Katholieke Universiteit Leuven

Departement Elektrotechniek

ESAT-SISTA/TR 09-253

A Modified Multi-Channel Wiener Filter based Noise Reduction in Hearing Aids[1]

Kim Ngo[2], Ann Spriet[2,3], Marc Moonen[2],
Jan Wouters [3], Søren Holdt Jensen [4]

April 2010
Internal Report

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.be* in the directory *pub/sista/kngo/reports/09-253.pdf*

[2]K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group SCD (SISTA) Kasteelpark Arenberg 10, 3001 Leuven, Belgium, Tel. +32 16 321797, Fax +32 16 321970, WWW: *http://homes.esat.kuleuven.be/~kngo*. E-mail: *kim.ngo@esat.kuleuven.be*. This research work was carried out at the ESAT Laboratory of Katholieke Universiteit Leuven, in the frame of the EST-SIGNAL Marie-Curie Fellowship program (http://est-signal.i3s.unice.fr) under contract No. MEST-CT-2005-021175, the Concerted Research Action GOA-Ambiorics, the Concerted Research Action GOA-MaNet, 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, and the K.U.Leuven Research Council CoE EF/05/006 (Optimization in Engineering; OPTEC). The scientific responsibility is assumed by its authors.

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

[4]Aalborg University, Department of Electronic Systems, MISP, Niels Jernes Vej 12 A6-3, 9220 Aalborg, Denmark, E-mail: shj@es.aau.dk

Abstract

In this paper, a multi-channel noise reduction algorithm is presented based on a Speech Distortion Weighted Multi-channel Wiener Filter (SDW-MWF) approach that incorporates a flexible weighting factor, which is an extension of the SDW-MWF incorporating the conditional Speech Presence Probability (SPP). A typical SDW-MWF uses a fixed weighting factor to trade-off between noise reduction and speech distortion without taking speech presence or speech absence into account. Consequently, the improvement in noise reduction comes at the cost of a higher speech distortion since the speech dominant segments and the noise dominant segments are weighted equally. Based on a two-state speech model with a noise-only and a speech+noise state, a solution is introduced that allows for a more flexible trade-off between noise reduction and speech distortion. Experimental results with hearing aid scenarios demonstrate that the proposed SDW-MWF incorporating the flexible weighting factor improves the signal-to-noise-ratio with lower speech distortion compared to a typical SDW-MWF and the SDW-MWF incorporating the conditional SPP.

A Modified Multi-Channel Wiener Filter based Noise Reduction in Hearing Aids

Kim Ngo*, *Student Member, IEEE*, Ann Spriet, Marc Moonen, *Fellow, IEEE*, Jan Wouters and Søren Holdt Jensen, *Senior Member, IEEE*

Abstract

In this paper, a multi-channel noise reduction algorithm is presented based on a Speech Distortion Weighted Multi-channel Wiener Filter (SDW-MWF) approach that incorporates a flexible weighting factor, which is an extension of the SDW-MWF incorporating the conditional Speech Presence Probability (SPP). A typical SDW-MWF uses a fixed weighting factor to trade-off between noise reduction and speech distortion without taking speech presence or speech absence into account. Consequently, the improvement in noise reduction comes at the cost of a higher speech distortion since the speech dominant segments and the noise dominant segments are weighted equally. Based on a two-state speech model with a noise-only and a speech+noise state, a solution is introduced that allows for a more flexible trade-off between noise reduction and speech distortion. Experimental results with hearing aid scenarios demonstrate that the proposed SDW-MWF incorporating the flexible weighting factor improves the signal-to-noise-ratio with lower speech distortion compared to a typical SDW-MWF and the SDW-MWF incorporating the conditional SPP.

This research work was carried out at the ESAT Laboratory of Katholieke Universiteit Leuven, in the frame of the EST-SIGNAL Marie-Curie Fellowship program (http://est-signal.i3s.unice.fr) under contract No. MEST-CT-2005-021175, the Concerted Research Action GOA-Ambiorics, the Concerted Research Action GOA-MaNet, 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, and the K.U.Leuven Research Council CoE EF/05/006 (Optimization in Engineering; OPTEC). The scientific responsibility is assumed by its authors.

K. Ngo and A. Spriet and M. Moonen are with the Department of Electrical Engineering, Katholieke Universiteit Leuven, ESAT-SCD, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium (e-mails: {kim.ngo,ann.spriet,marc.moonen}@esat.kuleuven.be).

J. Wouters and A. Spriet are with the Division of Experimental Otorhinolaryngology, Katholieke Universiteit Leuven, ExpORL, O.& N2, Herestraat 49/721, B-3000 Leuven, Belgium (e-mail: jan.wouters@med.kuleuven.be).

S.H. Jensen is with the Department of Electronic Systems, Aalborg University, Niels Jernes Vej 12, DK-9220 Aalborg, Denmark (e-mail: shj@es.aau.dk).

Index Terms

Multi-channel Wiener filter, noise reduction, distortion, speech presence probability, hearing aids.

I. INTRODUCTION

Background noise (from competing speakers, traffic etc.) is a significant problem for hearing impaired people who indeed have more difficulty understanding speech in noise and so in general need a higher signal-to-noise-ratio (SNR) than people with normal hearing [1]. To overcome this problem both single-channel and multi-channel noise reduction algorithms have been proposed and implemented in hearing aids. The objective of these noise reduction algorithms is to maximally reduce the noise while minimizing speech distortion.

