An Information-Theoretic View of Array Processing

Jacek Dmochowski, Jacob Benesty, and Sofiène Affes

Abstract—The removal of noise and interference from an array of received signals is a most fundamental problem in signal processing research. To date, many well-known solutions based on second-order statistics (SOS) have been proposed. This paper views the signal enhancement problem as one of maximizing the mutual information between the source signal and array output. It is shown that if the signal and noise are Gaussian, the maximum mutual information estimation (MMIE) solution is not unique but consists of an infinite set of solutions which encompass the SOS-based optimal filters. The application of the MMIE principle to Laplacian signals is then examined by considering the important problem of estimating a speech signal from a set of noisy observations. It is revealed that while speech (well modeled by a Laplacian distribution) possesses higher order statistics (HOS), the well-known SOS-based optimal filters maximize the Laplacian mutual information as well; that is, the Laplacian mutual information differs from the Gaussian mutual information by a single term whose dependence on the beamforming weights is negligible. Simulation results verify these findings.

Index Terms—Array signal processing, beamforming, information entropy, mutual information.

I. INTRODUCTION

One of the most fundamental problems in signal processing is that of removing noise and interference from a received sensor signal. There are two general approaches. Single-sensor methods such as the celebrated Wiener filter [1] enhance the signal by emphasizing frequencies with a high signal-to-noise ratio (SNR) while attenuating those with a low SNR. On the other hand, multichannel techniques employ an array of sensors that perform spatial discrimination (or *beamforming*) to aid in removing the unwanted noise [2], [3]. This correspondence focuses on the latter of the two categories: array processing with particular emphasis on microphone-array processing for speech enhancement.

We propose a framework for array processing based on Shannon's concept of entropy and mutual information [4]. The goal of the proposed maximum mutual information estimation (MMIE) is to maximize the mutual information between the source signal and the beamformer output—that is, to filter the received signals in such a way that the resulting output conveys as much information about the source signal as possible.

Recently, a paper based on minimizing the mutual information between two beamformer outputs has been proposed for microphone arrays [5]. Indeed, it is well-known from independent components analysis (ICA) and blind source separation (BSS) that one way of defining the independence between component signals is through the idea of mutual information. In this case, we seek to minimize the mutual information between the multiple output signals so as to maximize their independence [6]. ICA and BSS are concerned with separating multiple signals and say nothing about the removal of noise from a single source

signal. In many applications, one does not need to resort to BSS techniques since one has prior information about the array geometry and source location. For example, estimates of relative delays [7] allow for nonblind beamforming techniques. To that end, the concept of information entropy has been applied to the time delay estimation (TDE) problem in [8]. In the following, we present a treatment of sensor array beamforming from an information theory point-of-view.

This correspondence is structured as follows: Section II presents the signal model used throughout the paper. The classical second-order-statistics (SOS)-based optimal beamformers are reviewed in Section III. Section IV reviews the basics of information theory and develops the MMIE solutions for Gaussian and Laplacian signals. Simulation results are presented in Section V; concluding remarks are made in Section VI.

II. ARRAY MODEL

Consider the conventional signal model in which the N-element sensor array captures a convolved desired signal in some noise field

$$y_n(k) = h_n * s(k) + v_n(k), \quad n = 1, 2, \ldots, N \quad (1)$$

where $y_n(k)$ is the received signal at sensor n and discrete time sample k, $s(k)$ is the desired source signal, h_n is the impulse response from the source to sensor n, $*$ denotes the linear convolution operation, and $v_n(k)$ is the additive noise at sensor n which is uncorrelated with the source signal. All signals are assumed to have zero-mean. In most cases, the impulse responses to the array are unknown. It is thus common to model each impulse response by a single attenuated and delayed direct-path component

$$y_n(k) = a_n s(k - \tau_n) + v_n(k) \quad (2)$$

where a_n and τ_n denote the attenuation and propagation time of the direct-path from the source to sensor n. Transposing to the frequency-domain leads to

$$\mathbf{y}(\omega) = s(\omega)\mathbf{d}(\omega) + \mathbf{v}(\omega) \quad (3)$$

where $s(\omega)$ is the discrete-time Fourier transform of $s(k)$, and

$$\mathbf{y}(\omega) = [\, y_1(\omega) \quad y_2(\omega) \quad \cdots \quad y_N(\omega) \,]^T$$
$$\mathbf{v}(\omega) = [\, v_1(\omega) \quad v_2(\omega) \quad \cdots \quad v_N(\omega) \,]^T$$
$$\mathbf{d}(\omega) = [\, a_1 e^{-j\omega\tau_1} \quad a_2 e^{-j\omega\tau_2} \quad \cdots \quad a_N e^{-j\omega\tau_N} \,]^T.$$

where ^{T} denotes matrix transposition.

The array processing, or beamforming, is then performed by applying a complex weight to each sensor and summing across the aperture

$$z(\omega) = \mathbf{w}^H(\omega)\mathbf{y}(\omega) = \mathbf{w}^H(\omega)[s(\omega)\mathbf{d}(\omega) + \mathbf{v}(\omega)] \quad (4)$$

where $z(\omega)$ is the beamformer output and ^{H} denotes conjugate transposition. From this point on, the dependence on frequency is dropped and the beamformer output is written as $z = \mathbf{w}^H \mathbf{y} = \mathbf{w}^H(s\mathbf{d} + \mathbf{v})$. It should be noted that if the impulse responses are in fact known *a priori* or estimated, then $\mathbf{h}(\omega)$ may replace $\mathbf{d}(\omega)$ in (3) and (4).

III. SECOND-ORDER STATISTICAL METHODS

A. Multichannel Wiener Filter

In loose terms, the fundamental goal of array beamforming is to produce an array output z which matches the desired signal s. An obvious

criterion to measure the "similarity" is the mean-squared error (MSE). The error signal is

$$e(\mathbf{w}) = s - \mathbf{w}^H \mathbf{y}. \qquad (5)$$

The MSE follows as

$$
\begin{aligned}
J_{\mathrm{MSE}}(\mathbf{w}) &= E\left\{|e(\mathbf{w})|^2\right\} \\
&= \sigma_s^2 - 2\Re\left\{\mathbf{w}^H \mathbf{r}_{sy}\right\} + \mathbf{w}^H \mathbf{R}_{yy} \mathbf{w}
\end{aligned} \qquad (6)
$$

where $E\{\cdot\}$ denotes mathematical expectation, $\Re(\cdot)$ denotes "real part of," $\sigma_s^2 = E\{|s|^2\}$ is the variance of the source signal, $\mathbf{r}_{sy} = E\{s^* \mathbf{y}\} = \sigma_s^2 \mathbf{d}$, and $\mathbf{R}_{yy} = E\{\mathbf{y}\mathbf{y}^H\}$. The weight vector which minimizes (6) is the well-known minimum-MSE (MMSE) solution [1]

$$\mathbf{w}_{\mathrm{MMSE}} = \sigma_s^2 \mathbf{R}_{yy}^{-1} \mathbf{d}. \qquad (7)$$

B. Minimum Variance Distortionless Response Filter

It is interesting to compare the MMSE solution to that of minimum variance, distortionless response (MVDR) beamformer, which solves the constrained optimization problem given by [9]

$$\mathbf{w}_{\mathrm{MVDR}} = \arg\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_{yy} \mathbf{w} \text{ subject to } \mathbf{w}^H \mathbf{d} = 1. \qquad (8)$$

The constraint $\mathbf{w}^H \mathbf{d} = 1$ (or $\mathbf{w}^H \mathbf{h} = 1$ in the general reverberant case [10]) performs dereverberation and ensures a unity-gain response to the desired signal, while the remaining degrees of freedom in $\mathbf{w}$ are utilized to minimize the contribution of noise and interference to the array output. The MVDR solution is well-known and given by

$$\mathbf{w}_{\mathrm{MVDR}} = \frac{\mathbf{R}_{yy}^{-1}\mathbf{d}}{\mathbf{d}^H \mathbf{R}_{yy}^{-1}\mathbf{d}} \qquad (9)$$

where $\mathbf{R}_{vv}$ may replace $\mathbf{R}_{yy}$ in (9). The MVDR solution is equivalent to the MMSE solution up to a scaling factor, and is generalized by the linearly constrained minimum variance method (LCMV) [11].

