Dual-Source Transfer-Function Generalized Sidelobe Canceller

Gal Reuven, Sharon Gannot, *Senior Member, IEEE*, and Israel Cohen, *Senior Member, IEEE*

***Abstract*—Full-duplex hands-free man/machine interface often suffers from directional nonstationary interference, such as a competing speaker, as well as stationary interferences which may comprise both directional and nondirectional signals. The transfer-function generalized sidelobe canceller (TF-GSC) exploits the nonstationarity of the speech signal to enhance it when the undesired interfering signals are stationary. Unfortunately, the assumptions leading to the derivation of the TF-GSC are violated when a nonstationary interference is present. In this paper, we propose an adaptive beamformer, based on the TF-GSC, that is suitable for cancelling nonstationary interferences in noisy reverberant environments. We modify two of the TF-GSC components to enable suppression of the nonstationary undesired signal. A modified fixed beamformer (FBF) is designed to block the nonstationary interfering signal while maintaining the desired speech signal. A modified blocking matrix (BM) is designed to block both the desired signal and the nonstationary interference. We introduce a novel method for updating the blocking matrix in *double talk* scenarios, which exploits the nonstationarity of both the desired and interfering speech signals. Experimental results demonstrate the performance of the proposed algorithm in noisy and reverberant environments and show its superiority over the original TF-GSC.**

***Index Terms*—Array signal processing, interference cancellation, nonstationarity, speech enhancement.**

I. Introduction

In MANY practical environments, as in a conference call scenario, a desired speech signal, received by a microphone array, is contaminated by both nonstationary interfering signal (such as a competing speech), and by stationary noise. Furthermore, the received signals are often subject to distortion imposed by the room impulse response (RIR) of the acoustic environment. Beamforming is the most commonly used approach for extracting a desired source out of spatially distinct sources. The array beampattern can generally be designed to have a specified response. This can be done by properly setting the values of the multichannel filters' weights. However, the application of data independent design methods is very limited in dynamic acoustical environments. Statistically, optimal beamformers are designed based on the statistical properties of the desired and interference signals. In general, they aim at enhancing the desired signal, while rejecting the interference signals. Several criteria can be applied in the design of the beamformer, e.g., maximum signal-to-noise ratio (SNR), minimum mean-squared error (MMSE), and linearly constrained minimum variance (LCMV). A summary of several design criteria can be found in [1] and [2].

Frost [3] proposed the LCMV beamformer as a method for minimizing the array output power under a look-direction constraint. Several researchers (e.g., Er and Cantoni [4]) have proposed modifications to the LCMV for dealing with multiple linear constraints. Their work was motivated by the desire to apply further control to the array/beamformer beampattern, beyond that of steer-direction gain constraints. Hence, the LCMV can be applied for constructing a beampattern satisfying certain constraints for a set of directions, while minimizing the array response in all other directions. The LCMV is efficiently implemented in a generalized sidelobe canceller (GSC) [5] structure, which decouples the constraints and the minimization. Several contributions proved the equivalence of the LCMV and GSC structures [6]–[8].

The GSC has found numerous applications in the field of speech enhancement (e.g., [9]–[12]). In most speech enhancement applications, the beamformer is constrained to produce a dominant response towards the desired speech source location, while minimizing the response in all other directions. However, in reverberant environments a single direction of arrival cannot be determined since the desired signal and its reflections impinge on the array from several directions. This problem may be alleviated by using an acoustic transfer function (ATF) rather than just a simple delay for modeling the propagation of the speech signal in a reverberant room.

Affes and Grenier [9] proposed a subspace method for estimating and tracking ATFs in a reverberant environment. In [13], they further proposed a GSC structure, for situations where two speech signals are active simultaneously (referred to as *double talk* situation), encountered in the context of acoustic echo canceler (AEC). The GSC includes a distortionless FBF which is constrained to cancel the echo, and a BM which is constrained to block both the desired signal and echo signal. However, the estimation of the various blocks in the proposed structure necessitates the use of the echo signal, which is unavailable in many applications.

Gannot *et al.* [14] proposed to use the relative transfer function (RTF) between microphones in response to a desired source signal rather than the ATF themselves. A method for estimating the RTF, based on the background noise stationarity and the speech nonstationarity, is derived. Compared with the conventional GSC, the resulting TF-GSC is of practical utility when enhancing a speech signal deteriorated by *stationary* interfering

Manuscript received August 14, 2007; revised November 28, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hiroshi Sawada.

G. Reuven was with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel. He is now with Dune Medical Devices, Ltd., Caesarea 38900, Israel (e-mail: galrv@techunix.technion.ac.il).

S. Gannot is with the School of Engineering, Bar-Ilan University, Ramat-Gan 52900, Israel (e-mail: gannot@eng.biu.ac.il).

I. Cohen is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: icohen@ee.technion.ac.il).

Digital Object Identifier 10.1109/TASL.2008.917389

signals in a reverberant enclosure. However, in the presence of an additional *nonstationary* interference, the TF-GSC cannot distinguish between the desired signal and the interfering signal, which renders the proposed structure useless.

Benesty *et al.* [15] address beamforming structures for multiple input signals. In their contribution, derived in the time-domain, the microphone array is treated as a multiple-input multiple-output (MIMO) system. Among the beamforming structures considered in that paper, the LCMV with multiple constraints is of special relevance to our work. Performance bounds for the obtained speech distortion and noise reduction are calculated, and equivalence between several popular beamforming structures is proved. However, in the experimental study, it is assumed that the filters relating the sources and the microphones are *a priori* known, or alternatively, that the sources are not active simultaneously. This assumption cannot be met in our scenario, in which the noise signal is always active, and *double talk* situations can occasionally occur.

In this paper, we present[1] an adaptive beamformer, based on the TF-GSC [14]. The new structure is aimed at maintaining the desired speech signal while cancelling two interference signals: a stationary (noise) signal and nonstationary (competing speech) signal. The proposed beamformer is designed for reverberant environment application. In Section II, we formulate the dual interference problem for the general ATF case. In Section III, we derive a closed-form linearly constraint beamformer specifically designed for suppressing the undesired interference signals. In Section IV, we derive the equivalent GSC structure. We show that a modification of two of the TF-GSC components, namely the FBF and the BM, allows the beamformer to suppress the nonstationary interference. The modified fixed beamformer, denoted matched beamformer (MBF), is designed to block the nonstationary interference while maintaining the desired speech signal. Beampattern evaluation emphasizes the differences between the MBF and the conventional FBF. The modified BM blocks both the desired signal and the nonstationary interference. As in the original TF-GSC, the adaptive noise canceller (ANC) employs the reference signals generated by the BM to cancel the residual stationary noise at the MBF output. The resulting structure is named a dual transfer-function generalized sidelobe canceller (DTF-GSC). In Section V, we address the problem of RTF estimation. Specifically, a method for updating the BM in *double talk* situations is proposed. This method extends the RTF estimation method, applied in the conventional TF-GSC, by exploiting both the desired and interfering speech signals nonstationarity. In Section VI, we discuss the differences between the TF-GSC and the DTF-GSC structures. We stress the benefits gained by applying the novel structure. Experimental results, presented in Section VII, demonstrate the performance of the proposed algorithm in noisy and reverberant environments, with comparison to the TF-GSC structure.

II. Problem Formulation

Consider an array of sensors in a noisy and reverberant environment. We assume that the received signals include three components: a desired speech source, a directional nonstationary interference signal (e.g., competing speech), and a stationary noise signal, which can be either directional, diffused, or uncorrelated (spatially white). Our goal is to reconstruct the desired speech signal from received reverberated signals. Let $s_1(t)$ denote the desired speech signal, let $s_2(t)$ denote the nonstationary interfering signal, and let $a_m(t)$ and $b_m(t)$ represent the acoustical impulse responses of the mth microphone to the desired speech source and the nonstationary interference source, respectively. The mth microphone signal is given by

$$z_m(t) = a_m(t) * s_1(t) + b_m(t) * s_2(t) + n_m(t); \quad m = 1, \ldots, M \quad (1)$$

where $n_m(t)$ is the (directional or nondirectional) stationary noise signal at the mth microphone, and $*$ denotes convolution. The analysis frame duration is chosen such that the signal may be considered stationary over the analysis frame. Typically, the impulse responses $a_m(t)$ and $b_m(t)$ are slowly changing in time and can be considered time-invariant over the analysis frame.

In the short time Fourier transform (STFT) domain, (1) can be approximately rewritten[2] as

$$Z_m(t, e^{j\omega}) \approx A_m(e^{j\omega})S_1(t, e^{j\omega}) + B_m(e^{j\omega})S_2(t, e^{j\omega}) + N_m(t, e^{j\omega}) \quad m = 1, \ldots, M \quad (2)$$

where $Z_m(t, e^{j\omega})$, $S_1(t, e^{j\omega})$, $S_2(t, e^{j\omega})$ and $N_m(t, e^{j\omega})$ are the STFT of the respective signals. $A_m(e^{j\omega})$ and $B_m(e^{j\omega})$ are the ATFs from the desired source and interference source to the mth microphone, respectively, which are assumed hereinafter time-invariant over the observation period. A vector formulation of (2) is

$$\mathbf{Z}(t, e^{j\omega}) = \mathbf{A}(e^{j\omega})S_1(t, e^{j\omega}) + \mathbf{B}(e^{j\omega})S_2(t, e^{j\omega}) + \mathbf{N}(t, e^{j\omega}) \quad (3)$$

where

$$\begin{aligned}
&\mathbf{Z}(t, e^{j\omega}) \\
&\quad = [Z_1(t, e^{j\omega}) \quad Z_2(t, e^{j\omega}) \quad \cdots \quad Z_M(t, e^{j\omega})]^T \\
&\mathbf{A}(e^{j\omega}) \\
&\quad = [A_1(e^{j\omega}) \quad A_2(e^{j\omega}) \quad \cdots \quad A_M(e^{j\omega})]^T \\
&\mathbf{B}(e^{j\omega}) \\
&\quad = [B_1(e^{j\omega}) \quad B_2(e^{j\omega}) \quad \cdots \quad B_M(e^{j\omega})]^T \\
&\mathbf{N}(t, e^{j\omega}) \\
&\quad = [N_1(t, e^{j\omega}) \quad N_2(t, e^{j\omega}) \quad \cdots \quad N_M(t, e^{j\omega})]^T.
\end{aligned}$$

Our problem is to reconstruct the desired speech signal $S_1(t, e^{j\omega})$ (or a filtered version thereof) from the noisy observations $\mathbf{Z}(t, e^{j\omega})$.

[1] A preliminary version of this work was presented in [16].

[2] The approximation sign in (2) can be replaced with equality when the length of the frames is much larger compared with the length of the filter [17]. This assumption cannot be met in our case when considering the ATFs themselves. Equality is only required for the ATFs estimation method. We stress that the actual filtering is conducted with the RTFs rather than the ATFs. The former are regarded as much shorter filters. Finally, the filtering is implemented using the *overlap and save* procedure, and therefore only correct samples from the cyclic convolution results are used.

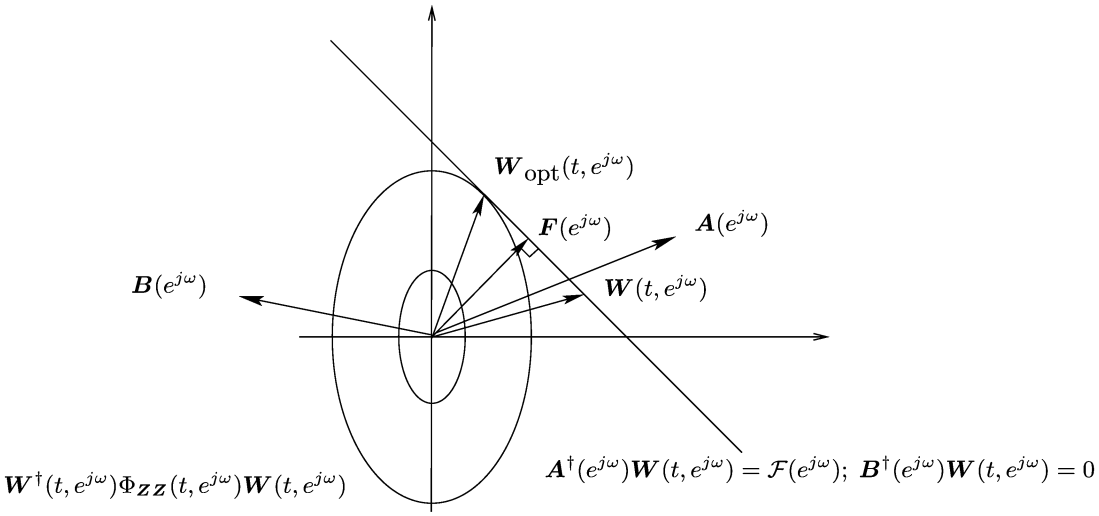

Fig. 1. Graphical interpretation of the constrained minimization in (6). The Ellipsoids depict equi-power surfaces. The line depicts the constraint plane and $\mathbf{F}(e^{j\omega})$ is its perpendicular. $\mathbf{W}(t,e^{j\omega})$ is a possible solution at time instant t, whereas $\mathbf{W}^{\mathrm{opt}}(t,e^{j\omega})$ is the optimal solution, namely it is the vector $\mathbf{W}(t,e^{j\omega})$ fulfilling the constraint while maintaining minimum output power.

III. Optimal Solution Based on Constrained Optimization

In this section, we derive a linearly constrained beamformer, specifically designed for suppressing undesired interference signals. We first obtain a closed-form linearly constrained minimum variance beamformer and then derive an adaptive solution. We initially assume that the ATFs are known, and in Section V we derive their estimates based on the nonstationarity of the speech signals.

