



ISSN 2278 – 0211 (Online)

Basic Characteristics of Speech Signal Analysis

S. Poornima

Assistant Professor, VlbJanakiammal College of Arts and Science, Coimbatore, Tamil Nadu, India

Abstract:

Speech signal include several fundamental characteristics and it is divided into two types of features such as time-domain speech signal features and frequency domain signal features which is mainly used for segmenting speech signals. Three important characteristics of speech signal is short time zero crossing, energy and auto correlation. The short time energy and short time zero crossing rates are important properties for detecting the end point of a speech signal analysis. Especially these two properties are used in voiced and unvoiced segmentation and classification.

Keyword: Short-time energy, short-time zero crossing rate, autocorrelation, spectral centroid, spectral flux.

1. Introduction

Speech processing is an interesting area of signal processing where speaker identification and speaker recognition are widely used applications. The first feature is Time-domain speech signal features are short-time energy, short time zero crossing rate, short time autocorrelation. Frequency domain features such as spectral centroid, spectral flux. The other characteristics of speech signals are pitch, stress, power spectral density, vowel duration, rhythm and intonation patterns. These characteristics mainly involved in speech segmentation which is recognizable and meaningful. Three important characteristics of speech signal is short time zero crossing, energy and auto correlation. [4].

2. Time-Domain Signal Features

Time-domain signal features can be used in speech segment extraction, which has algorithm for implementation and efficiency calculation. It contains three features such as short-time energy, short-time zero-crossing rate and autocorrelation. Short-time energy and short-time zero-crossing is most important features to detect the voiced and voiceless speech.

2.1. Short-Time Energy

Short time energy is a basic and important characteristic of speech processing. Energy is defined like strength of the signal. Speech signals are naturally differing in terms of energy with respect to time. Short time analysis is used to estimate the speech signal. In General, the speech signals are voiced, unvoiced, silence and noise regions. Analyzing speech signal based on energy will have higher pressure on identifying these regions of a speech signal [4]. Voiced segments of a speech signal will have higher short-time energy and it will be low when unvoiced speech occurs. It is too low when speech signal is silent. Short time energy is opposite process of zero crossing rates. It is calculated by the equation

$$E_n = \sum_{m=-inf}^{inf} [x(m)W(n - m)]^2$$
 (1) Where, E_n is an energy at sample n in the signal x , W is a window and m is number of frames will be occurring in the signal. Short time energy of sample speech signal for the sentence “My first speech processing work” is shown in figure 1.

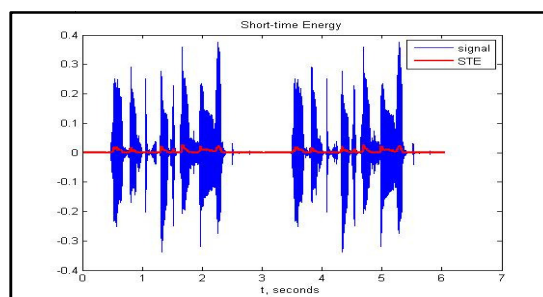


Figure 1: Short time energy of sample speech signal

In the above figure the speech signal is displayed in blue colour and short time energy is plotted by red colour which shows high energy in few point of speech signal [6]. Figure 2 shows short time energy of normal speech for the word “Already”.

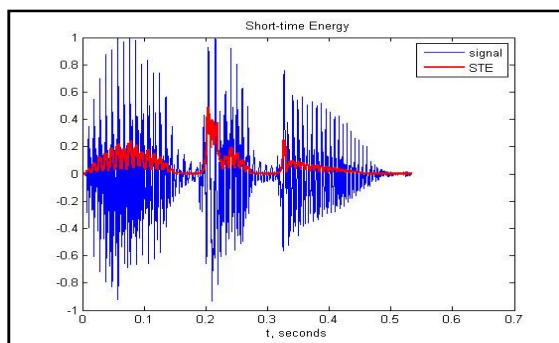


Figure 2: Short-time energy of normal speech word ‘Already’

The voiced speech signal has high energy power and unvoiced speech has low energy power. Silent is identified when energy is too low. Figure 3 shows the short-time energy of stuttering speech word “Already”.

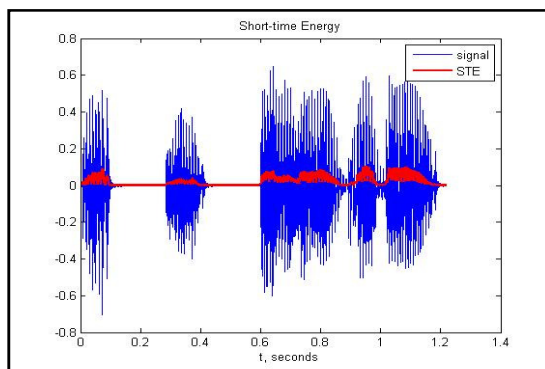


Figure 3: Short-time energy of stuttering speech word for the ‘Already’

In stuttering speech short-time energy can be find out in the silent region, voiced and unvoiced regions are foundby calculating the pressure of energy.

