

A Comparative Study of Sound Localization Algorithms for Energy Aware Sensor Network Nodes

Pedro Julián, *Member, IEEE*, Andreas G. Andreou, Larry Riddle, Shihab Shamma, David H. Goldberg, and Gert Cauwenberghs, *Member, IEEE*

Abstract—Sound localization using energy-aware hardware for sensor networks nodes is a problem with many applications in surveillance and security. In this paper, we evaluate four algorithms for sound localization using signals recorded in a natural environment with an array of commercial off-the-shelf microelectromechanical systems microphones and a specially designed compact acoustic enclosure. We evaluate performance of the algorithms and their hardware complexity which relates directly to energy consumption.

Index Terms—Direction of arrival estimation, intelligent sensors, networks.

I. INTRODUCTION

SOUND localization using compact sensor nodes deployed in networks [1] has applications in security, surveillance, and law enforcement [2]. Several groups have reported on coherent [3] and noncoherent [4] methods for sound localization, detection, classification, and tracking in sensor networks [5]. Coherent methods are based on the arrival time differences of the acoustic signal to the sensors [6]. In standard systems, microphones are separated to maximize accuracy, therefore, the nodes need to achieve synchronization to produce a valid estimate [4]. The need of synchronization implies a frequent communication which is expensive in terms of power consumption. Noncoherent methods like closest point of approach (CPA) [4] are not critical with respect to synchronization, but are sensitive to sensor mismatch and differences in the channels between the sound source and the sensors. The methods discussed in this paper are all coherent approaches at

Manuscript received March 25, 2003; revised December 2, 2003. This work was supported in part by the Defense Advanced Research Projects Agency/Office of Naval Research under Contract N00014-00-C-0315 and in part by Intelligent and Noise-Robust Interfaces for Microelectromechanical Systems Acoustic Sensors. The work of P. Julián, A. G. Andreou, and D. H. Goldberg was supported by the National Science Foundation under Grant EIA-0130812.

P. Julián is with the Electrical and Computer Engineering Department, The Johns Hopkins University, Baltimore, MD 21210 USA and also with Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Capital Federal CP 1033, Argentina, on leave from the Departamento de Ingeniería Eléctrica y Computadoras, Universidad Nacional del Sur, 8000 Bahía Blanca, Argentina (e-mail: pjulian@ieee.org).

A. G. Andreou, D. H. Goldberg, and G. Cauwenberghs are with the Electrical and Computer Engineering Department, The Johns Hopkins University, Baltimore, MD 21210 USA.

L. Riddle is with the Signal Systems Corporation, Annapolis, MD 21146 USA.

S. Shamma is with the Department of Electrical Engineering, University of Maryland, College Park, MD 20742 USA.

Digital Object Identifier 10.1109/TCSI.2004.826205

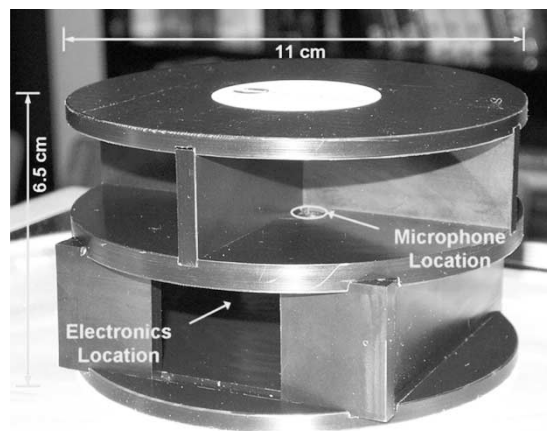


Fig. 1. Photograph of ASU enclosure.

the node level, therefore eliminating the need for synchronization. Indeed, one of the presented methods, the gradient flow algorithm, is capable of bearing estimation with subwavelength distances among sensors.

In the above-mentioned references, low-power commercial off-the-shelf (COTS) hardware are employed. However, even with low-power state-of-the-art hardware, COTS devices consume power at the milliwatt level. While in some applications, this is adequate, in truly autonomous nodes that harvest energy from the environment (sun) the power dissipation must be reduced many orders of magnitude to the microwatt level. This can only be attained by co-designing the algorithms with custom mixed analog–digital hardware [7]–[9].

In this paper, we evaluate four different algorithms for bearing estimation using signals recorded in a natural environment with an array of four microelectromechanical systems (MEMS) microphones, embedded in a custom designed acoustic enclosure (see Fig. 1). The algorithms are aimed towards a custom mixed analog–digital integrated circuit implementations and the complexity of the algorithm is related to the power dissipation in the final system.

II. STATEMENT OF PROBLEM

A general array of four microphones as illustrated in Fig. 2, pairwise separated by a distance d , will be considered. The objective pursued is the estimation of the bearing angle, i.e., the angle of the sound source with respect to either coordinate axis.

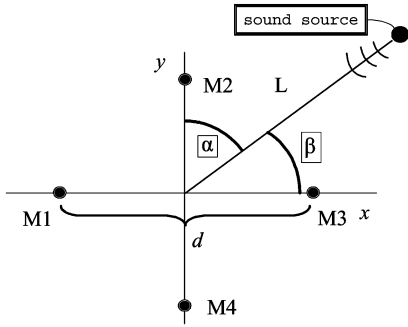


Fig. 2. Microphones array to measure the bearing angle.

For example, referring to Fig. 2, if we are using the pair of microphones M1 and M3, then, the bearing angle will be given by β , whereas if the pair of microphones is M2 and M4, then the bearing angle will be given by α .

The study is based on a particular application that employs an acoustic surveillance unit (ASU) enclosure (shown in Fig. 1) with an array of four Knowles SiSonic MEMS microphones. The distance between microphones is $d = 6$ cm; however, the acoustic enclosure produces an effective separation between microphones $d_s = 15.9$ cm. In all cases, we are assuming that the sound source is far away from the microphones ($d_s \ll L$). The ASU enclosure also includes a signal conditioning circuitry consisting of a low-pass filter and a gain stage. The original motivation for the study was to develop an algorithm able to localize a broad-band signal in the frequency range [20 Hz, 300 Hz] with an accuracy of one degree, using an estimation period of 1 s. In order to evaluate different algorithms, signals were recorded in a natural environment using a digital audio acquisition board at 2048 samples/s. In Section III, the different algorithms are introduced while in Section IV, the details of the experiment and the obtained results are described.

III. BEARING ESTIMATION ALGORITHMS

In this section, we summarize the four algorithms employed to estimate the bearing of the sound source with respect to the four microphones in the ASU. The algorithms are: 1) cross-correlation algorithm (CA) [6]; 2) cross-correlation derivative algorithm (CDA); 3) spatial-gradient algorithm (SGA) [10]; 4) stereausis algorithm (SA)[11].

All methods employ time-domain signal processing based on coherent localization, with the particular feature that nodes are spaced at subwavelength spacing. All methods are inspired by biological information processing structures.

A. Cross-Correlation Algorithm (CA)

Bearing estimation using time-domain CAs has been extensively studied in the literature (see [6] and [12]–[14]). A time-domain CA is also used by many animals such as the barn owl to provide azimuth information [15]. An analog very-large scale integration (VLSI) implementation of the barn owl azimuth localization system was reported by Lazzaro and Mead [16].

Consider one pair of microphones with signals $x_1(t)$ and $x_3(t)$ arriving at the two microphones given by

$$\begin{aligned} x_1(t) &= s(t) + n_1(t) \\ x_3(t) &= s(t - T_D) + n_3(t) \end{aligned} \quad (1)$$

where $s(\cdot)$ is the signal emitted by the source, $n_1(\cdot)$ and $n_3(\cdot)$ are uncorrelated noise signals, and T_D is the time delay between microphones. Under the assumption that the source is far away, the signal arriving at the two microphones can be approximated by a plane wave, and the following relation holds:

$$T_D = \frac{d}{c \cos(\beta)} = T_{D \max} \cos(\beta) \quad (2)$$

where $c = 345$ m/s is the speed of sound in air at ambient temperature and $T_{D \max} := d_s/c$ is the maximum delay. The correlation between signals x_1 and x_3 is given by

$$R_{x_1 x_3}(\tau) = \int_{-\infty}^{\infty} x_1(t) x_3(t + \tau) dt. \quad (3)$$

After replacing (1) into (3), and considering that n_1 and n_3 are uncorrelated, (3) can be rewritten as

$$R_{x_1 x_3}(\tau) = \int_{-\infty}^{\infty} s(t) s(t - T_D + \tau) dt. \quad (4)$$

This function will exhibit a maximum at $\tau = T_D$. Therefore, one way to estimate the time delay is to generate (3) numerically and calculate the time where the maximum is achieved.

In practice, the signal is sampled at a certain frequency $f_s = 1/T_s$ and the correlation is approximated using a discrete time version

$$\tilde{R}_{x_1 x_3}(iT_s) = \sum_{k=0}^K x_1(kT_s) x_3((k-i)T_s) \quad (5)$$

where K is such that $K \cdot T_s$ is the time window under consideration. From now on, we will discard T_s in the notation, and instead we will use discrete instants indexed with an integer. Operation (5) can be implemented in a digital fashion after quantization of the signals. Using experimental data, we found that a 1-bit quantization was sufficient to obtain accurate estimations, as will be shown later. From a hardware perspective, coding the signal with just 1 bit produces a dramatic reduction in complexity. The resulting architecture consists of a number of stages in the form of

$$y(i) = \sum_{k=0}^K x_1(k) x_3(k - i) \quad (6)$$

where i is an index to the stage number (see Fig. 3).

At this point, some practical considerations are in order. A sampling frequency of 200 kHz is required (see Appendix) to estimate the angle with an accuracy of one degree for angles in the range $\alpha \in [0, 50] \cup [+130, +180]$. As the test signals were sampled at 2048 Hz, for this particular case, it was necessary to interpolate and resample the signals at 200 kHz. This choice of sampling frequency implies that every discrete time delay is $T_s = 5 \mu\text{s}$. As the maximum possible delay, that corresponds to an angle $\beta = 90^\circ$ is $T_{D \max} = d_s/c = 460 \mu\text{s}$, 92 stages are necessary. Accordingly, index i in (6) ranges from 0 to 91.

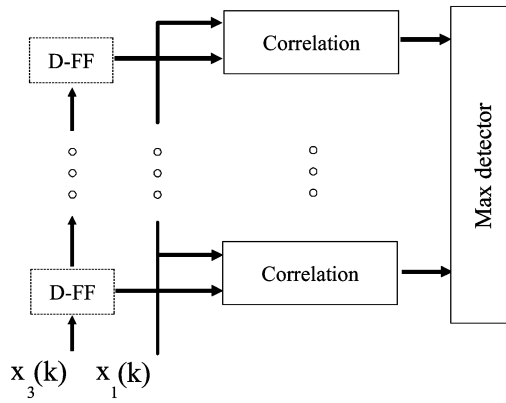


Fig. 3. Estimation architecture for CA.

From a hardware viewpoint, the digital implementation of (6) requires shift registers to generate the delayed versions of x_3 , a counter implementing the correlation operation and one block to determine where the maximum has occurred. Once the signal is quantized with 1 bit, the information corresponding to the time delay between both signals is encoded in the relative changes of state from zero to one and one to zero. Accordingly, no information is contained in those parts of the signal where there are no state changes. However, every stage (6) counts all the time at the speed set by the clock, regardless of input values. As the frequency of the clock is much higher than the frequency of the signal, this architecture will dissipate more power than what is actually necessary. This observation motivated the approach presented in Sections III-B–D. Another point that must be considered in this approach is the need to calculate the occurrence of the maximum of (6), which requires the implementation of additional circuitry (a winner-takes-all circuit or an equivalent digital circuit).

B. Cross-Correlation Derivative Algorithm (CDA)

As we said, the maximum of the correlation occurs when the delay produced by the shift register chain coincides with the relative delay between signals. Mathematically, detecting the maximum of the correlation function is equivalent to detecting the zero crossing of its derivative when the second derivative is negative. This has several advantages as we will show.

If we consider (6) and calculate the discrete difference between adjacent elements for every stage, we get

$$\begin{aligned} \Delta y(i) &:= y(i) - y(i-1) \\ &= \sum_{k=0}^l x_1(k)[x_3(k-i) - x_3(k-(i-1))]. \end{aligned} \quad (7)$$

A careful observation of (7) reveals that it is in fact an up/down (UP/DN) counter. The counter counts up when $x_1(k) = 1$ and the other signal satisfies $x_3(k-i) = 1$ and $x_3(k-(i-1)) = 0$; it counts down when $x_1(k) = 1$ and the other signal satisfies $x_3(k-i) = 0$ and $x_3(k-(i-1)) = 1$. Accordingly, the signals UP and DN commanding the counter can be written as

$$\begin{aligned} \text{UP} &= x_1(k) \cdot \left(x_3(k-i) \overline{x_3(k-i+1)} \right) \\ \text{DN} &= x_1(k) \cdot \left(\overline{x_3(k-i)} \cdot x_3(k-i+1) \right). \end{aligned} \quad (8)$$

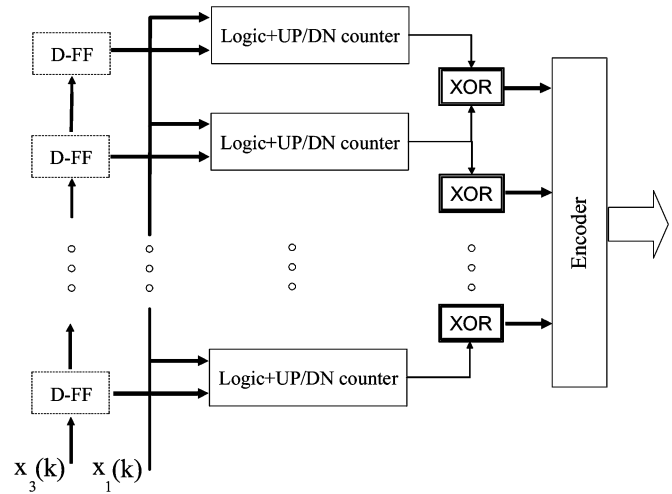


Fig. 4. Estimation architecture for CDA.

In this case, the count is only updated when one of the two signals changes its state, and it is idle the rest of the time. This mode of operation reduces the activity of the circuit and consequently implies a reduced power consumption and also smaller counters. In addition, to obtain the value of the delay, it is just necessary to read the position of the stage where the zero crossing occurred. Related to this, notice that all counters above the stage where the coincidence occurs will have a count of a given sign, whereas all counters below will have a count of the opposite sign. Therefore, the zero crossing can be detected by connecting the sign bit of every pair of adjacent blocks (7) to an XOR gate, in such a way that it will become active when two adjacent cells have a count of different sign (see Fig. 4). Then, if the XOR gates are connected to encoders the position of the zero crossing is converted to a binary number that gives the reading. Notice here that the use of the derivative in the calculation of the correlation eliminates the need to search for the maximum of the outputs, and instead provides a straightforward architecture to read the value of the delay.

A final observation regarding the number of maxima occurring. In this application, we are relying on the fact that the minimum period of the signal ($T = 1/300 = 3.3$ ms) is larger than the maximum delay, so that only one maximum will be noticeable in the range of times considered, given by $[0, 460 \mu\text{s}]$. Another assumption for this observation to be true is that only one source of sound is present. The introduction of multiple sources would give rise to several maxima (and minima).

C. Stereausis Approach (SA)

This approach is inspired in the stereausis network described in [11] that uses two cochlea channels to preprocess the input signals. In this case, every channel only reproduces the transfer function of the basilar membrane (The model presented in [17] also models the outer ear and fluid-cilia coupling stages). Following the work in [17] and [11], the output of every section of basilar membrane is modeled with an infinite impulse response (IIR) (bandpass) digital filter. The frequency responses (magnitude) of the 32 filters used are shown in Fig. 5. Analog VLSI implementations of the basilar membrane as bank filters have also been described in the literature [18].

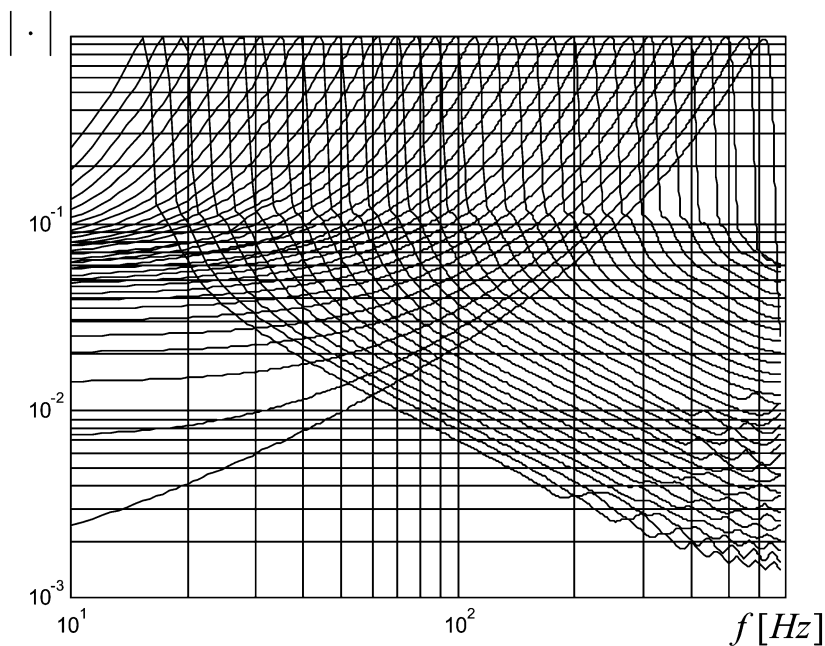


Fig. 5. Frequency response (in magnitude) of the stereoisis network bandpass filters.

In the stereoisis network, the sound from the left and right microphones are fed to the ipsi-lateral and contra-lateral cochlea channels, respectively. Then, all outputs are quantized to one bit and the outputs of every stage of one channel are digitally correlated with the outputs of the other channel. In this way, a spatial arrangement of elements results, which can be associated to an image, namely C , whose (i, j) element $C_{i,j} \geq 0$ is the correlation between the output of the i th element of the ipsi-lateral channel and the output of the j th element of the contra-lateral channel (see Fig. 6). When the left and right signals are equal, the resulting image C will have a significant density of nonzero elements along the main diagonal. However, if there is a delay in one of the signals, the image C will show a shift of the main diagonal toward one of the sides. This is illustrated in Fig. 7, which shows the response of the network to a set of real data where one of the inputs is delayed. The simulated network consists of a 32-stage cochlea with cutoff frequencies between 252 and 618 Hz. Notice that as a delay of τ s is equivalent to a phase shift of $\phi(f) = 2\pi f\tau$, the higher the frequency the more noticeable the unbalance of the image with respect to the main diagonal. Actually, the frequency range of the filter was adjusted to maximize the detection sensitivity through simulations.¹

The indication of time delay is calculated by measuring the unbalance of image C with respect to the main diagonal. This is done by computing the difference between the sum of upper diagonal elements and lower diagonal elements, i.e.,

$$\Upsilon = \sum_{i < j} C_{i,j} - \sum_{i > j} C_{i,j}. \quad (9)$$

¹At first sight, it might seem surprising that the cutoff frequencies of the filters are higher than the signal bandwidth. Regarding this, the reader should note that every cochlea filter section is a bandpass filter with a very long tail. Accordingly, the filter with the highest cutoff frequency (e.g., 618 Hz) still amplifies the contribution of lower frequency signals.

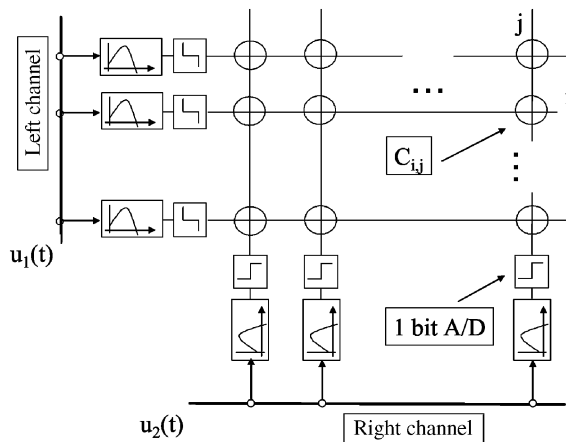


Fig. 6. Estimation architecture for stereoisis algorithm.

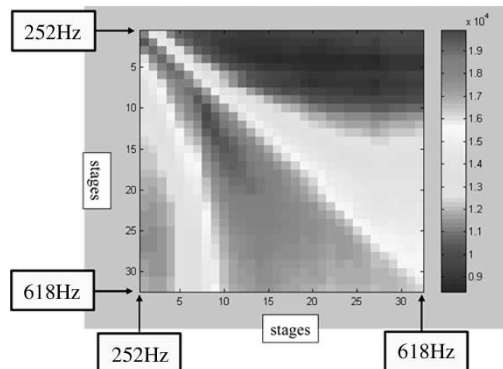


Fig. 7. Response of the stereoisis network to two signals with a relative delay.

D. Spatial Gradients Approach (SPGA)

In this approach, the signals recorded by the microphones are interpreted as samples of a field sound wave $s(\cdot)$ and the

bearing angle is estimated using first-order derivatives [10]. This algorithm, as opposed to the previous ones, takes full advantage of the four microphones for the time delay estimation. For the present situation let us consider the position of the microphones with respect to the center of the array. We will assume that for any given location \mathbf{r} in the plane, where $\mathbf{r} \in \mathbf{R}^2$, the magnitude $\tau(\mathbf{r})$ represents the time delay between the wavefront of the sound wave at \mathbf{r} and the wavefront of the sound wave at the center of the array. Using this convention, we can express the field $s(t + \tau(\mathbf{r}))$, $s : \mathbf{R}^1 \mapsto \mathbf{R}^1$ in a Taylor series around a neighborhood of the origin

$$s(t + \tau(\mathbf{r})) = s(t) + \tau(\mathbf{r}) \frac{d}{dt} s(t) + \frac{1}{2} \tau(\mathbf{r})^2 \frac{d^2}{dt^2} s(t) + O(\tau(\mathbf{r})^3). \quad (10)$$

To first order, and after geometric considerations, it can easily be seen that

$$\begin{aligned} x_1(t) &\approx s(t) + \tau_2 \dot{s} \\ x_4(t) &\approx s(t) + \tau_1 \dot{s} \\ x_3(t) &\approx s(t) - \tau_2 \dot{s} \\ x_2(t) &\approx s(t) - \tau_1 \dot{s} \end{aligned} \quad (11)$$

where $\tau_1 = (1/2)(d/C) \cos(\alpha)$, $\tau_2 = (1/2)(d/C) \cos(\beta)$ are the delays with respect to the coordinate axes. Then, a simple manipulation of the variables leads to

$$\begin{aligned} s(t) &= \frac{1}{4} (x_1(t) + x_3(t) + x_2(t) + x_4(t)) \\ \tau_1 \dot{s} &= \frac{1}{2} (x_4(t) - x_2(t)) \\ \tau_2 \dot{s} &= \frac{1}{2} (x_1(t) - x_3(t)). \end{aligned} \quad (12)$$

If we sample the signals with a sampling time $T_s = 1/f_s$, and assume that $\dot{s}(kT_s)$ can be adequately measured by filtering $s(kT_s)$, then, (12) is a standard least means-square (LMS) problem and τ_1 , τ_2 can be obtained independently after collecting $N + 1$ samples as²

$$\begin{aligned} \tau_1 &= (\mathbf{a}_1^T \mathbf{a}_1)^{-1} \mathbf{a}_1^T \mathbf{y}_1 \\ \tau_2 &= (\mathbf{a}_2^T \mathbf{a}_2)^{-1} \mathbf{a}_2^T \mathbf{y}_2 \end{aligned} \quad (13)$$

where

$$\begin{aligned} \mathbf{a}_1 &= \mathbf{a}_2 = \left[\dot{s}(k+j) \cdots \dot{s}(k+j+N) \right]^T \\ \mathbf{y}_1 &= [x_4(k+j) - x_2(k+j) \cdots x_4 \\ &\quad \times (k+j+N) - x_2(k+j+N)]^T \\ \mathbf{y}_2 &= [x_1(k+j) - x_3(k+j) \cdots x_1 \\ &\quad \times (k+j+N) - x_3(k+j+N)]^T \end{aligned} \quad (14)$$

and $j \geq 0$.

This approach heavily relies on the accuracy of the signals measurement, especially due to the need of an estimate of the derivative. For this reason, even though in this case the signal is sampled in time, its amplitude cannot be quantized. In practice, the original signal was used with the original sampling rate of 2048 samples/s, and the derivative was calculated using finite

²Similar results can be obtained using adaptive algorithms.

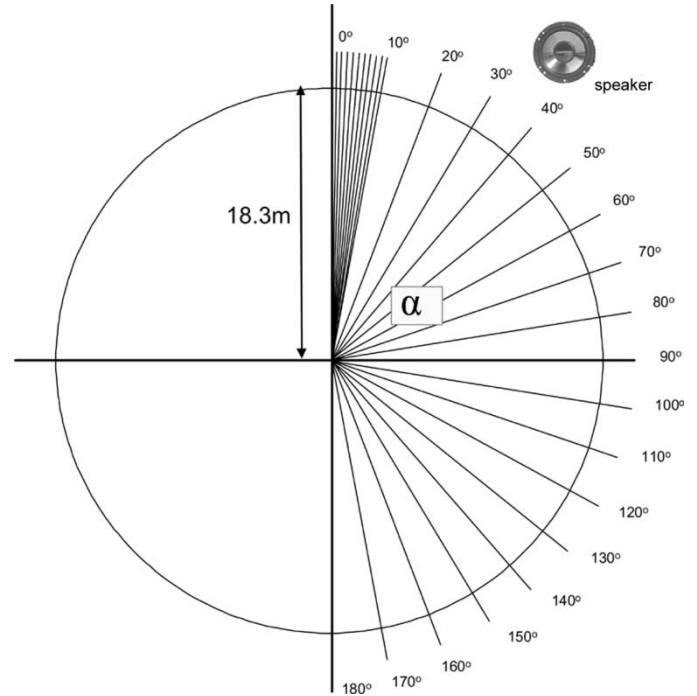


Fig. 8. Testing setup showing location of ASU, speaker, and measured angles.

differences. The derivative is obtained using a one-step difference equation $\dot{s}(k) \approx (s(k) - s(k-1))/T_s$. As is well known, this scheme might produce noise amplification at high frequencies. In this particular case, high frequencies are filtered by the anti-alias filter and by a high-order low-pass filter with cutoff frequency at 300 Hz.³ As can be seen from Fig. 13 (shown later), the amplitude of the signal is higher than the amplitude of the noise in most of the spectrum of interest ($f \in [0 \text{ Hz}, 800 \text{ Hz}]$). In the remaining part of the spectrum, both signal and noise are negligible. Integration of signal and noise power in the full frequency range $[0 \text{ Hz}, 1024 \text{ Hz}]$ indicates that the signal-to-noise ratio $\text{SNR}(s(w)) = 62$ and $\text{SNR}(\dot{s}(w)) = 220$. This observation agrees with the experimental results presented in [7]. Nevertheless, it must be pointed out that this scheme could produce noise amplification, for example, if noise conditions were different, or the signals were narrowband instead of broad-band, or the sampling rate were higher (producing a greater bandwidth). In these cases, more elaborate calculation schemes for the derivative should be chosen [19].

