

Far-Field Acoustic Source Localization and Bearing Estimation Using $\Sigma\Delta$ Learners

Amit Gore, Amin Fazel, *Student Member, IEEE*, and Shantanu Chakrabartty, *Senior Member, IEEE*

Abstract—Localization of acoustic sources using miniature microphone arrays poses a significant challenge due to fundamental limitations imposed by the physics of sound propagation. With sub-wavelength distances between the microphones, resolving acute localization cues become difficult due to precision artifacts. In this paper we propose a framework which overcomes this limitation by integrating signal-measurement (analog-to-digital conversion) with statistical learning (bearing estimation). At the core of the proposed approach is a min-max stochastic optimization of a regularized cost function that embeds manifold learning within $\Sigma\Delta$ modulation. As a result, the algorithm directly produces a quantized sequence of the bearing estimates whose precision can be improved asymptotically similar to a conventional $\Sigma\Delta$ modulators. In this paper we present a hardware implementation of a miniature acoustic source localizer which comprises of: (a) a common-mode canceling microphone array and (b) a $\Sigma\Delta$ integrated circuit which produces bearing parameters. The parameters are then combined in an estimation procedure that can achieve a linear range from 0° – 90° . Measured results from a prototype fabricated in a $0.5\ \mu\text{m}$ CMOS process demonstrate that the proposed localizer can reliably estimate the bearing of an acoustic source with a resolution less than 2° while consuming less than $75\ \mu\text{W}$ of power.

Index Terms—Analog-to-information converter, bearing estimation, high-dimensional array processing, microphone arrays, on-chip learning, oversampling converters, source localization, $\Sigma\Delta$ modulation.

I. INTRODUCTION

ONE of several emerging areas where micro/nano-scale integration promises significant breakthroughs is in the field of acoustic sensing. It is envisioned that next generation of intelligent hearing devices will integrate hundreds of micro/nano-scale microphones [1], separate speech from noise by localizing different acoustic sources. It is also envisioned that surveillance robots and sensor networks will utilize the microphone array to remotely localize, track and identify speakers of interest [2]–[4]. However, localization of acoustic sources using miniature microphone arrays poses a significant challenge due to fundamental limitations imposed by the physics of sound propagation [1], [5]. The smaller the distance between

the recording elements, the more difficult it is to measure localization cues [6], [7] which include: (a) inter-aural temporal difference (ITD) between the time-of-arrival of the acoustic wavefront at two recording elements (microphones); and (b) inter-aural level difference (ILD) between sound amplitude at different recording elements. For instance, resolving ITD cues at 8 bits resolution using microphones placed at a distance of 1 mm would require detecting 10 ns delays. However, since the position of the acoustic source varies slowly (typically less than 1 Hz) the localization cues can be inferred by measuring the correlation between the microphone signals. Source localization with correlated signals have been reported by several groups [8], [9] based on the coherent as well as the non-coherent methods. The non-coherent methods such as closest point of approach (CPA) [9] are not sensitive to synchronization but suffers from the sensor mismatch and the sensor response due to a distance between the source and the sensor. In contrast to non-coherent methods, the coherent methods poses stringent constraints on synchronization as they are based on the arrival time differences of the acoustic signal to the sensors. The coherent methods make use of signal correlation in estimating the bearing of the sound source with respect to the miniaturized microphone array. Examples of coherent techniques include algorithms based on the cross correlation (CA) [10]–[13], cross correlation derivative (CDA) [14]–[16], stereausis networks [17], [18] and matched filtering techniques [19]. Unfortunately, most of these algorithms are feed-forward in nature which decouple the signal measurement (analog-to-digital conversion) with the learning (bearing estimation) process. As a result, the variance of estimation is either limited by analog VLSI artifacts (mismatch or noise) or by quantization effects of the analog-to-digital converter.

One of the ways to improve the resolution of bearing estimation and alleviating effects of analog artifacts is to integrate learning algorithms directly with the signal/sensor measurement (analog-to-digital conversion) process. In fact, a similar principle is observed in nature, where the remarkable localization ability of the parasitoid fly *Ormia ochracea* (accuracy better than 2 degrees) [1], [20], [21] is attributed to coupling between its miniature differential eardrums (separated by less than 0.5 mm) and its neuronal circuitry (which acts as an equivalent ADC). This biological prototype has inspired neuromorphic architectures which includes the “gradient-flow” localization [22] technique that relies on a least mean square (LMS) learning technique method that processes acoustic gradients recorded by a differential microphone array. In this paper, we present a source localization technique that integrates bearing estimation (learning) within a $\Sigma\Delta$ modulation procedure. Traditionally, $\Sigma\Delta$ modulators have been the architecture of choice for any

Manuscript received December 08, 2008; revised April 22, 2009. First published November 03, 2009; current version published April 09, 2010. This work was supported in part by the National Institute of Health under Grant R21NS047516-01A2 and in part by the National Science Foundation under Grant IIS:0836278. This paper was recommended by Associate Editor Y. Lian.

A. Gore is with General Electric Global Research, Niskayuna, NY 12309 USA.

A. Fazel and S. Chakrabartty are with the Electrical and Computer Engineering Department, Michigan State University, East Lansing, MI 48823 USA (e-mail: shantanu@egr.msu.edu).

Digital Object Identifier 10.1109/TCSI.2009.2027627

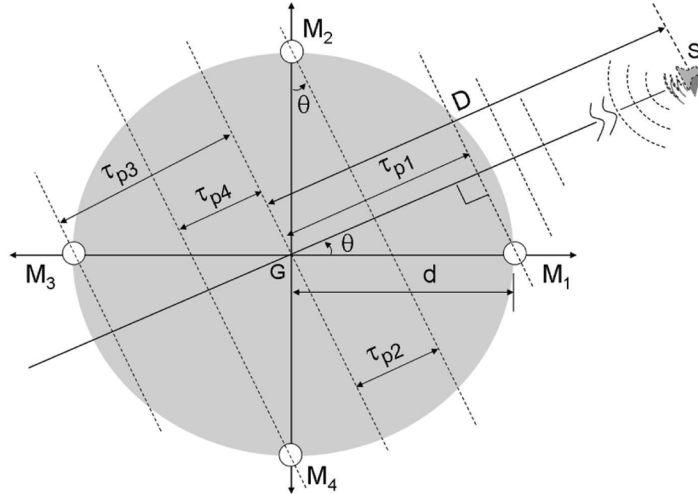


Fig. 1. Far field acoustic model on a 2-D four-microphone array.

audio based processing because of its robustness to analog imperfections allowing it to achieve very high resolution [23]. $\Sigma\Delta$ modulators also share significant similarities with neuronal circuits in terms of their robustness and noise-shaping capabilities allowing them to resolve acute changes in signal levels [24]. We have formulated $\Sigma\Delta$ modulation within the framework of statistical learning such that the algorithm tracks the bearing manifold which captures the direction of arrival of acoustic wavefront. A prototype $\Sigma\Delta$ source localizer has been fabricated in a standard CMOS process and has been interfaced with a differential microphone array. The architecture of the resulting $\Sigma\Delta$ localizer is shown in Fig. 2 which consists of an analog projection (manifold learning) operator whose parameters are adapted based on the output of a $\Sigma\Delta$ modulator array. The adapted manifold parameters are then combined together to estimate the bearing of the far-field acoustic source. We will show that this estimation procedure is robust to analog artifacts and improves the dynamic range compared to a simple correlation based bearing estimation.

