# A STUDY OF MULTICHANNEL NOISE REDUCTION LINEAR FILTERS IN THE TIME DOMAIN

*Jacob Benesty*[1]*, Mehrez Souden*[2]*, Jingdong Chen*[3]

[1]INRS-EMT, University of Quebec, Montreal, QC, Canada
[2]Communication Science Laboratories, NTT Corp., Kyoto, Japan
[3]Northwestern Polytechnical University, Xi'an, China

## ABSTRACT

We propose a new study of multichannel time-domain noise reduction filters where we decompose the *noise-free* microphone array observations into two components: the first is correlated with the target signal and perfectly coherent across the sensors while the second consists of residual interference. Then, alternative formulations of well known time-domain filters, namely, the minimum variance distortionless response (MVDR), the linearly constrained minimum variance (LCMV), the multichannel tradeoff, and Wiener filters are given. Besides, the performance of these time-domain filters are analytically investigated and new insights are presented. Numerical results are finally provided to corroborate our study.

***Index Terms—*** Noise reduction, speech enhancement, microphone array, Wiener filter, beamforming.

## 1. INTRODUCTION

Multichannel processing for noise reduction has recently become a very active field of research due to its great potential as compared to its single-channel counterpart and to the challenging propagation conditions, namely, the acoustic noise and reverberation characteristics, that can dramatically change from one application to another [1]. Prior to developing a robust multichannel processing scheme, which is the ultimate goal of all research efforts in this field, it is crucial to comprehend the fundamental functioning and achievable performance of widely utilized multichannel noise reduction approaches. This paper is devoted to the analysis of *time-domain* multichannel noise reduction techniques, namely the MVDR, LCMV, Wiener, and tradeoff filters.

In the traditional formulation of multichannel noise reduction, the objective is to recover a noise-free reference microphone signal by taking advantage of the spatio-temporal properties of the observed mixtures of sounds. This objective can be achieved in either time [2, 3] or some transform domains [4, 5, 6]. Good analysis and comparisons of noise reduction using different transforms including Fourier, Karhunen-Loève, cosine, and Hadamard can be found in [6]. Traditionally, multichannel noise reduction techniques are formulated by splitting the processed microphone observations into two terms: filtered noise-free speech and residual additive noise. The first term is treated as desired signal while the second is a nuisance. Then, the objective has typically been to reduce the nuisance while keeping the filtered speech as similar as possible to a reference signal. It turns out that this treatment of the overall filtered speech as the desired signal is inappropriate as it will become clear soon.

In this paper, we propose a new perspective on multichannel time-domain noise reduction. We demonstrate that it is possible to decompose the noise-free observations into two uncorrelated components: a desired signal which is fully coherent across the sensors and some additive interference. Subsequently, we derive many well-known filters,namely, the MVDR, LCMV, tradeoff, and Wiener filters and show how the time-domain MVDR, Wiener, and tradeoff filters are identical up to a scaling factor. Also we demonstrate an interesting connection between the LCMV and MVDR filter. It is worthwhile noting that the most widely used multichannel linear filters, in particular those investigated in this paper, are obtained by optimizing second-order-statistics-based criteria. Hence, using second-order-statistics-based metrics to evaluate their performance is the most natural and intuitive way to comprehend their functioning. In particular, the noise reduction factor, signal distortion index and output SNR [2, 7] are used for performance evaluation here. Using these metrics, we carry out a simplified yet rigorous analysis of the performance of the MVDR, LCMV, tradeoff, and Wiener filters.

## 2. SIGNAL MODEL

We consider the typical formulation of signal model in which an $N$-element microphone array captures a convolved source signal in some noise field. The received signals, at the discrete-time index $k$, are expressed as [1, 4]

$$y_n(k) = x_n(k) + v_n(k),\ n = 1, 2, \ldots, N, \quad (1)$$

where $x_n(k) = g_n(k) * s(k)$, $g_n(k)$ is the impulse response from the unknown speech source $s(k)$ location to the $n$th

microphone, $*$ stands for linear convolution, and $v_n(k)$ is the additive noise at microphone $n$. We assume that the signals $x_n(k)$ and $v_n(k)$ are uncorrelated and zero mean. The noise signals $v_n(k)$ are typically either partially coherent or incoherent across the array. All previous signals are assumed to be real, broadband, Gaussian, and stationary.

By processing the data by blocks of $L$ samples and considering all microphones, the signal model given in (1) can be put into a vector form as

$$\begin{aligned}\mathbf{y}(k) &= \left[\begin{array}{cccc}\mathbf{y}_1^T(k) & \mathbf{y}_2^T(k) & \cdots & \mathbf{y}_N^T(k)\end{array}\right]^T \\ &= \mathbf{x}(k)+\mathbf{v}(k), \end{aligned} \quad (2)$$

where $\mathbf{y}_n(k) = [y_n(k)\ y_n(k-1)\ ...\ y_n(k-L+1)]^T$. Vectors $\mathbf{x}(k)$ and $\mathbf{v}(k)$ of length $NL$ are defined in a similar way to $\mathbf{y}(k)$. Since $x_n(k)$ and $v_n(k)$ are uncorrelated by assumption, the correlation matrix (of size $NL \times NL$) of the microphone signals is

$$\mathbf{R}_{\mathbf{y}} = E\left[\mathbf{y}(k)\mathbf{y}^T(k)\right] = \mathbf{R}_{\mathbf{x}} + \mathbf{R}_{\mathbf{v}}, \quad (3)$$

where $E[\cdot]$ denotes mathematical expectation, and $\mathbf{R}_{\mathbf{x}} = E\left[\mathbf{x}(k)\mathbf{x}^T(k)\right]$ and $\mathbf{R}_{\mathbf{v}} = E\left[\mathbf{v}(k)\mathbf{v}^T(k)\right]$ are the correlation matrices of $\mathbf{x}(k)$ and $\mathbf{v}(k)$, respectively. With the above signal model, the objective of noise reduction is to estimate any one of the signals $x_n(k)$ [2, 3, 4, 5] say $x_1(k)$, without loss of generality.

## 3. LINEAR DECOMPOSITION OF THE MICROPHONE SIGNALS

Our objective is to recover an estimate of the desired signal as $\hat{x}_1(k) = \mathbf{h}^T\mathbf{y}(k)$ where $\mathbf{h}$ is an $NL$–dimensional linear filter. We decompose the estimated signal as

$$\hat{x}_1(k) = \mathbf{h}^T\left[\mathbf{x}(k)+\mathbf{v}(k)\right] = x_{\mathrm{f}}(k) + v_{\mathrm{rn}}(k), \quad (4)$$

where $x_{\mathrm{f}}(k) = \mathbf{h}^T\mathbf{x}(k)$ is the filtered speech signal and $v_{\mathrm{rn}}(k) = \mathbf{h}^T\mathbf{v}(k)$ is the residual noise. From (4), we see that $\hat{x}_1(k)$ depends on the vector $\mathbf{x}(k)$. However, our desired signal at time $k$ is only $x_1(k)$ [and not the whole vector $\mathbf{x}(k)$]. Actually, the vector $\mathbf{x}(k)$ may contain a component uncorrelated with $x_1(k)$, which we consider as an undesired component. Specifically,

$$\mathbf{x}(k) = \mathbf{x}_{\mathrm{d}}(k) + \mathbf{x}'(k), \quad (5)$$

where

$$\mathbf{x}_{\mathrm{d}}(k) = x_1(k)\boldsymbol{\gamma}_{\mathbf{x}} \quad (6)$$

is the desired signal vector (of length $NL$), $\mathbf{x}'(k)$ is orthogonal to $\mathbf{x}_{\mathrm{d}}(k)$,

$$\boldsymbol{\gamma}_{\mathbf{x}} = \left[\begin{array}{cccc}\boldsymbol{\gamma}_{\mathbf{x}_1}^T & \boldsymbol{\gamma}_{\mathbf{x}_2}^T & \cdots & \boldsymbol{\gamma}_{\mathbf{x}_N}^T\end{array}\right]^T$$

is the normalized [with respect to $x_1(k)$] cross-correlation vector (of length $NL$) between $x_1(k)$ and $\mathbf{x}(k)$,

$$\begin{aligned}\boldsymbol{\gamma}_{\mathbf{x}_n} &= \left[\begin{array}{ccc}\gamma_{\mathbf{x}_n,0} & \gamma_{\mathbf{x}_n,1} & \gamma_{\mathbf{x}_n,L-1}\end{array}\right]^T \\ &= \frac{E\left[x_1(k)\mathbf{x}_n(k)\right]}{E\left[x_1^2(k)\right]},\ n=1,2,\ldots,N \end{aligned} \quad (7)$$

is the cross-correlation vector (of length $L$) between $x_1(k)$ and $\mathbf{x}_n(k)$,

$$\gamma_{\mathbf{x}_n,l} = \frac{E\left[x_1(k)x_n(k-l)\right]}{E\left[x_1^2(k)\right]},\ l=0,1,\ldots,L-1 \quad (8)$$

is the cross-correlation coefficient between $x_1(k)$ and $x_n(k-l)$, and

$$\mathbf{x}'(k) = \mathbf{x}(k) - x_1(k)\boldsymbol{\gamma}_{\mathbf{x}}, \quad (9)$$

$$E\left[x_1(k)\mathbf{x}'(k)\right] = \mathbf{0}. \quad (10)$$

Substituting (5) into (4), we get

$$\begin{aligned}\hat{x}_1(k) &= \mathbf{h}^T\left[x_1(k)\boldsymbol{\gamma}_{\mathbf{x}} + \mathbf{x}'(k) + \mathbf{v}(k)\right], \\ &= x_{\mathrm{fd}}(k) + x'_{\mathrm{ri}}(k) + v_{\mathrm{rn}}(k), \end{aligned} \quad (11)$$

where $x_{\mathrm{fd}}(k) = x_1(k)\mathbf{h}^T\boldsymbol{\gamma}_{\mathbf{x}}$ is the filtered desired signal and $x'_{\mathrm{ri}}(k) = \mathbf{h}^T\mathbf{x}'(k)$ is the residual interference. We observe that the estimate of the desired signal at time $k$ is the sum of three terms: the first one is clearly the filtered desired signal while the two others are the filtered undesired signals (interference-plus-noise). Since the three terms are mutually uncorrelated, the variance of $\hat{x}_1(k)$ is $\sigma^2_{\hat{x}_1} = \sigma^2_{x_{\mathrm{fd}}} + \sigma^2_{x'_{\mathrm{ri}}} + \sigma^2_{v_{\mathrm{rn}}}$, where $\sigma^2_{x_{\mathrm{fd}}} = \sigma^2_{x_1}\left(\mathbf{h}^T\boldsymbol{\gamma}_{\mathbf{x}}\right)^2 = \mathbf{h}^T\mathbf{R}_{\mathbf{x}_{\mathrm{d}}}\mathbf{h}$, $\sigma^2_{x'_{\mathrm{ri}}} = \mathbf{h}^T\mathbf{R}_{\mathbf{x}'}\mathbf{h} = \mathbf{h}^T\mathbf{R}_{\mathbf{x}}\mathbf{h} - \sigma^2_{x_1}\left(\mathbf{h}^T\boldsymbol{\gamma}_{\mathbf{x}}\right)^2$, $\sigma^2_{v_{\mathrm{rn}}} = \mathbf{h}^T\mathbf{R}_{\mathbf{v}}\mathbf{h}$, $\sigma^2_{x_1} = E\left[x_1^2(k)\right]$ is the variance of the desired signal, $\mathbf{R}_{\mathbf{x}_{\mathrm{d}}} = \sigma^2_{x_1}\boldsymbol{\gamma}_{\mathbf{x}}\boldsymbol{\gamma}_{\mathbf{x}}^T$ is the correlation matrix (whose rank is equal to 1) of $\mathbf{x}_{\mathrm{d}}(k)$, and $\mathbf{R}_{\mathbf{x}'} = E\left[\mathbf{x}'(k)\mathbf{x}'^T(k)\right]$ is the correlation matrix of $\mathbf{x}'(k)$. Comparing the decomposition of the filtered microphone observations in (11), and recalling its narrowband counterpart in the frequency domain [4, 8] we clearly see that $\boldsymbol{\gamma}_{\mathbf{x}}$ plays the role of the steering vector of the desired source in the time domain.

## 4. PERFORMANCE MEASURES

Since our objective is to estimate $x_1(k)$, the first microphone has to be taken as a reference when defining our performance measures. The proposed definitions slightly differ from traditional (old) definitions that can be found in previous references, e.g., [2, 7], in the sense that the decomposition of the noise-free speech observations outlined above is first accounted for here. The first important measure is the input SNR defined as iSNR $= \frac{\sigma^2_{x_1}}{\sigma^2_{v_1}}$, where $\sigma^2_{v_1} = E\left[v_1^2(k)\right]$. We also quantify the level of noise remaining at the output of the filter by defining the output

SNR as

$$\text{oSNR}\left(\mathbf{h}\right) = \frac{\sigma_{x_{\text{fd}}}^2}{\sigma_{x'_{\text{ri}}}^2 + \sigma_{v_{\text{rn}}}^2} = \frac{\sigma_{x_1}^2 \left(\mathbf{h}^T \boldsymbol{\gamma}_{\mathbf{x}}\right)^2}{\mathbf{h}^T \mathbf{R}_{\text{in}} \mathbf{h}}, \tag{12}$$

where $\mathbf{R}_{\text{in}} = \mathbf{R}_{\mathbf{x}'} + \mathbf{R}_{\mathbf{v}}$ is the interference plus noise covariance matrix. The objective of the noise reduction filter is to make the output SNR greater than the input SNR to enhance the quality of the noisy signal.

The noise-reduction factor [2, 7] quantifies the amount of noise rejected by the filter. This quantity is defined as the ratio of the variance of the noise at the reference microphone over the variance of the interference-plus-noise remaining at the output of the filter

$$\xi_{\text{nr}}\left(\mathbf{h}\right) = \frac{\sigma_{v_1}^2}{\sigma_{x'_{\text{ri}}}^2 + \sigma_{v_{\text{rn}}}^2} = \frac{\sigma_{v_1}^2}{\mathbf{h}^T \mathbf{R}_{\text{in}} \mathbf{h}}. \tag{13}$$

The noise-reduction factor is expected to be lower bounded by 1 for optimal filters.

In practice, the FIR filter, $\mathbf{h}$, may distort the desired signal. In order to evaluate the level of this distortion, we consider the speech-distortion index [2, 7] defined as

$$\upsilon_{\text{sd}}\left(\mathbf{h}\right) = \frac{E\left\{\left[x_{\text{fd}}(k) - x_1(k)\right]^2\right\}}{\sigma_{x_1}^2} = \left(\mathbf{h}^T \boldsymbol{\gamma}_{\mathbf{x}} - 1\right)^2. \tag{14}$$

The speech-distortion index is always greater than or equal to 0 and should be upper bounded by 1 for optimal filters; so the higher is $\upsilon_{\text{sd}}\left(\mathbf{h}\right)$, the more distorted is the desired signal.

## 5. OPTIMAL NOISE REDUCTION FILTERS

By using the new decomposition of the noise-free microphone signals in Section 3, we demonstrate that, even in the time domain (in analogy to the frequency domain formulation, for example), the maximum SNR, Wiener, MVDR, and tradeoff filters are equivalent up to a scaling factor. Furthermore, by decomposing the additive noise term in a similar way, we can develop an alternative expression of the LCMV and demonstrate its connection to the MVDR.

### 5.1. Wiener Filter

The Wiener filter is easily derived by taking the gradient of the mean square error (MSE), which is defined as [2, 3]

$$\varepsilon(\mathbf{h}) = E\left[\left(\mathbf{h}^T \mathbf{y}(k) - x_1(k)\right)^2\right], \tag{15}$$

with respect to $\mathbf{h}$ and equating the result to zero. This leads to

$$\mathbf{h}_{\text{W}} = \mathbf{R}_{\mathbf{y}}^{-1} \mathbf{R}_{\mathbf{x}_{\text{d}}} \mathbf{i}_1, \tag{16}$$

where $\mathbf{i}_1 = [1\ 0\ ...\ 0]$ is $(NL)$–dimensional vector. Thanks to the new linear decomposition of the noise-free microphone observations of the speech signal, we demonstrate another interesting expression of the Wiener filter

$$\mathbf{h}_{\text{W}} = \frac{\mathbf{R}_{\text{in}}^{-1} \boldsymbol{\gamma}_{\mathbf{x}}}{\sigma_{x_1}^{-2} + \boldsymbol{\gamma}_{\mathbf{x}}^T \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\gamma}_{\mathbf{x}}}. \tag{17}$$

Using (17), we deduce that the output SNR is $\text{oSNR}\left(\mathbf{h}_{\text{W}}\right) = \sigma_{x_1}^2 \boldsymbol{\gamma}_{\mathbf{x}}^T \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\gamma}_{\mathbf{x}}$. For notational convenience, we will denote $\text{oSNR}\left(\mathbf{h}_{\text{W}}\right) = \lambda_{\max}$. Actually, $\lambda_{\max}$ is the maximum eigenvalue of $\mathbf{R}_{\text{in}}^{-1} \mathbf{R}_{\mathbf{x}_{\text{d}}}$. Again, thanks to the new decomposition, it becomes possible (and easy) to show that $\mathbf{h}_{\text{W}}$, indeed, achieves the maximum output SNR. This result is first outlined here.

Now, the speech-distortion index is a clear function of the output SNR: $\upsilon_{\text{sd}}\left(\mathbf{h}_{\text{W}}\right) = \frac{1}{(1+\lambda_{\max})^2} \leq 1$. The higher is the value of the output SNR, the less distorted is the desired signal.

Clearly, $\text{oSNR}\left(\mathbf{h}_{\text{W}}\right) \geq \text{iSNR}$, since the Wiener filter maximizes the output SNR. Finally, with the Wiener filter the noise-reduction factor is $\xi_{\text{nr}}\left(\mathbf{h}_{\text{W}}\right) = \frac{(1+\lambda_{\max})^2}{\text{iSNR} \cdot \lambda_{\max}} \geq \left(1 + \frac{1}{\lambda_{\max}}\right)^2$, meaning that this filter always reduces the additive noise.

### 5.2. MVDR Filter

Another important filter, initially proposed by Capon [9] and delineated in several forms [2, 4, 5], is the minimum variance distortionless response (MVDR) beamformer which is obtained by minimizing the variance of the interference-plus-noise at the beamformer output with the constraint that the desired signal is not distorted, i.e., $\mathbf{h}^T \boldsymbol{\gamma}_{\mathbf{x}} = 1$. In our case, we can demonstrate that this filter is expressed as

$$\mathbf{h}_{\text{MVDR}} = \frac{\mathbf{R}_{\text{in}}^{-1} \boldsymbol{\gamma}_{\mathbf{x}}}{\boldsymbol{\gamma}_{\mathbf{x}}^T \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\gamma}_{\mathbf{x}}} = \frac{\mathbf{R}_{\text{in}}^{-1} \mathbf{R}_{\mathbf{x}_{\text{d}}}}{\lambda_{\max}} \mathbf{i}_1. \tag{18}$$

By comparing (17) and (18), it is easy to see that the time-domain Wiener and MVDR filters are related as follows

$$\mathbf{h}_{\text{W}} = \alpha_0 \mathbf{h}_{\text{MVDR}}, \tag{19}$$

where $\alpha_0 = \frac{\lambda_{\max}}{1+\lambda_{\max}}$.

Obviously, the MVDR does not distort the desired signal in theory. It outputs the same SNR value $\lambda_{\max}$ as the Wiener filter and its noise reduction factor is given by $\xi_{\text{nr}}\left(\mathbf{h}_{\text{MVDR}}\right) = \frac{\lambda_{\max}}{\text{iSNR}} \geq 1$.

### 5.3. Tradeoff Filter

The tradeoff filter generalizes the filters discussed above and that is achieved by scaling the noise-plus-interference covariance matrix in the expression of the Wiener filter (16). This scaling factor is denoted as $\mu\ (\geq 0)$ and allows to tune noise-plus-interference reduction factor versus speech distortion. Following the linear decomposition of the desired

signal in Section 3, we demonstrate that the tradeoff filter is given by

$$\begin{aligned}\mathbf{h}_{\mathrm{T},\mu} &= \sigma_{x_1}^2 \left[\sigma_{x_1}^2 \boldsymbol{\gamma}_{\mathbf{x}} \boldsymbol{\gamma}_{\mathbf{x}}^T + \mu \mathbf{R}_{\mathrm{in}}\right]^{-1} \boldsymbol{\gamma}_{\mathbf{x}} \\ &= \frac{\mathbf{R}_{\mathrm{in}}^{-1} \boldsymbol{\gamma}_{\mathbf{x}}}{\mu \sigma_{x_1}^{-2} + \boldsymbol{\gamma}_{\mathbf{x}}^T \mathbf{R}_{\mathrm{in}}^{-1} \boldsymbol{\gamma}_{\mathbf{x}}}. \end{aligned} \quad (20)$$

In order to obtain a certain value $\beta\sigma_{v_1}^2$ of the residual noise-plus-interference reduction energy ($0 \leq \beta \leq 1$), we can demonstrate that the tuning parameter $\mu$ has to be chosen such that

$$\beta = \frac{\mathrm{iSNR}\ \lambda_{\max}}{\left(\mu + \lambda_{\max}\right)^2}. \quad (21)$$

In particular, we can see that for

- $\mu = 1$, $\mathbf{h}_{\mathrm{T},1} = \mathbf{h}_{\mathrm{W}}$, which is the Wiener filter,
- $\mu = 0$, $\mathbf{h}_{\mathrm{T},0} = \mathbf{h}_{\mathrm{MVDR}}$, which is the MVDR filter,
- $\mu > 1$, results in low residual noise at the expense of high speech distortion,
- $\mu < 1$, results in high residual noise and low speech distortion.

Again, we observe here as well that the tradeoff, Wiener, and MVDR filters are equivalent up to a scaling factor. As a result, the output SNR of the tradeoff filter is independent of $\mu$ and is $\mathrm{oSNR}\left(\mathbf{h}_{\mathrm{T},\mu}\right) = \lambda_{\max},\ \forall \mu$. The noise-plus-interference reduction factor is given by $\xi_{\mathrm{nr}}\left(\mathbf{h}_{\mathrm{T},\mu}\right) = \frac{\left(\mu+\lambda_{\max}\right)^2}{\mathrm{iSNR}\ \lambda_{\max}}$ and is a decreasing function of $\mu$, while the signal distortion index is given by $\upsilon_{\mathrm{sd}}\left(\mathbf{h}_{\mathrm{T},\mu}\right) = \frac{\left(\mu+\lambda_{\max}\right)^2}{\mathrm{iSNR}\cdot\lambda_{\max}}$, which is an increasing function of the same parameter.

### 5.4. The Linearly Constrained Minimum Variance (LCMV) Filter

We can derive an LCMV filter [10] which can handle more than one linear constraint, by exploiting the structure of the noise as well as the noise-free structures.

In Section 3, we decomposed the vector $\mathbf{x}(k)$ into two orthogonal components to extract the desired signal, $x_1(k)$. We can also decompose (but for a different objective as explained below) the noise signal vector, $\mathbf{v}(k)$, into two orthogonal terms:

$$\mathbf{v}(k) = v_1(k)\boldsymbol{\gamma}_{\mathbf{v}} + \mathbf{v}'(k), \quad (22)$$

where $\boldsymbol{\gamma}_{\mathbf{v}}$ and $\mathbf{v}'(k)$ are defined in a similar way to $\boldsymbol{\gamma}_{\mathbf{x}}$ and $\mathbf{x}'(k)$.

Our problem this time is the following. We wish to perfectly recover our desired signal, $x_1(k)$, and completely remove the correlated components, $v_1(k)\boldsymbol{\gamma}_{\mathbf{v}}$. Thus, the two constraints can be put together in a matrix form as $\mathbf{C}^T\mathbf{h} = \mathbf{i}$, where $\mathbf{C} = \left[\begin{array}{cc} \boldsymbol{\gamma}_{\mathbf{x}} & \boldsymbol{\gamma}_{\mathbf{v}} \end{array}\right]$ is our constraint matrix of size $NL \times 2$ and $\mathbf{i} = [1\ 0]^T$. Then, our optimal filter is obtained by minimizing the overall mixture of sounds energy at the filter output, with the constraints that the correlated noise components are canceled and the desired speech is preserved. The LCMV filter is then given by [10]

$$\mathbf{h}_{\mathrm{LCMV}} = \mathbf{R}_{\mathbf{y}}^{-1}\mathbf{C}\left[\mathbf{C}^T\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{C}\right]^{-1}\mathbf{i}. \quad (23)$$

In a similar way to [8], we demonstrate that the MVDR filter can be written as a linear combination of two beamformers

$$\mathbf{h}_{\mathrm{MVDR}} = \varrho\mathbf{h}_{\mathrm{LCMV}} + (1-\varrho)\mathbf{h}_{\mathrm{MATCH}} \quad (24)$$

where

$$\mathbf{h}_{\mathrm{MATCH}} = \frac{\mathbf{R}_{\mathrm{in}'}^{-1}\boldsymbol{\gamma}_{\mathbf{x}}}{\boldsymbol{\gamma}_{\mathbf{x}}^T\mathbf{R}_{\mathrm{in}'}^{-1}\boldsymbol{\gamma}_{\mathbf{x}}} \quad (25)$$

$$\varrho = \frac{\boldsymbol{\gamma}_{\mathbf{v}}^T\mathbf{R}_{\mathrm{in}'}^{-1}\boldsymbol{\gamma}_{\mathbf{v}}\left(1-\kappa^2\right)}{\sigma_{v_1}^{-2} + \boldsymbol{\gamma}_{\mathbf{v}}^T\mathbf{R}_{\mathrm{in}'}^{-1}\boldsymbol{\gamma}_{\mathbf{v}}\left(1-\kappa^2\right)} \quad (26)$$

$$\kappa = \frac{\left(\boldsymbol{\gamma}_{\mathbf{x}}^T\mathbf{R}_{\mathrm{in}'}^{-1}\boldsymbol{\gamma}_{\mathbf{v}}\right)^2}{\left(\boldsymbol{\gamma}_{\mathbf{v}}^T\mathbf{R}_{\mathrm{in}'}^{-1}\boldsymbol{\gamma}_{\mathbf{v}}\right)\left(\boldsymbol{\gamma}_{\mathbf{x}}^T\mathbf{R}_{\mathrm{in}'}^{-1}\boldsymbol{\gamma}_{\mathbf{x}}\right)}, \quad (27)$$

where $\mathbf{R}_{\mathrm{in}'} = \mathbf{R}_{\mathbf{v}'} + \mathbf{R}_{\mathbf{x}'}$ is the covariance matrix of all the incoherent interference plus noise components. $\varrho$ is a tradeoff parameter between $\mathbf{h}_{\mathrm{MATCH}}$ and $\mathbf{h}_{\mathrm{LCMV}}$ that are optimal in the absence of interference and non-coherent noise, respectively. $\kappa$ measures the collinearity[1] between the vectors $\boldsymbol{\gamma}_{\mathbf{x}}$ and $\boldsymbol{\gamma}_{\mathbf{v}}$ in some transform domain defined by the non-coherent noise whitening matrix $\mathbf{R}_{\mathrm{in}'}^{-1/2}$ [8]. We observe from (24) that when $\varrho$ tends to 0, the MVDR filter tends to the LCMV filter while when $\varrho$ tends to 1, the MVDR filter tends to the matched filter, thereby trading off the rejection of the coherent noise and the reduction of the other residual noise components [8]. This tradeoff is determined by the value of the collinearity factor, $\kappa$, and the generalized coherent-to-other noise components ratio, $\sigma_{v_1}^2\boldsymbol{\gamma}_{\mathbf{v}}^T\mathbf{R}_{\mathrm{in}'}^{-1}\boldsymbol{\gamma}_{\mathbf{v}}$.

We always have $\mathrm{oSNR}\left(\mathbf{h}_{\mathrm{LCMV}}\right) \leq \mathrm{oSNR}\left(\mathbf{h}_{\mathrm{MVDR}}\right)$, $\upsilon_{\mathrm{sd}}\left(\mathbf{h}_{\mathrm{LCMV}}\right) = 0$, and we can show that

$$\xi_{\mathrm{nr}}\left(\mathbf{h}_{\mathrm{LCMV}}\right) \leq \xi_{\mathrm{nr}}\left(\mathbf{h}_{\mathrm{MVDR}}\right) \leq \xi_{\mathrm{nr}}\left(\mathbf{h}_{\mathrm{W}}\right). \quad (28)$$

The LCMV filter is able to remove all the correlated noise but at the price of a decreased overall noise reduction factor.

## 6. SIMULATION RESULTS

In our setup, we use real channel impulse responses recorded in an acoustic chamber of size 6.7 m long by 6.1 m wide by 2.9 m high. The reverberation time is $T_{60} = 580$ ms. A linear array of 4 microphones is used. The microphone array lies along the first dimension with the first element at $(2.437, 0.5, 1.4)$ m and inter-element spacing of 0.1 m. A target speaker is located at $(1.337, 1.938, 1.6)$ and

[1] The larger is $\kappa$, the more collinear (or less orthogonal) are $\mathbf{R}_{\mathrm{in}'}^{-1/2}\boldsymbol{\gamma}_{\mathbf{x}}$ and $\mathbf{R}_{\mathrm{in}'}^{-1/2}\boldsymbol{\gamma}_{\mathbf{v}}$

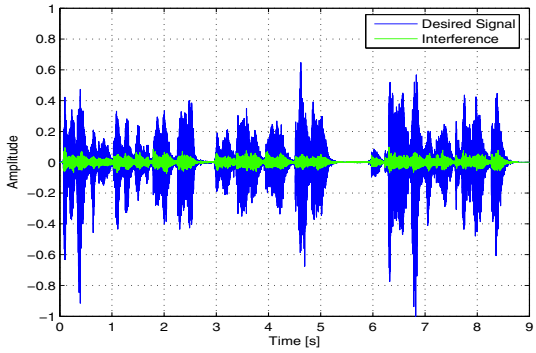


**Fig. 1**. Desired signal and interference at the output of the multichannel Wiener filter. The number of filter taps is $L = 100$.

generates a 12-second-long female speech (8 kHz sampling frequency) while an interference source is located at $(3.337, 1.938, 1.6)$ m and generates a ventilation (fan) noise. Segments of babble noise from the noisex database [11] are added to the noise-free observations. Our purpose here is to investigate the fundamental limits of the investigated filters. Thus, we put aside the problem of noise statistic estimation, and assume that the noise samples are available when estimating $\mathbf{R}_{\mathbf{v}}$. The statistics estimation and filtering are performed using batch processing with 256 ms-length windows and an overlap rate of $75\%$ between consecutive frames.

For completeness, we investigate the performance of the tradeoff filter for three conditions: $\mu = 5,\ 1$, and $0$. Recall that $\mu = 0$ corresponds to the MVDR and $\mu = 1$ corresponds to the traditional multichannel Wiener filter. We also include comparisons to the LCMV beamformer. The noise reduction factor, signal distortion index and output SNR are used for performance evaluation here. To demonstrate the relevance of the new definitions of our metrics in Section 4, we compare them to the old definitions where the interference (as described in Section 3) is included in the desired signal part (e.g., see [2, 7] for details). The results are presented for a variable number of filter taps at a constant overall input SNR (the speech-to-babble-noise ratio and the speech-to-fan-noise ratio are both equal to 10 dB).

We start our simulations with an illustration of the relevance of the new linear decomposition in Section 3 and the resulting residual interference as compared to the desired signal at the output of the multichannel Wiener filter in Figure 1. The interference level is remarkably lower than the desired signal, but taking it into account substantially alters the expected performance of the filters. Including the interference in the definition of the desired signal is certainly not correct since both components are, *by definition*, orthogonal.

Figures 2, 3, and 4 show the effect of the number of taps on the signal distortion index, noise reduction factor, and SNR at the output of the investigated filters. Note that there is a clear discrepancy between the old and new definitions of the three performance measures. For a given filter, the old metrics indicate larger noise reduction, output SNR, and signal distortion values than their new counterparts. This discrepancy is caused by the interference in the filtered speech signal which is traditionally included in the desired signal part.

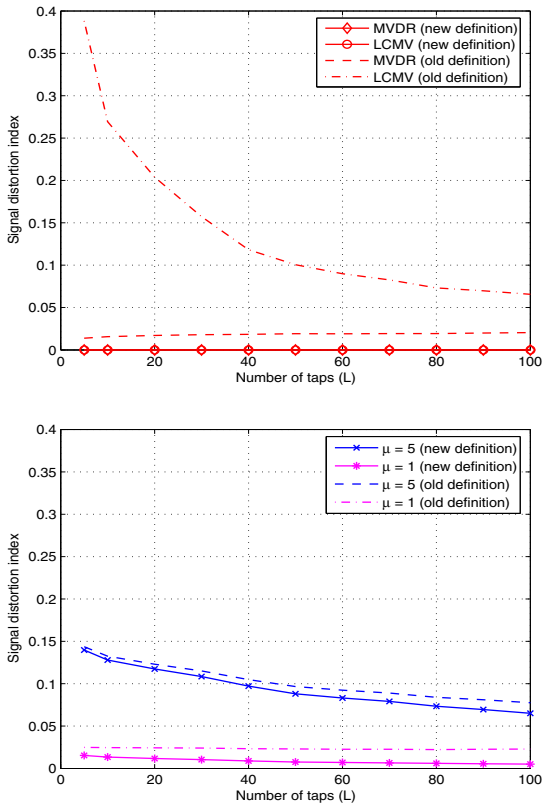


**Fig. 2**. Effect of the number of taps, $L$, on the signal distortion index. Top: MVDR and LCMV. Bottom: Tradeoff filter with $\mu = 5$ and $\mu = 1$. Old and new definitions are compared.

By observing the newly defined performance measures, we notice that as the tradeoff parameter increases from 1 to 5, more noise reduction and signal distortion are obtained at the output of the tradeoff filter. Comparing the top and bottom subplots in Figures 2 and 3, respectively, we see that the MVDR achieves less noise reduction than the Wiener filter but also zero distortion of the desired signal. These results agree with the theoretical analysis. The LCMV does not distort the desired signal but it achieves the lowest noise reduction values. The last result is also confirmed by the link between the MVDR and LCMV in (24). As for the output SNR, we expect from our analysis in Section 5 that its values would be the same for the tradeoff filter regardless of the value of $\mu$. The plots in Figure 4 do not perfectly agree with this theoretical finding. To explain this mismatch, recall that in all our analysis, we assumed the coexistence of the desired speech signal and noise at every time instant. This assumption is not always valid for speech

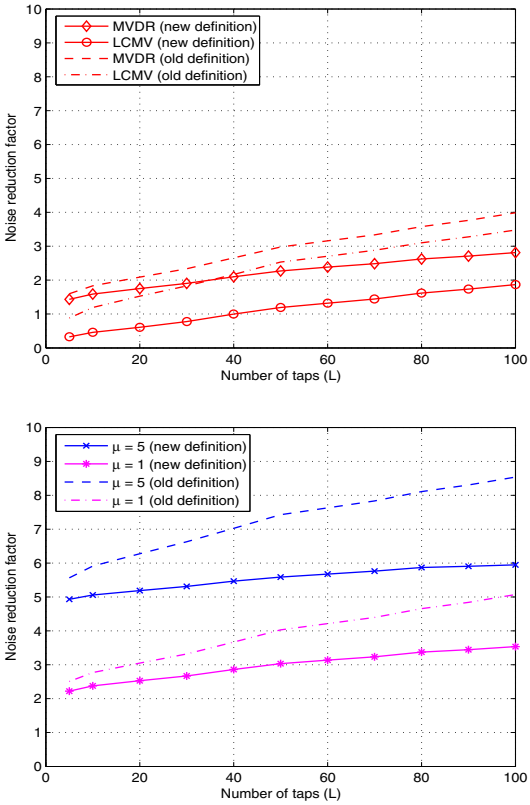


**Fig. 3**. Effect of the number of taps, $L$, on the noise reduction factor. Top: MVDR and LCMV. Bottom: Tradeoff filter with $\mu = 5$ and $\mu = 1$. Old and new definitions are compared.

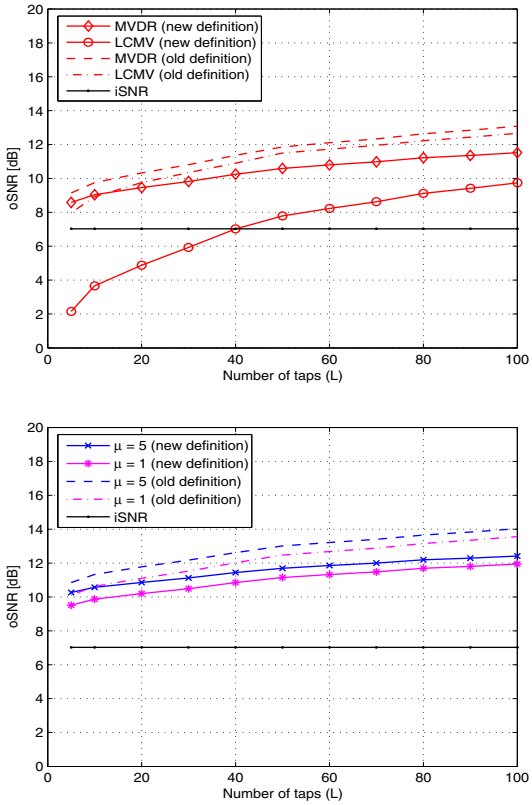


**Fig. 4**. Effect of the number of taps, $L$, on the output SNR. Top: MVDR and LCMV. Bottom: Tradeoff filter with $\mu = 5$ and $\mu = 1$. Old and new definitions are compared.

which is known to be non-stationary and its energy may frequently decay to zero. In noise-only frames, the filters corresponding to $\mu > 0$ attenuate all their outputs to 0 while the MVDR is regularized enough and does not suppress all signals in these frames. Besides, it can be empirically verified that when increasing $\mu$, the tradeoff filter becomes more aggressive in terms of signals suppression when the speech energy decays. Hence, less noise reduction is achieved by the MVDR followed by the Wiener filter in noise-only frames. This translates into larger SNR gains of the tradeoff filter with $\mu = 5$ as compared to the multichannel Wiener filter and MVDR.

## 7. CONCLUSIONS

In this paper, we presented a new perspective on well known time-domain multichannel noise reduction approaches. We demonstrated that the observed noise-free signals are composed of two main components: the first is perfectly coherent across the sensors and obtained by projecting the noise-free observations on the reference microphone speech signal, while the second is orthogonal to the desired signal and considered as interference. Thanks to this new decomposition, we re-formulated the multichannel Wiener, MVDR, LCMV, and tradeoff filters and expressed them in a simple way that allows for a better comprehension of their functioning. The results of this contribution can be easily extended to noise reduction in transform domains, in particular those investigated in [6] or the convolutive frequency-domain approach [5].

## 8. REFERENCES


[1] M. Brandstein and D. B. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Germany: Springer-Verlag, 2001.

[2] Y. Huang, J. Benesty, and J. Chen, "Analysis and comparison of multichannel noise reduction methods in a common framework," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, pp. 957–968, July 2008.

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

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

[5] R. Talmon, I. Cohen and S. Gannot, "Convolutive transfer function generalized sidelobe canceler," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, pp. 1420–1434, Sept. 2009.

[6] J. Benesty, J. Chen, and Y. Huang, "Noise reduction algorithms in a generalized transform domain *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, pp. 1109–1123, Aug. 2009.

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

[8] M. Souden, J. Benesty, and S. Affes, "A study of the LCMV and MVDR noise reduction filters," *IEEE Trans. Signal Process.*, vol. 18, pp. 4925–4935 , Sept. 2010.

[9] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.

[10] O. Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, pp. 926–935, Jan. 1972.

[11] A. P. Varga , H. J. M. Steenekan , M. Tomlinson, and D. Jones, "The noisex-92 study on the effect of additive noise on automatic speech recognition," *Tech. Rep., DRA Speech Research Unit,* 1992.