Volume 3, Issue 6, June 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

A Survey on Punjabi Speech Segmentation into Syllable-Like Units Using Group Delay

Anupriya Sharma, Amanpreet Kaur
RIMT-IET, Mandi Gobindgarh
India

***Abstract*— This paper describes the best method for automatic segmentation of Punjabi speech. Punjabi is one the most widely used language. Consequently, Punjabi is a syllabic language, so syllables are found to be the best units for the preparation of speech database. Further the automatic segmentation of speech is proved to be the better approach than the manual segmentation. The basic characteristics of the speech can be evaluated by the zero crossing rate and short term energy. Due to local energy fluctuation the STE functions are not reliable for continuous speech. Thus the approach used for the segmentation of speech is by using the negative derivative of Fourier transformations i.e. the Group Delay method. The peaks and valleys are better resolved when the signal is in minimum phase in case of group delay approach.**

***Keywords*— Short Term Energy, Zero crossing rate, Automatic speech segmentation, Group delay, Syllable units, Punjabi Syllables.**

I. Introduction

Spoken language is not just a means to access information, but itself information. In order to facilitate communication between humans and machines Automatic speech recognizers (ASR) are used. Speech recognizer is a machine which understands humans and their spoken words. For automatic recognition of continuous speech, it is required to perform the process of segmentation. The Segmentation of acoustic signal into basic units is an important stage. The syllables are very important unit of language. The syllables are composed of vowels and consonants. The short-term energy (STE) function contains useful information about syllable segment boundaries. By using short-time Fourier analysis the information in the speech signal can be extracted. The group delay function is found to be a better representative of the STE function for syllable boundary detection.

II. Units For Segmentation

The Speech recognition and synthesis systems always need a speech signal to be segmented into some basic units like Words, Phonemes, or syllables. Depending on the size of vocabulary the choice of representative units are made. Word is the most natural unit of segmentation. It's not appropriate to use words as the units for segmentation due to lack of generalization and more memory consumption [7]. Phonemes are the smallest segmental unit of sound employed to form meaning. The same phoneme in different words has different realization. There is overgeneralization of phonemes. So the combination of phone and words gives rise to next level basic unit of speech called as syllables. Syllable like units are defined by rules, a syllable must have a vowel called its nucleus, where as presence of consonant is optional.

III. Syllables For Punjabi Language

It was discussed in one of the paper (Hema A. Murthy and Ashwin Bellur in 2011) the scripts in Indian languages have originated from the ancient Brahmi script [9]. The basic units of the writing system are referred to as Aksharas. An Akshara is an orthographic representation of a speech sound in an Indian language, they are syllabic in nature, the typical forms of Akshara are V, CV, CCV and CCCV [9], thus have a generalized form of C*V where C and V are consonant vowel. The syllables are composed of vowel and consonants. Every syllable must have a vowel. The vowel is also known as nucleus, where as presence of consonant is optional. Vowel (V) is always the nucleus part and the left part is onset and the right part is coda that is consonant.
Following are the seven types of syllables recognized in Punjabi language:

TABLEL I
VARIOUS SYLLABLES IN PUNJABI

V	Vowel	ੳ	ੳ
VC	Vowel+ Consonant	ਇ+ਹ	ਇਹ
CV	Consonant +Vowel	ਗ+ਆ	ਗਾ
VCC	Vowel+ Consonant+ Consonant	ਜ+ ਗ+ ਗ	ਜੱਗ
CVC	Consonant +Vowel+ Consonant	ਰ+ਆ+ਤ	ਰਾਤ
CCVC	Consonant + Consonant +Vowel+ Consonant	ਸ+ਵ+ਏ+ਰ	ਸਵੇਰ
CVCC	Consonant +Vowel+ Consonant+ Consonant	ਸ+ਊ+ਰ+ਜ	ਸੂਰਜ

Punjabi language has thirty eight consonants, ten non-nasal vowels and same number of nasal vowels. Vowels can appear alone but consonants can only appear with vowels. Following are the list of consonants in Punjabi language:

ਸ ਹ ਕ ਖ ਗ ਘ ਙ ਚ ਛ ਜ ਝ ਞ
ਟ ਠ ਡ ਢ ਣ ਤ ਥ ਦ ਧ ਨ ਪ ਫ ਬ
ਭ ਮ ਯ ਰ ਲ ਵ ੜ ਸ਼ ਖ਼ ਗ ਜ ਫ਼ ਲ਼

List of Non-Nasal Vowels:

ਈ ਇ ਏ ਐ ਅ ਆ ਔ ਉ ਊ ਓ

The number of nasal vowels is same as non-nasal vowels and is represented by Bindi or Tippi over the Non-Nasal Vowels.

IV. General Characteristics Of Speech

A continuous speech signal consists of two main parts: one carries the speech information, and the other includes silent or noise sections that are between the utterances, without any verbal information. The verbal (informative) part of speech can be further divided into two categories: Voiced and Unvoiced speech. When air from the lungs passes through the larynx *Voiced sounds* are produced. The fundamental frequency is also called the pitch, and differs among people due to the differences in their larynx's anatomy. Men's pitch range occurs commonly in the interval between 50 to 250 Hz while women's lies between 120 and 500 Hz. *Unvoiced speech sounds* are produced by air passed directly through vocal tract formations. Unvoiced speech, contrary to voiced speech, does not exhibit periodicity, and is characterized by a noise-like signal. The speech production process involves generating voiced and unvoiced speech separated by a *silence region*. There is no excitation supplied to the vocal tract during silence region and hence no speech output. However, silence is an integral part of speech signal. Without the presence of silence region between voiced and unvoiced speech, the speech will not understandable [3].

V. Characteristic Features For Voice And Its Detection Criteria

The zero crossing rate and short term energy are the two characteristics features for voice. The rate at which the speech signal crosses zero can provide information about the source of its creation. The unvoiced speech has a much higher ZCR than voiced speech. The amplitude of unvoiced segments is noticeably lower than that of the voiced segments. The amplitude variation is reflected by short-time energy of speech signals. For the segmentation of speech the STE function of speech signal can be processed. This method can only give the information about the number of voiced segments, excluding the phonetic content of the speech. Such approach can be used for the language independent segmentation of multilingual speech as it cannot directly perform segmentation due to local energy fluctuations. The syllable centers are the high energy regions in the STE functions and the valleys at both ends of the syllable nuclei are the syllable boundaries. For continuous speech STE functions are not reliable [2].

VI. Representation Of Speech

Traditionally, the information in speech signal is represented in terms of features derived from Fourier analysis: Fourier transformation, Fast Fourier transformation, discrete fourier transformation, or Wavelets. The key difference between fourier transform and wavelets transform is that wavelet transform is a multi-resolution transform as it allows a form of time-frequency analysis. When using the fourier transform the result is a very precise analysis of its frequency contained in the signal, but no information about when certain features occurred and about the scale characteristics of the signal. Scale is similar to frequency. It is a measure of the amount of detail in the signal. Small scale means coarse details and large scale means fine details.

The information in speech signals can be represented in terms of features derived from short-time Fourier analysis. The information in the short-time FT phase function can be extracted by processing the negative derivative of the FT phase, i.e., the group delay function [9].

$$H(\omega) = H1(\omega) \cdot H2(\omega), \qquad (1)$$

the group delay function $\tau h(\omega)$ can be represented as

$$\tau h(\omega) = \frac{-\partial(\arg(H(\omega)))}{\partial\omega}$$

$$= \tau h1(\omega) + \tau h2(\omega). \qquad (2)$$

The equation (1) shows the multiplicative property of magnitude spectra where as equation (2) is in group delay domain it becomes an addition. The group delay spectrum has been declared better due to its additive property over magnitude spectra. It was observed that in case of the magnitude spectra the peaks are clearly visible, but the peaks are not resolved in a system where the two poles are combined together [2]. The research shows the disadvantage of multiplicative property of magnitude spectra. In case of group delay spectra the peaks and valleys are better resolved when the signal is in minimum phase [2].

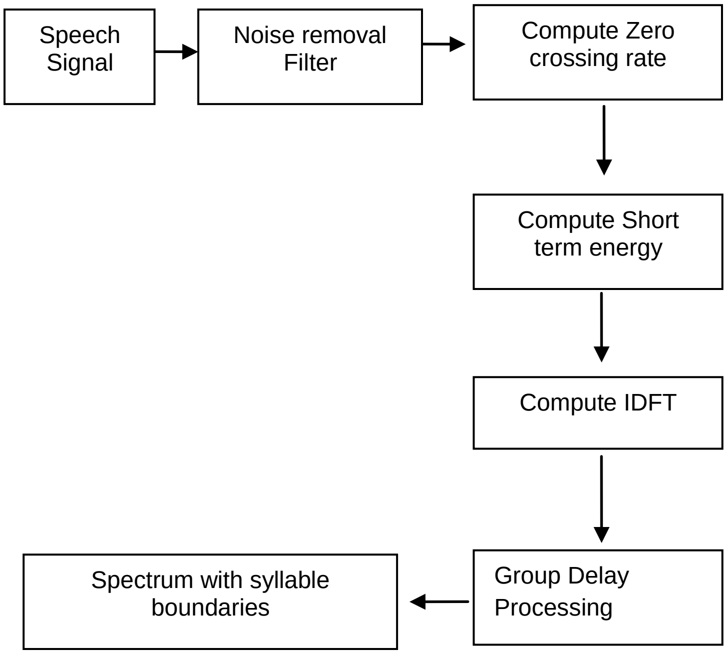
Speech Signal
Noise removal Filter
Compute Zero crossing rate
Compute Short term energy
Compute IDFT
Group Delay Processing
Spectrum with syllable boundaries

