FIRST-ORDER DIFFERENTIAL MICROPHONE ARRAYS FROM A TIME-DOMAIN BROADBAND PERSPECTIVE

Yaakov Buchris and Israel Cohen

Technion, Israel Institute of Technology
Technion City, Haifa 32000, Israel
{bucris@tx, icohen@ee}.technion.ac.il

Jacob Benesty

INRS-EMT, University of Quebec
800 de la Gauchetiere Ouest, Suite 6900
Montreal, QC H5A 1K6, Canada
benesty@emt.inrs.ca

ABSTRACT

Differential microphone arrays (DMAs) have a great potential to overcome some of the problems of additive arrays and provide high spatial gain relative to their small size. In this work, we present a time-domain formulation for implementing first-order DMAs, which is very important for some applications in which minimal delay is required, such as real-time communications. We present a design example for first-order DMAs illustrating some of the fundamental properties of the time-domain implementation as well as the equivalence to the frequency-domain implementation.

Index Terms— Microphone arrays, differential microphone arrays (DMAs), time-domain broadband beamforming.

1. INTRODUCTION

Differential microphone arrays (DMAs) can be integrated into several real-world beamforming applications involving speech signals, e.g., hands-free telecommunication. DMAs are characterized as superdirective [3] in general, small-size arrays, whose beampattern is almost frequency invariant, leading to greatly intelligible speech signals even in heavy reverberant and noisy environments. Due to these benefits, DMAs have attracted a significant amount of interest in the field of broadband microphone array processing during the past decade (see [4] - [11] and the references therein).

Broadband array processing algorithms can be implemented both in the time and frequency domains. Design in the time domain is important for applications that require small delays such as real-time communications [13]. Second, processing in the time domain circumvents the edge effects between successive snapshots of the incoming signals. Furthermore, in some cases the implementation of time-domain filters is computationally more efficient than the equivalent frequency-domain filters, especially when short filters are sufficient. The advantage of frequency-domain implementation is mainly due to the ability to implement some frequency dependent processing algorithms, like frequency-selective null-steering and efficient calculation of the sample matrix inversion (SMI) [12] used in several adaptive array processing applications.

In this work, we present a framework for a broadband time-domain implementation of first-order DMAs. First, the input array signals are manipulated and represented in a separable form as a product between a desired signal dependent term and a second term which depends only on the array geometry. This representation is beneficial because it enables to apply several array processing algorithms, originally developed in the frequency domain, into broadband time-domain DMAs. We derive a closed-form solution for time-domain first-order DMAs for any given number of sensors. Due to the DMA assumption, the derived solution is very simple with respect to other methods usually employed in the design of general arrays where some constraints that ensure the frequency invariance should be imposed. We also establish the equivalent time-domain expressions for several commonly-used quality measures like the beampattern, white noise gain (WNG), and directivity factor (DF). Finally, we evaluate the performance of the time-domain DMAs and compare it with that of the frequency-domain implementation recently proposed by Benesty et al. [5].

2. SIGNAL MODEL

We consider a broadband source signal, $s(n)$, in the far-field, where n is the discrete-time index, that propagates in an anechoic acoustic environment at the speed of sound, i.e., $c = 340$ m/s, and impinges on a uniform linear sensor array consisting of M omnidirectional microphones, where the distance between two successive sensors is equal to δ. The direction of the source signal to the array is parameterized by the angle θ, where $\theta = 0^\circ$ corresponds to the endfire direction. In the rest, microphone 1 is chosen as the reference sensor. In this scenario, the signal measured at the mth microphone is

This Research was supported by Qualcomm Research Fund and MAFAAT-Israel Ministry of Defense.

