Word and syllable Boundary Determination in Continuous Bodo Speech

Uzzal Sharma
Department of Computer Application,
SoT, Assam Don Bosco University, Guwahati, Assam
uzzal.sharma@dbuniversity.ac.in

Abstract*: *The idea of word boundary is very fundamental and one of the important factors as far as the study of any human language is concerned. In actual practice it seems, describing boundaries of word are somewhat easier to define, although it is a subject with limited spread. In this paper, we try to study the spoken Bodo language, we considered four common criteria for detection of boundaries of syllable or word namely quick initial speech, sluggish terminating speech, reset in pitch and pauses) and examined their existence in a fragment taken from a recorded database that contains approximately 23 similar units. The recorded speech was parsed by two linguistic experts and the result so obtained were examined acoustically to establish which were present at each separation of words. Later on the obtained result is compared with the result obtained from SVM classifier. It has been noticed that both approaches gives similar and comparative result. Further, it is found that only a few of the determined separation of words conformed to all cues. It is found that the sluggish terminating speech was most prevailing, followed by reset in pitch, then pauses, and finally quick initial speech. An intense study, involving large set of units and more speakers is still under process.

***Keywords*: Intonation, SVM, Syllable, Pitch, Prosody, SSR, FSR, Pitch Reset, Pause;**

(Article history: Received: 16th March 2019 and accepted 25th May 2019)

I. INTRODUCTION

The Bodo (pronounced as bɔ́ɾɔ̀z) community is a major ethnic and linguistic communities of North-East India including Nepal and adjoining areas and one of the early settlers of Assam state of India. The word 'Bodo' has a duel role, it denotes the language and the community. From mythological concept, the Bodos are the ancestors of Lord Vishnu and mother Earth. They were known as 'Kiratas' during the Epic period[1].

The Bodo speaking community is at present well-spread across the north-east India including Assam, Meghalaya, Arunachal Pradesh, Nagaland, Mizoram, Tripura, Manipur, northern part of West Bengal, Bihar and adjoining areas of Bangladesh, Nepal and Bhutan. The Bodo speaking people are primarily found in the Brahmaputra Valley in Assam, and its few adjacent areas of Jalpaiguri, in West Bengal. The concentration of Bodo speaking people is relatively less in the northern part of the Brahmaputra Valley.

Recently, for the last two to three decades, the research in the area of speech associated with different language has got lot of importance because of the demand and improvement in the technology related to the man-machine interaction. The interest of the mankind has shifted for ensuring easy access to the technology. As a result, Automatic Speech Processing (ASR) research for different languages are going on in full swing. As one of the primary steps for ASR, the identification of words in an utterance as well as syllables present in a word is very important. The words are the building blocks of a sentence and each word comprises of one or more syllables. In research related to speech processing the determination of word and syllable boundary plays a major and important role. The word and syllable boundary is collectively termed as Intonation boundary[2]. The aim of this paper is to discuss about the identification of word and syllable boundary present in the utterance of Bodo speech.

II. HYPOTHESIS

The primary hypothesis formulated in the present study is that the fundamental building blocks of any spoken language are the words. The term intonation unit instead of word was first used by Chafe in the model he proposed regarding flow of information associated with speech, where word boundary are basically the functions of our brain limited to the processes that occurs during articulation within the brain of both the speaker and the listener [2]. The word, along with its allied concepts, does not match with the researchers of other theoretical areas and various descriptions have been given to explicate the concept [3], [4], [5], [6]. The intonation units within a sentence are considered as the primary prosodic or syntactical unit of a language in its spoken form. According to the definition given by Chafe, the word is a unit speech that has high involvement with a "coherent intonation contour", an involvement which could be changed depending on the situation, which is found in many descriptions of intonation units. The problem with this sort of explanation is that "a coherent intonation contour" is extremely complex to define, neither is it simple to define a word by any other inherent criteria [6], [7] nor separation of different words. Therefore, a frequently held course of action to parse an utterance into word is as per their boundaries [6], [8], [9].

