Katholieke Universiteit Leuven

Departement Elektrotechniek

ESAT-SISTA/TR 2003-46a

Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction [1]

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

July 2003

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/sista/spriet/reports/03-46a.pdf.gz*

[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 the Belgian State, Prime Minister's Office - Federal Office for Scientific, Technical and Cultural Affairs - Interuniversity Poles of Attraction Programme (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'), IWT project 020540: 'Innovative Speech Processing Algorithms for Improved Performance of Cochlear Implants' 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.

Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction

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

[1]Katholieke Universiteit Leuven - ESAT/SCD
Kasteelpark Arenberg 10
B-3001 Leuven-Heverlee, Belgium
Tel. +32 16 32 17 95 Fax +32 16 32 19 70
{spriet,moonen}@esat.kuleuven.ac.be

[2]Katholieke Universiteit Leuven
ENT-Dept./Lab Exp ORL
Kapucijnenvoer 33
B-3000 Leuven, Belgium
jan.wouters@uz.kuleuven.ac.be

Abstract

In this paper we establish a generalized noise reduction scheme, called the Spatially Pre-processed Speech Distortion Weighted Multi-channel Wiener Filter (SP-SDW-MWF), that encompasses the Generalized Sidelobe Canceller (GSC) and a recently developed Multi-channel Wiener Filtering technique (MWF) as extreme cases. In addition, the scheme allows for in-between solutions such as the Speech Distortion Regularized GSC (SDR-GSC). The SDR-GSC adds robustness against signal model errors to the GSC by taking speech distortion explicitly into account in the design criterion of the adaptive stage. Compared to the widely studied GSC with Quadratic Inequality Constraint (QIC-GSC), the SDR-GSC achieves better noise reduction for small model errors, while guaranteeing robustness against large model errors. In addition, the extra filtering of the speech reference signal in the SP-SDW-MWF further improves the performance. In the absence of model errors and for infinite filter lengths, the SP-SDW-MWF corresponds to a cascade of an SDR-GSC with a speech distortion weighted single-channel Wiener filter. In contrast to the SDR-GSC and the QIC-GSC, its performance does not degrade due to microphone mismatch.

I. INTRODUCTION

In speech communication applications, such as teleconferencing, hands-free telephony and hearing aids, the presence of background noise and/or reverberation may significantly reduce 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 studied multi-channel adaptive noise reduction algorithm is the *Generalized Sidelobe Canceller* (GSC) [1]-[9]. The GSC consists of a fixed, spatial pre-processor, which includes a fixed beamformer and a blocking matrix, and an Adaptive Noise Canceller (ANC). The ANC should minimize the output noise power while the blocking matrix is designed to avoid so-called speech leakage into the noise references. The standard GSC assumes the desired speaker location, the microphone characteristics and positions to be known, and reflections of the speech signal to be absent. If these assumptions are satisfied, it provides an undistorted enhanced speech signal with minimum residual noise. However, in reality these assumptions are often violated, resulting in so-called speech leakage and hence speech distortion. To limit speech distortion, the ANC is adapted during periods of noise only [5], [8], [10], [11]. When used in combination with small-sized arrays, e.g., in hearing aid applications, an additional robustness constraint [7], [8], [12], [13], [14] is required to guarantee performance in the presence of small errors in the assumed signal model, such as microphone mismatch [15], [16]. However, this constraint goes at the expense of less noise reduction [16].

Recently, a *Multi-channel Wiener Filtering* (MWF) technique has been proposed that provides a Minimum Mean Square Error (MMSE) estimate of the speech component in one of the received microphone signals [17]-[21]. In contrast to the ANC of the GSC, the MWF is able to take speech distortion into account in its optimization criterion. The MMSE optimization criterion can also be generalized to allow for a trade-off between speech distortion and noise reduction [18]. We will refer to this generalization as Speech Distortion Weighted MWF (SDW-MWF). In [17], [18], [19], (recursive) implementations of the (SDW-)MWF are proposed based on a Generalized Singular Value Decomposition (GSVD) or a QR Decomposition (QRD) of an input data matrix and a noise data matrix. A subband implementation [20], [21] results in improved intelligibility at a significantly lower cost compared to the fullband approach, making it suitable for, e.g., hearing aids. The MWF technique 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 contrast to the GSC, the MWF does not make any a priori assumptions about the signal model so that no or a less severe robustness constraint is needed to guarantee performance when used in combination with small-sized arrays [15], [16]. Especially in complex noise scenarios such as multiple noise sources or diffuse noise, the MWF outperforms the GSC, even when the GSC is supplemented with a robustness constraint [16].

In this paper, we establish a common framework for the GSC and SDW-MWF. We show that both algorithms can be integrated into one generalized scheme. This scheme consists of a fixed, spatial pre-processor and an adaptive stage that is based on an SDW-MWF, hence the name *Spatially Pre-processed Speech Distortion Weighted Multi-channel Wiener Filter (SP-SDW-MWF)*. The SP-SDW-MWF encompasses the GSC and the (SDW-)MWF as extreme cases. Hence, it allows to understand the differences and links between the two algorithms more thoroughly. In addition, an in-between solution is found where -like in the GSC- no filtering is applied on the speech reference. This solution can be interpreted as a Speech Distortion Regularized GSC (SDR-GSC): the ANC design criterion is then supplemented with a regularization term that limits speech distortion due to signal model errors. In the literature, the robustness of the GSC against errors in the assumed signal model is often increased by imposing an Quadratic Inequality Constraint (QIC) on the ANC filters [7], [8], [12], [13], at the expense of noise reduction. We show that the SDR-GSC -in contrast to the QIC-GSC- guarantees robustness at large signal model errors while less affecting the noise reduction performance in the presence of small model errors. In addition, the extra filtering of the speech reference signal in the SP-SDW-MWF further improves the performance that is achieved by the SDR-GSC and the QIC-GSC in the presence of signal model errors.

The paper is organized as follows. Section II briefly reviews the GSC and the QIC-GSC. Section III reviews the SDW-MWF technique. In Section IV, we show how both algorithms can be integrated into one generalized scheme, called the Spatially Pre-processed SDW-MWF, with the GSC, the SDR-GSC and the SDW-MWF as particular cases. Section V evaluates the performance of the SP-SDW-MWF and its sensitivity to signal model errors for different parameter settings and compares it with the QIC-GSC and SDR-GSC. In the experiments, we focus on the case of small-sized arrays as used in hearing aids.

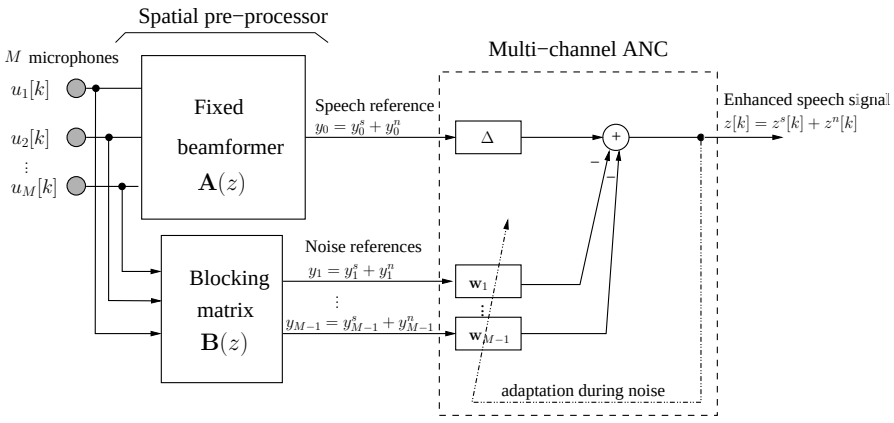


Fig. 1. Concept of the Generalized Sidelobe Canceller.

II. Generalized Sidelobe Canceller (GSC)

A. Concept

Figure 1 describes the concept of the GSC [2], which consists of a fixed, spatial pre-processor, i.e., a fixed beamformer $\mathbf{A}(z)$ and a blocking matrix $\mathbf{B}(z)$, and an ANC. Given M microphone signals

$$u_i[k] = u_i^s[k] + u_i^n[k],\ i = 1,\ ...,\ M \tag{1}$$

with $u_i^s[k]$ the desired speech contribution and $u_i^n[k]$ the noise contribution, the fixed beamformer $\mathbf{A}(z)$ (e.g., delay-and-sum) creates a so-called speech reference

$$y_0[k] = y_0^s[k] + y_0^n[k], \tag{2}$$

by steering a beam towards the direction of the desired signal, with a speech contribution $y_0^s[k]$ and a noise contribution $y_0^n[k]$. In the sequel an endfire array is assumed and the desired speaker is assumed to be in front at 0°. The blocking matrix $\mathbf{B}(z)$ creates $M-1$ so-called noise references

$$y_i[k] = y_i^s[k] + y_i^n[k],\ i = 1,\ ...,\ M-1 \tag{3}$$

by steering zeroes towards the front so that the noise contributions $y_i^n[k]$ are dominant compared to the speech leakage contributions $y_i^s[k]$. In the sequel, the superscripts s and n are used to refer to the speech and noise contribution of a signal. During periods of speech + noise, the references $y_i[k]$, $i = 0,\ ...,\ M-1$ contain speech + noise. During periods of noise only, the references only consist of a noise component, i.e., $y_i[k] = y_i^n[k]$. 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.

To design the fixed, spatial pre-processor, assumptions are made about the microphone characteristics, the speaker position and the microphone positions and furthermore reverberation is assumed to be absent. If these assumptions are satisfied, the noise references do not contain any speech, i.e., $y_i^s[k] = 0$, for $i = 1, ..., M-1$. However, in practice, the assumptions are often violated (e.g. due to microphone mismatch and reverberation) so that speech leaks into the noise references. To limit the effect of such speech leakage, the ANC[1] $\mathbf{w}_{1:M-1} \in \mathcal{C}^{(M-1)L\times 1}$

$$\mathbf{w}_{1:M-1}^H = \left[\begin{array}{cccc} \mathbf{w}_1^H & \mathbf{w}_2^H & ... & \mathbf{w}_{M-1}^H \end{array} \right] \tag{4}$$

where

$$\mathbf{w}_i = \left[\begin{array}{cccc} w_i[0] & w_i[1] & ... & w_i[L-1] \end{array} \right]^T, \tag{5}$$

is adapted during periods of noise only [5], [8], [10], [11]. Hence, the ANC $\mathbf{w}_{1:M-1}$ minimizes the output noise power, i.e.,

$$\boxed{\mathbf{w}_{1:M-1} = \arg\min_{\mathbf{w}_{1:M-1}} \mathcal{E}\{\left|y_0^n[k-\Delta] - \mathbf{w}_{1:M-1}^H[k]\mathbf{y}_{1:M-1}^n[k]\right|^2\}} \tag{6}$$

leading to

$$\boxed{\mathbf{w}_{1:M-1} = \mathcal{E}\{\mathbf{y}_{1:M-1}^n \mathbf{y}_{1:M-1}^{n,H}\}^{-1} \mathcal{E}\{\mathbf{y}_{1:M-1}^n y_0^{n,*}[k-\Delta]\},} \tag{7}$$

where

$$\mathbf{y}_{1:M-1}^{n,H}[k] = \left[\begin{array}{cccc} \mathbf{y}_1^{n,H}[k] & \mathbf{y}_2^{n,H}[k] & ... & \mathbf{y}_{M-1}^{n,H}[k] \end{array} \right] \tag{8}$$

$$\mathbf{y}_i^n[k] = \left[\begin{array}{cccc} y_i^n[k] & y_i^n[k-1] & ... & y_i^n[k-L+1] \end{array} \right]^T \tag{9}$$

and where Δ is a delay applied to the speech reference to allow for non-causal taps in the filter $\mathbf{w}_{1:M-1}$. The delay Δ is usually set to $\lceil \frac{L}{2} \rceil$, where $\lceil x \rceil$ returns the smallest integer equal or larger than x. The subscript $1 : M-1$ in $\mathbf{w}_{1:M-1}$ and $\mathbf{y}_{1:M-1}$ refers to the subscripts of the first and last channel component of the adaptive filter and input vector, respectively. In practice, the ANC is implemented using LMS or RLS updating.