978-1-5090-2007-2/16/$31.00

given by

$$y_m(n) = s\left[n - \Delta - f_s\tau_m(\theta)\right] + v_m(n), \quad (1)$$

where Δ is the propagation time from the position of the source $s(t)$ to sensor 1, f_s is the sampling frequency, $\tau_m(\theta) = (m-1)\frac{\delta \cos\theta}{c}$ is the delay between the first and the mth microphone which can be either positive or negative, and $v_m(n)$ is the noise picked up by the mth sensor. We can also express (1) as

$$y_m(n) = \mathbf{g}_m^T(\theta)\mathbf{s}\left(n - \Delta\right) + v_m(n), \quad (2)$$

where the superscript ^{T} is the transpose operator, and $\mathbf{g}_m(\theta)$ is a vector containing causal fractional delay filter coefficients [14] with a maximum value in the location suitable to the delay between the mth sensor and the reference sensor. The length of the vector $\mathbf{g}_m(\theta)$ is $L_g = 2L_D + L_{fd} + 1$ where L_{fd} is the length of the causal fractional delay filter, and $L_D = \lceil \frac{(M-1)\delta f_s}{c} \rceil$ is the maximal delay between the reference sensor and the extreme sensor. Note also that we introduce non-causality to the vector $\mathbf{g}_m(\theta)$ in order to support the scenario of signals that propagate from directions in which the signal arrives to the sensors before it arrives to the reference sensor. The signal vector $\mathbf{s}\left(n - \Delta\right)$ is a vector containing L_g successive samples of the signal $s\left(n - \Delta\right)$.

By considering L_h successive time samples of the mth microphone signal, (2) becomes a vector of length L_h:

$$\mathbf{y}_m(n) = \mathbf{G}_m(\theta)\mathbf{s}_L\left(n - \Delta\right) + \mathbf{v}_m(n), \quad (3)$$

where $\mathbf{G}_m(\theta)$ is a Sylvester matrix of size $L_h \times L$ created from the vector $\mathbf{g}_m(\theta)$, with $L = L_g + L_h - 1$, the vector $\mathbf{s}_L\left(t - \Delta\right)$ is the signal vector of length L containing the signal samples, and $\mathbf{v}_m(n)$ is a vector of length L_h containing the noise samples.

Now, by concatenating the observations from the M microphones, we get a vector of length ML_h:

$$\begin{aligned} \underline{\mathbf{y}}(n) &= \left[\begin{array}{cccc} \mathbf{y}_1^T(n) & \mathbf{y}_2^T(n) & \cdots & \mathbf{y}_M^T(n) \end{array}\right]^T \\ &= \underline{\mathbf{G}}(\theta)\mathbf{s}_L\left(n - \Delta\right) + \underline{\mathbf{v}}(n), \end{aligned} \quad (4)$$

where

$$\underline{\mathbf{G}}(\theta) = \left[\begin{array}{c} \mathbf{G}_1(\theta) \\ \mathbf{G}_2(\theta) \\ \vdots \\ \mathbf{G}_M(\theta) \end{array}\right] \quad (5)$$

is a matrix of size $ML_h \times L$ and

$$\underline{\mathbf{v}}(n) = \left[\begin{array}{cccc} \mathbf{v}_1^T(n) & \mathbf{v}_2^T(n) & \cdots & \mathbf{v}_M^T(n) \end{array}\right]^T \quad (6)$$

is a vector of length ML_h.

Like in DMAs, we assume that δ is much small with respect to the wavelength of the signal, and the desired signal propagates at the endfire, so that the observations are

$$\underline{\mathbf{y}}(t) = \underline{\mathbf{G}}(0)\mathbf{s}_L\left(t - \Delta\right) + \underline{\mathbf{v}}(t). \quad (7)$$

Then, our objective is to design all kind of broadband DMAs, where the main lobe is at the angle $\theta = 0$, with a real-valued spatiotemporal filter of length ML_h:

$$\underline{\mathbf{h}} = \left[\begin{array}{cccc} \mathbf{h}_1^T & \mathbf{h}_2^T & \cdots & \mathbf{h}_M^T \end{array}\right]^T, \quad (8)$$

where $\mathbf{h}_m,\ m = 1, \ldots, M$ are temporal filters of length L_h.

