TECHNICKÁ UNIVERZITA V LIBERCI
Fakulta mechatroniky, informatiky
a mezioborových studií

SKFM 2021
Studentská konference Fakulty mechatroniky,
informatiky a mezioborových studií

Semi-Blind Speaker Extraction Performed On-line on Dense Microphone Array

J. Čmejla, T. Kounovský, J. Janský, J. Málek, M. Rozkovec, and Z. Koldovský

We present an experimental device for speaker extraction and physical tracking and demonstrate its use in real conditions. The device is equipped with a dense planar array consisting of 64 microphones mounted on a rotating platform. Blind source extraction algorithms controlled by x-vector piloting are used to extract the desired speaker, which can be physically tracked by the rotating microphone array.

Keywords: source extraction, source tracking, x-vectors, Independent Vector Extraction

Introduction

We present the experimental device, shown in Fig.1, and demonstrate it's speaker extraction capabilities. The device consists of a rotating array of microphones, a stepper motor, and electronic circuitry for control, sampling, and signal transmission. The device is connected to a computer running MATLAB, where advanced methods for semi-blind speaker extraction are processing signals on-line. The device output yields an extracted signal of the desired speaker.

Algorithm

The algorithm's job is to take the input mixture, extract the selected source of interest (SOI), and estimate SOI's angular position w.r.t. the direction the array is currently facing. The source extraction can be performed by one of the three selected algorithms:

- CSV-AuxIVE [1],
- QuickIVE [2],
- FastDIVA [3].

All of these algorithms employ an advanced mixing model called CSV (Constant Separating Vector), which is a semi-time-variant linear mixing model that allows for movement of SOI within the current data context [3, 4]. The algorithms are used in a batch-online processing mode, i.e. they operate on a limited context of incoming data with overlaps between batches.

To ensure that the blind algorithm converges to the desired source (the so-called global permutation problem), a piloted version of the extraction algorithm is used. The pilot is a signal that is correlated with SOI's activity - in our case, we employ a speaker-identification system based on X-vectors. This system consists of a time-delayed neural network, which extracts speaker embeddings, an a PLDA classifier that compares new embeddings to known speaker embeddings (prerecorded one noiseless utterance of each speaker in advance) [5]. This allows the system to find frames where SOI is dominant, i.e. frames that allow the extraction algorithm to easily converge on the correct source.

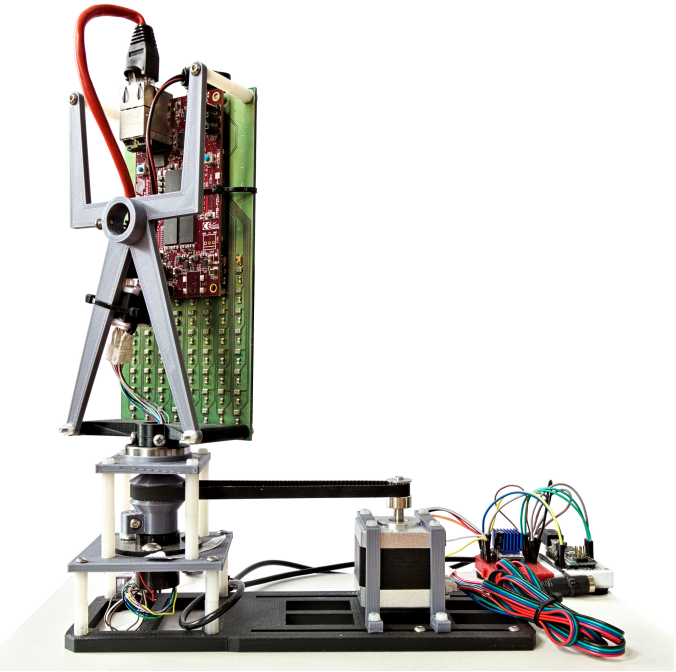

Figure 1: 64 microphone array board with a rotary platform. 20×20 cm version of the board.

Device

The device consists of 2 distinct parts: the microphone array and the rotary platform. The microphone array contains 64 MEMS microphones arranged in a vertical planar grid - 8×8 with 2 cm equidistant spacing. The sampled signals from the microphones are processed by an FPGA controller, which handles synchronization, batching, serialization and transfer to the main PC via Ethernet using the standard Transmission Control Protocol (TCP).

