Abstract

END-TO-END SPEAKER RECOGNITION USING DEEP NEURAL NETWORK

The present invention is directed to a deep neural network (DNN) having a triplet network architecture, which is suitable to perform speaker recognition. In particular, the DNN includes three feed-forward neural networks, which are trained according to a batch process utilizing a cohort set of negative training samples. After each batch of training samples is processed, the DNN may be trained according to a loss function, e.g., utilizing a cosine measure of similarity between respective samples, along with positive and negative margins, to provide a robust representation of voiceprints.

(FIG. 2A)