

A Speech Distortion and Interference Rejection Constraint Beamformer

Emanuël A. P. Habets, *Senior Member, IEEE*, Jacob Benesty, and Patrick A. Naylor, *Senior Member, IEEE*

Abstract—Signals captured by a set of microphones in a speech communication system are mixtures of desired and undesired signals and ambient noise. Existing beamformers can be divided into those that preserve or distort the desired signal. Beamformers that preserve the desired signal are, for example, the linearly constrained minimum variance (LCMV) beamformer that is supposed, ideally, to reject the undesired signal and reduce the ambient noise power, and the minimum variance distortionless response (MVDR) beamformer that reduces the interference-plus-noise power. The multichannel Wiener filter, on the other hand, reduces the interference-plus-noise power without preserving the desired signal. In this paper, a speech distortion and interference rejection constraint (SDIRC) beamformer is derived that minimizes the ambient noise power subject to specific constraints that allow a tradeoff between speech distortion and interference-plus-noise reduction on the one hand, and undesired signal and ambient noise reductions on the other hand. Closed-form expressions for the performance measures of the SDIRC beamformer are derived and the relations to the aforementioned beamformers are derived. The performance evaluation demonstrates the tradeoffs that can be made using the SDIRC beamformer.

Index Terms—Beamforming, frequency domain, linearly constrained minimum variance (LCMV), minimum variance distortionless response (MVDR), noise reduction, parameterized multichannel Wiener filter, speech enhancement.

I. INTRODUCTION

DISTANT or hands-free speech acquisition is required in many applications such as hearing aids and teleconferencing. Microphone arrays are often used for the acquisition and consist of sets of microphones that are arranged in specific geometries. The received microphone signals usually consist of a mixture of desired source signals, undesired source signals, and ambient noise. As the acoustic interference degrades the quality and intelligibility of the desired sources, the received signals are processed (i.e., filtered and summed) in order to extract the desired signals or in other words, reduce the interference.

Copyright (c) 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Manuscript received October 16, 2010; revised February 15, 2011 and July 21, 2011; accepted August 09, 2011. Date of publication September 01, 2011; date of current version January 18, 2012. This work was supported by the Marie Curie Intra European Fellowship within the 7th European Community Framework Programme under contract number PIEF-GA-2009-237246. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Michael Lewis Seltzer.

E. A. P. Habets is with the International Audio Laboratories Erlangen, 91058 Erlangen, Germany (e-mail: emanuel.habets@audiolabs-erlangen.de).

J. Benesty is with the INRS-EMT, University of Quebec, Montreal, QC H3C 3P8, Canada (e-mail: benesty@emt.inrs.ca).

P. A. Naylor is with Imperial College London, London SW7 2AZ, U.K. (e-mail: p.naylor@imperial.ac.uk).

In the last four decades, numerous spatio-temporal filters have been proposed to process the received microphone signals (see [1] and [2] and the references therein). The process of filtering the received microphone signals and summing up all filtered output signals is widely known as beamforming and the filter-and-sum system is called a beamformer. Many filters were originally developed for, and used in, wireless communication systems. More recently filters were developed specifically for speech communication systems. Existing beamformers can be divided into those that preserve or distort the desired signal. Beamformers that preserve the desired signal are, for example, the minimum variance distortionless response (MVDR) beamformer (also known as Capon's beamformer) [3] that reduces the interference-plus-noise power, and the linearly constrained minimum variance (LCMV) beamformer [4] that is supposed, ideally, to reject the undesired signal and reduce the ambient noise power. The LCMV is a generalization of the MVDR and commonly aims at minimizing the beamformer's output power while satisfying multiple constraints such as rejecting the interference and passing the desired signal through undistorted. Another beamformer that consists of a weighted sum of the LCMV and a matched filter (i.e., an MVDR that reduces ambient noise only) has been recently proposed by Souden et al. [5]. The proposed beamformer allows a tradeoff between the undesired signal and ambient noise reductions. The multichannel Wiener filter (MWF), on the other hand, reduces the interference-plus-noise power without exactly preserving the desired signal. In order to control the amount of distortion, the parameterized multichannel Wiener filter (also known as speech-distortion weighted multichannel Wiener filter) has been proposed [6], [7].

In addition to closed-form beamformers, adaptive beamformers have been proposed. In [8], Frost proposed an adaptive scheme of the MVDR beamformer, which is based on a constrained least-mean-square-type adaptation. To avoid the constrained adaptation of the MVDR beamformer, Griffiths and Jim [9] proposed the generalized sidelobe canceller (GSC) structure that separates the output power minimization and the application of the constraint. While Griffiths and Jim only considered one constraint, it was later shown in [10] that the GSC structure can also be used in the case of multiple constraints. The original GSC structure is based on the assumption that the different microphones receive a delayed version of the desired signal. The GSC structure was later re-derived in the frequency domain, and extended to deal with general acoustic transfer functions (ATFs) [11].

Besides the great endeavors to develop reliable beamformers to reduce interference, many contributions have been made

to understand their functioning and accurately quantify their gains and losses in terms of speech distortion and noise reduction. In this context, it is worthwhile noting that the choice of the desired signal can have a significant impact on the information required to compute the beamformer and on the performance of the beamformer. In [12], the authors analysed the performance of the MVDR in different noise fields (coherent and non-coherent) and studied the effectiveness of the MVDR when designed to remove both additive noise and reverberation. The study showed that a tradeoff between noise reduction and dereverberation has to be made. In [13], Bitzer et al. investigated the theoretical performance limits of the GSC in the case of a spatially diffuse noise. The tradeoff between noise reduction and speech distortion in the parameterized multichannel Wiener filtering was established in [7]. Another notable effort to understand the functioning of the TF ratios-based GSC beamformer was published by Gannot and Cohen in [2]. They found that it is theoretically possible to achieve infinite noise reduction when only a coherent interference is added to the desired source. In the presence of ambient noise, the latter can also be achieved by the LCMV beamformer. Unfortunately, rejecting the coherent interference comes at a price as the noise reduction of the LCMV is always smaller than that of the MVDR [14].

In this paper, we assume that both an undesired source and an ambient noise coexist with the desired source. This assumption is very plausible in the context of hearing aids and hands-free communication devices that are installed within a teleconferencing room. In the latter situation for instance, the desired sound can originate from a speaker while the undesired sound originates from a loudspeaker or devices such as an air conditioner or a computer located within the same room. In addition, ambient noise as well as sensor noise is always present and should therefore be taken into account. To design an effective speech acquisition system, a clear understanding of the functioning of noise reduction filters as well as the ability to control various tradeoffs is paramount. In this contribution, a novel speech distortion and interference rejection constraint (SDIRC) beamformer is derived that minimizes the ambient noise power subject to specific constraints that allow a tradeoff between speech distortion and interference-plus-noise reduction on the one hand, and undesired signal and ambient noise reductions on the other hand. Closed-form expressions for the performance measures of the SDIRC beamformer are derived and the relations to some of the aforementioned beamformers are deduced. The performance evaluation demonstrates the tradeoffs that can be made using the SDIRC beamformer.

This paper is organized as follows. Section II describes the signal model, definitions, and fundamental assumptions made in this paper. The proposed SDIRC beamformer is derived in Section III. In Section IV, the relationship between the SDIRC beamformer and the MVDR, LCMV, and parameterized MWF are established. In Section V performance measures are defined and closed-form expressions for the aforementioned beamformers are deduced. Section VI investigates the performance of the new SDIRC beamformer with a special focus on the tradeoff between speech distortion and interference-

plus-noise reduction on the one hand, and undesired signal and ambient noise reductions on the other hand. Finally, Section VII provides some concluding remarks.

II. SIGNAL MODEL

Consider the signal model in which an M -element microphone array captures one desired and one undesired coherent source signal in some ambient noise field. We assume that all signals are broadband, and that all source signals and ambient noise are mutually independent and zero mean. The desired and undesired source signals as received by the m th microphone can be expressed as

$$d_m(t) = g_{d,m}(t) * s_d(t) \quad (1)$$

and

$$u_m(t) = g_{u,m}(t) * s_u(t), \quad (2)$$

where $*$ is the convolution operation, and $g_{d,m}(t)$ and $g_{u,m}(t)$ are the acoustic impulse responses between the desired source signal $s_d(t)$ and undesired source signal $s_u(t)$, respectively. The microphone signals are given by

$$y_m(t) = d_m(t) + u_m(t) + v_m(t), \quad m = 1, 2, \dots, M, \quad (3)$$

where $v_m(t)$ denotes the ambient noise received by the m th microphone.

In the discrete-time Fourier transform (DTFT) domain the received signals are expressed as

$$Y_m(\omega) = G_{d,m}(\omega)S_d(\omega) + G_{u,m}(\omega)S_u(\omega) + V_m(\omega) \quad (4)$$

$$= D_m(\omega) + U_m(\omega) + V_m(\omega), \quad m = 1, 2, \dots, M,$$

where $Y_m(\omega)$, $G_{d,m}(\omega)$, $G_{u,m}(\omega)$, $S_d(\omega)$, $S_u(\omega)$, $D_m(\omega)$, $U_m(\omega)$, and $V_m(\omega)$ are the DTFTs of $y_m(t)$, $g_{d,m}(t)$, $g_{u,m}(t)$, $s_d(t)$, $s_u(t)$, $d_m(t)$, $u_m(t)$, and $v_m(t)$, at angular frequency ω ($-\pi < \omega \leq \pi$).

We now express the M microphone signals in the frequency domain as a function of the desired and undesired signals as received by the first microphone¹ in a vector notation as

$$\mathbf{y}(\omega) = \mathbf{a}_d(\omega)D_1(\omega) + \mathbf{a}_u(\omega)U_1(\omega) + \mathbf{v}(\omega) \quad (5)$$

$$= \mathbf{d}(\omega) + \mathbf{u}(\omega) + \mathbf{v}(\omega),$$

where

$$\mathbf{y}(\omega) = [Y_1(\omega) \quad Y_2(\omega) \quad \dots \quad Y_M(\omega)]^T, \quad (6)$$

$$\mathbf{a}_d(\omega) = \left[\begin{array}{ccc} 1 & \frac{G_{d,2}(\omega)}{G_{d,1}(\omega)} & \dots & \frac{G_{d,M}(\omega)}{G_{d,1}(\omega)} \end{array} \right]^T,$$

$$\mathbf{a}_u(\omega) = \left[\begin{array}{ccc} 1 & \frac{G_{u,2}(\omega)}{G_{u,1}(\omega)} & \dots & \frac{G_{u,M}(\omega)}{G_{u,1}(\omega)} \end{array} \right]^T, \quad (7)$$

$$\mathbf{d}(\omega) = [D_1(\omega) \quad D_2(\omega) \quad \dots \quad D_M(\omega)]^T,$$

$$\mathbf{u}(\omega) = [U_1(\omega) \quad U_2(\omega) \quad \dots \quad U_M(\omega)]^T,$$

$$\mathbf{v}(\omega) = [V_1(\omega) \quad V_2(\omega) \quad \dots \quad V_M(\omega)]^T,$$

and superscript T denotes transpose of a vector or a matrix.

It is important to note that $\mathbf{a}_d(\omega)$ and $\mathbf{a}_u(\omega)$ contain the relative transfer functions (RTFs) from the M microphones to

¹Here the first microphone was chosen as a reference. In principle, any other microphone can be used as a reference.

the reference microphone (in this case the first microphone) for the desired source and undesired source, respectively. These RTFs can be estimated online using for example the methods described in [11], [15].

The beamforming is then performed by applying a complex weight to each microphone and summing across all microphones:

$$\begin{aligned} Z(\omega) &= \mathbf{h}^H(\omega)\mathbf{y}(\omega) \\ &= \mathbf{h}^H(\omega) [\mathbf{a}_d(\omega)D_1(\omega) + \mathbf{a}_u(\omega)U_1(\omega) + \mathbf{v}(\omega)], \end{aligned} \quad (8)$$

where $Z(\omega)$ is the beamformer output, $\mathbf{h}(\omega) = [H_1(\omega) H_2(\omega) \dots H_M(\omega)]^T$ is the beamforming weight vector which is suitable for performing spatial filtering at frequency ω , and superscript H denotes transpose conjugation of a vector or a matrix.

