Katholieke Universiteit Leuven

Departement Elektrotechniek ESAT-SISTA/TR 2002-81

Robustness Analysis of Multi-channel Wiener Filtering and Generalized Sidelobe Cancellation for Multi-microphone Noise Reduction in Hearing Aid Applications [1]

Ann Spriet[2], Marc Moonen[3] ,Jan Wouters[4]

September 2002
Published in IEEE Transactions on Speech and Audio Processing, July 2005, vol. 13, no. 4, pp. 487-503.

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/sista/spriet/reports/02-81.pdf*

[2]K.U.Leuven, Dept. of Electrical Engineering (ESAT), SISTA, Kasteelpark Arenberg 10, 3001 Leuven-Heverlee, Belgium, Tel. 32/16/32 18 99, Fax 32/16/32 19 70, WWW: *http://www.esat.kuleuven.ac.be/sista*. E-mail: *ann.spriet@esat.kuleuven.ac.be*. K.U.Leuven, Lab. Exp. ORL, Dept. Neurowetenschappen, Kapucijnenvoer 33, 3000 Leuven, Belgium, Tel. 32/16/33 24 15, Fax 32/16/33 23 35, WWW: *http://www.kuleuven.ac.be/exporl/Lab/Default.htm*. From 1999 until 2003, Ann Spriet was supported by FWO-Vlaanderen. This research work was carried out at the ESAT laboratory and Lab. Exp. ORL of the Katholieke Universiteit Leuven, in the frame of Interuniversity Attraction Pole (2002-2007) - IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modelling'), the Concerted Research Action GOA-MEFISTO-666 (Mathematical Engineering for Information and Communication Systems Technology) of the Flemish Government, Research Project FWO nr. G.0233.01 ('Signal processing and automatic patient fitting for advanced auditory prostheses') and IWT project 020540 ('Innovative speech processing algorithms for improved performance of cochlear implants').

[3]K.U.Leuven, Dept. of Electrical Engineering (ESAT), SISTA, Kasteelpark Arenberg 10, 3001 Heverlee, Belgium, Tel. 32/16/32 17 09, Fax 32/16/32 19 70, WWW: *http://www.esat.kuleuven.ac.be/sista*. E-mail: *marc.moonen@esat.kuleuven.ac.be*. Marc Moonen is a professor at the Katholieke Universiteit Leuven.

[4]K.U.Leuven, Lab. Exp. ORL, Dept. Neurowetenschappen, Kapucijnenvoer 33, 3000 Leuven, Belgium, Tel. 32/16/33 23 42, Fax 32/16/33 23 35, WWW: *http://www.kuleuven.ac.be/exporl/Lab/Default.htm* E-mail: *jan.wouters@uz.kuleuven.ac.be*. Jan Wouters is a professor at the Katholieke Universiteit Leuven.

Robustness Analysis of Multi-channel Wiener Filtering and Generalized Sidelobe Cancellation for Multi-microphone Noise Reduction in Hearing Aid Applications

Ann Spriet*[1,2], *Student Member, IEEE,* Marc Moonen[1], *Member, IEEE,* Jan Wouters[2]

[1]ESAT/SCD-SISTA, K.U. Leuven

Kasteelpark Arenberg 10

B-3001 Leuven-Heverlee, Belgium

Tel.: +32 16 321795, Fax: +32 16 321970

Email: ann.spriet@esat.kuleuven.ac.be; marc.moonen@esat.kuleuven.ac.be

[2]Lab. Exp. ORL / ENT-Dept., K.U. Leuven

Kapucijnenvoer 33, B-3000 Leuven, Belgium

Email: jan.wouters@uz.kuleuven.ac.be

Manuscript received September 17, 2002; revised August 6, 2003. From 1999 until 2003, Ann Spriet was supported by FWO-Vlaanderen. This research work was carried out at the ESAT laboratory and Lab. Exp. ORL of the K. U. Leuven, in the frame of Interuniversity Attraction Pole (2002-2007) - IUAP P5/22 (‘Dynamical Systems and Control: Computation, Identification and Modelling’), the Concerted Research Action GOA-MEFISTO-666 (Mathematical Engineering for Information and Communication Systems Technology) of the Flemish Government, Research Project FWO nr. G.0233.01 (‘Signal processing and automatic patient fitting for advanced auditory prostheses’) and IWT project 020540 (’Innovative speech processing algorithms for improved performance of cochlear implants’). The scientific responsibility is assumed by its authors. The associate editor coordinating the review of this paper and approving it for publication was Prof. Maurizio Omologo.

A. Spriet and M. Moonen are with the Katholieke Universiteit Leuven, Department of Electrical Engineering (ESAT-SCD), B-3001 Leuven, Belgium (e-mail: ann.spriet@esat.kuleuven.ac.be, marc.moonen@esat.kuleuven.ac.be).

A. Spriet and J. Wouters are with the Katholieke Universiteit Leuven, ENT-Department (Lab. Exp. ORL), B-3000 Leuven, Belgium (e-mail: jan.wouters@uz.kuleuven.ac.be).

Abstract

For small-sized arrays such as hearing aids, noise reduction is obtained at the expense of an increased sensitivity to errors in the assumed signal model, i.e. microphone mismatch, variations in speaker and microphone positions, reverberation. However, the noise reduction algorithm should still be robust, i.e. insensitive to small signal model errors. In this paper, we evaluate the robustness of the Generalized Sidelobe Canceller (GSC) and a recently developed Multi-channel Wiener Filtering (MWF) technique for hearing aid applications both analytically and experimentally. The analysis reveals that robustness of the GSC is especially crucial in complicated noise scenarios and that microphone mismatch is particularly harmful to the GSC, even when the adaptive noise canceller is adapted during noise only. Hence, a constraint on the noise sensitivity of the GSC is essential, at the expense of less noise reduction. The MWF on the other hand, is not affected by microphone mismatch and has a potential benefit over the robust GSC with noise sensitivity constraint. However, the MWF is sensitive to the estimation accuracy of the second order statistics of speech and noise so that its benefit may be lost in non-stationary noise scenarios.

Index Terms

speech enhancement, robustness, generalized sidelobe canceller, multi-channel Wiener filtering, hearing aids

I. INTRODUCTION

Noise reduction algorithms are crucial in hearing aids to improve the intelligibility in background noise. Multi-microphone systems exploit spatial information in addition to temporal and spectral information of the desired signal and noise signal, and are thus preferred to single microphone procedures (such as spectral subtraction). In hearing aids, the array length is limited because of aesthetical and practical reasons. In [1], [2], it is shown that noise can be reduced even at low $f \cdot d$, i.e. $f \cdot d \ll \frac{c}{2}$ (with f the frequency [Hz], d the microphone interspacing [m] and c the velocity of sound (≈ 340 m/s)), be it at the expense of an increased noise sensitivity. The *noise sensitivity* is defined as the ratio of the spatially white noise gain to the gain of the desired signal [2], [3]. It is used to quantify the sensitivity to errors in the assumed signal model and processing errors (such as microphone mismatch, round off errors due to finite processing accuracy). A noise reduction algorithm should however be *robust*, i.e. insensitive to small signal model errors. The increased noise sensitivity shows that the *robustness* of multi-microphone noise reduction algorithms may be crucial for practical use in hearing aids.

A widely studied, adaptive noise reduction algorithm for hearing aids is the Linearly Constrained Minimum Variance (LCMV) beamformer [2]-[13], which is often implemented using the *Generalized Sidelobe Canceller (GSC)* [14]. The GSC consists of a fixed, spatial pre-processor, which includes a fixed beamformer and a blocking matrix, and an Adaptive Noise Canceller (ANC). The LCMV beamformer or GSC assumes the desired speaker location, as well as the microphone characteristics and positions to be known a priori, and reverberation to be absent. In reality, these conditions are seldom completely fulfilled so that leakage of speech into the noise references occurs and causes speech distortion by the ANC. Therefore, different variants of the standard GSC have been proposed. Speech leakage in the presence of misteer, microphone placement uncertainty and reverberation is reduced by incorporating multiple linear constraints in the design of the fixed spatial pre-processor [8], by using a spatial filter designed blocking matrix [15] or an adaptive blocking matrix [11], [16], [17]. Although these techniques reduce the amount of speech leakage, it can never be completely avoided. Hence, additional measures are required to limit speech distortion caused by the ANC: the adaptive filters are only adapted during periods of noise only [5], [7], [16] or/and the noise sensitivity is constrained by applying a quadratic inequality constraint to the ANC [3], [8], [11], [12], [16] or by injection of spatially uncorrelated noise [13]. However, this noise sensitivity constraint goes at the expense of less noise reduction. In the remainder, we concentrate on the widely studied *standard GSC with ANC adapted during noise only* with and without noise sensitivity constraint. In the sequel, we refer to the *GSC with Noise Sensitivity Constraint as GSC-NSC*.

Another category of adaptive multi-microphone noise reduction techniques is *Multi-channel Wiener Filtering* (MWF) [18]-[26]. These techniques provide a Minimum Mean Square Error (MMSE) estimate of the clean *source* speech signal, the (reverberant) speech component in one of the microphone signals or a reference signal. Among the different MWF techniques, we can distinguish between three classes, i.e., GSC with single-channel post-filter [18], [26], MWF using calibration sequences [24], [25] and MWF with unknown reference [19]-[23]. In [18], the *GSC is cascaded with a single-channel post-filter* to provide an MMSE estimate of the source speech signal. This MWF technique relies on the same assumptions about the desired signal model as the GSC and is thus equivalently robust against errors in the assumed signal model. In [24], [25], an *MWF is proposed that uses calibration sequences*. The calibration sequences are recorded through the actual system when no noise is present. The MWF then provides an MMSE estimate of a calibration observation measured at one of the microphones or at a separate microphone located at the source position. The calibration scenario is assumed to be representative of the practical scenario (i.e., same microphone characteristics, same nominal position and long-term Power Spectral Density (PSD) of the desired source, same room acoustics etc.). Like for the GSC, deviations from these assumptions

will result in performance loss. In this paper, the work is only concerned about the third class of MWF, i.e., *MWF with unknown reference*: in contrast to the GSC and the previous MWF techniques, neither a priori information about the desired signal model nor calibration is required [19]-[23]. This MWF provides an MMSE estimate of the (reverberant) speech component in one of the microphone signals. It is uniquely based on estimates of the second order statistics of the speech and the noise signal. To obtain these estimates, the MWF assumes that the second order statistics of the noise are sufficiently stationary so that they can be estimated during periods of noise only. Like for the GSC, a robust speech detection is thus needed. In [19]-[22], (recursive) implementations based on a Generalized Singular Value Decomposition (GSVD) or a QR Decomposition (QRD) of a speech + noise data matrix and a noise data matrix have been proposed. A subband implementation [23] results in a significantly lower cost compared to the fullband approach, making it suitable for future hearing aids. Simulations with a 5-microphone array of 20 cm showed that the GSVD based algorithm is indeed more robust than the GSC [19] in case of significantly violated assumptions. Compared to the GSC, no or a less severe constraint on the noise sensitivity is expected to be needed so that more noise may be reduced. However, since the MWF does not rely on a priori information, it is expected to be sensitive to the accuracy of the estimated statistics.

Analytical calculations have drawn the attention of many researchers to predict the optimal (i.e., stationary signals, infinite filter length) *noise reduction performance* achievable by single and multi-microphone noise reduction systems [2], [27]-[34]. In [2], [28], [29], [31], the theoretically optimal noise reduction performance of the GSC has been derived as a function of the complex coherence functions between the microphone noise signals. In [26], [29], a similar theoretical analysis evaluates a microphone array with post-filtering. In [13], [32], [33], [34, Chapter 6] the sensitivity of Minimum Variance Distortionless Response (MVDR) beamforming and GSC[1] against signal model errors has been analytically investigated for antenna array applications. Most of these analyses focus on an MVDR beamformer that minimizes the total output power and not the output noise power: then, performance degradation due to signal model errors especially occurs for high input SNR [8], [13], [32]-[35]. The sensitivity to errors is significantly reduced when only the output noise power is minimized [34, Chapter 6], which is obtained by the GSC with ANC adapted during noise only.

In this paper, we develop a unified theoretical framework based on the complex coherence functions to assess both the *broadband performance* of the GSC with ANC adapted during noise only, the GSC-NSC and the MWF and their *robustness* to different types of model errors for different noise scenarios. The

[1]Note that MVDR beamforming can be considered as a special case of LCMV beamforming or GSC [4].

robustness is analyzed for *small-sized arrays as used in hearing aids and in terms of speech intelligibility weighted performance measures*. The theoretical results are verified through experiments with a Behind-The-Ear (BTE) hearing aid.

The paper is organized as follows. Section II describes the notation and the signal model. In Section III, the optimal noise reduction performance of the GSC is briefly reviewed and the formulas are extended to also quantify noise sensitivity, speech distortion and intelligibility weighted performance measures. In addition, the optimal performance of the GSC-NSC is derived. Section IV reviews the concept of the MWF and derives formulas for its optimal performance. In Section V, theoretical analysis of the noise sensitivity determines under which conditions (e.g. noise scenario, microphone configuration, etc.) the noise sensitivity increases and hence robustness issues become crucial for the GSC: especially complicated noise scenarios and low $f \cdot d$ result in a high noise sensitivity. Among the different signal model errors, microphone gain mismatch in particular is harmful to the GSC, even if the ANC is adapted during noise-only. As a consequence, a noise sensitivity constraint is essential, which goes at the expense of less noise reduction. The MWF on the other hand, is not degraded by microphone mismatch, so that - compared to a robust GSC-NSC - more noise is reduced. The results above assume perfect estimation of the second order statistics of speech and noise by the GSC and the MWF. Because of the non-continuous estimation of the speech + noise and noise statistics, non-stationarities in the noise may cause estimation errors in the second order statistics. Simulations indicate that the MWF is quite sensitive to such estimation errors so that its benefit over the GSC-NSC may be lost in non-stationary noise scenarios. The theoretical analysis assumes stationary signals, free field propagation of speech and noise sources and infinite filter lengths. In practice, finite filter lengths are used, reverberation is present and signals are non-stationary. In addition, the second order statistics of speech + noise and noise only are estimated from a finite amount of incoming data. Therefore, the theory is validated in Section VI based on real recordings made in a reverberant office room with a 3-microphone BTE mounted on a dummy head. To allow for time-varying environments, exponential weighting is incorporated in the GSC and the MWF. The experimental results are found to be well predicted by the theoretical analysis. In addition, the experiments show that the MWF preserves its benefit for spectrally non-stationary noise, provided that the environment and long-term average noise PSD vary relatively slowly in time (on the order of seconds). In highly non-stationary environments, the benefit of the MWF over the GSC-NSC may be lost.

II. NOTATION AND SIGNAL MODEL

A. Notation

The complex conjugate of a scalar B is denoted as B^*. The superscript T and H denote the transpose and conjugate transpose operator, respectively.

