Far-field speech recognition

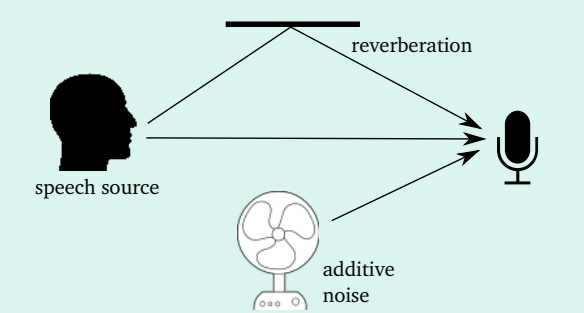
Kateřina Žmolíková (*izmolikova@fit.vutbr.cz*), Doc. Dr. Ing. Jan Černocký

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Problem

Speech recognition system nowadays achieve very good performance and start to be used in real world applications.

However, using **distant microphones** introduces lots of distortions which significantly degrade the accuracy. This degradation is caused mainly by **additive noise** and **reverberation**.



Solution

Common way to reduce the problem is the usage of microphone arrays instead of a single microphone. This enables to spatially select the signal of interest.

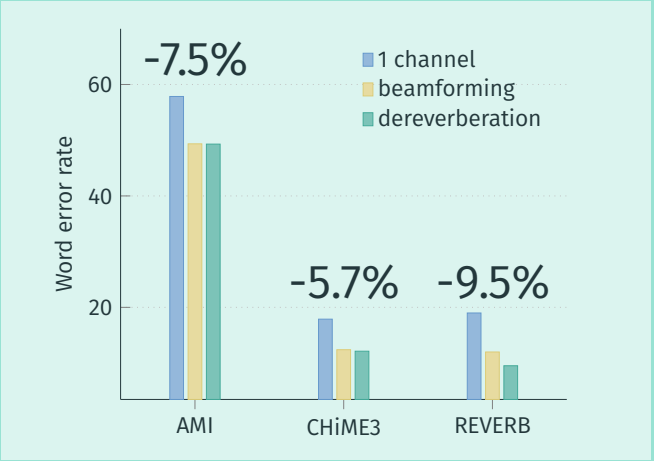
To combine the signals from multiple microphones we used **Delay-and-sum** and **Minimum variance distortionless response beamforming**.

To dereverberate the signal we applied **Weighted prediction error** method.

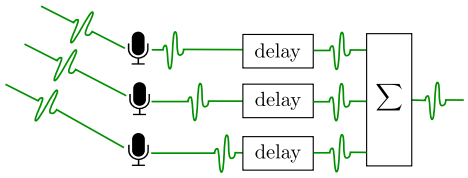
We also experimented with new approach of beamforming using **deep neural networks**.

Results

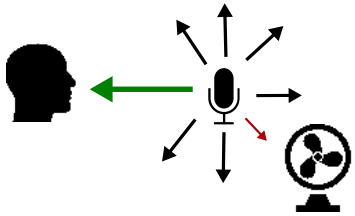
We evaluated the techniques on 3 speech recognition tasks - **CHiME3**, **AMI**, **REVERB**. The results have shown the effectiveness of used methods on all three datasets.



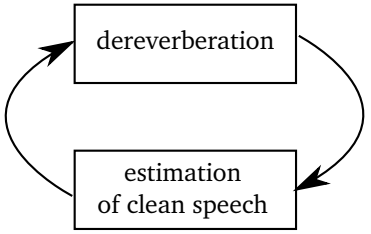
Used methods



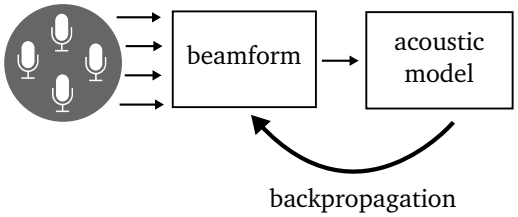
Delay-and-sum



Minimum variance distortionless response



Weighted prediction error



DNN beamforming