Let us define the power spectral density (PSD) of a random variable $B(\omega)$ and the PSD matrix of a column vector $\mathbf{b}(\omega)$ of length M as

$$\phi_b(\omega) = E [|B(\omega)|^2] \quad (10)$$

and

$$\Phi_{\mathbf{b}}(\omega) = E [\mathbf{b}(\omega)\mathbf{b}^H(\omega)], \quad (11)$$

with $E(\cdot)$ denoting mathematical expectation. The PSD of the received signal at the m th microphone can be expressed as

$$\phi_{y_m}(\omega) = \phi_{d_m}(\omega) + \phi_{u_m}(\omega) + \phi_{v_m}(\omega), \quad (12)$$

where $\phi_{y_m}(\omega)$, $\phi_{d_m}(\omega)$, $\phi_{u_m}(\omega)$, and $\phi_{v_m}(\omega)$ are the PSDs of the m th microphone signal, the desired signal at the m th microphone, the undesired signal at the m th microphone and the ambient noise signal at the m th microphone, respectively. The PSD of the beamformer output is given by

$$\begin{aligned} \phi_z(\omega) &= \mathbf{h}^H(\omega)\Phi_{\mathbf{d}}(\omega)\mathbf{h}(\omega) \\ &+ \mathbf{h}^H(\omega)\Phi_{\mathbf{u}}(\omega)\mathbf{h}(\omega) + \mathbf{h}^H(\omega)\Phi_{\mathbf{v}}(\omega)\mathbf{h}(\omega), \end{aligned} \quad (13)$$

where

$$\Phi_{\mathbf{d}}(\omega) = \phi_{d_1}(\omega)\mathbf{a}_d(\omega)\mathbf{a}_d^H(\omega), \quad (14)$$

$$\Phi_{\mathbf{u}}(\omega) = \phi_{u_1}(\omega)\mathbf{a}_u(\omega)\mathbf{a}_u^H(\omega), \quad (15)$$

and

$$\Phi_{\mathbf{v}}(\omega) = E [\mathbf{v}(\omega)\mathbf{v}^H(\omega)] \quad (16)$$

is the PSD matrix of the noise field. In the rest of this paper, we assume that $\Phi_{\mathbf{v}}(\omega)$ is a full-rank matrix such that its inverse exists.

In this paper, we assume that the ambient noise is sufficiently stationary and that we can identify periods where only the noise is active, other periods where either the desired or the undesired source is active (in addition to the noise), and other periods where all sources are active (in addition to the noise). Based on these assumptions, one can take advantage of the fact that the signals are mutually independent to calculate $\Phi_{\mathbf{d}}(\omega)$, $\Phi_{\mathbf{u}}(\omega)$, and $\Phi_{\mathbf{v}}(\omega)$ separately. These PSD matrices are required in the sequel to compute the beamformers.

III. SPEECH-DISTORTION AND INTERFERENCE-REJECTION CONSTRAINT BEAMFORMER

In this section, we derive a speech-distortion and interference-rejection constraint beamformer that is able to tradeoff between speech distortion and interference-plus-noise reduction on the one hand, and undesired signal and ambient noise reductions on the other hand.

In general, the aforementioned tradeoffs can be realized by modifying the optimization problem. In earlier work the speech distortion was controlled by augmenting the classical mean-square error (MSE) cost function [6] or by imposing a linear inequality constraint [7], [16], [17]. It should be noted that the inequality constraint on the PSD of the desired signal at the beamformer's output (as used in [7], [17]) is binding. Consequently, the same solution is obtained if the inequality constraint is replaced by an equality constraint. Therefore, we propose to control the distortion of the desired speech signal as received by the first microphone using the following equality constraint:

$$\mathbf{h}^H(\omega)\mathbf{a}_d(\omega) = \alpha(\omega), \quad (17)$$

where $\alpha(\omega)$ is a complex number. The closer is the value of $|\alpha(\omega)|^2$ to one, the less the amplitude response of desired signal is distorted; for $\alpha(\omega) = 1$, there is no distortion. When the phase response of $\alpha(\omega)$ is linear the desired signal at the beamformer's output is a delayed version of the desired signal as received by the reference microphone. For other phase responses of $\alpha(\omega)$ (unequal to zero) the desired signal might contain audible distortions. The parameter $\alpha(\omega)$ in (17) can be varied with respect to the frequency depending on, for example, the PSDs of the desired signal and ambient noise. It is worthwhile noting that multiplying $\alpha(\omega)$ by a frequency independent scale factor controls the gain of the desired signal at the output of the beamformer.

The same idea can be applied in order to tradeoff between reduction of the undesired signal and ambient noise. Thus, we have

$$\mathbf{h}^H(\omega)\mathbf{a}_u(\omega) = \beta(\omega), \quad (18)$$

where $\beta(\omega)$ is a complex number. The closer the value of $|\beta(\omega)|^2$ is to zero, the more the undesired signal is reduced; for $\beta(\omega) = 0$, the undesired signal is completely removed. Putting these constraints together we obtain

$$\mathbf{A}^H(\omega)\mathbf{h}(\omega) = \mathbf{q}(\omega), \quad (19)$$

where

$$\begin{aligned} \mathbf{A}(\omega) &= [\mathbf{a}_d(\omega) \quad \mathbf{a}_u(\omega)], \\ \mathbf{q}(\omega) &= [\alpha(\omega) \quad \beta(\omega)]^H. \end{aligned}$$

The SDIRC beamformer is obtained by minimizing the power at the beamformer output subject to (19), i.e.,

$$\begin{aligned} \mathbf{h}_{\text{SDIRC}}(\omega) &= \arg \min_{\mathbf{h}(\omega)} \mathbf{h}^H(\omega)\Phi_{\mathbf{y}}(\omega)\mathbf{h}(\omega) \\ &\text{subject to } \mathbf{A}^H(\omega)\mathbf{h}(\omega) = \mathbf{q}(\omega), \end{aligned} \quad (20)$$

for which the solution is given by²

$$\begin{aligned} \mathbf{h}_{\text{SDIRC}}(\omega) &= \Phi_{\mathbf{y}}^{-1}(\omega) \mathbf{A}(\omega) [\mathbf{A}^H(\omega) \Phi_{\mathbf{y}}^{-1}(\omega) \mathbf{A}(\omega)]^{-1} \mathbf{q}(\omega). \end{aligned} \quad (21)$$

Alternatively, one can minimize the power of the ambient noise at the beamformer output subject to (19) yielding

$$\begin{aligned} \mathbf{h}_{\text{SDIRC}}(\omega) &= \Phi_{\mathbf{v}}^{-1}(\omega) \mathbf{A}(\omega) [\mathbf{A}^H(\omega) \Phi_{\mathbf{v}}^{-1}(\omega) \mathbf{A}(\omega)]^{-1} \mathbf{q}(\omega). \end{aligned} \quad (22)$$

Using the Woodbury's matrix identity, it can be shown that (21) and (22) are mathematically equivalent.

The SDIRC beamformer can be interpreted as a two stage spatial processor that first computes two signals given by $\mathbf{A}^H(\omega) \Phi_{\mathbf{v}}^{-1}(\omega) \mathbf{y}(\omega)$. Finally, these signals are combined using $\mathbf{q}^H(\omega) [\mathbf{A}^H(\omega) \Phi_{\mathbf{v}}^{-1}(\omega) \mathbf{A}(\omega)]^{-1}$ to compute the output of the SDIRC beamformer $Z(\omega)$.

IV. RELATION TO LCMV, MVDR, AND PMWF

In this section, we show how the proposed SDIRC beamformer is related to the well known LCMV, MVDR, and parameterized multichannel Wiener filter (PMWF).

For convenience, we first rewrite (22) as

$$\mathbf{h}_{\text{SDIRC}}(\omega) = \alpha^*(\omega) \mathbf{h}_1(\omega) + \beta^*(\omega) \mathbf{h}_2(\omega), \quad (23)$$

where

$$\begin{aligned} \begin{bmatrix} \mathbf{h}_1(\omega) & \mathbf{h}_2(\omega) \end{bmatrix} &= \Phi_{\mathbf{v}}^{-1}(\omega) \mathbf{A}(\omega) [\mathbf{A}^H(\omega) \Phi_{\mathbf{v}}^{-1}(\omega) \mathbf{A}(\omega)]^{-1}. \end{aligned} \quad (24)$$

A. LCMV

The LCMV beamformer completely rejects the undesired signal while preserving the desired signal. This can be achieved using

$$\alpha_{\text{LCMV}}(\omega) = 1, \quad (25)$$

$$\beta_{\text{LCMV}}(\omega) = 0. \quad (26)$$

Using (23) we directly obtain the well-known solution of the LCMV beamformer [1], [18]:

$$\mathbf{h}_{\text{LCMV}}(\omega) = \mathbf{h}_1(\omega). \quad (27)$$

B. MVDR

The MVDR beamformer can be obtained by jointly preserving the desired signal (i.e., by setting $\alpha(\omega) = 1$) and minimizing either the total signal output power of beamformer or the interference-plus-noise power at the output of the beamformer. Therefore, the MVDR beamformer can be expressed as

$$\mathbf{h}_{\text{MVDR}}(\omega) = \mathbf{h}_1(\omega) + \beta_{\text{MVDR}}^*(\omega) \mathbf{h}_2(\omega). \quad (28)$$

Since $\mathbf{a}_d^H(\omega) \mathbf{h}_{\text{MVDR}}(\omega) = 1$ and $\mathbf{a}_d^H(\omega) \mathbf{h}_{\text{LCMV}}(\omega) = \mathbf{a}_d^H(\omega) \mathbf{h}_1(\omega) = 1$, we deduce that $\mathbf{a}_d^H(\omega) \mathbf{h}_2(\omega) = 0$ and

²Strictly speaking the SDIRC is also an LCMV beamformer. Similar to other works [1], [4], we follow the protocol of referring to the LCMV when we aim at completely suppressing the interference [i.e., $\beta(\omega) = 0 \forall \omega$].

therefore $\beta_{\text{MVDR}}(\omega)$ cannot be found from (28). However, since the output signal-to-interference-plus-noise ratio (SINR) is upper bounded and its maximum is obtained with the MVDR beamformer [1], we can derive $\beta_{\text{MVDR}}(\omega)$ using the output SINR of the SDIRC beamformer which is given by³

$$\begin{aligned} \text{oSINR}[\mathbf{h}_{\text{SDIRC}}] &= \frac{\mathbf{h}_{\text{SDIRC}}^H \Phi_{\mathbf{d}} \mathbf{h}_{\text{SDIRC}}}{\mathbf{h}_{\text{SDIRC}}^H [\Phi_{\mathbf{u}} + \Phi_{\mathbf{v}}] \mathbf{h}_{\text{SDIRC}}} \\ &= \frac{\phi_{d_1} |\alpha|^2}{\phi_{u_1} |\beta|^2 + \mathbf{h}_{\text{SDIRC}}^H \Phi_{\mathbf{v}} \mathbf{h}_{\text{SDIRC}}} \\ &= \frac{\phi_{d_1} |\alpha|^2}{m_1 |\alpha|^2 + [\phi_{u_1} + m_2] |\beta|^2 + 2 \text{Re}\{m_{12} \alpha \beta\}}, \end{aligned} \quad (29)$$

where

$$m_1(\omega) = \mathbf{h}_1^H(\omega) \Phi_{\mathbf{v}}(\omega) \mathbf{h}_1(\omega), \quad (30)$$

$$m_2(\omega) = \mathbf{h}_2^H(\omega) \Phi_{\mathbf{v}}(\omega) \mathbf{h}_2(\omega), \quad (31)$$

$$m_{12}(\omega) = \mathbf{h}_1^H(\omega) \Phi_{\mathbf{v}}(\omega) \mathbf{h}_2(\omega) \quad (32)$$

and $\text{Re}\{\cdot\}$ provides the real part of the argument.

