Mixed-signal VLSI Microsystem for Acoustic Source Separation

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

***Abstract*—We present an architecture of a microsystem designed for the task of blind source separation of acoustic sources that impinge on the miniature microphone array. The proposed architecture implements the source separation algorithm in a unique framework that combines wavefront sensing, subband signal processing and independent component analysis. The spatial gradients of the acoustic wavefront are computed in the continuous-time domain and decomposed in the 16 subbands. In each band, we perform static independent component analysis. We demonstrate that with the proposed system architecture 13 dB separation performance is achieved in moderate reverberant environment.**

I. Introduction

With the emergence and advance of microelectromechanical systems (MEMS) technology, the size of the acoustic sensors has been reduced down to a millimeter and micrometer range [1]. Smart sensing hearing aids is one of the areas where microscale integration using MEMS technology promises significant breakthrough. To comply with the requirement of small form factor and low power operation, it is desirable that the sensor array and the processing mixed-signal circuit are integrated on the same substrate. It is also critical for hearing aids and other acoustic applications to perform well in real room environments where the echos and reflections create multiple-path disturbances other than the direct-path signal.

There have been different architectures for the implementation of blind source separation algorithms. The implementations were constrained to the static independent component analysis in both analog and mixed-signal VLSI [2]–[4] and in digital domain [5], [6]. We have previously presented a system that performs source separation and localization in an anechoic environment, accounting only for the direct path signals. That microsystem comprises spatial gradient computation integrated circuit [7] and independent component analysis processor [8]. The system demonstrated 12 dB separation in mild reverberant environment. We present a system architecture that accounts for the reverberant environment by using subband signal processing and extend the separation performance to intelligible separation in the moderate reverberant environments.

II. Source Separation System Architecture

The blind source separation problem can be defined as a task to recover unknown source signals, that propagate through the unknown medium, from the signals observed by an array of sensors. The source signals are estimated using the assumption of their mutually independence. If the sources are linearly mixed, the sensors observations $\mathbf{x}$ are

$$\mathbf{x} = A\mathbf{s} \tag{1}$$

where A denotes the mixing matrix and $\mathbf{s}$ are the independent input signals. Matrix A is an $M \times N$ dimensional matrix, where N is the number of sources and M is number of sensors. The assumption is that $M > N$, since in the case of more sensors than sources prior information about sources is necessary for separation [9]. To recover the sources, the observed signals are processed by a transformation matrix W

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

and the coefficients of matrix W_{ij} are updated based on the learning rule defined from the cost function that imposes the condition of the independence of the recovered sources.

In a reverberant environment, the multi-path wave propagation contributes delayed mixture components to the observations, leading to model

$$x_i(t) = \sum_{j=1}^{M} h_{ij} * s_j(t), \tag{3}$$

where h_{ij} is the channel impulse response from the source j to the sensor i. In typical room environment, the channel impulse response is the room impulse response, which is dependent on reverberation and absorption characteristics of the room. An FIR filter representing typical room impulse response requires a large number of delay elements, typically few thousand significant delay coefficients [10]. Convolutive ICA techniques(*e.g.* [11]), that explicitly assume convolutive or delayed mixtures (3), are usually much more involved than static and require a large number of parameters and long adaptation time horizons for proper convergence, with no real-time microsystems proposed in the literature.

We propose an implementation of the microsystem for blind source separation that interface a miniature microphone array, where the distance between microphones is much smaller than the wavelength, as illustrated in Figure 1. In an anechoic environment, the microphone signals are mixtures of time delayed sources, where the delays between observed signal are proportional to the directional cosines of the impinging

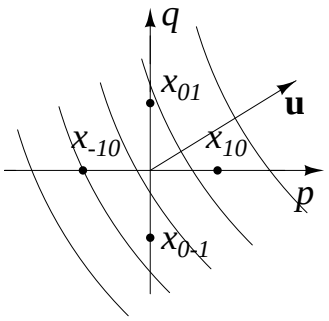

Fig. 1. Configuration of sensors for spatial gradient estimation..

