Marathi Isolated Word Recognition System using MFCC and DTW Features

Bharti W. Gawali[1], Santosh Gaikwad[2], Pravin Yannawar[3], Suresh C.Mehrotra[4]
[1,2,3,4]Department of Computer Science & Information Technology,
Dr.Babasaheb Ambedkar Marathwada University,
Aurangabad. 431001(MS) India.
Address: 1bharti_rokade@yahoo.co.in, , 2santosh.gaikwadcsit@gmail.com
3pravinyannawar@gmail.com , 4mehrotra_suresh@yahoo.com

***Abstract*— This paper presents a Marathi database and isolated Word recognition system based on Mel-frequency cepstral coefficient (MFCC), and Distance Time Warping (DTW) as features. For the extraction of the feature, Marathi speech database has been designed by using the Computerized Speech Lab. The database consists of the Marathi vowels, isolated words starting with each vowels and simple Marathi sentences. Each word has been repeated three times by the 35 speakers. This paper presents the comparative recognition accuracy of DTW and MFCC.**

***Index Terms*— CSL, MFCC, DTW, Spectrogram, Speech Recognition and statistical method**

I.. Introduction

The Speech is the most prominent and natural form of communication between humans. There are various spoken Languages thought the world [1]. Marathi is an Indo-Aryan Language, spoken in western and central India. There are 90 million of fluent speakers all over world [2].However; there is lot of scope to develop systems using Indian languages which are of different variations. Some work is done in this direction in isolated Bengali words, Hindi and Telugu [3].The amount of work in Indian regional languages has not yet reached to a critical level to be used it as real communication tool, as already done in other languages in developed countries. Thus, this work was taken to focus on Marathi language. It is important to see that whether Speech Recognition System for Marathi can be carried out similar pathways of reaserch as carried out in English [4, 5].In this paper we are presenting work consists of the creation of Marathi speech database and its speech recognition system for isolated words.

The paper is divided into five sections, Section 1, gives Introduction, Section 2 deals with details of creating Marathi speech database ,section 3,focuses on Recognition of isolated words using MFCC and DTW, Section 4 ,covers results and conclusion followed by section 5 with the References.

Dr.Bharti W.Gawali
Associate Professor
Department of Computer Science & Information Technology
Dr.Babasaheb Ambedkar Marathwada University Aurangabad.
Email. bharti_rokade@yahoo.co.in
DST Project Sanction: SR/FTP/ETA-0009/2010

II. MARATHI SPEECH DATABASE

For accuracy in the speech recognition, we need a collection of utterances, which are required for training and testing. The Collection of utterances in proper manner is called the database.

The generation of a corpus of Marathi Vowels, words and sentences as well as the collection of speech data are described below. The age group of speakers selected for the collection of database ranges from 22 to 35. Mother tongue of all the speakers was Marathi. The total number of speakers was 35 out of which 17 were Females and 18 were Males. The vocabulary size of the database consists of

- Marathi Vowels: 105 samples
- Isolated words stating with each vowel: 420 Samples
- Sentences: 175 samples.

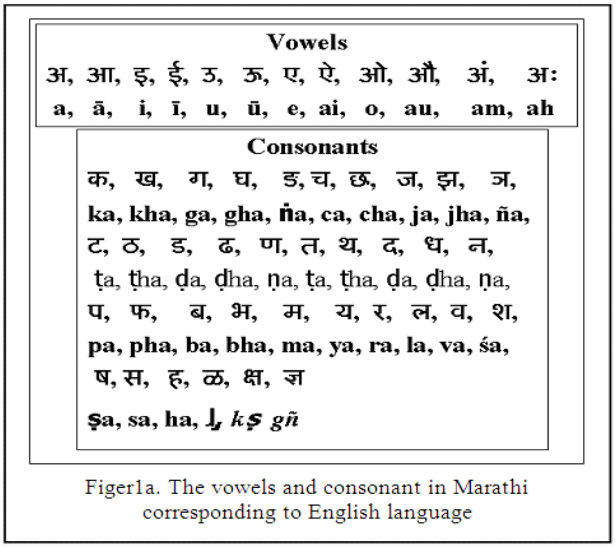
A. Acquisition setup

To achieve a high audio quality the recording took place in the10 X 10 rooms without noisy sound and effect of echo. The Sampling frequency for all recordings was 11025 Hz in the Room temperature and normal humidity. The speaker were Seating in front of the direction of the microphone with the Distance of about 12-15 cm [6]. The speech data is collected with the help of Computerized speech laboratory (CSL) using the single channel. The CSL is most advanced analysis system For speech and voice. It is a complete hardware and software system with specifications and performance. It is an input/output recording device for a PC, which has special features for reliable acoustic measurements [7].