Let $W_m^*(t,e^{j\omega}); m=1,\ldots,M$ be a set of M filters

$$\mathbf{W}^H(t,e^{j\omega})=[W_1^*(t,e^{j\omega}) \quad W_2^*(t,e^{j\omega}) \quad \cdots \quad W_M^*(t,e^{j\omega})]$$

where * denotes conjugation and ^{H} denotes conjugation transpose. A beamformer is realized by filtering each sensor output $Z_m(t,e^{j\omega})$ by $W_m^*(t,e^{j\omega}); m=1,\ldots,M$ and summing the outputs

$$\begin{aligned}&Y(t,e^{j\omega})\\&\quad=\mathbf{W}^H(t,e^{j\omega})\mathbf{Z}(t,e^{j\omega})\\&\quad=\mathbf{W}^H(t,e^{j\omega})\mathbf{A}(e^{j\omega})S_1(t,e^{j\omega})\\&\qquad+\mathbf{W}^H(t,e^{j\omega})\mathbf{B}(e^{j\omega})S_2(t,e^{j\omega})\\&\qquad+\mathbf{W}^H(t,e^{j\omega})\mathbf{N}(t,e^{j\omega})\\&\quad\triangleq Y_{s_1}(t,e^{j\omega})+Y_{s_2}(t,e^{j\omega})+Y_n(t,e^{j\omega})\end{aligned}\tag{4}$$

where $Y_{s_1}(t,e^{j\omega})$ represents the desired signal component, $Y_{s_2}(t,e^{j\omega})$ is the directional interference component and $Y_n(t,e^{j\omega})$ is the stationary noise component. The output power is given by

$$\begin{aligned}&E\{Y(t,e^{j\omega})Y^*(t,e^{j\omega})\}\\&\quad=E\{\mathbf{W}^H(t,e^{j\omega})\mathbf{Z}(t,e^{j\omega})\mathbf{Z}^H(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\}\\&\quad=\mathbf{W}^H(t,e^{j\omega})\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\end{aligned}$$

where $\Phi_{\mathbf{ZZ}}(t,e^{j\omega})=E\{\mathbf{Z}(t,e^{j\omega})\mathbf{Z}^H(t,e^{j\omega})\}$. We want to minimize the output power subject to the following constraints

$$\begin{aligned}Y_{s_1}(t,e^{j\omega})&=\mathbf{W}^H(t,e^{j\omega})\mathbf{A}(e^{j\omega})S_1(t,e^{j\omega})\\&=\mathcal{F}^*(e^{j\omega})S_1(t,e^{j\omega})\\Y_{s_2}(t,e^{j\omega})&=\mathbf{W}^H(t,e^{j\omega})\mathbf{B}(e^{j\omega})S_2(t,e^{j\omega})=0\end{aligned}\tag{5}$$

where $\mathcal{F}^*(e^{j\omega})$ is an arbitrary filter response. Hence, the following minimization problem is obtained:

$$\min_{\mathbf{W}}\{\mathbf{W}^H(t,e^{j\omega})\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\}$$

subject to

$$\mathbf{W}^H(t,e^{j\omega})\mathbf{A}(e^{j\omega})=\mathcal{F}^*(e^{j\omega})$$

and

$$\mathbf{W}^H(t,e^{j\omega})\mathbf{B}(e^{j\omega})=0.\tag{6}$$

The minimization in (6) is depicted in Fig. 1. The tangent point of the equi-power contours with the constraint line is the optimum vector of beamformer filters. Solution to the constrained minimization problem is obtained by minimizing the complex Lagrangian

$$\begin{aligned}\mathcal{L}(\mathbf{W})=&\mathbf{W}^H(t,e^{j\omega})\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\\&+\lambda_1[\mathbf{W}^H(t,e^{j\omega})\mathbf{A}(e^{j\omega})-\mathcal{F}^*(e^{j\omega})]\\&+\lambda_1^*[\mathbf{A}^H(e^{j\omega})\mathbf{W}(t,e^{j\omega})-\mathcal{F}(e^{j\omega})]\\&+\lambda_2\mathbf{W}^H(t,e^{j\omega})\mathbf{B}(e^{j\omega})+\lambda_2^*\mathbf{B}^H(e^{j\omega})\mathbf{W}(t,e^{j\omega}).\end{aligned}\tag{7}$$

Setting the derivative with respect to $\mathbf{W}^*$ to zero (see for instance [18]) we obtain

$$\begin{aligned}\nabla_{W^*}\mathcal{L}(\mathbf{W})=\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})+\lambda_1\mathbf{A}(e^{j\omega})\\+\lambda_2\mathbf{B}(e^{j\omega})=0\end{aligned}\tag{8}$$

and since $\Phi_{\mathbf{ZZ}}(t, e^{j\omega})$ is usually invertible,[3] $\mathbf{W}(t, e^{j\omega})$ can be written as

$$\mathbf{W}(t, e^{j\omega}) = -\Phi_{\mathbf{ZZ}}^{-1}(t, e^{j\omega})[\lambda_1 \mathbf{A}(e^{j\omega}) + \lambda_2 \mathbf{B}(e^{j\omega})]. \quad (9)$$

Imposing the constraints on $\mathbf{W}(t, e^{j\omega})$ and solving for the Lagrange multipliers yields (see Appendix A)

$$\mathbf{W}^{\text{opt}}(t, e^{j\omega}) = \mathcal{F}(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t, e^{j\omega}) \times \frac{\frac{\mathbf{A}(e^{j\omega})}{\|\mathbf{A}(e^{j\omega})\|_\Phi^2} - \rho_\Phi(e^{j\omega})\frac{\mathbf{B}(e^{j\omega})}{\|\mathbf{A}(e^{j\omega})\|_\Phi \|\mathbf{B}(e^{j\omega})\|_\Phi}}{1 - |\rho_\Phi(e^{j\omega})|^2} \quad (10)$$

where

$$\|\mathbf{X}(e^{j\omega})\|_\Phi^2 \triangleq \mathbf{X}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t, e^{j\omega})\mathbf{X}(e^{j\omega}) \quad (11)$$

denotes a weighted norm of a vector $\mathbf{X}(e^{j\omega})$ and equation (12) shown at the bottom of the page is the cosine of the angle between the vectors $\mathbf{A}(e^{j\omega})$ and $\mathbf{B}(e^{j\omega})$ in a weighted inner product space.

The closed-form solution to the constrained minimization problem $\mathbf{W}^{\text{opt}}(t, e^{j\omega})$ lacks the ability to track changes in the environment and is difficult to implement. Hence, we replace the closed-form solution with an adaptive one. Consider the following steepest descent recursive algorithm for minimizing the complex Lagrangian in (8)

$$\begin{aligned}\mathbf{W}(t+1, e^{j\omega}) &= \mathbf{W}(t, e^{j\omega}) - \mu\nabla_{W^*}\mathcal{L}(e^{j\omega}) \\ &= \mathbf{W}(t, e^{j\omega}) - \mu[\Phi_{\mathbf{ZZ}}(t, e^{j\omega})\mathbf{W}(t, e^{j\omega}) \\ &\quad + \lambda_1 \mathbf{A}(e^{j\omega}) + \lambda_2 \mathbf{B}(e^{j\omega})]. \end{aligned} \quad (13)$$

Imposing the constraints on $\mathbf{W}(t+1, e^{j\omega})$ yields (see Appendix B)

$$\begin{aligned}\mathbf{W}(t+1, e^{j\omega}) &= P(e^{j\omega})\mathbf{W}(t, e^{j\omega}) \\ &\quad - \mu P(e^{j\omega})\Phi_{\mathbf{ZZ}}(t, e^{j\omega})\mathbf{W}(t, e^{j\omega}) + \mathbf{F}(e^{j\omega}) \end{aligned} \quad (14)$$

where

$$\begin{aligned} P(e^{j\omega}) &= I - \alpha^{-1}[\|\mathbf{B}(e^{j\omega})\|^2 \mathbf{A}(e^{j\omega})\mathbf{A}^H(e^{j\omega}) \\ &\quad - \mathbf{A}(e^{j\omega})\mathbf{A}^H(e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega}) \\ &\quad - \mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega})\mathbf{A}^H(e^{j\omega}) \\ &\quad + \|\mathbf{A}(e^{j\omega})\|^2 \mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})] \\ \mathbf{F}(e^{j\omega}) &= \alpha^{-1}[\|\mathbf{B}(e^{j\omega})\|^2 \mathbf{A}(e^{j\omega}) \\ &\quad - \mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega})]\mathcal{F}(e^{j\omega}) \\ \alpha &\triangleq \|\mathbf{A}(e^{j\omega})\|^2 \|\mathbf{B}(e^{j\omega})\|^2 \\ &\quad - \mathbf{A}^H(e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega}). \end{aligned} \quad (15)$$

[3]As a small amount of uncorrelated sensor noise always exists, the invertibility of $\Phi_{\mathbf{ZZ}}(t, e^{j\omega})$ might be guaranteed in practical scenarios.

This forms the constrained recursive structure. Now, defining $\rho(e^{j\omega})$ as the cosine of the angle between the vectors $\mathbf{A}(e^{j\omega})$ and $\mathbf{B}(e^{j\omega})$ in an inner product space

$$\rho(e^{j\omega}) \equiv \frac{\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega})}{\|\mathbf{A}(e^{j\omega})\|\|\mathbf{B}(e^{j\omega})\|} \quad (16)$$

we obtain

$$\begin{aligned} &P(e^{j\omega}) \\ &= I - \frac{1}{1 - |\rho(e^{j\omega})|^2} \\ &\quad \times \left[\frac{\mathbf{A}(e^{j\omega})\mathbf{A}^H(e^{j\omega})}{\|\mathbf{A}(e^{j\omega})\|^2} - \rho^*(e^{j\omega})\frac{\mathbf{A}(e^{j\omega})\mathbf{B}^H(e^{j\omega})}{\|\mathbf{A}(e^{j\omega})\|\,\|\mathbf{B}(e^{j\omega})\|}\right. \\ &\quad \left. - \rho(e^{j\omega})\frac{\mathbf{B}(e^{j\omega})\mathbf{A}^H(e^{j\omega})}{\|\mathbf{A}(e^{j\omega})\|\,\|\mathbf{B}(e^{j\omega})\|} + \frac{\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})}{\|\mathbf{B}(e^{j\omega})\|^2}\right] \\ &\mathbf{F}(e^{j\omega}) \\ &= \frac{\frac{\mathbf{A}(e^{j\omega})}{\|\mathbf{A}(e^{j\omega})\|^2} - \rho(e^{j\omega})\frac{\mathbf{B}(e^{j\omega})}{\|\mathbf{A}(e^{j\omega})\|\|\mathbf{B}(e^{j\omega})\|}}{1 - |\rho(e^{j\omega})|^2}\mathcal{F}(e^{j\omega}). \end{aligned} \quad (17)$$

The meaning of $\rho(e^{j\omega})$ will be discussed in the next section.

IV. Dual-Source TF-GSC

Following Gannot *et al.* [14] footsteps, we now derive an unconstrained adaptive enhancement algorithm. The unconstrained algorithm is usually advantageous due to its superior computational efficiency and the ability to use the well-behaved normalized least mean squares (NLMS) scheme.

A. Generalized Sidelobe Canceller Interpretation

Consider the null space of $[\mathbf{A}(e^{j\omega}) \mid \mathbf{B}(e^{j\omega})]$, defined by

$$\mathcal{N}(e^{j\omega}) \triangleq \{\mathbf{W} \,|\, [\mathbf{A}(e^{j\omega}) \,|\, \mathbf{B}(e^{j\omega})]^H \mathbf{W}(e^{j\omega}) = [0 \,|\, 0]\}.$$

Define the constraint hyperplane

$$\Lambda(e^{j\omega}) \triangleq \{\mathbf{W} \,|\, [\mathbf{A}(e^{j\omega}) \,|\, \mathbf{B}(e^{j\omega})]^H \mathbf{W}(e^{j\omega}) = [\mathcal{F}(e^{j\omega}) \,|\, 0]\}$$

which is parallel to $\mathcal{N}(e^{j\omega})$. Furthermore, define the column space of $[\mathbf{A}(e^{j\omega}) \,|\, \mathbf{B}(e^{j\omega})]$ by

$$\mathcal{R}(e^{j\omega}) \triangleq \{\kappa_1 \mathbf{A}(e^{j\omega}) + \kappa_2 \mathbf{B}(e^{j\omega}) \,|\, \text{for any real } \kappa_1, \kappa_2 \neq 0\}.$$

Using the fundamental theorem of linear algebra [19], $\mathcal{R}(e^{j\omega}) \perp \mathcal{N}(e^{j\omega})$. The second line in (17) implies that $\mathbf{F}(e^{j\omega}) \in \mathcal{R}(e^{j\omega})$ (as κ_1 and κ_2 can be easily identified) and therefore $\mathbf{F}(e^{j\omega})$ is perpendicular to $\mathcal{N}(e^{j\omega})$. Furthermore

$$\begin{aligned} \mathbf{A}^H(e^{j\omega})\mathbf{F}(e^{j\omega}) &= \mathbf{A}^H(e^{j\omega}) \\ &\quad \times \frac{\frac{\mathbf{A}(\mathbf{e}^{\mathbf{j}\boldsymbol{\omega}})}{\|\mathbf{A}(e^{j\omega})\|^2} - \rho(e^{j\omega})\frac{\mathbf{B}(\mathbf{e}^{\mathbf{j}\boldsymbol{\omega}})}{\|\mathbf{A}(e^{j\omega})\|\|\mathbf{B}(e^{j\omega})\|}}{1 - |\rho(e^{j\omega})|^2} \\ &\quad \times \mathcal{F}(e^{j\omega}) = \mathcal{F}(e^{j\omega}) \end{aligned}$$

$$\rho_\Phi(e^{j\omega}) \triangleq \frac{\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t, e^{j\omega})\mathbf{A}(e^{j\omega})}{\sqrt{\mathbf{A}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t, e^{j\omega})\mathbf{A}(e^{j\omega})}\sqrt{\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t, e^{j\omega})\mathbf{B}(e^{j\omega})}} \quad (12)$$

and

$$\mathbf{B}^H(e^{j\omega})\mathbf{F}(e^{j\omega}) = \mathbf{B}^H(e^{j\omega}) \times \frac{\frac{\mathbf{A}(\mathbf{e^{j\omega}})}{||\mathbf{A}(e^{j\omega})||^2} - \rho(e^{j\omega})\frac{\mathbf{B}(\mathbf{e^{j\omega}})}{||\mathbf{A}(e^{j\omega})||||\mathbf{B}(e^{j\omega})||}}{1-|\rho(e^{j\omega})|^2} \times \mathcal{F}(e^{j\omega}) = 0.$$