sources [8]. This mixture of the delayed sources is transformed to a linear instantaneous mixture of time-differentiated source signals in the form of (1) by observing the spatial gradients

$$\begin{aligned} \dot{\xi}_{00}(t) &= \textstyle\sum_l \dot{s}^l(t) \\ \xi_{10}(t) &= \textstyle\sum_l \tau_1^l \dot{s}^l(t) \\ \xi_{01}(t) &= \textstyle\sum_l \tau_2^l \dot{s}^l(t) \end{aligned} \tag{4}$$

where $\dot{\xi}_{00}$ is the time derivative of the spatial common mode signal, ξ_{10} and ξ_{01} are the first-order spatial gradients in p and q directions around the origin ($p = q = 0$) and τ_1^l and τ_2^l are the time delays of the source signal l in p and q direction, respectively. The time delays uniquely determine the directions of the source signals. Therefore, by applying the static ICA on the three gradient signals, the source signals can be recovered along with the location of the sources.

In the reverberant environment, in addition to the direct path signals, we have addition of the multi-path copies of the source signals, modifying (4) with contribution of these signals. There are two options for the implementation of the convolutive ICA in this framework, in time domain and in frequency domain. After the extraction of the direct-path signals, in time domain the impulse response of the signals are very similar and the unmixing filters would have significantly reduced number of taps than in the case of the convolutive ICA (3) with the distance between microphones comparable to the signal wavelength. In the frequency domain, we propose subband ICA architecture that comprises static ICA separation applied on the unfiltered spatial gradient signals and static ICA applied separately in each frequency band [12]. 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.

As illustrated in the block diagram, the proposed architecture comprises implementation of the spatial and temporal sensing, the subband decomposition of the spatial gradient signals and application of static linear ICA in 16-channels along with the synthesis of the output signals. The details of the implementation of each of these blocks are outlined in the following subsections.

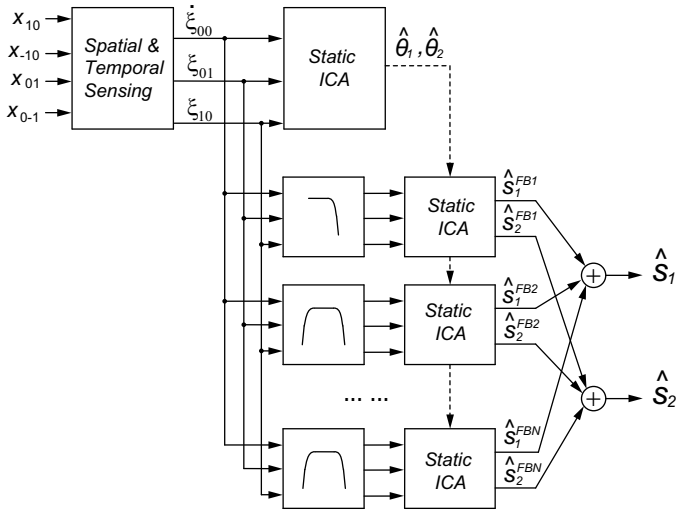

Fig. 2. Block diagram of the proposed subband gradient flow ICA architecture.

A. Spatial and Temporal Sensing

The spatial gradients are approximated on an planar array of four microphones:

$$\begin{aligned} \xi_{00} &\approx \tfrac{1}{4}\big(x_{-1,0} + x_{1,0} + x_{0,-1} + x_{0,1}\big) \\ \xi_{10} &\approx \tfrac{1}{2}\big(x_{1,0} - x_{-1,0}\big) \\ \xi_{01} &\approx \tfrac{1}{2}\big(x_{0,1} - x_{0,-1}\big) \end{aligned} \tag{5}$$

The required operation includes the computation of the average of the four signals with the computation of the time-derivative. The first-order spatial gradients include the computation of the difference between two microphone signals. The computation of the first-order signals is implemented using the difference amplifier.

B. 16-channel Filter Bank

For the subband processing of the spatial gradient signals, we have proposed a 16-channel continuous-time filter bank. The each channel implements a first-order band-pass filter with the factor Q being equal to 4. The center frequencies of the band-pass filters are selected according to the mel-scale, with the linear spacing between 150 Hz and 1 kHz and logarithmic spacing between 1 kHz and 10 kHz. The proposed implementation of the bandpass filter achieved an harmonic distortion of -59.6dB and dynamic range of 59.9dB [13].

