Katholieke Universiteit Leuven

Departement Elektrotechniek

ESAT-SISTA/TR 2004-25

Design of a robust multi-microphone noise reduction algorithm for hearing instruments

Simon Doclo, Ann Spriet, Marc Moonen, Jan Wouters [1]

January 30, 2004

in *Proc. of the International Symposium on Mathematical Theory of Networks and Systems (MTNS2004)*, Leuven, Belgium, July 5-9 2004.

[1]ESAT (SISTA) - Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Leuven (Heverlee), Belgium, Tel. 32/16/321899, Fax 32/16/321970, WWW: *http://www.esat.kuleuven.ac.be/sista*. E-mail: *simon.doclo@esat.kuleuven.ac.be*. Simon Doclo is a postdoctoral researcher funded by KULeuven-BOF. Marc Moonen is an Associate Professor at the Department of Electrical Engineering of the Katholieke Universiteit Leuven. Jan Wouters is an Associate Professor at the Laboratory for Exp. ORL of the Katholieke Universiteit Leuven. This research work was carried out at the ESAT laboratory and the Lab. Exp. ORL of the Katholieke Universiteit Leuven, in the frame of the F.W.O. Project G.0233.01, *Signal Processing and Automatic Patient Fitting for Advanced Auditory Prostheses*, the I.W.T. Project 020540, *Performance improvement of cochlear implants by innovative speech processing algorithms*, the I.W.T. Project 020476, *Sound Management System for Public Address systems (SMS4PA)*, the Concerted Research Action *Mathematical Engineering Techniques for Information and Communication Systems* (MEFISTO-666) of the Flemish Government, the Interuniversity Attraction Pole IUAP P5-22, *Dynamical systems and control: computation, identification and modelling*, and was partially sponsored by Cochlear. The scientific responsibility is assumed by its authors.

Design of a robust multi-microphone noise reduction algorithm for hearing instruments

Simon Doclo[1], *Ann Spriet*[1,2], *Marc Moonen*[1]
[1]Katholieke Universiteit Leuven
Dept. of Electrical Engineering (ESAT-SCD)
Kasteelpark Arenberg 10, 3001 Leuven, Belgium
{doclo,spriet,moonen}@esat.kuleuven.ac.be

Jan Wouters[2]
[2]Katholieke Universiteit Leuven
Laboratory for Exp. ORL
Kapucijnenvoer 33, 3000 Leuven, Belgium
jan.wouters@uz.kuleuven.ac.be

***Abstract*— This paper discusses the design and low-cost implementation of a robust multi-microphone noise reduction scheme, called the Spatially Pre-processed Speech Distortion Weighted Multi-channel Wiener Filter (SP-SDW-MWF). This scheme consists of two parts: a robust fixed spatial pre-processor and a robust adaptive Multi-channel Wiener Filter (MWF). Robustness against signal model errors is achieved by incorporating statistical information about the microphone characteristics into the design procedure of the spatial pre-processor and by taking speech distortion explicitly into account in the optimisation criterion of the MWF. Experimental results using a hearing aid show that the proposed scheme achieves a better noise reduction performance for a given maximum speech distortion level, compared to the widely studied Generalised Sidelobe Canceller (GSC) with Quadratic Inequality Constraint (QIC). For implementing the adaptive SDW-MWF, an efficient stochastic gradient algorithm in the frequency-domain can be derived, whose computational complexity and memory usage is comparable to the NLMS-based Scaled Projection Algorithm for implementing the QIC-GSC.**

I. Introduction

Noise reduction algorithms in hearing aids and cochlear implants are crucial for hearing impaired persons in order to improve speech intelligibility in background noise. Multi-microphone systems exploit spatial in addition to spectro-temporal information of the desired and the noise signals and are hence preferred to single-microphone systems. For small-sized microphone arrays such as typically used in hearing instruments, multi-microphone noise reduction however goes together with an increased sensitivity to errors in the assumed signal model such as microphone mismatch (gain, phase, position), reverberation, speech detection errors, etc. [1]–[8].

In [9] a generalised multi-microphone noise reduction scheme, called the Spatially Pre-processed Speech Distortion Weighted Multi-channel Wiener Filter (SP-SDW-MWF), has been proposed (cf. Section II), whose structure strongly resembles the widely used Generalised Sidelobe Canceller (GSC) [10]–[17]. It consists of a fixed spatial pre-processor, generating speech and noise references, and an adaptive stage, reducing the residual noise in the speech reference. This generalised scheme encompasses both the GSC and the MWF [18]–[20] as extreme cases and allows for in-between solutions such as the Speech Distortion Regularised GSC (SDR-GSC).

Both the fixed spatial pre-processor and the adaptive stage strongly rely on a-priori assumptions (e.g. about the microphone characteristics). When these assumptions are not satisfied, both the fixed and the adaptive stage give rise to undesired speech distortion and to a reduced noise reduction performance. Hence, for both stages the robustness against signal model errors needs to be improved. The robustness of the *fixed spatial pre-processor* can be improved e.g. by limiting the white noise gain [1] or by calibrating the used microphone array [3]. However, when statistical knowledge about the microphone deviations (gain, phase, position) is available, we propose to incorporate this knowledge directly into the design procedure [21], [22] (cf. Section III). The robustness of the *adaptive stage* can be improved e.g. by using a Quadratic Inequality Constraint (QIC) [5] or coefficient constraints [16] on the adaptive filter. However, these are quite conservative approaches since the constraint is not related to the amount of speech leakage actually present in the noise references. Hence, we propose to take speech distortion explicitly into account in the design criterion of the adaptive stage, resulting in the SDW-MWF and the SDR-GSC [9] (cf. Section IV). Experimental results using a hearing aid show that, compared to the QIC-GSC, the SP-SDW-MWF achieves a better noise reduction performance for a given maximum speech distortion level (cf. Section V).

Different implementations exist for updating the adaptive filter in the SDW-MWF. In [19], [20] recursive matrix-based implementations (using GSVD and QRD) have been proposed, while in [23], [24] cheap stochastic gradient implementations in the time-domain and the frequency-domain have been developed (cf. Section VI). The computational complexity and memory usage of the frequency-domain algorithm in [24] is comparable to the NLMS-based algorithm for implementing the QIC-GSC, while experimental results show that it preserves the robustness benefit of the SP-SDW-MWF.

II. General structure and notational conventions

The Spatially Pre-processed Speech Distortion Weighted Multi-channel Wiener Filter (SP-SDW-MWF) is depicted in Figure 1 and consists of a *fixed spatial pre-processor*, i.e. a fixed beamformer $\mathbf{A}(z)$ and a blocking matrix $\mathbf{B}(z)$, and an *adaptive Speech Distortion Weighted Multi-channel Wiener Filter (SDW-MWF)* [9]. Note that this structure strongly resembles the GSC-structure [10]–[17], where the standard adaptive filter has been replaced by an adaptive SDW-MWF.

The desired speaker is assumed to be in front of the microphone array (having M microphones), and an endfire array is used. The

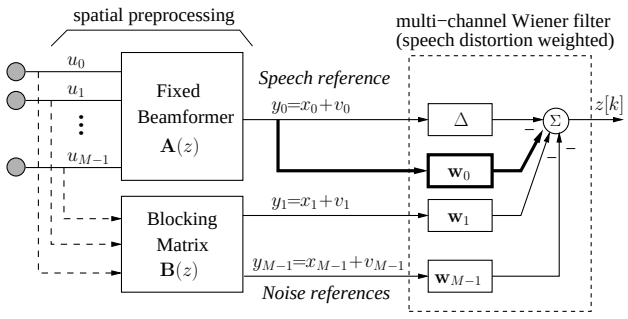


Fig. 1. General structure of the Spatially Pre-processed Speech Distortion Weighted Multi-channel Wiener Filter

fixed beamformer creates a so-called speech reference $y_0[k] = x_0[k] + v_0[k]$ (with $x_0[k]$ and $v_0[k]$ respectively the speech and the noise component of $y_0[k]$) by steering a beam towards the front, whereas the blocking matrix creates $M-1$ so-called noise references $y_i[k] = x_i[k]+v_i[k], i = 1 \ldots M-1$, by steering zeroes towards the front. During *speech-periods* these reference signals consist of speech and noise components, i.e. $y_i[k] = x_i[k] + v_i[k]$, whereas during *noise-only-periods* only the noise components $v_i[k]$ are observed. We assume that the second-order statistics of the noise are sufficiently stationary such that they can be estimated during noise-only-periods and used during subsequent speech-periods. This requires the use of a voice activity detection (VAD) mechanism [25]–[27], which determines whether speech is present or not.