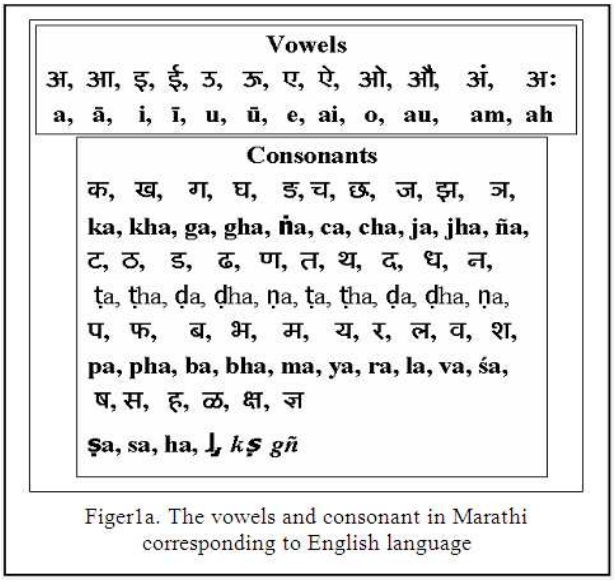
B. Vowels, Words and Sentences corpus

Marathi language uses Devanagari, a character based script. A character represents one vowel and zero or more consonants. There are 12 vowels and 36 consonants present in Marathi languages as shown in figure 1.a [8]. We have recorded vowels, isolated words starting from each vowel and the simple sentences which are used for communication in Government offices shown in figure 1.b. All sentences are interrogative and containing 5 to 6 words. The speech signals have been stored in form of wav file. Each vowels, words and sentences have been separated using CSL software and stored and given the labels of spoken words to the different files and folders.

DOI: 01.IJIT.01.01.73



Figer1a. The vowels and consonant in Marathi corresponding to English language



Figer1a. The vowels and consonant in Marathi corresponding to English language

III. WORD RECOGNITION SYSTEM

There are several kinds of parametric representation of the acoustic signals. Among of them the Mel-Frequency cepstral Coefficient (MFCC) is most widely used [9].There are many work on MFCC, on the improvement and accuracy in recognition [10].The recognition rate achieved by MCC, LPC and Combined are given 87.5%, 91.61% and 84.6% respectively for 312 number of testing patterns [11].We have developed the recognition system using MFCC and DTW.paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

A. Mel Frequency Cepstral Coefficient

It consists of various steps described below.

Speech Signal

The excitation signal is spectrally shaped by a vocal tract Equivalent filter. The outcome of this process is the sequence of exciting signal called speech

Pre-emphasis

The speech is first pre-emphasis with the pre-emphasis filter 1-az-1 to spectrally flatten the signal.

Framing and Windowing

A speech signal is assumed to remain stationary in periods of approximately 20 ms. Dividing a discrete signal s[n] into frames in the time domain truncating the signal with a window function w[n]. This is done by multiplying the signal, consisting of N samples, . The frame is shifted 10 ms so that the overlapping between two adjacent frames is 50% to avoid the risk of losing the information from the speech signal. After dividing the signal into frames that contain nearly stationary signal blocks, the windowing function is applied.

Fourier Transform

To obtain a good frequency resolution, a 512 point Fast Fourier Transform (FFT) is used [12, 13]

Mel-Frequency Filter Bank

A filter bank is created by calculating a number of peaks, Uniformly spaced in the Mel-scale and then transforming the back to the normal frequency scale where they are used a speaks for the filter banks.

Discrete Cosine Transform

As the Mel-cepstrum coefficients contain only real parts, the Discrete Cosine Transform (DCT) can be used to achieve the Mel- cepstrum coefficients. There were 24 Coefficients out of that only 13 coefficients have been selected for the recognition system.

Distance Measures

There are some commonly used distance measures i.e. Euclidean Distance. Euclidian Distance of two vectors x and p is used measured. The result of MFCC on the vector of words starting with अ are presented in Table 1.

B. Dynamic Time Warping

The simplest way to recognize an isolated word sample is to compare it against a number of stored word templates and determine the best match [14,15].DTW is an instance of the general class of algorithms and known as dynamic programming .its time and space complexity is merely linear in duration of speech sample and the vocabulary size. The algorithm makes a single pass though a matrix of frame scores while computing locally optimized segment of the global alignment path. The dynamic time warping algorithm provides a procedure to align in the test and reference patterns to give the average distance associated with the optimal warping path. The results of DTW are given in table 2.

IV. RESULTS AND CONCLUSION

The aim here is to compare the performance of MFCC (where 13 coefficients are used), and DTW. The speech data used in this experiment are the isolated words starting from vowel in Marathi, spoken by female speakers.The test pattern is compared with the reference pattern to get the best Match. The symmetric form of DTW algorithm is used to optimally align in time the test and reference patterns and

DOI: 01.IJIT.01.01.73

and to give average distance associated with optimal warping path. The recognition system uses the utterance of the spoken word For training and the remaining utterances for testing.

The Reference patterns are created for each word in the vocabulary from the data in the training set by averaging (after dynamic time warping) the patterns of utterances of the same word. The comparative recognition accuracy is presented in Table 3.

