

Performance analysis of dual source transfer-function generalized sidelobe canceller

Gal Reuven ^{a,*}, Sharon Gannot ^b, Israel Cohen ^a

^a Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel
^b School of Engineering, Bar-Ilan University, Ramat-Gan 52900, Israel

Received 22 January 2006; received in revised form 16 December 2006; accepted 16 December 2006

Abstract

In this work, we evaluate the performance of a recently proposed adaptive beamformer, namely Dual source Transfer-Function Generalized Sidelobe Canceller (DTF-GSC). The DTF-GSC is useful for enhancing a speech signal received by an array of microphones in a noisy and reverberant environment. We demonstrate the applicability of the DTF-GSC in some representative reverberant and non-reverberant environments under various noise field conditions. The performance is evaluated based on the power spectral density (PSD) deviation imposed on the desired signal at the beamformer output, the achievable noise reduction, and the interference reduction. We show that the resulting expressions for the PSD deviation and noise reduction depend on the actual acoustical environment, the noise field, and the estimation accuracy of the relative transfer functions (RTFs), defined as the ratio between each acoustical transfer function (ATF) and a reference ATF. The achievable interference reduction is generally independent of the noise field. Experimental results demonstrate the sensitivity of the system's performance to array misalignments.

© 2007 Elsevier B.V. All rights reserved.

1. Introduction

A robust and computationally efficient adaptive beamforming algorithm is an essential signal processing tool for enhancing a speech signal received by an array of microphones in a noisy and reverberant environment (Nordholm et al., 1993; Doclo and Moonen, 2002; Spiet et al., 2004). In most speech enhancement applications the beamformer is constrained to produce a dominant response in the direction of the desired speech source, while minimizing the response in all other directions. However, in reverberant environments the direction of arrival of the desired signal is insufficient for the design of the beamformer, due to the multiple signal's reflections. This problem may be alleviated by using an *room impulse response*

(RIR) rather than just a simple delay for modeling the propagation of the speech signal in a reverberant room.

Gannot et al. (2001) showed that the relative transfer function (RTF), defined as the ratio between the acoustical transfer functions (ATFs) is sufficient for constructing the beamformer. Assuming that the background noise is stationary, and exploiting the non-stationarity of the speech signal, an unbiased estimate of the RTF is obtained. Compared with the conventional generalized sidelobe canceller (GSC) (Griffiths and Jim, 1982), the resulting *transfer function generalized sidelobe canceller* (TF-GSC) is of practical importance when enhancing a speech signal deteriorated by *stationary* interfering signals in an arbitrary ATF enclosure. However, in the presence of an additional *non-stationary* interference, the TF-GSC is insufficient, since it cannot distinguish between the desired signal and the interfering signal.

Gannot et al. (2004) analyzed the performance of the TF-GSC by evaluating the power spectral density (PSD) deviation imposed on the desired signal at the beamformer

* Corresponding author. Tel.: +972 4 8294731; fax: +972 4 8295757.

E-mail addresses: galrv@techunix.technion.ac.il (G. Reuven), gannot@eng.biu.ac.il (S. Gannot), icohen@ee.technion.ac.il (I. Cohen).

output. They showed that for speech signals, PSD deviation is one of the main factors to degradation of speech quality. It is interesting to note that adaptive beamformers based on the GSC structure are often analyzed in the literature by using a measure of *noise reduction* (NR), while not considering any measure of distortion imposed on the desired signal. Analytical expression for a broadband beamformer using Wiener filters is given by Nordholm et al. (1992). Nordholm et al. (1999) also give an analytical description of an adaptive microphone array that facilitates a simple built-in environment calibration. Their analysis employs non-causal Wiener filters yielding compact and effective theoretical noise suppression limits. Nordholm and Leung (2000) further provide an analysis of the limits of the obtainable NR of the GSC in an isotropic noise field. Bitzer et al. (1999, 1998) derive an expression for the NR as a function of the noise field and evaluate the degradation as a function of the reverberation time (T_{60}). The additional NR due to the noise canceller branch of the GSC, implemented by a closed-form Wiener filter rather than the adaptive Widrow-LMS procedure, is evaluated in (Bitzer et al., 1999). Huarng and Yeh (1990) address the distortion issue by evaluating the leakage of the desired signal into the reference noise branch of the GSC. However, they assume delay-only ATFs and consider signal degradation due to steering errors alone. A theoretical assessment of the broadband performance of the GSC and its robustness to different types of model errors for different noise scenarios is presented by Spiert et al. (2005). However, this analysis is conducted in the context of hearing aids.

Recently, we have extended the TF-GSC to deal with scenarios where the desired signal is contaminated by directional non-stationary interference, such as a competing speaker, as well as by stationary interferences, which may comprise both directional and non-directional signals (Reuven et al., 2005, submitted for publication). The resulting beamformer, namely the Dual-Source TF-GSC (DTF-GSC), includes modified versions of *matched beamformer* (MBF) and *blocking matrix* (BM), which facilitate the suppression of the non-stationary interference. The matched beamformer is modified to block the non-stationary interference while maintaining the desired speech signal. The blocking matrix is modified to block both the desired signal and the non-stationary interference. We proposed an efficient method for updating the blocking matrix in double talk situations, by exploiting the non-stationarity of both the desired and interfering speech signals.

In this paper, we evaluate the performance of the DTF-GSC and demonstrate its applicability in some representative reverberant and non-reverberant environments under various noise field conditions. The performance is evaluated based on the PSD deviation (Gannot et al., 2004) imposed on the desired signal by the DTF-GSC algorithm, and based on additional quality measures, specifically the achievable noise reduction and interference reduction. We show that the resulting expressions for the PSD deviation and noise reduction depend on the actual ATFs, the noise

field, and the estimation accuracy of the RTFs, while the achievable interference reduction is generally independent of the noise field. Experimental results demonstrate the effect of estimation error of RTFs on the PSD deviation under reverberant and non-reverberant environmental conditions and various noise fields. The significance of this effect is related to the contradicting constraints imposed on the MBF in case the spatial distance between the desired source and the interfering source is too small. We evaluate the noise reduction performance and its relation to the noise field and the ATFs involved, and evaluate the interference reduction in reverberant and non-reverberant environments assuming certain errors in the RTFs estimates.

The paper is organized as follows. In Section 2, we briefly review the DTF-GSC algorithm. In Section 3, we derive a general expression for the power spectral density of the DTF-GSC output. In Section 5, we evaluate the deviation of the power spectral density of the desired signal from its nominal value for representative reverberant and non-reverberant environments and various noise fields. In Section 6, we evaluate the noise reduction performance. Finally, in Section 7 we evaluate the interference reduction in reverberant and non-reverberant environments assuming certain errors in the RTFs estimates.

2. Summary of the DTF-GSC

We assume an array of sensors in a noisy and reverberant environment that receives signals from three types of sources: a desired speech source, a directional non-stationary interference source (e.g. competing speech source) and a stationary noise source, which can be either directional, non-directional or a combination thereof. The DTF-GSC is designed with the objective of reconstructing the desired speech signal from the received reverberated signals. Let $s_1(t)$ denote the desired speech signal, let $s_2(t)$ denote the non-stationary interfering signal, and let $a_m(t)$ and $b_m(t)$ represent the room impulse responses (RIRs) of the m th microphone to the desired speech source and the non-stationary interference source, respectively. The m th microphone signal is given by

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

where $n_m(t)$ is the (directional or non-directional) stationary noise signal at the m th microphone, and $*$ denotes convolution. In the *short time Fourier transform* (STFT) domain, (1) can be approximately rewritten as¹

$$Z(t, e^{j\omega}) = \mathbf{A}(e^{j\omega})S_1(t, e^{j\omega}) + \mathbf{B}(e^{j\omega})S_2(t, e^{j\omega}) + N(t, e^{j\omega}), \quad (2)$$

¹ We assume that time variation of the impulse responses $a_m(t)$ and $b_m(t)$ over the observation interval is negligible, and that the signals the analysis frame is longer than the filter length.

where

$$\begin{aligned} \mathbf{Z}(t, e^{j\omega}) &= [Z_1(t, e^{j\omega}) \ Z_2(t, e^{j\omega}) \ \dots \ Z_M(t, e^{j\omega})]^T, \\ \mathbf{A}(e^{j\omega}) &= [A_1(e^{j\omega}) \ A_2(e^{j\omega}) \ \dots \ A_M(e^{j\omega})]^T, \\ \mathbf{B}(e^{j\omega}) &= [B_1(e^{j\omega}) \ B_2(e^{j\omega}) \ \dots \ B_M(e^{j\omega})]^T, \\ \mathbf{N}(t, e^{j\omega}) &= [N_1(t, e^{j\omega}) \ N_2(t, e^{j\omega}) \ \dots \ N_M(t, e^{j\omega})]^T, \end{aligned} \quad (3)$$

where ^T denotes the transpose operation. $Z_m(t, e^{j\omega})$, $S_1(t, e^{j\omega})$, $S_2(t, e^{j\omega})$ 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 acoustic transfer function (ATFs) relating the desired source and interference source and the m th microphone, respectively.

A block diagram of the DTF-GSC scheme is depicted in Fig. 1. The structure of the DTF-GSC is the same as that of the GSC. It comprises of three blocks: matched beamformer, blocking matrix, and multi-channel noise canceller. However, two of its components, namely the matched beamformer and the blocking matrix, are modified to enable suppression of the non-stationary interference.

It was shown in (Reuven et al., submitted for publication) that it is sufficient to use the ratio between the ATFs (RTFs) rather than the ATFs themselves in order to implement the suggested algorithm. Define the desired and interference signals' RTFs

$$\tilde{\mathbf{A}}(e^{j\omega}) \triangleq \frac{\mathbf{A}(e^{j\omega})}{A_1(e^{j\omega})} = \left[1 \ \frac{A_2(e^{j\omega})}{A_1(e^{j\omega})} \ \dots \ \frac{A_M(e^{j\omega})}{A_1(e^{j\omega})} \right]^T, \quad (4)$$

$$\tilde{\mathbf{B}}(e^{j\omega}) \triangleq \frac{\mathbf{B}(e^{j\omega})}{B_1(e^{j\omega})} = \left[1 \ \frac{B_2(e^{j\omega})}{B_1(e^{j\omega})} \ \dots \ \frac{B_M(e^{j\omega})}{B_1(e^{j\omega})} \right]^T, \quad (5)$$

where the first microphone is arbitrarily chosen as the reference microphone. The modified fixed beamformer is

designed to block the non-stationary interference while maintaining the desired speech signal, and is given by

$$\mathbf{W}_0(e^{j\omega}) = \frac{\tilde{\mathbf{A}}(e^{j\omega}) - \rho(e^{j\omega}) \frac{\tilde{\mathbf{B}}(e^{j\omega})}{\|\tilde{\mathbf{A}}(e^{j\omega})\| \|\tilde{\mathbf{B}}(e^{j\omega})\|}}{1 - |\rho(e^{j\omega})|^2} \mathcal{F}(e^{j\omega}). \quad (6)$$

$\rho(e^{j\omega})$ denotes the cosine of the angle between the vectors $\tilde{\mathbf{A}}(e^{j\omega})$ and $\tilde{\mathbf{B}}(e^{j\omega})$ in an inner product space

$$\rho(e^{j\omega}) \equiv \frac{\tilde{\mathbf{B}}^\dagger(e^{j\omega}) \tilde{\mathbf{A}}(e^{j\omega})}{\|\tilde{\mathbf{A}}(e^{j\omega})\| \|\tilde{\mathbf{B}}(e^{j\omega})\|}, \quad (7)$$

where ^{dagger} denotes the conjugate transpose operation. $\mathcal{F}(e^{j\omega})$ represents the desired filter response of the DTF-GSC with respect to the desired speech signal $s_1(t)$ (Reuven et al., submitted for publication). The modified $M \times (M-2)$ blocking matrix is designed to block both the desired signal and the non-stationary interference, and is given by

$$\mathcal{H}(e^{j\omega}) = \begin{bmatrix} Q_3(e^{j\omega}) & Q_4(e^{j\omega}) & \dots & Q_M(e^{j\omega}) \\ L_3(e^{j\omega}) & L_4(e^{j\omega}) & \dots & L_M(e^{j\omega}) \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \ddots & \ddots & \ddots \\ 0 & 0 & \dots & 1 \end{bmatrix}, \quad (8)$$

where

$$Q_m(e^{j\omega}) = -\frac{\tilde{A}_2^*(e^{j\omega}) \tilde{B}_m^*(e^{j\omega}) - \tilde{B}_2^*(e^{j\omega}) \tilde{A}_m^*(e^{j\omega})}{\tilde{A}_2^*(e^{j\omega}) - \tilde{B}_2^*(e^{j\omega})}, \quad m = 3, \dots, M \quad (9)$$

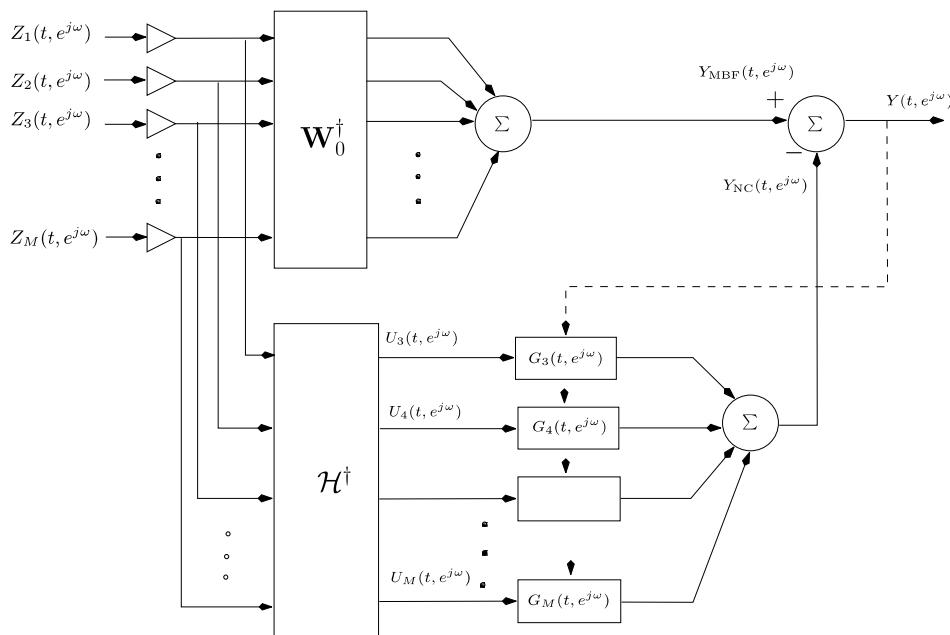


Fig. 1. GSC solution for the dual source case. Three blocks: a matched beamformer $\mathbf{W}_0^\dagger(t, e^{j\omega})$; a blocking matrix $\mathcal{H}^\dagger(e^{j\omega})$; and a multi channel noise canceller $\mathbf{G}(t, e^{j\omega})$ (Reuven et al., submitted for publication).

1) Matched beamformer:

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

2) Noise reference signals:

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

3) Output signal:

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

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

$$\begin{aligned} \tilde{G}_m(t+1, e^{j\omega}) &= G_m(t, e^{j\omega}) + \mu \frac{U_m(t, e^{j\omega}) Y^*(t, e^{j\omega})}{P_{\text{est}}(t, e^{j\omega})} \\ G_m(t+1, e^{j\omega}) &\xleftarrow{\text{FIR}} \tilde{G}_m(t+1, e^{j\omega}) \end{aligned}$$

where, $P_{\text{est}}(t, e^{j\omega}) = \eta P_{\text{est}}(t-1, e^{j\omega}) + (1-\eta) \|\mathbf{Z}(t, e^{j\omega})\|^2$

5) keep only non-aliased samples.

