IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 11, Issue 5, Ver. I (Sep.-Oct .2016), PP 124-129
www.iosrjournals.org

Frequency Analysis Of Speech Signals For Devanagari Script And Numerals

Vidhu Gupta[1], Neeraj Jain[2]
(Electronics & Communication Department, Modern Institute of Technology and Research Centre)
(Rajasthan Technical University Kota, INDIA)

Abstract: *This paper comprehends the analysis of frequency of spoken Devnagari script and Numerals from the primitive speech signals. Devnagari vowels and numerals play essential role in pronunciation of any word or counting. Particular vowel & number is dividing as starting, middle and end according to the time of occurrences in the word. The Devnagari script contains 12-vowels and 34-consonants and they are pre-owned in some Indian language like Hindi and 10 numerals (0-9) are worn in mathematics. Sound samples from multiple speakers were applied to obtain different features. Starting processing of data, that is normalizing and time-slicing was done using an amalgamation of Simulink and MATLAB. Later on, the same engines were used for calculation of fourier descriptions and correlations. The correlation permits comparison of the identical words or numeral uttered by the identical and different speakers. Hence the frequency has been calculated in statistical manner and produced a table among amplitude and frequencies. Standard deviation and mean can also be used in implementation of a voice-driven help setup at call centres of commercial organizations running in India and other foreign areas.*

Keywords*: Frequency Analysis Devnagari script, Fourier transforms Correlation, Feature extraction*

I. Introduction

Frequency analysis is used in many areas of research like as speech formation, speech analysis, speech recognition, speaker identification etc. The Devnagari vowels and numerals cannot be uttered in two forms as like in English language where phoneme and grapheme are different in most cases whereas in Devnagari it can be pronounced as one way only e.g. devnagari 12-vowels are divide with the phonetic transcription structure of phonemes according to organ used in generating the sound. Devnagari is placed on phonetics principles which are taken as Place of articulation (POA) vowels. These Devnagari vowels have analysis of frequency for speech signals are calculated in noisy environment (original signals) for analysis and synthesis [1-3]. The primitive speech signals are unbalanced to balance of an interval with the help of some feature extraction techniques. The basic objective is to estimating the pitch of Devnagari vowels and numerals with noisy environments speech signals.

When one glimpse at a person, car or house, one's brain tries to match the incoming frame with hundreds (or thousands) of frame that are previously stored in memory. In the speech recognition research literature, no work has been recorded on Devnagari speech processing and numerals. So we acknowledge our work to be the initial move in this direction. The process includes extraction of some unique characteristics of each word by using Fourier transforms and their correlations. The system is speaker-independent and is fairly tolerant to background noise.

II. Vowels In Devnagari Script

The 12-Devnagari vowels are classified as per IPA (International Phonetics Association) as shown in Table 1 & hindi character set in Table 2. They are used for the speech analysis and synthesis purpose. It describes in different categories such as follows:

(i) SHORT VOWELS

The short vowel is a individual vowel in a short word or syllable, that vowel basically makes a short sound. These short vowels generally occur at the starting of the word or among two consonants. E.g. the short vowels present character in Marathi and in Hindi.

(ii) LONG VOWELS

The long vowels a short word or syllable terminates with a vowel-consonant. The long vowels generally are an amalgamation of words in Devnagari. The `a' at the terminal of the word is mute. Long vowels when the word or syllable has a individual vowel and the vowel present at the end of the word or syllable, the vowels usually present makes the long sound in Hindi.

(iii) CONJUNCT VOWELS
The conjunct vowels are joining of short and long vowels. These phonemes are generated in Hindi e.g. as shown in Table 2.

(iv) NASAL VOWEL
A nasal vowel is yield with a low tune so that air pressures through nose as well as mouth. The title "nasal" is few air pressure which does not come uniquely out of the nose in nasal vowels.

(v) VISARG VOWEL
The Visarg symbol is used occasionally in Devnagari. The Visarg is uttered as the voiceless sound after the vowels. E.g. .in hindi.

Table 1. Devnagari Vowels Classified Into Five Types

TYPE OF DEVNAGARI VOWELS	1	2	3	4
SHORT	अ	इ	उ	-
LONG	आ	ई	ऊ	-
CONJUN-CT	अ+इ=ए	अ+ई=ऐ	अ+उ=ओ	अ+उ=औ
NASAL	अं	-	-	-
VISARG	अः		-	-