3. BROADBAND BEAMFORMING

By applying the filter $\underline{\mathbf{h}}$ to the observation vector $\underline{\mathbf{y}}(n)$, we obtain the output of the broadband beamformer:

$$z(n) = \sum_{m=1}^{M} \mathbf{h}_m^T \mathbf{y}_m(n) = \underline{\mathbf{h}}^T \underline{\mathbf{y}}(n) = x_{fd}(n) + v_{rn}(n), \quad (9)$$

where

$$\begin{aligned} x_{fd}(n) &= \sum_{m=1}^{M} \mathbf{h}_m^T \mathbf{G}_m(0)\mathbf{s}_L\left(n - \Delta\right) \\ &= \underline{\mathbf{h}}^T \underline{\mathbf{G}}(0)\mathbf{s}_L\left(n - \Delta\right) \end{aligned} \quad (10)$$

is the filtered desired signal and

$$v_{rn}(n) = \sum_{m=1}^{M} \mathbf{h}_m^T \mathbf{v}_m(n) = \underline{\mathbf{h}}^T \underline{\mathbf{v}}(n) \quad (11)$$

is the residual noise. We see from (10) that our desired signal is $s\left(n - \Delta\right)$. Therefore, the distortionless constraint is

$$\underline{\mathbf{h}}^T \underline{\mathbf{G}}(0) = \mathbf{i}^T, \quad (12)$$

where $\mathbf{i}$ is a column vector of length L with all its elements equal to zero except for the $(L_D + 1)$th element. This constraint is always required in the design of DMAs.

4. DESIGN OF FIRST-ORDER DMAS

It is well known that the design of a first-order DMA requires at least $M \geq 2$ microphones [5], [6]. For first-order design, we have two constraints to fulfill; the distortionless one given in (12) and a constraint with a null in the direction $\theta_1 \in \left[\frac{\pi}{2}, \pi\right]$, i.e.,

$$\underline{\mathbf{h}}^T \underline{\mathbf{G}}(\theta_1) = \mathbf{0}^T, \quad (13)$$

where $\mathbf{0}$ is a zero vector of length L. Combining these two constraints together, we get the following linear system:

$$\left[\begin{array}{cccc} \mathbf{G}_1^T(0) & \mathbf{G}_2^T(0) & \cdots & \mathbf{G}_M^T(0) \\ \mathbf{G}_1^T(\theta_1) & \mathbf{G}_2^T(\theta_1) & \cdots & \mathbf{G}_M^T(\theta_1) \end{array}\right] \underline{\mathbf{h}} = \underline{\mathbf{i}}_1, \quad (14)$$

or, equivalently,

$$\mathbf{C}_{1,M}(\boldsymbol{\theta})\underline{\mathbf{h}} = \underline{\mathbf{i}}_1, \quad (15)$$

where $\mathbf{C}_{1,M}(\boldsymbol{\theta})$ is a matrix of size $2L \times ML_h$, and

$$\underline{\mathbf{i}}_1 \triangleq \begin{bmatrix} \mathbf{i} \\ \mathbf{0} \end{bmatrix} \tag{16}$$

is a vector of length $2L$. We can solve (15) using the pseudo-inverse of $\mathbf{C}_{1,M}(\boldsymbol{\theta})$:

$$\underline{\mathbf{h}} = \mathbf{P}^{\dagger}_{\mathbf{C}_{1,M}}(\boldsymbol{\theta})\underline{\mathbf{i}}_1, \tag{17}$$

where

$$\mathbf{P}^{\dagger}_{\mathbf{C}_{1,M}}(\boldsymbol{\theta}) = \left[\mathbf{C}^T_{1,M}(\boldsymbol{\theta})\mathbf{C}_{1,M}(\boldsymbol{\theta}) + \lambda\mathbf{I}\right]^{-1}\mathbf{C}^T_{1,M}(\boldsymbol{\theta}) \tag{18}$$

is the pseudo-inverse of the matrix $\mathbf{C}_{1,M}(\boldsymbol{\theta})$ and the scalar λ is a regularization small parameter which provide stable inversion of the matrix. Later, in the simulation section we show that this simple solution yields a frequency-invariant beampattern although no specific constraints were imposed. This is due to the fact that we deal with the DMA model which inherently provide the frequency-invariance property.