Now we can find $\beta_{\text{MVDR}}(\omega)$ by taking $\alpha(\omega) = 1$ and maximizing $\text{oSINR}[\mathbf{h}_{\text{SDIRC}}(\omega)]$ with respect to $\beta(\omega)$. It is important to note that the narrowband output SINR is a real function with complex variables. Following Brandwood [19] the derivative with respect to $\beta(\omega)$ or $\beta^*(\omega)$ (where the subscript * denotes complex conjugation) can be used to find a stationary point of a real function. Subsequently, equating the derivative with respect to either $\beta(\omega)$ or $\beta^*(\omega)$ to 0 and solving $\beta(\omega)$ yields

$$\beta_{\text{MVDR}}(\omega) = -\frac{m_{12}(\omega)}{\phi_{u_1}(\omega) + m_2(\omega)}. \quad (33)$$

As a result, another way to write the MVDR is

$$\mathbf{h}_{\text{MVDR}}(\omega) = \mathbf{h}_1(\omega) - \frac{m_{12}^*(\omega)}{\phi_{u_1}(\omega) + m_2(\omega)} \mathbf{h}_2(\omega). \quad (34)$$

The latter expression shows that $\beta_{\text{MVDR}}(\omega)$ depends on $\mathbf{h}_1(\omega)$, $\mathbf{h}_2(\omega)$, $\Phi_{\mathbf{v}}(\omega)$, and the PSD of the undesired source as received by the first microphone, denoted by $\phi_{u_1}(\omega)$. It is worthwhile to note that in this form the MVDR uses the RTFs $\mathbf{a}_u(\omega)$ and $\phi_{u_1}(\omega)$ separately while the classic MVDR beamformer uses $\Phi_{\mathbf{u}}(\omega) = \phi_{u_1}(\omega) \mathbf{a}_u(\omega) \mathbf{a}_u^H(\omega)$.

C. PMWF

Finally, we show the relation between the SDIRC beamformer and the the PMWF that is given by [7]:

$$\begin{aligned} \mathbf{h}_{\text{PMWF}}(\omega; \gamma) &= (\Phi_{\mathbf{d}}(\omega) + \gamma [\Phi_{\mathbf{u}}(\omega) + \Phi_{\mathbf{v}}(\omega)])^{-1} \Phi_{\mathbf{d}}(\omega) \mathbf{i}_1 \\ &= \frac{[\Phi_{\mathbf{u}}(\omega) + \Phi_{\mathbf{v}}(\omega)]^{-1} \Phi_{\mathbf{d}}(\omega)}{\gamma + \text{tr}\{[\Phi_{\mathbf{u}}(\omega) + \Phi_{\mathbf{v}}(\omega)]^{-1} \Phi_{\mathbf{d}}(\omega)\}} \mathbf{i}_1, \end{aligned} \quad (35)$$

where γ ($\gamma \geq 0$) controls the amount of speech distortion and

$$\mathbf{i}_1 = [1 \ 0 \ \cdots \ 0]^T.$$

³Where ω has been omitted for conciseness.

To show the relation between the parameterized MWF filter and the SCIRC beamformer we use the following relation between the MVDR and PMWF:

$$\mathbf{h}_{\text{PMWF}}(\omega; \gamma) = H_{\text{PW}}(\omega; \gamma) \mathbf{h}_{\text{MVDR}}(\omega). \quad (36)$$

The gain $H_{\text{PW}}(\omega; \gamma)$ can be seen as a parameterized single-channel Wiener gain that is given by

$$H_{\text{PW}}(\omega; \gamma) = \frac{\text{tr} \left\{ [\Phi_{\mathbf{u}}(\omega) + \Phi_{\mathbf{v}}(\omega)]^{-1} \Phi_{\mathbf{d}}(\omega) \right\}}{\gamma + \text{tr} \left\{ [\Phi_{\mathbf{u}}(\omega) + \Phi_{\mathbf{v}}(\omega)]^{-1} \Phi_{\mathbf{d}}(\omega) \right\}}, \quad (37)$$

where $\text{tr}\{\cdot\}$ denotes the trace of a square matrix. When $\gamma = 0$ we obtain $H_{\text{PW}}(\omega; 0) = 1$, $\forall \omega$ such that $\mathbf{h}_{\text{PMWF}}(\omega; 0) = \mathbf{h}_{\text{MVDR}}(\omega)$.

Using (36) and (34), we can also deduce another form for the PMWF:

$$\begin{aligned} \mathbf{h}_{\text{PMWF}}(\omega; \gamma) &= H_{\text{PW}}(\omega; \gamma) \mathbf{h}_1(\omega) \\ &\quad - H_{\text{PW}}(\omega; \gamma) \frac{m_{12}^*(\omega)}{\phi_{u_1}(\omega) + m_2(\omega)} \mathbf{h}_2(\omega). \end{aligned} \quad (38)$$

Hence,

$$\alpha_{\text{PMWF}}(\omega; \gamma) = H_{\text{PW}}(\omega; \gamma), \quad (39)$$

$$\beta_{\text{PMWF}}(\omega; \gamma) = -H_{\text{PW}}(\omega; \gamma) \frac{m_{12}(\omega)}{\phi_{u_1}(\omega) + m_2(\omega)}. \quad (40)$$

To obtain results with the SDIRC beamformer equal to those obtained by the parameterized MWF we require $\Phi_{\mathbf{u}}(\omega)$ in addition to $\Phi_{\mathbf{v}}(\omega)$ and the RTFs $\mathbf{a}_{\mathbf{d}}(\omega)$ and $\mathbf{a}_{\mathbf{u}}(\omega)$.

V. PERFORMANCE MEASURES AND ANALYSES

We define and use different performance measure to analyze how the SDIRC beamforming operates in an acoustic environment. The performance measures will also be used for the performance evaluation in Section VI. In addition, we deduce the performance measures of existing beamformer using the relations derived in Section IV. The SDIRC beamformer and the aforementioned relations provide significantly simpler derivations of the important performance measures, some of which have been presented elsewhere [1], [2].

Since the signal we want to recover is the clean (but convolved) signal received at microphone 1, i.e., $d_1(t)$, microphone 1 is serving as the reference microphone.

A. Input and Output Signal-to-Interference-plus-Noise Ratios

We define the narrowband input signal-to-noise ratio (SNR), signal-to-interference ratio (SIR), and signal-to-interference-plus-noise ratio (SINR) as

$$\text{iSNR}(\omega) = \frac{\phi_{d_1}(\omega)}{\phi_{v_1}(\omega)}, \quad (41)$$

$$\text{iSIR}(\omega) = \frac{\phi_{d_1}(\omega)}{\phi_{u_1}(\omega)}, \quad (42)$$

and

$$\begin{aligned} \text{iSINR}(\omega) &= \frac{\phi_{d_1}(\omega)}{\phi_{u_1}(\omega) + \phi_{v_1}(\omega)} \\ &= \frac{\text{iSIR}(\omega) \text{iSNR}(\omega)}{\text{iSIR}(\omega) + \text{iSNR}(\omega)}, \end{aligned} \quad (43)$$

respectively.

Fullband performance measures can be obtained by integrating the numerator and denominator across ω , e.g., the fullband iSINR is given by

$$\text{iSINR} = \frac{\int_{-\pi}^{+\pi} \phi_{d_1}(\omega) d\omega}{\int_{-\pi}^{+\pi} [\phi_{u_1}(\omega) + \phi_{v_1}(\omega)] d\omega}. \quad (44)$$

To quantify the level of noise remaining in the output signal of the beamformer, $Z(\omega)$, we define the narrowband output SNR, SIR, and SINR as the ratio of the power of the filtered desired signal over the power of the residual noises, i.e.,

$$\begin{aligned} \text{oSNR}[\mathbf{h}(\omega)] &= \frac{\mathbf{h}^H(\omega) \Phi_{\mathbf{d}}(\omega) \mathbf{h}(\omega)}{\mathbf{h}^H(\omega) \Phi_{\mathbf{v}}(\omega) \mathbf{h}(\omega)} \\ &= \frac{\phi_{d_1}(\omega) |\mathbf{h}^H(\omega) \mathbf{a}_{\mathbf{d}}(\omega)|^2}{\mathbf{h}^H(\omega) \Phi_{\mathbf{v}}(\omega) \mathbf{h}(\omega)}, \end{aligned} \quad (45)$$

$$\begin{aligned} \text{oSIR}[\mathbf{h}(\omega)] &= \frac{\mathbf{h}^H(\omega) \Phi_{\mathbf{d}}(\omega) \mathbf{h}(\omega)}{\mathbf{h}^H(\omega) \Phi_{\mathbf{u}}(\omega) \mathbf{h}(\omega)} \\ &= \frac{\phi_{d_1}(\omega) |\mathbf{h}^H(\omega) \mathbf{a}_{\mathbf{d}}(\omega)|^2}{\phi_{u_1}(\omega) |\mathbf{h}^H(\omega) \mathbf{a}_{\mathbf{u}}(\omega)|^2}, \end{aligned} \quad (46)$$

and

$$\begin{aligned} \text{oSINR}[\mathbf{h}(\omega)] &= \frac{\mathbf{h}^H(\omega) \Phi_{\mathbf{d}}(\omega) \mathbf{h}(\omega)}{\mathbf{h}^H(\omega) [\Phi_{\mathbf{u}}(\omega) + \Phi_{\mathbf{v}}(\omega)] \mathbf{h}(\omega)} \\ &= \frac{\phi_{d_1}(\omega) |\mathbf{h}^H(\omega) \mathbf{a}_{\mathbf{d}}(\omega)|^2}{\phi_{u_1}(\omega) |\mathbf{h}^H(\omega) \mathbf{a}_{\mathbf{u}}(\omega)|^2 + \mathbf{h}^H(\omega) \Phi_{\mathbf{v}}(\omega) \mathbf{h}(\omega)}. \end{aligned} \quad (47)$$

Because the oSINR depends on both the speech distortion and the interference-plus-noise, it will be used later to demonstrate the tradeoff between these two quantities. The oSINR is related to the oSNR and oSIR in a similar way the iSINR is related to the iSNR and iSIR, i.e.,

$$\text{oSINR}[\mathbf{h}(\omega)] = \frac{\text{oSIR}[\mathbf{h}(\omega)] \text{oSNR}[\mathbf{h}(\omega)]}{\text{oSIR}[\mathbf{h}(\omega)] + \text{oSNR}[\mathbf{h}(\omega)]}. \quad (48)$$

Let us consider a scenario with one undesired source and spatially white ambient noise such that $\Phi_{\mathbf{u}}(\omega) = \phi_{u_1}(\omega) \mathbf{a}_{\mathbf{u}}(\omega) \mathbf{a}_{\mathbf{u}}^H(\omega)$ and $\Phi_{\mathbf{v}}(\omega) = \phi_{v_1}(\omega) \mathbf{I}$. We can now write $\mathbf{h}_1(\omega)$ and $\mathbf{h}_2(\omega)$ explicitly as:

$$\mathbf{h}_1(\omega) = \frac{\mathbf{a}_{\mathbf{d}}(\omega) \|\mathbf{a}_{\mathbf{u}}(\omega)\|^2 - \mathbf{a}_{\mathbf{u}}(\omega) \mathbf{a}_{\mathbf{u}}^H(\omega) \mathbf{a}_{\mathbf{d}}(\omega)}{\lambda(\omega)}, \quad (49)$$

$$\mathbf{h}_2(\omega) = \frac{\mathbf{a}_{\mathbf{u}}(\omega) \|\mathbf{a}_{\mathbf{d}}(\omega)\|^2 - \mathbf{a}_{\mathbf{d}}(\omega) \mathbf{a}_{\mathbf{d}}^H(\omega) \mathbf{a}_{\mathbf{u}}(\omega)}{\lambda(\omega)}, \quad (50)$$

with

$$\lambda(\omega) = \|\mathbf{a}_{\mathbf{d}}(\omega)\|^2 \|\mathbf{a}_{\mathbf{u}}(\omega)\|^2 - |\mathbf{a}_{\mathbf{d}}^H(\omega) \mathbf{a}_{\mathbf{u}}(\omega)|^2. \quad (51)$$