Table 2. Hindi Character Set

Vowels	अ	आ	इ	ई	उ	ऊ	ऋ	ए	ऐ	ओ	औ	अं	अः
	a	ā	i	ī	u	ū	r	e	ai	o	au	ań	ah
Gutturals (कवर्ग)	क	ख	ग	घ	ङ								
	ka	kha	ga	gha	ṅa								
Palatals (चवर्ग)	च	छ	ज	झ	ञ								
	ca	cha	ja	jha	ña								
Cerebrals (टवर्ग)	ट	ठ	ड	ढ	ण								
	ṭa	ṭha	da	dha	ṇa								
Dentals (तवर्ग)	त	थ	द	ध	न								
	ta	tha	da	dha	na								
Labials (पवर्ग)	प	फ	ब	भ	म								
	pa	pha	ba	bha	ma								
Semi-Vowels	य	र	ल	व									
	ya	ra	la	va									
Sibilants	श	ष	स										
	sa	pa	sa										
Aspirate	ह												
	Ha												

III. Modeling Of Speech By Applying Average Energy In Zero Crossing

The speech generating model advises that the energy of the voiced speech is pointing about 8 kHz, where as in the case of unvoiced speech, most of the energy is present at the higher Frequencies. As high frequency gives high zero crossing rate and low frequency gives low zero crossing rate, there is a solid correlation between zero crossing rate and energy distribution with frequency [4-5]. This inspires us to model the speech signal utilizing average energy in zero crossing interval of the signal. Take the speech segment

shown in Figure 1. The Z C_i^K views the ith zero crossing and Z C_{i+1}^K displays the(i+1)th zero crossing of kth observation window. The time interval among these two points is known as ith zero crossing intervals T_i^K in the kth observation window.

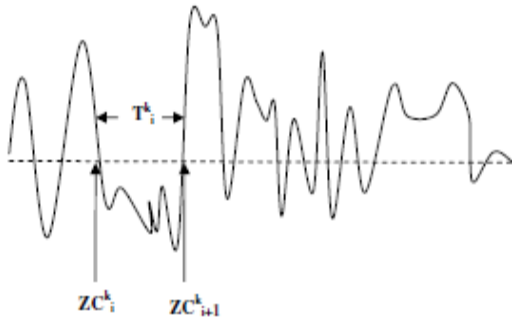

Figure1. Speech segment in kth observation

The average energy in the ith zero crossing intervals can be get by the expression

$$E_t^k = \frac{1}{T_t^k} \int_{Zc_i^k}^{Zc_{i+1}^k} X^2(t)dt$$

E_i^K Is the average energy of the signal in T_i^K th zero crossing interval of kth observation window and x(t) is the instantaneous signal amplitude. The goal of the current study is to catch a vigorous coefficient for speech identification application by applying the average energy in the zero crossing intervals (AEZI). An XY plot is created by plotting index number of zero crossing intervals parallel with X axis and Average Energy in the Zero crosses Interval (AEZI) along Y axis. Figure 1 represents the average energy in the zero crossing interval vs index number of the zero crossing intervals for the Hindi script.

IV. Acquisition Of Data And Processing

One of the considered methods of speech data acquisition is to have a person speak into an audio device such as microphone or telephone. This way of speaking generates a sound pressure wave that makes an acoustic signal. The microphone or telephone perceives the acoustic signal and changes it into an analog signal that can be learnt by an electronic system. Lastly, in order to retain the analog signal on a computer, it must be changed into a digital signal.

The statistics in this paper is obtained by speaking Hindi Word and numeral into a microphone joined to Windows-7 based PC. The data is saved into '.wav' format files by the use of MATLAB. The wave files are processed after passing through a (Simulink) filter, and are saved for future analysis like as FFT. The data is saved form speakers who spoke the identical word set, i.e. Devnagari Script & numerals. In common, the computerized speech waveform has a high dynamic range, and may suffer from additional noise [7-9]. So first, a Simulink model was used to extract and analyze the present data as shown in Fig. 2.

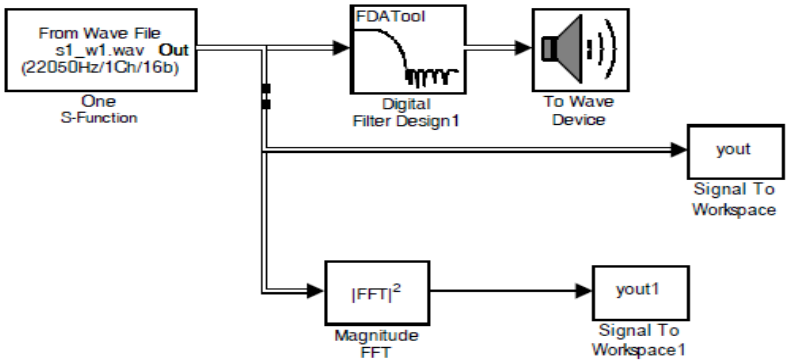

Figure2. Simulink model for analyzing hindi data

The Simulink model was designed for performing analysis like as standard deviation, mean, autocorrelation, magnitude of FFT, data matrix correlation. We will also experiment some other statistical techniques.

