Indian Journal of Science and Technology, *Vol 8(29), DOI: 10.17485/ijst/2015/v8i29/83616, November 2015*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Performance Analysis of Windowing Techniques in Automatic Speech Signal Segmentation

P. L. Chithra[1]* and R. Aparna[2]

[1]Department of Computer Science, University of Madras, Chennai - 600 005, Tamil Nadu, India; chitrasp2001@yahoo.com
[2]Department of Computer Science, MOP Vaishnav College, Chennai – 600034, Tamil Nadu, India; aparna_ilangovan@yahoo.co.in

Abstract

Background/Objectives: Automatic Speech Recognition (ASR) and Language Identification (LID) are the key areas of acoustic speech signal processing. Speech signals watermarking, steganography and cryptography are considered to be the emerging techniques to ensure information security in speech signal transmission. Efforts should be taken to retain every minute detail of the signal. In all above mentioned methods, it is necessary to preprocess the speech signal so as to get best results. To take into account, this paper presents the performance analysis of windowing techniques in automatic speech signal segmentation. **Methods/Statistical Analysis:** Speech signal is segmented into syllables as a first step. In the process of segmentation, a windowing technique is applied to enhance the syllable boundaries. Then window function is applied to the input signal before Discrete Fourier Transform has applied. There are many windowing techniques available. Proposed work is carried to analyze the performance of few windowing functions in order to retain every minute detail of the signal and to preprocess the speech signal effectively in the Automatic Speech Recognition (ASR) and Language Identification (LID). **Findings:** The results produced by the windowing and filtering techniques in segmentation process are plotted. The proposed method out performs well and the performance of different windowing and filtering techniques is analyzed. The number of peaks found is tabulated. **Application/Improvements:** Thus our experimental results shows the significance of segmenting speech signals effectively using windowing function with discrete filters than the adaptive filters. Further those segments can be used in the field of Automatic Speech Recognition (ASR), Language Identification (LID), Speech signals watermarking, steganography. The observed segments with windowing using discrete filters are highly useful for clustering and pattern matching techniques

Keywords: Filtering, Speech Signal Processing, Segmentation, Windowing Function

1. Introduction

Speech processing is nothing but to analyze and process the speech signals. Automatic language identification is the process by which the language of digitalized speech utterance is recognized by a computer. Characteristics of the speaker voices are identified in the voice recognition. Automatic Speaker recognition is the identification of the person who is speaking by Speech recognition supports in text to speech and speech to text conversion. Our previous research work focused on identifying segment boundaries[1] in a continuous speech signal. Segmentation is considered to be the most vital step in signal processing. Filtering is a technique to enhance the segmentation process and to find the segment boundaries that matches with phoneme boundaries. Discrete filters, adaptive filter and multirate filters[2] are applied to the proposed system and their performances are compared to find the most appropriate filter for segmentation.

Other concerns in the field of speech signal processing is secure transmission of signals. Hiding secret signals into other cover signal is known as watermarking. A unique electronic identifier embedded in an audio signal

*Author for correspondence

is known as an audio watermark. It is used to identify ownership of copyright. Image watermarking embeds information into a signal (e.g. audio, video or pictures) so as to hard to remove. While the signal is copied, then embedded information is carried forwarded to the signal. An audio Steganography is proposed by Sridevi R et al.[3] In this method the Least Significant Bit (LSB) is substituted with the secret data. Cryptography techniques are used to prevent the unauthorized access. Signal cryptography has two sub processes - Encryption and Decryption of digital signals. Many algorithms are available for such processes.

Two categories of Window functions are Fixed and Adjustable window functions. Frequently used fixed window functions are Rectangular window, Hanning window, Hamming window and Blackman window. A special kind of adjustable window is Kaiser Window function. Digital FIR filter designing and spectral performance analysis are involved with these different windows. Subhadeep Chakraborthy[4] described the significance of Blackman window over Hamming window and Datar A et al.[10] also proved the advantages of Blackman window for medical signal processing. Rajesh M H et al.[7] proposed the significance of the modified group delay feature in speech recognition. Doddington G8 discussed the Syllable-based speech recognition. The group delay function is the derivative of the FT (Fourier Transformation) phase and it is processed to extract the information in the short-time FT (Fourier Transform) phase function. Poles and Zeros in the minimum phase group delay function can be renowned easily as peaks correspond to poles while valleys correspond to zeros. Non-minimum phase signals do not show this property.

