ABSTRACT

Speech is an easy way for detecting the state of the speaker, and speech emotion recognition has been investigated widely in the field of human-machine interaction. Currently, most speech emotion recognition methods use existing parameters in OpenSmile. But, these methods cannot catch the dynamic states of the speaker because the parameter was taken from regression data. Our method tries to catch the dynamic state of emotions from the voice signal and extracts some parameters from it. This study proposes a Discrete Wavelet Transform (DWT) to decompose voice signals into several levels so that the level with minimum noise interference can be selected. The decomposition of DWT result shows that 4th until 6th levels provide a signal with the least noise level. It is assumed that emotion information is in the voice segment so that this study proposes to conduct voice segmentation to separate voiced-segments from unvoiced-ones for the parameter extraction process. The proposed method adopted the existing emotion parameters such as Zero-Crossing Rate, Energy, Peak, Fourier Coefficients, and Cepstrum. This study uses Neural Networks for classification. The experiment result shows that level 6 of DWT achieved eight emotions classification with 98% accuracy.

Keywords: Speech; Emotion Recognition; Discrete Wavelet Transform; Voice Segmentation; Neural Network; Parameter Extraction.