A QRD-RLS BASED FREQUENCY DOMAIN MULTICHANNEL WIENER FILTER ALGORITHM FOR NOISE REDUCTION IN HEARING AIDS

Bram Cornelis[1] , *Marc Moonen*[1], *Jan Wouters*[2]

[1]ESAT-SCD
Dept. of electrical engineering, K.U.Leuven
Kasteelpark Arenberg 10, 3001 Heverlee, Belgium
email: bram.cornelis@esat.kuleuven.be,
marc.moonen@esat.kuleuven.be

[2]ExpORL
Dept. of Neurosciences, K.U.Leuven
Herestraat 49/721, 3000 Leuven, Belgium
email: jan.wouters@med.kuleuven.be

ABSTRACT

In this paper a frequency domain multichannel Wiener filter algorithm is proposed for noise reduction in hearing aids. It is shown that a robust and efficient QR Decomposition Recursive Least Squares (QRD-RLS) based updating scheme can be derived, if a single target speech source is assumed. Moreover, the scheme also allows to include a trade-off between speech distortion and noise reduction, as with the Speech Distortion Weighted Multichannel Wiener Filter (SDW-MWF). The QRD-RLS based algorithm is compared with an adaptive SDW-MWF algorithm, for a binaural hearing aid setup with 4 microphones. Besides the fact that the QRD-RLS based algorithm achieves a further improvement in speech intelligibility weighted SNR, the computational efficiency and numerical robustness are also increased.

1. INTRODUCTION

Modern hearing aids make use of noise reduction algorithms to improve speech intelligibility in background noise. Hearing aids are usually fitted with multiple microphones, which generally leads to an improvement in noise reduction performance because spatial sound information can then be exploited in addition to spectral information. In the future, binaural hearing aids will emerge, which exchange microphone signals over a wireless radio link. As signals from both sides of the head are then available, an additional noise reduction performance increase will then be achieved.

An interesting approach to multichannel noise reduction, is based on multichannel Wiener filtering (for example, [1–3]). A Wiener filtering based approach eliminates the need for a fixed beamformer preprocessor, hence offers a very promising alternative to the Generalized Sidelobe Canceller (GSC) structure [4].

In [1], a class of adaptive noise reduction algorithms is introduced, which are frequency domain implementations of the Speech Distortion Weighted Multichannel Wiener Filter (SDW-MWF). A Recursive Least Squares (RLS)-type update procedure is adopted, where a weighted sum of a speech and a noise correlation matrix has to be inverted at every filter update. Moreover, an eigenvalue decomposition is calculated to ensure a positive definite speech correlation matrix, so that the algorithm is guaranteed not to diverge. When the number of input microphone signals is large (e.g. in binaural hearing aids), the complexity of these operations increases dramatically. Therefore, some simplifications were also proposed in [1], based on (block) diagonal approximations of the correlation matrices, which however decreases the performance.

In [2], a QR Decomposition Recursive Least Squares (QRD-RLS) based time domain implementation of the Wiener filter was introduced. Instead of the speech and noise correlation matrices, their Cholesky (square root) factors are stored and updated by a numerically robust procedure based on Givens transformations. As the Cholesky factors have half the dynamic range of the correlation matrices, the wordlength can be reduced in fixed point processing without loss of numerical accuracy. A problem with the QRD-RLS scheme of [2] is that it does not allow to include a trade-off between speech distortion and noise reduction, as in the SDW-MWF. This explicit trade-off is beneficial, as it allows increasing the global (broadband) output SNR [3]. Additionally, as the algorithm operates in the time domain, the computational complexity is prohibitive for a hearing aid application.

In this paper, it will be shown that, by assuming a single target speech source, an alternative SDW-MWF formula can be used which enables a frequency domain implementation of the SDW-MWF algorithm based on QRD-RLS. In section 2, the SDW-MWF and related filters are first reviewed. In section 3, the frequency domain implementation based on QRD-RLS is derived. In section 4, the QRD-RLS algorithm is compared with the adaptive SDW-MWF algorithm of [1]. It will be shown that the QRD-RLS algorithm obtains a higher speech intelligibility weighted SNR improvement than the algorithm in [1]. Additionally, in contrast to the algorithm in [2], a trade-off can be included between speech distortion and noise reduction. Also, as all processing is performed in the frequency domain (as is usually done in hearing aids), the computational efficiency is increased. Finally, it is demonstrated that the QRD-RLS algorithm indeed improves the numerical robustness so that the wordlength can be reduced.

