P: ISSN No. 0976 - 8602 RNI No.UPENG/2012/42622 VOL.-IV, ISSUE-III, July-2015
E: ISSN No. 2349 - 9443
Asian Resonance

Voice Recognition with Aid of Word Processing

Jasdeep Kaur
Assistant Professor,
Deptt.of Physics,
A. S College,
Khanna, Punjab

Abstract

The areas where application of Automatic speech recognition (ASR) is evolve technologies are as follows: Controlling the programs, probe based Information system such as travel information system, automatic telephone call processing and Biometrics etc. Our goal is to study various aspects of voice recognition software that can recognize Punjabi words by make use of MFCC and LPC. This paper aims to deals with characteristics of voice recognition system for Punjabi language. The system is trained for continuous Punjabi speech; data has been taken from different male and female speakers.

Keywords : Voice Recognition, Mel-Frequency Cepstral Coefficients (MFCC), Linear Predictive Coding (LPC).

Introduction

Speech consists of multifarious sound signals produced by the individual vocal apparatus-an apparatus, which has the aptitude of producing a wide variety of speech sounds [1]. What is important for us in our study of speech is that acoustic signal is completely recognizable; we can confine everything the listener perceive sounds in the form of a recording and then measure whichever aspect of the signal that we want to know about [2].

Fundamentals of Automatic Speech Recognition Systems (ASR)

Speech interface in local languages is the prerequisite for speech interface. Two major steps that involved by ASR are: Feature Extraction and Classification. In First step, given speech is preprocessed and salient features with help of digital processing techniques are computed. Second step involved classification, done with help of machine learning techniques [3].

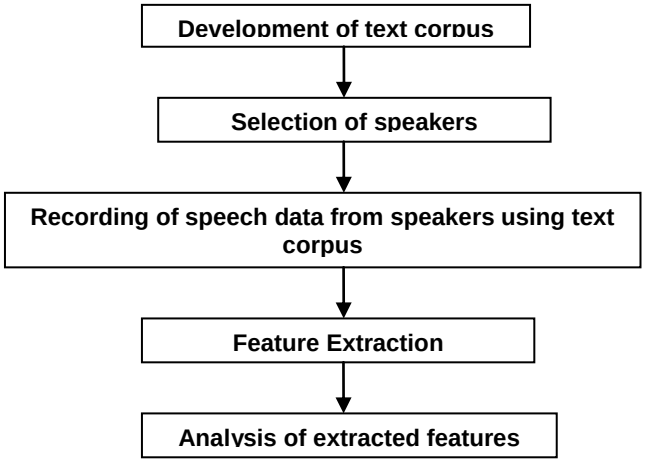
To convert an acoustic waveform into content equivalent to the information being revealed by the speaker, speech recognition system is used. Speech interface for local languages is the prerequisite for speech interface. Since, the syllables are very important unit of language. The syllable consists of vowels and consonants. The choice of representative units is made depending on the size of vocabulary. The smallest segmental units of sound are Phoneme. The next level basic unit of speech is syllables. Through a number of experimental studies, [4], [5] and [6] it can be said that articulatory and acoustic bases of various consonants are now fairly well understood, though the details of some specific aspects of individual language remain to be clarified. Speech recognition research work can be found for various Indian languages like Hindi [7], Kangri [8], Kannada [9], Tamil (10), Malayalam [11], Telugu (12) Bengali (13), Gujarati (14) Marathi (15) and Maithili (16). For Punjabi, not so much work was done so far. Few researchers [17], [18], [19] have worked on segmentation of speech into smallest units (syllable like). In another paper (20) on Punjabi language, authors studied speech synthesis architecture by make use of Hidden Markov Model (HMM). So, further documentation of Punjabi speech corpora is the need of hour.

