IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 20, Issue 6, Ver. II (Nov - Dec 2018), PP 69-74
www.iosrjournals.org

Isolated Pali Word (IPW) Feature Extraction using MFCC & KNN Based on ASR

Siddharth S More[1], Prashant kumar L. Borde[2] Sunil S Nimbhore[3]
[1](Dept. Of Computer Science & IT, Dr.Babasaheb Ambedkar Marathwada University, Aurangabad (MS) India)
[2](Dept. Of Computer Science & IT, Dr.Babasaheb Ambedkar Marathwada University, Aurangabad (MS) India)
[3](Dept. Of Computer Science & IT, Dr.Babasaheb Ambedkar Marathwada University, Aurangabad (MS) India)
Corresponding Author: Siddharth S More

Abstract: *Speech is most prominent & primary mode of Communication among human being. Past after years of research and development & the accuracy of automatic speech recognition remains the important challenges in research. The design of Speech Recognition system require careful attentions the following issues: Definition of various types of speech classes, speech representation, feature extraction techniques, speech classifier, and database and performance evaluation. The problems are existing in ASR and the various techniques to solve these problems constructed by various research work have been presented in a chronological order. The objective of this research paper is to summarize and compare some of the well-known methods used in various stages of speech recognition system and to identify research topic and applications which is front of this exciting and challenging field. In this paper is computed acoustic features using MFCC on 'IPW' Isolated Pali Word Database which contain collection of isolated number (0-10) and isolated month (Jan-Dec) words of 20 speakers. The performance of the isolated digit and month's words based on MFCC and KNN classification were 80.36% and 81.83% respectively.*

Date of Submission: 07-12-2018 Date of acceptance: 22-12-2018

I. Introduction

The Pali Language is most historical gift to India; this was widely used language for education, communication, and writing. Therefore Pali language literature is full of knowledge of social development, economical development, national and international political, environmental development of nation. As compared Pali language values in international and national level research is very few in Automatic Speech Recognition system (ASR). In this research article we have selected isolated word of Pali language to process and achieved better recognition rate. Pali language is a Middle Indo-Aryan language of north Indian origin. It is also known as Magadhi, although it was spoken, or at least well understood, in almost the whole of Northern India in the Buddha's time. Pali is a reverberating, sweet language. Besides these, there is some other literature comprising works on grammar, metrics, prosody, etymology, rhetoric, logic, astrology, polity, history, genealogy, medicine, pharmacology [1-2]. There is many works done in Indian language like Marathi, Hindi, Tamil, etc. But very few works have been done in Pali language. Pali language is among are of the historical Indian language, this language is very important in Indian culture like other language, the thoughts of Goutam Buddha and ideology available in the Pali literature this culture, speared all over in the world like Sri Lanka, Myanmar, Sayam (Thailand), Burma or also China , Japan these countries connected to India by the relation of friendship because of only Pali language ,there is some countries Taiwan or Myanmar this countries is known as their religion or local language and they are creating their literature in Pali language till today.
In this research paper, focused on isolated word of Pali Language. This paper is organized as follows: Section 1 about the introduction section 2 deals with Literature Review Section 3 Describes methodology and feature extraction Techniques. Section 5 deals with experimental work and section 5 describe the conclusion.

II. Related Work

Automatic speech recognition is most important application in various areas. Witch mention there are many parts of speech recognition system available till now and several feature extraction techniques for speech recognition system. [3] The various techniques for acoustic feature extraction like Mel Frequency Cepstrum Coefficient (MFCC), linear predictive coding (LPC) and so on. During his research, researcher got MFCC is well known techniques is used for speaker recognition system. The types of speech recognition techniques and algorithm for better speech to text conversion using MFCC feature extraction techniques and HMM was mostly used for forming speech is present ~~in~~ [4]. The automatic isolated speech recognition system using Dynamic Time Warp (DTW) and Hidden Markov Model approach shown in [5-6]. The studies of behavior of Long Short-

