On the Global Output SNR of the Parameterized Frequency-Domain Multichannel Noise Reduction Wiener Filter

Mehrez Souden, *Student Member, IEEE*, Jacob Benesty, and Sofiène Affes, *Senior Member, IEEE*

***Abstract*—The parameterized multichannel Wiener filter (PMWF) is known to allow for a flexible tuning of speech distortion and noise reduction. In addition, the output signal-to-noise ratio (SNR) is a natural metric that shows the true effect of such a filter on both residual noise and filtered speech. Earlier contributions have only shown that the output SNR of the Wiener filter is larger than the input SNR (as a proof of its effectiveness). However, the effect of the tuning parameter (denoted as β following the notation of [1]) on this metric is not yet understood. In this paper, we prove that the *global* (fullband) output SNR of the PMWF is an increasing function of β but remains below an asymptotic value. As a byproduct, a very simplified proof of the global output SNR improvement is provided.**

***Index Terms*—Microphone array, noise reduction, parameterized multichannel Wiener filter, signal-to-noise ratio (SNR).**

I. Introduction

THE Wiener filter has been extensively used in noise reduction applications due to its simplicity and effectiveness. When used with multiple microphones in speech acquisition systems, this filter takes advantage of both temporal and spatial dimensions to enhance the target speech and attenuate the undesired signals. Unfortunately, it is known that noise cannot be infinitely reduced (except in the case of spatially correlated noise) without the distortion of speech. Then, a tradeoff between noise reduction and speech distortion is commonly achieved by means of the so-called PMWF in the frequency domain [1] or the speech distortion weighted multichannel Wiener filter in the time domain [2].

Time-domain techniques are generally computationally demanding since large matrix calculations are involved and numerical problems are commonly encountered especially as the number of microphones and/or reverberation time increase. Conversely, frequency-domain techniques are generally preferred because each frequency bin can be processed apart from the others. This allows for easier calculations and interesting relationships can be easily found as compared to time-domain approaches. In [1], for instance, a frequency-domain framework for the analysis of the PMWF, of which the noise reduction minimum variance distortionless response (MVDR) beamformer [3], [4] is shown to be a particular case, was investigated. Local (frequency-bin-wise) performance measures, namely, speech distortion index, noise reduction factor, and output SNR of this filter were studied and the effect of the tuning parameter on the tradeoff of noise reduction versus speech distortion was highlighted. Since the effect of the tuning parameter was only seen in the scaling factor that related all filters, it was established that *locally* the tuning parameter has no effect on the output SNR. This result does not correspond to the actual perception of the filtered signal by human listeners.

This work builds upon reference [1]. Indeed, we consider the *global* output SNR of the PMWF and show that it is an increasing function of the tradeoff parameter. To the best of our knowledge, this result is first established in this paper. As a byproduct, we show in a very simple way that the SNR at the output of the PMWF is larger than or equal to the input SNR. Earlier contributions in this regard include [5], where Chen *et al.* demonstrated the effectiveness of the time-domain single-channel Wiener filter to increase the SNR at its output. In [6], Doclo and Moonen, considered the so-called speech distortion weighted multichannel Wiener filter and provided a quite involved (as compared to the one provided here) proof of the output SNR improvement. In [1], a magnitude-squared-coherence-based proof was given to show that the PMWF *locally* improves the output SNR.

This paper is organized as follows. Section II describes the data model, assumptions, and definitions. Section III reviews the optimization framework leading to the PMWF. Section IV investigates the performance of the PMWF in terms of the global output SNR. Finally, some concluding remarks are given in Section V.

