MULTICHANNEL NOISE REDUCTION WIENER FILTER IN THE KARHUNEN-LOÈVE EXPANSION DOMAIN

Yesenia Lacouture-Parodi, Emanuël A. P. Habets

International Audio Laboratories Erlangen[†]
91058 Erlangen, Germany
{yesenia.lacouture,emanuel.habets}@audiolabs-erlangen.de

Jacob Benesty

INRS-EMT, University of Quebec
Montreal, QC H5A 1K6, Canada
benesty@emt.inrs.ca

ABSTRACT

This paper explores the noise reduction problem in the Karhunen-Loève expansion (KLE) domain from a multichannel perspective. Based on formulations proposed for the design of optimal single-channel noise reduction in the KLE domain, we formulate the multichannel noise reduction in the KLE domain. Two different performance measures are presented: the noise reduction and speech distortion. The optimal multichannel Wiener filter is derived and its performance in terms of noise reduction and speech distortion is compared with the performance of the optimal single-channel Wiener filter. Experimental results show that a significant improvement in performance is obtained when using multiple microphone signals. The multichannel Wiener filter results also in better noise reduction in the presence of coherent noise sources.

***Index Terms*—** Noise reduction, speech enhancement, Karhunen-Loève expansion (KLE), multichannel, Wiener filter.

1. INTRODUCTION

Traditionally, the noise reduction problem is approached in either the time or frequency domain. The optimal filters are often estimated by minimizing the mean-square error between the clean signal and its estimate. The time domain approach can be sample based, estimating one speech sample at a time, while the frequency domain is often formulated on a frame basis, i.e. a block of noisy speech signal is transformed into the frequency domain using the discrete Fourier transform (DFT) and then a filter is estimated and applied to the frame [1]. The frequency-domain approaches are in general more flexible with respect to controlling the noise reduction performance versus the speech distortion, though special attention has to be paid to the aliasing distortion caused by the independent processing of subbands. The time domain approaches do not suffer from aliasing problems, but are less flexible regarding the performance and computational complexity [2]. The use of signal-dependent transforms has shown some advantages with regard to speech distortion and noise reduction [2–4]

Recently, single-channel noise reduction formulation in the Karhunen-Loève Expansion (KLE) domain, which employs a signal-dependent transform, has received special attention [2, 4, 5]. The basic advantages of using the KL transform is that if the covariance matrices are properly calculated, there will be no aliasing problems and that the desired speech and noise may be better separated as opposed to the frequency-domain methods [6]. A general formulation of the KLE domain approach and the design of different optimal filters has been previously proposed in [2] and [6]. In those studies, the clean speech signal is estimated from a noisy observation, which is obtained from a single microphone. It has been shown that a better noise reduction performance is achieved when properly choosing the parameters to calculate the filters.

In this study, we explore the possibility of using multiple microphone signals to improve the performance of the optimal noise reduction filters in the KLE domain [6]. One possible advantage of adding more channels, is the potential of extending the noise reduction problem in the KLE domain into spatial filtering techniques and being able to further exploit its benefits. We present here a formulation of the multichannel noise reduction problem in the KLE domain. As an example, we derive an optimal multichannel Wiener filter and compare its performance in terms of the noise reduction and speech distortion with that obtained with a single-channel Wiener filter. The optimal single-channel Wiener filter presented in the example corresponds to the Class I optimal Wiener filter described in [2].

2. PROBLEM FORMULATION

Let us first consider a microphone array with N microphones that captures a noisy signal $y(k)$, where k is the discrete-time index. The signals received at each microphone can be thus defined as [7]

