	**1000**	**8273**	**2145**	**19771**	**890**	**8**	**1:12**	**4.35**
Kannada	**1000**	**6652**	**2125**	**25004**	**851**	**6**	**1:41**	**6.05**
Malayalam	**1000**	**6356**	**2077**	**21620**	**1191**	**6**	**1:37**	**5.83**
Marathi	**1000**	**7601**	**2097**	**25558**	**660**	**7**	**1:56**	**6.98**
Tamil	**1000**	**7045**	**2182**	**23284**	**930**	**7**	**1:28**	**5.27**
Telugu	**1000**	**7347**	**2310**	**24743**	**997**	**7**	**1:31**	**5.47**

3.2 LDC-IL Speech Database

Telephone speech signal is recorded using 8 bit and 8 KHz sampling rate and Microphone speech signal is recorded using 16 bit and 16 KHz sampling rate. LDC-IL is supported by the Government of India (GOI) and responsible to create the database to develop speech application in various domains. LDC-IL has collected speech database for various Indian Languages as shown in Table 2. shown speech corpus details.

In LDC-IL Speech corpus is represented in NIST format and it is annotated using PRAAT and wave surfer. The speech signal as segmented in various levels such as phoneme, syllable, word and sentence level. The speech and text database of IIIT-H and LDC-IL are used. The speech and text databases are used for training and testing of

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 116 - 124

Acoustic Model and Language Model respectively inASR.

1) Table 2 LDC – IL Speech Corpus Details

Sl. No.	Language	Raw Data (Hours)	Segmented Data (Hours)
1.	Assamese	105:52:37	28:18:56
2.	Bengali	138:18:47	39:12:31
3.	Bodo	114:38:55	30:45:56
4.	Gujarati	146:23:04	02:39:39
5.	Hindi	163;25:47	80:01:48
6.	Kannada	137:53:28	69:05:56
7.	Kashmiri	44:59:07	06:28:25
8.	Konkani	205:01:48	37:00:00
9.	Maithili	43:33:42	30:10:40
10.	Malayalam	105:47:05	92:40:43
11.	Manipuri	107:10:30	109:48:27
12.	Nepali	145:04:46	39:33:34
13.	Oriya	45:10:25	62:33:15
14.	Punjabi	71:55:56	47:07:13
15.	Tamil	87:03:24	72:09:09
16.	Urdu	81:06:25	36:33:55

II. SYLLABLE SEGMENTATION

The basic problem with the short-term energy function-based speech segmentation is thresholding and local energy fluctuation due to the presence of consonants. The syllable has been defined linguistically as "a sequence of speech sounds having a maximum or peak of inherent sonority between two minima of sonority".

4.1 Properties of Group Delay

Discrete –time real signal $X(n)$, the z transforms is given by

$$\tau_p(\omega) = \frac{X_R(\omega)Y_R(\omega)+Y_1(\omega)X_1(\omega)}{|X(\omega)|^2} \qquad (3)$$

Where $\tau_p(\omega)$ is the group delay function and can be computed directly from the signal, the subscripts R and I denote the real and imaginary parts. $X(\omega)$and $Y(\omega)$ are the Fourier transform of $X(n)$ and $nx(n)$, respectively.

4.2 Group Delay Based Segmentation of Speech

The Figure 3 shows Group delay-based segmentation of speech into syllables.

4.3. Gammatone Function

The Gammatone function is a tone (sinusoid) modulated by a gamma distribution function and is expressed as

$$g(t) = t^{(N-1)}e^{-\alpha t}e^{j\omega_c t}u(t) \qquad (4)$$

Where t (in seconds) denotes time, α is the bandwidth parameter and it determines the effective duration of$g(t)$; the center frequency is ω_c (in radians/second),$u(t)$ is the unit step function and N is the order, which controls the rise and decay of the function andthe ω (in radians/second) is the angular frequency.

