Katholieke Universiteit Leuven

Departement Elektrotechniek ESAT-SISTA/TR 10-81

Performance analysis of multichannel Wiener filter based noise reduction in hearing aids under second order statistics estimation errors [1]

Bram Cornelis[2] , Marc Moonen[2], Jan Wouters[3]

Published in IEEE Transactions on
Audio, Speech and Language Processing,
Vol. 19, No. 5, July 2011

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/sista/bcorneli/reports/IEEETranASL_MWFperf.pdf*. DOI: 10.1109/TASL.2010.2090519. (c) 2010 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

[2]K.U.Leuven, Dept. of Electrical Engineering (ESAT), Kasteelpark Arenberg 10, 3001 Leuven, Belgium. Tel. +32 16 321797, Fax +32 16 321970, WWW: *http://www.esat.kuleuven.ac.be/sista*, E-mail: *bram.cornelis@esat.kuleuven.ac.be*. 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.

[3]K.U.Leuven, Dept. of Neurosciences, ExpORL, Herestraat 49/721, 3000 Leuven, Belgium.

Abstract

The Speech Distortion Weighted Multichannel Wiener Filter (SDW-MWF) is a promising multi-microphone noise reduction technique, in particular for hearing aid applications. Its benefit over other single and multi-microphone techniques has been shown in several previous contributions, theoretically as well as experimentally. In theoretical studies, it is usually assumed that there is a single target speech source. The filter can then be decomposed into a conceptually interesting structure, i.e. into a spatial filter (related to other known techniques) and a single-channel postfilter, which then also allows for a performance analysis. Unfortunately, it is not straightforward to make a robust practical implementation based on this decomposition. Instead, a general SDW-MWF implementation, which only requires a (relatively easy) estimation of speech and noise correlation matrices, is mostly used in practice. This paper features a theoretical study and experimental validation on a binaural hearing aid setup of this standard SDW-MWF implementation, where the effect of estimation errors in the second order statistics is analyzed. In this case, and for a single target speech source, the standard SDW-MWF implementation is found not to behave as predicted theoretically. Second, two recently introduced alternative filters, namely the rank-one SDW-MWF and the spatial prediction SDW-MWF, are also studied in the presence of estimation errors in the second order statistics. These filters implicitly assume a single target speech source, but still only rely on the speech and noise correlation matrices. It is proven theoretically and illustrated through experiments that these alternative SDW-MWF implementations behave close to the theoretical optimum, and hence outperform the standard SDW-MWF implementation.

Performance analysis of multichannel Wiener filter based noise reduction in hearing aids under second order statistics estimation errors

Bram Cornelis*, *Student Member, IEEE*,
Marc Moonen, *Fellow, IEEE*

Katholieke Universiteit Leuven
Department of Electrical Engineering (ESAT–SCD)
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
Tel: +32/16/321927
E-mail: bram.cornelis@esat.kuleuven.be;
marc.moonen@esat.kuleuven.be

Jan Wouters

Katholieke Universiteit Leuven
Dept. of Neurosciences, ExpORL
Herestraat 49/721, 3000 Leuven, Belgium
E-mail: jan.wouters@med.kuleuven.be

Copyright (c) 2010 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.

B. 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.

Abstract

The Speech Distortion Weighted Multichannel Wiener Filter (SDW-MWF) is a promising multi-microphone noise reduction technique, in particular for hearing aid applications. Its benefit over other single and multi-microphone techniques has been shown in several previous contributions, theoretically as well as experimentally. In theoretical studies, it is usually assumed that there is a single target speech source. The filter can then be decomposed into a conceptually interesting structure, i.e. into a spatial filter (related to other known techniques) and a single-channel postfilter, which then also allows for a performance analysis. Unfortunately, it is not straightforward to make a robust practical implementation based on this decomposition. Instead, a general SDW-MWF implementation, which only requires a (relatively easy) estimation of speech and noise correlation matrices, is mostly used in practice. This paper features a theoretical study and experimental validation on a binaural hearing aid setup of this standard SDW-MWF implementation, where the effect of estimation errors in the second order statistics is analyzed. In this case, and for a single target speech source, the standard SDW-MWF implementation is found not to behave as predicted theoretically. Second, two recently introduced alternative filters, namely the rank-one SDW-MWF and the spatial prediction SDW-MWF, are also studied in the presence of estimation errors in the second order statistics. These filters implicitly assume a single target speech source, but still only rely on the speech and noise correlation matrices. It is proven theoretically and illustrated through experiments that these alternative SDW-MWF implementations behave close to the theoretical optimum, and hence outperform the standard SDW-MWF implementation.

Index Terms

Noise reduction, speech enhancement, microphone arrays, hearing aids, binaural hearing aids, multichannel Wiener filtering

I. Introduction

A major problem for hearing aid users is the degradation of their speech understanding in a noisy environment. Sensorineural hearing loss is most often accompanied by a loss of spectral and temporal resolution in the auditory processing, which results in a SNR loss of about 4-10 dB [1], [2]. Noise reduction in hearing aids has therefore been an active area of research for several years. Besides hearing aids, noise reduction also has many other applications such as hands-free communications and teleconferencing.

The first noise reduction techniques applied in hearing aids were single-microphone techniques [3]. Although these techniques may improve the SNR, this comes at the price of a high speech distortion [4]. Maybe as a consequence, single-microphone noise reduction techniques do not seem to increase speech

intelligibility significantly [1], [5]. The noise reduction does increase the overall listening comfort, so that hearing aid users generally find single-microphone noise reduction useful [5]. Single-microphone noise reduction is therefore usually included as a postfiltering stage.

In order to achieve the initial goal of increasing the speech intelligibility, hearing aids are fitted with multiple microphones, so that spatial information can be utilized in addition to temporal and spectral information to reduce the noise. Theoretically, a SNR improvement can then be achieved without distorting the target speech signal [4]. In practice, a speech intelligibility improvement can indeed be obtained [1], unlike with single-microphone techniques.

A popular multi-microphone noise reduction technique is the linearly constrained minimum variance (*LCMV*) beamformer. The LCMV minimizes the output power while imposing linear constraints on the beamformer response towards a target direction. The problem can be transformed into an easier unconstrained optimization problem by using the Generalized Sidelobe Canceller (*GSC*) technique [6]. The initial approach assumed freefield propagation, but this was extended into arbitrary transfer functions in the Transfer Function GSC (*TF-GSC*) technique [7]. For hearing aid applications, the GSC technique may be viewed as the current state of the art, and it leads to a significant benefit in certain scenarios [8]. However, it relies on a priori knowledge or assumptions about the target signal location and microphone characteristics. These assumptions are usually violated in practice so that performance may degrade significantly [9].

A different class of multi-microphone noise reduction techniques is based on Multichannel Wiener Filtering (MWF) [10]–[12], which is basically a generalization of single-channel procedures [3], [13]. The MWF produces a minimum-mean-square-error (MMSE) estimate of the speech component in a reference microphone signal, by exploiting speech and noise correlation matrices. To provide an explicit tradeoff between speech distortion and noise reduction, the Speech Distortion Weighted Multichannel Wiener Filter (*SDW-MWF*) was also proposed in [10]–[12]. This extension is equivalent to applying additional single-channel noise reduction to the spatial filter output, which, as already mentioned, is generally considered useful by hearing aid users. As the SDW-MWF does not require a priori knowledge or assumptions about the target signal location and microphone characteristics unlike the GSC, it is expected to be more robust, which was indeed demonstrated in [9]. A frequency-domain adaptive implementation of the SDW-MWF was also proposed in [12]. This approach is computationally advantageous as every frequency bin can be processed separately. The SDW-MWF thus offers a robust and computationally efficient alternative to the GSC.

In the near-future, the number of microphone signals available in a hearing aid will increase as a

wireless link will allow for exchanging signals between a left and a right hearing aid [1]. Recent research work was therefore focused on noise reduction techniques for such binaural hearing aids [14]–[26]. The noise reduction should then also preserve the so-called binaural cues, which are used by the human brain to localize sounds [27]. Correct sound localization (of both speech and noise sources) is an important goal by itself, but can also further improve speech intelligibility [28]. In the context of binaural hearing aids, it was shown that the SDW-MWF can be extended so that both the speech and the noise binaural cues can be preserved [29], [24], [26]. Hence, the SDW-MWF also offers a valuable approach to binaural noise reduction.

The SDW-MWF and its related filters have been thoroughly studied in previous theoretical work (for example [4], [29]–[33]). In the analysis, it is often assumed that there is a single target speech source. As a consequence, the frequency-domain speech correlation matrices are rank-one matrices. A closed-form expression for the SDW-MWF can then be found, which explicitly depends on the speech power and steering vector [29]. Using this closed-form expression, the SDW-MWF can be related to the TF-GSC [7], i.e. the SDW-MWF is equivalent to the TF-GSC followed by a single-channel spectral postfilter [30]. Recent theoretical contributions provide alternative closed-form expressions (still assuming a single target source) for the TF-GSC [32] or more generally for the SDW-MWF [33], that are also structured as a spatial filter followed by a single-channel postfilter. These expressions do not depend explicitly on the speech power and steering vector, but only make use of the speech and noise second order statistics. Using these expressions, the trade-off between speech distortion and noise reduction was quantified in [33], in analogy to the single-channel case [4].

The decomposition into a spatial filter and spectral postfilter is conceptually interesting, as for example, the spatial filter and postfilter can then be updated at different rates, or extended independently with other features. Unfortunately the closed-form expression proposed in [29] does not allow for a practical implementation as the speech power and steering vector would have to be calibrated or somehow estimated. Alternatively, the SDW-MWF could be implemented as a TF-GSC followed by a single-channel Wiener postfilter [34], but such implementations would then suffer from the robustness issues of the GSC [9]. The standard SDW-MWF implementation is based on the general SDW-MWF expression as in [10]–[12]. As it does not assume a single target speech source, the filter is not in a decomposed structure. As a result, the implementation only requires estimation of the speech and noise correlation matrices. However, as the closed-form expressions in [33] also only depend on the speech and noise correlation matrices, a robust implementation can be derived which is similar to the standard SDW-MWF implementation, but then structured as a spatial filter and a spectral postfilter.

In this paper the performance of the standard SDW-MWF implementation and two implementations based on the closed-form expressions in [33] is studied theoretically and through experiments on a binaural hearing aid setup. The effect of estimation errors in the second order statistics is included in the theoretical analysis as well as in the experiments. It is shown that the filter implementations behave differently in the presence of these estimation errors.

For a single target source, it is proven that the standard SDW-MWF implementation does not behave as predicted theoretically. In particular, when estimation errors are present in the speech correlation matrix, the single-frequency SNR improvement (obtained by spatial filtering) is shown to be dependent on the so-called speech distortion parameter, in contrast to the theoretical performance. Moreover, the intelligibility improvement is smaller than expected, especially when a small speech distortion parameter value is chosen.

The performance of the two alternative SDW-MWF implementations, which assume a single target speech source and are structured as a spatial filter followed by a single-channel postfilter, is also studied. The alternative SDW-MWF's are referred to as the rank one SDW-MWF (*R1-MWF*), which is based on the filter in [33], and the Spatial Prediction SDW-MWF (*SP-MWF*), which is an extension of the filter in [35], [36]. It will be shown that these implementations perform close to the optimal (theoretical) performance, and moreover, that they outperform the standard SDW-MWF implementation. In particular, in order to obtain a similar amount of noise reduction, the standard SDW-MWF implementation introduces more speech distortion than the R1-MWF and SP-MWF implementations. A simulation on scenarios with more than one target source, where the rank-one assumption is violated, is also performed. The R1-MWF and SP-MWF implementations still obtain a large SNR improvement compared to the standard SDW-MWF implementation, but they introduce different distortions. It is shown that the R1-MWF implementation attenuates the dominant sources, while the SP-MWF implementation tries to preserve the dominant sources. The SP-MWF implementation therefore introduces the least overall distortion.