Furthermore, we obtain for $m_1(\omega)$, $m_2(\omega)$ and $m_{12}(\omega)$:

$$m_1(\omega) = \frac{\phi_{v_1}(\omega) \|\mathbf{a}_u(\omega)\|^2}{\lambda(\omega)}, \quad (52)$$

$$m_2(\omega) = \frac{\phi_{v_1}(\omega) \|\mathbf{a}_d(\omega)\|^2}{\lambda(\omega)}, \quad (53)$$

$$m_{12}(\omega) = -\frac{\phi_{v_1}(\omega) \mathbf{a}_d^H(\omega) \mathbf{a}_u(\omega)}{\lambda(\omega)}. \quad (54)$$

Using (29) and (52)-(54) we can now deduce the output SINR for the SDIRC beamformer:³

$$\begin{aligned} \text{oSINR} [\mathbf{h}_{\text{SDIRC}}] &= \frac{\phi_{d_1} |\alpha|^2}{\phi_{u_1} |\beta|^2 + \mathbf{h}_{\text{SDIRC}}^H \Phi_{\mathbf{v}} \mathbf{h}_{\text{SDIRC}}} \\ &= \frac{\phi_{d_1} |\alpha|^2}{m_1 |\alpha|^2 + [\phi_{u_1} + m_2] |\beta|^2 + 2 \text{Re}\{m_{12} \alpha \beta\}} \\ &= \frac{\phi_{d_1} |\alpha|^2}{\phi_{u_1} |\beta|^2 + \phi_{v_1} \lambda^{-1} \kappa}, \end{aligned} \quad (55)$$

where

$$\begin{aligned} \kappa(\omega) &= |\alpha(\omega)|^2 \|\mathbf{a}_u(\omega)\|^2 + |\beta(\omega)|^2 \|\mathbf{a}_d(\omega)\|^2 \\ &\quad - 2 \text{Re} \{ \mathbf{a}_d^H(\omega) \mathbf{a}_u(\omega) \alpha(\omega) \beta(\omega) \}. \end{aligned} \quad (56)$$

Using the previous definitions for $\alpha(\omega)$ and $\beta(\omega)$ given in Section IV, we find the output SINR for the MVDR, LCMV, and PMWF given by

$$\begin{aligned} \text{oSINR} [\mathbf{h}_{\text{LCMV}}(\omega)] &= \frac{\phi_{d_1}(\omega)}{\mathbf{h}_{\text{LCMV}}^H(\omega) \Phi_{\mathbf{v}}(\omega) \mathbf{h}_{\text{LCMV}}(\omega)} \\ &= \frac{\phi_{d_1}(\omega)}{\mathbf{h}_1^H(\omega) \Phi_{\mathbf{v}}(\omega) \mathbf{h}_1} \\ &= \frac{\phi_{d_1}(\omega) \lambda(\omega)}{\phi_{v_1}(\omega) \|\mathbf{a}_u(\omega)\|^2} \\ &= \text{iSNR}(\omega) \left(\|\mathbf{a}_d(\omega)\|^2 - \frac{|\mathbf{a}_d^H(\omega) \mathbf{a}_u(\omega)|^2}{\|\mathbf{a}_u(\omega)\|^2} \right), \end{aligned} \quad (57)$$

$$\begin{aligned} \text{oSINR} [\mathbf{h}_{\text{MVDR}}(\omega)] &= \frac{\phi_{d_1}(\omega)}{\mathbf{h}_{\text{MVDR}}^H(\omega) [\Phi_{\mathbf{u}}(\omega) + \Phi_{\mathbf{v}}(\omega)] \mathbf{h}_{\text{MVDR}}(\omega)} \\ &= \phi_{d_1}(\omega) \mathbf{a}_d^H(\omega) [\Phi_{\mathbf{u}}(\omega) + \Phi_{\mathbf{v}}(\omega)]^{-1} \mathbf{a}_d(\omega) \\ &= \text{iSNR}(\omega) \left(\|\mathbf{a}_d(\omega)\|^2 - \frac{|\mathbf{a}_d^H(\omega) \mathbf{a}_u(\omega)|^2}{\frac{\phi_{v_1}(\omega)}{\phi_{u_1}(\omega)} + \|\mathbf{a}_u(\omega)\|^2} \right), \end{aligned} \quad (58)$$

$$\begin{aligned} \text{oSINR} [\mathbf{h}_{\text{PMWF}}(\omega; \gamma)] &= \frac{H_{\text{PW}}^2(\omega) \phi_{d_1}(\omega)}{\mathbf{h}_{\text{PMWF}}^H(\omega) [\Phi_{\mathbf{u}}(\omega) + \Phi_{\mathbf{v}}(\omega)] \mathbf{h}_{\text{PMWF}}(\omega)} \\ &= \frac{H_{\text{PW}}^2(\omega) \phi_{d_1}(\omega)}{H_{\text{PW}}^2(\omega) \mathbf{h}_{\text{MVDR}}^H(\omega) [\Phi_{\mathbf{u}}(\omega) + \Phi_{\mathbf{v}}(\omega)] \mathbf{h}_{\text{MVDR}}(\omega)} \\ &= \text{oSINR} [\mathbf{h}_{\text{MVDR}}(\omega)]. \end{aligned} \quad (59)$$

where for oSINR $[\mathbf{h}_{\text{MVDR}}(\omega)]$ we have used the fact that

$$[\Phi_{\mathbf{u}}(\omega) + \Phi_{\mathbf{v}}(\omega)]^{-1} = \frac{1}{\phi_{v_1}(\omega)} \left(\mathbf{I} - \frac{\mathbf{a}_u(\omega) \mathbf{a}_u^H(\omega)}{\frac{\phi_{v_1}(\omega)}{\phi_{u_1}(\omega)} + \|\mathbf{a}_u(\omega)\|^2} \right). \quad (60)$$

Hence, for the considered scenario, the narrowband output SINRs of the MVDR and PMWF are equal. In addition, their value is always larger than or equal to the output SINR of the LCMV as the ratio $\phi_{v_1}(\omega)/\phi_{u_1}(\omega)$ is always positive, i.e.,

$$\begin{aligned} \text{oSINR} [\mathbf{h}_{\text{PMWF}}(\omega; \gamma)] &= \text{oSINR} [\mathbf{h}_{\text{MVDR}}(\omega)] \\ &\geq \text{oSINR} [\mathbf{h}_{\text{LCMV}}(\omega)]. \end{aligned} \quad (61)$$

To demonstrate the tradeoff between undesired signal and ambient noise reductions we define the input and output noise-to-interference ratios (NIRs) as

$$\text{iNIR}(\omega) = \frac{\phi_{v_1}(\omega)}{\phi_{u_1}(\omega)} \quad (62)$$

and

$$\text{oNIR} [\mathbf{h}(\omega)] = \frac{\mathbf{h}^H(\omega) \Phi_{\mathbf{v}}(\omega) \mathbf{h}(\omega)}{\mathbf{h}^H(\omega) \Phi_{\mathbf{u}}(\omega) \mathbf{h}(\omega)}. \quad (63)$$

B. Noise Reduction

The narrowband noise-reduction factor [20], or narrowband noise-rejection factor [21] quantifies the amount of noise being rejected by the beamformer. This quantity is defined as the ratio of the power of the noise at the reference microphone over the power of the noise remaining at the beamformer output, i.e.,

$$\begin{aligned} \xi_{\text{nr}} [\mathbf{h}(\omega)] &= \frac{\phi_{u_1}(\omega) + \phi_{v_1}(\omega)}{\mathbf{h}^H(\omega) [\Phi_{\mathbf{u}}(\omega) + \Phi_{\mathbf{v}}(\omega)] \mathbf{h}(\omega)} \\ &= \frac{\phi_{u_1}(\omega) + \phi_{v_1}(\omega)}{\phi_{u_1}(\omega) |\mathbf{h}^H(\omega) \mathbf{a}_u(\omega)|^2 + \mathbf{h}^H(\omega) \Phi_{\mathbf{v}}(\omega) \mathbf{h}(\omega)}. \end{aligned} \quad (64)$$

The noise-rejection factor is expected to be lower bounded by 1. Otherwise, the beamformer amplifies the noise received at the microphones. It should be noted however that in some scenarios the noise-reduction factor can become lower than 1. The higher the value of the noise-rejection factor, the more the noise is rejected.

For the SDIRC beamformer we have³

$$\begin{aligned} \xi_{\text{nr}} [\mathbf{h}_{\text{SDIRC}}] &= \frac{\phi_{u_1} + \phi_{v_1}}{m_1 |\alpha|^2 + [\phi_{u_1} + m_2] |\beta|^2 + 2 \text{Re}\{m_{12} \alpha \beta\}}. \end{aligned} \quad (65)$$

Hence, the narrowband noise-reduction factor of the SDIRC beamformer depends on $\alpha(\omega)$ and $\beta(\omega)$ and approaches infinity when $\alpha(\omega)$ and $\beta(\omega)$ approach zero.

Let us again investigate the previously discussed scenario when $\Phi_{\mathbf{u}}(\omega) = \phi_{u_1}(\omega) \mathbf{a}_u(\omega) \mathbf{a}_u^H(\omega)$ and $\Phi_{\mathbf{v}}(\omega) = \phi_{v_1}(\omega) \mathbf{I}$. For the LCMV beamformer we then obtain

$$\begin{aligned} \xi_{\text{nr}} [\mathbf{h}_{\text{LCMV}}(\omega)] &= \left(\frac{\phi_{u_1}(\omega)}{\phi_{v_1}(\omega)} + 1 \right) \\ &\quad \left(\|\mathbf{a}_d(\omega)\|_2^2 - \frac{|\mathbf{a}_d^H(\omega) \mathbf{a}_u(\omega)|^2}{\|\mathbf{a}_u(\omega)\|_2^2} \right). \end{aligned} \quad (66)$$

The narrowband noise reduction for the MVDR beamformer is given by

$$\xi_{\text{nr}}[\mathbf{h}_{\text{MVDR}}(\omega)] = \left(\frac{\phi_{u_1}(\omega)}{\phi_{v_1}(\omega)} + 1 \right) \left(\|\mathbf{a}_d(\omega)\|_2^2 - \frac{|\mathbf{a}_d^H(\omega)\mathbf{a}_u(\omega)|^2}{\frac{\phi_{v_1}(\omega)}{\phi_{u_1}(\omega)} + \|\mathbf{a}_u(\omega)\|_2^2} \right). \quad (67)$$

When $|\mathbf{a}_d^H(\omega)\mathbf{a}_u(\omega)|^2 \neq 0$ the narrowband noise-reduction factor of the MVDR and LCMV differ depending on the ratio $\phi_{v_1}(\omega)/\phi_{u_1}(\omega)$. From these expressions we also see that the noise-reduction factor of the MVDR and LCMV are similar when $\phi_{u_1}(\omega)$ is much larger than $\phi_{v_1}(\omega)$.

Finally, the narrowband noise reduction for the PMWF is given by [7]

$$\xi_{\text{nr}}[\mathbf{h}_{\text{PMWF}}(\omega; \gamma)] = \frac{1}{\text{iSINR}(\omega)} \cdot \frac{[\gamma + \rho(\omega)]^2}{\rho(\omega)}, \quad (68)$$

with

$$\rho(\omega) = \text{tr} \left\{ [\Phi_{\mathbf{u}}(\omega) + \Phi_{\mathbf{v}}(\omega)]^{-1} \Phi_{\mathbf{d}}(\omega) \right\}. \quad (69)$$

Using the matrix inversion lemma it can be shown that for the considered scenario

$$\rho(\omega) = \frac{\phi_{d_1}(\omega)}{\phi_{v_1}(\omega)} \left(\|\mathbf{a}_d(\omega)\|_2^2 - \frac{|\mathbf{a}_d^H(\omega)\mathbf{a}_u(\omega)|^2}{\frac{\phi_{v_1}(\omega)}{\phi_{u_1}(\omega)} + \|\mathbf{a}_u(\omega)\|_2^2} \right). \quad (70)$$

Analyzing (68) we find that for a given situation and hence $\rho(\omega)$ ($\rho(\omega) > 0$) and $\text{iSINR}(\omega)$, the noise reduction of the PMWF increases monotonically with γ .

