Katholieke Universiteit Leuven

Departement Elektrotechniek

ESAT-SISTA/TR 2002-163

The impact of speech detection errors on the noise reduction performance of multi-channel Wiener filtering and Generalized Sidelobe Cancellation [1]

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

November 2002

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/sista/spriet/reports/02-163.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/ ENT-Dept., 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 the frame of 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 was partially sponsored by Cochlear. 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.

The impact of speech detection errors on the noise reduction performance of multi-channel Wiener filtering and Generalized Sidelobe Cancellation

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

While the noise reduction performance of the Generalized Sidelobe Canceller (GSC) depends on the validity of a priori assumptions about the signal model, the recently developed multi-channel Wiener filter technique does not rely on any such a priori information. This provides a potential benefit of the latter over GSC. However, both techniques also rely on a speech detection algorithm. In this paper, we analyze the average effect of speech detection errors on the performance of the GSC and the multi-channel Wiener filtering technique both theoretically and experimentally. In the GSC case, it is the simultaneous presence of signal model errors and speech detection errors that affects performance. Theoretical analysis (for infinite filter lengths) shows that already in the presence of small signal model errors, the GSC is more affected by speech detection errors than the multi-channel Wiener filter, both with and without Adaptive Noise Canceller (ANC) postprocessing. Incorporating a constraint on the noise sensitivity of the GSC limits the drastic impact of speech detection errors at the expense of reduced noise reduction performance. It is shown that the multi-channel Wiener filter technique preserves its benefit over the GSC for a reasonable speech detection error rate of 20% or less, even when the GSC is supplied with a noise sensitivity constraint. Real data experiments confirm the theoretical results.

Index Terms

Multi-microphone noise reduction, speech detection, multi-channel Wiener filter, Generalized Sidelobe Canceller

I. INTRODUCTION

In speech communication applications, such as teleconferencing, handsfree telephony and hearing aids, the presence of background noise and/or reverberation seriously reduces the intelligibility of the desired speech signal. This stems from the large distance between the speaker and the microphone(s). Hence, the use of a noise reduction algorithm is necessary. 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). A widely used adaptive multi-microphone noise reduction technique is the Generalized Sidelobe Canceller (GSC) [1], [2], [3]. 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 fulfi lled so that leakage of speech into the noise references occurs, especially at high input Signal-to-Noise Ratios (SNR) [4], [5]. To avoid cancellation of speech, the Adaptive Noise Canceller (ANC) in the GSC is adapted during periods of noise only [6], [5], [7], hence the need for robust speech/noise detection. When used in combination with small-sized arrays, e.g. in hearing aid applications, an additional constraint on the noise sensitivity of the GSC [7], [8], [9], [10] is often included to guarantee performance in the presence of small errors in the assumed signal model, e.g. due to microphone mismatch [11], [12]. However, this constraint goes at the expense of less noise reduction.

Recently, a *multi-channel Wiener filtering* technique has been proposed that provides a Minimum Mean Square Error (MMSE) estimate of the desired signal portion in one of the received microphone signals [13], [14], [15], [16], [17], [18]. The procedure is uniquely based on estimates of the second order statistics of the recorded speech

signal and the noise signal. A robust speech detection is thus (again) needed. In [13], [14], the implementation of the multi-channel 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 [16]. A subband implementation [17] results in improved intelligibility at a significantly lower cost compared to the fullband approach, making it suitable for e.g. hearing aid applications. The multi-channel Wiener filter can be further extended with an ANC postprocessing stage [15]. In contrast to the GSC, the multi-channel Wiener filtering does not make any a priori assumptions about the signal model so that no or a less severe constraint on the noise sensitivity is needed to guarantee performance when used in combination with small-sized arrays [11], [12]. Especially in complicated noise scenarios such as multiple noise sources or diffuse noise, the multi-channel Wiener filter outperforms the GSC, even with noise sensitivity constraint [12].

The GSC and the multi-channel Wiener filtering technique both rely on a *speech detection algorithm*. In the GSC case, it is the simultaneous presence of signal model errors and speech detection errors that affects performance. In the multi-channel Wiener filtering case, as the procedure does not require any other a priori information, a reliable speech detection is expected to be crucial to achieve the potentially better performance. In this paper, we analyze the average effect of speech detection errors on the performance of the GSC (with and without noise sensitivity constraint) and the multi-channel Wiener filter technique both theoretically and experimentally. 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 [19]-[25]. In [19], [21], [22], [24], 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 [22], [25], a similar theoretical analysis is performed to evaluate a microphone array with postfiltering. In [11], [12], we have developed a theoretical framework to assess the performance of the GSC and the multi-channel Wiener filter and to evaluate their sensitivity to small signal model errors such as microphone mismatch. In this paper, we extend the theoretical formulas to predict the sensitivity of the GSC (with and without noise sensitivity constraint) and the multi-channel Wiener filtering (with and without ANC postprocessing) to speech detection errors. In the simulations and experiments, we focus on the harsh case of small-sized arrays as used in hearing aid applications.

The paper is organized as follows. In Section II and Section III, the concept and the theoretically optimal performance of the GSC and the multi-channel Wiener filter, respectively, are briefly reviewed. The formulas are extended to also assess the performance of the multi-channel Wiener filtering with ANC postprocessing stage. Section IV theoretically analyzes the average impact of an erroneous speech detection on the performance of the algorithms. It is shown that the multi-channel Wiener filter technique preserves its benefit over the GSC for a reasonable speech detection error rate of 20% or less, even when the GSC is supplied with a noise sensitivity constraint. In Section V, the theory is verified based on real recordings made in an office room. The experimental results are found to be well predicted by the theoretical analysis.

Notation

Most signals will be represented in the frequency domain. The microphone signals are $X_k(f)$, $k = 1, ..., M$ with M the number of microphones. The Power Spectral Density (PSD) of signal $X(f)$ 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 $\mathrm{Re}\{B\}$ and $\mathrm{Im}\{B\}$, respectively. The superscript T denotes the transpose operator, the superscript H the conjugate transpose.

II. Generalized Sidelobe Cancellation

A. Concept

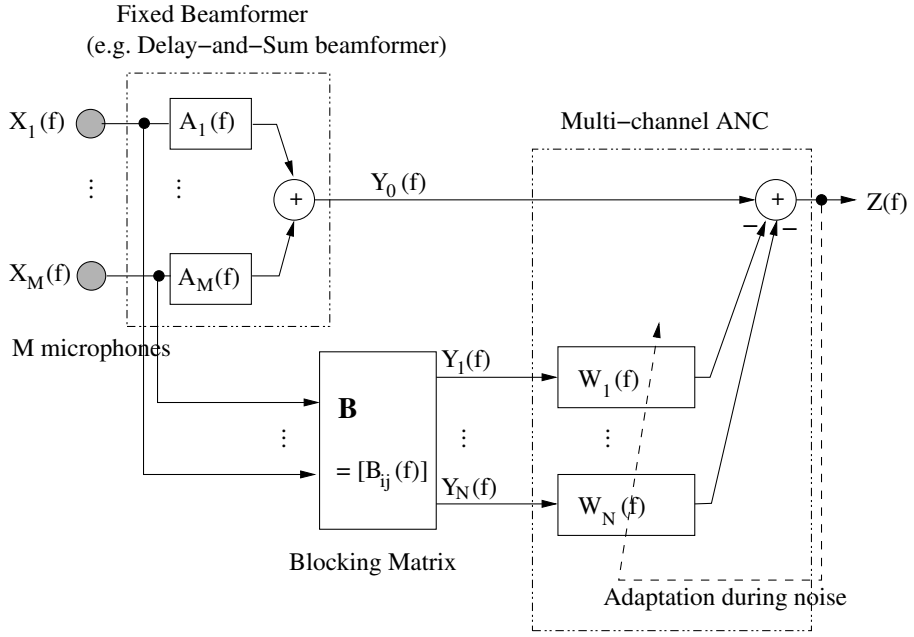


Fig. 1. Concept of the Generalized Sidelobe Canceller.

The Generalized Sidelobe Canceller (GSC) [1], [2], [3], depicted in Figure 1, consists of a fixed beamformer $\mathbf{A} \in \mathcal{C}^{1\times M}$ (e.g. delay-and-sum), a blocking matrix $\mathbf{B}$ and an Adaptive Noise Canceller (ANC) [26]. 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 the front at 0 degrees.

The $N \times M$ blocking matrix $\mathbf{B}$ creates N noise references $Y_j(f)$, $j = 1, \cdots N$ by steering zeroes to the front. The ANC filters $W_j(f)$, $j = 1, \cdots, N$ reduce the portion of 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 [6], [5], [7], using a speech/noise detector.