Under ideal conditions ($y_i^s[k] = 0$, $i = 1, ,,,. M-1$), the GSC minimizes the residual noise while not distorting the desired speech signal, i.e., $z^s[k] = y_0^s[k-\Delta]$. However, when used in combination

[1]In a time-domain implementation, the input signals of the adaptive filter $\mathbf{w}_{1:M-1}$ and the filter $\mathbf{w}_{1:M-1}$ are real. Hence, $\mathbf{w}_{1:M-1}^H = \mathbf{w}_{1:M-1}^T$. In the sequel, the formulas are generalized to complex input signals so that they can also be applied to a subband implementation.

with small-sized arrays, a small error in the assumed signal model (hence $y_i^s[k] \neq 0$, $i = 1, ..., M-1$) already suffices to produce a significantly distorted output speech signal $z^s[k]$

$$z^s[k] = y_0^s[k-\Delta] - \mathbf{w}_{1:M-1}^H \mathbf{y}_{1:M-1}^s[k], \tag{10}$$

even when only adapting during noise-only periods, so a robustness constraint on $\mathbf{w}_{1:M-1}$ is required [16]. In addition, the fixed beamformer $\mathbf{A}(z)$ should be designed so that the distortion in the speech reference $y_0^s[k]$ is minimal for all possible model errors. In the sequel, a delay-and-sum beamformer is used. For small-sized arrays, this beamformer offers sufficient robustness against signal model errors, as it minimizes the white noise gain or noise sensitivity [2]. Given statistical knowledge about the signal model errors that occur in practice, further optimized beamformers can be designed, e.g., using the techniques in [22].

B. Quadratic Inequality Constraint (QIC-GSC)

A common approach to increase the robustness of the GSC is to apply a Quadratic Inequality Constraint (QIC) [7]-[12], [13] to the ANC filters $\mathbf{w}_{1:M-1}$, so that the optimization criterion (6) of the GSC is modified into

$$\begin{aligned} \mathbf{w}_{1:M-1} \quad = \quad & \arg \min_{\mathbf{w}_{1:M-1}} \mathcal{E}\{\left|y_0^n[k-\Delta] - \mathbf{w}_{1:M-1}^H[k]\mathbf{y}_{1:M-1}^n[k]\right|^2\} \\ & \text{subject to } \mathbf{w}_{1:M-1}^H \mathbf{w}_{1:M-1} \leq \beta^2. \end{aligned} \tag{11}$$

The QIC avoids excessive growth of the filter coefficients $\mathbf{w}_{1:M-1}$. Hence, it reduces the undesired speech distortion when speech leaks into the noise references. In [12], [13], it is shown that - for a GSC with a blocking matrix $\mathbf{B}(f)$ that satisfies $\mathbf{B}^H(f)\mathbf{B}(f) = \mathbf{I}$ for each frequency f - the QIC on the ANC filters corresponds to a constraint on the noise sensitivity.

In [12], the QIC-GSC is implemented by using the adaptive *scaled projection algorithm*: at each update step, the quadratic constraint is applied to the newly obtained ANC filter by scaling the filter coefficients by $\beta / \left\|\mathbf{w}_{1:M-1}\right\|_2$ when $\mathbf{w}_{1:M-1}^H \mathbf{w}_{1:M-1} = \left\|\mathbf{w}_{1:M-1}\right\|_2^2$ exceeds β^2. Although this technique works well for LMS updating, it does not appear to be as effective for RLS as for LMS [13]. Recently, Tian et al. implemented the quadratic constraint by using *variable loading* [13]. For RLS, this technique provides a

[2]The noise sensitivity is defined as the ratio of the gain of spatially white noise to the gain of the desired signal [1]. Its reciprocal is often referred to as the white noise gain [12]. The noise sensitivity and the white noise gain are often used to quantify the sensitivity of an algorithm against errors in the assumed signal model [1], [12].

better approximation to the optimal solution (11) than the scaled projection algorithm. For LMS, variable loading does not appear to offer any performance advantage over the cheaper, scaled projection LMS.

III. Multi-channel Wiener filtering (MWF)

A. Concept

Recently, an MWF technique has been proposed that provides an MMSE estimate of the speech component in one of the received microphone signals [17], [19], [20], [21]. In contrast to the GSC, this filtering technique does not make any a priori assumptions about the signal model and is found to be more robust [15], [16], [17]. Especially in complex noise scenarios such as multiple noise sources or diffuse noise, the MWF outperforms the GSC, even when the GSC is supplied with a robustness constraint [16].

The MWF $\bar{\mathbf{w}}_{1:M} \in \mathcal{C}^{ML\times 1}$ minimizes the Mean Square Error (MSE) between a delayed version of the (unknown) speech component $u_i^s[k-\Delta]$ at the i-th (e.g., first) microphone signal and the sum $\bar{\mathbf{w}}_{1:M}^H \mathbf{u}_{1:M}[k]$ of the M filtered, received microphone signals, i.e.,

$$\boxed{\bar{\mathbf{w}}_{1:M} = \arg\min_{\bar{\mathbf{w}}_{1:M}} \mathcal{E}\left\{\left|u_i^s[k-\Delta] - \bar{\mathbf{w}}_{1:M}^H \mathbf{u}_{1:M}[k]\right|^2\right\},} \tag{12}$$

leading to

$$\boxed{\bar{\mathbf{w}}_{1:M} = \mathcal{E}\{\mathbf{u}_{1:M}[k]\mathbf{u}_{1:M}^H[k]\}^{-1}\mathcal{E}\{\mathbf{u}_{1:M}[k]u_i^{s,*}[k-\Delta]\},} \tag{13}$$

with

$$\bar{\mathbf{w}}_{1:M}^H = \left[\begin{array}{cccc} \bar{\mathbf{w}}_1^H & \bar{\mathbf{w}}_2^H & \cdots & \bar{\mathbf{w}}_M^H \end{array}\right], \tag{14}$$

$$\bar{\mathbf{w}}_i = \left[\begin{array}{cccc} \bar{\mathbf{w}}_i[0] & \bar{\mathbf{w}}_i[1] & \cdots & \bar{\mathbf{w}}_i[L-1] \end{array}\right]^T, \tag{15}$$

$$\mathbf{u}_{1:M}^H[k] = \left[\begin{array}{cccc} \mathbf{u}_1^H[k] & \mathbf{u}_2^H[k] & \cdots & \mathbf{u}_M^H[k] \end{array}\right], \tag{16}$$

$$\mathbf{u}_i[k] = \left[\begin{array}{cccc} u_i[k] & u_i[k-1] & \cdots & u_i[k-L+1] \end{array}\right]^T. \tag{17}$$

An equivalent approach consists in estimating a delayed version of the (unknown) noise component $u_i^n[k-\Delta]$ in the i-th microphone signal, resulting in

$$\boxed{\mathbf{w}_{1:M} = \arg\min_{\mathbf{w}_{1:M}} \mathcal{E}\left\{\left|u_i^n[k-\Delta] - \mathbf{w}_{1:M}^H \mathbf{u}_{1:M}[k]\right|^2\right\},} \tag{18}$$

and

$$\boxed{\mathbf{w}_{1:M} = \mathcal{E}\{\mathbf{u}_{1:M}[k]\mathbf{u}_{1:M}^H[k]\}^{-1}\mathcal{E}\{\mathbf{u}_{1:M}[k]u_i^{n,*}[k-\Delta]\},} \tag{19}$$

where

$$\mathbf{w}_{1:M}^H = \left[\begin{array}{cccc} \mathbf{w}_1^H & \mathbf{w}_2^H & \cdots & \mathbf{w}_M^H \end{array}\right]. \tag{20}$$

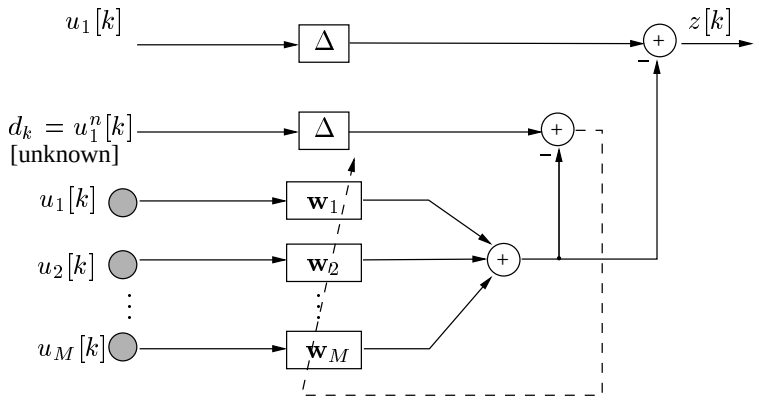


Fig. 2. Concept of multi-channel Wiener filtering.

The estimate $z[k]$ of the speech component $u_i^s[k-\Delta]$ is then obtained by subtracting the estimate $\mathbf{w}_{1:M}^H \mathbf{u}_{1:M}[k]$ of $u_i^n[k-\Delta]$ from the delayed, i-th microphone signal $u_i[k-\Delta]$, i.e.,

$$z[k] = u_i[k-\Delta] - \mathbf{w}_{1:M}^H \mathbf{u}_{1:M}[k]. \tag{21}$$

This is depicted in **Figure 2** for $u_i^n[k-\Delta] = u_1^n[k-\Delta]$. Using (13) and (19), it can be easily shown that

$$\mathbf{w}_{1:M} + \bar{\mathbf{w}}_{1:M} = \mathbf{e}_{(i-1)L+\Delta+1}, \tag{22}$$

with $\mathbf{e}_l$ the l-th canonical $ML \times 1$-dimensional vector, defined as

$$\mathbf{e}_l = \begin{bmatrix} 0 & \cdots & 0 & \underbrace{1}_{\text{position } l} & 0 & \cdots & 0 \end{bmatrix}^T. \tag{23}$$

This shows that the two approaches indeed lead to exactly the same speech signal estimate. A procedure for computing $\mathbf{w}_{1:M}$ or $\bar{\mathbf{w}}_{1:M}$ will be given in Section III-C.