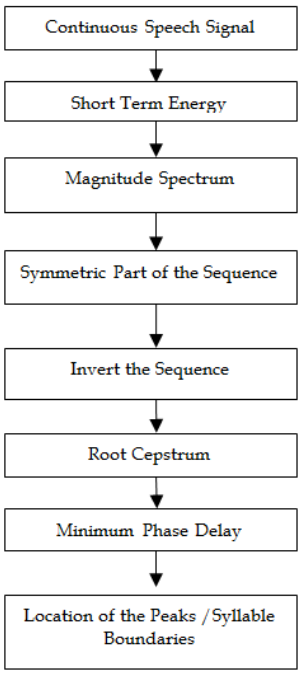


Figure 3Group delay-based segmentation of speech

January - February 2020
ISSN: 0193 - 4120 Page No. 116 - 124

4.4 Gammatone Wavelet Filter bank

The bandwidth of each filter is described as an Equivalent Rectangular Bandwidth (ERB) while Constructingthe Gammatonefilter bank. The expression chosen for calculating ERB (in Hz) at any frequencyf (in Hz) is as follows

$$\mathrm{ERB} = (\mathrm{f}) = \frac{\mathrm{f}}{9.26} + 24.7 \quad (5)$$

The calculation of center frequency f_c for a channel k, is based on the following expression,

$$f_c(k) = -C + e^{k\log\left(\frac{f_{min}+C}{f_{max}+C}\right)/k}.(f_{max} + C) \quad (6)$$

Where $1 \leq k \leq K$, K is the total number of filters, $C\,(\text{constant}) = 228.83$, f_{min} (in Hz) and f_{max} (in Hz) are lowest and highest cutoff frequencies of the filter bank. In our construction $f_{min} = 133$ Hz and $f_{max} = 4$ Hz (half the sampling rate of the input data), are chosen.Gammatonefilter bank provides a good fit auditory response than Mel Frequency Filter bank and Perceptual Linear Prediction (PLP) filter bank.The above method for segmentation the acoustic signal into syllable like units, in which a minimum phase signal is derived from the short-term energy function as if it was a magnitude spectrum.

4.5 Syllable Segmentation

Figure 4 shows that continuous speech signals are segmented into syllables using Gammatone filterbank-based Group Delay Function and the experiment results are given in Table 3.

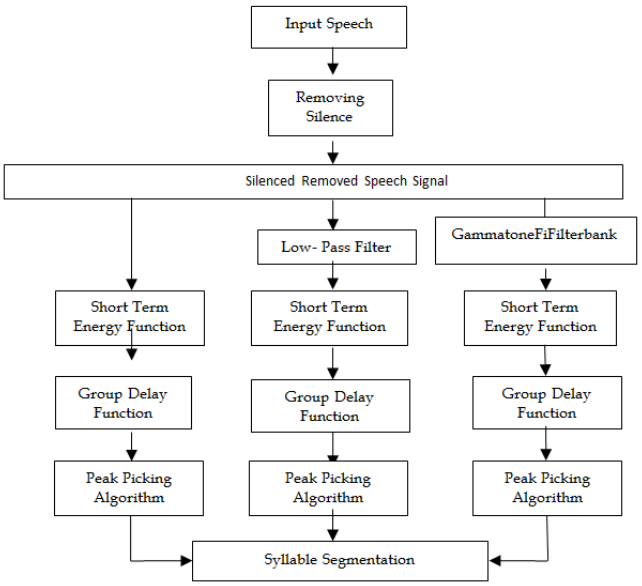


Table 3 Performance (in %) of the Group Delay based segmentation approach

error (in msec)	LDC-IL	IIIT-H
<25	74.81	87.18
25-40	26.56	20.31
40-60	19.84	18.95
60-80	10.31	16.96
Insertion	4.01	3.68
Deletion	5.08	3.02

III. FEATURE EXTRACTION

Gammatone based features yield better recognition performance at low Signal to Noise Ratio (SNR). State of the art ASR feature extraction technique such as Mel Frequency Cepstral Coefficient (MFCC) is ignoring the Fourier transform phase.Joint features derived from Modified Group Delay Function (MOGGDF) and MFCC (Hema A Murthy & Venkata Ramana Rao Gadde 2003)derived from Fourier Transform Magnitude for Continuous Speech Recognizer of Tamil language gave promisingASR performance and compared well with the MFCC.

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 116 - 124

1) 5.1 Feature Extraction using the Modified Group Delay Function

Toconvert theModified GroupDelayFunction(Rajesh M hegde et al. 2014) to some meaningful parameters, the group delay function is converted to cepstrausing the Discrete Cosine Transform (DCT).

$c(n) = \sum_{k=0}^{k=N_f} T_x(k) \cos(n(2k+1)\pi/N_f)$(7)

Where $N_f$is the DFT orderand T_x (k)isthegroupdelayfunction.

Joint Features

$p(\boldsymbol{x_t}|q_t) = \frac{p(q_t|\boldsymbol{x_t})p(\boldsymbol{x_t})}{p(q_t)}$ e following meth

- Compute 13 dimensional MODGDF and the MFCC streamsappended with velocity, acceleration and energyparameters.
-
- Use Feature stream combination to append the 42-dimensional MODGDF stream to the 42-dimensional MFCC stream to derive an84-dimensional joint feature stream.

