Journal of Acoustical Engineering Society of Iran, Vol. 10, No. 2, 2023

(Research Article)
Telephone robustness speaker verification using time delay neural network

M. Asgari *, N. Akbari, M. Aghagolzade, M. S. Mehrabikia
Broadcast Engineering Department, Islamic Republic of Iran Broadcasting University (IRIBU)

Received: 2021/12/15, Accepted: 2022/12/28

Abstract

In this research, TDNN model and x-vector are presented in order to robust noise and frequency filtering caused by telephone communication. MFCC is used as the speaker-related audio feature as input to this model. The output of neural network of this model is considered as an x-vector so that it can be used in the decision stage. In the decision stage, PLDA was used for scoring and comparison. In order to increase accuracy and reduce EER, the training dataset is a combination of relatively clean VoxCeleb 1,2 dataset and Callhome telephone dataset, as well as noise and telephone dataset obtained from the data augmentation method. The results of using this method for EER in the clean state are 3.09%, which has improved about 0.15% (3.24% has been obtained in previous works) in the worst case and 6.93% (10.2% has been obtained in previous works) in the best case compared to the base models. When training with Voxceleb1,2 and Callhome datasets was used as an adaptation, the EER was 4.95%. In the worst case, when only the Voxceleb1 data is converted to a telephone, the EER is 14.34%.

Keywords: Speaker verification, Time delay neural network, x-Vector, Mel frequency cepstral coefficients, Probability linear discriminant analysis.

pp. 11-20 (In Persian)

* Corresponding author E-mail: m.asgari@iribu.ac.ir

[DOR: 20.1001.1.23455748.1401.10.2.2.3]