

Abstract

Speech Emotion Recognition (SER) is a technology developed on several computer to realize a Human-Computer Interaction (HCI). It is a challenging task because the lack of data. Some data augmentation methods have been created to increase the data variations, but they do not significantly improve the SER accuracy. Therefore, a new additional data augmentation method called Segment Repetition based on High Amplitude (SRHA) is proposed to solve this problem. This method makes some repetitions on the segments those have the highest amplitude. An experiment of 10 times data augmentation, using five standard augmentations and the additional SRHA with a Long Short-Term Memory (LSTM) as the classifier, shows that the proposed SRHA significantly increases the SER accuracy from 95.88% to 98.16%. Other experiments for 20 and 40 times data augmentations also show that the SRHA outperforms the five standard augmentations. These indicate that the SRHA is a powerful data augmentation method for SER.

Keywords: Data augmentation; High amplitude; Long-short term memory; Segment repetition; Speech emotion recognition.