One of the first proposed single-channel noise reduction algorithms is spectral subtraction [2], which is based on the assumption that the noise is additive and that the clean speech spectrum can be obtained by subtracting an estimate of the noise spectrum from the noisy speech spectrum. The noise spectrum is updated during periods where the speech is absent, as detected by a Voice Activity Detection (VAD). Another well known single-channel noise reduction algorithm is the Ephraim and Malah noise suppressor [3] [4], which estimates the amplitude of the clean speech spectrum in the spectral or in the log-spectral domain based on a Minimum Mean Square Error (MMSE) criterion. Common to these techniques are usually noticeable artifacts known as musical noise [5] mainly caused by the short-time spectral attenuation, the nonlinear filtering and an inaccurate estimate of the noise characteristic. A limitation of single-channel noise reduction is that only differences in temporal and spectral signal characteristics can be exploited. For example, in a multiple speaker scenario also known as the cocktail party problem the speech (desired speaker) and the noise (competing speakers) considerably overlap in time and frequency. This makes it difficult for single-channel noise reduction algorithms to suppress the noise without reducing speech intelligibility and introducing speech distortion or musical noise.

However, in most scenarios, the desired speaker and the noise sources are physically located at different positions. Multi-channel noise reduction algorithms can then exploit both spectral and spatial characteristics of the speech and the noise. The Frost beamformer and the Generalized Sidelobe Canceler [6] [7] [8] are well known multi-channel noise reduction algorithms. The basic idea is to steer a beam towards the desired speaker while reducing the background noise coming from other directions. Another known multi-channel noise reduction algorithm is the Multi-channel Wiener filter (MWF) that provides an MMSE estimate of the speech component in one of the input signals. The extension from MWF

to Speech Distortion Weighted MWF (SDW-MWF) [9] [10] [11] allows for a trade-off between noise reduction and speech distortion.

Traditionally, these multi-channel noise reduction algorithms adopt a (short-time) fixed filtering under the implicit hypothesis that the speech is present at all time. However, while the noise can indeed be continously present, the speech signal typically contains many pauses. Furthermore, the speech may not be present at all frequencies even during speech segments. It has been shown in single-channel noise reduction schemes that by incorporating the conditional Speech Presence Probability (SPP) in the gain function or in the noise spectrum estimation a better performance can be achieved compared to traditional methods [4] [12] [13] [14]. In these approaches the conditional SPP is estimated for each frequency bin and each frame by a soft-decision approach, which exploits the strong correlation of speech presence in neighboring frequency bins of consecutive frames.

A typical SDW-MWF uses a fixed weighting factor to trade-off between noise reduction and speech distortion without taking speech presence or speech absence into account. This means that the speech dominant segments and the noise dominant segments are weighted equally in the noise reduction process. Consequently, the improvement in noise reduction comes at the cost of a higher speech distortion. In [15] [16] an SDW-MWF approach that incorporates the conditional SPP in the trade-off between noise reduction and speech distortion has been introduced. In speech dominant segments it is then desirable to have less noise reduction to avoid speech distortion, while in noise dominant segments it is desirable to have as much noise reduction as possible.

This paper presents an SDW-MWF approach that incorporates a flexible weighting factor based on a two-state speech model with a noise-only and a speech+noise state, which is an extension of the SDW-MWF incorporating the conditional SPP. The flexible weighting factor is introduced to allow for a more flexible trade-off between noise reduction and speech distortion. Experimental results with hearing aid scenarios demonstrate that the proposed SDW-MWF incorporating a flexible weighting factor improves the signal-to-noise-ratio with lower speech distortion compared to a typical SDW-MWF and the SDW-MWF incorporating the conditional SPP.

The paper is organised as follows. Section II describes the general set-up of a multi-channel noise reduction algorithm. Section III reviews the SDW-MWF incorporating the conditional SPP. Section IV explains the concept behind introducing the flexible weighting factor in the SDW-MWF. In Section V experimental results are presented. The work is summarized in Section VI.

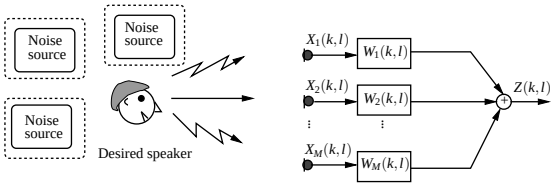


Fig. 1. Multi-channel noise reduction set-up in an environment with one or more noise sources and a desired speaker.

II. Speech Distortion Weighted Multi-channel Wiener Filter (SDW-MWF$_\mu$)

A general set-up of a multi-channel noise reduction is shown in figure 1, with M microphones in an environment with one or more noise sources and a desired speaker. Let $X_i(k,l), i = 1, ..., M$ denote the frequency-domain microphone signals

$$X_i(k,l) = X_i^s(k,l) + X_i^n(k,l) \tag{1}$$

where k is the frequency bin index, and l the frame index of a short-time Fourier transform (STFT), and the superscripts s and n are used to refer to the speech and the noise contribution in a signal, respectively. Let $\mathbf{X}(k,l) \in \mathbb{C}^{M\times 1}$ be defined as the stacked vector