C. Maximum SNR Filter

The maximum SNR filter attempts to maximize the ratio of the energy of the signal to that of the noise [12]

$$\mathbf{w}_{\mathrm{maxSNR}} = \arg\max_{\mathbf{w}} \frac{\mathbf{w}^H \mathbf{R}_{ss} \mathbf{w}}{\mathbf{w}^H \mathbf{R}_{vv} \mathbf{w}} \qquad (10)$$

where $\mathbf{R}_{ss} = \sigma_s^2 \mathbf{d}\mathbf{d}^H$ and $\mathbf{R}_{vv} = E\{\mathbf{v}\mathbf{v}^H\}$, with $\mathbf{R}_{yy} = \mathbf{R}_{ss} + \mathbf{R}_{vv}$.

Differentiating (10) with respect to $\mathbf{w}^H$ and setting the gradient to zero results in the generalized eigenvalue problem

$$\mathbf{R}_{vv}^{-1} \mathbf{R}_{ss} \mathbf{w}_{\mathrm{maxSNR}} = \lambda_{\max} \mathbf{w}_{\mathrm{maxSNR}} \qquad (11)$$

where $\lambda_{\max}$ is the maximum eigenvalue of $\mathbf{R}_{vv}^{-1}\mathbf{R}_{ss}$ and also the value of the maximum attainable SNR. Since $\mathbf{R}_{ss}$ is rank-one, $\mathbf{R}_{vv}^{-1}\mathbf{R}_{ss}$ is also rank-one and has only one nonzero eigenvalue. The lone eigenvector of $\mathbf{R}_{ss}$ is $\mathbf{d}$; it is then easy to show that the principal eigenvector of $\mathbf{R}_{vv}^{-1}\mathbf{R}_{ss}$ is indeed $\mathbf{R}_{vv}^{-1}\mathbf{d}$. As a result

$$\mathbf{w}_{\mathrm{maxSNR}} = \frac{\mathbf{R}_{vv}^{-1}\mathbf{d}}{\left\|\mathbf{R}_{vv}^{-1}\mathbf{d}\right\|} \qquad (12)$$

where the eigenvector is normalized to unit-norm. Thus, the maximum SNR filter is also a scaled version of the MMSE solution.

IV. MAXIMUM MUTUAL INFORMATION ESTIMATION

All of the classical methods outlined above employ a SOS criterion as the similarity measure between the beamformer output and desired signal. However, the beamformer output and source signal are random variables, and a more complete measure would compare the probability density functions (pdfs) of the respective random variables. By doing so, higher order statistics (HOS) are implicitly taken into account.

A. Information-Theoretic Concepts

The quantification of the "difference" between two random variables is not well defined. Closely related to one of such measures is information entropy, which for a continuous random variable x is given by

$$H(x) = -\int_{-\infty}^{\infty} p(x) \ln p(x) dx \qquad (13)$$

where $p(x)$ is the pdf of x. The joint differential entropy between two random variables x_1 and x_2 follows intuitively as

$$H(x_1, x_2) = -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2) \ln p(x_1, x_2) dx_1 dx_2 \qquad (14)$$

where $p(x_1, x_2)$ is the joint pdf of x_1 and x_2.

One way of quantifying the difference between the random variables x_1 and x_2 is through the information theoretic construct of mutual information between two random variables x_1 and x_2

$$I(x_1; x_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2) \ln \frac{p(x_1, x_2)}{p(x_1)p(x_2)} dx_1 dx_2. \qquad (15)$$

The mutual information is equal to the Kullback–Leibler divergence between the joint pdf $p(x_1, x_2)$ and the product of the marginal pdfs $p(x_1)p(x_2)$, which leads to zero mutual information if the two random variables are independent. A well-known property of the mutual information which follows by substituting (13) and (14) into (15) is

$$I(x_1; x_2) = H(x_1) + H(x_2) - H(x_1, x_2). \qquad (16)$$

The MMIE weights the sensors such that the mutual information between the desired signal and the beamformer output is maximized. As the quantities in the frequency-domain model are complex, the concept of mutual information must have meaning in the complex domain. In the following, the MMIE is derived as the weight vector which maximizes the sum of the mutual informations between the real and imaginary parts

$$\mathbf{w}_{\mathrm{MMIE}} = \arg\max_{\mathbf{w}} \left\{ I\left[\Re(s); \Re(\mathbf{w}^H \mathbf{y})\right] + I\left[\Im(s); \Im(\mathbf{w}^H \mathbf{y})\right] \right\} \qquad (17)$$

where $\Im(\cdot)$ denotes "imaginary part of." In order to apply the MMIE, it is required to model the marginal distributions of the desired signal and the beamformer output, as well as their joint density.

B. Gaussian Signals

The pdf of a Gaussian-distributed zero-mean random variable x with variance σ_x^2 is given by

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-x^2/2\sigma_x^2}. \qquad (18)$$

The entropy of Gaussian random variable x then follows as

$$H(x) = \ln\sqrt{2\pi e \sigma_x^2}. \qquad (19)$$

The zero-mean random vector $\mathbf{x}$ is said to be jointly Gaussian distributed if the joint pdf of $\mathbf{x}$ is equal to

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{N/2} |\mathbf{\Sigma}_{xx}|^{1/2}} e^{-1/2 \mathbf{x}^T \mathbf{\Sigma}_{xx}^{-1} \mathbf{x}} \tag{20}$$

where $\mathbf{\Sigma}_{xx} = E\left\{\mathbf{x}\mathbf{x}^T\right\}$ is the covariance matrix, $|\cdot|$ denotes the determinant of a matrix, and $E\left\{\mathbf{x}\right\} = \mathbf{0}$. The joint entropy of a jointly Gaussian distributed random vector $\mathbf{x}$ is then

$$H(\mathbf{x}) = \ln \sqrt{(2\pi e)^N |\mathbf{\Sigma}_{xx}|}. \tag{21}$$

We begin with the case of a zero-mean Gaussian source signal corrupted by zero-mean Gaussian noise

$$s_{\mathrm{R}} \triangleq \Re(s) \sim \mathcal{N}(0, \varsigma_s), \quad s_{\mathrm{I}} \triangleq \Im(s) \sim \mathcal{N}(0, \varsigma_s),$$
$$\mathbf{v}_{\mathrm{R}} \triangleq \Re(\mathbf{v}) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{vv}), \quad \mathbf{v}_{\mathrm{I}} \triangleq \Im(\mathbf{v}) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{vv})$$

where $\varsigma_s^2 = \sigma_s^2/2$ is the variance of the real and imaginary parts of the source signal, and $\mathbf{\Sigma}_{vv}$ is the covariance matrix of the real and imaginary components of the noise field, where it has been assumed that the real and imaginary parts of the source and noise are independent and identically distributed.