Term Memory (LSTM)-based neural networks on a specific task of automatic speech processing: speech detection. LSTM model were compared to two neural models: Multi-Layer Perceptron (MLP) and Elman's Recurrent Neural Network (RNN). Tests on five speech detection tasks show the efficiency of the Long Short-Term Memory (LSTM) model. All tests show that the LSTM model is more efficient than Elman MLP and RNN neuron networks describe in [7]. The paper [8] presents a spontaneous speech recognition system for Myanmar language. Author built recognizer for Myanmar interview speech by using the techniques Gaussian mixture model, based on non-linear distortion, accents and weakened articulation. The Gaussian densities in Myanmar Interview speech achieved the best Word Error Rate (WER) of 20.47%. In [9] research paper, author describe the speech recognition of Marathi digits using MFCC and DTW techniques and Vector Quantization is used to minimize the data of the extracted feature. In paper [10], proposed a new lossy algorithm to compress speech signal using DWT techniques. The growth of multimedia technology over the past decade, demand for digital information has increased. The only way to overcome this situation is to compress the information signal by removing the redundancies present in them. The compression ratio can be easily varied by using wavelets while other methods have fixed compression ratios. The importance and need for audio compression. Audio compression has become one of the most basic technology present in paper [11].

III. Methodology

The typical ASR system accepts the audio input as that is shown in figure 1. The audio data input is captured using Sennheiser standard audio mic. Once the input is acquired, it will be preprocessed for acoustic feature extraction and further used for recognition of utterance.

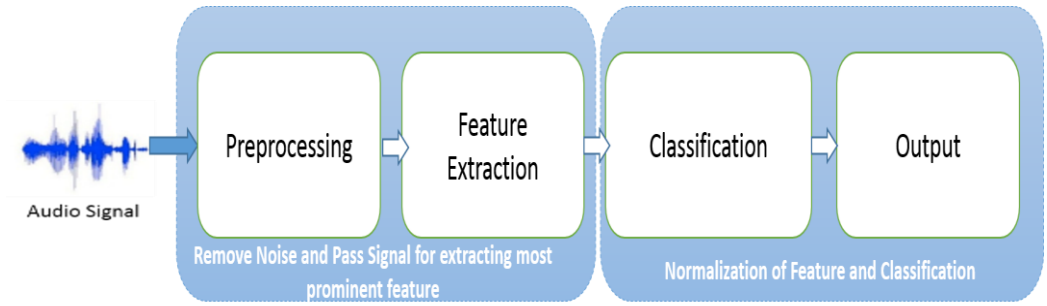

Figure 1 show the framework of Automatic Speech Recognition System

The figure shows the framework for developing robust speech recognition system. First step to collect the speech database, then pass it to preprocessing. The preprocessing part includes cleaning of signal and removal of silence which exists in the signal before utterance and after utterance. This procedure return the absolute signal representing only user utterance corresponding to isolated word. Then preprocessed acoustic signal will pass for feature extraction and finally classification techniques used to classify data.

1. Data Acquisition

In order to design automatic speech recognition system, the researchers have designed the database as per requirement that meets to their research problem in general. The Isolated Pali Words (IPW) database was developed through this research work. The 'IPW' speech database was developed at Natural Sounding Speech Recognition and Speech Synthesis Lab (NSSR&SS) of Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University. The database is collected from 20 native speakers i.e. 12 were male and 8 were female speaker in the age group of 25-35. The quality of digital audio signal is determined by discrete parameters that are the sample rate, bit capacity and number of channels. The recording was done in noisy environment (with noisy fans). The IPW database was recorded by using Sennheiser a standard headset microphone connected to the laptop and 16 KHz sampling rate were used. We have used PRAAT software tool to record the speech at 16 KHz sampling frequency and represented using 16 bits/sample [12]. The database consisting isolated numerals (0-10), Months (Jan-Dec) and weeks (Mon – Sun) were selected for the constitution of database as shown in table 1. The total volume of IPW speech database by native speakers is 6,000.

Class of Pali word	No of Words	Corpus Set of Pali words Language	Speakers	Total recorded Samples
Digit	11	{सुय्य, एक, व्दे, तयो, चतु चत्तारो, पञ्च, छ, सत्त, अट्ठ, नव, दस}	20	2,200
Week	07	{ चन्दवारो, कुंजवारो, बुधवारो, गुरुवारो, सुक्रवारो, सनिवारो, आदिच्चवारो }	20	1,400
Months	12	{ चित्तमासी, वेसाख, जेट्ठ, आसळह, सावन, भद्दपद, अस्सयुजो, कत्तिको, मागसिरो, पुस्स, माघो, फग्गुण्णो }	20	2,400
Total Volume of Database :				**6,000**

Table 1 shows the description of IPW speech corpus.

