Stony Brook University

Smart Sensing: Mixed-Signal VLSI Implementation of Gradient Flow Localization and Separation

A Dissertation Presented

by

Shuo Li

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Electrical Engineering

Stony Brook University

May 2015

Stony Brook University

The Graduate School

Shuo Li

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

Milutin Stanaćević – Dissertation Advisor
Associate Professor, Department of Electrical and Computer Engineering

Sangjin Hong – Chairperson of Defense
Professor, Department of Electrical and Computer Engineering

Emre Salman
Assistant Professor, Department of Electrical and Computer Engineering

Pelagia-Irene Gouma
Professor, Department of Material Science and Engineering
Stony Brook University

This dissertation is accepted by the Graduate School.

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

Smart Sensing: Mixed-Signal VLSI Implementation of Gradient Flow Localization and Separation

by

Shuo Li

Doctor of Philosophy

in

Electrical Engineering

Stony Brook University

2015

Acoustic source localization and separation are two crucial components in the acoustic signal processing field, which can be utilized in many applications such as smart hearing aids, tracking and surveillance devices. The small form factor of these applications urge the use of miniature sensor arrays. Nonetheless, the performance of traditional localization and separation techniques deteriorate with small-form-factor sensor arrays and they require higher sampling frequency to compensate for the performance loss, which means higher noise floor and more power consumption.

Gradient flow is a technique for localization of an acoustic source using miniature microphone arrays by relating temporal and spatial gradients of the impinging source signal. Time delays of the impinging sources can be estimated in this fashion along the two axes in the horizontal plane and the incidence angle can be derived. The performance of the gradient flow in noisy and reverberant environ-

ment is examined and quantified through simulations that incorporate additive measurement noise, directional interference signal and room acoustic model. The algorithm demonstrates robust performance for additive signal-to-noise ratio down to 5 dB and at the signal-to-interference ratio of 10 dB. In the echoic room, localization performance remains effective under moderate reverberation conditions. The experimental results from a miniature microphone array recordings in a conference-room environment corroborate the presented simulations.

Then I 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, I 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.

To my dear parents and loving wife...

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost I want to express my deepest gratitude to my advisor Prof. Milutin Stanaćević, for his excellent guidance, caring, patience, and providing me with an excellent atmosphere for doing research. It has been an honor to be his Ph.D. student. I appreciate all his contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating. The joy and enthusiasm he has for her research was contagious and motivational for me, even during tough times in the Ph.D. pursuit.

My thesis committee guided me through all these years. With utmost gratitude I am thankful to Prof. Sangjin Hong, Prof. Emre Salman, and Prof. Perena Gouma for serving as committee members and giving important advice and suggestions on my research. I was my honor taking Prof. Hong and Prof. Salman's courses and learn from their broad knowledge and ways of conducting research. Prof. Gouma gave me plenty of insight to sensor design and it was great pleasure collaborating with her team.

Past and current members of the integrated micro systems lab have greatly contributed to my research. I am very thankful to Xiao Yun, Jinghui Jian, Yingkan Lin and Yasha Karimi for their help and support during the years. It is my honor to have the opportunity to work and discuss my research with

them. Every idea and inspiration means a lot to me and I really appreciate your input.

Special thanks to Anthony Olivo for helping me out on the mechanical and electrical design. We have spent countless hours together to make the prototypes and testing environment work and Tony has never hesitated to help. I am also thankful to Deborah Kloppenburg, Rachel Ingrassia and Susan Hayden for taking care all the administrative issues and make my research life easier.

I would also like to thank Zhichao Chen, Yunlong Wang from Stony Brook University, Ruoxi Yang from Rochester Institute of Technology and Jie Liu from University of South Carolina for their insightful discussions with different aspects of my research. Many thanks to my close friends Yi Ding, Hui Li, Jiawei Zang, Fang Yuan and Zijie Liu for their mental support over the years. I cannot make it without them.

This work is dedicated to my family and my parents from whom I learned the most important values and principles in life. I am also very grateful to all my friends and relatives in China that have always supported me and shared my successes.

Finally, I would like to thank my wife, Chunmei Jia. She was always there cheering me up and stood by me through the good and the bad times. Her love, support and constant patience have taught me so much about sacrifice, discipline and compromise. Finishing my degree is significant to me, and I hope I can repay her by always standing by her side for the rest of my life.

Chapter 1

Introduction

1.1 Background

With the emergence and advance of micro-electro-mechanical systems (MEMS) technology, it is achievable that the acoustic sensors can be made very small, i.e. in millimeter range [5, 6]. Smart sensing hearing aids is one of the areas where nanoscale 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. Alternatively, three-dimension integrated circuit (3D IC) technology can be applied to stack the sensor layer and the signal processing circuitry layer to miniaturize the package size. The advancement and methodology are not just limited to the smart sensing hearing aids, but they can be easily extended to other acoustic applications such as tracking, surveillance and multimedia. Ideally most of the applications will be powered by batteries, so the low power consumption

feature of these applications also comes naturally. Most of the digital very-large-scale integration (VLSI) circuits consume miliwatts of power, which does not present itself as a perfect candidate for the application. On the other hand, analog/mixed-signal circuit, although has its limitation in noise margin and immunity, has its advantage in the ability to accomplish certain calculations and operations while consuming only microwatts of power.

Traditional hearing aids achieve solid performance in amplifying acoustic signals that lend the hearing-impaired people ability to better hear and comprehend speech. However, under the age-old 'cocktail party' acoustic scene where both multiple acoustic sources and environmental noise are presented, the speech intelligibility of traditional hearing aids are very limited. As the noise and irrelevant speech sources are equally amplified as the desired speech sources in these applications, the hearing aids just deliver a mixture of all acoustic observations and do not present any advantage to the hearing-impaired people to listen to only a preferred speaker and neglect other voices and attenuate the background noise. Therefore most traditional hearing aids fail in the acoustic scene of low signal-to-noise ratio (SNR) or with multiple sources. In order for hearing aids to obtain intelligibility, smart sensing is required to suppress the noise source or irrelevant sources based on the spatial location or frequency features. 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 remarkable results in segregating multiple sound sources, even under a very noisy environment. Some modern hearing aids utilize directional or multiple microphones to add some of the functionality of binaural sensing, e.g., the

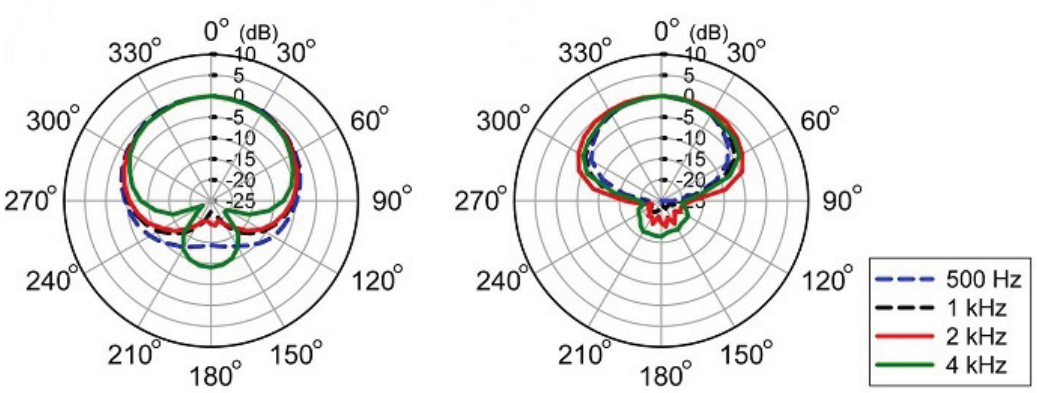

Figure 1.1: Directionality measured from a hearing aid programmed to a fixed directional mode (left) and a directional adaptive mode(right) [1].

Oticon Spatial Sound 2.0 technology. In this case, two microphones allow for one null angle in directionality pattern, and adaptive beamforming allows to steer the null to the noise source. For newer adaptive directional systems that can steer the directivity null to the noise source, the traditional method would result in inaccurate polar patterns and inflated directivity indexes, shown in Figure 1.1. Moreover, their performance still degrades significantly when multiple sources and noise are present. To effectively solve the signal of interest, both localization and separation of multiple acoustic sources using multiple microphones are required.

Moreover, it is critical for hearing aids and other acoustic applications to perform well under real room environments where the echos and reflections create multiple-path disturbances other than the direct-path signal. These delayed and attenuated replicas of the source signal due to reflections from the room boundaries turn the problem of localizing and separating instantaneous mixtures of the original signals to the problem of convolutive acoustic localization and separation. Plenty of studies worked on the problem [7] and many

techniques and algorithms such as the time domain methods [8, 9], frequency domain methods [10, 11] and even subband methods [12–14] are proposed. Yet few of them fit into the efficient implementation with analog/mixed-signal VLSI and smart analog-to-digital processing.

1.2 Inspiration

The human auditory system performs remarkably well in segregating multiple streams of acoustic sources. In relatively large animals, such as owls and humans [15], where the distance between the ears is substantial relative to the wavelength of sound, interaural time, and intensity differences are large enough to be detectable by the central nervous system. The brain constantly uses both ears to orient itself and know what is happening in the environment. It separates relevant sounds from competing noise and it has to know where to focus in noisy environments.

However, for small insects whose hearing organ has much smaller separation, which is in the sub-wavelength range, it could be very difficult for them to localize the acoustic source directly by sensing the interaural time delay. For example, the parasitoid fly has only 0.6mm of hearing organ separation comparing to a 170mm ear separation for a normal human being. Figure 1.2 shows the comparative interaural time delay for parasitoid fly and human being under different source incidence angle. Their interaural time delay has a roughly 1:250 ratio, and it's prohibitive for the parasitoid fly to directly sense the time delay of about 1us.

Experimental biologists observed the phenomenon that parasitoid fly, shown

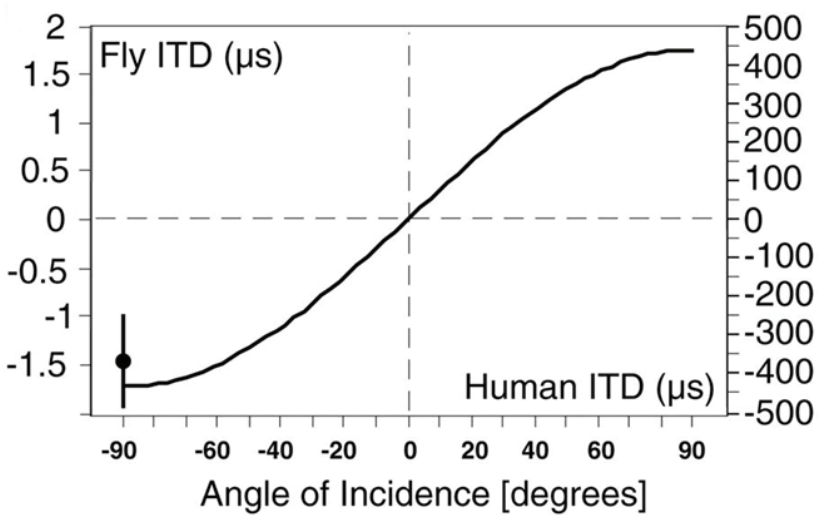

Figure 1.2: A comparison of human ITD and parasitoid fly ITD [2].

in Figure 1.3, and other creatures use tympanal organ which is in a sub-wavelength scale to infer the direction of sound propagation from sensing spatial and temporal gradients of the wave signal for localizing its singing cricket host [2, 3, 16]. The functional principle permitting this directionality may be of particular relevance for technological applications necessitating sensors that are low cost, low weight, and low energy. A small-size, low-noise differential microphone for directional hearing inspired by the structure of the parasitoid fly ears has been proposed, fabricated and tested [17]. The microphone is capable of detecting pressure gradients while diminishing the effect of electronic noise.

With the biological inspiration, I want to take the application of the principle further by combining the localization capability of the above mentioned gradient-sensing microphone sensor and also the more important source separation capability enabled by the directional information that is obtained in the

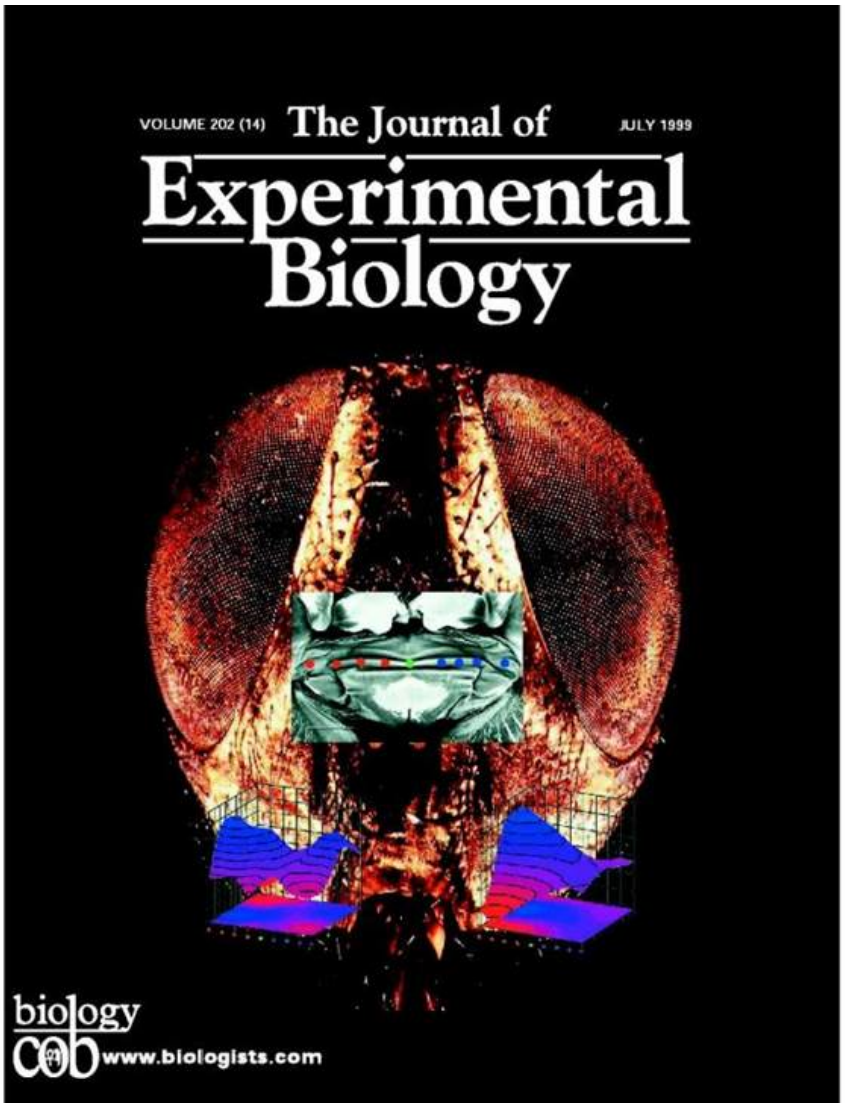

Figure 1.3: Parasitoid fly on the cover of Experimental Biology [3].

localization step. With the source separation, the interference noise sources could be significantly suppressed and thus the performance of the smart hearing aids could be improved to a whole new level.

1.3 Approach

Gradient flow is a signal conditioning technique for sensor arrays of very small separation, which solves interaural time differences (ITD) between signal observations by relating spatial gradients (interaural level differences, i.e., ILD) with the temporal gradients of the time-differentiated signal. Improved differential sensitivity of gradient sensing allows shrinking the aperture of the sensor array without sampling in excess of signal bandwidth, increasing power dissipation and noise bandwidth. Aside from its use in acoustic localization, gradient flow greatly simplifies the the problem of acoustic source separation by establishing the coefficients of the estimated mixing matrix as a combination of time delays from each individual incoming source. The convolutive acoustic source separation is resolved by transforming to a simpler problem of separating corresponding instantaneous mixtures of the time-differentiated signals in multiple subbands.

Gradient flow lends itself into efficient mixed-signal VLSI implementation. Switched-capacitor circuits, with amplifiers biased in subthreshold region for power efficiency, are ideal match for low-noise and low-frequency applications of the designed VLSI systems. Mixed-signal implementation offers best of both worlds: analog signal path avoids the need for high-resolution analog-to-digital conversion and digital adaptation offers flexibility and reconfigurability of the

learning rules, as well as desired outputs in digital format, performing smart analog-to-digital conversion. In the process, the analog/mixed-signal implementation consumes much less power, while still maintaining high linearity and accuracy. Moreover, the silicon size of the mixed-signal circuit is still comparable to the digital signal processing circuit counterpart in the digital domain. This will greatly help the integration process with the sensor silicon in the future.

1.4 Overview

The remainder of the thesis proposal describes the theory, algorithms, design procedures and circuit implementation of real-time blind source separation and localization systems using miniature microphone arrays in detail.

The remaining chapters of the thesis proposal are organized as follows. Chapter 2 describes the theory and design of gradient flow technique with its applications and simulation results in various acoustic environments. Chapter 3 describes the theory and design of subband gradient flow acoustic source separation implementation and its performance under echoic environments. In Chapter 2 and 3, the relationship between localization and separation performance and sensor array dimension is also discussed. Chapter 4 presents the mixed-signal VLSI implementation of the proposed gradient flow acoustic localizer. Chapter 5 presents the mixed-signal VLSI implementation of the proposed independent component analyzer for solving acoustic source separation. Chapter 6 produces the summary and conclusive remarks.

Chapter 2

Gradient Flow Technique and Localization

2.1 Introduction

Sound localization is the process of determining the spatial orientation and location of a sound source based on multiple observations of the emitted sound signal, which normally involves an array of multiple sensing microphones. Modern algorithms used for solving the localization problem include maximum likelihood estimation [18, 19], Gaussian mixture model [20], generalized hard clustering algorithm [21], and other techniques base on probability models [22–28].

Traditionally, sensor arrays with large inter-sensor distance are required to guarantee sufficient spatial separation across sensors to solve localization. It is common wisdom that the accuracy of delay-based source localization degrades with shrinking dimensions of the sensor array. Also, time delay of arrival

(TDOA) estimation techniques based on cross-correlation of the signals require high over-sampling ratios for estimating small time delays. Some researchers also explore the possibility of using inter-aural level difference (ILD) at the same time, but only for compensation of unsatisfactory high frequency performance while a significant separation between sensors is still needed [20, 23]. Moreover, most of the models involving computationally expensive signal processing and statistical techniques and utilizing relatively high-power DSP implementation structures.

Performance of source localization algorithms degrades in the presence of multiple noise sources and is typically limited by room reverberation. To access the localization performance under these adverse conditions is of great interest to a variety of applications in hearing enhancement, multimedia, communication and surveillance. It is required to use microphone arrays with substantial inter-sensor spacings to warrant sufficient spatial separation across sensors and the source localization algorithms are mostly based on estimation of the time delays between source observations [28, 29]. The effect of the reverberation on the performance of these types of algorithms has been studied extensively, with the conclusion that the performance significantly deteriorates above certain level of room reverberation [30, 31]. Ribeiro and etc. [27] has claimed their maximum likelihood algorithm has significantly improved the robustness under reverberant room environment, yet still falls short in some of the acoustic scenes under certain orientation of the source and reverberation time.

However, many applications call for localization of acoustic sources using miniature microphone arrays, where resolving temporal difference poses a sig-

nificant challenge due to the sub-wavelength distance between microphones. Inspired by optical flow for motion estimation in the field of imaging, the direction of sound propagation can be inferred directly from sensing spatial and temporal gradients of the wave signal on a sub-wavelength scale [32, 33]. This principle is also observed in biology, *e.g.,* for localizing prey transmission using differential eardrums and neuronal processing structure or reflecting high-frequency sound, and has already been implemented in biomimetic MEMS systems. Gradient flow technique lends itself into efficient implementation in low-power smart time-to-digital conversion mixed-signal VLSI system, amenable to miniature sensing and processing applications [34]. In this chapter, I study the performance of the gradient flow algorithm in noisy and reverberant conditions using synthetically generated microphone array signals and compare these simulation results with the measured performance in the office-room environment. Furthermore, the effect of sensor array dimensions on the localization performed is discussed.

2.2 Gradient Flow Acoustic Localization

Gradient flow [32, 34] is a signal conditioning technique for source localization designed for arrays of very small aperture, *i.e.,* of dimensions significantly smaller than the shortest wavelength in the sources. Consider a traveling acoustic wave impinging on an array of four microphones, in the configuration of Figure 2.1. The 3-D direction cosines of the traveling wave $\mathbf{u}$ are implied by propagation delays τ_1 and τ_2 in the source along directions p and q in the sensor plane. Direct measurement of these delays is problematic as they require

sampling in excess of the bandwidth of the signal, increasing noise floor and power requirements. However, indirect estimates of the delays are obtained, to first order, by relating spatial and temporal derivatives of the acoustic field.

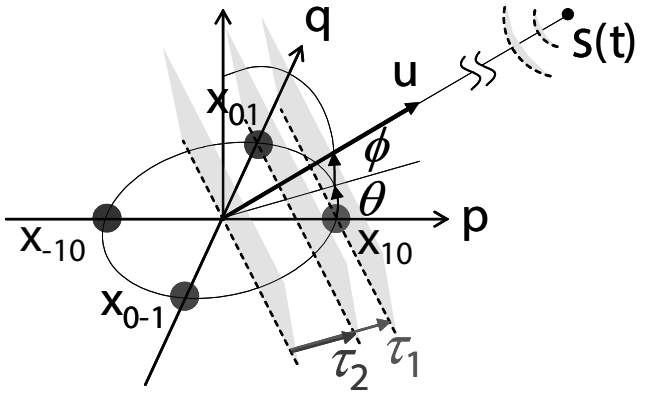

Figure 2.1: Configuration of sensors for spatial gradient estimation.

Since the acoustic source is a far-field source, and the sensor spacing is very small, the signal impinging at the sensor with position coordinates p and q can be expanded about the center of the array in the power series expansion as

$$x_{pq}(t) = s(t) + \tau_{pq}\dot{s}(t) + \frac{1}{2}(\tau_{pq})^2\ddot{s}(t) + ... + n_{pq}(t) \tag{2.1}$$

I concentrate on the first two terms in the series expansion, which is linear in the space coordinates.

$$x_{pq}(t) \approx s(t) + \tau_{pq}\dot{s}(t) + n_{pq}(t) \tag{2.2}$$

Then I take spatial derivatives of various orders i and j along the coordinates p and q, round the origin $p = q = 0$, and apply the above series expansion

model to obtain:

$$\begin{aligned} \xi_{ij}(t) &\equiv \frac{\partial^{i+j}}{\partial^i p \partial^j q} x_{pq}(t) \\ &= (\tau_1)^i (\tau_2)^j \frac{d^{i+j}}{d^{i+j} t} s(t) + \nu_{ij}(t) \end{aligned} \tag{2.3}$$

where ν_{ij} is the corresponding spatial derivatives of the sensor noise n_{pq} around the center. For the first-order case which $i+j=1$, the equation above becomes:

$$\begin{aligned} \xi_{10}(t) &\approx \tau_1 \dot{\xi}_{00}(t) \\ \xi_{01}(t) &\approx \tau_2 \dot{\xi}_{00}(t) \end{aligned} \tag{2.4}$$

where ξ_{10} and ξ_{01} represent spatial gradients in p and q directions around the origin ($p = q = 0$), ξ_{00} the spatial common mode, and $\dot{\xi}_{00}$ its time derivative. Estimates of ξ_{00}, ξ_{10} and ξ_{01} are obtained by finite difference gradient approximation on a grid from the sensor observations $x_{-1,0}$, $x_{1,0}$, $x_{0,-1}$ and $x_{0,1}$ as:

$$\begin{aligned} \xi_{00} &\approx \tfrac{1}{4}\big(x_{-1,0} + x_{1,0} + x_{0,-1} + x_{0,1}\big) \\ \xi_{10} &\approx x_{1,0} - x_{-1,0} \\ \xi_{01} &\approx x_{0,1} - x_{0,-1} \end{aligned} \tag{2.5}$$

The first-order spatial gradients are linearly scaled versions of temporal gradients of the signal measured at the center of the array. If I have a number of samples on the sensor signals thus the temporal and spatial gradients, the estimation of the time delays along the two axes are solved as the standard

least-square problem. The bearing estimates of τ_1 and τ_2 are

$$\hat{\tau}_1 = \frac{E[\xi_{10}\dot{\xi}_{00}]}{E[\dot{\xi}_{00}^2]} \tag{2.6}$$

$$\hat{\tau}_2 = \frac{E[\xi_{01}\dot{\xi}_{00}]}{E[\dot{\xi}_{00}^2]}. \tag{2.7}$$

These time delays τ_1 and τ_2 represent the direction cosines (or angle coordinates) of the source relative to the array, scaled by the speed of wave propagation and the unit dimensions of the array. The direction cosines can be decomposed in azimuth angle θ and elevation angle ϕ as indicated in Figure 2.1:

$$\begin{aligned} \tau_1 &= \frac{d}{c}|\mathbf{r}_1| \ \cos\theta\cos\phi \\ \tau_2 &= \frac{d}{c}|\mathbf{r}_2| \ \sin\theta\cos\phi, \end{aligned} \tag{2.8}$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are unit orthogonal vectors along p and q directions. From least-square estimates of the time delays, I can directly obtain estimates of azimuth angle θ and elevation angle ϕ according to (2.8). The estimate of azimuth angle can be simply calculated as the ratio of delay estimates $\hat{\tau}_1$ and $\hat{\tau}_2$:

$$\hat{\theta} = \arctan\frac{\hat{\tau}_2}{\hat{\tau}_1}, \tag{2.9}$$

and the estimate of elevation angle is

$$\hat{\phi} = \arccos\sqrt{\frac{c^2(\hat{\tau}_1^2 + \hat{\tau}_2^2)}{d^2(|\mathbf{r}_1|^2 + |\mathbf{r}_2|^2)}}. \tag{2.10}$$

2.3 Testing Data Generation

It is desirable to apply the gradient flow localization algorithm to artificial data first to evaluate the performance and then use real recorded data in an office room environment and check if the algorithm still performs robustly. This section discusses the process of producing these two types of testing data.

2.3.1 Artificial Data Generation

First, the direct-path signal at the microphone array is generated considered there is no reverberation involved. The sound pressure at the center of the microphone array emitted from a single frequency point source in free space can be modeled as:

$$P(\omega, R, t) = \frac{exp[i\omega(R/c) - t]}{4\pi R}, \tag{2.11}$$

which ω is the angular frequency, R is the distance between the source and the sensor, t is the time, and c is the speed of sound. The sound pressure is attenuated proportionally to the distance it travels within a certain media. It is straightforward to get the response from the attenuation and the time delay based on the relative location of the source and the sensor.