In the sequel, we refer to the output of the ANC as $Z(f)$, the filter of the fixed beamformer applied to the j-th microphone is called $A_j(f)$. We assume that a delay-and-sum beamformer is used, i.e. $A_j(f) = \frac{1}{M} e^{\mathrm{i}2\pi f \frac{d_j - d_1}{c}}$ with $d_j - d_1$ the interspacing between the j-th and the first microphone and c the velocity of sound in air ($c \approx 340$ m/s). The $N \times M$ blocking matrix $\mathbf{B}$ consists of the filters $B_{ij}(f)$, $i = 1, \cdots, N$, $j = 1, \cdots, M$. In the sequel, the number of noise references N equals $M - 1$ and for e.g. $M = 4$, $\mathbf{B}$ is set to

$$\mathbf{B} = \begin{bmatrix} 1 & -e^{\mathrm{i}2\pi f \frac{d_2 - d_1}{c}} & 0 & 0 \\ 1 & 0 & -e^{\mathrm{i}2\pi f \frac{d_3 - d_1}{c}} & 0 \\ 1 & 0 & 0 & -e^{\mathrm{i}2\pi f \frac{d_4 - d_1}{c}} \end{bmatrix}. \tag{1}$$

B. Theoretical analysis

In [21], [24], [27], 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 $G^n(f) = \frac{P_Z^n(f)}{P_X^n(f)} = \frac{1}{\mathrm{NR}(f)}$, also the PTF of the desired speech signal $G^s(f) = \frac{P_Z^s(f)}{P_X^s(f)}$ can be expressed as a function of $\Gamma_{kl}(f)$ [11], [12]. In [12], also the PTFs $\tilde{G}^s(f)$ and $\tilde{G}^n(f)$ of the GSC with noise sensitivity constraint have been derived. Based on the PTFs $G^s(f)$ and $G^n(f)$, broadband intelligibility weighted performance measures are computed to assess the overall performance of the algorithm. These results are reviewed in this section.

We assume all microphones have the same spatial characteristic and frequency characteristic. The actual characteristic then does not influence the coherence functions as well as $G^s(f)$ and $G^n(f)$. See below (Remark 1) for comments on microphone mismatch effects.

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} due to a *localized sound source* at an angle θ with respect to the front direction in free and far field equals

$$\Gamma_{kl}(f) = e^{-\mathrm{i}2\pi f \cos(\theta) \frac{d_k - d_l}{c}} \tag{3}$$

and the coherence Γ_{kl} due to 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 [21]. The diffuse sound fi eld is a good approximation of sound fi elds 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 $\Gamma_{kl}(f)$ due to a set of K *uncorrelated sound sources*, individually inducing coherence functions $\Gamma^j_{kl}(f)$ and PSDs $P^j(f)$, $j = 1, ..., K$ is given by

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

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 [22]

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

since $\Gamma^i_{kl} = \delta[k-l]$, 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. Typical values for $\beta(f)$ lie between -20 dB and -40 dB [22].

2) PTFs $G^s(f)$ and $G^n(f)$ of the GSC : The PTFs $G^s(f)$ and $G^n(f)$ can be expressed as a function of $\Gamma^s_{kl}(f)$ and $\Gamma^n_{kl}(f)$, the fi xed beamformer $\mathbf{A}(f)$ and the blocking matrix $\mathbf{B}(f)$ [12]. The ANC fi lters $\mathbf{W}(f)$ are given by:

$$\mathbf{W}(f) = \begin{bmatrix} P^n_{Y_1Y_1} & P^n_{Y_2Y_1} & \cdots & P^n_{Y_NY_1} \\ P^n_{Y_1Y_2} & P^n_{Y_2Y_2} & \cdots & P^n_{Y_NY_2} \\ \vdots & \vdots & \ddots & \vdots \\ P^n_{Y_1Y_N} & P^n_{Y_2Y_N} & \cdots & P^n_{Y_NY_N} \end{bmatrix}^{-1} \begin{bmatrix} P^n_{Y_0Y_1} \\ \vdots \\ P^n_{Y_0Y_N} \end{bmatrix}, \tag{7}$$

where $P^n_{Y_jY_i}$ is the cross-PSD of the noise contributions in $Y_j(f)$ and $Y_i(f)$ [27]. $P^n_{Y_jY_i}$ 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 fi lters $W_i(f)$ thus only depend on the spatial characteristics of the noise and not on $P^n_X(f)$.

Using (7)-(8), $G^s(f) = P^s_Z(f)/P^s_X(f)$ and $G^n(f) = P^n_Z(f)/P^n_X(f)$ can be expressed as a function of Γ^s_{kl} and Γ^n_{kl} [12]:

$$\begin{aligned} G^{n|s}(f) &= \Bigg[P^{n|s}_{Y_0}(f) - 2\sum_{i=1}^{N}\mathrm{Re}\{P^{n|s}_{Y_0Y_i}(f)W^*_i(f)\} \\ &\quad + \sum_{i=1}^{N}\sum_{j=1}^{N} W_i(f)W^*_j(f)P^{n|s}_{Y_iY_j}(f)\Bigg]/P^{n|s}_X(f), \end{aligned} \tag{9}$$

where $G^{n|s}(f)$ means $G^n(f)$ or $G^s(f)$, etc. (for compact notation). Here $P^{n|s}_{Y_0}$ is the PSD at the output of the fi xed

beamformer, i.e.

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

Remark 1: Equations (3)-(4) and (8)-(10) assume that the microphones are perfectly matched. Microphone mismatch can be easily incorporated by modifying Γ_{kl}^{s} and Γ_{kl}^{n} in (8)-(10) [12]. 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 $\tilde{\Gamma}_{kl}(f)$

$$\tilde{\Gamma}_{kl}(f) = \begin{cases} \Delta\Upsilon(f) e^{\mathrm{i}\Delta\Phi(f)\frac{\pi}{180}} \Gamma_{kl}(f) & \text{for } l \neq k; \\ |\Delta\Upsilon(f)|^2 \Gamma_{kl}(f) & \text{for } l = k. \end{cases} \tag{11}$$

3) PTFs $\tilde{G}^s(f)$ and $\tilde{G}^n(f)$ of the GSC with noise sensitivity constraint : To increase the robustness of the GSC against errors in the assumed signal model, a constraint on the noise sensitivity is often applied [7], [8], [9], [28]. Especially for small-sized arrays and in complicated noise scenarios, this constraint is found to be crucial to guarantee performance [12]. The *noise sensitivity* is defined as the ratio of the spatially white noise gain to the gain of the desired signal [8], [19]. It is used to quantify the sensitivity to errors in the assumed signal model (such as microphone mismatch) as well as to processing errors (such as round off errors due to finite processing accuracy) and determines how much the internal noise, e.g. sensor noise and quantization noise, is amplified. In [8], it is shown that a constraint on the noise sensitivity corresponds to a quadratic inequality constraint on the filters $\mathbf{W}(f)$, namely

$$\mathbf{W}^T(f)\mathbf{B}(f)\mathbf{B}^H(f)\mathbf{W}^*(f) \leq \zeta. \tag{12}$$

In [9], a similar constraint is imposed by inserting uncorrelated noise in the microphone signals[2] when designing the ANC filters $\mathbf{W}(f)$. The unaltered microphone signals are filtered. In [8], 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 (i.e. the PTFs $\tilde{G}^s(f)$, $\tilde{G}^n(f)$) is obtained by replacing the cross-PSD $P_X^n(f)\Gamma_{kl}^n(f)$ in the computation of $\mathbf{W}(f)$ by

$$P_X^n(f)\left(\Gamma_{kl}^n(f) + \eta(f)\delta[k-l]\right), \tag{13}$$

where $\eta(f)$ is the ratio of the injected noise power to the input noise power $P_X^n(f)$ [12]. The equivalent noise injection level $\eta(f)$ for a given ζ is scenario dependent.

4) Intelligibility weighted performance measures: To assess the effect of the obtained PTFs $G^s(f)$ and $G^n(f)$ on intelligibility, broadband intelligibility weighted measures have been proposed, such as the improvement in *intelligibility weighted Signal-to-Noise Ratio (SNR)* [29], which is defined as

$$\Delta\mathrm{SNR}_{\mathrm{intellig}} = \sum_i I_i(\mathrm{SNR}_{i,\mathrm{out}} - \mathrm{SNR}_{i,\mathrm{in}}), \tag{14}$$

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, $\mathrm{SNR}_{i,\mathrm{out}}$ is the output SNR (in dB) and $\mathrm{SNR}_{i,\mathrm{in}}$ is the input SNR (in dB) in the i-th

[2]Since $\mathbf{B}(f).\mathbf{A}^H(f) = 0$, this corresponds to adding a diagonal loading term to the spatiospectral correlation matrix of the noise in (7).

one-third octave band. The center frequencies f_i^c and the values I_i are defined in [30]; 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

$$\text{SNR}_{i,\text{in}} = 10\log_{10}\left(\int_{2^{-\frac{1}{6}}f_i^c}^{2^{\frac{1}{6}}f_i^c} P_X^s(f)df \Big/ \int_{2^{-\frac{1}{6}}f_i^c}^{2^{\frac{1}{6}}f_i^c} P_X^n(f)df\right) \quad (15)$$

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

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

$$\text{SD}_{\text{intellig}} = \sum_i I_i \text{SD}_i, \quad (17)$$

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

$$\text{SD}_i = \int_{2^{-\frac{1}{6}}f_i^c}^{2^{\frac{1}{6}}f_i^c} |10\log_{10} G^s(f)|\, df \Big/ \left[\left(2^{\frac{1}{6}} - 2^{-\frac{1}{6}}\right) f_i^c\right]. \quad (18)$$

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

III. Multi-channel Wiener filtering

A. Concept

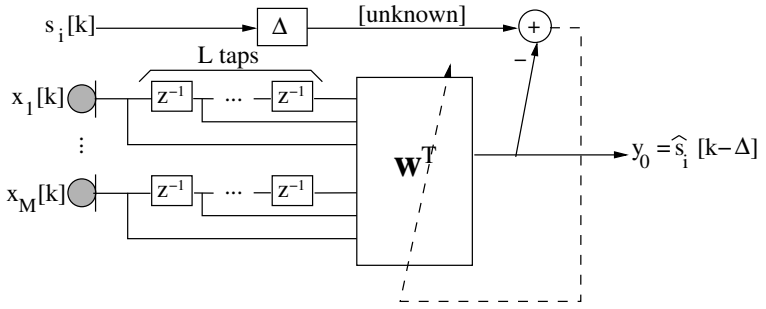


Fig. 2. Concept of the multi-channel Wiener filter.