Fig. 2. Summary of the DTF-GSC algorithm (Reuven et al., submitted for publication).

and

$$L_m(e^{j\omega}) = -\frac{\tilde{A}_m^*(e^{j\omega}) - \tilde{B}_m^*(e^{j\omega})}{\tilde{A}_2^*(e^{j\omega}) - \tilde{B}_2^*(e^{j\omega})}, \quad m = 3, \dots, M, \quad (10)$$

where $*$ denotes the conjugate operation. The steps involved in the application of the DTF-GSC algorithm are summarized in Fig. 2. It is proven in (Reuven et al., submitted for publication) that using the above GSC blocks the desired signal component at the output of the structure is given by

$$Y^{S_1}(t, e^{j\omega}) = S_1(t, e^{j\omega}) A_1(e^{j\omega}) \mathcal{F}(e^{j\omega}). \quad (11)$$

The third block, the multi-channel noise canceller, is similar to the corresponding TF-GSC block (only the number of parallel filters is changed to $M-2$ rather than $M-1$ in the TF-GSC)

$$\tilde{G}_m(t+1, e^{j\omega}) = G_m(t, e^{j\omega}) + \mu \frac{U_m(t, e^{j\omega}) Y^*(t, e^{j\omega})}{P_{\text{est}}(t, e^{j\omega})}, \quad (12)$$

$$G_m(t+1, e^{j\omega}) \xleftarrow{\text{FIR}} \tilde{G}_m(t+1, e^{j\omega})$$

for $m = 3, \dots, M$. The normalizing term is given by

$$P_{\text{est}}(t, e^{j\omega}) = \eta P_{\text{est}}(t-1, e^{j\omega}) + (1-\eta) \|\mathbf{Z}(t, e^{j\omega})\|^2 \quad (13)$$

and η is a forgetting factor (typically $0.8 < \eta < 1$).² Finally, the operator $\xleftarrow{\text{FIR}}$ 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

² Another possibility is to calculate P_{est} using the $\|\mathbf{U}(t, e^{j\omega})\|^2$ instead of $\|\mathbf{Z}(t, e^{j\omega})\|^2$. 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 $P_{\text{est}}(t, e^{j\omega})$ using the input sensor signals, $Z_m(t, e^{j\omega})$, as indicated in (13), 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.

FIR constraint is imposed. Third, the result is transformed back to the frequency domain. Performing the $\xleftarrow{\text{FIR}}$ operator avoids cyclic convolution effects.

3. Output power spectral density

In this section, we derive a general expression for the output PSD of the DTF-GSC, which is necessary for analyzing its performance. From the algorithm summary in Fig. 2, we have the DTF-GSC output

$$\begin{aligned} Y(t, e^{j\omega}) &= Y_{\text{MBF}}(t, e^{j\omega}) - \mathbf{G}^\dagger(t, e^{j\omega}) \mathbf{U}(t, e^{j\omega}) \\ &= \widehat{\mathbf{W}}_0^\dagger(e^{j\omega}) \mathbf{Z}(t, e^{j\omega}) - \mathbf{G}^\dagger(t, e^{j\omega}) \widehat{\mathcal{H}}^\dagger(e^{j\omega}) \mathbf{Z}(t, e^{j\omega}), \end{aligned} \quad (14)$$

where only estimates of the RTFs, rather than the exact values, are assumed to be known.³ Hence, the PSD of the output signal is given by

$$\begin{aligned} \phi_{YY}(t, e^{j\omega}) &= E\{Y(t, e^{j\omega}) Y^*(t, e^{j\omega})\} \\ &= E\left\{(\widehat{\mathbf{W}}_0^\dagger(e^{j\omega}) \mathbf{Z}(t, e^{j\omega}) - \mathbf{G}^\dagger(t, e^{j\omega}) \widehat{\mathcal{H}}^\dagger(e^{j\omega}) \mathbf{Z}(t, e^{j\omega}))\right. \\ &\quad \times (\widehat{\mathbf{W}}_0^\dagger(e^{j\omega}) \mathbf{Z}(t, e^{j\omega}) - \mathbf{G}^\dagger(t, e^{j\omega}) \widehat{\mathcal{H}}^\dagger(e^{j\omega}) \mathbf{Z}(t, e^{j\omega}))^\dagger\} \\ &= \widehat{\mathbf{W}}_0^\dagger(e^{j\omega}) \Phi_{ZZ}(t, e^{j\omega}) \widehat{\mathbf{W}}_0(e^{j\omega}) \\ &\quad - \mathbf{G}^\dagger(t, e^{j\omega}) \widehat{\mathcal{H}}^\dagger(e^{j\omega}) \Phi_{ZZ}(t, e^{j\omega}) \widehat{\mathbf{W}}_0(e^{j\omega}) \\ &\quad - \widehat{\mathbf{W}}_0^\dagger(e^{j\omega}) \Phi_{ZZ}(t, e^{j\omega}) \widehat{\mathcal{H}}(e^{j\omega}) \mathbf{G}(t, e^{j\omega}) \\ &\quad + \mathbf{G}^\dagger(t, e^{j\omega}) \widehat{\mathcal{H}}^\dagger(e^{j\omega}) \Phi_{ZZ}(t, e^{j\omega}) \widehat{\mathcal{H}}(e^{j\omega}) \mathbf{G}(t, e^{j\omega}), \end{aligned} \quad (15)$$

where $\Phi_{ZZ}(t, e^{j\omega}) = E\{\mathbf{Z}(t, e^{j\omega}) \mathbf{Z}^\dagger(t, e^{j\omega})\}$ is the input signal PSD matrix. The output PSD depends on the input signal $\mathbf{Z}(t, e^{j\omega})$ and the multi-channel filter $\mathbf{G}(t, e^{j\omega})$ derived in the sequel. Although signal leakage is possible due to errors in estimating $\tilde{\mathbf{A}}(e^{j\omega})$ and $\tilde{\mathbf{B}}(e^{j\omega})$, we can calculate the multi-channel Wiener filter $\mathbf{G}(t, e^{j\omega})$ during non-active periods of the desired signal and non-stationary interference, in which case $\mathbf{Z}(t, e^{j\omega}) = \mathbf{N}(t, e^{j\omega})$. Thus,

$$\begin{aligned} \mathbf{U}(t, e^{j\omega})|_{Z=N} &= \widehat{\mathcal{H}}^\dagger(e^{j\omega}) \mathbf{N}(t, e^{j\omega}), \\ Y_{\text{MBF}}|_{Z=N} &= \widehat{\mathbf{W}}_0^\dagger(e^{j\omega}) \mathbf{N}(t, e^{j\omega}). \end{aligned} \quad (16)$$

For calculating the multichannel Wiener filter $\mathbf{G}(t, e^{j\omega})$ we need the following PSD vector $\Phi_{UY_{\text{MBF}}}(t, e^{j\omega})$ and PSD matrix $\Phi_{UU}(t, e^{j\omega})$

$$\begin{aligned} \Phi_{UY_{\text{MBF}}}(t, e^{j\omega}) &= E\{\mathbf{U}(t, e^{j\omega}) Y_{\text{MBF}}^*(t, e^{j\omega})\} \\ &= E\left\{\widehat{\mathcal{H}}^\dagger(e^{j\omega}) \mathbf{N}(t, e^{j\omega}) (\widehat{\mathbf{W}}_0^\dagger(e^{j\omega}) \mathbf{N}(t, e^{j\omega}))^\dagger\right\} \\ &= \widehat{\mathcal{H}}^\dagger(e^{j\omega}) \Phi_{NN}(t, e^{j\omega}) \widehat{\mathbf{W}}_0(e^{j\omega}) \end{aligned} \quad (17)$$

³ To avoid excess notation we use from this point on $Y(t, e^{j\omega})$ and $Y_{\text{MBF}}(t, e^{j\omega})$ to denote the involved signals when only estimated values of the RTFs are given.

and

$$\begin{aligned}\Phi_{UU}(t, e^{j\omega}) &= E\{\mathbf{U}(t, e^{j\omega})\mathbf{U}^\dagger(t, e^{j\omega})\} \\ &= E\{\widehat{\mathcal{H}}^\dagger(e^{j\omega})\mathbf{N}(t, e^{j\omega})\mathbf{N}^\dagger(t, e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\} \\ &= \widehat{\mathcal{H}}^\dagger(e^{j\omega})\Phi_{NN}(t, e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega}),\end{aligned}\quad (18)$$

where $\Phi_{NN}(t, e^{j\omega}) = E\{\mathbf{N}(t, e^{j\omega})\mathbf{N}^\dagger(t, e^{j\omega})\}$. Hence, the Wiener filter is given by

$$\begin{aligned}\mathbf{G}(t, e^{j\omega}) &= (\widehat{\mathcal{H}}^\dagger(e^{j\omega})\Phi_{NN}(t, e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega}))^{-1} \\ &\quad \times \widehat{\mathcal{H}}^\dagger(e^{j\omega})\Phi_{NN}(t, e^{j\omega})\widehat{\mathbf{W}}_0(e^{j\omega}).\end{aligned}\quad (19)$$

Note that $\mathbf{G}(t, e^{j\omega})$ is only responsible for the stationary background noise reduction, while the competing speech signal $s_2(t)$ is blocked by the MBF and the BM blocks. Substituting (19) into (15) we have

$$\begin{aligned}\phi_{YY}(t, e^{j\omega}) &\left\{ \widehat{\mathbf{W}}_0^\dagger(e^{j\omega})\Phi_{ZZ}(t, e^{j\omega})\widehat{\mathbf{W}}_0(e^{j\omega}) \right. \\ &- \widehat{\mathbf{W}}_0^\dagger(e^{j\omega})\Phi_{NN}(t, e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\left(\widehat{\mathcal{H}}^\dagger(e^{j\omega})\Phi_{NN}(t, e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\right)^{-1} \\ &\times \widehat{\mathcal{H}}^\dagger(e^{j\omega})\Phi_{ZZ}(t, e^{j\omega})\widehat{\mathbf{W}}_0(e^{j\omega}) \\ &- \widehat{\mathbf{W}}_0^\dagger(e^{j\omega})\Phi_{ZZ}(t, e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\left(\widehat{\mathcal{H}}^\dagger(e^{j\omega})\Phi_{NN}(t, e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\right)^{-1} \\ &\times \widehat{\mathcal{H}}^\dagger(e^{j\omega})\Phi_{NN}(t, e^{j\omega})\widehat{\mathbf{W}}_0(e^{j\omega}) \\ &+ \widehat{\mathbf{W}}_0^\dagger(e^{j\omega})\Phi_{NN}(t, e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\left(\widehat{\mathcal{H}}^\dagger(e^{j\omega})\Phi_{NN}(t, e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\right)^{-1} \\ &\times \widehat{\mathcal{H}}^\dagger(e^{j\omega})\Phi_{ZZ}(t, e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\left(\widehat{\mathcal{H}}^\dagger(e^{j\omega})\Phi_{NN}(t, e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\right)^{-1} \\ &\left. \times \widehat{\mathcal{H}}^\dagger(e^{j\omega})\Phi_{NN}(t, e^{j\omega})\widehat{\mathbf{W}}_0(e^{j\omega}) \right\}.\end{aligned}\quad (20)$$

4. Evaluation methodology

The expression in (20) is used for evaluating three important attributes of the DTF-GSC algorithm: the desired signal PSD deviation imposed by the algorithm, the achievable noise reductions, and the interference cancellation.

The expression in (20) depends on the input signal PSD $\Phi_{ZZ}(t, e^{j\omega})$, the noise PSD $\Phi_{NN}(t, e^{j\omega})$, and the acoustical environment. The latter manifests itself through the signals' ATF (involved in the respective RTFs), which are employed through the fixed beamformer $\widehat{\mathbf{W}}_0(e^{j\omega})$ and the blocking matrix $\widehat{\mathcal{H}}(e^{j\omega})$.⁴

Using the desired, interference, and noise signals' independence, the *PSD deviation* imposed by the algorithm, its *noise reduction*, and its *interference reduction* can be calculated separately by deriving expressions for the output PSD in the following three cases:

$$\begin{aligned}\Phi_{YY}(t, e^{j\omega}) &= \begin{cases} \Phi_{YY}^S(t, e^{j\omega}) & \text{if } \mathbf{Z}(t, e^{j\omega}) = \mathbf{A}(e^{j\omega})\mathbf{S}_1(t, e^{j\omega}) \Rightarrow \text{desired signal deviation}, \\ \Phi_{YY}^N(t, e^{j\omega}) & \text{if } \mathbf{Z}(t, e^{j\omega}) = \mathbf{N}(t, e^{j\omega}) \Rightarrow \text{noise reduction}, \\ \Phi_{YY}^S(t, e^{j\omega}) & \text{if } \mathbf{Z}(t, e^{j\omega}) = \mathbf{B}(e^{j\omega})\mathbf{S}_2(t, e^{j\omega}) \Rightarrow \text{interference reduction}. \end{cases}\end{aligned}$$

⁴ Note that Gannot et al. (2004) obtained a similar expression for evaluating the TF-GSC deviation performance, however $\widehat{\mathbf{W}}_0(e^{j\omega})$ and $\widehat{\mathcal{H}}(e^{j\omega})$ are completely different for the DTF-GSC.

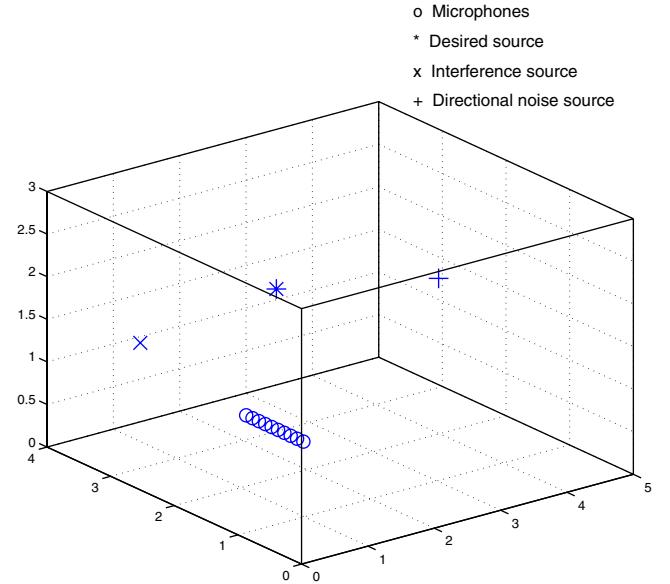


Fig. 3. Test scenario: 10 microphone linear array in a noisy and reverberated room. Room dimensions 450 cm × 330 cm × 420 cm. Microphones are arranged along the y-axis with inter-element distance of 10 cm. The first Microphone is located at coordinates [70 70 105] cm. The nominal desired source position is [245 295 170] cm, and the nominal interference signal position is [40 295 150] cm. Whenever a noise point source is used its nominal position is [245 40 270] cm.