B. Trade-off speech distortion versus noise reduction (SDW-MWF)

The residual error energy equals

$$\mathcal{E}\{|e[k]|^2\} = \mathcal{E}\{\left|u_i^s[k-\Delta] - \bar{\mathbf{w}}_{1:M}^H \mathbf{u}_{1:M}[k]\right|^2\}, \tag{24}$$

and can be decomposed into

$$\underbrace{\mathcal{E}\{\left|u_i^s[k-\Delta] - \bar{\mathbf{w}}_{1:M}^H \mathbf{u}_{1:M}^s[k]\right|^2\}}_{\epsilon_d^2} + \underbrace{\mathcal{E}\{\left|\bar{\mathbf{w}}_{1:M}^H \mathbf{u}_{1:M}^n[k]\right|^2\}}_{\epsilon_n^2} \tag{25}$$

where ϵ_d^2 equals the speech distortion energy and ϵ_n^2 the residual noise energy. The design criterion of the MWF can be generalized to allow for a trade-off between speech distortion and noise reduction, by incorporating a weighting factor μ [18] with $\mu \in [0, \infty]$

$$\boxed{\bar{\mathbf{w}}_{1:M} = \arg\min_{\bar{\mathbf{w}}_{1:M}} \mathcal{E}\{\left|u_i^s[k-\Delta] - \bar{\mathbf{w}}_{1:M}^H \mathbf{u}_{1:M}^s[k]\right|^2\} + \mu\mathcal{E}\{\left|\bar{\mathbf{w}}_{1:M}^H \mathbf{u}_{1:M}^n[k]\right|^2\}.} \tag{26}$$

The solution of (26) is given by

$$\boxed{\bar{\mathbf{w}}_{1:M} = \mathcal{E}\{\mathbf{u}_{1:M}^s[k]\mathbf{u}_{1:M}^{s,H}[k] + \mu\mathbf{u}_{1:M}^n[k]\mathbf{u}_{1:M}^{n,H}[k]\}^{-1}\mathcal{E}\{\mathbf{u}_{1:M}^s[k]u_i^{s,*}[k-\Delta]\},} \tag{27}$$

which corresponds to the Wiener formula (13) with an adjustable input noise level. Note that (13) is obtained with $\mu = 1$. The filter (27) corresponds to the time domain constrained estimator proposed in [23], which optimizes the following criterion:

$$\min_{\bar{\mathbf{w}}} \epsilon_d^2 \text{ subject to } \epsilon_n^2 \leq \alpha\mathcal{E}\{\mathbf{u}_{1:M}^{n,H}\mathbf{u}_{1:M}^n\} \tag{28}$$

where $0 \leq \alpha \leq 1$ and μ is the Lagrange-multiplier.

Equivalently, the optimization criterion for $\mathbf{w}$ in (19) can be modified into

$$\boxed{\mathbf{w}_{1:M} = \arg\min_{\mathbf{w}_{1:M}} \mathcal{E}\{\left|\mathbf{w}_{1:M}^H \mathbf{u}_{1:M}^s[k]\right|^2\} + \mu\mathcal{E}\{\left|u_i^n[k-\Delta] - \mathbf{w}_{1:M}^H \mathbf{u}_{1:M}^n[k]\right|^2\},} \tag{29}$$

resulting in

$$\boxed{\mathbf{w}_{1:M} = \mathcal{E}\{\mathbf{u}_{1:M}^n[k]\mathbf{u}_{1:M}^{n,H}[k] + \frac{1}{\mu}\mathbf{u}_{1:M}^s[k]\mathbf{u}_{1:M}^{s,H}[k]\}^{-1}\mathcal{E}\{\mathbf{u}_{1:M}^n[k]u_i^{n,*}[k-\Delta]\}.} \tag{30}$$

Note that (22) still applies. In the sequel, we will refer to (27)-(30) as the *Speech Distortion Weighted* Multi-channel Wiener Filter (SDW-MWF).

The factor $\mu \in [0, \infty]$ trades off speech distortion versus noise reduction. If $\mu = 1$, the MMSE criterion (12) or (18) is obtained. If $\mu > 1$, the residual noise level will be reduced at the expense of increased speech distortion. By setting μ to ∞, all emphasis is put on noise reduction and speech distortion is completely ignored. This results in $\bar{\mathbf{w}}_{1:M} = 0$ or $\mathbf{w}_{1:M} = \mathbf{e}_{(i-1)L+\Delta+1}$, which means that the output signal equals 0. Setting μ to 0 on the other hand, results in $\bar{\mathbf{w}}_{1:M} = \mathbf{e}_{(i-1)L+\Delta+1}$ or $\mathbf{w}_{1:M} = 0$ and hence in no noise reduction.

C. Implementation of MWF

In practice, the correlation matrix $\mathcal{E}\{\mathbf{u}_{1:M}^s[k]\mathbf{u}_{1:M}^{s,H}[k]\}$ is unknown. During periods of speech, the inputs $u_i[k]$ consist of speech + noise, i.e., $u_i[k] = u_i^s[k] + u_i^n[k]$, $i = 1, ..., M$. During periods of noise, only the noise component $u_i^n[k]$ is observed. The second order statistics of the noise are assumed to be sufficiently

stationary so that they can be estimated during periods of noise only. Assuming that the speech and noise signal are uncorrelated, $\mathcal{E}\{\mathbf{u}^s_{1:M}[k]\mathbf{u}^{s,H}_{1:M}[k]\}$ can be estimated as

$$\mathcal{E}\{\mathbf{u}^s_{1:M}[k]\mathbf{u}^{s,H}_{1:M}[k]\} = \mathcal{E}\{\mathbf{u}_{1:M}[k]\mathbf{u}^H_{1:M}[k]\} - \mathcal{E}\{\mathbf{u}^n_{1:M}[k]\mathbf{u}^{n,H}_{1:M}[k]\}, \tag{31}$$

where the second order statistics $\mathcal{E}\{\mathbf{u}_{1:M}[k]\mathbf{u}^H_{1:M}[k]\}$ are estimated during speech + noise and the statistics $\mathcal{E}\{\mathbf{u}^n_{1:M}[k]\mathbf{u}^{n,H}_{1:M}[k]\}$ during periods of noise only. Like for the GSC, a robust speech detection is thus needed. Using (31), (27) and (30) can be re-written as:

$$\begin{aligned}\bar{\mathbf{w}}_{1:M} &= \Big(\mathcal{E}\{\mathbf{u}_{1:M}[k]\mathbf{u}^H_{1:M}[k]\} + (\mu-1)\mathcal{E}\{\mathbf{u}^n_{1:M}[k]\mathbf{u}^{n,H}_{1:M}[k]\}\Big)^{-1} \\ &\quad \times \big(\mathcal{E}\{\mathbf{u}_{1:M}[k]u_i^*[k-\Delta]\} - \mathcal{E}\{\mathbf{u}^n_{1:M}[k]u_i^{n,*}[k-\Delta]\}\big)\end{aligned} \tag{32}$$

and

$$\mathbf{w}_{1:M} = \left(\frac{1}{\mu}\mathcal{E}\{\mathbf{u}_{1:M}[k]\mathbf{u}^H_{1:M}[k]\} + (1-\frac{1}{\mu})\mathcal{E}\{\mathbf{u}^n_{1:M}[k]\mathbf{u}^{n,H}_{1:M}[k]\}\right)^{-1} \mathcal{E}\{\mathbf{u}^n_{1:M}[k]u_i^{n,*}[k-\Delta]\}. \tag{33}$$

In [17], the Wiener filter is computed at each time instant k by means of a Generalized Singular Value Decomposition (GSVD) of a speech + noise and noise data matrix. A cheaper recursive alternative based on a QR-decomposition has been proposed in [24]. In [20], [21], a subband implementation has been developed to increase the resulting speech intelligibility and reduce complexity, making it suitable for hearing aid applications.

Finally note that instead of estimating $\mathcal{E}\{\mathbf{u}^s_{1:M}[k]\mathbf{u}^{s,H}_{1:M}[k]\}$ online using (31), a pre-determined estimate of $\mathcal{E}\{\mathbf{u}^s_{1:M}[k]\mathbf{u}^{s,H}_{1:M}[k]\}$ is sometimes used [25], [26]. In [25], this estimate is derived from clean speech recordings measured during an initial calibration phase. Additional recordings of the source speech signal allow to produce an estimate of the non-reverberant source speech signal instead of an estimate of the reverberant speech component in one of the microphone signals. However, since the room acoustics, the position of desired speaker and microphone characteristics may change over time, frequent re-calibration is required. In [26], a mathematical estimate of the correlation matrix and the correlation vector of the non-reverberant speech is exploited in which some signal model errors are taken into account.

IV. Spatially pre-processed SDW-MWF (SP-SDW-MWF)

A. Concept

The SDW-MWF depicted in Figure 2 has a structure similar to the multi-channel ANC block of Figure 1. Replacing the ANC in the GSC by the SDW-MWF results in the generalized scheme that is

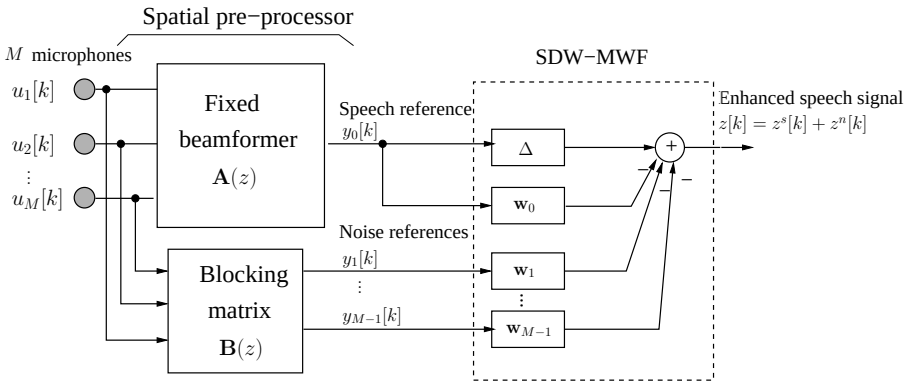


Fig. 3. Spatially Pre-processed SDW MWF.

depicted in Figure 3. In the sequel, we will refer to this generalized scheme as the *Spatially Pre-processed, Speech Distortion Weighted Multi-channel Wiener Filter (SP-SDW-MWF)*. The input signals to the multi-channel Wiener filter are now spatially pre-processed microphone signals $y_i[k]$, $0 \leq i \leq M-1$ and the speech contribution $y_0^s[k]$ at the output of the fixed beamformer is estimated. To preserve the robustness advantage of the MWF, the fixed beamformer $\mathbf{A}(z)$ should be robust against possible signal model errors. In the sequel a delay-and-sum beamformer is used. As mentioned before, this beamformer is sufficiently robust when used for small-sized arrays. Again, given statistical knowledge about the signal model errors that occur in practice, a further optimized robust fixed beamformer can be designed e.g., using [22].

Using (30), the expression for $\mathbf{w}_{0:M-1}$ becomes

$$\boxed{\mathbf{w}_{0:M-1} = \left(\frac{1}{\mu}\mathcal{E}\{\mathbf{y}^s_{0:M-1}[k]\mathbf{y}^{s,H}_{0:M-1}[k]\} + \mathcal{E}\{\mathbf{y}^n_{0:M-1}[k]\mathbf{y}^{n,H}_{0:M-1}[k]\}\right)^{-1}\mathcal{E}\{\mathbf{y}^n_{0:M-1}[k]y_0^{n,*}[k-\Delta]\},} \quad (34)$$

where

$$\mathbf{w}^H_{0:M-1} = \left[\begin{array}{cccc} \mathbf{w}_0^H & \mathbf{w}_1^H & ... & \mathbf{w}_{M-1}^H \end{array}\right] \quad (35)$$

$$\mathbf{y}^H_{0:M-1}[k] = \left[\begin{array}{cccc} \mathbf{y}_0^H[k] & \mathbf{y}_1^H[k] & ... & \mathbf{y}_{M-1}^H[k] \end{array}\right] \quad (36)$$

$$\mathbf{y}^{n,H}_{0:M-1}[k] = \left[\begin{array}{cccc} \mathbf{y}_0^{n,H}[k] & \mathbf{y}_1^{n,H}[k] & ... & \mathbf{y}_{M-1}^{n,H}[k] \end{array}\right] \quad (37)$$

$$\mathbf{y}^{s,H}_{0:M-1}[k] = \left[\begin{array}{cccc} \mathbf{y}_0^{s,H}[k] & \mathbf{y}_1^{s,H}[k] & ... & \mathbf{y}_{M-1}^{s,H}[k] \end{array}\right] \quad (38)$$

and where μ trades off distortion of the speech reference $y_0^s[k]$ versus noise reduction. For $\mu = 0$, $\mathbf{w}_{0:M-1} = \mathbf{0}$ and so $z[k]$ is equal to the output of the fixed beamformer $\mathbf{A}(z)$ delayed by Δ samples. In noise scenarios with very low Signal-to-Noise Ratio (SNR), e.g., -10 dB, a fixed beamformer output may

be preferred: for low SNR, the error rate of the speech detection mechanism may be too high, resulting in unacceptable speech distortion by the adaptive filter [27]. Adaptivity can then easily be reduced or excluded in the SP-SDW-MWF by decreasing μ to 0. Alternatively, adaptivity can be limited by applying an QIC to $\mathbf{w}_{0:M-1}$.

Note that when the fixed beamformer $\mathbf{A}(z)$ and the blocking matrix $\mathbf{B}(z)$ are set to

$$\mathbf{A}(z) = \begin{bmatrix} 1 & 0 & ... & 0 \end{bmatrix}^H \tag{39}$$

$$\mathbf{B}(z) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{bmatrix}^H, \tag{40}$$

we obtain the original SDW-MWF that operates on the received microphone signals $u_i[k]$, $i = 1, ..., M$.

Below, the different parameter settings of the SP-SDW-MWF are discussed. Depending on the setting of the parameter μ and presence or absence of the filter $\mathbf{w}_0$, the GSC, the (SDW-)MWF as well as in-between solutions may be obtained. We distinguish between two cases, i.e., the case where no filter $\mathbf{w}_0$ is applied to the speech reference (filter length $L_0 = 0$) and the case where an additional filter $\mathbf{w}_0$ is used ($L_0 \neq 0$).

The adaptive stage of the SP-SDW-MWF can be implemented using the recursive QRD-based implementation of the SDW-MWF [24]. Like for the SDW-MWF, complexity can be reduced by a subband implementation [20]. For $L_0 \neq 0$, also the GSVD based algorithm [18] can be applied. Also, cheaper stochastic gradient based algorithms have been developed [28].

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

First, consider *the case without* $\mathbf{w}_0$, i.e. $L_0 = 0$. The solution for $\mathbf{w}_{1:M-1}$ in (34) then reduces to

$$\boxed{\arg \min_{\mathbf{w}_{1:M-1}} \frac{1}{\mu} \underbrace{\mathcal{E}\{\left|\mathbf{w}_{1:M-1}^H \mathbf{y}_{1:M-1}^s[k]\right|^2\}}_{\varepsilon_d^2} + \underbrace{\mathcal{E}\{\left|y_0^n[k-\Delta] - \mathbf{w}_{1:M-1}^H \mathbf{y}_{1:M-1}^n[k]\right|^2\}}_{\varepsilon_n^2},} \tag{41}$$

leading to

$$\boxed{\mathbf{w}_{1:M-1} = \left(\frac{1}{\mu}\mathcal{E}\{\mathbf{y}_{1:M-1}^s \mathbf{y}_{1:M-1}^{s,H}[k]\} + \mathcal{E}\{\mathbf{y}_{1:M-1}^n \mathbf{y}_{1:M-1}^{n,H}[k]\}\right)^{-1} \mathcal{E}\{\mathbf{y}_{1:M-1}^n[k] y_0^{n,*}[k-\Delta]\},} \tag{42}$$

where ε_d^2 is the speech distortion energy and ε_n^2 the residual noise energy.

Remark: For $L_0 = 0$, it is readily seen that (22) does not hold, i.e., $\mathbf{w}_{1:M-1} + \bar{\mathbf{w}}_{1:M-1} \neq \mathbf{e}_{\Delta+1}$ where

$$\bar{\mathbf{w}}_{1:M-1} = \left(\mathcal{E}\{\mathbf{y}_{1:M-1}^s \mathbf{y}_{1:M-1}^{s,H}\} + \mu\mathcal{E}\{\mathbf{y}_{1:M-1}^n \mathbf{y}_{1:M-1}^{n,H}\}\right)^{-1} \mathcal{E}\{\mathbf{y}_{1:M-1}^{s,H} y_0^{s,*}[k-\Delta]\}, \tag{43}$$

because the speech component $\mathbf{y}_{1:M-1}^s[k]$ *in the input to the adaptive filter* $\mathbf{w}_{1:M-1}$ *does not contain the estimated speech signal* $y_0^s[k-\Delta]$.

Compared to the optimization criterion (6) of the GSC, a regularization term

$$\frac{1}{\mu}\mathcal{E}\{\left|\mathbf{w}_{1:M-1}^H\mathbf{y}_{1:M-1}^s[k]\right|^2\} \tag{44}$$

has been added. This regularization term limits the amount of speech distortion that is caused by the filter $\mathbf{w}_{1:M-1}$ when speech leaks into the noise references, i.e., $y_i^s[k] \neq 0$, $i = 1, ..., M-1$. In the sequel, we therefore refer to the SP-SDW-MWF with $L_0 = 0$ as the *Speech Distortion Regularized GSC (SDR-GSC)*. The smaller μ, the smaller the resulting amount of speech distortion will be. For $\mu = 0$, all emphasis is put on speech distortion and so $z[k]$ is equal to the output of the fixed beamformer $\mathbf{A}(z)$ delayed by Δ samples. For $\mu = \infty$, all emphasis is put on noise reduction and speech distortion is not taken into account. This corresponds to the GSC. Hence, the SDR-GSC encompasses the GSC as a special case.

The regularization term (44) with $\frac{1}{\mu} \neq 0$ 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.,* $y_i^s[k] = 0$, $i = 1, ..., M-1$, the regularization term equals $\mathbf{0}$ for all $\mathbf{w}_{1:M-1}$ and hence the residual noise energy ε_n^2 is effectively minimized. In other words, in the absence of speech leakage, the GSC solution is obtained.
- In the *presence of speech leakage*, i.e., $y_i^s[k] \neq 0$, $i = 1, ..., M-1$, speech distortion is explicitly taken into account in the optimization criterion (41) for the adaptive filter $\mathbf{w}_{1:M-1}$, limiting speech distortion while reducing noise. The larger the amount of speech leakage, the more attention is paid to speech distortion.

  To limit speech distortion alternatively, an QIC is often imposed on the filter $\mathbf{w}_{1:M-1}$ (see Section II-B). In contrast to the SDR-GSC, the QIC acts irrespective of the amount of speech leakage $\mathbf{y}_{1:M-1}^s[k]$ that is present. The constraint value β^2 in (11) has to be chosen based on the largest model errors that may occur. As a consequence, noise reduction performance is compromised even when no or very small model errors are present. Hence, the QIC is more conservative than the SDR-GSC. The experimental results in Section V confirm this.

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

Since the SDW-MWF (34) takes speech distortion explicitly into account in its optimization criterion, an additional filtering $\mathbf{w}_0$ on the speech reference $y_0[k]$ may be added (see Figure 3). The SDW-MWF (34) then solves the following more general optimization criterion

$$\mathbf{w}_{0:M-1} = \arg\min_{\mathbf{w}_{0:M-1}} \underbrace{\mathcal{E}\left\{\left|y_0^n[k-\Delta] - \begin{bmatrix} \mathbf{w}_0^H & \mathbf{w}_{1:M-1}^H \end{bmatrix} \begin{bmatrix} \mathbf{y}_0^n[k] \\ \mathbf{y}_{1:M-1}^n[k] \end{bmatrix}\right|^2\right\}}_{\varepsilon_n^2} + \frac{1}{\mu}\underbrace{\mathcal{E}\left\{\left|\begin{bmatrix} \mathbf{w}_0^H & \mathbf{w}_{1:M-1}^H \end{bmatrix} \begin{bmatrix} \mathbf{y}_0^s[k] \\ \mathbf{y}_{1:M-1}^s[k] \end{bmatrix}\right|^2\right\}}_{\varepsilon_d^2}, \tag{45}$$

where $\mathbf{w}_{0:M-1}^H = [\mathbf{w}_0^H \ \mathbf{w}_{1:M-1}^H]$ is given by (34).

Again, μ trades off speech distortion and noise reduction. For $\mu = \infty$, speech distortion ϵ_d^2 is completely ignored so that the solution becomes

$$\begin{aligned} \mathbf{w}_{0:M-1}^H &= \begin{bmatrix} \mathbf{w}_0^H & \mathbf{w}_{1:M-1}^H \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{e}_{\Delta+1}^H & \mathbf{0}^H \end{bmatrix}, \end{aligned} \tag{46}$$

which results in a zero output signal $z[k]$. For $\mu = 0$, all emphasis is put on speech distortion so that $z[k]$ is equal to the output of the fixed beamformer delayed by Δ samples.

In addition, we can make the following statements:

- In the *absence of speech leakage*, i.e., $y_i^s[k] = 0$ for $i = 1, ..., M-1$, and for infinitely long filters $\mathbf{w}_i$, $i = 0, ..., M-1$, the SP-SDW-MWF with $\mathbf{w}_0$ corresponds to a *cascade of a SDR-GSC and a SDW Single-channel WF (SDW-SWF) post-processor* [29], [30].

  Proof: In case of infinite filter lengths, the SP-SDW-MWF and its optimization criterion can be represented in the frequency domain[3]. For simplicity, but without loss of generality, we assume $\Delta = 0$.

$$\begin{aligned} \mathbf{W}_{0:M-1}(f) = \arg\min_{\mathbf{W}_{0:M-1}} &\mathcal{E}\left\{\left|\begin{bmatrix} (1-W_0^*) & -\mathbf{W}_{1:M-1}^H \end{bmatrix} \begin{bmatrix} Y_0^n(f) \\ \mathbf{Y}_{1:M-1}^n(f) \end{bmatrix}\right|^2\right\} \\ &+\frac{1}{\mu}\mathcal{E}\left\{\left|\begin{bmatrix} W_0^* & \mathbf{W}_{1:M-1}^H \end{bmatrix} \begin{bmatrix} Y_0^s(f) \\ \mathbf{Y}_{1:M-1}^s(f) \end{bmatrix}\right|^2\right\} \end{aligned} \tag{47}$$

  Decompose $\mathbf{W}_{1:M-1}(f)$ as

$$\mathbf{W}_{1:M-1}(f) = (1 - W_0(f))\,\mathbf{W}_d(f) \tag{48}$$

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

with $W_0(f)$ the single-channel filter applied to the speech reference and $\mathbf{W}_d(f)$ a multi-channel filter and define an intermediate output $V(f)$ (see also Figure 4) as

$$V(f) = Y_0(f) - \mathbf{W}_d^H(f)\mathbf{Y}_{1:M-1}(f). \tag{49}$$

Then, the cost function $J(W_0, \mathbf{W}_d)$ of (47) can be re-written as

$$J = \mathcal{E}\left\{\left|(1 - W_0^*)\, V^n(f)\right|^2\right\} + \frac{1}{\mu}\mathcal{E}\left\{\left|W_0^* V^s(f) + \mathbf{W}_d^H(f)\mathbf{Y}_{1:M-1}^s(f)\right|^2\right\}. \tag{50}$$

From $\frac{\partial}{\partial W_0} J(W_0, \mathbf{W}_d) = 0$, we find

$$\begin{aligned} W_0(f) &= \underbrace{\left(\mathcal{E}\{V^n V^{n,*}\} + \frac{1}{\mu}\mathcal{E}\{V^s V^{s,*}\}\right)^{-1} \mathcal{E}\{V^n V^{n,*}\}}_{W_{0,1}(f)} \\ &\quad + \underbrace{\left(\mu\mathcal{E}\{V^n V^{n,*}\} + \mathcal{E}\{V^s V^{s,*}\}\right)^{-1} \left(-\mathcal{E}\{V^s \mathbf{Y}_{1:M-1}^{s,H}\mathbf{W}_d\}\right)}_{W_{0,2}(f)} \end{aligned} \tag{51}$$

This *single-channel filter* $W_0(f)$ thus consists of two terms.

- The first term $W_{0,1}(f)$ estimates the noise component $V^n(f)$ in the intermediate output $V(f)$. The filter $1 - W_{0,1}$ then corresponds to an SDW-SWF that estimates the speech component $V^s(f)$ in the intermediate output $V(f)$.
- The second term $W_{0,2}(f)$ estimates the speech leakage filtered by $\mathbf{W}_d(f)$, i.e., $-\mathbf{W}_d^H\mathbf{Y}_{1:M-1}^s$. The speech component in the intermediate output $V(f)$ equals $V^s(f) = Y_0^s - \mathbf{W}_d^H\mathbf{Y}_{1:M-1}^s$. The filter $W_{0,2}(f)$ thus tries to compensate for the distortion $-\mathbf{W}_d^H\mathbf{Y}_{1:M-1}^s$ by adding an estimate of $\mathbf{W}_d^H\mathbf{Y}_{1:M-1}^s$ to the output of the SDW-SWF.In the absence of speech leakage (i.e., $\mathbf{Y}_{1:M-1}^s = \mathbf{0}$), the filter $W_{0,2}(f)$ equals zero.

From $\frac{\partial}{\partial \mathbf{W}_d} J(W_0, \mathbf{W}_d) = 0$, we obtain the following solution for $\mathbf{W}_d(f)$:

$$\begin{aligned} \mathbf{W}_d(f) &= \underbrace{\left(\mathcal{E}\{\mathbf{Y}_{1:M-1}^n \mathbf{Y}_{1:M-1}^{n,H}\} + \frac{1}{\mu}\mathcal{E}\{\mathbf{Y}_{1:M-1}^s \mathbf{Y}_{1:M-1}^{s,H}\}\right)^{-1} \mathcal{E}\{\mathbf{Y}_{1:M-1}^n Y_0^{n,*}\}}_{\mathbf{W}_{d,1}(f)} \\ &\quad \underbrace{- \left(\mu\mathcal{E}\{\mathbf{Y}_{1:M-1}^n \mathbf{Y}_{1:M-1}^{n,H}\} + \mathcal{E}\{\mathbf{Y}_{1:M-1}^s \mathbf{Y}_{1:M-1}^{s,H}\}\right)^{-1} \mathcal{E}\{\mathbf{Y}_{1:M-1}^s Y_0^{s,*} \frac{W_0}{1 - W_0}\}}_{\mathbf{W}_{d,2}(f)}. \end{aligned} \tag{52}$$

This *multi-channel filter* $\mathbf{W}_d(f)$ consists of two terms.

- The first term $\mathbf{W}_{d,1}(f)$ corresponds to the SDR-GSC and estimates the noise component $Y_0^n(f)$ at the output of the fixed beamformer.
- The second term $\mathbf{W}_{d,2}(f)$ tries to compensate for the speech distortion $-W_0^*(f)Y_0^s(f)$ caused by $W_0(f)$ by adding an estimate of $\frac{W_0^*(f)}{1-W_0^*(f)}Y_0^s(f)$ to the output of the SDR-GSC. Note that this

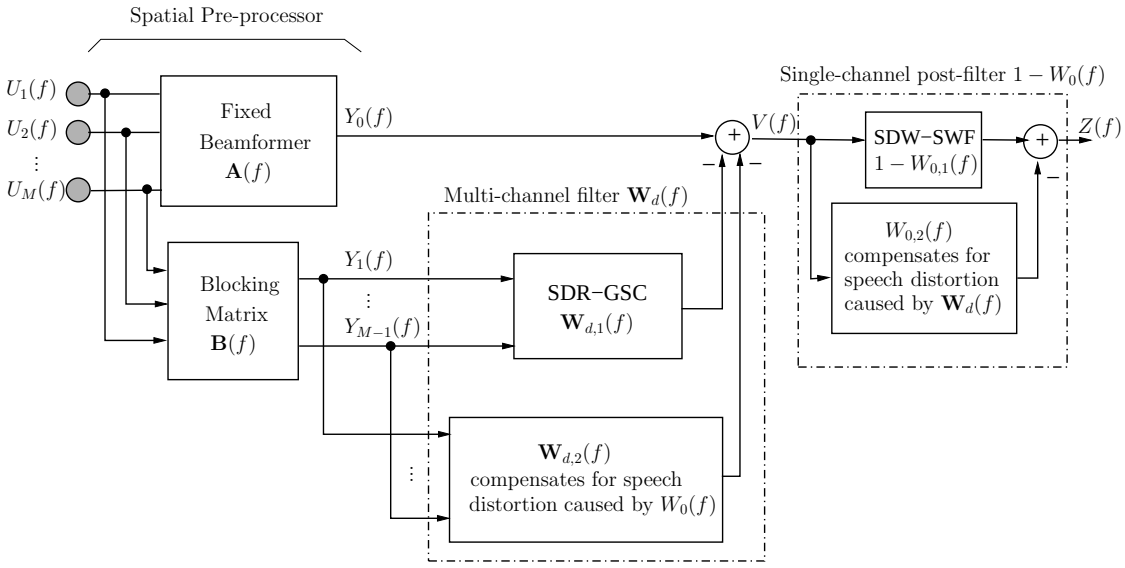


Fig. 4. Decomposition of SP-SDW-MWF with $\mathbf{w}_0$ in a multi-channel filter $\mathbf{w}_d$ and single-channel post-filter $\mathbf{e}_1 - \mathbf{w}_0$.

corresponds to adding an estimate of $W_0^*(f)Y_0^s(f)$ to the output $Z(f)$ of the SP-SDW-MWF. In the absence of speech leakage, $\mathbf{W}_{d,2}(f) = \mathbf{0}$.

Figure 4 graphically illustrates the solution for $\mathbf{W}_d(f)$ and $W_0(f)$ for $\Delta = 0$. In the absence of speech leakage, the filters $W_{0,2}(f)$ and $\mathbf{W}_{d,2}(f)$ equal zero, hence, the SP-SDW-MWF corresponds to an SDR-GSC (or GSC) cascaded with an SDW-SWF post-processor. This statement can also be proved in the time-domain, assuming sufficiently long filters $\mathbf{w}_{0:M-1}$ (see Appendix). The SDW-SWF increases noise suppression at the expense of speech distortion, especially for large μ and at frequencies where the Signal-to-Noise Ratio (SNR) of the intermediate output $V(f)$ is low. ■

- In the *presence of speech leakage,* the SP-SDW-MWF with $\mathbf{w}_0$ tries to preserve its performance: compared to an SDR-GSC with SDW-SWF post-processor, the SP-SDW-MWF then contains extra filtering operations (i.e., $W_{0,2}(f)$ and $\mathbf{W}_{d,2}(f)$) that compensate for the performance degradation of the SDR-GSC with SDW-SWF due to speech leakage (see Figure 4 and the proof above). We will consider the effects of microphone mismatch as an example: *for infinite filter lengths, the performance of the SP-SDW-MWF with* $\mathbf{w}_0$ *is not affected by microphone mismatch as long as the desired speech component at the output of the fixed beamformer* $\mathbf{A}(z)$ *remains unaltered.*

  Proof: In case of infinite filter lengths, the SDW-MWF filter $\mathbf{w}_{0:M-1}$ in (34) can be represented

in the frequency domain as[4]

$$\mathbf{W}(f) = \left(\mathcal{E}\{\mathbf{Y}^n\mathbf{Y}^{n,H}\} + \frac{1}{\mu}\mathcal{E}\{\mathbf{Y}^s\mathbf{Y}^{s,H}\}\right)^{-1}\mathcal{E}\{\mathbf{Y}^n Y_0^{n,*}\mathrm{e}^{\mathrm{i}2\pi f\Delta}\}. \tag{53}$$

Suppose that the received microphone signals $\tilde{U}_i(f)$, with $i = 1, ..., M$ are perturbed by $\Upsilon_i(f) = \upsilon_i \mathrm{e}^{\mathrm{i}\phi_i(f)}$ with υ_i a gain deviation and $\phi_i(f)$ a phase deviation, with respect to the signals $U_i(f)$ received in the absence of mismatch, i.e.,

$$\tilde{\mathbf{U}}_{1:M}(f) = \mathrm{diag}\{\Upsilon_1(f), ..., \Upsilon_M(f)\}\mathbf{U}_{1:M}(f), \tag{54}$$

with

$$\tilde{\mathbf{U}}_{1:M}(f) = [\tilde{U}_1(f) \cdots \tilde{U}_M(f)]^T \tag{55}$$

and with $\mathrm{diag}\{\Upsilon_1(f), ..., \Upsilon_M(f)\}$ of full rank, i.e., $\Upsilon_i(f) \neq 0$. Then, - assuming that the matrix $[\ \mathbf{A}(f) \quad \mathbf{B}(f)\ ]$ is of full rank- the input vector $\tilde{\mathbf{Y}}(f)$ to the adaptive filter $\mathbf{W}(f)$ can be described as

$$\tilde{\mathbf{Y}}(f) = \mathbf{T}(f)\mathbf{Y}(f) \tag{56}$$

where $\mathbf{Y}(f)$ is the input vector in the absence of mismatch and $\mathbf{T}(f) \in \mathbb{C}^{M\times M}$ is a full rank matrix that depends on $\mathbf{A}(f)$, $\mathbf{B}(f)$ and $\Upsilon_i(f)$. Plugging (56) into (53), we obtain a modified filter

$$\tilde{\mathbf{W}}(f) = \mathbf{T}^{-H}\left(\mathcal{E}\{\mathbf{Y}^n\mathbf{Y}^{n,H}\} + \frac{1}{\mu}\mathcal{E}\{\mathbf{Y}^s\mathbf{Y}^{s,H}\}\right)^{-1}\mathcal{E}\{\mathbf{Y}^n\mathbf{Y}^{n,H}\}\mathbf{T}^H(:,1)\mathrm{e}^{\mathrm{i}2\pi f\Delta}. \tag{57}$$

Hence, the SP-SDW-MWF compensates for the mismatch through the factor $\mathbf{T}^{-H}$. If the desired speech component at the output of the fixed beamformer $\mathbf{A}(z)$ remains unaltered, i.e., if

$$\mathbf{Y}^{s,H}(f)\mathbf{T}^H(:,1) = Y_0^{s,*}(f), \tag{58}$$

the output $Z(f)$ of the SP-SDW-MWF is not affected by the mismatch. Indeed, using (57), (58) and the frequency-domain equivalent of (22), we find

$$\begin{aligned} Z(f) &= \left(\mathrm{e}^{-\mathrm{i}2\pi f\Delta}\mathbf{T}(1,:)\mathbf{Y}(f) - \tilde{\mathbf{W}}^H(f)\tilde{\mathbf{Y}}(f)\right) \\ &= \mathrm{e}^{-\mathrm{i}2\pi f\Delta}\mathbf{T}(1,:)\frac{1}{\mu}\mathcal{E}\{\mathbf{Y}^s\mathbf{Y}^{s,H}\}\left(\mathcal{E}\{\mathbf{Y}^n\mathbf{Y}^{n,H}\} + \frac{1}{\mu}\mathcal{E}\{\mathbf{Y}^s\mathbf{Y}^{s,H}\}\right)^{-H}\mathbf{Y}(f) \\ &= \mathrm{e}^{-\mathrm{i}2\pi f\Delta}Y_0(f) - \mathbf{W}^H(f)\mathbf{Y}(f), \end{aligned} \tag{59}$$

which is the output obtained in the case where there is no microphone mismatch. ■

[4]The subscripts $0 : M-1$ in $\mathbf{W}_{0:M-1}(f)$ and $\mathbf{Y}_{0:M-1}(f)$ are omitted for conciseness.

It should be noted that, without $\mathbf{w}_0$, performance degrades in the presence of speech leakage. Indeed, the filter $\mathbf{w}_{1:M-1}$ should both suppress noise and keep the speech distortion $\frac{1}{\mu}\mathcal{E}\{\left|\mathbf{w}_{1:M-1}^H\mathbf{y}_{1:M-1}^s\right|^2\}$ limited at those frequencies where speech leakage occurs (see (41)): for $\mathbf{y}_{1:M-1}^s[k]=\mathbf{0}$, $\mathbf{w}_{1:M-1}$ can spend all efforts to noise suppression, whereas for $\mathbf{y}_{1:M-1}^s[k]\neq\mathbf{0}$ only the component $\mathbf{w}_n=\mathbf{w}_{1:M-1}-\mathbf{w}_s$ with $\mathbf{w}_n^H\mathbf{w}_s=0$ and $\mathbf{w}_s\in$ Range$\{\mathcal{E}\{\mathbf{y}_{1:M-1}^s\mathbf{y}_{1:M-1}^{s,H}\}\}$[5], can freely (i.e., without causing speech distortion) suppress noise. The component $\mathbf{w}_s$ should be kept small to avoid speech distortion. Hence, at those frequencies where speech leakage occurs, the filter $\mathbf{w}_{1:M-1}$ has less degrees of freedom to suppress noise than in the absence of speech leakage. In the absence of reverberation and internal noise (such as sensor noise) and assuming a single desired speech source, the filter $\mathbf{w}_n$ can still suppress $M-2$ localized noise sources coming from another direction than the desired speech source. Hence, performance degradation will especially occur when the total number of sound sources (i.e. speech and noise sources) exceeds the number of noise references ($=M-1$) and/or in the presence of significant reverberation. An additional filter $\mathbf{w}_0$ on the speech reference then compensates for this performance degradation (see also Section V).

V. EXPERIMENTAL RESULTS

This Section illustrates the theoretical results of Section IV by means of experimental results for a hearing aid application. Section V-A and Section V-B, respectively, describe the set-up and the performance measures that are used. In Section V-C, the impact of the different parameter settings of the SP-SDW-MWF on the performance and the sensitivity to signal model errors is evaluated. Comparison is made with the QIC-GSC.

A. Set-up

A three-microphone Behind-The-Ear (BTE) hearing aid with three omnidirectional microphones (Knowles FG-3452) has been mounted on a dummy head in an office room. The interspacing d between the first and the second microphone is about $d=1$ cm and the interspacing between the second and third microphone about 1.5 cm. The reverberation time $T_{60\,\text{dB}}$ is about 700 ms for a speech weighted noise signal. The desired speech signal and the noise signals are uncorrelated. Both the speech and the noise signal have a level of 70 dB SPL at the center of the head. The desired speech source and noise sources are positioned

[5] $\mathcal{E}\{\mathbf{y}_{1:M-1}^s\mathbf{y}_{1:M-1}^{s,H}\}$ is rank deficient and hence, $\mathbf{w}_n\neq\mathbf{0}$, if the speech leakage does not cover the full frequency spectrum and/or when the number of noise references exceeds one and the direct-to-reverberant ratio of the speech is high.

at a distance of 1 meter from the head: the speech source in front of the head, the noise sources at an angle θ w.r.t. the speech source. To get an idea of the average performance based on directivity only, stationary speech and noise signals with the same, average long-term power spectral density are used. The signals can be found on [31]. The total duration of the input signal is 10 seconds of which 5 seconds contains noise only and 5 seconds contain both the speech and noise signal. For evaluation purposes, the speech and noise signal have been recorded separately.

The microphone signals are pre-whitened prior to processing to improve intelligibility [32], and the output is accordingly de-whitened. In the experiments, the microphones have been calibrated by means of recordings of an anechoic speech weighted noise signal positioned at 0° measured while the microphone array was mounted on the head. A delay-and-sum beamformer is used as a fixed beamformer, since -in case of small microphone interspacing - it is robust to model errors. The blocking matrix $\mathbf{B}$ pairwise subtracts the time aligned calibrated microphone signals.

To investigate the effect of the different parameter settings (i.e. μ, $\mathbf{w}_0$) on the performance, the filter coefficients are computed using (34) where $\mathcal{E}\{\mathbf{y}^s_{0:M-1}\mathbf{y}^{s,H}_{0:M-1}\}$ is estimated by means of the clean speech contributions of the microphone signals. In practice, $\mathcal{E}\{\mathbf{y}^s_{0:M-1}\mathbf{y}^{s,H}_{0:M-1}\}$ is approximated using (31). The effect of approximation (31) on the performance was found to be small (i.e. differences of at most 0.5 dB in intelligibility weighted Signal-to-Noise ratio improvement) for the given data set. The QIC-GSC is implemented using variable loading RLS [13]. The filter length L per channel equals 96.

B. Performance measures

To assess the performance of the different approaches, the broadband intelligibility weighted signal-to-noise ratio improvement [33] is used, defined as

$$\Delta\text{SNR}_{\text{intellig}} = \sum_i I_i(\text{SNR}_{i,\text{out}} - \text{SNR}_{i,\text{in}}), \tag{60}$$

where the band importance function I_i expresses the importance of the i-th one-third octave band with center frequency f_i^c for intelligibility, $\text{SNR}_{i,\text{out}}$ is the output SNR (in dB) and $\text{SNR}_{i,\text{in}}$ is the input SNR (in dB) in the i-th one third octave band. The center frequencies f_i^c and the values I_i are defined in [34]. The intelligibility weighted signal-to-noise ratio reflects how much intelligibility is improved by the noise reduction algorithms, but does not take into account speech distortion.

To measure the amount of speech distortion, we define the following intelligibility weighted spectral distortion measure

$$\text{SD}_{\text{intellig}} = \sum_i I_i \text{SD}_\text{i} \tag{61}$$

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

$$\text{SD}_i = \int_{2^{-1/6} f_i^c}^{2^{1/6} f_i^c} |10 \log_{10} G^s(f)| \, df \Bigg/ \left[\left(2^{1/6} - 2^{-1/6} \right) f_i^c \right], \tag{62}$$

with $G^s(f)$ 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 w.r.t. the output of the fixed beamformer.

C. Experimental results

The impact of the different parameter settings for μ and $\mathbf{w}_0$ on the performance of the SP-SDW-MWF is illustrated for a five noise source scenario. The 5 noise sources are positioned at angles 75°, 120°, 180°, 240°, 285° w.r.t. the desired source at 0°. To assess the sensitivity of the algorithm against errors in the assumed signal model, the influence of microphone mismatch, e.g., gain mismatch of the second microphone, on the performance is depicted. Among the different possible signal model errors, microphone mismatch was found to be especially harmful to the performance of the GSC in a hearing aid application [16]. In hearing aids, microphones are rarely matched in gain and phase. In [35], gain and phase differences between microphone characteristics of up to 6 dB and 10°, respectively, have been reported.

1) SP-SDW-MWF without $\mathbf{w}_0$ (SDR-GSC): Figure 5 plots the improvement $\Delta\text{SNR}_{\text{intellig}}$ and the speech distortion $\text{SD}_{\text{intellig}}$ as a function of $\frac{1}{\mu}$ obtained by the SDR-GSC (i.e., the SP-SDW-MWF without filter $\mathbf{w}_0$) for different gain mismatches Υ_2 at the second microphone. *In the absence of microphone mismatch*, the amount of speech leakage into the noise references is limited. Hence, the amount of speech distortion is low for all μ. Since there is still a small amount of speech leakage due to reverberation, the amount of noise reduction and speech distortion slightly decreases for increasing $\frac{1}{\mu}$, especially for $\frac{1}{\mu} > 1$. *In the presence of microphone mismatch*, the amount of speech leakage into the noise references grows. For $\frac{1}{\mu} = 0$ (GSC), the speech gets significantly distorted. Due to the cancellation of the desired signal, also the improvement $\Delta\text{SNR}_{\text{intellig}}$ degrades. Setting $\frac{1}{\mu} > 0$, improves the performance of the GSC in the presence of model errors without compromising performance in the absence of signal model errors. Since in this example the number of localized noise sources exceeds $M - 2$, the improvement $\Delta\text{SNR}_{\text{intellig}}$ by the SDR-GSC significantly degrades in the presence of mismatch. For a single noise source scenario, this degradation is less pronounced. Figure 6 illustrates this for a noise source at 180°.

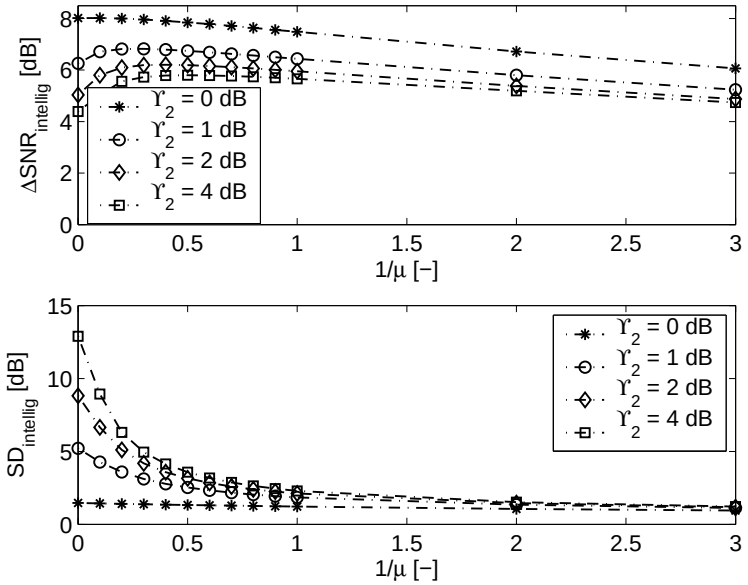


Fig. 5. Influence of $1/\mu$ on the performance of the SDR GSC for different gain mismatches Υ_2 at the second microphone.

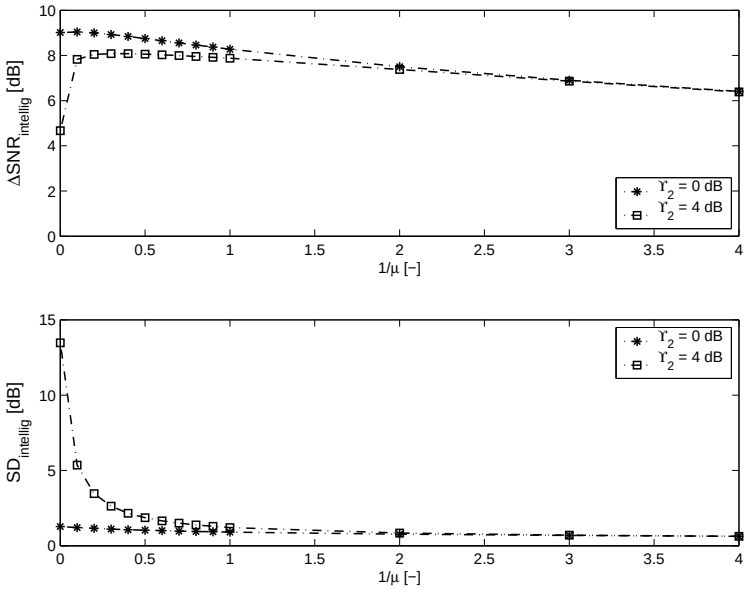


Fig. 6. Influence of $1/\mu$ on the performance of the SDR GSC for different gain mismatches Υ_2 at the second microphone, single noise source at 180°.

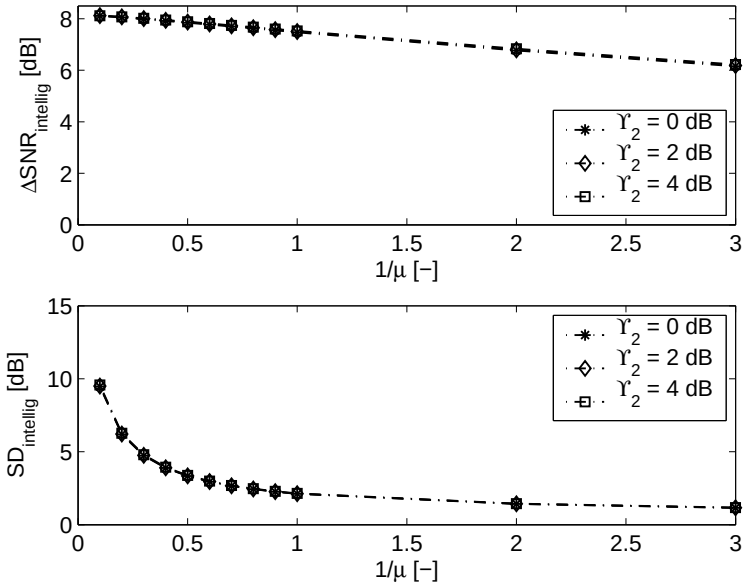


Fig. 7. Influence of $1/\mu$ on the performance of the SP SDW MWF with $\mathbf{w}_0$ for different gain mismatches Υ_2 at the second microphone.

2) SP-SDW-MWF with filter $\mathbf{w}_0$*:* Figure 7 plots the performance measures $\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$ of the *SP-SDW-MWF with filter* $\mathbf{w}_0$. In general, the amount of speech distortion and noise reduction grows for decreasing $\frac{1}{\mu}$. For $\mu = \infty$, all emphasis is put on noise reduction. As also illustrated by Figure 7, this results in a total cancellation of the speech and the noise signal and hence degraded performance. *In the absence of model errors*, the settings $L_0 = 0$ and $L_0 \neq 0$ result - except for $\frac{1}{\mu} = 0$ - in the same $\Delta\text{SNR}_{\text{intellig}}$ [6], while the distortion for the SP-SDW-MWF with $\mathbf{w}_0$ is higher due to the additional single-channel SDW-SWF. For $L_0 \neq 0$, the performance does -in contrast to $L_0 = 0$ - not degrade due to the microphone mismatch.

3) Comparison with QIC: Figure 8 depicts the improvement $\Delta\text{SNR}_{\text{intellig}}$ and the speech distortion $\text{SD}_{\text{intellig}}$, respectively, of the QIC-GSC as a function of β^2. Like the SDR-GSC, the QIC increases the robustness of the GSC. The QIC is independent of the amount of speech leakage. As a consequence, distortion grows fast with increasing gain deviation. The constraint value β should be chosen so that the maximum permissible speech distortion level is not exceeded for the largest possible model errors. This

[6]For $L_0 \neq 0$, the SNR improvement was larger thanks to the single channel SDW SWF post-processor (see Section IV). For other noise sources, e.g., a narrow band noise source, also a better improvement in $\text{SNR}_{\text{intellig}}$ can be achieved by $L_0 \neq 0$ thanks to the single channel spectral filtering.

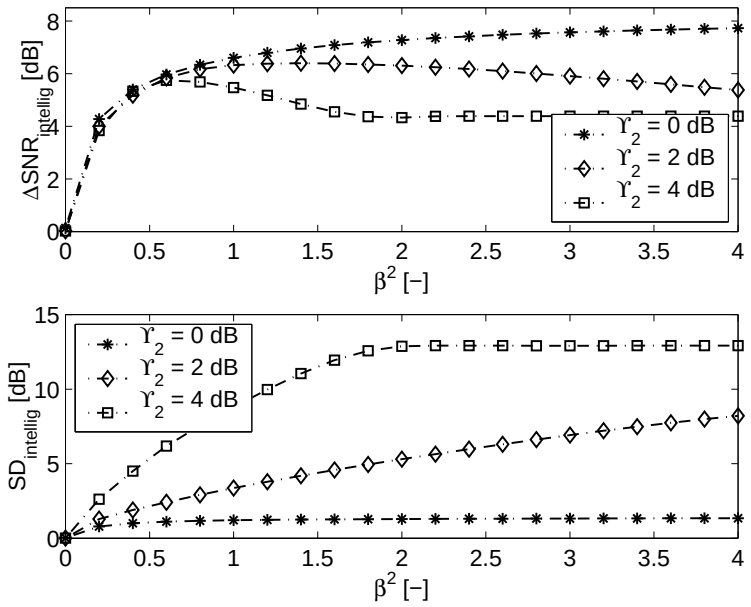


Fig. 8. $\Delta\text{SNR}_{\text{intellig}}$and $\text{SD}_{\text{intellig}}$ for QIC-GSC as a function of β^2 for different gain mismatches Υ_2 at the second microphone.

goes at the expense of reduced noise reduction for small model errors. The SDR-GSC on the other hand, keeps the speech distortion limited for all model errors (see Figure 5). Emphasis on speech distortion is increased if the amount of speech leakage grows. As a result, a better noise reduction performance is obtained for small model errors, while guaranteeing sufficient robustness for large model errors. In addition, Figure 7 demonstrates that an additional filter $\mathbf{w}_0$ significantly improves the performance of the SP-SDW-MWF in the presence of signal model errors.

VI. Conclusion

In [18], a SDW-MWF technique has been proposed for speech enhancement that provides an MMSE estimate of the speech signal portion in one of the microphone signals. In contrast to the GSC, it does not rely on any a priori assumptions about the signal model so that it is found to be less sensitive to errors in the assumed signal model.

In this paper, we showed that the GSC and the SDW-MWF can be integrated into one signal processing scheme, referred to as *Spatially pre-processed, Speech Distortion Weighted Multi-channel Wiener Filter (SP-SDW-MWF)*. This signal processing scheme consists of a fixed, spatial pre-processor and an adaptive stage that is based on an SDW-MWF. This new scheme encompasses the GSC and the SDW-MWF as extreme cases. In addition, it allows for an in-between solution that can be interpreted as a Speech

Distortion Regularized GSC (SDR_GSC). Depending on the setting of a trade-off *parameter* μ and *the presence or absence of the filter* $\mathbf{w}_0$ *on the speech reference*, the GSC, the SDR-GSC or a (SDW-)MWF is obtained.

In Section IV, the different parameter settings of the SP-SDW-MWF have been interpreted.

- Without $\mathbf{w}_0$, the SP-SDW-MWF corresponds to an SDR-GSC: the ANC design criterion is supplemented with a regularization term that limits the speech distortion due to signal model errors. The larger $\frac{1}{\mu}$, the smaller the amount of distortion. For $\frac{1}{\mu} = 0$, distortion is ignored completely, which corresponds to the GSC-solution. The SDR-GSC is then an alternative technique to the QIC-GSC to decrease the sensitivity of the GSC to signal model errors. In contrast to the QIC-GSC, the SDR-GSC shifts emphasis towards speech distortion when the amount of speech leakage grows. In the absence of signal model errors, the performance of the GSC is preserved. As a result, a better noise reduction performance is obtained for small model errors, while guaranteeing robustness against large model errors.
- Since the SP-SDW-MWF takes speech distortion explicitly into account, a filter $\mathbf{w}_0$ on the speech reference can be added. It is shown that -in the absence of speech leakage and for infinitely long filter lengths- the SP-SDW-MWF corresponds to a cascade of a SDR-GSC with a SDW-SWF post-processor. In the presence of speech leakage, the SP-SDW-MWF with $\mathbf{w}_0$ tries to preserve its performance: compared to an SDR-GSC with SDW-SWF post-processor, the SP-SDW-MWF then contains extra filtering operations that compensate for the performance degradation of the SDR-GSC with SDW-SWF due to speech leakage. In contrast to the SDR-GSC (and thus also the GSC), performance does not degrade due to microphone mismatch.

