Katholieke Universiteit Leuven

Departement Elektrotechniek

ESAT-SISTA/TR 2002-81a

Robustness Analysis of Multichannel Wiener Filtering and Generalized Sidelobe Canceller for Hearing Aid Applications [1]

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

May 2002

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/sista/spriet/reports/02-81a.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*. Ann Spriet is a Research Assistant supported by the Fonds voor Wetenschappelijk Onderzoek (FWO) - Vlaanderen. This research work was carried out at the ESAT laboratory and Lab. Exp. ORL of the Katholieke Universiteit Leuven, in the framework of the Concerted Research Action GOA-MEFISTO-666 (Mathematical Engineering for Information and Communication Systems Technology) of the Flemish Government, IUAP P4-02 (1997-2001) 'Modeling, Identification, Simulation and Control of Complex Systems' and FWO Research Project nr. G.0233.1 ('Signal processing and automatic patient fitting for advanced auditory prostheses'). The scientific responsibility is assumed by its authors.

[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

ESAT/SCD-SISTA, K.U. Leuven

Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee

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

Kapucijnenvoer 33, B-3000 Leuven

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

Email: ann.spriet@esat.kuleuven.ac.be

Marc Moonen

ESAT/SCD-SISTA, K.U. Leuven

Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee

Tel.: +32 16 321060, Fax: +32 16 321970

Email: marc.moonen@esat.kuleuven.ac.be

Jan Wouters

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

Kapucijnenvoer 33, B-3000 Leuven

Tel.: +32 16 332342, Fax: +32 16 332335

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

Abstract

Multi-microphone noise reduction techniques for hearing aid applications go together with the use of small-sized arrays. Considerable noise reduction can be achieved with such arrays, but 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 recently developed multi-channel Wiener filtering techniques for small-sized arrays both analytically and experimentally. The theoretical analysis reveals that robustness issues are especially crucial in complicated noise scenarios and that microphone mismatch is particularly harmful to the GSC. To increase the robustness of the GSC to signal model errors in complicated noise scenarios, a constraint on the noise sensitivity is essential, which goes at the expense of less noise reduction. The multi-channel Wiener filtering techniques on the other hand, are not degraded by signal model errors, so that - compared to a robust GSC with noise sensitivity constraint - a better performance is obtained. The theoretical results (for infinite filter lengths) are verified by real data experiments.

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 and/or reverberation. 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 aid applications, 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 $f.d$, i.e. $f.d \ll \frac{c}{2}$ (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) and determines how much the internal noise in the hearing aid system, e.g. sensor noise, is amplified. 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 the design of *fixed beamformers* [1], [4], [5], the noise sensitivity should be kept small for two reasons: the amplification of the sensor noise may not be audible to the user during periods of silence or low signal level and the fixed design must be robust enough to processing errors and possible deviations from a priori assumptions such as microphone mismatch, misteer and microphone placement uncertainty [1]. The noise sensitivity can be constrained by e.g. inserting spatially uncorrelated noise in the microphone signals used to design the filters [1]. Also *adaptive beamforming algorithms* may require constraints to guarantee robustness. A widely studied, adaptive noise reduction algorithm for hearing aid applications is the *Generalized Sidelobe Canceller (GSC)* [2]-[18]. The GSC assumes the desired speaker location, as well as the microphone characteristics and positions to be known a priori, and reflections of the speech signal to be absent. In reality, these conditions are seldom completely fulfilled such that leakage of speech into the noise references occurs. To avoid cancellation of speech, the Adaptive Noise Canceller (ANC) in the GSC is adapted during periods of noise only [12] and constraints are applied to increase robustness [3], [13], [14], [17], [18]. In [3], [13], [14], [17] the noise sensitivity of the GSC is constrained by applying a quadratic inequality constraint to the filters of the Adaptive Noise Canceller (ANC). In [18], a similar constraint is imposed by injection of spatially uncorrelated noise. In [13], [14], the robustness of the GSC against misteer and uncertainties in microphone positions is additionally increased by incorporating multiple linear constraints. However, all these constraints go at the expense of less noise reduction.

Recently, *multi-channel Wiener filtering* techniques have been proposed that provide a Mean Square Error (MSE) estimate of the desired signal portion in one of the received microphone signals [19], [20], [21], [22]. In [19], [20], the implementation of the Wiener filter is based on a Generalized Singular Value Decomposition (GSVD) of an input data matrix and a noise data matrix. A cheaper alternative based on a QR Decomposition (QRD) has been proposed in [21]. A subband implementation [22], [23] results in improved intelligibility at a significantly lower cost compared to the fullband approach, making it suitable for future hearing aids. In contrast to the GSC, the multi-channel Wiener filtering does not make any a priori assumptions about the microphone characteristics, speaker location, reverberation, ... Simulations with a 5-microphone array with interspacing $d = 5$ 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 a higher noise reduction performance may be achieved.

In this paper, we analyse the robustness of the GSC with and without noise sensitivity constraint in

comparison with the multi-channel Wiener filter techniques both theoretically and experimentally for *small-sized endfire arrays*. Analytical calculations have drawn the attention of many researchers to predict the optimal (i.e. infinite filter length) noise reduction performance achievable by different single and multi-microphone noise reduction systems [2], [24]-[29]. In [2], [25], [26], [28], 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 is performed to evaluate a microphone array with postfiltering.

We develop a theoretical framework to assess the performance of the GSC with and without noise sensitivity constraint compared to the multi-channel Wiener filter techniques and to investigate their robustness to different types of model errors. In the simulations and experiments, we focus on small-sized arrays as used in hearing aid applications.

The paper is organized as follows. In Section II, the optimal noise reduction performance of the GSC is briefly reviewed and the formulas are extended to also quantify noise sensitivity and speech distortion. In addition, the optimal performance of GSC with noise sensivity constraint is derived. Section III reviews the concept of the multi-channel Wiener filtering techniques and derives formulas of its optimal performance. In Section IV, theoretical analysis of the noise sensitivity determines under which conditions (e.g. noise scenario, microphone configuration, ...) the noise sensitivity increases and hence robustness issues become crucial. The effects of different types of signal model errors on the optimal performance of the algorithms are investigated, i.e. microphone mismatch, microphone placement uncertainty and variations in speaker location. We will show that, especially in complicated noise scenarios and at low $f.d$, robustness issues are crucial and that microphone mismatch in particular is harmful to the GSC. As a consequence, a constraint on the noise sensitivity is essential, which goes at the expense of decreased noise reduction performance. The multi-channel Wiener filtering techniques on the other hand, are not degraded by signal model errors, so that - compared to a robust GSC with noise sensitivity constraint- more noise is reduced in complicated noise scenarios. In Section V, the theory is validated based on real recordings made in an office room with a 4-microphone endfire array mounted on a dummy head. The experimental results are found to be well predicted by the theoretical analysis.

Notation

Signals will generally be represented in the frequency domain. The microphone signals will be called $X_k(f)$, $k = 1, ..., M$ with M the number of microphones. The Power Spectral Density (PSD) of signal

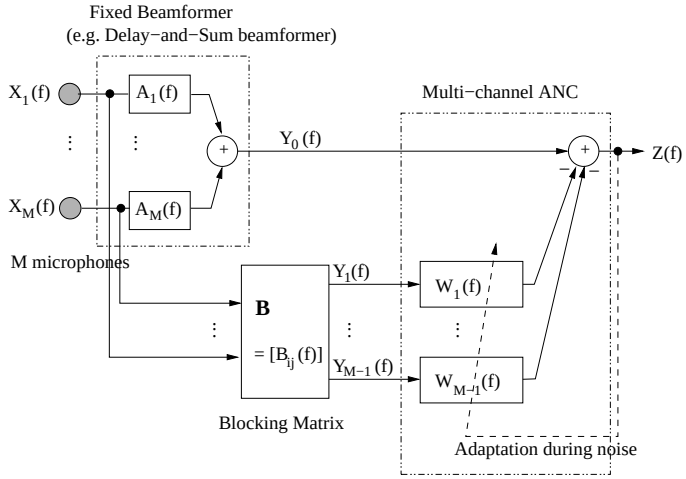


Fig. 1. Concept of the Generalized Sidelobe Canceller.

$X(f)$ is $P_X(f)$, the cross-PSD between signals $X(f)$ and $Y(f)$ is $P_{XY}(f)$. 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, ..., modelled as spatially white noise. We assume the PSD of the received speech, $P_X^s(f)$, the received noise, $P_X^n(f) = P_X^i(f) + P_X^e(f)$, the internal and external noise $P_X^i(f)$ and $P_X^e(f)$ and the received microphone signals $P_X(f) = P_X^s(f) + P_X^n(f)$ to be the same at each microphone.

The complex conjugate of a complex value B is denoted as B^*. The real and imaginary part of the complex value B will be denoted as $Re\{B\}$ and $Im\{B\}$, respectively. The superscript T denotes the transpose operator, the superscript H the conjugate transpose.

