Volume 5, Issue 4, 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Detection of Syllables in Continuous Punjabi Speech Signal and Extraction of Formant Frequencies of Vowels

Parwinder Kaur* CSE Department, RIMT India
Ranpreet Kaur CSE Department, RIMT India
Amanpreet Kaur CSE Department, BBSBEC India

Abstract— Speech is the fastest means of communication. If speech can be used to communicate with computer to give commands, it will be very easy and helpful, especially for people with certain handicaps to use computer. This has been made possible with automatic speech recognition systems. The integral part of such systems is automatic speech segmentation systems which segment the speech into its subunits like words, syllables or phonemes and then these segments are recognized using some effective procedure. In this paper, Punjabi speech signal has been segmented into syllable like units using short term energy and zero crossing rate and then formant frequencies of vowels are extracted which can be later useful for recognition purpose.

Keywords— ASR, Short term energy, Zero crossing rate, Formant frequencies, LPC.

I. INTRODUCTION

Speech is the most basic, common and efficient form of communication method for people to interact with each other. People are comfortable with speech therefore persons would also like to interact with computers via speech, rather than using primitive interfaces such as keyboards and pointing devices. This can be accomplished by developing an Automatic Speech Recognition (ASR) system which allows a computer to identify the words that a person speaks into a microphone or telephone and convert it into written text. Therefore, speech segmentation system is required which is the integral part of speech recognition system. Speech segmentation is the process of identifying the boundaries between words, syllables, or phonemes in spoken natural languages. The term applies both to the mental processes used by humans, and to artificial processes of natural language processing. Speech segmentation is an important sub problem of speech recognition, and cannot be adequately solved in isolation. Speech can be categorized into three activities.

A. *Voiced speech*

Vocal cords are tensed and vibrating periodically, resulting in a speech waveform that is quasi-periodic. All the information that is to extract is contained in the voiced part of the speech signal.

B. *Unvoiced speech*

Vocal cords are not vibrating, resulting in aperiodic or random speech waveform.

C. *Silence*:

When no speech is produced. Voiced and unvoiced region is usually separated by silence region. There is no waveform in the silence region of the speech signal. Hence this part can be easily detected in the signal.

II. SEGMENTATION UNITS

Most of time word is considered to be the most natural unit of speech. Every word in Punjabi or any language has its well defined boundaries. But there are other problems that arise by using word as a speech unit. Each word has to be trained individually and there any sharing of parameters cannot be possible among words. Therefore, it is essential to have a very large training set so that all words in the vocabulary are adequately trained. In addition to these, more memory is also required as the number of words grow which in turn increases the problem of memory management. So choosing word as a basic unit for segmentation is not a good choice. Another option for the segmentation unit is phoneme. There are about 50 phonemes in a language. So it is easy to train with a training set of reasonable size. It is a well known fact that the same phone in different words has different realizations [6]. The realization of a phone is strongly affected by its adjacent phones or in other words, phones are highly context dependent. Therefore, the acoustic variability of basic phonetic units due to context is sufficiently large and is not well understood in many languages. Thus, it can be observed that there is overgeneralization in phone models while word models lack in generalization [7]. So it is clear from the discussion that we need the segmentation unit that is in between word & phonemes i.e. third fundamental unit syllables. Combination of phonemes gives rise to next higher unit called syllables which is one of the most important units of a language. A syllable must have a vowel called its nucleus, whereas presence of consonant is optional [4].

TABLE I PUNJABI SYLLABLES

Type	Pattern	Example
V	Vowel	ਆ
VC	Vowel-Consonant	ਉਚ (ਉ+ਚ)
CV	Vowel-Consonant	ਖਾ (ਖ+ਆ)
VCC	Vowel-Consonant-Consonant	ਅੰਗ (ਅ+ਙ+ਗ)
CVC	Consonant-Vowel-Consonant	ਗੀਤ (ਗ+ੀ+ਤ)
CCVC	Consonant-Consonant-Vowel-Consonant	ਪ੍ਰੀਤ (ਪ+ਰ+ੀ+ਤ)
CVCC	Consonant-Vowel-Consonant-Consonant	ਧਰਮ (ਧ+ਰ+ਮ)

A. Punjabi vowels

There are ten types of vowels as shown in the figure below. Vowels are characterized by:

1) *Tongue height*: The tongue is the main articulator for vowels. Therefore, it is important on where it is placed. The height can be characterized by being either high, mid, or low.
2) *Tongue advancement*: The tongue is also characterized by being either front, central, or back.
3) *Lip rounding*: The lips are characterized by being either retracted or rounded. This is shown in the difference between "moon" and "mean".
4) *Tense/lax vowels*: The vowels can also be characterized by being either a tense vowel or a lax vowel. Tense vowel tends to be longer in duration and may require more effort.