From the array model (4), it follows that

$$z_{\mathrm{R}} \triangleq \Re(z) = b_{\mathrm{R}} s_{\mathrm{R}} + b_{\mathrm{I}} s_{\mathrm{I}} + \mathbf{w}_{\mathrm{R}}^T \mathbf{v}_{\mathrm{R}} + \mathbf{w}_{\mathrm{I}}^T \mathbf{v}_{\mathrm{I}} \tag{22}$$

$$z_{\mathrm{I}} \triangleq \Im(z) = b_{\mathrm{R}} s_{\mathrm{I}} - b_{\mathrm{I}} s_{\mathrm{R}} + \mathbf{w}_{\mathrm{R}}^T \mathbf{v}_{\mathrm{I}} - \mathbf{w}_{\mathrm{I}}^T \mathbf{v}_{\mathrm{R}} \tag{23}$$

where $b = b_{\mathrm{R}} + j b_{\mathrm{I}} = \mathbf{d}^H \mathbf{w}$, $\mathbf{w}_{\mathrm{R}} \triangleq \Re(\mathbf{w})$, $\mathbf{w}_{\mathrm{I}} \triangleq \Im(\mathbf{w})$, and $E\left\{z_{\mathrm{R}}\right\} = E\left\{z_{\mathrm{I}}\right\} = 0$. The variance of the real and imaginary components of z is given by

$$\mathrm{var}\,(z_{\mathrm{R}}) = \mathrm{var}\,(z_{\mathrm{I}}) = \frac{1}{2}\mathbf{w}^H \mathbf{R}_{yy} \mathbf{w} \tag{24}$$

where var denotes "variance of." The covariance between s_{R} and z_{R} is easily found as

$$E\left\{s_{\mathrm{R}} z_{\mathrm{R}}\right\} = E\left\{s_{\mathrm{I}} z_{\mathrm{I}}\right\} = \frac{1}{2}\Re(\mathbf{w}^H \mathbf{d})\sigma_s^2. \tag{25}$$

The mutual information between the real (or imaginary) parts of the source and array output may be written as

$$I(s_{\mathrm{R}}; z_{\mathrm{R}}) = H(s_{\mathrm{R}}) + H(z_{\mathrm{R}}) - H(s_{\mathrm{R}}, z_{\mathrm{R}}) = I(s_{\mathrm{I}}; z_{\mathrm{I}}). \tag{26}$$

From direct substitution, the mutual information between the desired signal and array output is given by (27), as shown at the bottom of the page. Taking the gradient of (27) with respect to $\mathbf{w}^H$ results in

$$\nabla_{\mathbf{w}^H}\left[I(s_{\mathrm{R}}; z_{\mathrm{R}}) + I(s_{\mathrm{I}}; z_{\mathrm{I}})\right]$$
$$= 2\left[\frac{\mathbf{R}_{yy}\mathbf{w}}{\mathbf{w}^H \mathbf{R}_{yy}\mathbf{w}} - \frac{\mathbf{R}_{yy}\mathbf{w} - \sigma_s^2 \Re(\mathbf{w}^H \mathbf{d})\mathbf{d}}{\mathbf{w}^H \mathbf{R}_{yy}\mathbf{w} - \sigma_s^2 \Re^2(\mathbf{w}^H \mathbf{d})}\right]. \tag{28}$$

Setting (28) to zero and simplifying leads to

$$\frac{\Re\left(\mathbf{w}^H \mathbf{d}\right)}{\mathbf{w}^H \mathbf{R}_{yy}\mathbf{w}}\mathbf{R}_{yy}\mathbf{w} = \mathbf{d}. \tag{29}$$
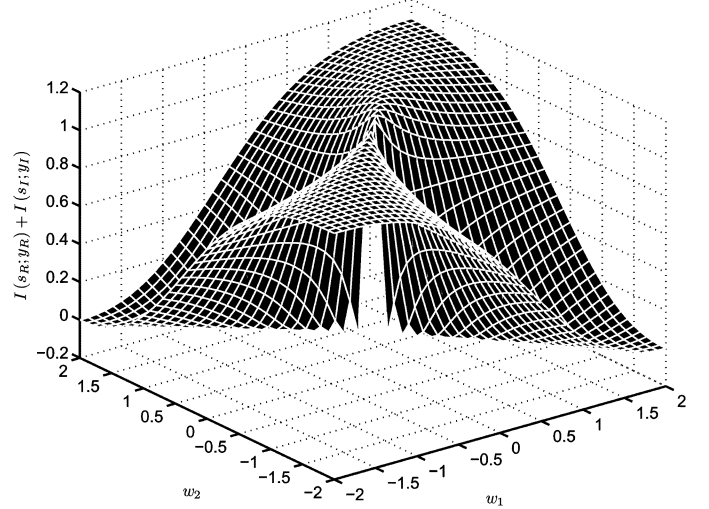

Fig. 1. Mutual information surface for Gaussian signals.

It easily follows from (29) that $\Re\left(\mathbf{w}^H \mathbf{d}\right) = \mathbf{w}^H \mathbf{d}$, and thus the MMIE solution is given by the set

$$\left\{\mathbf{w}_{\mathrm{MMIE,Gaussian}}\right\} \triangleq \left\{\mathbf{w} \in \mathbb{C}^N \,|\, \mathbf{g}(\mathbf{w}) = \mathbf{d}\right\} \tag{30}$$

where $\mathbf{g}(\mathbf{w}) = (\mathbf{w}^H \mathbf{d})/(\mathbf{w}^H \mathbf{R}_{yy}\mathbf{w})\mathbf{R}_{yy}\mathbf{w}$. It can easily be checked that one solution to (29) is indeed the Wiener filter

$$\mathbf{w}_{\mathrm{MMSE}} \in \left\{\mathbf{w}_{\mathrm{MMIE,Gaussian}}\right\}. \tag{31}$$

Moreover, it is evident that if $\mathbf{g}(\mathbf{w}) = \mathbf{d}$, then $\mathbf{g}(c\mathbf{w}) = \mathbf{d}$, where c is a complex constant. Thus, the MVDR and maximum SNR solutions are also MMIE solutions. We see that the mutual information criterion seems to unify the various SOS-based optimal beamformers into a common framework.

To illustrate, Fig. 1 shows the mutual information surface for a two-element array with an inter-element spacing of half the operating wavelength. The source is located at array broadside (far-field anechoic propagation), meaning that $\mathbf{d} = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$ without loss of generality. The sensors are corrupted with spatially diffuse Gaussian noise. The SNR is 0 dB. To enable visualization, the plot is shown for all weight vectors of the form $\mathbf{w} = \mathbf{w}_{\mathrm{R}}$ (the imaginary parts are all zero). Since the signals are already aligned, this is a reasonable space of weights, which includes the MMSE $\left(\mathbf{w}_{\mathrm{MMSE}} = \begin{bmatrix} 1/3 & 1/3 \end{bmatrix}^T\right)$, MVDR $\left(\mathbf{w}_{\mathrm{MVDR}} = \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}^T\right)$, and maximum SNR $\left(\mathbf{w}_{\mathrm{maxSNR}} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}^T\right)$ solutions. It is clear from the plot that the maxima of the mutual information surface lie on the line $w_1 = w_2$, which contains the SOS-based optimal beamformers.

From the surface of Fig. 1, it is evident that in the case of a Gaussian source signal corrupted by additive Gaussian noise, the well-known MVDR filter (and its equivalents) maximizes the mutual information between the source signal and array output. This is intuitively satisfying, as it is well known that the Gaussian distribution is specified completely by its SOS.

$$I(s_{\mathrm{R}}; z_{\mathrm{R}}) + I(s_{\mathrm{I}}; z_{\mathrm{I}}) = 2\left[\ln \sqrt{\pi e \sigma_s^2} + \ln \sqrt{\pi e \mathbf{w}^H \mathbf{R}_{yy}\mathbf{w}} - \ln \sqrt{(\pi e)^2 \left[\sigma_s^2 \mathbf{w}^H \mathbf{R}_{yy}\mathbf{w} - \sigma_s^4 \Re^2(\mathbf{w}^H \mathbf{d})\right]}\right] \tag{27}$$

C. Laplacian Signals

A Laplacian random variable x with zero-mean and variance σ_x^2 has a pdf given by

$$p(x) = \frac{\sqrt{2}}{2\sigma_x} e^{-\sqrt{2}|x|/\sigma_x}. \tag{32}$$

The resulting entropy is given by

$$H(x) = 1 + \ln\left(\sqrt{2}\sigma_x\right). \tag{33}$$

The elements of the zero-mean random vector $\mathbf{x}$ are said to be jointly Laplacian if their joint pdf is given by [13]

$$p(\mathbf{x}) = \frac{2\left[\frac{1}{2}\theta\right]^{P/2} K_P\left[\sqrt{2\theta}\right]}{\sqrt{(2\pi)^N |\boldsymbol{\Sigma}_{xx}|}} \tag{34}$$

where $\theta = \mathbf{x}^T \boldsymbol{\Sigma}_{xx}^{-1} \mathbf{x}$, $P = 1 - N/2$, and $K_P(\cdot)$ is the modified Bessel function of the second kind. It easily follows that the expression for joint Laplacian entropy is given by [8]

$$\begin{aligned}
H(\mathbf{x}) = &\frac{1}{2}\ln\left\{\frac{(2\pi)^N}{4}|\boldsymbol{\Sigma}_{xx}|\right\} \\
&- \frac{P}{2}E\left\{\ln\left(\frac{1}{2}\theta\right)\right\} - E\left\{\ln K_P\left(\sqrt{2\theta}\right)\right\}
\end{aligned} \tag{35}$$

where the expectation terms are evaluated with respect to $p(\mathbf{x})$ but lack closed forms.