In Section II, we describe a far-field acoustic localization model that formulates bearing estimation as an equivalent linear regression problem. Section III describes the $\Sigma\Delta$ learning algorithm which is used for source localization. Section IV describes the circuit level implementation of a differential microphone array and an integrated circuit implementation of a $\Sigma\Delta$ localizer. In Section V we present measured results obtained using the differential microphone array and a fabricated prototype of the $\Sigma\Delta$ learner. Section VI concludes the paper with some final remarks and future directions.

II. FAR FIELD SOURCE LOCALIZATION

Far-field acoustics has been extensively studied within the context of array processing plenacoustic models [25], [22]. In this section we describe a simplistic model which is applicable for source localization using miniature microphone array. Consider a sensor array shown in Fig. 1 that consists of four recording elements. If the inter-element distance is much less than the wavelength of the sensor signal of interest, the signals recorded at each of the sensor elements can be approximated

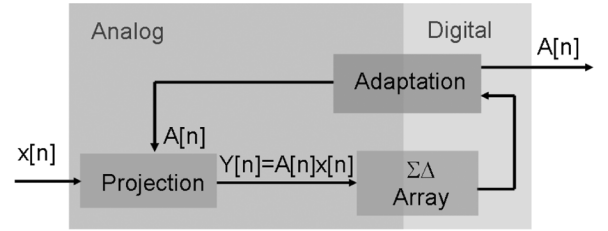


Fig. 2. Architecture of the $\Sigma\Delta$ learner.

using a far-field wave propagation model [25] where the acoustic wavefront can be assumed to planar (see Fig. 1). For acoustic signals with a frequency range of 100 Hz–20 kHz, this distance is typically less than $\lambda/10 \approx 3.4$ cm, where λ is the wavelength of the audio signal. Also, the distance to the source from the center G of the array is assumed to be larger than the inter-element distance. Therefore, the acoustic signal wavefront is considered planar as it reaches the microphone array as shown in Fig. 1. The signal $x(\mathbf{p}_j, t)$ recorded at the j th microphone (located at the position vector, $\mathbf{p}_j = (x, y, z)$ with respect to the center G of the array) can be expressed as a function of the bearing θ which is the angle between the position vector \mathbf{p}_j and the source vector \mathbf{u} . The signal $x(\mathbf{p}_j, t)$ is written as

$$x(\mathbf{p}_j, t) = a(\mathbf{p}_j)s(t - \tau(\mathbf{p}_j)) \quad (1)$$

where $a(\mathbf{p}_j)$ and $\tau(\mathbf{p}_j)$ denotes the attenuation and delay for the source $s(t)$ respectively, measured relative to the center of the microphone array. Equation (1) is expanded using Taylor's series as

$$x(\mathbf{p}_j, t) = a(\mathbf{p}_j) \sum_{k=0}^{\infty} \frac{(-\tau(\mathbf{p}_j))^k}{k!} s^{(k)}(t). \quad (2)$$

Under far-field conditions, $a(\mathbf{p}_j)$ is approximately constant and without loss of generality $a(\mathbf{p}_j) \approx 1$. Also, for far-field conditions the time delays $\tau(\mathbf{p}_j)$ can be assumed to be a linear projection between the vector \mathbf{p}_j and the unit vector \mathbf{u} oriented to-

wards the direction of arrival of the acoustic wavefront. Mathematically this is modeled as

$$\tau(\mathbf{p}_j) = \frac{1}{c} \mathbf{u}^T \cdot \mathbf{p}_j \quad (3)$$

where c is the speed of sound wave propagation. Ignoring the higher-order terms in the series expansion (under far-field assumptions), (2) is expressed as

$$x(\mathbf{p}_j, t) \approx s(t) - \tau(\mathbf{p}_j) \dot{s}(t) \quad (4)$$

implying that under far-field conditions, the signals recorded at the microphone array is linear with respect to the bearing parameter $\tau(\mathbf{p}_j)$.

In the proposed implementation the microphone array used for bearing estimation consists of four recording elements $j = 1, \dots, 4$. Thus, the recorded signals can be explicitly written as

$$\begin{aligned} x_{p1}(t) &= s(t - \tau_{p1}) = s(t) - \dot{s}(t) \frac{d}{c} \cos \theta \\ x_{p2}(t) &= s(t - \tau_{p2}) = s(t) - \dot{s}(t) \frac{d}{c} \sin \theta \\ x_{p3}(t) &= s(t + \tau_{p3}) = s(t) + \dot{s}(t) \frac{d}{c} \cos \theta \\ x_{p4}(t) &= s(t + \tau_{p4}) = s(t) + \dot{s}(t) \frac{d}{c} \sin \theta \end{aligned} \quad (5)$$

It can be easily verified that

$$x_{\text{cm}}(t) = \frac{x_{p1}(t) + x_{p2}(t) + x_{p3}(t) + x_{p4}(t)}{4} = s(t) \quad (6)$$

is the common-mode signal which is independent of the bearing angle θ . Therefore, a more desirable output from the microphone array are differential signals which are devoid of the common-mode factors $s(t)$ resulting in

$$\begin{aligned} \Delta x_1(t) &= -\dot{s}(t) \frac{d}{c} \cos \theta \\ \Delta x_2(t) &= -\dot{s}(t) \frac{d}{c} \sin \theta \\ \Delta x_3(t) &= +\dot{s}(t) \frac{d}{c} \cos \theta \\ \Delta x_4(t) &= +\dot{s}(t) \frac{d}{c} \sin \theta. \end{aligned} \quad (7)$$

Equation (7) shows that for an ideal differential microphone array (perfect matching of the microphones), the output signals are shifted in phase by 90° , 180° and 270° . In the next section we will show how these differential signals can be processed by a $\Sigma\Delta$ learner to estimate the bearing of the acoustic source.

III. $\Sigma\Delta$ LEARNING AND SOURCE LOCALIZATION

$\Sigma\Delta$ learning is based on an optimization framework that integrates $\Sigma\Delta$ modulation with statistical learning. A special case of the proposed framework was introduced in [26] where it was used for neural signal compression. Given a random input vector $\mathbf{x} \in \mathcal{R}^D$ and an internal state vector $\mathbf{w} \in \mathcal{R}^M$, a $\Sigma\Delta$ learner estimates the parameters of a linear transformation matrix $\mathbf{A} \in \mathcal{R}^D \times \mathcal{R}^M$ according to the following optimization criterion:

$$\max_{\mathbf{A} \in \mathcal{C}} (\min_{\mathbf{w}} f(\mathbf{w}, \mathbf{A})) \quad (8)$$

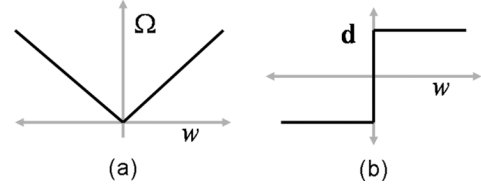


Fig. 3. (a) 1-D L_1 norm and (b) its derivative.

where

$$f(\mathbf{w}, \mathbf{A}) = \|\mathbf{w}\|_1 - \mathbf{w}^T \mathcal{E}_{\mathbf{x}}\{\mathbf{A}\mathbf{x}\}. \quad (9)$$

$\mathcal{E}_{\mathbf{x}}\{\cdot\}$ denotes an expectation operator with respect to the random vector \mathbf{x} . The term $\|\mathbf{w}\|_1$ bears similarity to the regularization which is extensively used in machine learning algorithms [27], [28]. However, the L_1 norm in (9) forms an important link in connecting the cost function to single bit quantizers. This is illustrated in Fig. 3 which shows an example of a 1-D regularization function $\|\mathbf{w}\|_1$. The piece-wise behavior of $\|\mathbf{w}\|_1$ leads to discontinuous gradient $\text{sgn}(\mathbf{w})$ (shown in Fig. 3(b)) where $\text{sgn}(\cdot)$ denotes a signum operation equivalent to a single bit quantization. Even though the framework can be extended to multi-bit quantization, in this paper we will only discuss a single bit operator. The minimization step in (8) ensures that the state vector \mathbf{w} is correlated with the transformed input signal $\mathbf{A}\mathbf{x}$ (signal tracking step) and the maximization step in (8) adapts the parameters of \mathbf{A} such that it minimizes the correlation (de-correlation step). The formulation bears similarities with game-theoretic approaches [29], [30] where signal tracking and de-correlation have been formulated as conflicting objectives. The uniqueness of the proposed approach, compared to other optimization techniques to solve (8) is the use of bounded gradients to generate $\Sigma\Delta$ limit-cycles about a minima. It can be verified that if $\|A\|_{\infty} \leq 1$ then $\mathbf{w}^* = \mathbf{0}$ is the location of the minima with $f(\mathbf{w}^*, \mathbf{A}) = 0$.

The link between optimization (8) and $\Sigma\Delta$ modulation is through a stochastic gradient minimization [31] of the cost function 9. Assuming that the input random vector \mathbf{x} is stationary and under the assumption that the probability density function of \mathbf{x} is well behaved (gradient of expectation operator is equal to the expectation of the gradient operator), the stochastic gradient step with respect to \mathbf{w} yields

$$\mathbf{w}[n] = \mathbf{w}[n-1] - \left. \frac{\partial f(\mathbf{w}, \mathbf{A})}{\partial \mathbf{w}} \right|_{(n-1)} \quad (10)$$

$$\mathbf{w}[n] = \mathbf{w}[n-1] + \mathbf{A}[n-1]\mathbf{x}[n-1] - \mathbf{d}[n] \quad (11)$$

where n signifies the time steps and $\mathbf{d}[n] = \text{sgn}(\mathbf{w}[n-1])$ denotes the quantized representation. Solving the discrete time recursion (11) leads to

$$\frac{1}{N} \sum_{n=1}^N \mathbf{d}[n] = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{A}[n]\mathbf{x}[n] + \frac{1}{N} (\mathbf{w}[n] - \mathbf{w}[0]) \quad (12)$$

The bounded property of $\mathbf{w}[n]$ leads to the following asymptotic property:

$$\mathcal{E}_n\{\mathbf{d}[n]\} \rightarrow \mathcal{E}_n\{\mathbf{A}[n]\mathbf{x}[n]\} \quad (13)$$

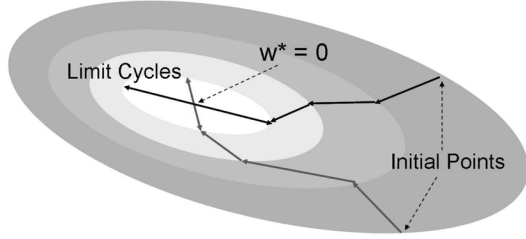


Fig. 4. Limit cycle behavior using bounded gradients.

where $\mathcal{E}_n\{\cdot\}$ denotes an empirical expectation with respect to time index n . Thus, the recursion (11) produces a quantized sequence whose mean asymptotically encodes the transformed input at infinite resolution. This is illustrated in Fig. 4 which shows a 2-D optimization contour. The objective of the $\Sigma\Delta$ learning is to follow the trajectory from an initial condition to the minima and induce limit-cycles about the minima \mathbf{w}^* . The dynamics of the limit-cycles then encodes the shape of the optimization contour and hence also encodes the estimation parameters.

The maximization step (de-correlation) in (8) yields updates for matrix \mathbf{A} according to:

$$\mathbf{A}[n] = \mathbf{A}[n-1] + \xi \left. \frac{\partial f(\mathbf{w}, \mathbf{A})}{\partial \mathbf{A}} \right|_{n-1} \quad (14)$$

which leads to

$$\mathbf{A}[n] = \mathbf{A}[n-1] - \xi \mathbf{w}[n-1] \mathbf{x}[n-1]^T; \mathbf{A}[n] \in \mathcal{C}. \quad (15)$$

The parameter ξ can be expressed in a binary form as

$$\xi = 2^{-P} \quad (16)$$

therefore, the adaptation step in (15) can be expressed by its digital equivalent as

$$\mathbf{A}[n] = \mathbf{A}[n-1] - 2^{-P} \mathbf{d}[n] \mathbf{sgn}(\mathbf{x}[n-1])^T; \mathbf{A}[n] \in \mathcal{C} \quad (17)$$

where we have used the relationship that $\mathbf{d}[n] = \mathbf{sgn}(\mathbf{w}[n-1])$. P is a parameter that controls the update rate and does not affect the stability of learning. As will be discussed in the later sections, the constraint space \mathcal{C} will be restricted to transforms represented by lower triangular matrices with diagonal elements set to unity which can be expressed as $a_{ij} = 0; \forall i < j; a_{ii} = 1$. Thus, to satisfy this constraint $\mathbf{A} \in \mathcal{C}$, only the lower diagonal elements are updated in (17). It can be seen from the (17) that if $\|\mathbf{A}\|_\infty \leq B > 0$ is satisfied, then the recursion (17) will asymptotically lead to

$$\mathcal{E}_n\{\mathbf{d}[n] \mathbf{sgn}(\mathbf{x}[n])^T\} \rightarrow 0 \quad (18)$$

for $\mathbf{A}_\infty \in \mathcal{C}$. The (18) show that the proposed $\Sigma\Delta$ learning algorithm produces quantized sequences that are mutually uncorrelated to a non-linear function of the input signals.

A. $\Sigma\Delta$ Learning for Bearing Estimation

To apply $\Sigma\Delta$ learning for bearing estimation, three of the four differential signals in (7) are chosen as inputs and the transformation matrix \mathbf{A} is chosen to be of the form

$$\mathbf{A} = \begin{bmatrix} a_{11} & 1 & 0 \\ a_{21} & a_{22} & 1 \end{bmatrix} \quad (19)$$

When applied to the three differential signals of the microphone array modeled by (7), the $\Sigma\Delta$ recursions in (11) leads to

$$\begin{aligned} w_1[n] &= w_1[n-1] + \Delta x_2[n] + a_{11}[n] \Delta x_1[n] - d_1[n] \\ w_2[n] &= w_2[n-1] + \Delta x_3[n] + a_{21}[n] \Delta x_1[n] \\ &\quad + a_{22} \Delta x_2[n] - d_2[n] \end{aligned} \quad (20)$$

with $d_1[n] = \text{sgn}(w_1[n])$ and $d_2[n] = \text{sgn}(w_2[n])$. The adaptation steps for the parameters a_{11}, a_{21}, a_{22} based on (17) can be expressed as

$$\begin{aligned} a_{11}[n] &= a_{11}[n-1] - 2^{-P} d_1[n] \mathbf{sgn}(\Delta x_1[n]) \\ a_{21}[n] &= a_{21}[n-1] - 2^{-P} d_2[n] \mathbf{sgn}(\Delta x_1[n]) \\ a_{22}[n] &= a_{22}[n-1] - 2^{-P} d_2[n] \mathbf{sgn}(\Delta x_2[n]). \end{aligned} \quad (21)$$

To demonstrate how the (20) and (21) can be used for bearing estimation, we will consider two different cases based on the quality of common-mode cancellation.

Case I: Perfect Common-Mode Cancellation: In this case it will be assumed that the differential microphones can completely suppress the common-mode signal $x_{cm}(t)$ in (7). Also we will assume that the bearing of the source is located in the positive quadrant $90^\circ > \theta > 0$ such that $\text{sign } \theta, \cos \theta > 0$. Inserting (7) into (20) and (21) the following expressions are obtained:

$$\begin{aligned} a_{11}[N] &= -2^{-P} \sum_{n=1}^N d_1[n] \mathbf{sgn}(\dot{s}[n]) \\ d_1[n] &= \dot{s}[n] \sin \theta + a_{11}[n] \dot{s}[n] \cos \theta \\ &\quad - (w_1[n] - w_1[n-1]) \end{aligned}$$

Combining (22) leads to

$$\begin{aligned} a_{11}[N] &= -2^{-P} \sum_{n=1}^N |\dot{s}[n]| (\sin \theta + a_{11}[n] \cos \theta) \\ &\quad + 2^{-P} \sum_{n=1}^N (w_1[n] - w_1[n-1]) \mathbf{sgn}(\dot{s}[n]). \end{aligned} \quad (22)$$

Let a_{11}^* denote the converged value such that $a_{11}[n] \xrightarrow{n \rightarrow \infty} a_{11}^*$. Equation (23) then leads to

$$\begin{aligned} &\frac{1}{N} \sum_{n=1}^N |\dot{s}[n]| (\sin \theta + a_{11}^* \cos \theta) \\ &\quad - \frac{1}{N} \sum_{n=1}^N (w_1[n] - w_1[n-1]) \mathbf{sgn}(\dot{s}[n]) \xrightarrow{N \rightarrow \infty} 0. \end{aligned} \quad (24)$$

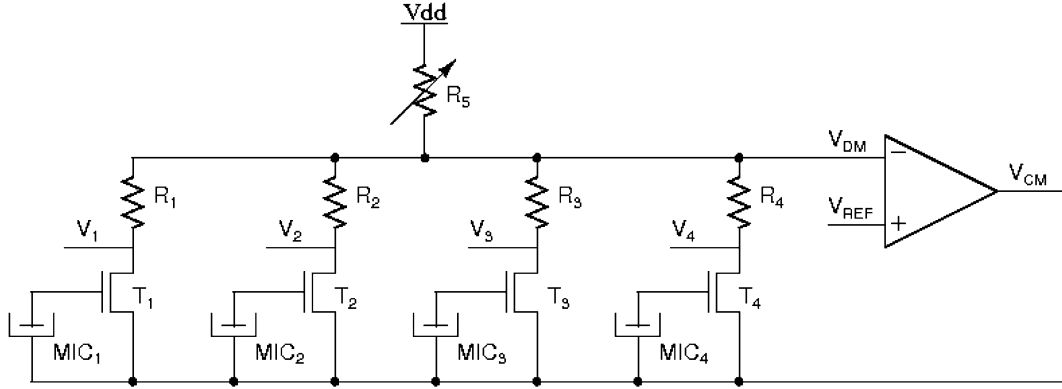


Fig. 5. Schematic of the pre-processing circuit used to separate the common mode and differential mode signal of each microphone.

The second part of the (24) converges to zero which implies that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N |\dot{s}[n]| (\sin \theta + a_{11}^* \cos \theta) = 0. \quad (25)$$

The solution to (25) is given by $a_{11}^* = -\tan \theta$ which leads the following estimate of the bearing

$$\hat{\theta} = \arctan(-a_{11}^*) \quad (26)$$

In a similar fashion, a_{21} and a_{22} can also be used to estimate the bearing of the source according to

$$\hat{\theta} = \arctan\left(\frac{1 - a_{21}}{a_{22}}\right) \quad (27)$$

Both the estimation procedures given by (26) and (27) uses $\Sigma \Delta$ learning to estimate the correlation between differential microphone signals. However, the correlation based approach does not yield robust results when the common-mode cancellation is imperfect which leads to the second case.

Case II: Imperfect Common-Mode Cancellation: In practice, microphone artifacts and mismatch in analog circuits limits the precision of common-mode cancellation. For this case the common-mode signal $s_c(t)$ are assumed to be equal for all the microphones in the array and is given by

$$\begin{aligned} \Delta x_1(t) &= s_c(t) - \dot{s}(t) \frac{d}{c} \cos \theta \\ \Delta x_2(t) &= s_c(t) - \dot{s}(t) \frac{d}{c} \sin \theta \\ \Delta x_3(t) &= s_c(t) + \dot{s}(t) \frac{d}{c} \cos \theta \\ \Delta x_4(t) &= s_c(t) + \dot{s}(t) \frac{d}{c} \sin \theta. \end{aligned} \quad (28)$$

This is a valid assumption since sensor mismatch only changes the location of the point of reference according to which the measurements are made. Then using similar assumptions and approach as in *Case I*, it can be verified that the estimate of the bearing is given by:

$$\hat{\theta} = \arctan\left(\frac{a_{21} - 2a_{21} - a_{11}a_{22} - 1}{a_{21} - a_{11}a_{22} + 1}\right) \quad (29)$$

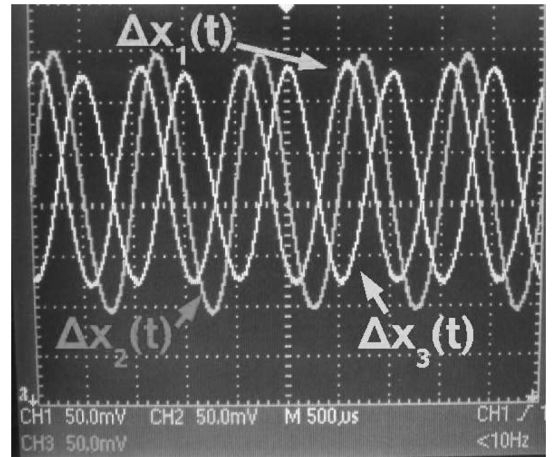


Fig. 6. Scope plot of measured signals obtained from 3 of the 4 microphones when a tone of 1 KHz is played.

In the next section, we will present the hardware implementation of the differential microphone array and an integrated circuit implementation for the updates (20) and (21).

IV. HARDWARE IMPLEMENTATION

A. Differential Microphone Array

The analog frontend circuit presented here is based on a commonly available electret microphone where the diaphragm is coupled to the gate of a field-effect transistor (FET). This is shown in circuit Fig. 6 which is used to cancel the common-mode and to sense and amplify the differential signals. The current through the microphone FET I_m is approximated by a linear form

$$I_m = g_m[V_g - V_s] \quad (30)$$

where g_m is the transconductance of the microphone FET and V_g is the gate-voltage of the FET and V_s is the source voltage of the FET. Given an array of electret microphones (indexed by $i = 1, 2, 3, 4$) used in a circuit shown in Fig. 6 the differential voltage V_i is given by

$$V_i(t) = V_{\text{ref}} - g_m[s_i(t) - V_{\text{cm}}(t)]R_i \quad (31)$$

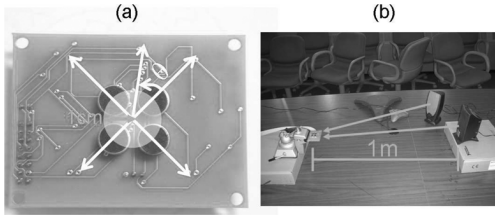


Fig. 7. (a) A 4-microphone array/grid consists electret condenser microphone separated by the distance of 1 cm. (b) The test setup shows the tonal source located at a distance of 1 m from the microphone array.

where $s_i(t)$ is the gate voltage proportional to the acoustic signal impinging on the i th microphone. The common-mode signal $V_{cm}(t)$ in (31) is obtained by summing up the microphone FET currents as

$$\sum_{i=1}^4 g_{mi}[s_i(t) - V_{cm}(t)] = \frac{V_{dd} - V_{ref}}{R_5} \quad (32)$$

which leads to

$$V_{cm}(t) = \frac{s_1(t) + s_2(t) + s_3(t) + s_4(t)}{4} + \frac{V_{dd} - V_{ref}}{g_m R_5}. \quad (33)$$

The operation amplifier in a feedback configuration maintains a virtual ground V_{ref} by adjusting the common-mode signal $V_{cm}(t)$. The outputs $\Delta x_1, \Delta x_2, \Delta x_3$ of the circuit, thus, tracks only the differential signals as required by (7) and hence can achieve good common-mode cancellation.

Fig. 7 shows a prototype of the acoustic array constructed using off-the-shelf electret microphones available from *Knowles*. The circuit shown Fig. 5 is built using discrete components and the differential output is presented as input to a $\Sigma\Delta$ learner integrated circuit. Fig. 6 show a sample differential output produced by the microphone array when a 1 KHz tone is played from a standard computer speaker (the experimental set up is shown in Fig. 5(b)). The scope trace in Fig. 7 clearly shows that the differential signal $\Delta x_2(t)$ is 90° out-of-phase with $\Delta x_1(t)$ and the differential signal $\Delta x_3(t)$ is 180° out-of-phase with respect to $\Delta x_1(t)$.

B. Circuit Implementation of $\Sigma\Delta$ Learner

A system and circuit level implementation of the $\Sigma\Delta$ localizer which implements the updates of (17) and (11) is shown in Fig. 8. The multiplication and addition operations in the updates (17) and (11) are implemented in analog domain using a transconductor array. This is shown in Fig. 8 where the output of the transconductors are connected together to a virtual ground which is determined by the integrator of a first-order $\Sigma\Delta$ modulator. The transconductor (schematic shown in the inset of Fig. 8) is source degenerated using transistor based bump circuit which increases the input linear range. Compared to resistive degeneration, the proposed degeneration consumes a significantly lower area in silicon. The parameter a_{ij} in (11) is determined by the bias current of the transconductor which is digitally controlled using a counter and a current-mode digital-to-analog converter (DAC). Additions in (20) is implemented using Kirchoff's current summation principle by

connecting the outputs of all the transconductors to virtual ground y_i .

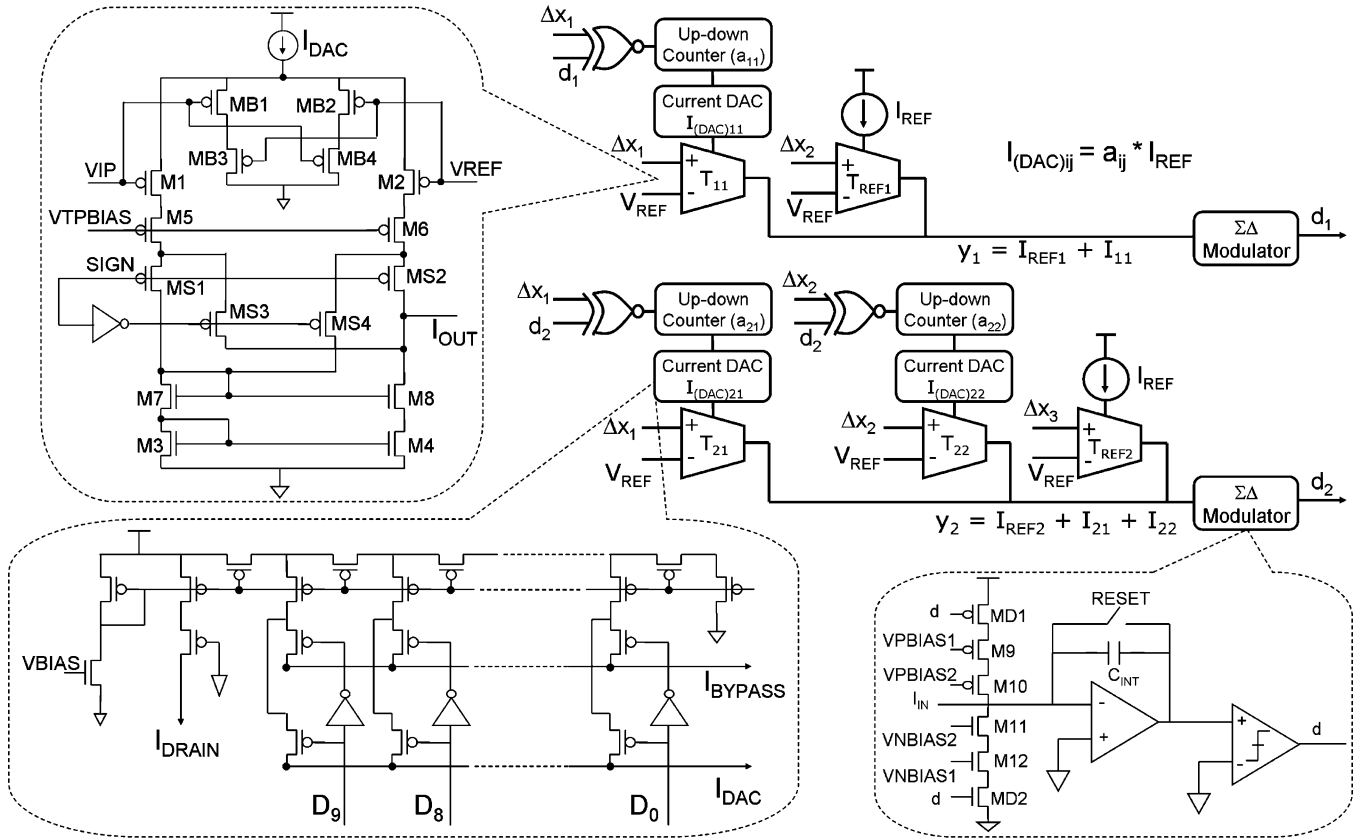
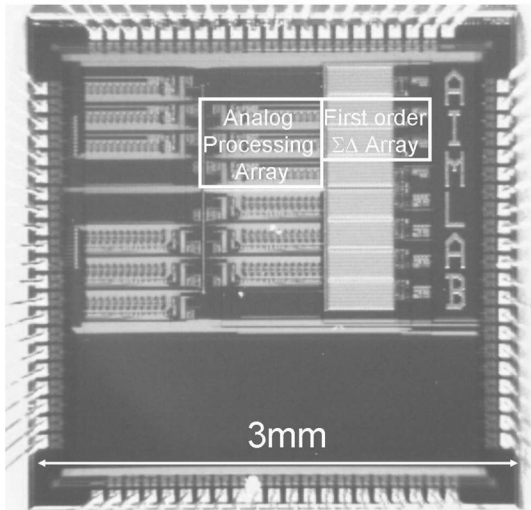
Updates of the parameter a_{ij} according to (15) are implemented using a digitally programmable current DAC (see Fig. 8 inset). An 11-bit up-down counter stores a digital representation of a_{ij} which is processed by a digital-to-analog converter that produces the bias current of the transconductor. Because the updates in (17) is binary, the multiplication operator in (17) is implemented using an XNOR logic (see Fig. 8 inset) whose output controls the UP/DOWN flag of the counter. The counter also incorporates a shift functionality where the contents of the counter can be initialized and accessed using a serial-chain interface. Note that for the diagonal transconductors the parameters a_{ii} are non-adaptive and hence does not use the counter and the current-DAC.

The 10 least significant bits of the Up-Down counter drive a 10-bit current DAC which is implemented using a standard MOS resistive network [32]. The output current I_{OUT} modulates the bias current of the transconductor shown in Fig. 8 whereas the most significant (11th) bit of the Up-Down counter controls the direction of the output current, which is useful for implementing a four-quadrant multiplier. The transconductor consists of a p-MOS input stage which drives a cascoded output stage. The transconductor uses a bump circuit [33] (Transistors *MB1* – *MB4*) to achieve source degeneration and hence increase the input voltage range (reduce the transconductance). The bump circuit operates by steering the output current in a manner that the transistor pair implements an equivalent resistor. The direction of the output current is digitally controlled by the *MSB* (*SIGN*) of the Up-Down counter. A standard folded cascode op-amp with open loop DC gain of 60 dB is used for the integrator. The reference current of the first order $\Sigma\Delta$ modulator passing through transistor devices *M9* – *M12* are tuned to accommodate the maximum input current of ($y_4 = I_4 = 4 \times \max(I_{OUT})$) to avoid overload condition. As shown in Fig. 8, the multiplication between the digital bit d and the reference current is implemented by switching (on/off) the cascoded current source (sink). Switching at the source as opposed to switching at the drain has several advantages [35] as it reduces the channel charge injection [35] and clock feed-through at the integration node.

V. MEASUREMENT RESULTS

A. Circuit Characterization

A prototype $\Sigma\Delta$ localizer implementing the circuit in Fig. 8 was fabricated in a $0.5 \mu\text{m}$ CMOS process. Fig. 9 shows the micrograph of the fabricated prototype and Table I summarizes its measured specifications. Even though the $\Sigma\Delta$ localizer has been verified to operate up to 250 KHz, all the measurements presented in this paper performed at a sampling frequency of 25 KHz. The $\Sigma\Delta$ learner has an input range of 300 mV which is determined by the linear operating range of the source degenerated transconductor shown in Fig. 8. The power dissipation of individual components of the $\Sigma\Delta$ learner are given in Table II. The power dissipation of a single $\Sigma\Delta$ modulator was measured to be $8.68 \mu\text{W}$ which consistent with the metric reported in [34]. The average power dissipation of a single transconductor along

Fig. 8. Circuit diagram of $\Sigma\Delta$ learner.Fig. 9. Microphotograph of $\Sigma\Delta$ learner showing the 2 channels of the system performing the source localization.

with the counter and the DAC was measured to be $7.15 \mu W$. The power dissipation metrics of each channels (corresponding to each $\Sigma\Delta$ modulator) is summarized in Table II, where the calculation is based on the following formula:

$$P_{Mod} + P_{DC} + (j - 1)(P_{NDC} + P_{CNT})\mu W \quad (34)$$

TABLE I
MEASURED SPECIFICATION OF $\Sigma\Delta$ LEARNER

Parameters	Values
Technology	0.5 μm 2P3M CMOS
Die Size	3 mm \times 3 mm
Supply	3.3 V
Channels	2
Input Range	300mV
Sampling Frequency	25 kHz
Total Power dissipation	75.11 μW at 25KHz (2-channel)
Active area of the 1 st order $\Sigma\Delta$ modulator	662 $\mu m \times$ 193 μm
Active area of Analog Processing Module	760 $\mu m \times$ 540 μm

TABLE II
MEASURED POWER DISSIPATION OF THE $\Sigma\Delta$ LEARNER. (A) SYSTEM COMPONENT POWER AT 25 kHz (B) SYSTEM CHANNEL POWER

(a)

1 st order mixed mode Modulator (P_{Mod})	8.68 μW
Diagonal cell (P_{DC})	6.6 μW
Non-diagonal Cell (P_{NDC})	7.7 μW
10-bit counter/shifter (P_{CNT})	7.15 μW

(b)

Channel 1	30.13 μW
Channel 2	44.98 μW

P_{Mod} represents the dynamic power dissipation of the modulator, P_{DC} represents the static power dissipation, P_{NDC} represents the power dissipation of a single transistor and

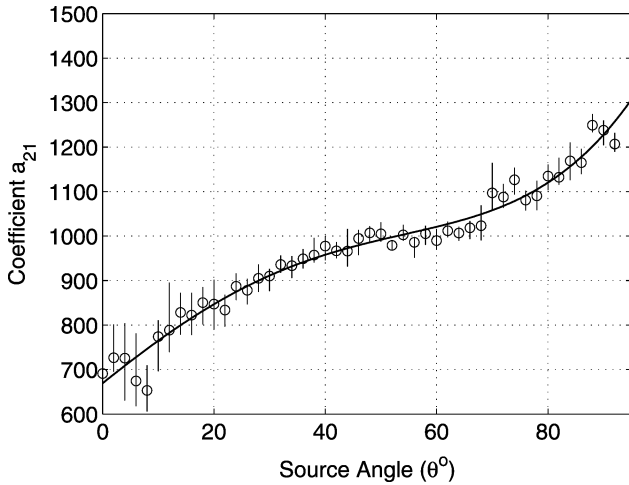


Fig. 10. Measured response demonstrating that the parameter a_{11} tracks the bearing angle θ of the sound source.

P_{CNT} represents the power dissipation of the counter and current DACs.

B. Acoustic Source Localization

The microphone array is formed with the omnidirectional electret condenser microphones (*Knowles Inc*). This condenser microphone has the sensitivity of -42 dB at 1 kHz. The Signal to Noise Ratio (SNR) of this microphone is 55 dB. The 4-microphone array is designed with each condenser microphone 1 cm apart is shown in Fig. 7(a). The far field effect is generated by keeping the acoustic source at a distance of 1 m from the microphone array. A 1 KHz tone was played through a standard loud speaker and the parameters a_{11} , a_{21} and a_{22} were obtained by accessing the counters on-chip (through the serial-chain interface). In all the experiments, the position of the speaker was fixed and the microphone array was rotated in steps of the desired angle. The source localization was then carried out using only three microphones which we have shown to be sufficient using (29) for compensating sensor mismatch. For each angular orientation of the source ten sets of measurements ($(a_{11}, a_{21}$ and $a_{22})$) were recorded every 80 ms using an Xilinx field-programmable gate array (FPGA) and (26), (27) and (29) were used to estimate the bearing angle. The power dissipation of the $\Sigma\Delta$ localizer operating at 25 KHz was measured to be $75.11 \mu\text{W}$ and it was observed that this metric reduces quadratically with a reduction in the sampling frequency.

Fig. 10 shows the mean and standard deviation corresponding to the parameter a_{11} as measured using the $\Sigma\Delta$ localizer. The response shows a monotonic behavior with respect to the bearing angle which is varied from 0° to 90° . For this experiment the bearing angle is varied by steps of 4° and the response was also verified to be antisymmetric with respect to the origin. The plot 10 shows a resemblance to a $\tan(\cdot)$ function with deviation at the two end points (0° and 90°) due to finite resolution of the DAC. The measured response is therefore consistent with the theoretical response given by (26).

Similarly, the measured response in Fig. 11 shows that the bearing estimate as obtained using (27) also demonstrates a monotonic response with respect to the bearing angle. However,

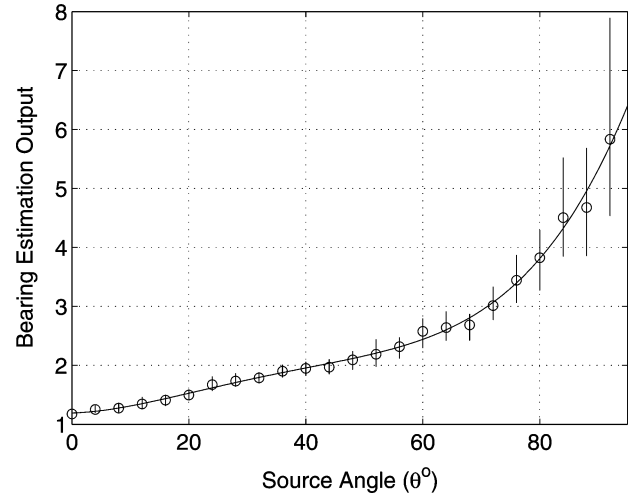


Fig. 11. Bearing estimates obtained using (27).

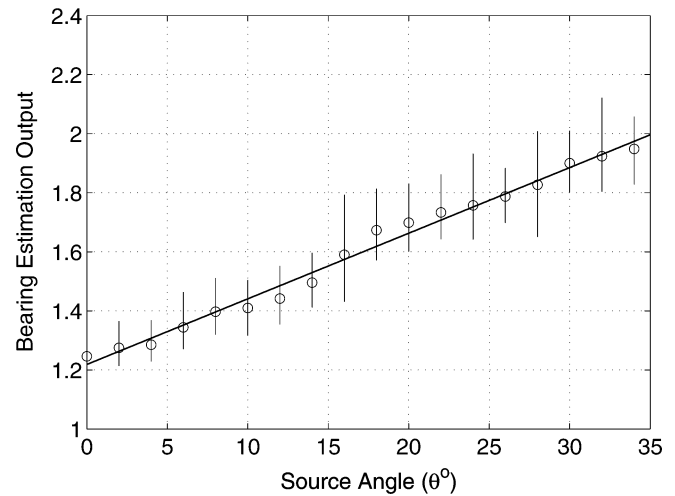


Fig. 12. Zoomed in response for Fig. 11 where from within the range of 0° – 34° , a resolution of 2° .

compared to the estimate obtained using the parameter a_{11} as shown in Fig. 10, the estimates using a_{21} and a_{22} demonstrates 7.3% lower standard deviation and hence provides a more reliable estimate. Fig. 12 shows the measured response when the source was varied from 0° to 35° . The mean response (computed using 10 sets of measurements) demonstrates a linear behavior with respect to the source angle. The resolution of the bearing estimates defined as the minimum angular separation that can be resolved was measured to be 2° .

In the next set of experiment all the three parameters a_{11} , a_{21} , a_{22} were used for bearing estimation according to the (29). Fig. 13 shows that (29) suppresses the non-linearity of the estimation process and significantly improves the linear dynamic range (from 34° to 90°) response. Also the standard deviation of the bearing estimate reduced by 73%. For this experiment, the bearing resolution was measured to be 4° . We conducted a similar experiment where bearing angle was varied from 0° to 50° . The measured response shown in Fig. 14 demonstrates a resolution of 2° which is comparable to the state-of-the-art acoustic localizers reported in literature. All the

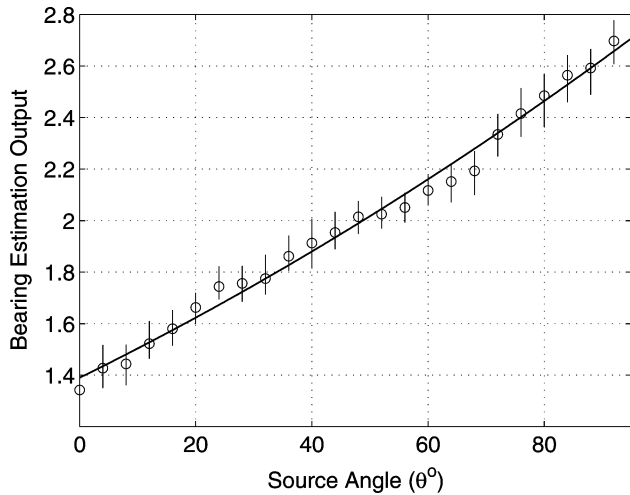


Fig. 13. Bearing estimates obtained using (29) for a dynamic range of 0° – 90° at a resolution of 4° .

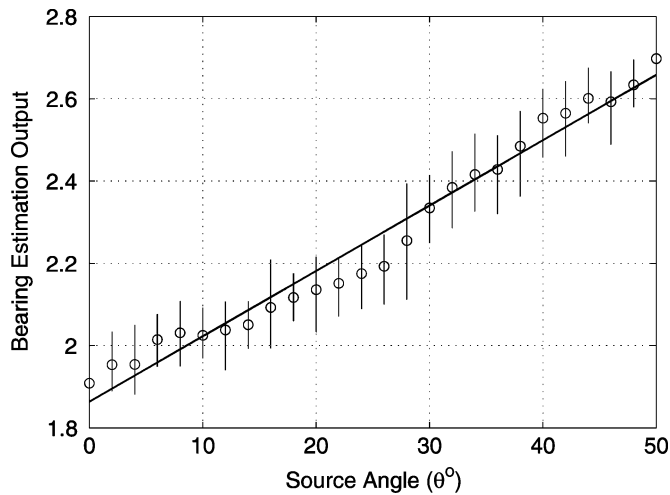


Fig. 14. Bearing estimates obtained using (29) for a dynamic range 0° – 50° at a resolution of 2° .

experiments presented in this section show that the $\Sigma\Delta$ learner can be used for localizing an acoustic source using a miniaturized microphone array while consuming only microwatts of power.

VI. CONCLUSION

In this paper we presented algorithm and hardware implementation of acoustic source localizer based on $\Sigma\Delta$ learning. At the core of the proposed approach is the integration of $\Sigma\Delta$ modulation with statistical learning that leads to robust estimates of the bearing angle of the acoustic source. We also presented a circuit level implementation of a miniature differential microphone array which uses discrete analog components to sense the far-field differential signals. These differential signals is processed by a $\Sigma\Delta$ localizer prototyped in silicon which produces the bearing parameters used for source localization. We demonstrated that by employing multiple parameters available from the $\Sigma\Delta$ localizer, the bearing estimates can be significantly improved and its linear dynamic range can be significantly extended. Even though in this paper, we have demon-

strated a bearing resolution of 2° , the proposed algorithm and architecture can be expanded to approach the fundamental limit of bearing estimates as determined by Cramer-Rao Lower Bound (CRLB) [14], [22] which for the relative spacing of the microphones (1 cm), sampling rate of 25 KHz and estimation duration of 100 ms, is computed to be 1° [36]. One of the ways to approach this limit is to increase the resolution of the DAC as well as use higher-order noise-shaping $\Sigma\Delta$ modulator in the localizer. However, this can only be achieved at the expense of larger silicon area and higher power dissipation. In the future, our focus will be to reduce the power dissipation of the localizer which is primarily dominated by the current DAC and the counter.

REFERENCES

- [1] R. N. Miles and R. R. Hoy, "The development of a biologically-inspired directional microphone for hearing aids," *Audiology Neuro-Otology*, vol. 11, no. 2, pp. 86–94, 2006.
- [2] S. Chowdhury, M. Ahmadi, and W. C. Miller, "Design of a MEMS acoustical beamforming sensor microarray," *IEEE J. Sens.*, vol. 2, no. 6, pp. 617–627, Dec. 2002.
- [3] H. Gharavi and S. P. Kumar, Eds., "Special issue on sensor networks and applications," *Proc. IEEE*, vol. 91, pp. 1151–1153, Aug. 2003.
- [4] R. Showen, "Operational gunshot location system," *Proc. SPIE*, vol. 2935, pp. 130–139, Feb. 1997.
- [5] B. Edwards, "The future of hearing aid technology," *Trends Amplification*, vol. 11, no. 1, pp. 31–45, Mar. 2007.
- [6] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA: MIT Press, 1997.
- [7] C. Lorenzi, S. Gatehouse, and C. Lever, "Sound localization in noise in hearing-impaired listeners," *J. Acoust. Soc. Am.*, vol. 105, no. 6, pp. 3454–3463, 1999.
- [8] J. C. Chen, J. E. L. Yip, H. Wang, D. Maniezzo, R. E. Hudson, K. Yao, and D. Estrin, "Coherent acoustic array processing and localization on wireless sensor networks," *Proc. IEEE*, pp. 1154–1162, Aug. 2003.
- [9] J. C. Chen, K. Yao, and R. E. Hudson, "Source localization and beamforming," *IEEE Signal Processing Mag.*, pp. 30–39, Mar. 2002.
- [10] G. C. Carter, "Time delay estimation for passive sonar signal processing," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. ASSP-29, no. 6, pp. 463–470, Jun. 1981.
- [11] G. C. Carter, "Coherence and time delay estimation," *Proc. IEEE*, vol. 75, pp. 236–255, Feb. 1987.
- [12] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-24, pp. 320–327, Aug. 1976.
- [13] A. H. Quazi, "An overview on the time delay estimate in active and passive systems for target localization," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, no. 6, pp. 527–533, Jun. 1981.
- [14] P. Julian, A. G. Andreou, L. Riddle, S. Shamma, D. H. Goldberg, and G. Cauwenberghs, "A comparative study of sound localization algorithms for energy aware sensor network nodes," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 4, pp. 640–648, Apr. 2004.
- [15] P. Julian, A. G. Andreou, and D. H. Goldberg, "A low-power correlation-derivative CMOS VLSI circuit for bearing estimation," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 14, no. 2, pp. 207–212, Feb. 2006.
- [16] P. Julian, F. N. M. Pirchio, and A. G. Andreou, "Experimental results for cascaded micropower time delay estimator," *IEEE Electron. Lett.*, vol. 42, no. 21, pp. 1218–1219, Oct. 2006.
- [17] S. Shamma, N. Shen, and P. Gopalaswamy, "Stereausic: Binaural processing without neural delays," *J. Acoust. Soc. Amer.*, vol. 86, pp. 989–1006, 1989.
- [18] V. Chan, S.-C. Liu, and A. van Schaik, "AER EAR: A matched silicon cochlea pair with address event representation interface," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 54, no. 1, pp. 48–59, Jan. 2007.
- [19] R. Z. Shi and T. K. Horiuchi, "A neuromorphic VLSI model of bat interaural level difference processing for azimuthal echolocation," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 54, no. 1, pp. 74–88, Jan. 2007.
- [20] A. C. Mason, M. L. Oshinsky, and R. R. Hoy, "Hyperacute directional hearing in a microscale auditory system," *Nature*, vol. 410, Apr. 2001.

- [21] D. Robert, J. Amoroso, and R. R. Hoy, "The evolutionary convergence of hearing in a parasitoid fly and its cricket host," *Science*, vol. 258, pp. 1135–1137, 1992.
- [22] M. Stanacevic and G. Cauwenberghs, "Micropower gradient flow acoustic localizer," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 52, no. 10, pp. 2148–2157, 2005.
- [23] , S. R. Norsworthy, R. Schreier, and G. C. Temes, Eds., *Delta-Sigma Data Converters*. Piscataway, NJ: IEEE Press, 1997.
- [24] D. J. Mar, C. C. Chow, W. Gerstner, R. W. Adams, and J. J. Collins, "Noise shaping in populations of coupled model neurons," *Proc. Natl. Acad. Sci. USA*, vol. 96, pp. 10450–10455, Aug. 1999.
- [25] M. N. Do, "Toward sound-based synthesis: The far-field case," in *Proc. IEEE Conf. Acoust. Speech, Signal Process. (ICASSP)*, 2004, vol. 2, pp. 601–604.
- [26] A. Gore and S. Chakraborty, "Large margin analog-to-digital converters with applications in neural prosthetics," in *Adv. Neural Inf. Process. Systems (NIPS)*. Cambridge, MA: MIT Press, 2006, vol. 19, pp. 532–539.
- [27] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
- [28] F. Girosi, M. Jones, and T. Poggio, "Regularization theory and neural networks architectures," *Neural Computation*, vol. 7, pp. 219–269, 1996.
- [29] D. Fudenberg and D. Levine, *The Theory of Learning in Games*. Cambridge, MA: MIT Press, 1998.
- [30] T. Basar and P. Bernhard, *H[∞] Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*. New York: Springer, 1995.
- [31] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [32] B. Linares-Barranco and T. Serrano-Gotarredona, "On the design and characterization of femtoampere current-mode circuits," *IEEE J. Solid-State Circuits*, vol. 38, no. 8, pp. 1353–1363, Aug. 2003.
- [33] P. M. Furth and H. A. Ommami, "Low-voltage highly-linear transconductor design in subthreshold CMOS," in *Proc. 40th Midwest Symp. Circuits Syst. (MWCAS)*, Aug. 1997, vol. 1, no. 3–6, pp. 156–159.
- [34] A. Gore, S. Chakraborty, S. Pal, and E. C. Alocilja, "A multichannel femtoampere-sensitivity potentiostat array for biosensing applications," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 53, no. 11, pp. 2357–2363, Nov. 2006.
- [35] G. Cauwenberghs and V. Pedroni, "A charge-based CMOS parallel analog vector quantizer," in *Adv. Neural Information Processing Systems (NIPS)*. Cambridge, MA: MIT Press, 1995, vol. 7, pp. 779–786.
- [36] B. Friedlander, "On the Cramer-Rao bound for time delay and Doppler estimation," *IEEE Trans. Inf. Theory*, vol. 30, no. 3, pp. 575–580, 1984.



Amit Gore received the B.E. degree in instrumentation engineering from the University of Pune, Pune, India, in 1998, and the M.S. and Ph.D. degrees in electrical and computer engineering from Michigan State University, East Lansing, in 2002 and 2008, respectively.

Currently, he is with General Electric Global Research, Niskayuna, New York. His research interests are low-power sigma-delta converters, analog signal processing and lowpower mixed-signal VLSI design.



Amin Fazel (S'07) received the B.Sc. degree in computer science and engineering from Shiraz University, Shiraz, Iran, in 2002 and the M.Sc. degree in computer engineering from Sharif University of Technology, Tehran, Iran, in 2005. Currently, he is pursuing the Ph.D. degree in electrical and computer engineering at Michigan State University.

His research interests include speech processing, speech/speaker recognition, acoustic source separation, and analog-to-feature converters.



Shantanu Chakraborty (M'96–SM'09) received the B.Tech. degree from Indian Institute of Technology, Delhi, India, in 1996, and the M.S. and Ph.D. degrees in electrical engineering from Johns Hopkins University, Baltimore, MD, in 2001 and 2004, respectively.

He is currently an Assistant Professor in the Department of Electrical and Computer Engineering at Michigan State University, East Lansing. From 1996–1999 he was with Qualcomm Incorporated, San Diego and during 2002 he was a visiting

researcher at University of Tokyo. His current research interests include low-power analog and digital VLSI systems, hardware implementation of machine learning algorithms with application to biosensors and biomedical instrumentation.

Dr. Chakraborty was a recipient of The Catalyst foundation fellowship from 1999–2004 and received the best undergraduate thesis award in 1996. He is currently a senior member for IEEE BioCAS technical committee, IEEE Circuits and Systems Sensors technical committee and serves as an associate editor for *Advances in Artificial Neural Systems* from Hindawi publications.