II. Problem Statement

Let $s(t)$ denote a speech signal impinging on an array of N microphones with an arbitrary geometry. The resulting observations are given by

$$\begin{aligned} y_n(t) &= g_n(t) * s(t) + v_n(t) \\ &= x_n(t) + v_n(t), \quad n = 1, 2, \ldots, N \end{aligned} \tag{1}$$

where $*$ is the convolution operator, $g_n(t)$ is the channel impulse response encountered by the source before impinging on the nth microphone, $x_n(t) = g_n(t) * s(t)$ is the noise-free reverberant speech component, and $v_n(t)$ is the noise at microphone n. The source and noise are zero-mean random processes and are assumed to be uncorrelated. The above data model is written in the frequency domain as

$$\begin{aligned} Y_n(\omega) &= G_n(\omega)S(\omega) + V_n(\omega) \\ &= X_n(\omega) + V_n(\omega), \quad n = 1, 2, \ldots, N \end{aligned} \tag{2}$$

where $Y_n(\omega)$, $G_n(\omega)$, $S(\omega)$, and $V_n(\omega)$ are discrete-time Fourier transforms (DTFT's) of $y_n(t)$, $g_n(t)$, $s(t)$, and $v_n(t)$, respectively, and ω is the angular frequency.

Manuscript received November 26, 2009; revised January 26, 2010. First published February 25, 2010; current version published March 12, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Rudolf Rabenstein.

The authors are with the Université du Québec, INRS-EMT, Montréal, QC H5A 1K6 Canada (e-mail: souden@emt.inrs.ca; benesty@emt.inrs.ca, affes@emt.inrs.ca).

Digital Object Identifier 10.1109/LSP.2010.2042520

Our aim is to recover one of the signal components, say $X_{n_0}(\omega)$, $n_0 \in \{1,\ldots,N\}$ is the index of the reference microphone, the best way we can (along some criteria to be defined later) by applying a linear filter $\mathbf{h}_{n_0}(\omega)$ to the vector $\mathbf{y}(\omega) = [\,Y_1(\omega) \quad Y_2(\omega) \quad \ldots \quad Y_N(\omega)\,]^T$ whose entries are the microphone observations in the frequency domain and where $(\cdot)^T$ denotes the transpose of a matrix or a vector. The output of this filter is given by

$$\begin{aligned} Z_{n_0}(\omega) &\triangleq \mathbf{h}_{n_0}^H(\omega)\mathbf{y}(\omega) \\ &= \mathbf{h}_{n_0}^H(\omega)\mathbf{x}(\omega) + \mathbf{h}_{n_0}^H(\omega)\mathbf{v}(\omega) \end{aligned} \tag{3}$$

where $\mathbf{x}(\omega) = [\,X_1(\omega) \quad X_2(\omega) \quad \ldots \quad X_N(\omega)\,]^T$, $\mathbf{v}(\omega) = [\,V_1(\omega) \quad V_2(\omega) \quad \ldots \quad V_N(\omega)\,]^T$, $\mathbf{h}_{n_0}^H(\omega)\mathbf{x}(\omega)$ is the filtered speech, $\mathbf{h}_{n_0}^H(\omega)\mathbf{v}(\omega)$ is the residual noise, and $(\cdot)^H$ denotes the transpose-conjugate operator. To proceed, we also need to define the so-called power spectrum density (PSD) matrix for a given vector $\mathbf{a}(\omega)$ as

$$\mathbf{\Phi}_{aa}(\omega) \triangleq E\left\{\mathbf{a}(\omega)\mathbf{a}^H(\omega)\right\}. \tag{4}$$

In practice, we assume that the noise is stationary enough [2]. Therefore, $\mathbf{\Phi}_{vv}(\omega)$ can be estimated during the periods of silence of the desired speech and used during its periods of activity. Subsequently, we use the fact that the speech and noise are uncorrelated to calculate $\mathbf{\Phi}_{xx}(\omega) = \mathbf{\Phi}_{yy}(\omega) - \mathbf{\Phi}_{vv}(\omega)$.