2. MULTICHANNEL WIENER FILTER REVIEW

2.1 Notation and correlation matrix estimation

We consider a microphone array consisting of N microphones. The nth microphone signal $Y_n(\omega)$ can be specified in the frequency domain as

$$Y_n(\omega) = X_n(\omega) + V_n(\omega), \quad n = 1 \dots N, \tag{1}$$

where $X_n(\omega)$ represents the speech component and $V_n(\omega)$ represents the noise component in the nth microphone. For conciseness, we omit the frequency variable ω from now on. The signals Y_n, X_n and V_n are stacked in the N-dimensional vectors $\mathbf{Y}$, $\mathbf{X}$ and $\mathbf{V}$, with $\mathbf{Y}$=$\mathbf{X}+\mathbf{V}$. The correlation matrix $\mathbf{R}_y$, the speech correlation matrix $\mathbf{R}_x$ and the noise correlation matrix $\mathbf{R}_v$ are then defined as

$$\mathbf{R}_y = \mathscr{E}\{\mathbf{Y}\mathbf{Y}^H\},\ \ \mathbf{R}_x = \mathscr{E}\{\mathbf{X}\mathbf{X}^H\},\ \ \mathbf{R}_v = \mathscr{E}\{\mathbf{V}\mathbf{V}^H\}\,, \tag{2}$$

where $\mathscr{E}$ denotes the expected value operator. It will be assumed that a voice activity detection (VAD) algorithm is available so that a distinction can be made between speech + noise and noise-only

Bram Cornelis is funded by a Ph.D. grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). This research work was carried out at the ESAT Laboratory of Katholieke Universiteit Leuven in the frame of the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office IUAP P6/04 (DYSCO, 'Dynamical systems, control and optimization', 2007-2011), Concerted Research Action GOA-MaNet and research project FWO nr. G.0600.08 ('Signal processing and network design for wireless acoustic sensor networks'). The scientific responsibility is assumed by its authors.

frames. The correlation matrix estimates $\mathbf{R}_y^{est}$ and $\mathbf{R}_v^{est}$ are then recursively updated (per frequency bin) as

$$\mathbf{R}_y^{est}[m+1] = \lambda_y \mathbf{R}_y^{est}[m] + (1-\lambda_y)\mathbf{Y}[m+1]\mathbf{Y}^H[m+1] \quad (3)$$

$$\mathbf{R}_v^{est}[m+1] = \lambda_v \mathbf{R}_v^{est}[m] + (1-\lambda_v)\mathbf{V}[m+1]\mathbf{V}^H[m+1] \quad (4)$$

in speech + noise frames and noise-only frames respectively. λ_y and λ_v are forgetting factors (usually chosen close to 1), and m is the frame-index. Assuming that the speech and the noise components are uncorrelated, the speech correlation matrix can be found as $\mathbf{R}_x^{est} = \mathbf{R}_y^{est} - \mathbf{R}_v^{est}$.

The noise reduction algorithms considered here are based on a linear filtering of the microphone signals by a filter $\mathbf{W}$ so that an output signal Z is obtained as $Z=\mathbf{W}^H\mathbf{Y}$. The goal of the noise reduction procedure is to minimize the distance between this output signal and the speech component in one of the microphone signals (unknown reference signal X_{ref}, e.g. $X_{\text{ref}} = X_1$).

2.2 SDW-MWF and related filters

In [3], it is shown that by minimizing a residual noise MSE cost function, while keeping the speech distortion below a certain threshold, the following filter is found:

$$\mathbf{W}_{SDW-MWF} = (\mathbf{R}_x + \mu\mathbf{R}_v)^{-1}\mathbf{R}_x\mathbf{u}\,, \quad (5)$$

where $\mathbf{u}$ is a vector with one entry equal to one and all other entries equal to zero, so that $\mathbf{u}^H\mathbf{X}{=}X_{\text{ref}}$. This filter was introduced as the Speech-Distortion Weighted Multichannel Wiener Filter (**SDW-MWF**) [1]. The parameter μ allows a trade-off between speech distortion and noise reduction.
If a single target speech source is assumed, the speech correlation matrix $\mathbf{R}_x$ is a rank one matrix. In [3], it is shown that an alternative (but theoretically equivalent in the single target speech source case) SDW-MWF formula can then be derived (denoted here as rank one MWF or **R1-MWF**), which still only depends on the speech and noise second order statistics, i.e.

