IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Volume-3, Special Issue-1 E-ISSN: 2347-2693

Word Boundary Detection in Assamese Language

Sudarshana Sarma*
Dept. of Computer Sc. & Engineering and IT
Assam Don Bosco University
Guwahati, India
sudarshanasarma8@gmail.com

Uzzal Sharma
Dept. of Computer Sc. & Engineering and IT
Assam Don Bosco University
Guwahati, India
uzzal.sharma@dbuniversity.ac.in

***Abstract*— Detection of word boundary in continuous speech is an important issue in speech recognition. This paper presents different methods for detecting word boundary in Assamese language and also discusses the possible future development to improve the accuracy. These methods are based on the behavior of pitch frequency across the sentences. Here we use Scilab, Praat and Wavesurfer software.**

Keywords—Word boundary detection(WBD),Hidden markov model(HMM), LPC, SOM, MLP

I. Introduction

Word boundaries are conventionally represented by pause between words. Word boundary detection (WBD) is very common and important issue in the field of speech synthesis and recognition. The recognition of continuous speech presents the listeners (human or machine) with a problem which does not arise in the recognition of isolated words. Recognition involves the conversion of a speech utterance to its orthographic representation.

This process involves [1]:

1. Segmentation of the speech utterance into lexical items (words).
2. Matching the items with the units present in the memory.
3. Performing an action depending upon the semantics of the unit matched.

Difficulties [1] are discussed below:

Language Independence: Several language features can be exploited for word boundary detection. But, clearly these are language dependent, and cannot be applied commonly for all languages .The method developed must use some universal characteristics of the languages rather than features that are language specific.

Length of word string:

The spoken utterance can either be some collected words, a phrase or a sentence. There is no constraint imposed on the length of the speech input. So a method used should be effective over all continuous speech input.

No explicit clues:

Continuous speech offers no explicit clues for placing the word boundaries. The speaker does not pause consciously between words to signify the word boundaries while speaking or reading continuously. So speech signal related clues must be used for marking the word boundaries.

I. EXISTING METHODOLOGY

A. Word Boundary Detection Using Pitch , intensity and duration

In paper [2], word boundary was detected using pitch frequency. Here found the behavior of the pitch frequency across the sentences. The pitch frequency is found to rise in a word and fall to the next word. They used word boundary hypothesization algorithm. According to the word boundary hypothesization algorithm [2], detect the peaks in the pitch contour of the utterance and hypothesize these vowels as the word final vowel. Then hypothesize a vowel as a word final vowel if its pitch frequency is larger than that of the next vowel. Their proposed method was defined two measures hit rate and false alarm rate. Hit rate means percentage of the word boundary correctly detected and false alarm rate means percentage of error in word boundary detection .Using pitch frequency, they had shown that more than 85% of the word boundaries are correctly detected in the Indian languages and more than 65% of the word boundaries are correctly detected in German languages.

The result of the paper study clearly indicate that the Indian language have similar prosody. Further studies can be conducted using a greater database in German language to establish the similarity between Indian and German language.

In paper [3], using two prosodic parameters, pitch and frequency, evaluate the performance of word boundary detection Hindi speech database. They proposed a word boundary detection algorithm where they analyzed intensity contour and pitch contour. They defined three prosodic parameters: "defined pitch contour", "undefined pitch on silence zone", and "intensity contour". Using a proposed word boundary detection algorithm [3], which analyzed intensity contour and pitch contour, 90.8% times algorithm recognized the correct word boundary and 80.1% times algorithm did not recognized the word boundary for actual non boundary.

In paper [4], using pitch and duration ,Word boundary was detected in continuous Hindi speech .Word boundary hypothesization technique[4] based on the durational knowledge for Hindi was used in this paper .Here also combined duration and pitch knowledge to show improvement of the accuracy of word boundary detection .The proposed algorithm was based on following feature of Hindi language . Firstly, In Hindi, very few words are ends in short vowel. Secondly, In Hindi, vowel occurs twice as often as consonant in position before a word boundary .Using pitch and duration detect word boundary in Hindi speech. Future task is to include third prosodic feature "intensity" in the word boundary hypothesization algorithm. Another task is to examine if the similar knowledge can be extracted for other Indian languages and to observe the similarities among the various Indian languages.