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

where $x_n(k)$ is the signal of interest and $v_n(k)$ is the additive noise captured by the nth microphone. We assume that $x_n(k)$ and $v_n(k)$ are uncorrelated and zero mean. By definition, $x_n(k)$ is coherent across the array and $v_n(k)$ is typically only partially coherent across the array. These signals are considered to be real, broadband and stationary. The latter is assumed to simplify the development and analysis of this work. By processing the data by blocks of L samples, (1) can be expressed in a vector form as

$$\mathbf{y}_n(m) = \mathbf{x}_n(m) + \mathbf{v}_n(m), \quad n = 1, 2, \ldots, N, \tag{2}$$

where $\mathbf{y}_n(m) = [y_n(mL)\ y_n(mL+1)\ \ldots\ y_n(mL+L-1)]^T$, the superscript ^{T} denotes transpose, $m \geq 0$ is the time-frame index, L is the frame length and $\mathbf{x}_n(m)$ and $\mathbf{v}_n(m)$ are defined similarly to $\mathbf{y}_n(m)$. Since $x_n(k)$ and $v_n(k)$ are uncorrelated by assumption, the $L \times L$ correlation matrix of the nth microphone is $\mathbf{R}_{\mathbf{y}_n} = \mathbf{R}_{\mathbf{x}_n} + \mathbf{R}_{\mathbf{v}_n}$, where $\mathbf{R}_{\mathbf{y}_n} = E[\mathbf{y}_n(m)\mathbf{y}_n^T(m)]$ is the correlation matrix of the noisy signal $\mathbf{y}_n(m)$, $\mathbf{R}_{\mathbf{x}_n}$ and $\mathbf{R}_{\mathbf{v}_n}$ are the correlation matrices of $\mathbf{x}_n(m)$ and $\mathbf{v}_n(m)$, respectively, and $E[\cdot]$ denotes mathematical expectation.

In this paper our desired signal is designated by the clean signal received at microphone 1, i.e., $x_1(k)$[1]. Thus, giving N mixtures of two uncorrelated signals $x_n(k)$ and $v_n(k)$, our aim is to preserve

[†] A joint institution of the University of Erlangen-Nuremberg and Fraunhofer IIS

[1] In principle, any signal $x_n(k)$ could be used as the reference.

978-1-4673-0046-9/12/$26.00 ©2012 IEEE

$x_1(k)$ while minimizing the contribution of the noise terms $v_n(k)$ at the array output [7].

3. KARHUNEN-LOÈVE EXPANSION (KLE)

In this section, we briefly recall the principle of the KLE, which can be applied to $\mathbf{y}_n(m)$, $\mathbf{x}_n(m)$, or $\mathbf{v}_n(m)$.

Let us first diagonalize the correlation matrix $\mathbf{R}_{\mathbf{y}_n}$

$$\mathbf{Q}_n^T \mathbf{R}_{\mathbf{y}_n} \mathbf{Q}_n = \boldsymbol{\Lambda}_n, \tag{3}$$

where $\mathbf{Q}_n = [\mathbf{q}_{n,1}\ \mathbf{q}_{n,2}\ \dots\ \mathbf{q}_{n,L}]$ and $\boldsymbol{\Lambda}_n = \mathrm{diag}(\lambda_{n,1}, \lambda_{n,2}, \dots, \lambda_{n,L})$ are, respectively, orthogonal and diagonal matrices. The orthonormal vector $\mathbf{q}_{n,l}$ is the eigenvector corresponding to the eigenvalue $\lambda_{n,l}$.

We can write the vector $\mathbf{y}_n(m)$ as a combination (expansion) of the eigenvectors of the correlation matrix $\mathbf{R}_{\mathbf{y}_n}$ [2] as follows:

$$\mathbf{y}_n(m) = \sum_{l=1}^{L} c_{y_n,l}(m) \mathbf{q}_{n,l}, \tag{4}$$

where

$$c_{y_n,l}(m) = \mathbf{q}_{n,l}^T \mathbf{y}_n(m), \quad l = 1, 2, \dots, L \tag{5}$$

are the coefficients of the expansion and l is the mode index. The representation of the random vector $\mathbf{y}_n(m)$ described by (4) and (5) is the KLE, where (4) and (5) are, respectively, the synthesis and analysis part of the expansion [8]. The signals $\mathbf{x}_n(m)$ and $\mathbf{v}_n(m)$ are synthesized and analyzed in a similar way as $\mathbf{y}_n(m)$. Left multiplying both sides of (2) by $\mathbf{q}_{n,l}^T$, the time-domain signal model is transformed into the KLE domain as

$$c_{y_n,l}(m) = c_{x_n,l}(m) + c_{v_n,l}(m), \quad l = 1, 2, \dots, L. \tag{6}$$

Thus, the multichannel noise reduction in the KLE domain consists basically of the estimation of the coefficients $c_{x_1,l}(m)$, $l = 1, 2, \dots, L$, from the observations $c_{y_n,l}(m)$, $n = 1, 2, \dots, N$, $l = 1, 2, \dots, L$.

