# Multimicrophone Noise Reduction Using Recursive GSVD-Based Optimal Filtering With ANC Postprocessing Stage

Simon Doclo, *Member, IEEE,* and Marc Moonen, *Member, IEEE*

***Abstract*—Recently, a generalized singular value decomposition (GSVD)-based optimal filtering technique has been proposed for enhancing multimicrophone speech signals degraded by additive colored noise. The GSVD-based optimal filtering technique has a better noise reduction performance than standard beamforming techniques provided that the used filter length is large enough. In this paper, it is shown that the same noise reduction performance can be obtained with shorter filter lengths at a lower computational complexity by incorporating the GSVD-based optimal filtering technique in a generalized sidelobe canceller type structure, i.e., by adding an adaptive noise cancellation (ANC) postprocessing stage. Even when using short filter lengths, the total computational complexity is essentially determined by the calculation of the GSVD of a speech and a noise data matrix. It is shown that the complexity can be significantly reduced by using recursive GSVD-updating algorithms and by using subsampling.**

**Simulations have been performed for various acoustic scenarios (different and multiple noise sources and different reverberation conditions), where both the improvement in signal-to-noise ratio and speech distortion have been analyzed. These simulations show that the GSVD-based optimal filtering technique with an ANC postprocessing stage has a better noise reduction performance than standard fixed and adaptive beamforming techniques while introducing an acceptable amount of speech distortion.**

***Index Terms*—Generalized sidelobe canceller, generalized singular value decomposition (GSVD), multichannel Wiener filter, optimal filtering, recursive algorithms, speech enhancement.**

## I. Introduction

IN MANY speech communication applications, such as hands-free mobile telephony, hearing aids, and voice-controlled systems, the recorded speech signals are often corrupted by a considerable amount of acoustic background noise. Since the desired speech signal and the undesired noise signal usually occupy overlapping frequency bands, single-microphone speech enhancement techniques (e.g., spectral subtraction [1], Kalman filtering [2], signal subspace-based techniques [3], [4]) generally have problems to reduce the background noise without introducing noticeable artifacts (e.g., musical noise) or speech distortion. However, when the speech and noise sources are physically located at different positions, this spatial diversity can be exploited by using a microphone array (see Fig. 1), such that both the spectral and the spatial characteristics of the signal sources can be used.

Well-known multimicrophone speech enhancement techniques are fixed and adaptive beamforming techniques [5]. A fixed delay-and-sum (DS) beamformer spatially aligns the microphone signals to the direction of the speech source. In a minimum-variance distortionless response (MVDR) beamformer, the energy of the output signal is minimized under the constraint that signals arriving from the look direction, i.e., the direction of the speech source, are processed without distortion. A well-known adaptive implementation of this beamformer is the generalized sidelobe canceller (GSC) [6], which consists of a fixed beamformer, creating a so-called speech reference signal; a blocking matrix, creating so-called noise reference signals; and a multichannel adaptive filter [7]–[8], eliminating the noise components in the speech reference signal which are correlated with the noise reference signals. However, because of room reverberation, microphone mismatch, and look direction error, the speech signal leaks into the noise references, such that signal cancellation occurs in the standard GSC. In order to limit signal cancellation, different variants of the standard GSC implementation exist, e.g., using a speech-controlled adaptation algorithm [9]–[12], a spatial filter designed blocking matrix [11], [13], norm-constrained [14] and coefficient-constrained adaptive filters [12], or incorporating a transfer function model [15].

Recently, a generalized singular value decomposition (GSVD)-based optimal filtering technique has been proposed for enhancing multimicrophone speech signals degraded by additive colored noise [16]–[18]. This optimal filtering technique makes a minimum mean square error (MMSE) estimate of the speech component in one of the microphone signals. Hence, the reverberation present in the microphone signals will not be suppressed and inevitably some (linear) speech distortion will be introduced. The GSVD-based optimal filtering technique is in fact a multimicrophone extension of the single-microphone signal subspace-based techniques for speech enhancement [3],

Manuscript received May 8, 2002; revised November 12, 2003. This work was performed at the ESAT Laboratory of the Katholieke Universiteit Leuven and was supported in part by the F.W.O. Research Project G.0233.01 (Signal processing and Automatic Patient fitting for Advanced Auditory Prostheses), the I.W.T. Project MUSETTE I-II (Multimicrophone Signal Enhancement Techniques for hands-free Telephony and Voice-controlled Systems), the Concerted Research Action GOA-MEFISTO-666 (Mathematical Engineering Techniques for Information and Communication Systems) of the Flemish Government, the Interuniversity Attraction Pole IUAP P5-22 (2002–2007), Dynamical Systems and Control: Computation, Identification, and Modeling, initiated by the Belgian State, Prime Minister's Office—Federal Office for Scientific, Technical, and Cultural Affairs, and in part by Philips ITCL. The associate editor coordinating the review of this paper and approving it for publication was Dr. Walter Kellermann.

The authors are with the Department of Electrical Engineering (ESAT-SCD), Katholieke Universiteit Leuven, B-3001 Leuven, Belgium (e-mail: simon.doclo@esat.kuleuven.ac.be; marc.moonen@esat.kuleuven.ac.be).

Digital Object Identifier 10.1109/TSA.2004.834462

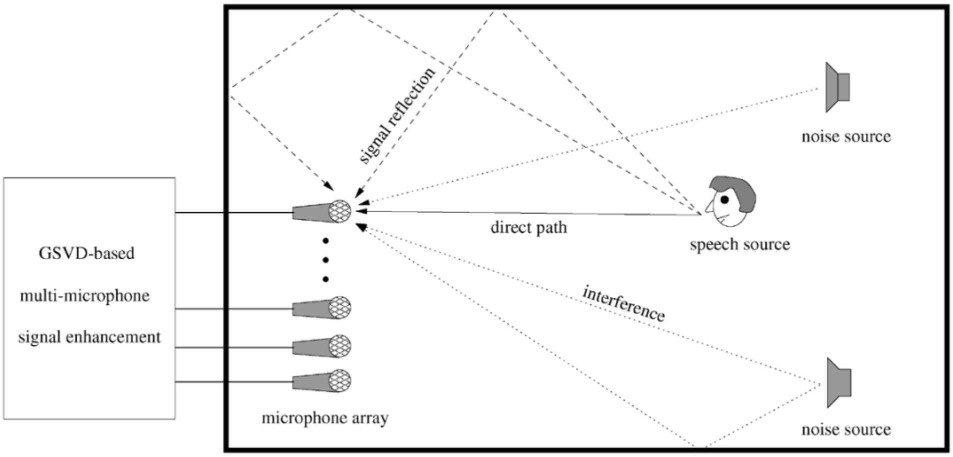


Fig. 1. Typical speech communication environment with desired speech source and undesired noise sources recorded with a microphone array.

[4], now combining the spatio-temporal information of the speech and noise sources. In [16], it has been shown that the optimal filter can be written as a function of the generalized singular vectors and generalized singular values of a so-called speech and noise data matrix, where the specific function used provides a means to trade off noise reduction versus speech distortion. It has also been shown that the GSVD-based optimal filtering technique has a better noise reduction performance than standard beamforming techniques (DS beamformer, GSC) for all reverberation times, if the used filter length is large enough. In addition, since the GSVD-based optimal filtering technique does not make any *a priori* assumptions about the location of the speaker, the microphone characteristics, and the room reverberation, it is more robust to deviations from the nominal situation [16], [19]. However, the computational complexity of this technique is quite high, since it requires calculating the GSVD of two matrices.

In this paper, several techniques are discussed for *reducing the total computational complexity* of the GSVD-based optimal filtering technique described in [16]. First, it is shown that the same noise reduction performance can be obtained with shorter filter lengths by incorporating the GSVD-based optimal filtering technique in a GSC-type structure, i.e., adding an ANC postprocessing stage. Second, several techniques are discussed for efficiently calculating the GSVD of the speech and the noise data matrix, making this multimicrophone noise reduction technique amenable to real-time implementation.

The paper is organized as follows. In Section II, the GSVD-based optimal filtering technique is briefly reviewed and it is shown that the optimal filter can be written as a function of the generalized singular vectors and generalized singular values of a speech and a noise data matrix. Section III describes how the GSVD-based optimal filtering technique can be incorporated in a GSC-type structure by creating speech and noise reference signals and by using these signals in an adaptive noise cancellation (ANC) postprocessing stage. The output of the GSVD-based optimal filtering technique is used as speech reference signal, while different possibilities exist for creating a noise reference. Since the total computational complexity is essentially determined by the calculation of the GSVD, Section IV describes several techniques for reducing the complexity by using recursive GSVD-updating algorithms (Section IV-B) and subsampling (Section IV-C). In Section IV-D the computational complexity is summarized for realistic parameter values, showing that the computational complexity can be significantly reduced such that the proposed signal enhancement technique indeed becomes suitable for real-time implementation. Section V describes several simulation results. In Sections V-A, V-B, and V-C, the used simulation environment and implementation issues of the different signal enhancement techniques are discussed. In Section V-D it is shown that the batch and the recursive version of the GSVD-based optimal filtering technique nearly have the same performance. Section V-E describes the effect of several parameters in the recursive GSVD-updating algorithms. In Section V-F, the effect of the ANC postprocessing stage on the noise reduction performance and the speech distortion is investigated for different filter lengths and number of noise references. It is shown that the decrease in noise reduction performance due to short filter lengths for the GSVD-based optimal filter can be fully compensated by adding the ANC postprocessing stage, at a lower total computational complexity and causing a small increase in speech distortion. In Sections V-G and V-H, simulations are performed for various acoustic scenarios, showing that the GSVD-based optimal filtering technique with an ANC postprocessing stage outperforms standard fixed and adaptive beamforming techniques.

## II. GSVD-Based Optimal Filtering

### A. Problem Formulation and Notation

Consider $N$ microphones, where each microphone signal $y_n[k], n = 0 \dots N-1$, at time $k$, consists of a filtered version of the clean speech signal $s[k]$ and additive noise

$$y_n[k] = h_n[k] \otimes s[k] + v_n[k] = x_n[k] + v_n[k] \qquad (1)$$

where $x_n[k]$ and $v_n[k]$ are, respectively, the speech and the noise component received at the $n$th microphone, $h_n[k]$ is the acoustic room impulse response between the speech source and the $n$th microphone and $\otimes$ denotes convolution. The additive noise can be colored and is assumed to be uncorrelated with the speech signal.

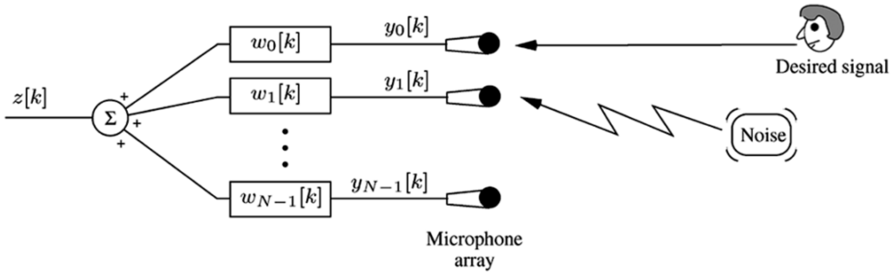


Fig. 2. Multimicrophone filtering for speech enhancement.

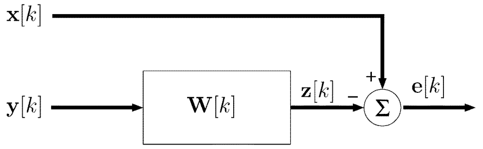


Fig. 3. Optimal filtering problem with desired response vector $\mathbf{x}[k]$.

The goal of multimicrophone speech enhancement is to compute filters $w_n[k], n = 0 \ldots N-1$ (see Fig. 2), such that the speech signal $s[k]$ (GSC) or one of the speech components $x_n[k]$ (GSVD-based optimal filtering technique) is recovered. Let the FIR filters $\mathbf{w}_n[k]$ have length $L$, and consider the $L$-dimensional data vectors $\mathbf{y}_n[k]$, the $M$-dimensional stacked filter $\mathbf{w}[k]$ (with $M = LN$) and the $M$-dimensional stacked data vector $\mathbf{y}[k]$, defined as

$$\mathbf{y}_n[k] = [y_n[k] \quad y_n[k-1] \quad \ldots \quad y_n[k-L+1]]^T \tag{2}$$

$$\mathbf{w}[k] = \left[\mathbf{w}_0^T[k] \quad \mathbf{w}_1^T[k] \quad \ldots \quad \mathbf{w}_{N-1}^T[k]\right]^T \tag{3}$$

$$\mathbf{y}[k] = \left[\mathbf{y}_0^T[k] \quad \mathbf{y}_1^T[k] \quad \ldots \quad \mathbf{y}_{N-1}^T[k]\right]^T \tag{4}$$

with $^T$ denoting transpose, such that the output signal $z[k]$ can be written as

$$z[k] = \sum_{n=0}^{N-1} \mathbf{w}_n^T[k]\,\mathbf{y}_n[k] = \mathbf{w}^T[k]\,\mathbf{y}[k]. \tag{5}$$

In the next section, a method will be described for computing the stacked filter $\mathbf{w}[k]$ such that $z[k]$ is an optimal estimate for one of the speech components $x_n[k]$.

### *B. Unconstrained Optimal Filtering*

Consider the filtering problem in Fig. 3: $\mathbf{y}[k]$ is the $M$-dimensional filter input vector, and $\mathbf{z}[k] = \mathbf{W}^T[k]\,\mathbf{y}[k]$ is the filter output vector, where $\mathbf{W}[k]$ is an $M \times M$-dimensional filter matrix. The $M$-dimensional vector $\mathbf{x}[k]$ is the desired response vector and $\mathbf{e}[k] = \mathbf{x}[k] - \mathbf{z}[k]$ is the estimation error vector. The mean square error (MSE) cost function leads to the well-known multidimensional Wiener filter [20]

$$\bar{\mathbf{W}}_{\mathrm{WF}}[k] = \bar{\mathbf{R}}_{yy}^{-1}[k]\,\bar{\mathbf{R}}_{yx}[k] \tag{6}$$

where $\bar{\mathbf{R}}_{yy}[k] = \mathcal{E}\{\mathbf{y}[k]\,\mathbf{y}^T[k]\}$ is the $M \times M$-dimensional correlation matrix of the input signal, $\bar{\mathbf{R}}_{yx}[k] = \mathcal{E}\{\mathbf{y}[k]\,\mathbf{x}^T[k]\}$ is the $M \times M$-dimensional cross-correlation matrix of the input signal and the desired signal, and $\mathcal{E}\{\cdot\}$ denotes the expectation operator.

