Rank-1 Approximation Based Multichannel Wiener Filtering Algorithms For Noise Reduction In Cochlear Implants

Romain Serizel\(^2\), Marc Moonen\(^2\), Bas Van Dijk\(^3\) and Jan Wouters\(^4\)


---

1This report is available by anonymous ftp from ftp.esat.kuleuven.be in the directory pub/sista/rserizel/reports/13-38.pdf

2K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group SCD (SISTA) Kasteelpark Arenberg 10, 3001 Leuven, Belgium, Tel. +32 16 32 9607, Fax +32 16 321970, E-mail: romain.serizel@esat.kuleuven.be. This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of the Marie-Curie Fellowship EST-SIGNAL program (http://est-signal.i3s.unice.fr) under contract No. MEST-CT-2005-021175, and the Concerted Research Action GOA-AMBioRICS. The scientific responsibility is assumed by its authors.

3Cochlear CTCE, Schalinhoevedreef 20, Building i-B, B-2800 Mechelen, Belgium

4Katholieke 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 multichannel Wiener filtering-based algorithms for noise reduction in cochlear implants. In a single speech scenario, the autocorrelation matrix of the speech signal can be approximated by a rank-1 matrix. It is then possible to derive noise reduction filters that deliver improved signal-to-noise ratio performance. The link between these different filters is investigated here and an eigenvalue decomposition based algorithm is demonstrated to be more stable at low input signal-to-noise ratio compared to previous algorithms.
1. INTRODUCTION

A major challenge in cochlear implant (CI) is to improve the speech understanding in noise [1] and so having an efficient front-end noise reduction (NR) is important. Therefore, during the past years, several NR algorithms have been developed and tested with CI recipients [2, 3, 4].

In general, CI users need a 10dB to 25dB higher signal-to-noise ratio (SNR) than normal hearing listeners to achieve similar speech intelligibility and how this provides an interpretation of the SDW-MWF and the SP-MWF. The EVD-based NR is introduced in Section 5. The performance of the original SDW-MWF and the EVD-based NR are compared in Section 6. Finally, Section 7 presents a summary of the paper.

2. RELATION TO PRIOR WORK

The work presented in this paper focuses on the analysis of the difference between the SDW-MWF [6, 7, 8] and the SP-MWF [9, 10] when the rank of the estimated autocorrelation matrix of the speech signal is higher than one. A new rank-1 approximation is also introduced. While the (implicit or explicit) rank-1 approximation in the previous work was based on a so-called first column decomposition, the new (explicit) rank-1 approximation presented here is based on the EVD.

3. BACKGROUND AND PROBLEM STATEMENT

3.1. Signal model

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

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

where $\omega = 2\pi f$ is the frequency-domain variable. For conciseness, $\omega$ will be omitted in all subsequent equations.

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 noise reduction. For the time being, a perfect VAD is assumed.

The compound vector gathering all microphone signals is:

$$X = [X_1 \ldots X_M]^T$$

An optimal (Wiener) filter $W = [W_1 \ldots W_M]^T$ will be designed and applied to the signals, which minimizes a Mean Squared Error.
(MSE) criterion:

\[ J_{\text{MSE}} = \mathbb{E} \{ |E|^2 \} \quad (3) \]

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:

\[ Z = W^H X \quad (4) \]

where \( W \) denotes the Hermitean transpose.

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

\[ D_{\text{NR}} = e^H_1 X_s \quad (5) \]

where \( e_1 \) is an all-zero vector expect for a one in the first position.

The autocorrelation matrices 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:

\[ R_s = \mathbb{E} \{ X X^H \} \quad (6) \]
\[ R_s = \mathbb{E} \{ X_s X_s^H \} \quad (7) \]
\[ R_n = \mathbb{E} \{ X_n X_n^H \} \quad (8) \]

\( R_s \) can be estimated during “noise only periods” and \( R_s \) 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, \( R_s \) can be estimated by using:

\[ R_s = R_z - R_n \quad (9) \]

3.2. MWF-based Noise Reduction

The MWF aims to minimize the squared difference between the filtered microphone signal and the desired speech signal. The corresponding MSE criterion is:

\[ J_{\text{MSE}} = \mathbb{E} \{ |E_{\text{MWF}}|^2 \} \quad (10) \]

\[ E_{\text{MWF}} = W^H X - e^H_1 X_s \quad (11) \]

The MWF solution is given as:

\[ W_{\text{MWF}} = (R_s + R_n)^{-1} R_s e_1 \quad (12) \]

The SDW-MWF has been proposed to provide an explicit trade-off between the SD and the NR [6, 7, 8]:

\[ J_{\text{MSE}} = \mathbb{E} \{ |W^H X - e^H_1 X_s|^2 \} + \mu \mathbb{E} \{ W^H X_n^2 \} \quad (13) \]

The SDW-MWF solution is then given as:

\[ W_{\text{SDW-MWF}} = (R_s + \mu R_n)^{-1} R_s e_1 \quad (14) \]

In a single speech source scenario, the autocorrelation matrix of the speech component of the microphone signals \( R_s \) is often assumed to be a rank-1 matrix, which can then be rewritten as:

\[ R_s = P^* A A^H \quad (15) \]

where \( P^* \) is the power of the speech source signal and \( A \) is the \( M \)-dimensional steering vector, containing the acoustic transfer functions from the speech source position to the hearing aid microphones (including room acoustics, microphone characteristics, and head shadow effect).

Based on this rank-1 assumption it is possible to derive the SP-MWF [9, 10]:

\[ W_{\text{SP-MWF}} = R_s^{-1} R_s e_1 + \mu e^H_1 R_s e_1 \quad (16) \]

The filters (14) and (16) are fully equivalent if \( R_s \) is rank-1. In practice, however rank\((R_s) > 1 \) even for a single speech source scenario and then (14) and (16) are different filters.

4. FIRST COLUMN DECOMPOSITION

When rank\((R_s) > 1 \) the speech autocorrelation matrix \( R_s \) can still be decomposed as:

\[ R_s = R_{s,1} + R_z \quad (17) \]

where \( R_{s,1} \) is a rank-1 matrix and \( R_z \) is a “remainder” matrix.

The most obvious choice for \( R_{s,1} \) is then a rank-1 extension of its first column and row, i.e.:

\[ R_s = dd^H \sigma_{1,1} + R_z \quad (18) \]

where

\[ \sigma_{1,1} = [R_s]_{1,1} \quad (19) \]
\[ d = \begin{bmatrix} \sigma_{2,1} & \cdots & \sigma_{1,n} \\ \sigma_{1,1} & \ddots \end{bmatrix} \quad (20) \]
\[ R_z = \begin{bmatrix} 0 & \cdots & 0 \\ 0 & \ddots & \ddots \\ \vdots & \ddots & \ddots \end{bmatrix} \quad (21) \]

and \( \sigma_{1,1} \) is the speech power in microphone 1. It is noted that:

\[ R_s e_1 = R_{s,1} e_1 + R_z e_1 \quad (22) \]

which means that the (rightmost) “desired signal part” \( R_s e_1 \) in (14) and (16) can (obviously) be replaced by the “rank-1 desired signal part” \( R_{s,1} e_1 \). The difference between the two approaches (14) and (16) then effectively depends on how \( R_z \) is treated.

4.1. SDW-MWF

Plugging (18) into the SDW-MWF formula (14) leads to:

\[ W_{\text{SDW-MWF}} = (R_{s,1} + \mu (R_n + \frac{1}{\mu} R_z))^{-1} R_{s,1} e_1 \quad (23) \]

This means that in the SDW-MWF \( R_s \) is replaced by \( R_{s,1} \) and the remainder matrix \( R_z \) is effectively treated as noise (up to a scaling with \( \frac{1}{\mu} \)).

4.2. SP-MWF

Plugging (18) into the SP-MWF formula (16) leads to:

\[ W_{\text{SP-MWF}} = (R_{s,1} + \mu R_n)^{-1} R_{s,1} e_1 \quad (24) \]

This means that the SP-MWF corresponds to the SDW-MWF (14) where \( R_s \) is replaced by \( R_{s,1} \) and the remainder matrix \( R_z \) is simply ignored. If \( R_z = 0 \) (rank-1 case) formulas (23) and (24) are again seen to be the same.
4.3. Speech autocorrelation matrix estimation

At low SNR and if the noise is non-stationary it is observed that:

\[ R_s \approx R_s \]  

(25)

and then the estimated speech autocorrelation matrix \( R_s = R_s - R_n \) can lose its positive definiteness, which has been observed to lead to filter instabilities. The first column decomposition-based filters suffer from the same estimation problem where then the estimated speech power in microphone 1 (\( \sigma_{1,1} \)) could become negative, i.e., \( R_{s,1} \) is non-positive definite and so that the desired signal is ill-defined.

5. EVD-BASED FILTERS

An alternative to the first column decomposition based rank-1 approximation would be to consider a rank-1 approximation based on an EVD of \( R_s \):

\[ R_s = d_{\text{max}} \lambda_{\text{max}} R_s + R_Z \]  

(26)

where \( \lambda_{\text{max}} \) is \( R_s \)'s largest eigenvalue and \( d_{\text{max}} \) is the corresponding normalized eigenvector and \( R_Z \) is again a remainder matrix. In this case, the rank-1 part \( R_{s,1} \) is positive definite if the dominant eigenvalue of \( R_s \) is positive (which is more likely than the first diagonal element \( \sigma_{1,1} \) of the matrix \( R_s \) being positive).

All the formulas introduced above can be modified based on this decomposition.

It is noted that now:

\[ R_s f_1 = R_{s,1} f_1 + R_Z f_1 \]  

(27)

\[ R_{s,1} f_1 = R_{s,1} e_1 \]  

(28)

to be compared to (22) where

\[ f_1 = d_{\text{max}} d_{\text{max}}(1)^* \]  

(29)

An analysis similar to the analysis of the first column decomposition in Section 4 can then be done where \( R_s \) is replaced by \( R_{s,1} \) and the remainder matrix \( R_Z \) is either treated as noise or ignored. Equivalently, one can start from a modified SDW-MWF criterion where the (arbitrary) \( e_1 \) is replaced by \( f_1 \):

\[ J_{\text{MSE}} = E\{ |W^H X^i - f^H X^i|^2 \} + \mu E\{ |W^H X^i|^2 \} \]  

(30)

5.1. SDW-MWF \text{EVD}

Plugging (27) into the SDW-MWF formula corresponding to (30) leads to:

\[ W_{\text{SDW-MWF}} = (R_{s,1} + \mu(R_n + \frac{1}{\mu} R_Z))^{-1} R_{s,1} f_1 \]  

(31)

This means that in the SDW-MWF \( R_s \) is replaced by the EVD-based \( R_{s,1} \) and the remainder matrix \( R_Z \) is effectively treated as noise (up to a scaling with \( \frac{1}{\mu} \)).

5.2. SP-MWF \text{EVD}

Plugging (27) into the SP-MWF formula corresponding to (30) leads to:

\[ W_{\text{SP-MWF}} = (R_{s,1} + \mu R_n)^{-1} R_{s,1} f_1 \]  

(32)

This means that in the SP-MWF \( R_s \) is replaced by the EVD-based \( R_{s,1} \) and the remainder matrix \( R_Z \) is simply ignored.

6. EXPERIMENTAL RESULTS

6.1. Experimental setup

The simulations were run on acoustic path measurements obtained in a reverberant room (RT60 = 0.61 s [15, 16]) with a CORTEX MK2 manikin head and torso equipped with two Cochlear SP15 behind-the-ear devices. Each device has two omnidirectional microphones. The sound sources (FOSTEX 6301B loudspeakers) were positioned at 1 meter from the center of the head.

The speech signal was composed of five consecutive sentences from the English Hearing-In-Noise Test (HINT) database [17] concatenated with five seconds silence periods. The noise was the multitalker babble from Auditec [18]. The speech source was located at 0° and the noise source at 45°. 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 DET size (\( L = 64 \)). When mentioned, the so-called input SNR is the SNR at the center of the head (excluding the head shadow effects).

6.2. Performance measures

The speech intelligibility-weighted SNR (SIW-SNR) [19] is used here to compute the \( \text{SIW-SNR improvement} \) which is defined as

\[ \Delta \text{SNR}_{\text{i,est}} = \sum \Delta \text{SNR}_{\text{i,est}} = \sum (\text{SNR}_{i,\text{out}} - \text{SNR}_{i,\text{ref}}) \]  

(33)

where \( I_i \) is the band importance function defined in [20] and \( \text{SNR}_{i,\text{out}} \) and \( \text{SNR}_{i,\text{ref}} \) represent the output SNR (at the considered ear) of one of the NR schemes and the SNR of the signal in the reference microphone (at the considered ear) of the \( i \)th band, respectively.

6.3. S0N45

Signals with input SNR varying from -15dB to 5dB are presented to the left and right hearing aid devices. The microphone signal is then filtered by several NR algorithms and the performance is compared. The SIW-SNR improvement at the left ear for bilateral NR filters is presented in Figure 1. The EVD-based rank-1 filters improve the SIW-SNR by about 2dB compared to the SDW-MWF. At low SNR, the behaviour of the SP-MWF is unpredictable, this is caused by the sensitivity of the SP-MWF to \( R_s \)'s that are not positive definite.

The next results demonstrate how the EVD-based NR can benefit from binaural setups. Three setups are considered: the bilateral setup, a so-called binaural “front” setup where only the signal from the front microphone of the contra-lateral ear is shared and a setup where both microphone signals from the contra-lateral ear are shared (this approach is referred to as “binaural” here).

Figure 2 presents percentage of estimated \( R_s \)'s that are not positive definite, at the left ear, as a function of the input SNR for bilateral, front and binaural SDW-MWF and SDW-MWF \text{EVD}. In the
SDW-MWF based NR the positive definiteness of $R_{s1}$ only depends on the first diagonal element of the speech autocorrelation matrix $R_{s1}(1,1)$, therefore, bilateral, front and binaural approaches return the same percentage of estimated $R_{s1}$’s that are not positive definite which can be as high as 65% at low SNR. In the SDW-MWF$_{EVD}$ on the other hand, the positive definiteness of $R_{s1}$ depends on $\lambda_{max}$ and each additional channel can help increasing this quantity. Therefore, whereas the bilateral SDW-MWF$_{EVD}$ can help to decrease the percentage of estimated $R_{s1}$’s that are not positive definite from about 65% to 60% at low SNR, the front and the binaural SDW-MWF$_{EVD}$ allow to decrease this percentage to around 40%.

Figures 3 presents the SIW-SNR improvement at the left ear. The front and the binaural SDW-MWF allow to improve the SIW-SNR from 2dB to 6dB depending on the input SNR. The bilateral SDW-MWF$_{EVD}$ provides an SIW-SNR improvement from 2dB to 4dB for any input SNR. This is 2dB better than the SIW-SNR improvement of the bilateral SDW-MWF. The front and the binaural SDW-MWF$_{EVD}$ provide an SIW-SNR improvement between 6dB and 12dB depending on the input SNR. This is 4dB to 6dB better than with the respective SDW-MWF approaches.

7. CONCLUSION
In this paper the difference between the SDW-MWF and the SP-MWF (which are equivalent when the autocorrelation matrix of the speech signal is a rank-1 matrix) is analysed when the rank of the autocorrelation matrix of the speech signal is higher than one. In this case, it is possible to decompose the autocorrelation matrix of the speech signal into the sum of a rank-1 matrix and a remainder matrix. The SDW-MWF and the SP-MWF then differs in the way this remainder matrix is treated.

At low input SNR, due to noise non-stationnarity, the speech autocorrelation matrix may become non-positive definite. An EVD-based rank-1 approach to SDW-MWF and to SP-MWF has been introduced. It is then again possible to decompose the autocorrelation matrix of the speech signal into the sum of a rank-1 matrix and a remainder matrix and the difference between the EVD-based SDW-MWF and SP-MWF depends in the way the remainder matrix is treated. It is demonstrated that the SDW-MWF$_{EVD}$ allows an improved SIW-NSR performance.

8. REFERENCES