Researchers across the globe has proposed many ways to distinguish words separately. Some of the most commonly used ways applied to identify the words are: (1) final lengthening of syllable or slow speaking rate at the ending of a word; (2) pause; (3) pitch reset ; and (4) fast speaking rate at the beginning of the next word. While languages vary

in its most important cue for division of word [10], the hierarchy given below is suggested for Bodo: (1) reset of pitch; (2) speech rate changed at cross-boundary; (3) pause [11], [12]. However, in the present research , this suggestion is based on recorded speech rather on naturally occurring speech which is spontaneous. A primary findings in recorded Bodo speech gives the intuition that rate of speech, principally final lengthening of syllable, may be pre-dominant in the hierarchy as compared to the other cues [13]. It is found that though, the different linguists sometime may have difference in their opinion of word boundaries, and it also many a times occur that the same transcribers change their way of thinking and gives difference of opinion about word boundary. Therefore, we put effort in investigating the correctness of intonation unit processed with quantitative acoustic tools. In this paper we are trying to present our initial findings and try to establish a relation between acoustic analyses of word boundaries set against human opinion and the areas of Artificial Intelligence.

III. TEXUAL DATA AND METHODOLOGY

The speech corpus under study is narrated by a group of speakers whose mother tongue is bodo and is qualified enough (either graduate or post-graduate in bodo). We decided to start with a recorded instead of a common chat, because it is characterized by the presence of more sequences of prominent word boundary spoken by the same speaker. The selection of a recorded speech further facilitates the establishment of relationship between theoretical boundaries and their acoustic counterpart, because normal conversation habitually contains abundant occurrences of speech which are broken, units which are truncated and overlaps and is often performed in an environment which is noisy, which demands a robust system. The recorded samples were given to two phonetic experts. As the impression about word is pretty hard for explanation to a layperson, these persons were people having sufficient expertise with the task. The informants were advised to make their parsing by ignoring the syntactic cues. The PRAAT is used to analyzed acoustically the perceptual segmentation.

IV. FINDINGS AND ANALYSIS

In the current study we analyzed 23 units altogether. The average duration of utterance of a word is 0.89 second, ranging from 0.31 second to 1.47 second.

A. Acoustic Findings

The four conditions, or the cues which are mentioned earlier for detection of boundaries of word were examined quantitatively: (1) fast speaking rate at the starting of a word; (2) final lengthening of syllable or slow rate of speaking at the end of a word; (3) reset in pitch; (4) pause. The results so obtained were verified by two researchers independent of each other. Among the 23 word boundary with theoretical consent among the speakers, 5 word boundaries (24%) were aligned to all the four cues. At the same time, two word boundaries did not even conform to any acoustic condition. This is because these two word boundary have continuing tone and thus found difficult to be determined [3], [4], [14]. Therefore, internal cues are also need to be considered seriously and should be given due importance. Among the remaining words, 4 had three boundary cues each, among which Slow Rate of Speaking (SRS) was found in all 4, and reset of pitch was found in 4 out of the 4 words. 9 words showed two of the pre-assumed cues, among which Slow Rate of Speaking (SRS) was found in 5 words and the same number also present in reset of pitch, even if it is not necessarily the same units. A single cue was associated in 4 words, where SRS was attested in 2 of them, reset of pitch in 1, and pause in 1, at the same time FRS was not found in any word. Table 1 displays the type as well as number of cues in the sample.

(a) Slow Rate of Speaking (SRS)

The condition of final lengthening of syllable or slow rate of speaking at the closure of a word was seen in 17 words (85%) out of the entire sample). The computation of SRS is the ratio of the average duration of the concluding syllable of word and the average syllable duration in that word. We termed it as lengthening whenever the ratio was >1.2 (i.e., final syllable duration is more than 12% of the average of duration of syllable in a word). The average duration of final syllable in the entire sample is 0.24 second.

(b) Reset inPitch

The reset of pitch was found in 16 words (69%). Intonation is characterized by the rises and falls of pitch and at the same time they are also the basic cues for stress in word [6], [10]. Bodo is a tone language having both unstressed and stressed syllables. A word of Bodo language has only one primary stress and it may or may not have any secondary stress. The position where a stress occurs has major implications on the syllable structure of a prosodic or lexical word [15] in Bodo language. According to some groundwork research conducted for Bodo language, all proposed factors were predominant in producing the word stress in Bodo [16], in the following sequence: (1) pitch, (2) intensity, (3) duration. Because of this, to identify whether there is a reset of pitch or not, a random check on the frequency difference is done to differentiate between unstressed and stressed syllables in the recorded sample. When the difference is greater than 15 Hz, it was considered that there is a reset in pitch.

The average reset in pitch is found to be 45 Hz. A downward in Reset was found in 13 words. Upwards in Reset was found in only 8 words (Table 1). In the present study any significant difference in property between terminal tones fall and terminal tones raise regarding reset in pitch could not be detected.

TABLE I. NUMBER AND TYPE OF CUES WITHIN AN INTONATION UNIT.