The multi-channel Wiener filter $\mathbf{w} \in \mathcal{C}^{ML\times 1}$, described in Figure 2, provides a Minimum Mean Square Error (MMSE) 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

$$\mathbf{w}^* = \mathcal{E}\{\mathbf{x}[k]\mathbf{x}^H[k]\}^{-1}\mathcal{E}\{\mathbf{x}[k]s_i^*[k-\Delta]\}, \quad (19)$$

$$\mathbf{x}[k] = [x_1[k]\cdots x_1[k-L+1]\cdots x_M[k]\cdots x_M[k-L+1]]^T \quad (20)$$

where $s_i[k-\Delta] = x_i[k-\Delta] - n_i[k-\Delta]$, is obviously unknown and where L is the number of filter taps per channel [14].

Assuming that the speech and noise signals are uncorrelated, (19) becomes:

$$\mathbf{w}^* = \mathcal{E}\{\mathbf{x}[k]\mathbf{x}^H[k]\}^{-1} \left(\mathcal{E}\{\underbrace{\mathbf{x}[k]x_i^*[k-\Delta]}_{\text{speech+noise}}\} - \mathcal{E}\{\underbrace{\mathbf{n}[k]n_i^*[k-\Delta]}_{\text{noise only}}\} \right). \tag{21}$$

The second order statistics of the noise signal are assumed to be sufficiently stationary so that they can be estimated during periods of noise only. A robust speech detection is thus (again) needed. Note that the multi-channel Wiener filter combines speech distortion and noise reduction in its MMSE optimization criterion

$$\min_{\mathbf{w}} \mathcal{E}\{\left(\mathbf{w}^T\mathbf{s}[k] - s_i[k-\Delta]\right)^2\} + \mathcal{E}\{\left(\mathbf{w}^T\mathbf{n}[k]\right)^2\}. \tag{22}$$

In [13], [14], 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 [16]. In [17], [18], a subband implementation of the GSVD based algorithm has been developed to increase intelligibility and at the same time reduce complexity. In contrast to the GSC, the multi-channel Wiener filter does not make any a priori assumptions about the signal model so that it is robust against small signal model errors such as microphone mismatch, reverberation, variations in speaker location and microphone positions [11], [13]. 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. $s_1[k-\Delta]$.

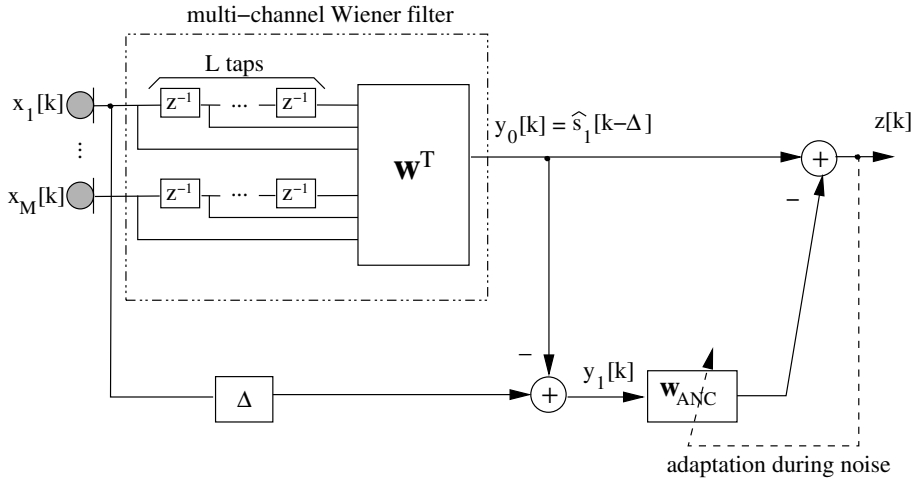


Fig. 3. Concept of the multi-channel Wiener filter with ANC postprocessing stage

The multi-channel Wiener filter can be further extended with an *ANC postprocessing stage* [15] as described in Figure 3. In this way, the number of taps used in the more complex Wiener filtering stage can be reduced. The potential performance degradation due to this reduced number of taps can be compensated for by the much cheaper

ANC postprocessing stage. The output of the multi-channel Wiener filter $y_0[k]$ is used as a speech reference in the ANC, whereas the noise reference can be created in different ways. An obvious way of creating the noise reference $y_1[k]$ at time instant k consists of subtracting the speech reference $y_0[k] = \hat{s}_i[k-\Delta]$ from the delayed i-th microphone signal $x_i[k-\Delta]$,

$$y_1[k] = x_i[k-\Delta] - y_0[k]. \tag{23}$$

As in the GSC, the ANC is adapted during periods of noise only.

B. Theoretical analysis

In [11], [12], the theoretically optimal performance of the multi-channel Wiener filter without ANC postprocessing stage, i.e. the PTF of noise and speech as well as the intelligibility weighted performance measures, has been expressed as a function of Γ_{kl}, Γ_{kl}^n and Γ_{kl}^s. These results are briefly reviewed in this section. In addition, they are extended to the multi-channel Wiener filter with ANC postprocessing stage.

1) Multi-channel Wiener filtering: If

$$\mathbf{W}(f) = \begin{bmatrix} W_1(f) & W_2(f) & \cdots & W_M(f) \end{bmatrix}^T \tag{24}$$

with $W_i(f)$ filtering the i-th microphone signal, $\mathbf{W}(f)$ can be written as[3]:

$$\mathbf{W}(f) = \begin{bmatrix} P_X & P_{X_2X_1} & \cdots & P_{X_MX_1} \\ P_{X_1X_2} & P_X & & P_{X_MX_2} \\ \vdots & & \ddots & \vdots \\ P_{X_1X_M} & \cdots & \cdots & P_X \end{bmatrix}^{-1} \begin{bmatrix} P_X^s \\ P_{X_1X_2}^s \\ \vdots \\ P_{X_1X_M}^s \end{bmatrix}, \tag{25}$$

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

$$P_{X_kX_l}^s(f) = P_{X_kX_l}(f) - P_{X_kX_l}^n(f), \tag{26}$$

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

$$P_{X_kX_l}(f) = P_{X_kX_l}^s(f) + P_{X_kX_l}^n(f) \tag{27}$$

formula (25) can be rewritten as:

$$\mathbf{W}(f) = \frac{1}{1+\alpha(f)} \begin{bmatrix} 1 & \Gamma_{21} & \cdots & \Gamma_{M1} \\ \Gamma_{12} & 1 & & \Gamma_{M2} \\ \vdots & & \ddots & \vdots \\ \Gamma_{1M} & \cdots & \cdots & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \Gamma_{12}^s \\ \vdots \\ \Gamma_{1M}^s \end{bmatrix} \tag{28}$$

[3]In the theoretical analysis, we assume, without loss of generality, that the delay Δ equals zero.

where, using (5):

$$\Gamma_{kl}(f) = \frac{\alpha(f)\Gamma^n_{kl} + \Gamma^s_{kl}}{1+\alpha(f)}; \tag{29}$$

$$\alpha(f) = \frac{P^n_X(f)}{P^s_X(f)}; \tag{30}$$

In contrast to the GSC, the filter $\mathbf{W}(f)$ not only depends on the spatial characteristics of the noise $\Gamma^n_{kl}(f)$, but also on the input SNR $1/\alpha(f)$ and the spatial characteristics of the speech signal $\Gamma^s_{kl}(f)$.

Note that the ratio $1/(1+\alpha(f))$ in (28) corresponds to a *spectral* Wiener filter, which results in additional performance if the spectra of the speech and noise signals differ. Hence, (28) combines spatial and spectral filtering.

The PSD of the noise $Y^n_0(f)$ at the output equals

$$P^n_{Y_0}(f) = P^n_X \left(\sum_{k=1}^{M} |W_k|^2 + \sum_{k=1}^{M} \sum_{l=k+1}^{M} 2.\mathrm{Re}\{W_k W^*_l \Gamma^n_{kl}\} \right) \tag{31}$$

and is determined by Γ^s_{kl}, Γ^n_{kl} and $\alpha(f)$. The expression for the PSD $P^s_{Y_0}(f)$ can be found by replacing the superscripts n by s in (31). Using (31), the PTFs $G^n(f) = P^n_{Y_0}(f)/P^n_X$ and $G^s(f) = \frac{P^s_{Y_0}(f)}{P^s_X(f)}$ can be easily derived [12]:

$$G^{n|s}(f) = \sum_{k=1}^{M} |W_k|^2 + \sum_{k=1}^{M} \sum_{l=k+1}^{M} 2.\mathrm{Re}\{W_k W^*_l \Gamma^{n|s}_{kl}\} \tag{32}$$

2) Multi-channel Wiener filtering with ANC postprocessing: The multi-channel Wiener filter with ANC post-processing stage corresponds to a GSC in which the fixed matrices $\mathbf{A}(f)$ and $\mathbf{B}(f)$ are replaced by the adaptive matrices $\mathbf{W}(f)$ and $(\mathbf{e}_1 - \mathbf{W}(f))^T$, where $\mathbf{e}_1$ is defined as $[\,1 \quad 0 \quad \cdots \quad 0\,]^T$, and in which the number of noise references (blocking matrix outputs) N equals 1. Hence, the formulas for the filter $W_{\mathrm{ANC}}(f)$ and the PSDs $P^n_Z(f)$, $P^s_Z(f)$ can be easily derived using (7), (9) and (10).