The rotary platform allows unlimited rotation of the array. The construction has been designed for and manufactured on a regular FDM 3D printer. The rotation is driven by a NEMA 17 stepper motor connected to the main shaft by a GT2 timing belt. Noiseless operation is achieved by using a TMC2209 silent stepper driver, controlled by an Arduino Nano microcontroller, which receives commands using a serial interface.

Software

Device control software (screenshot shown in Fig. 2) is implemented in MATLAB.

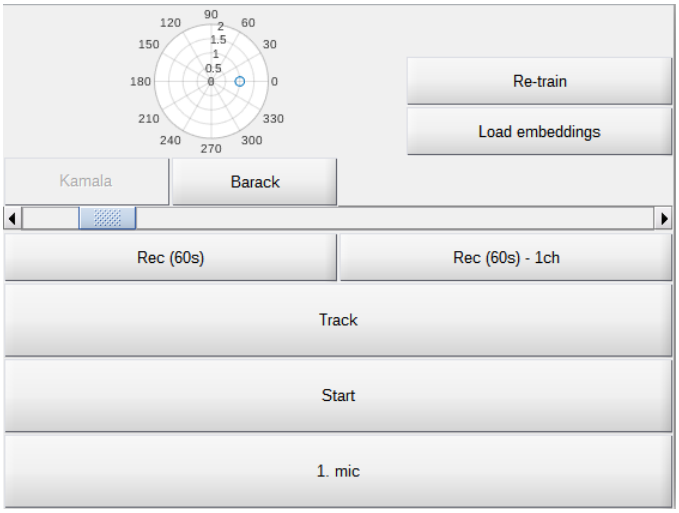

Figure 2: Device control software interface (Matlab GUI).

The software handles the following tasks:

- communication and data acquisition from the microphone board controller,
- on-line source extraction and source tracking calculations,
- communication with the stepper motor controller.

The GUI provides an easy way of selecting the current source of interest from a list of available speakers. This list contains distinct speakers for whom an enrollment data is available. This list can be modified at any time, i.e. adding new speakers on-the-fly is possible, provided that a short and relatively noiseless utterance can be obtained.

For easy comparison, it is possible to switch between two output audio modes - the user can listen to the mixture exactly as it is being recorded by the 1st microphone (i.e. noisy), or to the output of the source extraction algorithm.

The GUI also contains a switch that enables or disables the tracking functionality. Once tracking is enabled, the algorithm's estimate of the SOI's angular position relative to the current facing of the array (shown in a polar plot) will be sent to the motor controller, which will turn the array in towards the source.

On-line DEMO

The device will be accessible during the conference for a live presentation. Visitors will be able to inspect the device itself as well as the signal processing system to evaluate the extraction algorithm's performance.

Acknowledgment

This work was supported by the Student Grant Scheme of the Technical University of Liberec (2021).

References

[1] J. Janský; Z. Koldovský; J. Málek; T. Kounovský; and J. Čmejla. Auxiliary function-based algorithm for blind extraction of a moving speaker. *arXiv 2002.12619*, 2020

[2] Z. Koldovský; V. Kautský; T. Kounovský; and J. Čmejla. Algorithm for independent vector extraction based on semi-time-variant mixing model, *arXiv 1910.10242*, 2021.

[3] Z. Koldovský; V. Kautský; P. Tichavský; J.Čmejla, and J. Málek. Dynamic independent component/vector analysis: Time-variant linear mixtures separable by time-invariant beamformers, *IEEETransactions on Signal Processing*, vol. 69, pp. 2158–2173, 2021.

[4] N. Amor; J. Čmejla; V. Kautský; Z. Koldovský; and T. Kounovský. Blind extraction of moving sources via independent component and vector analysis: Examples, in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 3725–3729.

[5] J. Janský; J. Málek, J. Čmejla; T. Kounovský; Z. Koldovský; and J. Žďánský. Adaptive blind audio source extraction supervised by dominant speaker identification using x-vectors, in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 676–680.