4. LINEAR ARRAY MODEL

In the KLE domain, we are going to focus on the simplest linear model for array processing, which is realized by applying a real weight to the output of each microphone and summing across the aperture, i.e.,

$$\begin{aligned} c_{z,l}(m) &= \sum_{n=1}^{N} h_{l,n} c_{y_n,l}(m) \\ &= \mathbf{h}_l^T \mathbf{c}_{y_:,l}(m) \\ &= \mathbf{h}_l^T \mathbf{c}_{x_:,l}(m) + \mathbf{h}_l^T \mathbf{c}_{v_:,l}(m) \\ &= c_{x_\mathrm{f},l}(m) + c_{v_\mathrm{rn},l}(m), \quad l = 1, 2, \dots, L, \end{aligned} \tag{7}$$

where $c_{z,l}(m)$ is the estimate of $c_{x_1,l}(m)$, $\mathbf{h}_l = [h_{l,1}\ h_{l,2} \dots h_{l,N}]^T$ is the weight vector, and $\mathbf{c}_{y_:,l}(m) = [c_{y_1,l}(m)\ c_{y_2,l}(m) \dots c_{y_N,l}(m)]^T$ is a vector containing the observations from all microphones at time-frame m (the vectors $\mathbf{c}_{x_:,l}(m)$ and $\mathbf{c}_{v_:,l}(m)$ are defined in a similar way). The coefficients $c_{x_\mathrm{f},l}(m) = \mathbf{h}_l^T \mathbf{c}_{x_:,l}(m)$ and $c_{v_\mathrm{rn},l}(m) = \mathbf{h}_l^T \mathbf{c}_{v_:,l}(m)$ are the filtered desired speech signal and residual noise in the KLE-domain respectively.

At frame m our desired signal is $c_{x_1,l}(m)$. However, the vector $\mathbf{c}_{x_:,l}$ contains both the desired signal $c_{x_1,l}(m)$ and the components $c_{x_i,l}(m)$, $i = 2, 3, \dots, N$, which are correlated with $c_{x_1,l}(m)$. We should thus decompose $\mathbf{c}_{x_:,l}(m)$ into two orthogonal components corresponding to the part of the desired signal and interference, i.e.,

$$\begin{aligned} \mathbf{c}_{x_:,l}(m) &= c_{x_1,l}(m) \boldsymbol{\gamma}_{\mathbf{c}_{x_:,l}} + \mathbf{c}'_{x_:,l}(m) \\ &= \mathbf{c}_{x_\mathrm{d},l}(m) + \mathbf{c}'_{x_:,l}(m), \end{aligned} \tag{8}$$

where $\mathbf{c}_{x_\mathrm{d},l}(m) = c_{x_1,l}(m) \boldsymbol{\gamma}_{\mathbf{c}_{x_:,l}}$ is a signal vector depending on the desired signal $c_{x_1,l}(m)$, $\mathbf{c}'_{x_:,l}(m) = \mathbf{c}_{x_:,l}(m) - c_{x_1,l}(m) \boldsymbol{\gamma}_{\mathbf{c}_{x_:,l}}$ is the interference signal, and

$$\boldsymbol{\gamma}_{\mathbf{c}_{x_:,l}} = \frac{E\left[c_{x_1,l}(m) \mathbf{c}_{x_:,l}(m)\right]}{E\left[c_{x_1,l}^2(m)\right]} \tag{9}$$

is the partially normalized cross-correlation vector (of length N) between $c_{x_1,l}(m)$ and $\mathbf{c}_{x_:,l}$. In practice, we can estimate $\boldsymbol{\gamma}_{\mathbf{c}_{x_:,l}}$ by using $\boldsymbol{\gamma}_{\mathbf{c}_{y_:,l}}$ and $\boldsymbol{\gamma}_{\mathbf{c}_{v_:,l}}$, which can be estimated during noisy and noiseless periods.