Contemporary Part of Research: Introduction To Punjabi Vernacular

It is mandatory to spotlight language specific aspects of speech recognition. In present module, analysis of features of Punjabi language is under limelight by signal processing techniques. It further helps in recognition and understanding of words. The nature of Punjabi language is syllabic and it comprises 41 consonants (Vianjans). Punjabi language ensures seven groups of syllables: V, VC, CV, VCC, CVC, CCVC and CVCC

Methodology Adopted

Methodology of experiment involved following steps:



Implementation

Compilation Process of data: This stage gives description about steps involved to develop speech corpora.

A) Development of Text Corpus

Total of 50 words were recorded with help of PRAAT by 10 native speakers, both voiceless and voiced with abutted Punjabi vowel sounds to obtain 500 syllables.

B) Selection of Speakers

The data was collected from native speakers of Punjabi language. They were from different regions of Punjab. All speakers were classified on basis of gender.

C) Recording Ambiance

In this step, speech data was recorded using good quality microphones (Sennheiser PC 350) with help of PRAAT software.

Table 1: Detailed Information About Data Collection Procedure

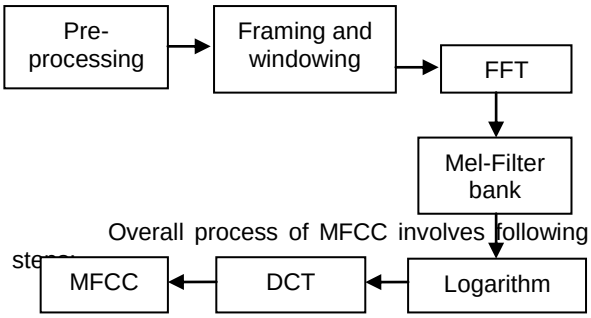
process	depiction
1. Speaker	Male = 5 Female =5
2. Analytic Tools	PRAAT
3. speech corpora	500 words
4. Utterance	Two utterances of each word
5.Sampling frequency	22050 Hz

D) Feature Extraction Techniques

Two most popular techniques used for extracting features from recorded speech .i.e. MFCC and LPC. Briefly description of these two techniques is as following:-

Mel Frequency Cepstral Coefficients (MFCC)

From part of filter bank analysis, MFCC are most widespread acoustic features found by Davis and Mermelstein (1980) [21] .If the original signal is being applied by more than one Fourier transform continuously, and then feature that is extracted is MFCC.



Overall process of MFCC involves following steps:

Pre-processing

This step involves passage of signal through a filter; consequently there will be enhancement in energy of signal at high frequency.

$$Y[n] = X[n]-0.95\ X\ [n-1]$$

Framing and Windowing

In it, framing (breaking signal with size into small chunks) is taken. Generally for speech recognition, frame size is in between 20ms to 40ms used, we have used size of 30ms.Windowing (multiply by a hamming function) is used as integrator for all closest frequency lines. Equation for hamming window is as

$$Y[n] = X[n] \times w[n];$$

Where $W[n] = 0.54-0.46\cos(2\Pi n/ N- 1)$

N = No. of samples in each frame.

Fast Fourier Transform (FFT)

Frame of N samples get converted from Time domain to Frequency Domain. The equation corresponds to following statements:

$$Y(w) = FFT[h(t) \times (t)] = H(w) \times X(w)$$

Mel-filter bank

The filter function involves three parameters: lower frequency f_1, central frequency f_c and higher frequency f_h [22]. These higher and lower values of frequency are converted into Mel scale by following formula, which include conversion between a frequency value in hertz (f) and Mel as:- $Mel(f) = 2595 \log10(1+ f/700)$

Discrete Cosine Transform (DCT)

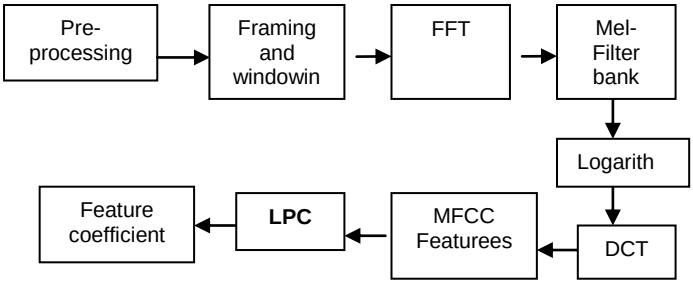
Through DCT log Mel spectrum may convert into time domain. Result of such conversion will give us Mel frequency cestrum coefficient (MFCC). Different sets of coefficients known as Acoustic vectors.