In Section V, experimental results for a hearing aid application illustrated the theoretical results of Section IV. The SP-SDW-MWF indeed increases the robustness of the GSC against signal model errors. Comparison with the QIC-GSC demonstrated that the SP-SDW-MWF achieves a better noise reduction performance for a given maximum allowable speech distortion level.

APPENDIX

In the absence of speech leakage, i.e. $y_l^s[k] = 0$, $l = 1, \cdots, M-1$ and for sufficiently long filter lengths, the SP-SDW-MWF with $\mathbf{w}_0[k] \neq \mathbf{0}$ corresponds to the cascade of a GSC and a single-channel SDW-SWF post-processor.

Proof: If $\mathbf{y}^s_{1:M-1}[k]=\mathbf{0}$, the cost function of the SP-SDW-MWF (45) can be written as

$$J(\mathbf{w}_{0:M-1})=\mathcal{E}\left\{\underbrace{\left|\begin{bmatrix}(\mathbf{e}_{\Delta+1}-\mathbf{w}_0)^H & -\mathbf{w}^H_{1:M-1}\end{bmatrix}\begin{bmatrix}\mathbf{y}^n_0\\ \mathbf{y}^n_{1:M-1}\end{bmatrix}\right|^2}_{\varepsilon_n^2}\right\}+\frac{1}{\mu}\mathcal{E}\left\{\underbrace{\left|\mathbf{w}_0^H\mathbf{y}_0^s\right|^2}_{\varepsilon_d^2}\right\}. \tag{63}$$

Without loss of generality, we assume that $\Delta=\lceil\frac{L}{2}\rceil-1$ where L is the filter length per channel. Assume that the filter length L is sufficiently long such that $\mathbf{w}_{0:M-1}[k]$ can be decomposed as depicted in **Figure 9**, i.e.,

$$\mathbf{w}_0[k] \;=\; \mathbf{e}_{\Delta_1+1}\otimes\mathbf{w}_b[k], \tag{64}$$

$$\mathbf{w}_l[k] \;=\; (\mathbf{e}_{\Delta_2+1}-\mathbf{w}_b[k])\otimes\mathbf{w}_{dl}[k],\quad l=1,\,...,\,M-1 \tag{65}$$

where $\Delta_1+\Delta_2=\Delta$, $\mathbf{w}_b[k]\in\mathbb{C}^{L_b\times 1}$ is a single-channel filter with length L_b and

$$\mathbf{w}_d[k]=\begin{bmatrix}\mathbf{w}^T_{d1}[k] & \mathbf{w}^T_{d2}[k] & \cdots & \mathbf{w}^T_{dM-1}[k]\end{bmatrix}^T\in\mathbb{C}^{L_d(M-1)\times 1} \tag{66}$$

is a multi-channel filter with filter length L_d per channel. Define the intermediate output $v[k]$ as the output of the multi-channel filter $\begin{bmatrix}\mathbf{e}^T_{\Delta_1+1} & -\mathbf{w}^T_d[k]\end{bmatrix}^T$, i.e.,

$$v[k] \;=\; y_0[k-\Delta_1]-\sum_{l=1}^{M-1}\mathbf{w}^H_{dl}[k]\mathbf{y}_l[k]. \tag{67}$$

Then, using $\mathbf{y}^s_{1:M-1}[k]=0$, the cost function J can be re-written as

$$J(\mathbf{w}_b,\mathbf{w}_d)=\mathcal{E}\left\{\left|(\mathbf{e}_{\Delta_2+1}-\mathbf{w}_b)^H\mathbf{v}^n[k]\right|^2\right\}+\frac{1}{\mu}\mathcal{E}\left\{\left|\mathbf{w}_b^H\mathbf{v}^s[k]\right|^2\right\} \tag{68}$$

where

$$\mathbf{v}^{n,H}[k] \;=\; \begin{bmatrix}v^n[k] & v^n[k-1] & \cdots & v^n[k-L_b+1]\end{bmatrix}^H, \tag{69}$$

$$\mathbf{v}^{s,H}[k] \;=\; \begin{bmatrix}v^s[k] & v^s[k-1] & \cdots & v^s[k-L_b+1\ ]\end{bmatrix}^H. \tag{70}$$

From $\frac{\partial}{\partial\mathbf{w}_b}J(\mathbf{w}_b,\mathbf{w}_d)=0$, we find

$$\mathbf{w}_b[k]=\left(\mathcal{E}\{\mathbf{v}^n\mathbf{v}^{n,H}\}+\frac{1}{\mu}\mathcal{E}\{\mathbf{v}^s\mathbf{v}^{s,H}\}\right)^{-1}\mathcal{E}\{\mathbf{v}^n[k]v^{n,*}[k-\Delta_2]\}. \tag{71}$$

Hence, the filter $(\mathbf{e}_{\Delta_2+1}-\mathbf{w}_b[k])$ corresponds to a single channel SDW-SWF that estimates the (delayed) speech component $v^s[k-\Delta_2]$ in the output $v[k]$ of the multi-channel filter $\begin{bmatrix}\mathbf{e}^T_{\Delta_1+1} & -\mathbf{w}^T_d[k]\end{bmatrix}^T$.

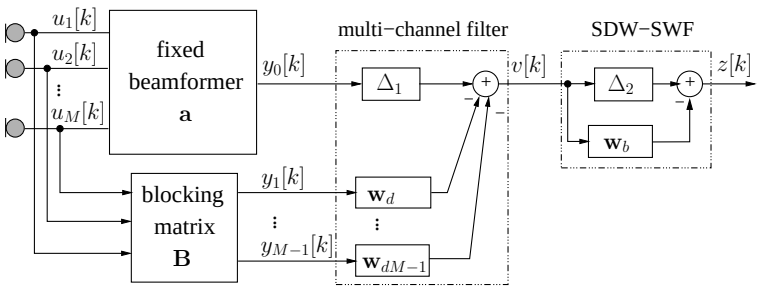


Fig. 9. Decomposition of the SP-SDW-MWF $\mathbf{w}_{0:M-1}[k]$ in a multi-channel filter $\mathbf{w}_d[k]$ and a single-channel post-filter $\mathbf{e}_{\Delta_2+1} - \mathbf{w}_a[k]$.

The solution for $\mathbf{w}_d$ can be found by setting the derivative $\frac{\partial}{\partial \mathbf{w}_d}\varepsilon_n^2$ (cf (63)) to zero[7]. Re-writting ε_n^2 as

$$\varepsilon_n^2 \quad = \quad \mathcal{E}\left\{ \left| \left[\begin{array}{cc} \mathbf{e}_{\Delta_1+1}^H & -\mathbf{w}_d^H \; [k] \end{array} \right] \mathbf{y}_{0:M-1,p}^n[k] \right|^2 \right\}, \tag{72}$$

$$\mathbf{y}_{0:M-1,p}^n \; [k] = \left[\begin{array}{cccc} \mathbf{y}_{0,p}^{n,H}[k] & \mathbf{y}_{1,p}^{n,H}[k] & \cdots & \mathbf{y}_{M-1,p}^{n,H}[k] \end{array} \right]^H, \tag{73}$$

with

$$\mathbf{y}_{l,p}^n = \begin{bmatrix} \boxed{(\mathbf{e}_{\Delta_2+1} - \mathbf{w}_b)^H} & 0 & \cdots & 0 \\ 0 & \boxed{(\mathbf{e}_{\Delta_2+1} - \mathbf{w}_b)^H} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \boxed{(\mathbf{e}_{\Delta_2+1} - \mathbf{w}_b)^H} \end{bmatrix} \begin{bmatrix} y_l[k] \\ y_l[k-1] \\ \vdots \\ y_l[k-L+1] \end{bmatrix}, \tag{74}$$

we find

$$\mathbf{w}_d[k] = \mathcal{E}\{\mathbf{y}_{0:M-1,p}^n[k]\mathbf{y}_{0:M-1,p}^{n,H}[k]\}^{-1}\mathcal{E}\{\mathbf{y}_{0:M-1,p}^n[k]y_{0,p}^n[k-\Delta_1]\}.$$

The filter coefficients $\mathbf{w}_d$ correspond to the ANC coefficients of a GSC where the speech and noise references have been pre-filtered with the single -channel SDW-SWF filter $\mathbf{e}_{\Delta_2+1} - \mathbf{w}_b$. For infinite filter lengths, $\mathbf{w}_d$ then equals the ANC operating on the original speech and noise references.

In short, in the absence of speech leakage, the SP-SDW-MWF equals to a GSC with a single channel SDW-SWF post-processor $\mathbf{e}_{\Delta_2+1} - \mathbf{w}_b$. ■

ACKNOWLEDGMENTS

Ann Spriet is a Research Assistant with the F.W.O. Vlaanderen. This research work was carried out at the ESAT laboratory and Lab. Exp. ORL of the Katholieke Universiteit Leuven, in the frame of the

[7]Note that for $\mathbf{y}^s = \mathbf{0}$, ε_d^2 in (63) does not depend on $\mathbf{w}_d[k]$.

Belgian State, Prime Minister's Office - Federal Office for Scientific, Technical and Cultural Affairs - Interuniversity Poles of Attraction Programme (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'), IWT project 020540: 'Innovative Speech Processing Algorithms for Improved Performance of Cochlear Implants' and was partially sponsored by Cochlear. The scientific responsibility is assumed by its authors.

REFERENCES

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

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

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

[4] K. M. Buckley, "Broad- band beamforming and the generalized sidelobe canceller," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 34, no. 5, pp. 1322–1323, Oct. 1986.

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

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

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

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

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

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

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

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

[13] 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.

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

[15] A. Spriet, M. Moonen, and 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, Oct. 2001, pp. 31–34.

[16] A. Spriet, M. Moonen, and 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), Sept. 2002, available at ftp://ftp.esat.kuleuven.ac.be/sista/spriet/reports/02-81.pdf.

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

[18] 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.

[19] 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.

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

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

[22] S. Doclo and M. Moonen, "Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics," Accepted for publication in IEEE Transactions on Signal Processing, Jan. 2003. Available at ftp://ftp.esat.kuleuven.ac.be/sista/doclo/reports/02-111.ps.gz.

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

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

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

[26] N. Grbíc and S. Nordholm, "Soft contrained subband beamforming for hands-free speech enhancement," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, May 2002, pp. 885–888.

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

[28] A. Spriet, M. Moonen, and J. Wouters, "Stochastic gradient based implementation of spatially pre-processed (speech distortion weighted) multi-channel Wiener filtering for noise reduction in hearing aids," Tech. Rep. ESAT-SISTA/TR 03-47, ESAT/SISTA, K.U. Leuven (Belgium), 2003, available at ftp://ftp.esat.kuleuven.ac.be/sista/spriet/reports/03-47.pdf.

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

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

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

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

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

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

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