Throughout the analysis, we will discuss two acoustical environments. The first is the, commonly used, free-space propagation scenario, in which the acoustical paths, relating directional sources and the sensors (microphones), are modeled as pure delays. The second is the, more realistic, reverberant enclosure, in which a complex RIR relates the sources and the sensors. The free-space propagation scenario can be seen as a special case of the reverberant scenario, where

$$\mathbf{A}(e^{j\omega}) = [1 \quad e^{-j\omega\tau_s} \quad \dots \quad e^{-j(M-1)\omega\tau_s}]^T.$$

In determining this steering vector we use the assumption that the microphones are forming an equally-spaced linear array. The relative delay between sensors, τ_s , is determined by the angle of arrival of the wave-front. Define the distance between microphones as ℓ_0 , c the speed of sound (342 m/S for air propagation), and θ_s the angle between the array axis and the sound wave direction. Then $\tau_s = \frac{\ell_0}{c} \cos(\theta_s)$.

For the evaluation of the reverberant scenario, we use RIRs simulated using Allen and Berkley's *image method*⁵ (Allen and Berkley, 1979). The scenario shown in Fig. 3 is studied. The enclosure is a room with dimensions 450 cm × 330 cm × 420 cm. An array of 10 microphones with inter-element distance of 10 cm, is used. The sampling frequency is set to 8 kHz, while reverberation time is set to T_{60} 300 ms. The room impulse response and the respective

⁵ The authors thank E.A.P. Habets from T.U. Eindhoven for providing an efficient implementation of the image method.

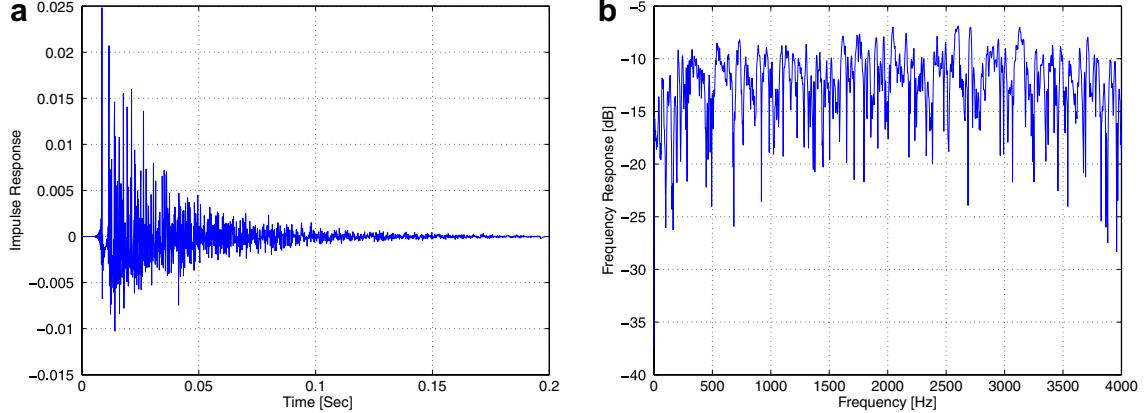


Fig. 4. Simulated room environment: (a) room impulse response (b) and acoustical transfer function relating the desired speech signal and the first microphone.

acoustical transfer function between the desired speech source and the first microphone are depicted in Fig. 4. All other RIRs and ATFs have similar structure.

For the evaluation of the free-space propagation scenario we used the same array configuration. However, the RIRs are replaced by pure delay impulse responses.

The noise PSD dependence was tested using three typical noise fields: coherent, diffused and incoherent (spatially white).

For the coherent noise field the noise signal is modeled as a single point source. Define $d_m(t)$, $m = 1, \dots, M$ as the RIRs relating the noise source and the m th sensor. Since the RIRs are assumed to be slowly time varying, the time dependence of their respective frequency response $D_m(e^{j\omega})$, $m = 1, \dots, M$ can be omitted. Define

$$N(t, e^{j\omega}) = \mathbf{D}(e^{j\omega})N(t, e^{j\omega}),$$

where

$$\mathbf{D}(e^{j\omega}) = [D_1(e^{j\omega}) \quad D_2(e^{j\omega}) \quad \dots \quad D_M(e^{j\omega})]^T.$$

The PSD matrix of the noise component at the sensors, $\Phi_{NN}(t, e^{j\omega})$, is thus given by

$$\Phi_{NN}(t, e^{j\omega}) = \phi_{NN}(t, e^{j\omega})\mathbf{D}(e^{j\omega})\mathbf{D}^\dagger(e^{j\omega}), \quad (21)$$

where $\phi_{NN}(t, e^{j\omega})$ is the input noise PSD. If a free-space propagation is assumed, the acoustical frequency response (ATF) simplifies to

$$\mathbf{D}(e^{j\omega}) = [1 \quad e^{-j\omega\tau_n} \quad \dots \quad e^{-j(M-1)\omega\tau_n}]^T,$$

where we used again the equal inter-element distance of the array. The relative delay between sensors, τ_n , is determined by the angle of arrival of the wave-front. Defining θ_n , the angle between the array axis and the sound wave direction, we obtain $\tau_n = \frac{\ell_0}{c} \cos(\theta_n)$.

In reverberant acoustical environment, such as a car enclosure, the noise field tends to be diffused (see for instance Dal-Degan and Prati, 1988; Bitzer et al., 1999). A diffused noise source is assumed to be equi-distributed on a sphere in the far field of the array. The cross-coherence function between signals received by two sensors

(i, j) with distance d_{ij} can be found in (Dal-Degan and Prati, 1988), and is given by

$$\Gamma_{N_i N_j}(e^{j\omega}) = \frac{\phi_{N_i N_j}(e^{j\omega})}{\sqrt{\phi_{N_i N_i}(e^{j\omega})\phi_{N_j N_j}(e^{j\omega})}} = \frac{\sin(\omega d_{ij}/c)}{\omega d_{ij}/c}, \quad (22)$$

where c is the speed of sound. For the equal inter-element distance case $d_{ij} = |i - j|\ell_0$, for $i, j = 0, \dots, M - 1$. Therefore, the coherence matrix is given by,

$$\Gamma(e^{j\omega}) = \begin{bmatrix} 1 & \Gamma_{N_1 N_2}(e^{j\omega}) & \dots & \Gamma_{N_1 N_M}(e^{j\omega}) \\ \Gamma_{N_2 N_1}(e^{j\omega}) & 1 & \dots & \vdots \\ \vdots & \ddots & \ddots & 1 \end{bmatrix}. \quad (23)$$

The noise PSD at the sensors input is thus,

$$\Phi_{NN}(t, e^{j\omega}) = \phi_{NN}(t, e^{j\omega})\Gamma(e^{j\omega}). \quad (24)$$

In our experiments we simulate the diffused noise field by equally spreading many uncorrelated computer-generated point sources around a sphere. Experiments show that the estimated coherence function approximates the theoretical Sinc-shaped curve.

For incoherent noise field we assume that the noise at the sensors is spatially white. Thus,

$$\Phi_{NN}(t, e^{j\omega}) = \phi_{NN}(t, e^{j\omega})I,$$

where I is an $M \times M$ identity matrix. In our experiments we use M uncorrelated computer-generated noise signals, with equal power $\phi_{NN}(t, e^{j\omega})$, to approximate the incoherent noise field.

In the following Sections 5–7 we analyze the PSD deviation, the noise reduction, and the interference cancellation separately and independently.

5. Evaluation of the PSD deviation

In this section, we evaluate the PSD deviation of the desired signal imposed by the algorithm for several

representative cases. We first derive an expression for its deviation from the nominal value for any RTFs estimates, and then determine the influence of estimation error on the PSD deviation under reverberant and non-reverberant environmental conditions and various noise fields.

The desired signal PSD deviation imposed by the algorithm can be calculated by the general expression given in

(20) when the microphone signals are $Z(t, e^{j\omega}) = A(e^{j\omega}) \times S_1(t, e^{j\omega})$ (Gannot et al., 2004). Denoting

$$\phi_{YY}^S(t, e^{j\omega}) = \phi_{YY}(t, e^{j\omega})|_{Z=AS_1} \quad (25)$$

and assuming exact knowledge of the RTFs $\tilde{A}(e^{j\omega})$ and $\tilde{B}(e^{j\omega})$, i.e. $\widehat{\mathcal{H}}(e^{j\omega}) = \mathcal{H}(e^{j\omega})$ and $\widehat{W}_0(e^{j\omega}) = W_0(e^{j\omega})$,

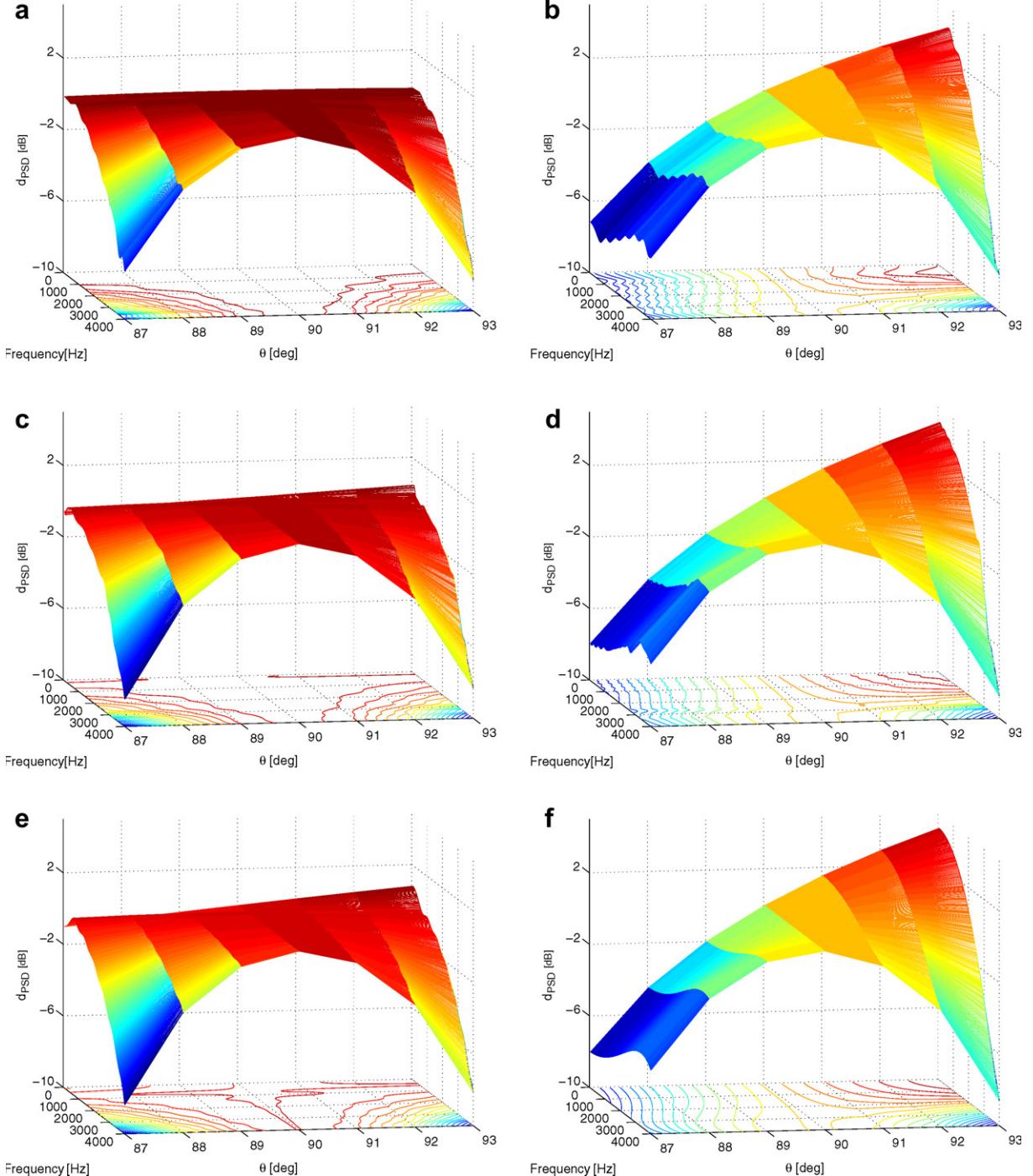


Fig. 5. Deviation as a function of the frequency and direction of arrival. Desired signal direction $\theta = 90^\circ$. $M = 10$ sensors. Remote sources: (a) directional noise field ($\theta = 120^\circ$), (c) diffused noise field, (e) incoherent noise field; close sources: (b) directional noise field ($\theta = 120^\circ$), (d) diffused noise field, (f) incoherent noise field.

calculated using (10), (9), (8) and (6). Then using the signal PSD expression $\Phi_{ZZ}(t, e^{j\omega}) = \phi_{S_1 S_1}(t, e^{j\omega}) A(e^{j\omega}) A^\dagger(e^{j\omega})$ and the identities $\mathcal{H}^\dagger(e^{j\omega}) A(e^{j\omega}) = 0$, $\mathcal{H}^\dagger(e^{j\omega}) B(e^{j\omega}) = 0$ and $A^\dagger(e^{j\omega}) W_0(e^{j\omega}) = A_1(e^{j\omega}) \mathcal{F}(e^{j\omega})$, expression (20) reduces to

$$\begin{aligned}\phi_{YY}^{S_1}(t, e^{j\omega})|_{\mathcal{H}=\mathcal{H}, \hat{W}_0=W_0} &= W_0^\dagger(e^{j\omega}) \Phi_{ZZ}(t, e^{j\omega}) W_0(e^{j\omega}) \\ &= \phi_{S_1 S_1}(t, e^{j\omega}) |W_0^\dagger(e^{j\omega}) A(e^{j\omega})|^2 \\ &= \phi_{S_1 S_1}(t, e^{j\omega}) |\mathcal{F}(e^{j\omega})|^2 |A_1(e^{j\omega})|^2.\end{aligned}\quad (26)$$

The filter $\mathcal{F}(e^{j\omega})$ is an arbitrary predetermined filter, so it should not be regarded as a deviation. The filter $A_1(e^{j\omega})$ is the ATF from the signal source to the first (arbitrarily chosen) sensor, which cannot be eliminated by the algorithm. Accordingly when the RTFs are known, the output PSD is the same as that of the arbitrary reference sensor. Therefore, the PSD deviation from the nominal value for any non-accurate RTFs' estimates is obtained by normalizing the output (Gannot et al., 2004)

