CURVATURE-BASED OPTIMIZATION OF THE TRADE-OFF PARAMETER IN THE SPEECH DISTORTION WEIGHTED MULTICHANNEL WIENER FILTER

Ina Kodrasi, Daniel Marquardt, Simon Doclo

University of Oldenburg, Department of Medical Physics and Acoustics,
and Cluster of Excellence Hearing4All, Oldenburg, Germany
ina.kodrasi@uni-oldenburg.de

ABSTRACT

The objective of the speech distortion weighted multichannel Wiener filter (MWF) is to reduce background noise while controlling speech distortion. This can be achieved by means of a trade-off parameter, hence, selecting an optimal trade-off parameter is of crucial importance.

Aiming at incorporating knowledge about the resulting speech distortion and noise power, in this paper we propose to compute the trade-off parameter as the point of maximum curvature of the parametric plot of noise power versus speech distortion. To determine a narrowband trade-off parameter, an analytical expression is derived for computing the point of maximum curvature, whereas to determine a broadband parameter an optimization routine is used. The speech distortion and the noise power terms can also be weighted in advance, e.g. based on perceptually motivated criteria. Experimental results show that using the proposed method instead of the MWF improves the intelligibility weighted SNR without significantly degrading the speech distortion.

Index Terms— noise reduction, speech distortion, MWF, trade-off parameter, L-curve

1. INTRODUCTION

In many speech communication applications such as teleconferencing applications, hearing aids, and voice-controlled systems, the microphone signals are often corrupted by additive background noise, which can significantly impair speech intelligibility. To tackle this problem several multichannel noise reduction techniques have been investigated, which exploit both spatial and spectro-temporal information to reduce the background noise while limiting speech distortion [1, 2, 3, 4, 5]. A commonly used noise reduction technique is multichannel Wiener filtering (MWF) which minimizes the mean-square error between the output signal and the speech component in one of the microphones [6, 7]. The error typically consists of a noise power term and a speech distortion term. While the MWF assigns equal importance to both terms, the speech distortion weighted MWF (MWF_{SDW}) incorporates a trade-off parameter which provides a trade-off between noise reduction and speech distortion [1, 2]. Due to the arising trade-off, the choice of this parameter in the MWF_{SDW} is of crucial importance.

Typically a fixed trade-off parameter, empirically selected, has been used which can be advantageous in preventing the filter coefficients from changing excessively, hence avoiding spectral peaks that might be perceived as musical noise. However, using a fixed parameter can be suboptimal since it does not reflect the typically changing speech and noise powers in different time-frequency bins [8, 9, 10, 11]. Hence, in [8, 9, 10] it has been proposed to use a soft voice activity detector [12] to weight the speech distortion term by the probability that speech is present and the noise power term by the probability that speech is absent. This principle has been further extended in [11] where an empirical strategy for the selection of a narrowband trade-off parameter has been proposed based on the instantaneous masking threshold [13].

In this paper a systematic method for selecting a narrowband trade-off parameter as well as a broadband one is established. Aiming at incorporating knowledge about the resulting speech distortion and noise power, it is proposed to use the parameter that yields small and approximately equal relative changes in both quantities. Mathematically this parameter is defined as the point of maximum curvature of the parametric plot of noise power versus speech distortion. Furthermore, the speech distortion and noise power terms can be weighted in advance, based on what is more important to the speech communication application under consideration or based on perceptually motivated criteria. An analytical expression in terms of the signal-to-noise ratio (SNR) is derived for the narrowband trade-off parameter, whereas an optimization routine needs to be used to compute the broadband trade-off parameter. The narrowband trade-off parameters in [8, 11] can then be derived within the proposed method by selecting appropriate weighting functions.