The remainder of the paper is organized as follows. In section II, the notation and noise reduction configuration is introduced. The SDW-MWF is briefly reviewed and the alternative SDW-MWF's (R1-MWF and SP-MWF) are also described. In section III, theoretical expressions are derived for the three implementations, where estimation errors in the second order statistics are taken into account. These expressions are used in section IV to obtain expressions for the output SNR's of the different filter implementations. The theoretical results are validated by experiments in section V, both for a single target speech source scenario as for a more general scenario. Finally, overall conclusions are drawn in section VI.

II. Configuration, notation and review of multichannel Wiener filter

A. SDW-MWF

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 target speech component and $V_n(\omega)$ represents the noise component in the nth microphone. For conciseness, we will 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 $\mathbf{y} = \mathbf{x} + \mathbf{v}$. One of the microphone signals is used as the so-called reference microphone signal for the noise reduction algorithms. The reference microphone signal is denoted as Y_{ref} and is then equal to $Y_{\text{ref}} = \mathbf{e}_{\text{ref}}^H \mathbf{y}$, where $\mathbf{e}_{\text{ref}} = [0 \dots 0 \;\; 1 \;\; 0 \dots 0]^T$ is an N-dimensional vector where the entry corresponding to the reference microphone is equal to one. The reference microphone signal can also be written as a sum of a speech and noise component, i.e. $Y_{\text{ref}} = X_{\text{ref}} + V_{\text{ref}}$. The correlation matrix $\mathbf{R}_y$, the speech correlation matrix $\mathbf{R}_x$ and the noise correlation matrix $\mathbf{R}_v$ are defined as

$$\mathbf{R}_y = \mathcal{E}\{\mathbf{y}\mathbf{y}^H\}, \quad \mathbf{R}_x = \mathcal{E}\{\mathbf{x}\mathbf{x}^H\}, \quad \mathbf{R}_v = \mathcal{E}\{\mathbf{v}\mathbf{v}^H\}\,, \tag{2}$$

where $\mathcal{E}$ denotes the expected value operator. Assuming that the speech and the noise components are uncorrelated, we have $\mathbf{R}_y = \mathbf{R}_x + \mathbf{R}_v$. 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 Multichannel Wiener Filter (MWF) produces a minimum-mean-square-error (MMSE) estimate of the speech component in the reference microphone, hence simultaneously reducing noise and limiting speech distortion. To provide a more explicit tradeoff between speech distortion and noise reduction, the Speech Distortion Weighted Multichannel Wiener Filter (SDW-MWF) has been proposed, which minimizes a weighted sum of the residual noise energy and the speech distortion energy [10]–[12]. The SDW-MWF[1] cost function is equal to:

$$J_{\text{MWF}} = \mathcal{E}\left\{\left|X_{\text{ref}} - \mathbf{w}^H\mathbf{x}\right|^2\right\} + \mu\, \mathcal{E}\left\{\left|\mathbf{w}^H\mathbf{v}\right|^2\right\}\,. \tag{3}$$

The trade-off parameter μ allows putting more emphasis on noise reduction, at the cost of a higher speech distortion. We will therefore refer to μ as the speech distortion parameter. The MMSE estimator

[1]For conciseness, SDW-MWF is abbreviated to MWF when used as a subscript.

is obtained for $\mu = 1$. The SDW-MWF which minimizes (3) is given by the following expression[2]:

$$\boxed{\mathbf{w}_{\mathrm{MWF}} = \left(\mathbf{R}_x + \mu \mathbf{R}_v\right)^{-1} \mathbf{R}_x \mathbf{e}_{\mathrm{ref}}} \tag{4}$$

The narrowband (single-frequency) *input SNR* is defined as the power ratio of the speech and noise component in the reference microphone signal, i.e.

$$\mathrm{SNR}_{\mathrm{in}} \quad = \quad \frac{\mathcal{E}\{|X_{\mathrm{ref}}|^2\}}{\mathcal{E}\{|V_{\mathrm{ref}}|^2\}} = \frac{\mathbf{e}_{\mathrm{ref}}^H \mathbf{R}_x \mathbf{e}_{\mathrm{ref}}}{\mathbf{e}_{\mathrm{ref}}^H \mathbf{R}_v \mathbf{e}_{\mathrm{ref}}}, \tag{5}$$

and the narrowband (single frequency) *output SNR* is defined as the power ratio of the speech and noise component in the output signal, i.e.

$$\mathrm{SNR}_{\mathrm{out}} \quad = \quad \frac{\mathcal{E}\{|Z_x|^2\}}{\mathcal{E}\{|Z_v|^2\}} = \frac{\mathbf{w}^H \mathbf{R}_x \mathbf{w}}{\mathbf{w}^H \mathbf{R}_v \mathbf{w}} \ . \tag{6}$$

The (single-frequency) SNR improvement is then calculated as

$$\Delta\mathrm{SNR} = \frac{\mathrm{SNR}_{\mathrm{out}}}{\mathrm{SNR}_{\mathrm{in}}}. \tag{7}$$

B. Special case: single target source

In the case of a single target speech source, the speech signal vector can be modeled as

$$\mathbf{x} = \mathbf{a} S \ , \tag{8}$$

where the N-dimensional steering vector $\mathbf{a}$ contains the acoustic transfer functions from the speech source to the microphones (including room acoustics, microphone characteristics and head shadow effect) and S denotes the speech signal.

The speech correlation matrix is then a rank-one matrix, i.e.

$$\mathbf{R}_x = P_s \mathbf{a}\mathbf{a}^H \ , \tag{9}$$

with $P_s = \mathcal{E}\{|S|^2\}$ the power of the speech signal.

By assuming a single speech source and by applying the matrix inversion lemma, it has been shown [29] that the SDW-MWF (4) reduces to the following optimal filter:

$$\boxed{\mathbf{w}^{\mathrm{opt.}} \quad = \quad \mathbf{R}_v^{-1}\mathbf{a} \ . \ \frac{P_s \ A_{\mathrm{ref}}^*}{\mu + \rho}} \tag{10}$$

[2]We note that all frequency-domain filter expressions in this paper yield non-causal filters. In a practical implementation, the filters thus have to be adjusted to be causal, as will be explained in section V.

with $A_{\text{ref}}^{*} = \mathbf{a}^H \mathbf{e}_{\text{ref}}$ and $\rho = P_s \mathbf{a}^H \mathbf{R}_v^{-1} \mathbf{a}$. Using definition (6), the narrowband output SNR is then equal to

$$\text{SNR}_{\text{out}}^{\text{opt.}} = \rho = P_s \mathbf{a}^H \mathbf{R}_v^{-1} \mathbf{a} \, . \tag{11}$$

As is shown in [11], [30], the rank-one filter (10) can be decomposed into a spatial filter, which is equivalent to the TF-GSC filter [7], and a single-channel postfilter. The speech distortion parameter μ only appears in the single-channel postfilter and fulfills the same role as in single-channel constrained Wiener filters [3], [13] or spectral oversubtraction [3], [37]. As the filters obtained for different values of μ are related by a scalar factor, the output SNR per frequency bin (11) is indeed independent of μ. However, larger values of μ allow further attenuation of the residual noise, thus leading to a better listening comfort, at the cost of a higher speech distortion. The fact that the spatial filter is independent of μ is a desirable property. However, we will show in the following section that this property is lost in a SDW-MWF implementation (4) with estimation errors in the speech correlation matrix $\mathbf{R}_x$.

C. Rank-one SDW-MWF

Formula (10) incorporates prior knowledge (single target speech source) to obtain an alternative for the general SDW-MWF expression (4). It requires explicit estimation (or prior knowledge) of the steering vector $\mathbf{a}$ and the speech power P_s. It is however possible to derive an alternative expression which only uses the speech and noise second order statistics [32], [33], similar to the general expression (4).

By rewriting ρ as

$$\begin{aligned} \rho &= P_s \ \mathbf{a}^H \mathbf{R}_v^{-1} \mathbf{a} \, , & (12) \\ &= P_s \ \text{Tr}\{\mathbf{R}_v^{-1} \mathbf{a}\mathbf{a}^H\} \, , & (13) \\ &= \text{Tr}\{\mathbf{R}_v^{-1} \mathbf{R}_x\} \, , & (14) \end{aligned}$$

where $\text{Tr}\{.\}$ is the trace operator, (10) is equivalent to the following rank-one SDW-MWF (R1-MWF) expression:

$$\boxed{\mathbf{w}_{\text{R1-MWF}} = \mathbf{R}_v^{-1} \mathbf{R}_x \mathbf{e}_{\text{ref}} \ . \ \frac{1}{\mu + \text{Tr}\{\mathbf{R}_v^{-1} \mathbf{R}_x\}}} \tag{15}$$

Although (15) is derived for the special case of a single target speech source, it can also be used when this assumption is not fulfilled. Otherwise, for a single target speech source case, it is completely equivalent to (10).

D. Spatial Prediction SDW-MWF

The minimum distortion Spatial Prediction MWF (SP-MWF) was discussed in [36], and was originally proposed in [35] under the name Distortionless Multichannel Wiener Filter. It can be viewed as a frequency-domain version of the spatial-temporal prediction approach [31], [38]. For a single target speech source this filter is theoretically equivalent to the TF-GSC approach [7], or the R1-MWF (10), (15), where $\mu = 0$.

It is assumed that the N speech components can be related to the speech component in the reference microphone signal, i.e. $X_n = \mathcal{H}_{n,\text{ref}}\ X_{\text{ref}}$, for $n = 1...N$, so that

$$\mathbf{x} = \begin{bmatrix} \mathcal{H}_{1,\text{ref}} \\ \vdots \\ \mathcal{H}_{N,\text{ref}} \end{bmatrix} X_{\text{ref}} = \mathbf{h}\ X_{\text{ref}}\ , \tag{16}$$

For a single target speech source, $\mathbf{h}$ is then equal to[3]

$$\mathbf{h}^{\text{opt.}} = \mathbf{a}\frac{1}{A_{\text{ref}}}\ . \tag{17}$$

We only make use of the spatial correlations between the speech components, hence only a spatial prediction is performed. The spatial prediction vector $\mathbf{h}$ can be found in the Wiener sense, i.e. by minimizing

$$\min_{\mathbf{h}}\ \ \mathcal{E}\left\{(\mathbf{x} - \mathbf{h}\ X_{\text{ref}})^H\ (\mathbf{x} - \mathbf{h}\ X_{\text{ref}})\right\} \tag{18}$$

which leads to

$$\mathbf{h} = \frac{1}{\mathbf{e}_{\text{ref}}^H \mathbf{R}_x \mathbf{e}_{\text{ref}}}\ \mathbf{R}_x \mathbf{e}_{\text{ref}} \tag{19}$$

i.e. one column of the speech correlation matrix is selected and divided by the speech component power in the reference microphone. We can now impose the speech distortion to be zero, which leads to the following constrained optimization problem [36]:

$$\min_{\mathbf{w}} \qquad \mathbf{w}^H \mathbf{R}_v \mathbf{w} \tag{20}$$

$$\text{s.t.} \qquad \mathbf{w}^H \mathbf{h} = 1\ . \tag{21}$$

It is easily shown that the optimal filter is equal to

$$\mathbf{w} = \mathbf{R}_v^{-1}\mathbf{h}\ .\ \frac{1}{\mathbf{h}^H \mathbf{R}_v^{-1}\mathbf{h}}\ , \tag{22}$$