So, each input utterance transformed into a sequence of acoustic vectors.

Linear Predictive Coding (LPC)

It is a tool which is used mostly by researchers for speech synthesis and speech analysis. It gives analysis for abort speech signal by measuring values of Formants, Intensity and frequency [23]. It also smoothens out the required signal. In it, each word is expressed as linear combination of previous sample, so called linear and therefore known as “Linear predictive coding”. Functioning of LPC is as follows:-

P: ISSN No. 0976 - 8602 RNI No.UPENG/2012/42622 VOL.-IV, ISSUE-III, July-2015
E: ISSN No. 2349 - 9443
Asian Resonance



Results and Discussion

The input voice signal of word "Yaad" is as follows (Figure 1):

This original signal gets transformed into Mel filter form by applying the algorithm (Discussed as above). Then the pictorial representation (Figure 2) of signal will be of this form:

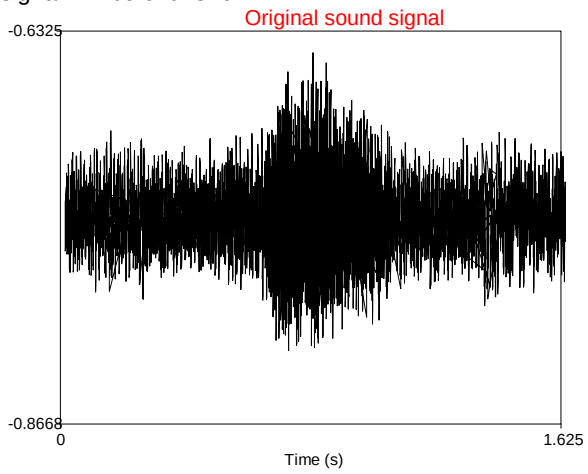


Figure 1. Original sound signal

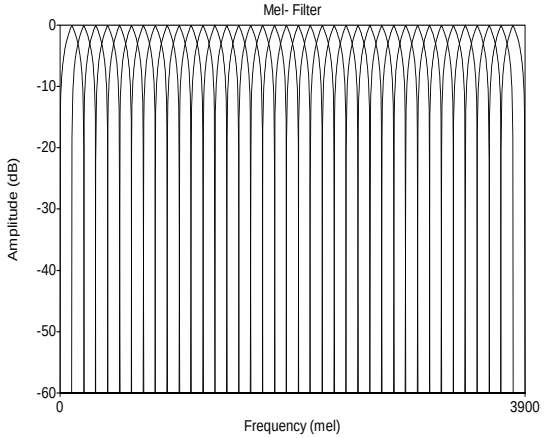


Figure 2. Mel- Filter Bank

After it, Mel- Frequency cepstral coefficients (MFCC) were calculated as the given relation. MFCC had following shape (Figure 3)

