WORD BOUNDARY DETECTION USING PITCH VARIATIONS

Ramana Rao G.V. and Srichand J.
Department of Computer Science and Engineering
Indian Institute of Technology
Madras 600 036, INDIA

May 2, 1996

ABSTRACT

This paper proposes a method for detecting word boundaries. This method is based on the behaviour of the pitch frequency across the sentences. The pitch frequency (F_0) is found to rise in a word and fall to the next word. The presence of this fall is proposed as a means of detecting word boundaries. Four major Indian languages are used and the results show that nearly 85% of the word boundaries were correctly detected. The same method used for German language shows that nearly 65% of the word boundaries were correctly detected. The implications of these result in the development of a continuous speech recognition system are discussed.

1. INTRODUCTION

The problem of word boundary detection can be stated as follows : Given a string of symbols representing a sentence, word boundaries are to be placed in the symbol string to convert it into a string of words. This problem is relevant in developing a continuous speech recognition system, as the speaker does not indicate the word boundaries consciously, while speaking. Reliable detection of word boundaries in continuous speech is an important problem in speech recognition. Continuous speech does not offer many clues for detecting word boundaries. One could identify and use clues based on speech and language characteristics. Prosodic features such as pitch, duration and pause can be used as speech clues to detect word boundaries, while language features such as phoneme sequence constraints and syntactic markers can also be exploited for placing word boundaries. this paper focuses on the use of speech clues for word boundary detection.

In most of the speech recognition systems, the word boundaries are detected as a part of the word hypothesization process in the lexical analyzer. But this is a tedious and complex procedure. However, *apriori* detection of the word boundaries offers several advantages. If word boundaries can be correctly detected, many of the techniques developed for Isolated Word Recognition systems can be adapted for Continuous Speech Recognition.

2. PITCH PATTERN CLUES

Phonologically pitch frequency can be explained as the rate of vibration of the vocal chords during speech production. Recent studies for Hindi, a widely spoken Indian language, show that pitch frequency variations can be used to hypothesize word boundaries in continuous speech[2][3]. These studies were conducted on simple declarative sentences. It was observed that the pitch frequency across a Hindi sentence declined from the beginning to the end of the sentence, as in the case of other languages. However, within this overall declination, there were local F_0 rise-fall patterns which are found to be related to the word boundaries in the sentence. A word boundary hypothesization algorithm based on these pitch patterns was developed for Hindi language(an Indian language), and it was found that the algorithm detected more than 70% of the word boundaries correctly [4].

3. WORD BOUNDARY DETECTION

3.1 Indian Languages

In the present study, the relation between pitch variations and word boundaries is examined for several Indian languages and the results are reported.

Speech data in four Indian languages, Hindi, Bengali, Marathi and Telugu, respectively spoken in the northern, eastern, western and southern regions of India, were used. The results of these studies have shown that all these languages, have similar F_0 variations and these variations can be used to identify word boundaries.

The F_0 patterns have global and local properties. The global characteristics of F_0 is observed over a unit of speech that includes a phrase or sentence. Here, the F_0 tends to decline across the sentence, in the case of declarative sentences. Also, we observe that there are local rises and falls of pitch across each word, contributing to local characteristics. It is observed that F_0 rises from the beginning of the word to the end, that is, it rises from the first vowel of each word to the final vowel. Thus, F_0 patterns over sentences tend to decline, with local falls and rises in the words.

The Figure 1. shows the pitch patterns for one utterance each in the Indian languages. In the Figure 1.1, which denotes a Bengali utterance, the points marked a, b, c, d, e, f when connected, represent the global characteristics. It can be observed that the pitch falls steadily from the beginning to the end. In the same figure, the points g, a and h, b when connected, represent the local characteristics. It can be seen that the pitch increases from the first vowel to the last vowel in the sentence. From Figure 1., it is observed that all the utterances, spoken in different languages have similar intonation, and display the same properties.