Because the ratio $\phi_{v_1}(\omega)/\phi_{u_1}(\omega)$ is always positive the following holds:

$$\xi_{\text{nr}}[\mathbf{h}_{\text{PMWF}}(\omega)] \geq \xi_{\text{nr}}[\mathbf{h}_{\text{MVDR}}(\omega)] \geq \xi_{\text{nr}}[\mathbf{h}_{\text{LCMV}}(\omega)]. \quad (71)$$

Hence, for the considered scenario the narrowband noise-reduction factor of the parameterized MWF is always larger or equal than the narrowband noise-reduction factor of the MVDR, which is always larger or equal to the narrowband noise reduction of the LCMV.

C. Speech Distortion

Many beamforming algorithms distort the desired signal. Even when the beamformer is designed to preserve the desired signal, distortions might be introduced due to estimation errors. In order to quantify the level of this distortion, we define the narrowband desired-signal-reduction factor [22] or narrowband desired-signal-cancellation factor [21] as the ratio of the variance of the desired signal at the reference microphone over the variance of the filtered desired signal at the beamformer output, i.e.,

$$\begin{aligned} \xi_{\text{dsc}}[\mathbf{h}(\omega)] &= \frac{\phi_{d_1}(\omega)}{\mathbf{h}^H(\omega)\Phi_{\mathbf{d}}(\omega)\mathbf{h}(\omega)} \\ &= \frac{1}{|\mathbf{h}^H(\omega)\mathbf{a}_d(\omega)|^2}. \end{aligned} \quad (72)$$

Broadband beamformers that do not cancel the broadband desired signal require that the desired-signal-cancellation factor

is equal to 1. When $\xi_{\text{dsc}}[\mathbf{h}(\omega)]$ is greater than 1 the desired signal is distorted.

Another useful performance measure is the speech-distortion index [20], [23] defined as

$$\begin{aligned} v_{\text{sd}}[\mathbf{h}(\omega)] &= \frac{E \left[|\mathbf{h}^H(\omega)\mathbf{a}_d(\omega)D_1(\omega) - D_1(\omega)|^2 \right]}{\phi_{d_1}(\omega)} \\ &= |\mathbf{h}^H(\omega)\mathbf{a}_d(\omega) - 1|^2. \end{aligned} \quad (73)$$

The speech-distortion index is always greater than or equal to 0 and should be upper bounded by 1; so the higher is the value of $v_{\text{sd}}[\mathbf{h}(\omega)]$, the more the desired signal is distorted.

By substituting (17) in (72) and (73), we obtain the distortion measures for the SDIRC beamformer

$$\xi_{\text{dsc}}[\mathbf{h}_{\text{SDIRC}}(\omega)] = \frac{1}{|\alpha(\omega)|^2}, \quad (74)$$

$$v_{\text{sd}}[\mathbf{h}_{\text{SDIRC}}(\omega)] = [\alpha(\omega) - 1]^2. \quad (75)$$

Hence, the distortion of the desired signal depends only on $\alpha(\omega)$.

For the LCMV and MVDR beamformer, we obtain

$$\xi_{\text{dsc}}[\mathbf{h}_{\text{LCMV}}(\omega)] = \xi_{\text{dsc}}[\mathbf{h}_{\text{MVDR}}(\omega)] = 1, \quad (76)$$

$$v_{\text{sd}}[\mathbf{h}_{\text{MVDR}}(\omega)] = v_{\text{sd}}[\mathbf{h}_{\text{SDIRC}}(\omega)] = 0. \quad (77)$$

Now using the fact that $\alpha(\omega) = H_{\text{PW}}(\omega; \gamma)$, we directly find

$$\xi_{\text{dsc}}[\mathbf{h}_{\text{PMWF}}(\omega; \gamma)] = \frac{[\gamma + \rho(\omega)]^2}{\rho^2(\omega)}. \quad (78)$$

$$v_{\text{sd}}[\mathbf{h}_{\text{PMWF}}(\omega; \gamma)] = \frac{\gamma^2}{[\gamma + \rho(\omega)]^2}. \quad (79)$$

For a given situation and hence $\rho(\omega)$, we find using (79) and (70) that the speech-distortion index increases monotonically with increasing γ .

D. Array Gain

The role of the beamformer is to produce a signal whose SINR is higher than that which was received at the reference microphone. To that end, the array gain is defined as the ratio of the output SINR (after beamforming) over the input SINR (at the reference microphone) [24]. It is therefore equal to the SINR improvement that is often presented. This leads to the following definition:

$$\begin{aligned} \mathcal{A}[\mathbf{h}(\omega)] &= \frac{\text{oSINR}[\mathbf{h}(\omega)]}{\text{iSINR}(\omega)} \\ &= \frac{|\mathbf{h}^H(\omega)\mathbf{a}_d(\omega)|^2 [\phi_{u_1}(\omega) + \phi_{v_1}(\omega)]}{\phi_{u_1}(\omega) |\mathbf{h}^H(\omega)\mathbf{a}_u(\omega)|^2 + \mathbf{h}^H(\omega)\Phi_{\mathbf{v}}(\omega)\mathbf{h}(\omega)} \\ &= \frac{|\mathbf{h}^H(\omega)\mathbf{a}_d(\omega)|^2 [1 + \text{iNIR}(\omega)]}{|\mathbf{h}^H(\omega)\mathbf{a}_u(\omega)|^2 + \phi_{u_1}^{-1}(\omega)\mathbf{h}^H(\omega)\Phi_{\mathbf{v}}(\omega)\mathbf{h}(\omega)}. \end{aligned} \quad (80)$$

By making the appropriate substitutions, one can derive the following relationship between the array gain, noise-rejection factor, and desired-signal-cancellation factor:

$$\mathcal{A}[\mathbf{h}(\omega)] = \frac{\xi_{\text{nr}}[\mathbf{h}(\omega)]}{\xi_{\text{dsc}}[\mathbf{h}(\omega)]}. \quad (81)$$

Hence, when no distortion occurs (i.e., $\xi_{\text{dsc}}[\mathbf{h}(\omega)] = 1$), the array gain coincides with the noise-reduction factor.

By substituting (43) and (29) in (80), we obtain the array gain of the SDIRC beamformer:³

$$\mathcal{A}[\mathbf{h}_{\text{SDIRC}}] = \frac{(\phi_{u_1} + \phi_{v_1})|\alpha|^2}{m_1|\alpha|^2 + [\phi_{u_1} + m_2]|\beta|^2 + 2\text{Re}\{m_{12}\alpha\beta\}}. \quad (82)$$

Clearly, the array gain depends on $\alpha(\omega)$ and $\beta(\omega)$.

VI. PERFORMANCE EVALUATION

The performance of the SDIRC beamformer in two environments, viz., anechoic and reverberant, is evaluated in this section. The tradeoff between speech distortion and interference-plus-noise reduction on the one hand, and undesired signal and ambient reductions on the other hand will be demonstrated. For comparison, we will show the performance obtained by the LCMV, MVDR, and PMWF.

The results of our simulations are presented in terms of the performance measures described in Section V. Most results are presented in terms of the fullband version of these performance measures, which are defined in a similar way the fullband input SINR is defined in (44). Because the filters are applied on a frame-by-frame basis, as will be explained shortly, the performance measures are evaluated per frame and subsequently averaged over all frames. The SNRs, SIRs, SINRs, and NIRs are averaged in the logarithm domain while other performance measures are averaged in the linear domain. In reverberant environments, the performance depends on the absolute spatial location of the sources and microphones in the enclosed space. To alleviate this spatial dependency, Monte Carlo simulations were conducted by rotating and translating the source-array geometry inside the room. The final results were obtained by averaging the results of 50 Monte Carlo trials.

It is important to note that the expressions of the frequency-domain filters involve divisions by some quantities that might approach zero due to speech absence or common-zeros between the channels. Therefore, all quantities in the denominators are kept above a certain threshold that was experimentally set to 10^{-6} .

A. Experimental Setup and Implementation

A uniform linear array (ULA) was used with 4 microphones and an inter-microphone distance of 2.5 cm. The distance between the floor and the microphone array is 1.6 m.

In the first scenario, it is assumed that the desired and undesired sources are located in the far-field with no reverberation. The corresponding steering vectors are given by $\mathbf{g}_d(\omega) = \begin{bmatrix} 1 & e^{-j\omega\delta c^{-1}\cos(\theta_d)} & \dots & e^{-j\omega(M-1)\delta c^{-1}\cos(\theta_d)} \end{bmatrix}^T$ and $\mathbf{g}_u(\omega) = \begin{bmatrix} 1 & e^{-j\omega\delta c^{-1}\cos(\theta_u)} & \dots & e^{-j\omega(M-1)\delta c^{-1}\cos(\theta_u)} \end{bmatrix}^T$, where $\delta = 0.025$ is equal to the distance between the microphones of the ULA, $c = 343 \text{ ms}^{-1}$ is the speed of sound, and θ_d and θ_u determine the azimuth of the desired and undesired sources as shown in Fig. 1. Here the attenuation of the propagating sounds, which depends on the source-receiver distance, is

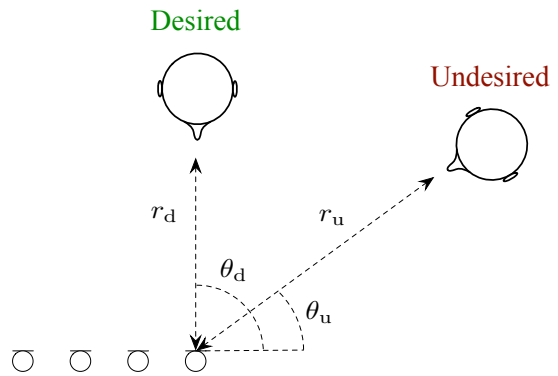


Fig. 1. Schematic drawing showing the array configuration and the location of the desired and undesired sources.

taken into account in the power of the sources rather than the steering vector. From (6) and (7) it follows that $\mathbf{a}_d(\omega) = \mathbf{g}_d(\omega)$ and $\mathbf{a}_u(\omega) = \mathbf{g}_u(\omega)$, respectively.

In the second scenario, we consider a reverberant environment. The acoustic impulse responses (each with a duration of 500 ms) were generated using an efficient implementation [25] of the image-method [26] with some necessary modifications that ensure correct inter-microphone phase delays by using fractional decays [27]. The room size is $5 \times 4 \times 6$ m (length \times width \times height) and the reflection coefficients of the walls, ceiling, and floor are set to achieve a reverberation time $T_{60} = 400$ ms measured using the backward integration method [28]. The distance between the first microphone and the desired and undesired sources are denoted respectively by r_d and r_u as shown in Fig. 1. The vectors $\mathbf{a}_d(\omega)$ and $\mathbf{a}_u(\omega)$ are computed per frequency using

$$\mathbf{a}_d(\omega) = \frac{E\{\mathbf{d}(\omega)D_1^*(\omega)\}}{\phi_{d_1}(\omega)} \quad (83)$$

and

$$\mathbf{a}_u(\omega) = \frac{E\{\mathbf{u}(\omega)U_1^*(\omega)\}}{\phi_{u_1}(\omega)}, \quad (84)$$

respectively.

The desired source consists of 10 minutes of male and female speech composed of data from the APLAWD speech database [29] sampled at 8 kHz and is located at ($\theta_d = 50$ degrees, $r_d = 1$ m). The undesired source, located at ($\theta_u = 130$ degrees, $r_u = 1.5$ m), consists of an USASI noise sequence that exhibits the same spectral properties as speech. The ambient noise consists of spatially homogeneous and spatially and temporally white Gaussian noise with a long-term input SNR of 20 dB, evaluated at the first microphone. The microphone signals are generated using (1), (2), and (3).

The microphone signals are processed in the discrete Fourier transform domain using the overlap-save [30] method. Specifically, the signals are divided into 50% overlapping frames of duration 512 ms. Once the microphone signals are filtered in the frequency domain, they are transformed into the time domain and only those samples are kept that correspond to a linear convolution [30]. Finally, a finite-impulse-response

(FIR) constraint was applied to the filter in the time domain [11], [31]. Because small estimation errors can result in pre-echoes (i.e., echoes that arrive before the arrival of the direct sound), an asymmetric FIR constraint was applied that was determined experimentally. Specifically, the length of the anticausal part was set to 64 ms and the length of the causal part was set to 128 ms.