Since we are taking the n_0th noise-free microphone signal as a reference, we define the local input SNR (at frequency ω) as [4]

$$\mathrm{SNR}(\omega) \triangleq \frac{\phi_{x_{n_0}x_{n_0}}(\omega)}{\phi_{v_{n_0}v_{n_0}}(\omega)} \tag{5}$$

where $\phi_{aa}(\omega) \triangleq E\left\{|A(\omega)|^2\right\}$ is the PSD of $a(t)$ [having $A(\omega)$ as DTFT]. The global input SNR is defined as [4]

$$\mathrm{SNR} \triangleq \frac{\int_{-\pi}^{\pi}\phi_{x_{n_0}x_{n_0}}(\omega)d\omega}{\int_{-\pi}^{\pi}\phi_{v_{n_0}v_{n_0}}(\omega)d\omega}. \tag{6}$$

Again, our aim is to have an optimal (in some sense that will be specified later) estimate of $X_{n_0}(\omega)$ at every frequency ω at the output of the linear filter $\mathbf{h}_{n_0}(\omega)$. Hence, we define the error signals [1], [3], [4]

$$\mathcal{E}_{x,n_0}(\omega) \triangleq [\mathbf{u}_{n_0} - \mathbf{h}_{n_0}(\omega)]^H\mathbf{x}(\omega), \tag{7}$$

$$\mathcal{E}_{v,n_0}(\omega) \triangleq \mathbf{h}_{n_0}^H(\omega)\mathbf{v}(\omega) \tag{8}$$

where $\mathbf{u}_{n_0} = \begin{bmatrix} 0 & \ldots & 0 & \underbrace{1}_{n_0\mathrm{th}} & 0 & \ldots & 0 \end{bmatrix}^T$ is an N-dimensional vector. Note that $\mathcal{E}_{x,n_0}(\omega)$ and $\mathcal{E}_{v,n_0}(\omega)$ are the estimation error signals that capture the speech distortion and residual noise at the output of $\mathbf{h}_{n_0}(\omega)$, respectively.

In this paper, we are interested in the analysis of the output SNR for a given filter $\mathbf{h}_{n_0}(\omega)$. The accurate quantification of this objective metric is important to have a clear idea of how much a filter of interest attenuates the noise relative to the output desired signal energy. It has also to be pointed out that in the investigated scenario, this metric can be defined in two different ways. First, we define the *local* SNR at the output of $\mathbf{h}_{n_0}(\omega)$ as [4]

$$\begin{aligned} \mathrm{SNR}_o[\mathbf{h}_{n_0}(\omega)] &\triangleq \frac{E\left\{\left|\mathbf{h}_{n_0}^H(\omega)\mathbf{x}(\omega)\right|^2\right\}}{E\left\{\left|\mathbf{h}_{n_0}^H(\omega)\mathbf{v}(\omega)\right|^2\right\}} \\ &= \frac{\mathbf{h}_{n_0}^H(\omega)\mathbf{\Phi}_{xx}(\omega)\mathbf{h}_{n_0}(\omega)}{\mathbf{h}_{n_0}^H(\omega)\mathbf{\Phi}_{vv}(\omega)\mathbf{h}_{n_0}(\omega)}. \end{aligned} \tag{9}$$

This metric quantifies the SNR variations *per frequency bin* that might give an idea about the output speech quality. However, after noise reduction, a synthesis step is commonly performed to reconstruct the (time-domain) estimate of the desired speech signal and it is more judicious to consider the *global* (or full-band) SNR after all the processing stage. This global output SNR is defined as

$$\mathrm{SNR}_o \triangleq \frac{\int_{-\pi}^{\pi}\mathbf{h}_{n_0}^H(\omega)\mathbf{\Phi}_{xx}(\omega)\mathbf{h}_{n_0}(\omega)d\omega}{\int_{-\pi}^{\pi}\mathbf{h}_{n_0}^H(\omega)\mathbf{\Phi}_{vv}(\omega)\mathbf{h}_{n_0}(\omega)d\omega}. \tag{10}$$

The analysis of this metric is relevant since it characterizes the relative variations of the output noise and speech energies. Consequently, it was considered in several previous contributions including [1], [4]–[6]. Another definition of the output SNR can be found in [7] and is, essentially, equal to the ratio of the desired clean signal energy divided by the energy of the estimation error signal $\mathcal{E}_{x,n_0}(\omega) - \mathcal{E}_{v,n_0}(\omega)$. The analysis of the latter falls beyond the scope of this work and we would rather focus on the definition in (10).