$$\mathbf{W}_{R1-MWF} = \mathbf{R}_v^{-1}\mathbf{R}_x\mathbf{u}\,.\,\frac{1}{\mu + \text{tr}\{\mathbf{R}_v^{-1}\mathbf{R}_x\}} \quad (6)$$

where tr$\{.\}$ is the trace operator. The fact that only $\mathbf{R}_v$ is inverted in this expression (in contrast to the general formula (5)) will be utilized in this paper to derive a robust QRD-RLS based algorithm. In [5], a related filter formula is analyzed, namely the spatial prediction MWF (**SP-MWF**). By first estimating a spatial prediction vector, the speech distortion can be forced to zero (corresponding to the case μ=0 in (6)), which results in:

$$\mathbf{W}_{SP-MWF} = \mathbf{R}_v^{-1}\mathbf{R}_x\mathbf{u}\,.\,\frac{\mathbf{u}^H\mathbf{R}_x\mathbf{u}}{\text{tr}\{\mathbf{R}_v^{-1}\mathbf{R}_x\mathbf{u}\mathbf{u}^H\mathbf{R}_x\}}\,. \quad (7)$$

It is also possible to incorporate a speech distortion parameter μ into the SP-MWF filter, thereby relaxing the minimum distortion hard constraint. By enforcing that the postfilters of the SP-MWF and R1-MWF are equal for a single target source, the speech distortion weighted SP-MWF becomes equal to:

$$\mathbf{W}_{SP-MWF} = \mathbf{R}_v^{-1}\mathbf{R}_x\mathbf{u}\,.\,\frac{\mathbf{u}^H\mathbf{R}_x\mathbf{u}}{\mu\;\mathbf{u}^H\mathbf{R}_x\mathbf{u} + \text{tr}\{\mathbf{R}_v^{-1}\mathbf{R}_x\mathbf{u}\mathbf{u}^H\mathbf{R}_x\}} \quad (8)$$

Here also, only $\mathbf{R}_v$ is inverted so that again a robust QRD-RLS based algorithm can be derived. When comparing (8) to (6), it can be seen that both filters can be decomposed into a spatial filter $\mathbf{R}_v^{-1}\mathbf{R}_x\mathbf{u}$, which is the same for both filters, and a single channel postfilter, which is different for both filters. As the postfilter in (8) does not require the full speech correlation matrix (only the reference column), in contrast to the postfilter in (6), the SP-MWF allows for a simpler QRD-RLS scheme.

3. FREQUENCY DOMAIN QRD-RLS NOISE REDUCTION

3.1 QRD-RLS implementation of R1-MWF

In [2], a QRD-RLS implementation based on the general filter formula (5) with $\mu = 1$ was proposed. Instead of the speech and noise correlation matrices, their Cholesky factors are stored and updated by a numerically robust procedure based on Givens transformations. A review of QRD updating and QRD-RLS can be found in [2]. As already mentioned, the problem with this approach is that it is derived for the particular case $\mu = 1$, such that effectively $\mathbf{R}_y$ is inverted. For $\mu \neq 1$, large circular noise buffers have to be used, which is not feasible in a hearing aid application. To work around this problem, we propose to use the R1-MWF formula (6) as a starting point. As only $\mathbf{R}_v$ is inverted, a QRD updating scheme is then possible even for $\mu \neq 1$.

By plugging $\mathbf{R}_x = \mathbf{R}_y - \mathbf{R}_v$ into (6), and by defining $\mathbf{M}_{vy} = \mathbf{R}_v^{-1}\mathbf{R}_y$, the following expression is obtained:

$$\mathbf{W}_{R1-MWF} = (\mathbf{M}_{vy} - \mathbf{I}_N)\mathbf{u}\,.\,\frac{1}{\mu + \text{tr}\{\mathbf{M}_{vy}\} - N}\,, \quad (9)$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix. The R1-MWF formula can thus be split into a spatial beamformer $(\mathbf{M}_{vy} - \mathbf{I}_N)\mathbf{u}$ followed by a (single channel) spectral postfilter, and both parts only depend on the unknown matrix $\mathbf{M}_{vy}$. By defining $\mathbf{R}_v = \mathbf{R}_{v\Delta}^H\mathbf{R}_{v\Delta}$ (i.e. $\mathbf{R}_{v\Delta}$ is the upper triangular Cholesky factor of $\mathbf{R}_v$) and $\mathbf{B} = \mathbf{R}_{v\Delta}^{-H}\mathbf{R}_y$, matrix $\mathbf{M}_{vy}$ is found by solving the following system of equations:

$$\mathbf{R}_{v\Delta}\,\mathbf{M}_{vy} = \mathbf{B} \quad (10)$$

As $\mathbf{R}_{v\Delta}$ is triangular, this can be done by backsubstitution. In the next section, it will be shown that $\mathbf{R}_{v\Delta}$ and $\mathbf{B}$ can be efficiently updated together by applying sequences of Givens rotations. As in other Wiener filtering based procedures, there are two modes of operation (noise-only and speech + noise), which will be described separately.