Digital Signal processing applications render many research topics. Speech signals are considered to be the important and complicated signals that needed to be processed intensively. Segmentation process plays a vital role in ASR, LID and also in information security techniques. DFT/DWT transformation of the wave is found. The most important discrete transform is DFT which is used to perform Fourier analysis in many practical applications. In digital signal processing, the function is any quantity or signal that varies over time, such as the pressure of a sound wave, a radio signal, or a speech signal sampled over a finite time interval defined by a window function. Hence windowing technique holds higher importance in building DSP applications especially for areas dealing with speech signal. Filtering technique and its corresponding windowing function is applied in the process of segmentation. Filtering technique incorporated in the segmentation process identifies the phoneme boundaries related to the word boundaries[1]. Discrete filters and adaptive filters[2] are being used, and the results produced by each filter are compared.

Figure 1 shows the original speech signal, its energy and its group delay. The original signal is filtered and the energy of the semivowels is focused at low formant frequencies. During this filtering the semivowels will be attenuated rigorously without affecting much of the vowel regions. Hence it is ensured that the peak will be present at semivowel segment too. Hence make the system more effective in finding the segment boundaries. Main application of window function is filter designing. So before filtering is done for any speech signal, the underlying window function should be thoroughly analyzed. The two sinusoids of different frequencies in the waveform are analyzed then spectrally distinguish them the leakage interferes. If the frequencies of the two sinusoids are unequal and one component is weaker, then leakage from the other larger component can unclear the weaker one's presence. But if the two sinusoids frequencies are equal, leakage can render them irresolvable. The rectangular window suits well for sinusoids of comparable strength, at the same time it is not a good choice for sinusoids of disparate amplitudes. This is known as low-dynamic-range.

The window with the poorest resolution is another problem to be addressed. Sensitivity is very meager in the high-dynamic-range low-resolution windows, this is, if the random noise is present in input waveform, close to the frequency of a sinusoid, the response to noise, compared to the sinusoid, will be higher than with a higher-resolution

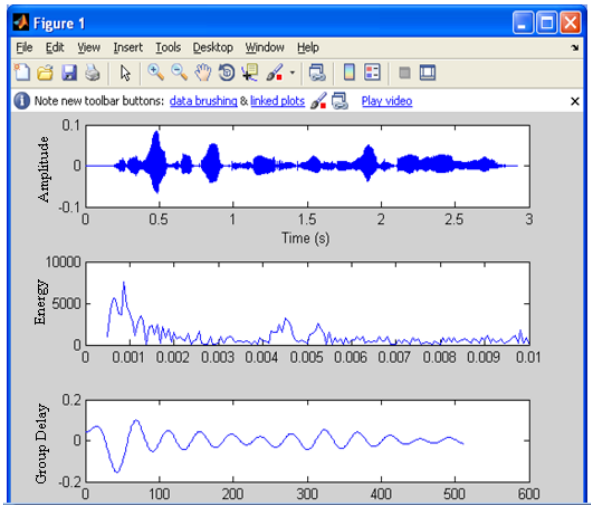

Figure 1. The original speech signal, its energy and its group delay.

window. That is the high-dynamic-range window reduces the ability to find weak sinusoids amidst the noise. High-dynamic-range windows are classified in wideband applications. Wideband applications are such applications that deal with the spectrum that contain many different components of various amplitudes. In between the extremes are Hamming and Hann windows which are known as moderate windows. They are commonly used in the spectrum of a telephone channel which is an example of narrowband application. Thus spectral analysis involves a compromise between resolving similar strength components with comparable frequencies and determining disparate strength components with dissimilar frequencies. The window function is chosen when tradeoff occurs.

The rest of the paper is organized as follows. Section II elaborates the methodology. Section III explains the windowing techniques applied. Section IV presents the experimental analysis and the performance analysis and Section V gives the conclusion of the research work performed.