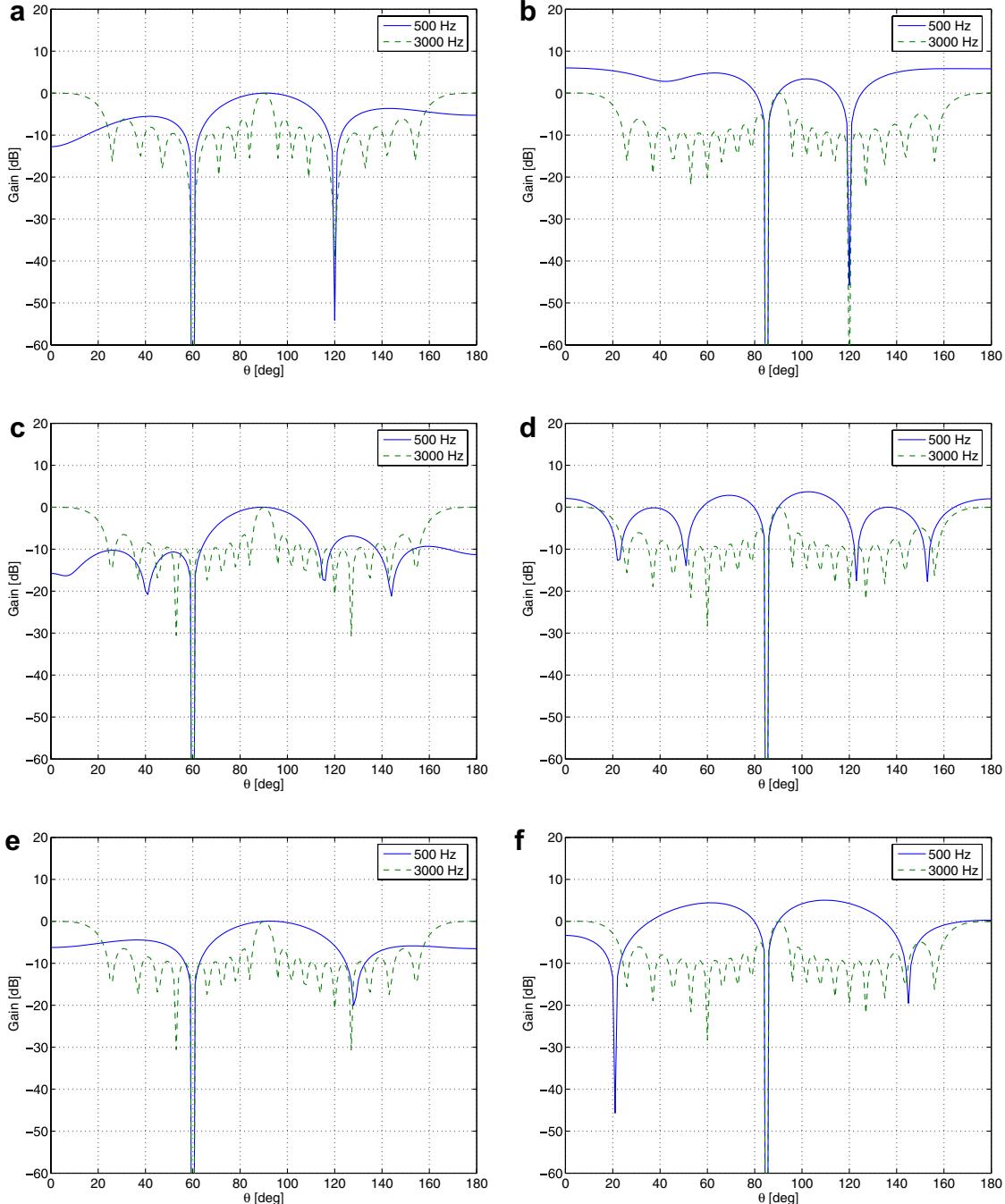


Fig. 6. Spatial beamformer response in 500 Hz and 3000 Hz for the remote sources scenario: (a) directional (c) diffused, (e) incoherent noise field; close sources scenario: (b) directional (d) diffused, (f) incoherent noise field.

$$\begin{aligned} d_{\text{PSD}}(t, e^{j\omega}) &= \frac{\phi_{YY}^{S_1}(t, e^{j\omega})}{\phi_{YY}^{S_1}(t, e^{j\omega})|_{\mathcal{H}=\mathcal{H}, \hat{W}_0=W_0}} \\ &= \frac{\phi_{YY}^{S_1}(t, e^{j\omega})}{|\mathcal{F}(e^{j\omega})|^2 |A_1(e^{j\omega})|^2 \phi_{S_1 S_1}(t, e^{j\omega})}. \end{aligned} \quad (27)$$

A value of $d_{\text{PSD}}(t, e^{j\omega}) = 1$ indicates no-deviation from the nominal value, which is achieved whenever an exact knowledge of the RTFs is available. Any other value (either larger or smaller than 1) indicates a PSD deviation.

We will now determine the sensitivity to errors in estimating the RTFs on the expected deviation imposed by the algorithm. The results depend on the desired signal's ATFs and the noise field. We begin with the free-space propagation case and determine the performance degradation as a function of the steering error. Then, we proceed to the more general reverberant enclosure.

5.1. Free-space propagation

Assume free space propagation, i.e., the ATFs from the desired source and interference source to the sensors are pure delays. Assume that the desired signal impinges on the array from $\theta = 90^\circ$ (in respect to the array axis) and that we use the microphone array depicted in 3. We will examine two scenarios. In the first, denoted *remote sources*, the non-stationary interference impinges on the array from $\theta = 60^\circ$, while in the second, denoted *close sources*, $\theta = 85^\circ$. The deviation imposed by the algorithm is evaluated by steering the array to the assumed desired signal angle of arrival ($\theta = 90^\circ$) while changing the real direction of the signal in the range $\theta \in [87^\circ, 93^\circ]$. We examine now the influence of the noise field on the imposed PSD deviation.

5.1.1. Directional noise signal

In the directional noise field case we optimize the array to cancel a noise source which impinges on the array from $\theta = 120^\circ$ (by optimization of the array we refer to designing the optimal Wiener filter in the noise cancellation branch).

In Fig. 5a and b we present the deviation of the output signal of the array as a function of the frequency and the steering angle, for the remote sources scenario and close sources scenario, respectively. It is clearly shown that for both scenarios the output signal is distorted as we move the steering angle of the array away from the desired signal direction $\theta = 90^\circ$. In both the remote and close sources scenarios the nominal, no distortion value of 0 dB, is obtained in the desired source direction, i.e. $\theta = 90^\circ$, as expected. In the remote sources scenario, the imposed deviation is not exceeding -4 dB for frequencies up to 3000 Hz. On the other hand, in the close sources scenario, the deviation may reach -8 dB, where maximum deviation is obtained only for the high frequency range.

The spatial beamformer response in 500 Hz and 3000 Hz is depicted in Fig. 6a and b, for the remote and close sources scenarios, respectively. It can be clearly seen that the main lobe around the desired direction is wider at 500 Hz than in 3000 Hz. This is in good agreement with the theory, since at $\omega = 0$ [rad/s] there is no phase difference between the signals at the sensors. Wide main lobe is reflected in low sensitivity to steering angle errors. In this case, only negligible impairment to the desired signal is obtained. However, the wider lobe limits the obtained performance of the other figures-of-merit, as will be shown in the sequel.

The close sources scenario impose larger PSD deviation due to the contradicting constraints.

5.1.2. Diffused noise field

If the array is designed to work with a diffused noise field, the deviation imposed on the desired signal is given in Fig. 5c and d. Again, for the remote sources scenario, no more than 5 dB deviation is imposed in the range $\theta \in [87^\circ, 93^\circ]$ for frequencies lower than 3000 Hz. Similar results to the directional noise field are observed for the close sources scenario.

Note that for the diffused scenario the deviation is non-symmetric as a function of the steering error, despite of the

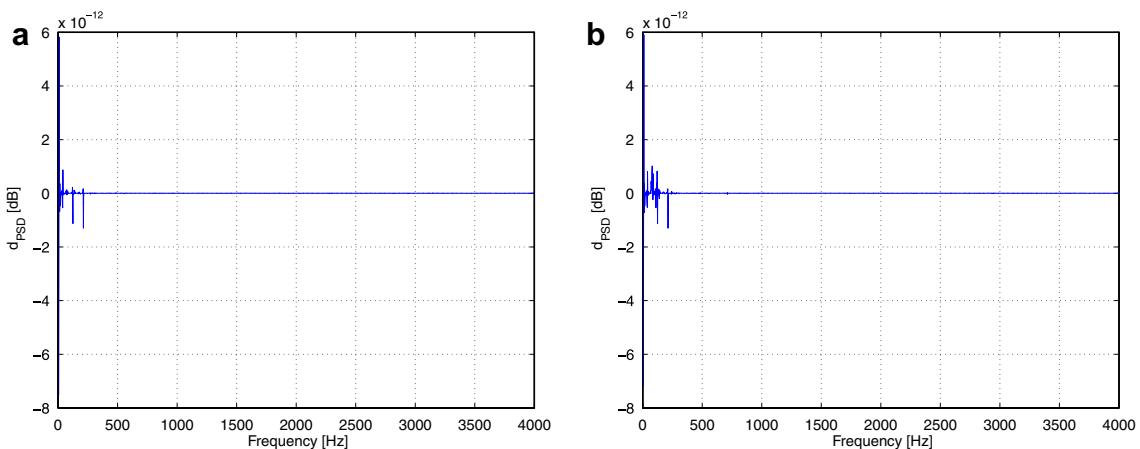


Fig. 7. Expected deviation performance for directional noise field: signal received from the desired direction. Desired signal PSD deviation in (a) the MBF output, and (b) the DTF-GSC output.

inherent symmetry of the noise field. This phenomenon is emphasized in the close sources scenario. This is due to the null directed towards the interference direction in the beamformer spatial response, implemented by both the MBF and BM. When the sources are remote, the interference reduction constraint has no effect on the deviation near the desired direction. On the other hand, when the sources are close, the contradicting constraints are hardly met causing a severe degradation around the desired direction.

The spatial beamformer response in 500 Hz and 3000 Hz is depicted in Fig. 6c and d, for both the remote and close sources scenarios. Similar trends to the directional noise field can be observed.

5.1.3. Incoherent noise field

When the array is designed to work with an incoherent noise field, the deviation imposed on the desired signal is given in Fig. 5e and f. Again, no PSD deviation is obtained at $\theta = 90^\circ$ for both close and remote sources scenarios. In

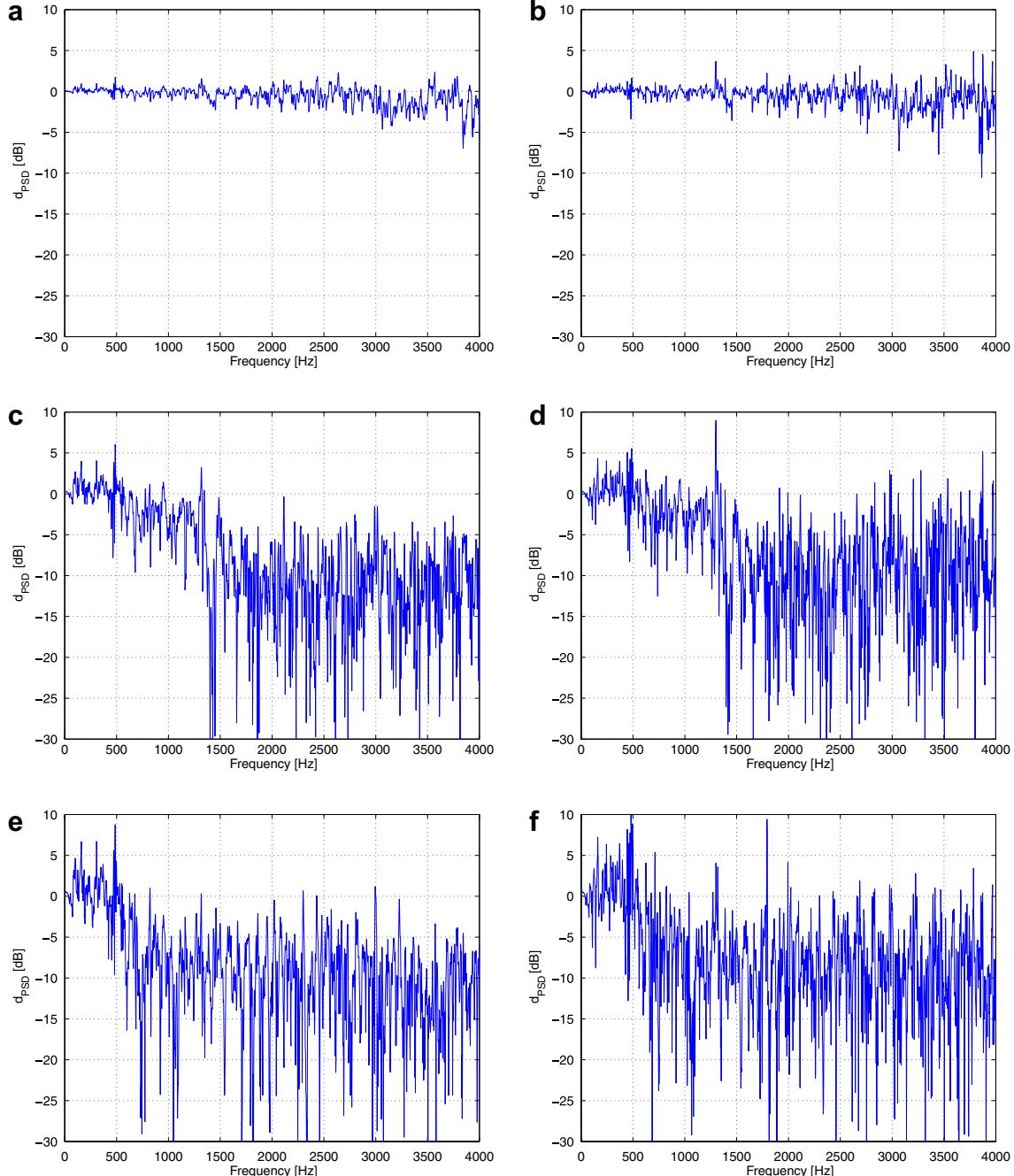


Fig. 8. Expected deviation performance for directional noise field. Desired signal received 1.7 cm away from the estimated point: desired signal PSD deviation in (a) the MBF output, and (b) the DTF-GSC output. 8.7 cm away off: (c) MBF output deviation, (d) DTF-GSC output deviation. 17.3 cm away off: (e) MBF output deviation, (f) DTF-GSC output deviation.

the remote sources scenario, no more than 5 dB deviation is imposed in the range $\theta \in [87^\circ, 93^\circ]$ for frequencies lower than 3000 Hz. On the other hand, in the close sources scenario, the deviation may reach -8 dB.

The spatial beamformer response in 500 Hz and 3000 Hz is depicted in Fig. 6e and f. Similar trends to the directional noise field can be observed.

5.2. Reverberant enclosure

In the reverberant enclosure case we limit our discussion to the directional noise case due to the weak dependence on the noise field. Nevertheless, the noise field type may have some influence on the estimation error and therefore affects the deviation. The algorithm was designed (i.e., $W_0(e^{j\omega})$, $\mathcal{H}(e^{j\omega})$ and $G(t, e^{j\omega})$ were determined) to receive signals from the desired direction, reject interference from the competing source direction and to cancel signals arriving from the noise direction. Results for the figure-of-merit introduced before (desired signal deviation $d_{PSD}(e^{j\omega})$) are presented. Note that the deviation is measured here twice. First, at the MBF output, and second at the DTF-GSC output. In Fig. 7, the array response to a signal source transmitted from the desired direction is introduced. The deviation from the desired response demonstrated in the interesting frequency band is negligible. It can be noticed that the MBF succeeds in maintaining the desired signal direction, while the lower branch has almost no effect on the signal distortion at the output.

Consider the case where, due to movement of the desired source or error in the estimation of the RTFs, the estimated desired source position is 1.7 cm away from the real point (1 cm away off in each axis). MBF deviation and array output deviation for that case are depicted in Fig. 8a and b, respectively. It can be noticed that the deviation is still minor. In Fig. 8c–f larger position estimation errors are presented, for misalignment of 8.7 cm and 17.3 cm off the correct position, respectively. Clear degradation in performance can be observed as the distance increases. While 4 dB average deviation can be observed in (b) for the higher frequency range, around 6 dB average deviation is depicted in (d) and (f) for the same frequency range.

