Katholieke Universiteit Leuven

Departement Elektrotechniek ESAT-SISTA/TR 13-208

Low-rank approximation based multichannel Wiener filter algorithms for noise reduction with application in cochlear implants

Romain Serizel[1][2], Marc Moonen[2],
Bas Van Dijk[3] and Jan Wouters[4]

April 2014
Published in the IEEE Transactions on Audio, Speech and Language Processing
Vol. 22, No. 4, pp.785-799, April 2014

[1]Fondazione Bruno Kessler-IRST, Human Language Technology Research Unit, Via Sommarive 18, 38123 Povo (TN), Italy (serizel@fbk.eu)

[2]K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group SCD (SISTA) Kasteelpark Arenberg 10, 3001 Leuven, Belgium. This research work was carried out at the ESAT Laboratory of Katholieke Universiteit Leuven, in the frame of K.U.Leuven Research Council CoE EF/05/006 Optimization in Engineering (OPTEC), PFV/10/002 (OPTEC), Concerted Research Action GOA-MaNet, the Belgian Programme on Interuniversity Attraction Poles initiated by theBelgian Federal Science Policy Office IUAP P6/04 (DYSCO, 'Dynamical systems, control and optimization', 2007-2011), Research Project FWO nr. G.0600.08 ('Signal processing and network design for wireless acoustic sensor networks'), EC-FP6 project SIGNAL: 'Core Signal Processing Training Program'. The scientific responsibility is assumed by its authors.

[3]Cochlear CTCE, Schaliënhoevedreef 20, Building i-B, B-2800 Mechelen, Belgium

[4]Katholieke Universiteit Leuven, Department of Neurosciences, ExpORL, O. & N2, Herestraat 49/721, 3000 Leuven, Belgium, E-mail: Jan.Wouters@med.kuleuven.be

Abstract

This paper presents low-rank approximation based multichannel Wiener filter algorithms for noise reduction in speech plus noise scenarios, with application in cochlear implants. In a single speech source scenario, the frequency-domain autocorrelation matrix of the speech signal is often assumed to be a rank-1 matrix, which then allows to derive different rank-1 approximation based noise reduction filters. In practice, however, the rank of the autocorrelation matrix of the speech signal is usually greater than one.

Firstly, the link between the different rank-1 approximation based noise reduction filters and the original speech distortion weighted multichannel Wiener filter is investigated when the rank of the autocorrelation matrix of the speech signal is indeed greater than one.

Secondly, in low input signal-to-noise-ratio scenarios, due to noise non-stationarity, the estimation of the autocorrelation matrix of the speech signal can be problematic and the noise reduction filters can deliver unpredictable noise reduction performance. An eigenvalue decomposition based filter and a generalized eigenvalue decomposition based filter are introduced that include a more robust rank-1, or more generally rank-R, approximation of the autocorrelation matrix of the speech signal. These noise reduction filters are demonstrated to deliver a better noise reduction performance especially in low input signal-to-noise-ratio scenarios. The filters are especially usefull in cochlear implants, where more speech distortion and hence a more agressive noise reduction can be tolerated.

Low-rank approximation based multichannel Wiener filter algorithms for noise reduction with application in cochlear implants

Romain Serizel, Marc Moonen, Bas Van Dijk and Jan Wouters

***Abstract*—This paper presents low-rank approximation based multichannel Wiener filter algorithms for noise reduction in speech plus noise scenarios, with application in cochlear implants. In a single speech source scenario, the frequency-domain autocorrelation matrix of the speech signal is often assumed to be a rank-1 matrix, which then allows to derive different rank-1 approximation based noise reduction filters. In practice, however, the rank of the autocorrelation matrix of the speech signal is usually greater than one.**

Firstly, the link between the different rank-1 approximation based noise reduction filters and the original speech distortion weighted multichannel Wiener filter is investigated when the rank of the autocorrelation matrix of the speech signal is indeed greater than one.

Secondly, in low input signal-to-noise-ratio scenarios, due to noise non-stationarity, the estimation of the autocorrelation matrix of the speech signal can be problematic and the noise reduction filters can deliver unpredictable noise reduction performance. An eigenvalue decomposition based filter and a generalized eigenvalue decomposition based filter are introduced that include a more robust rank-1, or more generally rank-R, approximation of the autocorrelation matrix of the speech signal. These noise reduction filters are demonstrated to deliver a better noise reduction performance especially in low input signal-to-noise-ratio scenarios. The filters are especially usefull in cochlear implants, where more speech distortion and hence a more agressive noise reduction can be tolerated.

I. Introduction

A major challenge in cochlear implant (CI) design is to improve the speech understanding in noise for CI recipients [1] and so having an efficient front-end noise reduction (NR) is important. Therefore, several NR algorithms have been developed and tested with CI recipients [2], [3], [4]. Recent commercial CIs usually include multiple microphones and allow for multichannel adaptive NR algorithms, such as the BEAM^{TM} in the Cochlear Freedom^{TM} device, which have been shown to greatly improve speech understanding for CI recipients [5].

In general, CI recipients need a 10dB to 25dB higher signal-to-noise-ratio (SNR) than normal hearing subjects to achieve a similar speech understanding performance [6] but they can tolerate a much higher speech distortion (SD) [7]. This motivates the use of more aggressive NR strategies. The speech distortion weighted multichannel Wiener filter (SDW-MWF) has been developed to allow to tune multichannel Wiener filter (MWF)-based NR and perform a more aggressive NR by allowing more SD [8], [9], [10], [11]. In the case of a single speech source the SDW-MWF performance can sometimes be improved if the filters are reformulated based on the assumption that the frequency-domain autocorrelation matrix of the speech signal is a rank-1 matrix, leading to the so-called spatial-prediction MWF (SP-MWF) [12], [13] and the rank-1 MWF (R1-MWF) [14]. In this paper, the difference is investigated between the original SDW-MWF and these two rank-1 approximation based NR filters when the rank of autocorrelation matrix of the speech signal is actually greater than one.

All these NR algorithms rely on the estimation of the autocorrelation matrix of the speech signal, which is based on a rank-1 approximation with a so-called first column decomposition, as well as on the assumption that the (unknown) speech signal and the noise are uncorrelated and that these signals are stationary. In low input SNR scenarios, if these assumptions are violated, the autocorrelation matrix of the speech signal can be wrongly estimated and become non positive semi-definite. The SDW-MWF as well as the rank-1 approximation based filters can then deliver unpredictable NR performance. This paper proposes a solution to this problem that is to select an alternative rank-1 approximation based on an eigenvalue decomposition (EVD) [15], or a generalized eigenvalue decomposition (GEVD) [16], [17], [18], of the autocorrelation matrix of the speech signal.

These alternative NR filters are demonstrated to deliver a better NR performance especially in low input SNR scenarios and are especially usefull in cochlear implants, where more SD and hence a more agressive NR can be tolerated. The GEVD

Copyright (c) 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

This research work was carried out at the ESAT and ExpORL Laboratories of KU Leuven, in the frame of KU Leuven Research Council CoE EF/05/006 Optimization in Engineering (OPTEC) and PFV/10/002 (OPTEC), IWT Project 'Signal processing and automatic fitting for next generation cochlear implants', Concerted Research Action GOA-MaNet, the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office IUAP P7/19 (DYSCO, 'Dynamical systems, control and optimization', 2012-2016), EC-FP6 project SIGNAL: 'Core Signal Processing Training Program'. The scientific responsibility is assumed by its authors.

R. Serizel is with Fondazione Bruno Kessler-IRST, Human Language Technology Research Unit, Via Sommarive 18, 38123 Povo (TN), Italy (serizel@fbk.eu)

R. Serizel and M. Moonen are with KU Leuven, Department of Electrical Engineering, ESAT-Stadius Center for Dynamical Systems, Signal Processing and Data Analytics, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium.

B. Van Dijk is with Cochlear CTCE, Schaliënhoevedreef 20, Building i-B, B-2800 Mechelen, Belgium

J. Wouters is with KU Leuven, Department of Neurosciences, Experimental Otorhinolaryngology (ExpORL), O.& N2, Herestraat 49/721, B-3000 Leuven, Belgium.

based NR filter is also extended to a rank-R approximation based filter, in which the rank reduction is shown to be equivalent to tuning the NR to be more agressive. The rank-1 approximation based filter then indeed represents the extreme case with the most agressive NR. A performance comparison is provided between the original SDW-MWF, the EVD based NR filter and the GEVD based NR filter applied on both bilateral and binaural set-ups [19], [20], [21], [22], [23].

The signal model and the SDW-MWF are described in Section II. The so-called first column decomposition and how this provides an interpretation of the SDW-MWF versus the SP-MWF and the R1-MWF is described in Section III. The EVD based NR filter is introduced in Section IV. The GEVD based NR filter is presented in Section V and is extended to a rank-R approximation based filter in Section VI. The performance of the original SDW-MWF, the EVD based NR filter and the GEVD based NR filter are compared in Section VII. Finally, a summary of the paper and conclusions are provided in Section VIII.

II. Background and problem statement

A. Signal model

Let M be the number of microphones (channels). The frequency-domain signal $X_m(\omega)$ for microphone m has a speech component $X_{m,s}(\omega)$ and an additive noise component $X_{m,n}(\omega)$, i.e.:

$$X_m(\omega) = X_{m,s}(\omega) + X_{m,n}(\omega) \quad m \in \{1 \dots M\} \tag{1}$$

where $\omega = 2\pi f$ is the frequency-domain variable. For conciseness, ω will be omitted in all subsequent equations. Subscripts "s" and "n" will also be used to denote the "speech" and "noise" component of other quantities.

Signal model (1) holds for so-called "speech plus noise periods". There are also "noise only periods" (*i.e.*, speech pauses), during which only a noise component is observed.

In practice, in order to distinguish between "speech plus noise periods" and "noise only periods" it is necessary to use a voice activity detector (VAD). The performance of the VAD can affect the performance of the NR. For the time being, a perfect VAD is assumed.

The compound vector gathering all microphone signals is:

$$\mathbf{X} = [X_1 \dots X_M]^T \tag{2}$$

The autocorrelation matrix of the microphone signals in "speech plus noise periods", and of the speech component and the noise component of the microphone signals are given by:

$$\mathbf{R}_x = \mathbb{E}\{\mathbf{X}\mathbf{X}\} \tag{3}$$
$$\mathbf{R}_s = \mathbb{E}\{\mathbf{X}_s{\mathbf{X}_s}^H\} \tag{4}$$
$$\mathbf{R}_n = \mathbb{E}\{\mathbf{X}_n{\mathbf{X}_n}^H\} \tag{5}$$

where ^{H} denotes the Hermitian transpose. $\mathbf{R}_n$ can be estimated during "noise only periods" and $\mathbf{R}_x$ can be estimated during "speech plus noise periods". If the speech and noise signals are assumed to be uncorrelated and if the noise signal is stationary, $\mathbf{R}_s$ can be estimated by using:

$$\mathbf{R}_s = \mathbf{R}_x - \mathbf{R}_n \tag{6}$$

In practice, the autocorrelation matrices are estimated recursively. The estimate of the autocorrelation matrix of the microphone signals is updated during "speech plus noise periods", using:

$$\tilde{\mathbf{R}}_x = \lambda\tilde{\mathbf{R}}_x + (1-\lambda)\mathbf{X}\mathbf{X}^H \tag{7}$$

where $\lambda \in [0,\ 1]$ is an exponential forgetting factor that depends on the number of past samples to be taken into account. Here $\lambda = 1 - \frac{1}{N_{\text{frame}}}$ with N_{frame} set so that the forgetting time is about 1 second. This clearly exceeds the spectral stationarity of speech signals (around 20ms) but not necessarily the spatial stationarity of the sources.

The estimate of the autocorrelation matrix of the noise component of the microphone signals is updated similarly during "noise only periods", using:

$$\tilde{\mathbf{R}}_n = \lambda\tilde{\mathbf{R}}_n + (1-\lambda)\mathbf{X}\mathbf{X}^H \tag{8}$$
$$= \lambda\tilde{\mathbf{R}}_n + (1-\lambda)\mathbf{X}_n{\mathbf{X}_n}^H \tag{9}$$

The estimate of the autocorrelation matrix of the speech component of the microphone signals is then given by:

$$\tilde{\mathbf{R}}_s = \tilde{\mathbf{R}}_x - \tilde{\mathbf{R}}_n \tag{10}$$

It is noted that in the sequel, NR filters are specified as functions of $\mathbf{R}_x$, $\mathbf{R}_n$ and/or $\mathbf{R}_s$, whereas in practice these matrices are replaced by their estimated versions $\tilde{\mathbf{R}}_x$, $\tilde{\mathbf{R}}_n$ and/or $\tilde{\mathbf{R}}_s$ (or modifications thereof).

B. MWF-based Noise Reduction

An MWF $\mathbf{W} = [W_1 \dots W_M]^T$ will be designed and applied to the microphone signals, which minimizes a Mean Squared Error (MSE) criterion:

$$J_{\text{MWF}} = \mathbb{E}\{|E|^2\} \tag{11}$$

where $\mathbb{E}$ is the expectation operator and E is an error signal to be defined next, depending on the scheme applied. The filter output signal Z is defined as:

$$Z = \mathbf{W}^H\mathbf{X} \tag{12}$$

The desired signal for the MWF is arbitrarily chosen to be the (unknown) speech component of the first microphone signal ($m = 1$). This can be written as:

$$D_{\text{MWF}} = \mathbf{e}_1^H\mathbf{X}_s \tag{13}$$

where $\mathbf{e}_1$ is an all-zero vector except for a one in the first position.