Next, the response at the sensor array in a reverberant environment is calculated. A virtual rectangular cuboid office room which has six surfaces with uniform reflection coefficient is modeled. The image method [35] is used to generate the imaginary sources, with their location, the incidence angle on the microphone array, and the number of reflections on each of the surfaces.

The reflected sound is propagated from the direction of the sound source of the mirror image room, and the image of the source in the mirror room is referred to as a virtual source. Figure 2.2 shows the illustration of first order virtual source with a single reflection of a surface and also the second order virtual sources. The higher order virtual sources can be deduced in the same manner. If I get the location and number of reflections of the virtual sources and the reflection coefficient, I can get an impulse response of the signal source in an echoic environment. The reflection coefficient is assumed to be frequency-independent and the same for all walls, ceiling and floor. For a given value of reverberation time, T_{60}, the reflection coefficient is computed using Eyring's formula [36]:

$$r = \exp(-\frac{13.82}{(\frac{1}{\mathrm{L_x}} + \frac{1}{\mathrm{L_y}} + \frac{1}{\mathrm{L_z}})\mathrm{cT_{60}}}). \tag{2.12}$$

In my experiment, corresponding room impulse responses is generated based on the calculated reflection coefficient according to the room dimensions and reverberation time. The signal received at the microphone x_{10} is represented as

$$x_{10}(t) = \sum_{i=1}^{N} a_i s(t - \tau^i) + n_{10}(t). \tag{2.13}$$

where a_i and τ^i are attenuation and time delay associated with the virtual source i and N is the number of virtual sources. I assume that the source location is fixed, so that the model is time-invariant.

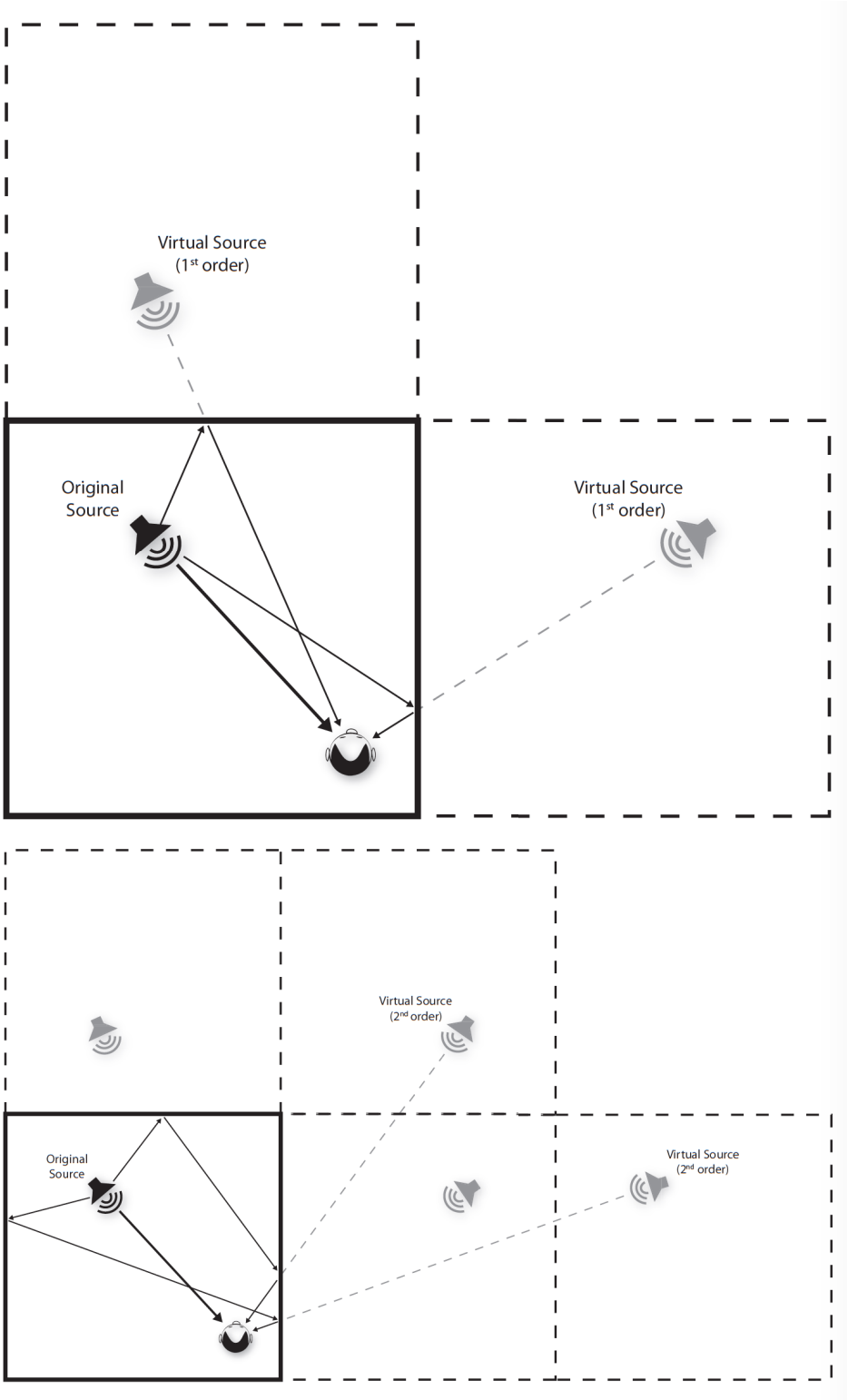

Figure 2.2: Illustration of the image method [4].

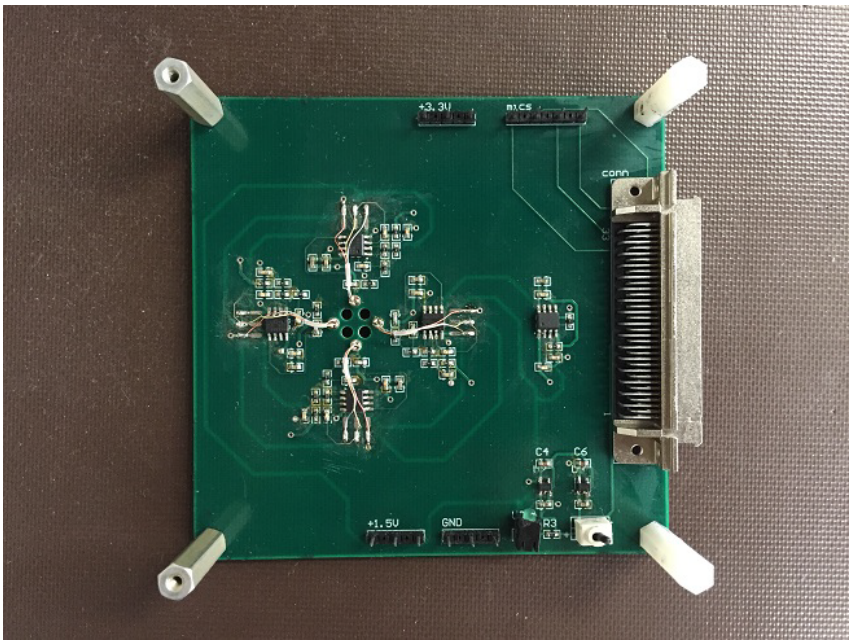

Figure 2.3: Photo of the custom designed PCB for collecting recorded data.

2.3.2 Recorded Data Collection

To collect real recorded data within a conference room environment, a custom designed PCB is fabricated to interface the sensor array with the National Instruments data acquisition card to acquire the acoustic signals of each sensory channel. The analog signals picked up by the microphones go through a bandpass prefilter and preamplifier, which has cutoff frequencies of 100Hz and 4kHz and a gain of 20dB, to filter out the frequency components that is not of interest in the human speech and amplify the rather weak microphone signals. The analog signals are sampled at 16kHz, and multiple samples in 10 seconds length are collected, which can be truncated into shorter length according to requirements in the localization simulations in the next section. The photo of the PCB is shown in Figure 2.3.

2.4 Simulated Localization Performance

The performance of gradient flow algorithm for source localization is quantified in simulated adverse acoustic conditions. I performed simulations with artificially generated microphone array signals with scaled additive measurement noise and different levels of background noise with different incidence angles of the source signal. The simulated room experiments were performed to determine the degrading effect of reverberation in the real room environment.

2.4.1 Localization Performance with Additive Measurement Noise and Directional Noise Source

I first investigate the effect of the additive measurement noise, that can originate from the sensor itself or the readout electronics. The assumption is that the mutually independent white Gaussian noise sources are scaled and added to each microphone signal to control the signal-to-noise ratio (SNR). The performance of the localization algorithm was analyzed given the results of a series of Monte Carlo simulation experiments with artificially generated source signals and additive noise. Two different types of the source signals is used, which one signal is a bandlimited (100Hz-4kHz) white Gaussian source signal, while the second source is the speech signal from the TIMIT database. The sampling frequency is 16 kHz, while the length of the signal is 1 s. The distance between microphones is set to 1 cm. The signal-to-noise ratio at the received sensor signals is varied from -5dB to 15dB for the incidence angle of the source signal of 0^o, 15^o, 30^o and 45^o. The choice of angles is justified by the symmetry of the four element microphone array. The Figure 2.4 shows the

standard deviation of localization error for different angles for band-limited white Gaussian random source, while the Figure 2.5 shows the standard deviation error for a speech source signal.

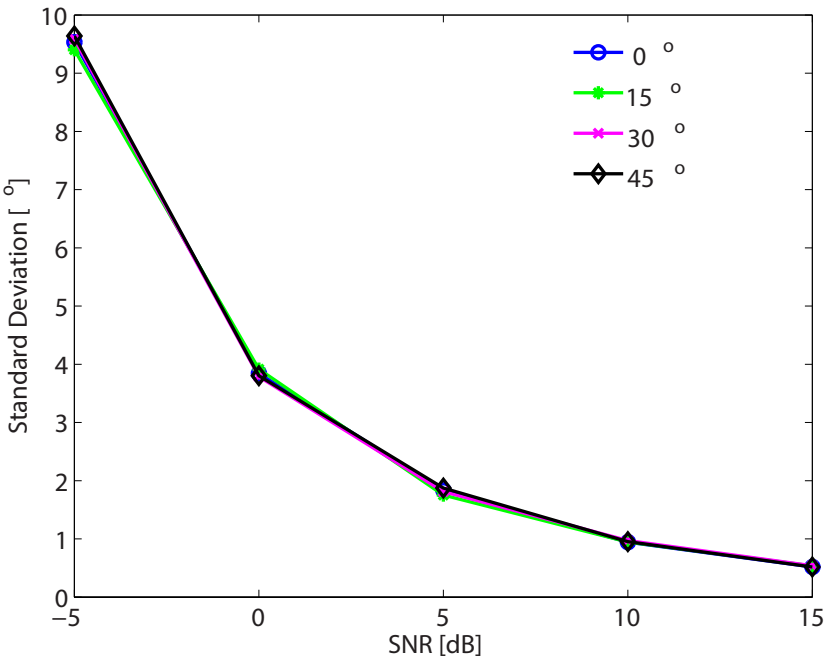

Figure 2.4: Standard deviation error of the estimated incidence angle under different SNR conditions and at different angular location of the source for bandlimited white Gaussian signal.

Despite the low aperture, the gradient flow demonstrates a robust performance with respect to acquisition noise. The performance is maintained to level of SNR as low as 5 dB, while it deteriorates significantly when the SNR drops down to 0 dB. I can also observe that the localization performance is not influenced by the incidence angle.

In the next set of simulations, I investigate the effect of a directional noise sound source on the localization performance. The source signal is a speech signal chosen from TIMIT database, while the interference source is a bandlim-

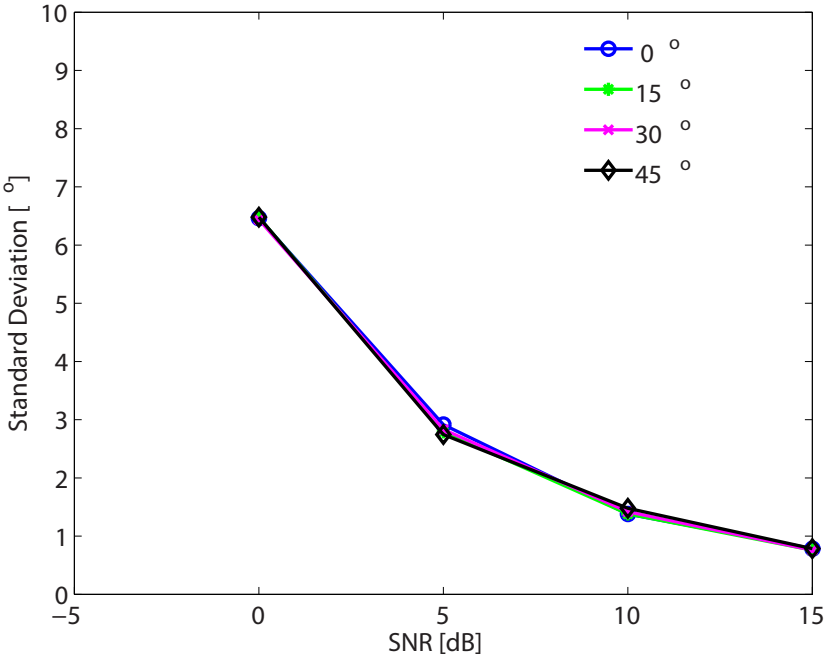

Figure 2.5: Standard deviation error of the estimated incidence angle under different SNR conditions and at different angular location of the source for speech signal.

ited white Gaussian signal. The source signal is located at incidence angle of 45^o, while the interferer is moved from the angle of 30^o to the angle of -45^o in step of 15^o, thus varying the angular distance from 15^o of separation to 90^o. The signal-to-interference ratio (SIR) is set at 10dB and 20dB, with the additive measurement noise of 20dB. For the SIR of 20dB the influence of the interferer on the localization performance is minimal, while at 10dB of SIR the influence of the interferer dominates the standard deviation of the error.

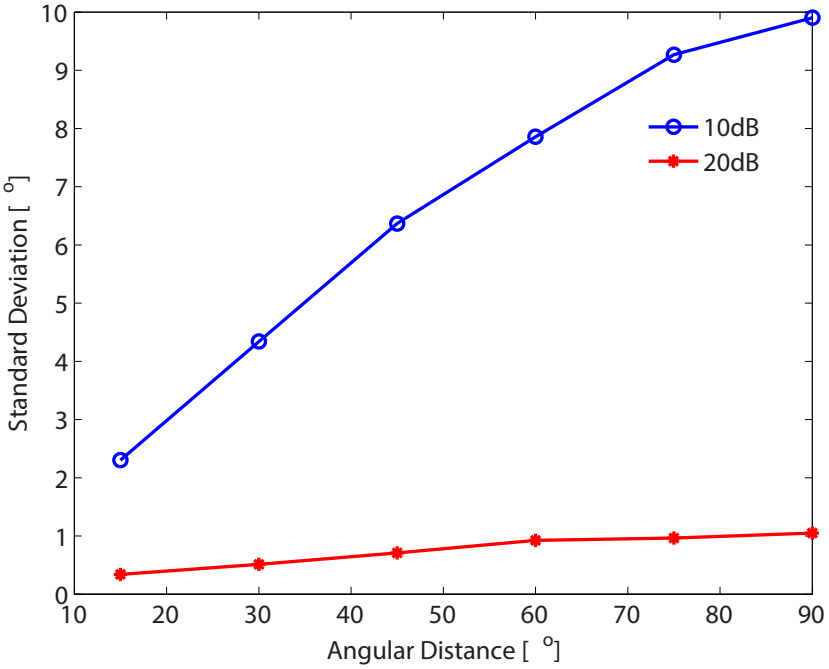

Figure 2.6: Standard deviation error of the estimated incidence angle if an interfering source is present in the environment.

2.4.2 Localization Performance under Reverberant Conditions

I want to understand how room reverberation affects the localization performance of the gradient flow technique. In the room environment, in addition to the direct-path signal received by the microphone array, the observed signals comprise multiple delayed and attenuated replicas of the source signal due to reflections from the room boundaries as walls, ceiling, and floor. The reflections also originate from different objects present in the room. The observer perceives the echo as radiating from a point past the wall from which it was reflected.

To quantify the localization performance under different reverberant con-

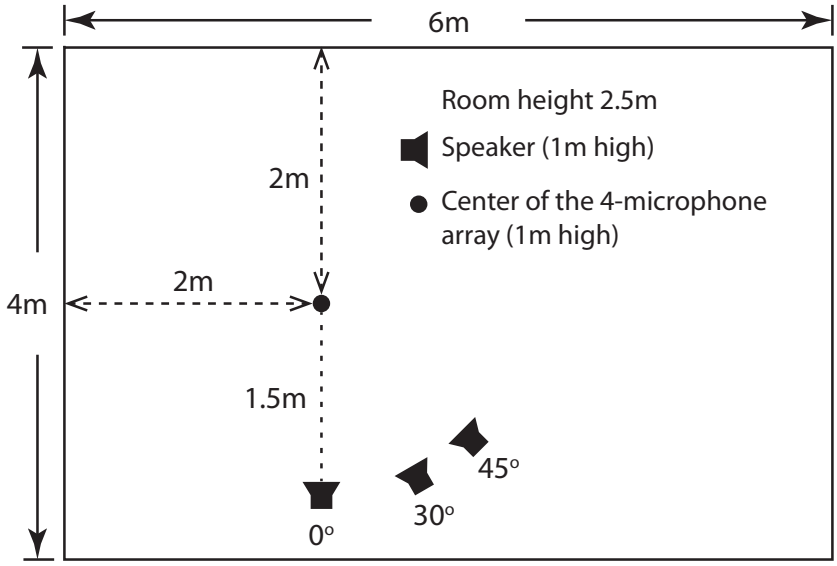

Figure 2.7: Reverberant room configuration.

ditions, I construct an artificial room using virtual source mapping model to generate observations at the microphone array in an echoic room environment. The dimensions of the room, with the location of the microphone array and the source signal, are illustrated in Figure 2.7. The distance between the array and source was kept at 1.5 m, while I investigate three different incidence source angles, 0^o, 30^o and 45^o. The additive measurement noise was mutually independent, white, Gaussian random signal with power varied so that the signal-to-noise ratio was varied from 0dB to 20dB.

Figure 2.8 and Figure 2.9 show the dependance of accuracy in the estimated angle of incidence, as the percentage of the total number of absolute errors smaller than 5^o and standard deviation, as a function of the reverberation time for different incidence angles and different level of additive measurement noise. I can conclude that the gradient flow demonstrates robust performance under mild reverberation. Under moderate reverberation, the gradient flow algorithm

performs well if the additive noise is on the order of 10 dB. As expected, the impact of the reverberation is strongest for the incidence angle of 45^o, which has the least amount of symmetry in this artificial room configuration.

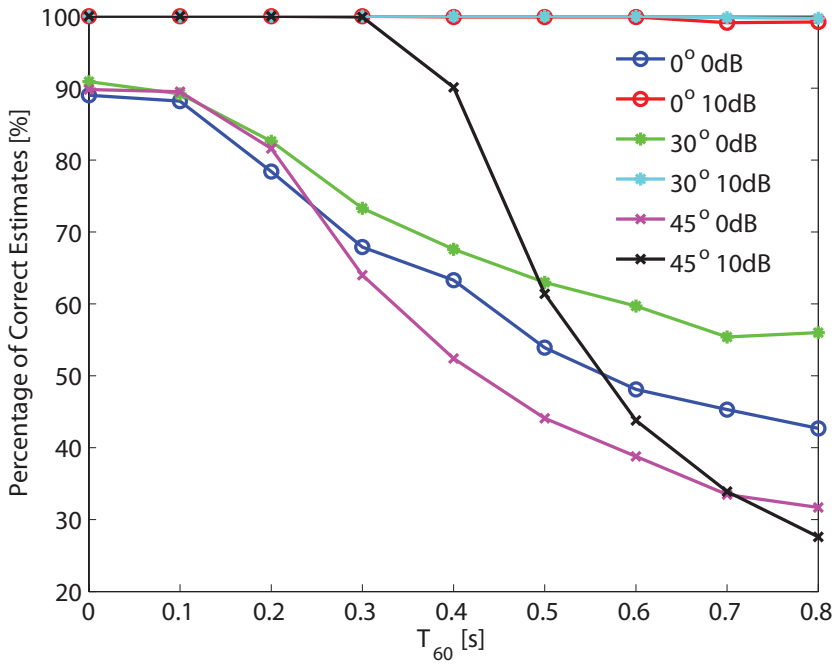

Figure 2.8: Percentage of correctness in the estimated angle of incidence as a function of the reverberation time for different incidence angle and different level of additive measurement noise.

2.5 Experimental Results

To verify the simulation results and to demonstrate the performance of the gradient flow algorithm, the planar array of four omnidirectional hearing aid miniature microphones was used for localization experiments in the room environment. A single acoustic source was presented through a loudspeaker positioned at 1 m distance from the array in a conference room with reverber-

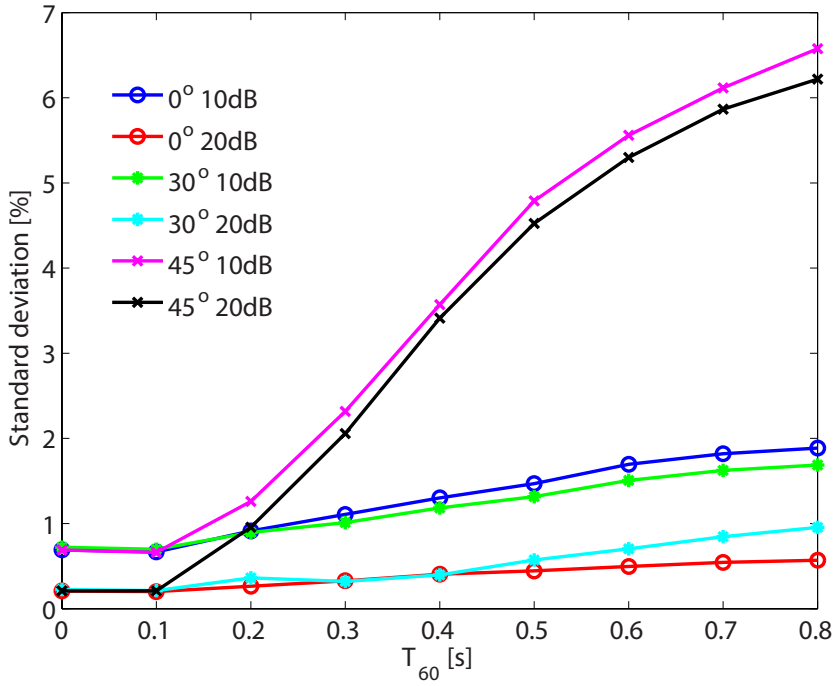

Figure 2.9: Standard deviation of the estimated angle of incidence as a function of the reverberation time for different incidence angle and different level of additive measurement noise.

ation time of 180 ms, while the microphone array was mounted on a rotating platform. The localization experiment was repeated for two different source signals, a broadband bandlimited (100-1000Hz) Gaussian signal and speech signal. The corresponding SNR was around 35 dB. The 10 estimates of the incidence angle were obtained for 1 s long source signals. The mean and variance of estimated incidence angles for Gaussian and speech signal are shown in Figure 2.10 and Figure 2.11. The gradient flow algorithm shows robust performance in real moderate reverberant room environment.

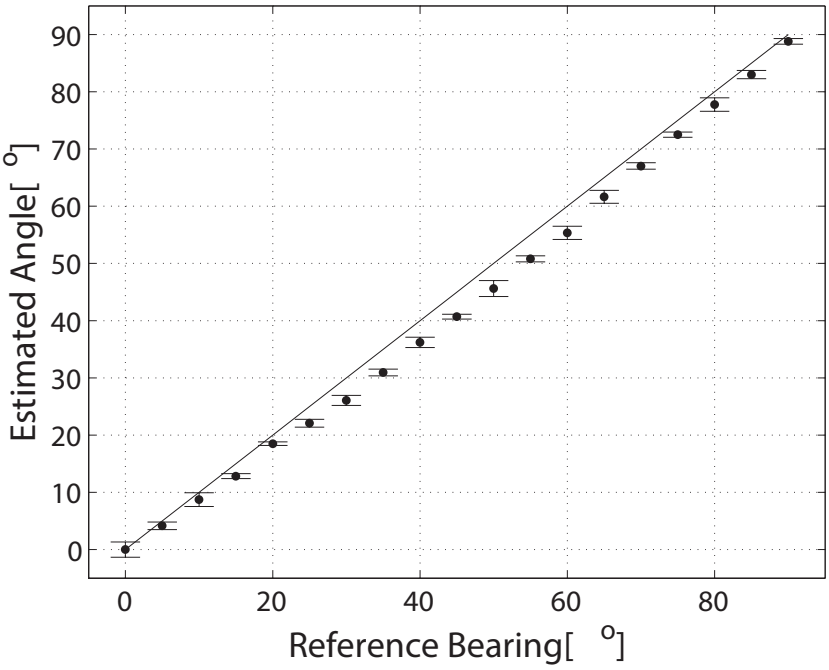

Figure 2.10: Localization of bandlimited (100-1000Hz) Gaussian source in room environment.

2.6 Study on Sensor Array Aperture

It is commonly recognized that sensor arrays with large inter-sensor spacing should be used for source separation and beamforming to warrant sufficient spatial difference across sensors to resolve time delays between source observations. When the dimensions of the sensor array shrink, traditional techniques require sampling in excess of the bandwidth of the signal, increasing noise floor and power requirements. These issues pose a critical barrier to make implementation such as smart hearing aids smaller in size and operating with low power.

Current works on acoustic localization with multiple-microphone arrays generally use arrays with inter-microphone spacing of at least 10cm [18, 28].