5. PERFORMANCE MEASURES

We define in this section measures that help us assessing the performance of the multichannel noise reduction Wiener filter in the KLE domain. Before defining the noise reduction factor, it is useful first to define the signal-to-noise ratio (SNR). Since the signal we want to recover is the clean signal received at microphone 1, i.e., $x_1(k)$, this signal will serve as the reference signal.

First, we define the mode input SNR as

$$\mathrm{iSNR}_l = \frac{\phi_{c_{x_1,l}}}{\phi_{c_{v_1,l}}} = \frac{\mathbf{q}_{1,l}^T \mathbf{R}_{\mathbf{x}_1} \mathbf{q}_{1,l}}{\mathbf{q}_{1,l}^T \mathbf{R}_{\mathbf{v}_1} \mathbf{q}_{1,l}}, \quad l = 1, 2, \dots, L, \tag{10}$$

where $\phi_{c_{x_1,l}}$ and $\phi_{c_{v_1,l}}$ are the variances of $c_{x_1,l}(m)$ and $c_{v_1,l}(m)$, respectively. The fullmode input SNR is thus

$$\mathrm{iSNR} = \frac{\sum_{l=1}^{L} \mathbf{q}_{1,l}^T \mathbf{R}_{\mathbf{x}_1} \mathbf{q}_{1,l}}{\sum_{l=1}^{L} \mathbf{q}_{1,l}^T \mathbf{R}_{\mathbf{v}_1} \mathbf{q}_{1,l}} = \frac{\sigma_{x_1}^2}{\sigma_{v_1}^2}, \tag{11}$$

where $\sigma_{x_1}^2 = E[x_1^2(k)]$ and $\sigma_{v_1}^2 = E[v_1^2(k)]$ are the variances of $x_1(k)$ and $v_1(k)$ respectively. It can be shown that $\mathrm{iSNR} \le \sum_{l=1}^{L} \mathrm{iSNR}_l$. The mode output SNR, i.e., the SNR after the filtering operation, is defined as[2]

$$\mathrm{oSNR}(\mathbf{h}_l) = \frac{\mathbf{h}_l^T \boldsymbol{\Phi}_{c_{x_\mathrm{d}},l} \mathbf{h}_l}{\mathbf{h}_l^T \boldsymbol{\Phi}_{\mathrm{in},l} \mathbf{h}_l} = \frac{\phi_{c_{x_1,l}} \left(\mathbf{h}_l^T \boldsymbol{\gamma}_{\mathbf{c}_{x_:,l}}\right)^2}{\mathbf{h}_l^T \boldsymbol{\Phi}_{\mathrm{in},l} \mathbf{h}_l}, \tag{12}$$

for $l = 1, 2, \dots, L$. The matrix $\boldsymbol{\Phi}_{c_{x_\mathrm{d}},l} = E[\mathbf{c}_{x_\mathrm{d},l}(m) \mathbf{c}_{x_\mathrm{d},l}^T(m)] = \phi_{c_{x_1,l}} \boldsymbol{\gamma}_{\mathbf{c}_{x_:,l}} \boldsymbol{\gamma}_{\mathbf{c}_{x_:,l}}^T$ is the correlation matrix of $\mathbf{c}_{x_\mathrm{d},l}(m)$ and $\boldsymbol{\Phi}_{\mathrm{in},l} = \boldsymbol{\Phi}_{\mathbf{c}_{x_:,l}} - \boldsymbol{\Phi}_{\mathbf{c}_{x_\mathrm{d},l}} + \boldsymbol{\Phi}_{\mathbf{c}_{v_:,l}}$ is the interference-plus-noise correlation matrix. The matrices $\boldsymbol{\Phi}_{\mathbf{c}_{x_:,l}}$ and $\boldsymbol{\Phi}_{\mathbf{c}_{v_:,l}}$ are the correlation matrices of $\mathbf{c}_{x_:,l}(m)$ and $\mathbf{c}_{v_:,l}(m)$ respectively.