C. Independent Component Analysis Implementation

In the proposed subband implementation, the static independent component analysis is performed in 16-channels, with the resulting output signals in each subband combined to recover the source signals. Based on the static ICA solutions on the fullband spatial gradient signals, the coefficients of the unmixing matrix in each subband are initialized. The initialization of the weight coefficients solves the problem of indeterminacy in the ordering of the source signals in the unmixed signals in the frequency domain.

The independent component analysis implementation comprises the vector-matrix multiplication $\mathbf{y} = \mathbf{W}\mathbf{x}$ and adaptation of the unmixing matrix coefficient according to an ICA learning rule. There is a wide variety of ICA learning rules proposed in the literature [9]. For the efficient implementation of the update rule, in order to avoid excessive matrix multiplications and inversions, an outer-product update rule was proposed in [4] with fixed diagonal terms $w_{ii} \equiv 1$, and with off-diagonal terms adapting according to

$$\Delta w_{ij} = -\mu\ f(y_i)g(y_j), \qquad i \neq j \tag{6}$$

This update rule can be seen as the gradient of the information maximization learning rule multiplied by $\mathbf{W}^T$, rather than the natural gradient multiplication factor $\mathbf{W}^T\mathbf{W}$ [9]. To obtain the full natural gradient in outer-product form, it is necessary to include a back-propagation path in the network architecture, and thus additional silicon resources, to implement the vector contribution $\mathbf{y}\mathbf{W}$.

The speech signals are approximately Laplacian distributed, leading to the optimal choice of the nonlinear scalar function $f(y)$ being $\mathrm{sign}(y)$. For efficient implementation, a linear function $g(y) \equiv y$ in the learning rule can be approximated by a 3-level staircase function $(-1, 0, +1)$. This function can be implemented using 2-bit quantization. The quantization of the f and g terms in the update rule (6) simplifies the implementation of the update rule to the function that can be implemented using a single transistor.

In the proposed implementation, the unmixing coefficients are stored differentially as a voltages W_{ij}^+ and W_{ij}^- on two complementary switched current sources [14], with the update rule (6) implemented using two transistor with the functions $f(y)$ and $g(y)$ time encoded, as illustrated in Figure 3. In subthreshold, the current during activation of the sources is exponential in the weights, implementing a coefficient

$$h = h_0(\exp(\kappa\beta w^+) - \exp(\kappa\beta w^-)) \tag{7}$$

where $\beta = \frac{q}{kT}$ and $h_0 = I_0 \exp(-\beta V_s)$. The advantage of this nonlinear transformation is that a wide dynamic range of coefficients is obtained over a limited linear range of voltages W_{ij}^+ and W_{ij}^-. The proposed implementation enables fine updates of the unmixing coefficients with both positive and negative increments. The 3-level staircase function $g(y)$ is approximated with the presence/absence of the voltage pulse and by the relative position of the pulse. The function $f(y)$ is coded as a two-level signal, with the $\mathrm{sign}(y)$ determining the order of the levels V_{lo} and V_{hi}. These voltage levels are applied externally, which control the value of adaptation rate μ. To reduce the required silicon area the C_w is implemented as a MOS capacitance. When the $update$ signals goes high, the charge on the small parasitic capacitance on the drain/source diffusion between transistors M_1 and M_2, denote as C_p, and C_w is shared. The resulting change on the capacitor is given by

$$W_{ij}^+[n+1] = W_{ij}^+[n] + \frac{C_p}{C_w + C_p}(V_{Aij}^+[n] - W_{ij}^+[n]) \tag{8}$$

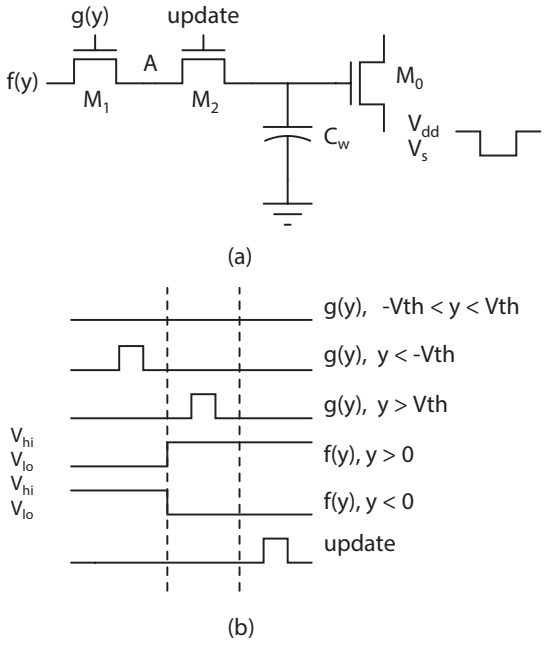

Fig. 3. Adaptation cell: (a) Circuit implementation. (b) Timing diagram.

