

Speech Detection Model Based on Deep Neural Networks Classifier for Speech Emotions Recognition

Authors : Aisultan Shoiynbek, Darkhan Kuanyshbay, Paulo Menezes, Akbayan Bekarystankyzy, Assylbek Mukhametzhanov, Temirlan Shoiynbek

Abstract : Speech emotion recognition (SER) has received increasing research interest in recent years. It is a common practice to utilize emotional speech collected under controlled conditions recorded by actors imitating and artificially producing emotions in front of a microphone. There are four issues related to that approach: emotions are not natural, meaning that machines are learning to recognize fake emotions; emotions are very limited in quantity and poor in variety of speaking; there is some language dependency in SER; consequently, each time researchers want to start work with SER, they need to find a good emotional database in their language. This paper proposes an approach to create an automatic tool for speech emotion extraction based on facial emotion recognition and describes the sequence of actions involved in the proposed approach. One of the first objectives in the sequence of actions is the speech detection issue. The paper provides a detailed description of the speech detection model based on a fully connected deep neural network for Kazakh and Russian. Despite the high results in speech detection for Kazakh and Russian, the described process is suitable for any language. To investigate the working capacity of the developed model, an analysis of speech detection and extraction from real tasks has been performed.

Keywords : deep neural networks, speech detection, speech emotion recognition, Mel-frequency cepstrum coefficients, collecting speech emotion corpus, collecting speech emotion dataset, Kazakh speech dataset

Conference Title : ICEER 2025 : International Conference on Emotions and Emotion Recognition

Conference Location : Oslo, Norway

Conference Dates : June 24-25, 2025