[3]We note that the denominator of this formula (and also of subsequent formula's) might decay to 0 at certain frequencies due to speech absence [33]. The denominator values are therefore kept above certain thresholds in a practical implementation.

or by plugging (19) into (22),

$$\mathbf{w} = \mathbf{R}_v^{-1}\mathbf{R}_x\mathbf{e}_{\text{ref}} \cdot \frac{\mathbf{e}_{\text{ref}}^H\mathbf{R}_x\mathbf{e}_{\text{ref}}}{\text{Tr}\{\mathbf{R}_v^{-1}\mathbf{R}_x\mathbf{e}_{\text{ref}}\mathbf{e}_{\text{ref}}^H\mathbf{R}_x\}} \tag{23}$$

Compared with the R1-MWF (15), expression (23) has the same spatial filter, but the single-channel postfilter is different. It is also possible to incorporate a speech distortion parameter μ into (23), thereby relaxing the minimum distortion hard constraint. This filter will be referred to as the SP-MWF. 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 expression is obtained as

$$\boxed{\mathbf{w}_{\text{SP-MWF}} = \mathbf{R}_v^{-1}\mathbf{R}_x\mathbf{e}_{\text{ref}} \cdot \frac{\mathbf{e}_{\text{ref}}^H\mathbf{R}_x\mathbf{e}_{\text{ref}}}{\mu\ \mathbf{e}_{\text{ref}}^H\mathbf{R}_x\mathbf{e}_{\text{ref}} + \text{Tr}\{\mathbf{R}_v^{-1}\mathbf{R}_x\mathbf{e}_{\text{ref}}\mathbf{e}_{\text{ref}}^H\mathbf{R}_x\}}} \tag{24}$$

For a single target source, (24) is thus again equivalent to (10).

III. Impact of speech correlation matrix estimation errors: Estimated filters

In this section, the impact of estimation errors in the speech correlation matrix $\mathbf{R}_x$ in the implementations of (4), (15) and (24), is investigated for a scenario with a single target speech source.

In practice, a voice activity detector (VAD) has to be implemented to distinguish between segments where speech and noise are both active and segments where only noise is active. The correlation matrix $\hat{\mathbf{R}}_y$ is then estimated[4] as

$$\hat{\mathbf{R}}_y(\omega) = \frac{1}{K}\sum_{k=1}^{K}\mathbf{y}(k,\omega)\ \mathbf{y}^H(k,\omega)\ , \tag{25}$$

where the summation only counts the segments where both speech and noise are active. In a similar fashion, the noise correlation matrix is estimated during the noise-only segments. The speech correlation matrix estimate is found as $\hat{\mathbf{R}}_x = \hat{\mathbf{R}}_y - \hat{\mathbf{R}}_v$. In order to obtain practical implementations, the estimated correlation matrices $\hat{\mathbf{R}}_v$ and $\hat{\mathbf{R}}_x$ are then plugged into (4), (15) and (24) instead of the theoretical $\mathbf{R}_v$ and $\mathbf{R}_x$.

Inaccurate estimation of the speech statistics occurs because of several reasons [9]. The speech and noise may be nonstationary, while $\hat{\mathbf{R}}_y$ and $\hat{\mathbf{R}}_v$ are estimated at different moments in time. Speech detection errors made by the VAD will also introduce estimation errors in both the speech and the noise correlation matrices.

[4]We note that in practice, the correlation matrix estimate (25) can also be recursively updated using an exponential weighting factor which is typically close to one [12].

In the case of a single target speech source, the speech correlation matrix $\mathbf{R}_x$ is a rank-one matrix and given by (9), and the SDW-MWF is given by (10). However, the estimated speech correlation matrix $\hat{\mathbf{R}}_x$ will be

$$\hat{\mathbf{R}}_x = P_s \mathbf{a}\mathbf{a}^H + \mathbf{\Delta} \ , \tag{26}$$

i.e. $\hat{\mathbf{R}}_x$ will be equal to the theoretical rank-one matrix plus a full rank (Hermitian) error matrix $\mathbf{\Delta}$. Formula (26) will be plugged into formula's (4), (15) and (24) to analyze the impact of speech correlation matrix estimation errors.

It is noted that the impact of estimation errors in the noise correlation matrix can be investigated in a similar fashion. However, simulations indicate that the SNR performance degradation caused by these errors is the same for the different filters, and independent of μ, in contrast to the impact of speech correlation matrix estimation errors. Therefore, and also to avoid overly complicated expressions, only speech correlation matrix estimation errors are included in the analysis.

A. SDW-MWF

By plugging (26) into (4) and applying the matrix inversion lemma, we obtain the following formula:

$$\begin{aligned} \hat{\mathbf{w}} &= \left(P_s \mathbf{a}\mathbf{a}^H + \mathbf{\Delta} + \mu \mathbf{R}_v\right)^{-1} \left(P_s \mathbf{a}\mathbf{a}^H + \mathbf{\Delta}\right) \mathbf{e}_{\text{ref}} && (27) \\ &= \frac{P_s}{1+\tilde{\rho}} (\mathbf{\Delta} + \mu \mathbf{R}_v)^{-1} \mathbf{a} \ A^*_{\text{ref}} \\ &\quad + \left[\mathbf{I} - \frac{P_s}{1+\tilde{\rho}} (\mathbf{\Delta} + \mu \mathbf{R}_v)^{-1} \mathbf{a}\mathbf{a}^H\right] (\mathbf{\Delta} + \mu \mathbf{R}_v)^{-1} \mathbf{\Delta} \ \mathbf{e}_{\text{ref}} && (28) \end{aligned}$$

with

$$\tilde{\rho} = P_s \mathbf{a}^H (\mathbf{\Delta} + \mu \mathbf{R}_v)^{-1} \mathbf{a} \ . \tag{29}$$

Two limit cases are now considered.

- $\mathbf{\Delta} >> \mu \mathbf{R}_v \implies (\mathbf{\Delta} + \mu \mathbf{R}_v)^{-1} \approx \mathbf{\Delta}^{-1}$

  In this case (28) reduces to

  $$\hat{\mathbf{w}} = \mathbf{e}_{\text{ref}} \ . \tag{30}$$

  This means that if a small μ parameter is chosen or if the input SNR is high, the estimated SDW-MWF reduces to the trivial filter $\mathbf{e}_{\text{ref}}$ (i.e. pass the reference microphone signal, no noise reduction).
- $\mathbf{\Delta} << \mu \mathbf{R}_v \implies (\mathbf{\Delta} + \mu \mathbf{R}_v)^{-1} \approx \frac{1}{\mu} \mathbf{R}_v^{-1}$ and $\tilde{\rho} \approx \frac{1}{\mu} \rho$

  In this case (28) reduces to

  $$\hat{\mathbf{w}} = \mathbf{w}^{\text{opt.}} + \frac{1}{\mu} \left(\mathbf{I} - \frac{P_s}{\mu + \rho} \mathbf{R}_v^{-1} \mathbf{a}\mathbf{a}^H\right) \mathbf{R}_v^{-1} \mathbf{\Delta} \ \mathbf{e}_{\text{ref}} \tag{31}$$

where $\mathbf{w}^{\text{opt.}}$ is given by formula (10). Hence, for large values of μ or for a low input SNR, the estimated SDW-MWF will be a combination of the optimal theoretical filter and an extra bias term, which causes performance degradation. A larger μ value (which will result in more speech distortion) puts less weight on the bias term, so that it can be expected that the performance degradation will be smaller for larger μ values. In section IV, the output SNR obtained with (31) will be calculated.

B. Rank-one SDW-MWF

By plugging (26) into (15) and applying the matrix inversion lemma, the following formula is obtained:

$$\hat{\mathbf{w}} = \frac{\mathbf{R}_v^{-1}\left(P_s\mathbf{a}\mathbf{a}^H + \mathbf{\Delta}\right)\mathbf{e}_{\text{ref}}}{\mu + \text{Tr}\{\mathbf{R}_v^{-1}P_s\mathbf{a}\mathbf{a}^H + \mathbf{R}_v^{-1}\mathbf{\Delta}\}} \tag{32}$$

$$= \frac{\mu+\rho}{\mu+\rho+\text{Tr}\{\mathbf{R}_v^{-1}\mathbf{\Delta}\}} \cdot \mathbf{w}^{\text{opt.}} + \frac{1}{\mu+\rho+\text{Tr}\{\mathbf{R}_v^{-1}\mathbf{\Delta}\}} \cdot \mathbf{R}_v^{-1}\mathbf{\Delta}\mathbf{e}_{\text{ref}} \tag{33}$$

$$= \left(P_s\mathbf{R}_v^{-1}\mathbf{a}A_{\text{ref}}^* + \mathbf{R}_v^{-1}\mathbf{\Delta}\mathbf{e}_{\text{ref}}\right)\frac{1}{\mu+\rho+\text{Tr}\{\mathbf{R}_v^{-1}\mathbf{\Delta}\}} \tag{34}$$

Equation (33) shows that the estimated SDW-MWF can still be written as a combination of the optimal filter (10) and a bias term. Equation (34) shows that the estimated filter can also be written as a spatial filter followed by a single-channel postfilter, where the parameter μ only occurs in the single-channel postfilter. Therefore, μ will not influence the obtained narrowband (single-frequency) output SNR (6), which is similar to the optimal case (11).

C. Spatial prediction SDW-MWF

By plugging (26) into (19) and (24), we find

$$\hat{\mathbf{h}} = \frac{1}{\tilde{P}_s} \cdot \left(P_s\mathbf{a}A_{\text{ref}}^* + \mathbf{\Delta}\mathbf{e}_{\text{ref}}\right) \tag{35}$$

and

$$\hat{\mathbf{w}} = \left(P_s\mathbf{R}_v^{-1}\mathbf{a}A_{\text{ref}}^* + \mathbf{R}_v^{-1}\mathbf{\Delta}\mathbf{e}_{\text{ref}}\right)\frac{\tilde{P}_s}{\mu\tilde{P}_s + \rho P_s|A_{\text{ref}}|^2 + \rho' P_s A_{\text{ref}} + (\rho')^* P_s A_{\text{ref}}^* + \rho''} \tag{36}$$

where

$$\tilde{P}_s = P_s|A_{\text{ref}}|^2 + \mathbf{e}_{\text{ref}}^H\mathbf{\Delta}\mathbf{e}_{\text{ref}} , \tag{37}$$

$$\rho' = \mathbf{a}^H\mathbf{R}_v^{-1}\mathbf{\Delta}\mathbf{e}_{\text{ref}} , \tag{38}$$

$$\rho'' = \mathbf{e}_{\text{ref}}^H\mathbf{\Delta}^H\mathbf{R}_v^{-1}\mathbf{\Delta}\mathbf{e}_{\text{ref}} . \tag{39}$$

The estimated SP-MWF (36) can thus be written as a spatial filter followed by a single-channel postfilter. The spatial filter in (36) is equal to the spatial filter in (34), so that the estimated SP-MWF will obtain the same narrowband (single-frequency) output SNR as the estimated R1-MWF.

IV. IMPACT OF SPEECH CORRELATION MATRIX ESTIMATION ERRORS: OUTPUT SNR

In this section, the impact of speech correlation matrix estimation errors on the obtained output SNR will be analyzed. The estimated filters, derived in the previous section, are plugged into the narrowband output SNR definition (6). The rank-one model (9) is again used for the speech correlation matrix.

A. SDW-MWF

Again, we consider the two limit cases of the previous section.