IV. EXPERIMENTS AND NUMERICAL RESULTS

In order to design and test the different algorithms, experimental data were collected in a field test in a public park in Severna Park, MD. The ASU was located in the center of a field, and one 30.5-cm (12 in) subwoofer was placed 18.3 m away. A Gaussian white noise signal was played through the speaker and the signals received at the four microphones were recorded using a sampling time of 2048 samples/s. For every angle, we played 30 s of data and obtained 30 different readings of time delay, corresponding to different estimations during a 1-s time window. Two different sets of data were collected. One set of

³This is the result of a second-order low-pass filter plus the low-pass filter action of the microphone itself.

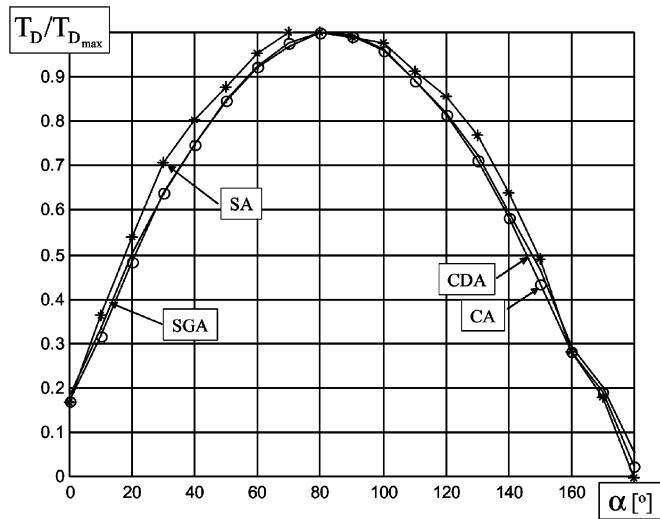


Fig. 9. Normalized mean value of the estimated time delay for the four algorithms in the range $[0^\circ, 180^\circ]$.

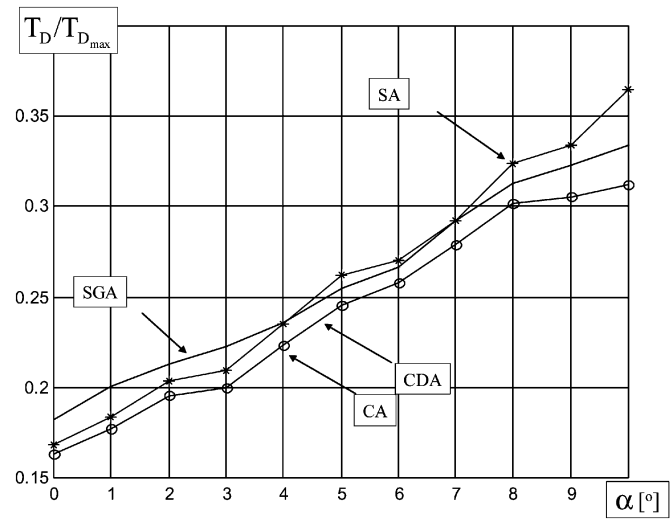


Fig. 11. Normalized mean value of the estimated time delay for the four algorithms in the range $[0^\circ, 10^\circ]$.

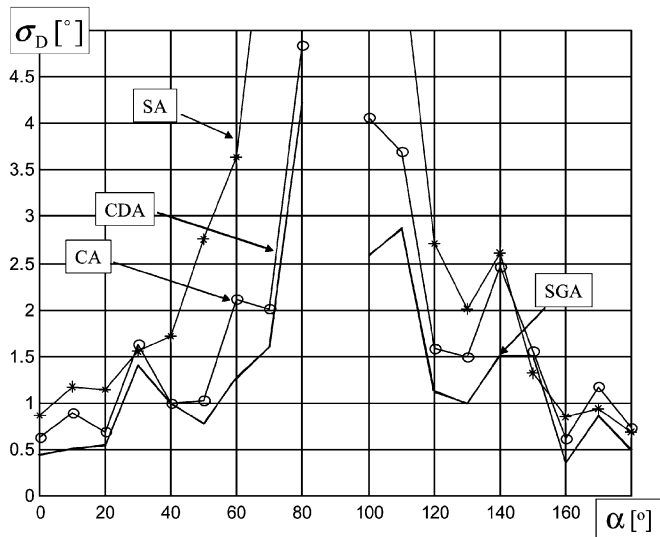


Fig. 10. Standard deviation in degrees for the four algorithms in the range $[0^\circ, 180^\circ]$.

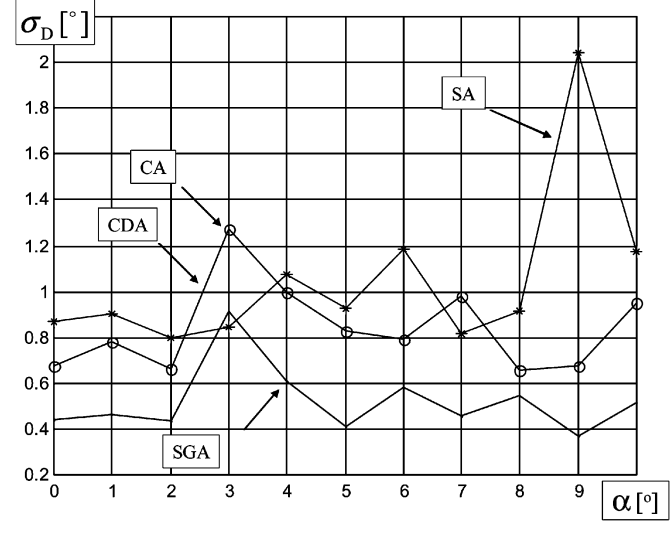


Fig. 12. Standard deviation in degrees for the four algorithms in the range $[0^\circ, 10^\circ]$.

data corresponding to angles in the range $[0^\circ, 180^\circ]$ in steps of 10° ; the other set of data corresponding to angles in the range $[0^\circ, 10^\circ]$ in steps of 1° (see Fig. 8). For every angle, we used the mean to define the transfer curve time delay versus angle, and the standard deviation to quantify the precision. The SNR has an approximate mean value of 60 in the range of interest ($f \in [20 \text{ Hz}, 300 \text{ Hz}]$).

As shown in the Appendix, the time-delay variation corresponding to a change of 1° at an angle β^* is

$$\Delta T_D|_{1^\circ} = \frac{T_{D_{max}} \sin(\beta^*)\pi}{180}. \quad (15)$$

However, in the present case, it is useful to quantify the error in degrees. Accordingly, using (15), we can show that if the reading of a certain time delay has a standard deviation of σ_T , then, the standard deviation in degrees is given by

$$\sigma_D := \frac{\sigma_T}{\frac{T_{D_{max}} \sin(\beta^*)\pi}{180}}. \quad (16)$$

TABLE I
ACCURACY OF ALGORITHMS (STD) IN DEGREES

Range	CA	CDA	SA	SGA
0-180	1.18	1.18	1.47	0.87
0-10	0.85	0.85	1.05	0.25

Figs. 9 and 10 show the mean value (normalized with respect to $T_{D_{max}}$) and standard deviation in degrees, resp., of the estimated time delay corresponding to the four algorithms in the range $[0^\circ, 180^\circ]$. The asymmetric characteristics seen in Figs. 9 and 10 are mainly due to misalignment in the location of the speaker with respect to the ASU. Figs. 11 and 12 show the mean values and the standard deviation, resp., corresponding to the four algorithms in the range $[0^\circ, 10^\circ]$. As can be appreciated from Figs. 9 and 11, the CA and CDA give indistinguishable results. Finally, Table I summarizes the average standard deviation of the four algorithms in both ranges.

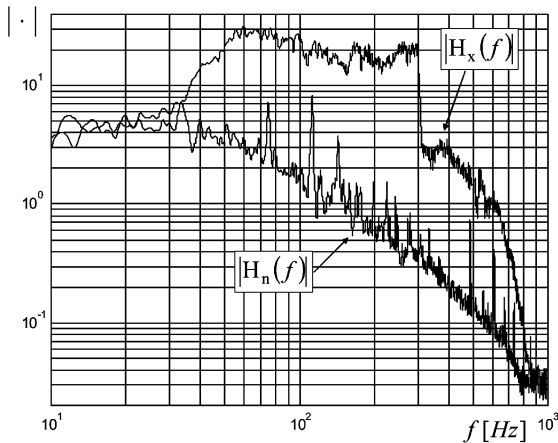


Fig. 13. Spectral amplitude of recorded ambient noise and a recorded signal.

V. REMARKS ABOUT ACCURACY

A natural question in the present study is: what is the maximum accuracy that can be achieved in the estimation of the time delay for a given recorded data? The answer is given by the Cramér–Rao lower bound (CRLB) that gives a lower bound for the variance of the time delay estimation as a function of the signal and noise power spectra. The CRLB is given by (see [12])

$$\sigma \geq \left[4\pi^2 T \int_0^B f^2 \frac{C_{AB}(f)}{1 - C_{AB}(f)} df \right]^{-1} \quad (17)$$

where T is the integration time, B is the bandwidth of the signal, and

$$C_{AB} = \left| \frac{R_{x_1 x_3}(f)}{R_{x_1 x_1}(f) R_{x_3 x_3}(f)} \right|^2 \quad (18)$$

is the so-called squared magnitude coherence (see [6]). Here, $R_{x_1 x_1}(f)$ and $R_{x_3 x_3}(f)$ are the power-spectral density of signals x_1 , x_3 , and $R_{x_1 x_3}(f)$ is the cross power-spectral density of signals x_1 and x_3 . Using the relations between signals and noise, we can rewrite (18) as

$$C_{AB} = \frac{R_{ss}(f)^2}{(R_{ss}(f) + R_{n_1 n_1}(f))(R_{ss}(f) + R_{n_3 n_3}(f))}. \quad (19)$$

We will assume that the noise at both microphones have the same power spectra distribution, i.e., $R_{n_1 n_1} = R_{n_3 n_3} = R_{nn}$, and that the signals received at both microphones also have the same power spectra distribution, i.e., $R_{x_1 x_1} = R_{x_3 x_3} = R_{xx}$. To calculate (18), we apply Welch's averaged periodogram method⁴ [20] to the measured data and obtain $R_{nn}(f) = |H_n(f)|^2$ and $R_{xx}(f) = |H_x(f)|^2$. Fig. 13 shows $|H_n(f)|$ and $|H_x(f)|$, where $H_n(f)$ corresponds to an experiment with ambient noise and no signal, and $H_x(f)$ corresponds to an experiment with both ambient noise and signal. After noticing that $R_{xx} = R_{ss} + R_{nn}$, some algebraic manipulation leads to the following expression for the CRLB as a function of only measured variables

$$\sigma \geq \left[4\pi^2 T \int_0^B f^2 \frac{\left(\frac{-1 + R_{xx}(f)}{R_{nn}(f)} \right)^2}{\left(\frac{-1 + 2R_{xx}(f)}{R_{nn}(f)} \right)} df \right]^{-1}. \quad (20)$$

⁴The length of the signals is 2^{16} . The fast Fourier transform length is 2^{14} and the Hanning window used has a length of 2^{11} .

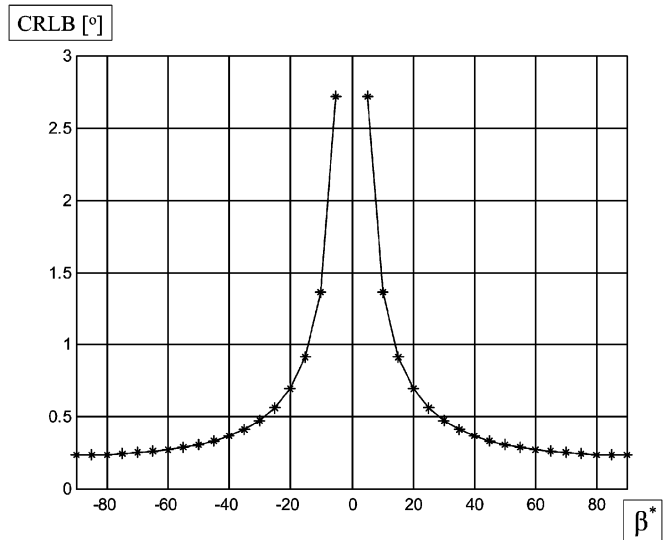


Fig. 14. CRLB in degrees as a function of the sound source position.

We calculated (20) using all the available data, considering that $T = 1$ s, $B = 300$ Hz, and we obtained a mean value $\overline{\text{CRLB}} = 1.66 \mu\text{s}$, with standard deviation $\sigma(\text{CRLB}) = 0.16 \mu\text{s}$ and a maximum value $\text{CRLB}^* = 1.9 \mu\text{s}$. We can translate this bound on the time-delay estimation accuracy into an equivalent bound on degree estimation accuracy by using the relation (22) derived in the Appendix. Accordingly, the minimum error in degrees as a function of the position of the source is given by

$$\text{CRLB}^*(\beta) = \frac{1.9 \mu\text{s}}{\frac{T_D \max \sin(\beta)\pi}{180}}. \quad (21)$$

This relation is plotted in Fig. 14 and shows that it is possible to achieve less than 1° error in the estimation. Actually, the worst case error for all angles is 0.33° . This observation is due to the fact that 0.33° is the maximum error in the range $[0^\circ, 45^\circ] \cup [135^\circ, 180^\circ]$, and the same applies to the range $[45^\circ, 135^\circ]$ considering the other pair of microphones. Therefore, if the two pair of microphones are used in their best range, the bound for all angles is 0.33° . The reader should notice that this is only a lower bound, and algorithms will exhibit in general greater errors.

VI. CONCLUSION

We have compared four different algorithms for sound localization using MEMS microphones and signals recorded in a natural environment. Two of the algorithms were previously reported in the literature, and the other two were developed specifically for this application. The SGA shows the best accuracy results and its implementation requires a sampled data analog architecture able to solve adaptively an LMS problem [7]. The SA shows acceptable results but requires a two-dimensional array of correlators in addition to the two cochlea filter channels (a simplified approach has been implemented in [21]). Finally, the CDA shows an accuracy very close to the SGA while offering a very convenient architecture, evidenced not only by its simplicity but also by the associated low temporal activity. Experimental results of an integrated circuit in standard CMOS technology that implements this approach can be found in [9].

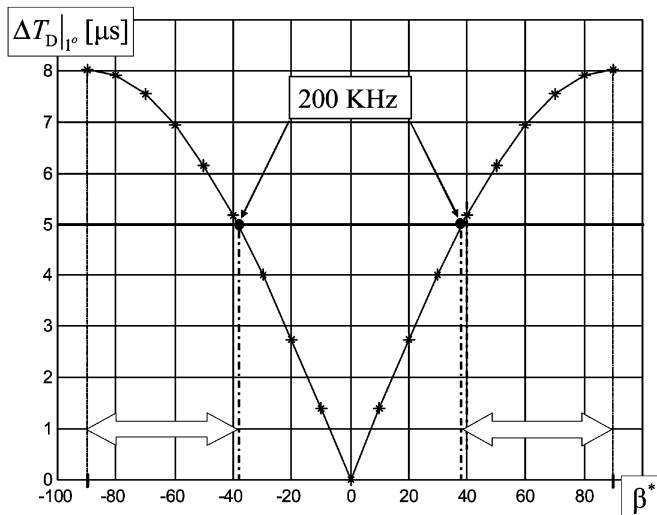


Fig. 15. Sampling time for a one degree discrimination as a function of the angle. The arrows show the range of interest.

All presented methods constitute an alternative to applications in sensor networks where power consumption is restricted. Moreover, they all implement coherent strategies at the sensor level, eliminating the need for synchronization.

APPENDIX DETERMINATION OF SAMPLING FREQUENCY

Let us consider a variation of the time delay (2) produced by a variation of the angle around a given angle β^*

$$\Delta T_D = T_{D \max} \sin(\beta^*) \Delta \beta. \quad (22)$$

Now, if we consider the time delay variation corresponding to a variation of one degree, i.e., $\Delta \beta = 1^\circ$, we obtain

$$\Delta T_D|_{1^\circ} = \frac{T_{D \max} \sin(\beta^*) \pi}{180}. \quad (23)$$

This formula basically tells us that to discriminate one degree at an angle β^* is equivalent to $\Delta T_D|_{1^\circ}$ s. A plot of $\Delta T_D|_{1^\circ}$ versus the angle is shown in Fig. 15. It can also be seen in this picture that a sampling frequency of $f_s = 200$ kHz permits to discriminate one degree for the range of angles $\beta \in [-90, -40] \cup [+40, +90]$, or equivalently $\alpha \in [0, 50] \cup [+130, +180]$. In practice, this is enough since there is another pair of microphones in quadrature, so whenever the angle is not in the range just mentioned, the algorithm can use the other pair of microphones. In this way, the angle is always in the appropriate range.

ACKNOWLEDGMENT

The authors want to thank Dr. P. Pouliquen for his helpful discussions regarding the correlation derivative approach implementation issues, and R. Rosasco for recording the signals.

REFERENCES

- [1] H. Gharavi and S. P. Kumar, "Scanning the issue: Special issue on sensor networks and applications," *Proc. IEEE*, vol. 91, pp. 1151–1153, Aug. 2003.
- [2] R. Showen, "Operational gunshot location system," *Proc. SPIE*, vol. 2935, pp. 130–139, Feb. 1997.

- [3] 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.
- [4] J. C. Chen, K. Yao, and R. E. Hudson, "Source localization and beam-forming," *IEEE Signal Processing Mag.*, pp. 30–39, Mar. 2002.
- [5] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, classification, and tracking of targets," *IEEE Signal Processing Mag.*, pp. 17–29, Mar. 2002.
- [6] G. C. Carter, "Time delay estimation for passive sonar signal processing," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-29, pp. 463–470, June 1981.
- [7] M. Stanacevic and G. Cauwenberghs, "Mixed-signal gradient flow bearing estimation," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 1, 2003, pp. 777–780.
- [8] —, "Micropower mixed-signal acoustic localizer," in *Proc. IEEE Eur. Solid State Circuits Conf.*, Estoril, Portugal, 2003.
- [9] P. Julian, A. G. Andreou, P. Mandolesi, and D. Goldberg, "A low-power CMOS integrated circuit for bearing estimation," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 5, 2003, pp. 305–308.
- [10] G. Cauwenberghs, M. Stanacevic, and G. Zweig, "Blind broad-band source localization and separation in miniature sensor arrays," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 3, 2001, pp. 193–196.
- [11] S. Shamma, N. Shen, and P. Gopalaswamy, "Stereaussis: Binaural processing without neural delays," *J. Acoust. Soc. Amer.*, vol. 86, pp. 989–1006, 1989.
- [12] G. C. Carter, "Coherence and time delay estimation," *Proc. IEEE*, vol. 75, pp. 236–255, Feb. 1987.
- [13] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-24, pp. 320–327, Aug. 1976.
- [14] A. H. Quazi, "An overview on the time delay estimate in active and passive systems for target localization," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-29, pp. 527–533, June 1981.
- [15] A. Moiseff and M. Konishi, "Neuronal and behavioral sensitivity to binaural time differences in the owl," *J. Neurosci.*, vol. 1, pp. 40–48, 1981.
- [16] J. P. Lazzaro and C. Mead, "Silicon models of auditory localization," *Neural Computation*, vol. 1, pp. 41–70, 1989.
- [17] S. Shamma, R. Chadwick, J. Wilbur, K. Moorish, and J. Rinzl, "A biophysical model of cochlear processing: Intensity dependence of pure tone responses," *J. Acoust. Soc. Amer.*, vol. 80, pp. 133–145, 1986.
- [18] E. Fragniere, A. V. Schaik, and E. Vittoz, "Design of an analogue VLSI model of an active cochlea," *Anal. Integr. Circuits Signal Processing*, vol. 12, pp. 19–35, 1997.
- [19] S. Diop, J. W. Grizzle, P. E. Moraal, and A. Stefanopoulou, "Interpolation and numerical differentiation for observer design," in *Proc. Amer. Control Conf.*, 1994, pp. 1329–1333.
- [20] P. D. Welch, "The use of fast fourier transforms for the estimation of power spectra: A method based on time averaging over short modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, pp. 70–73, 1967.
- [21] A. van Schaik and S. Shamma, "A neuromorphic sound localizer for a smart MEMS system," in *Proc. IEEE Int. Symp. Circuits Systems*, vol. 4, 2003, pp. 864–867.



Pedro Julián (S'94–M'00) was born in Bahía Blanca, Argentina, on June 28, 1970. He received the Ingeniero Electrónico degree and the Ph.D. degree in control de sistemas, from Universidad Nacional del Sur (UNS), Bahía Blanca, Argentina, in 1994 and 1999, respectively.

From 2000 to 2002, he was a Visiting Scholar in the Nonlinear Electronics Laboratory, University of California at Berkeley. From 2002 to 2003, he was a Visiting Scholar in the Sensory Communication and Microsystems Laboratory, The Johns Hopkins University, Baltimore, MD. He has been an Assistant Professor in the Departamento de Ingeniería Eléctrica y Computadoras (DIEC), UNS, since 2003. He also holds a position as an Assistant Researcher in the National Research Council of Argentina (CONICET) since 2002. His research interests areas are computation, circuits and systems theory and design. In particular, he is interested in practical and theoretical aspects of computation structures, including parallel systems like cellular neural networks and cellular automata, with a special emphasis in the design and analysis of very large-scale integration systems.

Dr. Julián serves as the Region 9 (Latin America) Vice President of the IEEE Circuits and Systems Society for 2004 and 2005.



Andreas G. Andreou received the Ph.D. degree in electrical engineering and computer science from The Johns Hopkins University, Baltimore, MD, in 1986.

Between 1986 and 1989, he held Post-Doctoral Fellow and Associate Research Scientist positions in the electrical and computer engineering department while also a member of the professional staff at the Johns Hopkins Applied Physics Laboratory. He became an Assistant Professor of Electrical and Computer Engineering in 1989, Associate

Professor in 1993, and Professor in 1996. In 1995 and 1997, he was a Visiting Associate Professor and Visiting Professor, respectively, in the computation and neural systems program at California Institute of Technology, Pasadena. In the summer of 2001, he was a Visiting Professor at Tohoku University, Tohoku, Japan, working on three-dimensional integration. He is a member of the Whitaker Bioengineering Institute, The Johns Hopkins University, and Founding Director of the Whitaker Fabrication and Lithography facility. He is the co-founder of the Center for Language and Speech Processing, The Johns Hopkins University. His research interests include integrated circuits, sensory information processing, and neural computation. He is a co-editor of the book *Adaptive Resonance Theory Microchips* (Norwell, MA: Kluwer, 1998). He is Associate Editor of the journal *Neural Networks*.

He now serves as a Member of the Board of Governors for the IEEE Circuits and Systems (CAS) Society. Dr. Andreou is a recipient of a National Science Foundation Research Initiation Award. He is the recipient of the 2000 IEEE CAS Society Darlington Award.



Larry Riddle received the B.S. degree in applied physics (with highest honors) from the Georgia Institute of Technology, Atlanta, and the M.S. and Ph.D. degrees in electrical and computer engineering, from The Johns Hopkins University, Baltimore, MD, in 1980, 1984, and 1987, respectively.

In 1995, he founded Signal Systems Corporation, Annapolis, MD, a high-technology small business specializing in acoustic signal processing, active noise control and integrated RF electronics signal processing. Under his leadership, the company

has performed extensive research and engineering services to develop state-of-the-art processing solutions for more than two dozen government and commercial customers and is a recognized leader in platform noise reduction and acoustic surveillance. Prior to 1995, Dr. Riddle served as Manager of Signal Processing Algorithms for Martin Marietta Laboratories and Program Manager for Martin Marietta Aero and Naval Systems where he was the Principal Investigator for numerous studies in acoustic signal processing and active noise control. He is also the Program Manager of the DARPA sponsored Smart Microphone Program. He has published numerous articles and holds a patent for method and apparatus for acoustic attenuation.

Dr. Riddle is a peer reviewer for the IEEE SENSORS JOURNAL and the *Journal of Underwater Acoustics* on topics of active sonar classification and adaptive beamforming.



Shihab Shamma received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1980.

He joined the Department of Electrical Engineering, University of Maryland, College Park, in 1984, where his research has dealt with issues in computational neuroscience and the development of microsensor systems for experimental research and neural prostheses. His primary focus has been on uncovering the computational principles underlying the processing and recognition sounds (speech and

music) in the auditory system and the relationship between auditory and visual processing, and includes the development of photolithographic arrays for recording and stimulation of neural signals, implementations of auditory processing algorithms, and algorithms for the detection, classification, and activity from multiple simultaneous sources.



David H. Goldberg received the B.Sc. degree from Brown University, Providence, RI, in 1998, the M.S.E. degree in electrical and computer engineering, The Johns Hopkins University, Baltimore, MD, in 2000, and is currently working toward the Ph.D. degree at the same university.

His research interests include computational neuroscience, signal processing, and very large-scale integration design.



Gert Cauwenberghs (S'89–M'94) received the Ph.D. degree in electrical engineering from the California Institute of Technology (Caltech), Pasadena, in 1994.

He is presently a Professor of Electrical and Computer Engineering at The Johns Hopkins University, Baltimore, MD. He was a Visiting Professor of Brain and Cognitive Science at the Massachusetts Institute of Technology, Cambridge, in 1999. His research covers very large-scale integration circuits, systems and algorithms for parallel signal processing,

adaptive neural computation, and low-power coding and instrumentation. He recently co-edited the book *Learning on Silicon* (Norwell, MA: Kluwer, 1999).

He was Francqui Fellow of the Belgian American Educational Foundation in 1988, and received the National Science Foundation Career Award in 1997, the Office of Naval Research Young Investigator Award in 1999, and the Presidential Early Career Award for Scientists and Engineers in 2000. He is an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II, and the IEEE SENSORS JOURNAL. He was Publication Chair at IEEE MWSCAS 2000, Analog Track Chair at IEEE International Symposium on Circuits and Systems (ISCAS) 2002, and Invited Sessions Co-Chair at ISCAS 2003. He chaired the Analog Signal Processing Technical Committee of the IEEE Circuits and Systems Society in 2001.