Let N be the number of input channels to the multi-channel Wiener filter in Figure 1 ($N = M$ if $\mathbf{w}_0$ is present, $N = M-1$ otherwise). Let the FIR filters $\mathbf{w}_i[k]$ have length L, and consider the L-dimensional data vectors $\mathbf{y}_i[k]$, the NL-dimensional stacked filter $\mathbf{w}[k]$ and the NL-dimensional stacked data vector $\mathbf{y}[k]$, defined as

$$\mathbf{y}_i[k] = \left[\begin{array}{cccc} y_i[k] & y_i[k-1] & \ldots & y_i[k-L+1] \end{array}\right]^T , \quad (1)$$

$$\mathbf{w}[k] = \left[\begin{array}{cccc} \mathbf{w}_{M-N}^T[k] & \mathbf{w}_{M-N+1}^T[k] & \ldots & \mathbf{w}_{M-1}^T[k] \end{array}\right]^T , \quad (2)$$

$$\mathbf{y}[k] = \left[\begin{array}{cccc} \mathbf{y}_{M-N}^T[k] & \mathbf{y}_{M-N+1}^T[k] & \ldots & \mathbf{y}_{M-1}^T[k] \end{array}\right]^T , \quad (3)$$

with ^{T} denoting transpose. The data vector $\mathbf{y}[k]$ can be decomposed into a speech and a noise component, i.e. $\mathbf{y}[k] = \mathbf{x}[k] + \mathbf{v}[k]$, with $\mathbf{x}[k]$ and $\mathbf{v}[k]$ defined similarly as in (3). The goal of the filter $\mathbf{w}[k]$ is to estimate the delayed noise component $v_0[k-\Delta]$ in the speech reference (cf. Section IV). As can be seen from Figure 1, the output signal $z[k]$ is then computed by subtracting the filtered (speech and noise) reference signals from the delayed speech reference, i.e.

$$z[k] = y_0[k-\Delta] - \mathbf{w}^T[k]\mathbf{y}[k] . \quad (4)$$

Hence, the speech component of the output signal $z_x[k]$ is equal to

$$z_x[k] = x_0[k-\Delta] - \mathbf{w}^T[k]\mathbf{x}[k] . \quad (5)$$

This equation shows that *speech distortion* in the output signal can originate both from distortion of the speech component in the speech reference $x_0[k]$ and from speech leakage into the noise references ($\mathbf{x}[k] \neq \mathbf{0}$), e.g. caused by microphone mismatch and reverberation. Section III describes a procedure for designing robust fixed beamformers, hence limiting speech distortion in the speech reference (and to some extent speech leakage into the noise references), while Section IV describes a procedure for limiting speech distortion caused by the term $\mathbf{w}^T[k]\mathbf{x}[k]$.

III. Robust fixed spatial pre-processor

This section describes a design procedure for making the fixed beamformer $\mathbf{A}(z)$ and the blocking matrix $\mathbf{B}(z)$ more robust against microphone mismatch (gain, phase, position) [21], [22], hence limiting speech distortion in the speech reference and reducing to some extent speech leakage into the noise references.

A. Broadband beamforming: configuration

Consider the linear microphone array depicted in Figure 2, with M microphones, M K-taps FIR filters $\mathbf{f}_m$ (with real coefficients) and d_m the distance between the mth microphone and the centre of the microphone array. Assuming far-field conditions[1], the spatial directivity pattern $H(\omega,\theta)$ for a source $S(\omega)$ with normalised frequency

[1]Although far-field conditions are usually valid for hearing instruments because of the small size of the used microphone array, the proposed methods can easily be extended to near-field conditions [28], [29].

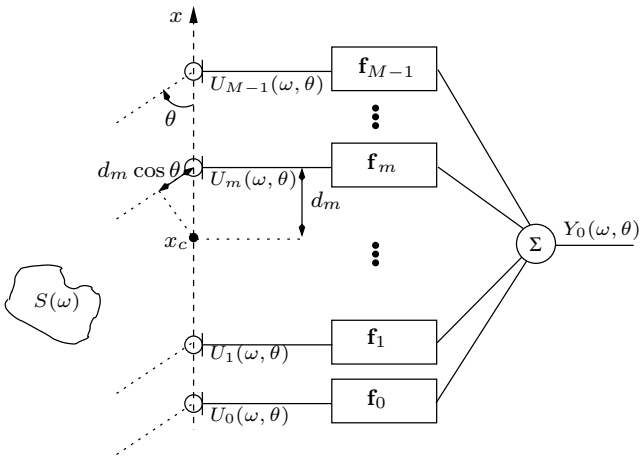


Fig. 2. Microphone array configuration (far-field assumption)

ω at an angle θ from the microphone array is defined as

$$H(\omega,\theta) = \mathbf{f}^T\mathbf{g}(\omega,\theta) , \quad (6)$$

with $\mathbf{f}$ the MK-dimensional real-valued stacked filter vector, $\mathbf{f} = \left[\begin{array}{cccc} \mathbf{f}_0^T & \ldots & \mathbf{f}_{M-1}^T \end{array}\right]^T$, and the steering vector $\mathbf{g}(\omega,\theta)$ equal to

$$\mathbf{g}(\omega,\theta) = \left[\begin{array}{c} \mathbf{e}(\omega)\, A_0(\omega,\theta)\, e^{-j\omega\tau_0(\theta)} \\ \vdots \\ \mathbf{e}(\omega)\, A_{M-1}(\omega,\theta)\, e^{-j\omega\tau_{M-1}(\theta)} \end{array}\right] , \quad (7)$$

with $\mathbf{e}(\omega) = \left[\begin{array}{cccc} 1 & e^{-j\omega} & \ldots & e^{-j(K-1)\omega} \end{array}\right]^T$ and

$$A_m(\omega,\theta) = a_m(\omega,\theta)\, e^{-j\psi_m(\omega,\theta)}, \quad m = 0 \ldots M-1 , \quad (8)$$

representing the frequency- and angle-dependent characteristics (gain, phase) of the mth microphone. The delay $\tau_m(\theta)$ is equal to

$$\tau_m(\theta) = \frac{d_m \cos\theta}{c} f_s , \quad (9)$$

with c the speed of sound ($340\frac{m}{s}$) and f_s the sampling frequency.

When a microphone position error occurs and the distance between the mth microphone and the centre of the array is $d_m + \delta_m$, this can be seen as a frequency- and angle-dependent phase shift $\omega\frac{\delta_m \cos\theta}{c} f_s$ for the mth microphone, which hence can be easily incorporated into the microphone characteristics in (8) as

$$A_m(\omega,\theta) = \underbrace{a_m(\omega,\theta)}_{\text{gain}} \underbrace{e^{-j\psi_m(\omega,\theta)}}_{\text{phase}} \underbrace{e^{-j\omega\frac{\delta_m\cos\theta}{c}f_s}}_{\text{position}} . \quad (10)$$

Using (7), (9) and (10), the ith element of $\mathbf{g}(\omega,\theta)$ is equal to

$$\mathbf{g}^i(\omega,\theta) = e^{-j\omega\left(k+\frac{d_m\cos\theta}{c}f_s\right)} a_m(\omega,\theta)\, e^{-j\psi_m(\omega,\theta)} e^{-j\omega\frac{\delta_m\cos\theta}{c}f_s} , \quad (11)$$

with $k = \text{mod}(i-1, K)$ and $m = \lfloor\frac{i-1}{K}\rfloor$. The steering vector $\mathbf{g}(\omega,\theta)$ can be decomposed into a real and an imaginary part, i.e. $\mathbf{g}(\omega,\theta) = \mathbf{g}_R(\omega,\theta) + j\mathbf{g}_I(\omega,\theta)$.

Using (6), the spatial directivity spectrum $|H(\omega,\theta)|^2$ is equal to

$$|H(\omega,\theta)|^2 = H(\omega,\theta)H^*(\omega,\theta) = \mathbf{f}^T\mathbf{G}(\omega,\theta)\mathbf{f} , \quad (12)$$

with $\mathbf{G}(\omega,\theta) = \mathbf{g}(\omega,\theta)\,\mathbf{g}^H(\omega,\theta)$. Using (11), the (i,j)-th element of $\mathbf{G}(\omega,\theta)$ is equal to

$$\mathbf{G}^{ij}(\omega,\theta) = e^{-j\omega\left((k-l)+\frac{(d_m-d_n)\cos\theta}{c}f_s\right)} a_m(\omega,\theta)\, a_n(\omega,\theta) \cdot e^{-j\left(\psi_m(\omega,\theta)-\psi_n(\omega,\theta)\right)} e^{-j\omega\frac{(\delta_m-\delta_n)\cos\theta}{c}f_s} , \quad (13)$$

with $l = \text{mod}(j-1, K)$ and $n = \lfloor \frac{j-1}{K} \rfloor$. The matrix $\mathbf{G}(\omega,\theta)$ can be decomposed into a real and an imaginary part $\mathbf{G}_R(\omega,\theta)$ and $\mathbf{G}_I(\omega,\theta)$. Since $\mathbf{G}_I(\omega,\theta)$ is anti-symmetric, $|H(\omega,\theta)|^2$ is equal to

$$|H(\omega,\theta)|^2 = \mathbf{f}^T \mathbf{G}_R(\omega,\theta)\mathbf{f} \; . \tag{14}$$

B. Weighted least-squares cost function

The design of a broadband beamformer consists of calculating the filter $\mathbf{f}$, such that $H(\omega,\theta)$ optimally fits the desired spatial directivity pattern $D(\omega,\theta)$, where $D(\omega,\theta)$ is allowed to be an arbitrary 2-dimensional function. Several design procedures exist, depending on the specific cost function that is optimised. In this paper, we only consider the weighted least-squares cost function. In [21], [28]–[32] also eigenfilter-based and non-linear cost functions are discussed.