Hence, $\mathbf{F}(e^{j\omega}) \in \Lambda(e^{j\omega})$. Now, since $\mathbf{F}(e^{j\omega}) \perp \mathcal{N}(e^{j\omega})$ and $\mathcal{N}(e^{j\omega})$ is parallel to $\Lambda(e^{j\omega})$, $\mathbf{F}(e^{j\omega}) \perp \Lambda(e^{j\omega})$. This implies that $\mathbf{F}(e^{j\omega})$ is the perpendicular from the origin to the constraint hyperplane, $\Lambda(e^{j\omega})$. The matrix $P(e^{j\omega})$ is the *projection matrix* to the null space of $[\mathbf{A}(e^{j\omega})\,|\,\mathbf{B}(e^{j\omega})]$, $\mathcal{N}(e^{j\omega})$. This is easily shown by the following arguments. Using (17) we have

$$\begin{aligned} P(e^{j\omega})\mathbf{A}(e^{j\omega}) &= \mathbf{A}(e^{j\omega}) \\ &\quad - \frac{1}{1-|\rho(e^{j\omega})|^2}\left[\frac{\mathbf{A}(e^{j\omega})\mathbf{A}^H(e^{j\omega})\mathbf{A}(e^{j\omega})}{||\mathbf{A}(e^{j\omega})||^2}\right. \\ &\quad - \rho^*(e^{j\omega})\frac{\mathbf{A}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega})}{||\mathbf{A}(e^{j\omega})||\,||\mathbf{B}(e^{j\omega})||} \\ &\quad - \rho(e^{j\omega})\frac{\mathbf{B}(e^{j\omega})\mathbf{A}^H(e^{j\omega})\mathbf{A}(e^{j\omega})}{||\mathbf{A}(e^{j\omega})||||\mathbf{B}(e^{j\omega})||} \\ &\quad \left. + \frac{\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega})}{||\mathbf{B}(e^{j\omega})||^2}\right]. \end{aligned} \tag{18}$$

The term in the brackets simplifies to

$$\begin{aligned} &\mathbf{A}(e^{j\omega}) - \frac{1}{||\mathbf{A}(e^{j\omega})||^2||\mathbf{B}(e^{j\omega})||^2} \\ &\quad \times [\mathbf{A}^H(e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{A}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega}) \\ &\quad - \mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{A}^H(e^{j\omega})\mathbf{A}(e^{j\omega}) \\ &\quad + \mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{A}^H(e^{j\omega})\mathbf{A}(e^{j\omega})] \\ &\quad = \mathbf{A}(e^{j\omega}) - \frac{\mathbf{A}^H(e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega})}{||\mathbf{A}(e^{j\omega})||^2||\mathbf{B}(e^{j\omega})||^2} \\ &\qquad \times \mathbf{A}(e^{j\omega}) \\ &\quad = \mathbf{A}(e^{j\omega})[1-|\rho(e^{j\omega})|^2] \end{aligned} \tag{19}$$

resulting in $P(e^{j\omega})\mathbf{A}(e^{j\omega}) = 0$. Similarly, it can be shown that $P(e^{j\omega})\mathbf{B}(e^{j\omega}) = 0$ as well.

Now, a vector in a linear space can be uniquely split into a sum of two vectors in mutually orthogonal subspaces (see for instance [19]). Hence

$$\mathbf{W}(t,e^{j\omega}) = \mathbf{W}_0(t,e^{j\omega}) - \mathbf{V}(t,e^{j\omega}) \tag{20}$$

where $\mathbf{W}_0(t,e^{j\omega}) \in \mathcal{R}(e^{j\omega})$ and $-\mathbf{V}(t,e^{j\omega}) \in \mathcal{N}(e^{j\omega})$. By the definition of $\mathcal{N}(e^{j\omega})$

$$\mathbf{V}(t,e^{j\omega}) = \mathcal{H}(e^{j\omega})\mathbf{G}(t,e^{j\omega}) \tag{21}$$

where $\mathcal{H}(e^{j\omega})$ is a full-rank $M \times (M-2)$ matrix, such that the columns of $\mathcal{H}(e^{j\omega})$ span the null space of $[\mathbf{A}(e^{j\omega})\,|\,\mathbf{B}(e^{j\omega})]$, i.e.,

$$\begin{aligned} \mathbf{A}^H(e^{j\omega})\mathcal{H}(e^{j\omega}) &= 0 \\ \mathbf{B}^H(e^{j\omega})\mathcal{H}(e^{j\omega}) &= 0. \end{aligned} \tag{22}$$

The vector $\mathbf{G}(t,e^{j\omega})$ is an $(M-2)\times 1$ vector of adjustable filters.

Using the geometrical interpretation of Frost's algorithm [3] (see Fig. 1)

$$\begin{aligned} \mathbf{W}_0(t,e^{j\omega}) &= \mathbf{F}(e^{j\omega}) \\ &= \frac{\frac{\mathbf{A}(\mathbf{e^{j\omega}})}{||\mathbf{A}(e^{j\omega})||^2} - \rho(e^{j\omega})\frac{\mathbf{B}(\mathbf{e^{j\omega}})}{||\mathbf{A}(e^{j\omega})||||\mathbf{B}(e^{j\omega})||}}{1-||\rho(e^{j\omega})|^2}\mathcal{F}(e^{j\omega}) \end{aligned} \tag{23}$$

(recall that $\mathbf{F}(e^{j\omega})$ is the perpendicular from the origin to the constraint hyperplane, $\Lambda(e^{j\omega})$). Now, using (4), (20), and (21) we obtain

$$Y(t,e^{j\omega}) = Y_{\mathrm{MBF}}(t,e^{j\omega}) - Y_{\mathrm{NC}}(t,e^{j\omega}) \tag{24}$$

where

$$\begin{aligned} Y_{\mathrm{MBF}}(t,e^{j\omega}) &= \mathbf{W}_0^H(t,e^{j\omega})\mathbf{Z}(t,e^{j\omega}) \\ Y_{\mathrm{NC}}(t,e^{j\omega}) &= \mathbf{G}^H(t,e^{j\omega})\mathcal{H}^H(e^{j\omega})\mathbf{Z}(t,e^{j\omega}). \end{aligned} \tag{25}$$

The solution structure is similar to [14], although the constraints are different. The output of the constrained beamformer is a difference of two terms, both operating on the input signal $\mathbf{Z}(t,e^{j\omega})$. $\mathbf{W}_0(t,e^{j\omega})$ in our problem, steers the beam towards the desired direction, while blocking the interference direction. In [14], $\mathbf{W}_0(t,e^{j\omega})$ is only responsible for steering the beam towards the desired direction. Furthermore, $\mathcal{H}(e^{j\omega})$ in the current contribution blocks both directions while in [14] it only blocks the desired direction. $\mathbf{G}(e^{j\omega})$ in both cases has similar functionality. However, its rank here is lower, allowing less degrees of freedom.

The first term $Y_{\mathrm{MBF}}(t,e^{j\omega})$ is dependent on the ATFs; hence, it can be regarded as a MBF. We now examine the second term, $Y_{\mathrm{NC}}(t,e^{j\omega})$. The *reference noise* signals are given by

$$\begin{aligned} \mathbf{U}(t,e^{j\omega}) &= \mathcal{H}^H(e^{j\omega})\mathbf{Z}(t,e^{j\omega}) \\ &= \mathcal{H}^H(e^{j\omega})[\mathbf{A}(e^{j\omega})S_1(t,e^{j\omega}) \\ &\quad + \mathbf{B}(e^{j\omega})S_2(t,e^{j\omega}) + \mathbf{N}(t,e^{j\omega})] \\ &= \mathcal{H}^H(e^{j\omega})\mathbf{N}(t,e^{j\omega}) \end{aligned} \tag{26}$$

where the last transition follows from (22). Both desired and competing signals' components are blocked by $\mathcal{H}^H(e^{j\omega})$ and therefore $\mathbf{U}(t,e^{j\omega})$ contains only noise. Hence, the noise term of $Y_{\mathrm{MBF}}(t,e^{j\omega})$ can be reduced by properly adjusting the filters $\mathbf{G}(t,e^{j\omega})$, using the minimum output power criterion. This adjustment problem is in fact the classical multichannel noise cancellation problem, that can be solved by using the Wiener filter. An adaptive least mean squares (LMS) solution to the problem was proposed by Widrow [20].

Recall that $\rho(e^{j\omega})$ as defined in (16) is the cosine of the angle between $\mathbf{A}(e^{j\omega})$ and $\mathbf{B}(e^{j\omega})$. When these vectors are perpendicular $\rho(e^{j\omega})$ vanishes. In this case, the resulting $\mathbf{F}(e^{j\omega})$ is exactly the single source MBF derived in [14], and the projection matrix reduces to $P(e^{j\omega}) = I - (\mathbf{A}(e^{j\omega})\mathbf{A}^H(e^{j\omega}))/(||\mathbf{A}(e^{j\omega})||^2) - (\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega}))/(||\mathbf{B}(e^{j\omega})||^2)$.

B. Detailed Structure

The solution comprises three building blocks. The first is an MBF, which satisfies the required constraints, i.e., the desired signal is kept undistorted while the nonstationary interfering signal is blocked. The second is a blocking matrix, that produces noise-only reference signals by blocking both the desired and interfering signals. The third block is an unconstrained LMS-type algorithm, that cancels the coherent noise in the MBF output.

1) Blocking Matrix: The blocking matrix should be designed to block both the desired and interfering signals, and yield noise-only components at its outputs. We propose to construct $\mathcal{H}(e^{j\omega})$ as a cascade of two blocking matrices, $\mathcal{H}(e^{j\omega}) = \mathcal{H}_1(e^{j\omega})\mathcal{H}_2(e^{j\omega})$. $\mathcal{H}_1(e^{j\omega})$ is designed to block signals arriving from the desired signal direction, while $\mathcal{H}_2(e^{j\omega})$ is designed to block the signals arriving from the interfering direction, after being rotated by the first matrix. As in [14], $\mathcal{H}_1(e^{j\omega})$ is defined by

$$\mathcal{H}_1(e^{j\omega}) = \begin{bmatrix} -\frac{A_2^*(e^{j\omega})}{A_1^*(e^{j\omega})} & -\frac{A_3^*(e^{j\omega})}{A_1^*(e^{j\omega})} & \cdots & -\frac{A_M^*(e^{j\omega})}{A_1^*(e^{j\omega})} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & \cdots & \ddots & \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{27}$$

Regarding $\mathcal{H}_2(e^{j\omega})$, we have

$$\begin{aligned} \mathbf{B}^H(e^{j\omega})\mathcal{H}(e^{j\omega}) &= \mathbf{B}^H(e^{j\omega})\left(\mathcal{H}_1(e^{j\omega})\mathcal{H}_2(e^{j\omega})\right) \\ &= \left(\mathbf{B}^H(e^{j\omega})\mathcal{H}_1(e^{j\omega})\right)\mathcal{H}_2(e^{j\omega}). \end{aligned} \tag{28}$$

Thus

$$\begin{aligned} \mathbf{B}^H(e^{j\omega})\mathcal{H}_1(e^{j\omega}) &= [B_1^*(e^{j\omega}) \quad B_2^*(e^{j\omega}) \quad \cdots \quad B_M^*(e^{j\omega})] \\ &\quad \cdot \begin{bmatrix} -\frac{A_2^*(e^{j\omega})}{A_1^*(e^{j\omega})} & -\frac{A_3^*(e^{j\omega})}{A_1^*(e^{j\omega})} & \cdots & -\frac{A_M^*(e^{j\omega})}{A_1^*(e^{j\omega})} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & \cdots & \ddots & \\ 0 & 0 & \cdots & 1 \end{bmatrix} \\ &= \begin{bmatrix} -B_1^*(e^{j\omega})\frac{A_2^*(e^{j\omega})}{A_1^*(e^{j\omega})} + B_2^*(e^{j\omega}) \\ -B_1^*(e^{j\omega})\frac{A_3^*(e^{j\omega})}{A_1^*(e^{j\omega})} + B_3^*(e^{j\omega}) \\ \cdots \\ -B_1^*(e^{j\omega})\frac{A_M^*(e^{j\omega})}{A_1^*(e^{j\omega})} + B_M^*(e^{j\omega}) \end{bmatrix}^T. \end{aligned}$$

This vector, multiplied by $\mathcal{H}_2(e^{j\omega})$, should yield a vector of zeros. Consider $\mathcal{H}_2(e^{j\omega})$ of the type

$$\mathcal{H}_2(e^{j\omega}) = \begin{bmatrix} L_3(e^{j\omega}) & L_4(e^{j\omega}) & \cdots & L_M(e^{j\omega}) \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & \cdots & \ddots & \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{29}$$

The following linear equation determines $L_m(e^{j\omega}); m = 3, \ldots, M$

$$\begin{aligned} &\left[-B_1^*(e^{j\omega})\frac{A_2^*(e^{j\omega})}{A_1^*(e^{j\omega})} + B_2^*(e^{j\omega})\right] L_m(e^{j\omega}) \\ &\quad + \left[-B_1^*(e^{j\omega})\frac{A_m^*(e^{j\omega})}{A_1^*(e^{j\omega})} + B_m^*(e^{j\omega})\right] = 0. \end{aligned} \tag{30}$$

Solving (30) we obtain

$$L_m(e^{j\omega}) = -\frac{\frac{A_m^*(e^{j\omega})}{A_1^*(e^{j\omega})} - \frac{B_m^*(e^{j\omega})}{B_1^*(e^{j\omega})}}{\frac{A_2^*(e^{j\omega})}{A_1^*(e^{j\omega})} - \frac{B_2^*(e^{j\omega})}{B_1^*(e^{j\omega})}}; \quad m = 3, \ldots, M. \tag{31}$$