Following the development of the previous section, we assume that the source and array output follow a joint Laplacian density. As a result, the source and array output are also marginally Laplacian. Speech has been shown to follow a Laplacian distribution [14]. However, since noise is typically Gaussian, the array output is a mixture of Laplacian and Gaussian random variables. Assuming a Laplacian distribution for the array output is questionable, particularly for low SNRs. However, it has been shown that assuming a Laplacian distribution for Gaussian noise-corrupted signals is valid [8]. As a result, we proceed with the Laplacian assumption.

We are now ready to define the mutual information between the source and beamformer output for jointly Laplacian signals. By substituting the SOS of the source signal and array output into the expressions (33) and (35), the mutual information between the source and beamformer output is given by

$$\begin{aligned}
&I(s_{\mathrm{R}}; z_{\mathrm{R}}) + I(s_{\mathrm{I}}; z_{\mathrm{I}}) \\
&= 4 + \ln \sigma_s^2 + \ln \mathbf{w}^H \mathbf{R}_{yy}\mathbf{w} - \ln\left(\pi^2 |\boldsymbol{\Sigma}_{sz}|\right) \\
&\quad + \frac{2}{\pi\sqrt{|\boldsymbol{\Sigma}_{sz}|}} \int_{-\infty}^{\infty} K_0\left(\sqrt{2\mathbf{x}^T \boldsymbol{\Sigma}_{sz}^{-1}\mathbf{x}}\right) \\
&\quad \times \ln\left[K_0\left(\sqrt{2\mathbf{x}^T \boldsymbol{\Sigma}_{sz}^{-1}\mathbf{x}}\right)\right] d\mathbf{x}
\end{aligned} \tag{36}$$

where

$$\boldsymbol{\Sigma}_{sz} = \begin{bmatrix} \frac{1}{2}\sigma_s^2 & \frac{1}{2}\Re\left(\mathbf{w}^H\mathbf{d}\right)\sigma_s^2 \\ \frac{1}{2}\Re\left(\mathbf{w}^H\mathbf{d}\right)\sigma_s^2 & \frac{1}{2}\mathbf{w}^H\mathbf{R}_{yy}\mathbf{w} \end{bmatrix}. \tag{37}$$

Notice that the gradient of the first four terms in (36) is equivalent up to scale to the gradient of the Gaussian mutual information (28)

$$\begin{aligned}
&\nabla_{\mathbf{w}^H}\left[4 + \ln \sigma_s^2 + \ln \mathbf{w}^H \mathbf{R}_{yy}\mathbf{w} - \ln\left(\pi^2 |\boldsymbol{\Sigma}_{sz}|\right)\right] \\
&= \frac{\mathbf{R}_{yy}\mathbf{w}}{\mathbf{w}^H \mathbf{R}_{yy}\mathbf{w}} - \frac{\mathbf{R}_{yy}\mathbf{w} - \sigma_s^2 \Re(\mathbf{w}^H\mathbf{d})\mathbf{d}}{\mathbf{w}^H \mathbf{R}_{yy}\mathbf{w} - \sigma_s^2 \Re^2(\mathbf{w}^H\mathbf{d})}.
\end{aligned} \tag{38}$$

Thus, the last term in (36) is the term which distinguishes the Laplacian MMIE from the Gaussian MMIE.

The gradient of this last term in (36) is difficult to derive in closed-form. In fact, symbolic mathematical packages are unable to yield a simplification. In this correspondence, we study the dependence of this last term on the beamforming weight vector $\mathbf{w}$ and investigate the possibility of selecting filter coefficients taking into account this term.

As an initial example, Fig. 2 illustrates the mutual information surface for Laplacian signals using the parameters described in the previous section, which specify an anechoic environment with additive diffuse noise. Fig. 2 (a) and (b) displays the surfaces pertaining to the first four and last terms, respectively, of the Laplacian mutual information (36). Fig. 2(c) displays the total mutual information surface. It is evident that there is a lack of dependence of the last term on the weight vector; note also the similarity of the Laplacian mutual information surface to that of the Gaussian case.

It appears that in the fundamental case of anechoic propagation and additive diffuse noise, the Laplacian MMIE is equivalent to the Gaussian MMIE. Note, however, that when considering weight vectors with nonzero imaginary components, the last term in (36) may vary with $\mathbf{w}$; however, it is not possible to visualize such behavior. This will be investigated in the simulation section.

D. Dependence of Real and Imaginary Parts

It has been suggested in [15] that the real and imaginary parts of a speech signal's discrete Fourier coefficient are uncorrelated but not independent. In the development of Section IV-B, it was stated that we assume independence for s_{R} and s_{I}; for Gaussian signals, this is equivalent to assuming that the real and imaginary parts are uncorrelated.

In the case of Laplacian signals, it is easily seen that the results of Section IV-C hold as long as s_{R} and s_{I} are uncorrelated. Even though the joint pdf

$$p(s_{\mathrm{R}}, s_{\mathrm{I}}) \neq p(s_{\mathrm{R}})p(s_{\mathrm{I}}) \tag{39}$$

need not factor into the product of the marginal pdfs, the uncorrelated assumption (in addition to zero-mean for s_{R} and z_{R}) leads to

$$E\{s_{\mathrm{R}}z_{\mathrm{R}}\} = \frac{1}{2}\Re(\mathbf{w}^H\mathbf{d})\sigma_s^2 + \Im(\mathbf{w}^H\mathbf{d})E\{s_{\mathrm{R}}s_{\mathrm{I}}\} \tag{40}$$

$$= \frac{1}{2}\Re(\mathbf{w}^H\mathbf{d})\sigma_s^2 \tag{41}$$

meaning that the joint Laplacian entropy takes the form

$$\begin{aligned}
H(s_{\mathrm{R}}, z_{\mathrm{R}}) = &\ln\left(\pi |\boldsymbol{\Sigma}_{sz}|^{1/2}\right) - \frac{1}{\pi\sqrt{|\boldsymbol{\Sigma}_{sz}|}} \\
&\times \int_{-\infty}^{\infty} K_0\left(\sqrt{2\mathbf{x}^T\boldsymbol{\Sigma}_{sz}^{-1}\mathbf{x}}\right) \\
&\times \ln\left[K_0\left(\sqrt{2\mathbf{x}^T\boldsymbol{\Sigma}_{sz}^{-1}\mathbf{x}}\right)\right] d\mathbf{x}
\end{aligned} \tag{42}$$