$$\mathbf{X}(k,l) = [X_1(k,l)\ X_2(k,l)\ ...\ X_M(k,l)]^T \tag{2}$$

$$= \mathbf{X}^s(k,l) + \mathbf{X}^n(k,l) \tag{3}$$

where the superscript T denotes the transpose. In addition, we define the speech+noise, the speech and the noise correlation matrices as

$$\mathbf{R}^x(k,l) = \varepsilon\{\mathbf{X}(k,l)\mathbf{X}^H(k,l)\} \tag{4}$$

$$\mathbf{R}^s(k,l) = \varepsilon\{\mathbf{X}^s(k,l)\mathbf{X}^{s,H}(k,l)\} \tag{5}$$

$$\mathbf{R}^n(k,l) = \varepsilon\{\mathbf{X}^n(k,l)\mathbf{X}^{n,H}(k,l)\} \tag{6}$$

where $\varepsilon\{\}$ denotes the expectation operator, H denotes Hermitian transpose. The MWF optimally estimates the speech signal, based on a Minimum Mean Squared Error (MMSE) criterion, i.e.,

$$\mathbf{W}_{\text{MMSE}}(k,l) = \arg\min_{\mathbf{W}}\ \varepsilon\{|X_1^s(k,l) - \mathbf{W}^H\mathbf{X}(k,l)|^2\} \tag{7}$$

where the desired signal in this case is the (unknown) speech component $X_1^s(k,l)$ in the first microphone signal. The MWF has been extended to the SDW-MWF$_\mu$ that allows for a trade-off between noise

reduction and speech distortion using a weighting factor μ [9] [10]. If the speech and the noise signals are statistically independent the design criterion of the SDW-MWF$_\mu$ is given by

$$\mathbf{W}_\mu(k,l) = \arg\min_{\mathbf{W}} \ \varepsilon\{|X_1^s(k,l) - \mathbf{W}^H\mathbf{X}^s(k,l)|^2\} + \mu\varepsilon\{|\mathbf{W}^H\mathbf{X}^n(k,l)|^2\}. \tag{8}$$

The SDW-MWF$_\mu$ is then given by

$$\mathbf{W}_\mu(k,l) = \left[\mathbf{R}^s(k,l) + \mu\mathbf{R}^n(k,l)\right]^{-1}\mathbf{R}^s(k,l)\mathbf{e}_1 \tag{9}$$

where the $M \times 1$ vector $\mathbf{e}_1$ equals the first canonical vector defined as $\mathbf{e}_1 = [1 \quad 0 \quad ... \quad 0]^T$ and the correlation matrices can be estimated as

$$\begin{aligned} H_0(k,l): &\begin{cases} \mathbf{R}^n(k,l) = \alpha_n\mathbf{R}_n(k,l) + (1-\alpha_n)\mathbf{X}(k,l)\mathbf{X}^H(k,l) \\ \mathbf{R}^x(k,l) = \mathbf{R}_x(k,l) \end{cases} \\ H_1(k,l): &\begin{cases} \mathbf{R}^x(k,l) = \alpha_x\mathbf{R}^x(k,l) + (1-\alpha_x)\mathbf{X}(k,l)\mathbf{X}^H(k,l) \\ \mathbf{R}^n(k,l) = \mathbf{R}^n(k,l) \end{cases} \end{aligned} \tag{10}$$

where $H_0(k,l)$ and $H_1(k,l)$ represent speech absence and speech presence in frequency bin k and frame l, respectively. The second-order statistics of the noise are assumed to be (short-term) stationary which means that $\mathbf{R}^s(k,l)$ can be estimated as $\mathbf{R}^s(k,l) = \mathbf{R}^x(k,l) - \mathbf{R}^n(k,\tilde{l})$ where $\mathbf{R}^x(k,l)$ and $\mathbf{R}^n(k,\tilde{l})$ are estimated (i.e. adapted) during periods of speech+noise (l) and periods of noise-only $(\tilde{l})$, respectively (and "frozen" otherwise). For μ = 1 the SDW-MWF$_\mu$ reduces to the MWF, while for $\mu > 1$ the residual noise level will be reduced at the cost of a higher speech distortion. The output $Z(k,l)$ of the SDW-MWF$_\mu$ can then be written as

$$Z(k,l) = \mathbf{W}_\mu^H(k,l)\mathbf{X}(k,l). \tag{11}$$

Some properties and drawbacks of the SDW-MWFμ are:

- A SDW-MWFμ is uniquely based on the second-order statistics and in the estimation of the speech+noise and the noise-only correlation matrices (containing spectral as well as spatial information) an averaging time window of 2-3 seconds is typically used to achieve a reliable estimate.
- In the SDW-MWF$_\mu$ the weighting factor μ is a fixed value for each frame and each frequency.
- As a consequence, any improvement in noise reduction (by setting μ to a larger value) comes at the cost of a higher speech distortion.

The SDW-MWF$_{\text{SPP}}$, which is reviewed in the next section, improves the spectral filtering such that the following items are taken into consideration [1] [15] [17] [18]:

- Speech is a spectrally non-stationary signal and can be considered stationary only in a short time window, shorther than the 2-3 seconds time window used for the estimation of $\mathbf{R}^s(k,l)$, and so the spectral non-stationarity is not captured in $\mathbf{R}^s(k,l)$.
- The speech signal contains many pauses while the noise is assumed to be continously present.