Table 4 Recognition Performance of the Baseline system for Tamil Language

Category	Joint Features
Whole Syllable	44.3%
Similar syllable	19.9%
Only vowel part	8.5%
Only consonant part	5.1%
Total	77.8%

$$p(v_i = 1|\mathbf{h}) = \sigma(a_i + \sum_j h_j w_{ij})$$

IV. ACOUSTIC MODEL

Gaussian Mixture Model (GMM) is a generative model and Deep Neural Network (DNN) is a discriminative model.There are many efforts on applying syllable that has the basic modeling units of ASR. The motivation comes from the fact that Tamil is Syllabic Language during the past decades several researches have been taken on Tamil Acoustical Modeling. DNN-HMM experiments obtained better Word Error Rate (WER) than a context dependent GMM-HMM(Timothy WONG et al. 2016).

6.1 Deep Neural Networks

Restricted Boltzmann machine was trained to have an input layer representing the features, which are non-linearly transformed through the hidden units. Gaussian-Bernoulli RBM conditional probabilities of visible unitsarecomputedasfollows(Xiangang Li et al. 2013). (8)

Where Z is the partition function, v and h denote visible units and hidden units respectively. Parameter wijdenotes the weight of the edge between visible units viand hidden units hj, and ai and bj are bias terms associated with the i-th visible unit and j-th hidden unit respectively.

Bernoulli-Bernoulli RBM is as follows:

(9)

The conditional probabilities of hidden units are computed as follows:

The k-step contrastive divergence (CD-k) runs a Gibbs sampling for k steps. Usually, k was set to 1. Therefore, Equation (10) is the napproxima

$$CD_k(\boldsymbol{\theta}, \boldsymbol{v}^{(0)}) = -\sum_{\boldsymbol{h}} p(\boldsymbol{h}|\boldsymbol{v}^{(0)}) \frac{\partial E(\boldsymbol{v}^{(0)}, \boldsymbol{h})}{\partial \boldsymbol{h}} + \sum_{\boldsymbol{h}} p(\boldsymbol{h}|\boldsymbol{v}^{(k)}) \frac{\partial E(\boldsymbol{v}^{(k)}, \boldsymbol{h})}{\partial \boldsymbol{\theta}}$$

tedusingthe followingequation.

If the training data v (0)is given, k step Gibbs sampling is performed using conditional probability of visible units and hidden units.

6.2 Hybrid DNN andHMM

Thek-stepsamplingiscomputed using p(h|v) $p(v_i = v|\mathbf{h}) = N(v|a_i + \sum_j h_j w_{ij}, 1)$ ork comput... ...q at frame t has beengiven the acoustic features xat t. Hence, the output probability of HMM at state q, frame t has been computed as follows

January - February 2020
ISSN: 0193 - 4120 Page No. 116 - 124

(11)

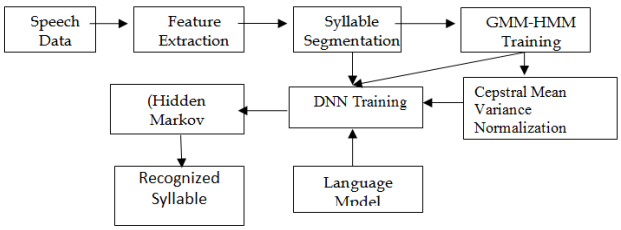
Because observation sequence x does not depend on transition states, p (xt) can be ignored. P (qt) is estimated by counting the occurrence number of "tied state" q in the training data.

A. 6.3 Deep Neural Network (DNN) for Syllable Based Acoustic Modeling

In the DNN based acoustic models, the DNN outputs the posterior probabilities of the acoustic modeling units over the input acoustic feature. In the HMM framework, the acoustic model is always formulated as:

$p(x|w) = \max_q \pi(q_0) \prod_{t=1}^{T} a_{qt-1}\, a_{qt} \prod_{t=0}^{T} p(x_t \backslash qt)$ (12)

In the GMMs based ASR systems, observation probability p(xt|qt) is directly modeled by GMMs.



B. Figure 5DNN based acoustic model

However, in the DNN based ASR systems shown in Figure 5, the observation probability p(xt|qt) is converted by p(xt|qt) = p(qt|xt) p(xt)/p(qt), and p(qt|xt) is the posterior probability modeled by DNN.Through the forced alignment, the input acoustic features are labeled, and then, the pre-trained neural network is fine-tuned discriminatively withBP.