Generally, noise reduction comes at the price of speech distortion. In [5], [7], it was shown that in the single-channel case, any noise reduction leads to speech distortion. However, when multiple microphones are deployed for speech acquisition, noise reduction can be achieved with no speech distortion [3], [4]. Alternatively, one may also relax the constraint on the output speech distortion [1], [2] to further reduce the noise. In either case, it is of paramount importance to accurately quantify the true gains and losses in terms of noise reduction and speech distortion. The global output SNR is a good metric to quantify these variations. Therefore, we investigate it in this work.

III. Parameterized Frequency-Domain Multichannel Wiener Filter

The frequency-domain optimization problem that we consider in this work consists in minimizing the residual noise while constraining the output speech distortion. Mathematically, this translates into the following

$$\begin{aligned} &\underset{\mathbf{h}_{n_0}(\omega)}{\mathrm{minimize}} \quad E\left\{|\mathcal{E}_{v,n_0}(\omega)|^2\right\} \\ &\text{subject to} \quad E\left\{|\mathcal{E}_{x,n_0}(\omega)|^2\right\} \le \sigma^2 \end{aligned} \tag{11}$$

where σ^2 represents the maximum allowable local signal distortion, and $\mathcal{E}_{x,n_0}(\omega)$ and $\mathcal{E}_{v,n_0}(\omega)$ are defined in (7) and (8), respectively. The solution to this optimization problem is [1]

$$\begin{aligned} \mathbf{h}_{W\beta,n_0}(\omega) &= \frac{\mathbf{\Phi}_{vv}^{-1}(\omega)\mathbf{\Phi}_{xx}(\omega)}{\beta+\lambda(\omega)}\mathbf{u}_{n_0} \\ &= \frac{\mathbf{\Phi}_{vv}^{-1}(\omega)\mathbf{\Phi}_{yy}(\omega) - \mathbf{I}_{N\times N}}{\beta+\lambda(\omega)}\mathbf{u}_{n_0} \end{aligned} \tag{12}$$

where $\mathbf{I}_{N\times N}$ is the identity matrix of size $N \times N$ and

$$\lambda(\omega) = \mathrm{tr}\left\{\mathbf{\Phi}_{vv}^{-1}(\omega)\mathbf{\Phi}_{xx}(\omega)\right\} = \mathrm{tr}\left\{\mathbf{\Phi}_{vv}^{-1}(\omega)\mathbf{\Phi}_{yy}(\omega)\right\} - N.$$

$\mathrm{tr}\{\cdot\}$ denotes the trace of a square matrix. Note that $\lambda(\omega)$ is real and positive. $\mathbf{h}_{W\beta,n_0}(\omega)$ is the PMWF whose performance is analyzed in what follows. β is the tuning parameter which is real and positive. It is equal to the inverse of the Lagrange multiplier associated with the optimization problem (11). In [1], we have also established that β is related to σ as

$$\beta \leq \frac{\tilde{\sigma}(\omega)}{1-\tilde{\sigma}(\omega)}\lambda(\omega),$$

where $\tilde{\sigma}(\omega) = \sigma\phi_{x_{n_0}x_{n_0}}^{-1/2}(\omega)$. From this optimization framework, it can be seen that the MVDR beamformer is a particular case of $\mathbf{h}_{W\beta,n_0}(\omega)$ that corresponds to the parameter $\sigma = 0$ (equivalently, $\beta = 0$). Increasing β leads to more noise reduction at the price of an increased signal distortion as it has been shown using the local performance measures (speech distortion index and noise reduction factor) in [1]. However, no previous study has shown the effect of this tuning parameter on the global output SNR that effectively quantifies the achieved noise reduction relative to the output speech energy.