TABLE II PUNJABI VOWELS

Vowel			Name		IPA
Ind.	Dep.	with /k/	Letter	Unicode	
ਅ	(none)	ਕ	Mukta	A	[ə]
ਆ	ਾ	ਕਾ	Kanna	AA	[ɑ]
ਇ	ਿ	ਕਿ	Sihari	I	[ɪ]
ਈ	ੀ	ਕੀ	Bihari	II	[i]
ਉ	ੁ	ਕੁ	Onkar	U	[ʊ]
ਊ	ੂ	ਕੂ	Dulankar	UU	[u]
ਏ	ੇ	ਕੇ	Lavan	EE	[e]
ਐ	ੈ	ਕੈ	Dulavan	AI	[ɛ]
ਓ	ੋ	ਕੋ	Hora	O	[o]
ਔ	ੌ	ਕੌ	Kanuara	AU	[ɔ]

III. FORMANT FREQUENCY OF VOWELS

A Formant is a concentration of acoustic energy around a particular frequency in a speech wave**.** In other words, these are the meaningful & distinguishable frequency components. There are several formants, each at a different frequency, roughly one in each 1000 Hz band. Each formant corresponds to a resonance in the vocal tract. The information that humans require to distinguish between vowels can be represented purely quantitatively by the frequency content of the vowel sounds. The formant with the lowest frequency is called f_1, the second f_2, and the third f_3. Most often the two first formants, f_1 and f_2, are enough to disambiguate the vowel. These two formants determine the quality of vowels in terms of the open/close and front/back dimensions (which have traditionally, though not entirely accurately, been associated with the position of the tongue). Thus the first formant f_1 has a higher frequency for an open vowel (such as [a]) and a lower frequency for a close vowel (such as [i] or [u]); and the second formant f_2 has a higher frequency for a front vowel (such as [i]) and a lower frequency for a back vowel (such as[u]). Vowels will almost always have four or more distinguishable formants; sometimes there are more than six. However, the first two formants are most important in determining vowel quality, and this is often displayed in terms of a plot of the first formant against the second formant, though this is not sufficient to capture some aspects of vowel quality. Spectrograms are used to visualize formants. In spectrograms shown in figure1, the dark energy bands are the formants of their respective vowels.