In the theoretical analysis, we assume stationary signals and infinite filters so that the signals and filters can be represented in the frequency domain. The microphone signals are $X_k(f)$, $k = 1, ..., M$ with M the number of microphones. The PSD of signal $X(f)$ is $P_X(f)$, the cross-PSD between signals $X(f)$ and $Y(f)$ is $P_{XY}(f) = \mathcal{E}\{X(f)Y^*(f)\}$. Where needed, the superscripts s and n are used to refer to the contribution of the speech and noise signal only. The noise signal consists of external noise (superscript e) and internal noise (superscript i) e.g. sensor noise, quantization noise, etc. modelled as spatially white noise. We assume a homogeneous speech and noise field, i.e., that the PSD of the received speech, $P_X^s(f)$, the internal and external noise $P_X^i(f)$ and $P_X^e(f)$, the received noise, $P_X^n(f) = P_X^i(f) + P_X^e(f)$, and the received microphone signals $P_X(f) = P_X^s(f) + P_X^n(f)$ are the same at each microphone.

The output of the noise reduction algorithm will be called $Z(f)$. The Power Transfer Functions (PTF) of the speech, the noise and the internal noise from the first microphone[2] to the algorithm's output are denoted as $G^s(f) = P_Z^s(f)/P_X^s(f)$, $G^n(f) = P_Z^n(f)/P_X^n(f)$ and $G^i(f) = P_Z^i(f)/P_X^i(f)$, respectively.

B. Signal Model

Each microphone signal $X_k(f)$, $k = 1, ..., M$, can be decomposed as

$$X_k(f) = X_k^s(f) + X_k^n(f), \tag{1}$$

The stacked vector $\mathbf{X}(f) = \begin{bmatrix} X_1(f) & X_2(f) & \cdots & X_M(f) \end{bmatrix}^T$ can then be written as

$$\mathbf{X}(f) = \mathbf{X}^s(f) + \mathbf{X}^n(f). \tag{2}$$

During periods of noise only, $\mathbf{X}^s(f) = 0$. In this paper, we assume that a robust speech detection algorithm is available. For the influence of speech detection errors on the performance of the GSC and MWF, we refer to [21], [36]. In the presence of a single speech source

$$\mathbf{X}^s(f) = \mathbf{H}(f)S_d(f), \tag{3}$$

where $\mathbf{H}(f) = \begin{bmatrix} H_1(f) & H_2(f) & \cdots & H_M(f) \end{bmatrix}^T$ and $H_k(f)$ is the acoustic transfer function between the speech source $S_d(f)$ and the k-th microphone.

[2]The PTFs are computed with respect to the first microphone because the MWF provides an estimate of the speech component in this microphone signal (see Section 2).

C. Complex coherence function $\Gamma_{kl}(f)$

The complex coherence $\Gamma_{kl}(f)$ between the k-th and l-th microphone signals with PSD respectively $P_{X_k}(f)$ and $P_{X_l}(f)$ and cross-PSD $P_{X_kX_l}(f)$, is defined as[3]:

$$\Gamma_{kl}(f) = \frac{P_{X_kX_l}}{\sqrt{P_{X_k}P_{X_l}}}. \tag{4}$$

For a homogeneous sound field, the correlation matrix $\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\}$ then equals[4]

$$\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\} = P_X(f)\mathbf{\Gamma}(f), \tag{5}$$

where $\mathbf{\Gamma}(f)$ is the coherence matrix defined as

$$\mathbf{\Gamma}(f) = \begin{bmatrix} 1 & \Gamma_{12}(f) & \cdots & \Gamma_{1M}(f) \\ \Gamma_{21}(f) & 1 & \cdots & \Gamma_{2M}(f) \\ \vdots & \vdots & \vdots & \vdots \\ \Gamma_{M1}(f) & \Gamma_{M2}(f) & \cdots & 1 \end{bmatrix}. \tag{6}$$

Similarly, we define the noise and speech correlation matrix as

$$\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\} = P_X^n(f)\mathbf{\Gamma}^n(f), \tag{7}$$

$$\mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\} = P_X^s(f)\mathbf{\Gamma}^s(f). \tag{8}$$

In the theoretical analysis, we assume perfect estimation of these statistics unless stated differently.

When the sound field is known, $\Gamma_{kl}(f)$ can be computed mathematically. E.g., the coherence Γ_{kl} of a *localized sound source* at an angle θ with respect to the front direction in free and far field equals

$$\Gamma_{kl}(f) = \exp\left(j2\pi f \cos\theta(d_l - d_k)/c\right) \tag{9}$$

and the coherence Γ_{kl} of a *diffuse sound field* equals

$$\Gamma_{kl}(f) = \frac{\sin(2\pi f(d_l - d_k)/c)}{2\pi f(d_l - d_k)/c} = \text{sinc}(2\pi f(d_l - d_k)/c) \tag{10}$$

with $d_l - d_k$ the interspacing between microphone l and k [28], [37]. The diffuse sound field is a good approximation of sound fields in strongly reverberant rooms, if the sound sources in the room have no direct path to the microphone array, i.e. for a low direct-to-reverberant ratio. The coherence function

[3]The frequency parameter f is often omitted in the sequel for the sake of conciseness.

[4]In Section V-B, the coherence matrix $\mathbf{\Gamma}$(f) in (5) is modified to include spectral differences between the microphone signals.

$\Gamma_{kl}(f)$ for a set of L *uncorrelated sound sources* with individual coherence functions $\Gamma^j_{kl}(f)$ and PSD $P^j(f)$, $j = 1, ..., L$ is given by

$$\Gamma_{kl}(f) = \frac{\sum_{j=1}^{L} \alpha_j(f)\Gamma^j_{kl}(f)}{\sum_{j=1}^{L} \alpha_j(f)}, \tag{11}$$

with $\alpha_j(f) = \frac{P^j(f)}{P^1(f)}$. E.g., in the presence of spatially uncorrelated noise such as sensor noise, the noise coherence $\Gamma^n_{kl}(f)$ is given by [29]

$$\Gamma^n_{kl}(f) = \frac{\Gamma^e_{kl}(f)}{1 + \beta(f)} \qquad \text{for } k \neq l, \tag{12}$$

since $\Gamma^i_{kl} = \delta[k - l]$, with Γ^e_{kl} the coherence function of the external noise and with $\beta(f) = \frac{P^i_X(f)}{P^e_X(f)} > 0$ the internal-to-external noise ratio.

D. Intelligibility weighted performance measures

To assess the effect of the PTFs $G^s(f)$ and $G^n(f)$ on intelligibility - which is the major goal of a noise reduction algorithm in hearing aid applications - broadband intelligibility weighted measures have been proposed. The improvement in intelligibility weighted SNR [38] is defined as

$$\Delta\text{SNR}_{\text{intellig}} = \sum_i I_i(\text{SNR}_{i,\text{out}} - \text{SNR}_{i,\text{in}}), \tag{13}$$

where the band importance function I_i expresses the importance of the i-th one-third octave band with center frequency f^c_i for intelligibility, $\text{SNR}_{i,\text{out}}$ is the output SNR (in dB) and $\text{SNR}_{i,\text{in}}$ is the input SNR (in dB) in the i-th one-third octave band. The center frequencies f^c_i and the values I_i are defined in [39]; frequencies around 1 kHz are most important while frequencies above 5 kHz are less important for speech intelligibility. Given $G^s(f)$ and $G^n(f)$, the improvement $\Delta\text{SNR}_{\text{intellig}}$ can be calculated using

$$\text{SNR}_{i,\text{out}} = 10\log_{10}\left(\int_{2^{-\frac{1}{6}}f^c_i}^{2^{\frac{1}{6}}f^c_i} P^s_X(f)G^s(f)df \Big/ \int_{2^{-\frac{1}{6}}f^c_i}^{2^{\frac{1}{6}}f^c_i} P^n_X(f)G^n(f)df\right) \tag{14}$$

$$\text{SNR}_{i,\text{in}} = 10\log_{10}\left(\int_{2^{-\frac{1}{6}}f^c_i}^{2^{\frac{1}{6}}f^c_i} P^s_X(f)df \Big/ \int_{2^{-\frac{1}{6}}f^c_i}^{2^{\frac{1}{6}}f^c_i} P^n_X(f)df\right). \tag{15}$$

Similarly, we define an intelligibility weighted spectral distortion measure (in dB), $\text{SD}_{\text{intellig}}$, of the desired speech signal as

$$\text{SD}_{\text{intellig}} = \sum_i I_i \text{SD}_i, \tag{16}$$

with SD_i the average spectral distortion (in dB) in the i-the one-third octave band, calculated as

$$\text{SD}_i = \int_{2^{-\frac{1}{6}}f^c_i}^{2^{\frac{1}{6}}f^c_i} |10\log_{10} G^s(f)|\, df \Big/ \left[\left(2^{\frac{1}{6}} - 2^{-\frac{1}{6}}\right) f^c_i\right]. \tag{17}$$

III. Generalized Sidelobe Cancellation

A. Principle and notation

The GSC [4], [14], depicted in Figure 1, consists of a fixed beamformer $\mathbf{A}$, a blocking matrix $\mathbf{B}$ and an ANC. The $M \times 1$ fixed beamformer $\mathbf{A} = [\ A_1(f) \quad \dots \quad A_M(f)\ ]^T$ creates the speech reference $Y_0(f) = \mathbf{A}^H(f)\mathbf{X}(f)$ by steering a beam towards the direction of the desired signal. In the sequel, an endfire microphone array is assumed and the desired speaker is assumed to be in front at 0 degrees. The $M \times (M-1)$ blocking matrix $\mathbf{B}$ creates noise references $Y_i(f) = \mathbf{B}^H(i,:)\mathbf{X}(f)$, $i = 1, \cdots, M-1$, by steering zeroes to the front. The ANC filters $W_i(f)$, $i = 1, \cdots, M-1$ reduce the noise in the speech reference that is correlated with the noise references. To avoid possible speech cancellation, the ANC is adapted during periods of noise only [5], [7], [11], [16]. In the sequel, a delay-and-sum beamformer is used, i.e. $A_i(f) = \frac{1}{M} e^{-j2\pi f \frac{d_i - d_1}{c}}$ with $d_i - d_1$ the interspacing between the i-th and the first microphone. For e.g., $M = 3$, the $M \times (M-1)$ blocking matrix $\mathbf{B}$ is set to

$$\mathbf{B} = \begin{bmatrix} 1 & -e^{j2\pi f \frac{d_2 - d_1}{c}} & 0 \\ 1 & 0 & -e^{j2\pi f \frac{d_3 - d_1}{c}} \end{bmatrix}^H . \tag{18}$$

B. Theoretical performance

In [28], [31], the theoretical noise reduction $NR(f)$ of the GSC has been expressed in terms of the complex coherence $\Gamma_{kl}(f)$ between the k-th and l-th microphone signals. In addition to the PTF of the noise signal $G^n(f) = \frac{1}{NR(f)}$, also the PTF of the desired speech signal $G^s(f)$ and the internal noise $G^i(f)$ can be expressed as a function of $\Gamma_{kl}(f)$. These results are reviewed in this section. In addition, the formulas for the more robust GSC-NSC are derived. From the PTFs $G^s(f)$, $G^n(f)$, $G^i(f)$, the broadband intelligibility weighted performance measures can be obtained.

1) PTFs $G^s(f)$, $G^n(f)$ and $G^i(f)$ of the GSC as a function of $\Gamma_{kl}(f)$: The PTFs $G^s(f)$, $G^n(f)$ and $G^i(f)$ can be expressed as a function of $\mathbf{\Gamma}^s(f)$ and $\mathbf{\Gamma}^n(f)$, the fixed beamformer $\mathbf{A}(f)$ and the blocking matrix $\mathbf{B}(f)$. The ANC filters $\mathbf{W}(f)$ are given by:

$$\begin{aligned} \mathbf{W}(f) &= \mathcal{E}\{\mathbf{B}^H(f)\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\mathbf{B}(f)\}^{-1}\mathcal{E}\{\mathbf{B}^H(f)\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\mathbf{A}(f)\} \\ &= \left(\mathbf{B}^H(f)\mathbf{\Gamma}^n(f)\mathbf{B}(f)\right)^{-1}\mathbf{B}^H(f)\mathbf{\Gamma}^n(f)\mathbf{A}(f). \end{aligned} \tag{19}$$

For a homogeneous noise field and infinite filter lengths, the filters $W_i(f)$ thus only depend on the spatial characteristics of the noise $\mathbf{\Gamma}^n(f)$ and not on $P_X^n(f)$.

The PSD of the noise signal at the ANC output $P_Z^n(f)$ equals

$$P_Z^n(f) = P_X^n(f)\left(\mathbf{A}^H(f) - \mathbf{W}^H(f)\mathbf{B}^H(f)\right)\mathbf{\Gamma}^n(f)\left(\mathbf{A}(f) - \mathbf{B}(f)\mathbf{W}(f)\right). \tag{20}$$

The speech PSD P_Z^s and internal noise PSD P_Z^i are obtained by replacing the superscripts n with s and i. Using (19)-(20) and $\Gamma_{kl}^i = \delta[k-l]$, G^s, G^n and G^i can be expressed as a function of $\mathbf{\Gamma}^s$ and $\mathbf{\Gamma}^n$:

$$G^{n|s}(f) = \left(\mathbf{A}^H(f) - \mathbf{W}^H(f)\mathbf{B}^H(f)\right)\mathbf{\Gamma}^{n|s}(f)\left(\mathbf{A}(f) - \mathbf{B}(f)\mathbf{W}(f)\right), \quad (21)$$

$$G^{i}(f) = \left(\mathbf{A}^H(f) - \mathbf{W}^H(f)\mathbf{B}^H(f)\right)\left(\mathbf{A}(f) - \mathbf{B}(f)\mathbf{W}(f)\right), \quad (22)$$

where $G^{n|s}(f)$ means $G^n(f)$ or $G^s(f)$, etc. (for compact notation).

In [21], [40], it has been shown that if $\mathbf{B}^H(f)\mathbf{A}(f) = \mathbf{0}$ and if $\left[\begin{array}{cc} \mathbf{B}(f) & \mathbf{A}(f) \end{array}\right]$ is a non-singular matrix of dimensions $M \times M$,

$$\mathbf{A}(f) - \mathbf{W}(f)\mathbf{B}(f) = \frac{\mathbf{A}^H(f)\mathbf{A}(f)}{\mathbf{A}^H(f)\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{A}(f)}\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{A}(f). \quad (23)$$

For $\mathbf{A}(f) = \frac{\mathbf{H}(f)}{\mathbf{H}^H(f)\mathbf{H}(f)}$, $\mathbf{A}(f) - \mathbf{W}(f)\mathbf{B}(f)$ corresponds to an optimum MVDR beamformer [34, Chapter 6] steered to $\mathbf{H}(f)$.

2) PTFs $\tilde{G}^s(f)$, $\tilde{G}^n(f)$ and $\tilde{G}^i(f)$ of the GSC-NSC: To increase the robustness of the GSC, a constraint on the noise sensitivity is often applied [3], [8], [11], [13]. In [3], it is shown that a constraint on the noise sensitivity corresponds to a quadratic inequality constraint on the filters $\mathbf{B}(f)\mathbf{W}(f)$, i.e.

$$\mathbf{W}^H(f)\mathbf{B}^H(f)\mathbf{B}(f)\mathbf{W}(f) \leq \zeta. \quad (24)$$

In [13], a similar constraint is imposed by inserting uncorrelated noise in the microphone signals[5] used to design the filters. In [3], it is demonstrated that the method of noise injection is equivalent with a quadratic inequality constraint for a certain level of noise injection and a certain value ζ. Hence, the optimal performance of GSC with noise sensitivity constraint is obtained by replacing $\Gamma_{kl}^n(f)$ in the computation of $\mathbf{W}(f)$ with

$$\Gamma_{kl}^n(f) + \eta(f)\delta[k-l], \quad (25)$$

where $\eta(f)$ is the ratio of the injected noise power to the input noise power $P_X^n(f)$.