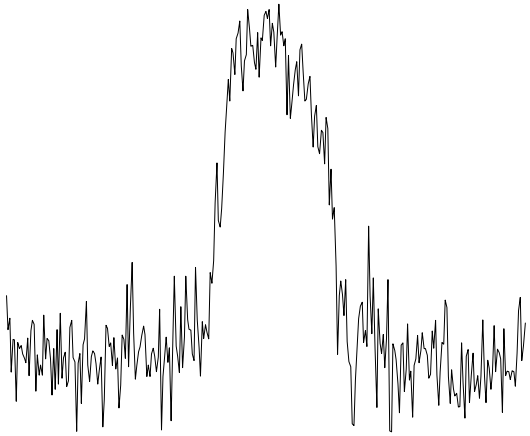


Figure. 3. MFCC

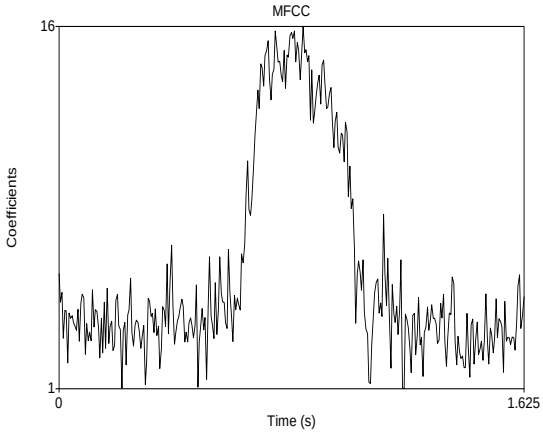


Figure 4. MFCC (16 Coefficients)

The MFCC can be calculated with 12 as well as 16 coefficients. The graph of 16 coefficients Vs. Time(s) had following form (Figure 4).

Next step to obtain LPC, Then by algorithm (above), LPC contour is as following (Figure 5)

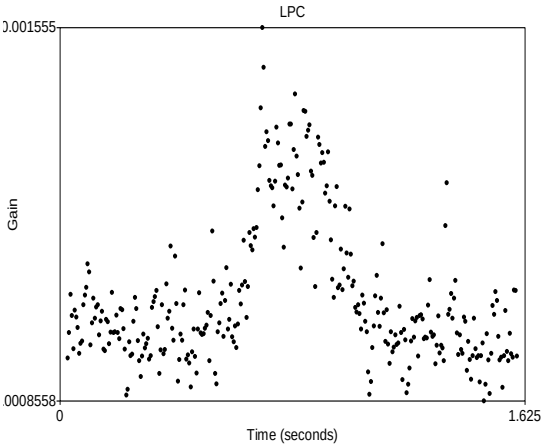


Figure 5: LPC

P: ISSN No. 0976 - 8602 RNI No.UPENG/2012/42622 VOL.-IV, ISSUE-III, July-2015
E: ISSN No. 2349 - 9443
Asian Resonance

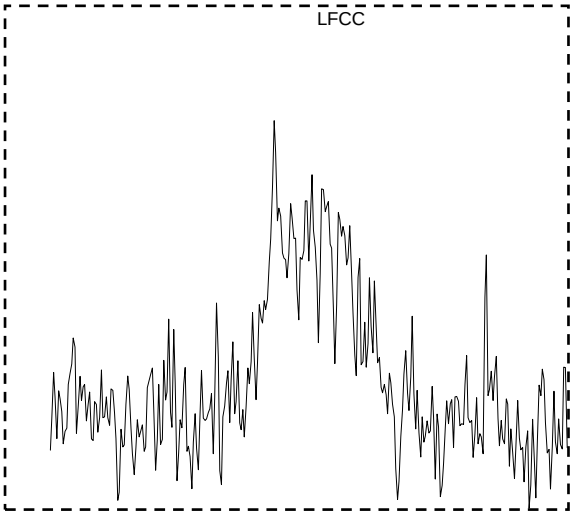


Figure 6: LFCC

An object of type LFCC (figure 6) represents Cepstral coefficients on a linear frequency scale as a function of time. When these coefficients (12 in no.) are represented in frames with constant sampling period, following form (figure 7) will appear:

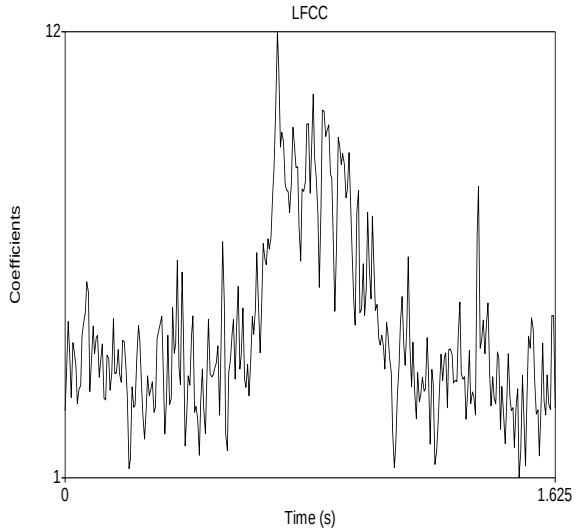


Figure 7: LFCC vs. Time(s)