Table 5 WER Using DNN-HMM on LDC-IL+ IIIT-H

DNN-HMM		Multiple models			Single Model		
		Male	Female	Average	Male	Female	Average
	CI	3.8	2.8	3.3	3.5	4	3.75
	TC3	5.1	4.2	4.65	2.9	5.6	4.25
Syllable	CD	6.5	5.4	5.95	4.9	5.6	5.25

V. LANGUAGE MODEL

Sub-word Level Language Model combines(Tomas Mikolov et al. 2012)advantages of both character and word level models.Neural Network based Language model can be order of magnitude which is smaller than compressed n-gram models at the same level of performance and by using sub-word units, the size can be reduced even more.Neural Network Language Model (NNLM) for Low Resource Language (LRL) proves that NNLMs learn better word probability that state of the art n-gram models even when the amount of training data is severely limited.

Recurrent neural network architecture has several layers of nonlinearities. In Syllable based Recurrent Neural Network Language Model (RNNLM) (Tomas Mikolov et al. 2010) architecture,the network is made deeper by adding hidden layers followed by hyper bolictangent nonlinearities.

Input vector x (t) is formed by concatenating vector z representing current syllable, and output from neurons in context layer s at time t-1. The input, hidden and output layers are then computed as follows

$$x(t) = z(t) + s(t) \quad (13)$$

$$s_j(t) = f\left(\sum_i x_i(t) u_{ji}\right) \quad (14)$$

$$y_k(t) = g\left(\sum_i s_j(t) v_{kj}\right) \quad (15)$$

January - February 2020
ISSN: 0193 - 4120 Page No. 116 - 124

Where$f(z)$ is the sigmoid activation function,

$$f(z) = \frac{1}{1+e^{-x}} \quad (16)$$

and$g(z)$ is softmax function

$$g(zm) = \frac{e^{zm}}{\sum_k e^{zk}} \quad (17)$$

For initialization, s(0) can be set to vector of small values, like 0.1 - when processing a large amount of data, initialization is not crucial. In the next subsequent steps, s(t+1) is a copy of s(t). At each training step, error vector is computed according to cross entropy criterion and weights are updated with the standard back propagation algorithm:

$$Error(t) = desired(t) - y(t) \quad (18)$$

Table 6 Performance of RNNLM

Model	WER [%]	Size (MB)	Size (MB) Compressed
RNN-80(text format)	12.98	130	-
RNN-80 (quantized)	13	-	-

VI. PERFORMANCE EVALUATION

- Joint features derived from the Modified Group Delay Function (MOD-GDF)(Pukhraj et al. 2012) and GammatoneCepstral Coefficients (GCC) for syllable based continuous Tamil Language speech recognition can be improved with local forced Viterbi realignment.
- TheHMMdecoderisusedtogeneratenbestlikelyalternative syllabic units corresponding to each isolated testsegment.
- All possible permutations with these n basic units at each position are enumerated and a list of possible syllable strings (sequences) is found.Thecorrectsyllable string (sequence) should be one among the list of possible syllable strings (sequences).
- Theproblemoffindingthestringofsyllabicunitswith maximum likelihood is now transformed to the problemoffindinganoptimalstatesequence.

Table 7 Recognition performance of the enhanced system for Tamil Language ASR

Category	Joint Features
Whole syllable	50.7%
Similar syllables	25.6%
Only vowel part	11.9%
Only consonant part	7.4%
Total	93.26%

Cochlear Filter Cepstral Coefficient works typically used to extract a feature that is different from standard MFCC or PLP and showed are the robustness aspects of GammatoneFilterbank.GammatoneCepstral Coefficients (GCC) provide better results compared to Gammatone wavelet Cepstral Coefficients (GWCC). Increasing the order of the derivative of Gammatone Function gives better accuracy in speech syllable segmentation.

Accuracy of syllable based acoustic Model can be improved by adding speaker adoption. Initial –Onset + nucleus and coda (ONC) based syllable units based acoustical model will have improved Word Error Rate (WER) than initial final based syllable for acoustical model. The Context independent syllables gave better results than context dependent syllable based acoustical model.

The Neural Network Language Model (NNLM) [2] performscomparatively better for both Limited Language Pack (Limited LP) and full Language Pack (Full LP). The Syllable based Recurrent Neural Network Language Model (RNNLM) gives almost same WER for both low resource and high resource Tamil text corpus.