Summarizing, given the GSC parameters $\mathbf{A}(f)$ and $\mathbf{B}(f)$, the coherences Γ_{kl}^s, Γ_{kl}^n and the PSD $P_X^s(f)$, $P_X^n(f)$, the intelligibility weighted performance measures (13) and (16) of GSC and GSC-NSC are obtained.

[5]Since $\mathbf{B}^H(f)\mathbf{A}(f) = 0$, this corresponds to adding a diagonal loading term to the coherence matrix $\mathbf{\Gamma}^n(f)$ of the noise in the denominator of (19).

IV. MULTI-CHANNEL WIENER FILTERING

A. Concept

The MWF $\mathbf{w} \in \mathcal{C}^{ML\times 1}$, described in Figure 2, provides an MMSE estimate of a delayed[6] version of the (unknown) reverberant speech component $x_i^s[k-\Delta]$ in the i-th (e.g. first) microphone signal, which is computed as $z[k] = \mathbf{w}^H\mathbf{x}[k]$, i.e. the sum of the M filtered microphone signals, where

$$\mathbf{w} = \mathcal{E}\{\mathbf{x}[k]\mathbf{x}^H[k]\}^{-1}\mathcal{E}\{\mathbf{x}[k]x_i^{s,*}[k-\Delta]\}, \tag{26}$$

$$\mathbf{x}^H[k] = \left[\begin{array}{cccc} \mathbf{x}_1^H[k] & \mathbf{x}_2^H[k] & \cdots & \mathbf{x}_M^H[k] \end{array}\right], \tag{27}$$

$$\mathbf{x}_i[k] = \left[\begin{array}{cccc} x_i[k] & x_i[k-1] & \cdots & x_i[k-L+1] \end{array}\right]^T, \tag{28}$$

and where L is the number of filter taps per channel [20].

Because of double-talk of speech and noise, the reference signal $x_i^s[k-\Delta]$ is unknown. If a robust speech detection is used, noise only observations can be made during speech pauses (denoted with time index k'), where $x_i[k'] = x_i^n[k']$. This allows to estimate the second order statistics of the noise. These noise statistics are assumed to be sufficiently stationary, i.e., $\mathcal{E}\{\mathbf{x}^n[k]x_i^{n,*}[k-\Delta]\} = \mathcal{E}\{\mathbf{x}^n[k']x_i^{n,*}[k'-\Delta]\}$. Assuming that the speech and noise signals are uncorrelated, (26) becomes:

$$\mathbf{w} = \mathcal{E}\{\mathbf{x}[k]\mathbf{x}^H[k]\}^{-1}\left(\mathcal{E}\{\mathbf{x}[k]x_i^*[k-\Delta]\} - \mathcal{E}\{\mathbf{x}^n[k]x_i^{n,*}[k-\Delta]\}\right), \tag{29}$$

where the second order statistics $\mathcal{E}\{\mathbf{x}[k]\mathbf{x}^H[k]\}$ and $\mathcal{E}\{\mathbf{x}^n[k]x_i^{n,*}[k-\Delta]\}$ are estimated online during periods of speech + noise and periods of noise only, respectively. In the sequel, we will assume, without loss of generality, that the speech contribution in the first microphone signal is estimated, i.e. $x_1[k-\Delta]$. In contrast to the GSC, the MWF does not make any a priori assumptions about the desired signal model. It is uniquely based on estimates of the second order statistics of the speech and noise signal.

In [19]-[22], recursive implementations based on a GSVD or a QRD of an speech + noise data matrix and a noise data matrix have been proposed for (29). A subband implementation [23] results in improved intelligibility at a significantly lower cost compared to the fullband approach, making it suitable for future hearing aids. To allow for tracking of changing scenarios, exponential weighting with forgetting factors λ_{s+n}^2 and λ_n^2 is incorporated in the update formulas for the speech + noise and noise only correlation matrix, respectively [21], [22]. For a sampling frequency $f_s = 16$ kHz, typical values for λ_{s+n}^2 and λ_n^2 are around 0.99998, which corresponds to a time-window $1/[f_s(1-\lambda_{s+n|n}^2)] \approx 3$ seconds, i.e., much longer than the average short-time stationarity of speech. Hence, the performance of the MWF is

[6]The delay Δ is applied to allow for non-causal taps of the filter $\mathbf{w}$.

largely dependent on the average, i.e. long-term, spectral and spatial characteristics of the speech and noise sources so that no short-time effects, such as residual musical noise, will occur in this MWF. This implies that the MWF can also be used to suppress spectrally non-stationary noise, provided that the long-term, spectral and spatial characteristics of the speech and the noise vary slowly in time [21]. Highly non-stationary noise environments complicate the estimation of the speech correlation matrix by the MWF and hence, are expected to affect its performance.

B. Theoretical analysis

The theoretically optimal performance of the MWF, i.e., the PTFs $G^s(f)$,$G^n(f)$ and $G^i(f)$ as well as the intelligibility weighted performance measures, can again be expressed as a function of $\mathbf{\Gamma}^n$ and $\mathbf{\Gamma}^s$.

The MWF that provides an MMSE estimate of[7] $X_1^s(f)$, can be written as:

$$\mathbf{W}(f) = \mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\}^{-1}\left(\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\} - \mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}\right)\mathbf{e}_1 \tag{30}$$

where $\mathbf{W}(f) = \left[\begin{array}{cccc} W_1(f) & W_2(f) & \cdots & W_M(f) \end{array}\right]^T$, where $W_i^*(f)$ filters the i-th microphone signal and where the $M \times 1$ vector $\mathbf{e}_1$ equals the first canonical vector defined as $\mathbf{e}_1 = \left[\begin{array}{cccc} 1 & 0 & \cdots & 0 \end{array}\right]^T$. First, we assume perfect estimation of the second order statistics of speech and noise, i.e.,

$$\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\} - \mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\} = \mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\}. \tag{31}$$

The sensitivity of the MWF to inaccurate estimation of the speech + noise and noise statistics is analyzed later in Section V-C. Using (5)-(8) and (31), formula (30) can be rewritten as:

$$\boxed{\mathbf{W}(f) = \frac{P_X^s(f)}{P_X^s(f) + P_X^n(f)}\mathbf{\Gamma}^{-1}(f)\mathbf{\Gamma}^s(f)\mathbf{e}_1,} \tag{32}$$

where, using (11):

$$\mathbf{\Gamma}(f) = \frac{\mathbf{\Gamma}^s(f) + \alpha(f)\mathbf{\Gamma}^n(f)}{1+\alpha(f)}; \tag{33}$$

$$\alpha(f) = \frac{P_X^n(f)}{P_X^s(f)}. \tag{34}$$

In contrast to the GSC, the filter $\mathbf{W}(f)$ not only depends on the spatial characteristics of the noise $\mathbf{\Gamma}^n(f)$, but also on the input noise-to-signal ratio $\alpha(f)$ and the spatial characteristics of the speech signal $\mathbf{\Gamma}^s(f)$.

Note that the ratio $\frac{P_X^s(f)}{P_X^s(f)+P_X^n(f)}$ in (32) corresponds to a *spectral* Wiener filter, which results in additional performance if the spectra of the desired and undesired signal differ. Hence, (32) combines

[7]In the derivation of $\mathbf{W}(f)$, we assume, without loss of generality, that the delay Δ equals zero.

spatial and spectral filtering. Using

$$P_Z^{n|s}(f) = P_X^{n|s}(f)\mathbf{W}^H(f)\mathbf{\Gamma}^{n|s}(f)\mathbf{W}(f), \tag{35}$$

and $\Gamma_{kl}^i = \delta[k-l]$, the PTFs $G^n(f)$, $G^s(f)$ and $G^i(f)$ can be easily derived:

$$\begin{aligned} G^{n|s}(f) &= \mathbf{W}^H(f)\mathbf{\Gamma}^{n|s}(f)\mathbf{W}(f), & (36)\\ G^i(f) &= \mathbf{W}^H(f)\mathbf{W}(f). & (37) \end{aligned}$$

In the presence of a *single speech source* with PSD $P_{S_d}(f)$, $\mathbf{\Gamma}^s(f)$ is rank-one and equals

$$\mathbf{\Gamma}^s(f) = \frac{P_{S_d}(f)}{P_X^s(f)}\mathbf{H}(f)\mathbf{H}^H(f). \tag{38}$$

Using (38) and the matrix inversion lemma, we obtain

$$\left(\mathbf{\Gamma}^s(f) + \alpha(f)\mathbf{\Gamma}^n(f)\right)^{-1} = \alpha^{-1}(f)\left[\mathbf{\Gamma}^{n^{-1}} - \frac{\mathbf{\Gamma}^{n^{-1}}\mathbf{H}(f)\mathbf{H}^H(f)\mathbf{\Gamma}^{n^{-1}}}{\mathbf{H}(f)^H\mathbf{\Gamma}^{n^{-1}}\mathbf{H}(f) + \frac{P_X^n(f)}{P_{S_d}(f)}}\right], \tag{39}$$

so that $\mathbf{W}(f)$ in (32) can be re-written as

$$\mathbf{W}(f) = \underbrace{\Lambda(f)\left(P_X^n(f)\mathbf{\Gamma}^n\right)^{-1}\mathbf{H}(f)}_{\mathbf{W}_1(f)}\underbrace{\left(\frac{P_{S_d}(f)}{P_{S_d}(f)+\Lambda(f)}\right)}_{W_2(f)}H_1^*(f), \tag{40}$$

with

$$\Lambda(f) = \left[\mathbf{H}(f)^H(P_X^n(f)\mathbf{\Gamma}^n)^{-1}\mathbf{H}(f)\right]^{-1}. \tag{41}$$

The MWF $\mathbf{W}(f)$ in (40) can be decomposed into three parts:

- The *multi-channel filter* $\mathbf{W}_1(f)$ corresponds to an optimum MVDR beamformer with steering vector $\mathbf{H}(f)$ [34, Chapter 6] and, hence, provides a distortionless response for the target signal $S_d(f)$, i.e.,

  $$\mathbf{W}_1^H P_X^s(f)\mathbf{\Gamma}^s(f)\mathbf{W}_1 = P_{S_d}(f). \tag{42}$$

  The residual noise at the output of the filter $\mathbf{W}_1(f)$, i.e., $\mathbf{W}_1^H(f)P_X^n(f)\mathbf{\Gamma}^n(f)\mathbf{W}_1(f)$ equals $\Lambda(f)$.
- The *single-channel filter* $W_2(f)$ corresponds to a single-channel Wiener filter at the output of the MVDR $\mathbf{W}_1(f)$. This filter increases noise suppression at the expense of distortion of the target signal. Note that the cascade $\mathbf{W}_1(f)W_2(f)$ corresponds to an MWF that provides an MMSE estimate of the source speech signal $S_d(f)$ [18].
- The single-channel filter $H_1^*(f)$ translates the MMSE estimate of the source signal $S_d(f)$ to an MMSE estimate of the speech component $X_1^s(f)$ in the first microphone signal.

Using (40), the PTF $G^n(f)$, $G^s(f)$ and $G^i(f)$ in (36) and (37) can be simplified to:

$$G^s(f) = \left(\frac{P_{S_d}(f)}{P_{S_d}(f)+\Lambda(f)}\right)^2, \quad (43)$$

$$G^n(f) = |H_1(f)|^2 \left(\frac{P_{S_d}(f)}{P_{S_d}(f)+\Lambda(f)}\right)^2 \frac{\Lambda(f)}{P_X^n(f)}, \quad (44)$$

$$G^i(f) = |H_1(f)|^2 \left(\frac{P_{S_d}(f)}{P_{S_d}(f)+\Lambda(f)}\right)^2 \frac{\mathbf{H}^H\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{H}}{\left(\mathbf{H}^H\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{H}\right)^2} \quad (45)$$

Speech distortion occurs at frequencies f where the SNR $P_{S_d}(f)/P_X^n(f)$ is low and where the amount of spatial noise reduction $\Lambda(f)/P_X^n(f)$ is limited (i.e., for small $f \cdot d$ and difficult noise scenarios). Theoretically, the MWF achieves the same SNR improvement $\frac{G^s(f)}{G^n(f)}$ with respect to the first microphone signal as an optimum MVDR beamformer steered to $\mathbf{H}(f)$ and - in the absence of errors in the signal model assumed by GSC such as reverberation - as the GSC.

V. Theoretical robustness analysis

Based on the formulas in Section III and Section IV, the theoretical performance of the GSC and the MWF can be evaluated analytically. In this section, we investigate how the broadband intelligibility weighted performance measures of GSC, GSC-NSC and MWF are affected by different types of model errors when used for small-sized arrays. The results are illustrated for a linear, 3-microphone endfire array with the dimensions of a BTE.

In contrast to the GSC, the MWF does not require a priori information about the signal model: theoretically, the MWF consists of an optimum MVDR cascaded with a single-channel spectral filter (see (40)) and is thus expected to be more robust than the GSC. To quantify the sensitivity to errors in the assumed signal model, the *noise sensitivity* is often used [1], [2], [3]. In Section V-A, we study under which conditions (noise scenario, microphone configuration) the noise sensitivity of the GSC and the MWF is high and, hence, robustness of the GSC is expected to be crucial. Section V-B analyzes the effect of a high noise sensitivity on the robustness of the GSC, the GSC-NSC and the MWF against different types of errors in the assumed signal model such as microphone mismatch.

In Section V-A and Section V-B, we assume perfect estimation of the second order statistics of speech and noise by the GSC and the MWF. Section V-C investigates the sensitivity of the MWF, the GSC and the GSC-NSC to estimation errors in the second order statistics, e.g., caused by non-stationarities in the noise.

A. Analysis of the noise sensitivity

Using (21), (22), (36) and (37), the noise sensitivity $\Psi(f)$ of the GSC and the MWF is derived as:

$$\boxed{\Psi(f) = \frac{G^i(f)}{G^s(f)}.} \tag{46}$$

Equations (22) and (37) for the internal noise PTF $G^i(f)$ show that, for both algorithms, large filter amplitudes $|W_i(f)|$ with respect to $G^s(f)$ correspond to a high noise sensitivity $\Psi(f)$. If the filter amplitudes of the ANC in the GSC are large, possible leakage of speech into the noise references will get amplified resulting in distortion of the desired signal.

