Katholieke Universiteit Leuven

Departement Elektrotechniek ESAT-SISTA/TR 2002-163b

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]

Published in Proc. of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003), Hong Kong, April 6-10 2003, pp. 501–504

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/sista/spriet/reports/02-163b.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

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

[1]K.U. Leuven, ESAT/SCD-SISTA
Kasteelpark Arenberg 10, 3001 Leuven, Belgium
{spriet,moonen}@esat.kuleuven.ac.be

[2]K.U. Leuven - Lab. Exp. ORL
Kapucijnenvoer 33, 3000 Leuven, Belgium
jan.wouters@uz.kuleuven.ac.be

ABSTRACT

The noise reduction performance of the Generalized Sidelobe Canceller (GSC) depends on the validity of a priori assumptions about the signal model, whereas the recently developed Multi-channel Wiener Filter (MWF) techniques do not, hence, their potential benefit. However, both techniques 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 MWF both theoretically and experimentally. It is shown that the MWF 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 robustness constraint.

1. INTRODUCTION

In speech communication applications, such as handsfree telephony and hearing aids, background noise reduces the intelligibility of the desired speech seriously, making a noise reduction algorithm necessary. Multi-microphone systems exploit spatial information in addition to temporal and spectral information of the desired and noise signal and are thus preferred to single microphone approaches.

Recently, Multi-channel Wiener Filtering (MWF) techniques have been proposed that provide a Minimum Mean Square Error (MMSE) estimate of the desired signal portion in one of the received microphone signals [1, 2, 3]. In contrast to the GSC [4], they do not make any a priori assumptions about the signal model (such as microphone characteristics, speaker and microphone positions, reverberation, ...) so that no robustness constraint [5, 6] is needed to guarantee its performance when applied in small-sized arrays [7]. Especially in complicated noise scenarios, the MWF outperforms the GSC with robustness constraint [7].

The MWF is uniquely based on estimates of the second order statistics of the speech and the noise. Both the MWF and the GSC need a robust speech detection to determine periods of noise only. Since the MWF does not require any other a priori information, the reliance on the 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 MWF both theoretically and experimentally and compare it with the GSC (with and without robustness constraint). In the simulations and experiments, we focus on the harsh case of small-sized arrays as used in hearing aid applications.

Notation

In the sequel, signals and filters 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 $P_{XY}(f) = \mathcal{E}\{X(f)Y^*(f)\}$. Where needed, the superscripts s and n are used to refer to the contribution of the speech and noise signal only. The noise signal consists of external noise (superscript e) and internal noise (superscript i) e.g. sensor noise, 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 $P_X(f)$ to be the same at each microphone.

2. MULTI-CHANNEL WIENER FILTER (MWF)

2.1. Concept

The MWF $\mathbf{W}(f) \in \mathcal{C}^{M\times 1}$ (with $W_k(f)$ the k-th entry of $\mathbf{W}(f)$) provides a MMSE estimate of the (unknown) speech signal $X_k^s(f)$ at the k-th (e.g. first) microphone[1], which is computed as $Y_0(f) = \mathbf{W}^T(f)\mathbf{X}(f)$, i.e. the sum of the M filtered microphone signals, where[2] [7]

$$\mathbf{X}(f) = \begin{bmatrix} X_1(f) & X_2(f) & \cdots & X_M(f) \end{bmatrix}^T. \quad (1)$$

$$\mathbf{W}(f) = \begin{bmatrix} P_X & P_{X_2X_1} & \cdots & P_{X_MX_1} \\ P_{X_1X_2} & P_X & & P_{X_MX_2} \\ \vdots & & \ddots & \vdots \\ P_{X_1X_M} & \cdots & \cdots & P_X \end{bmatrix}^{-1} \begin{bmatrix} P_X^s \\ P_{X_1X_2}^s \\ \vdots \\ P_{X_1X_M}^s \end{bmatrix}. \quad (2)$$