In Paper [11], using suprasegmental parameters of speech signal, word boundary was detected in continuous speech. Suprasegmental parameters are pitch, F0 fundamental frequency, duration, intensity and pause, which can play important role in finding some clues to detect the start and the end of the word from the spoken utterance of Hindi Language. They proposed an algorithm, which is based mainly on two prosodic parameters, pitch and intensity. A corpus is designed [11] consisting of a selection of 60 phonetically balanced and prosodically rich sentences from the interactive Hindi course books and Hindi newspapers. Total of 3 speakers, including one male and two female speakers, recorded 40 sentences each in five emotions like: Neutral, Happy,

Sad, Surprise, Anger (200 sentences per speaker and total 600 sentences). The sentences were recorded at 44.1 kHz using "PRAAT" speech synthesis software tool. The sentences consist of different categories like simple, complex, and interrogative, and declarative. Using of two important prosodic parameters pitch and intensity, they got strong clues to detect the start of the word, end of the word and no. of words in continuous speech of Hindi language. Future work will be to reduce the incorrect inclusion rate, and achieve more correction rate in word boundary detection algorithm and to realize the effects of other prosodic parameters on the algorithm of word boundary detection.

B. Word boundary detection using neural network

In paper [5], using time delay neural network, word boundary was detected in continuous speech .They used pruning algorithm. Pruning algorithm was used to the trained neural network to reduce over-fitting caused by limited data. They also used bootstrapping to achieve segmentation of new speech data. Time delay neural network was used to detect word boundaries in continuous speech. Future work will focus on combining the prediction of the TDNN with other method like EM clustering to improve the performance.

In paper [12], word boundary was detected using fuzzy logic. They proposed a fuzzy logic based boundary detection algorithm [12] that meets requirements of both computational simplicity and robustness to the background noise. The Fuzzy end point detector used a set of four simple differential parameters and a matching phase based on a set of fuzzy rule .The algorithm consists of pre-processing the speech signal, extracting its significant features, a matching phase, a post processing module and finally a decision block.

The Fuzzy End Point Detection Algorithm gives the most accurate result in word boundary detection than the traditional word boundary detection algorithm.

In paper [9], an effective method of detection of Assamese numerals under varied recording conditions and moods is presented here. The work also deals with gender variations. Combination of Linear Predictive Code (LPC) and Vector Quantization applied to a combination of Self Organizing Map (SOM) and Multi Layer Perception (MLP) based system. The SOM and MLPs are used to constitute a Learning Vector Quantization (LVQ) block which is necessitated by the fact that the LPC-VQ methods fail to produce the expected outcome while dealing with numerals of Assamese- a language spoken by a sizable population in the North-Eastern part of India.

In paper [10], using three layer perception network, classify the /i/ sound from different speakers. Classification accuracy is achieved 97%. A map of phonemes is used to trace trajectories of utterances. Self-organizing map is used to determine the inherent dimensionality of a set of points [10]. A three layer multilayer perception is used in this paper to classify one phoneme sound using a training set of isolated words spoken by different speaker. In this paper they also use self organizing map to segment a set of isolated digit.

C. Robust Word Boundary Detection in Spontaneous Speech

In paper [6], using acoustic and lexical cues, detect word boundary in spontaneous speech. The proposed system was used lexical feature along with the acoustic feature to improve the accuracy of the word boundary detection. Also used acoustic features: "short-time energy", "short-time zero crossing rate", "short-time pitch frequency" to detect word boundary. Using acoustic and lexical feature improves the word boundary detection accuracy.

D. A LPC-PEV Based VAD for Word Boundary Detection

Using LPC-PEV algorithm [7], word boundary was detected in the presence of noise. Here Yule Walker method is used to detect error. To identify the presence of speech in an input signal by marking the boundaries of speech and non speech segment, VAD is used. From future research aspect, it will also be interesting to study the LPC-PEV based VAD to improve the robustness of the speech recognition system and combine this feature with any margin based learning algorithm to improve the generalization capability of the acoustic model.

E. Foot Detection in Czech Using Pitch Information and HMM

Using pitch information and HMM [8], detect foot in Czech .This paper focused on modeling and detection of lexical stress group (foot) for Czech language and tried to train the Hidden Markov Model(HMM) for Czech feet information using only pitch information in the syllable nuclei. Although reached results are not much impressive, there is still a lot of possible future work and improvements [8] pending: 1) Collaboration with Czech language specialists (phonetics), which can lead into careful verification of used features and corresponding labels and also improvements of the used text-to-foot module. 2) Choice of features - except the pitch there are another features that are obtainable and worth to try (intensity, durations of syllables or vowels and spectral slope).

F. A Robust Algorithm for Detecting Speech Segments Using an Entropic Contrast

Using Entropy-Based Speech Segmentation Algorithm[13] ,detect word boundaries in noisy environment .This algorithm was to use a entropy based contrast function between the speech segments and the background noise .The entropy based contrast exhibits better behaved characteristics as compared to the energy-based method. An adaptive threshold is used to determine the candidate speech segments .The main advantage of using Entropy-Based Speech Background Contrast is for the endpoint detection where the energy based methods fails at time due to sub vocal or fricative sound .The new Entropy Based Algorithm shows better performance in noisy environment .Using this algorithm, much higher recognition rate can be achieved specially for small to medium vocabulary system. For large vocabulary continuous speech, it is useful in rejecting silence periods.

II. Proposed Methodology

The proposed method is based on detecting word boundary using pitch variation. This method is based on the behavior of pitch frequency across the sentences.

Step 1: Identify the sentence.
Step 2: Record the sentence from different informants.
Step 3: Create .Lab file.
Step 4: Calculate pitch frequency across the sentences using PRAAT software.
Step 5: Calculate Pitch contour using Autocorrelation method.
Step 6: Calculate Pitch contour using Cepstrum method.

III. Implementation

A. Calculate Pitch contour using PRAAT software

- Identify a sentence: "Ram bhal lora"
- Record the above sentence from three different persons.
- After recording, I get some waveform. From the waveform we can find pitch contour.
- Using pitch contour, we can detect word boundary.

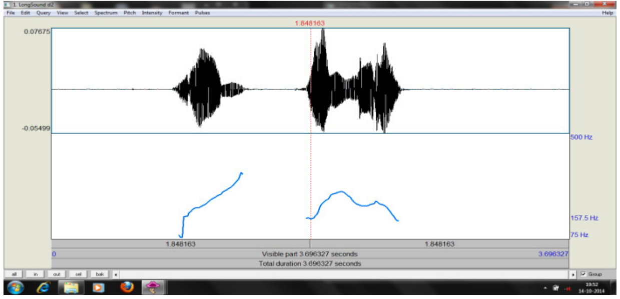
0.07675
-0.05499
1.848163
500 Hz
157.5 Hz
75 Hz
Visible part 3.696327 seconds
Total duration 3.696327 seconds

Fig 1 : Screenshot of waveform and pitch contour for speaker 1

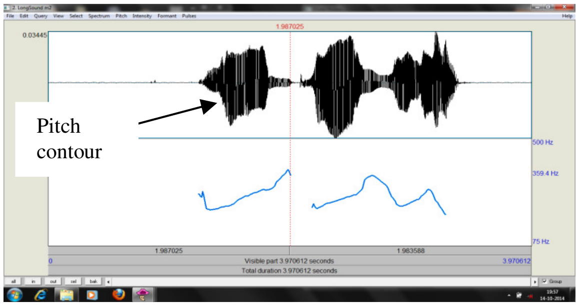
0.03445
1.987025
Pitch
contour
500 Hz
359.4 Hz
75 Hz
Visible part 3.970612 seconds
Total duration 3.970612 seconds

Fig 2 : Screenshot of waveform and pitch contour for speaker 2

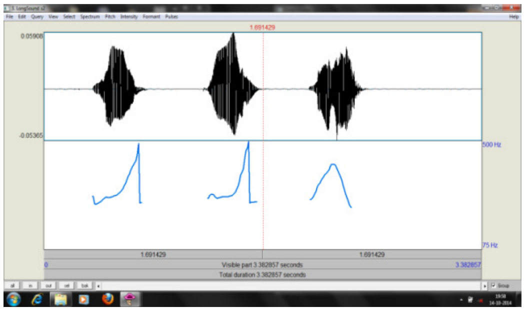
0.05908
-0.05365
1.691429
500 Hz
75 Hz
Visible part 3.382857 seconds
Total duration 3.382857 seconds

Fig 3 : Screenshot of waveform and pitch contour for speaker 3

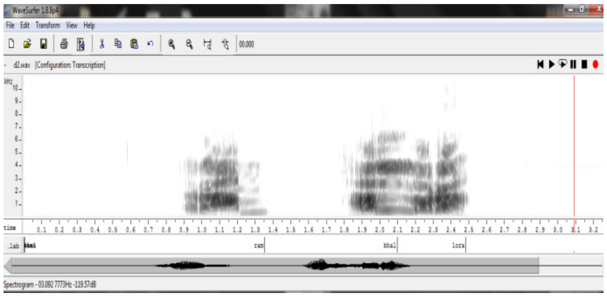
WaveSurfer
File Edit Transform View Help
d2.wav [Configuration: Transcription]
time
.lab
ram
bhal
lora

Fig 4 : Screenshot of .lab file in wavesurfer

B. Create .Lab file

Steps:

```
#
200     125     si
400     125     ram
600     125     bhal
800     125     lora
900     125     si
```

C. Calculate Pitch contour using Autocorrelation method

The objective of autocorrelation method is to calculate the pitch periods of a given speech signals by finding the time lag corresponds to the second largest peak from the central peak of autocorrelation sequence.

