Source Separation in Noisy and Reverberant Environment using Miniature Microphone Array

Shuo Li and Milutin Stanačević
Department of Electrical and Computer Engineering
Stony Brook University
Stony Brook, NY 11794–2350
Email: shuo.li,milutin.stanacevic@stonybrook.edu

Abstract—In the unique framework that combines spatial sensing and independent component analysis, we recover the impinging acoustic sources and estimate their direction of arrival on the miniature four microphone array. We examine and quantify the performance of the proposed algorithm under different acquisition noise and reverberant conditions. With artificially generated microphone signals using room acoustic model, the algorithm demonstrates over 10 dB separation in moderate reverberant environments. With the recordings from a miniature microphone array in a typical conference-room environment, the algorithm demonstrates over 10 dB separation performance if the angular distance between two speech sources is over 30 degrees.

I. INTRODUCTION

The human ears have the innate capability of producing remarkable results in separating multiple sound sources by resolving time delays and intensity differences between sound waves with binaural observations, and correlating these differences across various source components, even under very noisy environments [1]. To enable the hearing impaired people with speech intelligibility of an average person, the amplification of the microphone signal is not enough. Traditional hearing aids perform poorly in the rich acoustic environments where multiple sources or significant background noise exists. Although modern hearing aids have an added functionality of binaural sensing to a certain degree by utilization of the directional microphones, their performance still deteriorates significantly when multiple speakers and noise sources are present. Thus we need to fully enable the separation ability of the smart sensing hearing aids to improve their performance under complicated acoustic scenes. Moreover, the emergence and advancement of the MEMS technology empowers the promising future of nanoscale integration of the acoustic sensor array and processing circuitry [2], [3].

Common wisdom requires that sensor arrays with large inter-sensor spacing should be used for source separation and beamforming to guarantee sufficient spatial difference across sensors to resolve time delays between source observations [4]. Most of the current proposed implementations employ microphone arrays with at least 4cm inter-microphone spacing. For applications such as hearing aids, a small-form factor microphone arrays, with inter-sensor spacing much lower than the smallest wavelength are desirable. Gradient flow is a signal conditioning technique that can estimate the direction of sound propagation directly from sensing spatial and temporal gradients of the wave signal on a sub-wavelength scale [5]. The motivation comes from the parasitoid fly *Ormia ochracea*, which localizes its sound-emitting prey by a beamforming acoustic sensor of dimensions a factor 100 smaller than the acoustic signal wavelength [6]. It has been discovered that evolution has empowered the fly with a hearing mechanism that utilizes multiple vibration modes to amplify interaural time and level differences to get the directional information out of the sound its tympanal beamforming organ senses. We have evaluated the performance of the gradient flow in a localization of a single acoustic source in adverse noisy and reverberant environments. The simulation results demonstrate robust performance with mild to moderate reverberations with additive noise levels down to 10dB [7].

With the gradient flow representation in the case of multiple acoustic sources present, the acoustic mixture can be easily identified in the form of static independent component analysis (ICA) mixing process, with the mixing coefficients being time delays of each sources, thus exhibiting the directional information. We then apply static ICA techniques to solve the separation of the mixed signals captured by the acoustic sensor array in the moderate reverberation environment, while acquiring the geometric cues at the same time. The proposed ICA implementation is utilized on various synthesized and recorded speech signal under different acoustic scenes to validate the separation performance dependance on levels of reverberation and environmental noise. Both results are presented and illustrated in the following sections.

II. GRADIENT FLOW ICA ARCHITECTURE

Gradient flow [5] is a signal conditioning technique for sensor arrays of very small dimensions illustrated in Figure 1, which in anechoic environment converts time delays between signal observations into relative amplitudes of the time-differentiated signal by observing gradients (spatial differences). The 3-D directional vector of the propagation wave is uniquely defined by time delays $\tau_1$ and $\tau_2$ of the source along the $p$ and $p$ directions in the sensor plane. In the case of multiple sources impinging the microphone array, the first-order spatial gradients of the observed signals $\xi_{10}$ and $\xi_{01}$ in $p$ and $q$ directions around the origin $(p = q = 0)$ and the spatial
common mode $\xi_{00}$ are:

$$\xi_{00}(t) = \sum_l \tau^i_l s^l(t)$$
$$\xi_{10}(t) = \sum_l \tau^i_1 s^l(t)$$
$$\xi_{01}(t) = \sum_l \tau^i_2 s^l(t)$$

where $\tau^i_1$ and $\tau^i_2$ are the time delays of the source signal $l$ in $p$ and $q$ direction, respectively. Taking the time derivative of $\xi_{00}$ and observing the three spatial gradients, the mixture of delayed sources is converted into a linear instantaneous mixture of time-differentiated source signals in the form of

$$\begin{bmatrix}
\xi_0 \\
\xi_{10} \\
\xi_{01}
\end{bmatrix} =
\begin{bmatrix}
1 & \cdots & 1 \\
\tau^1_1 & \cdots & \tau^M_1 \\
\tau^1_2 & \cdots & \tau^M_2
\end{bmatrix}
\begin{bmatrix}
\dot{s}^1 \\
\dot{s}^2 \\
\vdots \\
\dot{s}^M
\end{bmatrix}.$$  

This form corresponds to the classic linear mixing form of $x = As$, which can be solved with ICA algorithm. In this case, the observation vector $x$ is the temporal and spatial derivatives acquired from the sensor observations $x_{-10}, x_{10}, x_{0-1}$ and $x_{01}$ as:

$$\begin{align*}
\xi_{00} & \approx \frac{1}{2}(x_{-10} + x_{10} + x_{0-1} + x_{01}) \\
\xi_{10} & \approx \frac{1}{2}(x_{10} - x_{-10}) \\
\xi_{01} & \approx \frac{1}{2}(x_{01} - x_{0-1}).
\end{align*}$$

The original source signal vector $s$ corresponds to the time differentiated version of $x = As$, which can be solved with ICA algorithm. In this case, the observation vector $x$ is the temporal and spatial derivatives acquired from the sensor observations $x_{-10}, x_{10}, x_{0-1}$ and $x_{01}$ as:

$$\begin{align*}
\xi_{00} & \approx \frac{1}{2}(x_{-10} + x_{10} + x_{0-1} + x_{01}) \\
\xi_{10} & \approx \frac{1}{2}(x_{10} - x_{-10}) \\
\xi_{01} & \approx \frac{1}{2}(x_{01} - x_{0-1}).
\end{align*}$$

For this linear mixing model, ICA solution is formulated as a linear transformation that minimizes the statistical dependence between components in the output signals $y$

$$y(t) = Wx(t),$$

where $W$ is $N \times M$ dimensional unmixing matrix, where $N$ is the number of source and $M$ is the number of observations. The unmixing matrix $W$ is not uniquely defined, with ambiguity in scaling and permutation. The energy of the source signals cannot be determined, since both $s$ and $A$ are unknown and any scalar multiplier in one of the sources could be canceled by dividing the corresponding column of $A$ by the same scalar. Most of the ICA learning algorithms proposed in the literature are based on optimizing a cost function defined as the measure of independence between the components of the output signals $[8]$. Different approaches, like maximization of entropy $[9]$, minimization of mutual information of the output signals $[8], [10]$ and the maximization of likelihood function $[11]$, lead to the same form of the cost function.

III. SIMULATION RESULTS

The effect of the additive sensor noise and the amount of reverberation on the separation performance using both synthesized and recorded data has been studied. In the performed simulations, we assume that source signals impinge a microphone array comprising four microphones in configuration shown in Figure 1, where the microphones are in planar orthogonal positions and the distance between the opposing microphones is 1 cm. To quantify the performance under different reverberant conditions, we first constructed an artificial room using virtual source mapping model $[12]$ to generate observations at the microphone array in an echoic room environment. The room is selected to be an ordinary office space and its dimensions are [6 m, 4 m, 2.5 m], while the distance between the microphone array and the source signals is 1.5 m. The relative location of the sensor array and the speech sources within the simulated room is illustrated in Figure 2. Reverberation coefficients of all six surfaces of the simulated room were considered uniform and generated based on the room dimensions and different reverberation times. Various synthesized sensor data was generated based on different reverberation times and additive noise level. The separation performance is defined through signal-to-interference ratio (SIR), which is computed as

$$SIR = -10 \log_{10} \min_i \sum_j \frac{y_{ij}^2}{\max_j \frac{y_{ij}^2}{\max_j < y_{ij}^2 >}},$$
where $y_{ij}$ is the contribution of the source signal $j$ to the estimated source signal $i$ and is considered as interference.