In all cases the PSD deviation is mainly due to the MBF branch, although considerable amount of distortion is imposed by the leakage to the ANC branch. Note, that this leakage may impose an unpleasant distortion for real speech signals, as the resulting speech signal is usually perceived as muffled.

5.3. Summary

In this section we evaluated the PSD deviation in two typical scenarios: free-space propagation and reverberant enclosure. For the former case, we evaluated the distortion due to array steering errors in three noise fields: directional, diffused, and incoherent. For the latter case, we evaluated the sensitivity to talker movements from the nominal posi-

tion. In this case we demonstrate the performance only for directional noise field, due to the weak dependence of the obtained performance on the noise field.

6. Evaluation of noise reduction

In this section, we evaluate the noise reduction performance and its relation to the noise field and the ATFs involved. To evaluate the noise reduction capability of the algorithm, we will use again the general expression for the output signal given by (20), this time with a noise signal $Z(t, e^{j\omega}) = N(t, e^{j\omega})$ as the input signal (the same noise signal used to calculate the optimal Wiener filter). Denoting $\phi_{YY}^N(t, e^{j\omega}) = \phi_{YY}(t, e^{j\omega})|_{Z=N}$, (20) reduces to

$$\begin{aligned}\phi_{YY}^N(t, e^{j\omega}) &= \widehat{W}_0^\dagger(e^{j\omega}) \Phi_{NN}(t, e^{j\omega}) \widehat{W}_0(e^{j\omega}) \\ &\quad - \widehat{W}_0^\dagger(e^{j\omega}) \Phi_{NN}(t, e^{j\omega}) \widehat{\mathcal{H}}(e^{j\omega}) \\ &\quad \times \left(\widehat{\mathcal{H}}^\dagger(e^{j\omega}) \Phi_{NN}(t, e^{j\omega}) \widehat{\mathcal{H}}(e^{j\omega}) \right)^{-1} \\ &\quad \times \widehat{\mathcal{H}}^\dagger(e^{j\omega}) \Phi_{NN}(t, e^{j\omega}) \widehat{W}_0(e^{j\omega}).\end{aligned}\quad (28)$$

The resulting expression depends on the noise PSD at the sensors. We therefore evaluate the achievable noise reduction for three different noise fields, namely coherent, diffused and incoherent noise fields. The noise reduction also depends on the ATFs involved. Hence for each noise field we assume either free-space propagation or reverberant enclosure.

6.1. Coherent noise field

For the coherent noise field the noise PSD is given by $\Phi_{NN}(t, e^{j\omega}) = \phi_{NN}(t, e^{j\omega}) \mathbf{D}(e^{j\omega}) \mathbf{D}^\dagger(e^{j\omega})$. (29)

For $\mathbf{D}(e^{j\omega}) \neq \mathbf{A}(e^{j\omega})$, the achievable noise reduction is infinite, i.e.

$$\phi_{YY}^N(t, e^{j\omega}) = 0 \quad \text{for } \mathbf{D}(e^{j\omega}) \neq \mathbf{A}(e^{j\omega}) \quad (30)$$

regardless of the estimation accuracy of the MBF and BM (provided that the RTFs are modeled as having infinite impulse responses. For a reverberant environment and $\mathbf{D}(e^{j\omega}) \neq \mathbf{A}(e^{j\omega})$ the suppression of a point-source noise signal, obtained by the proposed algorithm, is comparable to that obtainable for a non-reverberant environment using a simple GSC (see for instance Bitzer et al., 1999). The derivation of this result is given in Appendix A. Note, that this is not a surprising result, as the presence of at least single noise reference signal is sufficient for the Wiener filter to completely eliminate the noise component. This condition is met for $M \geq 3$. The infinite noise reduction is obtained for all ATFs $\mathbf{D}(e^{j\omega})$ except for the desired signal's ATFs $\mathbf{D}(e^{j\omega}) = \mathbf{A}(e^{j\omega})$. When the desired source and noise source are located at exactly the same point, namely $\mathbf{D}(e^{j\omega}) = \mathbf{A}(e^{j\omega})$, the noise and desired signal are indistinguishable and no noise reduction can be expected.

It is also interesting to evaluate the noise part of the MBF branch

$$\begin{aligned}
\phi_{\text{MBF}}^N(t, e^{j\omega}) &= E\{Y_{\text{MBF}}(t, e^{j\omega}) Y_{\text{MBF}}^*(t, e^{j\omega})\}|_{Z=N} \\
&= \widehat{\mathbf{W}}_0^\dagger(e^{j\omega}) \Phi_{NN}(t, e^{j\omega}) \widehat{\mathbf{W}}_0(e^{j\omega}) \\
&= \widehat{\mathbf{W}}_0^\dagger(e^{j\omega}) \phi_{NN}(t, e^{j\omega}) \mathbf{D}(e^{j\omega}) \mathbf{D}^\dagger(e^{j\omega}) \widehat{\mathbf{W}}_0(e^{j\omega}) \\
&= \phi_{NN}(t, e^{j\omega}) |\widehat{\mathbf{W}}_0^\dagger(e^{j\omega}) \mathbf{D}(e^{j\omega})|^2. \quad (31)
\end{aligned}$$

The term $|\widehat{\mathbf{W}}_0^\dagger(e^{j\omega}) \mathbf{D}(e^{j\omega})|^2$ depends on the estimated MBF and the ATF $\mathbf{D}(e^{j\omega})$ and may be greater than one. Although the desired signal is coherently combined in the MBF, the signal-to-noise-ratio does not necessarily increase. This phenomenon can be attributed to the particular structure of involved ATFs. The infinite noise reduction, obtained by the DTF-GSC in coherent noise field, is due to the noise canceller branch.

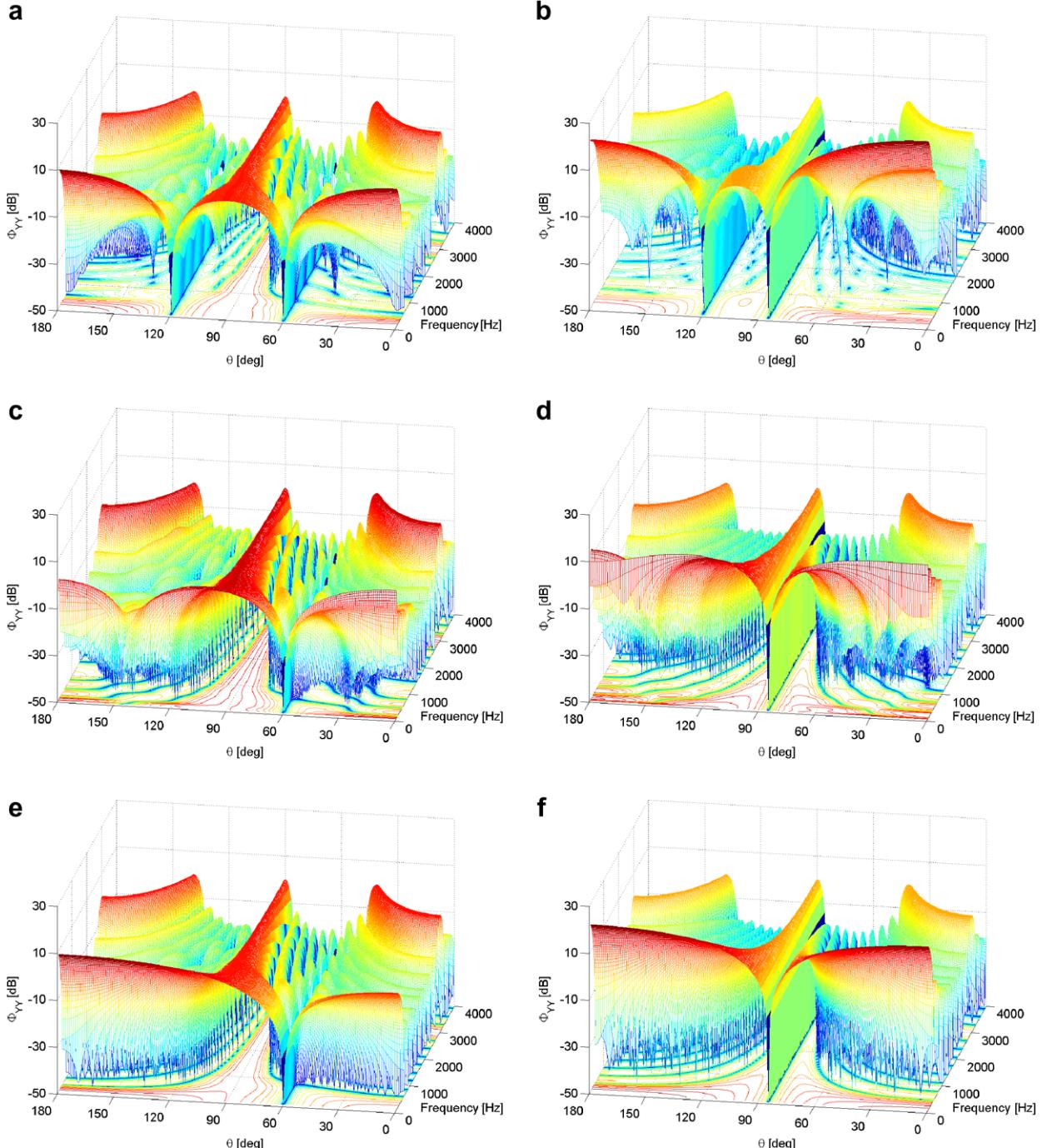


Fig. 9. Array output PSD ϕ_{YY} as a function of the frequency and direction of arrival. Desired signal direction $\theta = 90^\circ$. $M = 10$ sensors. Remote sources: (a) directional noise field ($\theta = 120^\circ$), (c) diffused noise field, (e) incoherent noise field; close sources: (b) directional noise field ($\theta = 120^\circ$), (d) diffused noise field, (f) incoherent noise field.

6.1.1. Free-space propagation

Assume free space propagation, i.e., the ATFs from the desired source and interference source to the sensors are pure delays. The desired signal is assumed to impinge on the array from $\theta = 90^\circ$. We will examine two scenarios. In the first, the non-stationary interference impinges on the array from $\theta = 60^\circ$ (denoted as remote sources scenario), while in the second, $\theta = 85^\circ$ (denoted as close sources scenario). We optimize the array to cancel a noise source from $\theta = 120^\circ$. In Fig. 9a and b we present the PSD of the array output as a function of the frequency and the array steering angle θ . For both scenarios, it is clearly shown that the main lobe is maintained, while null is constructed in all frequencies at the noise angle (in addition to null at the interference angle). Recall that the spatial beamformer response in 500 Hz and 3000 Hz was depicted in Fig. 6a and b for the remote and close sources scenarios, respectively.

It can be clearly seen that the main lobe around the desired direction is wider in 500 Hz than in 3000 Hz. When the directional noise source is away from the desired source, this phenomenon is meaningless. Nevertheless, when the noise arrives from a direction adjacent to the desired source, the BM, constrained to block this direction, prevents the ANC from cancelling the unwanted components. Therefore, since the MBF may enhance noise signals arriving from a direction adjacent to desired source, this could result in an amplification of the noise component at the beamformer output.

6.1.2. Reverberant enclosure

To eliminate the acoustic environment gain, the noise reduction (NR) is calculated by normalizing the output PSD by the ATF that relates the noise source and the first (reference) microphone. We assume that due to the microphones proximity, the arbitrary choice of the first microphone does not have any significant influence on the obtained results. We therefore define the NR in a way similar to the definition of the PSD deviation

$$\text{NR}(t, e^{j\omega}) = \frac{\phi_{YY}^N(t, e^{j\omega})}{|\mathcal{F}(e^{j\omega})|^2 |D_1(e^{j\omega})|^2 \phi_{NN}(t, e^{j\omega})}. \quad (32)$$

A value of $\text{NR}(t, e^{j\omega}) = 0$ indicates complete noise reduction. Note that $\text{NR}(t, e^{j\omega})$ is measured here twice. First, at the MBF output (denoted $\text{NR}_{\text{MBF}}(t, e^{j\omega})$), and second at the beamformer output (denoted $\text{NR}(t, e^{j\omega})$). Results for the two figures-of-merit are introduced in Fig. 10, for a signal source transmitted from the noise direction.

Noise reduction of more than 30 dB is demonstrated in most of the interesting frequency band. It can be noticed that the noise reduction is mainly attributed to the ANC branch. Some noise amplification can be observed at frequencies below 200 Hz. Nevertheless, it should be emphasized that speech signal usually contains low energy in this band.

Note, also, that opposed to the free-space propagation scenario only finite noise reduction is obtained. This degradation is due to the truncation of the impulse response, which corresponds to the RTF.

For the evaluation of the sensitivity of the NR to array misalignments, a beamformer is designed to reject noise signal according to the system introduced in Section 4, while the noise actually impinges the array from a different direction. This evaluation was conducted for three different distances between the assumed noise source and real point: 1.7 cm, 8.7 cm and 26 cm.

Degradation in NR performance for distance of 1.7 cm is depicted in Fig. 11a and b, as measured at the MBF output and beamformer output, respectively. While only marginal degradation can be observed in the MBF performance, severe degradation in performance can be observed at the beamformer output. We conclude that most of the degradation is due to the inaccurate ANC. The degradation in performance increases as the error in the estimation steps up, as depicted in Fig. 11c and d, measured for 8.7 cm estimation error. Similar results are obtained if the noise source is transmitted 26 cm away off the assumed point, as depicted in Fig. 11e and f.

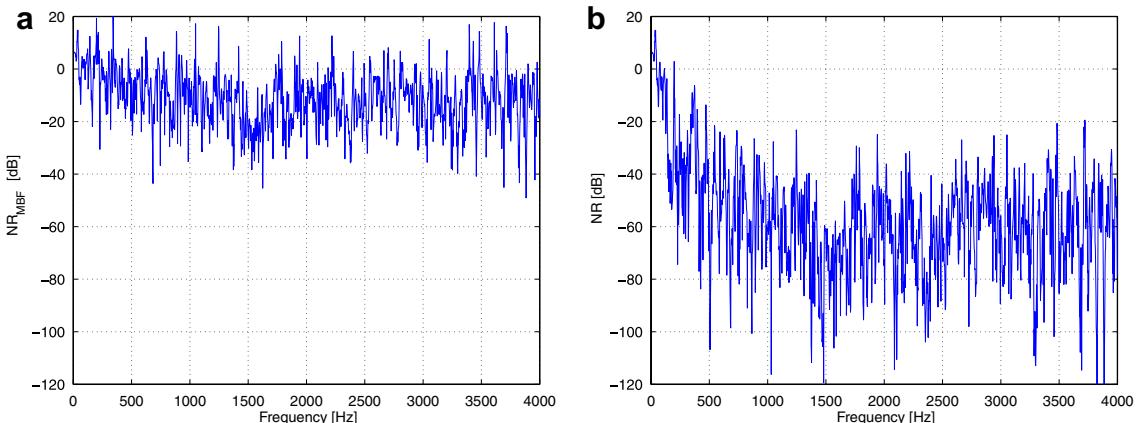


Fig. 10. Expected NR performance for directional noise field: signal received from the noise direction. (a) MBF output PSD and (b) array output PSD.