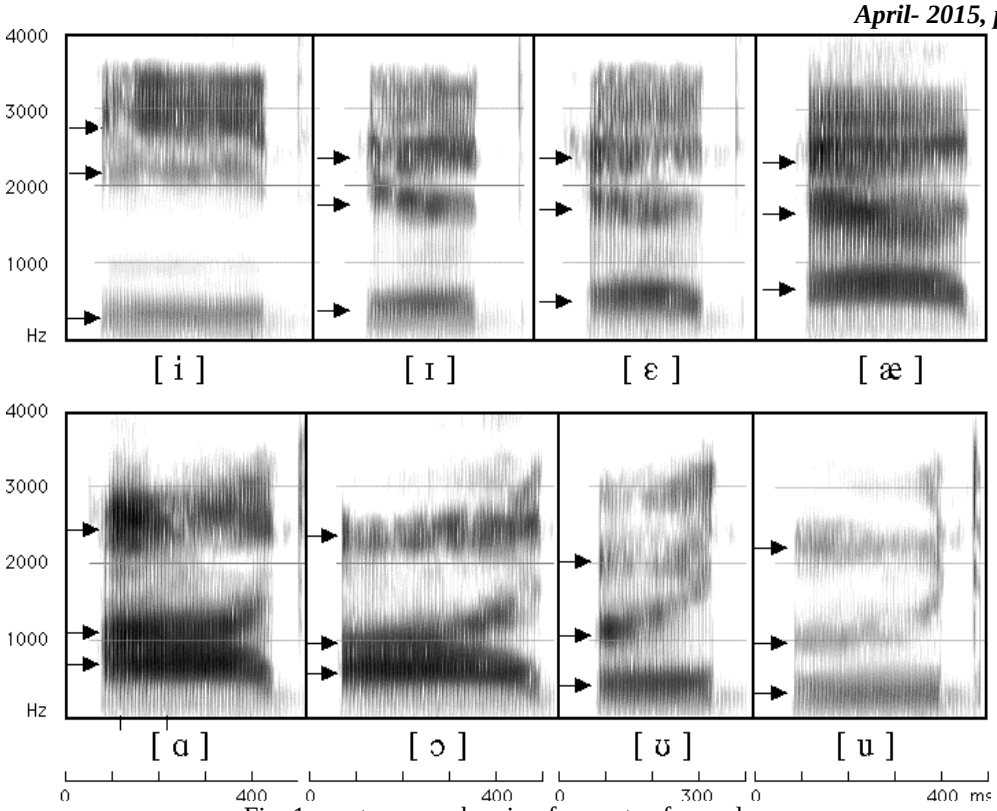

Fig. 1 spectrogram showing formants of vowels

The first formant (F1) in vowels is inversely related to vowel height. The higher is the vowel, the lower is the first formant (and vice versa). The second formant (F2) in vowels is somewhat related to degree of backness. The more front is the vowel, the higher is the second formant (but affected by lip-rounding). According to height and degree of backness, vowels are classified as shown in figure 2.

	Front	Central	Back
High	i ɪ		ʊ u
High-mid	e		o
Mid		ə	
Low-mid	ɛ		ɔ
Low		a	

Fig.2 classification of vowels

IV. METHODOLOGY

The proposed work has been divided into two parts. First part is automatic speech segmentation system and the second part is formant frequency estimation of vowels in segmented files. The automatic speech segmentation system consists of the following steps:

A. Speech recording

The continuous Punjabi speech signal is recorded using mice with the help of Sound Forge software at 16000 hz, 8 bit and mono channel.

B. Preprocessing

The recorded speech signal is preprocessed to make it suitable for further segmentation process. The preprocessing of speech signal includes the following steps:

1) *Noise removal:* Background noise elimination is the first step in the signal processing. By this process, background noise is removed from the data so that only speech samples are the input to the further processing.
2) *Pre-emphasis filtering:* Pre-emphasis of the speech signal at higher frequencies is a preprocessing step employed in various speech processing applications. Pre-emphasis of the speech signal is achieved by first ordering differencing of the speech signal. Although possessing relevant information, high frequency formants have smaller amplitude with respect to low frequency formants. A pre-emphasis of high frequencies is therefore required to obtain similar amplitude for all formants. This is usually obtained by filtering the speech signal with a first order FIR (Finite Impulse Response) filter, known as pre-emphasis filter.

3) *Framing:* In most processing tools, it is not appropriate to consider a speech signal as a whole for conducting calculations. A speech signal is often separated into a number of segments called frames. This process of separation is known as framing. Continuous speech signal has been blocked into N samples, with adjacent frames being separated by M (M<N). After the pre-emphasis, filtered samples have been converted into frames, having frame size of 20 msec. Each frame overlaps by 10 msec.
4) *Windowing:* The window, w (n), determines the portion of the speech signal that is to be processed by zeroing out the signal outside the region of interest. To reduce the edge effect of each frame segment windowing is done. Rectangular window has been used.

C. Segmentation

Step 1-The unvoiced and silence region of the signal is set to zero using zero crossing rate feature.
Step2- The short time energy of the speech signal is computed.
Step3-Then the signal is segmented into syllables using some threshold value of short term energy of signal.
Step4-Spectrum with marked syllable boundaries is stored in the database and also segmented wav files are saved.
Step5- Formant frequency algorithm is applied to these segmented files one by one.
Step6-Formant frequency values of first three formants are stored in excel sheet for later use in the process of recognition.

FORMANT FREQUENCY ESTIMATION ALGORITHM:

Step1- Load the segmented wave file.
Step2- Extract the segment from 0.1 sec to 0.15 sec.
Step3- Apply the hamming window on the segment.
Step4- Then preemphasis filter is applied to emphasize the high frequency components of the signal.
Step5-Calculate the LPC coefficients of predictor polynomial using formula 2n+2, where n is the number of formant frequencies required.
Step6-Then roots of the predictor polynomial are calculated which gives the required formant frequencies of vowel in the syllable.

V. RESULTS AND DISCUSSION

The speech signal containing the following sentence has been recorded for segmentation and formant frequency estimation.

1. ਆਪਣਾ ਕਮ ਸਮੇ ਸਿਰ ਕਰਨ ਦੀ ਆਦਤ ਪਾਓ ਤਾਂ ਹੀ ਕਾਮਯਾਬ ਹੋਵੋਗੇ।

This above mentioned speech signal has gone through various steps of segmentation and the starting and end points of syllables with its duration has been calculated by the system are shown below in table 3.