The fullmode output SNR is defined as

$$\mathrm{oSNR}(\mathbf{h}_:) = \frac{\sum_{l=1}^{L} \phi_{c_{x_1,l}} \left(\mathbf{h}_l^T \boldsymbol{\gamma}_{\mathbf{c}_{x_:,l}}\right)^2}{\sum_{l=1}^{L} \mathbf{h}_l^T \boldsymbol{\Phi}_{\mathrm{in},l} \mathbf{h}_l}. \tag{13}$$

The noise reduction factor gives more insight into the noise reduction performance of the filters [2, 9]. The mode and fullmode noise

[2] In this study, we consider the interference as part of the noise in the definitions of the performance measures.

reduction factors are defined as

$$\xi_{\text{nr}}(\mathbf{h}_l) = \frac{\phi_{c_{v_1,l}}}{\mathbf{h}_l^T \mathbf{\Phi}_{\text{in},l} \mathbf{h}_l}, \quad l = 1, 2, \ldots, L, \tag{14}$$

$$\xi_{\text{nr}}(\mathbf{h}_:) = \frac{\sum_{l=1}^{L} \phi_{c_{v_1,l}}}{\sum_{l=1}^{L} \mathbf{h}_l^T \mathbf{\Phi}_{\text{in},l} \mathbf{h}_l}. \tag{15}$$

To evaluate the amount of speech distortion we make use of the mode and fullmode speech distortion indexes [2,9], i.e.,

$$\upsilon_{\text{sd}}(\mathbf{h}_l) = \frac{E\left\{\left[c_{x_1,l}(m)\mathbf{h}_l^T \boldsymbol{\gamma}_{\mathbf{c}_{x:,l}} - c_{x_1,l}(m)\right]^2\right\}}{\phi_{c_{x_1,l}}}, \tag{16}$$

$$\upsilon_{\text{sd}}(\mathbf{h}_:) = \frac{\sum_{l=1}^{L} \phi_{c_{x_1,l}} \left(\mathbf{h}_l^T \boldsymbol{\gamma}_{\mathbf{c}_{x:,l}} - 1\right)^2}{\sum_{l=1}^{L} \phi_{c_{x_1,l}}}. \tag{17}$$

6. WIENER FILTER

By taking the gradient with respect to $\mathbf{h}_l$ of the mode mean-square error (MSE), which is defined as

$$J(\mathbf{h}_l) = E\left\{\left[\mathbf{h}_l^T \mathbf{c}_{y:,l}(m) - c_{x_1,l}(m)\right]^2\right\}, \tag{18}$$

and equating the result to zero, we can derive the multichannel Wiener filter:

$$\begin{aligned} \mathbf{h}_{\text{W},l} &= \mathbf{\Phi}_{\mathbf{c}_{y:,l}}^{-1} \mathbf{\Phi}_{\mathbf{c}_{x:,l}} \mathbf{i}_1 \\ &= \left(\mathbf{I}_N - \mathbf{\Phi}_{\mathbf{c}_{y:,l}}^{-1} \mathbf{\Phi}_{\mathbf{c}_{x:,l}}\right) \mathbf{i}_1 \qquad (19) \\ &= \phi_{c_{x_1,l}} \mathbf{\Phi}_{\mathbf{c}_{y:,l}}^{-1} \boldsymbol{\gamma}_{\mathbf{c}_{x:,l}}, \qquad (20) \end{aligned}$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix and $\mathbf{i}_1$ corresponds to the first column of $\mathbf{I}_N$. Note that for the particular case of $N = 1$, the equation (20) is equivalent to the Wiener filter of Class I described in [2]. It can be shown that $\mathbf{\Phi}_{\mathbf{c}_{y:,l}} = \phi_{c_{x_1,l}} \boldsymbol{\gamma}_{\mathbf{c}_{x:,l}} \boldsymbol{\gamma}_{\mathbf{c}_{x:,l}}^T + \mathbf{\Phi}_{\text{in},l}$, whose inverse can be determined with the Woodbury's identity:

$$\mathbf{\Phi}_{\mathbf{c}_{y:,l}}^{-1} = \mathbf{\Phi}_{\text{in},l}^{-1} - \frac{\mathbf{\Phi}_{\text{in},l}^{-1} \boldsymbol{\gamma}_{\mathbf{c}_{x:,l}} \boldsymbol{\gamma}_{\mathbf{c}_{x:,l}}^T \mathbf{\Phi}_{\text{in},l}^{-1}}{\phi_{c_{x_1,l}}^{-1} + \boldsymbol{\gamma}_{\mathbf{c}_{x:,l}}^T \mathbf{\Phi}_{\text{in},l}^{-1} \boldsymbol{\gamma}_{\mathbf{c}_{x:,l}}}. \tag{21}$$