When considering multimicrophone noisy speech signals, the input vector $\mathbf{y}[k]$ consists of the speech component and the additive noise component, $\mathbf{y}[k] = \mathbf{x}[k] + \mathbf{v}[k]$, with $\mathbf{y}[k]$ defined in (4) and $\mathbf{x}[k]$ and $\mathbf{v}[k]$ similarly defined. Since the desired signal $\mathbf{x}[k]$ is an unobservable signal, this poses a particular problem which may be solved based on the on/off characteristics of the speech signal. If we use a robust voice activity detection (VAD) algorithm [21], [22], noise-only observations can be made during speech pauses (time $k'$), where $\mathbf{y}[k'] = \mathbf{v}[k']$, which allows estimation of the spatio-temporal correlation properties of the noise signal. The output of the VAD-algorithm at time $k$ is represented by $\beta_k$, where $\beta_k = 1$ represents a speech-and-noise observation and $\beta_k = 0$ represents a noise-only observation.

We now make two *assumptions*: We assume that the second-order statistics of the noise signal are sufficiently stationary such that the noise correlation matrix $\bar{\mathbf{R}}_{vv}[k]$, which can be estimated during noise-only periods, can also be used during subsequent speech-and-noise periods, i.e.,

$$\bar{\mathbf{R}}_{vv}[k] = \mathcal{E}\{\mathbf{v}[k]\,\mathbf{v}^T[k]\} = \mathcal{E}\{\mathbf{v}[k']\,\mathbf{v}^T[k']\} = \bar{\mathbf{R}}_{vv}[k'] \tag{7}$$

and we also assume that the speech and noise signals are statistically independent, implying that

$$\bar{\mathbf{R}}_{xv}[k] = \mathcal{E}\{\mathbf{x}[k]\,\mathbf{v}^T[k]\} = \mathbf{0}. \tag{8}$$

From the second assumption it is easily verified that $\bar{\mathbf{R}}_{yy}[k] = \bar{\mathbf{R}}_{xx}[k] + \bar{\mathbf{R}}_{vv}[k]$ and $\bar{\mathbf{R}}_{yx}[k] = \bar{\mathbf{R}}_{xx}[k]$, such that the optimal filter matrix in (6) can be written as

$$\bar{\mathbf{W}}_{\mathrm{WF}}[k] = \bar{\mathbf{R}}_{yy}^{-1}[k]\,\left(\bar{\mathbf{R}}_{yy}[k] - \bar{\mathbf{R}}_{vv}[k]\right) \tag{9}$$

with $\bar{\mathbf{R}}_{yy}[k]$ estimated during speech-and-noise periods and $\bar{\mathbf{R}}_{vv}[k]$ estimated during noise-only periods.

For calculating the multidimensional Wiener filter $\bar{\mathbf{W}}_{\mathrm{WF}}[k]$, the expression in (9) can generally be used directly without any numerical problem. However, in [3] and [16], it has been shown that by using the joint diagonalization of the symmetric correlation matrices $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$, the low-rank model of the clean speech signal $s[k]$ can easily be taken into account and one can easily provide a tradeoff between noise reduction and speech distortion. The joint diagonalization of $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$ is defined as

$$\begin{cases} \bar{\mathbf{R}}_{yy}[k] = \bar{\mathbf{Q}}[k]\mathrm{diag}\left\{\bar{\sigma}_i^2[k]\right\}\bar{\mathbf{Q}}^T[k] \\ \bar{\mathbf{R}}_{vv}[k] = \bar{\mathbf{Q}}[k]\mathrm{diag}\left\{\bar{\eta}_i^2[k]\right\}\bar{\mathbf{Q}}^T[k] \end{cases} \tag{10}$$

where $\bar{\mathbf{Q}}[k]$ is an invertible, but not necessarily orthogonal, matrix [23]. Substituting (10) into (9) gives an expression for the optimal filter matrix

$$\bar{\mathbf{W}}_{\mathrm{WF}}[k] = \bar{\mathbf{Q}}^{-T}[k]\,\mathrm{diag}\left\{1 - \frac{\bar{\eta}_i^2[k]}{\bar{\sigma}_i^2[k]}\right\}\bar{\mathbf{Q}}^T[k]. \tag{11}$$

The enhanced speech vector $\hat{\mathbf{x}}[k] = \mathbf{z}[k]$ is obtained as $\hat{\mathbf{x}}[k] = \bar{\mathbf{W}}_{\mathrm{WF}}^T[k]\,\mathbf{y}[k]$, such that the $M$-dimensional vector $\hat{\mathbf{x}}[k]$ contains an estimate for all speech samples $x_n[k-l], n = 0 \ldots N-1, l = 0 \ldots L-1$, i.e., for all $L$ delayed versions of the speech components in all $N$ microphone signals. The $i$th element of $\hat{\mathbf{x}}[k]$, which is obtained by filtering the microphone signals with the $i$th column $\bar{\mathbf{w}}_{WF,i}[k]$ of $\bar{\mathbf{W}}_{\mathrm{WF}}[k]$, represents an optimal estimate for the speech component in the $m$th microphone signal with delay $\Delta$, i.e.,

$$\hat{x}_m[k-\Delta] = \mathbf{y}^T[k]\,\bar{\mathbf{w}}_{WF,i}[k] \tag{12}$$

with

$$m = \mathrm{div}(i-1, L), \quad \Delta = \mathrm{mod}(i-1, L) \tag{13}$$

where $\mathrm{div}(i-1,L)$ denotes the integer part of $(i-1)/L$ and $\mathrm{mod}(i-1,L)$ denotes the remainder of the division. The question now arises which of the $M$ columns of $\bar{\mathbf{W}}_{\mathrm{WF}}[k]$ yields the lowest MSE. In [16] it has been shown that the smallest diagonal element of the error covariance matrix $\bar{\mathbf{R}}_{ee}[k] = \mathcal{E}\{\mathbf{e}[k]\,\mathbf{e}[k]^T\}$, with $\mathbf{e}[k] = \mathbf{x}[k] - \hat{\mathbf{x}}[k]$, corresponds to the "best" estimator. However, computing $\bar{\mathbf{R}}_{ee}[k]$ at each time step and choosing the column corresponding to its smallest diagonal element is a computationally very demanding procedure. Simulations have indicated that taking a fixed value $i = L/2$, i.e., using the optimal estimate of the delayed speech component in the first microphone signal $x_0[k-(L/2)+1]$, does not have a significant effect on the noise reduction performance and the speech intelligibility [17].

In [16] it has already been indicated that, when using the GSVD-based optimal filter for noise reduction, some speech distortion cannot be avoided, since the estimation error $\mathbf{e}[k]$ is the sum of a term $\mathbf{e}_y[k]$ representing speech distortion and a term $\mathbf{e}_v[k]$ representing the residual noise, i.e.,

$$\begin{aligned}\mathbf{e}[k] &= \mathbf{x}[k] - \bar{\mathbf{W}}_{\mathrm{WF}}^T\,\mathbf{y}[k]\\ &= \underbrace{\left(\mathbf{I}_M - \bar{\mathbf{W}}_{\mathrm{WF}}^T\right)\,\mathbf{x}[k]}_{\mathbf{e}_y[k]} - \underbrace{\bar{\mathbf{W}}_{\mathrm{WF}}^T\,\mathbf{v}[k]}_{\mathbf{e}_v[k]}\end{aligned} \tag{14}$$

where $\mathbf{I}_M$ is the $M \times M$-dimensional identity matrix. In [16] it has also been shown that it is possible to provide a tradeoff between noise reduction and speech distortion.

### C. Practical Computation Using GSVD

In practice, the matrix $\bar{\mathbf{Q}}[k]$ and the diagonal elements $\bar{\sigma}_i^2[k]$ and $\bar{\eta}_i^2[k]$ can be estimated by a GSVD [24], [25] of a $p_k \times M$-dimensional speech data matrix $\mathbf{Y}[k]$, containing $p$ speech data vectors, and a $q_k \times M$-dimensional noise data matrix $\mathbf{V}[k]$, containing $q$ noise data vectors (with $p$ and $q$ typically much larger than $M$), i.e.,

$$\mathbf{Y}[k] = \begin{bmatrix}\beta_{k-p_k+1}\mathbf{y}^T[k-p_k+1]\\ \vdots\\ \beta_{k-1}\mathbf{y}^T[k-1]\\ \beta_k\mathbf{y}^T[k]\end{bmatrix} \tag{15}$$

$$\mathbf{V}[k] = \begin{bmatrix}(1-\beta_{k-q_k+1})\mathbf{y}^T[k-q_k+1]\\ \vdots\\ (1-\beta_{k-1})\mathbf{y}^T[k-1]\\ (1-\beta_k)\mathbf{y}^T[k]\end{bmatrix} = \begin{bmatrix}(1-\beta_{k-q_k+1})\mathbf{v}^T[k-q_k+1]\\ \vdots\\ (1-\beta_{k-1})\mathbf{v}^T[k-1]\\ (1-\beta_k)\mathbf{v}^T[k]\end{bmatrix} \tag{16}$$

where $p_k$ and $q_k$ are chosen such that

$$\sum_{l=k-p_k+1}^{k}\beta_l = p \qquad \sum_{l=k-q_k+1}^{k}(1-\beta_l) = q. \tag{17}$$

Recall that $\beta_k = 1$ for speech-and-noise observations $(\mathbf{y}[k] = \mathbf{x}[k] + \mathbf{v}[k])$, whereas $\beta_k = 0$ for noise-only observations $(\mathbf{y}[k] = \mathbf{v}[k])$. The correlation matrices $\bar{\mathbf{R}}_{yy}[k]$ and $\bar{\mathbf{R}}_{vv}[k]$ can now be approximated by the empirical correlation matrices $\mathbf{R}_{yy}[k] = \mathbf{Y}^T[k]\,\mathbf{Y}[k]/p$ and $\mathbf{R}_{vv}[k] = \mathbf{V}^T[k]\,\mathbf{V}[k]/q$.

The GSVD of the data matrices $\mathbf{Y}[k]$ and $\mathbf{V}[k]$ is defined as

$$\begin{cases}\mathbf{Y}[k] = \mathbf{U}_Y[k]\cdot\mathbf{\Sigma}_Y[k]\cdot\mathbf{Q}^T[k]\\ \mathbf{V}[k] = \mathbf{U}_V[k]\cdot\mathbf{\Sigma}_V[k]\cdot\mathbf{Q}^T[k]\end{cases} \tag{18}$$

with $\mathbf{\Sigma}_Y[k] = \mathrm{diag}\{\sigma_i[k]\}, \mathbf{\Sigma}_V[k] = \mathrm{diag}\{\eta_i[k]\}, \mathbf{U}_Y[k]$ and $\mathbf{U}_V[k]$ orthogonal matrices, $\mathbf{Q}[k]$ an invertible but not necessarily orthogonal matrix containing the generalized singular vectors and $\sigma_i[k]/\eta_i[k]$ the generalized singular values. Substituting these formulas into (9) gives an empirical estimate $\mathbf{W}_{\mathrm{WF}}[k]$ for the optimal filter matrix $\bar{\mathbf{W}}_{\mathrm{WF}}[k]$ at time $k$

$$\mathbf{W}_{\mathrm{WF}}[k] = \mathbf{Q}^{-T}[k]\mathrm{diag}\left\{1 - \frac{p}{q}\frac{\eta_i^2[k]}{\sigma_i^2[k]}\right\}\mathbf{Q}^T[k] \tag{19}$$

showing that the optimal filter matrix estimate $\mathbf{W}_{\mathrm{WF}}[k]$ can be written as a function of the generalized singular vectors and generalized singular values of the speech and the noise data matrices.

Since in practice the generalized singular values are estimated from the empirical correlation matrices, it may occur that some diagonal elements in (19) become negative. In [3] and [16] it has already been noted that negative values will always be obtained since an unbiased nonperfect estimator is used. Therefore, these negative values, which are in fact zero estimates, will be set to zero.

### D. Batch and Recursive Algorithm

In the *batch version* of the algorithm, the speech and the noise data matrices $\mathbf{Y}[k]$ and $\mathbf{V}[k]$ are constructed using all available speech and noise data vectors in the considered signal frame.

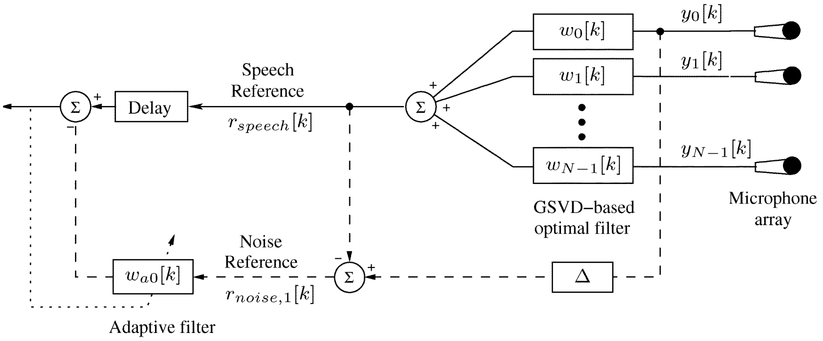


Fig. 4. GSVD-based optimal filtering technique with an ANC postprocessing stage.

The optimal filter matrix $\mathbf{W}_{\mathrm{WF}}[k]$ (which is then actually independent of $k$) is computed using the GSVD of $\mathbf{Y}[k]$ and $\mathbf{V}[k]$ in (19), and the total enhanced signal is obtained by filtering the microphone signals with the filter $\mathbf{w}_{WF,i}[k]$. The batch version is not suitable for real-time implementation because of the large delay introduced by the frame-based processing.

In the *recursive version*, the speech and the noise data matrices are updated for each time step $k$ with the newly available speech or noise data vector (depending on the output of the VAD-algorithm). Depending on the specific implementation, a fixed length data window (with length $p$ and $q$ for speech and noise respectively), or an exponential weighting window (with exponential weighting factors $\lambda_y$ and $\lambda_v$, cf. Section IV-B) can be used. For each time $k$, the GSVD of $\mathbf{Y}[k]$ and $\mathbf{V}[k]$ and the optimal filter matrix $\mathbf{W}_{\mathrm{WF}}[k]$ are recomputed and the enhanced signal at time $k$ is obtained by filtering the microphone signals with the filter $\mathbf{w}_{WF,i}[k]$. The recursive version introduces only a small processing delay equal to $\Delta = \frac{L}{2} - 1$ samples, and is able to track changing acoustic environments and signal statistics faster than the batch version. However, since at each time step the GSVD and the optimal filter need to be recalculated, the computational complexity is quite high. As will be shown in Section IV, the computational complexity can be drastically reduced by using recursive GSVD-updating algorithms.

In Sections V-D it will be shown using simulations that the batch and the recursive version of the GSVD-based optimal filtering technique nearly have the same performance.

### E. Other Implementations

Instead of using the discussed fullband GSVD-based implementation of the multichannel Wiener expression (9), other implementations exist, which exhibit a lower computational complexity and/or a better performance. In [26], a subband implementation of the GSVD-based optimal filtering technique has been proposed, leading to a better performance than the fullband implementation, since the MSE can be optimized in each individual subband, which is perceptually more relevant. In [27] and [28] a (fast) QRD-based implementation has been proposed, leading to a lower complexity scheme having nearly the same performance. However, in this QRD-based implementation it is not possible to incorporate the low-rank model of the speech signal. In [29]–[30] stochastic gradient LMS-based implementations in the frequency-domain have been proposed with an even lower computational complexity.

## III. ANC Postprocessing Stage

In [16]–[18] and [26], it has been shown that the GSVD-based optimal filtering technique has a better noise reduction performance than standard beamforming techniques, if the used filter length $L$ is large enough. However, the same noise reduction performance can be obtained with shorter filter lengths at a lower total computational complexity by incorporating the GSVD-based optimal filtering technique in a GSC-type structure, i.e., by adding an ANC postprocessing stage.

This postprocessing stage is a widely used structure in adaptive beamforming, where speech and noise reference signals are created and then used in an adaptive noise cancellation algorithm [5]–[15]. The objective is to create a speech reference signal having a higher signal-to-noise ratio (SNR) than the original microphone signals and to create one or more noise reference signals containing as little speech energy as possible. A multichannel adaptive filter (e.g., NLMS, APA, RLS [7]–[8]) then removes the remaining correlation between the (residual) noise component in the speech reference signal and the noise reference signals. In order to avoid signal cancellation and distortion, signal leakage into the noise reference (e.g., caused by reverberation, microphone mismatch, look direction error and spatially distributed sources) needs to be minimized and the effect of the signal leakage on the ANC adaptive filters needs to be limited. For adaptive beamformers, signal leakage can be reduced by e.g., using a spatial filter designed blocking matrix [11], [13], whereas, the effect of the signal leakage on the ANC adaptive filters can be limited by e.g., using a speech-controlled (VAD) adaptation algorithm [9]–[12] or constrained adaptive filters [12], [14].

However, instead of using a fixed beamformer to create the speech reference signal, it is also possible to use the GSVD-based optimal filtering technique. The complete noise reduction scheme, incorporating the GSVD-based optimal filter in a GSC-type structure with an ANC postprocessing stage, is depicted in Fig. 4. The output signal of the GSVD-based optimal filter is used as the *speech reference* signal

$$r_{\mathrm{speech}}[k] = \hat{x}_m[k-\Delta] = \mathbf{y}^T[k]\,\mathbf{w}_{WF,i}[k] \tag{20}$$

which is the optimal estimate for the speech component in the $m$th microphone signal (with delay $\Delta$), obtained by filtering the microphone signals with $\mathbf{w}_{WF,i}[k]$, with $i = mL + \Delta + 1$. The residual noise level in the speech reference signal depends on the filter length $L$ used for the GSVD-based optimal filter. For the creation of a *noise reference* different possibilities exist. An obvious choice consists in simply subtracting the speech reference signal from the delayed $m$th microphone signal, i.e.,

$$\begin{aligned} r_{\mathrm{noise},1}[k] &= y_m[k-\Delta] - r_{\mathrm{speech}}[k] \\ &= y_m[k-\Delta] - \hat{x}_m[k-\Delta]. \end{aligned} \tag{21}$$

Indeed, if $\mathbf{W}_{\mathrm{WF}}[k]$ is the optimal filter matrix for estimating the speech components in the microphone signals, i.e., $\hat{\mathbf{x}}[k] = \mathbf{W}_{\mathrm{WF}}^T[k]\,\mathbf{y}[k]$, then it is easily shown that $(\mathbf{I}_M - \mathbf{W}_{\mathrm{WF}}[k])$ is the optimal filter matrix for estimating the noise components in the microphone signals, i.e., $\hat{\mathbf{v}}[k] = (\mathbf{I}_M - \mathbf{W}_{\mathrm{WF}}^T[k])\,\mathbf{y}[k]$. The $i$th element of $\hat{\mathbf{v}}[k]$ is equal to the optimal estimate of the noise component in the $m$th microphone signal (with delay $\Delta$), i.e.,

$$\begin{aligned} \hat{v}_m[k-\Delta] &= \mathbf{y}^T[k](\mathbf{e}_i - \mathbf{w}_{WF,i}[k]) \\ &= y_m[k-\Delta] - \hat{x}_m[k-\Delta] \end{aligned} \tag{22}$$

where $\mathbf{e}_i$ is an $M$-dimensional vector with all zeros, except for the $i$th element which is equal to 1. Instead of only calculating a noise reference for one microphone signal, it is also possible to calculate noise references for all microphone signals, i.e.,

$$\begin{aligned} \mathbf{r}_{\mathrm{noise},2}[k] &= \begin{bmatrix} \hat{v}_0[k-\Delta] \\ \hat{v}_1[k-\Delta] \\ \vdots \\ \hat{v}_{N-1}[k-\Delta] \end{bmatrix} \\ &= \begin{bmatrix} y_0[k-\Delta] - \hat{x}_0[k-\Delta] \\ y_1[k-\Delta] - \hat{x}_1[k-\Delta] \\ \vdots \\ y_{N-1}[k-\Delta] - \hat{x}_{N-1}[k-\Delta] \end{bmatrix}. \end{aligned} \tag{23}$$

In order to construct $\mathbf{r}_{\mathrm{noise},2}[k]$, optimal estimates for the speech components in all microphone signals need to be computed.

Also for the ANC postprocessing stage of the GSVD-based optimal filtering technique, signal leakage into the noise reference will occur, since the estimate of the speech component $\hat{x}_m[k-\Delta]$ is generally not exactly equal to $x_m[k-\Delta]$. However, signal leakage can be reduced by using longer filter lengths $L$ for the GSVD-based optimal filter and the effect of the signal leakage on the ANC adaptive filters can be limited by using a speech-controlled (VAD) adaptation algorithm, where the ANC adaptive filters are only allowed to adapt during noise-only periods [9]–[12].

In Section V-F the noise reduction improvement and the additional speech distortion of the ANC postprocessing stage will be investigated experimentally for different filter lengths of the GSVD-based optimal filter and the ANC adaptive filter and for the two different noise references $r_{\mathrm{noise},1}[k]$ and $\mathbf{r}_{\mathrm{noise},2}[k]$. It will be shown that the SNR of the enhanced signal improves with increasing filter lengths and increasing number of noise reference signals. It will also be shown that the decrease in noise reduction performance due to short filter lengths $L$ can be fully compensated by adding the ANC postprocessing stage, at a lower total computational complexity. The ANC postprocessing stage will however give rise to a slight increase in speech distortion, which can be limited by using longer filter lengths for the GSVD-based optimal filter and for the ANC adaptive filter.

## IV. Recursive GSVD-Updating and Subsampling

As already stated in Section II-D, in the recursive version of the GSVD-based optimal filtering technique, the GSVD of the speech and the noise data matrices needs to be recomputed at each time step, giving rise to a high computational complexity, even when using short filter lengths $L$. This section describes several techniques for drastically reducing the computational complexity by using recursive Jacobi-type GSVD-updating algorithms and by using subsampling. In addition, a summary of the total computational complexity is given for realistic parameter values.

### *A. Jacobi-Type Algorithm for Computing GSVD*

For conciseness, the time index $k$ will be omitted in this section. The GSVD of two matrices $\mathbf{Y}$ and $\mathbf{V}$ can be computed as follows (for details, see [24] and [25]). First, the matrices $\mathbf{Y}$ and $\mathbf{V}$ are reduced to upper triangular form by a QR-decomposition

$$\mathbf{Y} = \mathbf{Q}_Y \cdot \mathbf{R}_Y, \quad \mathbf{V} = \mathbf{Q}_V \cdot \mathbf{R}_V \tag{24}$$

where $\mathbf{R}_Y$ and $\mathbf{R}_V$ are square upper triangular matrices, and $\mathbf{Q}_Y$ and $\mathbf{Q}_V$ have orthonormal columns, i.e., $\mathbf{Q}_Y^T \cdot \mathbf{Q}_Y = \mathbf{Q}_V^T \cdot \mathbf{Q}_V = \mathbf{I}_M$. The GSVD of $\mathbf{Y}$ and $\mathbf{V}$ readily follows from the GSVD of $\mathbf{R}_Y$ and $\mathbf{R}_V$, which is computed by carrying out an iterative procedure, where a series of orthogonal Givens transformations are applied to $\mathbf{R}_Y$ and $\mathbf{R}_V$ in order to yield square upper triangular factors $\mathbf{S}_Y$ and $\mathbf{S}_V$ with parallel rows, i.e.,

$$\begin{cases} \mathbf{U}_{R_Y}^T \cdot \mathbf{R}_Y \cdot \mathbf{Q}_R = \mathbf{S}_Y = \mathbf{\Sigma}_Y \cdot \mathbf{R} \\ \mathbf{U}_{R_V}^T \cdot \mathbf{R}_V \cdot \mathbf{Q}_R = \mathbf{S}_V = \mathbf{\Sigma}_V \cdot \mathbf{R} \end{cases} \tag{25}$$

where $\mathbf{U}_{R_Y}, \mathbf{U}_{R_V}$, and $\mathbf{Q}_R$ are orthogonal matrices, $\mathbf{\Sigma}_Y$ and $\mathbf{\Sigma}_V$ diagonal matrices and $\mathbf{R}$ a square upper triangular matrix. Combining (24) and (25), the GSVD of $\mathbf{Y}$ and $\mathbf{V}$ can be written as

$$\begin{cases} \mathbf{Y} = \mathbf{Q}_Y \cdot \mathbf{R}_Y = \mathbf{U}_Y \cdot \mathbf{S}_Y \cdot \mathbf{Q}_R^T \triangleq \mathbf{U}_Y \cdot \mathbf{\Sigma}_Y \cdot \mathbf{Q}^T \\ \mathbf{V} = \mathbf{Q}_V \cdot \mathbf{R}_V = \mathbf{U}_V \cdot \mathbf{S}_V \cdot \mathbf{Q}_R^T \triangleq \mathbf{U}_V \cdot \mathbf{\Sigma}_V \cdot \mathbf{Q}^T \end{cases} \tag{26}$$

with $\mathbf{U}_Y = \mathbf{Q}_Y \cdot \mathbf{U}_{R_Y}, \mathbf{U}_V = \mathbf{Q}_V \cdot \mathbf{U}_{R_V}$ and $\mathbf{Q}^T = \mathbf{R} \cdot \mathbf{Q}_R^T$.

The algorithm for computing the matrices $\mathbf{U}_{R_Y}, \mathbf{U}_{R_V}, \mathbf{S}_Y, \mathbf{S}_V$, and $\mathbf{Q}_R$ is presented below (typically only $\mathbf{S}_Y, \mathbf{S}_V$, and $\mathbf{Q}_R$ are stored).

```
1) Initialization:

    $\mathbf{S}_Y \Leftarrow \mathbf{R}_Y \quad \mathbf{U}_{R_Y} \Leftarrow \mathbf{I}_M \quad \mathbf{Q}_R \Leftarrow \mathbf{I}_M$
    $\mathbf{S}_V \Leftarrow \mathbf{R}_V \quad \mathbf{U}_{R_V} \Leftarrow \mathbf{I}_M$

2) Iterative GSVD-procedure
    for $j = 1 \ldots \alpha M$ (sweeps)
      for $i = 1 \ldots M-1$ (GSVD-steps)

        $\mathbf{S}_Y \Leftarrow \mathbf{\Theta}_{i,j}^T \cdot \mathbf{S}_Y \cdot \mathbf{Q}_{i,j} \quad \mathbf{U}_{R_Y} \Leftarrow \mathbf{U}_{R_Y} \cdot \mathbf{\Theta}_{i,j}$   (27)
        $\mathbf{S}_V \Leftarrow \mathbf{\Phi}_{i,j}^T \cdot \mathbf{S}_V \cdot \mathbf{Q}_{i,j} \quad \mathbf{U}_{R_V} \Leftarrow \mathbf{U}_{R_V} \cdot \mathbf{\Phi}_{i,j}$   (28)
        $\mathbf{Q}_R \Leftarrow \mathbf{Q}_R \cdot \mathbf{Q}_{i,j}$   (29)

      end
    end
```

The orthogonal matrices $\mathbf{\Theta}_{i,j}$ and $\mathbf{\Phi}_{i,j}$ in (27) and (28) represent plane Givens rotations with rotation angles $\theta_{i,j}$ and $\phi_{i,j}$ in the $(i, i+1)$-plane, i.e.,

$$\mathbf{\Theta}_{i,j} = \begin{bmatrix} \mathbf{I}_{i-1} & & & \\ & -\sin\theta_{i,j} & \cos\theta_{i,j} & \\ & \cos\theta_{i,j} & \sin\theta_{i,j} & \\ & & & \mathbf{I}_{M-i-1} \end{bmatrix}$$
$$\mathbf{\Phi}_{i,j} = \begin{bmatrix} \mathbf{I}_{i-1} & & & \\ & -\sin\phi_{i,j} & \cos\phi_{i,j} & \\ & \cos\phi_{i,j} & \sin\phi_{i,j} & \\ & & & \mathbf{I}_{M-i-1} \end{bmatrix}. \quad (30)$$

In each iteration, the computation of the rotation angles $\theta_{i,j}$ and $\phi_{i,j}$ and $\mathbf{Q}_{i,j}$, essentially reduces to the GSVD of the elementary $2 \times 2$-dimensional blocks $\{\mathbf{S}_Y\}_{i,i+1}$ and $\{\mathbf{S}_V\}_{i,i+1}$ on the main diagonal, where $\{\mathbf{A}\}_{i,i+1}$ denotes the $2 \times 2$-dimensional matrix on the intersection of rows $\{i, i+1\}$ and columns $\{i, i+1\}$ of the matrix $\mathbf{A}$. The pivot index $i$ repeatedly takes up all possible values $i = 1 \ldots M-1$ on the main diagonal. Here, one such sequence is referred to as a sweep ($= M-1$ GSVD-steps).

Since the GSVD of the upper triangular matrices $\mathbf{S}_Y$ and $\mathbf{S}_V$ corresponds to the SVD of the upper triangular matrix $\mathbf{S}_C = \mathbf{S}_Y \mathbf{S}_V^{-1}$, it is possible to implicitly apply a Jacobi-type SVD-algorithm to $\mathbf{S}_C$ without explicitly having to compute $\mathbf{S}_V^{-1}$ and $\mathbf{S}_C$ [24]. The GSVD of the $2 \times 2$-dimensional blocks $\{\mathbf{S}_Y\}_{i,i+1}$ and $\{\mathbf{S}_V\}_{i,i+1}$ corresponds to the SVD of the $2 \times 2$-dimensional upper triangular matrix $\{\mathbf{S}_C\}_{i,i+1}$

$$\begin{aligned} \{\mathbf{S}_C\}_{i,i+1} &= \{\mathbf{S}_Y\}_{i,i+1} \cdot \{\mathbf{S}_V^{-1}\}_{i,i+1} \\ &= \{\mathbf{S}_Y\}_{i,i+1} \cdot \{\mathbf{S}_V\}_{i,i+1}^{-1} \\ &= \begin{bmatrix} \frac{s_Y^{i,i}}{s_V^{i,i}} & \frac{s_Y^{i,i+1} s_V^{i,i} - s_Y^{i,i} s_V^{i,i+1}}{s_V^{i,i} s_V^{i+1,i+1}} \\ 0 & \frac{s_Y^{i+1,i+1}}{s_V^{i+1,i+1}} \end{bmatrix} \end{aligned} \quad (31)$$

which comes down to calculating the Givens rotation angles $\theta_{i,j}$ and $\phi_{i,j}$ (cf., [24], [31]) such that

$$\underbrace{\begin{bmatrix} \tilde{s}_C^{i,i} & 0 \\ 0 & \tilde{s}_C^{i+1,i+1} \end{bmatrix}}_{\{\mathbf{\Sigma}\}_{i,i+1}} = \underbrace{\begin{bmatrix} -\sin\theta_{i,j} & \cos\theta_{i,j} \\ \cos\theta_{i,j} & \sin\theta_{i,j} \end{bmatrix}}_{\{\mathbf{\Theta}_{i,j}^T\}_{i,i+1}} \cdot \underbrace{\begin{bmatrix} s_C^{i,i} & s_C^{i,i+1} \\ 0 & s_C^{i+1,i+1} \end{bmatrix}}_{\{\mathbf{S}_C\}_{i,i+1}} \cdot \underbrace{\begin{bmatrix} -\sin\phi_{i,j} & \cos\phi_{i,j} \\ \cos\phi_{i,j} & \sin\phi_{i,j} \end{bmatrix}}_{\{\mathbf{\Phi}_{i,j}\}_{i,i+1}}. \quad (32)$$

These orthogonal transformations are seen to parallelize the rows of $\{\mathbf{S}_Y\}_{i,i+1}$ and $\{\mathbf{S}_V\}_{i,i+1}$, i.e.,

$$\{\mathbf{\Theta}_{i,j}^T\}_{i,i+1} \cdot \{\mathbf{S}_Y\}_{i,i+1} = \{\mathbf{\Sigma}\}_{i,i+1} \cdot \{\mathbf{\Phi}_{i,j}^T\}_{i,i+1} \cdot \{\mathbf{S}_V\}_{i,i+1} \quad (33)$$

which then allows for a joint upper triangularizing orthogonal transformation $\{\mathbf{Q}_{i,j}\}_{i,i+1}$ in order to obtain the GSVD of $\{\mathbf{S}_Y\}_{i,i+1}$ and $\{\mathbf{S}_V\}_{i,i+1}$, i.e.,

$$\begin{aligned} &\underbrace{\{\mathbf{\Theta}_{i,j}^T\}_{i,i+1} \cdot \{\mathbf{S}_Y\}_{i,i+1} \cdot \{\mathbf{Q}_{i,j}\}_{i,i+1}}_{\text{uppertriangular}} \\ &= \{\mathbf{\Sigma}\}_{i,i+1} \cdot \underbrace{\{\mathbf{\Phi}_{i,j}^T\}_{i,i+1} \cdot \{\mathbf{S}_V\}_{i,i+1} \cdot \{\mathbf{Q}_{i,j}\}_{i,i+1}}_{\text{uppertriangular}}. \end{aligned} \quad (34)$$

TABLE I
SUMMARY OF TOTAL COMPLEXITY OF GSVD-BASED OPTIMAL FILTERING TECHNIQUE FOR BATCH AND RECURSIVE VERSIONS USING REALISTIC PARAMETER VALUES

| | Non-recursive/Batch | Recursive | Square root-free |
|---|---|---|---|
| | $\frac{16M^3+3(p+q)M^2}{s_g}$ | $\frac{23.5M^2}{s_g} + \frac{4M^2}{s_f}$ | $\frac{16.5M^2}{s_g} + \frac{4M^2}{s_f}$ |
| $s_f = s_g = 1$ | 7504 Gflops | 2.8 Gflops | 2.1 Gflops |
| $s_f = s_g = 20$ | 375 Gflops | 141 Mflops | 105 Mflops |

Since computing a full GSVD requires $\alpha M$ sweeps (with $\alpha$ typically $3 \ldots 5$ for convergence [32]), the total computational complexity, defined as the total number of additions and multiplications, amounts to $3M^2(p + q - 2M/3)$(QR-decomposition) $+ 18\alpha M^3$ (GSVD-procedure), such that the total complexity is equal to $(18\alpha - 2)M^3 + 3M^2(p+q)$. For typical values of $p$, $q$, and $M$, the complexity of this algorithm is too high to be suitable for real-time implementation (see Table I).

### B. Recursive GSVD-Updating Algorithm

Instead of recomputing the GSVD from scratch at each time step, recursive GSVD-updating algorithms compute the GSVD at time $k$ using the decomposition at time $k-1$. In [33] and [34] a Jacobi-type (G)SVD-updating algorithm has been described. Suppose that at time $k-1$ the upper triangular factors are reduced to $\mathbf{S}_Y[k-1]$ and $\mathbf{S}_V[k-1]$ having approximately parallel rows, cf. (25), shown in

$$\begin{cases} \mathbf{Y}[k-1] = \mathbf{U}_Y[k-1] \cdot \mathbf{S}_Y[k-1] \cdot \mathbf{Q}_R^T[k-1] \\ \qquad\quad \triangleq \mathbf{U}_Y[k-1] \cdot \mathbf{\Sigma}_Y[k-1] \cdot \mathbf{Q}^T[k-1] \\ \mathbf{V}[k-1] = \mathbf{U}_V[k-1] \cdot \mathbf{S}_V[k-1] \cdot \mathbf{Q}_R^T[k-1] \\ \qquad\quad \triangleq \mathbf{U}_V[k-1] \cdot \mathbf{\Sigma}_V[k-1] \cdot \mathbf{Q}^T[k-1] \end{cases} \quad (35)$$

of which only $\mathbf{S}_Y[k-1]$, $\mathbf{S}_V[k-1]$ and the orthogonal matrix $\mathbf{Q}_R[k-1]$ are stored and updated.

At time $k$, a new data vector $\mathbf{y}[k]$ is present, such that we need to recompute the GSVD of the updated data matrices $\mathbf{Y}[k]$ and $\mathbf{V}[k]$, which are constructed by weighting $\mathbf{Y}[k-1]$ and $\mathbf{V}[k-1]$ and adding the new data vector (when using fixed length data windows, also a down-date has to be performed, which is not numerically stable). If $\mathbf{y}[k]$ is classified by the VAD-algorithm as a speech-and-noise vector ($\beta_k = 1$), only the speech data matrix $\mathbf{Y}[k]$ is updated, i.e.,

$$\mathbf{Y}[k] = \begin{bmatrix} \lambda_y \cdot \mathbf{Y}[k-1] \\ \mathbf{y}^T[k] \end{bmatrix}, \quad \mathbf{V}[k] = \mathbf{V}[k-1] \quad (36)$$

whereas if $\mathbf{y}[k]$ is classified as a noise-only vector ($\beta_k = 0$), only the noise data matrix $\mathbf{V}[k]$ is updated, i.e.,

$$\mathbf{Y}[k] = \mathbf{Y}[k-1], \quad \mathbf{V}[k] = \begin{bmatrix} \lambda_v \cdot \mathbf{V}[k-1] \\ \mathbf{y}^T[k] \end{bmatrix} \tag{37}$$

with $\lambda_y$ the exponential weighting factor for speech and $\lambda_v$ the exponential weighting factor for noise (if $\lambda = 1$, no weighting is performed). Assuming that $\mathbf{y}[k]$ is classified as a speech-and-noise vector, the speech data matrix $\mathbf{Y}[k]$ can be rewritten as

$$\mathbf{Y}[k] = \left[\begin{array}{cc} \boxed{\mathbf{U}_Y[k-1]} & \boxed{\begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}} \\ \boxed{0 \;\cdots\; 0} & \boxed{1} \end{array}\right] \cdot \begin{bmatrix} \lambda_y \cdot \mathbf{S}_Y[k-1] \\ \mathbf{y}^T[k] \cdot \mathbf{Q}_R[k-1] \end{bmatrix} \cdot \mathbf{Q}_R^T[k-1]. \tag{38}$$

First, the upper triangular factor is restored by performing a QR-update with the transformed input vector $\tilde{\mathbf{y}}^T[k] = \mathbf{y}^T[k] \cdot \mathbf{Q}_R[k-1]$. QR-updating can be performed by using orthogonal Givens rotations, zeroing the elements on the bottom row, yielding the upper triangular matrix $\tilde{\mathbf{S}}_Y[k]$

$$\mathbf{Y}[k] = \underbrace{\left[\begin{array}{cc} \boxed{\mathbf{U}_Y[k-1]} & \boxed{\begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}} \\ \boxed{0 \;\cdots\; 0} & \boxed{1} \end{array}\right] \cdot \tilde{\mathbf{Q}}_Y[k]}_{\tilde{\mathbf{U}}_Y[k]} \cdot \tilde{\mathbf{S}}_Y[k] \cdot \mathbf{Q}_R^T[k-1]. \tag{39}$$

In this equation, $\tilde{\mathbf{Q}}_Y[k]$ is an $(M+1) \times M$-dimensional matrix with orthogonal columns, which does not need to be computed explicitly. The matrix $\mathbf{Q}_R[k-1]$ is not altered by the QR-update. If $\mathbf{y}[k]$ is classified as a noise-only vector, a similar procedure needs to be performed for $\mathbf{V}[k]$ instead of for $\mathbf{Y}[k]$.

Second, the iterative GSVD-procedure is resumed in order to further parallelise the rows of the square upper triangular matrices $\tilde{\mathbf{S}}_Y[k]$ and $\tilde{\mathbf{S}}_V[k]$. A fixed number of sweeps ($s$) is performed, where the pivot index $i$ takes up $r$ consecutive values. Typically one sweep is performed ($s = 1$), where the pivot index takes up all possible values along the main diagonal ($r = M - 1$).

The complete procedure at time $k$, where only the square upper triangular matrices $\mathbf{S}_Y[k]$ and $\mathbf{S}_V[k]$ and the orthogonal matrix $\mathbf{Q}_R[k]$ are stored and updated, can be summarized as follows:

```
1) matrix-vector multiplication and QR-update
   if $\beta_k = 1$ (speech-and-noise)
      $\mathbf{S}_Y[k] \Leftarrow \tilde{\mathbf{Q}}_Y^T[k] \cdot \begin{bmatrix} \lambda_y \cdot \mathbf{S}_Y[k-1] \\ \mathbf{y}^T[k] \cdot \mathbf{Q}_R[k-1] \end{bmatrix}$   (40)
      $\mathbf{S}_V[k] \Leftarrow \mathbf{S}_V[k-1]$   (41)
   else if $\beta_k = 0$ (noise-only)
      $\mathbf{S}_Y[k] \Leftarrow \mathbf{S}_Y[k-1]$   (42)
      $\mathbf{S}_V[k] \Leftarrow \tilde{\mathbf{Q}}_V^T[k] \cdot \begin{bmatrix} \lambda_v \cdot \mathbf{S}_V[k-1] \\ \mathbf{y}^T[k] \cdot \mathbf{Q}_R[k-1] \end{bmatrix}$   (43)
   end
   $\mathbf{Q}_R[k] \Leftarrow \mathbf{Q}_R[k-1]$
2) GSVD-update procedure
   $r_{k+1} = \mathrm{mod}(r_k + r - 1, M - 1) + 1$
   for $j = 1 \ldots s$ (sweeps)
     for $i = r_k \ldots r_{k+1} - 1$ (GSVD-steps)
        $\mathbf{S}_Y[k] \Leftarrow \mathbf{\Theta}_{i,j}^T[k] \cdot \mathbf{S}_Y[k] \cdot \mathbf{Q}_{i,j}[k]$   (44)
        $\mathbf{S}_V[k] \Leftarrow \mathbf{\Phi}_{i,j}^T[k] \cdot \mathbf{S}_V[k] \cdot \mathbf{Q}_{i,j}[k]$   (45)
        $\mathbf{Q}_R[k] \Leftarrow \mathbf{Q}_R[k] \cdot \mathbf{Q}_{i,j}[k]$   (46)
     end
   end
```

The computational complexity of one GSVD-update is equal to $2.5M^2$(matrix-vector multiplication) $+ 3M^2$(QR-update) $+ s \cdot r/(M-1) \cdot 18M^2$ (GSVD-update procedure). For $s = 1$ and $r = M - 1$, the total complexity amounts to $23.5M^2$.

The optimal filter matrix $\mathbf{W}_{\mathrm{WF}}[k]$ in (19) can now be computed as

$$\mathbf{W}_{\mathrm{WF}}[k] = \mathbf{Q}^{-T}[k] \, \mathrm{diag} \left\{ 1 - \frac{(1-\lambda_v^2)}{(1-\lambda_y^2)} \frac{\eta_i^2[k]}{\sigma_i^2[k]} \right\} \mathbf{Q}^T[k] \tag{47}$$

where the factor $p/q$ has been replaced by $(1-\lambda_v^2)/(1-\lambda_y^2)$, because exponential weighting is used. Upon convergence of the recursive GSVD-updating algorithm, it follows from (35) that $\mathbf{Q}^T[k] = \mathbf{\Sigma}_Y^{-1}[k] \cdot \mathbf{S}_Y[k] \cdot \mathbf{Q}_R^T[k]$ and $s_Y^{i,i}[k]/s_V^{i,i}[k] = \sigma_i[k]/\eta_i[k]$, since $\mathbf{S}_Y[k]$ and $\mathbf{S}_V[k]$ have parallel rows. Hence, $\mathbf{W}_{\mathrm{WF}}[k]$ can be computed as

$$\begin{aligned} \mathbf{W}_{\mathrm{WF}}[k] &= \mathbf{Q}_R[k] \cdot \mathbf{S}_Y^{-1}[k] \cdot \mathbf{\Sigma}_Y[k] \\ &\quad \times \mathrm{diag} \left\{ 1 - \frac{(1-\lambda_v^2)}{(1-\lambda_y^2)} \frac{\eta_i^2[k]}{\sigma_i^2[k]} \right\} \\ &\quad \times \mathbf{\Sigma}_Y^{-1}[k] \cdot \mathbf{S}_Y[k] \cdot \mathbf{Q}_R^T[k] &(48) \\ &= \mathbf{Q}_R[k] \cdot \mathbf{S}_Y^{-1}[k] \\ &\quad \times \mathrm{diag} \left\{ 1 - \frac{(1-\lambda_v^2)}{(1-\lambda_y^2)} \frac{\left(s_V^{i,i}[k]\right)^2}{\left(s_Y^{i,i}[k]\right)^2} \right\} \\ &\quad \times \mathbf{S}_Y[k] \cdot \mathbf{Q}_R^T[k]. &(49) \end{aligned}$$

Since only the $i$th column $\mathbf{w}_{WF,i}[k]$ of $\mathbf{W}_{\mathrm{WF}}[k]$ needs to be computed, this column can be computed as the solution of the linear set of equations

$$\mathbf{S}_Y[k] \cdot \underbrace{\mathbf{Q}_R^T[k] \cdot \mathbf{w}_{WF,i}[k]}_{\tilde{\mathbf{w}}[k]} = \underbrace{\mathrm{diag} \left\{ 1 - \frac{(1-\lambda_v^2)}{(1-\lambda_y^2)} \frac{\left(s_V^{i,i}[k]\right)^2}{\left(s_Y^{i,i}[k]\right)^2} \right\} \cdot \mathbf{S}_Y[k] \cdot \mathbf{q}_{R,i}[k]}_{\tilde{\mathbf{q}}[k]} \tag{50}$$

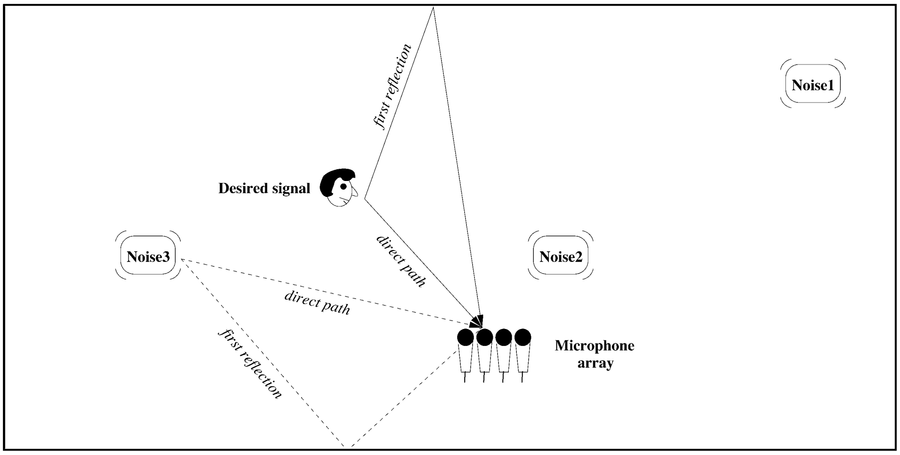


Fig. 5. Simulation environment.

where $\mathbf{q}_{R,i}[k]$ is the $i$th column of $\mathbf{Q}_R^T[k]$. The calculation of $\mathbf{w}_{WF,i}[k]$ consists of computing $\tilde{\mathbf{q}}[k]$, requiring $M^2$ operations (multiplication of triangular matrix with vector), solving the equation $\mathbf{S}_Y[k] \cdot \tilde{\mathbf{w}}[k] = \tilde{\mathbf{q}}[k]$ by back-substitution, requiring $M^2$ operations, and computing $\mathbf{w}_{WF,i}[k]$ as

$$\mathbf{w}_{WF,i}[k] = \mathbf{Q}_R[k] \cdot \tilde{\mathbf{w}}[k] \tag{51}$$

requiring $2M^2$ operations. Hence, the total computational complexity for computing $\mathbf{w}_{WF,i}[k]$ from the GSVD of $\mathbf{Y}[k]$ and $\mathbf{V}[k]$ amounts to $4M^2$.

The computational complexity can be further reduced by using a square root-free implementation for the QR-updates and for the calculation of the elementary $2 \times 2$-dimensional GSVD's. Since the above GSVD-schemes as such do not lend themselves to square root-free implementation, alternative schemes based on approximate formulas for the calculation of the rotation angles $\theta_{i,j}$ and $\phi_{i,j}$ have to be considered [31]. These schemes eventually yield square root-free SVD-updating algorithms [35], which can be easily extended to square root-free GSVD-updating algorithms [34]. It can be shown that the complexity of one square root-free GSVD-update is equal to $2.5M^2(\text{matrix-vector multiplication}) + 2M^2(\text{square root-free QR-update}) + s \cdot r/(M-1) \cdot 12M^2$ (square root-free GSVD-update procedure). For $s = 1$ and $r = M - 1$, the total computational complexity amounts to $16.5M^2$, which is less expensive than the "conventional" GSVD-updating procedure.

### C. Subsampling Techniques

For stationary acoustic environments the computational complexity can be reduced without any loss in performance by using subsampling techniques. In this context subsampling means that the GSVD of $\mathbf{Y}[k]$ and $\mathbf{V}[k]$ and the optimal filter $\mathbf{w}_{WF,i}[k]$ are not updated for every sample, but that the GSVD is updated every $s_g$ samples and that the optimal filter is updated every $s_f$ samples. If a higher subsampling factor is used, the convergence speed toward the converged optimal filter is slower, implying that the amount of subsampling should be limited in nonstationary acoustic environments.

### D. Total Computational Complexity

Table I summarizes the total complexity in floating point operations per second (flops) for the batch and the recursive version of the GSVD-based optimal filtering technique, assuming that $s = 1, r = M - 1$ and $\alpha = 1$. The numerical results are obtained for $N = 4$ microphones, filter length $L = 20$ $(M = 80)$, sampling frequency $f_s = 16$ kHz, data window lengths $p = 4000$ and $q = 20\,000$ (for the batch version) and are shown both in case of no subsampling and in case of subsampling with $s_g = s_f = 20$. By using the recursive version of the GSVD-based optimal filtering technique, the computational complexity can be significantly reduced such that the algorithm becomes suitable for real-time implementation.

## V. Simulation Results

This section discusses the performance (SNR improvement and speech distortion) of the GSVD-based optimal filtering technique with and without an ANC postprocessing stage. First, the simulation environment and the implementation details of the considered algorithms are described. Then, the performance difference between the batch and the recursive version of the GSVD-based optimal filtering technique and the effect of different parameters in the recursive GSVD-updating algorithms are discussed. The effect of the ANC postprocessing stage on the noise reduction performance and speech distortion is also analyzed. Finally, the performance of the recursive GSVD-based optimal filtering technique is compared with standard beamforming techniques for various simulated acoustic scenarios and a real-life recording.

### A. Simulation Environment

The simulation room is depicted in Fig. 5 and has dimensions 6 $\times$ 3 $\times$ 2.5 m. It consists of a microphone array, a speech source and 3 noise sources. Unless otherwise indicated, we will only use the noise source at position 1 (only in Section V-G, three simultaneous noise sources at different positions will be used). In our simulations we have used a linear equi-spaced microphone array with $N = 4$ microphones and the distance $d$ between two adjacent microphones is 5 cm. The speech source is located at 1.3 m from the centre of the microphone array at an angle of 56°. The used signals are a 16 kHz clean speech signal, consisting of english sentences from the "Hearing in Noise Test" [36], and three different noise signals: stationary white noise, stationary speech noise from the NOISEX-92 database [37], having the same long-term spectrum as speech, and a nonstationary classical music signal. The speech and the noise components received at the $n$th microphone are filtered versions of the clean speech and noise signals with simulated acoustic room impulse responses, constructed using the image method [38], [39] for different reverberation times $T_{60}$. The reverberation time $T_{60}$ can be expressed as a function of the absorption coefficient $\gamma$ of the walls, according to Eyring's formula [40]

$$T_{60} = \frac{0.163\, V}{-S \log(1-\gamma)} \tag{52}$$

where $V$is the volume of the room and $S$ the total surface of the room. Using simulated acoustic impulse responses, we can easily compare the performance for different reverberation conditions.

Since all described algorithmic operations (GSVD-based optimal filter, ANC postprocessing stage and fixed and adaptive beamforming) amount to linear filtering operations, the speech and the noise components of the output signal and all intermediate signals can be easily obtained by applying the computed filters separately to the speech and the noise components of the microphone signals. The performance of the GSVD-based optimal filtering technique will be described by the unbiased SNR improvement and by the introduced speech distortion. The *unbiased SNR* of a signal $z[k] = z_x[k] + z_v[k]$ can be computed during speech-and-noise periods as

$$\text{SNR} = 10 \log_{10} \frac{\sum_{\beta_k=1} z_x^2[k]}{\sum_{\beta_k=1} z_v^2[k]} \tag{53}$$

with $z_x[k]$ and $z_v[k]$, respectively, the speech and the noise component of the considered signal $z[k]$. *Speech distortion* will be analyzed by considering the Power Transfer Function (PTF) $G_x(\omega)$ between the speech component of the first microphone signal $x_0[k]$ and the speech component of the considered signal $z_x[k]$

$$G_x(\omega) = \frac{P_{z_x}(\omega)}{P_{x_0}(\omega)} \tag{54}$$

with $P_{x_0}(\omega)$ the power spectral density (PSD) of $x_0[k]$ and $P_{z_x}(\omega)$ the PSD of $z_x[k]$. The average speech distortion (SD) is computed as the average of the PTF $G_x(\omega)$ in dB over the full frequency band, i.e.,

$$\text{SD} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |10 \log_{10} G_x(\omega)|\, d\omega. \tag{55}$$

We also consider the Itakura-Saito distance [41] between the speech component of the first microphone signal $x_0[k]$ and the speech component of the considered signal $z_x[k]$. We have calculated this distance for consecutive frames of 480 samples (with an overlap of 360 samples) and an LPC-order of 12 and we will use the *average Itakura-Saito distance* over all frames as an additional measure for describing speech distortion.

In our simulations, we have constructed the noisy microphone signals such that the unbiased SNR of the first microphone signal $y_0[k]$ equals 0 dB. Figs. 6(a) and (b) depict the speech component $x_0[k]$ and the noisy microphone signal $y_0[k]$ for reverberation time $T_{60} = 300$ ms when using speech noise. Fig. 6(c) shows the enhanced signal $z[k]$ processed by the recursive GSVD-based optimal filtering technique with an ANC postprocessing stage using all noise reference signals.[1]

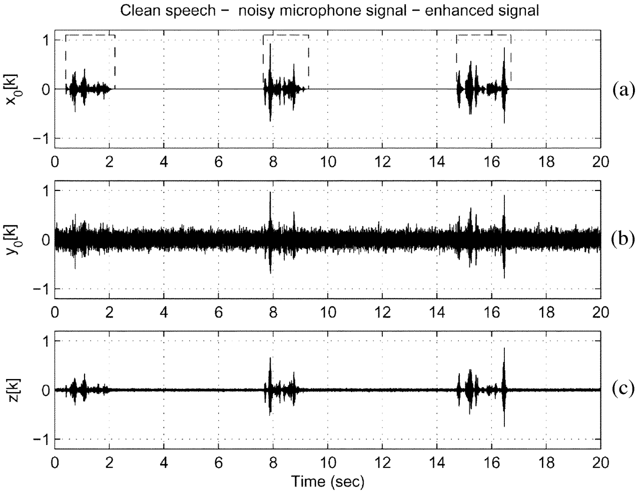


Fig. 6. (a) Speech component $x_0[k]$ and voice activity detection. (b) Noisy microphone signal $y_0[k]$ (speech noise, SNR $=$ 0 dB, $T_{60} = 300$ ms). (c) Enhanced signal $z[k]$ using recursive GSVD-based optimal filtering technique with an ANC postprocessing stage ($L = 20, L_{\text{ANC}} = 400$, no subsampling, all noise references).

### *B. Implementation Issues for the GSVD-Based Optimal Filtering Technique*

Both for the batch and for the recursive version of the GSVD-based optimal filtering technique, a voice activity detection (VAD) algorithm determines when speech is present. Fig. 6(a) shows the output of a perfect VAD algorithm on the speech component of the first microphone signal. In [42] the effect of speech detection errors on the performance has been theoretically analyzed, and it has been shown that the unbiased SNR improvement of the optimal filtering technique is not degraded by speech detection errors, neither when speech is wrongly detected as noise nor when noise is wrongly detected as speech. However, speech distortion dramatically increases with speech detection error rate when speech is wrongly detected as noise, whereas speech distortion only slightly increases when noise is wrongly detected as speech. Hence, the VAD should be tuned such that especially the speech-and-noise periods are correctly classified. Both the theoretical analysis in [42] and an experimental validation in [18] have shown that the effect of speech detection errors on the speech distortion remains small when the speech detection error rate is smaller than 20%. In this paper, we will generally assume that a perfect VAD is available. Only for the real-life recordings in Section V-H, a (nonperfect) energy-based VAD will be used [21].

In the *batch GSVD-based optimal filtering technique*, the speech and the noise data matrices $\mathbf{Y}[k]$ and $\mathbf{V}[k]$ are

[1] For this specific simulation, sound files, spectrograms, and power transfer functions are available at http://www.esat.kuleuven.ac.be/~doclo/SA00061/audio.html.

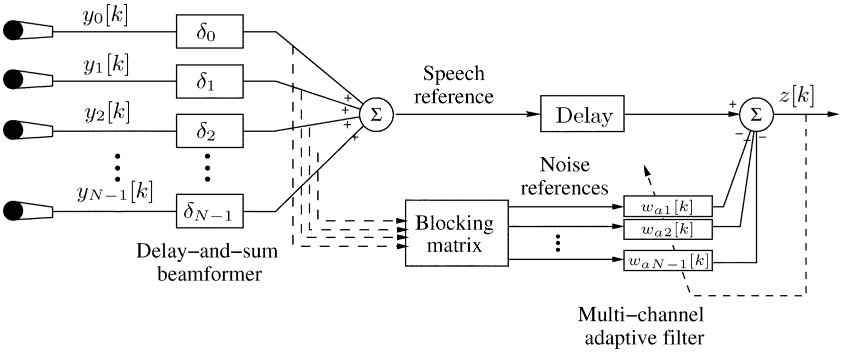


Fig. 7. Generalized sidelobe canceller (GSC).

constructed from the noisy microphone signals $y_n[k], n = 0 \ldots N-1$, using all available speech and noise samples. The filter length of the optimal filter is denoted by $L$. The optimal filter matrix $\mathbf{W}_{\mathrm{WF}}[k]$ is computed using (19), where all negative diagonal elements are put to zero. The stacked filter $\mathbf{w}[k]$ is determined as the $i$th column $\mathbf{w}_{WF,i}[k]$ of $\mathbf{W}_{\mathrm{WF}}[k]$, with the fixed value $i = L/2$ (cf. Section II-B). The total enhanced signal $z[k]$ is obtained by filtering the microphone signals with $\mathbf{w}_n[k], n = 0 \ldots N-1$.

In the *recursive GSVD-based optimal filtering technique*, the data matrices are updated according to (36) or (37), with $\lambda_y = 0.99999$ and $\lambda_v = 0.999995$. Using the recursive techniques of Section IV, the GSVD and the optimal filter are updated for every sample. The $i$th column $\mathbf{w}_{WF,i}[k]$ of $\mathbf{W}_{\mathrm{WF}}[k]$, with $i = L/2$, is computed using (50), and the enhanced signal at time $k$ is computed by filtering the microphone signals with $\mathbf{w}_n[k], n = 0 \ldots N-1$. When using subsampling, the GSVD and the optimal filter are updated, respectively, for every $s_g$ and $s_f$ samples. In order to avoid initial effects (initially no knowledge about either the speech nor the noise data matrix is available), signal segments twice as long as for the batch version are processed and only the last half is used for computing the performance measures.

For the *ANC postprocessing stage*, two possible noise references will be investigated: $r_{\mathrm{noise},1}[k]$ in (21) with one noise reference signal and $\mathbf{r}_{\mathrm{noise},2}[k]$ in (23) with $N = 4$ noise reference signals. The adaptive filter used in the ANC postprocessing stage is a time-domain NLMS algorithm [7]. The filter length of the adaptive filter is denoted by $L_{\mathrm{ANC}}$ and the step size $\mu = 0.05$. The speech reference signal is delayed by $L_{\mathrm{ANC}}/2$ samples in order for the adaptive filter to be able to model some acausal taps. As already mentioned in Section III, in order to limit signal cancellation and distortion, a speech-controlled adaptation algorithm will be used, where the ANC adaptive filter is only allowed to adapt during noise-only periods.

### C. Implementation Issues for the Fixed and Adaptive Beamforming Techniques

The performance of the GSVD-based optimal filtering technique will be compared with fixed and adaptive beamforming techniques. A fixed *delay-and-sum (DS) beamformer* spatially aligns the microphone signals to the direction of the speech source by delaying and summing the microphone signals, i.e.,

$$z[k] = \frac{1}{N} \sum_{n=0}^{N-1} y_n[k - \delta_n] \tag{56}$$

where the delays $\delta_n$ are computed as $\delta_n = -(nd \cos\theta_x / c) f_s$, with $\theta_x = 56°$ the direction of the speech source.

The standard *GSC* [6], depicted in Fig. 7, uses the output signal of a DS beamformer as speech reference signal, and creates a noise reference by combining the delayed microphone signals using a blocking matrix (e.g., Griffiths-Jim), blocking out signals arriving from the direction of the speech source

$$\mathbf{r}_{\mathrm{noise}}^{\mathrm{GSC}}[k] = \begin{bmatrix} y_0[k-\delta_0] - y_1[k-\delta_1] \\ y_1[k-\delta_1] - y_2[k-\delta_2] \\ \vdots \\ y_{N-2}[k-\delta_{N-2}] - y_{N-1}[k-\delta_{N-1}] \end{bmatrix}. \tag{57}$$

A multichannel adaptive filter then removes the correlation between the (residual) noise component in the speech reference signal and the noise reference signals. When using 1 noise reference signal, only the first element of $\mathbf{r}_{\mathrm{noise}}^{GSC}[k]$ is considered. The used adaptive filter is a time-domain NLMS algorithm [7], with filter length denoted by $L_{\mathrm{ANC}}$ and step size $\mu = 0.1$. The speech reference signal is delayed by $L_{\mathrm{ANC}}/2$ samples in order for the adaptive filter to be able to model some acausal taps.

As already mentioned in Section III, because of room reverberation, microphone mismatch and look direction error, signal leakage into the noise reference occurs. In order to limit the effect of the signal leakage on the adaptive filters, a speech-controlled (VAD) adaptation algorithm will be used, where the ANC adaptive filter is only allowed to adapt during noise-only periods [9]–[12].

However, it is also possible to reduce the amount of signal leakage in the noise reference by using a spatial filter designed blocking matrix [11], [13] instead of the standard Griffiths-Jim blocking matrix. We have designed a spatial filter for the blocking matrix using a nonlinear design criterion for far-field broadband beamformers [43] with stopband specifications $(\Omega_s, \Theta_s) = (0 - 7500\ \mathrm{Hz}, (\theta_x - 20°) - (\theta_x + 20°))$ and passband specifications $(\Omega_p, \Theta_p) = (0 - 7500\ \mathrm{Hz}, 0° - (\theta_x - 20°)$ and $(\theta_x + 20°) - 180°)$. We have designed this spatial filter with $L = 20$ taps per microphone and using $N - 1$ microphones, such that we are able to create two independent

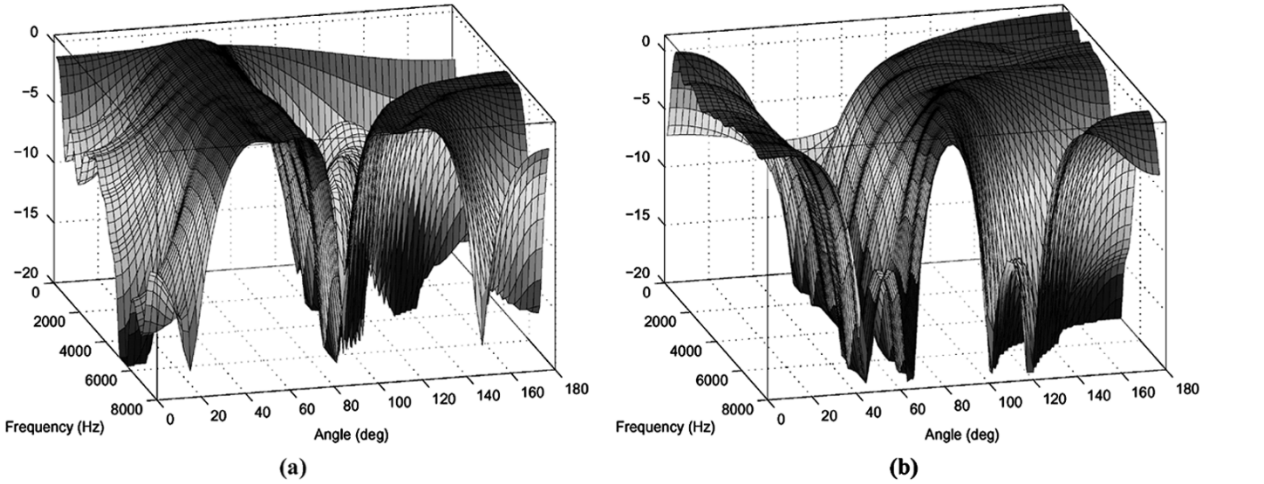


Fig. 8. Spatial directivity pattern of (a) fixed beamformer (speech reference) and (b) blocking matrix (noise reference).

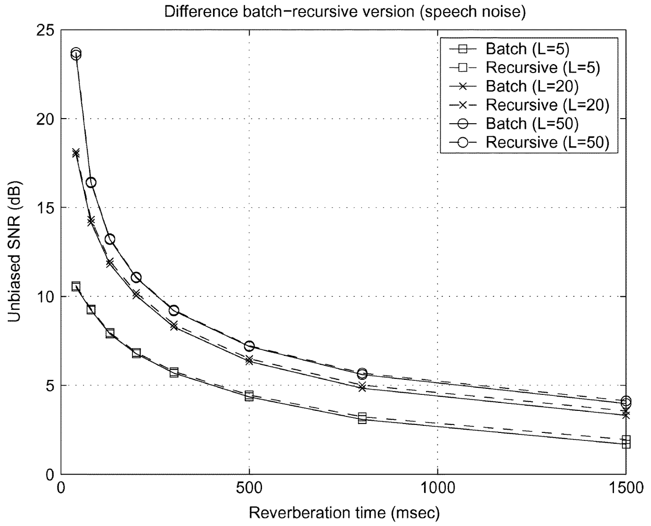


Fig. 9. Comparison of unbiased SNR for batch and recursive version of GSVD-based optimal filtering technique for different filter lengths $L$ (speech noise, no subsampling).

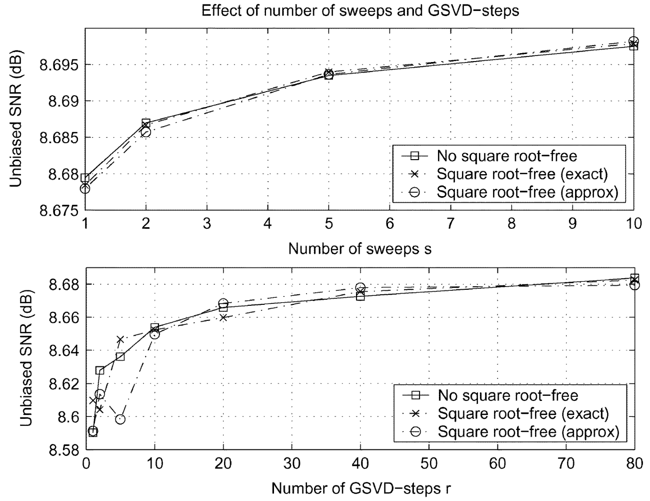


Fig. 10. Effect of number of sweeps, GSVD-steps and square root-free implementation on unbiased SNR for recursive GSVD-based optimal filtering technique (speech noise, $T_{60} = 300$ ms, $L = 20$, no subsampling).

noise reference signals [13]. The fixed beamformer for creating the speech reference signal has inverse stopband and passband specifications and is designed to be orthogonal to the blocking matrix, which can be achieved by imposing linear constraints in the design procedure [13]. The spatial directivity pattern of the fixed beamformer and the spatial filter designed blocking matrix are depicted in Fig. 8. Although the amount of signal leakage into the noise reference will be reduced, it can never be completely avoided (certainly not in highly reverberant acoustic environments). Therefore, we will still use a speech-controlled (VAD) adaptation algorithm, switching off the adaptation during speech-and-noise periods.

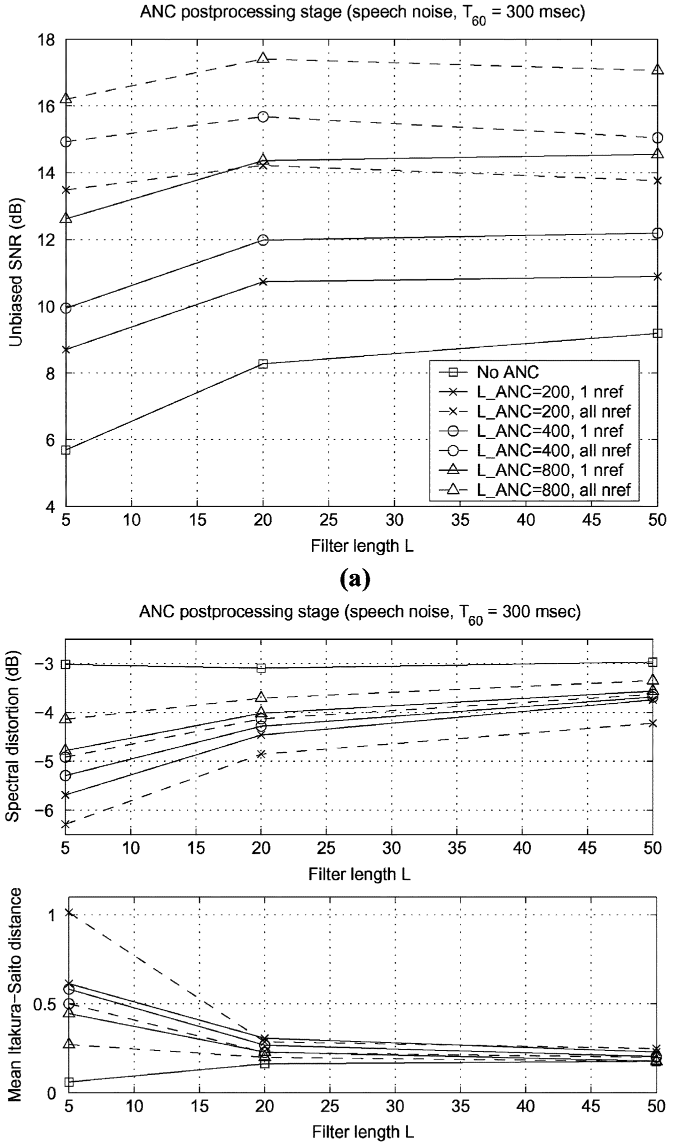
The speech distortion measures, defined in Section V-A, are not really useful for fixed and adaptive beamformers, since these speech distortion measures consider the speech component in the first microphone signal $x_0[k]$, whereas the DS beamformer and the GSC try to recover the signal $s[k]$.

### *D. Batch versus Recursive Version*

Fig. 9 compares the unbiased SNR of the enhanced signal for the batch and the recursive version of the GSVD-based optimal filtering technique (without an ANC postprocessing stage). The noisy microphone signals have been constructed using a speech noise source at position 1, and the simulations have been performed for different reverberation times $T_{60}$ and for different filter lengths $L$ of the GSVD-based optimal filter and without subsampling. As can be seen from Fig. 9, the unbiased SNR increases for higher filter lengths $L$ and for lower reverberation times $T_{60}$. This can be explained from the fact that in highly reverberant acoustic environments the GSVD-based optimal filtering technique will tradeoff noise reduction and cancellation of the reverberant part of the speech signal, in order to make an optimal estimate of the speech component $x_0[k]$. As can also be seen from this figure, the performance of the batch and the recursive version are practically equal for all reverberation times and filter lengths. The performance of the recursive version is even slightly better, because it is able to adapt to (local) changes in the spatio-temporal statistics of the speech and the noise sources.

### *E. Recursive GSVD-Updating Algorithms*

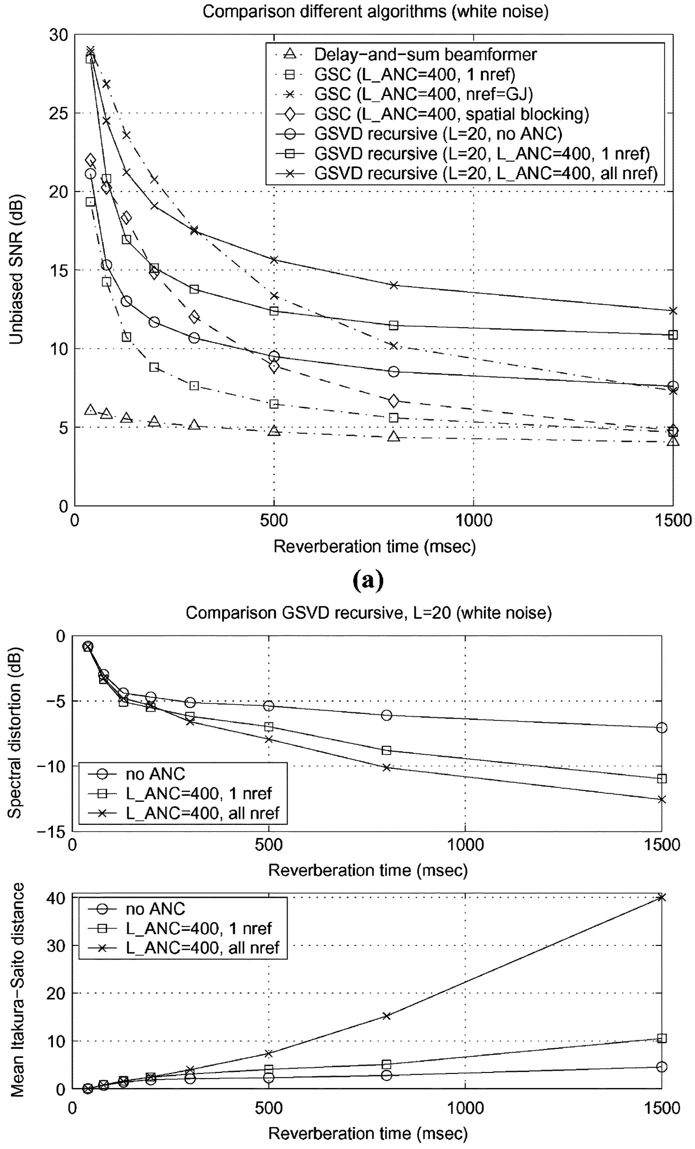
As discussed in Section IV-B, different implementations of the recursive GSVD-updating algorithm exist: a “conventional” implementation and a square root-free implementation, both



(a)

(b)

Fig. 11. Effect of the ANC postprocessing stage on (a) unbiased SNR and (b) speech distortion for different filter lengths and for different number of noise references (speech noise, $T_{60} = 300$ ms, batch version).



(a)

(b)

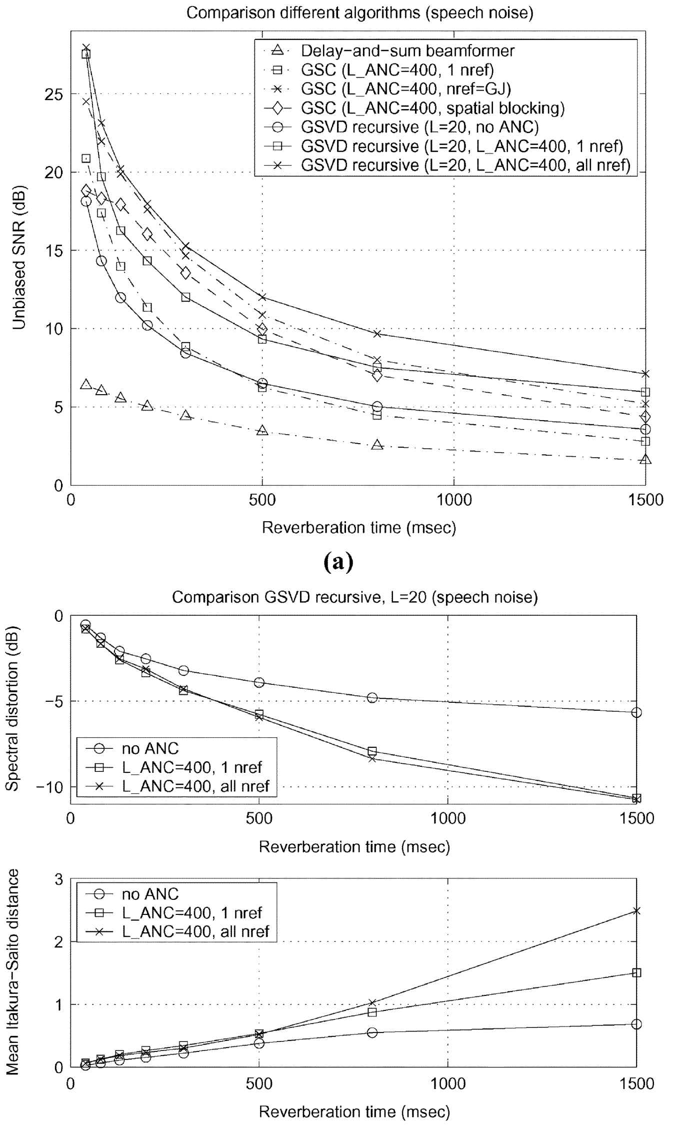
Fig. 12. Comparison of (a) unbiased SNR and (b) speech distortion for delay-and-sum, GSC and recursive GSVD-based optimal filtering technique with and without an ANC postprocessing stage (white noise, $L = 20, L_{\mathrm{ANC}} = 400$, no subsampling).

with the possibility to perform $s$ sweeps and $r$ GSVD-steps. Fig. 10 shows the unbiased SNR of the enhanced signal for different implementations of the recursive GSVD-updating algorithms and for a different number of sweeps and GSVD-steps. The noisy microphone signals have been constructed using a speech noise source at position 1 and for reverberation time $T_{60} = 300$ ms. The simulations have been performed with $L = 20$, without subsampling and without the ANC postprocessing stage. Fig. 10 shows that there is practically no difference in noise reduction performance between the "conventional" and the square root-free implementation. When performing more than one sweep, the SNR only marginally improves. When performing less than $M - 1$ GSVD-steps, the SNR gradually decreases.

### *F. Effect of the ANC Postprocessing Stage*

Fig. 11 investigates the effect of the ANC postprocessing stage on the noise reduction performance and the speech distortion for different filter lengths $L$ and $L_{\mathrm{ANC}}$ and for a different number of noise reference signals. The noisy microphone signals have been constructed using a speech noise source at position 1 and for reverberation time $T_{60} = 300$ ms, and simulations have been performed with the batch version of the GSVD-based optimal filtering technique.

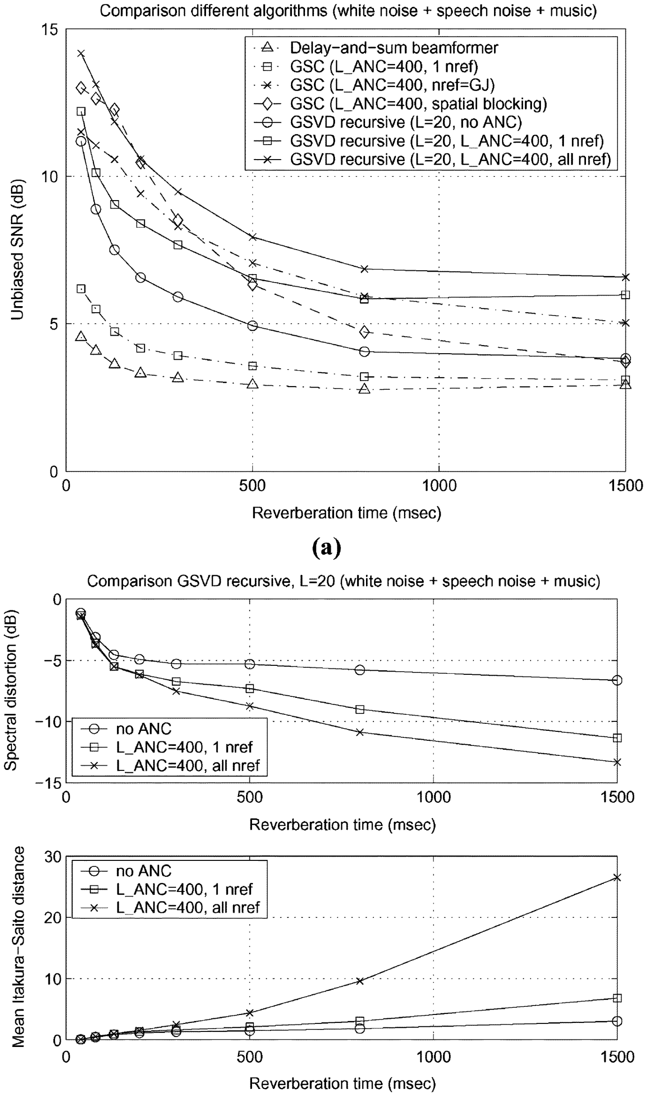
Fig. 11(a) shows that the SNR of the enhanced signal improves with increasing filter lengths $L$ and $L_{\mathrm{ANC}}$ and with increasing number of noise reference signals. In addition, this figure shows that the same noise reduction performance can be obtained either with large filter lengths $L$ without an ANC postprocessing stage or with short filter lengths $L$ and using anANC postprocessing stage. Since the total computational complexity is $\mathcal{O}(L^2) + \mathcal{O}(L_{\mathrm{ANC}})$, using short filter lengths $L$ with an ANC postprocessing stage gives rise to a lower computational complexity. The ANC postprocessing stage can therefore be used either for increasing the noise reduction performance or for com-



(a)

(b)

Fig. 13. Comparison of (a) unbiased SNR and (b) speech distortion for delay-and-sum, GSC and recursive GSVD-based optimal filtering technique with and without an ANC postprocessing stage (speech noise, $L = 20, L_{\mathrm{ANC}} = 400$, no subsampling).



(a)

(b)

Fig. 14. Comparison of (a) unbiased SNR and (b) speech distortion for delay-and-sum, GSC and recursive GSVD-based optimal filtering technique with and without an ANC postprocessing stage (three noise sources, $L = 20, L_{\mathrm{ANC}} = 400$, no subsampling).

putational complexity reduction without decreasing the performance.

Fig. 11(b) shows that the ANC postprocessing stage gives rise to a small increase in speech distortion (spectral distortion and Itakura–Saito distance), compared to not using the ANC postprocessing stage. However, speech distortion can be limited by using longer filter lengths $L$ (since signal leakage into the noise reference is then reduced) and longer filter lengths $L_{\mathrm{ANC}}$.

### G. Comparison for Simulated Acoustic Scenarios

In this section, the noise reduction performance and the speech distortion of the GSVD-based optimal filtering technique with and without an ANC postprocessing stage is compared with standard beamforming techniques for three simulated acoustic scenarios: a white noise source at position 1 (Fig. 12), a speech noise source at position 1 (Fig. 13) and three simultaneous noise sources (white+speech+music) at the three noise positions (Fig. 14). In all scenarios the noisy microphone signals are constructed such that the unbiased SNR of $y_0[k]$ is 0 dB. The following signal enhancement techniques are compared: DS-beamformer, GSC ($L_{\mathrm{ANC}} = 400$, 1 and all noise reference signals), GSC with spatial filter designed blocking matrix, recursive GSVD-based optimal filtering technique ($L = 20$, no subsampling) with and without an ANC postprocessing stage ($L_{\mathrm{ANC}} = 400$, 1 and all noise reference signals). This comparison is performed for different reverberation conditions. The scenario of the three simultaneous noise sources in a highly reverberant environment can actually be considered quite a good approximation of a diffuse noise field.

Figs. 12(a), 13(a), and 14(a) show that for low $T_{60}$ the SNR improvement of the GSC-based techniques is better than the SNR improvement of the GSVD-based optimal filtering technique without an ANC postprocessing stage. When adding the ANC postprocessing stage using all noise reference signals, the

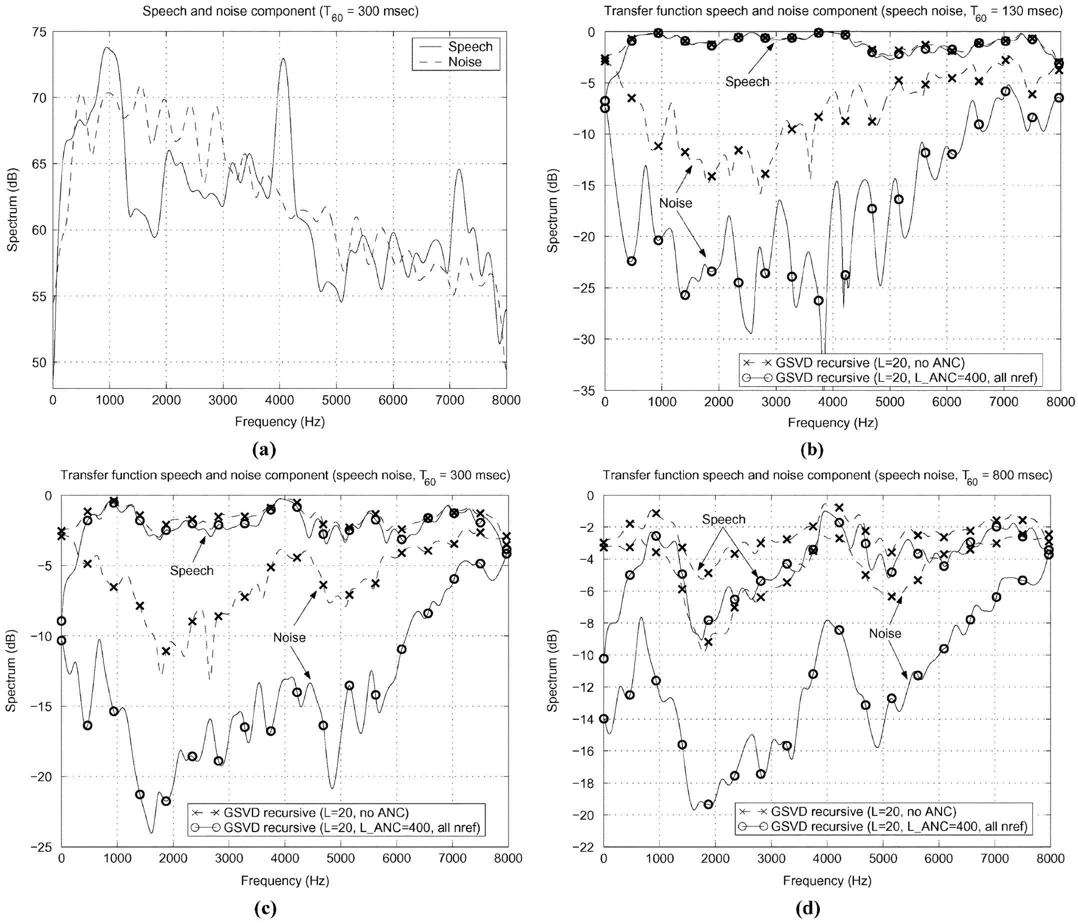


Fig. 15. (a) PSD of speech and noise components of first microphone signal ($T_{60} = 300$ ms), PTF of speech and noise components for recursive GSVD-based optimal filtering technique with and without an ANC postprocessing stage for (b) $T_{60} = 130$ ms, (c) $T_{60} = 300$ ms, (d) $T_{60} = 800$ ms (speech noise, $L = 20$, no subsampling, $L_{\rm ANC} = 400$, all noise references).

SNR improvement of the GSVD-based optimal filtering technique clearly outperforms the SNR improvement of the GSC (both Griffiths-Jim and spatial filter designed blocking matrix) for all reverberation times and all considered acoustic scenarios. In addition, the performance for the white noise source is better than for the speech noise source and the performance for a single noise source is better than for three simultaneous noise sources at different positions. This can be explained by the fact that the GSVD-based optimal filter can actually be decomposed as a spatial filtering operation, depending on the spatial characteristics (coherence) of the speech and the noise field, and a single-channel Wiener filter, depending on the spectral characteristics of the speech and the noise sources [19].

Figs. 12(b), 13(b), and 14(b) show the speech distortion (spectral distortion and Itakura–Saito distance) introduced by the recursive GSVD-based optimal filtering technique with and without an ANC postprocessing stage for different reverberation times. More speech distortion occurs for higher reverberation times and when using more noise reference signals. This can also be seen from Fig. 15, where the PSD and the PTF of the speech and the noise components have been plotted for three different reverberation times. For reverberation time $T_{60} = 300$ ms, Fig. 15(a) shows the PSD of the speech and the noise components of the first microphone signal and Fig. 15(c) shows the PTF for the speech and the noise components of the output signal of the recursive GSVD-based optimal filtering technique with and without an ANC postprocessing stage. As can be seen from Fig. 15(c), spectral distortion is limited, mainly occurs in frequency regions having a low SNR and is slightly higher when using an ANC postprocessing stage (which however also reduces a large amount of noise). Fig. 15(b) and (d) show the PTF's for reverberation times $T_{60} = 130$ ms and $T_{60} = 800$ ms. By comparing these figures, it is clear that more spectral distortion occurs for higher reverberation times (both in the GSVD-based optimal filter and in the ANC postprocessing stage).

### *H. Comparison for Real-Life Recording and Energy-Based VAD*

We have also compared the performance of the different speech enhancement algorithms for a real-life recording, performed in the Speech Lab at our department.[2] The reverberation time of the used room is approximately 500 ms. We have used a linear equi-spaced microphone array with $N = 3$ omnidirectional microphones (Sennheiser ME-102) and inter-microphone distance $d = 5$ cm. The speech source is located at approximately 1 m from the center of the microphone

[2] Sound files and results are available at http://www.esat.kuleuven.ac.be/~doclo/SA00061/audio.html

array at an angle of 110°, and three noise sources are located at different positions. We have used the same speech signal as for the simulated acoustic environments and speech noise from the NOISEX-92 database for all noise sources. We have used a nonperfect energy-based VAD [21] on the noisy microphone signal $y_0[k]$ and we have used a robust design procedure for the spatial filter designed blocking matrix [44], taking into account some gain and phase errors in the microphone characteristics. The parameters for the speech enhancement algorithms are the same as for the simulated acoustic environments (cf. Sections V-B and V-C).

The unbiased SNR of the first microphone signal is 0 dB, and the SNR's for the DS-beamformer, GSC with Griffiths-Jim blocking matrix and spatial filter designed blocking matrix are 0.46, 7.43, and 6.67 dB, respectively. For the recursive GSVD-based optimal filtering technique, the SNR is 6.25 dB, and when adding the ANC postprocessing stage using all noise reference signals, the SNR is 9.02 dB. The GSVD-based optimal filtering technique also introduces some amount of speech distortion, which increases when adding the ANC postprocessing stage.

## VI. Conclusion

In this paper, we have shown that the GSVD-based optimal filtering technique can be incorporated in a GSC-type structure, creating speech and noise reference signals and using these signals in an ANC postprocessing stage. This ANC postprocessing stage can either be used for increasing the noise reduction performance or for computational complexity reduction, since shorter filter lengths can be used for the GSVD-based optimal filter. In addition, the computational complexity of the GSVD-based optimal filtering technique can be drastically reduced by using recursive GSVD-updating algorithms and subsampling (in stationary acoustic environments) without any loss in performance.

Simulations have been performed for various acoustic scenarios, where both the SNR improvement and the speech distortion have been analyzed. These simulations show that the SNR improvement of the GSVD-based optimal filtering technique with an ANC postprocessing stage is better than the SNR improvement of standard fixed and adaptive beamforming techniques, while introducing an acceptable amount of speech distortion.

## Acknowledgment

The authors wish to thank the reviewers for their valuable comments and suggestions.

## References

[1] Y. Ephraim and D. Malah, "Speech enhancement using a minimun mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 443–445, Apr. 1985.

[2] S. Gannot, D. Burshtein, and E. Weinstein, "Iterative and sequential Kalman filter-based speech enhancement algorithms," *IEEE Trans. Speech Audio Processing*, vol. 6, pp. 373–385, Jul. 1998.

[3] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 4, pp. 251–266, Jul. 1995.

[4] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sørensen, "Reduction of broad-band noise in speech by truncated QSVD," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 439–448, Nov. 1995.

[5] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, pp. 4–24, Apr. 1988.

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

[7] S. Haykin, *Adaptive Filter Theory*, 4th ed, ser. Information and system sciences series. Englewood Cliffs, NJ: Prentice-Hall, 2001.

[8] J. Benesty *et al.*, "General Derivation of Frequency-Domain Adaptive Filtering," in *Advances in Network and Acoustic Echo Cancellation*. New York: Springer-Verlag, 2001, ch. 8, pp. 157–176.

[9] 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 (ICASSP)*, vol. 2, Albuquerque, NM, Apr. 1990, pp. 833–836.

[10] 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.

[11] S. Nordholm, I. Claesson, and B. Bengtsson, "Adaptive array noise suppression of handsfree speaker input in cars," *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 514–518, Nov. 1993.

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

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

[14] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 10, pp. 1365–1376, Oct. 1987.

[15] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Processing*, vol. 49, pp. 1614–1626, Aug. 2001.

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

[17] ——, "GSVD-based optimal filtering for multi-microphone speech enhancement," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. S. Brandstein and D. B. Ward, Eds. New York: Springer-Verlag, May 2001, ch. 6, pp. 111–132.

[18] S. Doclo, "Multimicrophone noise reduction and dereverberation techniques for speech applications," Ph.D. dissertation, Dept. Elect. Eng., Katholieke Univ. Leuven, Leuven, Belgium, May 2003.

[19] 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 Applications Signal Processing Audio Acoustics (WASPAA)*, New Paltz, NY, Oct. 2001, pp. 31–34.

[20] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*, First ed. Reading, MA: Addison Wesley, July 1991.

[21] S. Van Gerven and F. Xie, "A comparative study of speech detection methods," in *Proc. EUROSPEECH*, vol. 3, Rhodos, Greece, Sept. 1997, pp. 1095–1098.

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

[23] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: John Hopkins Univ. Press, 1996.

[24] F. T. Luk, "A parallel method for computing the generalized singular value decomposition," *J. Parallel Distrib. Comput.*, vol. 2, pp. 250–260, 1985.

[25] C. C. Paige, "Computing the generalized singular value decomposition," *SIAM J. Sci. Stat. Comput.*, vol. 7, pp. 1126–1146, 1986.

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

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

[28] ——, "Fast QRD-lattice-based unconstrained optimal filtering for acoustic noise reduction," *IEEE Trans. Speech Audio Process.*, to be published.

[29] A. Spriet, M. Moonen, and J. Wouters, “Stochastic gradient implementation of spatially pre-processed multi-channel Wiener filtering for noise reduction in hearing aids,” in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 4, Montreal, Canada, May 2004, pp. 57–60.

[30] S. Doclo, A. Spriet, and M. Moonen, “Efficient frequency-domain implementation of speech distortion weighted multi-channel Wiener filtering for noise reduction,” in *Proc. Eur. Signal Processing Conf. (EUSIPCO)*, Vienna, Austria., Sep. 2004, pp. 2007–2010.

[31] J. P. Charlier, M. Vanbegin, and P. Van Dooren, “On efficient implementations of Kogbetliantz’s algorithm for computing the singular value decomposition,” *Numerische Mathematik*, vol. 52, pp. 279–300, 1988.

[32] M. Moonen, “Jacobi-Type updating algorithms for signal processing, systems identification and control,” Ph.D. dissertation, Dept. Elect. Eng., Katholieke Univ. Leuven, Leuven, Belgium, 1990.

[33] M. Moonen, P. Van Dooren, and J. Vandewalle, “A singular value decomposition updating algorithm for subspace tracking,” *SIAM J. Matrix Anal. Applicat.*, vol. 13, no. 4, pp. 1015–1038, Oct. 1992.

[34] ——, “A systolic algorithm for QSVD updating,” *Signal Process.*, vol. 25, pp. 203–213, 1991.

[35] ——, “A systolic array for SVD updating,” *SIAM J. Matrix Anal. Appl.*, vol. 14, no. 2, pp. 353–371, 1993.

[36] M. Nilsson, S. D. Soli, and A. Sullivan, “Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise,” *J. Acoust. Soc. Amer.*, vol. 95, no. 2, pp. 1085–1099, Feb. 1994.

[37] A. Varga and H. J. M. Steeneken, “Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems,” *Speech Commun.*, vol. 12, no. 3, pp. 247–251, 1993.

[38] J. Allen and D. Berkley, “Image method for efficiently simulating small-room acoustics,” *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, Apr. 1979.

[39] P. M. Peterson, “Simulating the response of multiple microphones to a single acoustic source in a reverberant room,” *J. Acoust. Soc. Amer.*, vol. 80, no. 5, pp. 1527–1529, 1986.

[40] *The Master Handbook of Acoustics*, Fourth ed., McGraw Hill, New York, 2001.

[41] S. R. Quackenbush, T. P. Barnwell, and M. A. Clements, *Objective Measures of Speech Quality*. Englewood Cliffs, NJ: Prentice Hall, 1988.

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

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

[44] ——, “Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics,” *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2511–2526, Oct. 2003.

**Simon Doclo** (S’95–M’03) was born in Wilrijk, Belgium, in 1974. He received the M.Sc. degree in electrical engineering and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1997 and 2003, respectively.

Currently, he is a post-doctoral researcher with the Electrical Engineering Department, Katholieke Universiteit Leuven. His research interests are in microphone array processing for acoustic noise reduction, dereverberation and sound localization, adaptive filtering, speech enhancement, and hearing aid technology. He served as Guest Editor for a special issue on DSP in Hearing Aids and Cochlear Implants of the *EURASIP Journal on Applied Signal Processing*.

Dr. Doclo received the first prize “KVIV-Studentenprijzen” (with E. De Clippel) for his M.Sc. thesis in 1997, a Best Student Paper Award at the International Workshop on Acoustic Echo and Noise Control in 2001, and the EURASIP Signal Processing Best Paper Award 2003 (with M. Moonen). He was secretary of the IEEE Benelux Signal Processing Chapter from 1998 to 2002.

**Marc Moonen** (M’94) 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.

Since 2000, he has been an Associate Professor with the Electrical Engineering Department, Katholieke Universiteit Leuven, where he is currently heading a research team of 16 Ph.D. candidates and postdocs, working in the area of signal processing for digital communications, wireless communications, DSL, and audio signal processing. He is Editor-in-Chief for the *EURASIP Journal on Applied Signal Processing*, and a Member of the editorial board of *Integration, the VLSI Journal* and the *EURASIP Journal on Wireless Communications and Networking*.

Dr. Moonen received the 1994 KU Leuven Research Council Award and the 1997 Alcatel Bell (Belgium) Award (with P. Vandaele) and was a 1997 “Laureate of the Belgium Royal Academy of Science.” He was Chairman of the IEEE Benelux Signal Processing Chapter from 1998 to 2002, and is currently secretary/treasurer of European Association for Signal, Speech, and Image Processing (EURASIP). He is a Member of the editorial board of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II and the IEEE SIGNAL PROCESSING MAGAZINE.