3.2 Noise-only mode

In noise-only mode, the noise correlation matrix is updated as in (4). First, a standard QRD updating procedure [2] can be used to update the Cholesky factor of the noise correlation matrix estimate (4), i.e.

$$\begin{pmatrix} \mathbf{0}_{1\times N} \\ \hdashline \mathbf{R}_{v\Delta}[m+1] \end{pmatrix} = \mathbf{Q}^H[m+1]\begin{pmatrix} \sqrt{1-\lambda_v}\,\mathbf{V}^H[m+1] \\ \hdashline \sqrt{\lambda_v}\,\mathbf{R}_{v\Delta}[m] \end{pmatrix} \quad (11)$$

where $\mathbf{0}_{1\times N}$ is an all-zero N-dimensional row vector, and where $\mathbf{Q}^H[m+1]$ can be constructed as a series of N Givens transformations [2]. As the processing is performed in the frequency domain, complex Givens transformations have to be calculated, for example as in [6]. The transformation matrix $\mathbf{Q}$ is then unitary, i.e. $\mathbf{Q}^H\mathbf{Q} = \mathbf{Q}\mathbf{Q}^H = \mathbf{I}_{N+1}$.
The matrix $\mathbf{B}[m]$ can then be updated to $\mathbf{B}[m+1]$ using the same matrix $\mathbf{Q}^H[m+1]$ as in (11), which is explained as follows. As $\mathbf{Q}^H[m+1]$ is unitary, the following expression holds [7]:

$$\begin{pmatrix} \mathbf{0}_{N\times 1} & \vdots & \frac{1}{\sqrt{\lambda_v}}\mathbf{R}_{v\Delta}^{-1}[m] \end{pmatrix}\mathbf{Q}[m+1]\,\mathbf{Q}^H[m+1] \times \begin{pmatrix} \sqrt{1-\lambda_v}\,\mathbf{V}^H[m+1] \\ \hdashline \sqrt{\lambda_v}\,\mathbf{R}_{v\Delta}[m] \end{pmatrix} = \mathbf{I}_N\,. \quad (12)$$

By plugging (11) into (12), we find that:

$$\begin{pmatrix} \mathbf{0}_{N\times 1} & \vdots & \frac{1}{\sqrt{\lambda_v}}\mathbf{R}_{v\Delta}^{-1}[m] \end{pmatrix}\mathbf{Q}[m+1] = \begin{pmatrix} * & \vdots & \mathbf{R}_{v\Delta}^{-1}[m+1] \end{pmatrix}, \quad (13)$$

where $*$ indicates 'don't care' entries, i.e. values which will not be used. By taking the Hermitian transpose of expression (13), and by multiplying with $\mathbf{R}_y[m+1]$, we obtain the following expression:

$$\begin{pmatrix} * \\ \hline \mathbf{R}_{v\Delta}^{-H}[m+1] \end{pmatrix} \mathbf{R}_y[m+1] =$$

$$\mathbf{Q}^H[m+1] \begin{pmatrix} \mathbf{0}_{1\times N} \\ \hline \frac{1}{\sqrt{\lambda_v}}\mathbf{R}_{v\Delta}^{-H}[m] \end{pmatrix} \mathbf{R}_y[m+1] \; . \tag{14}$$

As in noise-only mode $\mathbf{R}_y[m+1] = \mathbf{R}_y[m]$, we thus find an update formula for $\mathbf{B} = \mathbf{R}_{v\Delta}^{-H}\mathbf{R}_y$, i.e.

$$\begin{pmatrix} * \\ \hline \mathbf{B}[m+1] \end{pmatrix} = \mathbf{Q}^H[m+1] \begin{pmatrix} \mathbf{0}_{1\times N} \\ \hline \frac{1}{\sqrt{\lambda_v}}\mathbf{B}[m] \end{pmatrix} . \tag{15}$$

In conclusion, we see that $\mathbf{R}_{v\Delta}$ and $\mathbf{B}$ can be updated together using a series of N complex Givens rotations, i.e.

$$\left(\begin{array}{c|c} \mathbf{0}_{1\times N} & * \\ \hline \mathbf{R}_{v\Delta}[m+1] & \mathbf{B}[m+1] \end{array}\right) \tag{16}$$
$$= \quad \mathbf{Q}^H[m+1] \left(\begin{array}{c|c} \sqrt{1-\lambda_v}\,\mathbf{V}^H[m+1] & \mathbf{0}_{1\times N} \\ \hline \sqrt{\lambda_v}\,\mathbf{R}_{v\Delta}[m] & \frac{1}{\sqrt{\lambda_v}}\,\mathbf{B}[m] \end{array}\right) .$$

With the updated $\mathbf{R}_{v\Delta}$ and $\mathbf{B}$, equation (10) can then be solved for $\mathbf{M}_{vy}$, so that the new optimal R1-MWF filter can be computed.

3.3 Speech+noise mode

In speech+noise mode, the speech+noise correlation matrix is updated as in (3). However, as we are tracking $\mathbf{B}$ instead of $\mathbf{R}_y$, an update procedure for $\mathbf{B}$ is needed. As the noise correlation matrix is not updated, we can set $\mathbf{R}_{v\Delta}[m+1] = \mathbf{R}_{v\Delta}[m]$, so that

$$\mathbf{B}[m+1] = \lambda_y\,\mathbf{B}[m] + (1-\lambda_y)\,\bar{\mathbf{Y}}[m+1]\mathbf{Y}^H[m+1] \; , \tag{17}$$

with $\bar{\mathbf{Y}}[m+1] = \mathbf{R}_{v\Delta}^{-H}[m]\mathbf{Y}[m+1]$. In this update, $\bar{\mathbf{Y}}[m+1]$ can be efficiently calculated by solving

$$\mathbf{R}_{v\Delta}^{H}[m]\,\bar{\mathbf{Y}}[m+1] = \mathbf{Y}[m+1] \tag{18}$$

by a single backsubstitution.

Similarly to the adaptive algorithms in [1], the MWF will be kept fixed in speech+noise mode, however, this need not be the case.

3.4 QRD-RLS implementation of SP-MWF

In a similar manner, the SP-MWF can be realized with a QRD-RLS scheme. By working out (8) as in section 3.1, the following expression is found:

$$\mathbf{W}_{SP-MWF} = (\mathbf{m}_{vy} - \mathbf{u}) \, . \, \frac{\mathbf{r}_x^H\mathbf{u}}{\mathbf{r}_x^H\left(\mathbf{m}_{vy} + (\mu-1)\mathbf{u}\right)} \; , \tag{19}$$

where $\mathbf{r}_x$ is a column of the speech correlation matrix ($\mathbf{r}_x = \mathbf{R}_x\mathbf{u}$), and $\mathbf{m}_{vy}$ is a column of $\mathbf{M}_{vy}$ ($\mathbf{m}_{vy} = \mathbf{M}_{vy}\mathbf{u}$). Then, as $\mathbf{r}_x = \mathbf{R}_v(\mathbf{m}_{vy} - \mathbf{u})$, $\mathbf{R}_v = \mathbf{R}_{v\Delta}^H\mathbf{R}_{v\Delta}$ and $\mathbf{R}_{v\Delta}\mathbf{m}_{vy} = \mathbf{B}\mathbf{u} = \mathbf{b}$, this can finally be written as:

$$\mathbf{W}_{SP-MWF} = (\mathbf{m}_{vy} - \mathbf{u}) \, . \, \frac{<\mathbf{R}_{v\Delta}\mathbf{u} \, , \, \mathbf{b} - \mathbf{R}_{v\Delta}\mathbf{u}>}{<\mathbf{b} + (\mu-1)\mathbf{R}_{v\Delta}\mathbf{u} \, , \, \mathbf{b} - \mathbf{R}_{v\Delta}\mathbf{u}>} \; , \tag{20}$$

where the dotproduct $<\mathbf{v}_1, \mathbf{v}_2> = \mathbf{v}_2^H\mathbf{v}_1$.

Vector $\mathbf{m}_{vy}$ can be updated during noise-only periods in a similar way as matrix $\mathbf{M}_{vy}$ is updated for the R1-MWF filter. However, complexity is reduced here as only one column of $\mathbf{M}_{vy}$ is needed, so that only a single column of $\mathbf{B}$ has to be stored and updated. In contrast, the R1-MWF (9) requires the full matrix $\mathbf{M}_{vy}$ in order to calculate $\mathrm{tr}\{\mathbf{M}_{vy}\}$ in the single channel postfilter. The single channel postfilter of the SP-MWF requires the calculation of two dot-products, using vectors that are easily obtained from the (reduced) QRD-RLS scheme.

3.5 Residual extraction

In noise-only mode, it is also possible to obtain the output of the (spatial) filtering $(\mathbf{m}_{vy} - \mathbf{u})^H\mathbf{Y} = (\mathbf{m}_{vy} - \mathbf{u})^H\mathbf{V}$ directly from the QRD-RLS scheme, without having to solve (10). Namely, by extracting the least squares residuals as in [8], it can be shown that:

$$(\mathbf{m}_{vy} - \mathbf{u})^H\mathbf{V} = -V_{\mathrm{ref}} - \frac{1}{\sqrt{1-\lambda_v}}\,\varepsilon \prod_{n=1}^{N} \cos\theta_n \; , \tag{21}$$

where the $\cos\theta_n$ are found in the Givens rotation matrices, and where ε is a by-product of the QRD-RLS scheme, i.e. the value which was indicated with a $*$ in (16), above the reference column of $\mathbf{B}$. The final output is then found by multiplying (21) with the single channel postfilter of (9) or (20). As the postfilter of the R1-MWF (9) requires $\mathbf{M}_{vy}$ so that (10) still has to be solved, the residual extraction does not yield any benefit. This is however not the case for the SP-MWF, so that the SP-MWF allows for a further reduction of the computational complexity compared to the R1-MWF.