Assuming that the speech and noise signals are uncorrelated, $P_{X_kX_l}^s(f)$ is estimated as

$$P_{X_kX_l}^s(f) = P_{X_kX_l}(f) - P_{X_kX_l}^n(f), \quad (3)$$

with $P_{X_kX_l}(f)$ and $P_{X_kX_l}^n(f)$ estimated during periods of *speech + noise* and periods of *noise only*, respectively. The second order statistics of the noise are assumed to be sufficiently stationary so that they can be estimated during periods of *noise only*. Like for

*Ann Spriet is a Research Assistant supported by F.W.O. This research was carried out at the ESAT laboratory and Lab. Exp. ORL of the K. U. Leuven, in the frame of IUAP P5/22 (2002-2007) ('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, FWO Project 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.

[1]In the sequel, we assume without loss of generality that the first microphone signal is estimated.

[2]The parameter f is often omitted for the sake of conciseness.

the GSC, a robust speech detection is thus needed.

In [1], the MWF is implemented in the time-domain by means of a Generalized Singular Value Decomposition (GSVD) of an input and noise data matrix. Cheaper alternatives based on a QR Decomposition and/or a subband implementation have been proposed in [2, 3]. In contrast to the GSC, the MWF does not make any a priori assumptions about the signal model so it is more robust to small signal model errors [1, 7].

2.2. Theoretical performance

2.2.1. Power transfer functions $G^n(f)$ and $G^s(f)$

The theoretical performance (i.e. assuming infinite filter lengths) of the MWF $\mathbf{W}(f)$, i.e. the Power Transfer Function (PTF) of the noise signal $G^n(f) = P^n_{Y_0}(f)/P^n_X(f)$ and the desired speech signal $G^s(f) = P^s_{Y_0}(f)/P^s_X(f)$, can be expressed as a function of the so-called *complex coherence* $\Gamma_{kl}(f)$ between the k-th and l-th microphone, defined as:

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

Using definition (4), $\mathbf{W}(f)$ can be rewritten as [7]:

$$\frac{P^s_X(f)}{P^s_X(f)(1+\alpha(f))} \begin{bmatrix} 1 & \Gamma_{21} & \cdots & \Gamma_{M1} \\ \Gamma_{12} & 1 & & \Gamma_{M2} \\ \vdots & & \ddots & \vdots \\ \Gamma_{1M} & \cdots & \cdots & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \Gamma^s_{12} \\ \vdots \\ \Gamma^s_{1M} \end{bmatrix} \tag{5}$$

where

$$\Gamma_{kl}(f) = \frac{\alpha(f)\Gamma^n_{kl} + \Gamma^s_{kl}}{1+\alpha(f)}, \ \alpha(f) = \frac{P^n_X(f)}{P^s_X(f)}, \tag{6}$$

with $1/\alpha(f)$ the input SNR. The PTFs $G^n(f)$ and $G^s(f)$ can be easily expressed as a function of Γ_{kl}, Γ^n_{kl} and Γ^s_{kl} [7]:

$$G^{n|s}(f) = \sum_{k=1}^{M} |W_k|^2 + \sum_{k=1}^{M} \sum_{l=k+1}^{M} 2.Re\{W_k W_l^* \Gamma^{n|s}_{kl}\}, \tag{7}$$

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

Remark: Equations (5)-(7) assume that the microphones are perfectly matched. Suppose that the k-th microphone has a gain mismatch $\Delta\Upsilon(f)$ and a phase mismatch $\Delta\Phi(f)$ (in degrees) w.r.t. microphone l. The coherence functions $\Gamma^{s,n}_{kl}(f)$ in (5)-(7) should then be replaced by $\tilde{\Gamma}^{n|s}_{kl}(f)$

$$\tilde{\Gamma}^{n|s}_{kl}(f) = \begin{cases} \Delta\Upsilon(f) e^{j\Delta\Phi(f)\frac{\pi}{180}} \Gamma^{n|s}_{kl}(f) & \text{for } k \neq l; \\ |\Delta\Upsilon(f)|^2 \Gamma^{n|s}_{kl}(f) & \text{for } k = l. \end{cases} \tag{8}$$