Fig 1: Steps involved in finding syllable boundaries

VII. METHODS OF SEGMENTATION

There are two ways of segmentation: Manual segmentation (hand labeling) and automatic segmentation (ASR). It has also been observed in some studies that, the deviation between manual and automatic segmentation had been calculated for the onset and offset values for the syllable boundaries the method was implemented and analyzed for different Punjabi speech signals. Results proved that the boundaries of syllables were marked automatically with accuracy. It had also been observed that the difference between two (Automatic and Manual) techniques is very much negligible and the boundaries of syllables marked by automatic technique were comparatively much accurate [7].

VIII. Conclusions

From the above survey it is clear that in order to segment the continuous speech, syllable as units are best suited for Indian languages. The group delay approach is best suited for the process of segmentation. The tools developed so far works only for Tamil, Hindi, Bengali, Malayalam, Telugu and Marathi. This survey points to the creation of speech database for Punjabi language by automatic segmentation of continuous speech into syllable like units using group delay approach so that it can further be used for recognition purpose.

References

[1] T.Nagarajan et al. "*Segmentation of speech into syllable-like units*," in Eurospeech Sixth biennial conference of signal processing, Geneva, 2003.

[2] T. Nagarajan and H. A. Murthy, "*Subband-Based Group Delay Segmentation of Spontaneous Speech into Syllable-Like Units*," in Eurasip Journal on Applied Signal Processing , Hindawi Publishing Corporation 2004:17, pp. 2614–2625.

[3] N. Mikael, E. Marcus, "*Speech Recognition using Hidden Markov Model, Performance evaluation in noisy environment*", Degree of master of science in Electrical Engineering, Department of telecomminications and engineering, Blekinge Institute of Technology, March 2002.

[4] G. Pradeep "*Text-to-Speech Synthesis for Punjabi Language*", Thesis degree of Master of Engineering in Software Engineering submitted in Computer Science and Engineering Department of Thapar Institute of Engineering and Technology (Deemed University), Patiala, May 2006.

[5] V. Kamakshi Prasad, T. Nagarajan, Hema A. Murthy, "*Automatic segmentation of continuous speech, using minimum phase group delay functions*," in the proceedings of science direct, Speech Communication 42, 2004, pp. 429–446.

[6] G Lakshmi Sar ada, et al. "*Automatic transcription of continuous speech into syllable-like units for Indian languages*," in Sadhana, Vol. 34, Part 2, April 2009, pp. 221–233

[7] K. Amanpreet, and S. Tarandeep, "*Segmentation of Continuous Punjabi Speech Signal into Syllables*," in the Proceedings of the World Congress on Engineering and Computer Science 2010 Vol I, WCECS 2010, San Francisco, USA, October 20-22, 2010.

[8] S. Parminder, L. Gurpreet, "*Corpus Based Statistical Analysis of Punjabi Syllables for Preparation of Punjabi Speech Database*," in International Journal of Intelligent Computing Research (IJICR), Volume 1, Issue 3, June 2010.

[9] A.Hema, and B.Yegnanarayan, "*Group delay functions and its applications in speech technology*," in Sadhana, Vol. 36, Part 5, October 2011, pp. 745–782.

[10] S. Nishi, and S. Parminder, "*Automatic Segmentation of Wave File*," in International Journal of Computer Science & Communication Vol. 1, No. 2, July-December 2010, pp. 267-270.

[11] A. Hema, B.Ashwin, et al., IIT-Madras, IIT-Kharagpur, CDAC-Trivandrum, CDAC- mumbai, IIIT-Hyderabad, "*Building Unit Selection Speech Synthesis in Indian Languages*," An Initiative by an Indian Consortium, 2009.

[12] Zhihong Hu, Johan Schalkwyk, Etienne Barnard, Ronald Cole, "*Speech Recognition Using Syllable-Like Units*," Center for Spoken Language Understanding, Oregon Graduate Institute of Science and Technology, September 2008, pp. 218-222.

[13] R.G. Bachu, S. Kopparthi, B. Adapa, B.D. Barkana, " *Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal*," Electrical ¬Engineering Department School of Engineering, University of Bridgeport, March 2010, volume 7340, 2012, pp. 539-546.