2. Methodology

First step in segmentation of continuous speech signal is to digitalize the signal. The short-term energy function for the digitalized signal is calculated. FFT (Fast Fourier Transformation) of the energy function is found. Length of the window[4] is approximately adjusted to the length of the signal so that obtained result is closer to the number of phonemes in the taken speech signal. Various windowing techniques are applied to the continuous speech signal and its performance is evaluated. The resultant FFT is raised to the power of 0.01 so that the magnitude spectrum calculated is brought to the optimized value. Next step is to invert the derived signal. For the inverted signal IFFT (Inverse Fast Fourier Transformation) is found. Discrete and Adaptive filters are applied, and then the minimum phase group delay[5] is found for the filtered signal. The results are compared to conclude with the best filter. Hence, the graph is plotted, which is the minimum phase group delay function of the signal. Positive peaks on the graph relate to the phoneme boundary. The research work is carried to analyze the performance of different windowing techniques with discrete and adaptive filters.

2.1 Speech Signal Segmentation Algorithm

1. Digitalize the speech signal
2. Calculate the short-term energy function for the speech signal
3. Construct the symmetric part of the sequence by generating a lateral inversion of this sequence which is an arbitrary magnitude spectrum
4. The resultant is raised to the power of γ where γ is $0 < \gamma \leq 2$. (Specifically, the value of γ has been optimized to 0.01.)
5. Invert the function
6. Calculate the inverse DFT of the function, which is the root cepstrum and the causal portion of it, has minimum phase properties.
7. Apply appropriate filter as required
8. Calculate the minimum phase group delay function
9. The positive peaks in the minimum phase group delay function which approximately corresponds to sub-word/syllable boundaries are found.

Figure 2 describes the above algorithm steps for a sample data. Filtering technique included in the segmentation process identifies the syllable boundaries related to the word boundaries[9,10]. Discrete filters and adaptive filters are being used, and the results produced by each filter are compared. The technique of screening out a message before it is passed on to some other process is known as filtering. To perform filtering in a more eminent way: 1. More than one communication channel can be established, 2. As many intermediaries as possible can be eliminated, and 3. Distortion can be decreased by condensing message information to the bare essentials. The fricative in the input speech signal will reduce the significance of peaks in the signal, as the fricative produces false peaks in the signal. This will be manifested in the group-delay domain[5] also, which is a spurious peak. The signal

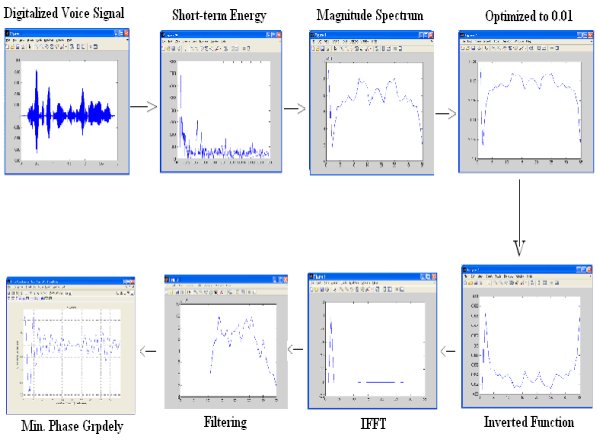

Figure 2. The Above Algorithm Steps for a Sample Data.

is low-pass filtered to avoid this kind of misinterpretation and to eliminate the high frequency fricatives. This also causes slight variation in the segment boundary[6]. Therefore the group delay function derived with the fricatives should not be treated as the reference. It is essential to ensure whether the peak present in the group delay function is due to the fricative or by the original signal. If a band-pass filter[4] is applied to the original signal, since the energy of the semivowels are concentrated at low formant frequencies alone, the semivowels will be attenuated severely without affecting the vowel regions much[11]. This will ensure that a peak will be present at semivowel segment also.