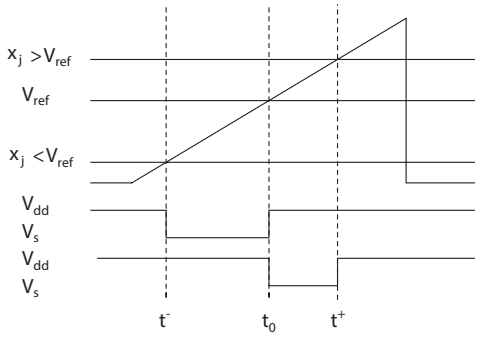

Fig. 4. Pulse-width modulation of the input signals.

The common mode component $\frac{1}{2}(W_{ij}^+ + W_{ij}^-)$ is regulated by the weight decay term on the right side of (8), pulling the values towards the center of the range.

The vector-matrix multiplication $\mathbf{y} = W\mathbf{x}$ is implemented by integrating switched currents controlled by a pulse-width modulation of input signal and gate voltages of a pair of CMOS current sources. The realized multiplication is four-quadrant, with differential weights and bipolar input signal. The source voltage of the transistor, whose gate voltage is controlled by W_{ij} is pulsed, where the width of the pulse is proportional to absolute value of input signal, x_j. The pulse width-modulation of the input signal x_j is illustrated in Figure 4, with the position of the pulse with respect to the reference time point t_0 encoding the polarity of the input signal, relative to reference voltage V_{ref}. Active low source voltage controls the amount of transistor current.

Current pulses are integrated on the output capacitor C_{int}. Due to time encoding of the polarity of the input signal to account for the polarity of the contributions, the voltage

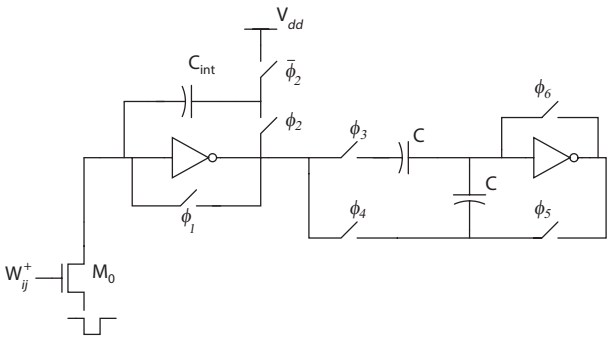

Fig. 5. Integration of switched currents and weighted subtraction.

obtained by the integration before time t_o is subtracted twice from the final integrated value [14]

$$\begin{aligned} y_i^+ - y_i^- &= \sum_{j=1}^{M} \int_{t_0}^{t^+} \frac{I_{ij}}{C_{int}} dt - \sum_{j=1}^{M} \int_{t^-}^{t_0} \frac{I_{ij}}{C_{int}} dt \\ &= \sum_{j=1}^{M} \int_{t^-}^{t^+} \frac{I_{ij}}{C_{int}} dt - 2 \sum_{j=1}^{M} \int_{t^-}^{t_0} \frac{I_{ij}}{C_{int}} dt \quad (9) \end{aligned}$$

The circuit for implementing this weighted subtraction is shown in Figure 5 and is followed by a standard SC subtraction stage.