2. CONFIGURATION AND NOTATION

Consider an M-channel acoustic system, where the m-th microphone signal $Y_m(k,l)$ at frequency index k and time index l consists of a speech component $X_m(k,l)$ and a noise component $V_m(k,l)$, i.e., $Y_m(k,l) = X_m(k,l) + V_m(k,l)$. For the sake of readability the time index l will be omitted in the remainder of this paper, except where explicitly required. In vector notation, the M-dimensional vector $\mathbf{y}(k)$ of the received microphone signals can be written as

$$\boxed{\mathbf{y}(k) = \mathbf{x}(k) + \mathbf{v}(k)} \tag{1}$$

with $\mathbf{y}(k) = [Y_1(k) \ \ldots Y_M(k)]^T$, and the speech and noise vectors $\mathbf{x}(k)$ and $\mathbf{v}(k)$ similarly defined. Defining the vector of filter coefficients $\mathbf{w}(k)$ similarly as $\mathbf{y}(k)$, the output signal $Z(k)$ is given by

$$\boxed{Z(k) = \mathbf{w}^H(k)\mathbf{y}(k) = \mathbf{w}^H(k)\mathbf{x}(k) + \mathbf{w}^H(k)\mathbf{v}(k)} \tag{2}$$

The MWF aims at noise reduction by minimizing the mean-square error between the output signal and the received speech component in the m-th microphone, i.e., reference microphone. In the MWF_{SDW} a trade-off parameter $\mu(k)$ has been incorporated, which allows to

This work was supported in part by a Grant from the GIF, the German-Israeli Foundation for Scientific Research and Development, the Cluster of Excellence 1077 Hearing4All, funded by the German Research Foundation (DFG), and the Marie Curie Initial Training Network DREAMS (Grant no. 316969).

trade-off between noise reduction and speech distortion [1, 2]. Assuming that the speech and noise components are uncorrelated, the MWF$_{\text{SDW}}$ cost function can be written as

$$\boxed{\min_{\mathbf{w}(k)} \underbrace{\mathcal{E}\{|\mathbf{w}^H(k)\mathbf{x}(k) - \mathbf{e}_m^T\mathbf{x}(k)|^2\}}_{\psi_{\mathbf{x}}(k)} + \mu(k)\underbrace{\mathcal{E}\{|\mathbf{w}^H(k)\mathbf{v}(k)|^2\}}_{\psi_{\mathbf{v}}(k)}} \tag{3}$$

with $\mathcal{E}$ the expected value operator, $\mathbf{e}_m$ the M-dimensional selector vector, i.e., a vector of which the m-th element is equal to 1 and all other elements are equal to 0, $\psi_{\mathbf{x}}(k)$ the speech distortion, and $\psi_{\mathbf{v}}(k)$ the noise power. The filter minimizing the cost function in (3) is given by

$$\mathbf{w}(k) = \left[\mathbf{R}_{\mathbf{x}}(k) + \mu(k)\mathbf{R}_{\mathbf{v}}(k)\right]^{-1}\mathbf{R}_{\mathbf{x}}(k)\mathbf{e}_m, \tag{4}$$

with $\mathbf{R}_{\mathbf{x}}(k)$ and $\mathbf{R}_{\mathbf{v}}(k)$ being the speech and noise correlation matrices respectively, defined as

$$\mathbf{R}_{\mathbf{x}}(k) = \mathcal{E}\{\mathbf{x}(k)\mathbf{x}^H(k)\} = P_s(k)\mathbf{a}(k)\mathbf{a}^H(k), \tag{5}$$

$$\mathbf{R}_{\mathbf{v}}(k) = \mathcal{E}\{\mathbf{v}(k)\mathbf{v}^H(k)\}, \tag{6}$$

where $P_s(k) = \mathcal{E}\{|S(k)|^2\}$ is the power spectral density of the speech source and $\mathbf{a}(k) = [A_1(k) \ \ldots \ A_M(k)]^T$ is the vector of the acoustic transfer functions (ATFs). The MWF in (4) can be decomposed into a Minimum Variance Distortionless Response Beamformer (MVDR) $\mathbf{w}_{\text{MVDR}}(k)$ and a single channel Wiener postfilter $G(k)$ applied to the MVDR output [14], i.e.,

$$\boxed{\mathbf{w}(k) = \underbrace{A_m^*(k)\frac{\mathbf{R}_{\mathbf{v}}^{-1}(k)\mathbf{a}(k)}{\mathbf{a}^H(k)\mathbf{R}_{\mathbf{v}}^{-1}(k)\mathbf{a}(k)}}_{\mathbf{w}_{\text{MVDR}}(k)}\underbrace{\frac{\rho(k)}{\mu(k)+\rho(k)}}_{G(k)}} \tag{7}$$

with $\rho(k)$ being the SNR at the output of the MVDR beamformer, i.e.,

$$\boxed{\rho(k) = P_s(k)\mathbf{a}^H(k)\mathbf{R}_{\mathbf{v}}^{-1}(k)\mathbf{a}(k)} \tag{8}$$

Using $\mu(k) = 0$ in (7), the MWF$_{\text{SDW}}$ yields the MVDR beamformer, which reduces the noise while keeping the speech component in the reference microphone undistorted, i.e., $\mathbf{w}_{\text{MVDR}}^H\mathbf{a}(k) = A_m(k)$. Using $\mu(k) \neq 0$, the residual noise at the output of the MVDR beamformer can be further suppressed at the cost of introducing speech distortion. Using $\mu(k) = 1$, the MWF$_{\text{SDW}}$ results in the MWF which assigns equal importance to the speech distortion and noise power terms. If $\mu(k) > 1$, the noise power is reduced further in comparison to the MWF at the expense of increased speech distortion. On the contrary, if $\mu(k) < 1$ speech distortion is reduced further at the expense of increased noise power. Hence the selection of the trade-off parameter in the MWF$_{\text{SDW}}$ is of crucial importance.

3. SELECTION OF THE TRADE-OFF PARAMETER

In the following, the L-curve method used for the automatic selection of the regularization parameter in least squares problems [15, 16] is adapted to select a trade-off parameter in the MWF$_{\text{SDW}}$.

3.1. Narrowband trade-off parameter

Applying the filter $\mathbf{w}(k)$ from (7) and using the definition of $\mathbf{R}_{\mathbf{x}}(k)$ in (5), the speech distortion $\psi_{\mathbf{x}}(k)$ can be expressed as

$$\boxed{\psi_{\mathbf{x}}(k) = P_s(k)|A_m(k)|^2\frac{\mu^2(k)}{[\mu(k)+\rho(k)]^2}} \tag{9}$$

Furthermore, the noise power can be expressed as

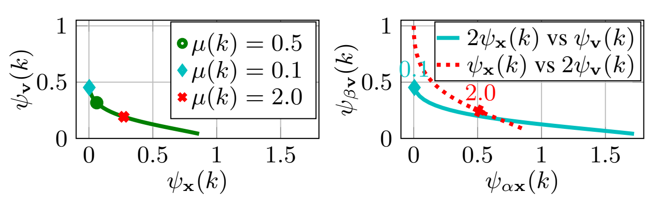
$$\boxed{\psi_{\mathbf{v}}(k) = P_s(k)|A_m(k)|^2\frac{\rho(k)}{[\mu(k)+\rho(k)]^2}} \tag{10}$$

Clearly it is desirable to use a trade-off parameter $\mu(k)$ that yields no speech distortion and no noise power, i.e., perfect noise reduction. However, given the inversely proportional relationship between $\psi_{\mathbf{x}}(k)$ and $\psi_{\mathbf{v}}(k)$, this is not achievable. Fig. 1a depicts a typical parametric plot of $\psi_{\mathbf{v}}(k)$ versus $\psi_{\mathbf{x}}(k)$ for 50 trade-off parameters linearly spaced between 10^{-4} and 5, with the marked points showing the exact value of $\mu(k)$ at the given positions. Due to the arising trade-off between $\psi_{\mathbf{v}}(k)$ and $\psi_{\mathbf{x}}(k)$, this parametric plot has an L-shape, with the corner (i.e., point of maximum curvature) located where the filter $\mathbf{w}(k)$ changes in nature from being dominated by large noise power to being dominated by large speech distortion. At the point of maximum curvature, i.e., $\mu(k) = 0.5$ in the depicted example, speech distortion and noise power are simultaneously minimized. Hence we propose to select the trade-off parameter $\mu(k)$ as the point of maximum curvature of the parametric plot of $\psi_{\mathbf{v}}(k)$ versus $\psi_{\mathbf{x}}(k)$.

Using such a parameter inherently implies that maintaining a low speech distortion and a high noise reduction performance is equally valuable to the speech communication system. However, in certain systems speech intelligibility is of central importance, hence one could allow for a higher noise reduction performance at the cost of increased speech distortion. In communication systems where speech quality is of central importance, noise reduction could be sacrificed to maintain a lower speech distortion. Furthermore, the importance of maintaining a low speech distortion or a high noise reduction performance also varies between different frequency bands, e.g. based on auditory masking properties. To account for these differences, we propose introducing a weighting function to the speech distortion and noise power terms, i.e.,

$$\psi_{\alpha\mathbf{x}}(k) = \alpha(k)\psi_{\mathbf{x}}(k) \ \text{ and } \ \psi_{\beta\mathbf{v}}(k) = \beta(k)\psi_{\mathbf{v}}(k), \tag{11}$$

with $\alpha(k)$ and $\beta(k)$ being the speech distortion and noise power weighting functions, defined e.g. based on psychoacoustically motivated measures such as average masking threshold [13] or speech intelligibility weighting [17] (cf. Section 4). Introducing a weighting function changes the point of maximum curvature. Fig. 1b depicts the parametric plot of $\psi_{\beta\mathbf{v}}(k)$ versus $\psi_{\alpha\mathbf{x}}(k)$ when the speech distortion term is weighted more, i.e., $\alpha(k) = 2$, $\beta(k) = 1$, and when the noise power is weighted more, i.e., $\alpha(k) = 1$, $\beta(k) = 2$. As it can be seen, putting more emphasis on the speech distortion term yields a lower trade-off parameter, i.e., the point of maximum curvature is $\mu(k) = 0.1$. On the other hand putting more emphasis on the noise power yields a higher trade-off parameter, i.e., $\mu(k) = 2$. The location of these points is also marked in the original plot in Fig. 1a,



(a) (b)

Fig. 1: Typical parametric plot of (a) noise power versus speech distortion and (b) (weighted) noise power versus (weighted) speech distortion

showing how weighting the speech distortion or noise power more changes the resulting trade-off in comparison to when no weights are applied.

The curvature $\kappa(k)$ of the parametric plot of $\psi_{\beta\mathbf{v}}(k)$ versus $\psi_{\alpha\mathbf{x}}(k)$ is defined as [18]

$$\kappa(k) = \frac{\psi'_{\alpha\mathbf{x}}(k)\psi''_{\beta\mathbf{v}}(k) - \psi''_{\alpha\mathbf{x}}(k)\psi'_{\beta\mathbf{v}}(k)}{\{[\psi'_{\alpha\mathbf{x}}(k)]^2 + [\psi'_{\beta\mathbf{v}}(k)]^2\}^{\frac{3}{2}}}, \tag{12}$$

where $\{\cdot\}'$ and $\{\cdot\}''$ denote the first and second derivative with respect to $\mu(k)$ respectively. The computation of the derivatives yields

$$\psi'_{\alpha\mathbf{x}}(k) = 2\alpha(k)P_s(k)|A_m(k)|^2\frac{\mu(k)\rho(k)}{[\mu(k)+\rho(k)]^3}, \tag{13}$$

$$\psi''_{\alpha\mathbf{x}}(k) = 2\alpha(k)P_s(k)|A_m(k)|^2\frac{\rho(k)[-2\mu(k)+\rho(k)]}{[\mu(k)+\rho(k)]^4}, \tag{14}$$

$$\psi'_{\beta\mathbf{v}}(k) = -2\beta(k)P_s(k)|A_m(k)|^2\frac{\rho(k)}{[\mu(k)+\rho(k)]^3}, \tag{15}$$

$$\psi''_{\beta\mathbf{v}}(k) = 6\beta(k)P_s(k)|A_m(k)|^2\frac{\rho(k)}{[\mu(k)+\rho(k)]^4}. \tag{16}$$

Substituting (13) to (16) in (12), the expression for the curvature can be simplified to

$$\boxed{\kappa(k) = \frac{\alpha(k)\beta(k)[\mu(k)+\rho(k)]^3}{2P_s(k)|A_m(k)|^2\rho(k)[\alpha^2(k)\mu^2(k)+\beta^2(k)]^{\frac{3}{2}}}} \tag{17}$$

To compute the optimal trade-off parameter $\mu(k)$, the curvature in (17) is maximized by setting its derivative to 0, i.e.,

$$\kappa'(k)=\frac{3\alpha(k)\beta(k)[\mu(k)+\rho(k)]^2[\beta^2(k)-\alpha^2(k)\mu(k)\rho(k)]}{2P_s(k)|A_m(k)|^2\rho(k)[\alpha^2(k)\mu(k)+\beta^2(k)]^{\frac{5}{2}}}=0$$
$$\Rightarrow \beta^2(k) - \alpha^2(k)\mu(k)\rho(k) = 0. \tag{18}$$

The solution to (18) yields

$$\boxed{\mu_{\mathrm{o}}(k) = \frac{\beta^2(k)}{\alpha^2(k)\rho(k)}} \tag{19}$$

It should be noted that $\mu_{\mathrm{o}}(k)$ only depends on the weighting functions $\alpha(k)$, $\beta(k)$, and on the SNR at the output of the MVDR beamformer $\rho(k)$. The SNR can be estimated using e.g. the decision-directed approach in [19] or the cepstro-temporal smoothing-based estimator in [20].

Fig. 2 depicts the postfilter gain $G(k)$ for different choices of the trade-off parameter as the SNR varies from -20 dB to 20 dB. For SNRs lower than 0 dB, using the proposed trade-off parameter when no weights are applied, i.e., $\alpha(k) = 1, \beta(k) = 1$, yields a more aggressive gain function than the MWF, i.e., a higher noise reduction performance as well as a higher speech distortion. For SNRs greater than 0 dB the proposed method yields a less aggressive gain function than the MWF, i.e., a lower noise reduction performance as well as lower speech distortion. Weighting the speech distortion term more, i.e., $\alpha(k) = 2$, $\beta(k) = 1$, or the noise power term more, i.e., $\alpha(k) = 1$, $\beta(k) = 2$, shifts the gain function to the left or right respectively.

3.2. Broadband trade-off parameter

Using the narrowband parameter in (19) is advantageous in order to account for the SNR differences in different frequency bands.

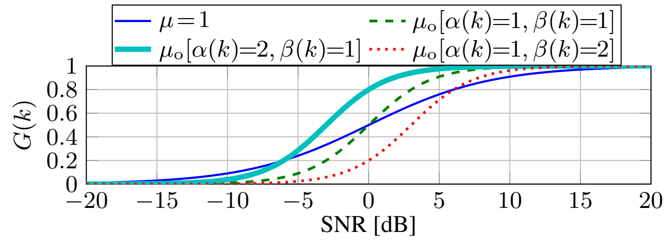


Fig. 2: Postfilter gain as a function of the SNR for $\mu(k) = 1$ and for the proposed parameter $\mu_{\mathrm{o}}(k)$ with different choices of the weighting functions

However in case of large SNR differences the trade-off parameter might vary significantly, resulting in large variations in $G(k)$. Such large variations might lead to undesirable spectral outliers. Hence in the following, we propose extending the method discussed above to compute a broadband trade-off parameter μ.

The weighted broadband speech distortion $\Psi_{\alpha\mathbf{x}}$ and the weighted broadband noise power $\Psi_{\beta\mathbf{v}}$ are defined as the summation of their respective narrowband counterparts, i.e.,

$$\boxed{\Psi_{\alpha\mathbf{x}} = \sum_{k=0}^{K-1}\psi_{\alpha\mathbf{x}}(k) \;\text{ and }\; \Psi_{\beta\mathbf{v}} = \sum_{k=0}^{K-1}\psi_{\beta\mathbf{v}}(k)} \tag{20}$$

with K denoting the total number of frequency bins and $\psi_{\alpha\mathbf{x}}(k)$ and $\psi_{\beta\mathbf{v}}(k)$ expressed as a function of μ. The curvature of the parametric plot of $\Psi_{\beta\mathbf{v}}$ versus $\Psi_{\alpha\mathbf{x}}$ is defined similarly as in (12), where the derivatives can be computed as the summation of the respective narrowband derivatives in (13) to (16). Since no analytical solution can be found for the parameter μ that maximizes the curvature of $\Psi_{\beta\mathbf{v}}$ versus $\Psi_{\alpha\mathbf{x}}$, an iterative optimization technique has been used. The analytical expression for the gradient of the curvature has been provided to the optimization routine in order to improve its numerical robustness and convergence speed. However, this expression has been omitted here due to space constraints.

4. EXPERIMENTAL RESULTS

In this section the performance when using the MWF, i.e., $\mu = 1$, is compared to the performance when using the proposed method for the selection of the trade-off parameter in the $\mathrm{MWF}_{\mathrm{SDW}}$.

4.1. Trade-off parameters

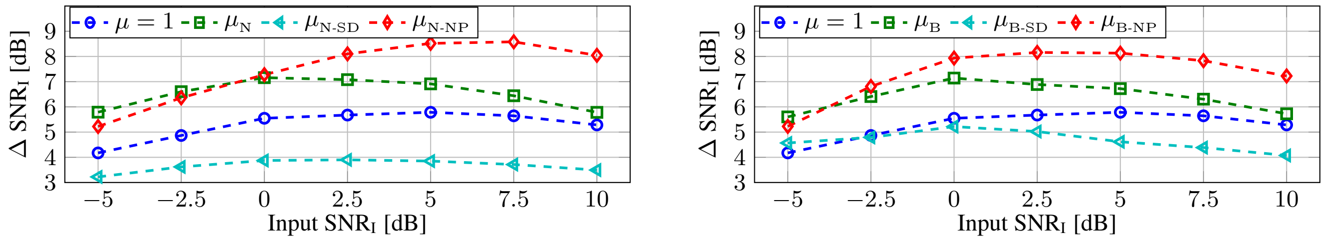
Within the proposed method, the following 3 alternative choices of the narrowband trade-off parameter are evaluated:

i) no weights are applied to the speech distortion and noise power terms, i.e., $\mu_{\mathrm{N}} = 1/\rho(k,l)$,

ii) the speech distortion term is weighted more, i.e., $\mu_{\text{N-SD}} = 1/[\alpha^2(k)\rho(k,l)]$ with $\alpha(k) \geq 1$,

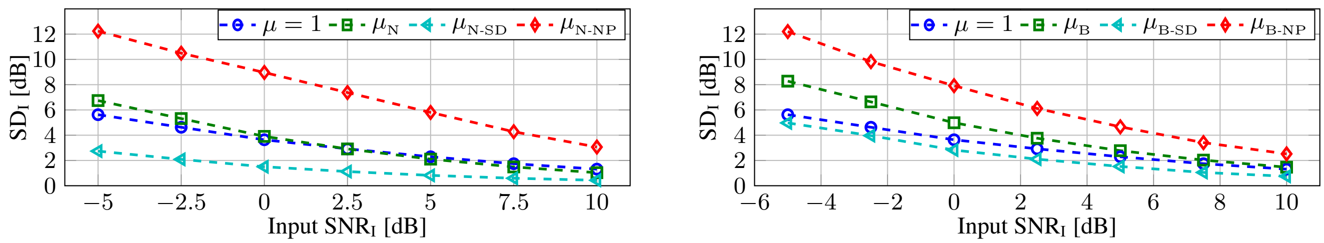
iii) the noise power term is weighted more, i.e., $\mu_{\text{N-NP}} = \beta^2(k)/\rho(k,l)$ with $\beta(k) \geq 1$,

with $\alpha(k)$ and $\beta(k)$ determined using the following simple approach based on the speech intelligibility index [17]. In [17] each frequency bin is assigned an intelligibility index to reflect how much a performance improvement in that bin contributes to the overall speech intelligibility improvement. In this work the intelligibility indexes are scaled between 1 and 10, which are lower and upper bounds selected such that the trade-off parameter stays within reasonable values. By setting $\alpha(k)$ and $\beta(k)$ to the scaled intelligibility indexes, speech distortion or noise power are weighted more in frequency bins with



(a) Narrowband trade-off parameter (b) Broadband trade-off parameter

Fig. 3: Intelligibility weighted SNR improvement using the fixed parameter $\mu = 1$ and the proposed parameters



(a) Narrowband trade-off parameter (b) Broadband trade-off parameter

Fig. 4: Intelligibility weighted speech distortion using the fixed parameter $\mu = 1$ and the proposed parameters

a high speech intelligibility index. Furthermore, using the method described in Section 3.2 also the broadband trade-off parameters for cases i) – iii) have been computed, referred to as μ_{B}, $\mu_{\mathrm{B\text{-}NP}}$, and $\mu_{\mathrm{B\text{-}SD}}$.

4.2. Setup and performance measures

We have considered a scenario with $M = 2$ microphones placed 5 cm apart and a single speech source located at 0°. The speech components were generated using measured room impulse responses with reverberation time $T_{60} \approx 450$ ms [21]. The noise components consisted of nonstationary babble speech generated using the algorithm in [22] under the assumption that the sound field is diffuse. The performance for several intelligibility weighted input SNRs ranging from -5 dB to 10 dB has been investigated. The MVDR filter coefficients have been computed using anechoic steering vectors assuming knowledge of the direction-of-arrival of the speech source and a theoretically diffuse noise correlation matrix. The signals were processed at a sampling frequency $f_s = 16$ kHz using a weighted overlap-add framework with a block size of 512 samples and an overlap of 50% between successive blocks. The cepstro-temporal smoothing-based approach in [20] has been used to estimate the SNR at the output of the MVDR beamformer. The minimum gain of the postfilters has been set to -10 dB. In order to avoid temporal outliers, a moving average smoothing over 5 time blocks has been applied to the obtained trade-off parameters. The minimum value allowed for the trade-off parameters has been set to 0.1. Since the performance of the MVDR beamformer is not relevant within the scope of this paper, the performance for the different trade-off parameters has been evaluated with respect to the beamformer output using the intelligibility weighted SNR improvement ΔSNR$_{\mathrm{I}}$ and the intelligibility weighted speech distortion SD$_{\mathrm{I}}$ computed as in [6].

4.3. Results

Fig. 3a and 4a depict the ΔSNR$_{\mathrm{I}}$ and the SD$_{\mathrm{I}}$ values for $\mu = 1$ and for the different choices of the narrowband parameter. It is shown that using μ_{N} results in a systematic improvement of 1 dB or higher in intelligibility weighted SNR in comparison to using $\mu = 1$. For high input SNRs, this improvement causes no additional speech distortion as can be seen in Fig. 4a. However, for low input SNRs using μ_{N} causes a higher speech distortion than $\mu = 1$ since the applied gain function is more aggressive. Furthermore, putting more emphasis on the speech distortion term, i.e., using $\mu_{\mathrm{N\text{-}SD}}$, yields a lower ΔSNR$_{\mathrm{I}}$ in comparison to using $\mu = 1$ while decreasing the speech distortion. On the other hand, putting more emphasis on the noise power, i.e., using $\mu_{\mathrm{N\text{-}NP}}$, results in a significantly higher improvement in intelligibility weighted SNR at the cost of increased speech distortion. At low input SNRs however, ΔSNR$_{\mathrm{I}}$ using $\mu_{\mathrm{N\text{-}NP}}$ is not higher than when using μ_{N}, which we believe occurs due to errors in the SNR estimation at low input SNRs.

In order to evaluate the performance when using the proposed method to select a broadband trade-off parameter, Fig. 3b and 4b depict the ΔSNR$_{\mathrm{I}}$ and the SD$_{\mathrm{I}}$ for $\mu = 1$ and for the different choices of the broadband parameter. Similarly as for the narrowband comparisons, using μ_{B} yields a higher ΔSNR$_{\mathrm{I}}$ than $\mu = 1$ at the cost of increased speech distortion. When more emphasis is put on the speech distortion term, i.e., using $\mu_{\mathrm{B\text{-}SD}}$, the noise reduction performance and the speech distortion are slightly decreased in comparison to using $\mu = 1$. On the other hand, when the noise power term is weighted more, i.e., using $\mu_{\mathrm{B\text{-}NP}}$, the noise reduction performance is increased at the cost of increased speech distortion.