2. Feature Extraction

Mel-Frequency Cepstral Coefficients (MFCC) approach is the most popular because it uses the spectral based base as parameters for recognition. MFCC's are the coefficients, which represent audio based on perception of human auditory systems. The reason behind selection of MFCC for recognition purpose due to its peculiar difference between the operations of FFT/DCT. In the MFCC, the frequency bands are positioned logarithmically (on the Mel scale) which approximates the human auditory system's response more closely than the linearly spaced frequency bands of FFT or DCT [13]. Figure 2 shows the block diagram of MFCC features extraction process. The After acquisition it is necessary to remove noise contained in speech signal and it will be removed by using first order high pass filter. This process helps in cleaning acoustic-input. Frame blocking phase converts segmented concatenated voiced speech signals in to frames. The discontinuities contained in the segmented signal is checked at the beginning and end of each frame using hamming window and this was performed by windowing step. Fast Fourier Transform (FFT) brings the each frame of signal from time domain to frequency domain and the result is said to be 'spectrum'.

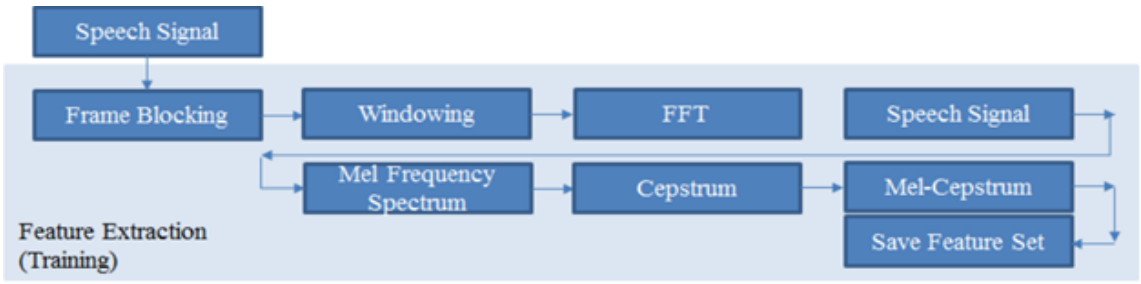

Figure 2 Block diagram of MFCC Feature extraction

IV. Experimental Work

1. Recognition of Isolated Pali Digits:

In this experiment, (Isolated Pali Words) 'IPW' dataset of isolated digit were utilized. The Pali language number word corpus included number (0-10) such as {सुय्य, एक, व्दे, तयो, चतु चत्तारो, पञ्च, छ, सत्त्त, अट्ठ, नव, दस} was considered for processing. The digit data set consists of 20 speakers (12 male and 8 female) and uttering isolated word 10 times. Therefore the database comprised of 2200 utterance (20*10*11) of these independent standard words. The isolated digit words have been processed for acoustic feature extraction using MFCC feature extraction techniques. MFCC features for all audio speech samples from training set and test set were computed is shown in table 2.

Isolated Words	1	2	3	4	5	6	7	…	13
Zero1	0.634696	0.825286	0.032174	-0.32119	-0.63235	-0.60462	-0.19511	…	-0.24133
Zero2	0.700435	1.306887	0.322338	-0.15445	-0.48658	-0.47775	-0.39196	…	-0.27303
Zero3	0.656792	1.147747	0.249968	-0.05726	-0.51077	-0.66095	-0.35586	…	-0.30978
Zero4	0.595056	0.888343	0.021677	-0.16739	-0.69384	-0.51094	-0.26505	…	-0.19809
Zero5	0.644909	1.01405	0.153434	-0.07311	-0.55919	-0.63904	-0.46874	…	-0.17955
Zero6	0.513662	1.375368	0.432333	0.102496	-0.43633	-0.51165	-0.37919	…	-0.21966
Zero7	0.672397	1.066945	0.210333	-0.09344	-0.61385	-0.6616	-0.34062	…	-0.24838
Zero8	0.528119	0.885022	-0.01486	-0.15133	-0.49166	-0.63008	-0.25165	…	-0.25724
Zero9	0.273876	0.812785	-0.34486	-0.36058	-0.61108	-0.56058	-0.14538	…	-0.04965
Zero10	0.870321	0.95996	-0.08818	-0.45585	-0.81949	-0.63937	-0.29886	…	-0.1636
One1	0.569871	0.564579	-0.32868	-0.38984	-0.67472	-0.54136	-0.14189	…	-0.29226
One2	0.469035	0.661252	-0.22991	-0.41401	-0.80978	-0.58228	-0.25778	…	-0.3016
One3	0.630776	0.795747	-0.06933	-0.28194	-0.67698	-0.62101	-0.2359	…	-0.29512

…	…	…	…	…	…	…	…	…	…
Ten10	0.453026	0.77977	-0.06763	-0.21283	-0.74485	-0.64835	-0.22525	…	-0.24982

Table 2 shows the MFCC Feature set of digit dataset.