The ANC postprocessing filter $W_{\mathrm{ANC}}(f)$, adapted during periods of noise only, equals

$$W_{\mathrm{ANC}}(f) = \left(P^n_{Y_1}(f)\right)^{-1} P^n_{Y_0Y_1}(f) \tag{33}$$

with

$$P^n_{Y_1}(f) = P^n_X(f) + P^n_{Y_0}(f) - 2\,\mathrm{Re}\{P^n_{Y_0X_1}(f)\} \tag{34}$$

$$P^n_{Y_0Y_1}(f) = -P^n_{Y_0}(f) + P^n_{Y_0X_1}(f) \tag{35}$$

$$P^n_{Y_0X_1}(f) = P^n_X(f) \sum_{k=1}^{M} W_k(f)\Gamma^n_{k1}(f) \tag{36}$$

and $P^n_{Y_0}(f)$ defined in (31).

The PSD of the noise $P^n_Z(f)$ at the output of the ANC equals

$$P^n_Z(f) = P^n_{Y_0}(f) \left(1 + |W_{\mathrm{ANC}}|^2 \frac{P^n_{Y_1}(f)}{P^n_{Y_0}(f)} - 2\,\mathrm{Re}\{W^*_{\mathrm{ANC}} \frac{P^n_{Y_0Y_1}(f)}{P^n_{Y_0}(f)}\}\right), \tag{37}$$

which -by using (33)- can also be written as

$$P^n_Z(f) = P^n_{Y_0}(f) - |W_{\mathrm{ANC}}|^2 P^n_{Y_1}(f) = P^n_{Y_0}(f)\left[1 - \left(\Gamma^n_{Y_0Y_1}(f)\right)^2\right], \tag{38}$$

with $\Gamma^n_{Y_0Y_1}(f) \in [0,\ 1]$ the complex coherence between the signals $Y_0^n(f)$ and $Y_1^n(f)$. The expressions for the PSD $P_Z^s(f)$ can be found by replacing the superscript n in (37) by the superscript s. Note that the simple equation (38) only holds for $P_Z^n(f)$ and for $W_{\mathrm{ANC}}(f)$ given by (33). From (37), the PTFs $\bar{G}^n(f)$ and $\bar{G}^s(f)$ from the microphone signal to the output of the ANC can be derived.

$$\boxed{\bar{G}^{n|s}(f) = G^{n|s}(f)\left(1 + |W_{\mathrm{ANC}}|^2\frac{P^{n|s}_{Y_1}(f)}{P^{n|s}_{Y_0}(f)} - 2\,\mathrm{Re}\{W^*_{\mathrm{ANC}}\frac{P^{n|s}_{Y_0Y_1}(f)}{P^{n|s}_{Y_0}(f)}\}\right),} \tag{39}$$

where $G^n(f)$ and $G^s(f)$ are the PTFs of the multi-channel Wiener filter given by (32). Also here, microphone mismatch can be incorporated in the theoretical formulas by modifying Γ^s_{kl} and Γ^n_{kl} according to (11) (see Remark 1 in Section II).

IV. Sensitivity to speech detection errors

The theoretical formulas derived in Section II and Section III allow to predict the average effect of speech detection errors on the performance of the algorithms. In [11], [12], it is shown that the multi-channel Wiener filtering technique -in contrast to the GSC - does not degrade in the presence of small signal model errors such as microphone mismatch. Especially in complicated noise scenarios, the multi-channel Wiener filtering is found to outperform the GSC as well as the more robust GSC with noise sensitivity constraint [12]. However, like the GSC, the multi-channel Wiener filtering technique relies on a speech detection algorithm. Since no other, a priori information is used, a reliable speech detection is expected to be crucial to achieve the potentially better performance. In this section, the effect of speech + noise detected as noise only and noise only detected as speech + noise on the performance of the algorithms are both investigated.

The results are illustrated for a small-sized uniform 4-microphone endfire array with microphone interspacing $d = 2$ cm and a diffuse noise field with internal-to-external noise ratio $\beta(f) = -30$ dB. Since in practice, the desired and interfering signals are often speech-like, the improvement $\Delta\mathrm{SNR}_{\mathrm{intellig}}$ is computed using the average long-term speech PSD of the signals found on [31] for $P_X^s(f)$ and $P_X^n(f)$. In the simulations, the second and the third microphone have a small gain mismatch $\Delta\Upsilon$ of 1 dB and -1 dB w.r.t. the first microphone. In [32], gain and phase differences between microphone characteristics of up to 6 dB and $10°$, respectively, have been reported. In addition, the microphone characteristics may drift over time, making perfect calibration practically impossible.

A. Generalized Sidelobe Canceller

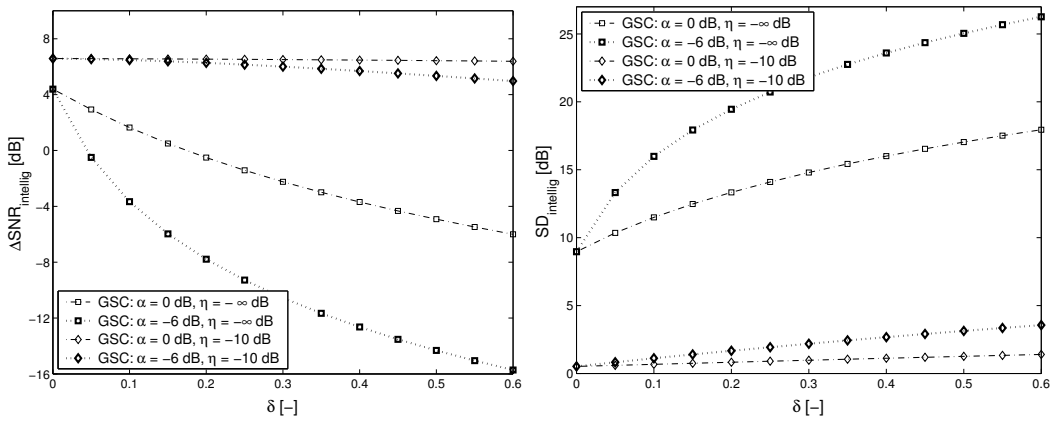
If the blocking matrix $\mathbf{B}$ is a perfect blocking matrix so that the noise references do not contain any speech, the performance of the GSC is roughly independent of the speech detection. In practice, the a priori assumptions of the GSC are violated due to e.g. the presence of reverberation, microphone mismatch, ... so that speech leaks into the noise references. To then prevent speech cancellation at high input SNR [4], the ANC is adapted during periods of noise only [6], [5], [7] and/or a constraint on the noise sensitivity of the GSC is applied [7], [8], [28]. For small-sized arrays, the latter is found to be crucial to guarantee performance, even when the ANC is only adapted

during noise [12]. Still, in the event of signal leakage, and whenever speech is detected as noise, the ANC may partially cancel the desired speech signal. On the other hand, whenever noise is detected as speech, the performance of the GSC will -apart from a slower convergence rate- not degrade, since the ANC is not adapted during periods of speech.

Suppose that $\delta \times 100\%$ of the noise samples used to compute the ANC filters actually contain speech. Suppose that the average PSD of the speech in the wrongly detected samples equals $\kappa(f)P_X^s(f)$, with $P_X^s(f)$ the average PSD of the speech in the correctly detected *speech + noise* samples. In the simulations, we set $\kappa(f) = 1$ to predict the average impact of speech detection errors. The formulas of the ANC filters are then obtained by replacing $P_{Y_j Y_i}^n(f)$ and $P_{Y_0 Y_i}^n(f)$ in (7)-(8) with $\tilde{P}_{Y_j Y_i}^n(f)$ and $\tilde{P}_{Y_0 Y_i}^n(f)$

$$\begin{aligned}\tilde{P}_{Y_j Y_i}^n(f) &= \sum_{k=1}^{M}\sum_{l=1}^{M} B_{jk}B_{il}^* \left(P_{X_k X_l}^n + \delta\kappa(f)P_X^s(f)\Gamma_{kl}^s\right) \\ &= P_X^n(f)\sum_{k=1}^{M}\sum_{l=1}^{M} B_{jk}B_{il}^* \left(\Gamma_{kl}^n + \frac{\delta\kappa(f)}{\alpha(f)}\Gamma_{kl}^s\right) \qquad (40)\\ \tilde{P}_{Y_0 Y_i}^n(f) &= P_X^n(f)\sum_{k=1}^{M}\sum_{l=1}^{M} A_k B_{il}^* \left(\Gamma_{kl}^n + \frac{\delta\kappa(f)}{\alpha(f)}\Gamma_{kl}^s\right). \qquad (41)\end{aligned}$$

The impact of an erroneous speech detection on the GSC depends on the quality of the noise references $Y_j = \sum_{k=1}^{M} B_{jk}X_k$, $j = 1, ..., N$: the larger the ratio of speech leakage to noise in the noise references, the larger the impact. From (40)-(41), it follows that the higher the SNR, i.e. the lower $\alpha(f)$, the more the ANC filters and hence the performance will be affected when model errors are present.



(a) (b)

Fig. 4. Theoretical improvement in $\mathrm{SNR}_{\mathrm{intellig}}$ (a) and spectral distortion $\mathrm{SD}_{\mathrm{intellig}}$ (b) of the GSC in a diffuse noise field with $\beta(f) = -30$ dB and in the case of a small microphone mismatch, when $(\delta \times 100)\%$ of the noise samples used to compute the ANC, actually contain speech.