Considering the least-squares (LS) error $|H(\omega,\theta) - D(\omega,\theta)|^2$, the weighted LS cost function is defined as

$$J_{LS}(\mathbf{w}) = \int_\Theta \int_\Omega F(\omega,\theta) |H(\omega,\theta) - D(\omega,\theta)|^2 d\omega d\theta \; , \tag{15}$$

where $F(\omega,\theta)$ is a positive real weighting function, assigning more or less importance to certain frequencies and angles. This cost function can be written as the quadratic function

$$J_{LS}(\mathbf{f}) = \mathbf{f}^T \mathbf{Q}_{LS} \mathbf{f} - 2\mathbf{f}^T \mathbf{a} + d_{LS} \; , \tag{16}$$

with (assuming $D(\omega,\theta)$ to be real-valued)

$$\mathbf{Q}_{LS} = \int_\Theta \int_\Omega F(\omega,\theta)\, \mathbf{G}_R(\omega,\theta)\, d\omega d\theta \tag{17}$$

$$\mathbf{a} = \int_\Theta \int_\Omega F(\omega,\theta) D(\omega,\theta)\, \mathbf{g}_R(\omega,\theta)\, d\omega d\theta \tag{18}$$

$$d_{LS} = \int_\Theta \int_\Omega F(\omega,\theta) D^2(\omega,\theta)\, d\omega d\theta \; . \tag{19}$$

The filter $\mathbf{f}_{LS}$, minimising the weighted LS cost function, is given by

$$\mathbf{f}_{LS} = \mathbf{Q}_{LS}^{-1}\, \mathbf{a} \; . \tag{20}$$

C. Robustness against microphone mismatch

Using the procedure described in Section III-B, it is possible to design beamformers when the microphone characteristics (gain, phase, position) are exactly known. However, small deviations from the assumed characteristics can lead to large deviations from the desired spatial directivity pattern [1]–[4]. Since in practice it is difficult to manufacture microphones with the same nominal gain and phase characteristics and microphone position errors frequently occur, a measurement or calibration procedure is required in order to obtain the true microphone characteristics [3]. However, after calibration the microphone characteristics can still drift over time [33].

When statistical knowledge, e.g. a probability density function (pdf), is available for the gain, phase and position errors, this knowledge can be incorporated into a robust design procedure. In [21], [22] two robust design procedures have been presented. Considering all feasible characteristics, the first design procedure optimises the *mean performance*, i.e. the weighted sum of the cost functions, using the probability of the microphone characteristics as weights, whereas the second design procedure optimises the *worst-case performance*, i.e. the maximum cost function.

The same problem of gain and phase errors has been studied in [6], where however only the narrowband case for a specific directivity pattern and a uniform pdf has been considered. The approach presented here is more general because we consider broadband beamformers with an arbitrary spatial directivity pattern, arbitrary probability density functions and we also take into account microphone position errors.

D. Mean performance criterion

Applied to the weighted LS cost function of Section III-B, the mean performance cost function can be written as

$$J_{LS}^t(\mathbf{f}) = \int_{A_0} \dots \int_{A_{M-1}} J_{LS}(\mathbf{f},\mathbf{A})\, f_{\mathcal{A}}(A_0) \dots f_{\mathcal{A}}(A_{M-1}) \; dA_0 \dots dA_{M-1} \; , \tag{21}$$

with $J_{LS}(\mathbf{f},\mathbf{A})$ the weighted LS cost function (16) for a specific microphone characteristic $\{A_0, \dots, A_{M-1}\}$ and $f_{\mathcal{A}}(A)$ the joint pdf of the stochastic variables a (gain), ψ (phase) and δ (position error). Without loss of generality, we assume that all microphone characteristics $A_m, m = 0 \dots M-1$, are described by the same pdf and that a, ψ and δ are independent stochastic variables, such that the joint pdf is separable, i.e. $f_{\mathcal{A}}(A) = f_\alpha(a) f_\Psi(\psi) f_\Delta(\delta)$, with $f_\alpha(a)$ the gain pdf, $f_\Psi(\psi)$ the phase pdf and $f_\Delta(\delta)$ the position error pdf.

By combining (16) and (21), the mean performance cost function can be written as

$$J_{LS}^t(\mathbf{f}) = \mathbf{f}^T \mathbf{Q}_t \mathbf{f} - 2\mathbf{f}^T \mathbf{a}_t + d_{LS} \; , \tag{22}$$

which has the same form as (16), with

$$\mathbf{a}_t = \int_{A_0} \dots \int_{A_{M-1}} \mathbf{a}\, f_{\mathcal{A}}(A_0) \dots f_{\mathcal{A}}(A_{M-1})\, dA_0 \dots dA_{M-1} \; ,$$

$$\mathbf{Q}_t = \int_{A_0} \dots \int_{A_{M-1}} \mathbf{Q}_{LS}\, f_{\mathcal{A}}(A_0) \dots f_{\mathcal{A}}(A_{M-1})\, dA_0 \dots dA_{M-1} \; .$$

The calculation of these expressions (both for uniform and Gaussian pdfs) has been thoroughly discussed in [21], [22], [29].

E. Minimax criterion

When optimising the mean performance, it is however still possible - although typically with a low probability - that for some specific microphone mismatch the cost function is quite high. If this is considered to be a problem, the worst-case performance should be optimised using the minimax criterion.

For the minimax criterion, we first have to define a (finite) set of microphone characteristics (K_a gain values, K_γ phase values and K_δ position error values),

$$\{a_{min} = a_1, \dots, a_{K_a} = a_{max}\},\ \{\gamma_{min} = \gamma_1, \dots, \gamma_{K_\gamma} = \gamma_{max}\},$$
$$\{\delta_{min} = \delta_1, \dots, \delta_{K_\delta} = \delta_{max}\} \; , \tag{23}$$

as an approximation for the continuum of feasible microphone characteristics, and use this set of gain, phase and position error values to construct the $(K_a K_\gamma K_\delta)^M$-dimensional vector $\mathbf{F}(\mathbf{f})$,

$$\mathbf{F}(\mathbf{f}) = \left[\; F_1(\mathbf{f}) \quad F_2(\mathbf{f}) \quad \dots \quad F_{(K_a K_\gamma K_\delta)^M}(\mathbf{f}) \; \right]^T \; , \tag{24}$$

which consists of the used cost function (weighted LS or any other cost function) at each possible combination of gain, phase and position error values. The goal then is to minimise the L_∞-norm of $\mathbf{F}(\mathbf{f})$, i.e. the maximum value of the elements $F_k(\mathbf{f})$,

$$\min_{\mathbf{f}} \|\mathbf{F}(\mathbf{f})\|_\infty = \min_{\mathbf{w}} \max_k F_k(\mathbf{f}) \; , \tag{25}$$

which can e.g. be done using a sequential quadratic programming (SQP) method [34]. In order to improve the numerical robustness and the convergence speed, the gradient

$$\left[\; \frac{\partial F_1(\mathbf{f})}{\partial \mathbf{f}} \quad \frac{\partial F_2(\mathbf{f})}{\partial \mathbf{f}} \quad \dots \quad \frac{\partial F_{(K_a K_\gamma K_\delta)^M}(\mathbf{f})}{\partial \mathbf{f}} \; \right] \; , \tag{26}$$

which is an $MK \times (K_a K_\gamma K_\delta)^N$-dimensional matrix, can be supplied analytically. As can be seen, the larger K_a, K_γ and K_δ, the denser

the grid of feasible microphone characteristics, and the higher the computational complexity for solving the minimax problem.

When only considering gain errors and using the weighted LS cost function, it has been proved in [21] that for any $\mathbf{f}$ the maximum value of $\mathbf{F}(\mathbf{f})$ occurs on a boundary point of an M-dimensional hypercube, i.e. $a_m = a_{min}$ or $a_m = a_{max}$, $m = 0 \ldots M-1$. This implies that $K_a = 2$ suffices and $\mathbf{F}(\mathbf{f})$ consists of 2^M elements.

F. Simulations

We have performed simulations using a small-sized non-uniform linear microphone array consisting of $M = 3$ microphones at positions $\begin{bmatrix} -0.01 & 0 & 0.015 \end{bmatrix}$ m. We have designed an endfire broadband beamformer with passband specifications $(\Omega_p, \Theta_p) = (300\text{–}4000\,\text{Hz}, 0^\circ\text{–}60^\circ)$ and stopband specifications $(\Omega_s, \Theta_s) = (300\text{–}4000\,\text{Hz}, 80^\circ\text{–}180^\circ)$ and $f_s = 8\,\text{kHz}$. The filter length $L = 20$ and the weighting function $F(\omega, \theta) = 1$.