where $\boldsymbol{\Sigma}_{sz}$ is specified by (37). Notice that the joint Laplacian entropy is specified entirely by the second-order covariance matrix $\boldsymbol{\Sigma}_{sz}$.
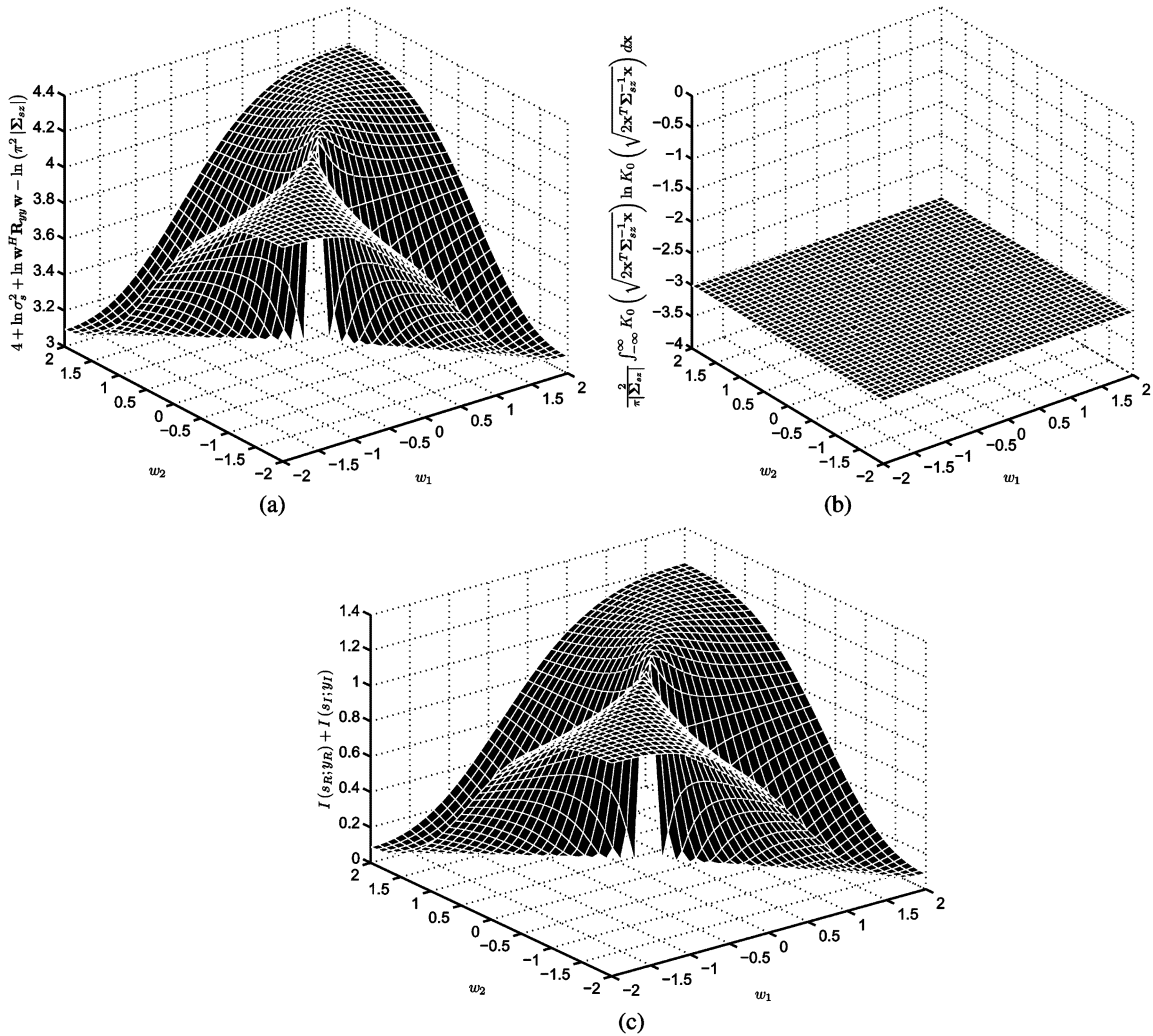
Fig. 2. Laplacian mutual information surfaces. (a) First four terms. (b) Last term. (c) Total mutual information.

V. SIMULATION STUDY

A. Simulation Model

A simulation study using the image method model [16] was conducted. The simulations employ a three-element uniform linear array (ULA) with an inter-element spacing of 17 cm. The first-element is located at (152.4,19.05,101.6) cm; the second at: (169.4,19.05,101.6) cm; the third at: (186.4,19.05,101.6) cm. In the simulations, we utilize the ULA formed by elements 1 and 2 as a two-element ULA optimized (half-wavelength spacing) for 1000 Hz. The ULA formed by elements 1 and 3 is employed as a ULA optimized for 500 Hz. The source is located at (50.8,304.8,101.6) cm and consists of female English speech. The room has dimensions 304.8-by-457.2-by-381 cm.

In the first simulation scenario, spatially white noise is added at the microphones with a SNR of 20 dB. In the second scenario, an interference source located at (254,304.8,101.6) cm plays temporally white noise with a signal-to-interference ratio (SIR) of 0 dB (the additive noise is still present). For both scenarios, an anechoic and reverberant simulation is performed; for the reverberant simulations, the reflection coefficients of the floors, ceiling, and walls are adjusted to achieve a 60 dB reverberation decay time of 300 ms. The layout of the experimental room is shown in Fig. 3.

In the reverberant simulations, it is assumed that the impulse responses to the array are known *a priori*. Sample estimates of the correlation matrices are computed every 8-ms frame and then averaged to
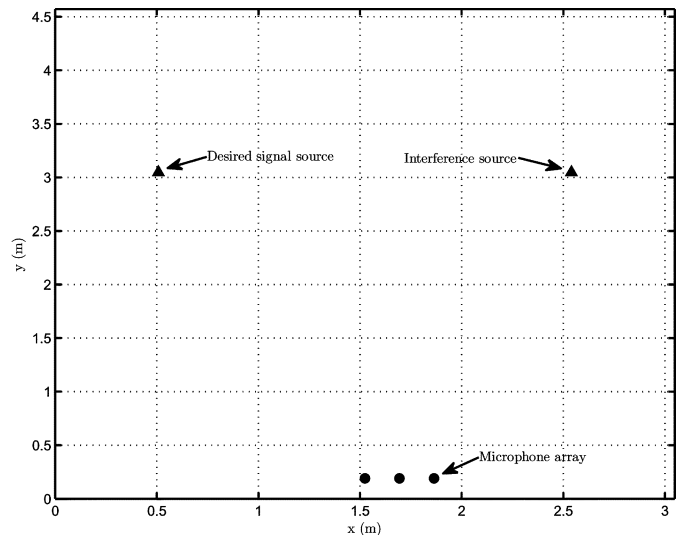

Fig. 3. Layout of experimental room.

arrive at the final SOS which are then used for computing the optimal weight vectors: MVDR, MMSE, and maximum SNR.