II. Generalized Sidelobe Cancellation

A. Principle and notation

The Generalized Sidelobe Canceller (GSC) [6], [7], [8], depicted in Figure 1, consists of a fixed beamformer $\mathbf{A}$ (e.g. delay-and-sum), a blocking matrix $\mathbf{B}$ and an Adaptive Noise Canceller (ANC) [30]. The fixed beamformer creates the speech reference $Y_0(f)$ 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 $(M-1) \times M$ blocking matrix $\mathbf{B}$ creates noise references $Y_i(f)$, $i = 1, \cdots M-1$, with M the number of microphones, 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 cancellation of the speech, the ANC is adapted during periods of noise only

[9], [12], [17].

In the sequel, we refer to the output of the ANC as $Z(f)$, the filter of the fixed beamformer applied to the i-th microphone is called $A_i(f)$. 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. The $(M-1) \times M$ blocking matrix $\mathbf{B}$ consists of the filters $B_{ij}(f)$, $i = 1, \cdots, M-1$, $j = 1, \cdots, M$. For e.g. $M = 4$, $\mathbf{B}$ is set to

$$\mathbf{B} = \begin{bmatrix} 1 & -e^{j2\pi f \frac{d_2 - d_1}{c}} & 0 & 0 \\ 1 & 0 & -e^{j2\pi f \frac{d_3 - d_1}{c}} & 0 \\ 1 & 0 & 0 & -e^{j2\pi f \frac{d_4 - d_1}{c}} \end{bmatrix}. \tag{1}$$

B. Theoretical performance

In [25], [28], [31], the theoretical (i.e. assuming infinite filter length) noise reduction $NR(f)$ of the GSC has been expressed in terms of the frequency domain, so-called *complex coherence* $\Gamma_{kl}(f)$ between the k-th and l-th microphone signals, $k = 1, ..., M$, $l = 1, ..., M$. In addition to the Power Transfer Function (PTF) of the noise signal $G^n(f) = \frac{P_Z^n(f)}{P_X^n(f)} = \frac{1}{NR(f)}$, also the PTF of the desired speech signal $G^s(f) = \frac{P_Z^s(f)}{P_X^s(f)}$ and the PTF of the internal noise $G^i(f) = \frac{P_Z^i(f)}{P_X^i(f)}$ can be expressed as a function of $\Gamma_{kl}(f)$. These results are reviewed in this section. In addition, the formulas for GSC with noise sensitivity constraint are derived. The PTFs from the microphones to the algorithm's output $G^s(f)$, $G^n(f)$, $G^i(f)$ will be used in Section IV. To evaluate the overall performance of the algorithm, broadband intelligibility weighted performance measures are computed based on the PTFs $G^s(f)$ and $G^n(f)$.

1) 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 $P_{X_k X_l}(f) = \mathcal{E}\{X_k(f) X_l^*(f)\}$, is defined as[1]:

$$\Gamma_{kl}(f) = \frac{P_{X_k X_l}}{\sqrt{P_{X_k} P_{X_l}}}. \tag{2}$$

The coherence function Γ_{kl} of a *localized sound source* at an angle θ with respect to the front direction in free and far field equals

$$\Gamma_{kl}(f) = \exp\left(-j2\pi f \cos\theta (d_k - d_l)/c\right) \tag{3}$$

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

$$\Gamma_{kl}(f) = \frac{\sin(2\pi f (d_k - d_l)/c)}{2\pi f (d_k - d_l)/c} \tag{4}$$

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

with $d_k - d_l$ the interspacing between microphone k and l [25]. 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 $\Gamma_{kl}(f)$ 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

$$\Gamma_{kl}(f) = \frac{\sum_{j=1}^{L} \alpha_j(f)\Gamma^j_{kl}(f)}{\sum_{j=1}^{L} \alpha_j(f)}, \tag{5}$$

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 [26]

$$\Gamma^n_{kl}(f) = \frac{\Gamma^e_{kl}(f)}{1 + \beta(f)} \qquad \text{for } k \neq l, \tag{6}$$

since $\Gamma^i_{kl} = \delta[k-l]$ and 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.

2) 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 now be expressed as a function of $\Gamma^s_{kl}(f)$ and $\Gamma^n_{kl}(f)$, the fixed beamformer $\mathbf{A}(f)$ and the blocking matrix $\mathbf{B}(f)$. The ANC filters $\mathbf{W}(f)$ are given by:

$$\mathbf{W}(f) = \begin{bmatrix} P^n_{Y_1Y_1} & P^n_{Y_2Y_1} & \cdots & P^n_{Y_{M-1}Y_1} \\ P^n_{Y_1Y_2} & P^n_{Y_2Y_2} & \cdots & P^n_{Y_{M-1}Y_2} \\ \vdots & \vdots & \ddots & \vdots \\ P^n_{Y_1Y_{M-1}} & P^n_{Y_2Y_{M-1}} & \cdots & P^n_{Y_{M-1}Y_{M-1}} \end{bmatrix}^{-1} \begin{bmatrix} P^n_{Y_0Y_1} \\ \vdots \\ P^n_{Y_0Y_{M-1}} \end{bmatrix}, \tag{7}$$

where $P^n_{Y_jY_i}$ is the cross-PSD of the noise contributions in the i-th and j-th noise reference and Y_0 is the speech reference [31]. $P^n_{Y_jY_i}$ and $P^n_{Y_0Y_i}$ can be expressed as a function of Γ^n_{kl}:

$$\begin{aligned} P^n_{Y_jY_i}(f) &= P^n_X(f) \sum_{k=1}^{M}\sum_{l=1}^{M} B_{jk}B^*_{il}\Gamma^n_{kl}(f) \\ P^n_{Y_0Y_i}(f) &= P^n_X(f) \sum_{k=1}^{M}\sum_{l=1}^{M} A_kB^*_{il}\Gamma^n_{kl}(f). \end{aligned} \tag{8}$$

The filters $W_i(f)$ thus only depend on the spatial characteristics of the noise and not on $P^n_X(f)$.

The PSD of the noise signal at the ANC output P^n_Z equals

$$\begin{aligned} P^n_Z(f) &= P^n_{Y_0} - 2\sum_{i=1}^{M-1} Re\{P^n_{Y_0Y_i}W^*_i\} \\ &\quad + \sum_{i=1}^{M-1}\sum_{j=1}^{M-1} W_iW^*_jP^n_{Y_iY_j} \end{aligned} \tag{9}$$

where $P_{Y_0}^n$ is the PSD at the output of the fixed beamformer, i.e.

$$P_{Y_0}^n(f) = P_X^n(f) \sum_{k=1}^{M} \sum_{l=1}^{M} A_k A_l^* \Gamma_{kl}^n. \tag{10}$$

Similar expressions can be found for the speech PSDs $P_{Y_0}^s$, P_Z^s and the internal noise PSDs $P_{Y_0}^i$, P_Z^i by replacing the superscripts n with s and i.

Using (7)-(10) and $\Gamma_{kl}^i = \delta[k-l]$, G^s, G^n and G^i can be expressed as a function of Γ_{kl}^s and Γ_{kl}^n:

$$\begin{aligned} G^{n,s}(f) &= \Bigg[P_{Y_0}^{n,s}(f) - 2 \sum_{i=1}^{M-1} Re\{P_{Y_0 Y_i}^{n,s}(f) W_i^*(f)\} \\ &\quad + \sum_{i=1}^{M-1} \sum_{j=1}^{M-1} W_i(f) W_j^*(f) P_{Y_i Y_j}^{n,s}(f) \Bigg] / P_X^{n,s}(f) \end{aligned} \tag{11}$$

$$\begin{aligned} G^i(f) &= \sum_{k=1}^{M} |A_k|^2 - 2 \sum_{i=1}^{M-1} Re\{W_i^*(f) \sum_{k=1}^{M} A_k B_{ik}^*\} \\ &\quad + \sum_{i=1}^{M-1} \sum_{j=1}^{M-1} W_i(f) W_j^*(f) \sum_{k=1}^{M} B_{ik} B_{jk}^*. \end{aligned} \tag{12}$$

3) PTFs $\tilde{G}^s(f)$, $\tilde{G}^n(f)$ and $\tilde{G}^i(f)$ of the GSC with noise sensitivity constraint : To increase the robustness of the GSC, a constraint on the noise sensitivity is often applied [3], [13], [17], [18]. In [3], it is shown that a constraint on the noise sensitivity corresponds to a quadratic inequality constraint on the filters $\mathbf{B}^H(f)\mathbf{W}^*(f)$, i.e.

$$\mathbf{W}^T(f)\mathbf{B}(f)\mathbf{B}^H(f)\mathbf{W}^*(f) \leq \zeta. \tag{13}$$