January - February 2020
ISSN: 0193 - 4120 Page No. 116 - 124

VII. CONCLUSION

All the experiments have been conducted by using IIIT-H and LDC-IL speech Databases. Most of the Indian Languages are syllable based ones. Syllable based ASR functional units such as Speech segmentation, Feature Extraction, Acoustical Model and Language Model are investigated in this work and concluded that syllable-based Speech Recognition system outperforms the phoneme and word-based ASR for Tamil language.

The proposed Tamil language speech syllable segmentation approach gives low error rate in IIIT-H and LDC-IL Tamil Language speech databases. The experiments have shown that 60-80 msec time duration speech syllables are segmented with low error rate and 76.18% of recognition performance of ASR system has been obtained for Tamil Language. GMM-HMM based ASR accuracy is decreased due to diversity of speakers. Syllable based acoustical Model for Tamil Language ASR using DNN-HMM with various speaker produces better resultsthan GMM-HMM Model. RNNLM of Low Resource Languagelike Tamil gives comparable perplexity and WER to n-gram Language Model for high Resource Language like English. Long Short-Term Memory (LSTM) RNN based acoustical model and Deep Neural Network (DNN) based language model for Tamil language can be attempted to obtain further improvement.

REFERENCES

[1] Aniruddha Adiga, Mathew Magimai-Doss & Chandra Sekhar Seelamantula 2013, 'Gammatone Wavelet Cepstral Coefficients for Robust Speech Recognition', IEEE Region 10 Conference (31194), pp. pp.1 – 4,TENCON

[2] Ankur Gandhe, Florian Metze, Ian Lane 2016, 'Neural Network Language Models For LowResourceLanguages'.http://www.cs.cmu.edu/~fmetze/interACT/Publications/IS14full_NNLM_lowResource_CameraReady.pdf.[30 JUNE 2016].

[3] Hema A Murthy &Venkata Ramana Rao Gadde 2003,'The Modified Group Delay Function And Its Application To Phoneme Recognition', IEEE Transaction on signal Processing, vol. 40, pp.2281-2289.

[4] Hiroshi Seki, Kazumasa Yamamoto&Seiichi Nakagawa 2014, 'Comparison of Syllable-based and Phoneme-based DNN-HTML in Japanese Speech Recognition'.

[5] kishore Prahallad, E, Naresh Kumar, Venkatesh Keri, S Rajendran & Alan W Black, 'The IIIT-H Indic Speech Databases',pp.5-10. <http://speech.iiit.ac.in/svlpubs/conference/IIITHIndic_Speech_Databases.pdf.

[6] Lakshmi A Hema a Murthy 2006, 'A Syllable Based Continuous Speech Recognizer for Tamil', in INTERSPEECH , International Conference on Spoken Language Processing – ICSLP.

[7] Matti Karjalainen 1987,'Auditory Models For Speech Processing', Proc. Int. Congr. Phon. Sciences. Tallinn

[8] Nagarajan,T, Hema AMurthy & Rajesh M Hegde 2003, Segmentation of Speech into Syllable-like Units',EUROSPEECH.

[9] Pukhraj P Shrishrimal, Rathadeep R Deshmukh & Vishal B Waghmare 2012, 'Indian Language Speech Database : A Review', vol.47, no.5.

[10] Rajesh M Hegde, Hema A Murthy & Gadde VRAmana Roa 2014,'Continuous Speech Recognition Using Joint Features Derived From The Modified Group Delay Function And MFCC',INTERPEECH –ICSLP, 8th International Conference on Spoken Language Processing.

[11] Timothy Wong, Claire Wyli, Sam Lam, Billy Chiu, Qin Lu, Mingelei Li, Dan Xiong, Roy S,Yu& Vineent, TYNG 2016,'Syllable Based

January - February 2020
ISSN: 0193 - 4120 Page No. 116 - 124

DNN-HMM Cantonese Speech-to-Text System', Irec conference.

[12] Tomas Mikolov, Ilya Sutskever, Anoop Deoras, Hai-Son Le, Stefan Kombrink & Jan Cernacky 2012, 'Subword Language Modeling With Neural Networks', Preprint.

[13] Tomas Mikolov, Martin Karafiat & Lukas Burget 2010,'Recurrent Neural Network Based Language Model', Interspeech, pp.26-30.

[14] Xiangang Li, Caifu Hong, Yuning Yang & Xihong Wu 2013,'Deep Neural Networks For Syllable Based Acoustic Modeling In Chinese Speech Recognition',signal and information Processing Association Annual Summit and Conference (APSIPA),Asia-Pacific.