Figure 4(a) and (b) depict the average impact ($\kappa(f) = 1$) of speech detection errors on the improvement in intelligibility weighted SNR and on the speech distortion $\mathrm{SD}_{\mathrm{intellig}}$ obtained by the GSC in the case of a small microphone

mismatch. Two input SNR values $1/\alpha(f)$, i.e. $\alpha(f) = 0$ dB and $\alpha(f) = -6$ dB, are considered. In addition, the performance of the more robust GSC with noise sensitivity constraint (described in Section II) is depicted for a noise injection level $\eta(f) = -10$ dB. This noise injection level has been found to provide sufficient robustness against gain mismatches $\Delta\Upsilon$ up to ± 3 dB and phase mismatches $\Delta\Phi$ up to $\pm 6^\circ$ for the given microphone array and the given noise scenario [12]. For larger model errors, a more severe constraint will be needed to keep distortion limited, at the expense of less noise reduction [12]. Without the noise sensitivity constraint, the performance of the GSC is strongly deteriorated by an erroneous speech detection, especially at high input SNR $1/\alpha(f)$. In addition to increasing robustness to model errors [8], [12], a noise sensitivity constraint is seen to reduce the impact of speech detection errors. This can be understood as follows. If the model errors are not too large, the noise reference of the GSC contains less speech than the speech reference. To cancel the speech effectively the ANC filters have to amplify the speech in the noise reference, without excessively amplifying internal noise. Large ANC filter amplitudes, however, result in a high noise sensitivity [8] and thus a large amplification of internal noise. As a consequence, the maximum amplification of the ANC filters and hence speech cancellation is determined by the internal-to-external noise ratio $\beta(f)$ and the constraint on the noise sensitivity, which is determined by the noise injection level $\eta(f)$.

B. Multi-channel Wiener filtering

1) Speech + noise detected as noise only: The multi-channel Wiener filter estimates the cross-PSDs $P_{X_k X_l}(f)$ with $k = 1, ..., M$ and $l = 1, ..., M$ during periods of *speech + noise* and the cross-PSDs $P^n_{X_k X_l}(f)$, $k = 1, ..., M$, $l = 1, ..., M$ during periods of *noise only*. Assume that the speech and noise signal are uncorrelated and that the second order statistics of the noise vary slowly in time. The cross-PSDs $P^s_{X_k X_l}(f)$ are then estimated using

$$P^s_{X_k X_l}(f) = P_{X_k X_l}(f) - P^n_{X_k X_l}(f). \tag{42}$$

In the case of a perfect speech detection, (42) applies and the multi-channel Wiener filter is given by (28). Suppose now that $(\delta \times 100)\%$ of the *noise only* samples used to estimate $P^n_{X_k X_l}(f)$ and to compute the multi-channel Wiener filter $\mathbf{W}(f)$ actually contain speech. Assume that the average PSD of the speech in the $(\delta \times 100)\%$ wrongly detected samples equals $\kappa(f) P^s_X(f)$, with $P^s_X(f)$ the average PSD of the speech in the correctly detected *speech + noise* samples. Then, the estimated cross-PSD $\hat{P}^n_{X_k X_l}(f)$ equals[4]

$$\begin{aligned} \hat{P}^n_{X_k X_l}(f) &= P^n_X(f)\Gamma^n_{kl} + \delta\kappa(f) P^s_X(f)\Gamma^s_{kl}(f) \\ &= P^s_X(f)\left(\alpha(f)\Gamma^n_{kl} + \delta\kappa(f)\Gamma^s_{kl}(f)\right) \end{aligned} \tag{43}$$

instead of

$$\hat{P}^n_{X_k X_l}(f) = P^n_{X_k X_l}(f) = P^i_X(f)\Gamma^i_{kl}(f) + P^e_X \Gamma^e_{kl}(f). \tag{44}$$

Using (42)-(43) and

$$\hat{P}_{X_k X_l}(f) = P_{X_k X_l}(f) = P^n_X(f)\Gamma^n_{kl}(f) + P^s_X(f)\Gamma^s_{kl}(f), \tag{45}$$

[4]Again, we assume that the second order statistics of the noise are quite stationary, so that the average PSD $P^n_X(f)$ of the noise in the wrongly and correctly detected speech + noise samples equal the average PSD of the noise only samples.

the estimated cross-PSD $\hat{P}^s_{X_k X_l}(f)$ becomes

$$\hat{P}^s_{X_k X_l}(f) = P^s_X(f)(1-\delta\kappa(f))\Gamma^s_{kl}(f) \tag{46}$$

instead of

$$\hat{P}^s_{X_k X_l}(f) = P^s_X(f)\Gamma^s_{kl}(f). \tag{47}$$

It can be shown with (25), (45) and (46) that (28) changes into

$$\boxed{\mathbf{W}(f) = \frac{(1-\delta\kappa(f))}{(1+\alpha(f))} \begin{bmatrix} 1 & \Gamma_{21} & \cdots & \Gamma_{M1} \\ \Gamma_{12} & 1 & & \Gamma_{M2} \\ \vdots & & \ddots & \vdots \\ \Gamma_{1M} & \cdots & \cdots & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \Gamma^s_{12} \\ \vdots \\ \Gamma^s_{1M} \end{bmatrix},} \tag{48}$$

with Γ_{kl} and Γ^s_{kl} the coherence functions between the k-th and l-th microphone in case of a perfect speech detection. **Hence, the PTFs $G^s(f)$ and $G^n(f)$ are both multiplied by $(1-\delta\kappa(f))^2$. This additional attenuation is independent of the input SNR and the noise scenario.** Consequently, a higher internal-to-external noise ratio $\beta(f)$ or a constraint on the noise sensitivity will not reduce the additional distortion that is caused by $\delta \neq 0$. In contrast to the GSC, the improvement in SNR $G^s(f)/G^n(f)$ at frequency f, and hence the improvement $\Delta\text{SNR}_{\text{intellig}}$, remains unchanged. In the simulations, we assume $\kappa(f) = 1$, illustrating the average effect of speech detection errors.

2) Noise detected as speech + noise: When $(\delta \times 100)\%$ of the *speech + noise* samples used in the computation of $P_{X_k X_l}(f)$ and $\mathbf{W}(f)$ actually contain noise only, the estimate $\hat{P}_{X_k X_l}(f)$ and as a consequence the estimate $\hat{P}^s_{X_k X_l}(f)$ become[5]

$$\begin{aligned} \hat{P}_{X_k X_l}(f) &= P^n_{X_k X_l}(f) + (1-\delta)P^s_{X_k X_l}(f) \\ &= P^s_X(f)(1-\delta)\left(\frac{\alpha(f)}{1-\delta}\Gamma^n_{kl}(f) + \Gamma^s_{kl}(f)\right) \qquad (49) \\ \hat{P}^s_{X_k X_l}(f) &= P^s_X(f)(1-\delta)\Gamma^s_{kl}(f), \qquad (50) \end{aligned}$$

so that the multi-channel Wiener fi lter equals

$$\boxed{\mathbf{W}(f) = \frac{1}{1+\tilde{\alpha}(f)} \begin{bmatrix} 1 & \tilde{\Gamma}_{21} & \cdots & \tilde{\Gamma}_{M1} \\ \tilde{\Gamma}_{12} & 1 & & \tilde{\Gamma}_{M2} \\ \vdots & & \ddots & \vdots \\ \tilde{\Gamma}_{1M} & \cdots & \cdots & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \Gamma^s_{12} \\ \vdots \\ \Gamma^s_{1M} \end{bmatrix},} \tag{51}$$

[5]Since the second order statistics of the noise are assumed to be quite stationary, the average PSD of the noise in the wrongly detected speech + noise samples, the average PSD of the noise in the correctly detected speech + noise samples and the average PSD of the noise only samples are assumed to be the same.

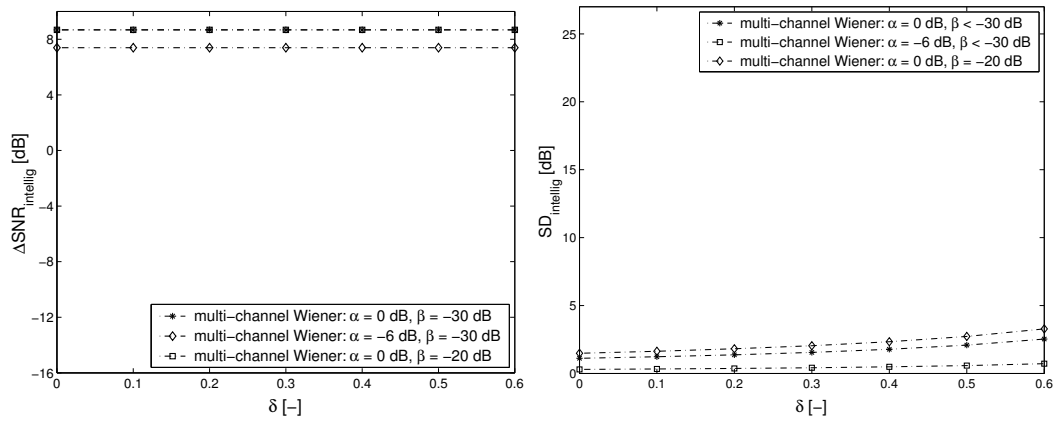
with

$$\tilde{\alpha}(f) = \frac{\alpha(f)}{1-\delta} \quad (52)$$

$$\tilde{\Gamma}_{kl}(f) = \frac{\tilde{\alpha}(f)\Gamma^{n}_{kl} + \Gamma^{s}_{kl}}{1+\tilde{\alpha}(f)}. \quad (53)$$