In [18], a similar constraint is imposed by inserting uncorrelated noise in the microphone signals 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

$$\Gamma_{kl}^n(f) + \eta(f)\delta[k-l], \tag{14}$$

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

4) 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, such as the improvement in intelligibil-

ity weighted Signal-to-Noise Ratio (SNR) [32], which is defined as

$$\Delta\text{SNR}_{\text{intellig}} = \sum_i I_i(\text{SNR}_{i,\text{out}} - \text{SNR}_{i,\text{in}}), \tag{15}$$

where the band importance function I_i expresses the importance of the i-th one-third octave band with center frequency f_i^c 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_i^c and the values I_i are defined in [33]; 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

$$\begin{aligned}\text{SNR}_{i,\text{out}} &= \\ &10\log_{10}\left(\int_{2^{-\frac{1}{6}}f_i^c}^{2^{\frac{1}{6}}f_i^c} P_X^s(f)G^s(f)df \Big/ \int_{2^{-\frac{1}{6}}f_i^c}^{2^{\frac{1}{6}}f_i^c} P_X^n(f)G^n(f)df\right) \qquad (16)\\ \text{SNR}_{i,\text{in}} &= 10\log_{10}\left(\int_{2^{-\frac{1}{6}}f_i^c}^{2^{\frac{1}{6}}f_i^c} P_X^s(f)df \Big/ \int_{2^{-\frac{1}{6}}f_i^c}^{2^{\frac{1}{6}}f_i^c} P_X^n(f)df\right). \qquad (17)\end{aligned}$$

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

$$\text{SD}_{\text{intellig}} = \sum_i I_i \text{SD}_i, \tag{18}$$

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

$$\text{SD}_i = \int_{2^{-\frac{1}{6}}f_i^c}^{2^{\frac{1}{6}}f_i^c} |10\log_{10} G^s(f)|\, df \Big/ \left[\left(2^{\frac{1}{6}} - 2^{-\frac{1}{6}}\right) f_i^c\right]. \tag{19}$$

Summarizing, given the GSC parameters $\mathbf{A}(f)$ and $\mathbf{B}(f)$, the coherence functions Γ_{kl}^s, Γ_{kl}^n and the PSD $P_X^s(f)$, $P_X^n(f)$ and the intelligibility weighted performance measures can be obtained.

III. Multi-channel Wiener filtering

A. Concept

The multi-channel Wiener filter $\mathbf{w} \in \mathcal{C}^{ML\times 1}$, described in Figure 2, provides a Mean Square Error (MSE) estimate of a delayed version of the (unknown) speech signal $s_i[k-\Delta]$ at the i-th (e.g. first) microphone, which is computed as $y_0[k] = \hat{s}_i[k-\Delta] = \mathbf{w}^T\mathbf{x}[k]$, i.e. the sum of the M filtered microphone signals, where

$$\mathbf{w}^* = \mathcal{E}\{\mathbf{x}[k]\mathbf{x}^H[k]\}^{-1}\mathcal{E}\{\mathbf{x}[k]s_i^*[k-\Delta]\}, \tag{20}$$

$$\mathbf{x}[k] = [x_1[k]\cdots x_1[k-L+1]\cdots x_M[k]\cdots x_M[k-L+1]]^T \tag{21}$$

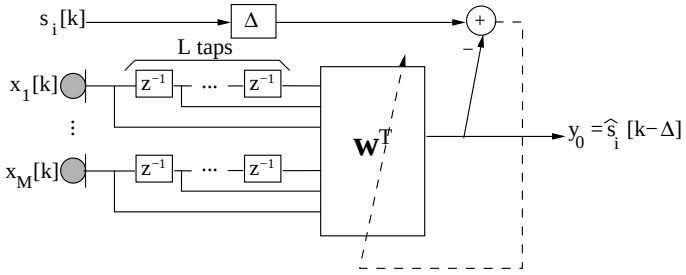


Fig. 2. Concept of the multi-channel Wiener filter.

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

Assuming that the speech and noise signals are uncorrelated, (20) becomes:

$$\mathbf{w}^* = \mathcal{E}\{\mathbf{x}[k]\mathbf{x}^H[k]\}^{-1} \left(\mathcal{E}\{\mathbf{x}[k]x_i^*[k-\Delta]\} - \mathcal{E}\{\mathbf{n}[k]n_i^*[k-\Delta]\}\right). \tag{22}$$

The second order statistics of the noise signal are assumed to be sufficiently stationary such that they can be estimated during periods of noise only. A robust speech detection is thus (again) needed. In [19], [20], the Wiener filter is approximated at time instant k by means of a Generalized Singular Value Decomposition (GSVD) of an input and noise data matrix. A cheaper alternative based on a QR Decomposition (QRD) has been proposed in [21]. In [22], a subband implementation of the GSVD based algorithm has been developed to increase intelligibility and reduce complexity. In contrast to the GSC, the multi-channel Wiener filter does not make any a priori assumptions about microphone characteristics and positions, speaker location, reflections and reverberation. It is uniquely based on estimates of the second order statistics of the speech and noise signal. In the sequel, we will assume, without loss of generality, that the speech contribution as received at the first microphone signal is estimated, i.e. $s_1[k-\Delta]$.

B. Theoretical analysis

The theoretically optimal performance of the multi-channel Wiener filter, i.e. the PTF of noise, speech and internal noise as well as the intelligibility weighted performance measures, can again be expressed as a function of Γ_{kl}, Γ_{kl}^n and Γ_{kl}^s, as will be detailed in this section.

If

$$\mathbf{W}(f) = \left[\begin{array}{cccc} W_1(f) & W_2(f) & \cdots & W_M(f) \end{array}\right]^T \tag{23}$$

with $W_i(f)$ filtering the i-th microphone signal, $\mathbf{W}(f)$ can be written as[2]:

$$\mathbf{W}(f) = \begin{bmatrix} P_X & P_{X_2X_1} & \cdots & P_{X_MX_1} \\ P_{X_1X_2} & P_X & & P_{X_MX_2} \\ \vdots & & \ddots & \vdots \\ P_{X_1X_M} & \cdots & \cdots & P_X \end{bmatrix}^{-1} \begin{bmatrix} P_X^s \\ P_{X_1X_2}^s \\ \vdots \\ P_{X_1X_M}^s \end{bmatrix}, \tag{24}$$

where $P_{X_kX_l}^s(f)$ is estimated as

$$P_{X_kX_l}^s(f) = P_{X_kX_l}(f) - P_{X_kX_l}^n(f), \tag{25}$$

with $P_{X_kX_l}(f)$ estimated during periods of *speech + noise* and $P_{X_kX_l}^n(f)$ estimated during periods of *noise only*. Using definition (2) of the complex coherence function Γ_{kl} and

$$P_X(f) = P_X^s(f) + P_X^e(f) + P_X^i(f), \tag{26}$$

formula (24) can be rewritten as:

$$\begin{aligned} \mathbf{W}(f) &= \frac{P_X^s(f)}{P_X^s(f) + P_X^e(f) + P_X^i(f)} \times \\ &\begin{bmatrix} 1 & \Gamma_{21} & \cdots & \Gamma_{M1} \\ \Gamma_{12} & 1 & & \Gamma_{M2} \\ \vdots & & \ddots & \vdots \\ \Gamma_{1M} & \cdots & \cdots & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \Gamma_{12}^s \\ \vdots \\ \Gamma_{1M}^s \end{bmatrix} \end{aligned} \tag{27}$$

where, using (5):

$$\Gamma_{kl}(f) = \frac{\alpha(f)\Gamma_{kl}^e + \gamma(f)\Gamma_{kl}^i + \Gamma_{kl}^s}{1 + \alpha(f) + \gamma(f)}; \tag{28}$$

$$\alpha(f) = \frac{P_X^e(f)}{P_X^s(f)}; \tag{29}$$

$$\gamma(f) = \frac{P_X^i(f)}{P_X^s(f)}; \tag{30}$$

In contrast to the GSC, the filter $\mathbf{W}(f)$ not only depends on the spatial characteristics of the external noise $\Gamma_{kl}^e(f)$ and the internal-to-external noise ratio $\beta(f)$, but also on the ratio $\alpha(f)$ and the spatial characteristics of the speech signal $\Gamma_{kl}^s(f)$.

Note that the ratio $\frac{P_X^s(f)}{P_X^s(f)+P_X^e(f)+P_X^i(f)}$ in (27) corresponds to a *spectral* Wiener filter, which results in additional performance if the spectra of the desired and undesired signal differ. Hence, (27) combines spatial and spectral filtering.

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

The PSD of the noise $Y_0^n(f)$ at the output equals

$$P_{Y_0}^n(f) = P_X^n \left(\sum_{k=1}^{M} |W_k|^2 + \sum_{k=1}^{M} \sum_{l=k+1}^{M} 2.Re\{W_k W_l^* \Gamma_{kl}^n\} \right) \tag{31}$$

and is determined by Γ_{kl}^s, Γ_{kl}^n, $\gamma(f)$ and $\alpha(f)$. The expressions for the PSD $P_{Y_0}^s(f)$ and $P_{Y_0}^i(f)$ can be found by substituting the superscripts n by s and i in (31). Using (31) and $\Gamma_{kl}^i = \delta[k-l]$, the PTFs $G^n(f)$, $G^s(f)$ and $G^i(f)$ can be easily derived:

$$G^{n,s}(f) = \sum_{k=1}^{M} |W_k|^2 + \sum_{k=1}^{M} \sum_{l=k+1}^{M} 2.Re\{W_k W_l^* \Gamma_{kl}^{n,s}\} \tag{32}$$

$$G^i(f) = \sum_{k=1}^{M} |W_k|^2. \tag{33}$$

IV. Theoretical robustness analysis

Based on the formulas in Section II and Section III, the performance of the GSC and the multi-channel Wiener filtering can be evaluated analytically. A robust noise reduction algorithm should be insensitive to small signal model errors such as microphone mismatch. In contrast to the GSC, the multi-channel Wiener filtering technique does not require a priori information about the signal model and is thus expected to be more robust. In this Section, the robustness of the GSC and the multi-channel Wiener filter against different types of model errors is investigated for small-sized arrays. To quantify the sensitivity to errors in the assumed signal model, the *noise sensitivity* is often used, which is defined as the ratio of the spatially white noise gain to the gain of the desired signal [1], [2], [3]. In Section IV-A, we study the influence of the noise scenario and the microphone configuration on the noise sensitivity of the GSC and the multi-channel Wiener filter, based on the formulas in Section II and Section III. This allows us to determine under which conditions robustness issues are expected to become crucial. Section IV-B analyses the effect of a high noise sensitivity on the robustness of the GSC and the multi-channel Wiener filter against different types of model errors such as microphone sensitivity mismatch, deviations in microphone positions and speaker location, ... The effect of constraint on the noise sensitivity of the GSC is studied too. The results are illustrated for a linear, uniform 4-microphone endfire array with microphone interspacing $d = 2$ cm.

