Speech Signal Enhancement for Audio Forensics: An Initial Study

G. Korvel[1], G. Tamulevičius[1], J. Devenson[2]

[1] Institute of Data Science and Digital Technologies, Vilnius University

[2] Forensic Science Centre of Lithuania

Introduction

Audio Forensics is the field of Forensic Science relating to analyzing, processing, and evaluating audio recordings. The primary aspects of Audio Forensics are:

- The authenticity of audio evidence and integrity of audio recordings;
- Reconstructing crime or accident scenes and timelines based on acoustic events;
- Identification of speaking persons;
- Transcription of speech content;
- Performing enhancement of audio recordings to improve speech intelligibility and the audibility of noisy or low-level sounds, etc.

Challenges

Noise is one of the most common problems in audio forensics. Noise is a major hindrance to person identification, the determination of the content of the phonogram, and also can be used to mask editing traces.

The main noise types and sources in the recordings:

- Additional sources of high-energy speech and noise;
- Various impulse noises;
- Echo and reverberation effects;
- Nonlinear distortions;
- Lost segments of the target signal;
- Limited recording quality.

In most cases the recordings are single-channel, therefore, no more additional noise data is available.

Considering the aforementioned, speech signal enhancement becomes a severe challenge. We need efficient enhancement techniques to provide a reliable audio forensic analysis.

State-of-the-art

Audio Forensic practice is based on specialized tools (e.g., iZotope), and the speech enhancement process consists of 2 main steps: identification of noise source/type and the removal of it. This task is carried out by human experts and requires a great deal of experience and concentration, labor time.

In the academic research domain, a great variety of different speech enhancement techniques is proposed:

- Filtering approach. Various filters are applied to remove the noise from the target signal. These include Wiener-type filters, adaptive filters.
- Spectral restoration approach: spectral subtraction technique, template-based noise removal, minimum MSE short-time spectral amplitude estimator.
- Model-based approach. Parametric models are applied to generate the clean speech signal: linear predictive coding model, autoregressive model.
- Artificial intelligence approach. It exploits deep learning neural networks, which have become dominant during the last decades. The idea is based on the data-driven paradigm that all differences and similarities of clean and noisy speech signals can be generalized by nonlinear network structures, which can be applied to speech signal enhancement.

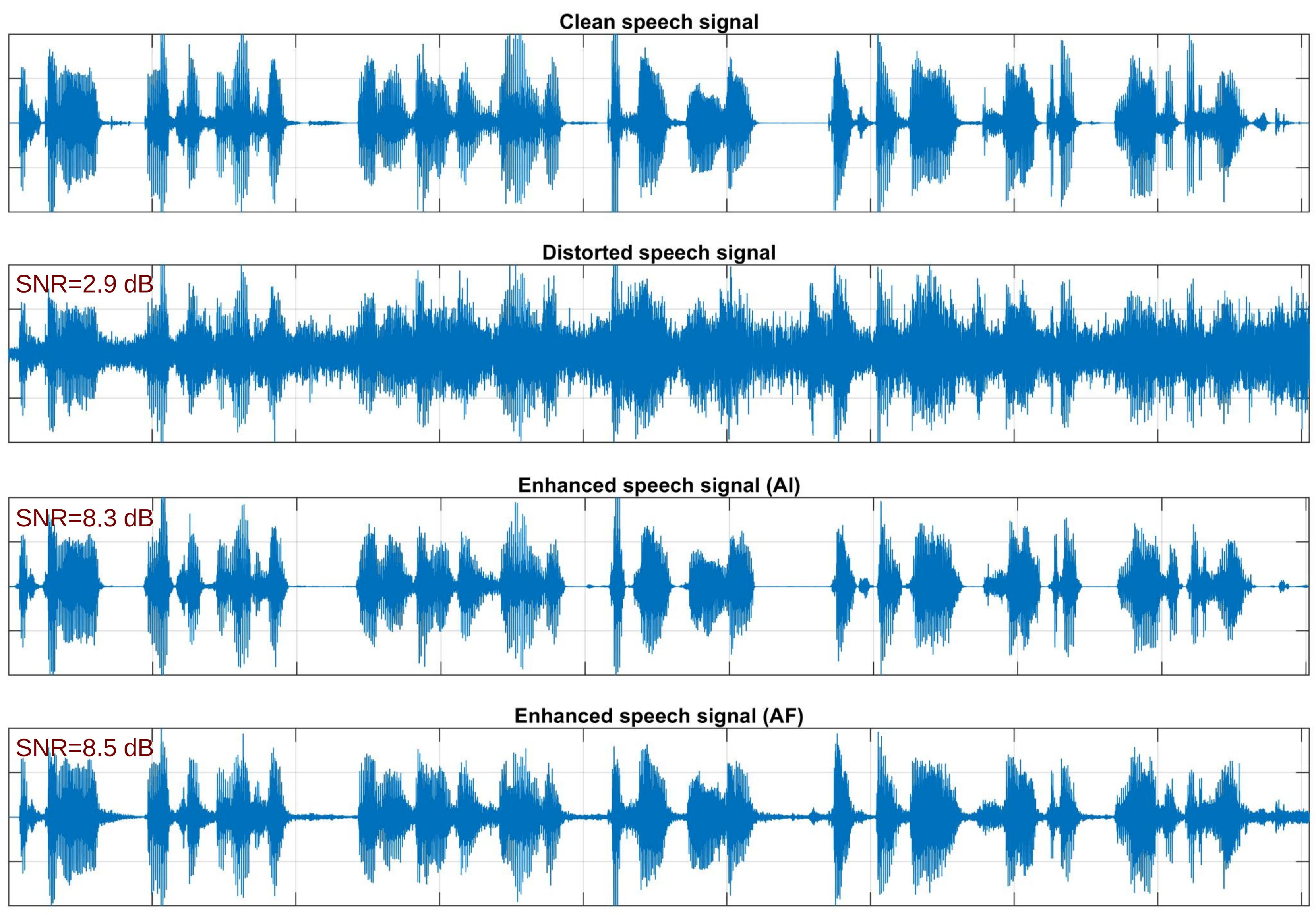
Still, the task of efficient speech enhancement is open.

Experiment case

The main objective of speech enhancement is to increase speech signal quality and intelligibility.

During the initial study two different enhancement approaches were applied for speech signal with additive babble noise:

- The artificial intelligence approach: fully-connected 4-layer (513-513-513-513 neurons, ReLU activated) feed-forward neural network, working on short-term magnitude spectra data.
- The filtering approach: 220th order adaptive FIR filter (LMS estimation).



The exemplar case of signal enhancement using the Artificial intelligence approach (AI) and the filtering approach (AF)

Objective results:

- The SNR values have increased.

Subjective results:

- The intelligibility has remained unchanged or decreased.
- The artificial intelligence approach eliminates too many speech signal harmonics.
- The filtering approach does not sufficiently remove noise components.

Both applied approaches require adaptation/training steps and a priori data on clean speech (noise).

Research directions

Directions for future research:

- Analysis of real-world data and generation of artificial data;
- Investigation of artificial intelligence-based techniques;
- Investigation of filters and model-based techniques;
- Combination of techniques: artificial intelligence-based estimation of filter parameters, the data filtering for training purposes, cascading (different) techniques, fuzzy connection of techniques.
- Comparison and evaluation of speech signal quality measures.

The team

Vilnius University
Institute of Data Science and Digital Technologies
Image and Signal Analysis Group

https://www.mii.lt/en/