

“A Review: Different methods of segmenting a continuous speech signal into basic units”

Manpreet Kaur, Amanpreet Kaur

Department of Computer Science, RIMT Mandi Gobindgarh, Punjab, India, preetjot.gill04@gmail.com

Department of Computer Science, BSBEC Fatehgarh Sahib, Punjab, India, er.amanpreet.cse@gmail.com

Abstract: Speech is the medium through which human beings can communicate. Segmentation of speech is required for better speech recognition. Segmentation of speech can be done into basic units like words, phonemes or syllables. The two main methods used for segmentation of speech signals are manual segmentation and automatic segmentation. But manual segmentation is not favoured as it is tedious, time consuming and results are not reproducible. So, automatic segmentation is used for segmentation of speech. This paper describes a number of methods for automatic segmentation of speech into units like phonemes or Syllables.

Keywords: methods of segmentation, short term energy, minimum phase group delay, Word Chopper based, DWT.

1. INTRODUCTION

Speech is the way of communication between human beings. For recognition of speech, it is required to segment the speech into basic units such as words, phonemes or syllables. This is also required to make the speech recognition better between humans and machines as like as between humans. Speech segmentation makes the speech more understandable and also improves interpretation. Speech can be categorized into voiced, unvoiced and silence. Silence represents separation between the voiced speeches. Speech segmentation is act as sub problem of speech recognition. Speech Synthesis and recognition systems are required to be segmented into basic units like words, phonemes or syllables [1].

2. METHODS OF SEGMENTATION

Speech segmentation is the process of dividing the continuous speech into basic units having finest boundaries. It is an important step in speech recognition [4]. It also plays an important role in certain applications. Speech segmentation can also be used for speech recognition

systems [9]. Generally there are two methods of speech segmentation which are discussed as below:

- a) Manual Segmentation
- b) Automatic

Manual Segmentation: Segmentation of speech can be done manually by examining the waveform of speech signal with spectrogram. But this process is very tedious, time consuming, error prone and results cannot be reproduced [4].

Automatic Segmentation: Then, another method which is considered very convenient is an automatic segmentation. The speech can be automatically segmented into sub word units which are defined acoustically [1]. A number of methods can be used for automatic segmentation.

3. AUTOMATIC SEGMENTATION

Automatic segmentation is considered better than manual segmentation [1]. Speech can be efficiently segmented into its units which are words, phonemes and syllables using automatic segmentation. The various methods involving automatic segmentation are described below.

3.1 FOURIER TRANSFORM

Traditional method of segmentation of speech is Fourier Transform [4]. This is a constant bandwidth method and is used for stationary signals. In this, time and frequency cannot be used at a time. To overcome this problem a number of methods for segmentation of speech into various units are available.

3.2 SHORT TERM ENERGY

Mark Greenwood [6] has presented in his paper a technique of Short Term Energy of Speech Signal which is the energy of the speech signal at a particular instant of time. It differentiates between voiced, unvoiced and silence part of the speech. The short term energy is high for voiced speech, low for unvoiced speech and zero for silence. The short term energy of sentence ਰਾਮ ਖੇਡਦਾ ਹੈ is shown in figure 1.

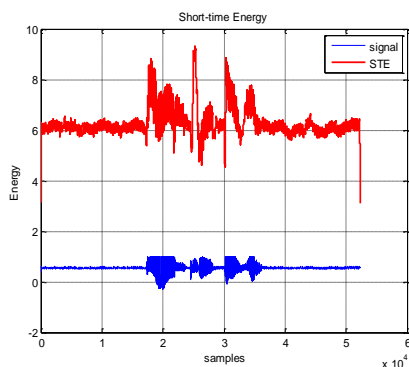


Fig 1: Short Term Energy of Speech

Amanpreet Kaur [1] has described a method for segmentation of speech into syllables using Short Term Energy of speech. In this method some threshold value is selected. The starting point of syllable is detected with value having more than thresholding value. The values less than threshold value are counted as zeros and constant zeros represent end of syllable.

Runshen Cai [8] has presented a technique for segmentation of speech into syllables based on phonetic transcriptions and features like short time average energy, short time zero crossing rate, product of above two features, ratio of first feature to the second and ratio of low frequency average energy to the total average energy.

3.3 MINIMUM PHASE GROUP DELAY METHOD

T. Nagarajan [2] has discussed Minimum Phase Group Delay method for segmentation of speech into syllables. The short term energy function involves the problems of thresholding and local energy fluctuations. To resolve these problems, the segmentation of speech signal is done with minimum phase group delay approach. The minimum phase signal is derived from short term energy. The negative derivative of Fourier Transform phase is used to derive 'group delay' function. The syllables present in a speech signal depend on the voiced segments. The short term energy of a syllable is high in voiced region. The group delay function of minimum phase signal performs better than short term energy to do segmentation.

3.4 WAVELET METHOD

Johnson Ihyeh Agbinya [7] offered a method for speech compression by using wavelet techniques. Voiced and Unvoiced speech and selection of type of wavelets are used for speech compression. In this method wavelets use low frequency coefficients. The voiced and unvoiced part in a speech signal is detected with energy in bands.