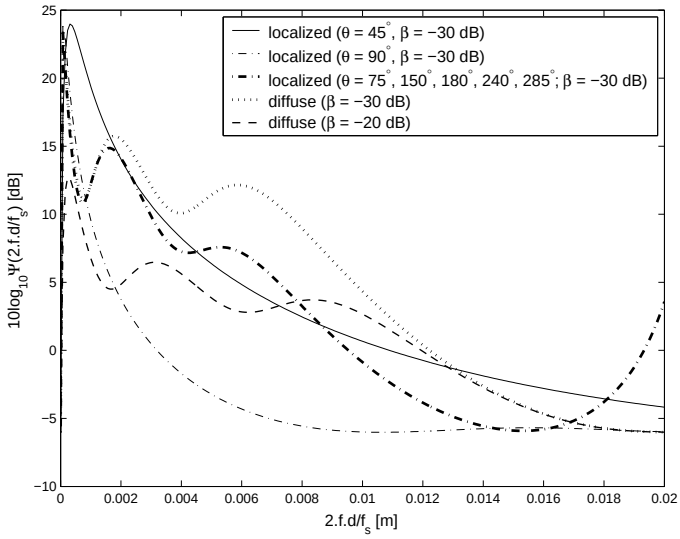


Fig. 3. Influence of the noise scenario on the noise sensitivity $\Psi(f)$ (in dB).

A. Analysis of the noise sensitivity

Using (11), (31) and (33), the noise sensitivity $\Psi(f)$ of the algorithms that are described in Section II and Section III, can be derived as:

$$\boxed{\Psi(f) = \frac{G^i(f)}{G^s(f)}} \tag{34}$$

In the absence of speech distortion, $G^s(f) = 1$.

Equations (11) and (33) for $G^i(f)$ show that, for both algorithms, a high noise sensitivity $\Psi(f)$ corresponds to large filter amplitudes $|W_i(f)|$ w.r.t. $G^s(f)$. 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. For both algorithms, $\Psi(f)$ depends on the coherence function $\Gamma^n_{kl}(f)$ of the noise field and hence, on the noise scenario and the microphone configuration. For a free, far field speech signal at an angle $\theta = 0^\circ$, the noise sensitivity of the GSC and the multi-channel Wiener filtering are found to coincide, i.e.

$$\boxed{\Psi_{\text{GSC}}(f) = \Psi_{\text{WF}}(f).} \tag{35}$$

Therefore, the subscripts WF and GSC will be omitted in Figure 3 and Figure 4.

Figure 3 investigates the influence of the noise scenario on the noise sensitivity $\Psi(f)$. The sensitivity $\Psi(f)$ (in dB) of the GSC is plotted as a function of $\frac{2f}{f_s}d$ (in meters) for a linear, uniform 4-microphone endfire array, with sampling frequency $f_s = 16$ kHz. Five different noise scenarios are shown: a localized

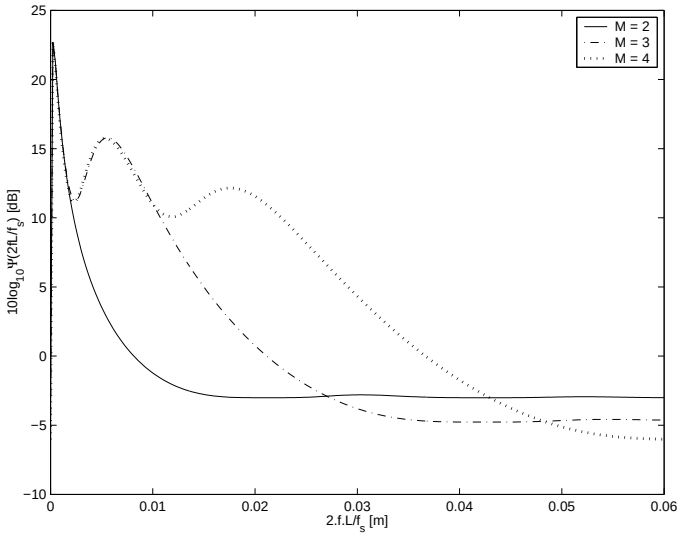


Fig. 4. Influence of the number of microphones M on the noise sensitivity $\Psi(f)$ (in dB) of the GSC and the multi-channel Wiener filter for a diffuse noise field and $\beta(f) = -30$ dB.

single noise source at 45° and 90° w.r.t. 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 [26]. Figure 3 shows that the noise sensitivity $\Psi(f)$ grows with decreasing $f.d$, 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 $\Psi(f)$ of the GSC as a function of $\frac{2f}{f_s}L$ (in meters) for a diffuse noise field and $\beta(f) = -30$ dB. 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. At very low $\frac{2f}{f_s}L$, the noise sensitivity reaches a maximum and does not further increase by adding more microphones. The external noise reduction and maximum internal noise amplification will be lower for high internal-to-external ratio $\beta(f)$. 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

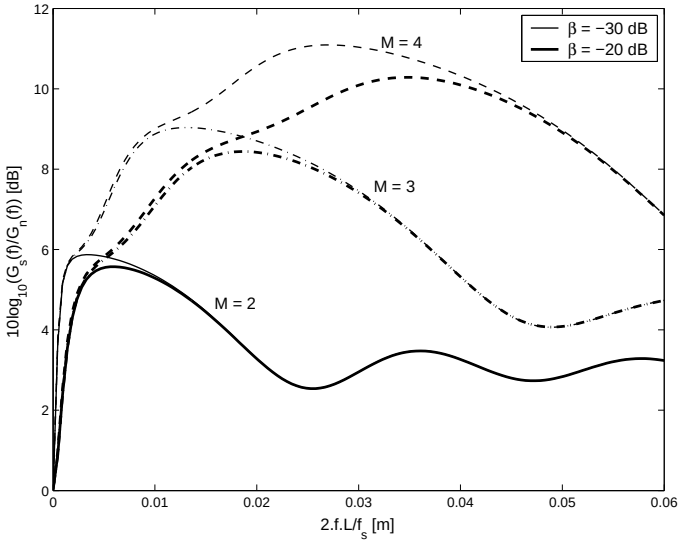


Fig. 5. Contribution of additional microphones to the improvement in signal-to-noise ratio $\frac{G^s(f)}{G^n(f)}$ (in dB) for the GSC and the multi-channel Wiener filtering in a diffuse noise field with $\beta(f) = -20$ dB and $\beta(f) = -30$ dB respectively.

noise ratio $\beta(f)$ [2] and by the noise sensitivity level $\Psi(f)$ tolerated by the noise reduction algorithm (see Section IV-B).

Figure 5 plots the improvement in signal-to-noise ratio, calculated as $\frac{G^s(f)}{G^n(f)}$ (in dB), as a function of $\frac{2f}{f_s}L$ (in meters) for a diffuse noise field and $\beta(f) = -20$ dB and -30 dB respectively. For a free and far field speech source at $\theta = 0^\circ$, both algorithms result in the same optimal improvement $\frac{G^s(f)}{G^n(f)}$. At low $\frac{2f}{f_s}L$ and for high $\beta(f)$, no improvement is gained by augmenting the number of microphones. For an array with a length $L = 3$ cm, which is similar to the dimensions of a commercial Behind-The-Ear (BTE) hearing aid, no significant 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 analysis

Section IV-A shows that for small-sized arrays, complex noise scenarios and an increasing number of microphones result in a high noise sensitivity. Hence, robustness issues are expected to become crucial under these conditions. In this section, we evaluate the effect of different types of model errors, i.e. microphone mismatch, robustness to microphone placement uncertainty and robustness to variations in speaker location, on the optimal performance of the GSC and multi-channel Wiener filtering algorithm. Since the GSC - in contrast to the multi-channel Wiener filtering - relies on a priori information about the

signal model, a constraint on its noise sensitivity may be needed at the expense of less noise reduction. The performance of the GSC with a noise sensitivity constraint will also be studied. 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 uniform 4-microphone endfire array and a diffuse noise source scenario with $\beta(f) = -30$ dB. Both the influence on the PTFs of speech and noise as well as the influence on the intelligibility weighted performance measures 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 [34] 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 [35], gain and phase differences between microphone characteristics of up to 6 dB and 10°, respectively, have been reported. The first, 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 [35]. 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_{kl}^s, Γ_{kl}^n and Γ_{kl}^i. Suppose that the k-th microphone has a gain mismatch $\Delta\Upsilon(f)$ and a phase mismatch $\Delta\Phi(f)$ (in degrees) at frequency f w.r.t. microphone l. The matched microphones coherence function $\Gamma_{kl}(f)$ should then be replaced by $\tilde{\Gamma}_{kl}(f)$

$$\tilde{\Gamma}_{kl}(f) = \begin{cases} \Delta\Upsilon(f)\exp(j\Delta\Phi(f)\frac{\pi}{180})\Gamma_{kl}(f) & \text{for } k \neq l; \\ |\Delta\Upsilon(f)|^2\Gamma_{kl}(f) & \text{for } k = l. \end{cases} \tag{36}$$

Note that in case of the GSC, the gain and phase mismatch could equivalently be incorporated by modifying $A_i(f)$ and $B_{ij}(f)$, $i = 1, \cdots, M-1$, $j = 1, \cdots, M$.