Multiplying $\mathcal{H}_1(e^{j\omega})$ by $\mathcal{H}_2(e^{j\omega})$ and rearranging terms yields

$$\mathcal{H}(e^{j\omega}) = \begin{bmatrix} Q_3(e^{j\omega}) & Q_4(e^{j\omega}) & \cdots & Q_M(e^{j\omega}) \\ L_3(e^{j\omega}) & L_4(e^{j\omega}) & \cdots & L_M(e^{j\omega}) \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & \cdots & \ddots & \\ 0 & 0 & \cdots & 1 \end{bmatrix} \tag{32}$$

where

$$Q_m(e^{j\omega}) = -\frac{\frac{A_2^*(e^{j\omega})}{A_1^*(e^{j\omega})}\frac{B_m^*(e^{j\omega})}{B_1^*(e^{j\omega})} - \frac{B_2^*(e^{j\omega})}{B_1^*(e^{j\omega})}\frac{A_m^*(e^{j\omega})}{A_1^*(e^{j\omega})}}{\frac{A_2^*(e^{j\omega})}{A_1^*(e^{j\omega})} - \frac{B_2^*(e^{j\omega})}{B_1^*(e^{j\omega})}}, \quad m = 3, \ldots, M. \tag{33}$$

We now verify that $\mathcal{H}(e^{j\omega})$ satisfies both constraints in (22), as shown by (34) at the bottom of the page. Calculating the mth element of the right-hand side of the equation

$$\begin{aligned} &A_1^*(e^{j\omega})Q_m(e^{j\omega}) + A_2^*(e^{j\omega})L_m(e^{j\omega}) + A_m^*(e^{j\omega}) \\ &\quad = \frac{1}{\frac{A_2^*(e^{j\omega})}{A_1^*(e^{j\omega})} - \frac{B_2^*(e^{j\omega})}{B_1^*(e^{j\omega})}} \\ &\qquad \cdot \left\{-A_2^*(e^{j\omega})\frac{B_m^*(e^{j\omega})}{B_1^*(e^{j\omega})} + A_m^*(e^{j\omega})\frac{B_2^*(e^{j\omega})}{B_1^*(e^{j\omega})}\right. \end{aligned}$$

$$A^H(e^{j\omega})\mathcal{H}(e^{j\omega}) = \begin{bmatrix} A_1^*(e^{j\omega})Q_3(e^{j\omega}) + A_2^*(e^{j\omega})L_3(e^{j\omega}) + A_3^*(e^{j\omega}) \\ A_1^*(e^{j\omega})Q_4(e^{j\omega}) + A_2^*(e^{j\omega})L_4(e^{j\omega}) + A_4^*(e^{j\omega}) \\ \vdots \\ A_1^*(e^{j\omega})Q_M(e^{j\omega}) + A_2^*(e^{j\omega})L_M(e^{j\omega}) + A_M^*(e^{j\omega}) \end{bmatrix}^T \tag{34}$$

$$- A_2^*(e^{j\omega})\frac{A_m^*(e^{j\omega})}{A_1^*(e^{j\omega})} + A_2^*(e^{j\omega})\frac{B_m^*(e^{j\omega})}{B_1^*(e^{j\omega})} + A_2^*(e^{j\omega})\frac{A_m^*(e^{j\omega})}{A_1^*(e^{j\omega})} - A_m^*(e^{j\omega})\frac{B_2^*(e^{j\omega})}{B_1^*(e^{j\omega})}\Bigg\} = 0 \tag{35}$$

the required solution is obtained. Similarly, $\mathcal{H}(e^{j\omega})$ satisfies the second constraint, $B^H(e^{j\omega})\mathcal{H}(e^{j\omega}) = 0$. Therefore, $\mathcal{H}(e^{j\omega})$ is a valid blocking matrix which is suitable for generating the reference noise signals. Using (26), we have

$$U_m(t,e^{j\omega}) = Q_m(e^{j\omega})Z_1(t,e^{j\omega}) + L_m(e^{j\omega})Z_2(t,e^{j\omega}) + Z_m(t,e^{j\omega}) \quad m = 3,\ldots,M. \tag{36}$$

Thus, the knowledge of both $(A_m(e^{j\omega}))/(A_1(e^{j\omega}))$ and $(B_m(e^{j\omega}))/(B_1(e^{j\omega}))$, or directly $Q_m(e^{j\omega})$ and $L_m(e^{j\omega})$, is sufficient for generating the noise reference signals.

2) Matched Beamformer: It was shown in Section IV-A that the MBF $\mathbf{W}_0(e^{j\omega}) = \mathbf{F}(e^{j\omega})$, given by (17), satisfies the required constraints. If in (15), the actual ATFs are replaced by the RTFs, then we have (37), as shown at the bottom of the page. When using the RTFs instead of $\mathbf{A}(e^{j\omega})$ and $\mathbf{B}(e^{j\omega})$, we obtain (38), as shown at the bottom of the page. Namely, the desired signal is only distorted by the first ATF $A_1(e^{j\omega})$, which can be absorbed into $\mathcal{F}^*(e^{j\omega})$. In a similar way, it can be shown that the nonstationary interference is completely blocked. Thus, the knowledge of the RTFs is sufficient for implementing the sidelobe canceller.

It should be noticed that maximum directivity of the MBF can be obtained at directions other than the desired signal direction. Consider the following example, as depicted in Fig. 2. Polar plots of directivity patterns are computed for five and ten microphones arrays for several frequencies for the simple delay-only case. The desired source signal is arriving from direction 90°, while the nonstationary interference impinges on the array from direction 100°. It is clear that the beamformers satisfy both constraints in all plots, namely the gain is 1 in the desired direction and 0 in the interference direction. However, the array with ten microphones outperforms the one with $M = 5$. For example, at 500 Hz the array with five microphones has maximum gain of 3.3 at direction of 45°, while using ten microphones maximum gain of 1.8 is achieved at direction of 75°, i.e., the peak is closer to the desired direction and to unity gain. Hence, interfering signals may be emphasized when not using enough microphones, especially in the low frequencies, which may deteriorate the ability of the ANC to cancel the stationary noise signals.

3) Multichannel Noise Canceller: Recall that our goal is to minimize the output power under constraints on the response at the desired signal direction and at the competing signal direction. By setting $\mathbf{W}_0(t,e^{j\omega})$ according to (37), the constraints are satisfied. Hence, minimization of the output power is achieved by adjusting the filters $\mathbf{G}(t,e^{j\omega})$. This is an unconstrained minimization, exactly as in Widrow's classical problem [20]. It can be implemented by using the multichannel Wiener filter. Recall (24), our goal is to set $\mathbf{G}(t,e^{j\omega})$ to minimize

$$\mathrm{E}\{||Y_{\mathrm{MBF}}(t,e^{j\omega}) - \mathbf{G}^H(t,e^{j\omega})\mathbf{U}(t,e^{j\omega})||^2\}.$$

Let

$$\Phi_{\mathbf{U}Y}(t,e^{j\omega}) = E\{\mathbf{U}(t,e^{j\omega})Y_{\mathrm{MBF}}^*(t,e^{j\omega})\}$$
$$\Phi_{\mathbf{U}\mathbf{U}}(t,e^{j\omega}) = E\{\mathbf{U}(t,e^{j\omega})\mathbf{U}^H(t,e^{j\omega})\}.$$

Then the multichannel Wiener filter is given by [10], [21]

$$\mathbf{G}(t,e^{j\omega}) = \Phi_{\mathbf{U}\mathbf{U}}^{-1}(t,e^{j\omega})\Phi_{\mathbf{U}Y}(t,e^{j\omega}). \tag{39}$$

In order to be able to track changes, the signals are processed by segments. The following frequency domain LMS algorithm is used. Let the residual signal be

$$Y(t,e^{j\omega}) = Y_{\mathrm{MBF}}(t,e^{j\omega}) - \mathbf{G}^H(t,e^{j\omega})\mathbf{U}(t,e^{j\omega}).$$

$$\mathbf{W}_0(e^{j\omega}) = \frac{\frac{\|\mathbf{B}(e^{j\omega})\|^2}{|B_1(e^{j\omega})|^2}\frac{\mathbf{A}(e^{j\omega})}{A_1(e^{j\omega})} - \frac{\mathbf{B}(e^{j\omega})}{B_1(e^{j\omega})}\frac{\mathbf{B}^H(e^{j\omega})}{B_1^*(e^{j\omega})}\frac{\mathbf{A}(e^{j\omega})}{A_1(e^{j\omega})}}{\frac{\|\mathbf{A}(e^{j\omega})\|^2}{|A_1(e^{j\omega})|^2}\frac{\|\mathbf{B}(e^{j\omega})\|^2}{|B_1(e^{j\omega})|^2} - \frac{\mathbf{A}^H(e^{j\omega})}{A_1^*(e^{j\omega})}\frac{\mathbf{B}(e^{j\omega})}{B_1(e^{j\omega})}\frac{\mathbf{B}^H(e^{j\omega})}{B_1^*(e^{j\omega})}\frac{\mathbf{A}(e^{j\omega})}{A_1(e^{j\omega})}}\mathcal{F}(e^{j\omega}) \tag{37}$$

$$\begin{aligned}\mathbf{W}_0^H(e^{j\omega})\mathbf{A}(e^{j\omega}) &= \left[\frac{\frac{\|\mathbf{B}(e^{j\omega})\|^2}{|B_1(e^{j\omega})|^2}\frac{\mathbf{A}(e^{j\omega})}{A_1(e^{j\omega})} - \frac{\mathbf{B}(e^{j\omega})}{B_1(e^{j\omega})}\frac{\mathbf{B}^H(e^{j\omega})}{B_1^*(e^{j\omega})}\frac{\mathbf{A}(e^{j\omega})}{A_1(e^{j\omega})}}{\frac{\|\mathbf{A}(e^{j\omega})\|^2}{|A_1(e^{j\omega})|^2}\frac{\|\mathbf{B}(e^{j\omega})\|^2}{|B_1(e^{j\omega})|^2} - \frac{\mathbf{A}^H(e^{j\omega})}{A_1^*(e^{j\omega})}\frac{\mathbf{B}(e^{j\omega})}{B_1(e^{j\omega})}\frac{\mathbf{B}^H(e^{j\omega})}{B_1^*(e^{j\omega})}\frac{\mathbf{A}(e^{j\omega})}{A_1(e^{j\omega})}}\right]^H \mathcal{F}^*(e^{j\omega})\mathbf{A}(e^{j\omega}) \\ &= \frac{\frac{1}{|B_1(e^{j\omega})|^2}\frac{1}{A_1^*(e^{j\omega})}\left[||\mathbf{B}(e^{j\omega})||^2||\mathbf{A}(e^{j\omega})||^2 - \mathbf{A}^H(e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega})\right]^*}{\frac{1}{|A_1(e^{j\omega})|^2}\frac{1}{|B_1(e^{j\omega})|^2}\left[||\mathbf{B}(e^{j\omega})||^2||\mathbf{A}(e^{j\omega})||^2 - \mathbf{A}^H(e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega})\right]^*}\mathcal{F}^*(e^{j\omega}) \\ &= A_1(e^{j\omega})\mathcal{F}^*(e^{j\omega}).\end{aligned} \tag{38}$$

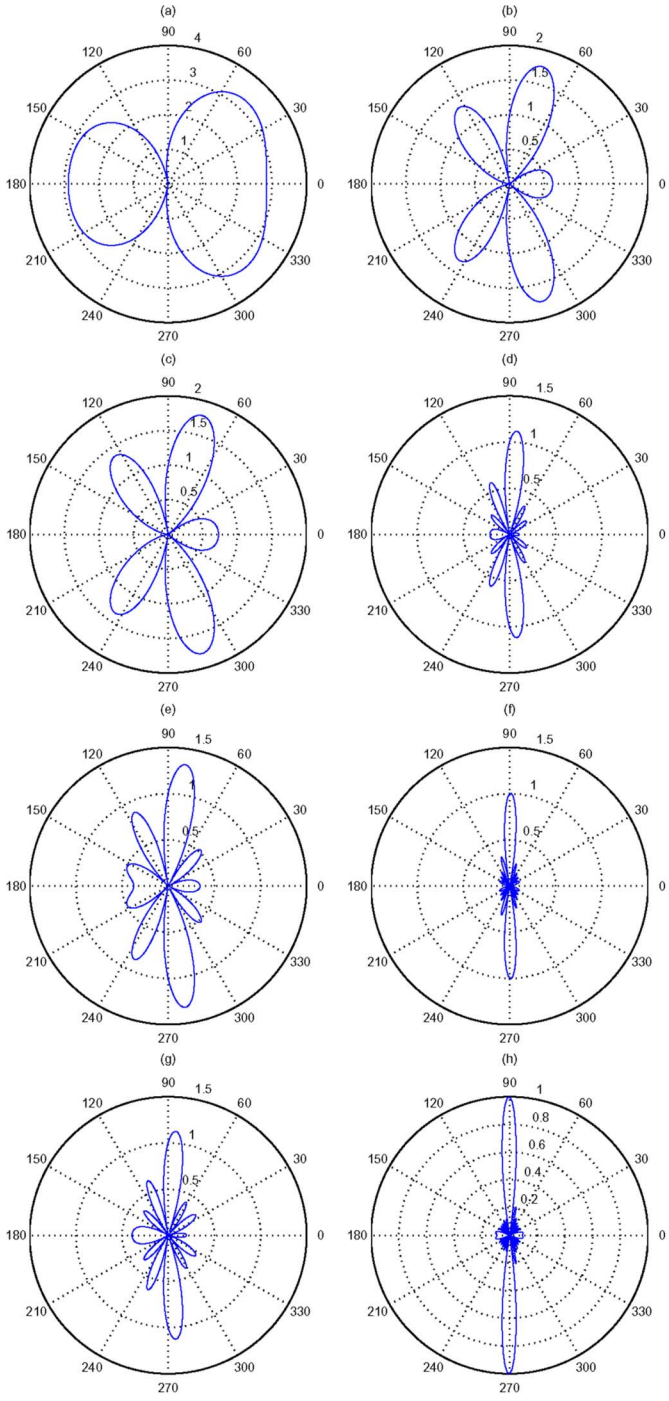

Fig. 2. MBF directivity patterns for several scenarios: $M = 5$, (a) $f = 500$ Hz, (c) $f = 1000$ Hz, (e) $f = 1500$ Hz, (g) $f = 2000$ Hz; $M = 10$, (b) $f = 500$ Hz, (d) $f = 1000$ Hz, (f) $f = 1500$Hz, (h) $f = 2000$ Hz.

Note that the residual signal is also the output of the enhancement algorithm. Using the orthogonality principle (see, e.g., [22]), the error is orthogonal to the measurements. Thus

$$E\{\mathbf{U}(t,e^{j\omega})Y^*(t,e^{j\omega})\} = 0. \tag{40}$$

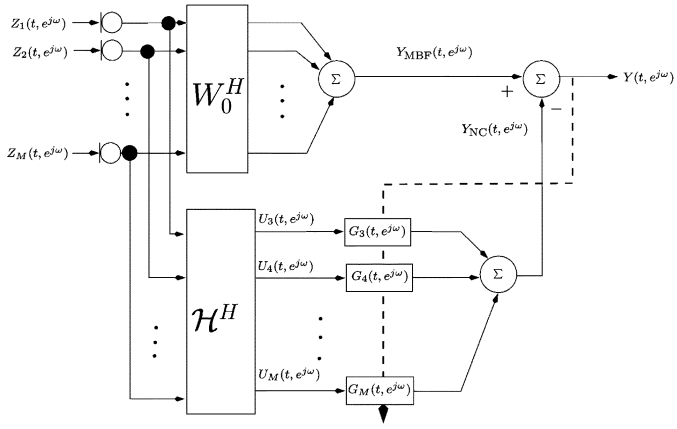

Fig. 3. GSC solution for the dual-source case. Three blocks: an MBF $\boldsymbol{W}_0^H(t,e^{j\omega})$; a BM $\mathcal{H}^H(e^{j\omega})$; and a multichannel ANC $\boldsymbol{G}(t,e^{j\omega})$.

Following the standard Widrow procedure, the solution is given by

$$\mathbf{G}(t+1,e^{j\omega}) = \mathbf{G}(t,e^{j\omega}) + \mu\mathbf{U}(t,e^{j\omega})Y^*(t,e^{j\omega}).$$

Usually, a more stable solution is obtained by using the NLMS algorithm, in which case each frequency is normalized separately, yielding

$$G_m(t+1,e^{j\omega}) = G_m(t,e^{j\omega}) + \mu\frac{U_m(t,e^{j\omega})Y^*(t,e^{j\omega})}{P_{\mathrm{est}}(t,e^{j\omega})}, \quad m = 2,\ldots,M$$

where

$$P_{\mathrm{est}}(t,e^{j\omega}) = \eta P_{\mathrm{est}}(t-1,e^{j\omega}) + (1-\eta)\sum_m |Z_m(t,e^{j\omega})|^2 \tag{41}$$

and η is a forgetting factor (typically $0.8 < \eta < 1$).[4]

The filter update is now given by

$$\tilde{G}_m(t+1,e^{j\omega}) = G_m(t,e^{j\omega}) + \mu\frac{U_m(t,e^{j\omega})Y^*(t,e^{j\omega})}{P_{\mathrm{est}}(t,e^{j\omega})}$$
$$G_m(t+1,e^{j\omega}) \overset{\mathrm{FIR}}{\longleftarrow} \tilde{G}_m(t+1,e^{j\omega}) \tag{42}$$

for $m = 3,\ldots,M$. The operator $\overset{\mathrm{FIR}}{\longleftarrow}$ includes the following three stages. First, $\tilde{G}_m(t+1,e^{j\omega})$ is transformed to the time domain. Second, the resulting impulse response is truncated, namely an FIR constraint is imposed. Third, the result is transformed back to the frequency domain. Performing the $\overset{\mathrm{FIR}}{\longleftarrow}$ operator avoids cyclic convolution. A block diagram of the GSC solution is depicted in Fig. 3, and the proposed algorithm is summarized in Algorithm 1.

[4]Another possibility is to calculate P_{est} using the $|U_m(t,e^{j\omega})|^2$ rather than using $|Z_m(t,e^{j\omega})|^2$. However, in that case an energy detector is required, so that $\mathbf{G}(t,e^{j\omega})$ is updated only when there is no active signal. If on the other hand, we calculate $P_{\mathrm{est}}(t,e^{j\omega})$ using the input sensor signals, $Z_m(t,e^{j\omega})$, as indicated in (41), then an energy detector may be avoided. This is due to the fact that the adaptation term becomes relatively small during periods of active input signal.

Algorithm 1—Summary of the DTF-GSC Algorithm

1) Matched beamformer:

$$Y_{\mathrm{MBF}}(t,e^{j\omega}) = \mathbf{W}_0^H(e^{j\omega})\mathbf{Z}(t,e^{j\omega})$$

$\mathbf{W}_0(e^{j\omega})$ is defined in (37) and its components are estimated using (49).

2) Noise reference signals:

$$\mathbf{U}(t,e^{j\omega}) = \mathcal{H}^H(e^{j\omega})\mathbf{Z}(t,e^{j\omega})$$

$\mathcal{H}(e^{j\omega})$ is defined in (32) and its components are estimated using either (49) or (53).

3) Output signal:

$$Y(t,e^{j\omega}) = Y_{\mathrm{MBF}}(t,e^{j\omega}) - \mathbf{G}^H(t,e^{j\omega})\mathbf{U}(t,e^{j\omega})$$

4) Filters update, for $m = 3,\ldots,M$:

$$\tilde{G}_m(t+1,e^{j\omega}) = G_m(t,e^{j\omega}) + \mu\frac{U_m(t,e^{j\omega})Y^*(t,e^{j\omega})}{P_{\mathrm{est}}(t,e^{j\omega})}$$

$$G_m(t+1,e^{j\omega}) \overset{\mathrm{FIR}}{\longleftarrow} \tilde{G}_m(t+1,e^{j\omega})$$

where $P_{\mathrm{est}}(t,e^{j\omega}) = \eta P_{\mathrm{est}}(t-1,e^{j\omega}) + (1-\eta)\sum_m |Z_m(t,e^{j\omega})|^2$

5) Keep only nonaliased samples.

V. RTF Estimation

The RTFs $(\mathbf{A}(e^{j\omega}))/(A_1(e^{j\omega}))$ and $(\mathbf{B}(e^{j\omega}))/(B_1(e^{j\omega}))$ are required for calculating the MBF and the blocking matrix. Until this point, the RTFs were assumed to be known. However, in practice, they are estimated from the observed noisy signals. We assume that the RTFs are slowly changing in time compared with the time variations of the desired signal and the competing speech signal. We also assume that the statistics of the noise signal is slowly changing compared with the statistics of both the desired signal and the competing speech signal.

A. Matched Beamformer Estimate

Estimation of the MBF is carried out in two steps. First, the RTFs $(\mathbf{A}(e^{j\omega}))/(A_1(e^{j\omega}))$ and $(\mathbf{B}(e^{j\omega}))/(B_1(e^{j\omega}))$ are estimated separately, using the system identification procedure described in [14]. Second, $\mathbf{W}_0(t,e^{j\omega})$ is estimated using (37), where the RTFs are used instead of the real ATFs. Since the system identification algorithm is designed for estimating a single system at a time, the two ratios cannot be estimated simultaneously. Therefore, only frames in which both signals are not simultaneously active are used.

We will now briefly describe the system identification algorithm. The observation period is divided into frames such that the desired or the competing speech signals may be considered stationary during each kth frame. Define $H_m(e^{j\omega}) \triangleq (A_m(e^{j\omega}))/(A_1(e^{j\omega}))$. Note that when no competing speech signal is in present, (2) becomes

$$Z_m(t,e^{j\omega}) = A_m(e^{j\omega})S(t,e^{j\omega}) + N_m(t,e^{j\omega}), \quad m = 1,\ldots,M. \tag{43}$$

It is shown in [14] that the following $U_m(t,e^{j\omega})$ are proper noise reference signals

$$U_m(t,e^{j\omega}) = Z_m(t,e^{j\omega}) - \frac{A_m(e^{j\omega})}{A_1(e^{j\omega})}Z_1(t,e^{j\omega}), \quad m = 2,\ldots,M. \tag{44}$$

Rearranging terms in (44) yields

$$Z_m(t,e^{j\omega}) = H_m(e^{j\omega})Z_1(t,e^{j\omega}) + U_m(t,e^{j\omega}). \tag{45}$$

Consider some analysis interval for which the RTFs are assumed to be time-invariant and the noise signal stationary. We divide this analysis interval into frames, such that the desired signal may be considered stationary during each frame. Consider the kth frame, using (45) we have

$$\Phi^{(k)}_{z_m z_1}(e^{j\omega}) = H_m(e^{j\omega})\Phi^{(k)}_{z_1 z_1}(e^{j\omega}) + \Phi_{u_m z_1}(e^{j\omega}), \quad k = 1,\ldots,K \tag{46}$$

where K is the number of frames used and $\Phi^{(k)}_{z_i z_j}(e^{j\omega})$ is the cross-PSD between z_i and z_j during the kth frame. $\Phi_{u_m z_1}(e^{j\omega})$ is the cross-PSD between u_m and z_1. Using (43) and (44)

$$U_m(t,e^{j\omega}) = N_m(t,e^{j\omega}) - H_m(e^{j\omega})N_1(t,e^{j\omega}) \tag{47}$$

$$Z_1(t,e^{j\omega}) = A_1(e^{j\omega})S(t,e^{j\omega}) + N_1(t,e^{j\omega}) \tag{48}$$

Since $N_m(t,e^{j\omega}), m = 1,\ldots,M$ are assumed stationary over the analysis interval and since $S(t,e^{j\omega})$ is independent of $N_m(e^{j\omega})$, it follows that $\Phi_{u_m z_1}(e^{j\omega})$ is independent of the frame index k.

Let $\hat{\Phi}^{(k)}_{z_1 z_1}(e^{j\omega})$, $\hat{\Phi}^{(k)}_{z_m z_1}(e^{j\omega})$ and $\hat{\Phi}^{(k)}_{u_m z_1}(e^{j\omega})$ be estimates of $\Phi^{(k)}_{z_1 z_1}(e^{j\omega})$, $\Phi^{(k)}_{z_m z_1}(e^{j\omega})$ and $\Phi_{u_m z_1}(e^{j\omega})$, respectively. The estimates are obtained by replacing expectations with averages.

Hence

$$\hat{\Phi}^{(k)}_{z_m z_1}(e^{j\omega}) = H_m(e^{j\omega})\hat{\Phi}^{(k)}_{z_1 z_1}(e^{j\omega}) + \Phi_{u_m z_1}(e^{j\omega}) + \varepsilon^{(k)}_m(e^{j\omega})$$

where $\varepsilon^{(k)}_m(e^{j\omega}) = \hat{\Phi}^{(k)}_{u_m z_1}(e^{j\omega}) - \Phi_{u_m z_1}(e^{j\omega})$ denotes the estimation error of the cross-PSD between z_1 and u_m in the kth frame. Hence, an unbiased estimate of H_m $(m = 2,\ldots,M)$ can be obtained by applying the least squares criterion to the following set of over-determined equations

$$\begin{bmatrix} \hat{\Phi}^{(1)}_{z_m z_1}(e^{j\omega}) \\ \hat{\Phi}^{(2)}_{z_m z_1}(e^{j\omega}) \\ \vdots \\ \hat{\Phi}^{(K)}_{z_m z_1}(e^{j\omega}) \end{bmatrix} = \begin{bmatrix} \hat{\Phi}^{(1)}_{z_1 z_1}(e^{j\omega}) & 1 \\ \hat{\Phi}^{(2)}_{z_1 z_1}(e^{j\omega}) & 1 \\ \vdots & \\ \hat{\Phi}^{(K)}_{z_1 z_1}(e^{j\omega}) & 1 \end{bmatrix}$$

$$\times \begin{bmatrix} H_m(e^{j\omega}) \\ \Phi_{u_m z_1}(e^{j\omega}) \end{bmatrix} + \begin{bmatrix} \varepsilon_m^{(1)}(e^{j\omega}) \\ \varepsilon_m^{(2)}(e^{j\omega}) \\ \vdots \\ \varepsilon_m^{(K)}(e^{j\omega}) \end{bmatrix} \quad (49)$$

where a separate set of equations is used for $m = 2, \ldots, M$. Note that the ratio $(\mathbf{B}(e^{j\omega}))/(B_1(e^{j\omega}))$ can be estimated in a similar manner.

B. Blocking Matrix Estimate

Inspecting (31) and (33), we note that the filters $Q_m(e^{j\omega})$ and $L_m(e^{j\omega})$ can be estimated by first estimating the RTFs and then substituting these estimates into the BM given in (32). However, in this section we propose a direct estimation method for $Q_m(e^{j\omega})$ and $L_m(e^{j\omega})$, which is applicable during *double talk* periods only. This facilitates tracking the BM when both the desired and interfering signals are active.

Rearranging terms in (36) yields

$$Z_m(t, e^{j\omega}) = -Q_m(e^{j\omega}) Z_1(t, e^{j\omega}) - L_m(e^{j\omega}) Z_2(t, e^{j\omega}) + U_m(t, e^{j\omega}). \quad (50)$$

We choose an observation period in which both the desired and competing speech signals are active simultaneously, and divide this period into frames such that the desired and directional interference signals may be considered stationary during each kth frame. Using (50) we obtain a system identification procedure

$$\Phi_{z_m z_1}^{(k)}(e^{j\omega}) = -Q_m(e^{j\omega}) \Phi_{z_1 z_1}^{(k)}(e^{j\omega}) - L_m(e^{j\omega}) \Phi_{z_2 z_1}^{(k)}(e^{j\omega}) + \Phi_{u_m z_1}(e^{j\omega}), \quad k = 1, \ldots, K \quad (51)$$

where K is the number of frames in the interval, $\Phi_{z_i z_j}^{(k)}(e^{j\omega})$ is the cross-power spectral density (PSD) between z_i and z_j during the kth frame, and $\Phi_{u_m z_1}(e^{j\omega})$ is the cross-PSD between u_m and z_1 (note that $\Phi_{u_m z_1}(e^{j\omega})$ is independent of the frame index k [14]). By replacing real PSD values with their estimates, calculated using time-averages, the following set of equations is obtained:

$$\hat{\Phi}_{z_m z_1}^{(k)}(e^{j\omega}) = -Q_m(e^{j\omega}) \hat{\Phi}_{z_1 z_1}^{(k)}(e^{j\omega}) - L_m(e^{j\omega}) \hat{\Phi}_{z_2 z_1}^{(k)}(e^{j\omega}) + \hat{\Phi}_{u_m z_1}(e^{j\omega}) + \varepsilon_m^{(k)}(e^{j\omega}), \quad k = 1, \ldots, K \quad (52)$$

where $\varepsilon_m^{(k)}(e^{j\omega})$ represents the estimation error in the kth frame. An unbiased estimate of $Q_m(e^{j\omega})$ and $L_m(e^{j\omega})$ is obtained by applying the least squares criterion to the following over-determined set of equations

$$\begin{bmatrix} \hat{\Phi}_{z_m z_1}^{(1)}(e^{j\omega}) \\ \hat{\Phi}_{z_m z_1}^{(2)}(e^{j\omega}) \\ \vdots \\ \hat{\Phi}_{z_m z_1}^{(K)}(e^{j\omega}) \end{bmatrix} = \begin{bmatrix} \hat{\Phi}_{z_1 z_1}^{(1)}(e^{j\omega}) & \hat{\Phi}_{z_2 z_1}^{(1)}(e^{j\omega}) & 1 \\ \hat{\Phi}_{z_1 z_1}^{(2)}(e^{j\omega}) & \hat{\Phi}_{z_2 z_1}^{(2)}(e^{j\omega}) & 1 \\ \vdots & & \\ \hat{\Phi}_{z_1 z_1}^{(K)}(e^{j\omega}) & \hat{\Phi}_{z_2 z_1}^{(K)}(e^{j\omega}) & 1 \end{bmatrix} \times \begin{bmatrix} -Q_m(e^{j\omega}) \\ -L_m(e^{j\omega}) \\ \Phi_{u_m z_1}(e^{j\omega}) \end{bmatrix} + \begin{bmatrix} \varepsilon_m^{(1)}(e^{j\omega}) \\ \varepsilon_m^{(2)}(e^{j\omega}) \\ \vdots \\ \varepsilon_m^{(K)}(e^{j\omega}) \end{bmatrix} \quad (53)$$

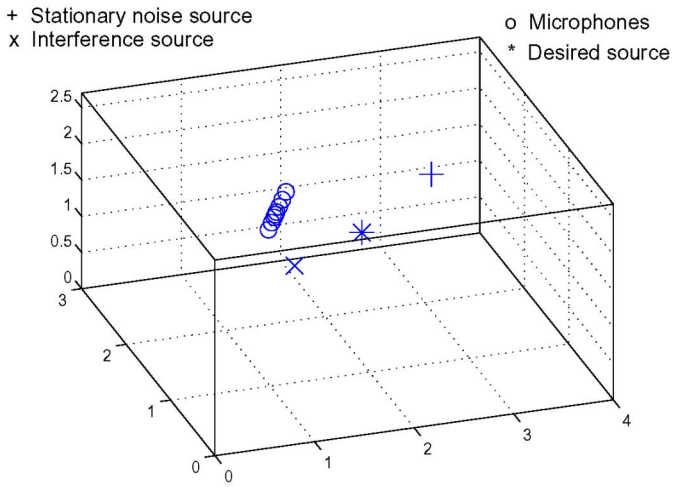

Fig. 4. Room configuration: desired speech signal, interference source (competing speaker), stationary noise source, and the microphone locations.

where a separate set of equations is used for $m = 3, \ldots, M$.

We assume that perfect voice activity and double-talk detectors are available. In periods when only the desired signal is active, an estimate of $(\mathbf{A}(e^{j\omega}))/(A_1(e^{j\omega}))$ is updated, while estimates of $(\mathbf{B}(e^{j\omega}))/(B_1(e^{j\omega}))$ are updated when only the interference is active. These estimates can be combined in the indirect calculation of the BM elements. The update of the BM by direct estimation of $Q_m(e^{j\omega})$ and $L_m(e^{j\omega})$ can be performed only during *double talk* periods, when no other update is applicable.

VI. Discussion

Two paradigms can be adopted for designing a beamformer for enhancing a desired signal contaminated by both noise and interference. These paradigms differ in their treatment of the nonstationary interference (competing speech). The straightforward alternative is to apply the single constraint beamformer (Griffiths and Jim [5] efficient implementation of Frost's LCMV [3], or the TF-GSC [14]), in which a beam is steered towards the desired signal, while all other interference signals are treated by the ANC. Buckley and Van Veen [1], in their beamforming survey, explicitly discuss a LCMV, constrained to null out an interference signal while maintaining the response towards the desired source. In this implementation, the ANC is only responsible for the stationary noise. In our proposed method, we adopt this structure and propose a practical implementation for the multilinear constraint, denoted DTF-GSC. The proposed structure is capable of steering a beam towards the desired signal while simultaneously steering a null towards the interferer.

Both the TF-GSC and the DTF-GSC reduce the stationary interference by applying an LMS-based ANC to the BM outputs. However, a major difference between the structures lies in the adaptation mechanism of the ANC. While in the TF-GSC the input signal of the ANC is comprised of both nonstationary competing speech signal and a stationary noise signal, the input to the corresponding DTF-GSC block is comprised of only the latter. The need of the ANC to adapt during both the stationary and nonstationary signals imposes contradicting requirements on the adaptation rate. On the one hand, the adaptation factor

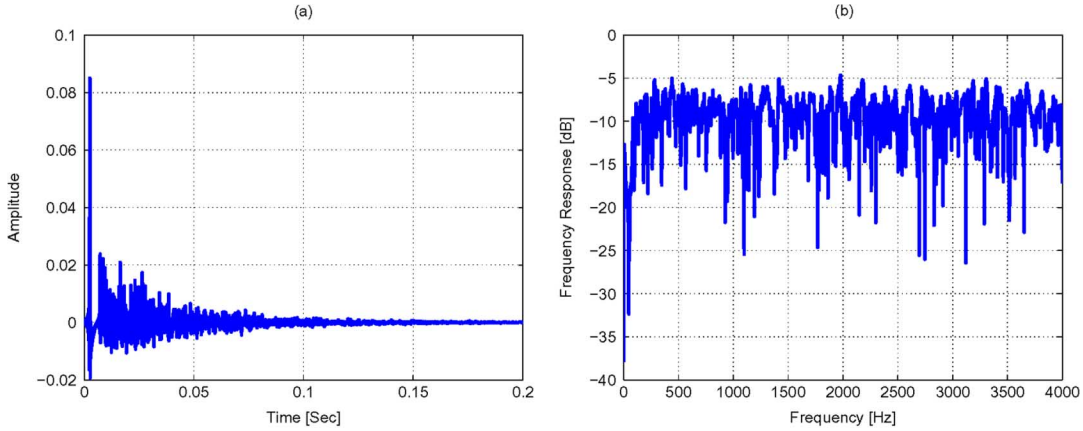

Fig. 5. (a) Impulse response and (b) frequency response relating the desired source and the first microphone.

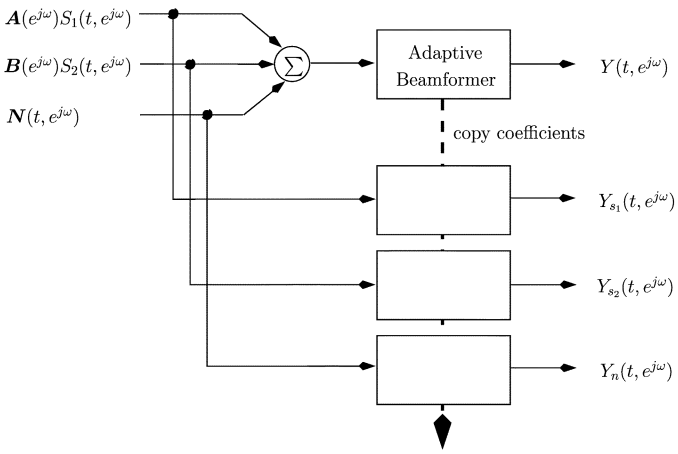

Fig. 6. Test procedure for evaluating the DTF-GSC in comparison with the TF-GSC.

μ should be high enough to allow for tracking of a fast-varying signal, and on the other hand, it should be low enough to enable sufficient reduction of the stationary noise level. Such a requirement on the TF-GSC necessitates the use of an algorithm for distinguishing between the two nonstationary signals, namely, the desired and interference signals. On the other hand, differentiating between slowly varying noise signal and the highly nonstationary speech signal is a considerably easier task [23], [24].

As shown in the experimental study in Section VII, the residual noise level of the TF-GSC structure is fluctuating over time (in accordance with the activity periods of the interference signal), while the corresponding signal level at the DTF-GSC output is much more stable. As beamformer algorithms are very often followed by a postfilter [25], [26], and since postfilters are sensitive to nonstationary noise signals, it is crucial to maintain the residual noise level as stable as possible. The application of a postfilter is beyond the scope of this contribution.

The design of the null in the beampattern towards the competing speech is another difference between the TF-GSC and the DTF-GSC. While in the former the null is built adaptively, using only power consideration, in the latter the null is established by exploiting the speech characteristics. The use of speech nonstationarity yields a more robust null and thus allowing for an improved interference cancellation. Furthermore, the estimation method derived in (53) allows for estimating BM even in *double talk* scenarios. Another drawback of the TF-GSC is its dependence on the FBF beampattern. If the interference signal level at the FBF output is significantly reduced while maintaining high level at the BM output, the ANC might increase the amount of interference leakage to the total output. In the DTF-GSC, where the ANC block is only responsible for the noise signal, this problem cannot be encountered. A comparison of the computational burden of both algorithms shows an advantage to the TF-GSC over the DTF-GSC, since two sets of RTFs should be estimated for the latter. However, this difference is not of crucial importance.

VII. Experimental Study

A. Test Scenario

The proposed algorithm was tested in a simulated room environment. The desired and competing speech signals were drawn from the TIMIT [27] database, while a speech-like noise from NOISEX-92 [28] database was used as a stationary noise source. All three signals were filtered by simulated RIRs, resulting in directional signals, which are received by $M = 8$ microphones. The microphone locations were set to

$$\begin{aligned}
\boldsymbol{m}_1^T &= [1.33 \quad 1.77 \quad 1.50] \\
\boldsymbol{m}_2^T &= [1.41 \quad 1.88 \quad 1.50] \\
\boldsymbol{m}_3^T &= [1.47 \quad 1.96 \quad 1.50] \\
\boldsymbol{m}_4^T &= [1.50 \quad 2.00 \quad 1.50] \\
\boldsymbol{m}_5^T &= [1.53 \quad 2.04 \quad 1.50] \\
\boldsymbol{m}_6^T &= [1.59 \quad 2.12 \quad 1.50] \\
\boldsymbol{m}_7^T &= [1.67 \quad 2.23 \quad 1.50] \\
\boldsymbol{m}_8^T &= [1.76 \quad 2.35 \quad 1.50]. \qquad (54)
\end{aligned}$$

The desired source is located at $\boldsymbol{r}_s^T = [2.1 \ 1.4 \ 1.6]$, the interference source at $\boldsymbol{r}_i^T = [1.2 \ 0.9 \ 1.7]$, and the stationary noise source at $\boldsymbol{r}_n^T = [3.2 \ 2.3 \ 1.5]$. The test scenario is depicted in Fig. 4.

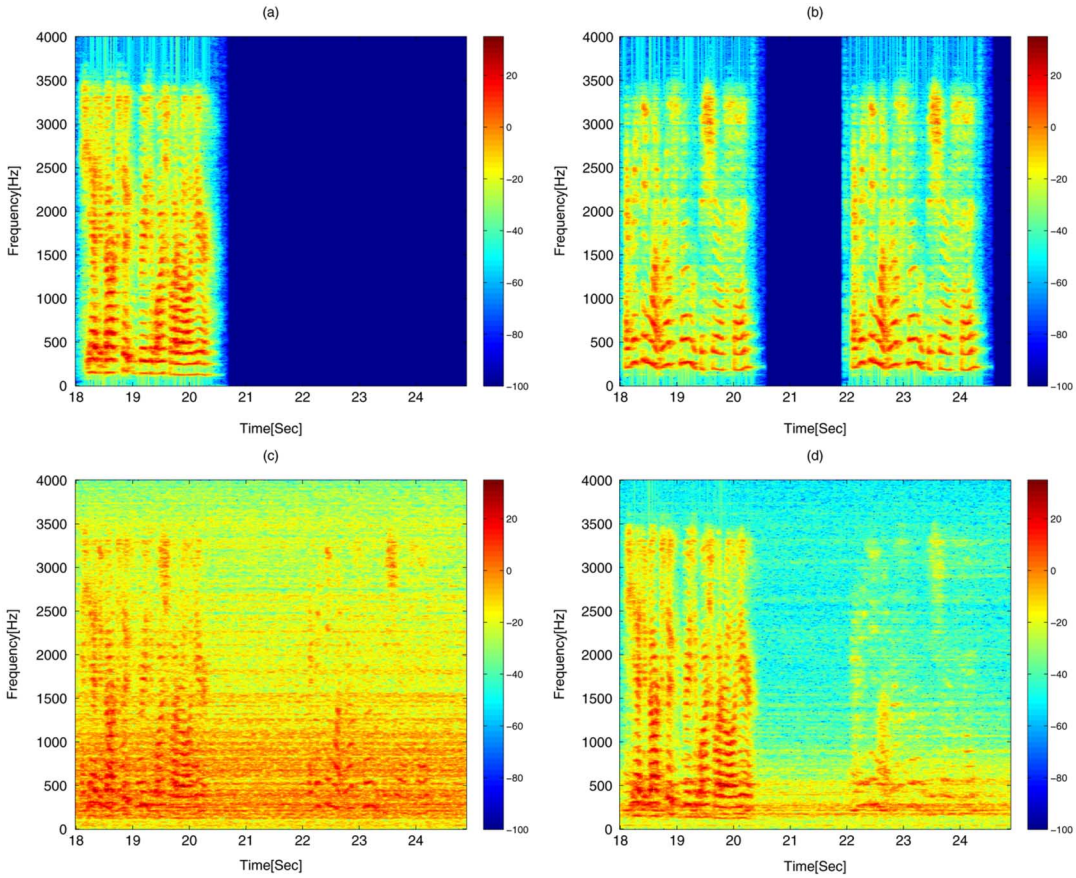

Fig. 7. Sonograms: (a) Desired speech signal at microphone #1. (b) Nonstationary interference signal at microphone #1. (c) Noisy signal at microphone #1. (d) Enhanced signal.

The RIRs were simulated with a modified version [29] of Allen and Berkley's *image method* [30]. The reverberation time was set to $T_{60} = 300$ ms. In Fig. 5, the impulse response relating the desired source and the first microphone, and its respective frequency response, are depicted. The lengths of the filters in the MBF, the BM, and the ANC were set to 250, 250, and 750 taps, respectively. Segments of 2048 samples were used to implement the *overlap and save* procedure. The sampling frequency was set to 8 kHz. The desired speech to competing speech [denoted signal-to-interference ratio (SIR)] was set to 5 dB. The desired speech to the stationary noise (denoted SNR) was set to 5 dB as well.

For comparison, we applied the TF-GSC [14] to the same signals. For the TF-GSC, the lengths of the filters in the MBF, the BM, and ANC were set to 250, 250, and 550 taps, respectively. Recall, that the ANC in the TF-GSC is comprised of $M - 1 = 7$ channels, while the dimension of respective block in the DTF-GSC is only $M - 2 = 6$. Hence, the computational burden imposed by both algorithms is comparable.

For evaluating and comparing the performance of the two beamformers, we applied the algorithms in two phases. In the first phase, the beamformers were applied to an input signal, comprised of the sum of the desired speech, the competing speech, and the stationary noise (with gains in accordance with the respective SNR and SIR). In this phase, the beamformers were allowed to adapt yielding $Y(t, e^{j\omega})$, the actual algorithm output. In the second phase, the beamformers were *not* allowed to adapt. Instead, a copy of the coefficients, obtained in the first phase, was used. As the beamformers coefficients are time varying, we used in each time instant the corresponding copy of the coefficients. This procedure enabled the demonstration of the algorithm convergence rate. Each beamformer was applied three times in the second phase, yielding three different signals. Denote the beamformer response to the desired signal $\mathbf{A}(e^{j\omega})s_1(t, e^{j\omega})$ alone as $Y_{s_1}(t, e^{j\omega})$, the response to the competing speech $\mathbf{B}(e^{j\omega})s_2(t, e^{j\omega})$ alone as $Y_{s_2}(t, e^{j\omega})$, and the response to the stationary noise signals at the microphones $\mathbf{N}(e^{j\omega})$ as $Y_n(t, e^{j\omega})$. The entire test procedure is depicted in Fig. 6.

For both algorithms, we assumed the existence of a perfect voice activity detector (VAD), allowing for ANC adaptation only during nonactive periods of the desired speech. Applying a simple energy-based VAD might suffice in moderate or high SNR levels. As discussed in Section VI, the ANC block of the TF-GSC is more sensitive to *double talk* situations than its respective DTF-GSC block. Moreover, we showed in Section V a method for estimating the BM in *double talk* scenario. It should be noted, however, that the MBF can be updated only when one speech signal is active. Major errors in the VAD and the double talk detector (DTD) decisions might deteriorate the performance of both algorithms.

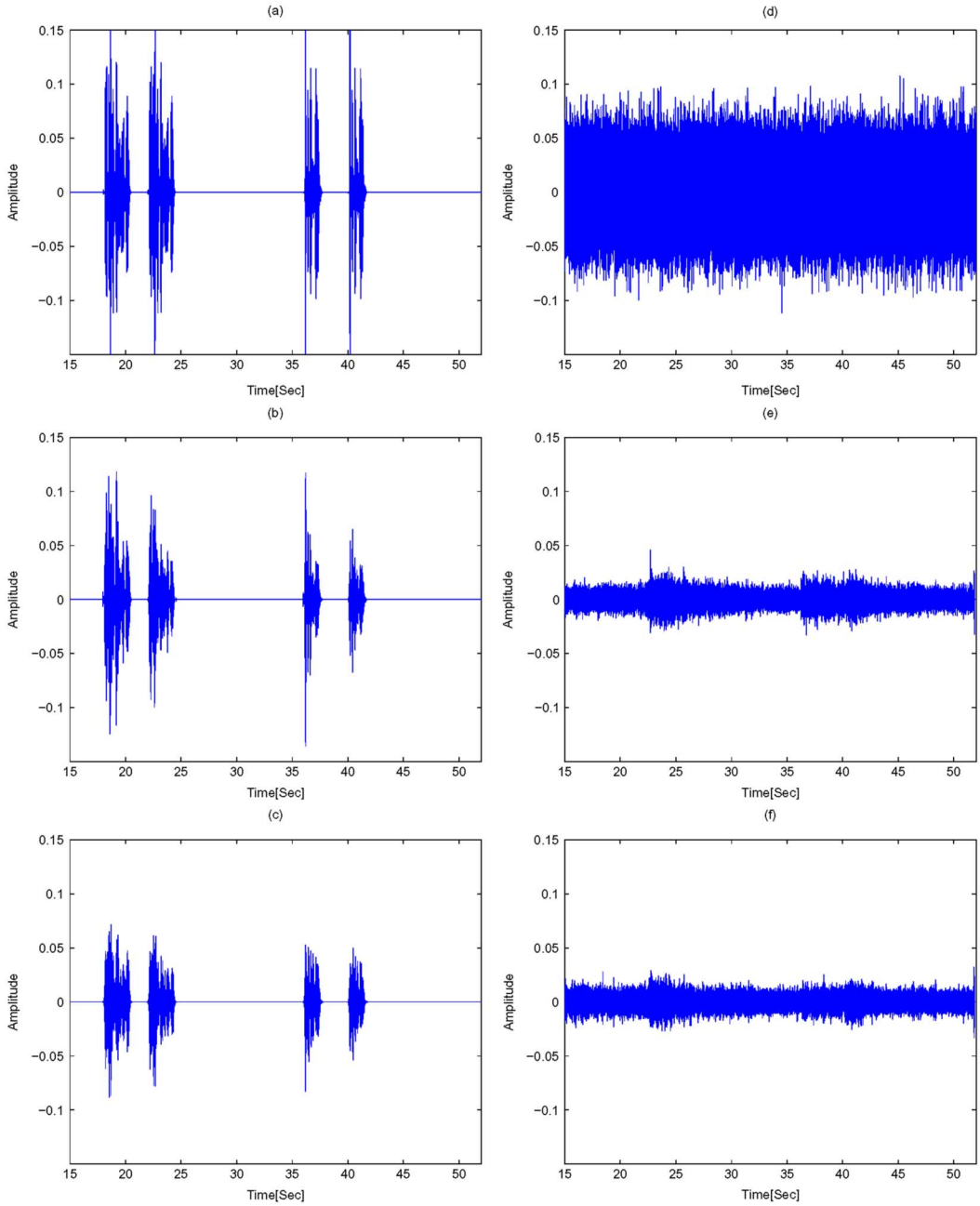

Fig. 8. Signal waveforms: (a) Interference signal at microphone #1. (b) Interference signal at the TF-GSC output. (c) Interference signal at the DTF-GSC output. (d) Noise signal at microphone #1. (e) Noise signal at the TF-GSC output. (f) Noise signal at the DTF-GSC output.

Finally, we draw the attention of the reader to a comprehensive theoretical performance analysis of the DTF-GSC, which can be found in [31], where we demonstrate the applicability of the DTF-GSC in some representative reverberant and non-reverberant environments under various noise field conditions. Three figures-of-merit are used for evaluating the algorithm performance, namely the PSD deviation imposed on the desired signal at the beamformer output, the achievable noise reduction, and the interference reduction.

B. Results

Fig. 7 shows sonograms of the desired signal, the interference signal, the noisy signal, all at microphone #1, and the DTF-GSC output. It can be seen that both noise and interference signals are well suppressed, especially in frequencies above 500 Hz. Moreover, no self-cancellation or other deviation can be noticed even during the *double talk* case.

In Fig. 8, a comparison between the TF-GSC and the DTF-GSC is presented. According to the outline of Fig. 6, each signal is analyzed separately. In the left column, the interference component at the input, at the TF-GSC output, and at the DTF-GSC output are presented. In the right column, the noise component at the respective signals are depicted. All segments depict the signal after the noise canceller converged. It is clearly seen that the interference signal and noise signal levels are significantly reduced for both algorithms. However,

TABLE I
NOISE AND INTERFERENCE REDUCTION FOR THE TF-GSC (TOP) AND THE TF-GSC (BOTTOM)

Algorithm	Input		FBF Output		Total Output	
	SIR	SNR	SIR	SNR	SIR	SNR
TF-GSC	5.1	4.7	10.7	4.6	10.7	14.3
DTF-GSC	5.1	4.7	15.7	4.4	15.7	17.0

while the TF-GSC is adaptively building the null towards the interference signal, the amount of interference reduction of the DTF-GSC is kept high for the entire waveform. The noise level at the output of both algorithms is comparable. However, while the noise signal at the TF-GSC output is fluctuating according to the interferer activity, it is kept relatively constant at the DTF-GSC output.

The SNR and SIR improvement for different stages of the GSC structure is given in Table I, for both the TF-GSC and DTF-GSC algorithms. A clear advantage of the DTF-GSC algorithm over the TF-GSC algorithm is evident.

Informal listening tests confirm that the perceptual quality of the desired speech signal (for the directional noise field case) is retained in the enhanced signal, while the stationary and nonstationary interferences are well suppressed (audio sample files are available online.[5]) The low PSD deviation imposed on the speech component at the system's output confirms the theoretical analysis presented in [31].

VIII. CONCLUSION

The TF-GSC interference reduction severely deteriorates in the presence of a second directional, nonstationary, interference signal. Theoretically, the ANC block should have been able to eliminate the interference speech signal component at the FBF output. However, the existence of nonstationary signals at its inputs significantly impairs its ability to converge. Moreover, the null towards the competing signal is adaptively built. This adaptation causes both a time-varying attenuation of the interference signal, and fluctuating residual noise level. These drawbacks of the TF-GSC algorithm limit its use in cases where the desired signal is contaminated by both stationary and nonstationary interfering signals. The proposed DTF-GSC structure avoids these difficulties. The nonstationary interference is treated separately by exploiting the speech characteristics. The FBF and BM are correspondingly updated to incorporate the extended set of constraints on the beampattern.

While the RTF estimation procedure is adopted from the TF-GSC when only one speech signal is active, a new system identification procedure for *double talk* segments was derived, which enables direct estimate of the BM components. Experimental results demonstrate the advantage of the novel DTF-GSC structure over the conventional TF-GSC structure in cases where both stationary and nonstationary interference signals are present at the beamformer inputs.

[5][Online]. Available: http://www.eng.biu.ac.il/gannot.

APPENDIX A
PROOF OF (10)

Imposing the constraints defined in (6) on (9) yields

$$\begin{aligned}-\mathbf{A}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\left[\lambda_1\mathbf{A}(e^{j\omega})+\lambda_2\mathbf{B}(e^{j\omega})\right]&=\mathcal{F}(e^{j\omega})\\ -\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\left[\lambda_1\mathbf{A}(e^{j\omega})+\lambda_2\mathbf{B}(e^{j\omega})\right]&=0.\end{aligned}$$

Solving for the Lagrange multipliers yields

$$\begin{bmatrix}\lambda_1\\ \lambda_2\end{bmatrix}=C^{-1}\begin{bmatrix}-\mathcal{F}(e^{j\omega})\\ 0\end{bmatrix}$$

where C is defined as

$$C\equiv\begin{bmatrix}\mathbf{A}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{A}(e^{j\omega}) & \mathbf{A}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{B}(e^{j\omega})\\ \mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{A}(e^{j\omega}) & \mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{B}(e^{j\omega})\end{bmatrix}.$$

The solution for the linear equations is then given by

$$\begin{aligned}\lambda_1&=-\alpha^{-1}\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{B}(e^{j\omega})\mathcal{F}(e^{j\omega})\\ \lambda_2&=\alpha^{-1}\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{A}(e^{j\omega})\mathcal{F}(e^{j\omega})\end{aligned}$$

where

$$\begin{aligned}\alpha\equiv\ &\mathbf{A}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\\ &\cdot[\mathbf{A}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{B}(e^{j\omega})\\ &-\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{A}(e^{j\omega})].\end{aligned}$$

Therefore, the optimal solution is given in the first equation at the bottom of the page. Dividing the nominator and denominator of last term by

$$(\mathbf{A}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{A}(e^{j\omega}))(\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{B}(e^{j\omega}))$$

yields the second equation at the bottom of the next page. Rearranging terms and using definitions (11) and (12) yields (10).

$$\mathbf{W}^{\mathrm{opt}}(t,e^{j\omega})=-\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\left(\lambda_1\mathbf{A}(e^{j\omega})+\lambda_2\mathbf{B}(e^{j\omega})\right)=\frac{\mathcal{F}(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})}{\mathbf{A}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})}\cdot\frac{\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{A}(e^{j\omega})-\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{A}(e^{j\omega})\mathbf{B}(e^{j\omega})}{[\mathbf{A}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{B}(e^{j\omega})-\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{A}(e^{j\omega})]}$$

APPENDIX B
PROOF OF (14)

Imposing our constraints on (13) yields the following linear equations:

$$\begin{aligned}&\mathbf{A}^H(e^{j\omega})[\mathbf{W}(t,e^{j\omega}) - \mu(\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\\&\quad + \lambda_1\mathbf{A}(e^{j\omega}) + \lambda_2\mathbf{B}(e^{j\omega}))] = \mathcal{F}(e^{j\omega})\\&\mathbf{B}^H(e^{j\omega})[\mathbf{W}(t,e^{j\omega}) - \mu(\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\\&\quad + \lambda_1\mathbf{A}(e^{j\omega}) + \lambda_2\mathbf{B}(e^{j\omega}))] = 0.\end{aligned} \tag{55}$$

Rearranging terms in (55) yields

$$\begin{aligned}&\lambda_1\mu\|\mathbf{A}(e^{j\omega})\|^2 + \lambda_2\mu\mathbf{A}^H(e^{j\omega})\mathbf{B}(e^{j\omega})\\&\quad = \mathbf{A}^H(e^{j\omega})\mathbf{W}(t,e^{j\omega})\\&\qquad - \mu\mathbf{A}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mathcal{F}(e^{j\omega})\end{aligned}$$

and

$$\begin{aligned}&\lambda_1\mu\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega}) + \lambda_2\mu\|\mathbf{B}(e^{j\omega})\|^2\\&\quad = \mathbf{B}^H(e^{j\omega})\mathbf{W}(t,e^{j\omega})\\&\qquad - \mu\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\end{aligned}$$

and therefore yields (56), shown at the bottom of the page.