- For the case $\mathbf{\Delta} >> \mu\mathbf{R}_v$, the trivial filter $\mathbf{e}_{\text{ref}}$ is obtained, so that $\widehat{\text{SNR}}_{\text{out}} = \text{SNR}_{\text{in}}$.
- For the case $\mathbf{\Delta} << \mu\mathbf{R}_v$, we can plug formula (28) into (6) to obtain:

$$\widehat{\text{SNR}}_{\text{out}} = \frac{\frac{1}{(\mu+\rho)^2}\left(P_s|A_{\text{ref}}|^2\rho^2 + P_s A_{\text{ref}}(\rho^{'})\rho + P_s A^*_{\text{ref}}(\rho^{'})^*\rho - \frac{2\mu+\rho}{\mu^2}P_s|\rho^{'}|^2\rho\right) + \frac{1}{\mu^2}P_s|\rho^{'}|^2}{\frac{1}{(\mu+\rho)^2}\left(P_s|A_{\text{ref}}|^2\rho + P_s A_{\text{ref}}(\rho^{'}) + P_s A^*_{\text{ref}}(\rho^{'})^* - \frac{2\mu+\rho}{\mu^2}P_s|\rho^{'}|^2\right) + \frac{1}{\mu^2}\rho^{''}} , \quad (40)$$

$$= \rho - \frac{\rho\rho^{''} - P_s|\rho^{'}|^2}{\frac{\mu^2}{(\mu+\rho)^2}\left(P_s|A_{\text{ref}}|^2\rho + P_s A_{\text{ref}}(\rho^{'}) + P_s A^*_{\text{ref}}(\rho^{'})^*\right) - \frac{2\mu+\rho}{(\mu+\rho)^2}P_s|\rho^{'}|^2 + \rho^{''}} , \quad (41)$$

$$= \rho - \frac{\rho\rho^{''} - P_s|\rho^{'}|^2}{\frac{\mu^2}{(\mu+\rho)^2}\left[P_s|A_{\text{ref}}|^2\rho + P_s A_{\text{ref}}(\rho^{'}) + P_s A^*_{\text{ref}}(\rho^{'})^* + \rho^{''} + \frac{2\mu+\rho}{\mu^2}(\rho\rho^{''} - P_s|\rho^{'}|^2)\right]} , \quad (42)$$

where $\rho^{'}$ and $\rho^{''}$ were defined in (38) and (39). Formula (42) shows that the obtained output SNR is equal to the optimal output SNR (11) minus a bias term. By defining matrix $\tilde{\mathbf{V}}$ such that $\tilde{\mathbf{V}}\tilde{\mathbf{V}}^H = \mathbf{R}_v^{-1}$, the bias term can also be written as

$$\frac{\frac{(\mu+\rho)^2}{\mu^2}(\rho\rho^{''} - P_s|\rho^{'}|^2)}{||P_s A_{\text{ref}}\mathbf{a}^H\tilde{\mathbf{V}} + \mathbf{e}_{\text{ref}}^H\mathbf{\Delta}^H\tilde{\mathbf{V}}||^2 + \frac{2\mu+\rho}{\mu^2}(\rho\rho^{''} - P_s|\rho^{'}|^2)} , \quad (43)$$

which clearly shows that the bias term is always positive if $\left(\rho\rho^{''} - P_s|\rho^{'}|^2\right)$ is positive. By using the Cauchy-Schwarz inequality, it can indeed be shown that $\left(\rho\rho^{''} - P_s|\rho^{'}|^2\right)$ is positive[5]:

$$\begin{aligned} \rho\rho^{''} &= P_s\left(\mathbf{a}^H\mathbf{R}_v^{-1}\mathbf{a}\right)\left(\mathbf{e}_{\text{ref}}^H\mathbf{\Delta}^H\mathbf{R}_v^{-1}\mathbf{\Delta}\mathbf{e}_{\text{ref}}\right) , \\ &\geq P_s|\mathbf{a}^H\mathbf{R}_v^{-1}\mathbf{\Delta}\mathbf{e}_{\text{ref}}|^2 = P_s|\rho^{'}|^2 . \end{aligned} \quad (44)$$

These results indicate that there will always be a performance degradation, which is moreover dependent on μ. As can be seen from (41), the denominator of the bias term monotonically increases

[5]The case $\rho\rho^{''} = P_s|\rho^{'}|^2$ occurs if $\mathbf{\Delta}\mathbf{e}_{\text{ref}}$ is a scaled version of the steering vector $\mathbf{a}$. Thus, if the speech correlation matrix is known up to a scaling factor, the optimal performance can be achieved.

when μ increases:

$$\begin{aligned} & \frac{\partial}{\partial \mu}\left(\frac{\mu^2}{(\mu+\rho)^2}\left[P_s|A_{\mathrm{ref}}|^2\rho + P_s A_{\mathrm{ref}}(\rho^{'}) + P_s A_{\mathrm{ref}}^*(\rho^{'})^*\right] - \frac{2\mu+\rho}{(\mu+\rho)^2}P_s|\rho^{'}|^2 + \rho^{''}\right) \\ = & \quad \frac{2\mu\rho}{(\mu+\rho)^3}\left[P_s|A_{\mathrm{ref}}|^2\rho + P_s A_{\mathrm{ref}}(\rho^{'}) + P_s A_{\mathrm{ref}}^*(\rho^{'})^*\right] + \frac{2\mu}{(\mu+\rho)^3}P_s|\rho^{'}|^2 \\ = & \quad \frac{2\mu}{(\mu+\rho)^3}P_s|\rho A_{\mathrm{ref}} + (\rho^{'})^*|^2 \ > 0\ . \end{aligned} \tag{45}$$

Therefore, the obtained output SNR monotonically increases as μ increases. By calculating the limit for $\mu \to \infty$, i.e.

$$\lim_{\mu\to\infty} \widehat{\mathrm{SNR}}_{\mathrm{out}} = \rho - \frac{\rho\rho^{''} - P_s|\rho^{'}|^2}{P_s|A_{\mathrm{ref}}|^2\rho + P_s A_{\mathrm{ref}}(\rho^{'}) + P_s A_{\mathrm{ref}}^*(\rho^{'})^* + \rho^{''}}\ , \tag{46}$$

we see that there is a maximum obtainable output SNR, which is equal to the theoretical output SNR minus a fixed bias term. Thus, to achieve this maximum output SNR with the SDW-MWF implementation, a large μ value should be used. This can however introduce too much speech distortion. Using a small μ value (low distortion) can also be unsatisfactory, as this will cause SNR performance degradation, especially so when $\mathbf{\Delta} >> \mu\mathbf{R}_v$.

B. Rank-one SDW-MWF and Spatial Prediction SDW-MWF

As was stated previously, the filters (34) and (36) consist of a spatial filter followed by a single-channel postfilter. The obtained narrowband output SNR only depends on the spatial filter part (which is independent of μ). As both filters have the same spatial filter, they can thus be treated together.
By plugging the spatial filter

$$\hat{\mathbf{w}} = P_s\mathbf{R}_v^{-1}\mathbf{a}A_{\mathrm{ref}}^* + \mathbf{R}_v^{-1}\mathbf{\Delta}\mathbf{e}_{\mathrm{ref}} \tag{47}$$

into definition (6), the following output SNR is obtained:

$$\widehat{\mathrm{SNR}}_{\mathrm{out}} = \rho - \frac{\rho\rho^{''} - P_s|\rho^{'}|^2}{P_s|A_{\mathrm{ref}}|^2\rho + P_s A_{\mathrm{ref}}(\rho^{'}) + P_s A_{\mathrm{ref}}^*(\rho^{'})^* + \rho^{''}}\ . \tag{48}$$

Remarkably, this output SNR is equal to the limit case (46) achieved by the SDW-MWF implementation for large values of μ.

C. Discussion

Under speech correlation matrix estimation errors, and for a scenario with a single target speech source, the R1-MWF (15) and SP-MWF (24) implementations always achieve the limit output SNR (46) (for any value of μ), whereas the SDW-MWF implementation (4) will only achieve this for large μ. In scenarios where only a single target speech source is present and where only limited speech distortions

(small μ's) are allowed, the SDW-MWF implementation is therefore outperformed by the other filter implementations. In particular, in order to obtain a similar output SNR, the SDW-MWF implementation will introduce a higher speech distortion. Experimental results will show that even for moderate values of μ (for example the standard MSE cost function, $\mu = 1$), a significant performance degradation occurs when using the SDW-MWF. This was also observed in [36], where the SP-MWF implementation (with $\mu = 0$) clearly outperformed the SDW-MWF implementation in both SNR and speech distortion. These practical results are now better explained by the above theoretical analysis.

For a scenario with multiple target speech sources (i.e. speech correlation matrix not a rank-one matrix), the R1-MWF and SP-MWF formula's are no longer theoretically equivalent to the general SDW-MWF formula. A theoretical analysis as for the rank-one case is not straightforward, but it can be expected that there will again be a performance degradation. To investigate this, the next section includes simulations with multiple target speech sources. These simulations show that the R1-MWF and SP-MWF implementations still obtain superior SNR's compared to the SDW-MWF implementation. The R1-MWF implementation will introduce more speech distortion however, especially when the number of target sources is large. Remarkably, the simulations also show that the SP-MWF implementation only introduces distortion in the less dominant target sources. As a consequence, the overall distortion (on the sum of all speech components) will still be low for this filter.

V. Experimental results

A. Setup and stimuli

We consider a binaural hearing aid configuration, i.e. two hearing aids connected by a wireless link [14]–[26]. It is assumed here that the link imposes no restrictions in terms of bandwidth and power consumption. We therefore assume that all microphone signals are available to the noise reduction procedure, where two microphones are at the left ear and two at the right ear, giving a total of $N = 4$. The left front and right front microphones are chosen as reference microphones to generate the left and right filters and output signals.

Head-related transfer functions (HRTF's) were measured in two acoustical environments (reverberation times RT_{60} = 0.21s and RT_{60} = 0.61s, [24], [25]) on a dummy-head, so that the head-shadow effect is taken into account. The behind-the-ear hearing aids have two omnidirectional microphones on each device, with an intermicrophone distance of approximately 1 cm. 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.

In all experiments, multi-talker babble (Auditec [39]) was used as noise signal. In the scenarios with multiple noise sources, different time-shifted versions of this signal were generated to obtain uncorrelated noise sources. In the single target speech source scenarios, a signal consisting of 6 instances of speech-shaped noise with periods of silence, was used as target signal (12 seconds of speech, total length 24 seconds), as in [24], [25]. The speech-shaped noise was obtained from the average spectrum of a Dutch male speaker of the VU test material [40]. In the multiple target speech sources scenarios, different lists of the VU test material were used for the different target signals, with each signal consisting of 4 sentences with periods of silence (total length 26 seconds). The different target speech sources are simultaneously active so that the SNR and distortion can be calculated on the global target signal, i.e. the sum of the speech components.

In all experiments, the signals are sampled at $f_s = 20480$ Hz, and the filter length (= DFT size) is $L = 128$. A batch procedure[6] was implemented where the speech and noise correlation matrices are estimated off-line using the complete microphone signals, as in (25). The microphone signals were hereby cut into frames of L samples with 50% overlap, and windowed by a Hann window. The value K in (25) is set to the total number of frames where speech is active. The noise correlation matrix is calculated on the noise-only frames in a similar manner. The voice activity detector (VAD) is assumed to be perfect. The filters (4), (15) and (24) are calculated using the estimated correlation matrices, as explained in section III. Following the approach in [42], the frequency-domain filters are then transformed into corresponding time-domain filters which do not rely on the circularity effect. An extra phase shift is also applied in the IDFT formula as in [42], so that the resulting filters are causal.