Figure 6 plots the PTF of the speech and noise signal of the **GSC** when the second and third microphone have a gain mismatch $\Delta\Upsilon$ of υ dB and $-\upsilon$ dB, respectively, with $\upsilon = 1$ dB or a phase mismatch $\Delta\Phi$ (in degrees) of ϕ and $-\phi$, respectively, with $\phi = 5^\circ$ w.r.t. the first microphone. The desired signal seems to be 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 signal, internal noise may become audible. The PTFs of speech and noise signal of the multi-channel Wiener filter (for $\alpha(f) = 0$ dB) are also shown. The multi-channel Wiener filter is independent of a deviation in microphone gain and microphone phase. The independency of microphone gain mismatch has previously been shown in [19]. Note that the **multi-channel Wiener filter** attenuates the speech at frequencies where the output SNR is small, which is typical of a Wiener filter, i.e. in this

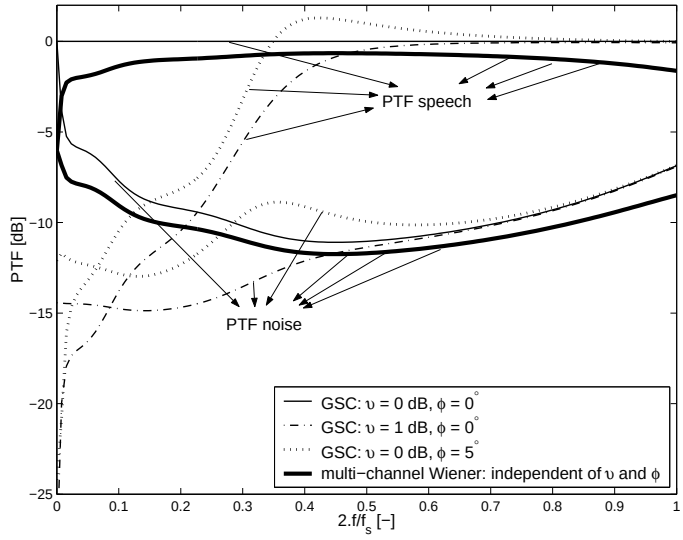


Fig. 6. Effect of microphone gain mismatch υ and phase mismatch ϕ on the PTF of speech and noise of GSC and multi-channel Wiener filtering for a diffuse noise field and $\beta(f) = -30$ dB.

example at low and high f. This attenuation will be high at f where the (average) input SNR $\frac{1}{\alpha(f)}$ is low (see also (27)) and at f where the amount of spatial noise reduction is limited (i.e. e.g. at low $f.d$ in complicated noise scenarios).

In Figure 7(a) to Figure 7(d), the degradation of the broadband performance $\Delta\text{SNR}_{\text{intellig}}$ (Figure 7(a)-(c)) and $\text{SD}_{\text{intellig}}$ (Figure 7(b)-(d)) due to gain and phase mismatch is shown as a function of the amount of mismatch υ and ϕ, respectively. The performance of a more robust GSC with noise sensitivity constraint, described in Section II, 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 mismatch in microphone sensitivity leads to an unacceptable degradation of the GSC performance in complicated noise scenarios. Constraining the noise sensitivity $\Psi(f)$, results in a higher robustness at the expense of noise reduction. The performance of the **multi-channel Wiener filtering** 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 multi-channel Wiener filter outperforms the robust GSC with noise sensitivity constraint when $\Psi(f)$ is large, i.e. in complicated noise scenarios and at low $f.d$. 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. The noise reduction algorithm should be robust to small deviations

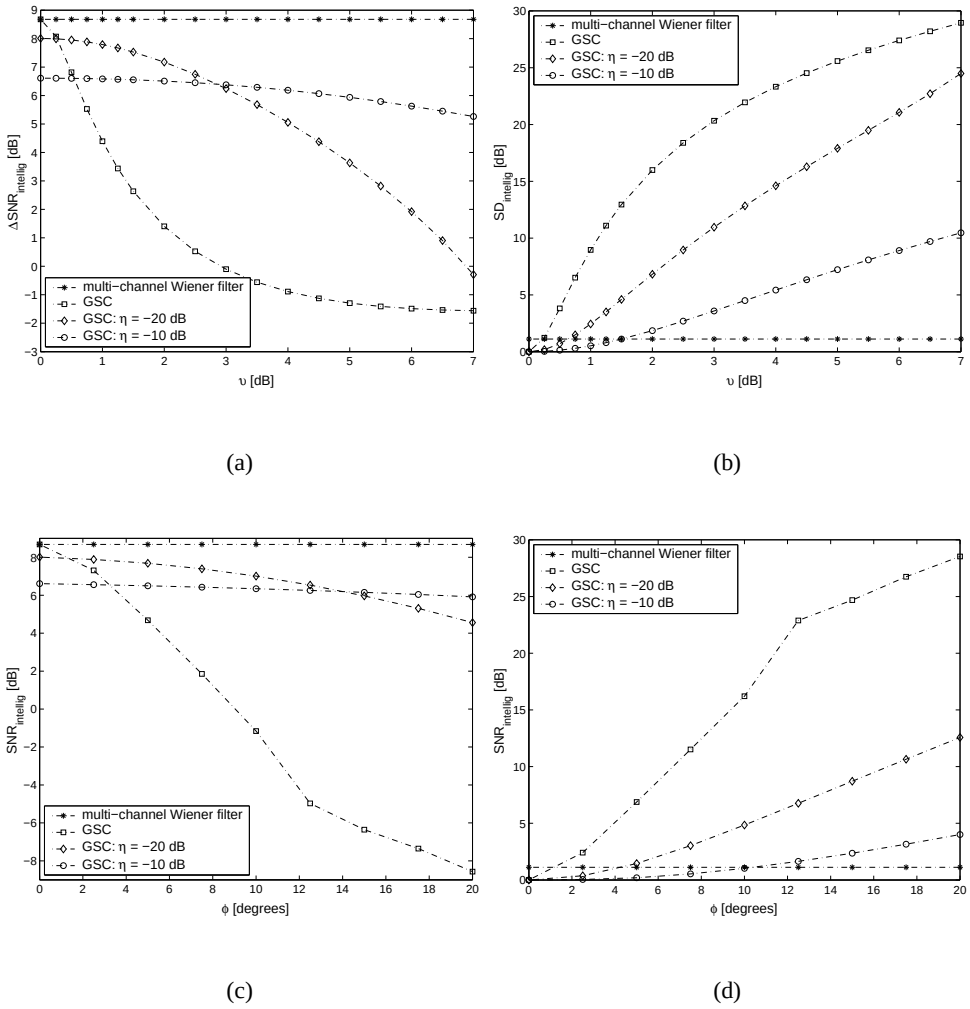


Fig. 7. Performance ((a)-(c) $\Delta\text{SNR}_{\text{intellig}}$; (b)-(d) $\text{SD}_{\text{intellig}}$) of multi-channel Wiener filtering, GSC and GSC with noise sensivity constraint in a diffuse noise field as a function of microphone sensitivity mismatch; (a)-(b) Microphone gain mismatch υ (in dB); (c)-(d) Microphone phase mismatch ϕ (in degrees).

in microphone positions. A change Δd (in meters) in the microphone position of the k-th microphone w.r.t 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 8 depicts the PTF of speech and noise for both algorithms when the 2nd microphone is moved by $\Delta d = 0.001$ m and $\Delta d = 0.002$ m, corresponding to a deviation in microphone position w.r.t. d_k of 5 % and 10 % 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. A change by Δd in (3) corresponds to a frequency dependent phase change in the k-th microphone frequency characteristic