2.2.2. Intelligibility weighted performance measures

To assess the effect of the obtained PTFs $G^s(f)$ and $G^n(f)$ on intelligibility - which is the major goal of a noise reduction algorithm in applications as hearing aids -the improvement in intelligibility weighted Signal-to-Noise Ratio (SNR) has been proposed [8]:

$$\Delta\text{SNR}_{\text{intellig}} = \sum_i I_i (\text{SNR}_{i,\text{out}} - \text{SNR}_{i,\text{in}}), \tag{9}$$

with:

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

$$\text{SNR}_{i,\text{in}} = 10\log_{10}\left(\int_{2^{-\frac{1}{6}} f_i^c}^{2^{\frac{1}{6}} f_i^c} P^s_X(f) df \Big/ \int_{2^{-\frac{1}{6}} f_i^c}^{2^{\frac{1}{6}} f_i^c} P^n_X(f) df \right), \tag{11}$$

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 [9], and where $\text{SNR}_{i,\text{out}}$ and $\text{SNR}_{i,in}$ is the output and input SNR (in dB) in the i-th one-third octave band, respectively.

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

$$\text{SD}_{\text{intellig}} = \sum_i I_i \text{SD}_i, \tag{12}$$

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

$$\text{SD}_i = \int_{2^{-\frac{1}{6}} f_i^c}^{2^{\frac{1}{6}} f_i^c} |10\log_{10} G^s(f)| \, df \Big/ \left[\left(2^{\frac{1}{6}} - 2^{-\frac{1}{6}} \right) f_i^c \right]. \tag{13}$$

3. SENSITIVITY TO SPEECH DETECTION ERRORS

Based on the formulas derived in Section 2.2, we can predict the average effect of speech detection errors on the MWF.

3.1. Multi-channel Wiener filtering

3.1.1. Speech + noise being erroneously detected as noise only

In case of a perfect speech detection, (3) applies and the multi-channel Wiener filter is given by (5). Suppose now that $(\delta \times 100)\%$ of the *noise only* samples used to estimate $P^n_{X_k X_l}(f)$ actually contain speech. Assuming that the average PSDs $P^s_X(f)$ of the $(\delta \times 100)\%$ wrongly detected samples and the correctly detected samples are the same[3], the estimated cross-PSD $\hat{P}^n_{X_k X_l}(f)$ equals

$$\begin{aligned} \hat{P}^n_{X_k X_l}(f) &= P^n_X(f)\Gamma^n_{kl} + \delta P^s_X(f)\Gamma^s_{kl} \\ &= P^s_X(f)\left(\alpha(f)\Gamma^n_{kl} + \delta\Gamma^s_{kl}\right), \end{aligned} \tag{14}$$

with $\alpha(f)$ defined in (6). Using (3), the estimated cross-PSD $\hat{P}^s_{X_k X_l}(f)$ becomes

$$\hat{P}^s_{X_k X_l}(f) = P^s_X(f)(1-\delta)\Gamma^s_{kl}(f). \tag{15}$$

It can be shown with (2) and (15) that $\mathbf{W}(f)$ changes into

$$\boxed{\frac{P^s_X(f)(1-\delta)}{P^s_X(f) + P^n_X(f)} \begin{bmatrix} 1 & \Gamma_{21} & \cdots & \Gamma_{M1} \\ \Gamma_{12} & 1 & & \Gamma_{M2} \\ \vdots & & \ddots & \vdots \\ \Gamma_{1M} & \cdots & \cdots & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \Gamma^s_{12} \\ \vdots \\ \Gamma^s_{1M} \end{bmatrix}}, \tag{16}$$

with Γ_{kl} and Γ^s_{kl} the coherence functions in case of a perfect speech detection.

Hence, the PTFs $G^s(f)$ and $G^n(f)$ are both scaled by $(1-\delta)^2$, independently of the input SNR and the noise scenario. In practice, the average PSD $P^s_X(f)$ of the correctly and wrongly detected samples will differ, resulting in a frequency dependent $\delta(f)$ and attenuation/distortion (see also footnote (3)), with an average effect as derived above. The SNR improvement $G^s(f)/G^n(f)$ at frequency f remains unchanged.

3.1.2. Noise being erroneously detected as speech + noise

A similar reasoning can be applied 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* [10]. The estimate $\hat{P}_{X_k X_l}(f)$ and as a consequence the estimate $\hat{P}^s_{X_k X_l}(f)$ then become[4]

[3] In practice, the average PSDs $P^s_X(f)$ can be different. This corresponds to replacing δ by a frequency dependent $\delta(f) = \delta.F(f)$, where $F(f)$ is determined by the ratio of the average PSDs (see [10]).

[4] Here, δ is f-independent, since the 2nd order statistics of the noise are assumed to be stationary.

$$\hat{P}_{X_k X_l}(f) = P^n_{X_k X_l}(f) + (1-\delta)P^s_{X_k X_l}(f) \quad (17)$$

$$\hat{P}^s_{X_k X_l}(f) = P^s_X(f)(1-\delta)\Gamma^s_{kl}(f), \quad (18)$$

so that $\mathbf{W}(f)$ equals

$$\boxed{\frac{P^s_X(f)}{P^s_X(f)(1+\tilde{\alpha}(f))}\begin{bmatrix} 1 & \tilde{\Gamma}_{21} & \cdots & \tilde{\Gamma}_{M1} \\ \tilde{\Gamma}_{12} & 1 & & \tilde{\Gamma}_{M2} \\ \vdots & & \ddots & \vdots \\ \tilde{\Gamma}_{1M} & \cdots & \cdots & 1 \end{bmatrix}^{-1}\begin{bmatrix} 1 \\ \Gamma^s_{12} \\ \vdots \\ \Gamma^s_{1M} \end{bmatrix}}, \quad (19)$$

where

$$\tilde{\Gamma}_{kl}(f) = \frac{\tilde{\alpha}(f)\Gamma^n_{kl} + \Gamma^s_{kl}}{1+\tilde{\alpha}(f)}, \ \tilde{\alpha}(f) = \frac{\alpha(f)}{1-\delta}. \quad (20)$$

The MWF acts as if the input SNR $1/\alpha(f)$ is reduced by a factor $1/(1-\delta)$. As a consequence, the filter will pay more attention to noise reduction than to speech distortion, so that 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 found to be limited [10]. Hence, the MWF is especially affected when speech + noise is detected as noise only.

3.2. Comparison with GSC

In this section, we compare the theoretical performance of the MWF in case of an erroneous speech detection to the performance of the GSC with(out) robustness constraint. The results are illustrated for a small-sized uniform 4-microphone endfire array with microphone interspacing $d = 2$ cm. The desired signal is assumed to be in front of the microphone array at $\theta = 0^\circ$.

3.2.1. Theoretical performance GSC

The GSC [4] consists of a fixed beamformer $\mathbf{A} \in \mathcal{C}^{M\times 1}$,which creates a speech reference $Y_0(f)$, a blocking matrix $\mathbf{B} \in \mathcal{C}^{N\times M}$, which creates N noise references $Y_i(f)$, $i = 1, ..., N$, and an Adaptive Noise Canceller (ANC). A delay-and-sum beamformer is used, i.e. $A_i(f) = \frac{1}{M}e^{j2\pi f\frac{d_i-d_1}{c}}$, and the matrix $\mathbf{B}$ is set to

$$\mathbf{B} = \begin{bmatrix} 1 & -e^{j2\pi f\frac{d_2-d_1}{c}} & 0 & 0 \\ 1 & 0 & -e^{j2\pi f\frac{d_3-d_1}{c}} & 0 \\ 1 & 0 & 0 & -e^{j2\pi f\frac{d_4-d_1}{c}} \end{bmatrix}. \quad (21)$$

If $\mathbf{B}$ is a perfect blocking matrix, the GSC is roughly independent of speech detection. In practice however, the a priori assumptions of the GSC are seldom fulfilled. To avoid possible cancellation of the speech, the ANC is adapted during periods of noise only. Hence, the GSC is -apart from a reduced convergence rate- not affected when *noise only* samples are detected as *speech + noise*. The PTFs $G^s(f)$, $G^n(f)$ of the GSC as a function of Γ^s_{kl}, Γ^n_{kl} and the percentage $(\delta \times 100\%)$ of *speech + noise* samples detected as *noise only*, can be obtained in a similar manner as in Section 2.2 and Section 3.1:

$$G^{n|s}(f) = \Bigg[P^{n|s}_X \sum_{k=1}^{M}\sum_{l=1}^{M} A_k A^*_l \Gamma^{n|s}_{kl} - 2\sum_{k=1}^{N} Re\{P^{n|s}_{Y_0 Y_k} W^*_k\} + \sum_{k=1}^{N}\sum_{l=1}^{N} W_k W^*_l P^{n|s}_{Y_k Y_l}\Bigg] / P^{n|s}_X, \quad (22)$$

where

$$P^{n|s}_{Y_i Y_j}(f) = P^{n|s}_X(f)\sum_{k=1}^{M}\sum_{l=1}^{M} B_{ik}B^*_{jl}\Gamma^{n|s}_{kl}(f) \quad (23)$$

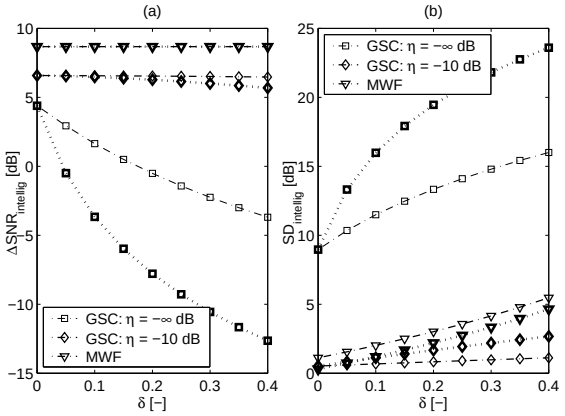


Fig. 1. Theoretical performance ((a) ΔSNR$_{\text{intellig}}$; (b) SD$_{\text{intellig}}$) of GSC (with(out) robustness constraint) and MWF, when $(\delta\times 100)$% of the noise samples used to compute the filters actually contain speech. Two input SNRs $1/\alpha(f)$ are depicted: $\alpha(f) = 0$ dB (dashed-dotted lines) and $\alpha(f) = -6$ dB (dotted lines).

$$P^{n|s}_{Y_0 Y_i}(f) = P^{n|s}_X(f)\sum_{k=1}^{M}\sum_{l=1}^{M} A_k B^*_{il}\Gamma^{n|s}_{kl}(f), \quad (24)$$

and $\mathbf{W}(f)$ equals

$$\begin{bmatrix} P^n_{Y_1Y_1} + \delta P^s_{Y_1Y_1} & \cdots & P^n_{Y_NY_1} + \delta P^s_{Y_NY_1} \\ \vdots & \ddots & \vdots \\ P^n_{Y_1Y_N} + \delta P^s_{Y_1Y_N} & \cdots & P^n_{Y_NY_N} + \delta P^s_{Y_NY_N} \end{bmatrix}^{-1}\begin{bmatrix} P^n_{Y_0Y_1} \\ \vdots \\ P^n_{Y_0Y_N} \end{bmatrix}. \quad (25)$$

For the derivation, we refer to [10].

