# STOCHASTIC GRADIENT IMPLEMENTATION OF SPATIALLY PRE-PROCESSED MULTI-CHANNEL WIENER FILTERING FOR NOISE REDUCTION IN HEARING AIDS

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

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

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

### ABSTRACT

Recently, a generalized noise reduction scheme has been proposed, called the Spatially Pre-processed Speech Distortion Weighted Multi-channel Wiener Filter (SP-SDW-MWF). Compared to GSC with Quadratic Inequality Constraint (QIC-GSC), the SP-SDW-MWF reduces more noise, for a given maximum speech distortion level. In this paper, we develop time-domain and frequency-domain stochastic gradient implementations of the SP-SDW-MWF. Experimental results with a hearing aid show that the proposed stochastic gradient algorithm preserves the benefit of the SP-SDW-MWF over the QIC-GSC, while its computational cost is comparable to the NLMS based Scaled Projection Algorithm (SPA) for QIC-GSC.

## 1. INTRODUCTION

Noise reduction algorithms are crucial for hearing impaired people to improve speech intelligibility in background noise. Multi-microphone systems exploit spatial in addition to temporal and spectral information of the desired and noise signal and are thus preferred to single microphone procedures. For small-sized arrays such as hearing aids, multi-microphone noise reduction goes together with an increased sensitivity to errors in the assumed signal model such as microphone mismatch, reverberation, etc. [1]

In [2], a generalized noise reduction scheme has been proposed, called the Spatially Pre-processed, Speech Distortion Weighted, Multi-channel Wiener Filter (SP-SDW-MWF). It encompasses the GSC and an MWF technique [3, 4] as extreme cases and allows for inbetween solutions such as the Speech Distortion Regularized GSC (SDR-GSC). The SDR-GSC or more general the SP-SDW-MWF adds robustness against 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 QIC-GSC, the SP-SDW-MWF achieves a better noise reduction performance, for a given maximum speech distortion level.

The recursive matrix-based implementations of the SDW-MWF [3, 4, 5] can be applied to implement the SP-SDW-MWF [2]. However, in contrast to the GSC and the QIC-GSC [6], no cheap stochastic gradient implementation is available yet. In this paper, we derive time-domain and frequency-domain stochastic gradient algorithms for the SP-SDW-MWF and compare their performance to the NLMS based SPA [6]. Experimental results demonstrate that the proposed stochastic gradient based SP-SDW-MWF outperforms the SPA, while its computational cost is comparable.

*This research was carried out at ESAT and Lab. Exp. ORL of K.U. Leuven, in the frame of IUAP P5/22, the Concerted Research Action GOA-MEFISTO-666, FWO Project nr. G.0233.01, *Signal Processing and automative patient fitting of auditory prostheses*, IWT project 020540, *Innovative speech processing algorithms for improved performance in cochlear implants* and was sponsored by Cochlear. The scientific responsibility is assumed by its authors.

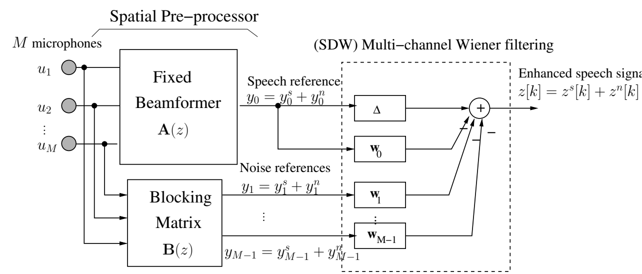


**Fig. 1**. Spatially Pre-processed SDW-MWF.

## 2. SPATIALLY PRE-PROCESSED SDW-MWF

The SP-SDW-MWF [2], described in Figure 1, 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 SDW-MWF [2, 3, 4]. In the sequel, an endfire array is assumed and the desired speaker is assumed to be in front at $0^\circ$. Given $M$ microphone signals[1]

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

the *fixed beamformer* $\mathbf{A}(z)$ creates a so-called speech reference $y_0[k] = y_0^s[k] + y_0^n[k]$, by steering a beam towards the front and the *blocking matrix* $\mathbf{B}(z)$ creates $M-1$ so-called noise references $y_i[k] = y_i^s[k] + y_i^n[k]$, $i = 1, \ldots, M-1$ by steering zeroes towards the front. During periods of speech, the references $y_i[k]$ consist of speech + noise, i.e., $y_i[k] = y_i^s[k] + y_i^n[k]$, $i = 0, \ldots, M-1$. During periods of noise, only the noise component $y_i^n[k]$ is observed. We assume that the second order statistics of the noise are sufficiently stationary so that they can be estimated during periods of noise only.

*The SDW-MWF filter* $\mathbf{w}_k \in \mathbb{R}^{ML\times 1}$[2] provides an estimate $\mathbf{w}_k^T \mathbf{y}_k$ of the noise contribution $y_0^n[k-\Delta]$ in the speech reference by minimizing the cost function $J(\mathbf{w}_k)$

$$\boxed{J(\mathbf{w}_k) = \frac{1}{\mu} \underbrace{\mathcal{E}\{\left|\mathbf{w}_k^T \mathbf{y}_k^s\right|^2\}}_{\varepsilon_d^2} + \underbrace{\mathcal{E}\{\left|y_0^n[k-\Delta] - \mathbf{w}_k^T \mathbf{y}_k^n\right|^2\}}_{\varepsilon_n^2}.} \tag{2}$$

[1]In the sequel, the superscripts $s$ and $n$ are used to refer to the speech and noise contribution of a signal.

with

$$\mathbf{w}_k^T = \begin{bmatrix} \mathbf{w}_0^T[k] & \mathbf{w}_1^T[k] & \dots & \mathbf{w}_{M-1}^T[k] \end{bmatrix}, \quad (3)$$

$$\mathbf{w}_i[k] = \begin{bmatrix} w_i[0] & w_i[1] & \dots & w_i[L-1] \end{bmatrix}^T, \quad (4)$$

$$\mathbf{y}_k^T = \begin{bmatrix} \mathbf{y}_0^T[k] & \mathbf{y}_1^T[k] & \dots & \mathbf{y}_{M-1}^T[k] \end{bmatrix}, \quad (5)$$

$$\mathbf{y}_i[k] = \begin{bmatrix} y_i[k] & y_i[k-1] & \dots & y_i[k-L+1] \end{bmatrix}^T, (6)$$

This estimate is then subtracted from the speech reference, as indicated in Figure 1, to obtain a better speech signal $z[k]$. The term $\varepsilon_d^2$ represents the speech distortion energy and $\varepsilon_n^2$ the residual noise energy. The parameter $\mu \in [0, \infty)$ trades off between noise reduction and speech distortion. Depending on the setting of $\frac{1}{\mu}$ and the presence of the filter $\mathbf{w}_0$ on the speech reference, the GSC, the (SDW-)MWF or the SDR-GSC is obtained [2].

- Without $\mathbf{w}_0$, the SP-SDW-MWF corresponds to an SDR-GSC: the ANC design criterion is supplemented with a regularization term $\frac{1}{\mu}\varepsilon_d^2$ that limits speech distortion due to signal model errors. For $\mu = \infty$, the GSC solution is obtained. Compared to the QIC-GSC, the SDR-GSC obtains better noise reduction for small signal model errors, while guaranteeing robustness against large model errors.
- Since the SP-SDW-MWF takes speech distortion explicitly into account in the design criterion, a filter $\mathbf{w}_0$ on the speech reference can be added. For $\mu = 1$, we obtain an MWF. Compared to the SDR-GSC, performance is less affected by model errors.

## 3. STOCHASTIC GRADIENT ALGORITHM (SG)

### 3.1. Time-Domain (TD) implementation

A stochastic gradient algorithm approximates the steepest descent algorithm

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \rho \left( -\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} \right)_{\mathbf{w}=\mathbf{w}_n}, \quad (7)$$

using an instantaneous gradient estimate. Replacing the iteration index $n$ by a time index $k$ and leaving out the expectation values, we obtain the following update equation for the cost function (2):

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \rho \left\{ \mathbf{y}_k^n (y_0^n[k-\Delta] - \mathbf{y}_k^{n,T}\mathbf{w}_k) - \mathbf{r}_k \right\}, (8)$$

$$\mathbf{r}_k = \frac{1}{\mu}\mathbf{y}_k^s \mathbf{y}_k^{s,T} \mathbf{w}_k, \quad (9)$$

with $\mathbf{w}_k, \mathbf{y}_k \in \mathbb{R}^{NL \times 1}$, where $N$ denotes the number of input channels to the adaptive filter ($N = M$ if $\mathbf{w}_0$ is present, $N = M-1$ if $\mathbf{w}_0$ is absent). For $\frac{1}{\mu} = 0$ and no filter $\mathbf{w}_0$, (8) reduces to an LMS type update formula often used in GSC, which is then operated during periods of *noise only*. The additional term $\mathbf{r}_k$ in (8) limits speech distortion due to signal model errors.

Equation (8) requires knowledge of the correlation matrix $\mathbf{y}_k^s \mathbf{y}_k^{s,T}$ or $\mathcal{E}\{\mathbf{y}_k^s \mathbf{y}_k^{s,T}\}$ of the clean speech. In practice, this information is not available. To avoid the need for calibration, $L \times 1$-dimensional *speech + noise* signal vectors $\mathbf{y}_i[k]$, $i = M-N, \dots, M-1$ are stored in a circular *speech + noise* buffer $\mathbf{B}_1 \in \mathbb{R}^{L_{buf_1} \times N}$ during processing as in [7]. *During periods of noise only* (i.e., when $y_i[k] = y_i^n[k]$, $i = 0, \dots, M-1$), the filter $\mathbf{w}_k$ is updated using the following approximation for (9):

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \rho \left\{ \mathbf{y}_k^n (y_0^n[k-\Delta] - \mathbf{y}_k^{n,T}\mathbf{w}_k) - \mathbf{r}_k \right\}, \quad (10)$$

$$\mathbf{r}_k = \tilde{\lambda}\mathbf{r}_{k-1} + (1-\tilde{\lambda})\frac{1}{\mu}\left( \mathbf{y}_k^{buf_1} \mathbf{y}_k^{buf_1,T} - \mathbf{y}_k^n \mathbf{y}_k^{n,T} \right) \mathbf{w}_k, (11)$$

where $\mathbf{y}_k^{buf_1}$ is a speech + noise vector constructed from data in the buffer $\mathbf{B}_1$. In the sequel, a normalized step size $\rho$ is used:

$$\rho = \frac{\rho'}{\zeta_k + \mathbf{y}_k^{n,T}\mathbf{y}_k^n + \delta} \quad (12)$$

$$\zeta_k = \tilde{\lambda}\zeta_{k-1} + (1-\tilde{\lambda})\frac{1}{\mu}\left| \mathbf{y}_k^{buf_1,T}\mathbf{y}_k^{buf_1}[k] - \mathbf{y}_k^{n,T}\mathbf{y}_k^n \right|. (13)$$

Additional storage of noise only vectors $\mathbf{y}_i^n$, $i = 0, \cdots, M-1$ in a second buffer $\mathbf{B}_2 \in \mathbb{R}^{L_{buf_2} \times M}$ allows to adapt $\mathbf{w}_k$ also *during periods of speech + noise,* using

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \rho \left\{ \mathbf{y}_k^{buf_2} (y_0^{buf_2}[k-\Delta] - \mathbf{y}_k^{buf_2,T}\mathbf{w}_k) - \mathbf{r}_k \right\}, (14)$$

$$\mathbf{r}_k = \tilde{\lambda}\mathbf{r}_{k-1} + (1-\tilde{\lambda})\frac{1}{\mu}\left( \mathbf{y}_k \mathbf{y}_k^T - \mathbf{y}_k^{buf_2}\mathbf{y}_k^{buf_2,T} \right)\mathbf{w}_k, \quad (15)$$

with $\mathbf{y}_k^{buf_2}$ a noise vector constructed from data in the buffer $\mathbf{B}_2$.

*Remark: For $\tilde{\lambda} = 0$ and $\mu > 1$, an alternative stochastic gradient algorithm similar to [7] can be derived from (10)-(15) by invoking some independence assumptions. However, its performance was found to be worse than algorithm (10)-(15) [8].*

For $\tilde{\lambda} = 0$, the estimate (11), (15) of $\mathbf{r}_k$ is quite bad due to large differences between the rank-one matrices $\mathbf{y}_i^n \mathbf{y}_i^{n,T}$ and $\mathbf{y}_j^n \mathbf{y}_j^{n,T}$ at different time instants $i$ and $j$. This results in a large excess error, especially for small $\mu$ and large step sizes $\rho'$ [8]. Using an estimate of the average correlation matrix $\mathcal{E}\{\mathbf{y}_k^s \mathbf{y}_k^{s,T}\}$ in (9), i.e.,

$$\mathbf{r}_k = \frac{1}{\mu}\frac{1}{K}\left( \sum_{l=k-K+1}^{k} \mathbf{y}_l^{buf_1}\mathbf{y}_l^{buf_1,T} - \sum_{l=k-K+1}^{k} \mathbf{y}_l^n \mathbf{y}_l^{n,T} \right)\mathbf{w}_k, \quad (16)$$

would significantly improve the performance, but requires expensive matrix operations. Therefore, assuming that $\mathbf{w}_k$ varies slowly in time, (11), (15) is - especially for small $\tilde{\lambda}$ - a good approximation of (16) without matrix operations. For *stationary noise*, a small $K$ or $\tilde{\lambda}$ (i.e., $K = 1/(1-\tilde{\lambda}) \sim ML$) suffices [8]. In practice, the speech and the noise signals are often *spectrally highly non-stationary* (e.g., multi-talker babble noise) while their *long-term* spectral and spatial characteristics such as the positions of the sources usually vary more slowly in time. Spectrally highly non-stationary noise can then still be spatially suppressed by using an estimate of the *long-term* speech correlation matrix in $\mathbf{r}_k$ (see (9)), i.e., by setting $K = 1/(1-\tilde{\lambda}) \gg ML$.

### 3.2. Frequency-Domain (FD) implementation

To speed-up convergence and reduce complexity, the stochastic gradient algorithm (10)-(14) is implemented in the frequency-domain, using overlap-save. Algorithm 1 summarizes the FD implementation. Note that the FD-SG algorithm implicitly averages the gradient estimate and hence, (16) over $K = L$ samples. To obtain the same time constant in the averaging operation of $\mathbf{R}_i[k]$ as in the TD-SG algorithm, $\lambda$ should equal $\tilde{\lambda}^L$.

## 4. COMPUTATIONAL COST

Table 1 summarizes the computational cost (expressed in number of real operations[2] per second (Ops/s)) of the TD-SG and FD-SG implementation of the SP-SDW-MWF. The sampling frequency $f_s$ equals 16 kHz. We assume that one complex multiplication is equivalent to 4 real multiplications and 2 real additions. A $2L$-point FFT of a real input vector requires $2L\log_2 2L$ real MACs

[2] Counted as the number of real multiply-accumulates, divisions, etc.

**Algorithm 1** Frequency-domain implementation

**Initialization and matrix definitions:**

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

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

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

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

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

**For each new block of $ML$ input samples:**

*If noise detected:*

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

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

Store input data $\mathbf{Y}_i^n[k]$, $\mathbf{d}[k]$ in noise buffer $\mathbf{B}_2$

Create $\mathbf{Y}_i[k]$ from data in speech+noise buffer $\mathbf{B}_1$

*If speech detected:*

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

Store input data $\mathbf{Y}_i[k]$ in speech + noise buffer $\mathbf{B}_1$

Create $\mathbf{Y}_i^n[k]$, $\mathbf{d}[k]$ using data from noise buffer $\mathbf{B}_2$

*Update formula:*

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

$$\mathbf{R}_i[k] = \lambda\mathbf{R}_i[k-1] + (1-\lambda)\frac{1}{\mu}\left(\mathbf{Y}_i^H[k]\mathbf{E}_2[k] - \mathbf{Y}_i^{n,H}[k]\mathbf{E}_1[k]\right)$$

with

$$\mathbf{E}[k] = \mathbf{F}\mathbf{k}^T\left(\mathbf{d}[k] - \mathbf{k}\mathbf{F}^{-1}\sum_{j=M-N}^{M-1}\mathbf{Y}_j^n[k]\mathbf{W}_j[k]\right)$$

$$\mathbf{E}_1[k] = \mathbf{F}\mathbf{k}^T\mathbf{k}\mathbf{F}^{-1}\sum_{j=M-N}^{M-1}\mathbf{Y}_j^n[k]\mathbf{W}_j[k] = \mathbf{F}\mathbf{k}^T\mathbf{e}_1[k]$$

$$\mathbf{E}_2[k] = \mathbf{F}\mathbf{k}^T\mathbf{k}\mathbf{F}^{-1}\sum_{j=M-N}^{M-1}\mathbf{Y}_j[k]\mathbf{W}_j[k] = \mathbf{F}\mathbf{k}^T\mathbf{e}_2[k]$$

*Step size* $\mathbf{\Lambda}[k]$:

$$\mathbf{\Lambda}[k] = \frac{2\rho'}{L}\text{diag}\left\{P_0^{-1}[k], ..., P_{2L-1}^{-1}[k]\right\}$$

$$P_m[k] = \gamma P_m[k-1] + (1-\gamma)\left(P_{1,m}[k] + P_{2,m}[k]\right)$$

$$P_{1,m}[k] = \sum\left|Y_{j,m}^n\right|^2$$

$$P_{2,m}[k] = \lambda P_{2,m}[k-1] + (1-\lambda)\frac{1}{\mu}\left|\sum\left(\left|Y_{j,m}\right|^2 - \left|Y_{j,m}^n\right|^2\right)\right|$$

*Output* $\mathbf{z}[k]$*:*

$\mathbf{y}_0[k] = \left[\begin{array}{ccc} y_0[kL-\Delta] & \cdots & y_0[kL-\Delta+L-1] \end{array}\right]^T$

- If noise detected: $\mathbf{z}[k] = \mathbf{y}_0[k] - \mathbf{e}_1[k]$
- If speech detected: $\mathbf{z}[k] = \mathbf{y}_0[k] - \mathbf{e}_2[k]$

(assuming the radix-2 FFT algorithm). Comparison[3] is made with standard NLMS based ANC and the NLMS based SPA [6]. The NLMS based SPA is translated to the frequency domain by the following equations:

[3]The complexity of the NLMS ANC and NLMS based SPA represents the complexity when the adaptive filter is only updated during noise only periods. If the adaptive filter is also updated during speech + noise periods additional operations are required to compute the output [8].

| Algorithm | Complexity (ops/s) (e.g., $M = 3$, $L = 32$, $f_s = 16$ kHz) | Mops/s |
|---|---|---|
| TD-ANC | $(3(M-1)L+2)f_s$ | 3.1 |
| TD-SPA | $(5(M-1)L+4)f_s$ | 5.2 |
| TD-SG | $(9NL+10)f_s$ | 9.4[(a)], 14.0[(b)] |
| FD-ANC | $[(6M-2)\log_2 2L + (12M-4)]f_s$ | 2.0 |
| FD-SPA | $[(6M-2)f_s\log_2 2L + (16M-8)]f_s$ | 2.2 |
| FD-SG | $[(6N+10)\log_2 2L + (30N+12)]f_s$ | 3.3[(a)], 4.3[(b)] |

**Table 1**. Complexity of the TD-SG and FD-SG SP-SDW-MWF ((a) $N = M-1$, (b) $N = M$) compared to ANC and SPA.

$$\|\mathbf{w}[k]\|_2^2 = \mathbf{w}^T[k]\mathbf{w}[k] = \frac{1}{2L}\sum_{i=1}^{M-1}\mathbf{W}_i^H[k]\mathbf{W}_i[k], \quad (17)$$

$$\text{If } \|\mathbf{w}[k]\|_2^2 \geq \beta^2 : \mathbf{W}_i[k] \leftarrow \beta\frac{\mathbf{W}_i[k]}{\|\mathbf{w}[k]\|_2}. \quad (18)$$

Table 1 indicates that the TD-SG SDR-GSC (i.e., without filter $\mathbf{w}_0$ and hence, $N = M-1$) is about twice as complex as the NLMS-based SPA and about three times as complex as the standard ANC. The SP-SDW-MWF with extra filter $\mathbf{w}_0$ is a bit more complex. The increase in complexity of the frequency-domain implementations is smaller. For $M = 3$ and $L = 32$, the FD-SG SDR-GSC and SP-SDW-MWF only require 3.3 Mops/s and 4.3 Mops/s, respectively.

## 5. EXPERIMENTAL RESULTS

This section compares the performance of the FD-SG SP-SDW-MWF and the FD-NLMS SPA for different parameter settings (i.e., $1/\mu$ and $\beta^2$), based on experimental results with a Behind-The-Ear (BTE). For a fair comparison, the FD-NLMS SPA is - like the FD-SG SP-SDW-MWF -also adapted during speech + noise using data from a noise buffer.

### 5.1. Set-up and performance measures

A three-microphone BTE has been mounted on a dummy head in an office room. The desired source is positioned in front of the head (i.e., at 0°) and consists of sentences spoken by a male speaker. The noise scenario consists of three multi-talker babble noise sources, positioned at 75°, 180°and 240°. The desired signal and the total noise signal both have a level of 70 dB SPL at the center of the head. For evaluation purposes, the speech and noise signal have been recorded separately. In the experiments, the microphones have been calibrated in an anechoic room while the BTE was mounted on the head. A delay-and-sum beamformer is used as a fixed beamformer. The blocking matrix $\mathbf{B}$ pairwise subtracts the time aligned calibrated microphone signals. The filter length $L = 32$, the step size $\rho' = 0.8$ (with $\gamma = 0.95$) and $\lambda = 0.999$.

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

$$\Delta\text{SNR}_{\text{intellig}} = \sum_i I_i(\text{SNR}_{i,\text{out}} - \text{SNR}_{i,\text{in}}), \quad (19)$$

where $I_i$ expresses the importance of the $i$-th one-third octave band with center frequency $f_i^c$ for intelligibility [9], and where $\text{SNR}_{i,\text{out}}$ and $\text{SNR}_{i,in}$ is the output and input SNR (in dB) in that band, respectively. Similarly, we define an intelligibility weighted spectral distortion measure, called $\text{SD}_{\text{intellig}}$, of the desired signal as

$$\text{SD}_{\text{intellig}} = \sum_i I_i\text{SD}_i \quad (20)$$

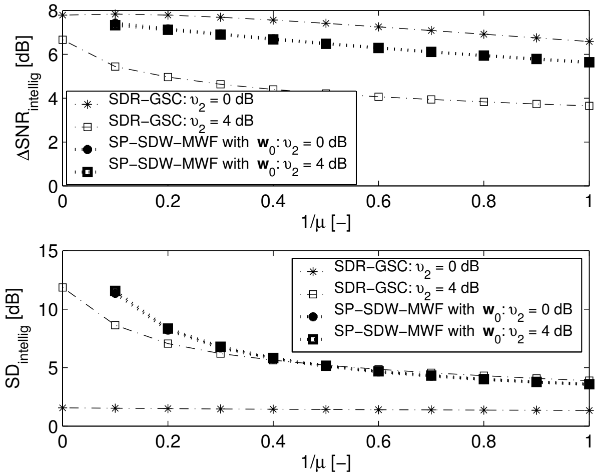


**Fig. 2**. Performance of FD-SG SP-SDW-MWF in a multiple noise source scenario.

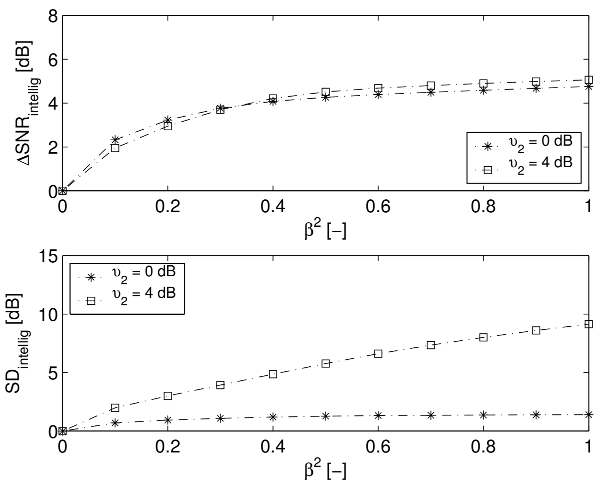


**Fig. 3**. Performance of FD-NLMS SPA in a multiple noise source scenario.

with $\mathrm{SD}_i$ the average spectral distortion (dB) in $i$-th one-third band, calculated as

$$\mathrm{SD}_i = \frac{1}{(2^{1/6} - 2^{-1/6})\, f_i^c} \int_{2^{-1/6} f_i^c}^{2^{1/6} f_i^c} |10 \log_{10} G^s(f)|\, df, \quad (21)$$

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.

### 5.2. Experimental results

Figure 2 depicts $\Delta\mathrm{SNR}_{\mathrm{intellig}}$ and $\mathrm{SD}_{\mathrm{intellig}}$ of the FD-SG SDR-GSC and SP-SDW-MWF with $\mathbf{w}_0$ as a function of the trade-off parameter $\frac{1}{\mu}$. The effect of a gain mismatch $\upsilon_2$ of $4$ dB at the second microphone is depicted too. Figure 3 shows the results of the FD-NLMS based SPA of (17)-(18) for different constraint values $\beta^2$.

In this scenario, the GSC still offers good noise suppression for a mismatch of 4 dB, at the expense of a large distortion. Both, the SPA and the stochastic gradient based SP-SDW-MWF increase the robustness of the GSC (i.e., the SDR-GSC with $\frac{1}{\mu} = 0$): distortion decreases with increasing $\frac{1}{\mu}$ and decreasing $\beta^2$. The SPA is more conservative than the SDR-GSC: the constraint value $\beta^2$ should be chosen so that the maximum permissible speech distortion is not exceeded for the largest model error, e.g., $5$ dB $\mathrm{SD}_{\mathrm{intellig}}$ for a gain mismatch up to $4$ dB. This goes at the expense of less noise reduction in case of smaller model errors (e.g., $\Delta\mathrm{SNR}_{\mathrm{intellig}} = 4\,\mathrm{dB}$ for $\beta^2 = 0.4$). The SDR-GSC on the other hand only puts emphasis on speech distortion if required, i.e., when the amount of speech leakage is large, so that a better noise reduction is obtained for small model errors (e.g., $\Delta\mathrm{SNR}_{\mathrm{intellig}}$ between $4$ dB and $7.4$ dB for $\frac{1}{\mu} = 0.5$). The SP-SDW-MWF offers more noise suppression at even larger model errors: the SP-SDW-MWF with $\mathbf{w}_0$ is -in contrast to the SDR-GSC and the SPA- hardly affected by microphone mismatch. In the absence of model errors, the SP-SDW-MWF with $\mathbf{w}_0$ achieves a slightly worse performance than the SDR-GSC. With $\mathbf{w}_0$, the estimate (11)-(15) of $\frac{1}{\mu}\mathcal{E}\{\mathbf{y}^s\mathbf{y}^{s,T}\}\mathbf{w}_k$ is less accurate due to the larger dimensions of $\frac{1}{\mu}\mathcal{E}\{\mathbf{y}^s\mathbf{y}^{s,T}\}$ and the large contribution of the speech reference in $\frac{1}{\mu}\mathcal{E}\{\mathbf{y}^s\mathbf{y}^{s,T}\}$.

In short, the proposed stochastic gradient based SP-SDW-MWF preserves the benefit of the exact SP-SDW-MWF over the QIC-GSC, while its complexity is comparable to NLMS-SPA.

## 6. REFERENCES

[1] R. W. Stadler and W. M. Rabinowitz, “On the potential of fixed arrays for hearing aids,” *J. Acoust. Soc. Amer.*, vol. 94, no. 3, pp. 1332–1342, Sept. 1993.

[2] 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. of IWAENC*, Kyoto, Japan, Sept. 2003.

[3] S. Doclo and M. Moonen, “Multi-microphone noise reduction using recursive GSVD-based optimal filtering with ANC post-processing stage,” To appear in IEEE Trans. ASSP, available at ftp://ftp.esat.kuleuven.ac.be/pub/sista/doclo/reports/02-04.ps.gz.

[4] 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.

[5] A. Spriet, M. Moonen, and J. Wouters, “A multi-channel sub-band GSVD approach to speech enhancement,” *ETT*, vol. 13, no. 2, pp. 149–158, Mar.-Apr. 2002.

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

[7] D. A. Florêncio and H. S. Malvar, “Multichannel filtering for optimum noise reduction in microphone arrays,” in *Proc. of ICASSP*, Salt Lake City, Utah, May 2001.

[8] 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, K.U. Leuven (Belgium), 2003, available at ftp://ftp.esat.kuleuven.ac.be/pub/sista/spriet/reports/03-47.pdf

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