4. SIMULATIONS

4.1 Setup

We consider a binaural hearing aid configuration, i.e. two hearing aids connected by a wireless link. The link is assumed to be ideal in terms of bandwidth and power consumption. We therefore assume that all microphone signals are available as inputs to the noise reduction procedure. Two microphones are used in the left ear device and two in the right ear device, giving a total of $N=4$. The binaural procedure produces a stereo output, but only the output for the left ear device will be shown. The left-front microphone is then chosen as the reference microphone.

Head-related transfer functions (HRTF's) were measured in a reverberant room (reverberation time RT_{60}=0.61s, cfr. [9]) on a dummy-head, so that the head-shadow effect is taken into account. To generate the microphone signals, the noise and speech signals are convolved with the HRTF's corresponding to their angles of arrival, before being added together. 11 different speech-noise configurations were tested, where the azimuthal angles (defined clockwise with 0° as frontal direction) of the noise source(s) are varied. The speech source is always at 0°, except for the last scenario where it is at 270° (to the left of the head). The first six scenarios have a single noise source at an angle between 60° and 300°. Scenarios N2a, N2b and N2c have two noise sources at $[-60°, 60°]$, $[-120°, 120°]$ and $[120°, 210°]$ respectively. Scenario N4 has four noise sources at 60°, 120°, 180° and 210°. Finally, for scenario S270N180 the target speech source is at 270° and the noise source is at 180°.

For the noise (interference) signal(s) multitalker babble noise is used. The target speech signal consists of 6 instances of speech-shaped noise, with periods of silence (12 s of speech, total signal length 24s). Average spectra of the target and interference signals can be found in [9]. The stimuli were scaled to obtain an input SNR of 0 dBA.

To assess the impact on speech intelligibility, a speech intelligibility (SI) weighted SNR improvement is calculated [10], i.e.

$$\Delta\mathrm{SNR}_{SI} = \sum_i I_i\,(SNR_{i,\mathrm{out}} - SNR_{i,\mathrm{in}}) \; , \tag{22}$$

where the band importance function I_i expresses the importance of the ith one-third octave band with center frequency f_i^c for intelligibility. The last 12 seconds of the output signals are selected to measure the obtained output SNR, so that the performance after convergence can be assessed.

A comparison is made between the QRD-RLS algorithms proposed in this paper, and the adaptive frequency domain SDW-MWF algorithm in [1] (unconstrained block-structured step size implementation). All algorithms are implemented in a weighted overlap-add (WOLA) filterbank framework [11], as this is a flexible framework suitable for hearing aid applications. The signals are sampled at 20480 Hz, and are processed by 128-point FFT's (with a frame-overlap of 32 samples). The MWF-based algorithms are also compared with a (time-domain) implementation of the GSC [4]. The fixed beamformer and blocking matrix of the GSC preprocessing stage are calibrated assuming the target speech source is located at 0°. To avoid speech cancellation, the GSC filters are only updated in periods where the target speech source is inactive. The filterlength was chosen so that the total input-output delay of the GSC algorithm is equal to the input-output delay of the WOLA filterbank.

All tested algorithms require voice activity detection (VAD), which will be assumed to be perfect in these simulations.

4.2 SI weighted SNR improvement

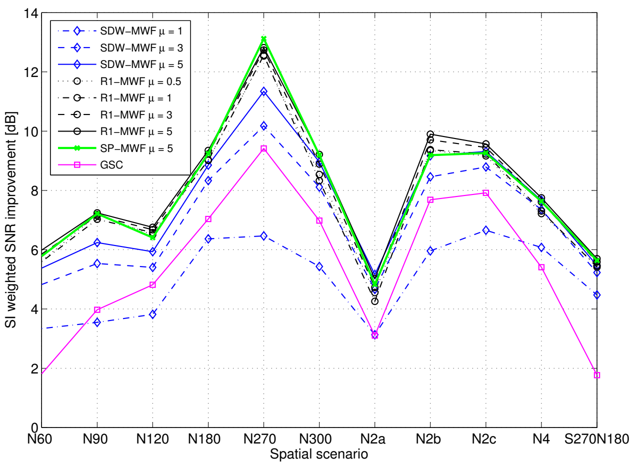

Figure 1: SI weighted SNR improvement at left output

In figure 1, the SI-weighted SNR improvement for the 11 different speech-noise scenarios is shown. The curves denoted with *SDW-MWF* are the performances obtained with the algorithm in [1], which is based on the general SDW-MWF formula (5). The curves denoted with *R1-MWF* and *SP-MWF* are the performances obtained with the QRD-RLS based algorithms for the filters (9) and (20) respectively. In order not to overload the figure, only the case $\mu = 5$ is shown for the SP-MWF. Finally, the curve denoted with *GSC* is the performance obtained with the GSC algorithm [4].