No. of cues	SRR	Reset of pitch	pause	FSR	Total
1	2	1	1	0	4
2	6	6	3	1	8
3	4	4	3	1	4
4	5	5	5	5	5
Total	17	16	11	7	21

(c) Pauses

Pauses are the characteristics which are found at the closure of a word. A pause was established as important and significant if it is at least last for 0.02 seconds or more. This

condition was found in 46% of the sample's word i.e., 11 words.

(d) Fast Rate of Speaking (FRS)

The condition of fast rate of speaking at the starting of a word was found in 7 words (30%; In the present study, we did not consider words, which has less than four syllables). The FRS is calculated as the ratio between the average of duration of syllable before the major syllable along with the average duration of syllable in that word under study. We considered a word with FRS whenever the ratio is greater than 0.9 (i.e., the average duration of syllable before the major syllable is less than 90% of the average duration of syllable in that word). The average duration of syllables in the whole sample is 0.16 second, while the average duration of a FRS syllable is 0.12 seconds. It must be noted that the number of unstressed syllables at the starting of an utterance [6] cannot be regarded at this phase of research as a important aspect in checking perceptual acoustic relation in words parsing in Bodo Language.

In any occurrences, FRS was not found to be the primary cue. Further, FRS was found in one of 3-4 cues, except for one word where it appears with another cue, i.e. reset of pitch. 95% of FRS along with SRS, and as it was mentioned earlier, along with at least another cue. While looking at its occurrences of SRS at the end of a word and FRS at the starting of the following word, i.e., looking at speech rate as a cross-boundary cue, we found only nine such mutual occurrences (56 % of FRS attestations).

(e) Pitch, Duration and Intensity

At the end let us throw some light into three important parameters, i.e. Pitch, Duration and Intensity. Although, intensity is one of the important factors in determining the word boundary, we are not considering it at this stage of study. But it plays a very important role in finding out word boundary more accurately. Intensity is associated to both pitch and duration, where pitch gives prominence and duration gives stress. Figure.1 illustrates another instance of the correlation between the three prosodic units in the internal structure of a word.

The important syllable in the word considered as example has been uttered by the two informants to be the last but one syllable, i.e., kh. Obviously, the contour formed by pitch does not qualify to this perceptual decision. Moreover, the curve formed by pitch as shown in Figure 1, may suggest three words, as opposed to the perceptual judgments. Since the final word in this case has a stress in penultimate position, duration of the major syllable is less important in this analysis since its measured length was close to the average duration of the syllable and therefore seems invalid cue neither for a word boundary nor for the accented syllable. Still, intense further research is required in order to portray any conclusions concerning the relationship between intensity, duration and pitch and their prosodic influence in Bodo language.

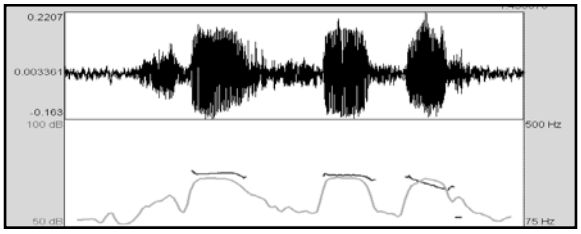

Figure 1: Pitch and Intensity correlates in the intonation unit "su su khAr".

V. Comparison with SVM Based Approach

The same speech corpora is also subjected to SVM classifier with polynomial kernel function. The order of the kernel function is considered as 3. Out of the 5 cues, 3 cues are considered. They are SSR, Pitch reset and pause. The reason for considering the above mentioned three cues is that they have given better result in the proposed approach. The result obtained from proposed approach is verified with the result obtained from the SVM approach, which is depicted in the table II.

TABLE II. A COMPARATIVE STUDY OF PROPOSED APPROACH AND SVM

No. of cues	SSR		pitch reset		pause		Total	
	Prop Approach	SVM	Prop Approach	SVM	Prop Approach	SVM	Prop Approach	SVM
1	2	2	1	0	1	1	4	3
2	6	5	6	5	3	3	8	13
3	4	4	4	4	3	3	4	11
4	5	5	5	4	5	4	5	13
Total	17	16	16	13	12	11	21	

VI. Conclusion

Several remarks can be made in conclusion. On one hand, in some ambiguous words, not more than two of the acoustic cues considered here were present. The three acoustic cues: final lengthening or SRS, reset in pitch and pause were found in more than 52% of the words. SRS and reset in pitch both were found in more than 71%, leading us to the conclusion that the theoretical word segmentation is indeed influenced mostly by its boundaries. Still, there is scope of much work to compare duration of syllables which are stressed and accented (prominent) syllables and assessing their relationship with SRS [17]. Nevertheless, we cannot evade the conclusion that research on inherent acoustic cues needs more attention. FRS – confirms its lowest hierarchical status – may suggest its consideration together with SRS as an internal condition and not as an external one. The frequency at which the cues occurs compels us to reconsider the hierarchy of cues given in earlier research on properties of word in Bodo language [11], [12].

