Subband Gradient Flow Acoustic Source Separation for Moderate Reverberation Environment

Shuo Li and Milutin Stanaćević
Department of Electrical and Computer Engineering
Stony Brook University
Stony Brook, NY 11794–2350
Email: shuoli,milutin@ece.sunysb.edu

***Abstract*— We present a subband source separation algorithm for miniature microphone arrays with dimensions smaller than the wavelength. By relating temporal and spatial gradients of the observed microphone signals in an anechoic environment, gradient flow converts the mixture of delayed sources to linear instantaneous mixture of the time-differentiated source signals, that can be then localized and separated using static linear independent component analysis algorithms. For source separation in multi-path environment, we propose subband decomposition of the spatial gradients estimated over an array of 4 microphones. The static ICA algorithms are applied in each frequency band and the localization results obtained from the ICA applied on the unfiltered spatial gradients resolve the scaling and permutation indeterminacy. The simulations with the room acoustic model and experimental results with conference room recordings demonstrate over 12dB separation in moderate reverberation environment.**

I. Introduction

Smart sensing hearing aids is one of the areas where nanoscale integration using MEMS technology promises significant breakthrough [1], [2]. The speech intelligibility of traditional hearing aids is limited with multiple sources and environmental noise present in the acoustic scene. In order for hearing aids to obtain intelligibility, smart sensing is required to suppress the noise sources based on the spatial location or spectral content. The human auditory system resolves time delays and intensity differences between sound waves of binaural observations, and correlates these differences across various source components to produce incredible results in segregating multiple sound sources, even under a very noisy environment [3]. Modern hearing aids utilize directional microphone arrays to add some of the functionality of binaural sensing. Yet their performance still degrades significantly when multiple sources and noise are present [4]. To effectively solve the signal of interest, both localization and separation of multiple acoustic sources are required.

In a typical room environment the acoustic signals observed by microphone array are convolutive mixtures of source signals due to multi-path wave propagation [5]. Implementation of the time-domain blind source separation algorithms requires resolving of a large number of unmixing filter coefficients with high computational cost and degrading algorithm convergence [6]. To alleviate these issues, frequency domain algorithms have been introduced. However, these algorithms suffer from the inherent ambiguity of permutation and scaling of independent component analysis (ICA). To solve the permutation and scaling indeterminacy, the source location information obtained through adaptive beamforming has been used in the frequency domain algorithms [7], [8], [9]. Conventional knowledge dictates that sensor arrays with large inter-sensor distance should be used for source separation and beamforming to warrant sufficient spatial diversity across sensors to resolve time delays between source observations. Most of the proposed methodologies employ microphone arrays with at least 4 cm inter-microphone spacing. For applications like hearing aids, a small-form factor microphone arrays, with the spacing much lower than the wavelength are required. 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 [10]. The inspiration comes from the parasitoid fly, which localizes its sound-emitting prey by a beamforming acoustic sensor of dimensions a factor 100 smaller than the acoustic signal wavelength [11]. Its tympanal beamforming organ senses acoustic pressure gradient, rather than time delays, in the incoming wave.

Using the gradient flow representation, we propose the subband ICA architecture to improve the separation of the mixed signals beyond the direct path signal separation in the moderate reverberation environment. The proposed technique consists of static ICA separation applied on the unfiltered spatial gradient signals and static ICA applied separately in each frequency band. Due to the localization performed inherently by the static ICA on the unfiltered spatial gradients, the permutation and scaling ambiguity of the subband ICA is resolved in the gradient flow representation and provide improved separation performance under moderate reverberations.

II. Spatial Wavefront Sensing and Linear Static ICA

In the case of instantaneous linear mixing of the source signals the observations at the sensor array can be written in a form

$$\mathbf{x} = A\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{x}$ is the vector of M observation signals at the sensor array, $\mathbf{s}$ is the vector of the original N source signals and

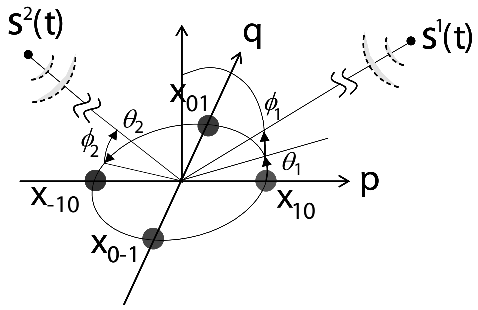

Fig. 1. Miniature microphone array used in gradient flow technique.

