Summary report for project

RATS - Robust Automatic Transcription of Speech

For year 2015

Submitted to Raytheon BBN Technologies

By Brno University of Technology

Lead author: Dr. Pavel Matějka

We present a design of a DNN-based autoencoder for speech enhancement and its use for speaker recognition systems in a domain of distant microphone and noisy data. We started with augmenting the Fisher database with artificially noised and reverberated data and trained the autoencoder to map noisy and reverberated speech to its clean version. We use the autoencoder as a preprocessing step in the later stage of modelling in state-of-the-arttest-dependent and text-independent speaker recognition systems. We report relative improvements up to 50% for the text-dependent system and up to 48% for the text-independent system. With textindependent system we present a more detailed analysis on various conditions of NIST SRE 2010 and PRISM suggesting that the proposed preprocessig is a promising and efficient way to build a robust speaker recognition system for distant microphone and noisy data.

We also studied the usage of the Deep Neural Network (DNN) Bottleneck (BN) features together with the traditional MFCC features in the task of i-vector-based speaker recognition. We decouple the sufficient statistics extraction by using separate GMM models for frame alignment, and for statistics normalization and we analyze the usage of BN andMFCC (together with their tandem variant) features in the two stages. We also show the effect of using full-covariance GMMmodels, and, as a contrast, we compare the result to the recent DNN-alignment approach. On the NIST SRE2010, telephone condition, we show 60% relative gain over the traditional MFCC baseline for EER (and similar for the NIST DCF metrics), resulting in 0.94% EER.