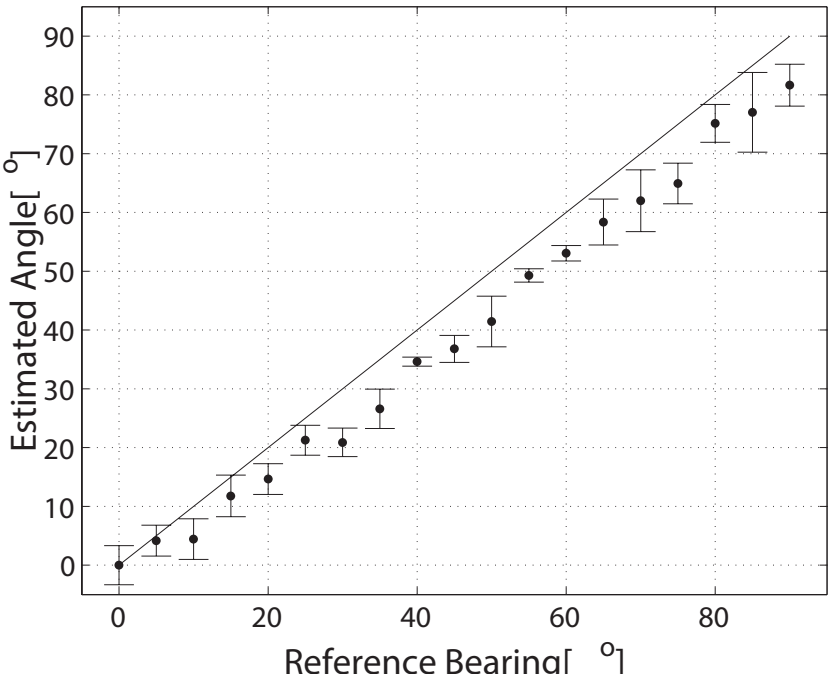

Figure 2.11: Localization of speech source signal in room environment.

Considering most of the arrays are linear arrays, the overall size of the arrays are not ideal candidates for small-form-factor and low-power applications. Gradient flow is a signal conditioning technique designed for arrays of very small aperture [32, 33]. The direction of sound propagation can be inferred directly from sensing spatial and temporal gradients of the wave signal on a sub-wavelength scale. Previous study exhibits very favorable localization performance with orthogonal four-microphone array of 1cm inter-sensor spacing under noisy and echoic room environments [37].

Nonetheless, the differential signals in the microphone observations are based on a large common-mode pedestal and are subject to the interference of noise originated from the sensor or the readout electronics. Especially in processing low-frequency signals when the wavelength is in the range of several

meters, a microphone array 1cm in diameter could produce very little differential signal because of the small spatial diversity across the sensors. Thus I need to look into the possibility of increasing the sensor dimension to help the localization performance in the low-frequency signals and improve the overall performance on the broadband signal.

2.6.1 Single Frequency Analysis

First, theoretical derivations are deduced to lay a foundation for better understanding of the effect sensor array dimension has on the localization performance when the incoming wave is a single frequency sine wave. Here, I will go back to the array configuration shown in Figure 2.1. Consider a single frequency sine wave impinging on the array. If the signal at the center of the array is assumed to be $x(t) = \sin(2\pi ft)$, which f will be the frequency of the sine wave, then the signal at the four sensors will be a delayed version of the sine wave. I assume the amplitude of the signal to be unity and there is no phase component for easier derivation and illustration. The assumption will not affect the results. The common mode signal can be decomposed as:

$$\begin{aligned} \xi_{00} &= \tfrac{1}{4}\big(x_{-1,0} + x_{1,0} + x_{0,-1} + x_{0,1}\big) \\ &= \tfrac{1}{4}[\sin(2\pi f(t - \tau_1/2)) + \sin(2\pi f(t + \tau_1/2)) \\ &\quad + \sin(2\pi f(t - \tau_2/2)) + \sin(2\pi f(t + \tau_2/2))]. \end{aligned} \tag{2.14}$$

The temporal derivative can be derived according to the trigonometric angle sum and difference identities as:

$$\begin{aligned}\dot{\xi}_{00} &= \tfrac{1}{4}(2\pi f)[\cos(2\pi f(t-\tau_1/2)) + \cos(2\pi f(t-\tau_1/2)) \\ &\quad + \cos(2\pi f(t-\tau_2/2)) + \sin(2\pi f(t+\tau_2/2))] \\ &= \pi f \cos(2\pi f t)[\cos(2\pi f \tau_1/2) + \cos(2\pi f \tau_2/2)]. \end{aligned} \tag{2.15}$$

The spatial derivative ξ_{10} can be derived as:

$$\begin{aligned}\xi_{10} &= x_{1,0} - x_{-1,0} \\ &= \sin(2\pi f(t+\tau_1/2)) - \sin(2\pi f(t-\tau_1/2)) \\ &= 2\cos(2\pi f t)\sin(2\pi f \tau_1/2). \end{aligned} \tag{2.16}$$

Similarly, spatial derivative along the other axis ξ_{01} will be:

$$\begin{aligned}\xi_{01} &= x_{0,1} - x_{0,-1} \\ &= \sin(2\pi f(t+\tau_2/2)) - \sin(2\pi f(t-\tau_2/2)) \\ &= 2\cos(2\pi f t)\sin(2\pi f \tau_2/2). \end{aligned} \tag{2.17}$$

Assume the incoming signal is on the same horizontal plane as the sensor array for simplicity, $\phi = 0$, the azimuth angle θ that is estimated using the gradient

flow algorithm will be:

$$\begin{aligned}\hat{\theta} &= \arctan\frac{E[\xi_{01}\dot{\xi}_{00}]}{E[\xi_{10}\dot{\xi}_{00}]} \\ &= \arctan\frac{2\sin(2\pi f\tau_2/2)E[\pi f\cos^2(2\pi ft)]}{2\sin(2\pi f\tau_1/2)E[\pi f\cos^2(2\pi ft)]} \\ &= \arctan\frac{\sin(\pi f\frac{d}{c}\sin\theta)}{\sin(\pi f\frac{d}{c}cos\theta)}. \end{aligned} \tag{2.18}$$

It could be observed from 2.18 that if both $\pi f\frac{d}{c}sin\theta$ and $\pi f\frac{d}{c}cos\theta$ approaches 0, then $\hat{\theta}$ approaches θ. The error will be dependent on the frequency f and the incidence angle θ of the incoming signal, and the spacing between the sensors d. Some following factor analysis will show the relationship between the error and the factors.

Similar results could also be obtained from frequency domain analysis. Consider the time domain spatial derivative

$$\xi_{10}(t) = x(t+\tau_1/2) - x(t-\tau_1/2). \tag{2.19}$$

The fourier transform of the derivative in frequency domain is

$$\begin{aligned}\Xi_{10}(\omega) &= X(\omega)e^{j\omega\tau_1/2} - X(\omega)e^{-j\omega\tau_1/2} \\ &= X(\omega)2j\sin(\omega\tau_1/2). \end{aligned} \tag{2.20}$$

Since the fourier transform of the temporal derivative is

$$\dot{\Xi}_{00}(\omega) = j\omega X(\omega)[\cos(\omega\tau_1/2) + \cos(\omega\tau_2/2)], \tag{2.21}$$

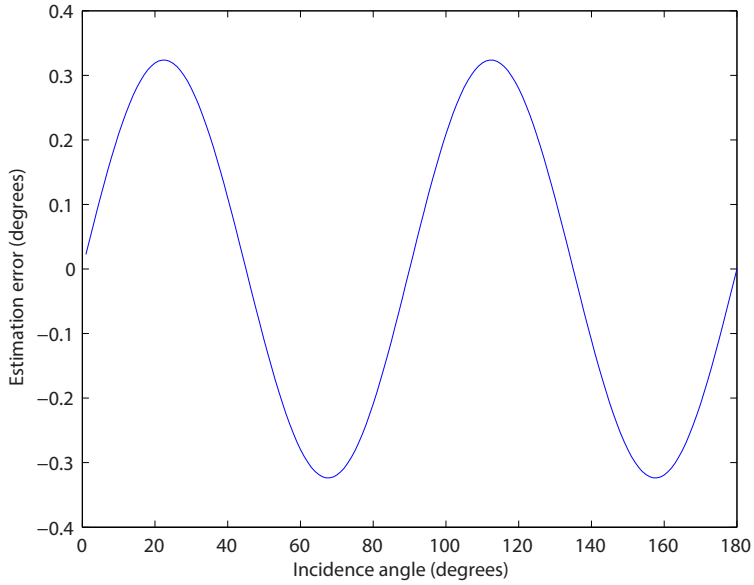

Figure 2.12: Estimation error against incidence angle for 4kHz signal and 1cm array.

the spatial derivative becomes

$$\Xi_{10}(\omega) = \dot{\Xi}_{00} \frac{\sin(\omega\tau_1/2)}{\omega/2}. \tag{2.22}$$

Similarly,

$$\Xi_{01}(\omega) = \dot{\Xi}_{00} \frac{\sin(\omega\tau_2/2)}{\omega/2}, \tag{2.23}$$

which leads us to the same result as in Equation 2.18.

Figure 2.12 shows the dependence of estimation error on the different incidence angle under 4kHz signal and a 1cm sensor array. It is observed that the worst case scenarios are when the incidence angle is at 22.5°, 67.5°, etc., while the estimation error is 0 when incidence angle is at 0°, 45° and 90°. This is very intuitive trend according to the partial derivative of 2.18 with respect to

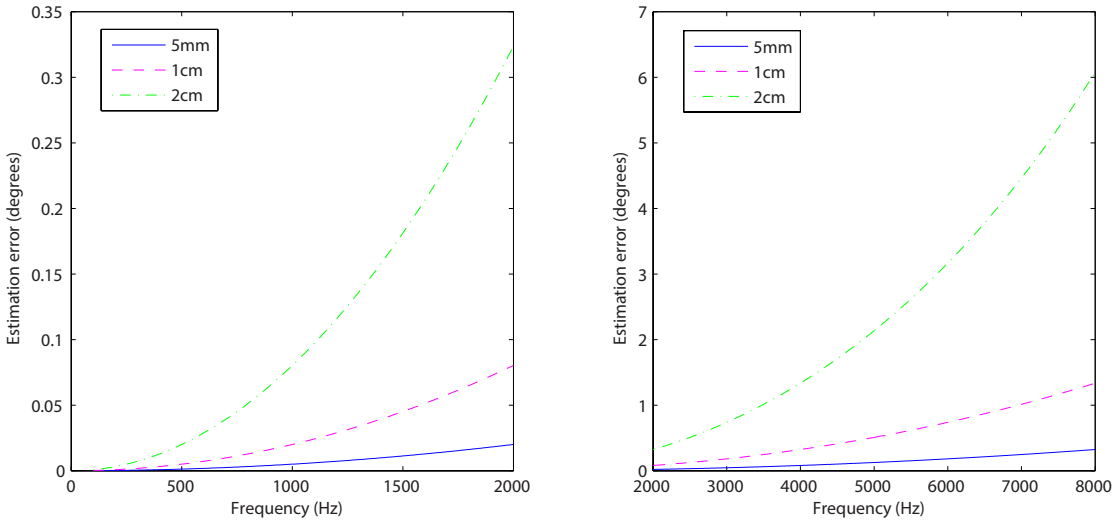

Figure 2.13: Estimation error against input frequency for 5mm, 1cm and 2cm array.

θ.

Figure 2.13 shows the dependence of estimation error on the input sinusoidal signal frequency under 5mm, 1cm and 2cm sensor array respectively. In this analysis, the worst case incidence angle 22.5° is selected to illustrate the tendency more clearly. It is observed that for frequencies under 2kHz, the difference in the estimation error is not obvious since it is very small in all three cases. Yet when the frequency increases, the difference starts to display, and the 5mm array compensates the estimation very well and produces the least amount of estimation error. The result corresponds to 2.18 accurately. Since $\pi f \frac{d}{c} sin\theta$ and $\pi f \frac{d}{c} cos\theta$ have to be small to ensure the estimation accuracy, when frequency rises thus the wavelength shrinks, the sensor array has to be made smaller to keep the gradient flow localization algorithm robust in anechoic and noiseless environment.

2.6.2 Simulated Localization Performance Comparison

All the previous studies were conducted with an orthogonal microphone array that has the inter-sensor spacing of 1cm. In this study I evaluate the localization performance difference between microphone arrays of 1cm and 2cm in diameter when the incoming acoustic signals are in low frequency. I performed simulations with artificially generated microphone array signals with scaled additive measurement noise and simulated room environments with various reverberation times.

I investigate the performance difference with simulated sensor observations in an artificially generated room impulse response. The dimensions of the room, with the location of the microphone array and the source signal, are illustrated in Figure 2.7. The distance between the array and source was kept at 1.5m. Mutually independent white Gaussian additive noise are scaled and added to each microphone observation to control the SNR. The performance of the localization algorithm was analyzed over the results of a series of Monte Carlo simulation experiments with various additive noise of the same SNR. I use a speech signal from the TIMIT database as the source signal. The sampling frequency is 16 kHz, while the length of the signal is 1 s.

First I am attempting to differentiate the effect of noise and reverberation on the localization results. Three sets of simulations were conducted when there is only additive noise presented, only reverberation presented, and under both noisy and reverberant conditions. The incidence angle used in the shown results is 10^o, yet the result is constant over different incidence angles in our test.

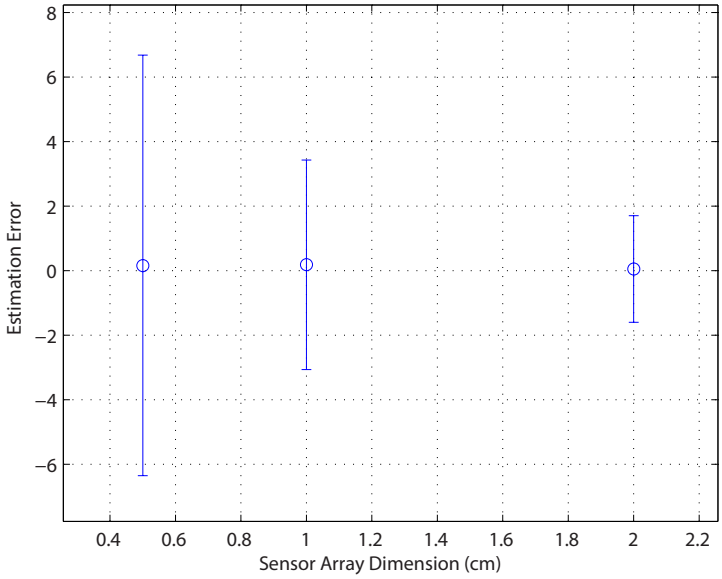

Figure 2.14: Error plot of different sensor dimensions when there is only noise (SNR=30dB) for low frequency source.

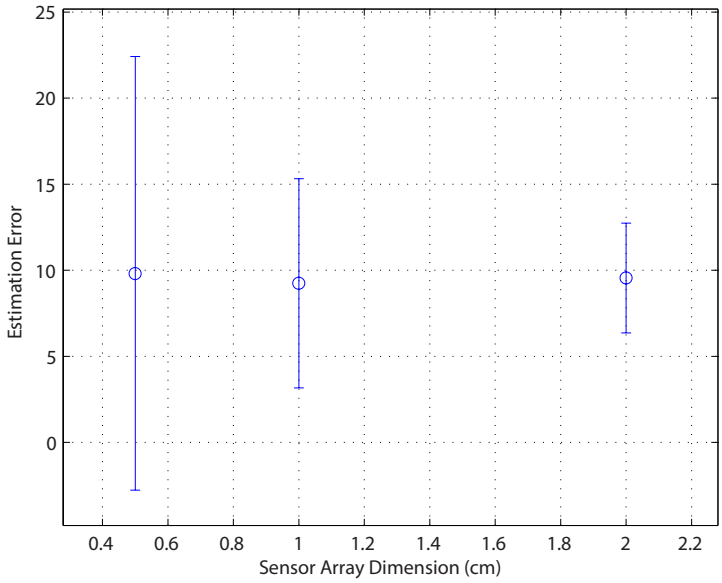

Figure 2.15: Error plot of different sensor dimensions when both noise (SNR=30dB) and reverberation (Rt60=300ms) for low frequency source.

I used 100Hz to 400Hz low frequency bandlimited speech signals in the first set of three graphs. Figure 2.14 shows the error plot of different sensor dimensions when there is only a noise of 30dB SNR. The standard deviation decreases when the sensor array dimension increases, showing that the bigger spatial diversity can produce larger derivatives so the results can be more immune to the effect of additive noise. In the mean time, the mean error stays at the same level, corresponding to the trend illustrated in the theoretical analysis that the difference in mean error is very small in low frequency. Figure 2.15 shows the error plot of different sensor dimensions under an environment when there is both reverberation and noise presented. The mean error does not change much under different inter-sensor spacings. The result can be expected that the derivatives are accordingly amplified for both direct path signals and inferences as spacing increases, so the ratio of signal over interference is not changed much that the performance stays similar. It also corroborates our previous theoretical analysis that in low frequency, the performance difference between various sensor array dimensions is negligible. As for the standard deviation, array with larger spacing still performs better, which is the same as the case with just white noise added. The 2cm microphone array still shows an advantage because of its superior ability to handle noisy situations.

For the second set of three graphs, I used 4kHz to 6kHz high frequency bandlimited speech signals as input. Figure 2.16 shows the error plot of different sensor dimensions when there is only a noise of 30dB SNR. The mean error increases when the sensor array dimension increases, corroborating our theoretical findings in the high frequency range. The standard deviation de-

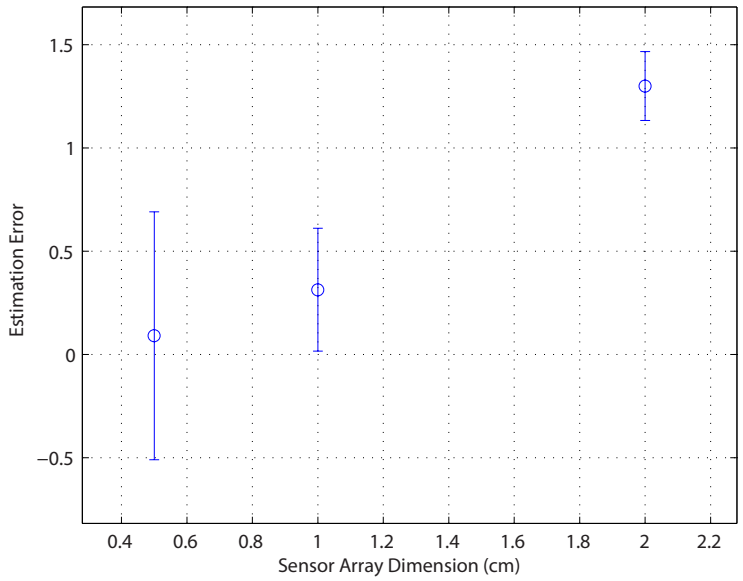

Figure 2.16: Error plot of different sensor dimensions when there is only noise (SNR=30dB) for high frequency source.

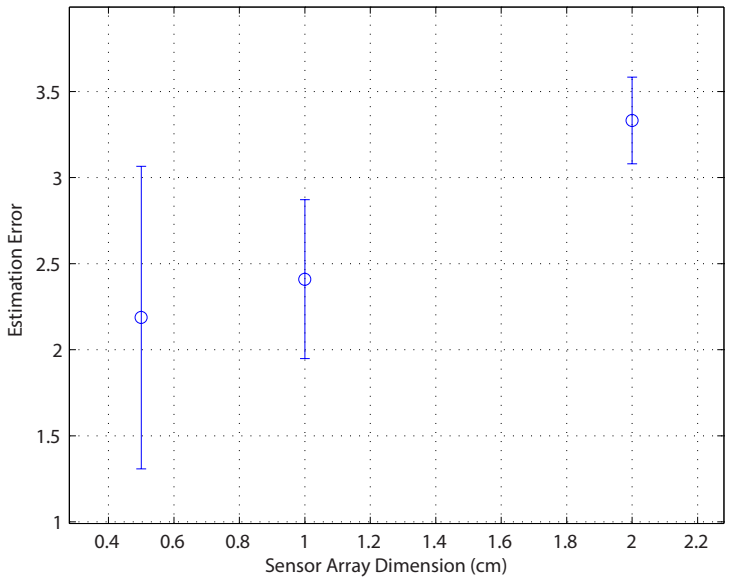

Figure 2.17: Error plot of different sensor dimensions when both noise (SNR=30dB) and reverberation (Rt60=300ms) for high frequency source.

creases when the sensor array dimension increases, which is the same as the low frequency case. Figure 2.17 shows the error plot of different sensor dimensions under an environment when there is both reverberation and noise presented. A trend predicted by the theoretical analysis was observed that the mean error increases when the sensor array dimension increases. The 5mm microphone presents smaller mean error, while the 2cm microphone array shows an advantage because of its superior ability to handle noisy situations.

Then I choose two most representative incidence angles 0^o and 45^o to investigate the change of localization performance based on different sensor dimension, reverberation time, and SNR level. The source used is 100Hz to 400Hz bandlimited speech signal. Figure 2.18 and Figure 2.19 show the comparison of the standard deviation in the estimated angle of incidence as a function of the reverberation time for 0^o and 45^o incidence angles and different level of additive measurement noise with a speech signal.

From the comparisons I could conclude that the 2cm sensor array dimension compensates the lack of spatial diversity in 1cm sensor array in a moderate degree. Under all the conditions with different incidence angle, reverberation time and SNR combinations, the 2cm sensor array cuts the standard deviation error almost by half in most cases. Nonetheless, in the 45^o case, when reverberation is high, the 2cm array does not exhibit much advantage probably due to its lack of capability to handle revelation as shown in Figure 2.14.

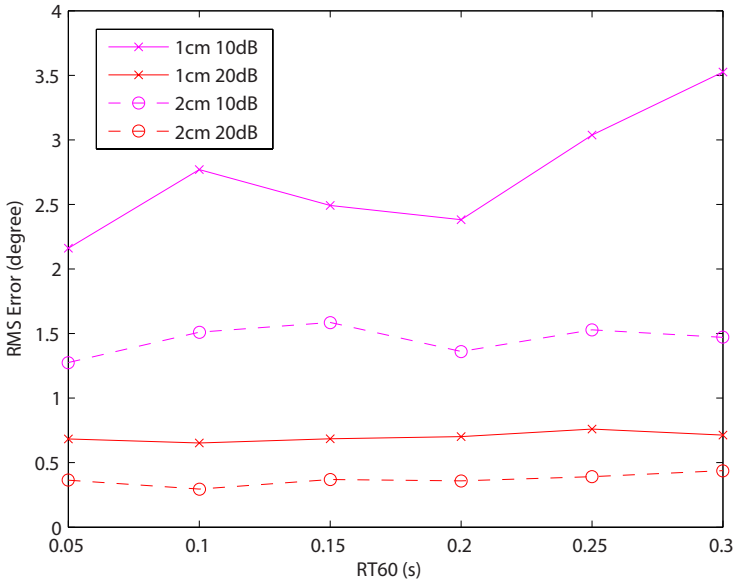

Figure 2.18: Standard deviation of the estimated angle of incidence as a function of the reverberation time for 1cm and 2cm sensor array and different level of additive measurement noise at 0 degree incidence angle with speech signal.

2.7 Conclusion

The gradient flow acoustic localization is presented and evaluated in adverse noisy and reverberant environments. The simulation results demonstrate robust performance with mild to moderate reverberations with additive noise levels down to 10dB. As the gradient flow technique in combination with independent component analysis provides a unique approach to the task of blind acoustic source separation, the separation performance will be investigated and quantified in the similar environments using the models developed in the presented work.

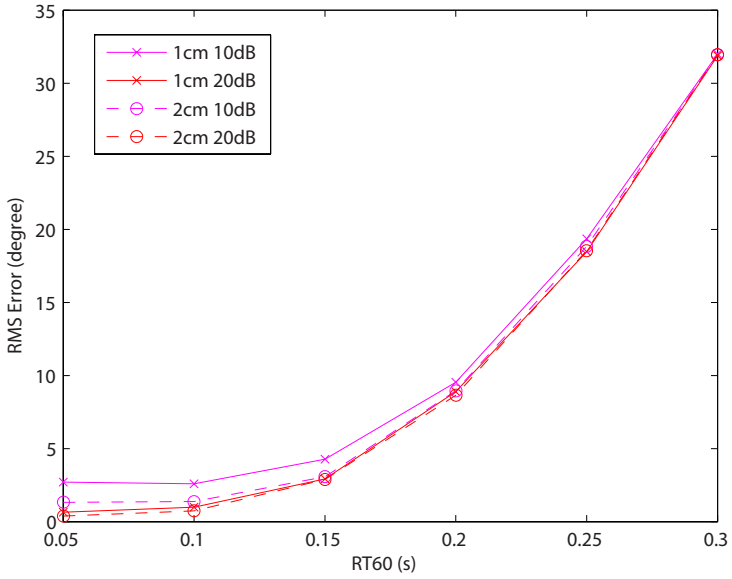

Figure 2.19: Standard deviation of the estimated angle of incidence as a function of the reverberation time for 1cm and 2cm sensor array and different level of additive measurement noise at 45 degrees incidence angle with speech signal.

Chapter 3

Gradient Flow Separation

3.1 Introduction

In a typical room environment the acoustic signals observed by microphone array are convolutive mixtures of source signals due to multi-path wave propagation [7]. 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 [38]. 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 [39–41]. Other works solve the permutation ambiguity based on neighboring correlations and exploiting the harmonic structure [42].Conventional knowledge dictates that sensor arrays with large inter-sensor distance should be used for source sepa-

ration 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 [11, 43]. 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 [32]. Using the gradient flow representation, I 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.

3.2 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

$$\mathbf{x} = A\mathbf{s} + \mathbf{n}, \tag{3.1}$$

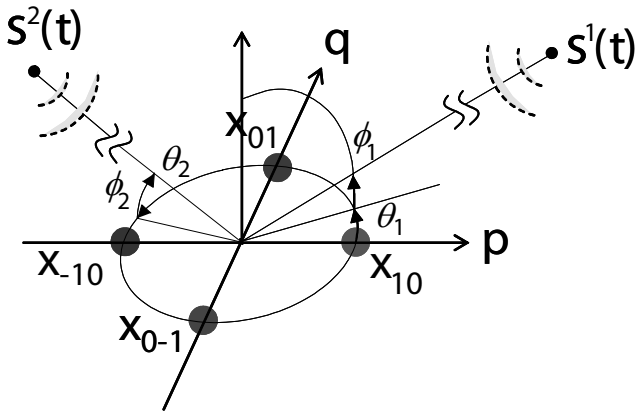

Figure 3.1: Miniature microphone array used in gradient flow technique.