5. PERFORMANCE MEASURES

Herein, we establish some measures which we use in the simulation section in order to assess the performance. Assuming microphone 1 to be the reference sensor, the gain in signal-to-noise ratio (SNR) is

$$\mathcal{G}(\underline{\mathbf{h}}) = \frac{\text{oSNR}(\underline{\mathbf{h}})}{\text{iSNR}} = \frac{\underline{\mathbf{h}}^T\underline{\mathbf{G}}(0)\underline{\mathbf{G}}^T(0)\underline{\mathbf{h}}}{\underline{\mathbf{h}}^T\mathbf{\Gamma}_{\underline{\mathbf{v}}}\underline{\mathbf{h}}}. \tag{19}$$

where $\mathbf{\Gamma}_{\underline{\mathbf{v}}} = \frac{\mathbf{R}_{\underline{\mathbf{v}}}}{\sigma^2_{v_1}}$ is the pseudo-correlation matrix of $\underline{\mathbf{v}}(t)$.

The WNG is obtained by taking $\mathbf{\Gamma}_{\underline{\mathbf{v}}} = \mathbf{I}_{ML_h}$, where $\mathbf{I}_{ML_h}$ is the $ML_h \times ML_h$ identity matrix, i.e.,

$$\mathcal{W}(\underline{\mathbf{h}}) = \frac{\underline{\mathbf{h}}^T\underline{\mathbf{G}}(0)\underline{\mathbf{G}}^T(0)\underline{\mathbf{h}}}{\underline{\mathbf{h}}^H\underline{\mathbf{h}}}. \tag{20}$$

We can also define the broadband beampattern or broadband directivity pattern as

$$|\mathcal{B}(\underline{\mathbf{h}},\theta)|^2 = \underline{\mathbf{h}}^T\underline{\mathbf{G}}(\theta)\underline{\mathbf{G}}^T(\theta)\underline{\mathbf{h}}. \tag{21}$$

Finally, we define the DF of the array which is the gain in SNR for the case of spherical diffuse noise. One way to calculate the DF is to use (19) and substitute the time-domain version of $\mathbf{\Gamma}_{\underline{\mathbf{v}}}$ for diffuse noise. Yet, an explicit expression for $\mathbf{\Gamma}_{\underline{\mathbf{v}}}$ in the time domain is unavailable. Instead, we can use directly the definition of the DF (see for example [15, ch.2]):

$$\mathcal{D}(\underline{\mathbf{h}}) = \frac{2}{\int_0^{\pi}|\mathcal{B}(\underline{\mathbf{h}},\theta)|^2\sin\theta d\theta}, \tag{22}$$

where $\mathcal{B}(\underline{\mathbf{h}},\theta)$ is defined in (21).

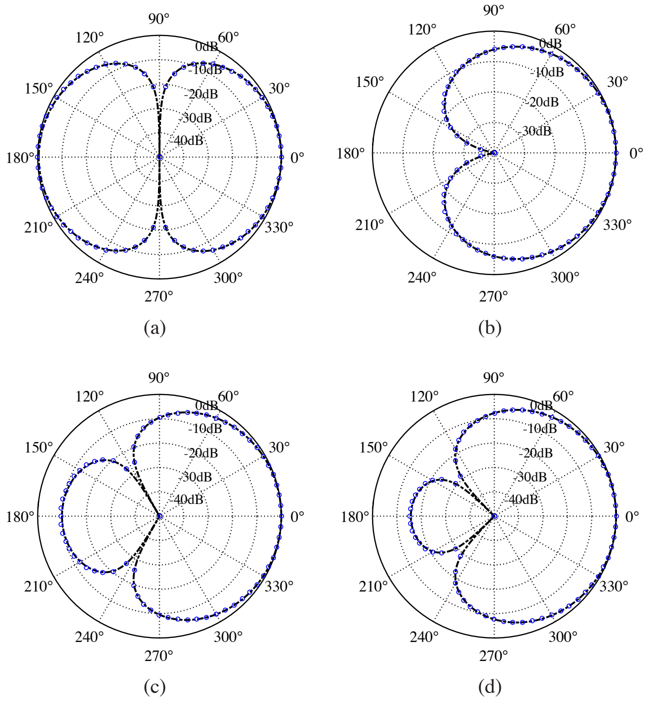

Fig. 1: Beampatterns for the four basic shapes of first-order DMAs produced by the time-domain implementation (dark dashed line): (a) dipole, (b) cardioid, (c) hypercardioid, and (d) supercardioid. The theoretical patterns are also presented (blue circles line).