IV. Study of the Global Output SNR

A. *Effect of the Tuning Parameter on the Global Output SNR*

Using the definition in (9) and the fact that the residual noise and filtered speech energies are given by [1]

$$\mathbf{h}_{W\beta,n_0}^{H}(\omega)\mathbf{\Phi}_{vv}(\omega)\mathbf{h}_{W\beta,n_0}(\omega) = \frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda(\omega)}{[\beta+\lambda(\omega)]^2} \quad (13)$$

and

$$\mathbf{h}_{W\beta,n_0}^{H}(\omega)\mathbf{\Phi}_{xx}(\omega)\mathbf{h}_{W\beta,n_0}(\omega) = \frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda^2(\omega)}{[\beta+\lambda(\omega)]^2} \quad (14)$$

respectively, we establish that the local output SNR of the PMWF is

$$\mathrm{SNR}_o[\mathbf{h}_{W\beta,n_0}(\omega)] = \lambda(\omega). \quad (15)$$

It is clear that the local output SNR is independent of β. Indeed, this parameter only appears in the scaling factor of $\mathbf{h}_{W\beta,n_0}(\omega)$ whose effect vanishes in the definition (9). It has to be pointed out, however, that this scaling factor is frequency dependent and it is known that β markedly affects the global performance of the parameterized Wiener filter in terms of the true levels of residual noise and filtered speech. To solve this dilemma, we have to consider the global output SNR[1]. We propose the following.

Proposition 1: The global SNR at the output of the PMWF defined in (12) is an increasing function of the tuning parameter β.

Proof: Inserting (13) and (14) into (10), we obtain

$$\mathrm{SNR}_o(\beta) = \frac{\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda^2(\omega)}{[\beta+\lambda(\omega)]^2}d\omega}{\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda(\omega)}{[\beta+\lambda(\omega)]^2}d\omega}. \quad (16)$$

[1]In what follows, the global output SNR of the PMWF will be denoted as $\mathrm{SNR}_o(\beta)$ to explicit its dependence on β.

In order to determine the variations of $\mathrm{SNR}_o(\beta)$ with respect to the parameter β, we check the sign of the following differentiation with respect to β (see (17) at the bottom of the page). We only focus on the numerator of the above derivative to see the variations of the the global output SNR since the denominator is always positive. Multiplying and dividing by $[\beta+\lambda(\omega)]$, this numerator can be rewritten (up to the scaling factor 2 that has no effect on the sign) as

$$\begin{aligned}
Num(\beta) =& -\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda(\omega)[\beta+\lambda(\omega)]}{[\beta+\lambda(\omega)]^3}d\omega \\
&\times\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda^2(\omega)}{[\beta+\lambda(\omega)]^3}d\omega \\
&+\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda^2(\omega)[\beta+\lambda(\omega)]}{[\beta+\lambda(\omega)]^3}d\omega \\
&\times\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda(\omega)}{[\beta+\lambda(\omega)]^3}d\omega \\
=& -\left[\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda^2(\omega)}{[\beta+\lambda(\omega)]^3}d\omega\right]^2 \\
&-\beta\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda(\omega)}{[\beta+\lambda(\omega)]^3}d\omega \\
&\times\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda^2(\omega)}{[\beta+\lambda(\omega)]^3}d\omega \\
&+\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda^3(\omega)}{[\beta+\lambda(\omega)]^3}d\omega \\
&\times\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda(\omega)}{[\beta+\lambda(\omega)]^3}d\omega \\
&+\beta\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda(\omega)}{[\beta+\lambda(\omega)]^3}d\omega \\
&\times\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda^2(\omega)}{[\beta+\lambda(\omega)]^3}d\omega \\
=& -\left[\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda^2(\omega)}{[\beta+\lambda(\omega)]^3}d\omega\right]^2 \\
&+\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda^3(\omega)}{[\beta+\lambda(\omega)]^3}d\omega \\
&\times\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda(\omega)}{[\beta+\lambda(\omega)]^3}d\omega. \quad (18)
\end{aligned}$$

As far as β, $\lambda(\omega)$, and $\phi_{x_{n_0}x_{n_0}}(\omega)$ are positive $\forall\,\omega$, we can use the Cauchy–Schwarz inequality

$$\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda^3(\omega)}{[\beta+\lambda(\omega)]^3}d\omega\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda(\omega)}{[\beta+\lambda(\omega)]^3}d\omega$$

$$\frac{d\mathrm{SNR}_o(\beta)}{d\beta} = \frac{-2\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda(\omega)}{[\beta+\lambda(\omega)]^2}d\omega\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda^2(\omega)}{[\beta+\lambda(\omega)]^3}d\omega + 2\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda^2(\omega)}{[\beta+\lambda(\omega)]^2}d\omega\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda(\omega)}{[\beta+\lambda(\omega)]^3}d\omega}{\left[\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda(\omega)}{[\beta+\lambda(\omega)]^2}d\omega\right]^2} \quad (17)$$

$$\geq \left[\int_{-\pi}^{\pi}\sqrt{\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda^3(\omega)}{[\beta+\lambda(\omega)]^3}}\sqrt{\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda(\omega)}{[\beta+\lambda(\omega)]^3}}d\omega\right]^2$$
$$= \left[\int_{-\pi}^{\pi}\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda^2(\omega)}{[\beta+\lambda(\omega)]^3}d\omega\right]^2. \quad (19)$$

Inserting (19) into (18), we conclude that

$$\frac{d\mathrm{SNR}_o(\beta)}{d\beta} \geq 0 \quad (20)$$

thereby proving that the global output SNR is increasing with respect to β. □

Note also that the Cauchy–Schwarz inequality in (19) turns into equality if and only if there exists a contant α such that

$$\sqrt{\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda^3(\omega)}{[\beta+\lambda(\omega)]^3}} = \alpha\sqrt{\frac{\phi_{x_{n_0}x_{n_0}}(\omega)\lambda(\omega)}{[\beta+\lambda(\omega)]^3}}, \quad \forall\,\omega\in(-\pi,\pi]$$

i.e., $\lambda(\omega) = \alpha, \forall\,\omega \in (-\pi,\pi]$. This means that when $\lambda(\omega)$ is constant, increasing the factor β does not increase the output SNR. Nevertheless, in the general case of speech corrupted with independently distributed noise, $\lambda(\omega)$ is not constant and the global output SNR is strictly increasing with respect to β. As a straightforward application, for example, we conclude that generally for speech signals, the MVDR provides *globally* a lower output SNR than the multichannel Wiener filter.

Now, since the global output SNR is an increasing function of β, it is quite natural to ask whether it can be infinitely increased. First, we have to mention that in the particular case of a perfectly spatially coherent noise field, the local output SNR is infinite [1], i.e., $\lambda(\omega)=\infty$, thereby meaning that a perfect recovery of the target speech signal by means of the PMWF is possible in this case regardless of the tuning parameter. However, this assumption is generally violated since other types of noise (e.g, spatially white and diffuse noise) are ubiquitous in acoustic rooms. Therefore, we consider the case where $\lambda(\omega) < \infty\;\forall\,\omega$, which is more plausible in practice. Consequently, we can easily see from (16) that the output SNR cannot be infinitely improved by increasing β and its upper bound is nothing but the following asymptotic value

$$\lim_{\beta\to\infty}\mathrm{SNR}_o(\beta) = \frac{\int_{-\pi}^{\pi}\phi_{x_{n_0}x_{n_0}}(\omega)\lambda^2(\omega)d\omega}{\int_{-\pi}^{\pi}\phi_{x_{n_0}x_{n_0}}(\omega)\lambda(\omega)d\omega}. \quad (21)$$

B. Simplified Proof of the Increase of SNR at the Output of the PMWF

By virtue of Proposition 1, we are able to propose the following result.

Proposition 2: The parameterized multichannel Wiener filter improves the global SNR at its output. In other words,

$$\mathrm{SNR}_o(\beta) \geq \mathrm{SNR},\;\forall\,\beta\geq 0. \quad (22)$$

Proof: We first recall that we demonstrated in [1] that the local output SNR satisfies

$$\mathrm{SNR}_o[\mathbf{h}_{W\,\beta,n_0}(\omega)] \geq \mathrm{SNR}(\omega). \quad (23)$$

Also, we have

$$\mathrm{SNR}_o(\beta) = \frac{\int_{-\pi}^{\pi}\phi_{x_{n_0}x_{n_0}}(\omega)\frac{\lambda^2(\omega)}{[\beta+\lambda(\omega)]^2}d\omega}{\int_{-\pi}^{\pi}\phi_{x_{n_0}x_{n_0}}(\omega)\frac{\lambda(\omega)}{[\beta+\lambda(\omega)]^2}d\omega} = \frac{\int_{-\pi}^{\pi}\phi_{x_{n_0}x_{n_0}}(\omega)\frac{\lambda^2(\omega)}{[\beta+\lambda(\omega)]^2}d\omega}{\int_{-\pi}^{\pi}\phi_{v_{n_0}v_{n_0}}(\omega)\frac{\mathrm{SNR}(\omega)\lambda(\omega)}{[\beta+\lambda(\omega)]^2}d\omega}. \quad (24)$$

Using Proposition 1, we have

$$\mathrm{SNR}_o(\beta) \geq \mathrm{SNR}_o(0), \quad \forall\,\beta\geq 0$$

and

$$\mathrm{SNR}_o(0) = \frac{\int_{-\pi}^{\pi}\phi_{x_{n_0}x_{n_0}}(\omega)d\omega}{\int_{-\pi}^{\pi}\phi_{v_{n_0}v_{n_0}}(\omega)\frac{\mathrm{SNR}(\omega)}{\lambda(\omega)}d\omega}. \quad (25)$$

Using (23) and (25), we obtain

$$\mathrm{SNR}_o(\beta) \geq \mathrm{SNR}_o(0) \geq \frac{\int_{-\pi}^{\pi}\phi_{x_{n_0}x_{n_0}}(\omega)d\omega}{\int_{-\pi}^{\pi}\phi_{v_{n_0}v_{n_0}}(\omega)d\omega} = \mathrm{SNR}, \quad \forall\,\beta\geq 0. \quad (26)$$

This completes the proof. □

A similar proposition was previously made in [5] in the case of time-domain single-channel Wiener filter where the generalized eigenvalue decomposition of the speech and noise correlation matrices was used jointly with an inductive reasoning. In [6], Doclo and Moonen provided another proof in the time-domain multichannel case where the so-called speech distortion weighted Wiener filter is used. Both proofs are relatively complicated as compared to the straightforward one that we propose here.

V. Conclusion

In this letter, we considered the parameterized multichannel Wiener filter that is commonly used for noise reduction in speech communication applications. We have established that the global SNR at its output is an increasing function of the tuning parameter (that simultaneously increases the signal distortion and noise reduction). An asymptotic finite value of this performance metric is given. Finally, a very simplified proof, compared to earlier works, namely, [5], [6], showing that the global output SNR is larger than the input SNR is outlined.

References

[1] M. Souden, J. Benesty, and S. Affes, "On optimal frequency-domain multichannel linear filtering for noise reduction," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, pp. 260–276, Feb. 2010.

[2] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Process.*, vol. 50, pp. 2230–2244, Sep. 2002.

[3] S. Gannot, D. Burstein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Process.*, vol. 49, pp. 1614–1626, Aug. 2001.

[4] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.

[5] J. Chen, J. Benesty, and Y. Huang, "New insights into the noise reduction Wiener filter," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, pp. 1218–1234, Jul. 2006.

[6] S. Doclo and M. Moonen, "On the output SNR of the speech-distortion weighted multichannel Wiener filter," *IEEE Signal Process. Lett.*, vol. 12, pp. 809–811, Dec. 2005.

[7] P. C. Loizou, *Speech Enhancement: Theory and Practice*. New York: CRC, 2007.