Global Journal of Computer Science and Technology
Software & Data Engineering
Volume 13 Issue 12 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Automatic Segmentation of Punjabi Speech Signal using Group Delay

By Anupriya Sharma & Amanpreet Kaur
RIMT, Gobindgrah, India

Abstract- This paper describes the concept of automatic segmentation of continuous speech signal. The language used for segmentation is the most widely spoken language i.e. Punjabi. Like all other Indian languages, Punjabi is a syllabic language, thus syllables are selected as the basic unit of segmentation. The traditional way of representing the speech signal is in terms of features derived from short-time Fourier analysis. It is difficult to compute the phase and processing the phase function from the FT phase. By processing the derivative of the FT phase, the information in the short-time FT phase function can be extracted. This paper describes the process of automatic segmentation of speech using group delay technique. This includes segmentation of continuous Punjabi speech into syllable like units by using the high resolution properties of group delay. This group delay function is found to be a better representative of the STE function for syllable boundary detection.

Keywords: *digital signal processing, speech signal, automatic speech segmentation, punjabi speech segmentation, asr, ste, syllables, units of speech.*

GJCST-C Classification : *H.5.5*

Strictly as per the compliance and regulations of:

© 2013. Anupriya Sharma & Amanpreet Kaur. This is a research/review paper, distributed under the terms of the Creative Commons Attribution-Noncommercial 3.0 Unported License http://creativecommons.org/licenses/by-nc/3.0/), permitting all non-commercial use, distribution, and reproduction inany medium, provided the original work is properly cited.

Automatic Segmentation of Punjabi Speech Signal using Group Delay

Anupriya Sharma [α] & Amanpreet Kaur [σ]

Abstract- This paper describes the concept of automatic segmentation of continuous speech signal. The language used for segmentation is the most widely spoken language i.e. Punjabi. Like all other Indian languages, Punjabi is a syllabic language, thus syllables are selected as the basic unit of segmentation. The traditional way of representing the speech signal is in terms of features derived from short-time Fourier analysis. It is difficult to compute the phase and processing the phase function from the FT phase. By processing the derivative of the FT phase, the information in the short-time FT phase function can be extracted. This paper describes the process of automatic segmentation of speech using group delay technique. This includes segmentation of continuous Punjabi speech into syllable like units by using the high resolution properties of group delay. This group delay function is found to be a better representative of the STE function for syllable boundary detection.

Keywords: digital signal processing, speech signal, automatic speech segmentation, punjabi speech segmentation, asr, ste, syllables, units of speech.

I. Introduction

The Automatic speech recognizers (ASR) are used to facilitate communication between humans and machines. So it's a machine which understands human and the words spoken by them. The process of segmentation is one of the most important phases in the automatic recognition of speech. There are various units of speech into which it can be segmented, but syllables are found to be one of the most efficient units for automatic speech segmentation. The characteristics features of speech can be expressed by using STE and ZCR. The STE function also known as Short Term Energy function is known to be the better representative of speech segment boundaries. By computing the short-time Fourier analysis information in the speech signal can be extracted. But due to difficulty in computing the phase and also in processing the phase function over the past few decades the features of the FT phase were not exploited fully. By processing the derivative of the FT phase, the information in the short-time FT phase function can be extracted. There are various units of speech. The syllables are found to be the most suitable unit for automatic speech segmentation. A single component in the syllable is called as nucleus. The nucleus is found to be vowel while the onset and coda are usually consonantal in form. The energy peak in the nucleus region can be viewed as the syllable; the consonants can be viewed as the valleys at both the ends. Many languages been spoken around the world posses a syllabic structure [10]. Mostly the syllable contains two phonetic segments of type CV such as in Japanese language. In contrast, English and German possess a more highly heterogeneous syllable structure [2].

Author α: Department of computer science and engineering RIMTMandi Gobindgarh, India. e-mail: er.anupriya33@gmail.com

Author σ: Department of computer science and engineering BBSB Fatehgarh Sahib, India. e-mail: er.amanpreet.cse@gmail.com

II. Research Background

a) Language Units of Speech in Punjabi

Punjabi is an Aryan language that is spoken by more than hundred million people those are inhabitants of the historical Punjab region (in north western India and Pakistan) and in the Diaspora, particularly Britain, Canada, North America, East Africa and Australasia [8].