A is the MxN mixing matrix. $\mathbf{n}$ is the additive noise at the sensor array. The problem of the blind source separation can be formulated as the search for a linear unmixing matrix W in order to estimate the original sources with no priori knowledge on the source signals and the mixing medium

$$\mathbf{y} = W\mathbf{x} \tag{2}$$

In this well-defined classic blind source separation problem ICA techniques can solve the separation problem very well under a fair amount of additive noise. However, when the real acoustic scenes is considered, the classic linear mixture model will no longer be valid.

In the case of the travelling acoustic wave signals impinging on an array of four microphones, as illustrated in Figure 1, the signals observed at the sensor array are mixture of the delayed source signals. Gradient flow [10] is a signal conditioning technique for source localization and separation designed for sensor arrays of very small aperture, of which the dimensions are significantly smaller than the shortest wavelength in the sources. The 3-D directional vector of the traveling wave is uniquely defined by propagation delays τ_1 and τ_2 of the source along the p and q directions in the sensor plane. In the case of a single source, direct calculation of these small interaural time difference (ITD) is troublesome as they require sampling in excess of the bandwidth of the signal, increasing noise floor and power consumption. However, indirect estimates of the delays are obtained through least-square regression as the first order spatial gradients along the p and q direction are proportional to the temporal derivatives of the average signal at the center of the array, where the linear coefficients are the propagation delays τ_1 and τ_2.

In the case of multiple sources impinging the microphone array, the first-order spatial gradients of the observed signals ξ_{10} and ξ_{01} in p and q directions around the origin ($p = q = 0$) and the spatial common mode ξ_{00} are:

$$\begin{aligned} \xi_{00}(t) &= \textstyle\sum_l s^l(t) \\ \xi_{10}(t) &= \textstyle\sum_l \tau_1^l \dot{s}^l(t) \\ \xi_{01}(t) &= \textstyle\sum_l \tau_2^l \dot{s}^l(t) \end{aligned} \tag{3}$$

where τ_1^l and τ_2^l are the time delays of the source signal l in p and q direction, respectively. Taking the time derivative of ξ_{00} and observing the three spatial gradients, the mixture of delayed source sources is converted into a linear instantaneous mixture of time-differentiated source signals in the form of classic linear static ICA (1)

$$\begin{bmatrix} \dot{\xi}_{00} \\ \xi_{10} \\ \xi_{01} \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 \\ \tau_1^1 & \cdots & \tau_1^M \\ \tau_2^1 & \cdots & \tau_2^M \end{bmatrix} \begin{bmatrix} \dot{s}^1 \\ \vdots \\ \dot{s}^M \end{bmatrix}. \tag{4}$$

The mixing matrix A has the special form, with its coefficients representing the time delays that uniquely determine the directions of the source signals. Therefore, by applying the static ICA on the three gradient signals, along the recovery of the source signals, the location of the sources is simultaneously obtained.

III. Subband Gradient Flow ICA Architecture

In a real room environment, reverberations will introduce a series of attenuated, time-delayed components to the original direct-path signals observed on the microphone array leading to convolutive mixing source separation problem. In the convolutive mixing model, each element of the mixing matrix A in the model (1) is a filter instead of a scalar and the i-th observed signal can be written as

$$x_i(t) = \sum_{j=1}^{n} \sum_{k} a_{ijk} s_j(t-k). \tag{5}$$

Frequency domain techniques are attractive for solving the convolutive mixtures, as the convolution becomes product in the frequency domain

$$X_i(\omega) = \sum_{j=1}^{n} A_{ij}(\omega) S_j(\omega), \tag{6}$$

where $X_i(\omega)$, $S_j(\omega)$ and $A_{ij}(\omega)$ are the Fourier transforms of i-th observation signal $x_i(t)$, j-th source signal $s_j(t)$ and mixing filter that describes the contribution of j-th source to i-th observation $a_{ij}(t)$. The convolutive mixture model is transformed into an instantaneous linear ICA model in each frequency bin and linear static ICA techniques to determine coefficients $A_{ij}(\omega)$ can be applied. However, due to the inherent ambiguity of permutation and scaling of linear ICA solution, putting the reconstructed signal together from the separated signals in each frequency bin is not defined without additional information on the source signals. The information on the estimated location of the source signals obtained by applying the static ICA on the spatial gradient signals can be used as prior information and enable the reconstruction of the estimated source signals.

