Summary report for project

NTT - Speech enhancement front-end for robust automatic speech recognition with large amount of training data

For year 2017

Submitted to NTT Corporation

By Brno University of Technology

Lead author: Katerina Zmolikova

Recently, schemes employing deep neural networks (DNNs) for extracting speech from noisy observation have demonstrated great potential for noise robust automatic speech recognition. However, these schemes are not well suited when the interfering noise is another speaker. To enable extracting a target speaker from a mixture of speakers, we have recently proposed to inform the neural network using speaker information extracted from an adaptation utterance from the same speaker. In our previous work, we explored ways how to inform the network about the speaker and found a speaker adaptive layer approach to be suitable for this task. In our experiments, we used speaker features designed for speaker recognition tasks as the additional speaker information, which may not be optimal for the speaker extraction task. In this paper, we propose a usage of a sequence summarizing scheme enabling to learn the speaker representation jointly with the network. Furthermore, we extend the previous experiments to demonstrate the potential of our proposed method as a front-end for speech recognition and explore the effect of additional noise on the performance of the method.