The input SNR (measured on the clean signals, i.e. at the loudspeaker) is 0 dB in all experiments. Due to the headshadow effect, the input SNR's at the left and right reference microphones then depend on the spatial scenario, i.e. the positions of the target speech source(s) and noise source(s).

B. Narrowband (single-frequency) SNR improvement

In a first experiment, the SNR improvement obtained in a single frequency bin (7) is calculated for the different filter implementations. The fifth frequency bin (corresponding to $f = 640$ Hz) is selected for the performance comparison. At this frequency, the speech (and noise) signals contain an intermediate amount of energy (see [24] for the average input spectra), while this frequency also has an intermediate importance for speech intelligibility [43]. A single noise source is placed at 120° (with 0° the front

[6]The R1-MWF and SP-MWF can also be implemented by an adaptive QRD-RLS scheme as is discussed in [41].

direction, 90° to the right of the head), the single target speech source is placed in front of the head at 0°.

The narrowband input and output SNR's are calculated by plugging the resulting filters into expressions (5) and (6), which make use of the (theoretical) rank one speech correlation matrix (9). The steering vector $\mathbf{a}$ is therefore constructed using the HRTF data, and P_s is estimated by calculating the PSD of the clean speech signal. The steering vector is also used to calculate the theoretical optimal filter (10). While the steering vector is calculated using a large DFT size, the noise correlation matrix used in the optimal filter is the same as for the other filters, and calculated at $L = 128$. In order to quantify the speech correlation matrix estimation error, we propose the error measure

$$\delta(\hat{\mathbf{R}}_x) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{|\mathbf{\Delta}_{ij}|}{|\mathbf{R}_{x,ij}|} \ , \tag{49}$$

where $\mathbf{\Delta}_{ij}$ is the entry in the ith row and jth column of matrix $\mathbf{\Delta}$, and similarly for $\mathbf{R}_{x,ij}$. For the case $N = 2$ (no binaural link), the estimation errors (average error of the left and right correlation matrix estimates) are then equal to $\delta(\hat{\mathbf{R}}_x) = 0.01$ for the low-reverberant environment and $\delta(\hat{\mathbf{R}}_x) = 0.05$ for the high-reverberant environment. For the case $N = 4$ (binaural setup, i.e. all microphone signals are available), the estimation errors are equal to $\delta(\hat{\mathbf{R}}_x) = 0.07$ for the low-reverberant environment, and $\delta(\hat{\mathbf{R}}_x) = 0.41$ for the high-reverberant environment. In a high-reverberant environment and for a binaural setup, a significant amount of estimation errors are thus introduced, so that a performance degradation can be expected.

Figures 1 (low reverberation) and 2 (high reverberation) show the SNR improvements obtained at the left and the right side as a function of the speech distortion parameter μ, for a setup with only 2 used microphones (i.e. no binaural link) and a setup where all 4 microphones are used to generate the outputs.

As discussed in section IV, there is a fixed performance degradation between the *optimal* performance and the performance of the *R1-MWF* implementation (48), especially so for the high-reverberant environment (figure 2). The SP-MWF implementation obtains the same SNR improvement as the R1-MWF, and is therefore omitted from the figures. As was explained, the trade-off parameter μ does not influence the narrowband SNR improvement for the R1-MWF and SP-MWF implementations. For the low-reverberant environment (figure 1), the estimation errors are small so that the R1-MWF performance is very close to the optimal performance. In the high-reverberant environment (figure 2), more errors are introduced so that the R1-MWF performance is visibly degraded compared to the optimal performance.

In contrast to the R1-MWF and SP-MWF, the performance of the *SDW-MWF* implementation (4) does

depend on μ. For small values of μ, the filter degrades to the trivial filter (no SNR improvement), while for larger values it converges towards the performance of the R1-MWF implementation. The performance degradation for small values of μ is significant for both the low-reverberant and the high-reverberant environment. As a check, the *predicted* output SNR for larger values of μ (42) was also calculated and shown on the figures.

When comparing the 4-microphone (binaural) setup to the 2-microphone (no binaural link) setup, it can be seen that the SNR improvement is significantly higher when using a binaural setup. By using more microphone signals, a better noise reduction performance is thus effectively achieved. However, it can also be seen that the performance degradation versus the optimal filter (and of the SDW-MWF versus the R1-MWF implementation) is larger for the binaural setup, especially so in the high-reverberant environment. As larger correlation matrices have to be estimated when the number of microphones increases, more estimation errors are introduced, thus leading to a larger degradation.

Remarkably, for the MSE cost function ($\mu = 1$), the difference between the general SDW-MWF and R1-MWF is quite large for the binaural setup. In other work using the general SDW-MWF formula, a value of $\mu = 5$ was therefore usually selected [25]. Figures 1 and 2 indeed illustrate that this is a sound choice: larger values of μ will not yield much SNR improvement anymore, but merely introduce more speech distortion.

While larger filterlengths (or DFT sizes) than $L = 128$ could be inappropriate for a binaural hearing aid application due to a prohibitive computational complexity and input-output latency, it is interesting from a theoretical point of view to study the effect of increasing L. In figure 3, the left and right narrowband SNR improvements for the 4-microphone setup are shown as a function of L. As before, the appropriate frequency bin (corresponding to $f = 640$ Hz) is selected for the performance evaluation, and the results for the SP-MWF are omitted.

As the noise correlation matrix is estimated with a better resolution when L increases, the performances of the different filters (also of the optimal filter) improve. It can also be observed that the performance of the R1-MWF gets closer to the optimal performance at larger values of L. This is due to the fact that the rank-one assumption is then better satisfied: for $L = 64$, the ratio of the dominant eigenvalue and second-highest eigenvalue is only 2.6, while for $L = 4096$ this ratio is equal to 30.4. The $\hat{\mathbf{R}}_x$ estimation error quantified by (49) goes down from $\delta(\hat{\mathbf{R}}_x) = 0.48$ for $L = 64$ to $\delta(\hat{\mathbf{R}}_x) = 0.22$ for $L = 4096$. The remaining errors at $L = 4096$ are mostly scaling errors, which do not degrade the performance according to (48), i.e. the numerator $(\rho\rho^{''} - P_s|\rho^{'}|^2)$ is almost equal to zero.

Finally, although the estimated $\hat{\mathbf{R}}_x$ better approximates the ideal $\mathbf{R}_x$ for higher values of L (as was just discussed), figure 3 illustrates that the performance degradation of the SDW-MWF implementation (4) is still significant for lower values of μ. In conclusion, even if larger filterlengths could be applied, the R1-MWF still offers a significant benefit over the standard SDW-MWF.

C. Broadband SNR improvement: effect of the single-channel postfilter

This section features more elaborate experiments for 21 different spatial scenarios, summarized in table I. To illustrate the effect of the single-channel SDW postfilter, the broadband output SNR is calculated in

Notation	**Description**
S0 Nx	target at 0°, single noise source at x° (0° : 30° : 330°)
S0 N2a	target at 0°, noise sources at -60° and 60°
S0 N2b	target at 0°, noise sources at -120° and 120°
S0 N2c	target at 0°, noise sources at 120° and 210°
S0 N3	target at 0°, noise sources at 90°, 180° and 270°
S0 N4a	target at 0°, noise sources at 60°, 120°, 180° and 210°
S0 N4b	target at 0°, noise sources at 60°, 120°, 180° and 270°
S90 N180	target at 90°, single noise source at 180°
S90 N270	target at 90°, single noise source at 270°
S45 N315	target at 45°, single noise source at 315°

TABLE I

SINGLE TARGET SOURCE, SPATIAL SCENARIOS

this experiment. As in [42], the frequency-domain filters are transformed into causal time-domain filters. The (time-domain) speech and noise components of the microphone signals are then filtered separately in order to calculate the broadband SNR. Figure 4 shows the broadband SNR improvements obtained by the SP-MWF implementation for a setup with only 2 used microphones (i.e. no binaural link) and a setup where all 4 microphones are used to generate the output. Only the left output SNR is shown as the right output SNR leads to similar conclusions. The performance of the SP-MWF is now shown, while the R1-MWF is omitted as it provides similar results. In section V-B it was already shown that the narrowband SNR performance obtained with the SDW-MWF is dependent on μ. Its broadband SNR improvement is therefore also dependent on μ. As in the narrowband case, it is consistently outperformed by the R1-MWF and SP-MWF, i.e. for a given μ, the broadband SNR improvement is smaller than the SNR improvements of the R1-MWF and SP-MWF. As this performance degradation will also be illustrated in the next section (where intelligibility weighted SNR's are calculated), these results are omitted here.

The postfilter attenuates frequency bins with significant residual noise (the denominator of (10) is small), so that the broadband output SNR is increased by the postfilter. Larger values of μ lead to larger broadband SNR improvements, as intuitively, the increase of μ globally attenuates the noise power at a higher rate than the target signal [33]. This SNR increase comes at the cost of a higher speech distortion as will be illustrated in the next section. When comparing the 2-microphone to the 4-microphone results, it can be seen that the relative SNR increase by postfiltering is larger in the 2-microphone case for a few particular spatial scenarios (for example S0N30, S0N330). When the angle between speech and noise sources is small, a closely spaced 2-microphone array cannot form a sufficiently narrow beam, so that the performance of the spatial filter is insufficient (see also [25]). In these scenarios, the postfilter can help increase the broadband SNR performance to a larger extend than in the 4-microphone (binaural) case.

As the average spectra of the multi-talker babble noise and speech-shaped VU noise overlap to a great extend [24], the achievable performance increase by postfiltering is limited. For non-overlapping spectra, the improvement by postfiltering can be much larger (see for example [33], where white Gaussian noise was used as interfering noise signal).

D. SI-weighted SNR improvement and distortion

To assess speech intelligibility improvements, broadband speech intelligibility (SI) weighted measures have been proposed. As for the broadband SNR, the (time-domain) speech and noise components of the microphone signals are filtered separately. The signals are then filtered by one-third octave band bandpass filters, and the SNR is calculated per band. The SI-weighted SNR improvement (in dB) [9], [44] is then defined as:

$$\Delta\mathrm{SNR}_{\mathrm{SI}} = \sum_i I_i \ (\mathrm{SNR}_{i,\mathrm{out}} - \mathrm{SNR}_{i,\mathrm{in}}) \ , \tag{50}$$

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, and where $\mathrm{SNR}_{i,\mathrm{out}}$ and $\mathrm{SNR}_{i,\mathrm{in}}$ are the output SNR and input SNR (in dB) in this band. The center frequencies f_i^c and the values I_i are defined in [43]. Similarly, an intelligibility weighted distortion measure was defined in [9]:

$$\mathrm{SD}_{\mathrm{SI}} = \sum_i I_i \ \mathrm{SD}_i \ , \tag{51}$$

where SD_i is the average spectral distortion in the ith one-third octave band, calculated as

$$\mathrm{SD}_i = \int_{2^{-1/6} f_i^c}^{2^{1/6} f_i^c} |10 \ \log_{10} \ G^s(f)| \, df \ , \tag{52}$$

with $G^s(f)$ the power transfer function of the speech component from the input to the output of the noise reduction algorithm. A distortion value of 0 dB corresponds to an undistorted signal, while larger distortion values correspond to more introduced speech distortion.

The SI-weighted SNR improvements for the left output are shown in figure 5, where the spatial scenarios of table I are again tested. For the R1-MWF and SP-MWF implementations, the SI-weighted SNR does not depend strongly on μ. This is actually expected as it was shown that the output SNR per frequency bin, which determines speech intelligibility, is independent of μ for the R1-MWF and SP-MWF implementations. The speech distortion parameter can therefore be set to $\mu = 0$ in this experiment, while keeping in mind that larger values can be used if a larger (not SI-weighted) broadband SNR, or thus listening comfort, is required (cfr. section V-C). The SDW-MWF (4) needs larger values of μ in order to achieve the same SNR performance as the other filter implementations, which is in accordance with the theoretical analysis.