We propose the following architecture for subband gradient flow ICA shown in Figure 2. First, the temporal and spatial gradients are computed as the finite differences of the field on the sensor grid comprising four microphones in a configuration illustrated in Figure 1. The issue of permutation ambiguity is an obstacle for application of frequency-domain ICA techniques. We propose to solve the problem by applying the estimated localization results from static ICA as a preprocessing technique to assist alignment of estimated

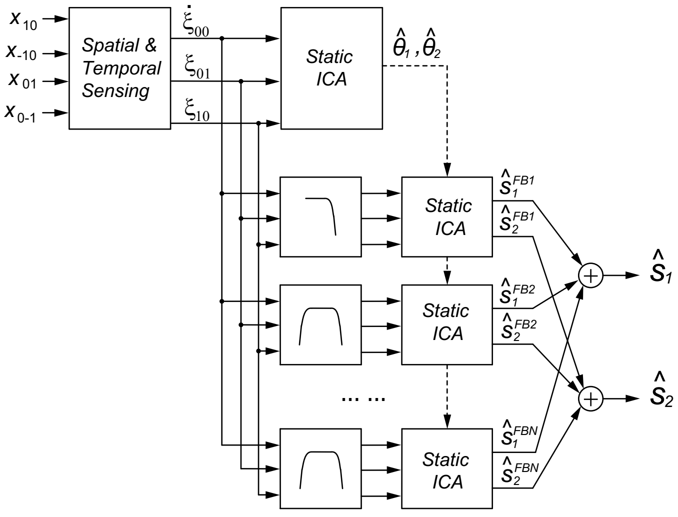

Fig. 2. Block diagram of the proposed subband gradient flow ICA architecture.

results from each frequency bin. The static ICA provides a rough estimation on the directional pattern of the incoming sources. Here we utilize it to assist the matching of the signal coming from the same direction. 16-channel filterbank is used to decompose the spatial gradient signals and static ICA algorithm is used in each frequency band to obtain the unmixing matrix and signal estimation. The solution of the static ICA applied to unfiltered spatial gradients is used as a initial point for the static ICA in each frequency band. In the moderate reverberation environment, the direct path will be the strongest source signal path and the directional information is pertained across the frequency bands. However, if the directional information in specific band strongly deviates from the direction obtained in the unfiltered static ICA, we assume that signal is not present in that frequency band. The unmixing coefficients in that frequency band are set to the initial unmixing coefficients. The scaling ambiguity has to be resolved as well because although the inconsistency of audio intensity in different frequency bins does not affect the level of separation, the aural perception can be significantly affected. In the proposed subband technique, we choose the first row of the estimated mixing matrix as the scaling factor to resolve the scaling ambiguity. Figure 3 shows the subband ICA model using the case of two sources and two microphones as an example. After applying the unmixing matrix, the separated sources are multiplied with the corresponding scalar in the first row of the estimated mixing matrix, so the final source estimated are actually the component of each source in the first observation. Then the amplitude of each estimation is uniform across multiple frequency bins. Finally, we align and synthesize the estimated signals from each frequency bin back into full-band estimations based on the preprocessing static ICA directional pattern.

Gradient flow techniques lend us the opportunity to utilize smart sensing mixed-signal circuits to achieve signal processing efficiently. While borrowing concepts from frequency-domain algorithms contributes to improved separation.

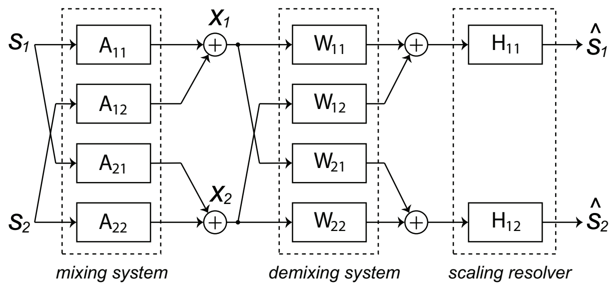

Fig. 3. Block diagram of ICA algorithm in one frequency bin.