3. Windowing Technique

Windowing techniques are mainly used in the process of designing digital filters. In order to convert an impulse response of infinite duration to a Finite Impulse Response (FIR) filter design[4] windowing is performed. Symmetrical sequences of Window functions generated for digital filter design. Those window functions are usually an odd length with a single maximum at the center. For spectral analysis[5], Windows for DFT/FFT are formed by removing the right-most coefficient of an odd-length, symmetrical window. Truncated sequences are known as periodic. When the truncated sequence is periodically extended, the deleted coefficient is commendably restored (by a virtual copy of the symmetrical left-most coefficient). Window technique consists of a function called window function which is nothing but if some interval is chosen, it returns with finite non-zero value inside that interval and zero value outside that interval

Mathematical expressions:

$$Y(z) = X(z)H(z) \tag{1}$$

Where H(z) is a system with input X(z) and output Y(z). After windowing the input signal, we get,

$$X'(z) = X(z)H(z) \tag{2}$$

Now the windowed signal is passed as the input to the system H(z).

$$Y'(z) = X'(z)H(z) \tag{3}$$

Equation (1) – (3) describes the window functions. For the performance analysis, Ten different window functions are applied and the corresponding output is noted.

3.1 Haan Window

The following Hann function is used to select a subset of a series of samples in order to perform a Fourier transform or other calculations:

$$w(n) = 0.5\left(1 - \cos\left(\frac{2\pi n}{N-1}\right)\right), 0 \le n \le N \tag{4}$$

3.2 Hamming Window

Hamming window is used to optimize the window to minimize the maximum (nearest) side lobe, giving it a height of about one-fifth that of the Hann window.[22] [23]

$$w(n) = \alpha + \beta\cos\left(\frac{2\pi n}{N-1}\right) \tag{5}$$

where, $\alpha = 0.54$, $\beta = 1-\alpha = 0.46$

3.3 Blackman Window

Blackman windows [12] are defined as:

$$w(n) = a_0 + a_1\cos\left(\frac{2\pi n}{N-1}\right) + a_2\cos\left(\frac{4\pi n}{N-1}\right) \tag{6}$$

where,

$$a_0 = \frac{1-\alpha}{2};\ a_1 = \frac{1}{2};\ a_2 = \frac{\alpha}{2}$$

3.4 Kaiser Window

The Kaiser, or Kaiser-Bessel, window is a simple approximation of the DPSS window using Bessel functions.

$$w(n) = \frac{I_0\left(\pi\alpha\sqrt{1-\left(\frac{2n}{N-1}-1\right)}\right)^2}{I_0(\pi\alpha)} \tag{7}$$

where I_0 is the zero-th order modified Bessel function. 'α' is the variable parameter used to determine the tradeoff between main lobe width and side lobe levels of the spectral leakage pattern.

3.5 Rectangular Window

The rectangular window is considered to be the simplest window, equivalent to replacing all but N values of a data sequence by zeros, making it appear as though the waveform suddenly turns on and off:

$$w(n) = 1 \tag{8}$$

3.6 Triangular Window

The triangular window is nothing but the 2nd order B-spline window which can be given by,

$$w(n) = 1 - \left| \frac{n - \frac{N-1}{2}}{\frac{L}{2}} \right| \tag{9}$$

3.7 Flat Top Window

A flat top window is a partially negative-valued window that has a flat top in the frequency domain,

$$w(n) = a_0 - a_1 \cos\left(\frac{2\pi n}{N-1}\right) + a_2 \cos\left(\frac{4\pi n}{N-1}\right) - a_3 \cos\left(\frac{6\pi n}{N-1}\right) + a_4 \cos\left(\frac{8\pi n}{N-1}\right) \tag{10}$$

$a_0 = 1$, $a_1 = 1.93$; $a_2 = 1.29$; $a_3 = 0.388$; $a_4 = 0.028$

3.8 Gaussian Window

The following is the Gaussian window which is used to produces a parabola, this can be used for nearly exact quadratic interpolation infrequency estimation.

$$w(n) = e - \frac{1}{2}\left(\frac{n - \frac{N-1}{2}}{\sigma \frac{N-1}{2}} \right)^2, \sigma \leq 0.5 \tag{11}$$