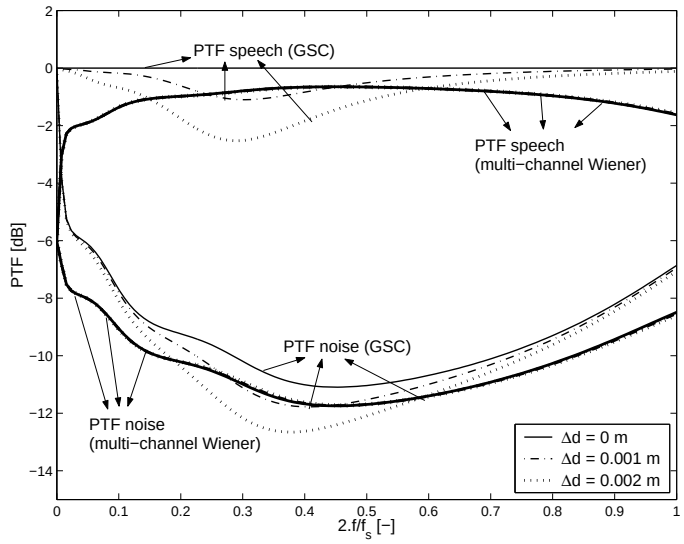


Fig. 8. Effect of a deviation Δd (in meters) in the position d_k of the 2nd microphone on the PTF of speech and noise for the GSC and multi-channel Wiener filter.

for a localized sound source at angle θ of

$$2\pi f \frac{\cos\theta}{c}\Delta d \tag{37}$$

(in rad). 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. This explains the limited effect of a change in microphone position. In addition, the microphones are quite accurately positioned in a hearing aid so that microphone placement uncertainty is expected to be small. Note that in applications where the uncertainties Δd are larger, a noise sensitivity constraint on the GSC may be required. In contrast to Section IV-B.1, the largest effect of a change in microphone position is not visible at the lowest frequencies. On one hand, the impact of phase mismatch is largest for high $\Psi(f)$ and hence at the lowest frequencies, on the other hand the amount of phase mismatch (37) grows with increasing frequency. Note that the impact still increases with increasing $\Psi(f)$. Figure 8 confirms that the PTF of speech and noise for the multi-channel Wiener filtering technique are not modified when the microphones are slightly moved.

3) Robustness to variations in speaker location: In addition to be 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 $\Delta\Theta$ w.r.t. the front. The effect of a change in speaker location can easily be studied analytically by incorporating the angle $\Delta\Theta$ w.r.t. 0° in $\Gamma^s_{kl}(f)$, i.e.

$$\Gamma^s_{kl}(f) = \exp\left(j2\pi f \cos(\Delta\Theta)\frac{d_k - d_l}{c}\right). \tag{38}$$

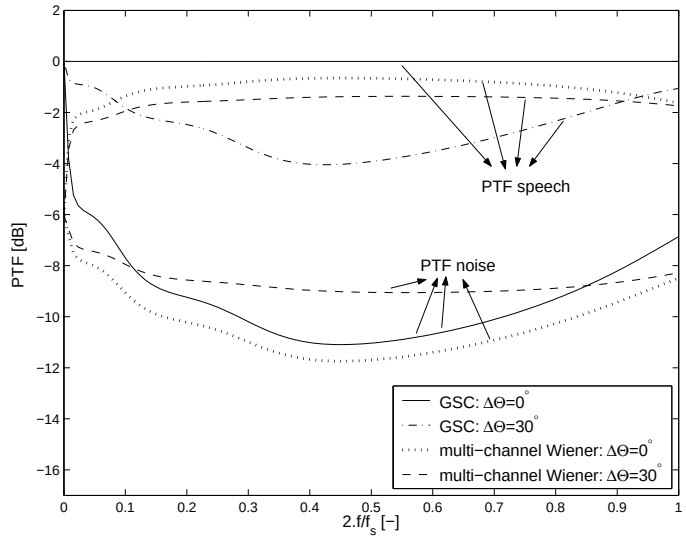


Fig. 9. Influence of a variation in speaker location $\Delta\Theta$ (in degrees) on the PTF of the noise (the solid and dash-dotted line coincide) and desired speech signal of the GSC and the multi-channel Wiener filtering.

Figure 9 depicts the PTF of speech and noise of the GSC and the multi-channel Wiener filtering for a variation in speaker location $\Delta\Theta$ of 30° in the presence of a diffuse noise field with $\beta = -30$ dB and $\alpha = 0$ dB. Since the ANC filters of the GSC are completely determined by $\Gamma_{kl}^{n}(f)$ and do not depend on $\Gamma_{kl}^{s}(f)$, the PTFs of the noise coincide. Its effect is less detrimental than microphone mismatch. In case of the **GSC**, speech distortion is observed over the full frequency range. In addition to $\Psi(f)$, the directivity pattern of the noise references determines the amount of speech distortion. The zero of the noise reference towards the front direction is more narrow at high f than at low f, hence the distortion. At low f, where $\Psi(f)$ is high, the zeros of the noise references are rather broad due to the small number of microphones that are used to create them and the small microphone interspacings. Hence, for larger arrays, the effect may be worse such that additional linear constraints should be incorporated in the design of the blocking matrix $\mathbf{B}$ [14]. The **multi-channel Wiener filter** also attenuates speech arriving from $\Delta\Theta = 30^\circ$ more than speech arriving from $\Delta\Theta = 0^\circ$. The multi-channel Wiener filter minimizes the MSE and is thus a compromise between noise reduction and speech distortion. It is more difficult to suppress diffuse noise with an endfire array if the speaker is positioned at 30° w.r.t. front direction, hence the larger amount of speech attenuation.

Figure 10 gives the effect on the broadband performance measures $\Delta\text{SNR}_{\text{intellig}}$ [dB] and $\text{SD}_{\text{intellig}}$ [dB] of speaker variations up to 45° for $\alpha(f) = 0$ dB and $\alpha(f) = 3$ dB, respectively. The improvement $\Delta\text{SNR}_{\text{intellig}}$ of an endfire array in a diffuse noise field decreases with increasing $\Delta\Theta$. For angles up

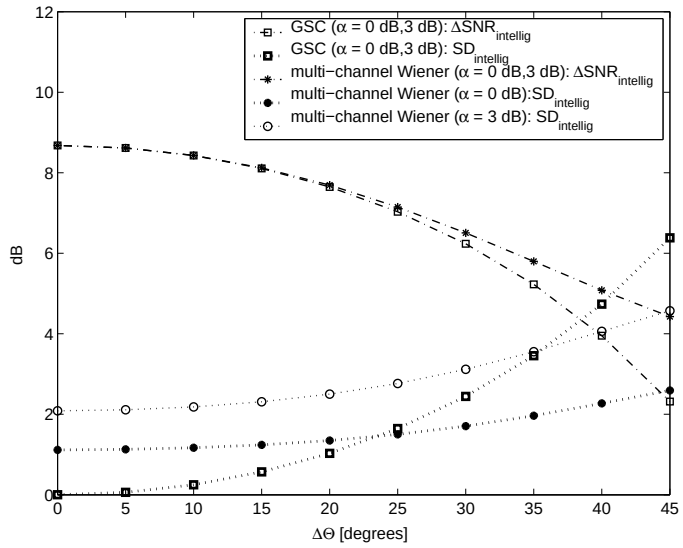


Fig. 10. Influence of a variation in speaker location on the broadband performance of the GSC and the multi-channel Wiener filtering.

to 30°, the GSC and the multi-channel Wiener filter offer about the same improvement in $\text{SNR}_{\text{intellig}}$. The amount of speech distortion $\text{SD}_{\text{intellig}}$ caused by the Wiener filter grows with increasing $\Delta\Theta$ and increasing $\alpha(f)$. The speech attenuation of the Wiener filter increases with decreasing output SNR: a low input SNR and more complicated noise scenarios both result in a low output SNR. In free field, the enhancement is the same for positive and negative $\Delta\Theta$. When mounted on the head, the input signal-to-noise ratio 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 mismatch when $\Psi(f)$ is high, i.e. at low $f.d$ 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 not found to be that detrimental for small-sized arrays. To guarantee performance in the presence of microphone mismatch, a constraint on the noise sensitivity should be imposed at the expense of noise reduction. The multi-channel Wiener filtering 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 in complicated noise scenarios and at low $\beta(f)$.

V. VALIDATION THROUGH EXPERIMENTAL RESULTS

The theoretical robustness analysis, presented in Section IV, holds for a localized speech source and localized or diffuse noise sources in the far and free field and assumes infinite filter lengths. In this section, we verify the major conclusions of Section IV based on real recordings in a reverberant room and when finite filters are used. The results are illustrated for a diffuse noise field. A linear endfire array with 4 omnidirectional microphones (Knowles EM-4368) with interspacing $d = 0.02$ 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. To get an idea of the average performance based on directivity only, stationary desired and noise signals are used with the same, average long-term power spectral density. The signals can be found on [34]. The desired and noise signal are uncorrelated and have a level of about 70 dB SPL at the center of the head. The input SNR at the first microphone equals 0 dB. The total duration of the input signal is 10 seconds. During the first 5 seconds only noise is present, during the last 5 seconds the desired and noise signal are both present. The subband implementation [22], [23] of the multi-channel Wiener filter is used. The fixed stage of the GSC has been designed for the 4-microphone array mounted on the dummy head so that - for an anechoic signal coming from $0°$- the energy of the noise references is minimized. The ANC filters are implemented using the Wiener formula to exclude the influence of a potentially slow convergence rate on the overall performance. The filters are sufficiently long so that a further increase in filter length does not result in a significant increase in performance. To improve intelligibility, the microphone signals are pre-emphasized in the GSC prior to processing [36]. At the GSC output, de-emphasis is performed. In the subband GSVD algorithm, the subband implementation intrinsically performs a pre-whitening operation [22], [23]. In the sequel, the robustness to two type of model errors will be investigated, i.e. microphone mismatch and variation in speaker location.