When used in combination with small-sized arrays, e.g. in hearing aid applications, an additional robustness constraint on the ANC of the GSC [5, 6] is required to guarantee performance in the presence of small signal model errors [7]. In [6], this constraint is imposed by inserting uncorrelated noise in the microphone signals used to design $\mathbf{W}(f)$. This constraint can be easily taken into account by replacing $\Gamma^n_{kl}(f)$ in the computation of $\mathbf{W}(f)$ by

$$\Gamma^n_{kl}(f) + \eta(f)\delta[k-l], \quad (26)$$

where $\eta(f)$ is the ratio of the injected noise power to the input noise power $P^n_X(f)$ [10].

3.2.2. Simulation results

The simulations are illustrated for a diffuse noise field with internal-to-external noise ratio $\beta(f) = P^i_X(f)/P^e_X(f) = -30$ dB. Typical values for $\beta(f)$ lie between -20 dB and -40 dB. The coherence Γ^n_{kl} of this noise field equals

$$\Gamma^n_{kl}(f) = \frac{\sin(2\pi f(d_k-d_l)/c)}{(1+\beta(f))2\pi f(d_k-d_l)/c} \qquad \text{for } k \neq l, \quad (27)$$

with $d_k - d_l$ the interspacing between the k-th and l-th microphone and c the velocity of sound in air ($c \approx 340$ m/s). The coherence function $\Gamma^s_{kl}(f)$ of the localized speech signal equals

$$\Gamma^s_{kl}(f) = e^{-j2\pi f\cos\theta(d_k-d_l)/c}. \quad (28)$$

Since in practice, the desired and interfering signals are often speech-like, ΔSNR$_{\text{intellig}}$ is computed using a model of the average long-term speech PSD for $P^s_X(f)$ and $P^n_X(f)$. In the simulations, the 2nd and 3rd microphone have a small gain mismatch $\Delta\Upsilon$ of 1 dB and -1 dB w.r.t. the first microphone. In [11], gain and phase differences 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.

Figure 1 depicts the impact of speech detection errors on $\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$ obtained by the GSC and the MWF, when $\delta \times 100\%$[5] of the noise samples used to compute the filters actually contain speech. Two input SNRs $1/\alpha(f)$, i.e. $\alpha(f) = 0\,\text{dB}$ and $\alpha(f) = -6$ dB are considered. The performance of the GSC with robustness constraint is depicted too for a noise injection ratio $\eta(f) = -10$ dB. This ratio 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 [7]. For larger model errors, a more severe constraint is needed, at the expense of less noise reduction [7].

In addition to increasing robustness to model errors [5, 7], the robustness constraint reduces the drastic impact of speech detection errors on the GSC, especially at high input SNR $1/\alpha(f)$. The MWF additionally distorts the speech by $10\log_{10}(1-\delta)^2$ but conserves the improvement $\Delta\text{SNR}_{\text{intellig}}$ if speech + noise is detected as noise only. In contrast to the GSC, this additional distortion (i.e. $10\log(1-\delta)^2$), is independent of the input SNR. Compared to the GSC with robustness constraint, the speech signal is attenuated more, especially for large δ. For error rates $\delta \times 100\%$ up to 20%, the distortion is limited, while more noise is reduced.

