The combination of spectral entropy, zero crossing rate, short time energy and linear prediction error for voice activity detection

Thein Htay Zaw
theinhtayzaw.cu@gmail.com
University of Computer Studies, Mandalay, UCSM

Nu War
nuwar81@gmail.com
University of Computer Studies, Mandalay, UCSM

In this paper, an efficient classification of voice segment from the silence segment, unvoiced segment algorithm, which is both more accurate and laid-back to implement is proposed by comparing to some previous algorithms. The proposed algorithm uses spectral entropy and short time features such as zero crossing rate, short time energy, linear prediction error are used for voice activity detection (VAD). A compound parameter, D, is calculated by using all these four parameters. Dmax is calculated from all the frames of the signal. Then the value of D/Dmax is used to determine whether the frames are classified as speech and non-speech and silence frames. The threshold values have to be obtained empirically. Experimental results show that the method of this paper can detect end-points of voice signal more accurately and outperforms the conventional VAD algorithms. The method we used in this work was evaluated on TIMIT Acoustic-Phonetic Continuous Speech Corpus. This corpus is mostly used for speech recognition application and contains clean speech data and is compared with some of the most recent proposed algorithms.