# Abstract

Speech is an efficient medium of communication in human-human, human-machine and machine-mediated human-human interactions. The development of accurate, robust and versatile speech recognition models is an integral part of automating machine operations in human-machine interactions. Although state-of-the-art ASR models perform well in case of isolated word recognition, their performance level is not satisfactory in case of continuous speech. The conversion between spoken and written language associated with ASR systems using the written language lexicon involve lot of ambiguity at the lexical and syntactic levels. It is, therefore, desirable to develop a spoken language lexicon based on higher level linguistic information that can be satisfactorily used by ASR systems handling continuous speech. However, the accurate detection of spoken word boundary in continuous speech poses serious challenge. The suprasegmental parameters or prosodic features of speech like fundamental frequency, stress, duration, etc. indicate the prosodic word boundaries which exactly match with spoken word boundaries but may not coincide with the written word boundaries. In this thesis, we propose a new method for the detection of prosodic word boundaries in Bengali continuous speech using the fundamental frequency contour.

Since Bengali is a bound-stressed language where the first syllable is always stressed, the fundamental frequency contour in continuous speech is characterized by a rising pattern at the beginning of every prosodic word. This feature is utilized to detect the prosodic word boundaries by applying the empirical mode decomposition technique on the logarithmic fundamental frequency pattern. In the present work, we investigate the improvements that can be achieved in phoneme recognition and classification for continuous speech by incorporating the phonological features such as place and manner of articulation into the recognition model. It is found that the classification of phonemes based on the place of articulation is difficult due to the relatively small duration of the transitory segment in the phonemes. Such difficulties in phoneme classification are reduced by using the manner of articulation as a basis. The phonemes are classified into 15 separate groups comprising one or more members based on distinct sets of robustly identified manners. These groups are used in labeling the prosodic words to generate pseudo words that redefine the basic units of the prosodic word dictionary. These newly defined units, which may contain one or more prosodic words, are known as cohorts. While an unique cohort (consisting of a single prosodic word) directly identifies the constitutive prosodic word, we have developed a lexical expert system based on the classification of vowels that can identify the prosodic words in all other cohorts (consisting of multiple prosodic words).

Our speech recognition model for Bengali continuous speech gives an overall accuracy of 92.07% in prosodic word boundary detection, and, 87.8% and 82.5% respectively in the training and testing phases of phoneme classification. We have achieved an accuracy of 98.9% while classifying phonological features based on the manner of articulation as compared to only 50.2% while classifying based on the place of articulation. We find that 4633 cohorts represent the 5031 prosodic words of the developed spoken language lexicon. Our method of labeling the prosodic words based on the manner of articulation and use of the lexical expert system based on vowel classification is found to give an overall recognition accuracy of 93.9%.



**Keywords:** Prosodic features; fundamental frequency; prosodic word boundary; spoken word lexicon; phoneme confusion matrix; manner-based labeling.