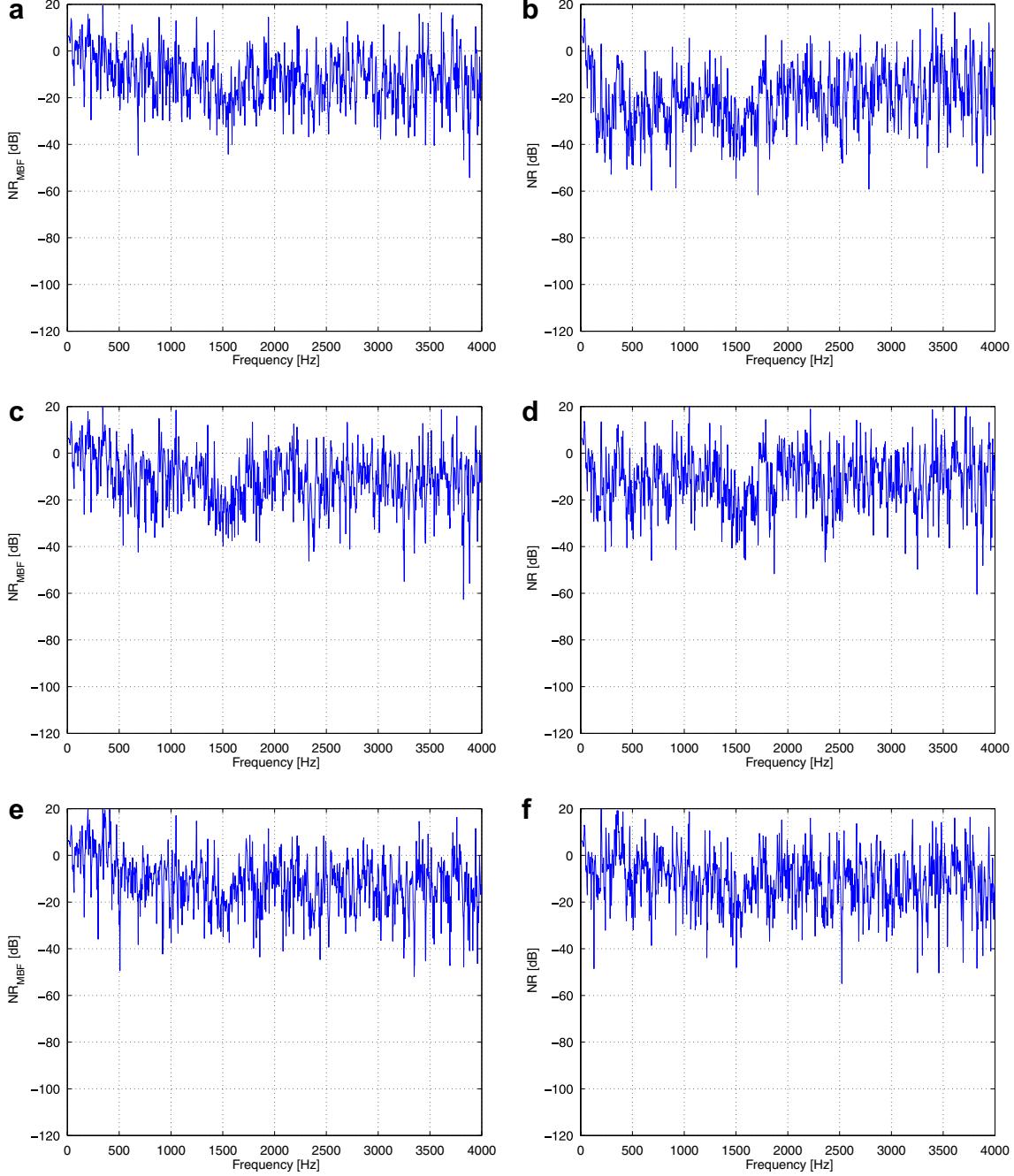


Fig. 11. Expected NR degradation for directional noise field: signal received from 1.7 cm near the noise source: (a) MBF NR degradation, (b) output NR degradation; 8.7 cm near the noise source: (c) MBF NR degradation, (d) output NR degradation; 26 cm near the noise source: (e) MBF NR degradation, (f) output NR degradation.

6.2. Diffused noise field

In the diffused noise field the noise PSD is given by

$$\Phi_{NN}(t, e^{j\omega}) = \phi_{NN}(t, e^{j\omega}) \Gamma(e^{j\omega}). \quad (33)$$

Using the second line in (31) the noise PSD at the MBF output is given by

$$\phi_{MBF}^N(t, e^{j\omega}) = \phi_{NN}(t, e^{j\omega}) \widehat{\mathbf{W}}_0^\dagger(e^{j\omega}) \Gamma(e^{j\omega}) \widehat{\mathbf{W}}_0(e^{j\omega}) \quad (34)$$

and using (28) the noise PSD at the beamformer output is obtained

$$\begin{aligned} \phi_{YY}^N(t, e^{j\omega}) &= \phi_{NN}(t, e^{j\omega}) \widehat{\mathbf{W}}_0^\dagger(e^{j\omega}) \Gamma(e^{j\omega}) \widehat{\mathbf{W}}_0(e^{j\omega}) \\ &\quad - \phi_{NN}(t, e^{j\omega}) \widehat{\mathbf{W}}_0^\dagger(e^{j\omega}) \Gamma(e^{j\omega}) \widehat{\mathcal{H}}(e^{j\omega}) \\ &\quad \times \left(\widehat{\mathcal{H}}^\dagger(e^{j\omega}) \Gamma(e^{j\omega}) \widehat{\mathcal{H}}(e^{j\omega}) \right)^{-1} \widehat{\mathcal{H}}^\dagger(e^{j\omega}) \Gamma(e^{j\omega}) \widehat{\mathbf{W}}_0(e^{j\omega}). \end{aligned} \quad (35)$$

This expression depends on the RTFs and the coherence function $\Gamma(e^{j\omega})$. Since the dependence on the RTFs is

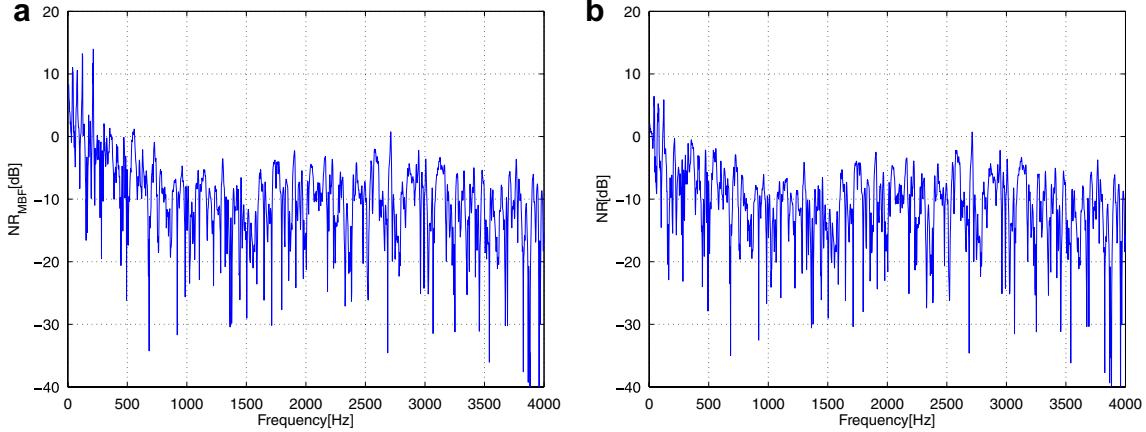


Fig. 12. Expected NR performance for diffused noise field: (a) MBF output PSD, (b) array output PSD.

empirically shown to be marginal compared to the dependence on the noise field, we assume that the correct RTFs are available.

Note that in frequencies above the cutoff frequency $\frac{c}{2\ell_0}$, the elements of $\Gamma(e^{j\omega})$ are almost zero besides the main diagonal elements. In this case, $\phi_{YY}^N(t, e^{j\omega})$ is close to the incoherent noise case, as will be shown in Section 6.3. On the other hand, in lower frequencies, $\Gamma(e^{j\omega})$ components are almost one. In that case, $\phi_{YY}^N(t, e^{j\omega})$ is close to the coherent noise case. In our simulations, the inter-element distance between the microphones is set to 10 cm, and therefore the cutoff frequency is about 1700 Hz.

6.2.1. Free-space propagation

The array was analyzed with diffused noise field by using (35). The case of $M = 10$ sensors is shown in Fig. 9c and d for various steering angles and the entire frequency band. Recall that the diffused noise impinges on the array from all directions. Therefore, the beamformer attenuates almost all directions, except the desired one. More than 10 dB attenuation can be observed for most frequencies in the interesting frequency band. It can be observed that in low frequencies, low attenuation is obtained in the remote sources scenario and even some amplification in the close sources scenario. Note that when considering noise reduction in the presence of diffused noise, wide main lobe as seen in Fig. 6c means that more noise would leak into the beamformer output.

6.2.2. Reverberant enclosure

The MBF output PSD and the beamformer output PSD are given in Fig. 12 for the diffused noise field. It is clearly shown, that the expected performance of the algorithm in diffused field is inferior to the expected performance for point source noise field. This is verified by the experimental study conducted in (Reuven et al., submitted for publication). Most of the noise reduction is obtained by the MBF. Note that the MBF amplifies frequencies below 300 Hz, while some noise reduction is obtained by the lower branch in frequencies below 500 Hz. Nevertheless, it should be emphasized that a speech signal usually con-

tains low energy in this band. Hence, the lower branch renders useless for most of the interesting band in this case.

6.3. Incoherent noise field

In the incoherent noise field the noise PSD is given by $\Phi_{NN}(t, e^{j\omega}) = \phi_{NN}(t, e^{j\omega})I$.

Using (28) and the pre-specified $\Phi_{NN}(t, e^{j\omega})$ we obtain,

$$\begin{aligned} \phi_{YY}^N(t, e^{j\omega}) &= \phi_{NN}(t, e^{j\omega}) \widehat{\mathbf{W}}_0^\dagger(t, e^{j\omega}) \\ &\quad \times \left\{ I - \widehat{\mathcal{H}}(e^{j\omega})(\widehat{\mathcal{H}}^\dagger(e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega}))^{-1}\widehat{\mathcal{H}}^\dagger(e^{j\omega}) \right\} \widehat{\mathbf{W}}_0(t, e^{j\omega}). \end{aligned} \quad (36)$$

Using (6)–(10) it can be verified that $\widehat{\mathbf{A}}^\dagger(e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega}) = \mathbf{0}_{1 \times (M-2)}$, where $\mathbf{0}_{1 \times (M-2)}$ denotes a row vector of $M-2$ zeros. Hence, we observe that

$$\begin{aligned} &\widehat{\mathbf{W}}_0^\dagger(t, e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega}) \\ &= \left[\begin{array}{c} \widehat{\mathbf{A}}^\dagger(e^{j\omega}) - \frac{\widehat{\mathbf{B}}^\dagger(e^{j\omega})\widehat{\mathbf{B}}(e^{j\omega})\widehat{\mathbf{A}}^\dagger(e^{j\omega})}{\|\widehat{\mathbf{A}}^\dagger(e^{j\omega})\|^2 + \|\widehat{\mathbf{B}}^\dagger(e^{j\omega})\|^2} \\ \frac{\mathcal{F}(e^{j\omega})}{1 - \frac{\|\widehat{\mathbf{B}}^\dagger(e^{j\omega})\widehat{\mathbf{A}}^\dagger(e^{j\omega})\|^2}{\|\widehat{\mathbf{A}}^\dagger(e^{j\omega})\|^2 + \|\widehat{\mathbf{B}}^\dagger(e^{j\omega})\|^2}} \end{array} \right]^\dagger \widehat{\mathcal{H}}(e^{j\omega}) \\ &= \mathbf{0}_{1 \times (M-2)}. \end{aligned} \quad (37)$$

Note that the last transition is valid regardless of the estimation accuracy, when both $\widehat{\mathbf{W}}_0(t, e^{j\omega})$ and $\widehat{\mathcal{H}}(e^{j\omega})$ are calculated using the same RTFs estimate $\widehat{\mathbf{A}}(e^{j\omega})$ and $\widehat{\mathbf{B}}(e^{j\omega})$. When $\widehat{Q}_m(e^{j\omega})$ and $\widehat{L}_m(e^{j\omega})$ (the non-trivial terms of $\widehat{\mathcal{H}}(e^{j\omega})$) are estimated directly this may not be guaranteed.⁶

Furthermore, as $\widehat{\mathcal{H}}^\dagger(e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})$ is always a positive matrix, its inverse always exists. Hence, the contribution of the noise cancelling branch is zero, and the noise reduc-

⁶ In (Reuven et al., 2005, submitted for publication) we showed that $Q_m(e^{j\omega})$ and $L_m(e^{j\omega})$ can be estimated in two ways. During double-talk situations these terms can be estimated directly. Whenever only one of the speech signals is present, $\widehat{Q}_m(e^{j\omega})$ and $\widehat{L}_m(e^{j\omega})$ can be calculated by inserting the estimated $\widehat{\mathbf{A}}(e^{j\omega})$ and $\widehat{\mathbf{B}}(e^{j\omega})$ in (9) and (10).

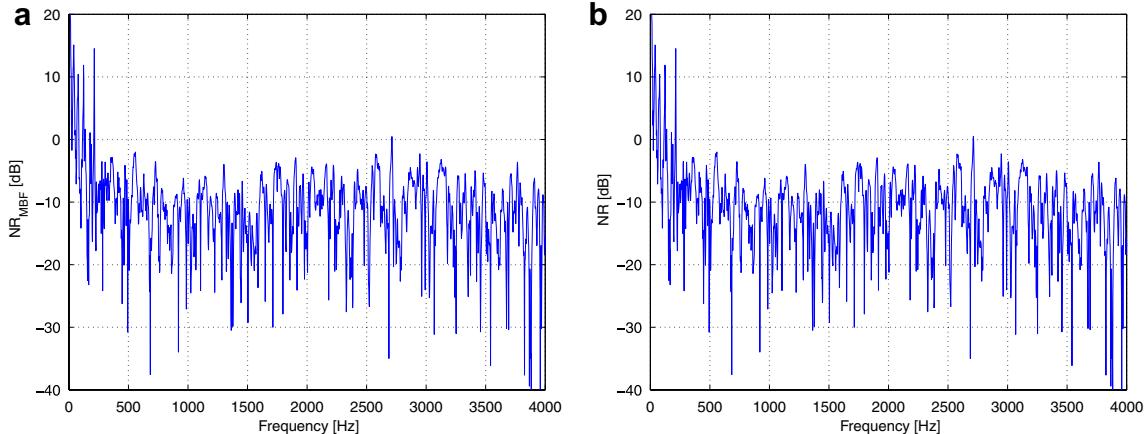


Fig. 13. Expected NR performance for incoherent noise field: (a) MBF output PSD, (b) array output PSD.

tion is only due to the matched beamformer. The noise power at the output is thus

$$\begin{aligned}\phi_{YY}^N(t, e^{j\omega}) &= \phi_{\text{MBF}}^N(t, e^{j\omega}) \\ &= \phi_{NN}(t, e^{j\omega}) \widehat{\mathbf{W}}_0^\dagger(e^{j\omega}) \widehat{\mathbf{W}}_0(e^{j\omega}).\end{aligned}\quad (38)$$

Again, no noise reduction is guaranteed by this structure, and the obtained result depends on the RTFs involved.

6.3.1. Free-space propagation

Array output PSD is depicted in Fig. 9e and f for various steering angles and the entire frequency band. Less than -10 dB white noise gain can be observed for most of the interesting frequency band.