3.9 Welch Window

The Welch window is given by,

$$w(n) = 1 - \left(\frac{n - \frac{N-1}{2}}{\frac{N-1}{2}} \right)^2 \tag{12}$$

3.10 Nuttall Window

The expression for Nuttall window is,

$$w(n) = a_0 - a_1 \cos\left(\frac{2\pi n}{N-1}\right) + a_2 \cos\left(\frac{4\pi n}{N-1}\right) - a_3 \cos\left(\frac{6\pi n}{N-1}\right) \tag{13}$$

where $a_0 = 0.355768$; $a_1 = 0.487396$; $a_2 = 0.144232$

4. Experimental Analysis

The objective of this proposed research is to analyze the performance of various above discussed windowing techniques in the process of segmentation. As segmentation[1,2] is considered to be the crucial step in signal processing areas, the best windowing function should be used so as to proceed with building the ASR, LID or any other signal processing applications. As explained in the background section, the after digitalizing the continuous speech signal the second step is to compute the short-term energy for the signal. The short- term energy calculation requires the continuous speech signal to be in a range. Hence, the speech signal has to be passed into a window, to make them adjust in that particular range without losing any detail of the speech signal. And the above mentioned windowing functions are applied to the speech signal with discrete and adaptive filters in the process of segmentation. The results produced by the windowing and filtering techniques in segmentation process are plotted. The proposed method out performs well and the performance of different windowing and filtering techniques is analyzed.

Figure 3 shows the input speech signal. Figure 4 shows the processed signal, in which the peaks can be easily identified. Those peaks are considered as the segment boundaries.

This research work is carried on the audio signal from TIMIT Acoustic – Phonetic Continuous Speech Corpus. http://www.Idc.upenn.eduCatlogdescaddendaLDC93S1.wav. It has the following eleven words which can be

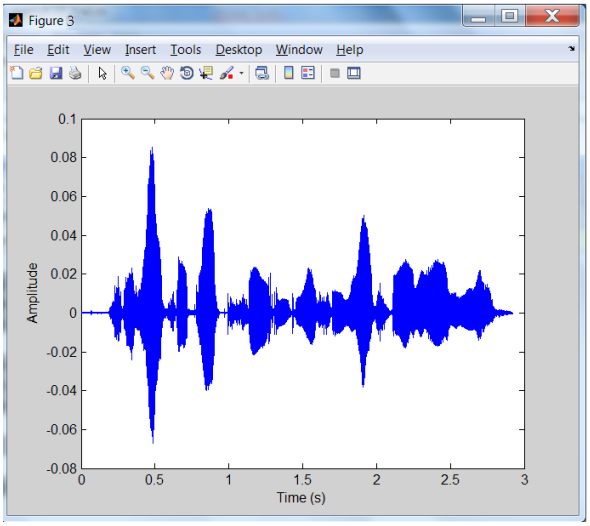

Figure 3. Input Speech Signal.

grouped into 37 phonemes. Sample data is She had your dark suit in greasy wash water all year

Discrete filters and Adaptive filters are taken for consideration in the process of segmentation.

The number of peaks found is tabulated in Table 1. Poles and Zeroes in the minimum phase group delay functions can be distinguished easily as the peaks correspond to poles and the valleys correspond to zeroes. Finding peaks is nothing but identifying the syllable boundaries for segmentation. If the number of peaks matches with number of phonemes, then the peaks can be considered segment boundaries.

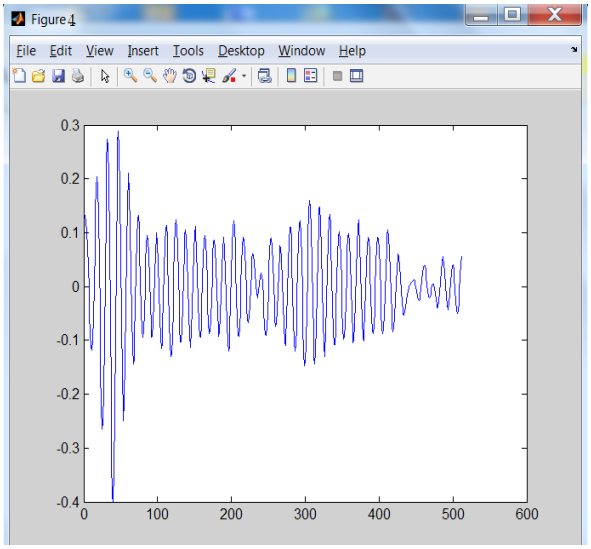

Figure 4. The Processed Signal.