A. Robustness to microphone sensitivity mismatch

1) Influence on PTFs of speech and noise: Figure 11 shows the PTF of the speech and noise signal when the second and third microphone have a gain mismatch of υ dB and $-\upsilon$ dB, respectively, with $\upsilon = 1$ dB or a phase mismatch of ϕ and $-\phi$, respectively, with $\phi = 5°$ w.r.t. the first microphone. Even in the absence of deviations, the GSC distorts the speech at low $f.d$: a small model error in the assumptions of GSC (e.g. presence of reverberation) is already detrimental to its performance, due to the high noise level (hence small $\beta(f)$) and the complicated noise scenario. An additional constraint

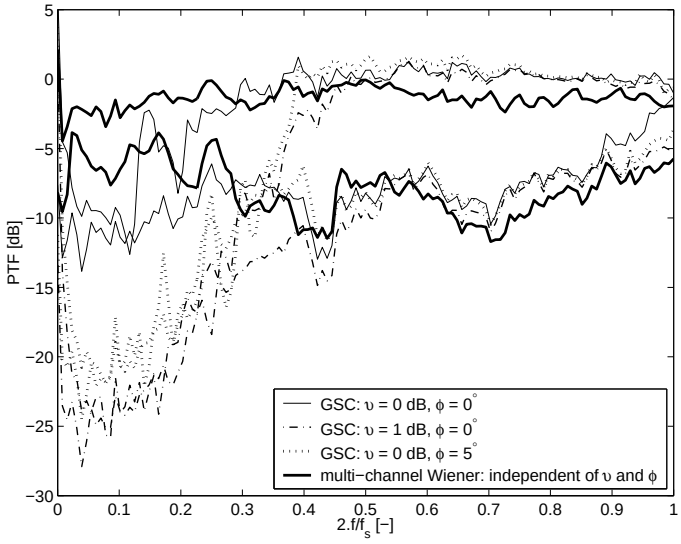


Fig. 11. Effect of microphone gain mismatch υ and phase mismatch ϕ respectively on the PTF of speech and noise of GSC and multi-channel Wiener filtering for a diffuse noise field.

on the noise sensitivity is indispensable to guarantee robustness in complicated noise scenarios. The multi-channel Wiener filter on the other hand is independent of the applied deviations. Provided enough processing accurracy is available, the multi-channel Wiener filter needs no or a less severe constraint, resulting in improved noise reduction.

Figure 12 plots the noise sensitivity $\Psi(f)$ for both algorithms[3]. For both algorithms, $\Psi(f)$ has a shape, similar to Figure 3. Compared to Figure 3, the noise sensitivity $\Psi(f)$ of the GSC is larger, especially at low $f.d.$ At low $f.d$, the internal-to-external noise ratio $\beta(f)$ 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 and the degradation in the performance of the GSC due to microphone mismatch is more significant at low frequencies than in Figure 6. In this example, the noise sensitivity of the multi-channel Wiener filtering technique is 2 dB to 5 dB lower than the sensitivity of the GSC. The same holds for the attenuation of the noise. This is due to the reverberation in the desired speech of the multi-channel Wiener filter. In contrast to the GSC, the Wiener filter does not dereverberate but tries to estimate the reverberant speech signal, i.e. as received at the first microphone. In a diffuse noise field, the filter trades off noise reduction and cancellation of the reverberant part in the speech signal.

[3]In the calculation of the noise sensitivity of the GSC, $G_s(f)$ is set to 1 to compare with Figure 3. Due to the small model error, $G_s(f) < 1$ (see Figure 11), which results in a larger $\Psi(f)$ than depicted in Figure 12.

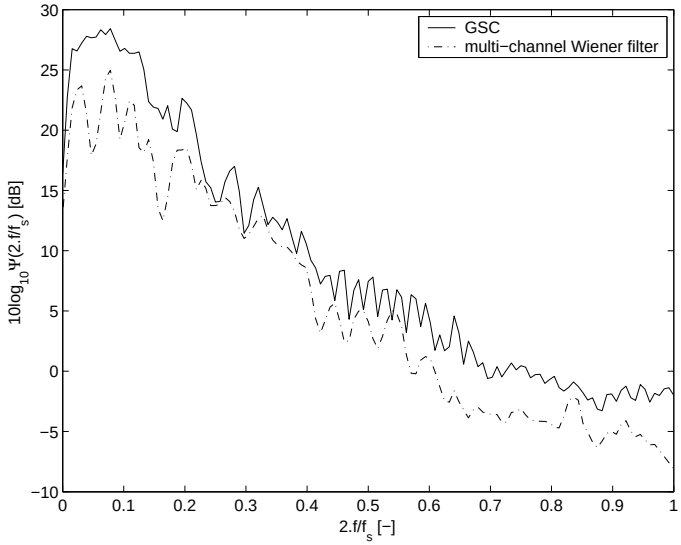


Fig. 12. Experimentally obtained noise sensitivity of the GSC and the multi-channel Wiener filter in a diffuse noise field.

2) Influence on intelligibility weighted performance measures: In Figure 13(a)-(b) and Figure 13(c)-(d), the performance measures $\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$ are given as a function of the amount of microphone gain mismatch υ and microphone phase mismatch ϕ, respectively. The performance of the multi-channel Wiener filter is indeed independent of υ and ϕ, whereas a small amount of gain and phase mismatch suffice to deteriorate the performance of the GSC. Due to the presence of reverberation, the performance of the multi-channel Wiener filter is lower than the theoretical performance. The performance curves of the GSC with a constraint on the noise sensitivity 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 w.r.t. the power of the noise in the pre-emphasized, first microphone signal. The curves for $\eta \approx -20$ dB and -10 dB illustrate that a noise sensitivity constraint results in a smaller performance degradation in case of gain and/or phase mismatch at the expense in less noise reduction. In complicated noise scenarios such as diffuse noise, and at low $f.d$, the multi-channel Wiener filtering outperforms the robust GSC with noise sensitivity constraint. Note that the curves for $\eta \approx -20$ dB and $\eta \approx -10$ dB resemble well the ones predicted in Figure 6, especially the ones depicting the distortion $\text{SD}_{\text{intellig}}$. The curves for the GSC without noise sensitivity constraint are different then the ones drawn in Figure 6 for $\beta = -30$ dB due to the different internal-to-external noise level $\beta(f)$: indeed, the experimentally found performance of the GSC with $\beta(f)$ increased to about -30 dB, depicted in Figure 13, is similar to the one shown in Figure 6. Given the amount of gain and phase mismatch that may occur, the appropriate amount of spatially uncor-

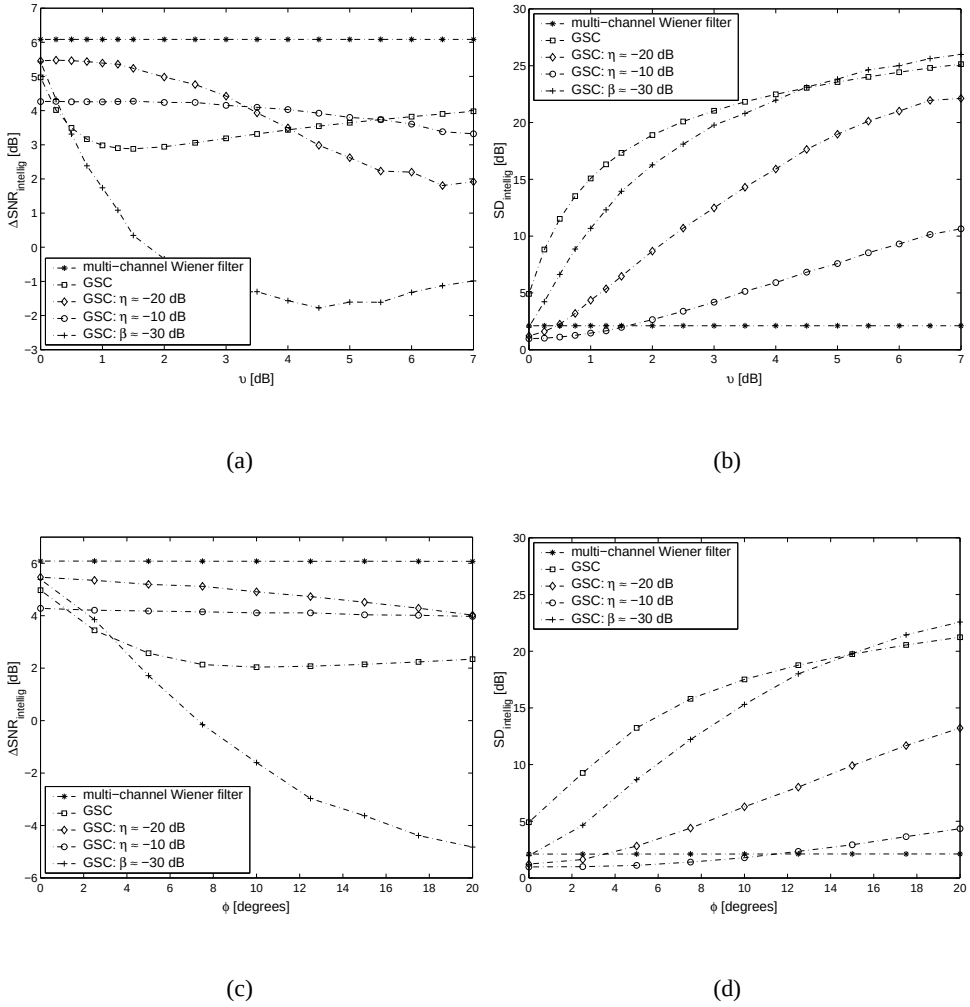


Fig. 13. Broadband performance measures ((a)-(c) $\Delta\text{SNR}_{\text{intellig}}$; (b)-(d) $\text{SD}_{\text{intellig}}$) of GSC and multi-channel Wiener filter in a diffuse noise field as a function of the amount of microphone sensitivity mismatch; (a)-(b) Microphone gain mismatch υ (in dB); (c)-(d) Microphone phase mismatch ϕ (in degrees).

