

# Automatic Segmentation of Punjabi Speech into Syllable-Like Units using Group Delay A Review

Anupriya Sharma  
CSE Deptt  
RIMT-IET  
Mandi Gobindgarh  
er.anupriya33@gmail.com

Amanpreet Kaur  
CSE Deptt  
RIMT-IET  
Mandi Gobindgarh  
er.amanpreet.cse@gmail.com

**Abstract:** The basic building blocks of a speech segmentation system are its units. Thus it's an important stage to select appropriate units into which the continuous speech needs to be segmented. The syllable like units is found to be the better representative for Indian languages. Punjabi is the most widely used language, thus this paper describes the automatic segmentation of Punjabi speech into syllable like units using group delay method. The STE (Short Term Energy) and ZCR (Zero Crossing Rate) are used to evaluate the basic characteristic of speech. In case of continuous speech group delay approach is used i.e. the negative derivative of Fourier transformations.

**Keywords:** Short Term Energy, Zero Crossing Rate, Automatic speech segmentation, Group delay, Syllable units, Punjabi Syllables.

## I. INTRODUCTION

Automatic speech recognizers (ASR) are used to facilitate communication between humans and machines. The Speech recognizer understands humans and their spoken words. It is required to perform the process of segmentation for automatic recognition of continuous speech. As the Segmentation of acoustic signal into basic units is an important stage. Consequently, the syllables are very important unit of language. The syllable consists of vowels and consonants. The signal information in the speech can be extracted by short-time Fourier analysis. For the detection of syllable boundary the group delay function is a better representative of the STE function.

## II. UNITS FOR SEGMENTATION

The basic units of speech recognition and segmentation are Words, Phonemes, or syllables. The choice of representative units is made depending on the size of vocabulary. One of the most natural units of segmentation is Word. Due to lack of generalization and more memory consumption words are not selected as units of segmentation. The smallest segmental units of sound are Phoneme. Due to overgeneralization of phonemes, they are not selected for the automatic speech segmentation. The next level basic unit of speech is syllables. Syllable is defined by rules, a syllable must have a vowel called its nucleus, and the presence of consonant is optional.

## III. UNITS OF SPEECH FOR PUNJABI LANGUAGE

The syllable comprises vowel and consonants. The presence of vowel is must in a syllable. The vowel is the nucleus, 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.

The seven types of syllables recognized in Punjabi language are as follows:

V, VC, CV, VCC, CVC, CCVC, CVCC

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

ਸ ਹ ਕ ਖ ਗ ਘ ਙ ਚ ਛ ਜ ਝ ਵ  
ਟ ਠ ਡ ਟ ਠ ਡ ਤ ਥ ਦ ਧ ਨ ਪ ਫ ਬ  
ਭ ਮ ਯ ਰ ਲ ਵ ਝ ਸ ਖ ਗ ਜ ਛ ਲ

List of Non-Nasal Vowels:

ਈ ਇ ਏ ਐ ਅ ਆ ਔ ਊ ਊ ਓ

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

TABLE 1: LIST OF SYLLABLES IN PUNJABI

V	Vowel	ਅ	ਅ
VC	Vowel+ Consonant	ਇ+ਹ	ਇਹ
CV	Consonant +Vowel	ਜ+ਆ	ਜਾ
VCC	Vowel+ Consonant+ Consonant	ਪ+ ਗ+ ਗ	ਪਗ
CVC	Consonant +Vowel+ Consonant	ਬ+ਆ+ਤ	ਬਾਤ
CCVC	Consonant + Consonant +Vowel+ Consonant	ਹ+ਨ+ਏ+ਰ	ਹਨੇਰ
CVCC	Consonant +Vowel+ Consonant+	ਪ+ਊ+ਰ+ਬ	ਪੂਰਬ

#### IV. FEATURES OF SPEECH

The verbal part of speech which contains the information of speech can be divided into two categories: Voiced and Unvoiced speech. Voiced sounds are generated when air from the lungs passes through the larynx where as Unvoiced speech sounds are produced by air passed directly through vocal tract formations. Unvoiced speech sounds are characterized by a noise-like signal. Silence is an integral part of speech signal. The voiced and unvoiced speech is separated by a silence region.

#### VI. REPRESENTATION OF SPEECH

Conventionally, the information in speech signal is represented in terms of features derived from Fourier analysis. When using the Fourier transform the result is a very exact analysis of its frequency contained in the signal, but no information about the time of occurrence of features and about the scale characteristics of the signal. The amount of detail in the signal can be measured by the scale (the frequency content of signal).

Short-time Fourier analysis can be used to derive the information in speech signals. By processing the negative derivative of the FT phase, i.e., the group delay function the information in the short-time Fourier Transform phase function can be extracted [9].

$$H(\omega) = H1(\omega) \cdot H2(\omega), \tag{1}$$

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

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

$$= \tau_{h1}(\omega) + \tau_{h2}(\omega). \tag{2}$$

The multiplicative property of magnitude spectra is shown in equation (1), whereas in group delay domain it becomes an addition as shown in equation (2). Due to the additive property of group delay spectrum it has been declared better than magnitude spectrum. The main problem faced in case of magnitude spectra was that the peaks are not resolved in a system where the two poles are combined together. On the other hand in case of group delay spectra the peaks and valleys are better resolved when the signal is in minimum phase [2].

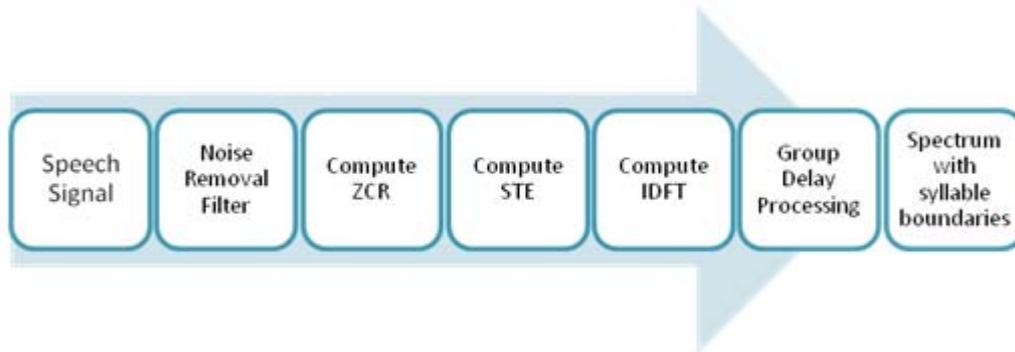


Figure 1: Steps for finding syllable boundaries

## VII. METHODS OF AUTOMATIC SPEECH SEGMENTATION

There are two ways of segmentation one is Manual segmentation (hand labeling) and other is automatic segmentation (ASR). It has been observed in some studies that the method was implemented and analyzed for different Punjabi speech signals, the deviation between manual and automatic segmentation had been calculated for the onset and offset values for the syllable boundaries. It was proved with the results that the boundaries of syllables were marked automatically with accuracy, thus the automatic segmentation technique was found to be the better in comparison to the manual segmentation [7].

## VII. CONCLUSIONS

According to the survey the syllables are found to be the best units for the automatic segmentation of speech as these are best suited for Indian Languages. Fourier analysis is used to represent the information in the speech signal. The negative derivative of Fourier transformation i.e. group delay approach is best suited for the process of segmentation for the continuous speech. In order to make the system capable for recognition it is required to maintain a speech database.

## 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 telecommunications 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.