2.2. Short-Time Zero Crossing

Short time zero crossing is another most popular characteristic of speech signal. It is identified by point where the positive sign change to negative sign in a graph of speech signal. The zero crossing is defined as “Number of times an amplitude of the sign wave changes in a sound sample”. Zero crossing rates are used to identifying the voiced and unvoiced speech signal and end point detection. Zero crossing value is high when unvoiced speech or silent occurring in speech signal [2].It is measured by calculating how many times amplitude of the speech signals passes through a value of zero in a particular time interval. Short time zero crossing rates will be calculated by

$$Z_n = \sum_{m=-inf}^{inf} |sgn[x(m) - sgn[x(m - 1)]]|W(n - m)(2)$$

Where, Z_n is a zero crossing rate at sample n in the signal x, sgn is a signum function and m is a number of frame to be occur in speech signal. Short-time zero crossing rate of sampling speech signal is shown in figure 4.

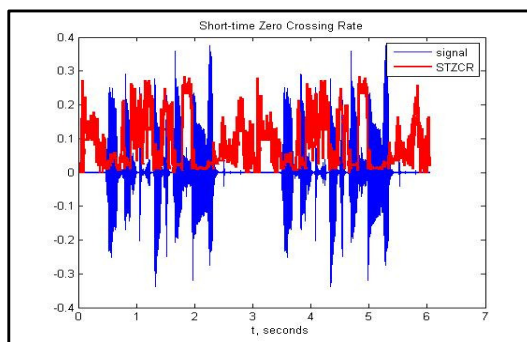


Figure 4: Short time zero crossing rate of sample speech signal.

The short time zero crossing is shown in in red colour and it is high when unvoiced signals compacted [3]. Figure 5 shows the short-time zero-crossing for normal speech for the word “Already”.

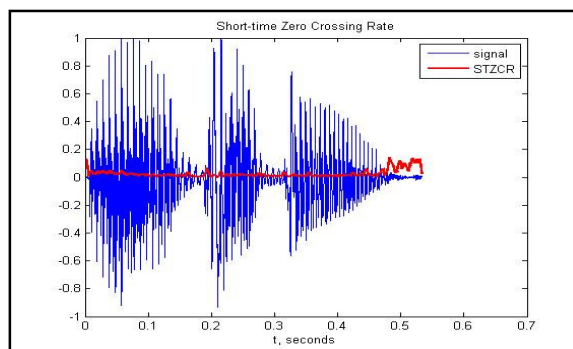


Figure 5: Short-time zero-crossing of normal speech word ‘Already’

Zero-crossing value is low when normal voiced speech occurs. Figure 6 shows short-time zero-crossing of stuttering speech word ‘Already’.

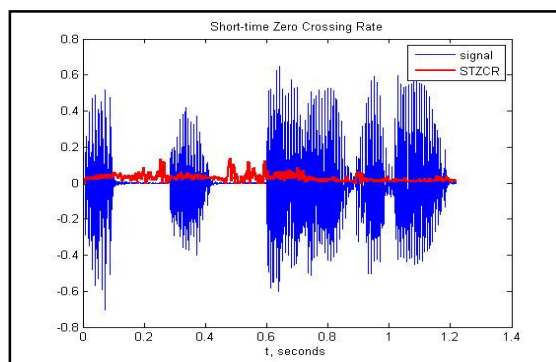


Figure 6: Short-time zero crossing rate of stuttering speech word ‘Already’

Silent regions will have very low energy and very low zero crossing values. So it will be removed. The number of zero crossing is processed nearly equal to zero.

2.3. Autocorrelation

Autocorrelation is a system analysis function and it is also called as serial correlation. It is computed by the correlation of time series compared and identifying similarities between with its own past and future values. It is used for detecting the repetition or periodicity present in the signals. Past values are the values before respective autocorrelation frame, and future values are the values after respective autocorrelation frame [6]. Short-time autocorrelation will be calculated by,

$$R_n(k) = \sum_{m=0}^{N-1-k} [x(n+m)w'(m)][x(n+m+k)w'(k+m)](3)$$

Where, $R_n(k)$ is a short time autocorrelation at sample n in the signal x. W is a window. Autocorrelation function is very useful tool in speech processing and used to identify the similarities of speech characteristics with respect to time. Short-time autocorrelation of sampling signal is shown in figure 7.

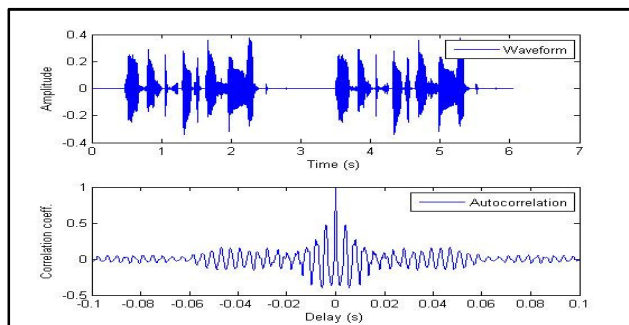


Figure 7: Short-time auto correlation of sampling speech signal

In voiced segmentation where signal is periodic and it is used to finding the high peaks for deciding voiced and unvoiced speech by using auto correlation function peaks [4].