In the *first experiment*, we have investigated the effect of only gain and phase errors, hence assuming no microphone position errors ($\delta_m = 0,\ m = 0 \ldots M-1$). We have designed several types of broadband beamformers using the weighted LS cost function:

1) a non-robust beamformer (i.e. assuming no mismatch)
2) a robust beamformer using a uniform gain pdf $(0.85, 1.15)$, and a uniform phase pdf $(-5^\circ, 10^\circ)^2$
3) a robust beamformer using the minimax criterion (only gain errors, $a_{min} = 0.85$, $a_{max} = 1.15$, $K_a = 2$)

Figure 3 shows the spatial directivity plots of the non-robust, the gain/phase-robust and the minimax beamformer for several frequencies, when no gain and phase errors occur. As can be seen, the performance of the non-robust beamformer is the best, but the performance of the robust beamformers is certainly acceptable. Figure 4 shows the spatial directivity plots in case of (small) gain and phase errors (microphone gains = $\begin{bmatrix} 0.9 & 1.1 & 1.05 \end{bmatrix}$ and phases = $\begin{bmatrix} 5^\circ & -2^\circ & 5^\circ \end{bmatrix}$). As can be seen, the performance of the non-robust beamformer deteriorates considerably. Certainly for the low frequencies, the spatial directivity pattern is almost omni-directional and the amplification is very high. On the other hand, the robust beamformers retain the desired spatial directivity pattern, even when gain and phase errors occur.

In the *second experiment*, we have investigated the effect of only microphone position errors, hence the microphones are assumed to be omni-directional microphones with a frequency response equal to 1, i.e. $a_m(\omega, \theta) = 1$ and $\psi_m(\omega, \theta) = 0$, $m = 0 \ldots M-1$. We have designed 2 types of broadband beamformers:

1) a non-robust beamformer, i.e. assuming no microphone position errors ($\delta_m = 0$, $m = 0 \ldots M-1$)
2) a robust beamformer using a Gaussian microphone position error pdf

$$f_\Delta(\delta) = \frac{1}{\sqrt{2\pi s_\delta^2}} e^{-\frac{(\delta - u_\delta)^2}{2 s_\delta^2}}, \qquad (27)$$

with $u_\delta = 0$ and $s_\delta = 0.003$[2].

Figures 5 and 6 show the spatial directivity plots of the non-robust beamformer and the robust beamformer for several frequencies, both when no microphone position errors occur and when (small) microphone position errors $\begin{bmatrix} 0.002 & -0.002 & 0.002 \end{bmatrix}$ m occur. When no errors occur, the performance of the non-robust beamformer is the best, but the performance of the robust beamformer is certainly acceptable. However, when microphone position errors occur, the performance of the non-robust beamformer deteriorates considerably, certainly at low frequencies. On the other hand, the robust beamformer retains the desired spatial directivity pattern, even when microphone position errors occur.

[2]These values for the probability density functions depend on the accuracy of the manufacturing process of the microphone arrays.

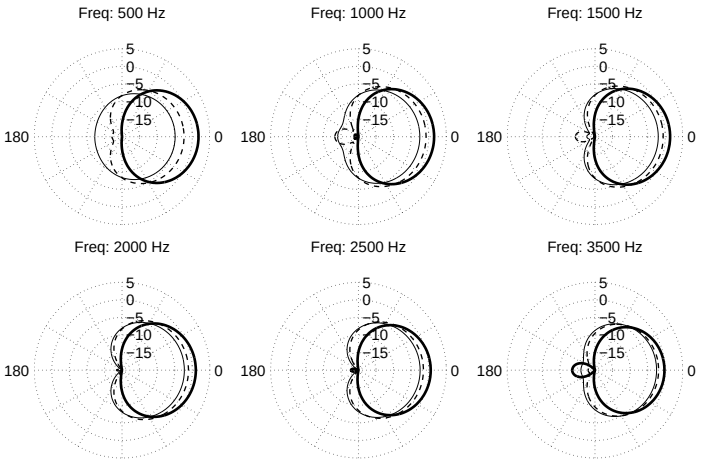


Fig. 3. Spatial directivity plots, no gain and phase errors (non-robust: thick solid, gain/phase-robust: dashed, minimax: solid)

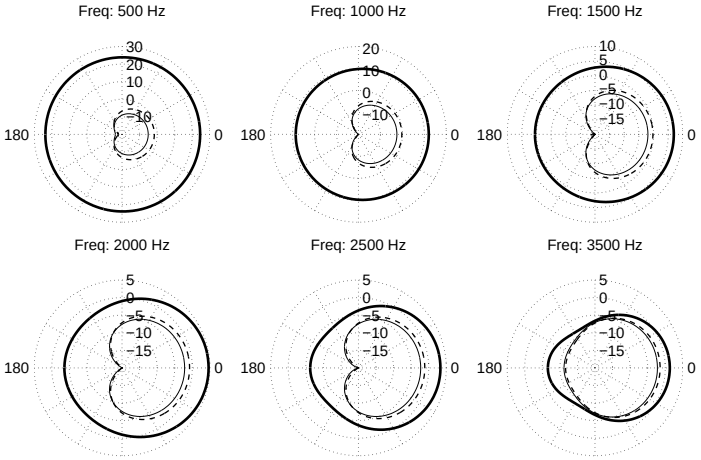


Fig. 4. Spatial directivity plots, gain and phase errors (non-robust: thick solid, gain/phase-robust: dashed, minimax: solid)

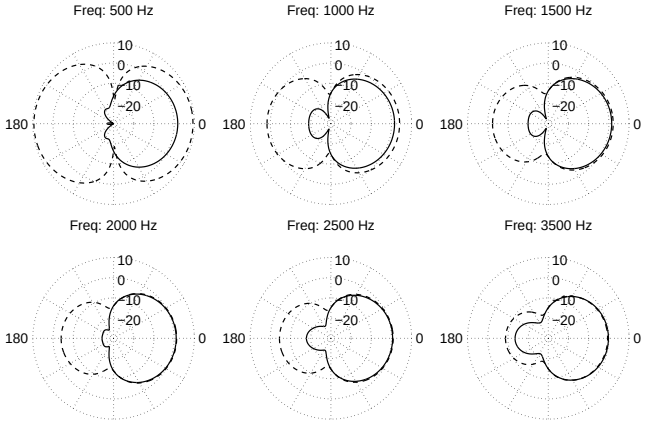


Fig. 5. Spatial directivity plots for non-robust beamformer (no errors: solid line, microphone position errors: dashed line)

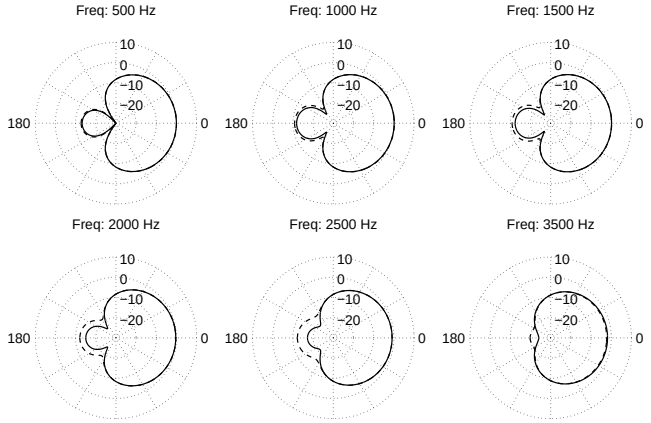


Fig. 6. Spatial directivity plots for robust beamformer (no errors: solid line, microphone position errors: dashed line)

IV. Robust adaptive stage: Speech distortion weighted Multi-channel Wiener Filter

This section describes a procedure for limiting speech distortion in the output signal due to the term $\mathbf{w}^T[k]\mathbf{x}[k]$ in the adaptive stage of the SP-SDW-MWF, cf. (5). A common approach to limit this term is to use a Quadratic Inequality Constraint (QIC) on the norm of the filter [5], i.e. $||\mathbf{w}[k]|| \leq \beta$. However - as will be shown in the simulations in Section V - this is a conservative approach, since the constraint is not dependent on the actual amount of speech leakage $\mathbf{x}[k]$ present in the noise references. In [9] a novel approach has been presented where speech distortion is taken directly into account in the optimisation criterion of the adaptive stage. The goal of the Speech Distortion Weighted Multi-channel Wiener Filter (SDW-MWF) in Figure 1 is to provide an estimate of the delayed noise component $v_0[k-\Delta]$ in the speech reference by minimising the cost function

$$J(\mathbf{w}[k]) = \frac{1}{\mu}\underbrace{\mathcal{E}\left\{\left|\mathbf{w}^T[k]\mathbf{x}[k]\right|^2\right\}}_{\varepsilon_x^2} + \underbrace{\mathcal{E}\left\{\left|v_0[k-\Delta]-\mathbf{w}^T[k]\mathbf{v}[k]\right|^2\right\}}_{\varepsilon_v^2} \tag{28}$$

where ε_x^2 represents the speech distortion energy, ε_v^2 represents the residual noise energy and the regularisation parameter $\mu \in [0,\infty)$ provides a trade-off between noise reduction and speech distortion [19], [35]. The filter $\mathbf{w}[k]$ minimising this cost function is given by

$$\mathbf{w}[k]=\left(\frac{1}{\mu}\mathcal{E}\left\{\mathbf{x}[k]\mathbf{x}^T[k]\right\}+\mathcal{E}\left\{\mathbf{v}[k]\mathbf{v}^T[k]\right\}\right)^{-1}\mathcal{E}\left\{\mathbf{v}[k]v_0[k-\Delta]\right\}. \tag{29}$$