Using (23) in (21) and (22), $\Psi_{\text{GSC}}(f)$ of the GSC can be written as:

$$\Psi_{\text{GSC}}(f) = \frac{\mathbf{A}^H(f)\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{A}(f)}{\mathbf{A}^H(f)\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{\Gamma}^{s}(f)\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{A}(f)}. \tag{47}$$

Using (38), (43), (45) and (47), it can be shown that *in case of a single speech source*, i.e., $\mathbf{\Gamma}^{\boldsymbol{s}}(f) = P_{S_d}(f)\mathbf{H}^H\mathbf{H}(f)/P_X^s(f)$, *and in the absence of errors in the signal model assumed by GSC*, i.e. $\mathbf{A}(f) = \alpha\mathbf{H}(f)$ with α a constant, the noise sensitivity of GSC and MWF coincide [21, Chapter 3], i.e.,

$$\boxed{\Psi_{\text{GSC}}(f) = \Psi_{\text{MWF}}(f) = |H_1(f)|^2 \frac{\mathbf{H}^H\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{H}}{\left(\mathbf{H}^H\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{H}\right)^2}.} \tag{48}$$

For both algorithms, $\Psi(f)$ depends on the coherence matrix $\mathbf{\Gamma}^n(f)$ of the noise field and hence, on the noise scenario and the microphone configuration.

Figure 3 investigates the *influence of the noise scenario on the noise sensitivity* $\Psi(f)$. The sensitivity $\Psi(f)$ (in dB) is plotted as a function of $\frac{2f}{f_s}d$ (in meters) for a linear, uniform 3-microphone endfire array, with sampling frequency $f_s = 16$ kHz. Five different noise scenarios are shown: a localized single noise source at 45° and 90° with respect to the speech source, both with internal-to-external noise ratio $\beta = -30$ dB, a multiple noise source scenario with five localized noise sources at 75°, 150°, 180°, 240°, 285° with the same PSD and a diffuse noise field with $\beta = -20$ dB and $\beta = -30$ dB. Typical values for $\beta(f)$ lie between -20 dB and -40 dB [29]. Figure 3 shows that the noise sensitivity $\Psi(f)$ grows with decreasing $f \cdot d$, decreasing internal-to-external noise ratio $\beta(f)$ and increasing complexity of the noise scenario. Diffuse noise, multiple noise sources and noise sources close to the speech source result in a high $\Psi(f)$ compared to e.g. a single, localized noise source at 90°.

Figure 4 depicts the *influence of the number of microphones* M *per array length* $L = (M-1)d$ on the noise sensitivity $\Psi(f)$ (in dB) and the improvement in SNR $\frac{G^s(f)}{G^n(f)}$ (in dB) of the GSC and the MWF as a function of $\frac{2f}{f_s}L$ (in meters) for a diffuse noise field with $\beta(f) = -30$ dB. For a free and far field speech source at $\theta = 0^\circ$, both algorithms result in the same optimal improvement $\frac{G^s(f)}{G^n(f)}$ (see also Section IV-B). For small-sized arrays (i.e. small $\frac{2f}{f_s}L$), an increase of the number of microphones

per array length corresponds to an equal or higher noise sensitivity $\Psi(f)$. Hence, the larger the number of microphones per array length, the more robust the noise reduction algorithm should be. At low $\frac{2f}{f_s}L$, external noise is reduced at the expense of internal noise amplification: there is indeed an overall noise reduction $NR(f)$ that optimally trades off internal noise amplification G^i and external noise reduction. The external noise reduction and maximum internal noise amplification will be low for high internal-to-external ratio $\beta(f)$. At very low $\frac{2f}{f_s}L$, the noise sensitivity and improvement in SNR reach a maximum and do not further increase by adding more microphones. In Figure 4, this maximum is reached at low $\frac{2f}{f_s}L$, below 0.01 m. As a consequence, the number of useful microphones per array length is limited by the internal-to-external noise ratio $\beta(f)$ [2] and by the noise sensitivity level $\Psi(f)$ tolerated by the noise reduction algorithm (see Section V-B). For an array with a length $L = 3$ cm, which is similar to the dimensions of a commercial BTE hearing aid, no significant SNR improvement is obtained with an additional fourth microphone at frequencies below 4 kHz, i.e. for $\frac{2fL}{f_s} \leq 0.015$ m. The number of useful microphones in a commercial BTE is limited to three, provided the noise reduction algorithm is robust enough so that no constraint on $\Psi(f)$ is needed.

B. Robustness against errors in the assumed signal model

Section V-A shows that for small-sized arrays, complex noise scenarios and an increasing number of microphones result in a high noise sensitivity. In this section, we evaluate the robustness of the GSC, the GSC-NSC and the MWF to different types of errors in the assumed signal model, i.e. microphone mismatch, robustness to microphone placement uncertainty and robustness to variations in speaker location. The theoretical robustness analysis allows to identify which type of model errors will be most detrimental to the GSC when used for hearing aid applications. The results are illustrated for a 3-microphone endfire array with microphone interspacings $d_2 - d_1 = 0.01$ m and $d_3 - d_2 = 0.015$ m and a multiple noise source scenario with five localized noise sources at $75°$, $150°$, $180°$, $240°$, $285°$ and $\beta(f) = -30$ dB. Both the influence on the PTFs of speech and noise as well as the influence on the intelligibility weighted performance measures (13)-(16) are analyzed. Since in practice, the desired and interfering signals will be speech-like, the improvement $\Delta\text{SNR}_{\text{intellig}}$ is computed using the average long-term speech PSD of the signals found on [41] for $P_X^s(f)$ and $P_X^n(f)$.

1) Robustness to microphone mismatch: In practice, microphones are rarely matched in gain and phase. In [42], gain and phase differences between microphone characteristics of up to 6 dB and $10°$, respectively, have been reported. The fixed stage of the GSC can be designed to compensate for the differences in microphone characteristics. However, calibrating the fixed stage for each hearing aid individually is

cumbersome and expensive. In addition, microphone characteristics may drift over time [42]. Hence, the robustness of the algorithm against gain and phase mismatch is crucial.

We mimic a deviation from matched microphone characteristics by modifying the coherence functions Γ^s_{kl}, Γ^n_{kl} and Γ^i_{kl}. Suppose that the k-th microphone has a gain mismatch $\upsilon(f)$ and a phase mismatch $\phi(f)$ (in degrees) at frequency f with respect to microphone l. The matched microphones coherence function $\Gamma_{kl}(f)$ should then be replaced by $\tilde{\Gamma}_{kl}(f)$

$$\tilde{\Gamma}_{kl}(f) = \begin{cases} \upsilon(f)\exp(j\phi(f)\frac{\pi}{180})\Gamma_{kl}(f) & \text{for } l \neq k; \\ |\upsilon(f)|^2\Gamma_{kl}(f) & \text{for } l = k. \end{cases} \tag{49}$$

Figure 5 plots the PTF $G^s(f)$ and $G^n(f)$ of the **GSC** when the second microphone has a gain mismatch $\upsilon(f) = 2$ dB or a phase mismatch $\phi(f) = 5^\circ$ with respect to the first microphone. The desired signal gets cancelled at lower f where $\Psi(f)$ is high (see Figure 3). Due to the high $\Psi(f)$ and the cancellation of the desired speech, internal noise may become audible. The PTFs $G^s(f)$ and $G^n(f)$ of the MWF (for noise-to-signal ratio $\alpha(f) = 0$ dB) are also shown. As demonstrated in Section IV-B the **MWF** attenuates the speech at frequencies where the output SNR is small, i.e., in this example at low f.

Theorem 1: Theoretically, the PTF $G_s(f)$ and $G_n(f)$ of the MWF are independent of a deviation in microphone gain and phase.

Proof: Suppose that the received microphone signals $\tilde{X}_i(f)$, with $i = 1, ..., M$ are perturbed by $\Upsilon_i(f) = \upsilon_i(f)\exp(j\phi_i(f)\frac{\pi}{180^\circ})$ with $\upsilon_i(f)$ a gain deviation and $\phi_i(f)$ a phase deviation, with respect to the signals $X_i(f)$ received in the absence of microphone mismatch, i.e.,

$$\tilde{\mathbf{X}}(f) = \mathbf{D}(f)\mathbf{X}(f), \tag{50}$$

with

$$\mathbf{D}(f) \;=\; \text{diag}\{\Upsilon_1(f), ..., \Upsilon_M(f)\} \tag{51}$$

and with $\mathbf{D}(f)$ of full rank, i.e., $\Upsilon_i(f) \neq 0$. The MWF $\tilde{\mathbf{W}}(f)$ in (30) then equals

$$\begin{aligned} \tilde{\mathbf{W}}(f) &= \mathcal{E}\{\tilde{\mathbf{X}}(f)\tilde{\mathbf{X}}^H(f)\}^{-1}\left(\mathcal{E}\{\tilde{\mathbf{X}}(f)\tilde{\mathbf{X}}^H(f)\} - \mathcal{E}\{\tilde{\mathbf{X}}^n(f)\tilde{\mathbf{X}}^{n,H}(f)\}\right)\mathbf{e}_1, \\ &= \mathbf{D}^{-H}\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\}^{-1}\left(\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\} - \mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}\right)\mathbf{D}^H\mathbf{e}_1. \end{aligned} \tag{52}$$

This results in the output $\tilde{Z}(f) = \Upsilon_1(f)\mathbf{W}^H(f)\mathbf{X}(f)\mathbf{e}_1$, which is a scaled version of the output $Z(f) = \mathbf{W}^H(f)\mathbf{X}(f)\mathbf{e_1}$ that would be obtained in the absence of mismatch. Both, the input $\tilde{X}_1(f)$ and the output $\tilde{Z}(f)$ of the MWF are scaled by $\Upsilon_1(f)$. Hence, the PTFs $G^s(f)$, $G^n(f)$ of the MWF are not affected by the mismatch. ■

In Figure 6(a) to Figure 6(d), the performance measures $\Delta\text{SNR}_{\text{intellig}}$ (Figure 6(a)-(c)) and $\text{SD}_{\text{intellig}}$ (Figure 6(b)-(d)) are shown as a function of the amount of gain mismatch υ and phase mismatch ϕ, respectively, at the second microphone. The performance of the more robust GSC-NSC, described in Section III, is depicted too for two levels of noise injection $\eta(f)$, $\eta(f) = -20$ dB and $\eta(f) = -10$ dB. Without noise sensitivity constraint, a small amount of mismatch leads to an unacceptable degradation of the GSC performance in complicated noise scenarios and at low $f \cdot d$, even if the ANC is adapted during noise only. Phase mismatch is less detrimental than gain mismatch. Constraining the noise sensitivity $\Psi(f)$, results in a higher robustness at the expense of less noise reduction. The performance of the **MWF** on the other hand is again seen to be robust to gain and phase mismatch, so that -depending on the processing accuracy- no (or a less severe) constraint on the noise sensitivity is needed. As a consequence, the MWF outperforms the robust GSC-NSC. The benefit of augmenting the number of microphones per array length, which is determined by $\beta(f)$ and the allowable noise sensitivity $\Psi(f)$, will also be more significant.

2) Robustness to deviations in microphone positions: The positions of the microphones vary slightly from one hearing aid to the other. Since microphones are quite accurately positioned in a hearing aid, the microphone placement uncertainty is expected to be small. A change Δd (in meters) in the microphone position of the k-th microphone with respect to the assumed position d_k can be incorporated in the coherence functions $\Gamma_{kl}(f)$, $\Gamma_{kl}^{n}(f)$ and $\Gamma_{kl}^{s}(f)$ by replacing d_k with $d_k + \Delta d$. Figure 7 depicts the PTF of speech and noise for GSC and MWF when the 2nd microphone is moved by $\Delta d = -0.001$ m and $\Delta d = -0.002$ m, respectively. The performance of the **GSC** is not strongly affected by such a small change in microphone positions, even when $\Psi(f)$ is rather large. The speech and noise are slightly attenuated, but the amount of noise reduction remains quite constant. In contrast to microphone mismatch, the largest effect of a change in microphone position does not occur at the lowest frequencies. A change by Δd in (9) corresponds to a frequency dependent phase change in the k-th microphone frequency characteristic for a localized sound source at angle θ of

$$2\pi f \frac{\cos\theta}{c} \Delta d \tag{53}$$

(in rad). On one hand, the impact of phase mismatch is largest for high $\Psi(f)$ and hence at low f, on the other hand the amount of phase mismatch (53) grows with increasing frequency. At low f where $\Psi(f)$ is high, the phase change is very small for $\Delta d = -0.001$ m and $\Delta d = -0.002$ m, hence the limited effect of a change in microphone position. Note that the impact still increases with increasing $\Psi(f)$. In applications where the uncertainties Δd are larger, a constraint on the noise sensitivity of GSC may be

required. Figure 7 shows that the PTF of speech and noise of the MWF are hardly modified when the microphones are slightly moved.

3) Robustness to variations in speaker location: In addition to being robust to microphone mismatch and variations in microphone positions, the algorithm should also be effective when the signal-of-interest does not arrive from the front but from an angle $\Delta\Theta$ with respect to the front. The effect of a change $\Delta\Theta$ can easily be studied analytically by incorporating the angle $\Delta\Theta$ with respect to 0° in $\Gamma^s_{kl}(f)$, i.e.

$$\Gamma^s_{kl}(f) = \exp\left(j2\pi f \cos(\Delta\Theta)\frac{d_l - d_k}{c}\right). \tag{54}$$

Since the ANC filters of the GSC are completely determined by $\Gamma^n_{kl}(f)$ and do not depend on $\Gamma^s_{kl}(f)$, the PTF $G^n(f)$ of the GSC is not affected by $\Delta\Theta$.

Figure 8 gives the effect of speaker variations up to 45° for noise-to-signal ratios $\alpha(f) = 0$ dB and $\alpha(f) = 3$ dB, respectively, on the performance measures $\Delta\text{SNR}_{\text{intellig}}$ [dB] and $\text{SD}_{\text{intellig}}$ [dB] of GSC and MWF. In contrast to microphone mismatch, this effect is not detrimental to GSC: for small to moderate variations up to 20°, the GSC even outperforms the MWF because of its smaller amount of speech distortion. The improvement $\Delta\text{SNR}_{\text{intellig}}$ decreases with increasing $\Delta\Theta$. The amount of speech distortion $\text{SD}_{\text{intellig}}$ caused by the MWF grows with decreasing output SNR, i.e., with increasing $\Delta\Theta$ and increasing $\alpha(f)$. In free field, the enhancement $\Delta\text{SNR}_{\text{intellig}}$ is the same for positive and negative $\Delta\Theta$. When mounted on the head, the input SNR and hence $\alpha(f)$ will be different for positive and negative $\Delta\Theta$ due to the head shadow effect: when speech arrives from an angle $-\Delta\Theta$ being shielded by the head, $\alpha(f)$ will be larger than when speech arrives from an angle $\Delta\Theta$. Hence, the speech will in practice be more attenuated for $-\Delta\Theta$.

4) Conclusion: In conclusion, the GSC seems extremely sensitive to microphone gain mismatch when $\Psi(f)$ is high, i.e. at low $f \cdot d$ and in complicated noise scenarios. The effect of variations in speaker position and microphone positions on the performance also grows for increasing $\Psi(f)$, however, even for high $\Psi(f)$, they are found to be not that detrimental for small-sized arrays. To guarantee performance in the presence of microphone gain mismatch, a constraint on the noise sensitivity should be imposed at the expense of less noise reduction. The MWF does not suffer from performance degradation in the presence of microphone mismatch and variations in speaker and microphone positions, so that, depending on the processing accuracy, no or a less severe constraint on the noise sensitivity is needed and hence, an improved performance is achieved.