In this study, we put aside the problem of estimating the statistics of the sources and ambient noise because we are interested in assessing the performance of various beamformers and the different tradeoffs. The PSD of the source signals $\Phi_d(k)$ and $\Phi_u(k)$, where k denotes the discrete frequency bin, are estimated recursively using the signal vectors $\mathbf{d}(k)$ and $\mathbf{u}(k)$ with a forgetting factor of 0.98. The ambient noise PSD matrix $\Phi_v(k)$ is estimated from the signal vector $\mathbf{v}(k)$ using the Welch's modified periodogram [32].

For the SDIRC beamformer we have two design parameters, viz., $\alpha(\omega)$ and $\beta(\omega)$. These parameters can be chosen in many ways, some of which perceptually motivated. Rather than evaluating the beamformer over the entire parameter space, we propose to investigate another parameter space such that it includes the MVDR, LCMV, and PMWF. Specifically, we investigate the influence of γ and therefore define $\alpha(\omega)$ as follows:

$$\alpha(\omega; \gamma) = H_{PW}(\omega; \gamma). \quad (85)$$

Furthermore, we define $\beta(\omega)$ as follows:

$$\beta(\omega; \beta') = -\beta' \frac{m_{12}(\omega)}{\phi_{u_1}(\omega) + m_2(\omega)}, \quad (86)$$

such that we can investigate the behavior of the SDIRC beamformer as a function of β' . For $(\gamma = 0, \beta' = 0)$, we then obtain the LCMV and for $(\gamma = 0, \beta' = 1)$, we then obtain the MVDR. For $(\gamma \geq 0, \beta' = 1)$, we obtain solutions of the PMWF. Previously unexplored solutions that allow a tradeoff between speech distortion and interference-plus-noise reduction on the one hand, and undesired signal and ambient noise reductions on the other hand are given by $(\gamma \geq 0, 0 < \beta' < 1)$.

B. Anechoic Environment

Let us first investigate the beamformer's performance in terms of the SINR improvement (i.e., array gain), NIR improvement and speech-distortion index as a function of γ and β' in an anechoic environment with a SIR of 10 dB and -5 dB.

The results obtained with an SIR of 10 dB are shown in Fig. 2. The SINR improvement shown in Fig. 2(a) demonstrates that the SINR increases monotonically with increasing β' and/or γ . As mentioned in Section V, the SINR of the LCMV ($\gamma = 0, \beta' = 0$) is smaller than that of the MVDR ($\gamma = 0, \beta' = 1$). The effect of β' becomes evident when analyzing the NIR shown in Fig. 2(b) (note that the axes have been changed to improve the visualization). While γ has almost no influence on the NIR, the NIR decreases monotonically with increasing β' . When $\beta' = 0$, the beamformer is designed to cancel the undesired source completely and in theory the NIR is infinity. The speech-distortion index is shown in Fig. 2(c) and demonstrates that only γ has an

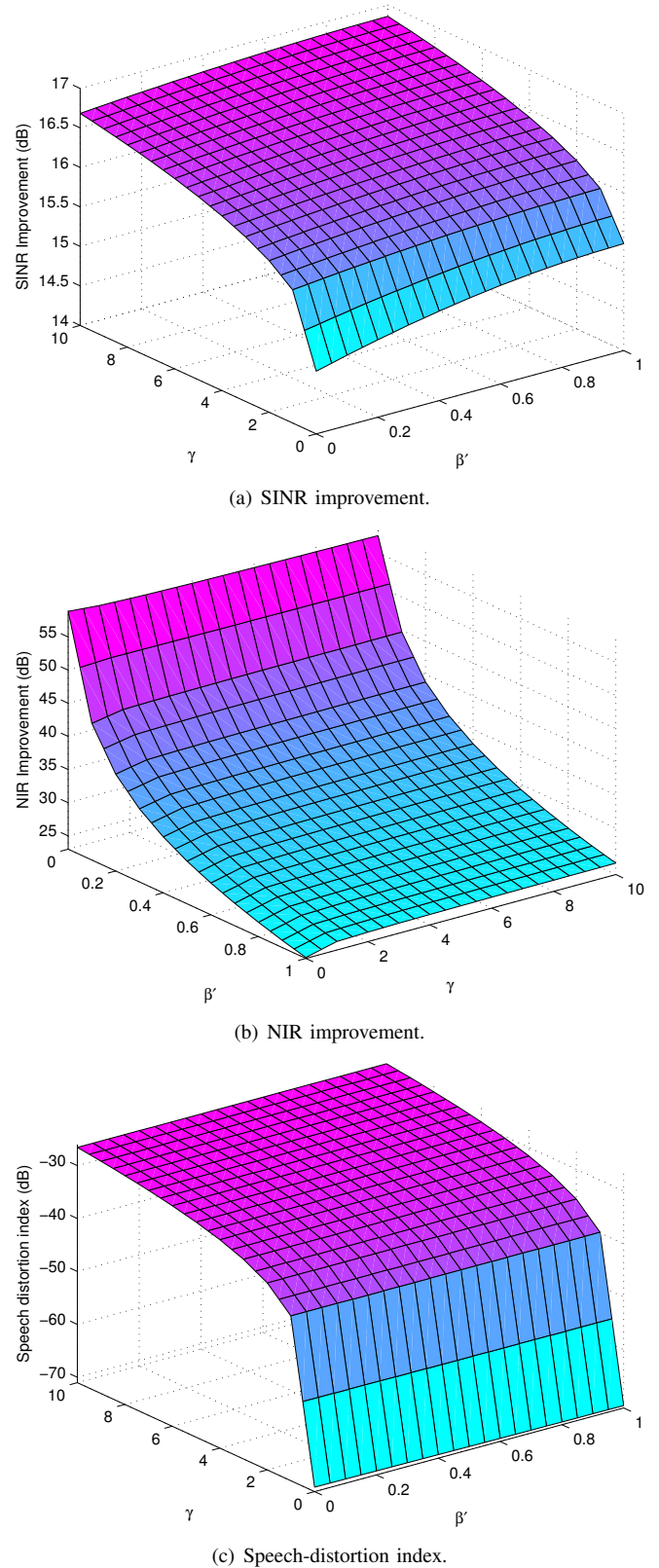


Fig. 2. Performance measures as a function of β' and γ for $M = 4$, $T_{60} = 0$ (i.e., anechoic), $\text{SNR} = 20$ dB and $\text{SIR} = 10$ dB.

influence on the speech distortion. As discussed in Section V the speech distortion increases monotonically with increasing

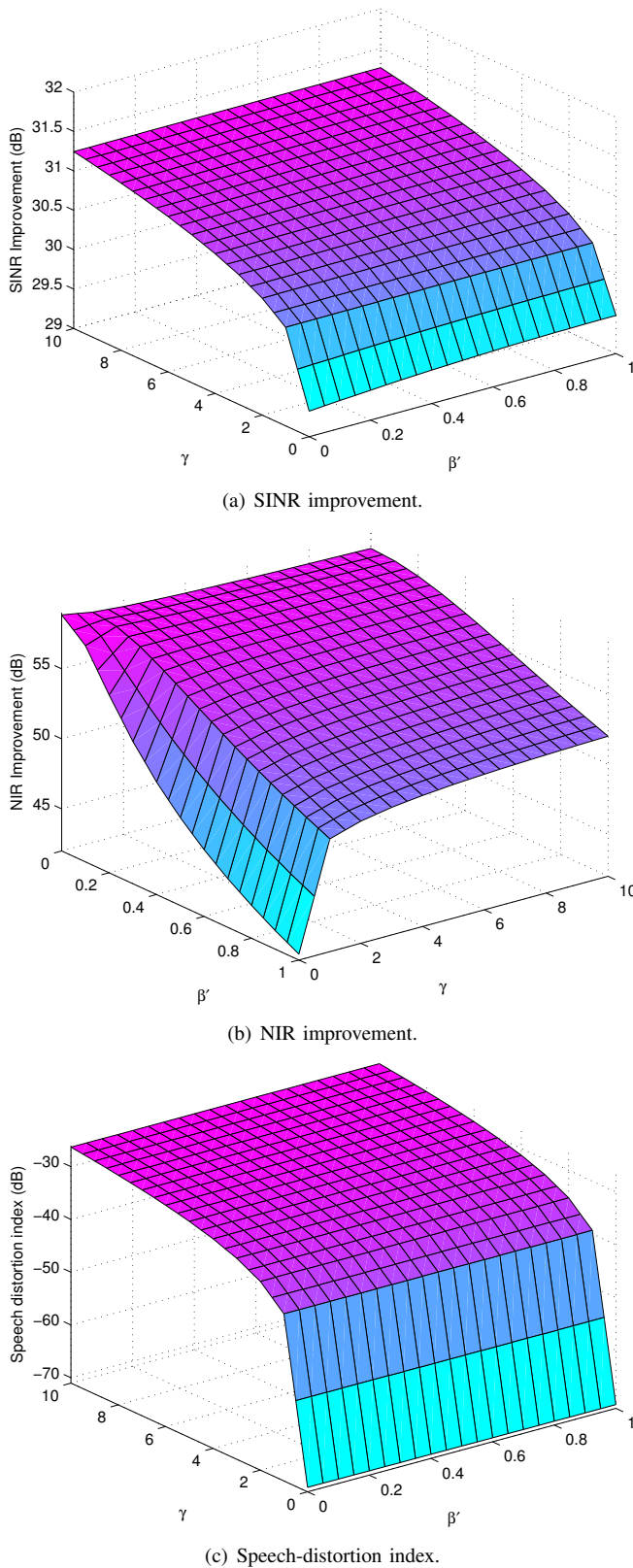


Fig. 3. Performance measures as a function of β' and γ for $M = 4$, $T_{60} = 0$ (i.e., anechoic), $\text{SNR} = 20$ dB and $\text{SIR} = -5$ dB.

γ . These results demonstrate that we can control independently the undesired signal and ambient noise reductions and the

speech distortion and interference-plus-noise reduction.

The results obtained with an SIR of -5 dB are shown in Fig. 3. In this scenario, the power of the undesired signal is much larger than the power of the ambient noise. While the results for the SINR improvement and the speech-distortion index have not changed much, we see that the influence of β' on the NIR improvement changed. While the trend remains the same, we observe that the range in which the NIR varies as a function of β' has reduced. As mentioned in Section V, the MVDR ($\gamma = 0, \beta' = 1$) maximizes the output SINR, and in case the power of the undesired source is much larger than the power of ambient noise the performance of the LCMV ($\gamma = 0, \beta' = 0$) and MVDR ($\gamma = 0, \beta' = 1$) are comparable.

C. Simulated Reverberant Environment

Now we investigate the beamformer's performance in terms of the SINR improvement, NIR improvement and speech-distortion index as a function of γ and β' in a simulated reverberant environment with $T_{60} = 400$ ms and a SIR of 10 dB. The obtained results are shown in Fig. 4. The behavior of the SINR improvement, NIR improvement and speech-distortion index is very similar to the beamformer's behavior observed in the anechoic environment discussed in the previous subsection. For the considered scenario the range in which we can vary the NIR using β' is about 8 dB. In general, we found that the amount of speech distortion obtained in the reverberant environment is higher than that obtained in an anechoic environment. This effect can be attributed to the finite filter length used in our simulations.

D. Real Reverberant Environment

Finally, we studied the performance of the SDIRC beamformer in a real reverberant environment. The room impulse responses were measured in the varechoic chamber at Bell Labs [33]. The chamber is a rectangular room ($6.7 \text{ m} \times 6.1 \text{ m} \times 2.9 \text{ m}$) with 368 electronically controlled panels that vary the acoustic absorption of the walls, floor, and ceiling [34]. Therefore, the level of room reverberation is well controlled by the percentage of open panels. In this experiment, 89% of the panels are open, which leads to a reverberation time of approximately 130 ms. We consider a scenario with one desired source that consists of a female speaker and one undesired source ($\text{SIR} = 5$ dB) that consists of a male speaker – both source signals are taken from the APLAWD speech database [29]. The desired source is active between 0 and 2 s and between 3.5 and 5 s, and the undesired source is active between 2 and 5 s. Therefore, both the desired and undesired sources are active between 3.5 and 5 s. The ambient noise is spatially homogeneous and spatially and temporally white Gaussian ($\text{SNR} = 15$ dB). The source and interference were positioned at approximately the same location with respect to the array as in our previous experiments and correspond to locations v12 and v14 in [33].