The recognition of acoustic signal, the data set of isolated digit words was divided in to 50% known sample to be used for training systems and 50% unknown sample to be used at the time of testing respectively were consider for classification using K-nearest Neighbor (KNN). The k-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms [14]. An object is classified by a majority vote of its neighbors, with the object being assigned the class most common amongst its k nearest neighbors. k is a positive integer, typically small. If k = 1, then the object is simply assigned the class of its nearest neighbor. In binary classification (i.e. two class classification) problems, it is helpful to choose k to be an odd number.

The same method can be used for regression, by simply assigning the property value for the object to be the average of the values of its k nearest neighbors. It can be useful to weight the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. The neighbors are taken from a set of objects for which the correct classification is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. In order to identify neighbors, the objects are represented by position vectors in a multidimensional feature space. It is usual to use the Euclidean distance, though other distance measures, such as the Manhattan distance could in principle be used instead. The k-nearest neighbor algorithm is sensitive to the local structure of the data.

The KNN classifier was used for measuring the similarity between test (1100 Samples) and train (1100 samples). It was seen that out of 1100 testing samples 885 samples were correctly recognized and 215 samples were misclassified. The study reveals that result in 80.45% correct classification over digit dataset. The confusion matrix of KNN classification is shown in table 3.

Subject	**Test Samples**	**Pali Digit Dataset (0-10)**											**Classification Result**		
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Recognized**	**Miss**	**Result**
Amardip	55	2	4	2	5	5	3	4	3	4	4	4	41	14	75.54
Anand	55	4	3	3	5	5	0	4	4	5	5	5	43	12	78.78
Kavita	55	5	1	4	4	5	4	5	3	1	5	5	42	13	76.36
Chetan	55	5	3	2	5	4	5	4	5	4	3	5	45	10	81.81
Ganesh	55	3	3	2	2	5	5	5	2	5	2	5	40	15	72.72
Priyanka	55	4	3	4	4	4	5	5	3	3	5	4	46	9	83.63
Arshad	55	5	4	5	3	5	4	4	2	5	5	3	45	10	81.81
Khandare	55	4	4	5	5	5	3	5	3	4	5	5	48	7	82.27
Sarita	55	5	3	5	5	4	5	5	3	2	4	3	44	11	80
Minakashi	55	5	3	5	2	5	2	2	5	4	5	5	43	12	78.18
Shruti	55	5	5	4	4	5	4	3	3	3	4	5	45	10	81.81
Rahul	55	4	5	2	5	5	3	5	3	4	3	5	44	11	80
Sariput	55	5	4	5	4	4	3	3	5	4	5	3	45	10	81.81
Aishwini	55	4	5	2	5	4	2	3	7	4	4	5	43	12	78.18
Manisha	55	4	4	5	5	3	5	4	3	4	5	3	45	10	81.81
Siddharth	55	2	3	5	3	5	5	4	5	5	5	3	45	10	81.81
Umesh	55	3	4	3	4	5	5	5	3	5	5	4	46	9	83.63
Komal	55	5	1	5	2	4	4	5	3	4	5	5	43	12	76.36
Vipin	55	4	4	2	4	5	4	3	5	5	4	5	45	10	81.81
Vishal	55	3	5	2	5	5	4	6	5	5	4	4	47	8	85.45
Total Test Samples	**1100**	**Final Result :**											**885**	**215**	**80.36 %**

Table 3 show the confusion matrix of Pali digit database using KNN classification.

2. Recognition of Isolated Pali Words (Months)

In this experiment, (Isolated Pali Words) 'IPW' dataset of isolated month were utilized. The Pali language month corpus included number (April - March) such as {चित्तमासी, वेसाख, जेट्ठ, आसळह, सावन, भद्दपद, अस्सयुजो, कत्तिको, मागसिरो, पुस्स, माघो, फग्गुण्णो} was considered for processing. The Month data set consists of 20 speakers (12 male and 8 female) and uttering isolated word 10 times. Therefore the database comprised of 2400 utterance (20*10*12) of these independent standard words. The isolated month words have been processed for acoustic feature extraction using Mel-Frequency Cepstral Coefficients (MFCC) as describe in experiment no 1 and MFCC feature of isolated month word are shown in table 4.