Table 2. Voice Report of Three Different Vowels

Vowels	pitch	Time range	Pulses	Harmonicity	MFCC
L(o) ck	Median pitch: **107.357 Hz** Mean pitch: **99.966 Hz** Standard deviation: **12.834 Hz** Minimum pitch: **72.609 Hz** Maximum pitch: **118.566 Hz**	From 1.109928 to 1.338780 seconds (duration: **0.228852 seconds)**	Number of pulses: **23** Number of periods: **22** Mean period: **9.956293 seconds** Standard deviation of period: **1.375201 seconds**	Mean autocorrelation: **0.883162** Mean noise-to-harmonics ratio: **0.135210** Mean harmonics-to-noise ratio: **9.069 dB**	Frame 1 C1 = 230.996325 C2 = 72.378997 C3 = 102.882929 C4 = 80.6576837 C5 = 94.167461 C6 = 63.1087144 C7 = 87.9047853 C8 = 64.4112552 C9 = 91.3817123 C10 = 37.5277151 C11 = -7.59332954 C12 = 16.3956819
Y(aa)d	Median pitch: **83.482 Hz** Mean pitch: **95.399 Hz** Standard deviation: **13.826 Hz** Minimum pitch: **75.520 Hz** Maximum pitch: **122.012 Hz**	From 0.938660 to 1.209860 seconds (duration: **0.271201 seconds)**	Number of pulses: **25** Number of periods: **24** Mean period: **10.552073 seconds** Standard deviation of period: **1.530205 seconds**	Mean autocorrelation: **0.889465** Mean noise-to-harmonics ratio: **0.127695** Mean harmonics-to-noise ratio: **9.485 dB**	Frame 1 C1 = 254.717929 C2 = 58.4834733 C3 = 65.7884198 C4 = 69.3299962 C5 = 109.374965 C6 = 87.4342877 C7 = 67.5971689 C8 = 60.0269596 C9 = 96.2047905 C10 = 60.1870632 C11 = 14.1875432 C12 = 8.55698759
R(ee)t	Median pitch: **107.109 Hz** Mean pitch: **104.349 Hz** Standard deviation: **9.263 Hz** Minimum pitch: **73.684 Hz** Maximum pitch: **116.077 Hz**	From 0.401699 to 0.849012 seconds (duration: **0.447313 seconds)**	Number of pulses: **47** Number of periods: **45** Mean period: **9.601312 seconds** Standard deviation of period: **0.963873 seconds**	Mean autocorrelation: **0.833750** Mean noise-to-harmonics ratio: **0.244341** Mean harmonics-to-noise ratio: **8.124 dB**	Frame 1 C1 = 204.752876 C2 = 89.2392887 C3 = 117.763348 C4 = 94.2504543 C5 = 119.307957 C6 = 65.4745078 C7 = 84.2435437 C8 = 52.4509342 C9 = 68.6283114 C10 = 32.8710504 C11 = 7.60409447 C12 = 10.9275941

Conclusion:

The present paper has discussed the algorithms of two techniques MFCC and LPC, which are important for improvement in voice recognition performance. This technique enables to authenticate particular speaker that is based on the individual information from voice reports. MFCCs are used in systems which can automatically recognize numbers spoken into a telephone and also in music information retrieval.The table show values of pitch, pulses, Harmonicity and MFCC for three Punjabi vowels. These suprasegmental parameters could use effectively for recognition purpose in ASR Several other techniques such as Hidden Markov model (HMM), Artificial Neural Network (ANN) are currently being investigated using MFCC. The findings will be presented in future publications.