Substituting (21) into (20), we obtain another interesting formulation of the Wiener filter:

$$\begin{aligned} \mathbf{h}_{\text{W},l} &= \frac{\mathbf{\Phi}_{\text{in},l}^{-1} \mathbf{\Phi}_{\mathbf{c}_{y:,l}} - \mathbf{I}_N}{1 - N + \text{tr}\left(\mathbf{\Phi}_{\text{in},l}^{-1} \mathbf{\Phi}_{\mathbf{c}_{y:,l}}\right)} \\ &= \frac{\mathbf{\Phi}_{\text{in},l}^{-1} \mathbf{\Phi}_{\mathbf{c}_{x_{\text{d}},l}}}{1 + \lambda_{\text{max},l}} \mathbf{i}_1, \qquad (22) \end{aligned}$$

where $\lambda_{\text{max},l} = \phi_{c_{x_1,l}} \boldsymbol{\gamma}_{\mathbf{c}_{x:,l}}^T \mathbf{\Phi}_{\text{in},l}^{-1} \boldsymbol{\gamma}_{\mathbf{c}_{x:,l}}$, $l = 1, 2, \ldots, L$ is the maximum eigenvalue of the matrix $\mathbf{\Phi}_{\text{in},l}^{-1} \mathbf{\Phi}_{\mathbf{c}_{x_{\text{d}},l}}$. From (22) it can be deduced that the mode output SNR for the wiener filter is $\text{oSNR}(\mathbf{h}_{\text{W},l}) = \lambda_{\text{max},l} = \text{tr}(\mathbf{\Phi}_{\text{in},l}^{-1} \mathbf{\Phi}_{\mathbf{c}_{y:,l}}) - N$, and the mode speech distortion index is a clear function of the mode output SNR:

$$\upsilon_{\text{sd}}(\mathbf{h}_{\text{W},l}) = \frac{1}{\left[1 + \text{oSNR}(\mathbf{h}_{\text{W},l})\right]^2}. \tag{23}$$

That is, the higher the value of $\text{oSNR}(\mathbf{h}_{\text{W},l})$, the less the desired signal is distorted. It can be shown that $\text{oSNR}(\mathbf{h}_{\text{W},l}) \geq \text{iSNR}_l$, since the Wiener filter maximizes the mode output SNR. The mode noise reduction factor is

$$\begin{aligned} \xi_{\text{nr}}(\mathbf{h}_{\text{W},l}) &= \frac{\left[1 + \text{oSNR}(\mathbf{h}_{\text{W},l})\right]^2}{\text{iSNR}_l \cdot \text{oSNR}(\mathbf{h}_{\text{W},l})} \\ &\geq \left[1 + \frac{1}{\text{oSNR}(\mathbf{h}_{\text{W},l})}\right]^2. \qquad (24) \end{aligned}$$

The fullmode oSNR is then

$$\text{oSNR}(\mathbf{h}_{\text{W},:}) = \frac{\sum_{l=1}^{L} \phi_{c_{x_1,l}} \frac{\text{oSNR}^2(\mathbf{h}_{\text{W},l})}{\left[1 + \text{oSNR}(\mathbf{h}_{\text{W},l})\right]^2}}{\sum_{l=1}^{L} \phi_{c_{x_1,l}} \frac{\text{oSNR}(\mathbf{h}_{\text{W},l})}{\left[1 + \text{oSNR}(\mathbf{h}_{\text{W},l})\right]^2}}. \tag{25}$$

Property: With the optimal KLE-domain Wiener filter given in (19), $\text{oSNR}(\mathbf{h}_{\text{W},:}) \geq \text{iSNR}$. Given the limitation of space, the proof is not presented in this paper.

7. EXPERIMENTAL RESULTS

In this section, we present the results of a set of experiments carried out to evaluate the performance of the multichannel noise reduction Wiener filter in the KLE domain. The results are compared with the performance obtained with the single-channel Wiener filter in the KLE domain, i.e., $N = 1$.

The clean signal used in the experiments was an anechoic recording of a female speaker with a length of 35 s. The sampling rate of the signal is 8 kHz. The clean signal was corrupted by a coherent noise source and incoherent noise (sensor noise). The coherent noise source was an anechoic recording of a different female speaker. Different coherent noise sources were evaluated during our studies and results showed similar trends. The incoherent noise used was low-pass filtered stationary white Gaussian noise. The noisy signal is then the addition of the clean anechoic speech, the incoherent and coherent noise.

In the simulations the microphone(s) and sources are located in a room of dimensions $x = 5$, $y = 6$ and $z = 4$ m. The room's reverberation time was set to 0.6 s. The room impulse responses were calculated using the image method [10]. For the multichannel case, we simulated an array of four microphones ($N = 4$) uniformly space with a distance $d = 5$ cm between microphones. For the single-channel case, we used only the reference microphone signal $x_1(k)$. The desired signal was simulated to be located 1 m away from the array at 40° azimuth and 2° elevation, where the point $(0^\circ, 0^\circ)$ is located right in front of the center array. The coherent noise source was simulated to be located 1.5 m away from the array at -40° azimuth and -2° elevation.

The implementation of the optimal noise reduction algorithms for the single-channel and multichannel cases was done in a similar fashion as described in [2]. In order to estimate the filter coefficients, we need to calculate the correlation matrices $\mathbf{\Phi}_{\mathbf{c}_{y:,l}}$ and $\mathbf{\Phi}_{\mathbf{c}_{v:,l}}$. At time frame m, the correlation matrix $\mathbf{\Phi}_{\mathbf{c}_{y:,l}}$ is estimated using the same recursive approach as the one presented in [2], namely

$$\mathbf{\Phi}_{\mathbf{c}_{y:,l}} = \alpha_y \mathbf{\Phi}_{\mathbf{c}_{y:,l-1}} + (1 - \alpha_y) \mathbf{c}_{y:,l}(m) \cdot \mathbf{c}_{y:,l}^T(m), \tag{26}$$

where α_y is a forgetting factor[3]. To estimate $\mathbf{\Phi}_{\mathbf{c}_{v:,l}}$ we would need in practice a noise estimator or a voice activity detector (VAD) to be able to compute the coefficients $\mathbf{c}_{v:,l}$. In order not to include the influence of possible errors from the noise estimator or the VAD, we calculated the coefficients $\mathbf{c}_{v:,l}$ directly from the noise signals. The estimation of $\mathbf{\Phi}_{\mathbf{c}_{v:,l}}$ is done in a similar fashion as in (26)

[3] The forgetting factors were set to $\alpha_y = 0.85$ and $\alpha_v = 0.91$, which were found to be optimal in terms of noise reduction and speech distortion.

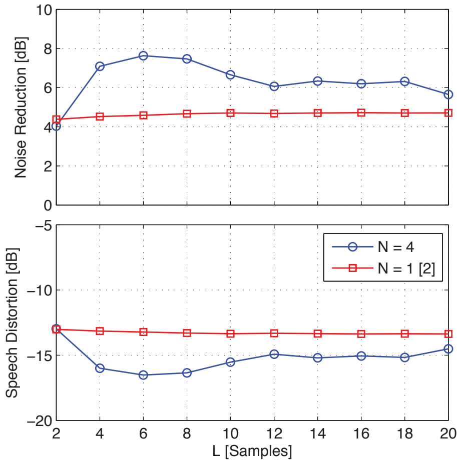

Fig. 1. Noise reduction and speech distortion as a function of L for a desired speech signal corrupted by another speech signal and stationary white Gaussian noise; iSINR $= 10$ dB, iSCNR $= 15$ dB, RT$_{60}$ $= 0.6$ s, and $N = 1, 4$.