Finally, comparing the performance of the narrowband and broadband parameters, it can be said that using a narrowband trade-off parameter is more advantageous since it typically yields a higher noise reduction performance (cf. Fig. 3a and 3b) at a lower speech distortion (cf. Fig. 4a and 4b). However, subjective listening tests are necessary in order to establish whether these differences are significant.

5. CONCLUSION

In this paper it has been proposed to select the trade-off parameter in the MWF$_{\mathrm{SDW}}$ as the one that maximizes the curvature of the parametric plot of noise power versus speech distortion. The speech distortion and the noise power terms can be weighted in advance, e.g. based on perceptually motivated criteria. Experimental results have shown that in comparison to the MWF, using the proposed trade-off parameter improves the intelligibility weighted SNR without significantly affecting the speech distortion at positive SNRs.

6. REFERENCES

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

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

[3] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, Springer, Berlin, Germany, 2008.

[4] S. Gannot and I. Cohen, "Adaptive beamforming and postfiltering," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Springer, Berlin, Germany, 2008.

[5] M. Souden, J. Benesty, and S. Affes, "A study of the LCMV and MVDR noise reduction filters," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4925–4935, Sept. 2010.

[6] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction," in *Speech Communication, special issue on Speech Enhancement*, Jul.-Aug. 2007, vol. 49, pp. 636–656.

[7] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*, S. Haykin and K. J. R. Liu, Eds. Wiley, USA, 2010.

[8] K. Ngo, A. Spriet, M. Moonen, J. Wouters, and S. H. Jensen, "Variable speech distortion weighted multichannel Wiener filter based on soft output voice activity detection for noise reduction in hearing aids," in *Proc. International Workshop on Acoustic Echo and Noise Control*, Seattle, Washington USA, Sept. 2008.

[9] K. Ngo, M. Moonen, S. H. Jensen, and J. Wouters, "A flexible speech distortion weighted multi-channel Wiener filter for noise reduction in hearing aids," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic, May 2011, pp. 2528–2531.

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

[11] B. Defraene, K. Ngo, T. van Waterschoot, M. Diehl, and M. Moonen, "A psychoacoustically motivated speech distortion weighted multi-channel Wiener filter for noise reduction," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, Kyoto, Japan, Mar. 2012, pp. 4637–4640.

[12] S. Gazor and W. Zhang, "A soft voice activity detector based on a Laplacian-Gaussian model," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 498–505, Sept. 2003.

[13] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proc. of the IEEE*, vol. 88, no. 4, pp. 451–515, Apr. 2000.

[14] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," in *Microphone Arrays*, M. Brandstein and D. Ward, Eds. Springer, Berlin, Germany, 2001.

[15] P. Hansen and D. O'Leary, "The use of the L-Curve in the regularization of discrete ill-posed problems," *SIAM Journal on Scientific Computing*, vol. 14, no. 6, pp. 1487–1503, Jan. 1993.

[16] I. Kodrasi, S. Goetze, and S. Doclo, "Regularization for partial multichannel equalization for speech dereverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 9, pp. 1879–1890, Sept. 2013.

[17] Acoustical Society of America, ANSI S3.5-1997, *American National Standard Methods for Calculation of the Speech Intelligibility Index*, June 1997.

[18] S. Sternberg, *Curvature in Mathematics and Physics*, Dover Pubications, USA, 2012.

[19] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[20] C. Breithaupt, T. Gerkmann, and R. Martin, "A novel a priori SNR estimation approach based on selective cepstro-temporal smoothing," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, USA, Mar. 2008, pp. 4897–4900.

[21] J. Wen, N. D. Gaubitch, E. A. P. Habets, T. Myatt, and P. A. Naylor, "Evaluation of speech dereverberation algorithms using the MARDY database," in *Proc. International Workshop on Acoustic Echo and Noise Control*, Paris, France, Sept. 2006.

[22] E. A. P. Habets, I. Cohen, and S. Gannot, "Generating nonstationary multisensor signals under a spatial coherence constraint," *Journal of the Acoustical Society of America*, vol. 124, no. 5, pp. 2911–2917, Nov. 2008.