C. Sensitivity to inaccurate estimation of the second order statistics

In the previous sections, we assumed perfect estimation of the second order statistics $\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}$ of the noise by the GSC and perfect estimation of the second order statistics $\mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\}$ and $\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}$ of the speech and the noise by the MWF.

In practice, the estimation of the second order statistics may be inaccurate because of several reasons. Because of double-talk of speech and noise, the clean speech signal is not available. The MWF estimates the speech correlation matrix $\mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\}$ by

$$\mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\} = \mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\} - \mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}, \tag{55}$$

where $\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\}$ and $\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}$ are estimated online during periods of speech + noise and periods of noise only, respectively. The correlation matrices $\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\}$ and $\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}$ are estimated from a *finite amount of incoming data* and *at a different moment in time* so that (55) does not exactly hold true. This estimation error is included in all experiments of Section VI. To allow for tracking of time-varying environments, the amount of data (time window) for the estimation of the second order statistics should be kept small. However, the smaller the amount of data, the worse the accuracy of the estimated second order statistics. Section VI-C studies the influence of the amount of data on the performance of MWF and GSC for a spatially and long-term spectrally stationary noise scenario.

Both the GSC and the MWF rely on a speech detection algorithm. *Speech detection errors* deteriorate the accuracy of the estimated second order statistics by the GSC and the MWF and hence, their performance. For the influence of speech detection errors on the GSC and MWF, we refer to [21], [36]: for a speech detection error rate of 20% or less, the MWF preserves its benefit over the GSC and the GSC-NSC.

Thirdly, because of the non-continuous estimation of the noise and speech + noise correlation matrices by the GSC[8] and the MWF, tracking of non-stationarities of the noise is reduced. *Non-stationarities of the noise* lead to a wrong estimate of $\mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\}$ by the MWF. Since the MWF does not rely on a priori information, the MWF is expected to be sensitive to these errors.

As an example to illustrate the sensitivity of the MWF, the GSC and GSC-NSC to erroneous estimation of the second order statistics due to non-stationarities in the noise, we assume that the noise correlation matrix $\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}$ is wrongly estimated because of a change in noise scenario. Suppose that the

[8]Note that the ANC of the GSC is adapted during noise only to avoid speech distortion. When non-stationarities of the noise occur during speech + noise, these are not immediately tracked by the GSC.

estimated noise correlation matrix $\mathcal{E}\{\mathbf{X}_{est}^{n}(f)\mathbf{X}_{est}^{n,H}(f)\}$ equals

$$\mathcal{E}\{\mathbf{X}_{est}^{n}(f)\mathbf{X}_{est}^{n,H}(f)\} = \mathcal{E}\{\mathbf{X}^{n}(f)\mathbf{X}^{n,H}(f)\} + \delta\left(\mathcal{E}\{\tilde{\mathbf{X}}^{n}(f)\tilde{\mathbf{X}}^{n,H}(f)\} - \mathcal{E}\{\mathbf{X}^{n}(f)\mathbf{X}^{n,H}(f)\}\right), \quad (56)$$

where $\mathcal{E}\{\mathbf{X}^{n}(f)\mathbf{X}^{n,H}(f)\}$ corresponds to the new noise correlation matrix after the change in noise scenario, $\mathcal{E}\{\tilde{\mathbf{X}}^{n}(f)\tilde{\mathbf{X}}^{n,H}(f)\}$ corresponds to the correlation matrix of the previous noise scenario and the scalar $\delta \in [0,1]$ measures how much $\mathcal{E}\{\mathbf{X}_{est}^{n}(f)\mathbf{X}_{est}^{n,H}(f)\}$ is perturbed by the non-stationarity. The parameter δ can be interpreted as a measure for the amount of non-stationarity. The error (56) in the noise correlation matrix leads to a wrong estimate $\mathcal{E}\{\mathbf{X}_{est}^{s}(f)\mathbf{X}_{est}^{s,H}(f)\}$ of $\mathcal{E}\{\mathbf{X}^{s}(f)\mathbf{X}^{s,H}(f)\}$ in the MWF $\mathbf{W}(f)$, i.e.,

$$\mathcal{E}\{\mathbf{X}_{est}^{s}(f)\mathbf{X}_{est}^{s,H}(f)\} = \mathcal{E}\{\mathbf{X}^{s}(f)\mathbf{X}^{s,H}(f)\} + \delta\left(\mathcal{E}\{\mathbf{X}^{n}(f)\mathbf{X}^{n,H}(f)\} - \mathcal{E}\{\tilde{\mathbf{X}}^{n}(f)\tilde{\mathbf{X}}^{n,H}(f)\}\right), \quad (57)$$

resulting in

$$\begin{aligned}\mathbf{W}(f) &= \mathcal{E}\{\mathbf{X}(f)\mathbf{X}^{H}(f)\}^{-1}\mathcal{E}\{\mathbf{X}^{s}(f)\mathbf{X}^{s,H}(f)\}\mathbf{e_1} + \boldsymbol{\Delta}\mathbf{W}(f),\\ \Delta\mathbf{W}(f) &= \mathcal{E}\{\mathbf{X}(f)\mathbf{X}^{H}(f)\}^{-1}\delta\left(\mathcal{E}\{\mathbf{X}^{n}(f)\mathbf{X}^{n,H}(f)\} - \mathcal{E}\{\tilde{\mathbf{X}}^{n}(f)\tilde{\mathbf{X}}^{n,H}(f)\}\right)\mathbf{e}_1. \quad (58)\end{aligned}$$

As an example, Figure 9 plots the performance measures $\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$ of the MWF, the GSC and the GSC-NSC with $\eta = -10$ dB as a function of δ with $\mathcal{E}\{\mathbf{X}^{n}(f)\mathbf{X}^{n,H}(f)\}$ and $\mathcal{E}\{\tilde{\mathbf{X}}^{n}(f)\tilde{\mathbf{X}}^{n,H}(f)\}$ computed using

$$P_{X_k X_l}^{n}(f) = P_{X}^{s}(f)\frac{0.8\exp(j2\pi f\cos\theta(d_l-d_k)/c) + 0.2\text{sinc}(2\pi f(d_l-d_k)/c)}{1+\beta(f)}, \quad (59)$$

$$P_{X_k \tilde{X}_l}^{n}(f) = \Delta\alpha(f)P_{X}^{s}(f)\frac{0.8\exp(j2\pi f\cos(\theta+\Delta\theta)(d_l-d_k)/c) + 0.2\text{sinc}(2\pi f(d_l-d_k)/c)}{1+\beta(f)\Delta\alpha(f)}, \quad (60)$$

where $\Delta\alpha(f) = P_{\tilde{X}}^{n}(f)/P_{X}^{n}(f)$ represents the difference in PSD between the new and the previous noise scenario and $\Delta\theta$ the difference in angle of the directional noise component. Figure 9 depicts the results for $\theta = 180°$, $\beta(f) = -30$ dB with $\Delta\theta = 15°$ and $\Delta\alpha(f) = 0$ dB and $\Delta\alpha(f) = 2$ dB, respectively. As expected, the MWF is more sensitive to a wrongly estimated noise correlation matrix than the GSC and the GSC-NSC because of the error in $\mathcal{E}\{\mathbf{X}_{est}^{s}(f)\mathbf{X}_{est}^{s,H}(f)\}$ (see (57)). Note that the GSC will perform worse than in Figure 9 if other signal model errors such as mismatch, occur. In particular, a change in PSD, i.e., $\Delta\alpha(f) \neq 0$ dB, causes a large error in $\mathcal{E}\{\mathbf{X}_{est}^{s}(f)\mathbf{X}_{est}^{s,H}(f)\}$ and hence, a significant performance degradation. Especially the noise reduction performance $\Delta\text{SNR}_{\text{intellig}}$ is affected. The increase in distortion $\text{SD}_{\text{intellig}}$ is small for this scenario: the deviation $\Delta\mathbf{W}(f)$ in (58) from the ideal MWF corresponds to an MWF that suppresses the desired speech signal $\mathbf{X}^{s}(f)$ so that the distortion $\text{SD}_{\text{intellig}}$ is limited. Simulations indicated that -in contrast to microphone mismatch for the

GSC- inaccurate estimation of the second order statistics does not degrade the MWF performance more for a high noise sensitivity $\Psi(f)$ (or at low frequencies f) than for a low $\Psi(f)$ (or at high frequencies f).

Figure 9 illustrates that the MWF may lose its benefit over the GSC-NSC in non-stationary noise scenarios due to wrong estimation of the second order statistics. The loss in performance of the MWF will depend on the non-stationarity of the noise scenario and the implementation of the MWF (e.g., the forgetting factor that is used in the update formula of the speech + noise and noise correlation matrix and the speech detection algorithm).

VI. Validation through experimental results

The theoretical robustness analysis, presented in Section V, assumes stationary signals, free field propagation of speech and noise, infinite filter lengths and -apart from Section V-C- perfect estimation of the second order statistics of speech and noise. In practice, finite filter lengths are used, reverberation is present and signals are non-stationary. Therefore, in this section, we verify the major conclusions of Section V based on real non-stationary recordings in a reverberant room and when finite filters are used. In addition, the second order statistics of speech + noise and noise are estimated from a finite amount of incoming data and at a different moment in time. In the experiments, we assume a perfect speech detection.

The results are illustrated for a multiple noise source scenario consisting of 5 spectrally non-stationary, multi-talker babble noise sources positioned at 75°, 120°, 180°, 240° and 285°. A BTE with 3 omnidirectional microphones (Knowles FG-3452) with interspacings $d_2 - d_1 = 0.01$ m and $d_3 - d_2 = 0.015$ m has been mounted on a dummy head in an office room. The reverberation time $T_{60\,\mathrm{dB}}$ is about 700 ms for a speech weighted noise. The desired signal source is positioned at a distance of 1 meter in front of the head and consists of English sentences spoken by a male speaker from the HINT-database [43]. The sampling frequency is $f_s = 16$ kHz. The desired and noise signals are uncorrelated. The desired signal and the total noise signal have a level of 70 dB SPL at the center of the head. The input SNR at the first microphone equals 0 dB.

The fixed stage of the GSC has been designed for the BTE mounted on the dummy head so that -for an anechoic signal coming from 0°- the energy of the noise references is minimized. To improve intelligibility, the microphone signals are pre-emphasized in the GSC prior to processing [44]. At the GSC output, de-emphasis is performed. The ANC filters are implemented using exponentially weighted RLS with forgetting factor λ^2. A recursive subband GSVD-based implementation of the MWF is used with forgetting factors $\lambda_{s+n}^{2D} = \lambda_n^{2D}$ where λ_{s+n}^2 and λ_n^2 are the equivalent forgetting factors in a fullband

implementation and $D = 12$ is the downsampling ratio [23]. In the subband GSVD algorithm, the subband implementation intrinsically performs a pre-whitening operation [23]. The filters are sufficiently long so that a further increase in filter length does not result in a significant increase in performance.

The robustness of the GSC and the MWF to two types of model errors, i.e., microphone mismatch and variation in speaker location is investigated in Section VI-A and Section VI-B. In the experiments, the forgetting factors $\lambda^2, \lambda_{s+n}^2$ and λ_n^2 of MWF and GSC are set to 0.99998. This implies the assumption of a quasi-stationary environment during at least $2/[f_s(1-0.99998)] \approx 6$ seconds. A larger λ^2 results in a better estimation of the second order statistics in quasi-stationary environments, while for tracking in highly non-stationary environments, a smaller λ^2 may be preferred. Section VI-C studies the effect of the forgetting factor λ^2, λ_{s+n}^2 and λ_n^2 on the performance of the GSC and the MWF. In addition, the influence of non-stationary noise on the performance of the MWF and the GSC is investigated.

A. Robustness to microphone mismatch

1) Influence on PTFs of speech and noise: Figure 10 shows the PTF $G^s(f)$ and $G^n(f)$ of GSC and MWF when the second microphone has a gain mismatch of $\upsilon(f) = 2$ dB with respect to the first microphone. A small amount of gain mismatch is indeed detrimental to the performance of GSC at low $f \cdot d$. An additional constraint on the noise sensitivity is indispensable to guarantee robustness in complicated noise scenarios. The MWF on the other hand is independent of the applied deviation, resulting in improved noise reduction. Compared to Figure 5, the MWF attenuates frequencies around $2f/f_s = 0.2$ because the input SNR of the real speech and noise is lower around $2f/f_s = 0.2$ than at other f, resulting in attenuation by the spectral Wiener filter $W_2(f)$ (see (40)).

Figure 11 plots the noise sensitivity $\Psi(f)$ for both algorithms. For comparison, the theoretical $\Psi(f)$ for internal-to-external noise ratio $\beta = -30$ dB is depicted too. Theoretically, the noise sensitivity $\Psi(f)$ of GSC and MWF coincide (see Section V-A): the small differences are due to reverberation. At low f, the noise sensitivity $\Psi(f)$ is larger than the theoretical $\Psi(f)$ for $\beta = -30$ dB as $\beta(f)$ is lower than -30 dB due to the high energy of the speech-like external noise source at low frequencies. As a consequence, the noise signal is attenuated more than in Figure 5.

2) Influence on intelligibility weighted performance measures: In Figure 12(a)-(b) and Figure 12(c)-(d), the performance measures $\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$ of GSC and MWF are given as a function of the amount of microphone gain mismatch υ and microphone phase mismatch ϕ, respectively, at the second microphone. For comparison, the experimentally found performance of the GSC with $\beta(f)$ increased to about -30 dB is also shown. Although differences occur between Figure 5(a)-(d) and Figure 12(a)-(d) due

to reverberation, a different $\beta(f)$, the presence of the head and non-stationary signals, the experimental results are quite well predicted by the theory: the performance of the MWF is indeed independent of v and ϕ, whereas a small amount of gain mismatch suffices to deteriorate the performance of the GSC. Again, phase mismatch is found to be less detrimental than gain mismatch. In the absence of mismatch, the performance of the MWF is a bit worse than the one of GSC: as explained in Section IV, the MWF tries to estimate the reverberant speech component in the first microphone signal and, hence, trades off noise reduction and cancellation of the reverberant part in the speech. The GSC on the other hand only tries to estimate the direct path. The speech distortion of the MWF is larger than in Figure 5, due to the spectral filtering (see also Figure 10). The performance curves of the GSC-NSC are depicted too. They have been obtained by injecting spatially uncorrelated noise in the pre-emphasized microphone signals that are used to compute the filters. The noise injection levels are approximately set to $\eta(f) \approx -20$ dB and $\eta(f) \approx -10$ dB with respect to the power of the noise in the pre-emphasized, first microphone signal. The curves for $\eta \approx -20$ dB and -10 dB confirm the ones predicted in Figure 5: a noise sensitivity constraint results in a smaller performance degradation in case of gain and/or phase mismatch at the expense of less noise reduction. Given the amount of mismatch that may occur, an appropriate constraint on $\Psi(f)$ should be chosen so that the performance is not severely affected in the presence of the mismatch: to allow for a gain mismatch up to 6 dB, $\eta(f) = -10$ dB is a good choice. The severer the constraint, the less the noise reduction that is achieved with the robust GSC-NSC. The MWF outperforms the robust GSC-NSC, even when the noise is spectrally non-stationary.