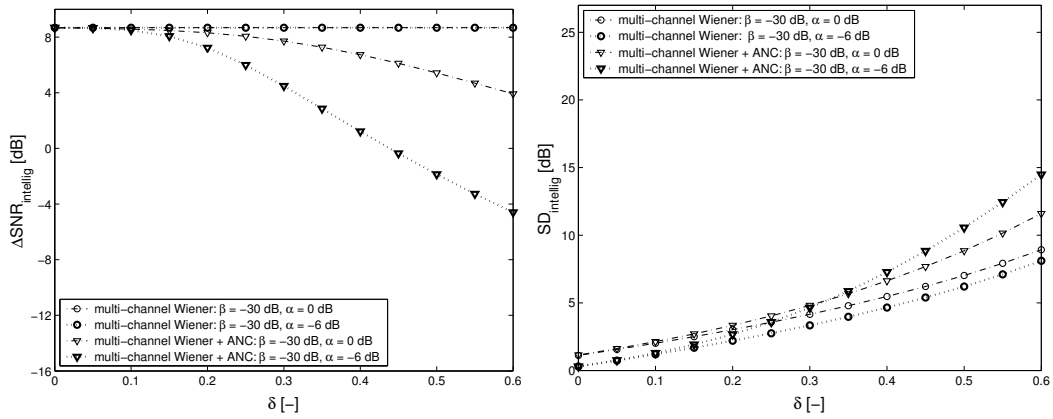
The multi-channel Wiener filter acts as if the input PSD of the speech signal and hence the input SNR $1/\alpha(f)$ is decreased by a factor $1-\delta$. As a consequence, the filter will pay more attention to noise reduction than to speech distortion, so that eventually the speech signal is attenuated more than in case of a perfect speech detection. For an excessive error rate $(\delta \times 100\%) = 50\%$, the effect corresponds to a decrease in input SNR of only 3 dB, so that the effect of noise detected as speech + noise is expected to be limited.

Figure 5(a) and (b) plot the performance as a function of δ for two values of internal-to-external noise ratio $\beta(f)$ (i.e. $\beta(f) = -30$ dB and $\beta(f) = -20$ dB) and external noise-to-signal ratio $\alpha(f)$ (i.e. $\alpha(f) = 0$ dB and $\alpha(f) = -6$ dB). In free field, the improvement $\Delta\text{SNR}_{\text{intellig}}$ is not affected by the erroneous speech detection. The speech distortion $\text{SD}_{\text{intellig}}$ slightly grows with increasing δ. The increase in distortion as a function of δ is more significant for scenarios where the output SNR is limited, e.g. due to low input SNR $1/\alpha(f)$, difficult noise scenarios and large $\beta(f)$ at low fd, where f is the frequency and d the microphone interspacing. **The effect of noise being detected as speech + noise is less disadvantageous than the effect of speech + noise being detected as noise only.** Note that in the latter case, the speech distortion would increase with 1.9 dB for $\delta = 0.2$ and 4.4 dB for $\delta = 0.4$ (see also in Figure 6). Hence, if the detection of noise only in a sample is uncertain, the sample should not be used in the computation of the multi-channel Wiener filter.



(a) (b)

Fig. 5. Theoretically optimal performance ($\Delta\text{SNR}_{\text{intellig}}$ (a) and $\text{SD}_{\text{intellig}}$ (b)) of the multi-channel Wiener filter in a diffuse noise field and in case of a small microphone mismatch, when $(\delta \times 100)$ percent of the speech and noise samples used to compute the multi-channel Wiener filter actually contain noise only.



(a) (b)

Fig. 6. Theoretical improvement in $\mathrm{SNR}_{\mathrm{intellig}}$ (a) and spectral distortion $\mathrm{SD}_{\mathrm{intellig}}$(b) of the multi-channel Wiener filter with(out) ANC in a diffuse noise field with $\beta(f) = -30$ dB and in case of a small microphone mismatch, when $(\delta \times 100)$ percent of the noise samples used to compute the multi-channel Wiener filter with(out) ANC actually contain speech.

C. Multi-channel Wiener filtering with ANC postprocessing

The influence of an erroneous speech detection on the optimal performance of the multi-channel Wiener filter with ANC postprocessing stage can be investigated by replacing the fixed matrices $\mathbf{B}(f)$ and $\mathbf{A}(f)$ in (40) and (41) with $\mathbf{W}(f)$ and $(\mathbf{e}_1 - \mathbf{W}(f))^T$, respectively. Hence, when speech + noise is detected as noise only, the performance of the ANC stage degrades with increasing input SNR and decreasing quality of speech and noise references. The latter depends on the performance of the multi-channel Wiener filter. In contrast to the fixed stages of the GSC, the multi-channel Wiener filter compensates for small model errors such as microphone mismatch, reverberation, speaker variations, ... and thus creates better speech and noise references when model errors occur. An erroneous speech detection will however also deteriorate their quality. Substituting the superscript n by the superscript s in (34) and (36), it follows that the PSD $P_{Y_1}^s(f)$ of the speech in the noise reference equals

$$P_{Y_1}^s(f) = P_X^s(f) + P_{Y_0}^s(f) - 2P_X^s(f)\,\mathrm{Re}\{\sum_{k=1}^{M} W_k(f)\Gamma_{k1}^s(f)\}. \tag{54}$$

- When speech + noise is detected as noise only, $\mathbf{W}(f)$ is scaled with $(1 - \delta\kappa(f))$ (see (48)) so that the power $P_{Y_0}^s$ of the speech leakage in the noise references will increase with increasing $\delta\kappa(f)$.
- When on the other hand noise only is detected ad speech + noise, the quality of the speech reference $Y_0(f)$ and hence also the noise reference $Y_1(f)$ is only slightly degraded (see Figure 5), so that the performance of the ANC that is adapted during periods of noise only, is hardly affected.

Figure 6(a) and Figure 6(b) plot the improvement $\Delta\mathrm{SNR}_{\mathrm{intellig}}$ (in dB) and the spectral distortion $\mathrm{SD}_{\mathrm{intellig}}$ (in dB), respectively, of the multi-channel Wiener filtering with and without ANC as a function of δ for $\alpha(f) = 0$ dB and -6 dB. In the simulations, $\kappa(f)$ is set to 1, illustrating the average effect. Again, the second and the third

microphone have a gain mismatch of 1 dB and -1 dB with respect to the first microphone. As mathematically derived in Section IV-B, the multi-channel Wiener filtering additionally distorts the speech by $10\log_{10}(1-\delta\kappa(f))^2$ but conserves the improvement $\Delta\text{SNR}_{\text{intellig}}$ if speech + noise is detected as noise only. The higher the output SNR, the smaller the distortion that is caused by the multi-channel Wiener filter. In contrast to the GSC, the additional distortion that is due to erroneous speech detection (i.e. $10\log(1-\delta)^2$), is independent of the input SNR. **When using an ANC postprocessing stage, speech detection errors additionally degrade both the noise reduction performance, $\Delta\text{SNR}_{\text{intellig}}$ and the amount of speech distortion $\text{SD}_{\text{intellig}}$. For $\delta < 0.2$ this extra degradation is limited.** Like the GSC, the performance decreases with increasing input SNR $1/(\alpha(f)+\gamma(f))$. In this example, the performance is less affected than for the GSC thanks to the better speech and noise references. Compared to the GSC with noise sensitivity constraint, the speech signal is attenuated more, especially at large error rates δ. For error rates $\delta \times 100\%$ up to 20% the distortion is limited, while considerably more noise is reduced.

D. Conclusion

In short, the noise reduction performance $\Delta\text{SNR}_{\text{intellig}}$ of the **multi-channel Wiener filtering** is not degraded by speech detection errors, but the desired signal gets distorted when speech + noise samples are detected as noise only. The average distortion equals $|20\log_{10}(1-\delta)|$ dB, independently of the noise scenario and the input SNR. For reasonable error rates $\delta \times 100\%$ up to 20%, this distortion is limited. The **multi-channel Wiener filtering with ANC postprocessing** is stronger affected by speech detection errors: the noise reduction degrades and the distortion grows with increasing error rate δ and increasing input SNR. Already in the presence of small model errors, the performance of the **GSC** is much more deteriorated than the performance of the multi-channel Wiener filtering technique. A constraint on the noise sensitivity of the GSC significantly reduces the influence of speech detection errors. However, compared to the multi-channel Wiener filter without ANC and - for a low error rate δ, also the multi-channel Wiener filter with ANC- less noise is reduced.

V. Validation through experimental results

The theoretical 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 based on real recordings in a reverberant room with reverberation time $T_{60\,\text{dB}} \approx 700$ ms for a speech weighted noise, and when finite filters are used.