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 A is the $M \times N$ 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

$$\mathbf{y} = W\mathbf{x} \tag{3.2}$$

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 3.1, the signals observed at the sensor array are mixture of the delayed source signals. Gradient flow [32] 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:

$$\begin{aligned} \xi_{00}(t) &= \textstyle\sum_l s^l(t) \\ \xi_{10}(t) &= \textstyle\sum_l \tau_1^l \dot{s}^l(t) \\ \xi_{01}(t) &= \textstyle\sum_l \tau_2^l \dot{s}^l(t) \end{aligned} \tag{3.3}$$

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 (3.1)

$$\begin{bmatrix} \dot{\xi}_{00} \\ \xi_{10} \\ \xi_{01} \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 \\ \tau_1^1 & \cdots & \tau_1^M \\ \tau_2^1 & \cdots & \tau_2^M \end{bmatrix} \begin{bmatrix} \dot{s}^1 \\ \vdots \\ \dot{s}^M \end{bmatrix}. \tag{3.4}$$

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.

3.3 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 (3.1) is a filter instead of a scalar and the i-th observed signal can be written as

$$x_i(t) = \sum_{j=1}^{n} \sum_{k} a_{ijk} s_j(t-k). \tag{3.5}$$

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

$$X_i(\omega) = \sum_{j=1}^{n} A_{ij}(\omega) S_j(\omega), \tag{3.6}$$

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, synthesizing 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.

I propose the following architecture for subband gradient flow ICA shown in Figure 3.2. The whole process is finished in the following four steps.

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 3.1. Although the static ICA suffers from the problem of poor convergence under strong reverberant environments, it provides a rough estimation on the directional pattern of the incoming sources. Here I utilize it to assist the matching of the signal coming from the same direction. The unmixing matrix solution is also fed to the next stage as initial unmixing matrix in each frequency band to help strength the independence assumption in subband ICA.

Second, 16-channel filterbank is used to decompose the spatial gradient signals. The filterbank is implemented in Mel scale, with linear-divided banks

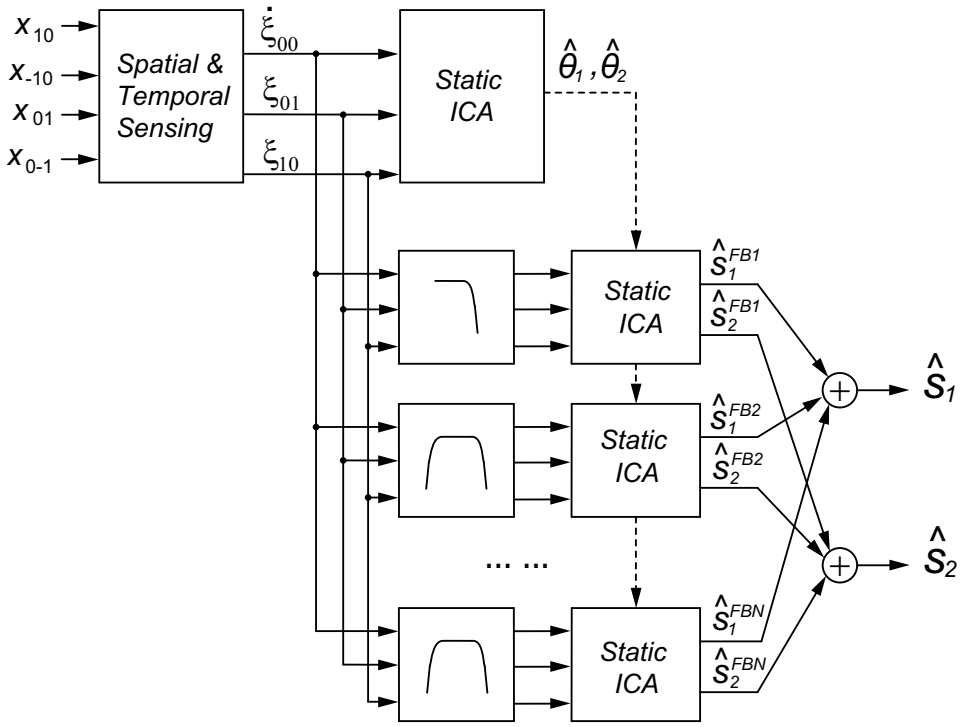

Figure 3.2: Block diagram of the proposed subband gradient flow ICA architecture.

from 0 to 1kHz and log-divided banks from 1kHz to 8kHz. The subband analysis filter bank is effectively constructed as the cosine modulated version of a prototype filter $h_0(n)$ of length N and cutoff frequency such that [13, 44]

$$h_k(n) = h_0(n)cos(f_c n\pi), \tag{3.7}$$

in which f_c is the normalized center frequency of each filter bank. In the prototype filter design, a truncated sinc function weighted by a Hamming window $w(n)$ is created, such that

$$h_0(n) = sinc(\frac{BW}{2}n)\frac{BW}{2}w(n), \tag{3.8}$$

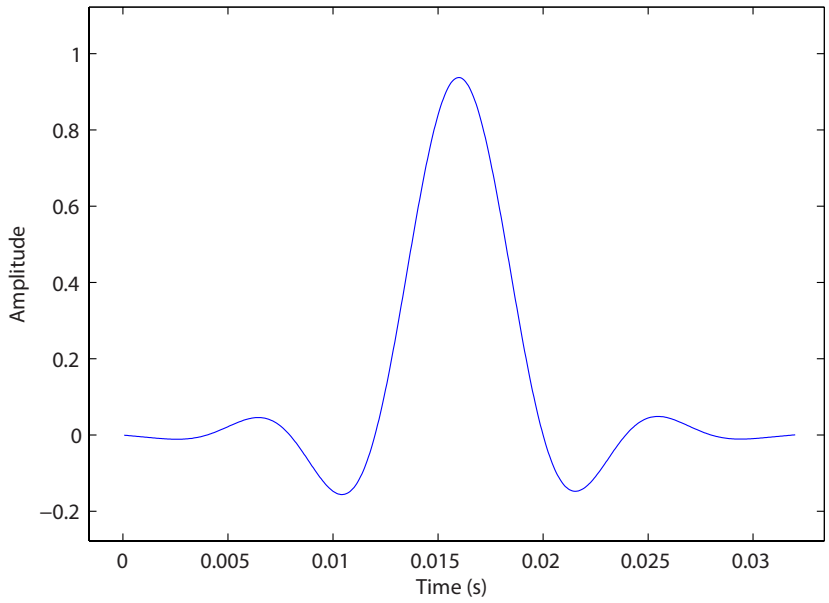

Figure 3.3: Impulse response of the low-pass FIR analysis filter with N = 512 and Fs = 16kHz.

with $w(n) = 0.54 - 0.46cos(2n\pi/N)$, and BW the normalized bandwidth of each filter bank. The length N of the filter bank is chosen to be 512. Figure 3.3 shows the impulse response of the low-pass FIR analysis filter. Figure 3.4 shows the frequency response of all the 16 filter banks implemented. The stopband attenuation is about 120dB.

Third, static ICA algorithm is used in each frequency band to obtain the unmixing matrix and signal estimation. To help convergence of single band ICA get to the global optima, 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

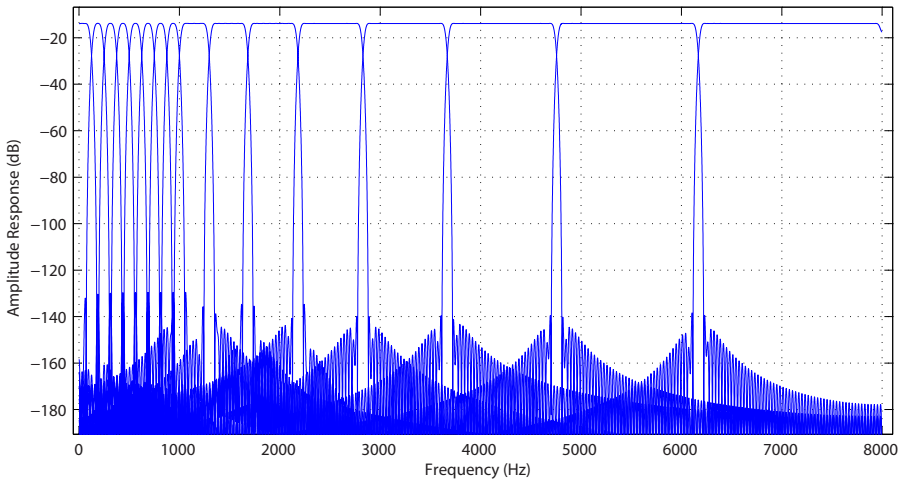

Figure 3.4: Frequency response characteristics of the implemented 16 filter bank .

a specific band strongly deviates from the direction obtained in the unfiltered static ICA, it is assumed that signal is not present in that frequency band. The initial unmixing coefficients from fullband ICA overwrite the unmixing coefficients of subband ICA in that frequency band. 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, I choose the first row of the estimated mixing matrix as the scaling factor to resolve the scaling ambiguity. Figure 3.5 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

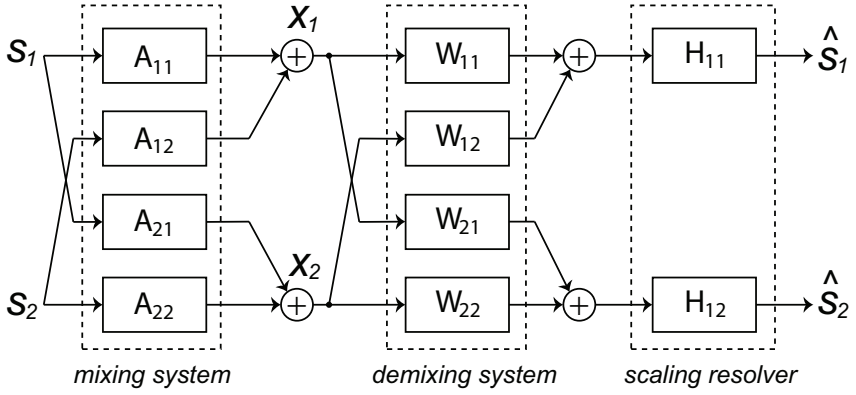

Figure 3.5: Block diagram of ICA algorithm in one frequency bin.

bins.

Finally, I align and synthesize the estimated signals from each frequency bin back into full-band estimations based on the preprocessing static ICA directional pattern. The issue of permutation ambiguity is an obstacle for application of frequency-domain ICA techniques. I propose to solve the problem by applying the estimated localization results from static ICA as a geographical cue to assist alignment of estimated results from each frequency bin. Estimated signal with similar incidence angles are merged into the same fullband signal bin so that homogeneity can be guaranteed.

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.

3.4 Simulation Results

3.4.1 Gradient Flow ICA Simulation Results

The effect of the additive sensor noise and the amount of reverberation on the separation performance using both synthesized data 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 3.1, where the microphones are in planar orthogonal positions and the distance between the opposing microphones is 1cm. To quantify the performance under different reverberant conditions, we first constructed an artificial room using the same virtual source mapping model in the previous chapter 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 [6m, 4m, 2.5m], 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 3.6. 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

$$SIR = -10 \log_{10} \min_i \frac{\sum_j < y_{ij}^2 > - \max_j < y_{ij}^2 >}{\max_j < y_{ij}^2 >}, \tag{3.9}$$

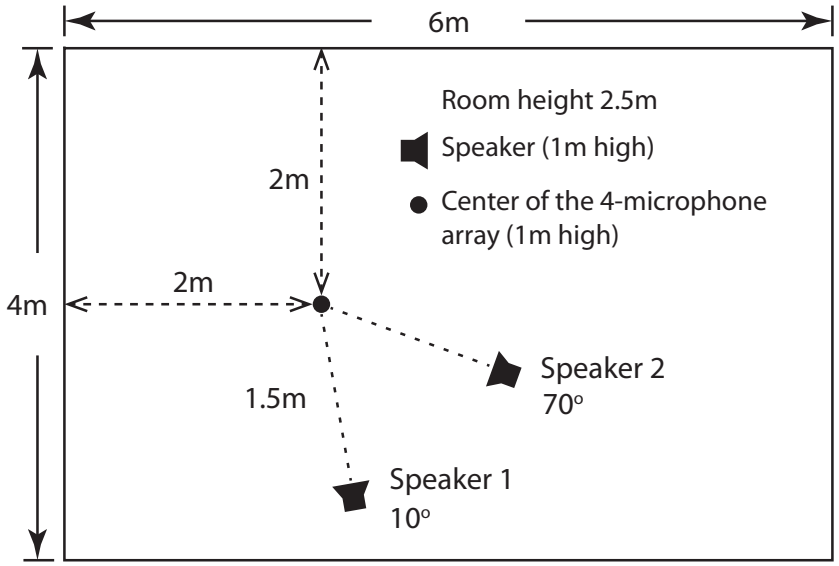

Figure 3.6: Simulated room dimensions and location of the sensor array and speakers.

where y_{ij} is the contribution of the source signal j to the estimated source signal i and is considered as interference.

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 [45] was used. The same elevation angles were assumed for both sources and the azimuth angle of the first source was set at $\theta_1 = 30^o$. The az-

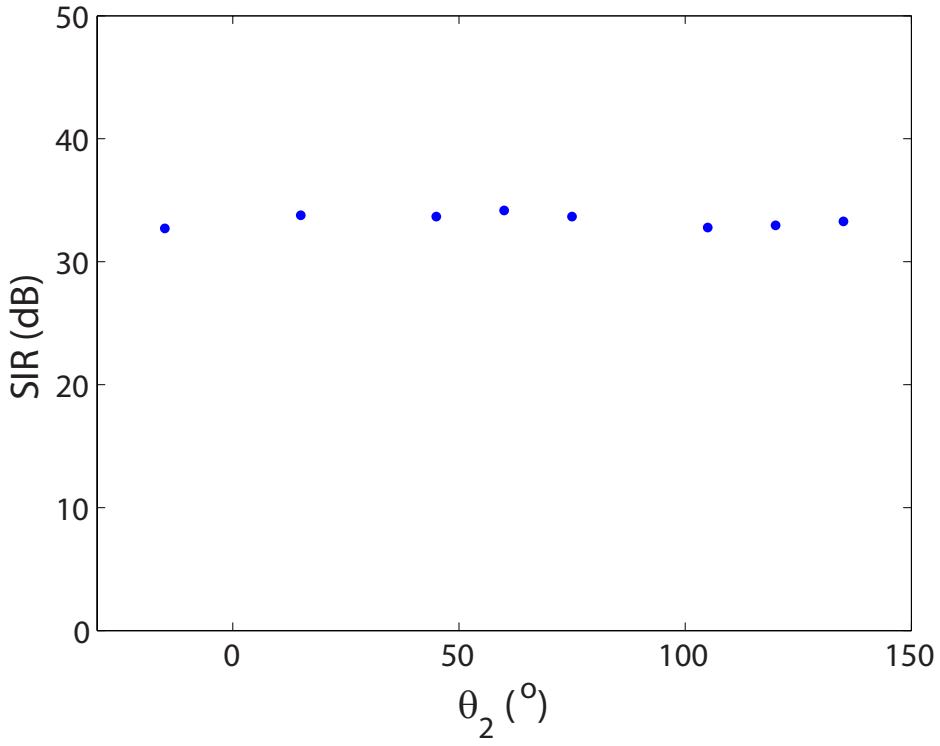

Figure 3.7: SIR for separation of two sources incident on the microphone array. The azimuth angle of the first source is fixed at θ_1=30^o and the azimuth angle of the second source θ_2 is varied from -15^o to 135^o.

imuth angle of the second source θ_2 was varied from -15^o to 135^o in increments of 15^o. We have omitted the locations of the second source at 0^o and 90^o where the separation is trivial. The separation performance, SIR, as a function of the azimuth angle of the rotating source is shown in Figure 3.7. The measurement results demonstrate that the separation performance does not depend strongly on the angular separation in anechoic and noiseless environment.

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^o and 70^o. The results for different signal-

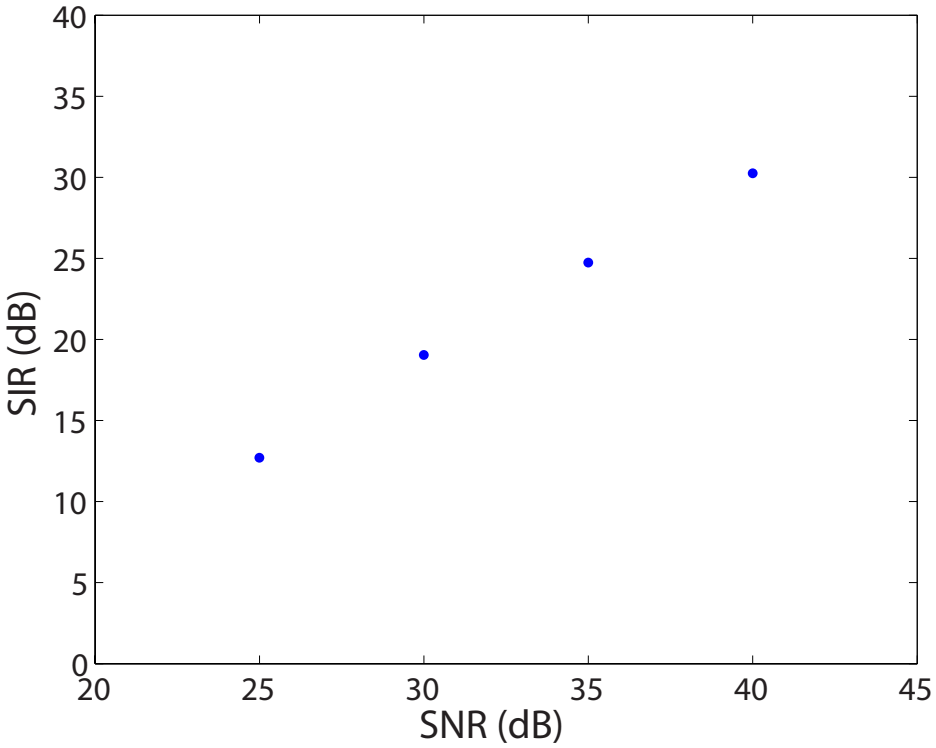

Figure 3.8: Separation performance as a function of spatially uncorrelated sensor noise when the source signals are impinging the array at 30^o and 70^o.

to-noise ratios (SNR) are presented in Figure 3.8. 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 (3.3), 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 3.9 shows the separation performance under different reverberation conditions, with different levels of

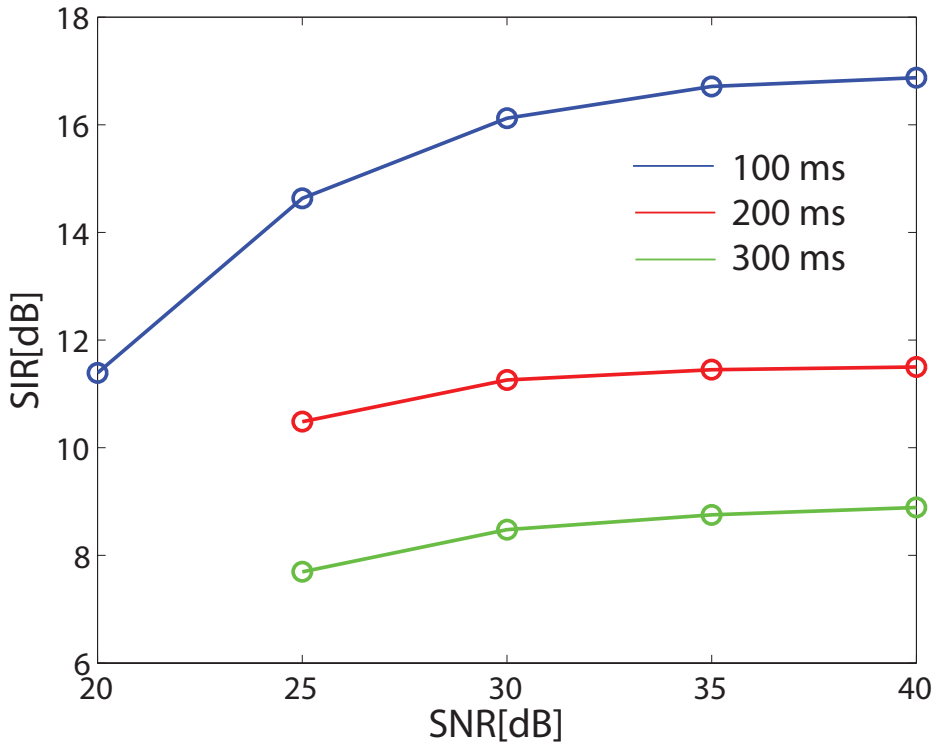

Figure 3.9: Separation performance with different reverberation times in the simulated room environment as a function of additive sensor noise.

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.

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 op-

posing microphones is kept at 1cm. 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 different 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^o and the incidence angle of the second source was swept from 15^o to -45^o.

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 3.10. 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^o, the SIR could stay above 10dB which is very impressive for a rather noisy real environment source separation.

3.4.2 Subband ICA Simulation Results

Next, the performance of the proposed gradient flow subband ICA model is tested and quantified in simulated adverse acoustic conditions. I 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

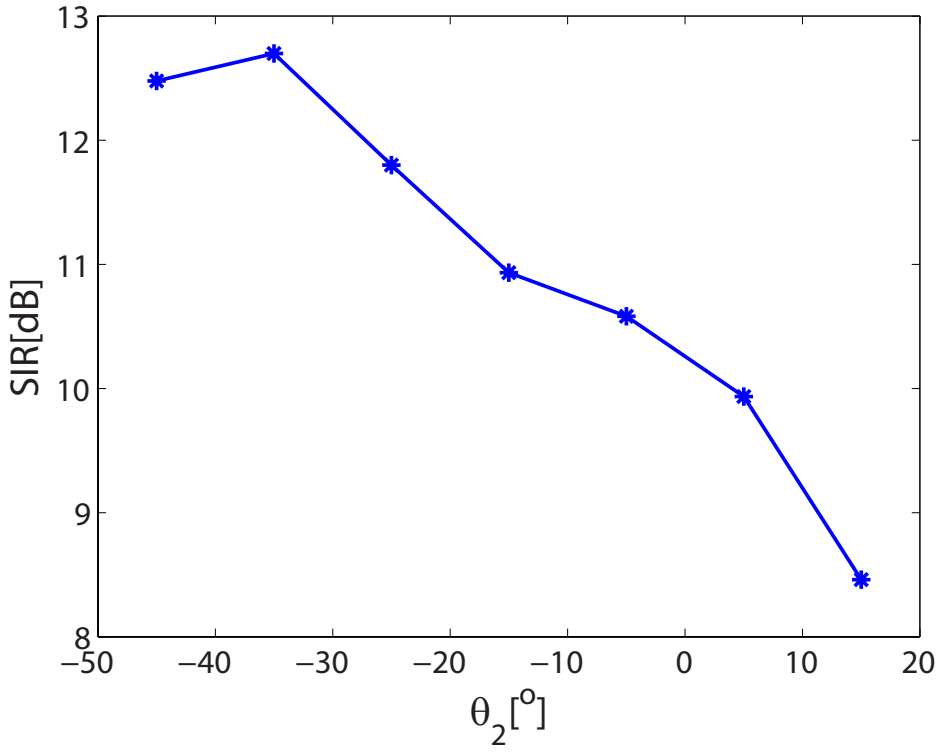

Figure 3.10: The separation performance for two speech signals recorded using miniature microphone array in a typical conference room environment as a function of incidence angle θ_2. The incidence angle of the first source is $\theta_1 = 30^o$.

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) [45]. 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 [35]. Simulated room dimensions correspond to an ordinary office space with the room dimensions and the relative location of the sensor array and the

Table 3.1: 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

speech sources shown in Figure 3.6. The configuration of the microphone array is orthogonal, with 1 cm inter-microphone spacing as illustrated in Figure 3.1. 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 static ICA and subband ICA algorithms is executed under various reverberant environments. Signal-to-interference ratio is selected to be the evaluation metric, which is based on [46]. Table 3.1 summarizes the 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 four hearing aid microphone array. The relative directional angles of the

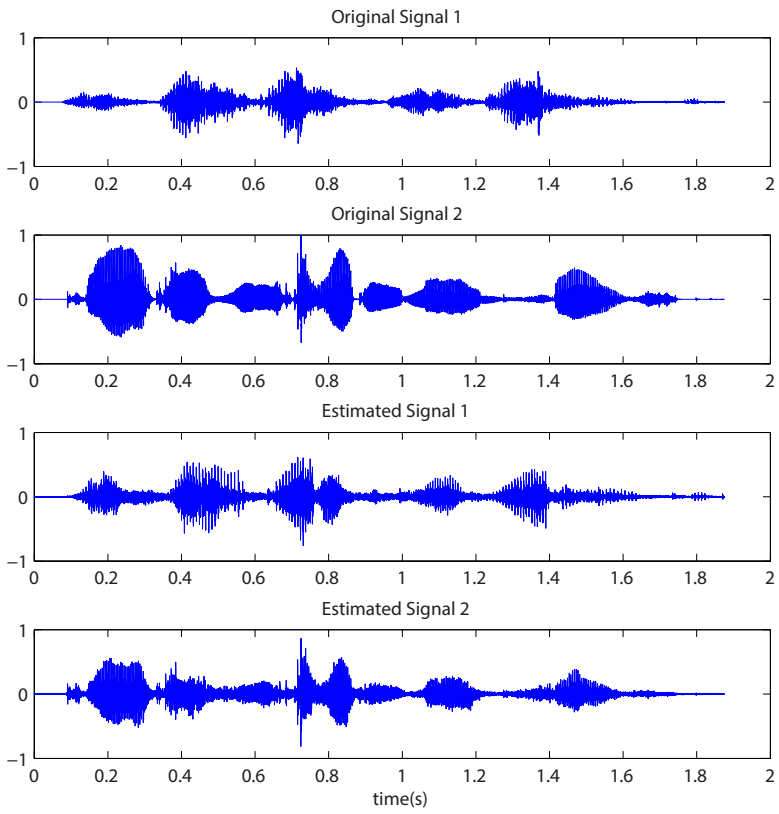

Figure 3.11: Time waveforms of the presented speech sources and signals recovered by the subband ICA algorithm.

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 3.11 and demonstrate the separation of around 15dB.

3.5 Study on Sensor Array Aperture

In terms of acoustic separation of multiple sources, the gradient flow technique converts the standard linear static ICA formality into a format with clear-defined physical meaning, by turning the mixing matrix into time delay matrix that uniquely determine the directions of the source signals. We further adopted the subband ICA methodology and process the signals in time domain within each band to utilize the benefits of both time-domain ICA and frequency-domain ICA. Futhermore, we introduced the permutation and scaling resolver to align and reconstruct the estimated signals from each subband [47]. The algorithm exhibits satisfactory performance advantage over the static ICA algorithm under moderate reverberant environments.

Nonetheless, the differential signals in the microphone observations are based on a large common-mode pedestal and are subject to the interference of noise originated from the sensor or the readout electronics. Especially in processing low-frequency signals when the wavelength is in the range of several meters, a microphone array 1cm in diameter could produce very little differential signal because of the small spatial diversity across the sensors. Also, when the frequency is high and the wavelength is in the centimeter range, a smaller array could potentially produce more accurate spatial derivatives and have a positive effect on the signal separation. Thus we need to look into the effect that changing the sensor dimension could have on the separation performance in the low-frequency and high-frequency subbands and improve the overall performance on the fullband signal.

3.5.1 Single Frequency Analysis

First, an analysis on sensor dimension using single frequency sinusoidal input signal is carried out in the similar fashion as section 2.6.1. Consider two sinusoidal input signals of the same frequency were mixed at the sensor array. Assume the incoming signal is on the same horizontal plane as the sensor array for simplicity, so $\phi = 0$. Apply 2.8 to 3.4, I could rearrange the form of the estimation of the mixing matrix A to

$$A = \begin{bmatrix} 1 & 1 \\ \tau_1^1 & \tau_1^2 \\ \tau_2^1 & \tau_2^2 \end{bmatrix} = \tau_{max} \begin{bmatrix} 1 & 1 \\ \cos\theta_1 & \cos\theta_2 \\ \sin\theta_1 & \sin\theta_2 \end{bmatrix}, \tag{3.10}$$

in which $\tau_{max} = d/c$. Since it is desirable that the unmixing matrix W will satisfy the requirement $WA = I$, a natural solution of W to the mixing matrix A above will be:

$$W = \begin{bmatrix} -\cos(2\theta_2) & \cos\theta_2 & -\sin\theta_2 \\ -\cos(2\theta_1) & \cos\theta_1 & -\sin\theta_1 \end{bmatrix}. \tag{3.11}$$

However, the actual mixing matrix $\hat{A}$ according to how the temporal and

spatial derivatives were formulated in 2.15, 2.16, and 2.17 is

$$\hat{A} = \begin{bmatrix} 1 & 1 \\ \frac{\sin(\pi f \tau_1^1)}{\pi f[\cos(\pi f \tau_1^1)+\cos(\pi f \tau_2^1)]} & \frac{\sin(\pi f \tau_1^2)}{\pi f[\cos(\pi f \tau_1^2)+\cos(\pi f \tau_2^2)]} \\ \frac{\sin(\pi f \tau_2^1)}{\pi f[\cos(\pi f \tau_1^1)+\cos(\pi f \tau_2^1)]} & \frac{\sin(\pi f \tau_2^2)}{\pi f[\cos(\pi f \tau_1^2)+\cos(\pi f \tau_2^2)]} \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 1 \\ \frac{\sin(P\cos\theta_1)}{\pi f[\cos(P\cos\theta_1)+\cos(P\sin\theta_1)]} & \frac{\sin(P\cos\theta_2)}{\pi f[\cos(P\cos\theta_2)+\cos(P\sin\theta_2)]} \\ \frac{\sin(P\sin\theta_1)}{\pi f[\cos(P\cos\theta_1)+\cos(P\sin\theta_1)]} & \frac{\sin(P\sin\theta_2)}{\pi f[\cos(P\cos\theta_2)+\cos(P\sin\theta_2)]} \end{bmatrix}, \quad (3.12)$$

in which $P = \pi f \frac{d}{c}$ for simplicity. In this case, the product $G = W\hat{A}$ will not be in the form of an identity matrix, because the off-diagonal term of G will not be zero. So I will define the SIR in this analysis as

$$SIR = 10\log\left[\frac{G(2,2)}{G(2,1)}\right]^2, \quad (3.13)$$

which is the amount signal source 1 gets attenuated in the recovered signal source 2 and study the SIR change according to the change in incidence angle and input signal frequency under different array dimensions.

Figure 3.12 shows the dependence of SIR on the different incidence angle of the first source under 2kHz signal and a 1cm sensor array. The first source is moved from 0° to 180°, and the second source is fixed at 30°. A minimum can be observed when the first source is also at 30°. When the two sources overlap, it can be predicted there will not be any separation between the two sources, thus the SIR drops to 0dB. As the angular difference between the two sources increases, the SIR also increases, which is intuitive.

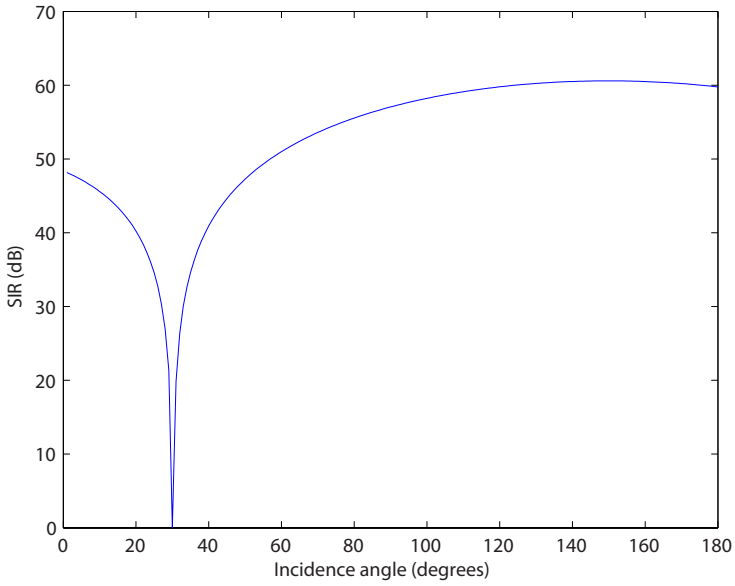

Figure 3.12: SIR against incidence angle for 2kHz signal and 1cm array.

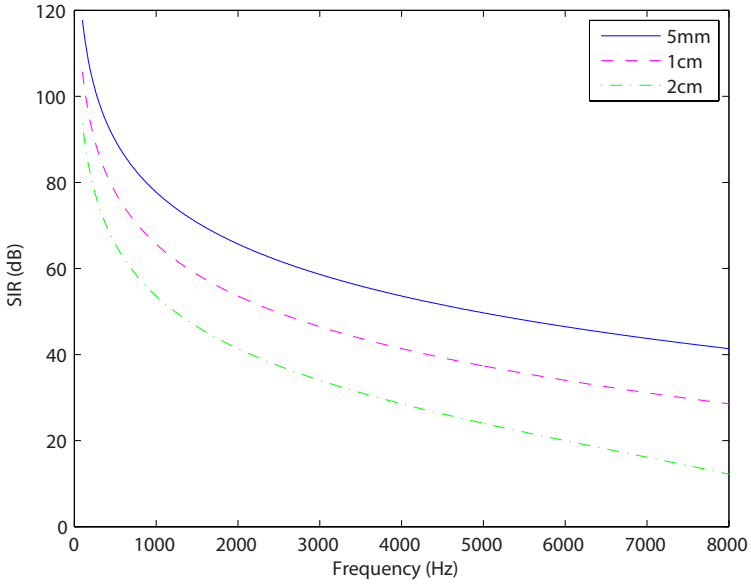

Figure 3.13: SIR variation against input frequency for 5mm, 1cm and 2cm array.

Figure 3.13 shows the dependence of SIR on the input sinusoidal signal frequency under 5mm, 1cm and 2cm sensor array respectively. In this analysis, incidence angles of 30° and 70° are selected to represent typical angular separation in the acoustic mixing scene. Similar trend as the localization performance is observed. Across all the frequencies, the separation performance will improve when the array dimension becomes smaller. For lower frequencies, the SIR is larger, and the difference in the SIR is not as significant as when the frequency rises. Thus the need to use smaller array dimension is further justified. To ensure better separation, when frequency gets higher thus the wavelength shrinks, the sensor array has to be decreased to achieve a good level of separation in anechoic and noiseless environment.

3.5.2 Simulated Low-frequency Separation Performance Comparison

The impact of the 2cm sensor array on low-frequency signal separation performance is evaluated. The room configuration used in the simulation is the same as illustrated in Figure 3.6. Uniform reflection coefficient is calculated based on the given reverberation time to simulate a room impulse response. The reverberation time used in the simulation is 200ms and 300ms. One speaker is located at 10^o 1.5 meters away from the center of the array and the other speaker is located at 70^o with the same distance. The sampling rate is 16kHz. The signal source is a 100Hz to 400Hz bandlimited speech signal from the TIMIT speech corpus. The length of the signal is 1.5 s.

Table 3.2 summarizes the SIR for both 1cm and 2cm sensor array under

Table 3.2: Comparison of 1cm and 2cm sensor array results

	RT_{60} = 200ms		RT_{60} = 300ms	
	SIR1	SIR2	SIR1	SIR2
1cm Array	7.36dB	10.75dB	4.26dB	9.80dB
2cm Array	8.67dB	13.20dB	4.77dB	11.16dB

200ms and 300ms reverberation times. We can conclude from the table that the 2cm microphone array is very effective in boosting the separation performance, with an average SIR increase of 2dB.

3.5.3 Fullband Separation Performance Comparison

First, we simulated similar situations as the localization part to study the effect of sensor array dimension on separation performance across all the filter banks. The signal source is still a speech signal from the TIMIT speech corpus and the length is truncated to 1.5 s. Both speakers are 1.5 meters in distance from the center of the array and their incidence angle is 10^o and 70^o.

Figure 3.14 shows the separation performance improves in the 2cm case over all the filter banks when there is only noise. Figure 3.15 shows there is almost no performance improvement in the 2cm case when there is only reverberation. The two plots corresponds to our findings in the localization part, that the increase of sensor array dimension only assists better performance when there is additive noise instead of interference noise. Figure 3.16 demonstrates the separation performance only improves in the first several low-frequency filter banks in the 2cm case when in both noisy and reverberant environment.

Figure 3.14: SIR improvement of 2cm over 1cm sensor array across all filter banks when there is only noise (SNR=40dB).

Then the separation performance of two configuration is tested. Configuration 1 is comprised of observations all from the 1cm sensor array. In configuration 2, we use the observation from the first 4 filter banks from the 2cm sensor array to improve the low-frequency performance, and for the rest of the banks, observations from 1cm array is used. Table 3.3 summarizes the SIR for both array configurations under 200ms and 300ms reverberation times. We can conclude from the table that the advantage in boosting the separation performance with 2cm microphone array in low frequency is very limited, with an average SIR increase of only around 0.3dB.

The 2cm microphone array shows superior results in source localization and separation when there is only noise in the environment. Under echoic situations

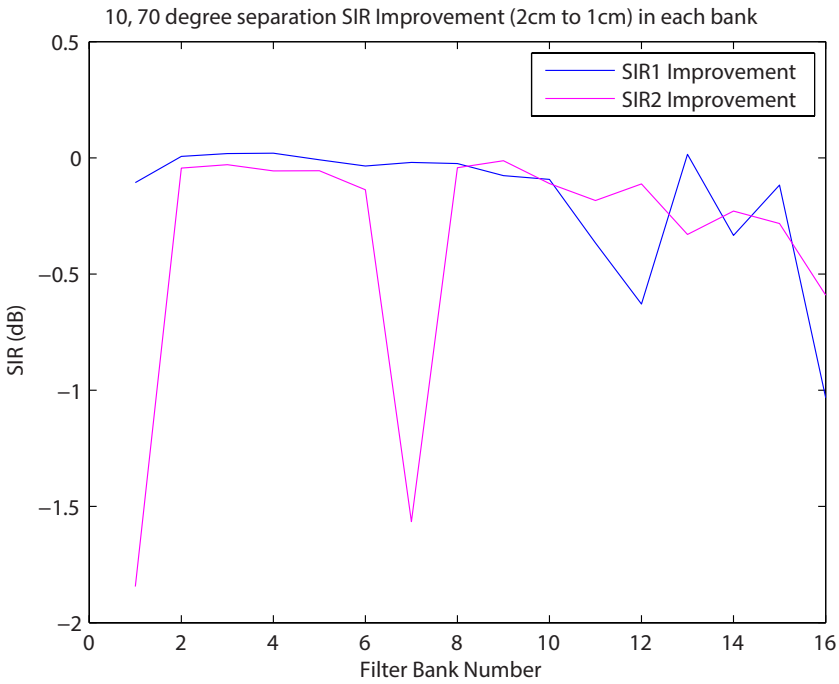

Figure 3.15: SIR improvement of 2cm over 1cm sensor array across all filter banks when there is only reverberation (Rt60=200ms).

the increase in the inter-sensor spacing does not contribute much improvement in separation performance. Because the 2cm microphone array did not show convincing improvement in real 16-band filterbank separation, and the size of the device is going to increase dramatically, we make the compromise to adopt just 1cm and 5mm sensor array in our implementation. We will look for other possibilities such as using hybrid derivative computation to boost the SNR in the future works.

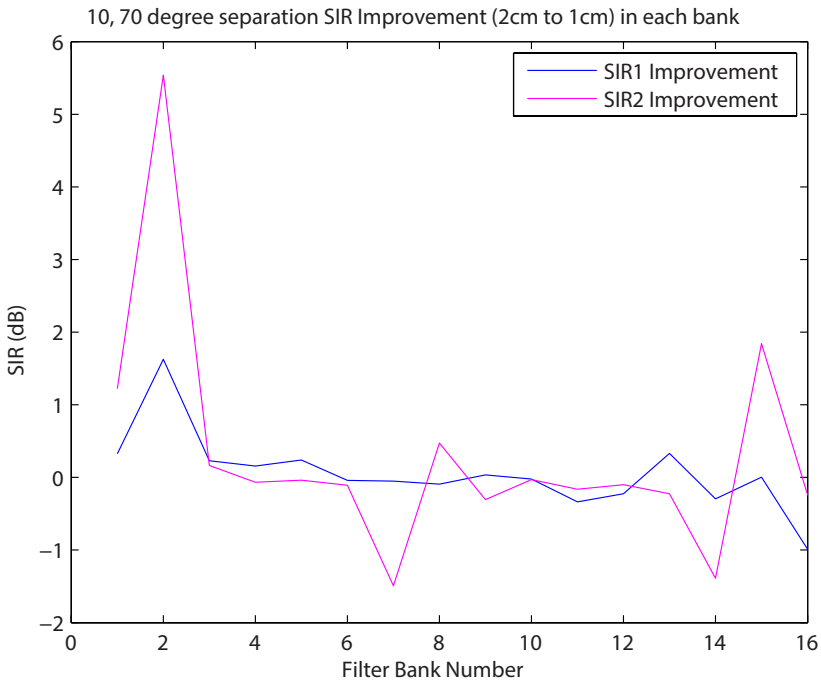

Figure 3.16: SIR improvement of 2cm over 1cm sensor array across all filter banks when both noise (SNR=40dB) and reverberation (Rt60=200ms) is presented.

3.6 Conclusion

The gradient flow subband acoustic separation architecture is proposed 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.

Table 3.3: Comparison of 1cm and 2cm sensor array results

	RT_{60} = 200ms		RT_{60} = 300ms	
	SIR1	SIR2	SIR1	SIR2
Configuration 1 Array	13.66dB	13.90dB	6.49dB	10.64dB
Configuration 2 Array	13.81dB	14.45dB	6.40dB	10.97dB

Chapter 4

Mixed-signal VLSI Gradient Flow Acoustic Localizer

4.1 Introduction

Precise and robust localization and tracking of acoustic sources using miniature microphone arrays in small form-factor and low power is of interest to a variety of applications, from hearing aids to surveillance and multimedia. There have been a number of VLSI systems for acoustic direction finding reported in the literature [34, 48, 49], but not a complete system-on-chip solution that can be integrated directly with microphone array.

Gradient flow [34] is a signal conditioning technique for source localization designed for arrays of very small aperture, *i.e.,* of dimensions significantly smaller than the shortest wavelength in the sources. In this chapter, the gradient flow principle was implemented in mixed-signal VLSI. By adopting a mixed-signal architecture in the implementation, the advantages of both

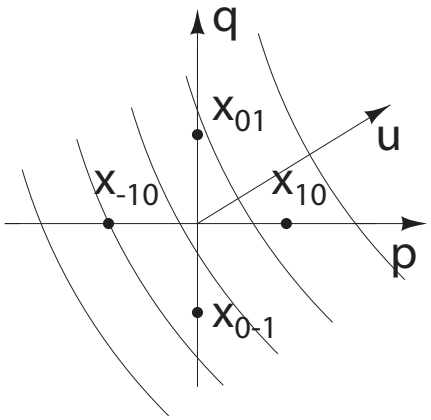

Figure 4.1: Configuration of sensors for spatial gradient estimation.

approaches are combined: an analog data path directly interfaces with inputs without the need for data conversion; and digital adaptation offers the desired outputs as digital values, performing 'smart' analog-to-digital conversion. Estimation of the gradients is implemented using sampled-data switched-capacitor (SC) circuits. The advantage of this implementation is application of correlated-double sampling (CDS) that significantly reduces common-mode offsets and $1/f$ noise [50], which is the dominant noise source at low frequency. The spatial gradients are computed in fully differential mode, to provide increased clock and supply feedthrough rejection. A CMOS cascoded amplifier implements a simple high-gain amplifier in all of SC circuits, supporting high density of integration, and low power consumption.

The system block diagram implementing gradient flow for bearing estimation is shown in Figure 4.2. Filtering and preamplification stage enables direct interface to microphone array and design of low-power system-on-chip matched for hearing aid applications. Spatial gradients are approximated by evaluating finite differences over the four sensors on the planar grid shown in

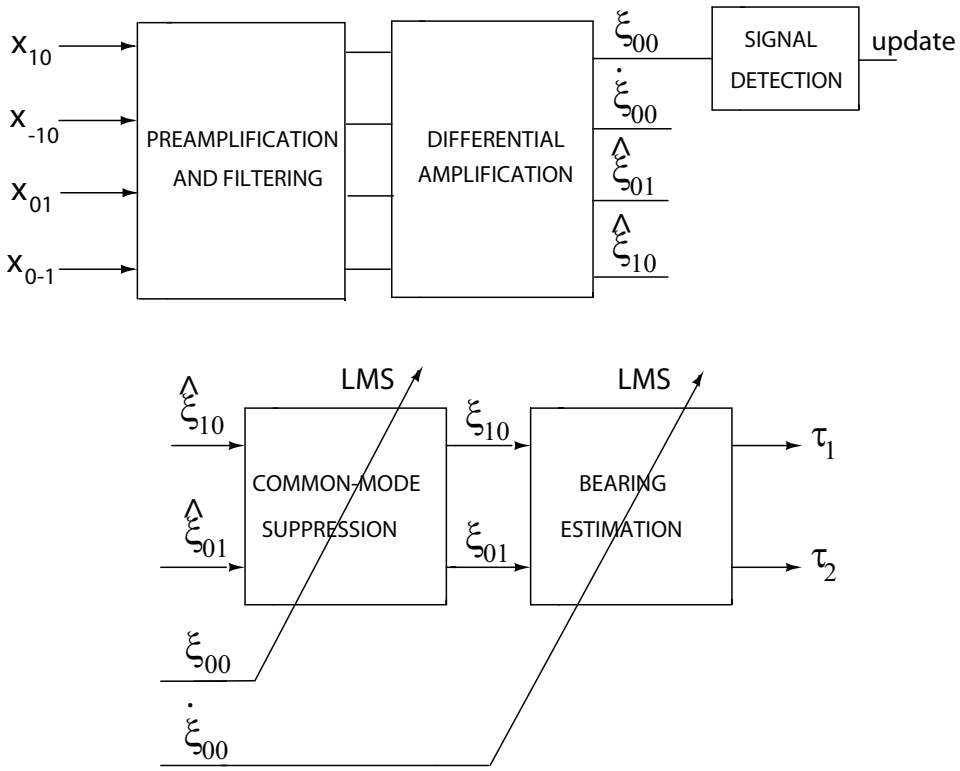

Figure 4.2: System block diagram.

Figure 4.1. Two stages of mixed-signal adaptation compensate for common mode errors in the differential amplification, and produce digital estimates of delays τ_1 and τ_2 from the spatial and temporal differentials. Signal detection, beside saving in power consumption, improves bearing estimation convergence speed and stability.

4.2 Circuit implementation

Gradient flow lends itself into efficient implementation in mixed-signal VLSI. The continuous-time $Gm - C$ circuits provide low-power and small-form factor implementation for filtering and preamplification of microphone signals.

Switched-capacitor circuits are ideal match for low-noise and low-frequency spatial gradient computation. Digital adaptation for bearing estimation offers flexibility and reconfigurability of the learning rules with estimates in digital format, avoiding need for high-resolution analog-to-digital conversion.

4.2.1 Bandpass Filtering and Preamplification of Microphone Signals

The proposed system-on-chip is intended for direct interface with microphone array in order to achieve small-form factor and improve power consumption. Performance of the gradient flow algorithm depends on spatial gradient signal acquisition, which is a small differential signal on large common-mode pedestal. Improved differential sensitivity of gradient sensing allows to shrink the aperture of the sensor array without degrading signal-to-noise ratio. The microphone signals have to be amplified and filtered before the spatial gradient computation. Instead of using discrete implementation of bandpass filter, that covers the bandwidth of the signal of interest, for better matching and improved power consumption, we propose bandpass filter on chip implementation as continuous time filter. Mismatch in amplitude and phase characteristics of filters in different channels leads to leakage of differential signal into common-mode signal and degrades performance of gradient flow algorithm.

The filtering of microphone signals is matched to the speech signal and implemented as second-order bandpass filtering, with low-frequency cutoff set at 100 Hz and high-frequency cutoff set at 8 kHz. The signals are also amplified by a factor of 10. A continuous time $Gm - C$ filter implements the bandpass

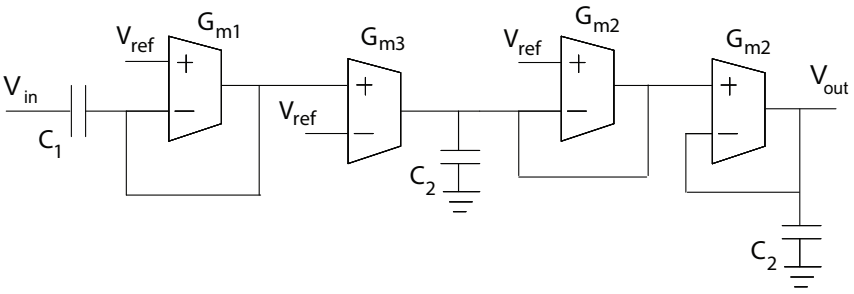

Figure 4.3: Continuous-time bandpass filter and pre-amplifier implementation.

filtering and preamplification. The design is shown in Figure 4.3. The transfer function of the filter is

$$H(s) = \frac{Gm_3}{Gm_2} \frac{\frac{Gm_1}{C_1}s}{\frac{Gm_1}{C_1}s + 1} \frac{1}{(\frac{Gm_2}{C_2}s + 1)^2} \tag{4.1}$$

Gm_1 and Gm_3 are nominally set to equal values. Gm_1/C_1 sets the low-frequency cutoff. Since this frequency is small, a large value of capacitance C_1 is required and this capacitance is implemented as an off-chip capacitor. The ratio of Gm_2 and Gm_3 sets the gain of the filter, while the ratio of Gm_2/C_2 sets the high-frequency cutoff. The implementation of OTA used in proposed $Gm - C$ filter implementation is shown in Figure 4.4.

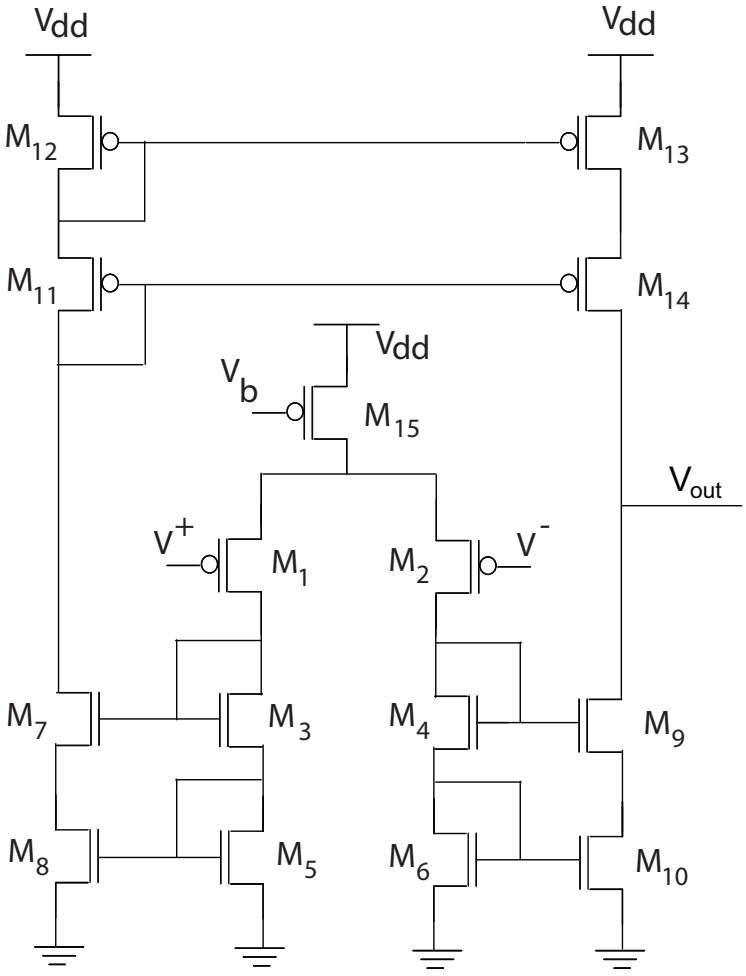

Figure 4.4: OTA implementation.

4.2.2 Spatial Gradient Calculation

Estimates of ξ_{00}, ξ_{10} and ξ_{01} are obtained from the sensor observations $x_{-1,0}$, $x_{1,0}$, $x_{0,-1}$ and $x_{0,1}$ as:

$$\begin{aligned} \xi_{00} &\approx \tfrac{1}{4}\big(x_{-1,0} + x_{1,0} + x_{0,-1} + x_{0,1}\big) \\ \xi_{10} &\approx \tfrac{1}{2}\big(x_{1,0} - x_{-1,0}\big) \\ \xi_{01} &\approx \tfrac{1}{2}\big(x_{0,1} - x_{0,-1}\big) \end{aligned} \tag{4.2}$$

Computation of the gradients is implemented using sampled-data switched-capacitor (SC) circuits. The advantage of this realization is application of correlated-double sampling (CDS) to significantly reduce common-mode offsets and $1/f$ noise [50]. A cascoded amplifier, biased in subthreshold regime, is used as high-gain amplifier in these and subsequent SC circuits, supporting high density of integration, and high energetic efficiency.