B. Robustness to variations in speaker location

Figure 13 illustrates the effect of a variation $\Delta\Theta$ (in degrees) in speaker location with respect to $\Theta = 0^\circ$ on the improvement $\Delta\text{SNR}_{\text{intellig}}$ and on the speech distortion $\text{SD}_{\text{intellig}}$ for the MWF, the GSC and the robust GSC-NSC with $\eta(f) = -10$ dB. As predicted theoretically, no significant degradation of the performance of the GSC occurs for speaker variations $\Delta\Theta$ up to 45° and -45°. In contrast to Figure 8, the performance measures are asymmetric for positive and negative $\Delta\Theta$. As explained in Section V-B, this is due to the head shadow effect. Speech arriving from negative angles $\Delta\Theta$ is shielded by the head, hence negative $\Delta\Theta$ result in a higher input noise-to-signal ratio $\alpha(f)$. As shown in Figure 8, a higher $\alpha(f)$ induces a larger amount of speech distortion by the MWF.

C. Estimation of second order statistics

In the previous experiments, the forgetting factors $\lambda^2, \lambda_{s+n}^2$ and λ_n^2 of GSC and MWF were set to 0.99998 to allow for tracking of slowly time-varying environments. A larger λ^2 results in a more accurate estimation of the second order statistics in quasi-stationary environments, while for tracking in highly non-stationary environments, a smaller λ^2 is preferred.

Figure 14 plots the improvement $\Delta\text{SNR}_{\text{intellig}}$ of GSC and MWF as a function of the time window $1/\left[(1-\lambda^2)f_s\right]$ (in seconds) with $\lambda^2 = \lambda_{s+n}^2 = \lambda_n^2$ for spectrally non-stationary (i.e., multi-talker babble) as well as stationary (i.e., stationary speech-like) noise sources. For *spectrally stationary noise sources*, $\Delta\text{SNR}_{\text{intellig}}$ for GSC and MWF is not affected by a decrease in λ^2. However, for small λ_{s+n}^2 and λ_n^2, the MWF depends on the short-term characteristics of the speech and noise sources so that short-time effects such as musical noise may occur. For *spectrally non-stationary noise sources*, the MWF is affected more than the GSC by a decrease in λ^2, especially for small λ^2: because of the short-time non-stationary spectrum, the noise PSD estimated during speech + noise differs from the noise PSD estimated during noise only, especially when λ^2 is small. As demonstrated in Section V-C, this results in a large error in the speech correlation matrix estimated by the MWF, and consequently, in a decrease in $\Delta\text{SNR}_{\text{intellig}}$. For time windows $1/[(1-\lambda^2)f_s]$ larger than 0.5 s, the performance loss of MWF is limited: the MWF then depends on the average, long-term characteristics of speech and noise so that spectrally non-stationary noise is successfully suppressed. Hence, the MWF preserves its benefit in robustness and performance over the GSC and the GSC-NSC for spectrally non-stationary noises, provided that the environment and long-term average PSD of the noise vary relatively slowly in time (i.e., on the order of seconds).

In *non-stationary environments*, the benefit of the MWF over the GSC-NSC may be lost. Due to the non-continuous estimation of the second order statistics, the tracking capability of non-stationarities of the noise is reduced, which leads to inaccurate estimates of the second order statistics. As an illustration, Figure 15 depicts the speech cancellation and noise suppression by the MWF and the GSC-NSC with $\eta \approx -20$ dB and $\lambda^2 = \lambda_{s+n}^2 = \lambda_n^2 = 0.99995$ for a sudden change in position and power of a stationary speech-like noise source. The position of the noise source changes at sample 1.8×10^5 from 180° to 150° and its power increases by 1.5 dB. Between sample 1.68×10^5 and 2×10^5, speech + noise is present; the other samples contain noise only. The speech cancellation/noise suppression are measured as the ratio of the output speech/noise power over the speech/noise power at the first microphone. As predicted in Section V-C, especially the noise suppression of the MWF is degraded by the change: during speech + noise, the noise correlation matrix is not updated, which results in a wrong estimate of the

speech correlation matrix by the MWF. The speech cancellation by the MWF is less affected.

VII. Conclusions

In this paper, we presented a unified theoretical framework to assess both the broadband performance of the GSC, the GSC-NSC and the MWF and their robustness to different types of model errors for different noise scenarios. The robustness has been analyzed for small-sized arrays as used in hearing aids and in terms of speech intelligibility weighted performance measures. The theoretical results are verified through experiments with a BTE hearing aid.

Theoretical analysis showed that sensitivity to errors in the assumed signal model grows with an increasing complexity of the noise scenario, with decreasing $f \cdot d$ and increasing number of microphones per array length. Among the different model errors, microphone gain mismatch is particularly harmful to GSC, even if the ANC is adapted during noise only. As a consequence, a noise sensitivity constraint is essential, which goes at the expense of less noise reduction. The MWF on the other hand, is not degraded by signal model errors, so that - compared to a robust GSC-NSC - more noise is reduced. However, the MWF is quite sensitive to wrong estimation of the second order statistics of speech and noise so that its benefit over the GSC-NSC may be lost in non-stationary noise scenarios.

The theoretical results have been validated for real, non-stationary recordings in an office room made with a 3-microphone BTE. For all types of model errors that have been studied, the experimental results confirm the theoretical ones. Hence, the analytical framework offers an easy way to predict the consequences of model errors on the performance of noise reduction algorithms. In addition, the experiments showed that the MWF preserves its benefit in robustness and performance over the GSC and the GSC-NSC for spectrally short-time non-stationary noise, provided that the environment and the long-term average PSD of the noise vary relatively slowly in time (on the order of seconds). In non-stationary environments, the benefit of the MWF over the GSC-NSC may be lost.

Acknowledgments

The authors would like to thank the reviewers for their valuable comments and suggestions.

References

[1] R. W. Stadler and W. M. Rabinowitz, "On the potential of fixed arrays for hearing aids," *J. Acoust. Soc. Amer.*, vol. 94, no. 3, pp. 1332–1342, Sept. 1993.

[2] P. M. Peterson, *Adaptive array processing for multiple microphone hearing aids*, Ph.D. thesis, Dept. Elect. Eng. and Comp. Sci., M.I.T, Cambridge, MA, 1989, available as Res. Lab. Elect. Techn. Rept. 541.

[3] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust Adaptive Beamforming," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 35, no. 10, pp. 1365–1376, 1987.

[4] B. D. Van Veen and K. M. Buckley, "Beamforming: A Versatile Approach to Spatial Filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[5] J. E. Greenberg and P. M. Zurek, "Evaluation of an Adaptive Beamforming Method for Hearing Aids," *J. Acoust. Soc. Amer.*, vol. 91, no. 3, pp. 1662–1676, Mar. 1992.

[6] J. Vanden Berghe and J. Wouters, "An adaptive noise canceller for hearing aids using two nearby microphones," *J. Acoust. Soc. Amer.*, vol. 103, pp. 3621–3626, June 1998.

[7] D. Van Compernolle, "Switching Adaptive Filters for Enhancing Noisy and Reverberant Speech from Microphone Array Recordings," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Albuquerque, Apr. 1990, vol. 2, pp. 833–836.

[8] M. W. Hoffman and K. M. Buckley, "Robust time-domain processing of broadband acoustic data," *IEEE Trans. Speech, Audio Processing*, vol. 3, pp. 193–203, May 1995.

[9] D. P. Welker, J. E. Greenberg, J. G. Desloge, and P. M. Zurek, "Microphone-Array Hearing Aids with Binaural Output–Part II: A Two-Microphone Adaptive System," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 5, no. 6, pp. 543–551, Nov. 1997.

[10] J. M. Kates and M. R. Weiss, "A comparison of hearing-aid array-processing techniques," *J. Acoust. Soc. Amer.*, vol. 99, no. 5, pp. 3138–3148, May 1996.

[11] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A Robust Adaptive Beamformer for Microphone Arrays with a Blocking Matrix Using Constrained Adaptive Filters," *IEEE Trans. Signal Processing*, vol. 47, pp. 2677–2683, 1999.

[12] F. Luo, J. Yang, C. Pavlovic, and A. Nehorai, "Adaptive Null-Forming Scheme in Digital Hearing Aids," *IEEE Trans. Signal Processing*, vol. 50, no. 7, pp. 1583–1590, July 2002.

[13] N. K. Jablon, "Adaptive beamforming with the Generalized Sidelobe Canceller in the presence of array imperfections," *IEEE Trans. Antennas Propag.*, vol. 34, pp. 996–1012, Aug. 1986.

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

[15] S. Nordebo, I. Claesson, and S. Nordholm, "Adaptive beamforming: spatial filter designed blocking matrix," *IEEE Journal of Oceanic Engineering*, vol. 19, no. 4, pp. 583–590, Oct. 1994.

[16] W. Herbordt, H. Buchner, and W. Kellermann, "An Acoustic Human-Machine Front-End for Multimedia Applications," *EURASIP journal on Applied Signal Processing*, vol. 1, pp. 21–31, Jan. 2003.

[17] W. Herbordt and W. Kellermann, *Adaptive Beamforming for Audio Signal Acquisition*, chapter 6 in "Adaptive Signal Processing: Applications to Real-World Problems" (Benesty, J. and Huang, Y., Eds.), pp. 155–188, Springer-Verlag, 2003.

[18] K. U. Simmer, J. Bitzer, and C. Marro, *Post-filtering techniques*, chapter 3 in "Microphone Arrays: Signal Processing Techniques and Applications" (Brandstein, M. S. and Ward, D. B., Eds.), pp. 39–57, Springer-Verlag, May 2001.

[19] S. Doclo and M. Moonen, *GSVD-Based Optimal Filtering for Multi-Microphone Speech Enhancement*, chapter 6 in "Microphone Arrays: Signal Processing Techniques and Applications" (Brandstein, M. S. and Ward, D. B., Eds.), pp. 111–132, Springer-Verlag, May 2001.

[20] S. Doclo and M. Moonen, "GSVD-Based Optimal Filtering for Single and Multimicrophone Speech Enhancement," *IEEE Trans. Signal Processing*, vol. 50, no. 9, pp. 2230–2244, Sept. 2002.

[21] S. Doclo, *Multi-microphone noise reduction and dereverberation techniques for speech applications*, Ph.D. thesis, Faculty of Engineering, K.U. Leuven, Leuven, Belgium, May 2003, available as Tech. Rep. ESAT-SISTA/TR 03-77.

[22] G. Rombouts and M. Moonen, "QRD-based unconstrained optimal filtering for acoustic noise reduction," *Signal Processing*, vol. 83, no. 9, pp. 1889–1904, Sept. 2003.

[23] A. Spriet, M. Moonen, and J. Wouters, "A multi-channel subband generalized singular value decomposition approach to speech enhancement," *European Transactions on Telecommunications*, vol. 13, no. 2, pp. 149–158, Mar.-Apr. 2002.

[24] S. Nordholm, I. Claesson, and M. Dahl, "Adaptive microphone array employing calibration signals: an analytical evaluation," *IEEE Trans. Speech, Audio Processing*, vol. 7, no. 3, pp. 241–22, May 1999.

[25] S. Nordholm, I. Claesson, and N. Grbíc, *Optimal and adaptive microphone arrays for speech input in automobiles*, chapter 14 in "Microphone Arrays: Signal Processing Techniques and Applications" (Brandstein, M. S. and Ward, D. B., Eds.), pp. 307–326, Springer-Verlag, May 2001.

[26] C. Marro, Y. Mahieux, and K. U. Simmer, "Analysis of Noise Reduction and Dereverberation Techniques Based on Microphone Arrays with Postfiltering," *IEEE Trans. Speech, Audio Processing*, vol. 6, no. 3, pp. 240–259, May 1998.

[27] J. Bitzer, K. U. Simmer, and K.-D. Kammeyer, "Multichannel noise reduction - Algorithms and theoretical limits-," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Rhodes, Greece, Sept. 1999, vol. 1, pp. 105–108.

[28] J. Bitzer, K. U. Simmer, and K.-D. Kammeyer, "Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Phoenix, Arizona, USA, May 1999, vol. 5, pp. 2965–2968.

[29] J. Bitzer, K. U. Simmer, and K.-D. Kammeyer, "Multi-microphone noise reduction techniques as front-end devices for speech recognition," *Speech Communication*, vol. 34, pp. 3–12, Apr. 2001.

[30] S. Gannot, Burshtein D., and Weinstein E., "Theoretical analysis of the general transfer function GSC," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Darmstadt, Germany, Sept. 2001, pp. 103–106.

[31] S. E. Nordholm and Y. H. Leung, "Performance limits of the broadband generalized sidelobe cancelling structure in an isotropic noise field," *J. Acoust. Soc. Amer.*, vol. 107, no. 2, pp. 1057–1060, Feb. 2000.

[32] H. Cox, R. M. Zeskind, and T. Kooij, "Resolving power and sensitivity to mismatch of optimum array processors," *Journal of the Acoustical Society of America*, vol. 54, no. 3, pp. 771–785, 1974.

[33] A. M. Vural, "Effects of perturbations on the performance of optimum/adaptive arrays," *IEEE Trans. Aerospace and Electronic Systems*, vol. 15, no. 1, pp. 76–87, Jan. 1979.

[34] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation and Modulation Theory*, John Wiley, 2002.

[35] P. M. Peterson, A.-M. Wie, W. M. Rabinowitz, and P. M. Zurek, "Robustness of an adaptive beamforming method for hearing aids," *Acta Otolaryngol. Suppl.*, vol. 469, pp. 85–90, 1990.

[36] A. Spriet, M. Moonen, and J. Wouters, "The impact of speech detection errors on the noise reduction performance of multi-channel Wiener filtering and Generalized Sidelobe Cancellation," in *Signal Processing*, to be published. Short version in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Hong Kong, China, Apr. 2003, vol. 5, pp. 501–504.

[37] M. J. Jacobson, "Space-time correlation in spherical and circular noise fields," *Journal of the Acoustical Society of America*, vol. 34, no. 7, pp. 971–978, July 1962.

[38] J. E. Greenberg, P. M. Peterson, and P. M. Zurek, "Intelligibility-weighted measures of speech-to-interference ratio and speech system performance," *J. Acoust. Soc. Amer.*, vol. 94, no. 5, pp. 3009–3010, Nov. 1993.

[39] Acoustical Society of America, "ANSI S3.5-1997 American National Standard Methods for calculation of the speech intelligibility index," June 1997.

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

[41] International Collegium of Rehabilitative Audiology, "Noise Signals ICRA Compact Disc," Ver 0.3 1997.

[42] L. B. Jensen, "Hearing Aid with adaptive Matching of Input Transducers," U.S. patent 0041696, Apr. 2002.

[43] M. Nilsson, S. D. Soli, and A. Sullivan, "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Amer.*, vol. 95, no. 2, pp. 1085–1099, Feb. 1994.

[44] M. J. Link and K. M. Buckley, "Prewhitening for intelligibility gain in hearing aid arrays," *J. Acoust. Soc. Amer.*, vol. 93, no. 4, pp. 2139–2140, Apr. 1993.