Solving the linear equations yields (57), shown at the bottom of the page. Define

$$\begin{aligned}\alpha &\equiv \|\mathbf{A}(e^{j\omega})\|^2\|\mathbf{B}(e^{j\omega})\|^2\\&\quad - \mathbf{A}^H(e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega})\\&= (1 - \|\rho(e^{j\omega})\|^2)\|\mathbf{A}(e^{j\omega})\|^2\|\mathbf{B}(e^{j\omega})\|^2\end{aligned} \tag{58}$$

then

$$\begin{aligned}\lambda_1 &= (\alpha\mu)^{-1}\\&\quad\cdot[\|\mathbf{B}(e^{j\omega})\|^2(\mathbf{A}^H(e^{j\omega})\mathbf{W}(t,e^{j\omega})\\&\quad - \mu\mathbf{A}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mathcal{F}(e^{j\omega}))\\&\quad - \mathbf{A}^H(e^{j\omega})\mathbf{B}(e^{j\omega})(\mathbf{B}^H(e^{j\omega})\mathbf{W}(t,e^{j\omega})\\&\quad - \mu\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega}))]\end{aligned} \tag{59}$$

$$\begin{aligned}\lambda_2 &= (\alpha\mu)^{-1}\\&\quad\cdot[-\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega})(\mathbf{A}^H(e^{j\omega})\mathbf{W}(t,e^{j\omega})\\&\quad - \mu\mathbf{A}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mathcal{F}(e^{j\omega}))\\&\quad + \|\mathbf{A}(e^{j\omega})\|^2(\mathbf{B}^H(e^{j\omega})\mathbf{W}(t,e^{j\omega})\\&\quad - \mu\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega}))].\end{aligned} \tag{60}$$

The following expression is utilized in calculating $\mathbf{W}(t+1,e^{j\omega})$:

$$\begin{aligned}&\mu\left[\lambda_1\mathbf{A}(e^{j\omega}) + \lambda_2\mathbf{B}(e^{j\omega})\right] = \alpha\left\|\mathbf{B}(e^{j\omega})\right\|^2\\&\quad\cdot[\mathbf{A}(e^{j\omega})\mathbf{A}^H(e^{j\omega})\mathbf{W}(t,e^{j\omega})\\&\quad - \mu\mathbf{A}(e^{j\omega})\mathbf{A}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\\&\quad - \mathbf{A}(e^{j\omega})\mathcal{F}(e^{j\omega})]\\&\quad - \alpha[\mathbf{A}(e^{j\omega})\mathbf{A}^H(e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\mathbf{W}(t,e^{j\omega})\\&\quad - \mu\mathbf{A}(e^{j\omega})\mathbf{A}^H(e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{B}^H\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})]\\&\quad - \alpha\left[\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega})\mathbf{A}^H(e^{j\omega})\mathbf{W}(t,e^{j\omega})\right.\\&\quad - \mu\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega})\\&\quad\left.\times\,\mathbf{A}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\right]\\&\quad + \alpha[\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega})\mathcal{F}(e^{j\omega})\\&\quad + \|\mathbf{A}(e^{j\omega})\|^2(\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\mathbf{W}(t,e^{j\omega})\\&\quad - \mu\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega}))].\end{aligned} \tag{61}$$

$$\mathbf{W}^{\mathrm{opt}}(t,e^{j\omega}) = \mathcal{F}(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\cdot\frac{\frac{\mathbf{A}(e^{j\omega})}{\mathbf{A}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{A}(e^{j\omega})} - \frac{\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{A}(e^{j\omega})\mathbf{B}(e^{j\omega})}{(\mathbf{A}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{A}(e^{j\omega}))(\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{B}(e^{j\omega}))}}{1 - \frac{\left|\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{A}(e^{j\omega})\right|^2}{(\mathbf{A}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{A}(e^{j\omega}))(\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}^{-1}(t,e^{j\omega})\mathbf{B}(e^{j\omega}))}}$$

$$\begin{bmatrix}\|\mathbf{A}(e^{j\omega})\|^2 & \mathbf{A}^H(e^{j\omega})\mathbf{B}(e^{j\omega})\\ \mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega}) & \|\mathbf{B}(e^{j\omega})\|^2\end{bmatrix}\begin{bmatrix}\lambda_1\\ \lambda_2\end{bmatrix} = \frac{1}{\mu}\begin{bmatrix}\mathbf{A}^H(e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mu\mathbf{A}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mathcal{F}(e^{j\omega})\\ \mathbf{B}^H(e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mu\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\end{bmatrix} \tag{56}$$

$$\begin{aligned}\begin{bmatrix}\lambda_1\\ \lambda_2\end{bmatrix} &= \frac{1}{\mu}\frac{\begin{bmatrix}\|\mathbf{B}(e^{j\omega})\|^2 & -\mathbf{A}^H(e^{j\omega})\mathbf{B}(e^{j\omega})\\ -\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega}) & \|\mathbf{A}(e^{j\omega})\|^2\end{bmatrix}}{\|\mathbf{A}(e^{j\omega})\|^2\|\mathbf{B}(e^{j\omega})\|^2 - \mathbf{A}^H(e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega})}\\&\quad\times\begin{bmatrix}\mathbf{A}^H(e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mu\mathbf{A}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mathcal{F}(e^{j\omega})\\ \mathbf{B}^H(e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mu\mathbf{B}^H(e^{j\omega})\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\end{bmatrix}\end{aligned} \tag{57}$$

Rearranging terms in (61) yields

$$\begin{aligned}
&\mu[\lambda_1\mathbf{A}(e^{j\omega})+\lambda_2\mathbf{B}(e^{j\omega})]\\
&\quad=\alpha[\mathbf{A}(e^{j\omega})\mathbf{A}^H(e^{j\omega})\\
&\qquad\times(\|\mathbf{B}(e^{j\omega})\|^2\mathbf{I}-\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega}))\\
&\qquad-\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\\
&\qquad\times(\mathbf{A}(e^{j\omega})\mathbf{A}^H(e^{j\omega})-\|\mathbf{A}(e^{j\omega})\|^2\mathbf{I})]\mathbf{W}(t,e^{j\omega})\\
&\qquad-\alpha\mu[\mathbf{A}(e^{j\omega})\mathbf{A}^H(e^{j\omega})\\
&\qquad\times(\|\mathbf{B}(e^{j\omega})\|^2\mathbf{I}-\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega}))\\
&\qquad-\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})(\mathbf{A}(e^{j\omega})\mathbf{A}^H(e^{j\omega})\\
&\qquad-\|\mathbf{A}(e^{j\omega})\|^2\mathbf{I})]\cdot\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\\
&\qquad+\alpha[-\|\mathbf{B}(e^{j\omega})\|^2\mathbf{A}(e^{j\omega})\\
&\qquad+\mathbf{B}(e^{j\omega})\mathbf{B}^H(e^{j\omega})\mathbf{A}(e^{j\omega})]\mathcal{F}(e^{j\omega}). \qquad (62)
\end{aligned}$$

Substituting (62) into (13), we obtain (14).

ACKNOWLEDGMENT

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

REFERENCES

[1] B. V. Veen and K. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Acoust., Speech, Signal Process. Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[2] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 10, pp. 1365–1376, Oct. 1987.

[3] O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[4] M. Er and A. Cantoni, "Derivative constraints for broadband element space antenna array processors," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, no. 6, pp. 1378–1393, Dec. 1983.

[5] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol. 30, no. 1, pp. 27–34, Jan. 1982.

[6] C. Jim, "A comparison of two LMS constrained optimal array structures," *Proc. IEEE*, vol. 65, no. 12, pp. 1730–1731, Dec. 1977.

[7] K. Buckley, "Braodband beamformingand the generalized sidelobe canceller," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 6, pp. 1322–1323, Oct. 1986.

[8] K. Buckley, "A short proof of the equivalence of LCMV and GSC beamforming," *IEEE Signal Process. Lett.*, vol. 9, no. 6, pp. 168–169, Jun. 2002.

[9] S. Affes 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, Sep. 1997.

[10] S. Nordholm, I. Claesson, and P. Eriksson, "The broadband Wiener solution for Griffiths–Jim beamformers," *IEEE Trans. Signal Process.*, vol. 40, no. 2, pp. 474–478, Feb. 1992.

[11] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2677–2684, Oct. 1999.

[12] J. Bitzer, K. U. Simmer, and K. D. Kammeyer, "Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement," in *Proc. 24th IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Phoenix, AZ, Mar. 15–19, 1999, pp. 2965–2968.

[13] S. Affes and Y. Grenier, "A source subspace tracking array of microphones for double talk situations," in *Proc. 22nd IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Munich, Germany, Apr. 20–24, 1997, pp. 269–272.

[14] 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.

[15] J. Benesty, J. Chen, and Y. Huang, "On microphone-array beamforming from a MIMO acoustic signal processing perspective," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 3, pp. 1053–1065, Mar. 2007.

[16] G. Reuven, S. Gannot, and I. Cohen, "Dual source TF-GSC and its application to echo cancellation," in *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, Eindhoven, The Netherlands, Sep. 2005, pp. 27–30.

[17] Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time fourier transform domain," *IEEE Signal Process. Lett.*, vol. 14, no. 5, pp. 337–340, May 2007.

[18] D. Brandwood, "A complex gradient operator and its application in adaptive array theory," *Inst. Elect. Eng. Proc.*, vol. 130, no. 1, pt. F and H, pp. 11–16, Feb. 1983.

[19] G. Strang, *Linear Algebra and Its Application*, 2nd ed. New York: Academic, 1980.

[20] B. Widrow, J. G. , Jr., J. McCool, J. Kaunitz, C. Williams, R. Hearn, J. Zeider, E. D. , Jr., and R. Goodlin, "Adaptive noise cancelling: Principals and applications," *Proc. IEEE*, vol. 63, no. 12, pp. 1692–1716, Dec. 1975.

[21] J. Bitzer, K.-D. Kammeyer, and K. Simmer, "An alternative implementation of the superdirective beamformer," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust.*, New Paltz, NY, Oct. 1999, pp. 7–10.

[22] S. Haykin, *Adaptive Filter Theory*, ser. Information and System Sciences, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2002.

[23] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 2, pp. 443–445, Apr. 1985.

[24] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Process.*, vol. 81, no. 11, pp. 2403–2418, Oct. 2001.

[25] K. U. Simmer, J. Bitzer, and C. Marro, *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Germany: Spfiriger-Verlag, 2001, ch. Post-Filtering Techniques, pp. 39–60.

[26] S. Gannot, D. Burshtein, and E. Weinstein, "Analysis of the power spectral deviation of the general transfer function GSC," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1115–1121, Apr. 2004.

[27] J. S. Garofolo, "Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database," National Inst. Stand. Technol. (NIST), Gaithersburg, MD, 1988, Tech. Rep., prototype as of December 1988.

[28] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Commun.*, vol. 12, no. 3, pp. 247–251, Jul. 1993.

[29] E. Habets, "Room impulse response (RIR) generator," Jul. 2006 [Online]. Available: http://home.tiscali.nl/ehabets/rir_generator.html.

[30] 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.

[31] G. Reuven, S. Gannot, and I. Cohen, "Performance analysis of dual source transfer-function generalized sidelobe canceller," *Speech Commun.*, vol. 49, pp. 602–622, 2007.

Gal Reuven received the B.Sc. degree from Tel-Aviv University, Tel-Aviv, Israel, and the M.Sc. degree from the Technion—Israel Institute of Technology, Haifa, Israel, both in electrical engineering, in 1998 and 2006, respectively.

From 1998 to 2004, he served as a DSP engineer and system engineer in an R&D center of the Israel Defense Forces. Since 2005, he has been a senior DSP and real-time embedded engineer in Dune Medical Devices, Ltd., Caesarea, Israel. His research interests include statistical signal processing and speech enhancement using microphone arrays.

Sharon Gannot (S'92–M'01–SM'06) received the B.Sc. degree (summa cum laude) from the Technion—Israel Institute of Technology, Haifa, Israel, in 1986 and the M.Sc. (cum laude) and Ph.D. degrees from Tel-Aviv University, Israel, in 1995 and 2000, respectively, all in electrical engineering.

In 2001, he held a postdoctoral position at the Department of Electrical Engineering (SISTA), Katholieke Universiteit Leuven, Leuven, Belgium. From 2002 to 2003, he held a research and teaching position at the Faculty of Electrical Engineering, Technion-Israel Institute of Technology, Haifa, Israel. Currently, he is a Senior Lecturer at the School of Engineering, Bar-Ilan University, Ramat-Gan, Israel. He is an Associate Editor of the *EURASIP Journal of Applied Signal Processing*, an Editor of a special issue on Advances in Multimicrophone Speech Processing of the same journal, a Guest Editor of *Elsevier Speech Communication Journal* and a reviewer of many IEEE journals and conferences. His research interests include parameter estimation, statistical signal processing, and speech processing using either single- or multimicrophone arrays.

Dr. Gannot has been a member of the Technical and Steering Committee of the International Workshop on Acoustic Echo and Noise Control (IWAENC) since 2005.

Israel Cohen (M'01–SM'03) received the B.Sc. (summa cum laude), M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, Haifa, Israel, in 1990, 1993, and 1998, respectively.

From 1990 to 1998, he was a Research Scientist with RAFAEL Research Laboratories, Haifa, Israel Ministry of Defense. From 1998 to 2001, he was a Postdoctoral Research Associate with the Computer Science Department, Yale University, New Haven, CT. In 2001, he joined the Electrical Engineering Department of the Technion, where he is currently an Associate Professor. His research interests are statistical signal processing, analysis and modeling of acoustic signals, speech enhancement, noise estimation, microphone arrays, source localization, blind source separation, system identification, and adaptive filtering. He is a coeditor of the Multichannel Speech Processing section of the *Springer Handbook of Speech Processing* and as Guest Editor of a special issue of the *EURASIP Journal on Advances in Signal Processing* on Advances in Multimicrophone Speech Processing and a special issue of the *EURASIP Speech Communication Journal* on Speech Enhancement.

Dr. Cohen received in 2005 and 2006 the Technion Excellent Lecturer awards. He serves as Associate Editor of the IEEE SIGNAL PROCESSING LETTERS and served as Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING.