2012 IEEE 27-th Convention of Electrical and Electronics Engineers in Israel

A Weighted Multichannel Wiener Filter for Multiple Sources Scenarios

Shmulik Markovich-Golan
Faculty of Engineering
Bar-Ilan University
Ramat-Gan, 52900, Israel
shmulik.markovich@gmail.com

Sharon Gannot
Faculty of Engineering
Bar-Ilan University
Ramat-Gan, 52900, Israel
sharon.gannot@biu.ac.il

Israel Cohen
Department of Electrical Engineering
Technion – Israel Institute of Technology
Technion City, Haifa 32000, Israel
icohen@ee.technion.ac.il

***Abstract*—The scenario of P speakers received by an M microphone array in a reverberant enclosure is considered. We extend the single source speech distortion weighted multichannel Wiener filter (SDW-MWF) to deal with multiple speakers. The mean squared error (MSE) is extended by introducing P weights, each controlling the distortion of one of the sources. The P weights enable further control in the design of the beamformer (BF). Two special cases of the proposed BF are the SDW-MWF and the linearly constrained minimum variance (LCMV)-BF. We provide a theoretical analysis for the performance of the proposed BF. Finally, we exemplify the ability of the proposed method to control the tradeoff between noise reduction (NR) and distortion levels of various speakers in an experimental study.**

I. Introduction

Beamforming, by utilizing the spatial diversity, extends the classic time-frequency filtering, and allows to cope with complicated scenarios of multiple speakers and interferences. Considering the single desired speaker scenario, several beamformer (BF) design criteria for optimizing the performance exist. The minimum mean square error (MMSE) beamformer, also known as the multichannel Wiener filter (MWF) [1] minimizes the variance of the error between the output and the desired signal. The minimum variance distortionless response (MVDR) beamformer [1] minimizes the noise power at the output while maintaining the desired signal undistorted. The speech distortion weighted multichannel Wiener filter (SDW-MWF) [2][3], generalizes both criteria. By identifying the two sources of error as distortion and residual noise, and weighting the residual noise component in the MMSE minimization by a factor μ, it is possible to control the tradeoff between the two error sources. By setting $\mu = 1$ or $\mu = 0$, the MWF and the MVDR are obtained as special cases of the SDW-MWF, respectively. Doclo et al. [4] show that the SDW-MWF is equivalent to the MVDR followed by a single channel SDW-MWF post-filter.

In more complicated scenarios, where several speakers exist, and more control over the beampattern is required, the linearly constrained minimum variance (LCMV) beamformer [5], which is an extension of the MVDR to multiple constraints, is an appropriate solution. Strictly maintaining the constraints set increases the noise power at the output of the LCMV, compared with the MWF. In a recent contribution, Habets and Benesty [6], applied the LCMV-BF in a scenario with a single desired speaker and an interference with a general covariance matrix. They suggested using two constraints, one for maintaining the desired speaker undistorted and the second for suppressing one of the coherent noise components. Furthermore, they showed that by a properly designing the desired response of each signal of interest (SOI), the tradeoff between distortion of the desired speaker and suppression of the coherent noise as well as the residual noise can be controlled.

In the current contribution, an extension of the SDW-MWF for the case of multiple sources is derived. We identify the various sources of error at the output as residual noise and distortion components. For each source, the distortion is defined as the variance of the error between the desired and actual responses. We propose to apply individual weights to each of the distortion components. We prove that the LCMV-BF is a special case of the proposed beamformer, denoted as multiple speech distortions weighted multichannel Wiener filter (MSDW-MWF).

II. Problem formulation

The problem is formulated in the short time Fourier transform (STFT) domain, where ℓ and k are time-frame and frequency bin indices, respectively. Consider a microphone array located in a reverberant enclosure. The signals received by the microphone array are categorized in two groups. The first group comprises sources for which a desired response is designated. A source belonging to this group is denoted SOI. The second group comprises interferences that we wish to mitigate. Consider P coherent SOIs, denoted $s_1(\ell,k), .., s_P(\ell,k)$. Denote by $\mathbf{h}_p(\ell,k)$, for $p = 1,..,P$, the acoustic transfer function (ATF) relating the pth SOI and the microphone signals. The received microphone signals are given by:

$$\mathbf{z}(\ell,k) \triangleq \mathbf{H}(\ell,k)\,\mathbf{s}(\ell,k) + \mathbf{v}(\ell,k) \tag{1}$$

where $\mathbf{s}(\ell,k) \triangleq \begin{bmatrix} s_1(\ell,k) & \cdots & s_P(\ell,k) \end{bmatrix}^T$ is a vector comprising all the SOIs, $\mathbf{H}(\ell,k) \triangleq \begin{bmatrix} \mathbf{h}_1(\ell,k) & \cdots & \mathbf{h}_P(\ell,k) \end{bmatrix}$ is an $M \times P$ matrix of the ATFs relating the SOIs and the microphones and $\mathbf{v}(\ell,k)$ denotes the received interferences. Next, we define the covariance matrices of the SOIs and interfering signals as $\mathbf{\Phi}_{ss}(\ell,k) \triangleq \mathrm{E}\left\{\mathbf{s}(\ell,k)\,\mathbf{s}^\dagger(\ell,k)\right\}$ and $\mathbf{\Phi}_{vv}(\ell,k) \triangleq \mathrm{E}\left\{\mathbf{v}(\ell,k)\,\mathbf{v}^\dagger(\ell,k)\right\}$, respectively. For brevity, hereafter the frequency bin index k is omitted and all derivations are valid for all $k = 1,..,K$ frequency bins. Moreover, we omit the frame index from $\mathbf{H}$, $\mathbf{\Phi}_{ss}$ and $\mathbf{\Phi}_{vv}$.

The covariance matrix of the received signals is given by:

$$\mathbf{\Phi}_{zz} \triangleq \mathbf{H}\mathbf{\Phi}_{ss}\mathbf{H}^{\dagger} + \mathbf{\Phi}_{vv}. \tag{2}$$

The desired response vector is denoted by $\mathbf{g}$ and the desired signal at the output of the BF is defined as:

$$d(\ell) \triangleq \mathbf{g}^{\dagger}\mathbf{s}(\ell). \tag{3}$$

The output of a BF $\mathbf{w}$ is denoted by:

$$y(\ell) = \mathbf{w}^{\dagger}\mathbf{z}(\ell) \tag{4}$$

and the MSE between the desired signal (3) and the BF's output (4) is:

$$J_w \triangleq \mathrm{E}\left\{|d(\ell) - y(\ell)|^2\right\}. \tag{5}$$

In the following section, we present the proposed algorithm.

III. Multiple speech distortions weighted multichannel Wiener filter

Substituting (1),(3),(4) in (5) and noting that $\mathbf{s}(\ell)$ and $\mathbf{v}(\ell)$ are statistically independent signals yields:

$$J_w = \left(\mathbf{g} - \mathbf{H}^{\dagger}\mathbf{w}\right)^{\dagger}\mathbf{\Phi}_{ss}\left(\mathbf{g} - \mathbf{H}^{\dagger}\mathbf{w}\right) + \mathbf{w}^{\dagger}\mathbf{\Phi}_{vv}\mathbf{w}. \tag{6}$$

We denote the component $\left(\mathbf{g} - \mathbf{H}^{\dagger}\mathbf{w}\right)^{\dagger}\mathbf{\Phi}_{ss}\left(\mathbf{g} - \mathbf{H}^{\dagger}\mathbf{w}\right)$ as the total distortion and the component $\mathbf{w}^{\dagger}\mathbf{\Phi}_{vv}\mathbf{w}$ as the residual noise. The SDW-MWF criterion introduces the parameter μ which controls the tradeoff between the total distortion and the noise reduction:

$$\begin{aligned} J_{\text{SDW-MWF}} \triangleq \min_{\mathbf{w}'} & \left(\mathbf{g} - \mathbf{H}^{\dagger}\mathbf{w}'\right)^{\dagger}\mathbf{\Phi}_{ss}\left(\mathbf{g} - \mathbf{H}^{\dagger}\mathbf{w}'\right) \\ & + \mu\left(\mathbf{w}'\right)^{\dagger}\mathbf{\Phi}_{vv}\mathbf{w}'. \end{aligned} \tag{7}$$

In the current contribution we propose to utilize individual parameters, one for each source, for controlling the distortion of each of the sources separately. Explicitly, the proposed MSE criterion is given by extending (7):

$$\begin{aligned} J_{\text{MSDW-MWF}} \triangleq \min_{\mathbf{w}'} & \left(\mathbf{g} - \mathbf{H}^{\dagger}\mathbf{w}'\right)^{\dagger}\mathbf{\Lambda}\mathbf{\Phi}_{ss}\left(\mathbf{g} - \mathbf{H}^{\dagger}\mathbf{w}'\right) \\ & + \left(\mathbf{w}'\right)^{\dagger}\mathbf{\Phi}_{vv}\mathbf{w}' \end{aligned} \tag{8}$$

where $\mathbf{\Lambda} \triangleq \mathrm{diag}\{\lambda_1, .., \lambda_P\}$, a diagonal matrix with the parameters λ_p for $p = 1, .., P$ on its diagonal, and the BF which minimizes (8) is denoted $\mathbf{w}_{\text{MSDW-MWF}}$, the MSDW-MWF. The closed-form solution of (8) is given by:

$$\mathbf{w} = \left(\mathbf{H}\mathbf{\Lambda}\mathbf{\Phi}_{ss}\mathbf{H}^{\dagger} + \mathbf{\Phi}_{vv}\right)^{-1}\mathbf{H}\mathbf{\Lambda}\mathbf{\Phi}_{ss}\mathbf{g}. \tag{9}$$

Note, that for a single desired speaker scenario the SDW-MWF can be obtained as special case of the MSDW-MWF by setting:

$$\mathbf{\Lambda} = \mu^{-1}\mathbf{I}_{P\times P} \tag{10}$$

where $\mathbf{I}_{P\times P}$ is a $P \times P$ identity matrix. In the following sections we analyze the distortion of the SOIs and the noise level at the output of the proposed BF. In Sec. III-C we show that the well-known LCMV-BF is also a special case of the MSDW-MWF.

A. Distortion analysis

Two distortion figures of merit are analyzed. The first is the total distortion, defined as:

$$\begin{aligned} D_T \triangleq & \mathrm{E}\left\{|d(\ell) - y(\ell)|^2\right\} \\ = & \left(\mathbf{g} - \mathbf{H}^{\dagger}\mathbf{w}\right)^{\dagger}\mathbf{\Phi}_{ss}\left(\mathbf{g} - \mathbf{H}^{\dagger}\mathbf{w}\right) \\ = & \|\mathbf{\Phi}_{ss}^{1/2}\left(\mathbf{g} - \mathbf{H}^{\dagger}\mathbf{w}\right)\|^2 \end{aligned} \tag{11}$$

where $\mathbf{\Phi}_{ss} = \left(\mathbf{\Phi}_{ss}^{1/2}\right)^{\dagger}\mathbf{\Phi}_{ss}^{1/2}$ is the Cholesky decomposition. The second is the individual distortion of the pth source for $p = 1, .., P$:

$$\begin{aligned} D_p \triangleq & \mathrm{E}\left\{|g_p^* s_p(\ell) - \mathbf{w}^{\dagger}\mathbf{h}_p s_p(\ell)|^2\right\} \\ = & |g_p - \mathbf{h}_p^{\dagger}\mathbf{w}|^2\phi_{ss,p} \end{aligned} \tag{12}$$

where $\phi_{ss,p}$ is the variance of the pth source, and $\mathbf{\Phi}_{ss} \triangleq \mathrm{diag}\{\phi_{ss,1}, .., \phi_{ss,P}\}$. Note, that since $\mathbf{\Phi}_{ss}$ is diagonal, $\mathbf{\Phi}_{ss}^{1/2}$ is also diagonal and therefore $D_p = \left|\left(\mathbf{\Phi}_{ss}^{1/2}\left(\mathbf{g} - \mathbf{H}^{\dagger}\mathbf{w}\right)\right)_p\right|^2$, where $(\bullet)_p$ denotes the pth element of a vector.

Let

$$\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^{\dagger} \tag{13}$$

be the singular value decomposition (SVD) of $\mathbf{H}$. Substituting (13) in (9) yields:

$$\mathbf{w} = \mathbf{U}\left(\mathbf{S}\mathbf{V}^{\dagger}\mathbf{\Lambda}\mathbf{\Phi}_{ss}\mathbf{V}\mathbf{S}^{\dagger} + \mathbf{U}^{\dagger}\mathbf{\Phi}_{vv}\mathbf{U}\right)^{-1}\mathbf{S}\mathbf{V}^{\dagger}\mathbf{\Lambda}\mathbf{\Phi}_{ss}\mathbf{g}. \tag{14}$$

Note that $\mathbf{S}$ is an $M \times P$ matrix of the form:

$$\mathbf{S} \triangleq \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{0}_{(M-P)\times P} \end{bmatrix} \tag{15}$$

where $\mathbf{S}_1$ is a $P \times P$ diagonal real matrix. Hence, the expression $\mathbf{S}\mathbf{V}^{\dagger}\mathbf{\Lambda}\mathbf{\Phi}_{ss}\mathbf{V}\mathbf{S}^{\dagger}$ in (14) equals:

$$\mathbf{S}\mathbf{V}^{\dagger}\mathbf{\Lambda}\mathbf{\Phi}_{ss}\mathbf{V}\mathbf{S}^{\dagger} = \left[\begin{array}{c|c} \mathbf{S}_1\mathbf{V}^{\dagger}\mathbf{\Lambda}\mathbf{\Phi}_{ss}\mathbf{V}\mathbf{S}_1 & \mathbf{0}_{P\times(M-P)} \\ \hline \mathbf{0}_{(M-P)\times P} & \mathbf{0}_{(M-P)\times(M-P)} \end{array}\right]. \tag{16}$$

Let $\mathbf{U}_1$ be an $M \times P$ matrix comprising the first P columns of U which span the column-space of $\mathbf{H}$, and let $\mathbf{U}_0$ be an $M \times (M - P)$ matrix comprising of the last $M - P$ columns of $\mathbf{U}$ which span the null-space of $\mathbf{H}$. I.e.,

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_0 \end{bmatrix}. \tag{17}$$

By substituting (17) in the expression $\mathbf{U}^{\dagger}\mathbf{\Phi}_{vv}\mathbf{U}$ in (14), we obtain the following block-matrix structure:

$$\mathbf{U}^{\dagger}\mathbf{\Phi}_{vv}\mathbf{U} = \left[\begin{array}{c|c} \mathbf{\Gamma}_A & \mathbf{\Gamma}_B \\ \hline \mathbf{\Gamma}_B^{\dagger} & \mathbf{\Gamma}_C \end{array}\right] \tag{18}$$

where we define:

$$\mathbf{\Gamma}_A \triangleq \mathbf{U}_1^{\dagger}\mathbf{\Phi}_{vv}\mathbf{U}_1 \tag{19a}$$

$$\mathbf{\Gamma}_B \triangleq \mathbf{U}_1^{\dagger}\mathbf{\Phi}_{vv}\mathbf{U}_0 \tag{19b}$$

$$\mathbf{\Gamma}_C \triangleq \mathbf{U}_0^{\dagger}\mathbf{\Phi}_{vv}\mathbf{U}_0. \tag{19c}$$

Now, applying the block-matrix inversion formula to the sum of (16) and (18) and substituting in (14) yields the following simplified expression:

$$\mathbf{w} = \mathbf{\Psi}\left(\mathbf{I} + \mathbf{\Lambda}^{-1}\mathbf{\Phi}_{ss}^{-1}\mathbf{\Theta}\right)^{-1}\mathbf{g} \tag{20}$$

where

$$\mathbf{\Psi} \triangleq \left(\mathbf{U}_1 - \mathbf{U}_0\mathbf{\Gamma}_C^{-1}\mathbf{\Gamma}_B^\dagger\right)\mathbf{S}_1^{-1}\mathbf{V}^\dagger \tag{21a}$$

$$\mathbf{\Theta} \triangleq \mathbf{V}\mathbf{S}_1^{-1}\left(\mathbf{\Gamma}_A - \mathbf{\Gamma}_B\mathbf{\Gamma}_C^{-1}\mathbf{\Gamma}_B^\dagger\right)\mathbf{S}_1^{-1}\mathbf{V}^\dagger. \tag{21b}$$

A more simplified expression can be obtained for cases in which low distortion is required. In these cases $\|\mathbf{\Lambda}\| \gg 1$, hence, we can assume that $\|\mathbf{\Lambda}^{-1}\mathbf{\Phi}_{ss}^{-1}\mathbf{\Theta}\| \ll 1$, and replace $\left(\mathbf{I} + \mathbf{\Lambda}^{-1}\mathbf{\Phi}_{ss}^{-1}\mathbf{\Theta}\right)^{-1}$ in (20) with its first order Taylor series approximation:

$$\mathbf{w} \approx \mathbf{\Psi}\left(\mathbf{I} - \mathbf{\Lambda}^{-1}\mathbf{\Phi}_{ss}^{-1}\mathbf{\Theta}\right)\mathbf{g}. \tag{22}$$

Finally, the total distortion is obtained by substituting (20) in (11):

$$D_T = \|\mathbf{\Phi}_{ss}^{1/2}\left(\mathbf{I} - \left(\mathbf{I} + \mathbf{\Lambda}^{-1}\mathbf{\Phi}_{ss}^{-1}\mathbf{\Theta}\right)^{-1}\right)\mathbf{g}\|^2 \tag{23}$$

Applying to (23) a similar approximation as in (22) yields an approximated expression for low distortion:

$$D_T \approx \|\mathbf{\Lambda}^{-1}\mathbf{\Phi}_{ss}^{-1/2}\mathbf{\Theta}\mathbf{g}\|^2. \tag{24}$$

Considering (24) and the relation between D_p and D_T, the following approximation holds:

$$D_p \approx \frac{|\boldsymbol{\theta}_p^\dagger\mathbf{g}|^2}{\lambda_p^2\phi_{ss,p}} \tag{25}$$

where $\boldsymbol{\theta}_p$ is the pth column of the matrix $\mathbf{\Theta}$. Next, we define the various sources distortion measures. Define d_p as the distortion level of the pth source normalized by its power:

$$d_p \triangleq \frac{D_p}{\phi_{ss,p}}. \tag{26}$$

Define the set of desired distortion levels as $\dot{d}_p$ for $p = 1, .., P$. Given such a set of desired distortion levels, a BF which satisfies them and minimizes the noise level can be obtained by using the proposed MSDW-MWF (9) with a proper $\mathbf{\Lambda}$ matrix whose diagonal elements are given by:

$$\lambda_p = \frac{|\boldsymbol{\theta}_p^\dagger\mathbf{g}|}{\sqrt{\dot{d}_p\phi_{ss,p}}};\ p = 1, ..P. \tag{27}$$

B. Noise analysis

The noise level at the output of the MSDW-MWF is defined as:

$$\begin{aligned} N \triangleq& \mathbf{w}^\dagger\mathbf{\Phi}_{vv}\mathbf{w} \\ =& \|\mathbf{\Phi}_{vv}^{1/2}\mathbf{w}\|^2 \end{aligned} \tag{28}$$

where $\mathbf{\Phi}_{vv} = \left(\mathbf{\Phi}_{vv}^{1/2}\right)^\dagger\mathbf{\Phi}_{vv}^{1/2}$ is the Cholesky decomposition of the noise correlation matrix. Substituting (20) in (28) yields:

$$N = \|\mathbf{\Phi}_{vv}^{1/2}\mathbf{\Psi}\left(\mathbf{I} + \mathbf{\Lambda}^{-1}\mathbf{\Phi}_{ss}^{-1}\mathbf{\Theta}\right)^{-1}\mathbf{g}\|^2. \tag{29}$$

In case that a low distortion of the SOIs is required the following approximation can be obtained, by substituting (22) in (28):

$$N \approx \|\mathbf{\Phi}_{vv}^{1/2}\mathbf{\Psi}\left(\mathbf{I} - \mathbf{\Lambda}^{-1}\mathbf{\Phi}_{ss}^{-1}\mathbf{\Theta}\right)\mathbf{g}\|^2. \tag{30}$$

C. The LCMV-BF special case

In this section, we show that the LCMV-BF is a special case of the MSDW-MWF. Consider the MSDW-MWF formula in (9), and the following choice of $\mathbf{\Lambda} \triangleq \mu^{-1}\mathbf{\Phi}_{ss}^{-1}$. Substituting the latter choice of $\mathbf{\Lambda}$ in (9) yields:

$$\begin{aligned} \mathbf{w} &= \left(\mu^{-1}\mathbf{H}\mathbf{H}^\dagger + \mathbf{\Phi}_{vv}\right)^{-1}\mu^{-1}\mathbf{H}\mathbf{g} \\ &= \left(\mathbf{H}\mathbf{H}^\dagger + \mu\mathbf{\Phi}_{vv}\right)^{-1}\mathbf{H}\mathbf{g}. \end{aligned} \tag{31}$$

By applying the Woodbury identity to (31) and after some manipulation we obtain:

$$\begin{aligned} \mathbf{w} = \Bigg(&\mu^{-1}\mathbf{\Phi}_{vv}^{-1}\mathbf{H} - \mu^{-1}\mathbf{\Phi}_{vv}^{-1}\mathbf{H} \\ &\times\left(\mathbf{I} + \mu\left(\mathbf{H}^\dagger\mathbf{\Phi}_{vv}^{-1}\mathbf{H}\right)^{-1}\right)^{-1}\Bigg)\mathbf{g}. \end{aligned} \tag{32}$$

Assuming that μ is "small" such that $\|\mu\left(\mathbf{H}^\dagger\mathbf{\Phi}_{vv}^{-1}\mathbf{H}\right)^{-1}\| \ll 1$, we can replace $\left(\mathbf{I} + \mu\left(\mathbf{H}^\dagger\mathbf{\Phi}_{vv}^{-1}\mathbf{H}\right)^{-1}\right)^{-1}$ by its first order Taylor series expansion $\mathbf{I} - \mu\left(\mathbf{H}^\dagger\mathbf{\Phi}_{vv}^{-1}\mathbf{H}\right)^{-1}$. Finally, by substituting the latter approximation in (32), the LCMV-BF which satisfies the constraint set $\mathbf{H}^\dagger\mathbf{w} = \mathbf{g}$ is obtained:

$$\mathbf{w} \approx \mathbf{\Phi}_{vv}^{-1}\mathbf{H}\left(\mathbf{H}^\dagger\mathbf{\Phi}_{vv}^{-1}\mathbf{H}\right)^{-1}\mathbf{g}. \tag{33}$$

D. A modified MSDW-MWF

In practice, it is a cumbersome task to estimate the ATFs, $\mathbf{H}$, and the covariance matrix of the SOIs, $\mathbf{\Phi}_{ss}$. In this section, we obtain a modified MSDW-MWF which makes use of the relative transfer functions (RTFs) and the covariance matrix of SOIs as received by some reference microphone. Without loss of generality let us define the RTFs of the SOIs with respect to the first microphone. Define the RTF of the pth source by $\tilde{\mathbf{h}}_p = \frac{\mathbf{h}_p}{h_{p,1}}$, and the RTF matrix by:

$$\tilde{\mathbf{H}} \triangleq \begin{bmatrix} \tilde{\mathbf{h}}_1 & \cdots & \tilde{\mathbf{h}}_P \end{bmatrix}. \tag{34}$$

Next, we redefine the SOIs as their respective components in the first microphone, i.e., the pth modified SOI is given by $\tilde{s}_p(\ell) = h_{p,1}s_p(\ell)$, for $p = 1, \ldots, P$. The corresponding modified SOIs covariance matrix equals

$$\tilde{\mathbf{\Phi}}_{ss} \triangleq \operatorname{diag}\left\{|h_{1,1}|^2\phi_{ss,1}, \ldots, |h_{P,1}|^2\phi_{ss,P}\right\}. \tag{35}$$

Finally, substituting (34) and (35) in (9) yields the modified MSDW-MWF:

$$\tilde{\mathbf{w}} = \left(\tilde{\mathbf{H}}\mathbf{\Lambda}\tilde{\mathbf{\Phi}}_{ss}\tilde{\mathbf{H}}^\dagger + \mathbf{\Phi}_{vv}\right)^{-1}\tilde{\mathbf{H}}\mathbf{\Lambda}\tilde{\mathbf{\Phi}}_{ss}\mathbf{g}. \tag{36}$$

For estimating the RTFs we use a similar subspace based procedure as in [5], and for the estimation the SOIs covariance matrix, we use a spectral substraction technique as in [2].

IV. Experimental study

Here, we examine the MSDW-MWF in the case of a wideband stationary noise and two speech SOIs, a desired speaker and an interfering speaker. The dimensions of the simulated room are $4\text{m} \times 3\text{m} \times 3\text{m}$, the reverberation time is set to 0.3s, an array of 8 microphones is located next to one of the walls and the sampling rate is set to 8kHz. The input signal to interference ratio (SIR) and signal to noise ratio (SNR) levels are set to 0dB and 13dB, respectively, and the measured performance criteria are the output SIR, the NR, defined as the ratio $\frac{N}{\Phi_{vv}(1,1)}$ and the distortion levels of the SOIs d_1 and d_2. Note that the second SOI is an interfering source, which we would like to suppress. The distortion criterion (26) is suitable also for interfering sources. However, we would like to emphasize its meaning for this case. Usually, the desired response of an interfering SOI is zero, therefore the distortion is actually the power ratio of the interference at the output and the input. Explicitly, lower distortion means higher suppression. The performance is measured for various values of desired distortion levels, $\dot{d}_1$ and $\dot{d}_2$, in the range $[-25\text{dB}, -20\text{dB}, \ldots, 0\text{dB}]$. For each pair of $\dot{d}_1$, $\dot{d}_2$, the performance figures of merit are averaged over 20 Monte-Carlo experiments, in which the locations of sources are randomly selected. We use a window size of 4096 samples with 75% overlap. The distortions of the desired and interfering sources are depicted in Figs. 1,2, respectively. Clearly, from these figures, the MSDW-MWF allows for controlling individual distortion levels of the various SOIs. Note that the distortion level of the desired source, d_1 is lower bounded by -15dB, due to estimation errors of the RTF. As the approximated distortion levels (25) are valid for low distortion, the measured distortion on the interfering source in Fig. 2 differs from the desired one for higher levels of distortion $\dot{d}_1$. The NR versus the SOIs desired distortions is depicted in Fig. 3. This figure exemplifies that the NR can be controlled by sacrificing the distortion of just a sub-group of the SOIs. The average output SIR versus the desired SOIs distortion levels is depicted in Fig. 4. Note, that since the desired response of the desired source is 1, the variation in its output power is small for desired distortion levels $\dot{d}_1 \ll 0$dB. Therefore, as evidently seen in this figure, the output SIR is mainly determined by the desired distortion level of the interfering source.

V. Conclusion

We have considered the multiple SOIs in a noisy and reverberant environment scenario and extended the SDW-MWF for this case. The proposed method, denoted MSDW-MWF, allows for a better control of the tradeoff between NR and distortion levels of SOIs. We derive the SDW-MWF and the LCMV-BF as two special cases of the proposed method. We analyze the distortion levels of the various SOIs as well as the NR, and derive a more compact and simple approximation for the latter figures of merit, in the case of designing a low distortion MSDW-MWF. Finally, we exemplify the extended control over the NR versus distortion tradeoff in an experimental study.

References

[1] B. D. Van Veen and K. M. Buckley, “Beamforming: A versatile approach to spatial filtering,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[2] A. Spriet, M. Moonen, and J. Wouters, “Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction,” *Signal Processing*, vol. 84, no. 12, pp. 2367–2387, Dec. 2004.

[3] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, “Speech distortion weighted multichannel Wiener filtering techniques for noise reduction,” in *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds., pp. 199–228. Springer, 2005.

[4] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, “Acoustic beamforming for hearing aid applications,” in *Handbook on Array Processing and Sensor Networks*, S. Haykin and K. Ray Liu, Eds. Wiley, 2008.

[5] S. Markovich-Golan, S. Gannot, and I. Cohen, “Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals,” *IEEE Trans. Audio, Speech and Language Processing*, vol. 17, no. 6, pp. 1071–1086, Aug. 2009.

[6] E.A.P. Habets and J. Benesty, “A perspective on frequency-domain beamformers in room acoustics,” *IEEE Trans. Audio, Speech and Language Processing*, vol. 20, no. 3, pp. 947–960, Mar. 2012.

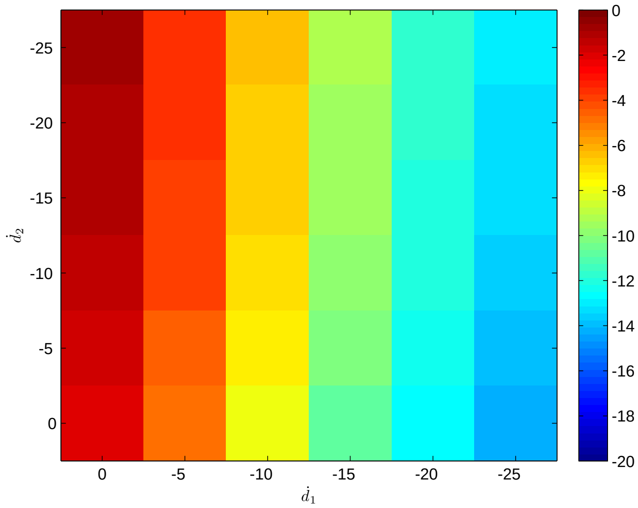

Fig. 1. The distortion of the desired source, d_1, versus the $\dot{d}_1$, $\dot{d}_2$.

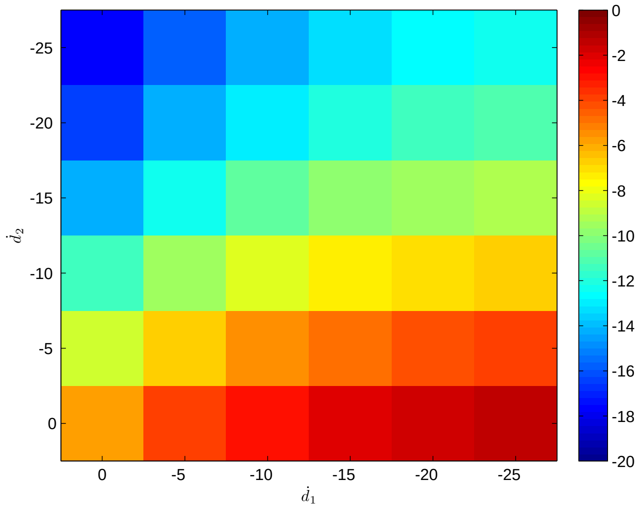

Fig. 2. The distortion of the interfering source, d_2, versus $\dot{d}_1$, $\dot{d}_2$.

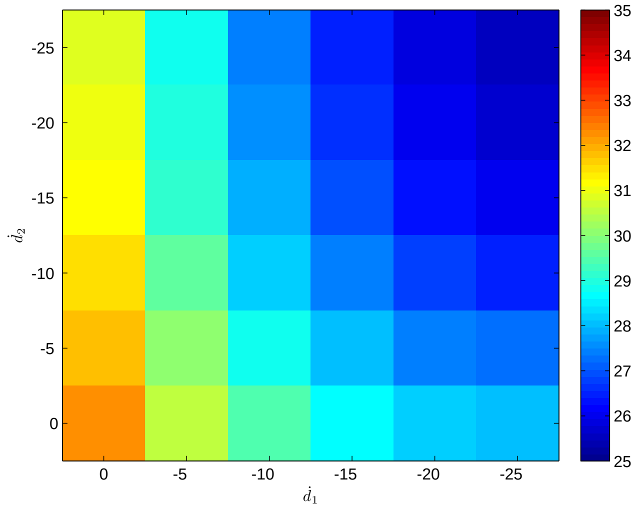

Fig. 3. The NR versus the desired distortion levels $\dot{d}_1$, $\dot{d}_2$.

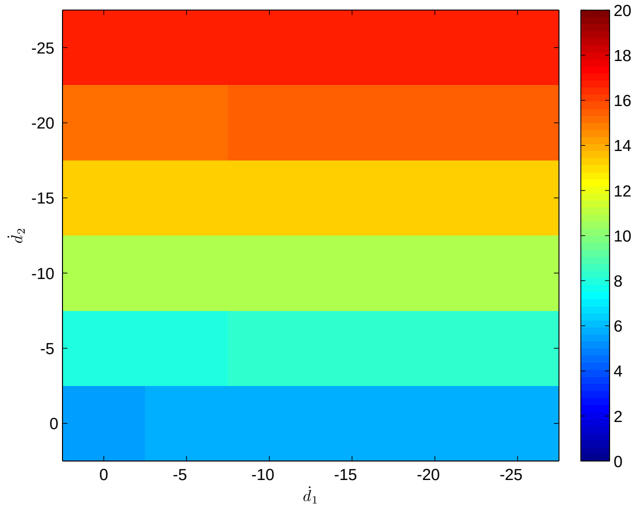

Fig. 4. The output SIR versus the desired distortion levels $\dot{d}_1$, $\dot{d}_2$.