The common-mode component is decomposed in differential form $\xi_{00} = \xi_{00}^{+}[n] - \xi_{00}^{-}[n]$ with

$$\begin{aligned} \xi_{00}^{+}[n] &= \frac{1}{8}(x_{10}[n-\frac{1}{2}] + x_{-10}[n-\frac{1}{2}] \\ &\quad + x_{01}[n-\frac{1}{2}] + x_{0-1}[n-\frac{1}{2}]) \\ \xi_{00}^{-}[n] &= -\frac{1}{8}(x_{10}[n] + x_{-10}[n] + x_{01}[n] + x_{0-1}[n]). \end{aligned} \tag{4.3}$$

The contribution ξ_{00}^{+} to ξ_{00} represents the estimate of the average signal at time instance $nT - \frac{T}{2}$, while the contribution ξ_{00}^{-} represents the inverted estimate at time instance nT. The difference between both contributions signals

hence produces an unbiased estimate of ξ_{00} centered at time $nT - \frac{T}{4}$. The corresponding switched-capacitor realization is given in Figure 4.5. Instead of using a differential operational amplifier as high gain element in SC circuit, we have chosen to use a lower power, single ended cascoded amplifier in pseudo-differential implementation. In this implementation, the offset between virtual grounds of two branches directly contributes to the differential signal. In the previously reported structure [34], the virtual ground was defined by the inverting point of the cascoded amplifier and was prone to mismatch errors. The proposed structure contains the virtual ground capacitor C_s, that defines the virtual ground. C_s is precharged to the mid-point voltage V_{ref} in clock phase ϕ_1, and voltage V_{ref} is the value of virtual ground in both branches, eliminating dependance on input transistor parameters. The clocks ϕ_1 and ϕ_2 are nonoverlapping, and ϕ_{1e} replicates ϕ_1 with its falling edge slightly preceding the falling edge of ϕ_1.

An estimate of the temporal derivative signal $\dot{\xi}_{00}$ centered at same time instance $nT - \frac{T}{4}$ is computed differentially in similar manner, by differencing signal averages (or the reference level) at time instances $nT - \frac{T}{2}$ and nT,

$$\begin{aligned}
\dot{\xi}^{+}_{00}[n] &= 0 \\
\dot{\xi}^{-}_{00}[n] &= -\frac{1}{8}(x_{10}[n] + x_{-10}[n] + x_{01}[n] + x_{0-1}[n]) \\
&\quad + x_{10}[n - \frac{1}{2}] + x_{-10}[n - \frac{1}{2}] \\
&\quad + x_{01}[n - \frac{1}{2}] + x_{0-1}[n - \frac{1}{2}].
\end{aligned} \tag{4.4}$$

The spatial gradients are computed in fully differential mode, to provide in-

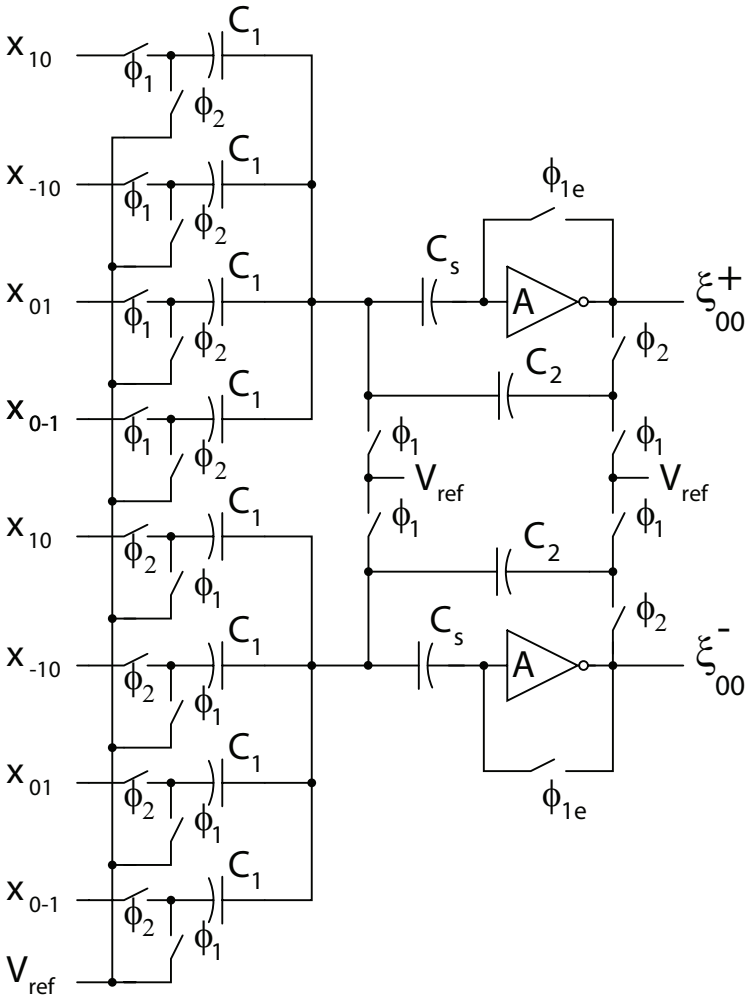

Figure 4.5: Pseudo-differential structure for computation of common-mode component.

creased clock and supply feedthrough rejection. The first-order spatial gradient ξ_{10}, likewise centered at time $nT - \frac{T}{4}$, is computed by differencing estimates of ξ_{10} at time instances $nT - \frac{T}{2}$ and nT

$$\begin{aligned} \xi_{10}^{+}[n] &= \frac{1}{4}(x_{10}[n - \frac{1}{2}] - x_{-10}[n]) \\ \xi_{10}^{-}[n] &= \frac{1}{4}(x_{-10}[n - \frac{1}{2}] - x_{10}[n]). \end{aligned} \tag{4.5}$$

The first-order spatial gradient in the q direction, ξ_{01}, is computed in identical fashion.

4.2.3 Common-mode suppression

Gain mismatch in the microphones and preamplification stage leads to contribution of common mode signal into the spatial gradient estimate. To estimate the leakage coefficient ε_1 of common-mode signal into spatial gradient estimation, a digital sign-sign LMS (SS-LMS) adaptation rule is used [34]. ε_1 is stored as digital value in a 12-bit counter and it is represented in two's complement. The update is performed by incrementing or decrementing the counter based on sign of spatial gradient and average signal

$$\begin{aligned} \varepsilon_1^{+}[n+1] &= \varepsilon_1^{+}[n] \\ &\quad +\text{sgn}(\xi_{10}^{+}[n] - \xi_{10}^{-}[n])\text{sgn}(\xi_{00}^{+}[n] - \xi_{00}^{-}[n]) \\ \varepsilon_1^{-}[n+1] &= 2^{12} - 1 - \varepsilon_1^{+}[n+1]. \end{aligned} \tag{4.6}$$

The 8 most significant bits are presented to a multiplying D/A capacitor array to construct the LMS error signal, for the case of ξ_{10}

$$\begin{aligned} \xi_{10}^{+}[n] &= \hat{\xi}_{10}^{+}[n] - (\varepsilon_1^{+}\xi_{00}^{+}[n] + \varepsilon_1^{-}\xi_{00}^{-}[n]) \\ \xi_{10}^{-}[n] &= \hat{\xi}_{10}^{-}[n] - (\varepsilon_1^{-}\xi_{00}^{+}[n] + \varepsilon_1^{+}\xi_{00}^{-}[n]). \end{aligned} \tag{4.7}$$

4.2.4 Bearing estimation

For implementation of bearing time-delay estimation, digital SS-LMS differential on-line adaptation is used similar to common-mode error correction. Bearing estimates are represented as 12-bit values in two's complement [34]

$$\begin{aligned} \tau_1^{+}[n+1] &= \tau_1^{+}[n] \\ &\quad +\mathrm{sgn}(e_{10}^{+}[n] - e_{10}^{-}[n])\mathrm{sgn}(\xi_{10}^{+}[n] - \xi_{10}^{-}[n]) \\ \tau_1^{-}[n+1] &= 2^{12} - 1 - \tau_1^{+}[n+1], \end{aligned} \tag{4.8}$$

with the 8 most significant bits used for computation of LMS error signal

$$\begin{aligned} e_{10}^{+}[n] &= \xi_{10}^{+}[n] - (\tau_1^{+}\dot{\xi}_{00}^{+}[n] + \tau_1^{-}\dot{\xi}_{00}^{-}[n]) \\ e_{10}^{-}[n] &= \xi_{10}^{-}[n] - (\tau_1^{-}\dot{\xi}_{00}^{+}[n] + \tau_1^{+}\dot{\xi}_{00}^{-}[n]). \end{aligned} \tag{4.9}$$

4.2.5 Signal Detection

For acoustic localization, detection of the presence of the signal is critical for bearing estimation process, specially in the case of speech like signals, that have a long pauses in the signal. It is also important for power saving, as

certain parts of the chip can be shut-off if the signal is not present. The simple thresholding function of the common-mode signal is chosen for signal detection. As the sign of common-mode signal is already necessary for update computation in common-mode suppression, the thresholding function is incorporated in the same comparator structure. Implementation of comparator that is able to compare signal with variable level, zero for the sign and positive and negative thresholding value, is shown in Figure 4.6(a). While the output signals are valid, ξ_{00}^{+} is sampled in phase ϕ_1^c on capacitor C_3. The sign of the comparison of the common-mode signal ξ_{00} with variable comparison level V_{th} is computed in the evaluate phase ϕ_2^c, through capacitive coupling into the amplifier input node. Sign of ξ_{00} would be computed in standard comparator implementation if capacitor C_4 was omitted. The addition of C_4 allows the computation of multiple level comparisons in a single cycle. Sign of ξ_{00} is computed first for reference level of V_{th} and the comparison with threshold level is enabled by the change in voltage V_{th} in phase ϕ_2^c in a single clock cycle. The clocks ϕ_1^c and ϕ_2^c are non-overlapping and their relative timing with respect to clocks ϕ_1 and ϕ_2 is shown in Figure 4.6(b), as well as the changes in voltage V_{th} for the sign and threshold comparison. ϕ_{1e}^c replicates ϕ_1^c with its falling edge slightly preceding the falling edge of ϕ_1^c.

4.3 Simulation results

The proposed architecture was implemented in 0.5 μm CMOS technology. Frequency response of the designed bandpass filter and preamplifier is shown in Figure 4.7. The simulated dynamic range of implemented OTA is 51 dB.

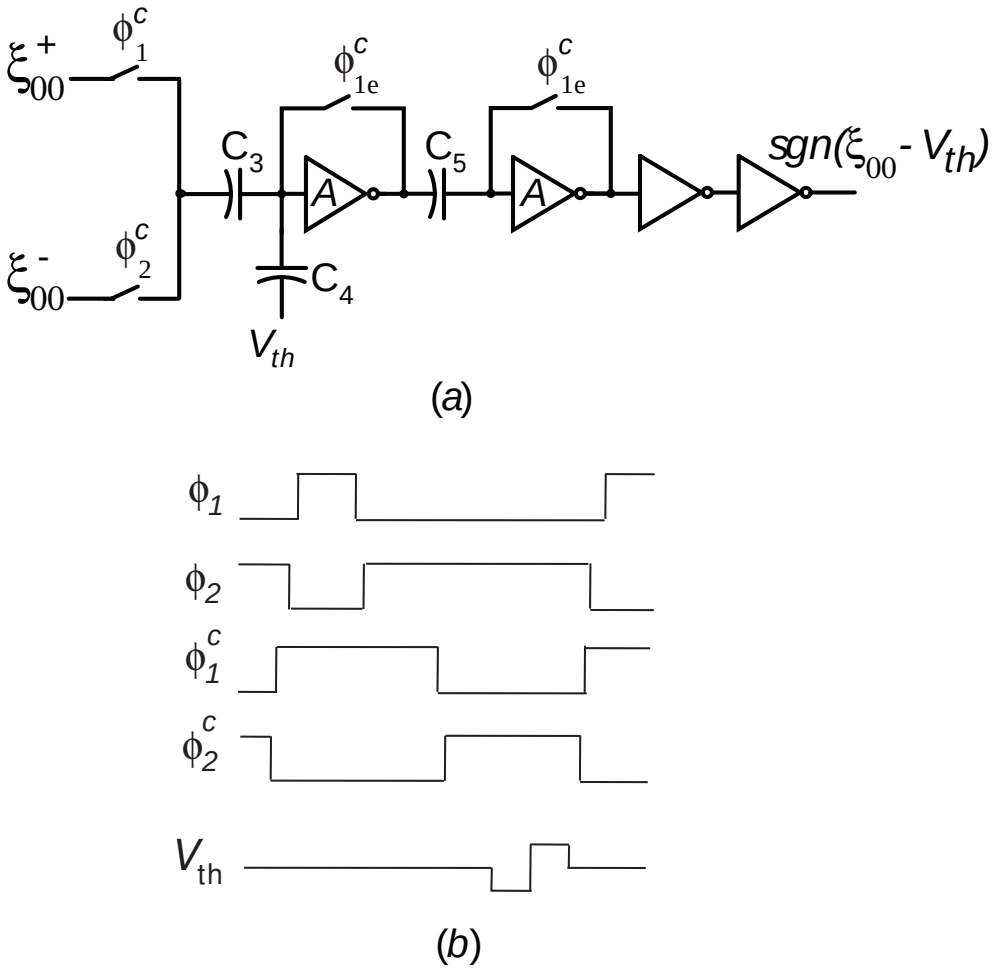

Figure 4.6: (a) Comparator design for sign and threshold comparison. (b) Illustration of relative position of clock signals and variable comparison level V_{th}.

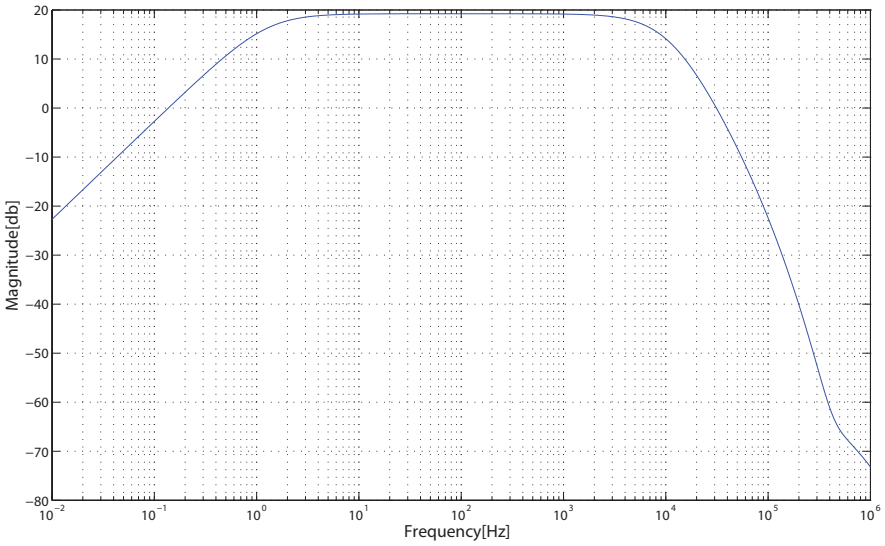

Figure 4.7: Frequency response of the designed bandpass filter and preamplifier.

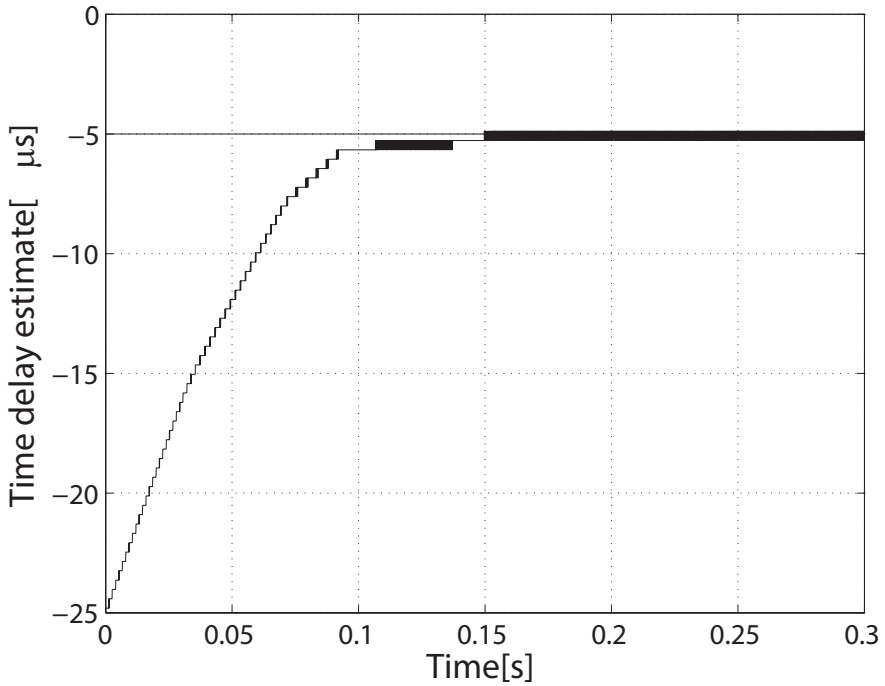

Figure 4.8: Bearing estimation adaptation for estimation of delay between two sine wave input signals.

The bearing estimation process is demonstrated using sine wave input signal at frequency of 1 kHz, with sampling frequency of 16 kHz. The signal was presented to x_{10} and x_{01}, and under 5 μs delay to x_{-10} and x_{0-1}. The update of time-delay estimation process is shown in Figure 4.8. The total power consumption of the proposed system-on-chip is estimated at a total of 150 μW, where 100 μW represents the power consumption of the bandpass filter, while 50 μW represents the power consumption of the spatial gradient computation, signal detection and bearing estimation.

4.4 Conclusion

The proposed system-on-chip solution for acoustic localization is implemented in small-form factor and operates at microwatt power consumption, making it suitable for direct integration with microphone array. The chip well matches requirements for hearing aid and acoustic surveillance applications.

Chapter 5

Mixed-signal VLSI Independent Component Analyzer

5.1 Introduction

A person can seamlessly focus and understand a specific speaker under various levels of the background noise. However, the performance of the current state-of-the-art hearing aids, as well as the speech recognition software, significantly deteriorates with other speakers present in the background. Smart sensing hearing aids could greatly benefit from robust speech separation in adverse acoustic environments.

I have proposed an algorithm that combines in a unique framework the spatial sampling, sub-band processing and independent component analysis to achieve improvement in the separation performance in moderate reverberant acoustic environments [47, 51]. The hardware implementation of the proposed algorithm requires a 16-channel 3x3 linear static ICA architecture leading to

stringent constraints on the chip area and power consumption of a single ICA channel. These constraints are not met in various proposed implementations of static ICA in either analog [52, 53] or digital domain [54, 55]. In the proposed implementation, I exploit the pulse width modulation to implement matrix-vector multiplication [56, 57] in order to meet the constraints on area and power consumption.

The implementation architecture and circuit blocks are discussed in the next sections, followed by the simulation results of the system.

5.2 Independent Component Analyzer Architecture

The independent component analysis is the signal processing technique for obtaining independent directions in multivariate data. In many cases it is exploited for the blind source separation (BSS), where the task is to recover the unknown sources $\mathbf{s}$ from their mixtures $\mathbf{x}$, without prior information on them, except their independence. I assume that the mixing is linear

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n}, \tag{5.1}$$

where $\mathbf{A}$ is dimensional mixing matrix and $\mathbf{n}$ is the additive noise. When the number of source N is greater than the number of observations M, $N > M$ some priori information on the source signals is necessary for solving the problem. For this mixing model (5.1), ICA solution is formulated as a linear transformation that minimizes the statistical dependence between components

in the output signals $\mathbf{y}$

$$\mathbf{y} = \mathbf{W}\mathbf{x}, \tag{5.2}$$

where $\mathbf{W}$ is $N \times M$ dimensional unmixing matrix. The unmixing matrix $\mathbf{W}$ is not uniquely defined, with ambiguity in scaling and permutation. The energy of the source signals cannot be determined, since both $\mathbf{s}$ and $\mathbf{A}$ are unknown and any scalar multiplier in one of the sources could be canceled by dividing the corresponding column of $\mathbf{A}$ by the same scalar. As the order of the independent sources is not predefined, any permutation of the columns of the matrix $\mathbf{W}$ is a valid solution of the separation problem.

A wide variety of ICA learning rules have been proposed in the literature [58, 59]. Most of the ICA learning algorithms are based on optimizing a cost function defined as the measure of independence between the components of the output signals [60]. Different approaches, like maximization of entropy [61], minimization of mutual information of the output signals [60, 62] and the maximization of likelihood function [63], lead to the same form of the cost function

$$L(\mathbf{W}) = -N \log(|\det(\mathbf{W})|) - \sum_{i=1}^{N} \sum_{k=1}^{K} \log(p_i(y_i(k))), \tag{5.3}$$

where $p_i(y_i)$ are the marginal probability density functions (pdfs) of output signals. Term $\det(\mathbf{W})$ represent the volume conserving property of the linear transformation [64]. ICA learning rule is derived by applying the stochastic gradient descent to (5.3)

$$\Delta\mathbf{W} = \mu \, ([\mathbf{W}^T]^{-1} - f(\mathbf{y})\mathbf{x}^T), \tag{5.4}$$

where $f(\mathbf{y})$ is the cumulative distribution function

$$f(y_i) = -\frac{dp_i(y_i)/dy_i}{p_i(y_i)}. \tag{5.5}$$

This update rule was first derived as *InfoMax* learning rule in [61] by maximization of entropy of transformed output signals. Selection of $f(\mathbf{y})$ as a non-linearity that approximately matches the input cdf's does not affect the performance of the algorithm. The uniform and robust convergence is obtained by using Amari's natural gradient [65], which has the simple form in the space of matrices, $\mathbf{W}\mathbf{W}^T$. Multiplying (5.4) by $\mathbf{W}\mathbf{W}^T$ leads to the learning rule without matrix inversion

$$\Delta\mathbf{W} = \mu\ [\mathbf{I} - f(\mathbf{y})\mathbf{y}^T]\mathbf{W} = \mu(\mathbf{W} - f(\mathbf{y})\mathbf{z}^T). \tag{5.6}$$

The convergence of (5.6) implies E$\{f_i(y_i)y_i\}$=1 as a constraint on the reconstructed signals. To avoid numerical instability due to non-stationarity in the sources, the Cichocki-Unbehauen (C-U) algorithm [66] introduces a non-holonomic constraint in the natural gradient learning rule (5.6), by fixing the diagonal terms of the unmixing matrix $\mathbf{W}$:

$$\Delta\mathbf{W} = \mu\ [\Lambda - f(\mathbf{y})\mathbf{y}^T]\mathbf{W}, \tag{5.7}$$

where Λ is a diagonal scaling matrix. Convergence of the C-U algorithm implies $\Lambda_{ii} = E[f(y_i)y_i]$.

The similar algorithm can be derived based on non-linear decorrelation that introduces higher-order statistics into the solution method. This first

formulation of ICA, inspired by biomimetic principles, was derived by Herault-Jutten (H-J) [67] and was based on a feedback network topology

$$\mathbf{y} = -\mathbf{W}\mathbf{y} + \mathbf{x}, \tag{5.8}$$

with zero diagonal terms ($w_{ii} \equiv 0, \forall i$). An independence criterion with non-linear correlation between output signals yields the on-line learning rule for the off-diagonal terms

$$\Delta w_{ij} = -\mu \; f(y_i)g(y_j), \qquad i \neq j \tag{5.9}$$

where $f(.)$ and $g(.)$ are appropriately chosen, odd-symmetric functions. Good example is the function that matches cumulative distribution function of source signals for f and linearity $g(y) \equiv y$ for function g.

Efficient implementation of ICA in feedforward parallel architecture requires casting learning rules (5.6) and (5.9) in a form of the outer-product update rule. H-J learning rule (5.9) is in the form of the outer-product, however the learning rule is defined for recurrent architecture (5.8). To map the recurrent architecture onto a feedforward form, the following approximation is applied:

$$\mathbf{y} = (\mathbf{I} + \mathbf{W})^{-1}\mathbf{x} \approx (\mathbf{I} - \mathbf{W})\mathbf{x} \tag{5.10}$$

In other words, the H-J rule can be implemented with linear feedforward network of the type $\mathbf{y} = \mathbf{W}\mathbf{x}$ with fixed diagonal terms $w_{ii} \equiv 1$, and with off-diagonal terms adapting according to (5.9) [53, 68]. Equivalently, the im-

plemented update rule can be seen as the gradient of *InfoMax* (5.4) multiplied by $\mathbf{W}^T$, rather than the natural gradient multiplication factor $\mathbf{W}^T\mathbf{W}$. Interestingly, in the special case of a 2 $\times$ 2 network (2 sources and 2 observations) the update rule (5.9) reduces to non-holonomic (zero-diagonal) form of the C-U rule (5.7).

To obtain the natural gradient update rule (5.6) in outer-product form, it is necessary to include a back-propagation path in the network architecture to implement the vector contribution $\mathbf{z} = \mathbf{W}^T\mathbf{y}$. The learning rule can then be represented in the form of the decay term and the outer-product term

$$\Delta w_{ij} = \mu\ w_{ij} - \mu\ f(y_i)z_j \tag{5.11}$$

Through quantization of the vector terms in the outer-product rules (5.9) and (5.11), as well as the quantization of the decay term in (5.11), the update rules in the proposed implementation are simplified to the charge transfer operations. In the case of speech signals that are approximately Laplacian distributed, the optimal nonlinear scalar function $f(y_i)$ can be approximated by $\text{sign}(y_i)$, which requires a single bit quantization. Conversely, vectors $\mathbf{y}$ in rule (5.9) and feedback signal $\mathbf{z} = \mathbf{W}^T\mathbf{y}$ in rule (5.11) are approximated by a 3-level staircase function $(-1, 0, +1)$ using 2-bit quantization denoted as $q(\mathbf{y})$ and $q(\mathbf{z})$. The quantization of the signal $\mathbf{y}$ and $\mathbf{z}$ in the update rules (5.9) and (5.11) simplifies the implementation of the update rule to a single-bit outer-product.

In this work, I implemented the natural gradient learning rule. The independent component analysis implementation comprises the vector-matrix

multiplication y = Wx and adaptation of the unmixing matrix coefficients according to learning rule (5.11). The block diagram of the proposed architecture for the implementation of the ICA algorithm is shown in Figure 5.1. In the Figure 5.1, the notation $< x_i >$ and $< y_i >$ denotes the pulse-width modulated signals controlled by the input signal x_i and the output signal y_i, respectively.

5.3 Circuit Implementation

The independent component analysis implementation comprises the vector-matrix multiplications $\mathbf{y} = \mathbf{W}\mathbf{x}$ and $\mathbf{z} = \mathbf{W}^T\mathbf{y}$ along with the adaptation of the unmixing matrix coefficients according to learning rule (5.6). There are two main circuit blocks that will be described, the adaptation cell and the voltage-to-time conversion circuit following the current integration.

5.3.1 Learning Rule Implementation

In the proposed implementation, the unmixing coefficients $\mathbf{W}_{ij}$ are stored differentially as voltages V_{ij}^+ and V_{ij}^- on two complementary switched current sources [69] as shown in Figure 5.2(a). For clarity of the Figure, I have omitted the replica of the current source M_0 that contributes to the i_i^- and i_j^- currents through the switches controlled by $< x_j^- >$ and $< y_i^- >$, respectively. The outer-product update rule (5.6) is implemented using two transistors with the functions $f(\mathbf{y})$ and $q(\mathbf{z})$ time encoded, as illustrated in Figure 5.2(b).

The proposed implementation enables fine updates of the unmixing coefficients with both positive and negative increments. The 3-level staircase

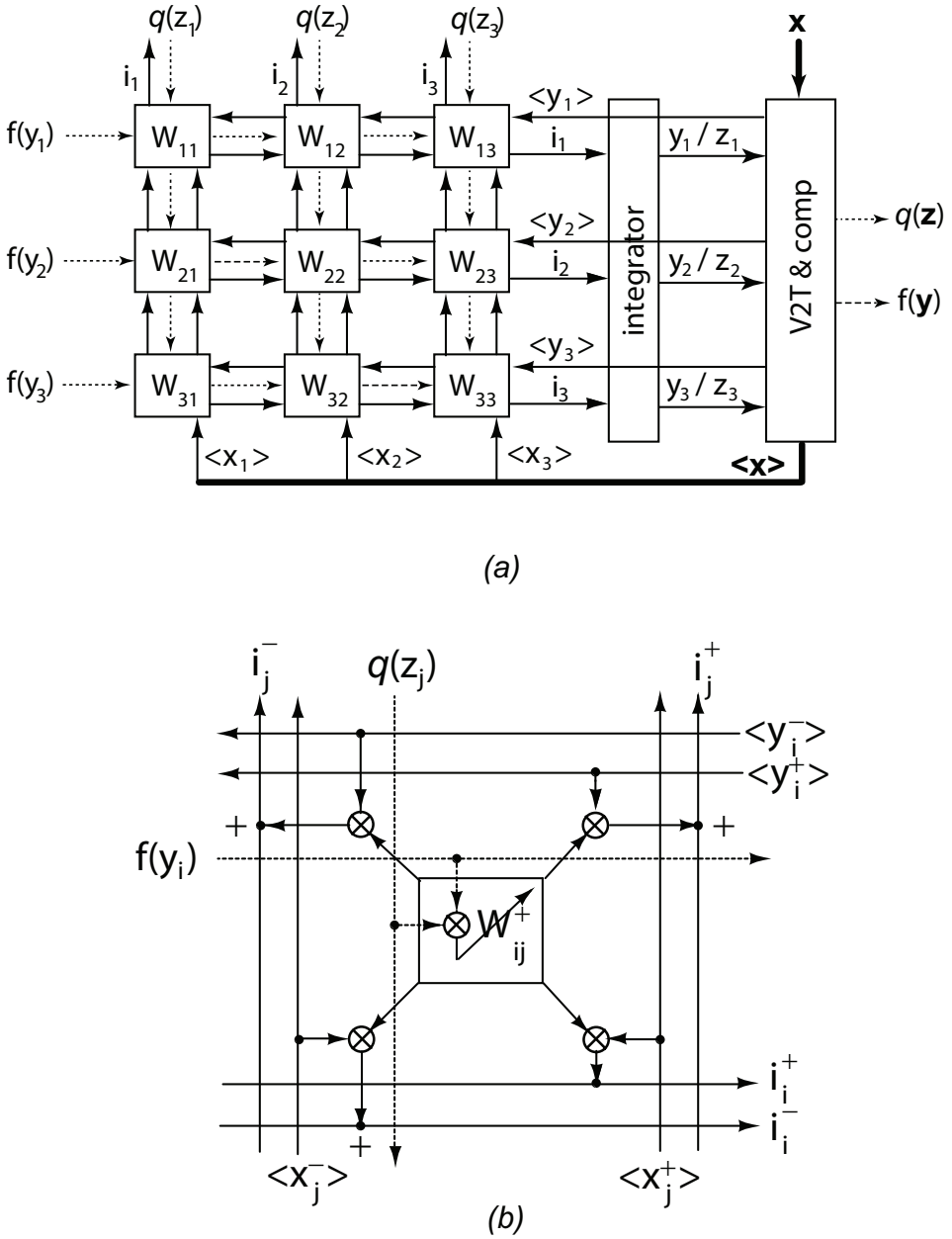

Figure 5.1: Block diagram of the architecture of the Independent Component Analyzer with 3x3 unmixing matrix.

function $q(\mathbf{z})$ is approximated with the presence/absence of the voltage pulse and by the relative position of the pulse. The function $f(\mathbf{y})$ is coded as a two-level signal, with the sign($\mathbf{y}$) determining the order of the levels V_{lo} and V_{hi}. These voltage levels are applied externally and control the value of the adaptation rate μ. To reduce the required silicon area the C_w is implemented as a MOS capacitance with the total capacitance of 2 pF. 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, denoted as C_p, and C_w is shared. The resulting voltage change on the capacitor C_w is given by

$$V_{ij}^{+}[n+1] = V_{ij}^{+}[n] + \frac{C_p}{C_w + C_p}(V_{Aij}^{+}[n] - V_{ij}^{+}[n]) \tag{5.12}$$

The common mode component $\frac{1}{2}(W_{ij}^{+} + W_{ij}^{-})$ is regulated by the weight decay term on the right side of (5.12), pulling the values towards the center of the range.

The effect of the charge injection and the clock feedthrough on the adaptation can be modeled as a constant offset plus the contribution that is dependant on the voltage V_{ij} which scales the decay term in the learning rule. The current source transistor M_0 is sized to operate in the subthreshold region of operation with the range of currents that represent W_{ij} from 100 pA to 100 nA. The advantage of the nonlinear transformation of stored voltages V_{ij} into a current representing the unmixing coefficient W_{ij} is a wide dynamic range of coefficients over a limited linear range of stored voltages.

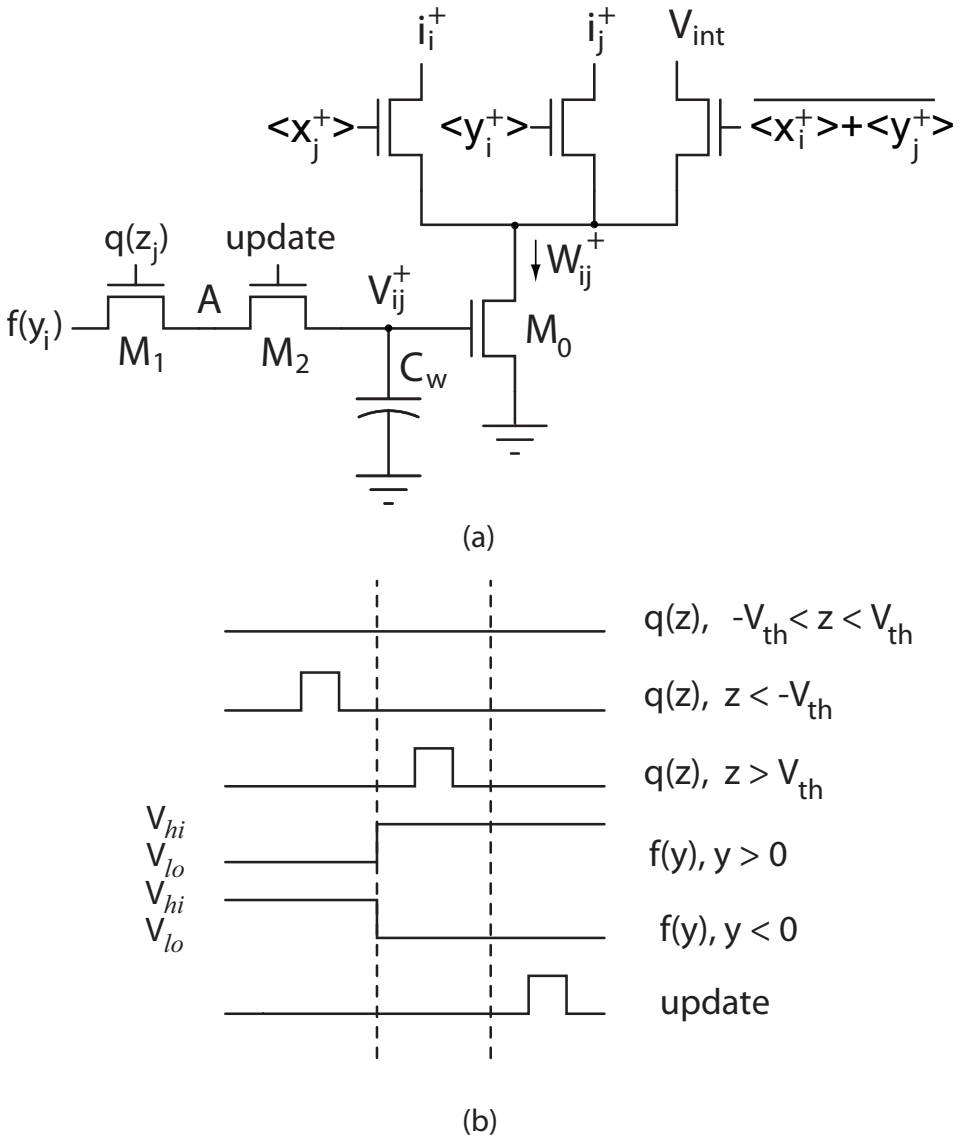

Figure 5.2: (a) Circuit implementation of the ICA learning rule. (b) Time-encoding of the functions $f(\mathbf{y})$ and $q(\mathbf{z})$ along with the timing of the update pulse.

5.3.2 Matrix-Vector Multiplication

The vector-matrix multiplications $\mathbf{y} = \mathbf{W}\mathbf{x}$ and $\mathbf{z} = \mathbf{W}^T\mathbf{y}$ are implemented by integrating switched currents controlled by a pulse-width modulated signal. To minimize the chip area, the two multiplications and the quantization of the signal $\mathbf{z}$ are implemented in three phases using the same the integration and voltage-to-time conversion circuitry. In the first phase, $\mathbf{y}$ is computed along with the voltage-to-time conversion of $\mathbf{x}$; in the second phase, $\mathbf{z}$ is computed along with the voltage-to-time conversion of $\mathbf{y}$ and in the third phase, the quantization of the signal $\mathbf{z}$ through voltage-to-time conversion is performed.

The implementation of the integration and voltage-to-time conversion is illustrated in Figure 5.3(a), with the corresponding clock timings of each switch shown in Figure 5.3(b). Clocks ϕ_1 and ϕ_2, as well as ϕ_5 and ϕ_6 are non-overlapping clocks. Both the input and the output signal are differential, as well as the coefficients of the mixing matrix:

$$y_i^+ = \sum_{j=1}^{3}(W_{ij}^+ x_j^+ + W_{ij}^- x_j^-) \tag{5.13}$$

$$y_i^- = \sum_{j=1}^{3}(W_{ij}^- x_j^+ + W_{ij}^+ x_j^-) \tag{5.14}$$

Current pulses are integrated on the capacitor C_{int} and the size of capacitor C_{int} is 2 pF. In the voltage-to-time converter, the input voltage signal precharges the integration capacitor C_t. The current fed into the input node of the inverting high-gain amplifier discharges the capacitor. A comparison of the decreasing voltage ramp signal at the output node of the amplifier with a reference voltage V_{comp} generates a pulsed signal with a pulse width pro-

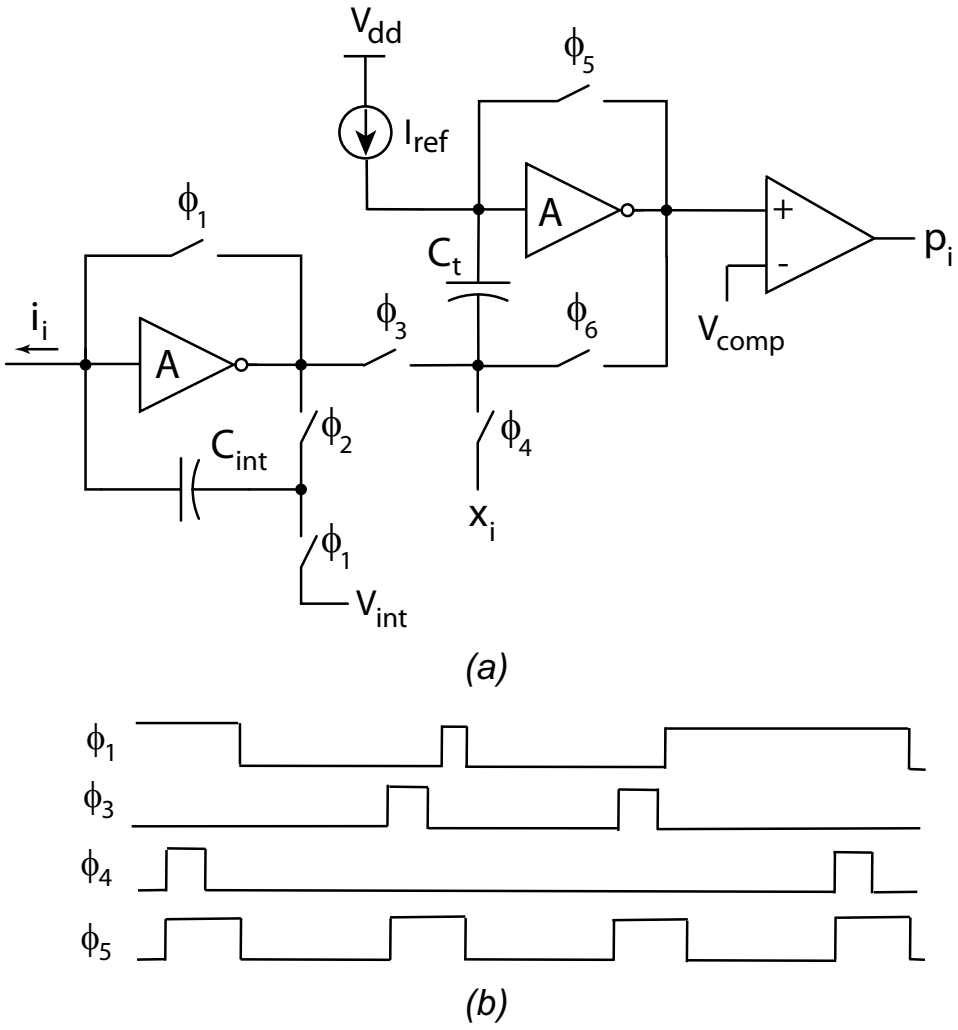

Figure 5.3: Integration of switched currents and voltage-to-time conversion circuitry.

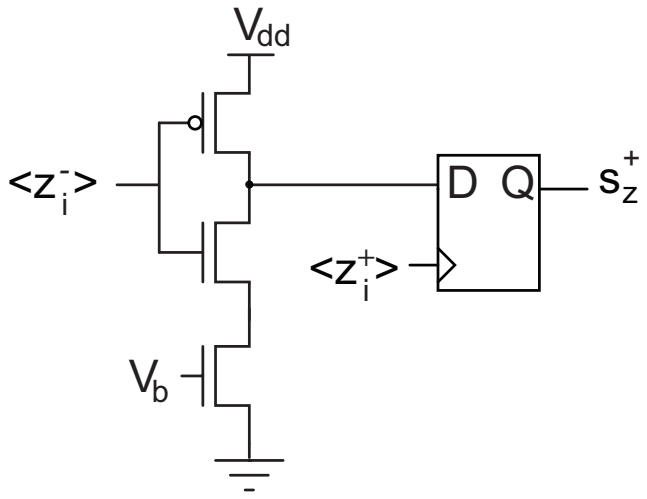

Figure 5.4: Implementation of the comparison of signal z_i with a threshold voltage for generation of the quantized signal $q(z_i)$.

portional to the input voltage. The high-gain amplifiers are implemented as cascoded amplifiers operating in sub-threshold region of operation with the input PMOS transistor. The integrator is also followed by the sample-and-hold circuit that holds the output signal $\mathbf{y}_i$.

As the pulse-width modulated output signals $< y_i^+ >$ and $< y_i^- >$ are available, with a single D-latch the sign of the y_i is determined. To generate the quantized signal $q(z_i)$, a comparison with a positive and a negative threshold voltage V_{th} is required. As in the case of the output signal $\mathbf{y}_i$, both pulse-width modulated z_i^+ and z_i^- signals are available. The comparison with the threshold voltage is performed by delaying one of these pulses before the connection to the input of the D-latch. Voltage V_b controls the threshold voltage by controlling the delay time [70]. In Figure 5.4, a single comparison of signal z_i with a threshold voltage is shown.

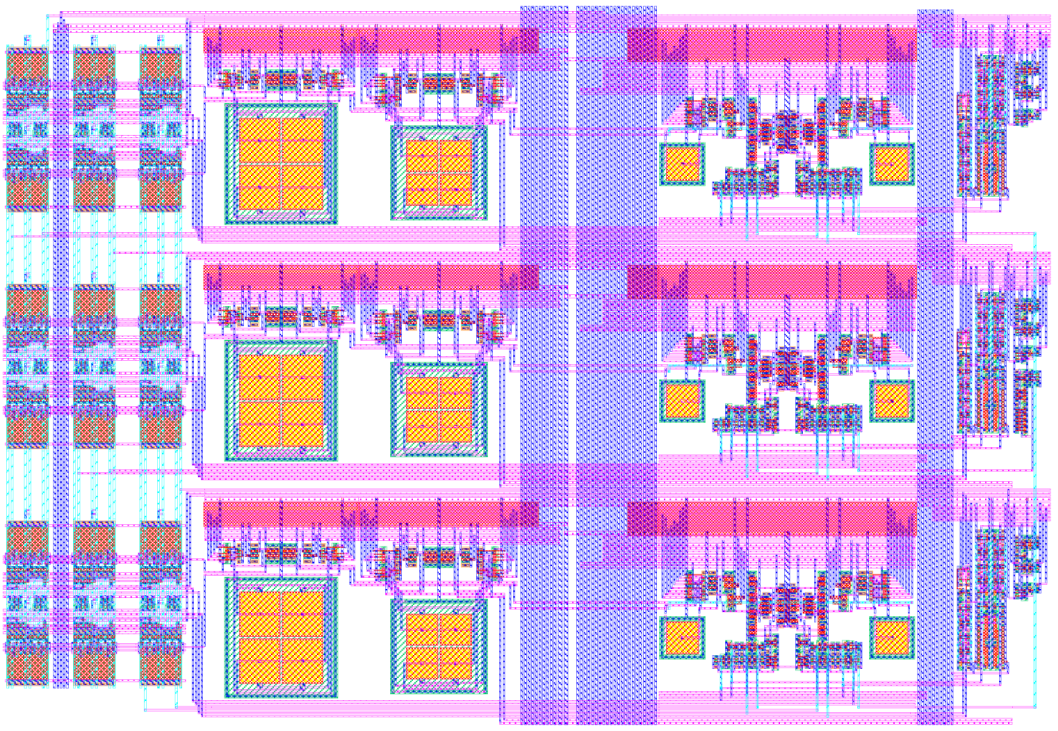

Figure 5.5: Layout of the proposed implementation in 0.5μm CMOS technology.

5.4 Simulation Results

The proposed architecture was implemented in 0.5μm 3M2P CMOS technology and the layout is shown in Figure 5.5. The total area of the 3x3 static ICA implementation is 0.49mm^2. The simulation of the circuit was performed on the extracted layout.

To demonstrate the adaptation process, the adaptation cell shown in Figure 5.2 has been simulated, with a constant sign of the update. The incremental values of the unmixing coefficient as the current of transistor M_0 are shown in Figure 5.7.

The output voltage of the integrator y_1 is shown in Figure 5.8 for three dif-

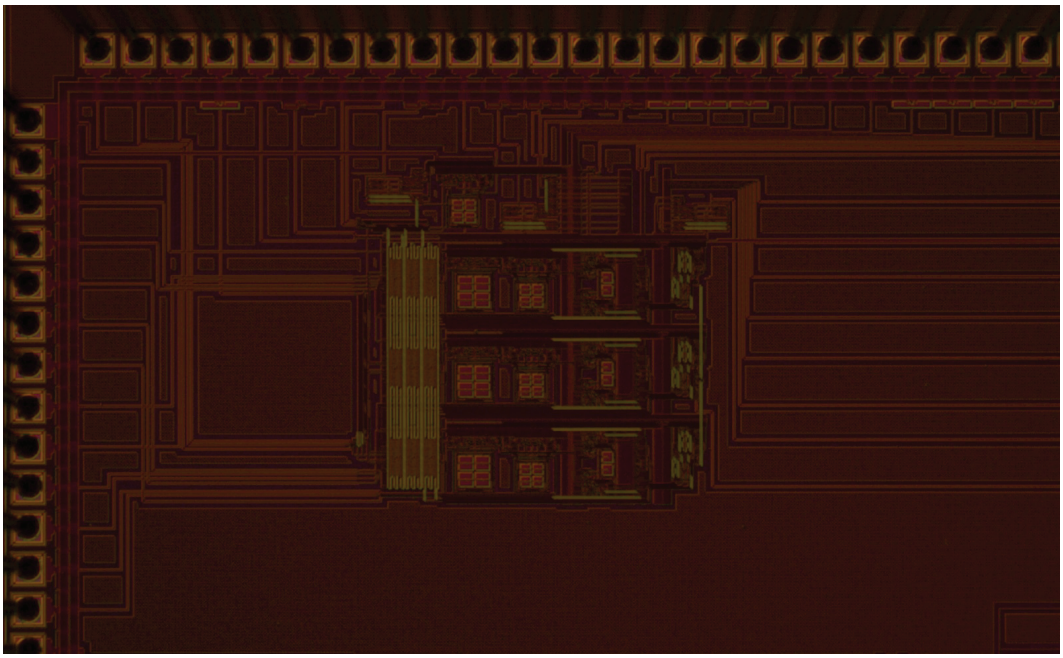

Figure 5.6: Micrograph of the proposed implementation in 0.5μm CMOS technology.

ferent values of the unmixing coefficient W_{11} while the other current sources representing unmixing coefficients are switched off. The input voltage x_1 is varied from 1 V to 4 V. The measured linearity of the matrix-vector multiplication is 0.05%.

The proposed implementation of the ICA algorithm for the acoustic source separation in hearing aid applications was modeled in MATLAB. To demonstrate the separation performance, the speech signals originating from two sources were artificially generated as received on the four microphone array with the distance between opposing microphone pairs set at 1 cm and with the sampling frequency of 16 kHz. The incidence angles of the two speech sources were 30^o and 70^o. Two first-order spatial gradient signals were obtained [51] and used as inputs to the model of the proposed ICA implementation. A white,

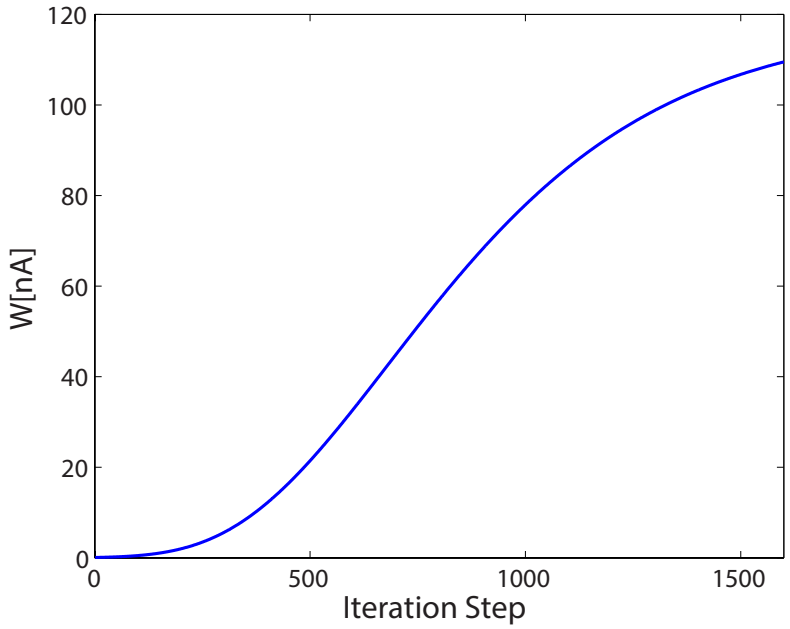

Figure 5.7: Successive incremental updates of the unmixing coefficient in a single direction.

spatially uncorrelated Gaussian noise sources were added to each sensor. The separation performance is quantized as the signal-to-interference ratio (SIR) in the output signals. The signal-to-interference ratio is computed as

$$SIR = -10 \log_{10} \min_i \frac{\sum_j < y_{ij}^2 > - \max_j < y_{ij}^2 >}{\max_j < y_{ij}^2 >}, \tag{5.15}$$

where y_{ij} is the contribution of the signal j to the output signal i. SIR, for different signal-to-noise ratio (SNR) in the sensor signals, is shown in Figure 5.9.

In Figure 5.10, the input and the output signals of the ICA processor, that is the spatial gradients (3.3) and the reconstructed source signals, are presented for incidence angles of θ_1=30^o and θ_2=105^o.

Table 5.1: ICA Processor Characteristics

Technology	0.5 μm 2P3M CMOS
Area	0.49mm^2
Supply	5 V
Room separation	8dB - 13dB
Power dissipation	80 μW at 16 kHz

5.5 Conclusion

I have presented an architecture and circuit implementation of an independent component analyzer for the use in a blind acoustic source separation microsystem using microphone array for hearing aid applications. The proposed pulse width modulation implementation allows a power and silicon area efficient application that can be used to realize multi-channel subband blind source separation and extended to other neural network applications. The measured characteristics of the fabricated chip are summarized in Table 5.1.