4. VALIDATION THROUGH EXPERIMENTAL RESULTS

In this section, we verify the conclusions of Section 3.2 based on real recordings for a diffuse noise field. A uniform endfire array with 4 microphones (Knowles EM-4368) and $d = 0.02$ m has been mounted on a dummy head in an office room with reverberation time $T_{60\,\text{dB}} \approx 700$ ms for a speech weighted noise. The desired source is positioned at a distance of 1 meter in front of the head. The speech and noise signal are uncorrelated, stationary and speech-like. 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 is speech-like, $\beta(f)$ is smaller than -30 dB at low f. The level of the speech signal is adjusted so that the input SNR at the first microphone equals 0 dB. 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 subband GSVD based algorithm is used [3] and the fixed beamformer in the GSC has been optimized (using a free field signal coming from 0°) for the 4-microphone array used . A gain deviation of 1 dB and -1 dB has been applied to the 2nd and 3rd microphone.

Figure 2 shows the performance of the GSC (with $\eta = -\infty$ dB and $\eta = -10$ dB, respectively) and the MWF when $(\delta \times 100)\%$ of the noise samples used to compute the filters, actually contain speech. The results are well predicted by Figure 1. Compared to Figure 1, less noise is reduced due to the presence of reverberation. In addition, $\beta(f) < -30$ dB at low f results in a larger distortion by the GSC at $\delta = 0$. The improvement $\Delta\text{SNR}_{\text{intellig}}$ of the MWF is (again) hardly affected by erroneous speech detection, while $\text{SD}_{\text{intellig}}$ increases with $10\log_{10}(1-\delta)^2$.

In conclusion, for a reasonable speech detection error rate of 20% or less, the MWF outperforms the GSC, even when the latter is supplied with a robustness constraint.

5. REFERENCES

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

[2] G. Rombouts and M. Moonen, "QRD-based optimal filtering for acoustic noise reduction," in *Proc. of EUSIPCO*, 2002, vol. 3, pp. 301–304.

[3] A. Spriet, M. Moonen, and J. Wouters, "A multi-channel subband gsvd approach to speech enhancement," *ETT*, vol. 13, no. 2, pp. 149–158, 2002.

[4] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. AP*, vol. 30, pp. 27–34, Jan. 1982.

[5] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust Adaptive Beamforming," *IEEE Trans. ASSP*, vol. 35, no. 10, pp. 1365–1376, Oct. 1987.

[6] N. K. Jablon, "Adaptive beamforming with the generalized sidelobe canceller in the presence of array imperfections," *IEEE Trans. AP*, vol. 34, pp. 996–1012, 1986.

[7] A. Spriet, M. Moonen, and J. Wouters, "Robustness analysis of GSVD based optimal filtering and GSC for hearing aid applications," in *Proc. of WASPAA*, 2001.

[8] 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, 1993.

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

[10] A. Spriet, M. Moonen, and J. Wouters, "The impact of speech detection errors on the noise reduction performance of multi-channel wiener filtering and GSC," Tech. Rep. ESAT-SISTA/TR 02-163, 2002, available at ftp://ftp.esat.kuleuven.ac.be/sista/spriet/reports/02-163.ps.gz.

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

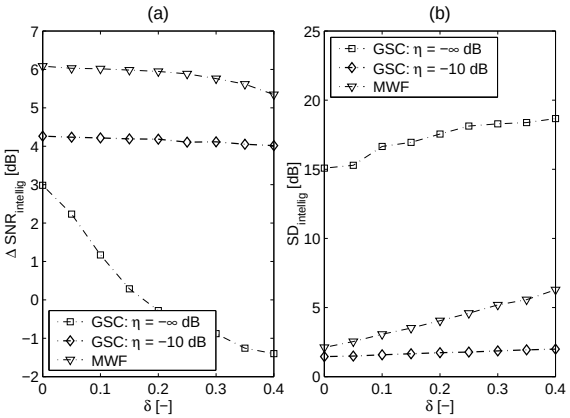


Fig. 2. Performance ((a) $\Delta\text{SNR}_{\text{intellig}}$; (b) $\text{SD}_{\text{intellig}}$) of GSC and MWF, when $(\delta \times 100)\%$ of the noise samples actually contain speech. Experimental results.

[5] For simplicity, we assume δ in (16) to be f-independent.