At the same time, when the same dataset is subjected to SVM, a near similar result is obtained with a very minimal variations.

So, finally, it could be concluded that the proposed approach is capable of correctly detecting the word and the syllable boundary.

REFERENCES

[1] Bhattacharya, P.C., (1997) *"A descriptive analysis of the boro language"*. Department of Publication, G.U.: The registrar, G.U.

[2] Chafe, W., 1994. *Discourse, Consciousness, and Time:The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.

[3] Halliday, M.A.K., 1989. poken and Wrtitten Language. Second edition. Oxford: Oxford University Press.

[4] Halliday M.A.K., 1994. *An Introduction to Functional Grammar*. Second edition. London: Arnold.

[5] Brazil D., 1997. *The Communicative Value of Intonation in English*. Cambridge: Cambridge University Press.

[6] Cruttenden, A., 1997. *Intonation*. 2nd edition. (Cambridge Textbook in Linguistics.) Cambridge: Cambridge University Press.

[7] Ladd, D. R., 1986. Intonational Phrasing: The Case for Recursive Prosodic Structure. *Phonology Yearbook* 3,pp. 311-340.

[8] Du Bois, J. W.; Cumming S.; Schuetze-Coburn S.; Paolino D.,1992. *Discourse Transcription*. (Santa Barbara Papers in Linguistics, 4.) Santa Barbara, CA: Department of Linguistics, University of California, Santa Barbara.

[9] Du Bois, J. W.; Cumming S.; Schuetze-Coburn S.; Paolino D., 1993. Outline of Discourse Transcription. In *Talking Data: Transcription and Coding in Discourse Research*, J. A. Edwards and M. D. Lampert (eds.). Hillsdale, New Jersey: Lawrence Erlbaum Associates. pp. 45- 89.

[10] Hirst, D.J.; Di Cristo, A., 1998. *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press.

[11] Laufer, A., 1987. *Hangana ('Intonation')*. Jerusalem: Institute for Judaic Studies, Hebrew University of Jerusalem. (Hebrew).

[12] Laufer, A., 1996. Pauses in Fluent Speech and Punctuation. In Studies in *Hebrew and Jewish Languages Presented to Shelomo Morag*, M. Bar-Asher (ed.). Jerusalem: The Center for Jewish anguages and Literatures, The Hebrew University of Jerusalem and The Bialik Institute, pp. 277-294. (Hebrew).

[13] Prosody and Intonation 9.59 / 24.905 April 14, 2005Ted.

[14] Miller, J.; Weinert, R., 1998. *Spontaneous Spoken Language: Syntax and Discourse*. Oxford: Oxford University Press.

[15] Bolozky, Sh., 2003. Phonological and Morphological Variation in Spoken Hebrew. In *Corpus Linguistics and Modern Hebrew: Towards the Compilation of The Corpus of Spoken Israeli Hebrew (CoSIH)*, B. Hary (ed.). Tel Aviv: Tel Aviv University, The Chaim Rosenberg School of Jewish Studies, pp. 119-156.

[16] Brahma, Aleendra, (2007*),"The verb phrase in bodo"*,An M.A. dissertation submitted in the Deptt. of Linguistics, Gauhati University.

[17] Campbell, N., 1993. Automatic detection of prosodic boundaries in speech. *Speech Communication* 13, pp. 343- 354.

AUTHOR PROFILE

Uzzal Sharma

The author has obtained his MCA from IGNOU and completed PhD from Gauhati University. He has over 15 years of experience in the field of academics and in industry. His research area includes Speech Signal Processing and Software Engineering and Data Engineering. He has produced one PhD scholar and currently guiding 4 research scholars for their PhD degree. He has also guided many MTech, BTech and MCA students for their projects in different areas. He has published more than 25 research papers in journals (International and National) and conference proceedings (International and National). He also has 11 book chapters to his credit in edited book. He has also published five books as a sole author. Currently he is an Assistant Professor - Stage 2 at Assam Don Bosco University, Guwahati, India.