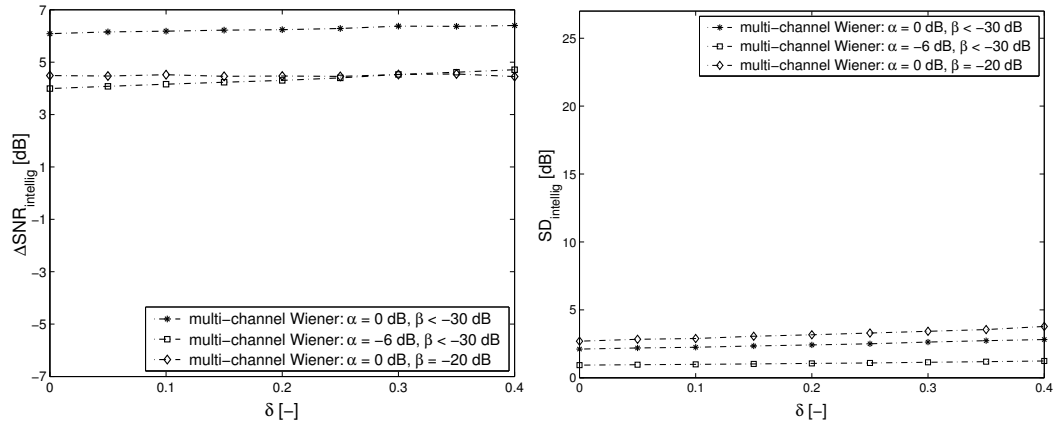
The results are illustrated for a diffuse noise field. A uniform, linear endfire array with 4 omnidirectional microphones (Knowles EM-4368) with interspacing $d = 0.02$ m has been mounted on a dummy head. The desired signal source is positioned at a distance of 1 meter in front of the head. To get an idea of the average performance based on directivity only, stationary speech and noise signals are used with the same, average long-term power spectral density. The signals can be found on [31]. The speech and noise signal are uncorrelated. The external noise signal has a level of about 70 dB SPL at the center of the head. Since the microphones have an internal noise level of about $25-28$ dB SPL and the external noise signal has the highest energy at low frequencies, the internal-to-external

noise level $\beta(f)$ is expected to be smaller than -30 dB at low frequencies. The level of the speech signal is adjusted so that the input SNR at the first microphone equals 0 dB and 6 dB, respectively. During the first 5 seconds only noise is present, during the last 5 seconds the speech and noise signal are both present.

In the experiments, the multi-channel Wiener filter with and without ANC postprocessing is implemented in subbands using an oversampled, uniform, DFT modulated filterbank with 16 subbands and a downsampling factor of 12 [17], [18]. 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. To improve intelligibility, the microphone signals are pre-emphasized in the GSC prior to processing [33]. At the GSC output, de-emphasis is performed. In the subband GSVD algorithm, the subband implementation intrinsically performs a pre-whitening operation [17]. The filters of the GSC and the multi-channel Wiener filtering technique are sufficiently long so that a further increase in filter length does not result in a significant increase in performance for the given noise scenario. The ANC of the GSC has 180 taps per noise reference. The subband, multi-channel Wiener filter with and without ANC uses 8 taps per subband and per channel for the multi-channel Wiener filter and 15 taps per subband for the ANC.

The effect of noise being detected as speech + noise and speech + noise being detected as noise only are both investigated. Since the ANC filters are adapted during periods of noise only, the former effect is studied only for the multi-channel Wiener filter. Again, a small gain mismatch of 1 dB and -1 dB has been applied to the second and third microphone w.r.t. the first microphone.

A. Noise detected as speech + noise

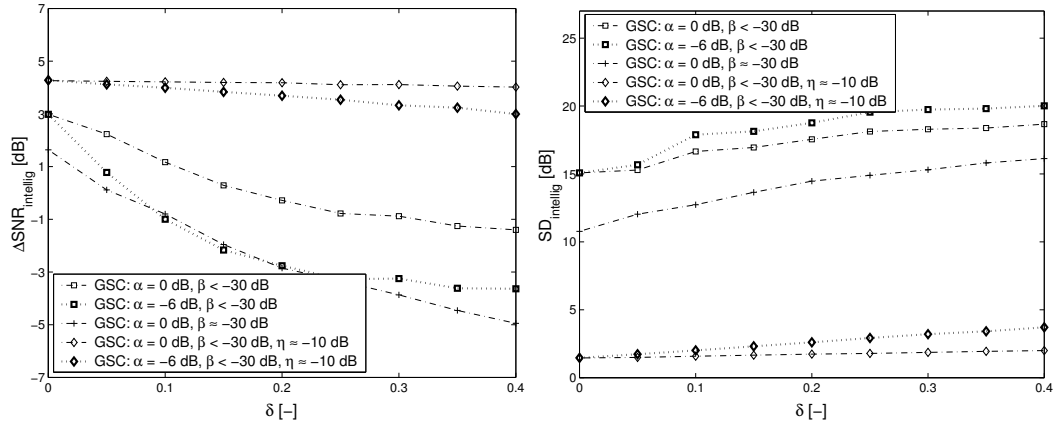


(a) (b)

Fig. 7. Effect of noise samples detected as speech + noise on the performance (i.e. $\Delta\text{SNR}_{\text{intellig}}$ (a) and $\text{SD}_{\text{intellig}}$ (b)) of the multi-channel Wiener filter in a diffuse noise field with internal-to-external noise ratio $\beta(f)$ and input SNR $1/\alpha(f)$; Experimental results.

Figure 7(a) and (b) study the effect of noise only samples detected as speech + noise on the performance of the multi-channel Wiener filter. The performance measures $\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$ are depicted as a function of δ with $(\delta \times 100)\%$ the percentage of samples detected as speech + noise that actually contain noise only. Two input noise-to-signal ratios $\alpha(f)$ are considered: $\alpha(f) = 0$ dB and $\alpha(f) = -6$ dB. The influence of a higher internal-to-external noise ratio $\beta(f) \approx -20$ dB is shown too. The behavior of the curves is quite well predicted by Figure 5: an increased error rate δ only leads to a small increase in distortion $\text{SD}_{\text{intellig}}$. The increase in distortion $\text{SD}_{\text{intellig}}$ grows with increasing $\beta(f)$ and increasing $\alpha(f)$. The noise reduction performance $\Delta\text{SNR}_{\text{intellig}}$ is not significantly affected by the speech detection errors. In contrast to the results of Figure 5, the improvement $\Delta\text{SNR}_{\text{intellig}}$ drops for decreasing α. This difference is due to the presence of reverberation in the speech signal. The Wiener filter trades off reduction of diffuse noise and cancellation of the reverberant part in the speech signal. The filter will pay more attention to noise reduction for large $\alpha(f)$. The improvement $\Delta\text{SNR}_{\text{intellig}}$ also slightly increases as a function of δ since a larger δ corresponds to an increase of $\alpha(f)$ (see (51)).

B. Speech + noise detected as noise only

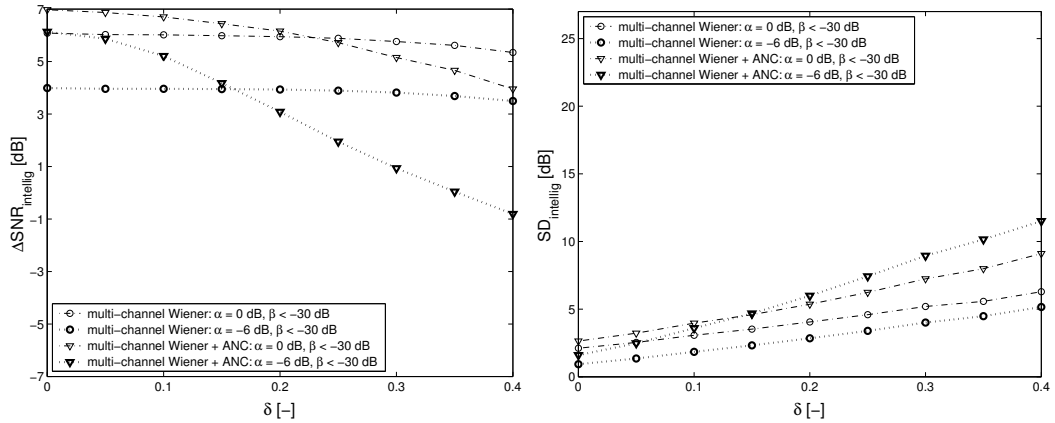


(a) (b)

Fig. 8. Performance of the GSC without and with noise sensitivity constraint (noise injection level $\eta(f) \approx -10$ dB) in a diffuse noise field with input SNR $= 1/\alpha(f)$ and in case of a small microphone mismatch, when $(\delta \times 100)\%$ of the assumed noise only samples actually contain speech; Experimental results. (a) Improvement $\Delta\text{SNR}_{\text{intellig}}$; (b) Spectral distortion $\text{SD}_{\text{intellig}}$.

1) GSC: Figure 8(a) and Figure 8(b) plot the improvement $\Delta\text{SNR}_{\text{intellig}}$ and the spectral distortion $\text{SD}_{\text{intellig}}$, respectively, of the GSC in case $(\delta \times 100)$ percent of the noise samples used to compute the filters, actually contain speech. Two input signal-to-noise ratios $1/\alpha(f)$ are depicted: $\alpha = 0$ dB and $\alpha = 6$ dB. As has been predicted by Figure 4 in Section IV, the improvement $\Delta\text{SNR}_{\text{intellig}}$ of the GSC is strongly affected by erroneous speech detection, especially for high input SNR. Compared to Figure 4, the distortion $\text{SD}_{\text{intellig}}$ at $\delta = 0$ is worse due to a lower internal-to-external noise ratio $\beta(f) < -30$ dB at low frequencies and the presence of reverberation.

Indeed, the experimentally found performance of the GSC with $\beta(f)$ increased to about -30 dB is more similar to the theoretical predicted one in Figure 4. In Figure 8, also the performance of the more robust GSC with a noise sensitivity constraint is depicted for a noise injection level $\eta(f) \approx -10$ dB. The drastic impact of speech detection errors on the GSC is then tremendously reduced.



(a) (b)

Fig. 9. Performance of the multi-channel Wiener filter with(out) ANC in a diffuse noise field with input SNR $= 1/\alpha(f)$ and in case of a small microphone mismatch, when $(\delta \times 100)\%$ of the assumed noise only samples actually contain speech; Experimental results. (a) Improvement $\Delta\text{SNR}_{\text{intellig}}$; (b) Spectral distortion $\text{SD}_{\text{intellig}}$.