Ann Spriet (S'01) was born in Izegem, Belgium, in 1976. She received the M.Sc. degree in electrical engineering degree and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1999 and 2004, respectively. From 1999 until 2003, she was a Research Assistant of the Fonds voor Wetenschappelijk Onderzoek - Vlaanderen. Currently, she is a a postdoctoral researcher at the Electrical Engineering Department (ESAT) of the Katholieke Universiteit Leuven, in the framework of a collaboration with Katholieke Universiteit Leuven - Laboratory Experimental ORL. Her research interests are in digital signal processing for speech enhancement in hearing aids and cochlear implants. She received the 2001 Lapperre Award and a Best Student Paper Award at the International Workshop on Acoustic Echo and Noise Control in 2003.



Marc Moonen (M'94) received the Electrical Engineering degree and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1986 and 1990, respectively. Since 2000 he has been an Associate Professor with the Electrical Engineering Department, Katholieke Universiteit Leuven, where he is currently heading a research team of sixteen Ph.D. candidates and postdocs, working in the area of signal processing for digital communications, wireless communications, DSL, and audio signal processing. Dr. Moonen received the 1994 KU Leuven Research Council Award, the 1997 Alcatel Bell (Belgium) Award (with P. Vandaele) and was a 1997 "Laureate of the Belgium Royal Academy of Science." He was Chairman of the IEEE Benelux Signal Processing Chapter (1998-2002), and is currently a EURASIP AdCom Member (European Association for Signal, Speech, and Signal Processing, 2000). He is Editor-in Chief for the *EURASIP Journal on Applied Signal Processing* (2003), and a member of the editorial board of *Integration, the VLSI Journal, IEEE Transactions on Circuits and Systems II, EURASIP Journal on Wireless Communications and Networking, and IEEE Signal Processing Magazine*.



Jan Wouters was born in Leuven, Belgium, in 1960. He received the physics degree and the Ph.D. degree in sciences / physics from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1982 and 1989, respectively. From 1989 till 1992 he has been a Research Fellow with the Belgian NFWO (National Fund for Scientific Research) at the Institute of Nuclear Physics (UCL Louvain-la-Neuve and KU Leuven) and at NASA Goddard Space Flight Center (USA). From 1993 he is a Professor at the Neuro sciences Department of the Katholieke Universiteit Leuven. His research activities are about audiology and the auditory system, signal processing for cochlear implants and hearing aids. Dr. Wouters received in 1989 an Award of the Flemish Ministry, in 1992 a Fullbright Award and a NATO Research Fellowship, the 1996 Flemish VVL Speech therapy - Audiology Award. He is member of the International Collegium for Rehabilitative Audiology, a Board Member of the NAG (Dutch Acoustical Society) and responsible for the Lab. Exp. ORL, author of 75 articles in international peer-review journals, reviewer for several international journals.

LIST OF FIGURES

1 Concept of the Generalized Sidelobe Canceller. 32

2 Concept of the multi-channel Wiener filter. 32

3 Influence of the noise scenario on the noise sensitivity $\Psi_{\text{GSC}}(f) = \Psi_{\text{MWF}}$ (in dB). 32

4 Influence of the number of microphones M on the noise sensitivity $\Psi_{\text{GSC}}(f) = \Psi_{\text{MWF}}$ (in dB) (upper plot) and the improvement in SNR $\frac{G^s(f)}{G^n(f)}$ (in dB) (lower plot) for the GSC and the MWF in a diffuse noise field with $\beta(f) = -30$ dB. 33

5 Effect of microphone gain mismatch υ and phase mismatch ϕ on the PTF $G^s(f)$ and $G^n(f)$ of GSC and MWF for a multiple noise source scenario and $\beta(f) = -30$ dB. 33

6 Performance ((a)-(c) $\Delta\text{SNR}_{\text{intellig}}$; (b)-(d) $\text{SD}_{\text{intellig}}$) of GSC, GSC-NSC and MWF in a multiple noise source scenario as a function of microphone mismatch; (a)-(b) Microphone gain mismatch υ (in dB); (c)-(d) Microphone phase mismatch ϕ (in degrees). 34

7 Effect of a deviation Δd (in meters) in the position d_k of the 2nd microphone on the PTF $G_s(f)$ and $G_n(f)$ of the GSC and the MWF. 35

8 Influence of a variation in speaker location on the broadband performance of the GSC and the MWF for noise-to-signal ratios $\alpha = 0$ dB and $\alpha = 3$ dB. 35

9 Influence of inaccurate estimation of the noise correlation matrix on the performance of the MWF, the GSC and the GSC-NSC ($\eta = -10$ dB). 36

10 Effect of microphone gain mismatch υ on the PTF $G^s(f)$ and $G^n(f)$ of GSC and MWF for a non-stationary multiple noise source scenario. 36

11 Experimentally obtained noise sensitivity of the GSC and the MWF in a multiple noise-source scenario. 37

12 Broadband performance measures ((a)-(c) $\Delta\text{SNR}_{\text{intellig}}$; (b)-(d) $\text{SD}_{\text{intellig}}$) of GSC, GSC-NSC and MWF in a multiple noise source scenario as a function of the amount of microphone mismatch; (a)-(b) Microphone gain mismatch υ (in dB); (c)-(d) Microphone phase mismatch ϕ (in degrees). 38

13 Influence of a variation $\Delta\Theta$ (in degrees) in speaker location on the performance of the GSC, the GSC-NSC and the MWF. 39

14 Influence of forgetting factor $\lambda^2 = \lambda_{s+n}^2 = \lambda_n^2$ on the improvement $\Delta\text{SNR}_{\text{intellig}}$ of GSC and MWF for spectrally non-stationary and stationary noise sources, respectively. 39

15 Influence of change in noise scenario on the speech cancellation and noise suppression by the GSC-NSC (with $\eta \approx -20$ dB) and the MWF. 40

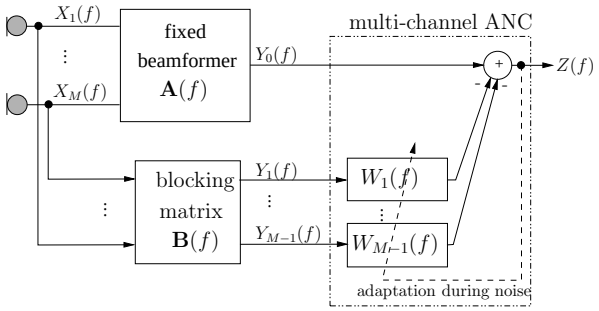


Fig. 1. Concept of the Generalized Sidelobe Canceller.

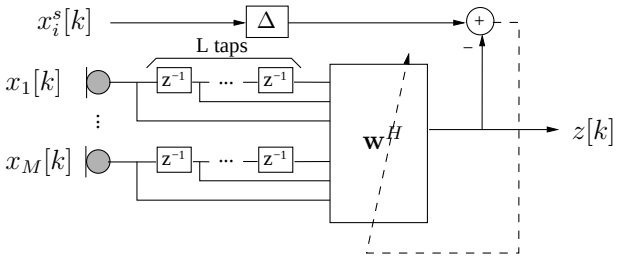


Fig. 2. Concept of the multi-channel Wiener filter.

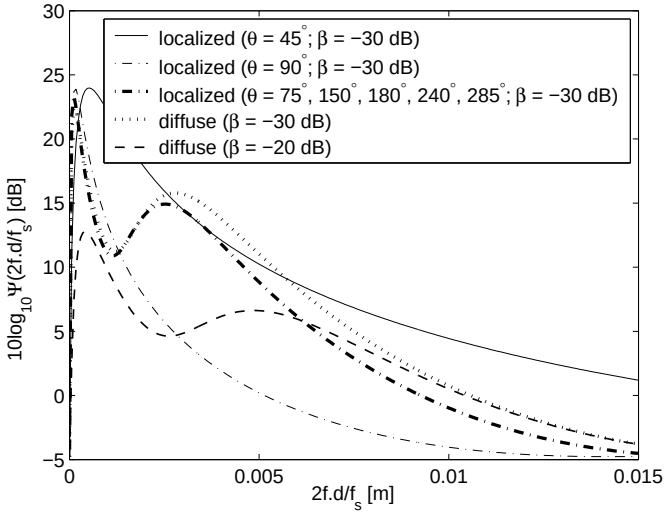


Fig. 3. Influence of the noise scenario on the noise sensitivity $\Psi_{\mathrm{GSC}}(f) = \Psi_{\mathrm{MWF}}$ (in dB).

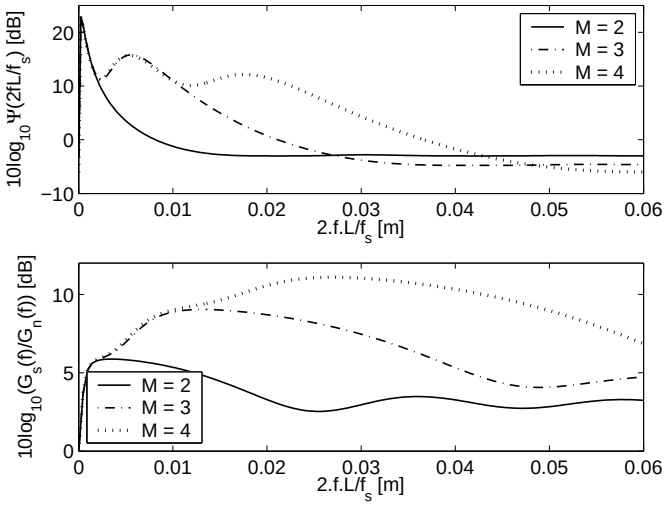


Fig. 4. Influence of the number of microphones M on the noise sensitivity $\Psi_{\text{GSC}}(f) = \Psi_{\text{MWF}}$ (in dB) (upper plot) and the improvement in SNR $\frac{G^s(f)}{G^n(f)}$ (in dB) (lower plot) for the GSC and the MWF in a diffuse noise field with $\beta(f) = -30$ dB.

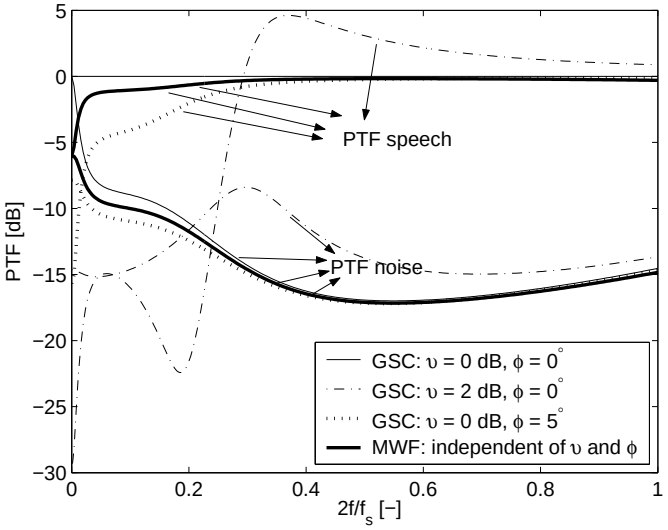


Fig. 5. Effect of microphone gain mismatch υ and phase mismatch ϕ on the PTF $G^s(f)$ and $G^n(f)$ of GSC and MWF for a multiple noise source scenario and $\beta(f) = -30$ dB.

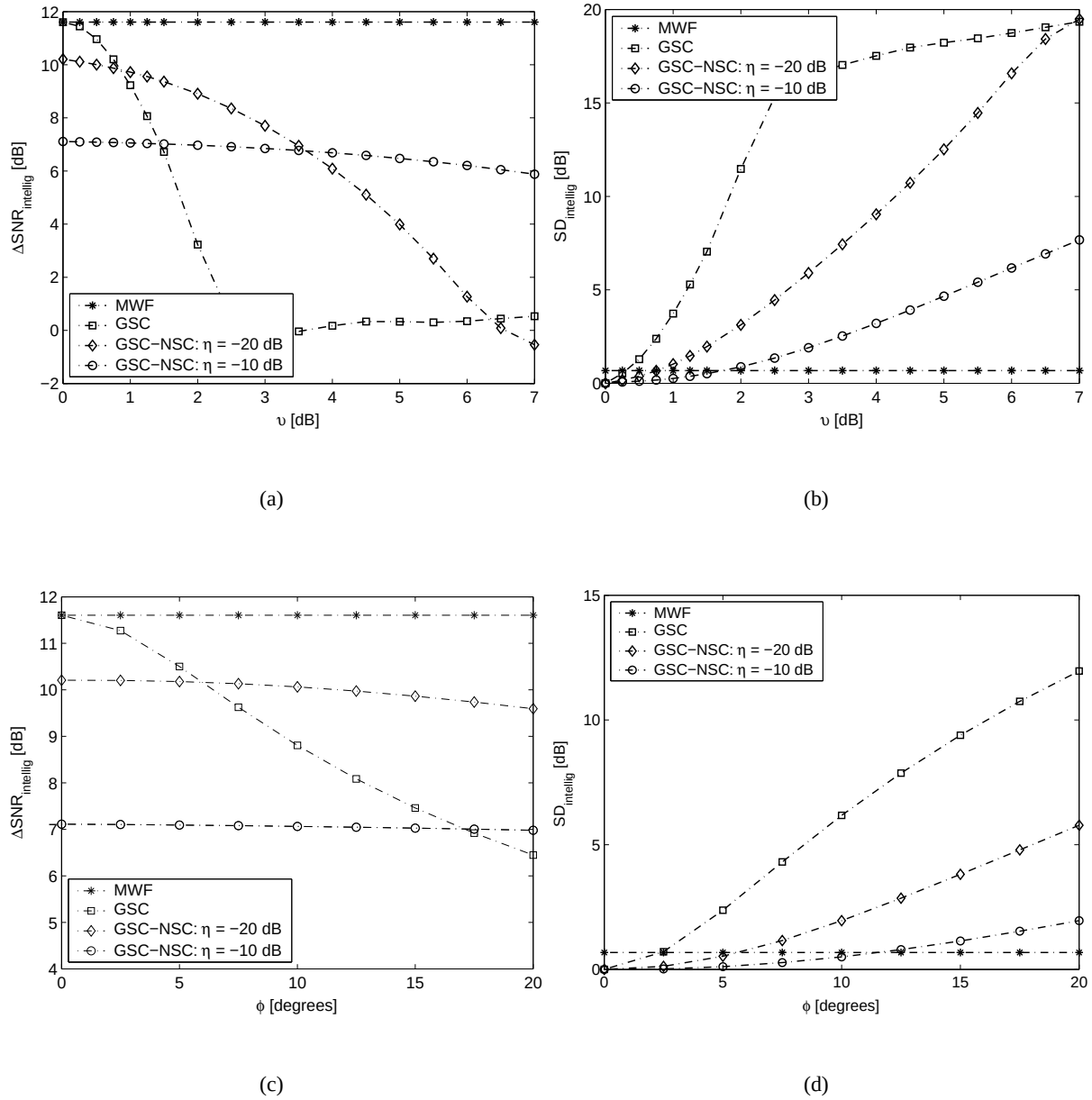


Fig. 6. Performance ((a)-(c) $\Delta\text{SNR}_{\text{intellig}}$; (b)-(d) $\text{SD}_{\text{intellig}}$) of GSC, GSC-NSC and MWF in a multiple noise source scenario as a function of microphone mismatch; (a)-(b) Microphone gain mismatch υ (in dB); (c)-(d) Microphone phase mismatch ϕ (in degrees).