Like other Indian languages the Punjabi language also contains segmental phonemes. The three basic units into which the speech can be segmented are: Words, Phonemes and Syllables. The syllable is the most important and widely used unit for automatic speech segmentation. Punjabi is a syllabic language thus syllables are selected as the basic units for segmentation.

b) Syllables as Basic unit of speech

Aksharas is the basic units of the writing system. An Akshara is an orthographic representation of a speech sound in an Indian language. Basically they are syllabic in nature; the typical forms of akshara are V, CV, CCV and CCCV type, where C and V are consonant vowel respectively [9]. There are thirty eight consonants in Punjabi language. Where ten are non-nasal and ten are nasal vowels. Vowels can appear alone but consonants can only appear with vowels. The number of nasal vowels is same as non-nasal vowels and is represented by Bindi or Tippi over the Non-Nasal Vowels. Following is the list of consonants in Punjabi language:

© 2013 Global Journals Inc. (US)

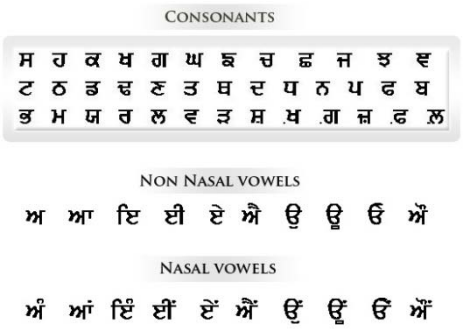

Figure 1 : Respresentation of 38 consonants and 20 vowels in Punjabi Language

As already mentioned syllables are the basic and most recommended used units of speech. Syllables are composed of vowel and consonants. Every syllable must have a vowel also known as its 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 which is always a consonant. The seven types of syllables recognized in Punjabi language are represented in the following figure:

Syllable	Syllable Description	Word	Segments	Word	Segments
V	Vowel	ਓ	ਓ	ਈ	ਈ
VC	Vowel-consonant	ਉਡ	ਉ+ ਡ	ਇਸ	ਇ+ ਸ
CV	Consonant-Vowel	ਜਾ	ਜ + ਆ	ਗਾ	ਗ + ਆ
VCC	Vowel-consonant-consonant	ਉਤਰ	ਉ+ ਤ+ਰ	ਅੰਦਰ	ਅੰ +ਦ+ਰ
CVC	Consonant-vowel-consonant	ਰਾਤ	ਰ + ਆ + ਤ	ਬਾਤ	ਬ + ਆ + ਤ
CVCC	Consonant-vowel-consonant-consonant	ਜੋਤਸ਼	ਜ + ਓ + ਤ +ਸ਼	ਪੁਰਬ	ਪ + ੁ + ਰ + ਬ
CCVC	Consonant-consonant-vowel-consonant	ਤਰੇਲ	ਤ + ਰ + ੇ +ਲ	ਸਵੇਰ	ਸ + ਵ + ੇ + ਰ

Figure 2 : Syllables in Punjabi language

III. Three State Respresentation of Speech

The continuous speech signals composed of two elements one includes the speech information, and the other carries noise or silent sections. The verbal part of the speech can be further divided into two categories: voiced and unvoiced speech.

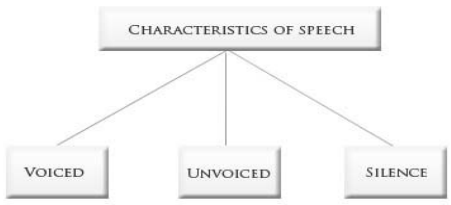

Figure 3 : Block diagram of characteristic features of voice

Moment the air from the lungs passes through the larynx voiced sound is produced. With the passage of air directly through the vocal tract formations the unvoiced speech sounds are produced. The speech production process is incomplete without the detection of voiced and unvoiced speech that is separated by a silence region. In case of silence region no excitation is supplied to the vocal tract and thus, no speech is produced. A regular speech is incomplete inaccurate without silence region. It helps to make the speech understandable [3].

IV. Characterization of Speech

In order to segment continuous speech it is required to check its basic content, whether the signal is voiced or unvoiced. The two characteristics features of voice are the zero crossing rate (ZCR) and short term energy (STE) [13].

a) Zero Crossing Rate

The rate at which the signal crosses zero provides the information regarding its (source of creation) i.e. zero crossing rate. Unvoiced speech has higher zero crossing rate. Whereas in case of voiced speech the zero crossing rate is low. Thus, the amplitude of unvoiced segments is lower than that of the voiced segments.

ZCR can be defined as:

$$Z_n = \sum_{m=-\infty}^{\infty} \left|\text{sgn}[x(m)] - \text{sgn}[x(m-1)]\right| w(n-m) \quad (1)$$

Where

$$\text{sgn}[x(n)] = \begin{cases} 1, x(n) \geq 0 \\ -1, x(n) < 0 \end{cases}$$

b) Short Term Energy (STE)

Short-time energy of speech signals reflects the amplitude variation. By processing STE function the speech can be segmented. STE shows the voiced content of the signal [13].

The STE can be defined as follows:

$$E_n = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2 \quad (2)$$

The STE of voiced signal is always much greater than that of unvoiced signals. In a speech signal where there are voiced signal its STE will be high, the peaks in the signal represents nucleus that is denoted as vowel where as the valleys at both the ends represents the coda.

V. Segmentation of Speech

The syllable is composed of three parts, the onset, rime (nucleus) and coda. The rime also known as

© 2013 Global Journals Inc. (US)

nuclei, where as the onset and coda consist of consonants. The high energy regions are represented by the nuclei where as the valleys at both ends corresponds to syllable boundaries. The vowel region corresponds to much higher energy region compared to that of a consonant region [9]. In case of spontaneous speech, the definition of a syllable in terms of short-term energy function is suitable for almost all the languages.

Due to local energy fluctuations the STE function alone cannot be directly used to perform segmentation. Techniques such as fixed or even adaptive threshold will not work when the energy variation across the signal is quite high [1].

To overcome the problems of local energy fluctuations, the STE function should be smoothed. 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 computing the group delay function [9].

$$H(\omega) = H1(\omega) \bullet H2(\omega), \tag{3}$$

group delay function can be represented as

$$\tau h(\omega) = -\frac{\partial(\arg(H(\omega)))}{\partial\omega} = \tau h1(\omega) + \tau h2(\omega). \tag{4}$$

The equation (1) shows the multiplicative property of magnitude spectra where as equation (2) is in group delay domain it shows an addition. The group delay spectrum has been found better due to its additive. It was observed that in case of the magnitude spectra the peaks are clearly visible, but when the two poles are combined together the peaks are not resolved. 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].

For any syllable, the STE function of the voiced region, the energy is quite high and diminishes at the ends, representing the consonants, due to which local energy fluctuations. If these local variations are smoothed, then the minima at both ends of a voiced region correspond to syllable boundaries [9].

The algorithm for group delay based segmentation

Step 1 - Let x[n] be continuous speech signal.

Step 2 - Compute N, the length(x) of the input signal.

Step 3 - Calculate the STE function E[m], where m=1,2,…,M is the number of frames.

Step 4 - Inverse the STE i.e E(i)= 1/E(m)

Step 5 - Compute the IFFT of E(i), It gives the magnitude of the input signal in form of complex function i.e. a+ib.

Step 6 - The phase angle is computed from the above values, i.e. ϕ= tan-1(b/a).

Step 7 - Compute the negative derivative of Fourier transformation i.e. the group delay function.

Step 8 - Compute the minimum phase of group delay, i.e. phase(n) - phase(n - 1), let the signal be of length n. Locate the positive peaks in the minimum phase group delay function, (Ei gd[f]). If Ei gd[f] is positive, and Ei gd[f-1]<Ei gd[f]<Ei gd[f+1] then Ei gd[f] is considered as a peak. These peaks represent the syllable boundaries.

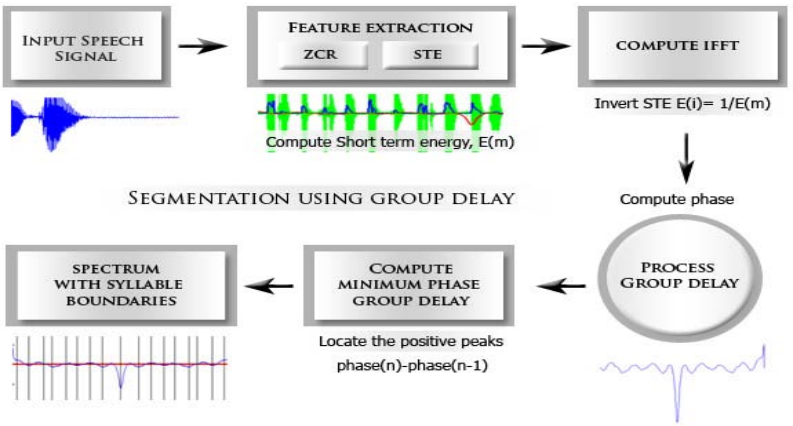

Figure 4 : Steps involved in finding syllable boundaries

VI. Results and Discussions

The technique of automatic segmentation is applied on the continuous Punjabi speech. The method was implemented in Matlab. The group delay algorithm is applied to segment the continuous Punjabi speech waveform. The following sentence is given as an input to the system.

ਅੰਮ੍ਰਿਤਸਰ ਸਿੱਖਾਂ ਦਾ ਸਭ ਤੋਂ ਉੱਚਾ ਧਾਰਮਿਕ ਸਥਾਨ ਹੈ

The system has efficiently marked the syllable boundaries. The onset and offset values are shown in the following table.

© 2013 Global Journals Inc. (US)

Tablel 1 : Results of Segmentation Obtained with Onset and Offset Syllable Values

Sentence	Onset	Offset	Duration
AMimR	0.576	1.664	1.088
qsr	1.664	2.88	1.216
isW	2.88	3.904	1.024
KF	3.904	5.056	1.152

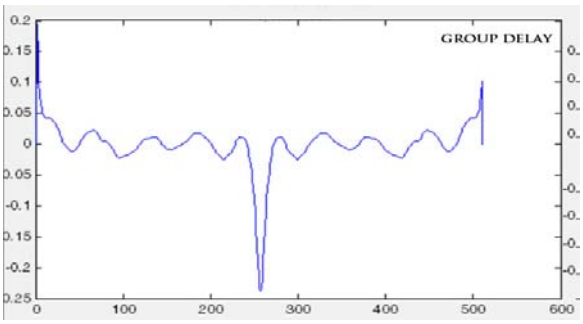

Figure 5 : Signal Representing the Group delay

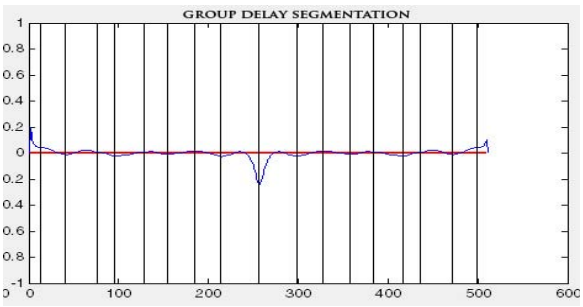

Figure 6 : Group delay based segmentation with marked syllable boundaries

References Références Referencias

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. 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.
14. B. Shneiderman, "The Limits of Speech Recognition," Communications of the ACM, vol. 43, no. 9, pp. 63–65, September 2000.
15. O. Fujimura, "Syllable as a unit of speech recognition," in IEEE Trans.Acoust., Speech, Signal Processing, vol. 23, February 1975, pp. 82–87.
16. J. L. Gauvian, "A syllable-based isolated word recognition experiment," in IEEE International Conf. on ICASSP'86, vol. 11, April 1986, pp. 57–60.

© 2013 Global Journals Inc. (US)