

Analysis of Various Parameters in Speech Signal

Balaji.B¹, Hari Prasanna.A², Sathish Kumar.V³, Vinodh Kumar.M⁴, Chidambaram.S⁵

UG Scholars, Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India^{1,2,3,4}

Asst. Professor, Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India⁵

ABSTRACT: This paper contains analysis of different speech signal such as identifying the voiced/unvoiced/silence regions of speech from their time domain and frequency domain representations. Analyse the Non-stationary nature of speech signal for single tone and multi tone synthesis operations. Identification of different sounds of a language, the language should be alphabets sounds (Vowels & consonants, short vowels, long vowels & diphthongs, stop consonants, fricatives, Affricates, Nasal, and Semivowels). In speech processing we can measure time domain, frequency domain and time-frequency representations of an alphabets.

KEYWORDS: Sampling, pitch frequency, Discrete Fourier Transform (DFT), Non-stationary, Diphthongs, Affricates.

I. INTRODUCTION

Speech is an acoustic signal produced from a speech production system. From our understanding of signals and systems, the system characteristic depends on the design of the system. For the case of linear time invariant system, this is completely characterized in terms of its impulse response. However, the nature of response depends on the type of input excitation to the system. For instance, we have impulse response, step response, sinusoidal response and so on for a given system. Each of these output responses are used to understand the behavior of the system under different conditions. A similar phenomenon happens in the production of speech also. Based on the input excitation phenomenon, the speech production can be broadly categorized into three activities. The first case where the input excitation is nearly periodic in nature, the second case where the input excitation is random noise-like in nature and third case where there is no excitation to the system.

Signal is a physical quantity that is measurable. System is a physical entity that exists. Signal is produced from a system. Depending on the nature of signal, it is categorized into several classes based on some criterion. Some of the classification includes continuous v/s discrete, periodic v/s aperiodic, energy v/s power and deterministic v/s random, stationary v/s non-stationary and so on. In digital signal processing, emphasis was not provided on the stationary v/s non-stationary classification of signals. Speech signal processing deviates in this aspect. This is because, speech is an example for non-stationary signal where as conventional synthetic signals like sine wave, triangular wave, square wave and so on are stationary in nature.

Speech generated from the speech production system consists of a sequence of basic sound units of a particular language. The need for studying the basic alphabet set (orthographic representation) of any language is to be able to express message in written form. On the similar lines we need to study the basic sound units set (acoustic representation) of any language for producing message in oral form. Every language is provided with unique set of alphabet set and sound units set. In most of the Indian languages we have about 40-50 distinct alphabets set and also of nearly same number of sound units set.

II. RELATED WORK

The system responds to the input signal/ excitation and produces output signal/ response. For given design of the system, the output response depends on the type of input excitation. Accordingly, we can have different output responses. For the case of speech, it can be modified. The speech production system responds to the input excitation by producing speech signal.



Figure 1: block diagram representing relation between signal and system

The schematic of human speech production mechanism is shown in fig 2. The speech production organs include lungs, trachea, glottis, pharynx, oral cavity and nasal cavity. The lungs supply the required air during exhalation for producing speech. Trachea also termed as wind pipe connects the lungs to the glottis. The glottis consists of two thin membranes termed as vocal folds or chords and obstructs airflow during specific category speech to generate the required excitation signal for speech production. The organs above glottis constitute the system part for speech production.

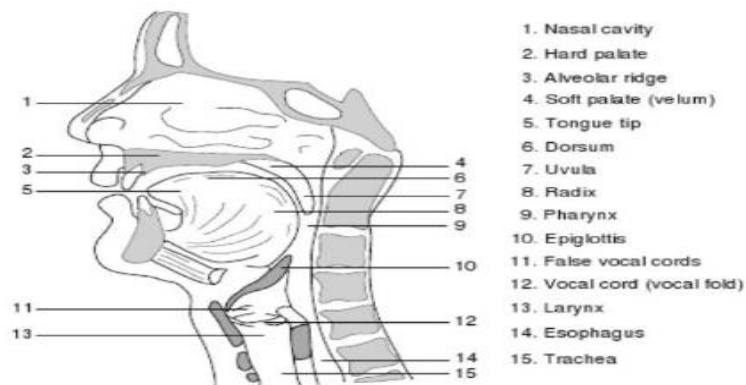


Fig 2: Schematic diagram of human speech production system

A signal is said to be stationary if its frequency or spectral contents are not changing with respect to time. This is very important point, pause a while and try to understand. This is because when we generate a sine wave using either a function generator or software, we selected the frequency value and kept it constant forever. Thus the frequency content of the sine wave will not change with time and hence is an example for stationary signal.

The important steps in speech processing are to get feel about the different sounds used for speech production. From signal processing point of view we need to get a feel of the time domain, frequency domain and time-frequency representations of these sounds. Perceptually we have been exposed to different sounds of mother tongue day in and day out and we agree that we can discriminate them based on their perceptual difference.

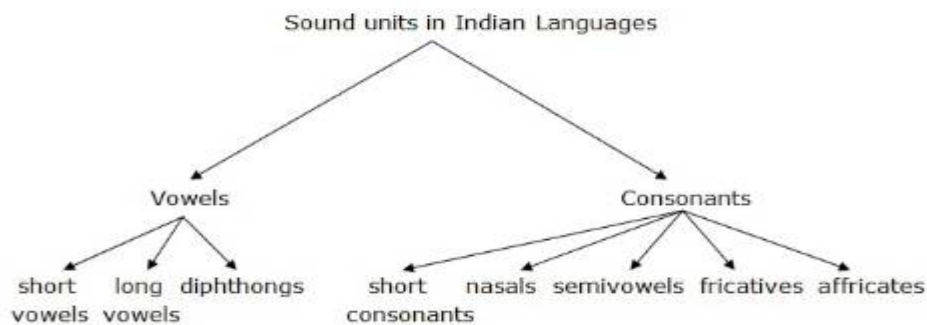


Fig 3: Classification of sound units in Indian languages

III. EXPERIMENTAL RESULTS

Figures 4 shows the results of voiced and its magnitude spectrum, there frequency components repeating at regular intervals indicating the presence of harmonic structure. In the frequency domain, the presence of this harmonic structure is the main distinguishing factor for voiced speech.

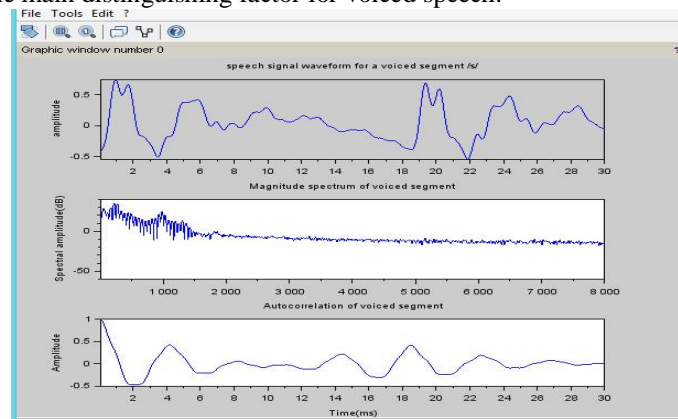


FIG 4: VOICED SEGMENT SPEECH AND ITS LOG MAGNITUDE SPECTRUM

The frequency domain, the absence of this harmonic structure is the main distinguishing factor for unvoiced speech in fig

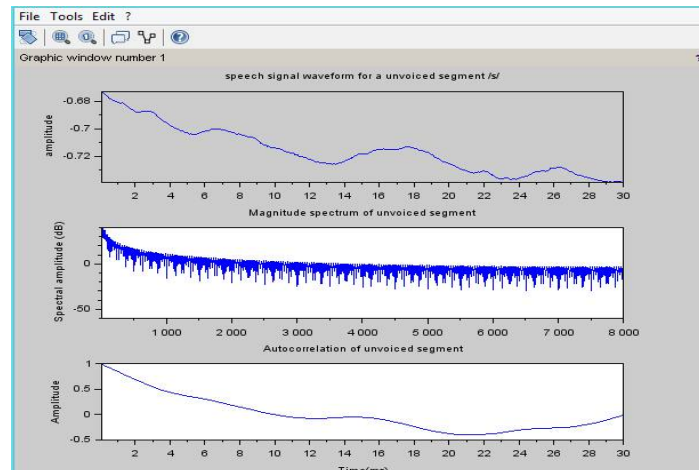


Fig 5: Unvoiced segment speech and its log magnitude spectrum

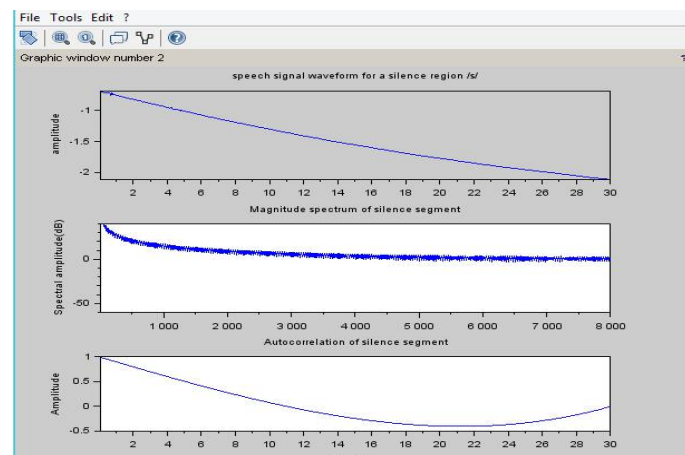


Fig 6: Silence region

The frequency contents of the speech signal will have many frequency contents and these components change continuously with time. For example consider a speech signal for the Hindi word SAKSHAAT shown in fig

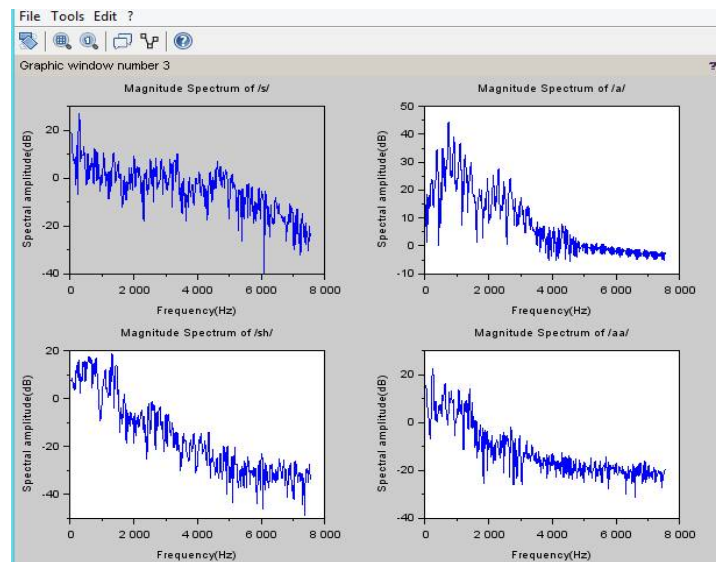


Fig 7: Magnitude spectrum of non-stationary signal

The voiced speech segment is characterized by the periodic nature, relatively high energy, less number of zero crossing and more correlation among successive samples. The unvoiced speech segment is characterized by the non-periodic and noise-like nature, relatively low energy compared to voiced speech, more number of zero crossings and relatively less correlation among successive samples. The silence region is characterized by the absence of any signal characteristics, lowest energy compared to unvoiced and voiced speech segments, relatively more number of zero crossing compared to unvoiced and no correlation among successive samples. The spectrum of the non-stationary signal will be meaningful if it is computed over regions that can be treated as stationary.

In the languages sounds vowels and consonants segments should be a fricative sound /s/ obtained from the syllable-like unit /sa/ and a long vowel sound /A/. The waveform of the short vowels, long vowels and diphthongs varying spectra of short vowel /a/, long vowel /A/, short vowel /i/ and diphthong /ai/. In stop consonants, velar consonants /k/, /kh/, /g/, /gh/ respectively. The varying spectra of segment fricative sound should be /sh/ and fricative sounds include /s/, /shh/ and /h/. A variation of the signal should be a /ch/, /chh/, /j/ and /jh/. In Indian languages the nasal sounds should be a /ng/, /nj/, /N/, /n/ and /m/. The magnitude spectrum of the signal should be varied. In semivowels in the Indian languages are /y/, /r/, /l/ and /w/, the waveform of the magnitude spectrum and time varying spectra of the semivowels.

REFERENCES

- [1] iitg.vlab.co.in., (2011). Identification of voice/unvoiced/silence regions of speech
- [2] iitg.vlab.co.in., (2011). Non-Stationary Nature of Speech Signal.
- [3] iitg.vlab.co.in., (2011). Different Sounds In language. Retrived 19
- [4] D. Gabor, W.P.L. Wilby, and R. Woodcock, "An universal non-linear filter, predictor and simulator which optimizes itself by a learning process," Proc Inst. Elec. Eng., vol. 108, pp. 422-438, 1961.
- [5] L. Li and S. Haykin, "A cascaded recurrent neural network for real time nonlinear adaptive filtering," in Proc. International Conference on Neural Networks' adaptive filtering," in Proc. International Conference on Neural Networks (San Francisco), 1993, pp. 857-862.
- [6] R.J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," Neural Computer, vol. 1, pp.270-280.
- [7] L.E. McBride, Jr. and K.S. Narendra, "Optimization of time-varying systems," IEEE Trans. Automat. Contr., vol. 10, pp. 289-294.

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization

Volume 5, Special Issue 2, March 2016

4th National Conference on Frontiers in Communication and Signal Processing Systems (NCFCSPPS '16)

10th-11th March 2016

Organized by

Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India.

- [8] D.C. Van Essen, C.H. Anderson, and D.J. Felleman, "Information processing in the primate visual system: An integrated systems perspective," Sci., vol.1 pp. 419-423.
- [9] J.C. Houk, " Learning in modular networks," in Proc. Int. Workshop Adaptive Learning Syst., vol.2 pp. 80-84.
- [10] Y.Qi, and B.R. Hunt, "Voiced-Unvoiced-Silence Classifications of Speech using Hybrid Features and a Network Classifier," IEEE Trans. Speech Audio Processing, vol. 1 No.2, pp. 250-255.
- [11] D.G. Childers, M. Hahn, and J.N. Larar, "Silent and Voiced Classification of Speech," IEEE Trans. On ASSP, vol.37 No.11, pp.1771-1774.
- [12] J.D. Markel, "The SIFT algorithm for fundamental frequency estimation," IEEE Trans. Audio Electroacoust., vol.AU-20, pp. 367-377.