The F_0 across each word increases from left to right. As the overall F_0 falls from left to right in a simple sentence, the fall will occur at the boundary between two words. The word boundary hypothesization algorithm is as follows :

1. Detect the peaks in the pitch contour of the utterance and hypothesize these vowels as the word final vowels.

2. Hypothesize a vowel as a word final vowel if its F_0 is larger than that of the next vowel.

The studies were conducted on utterances by ten speakers per language, each speaker uttering 50 sentences. The sentences were declarative, and of varying length, from 5 words to 15 words.
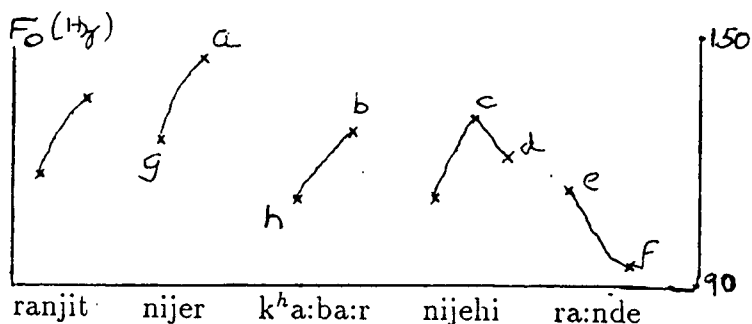

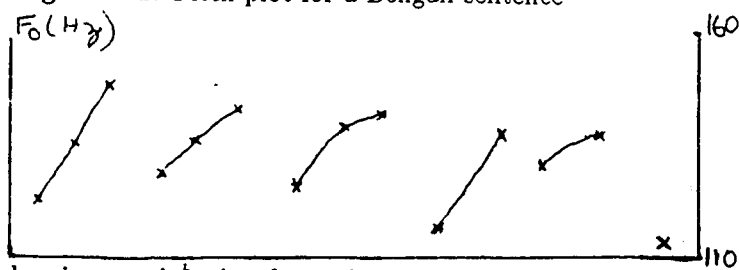Figure 1.1: Pitch plot for a Bengali sentence

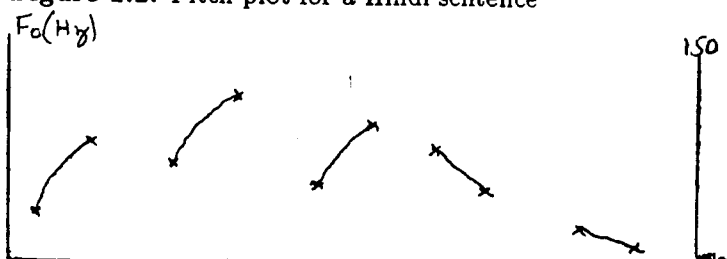Figure 1.2: Pitch plot for a Hindi sentence
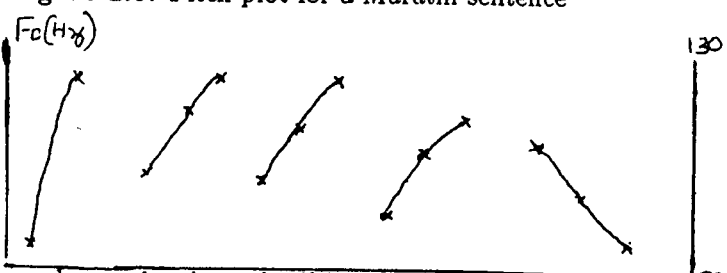
Figure 1.3: Pitch plot for a Marathi sentence

Figure 1.4: Pitch plot for a Telugu sentence

Figure 1: The pitch pattern variations of one sentence each of the Indian languages.

We define two measures, *hit rate* and *false alarm rate*. The *hit rate* gives the percentage of the word boundaries correctly detected and the *false alarm rate* specifies the percentage error in the word boundary hypotheses generated by the above algorithm. They are defined as follows.

WB - Word Boundaries

$$hit\ rate = \frac{No.\ of\ WB\ correctly\ hypothesised}{No.\ of\ WB\ present\ in\ the\ utterances}$$

$$false\ alarm\ rate = \frac{No.\ of\ erroneous\ hypotheses}{No.\ of\ WB\ hypotheses}$$

The following tables show the results for ten speakers, in each language.