III. Simulation Results

The spatial gradient sensing, subband signal decomposition and ICA processing were implemented in 0.5μm 3M2P CMOS technology. To demonstrate the feasibility and the performance of the proposed architecture, we have modeled the complete system in Matlab using the simulation results of all three building blocks in Cadence Virtuso. The simulation were performed on the artificial microphone array signals generated based on the image model.Simulated room dimensions correspond to an ordinary office space with the room dimensions 4m×6m and the room height of 2.5m. Both speakers playing the speech sources are kept 1.5m away from the center of the sensor array. The four microphones in the array is orthogonal aligned, with 1 cm inter-microphone spacing as illustrated in Figure 1. Two speech source signals from the TIMIT speech corpus 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 static ICA and subband ICA algorithms is executed under various reverberant environments. Table I summarizes the signal-to-interference ratio (SIR) for room conditions with reverberation time of 200ms and 300ms. Not much differentiation can be observed under mild reverberation situation, because the static ICA already produces very high level of separation. When the condition is more reverberant, the performance advantage of the subband ICA over static ICA starts to exhibit. When the reverberation time is 300ms, which corresponds to a 0.59 uniform reflection coefficient in the simulated room environment, the SIR improvement is over 3dB.

TABLE I
Comparison of Static ICA and Proposed Filterbank ICA results

	RT_{60} = 200ms		RT_{60} = 300ms	
	Static ICA	Subband ICA	Static ICA	Subband ICA
SIR1	25.30dB	25.43dB	6.95dB	10.07dB
SIR2	26.91dB	27.85dB	12.80dB	15.52dB

IV. Conclusion

We have presented an architecture of a microsystem for blind source separation of source signals impinging microphone array with the distance between microphones much smaller than the wavelength. The proposed microsystem enables a wide range of new applications of the miniature microphone arrays where the power consumption and size of the digital implementation were the limiting factors.

V. Acknowledgment

This work was supported by NSF CAREER Award 0846265.

References

[1] R. N. Miles and R. R. Hoy, "The development of a biologically-inspired directional microphone for hearing aids," *Audiology Neuro-Otology,* vol. 11, no. 2, pp. 8694, 2006.

[2] Cohen, M.H., Andreou, A.G. "Analog CMOS Integration and Experimentation with an Autoadaptive Independent Component Analyzer," *IEEE Trans. Circuits and Systems II,* vol 42 (2), pp 65-77, Feb. 1995.

[3] Gharbi, A.B.A., Salam, F.M.A. "Implementation and Test Results of a Chip for the Separation of Mixed Signals," *Proc. Int. Symp. Circuits and Systems (ISCAS'95)*, May 1995.

[4] A. Celik, M. Stanaćević and G. Cauwenberghs, "Mixed-signal real-time adaptive blind source separation," *Proc. IEEE Int. Symp. Circuits Syst.*, pp 760-763, 2004.

[5] K.K. Shyu, M.H. Lee, Y.T. Wu and P.L. Lee, "Implementation of pipelined fastICA on FPGA for real-time blind source separation," *IEEE Trans. on Neural Network*, vol 19 (6), pp 958-970, 2008.

[6] L.-D. Van, D-Y. Wu and C.-S. Chen, "Energy-Efficient FastICA Implementation for Biomedical Signal Separation," *IEEE Trans. on Neural Network*, vol 22 (11), pp 1809-1822, 2011.

[7] M. Stanaćević and G. Cauwenberghs, "Micropower Gradient Flow VLSI Acoustic Localizer", *IEEE Transactions on Circuits and Systems I : Regular Papers*, vol. **52** (10), pp. 2148 2157, 2005.

[8] A. Celik, M. Stanaćević and G. Cauwenberghs, "Gradient Flow Independent Component Analysis in Micropower VLSI", *Adv. Neural Information Processing Systems (NIPS'2005)*, Cambridge: MIT Press, vol. **18**, 2005.

[9] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, New York: John Wiley, 2002.

[10] A. Westner and V.M. Bove, "Blind separation of real world audio signals using overdetermined mixtures", *Proceedings of ICA*, Aussois, France, 1999.

[11] 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)*, Münich, 1997.

[12] S. Li and M. Stanaćević, "Subband Gradient Flow Acoustic Source Separation for Moderate Reverberation Environment," *Conf. Rec. of the 46th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove CA, Nov 2012.

[13] Y. Lin and M. Stanaćević, "A Low-Power, High-Linearity Filter Bank for Auditory Signal Processing Microsystem," *Proc. 56th. IEEE Midwest Symp. on Circuits and Systems (MWSCAS'2013)*, Columbus OH, 2013.

[14] M. Stanaćević and G. Cauwenberghs, "Charge-Based CMOS FIR Adaptive Filter," *Proc. 43rd IEEE Midwest Symp. Circuits and Systems (MWSCAS'2000)*, Lansing MI, August 8-11, 2000.