The MMIE solutions for both Gaussian and Laplacian distributions lack a closed-form solution; thus, to determine the MMIE solution, a
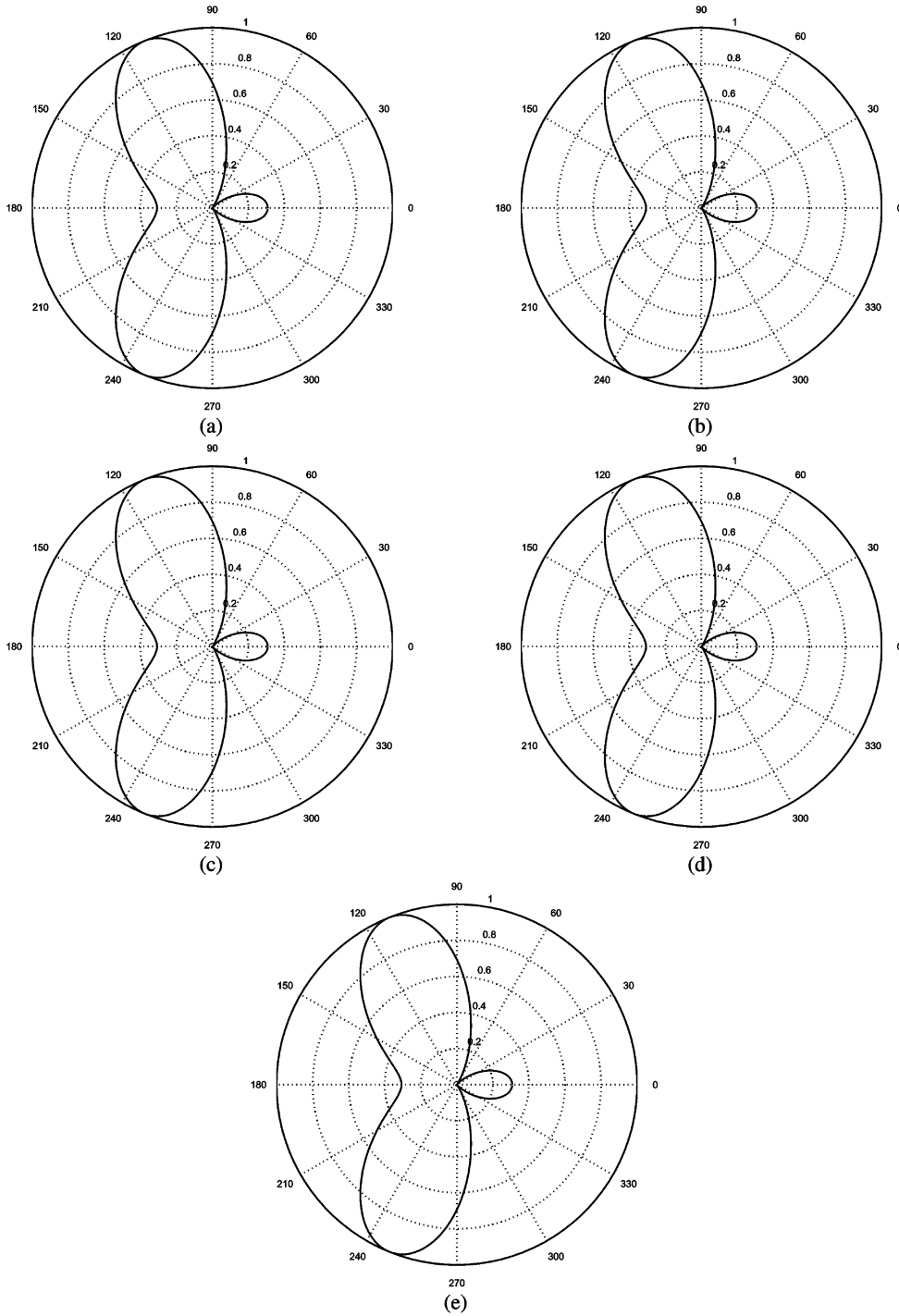
Fig. 4.   Beampatterns obtained in the anechoic case for scenario 1 (500 Hz). (a) MVDR. (b) MMSE. (c) maxSNR. (d) MMIE. (Gaussian). (e) MMIE (Laplacian).

comprehensive search of the space $\{\mathbf{w} \in \mathbb{C}^2 \mid \mathbf{w}^H\mathbf{w} = 1\}$ sampled at 15 625 locations was performed. With a two-microphone ULA, this corresponds to sampling the surface of a unit hypersphere [17] or three-sphere of the form

$$w_{1,R}^2 + w_{1,I}^2 + w_{2,R}^2 + w_{2,I}^2 = 1. \tag{43}$$

Notice that the coordinates of a unit three-sphere may be expressed in hyperspherical coordinates

$$w_{1,R} = \cos\psi$$

$$w_{1,I} = \cos\phi\sin\theta\sin\psi$$
$$w_{2,R} = \sin\phi\sin\theta\sin\psi$$
$$w_{2,I} = \cos\theta\sin\psi$$

where ψ, ϕ, and θ are the four-dimensional analogs of the spherical azimuth and elevation, where ψ and θ range over $(0,\pi)$ while ϕ ranges over $(0,2\pi)$. In order to create the search space, we sample each of ψ, ϕ, and θ uniformly with 25 samples; by forming the outer product of the three sets of samples, we arrive at the final 15 625-sample search space. Moreover, the MVDR, MMSE, and maximum SNR weights are added to the search space for completeness. In order to compute the
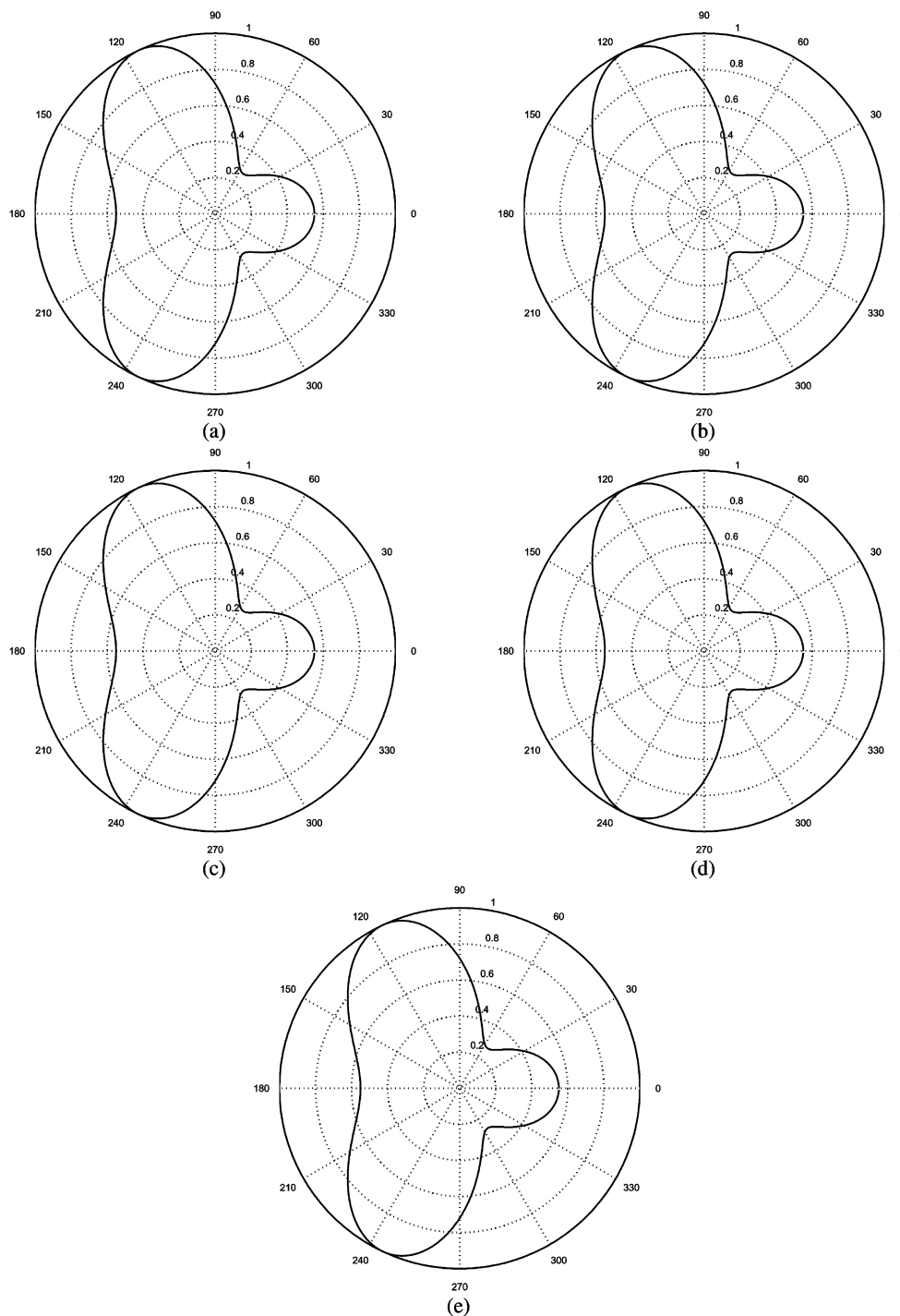
Fig. 5.   Beampatterns obtained in the reverberant case for scenario 1 (500 Hz). (a) MVDR. (b) MMSE. (c) maxSNR. (d) MMIE. (Gaussian). (e) MMIE (Laplacian).