Table 1. The Number of Peaks Found in the Sample Data

#	Window Function	Discrete Filter	Adaptive Filter
1	Hann	38	28
2	Hamming	37	23
3	Blackman	38	30
4	Kaiser	35	22
5	Rectangular	35	21
6	Triangular	37	25
7	Flat top	36	29
8	Gaussian	38	26
9	Welch	38	29
10	Nuttall	38	29

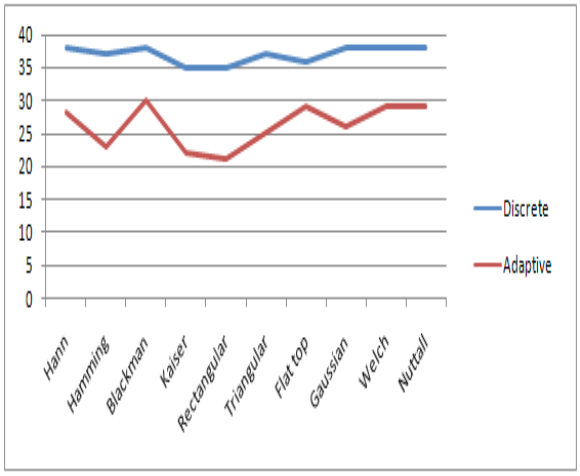

Figure 5. Windowing Function Plot with Discrete and Adaptive Filters.

4.1 Performance Analysis

The above tabulated result is plotted in a graph for easier comparison. The peaks identified using discrete filters matches with the number of phonemes in the speech signal.

From the above graph, it is found that windowing function performs best with discrete filters than adaptive filters.

5. Conclusions

This paper concluded that the taken windowing functions out performs well with discrete filters in which the peaks identified using discrete filters matches with the number of phonemes in the speech signal than the adaptive filters so as to find the appropriate peaks. With the peaks found, the speech signal can be easily segmented. Thus our experimental results shows the significance of segmenting speech signals effectively using windowing function with discrete filters. Further those segments can be used in the field of Automatic Speech Recognition (ASR), Language Identification (LID), Speech signals watermarking, steganography and for clustering and pattern matching techniques.

6. Acknowledgement

We owe sincere thankfulness to M.O.P Vaishnav College for women and University of Madras for being supportive throughout our research work.

7. References

1. Aparna R, Chithra PL. An effective method for continuous speech segmentation using filters. National Conference on Computing and Intelligence Systems. 2012; 1(1):17–23.
2. Aparna R, Chithra PL. A comparative study on various types of filters in audio signal processing. International Conference on Pattern Recognition Applications and Techniques Proceeding; 2013; 1(1):134–41.
3. Sridevi R, Damodaram A, Narasimham SVL. Efficient method of audio steganography by modified LSB algorithm and strong encryption key with enhanced security. Journal of Theoretical and Applied Information Technology. 2009; 5(6):768–71.
4. Subhadeep C. Advantages of blackman window over hamming window method for designing FIR filter. International Journal of Computer Science and Engineering Technology. 2013; 4(8):1181–9.
5. Nagarajan T, Kamakshi P, Hema AM. Minimum phase signal derived from root cepstrum. IEE Electronics Letters. 2003; 39(12):941–2.
6. Juang EB, Lee S, Soong F. Statistical segmentation and word modeling techniques in isolated word Recognition. Proceeding IEEE International Conference of Acoustics, Speech, Signal Processing. 1990; 2(1):745–8.
7. Rajesh M H, Hema AM, Gadde VRR. Significance of the modified group delay feature in speech recognition. IEEE Transactions on Audio, Speech and Language Processing. 2007; 15(1):190–202.
8. Doddington G. Syllable-based speech recognition. WS'97 final technical report, Center Lang. Speech Process; Johns Hopkins Univ., MD; 1997. p. 264–84.
9. Sree HK, Padmanabhan, Hema AM. Robust voice activity detection using group delay functions. Proceeding of IEEE Intl. Conf. Industrial Technology; 2006. p. 2603–7.
10. Datar A, Jain A, Sharma PC. Performance of Blackman window family in M-channel cosine modulated filter bank for ECG signal. Multimedia, Signal Processing and Communication Technologies, International, IEEE Conference; 2009 Mar 14-16; Aligarh. 2009. p. 98–101. ISBN: 978-1-4244-3602-6.
11. Rajput SS, Bhadauria SS. Implementation of fir filter using efficient window function and its application in filtering a speech signal. International Journal of Electrical, Electronics and Mechanical Controls. 2012 Nov; 1(1).