It can be observed that the R1-MWF (and SP-MWF) seems insensitive to changes in μ, with respect to speech intelligibility. This is actually expected, as in theory, the output SNR per frequency bin is independent of μ [5]. Therefore the intelligibility weighted SNR, where SNR values are measured per one-third octave band, should indeed not change significantly as μ changes.

In theory the SDW-MWF is equivalent to the R1-MWF and SP-MWF for a single target speech source so that its performance should also be independent of μ. However, figure 1 illustrates that in practice, the performance of the SDW-MWF algorithm is highly dependent on μ, i.e. if μ is chosen too small, the performance degrades. This effect was also observed in [5] where the performances of the batch filters were studied. The batch results indicated that the R1-MWF and SP-MWF are inherently more robust to errors in the estimated speech statistics than the SDW-MWF. The same effect is now also observed in the performance of the adaptive implementations.

Figure 1 also illustrates that the GSC algorithm is outperformed by the MWF algorithms. It was demonstrated in [12] that the GSC is particularly sensitive to microphone mismatch, in contrast to the MWF. In practice, microphones are rarely matched in phase and gain, even in a single hearing aid. For a binaural hearing application where the microphone signals of two separate hearing aids are combined, the microphone mismatch may be even more severe, which can explain the lower performance of the GSC in these simulations. Additionally, when the target speech location deviates from the assumed speech location (as for scenario S270N180), it can be seen from figure 1 that the GSC performance also degrades. Finally, we note that algorithms such as the GSC which make use of a fixed preprocessing stage, may also degrade localization performance, whereas a binaural MWF algorithm enables correct localization [9].

4.3 Impact of μ: single channel postfilter

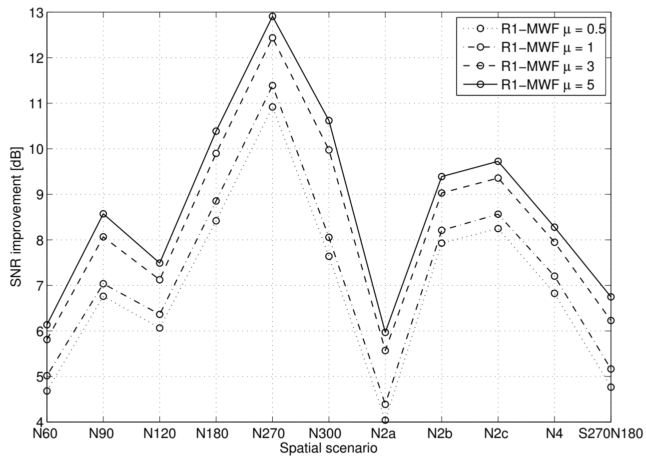

Figure 2: Broadband SNR improvement at left output

In figure 2, the broadband SNR improvement (i.e. the SNR calculated on the broadband time domain output signals, without SI weighting per one-third octave band) is shown for the R1-MWF[1], for different values of μ. As can be seen from (9), μ appears in the single channel spectral postfilter part, and therefore acts as in single microphone spectral subtraction algorithms [13]. If μ is increased, more residual noise is attenuated, hence increasing the broadband SNR by a few dB's. Although speech intelligibility is not improved (cfr. previous section), the listening comfort can be increased at the cost of more speech distortion.

A problem may arise when the estimated $\mathrm{tr}\{\mathbf{M}_{vy}\}$ takes too large or too small values. Constraining the postfilter between an upper and lower bound in this case, can give rise to musical noise artifacts, as is explained in [13]. This is especially the case when a small value of μ is chosen, as the postfilter value is then more dependent on $\mathrm{tr}\{\mathbf{M}_{vy}\}$. A possible solution would be to make μ dependent on the conditional speech presence probability as in [14]. In frequency bins where speech is absent, μ can be increased so that the residual noise is reduced and musical noise artifacts are also avoided, while the speech signal is not affected.