This repetitive and changing-nature of models ultimately take us to MATLAB's (text-based) .m files. These files are generated semi-automatically by applying Hindi-language script; the script was designed particularly for this agenda. Three important data pre-processing steps were needed before the data could be consumed for analysis.

(i) PRE-EMPHASIS

By pre-emphasis, we use the application of a normalization technique, which is done by parting the speech data vector by its peak magnitude.

(ii) LENGTH ADJUSTMENT OF DATA

FFT execution time lay on specific number of the samples (N) in the data sequence [xK], and that the execution time is minimal and proportional to N*log2 (N), where N is defined as power of two. Hence, it is generally useful to select the data length comparable to a power of two.

(iii) DETECTION OF END POINT

The aim of endpoint detection is to discriminate the word to be detected from the background noise. It is essential to trim the word pronounced to its tightest limits, in a way to avoid errors in the modeling of subsequent pronunciations of the same word. As shown in Fig. 2 from upper part, a threshold has been enforced at both ends of the waveform. The leading threshold is assigned to a value that all the spoken numbers trim to a maximum value. These values were find out after observing the nature of the waveform and noise in a specific environment. We can see the discrimination in frequency characteristics of the words.

(iv) FOURIER TRANSFORM

The MATLAB algorithm for the two dimensional FFT routine is given as: fft2(x) =fft(fft(x)); Hence the two dimensional FFT is calculated by first calculating the FFT of x, i.e the FFT of every column of x, and then calculating the FFT of every row of the result. Note that as the application of fft2 command generates even symmetric data, we only displays the lower half of the frequency spectrum in our intended graphs.

(v) CORRELATION

Computation for correlation coefficients of various speakers was performed. As convention, the cross-correlation of the identical speaker for the identical word did obtain to be 1. The correlation matrix of a spoken number was produced in a three-dimensional form for producing various simulations and graphs.

V. Results & Analysis

It is observed that Fourier descriptor trait was independent for the spoken Devnagari Script and numerals with the consolidation of the Fourier transform and correlation method commands applied in MATLAB, a high accuracy recognition structure can be realized. Recorded data was applied in Simulink model for initiatory analysis.

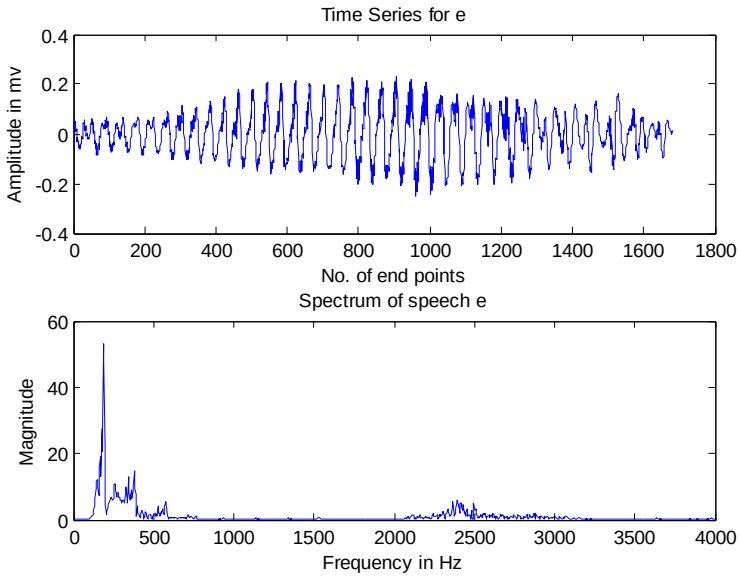

Figure 3. The fft waveform of the word ष in devnagari script

Figure 4 shows X = 1700 ,It has 1700 numbers of data points. It is given by X. and have 5 peaks values for each & every word identical for WORD ङ in Devnagari script.