The spectrograms and waveforms of the desired source signal and the first microphone signal are depicted in Fig. 5(a) and (b), respectively. The results obtained after processing the

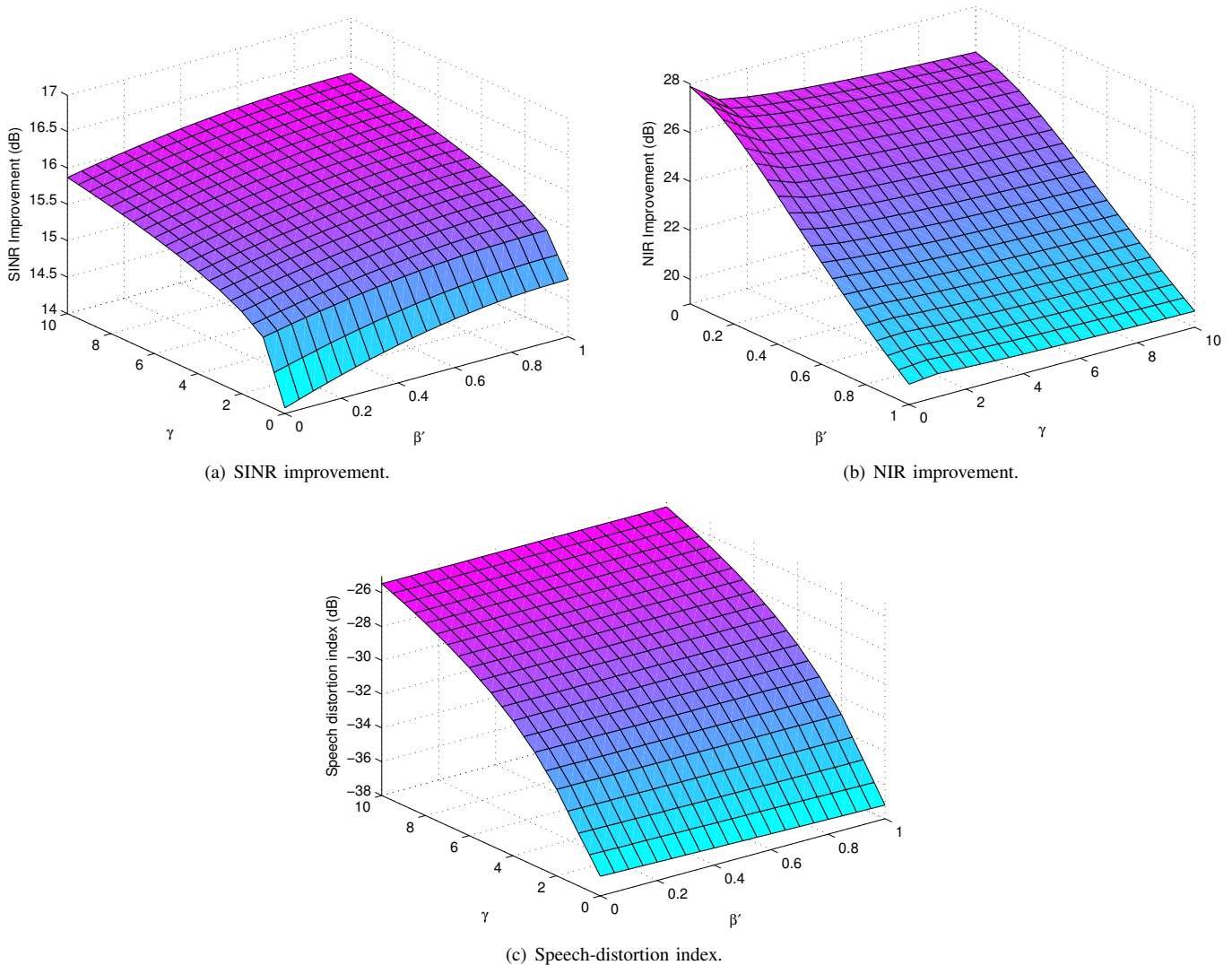


Fig. 4. Performance measures as a function of β' and γ for $M = 4$, $T_{60} = 400$ ms, $\text{SNR} = 20$ dB and $\text{SIR} = 10$ dB.

received signals using the proposed SDIRC beamformer with $\gamma = 0$ and $\beta' \in \{0, 0.5, 1\}$ are shown in Fig. 5(c)-(e) and the result obtained with $\gamma = 1$ and $\beta' = 1$ is shown in Fig. 5(f).

The output signal shown in Fig. 5(c) corresponds to the output of an LCMV beamformer. While the undesired speech signal (between 2 and 3.25 s) is almost canceled we can see that the ambient noise in the low frequencies has increased significantly. Results of an informal listening test, conducted using 5 subjects and a Sennheiser HD650 headphone, showed that undesired signal was inaudible. The residual noise at the output of the LCMV beamformer is very stationary (as it mainly consists of residual ambient noise that is stationary) but the power at low frequencies is significant. Reducing the residual noise power to an acceptable level using $\alpha(\omega)$ would cause significant speech distortion. The output signal shown in Fig. 5(d) corresponds to the output of the proposed SDIRC beamformer with $(\gamma = 0, \beta' = 0.5)$. We notice that the ambient noise has not been amplified as much compared to the LCMV, and we notice a small amount of residual undesired signal which was hardly audible. In Fig. 5(e), the

output signal is shown that corresponds to the output of an MVDR beamformer. While this signal has the largest output SINR with low speech distortion we can clearly see the residual undesired signal. The results of the informal listening test showed that the undesired signal was clearly audible. Further reducing the interference-plus-noise power is possible at the cost of increased speech distortion. Unfortunately, the residual interference-plus-noise is highly nonstationary and further reducing it using $\alpha(\omega)$ might be difficult as we need to estimate the PSD of the residual noise first. The interference-plus-noise obtained using $(\gamma = 0, \beta' = 0.5)$ is fairly stationary and less strong compared to that observed at the output of the LCMV. Using $\alpha(\omega)$ we can further reduce the residual interference-plus-noise power to an acceptable level without introducing too much speech distortion. As an example the output signal for $\gamma = 1, \beta' = 0.5$ is shown in Fig. 5(f). We observe that the ambient noise at very low and very high frequencies is significantly reduced compared to all other beamformers. Compared to the result obtained for $\gamma = 0$ and $\beta' = 0.5$, we observe that the interference-plus-noise power

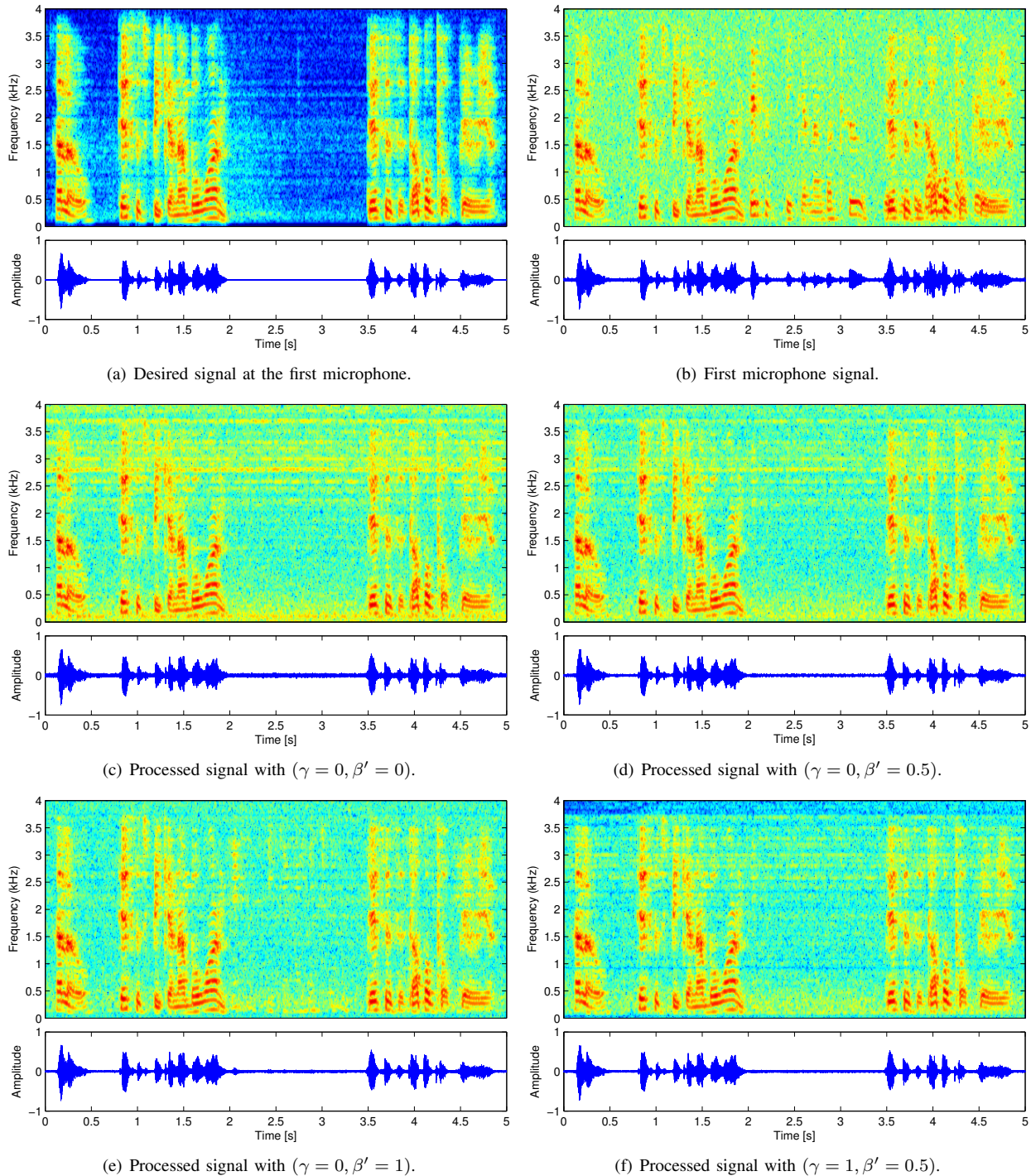


Fig. 5. Spectrogram and waveform of the signal received by the first microphone, the SDIRC beamformer with $\beta' \in \{0, 0.5, 1\}$ and $\gamma = \{0, 1\}$ for $M = 4$, $T_{60} = 130$ ms, SNR = 15 dB and SIR = 5 dB.

is also reduced between 0.1 and 3.75 kHz. The results of the informal listening test confirms that the ambient noise is significantly reduced and the undesired source remains inaudible.

We now analyze the speech-distortion index when both the desired and undesired source are active (i.e., between 3.5 and 5 s). In Fig. 6(a), the speech-distortion index is depicted as a function of time (top figure) and frequency (bottom figure).