6.3.2. Reverberant enclosure

MBF output and array output PSD for incoherent noise field are depicted in Fig. 13. Since only a minor difference between the two plots can be observed, it can be assumed that no noise reduction is achieved by the noise cancelling branch, and the negligible noise reduction is due to the MBF branch only. Note, that the noise amplification in the lower frequency branch is meaningless, since the speech power at this band is negligible.

6.4. Summary

In this section we evaluated the expected DTF-GSC noise reduction as a function of the noise field. We evaluated three noise fields, namely, directional, diffused, and incoherent. For each noise field we demonstrated the results for both free-space propagation and reverberant enclosure. It is clearly evident that the system is expected to perform very well under coherent (directional) noise field for both environments. However, in the other noise fields the performance of the system is expected to severely degrade.

7. Evaluation of interference reduction

In this section, we evaluate the interference reduction obtained by the proposed algorithm for both reverberant

and non-reverberant environments, and the performance sensitivity to array misalignment.

The interference reduction of the algorithm can be calculated by the general expression given in (20) for a signal $Z(t, e^{j\omega}) = \mathbf{B}(e^{j\omega})S_2(t, e^{j\omega})$. Assume we have an exact knowledge of the RTFs $\widetilde{\mathbf{A}}(e^{j\omega})$ and $\widetilde{\mathbf{B}}(e^{j\omega})$, i.e. $\widetilde{\mathcal{H}}(e^{j\omega}) = \mathcal{H}(e^{j\omega})$ and $\widetilde{\mathbf{W}}_0(e^{j\omega}) = \mathbf{W}_0(e^{j\omega})$. Using $\mathbf{B}^\dagger(e^{j\omega})\mathbf{W}_0(e^{j\omega}) = 0$ and $\mathcal{H}^\dagger(e^{j\omega})\mathbf{W}_0(e^{j\omega}) = 0$, $\phi_{YY}(t, e^{j\omega}) = 0$ can be obtained. Note, that the above result does not depend on the noise field. Furthermore, opposed to noise reduction, the interference reduction is due to the constraints imposed on the MBF and BM alone, rather than minimization or adaptive filtering.

On the other hand, when considering real scenarios, the RTFs estimates are not accurate, and $\widehat{\mathcal{Q}}_m(e^{j\omega})$ and $\widehat{\mathcal{L}}_m(e^{j\omega})$ are not exact estimates of $\mathcal{Q}_m(e^{j\omega})$ and $\mathcal{L}_m(e^{j\omega})$ ($m = 3, \dots, M$), respectively. In this case, the noise field may affect the estimation performance and consequently may affect the interference blocking ability of the BM, producing an interference leakage into the noise reference signals.

We now evaluate the interference reduction sensitivity to array misalignment. Again, we will test the algorithm for both free-space and reverberant environment.

7.1. Free-space propagation

Assume free space propagation, i.e., the ATFs from the desired source and interference source to the sensors are pure delays. Desired signal is assumed to impinge on the array from $\theta = 90^\circ$, while the inter-element distance is 10 cm. We will examine two scenarios. In the first, the interference signal impinges on the array from $\theta = 60^\circ$ (spatially remote sources), while in the second, $\theta = 85^\circ$ (close sources).

The array output PSD in the presence of directional noise is depicted in Fig. 14a and b, for both tested scenarios. In both, a 60 dB null can be observed in the interference direction, $\theta = 60^\circ$ for the remote sources case and $\theta = 85^\circ$ for the close sources case. If the array steering angle will be shifted from the assumed interference direction the

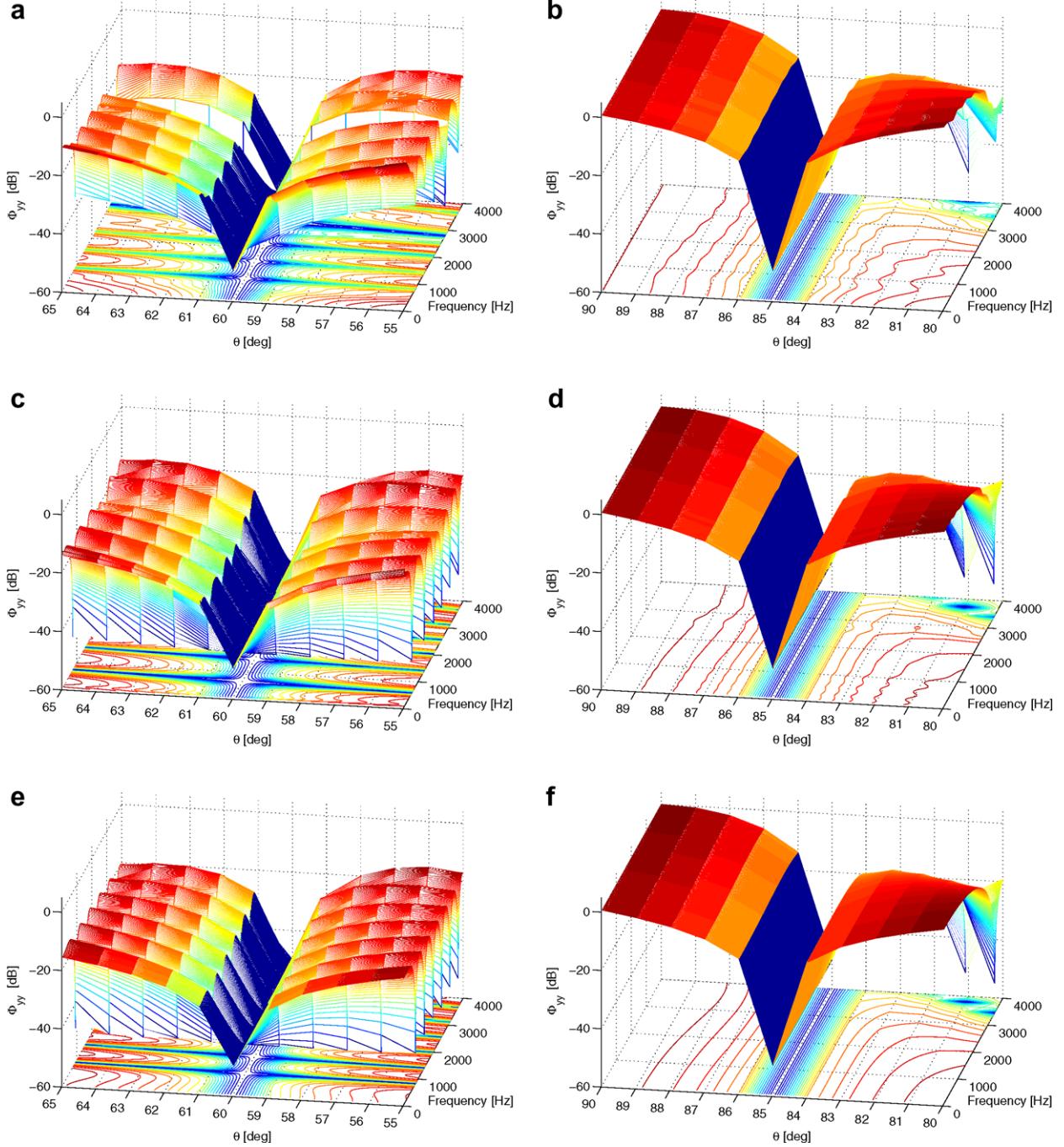


Fig. 14. Array output PSD ϕ_{YY} as a function of the frequency and direction of arrival. $M = 10$ sensors. Remote sources, interference signal direction $\theta = 60^\circ$: (a) directional noise field ($\theta = 120^\circ$), (c) diffused noise field, (e) incoherent noise field; close sources, interference signal direction $\theta = 85^\circ$: (b) directional noise field ($\theta = 120^\circ$), (d) diffused noise field, (f) incoherent noise field.

amount of interference reduction will be reduced. This phenomenon is emphasized in the close sources scenario, since the beampattern is rapidly changing from 0 dB (no distortion) to $-\infty$ (interference cancellation).

The array output PSD in the presence of diffused noise is depicted in Fig. 14c and d, for both tested scenarios. As for the directional noise, a 60 dB null can be observed in the interference direction. Only minor changes can be noticed, compared to the directional noise field. The array output

PSD in the presence of incoherent noise is depicted in Fig. 14e and f, for both tested scenarios. The same trends can be observed for the incoherent noise field.

7.2. Reverberant enclosure

A beamformer was designed to reject an interference signal for the nominal system depicted in Fig. 3, while the actual interference source was located in a different

location. This evaluation was conducted for three different distances between the assumed interference source and real source location: 1.7 cm, 8.5 cm and 17.3 cm.

Note that similar to noise signal, the interference signal may be attenuated due to the acoustic environment. Therefore, the interference reduction is calculated by normalizing the output PSD by the ATF that relates the interference source and the first microphone. We therefore define the normalized interference reduction by

$$\text{IR}(t, e^{j\omega}) = \frac{\phi_{YY}^{S_2}(t, e^{j\omega})}{|\mathcal{F}(e^{j\omega})|^2 |B_1(e^{j\omega})|^2 \phi_{S_2 S_2}(t, e^{j\omega})}. \quad (39)$$

A value of $\text{IR}(t, e^{j\omega}) = 0$ indicates complete interference elimination. This value is obtained whenever an exact knowledge of the RTFs is available.

The results for the three cases are depicted in Fig. 15. The normalized interference reduction $\text{IR}(e^{j\omega})$ is measured at the MBF output (denoted $\text{IR}_{\text{MBF}}(e^{j\omega})$), and at the

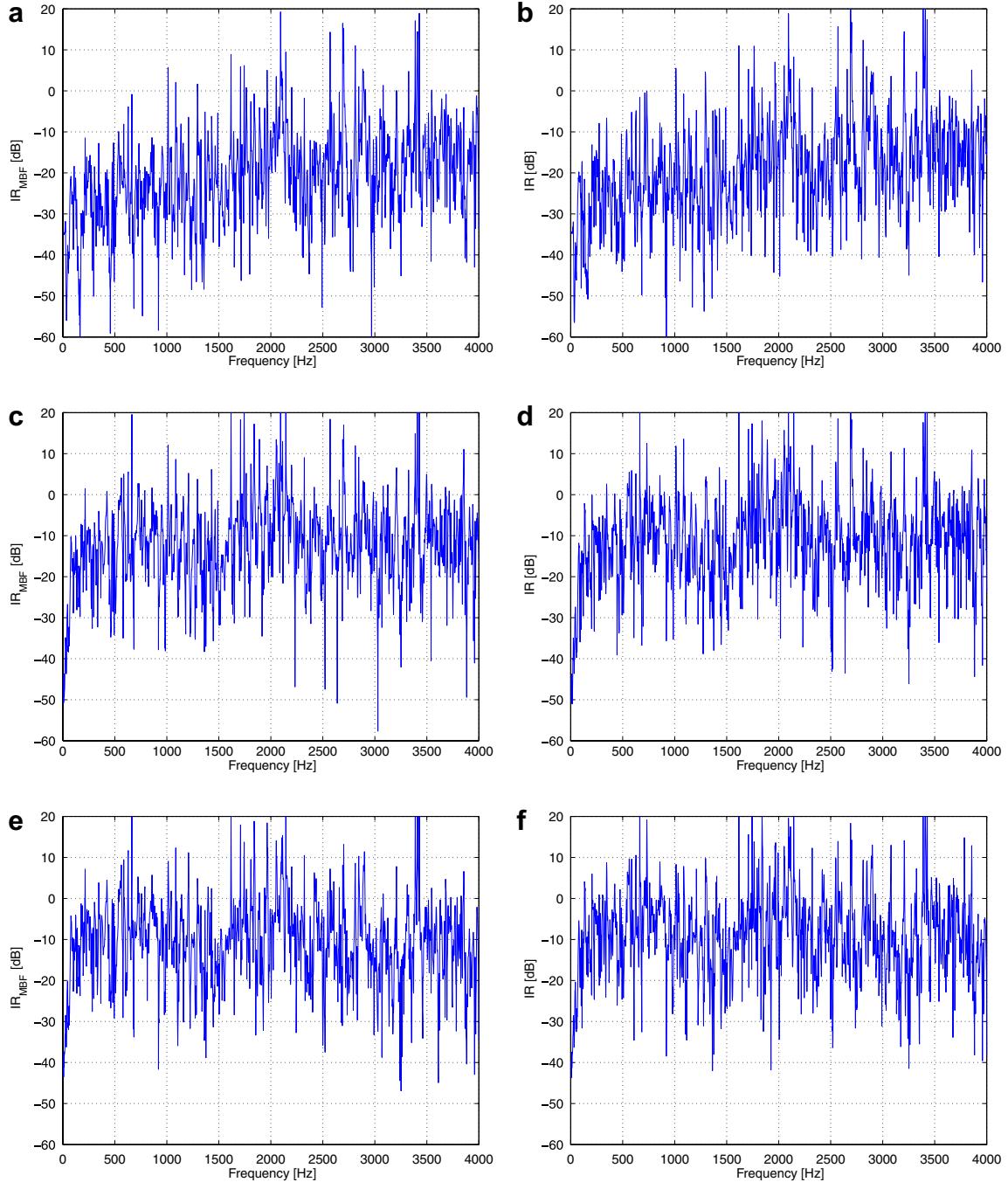


Fig. 15. Expected normalized IR performance for directional noise field: 1.7 cm near interference source (a) MBF output PSD, (b) array output PSD; 8.5 cm near interference source, (c) MBF output PSD, (d) array output PSD; 17.3 cm near interference source (e) MBF output PSD, (f) array output PSD.

beamformer output (denoted $\text{IR}(e^{j\omega})$). The array is designed assuming directional noise field. Note that for all three distances, the interference reduction is mainly due to the MBF alone, since the BM still blocks the interference preventing the ANC from affecting the output. The average normalized interference reduction for the 8.5 cm and 17.3 cm is less than 10 dB. No interference reduction can be measured when the actual interference source is positioned more than 20 cm away from the assumed point.

7.3. Summary

In this section we evaluated the IR for both free-space propagation and reverberant enclosure for three noise fields: directional, diffused, and incoherent. The IR in reverberant environment demonstrates significant sensitivity to array misalignments.

8. Discussion and summary

We have presented a performance evaluation of the recently proposed DTF-GSC (Reuven et al., 2005), aimed at simultaneous interference reduction and noise cancellation. We used three figures-of-merit, namely PSD deviation of the desired signal, noise cancellation, and interference reduction. Due to signal source movements, changes in the acoustic environment, and estimation errors, the DTF-GSC algorithm may employ inaccurate acoustical information. The purpose of this contribution is to provide a comprehensive performance and sensitivity analysis, which can be used in the beamformer design.

We started by deriving a general expression for the DTF-GSC output PSD. From the general expression we derived expressions for the PSD deviation imposed on the desired signal and the amount of achievable noise cancellation and interference reduction. For each figure-of-merit we analyzed the influence of the acoustical environment (namely, either free-space propagation or reverberant enclosure), and the noise field (namely, directional, diffused, and incoherent). We showed that the PSD deviation and interference reduction depend on the estimation accuracy of the RTFs. The sensitivity of the array to inaccurate position estimation is emphasized in the reverberant environment. Actually, even small movement of either the desired or interference sources, may significantly degrade the obtained performance.