In practice, the clean speech correlation matrix $\mathcal{E}\left\{\mathbf{x}[k]\mathbf{x}^T[k]\right\}$ obviously is unknown. Assuming that speech and noise are uncorrelated, this correlation matrix can be estimated as

$$\mathcal{E}\left\{\mathbf{x}[k]\mathbf{x}^T[k]\right\} = \mathcal{E}\left\{\mathbf{y}[k]\mathbf{y}^T[k]\right\} - \mathcal{E}\left\{\mathbf{v}[k]\mathbf{v}^T[k]\right\}\ , \tag{30}$$

where $\mathcal{E}\left\{\mathbf{y}[k]\mathbf{y}^T[k]\right\}$ is estimated during speech-periods and $\mathcal{E}\left\{\mathbf{v}[k]\mathbf{v}^T[k]\right\}$ is estimated during noise-only-periods. The second-order statistics of the noise are assumed to be quite stationary, such that they can be estimated during noise-only-periods and used during subsequent speech-periods. Similarly as for the GSC, a robust VAD-mechanism is hence required [25]–[27].

As depicted in Figure 1, the noise estimate $\mathbf{w}^T[k]\mathbf{y}[k]$ is then subtracted from the speech reference in order to obtain the enhanced output signal $z[k]$. Depending on the setting of μ and the presence/absence of the filter $\mathbf{w}_0$ on the speech reference, different algorithms are obtained [9].

A. SP-SDW-MWF without filter $\mathbf{w}_0$ (SDR-GSC)

When no filter $\mathbf{w}_0$ is present, the Speech Distortion Regularised GSC (SDR-GSC) is obtained, where the standard adaptive noise cancellation design criterion of the GSC (i.e. minimising the residual noise energy ε_v^2) is supplemented with a *regularisation term* $\frac{1}{\mu}\varepsilon_x^2$ that takes into account speech distortion due to signal model errors. For $\mu = \infty$, the standard GSC is obtained, and speech distortion is completely ignored. When $\mu \neq \infty$, the regularisation term adds robustness to the GSC, while not affecting the noise reduction performance in the absence of speech leakage:

- In the *absence of speech leakage*, i.e. $\mathbf{x}[k] = \mathbf{0}$, the regularisation term equals $\mathbf{0}$ for all $\mathbf{w}[k]$. Hence the residual noise energy ε_v^2 is effectively minimised or, in other words, the GSC-solution is obtained.
- In the *presence of speech leakage*, i.e. $\mathbf{x}[k] \neq \mathbf{0}$, speech distortion is explicitly taken into account in the optimisation criterion, hence limiting speech distortion while reducing noise. The larger the amount of speech leakage, the more attention is paid to speech distortion.

In contrast to the SDR-GSC, the QIC acts irrespective of the amount of speech leakage present. The constraint value β has to be chosen based on the largest model errors that may occur. Hence, noise reduction performance is compromised even when no or very small model errors are present, such that the QIC is more conservative than the SDR-GSC (cf. Section V).

B. SP-SDW-MWF with filter $\mathbf{w}_0$

When the filter $\mathbf{w}_0$ is present, the SP-SDW-MWF is obtained. Again, the regularisation parameter μ trades off speech distortion and noise reduction (for $\mu = 1$, we obtain an MWF, where the output signal $z[k]$ is the MMSE estimate of the speech component $x_0[k-\Delta]$ in the speech reference). In addition, we can make the following statements:

- In the *absence of speech leakage* and for infinitely long filters, the SP-SDW-MWF corresponds to a cascade of an SDR-GSC and an SDW single-channel Wiener postfilter [36], [37].
- In the *presence of speech leakage*, the SP-SDW-MWF tries to preserve its performance, i.e. the SP-SDW-MWF then contains extra filtering operations that compensate for the performance degradation of the SDR-GSC with postfilter due to speech leakage [9]. It can be proved that for infinite filter lengths, the SP-SDW-MWF is not affected by microphone mismatch as long as the speech component in the speech reference remains unaltered.

V. Experimental results

In this section it is shown by experimental results using hearing aid recordings that in comparison with the QIC-GSC, the SDR-GSC obtains a better noise reduction performance for small model errors, while guaranteeing robustness against large model errors, and that in comparison with the SDR-GSC, the performance of the SP-SDW-MWF is even less affected by signal model errors.

A. Set-up and performance measures

A 3-microphone BTE ('behind the ear') hearing aid has been mounted on a dummy head in an office room. The desired signal and the noise signals are uncorrelated, stationary and speech-like. The desired signal and the total noise signal both have a level of

70 dB SPL at the centre of the head. The desired source is positioned in front of the head (at 0°). Five noise sources are positioned at 75°, 120°, 180°, 240° and 285°. For evaluation purposes, the speech and the noise signals have been recorded separately. In the experiments, the microphones have been calibrated in an anechoic room with the BTE mounted on the head. A delay-and-sum beamformer is used as fixed beamformer $\mathbf{A}(z)$. The blocking matrix $\mathbf{B}(z)$ pairwise subtracts the time-aligned calibrated microphone signals. In order to investigate the effect of different parameter settings (i.e. μ, $\mathbf{w}_0$) on the performance of the SP-SDW-MWF, the filter coefficients are computed using (29) where $\mathcal{E}\left\{\mathbf{x}[k]\mathbf{x}^T[k]\right\}$ is estimated by means of the clean speech components of the microphone signals. In practice, $\mathcal{E}\left\{\mathbf{x}[k]\mathbf{x}^T[k]\right\}$ is approximated using (30). The effect of the approximation (30) on the performance was found to be small for the given data set. The used filter length $L = 96$. The QIC-GSC has been implemented using variable loading RLS [38].

To assess the performance, the intelligibility weighted signal-to-noise ratio improvement $\Delta\text{SNR}_{\text{intellig}}$ is used, defined as

$$\Delta\text{SNR}_{\text{intellig}} = \sum_i I_i(\text{SNR}_{i,\text{out}} - \text{SNR}_{i,\text{in}}), \tag{31}$$

where I_i expresses the importance for intelligibility of the i-th one-third octave band with centre frequency f_i^c [39], and where $\text{SNR}_{i,\text{out}}$ and $\text{SNR}_{i,in}$ are respectively the output and the input SNR (in dB) in this band. Similarly, we define an intelligibility weighted spectral distortion measure, called $\text{SD}_{\text{intellig}}$, of the desired signal as

$$\text{SD}_{\text{intellig}} = \sum_i I_i \text{SD}_i \tag{32}$$

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

$$\text{SD}_i = \frac{1}{(2^{1/6} - 2^{-1/6})\, f_i^c} \int_{2^{-1/6} f_i^c}^{2^{1/6} f_i^c} |10 \log_{10} G_x(f)|\, df, \tag{33}$$

with $G_x(f)$ the power transfer function of speech from the input to the output of the noise reduction algorithm. To exclude the effect of the spatial pre-processor, the performance measures are calculated with respect to the output of the fixed beamformer, i.e. the speech reference.

B. Experimental results

Figure 7 depicts the SNR improvement and the speech distortion of the SDR-GSC (without $\mathbf{w}_0$) and the SP-SDW-MWF (with $\mathbf{w}_0$) as a function of the regularisation parameter $1/\mu$. These figure also depict the effect of a gain mismatch Υ_2 of 4 dB at the second microphone. For comparison, Figure 8 plots the performance of the QIC-GSC with QIC $\mathbf{w}^T[k]\mathbf{w}[k] \leq \beta^2$, as a function of β^2.

From these figures, it can be observed that the standard GSC (i.e. the SDR-GSC with $1/\mu = 0$ or the QIC-GSC with $\beta^2 = \infty$) gives rise to a smaller SNR improvement and a large speech distortion when a gain mismatch of 4 dB occurs. Both the SP-SDW-MWF and the QIC-GSC increase the robustness of the standard GSC, since the speech distortion in the presence of signal model errors is reduced with increasing $1/\mu$ or decreasing β^2.

However, the QIC-GSC is more conservative than the SDR-GSC and the SP-SDW-MWF, since the constraint value β^2 is not dependent on the amount of speech leakage actually present in the noise references. E.g. suppose that the maximum allowable speech distortion $\text{SD}_{\text{intellig}}$ is 3 dB for a gain mismatch up to 4 dB. From Figure 8 it can be observed that $\beta^2 < 0.25$, such that the maximum SNR improvement $\Delta\text{SNR}_{\text{intellig}}$ is 4 dB (even when no gain

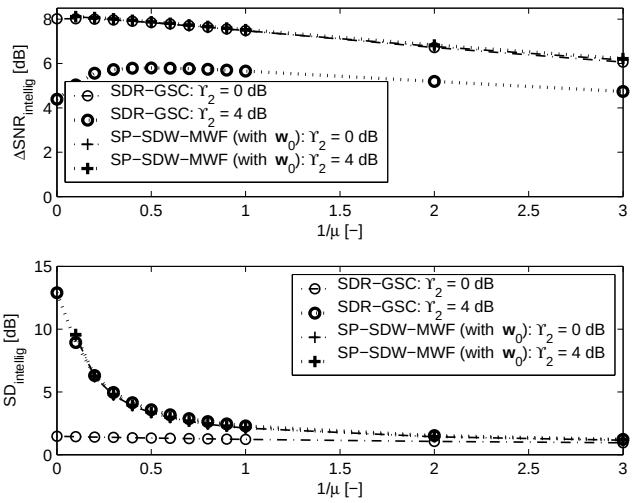


Fig. 7. SNR improvement and speech distortion of the SDR-GSC and the SP-SDW-MWF

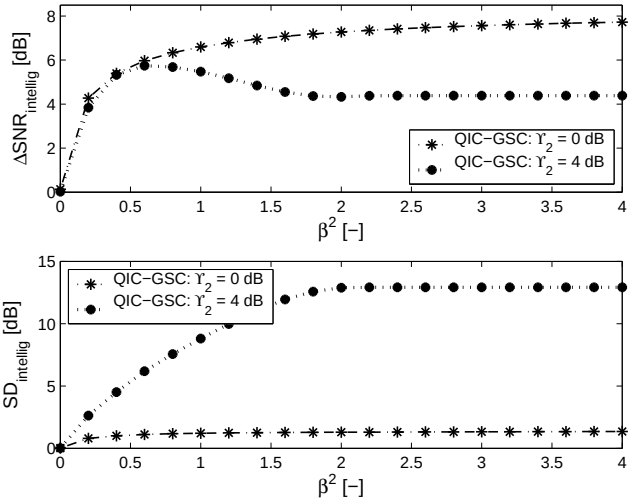


Fig. 8. SNR improvement and speech distortion of the QIC-GSC

mismatch occurs). Similarly, for the same maximum allowable speech distortion, it can be observed from Figure 7 that $1/\mu > 0.6$, such that the maximum SNR improvement for the SDR-GSC is equal to 6 dB with gain mismatch and 7.5 dB without gain mismatch, while the SNR improvement for the SP-SDW-MWF is equal to 7.5 dB (with and without gain mismatch). This can be explained by the fact that the SDR-GSC and the SP-SDW-MWF only put emphasis on speech distortion when actually required, i.e. when the amount of speech leakage is large.

Hence, for a given maximum allowable distortion, the SDR-GSC and the SP-SDW-MWF achieve a better noise reduction performance than the QIC-GSC. Furthermore, the performance of the SP-SDW-MWF is - in contrast to the SDR-GSC and the QIC-GSC - hardly affected by microphone mismatch.

VI. Efficient implementation using stochastic gradient (SG) algorithms

Different implementations exist for computing and updating the filter $\mathbf{w}[k]$. In [19], [20] recursive matrix-based implementations

(using GSVD and QRD) have been proposed, while in [23], [24] efficient stochastic gradient implementations in the time-domain and in the frequency-domain have been developed.

A. Time-Domain (TD) implementation

In [23] a stochastic gradient algorithm in the time-domain has been developed for minimising the cost function $J(\mathbf{w}[k])$ in (28), i.e.

$$\mathbf{w}[k+1] = \mathbf{w}[k] + \rho \left[\mathbf{v}[k](v_0[k-\Delta] - \mathbf{v}^T[k]\mathbf{w}[k]) - \mathbf{r}[k]\right] \quad (34)$$

$$\mathbf{r}[k] = \frac{1}{\mu}\mathbf{x}[k]\mathbf{x}^T[k]\mathbf{w}[k] \quad (35)$$

$$\rho = \frac{\rho'}{\mathbf{v}^T[k]\mathbf{v}[k] + \frac{1}{\mu}\mathbf{x}^T[k]\mathbf{x}[k] + \delta} , \quad (36)$$

with ρ the normalised step size of the adaptive algorithm, δ a small positive constant, and $\mathbf{w}[k]$, $\mathbf{v}[k]$, $\mathbf{x}[k]$ and $\mathbf{r}[k]$ NL-dimensional vectors. For $1/\mu = 0$ and no filter $\mathbf{w}_0$ present, (34) reduces to an NLMS-type update formula often used in GSC, *operated during noise-only-periods* [11]–[13]. For $1/\mu \neq 0$, the additional regularisation term $\mathbf{r}[k]$ limits speech distortion due to signal model errors.

In order to compute (35), knowledge about the (instantaneous) correlation matrix $\mathbf{x}[k]\mathbf{x}^T[k]$ of the clean speech signal is required, which is obviously not available. In order to avoid the need for calibration, it is suggested in [23] to store L-dimensional speech+noise-vectors $\mathbf{y}_i[k]$, $i = M-N \ldots M-1$ during speech-periods in a circular *speech+noise-buffer* $\mathbf{B}_y \in \mathbb{R}^{NL \times L_y}$ (similar as in [40])[3] and to adapt the filter $\mathbf{w}[k]$ using (34) during *noise-only-periods*, based on approximating the regularisation term in (35) by

$$\mathbf{r}[k] = \frac{1}{\mu}\left[\mathbf{y}_{B_y}[k]\mathbf{y}_{B_y}^T[k] - \mathbf{v}[k]\mathbf{v}^T[k]\right]\mathbf{w}[k] , \quad (37)$$

with $\mathbf{y}_{B_y}[k]$ a vector from the circular speech+noise-buffer $\mathbf{B}_y$. However, this estimate of $\mathbf{r}[k]$ is quite bad, resulting in a large excess error, especially for small μ and large ρ'. Hence, it has been suggested to use an estimate of the average clean speech correlation matrix $\mathcal{E}\{\mathbf{x}[k]\mathbf{x}^T[k]\}$ in (35), such that $\mathbf{r}[k]$ can be computed as

$$\mathbf{r}[k] = \frac{1}{\mu}(1-\bar{\lambda})\sum_{l=0}^{k}\bar{\lambda}^{k-l}\left[\mathbf{y}_{B_y}[l]\mathbf{y}_{B_y}^T[l] - \mathbf{v}[l]\mathbf{v}^T[l]\right]\cdot\mathbf{w}[k] , \quad (38)$$

with $\bar{\lambda}$ an exponential weighting factor and the step size ρ in (36) now equal to

$$\rho = \frac{\rho'}{\mathbf{v}^T[k]\mathbf{v}[k] + \frac{1}{\mu}(1-\bar{\lambda})\sum_{l=0}^{k}\bar{\lambda}^{k-l}\left|\mathbf{y}_{B_y}^T[l]\mathbf{y}_{B_y}[l] - \mathbf{v}^T[l]\mathbf{v}[l]\right| + \delta} .$$

For *stationary noise* a small $\bar{\lambda}$, i.e. $1/(1-\bar{\lambda}) \sim NL$, suffices. However, in practice the speech and the noise signals are often *spectrally highly non-stationary* (e.g. multi-talker babble noise), whereas their *long-term* spectral and spatial characteristics usually vary more slowly in time. Spectrally highly non-stationary noise can still be spatially suppressed by using an estimate of the *long-term* correlation matrix in $\mathbf{r}[k]$, i.e. $1/(1-\bar{\lambda}) \gg NL$.

In order to avoid expensive matrix operations for computing (38), it is assumed in [23] that $\mathbf{w}[k]$ varies slowly in time, i.e. $\mathbf{w}[k] \approx \mathbf{w}[l]$, such that (38) can be approximated without matrix operations as

$$\mathbf{r}[k] = \bar{\lambda}\mathbf{r}[k-1] + (1-\bar{\lambda})\frac{1}{\mu}\left[\mathbf{y}_{B_y}[k]\mathbf{y}_{B_y}^T[k] - \mathbf{v}[k]\mathbf{v}^T[k]\right]\mathbf{w}[k] . \quad (39)$$

[3]In [23] it has been shown that storing noise-only-vectors $\mathbf{v}_i[k]$, $i = M-N \ldots M-1$ during noise-only-periods in a circular *noise-buffer* $\mathbf{B}_v \in \mathbb{R}^{ML \times L_v}$ additionally allows adaptation during speech+noise-periods. However, as will be shown in the next paragraph, this assumption is actually not required in a frequency-domain implementation.

B. Efficient Frequency-Domain (FD) implementation

In [23] the SG-TD algorithm has been converted to a frequency-domain implementation by using a block-formulation and overlap-save procedures (similar to standard FD adaptive filtering techniques [41]). However, the SG-FD algorithm in [23] (**Algorithm 1**) requires the storage of large data buffers (with typical buffer lengths $L_y = 10000 \ldots 20000$). In [24] it has been shown that a substantial memory (and computational complexity) reduction can be achieved by the following two steps:

- When using (38) instead of (39) for calculating the regularisation term, *correlation matrices* instead of data buffers need to be stored. The FD implementation of the total algorithm is then summarised in **Algorithm 2**, where $2L \times 2L$-dimensional speech and noise correlation matrices $\mathbf{S}_y^{ij}[k]$ and $\mathbf{S}_v^{ij}[k]$, $i,j = M-N \ldots M-1$ are used for calculating the regularisation term $\mathbf{R}_i[k]$ and (part of) the step size $\mathbf{\Lambda}[k]$. These correlation matrices are updated respectively during speech-periods and noise-only-periods[4]. However, this first step does not necessarily reduce the memory usage (NL_y for data buffers vs. $2(NL)^2$ for correlation matrices) and will even increase the computational complexity, since the correlation matrices are not diagonal.
- The correlation matrices in the frequency-domain can be approximated by diagonal matrices, since $\mathbf{F}\mathbf{k}^T\mathbf{k}\mathbf{F}^{-1}$ in Algorithm 2 can be well approximated by $\mathbf{I}_{2L}/2$ [42]. Hence, the speech and the noise correlation matrices are updated as

$$\mathbf{S}_y^{ij}[k] = \lambda\mathbf{S}_y^{ij}[k-1] + (1-\lambda)\mathbf{Y}_i^H[k]\mathbf{Y}_j[k]/2 , \quad (40)$$

$$\mathbf{S}_v^{ij}[k] = \lambda\mathbf{S}_v^{ij}[k-1] + (1-\lambda)\mathbf{V}_i^H[k]\mathbf{V}_j[k]/2 , \quad (41)$$

  leading to a significant reduction in memory usage (and computational complexity), cf. Section VI-C. We will refer to this algorithm as **Algorithm 3**. This algorithm is in fact quite similar to [43], which is derived directly from a frequency-domain cost function. Some major differences however exist, e.g. in [43] the regularisation term $\mathbf{R}_i[k]$ is absent, the term $\mathbf{F}\mathbf{g}\mathbf{F}^{-1}$ is also approximated by $\mathbf{I}_{2L}/2$ and the speech and the noise correlation matrices are block-diagonal.

In [24] it has been shown by simulations that approximating the regularisation term in Algorithm 3 only results in a small performance difference (smaller than $0.5\,$dB) in comparison with Algorithm 1. For some scenarios the performance is even better for Algorithm 3 than for Algorithm 1, probably since in Algorithm 1 it is assumed that the filter $\mathbf{w}[k]$ varies slowly in time. Hence, when implementing the SDW-MWF using Algorithm 3, it still preserves its robustness benefit over the GSC (and the QIC-GSC).

C. Memory and computational complexity

Table I summarises the computational complexity and the memory usage for the FD implementation of the QIC-GSC (computed using the NLMS-based Scaled Projection Algorithm (SPA)[5] [5]) and the SDW-MWF (Algorithm 1 and 3). The computational complexity is expressed as the number of operations (i.e. real multiplications and additions (MAC) per second) in MIPS and the memory usage is

[4]When using correlation matrices, filter adaptation can only take place during noise-only-periods, since during speech-periods the desired signal $\mathbf{d}[k]$ cannot be constructed from the noise-buffer $\mathbf{B}_v$ any more.

[5]The complexity of the FD GSC-SPA also represents the complexity when the adaptive filter is only updated during noise-only-periods.

Algorithm 2 FD implementation (without approximation)

Initialisation and matrix definitions:

$\mathbf{W}_i[0] = \left[\begin{array}{cccc} 0 & \cdots & 0 \end{array}\right]^T$, $i = M-N \ldots M-1$

$P_m[0] = \delta_m$, $m = 0 \ldots 2L-1$

$\mathbf{F} = 2L \times 2L$-dimensional DFT matrix

$\mathbf{g} = \left[\begin{array}{cc} \mathbf{I}_L & \mathbf{0}_L \\ \mathbf{0}_L & \mathbf{0}_L \end{array}\right]$, $\quad \mathbf{k} = \left[\begin{array}{cc} \mathbf{0}_L & \mathbf{I}_L \end{array}\right]$

$\mathbf{0}_L = L \times L$ matrix with zeros, $\mathbf{I}_L = L \times L$ identity matrix

For each new block of L samples (per channel):

$\mathbf{d}[k] = \left[\begin{array}{cccc} y_0[kL-\Delta] & \cdots & y_0[kL-\Delta+L-1] \end{array}\right]^T$

$\mathbf{Y}_i[k] = \text{diag}\left\{\mathbf{F}\left[\begin{array}{cccc} y_i[kL-L] & \cdots & y_i[kL+L-1] \end{array}\right]^T\right\}$

Output signal:

$$\mathbf{e}[k] = \mathbf{d}[k] - \mathbf{k}\mathbf{F}^{-1}\sum_{j=M-N}^{M-1}\mathbf{Y}_j[k]\mathbf{W}_j[k], \quad \mathbf{E}[k] = \mathbf{F}\mathbf{k}^T\mathbf{e}[k]$$

If speech detected:

$$\mathbf{S}_y^{ij}[k] = (1-\lambda)\sum_{l=0}^{k}\lambda^{k-l}\mathbf{Y}_i^H[l]\mathbf{F}\mathbf{k}^T\mathbf{k}\mathbf{F}^{-1}\mathbf{Y}_j[l]$$

If noise detected: $\mathbf{V}_i[k] = \mathbf{Y}_i[k]$

$$\mathbf{S}_v^{ij}[k] = (1-\lambda)\sum_{l=0}^{k}\lambda^{k-l}\mathbf{V}_i^H[l]\mathbf{F}\mathbf{k}^T\mathbf{k}\mathbf{F}^{-1}\mathbf{V}_j[l]$$

Update formula (only during noise-only-periods):

$$\mathbf{R}_i[k] = \frac{1}{\mu}\sum_{j=M-N}^{M-1}\left[\mathbf{S}_y^{ij}[k] - \mathbf{S}_v^{ij}[k]\right]\mathbf{W}_j[k]$$

$$\mathbf{W}_i[k+1] = \mathbf{W}_i[k] + \mathbf{F}\mathbf{g}\mathbf{F}^{-1}\mathbf{\Lambda}[k]\left\{\mathbf{V}_i^H[k]\mathbf{E}[k] - \mathbf{R}_i[k]\right\}$$

with

$$\begin{aligned} \mathbf{\Lambda}[k] &= \frac{2\rho'}{L}\,\text{diag}\,\left\{P_0^{-1}[k], \ldots, P_{2L-1}^{-1}[k]\right\} \\ P_m[k] &= \gamma P_m[k-1] + (1-\gamma)\left(P_{v,m}[k] + P_{x,m}[k]\right) \\ P_{v,m}[k] &= \sum_{j=M-N}^{M-1}\left|V_{j,m}[k]\right|^2 \\ P_{x,m}[k] &= \frac{1}{\mu}\left|\sum_{j=M-N}^{M-1}S_{y,m}^{jj}[k] - S_{v,m}^{jj}[k]\right| \end{aligned}$$

expressed in kWords. We assume that one complex multiplication is equivalent to 4 real multiplications and 2 real additions and that a $2L$-point FFT of a real input vector requires $2L\log_2 2L$ real MACs (assuming the radix-2 FFT algorithm). From this table we can draw the following conclusions:

- The *computational complexity* of the SDW-MWF (Algorithm 1) with filter $\mathbf{w}_0$ is about twice the complexity of the GSC-SPA (and even less without $\mathbf{w}_0$). The approximation in the SDW-MWF (Algorithm 3) further reduces the complexity. However, this only remains true for a small number of input channels, since the approximation introduces a quadratic term $\mathcal{O}(N^2)$.
- Due to the storage of the speech+noise-buffer, the *memory usage* of the SDW-MWF (Algorithm 1) is quite high in comparison with the GSC-SPA (depending on the size of the data buffer L_y of course). By using the approximation in the SDW-MWF (Algorithm 3), the memory usage can be drastically reduced. Note however that also for the memory usage a quadratic term $\mathcal{O}(N^2)$ is introduced.