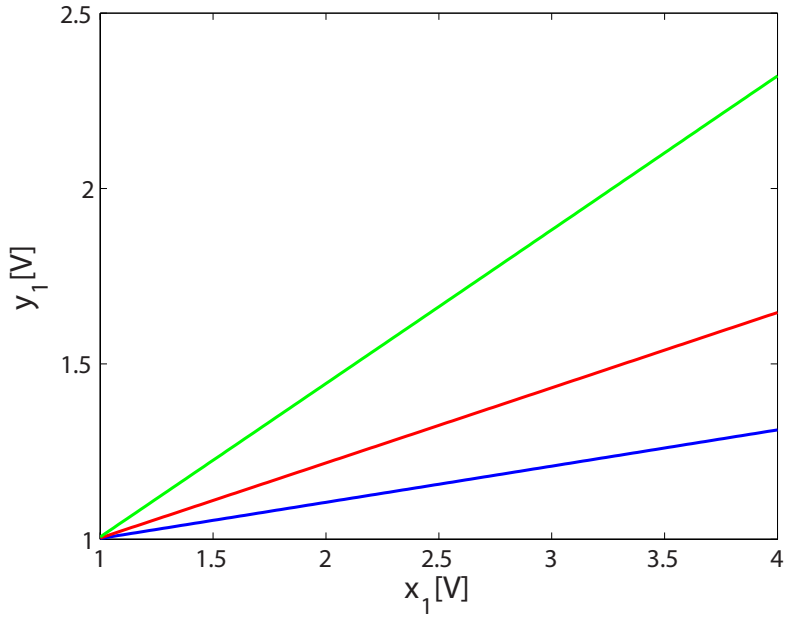

Figure 5.8: Linearity of the matrix-vector multiplication for three different values of the unmixing coefficient.

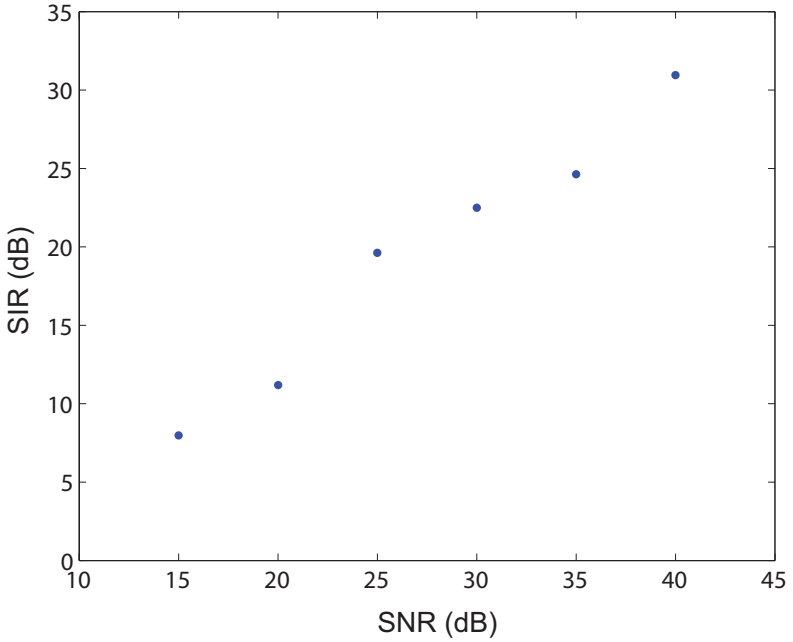

Figure 5.9: The separation performance expressed as SIR in the output signals for two incident speech signals on miniature microphone array.

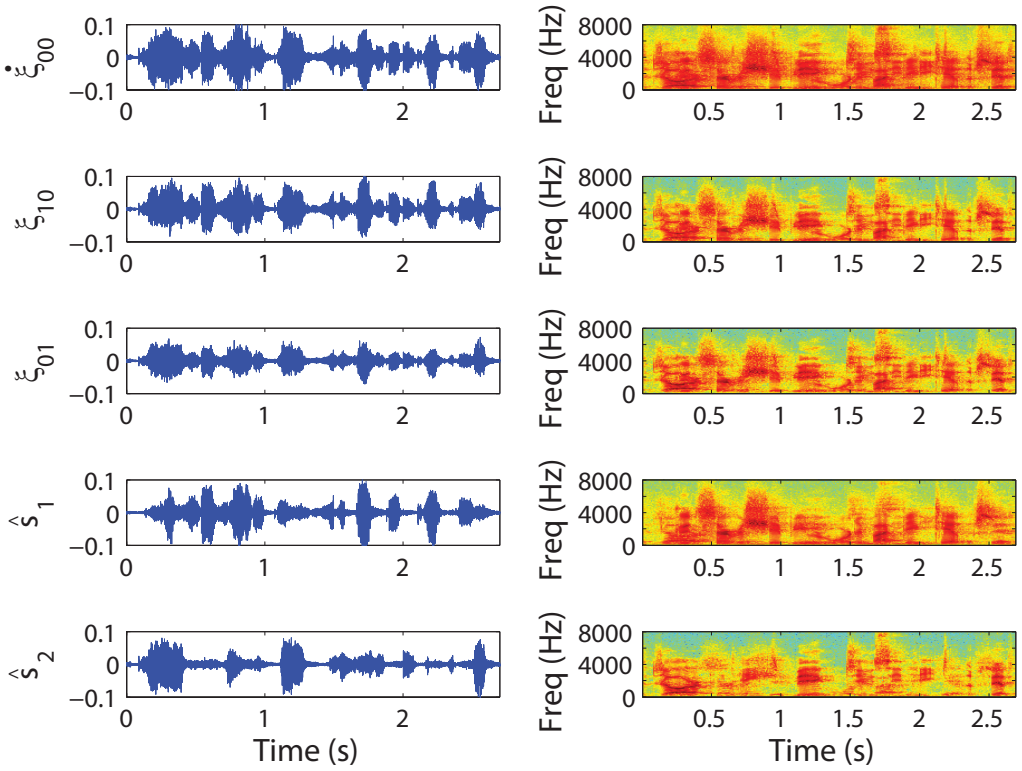

Figure 5.10: Time waveforms and spectrograms of spatial gradients (3.3), input signals to the ICA processor, and reconstructed source signals, output signals of the ICA processor for the case of two source signals with azimuth angles of θ_1=30^o and θ_2=105^o.

Chapter 6

Conclusions

6.1 Conclusion

In this work, I address the problem of acoustic source localization and separation. This problem is well-established with a long history and closely related to people's everyday lives. The 'cocktail party' problem is one of the most interesting yet intricate problems in the field of digital signal processing. Being able to localize and separate multiple sound sources with no priori information draws the interest from many researchers while few methodologies or frameworks are proposed under the premise of low-power, low-noise mixed-signal VLSI implementation.

I start from gradient flow, a technique that was inspired by both optical flow in image processing and biological observations on parasitoid fly. This technique can estimate the time delay of arrival by relating the spatial and temporal derivatives of a sensor array, avoiding the high sampling frequency and noise floor of traditional direct estimating methods. The technique can be

immune of interference noise from other sources when the aperture of the sensor array is shrinking, yet can be subject to the effect of additive noise which can be introduced from the sensor signal acquisition. A framework of generating synthetic sensor data under both anechoic environment and reverberant environment is proposed. Real world recordings in a room environment are made by using a custom made PCB interfacing the sensor array and the data acquisition card. The gradient flow technique is evaluated in a typical office room environment and simulated its performance under various noisy, interfering and reverberant environments. Results show the gradient flow localization can work robustly under mild to moderate reverberations with additive noise levels down to 10dB.

Then I move on to the acoustic source separation problem. While there are plenty of previous works on solving the problem in time-domain, frequency-domain or hybrid method combining two domains together, none of them fits into the low-power, low-noise mixed-signal VLSI implementation. I proposed my own subband gradient flow ICA architecture with 16 subbands. The subband methodology improves convergence under reverberant environment, while the time-domain ICA in each bank guarantees effective adaptation with mixed-signal circuits. The permutation and scaling ambiguity is solved by utilizing the static fullband ICA as a preprocessing technique to help alignment and better convergence, and processing with scaling solver to ensure uniform estimation across all the subbands. Simulation shows the proposed methodology produces around 13dB separation in SIR, which is a 3dB improvement over the static fullband ICA implementation.

Then I conduct a study on the effects that changing sensor array aper-

ture can have on localization and separation performances. Research shows increasing the inter-sensor spacing can produce diminishing errors in localization when there is only additive noise presented, but the error is kept at a similar level under reverberant environments. As separation is concerned, single band separation can still experience a better SIR in low frequency bands when the inter-microphone spacing is increasing. But when the results from multiple bands are combined, this slight advantage is submerged due to the lack of improvement in other bands. Therefore I am keeping the inter-sensor spacing to be 1cm with the consideration of better performance and small-form-factor of the device.

A micropower mixed-signal IC for localization of a broadband acoustic source has been designed. Continuous-time bandpass filtering and amplification at interfacing with microphone array enable low-power and small-form factor implementation. Time delays between acoustic signals observed over a planar geometry of four microphones are obtained by relating spatial and temporal differentials and estimated through mixed-signal least-square digital adaptation. The IC is fabricated in 0.5 μm CMOS technology and the dimension is 3mm $\times$ 3mm. At 16kHz sampling rate, the total power consumption of the proposed system-on-chip is estimated at total of 150 μW from a 5V supply, where 100 μW represents the power consumption of the bandpass filter, while 50 μW represents the power consumption of the spatial gradient computation, signal detection and bearing estimation.

Last but not least, an analog architecture for the implementation of independent component analyzer designed for the task of blind source separation of acoustic sources interfacing miniature microphone array is presented. The nat-

ural gradient ICA adaptation rule is selected because it's a great fit for analog implementation. The matrix-vector multiplication is implemented through integration of switched current sources, representing coefficients of the unmixing matrix, controlled by pulse-width modulated input signals. The adaptation is implemented by updating the charge pump at the gate of the NMOS transistor which controls the amount of current that is integrated. The proposed architecture implementing 3×3 static ICA in 0.5um CMOS technology occupies chip area of $0.49mm^2$ with the power consumption of 80uW at 5V supply voltage.

6.2 Future Directions

This work till now has constructed a solid model and framework for VLSI implementation for the acoustic source localization and separation. In the future, this work can still have some room to improve in the following area.

First, the usage of more microphones to improve the performance and flexibility of the architecture can be explored. Currently the structure of the array is four microphones with 1cm inter-microphone spacing in a orthogonal alignment. A prototype PCB adding four more microphones with 5.75mm spacing with a 45° rotation without affecting the size of the sensor array is constructed and illustrated in Figure 6.1. The PCB is tested using the four added microphones alone and the localization performance remains solid. There are two main reasons for adding more microphones to the array. One is more microphones are necessary for future researchers to exploit the possibility of involving second-order derivatives to expand the number of sources the archi-

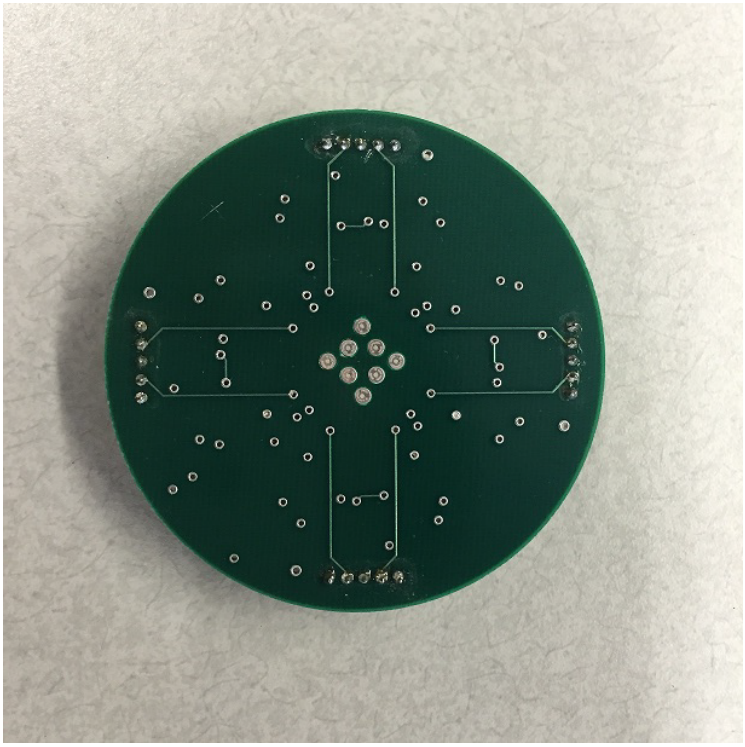

Figure 6.1: An eight microphone sensor array.

tecture can separate from three to five while also improving the localization accuracy. The other reason is that more microphones can be used to form a hybrid spatial derivatives so the SNR can be increased to alleviate the effect of additive noise when maintaining the device dimension.

Second, in this work, electret microphones are used because of their small dimensions and round shape in their package and the strict requirement on the array dimensions is what made this work unique. Yet, electret microphones suffer greatly from poor characteristic uniformity and their frequency response is uneven in both magnitude and phase which is very important to this work. Plenty of time must be spent on manually testing the frequency response of the microphones and handpicking matching microphones across the channels.

In the future, it is desirable to switch to MEMS microphones with equally small size and symmetrical shape because the MEMS microphones can have better uniformity across different units from the same batch and it can save researchers time on identifying matching microphones. Also more details on the characteristics of the MEMS microphones can be learned so a future integration of the sensor and the processing circuitry can be made as smooth as possible.

Last, an analog IC implementation of the independent component analyzer using current pulse-width modulation methodology is fabricated and tested in this work. Comparing to the previous mixed-signal IC implementation with D/A capacitor array for adaptation, the silicon area is significantly shrinked, thus it is desirable to fit the new IC into the 16-bank subband ICA architecture and integrate the filters and the independent component analyzers onto the same die. I also wish that the localizer circuit can be updated so that digital adaptation of the geometric cues can better help the subband ICA in permutation solving and unmixing matrix initialization. In the future, it will be great if the MEMS sensor array, the localizer and the independent component analyzer can be integrated into a single, small package and tested on hearing impaired people in real environment.

Bibliography

[1] Y.-H. Wu, "20q: Gaining new insights about directional hearing aid performance," *AudiologyOnline*, Feb 2013. Article 11617.

[2] D. Robert, "Innovative biomechanics for directional hearing in small flies," *The Biological Bulletin*, vol. 200, pp. 190 – 194, Apr 2001.

[3] D. Robert, R. Miles, and R. Hoy, "Tympanal hearing in the sarcophagid parasitoid fly emblemasoma sp.: the biomechanics of directional hearing," *J. Experimental Biology*, vol. 202, pp. 1865 – 1876, Jul 1999.

[4] M. Queiroz, F. Iazzetta, F. Kon, M. H. A. Gomes, F. L. Figueiredo, B. Masiero, L. K. Ueda, L. Dias, M. H. C. Torres, and L. F. Thomaz, "AcMus: an open, integrated platform for room acoustics research," *Journal of the Brazilian Computer Society*, vol. 14, no. 3, pp. 87–103, 2008.

[5] J. Duque-Carrillo, P. Malcovati, F. Maloberti, R. Perez-Aloe, A. Reyes, E. Sanchez-Sinencio, G. Torelli, and J. Valverde, "Verdi: an acoustically programmable and adjustable cmos mixed-mode signal processor for hearing aid applications," *Solid-State Circuits, IEEE Journal of*, vol. 31, pp. 634 –645, May 1996.

[6] F. Serra-Graells, L. Gomez, and J. Huertas, "A true-1V 300-μW CMOS-subthreshold log-domain hearing-aid-on-chip," *Solid-State Circuits, IEEE Journal of*, vol. 39, pp. 1271 – 1281, Aug 2004.

[7] M. Pedersen, J. Larsen, U. Kjems, and L. Para, "A survey of convolutive blind source separation methods," in *Springer Multichannel Speech Processing Handbook*, pp. 1065–1084, IEEE, 2007.

[8] R. Aichner, H. Buchner, S. Araki, and S. Makino, "On-line time-domain blind source separation of nonstationary convolved signals," in *4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, 2003.

[9] Z. Koldovsky and P. Tichavsky, "Time-domain blind audio source separation using advanced component clustering and reconstruction," in *Hands-Free*

Speech Communication and Microphone Arrays, 2008. HSCMA 2008, pp. 216–219, 2008.

[10] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing Journal*, vol. 22, pp. 21–34, 1998.

[11] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 2, pp. 109–116, 2003.

[12] T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind source separation of acoustic signals based on multistage ICA combining frequency-domain ICA and time-domain ICA," *IEICE Trans. Fundamentals.*, vol. E86-A, pp. 846–858, Apr 2003.

[13] K. Kokkinakis and P. Loizou, "Subband-based blind signal processing for source separation in convolutive mixtures of speech," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 4, pp. 917 – 920, Apr 2007.

[14] S. Ikeda and N. Murata, "A method of ICA in time-frequency domain," in *Proc. Int. Workshop on Independent Component Analysis and Signal Separation*, pp. 365–371, 1999.

[15] M. Konishi, "Study of sound localization by owls and its relevance to humans," *Comparative Biochemistry and Physiology Part A*, vol. 126, pp. 459–469, Aug 2000.

[16] R. N. Miles, D. Robert, and R. Hoy, "Mechanically coupled ears for directional hearing in the parasitoid fly ormia ochracea," *J. Acoust. Soc. Am.*, vol. 98, no. 6, pp. 3059–3070, 1995.

[17] R. N. Miles, Q. Su, W. Cui, M. Shetye, F. 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. Amer.*, vol. 125, no. 5, pp. 2013–2026, 2009.

[18] C. Zhang, D. Florencio, D. Ba, and Z. Zhang, "Maximum likelihood sound source localization and beamforming for directional microphone arrays in distributed meetings," *IEEE Transactions on Multimedia*, vol. 10, no. 3, pp. 538–548, 2008.

[19] D. Levin, E. A. P. Habets, and S. Gannot, "Maximum likelihood estimation of direction of arrival using an acoustic vector-sensor," *The Journal of the Acoustical Society of America*, vol. 131, no. 2, pp. 1240–1248, 2012.

[20] T. May, S. van de Par, and A. Kohlrausch, "A probabilistic model for robust localization based on a binaural auditory front-end," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 1, pp. 1–13, 2011.

[21] W. Zhang and B. Rao, "A two microphone-based approach for source localization of multiple speech sources," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, pp. 1913–1928, Nov 2010.

[22] M. Fallon and S. Godsill, "Acoustic source localization and tracking using track before detect," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, pp. 1228–1242, Aug 2010.

[23] M. Raspaud, H. Viste, and G. Evangelista, "Binaural source localization by joint estimation of ild and itd," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, pp. 68–77, Jan 2010.

[24] B. Mungamuru and P. Aarabi, "Enhanced sound localization," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 34, pp. 1526–1540, Jun 2004.

[25] S.-Y. Lee and H.-M. Park, "Multiple reverberant sound localization based on rigorous zero-crossing-based itd selection," *Signal Processing Letters, IEEE*, vol. 17, pp. 671–674, Jul 2010.

[26] W.-J. Zeng and X.-L. Li, "High-resolution multiple wideband and nonstationary source localization with unknown number of sources," *Signal Processing, IEEE Transactions on*, vol. 58, pp. 3125–3136, Jun 2010.

[27] F. Ribeiro, C. Zhang, D. A. Florencio, and D. Ba, "Using reverberation to improve range and elevation discrimination for small array sound source localization," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, pp. 1781–1792, Sept 2010.

[28] D. Halupka, N. Mathai, P. Aarabi, and A. Sheikholeslami, "Robust sound localization in 0.18 μm CMOS," *IEEE Transactions on Signal Processing*, vol. 53, no. 6, pp. 2243–2250, 2005.

[29] Y. Huang, J. Benesty, G. Elko, and R. Mersereau, "Real-time passive source localization: A practical linear-correction least-squares approach," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, pp. 943–956, 2001.

[30] B. Champagne, S. Bedard, and A. Stephenne, "Performance of time-delay estimation in the presence of room reverberation," *IEEE Transactions on Speech and Audio Processing*, vol. 4, no. 2, pp. 148–152, 1996.

[31] T. Gustafsson, B. Rao, and M. Trivedi, "Source localization in reverberant environments: Modeling and statistical analysis," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 791–803, 2003.

[32] G. Cauwenberghs, M. Stanacevic, and G. Zweig, "Blind broadband source localization and separation in miniature sensor arrays," in *Circuits and Systems, 2001. ISCAS 2001. The 2001 IEEE International Symposium on*, vol. 3, pp. 193 –196 vol. 2, May 2001.

[33] M. Stanaćević, G. Cauwenberghs, and G. Zweig, "Gradient flow adaptive beamforming and signal separation in a miniature microphone array," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 4, pp. IV–4016–IV–4019, 2002.

[34] M. Stanaćević and G. Cauwenberghs, "Micropower gradient flow vlsi acoustic localizer," *IEEE Transactions on Circuits and Systems I : Regular Papers*, vol. 52, no. 10, pp. 2148–2157, 2005.

[35] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.

[36] H. Kuttruff, *Room Acoustics*. New York: Wiley, 1973.

[37] S. Li and M. Stanaćević, "Gradient flow source localization in noisy and reverberant environments," in *Signals, Systems and Computers (ASILOMAR), 2012 Conference Record of the Forty Sixth Asilomar Conference on*, pp. 257–260, 2012.

[38] R. Lambert and A. Bell, "Blind separation of multiple speakers in a multipath environment," in *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, vol. 1, pp. 423 –426 vol.1, Apr 1997.

[39] 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, no. 5, pp. 530–538, 2004.

[40] 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, no. 2, pp. 666–678, 2006.

[41] L. Parra and C. Alvino, "Geometric source separation: Merging convolutive source separation with geometric beamforming," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 6, pp. 352–362, 2002.

[42] 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," *Speech and Audio Processing, IEEE Transactions on*, vol. 12, no. 5, pp. 530–538, 2004.

[43] H.-M. Park, C. Dhir, S.-H. Oh, and S.-Y. Lee, "A filter bank approach to independent component analysis for convolved mixtures," *Neurocomputing*, vol. 69, no. 16-18, pp. 2065–2077, 2006.

[44] K. Nayebi, T. Barnwell, and M. J. T. Smith, "Time-domain filter bank analysis: a new design theory," *Signal Processing, IEEE Transactions on*, vol. 40, no. 6, pp. 1412–1429, 1992.

[45] 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, no. 5, pp. 1265–1277, 2006.

[46] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 14, no. 4, pp. 1462–1469, 2006.

[47] S. Li and M. Stanaćević, "Subband gradient flow acoustic source separation for moderate reverberation environment," in *Signals, Systems and Computers (ASILOMAR), 2012 Conference Record of the Forty Sixth Asilomar Conference on*, pp. 253–256, 2012.

[48] J. Lazzaro and C. Mead, "A silicon model of auditory localization," *Neural Computation*, vol. 1, pp. 47–57, 1989.

[49] P. Julian, P. Andreou, P. Mandolesi, and D. Goldberg, "A low-power CMOS integrated circuit for bearing estimation," *Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS'2003)*, vol. 5, pp. 305–308, May 2003.

[50] C. Enz and G. Temes, "Circuit techniques for reducing the effects of op-amp imperfections: Autozeroing, correlated double sampling, and chopper stabilization," *Proceedings of the IEEE*, vol. 84, pp. 1584–1614, Nov 1996.

[51] S. Li, Y. Lin, and M. Stanacevic, "Mixed-signal vlsi microsystem for acoustic source separation," in *Circuits and Systems (MWSCAS), 2013 IEEE 56th International Midwest Symposium on*, pp. 1176–1179, Aug 2013.

[52] M. Cohen and A. Andreou, "Analog cmos integration and experimentation with an autoadaptive independent component analyzer," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 42, pp. 65–77, Feb 1995.

[53] A. Celik, M. Stanacevic, and G. Cauwenberghs, "Mixed-signal real-time adaptive blind source separation," in *Circuits and Systems, 2004. ISCAS '04. Proceedings of the 2004 International Symposium on*, vol. 5, pp. V–760–V–763 Vol.5, May 2004.

[54] 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," *Neural Networks, IEEE Transactions on*, vol. 19, pp. 958–970, Jun 2008.

[55] L.-D. Van, D.-Y. Wu, and C.-S. Chen, "Energy-efficient fastica implementation for biomedical signal separation," *Neural Networks, IEEE Transactions on*, vol. 22, pp. 1809–1822, Nov 2011.

[56] K. Papathanasiou, T. Brandtner, and A. Hamilton, "Palmo: pulse-based signal processing for programmable analog vlsi," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 49, pp. 379–389, Jun 2002.

[57] J.-C. Bor and P. Chung-Yu Wu, "Realization of the cmos pulsewidth-modulation (pwm) neural network with on-chip learning," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 45, pp. 96–107, Jan 1998.

[58] A. Hyvrinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Wiley-Interscience, 2001.

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

[60] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287 – 314, 1994. Higher Order Statistics.

[61] A. Bell and T. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129 – 1159, 1995.

[62] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems*, pp. 757–763, MIT Press, 1996.

[63] J.-P. Nadal and N. Parga, "Non linear neurons in the low noise limit: A factorial code maximizes information transfer," *Network*, vol. 5, pp. 565 – 581, 1994.

[64] D. Obradovic and G. Deco, "Information maximization and independent component analysis: Is there a difference?," *Neural Computation*, vol. 10, no. 8, pp. 2085–2101, 1998.

[65] A. Cichocki, R. Unbehauen, L. Moszcnski, and E. Rummert, "A new on-line adaptive learning algorithm for blind separation of sources," *Int. Symp. Artificial Neural Networks ISANN-94*, vol. 10, pp. 406–411, Dec 1994.

[66] A. Cichocki and R. Unbehauen, "Robust learning algorithm for blind separation of signals," *Electronics Letters*, vol. 30, pp. 1386–1387, Aug 1994.

[67] C. Jutten and J. Herault, "Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, no. 1, pp. 1 – 10, 1991.

[68] A. Celik, M. Stanacevic, and G. Cauwenberghs, "Gradient flow independent component analysis in micropower vlsi," in *Advances in Neural Information Processing Systems 18* (Y. Weiss, B. Schlkopf, and J. Platt, eds.), pp. 187–194, Cambridge, MA: MIT Press, 2005.

[69] M. Stanacevic and G. Cauwenberghs, "Charge-based cmos fir adaptive filter," in *Circuits and Systems, 2000. Proceedings of the 43rd IEEE Midwest Symposium on*, vol. 3, pp. 1410–1413 vol.3, 2000.

[70] M. Maymandi-Nejad and M. Sachdev, "A monotonic digitally controlled delay element," *Solid-State Circuits, IEEE Journal of*, vol. 40, pp. 2212–2219, Nov 2005.