Using autocorrelation method, we can find pitch, as shown in fig 5, fig 6 and fig 7. Here we find (x,y) coordinate point, from that we can find whether there is a word or not. If y=0, there is no pitch as shown in fig 5, fig 6 and fig 7, then we can assume that there is no word. .It means there is silence. If y≠0, that means there is some word exist. Also as shown in fig 5, fig 6 and fig 7, y value is decreasing in some area .In that case, we also consider that there is a word boundary.

For Speaker 1

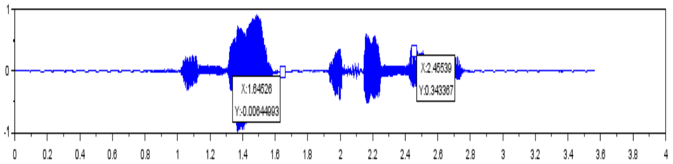

Fig 5 : Screenshot of waveform using Autocorrelation method for speaker1

For Speaker 2

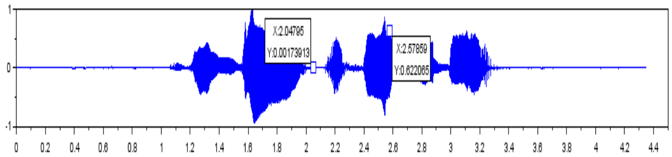

Fig 6 : Screenshot of waveform using Autocorrelation method for speaker2

For Speaker 3

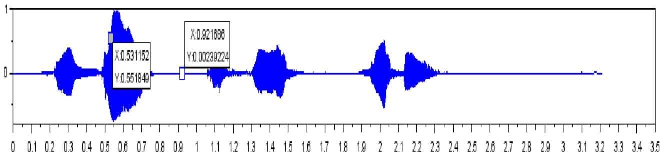

Fig 7: Screenshot of waveform using Autocorrelation method for speaker3

D. Calculate Pitch contour using Cepstrum method

- In first step, the original time signal is transformed using a Fast Fourier Transform (FFT).
- In second step, the resulting spectrum is converted to a logarithmic scale.
- In third step, the log scale spectrum is transformed using the same FFT to obtain the power cepstrum. The power cepstrum reverts to the time domain and exhibits peaks corresponding to the period of the frequency spacings common in the spectrum.

$c(n)=|F(\log|F(x(t)))|^2)|^2$ ....[1]

$c(n)= F^{-1}(\log|F(x(t))|^2)$...............[2]

For Speaker 1

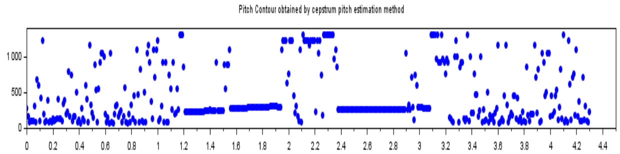

Fig 8 : Screenshot of pitch contour using cepstrum method for speaker 1

For Speaker 2

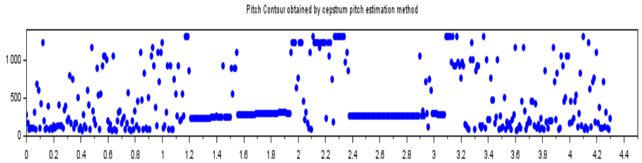

Fig 9 : Screenshot of pitch contour using cepstrum method for speaker 2

For Speaker 3

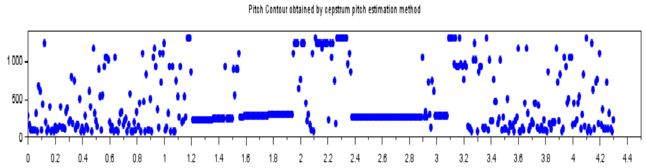

Fig 10 : Screenshot of pitch contour using cepstrum method for speaker 3

IV. EXPERIMENT RESULTS

In this paper we use four methods to detect word boundary in Assamese language. In first method we find pitch contour using Praat software. But after the experiment we found that this is not very efficient and the accuracy is less. Because, by observing the above figure 1, figure 2 and figure 3 we are not able to find accurate word boundary for large database.

In second method we create .lab file from our speech file. Using spectrogram and .lab file we find word boundary. As it is difficult to detect less concentrated areas in the spectrogram with naked eyes, there is a less scope of getting accurate result.