Isolated Month Words	1	2	3	4	5	6	7	…	13
चित्तमासी1	2.205189	0.391747	-0.281	-0.26266	-0.58213	-0.57314	-0.38832	…	-0.37938
चित्तमासी2	2.012579	0.378128	-0.17781	-0.22621	-0.52381	-0.59609	-0.47928	…	-0.36396
चित्तमासी3	1.936622	0.422028	-0.12959	-0.17625	-0.39493	-0.52358	-0.5367	…	-0.30502
चित्तमासी4	2.004344	0.417594	-0.19219	-0.18191	-0.4553	-0.48639	-0.41783	…	-0.40226
चित्तमासी5	2.130383	0.526543	-0.17719	-0.21627	-0.41276	-0.55851	-0.51981	…	-0.32686
चित्तमासी6	2.112196	0.487839	-0.2084	-0.26025	-0.53452	-0.56631	-0.48088	…	-0.34125
चित्तमासी7	2.125126	0.611297	-0.02485	-0.20924	-0.49558	-0.51838	-0.49282	…	-0.32397
चित्तमासी8	2.136876	0.611714	0.012787	-0.2058	-0.57784	-0.53796	-0.43391	…	-0.4289
चित्तमासी9	2.209927	0.447056	-0.2319	-0.23845	-0.53743	-0.5477	-0.41563	…	-0.41193
चित्तमासी10	2.114447	0.710005	0.030526	-0.22706	-0.5829	-0.56351	-0.52579	…	-0.37747
वेसाख1	1.723191	-0.28832	-0.80793	-0.36303	-0.61142	-0.66428	-0.33568	…	-0.40029
वेसाख2	1.813015	0.008201	-0.4599	-0.35678	-0.66625	-0.67228	-0.38959	…	-0.47381
वेसाख3	1.897467	0.150254	-0.38147	-0.15959	-0.42511	-0.56031	-0.45454	…	-0.38904
…	…	…	…	…	…	…	…	…	…
फग्गुण्णो10	2.417374	0.754378	0.009318	0.028987	-0.33449	-0.42118	-0.42405	…	-0.41832

Table 4 show the MFCC features of Isolated Month Database.

The recognition of acoustic signal, the data set of isolated month was divided in to 50% known sample (1200 samples) to be used for training systems and 50% unknown sample (1200 samples) to be used at the time of testing respectively were consider for classification using K-nearest Neighbor (KNN). After Appling KNN classifier on month dataset, it was seen that out of 1200 testing samples 982 samples were correctly recognized and 218 samples were misclassified. The study reveals that result in 81.83% correct classification over digit dataset. The confusion matrix shown in table 5.

Spaker	**Test Samples**	**Pali Month Dataset (Jan-Dec)**												**Classification Result**		
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Recognized**	**Miss**	**Result**
Amardip	60	4	4	3	3	4	4	5	4	4	4	5	5	49	11	81.66
Anand	60	3	5	3	4	2	5	5	5	5	4	5	5	51	9	85
Arshad	60	5	5	5	5	3	5	3	3	1	5	5	5	50	10	83.33
Chetan	60	3	4	5	5	5	5	5	4	5	4	4	5	54	6	90
Minakashi	60	4	5	5	3	4	4	3	5	4	5	5	5	52	8	86.67
Ganesh	60	3	5	1	5	3	5	5	5	2	5	4	3	45	15	75
Lalit	60	3	4	4	4	5	4	5	5	2	5	3	4	48	12	80
Siddharth	60	5	4	5	5	5	4	3	5	3	5	5	5	54	6	90
Khandare	60	2	3	1	2	5	5	3	5	5	5	5	5	46	14	76.66
Priyanka	60	4	5	3	4	5	5	5	3	2	4	1	4	45	15	75
Shruti	60	3	4	4	5	5	5	4	5	5	3	3	5	51	9	85
Sariput	60	3	4	3	5	4	5	4	4	3	5	4	4	48	12	80
Vipin	60	2	4	5	4	4	5	5	4	5	5	5	5	53	7	88.33
Sarita	60	5	2	5	3	5	4	2	3	4	5	5	5	48	12	80
Komal	60	2	5	1	5	4	5	5	3	3	5	4	4	46	14	76.66
Kavita	60	3	3	5	4	5	4	3	3	4	5	3	3	45	15	75
Rahul	60	3	4	4	5	4	5	4	4	3	5	4	5	50	10	83.33
Manisha	60	4	3	5	4	4	3	5	4	3	3	5	5	48	12	80
Umesh	60	5	4	3	2	4	4	2	3	3	4	5	5	44	16	73.33
Vishal	60	5	5	4	5	4	5	4	4	5	5	5	4	55	5	91.66
Total Test Samples	**1200**	**Final Result:**												**982**	**218**	**81.83 %**

Table 5 shown the confusion matrix of month dataset.