ACKNOWLEDGMENT

The authors would like to thank the university authorities for providing infrastructure to carry out experiment. This work has been supported by DST under Fast Track Scheme entitled as Design and Development of Marathi Speech Interface System

TABLE I
THE DISTANCE MATRIX OF TWO SUBJECTS USING MFCC

	Abhyas	Ajay	Akara	Amar	Ananas	Ati	Avidya
Abhyas	0.03	0.8098	0.9827	1.1824	1.2455	1.6888	0.5257
Ajay	0.3878	0.0839	1.0403	1.240	1.3031	1.7464	0.1114
Akara	0.5198	0.9996	0.0385	1.3722	1.4353	1.8786	0.7155
Amar	0.1813	0.5155	0.6844	0.1783	0.9471	1.3904	0.2274
Ananas	0.0943	0.5740	0.7470	0.9466	0.4530	1.4530	0.2899
Ati	0.2978	0.2197	0.3297	0.5923	1.0982	0.0346	0.0985
Avidya	0.3837	0.8635	1.0364	1.24	1.2991	1.7424	0.0702

TABLE II
THE DISTANCE MATRIX OF TWO SUBJECTS USING DTW

	Abhyas	Ajay	Akara	Amar	Ananas	Ati	Avidya
Abhyas	24.12	28.20	27.53	27.93	28.96	29.52	26.89
Ajay	24.21	24.01	34.04	23.07	30.25	28.68	27.67
Akara	22.65	27.30	20.32	24.91	26.77	23.39	21.78
Amar	29.19	24.70	28.43	24.04	25.78	29.20	32.12
Ananas	31.03	23.78	25.94	26.11	20.25	20.52	29.10
Ati	28.60	29.89	24.47	26.73	27.27	24.27	30.39
Avidya	24.42	24.83	25.76	27.40	25.89	30.11	22.49

TABLE III
COMPARATIVE RECOGNITION ACCURACY FOR MFCC AND DTW

Algorithm	Vector size	Percentage of variance	Recognition accuracy
MFCC	49	5.35	94.65
DTW	49	26.75	73.25

DOI: 01.IJIT.01.01.73

References

[1] (2010) The Wikipedia website [Online]. Available: http://en.wikipedia.org/wiki/List_of_languages_by_number_of_native _speakers. Viewed 10 Jan 2010.

[2] (2010)The Wikipedia website [Online]. Available: http://en.wikipedia.org/wiki/Marathi_language. Viewed 12 Jan 2010.

[3] Samudravijaya K, P.V.S. Rao and S.S. Agrawal, "Hindi Speech database", Proc. Int. Conf. Spoken language processing, ICSLP00,October Beijing 2000.

[4] F. Jelinek, L.R. Bahl, and R. L, Mercer, "Design of a linguistic statistical decoder for the recognition of continuous speech", IEEE Trans. Informat. Theory, vol. IT-21, PP. 250-250, 1975.

[5] Michael Grinm, Kristian Kroschel and Shrikanth Narayanan, "The Vera AM Mittag German Audio " Visual Emotional Speech Database.

[6] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, B. Weiss, "A database of German Emotion Speech",INTERSPEECH 2005, September, 4-8, Lisbon, Portugal

[7] The website for The Disordered Voice Database Available:http://www.kayelemetrics.com/Product%20Info/CSL%20Family/4500/4500.htm.

[8] [2010] The Wikipedia website[Online] Available: http://en.wikipedia.org/wiki/Devanagari

[9] Steven B. Davis and Paul Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken sentences", IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-28, No.4, August 1980.

[10] Qi, Li, Frank K, Soong and Olivier Siohan, "A High-Performance Auditory Feature for Robust Speech Recognition", 6th International Conferences on Spoken languages processing, Beijing, October 2000

[11] K. R. Aida "Zade, Ardil and S.S. Rustamov,"Investigation of Combined used of MFCC and LPC features in Speech Recognition Systems", PWASET VOLUME 13 MAY 2006 ISSN 1307-6884, Proceeding of world academy of science, engineering and technology volume.

[12] Hiromi Sakaguchi and Naoaki Kawaguchi, "Mathematical Modeling of Human Speech Processing Mechanism Based on the Principle of Bain Internal Model of Vocal Tract"Journal of the faculty of Engineering, Honshu University, No. 75,1995

[13] Steven W. Smith, "the Scientist and Engineer"s Guide to Digital Signal Processing", California Technical Publishing, 1977, page (s): 169-174

[14] Wei HAN, Cheong-Fat CHAN, Chiu-Sing CHOY and Kong-Pang PUN, "An efficient MFCC Extraction Method in Speech Recognition", 0-7803-9390-2/06/2006 IEEE

[15] K.K. PALIWAL, Anant AGARWAL and Sarvajit S. SINHA "Amodification over Sakoe and Chiba"s dynamic Time Warping Algorithm for isolated word recognition." Signal Processing 4.

DOI: 01.IJIT.01.01.73