TABLE III START AND END POINTS OF SYLLABLES

syllables	start	End	Duration
ਆਪ	0.8954	1.1516	0.2562
ਣਾ	1.36	1.499	0.14
ਕਮ	2.119	2.2978	0.1788
ਸਮੇ	2.895	3.0041	0.1091
ਸਿਰ	3.7013	3.9086	0.2073
ਕਰਨ	4.5012	4.6378	0.1366
ਦੀ	5.4303	5.5359	0.1057
ਆਦ	6.2971	6.4711	0.174
ਪਾ	7.184	7.5611	0.377
ਤਾਂ	8.1944	8.5044	0.31
ਹੀ	9.0244	9.1465	0.1221
ਕਾਮ	9.8395	10.098	0.2585
ਯਾਬ	10.702	11.048	0.346
ਹੋ	11.728	11.868	0.1400
ਵੋ	12.1761	12.3428	0.1666

ਗੋ	12.4878	12.6031	0.1153

The second step is formant frequency estimation of these segmented wav files. The output of the ASS is fed as input to the formant frequency estimation algorithm. The following table 4 shows the formant frequency values extracted for the vowels in these segmented wave files.

TABLE IV FORMANT FREQUENCY VALUES

F/S	ਆਪ	ਕਮ	ਸਿਰ	ਆਦ	ਪਾ	ਤਾਂ	ਕਾਮ	ਯਾਬ	ਵੋ
F1	986	388.5	494	798	1119	561	388	759	444
F2	1367	1241.3	2164	1636	2080	1250	1241	1546	914
F3	2212	2396.8	2898	2638	3031	2251	2396	2404	1801

The formant frequency values of segmented files which are smaller than 15 ms have not been generated by this algorithm. This can be the future scope of this work how to calculate formant frequencies of these small files and then this data can be used in vowel recognition later.

REFERENCES

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

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

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

[4] C. Vimala and V.Radha, "A Review on Speech Recognition Challenges and Approaches", World of Computer Science and Information Technology Journal, Vol. 2, No.1, 2012, pp. 1-7.

[5] Prica Biljana et al. "Recognition of vowels in Continuous Speech by using Formants", Facta Universitatis, Vol. 23, No. 3, December 2010, pp. 379-393.

[6] Anwar Jamil Muhammad et al. "Automatic Arabic Speech Segmentation System", International Journal of Information Technology", Vol. 12, No. 6, 2006, pp. 102-111.

[7] Natarajan Anantha V. et al. "Segmentation of Continuous Speech into Consonant and Vowel Units using Formant Frequencies", International Journal of Computer Applications, Vol. 56, No.15, October 2012, pp. 24-27.

[8] Rao Preeti et al. "Speech formant frequency estimation: evaluating a nonstationary analysis method", Elsevier, 2000, pp. 1655-1667.

[9] Rahman Mijanaur Md. et al. "Continuous Bangla Speech Segmentation using Short-term Speech Features Extraction Approaches", International Journal of Advanced Computer Science and Applications, Vol. 3, No. 11, 2012,pp. 131-138.

[10] Greenberg S. "Strategies for Automatic multi-tier annotation of spoken language corpora", Proceedings of the 8th European Conference on Speech Communication and Technology, 2003, pp. 45–48.

[11] Sharma M. & Mammone R. "Blind speech segmentation: Automatic segmentation of speech without linguistic knowledge", ICSLP Proceedings, Vol. 2, 1996, pp. 1237-1240.

[12] Tolba F. M. et al. "A Novel Method for Arabic Consonant/Vowel Segmentation using Wavelet Transform", International Journal of Intelligent Computing and Information Sciences, Vol. 5, No. 2, 2005, pp. 353-364.

[13] 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

[14] Thangarajan R, Natarajan A.M. "Syllable Based Continuous Speech Recognition for Tamil", in South Asian Language Review VOL. XVIII. No. 1, 2008.

[15] H. Iqbal, M. Awais, S. Masud, and S. Shamail, "On vowels segmentation and identification using formant transitions in continuous recitation of quranic arabic," in *New Challenges in Applied Intelligence Technologies*, ser. Studies in Computational Intelligence. Springer Berlin / Heidelberg, 2008, vol. 134, pp. 155–162.

[16] Y. A. Alotaibi and A. Hussain, "Comparative analysis of arabic vowels using formants and an automatic speech recognition system," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 3, 2010.

[17] S. A. M. Yusof, P. M. Raj, and S. Yaacob, "Speech recognition application based on malaysian spoken vowels using autoregressive model of the vocal tract," in *Proceedings of the International Conference on Electrical Engineering and Informatics*. Bandung, Indonesia: Institut Teknologi Bandung, June 2007.

[18] G. N. Kodandaramaiah, M. N. Giriprasad, and M. M. Rao, "Independent speaker recognition for native english vowels," *International Journal of Electronic Engineering Research*, vol. 2, 2010.

[19] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, 2005.