Another method we implemented is autocorrelation technique. Here we find (x,y) point. The main disadvantage of autocorrelation method is that there may be wrong estimation of pitch due to the excitation source of the vocal tract .Because in autocorrelation method we do not separate the vocal tract and excitation source related information in the speech signal.

By keeping in mind the above facts, we try to apply the cepstrum method. In cepstrum method we have done cepstral analysis. The ceptral analysis of speech provides better pitch estimation and in turn word boundary can be detected much accurately.

V. CONCLUSION AND FUTURE WORK

In future work, we will use Artificial Neural Network, to develop speaker independent word boundary detection system. Speaker dependent system is used by a single speaker, but a speaker independent system is used by any speaker. Also, the speech processing process is still facing a lot of problems .In general some of the most common ones are:

- *Speaker variation:* in this case exactly the same word is pronounced differently by different people because of age, sex, anatomic variations, speed of speech, emotional condition of the speaker and dialect variations.

- *Background noise:* a noise environment can add noise to the signal. Even the speaker himself can add noise by the way he speaks.

To overcome all the above problems, I will use artificial neural network.

REFERENCES

[1] Srichand "Word Boundary Detection in Indian Languages and Application to Keyword Spotting," Department of Computer Science and Engineering Indian Institute of Technology, Madras- 600 03G, India, August 1996.

[2] Ramana Rao G.V. and Srichand J. ,"Word Boundary Detection Using Pitch Variation," Department of Computer Science and Engineering, Indian Institute of Technology, May 1996.

[3] Anurag Jain, S.S. Agrawal, Nupur Prakash ,"Performance evaluation of word boundary detection for Hindi speech database," GGSIP University, Delhi India ,2CDAC Noida India.

[4] G.V.Ramana Rao "Detection of word boundaries in continuous hindi speech using pitch and duration,",Department of computer science and engineering,Indian Institute of Technology,Madras 600036,India.

[5] Colin Keng-Yan TAN, Kim-Teng LUA,"Learning of Word Boundaries In Continuous Speech Using Time Delay Neural Networks," Laboratory for Computational Linguistics, Department of Computer Science, School of Computing, National University of Singapore.

[6] Andreas Tsiartas, Prasanta Kumar Ghosh, Panayiotis Georgiou and Shrikanth Narayanan ,"Robust Word Boundary Detection in Spontaneous Speech using Acoustic and Lexical Cues,"Speech Analysis and Interpretation Laboratory,Department of Electrical Engineering, University of Southern California, Los Angeles,CA90089.

[7] Syed Abbas Ali,Najmi Ghani Haider and Mahmood Khan Pathan, "A LPC-PEV Based VAD for Word Boundary Detection,", Faculty of Computer &Information Systems Engineering, N.E.D University of Engg.&Tech.,Karachi.

[8] Jan Bartoˇsek and V´aclav Hanˇzl ,"Foot Detection in Czech Using Pitch Information and HMM,",Department of Circuit Theory, FEE CTU in Prague, Technick´a 2, 166 27 Praha 6 - Dejvice, Czech Republic.

[9] Manash Pratim Sarma and Kandarpa Kumar Sarma, "Speech Recognition of Assamese Numerals Using Combinations of LPC - Features and Heterogenous ANNs", Department of Electronics And Communication Technology,Gauhati University, Assam, India, Springer-Verlag Berlin Heidelberg 2010, pp. 8–12.

[10] Robert Togneri department of Electrical and Electronics Engineering , M.D. Alder , "Speech processing using artificial neural network", Department of mathematics Yianni Attikiouzel Department of electrical and electronics engineering, the university of western Australia.

[11] Archana Agarwal, Anurag Jain, Nupur Prakash and S.S.Agrawal," Word Boundary Detection in Continuous Speech based on Suprasegmental Features for Hindi Language,"University School of Information Technology, G.G.S Indraprastha University, Delhi, India, 2010 2nd International Conference on Signal Processing Systems (ICSPS).

[12] Francesco Beritelli, "Robust Word Boundary Detection Using Fuzzy Logic,"(Istituto di Informatica e Telecomunicazioni -University of Catania, V.le A. Doria 6, 95125 Catania, Italy).

[13] Khurram Waheed, Kim Weaver and Fathi M. Salam, "A Robust Algorithm for Detecting Speech Segments using an Entropic Contrast," Circuits, Systems and Artificial Neural Networks Laboratory ,Michigan State University, East Lansing, MI 48824-1226.

[14] Joe Tebelskis, "Speech Recognition using Neural Networks",School of Computer Science,Carnegie Mellon University,Pittsburgh, Pennsylvania 15213-3890,May 1995.