Across time and frequency we see that the speech-distortion index is almost independent of β' (only a slight increase in the the speech-distortion index is observed when β' decreases from one to zero). Finally, we analyze the NIR improvement when both the undesired source and the ambient noise are present (i.e., between 2 and 5 s). In Fig. 6(b), the NIR improvement is depicted as a function of time (top figure) and frequency (bottom figure). In general we observe that NIR

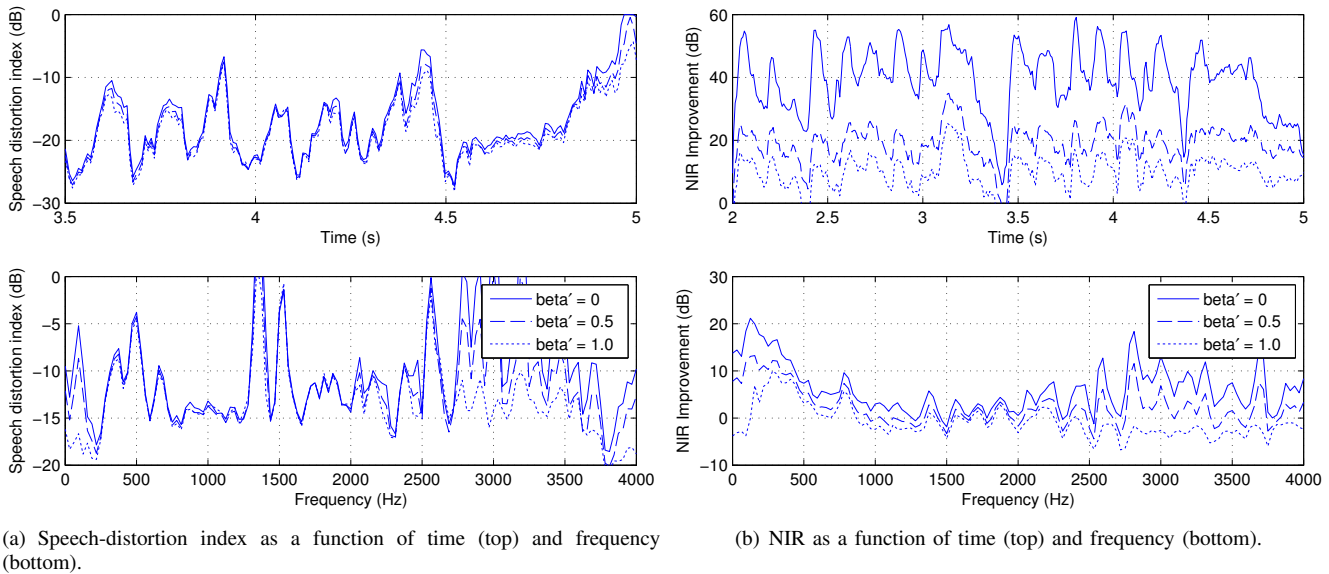


Fig. 6. NIR and speech-distortion index as a function of time and frequency for the SDIRC beamformer with $\beta' \in \{0, 0.5, 1\}$ and $\gamma = 0$ for $M = 4$, $T_{60} = 130$ ms, SNR = 15 dB and SIR = 5 dB.

improvement increases when β' decreases. Hence, using β' we can control the undesired signal and ambient noise reductions without significantly affecting the speech distortion.

VII. CONCLUSIONS

A speech distortion and interference rejection constraint beamformer was proposed that generalizes existing beamformers and allows a tradeoff between speech distortion and interference-plus-noise reduction on the one hand, and undesired signal and ambient noise reductions on the other hand. The conducted performance evaluation has demonstrated that this tradeoff can be achieved. Furthermore, relations between the proposed SDIRC beamformer and existing beamformers have been derived and frequently used performance measures were deduced and analyzed. In addition, it was shown that these relations provide an alternative way to compute performance measures of existing beamformers. The tradeoffs facilitated by the proposed beamformer can be used in the development of effective speech acquisition systems and in the design of perceptually motivated beamformers.

VIII. ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments and helpful suggestions.

REFERENCES

- [1] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.
- [2] S. Gannot and I. Cohen, "Adaptive beamforming and postfiltering," in *Springer Handbook of Speech Process.*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Berlin, Germany: Springer-Verlag, 2008, ch. 47.
- [3] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.
- [4] M. Er and A. Cantoni, "Derivative constraints for broad-band element space antenna array processors," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, no. 6, pp. 1378–1393, Dec. 1983.
- [5] M. Souden, J. Benesty, and S. Affes, "A study of the LCMV and MVDR noise reduction filters," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4925–4935, Sep. 2010.
- [6] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," *Signal Process.*, vol. 84, no. 12, pp. 2367–2387, Dec. 2004.
- [7] M. Souden, J. Benesty, and S. Affes, "On optimal frequency-domain multichannel linear filtering for noise reduction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 2, pp. 260–276, Feb. 2010.
- [8] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.
- [9] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol. AP-30, no. 1, pp. 27–34, Jan. 1982.
- [10] K. M. Buckley, "Broad-band beamforming and the generalized sidelobe canceller," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 5, pp. 1322–1323, Oct. 1986.
- [11] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.
- [12] E. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, "New insights into the MVDR beamformer in room acoustics," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, pp. 158–170, Jan. 2010.
- [13] J. Bitzer, K. Simmer, and K.-D. Kammeyer, "Theoretical noise reduction limits of the generalized sidelobe canceller for speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 1999, vol. 5, pp. 2965–2968.
- [14] E. A. P. Habets, J. Benesty, S. Gannot, P. A. Naylor, and I. Cohen, "On the application of the LCMV beamformer to speech enhancement," in *Proc. IEEE Workshop on Appl. Signal Process. Audio and Acoust.*, New York, USA, Oct. 2009, pp. 141–144.
- [15] I. Cohen, "Relative transfer function identification using speech signals," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 451–459, Sep. 2004.
- [16] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 4, pp. 251–266, Jul. 1995.
- [17] E. Habets, J. Benesty, S. Gannot, and I. Cohen, "The MVDR beamformer for speech enhancement," in *Speech Processing in Modern Communication: Challenges and Perspectives*, ser. Springer Topics in Signal Processing, I. Cohen, J. Benesty, and S. Gannot, Eds. Berlin, Germany: Springer, Jan. 2010, ch. 9, pp. 225–254.
- [18] H. L. van Trees, *Optimum Array Processing*, ser. Detection, Estimation and Modulation Theory. New York: Wiley, 2002.
- [19] D. H. Brandwood, "A complex gradient operator and its application in

adaptive array theory," *Proc. IEE*, vol. 130, no. 1, pt. F and H, pp. 11–16, Feb. 1983.

- [20] J. Benesty, S. Makino, and J. Chen, Eds., *Speech Enhancement*. New York: Springer, 2005.
- [21] W. Herbordt, W. Kellermann, and S. Nakamura, "Joint optimization of LCMV beamforming and acoustic echo cancellation," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Vienna, Austria, Sep. 2004, pp. 2003–2006.
- [22] J. Benesty, J. Chen, and Y. Huang, "On the importance of the Pearson correlation coefficient in noise reduction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 4, pp. 757–765, May 2008.
- [23] J. Chen, J. Benesty, and Y. Huang, "Time delay estimation in room acoustic environments: An overview," *EURASIP J. Appl. Signal Process.*, Special issue on advances in multimicrophone speech processing, pp. 1–19, 2006.
- [24] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [25] E. A. P. Habets, "Room impulse response generator," Technische Universiteit Eindhoven, 2006, Tech. Rep..
- [26] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.
- [27] P. M. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *J. Acoust. Soc. Amer.*, vol. 80, no. 5, pp. 1527–1529, Nov. 1986.
- [28] M. R. Schroeder, "Integrated-impulse method measuring sound decay without using impulses," *J. Acoust. Soc. Amer.*, vol. 66, no. 2, pp. 497–500, 1979.
- [29] G. Lindsey, A. Breen, and S. Nevard, "SPAR's archivable actual-word databases," University College London, Technical Report, Jun. 1987.
- [30] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, no. 1, pp. 14–37, Jan. 1992.
- [31] S. Affès and Y. Grenier, "A signal subspace tracking algorithm for microphone array processing of speech," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 5, pp. 425–437, 1997.
- [32] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck, *Discrete-Time Signal Processing*, ser. signal processing., 2nd ed., Upper Saddle River, NJ: Prentice-Hall, 1999.
- [33] A. Härmä, "Acoustic measurement data from the varechoic chamber," Agere Systems, Nov. 2001, Tech. Rep..
- [34] W. Ward, G. Elko, R.A. Kubli, and W. McDougald, "The new varechoic chamber at AT&T Bell Labs," in *Proc. Wallace Clement Sabine Centennial Symp.*, 1994, pp. 343–346.



Emanuel A.P. Habets (S'02-M'07-SM'11) received the B.Sc. degree in electrical engineering from the Hogeschool Limburg, The Netherlands, in 1999, and the M.Sc. and Ph.D. degrees in electrical engineering from the Technische Universiteit Eindhoven, The Netherlands, in 2002 and 2007, respectively.

From March 2007 until February 2009, he was a Postdoctoral Fellow at the Technion - Israel Institute of Technology and at the Bar-Ilan University in Ramat-Gan, Israel. From February 2009 until November 2010, he was a Research Fellow in the Communication and Signal Processing Group at Imperial College London, London, U.K. Since November 2010, he has been an Associate Professor at the International Audio Laboratories Erlangen (a joint institution of the University of Erlangen and Fraunhofer IIS) and a Chief Scientist Spatial Audio Processing at Fraunhofer IIS, Germany. His research interests are in signal processing for acoustical applications, more specifically microphone array processing for noise reduction, dereverberation, and source localization.

Dr. Habets was a member of the organization committee of the Ninth International Workshop on Acoustic Echo and Noise Control (IWAENC), Eindhoven, in 2005. He is a member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing (2011-2013).



Jacob Benesty was born in 1963. He received the M.Sc. degree in microwaves from Pierre & Marie Curie University, Paris, France, in 1987, and a Ph.D. degree in control and signal processing from Orsay University, Orsay, France, in April 1991.

During the Ph.D. degree (from November 1989 to April 1991), he worked on adaptive filters and fast algorithms at the Centre National d'Etudes des Telecommunications (CNET), Paris. From January 1994 to July 1995, he worked at Telecom Paris University on multichannel adaptive filters and acoustic echo cancellation. From October 1995 to May 2003, he was first a Consultant and then a Member of the Technical Staff at Bell Laboratories, Murray Hill, NJ. In May 2003, he joined the University of Quebec, INRS-EMT, in Montreal, QC, Canada, as a Professor. His research interests are in signal processing, acoustic signal processing, and multimedia communications. He is the inventor of many important technologies. In particular, he was the lead researcher at Bell Labs who conceived and designed the world-first real-time hands-free full-duplex stereophonic teleconferencing system. Also, he and T. Gaensler conceived and designed the world-first PC-based multi-party hands-free full-duplex stereo conferencing system over IP networks. He is the editor of the book series Springer Topics in Signal Processing. He has coauthored and coedited/coauthored many books in the area of acoustic signal processing. He is also the lead editor-in-chief of the reference Springer Handbook of Speech Processing (Springer-Verlag, 2007).

Dr. Benesty was the cochair of the 1999 International Workshop on Acoustic Echo and Noise Control and the general cochair of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. He was a member of the IEEE Signal Processing Society Technical Committee on Audio and Electroacoustics and a member of the editorial board of the EURASIP Journal on Applied Signal Processing. He is the recipient, with Morgan and Sondhi, of the IEEE Signal Processing Society 2001 Best Paper Award. He is the recipient, with Chen, Huang, and Doclo, of the IEEE Signal Processing Society 2008 Best Paper Award. He is also the coauthor of a paper for which Y. Huang received the IEEE Signal Processing Society 2002 Young Author Best Paper Award. In 2010, he received the "Gheorghe Cartianu Award" from the Romanian Academy.



Patrick A. Naylor (M'89-SM'07) received the B.Eng. degree in electronic and electrical engineering from the University of Sheffield, Sheffield, U.K., in 1986 and the Ph.D. degree from Imperial College, London, U.K., in 1990.

Since 1990, he has been a Member of Academic Staff in the Department of Electrical and Electronic Engineering, Imperial College London, where he is also Director of Postgraduate Studies. His research interests are in the areas of speech, audio and acoustic signal processing. He has worked in particular on adaptive signal processing for dereverberation, blind multichannel system identification and equalization, acoustic echo control, speaker identification, single and multi-channel speech enhancement and speech production modeling with particular focus on the analysis of the voice source signal. In addition to his academic research, he enjoys several fruitful links with industry in the U.K., USA and in mainland Europe.

Dr. Naylor is an Associate Editor of the IEEE Transactions on Audio Speech and Language Processing and an associate member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing.