SVD-Based Optimal Filtering Technique for Noise Reduction in Hearing Aids Using Two Microphones

Jean-Baptiste Maj

Lab. Exp. ORL, Katholieke Universiteit Leuven, Kapucijnenvoer 33, 3000 Leuven, Belgium

ESAT-SISTA, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B 3001 Leuven, Belgium
Email: Jean-Baptiste.Maj@uz.kuleuven.ac.be

Marc Moonen

ESAT-SISTA, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B 3001 Leuven, Belgium
Email: Marc.Moonen@esat.kuleuven.ac.be

Jan Wouters

Lab. Exp. ORL, Katholieke Universiteit Leuven, Kapucijnenvoer 33, 3000 Leuven, Belgium
Email: Jan.Wouters@uz.kuleuven.ac.be

We introduce a new SVD-based (singular value decomposition) strategy for noise reduction in hearing aids. This technique is evaluated for noise reduction in a behind-the-ear (BTE) hearing aid where two omnidirectional microphones are mounted in an endfire configuration. The behaviour of the SVD-based technique is compared to a two-stage adaptive beamformer for hearing aids developed by Vanden Berghe and Wouters (1998). The evaluation and comparison is done with a performance metric based on the speech intelligibility index (SII). The speech and noise signals are recorded in reverberant conditions with a signal-to-noise ratio of 0 dB and the spectrum of the noise signals is similar to the spectrum of the speech signal. The SVD-based technique works without initialization nor assumptions about a look direction, unlike the two-stage adaptive beamformer. Still, for different noise scenarios, the SVD-based technique performs as well as the two-stage adaptive beamformer, for a similar filter length and adaptation time for the filter coefficients. In a diffuse noise scenario, the SVD-based technique performs better than the two-stage adaptive beamformer and hence provides a more flexible and robust solution under speaker position variations and reverberant conditions.

Keywords and phrases: SVD-based technique, beamforming, noise reduction, hearing aids, speech intelligibility, diffuse noise.

1. INTRODUCTION

A major problem for hearing impaired listeners is the understanding of speech in noise. Indeed, their speech reception threshold (SRT: defined as the sound-pressure level of speech at which 50% of the speech is correctly understood by the listener) in noise, interfering noises or competing speakers, is higher than for normal hearing subjects [1]. To compensate for this difference, several noise reduction strategies with one or multiple microphones have been developed.

As for *single microphone approaches*, a noise reduction system in hearing aids is typically based on a hardware directional microphone. Some studies have shown that the directional microphone may give an SRT improvement of about 3 dB in difficult listening conditions [2, 3]. With other methods, such as spectral subtraction or Wiener filtering, an improvement in physical signal-to-noise ratio (SNR) has

been found, but unfortunately a similar improvement of the speech reception thresholds has not been observed [4].

As for *multiple microphone approaches*, beamforming is by far the most developed method. There are fixed and adaptive beamformers. Fixed beamformers focus on a target direction independent of the interfering signals. This has given significant improvement of the SRT in anechoic conditions. Soede et al. [5] obtained an SRT improvement of 7 dB with an endfire array configuration (an endfire array has its microphones colinear with the target direction), where five cardioid microphones were spaced on a 10 cm-long array, as compared with an omnidirectional microphone system. Similar approaches to the Soede strategy have been investigated [6, 7] and comparisons between these different methods have shown improvements of speech intelligibility in noise [8, 9].

Widrow and Stearns [10] introduced adaptive noise cancellation (ANC), where the aim is to null out the interfering

noise source(s). Griffiths and Jim gave an extension of the ANC, commonly known as "Griffiths-Jim beamforming," and for this approach, several studies have shown improvements in speech intelligibility for directional jammers [11, 12]. These studies were carried out with different microphone arrays and with monaural or binaural conditions. Adaptive systems are known to give a significant improvement in speech intelligibility when the noise is located, but not when it is diffuse [12, 13]. Generally, these studies tested systems with a microphone array larger than the behind-the-ear (BTE) device (which is a practical impediment) and without any limitations on the available computing power.

Based on the work by Van Compernolle [14], Vanden Berghe and Wouters [15] have developed a two-stage adaptive beamformer for BTE devices. This technique was implemented on a wearable digital signal processor (DSP) platform and tested with a hearing aid that contained a hardware directional microphone and an omnidirectional microphone spaced by 1.5 cm. A significant SRT improvement of about 11 dB was obtained between a hardware directional microphone and the output of the two-stage adaptive beamformer for a directional jammer [16]. Unfortunately, this performance is dependent on the microphone configuration used.

Here, we study the behaviour of a new multiple microphone signal enhancement technique based on SVD [17]. Signals coming from a normal-sized BTE device are used to evaluate offline the speech enhancement using a performance metric based on the so-called SII [18]. The hearing aid contains *two omnidirectional microphones*, spaced by 2 cm, and a dual microphone technique is implemented to create a directional microphone (see Figure 1). The signals of the software directional microphone and the rear omnidirectional microphone are used as input of the noise reduction algorithms. The SVD-based technique is compared to the two-stage adaptive beamformer of Vanden Berghe in different noise conditions. These noises have the same spectrum as the speech signals and the SNR of the recorded signals was 0 dB. In addition, a comparison has been performed for these algorithms in reverberant and diffuse noise conditions. Also, different parameters of the SVD-based technique were optimized, to improve the speech intelligibility.

2. METHODS

2.1. SVD-based multi-microphone noise reduction

2.1.1 Basic principle

Assume that the microphone samples processed at time k are stacked into a vector ($N \times 1$) $\mathbf{u}_k$, and that $\mathbf{u}_k = \mathbf{s}_k + \mathbf{n}_k$, where $\mathbf{s}_k$ is the speech part and $\mathbf{n}_k$ is the noise part. To extract the speech part $\mathbf{s}_k$ from $\mathbf{u}_k$, McAulay and Malpass [19] used the well-known Wiener filter which minimizes the mean-square error (MSE). The estimate of the speech $\tilde{\mathbf{s}}_k$ takes the form

$$\tilde{\mathbf{s}}_k = \mathbf{W}^T \mathbf{u}_k \tag{1}$$

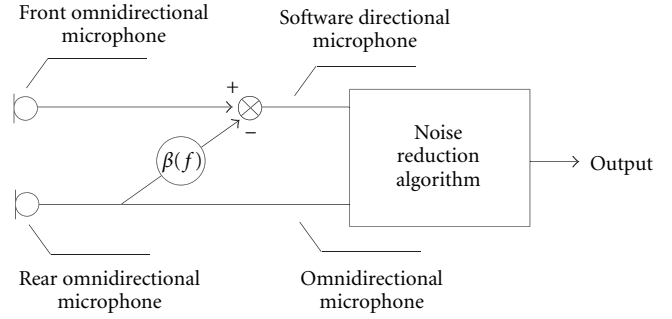with $\mathbf{W} \in \mathbb{R}^{N \times N}$ the optimal filter (with N the size of the filter

FIGURE 1: Strategy for noise reduction for hearing aids.

per channel), $\mathbf{u}_k \in \mathbb{R}^N$ being the filter input vector at time k, and $\tilde{\mathbf{s}}_k \in \mathbb{R}^N$ the filter output vector. The ith column of $\mathbf{W}$ is then an optimal filter for the ith component of $\mathbf{s}_k$. Then, the error of the estimation at time k is defined by $\mathbf{e}_k = \mathbf{s}_k - \tilde{\mathbf{s}}_k$, and the MSE cost function for optimal filtering is

$$
\begin{aligned}
\mathbf{J}_{\text{MSE}} &= \mathscr{E}\{\|\mathbf{e}_k\|_2^2\} \\
&= \mathscr{E}\{\|(\mathbf{s}_k - \tilde{\mathbf{s}}_k)\|_2^2\} \\
&= \mathscr{E}\{\|(\mathbf{s}_k - \mathbf{W}^T \mathbf{u}_k)\|_2^2\} \\
&= \mathscr{E}\{(\mathbf{s}_k - \mathbf{W}^T \mathbf{u}_k)^T (\mathbf{s}_k - \mathbf{W}^T \mathbf{u}_k)\} \\
&= \mathscr{E}\{\mathbf{s}_k^T \mathbf{s}_k\} - 2\mathscr{E}\{\mathbf{u}_k^T \mathbf{W} \mathbf{s}_k\} + \mathscr{E}\{\mathbf{u}_k^T \mathbf{W} \mathbf{W}^T \mathbf{u}_k\}.
\end{aligned}
\tag{2}
$$

The optimal filter is found by setting the derivative $\partial \mathbf{J}_{\text{MSE}} / \partial \mathbf{W}$ equal to zero

$$\frac{\partial \mathbf{J}_{\text{MSE}}}{\partial \mathbf{W}} = -2\mathscr{E}\{\mathbf{u}_k \cdot \mathbf{s}_k^T\} + 2\mathscr{E}\{\mathbf{u}_k \cdot \mathbf{u}_k^T\}\mathbf{W} = 0. \tag{3}$$

Hence, the optimal filter $\mathbf{W}_{\text{WF}}$ is

$$\mathbf{W}_{\text{WF}} = \mathscr{E}\{\mathbf{u}_k \cdot \mathbf{u}_k^T\}^{-1} \cdot \mathscr{E}\{\mathbf{u}_k \cdot \mathbf{s}_k^T\}. \tag{4}$$

However, obtaining an estimate for $\mathscr{E}\{\mathbf{u}_k \cdot \mathbf{s}_k^T\}$ is not straightforward as $\mathbf{s}_k$, obviously, is unknown. If we use a robust speech/noise detection algorithm, and noise-only observations can be made during speech pauses at time $k'(\mathbf{u}_{k'} = 0 + \mathbf{n}_{k'})$, then we can use such observations to estimate

$$\mathscr{E}\{\mathbf{n}_k \cdot \mathbf{n}_k^T\} = \mathscr{E}\{\mathbf{n}_{k'} \cdot \mathbf{n}_{k'}^T\} = \mathscr{E}\{\mathbf{u}_{k'} \cdot \mathbf{u}_{k'}^T\}, \tag{5}$$

which means that we are able to estimate $\mathscr{E}\{\mathbf{n}_k \cdot \mathbf{n}_k^T\}$. The leftmost term, $\mathscr{E}\{\mathbf{n}_k \cdot \mathbf{n}_k^T\}$, indicates that short-term noise stationarity is assumed. During *speech activity* (at time k), we observe both the signal-of-interest and the noise signal, $\mathscr{E}\{\mathbf{u}_k \cdot \mathbf{u}_k^T\}$. If we assume that $\mathbf{s}_k$ and $\mathbf{u}_k$ are statistically independent of each other ($\mathscr{E}\{\mathbf{s}_k \cdot \mathbf{n}_k^T\} = 0$), then

$$
\begin{aligned}
\mathscr{E}\{\mathbf{u}_k \cdot \mathbf{u}_k^T\} &= \mathscr{E}\{\mathbf{s}_k \cdot \mathbf{s}_k^T\} + \mathscr{E}\{\mathbf{s}_k \cdot \mathbf{n}_k^T\} \\
&\quad + \mathscr{E}\{\mathbf{n}_k \cdot \mathbf{s}_k^T\} + \mathscr{E}\{\mathbf{n}_k \cdot \mathbf{n}_k^T\} \\
&= \mathscr{E}\{\mathbf{n}_k \cdot \mathbf{n}_k^T\} + \mathscr{E}\{\mathbf{s}_k \cdot \mathbf{s}_k^T\}.
\end{aligned}
\tag{6}
$$

Given $\mathcal{E}\{\mathbf{u}_k \cdot \mathbf{u}_k^T\}$ and $\mathcal{E}\{\mathbf{n}_k \cdot \mathbf{n}_k^T\}$, we can thus compute $\mathcal{E}\{\mathbf{s}_k \cdot \mathbf{s}_k^T\}$. Finally, from the assumed independence of $\mathbf{s}_k$ and $\mathbf{n}_k$, it also follows that

$$\mathcal{E}\{\mathbf{u}_k \cdot \mathbf{s}_k^T\} = \mathcal{E}\{\mathbf{s}_k \cdot \mathbf{s}_k^T\} + \mathcal{E}\{\mathbf{s}_k \cdot \mathbf{n}_k^T\} = \mathcal{E}\{\mathbf{s}_k \cdot \mathbf{s}_k^T\}, \qquad (7)$$

so that the optimal filter $\mathbf{W}_{\text{WF}}$ is finally given by

$$\begin{aligned}\mathbf{W}_{\text{WF}} &= \mathcal{E}\{\mathbf{u}_k \cdot \mathbf{u}_k^T\}^{-1} \cdot \mathcal{E}\{\mathbf{s}_k \cdot \mathbf{s}_k^T\} \\ &= \mathcal{E}\{\mathbf{u}_k \cdot \mathbf{u}_k^T\}^{-1} \cdot (\mathcal{E}\{\mathbf{u}_k \cdot \mathbf{u}_k^T\} - \mathcal{E}\{\mathbf{n}_{k'} \cdot \mathbf{n}_{k'}^T\}).\end{aligned} \qquad (8)$$

Also, an approach often used to approximate the Wiener filter in the frequency domain is (see [20])

$$\mathbf{W}_{\text{WF}}(f) = (\mathcal{E}\{|S(f)|^2\} + \mathcal{E}\{|N(f)|^2\})^{-1} \cdot \mathcal{E}\{|S(f)|^2\}. \qquad (9)$$

The function $\mathcal{E}\{|N(f)|^2\}$ is obtained by averaging many frames during silence intervals in which the statistics of the background noise can be assumed stationary. To estimate $\mathcal{E}\{|S(f)|^2\}$, there are different possibilities [21, 22].

Doclo and Moonen [17] use an interesting and useful simplification in formula (8), of which more theoretical justification can be found in [23]. The matrix $\mathbf{W}_{\text{WF}}$ is derived from the joint diagonalization *(generalized eigenvalue decomposition)* of the symmetric matrices $\mathcal{E}\{\mathbf{u}_k \cdot \mathbf{u}_k^T\}$ and $\mathcal{E}\{\mathbf{n}_{k'} \cdot \mathbf{n}_{k'}^T\}$,

$$\begin{aligned}\mathcal{E}\{\mathbf{u}_k \cdot \mathbf{u}_k^T\} &= \mathbf{X} \cdot \text{diag}\{\sigma_i^2\} \cdot \mathbf{X}^T, \\ \mathcal{E}\{\mathbf{n}_{k'} \cdot \mathbf{n}_{k'}^T\} &= \mathbf{X} \cdot \text{diag}\{\eta_i^2\} \cdot \mathbf{X}^T,\end{aligned} \qquad (10)$$

where $\mathbf{X}$ is an invertible (but not necessarily orthogonal) matrix. Note that $\text{diag}\{\sigma_i^2\}$ represents a diagonal matrix with diagonal elements σ_i^2, $i = 1, \ldots, N$, and that $\text{diag}\{\eta_i^2\}$ is similarly defined. In practice, $\mathbf{X}$, σ_i^2, η_i^2 are computed by means of a *generalized singular value decomposition* of the data matrices $\mathbf{U}_k \in \mathbb{R}^{p \times N}$ and $\mathbf{N}_k \in \mathbb{R}^{q \times N}$ (with p and q typically larger than N),

$$\mathbf{U}_k = \begin{bmatrix} \mathbf{u}_k^T \\ \mathbf{u}_{k+1}^T \\ \vdots \\ \mathbf{u}_{k+p-1}^T \end{bmatrix}, \qquad \mathbf{N}_k = \begin{bmatrix} \mathbf{n}_{k'}^T \\ \mathbf{n}_{k'+1}^T \\ \vdots \\ \mathbf{n}_{k'+p-1}^T \end{bmatrix} \qquad (11)$$

such that $\mathcal{E}\{\mathbf{u}_k \cdot \mathbf{u}_k^T\} \Rightarrow (\mathbf{U}_k^T \cdot \mathbf{U}_k)/p$ and $\mathcal{E}\{\mathbf{n}_{k'} \cdot \mathbf{n}_{k'}^T\} \Rightarrow (\mathbf{N}_k^T \cdot \mathbf{N}_k)/q$. The generalized singular value decomposition of the matrices $\mathbf{U}_k$ and $\mathbf{N}_k$ is defined as

$$\begin{aligned}\mathbf{U}_k &= \mathbf{Y} \cdot \text{diag}\{\sigma_i\} \cdot \mathbf{X}^T, \\ \mathbf{N}_k &= \mathbf{Z} \cdot \text{diag}\{\eta_i\} \cdot \mathbf{X}^T,\end{aligned} \qquad (12)$$

where $\mathbf{Y} \in \mathbb{R}^{p \times N}$ and $\mathbf{Z} \in \mathbb{R}^{q \times N}$ are orthogonal matrices, $\mathbf{X} \in \mathbb{R}^{N \times N}$ is an invertible matrix and σ_i/η_i are the generalized singular values. By substituting the above formulas in (8), we obtain

$$\mathbf{W}_{\text{WF}} = \mathbf{X}^{-T} \cdot \text{diag}\left\{\frac{\sigma_i^2 - (p/q) \times \eta_i^2}{\sigma_i^2}\right\} \cdot \mathbf{X}^T. \qquad (13)$$

2.1.2 Single microphone applications

In single microphone speech processing, generally the vector $\mathbf{u}_k$ is taken from a time series $u(k)$, that is,

$$\mathbf{u}_k = \begin{bmatrix} u(k) & u(k-1) & u(k-2) & \cdots & u(k-N+1) \end{bmatrix}^T \qquad (14)$$

and similarly

$$\begin{aligned}\mathbf{s}_k &= \begin{bmatrix} s(k) & s(k-1) & s(k-2) & \cdots & s(k-N+1) \end{bmatrix}^T, \\ \mathbf{n}_k &= \begin{bmatrix} n(k) & n(k-1) & n(k-2) & \cdots & n(k-N+1) \end{bmatrix}^T.\end{aligned} \qquad (15)$$

The data matrices $\mathbf{U}_k \in \mathbb{R}^{pN}$ and $\mathbf{N}_k \in \mathbb{R}^{qN}$, as defined in (11), are then Toeplitz matrices, for example,

$$\begin{aligned}\mathbf{U}_k &= \begin{bmatrix} \mathbf{u}_k^T \\ \mathbf{u}_{k+1}^T \\ \vdots \\ \mathbf{u}_{k+p-1}^T \end{bmatrix} \\ &= \begin{bmatrix} u(k) & u(k-1) & \cdots & u(k-N+1) \\ u(k+1) & u(k) & \cdots & u(k-N+2) \\ \vdots & \vdots & & \vdots \\ u(k+p-1) & u(k+p-2) & \cdots & u(k+p-N) \end{bmatrix}.\end{aligned} \qquad (16)$$

The optimal filtering approach of (1) then provides an optimal filtered estimate $\tilde{s}(k)$ at time k, but in addition optimal smoothed estimates $\tilde{s}(k-1), \tilde{s}(k-2), \ldots, \tilde{s}(k-N+1)$ (all at time k). Conversely, optimal estimates for $s(k)$ are obtained at time $k, k+1, \ldots, k+N-1$, corresponding to different columns of $\mathbf{W}_{\text{WF}}$, applied to, respectively, $\mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_{k+N-1}$. The question then arises which estimate should be picked. A *first solution* consists in computing the error covariance matrix. The estimation error $\mathbf{e}_k$ is defined as

$$\mathbf{e}_k = \mathbf{s}_k - \tilde{\mathbf{s}}_k = \mathbf{s}_k - \mathbf{W}_{\text{WF}}^T \mathbf{u}_k. \qquad (17)$$

It is easily shown that the error covariance matrix is given as

$$\mathcal{E}\{\mathbf{e}_k \cdot \mathbf{e}_k^T\} = \mathcal{E}\{\mathbf{n}_k \cdot \mathbf{n}_k^T\} \cdot \mathbf{W}_{\text{WF}}. \qquad (18)$$

In particular, we are interested in the diagonal elements of the error covariance matrix $\{\mathcal{E}\{\mathbf{n}_k \cdot \mathbf{n}_k^T\} \cdot \mathbf{W}_{\text{WF}}\}_{ii}$. Indeed, the smallest element on the main diagonal of this error covariance matrix corresponds to the best estimator $\mathbf{w}_1 = \mathbf{W}_{\text{WF}}(:, i')$, where $i' = \arg\min_i(\mathcal{E}\{\mathbf{e}_k \cdot \mathbf{e}_k^T\})_{ii}$. A *second solution* consists in computing an average over all available estimates. This technique is often applied to rank truncation based estimation [24, 25]. The corresponding filter is then given as

$$\mathbf{w}_2 = \left[\frac{1}{N} \quad \frac{1}{N} \quad \cdots \quad \frac{1}{N} \quad \frac{1}{N} \right]$$

$$\cdot \begin{bmatrix} 0 & 0 & \cdots & 0 & w(1,1) & \cdots & w(N-1,1) & w(N,1) \\ 0 & 0 & \cdots & w(1,2) & w(2,2) & \cdots & w(N,2) & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & w(1,N-1) & \cdots & w(N-2,N-1) & w(N-1,N-1) & \cdots & 0 & 0 \\ w(1,N) & w(2,N) & \cdots & w(N-1,N) & w(N,N) & \cdots & 0 & 0 \end{bmatrix} \quad (19)$$

with $w(i,j)$ denoting the (i,j)-element of $\mathbf{W}_{\text{WF}}$. A *third solution* is to simply take an arbitrary column of $\mathbf{W}_{\text{WF}}$ and use that as a noise reduction filter, $\mathbf{w}_3 = \mathbf{W}_{\text{WF}}(:,i)$ for random $i \in \{1, \ldots, N\}$. A comparison between these three different strategies will be given in Section 3.1.

2.1.3 Multiple microphones applications

In a multiple microphone application, the vector $\mathbf{u}_k \in \mathbb{R}^{MN}$ takes the form

$$\mathbf{u}_k = \begin{bmatrix} \mathbf{u}_{1k} \\ \mathbf{u}_{2k} \\ \vdots \\ \mathbf{u}_{Mk} \end{bmatrix} \quad (20)$$

with

$$\mathbf{u}_{jk} = \begin{bmatrix} u_j(k) & u_j(k-1) & \cdots & u_j(k-N+1) \end{bmatrix}^T, \quad (21)$$

where j refers to the jth microphone and M is the number of the microphone inputs. In our case we have two microphones, so $j = 1, 2$. The vector $\mathbf{n}_k$ is similarly defined. The data matrices $\mathbf{U}_k \in \mathbb{R}^{(p \times MN)}$ and $\mathbf{N}_k \in \mathbb{R}^{(q \times MN)}$ as defined in (11) then take the form

$$\mathbf{U_k} = \begin{bmatrix} \mathbf{U}_{1k} & \mathbf{U}_{2k} \end{bmatrix},$$

$$\mathbf{U}_{jk} = \begin{bmatrix} u_j(k) & u_j(k-1) & \cdots & u_j(k-N+1) \\ u_j(k+1) & u_j(k) & \cdots & u_j(k-N+2) \\ \vdots & \vdots & \vdots & \vdots \\ u_j(k+p-1) & u_j(k+p-2) & \cdots & u_j(k+p-N) \end{bmatrix}. \quad (22)$$

Using similar formulas as in the one-channel case, the optimal filter $\mathbf{W}_{\text{WF}}$ and then the $(2 \times N)$-taps estimator $\mathbf{w}$ can be computed, so that the estimated signal $\bar{\mathbf{s}}_k$ is given as

$$\hat{\mathbf{s}}_k = \begin{bmatrix} \hat{s}(k) \\ \hat{s}(k+1) \\ \vdots \\ \hat{s}(k+p-1) \end{bmatrix} = \begin{bmatrix} \mathbf{U}_{1k} & \mathbf{U}_{2k} \end{bmatrix} \cdot \mathbf{w}, \quad (23)$$

where $\bar{\mathbf{s}}_k$ is an estimate for the (delayed version of the) speech part of either microphone signal 1 or microphone signal 2 depending on the choice for $\mathbf{w}$.
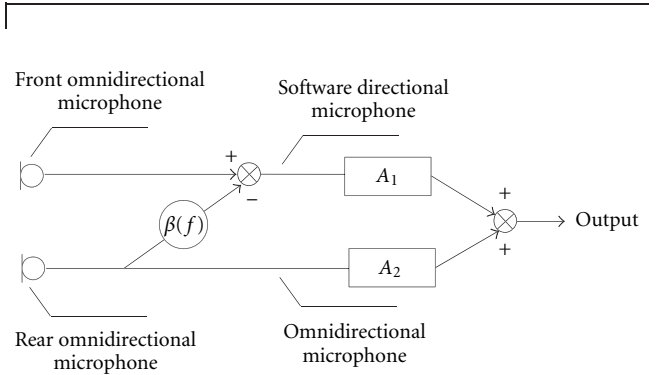

Figure 2: Scheme of the SVD-based strategy.

This filter can be considered as a two-channel filter (see Figure 2), where each microphone is filtered with an N-taps filter A_j

$$\mathbf{w}_{\text{WF}} = \begin{bmatrix} A_1^T & A_2^T \end{bmatrix}^T. \quad (24)$$

2.2. Two-stage adaptive beamformer

The behaviour of the SVD-based technique is compared with another noise reduction algorithm. This algorithm is a two-stage adaptive beamformer which has been developed by Vanden Berghe and Wouters [15]. The complete noise reduction strategy is depicted in Figure 3. The two-stage adaptive beamformer, which is based on adaptive noise cancellation (ANC), attempts to model noise during noise periods (where only noise signal is present), and subtracts noise from speech plus noise when speech is present. A speech detection algorithm is implemented to decide whether the signal contains speech plus noise or noise only. The sum and subtraction (middle part of Figure 3) based on the scheme of Griffiths and Jim [26] improve the noise reference of the ANC. Van Compernolle [14] used a second adaptive filter instead of the fixed filter, which adapts when desired speech is present and compensates for the different transfer functions between the speaker and the microphone array. It has been shown that this procedure experiences robustness problems. For example, it can converge to a wrong direction when different speakers are around the listener or if loud noise is wrongly detected as speech by the speech detection algorithm. For this reason, the first filter will be kept fixed, under the assumption that the speaker is always in front of the listener. In fact, we gave a specific look direction to the two-stage adaptive beamformer, namely the direction of the desired signal,
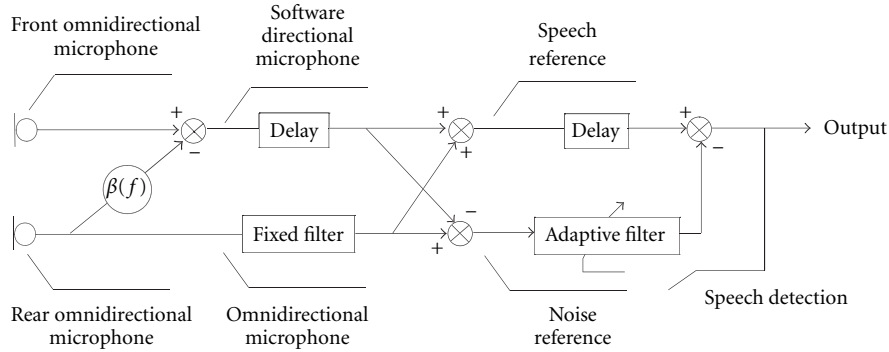
Figure 3: Scheme of the two-stage adaptive beamformer strategy.

at 0°. To determine the coefficients of the first filter, we used a stationary ICRA-signal as described below. In anechoic conditions, this signal was presented by a loudspeaker 1 meter in front of the dummy head with the hearing aid. The first filter's coefficients were adapted, by means of a normalized least mean squares procedure (NLMS) [27], during few seconds and then the adaptation was stopped. The obtained coefficients are used from then on to create the fixed first filter. Thus, a front cardioid and a back cardioid are obtained, respectively, at the speech reference and the noise reference. The number of coefficients are 10 and 30, respectively, for the first and the second filter. The additional delays (Figure 3) actually allow to have noncausal filters, and their values are set to half of the size of the filters (5 and 15). The second filter is an adaptive filter, also adapted by an NLMS procedure.

2.3. Performance metric

To evaluate the improvement of the speech intelligibility of noise reduction algorithms different performance metrics have been developed, which are mostly based on an averaged intelligibility gain. The basic idea was introduced by Peterson [11]. Greenberg et al. [28] developed an SNR measure where signals of output and input were decomposed in third octave bands and for each frequency band, weights were applied as defined in the articulation index calculation [29]. Hoffman et al. [30] introduced an improvement by limiting the speech peak-to-noise ratios to the range of 0 to 30 dB. Unfortunately, these performance metrics do not take into account the masking and the reverberation effect and, respectively, Kates and Weiss [8], Saunders and Kates [9], gave solutions that also consider these two effects. Since 1997, an extension of the articulation index calculation has been suggested. This extension is known as the SII [18], and comprises a variety of adverse listening conditions, such as noise masking, filtering, distortion, and low reverberation. Thus, we define a performance metric based on the SII

$$\text{SNR}_{\text{weigthed}} = \sum_{i=1}^{k} I_i \cdot A_i \cdot \text{SNR}_i, \tag{25}$$

where SNR_i is the signal-to-noise ratio measured (in dB-SPL) in the ith third octave band; I_i and A_i are the weights for the importance of the band and the audibility function, respectively, as described by [18]. Thus, $\text{SNR}_{\text{weigthed}}$ is a weighted SNR, where the weights reflect speech intelligibility. As Greenberg et al. [28], we estimate speech reception threshold (SRT) improvement between the input, the omnidirectional microphone in our case, and the output of the two-stage adaptive beamformer or the SVD-based techniques

$$\text{SNR}_{\text{SII}} = \text{SNR}_{\text{weigthed, output}} - \text{SNR}_{\text{weigthed, input}}. \tag{26}$$

Maj et al. [13] obtained a mean of 0.63 dB for the difference between the SRT values obtained by performing listening tests with subjects and the computed SNR_{SII}. For this evaluation to be valid, a linear relation has to be assumed between the input and the output of the system. This means that the hearing aid must have a linear amplification and the noise reduction system, the two-stage adaptive beamformer or the SVD-based technique, must implement a linear transformation of the input signals. Also, to measure the performance, the system must not be in transient conditions. As for the linearity conditions, we know that the hearing aid has a linear amplification, and that the noise reduction systems introduce a linear transformation of the input signals. As for the no transient conditions, the evaluation is done with stationary speech weighted noise and we took care that the different algorithms were given enough time to converge.

2.4. Hearing aid

To study the behaviour of the SVD-based technique on speech intelligibility improvement, we have applied it to signals which were recorded with a behind-the-ear hearing aid. The hearing aid is a Danavox-163D hearing aid housing where two omnidirectional microphones (Knowles Electronics-EM4368) are mounted in an endfire array configuration spaced 2 cm apart. Furthermore, in the first signal processing stage, a software directional microphone [31] is created with the omnidirectional microphones (Figure 1). The directional microphone signal is computed as the difference between the signal from the front microphone and the delayed-weighted signal of the rear microphone, resulting in a response comparable to a hardware directional microphone. The microphone parameters are the interport

distance d, internal delay τ, and the weight factor for the back port is $\beta(f) = a \cdot e^{-j2\pi f \tau}$. The delay τ and the weight a have been chosen to give a hypercardioid spatial characteristic in anechoic conditions. The hearing aid has a linear amplification and does not have systems for compression or feedback control.

2.5. Signals

Two neighboring rooms were used for the experiments. The first room simulates the acoustics of a living room situation with a volume of $63\,\mathrm{m}^3$ and a reverberation time of 0.76 s. The estimated direct-to-reverberant ratio is 5.20 dB at 1 m [32]. In this room, the hearing aid was positioned on a dummy head and loudspeakers (Yamaha CBX-S3) were situated at 1m from the dummy head (see Figures 4 and 5).

The signals of the two microphones of the hearing aid were amplified and connected to the second room where further signal processing took place. The signal of the microphones were amplified with a two-channel Larson Davis 2200 C amplifier and were digitized, at a sampling frequency of 16000 Hz, with a PC-platform using a Texas Instruments TMS320C40 digital signal processor (DSP) and two input channels with 16-bit analog-to-digital conversion (ADC). The amplifier of the hearing aid and the different devices have been checked to ensure that they were not saturating.

The speech and noise signals, which were sent to the loudspeakers, were stationary speech weighted noises, namely ICRA, BLU, NVA, and BRUGSE. The ICRA-signal (International Collegium of Rehabilitative Audiology) is based on the multilanguage long-term average speech spectrum (LTASS). It has been produced by the Hearing Aid Clinical Test Environment Standardization Work Group, and is a white noise signal filtered in close accordance with the LTASS [33] and ANSI S3.79 [18]. The signal is unmodulated random Gaussian noise representative of a male weighted idealized speech spectrum with a normal effort [34]. The BLU-signal, NVA-signal, and BRUGSE-signal are described and available on the Compact-Disk of Wouters [35]. They are also stationary speech-weighted noises but they do not have the same spectrum as the ICRA signal.

2.6. Optimization of the SVD-based technique

In these experiments, one loudspeaker (L_0) was situated in front of the dummy head (at 0°) and presents the speech material, and a second (L_{90}) at 90° (on the side of the hearing aid) presents noise material (see Figure 4).

We used an ICRA-signal at 0° for speech and a similar ICRA-signal at 90° for noise, implying that the noise and speech signals have the same spectrum. These two signals were recorded separately, that is, the speech was recorded when only the loudspeaker in front of the dummy head was active, and the noise was recorded when only the loudspeaker on the side of the dummy head was active. This provides additional experimental flexibility. The two loudspeakers have the same frequency response and were matched to get a level of 70 dBSPL at the center of the dummy head. We recorded the signals during 10 seconds, thus we got two signals with
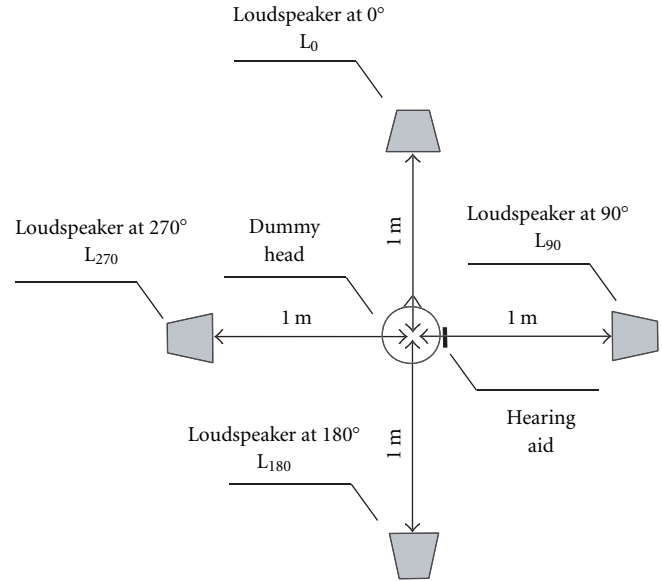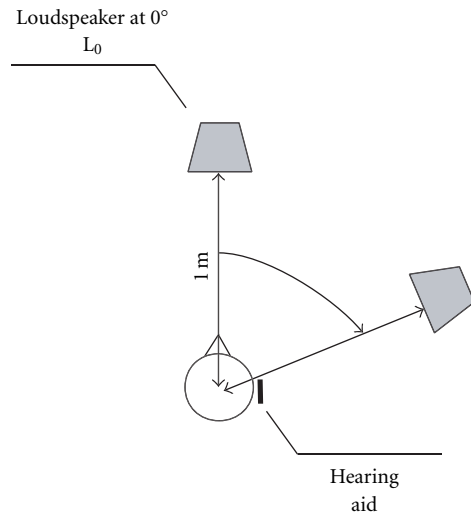

Figure 4: Setup for the recording of the signals.

Figure 5: Setup for the recording of the signals in one noise source scenario.

160000 samples for speech as well as for noise. From these signals, a noise frame was created only with the noise signal, and a speech-and-noise frame by the addition of the speech and the noise signals.

We define a point inside the original 160000 samples frames referred to as "*start speech.*" "*Start speech*" can be interpreted as a perfect "speech detection" and means that before the point "*start speech*" we have only noise (we use the noise frame) and after this point, we have speech and noise (we use the speech-and-noise frame).

With the SVD-based algorithm, the optimal filter $\mathbf{w}_{\mathrm{WF}}$ can be viewed as a two-channel filter, where each microphone is filtered with an N-taps filter A_j, see formula (24).

The resulting estimated signal is computed by filtering, and summing the microphone signals s_k and n_k with the filters A_j (see Figure 2). The SNR_{SII} computation uses these two filtered signals separately to make the evaluation of the algorithm performance. For the two-stage adaptive beamformer the coefficients of the second filter were adapted on the noise signal until convergence and then they were kept fixed. Then, the speech and noise signals were filtered separately with the fixed coefficients, and the evaluation of the SNR_{SII} performance was measured in the same fashion as for the SVD-based technique.

2.6.1 Optimal estimate

As already mentioned, $\mathbf{W}_{WF}$ (formula (13)) contains a collection of $2N$-taps filters. We need to make a choice for $\mathbf{w}_{WF}$ to obtain an estimate for $\mathbf{s}(k)$. In Section 2.1.2, we have described three different methods, an error-covariance matrix based method, an averaging over all available estimates, and an arbitrary choice between the columns of the filter matrix $\mathbf{W}_{WF}$. The SVD-based algorithm was applied with a speech-and-noise frame and noise frame of 2000 samples and the filter length N was 20 per channel.

2.6.2 Influence of the design parameters

The SVD-based algorithm depends on three parameters: the size of the noise frame (q), the size of the speech-and-noise frame (p), and the length of the filter for each channel (N). In the next paragraph, we describe the experiments to study the influence of these parameters on the speech intelligibility performance through the SNR_{SII}.

The signals $\mathbf{u}_k$ and $\mathbf{n}_k$ are stationary, which implies that for different values of "*start speech,*" we can expect to obtain the same SNR_{SII} improvement. We applied the SVD-based algorithm with the strategy described in Section 2.3 for fourteen different values of "*start speech.*" With the beamformer, the second stage adapted on the 10000 samples of the noise frame before the "*start speech.*" We calculated the standard deviation and the mean of the improvement with the fourteen different values for "*start speech*" for both algorithms. Also, with the SVD-based technique, we varied the length of the speech-and-noise frame and the noise frame, while the filter size is kept fixed ($N = 20$).

Next, we investigate the influence of the filter size N for each channel. A size of 5000 samples for the speech-and-noise frame and a size of 15000 samples for the noise frame are chosen. The "*start speech*" is taken in the middle of the file, at sample 80000 (= 5 s).

It is instructive to also investigate the improvement that can be obtained with a single microphone configuration. The method is the same as for the two microphones configuration. We have chosen the middle column of the matrix $\mathbf{W}_{WF}$ as the optimal filter, and we have varied the different lengths of the speech-and-noise frame, noise frame, and the size of the filter.

Finally, it is important to know if the algorithms distort or have an influence on the speech signal itself. To investigate this effect, we analyze the transfer function between the system input and output. For this and the next experiments, the "*start speech*" is taken in the middle of the file. The adaptation of the beamformer second filter was performed during 10000 samples of the noise period. With the SVD-based technique, we have calculated the filter with a length of 5000 samples for the speech frame and 15000 samples for the speech-and-noise frame. The length of the filter for each channel was 20.

2.7. Noise scenarios

2.7.1 One noise source

In this experiment, the recordings have been made with the same "ICRA" signals and under the same conditions as described in Section 2.6. The signals were recorded for different locations of the noise source, corresponding to angles between 0° and 345° in steps of 15° (90° is the side of the hearing aid). The desired speech source is always located at 0° (see Figure 5).

2.7.2 Multiple noise sources

For scenarios with multiple and different noise sources, we used four loudspeakers (see Figure 4), one for the speech signal (L_0) and three for the noise signals (L_{90}, L_{180}, and L_{270}). For the speech signal, we have again used "ICRA" signals, but for the noises we have utilized the ICRA-signal, BLU-signal, NVA-signal, and BRUGSE-signal. The sum of the signals has been recorded with the same level (70 dBSPL) at the middle of the dummy head.

2.7.3 Diffuse noise

Until now, we have focused on localized noise sources. However, the sound field in a large room or in a car may appear as diffuse noise. A diffuse noise does not have a well-specified direction. To create the diffuse noise, we have made a setup as defined by Veit and Sander [36].

The reverberation time of the room was $T_{60} = 0.88$ s for a speech-weighted noise. The estimated direct to reverberant ratio is 4.55 dB at 1 m. The point M is in the middle of eight loudspeakers. The level of each loudspeaker has been fitted to give the same sound level at the point M (61 dB) and when the eight loudspeakers work together we obtain a sound level of 70 dB at M. For the measurements, we have placed the center of the dummy head at the point M. To create the speech signal, a loudspeaker was put in front of the dummy head at 1 m at different angles between −45° and +45° in steps of 15°. The two signal contributions, speech, and diffuse noise, were recorded separately.

3. RESULTS AND DISCUSSION

3.1. Optimal estimate

The results are shown in Figure 6. The error covariance matrix method (ECM) clearly performs better than the averaging method and selects the optimal estimator (column 10). The improvement for each column of $\mathbf{W}_{WF}$ shows two different parts, namely between the columns 1 to 20, and the columns 21 to 40. In fact, in the first part we have the data of
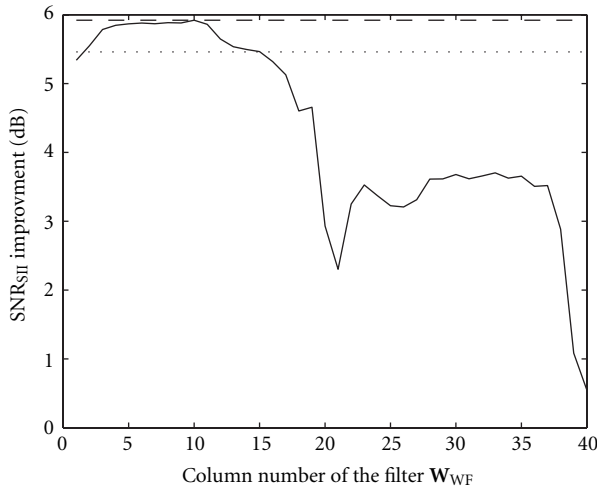
FIGURE 6: Results for the different methods to determine which column of $\mathbf{W}_{WF}$ corresponds to the best estimator. (—) shows the results obtained with the different column of $\mathbf{W}_{WF}$, (−−) with the error covariance matrix method, and ($\cdots$) with the average of the estimates.

the directional microphone, and in the second part we have the data of the omnidirectional microphone. In these two parts, the maximum improvement is reached when we take the column exactly in the middle, for the two kinds of microphones (omnidirectional and directional). Obviously, we have a better improvement with the directional part because the data of the directional microphone contains less noise than the data of the omnidirectional microphone when the noise source is located at 90° (remember that $\text{SNR}_{\text{weigthed, input}}$ input in formula (26) corresponds to the front omnidirectional microphone). The ECM method shows that column 10 gives the best result. Hence, from now on (and for the sake of simplicity) we will always take column 10 of the $\mathbf{W}_{WF}$.

3.2. Influence of the design parameters

Figures 7 and 8 show the effect of the size of the noise frame and the speech-and-noise frame on the mean and the standard deviation of the SNR_{SII} improvement. The values of the beamformer are taken as references, that is, 6.32 dB for the mean and 0.12 dB for the standard deviation. In the case of a large standard deviation, it is found that the signal has fluctuations which can correspond to unnatural sounds. The two-stage adaptive beamformer with this standard deviation (0.12 dB) is known to give satisfactory results in this respect.

Up to 5000 samples, if we increase the size of the speech and noise frame, we increase the SNR_{SII} improvement and the standard deviation drops. After 5000 samples, there is only a marginal improvement. For the noise frame, when we increase the length, we have the same effect as with the speech-and-noise frame. The SNR_{SII} improvement increases and the standard deviation decreases. So the longer the size of the noise frame, the larger the SNR_{SII} improvement. To obtain the same performance as the beamformer (with adaptation of the second filter on 10000 samples) for the standard deviation, we need to take a noise frame longer than
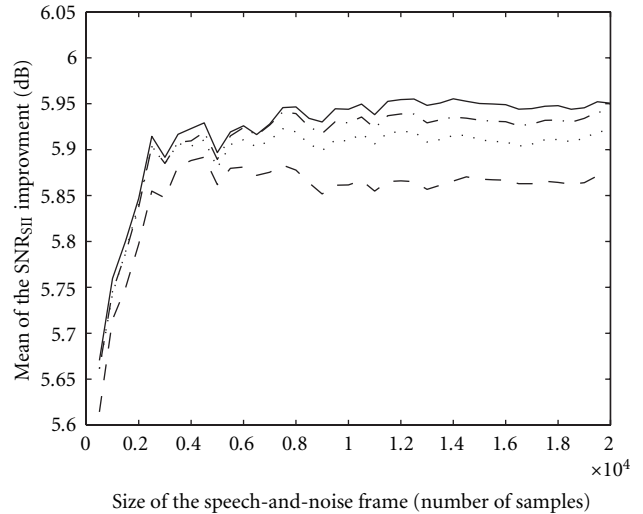

FIGURE 7: Influence of the length of the noise and the speech-and-noise frame on the mean of the SNR_{SII} improvement. The size of the filters for one channel is $N = 20$. (−−) shows the results obtained with a size of the noise frame = 5000 samples, ($\cdots$) with a size of the noise frame = 10000 samples, (−·) with a size of the noise frame = 15000 samples, and (—) with a size of the noise frame = 20000 samples.
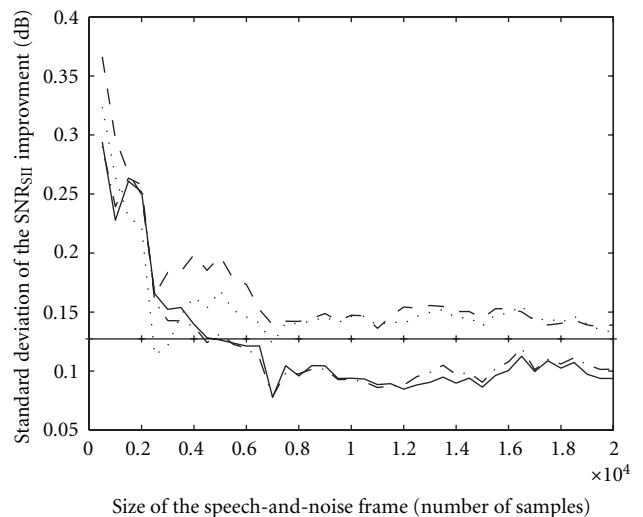

FIGURE 8: Influence of the length of the noise and the speech-and-noise frame on the standard deviation of the SNR_{SII} improvement. (−−) shows the results obtained with a size of the noise frame = 5000 samples, ($\cdots$) with a size of the noise frame = 10000 samples, (−·) with a size of the noise frame = 15000 samples and (—) with a size of the noise frame = 20000 samples. (++) shows the standard deviation obtained with the two-stage adaptive beamformer.

10000 samples (15000 samples). Furthermore, for the size of frame that we used, the SVD-based algorithm performs slightly worse (0.4 dB) than the beamformer.

Figure 9 shows the influence of the filter size N for each channel. For a length $N > 20$, the SNR_{SII} improvement stays
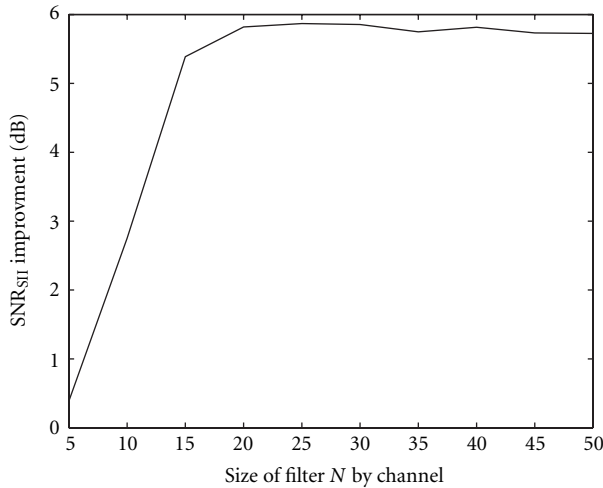
FIGURE 9: Influence of the filter size for each channel on the SNR_{SII} improvement. The length of the speech frame is 5000, the length of the noise frame is 15000.
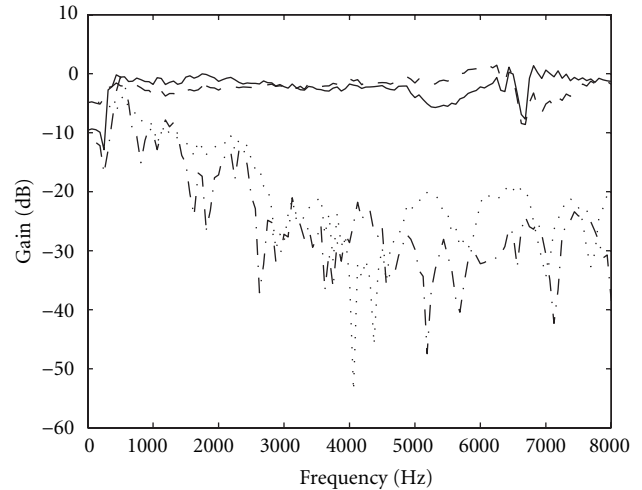

FIGURE 10: Magnitude of the transfer function estimate between the directional microphone and the output of the algorithms for speech only and for noise only. (——) shows the transfer function of the two-stage adaptive beamformer for speech and (−·) for noise. (−−) shows the transfer function of the SVD-based technique for speech and (···) for noise.
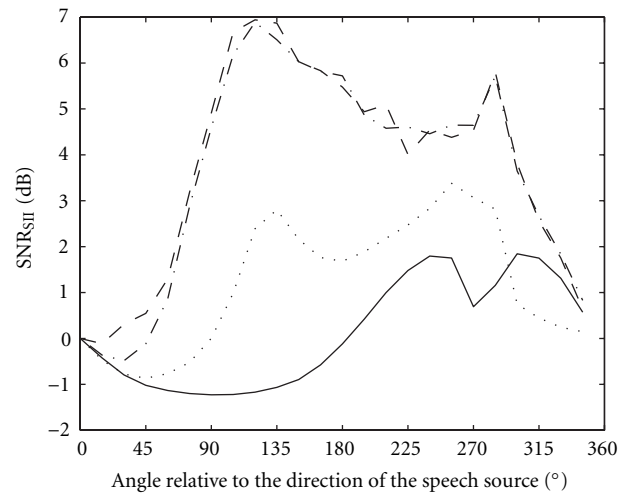

FIGURE 11: The curves show the SNR_{SII} at the omnidirectional microphone (——) as a function of angle relative to the direction of the speech source, at the directional microphone (···), at the output of the two-stage adaptive beamformer (−−) and at the output of the SVD-based technique (−·).

roughly the same, while for $N < 20$ the SNR_{SII} is seen to drop significantly.

In the case of a single microphone configuration, if the noise frame is equal to 15000 samples, the filter length is 20 and the length of the speech-and-noise frame is varied from 5000 samples to 20000 samples by steps of 5000 samples, an SNR_{SII} improvement < 0.1 dB is found. If the speech-and-noise frame is equal to 5000 samples, the filter length to 20 and the length of the noise frame is varied from 5000 samples to 20000 samples by steps of 5000 samples, again an SNR_{SII} improvement < 0.1 dB is found. Finally, when the speech-and-noise frame is equal to 5000 samples, the noise frame is equal to 15000 samples and the length of the filter is varied from 10 to 50 taps, no significant SNR_{SII} improvement is measured. The results of these different configurations show that the SVD-based algorithm for a single microphone configuration does not significantly improve the SNR_{SII}, which is a good predictor for the speech intelligibility.

Figure 10 shows the transfer function between the directional microphone and the output of the two algorithms. The SVD-based technique performs as well as the beamformer and the transfer function for speech is around 0 dB for the frequencies between 500 Hz and 6500 Hz. Thus, we can conclude that we do not have any distortion of the speech signal for each algorithm.

From these experiments, we can conclude that the SVD-based technique roughly performs as well as the two-stage adaptive beamformer. Indeed, with a filter size of 2×20 coefficients for the SVD-based technique and $10 + 30$ coefficients for the beamformer, the speech intelligibility improvement is only 0.4 dB higher for the beamformer. Furthermore, this improvement can be obtained with a similar adaptation time for the coefficients (10000 samples) for both algorithms. However, to obtain a similar standard deviation in the SNR_{SII} we prefer to take the adaptation time a bit longer for the SVD-based technique.

3.3. Behaviour with different noise scenarios

3.3.1 One noise source

Figure 11 shows the behaviour of the omnidirectional microphone, the software directional microphone, the output of the two-stage adaptive beamformer and the output of the SVD-based technique. Ideally, the omnidirectional microphone has the same sensitivity for all angles (around 0 dB in our case). However, the sensitivity is seen to be a function of the angle, mainly due to the effect of the dummy

TABLE 1: SNR$_{SII}$ improvement (in dB) for multiple noise sources.

			Beamformer	SVD-based	Dir. mic.
L_{90}	L_{180}	L_{270}	SNR$_{SII}$	SNR$_{SII}$	SNR$_{SII}$
ICRA			6.04	5.74	1.22
	ICRA		5.57	5.72	1.72
		ICRA	3.85	3.92	3.00
ICRA	BLU		3.87	3.85	2.06
ICRA		BLU	5.57	5.33	2.25
NVA	BLU		4.15	4.01	1.74
NVA		BLU	6.05	5.30	2.14
ICRA	ICRA		2.80	2.77	1.28
ICRA		ICRA	4.92	4.66	2.21
ICRA	BLU	NVA	3.35	3.45	2.25
NVA	BLU	BRUG	4.04	3.92	1.97
ICRA	ICRA	ICRA	2.84	2.87	1.88

head. Figure 11 shows that the beamformer and the SVD-based procedure have a very similar behaviour, and so we can say that the improvement is the same for the two algorithms. These algorithms always give a larger improvement than the omnidirectional and the directional microphone. Remarkably, the directional microphone can perform worse than the omnidirectional microphone, namely between angles +285° and +345°.

3.3.2 Multiple noise sources

For scenarios with multiple and different noise sources, the results are shown in Table 1. Evidently, with two and three noise sources, both the beamformer and the SVD-based algorithm perform worse than in the one noise source case. These two strategies give better results than the directional microphone and have roughly the same SNR$_{SII}$ improvement for the different situations. As in the one noise source configuration, the head diffraction contributes to the improvement. Indeed, the SNR$_{SII}$ improvement is larger when we have a noise source at 270° instead of 180°. Also, we can observe a better SNR$_{SII}$ improvement when the noise source does not contain an ICRA-signal spectrum.

3.3.3 Diffuse noise

In this section, we worked with high reverberations in the room and hence we cannot apply the SNR$_{SII}$ procedure properly [18]. For this reason we modified the SNR$_{weighted}$, as introduced by Greenberg et al. [28] and defined by (27), where now we only applied the weights for the importance of the band for the speech intelligibility. Thus, the SNR$_{weighted}$ becomes

$$SNR_{weigthed} = \sum_{i=1}^{k} I_i \cdot SNR_i \qquad (27)$$

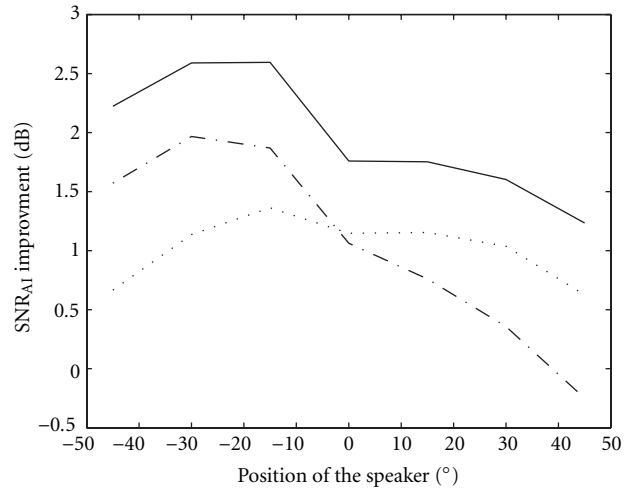and the SNR improvement, between the signal of the front omnidirectional microphone (input) and the signal of a

FIGURE 12: In a diffuse noise case, the SNR$_{AI}$ improvement between the omnidirectional microphone and the directional microphone ($\cdots$), the two-stage adaptive beamformer (–·) and the SVD-based technique (—).

noise reduction algorithm (output), is given by

$$SNR_{AI} = SNR_{weigthed,\,output} - SNR_{weigthed,\,input}. \qquad (28)$$

The SNR$_{AI}$ improvement of the SVD-based technique is always better than the directional microphone and the two-stage adaptive beamformer (see Figure 12). For angles above 0° the two-stage adaptive beamformer performs worse than the directional microphone and the omnidirectional microphone, in the case where the speaker is at +45°. The difference between the directional microphone and the SVD-based technique is small but still important for hearing-aid users, because in critical listening conditions (close to 50 per cent of speech understood by the listener) an improvement of 1 dB in SNR corresponds to an increase of speech understanding of about 15% in every day speech communication [37].

4. COMPUTATIONAL COMPLEXITY

The computational complexity of the two-stage adaptive beamformer equals $(2 \cdot N_1 + 7 \cdot N_2) \times fs$ operations (multiplications or additions) per second (ops/s). N_1 represents the size of the fixed filter, N_2 the size of the adaptive filter and fs the sampling frequency. In our case, $N_1 = 10$ taps and $N_2 = 30$ taps, thus the two-stage adaptive beamformer has a computational complexity of 3.7 Mops/s at sampling frequency 16.000 Hz.

Doclo and Moonen [23] suggests an implementation using a recursive and approximate SVD-based technique. The cost of calculation of this technique equals $(17.5 \cdot N^2/dg + 4 \cdot N^2/df) \times fs$ ops/s. N represents the size of the filters per channel, dg and df indicate the number of samples between respectively two GSVD and filter updates and they trade off convergence speed and cost of calculation. An optimal size of $N = 20$ taps has been found (see Section 2.6.2) and if $dg = df = 1$ the cost of calculation equals 550 Mops/s.

Also, an interesting implementation of the SVD-based technique has been proposed by Spriet [38], which is based on the Doclo's implementation. The difference between both implementations is that Doclo presents a full band implementation and Spriet a subband implementation where an important reduction of the computational complexity can be obtained.

5. CONCLUSION

In this paper, we have assessed the performance of a SVD-based multimicrophone enhancement procedure in the context of two microphone hearing aids. Through the different evaluations, the SVD-based noise reduction technique of [17] is seen to achieve roughly the same improvement of the speech intelligibility as the two-stage adaptive beamformer of [15] in the case of located noise sources. However, in the diffuse noise case, the SVD-based technique performs always better than the two-stage adaptive beamformer and the directional microphone. The SVD-based algorithm, furthermore, works without initializations nor assumptions about a look direction, unlike the two-stage adaptive beamformer. Indeed, with the beamformer, we have to give a look direction to the algorithm, which is not a simple task, and then assume that the speaker is always in front of the listener. This beamformer initialization is a function of the microphone characteristics and displacement and should be adapted to each and every hearing aid.

An important result is that the SVD-based technique does not distort speech during noise suppression. Indeed, the transfer function between the unprocessed and the processed speech signals is flat (0 dB) and we get roughly the same behaviour as for the beamformer.

The main disadvantage of the SVD-based technique is the cost of calculation but even this can be reduced to an acceptable level [23]. In the future, the technique will be implemented in real time on a PC-platform, and perceptual test will be done with normal hearing and hearing impaired listeners.

ACKNOWLEDGMENTS

REFERENCES

[1] R. Plomp, "Noise, amplification, and compression: considerations of three main issues in hearing aid design," *Ear and Hear*, vol. 15, no. 1, pp. 2–12, 1994.

[2] B. Hawkins and W. Yacullo, "Signal-to-noise ratio advantage of binaural hearing aids and directional microphones under different levels of reverberation," *Journal of Speech and Hearing Disorders*, vol. 49, no. 3, pp. 278–286, 1984.

[3] A. R. Leeuw and W. A. Dreschler, "Advantages of directional hearing aid microphones related to room acoustics," *Audiology*, vol. 30, no. 6, pp. 330–344, 1991.

[4] H. Levitt, A. Neuman, R. Mills, and T. Schwander, "A digital master hearing aid," *J. Rehab. Res. Dev.*, vol. 23, no. 1, pp. 79–87, 1986.

[5] W. Soede, F. A. Bilsen, and A. J. Berkhout, "Assessment of a directional microphone array for hearing-impaired listeners," *Journal of the Acoustical Society of America*, vol. 94, no. 2, pp. 799–808, 1993.

[6] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 35, no. 10, pp. 1365–1376, 1987.

[7] J. M. Kates, "Superdirective arrays for hearing aids," *Journal of the Acoustical Society of America*, vol. 94, no. 4, pp. 1930–1933, 1993.

[8] J. M. Kates and M. R. Weiss, "A comparison of hearing-aid array-processing techniques," *Journal of the Acoustical Society of America*, vol. 99, no. 5, pp. 3138–3148, 1996.

[9] G. H. Saunders and J. M. Kates, "Speech intelligibility enhancement using hearing-aid array processing," *Journal of the Acoustical Society of America*, vol. 102, no. 3, pp. 1827–1837, 1997.

[10] B. Widrow and S. Stearns, *Adaptive Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, USA, 1985.

[11] P. M. Peterson, "Adaptive array processing for multiple microphone hearing aids," Tech. Rep. 541, Research Laboratory of Electronics, MIT, Cambridge, MA, 1989.

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

[13] J. B. Maj, J. Wouters, and M. Moonen, "Field trial of a two-stage adaptive beamformer for a dual microphone hearing aid," in preparation.

[14] D. Van Compernolle, "Switching adaptive filters for enhancing noisy and reverberant speech from microphone array recordings," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 833–836, Albuquerque, Greece, September 1990.

[15] J. Vanden Berghe and J. Wouters, "An adaptive noise canceller for hearing aids using two nearby microphones," *Journal of the Acoustical Society of America*, vol. 103, no. 6, pp. 3621–3626, 1998.

[16] J. Vanden Berghe, J. Wouters, and J. B. Maj, "Evaluation of real-time adaptive noise suppression for a dual microphone hearing aid," to appear in *International Journal of Audiology*.

[17] S. Doclo and M. Moonen, "SVD-based optimal filtering with applications to noise reduction in speech signals," in *Proc. 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 143–146, New Paltz, USA, October 1999.

[18] American National Standards Institute, *American National*

Standard Methods for Calculation of the Speech Intelligibility Index, ANSI S3.5. 1997.

[19] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 28, no. 2, pp. 137–145, 1980.

[20] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proceedings of the IEEE*, vol. 67, no. 12, pp. 1586–1604, 1979.

[21] M. W. Callahan, *Acoustic signal processing based on the short-time spectrum*, Dep Compt Sci, University of Utah, Salt Lake City, 1976.

[22] W. K. Pratt, *Digital Image Processing*, John Wiley and Sons, New York, NY, USA, 1978.

[23] S. Doclo and M. Moonen, *Microphone Arrays: Signal Processing Techniques and Aplications*, Chapter 6, pp. 111–132, Springer-Verlag, 2001.

[24] P. C. Hansen and S. H. Jensen, "FIR filter representations of reduced-rank noise reduction," *IEEE Trans. Signal Processing*, vol. 46, no. 6, pp. 1737–1741, 1998.

[25] P. S. K. Hansen, *Signal subspace methods for speech enhancement*, Ph.D. thesis, IMM, Technical University of Denmark, DK-2800 Lyngby, Denmark, September 1997.

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

[27] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1996.

[28] J. E. Greenberg, P. M. Peterson, and P. M. Zurek, "Intelligibilty-weighted measures of speech-interference ratio and speech system performance," *Journal of the Acoustical Society of America*, vol. 94, no. 5, pp. 3009–3010, 1993.

[29] American National Standards Institute, *American national standard methods for the calculation of the articulation index*, ANSI-S3.5. 1969.

[30] M. W. Hoffman, T. D. Trine, K. M. Buckley, and D. J. Van Tasell, "Robust adaptive microphone array processing for hearing aids: realistic speech enhancement," *Journal of the Acoustical Society of America*, vol. 96, no. 2, pp. 759–770, 1994.

[31] S. C. Thompson, "Dual microphones or directional-plus-omni: which is the best?," *The Hearing Review*, vol. 3, pp. 31–35, 1999.

[32] U. M. A. Peutz, "Articulation loss of consonants as a criterion for speech transmission in a room," *Journal of the Audio Engineering Society*, vol. 19, no. 11, pp. 915–919, 1971.

[33] D. Byrne, H. Dillon, and K. Tran, "An international comparison of long-term average speech spectra," *Journal of the Acoustical Society of America*, vol. 96, no. 4, pp. 2108–2120, 1994.

[34] Noise Signals ICRA, "International collegium of reabilitative audiology," ver 0.3, 1997, Compact Disk.

[35] J. Wouters, W. Damman, and A. Bosman, "Vlaamse opname van woordenlijsten voor spraakaudiometrie," *Logopedie*, vol. 7, no. 6, pp. 28–33, 1994.

[36] I. Veit and H. Sander, "Production of spatially limited "diffuse" sound field in an anechoic room," *Journal of the Audio Engineering Society*, vol. 35, no. 3, pp. 138–142, 1987.

[37] R. Plomp and A. M. Mimpen, "Improving the reliability of testing the speech reception threshold for sentences," *Audiology*, vol. 18, pp. 43–52, 1979.

[38] A. Spriet, M. Moonen, and J. Woutersq, "A multichannel subband GSVD based approach for speech enhancement in hearing aids," in *Proc. IEEE International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Darmstadt, Germany, September 2001.

Jean-Baptiste Maj was born in Dijon, France, on the 8th of October 1973. In 1993, he graduated from the École Spéciale de Radioélectricité in Nancy (France). Between 1993 and 1998 he studied Applied Sciences at the University of Nancy (Université Point Carré), where he received his DEA degree in Electrical Engineering. In April 1999, he started research (Ph.D. student) at the Electrical Engineering Department and the Laboratory for Experimental ORL of the Katholieke Universiteit Leuven, Leuven, Belgium.

Marc Moonen received the electrical engineering degree and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1986 and 1990, respectively. He has been a Research Associate with the Belgian National Fund for Scientific Research, since 1994, and an Associate Professor, since 2000, both at the Electrical Engineering Department of the Katholieke Universiteit Leuven. His research activities are in digital signal processing, digital communications and audio signal processing. He received the 1994 K.U. Leuven Research Council Award, the 1997 Alcatel Bell (Belgium) Award (with Piet Vandaele), and was a 1997 "Laureate of the Belgium Royal Academy of Science." He is Chairman of the IEEE Benelux Signal Processing Chapter, a EURASIP AdCom Member (European Association for Signal, Speech and Image Processing), and a member of the editorial board of "Integration, the VLSI Journal" and "Applied Signal Processing."

Jan Wouters was born in Leuven, Belgium, in 1960. He received the physics degree and the Ph.D. degree in sciences/physics from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1982 and 1989, respectively. From 1989 till 1992 he has been a Research Fellow with the Belgian NFWO (National Fund for Scientific Research) at the Institute of Nuclear Physics (UCL Louvain-la-Neuve and KULeuven) and at NASA Goddard Space Flight Center (USA). From 1993 he is a Professor at the Neurosciences Department of the Katholieke Universiteit Leuven. His research activities are about audiology and the auditory system, signal processing for cochlear implants and hearing aids. Dr. Wouters received in 1989 an Award of the Flemish Ministry, in 1992 a Fullbright Award and a NATO Research Fellowship, the 1996 Flemish VVL Speech therapy - Audiology Award. He is member of the International Collegium for Rehabilitative Audiology, a Board Member of the NAG (Dutch Acoustical Society) and President Scientific Committee of the Belgian Society for Audiology, author of 75 articles in international peer-review journals, reviewer for several international journals.