4.4 Robustness: effect of fixed wordlength

Figure 3 illustrates the effect of quantizing the values of the noise correlation matrix (or its Cholesky factor), for the spatial scenario N270 and for $\mu = 5$. The QRD-RLS based implementation of

[1] The SP-MWF with speech distortion extension (8) behaves similarly to R1-MWF, but seems slightly less aggressive. Namely, for the same value of μ, although less SNR improvement is obtained, the filter introduces less speech distortion.

the R1-MWF is compared to an algorithm without QRD-RLS, i.e. where the filter is calculated as in (6), using the noise correlation matrix estimate (4). The SNR performance of the QRD-RLS algorithm stays close to the optimal performance (i.e. the performance obtained without quantization, as shown in figure 2) when the wordlength is reduced, whereas the performance of the algorithm without QRD-RLS degrades.

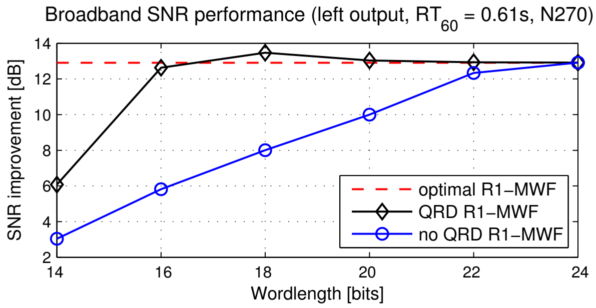

Figure 3: Effect of fixed wordlength in the noise correlation matrix

5. CONCLUSION

In this paper, we have shown that the adaptive frequency domain SDW-MWF can be realized with an efficient and robust QRD-RLS updating scheme.

Simulations on a binaural 4-microphone hearing aid setup show an improved speech intelligibility weighted SNR compared to the adaptive algorithm in [1], especially for small values of the trade-off parameter μ. Moreover, in contrast to the algorithm in [2], μ can be different from 1 without needing large circular buffers. The QRD-RLS algorithm can thus be used for smaller values of μ (low distortion beamforming), as it does not suffer from the same performance decrease as [1], but can also be used for larger values of μ, if the broadband SNR (and thus listening comfort) should be increased. Additionally, as the processing is performed in the frequency domain in contrast to the algorithm in [2], computational efficiency is increased. Finally, it was demonstrated that the QRD-RLS algorithm has a higher numerical robustness so that the wordlength can be reduced.

REFERENCES

[1] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, “Frequency-Domain Criterion for Speech Distortion Weighted Multichannel Wiener Filter for Robust Noise Reduction,” *Speech Communication, special issue on Speech Enhancement*, vol. 49, no. 7–8, pp. 636–656, Jul.-Aug. 2007.

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

[3] M. Souden, J. Benesty, and S. Affes, “On optimal frequency-domain multichannel linear filtering for noise reduction,” *IEEE Trans. Audio, Speech and Language Processing*, vol. 18, no. 2, pp. 260–276, Feb. 2010.

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

[5] B. Cornelis, M. Moonen, and J. Wouters, “Comparison of frequency domain noise reduction strategies based on Multichannel Wiener Filtering and Spatial Prediction,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Tai Pei, Taiwan, April 2009, pp. 129–132.

[6] D. Bindel, J. Demmel, W. Kahan, and O. Marques, “On computing Givens rotations reliably and efficiently,” *ACM Transactions on Mathematical Software (TOMS)*, vol. 28, no. 2, pp. 206–238, June 2002.

[7] C. Pan and R. Plemmons, “Least squares modifications with inverse factorization: parallel implications,” *J. Computational and Applied Mathematics*, vol. 27, no. 1-2, pp. 109–127, 1989.

[8] J. G. McWirther, “Recursive least squares mimimization using a systolic array,” *Proc. SPIE Real Time Signal Processing IV*, vol. 431, pp. 105–112, 1983.

[9] T. Van den Bogaert, S. Doclo, M. Moonen, and J. Wouters, “The effect of multi-microphone noise reduction systems on sound source localization in binaural hearing aids,” *Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 484–497, 2008.

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

[11] R. Brennan and T. Schneider, “A flexible filterbank structure for extensive signal manipulations in digital hearing aids,” in *Proc. IEEE Int. Symposium on Circuits and Systems (ISCAS)*, vol. 6, May-Jun 1998, pp. 569–572.

[12] A. Spriet, M. Moonen, and J. Wouters, “Robustness analysis of Multi-channel Wiener Filtering and Generalized Sidelobe Cancellation for multi-microphone noise reduction in hearing aid applications,” *IEEE Trans. Speech and Audio Processing*, vol. 13, no. 4, pp. 487–503, July 2005.

[13] P. C. Loizou, *Speech enhancement: Theory and Practice*. CRC press, New York, USA, 2007.

[14] K. Ngo, A. Spriet, M. Moonen, J. Wouters, and S. 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*, pp. 1–11, June 2009.