3. Frequency-Domain Signal Features

Frequency domain features are used to analyse the signal with respect to the frequency which is extracted by Discrete Fourier transform. The frequency-domain signal features consist of two important features such as spectral centroid and spectral flux.

3.1. Spectral Centroid

Spectral Centroid is generally correlated with measure of sound brightness which characterizes the spectrum. A measure that specifies the spectrum is 'center of gravity' used by frequency of Fast Fourier Transform and magnitude information. It is calculated by dividing the average weighted frequency of amplitudes which is divided by the sum of amplitude. The spectral centroid is calculated by,

$$\text{SpectralCentroid} = \frac{\sum_{m=0}^{N-1} f(m)X(m)}{\sum_{m=0}^{N-1} X(m)} \quad (4)$$

Where $f(m)$ represent the center frequency of the bin m with length N and $X(m)$ is weighted frequency value or DFT spectrum amplitude of the bin m . In speech signal spectral centroid and energy is low when silent regions are occurred [4]. Figure 8 shows the spectral centroid of normal speech word 'Already',

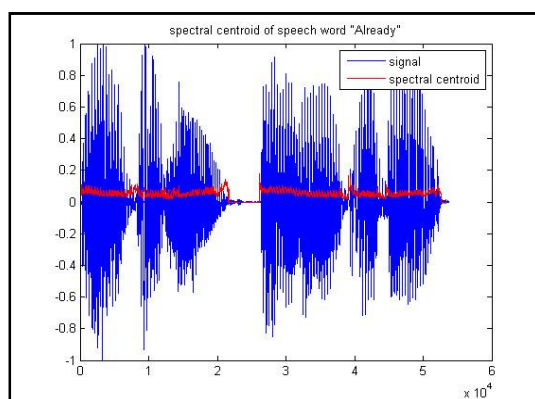


Figure 8: Spectral centroid of speech word 'Already'

Where original speech is shown in blue colour and red colour is spectral centroid for sample speech word "Already".

3.2. Spectral Flux

Spectral flux is measured the changing of power spectrum of the signal and it is computed by power spectrum comparison. It is most important feature to separate the music from speech signal. It is also defined as squared difference between two normalized magnitude of successive spectral distribution and it represent the successive signal frames. It is calculated by the equation

$\text{SpectralFlux}_i = \sum_{k=1}^{N/1} (|X_i(k)| - |X_i(k-1)|)^2 \quad (5)$ Where $X(k)$ is the DFT coefficient of i^{th} short term frame with length N . Spectral flux is used to find out the tone of audio signal [1].

4. Other Speech Signal Characteristics

Basic speech signal characteristics such as pitch and intonation which is identified by producing the speech. The speech is produced by air pressure come from lungs through vocal cords via vocal track. When vocal cord does not vibrate unvoiced sound is produced. Voiced sounds are produced when vocal card vibrate correctly.

4.1. Pitch and Intonation

Pitch frequency is an important parameter of speech processing. Vocal cord produced voiced and unvoiced sounds based on it vibration. A vibration sounds are delivered with glottal pulse, it has fundamental frequency and harmonics. The fundamental frequency of glottal pulse is called pitch. Basically pitch is also known as frequency of sound. Sound can be characterized based on pitch value, loudness and quality. It is compared like high or low in musical sounds. The pitch is just ear response of frequency. Human can hear range of sound between 3Hz to 3,000 Hz. Intonation speeches is basically a matter of vibrating in the pitch level of the voice [7].

5. References

- i. Abdullah I. Al-Shoshan, "Speech and Music Classification and Separation: A Review", J. King Saud Univ., Vol. 19, Eng. Sci. (1), pp. 95-133, Riyadh (147H./302).

- ii. A. Milton, S. Ashitha Dayana, S. Tamil selvi, "Voiced and unvoiced classification of speech signal using Average Zero Crossing Index Difference Function", International Journal of Advanced Information Science and technology (IJAIST), ISSN: 1319:1281.
- iii. Bachu R.G., Kopparthi S., Adapa B., Barkana B.D. "Separation of Voiced and Unvoiced using Zero Crossing Rate and Energy of the Speech Signal", Advanced Techniques in Computer Science and Software Engineering 33, pp:179-181.
- iv. Md.Mijanur Rahman, Md.A1-Amin Bhuiyan, "Continuous Bangla Speech Segmentation using Short-term Speech Features Extraction Approaches", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol.3, No. 11,311.
- v. Mojtaba Radnard, Mahdi Hadavi, Mohammad Mahdi Nayebi, "A new method of voiced and unvoiced classification based on clustering", Journal of Signal and Information Processing, 311, 1,332-347.
- vi. Paulraj M.P, Sazali Bin Yaacob, Ahamad Nazri Abdullah, Sathees Kumar Natraj, "Segmentation of voice portion for voice pathology classification using Fuzzy logic", Challenges and innovation in information technology, 33.
- vii. <http://www.physicsclassroom.com/class/sound/Lesson-2/Pitch-and-Frequency>