2) Multi-channel Wiener filter with(out) ANC postprocessing: Figure 9(a) and (b) illustrate the effect of an erroneous speech detection on the performance of the multi-channel Wiener filter with(out) ANC postprocessing. The improvement in $\text{SNR}_{\text{intellig}}$ of the multi-channel Wiener filter is hardly affected by speech detection errors, while the average speech distortion $\text{SD}_{\text{intellig}}$ that is due to erroneous speech detection increases with about $|10\log_{10}(1-\delta)^2|$, independently of the input SNR. In contrast to Figure 6, the improvement $\Delta\text{SNR}_{\text{intellig}}$ depends again on α due to reverberation: a lower α results in less noise reduction and less speech distortion.

As also shown analytically, the noise reduction performance of the multi-channel Wiener filter with ANC post-processing depends -in contrast to the multi-channel Wiener filter without postprocessing- on δ and is especially affected in case of high SNR. At a low error rate δ, the postprocessing stage offers an increase in $\Delta\text{SNR}_{\text{intellig}}$ at the expense of a small increase in speech distortion. In contrast to Section III, a finite filter length is used in the first Wiener filtering stage and a reverberant speech signal is estimated. The ANC stage additionally attenuates part of the reverberant speech portion.

Already in the presence of small model errors, the multi-channel Wiener filter with(out) ANC postprocessing is less sensitive to an erroneous speech detection than the GSC without noise sensitivity constraint. Compared to the more robust GSC with noise sensitivity constraint, the multi-channel Wiener filter and for small δ, also the multi-channel Wiener filter with ANC, reduce more noise, especially at low input SNR, while the distortion is limited

for $\delta < 0.2$. In addition, in the presence of larger model errors the performance of the GSC with noise sensitivity constraint will further degrade. The multi-channel Wiener filter on the other hand, is insensitive to model errors [12].

VI. Conclusions

We have analyzed the sensitivity to speech detection errors of the GSC and the recently developed multi-channel Wiener filtering technique both theoretically and experimentally. Since the latter does not rely on any a priori information about the signal model, a reliable speech detection is expected to be crucial to achieve their potentially better performance. In Section II and Section III formulas for the optimal performance of both algorithms have been derived.

In Section IV, the formulas have been extended to assess the average effect of speech detection errors on the performance. In the simulations, we focussed on small-sized arrays as e.g. used in hearing aid applications. Theoretical analysis showed that the noise reduction performance of the multi-channel Wiener filter (without ANC) is - in contrast to the GSC- not degraded by speech detection errors, but the speech gets distorted. The effect of noise detected as speech + noise is negligible. When speech + noise samples are detected as noise only, the speech signal gets on the average attenuated, hence distorted, by a factor $1/(1-\delta)$, where $\delta \times 100\%$ equals the percentage of samples that are detected as noise only but actually contain speech. This average attenuation is independent of the noise scenario and the input SNR and for error rates up to 20%, this distortion is limited. The multi-channel Wiener filter with ANC postprocessing is affected more by speech detection errors, especially for $\delta \times 100\% > 20\%$. The noise reduction deteriorates and the distortion grows with increasing error rate δ and increasing input SNR. Already in the presence of small model errors, the multi-channel Wiener filter with and without ANC postprocessing stage are found to be less sensitive to an erroneous speech detection than the GSC, despite their only dependence on a good performing speech detection. A constraint on the noise sensitivity of the GSC is not only essential to increase robustness to small signal model errors when small-sized arrays are used, but also to limit the drastic impact of speech detection errors. For a certain constraint and depending on the size of the model error, the robust GSC with noise sensitivity constraint distorts the speech signal less than the multi-channel Wiener filter with(out) ANC postprocessing stage, especially at large δ. However, compared to the multi-channel Wiener filter and - for a low error rate $\delta \times 100\%$- also the multi-channel Wiener filter with ANC, less noise is reduced. In short, the Wiener filtering technique preserves their potential benefit over the GSC for a reasonable speech detection error rate of 20% or less, even when the GSC is supplied with a noise sensitivity constraint.

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 theoretical results have been validated for real recordings in an office room and finite filter lengths. The experimental results confirm the theory.

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 the Belgian State, Prime Minister's Office - Federal Office for Scientific, Technical and Cultural Affairs - 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 was partially sponsored by Cochlear. The scientific responsibility is assumed by its authors.

REFERENCES

[1] L. J. Griffi ths, C. W. Jim, An alternative approach to linearly constrained adaptive beamforming, IEEE Trans. Antennas Propag. 30 (1) (1982) 27–34.

[2] B. D. Van Veen, K. M. Buckley, Beamforming: A Versatile Approach to Spatial Filtering, IEEE ASSP Magazine 5 (2) (1988) 4–24.

[3] K. M. Buckley, Broad-Band Beamforming and the Generalized Sidelobe Canceller, IEEE Trans. Acoust., Speech, and Signal Processing 34 (5) (1986) 1322–1323.

[4] P. M. Peterson, A.-M. Wie, W. M. Rabinowitz, P. M. Zurek, Robustness of an adaptive beamforming method for hearing aids, Acta Otolaryngol. Suppl. 469 (1990) 85–90.

[5] J. E. Greenberg, P. M. Zurek, Evaluation of an Adaptive Beamforming Method for Hearing Aids, J. Acoust. Soc. Amer. 91 (3) (1992) 1662–1676.

[6] 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), Vol. 2, Albuquerque, 1990, pp. 833–836.

[7] O. Hoshuyama, A. Sugiyama, A. Hirano, A Robust Adaptive Beamformer for Microphone Arrays with a Blocking Matrix Using Constrained Adaptive Filters, IEEE Trans. Signal Processing 47 (1999) 2677–2683.

[8] H. Cox, R. M. Zeskind, M. M. Owen, Robust Adaptive Beamforming, IEEE Trans. Acoust., Speech, and Signal Processing 35 (10) (1987) 1365–1376.

[9] N. K. Jablon, Adaptive beamforming with the generalized sidelobe can in the presence of array imperfections, IEEE Trans. Antennas Propag. 34 (1986) 996–1012.

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

[11] A. Spriet, M. Moonen, J. Wouters, Robustness analysis of GSVD based optimal Filtering and generalized Sidelobe Canceller for Hearing Aid Applications, in: Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 2001, pp. 31–34.

[12] A. Spriet, M. Moonen, J. Wouters, Robustness Analysis of Multi-channel Wiener Filtering and Generalized Sidelobe Cancellation for Multi-microphone Noise Reduction in Hearing Aid Applications, Tech. Rep. ESAT-SISTA/TR 02-81, ESAT-SCD/SISTA, KU Leuven (Belgium), available at ftp://ftp.esat.kuleuven.ac.be/sista/spriet/reports/02-81.pdf (Sep. 2002).

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

[14] S. Doclo, M. Moonen, GSVD-Based Optimal Filtering for Single and Multimicrophone Speech Enhancement, IEEE Trans. Signal Processing 50 (9) (2002) 2230–2244.

[15] S. Doclo, E. De Clippel, M. Moonen, Multi-microphone noise reduction using GSVD-based optimal fi ltering with ANC postprocessing stage, in: Proc. of DSP2000 Workshop, Hunt TX, USA, 2000.

[16] G. Rombouts, M. Moonen, QRD-based optimal fi ltering for acoustic noise reduction, in: Proc. European Signal Processing Conf. (EUSIPCO), Vol. 3, Toulouse, France, 2002, pp. 301–304.

[17] A. Spriet, M. Moonen, J. Wouters, A multi-channel subband generalized singular value decomposition approach to speech enhancement, European Transactions on Telecommunications 13 (2) (2002) 149–158.

[18] A. Spriet, M. Moonen, 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, 2001, pp. 187–191.

[19] 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, available as Res. Lab. Elect. Techn. Rept. 541 (1989).

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

[21] J. Bitzer, K. U. Simmer, 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), Vol. 5, Phoenix, Arizona, USA, 1999, pp. 2965–2968.

[22] J. Bitzer, K. U. Simmer, K.-D. Kammeyer, Multi-microphone noise reduction techniques as front-end devices for speech recognition, Speech Communication 34 (2001) 3–12.

[23] S. Gannot, B. D., W. E., Theoretical analysis of the general transfer function GSC, in: Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC), Darmstadt, Germany, 2001, pp. 103–106.

[24] S. E. Nordholm, Y. H. Leung, Performance limits of the broadband generalized sidelobe cancelling structure in an isotropic noise fi eld, J. Acoust. Soc. Amer. 107 (2) (2000) 1057–1060.

[25] C. Marro, Y. Mahieux, K. U. Summer, Analysis of Noise Reduction and Dereverberation Techniques Based on Microphone Arrays with Postfi ltering, IEEE Trans. Speech, Audio Processing 6 (3) (1998) 240–259.

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

[27] J. Bitzer, K. U. Simmer, 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, 1999, pp. 7–10.

[28] M. W. Hoffman, K. M. Buckley, Robust time-domain processing of broadband acoustic data, IEEE Trans. Speech, Audio Processing 3 (1995) 193–203.

[29] J. E. Greenberg, P. M. Peterson, P. M. Zurek, Intelligibility-weighted measures of speech-to-interference ratio and speech system performance, J. Acoust. Soc. Amer. 94 (5) (1993) 3009–3010.

[30] Acoustical Society of America, ANSI S3.5-1997 American National Standard Methods for Calculation of the Speech Intelligibility Index (Jun. 1997).

[31] International Collegium of Rehabilitative Audiology, Noise Signals ICRA Compact Disc (Ver 0.3 1997).

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

[33] M. J. Link, K. M. Buckley, Prewhitening for intelligibility gain in hearing aid arrays, J. Acoust. Soc. Amer. 93 (4) (1993) 2139–2140.