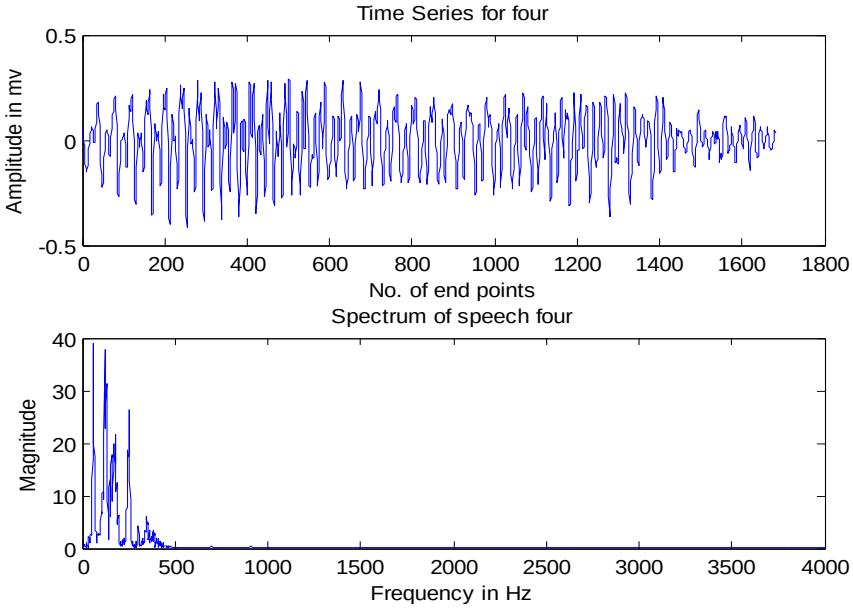

FIGURE 4. The Fft Waveform Of The Four In Numerals

Figure 5 shows X=1670It has 1670 numbers of data points. It is presented by X. and have a 5 peaks values for each & every word identical for FOUR in Numerals

TABLE.3. Peaks And Its Corresponding Frequencies

SR.NO	SPEECH WORD	PEAK		FREQUENCY IN (HZ)	
1	*FOR E*	P1	8.5455	F1	194
		P2	7.6778	F2	188
		P3	6.3280	F3	235
		P4	6.810	F4	182
		P5	6.810	F5	200
2	*FOR 4*	*P1*	*15.7404*	F1	487
		P2	*10.6615*	F2	481
		P3	*9.2795*	F3	474
		P4	*8.3985*	F4	327
		P5	*7.0171*	F5	467

VI. Conclusion And Future Work

An effective, abstract and fast ASR system for regional languages like Hindi is demand of the present. The work presented in the paper is a step towards the expansion of such type of systems. The work can further be increased to huge vocabulary size and to continuous speech recognition. As presented in results, the system is prone to changing spoken techniques and dynamic scenarios, so the accuracy of the system is a vindictive field to work upon. Hence, different Speech enhancements and noise reduction methods can be applied for making system more effective, appropriate and fast.

References

[1]. D. O'Shaughnessy, "*Interacting with Computers by Voice-Automatic Speech Recognitions and Synthesis"*, (Invited Paper), Proceedings of the IEEE, *Vol. 91, No. 9, 2003*, pp. 1272-1305.

[2]. Jun Cai, Ghazi Bouselmi, Yves Laprie, and Jean- Paul Haton, *"Efficient Likelihood Evaluation and Dynamic Gaussian Selection for HMM-Based Speech Recognition*", Computer Speech and Language, *Vol.23, 2009*, pp.147–164

[3]. Ingrid Verbauwhede; Patrick Schaumont; Christian Piguet; Bart Kienhuis (2005-12-24). "Architectures and Design techniques for energy efficient embedded DSP and multimedia processing" (PDF). rijndael.ece.vt.edu. Retrieved 2014

[4]. B. H.; Rabiner, Lawrence R. *"Automatic speech recognition–a brief history of the technology development" : Retrieved 17 January 2015.*

[5]. Richard Salomon *(2014), Indian Epigraphy, Oxford University Press, ISBN 978-0195356663, page 71.*

[6]. Speech Communication Technical Committee. *"Speech Communication"*. Acoustical Society of America. Retrieved 22 May 2013.

[7]. M. Habibullah Pagarkar, Lakshmi Gopalakrishnan*, et.al. "Language Independent Speech Compression using Devnagari Phonetics", 2002.*

[8]. S K Hasnain, Perez Akhter, "*Digital Signal Processing, Theory and Worked Examples*", January 2007.
[9]. Samuel D Stearns, Ruth A David, "*Signal Processing Algorithms in MATLAB*,"(Prentice Hall, 1996).
[10]. Redford, M. A. (2015). The *handbook of speech production." Chichester, West Sussex ; Malden, MA* ": John Wiley & Sons, Ltd, 2015.
[11]. C.H. Lee, J. L. Gauvain, R. Pieraccini, and L. R. Rabiner, *"Large Vocabulary Speech Recognition using Subword Units*", Speech Communication, Vol.13, 1993, pp. 263-279
[12]. S. Molau, F. Hilger and H. Ney, "*Feature Space Normalization in Adverse Acoustic Conditions*", Proc. of ICASSP, 2003, pp. 656-659.
[13]. *Fromkin, Victoria; Berstien Ratner, Nan. Chapter 7 Speech Production. Harcourt Brace College. pp. 322–327*
[14]. S .K Hasnain, Nighat Jamil, "*Implementation of Digital Signal Processing real time Concepts Using Code Composer Studio 3.1, TI DSK TMS 320C6713 and DSP Simulink Blocksets*," IC-4 conference, Indian Navy Engineering College, Goa, Nov. 2007