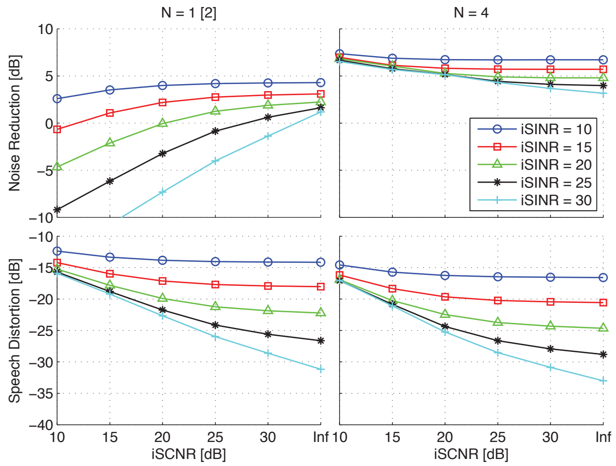

Fig. 2. Noise reduction and speech distortion as a function of iSCNR and iSINR a desired speech signal corrupted by another speech signal and stationary white Gaussian noise; RT$_{60}$ $= 0.6$ s, $N = 1, 4$, and $L = 6$.

but with a different forgetting factor α_v. The performance measures described in Section 5 were calculated as a function of different parameters such as the frame length L, number of microphones N, input signal-to-coherent-noise ratio (iSCNR) and input signal-to-incoherent-noise ratio (iSINR), among others. Here we present some of the results obtained for different frame lengths and the performance as a function of the iSCNR and iSINR.

Figure 1 shows the noise reduction and speech distortion as a function of the frame length L for the optimal single-channel and multichannel Wiener filter. It is clear from the figure that the performance of the filters is improved when using more microphones. For the multichannel case, a rather good performance is obtained when L is around 6 and decreases slightly when L is larger than 8, while for the single-channel case the frame length does not have a significant influence in the overall performance.

Figure 2 presents the noise reduction and speech distortion as a function of iSCNR and iSINR. It is interesting to notice that, in the presence of coherent noise, the performance of the single-channel case is in most instances not acceptable ($\xi_{\text{nr}}(\mathbf{h}_{\text{W},:}) < 0$ dB) when the iSINR is larger than the iSCNR. On the other hand, the multichannel case shows a rather stable performance for the different combinations of iSCNR and iSINR and, in general, the noise reduction values are significantly larger that those obtained with a single microphone. Both filters perform similarly with respect to speech distortion, though there is an improvement when employing multiple microphone signals.

8. CONCLUSIONS

We presented in this paper the multichannel noise reduction problem in the KLE domain. We formulated the KL transform for multiple receivers and presented the optimal Wiener filter for this case. Two different performance measures were defined and used to evaluate the performance of the optimal filters. Experimental results were then compared with the performance obtained with the optimal single-channel Wiener filter. The results clearly show an improvement in performance when employing multiple microphone signals. Additionally, while the single-channel approach showed to be sensitive to coherent noise, the multichannel case showed better noise reduction. We thus believe that there is a great potential for the use of multiple microphone signals to further exploit the advantages of the noise reduction problem in the KLE domain. The next step would be to explore the potential benefits of exploiting the correlation between subsequent time-frames [2].

9. REFERENCES

[1] J. Benesty, S. Makino, and J. Chen, *Speech Enhancement*. Springer-Verlag, Berlin, Germany, 2005.

[2] J. Chen, J. Benesty, and Y. Huang, "Study of the noise-reduction problem in the Karhunen-Loève expansion domain," *IEEE Trans. Audio, Speech, & Language Process.*, vol. 17, no. 4, pp. 787–802, 2009.

[3] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sørensen, "Reduction of broad-band noise in speech by truncated QSVD," *IEEE Trans. Speech & Audio Process.*, vol. 3, no. 6, pp. 439–448, 1995.

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

[5] U. Mittal and N. Phamdo, "Signal/noise KLT based approach for enhancing speech degraded by colored noise," *IEEE Trans. Speech & Audio Process.*, vol. 8, no. 2, pp. 159–167, 2000.

[6] J. Benesty, J. Chen, and Y. Huang, *Speech Enhancement in the Karhunen–Loève Expansion Domain*. Synthesis Lectures on Speech and Audio Processing, Morgan & Claypool, 2011.

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

[8] S. Haykin, *Adaptive Filter Theory*. Prentice-Hall, Upper Saddle River, NJ, fourth edition, 2002.

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

[10] J. B. Allen, "Image method for efficiently simulating small-room acoustics," *Journal of The Acoustical Society of America*, 1979.