V. Conclusion

In conclusion, an efficient, abstract and fast ASR system for Pali Language is need of the hour. The work implemented in the paper is a step towards the development of robust automatic speech recognition system for Pali Language. The work may further be extended to large vocabulary size of Pali language and to continuous speech recognition system. The work highlighted in this research work, provides an idea about

isolated Pali word recognition using MFCC features. The success rate of classification were 80.36% using MFCC & 81.83 using KNN for (IPW) digits & month.

Acknowledgement

The Authors gratefully acknowledge support by the **BARTI (Dr. Babasaheb Ambedkar Research and Training Institute)** for providing financial assistance and the authorities of **Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS) India**, for providing the infrastructure for this research work.

References

[1]. Hammarström, Harald; Forkel, Robert; Haspelmath, Martin, eds. (2017). "Pali". Glottolog 3.0. Jena, Germany: Max Planck Institute for the Science of Human History.
[2]. Hazra, Kanai Lal. Pāli Language and Literature; a systematic survey and historical study.D.K. Printworld Lrd. New Delhi, 1994.
[3]. Shipra Gupta, Application of MFCC in Text Independent Speaker Recognition, IJARCSSE, ISSN: 2277 128X, Volume 6, Issue 5, May 2016.
[4]. Mridula Shanbhogue, Shreya Kulkarni, R Suprith, Tejas K I, Nagarathna N, A Study on Speech Recognition, IJRASET, ISSN: 2321-9653, Volume 4 Issue III, March 2016.
[5]. Doda, Er Shweta, and Er Rajni Mehta. "Speech Recognition Techniques: A Review." JCD Vidyapeeth, Sirsa, India 4, no. 8 (2014).
[6]. Asha Shajee, Dhruv Patel, Rahul Mishra, Himangshu Saikia, Narendran.M, A SURVEY: SPEECH RECOGNITION APPLICATION, International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835, Volume-4, Issue-11, Nov.-2017.
[7]. Al-Rababah, Mohamad AA, Abdusamad Al-Marghilani, and Akram Aref Hamarshi. "Automatic Detection Technique for Speech Recognition based on Neural Networks Inter-Disciplinary." INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS 9, no. 3 (2018): 179-184.
[8]. Naing, Hay Mar Soe, and Win Pa Pa. "Automatic Speech Recognition on Spontaneous Interview Speech".
[9]. Gedam, Yogesh K., Sujata S. Magare, Amrapali C. Dabhade, and Ratnadeep R. Deshmukh. "Development of Automatic Speech Recognition of Marathi Numerals-A Review." International Journal of Engineering and Innovative Technology (IJEIT) 3, no. 9 (2014).
[10]. HatemElaydi and Mustafi I. Jaber and Mohammed B. Tanboura, "Speech compression using Wavelets", International Journal for Applied Sciences, Vol 2, 1-4, Sep 2011.
[11]. Othman O. Khalifa, SeringHabib Harding & Aisha-Hassan A. Hashim "Compression using Wavelet Transform" in Signal Processing: An International Journal, Volume (2): Issue (5).
[12]. Standards for Speech Data Capturing and Annotation, available online http://www.ldcil.org/download/samplingstandards.pdf, 2016.
[13]. Clarence Goh Kok Leon, Politeknik Seberang Perai, Jalan Permatang Pauh, "Robust Computer Voice Recognition Using Improved MFCC Algorithm", International Conference on New Trends in Information and Service Sci-ence,2009.
[14]. Y. Kim and E. Mower Provost, "Emotion classification via utterance-level dynamics: A pattern-based approach to characterizing affective expressions," in Proceedings of IEEE ICASSP 2013. IEEE, 2013.

Siddharth S More. " Isolated Pali Word (IPW) Feature Extraction using MFCC & KNN Based on ASR." IOSR Journal of Computer Engineering (IOSR-JCE) 20.6 (2018): 69-74.