The introduced SI-weighted distortion is shown in figure 6. If the SDW-MWF implementation is used, a value of $\mu = 5$ (or even larger for some scenarios) should be used in order to achieve the same SNR performance as the other filter implementations. Figure 6 shows that this will introduce more speech distortion compared to the other filter implementations (where $\mu = 0$ can be used). Another remarkable observation is that the R1-MWF and SP-MWF, which are theoretically equivalent, introduce different speech distortions. This is due to the fact that their postfilters are different if estimation errors are present (section III). Various simulations have shown that for the same value of μ, the SP-MWF is more conservative than the R1-MWF, thereby introducing less speech distortion.

E. Multiple target speakers

The SI-weighted performances of the different implementations are now tested in multiple target speech sources scenarios. That is, there are multiple desired speech signals in different directions that should be preserved by the filter. As a consequence, the performance of the R1-MWF and SP-MWF (which implicitly assume a single target source) is expected to degrade.

Sixteen different spatial scenarios are tested, cfr. table II for the used notation. The obtained SI-weighted SNR improvements are shown in figure 7. The R1-MWF and SP-MWF implementations still outperform the SDW-MWF implementation, even for scenarios with 4 target speakers (full-rank speech correlation matrix). There is no noticeable difference between the performance of the R1-MWF and SP-MWF. Again, as larger values of μ do not affect the SI-weighted SNR improvements of the R1-MWF and SP-MWF, only the performance for $\mu = 0$ is shown. Larger values can however be used to increase

Notation	Description
S1	1 target at -30°
S2	2 targets at -30° and 45°
S3	3 targets at -30°, 45° and 150°
S4	4 targets at -90°, -30°, 45° and 150°
N1a	1 noise source at 0°
N1b	1 noise source at -90°
N2	2 noise sources at -60° and 120°
N4	4 noise sources at -90°, -60°, 120°, 180°

TABLE II

MULTIPLE TARGET SOURCES, SPATIAL SCENARIOS (ALL COMBINATIONS OF S AND N ARE MADE).

the broadband SNR as was shown in section V-C.

The SI-weighted distortion is shown in figure 8. Again, the distortion increases as μ increases. For the R1-MWF, a performance degradation can be observed: as the number of target sources increases, more speech distortion is introduced. Even for $\mu = 0$, a large increase in distortion can be observed. The violation of the rank-one assumption will thus introduce speech distortion, but not affect the SNR performance. Remarkably, the SP-MWF does not seem to introduce more speech distortion as the number of target sources increases.

To further explain these results, distortion measures were also calculated for each target source separately. The results for the scenario S4N1a (targets at 270, 330, 45 and 150 degrees, noise in front of the head) are shown in figure 9, where (a) shows the distortion in the left hearing aid output, and (b) shows the distortion in the right hearing aid output. The results for the lower reverberation environment (RT_{60} = 0.21s) are used here, because the overall tendencies are more clearly visible in the figures.

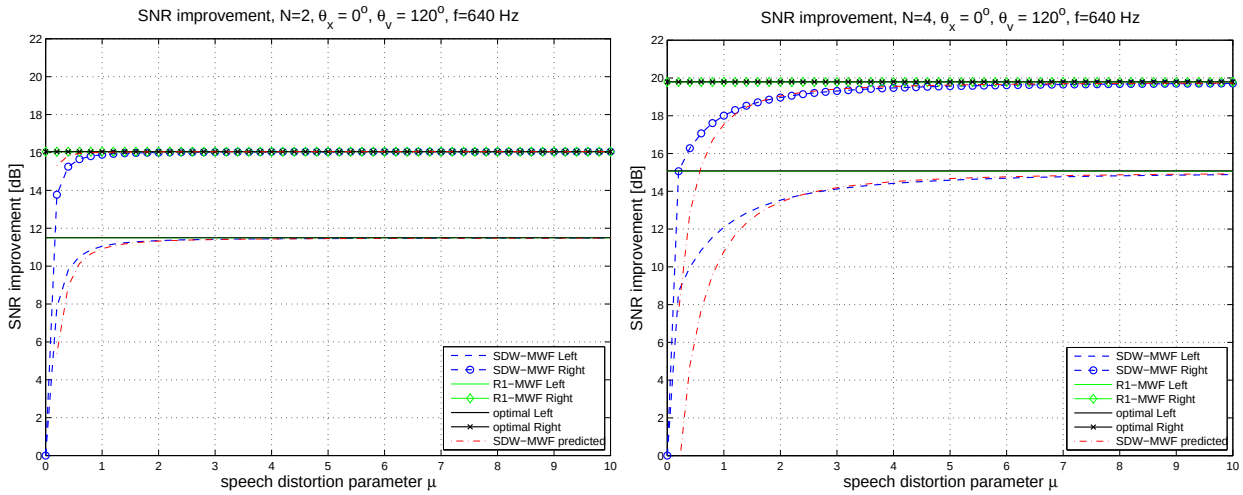
The target speech signals were scaled so as to have the same average input power at the loudspeakers, but because of the headshadow effect, the signals have different average input powers at the reference microphones. As a consequence, S270 and S330 will be dominant in the left output (a), whereas S45 and S150 are dominant in the right output (b). Figure 9 illustrates that the different filter implementations introduce different distortions in these signals. The R1-MWF introduces a lot of distortion, especially for the dominant sources: S270 and S330 have the largest distortion values in the left output (a), S45 and S150 have the largest values in the right output (b). The SP-MWF on the other hand introduces more distortion in the signals with the lowest input power, whereas the distortion on the dominant sources is very small. This is a beneficial effect, as the distortion introduced on the low power sources will be less audible, while the distortion on the dominant sources should be kept as small as possible. These results

also explain why the overall distortion in figure 8 (on the sum of the target signals) increased for the R1-MWF, but remained small for the SP-MWF. Figure 9 also shows that the SDW-MWF introduces a more or less even amount of distortion in the different target signals. Larger values of μ can be chosen as the introduced distortions are still reasonable, in contrast to the R1-MWF, where $\mu = 0$ already introduces a large amount of distortion.

VI. Conclusion

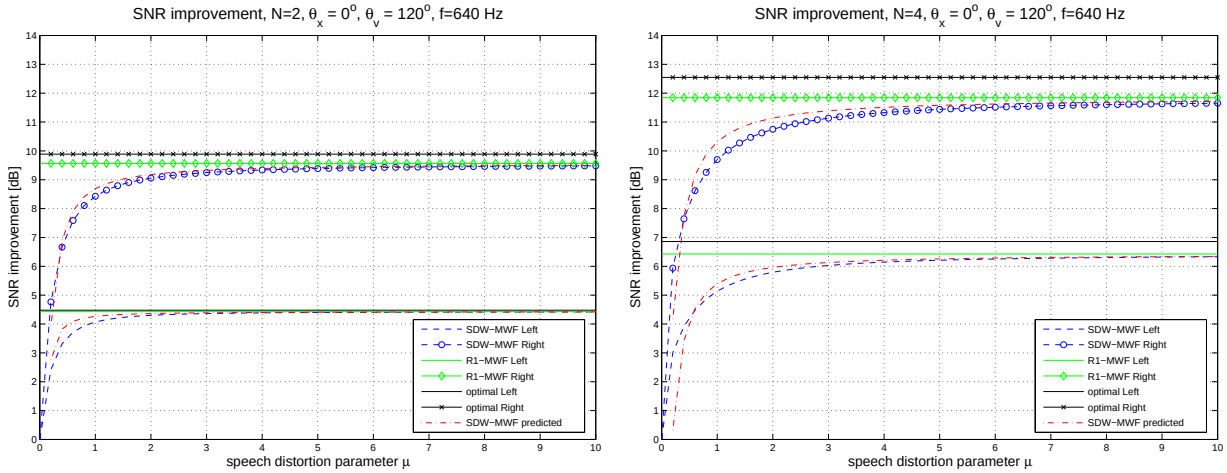
The theoretical analysis and the simulations illustrate that for a single target speech source, the R1-MWF and SP-MWF implementations achieve a better noise reduction performance than the standard SDW-MWF implementation. Moreover, they do not lose this performance for smaller values of μ (less speech distortion). For applications where only a single target speech source is present, the R1-MWF and SP-MWF thus have a clear advantage over the SDW-MWF, especially if only a limited amount of speech distortion is allowed.

As the number of target speech sources increases, the simulations show that the R1-MWF loses some of its benefit over the SDW-MWF. It still achieves large (SI-weighted) SNR improvements, but more speech distortion is introduced as the number of target speakers increases, especially for the dominant sources. The SP-MWF on the other hand only introduces distortions in the signals with low input powers, which is less audible, in addition to providing a large (SI-weighted) SNR improvement.



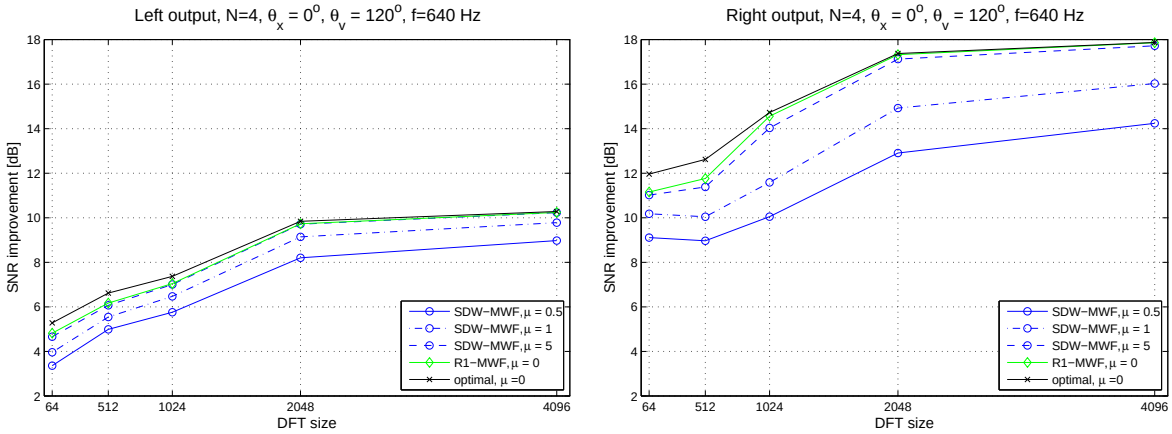
(a) $N = 2$ (no binaural link) (b) $N = 4$

Fig. 1. Narrowband SNR improvement as a function of μ for SDW-MWF (4) and R1-MWF (15); $\mathrm{RT}_{60} = 0.21$s (low reverberation); (a) $N = 2$ microphones (no binaural link), (b) $N = 4$ microphones



