

Improved Speech Activity Detection Model Using Convolutional Neural Networks

Friday 22 September 2023 11:20 (20 minutes)

In this study, the performance of four speech/voice activity detection (VAD) models, namely SpeechBrain, Picovoice, InaSpeechSegmenter and WebRTC, is compared using data collected from KAIST (Korea Advanced Institute of Science and Technology). The goal is to develop an improved VAD model based on the analysis of SpeechBrain's performance and provide a detailed perspective on the differences and similarities between these models. The study focuses on the technologies and algorithms used by each model and utilizes experimental methods to evaluate their performance. The experimental results show that SpeechBrain performs the best, with an average recall value of 0.97 and a precision value of 0.96. Our research endeavors have led to the refinement of the existing VAD model, resulting in even more compelling performance metrics. The improved model achieves a recall value of 0.98 and precision value of 0.97, signifying its enhanced capability to accurately detect speech activity. These findings hold promise for the future advancement of VAD models and their application in various speech-processing domains. This research can be used to enhance future VAD models and develop more advanced speech processing applications.

Index Terms—Improved Speech Activity Detection Model, Convolutional Neural Networks (CNN), VAD (Voice Activity Detection).

Authors: DEONISIE, Costin-Alexandru; POP, Florin (University Politehnica of Bucharest, Romania / National Institute for Research & Development in Informatics –ICI Bucharest, Romania); COCONU, Taisia-Maria; REBEDEA, Traian (University Politehnica of Bucharest)

Presenter: POP, Florin (University Politehnica of Bucharest, Romania / National Institute for Research & Development in Informatics –ICI Bucharest, Romania)

Session Classification: Session C

Track Classification: Pervasive Systems and Computing