A. Separation Performance with Additive Measurement Noise and Different Reverberation Levels

To demonstrate a benchmark performance of the algorithm, we first examined the effect of the angular distance between two sources on the separation performance. The two source signals used in the characterization of the algorithm are two speech segments chosen from the TIMIT database with approximately the same signal power. The sampling frequency is 16 kHz, while the length of the signal is 1 s. As the static ICA algorithm, the efficient FastICA(EFICA) algorithm [13] was used. The same elevation angles were assumed for both sources and the azimuth angle of the first source was set at $\theta_1 = 30^\circ$. The azimuth angle of the second source $\theta_2$ was varied from $-15^\circ$ to $135^\circ$ in increments of $15^\circ$. We have omitted the locations of the second source at $0^\circ$ and $90^\circ$ where the separation is trivial. The separation performance, SIR, as a function of the azimuth angle of the rotating source is shown in Figure 3. The measurement results demonstrate that the separation performance does not depend strongly on the angular separation.

In the second experiment, the effect of the acquisition noise on the separation performance is investigated. We assumed that a white, spatially uncorrelated Gaussian noise sources are added to each sensor. The source signals were located at the incidence angles of $30^\circ$ and $70^\circ$. The results for different signal-to-noise ratios(SNR) are presented in Figure 4. The SIR strongly depends on the measurement SNR and we can notice that the separation performance is above 10 dB until the SNR reaches 20 dB.

In the mixing model adopted in the gradient flow representation (2), the anechoic environment is assumed. However, in the real room environment, due to reverberations the observed microphone signal is a sum of multi-path replicas of the source signal, that is a sum of time-delayed and attenuated source signals, where the delays and attenuations depend on the room geometry and the reflection coefficient of the walls. The Figure 5 shows the separation performance under different reverberation conditions, with different levels of additive sensor/acquisition noise in a simulated echoic room environment. The separation performance degrades with the increase of the reverberations, but satisfying performance is demonstrated in mild reverberant conditions. The separation is sustained as long as the direct path signal is stronger than the multi-path signals.
The separation performance for two speech signals recorded using miniature microphone array in a typical conference room environment as a function of incidence angle $\theta_2$. The incidence angle of the first source is $\theta_1 = 30^\circ$.

### B. Room Speech Separation Experiments

To further examine the performance of the gradient flow algorithm, we devised a planar array of four omnidirectional hearing aid miniature microphones to record speech signals in a typical conference room. The spacing between opposing microphones is kept at 1 cm. A single acoustic source was presented through a loudspeaker positioned at 1 m distance from the array and the recordings were repeated for different speech signals at various incident angles with respect to the array. The corresponding SNR was around 35 dB. The two recordings from two different speech sources at each microphone were added. This mixing scenario, although more noisy, enables the quantification of the separation performance as opposed to simultaneous recordings of two signals. The incident angle of the first source was kept at $30^\circ$ and the incidence angle of the second source was swept from $15^\circ$ to $-45^\circ$.

The contribution of the second source in the estimated first source as a function of the direction of the second source is shown in Figure 6. As we could observe from the plot, the level of separation decreases as the angular difference becomes smaller and two sources are getting closer, which is corresponding to common sense. Yet as long as the angular difference is larger than $30^\circ$, the SIR could stay above 10 dB which is very impressive for a rather noisy real environment source separation.

### IV. Conclusion

The proposed gradient flow acoustic source separation technique is presented and evaluated under adverse reverberant and noisy situations. The demonstrated source separation performance using miniature microphone array is comparable to the performance of the state-of-the-art hearing aids. These results suggest application of the gradient flow system integrating spatial sensing and static ICA with miniature microphone arrays to intelligent hearing aids with adaptive suppression of interfering signals and nonstationary noise. With the amenable low-power analog VLSI implementation, the proposed gradient flow algorithm presents ideal candidate for the adverse set of applications of the miniature microphone arrays. The separation performance of the proposed system can be extended to moderate reverberation environment by using subband decomposition of spatial gradient signals and static ICA applied in each frequency band.

### V. Acknowledgement

This work was supported by NSF CAREER Award 0846265.

### References