Laplacian mutual information, we numerically compute the integral in (36) using MATLAB's dblquad function.

B. Discussion

The mutual information obtained using the SOS-based and MMIE solutions in scenarios 1 and 2 is shown in Tables I and III, respectively. Table II displays the input and output SNRs obtained by the various optimal beamformers in the first simulation scenario; similarly, Table IV displays the input and output signal-to-noise-and-interference ratios (SINRs) obtained in the second scenario. Figs. 4–7 display the beam-patterns corresponding to the SOS-based optimal filters and MMIE solutions for the various simulation parameters.

Consider the findings presented in Tables I and II: the SOS-based beamformers attain the same level of mutual information as that of the MMIE solution for both the Gaussian and Laplacian cases. Furthermore, the SNRs produced by both SOS-based filters and the MMIE solutions are identical. From the beampatterns of Figs. 4 and 5, the directivities produced by the classical beamformers are very similar to those corresponding to the MMIE filters—a beam is formed at approximately $110°$ (measured counter-clockwise from the array axis), which is the bearing of the desired signal source.
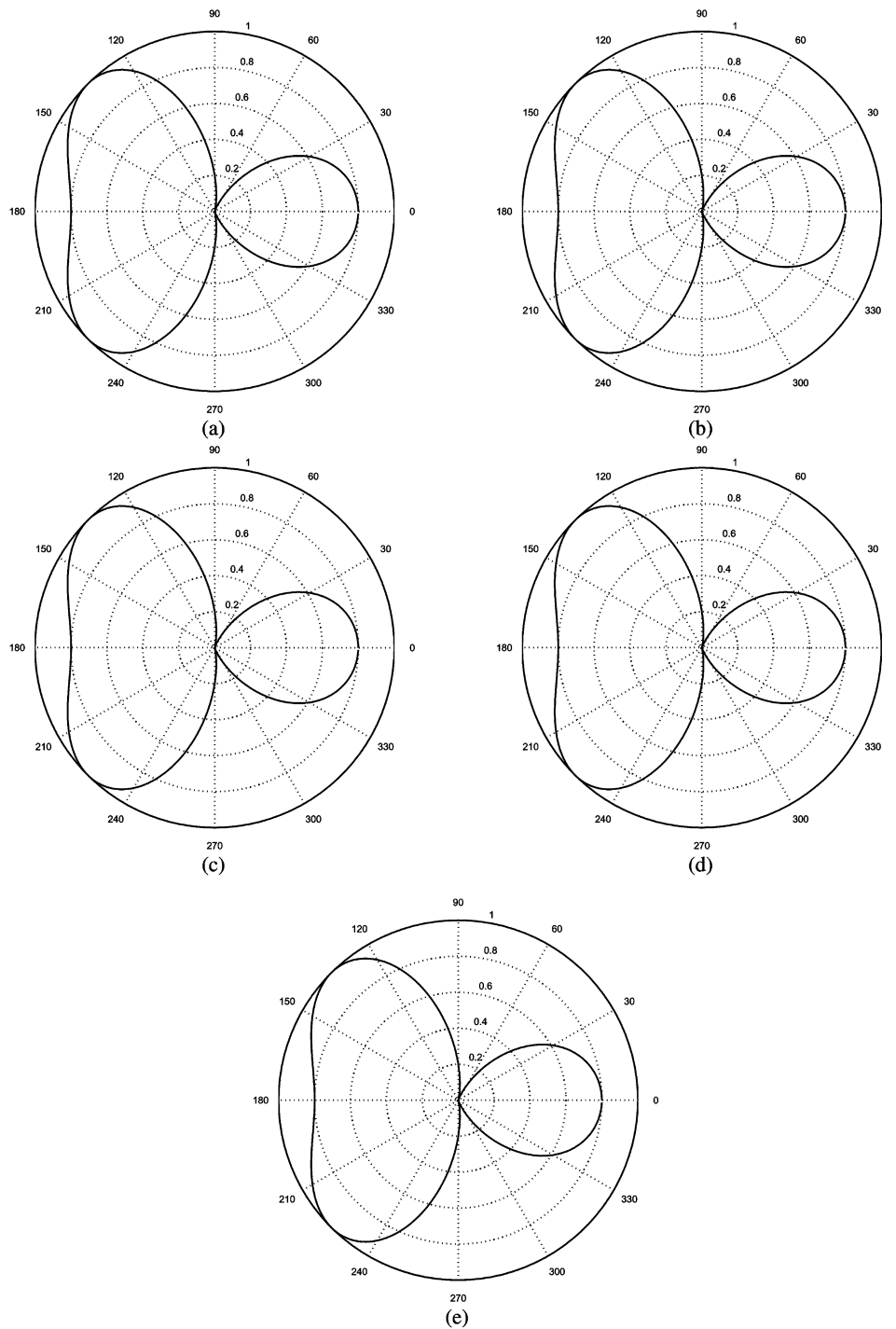
Fig. 6. Beampatterns obtained in the anechoic case for scenario 2 (500 Hz). (a) MVDR. (b) MMSE. (c) maxSNR. (d) MMIE. (Gaussian). (e) MMIE (Laplacian).

From Tables III and IV, it is seen that for the case of a point interference source, the MMIE solution is again equivalent to the classical SOS-based beamformers in terms of both the mutual information attained and the output SINR. From the beampatterns of Fig. 6, we see that both the conventional adaptive beamformers and MMIE solutions steer a null in the vicinity of the interference (i.e., approximately $70°$). The beampatterns are virtually identical.

The findings of the experimental evaluation confirm those predicted by Section IV: the MVDR, MMSE, and maximum SNR filters comprise the MMIE solution set in the case of Gaussian signals. Moreover, the results support the notion that the HOS of a speech signal