References

1. Ladefoged. P., Vowel and Consonants-An Introduction to Sounds of Languages, Wiley-Blackwell, (2001), ISBN 0-631-21412-7, 1-2.
2. Kumar. K, Aggarwal. R.K, A Hindi speech recognition system for connected words using HTK, Int. J. Computational Systems Engineering, 1,(2012).
3. Dua. M, Kaur. P, Saini. P, Hindi Automatic Speech Recognition Using HTK, International Journal of Engineering Trends and Technology (IJETT), 2013, 4, 2223.
4. Diehl. R. L, The role of Phonetics within the study of language, Phonetica, 1991, 48, 120-134.
5. Honda. M, Human Speech Production Mechanisms, Selected Papers, 2003, 1(2), 24-29.
6. Keating. P. A, Phonetic representations in a generative grammar, Journal of phonetics, 1990, 18, 321- 334.
7. Choudhary. A, Chauhan, R.S, Gupta, G, Automatic Speech Recognition System for Isolated & Connected Words of Hindi Language by Using Hidden Markov Model Toolkit (HTK).
8. Eaton.R, Kangri in context: An areal perspective, Thesis submitted to University of Texas at Arlington.
9. Hegde. S, Achary. K. K, Shetty.S, Statistical analysis of features and classification of alphasyllabary sounds in Kannada language, Int J speech technol, 2014.
10. Thangarajan. R, Natarajan.A.M, Selyam.M, Syllable modeling in continuous speech recognition for Tamil language, International Journal of speech technology, 2009, 12(1),47-57.
11. Soavithri.S.R, Jayaram.M, Rajasudhakar.R, Venugopal .M.B, A comparative study of base of articulation in Dravidian and indo- Aryan languages, J. Acous.Soc.Ind. 33(2005).
12. Bhaskar.P.V, Mohan .Rao.R, Gopi.A, HTK Based Telugu Speech recognition, International Journal of Advanced research in Computer Science and Software Engineering (ijarcsse) ,2012,2(12).
13. Das.B, Mandal.S, Mitra.P, Basu. A, Effect of aging on speech features and phonemic recognition: A study of Bengali voicing vowels, Int J speech technol, 2013, 16, 19-31.
14. Vuppala. A.K, Bhaskararao.P, Automatic detection of breathy voiced vowels in Gujarati speech, Int J speech technol, 2014, 17, 75-82.
15. Nathoosing.K.C, Isolated Word recognition for Marathi language using VQ and HMM, Science research reporter, 2012, 2(2), 161-165.
16. Yadav. R, (1980), The influence of aspiration on Vowel duration in Maithili, CNAS, 7 (1 and 2).
17. Dua.M, Aggarwal.R.K, Kadyan.V, Dua.S, Punjabi automatic speech recognition using HTK, IJCSI, 2012,9(4).
18. K.Amanpreet, S.Tarandeep, Segmentation of continuous Punjabi speech signal into syllables, in the proceedings of the world congress on engineering and computer science 2010 (WCECS 10), vol 1.
19. Kumar.R, Comparison of HMM and DTW for Isolated word recognition of Punjabi language, in proceedings of progress in pattern recognition, Image analysis, Computer vision and applications, springer Verlag, 2010, 6419, 244-252.
20. Bansal.D, Goel.A, Jindal.K, Punjabi speech synthesis system using HTK, International journal of information sciences and techniques (IJIST), 2(4),2012.
21. Davis.S.B, Mermelstein, Comparison of parametric representation for monosyllabic word recognition in continuously spoken sentences, IEEE Trans on ASSP, 1980, 28, 357-366.
22. Rabiner.R, Juang.B.H, Fundamentals of Speech Recognition, Prentice-Hall International, New Jersey, 1993.
23. Patil.H.A, Basu.T.K, Development of speech corpora for speaker recognition research and evaluation in Indian languages, Int J Speech Technol,11(2008) ,17-32.