The MWF aims to minimize the squared distance between the filtered microphone signal (12) and the desired signal (13). The corresponding MSE criterion is:

$$J_{\text{MWF}} = \mathbb{E}\{|\mathbf{W}^H\mathbf{X} - \mathbf{e}_1^H\mathbf{X}_s|^2\} \tag{14}$$

which is equivalent to:

$$J_{\text{MWF}} = \mathbf{W}^H\mathbf{R}_s\mathbf{W} - \mathbf{W}^H\mathbf{R}_s\mathbf{e}_1 - \mathbf{e}_1^H\mathbf{R}_s\mathbf{W} + \mathbf{e}_1^H\mathbf{R}_s\mathbf{e}_1 + \mathbf{W}^H\mathbf{R}_n\mathbf{W} \tag{15}$$

The MWF solution is given as:

$$\mathbf{W}_{\text{MWF}} = (\mathbf{R}_s + \mathbf{R}_n)^{-1}\mathbf{R}_s\mathbf{e}_1 \tag{16}$$

The SDW-MWF has been proposed to provide an explicit trade-off between the NR and the SD [8], [9], [10], [11]. Changing the optimization problem to a constrained optimization problem, the MSE criterion effectively becomes:

$$J_{\mathrm{SDW-MWF}} = \mathbb{E}\{|\mathbf{W}^H\mathbf{X}^s - \mathbf{e}_1^H\mathbf{X}^s|^2\} + \mu\mathbb{E}\{|\mathbf{W}^H\mathbf{X}^n|^2\} \quad (17)$$

where μ is a trade-off parameter. The SDW-MWF solution is then given as:

$$\mathbf{W}_{\mathrm{SDW-MWF}} = (\mathbf{R}_s + \mu\mathbf{R}_n)^{-1}\mathbf{R}_s\mathbf{e}_1 \quad (18)$$

In a single speech source scenario, the autocorrelation matrix of the speech component of the microphone signals $\mathbf{R}_s$ is often assumed to be a rank-1 matrix ($rank(\mathbf{R}_s) = 1$) and can then be rewritten as:

$$\mathbf{R}_s = P_s\mathbf{A}\mathbf{A}^H \quad (19)$$

where P_s is the power of the speech source signal and $\mathbf{A}$ is the M-dimensional steering vector, containing the acoustic transfer functions from the speech source position to the microphones (including the microphone characteristics).

Based on this rank-1 assumption it is possible to derive the so-called spatial-prediction MWF (SP-MWF) [12], [13]:

$$\mathbf{W}_{\mathrm{SP-MWF}} = \mathbf{R}_n^{-1}\mathbf{R}_s\mathbf{e}_1\frac{\mathbf{e}_1^H\mathbf{R}_s\mathbf{e}_1}{\mu\mathbf{e}_1^H\mathbf{R}_s\mathbf{e}_1 + \mathrm{Tr}\{\mathbf{R}_n^{-1}\mathbf{R}_s\mathbf{e}_1\mathbf{e}_1^H\mathbf{R}_s\}} \quad (20)$$

and the rank-1 MWF (R1-MWF) [14]:

$$\mathbf{W}_{\mathrm{R1-MWF}} = \mathbf{R}_n^{-1}\mathbf{R}_s\mathbf{e}_1\frac{1}{\mu + \mathrm{Tr}\{\mathbf{R}_n^{-1}\mathbf{R}_s\}} \quad (21)$$

The filters (18), (20) and (21) are fully equivalent if $rank(\mathbf{R}_s) = 1$. In practice, however, (19) may not hold, i.e., $rank(\mathbf{R}_s) > 1$ even for a single speech source scenario and then (18), (20) and (21) are different filters.

III. First column decomposition

When $rank(\mathbf{R}_s) > 1$ the matrix $\mathbf{R}_s$ can be decomposed as:

$$\mathbf{R}_s = \mathbf{R}_{s_{r1}} + \mathbf{R}_Z \quad (22)$$

where $\mathbf{R}_{s_{r1}}$ is a rank-1 approximation of $\mathbf{R}_s$ and $\mathbf{R}_Z$ is a "remainder" matrix. The decomposition is not unique and so several choices for $\mathbf{R}_{s_{r1}}$ can be considered.

The most obvious choice for $\mathbf{R}_{s_{r1}}$ is a rank-1 extension of the first column and row of $\mathbf{R}_s$, *i.e.*:

$$\mathbf{R}_s = \underbrace{\mathbf{d}\mathbf{d}^H\sigma_{1,1}}_{\mathbf{R}_{s_{r1}}} + \underbrace{\left[\begin{array}{c|ccc} 0 & 0 & \cdots & 0 \\ \hline 0 & x & \cdots & x \\ \vdots & \vdots & & \vdots \\ 0 & x & \cdots & x \end{array}\right]}_{\mathbf{R}_Z} \quad (23)$$

where

$$\sigma_{i,j} = [\mathbf{R}_s]_{i,j} \quad (24)$$

$$\mathbf{d} = [1 \ \frac{\sigma_{2,1}}{\sigma_{1,1}} \dots \frac{\sigma_{1,N}}{\sigma_{1,1}}]^T \quad (25)$$

and $\sigma_{1,1}$ is the speech power in microphone 1. This decomposition will be referred to as the "first column decomposition". It allows to pinpoint the differences between the filters (18), (20) and (21) whenever $rank(\mathbf{R}_s) > 1$. This decomposition has also been exploited in [24].

It is noted that:

$$\mathbf{R}_s\mathbf{e}_1 = \mathbf{R}_{s_{r1}}\mathbf{e}_1 + \underbrace{\mathbf{R}_Z\mathbf{e}_1}_{=\mathbf{0}} \quad (26)$$

which means that the (rightmost) "desired signal part" $\mathbf{R}_s\mathbf{e}_1$ in (18), (20) and (21) can be (obviously) replaced by the "rank-1 approximation desired signal part" $\mathbf{R}_{s_{r1}}\mathbf{e}_1$. The difference between the filters (18), (20) and (21) then effectively depends on how $\mathbf{R}_Z$ is treated, as will be explained next. Note that when $rank(\mathbf{R}_s) = 1$, then $\mathbf{R}_Z = 0$ and so it is again seen that the filters are fully equivalent.

A. SDW-MWF

Plugging (23) into the SDW-MWF formula (18) leads to:

$$\mathbf{W}_{\mathrm{SDW-MWF}} = (\mathbf{R}_{s_{r1}} + \mu(\mathbf{R}_n + \frac{1}{\mu}\mathbf{R}_Z))^{-1}\mathbf{R}_{s_{r1}}\mathbf{e}_1 \quad (27)$$

This means that in the SDW-MWF (18) $\mathbf{R}_s$ can be replaced by $\mathbf{R}_{s_{r1}}$ and then the remainder matrix $\mathbf{R}_Z$ is effectively treated as noise (up to a scaling with $\frac{1}{\mu}$).

To avoid the scaling with $\frac{1}{\mu}$ an alternative approach is to start from the MSE criterion (15). Plugging (23) into (15), merging $\mathbf{R}_Z$ with the noise and (only then) introducing the trade-off factor μ, leads to:

$$\begin{aligned} J^{\diamond}_{\mathrm{SDW-MWF}} = {} & \mathbf{W}^H\mathbf{R}_{s_{r1}}\mathbf{W} - \mathbf{W}^H\mathbf{R}_{s_{r1}}\mathbf{e}_1 - \mathbf{e}_1^H\mathbf{R}_{s_{r1}}\mathbf{W} \\ & + \mathbf{e}_1^H\mathbf{R}_{s_{r1}}\mathbf{e}_1 + \mu(\mathbf{W}^H\mathbf{R}_Z\mathbf{W} + \mathbf{W}^H\mathbf{R}_n\mathbf{W}) \end{aligned} \quad (28)$$

where superscript $^{\diamond}$ is used to denote the alternative formulation where the trade-off factor μ is introduced after $\mathbf{R}_Z$ and $\mathbf{R}_n$ are merged.

The filter minimizing (28) is then:

$$\mathbf{W}^{\diamond}_{\mathrm{SDW-MWF}} = (\mathbf{R}_{s_{r1}} + \mu(\mathbf{R}_n + \mathbf{R}_Z))^{-1}\mathbf{R}_{s_{r1}}\mathbf{e}_1 \quad (29)$$

This again means that in the SDW-MWF (18) $\mathbf{R}_s$ is replaced by $\mathbf{R}_{s_{r1}}$ and then the remainder matrix $\mathbf{R}_Z$ is effectively returned to noise, i.e, $\mathbf{R}_n$ is replaced by $\mathbf{R}_n + \mathbf{R}_Z$. The initial speech plus noise decomposition $\mathbf{R}_x = \mathbf{R}_s + \mathbf{R}_n$ is then effectively reshuffled into $\mathbf{R}_x = \mathbf{R}_{s_{r1}} + (\mathbf{R}_n + \mathbf{R}_Z)$.

It is seen that (27) and (29) only differ in the weighting applied to $\mathbf{R}_Z$. While (27) is fully equivalent to (18), (29) adopts a weighting that is intuitively more appealing if $\mathbf{R}_Z$ is considered to be a noise contribution. However, $\mathbf{R}_Z$ can come not only from noise estimate leaking into the speech estimate but also from various factors such as VAD errors, over/understimation of the noise during speech periods, correlation between speech and noise... Therefore, it is unclear which of noise weighting strategies (27) and (29) is the more appropriate.

For $\mu = 1$, filters (27) and (29) are obviously equivalent.

B. SP-MWF

Plugging (23) into the SP-MWF formula (20) leads to:

$$\mathbf{W}_{\mathrm{SP-MWF}} = \mathbf{R}_n^{-1}\mathbf{R}_{s_{r1}}\mathbf{e}_1\frac{1}{\mu + \mathrm{Tr}\{\mathbf{R}_n^{-1}\mathbf{R}_{s_{r1}}\}} \tag{30}$$

and

$$\mathbf{W}_{\mathrm{SP-MWF}} = (\mathbf{R}_{s_{r1}} + \mu\mathbf{R}_n)^{-1}\mathbf{R}_{s_{r1}}\mathbf{e}_1 \tag{31}$$

This means that the SP-MWF effectively corresponds to the SDW-MWF (18) where $\mathbf{R}_s$ is replaced by $\mathbf{R}_{s_{r1}}$ and the remainder matrix $\mathbf{R}_Z$ is simply ignored.

C. R1-MWF

Plugging (23) into the R1-MWF formula (21) leads to:

$$\mathbf{W}_{\mathrm{R1-MWF}} = \mathbf{R}_n^{-1}\mathbf{R}_{s_{r1}}\mathbf{e}_1\frac{1}{\mu + \mathrm{Tr}\{\mathbf{R}_n^{-1}(\mathbf{R}_{s_{r1}} + \mathbf{R}_Z)\}} \tag{32}$$

and

$$\mathbf{W}_{\mathrm{R1-MWF}} = (\mathbf{R}_{s_{r1}} + \bar{\mu}\mathbf{R}_n)^{-1}\mathbf{R}_{s_{r1}}\mathbf{e}_1 \tag{33}$$

where

$$\bar{\mu} = \mu + \mathrm{Tr}\{\mathbf{R}_n^{-1}\mathbf{R}_Z\} \neq \mu \tag{34}$$

By comparing (33) with (27) and (31), it is seen that the R1-MWF represents an intermediate approach between the SDW-MWF and the SP-MWF. Indeed, in the R1-MWF the remainder matrix $\mathbf{R}_Z$ is ignored in the spatial filter ($\mathbf{R}_n^{-1}\mathbf{R}_{s_{r1}}\mathbf{e}_1$) as it is also the case for the SP-MWF filter (see (30) and (32)). The remainder matrix $\mathbf{R}_Z$ changes the trade-off parameter from μ to $\bar{\mu}$ which effectively changes the spectral postfilter in (30) and (32). If $\mathbf{R}_Z$ is positive semi-definite, $\bar{\mu} > \mu$ which corresponds to putting a higher weight on the noise. This is similar to $\mathbf{R}_Z$ being treated as noise in the SDW-MWF case.

D. Speech autocorrelation matrix estimation

In low input SNR scenarios it is observed that:

$$\tilde{\mathbf{R}}_x \approx \tilde{\mathbf{R}}_n \tag{35}$$

and then, especially so if the noise is non-stationary, the estimated $\tilde{\mathbf{R}}_s = \tilde{\mathbf{R}}_x - \tilde{\mathbf{R}}_n$ can loose its positive semi-definiteness, which is problematic and indeed has been observed to lead to unpredictable NR performance. The first column decomposition in particular suffers from this estimation problem where then the estimated speech power in microphone 1 ($\tilde{\sigma}_{1,1} \triangleq [\tilde{\mathbf{R}}_s]_{1,1} = [\tilde{\mathbf{R}}_x]_{1,1} - [\tilde{\mathbf{R}}_n]_{1,1}$) can become negative (which is meaningless) so that $\tilde{\mathbf{R}}_{s_{r1}}$ is negative semi-definite and hence the desired signal is ill-defined. This explains why the first column decomposition based filters often provide poor NR performance in low input SNR scenarios. In addition, if $\mathbf{R}_Z$ is non positive definite, then the $\bar{\mu}$ in the R1-MWF (33) may be spuriously decreased instead of increased compared to the μ in the SP-MWF (30).

IV. EVD BASED NR FILTERS

An alternative to the first column decomposition based rank-1 approximation is a rank-1 approximation based on an EVD of $\mathbf{R}_s$, as also introduced in [15]:

$$\mathbf{R}_s = \underbrace{\mathbf{d}_{\max}\mathbf{d}_{\max}^H\lambda_{\max}}_{\mathbf{R}_{s_{r1}}} + \mathbf{R}_Z \tag{36}$$

where $\lambda_{\max}$ is $\mathbf{R}_s$'s (real-valued) largest eigenvalue, $\mathbf{d}_{\max}$ is the corresponding normalized eigenvector and $\mathbf{R}_Z$ is again a remainder matrix. When $rank(\mathbf{R}_s) = 1$, then $\mathbf{R}_Z = 0$ and $\mathbf{R}_{s_{r1}}$ is the same as in the first column decomposition. When $rank(\mathbf{R}_s) > 1$, then the rank-1 estimated part $\tilde{\mathbf{R}}_{s_{r1}}$ is positive semi-definite if the dominant eigenvalue of $\tilde{\mathbf{R}}_s$ is positive (which is more likely than the first diagonal element $\tilde{\sigma}_{1,1}$ of $\tilde{\mathbf{R}}_s$ being positive as needed in the first column decomposition approach).

It is noted that:

$$\mathbf{R}_s\mathbf{f}_1 = \mathbf{R}_{s_{r1}}\mathbf{f}_1 + \underbrace{\mathbf{R}_Z\mathbf{f}_1}_{=0} = \mathbf{R}_{s_{r1}}\mathbf{e}_1 \tag{37}$$

to be compared to (26), where

$$\mathbf{f}_1 = \mathbf{d}_{\max}\mathbf{d}_{\max}(1)^* \tag{38}$$

with $\mathbf{d}_{\max}(1)$ is the first element of $\mathbf{d}_{\max}$.

An analysis similar to the analysis for the first column decomposition in Section III can then be done where $\mathbf{R}_s$ is replaced by the rank-1 approximation $\mathbf{R}_{s_{r1}}$ and the remainder matrix $\mathbf{R}_Z$ is either treated as noise or ignored. Equivalently, one can start from a modified MSE criterion where, compared to (17), the (arbitrary) $\mathbf{e}_1$ is replaced by $\mathbf{f}_1$:

$$J_{\mathrm{EVD-SDW-MWF}} = \mathbb{E}\{|\mathbf{W}^H\mathbf{X}^s - \mathbf{f}_1^H\mathbf{X}^s|^2\} + \mu\mathbb{E}\{|\mathbf{W}^H\mathbf{X}^n|^2\} \tag{39}$$

Replacing the desired signal $\mathbf{e}_1\mathbf{X}_s$ by $\mathbf{f}_1\mathbf{X}_s$ is indeed equivalent to replacing $\mathbf{R}_s$ by the EVD based $\mathbf{R}_{s_{r1}}$ as demonstrated by (37).

A. EVD-SDW-MWF

The filter minimizing (39) is given as:

$$\mathbf{W}_{\mathrm{EVD-SDW-MWF}} = (\mathbf{R}_s + \mu\mathbf{R}_n)^{-1}\mathbf{R}_s\mathbf{f}_1 \tag{40}$$

Plugging (36) and (37) into (40) leads to:

$$\mathbf{W}_{\mathrm{EVD-SDW-MWF}} = (\mathbf{R}_{s_{r1}} + \mu(\mathbf{R}_n + \frac{1}{\mu}\mathbf{R}_Z))^{-1}\mathbf{R}_{s_{r1}}\mathbf{e}_1 \tag{41}$$

This means that in the SDW-MWF (18) $\mathbf{R}_s$ is replaced by the EVD based $\mathbf{R}_{s_{r1}}$ and the remainder matrix $\mathbf{R}_Z$ is effectively treated as noise (up to a scaling with $\frac{1}{\mu}$).

To avoid the scaling with $\frac{1}{\mu}$, the same alternative derivation as for (29) can be applied leading to:

$$\mathbf{W}^{\diamond}_{\mathrm{EVD-SDW-MWF}} = (\mathbf{R}_{s_{r1}} + \mu(\mathbf{R}_n + \mathbf{R}_Z))^{-1}\mathbf{R}_{s_{r1}}\mathbf{e}_1 \tag{42}$$

This means that in the SDW-MWF (18) the desired signal vector $\mathbf{R}_s$ is replaced by the EVD based $\mathbf{R}_{s_{r1}}$ and the remainder matrix $\mathbf{R}_Z$ is effectively returned to noise.

B. EVD-SP-MWF

Based on the MSE criterion (39) it is also possible to derive the SP-MWF:

$$\mathbf{W}_{\mathrm{EVD-SP-MWF}} = \mathbf{R}_n^{-1}\mathbf{R}_s\mathbf{f}_1 \frac{\mathbf{f}_1^H\mathbf{R}_s\mathbf{f}_1}{\mu\mathbf{f}_1^H\mathbf{R}_s\mathbf{f}_1 + \mathrm{Tr}\{\mathbf{R}_n^{-1}\mathbf{R}_s\mathbf{f}_1\mathbf{f}_1^H\mathbf{R}_s\}} \tag{43}$$

Plugging (36) into the EVD-SP-MWF formula (43) leads to:

$$\mathbf{W}_{\mathrm{EVD-SP-MWF}} = \mathbf{R}_n^{-1}\mathbf{R}_{s_{r1}}\mathbf{f}_1 \frac{1}{\mu + \mathrm{Tr}\{\mathbf{R}_n^{-1}\mathbf{R}_{s_{r1}}\}} \tag{44}$$

and

$$\mathbf{W}_{\mathrm{EVD-SP-MWF}} = (\mathbf{R}_{s_{r1}} + \mu\mathbf{R}_n)^{-1}\mathbf{R}_{s_{r1}}\mathbf{e}_1 \tag{45}$$

This means that in the SDW-MWF (18) $\mathbf{R}_s$ is replaced by the EVD based $\mathbf{R}_{s_{r1}}$ and the remainder matrix $\mathbf{R}_Z$ is simply ignored.

The EVD-R1-MWF derivation is omitted for conciseness.

C. A matrix approximation based derivation of EVD-SDW-MWF and EVD-SP-MWF

From a given $\mathbf{R}_x$ and $\mathbf{R}_n$ the autocorrelation matrix of the speech component can be computed as $\mathbf{R}_s = \mathbf{R}_x - \mathbf{R}_n$ and these matrices can be plugged in the SDW-MWF formula (18). It has been mentioned in Section III-D that this may result in poor NR performance, in particular in low input SNR scenarios, where then the estimated $\tilde{\mathbf{R}}_s$ is oftentimes indefinite rather than positive semi-definite. To avoid this, an alternative approach can be followed where first a better autocorrelation matrix of the speech component is computed (call it $\mathbf{R}_{s_{r1}}$) together with a better autocorrelation matrix of the noise component (call it $\mathbf{R}_{n_{r1}}$). To compute the $\{\mathbf{R}_{s_{r1}}, \mathbf{R}_{n_{r1}}\}$, a matrix approximation problem is formulated, specifying that $\mathbf{R}_{n_{r1}}$ should provide a good approximation to the given $\mathbf{R}_n$, while $(\mathbf{R}_{n_{r1}} + \mathbf{R}_{s_{r1}})$ should provide a good approximation to the given $\mathbf{R}_x$. In addition, "a priori knowledge" is incorporated, namely that $\mathbf{R}_{s_{r1}}$ should be a rank-1 matrix. The so obtained $\{\mathbf{R}_{s_{r1}}, \mathbf{R}_{n_{r1}}\}$ can then be used in the SDW-MWF formula (18). It is demonstrated in this section that this approach indeed leads to the EVD-SDW-MWF and EVD-SP-MWF, and so provides an alternative interpretation of these filters.

It is noted that the rank-1 condition for the autocorrelation matrix of the speech component is generalized to a rank-K condition in section VI. The rank condition is then also seen to be a crucial ingredient, where in the extreme case of $K = M$ (i.e., effectively no rank condition) the solution to the matrix approximation problem is merely $\{\mathbf{R}_x - \mathbf{R}_n, \mathbf{R}_n\}$, i.e., the autocorrelation matrices remain unchanged.

The $\{\mathbf{R}_{s_{r1}}, \mathbf{R}_{n_{r1}}\}$ should minimize the following criterion:

$$J_{\mathrm{r1}} = \alpha||\mathbf{R}_x - (\mathbf{R}_{n_{r1}} + \mathbf{R}_{s_{r1}})||_F^2 + (1-\alpha)||\mathbf{R}_n - \mathbf{R}_{n_{r1}}||_F^2 \tag{46}$$

with $||.||_F$ the Frobenius norm. Here, $\mathbf{R}_{n_{r1}}$ and $\mathbf{R}_{s_{r1}}$ are positive semi-definite matrices and $\mathbf{R}_{s_{r1}}$ is a rank-1 matrix. The two approximations may be given a different weight, i.e., α and $(1-\alpha)$, where α is a constant $(0 < \alpha < 1)$. In the case of estimated autocorrelation matrices, for instance, it may make sense to give a smaller weight to the approximation of the noise autocorrelation matrix (i.e. $\alpha > 0.5$), as this is estimated in older (hence possibly more outdated) "noise only" frames whenever a noise reduction is computed in a "speech plus noise" frame.

It is easy to check that when an optimal $\mathbf{R}_{s_{r1}}$ is given, the optimal solution for $\mathbf{R}_{n_{r1}}$ is:

$$\mathbf{R}_{n_{r1}} = \alpha(\mathbf{R}_x - \mathbf{R}_{s_{r1}}) + (1-\alpha)\mathbf{R}_n \tag{47}$$

with the positive semi-definiteness of $\mathbf{R}_{n_{r1}}$ yet to be checked. As $\mathbf{R}_n$ is positive semi-definite by construction, it remains to check if $\mathbf{R}_x - \mathbf{R}_{s_{r1}}$ is positive semi-definite (see below).

The $\mathbf{R}_{n_{r1}}$ can then be eliminated from the optimization problem by plugging (47) into (46). Therefore, after some simple manipulation, $\mathbf{R}_{s_{r1}}$ should minimize the following criterion:

$$J_{\mathrm{s_{r1}}} = \alpha(1-\alpha)||\mathbf{R}_x - \mathbf{R}_n - \mathbf{R}_{s_{r1}}||_F^2 \tag{48}$$

The optimal solution is then known to be:

$$\mathbf{R}_{s_{r1}} = \mathbf{d}_{\max}\mathbf{d}_{\max}^H \max(\lambda_{\max}, 0) \tag{49}$$

as defined in (36) (assuming $\lambda_{\max}$ is non-negative). For this $\mathbf{R}_{s_{r1}}$, the matrix $\mathbf{R}_x - \mathbf{R}_{s_{r1}}$ is indeed seen to be positive semi-definite, as required.

Once $\mathbf{R}_{s_{r1}}$ is defined according to (49), $\mathbf{R}_{n_{r1}}$ is computed based on (47). Two extreme cases can then be considered, as follows:

- If $\alpha \rightarrow 1$, which means that $\mathbf{R}_{n_{r1}} + \mathbf{R}_{s_{r1}}$ is to give the best possible approximation to $\mathbf{R}_x$ (first term in the original optimization function (46)), then $\mathbf{R}_{n_{r1}} = \mathbf{R}_x - \mathbf{R}_{s_{r1}} = \mathbf{R}_n + \mathbf{R}_Z$ with $\mathbf{R}_Z$ defined in (36). By replacing $\{\mathbf{R}_s, \mathbf{R}_n\}$ by this $\{\mathbf{R}_{s_{r1}}, \mathbf{R}_{n_{r1}}\}$ in formula (18), the EVD-SDW-MWF formula (42) is obtained.
- If $\alpha \rightarrow 0$, which means that $\mathbf{R}_{n_{r1}}$ is to give the best possible approximation to $\mathbf{R}_n$ (second term in the original optimization function (46)), then $\mathbf{R}_{n_{r1}} = \mathbf{R}_n$. By replacing $\{\mathbf{R}_s, \mathbf{R}_n\}$ by this $\{\mathbf{R}_{s_{r1}}, \mathbf{R}_{n_{r1}}\}$ in formula (18), the EVD-SP-MWF formula (45) is obtained.

V. GEVD BASED NR FILTERS

A second alternative to the first column decomposition based rank-1 approximation is a rank-1 approximation based on the GEVD [16], [17], [18] of the matrix pencil $\{\mathbf{R}_x, \mathbf{R}_n\}$:

$$\begin{aligned} \mathbf{R}_n &= \mathbf{Q}\mathbf{\Sigma}_n\mathbf{Q}^H \\ \mathbf{R}_x &= \mathbf{Q}\mathbf{\Sigma}_x\mathbf{Q}^H \\ \Rightarrow \mathbf{R}_n^{-1}\mathbf{R}_x &= \mathbf{Q}^{-H}(\mathbf{\Sigma}_n^{-1}\mathbf{\Sigma}_x)\mathbf{Q}^H = \mathbf{Q}^{-H}\mathbf{\Sigma}\mathbf{Q}^H \end{aligned} \tag{50}$$

where $\mathbf{Q}$ is an invertible matrix, the columns of which are normalized and define the generalized eigenvectors. $\mathbf{\Sigma}_x$, $\mathbf{\Sigma}_n$ and $\mathbf{\Sigma}$ are real-valued diagonal matrices with $\mathbf{\Sigma}_x = \mathrm{diag}\{\sigma_{x_1} \cdots \sigma_{x_M}\}$, $\mathbf{\Sigma}_n = \mathrm{diag}\{\sigma_{n_1} \cdots \sigma_{n_M}\}$ and $\mathbf{\Sigma} = \mathrm{diag}\{\frac{\sigma_{x_1}}{\sigma_{n_1}} \cdots \frac{\sigma_{x_M}}{\sigma_{n_M}}\}$ (with ordering $\frac{\sigma_{x_1}}{\sigma_{n_1}} \geq \frac{\sigma_{x_2}}{\sigma_{n_2}} \geq \cdots \geq \frac{\sigma_{x_M}}{\sigma_{n_M}}$) defining the generalized eigenvalues.

The $\mathbf{R}_s$ is then obtained as:

$$\mathbf{R}_s = \mathbf{R}_x - \mathbf{R}_n = \mathbf{Q}(\underbrace{\mathbf{\Sigma}_x - \mathbf{\Sigma}_n}_{\mathbf{\Sigma}_s})\mathbf{Q}^H \tag{51}$$

where $\mathbf{\Sigma}_s = \text{diag}\{\sigma_{s_1} \cdots \sigma_{s_M}\}$ and $\text{SNR}_i = \frac{\sigma_{s_i}}{\sigma_{n_i}} = \frac{\sigma_{x_i}}{\sigma_{n_i}} - 1$ is the SNR in the ith "mode".

The rank-1 approximation is then based on the decomposition:

$$\mathbf{R}_s = \underbrace{\mathbf{q}_1 \mathbf{q}_1^H \sigma_{s_1}}_{\mathbf{R}_{s_{r1}}} + \mathbf{R}_Z \tag{52}$$

where $\mathbf{q}_1$ is the first column of the matrix $\mathbf{Q}$, which corresponds to the highest SNR mode and $\mathbf{R}_Z$ is again a remainder matrix. The decomposition can then be sumarized as follows:

$$\begin{aligned} \mathbf{R}_s &= \mathbf{Q}\text{diag}\{\sigma_{s_1}, \sigma_{s_2}, \ldots, \sigma_{s_M}\}\mathbf{Q}^H \\ \mathbf{R}_{s_{r1}} &= \mathbf{Q}\text{diag}\{\sigma_{s_1}, 0, \ldots, 0\}\mathbf{Q}^H \quad (53) \\ \mathbf{R}_Z &= \mathbf{Q}\text{diag}\{0, \sigma_{s_2}, \ldots, \sigma_{s_M}\}\mathbf{Q}^H \quad (54) \end{aligned}$$

When $rank(\mathbf{R}_s) = 1$, then $\mathbf{R}_Z = 0$ and $\mathbf{R}_{s_{r1}}$ is the same as in the first column decomposition. When $rank(\mathbf{R}_s) > 1$, then the estimated rank-1 approximation $\tilde{\mathbf{R}}_{s_{r1}}$ is positive semi-definite if the estimated $\tilde{\sigma}_{s_1} = \tilde{\sigma}_{x_1} - \tilde{\sigma}_{n_1}$ is positive (which is again more likely than the first diagonal element $\tilde{\sigma}_{1,1}$ of the matrix $\tilde{\mathbf{R}}_s$ being positive as needed in the first column decomposition approach).

It is noted that:

$$\mathbf{R}_s \mathbf{t}_1 = \mathbf{R}_{s_{r1}} \mathbf{t}_1 + \underbrace{\mathbf{R}_Z \mathbf{t}_1}_{=0} = \mathbf{R}_{s_{r1}} \mathbf{e}_1 \tag{55}$$

to be compared to (26), where

$$\mathbf{t}_1 = \mathbf{Q}^{-H} \mathbf{e}_1 \mathbf{q}_1(1)^* \tag{56}$$

with $\mathbf{q}_1(1)$ is the first element of $\mathbf{q}_1$.

An analysis similar to the analysis for the first column decomposition in Section III and the EVD based decomposition in Section IV can then be done where $\mathbf{R}_s$ is replaced by the rank-1 approximation $\mathbf{R}_{s_{r1}}$ and the remainder matrix $\mathbf{R}_Z$ is either treated as noise or ignored. Equivalently, one can start from a modified MSE criterion where, compared to (17), the (arbitrary) $\mathbf{e}_1$ is replaced by $\mathbf{t}_1$:

$$\begin{aligned} J_{\text{GEVD-SDW-MWF}} = &\mathbb{E}\{|\mathbf{W}^H \mathbf{X}^s - \mathbf{t}_1^H \mathbf{X}^s|^2\} \\ &+ \mu \mathbb{E}\{|\mathbf{W}^H \mathbf{X}^n|^2\} \end{aligned} \tag{57}$$

Replacing the desired signal $\mathbf{e}_1 \mathbf{X}_s$ by $\mathbf{t}_1 \mathbf{X}_s$ is indeed equivalent to replacing $\mathbf{R}_s$ by the GEVD based $\mathbf{R}_{s_{r1}}$ as demonstrated by (55).

A. GEVD-SDW-MWF

The filter minimizing (57) is given as:

$$\mathbf{W}_{\text{GEVD-SDW-MWF}} = (\mathbf{R}_s + \mu \mathbf{R}_n)^{-1} \mathbf{R}_s \mathbf{t}_1 \tag{58}$$

Plugging (52) and (55) into (58) leads to:

$$\mathbf{W}_{\text{GEVD-SDW-MWF}} = (\mathbf{R}_{s_{r1}} + \mu(\mathbf{R}_n + \frac{1}{\mu}\mathbf{R}_Z))^{-1} \mathbf{R}_{s_{r1}} \mathbf{e}_1 \tag{59}$$

To avoid the scaling with $\frac{1}{\mu}$, the same alternative derivation as for (29) can be applied leading to:

$$\mathbf{W}^{\diamond}_{\text{GEVD-SDW-MWF}} = (\mathbf{R}_{s_{r1}} + \mu(\mathbf{R}_n + \mathbf{R}_Z))^{-1} \mathbf{R}_{s_{r1}} \mathbf{e}_1 \tag{60}$$

This means that in the SDW-MWF (18) the desired signal vector $\mathbf{R}_s$ is replaced by the GEVD based $\mathbf{R}_{s_{r1}}$ and the remainder matrix $\mathbf{R}_Z$ is treated as noise.

Plugging (50), (53) and (54) into the GEVD-SDW-MWF formula (59) also leads to:

$$\mathbf{W}_{\text{GEVD-SDW-MWF}} = \mathbf{Q}^{-H} \left[\begin{array}{c|c} \frac{\sigma_{s_1}}{\sigma_{n_1}} & \mathbf{0} \\ \hline \frac{\mu + \frac{\sigma_{s_1}}{\sigma_{n_1}}}{\mathbf{0}} & \mathbf{0} \end{array}\right] \mathbf{Q}^H \mathbf{e}_1 \tag{61}$$

Note that (61) is still true if (59) is replaced by (60).

From (50) and (53) it appears that (61) can be reformulated as follows:

$$\mathbf{W}_{\text{GEVD-SDW-MWF}} = (\mathbf{R}_{s_{r1}} + \mu \mathbf{R}_n)^{-1} \mathbf{R}_{s_{r1}} \mathbf{e}_1 \tag{62}$$

By comparing (62) to (59) it is seen that the remainder matrix $\mathbf{R}_Z$ actually has no influence on the GEVD-SDW-MWF (see also Section V-B).

B. GEVD-SP-MWF

Based on the MSE criterion (57) it is also possible to derive the SP-MWF:

$$\mathbf{W}_{\text{GEVD-SP-MWF}} = \mathbf{R}_n^{-1} \mathbf{R}_s \mathbf{t}_1 \frac{\mathbf{t}_1^H \mathbf{R}_s \mathbf{t}_1}{\mu \mathbf{t}_1^H \mathbf{R}_s \mathbf{t}_1 + \text{Tr}\{\mathbf{R}_n^{-1} \mathbf{R}_s \mathbf{t}_1 \mathbf{f}_1^H \mathbf{R}_s\}} \tag{63}$$

Plugging (52) into the GEVD-SP-MWF formula (63) leads to:

$$\mathbf{W}_{\text{GEVD-SP-MWF}} = \mathbf{R}_n^{-1} \mathbf{R}_{s_{r1}} \mathbf{t}_1 \frac{1}{\mu + \text{Tr}\{\mathbf{R}_n^{-1} \mathbf{R}_{s_{r1}}\}} \tag{64}$$

and

$$\mathbf{W}_{\text{GEVD-SP-MWF}} = (\mathbf{R}_{s_{r1}} + \mu \mathbf{R}_n)^{-1} \mathbf{R}_{s_{r1}} \mathbf{e}_1 \tag{65}$$

This means that in the SDW-MWF (18) $\mathbf{R}_s$ is replaced by the GEVD based $\mathbf{R}_{s_{r1}}$ and the remainder matrix $\mathbf{R}_Z$ is simply ignored. From equations (62) and (65) it appears that the GEVD-SDW-MWF and the GEVD-SP-MWF are fully equivalent.

$$\mathbf{W}_{\text{GEVD-SDW-MWF}} = \mathbf{W}_{\text{GEVD-SP-MWF}} \tag{66}$$

The good news here is that the question as to whether $\mathbf{R}_Z$ should be either treated as noise (GEVD-SDW-MWF) or ignored (GEVD-SP-MWF) becomes void, as the corresponding solutions are indeed the same.

C. A matrix approximation based derivation of GEVD-SDW-MWF and GEVD-SP-MWF

In matrix approximation problem (46), rather than using an unweighted Frobenius norm, where absolute (squared) approximation errors are summed, it may be more appropriate to consider relative approximation errors, where larger errors are tolerated in places where there is a lot of noise. This is standardly done by including a noise prewhitening operation. From the GEVD (50) it follows that:

$$\mathbf{R}_n = (\mathbf{Q}\mathbf{\Sigma}_n^{1/2})(\mathbf{Q}\mathbf{\Sigma}_n^{1/2})^H \tag{67}$$

The noise prewhitening is then done by premultiplying each vector with $(\mathbf{Q}\mathbf{\Sigma}_n^{1/2})^{-1}$. Each autocorrelation matrix is premultiplied with $(\mathbf{Q}\mathbf{\Sigma}_n^{1/2})^{-1}$ and postmultiplied with $(\mathbf{Q}\mathbf{\Sigma}_n^{1/2})^{-H}$ (so that for instance $\mathbf{R}_n$ is prewhitened into $\mathbf{I}$).

The criterion (46) is then replaced by:

$$\begin{aligned} J_{\text{pw-r1}} =& \alpha||(\mathbf{Q}\mathbf{\Sigma}_n^{1/2})^{-1}[\mathbf{R}_x - (\mathbf{R}_{n_{r1}} + \mathbf{R}_{s_{r1}})](\mathbf{Q}\mathbf{\Sigma}_n^{1/2})^{-H}||_F^2 \\ &+ (1-\alpha)||(\mathbf{Q}\mathbf{\Sigma}_n^{1/2})^{-1}[\mathbf{R}_n - \mathbf{R}_{n_{r1}}](\mathbf{Q}\mathbf{\Sigma}_n^{1/2})^{-H}||_F^2 \end{aligned} \tag{68}$$

where now $\mathbf{R}_{s_{r1}}$ and $\mathbf{R}_{n_{r1}}$ are sought such that after the prewhitening the Frobenius norms are minimal.

It can be verified that the prewhitening does not change (47), i.e., when an optimal $\mathbf{R}_{s_{r1}}$ is given, the optimal solution for $\mathbf{R}_{n_{r1}}$ is still given by (47).

The $\mathbf{R}_{n_{r1}}$ can then again be eliminated from the optimization problem by plugging (47) into (68). Therefore, after some simple manipulation, $\mathbf{R}_{s_{r1}}$ should minimize the criterion (68).

The optimal solution is then shown to be:

$$\begin{aligned} \mathbf{R}_{s_{r1}} &= \mathbf{q}_{\max}\mathbf{q}_{\max}^H \max(\sigma_{x_1} - \sigma_{n_1}, 0) \\ &= \mathbf{Q}\text{diag}\{\max(\sigma_{x_1} - \sigma_{n_1}, 0), 0, \ldots, 0\}\mathbf{Q}^H \end{aligned} \tag{70}$$

Assuming $\sigma_{x_1} - \sigma_{n_1}$ is non-negative, this corresponds to (55).

Once $\mathbf{R}_{s_{r1}}$ is defined according to (70), $\mathbf{R}_{n_{r1}}$ is computed based on (47), leading to (71).

Again, two extreme cases can be considered, as follows:

- If $\alpha \rightarrow 1$ then $\mathbf{R}_{n_{r1}} = \mathbf{R}_x - \mathbf{R}_{s_{r1}} = \mathbf{R}_n + \mathbf{R}_Z$ with $\mathbf{R}_Z$ defined in (52). By replacing $\{\mathbf{R}_s, \mathbf{R}_n\}$ by this $\{\mathbf{R}_{s_{r1}}, \mathbf{R}_{n_{r1}}\}$ in formula (18), the EVD-SDW-MWF formula (62) is obtained.
- If $\alpha \rightarrow 0$ then $\mathbf{R}_{n_{r1}} = \mathbf{R}_n$. By replacing $\{\mathbf{R}_s, \mathbf{R}_n\}$ by this $\{\mathbf{R}_{s_{r1}}, \mathbf{R}_{n_{r1}}\}$ in formula (18), the EVD-SP-MWF formula (65) is obtained.

It is reiterated that the GEVD-SDW-MWF and GEVD-SP-MWF are found to be fully equivalent (formula (66)), so that in this case, the selection of a good α, remarkably, becomes irrelevant.

VI. Rank-R approximation GEVD based NR filters

The GEVD based rank-1 approximation in (52) can be seen as an extreme case of a more general rank-R approximation, which then leads to more general rank-R approximation based NR filters.

Plugging (50) and (51) into the SDW-MWF formula (18) leads to:

$$\mathbf{W}_{\text{SDW-MWF}} = \mathbf{Q}^{-H}(\text{diag}\{\frac{\frac{\sigma_{s_i}}{\sigma_{n_i}}}{\mu + \frac{\sigma_{s_i}}{\sigma_{n_i}}}\})\mathbf{Q}^H\mathbf{e}_1 \tag{72}$$

Considering the gains in the diagonal matrix in (72):

$$1 \geq \frac{\frac{\sigma_{s_1}}{\sigma_{n_1}}}{\mu + \frac{\sigma_{s_1}}{\sigma_{n_1}}} \geq \frac{\frac{\sigma_{s_2}}{\sigma_{n_2}}}{\mu + \frac{\sigma_{s_2}}{\sigma_{n_2}}} \geq \cdots \geq \frac{\frac{\sigma_{s_M}}{\sigma_{n_M}}}{\mu + \frac{\sigma_{s_M}}{\sigma_{n_M}}} \geq 0 \tag{73}$$

It has been demonstrated that CI recipients can tolerate a much higher SD than normal hearing subjects. This means that the NR can be tuned to be more aggressive, which in the SDW-MWF corresponds to increasing the trade-off parameter μ. Following (73), by increasing μ, a relatively larger weight is given to the modes with the highest SNR. The modes with the lowest SNR are eventually set to 0.

This can also be pursued more explicitly by setting the $M-R$ modes with the lowest SNR to 0, leading to a rank-R approximation based NR filter:

$$\mathbf{W}_{\text{GEVD-R}} = \mathbf{Q}^{-H}\left[\begin{array}{ccc|c} \frac{\sigma_{s_1}}{\sigma_{n_1}} & & & \\ \overline{\mu+\frac{\sigma_{s_1}}{\sigma_{n_1}}} & & \mathbf{0} & \\ & \ddots & & \mathbf{0} \\ \mathbf{0} & & \frac{\sigma_{s_R}}{\sigma_{n_R}} & \\ & & \overline{\mu+\frac{\sigma_{s_R}}{\sigma_{n_1}}} & \\ \hline & \mathbf{0} & & \mathbf{0} \end{array}\right]\mathbf{Q}^H\mathbf{e}_1 \tag{74}$$

This NR filter is then equivalent to (72) where the trade-off parameter μ is mode-dependent. In the $M-R$ modes with the lowest SNR the trade-off parameter $\mu = \infty$ whereas in the R modes with the highest SNR μ is set to a real value. This approach then again corresponds to tuning the SDW-MWF to perform a more agressive NR, which indeed makes sense for CI recipients. Note that for $R = 1$ (rank-1 approximation case) only the mode with the highest SNR is not set to 0 and then (74) reduces to (61), i.e.,

$$\mathbf{W}_{\text{GEVD}-1} = \mathbf{W}_{\text{GEVD}-\text{SDW}-\text{MWF}} = \mathbf{W}_{\text{GEVD}-\text{SP}-\text{MWF}} \tag{75}$$

For $R = M$, none of the modes is set to 0 and hence

$$\mathbf{W}_{\text{GEVD}-\text{M}} = \mathbf{W}_{\text{SDW}-\text{MWF}} \tag{76}$$

The rank-R approximation is effectively based on the decomposition:

$$\mathbf{R}_s = \mathbf{R}_{s_{rR}} + \mathbf{R}_Z \tag{77}$$

where $\mathbf{R}_{s_{rR}}$ is a rank-R approximation of $\mathbf{R}_s$. For $R > 1$ the matrices can be expressed as follows:

$$\mathbf{R}_s = \mathbf{Q}\text{diag}\{\sigma_{s_1}, \sigma_{s_2}, \ldots, \sigma_{s_M}\}\mathbf{Q}^H \tag{78}$$

$$\mathbf{R}_{s_{rR}} = \mathbf{Q}\text{diag}\{\sigma_{s_1}, \ldots, \sigma_{s_R}, 0, \ldots, 0\}\mathbf{Q}^H \tag{79}$$

$$\mathbf{R}_Z = \mathbf{Q}\text{diag}\{0, \ldots, 0, \sigma_{s_{(R+1)}}, \ldots, \sigma_{s_M}\}\mathbf{Q}^H \tag{80}$$

The rank-R approximation can be further decomposed into a sum of rank-1 terms:

$$\mathbf{R}_{s_{rR}} = \sum_{i=1}^{R} \underbrace{\mathbf{q}_i\mathbf{q}_i^H\sigma_{s_i}}_{\mathbf{R}_{s_{Ri}}} \tag{81}$$

where $\mathbf{q}_i$ is the ith column of the matrix $\mathbf{Q}$, which corresponds to the ith mode, leading to:

$$\mathbf{R}_{s_{Ri}} = \mathbf{Q}\text{diag}\{0, \ldots, \sigma_{s_i}, \ldots, 0\}\mathbf{Q}^H \tag{82}$$

It is then noted that:

$$\mathbf{R}_s\sum_{i=1}^{R}\mathbf{t}_i = \sum_{i=1}^{R}(\mathbf{R}_{s_{ri}}\mathbf{t}_i + \underbrace{(\mathbf{R}_Z + \sum_{\substack{j=1 \\ j\neq i}}^{R}\mathbf{R}_{s_{ri}})\mathbf{t}_i}_{=0}) = \mathbf{R}_{s_{rR}}\mathbf{e}_1 \tag{83}$$

$$J_{\mathrm{s_{pw-r1}}} = \alpha(1-\alpha)||(\mathbf{Q}\mathbf{\Sigma}_n^{1/2})^{-1}[\mathbf{R}_x - \mathbf{R}_n - \mathbf{R}_{s_{r1}}](\mathbf{Q}\mathbf{\Sigma}_n^{1/2})^{-H}||_F^2 = \alpha(1-\alpha)||(\mathbf{\Sigma}_n^{-1}\mathbf{\Sigma}_s - \mathbf{I} - (\mathbf{Q}\mathbf{\Sigma}_n^{1/2})^{-1}\mathbf{R}_{s_{r1}}(\mathbf{Q}\mathbf{\Sigma}_n^{1/2})^{-H}||_F^2 \tag{68}$$

$$\mathbf{R}_{n_{r1}} = \mathbf{Q}\mathrm{diag}\{\sigma_{n_1}, \alpha\sigma_{x_2} + (1-\alpha)\sigma_{n_2}, \ldots, \alpha\sigma_{x_M} + (1-\alpha)\sigma_{n_M}\}\mathbf{Q}^H \tag{71}$$

to be compared to (26), where

$$\mathbf{t}_i = \mathbf{Q}^{-H}\mathbf{e}_i\mathbf{q}_i(1)^* \tag{84}$$

with $\mathbf{q}_i(1)$ is the first element of $\mathbf{q}_i$ and $\mathbf{e}_i$ an all-zero vector except for a one in the ith position.

An analysis similar to the analysis for the first column decomposition in Section III, the EVD based decomposition in Section IV and the GEVD based decomposition in Section V can then be done where $\mathbf{R}_s$ is replaced by the rank-R approximation $\mathbf{R}_{s_{rR}}$ and the remainder matrix $\mathbf{R}_Z$ is either treated as noise or ignored. Equivalently, one can start from a modified MSE criterion where, compared to (17), the (arbitrary) $\mathbf{e}_1$ is replaced by $\mathbf{t}_{rR}$:

$$J_{\mathrm{GEVD-R}} = \mathbb{E}\{|\mathbf{W}^H\mathbf{X}^s - \mathbf{t}_{rR}^H\mathbf{X}^s|^2\} + \mu\mathbb{E}\{|\mathbf{W}^H\mathbf{X}^n|^2\} \tag{85}$$

where:

$$\mathbf{t}_{rR} = \sum_{i=1}^{R}\mathbf{t}_i \tag{86}$$

Replacing the desired signal $\mathbf{e}_1\mathbf{X}_s$ by $\mathbf{t}_{rR}\mathbf{X}_s$ is indeed equivalent to replacing $\mathbf{R}_s$ by the GEVD based $\mathbf{R}_{s_{rR}}$ as demonstrated by (83). Note that $\mathbf{t}_{rM} = \mathbf{e}_1$ and so

$$J_{\mathrm{GEVD-M}} = J_{\mathrm{SDW-MWF}} \tag{87}$$

again leading to (76).

As in the rank-1 approximation case (Section V), it can easily be shown that the $\mathbf{R}_Z$ can be either treated as noise or ignored as the corresponding NR filters are both equal to $\mathbf{W}_{\mathrm{GEVD-R}}$ as given in (74).

VII. Experimental results

A. Experimental setup

The simulations were run on acoustic path measurements obtained in a reverberant room ($\mathrm{RT}_{60} = 0.61s$ [25], [26]) with a CORTEX MK2 manikin equipped with two Cochlear SP15 behind-the-ear devices. Each device has two omnidirectional microphones. The manikin head is used so that the head shadow effects are taken into account. The sound sources (FOSTEX 6301B loudspeakers) were positioned at 1 meter from the center of the head. The system was calibrated with a microphone placed at the position of the center of the head. The input SNR is then the SNR at the center of the head.

In each experiment, the speech signal was composed of five consecutive sentences from the English Hearing-In-Noise Test (HINT) database [27] concatenated with five second silence periods. The noise was the multitalker babble signal from Auditec [28]. Three spatial scenarios were considered, two single noise source scenarios (S0N45 and S90N270) and one scenario with multiple noise sources (S0N90-180-270) where the speech source (S) and the noise source(s) (N) are located at the specified angle. When multiple noise sources are present, different time shifted versions of the multitalker babble signal were used to ensure uncorrelated noise sources. In each scenario, signals with input SNR varying from -15dB to 5dB are presented to the left and right devices. The microphone signals are then filtered by several NR algorithms and the performance is compared.

All the signals were sampled at 20480Hz. The filter lengths and DFT size were set to $N = 128$ and the frame overlap was set to half of the DFT size ($L = 64$). When mentioned, the so-called input SNR is the SNR at the center of the head (excluding the HRTF effects).

B. Performance measures

An intelligibility weighted SD (SIW-SD) measure is used defined as

$$\mathrm{SIW-SD} = \sum_i I_i \mathrm{SD}_i \tag{88}$$

where I_i is the band importance function defined in [29] and SD_i the average SD (in dB) in the *i*-th one third octave band,

$$\mathrm{SD}_i = \frac{1}{(2^{1/6} - 2^{-1/6})f_i^c}\int_{2^{-1/6}f_i^c}^{2^{1/6}f_i^c} |10\log_{10} G^s(f)|df \tag{89}$$

with center frequencies f_i^c and $G^s(f)$ is given by:

$$G^s(f) = \frac{P_{X_s}(f)}{P_{Z_s}(f)} \tag{90}$$

where $P_{X_s}(f)$ and $P_{Z_s}(f)$ are the power, for the frequency f, of the speech component of the input signal X_s and the speech component output signal Z_s, respectively.

The speech intelligibility-weighted SNR (SIW-SNR) [30] is used here to compute the *SIW-SNR improvement* which is defined as

$$\Delta\mathrm{SNR}_{\mathrm{intellig}} = \sum_i I_i(\mathrm{SNR}_{i,\mathrm{out}} - \mathrm{SNR}_{i,\mathrm{front}}) \tag{91}$$

where $\mathrm{SNR}_{i,\mathrm{out}}$ and $\mathrm{SNR}_{i,\mathrm{front}}$ represent the output SNR (at the considered ear) of the NR filter and the SNR of the signal in the front microphone (at the considered ear) of the *i*th band, respectively.

The percentage of estimated $\mathbf{R}_{s_{r1}}$'s that are not positive semi-definite (%NPD) is defined as follows:

$$\%\mathrm{NPD} = \frac{NPD}{N_{\mathrm{Mat}}} * 100 \tag{92}$$

where NPD is the number of estimated $\mathbf{R}_{s_{r1}}$'s that are not positive semi-definite and N_{Mat} is the total number of estimated $\mathbf{R}_{s_{r1}}$'s.

C. Algorithms tested

For the first spatial scenario (S0N45) the SDW-MWF (18), the SP-MWF (20), the EVD-SDW-MWF (40) and the GEVD-SDW-MWF (61) are first tested (Figures 1 and 2), then only the performance of the SDW-MWF and the GEVD-MWF are investigated further (Figures 3 and 4). For the other two spatial scenarios (S90N270 and S0N90-180-270) only the performance of the SDW-MWF and the GEVD-MWF are compared (Figures 5 and 6, respectively Figures 7 and 8).

For each algorithm (SDW-MWF and GEVD-SDW-MWF), three cases are then considered: a bilateral (2+0) system, a binaural (2+1) system where only the signal from the front microphone of the contra-lateral device is used (this is referred to as "front") and a binaural (2+2) system where both microphone signals from the contra-lateral device are used (this is referred to as "binaural").

D. Speech source at 90°, single noise source at 45° (S0N45)

In the first spatial scenario, the speech source is located at 0° and the noise source at 45°. The aim of this scenario is to investigate to which extent an MWF-based NR can benefit from the EVD or the GEVD based approach when the speech source and the noise source are closely spaced.

The SIW-SNR improvement at the left ear for bilateral NR filters (where each device uses only its own microphones) is presented in Figure 1. The EVD based NR filters and the GEVD based NR filter exhibit similar SIW-SNR improvement performance and improve the SIW-SNR by about 2dB compared to the SDW-MWF. In this particular scenario, at low input SNR, the behaviour of the SP-MWF is unpredictable, this is caused by the sensitivity of the SP-MWF to the estimated $\mathbf{R}_{s_{r1}}$.

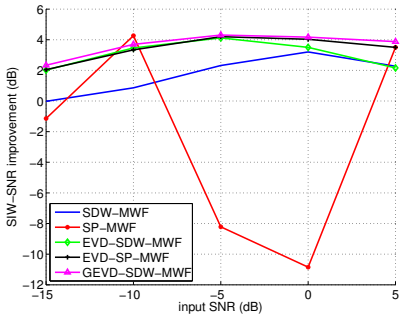


Fig. 1: SIW-SNR performance at the left ear (bilateral filters)

Figure 2 presents the %NPD, for the left ear, as a function of the input SNR. At -15dB input SNR, direct estimation leads to a %NPD of about 70%. EVD and GEVD based approaches allow to decrease this to 60% and 50% ,respectively.

From now on, only the SDW-MWF and the GEVD-SDW-MWF are going to be considered as, for this particular scenario, the SP-MWF has been seen to provide unpredictable NR performance and the EVD based NR does not provide any improvement compared to the GEVD-SDW-MWF.

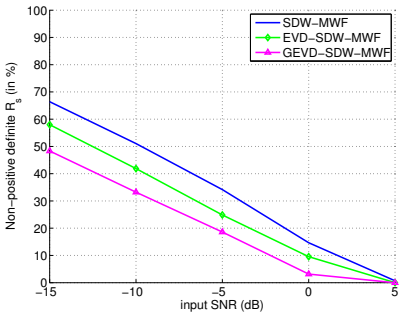
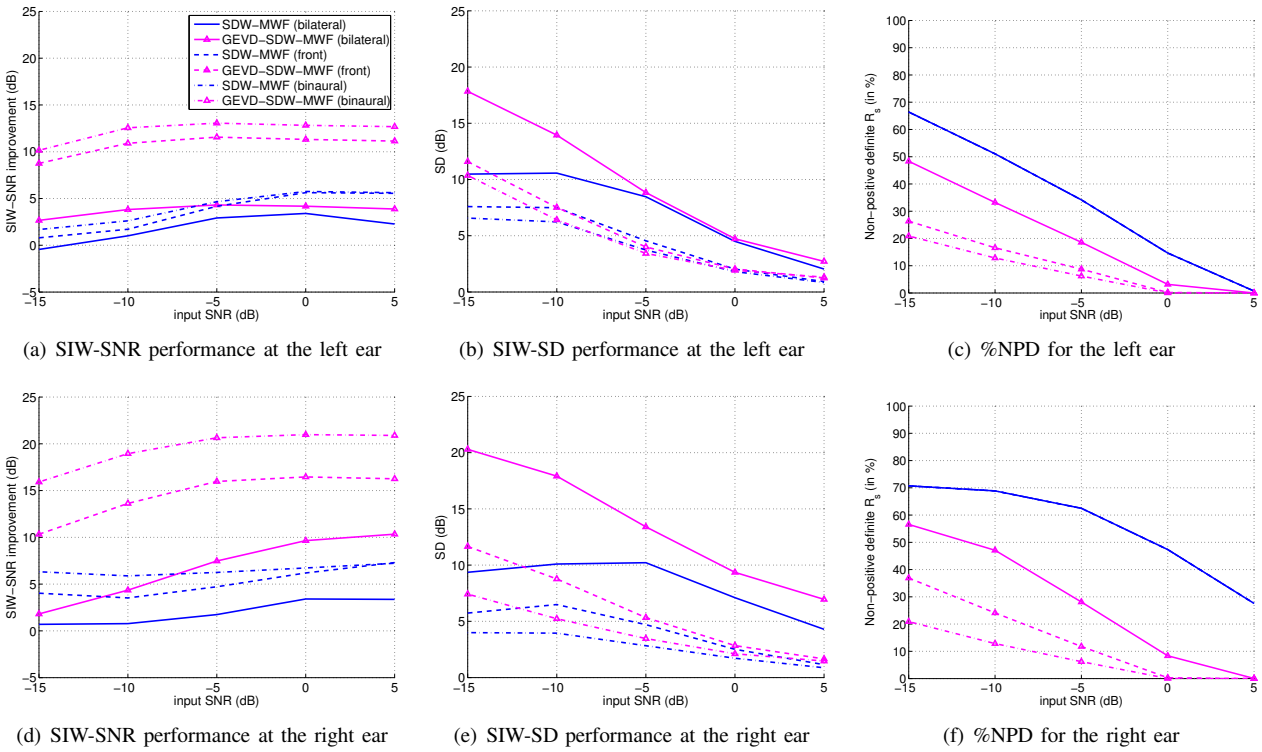


Fig. 2: %NPD for the left ear (bilateral filters)

The next results demonstrate the benefit from the GEVD-SDW-MWF for binaural systems (Figure 3). Figure 3(c) presents the %NPD, at the left ear, as a function of the input SNR for bilateral, front and binaural SDW-MWF and GEVD-SDW-MWF. For the SDW-MWF's the first diagonal element of $\mathbf{R}_{s_{r1}}$, i.e., $\mathbf{R}_{s_{r1}}(1,1)$ is the same in all three cases, therefore, bilateral, front and binaural return the same %NPD that can be as high as 65% at -15dB input SNR. For the GEVD-SDW-MWF's on the other hand, the positive semi-definiteness of $\mathbf{R}_{s_{r1}}$ depends on $\sigma_{\max} = \max\frac{\sigma_{x_i}}{\sigma_{n_i}}$ and each additional channel can help to improve this. Therefore, whereas the bilateral GEVD-SDW-MWF already decreases the %NPD to 50% at -15dB input SNR, the front GEVD-SDW-MWF and the binaural GEVD-SDW-MWF allow to further decrease the %NPD to 25% and 20%, respectively.

Figures 3(a) and 3(b) present the SIW-SNR improvement and the SIW-SD introduced at the left ear, respectively. At low input SNR, the bilateral SDW-MWF barely gives any SIW-SNR improvement while it is still introducing about 10dB SD. The front and the binaural SDW-MWF allow to improve the SIW-SNR from 2dB to 6dB depending on the input SNR while introducing lower SIW-SD than the bilateral SDW-MWF. It is important to notice that in this scenario, the left ear is the so-called best ear (*i.e.*, the ear with the highest input SNR) and an improvement of the SIW-SNR of around 2dB to 6dB at the best ear can already improve comfort and speech understanding tremendously. The GEVD-SDW-MWF provides an SIW-SNR improvement that is about 2dB to 6dB higher than the improvement for the SDW-MWF but at the cost of a higher SD. For input SNR higher than -10dB, the GEVD-SDW-MWF and the SDW-MWF are introducing a similar amount of SD.

Figure 3(f) presents the %NPD, at the right ear, as a function of the input SNR for bilateral, front and binaural SDW-MWF and GEVD-SDW-MWF. The SDW-MWF returns a %NPD that can be as high as 70% at -15dB input SNR whereas the GEVD-SDW-MWF can decrease this percentage down to about 20%. In this scenario, as the right ear in the worst ear, the $\mathbf{R}_{s_{r1}}$ can benefit from the higher SNR of the signal from the contra-lateral device. This is especially the case for the binaural GEVD-SDW-MWF that is delivering a %NPD as low as 20%



(a) SIW-SNR performance at the left ear (b) SIW-SD performance at the left ear (c) %NPD for the left ear

(d) SIW-SNR performance at the right ear (e) SIW-SD performance at the right ear (f) %NPD for the right ear

Fig. 3: Performance for the S0N45 scenario, comparison between SDW-MWF and GEVD-SDW-MWF

at -15dB SNR, which is the same figure as for the best ear (see also Figure 3(c)).

Figures 3(d) and 3(e) present the SIW-SNR improvement and the SIW-SD introduced at the right ear, respectively. The bilateral SDW-MWF provides an SIW-SNR improvement below 4dB while it is still introducing 5db to 10dB SD. The front and the bilateral SDW-MWF allow to improve the SIW-SNR from 4dB to 7dB depending on the input SNR while introducing less than 6dB SD. The GEVD-SDW-MWF provides an SIW-SNR improvement that is up to 10 dB higher than the improvement for the corresponding SDW-MWF, at a cost of a higher SIW-SD at low input SNR (up to 5dB). For input SNR higher than -10dB, however, the GEVD-SDW-MWF and the SDW-MWF are introducing a similar SD. In this scenario, as the right ear in the worst ear, the NR can benefit from the higher SNR of the signals from the contra-lateral device which is especially the case for the GEVD-SDW-MWF.

The next two experiments support the claim that the GEVD-SDW-MWF allows to increase the SIW-SNR while introducing only a controlled SD (Figure 4). In the first experiment, the trade-off parameter μ in the GEVD-SDW-MWF is set such that the same amount of SIW-SD is introduced as with the corresponding SDW-MWF with a $\mu = 1$. Figures 4(a) and 4(b) present the SIW-SNR improvement at the left and right ear, respectively, for the SDW-MWF and the GEVD-SDW-MWF. In all cases the SIW-SNR performance of the GEVD-SDW-MWF is decreased by about 1dB compared to Figures 3(a) and 3(b). This means that, the GEVD-SDW-MWF still allows to improve the SIW-SNR by up to 10dB compared to the SDW-MWF.

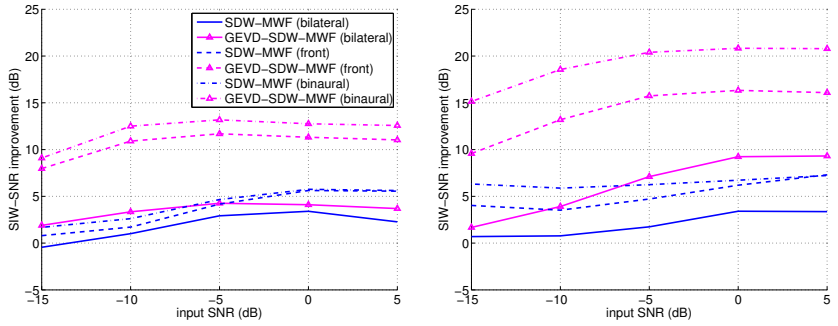
In the second experiment, the trade-off parameter μ in the SDW-MWF is set such that the SDW-MWF delivers the same SIW-SNR improvement as the corresponding GEVD-SDW-MWF with $\mu = 1$. Figures 4(c) and 4(d) present the SIW-SD introduced by the SDW-MWF and the GEVD-SDW-MWF at the left and right ear, respectively. In order to deliver a similar SIW-SNR performance, the SDW-MWF has to introduce 5dB to 10dB more SIW-SD than the corresponding GEVD-SDW-MWF at low input SNR.

E. Speech source at 0°, single noise source at 270° (S90N270)

In the second spatial scenario, the speech source is located at 90° and the noise source at 270°. The aim of this scenario is to investigate to which extent the GEVD based NR can improve the SIW-SNR performance at the best ear and the worst ear (Figure 5).

Figure 5(c) presents the %NPD, at the left ear, for the SDW-MWF and GEVD-SDW-MWF. The SDW-MWF returns a %NPD that can be as high as 70% at -15dB input SNR. The GEVD-SDW-MWF can decrease the %NPD to about 20% in the binaural case. In this scenario, as the left ear in the worst ear, the $\mathbf{R}_{s_{r1}}$ can benefit from the higher SNR of the signals from the contra-lateral device. This is especially the case for the front and the binaural GEVD-SDW-MWF that are delivering the same %NPD as the for the best ear (see also Figure 5(f)).

Figures 5(a) and 5(b) present the SIW-SNR improvement and the SIW-SD introduced at the left ear, respectively. The bilateral SDW-MWF provides an SIW-SNR improvement varying between 3dB and 7dB depending on the input SNR. The front and the binaural SDW-MWF allow to improve



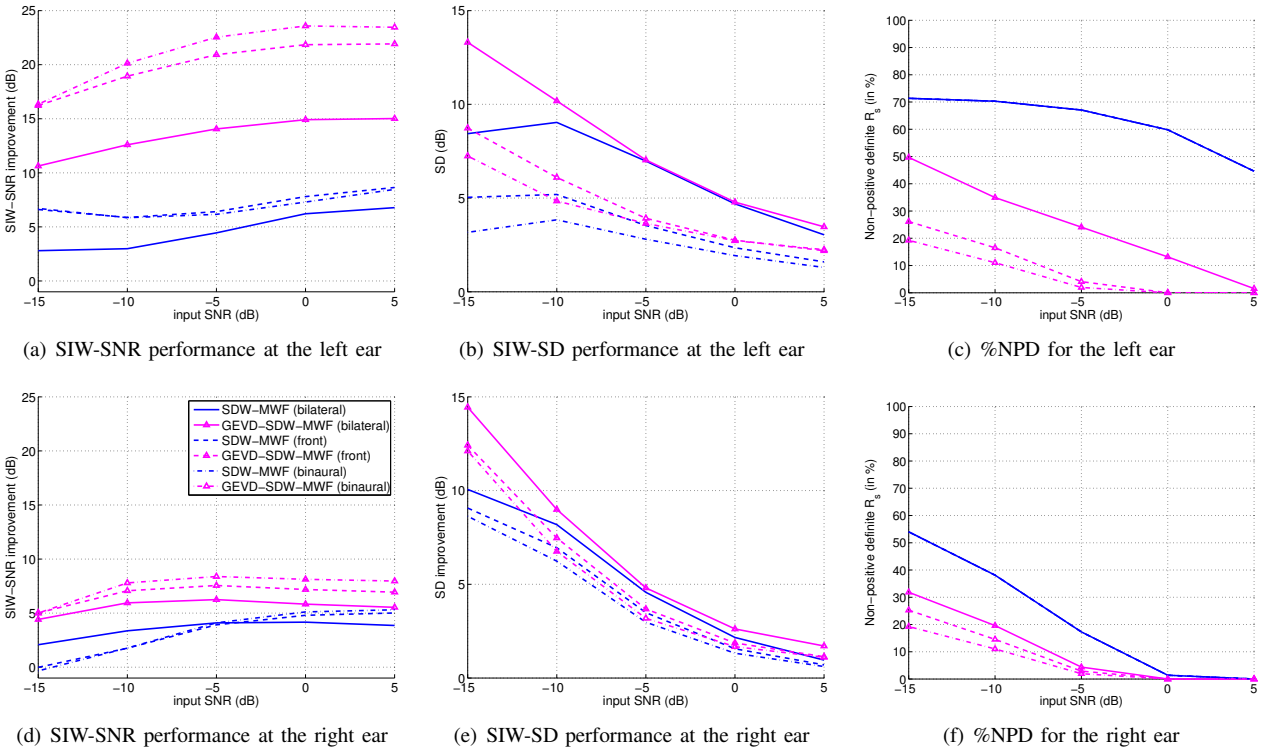
(a) SIW-SNR performance at the left ear

(b) SIW-SNR performance at the right ear

(c) SIW-SD performance at the left ear

(d) SIW-SD performance at the right ear

Fig. 4: Performance for the S0N45 scenario, comparison between SDW-MWF and GEVD-SDW-MWF with equal SIW-SD (a) and (b) and with equal SIW-SNR improvement (c) and (d).



(a) SIW-SNR performance at the left ear

(b) SIW-SD performance at the left ear

(c) %NPD for the left ear

(d) SIW-SNR performance at the right ear

(e) SIW-SD performance at the right ear

(f) %NPD for the right ear

Fig. 5: Performance for the S90N270 scenario, comparison between SDW-MWF and GEVD-SDW-MWF

the SIW-SNR by 7dB to 8dB while introducing lower SIW-SD than the bilateral SDW-MWF. The GEVD-SDW-MWF provides an SIW-SNR improvement 7dB to 15dB higher than the improvement for the corresponding SDW-MWF, at the cost of higher SIW-SD (up to 4dB) at low input SNR.

Figure 5(f) presents the %NPD, at the right ear, for the SDW-MWF and and the GEVD-SDW-MWF. The SDW-MWF returns a %NPD that can be as high as 55% at -15dB input SNR whereas the GEVD-SDW-MWF can decrease the %NPD to about 20% in the binaural case.

Figures 5(d) and 5(e) present the SIW-SNR improvement and the SIW-SD introduced at the right ear, respectively. The bilateral SDW-MWF delivers an SIW-SNR improvement from 2dB to 4dB. At low input SNR, the front and the binaural SDW-MWF suffer from the low input SNR of the signals from the contra-lateral device and deliver a lower SIW-SNR than the bilateral SDW-MWF. The GEVD-SDW-MWF delivers an SIW-SNR up to 6dB higher than the improvement for the corresponding SDW-MWF, at the cost of a higher SIW-SD (up to 5dB at low input SNR). It is important to note that, at low input SNR, the front and the binaural GEVD-SDW-MWF still deliver a better SIW-SNR improvement than the bilateral GEVD-SDW-MWF and are therefore less affected by the low SNR of the signals from the contra-lateral device than the corresponding SDW-MWF.

Figures 6(a) and 6(b) present the SIW-SNR improvement at the left and right ear, respectively, for the SDW-MWF and the GEVD-SDW-MWF when the trade-off parameter μ is set so that the GEVD-SDW-MWF introduces the same amount of SIW-SD as the corresponding SDW-MWF (with $\mu = 1$). In this case, the GEVD-SDW-MWF still outperforms the corresponding SDW-MWF and allows to improve the SIW-SNR by up to 8dB.

Figures 6(c) and 6(d) present the SIW-SD introduced by the SDW-MWF and the GEVD-SDW-MWF at the left and right ear, respectively when the trade-off parameter μ is set so that the SDW-MWF delivers the same SIW-SNR as the corresponding GEVD-SDW-MWF (with $\mu = 1$). The SDW-MWF then introduces at least 5dB more SIW-SD than the corresponding GEVD-SDW-MWF at low input SNR. Once again, at the best ear, the front and the binaural SDW-MWF suffer from the poor input SNR of signals from the contra-lateral device and introduce an extensive amount of SD.

F. Speech source at 0°, multiple noise sources (S0N90-180-270)

In the third spatial scenario, the speech source is located at 0° and the uncorrelated noise sources at 90°, 180° and 270°. The aim of this scenario is to investigate to which extent the GEVD based approach can improve the robustness in multiple noise sources scenarios. The scenario is spatially symmetrical so the NR will perform similarly in both ears. Therefore, only the results for the right ear are presented here (Figure 7).

Figure 7(c) presents the %NPD, as a function of the input SNR for the SDW-MWF and the GEVD-SDW-MWF. The SDW-MWF returns a %NPD that can be up to 65% at -15dB SNR. The GEVD-SDW-MWF can decrease this percentage to less than 10%.

Figures 7(a) and 7(b) present the SIW-SNR improvement and the SIW-SD introduced at the right ear, respectively. The SDW-MWF delivers a SIW-SNR improvement varying between 3dB and 7dB depending on the input SNR. As the scenario is symmetrical, the input SNR is similar at both ears and the front and the binaural NR cannot benefit from the higher input SNR at the contra-lateral device. There is no clear benefit either from the increased number of channels in the case of the SDW-MWF. The GEVD-SDW-MWF provides an SIW-SNR improvement 3dB to 4dB higher than the improvement for the corresponding SDW-MWF. The front and the binaural GEVD-SDW-MWF deliver an SIW-SNR improvement up to 2dB higher than the bilateral GEVD-SDW-MWF, which shows the benefits from the increased number of channels in the case of the GEVD-SDW-MWF.

When the trade-off parameter μ is set so that the GEVD-SDW-MWF introduces the same amount of SIW-SD as the corresponding SDW-MWF (with $\mu = 1$), the GEVD-SDW-MWF performance is just slightly reduced and still better than the SDW-MWF performance. Figure 8 presents the SIW-SD introduced by the SDW-MWF and the GEVD-SDW-MWF at the right ear when the trade-off parameter μ is set so that the SDW-MWF delivers the same SIW-SNR as the corresponding GEVD-SDW-MWF (with $\mu = 1$). The SDW-MWF then introduces up to 20dB more SIW-SD than the GEVD-SDW-MWF.

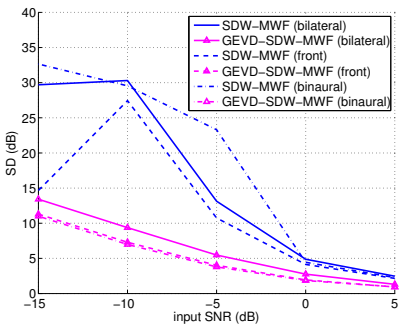
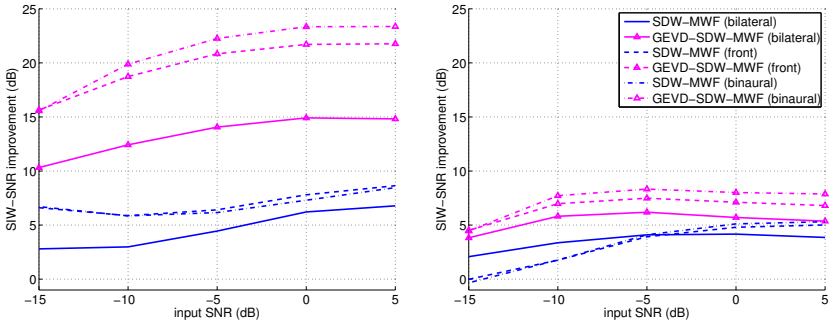


Fig. 8: SIW-SD performance at the right ear, comparison between SDW-MWF and GEVD-SDW-MWF with equal SIW-SNR improvement

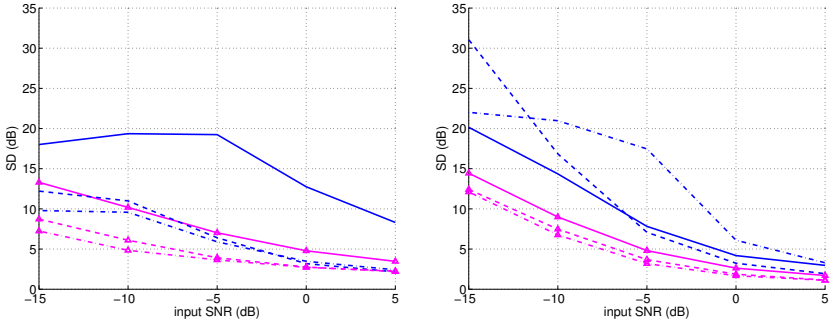
VIII. Conclusions

In this paper first the difference between the SDW-MWF, the R1-MWF and the SP-MWF (which are equivalent when the autocorrelation matrix of the speech signal is a rank-1 matrix) has been analysed when the rank of autocorrelation matrix of the speech signal is effectively greater than one. In this case, it is possible to decompose the autocorrelation matrix of the speech signal into the sum of a rank-1 approximation and a remainder matrix. The SDW-MWF, the R1-MWF and the SP-MWF then differ in the way this remainder matrix is treated.

At low input SNR, due to noise non-stationarity, the estimated autocorrelation matrix of the speech signal may not be positive semi-definite. To tackle this problem, an EVD based rank-1 approximation approach to SDW-MWF and to SP-MWF has been introduced. It is then again possible to

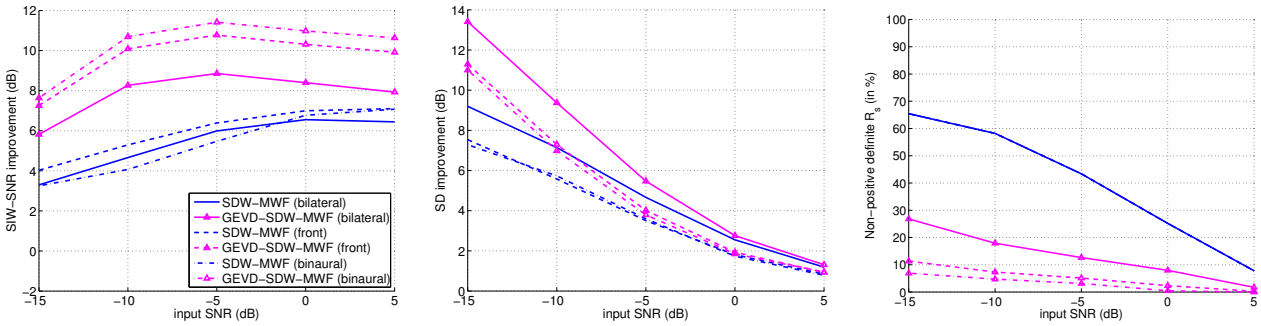


(a) SIW-SNR performance at the left ear (b) SIW-SNR performance at the right ear



(c) SIW-SD performance at the left ear (d) SIW-SD performance at the right ear

Fig. 6: Performance for the S90N270 scenario, comparison between SDW-MWF and GEVD-SDW-MWF with equal SIW-SD (a) and (b) and with equal SIW-SNR improvement (c) and (d).



(a) SIW-SNR performance at the right ear (b) SIW-SD performance at the right ear (c) %NPD for the right ear

Fig. 7: Performance for the S0N90-180-270 scenario, comparison between SDW-MWF and GEVD-SDW-MWF

decompose the autocorrelation matrix of the speech signal into the sum of a rank-1 approximation and a remainder matrix and the difference between the EVD based SDW-MWF and SP-MWF again depends in the way the remainder matrix is treated. It has been demonstrated that the EVD-SDW-MWF provides an improved SIW-NSR performance.

A GEVD based rank-1 approximation approach to SDW-MWF and to SP-MWF has finally been proposed. The rank-1 approximation based SDW-MWF and SP-MWF have then been shown to be fully equivalent even when the rank of the autocorrelation matrix of the speech signal is greater than one. As it effectively selects the mode with the highest SNR this approach has been shown to allow a more reliable estimation of the autocorrelation matrix of the speech signal than both the original SDW-MWF an SP-MWF approaches and the EVD based approaches, fully taking advantage of the high input SNR at best ear in the case of a binaural system.

The GEVD-SDW-MWF has been shown to deliver a better SIW-SNR than the corresponding SDW-MWF while introducing the same SD. Similarly, it has been shown that if the SDW-MWF was to be set to deliver similar SIW-SNR as the corresponding GEVD-SDW-MWF, it introduces a large amount of SD. Finally, the rank-1 approximation based GEVD-SDW-MWF has been generalised to a rank-R approximation based approach (GEVD-R), which encompasses the GEVD-SDW-MWF (GEVD-1) and the SDW-MWF (GEVD-M) as extreme cases.

As a final remark, in this paper a perfect VAD is used and the benefits of the presented algorithms might be limited by the need of a VAD at SNR ranging from -15dB to 5dB. However,

the improvement due to the GEVD-SDW-MWF happens not only at -15dB SNR but also at 5dB SNR where typically, VADs are functional. Also, binaural VADs have been shown to be functional down to about -5dB [31] SNR whereas a monaural VAD usually stops being efficient around 0dB SNR.

REFERENCES

[1] Y. Hu and P. C. Loizou, "A new sound coding strategy for suppressing noise in cochlear implants," *The Journal of the Acoustical Society of America*, vol. 124, no. 1, p. 498, 2008. [Online]. Available: http://link.aip.org/link/JASMAN/v124/i1/p498/s1&Agg=doi

[2] V. Hamacher, W. Doering, G. Mauer, H. Fleischmann, J. Hennecke *et al.*, "Evaluation of noise reduction systems for cochlear implant users in different acoustic environment." *The American journal of otology*, vol. 18, no. 6 Suppl, p. S46, 1997.

[3] R. J. M. van Hoesel, "Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients," *The Journal of the Acoustical Society of America*, vol. 97, no. 4, p. 2498, 1995. [Online]. Available: http://link.aip.org/link/JASMAN/v97/i4/p2498/s1&Agg=doi

[4] M. Weiss *et al.*, "Effects of noise and noise reduction processing on the operation of the nucleus-22 cochlear implant processor," *Journal of rehabilitation research and development*, vol. 30, pp. 117–117, 1993.

[5] A. Spriet, L. Van Deun, K. Eftaxiadis, J. Laneau, M. Moonen, B. van Dijk, A. van Wieringen, and J. Wouters, "Speech understanding in background noise with the two-microphone adaptive beamformer beam in the nucleus freedom cochlear implant system," *Ear Hearing*, vol. 28, no. 1, pp. 62–72, Feb 2007.

[6] J. Wouters and J. Vanden Berghe, "Speech recognition in noise for cochlear implantees with a two-microphone monaural adaptive noise reduction system," *Ear and Hearing*, vol. 22, no. 5, pp. 420–430, Oct 2001.

[7] O. R. Qazi, B. Van Dijk, M. Moonen, and W. J., "Noise rejection versus speech distortion for optimal speech perception of cochlear implant users in noise," *Hearing Research*, vol. Submitted, 2012.

[8] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-domain criterion for the speech distortion weighted multichannel wiener filter for robust noise reduction," *Speech Communication*, vol. 49, no. 7-8, pp. 636–656, Jul 2007. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0167639307000313

[9] Y. Ephraim and H. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 4, pp. 251–266, Jul 1995. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=397090

[10] K. Ngo, A. Spriet, M. Moonen, J. Wouters, and S. H. Jensen, "Incorporating the conditional speech presence probability in multi-channel wiener filter based noise reduction in hearing aids," *EURASIP Journal on Advances in Signal Processing, Special Issue on Digital Signal Processing for Hearing Instruments*, vol. 2009, 2009.

[11] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel wiener filtering for noise reduction," *Signal Processing*, vol. 84, no. 12, pp. 2367–2387, Dec 2004. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0165168404002002

[12] B. Cornelis, M. Moonen, and J. Wouters, "Performance analysis of multichannel wiener filter-based noise reduction in hearing aids under second order statistics estimation errors," *IEEE Transactions on Audio Speech and Language Processing*, vol. 19, no. 5, pp. 1368–1381, Jul 2011. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5617258

[13] *Noncausal (Frequency-Domain) Optimal Filters*, ser. Springer Topics in Signal Processing. Springer Berlin Heidelberg, 2008, vol. 1, ch. chapter 6, pp. 115–137. [Online]. Available: http://link.springer.com/10.1007/978-3-540-78612-2_6

[14] M. Souden, J. Benesty, and S. Affes, "On optimal frequency-domain multichannel linear filtering for noise reduction," *IEEE Transactions on Audio Speech and Language Processing*, vol. 18, no. 2, pp. 260–276, Feb 2010. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5089420

[15] R. Serizel, B. Moonen, M. Van Dijk, and J. Wouters, "Rank-1 approximation based multichannel wiener filtering algorithms for noise reduction in cochlear implants," in *The 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013)*, Vancouver, Canada, May 2013.

[16] M. Dendrinos, S. Bakamidis, and G. Garayannis, "Speech enhancement from noise: A regenerative approach," *Speech Commun.*, vol. 10, pp. 45–57, Feb 1991.

[17] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Transactions on Signal Processing [see also IEEE Transactions on Acoustics, Speech, and Signal Processing]*, vol. 50, no. 9, pp. 2230–2244, 2002.

[18] S. Jensen, P. Hansen, S. Hansen, and J. Sorensen, "Reduction of broad-band noise in speech by truncated qsvd," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 6, pp. 439–448, 1995. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=482211

[19] S. Doclo, T. J. Klasen, T. Van den Bogaert, J. Wouters, and M. Moonen, "Theoretical analysis of binaural cue preservation using multi-channel wiener filtering and interaural transfer functions," in *Proceedings International Workshop on Acoustic Echo and Noise Control (IWAENC'06)*, 2006, pp. 1–4.

[20] Y. Hamacher, "Comparison of advanced monaural and binaural noise reduction algorithms for hearing aids," in *IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP'02)*, vol. 4. IEEE, 2002.

[21] T. J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1579–1585, Apr 2007. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4133039

[22] A. Markides. Academic, 1977.

[23] T. Van den Bogaert, "Preserving binaural cues in noise reduction algorithms for hearing aids." Ph.D. dissertation, Katholieke Universiteit Leuven, Leuven, Belgium, Faculty of Engineering, 2008.

[24] J. Benesty, J. Chen, Y. (Arden) Huang, and T. Gaensler, "Time-domain noise reduction based on an orthogonal decomposition for desired signal extraction," *The Journal of the Acoustical Society of America*, vol. 132, no. 1, p. 452, 2012. [Online]. Available: http://link.aip.org/link/JASMAN/v132/i1/p452/s1&Agg=doi

[25] T. Van den Bogaert, S. Doclo, J. Wouters, and M. Moonen, "The effect of multimicrophone noise reduction systems on sound source localization by users of binaural hearing aids," *The Journal of the Acoustical Society of America*, vol. 124, no. 1, p. 484, 2008. [Online]. Available: http://link.aip.org/link/JASMAN/v124/i1/p484/s1&Agg=doi

[26] ——, "Speech enhancement with multichannel wiener filter techniques in multimicrophone binaural hearing aids," *The Journal of the Acoustical Society of America*, vol. 125, no. 1, p. 360, 2009. [Online]. Available: http://link.aip.org/link/JASMAN/v125/i1/p360/s1&Agg=doi

[27] M. Nilsson, S. D. Soli, and J. 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, pp. 1085–1099, 1994.

[28] Auditec, "Auditory Tests (Revised), Compact Disc, Auditec, St. Louis," St. Louis, 1997.

[29] A.-S. (1997), "ANSI S3.5-1997 American National Standard Methods for calculation of the speech intelligibility index," *Acoustical Society of America*, June 1997.

[30] J. E. Greenberg, P. M. Peterson, and P. M. Zurek, "Intelligibility-weighted measures of speech-to-interference ratio and speech system performance," *J. Acoust. Soc. Amer.*, vol. 94, pp. 3009–3010, Nov 1993.

[31] B. Cornelis, M. Moonen, and J. Wouters, "Binaural voice activity detection for mwf-based noise reduction in binaural hearing aids," in *Proceedings of the European Signal Processing Conference (EUSIPCO), Barcelona, Spain*, 2011, pp. 486–490.



Romain Serizel Romain Serizel received the M.Eng. degree in Automatic System Engineering from ENSEM (Nancy, France) in 2005 and the M.Sc. degree in Signal Processing from Universit Rennes 1 (Rennes, France) in 2006. He received the Ph.D. degree in Engineering Sciences from the Katholieke Universiteit Leuven (KUL), Belgium in June 2011.

From 2011 till 2012 he was a research assistant with Prof. Marc Moonen research group at the Electrical Engineering Department (ESAT-SCD) of the KUL. Since 2013 he a post-doctoral researcher with Prof. Marcello Federico at the Human Language Technologies reseach group of the Fondazione Bruno Kessler (FBK-ICT), Trento, Italy. His research interests include hearing aids systems, cochlear implants digital signal processing for audio, automatic speech recognition and neural networks.



Marc Moonen Marc Moonen (M'94, SM'06, F'07) received the electrical engineering degree and the PhD degree in applied sciences from Katholieke Universiteit Leuven, Belgium, in 1986 and 1990 respectively.

Since 2004 he is a Full Professor at the Electrical Engineering Department of Katholieke Universiteit Leuven, where he is heading a research team working in the area of numerical algorithms and signal processing for digital communications, wireless communications, DSL and audio signal processing.

He received the 1994 K.U.Leuven Research Council Award, the 1997 Alcatel Bell (Belgium) Award (with Piet Vandaele), the 2004 Alcatel Bell (Belgium) Award (with Raphael Cendrillon), and was a 1997 "Laureate of the Belgium Royal Academy of Science". He received a journal best paper award from the IEEE Transactions on Signal Processing (with Geert Leus) and from Elsevier Signal Processing (with Simon Doclo).

He was chairman of the IEEE Benelux Signal Processing Chapter (1998-2002), and is currently Past-President of EURASIP (European Association for Signal Processing) and a member of the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications.

He has served as Editor-in-Chief for the "EURASIP Journal on Applied Signal Processing" (2003-2005), and has been a member of the editorial board of "IEEE Transactions on Circuits and Systems II" (2002-2003) and "IEEE Signal Processing Magazine" (2003-2005) and "Integration, the VLSI Journal". He is currently a member of the editorial board of "EURASIP Journal on Applied Signal Processing", "EURASIP Journal on Wireless Communications and Networking", and "Signal Processing".



Bas Van Dijk Bas van Dijk was born in Amsterdam, the Netherlands in 1968. He received his masters degree in computer sciences and physics in 1993 from the University of Amsterdam. In 1998 he received his Ph.D. from the Biophysics department of the University of Leiden (the Netherlands). From 1998 to 2001 Dr. Bas van Dijk worked as Research Audiologist in the audiology department of the University Medical Centre Utrecht (the Netherlands) were he worked on clinical and scientific aspects of cochlear implant fitting and signal processing. In 2001 he joined Cochlear Ltd, initially as clinical advisor to the R&D teams in the Cochlear Technology Centre Belgium (Mechelen).

In recent years, he is responsible for coordinating research projects for Cochlear Ltd. on signal processing worldwide, and in addition he is teamleader of the research team in Mechelen, working mainly on signal processing, clinical validation, and development of research tools. He (co-)developed many innovations in the commercial cochlear implant system and is the (co-)inventor more than 10 international patents and (co-)author of 9 peer-reviewed publications in the field of cochlear implants.



Jan Wouters 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 was a Research Fellow with the Belgian National Fund for Scientific Research (FWO) at the Institute of Nuclear Physics (UCL Louvain-la-Neuve and K.U.Leuven) and at NASA Goddard Space Flight Center (USA). Since 1993 he is a Professor at the Neurosciences Department of the K.U.Leuven (Full Professor since 2001). His research activities center around audiology and the auditory system, signal processing for cochlear implants and hearing aids. He is author of about 145 articles in international peer-reviewed journals and is a reviewer for several international journals.

Dr. Wouters received an Award of the Flemish Ministery in 1989, a Fullbright Award and a NATO Research Fellowship in 1992, and the Flemish VVL Speech therapy - Audiology Award in 1996. He is member of the International Collegium for ORL (CORLAS), a Board Member of the International Collegium for Rehabilitative Audiology (ICRA) and is responsible for the Laboratory for Experimental ORL and the audiology program at K.U.Leuven.