The noise reduction depends on the accuracy of estimating the involved RTFs, on the noise field, as well as on the spatial distance between the desired signal and the interference signal. For close sources scenario, the sensitivity to noise source movements is lower than for the remote sources scenario, due to the contradicting constraints imposed on the beampattern in the latter scenario. High noise reduction is obtainable in a coherent noise field, while in an incoherent noise field only the MBF branch is responsible for noise reduction. Although our analysis is predicting infinite NR for the directional noise field in free-space

propagation, in practice this cannot be obtained due to sensor noise. In the reverberant environment the amount of NR is limited due to the truncation of the MBF and BM impulse responses. In a diffused noise field we expect a significant noise reduction due to the ANC branch only in the lower frequency band. However, since the MBF branch might amplify the noise level at this band, the expected overall noise reduction in the lower frequency band is only marginal (in some cases the noise might be even amplified). Fortunately, due to the low speech power in this band, we may overcome this degradation by applying appropriate low-pass filter. In reverberant environments, the NR is more sensitive to array misalignment than in the free-space propagation. This may cause a severe performance degradation when the array stops tracking the sources.

We note that in real-time applications, the DTF-GSC algorithm must use sequential updates of the filters involved rather than the closed form optimal solution. This may degrade the performance of the algorithm and produce inferior results compared to those presented in this study.

Appendix A. Noise reduction for a coherent noise field

Recall (28) and define,

$$\mathcal{X}(t, e^{j\omega}) = \Phi_{NN}(t, e^{j\omega}) \widehat{\mathcal{H}}(e^{j\omega}) \left(\widehat{\mathcal{H}}(e^{j\omega})^\dagger \Phi_{NN}(t, e^{j\omega}) \widehat{\mathcal{H}}(e^{j\omega}) \right)^{-1} \times \widehat{\mathcal{H}}^\dagger(e^{j\omega}) \Phi_{NN}(t, e^{j\omega}),$$

For clarity and simplicity we omit the time and frequency dependencies. Thus \mathcal{X} can be written as a multiplication of three terms

$$\mathcal{X} = \Phi_{NN} \widehat{\mathcal{H}} \left(\widehat{\mathcal{H}}^\dagger \Phi_{NN} \widehat{\mathcal{H}} \right)^{-1} \widehat{\mathcal{H}}^\dagger \Phi_{NN} = \mathcal{K} \times \mathcal{L} \times \mathcal{M},$$

where $\mathcal{K} \triangleq \Phi_{NN} \widehat{\mathcal{H}}$, $\mathcal{L} \triangleq (\widehat{\mathcal{H}}^\dagger \Phi_{NN} \widehat{\mathcal{H}})^{-1}$ and $\mathcal{M} \triangleq \widehat{\mathcal{H}}^\dagger \Phi_{NN}$.

The PSD matrix of the directional noise component at the sensors is given by

$$\Phi_{NN} = \phi_{NN} \mathbf{D} \mathbf{D}^\dagger + \varepsilon \mathbf{I}$$

where \mathbf{I} is an $M \times M$ identity matrix, and $\varepsilon \rightarrow 0$. The last term is added to allow for the calculation of the involved matrix inversion operations.

We start by calculating \mathcal{L}

$$\begin{aligned} \mathcal{L} &= \left(\widehat{\mathcal{H}}^\dagger (\phi_{NN} \mathbf{D} \mathbf{D}^\dagger + \varepsilon \mathbf{I}) \widehat{\mathcal{H}} \right)^{-1} \\ &= \left(\phi_{NN} (\widehat{\mathcal{H}}^\dagger \mathbf{D}) (\widehat{\mathcal{H}}^\dagger \mathbf{D})^\dagger + \varepsilon \widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}} \right)^{-1}. \end{aligned}$$

If $\mathbf{D}(e^{j\omega}) = \mathbf{A}(e^{j\omega})$, i.e. the noise source is located exactly at the desired signal location, then, provided that the blocking matrix $\widehat{\mathcal{H}}(e^{j\omega})$ is an accurate estimate of $\mathcal{H}(e^{j\omega})$, $\widehat{\mathcal{H}}^\dagger(e^{j\omega}) \mathbf{D}(e^{j\omega}) = 0$. Hence, the involved calculations simplify to $\mathcal{L} = (\varepsilon \widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1}$, $\mathcal{K} = \varepsilon \widehat{\mathcal{H}}$ and $\mathcal{M} = \varepsilon \widehat{\mathcal{H}}^\dagger$. Collecting terms, and using the invertibility of $\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}}$ we obtain, $\mathcal{X} = \frac{\varepsilon^2}{\varepsilon} \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger \xrightarrow{\varepsilon \rightarrow 0} 0$. Thus, no extra noise reduction is produced by the noise canceller branch, as expected, and the total noise PSD at the output is given by

$$\phi_{YY}^N(t, e^{j\omega}) = \phi_{NN}(t, e^{j\omega}) |\mathcal{F}(e^{j\omega})|^2 |A_1(e^{j\omega})|^2,$$

which is exactly the signal contribution derived in Section 5, with the input signal PSD $\phi_{SS}(t, e^{j\omega})$ replaced by the respective noise PSD $\phi_{NN}(t, e^{j\omega})$.

For the general case, when $D(e^{j\omega}) \neq A(e^{j\omega})$, we use the *Matrix Inversion Lemma*, yielding

$$\mathcal{L} = \left(\frac{1}{\varepsilon} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} - \frac{\frac{1}{\varepsilon} \phi_{NN} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1}}{1 + \frac{1}{\varepsilon} \phi_{NN} D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D} \right).$$

Now, using the approximation $\frac{1}{1+\mu} \approx 1 - \mu$, for $\mu \rightarrow 0$ (μ properly defined), yields,

$$\begin{aligned} \mathcal{L} = & \left(\frac{1}{\varepsilon} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} - \frac{\frac{1}{\varepsilon} \phi_{NN} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1}}{\phi_{NN} D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D} \right. \\ & \left. + \frac{\phi_{NN} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1}}{(\phi_{NN} D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D)^2} \right). \end{aligned}$$

Now, calculating \mathcal{X} ,

$$\begin{aligned} \mathcal{X} &= \mathcal{K} \mathcal{L} \mathcal{M} = \Phi_{NN} \widehat{\mathcal{H}} \mathcal{L} \widehat{\mathcal{H}}^\dagger \Phi_{NN} \\ &= (\phi_{NN} D D^\dagger + \varepsilon \mathbf{I}) \widehat{\mathcal{H}} \mathcal{L} \widehat{\mathcal{H}}^\dagger (\phi_{NN} D D^\dagger + \varepsilon \mathbf{I}) \\ &\triangleq (\mathcal{K}_1 + \mathcal{K}_2) \widehat{\mathcal{H}} (\mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3) \widehat{\mathcal{H}}^\dagger (\mathcal{M}_1 + \mathcal{M}_2) \end{aligned}$$

with the obvious definitions of \mathcal{K}_1 , \mathcal{K}_2 , \mathcal{L}_1 , \mathcal{L}_2 , \mathcal{L}_3 , \mathcal{M}_1 , \mathcal{M}_2 . Opening the brackets we have 12 terms:

- (I) $\mathcal{K}_1 \widehat{\mathcal{H}} \mathcal{L}_1 \widehat{\mathcal{H}}^\dagger \mathcal{M}_1 = \frac{1}{\varepsilon} \phi_{NN}^2 D D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D D^\dagger,$
- (II) $\mathcal{K}_1 \widehat{\mathcal{H}} \mathcal{L}_2 \widehat{\mathcal{H}}^\dagger \mathcal{M}_1 = -\frac{1}{\varepsilon} \phi_{NN}^3$
 $\times \frac{D D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} D D^\dagger}{\phi_{NN} D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D}$
 $= -\frac{1}{\varepsilon} \phi_{NN}^2 D D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D D^\dagger,$
- (III) $\mathcal{K}_1 \widehat{\mathcal{H}} \mathcal{L}_3 \widehat{\mathcal{H}}^\dagger \mathcal{M}_1 = \frac{\phi_{NN}^3 D (D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D)^2 D^\dagger}{(\phi_{NN} D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D)^2} \phi_{NN} D D^\dagger,$
- (IV) $\mathcal{K}_2 \widehat{\mathcal{H}} \mathcal{L}_1 \widehat{\mathcal{H}}^\dagger \mathcal{M}_1 = \phi_{NN} \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D D^\dagger,$
- (V) $\mathcal{K}_2 \widehat{\mathcal{H}} \mathcal{L}_2 \widehat{\mathcal{H}}^\dagger \mathcal{M}_1$
 $= -\phi_{NN}^2 \frac{\widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} D D^\dagger}{\phi_{NN} D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D}$
 $- \phi_{NN} \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D D^\dagger,$
- (VI) $\mathcal{K}_2 \widehat{\mathcal{H}} \mathcal{L}_3 \widehat{\mathcal{H}}^\dagger \mathcal{M}_1$
 $= \frac{\phi_{NN}^2 \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D D^\dagger}{(\phi_{NN} D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D)^2}$
 $\times \varepsilon \frac{\widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D D^\dagger}{D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D},$
- (VII) $\mathcal{K}_1 \widehat{\mathcal{H}} \mathcal{L}_1 \widehat{\mathcal{H}}^\dagger \mathcal{M}_2 = \phi_{NN} D D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger,$

- (VIII) $\mathcal{K}_1 \widehat{\mathcal{H}} \mathcal{L}_2 \widehat{\mathcal{H}}^\dagger \mathcal{M}_2$
 $= -\phi_{NN}^2 \frac{D D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger}{\phi_{NN} D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D}$
 $= -\phi_{NN} D D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger,$
- (IX) $\mathcal{K}_1 \widehat{\mathcal{H}} \mathcal{L}_3 \widehat{\mathcal{H}}^\dagger \mathcal{M}_2$
 $= \varepsilon \frac{\phi_{NN}^2 D D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger}{(\phi_{NN} D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D)^2}$
 $\times \varepsilon \frac{D D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger}{D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger},$
- (X) $\mathcal{K}_2 \widehat{\mathcal{H}} \mathcal{L}_1 \widehat{\mathcal{H}}^\dagger \mathcal{M}_2$
 $= \varepsilon \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger,$
- (XI) $\mathcal{K}_2 \widehat{\mathcal{H}} \mathcal{L}_2 \widehat{\mathcal{H}}^\dagger \mathcal{M}_2$
 $= -\varepsilon \frac{\widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger}{D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D},$
- (XII) $\mathcal{K}_2 \widehat{\mathcal{H}} \mathcal{L}_3 \widehat{\mathcal{H}}^\dagger \mathcal{M}_2$
 $= \varepsilon^2 \frac{\phi_{NN} \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger}{(\phi_{NN} D^\dagger \widehat{\mathcal{H}} (\widehat{\mathcal{H}}^\dagger \widehat{\mathcal{H}})^{-1} \widehat{\mathcal{H}}^\dagger D)^2}$

Note that terms (I) and (II), the terms (IV) and (V), and terms (VII) and (VIII) eliminate each other, respectively. Note also, that since $\widehat{\mathcal{H}}^\dagger D \neq 0$, the terms (VI), (IX), (X), (XI) and (XII) approach zero as $\varepsilon \rightarrow 0$. Therefore only term (III) remains, resulting in $\mathcal{X} = \phi_{NN} D D^\dagger$. Substituting \mathcal{X} in (28) we have

$$\begin{aligned} \phi_{YY}^N(t, e^{j\omega}) &= \widehat{\mathbf{W}}_0^\dagger(e^{j\omega}) (\Phi_{NN}(t, e^{j\omega}) - \phi_{NN}(t, e^{j\omega}) D(e^{j\omega}) D^\dagger \\ &\quad \times (e^{j\omega})) \widehat{\mathbf{W}}_0(e^{j\omega}) = 0. \end{aligned}$$

Namely, the noise part is completely eliminated, regardless of the estimation accuracy of the MBF and BM. In actual scenarios, where only a truncated version of the involved filters is used, the infinite NR can be only approximated.

References

- Allen, J., Berkley, D., 1979. Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Amer.* 65 (4), 943–950.
- Bitzer, J., Simmer, K., Kammeyer, K., 1998. Multichannel noise reduction – algorithms and theoretical limits. In: Proc. European Signal Processing Conf. (EUSIPCO), Vol. I, Rhodes, Greece, pp. 105–108.
- Bitzer, J., Kammeyer, K.-D., Simmer, K., 1999. An alternative implementation of the superdirective beamformer. In: IEEE Workshop on Application of Signal Processing to Audio and Acoustics, New Paltz, New York, USA.
- Bitzer, J., Simmer, K., Kammeyer, K., 1999. Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement. Proc. Internat. Conf. on Acoustics, Speech and Signal Processing (ICASSP). IEEE, Phoenix, AZ, USA, pp. 2965–2968.
- Dal-Degan, N., Prati, C., 1988. Acoustic noise analysis and speech enhancement techniques for mobile radio application. *Signal Process.* 15 (4), 43–56.
- Doclo, S., Moonen, M., 2002. GSVD-based optimal filtering for single and multimicrophone speech enhancement. *IEEE Trans. Speech Process.* 50 (9), 2230–2244.
- Gannot, S., Burshtein, D., Weinstein, E., 2001. Signal enhancement using beamforming and nonstationarity with applications to speech. *IEEE Trans. Signal Process.* 49 (8), 1614–1626.

- Gannot, S., Burshtein, D., Weinstein, E., 2004. Analysis of the power spectral deviation of the general transfer function GSC. *IEEE Trans. Signal Process.* 52 (4), 1115–1121.
- Griffiths, L.J., Jim, C.W., 1982. An alternative approach to linearly constrained adaptive beamforming. *IEEE Trans. Antennas Propagat.* 30 (1), 27–34.
- Huang, K.-C., Yeh, C.-C., 1990. Performance analysis of derivative constraint adaptive arrays with pointing errors. *IEEE Trans. Acoust. Speech Signal Process.* 38 (2), 209–219.
- Nordholm, S., Leung, Y.H., 2000. Performance limits of the broadband generalized sidelobe cancelling structure in an isotropic noise field. *J. Acoust. Soc. Amer.* 107 (2), 1057–1060.
- Nordholm, S., Claesson, I., Eriksson, P., 1992. The broadband Wiener solution for Griffiths–Jim beamformers. *IEEE Trans. Signal Process.* 40 (2), 474–478.
- Nordholm, S., Claesson, I., Bengtsson, B., 1993. Adaptive array noise suppression of handsfree speaker input in cars. *IEEE Trans. Vehicular Tech.* 42 (4), 514–518.
- Nordholm, S., Claesson, I., Dahl, M., 1999. Adaptive microphone array employing calibration signals: an analytical evaluation. *IEEE Trans. Speech Audio Process.* 7 (3), 241–252.
- Reuven, G., Gannot, S., Cohen, I., 2005. Dual source TF-GSC and its application to echo cancellation. In: The Internat. Workshop on Acoustic Echo and Noise Control (IWAENC), Eindhoven, The Netherlands, pp. 27–30.
- Reuven, G., Gannot, S., Cohen, I., submitted for publication. Dual source transfer-function generalized sidelobe canceller. *IEEE Trans. Speech Audio Process.*
- Spriet, A., Moonen, M., Wouters, J., 2004. Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction. *Signal Process.* 84 (12), 2367–2387.
- Spriet, A., Moonen, M., Wouters, J., 2005. Robustness analysis of multichannel Wiener filtering and generalized sidelobe cancellation for multimicrophone noise reduction in hearing aid applications. *IEEE Trans. Speech Audio Process.* 13 (4), 487–503.