6. A DESIGN EXAMPLE

In this section, we study the design of first-order standard DMA directivity patterns: dipole, cardioid, hypercardioid, and supercardioid, each with one distinct null in the following directions: $\theta_{\text{Dp}} = \frac{\pi}{2}, \theta_{\text{Cd}} = \pi, \theta_{\text{Hc}} = \frac{2\pi}{3}$, and $\theta_{\text{Sc}} = \frac{3\pi}{4}$. We choose a sensor spacing of $\delta = 1$cm and examine the case of $M = 2$ sensors. We choose $L_{\text{fd}} = 7$ taps and get $L_g = 10$ taps. We choose the filter length $L_h = 12$ taps and the sampling frequency to be $f_s = 8000$ Hz. The regularization parameter is set to be $\lambda = 10^{-3}$.

Figure 1 shows a comparison between the broadband beampattern of the time-domain implementation (21) (dark dashed line), to the theoretical beampattern [5, ch.2] (blue circles line). These patterns were also achieved by the frequency-domain implementation in [5, ch.3]. Comparing both patterns, one can obviously notice the equivalence between the time-domain and frequency-domain implementations.

The obtained filters were tested by simulating a white noise signal impinging towards the DMA and received by the sensors according to the model presented at (3). The received vector was fed into the temporal filters (17). Figure 2 shows the time-domain waveform (dark blue line) of the signals arrived from the endfire direction, null direction, and arbitrary direction of $88°$. It also shows the waveforms of the recovered signals in the output of the DMA (light red line). One can see that the derived filters provide perfect recovery of the desired endfire signal while suppressing the signal coming from the null direction. For signals impinging from an arbitrary direction which the DMA was not designed to suppress at all, the output signal is reasonably suppressed as compared to the input signal.

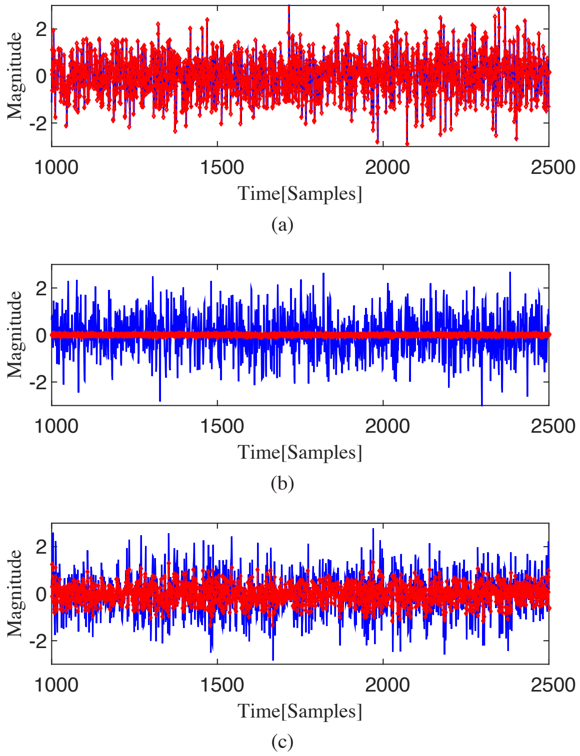

Fig. 2: Time-domain waveforms of the original signals (dark blue line) and output recovered signals (light red line) for the supercardioid: (a) source in the endfire direction, (b) source in the null direction, and (c) source in an arbitrary direction of $88°$.

We also plot in Fig. 3 the time-domain WNG and the time-domain DF as a function of the number of sensors, M, for the case of a first-order hypercardioid. One can see that the WNG increases with the number of sensors, while the DF is slightly above the value of 6 dB and does not vary at all. This result is expected since from theoretical point of view, we know that the directivity index is proportional to $(N+1)^2$, where N is the order. In addition, from this figure we can see that one of the effective ways to increase the robustness of the beamformer is to increase the number of sensors.

The results presented in this section show equivalence between time-domain and frequency-domain implementations of DMAs. Moreover, testing the time-domain filters with actual broadband signals confirms that the desired endfire signal is perfectly recovered while undesired signals, even if not arriving from null directions, are significantly suppressed.