S. Ratsameewichai [5] has presented a technique of Dual Band Energy Contour for segmentation of speech into phonemes. In this method, the speech signal is decomposed into low frequency and high frequency components using a wavelet decomposition technique. Then, the energy contour of decomposed speech signal is computed. By using energy contour of speech, the boundaries of phonemes can be located.

Bartosz Ziólko [3] described Wavelet method in their paper which is used for the detection of phonemes based on the power variations. The information from the speech signals can be efficiently extracted by using spectral analysis of speech. Spectral analysis can be done with the help of Discrete Wavelet Transform (DWT). The power is analyzed in various frequency sub bands to locate the beginning and end of phonemes. The boundaries of phonemes can be detected using power transitions in wavelets.

M. F. Tolba [4] has made an attempt on Discrete Wavelet Transform technique which is used for the segmentation of

speech signals into consonants and vowels. The frequency of the signal is used for the segmentation of speech in this method. The speech signal is decomposed into two parts which are approximation coefficients and detailed coefficients. The approximation components contain low frequency components and detailed coefficients contain high frequency components. The decomposition can be done at 1, 2, 3 levels. This can be done by applying Wavelet Transform individually on both approximation coefficients and detailed coefficients. The following diagram shows the one level decomposition of wavelet into detailed coefficients and approximation coefficients.

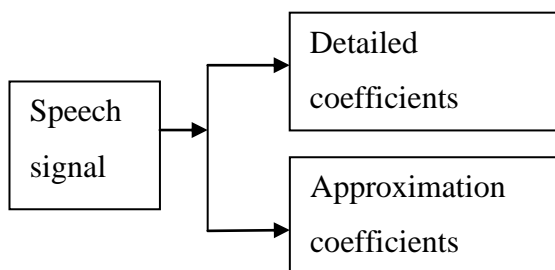


Figure 2: One level wavelet decomposition

DWT is non parametric method which can be used for various speech related problems. It can also be used for non stationary signals.

3.5 WORD CHOPPER TECHNIQUE

Nishi Sharma [9] has discussed the technique of segmentation of speech into syllables using Word Chopper. This paper involves Feature Extraction of speech like length, number of channels, sample rate, data length and bits per sample. Then, Word Chopper is used for features extraction. With this technique, silence regions can be detected and cut off and syllables are detected within the speech signal.

4. CONCLUSION

In this paper, techniques for segmentation of speech like short term energy, minimum phase group delay approach, Word Chopper based and discrete wavelet transform are described. The short term energy and energy contour

methods are suffered from the problem of thresholding and minimum phase group delay method faces problem at silence regions and semi vowels. Also, the Word Chopper based segmentation cannot be used for segmentation of all words. So, the robust method for segmentation of speech is Discrete Wavelet Transform (DWT) because this method uses the frequency and time simultaneously.

References

- [1] Amanpreet Kaur and Tarandeep Singh 'Segmentation of continuous speech into syllables' in the proceedings of world congress on Engineering and Computer Science, San Francisco, USA, 2010.
- [2] T. Nagarajan, Hema A. Murthy and Rajesh M. Hegde 'Segmentation of speech into syllable-like units' in Proc. EUROSPEECH-03, Geneva, Switzerland, Sep. 2003, pp.2893-2896
- [3] Bartosz Ziolko, Suresh Manandhar, Richard C. Wilson and Mariusz Ziolko 'Wavelet method of speech segmentation' department of Computer Science, University of York, Heslington, YO10 5DD, York, UK.
- [4] M. F. Tolba, T. Nazmy, A. A. Abdelhamid, M. E. Gadallah 'A Novel method for Arabic Consonant/Vowel segmentation using wavelet transform' Faculty of computer and information sciences, Military Technical College, Ain shams University, IJICIS, Vol. 5, No. 1, July 2005.
- [5] S. Ratsameewichai, N. Theera-Umpon, J. Vilasdechanon, S. Uatrongjit, and K. Likit-Anurucks 'Thai Phoneme Segmentation using Dual-Band Energy Contour' Electronic Circuits and Systems Research Laboratory, Department of Electrical Engineering Chiang Mai University, Thailand.
- [6] Mark Greenwood, Andrew Kinghorn 'Automatic silence/unvoiced/voiced classification of speech' Department of Computer Science, University of Sheffield, UK
- [7] Johnson Ihyeh Agbinya 'Discrete Wavelet Transform Techniques in Speech Processing' in Proc. Tencen Digital Signal Processing Applications New South Wales, Australia, 1996.
- [8] Runshen Cai 'An Automatic Syllable Segmentation Method for Mandarin Speech' Computer Science & Information Engineering College, Tianjin University of Science and Technology, Tianjin, China.
- [9] Nishi Sharma and Parminder Singh 'Automatic Segmentation of Wave File' in the International Journal of Computer Science and Communication, vol. 1, no. 2, July- December 2010.