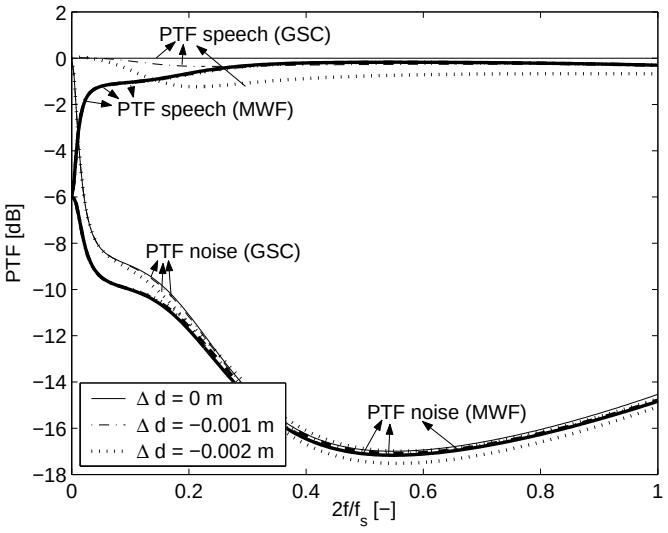


Fig. 7. Effect of a deviation Δd (in meters) in the position d_k of the 2nd microphone on the PTF $G_s(f)$ and $G_n(f)$ of the GSC and the MWF.

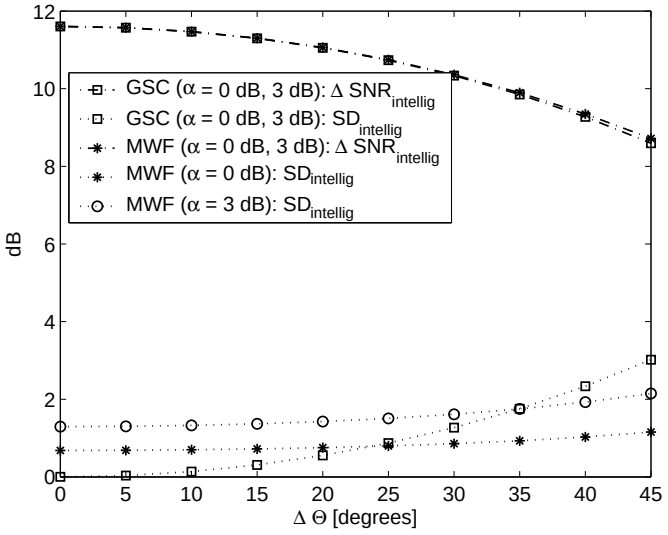


Fig. 8. Influence of a variation in speaker location on the broadband performance of the GSC and the MWF for noise-to-signal ratios $\alpha = 0$ dB and $\alpha = 3$ dB.

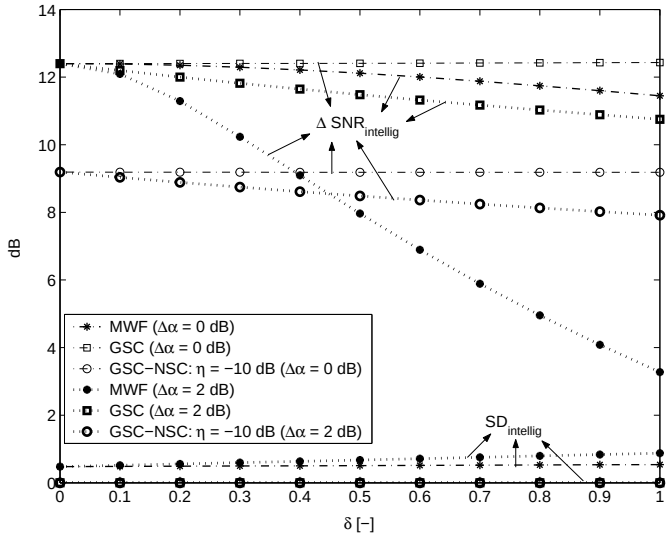


Fig. 9. Influence of inaccurate estimation of the noise correlation matrix on the performance of the MWF, the GSC and the GSC-NSC ($\eta = -10$ dB).

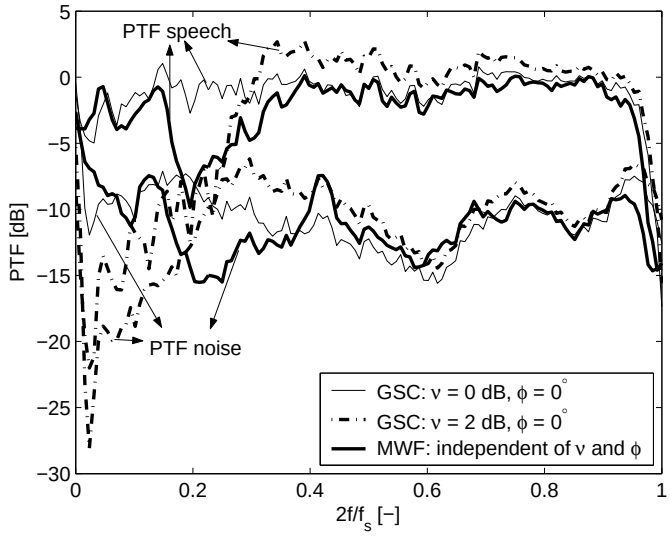


Fig. 10. Effect of microphone gain mismatch v on the PTF $G^s(f)$ and $G^n(f)$ of GSC and MWF for a non-stationary multiple noise source scenario.

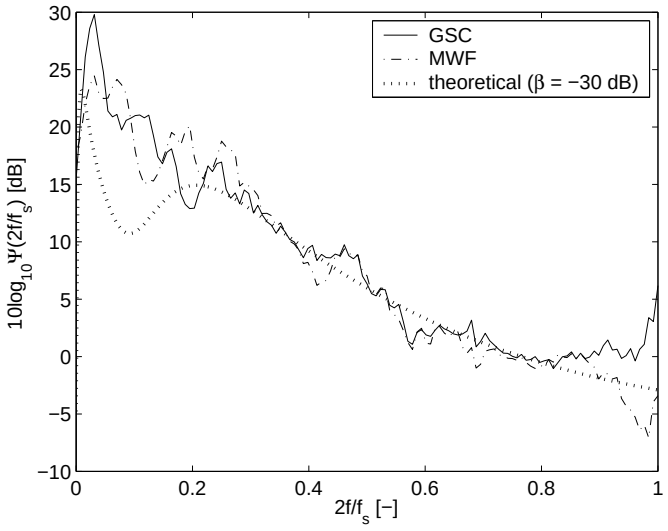


Fig. 11. Experimentally obtained noise sensitivity of the GSC and the MWF in a multiple noise-source scenario.

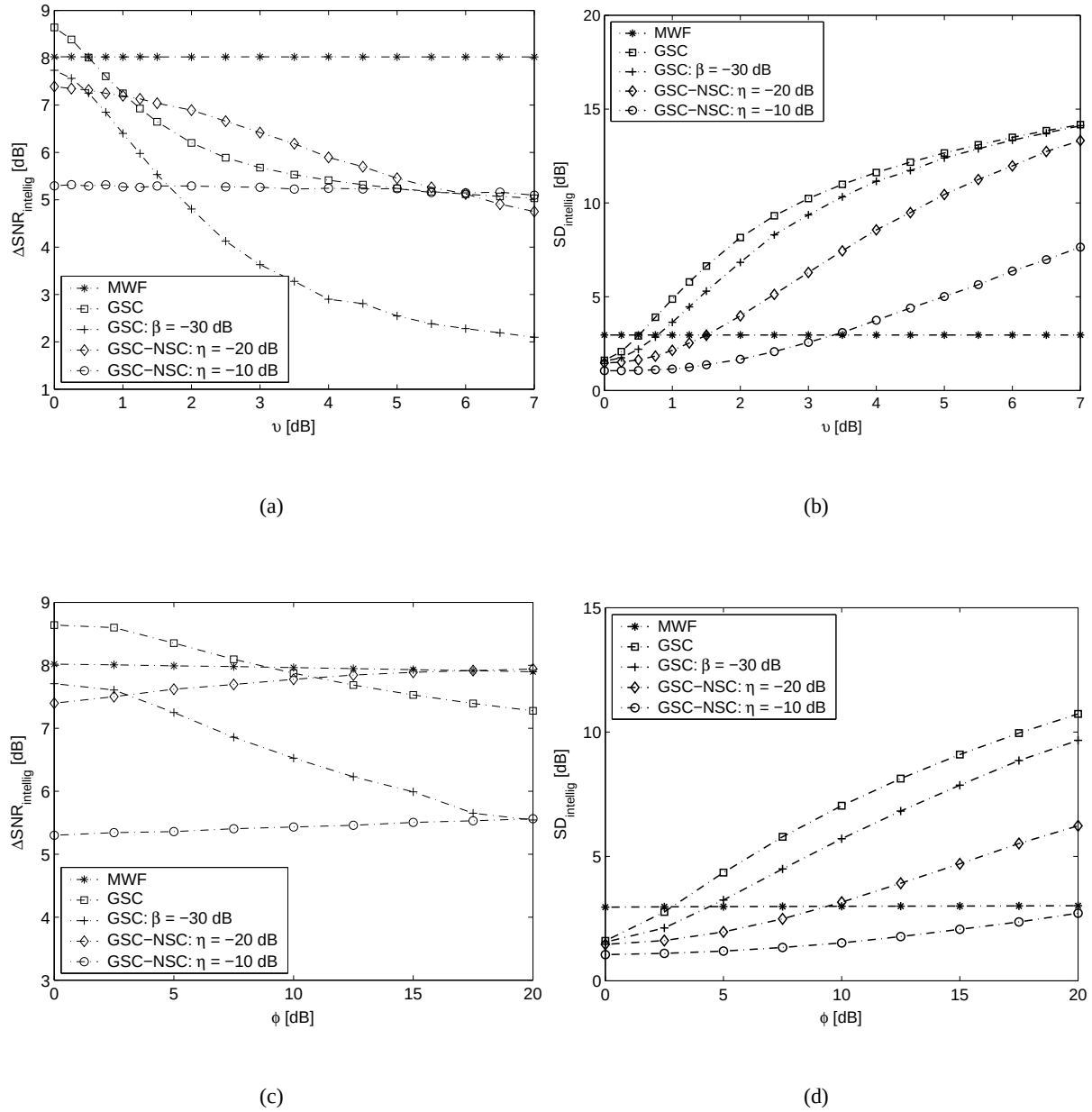


Fig. 12. Broadband performance measures ((a)-(c) $\Delta\text{SNR}_{\text{intellig}}$; (b)-(d) $\text{SD}_{\text{intellig}}$) of GSC, GSC-NSC and MWF in a multiple noise source scenario as a function of the amount of microphone mismatch; (a)-(b) Microphone gain mismatch υ (in dB); (c)-(d) Microphone phase mismatch ϕ (in degrees).

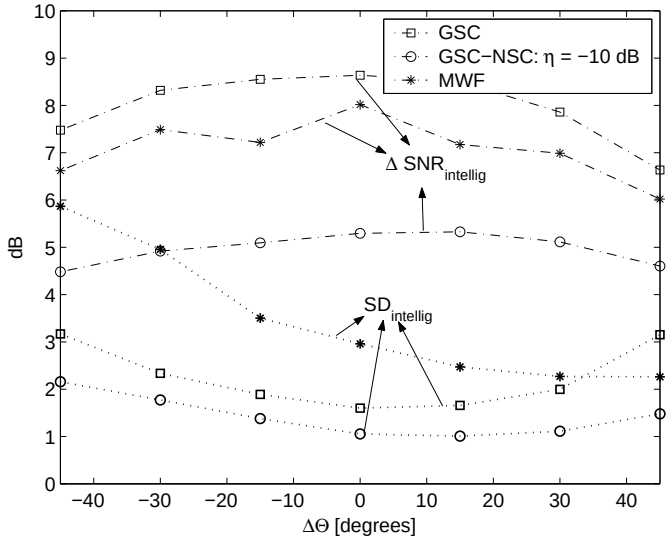


Fig. 13. Influence of a variation $\Delta\Theta$ (in degrees) in speaker location on the performance of the GSC, the GSC-NSC and the MWF.

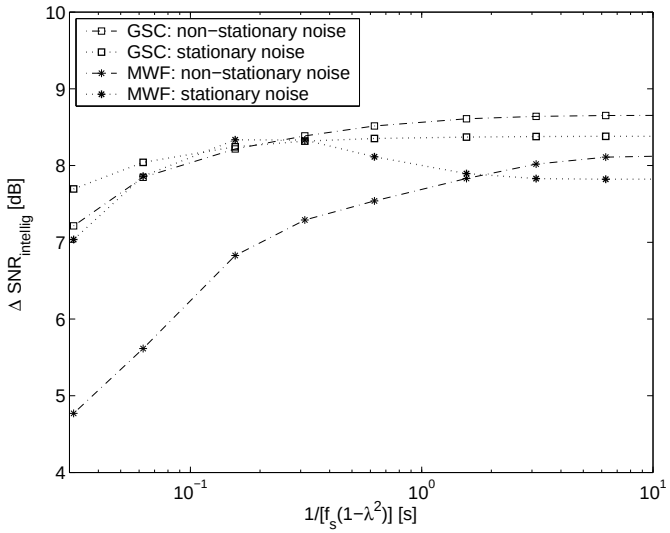


Fig. 14. Influence of forgetting factor $\lambda^2 = \lambda_{s+n}^2 = \lambda_n^2$ on the improvement $\Delta\text{SNR}_{\text{intellig}}$ of GSC and MWF for spectrally non-stationary and stationary noise sources, respectively.

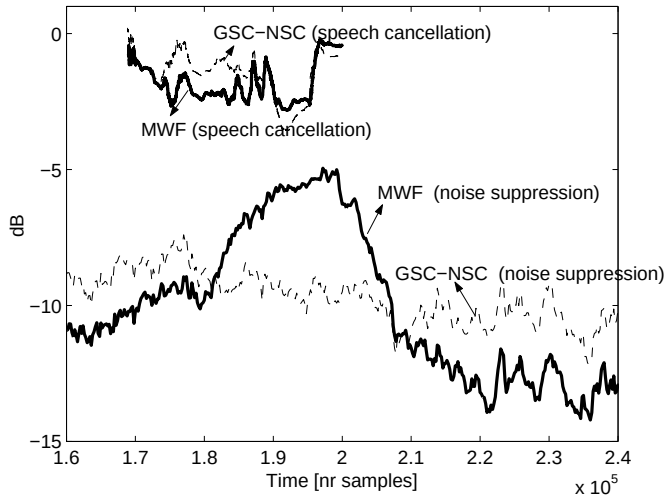


Fig. 15. Influence of change in noise scenario on the speech cancellation and noise suppression by the GSC-NSC (with $\eta \approx -20$ dB) and the MWF.