may not be exploited by a linear filter-and-sum beamformer to generate "cleaner" estimates (i.e., containing more mutual information) of the source signal, assuming that the source signal and array output are well-modeled by a joint Laplacian distribution. As witnessed in Fig. 2, the fifth term of the Laplacian mutual information (36) does not appear to be dependent on the beamforming weights. The MVDR filter (and its equivalents) attains the maximum possible level of mutual information between the source and beamformer output. Several conclusions may be drawn from this finding: perhaps we have reached the limit of linear filtering for noise and interference reduction. More sophisticated and possibly nonlinear beamforming structures may do a better
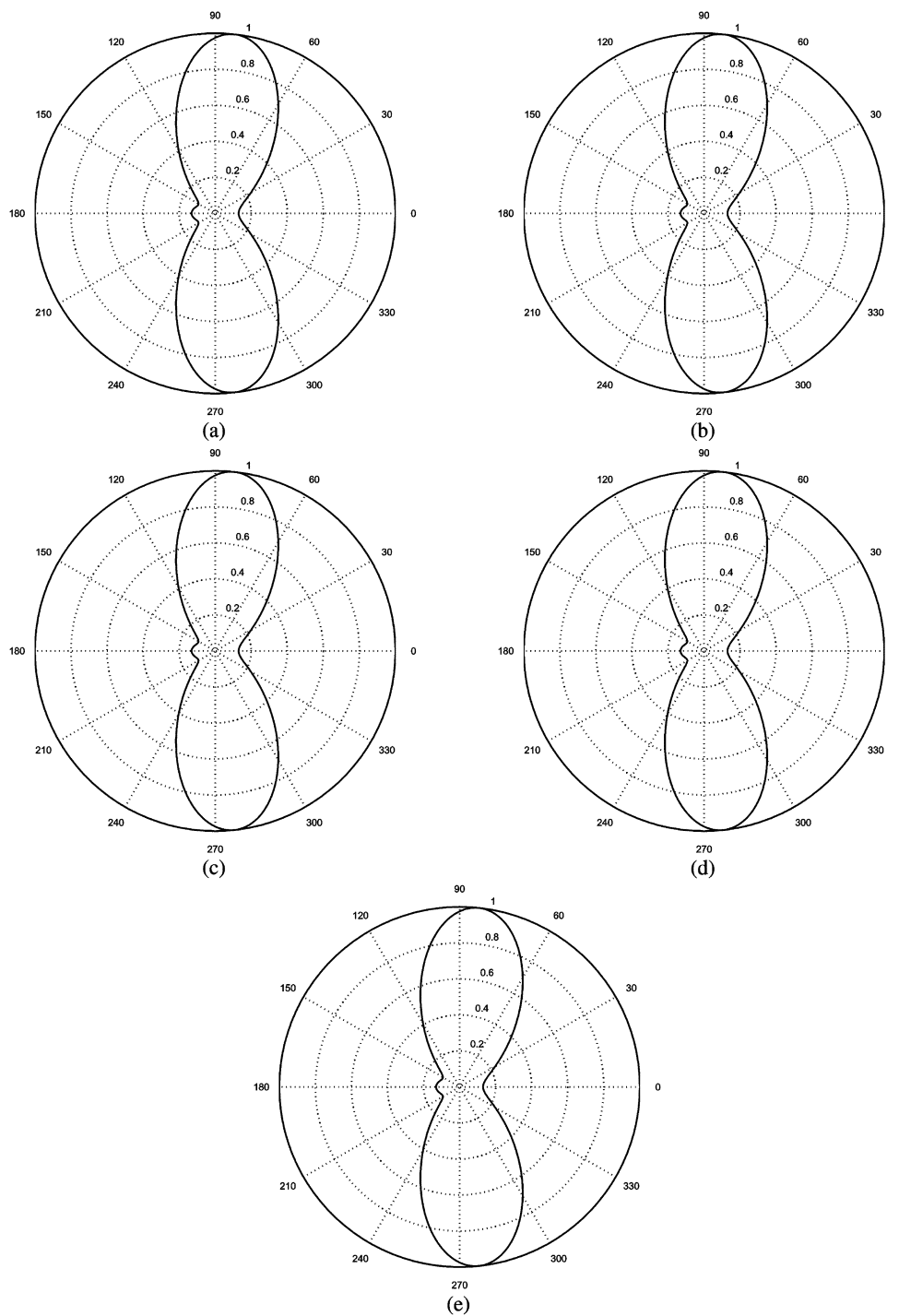
Fig. 7.   Beampatterns obtained in the reverberant case for scenario 2 (500 Hz). (a) MVDR. (b) MMSE. (c) maxSNR. (d) MMIE. (Gaussian). (e) MMIE (Laplacian).

TABLE I
SIMULATION SCENARIO 1: OPTIMAL BEAMFORMERS AND THE RESULTING MUTUAL INFORMATION OBTAINED

T_{60} (ms)	f (Hz)	Gaussian				Laplacian			
		MVDR	MMSE	maxSNR	MMIE	MVDR	MMSE	maxSNR	MMIE
0	500	8.37	8.37	8.37	8.37	11.47	11.47	11.47	11.47
	1000	8.11	8.11	8.11	8.11	11.21	11.21	11.21	11.21
300	500	8.58	8.58	8.58	8.58	11.67	11.67	11.67	11.67
	1000	6.77	6.77	6.77	6.77	9.86	9.86	9.86	9.86

TABLE II
SIMULATION SCENARIO 1: OPTIMAL BEAMFORMERS AND THE RESULTING SNRs OBTAINED

T_{60} (ms)	f (Hz)	input (dB)	MVDR (dB)	MMSE (dB)	maxSNR (dB)	MMIE-G (dB)	MMIE-L (dB)
0	500	33.30	36.37	36.37	36.37	36.37	36.37
	1000	32.20	35.24	35.24	35.24	35.24	35.24
300	500	34.18	37.25	37.25	37.25	37.25	37.25
	1000	26.39	29.38	29.38	29.38	29.38	29.38

TABLE III
SIMULATION SCENARIO 2: OPTIMAL BEAMFORMERS AND THE RESULTING MUTUAL INFORMATION OBTAINED

T_{60} (ms)	f (Hz)	Gaussian				Laplacian			
		MVDR	MMSE	maxSNR	MMIE	MVDR	MMSE	maxSNR	MMIE
0	500	8.06	8.06	8.06	8.06	11.15	11.15	11.15	11.15
	1000	7.84	7.84	7.84	7.84	10.93	10.93	10.93	10.93
300	500	7.09	7.09	7.09	7.09	10.18	10.18	10.18	10.18
	1000	6.28	6.28	6.28	6.28	9.37	9.37	9.37	9.37

TABLE IV
SIMULATION SCENARIO 2: OPTIMAL BEAMFORMERS AND THE RESULTING SINRs OBTAINED

T_{60} (ms)	f (Hz)	input (dB)	MVDR (dB)	MMSE (dB)	maxSNR (dB)	MMIE-G (dB)	MMIE-L (dB)
0	500	16.73	35.00	35.00	35.00	35.00	35.00
	1000	12.13	34.03	34.03	34.03	34.03	34.03
300	500	10.67	30.80	30.80	30.80	30.80	30.80
	1000	−1.81	27.25	27.25	27.25	27.25	27.25

job at matching the array output to the desired source signal. The results of this correspondence call into question the importance of HOS in array processing. From a spatial filtering standpoint, the MMIE filters operate in the same manner as the well-known SOS-based optimal filters: a beam is steered to the source, while a null is steered to the interference. If we are to truly combat the major problems in microphone arrays, it seems that HOS will not have much to do with the solution.

VI. CONCLUSION

This correspondence has presented an information-theoretic view of beamforming in which the criterion is the mutual information between source and array output. It was shown that the optimal SOS-based beamformers maximize this criterion under the assumption of Gaussian signals. For jointly Laplacian signals, it turns out that classical adaptive beamformers such as the MVDR filter also maximize the mutual information between the source and beamformer output, and thus appear to be optimal linear estimators of speech.

REFERENCES

[1] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. New York: Wiley, 1949.
[2] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing*. Englewood Cliffs, New Jersey: Prentice-Hall, 1993.
[3] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.
[4] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.
[5] K. Kumatani, T. Gehrig, U. Mayer, E. Stoimenov, J. McDonough, and M. Wölfel, "Adaptive beamforming with a minimum mutual information criterion," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, pp. 2527–2541, Nov. 2007.
[6] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, pp. 411–430, 2000.
[7] J. Chen, J. Benesty, and Y. Huang, "Time delay estimation in room acoustic environments: An overview," *EURASIP J. Appl. Signal Process.*, vol. 2006, 2006, article ID 26503.
[8] J. Benesty, J. Chen, and Y. Huang, "Time-delay estimation via minimum entropy," *IEEE Signal Process. Lett.*, vol. 14, pp. 157–160, Mar. 2007.
[9] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.
[10] S. Affes and Y. Grenier, "A signal subspace tracking algorithm for microphone array processing of speech," *IEEE Trans. Speech Audio Process.*, vol. 5, pp. 425–437, Sep. 1997.
[11] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.
[12] R. Monzingo and T. Miller, *Introduction to Adaptive Arrays*. New York: Wiley, 1980.
[13] T. Eltoft, T. Kim, and T. W. Lee, "On the multivariate Laplace distribution," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 300–303, May 2006.
[14] S. Gazor and W. Zhang, "Speech probability distribution," *IEEE Signal Process. Lett.*, vol. 10, no. 7, pp. 204–207, Jul. 2003.
[15] J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, "Minimum mean-square error estimation of discrete Fourier coefficients with generalized gamma priors," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 5, pp. 1741–1752, Aug. 2007.
[16] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, Apr. 1979.
[17] M. A. Peterson, "Dante and the 3-sphere," *Amer. J. Phys.*, vol. 47, pp. 1031–1035, 1979.