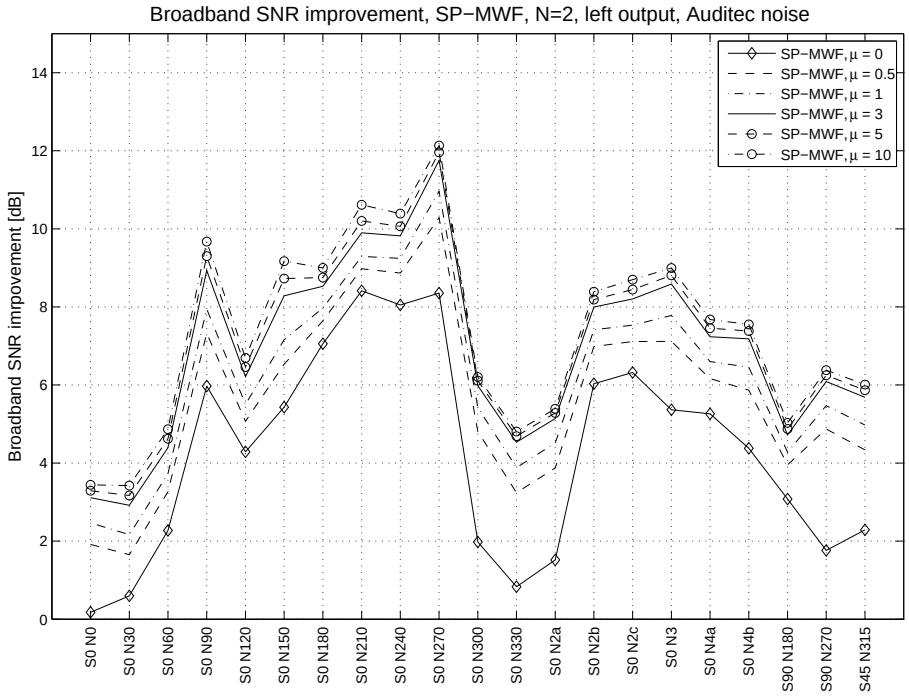
(a) $N = 2$ (no binaural link) (b) $N = 4$

Fig. 2. Narrowband SNR improvement as a function of μ for SDW-MWF (4) and R1-MWF (15); $\mathrm{RT}_{60} = 0.61$s (high reverberation); (a) $N = 2$ microphones (no binaural link), (b) $N = 4$ microphones

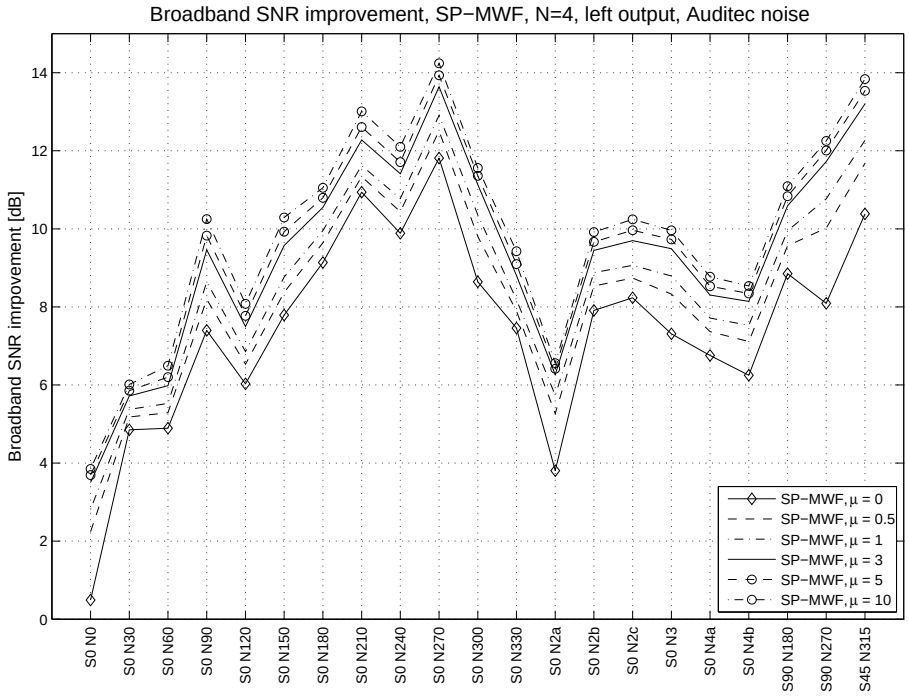


(a) Left output (b) Right output

Fig. 3. Narrowband SNR improvement as a function of DFT size for SDW-MWF (4) and R1-MWF (15); $N = 4$ microphones; $\mathrm{RT}_{60} = 0.61$s (high reverberation); (a) Left output, (b) Right output



(a) $N = 2$ (no binaural link)



(b) $N = 4$

Fig. 4. Broadband SNR improvement for SP-MWF; left output, RT_{60} = 0.61s, Multi-talker babble noise (Auditec); (a) $N = 2$ microphones (no binaural link), (b) $N = 4$ microphones

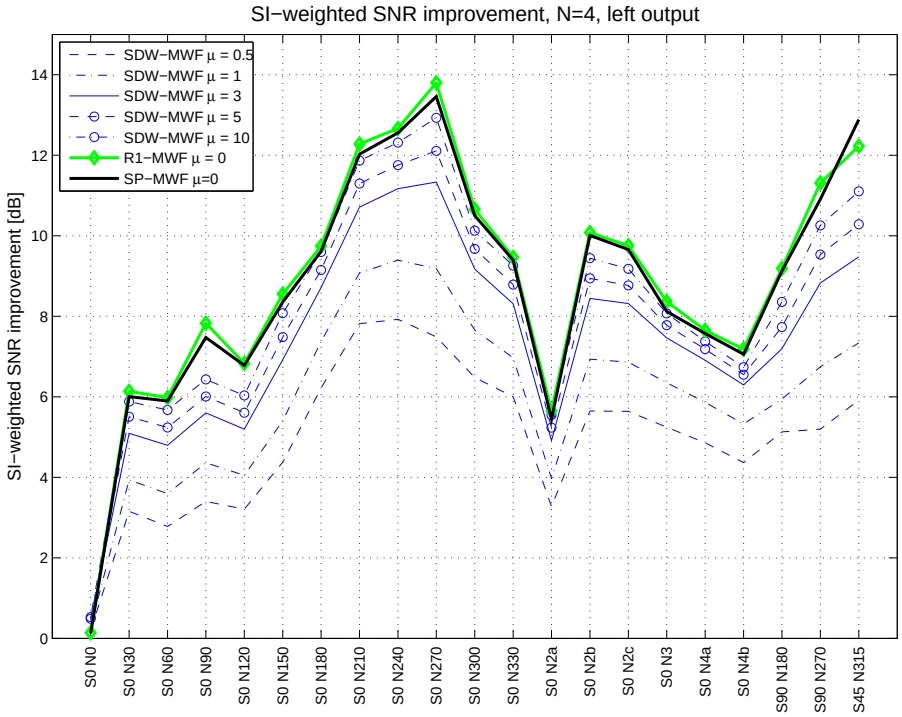


Fig. 5. SI-weighted SNR improvement for SDW-MWF, R1-MWF, SP-MWF; left output, RT_{60} = 0.61s, 4 microphones

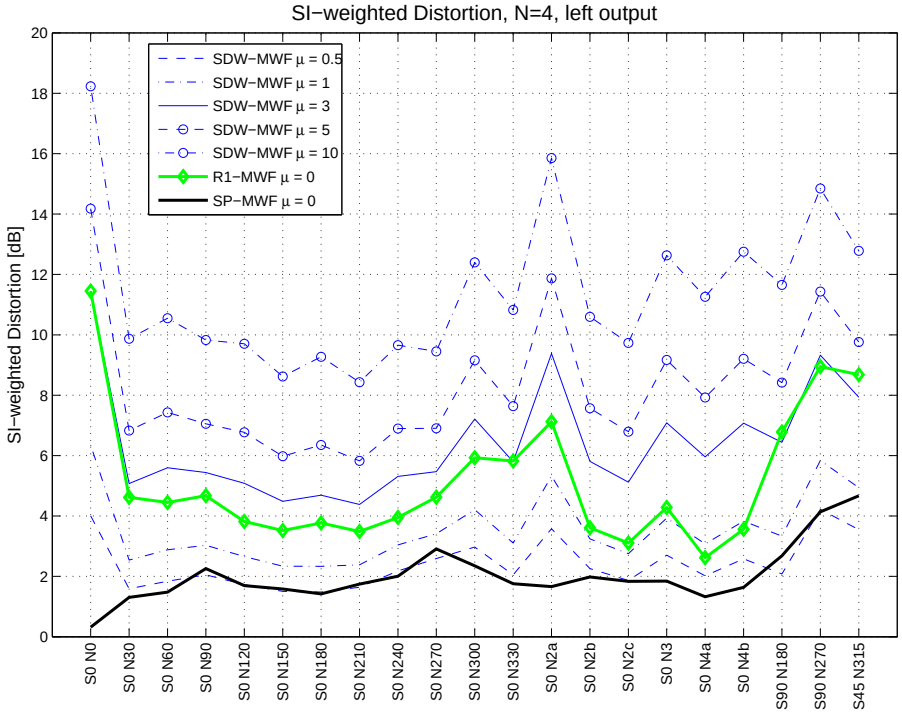


Fig. 6. SI-weighted Distortion for SDW-MWF, R1-MWF, SP-MWF; left output, RT_{60} = 0.61s, 4 microphones

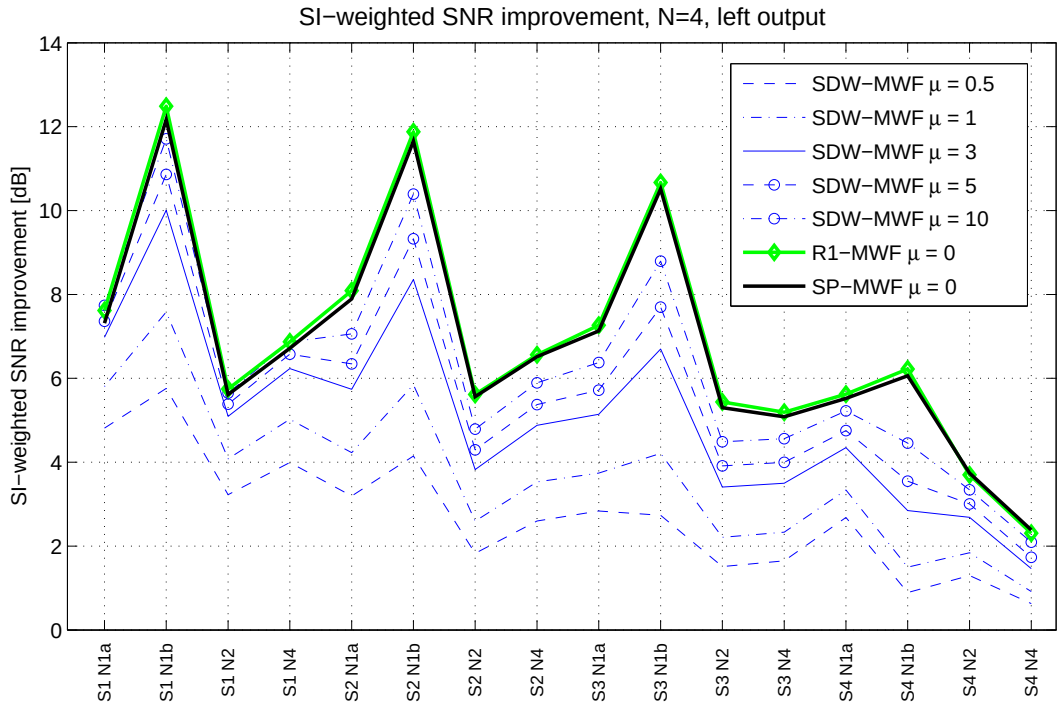


Fig. 7. SI-weighted SNR improvement for SDW-MWF, R1-MWF, SP-MWF; left output, RT_{60} = 0.61s, 4 microphones

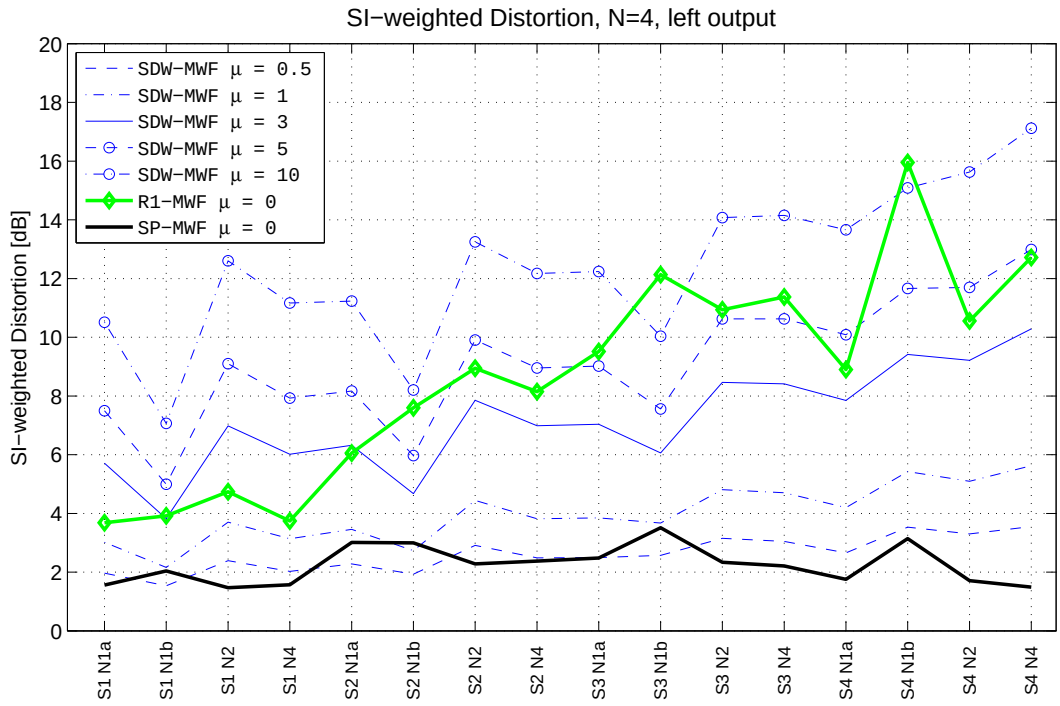
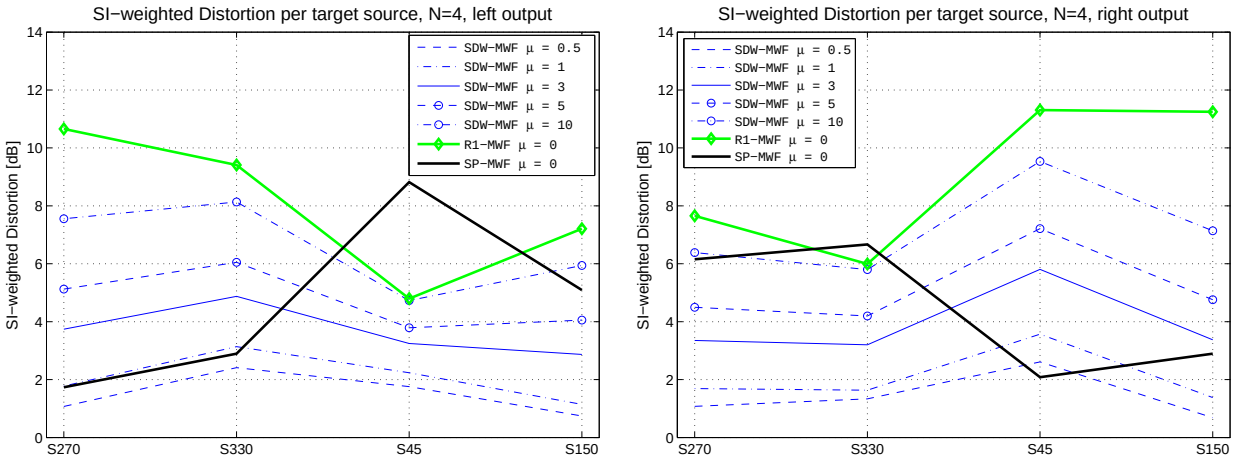


Fig. 8. SI-weighted Distortion for SDW-MWF, R1-MWF, SP-MWF; left output, RT_{60} = 0.61s, 4 microphones



(a) Left output (b) Right output

Fig. 9. SI-weighted Distortion per target source; S4N1a, RT_{60} = 0.21s, 4 microphones

REFERENCES

[1] V. Hamacher, J. Chalupper, J. Eggers, E. Fischer, U. Kornagel, H. Puder, and U. Rass, "Signal processing in high-end hearing aids: state of the art, challenges, and future trends," *EURASIP J. Appl. Signal Process.*, pp. 2915–2929, 2005.

[2] H. Dillon, *Hearing Aids*. Boomerang Press, Australia, 2001.

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

[4] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1218–1234, July 2006.

[5] T. D. Trine and D. Van Tasell, "Digital hearing aid design: Fact vs. fantasy," *The hearing Journal*, vol. 55, no. 2, pp. 36–38, 40–42, 2002.

[6] 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.

[7] S. Gannot, D. Burshtein, and E. Weinstein, "Signal Enhancement Using Beamforming and Non-Stationarity with Applications to Speech," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[8] 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 Hear.*, vol. 28, no. 1, pp. 62–72, 2007.

[9] 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 Audio Process.*, vol. 13, no. 4, pp. 487–503, July 2005.

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

[11] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," *Signal Process.*, vol. 84, no. 12, pp. 2367–2387, Dec. 2004.

[12] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-Domain Criterion for Speech Distortion Weighted Multichannel Wiener Filter for Robust Noise Reduction," *Speech Commun.*, vol. 49, no. 7–8, pp. 636–656, Jul.-Aug. 2007.

[13] Y. Ephraim and H. L. Van Trees, "A Signal Subspace Approach for Speech Enhancement," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 4, pp. 251–266, July 1995.

[14] B. Kollmeier, J. Peissig, and V. Hohmann, "Real-time multiband dynamic compression and noise reduction for binaural hearing aids," *J. Rehabil. Res. Develop.*, vol. 30, no. 1, pp. 82–94, 1993.

[15] J. Desloge, W. Rabinowitz, and P. Zurek, "Microphone-array hearing aids with binaural output–Part I: Fixed-processing systems," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 6, pp. 529–542, Nov. 1997.

[16] D. Welker, J. Greenberg, J. Desloge, and P. Zurek, "Microphone-array hearing aids with binaural output–Part II: A two-microphone adaptive system," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 6, pp. 543–551, Nov. 1997.

[17] I. Merks, M. Boone, and A. Berkhout, "Design of a broadside array for a binaural hearing aid," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust. (WASPAA)*, New Paltz NY, USA, Oct. 1997.

[18] V. Hamacher, "Comparison of advanced monaural and binaural noise reduction algorithms for hearing aids," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Orlando FL, USA, May 2002, pp. 4008–4011.

[19] R. Nishimura, Y. Suzuki, and F. Asano, "A new adaptive binaural microphone array system using a weighted least squares algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Orlando FL, USA, May 2002, pp. 1925–1928.

[20] T. Wittkop and V. Hohmann, "Strategy-selective noise reduction for binaural digital hearing aids," *Speech Commun.*, vol. 39, no. 1-2, pp. 111–138, Jan. 2003.

[21] M. Lockwood, D. Jones, R. Bilger, C. Lansing, W. O'Brien, B. Wheeler, and A. Feng, "Performance of time- and frequency-domain binaural beamformers based on recorded signals from real rooms," *J. Acoust. Soc. Amer.*, vol. 115, no. 1, pp. 379–391, Jan. 2004.

[22] T. Lotter and P. Vary, "Dual-channel speech enhancement by superdirective beamforming," *EURASIP J. Appl. Signal Process.*, pp. 1–14, 2006.

[23] O. Roy and M. Vetterli, "Rate-constrained beamforming for collaborating hearing aids," in *Proc. International Symposium on Information Theory (ISIT)*, Seattle WA, USA, July 2006, pp. 2809–2813.

[24] 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," *J. Acoust. Soc. Amer.*, vol. 124, no. 1, pp. 484–497, 2008.

[25] ——, "Speech enhancement with multichannel Wiener filter techniques in multi-microphone binaural hearing aids," *J. Acoust. Soc. Amer.*, vol. 125, no. 1, pp. 360–371, 2009.

[26] B. Cornelis, S. Doclo, T. Van den Bogaert, M. Moonen, and J. Wouters, "Theoretical analysis of binaural multimicrophone noise reduction techniques," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 2, pp. 342–355, Feb. 2010.

[27] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localisation* (Revised Edition). MIT Press, 1996.

[28] P. Zurek, *Binaural advantages and directional effects in speech intelligibility*. 2nd Ed., Boston, MA, Allyn and Bacon, 1992, ch. 15 in "Acoustical factors affecting hearing aid performance", pp. 255–276.

[29] 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 *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, Paris, France, Sept. 2006.

[30] A. Spriet, S. Doclo, M. Moonen, and J. Wouters, "A unification of adaptive multi-microphone noise reduction systems," in *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, Paris, France, Sep. 2006.

[31] Y. Huang, J. Benesty, and J. Chen, "Analysis and comparison of multichannel noise reduction methods in a common framework," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 5, pp. 957–968, July 2008.

[32] E. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, "New insights into the MVDR beamformer in room acoustics," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 1, pp. 158–170, Jan. 2010.

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

[34] K. U. Simmer, J. Bitzer, and C. Marro, *Post-Filtering Techniques*. Springer-Verlag, May 2001, ch. 3 in "Microphone Arrays: Signal Processing Techniques and Applications" (Brandstein, M. S. and Ward, D. B., Eds.), pp. 39–60.

[35] J. Benesty, J. Chen, and Y. Huang, *Noncausal (frequency-domain) optimal filters*. Springer-Verlag, 2008, ch. 6 in "Microphone Array Signal Processing", pp. 115–137.

[36] 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. Acoust., Speech, Signal Process. (ICASSP)*, Tai Pei, Taiwan, April 2009, pp. 129–132.

[37] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 4, April 1979, pp. 208–211.

[38] J. Chen, J. Benesty, and Y. Huang, "A minimum distortion noise reduction algorithm with multiple microphones," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 3, pp. 481–493, March 2008.

[39] Auditec, "Auditory tests (revised)," *Compact Disc, Auditec, St. Louis*, 1997.

[40] N. J. Versfeld, L. Daalder, J. M. Festen, and T. Houtgast, "Method for the selection of sentence materials for efficient measurement of the speech reception threshold," *J. Acoust. Soc. Amer.*, vol. 107, no. 3, pp. 1671–1684, 2000.

[41] B. Cornelis, M. Moonen, and J. Wouters, "A QRD-RLS based frequency-domain Multichannel Wiener Filter algorithm for noise reduction in hearing aids," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Aalborg, Denmark, Aug. 2010, pp. 1953–1957.

[42] H. Sawada, R. Mukai, S. Araki, and S. Makino, *Frequency-domain blind source separation*. Springer-Verlag, 2005, ch. 13 in "Speech Enhancement" (Benesty, J. and Makino, S. and Chen, J., Eds.), pp. 299–327.

[43] ANSI-SII, "ANSI S3.5-1997 American National Standard Methods for Calculation of the Speech Intelligibility Index," *Acoust. Soc. Amer., Melville, NY*, June 1997.

[44] 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, no. 5, pp. 3009–3010, Nov. 1993.