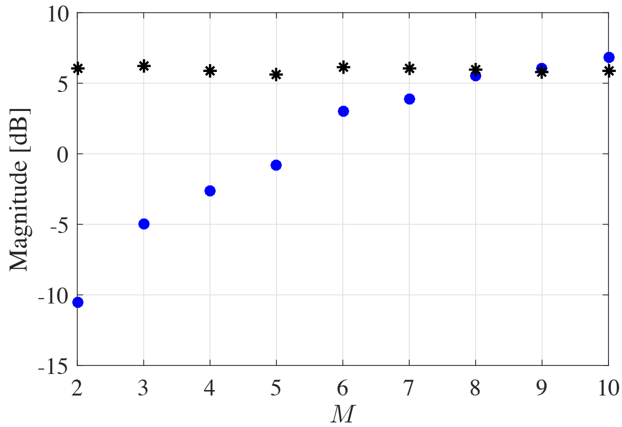

Fig. 3: WNG (circles) and DF (stars) vs. M, for the case of a first-order hypercardioid.

7. CONCLUSIONS

We have presented a framework for time-domain implementation of first-order DMAs, which is desirable in some applications such as real-time communications. Due to the DMA assumption, we get a very simple solution that provides a frequency-invariant beampattern. The quality measures widely used for assessment of beamformers were also defined in the time domain. Simulation results of the proposed implementation demonstrate that it is equivalent to the frequency-domain implementation, while providing more flexibility in the design considerations of practical systems employing DMAs.

8. REFERENCES

[1] H. H. Olson,"Gradient Microphones," *J. Acoust. Soc. Amer.*, vol. 17, pp. 192–198, Jan. 1946.

[2] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.

[3] G. W. Elko and J. Meyer, *Microphone arrays*, in Springer Handbook of Speech Processing, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Berlin, Germany: Springer-Verlag, 2008, Chapter 50, Part I, pp. 1021-1041.

[4] J. Benesty, M. Souden, and Y. Huang, "A perspective on differential microphone arrays in the context of noise

reduction," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 2, pp. 699-704, Feb. 2012.

[5] J. Benesty and J. Chen, *Study and Design of Differential Microphone Arrays*. Berlin, Germany: Springer-Verlag, 2012.

[6] G. W. Elko, "Superdirectional microphone arrays," in *Acoustic Signal Processing for Telecommunication*, S. L. Gay and J. Benesty, Eds. Boston, MA: Kluwer Academic Publishers, 2000, Chapter 10, pp. 181–237.

[7] G. W. Elko and A. T. N. Pong, "A simple first-order differential microphone,", in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust. (WASPAA)*, Oct. 1995, pp. 169-172.

[8] G. W. Elko and A. T. N. Pong, "A steerable and variable firstorder differential microphone array," in *Proc. ICASSP*, 1997, pp.223-226.

[9] R. M. M. Derkx and K. Janse, "Theoretical analysis of a first-order azimuth-steerable superdirective microphone array," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, pp. 150-162, Jan. 2009.

[10] J. Chen and J. Benesty, "A general approach to the design and implementation of linear differential microphone arrays," in *Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Oct, 2013 Asia-Pacific, pp. 1–7.

[11] L. Zhao, J. Benesty, and J. Chen,"Design of robust differential microphone arrays," *IEEE Trans. Audio, Speech, Language Process.*, vol. 22, no. 10, pp. 1455-1464, Oct. 2014.

[12] R. T. Compton, "The relationship between tapped delay-line and FFT processing in adaptive arrays," *IEEE Trans. Antennas and Propagation*, vol. 36, no. 1, pp. 15–26, Jan. 1988

[13] L. C. Godara, "Application of the fast Fourier transform to broadband beamforming," *J. Acoust. Soc. Amer.*, vol. 98, no. 1, pp. 230-240, July 1995.

[14] T. I. Laakso, V. Välimäki, M. Karjalainen and U. K. Laine, "Splitting the unit delay," *IEEE Signal Processing Mag.*, vol. 13, pp. 30–60, Jan. 1996.

[15] H. L. Van-trees, *Detection, Estimation and Modulation Theory, Part IV - Optimum Array Processing*. New York: Wiley Interscience, 2002.