III. SDW-MWF incorporating the conditional Speech Presence probability (SDW-MWF$_{\text{SPP}}$)

The SDW-MWF$_{\text{SPP}}$ [15] incorporates the conditional SPP in the SDW-MWFμ to allow for a faster tracking of the spectral non-stationarity of the speech, as well as for exploiting the fact that speech may not be present at all time.

A. Conditional Speech Presence Probability

The conditional SPP is estimated for each frequency bin and each frame by a soft-decision approach [13] [17] [19], which exploits the strong correlation of speech presence in neighboring frequency bins of consecutive frames. A two-state model for speech events can be expressed given two hypotheses $H_0(k,l)$ and $H_1(k,l)$ which represent speech absence and speech presence in frequency bin k and frame l, respectively, i.e.,

$$\begin{aligned} H_0(k,l) &: X_i(k,l) = X_i^n(k,l) + 0 \cdot X_i^s(k,l) \\ H_1(k,l) &: X_i(k,l) = X_i^n(k,l) + 1 \cdot X_i^s(k,l), \end{aligned} \tag{12}$$

where the i-th microphone signal is used as a reference (in our case the first microphone signal $X_1(k,l)$ is used). The inclusion of the second term in the definition of H_0 will be explained in Section IV. Assuming a complex Gaussian distribution of the STFT coefficients for both the speech and the noise, the conditional probability density functions (PDF) of the observed signals are given by

$$\begin{aligned} p(X_i(k,l)|H_0(k,l)) &= \frac{1}{\pi\lambda_i^n(k,l)} \exp\left\{-\frac{|X_i(k,l)|^2}{\lambda_i^n(k,l)}\right\} \\ p(X_i(k,l)|H_1(k,l)) &= \frac{1}{\pi(\lambda_i^s(k,l)+\lambda_i^n(k,l))} \\ &\quad \exp\left\{-\frac{|X_i(k,l)|^2}{\lambda_i^s(k,l)+\lambda_i^n(k,l)}\right\} \end{aligned} \tag{13}$$

where $\lambda_i^s(k,l) \triangleq \varepsilon\{|X_i^s(k,l)|^2\}$ and $\lambda_i^n \triangleq \varepsilon\{|X_i^n(k,l)|^2\}$ denote the power spectrum of the speech and the noise, respectively. Applying Bayes rule, the conditional SPP $p(k,l) \triangleq P(H_1(k,l)|X_i(k,l))$ can be written as [4]

$$p(k,l) = \left\{1 + \frac{q(k,l)}{1-q(k,l)}(1+\xi(k,l))\exp(-\upsilon(k,l))\right\}^{-1} \tag{14}$$

where $q(k,l) \triangleq P(H_0(k,l))$ is the a priori speech absence probability (SAP), $\xi(k,l)$ and $\gamma(k,l)$ denote the a priori SNR and the a posteriori SNR, respectively. Details on the estimation of the SAP, the a priori SNR and the a posteriori SNR can be found in [13] [15] [17] [19].

For the sake of conciseness the frequency bin index k and frame index l are omitted from now on in $\mathbf{X}(k,l)$, $\mathbf{X}^s(k,l)$, $\mathbf{X}^n(k,l)$ and $X_1^s(k,l)$.

B. Derivation of SDW-MWF$_{SPP}$

The conditional SPP in (14) and the two-state model in (12) for speech events can be incorporated into the optimization criterion of the SDW-MWFμ, leading to a weighted average where the first term corresponds to H_1 and is weighted by the probability that speech is present, while the second term corresponds to H_0 and is weighted by the probability that speech is absent, i.e.,

$$\begin{aligned}\mathbf{W}_{\text{SPP}}(k,l) = \arg\min_{\mathbf{W}}\ & p(k,l)\varepsilon\{|X_1^s - \mathbf{W}^H\mathbf{X}|^2|H_1\} \\ & + (1-p(k,l))\varepsilon\{|\mathbf{W}^H\mathbf{X}|^2|H_0\}\end{aligned} \tag{15}$$

where $p(k,l)$ is the conditional probability that speech is present and $(1-p(k,l))$ is the conditional probability that speech is absent. The solution is then given by

$$\begin{aligned}\mathbf{W}_{\text{SPP}}(k,l) &= \Big[p(k,l)\varepsilon\{\mathbf{X}\mathbf{X}^H|H_1\} + (1-p(k,l))\varepsilon\{\mathbf{X}\mathbf{X}^H|H_0\}\Big]^{-1} \\ &\quad p(k,l)\varepsilon\{\mathbf{X}^s X_1^{s,H}\}, \\ &= \Big[p(k,l)\varepsilon\{\mathbf{X}^s\mathbf{X}^{s,H}\} + p(k,l)\varepsilon\{\mathbf{X}^n\mathbf{X}^{n,H}\} + \\ &\quad (1-p(k,l))\varepsilon\{\mathbf{X}^n\mathbf{X}^{n,H}\}\Big]^{-1} p(k,l)\varepsilon\{\mathbf{X}^s X_1^{s,H}\},\end{aligned} \tag{16}$$

which can be written as

$$\mathbf{W}_{\text{SPP}}(k,l) = \left[\mathbf{R}^s(k,l) + \left(\tfrac{1}{p(k,l)}\right)\mathbf{R}^n(k,l)\right]^{-1}\mathbf{R}^s(k,l)\mathbf{e}_1. \tag{17}$$

Compared to (9) the fixed weighting factor μ is replaced by $\frac{1}{p(k,l)}$, which is now adjusted for each frequency bin k and for each frame l, making the SDW-MWF$_{\text{SPP}}$ change with a faster dynamic. The SDW-MWF$_{\text{SPP}}$ offers more noise reduction when $p(k,l)$ is small, i.e., for noise dominant segments, and less noise reduction when $p(k,l)$ is large, i.e., for speech dominant segments.

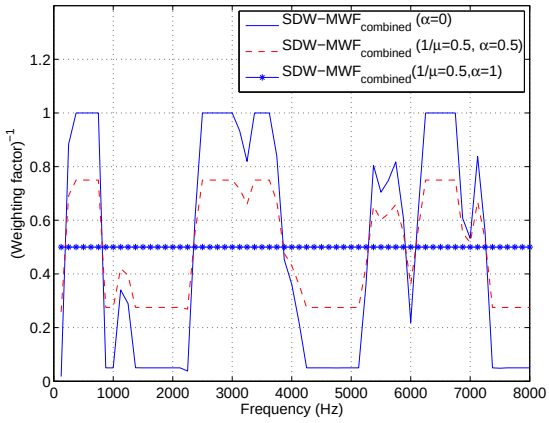


Fig. 2. Illustration of the different configuration of (weighting factor)$^{-1}$ used in (18).

C. Combined solution

A problem with the SDW-MWF$_{\text{SPP}}$ derived in (17) is that the conditional SPP tends to show significant variation over different frequency bins which then also causes the noise reduction to vary significantly over different frequency bins. In [15] a combined solution SDW-MWF$_{\text{combined}}$ is also proposed, which in one extreme case corresponds to the SDW-MWF$_{\text{SPP}}$ and in the other extreme case corresponds to the SDW-MWF$_{\mu}$. The combined solution can be written as

$$\mathbf{W}_{\text{combined}}(k,l) = \left[\mathbf{R}^s(k,l) + \left(\frac{1}{\alpha(\frac{1}{\mu})+(1-\alpha)p(k,l)}\right)\mathbf{R}^n(k,l)\right]^{-1} \mathbf{R}^s(k,l)\mathbf{e}_1 \tag{18}$$

where α is a trade-off factor between SDW-MWF$_{\mu}$ and SDW-MWF$_{\text{SPP}}$. The (weighting factor)$^{-1}$ i.e. $\alpha(\frac{1}{\mu}) + (1-\alpha)p(k,l)$ used in (18) is shown in figure 2 for different configurations, i.e., SDW-MWF$_{\text{combined}}$ (α=1) and SDW-MWF$_{\text{combined}}$ (α=0) and SDW-MWF$_{\text{combined}}$ (α=0.5). The dashed line shows the (weighting factor)$^{-1}$ when $1/\mu$=0.5 and α=0.5. This clearly shows that the combined solution corresponds to a smoothing of the conditional SPP. Since the variations between the speech dominant segments and the noise dominant segments are reduced, the distortion is also reduced. When α=1 the solution corresponds to a fixed μ shown with the (*) marker line. In Section IV a flexible trade-off between noise reduction and speech distortion is introduced.

IV. SDW-MWF INCORPORATING A FLEXIBLE WEIGHTING FACTOR (SDW-MWF$_{\text{FLEX}}$)

Based on the observations made above the SDW-MWF$_{\text{SPP}}$ is modified to further exploit the properties of the conditional SPP and to fully exploit the definition of the two-state speech model. The two-state

speech model given in (12) describes the different states of the noisy speech which leads to the following observations. First, it is clear that the noise reduction in the H_0 state and the H_1 state have a different interpretation, i.e.,

- Reducing the noise in the H_0 state can be related to increasing listening comfort, since speech is not present in the H_0 state, which means that a greater attenuation can be applied.
- Reducing the noise in the H_1 state is a more challenging task since this relates to speech intelligibility and hence the speech distortion weighted concept truly only makes sense in the H_1 state.

Secondly, as described in Section II, the speech correlation matrix $\mathbf{R}^s(k,l)$ and the noise correlation matrix $\mathbf{R}^n(k,\tilde{l})$ are estimated during periods of speech+noise (l) and periods of noise-only $(\tilde{l})$, respectively. This means that,

- In theory the SDW-MWF could be an all zero vector during noise-only periods since then $\mathbf{R}^s(k,l) = 0$.
- In practice $\mathbf{R}^s(k,l)$ is "frozen" during noise-only periods where $\mathbf{R}^n(k,\tilde{l})$ is updated. In fact this is in line with the definition of H_0 in (12), where the "0" indicate, that the speech X_i^s can have a non-zero $\mathbf{R}^s(k,l)$ in H_0, but is not transmitted into X_i.

We then suggest, that if the H_0 state and the H_1 state can be properly detected a more flexible trade-off between noise reduction and speech distortion can be achieved.

To this aim, the parameter $P(l)$ is introduced, which is a binary decision, obtained by averaging the conditional SPP $p(k,l)$ over all frequency bins k

$$P(l) = \begin{cases} 1 & \text{if} \quad \dfrac{1}{K}\displaystyle\sum_{k=1}^{K} p(k,l) \geq \alpha_{\text{frame}} \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

where $P(l) = 1$ means the H_1 state is detected and $P(l) = 0$ means the H_0 state is detected, and α_{frame} is a detection threshold. This $P(l)$ will be used in the operation of SDW-MWF$_{\text{Flex}}$. In Fig. 3 $P(l)$ is plotted for a given speech segment which shows that even in H_1 state there are some frames/samples where the conditional SPP is low. Notice that in this case the noise correlation matrice is kept fixed whereas $p(k,l)$ and $P(l)$ are continously updated. The two key ingredients of the proposed SDW-MWF$_{\text{Flex}}$ are now as follows:

- A weighting factor μ_{H_1} is introduced, which is a function of $p(k,l)$, that defines the amount of noise reduction that can be applied in the H_1 state.
- A weighting factor μ_{H_0} is introduced, which is a constant weighting factor, that defines the amount of noise reduction that can be applied in the H_0 state.

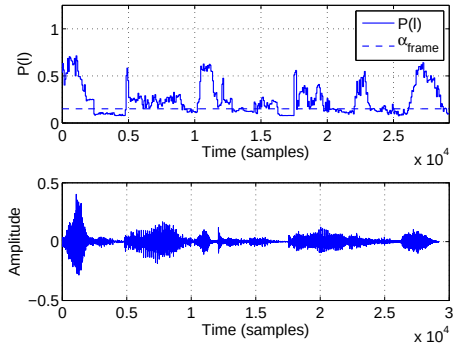


Fig. 3. Illustration of $P(l)$ for a given speech segment.

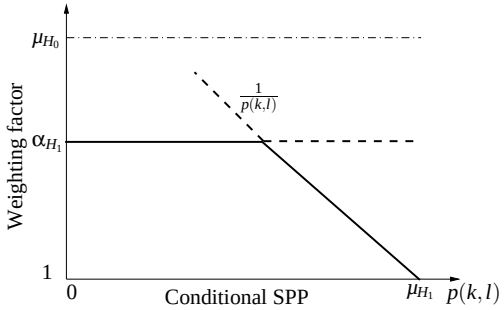


Fig. 4. The weighting factor used in SDW-MWF$_{\text{Flex}}$ as a function of the conditional SPP.

The SDW-MWF$_{\text{Flex}}$ weighting strategy is illustrated in figure 4 which shows the weighting factor as a function of $p(k,l)$. Notice that μ_{H_1} is defined here as $\min(\frac{1}{p(k,l)}, \alpha_{H_1})$, i.e., a function of the conditional SPP $\frac{1}{p(k,l)}$ and a lower threshold α_{H_1}, which is introduced since speech may not be present in all frequency bins even in state H_1. The optimization criterion for SDW-MWF$_{\text{Flex}}$ is given by

$$
\begin{aligned}
\mathbf{W}_{\text{Flex}}(k,l) = \arg\min_{\mathbf{W}} & \\
& P(l)\Big[\max(p(k,l),\tfrac{1}{\alpha_{H_1}})\varepsilon\{|X_1^s - \mathbf{W}^H\mathbf{X}|^2|H_1\} \\
& + (1-\max(p(k,l),\tfrac{1}{\alpha_{H_1}}))\varepsilon\{|\mathbf{W}^H\mathbf{X}|^2|H_0\}\Big] + \\
& (1-P(l))\Big[\tfrac{1}{\mu_{H_0}}\varepsilon\{|X_1^s - \mathbf{W}^H\mathbf{X}^s|^2\} + \varepsilon\{|\mathbf{W}^H\mathbf{X}^n|^2\}\Big] \\
& = \arg\min_{\mathbf{W}} \\
& \Big[P(l)\max(p(k,l),\tfrac{1}{\alpha_{H_1}}) + (1-P(l))\tfrac{1}{\mu_{H_0}}\Big] \\
& \varepsilon\{|X_1^s - \mathbf{W}^H\mathbf{X}^s|^2\} + \Big[P(l) + (1-P(l))\Big]\varepsilon\{|\mathbf{W}^H\mathbf{X}^n|^2\} \\
& = \arg\min_{\mathbf{W}} \\
& \Big[P(l)\max(p(k,l),\tfrac{1}{\alpha_{H_1}}) + (1-P(l))\tfrac{1}{\mu_{H_0}}\Big] \\
& \varepsilon\{|X_1^s - \mathbf{W}^H\mathbf{X}^s|^2\} + \varepsilon\{|\mathbf{W}^H\mathbf{X}^n|^2\}
\end{aligned} \tag{20}
$$

The solution is given by

$$
\mathbf{W}_{\text{Flex}}(k,l) = \Big[\varepsilon\{\mathbf{X}^s\mathbf{X}^{s,H}\} + \gamma(k,l)\varepsilon\{\mathbf{X}^n\mathbf{X}^{n,H}\}\Big]^{-1}\varepsilon\{\mathbf{X}^s X_1^{s,H}\} \tag{21}
$$

with the weighting factor defined as

$$
\begin{aligned}
\gamma(k,l) &= \Big[P(l)\max(p(k,l),\tfrac{1}{\alpha_{H_1}}) + (1-P(l))\tfrac{1}{\mu_{H_0}}\Big]^{-1} \\
&= \Big[P(l)\min(\tfrac{1}{p(k,l)},\alpha_{H_1}) + (1-P(l))\mu_{H_0}\Big].
\end{aligned} \tag{22}
$$

The SDW-MWF$_{\text{Flex}}$ can then be written as

$$
\mathbf{W}_{\text{Flex}}(k,l) = \Big[\mathbf{R}^s + \gamma(k,l)\mathbf{R}^n\Big]^{-1}\mathbf{R}^s\mathbf{e}_1 \tag{23}
$$

The SDW-MWF$_{\text{Flex}}$ is summarized in algorithm 1:

V. Experimental Results

In this section, experimental results for the proposed SDW-MWF$_{\text{Flex}}$ are presented and compared to SDW-MWF$_{\text{SPP}}$ and SDW-MWF$_{\mu}$.

A. Experimental set-up

Simulations have been performed with a 2-microphone behind-the-ear hearing aid mounted on a CORTEX MK2 manikin. The loudspeakers (FOSTEX 6301B) are positioned at 1 meter from the center

Algorithm 1 SDW-MWF$_{\text{Flex}}$

for each frame l and each frequency k **do**
  Estimate noise correlation matrix $\mathbf{R}^n(k,l)$ (4)
  Estimate speech correlation matrix $\mathbf{R}^s(k,l)$ (6)
  Estimate speech presence probability $p(k,l)$ (14)
  H_1 and H_0 detection $P(l)$ (19)
  if $P(l)=1$ **(per-frame decision) then**
    H_1 **state detected**
    Estimate SDW-MWF$_{\text{Flex}}$ (23)
    with $\gamma(k,l)=\min(\frac{1}{p(k,l)},\alpha_{H_1})$ (22)
  else
    H_0 **state detected**
    Estimate SDW-MWF$_{\text{Flex}}$ (23)
    with $\gamma(k,l)=\mu_{H_0}$ (22)
  end if
end for

of the head. The reverberation time $T_{60}=0.21s$. The speech is located at 0° and the two multi-talker babble noise sources are located at 120° and 180°. The speech signal consists of male sentences from Hearing in Noise Test (HINT) for the measurement of speech reception thresholds in quiet and in noise [20] and the noise signal consists of a multi-talker babble from Auditory Tests (Revised), Compact Disc, Auditec [21]. The signals are sampled at 16kHz. An FFT length of 128 with 50% overlap was used. The parameters for estimating the conditional SPP are similiar as in [15].

B. Performance measures

To assess the noise reduction performance the intelligibility-weighted signal-to-noise ratio (SNR) [22] is used which is defined as

$$\Delta\text{SNR}_{\text{intellig}}=\sum_i I_i(\text{SNR}_{i,out}-\text{SNR}_{i,in}) \tag{24}$$

where I_i is the band importance function defined in ANSI S3.5-1997 [23] and where $\text{SNR}_{i,out}$ and $\text{SNR}_{i,in}$ represent the output SNR and the input SNR (in dB) of the *i*-th band, respectively. For measuring the

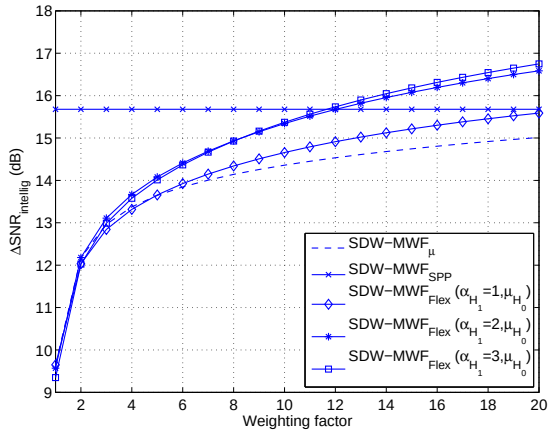


Fig. 5. SNR improvement for α_{H_1}=1,2 and 3 with variable μ_{H_0} compared to SDW-MWF$_\mu$ and SDW-MWF$_{\text{SPP}}$.

signal distortion a frequency-weighted log-spectral signal distortion (SD) is used defined as

$$\text{SD} = \frac{1}{K}\sum_{k=1}^{K}\sqrt{\int_{f_l}^{f_u} w_{ERB}(f)\Big(10log_{10}\frac{P^s_{out,k}(f)}{P^s_{in,k}(f)}\Big)^2\,df} \tag{25}$$

where K is the number of frames, $P^s_{out,k}(f)$ is the output power spectrum of the kth frame, $P^s_{in,k}(f)$ is the input power spectrum of kth frame and f is the frequency index. The SD measure is calculated with a frequency-weighting factor $w_{ERB}(f)$ giving equal weight for each auditory critical band, as defined by the equivalent rectangular bandwidth (ERB) of the auditory filter [24]. Notice that the intelligibility-weighted SNR and the spectral distortion are only computed during frames of speech+noise.

C. Results

In this experiment, for the SDW-MWF$_{\text{Flex}}$, the α_{H_1} is fixed to 1,2 and 3, μ_{H_0} is increased from 1 to 20 and the conditional SPP $p(k,l)$ is estimated according to (14). For SDW-MWF$_\mu$, μ is increased from 1 to 20. The SNR improvement is shown in figure 5 and the speech distortion is shown figure 6. This shows, that the SDW-MWF$_{\text{Flex}}$ outperforms the SDW-MWF$_\mu$ and SDW-MWF$_{\text{SPP}}$ both in SNR improvement and in terms of speech distortion, when the weighting factor μ_{H_0} is increased. Increasing α_{H_1} does show a further improvement in SNR using SDW-MWF$_{\text{Flex}}$ with a small increase in speech distortion.

VI. Conclusion

In this paper a noise reduction procedure SDW-MWF$_{\text{Flex}}$ has been presented that incorporates a flexible weighting factor to trade-off between noise reduction and speech distortion, which is an extension of the SDW-MWF$_{\text{SPP}}$ incorporating the conditional SPP. Based on a two-state speech model, with a noise-only

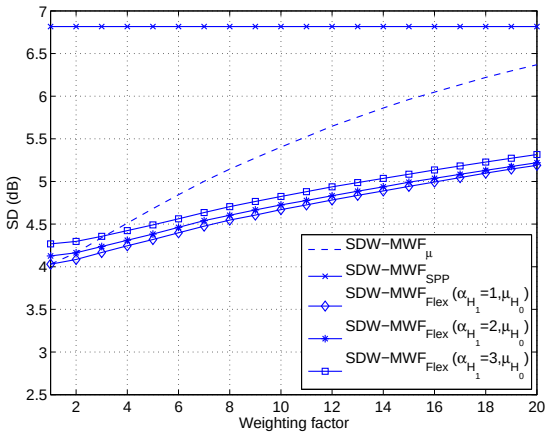


Fig. 6. Speech distortion for α_{H_1}=1,2 and 3 with variable μ_{H_0} compared to SDW-MWF$_\mu$ and SDW-MWF$_{\text{SPP}}$.

(H_0) and a speech+noise (H_1) state, the goal of the SDW-MWF$_{\text{Flex}}$ is to apply an equal amount of noise reduction as in a typical SDW-MWF$_\mu$ in the H_0 state, while in the H_1 state, the goal is to preserve the speech by exploiting the conditional SPP. The SDW-MWF$_{\text{Flex}}$ is found to significantly improve the SNR while the speech distortion is kept low compared to SDW-MWF$_\mu$ and SDW-MWF$_{\text{SPP}}$.

REFERENCES

[1] H. Dillon, *Hearing Aids*. Turramurra, Australia: Boomerang Press, 2001.

[2] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 27, no. 2, pp. 113–120, Apr 1979.

[3] Y. Ephraim and D. Malah, "Speech enhancement using optimal non-linear spectral amplitude estimation," *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '83.*, vol. 8, pp. 1118–1121, Apr 1983.

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

[5] O. Cappe, "Elimination of the musical noise phenomenon with the ephraim and malah noise suppressor," *Speech and Audio Processing, IEEE Transactions on*, vol. 2, no. 2, pp. 345–349, Apr 1994.

[6] I. Frost, O.L., "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[7] L. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Transactions on Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, Jan 1982.

[8] B. Van Veen and K. Buckley, "Beamforming: a versatile approach to spatial filtering," *ASSP Magazine, IEEE*, vol. 5, no. 2, pp. 4–24, April 1988.

[9] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-domain criterion for the speech distortion weighted multichannel wiener filter for robust noise reduction," *Speech Communication*, vol. 7-8, pp. 636–656, Jul. 2007.

[10] A. Spriet, M. Moonen, and J. Wouters, "Stochastic gradient based implementation of spatially pre-processed speech distortion weighted multi-channel wiener filtering for noise reduction in hearing aids," *IEEE Transactions on Signal Processing*, vol. 53, no. 3, pp. 911–625, Mar. 2005.

[11] B. Cornelis, M. Moonen, and J. Wouters, "Comparison of frequency-domain noise reduction strategies based on multichannel wiener filtering and spatial prediction," in *Proc. of the international conference on acoustics, speech and signal processing (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 129–132.

[12] R. McAulay and M. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 28, no. 2, pp. 137–145, Apr 1980.

[13] I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator," *Signal Processing Letters, IEEE*, vol. 9, no. 4, pp. 113–116, Apr 2002.

[14] ——, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 5, pp. 466–475, Sept. 2003.

[15] 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, Article ID 930625, 11 pages, 2009, doi:10.1155/2009/930625.

[16] K. Ngo, A. Spriet, M. Moonen, J. Wouters, and S. Jensen, "Variable speech distortion weighted multichannel wiener filter based on soft output voice activity detection for noise reduction in hearing aids," in *Proc. 11th IWAENC*, Seattle, USA, 2008.

[17] P. Loizou, *Speech Enhancement: Theory and Practice*. Boca Raton, FL: CRC Press, 2007.

[18] H. Levitt, "Noise reduction in hearing aids: a review," *Journal of Rehabilitation Research and Development*, vol. 38, no. 1, pp. 111–121, Jan. 2001.

[19] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Process*, vol. 81, no. 11, pp. 2403–2418, Nov. 2001.

[20] M. Nilsson, S. D. Soli, and A. Sullivan, "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," *jasa*, vol. 95, no. 2, pp. 1085–1099, Feb. 1994.

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

[22] J. E. Greenberg, P. M. Peterson, and P. M. Zurek, "Intelligibility-weighted measures of speech-to-interference ratio and speech system performance," *J. Acoustic. Soc. Am.*, vol. 94, no. 5, pp. 3009–3010, Nov. 1993.

[23] Acoustical Society of America, "ANSI S3.5-1997 American National Standard Methods for calculation of the speech intelligibility index," June 1997.

[24] B. Moore, *An Introduction to the Psychology of Hearing*, 5th ed. Academic Press, 2003.