UNIVERSITI TEKNOLOGI MARA

THE DOCTORAL RESEARCH ABSTRACTS

Volume: 10, Issue 10 October 2016

TENTH ISSUE

INSTITUTE of GRADUATE STUDIES

IGS Biannual Publication

16

Name : RASEEDA HAMZAH

Title : DISCRIMINATIVE CLASSIFICATION MODEL OF FILLED PAUSE AND ELONGATION FOR MALAY LANGUAGE SPONTANEOUS SPEECH

Supervisor : ASSOC. PROF. DR. NURSURIATI JAMIL (MS)
DR. NORAINI SEMAN (CS)

Automated speech recognition (ASR) for spontaneous speech poses extra challenge compared to read speech as it contains varied speaking rates, poor phonation and disfluencies. Studies have shown that filled pause is one of the most common disfluencies of spontaneous speech characteristic where it presents considerable problems for ASR performance. In many filled pause studies, the hindering factor is that filled pause being often recognized as short words which particularly has semantic meaning, such as „um" can be recognized as „thumb" or „arm". This problem becomes especially pertinent where a vowel sound of normal word being relatively long at any position in an utterance, both within a word as well as between words which formerly known as elongation. The existence of elongation causes normal word falsely detected as filled pause due to their similar acoustical feature patterns. Classifying elongation as filled pause affects ASR"s performance as eliminating normal words from recognition may modify the intended context of a speech. Therefore, the main aim of this research is to classify filled pause and elongation into its own classes by constructing a discriminative classification model from the extracted acoustical features. A large number of signal features have been employed for the problem of discriminating filled pause and elongation. Several well-established features such as Formant Frequency (FF), Fundamental Frequency (F0), Mel Frequency Cepstral Coefficients (MFCC), Zero Crossing Rates (ZCR) and Short Time Energy (STE) were used in this research. These features are carefully chosen to emphasize signal characteristics that differ between filled pause and elongation. In most speech research, extracting speech energy feature is still remains as challenging task due to it typically has a great deal of variance which include loudness as well as the variance in the signal energy between different phoneme which contains vowel or/and consonant sounds. One of the ways of detecting vowel and consonant is through its energy level. Beside the common way of quantifying the speech energy by calculating the sum of energy of the short interval centered on each interval, we proposed new technique namely, Local Maxima Energy (LM-E) to exploit the speech energy feature of filled pause and elongation. Experimentally, this can be done by measuring its amplitude transition from one frame to another by setting a threshold as height difference between peaks of the speech signal. Unlike other acoustical features, LM-E has shown its performance to classify elongation better by detecting the expressive contour of the elongation that is caused by the transition from consonant to vowel of the elongation. A rigorous feature performance evaluation shows that LM-E significantly increased the classification performance when fused with ZCR. Therefore, these two features are incorporated into discriminative Naïve-Bayes model for filled pause and elongation classification. The discriminative model of LM-E and ZCR improved the classification performance by 7% error rate reduction, and average of 7% accuracy increments compared to single feature classification performance. This model can further be used to improve disfluencies detection for a better ASR performance.