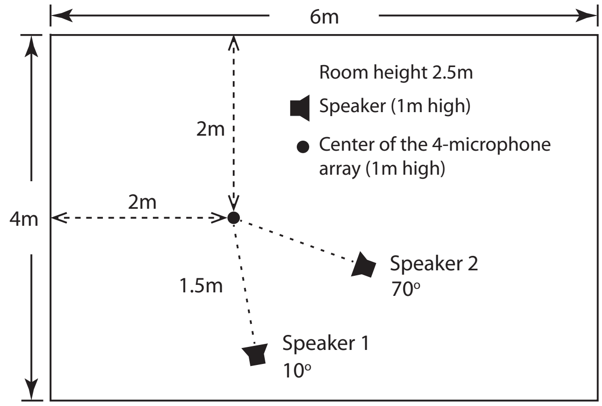

Fig. 4. Simulated room dimensions and location of the sensor array and speakers.

IV. Simulation Results

The performance of the proposed gradient flow subband ICA model is tested and quantified in simulated adverse acoustic conditions. We performed simulations with artificially synthesized microphone array signals with different reverberation times and incidence angles of the incoming source signals. The results determine the dependence of the separation results on reverberation in the room environment. As a benchmark for characterization of subband gradient flow ICA, the results obtained by the static gradient flow ICA model are also presented. In all simulations, the implemented static ICA algorithm implemented was the efficient FastICA algorithm (EFICA) [12]. EFICA is asymptotically efficient with computational complexity only slightly higher than the standard symmetric FastICA.

The artificial microphone array signals are generated based on the image model [13]. Simulated room dimensions correspond to an ordinary office space with the room dimensions and the relative location of the sensor array and the speech sources shown in Figure 4. The configuration of the microphone array is orthogonal, with 1 cm inter-microphone spacing as illustrated in Figure 1. Two speech source signals from the TIMIT database are used with the length truncated to 1.5s. The sampling frequency is 16 kHz. The incidence angle of the two sources are set to be 10^o and 70^o. A comparative simulation of

TABLE I
COMPARISON OF STATIC ICA AND FILTERBANK ICA RESULTS

	RT_{60} = 200ms		RT_{60} = 300ms	
	SIR1	SIR2	SIR1	SIR2
Static ICA	25.30dB	26.91dB	6.95dB	12.80dB
Subband ICA	24.65dB	28.99dB	10.12dB	15.75dB

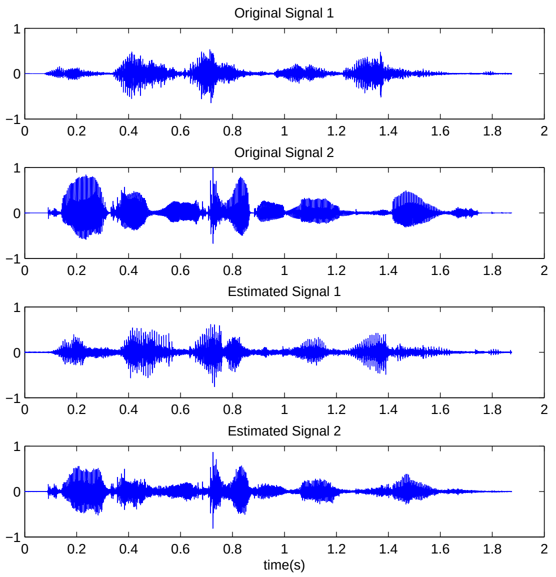

Fig. 5. Time waveforms of the presented speech sources and signals recovered by the subband ICA algorithm.

static ICA and subband ICA algorithms is executed under various reverberant environments. Table I summarizes the signal-to-interference ratio(SIR) for 200ms and 300ms reverberation times. Under mild reverberation situation, the improvement is limited because the static ICA already has a satisfying level of separation. When reverberation increases, the performance of the subband ICA is significantly higher than static ICA. Under 300ms reverberation time, which corresponds to a 0.59 uniform reflection coefficient in the simulated room environment, the SIR improvement is over 3dB.

The subband ICA is also been applied on the real-world recordings of two speech signals in a typical conference room. The speech signals were presented through loudspeakers positioned at 1.5 m distance from the array and recorded using 4 hearing aid microphone array. The relative directional angles of the speakers with respect to the center of the array were -30^o and 40^o. The distance between opposing omnidirectional miniature microphones (Knowles FG-3629) was 1 cm. The sampling frequency was set to 16 kHz. The separated signals along with the original source signals are shown in Figure 5 and demonstrate the separation of around 15dB.

V. CONCLUSION

The proposed gradient flow subband acoustic separation technique is presented and evaluated under adverse reverberant situations. The improvement over the static ICA algorithm in the gradient flow representation in moderate reverberation environments is demonstrated. The technique lends itself in mixed-signal VLSI implementation amenable to low-power, small-form-factor hearing aids and other acoustic tracking and separation applications.

VI. ACKNOWLEDGMENT

This work was supported by NSF CAREER Award 0846265.

REFERENCES

[1] R. N. Miles, Q. Su, W. Cui, M. Shetye, F.L. Degertekin, B. Bicen, C. Garcia, S. Jones and N. Hall, "A low-noise differential microphone inspired by the ears of the parasitoid fly Ormia ochracea", *J. Acoust. Soc. Am.*, vol. **125** (5), pp. 2013-2026, 2009.

[2] S. Ando, T. Kurihara, K. Watanabe, Y. Yamanishi and T. Ooasa, "Novel theoretical design and fabrication test of biomimicry directional microphone", *International Solid-State Sensors, Actuators and Microsystems Conference TRANSDUCERS 2009*, pp. 1932-1935, 2009.

[3] A.S. Bregman, *Auditory Scene Analysis, The Perceptual Organization of Sound*, Cambridge MA: MIT Press, 1990.

[4] V. Hamacher, J. Chalupper, J. Eggers, E. Fisher, U. Kornagel, H. Puder and U. Rass, "Signal Processing in High-End Hearing Aids: State of the Art, Challenges, and Future Trends", *EURASIP Journal on Applied Signal Processing*, vol. **18**, pp. 2915-2929, 2005.

[5] M.S. Pedersen, J. Larsen, U. Kjems and L.C. Para, "A Survey of Convolutive Blind Source Separation Methods", *Springer Multichannel Speech Processing Handbook*, pp. 1065-1084, 2007.

[6] R. Lambert and A. Bell, "Blind separation of multiple speakers in a multipath environment," *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP'97)*, Mϋnich, 1997.

[7] H. Sawada, R. Mukai, S. Araki and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation", *IEEE Transactions on Speech and Audio Processing*, vol. **12** (5), pp. 530-538, 2004.

[8] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee and K. Shikano, "Blind Source Separation Based on a Fast-Convergence Algorithm Combining ICA and Beamforming", *IEEE Transactions on Audio, Speech and Language Processing*, vol. **14** (2), pp. 666-678, 2006.

[9] L. Parra and C. Alvino, "Geometric Source Separation: Merging Convolutive Source Separation with Geometric Beamforming", *IEEE Transactions on Speech and Audio Processing*, vol. **10** (6), pp. 352-362, 2002.

[10] G. Cauwenberghs, M. Stanacevic, and G. Zweig, "Blind Broadband Source Localization and Separation in Miniature Sensor Arrays," *Proc.* IEEE *Int. Symp. Circuits and Systems (ISCAS'2001)*, Sydney, Australia, May 6-9, 2001.

[11] R. N. Miles, D. Robert and R.R. Hoy, "Mechanically coupled ears for directional hearing in the parasitoid fly Ormia ochracea," *J. Acoust. Soc. Am.*, vol. **98**, pp. 30593070., 1995.

[12] Z. Koldovsky, P. Tichavsky and E. Oja, "Efficient Variant of Algorithm FastICA for Independent Component Analysis Attaining the Cramer-Rao Lower Bound," *IEEE Trans. on Neural Networks*, vol. **17** (5), pp. 1265-1277, 2006.

[13] J.B. Allen and D.A. Berkley, "Image method for efficiently simulating small-room acoustics", *J. Acoust. Soc. Amer.*, vol. **65**, pp. 943950, Apr. 1979.