Algorithm	**Complexity**	**MIPS**
GSC-SPA	$(3M-1)$FFT $+ 14M - 12$	2.02
MWF (Algo1)	$(3N+5)$FFT $+ 28N + 6$	$3.10^{(a)}$, $4.13^{(b)}$
MWF (Algo3)	$(3N+2)$FFT $+ 8N^2 + 14N + 3$	$2.54^{(a)}$, $3.98^{(b)}$
	Memory	**kWords**
GSC-SPA	$4(M-1)L + 6L$	0.45
MWF (Algo1)	$2NL_y + 6LN + 7L$	$40.61^{(a)}$, $60.80^{(b)}$
MWF (Algo3)	$4LN^2 + 6LN + 7L$	$1.12^{(a)}$, $1.95^{(b)}$

TABLE I

COMPUTATIONAL COMPLEXITY AND MEMORY USAGE FOR $M = 3$, $L = 32$, $f_s = 16$ KHZ, $L_y = 10000$, (A) $N = M-1$, (B) $N = M$

VII. CONCLUSION

In this paper we have presented a robust multi-microphone noise reduction technique, called the Spatially Pre-processed Speech Distortion Weighted Multi-channel Wiener Filter (SP-SDW-MWF), which consists of a robust fixed spatial pre-preprocessor and a robust adaptive stage. Robustness in the fixed spatial pre-preprocessor is achieved by incorporating statistical information about the microphone characteristics into the design procedure, while robustness in the adaptive stage is achieved by taking speech distortion explicitly into account in the optimisation criterion of the MWF. For the implementation of the adaptive SDW-MWF an efficient stochastic gradient algorithm in the frequency-domain has been developed. Using simulations with hearing aid recordings we have demonstrated the robustness benefit of the presented multi-microphone noise reduction technique against microphone mismatch.

ACKNOWLEDGEMENTS

Simon Doclo is a postdoctoral researcher funded by KULeuven-BOF. This work was supported in part by F.W.O. Project G.0233.01, *Signal processing and automatic patient fitting for advanced auditory prostheses*, I.W.T. Project 020540, *Performance improvement of cochlear implants by innovative speech processing algorithms*, I.W.T. Project 020476, *Sound Management System for Public Address systems (SMS4PA)*, Concerted Research Action *Mathematical Engineering Techniques for Information and Communication Systems* (GOA-MEFISTO-666) of the Flemish Government, Interuniversity Attraction Pole IUAP P5-22 *Dynamical Systems and Control: Computation, Identification and Modelling*, initiated by the Belgian State, Prime Minister's Office - Federal Office for Scientific, Technical and Cultural Affairs, and was partially sponsored by Cochlear.

REFERENCES

[1] H. Cox, R. Zeskind, and T. Kooij, "Practical supergain," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 3, pp. 393–398, June 1986.

[2] R. W. Stadler and W. M. Rabinowitz, "On the potential of fixed arrays for hearing aids," *Journal of the Acoustical Society of America*, vol. 94, no. 3, pp. 1332–1342, Sept. 1993.

[3] C. Sydow, "Broadband beamforming for a microphone array," *Journal of the Acoustical Society of America*, vol. 96, no. 2, pp. 845–849, Aug. 1994.

[4] M. Buck, "Aspects of first-order differential microphone arrays in the presence of sensor imperfections," *European Transactions on Telecommunications, special issue on Acoustic Echo and Noise Control*, vol. 13, no. 2, pp. 115–122, Mar-Apr 2002.

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

[6] M. H. Er, "A robust formulation for an optimum beamformer subject to amplitude and phase perturbations," *Signal Processing*, vol. 19, no. 1, pp. 17–26, 1990.

[7] A. Spriet, M. Moonen, and J. Wouters, "Robustness Analysis of Multichannel Wiener Filtering and Generalized Sidelobe Cancellation for Multi-microphone Noise Reduction in Hearing Aid Applications," *IEEE Trans. on Speech and Audio Processing*, in press, 2004.

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

[9] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction in hearing aids," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Kyoto, Japan, Sept. 2003, pp. 147–150.

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

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

[12] J. E. Greenberg and P. M. Zurek, "Evaluation of an adaptive beamforming method for hearing aids," *Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1662–1676, Mar. 1992.

[13] S. Nordholm, I. Claesson, and B. Bengtsson, "Adaptive Array Noise Suppression of Handsfree Speaker Input in Cars," *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 514–518, Nov. 1993.

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

[15] J. Vanden Berghe and J. Wouters, "An adaptive noise canceller for hearing aids using two nearby microphones," *Journal of the Acoustical Society of America*, vol. 103, pp. 3621–3626, 1998.

[16] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. Signal Processing*, vol. 47, no. 10, pp. 2677–2684, Oct. 1999.

[17] S. Gannot, D. Burshtein, and E. Weinstein, "Signal Enhancement Using Beamforming and Non-Stationarity with Applications to Speech," *IEEE Trans. Signal Processing*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[18] A. Spriet, M. Moonen, and J. Wouters, "A Multi-Channel Subband Generalized Singular Value Decomposition Approach to Speech Enhancement," *European Transactions on Telecommunications, special issue on Acoustic Echo and Noise Control*, vol. 13, no. 2, pp. 149–158, Mar-Apr 2002.

[19] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Processing*, vol. 50, no. 9, pp. 2230–2244, Sept. 2002.

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

[21] S. Doclo and M. Moonen, "Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics," *IEEE Trans. Signal Processing*, vol. 51, no. 10, pp. 2511–2526, Oct. 2003.

[22] S. Doclo and M. Moonen, "Design of broadband beamformers robust against microphone position errors," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Kyoto, Japan, Sept. 2003, pp. 267–270.

[23] A. Spriet, M. Moonen, and J. Wouters, "Stochastic gradient implementation of spatially pre-processed multi-channel Wiener filtering for noise reduction in hearing aids," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, Canada, May 2004.

[24] S. Doclo, A. Spriet, and M. Moonen, "Efficient frequency-domain implementation of speech distortion weighted multi-channel Wiener filtering for noise reduction," Submitted to *European Signal Processing Conference (EUSIPCO)*, Vienna, Austria, Sept. 2004.

[25] S. Van Gerven and F. Xie, "A Comparative Study of Speech Detection Methods," in *Proc. EUROSPEECH*, Rhodos, Greece, Sept. 1997, vol. 3, pp. 1095–1098.

[26] J. Sohn, N. S. Kim, and W. Sung, "A Statistical Model-Based Voice Activity Detection," *IEEE Signal Processing Lett.*, vol. 6, no. 1, pp. 1–3, Jan. 1999.

[27] S. G. Tanyer and H. Özer, "Voice activity detection in nonstationary noise," *IEEE Trans. Speech and Audio Processing*, vol. 8, no. 4, pp. 478–482, July 2000.

[28] S. Doclo and M. Moonen, "Design of far-field and near-field broadband beamformers using eigenfilters," *Signal Processing*, vol. 83, no. 12, pp. 2641–2673, Dec. 2003.

[29] S. Doclo, *Multi-microphone noise reduction and dereverberation techniques for speech applications*, Ph.D. thesis, ESAT, Katholieke Universiteit Leuven, Belgium, May 2003.

[30] S. Nordebo, I. Claesson, and S. Nordholm, "Weighted Chebyshev approximation for the design of broadband beamformers using quadratic programming," *IEEE Signal Processing Lett.*, vol. 1, no. 7, pp. 103–105, July 1994.

[31] M. Kajala and M. Hämäläinen, "Broadband beamforming optimization for speech enhancement in noisy environments," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz NY, USA, Oct. 1999, pp. 19–22.

[32] B. K. Lau, Y. H. Leung, K. L. Teo, and V. Sreeram, "Minimax filters for microphone arrays," *IEEE Trans. Circuits Syst. II*, vol. 46, no. 12, pp. 1522–1525, Dec. 1999.

[33] L. B. Jensen, "Hearing aid with adaptive matching of input transducers," United Stated Patent No. 2002/0041696 A1, Apr. 11 2002.

[34] R. Fletcher, *Practical Methods of Optimization*, Wiley, New York, 1987.

[35] Y. Ephraim and H. L. Van Trees, "A Signal Subspace Approach for Speech Enhancement," *IEEE Trans. Speech and Audio Processing*, vol. 3, no. 4, pp. 251–266, July 1995.

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

[37] K. U. Simmer, J. Bitzer, and C. Marro, *Post-Filtering Techniques*, chapter 3 in "Microphone Arrays: Signal Processing Techniques and Applications" (Brandstein, M. S. and Ward, D. B., Eds.), pp. 39–60, Springer-Verlag, May 2001.

[38] Z. Tian, K. L. Bell, and H. L. Van Trees, "A Recursive Least Squares Implementation for LCMP Beamforming Under Quadratic Constraint," *IEEE Trans. Signal Processing*, vol. 49, no. 6, pp. 1138–1145, June 2001.

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

[40] D. A. Florêncio and H. S. Malvar, "Multichannel filtering for optimum noise reduction in microphone arrays," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Salt Lake City UT, USA, May 2001, pp. 197–200.

[41] J. J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering," *IEEE Signal Processing Magazine*, pp. 15–37, Jan. 1992.

[42] J. Benesty and D. R. Morgan, "Frequency-domain adaptive filtering revisited, generalization to the multi-channel case, and application to acoustic echo cancellation," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Istanbul, Turkey, May 2000, pp. 789–792.

[43] R. Aichner, W. Herbordt, H. Buchner, and W. Kellermann, "Least-squares error beamforming using minimum statistics and multichannel frequency-domain adaptive filtering," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Kyoto, Japan, Sept. 2003, pp. 223–226.