Speaker	Hit rate	False Alarm rate
1	87%	27%
2	90%	26%
3	89%	28%
4	87%	29%
5	88%	27%
6	86%	28%
7	89%	25%
8	90%	29%
9	88%	28%
10	88%	27%

Table 1: The results of the word boundary detection studies in Bengali.

Speaker	Hit rate	False Alarm rate
1	89%	19%
2	89%	19%
3	86%	19%
4	84%	18%
5	87%	21%
6	86%	20%
7	83%	18%
8	89%	21%
9	88%	20%
10	84%	23%

Table 2: The results of the word boundary detection studies in Hindi.

Speaker	Hit rate	False Alarm rate
1	89%	27%
2	92%	25%
3	87%	26%
4	85%	26%
5	91%	28%
6	87%	28%
7	89%	27%
8	88%	28%
9	86%	25%
10	87%	27%

Table 3: The results of the word boundary detection studies in Marathi.

Speaker	Hit rate	False Alarm rate
1	85%	31%
2	84%	26%
3	89%	25%
4	85%	28%
5	98%	29%
6	83%	26%
7	87%	27%
8	85%	26%
9	88%	27%
10	84%	25%

Table 4: The results of the word boundary detection studies in Telugu.

3.2 German Language

Based on literature [1] which places German and North-Indian languages in the same family, the study was extended to German. Data was collected from native speakers and the pitch contour of the utterances were observed for similar behaviour as the Indian languages. On, observation, the intonation of the German sentences was found to be close to the intonation of the Indian languages. There is a gradual fall of F_0 from the beginning of the sentence to the end. And across the words, there is a rise of F_0 from the first vowel to the last vowel. Thus, the pitch pattern in a series of rises and falls , with an overall fall across the utterance. So, the same algorithm can be applied to place word boundaries in German sentences also.

For German language, a total of five speakers were used, each speaking ten sentences. The sentences were declarative sentences, and had an average of 10 words per sentence. The dearth of native speakers has restricted our studies to a great extent. The results of the studies for the speakers are listed as follows.

Speaker	Hit rate	False Alarm rate
1	63%	24%
2	67%	29%
3	65%	25%
4	70%	26%
5	66%	24%

Table 5: The results of the word boundary detection studies in German.

4. ANALYSIS AND CONCLUSIONS

From the results it can be seen that more than 85% of the word boundaries are correctly detected in the Indian languages, and more than 65% of the word boundaries are detected in German using the specified method. In the Indian languages, the misses in the word boundaries were caused by the presence of compound sentences. And the false alarms were caused by words which were foreign to the user's vocabulary, and at the end of the utterance, where the fall in F_0 is rapid. In German, it was observed that the misses occurred less frequently after words having two or more vowels in them, and occurred more frequently between words with with only one vowel in it and the next word.

The results of our study clearly indicate that Indian languages have similar prosody, unlike European languages. Linguistic studies [1], place Bengali, German and Hindi in the same family of languages, that is, Indo-European, and Marathi and Telugu are placed in the Dravidian family of languages. It is interesting to note that though the languages belong to different families, they have similar F_0 variations.

As shown in our studies, the intonation patterns can be used to perform word boundary hypothesis. The detection of word boundaries, is very useful in developing a robust and efficient speech recognition system, as shown in earlier studies [4].

Since Indian languages have a similar F_0 pattern, synthesis of speech can also be done using a common model of prosody. Our results provide a starting point for developing such a common prosodic model for all Indian languages. Further studies can be conducted using a greater database of German sentences to establish the extent of similarity between German and the Indian languages.

5. REFERENCES

1. Comrie, B., The Worlds Major Languages, Croom Helm, (1987).

2. Madhukumar A.S, Intonation knowledge for speech systems for an Indian language, PhD thesis, IIT Madras, India (1992).

3. Ramana Rao G.V, Detection of word boundaries in continuous speech using pitch and duration, Proceedings of the Fourth Australian international conference on Speech Science and Technology (SST - 92), Brisbane, Australia (1992).

4. Ramana Rao G.V, Word boundary hypothesization in Hindi speech, PhD thesis, IIT Madras, India (1994).