related noise injection or the appropriate constraint on $\Psi(f)$ should be chosen so that the performance is not severely affected in the presence of the mismatch. The more severe the constraint that is needed, the less the noise reduction that is achieved with the robust GSC.

B. Robustness to variations in speaker location

Figure 14 illustrates the effect of a variation $\Delta\Theta$ (in degrees) in speaker location w.r.t. $\Theta = 0^\circ$ on the improvement $\Delta\text{SNR}_{\text{intellig}}$ and on the speech distortion $\text{SD}_{\text{intellig}}$ for the multi-channel Wiener filtering and the GSC. As predicted theoretically, no significant degradation of the performance of the GSC occurs for speaker variations $\Delta\Theta$ up to 30° and -30°. In contrast to Figure 10, the performance measures

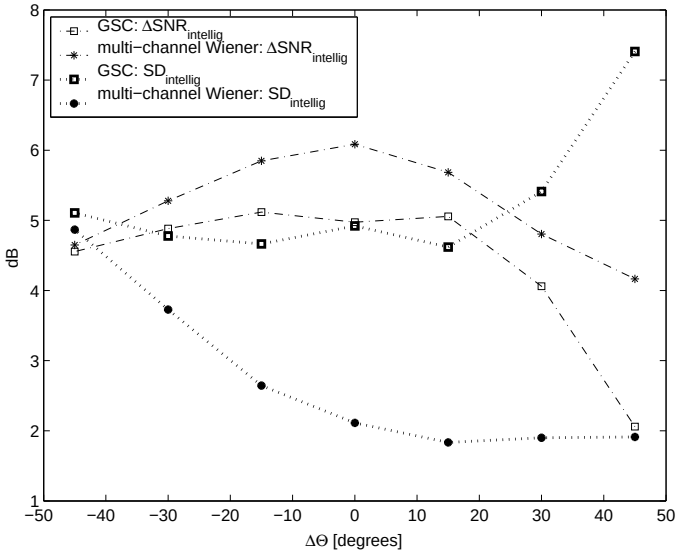


Fig. 14. Influence of a variation $\Delta\Theta$ (in degrees) in speaker location on the performance of the GSC and the multi-channel Wiener filtering.

are asymmetric for positive and negative $\Delta\Theta$. 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 $\alpha(f)(1+\beta(f)) \approx \alpha(f)$. As shown in Figure 10, a higher $\alpha(f)$ induces a larger amount of speech distortion by the multi-channel Wiener filter. The distortion caused by the GSC on the other hand, is smaller for negative $\Delta\Theta$ than for positive $\Delta\Theta$. Beside the blocking matrix of the GSC, the head additionaly attenuates speech signals coming from negative $\Delta\Theta$ compared to signals from positive $\Delta\Theta$. Hence, the amount of speech leakage into the noise references is smaller for speech coming from negative angles $\Delta\Theta$.

VI. Conclusions

Multi-microphone noise reduction techniques for hearing aid applications go together with the use of small-sized arrays. Considerable noise reduction can be achieved with such arrays, but at the expense of an increased noise sensitivity and hence an expected increased sensitivity to errors in the assumed signal model. A robust algorithm should be insensitive to small model errors such as microphone mismatch, variations in speaker and microphone positions. Recently, multi-channel Wiener filtering techniques [19], [20], [21], [22] have been proposed. In contrast to the GSC, they do not require prior knowledge of the signal mode and are thus expected to be more robust. We have evaluated the robustness of the GSC and the multi-channel Wiener filtering techniques against different types of model errors theoretically and

verified the results experimentally.

In Section II, the theoretical optimal noise reduction performance of the GSC has briefly been reviewed and extended to also qualify speech distortion and noise sensitivity. In addition, the optimal performance of GSC with noise sensitivity constraint is derived. Section III reviewed the concept of the multi-channel Wiener filtering and formulas describing its optimal performance have been derived. Based on the theoretical formulas of Section II and Section III, the robustness of both algorithms has been investigated. The theoretical analysis showed that the noise sensitivity grows with an increasing complexity of the noise scenario, with decreasing $f.d$ and increasing number of microphones per array length (see Section IV). Under these conditions, robustness issues are found to be crucial. The effect of different types of model errors on the performance of the GSC and multi-channel Wiener filter have been investigated. Among them, microphone mismatch is particularly harmful to GSC. To increase the robustness of the GSC a constraint on the noise sensitivity is needed, which goes at the expense of less noise reduction. The multi-channel Wiener filter on the other hand, is not degraded by small model errors. Hence, it needs a less severe or no constraint on the noise sensitivity so that- compared to the robust GSC with noise sensitivity constraint- a better performance is achieved, especially in complicated noise scenarios.

The theoretical analysis assumes infinite filter lengths and a localized speech source and localized or diffuse noise sources in the far and free field. In Section V, the different theoretical results have been validated for real recordings in an office room made with a 4-microphone endfire array mounted on a dummy head. Different model errors have been studied: microphone mismatch and variation in speaker location. For all type of model errors, the experimental results confirm the optimal ones. Hence, the analytical framework offers an easy way to predict the consequences of model errors on the performance of the noise reduction algorithms.

ACKNOWLEDGEMENTS

Ann Spriet is a Research Assistant supported by the F.W.O. 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.0295.97 (‘Design and implementation of adaptive digital signal processing algorithms for broadband applications’), Research Project FWO nr. G.0233.01 (‘Signal processing and automatic patient fitting for advanced auditory prostheses’). The scientific responsibility is assumed by its authors.

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] W. Soede, A. J. Berkhout, and F. A. Bilsen, "Development of a Directional Hearing Instrument Based on Array Technology," *J. Acoust. Soc. Amer.*, vol. 94, no. 2, pp. 785–798, Aug. 1993.

[5] J. M. Kates, "Superdirective Arrays for Hearing Aids," *J. Acoust. Soc. Amer.*, vol. 94, no. 4, pp. 1930–1933, Oct. 1993.

[6] 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.

[7] 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.

[8] K. M. Buckley, "Broad-Band Beamforming and the Generalized Sidelobe Canceller," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 34, no. 5, pp. 1322–1323, Oct. 1986.

[9] 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.

[10] M. Kompis and N. Dillier, "Noise Reduction for Hearing Aids: Combining Directional Microphones with an Adaptive Beamformer," *J. Acoust. Soc. Amer.*, vol. 96, no. 3, pp. 1910–1913, Sept. 1994.

[11] 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.

[12] 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.

[13] M. W. Hoffman, T. D. Trine, K. M. Buckley, and D. J. Van Tasell, "Robust Adaptive microphone array processing for hearing aids: realistic speech enhancement predictions," *J. Acoust. Soc. Amer.*, vol. 96, pp. 759–770, 1994.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] N. K. Jablon, "Adaptive beamforming with the generalized sidelobe can in the presence of array imperfections," *IEEE Trans. Antennas Propag.*, vol. 34, pp. 996–1012, Aug. 1986.

[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] G. Rombouts and M. Moonen, "QRD-based optimal filtering for acoustic noise reduction," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Toulouse, France, Sept. 2002, vol. 3, pp. 301–304.

[22] 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.

[23] A. Spriet, M. Moonen, and J. Wouters, "A multichannel subband GSVD based approach for speech enhancement in hearing aids," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Darmstadt, Germany, Sept. 2001, pp. 187–191.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] C. Marro, Y. Mahieux, and K. U. Summer, "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.

[30] B. Widrow and S. Stearns, *Adaptive Signal Processing*, Prentice Hall, Englewood Cliffs, 1985.

[31] J. Bitzer, K. U. Simmer, and K.-D. Kammeyer, "An Alternative Implementation of the Superdirective Beamformer," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz NY, USA, Oct. 1999, pp. 7–10.

[32] 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.

[33] Acoustical Society of America, "ANSI S3.5-1997 American National Standard Methods for Calculation of the Speech Intelligibility Index," June 1997.

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

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

[36] 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.