ON NOISE REDUCTION IN THE KARHUNEN-LOÈVE EXPANSION DOMAIN

Jacob Benesty¹, Jingdong Chen², and Yiteng (Arden) Huang³

¹: INRS-EMT, University of Quebec Montreal, QC H5A 1K6, Canada benesty@emt.inrs.ca ²: Bell Labs, Alcatel-Lucent Murray Hill, NJ 07974, USA jingdong@research.bell-labs.com ³: WeVoice, Inc. Bridgewater, NJ 08807, USA ardenhuang@gmail.com

ABSTRACT

In this paper, we study the noise-reduction problem in the Karhunen-Loève expansion domain. We develop two classes of optimal filters. The first class estimates a frame of speech by filtering the corresponding frame of the noisy speech. We will show that several well-known existing methods belong or are closely related to this category. The second class, which has not been studied before, obtains noise reduction by filtering not only the current frame, but also a number of previous consecutive frames of the noisy speech. We will discuss how to design the optimal noise-reduction filters in each class and demonstrate the properties of the deduced optimal filters.

Index Terms— Noise reduction, Karhunen-Loève expansion, Pearson correlation coefficient, Wiener filter, tradeoff filter.

1. PROBLEM FORMULATION

The noise-reduction problem considered in this paper is to recover a signal of interest (clean speech or desired signal) x(k) of zero mean from the noisy observation (microphone signal)

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

where k is the discrete time index, and v(k) is the unwanted additive noise, which is assumed to be a zero-mean random process (white or colored) and uncorrelated with x(k). In a vector form, the signal model given in (1) can be written as

 $\mathbf{y}(k) = \mathbf{x}(k) + \mathbf{v}(k),$ (2) where $\mathbf{y}(k) = \begin{bmatrix} y(k) & y(k-1) & \cdots & y(k-L+1) \end{bmatrix}^T$, superscript ^T denotes transpose of a vector or a matrix, L is the frame length, and $\mathbf{x}(k)$ and $\mathbf{v}(k)$ are defined similarly to $\mathbf{y}(k)$. Since x(k)and v(k) are uncorrelated, the correlation matrix of the noisy signal is equal to the sum of the correlation matrices of the speech and noise signals, i.e., $\mathbf{R}_{yy} = \mathbf{R}_{xx} + \mathbf{R}_{vv}$, where $\mathbf{R}_{yy} = E[\mathbf{y}(k)\mathbf{y}^T(k)]$ is the correlation matrix of the signal y(k), and \mathbf{R}_{xx} and \mathbf{R}_{vv} are the correlation matrices of the signals x(k) and v(k) respectively, which are defined similarly to \mathbf{R}_{yy} . With this vector form of the signal model, the noise-reduction problem becomes one of estimating $\mathbf{x}(k)$ from the observation vector $\mathbf{y}(k)$, which is generally achieved by the following filtering process [1]–[3]:

$$\mathbf{t}(k) = \mathbf{H}\mathbf{y}(k) = \mathbf{H}\left[\mathbf{x}(k) + \mathbf{v}(k)\right],$$
(3)

where **H** is a filtering matrix of size $L \times L$. So, the core problem of noise reduction is to find a matrix **H** that would attenuate the noise as much as possible while keeping the distortion of the clean speech low.

2. KARHUNEN-LOÈVE EXPANSION AND ITS DOMAIN

In this section, we briefly recall the basic principle of the so-called Karhunen-Loève expansion (KLE) and show how we can work in the KLE domain.

Let the $L \times 1$ vector $\mathbf{x}(k)$ denote a data sequence drawn from a zero-mean stationary process with the correlation matrix \mathbf{R}_{xx} . This matrix can be diagonalized as follows [4]:

$$\mathbf{Q}^T \mathbf{R}_{xx} \mathbf{Q} = \mathbf{\Lambda},\tag{4}$$

where $\mathbf{Q} = [\mathbf{q}_1 \ \mathbf{q}_2 \ \cdots \ \mathbf{q}_L]$ and $\mathbf{\Lambda} = \text{diag}[\lambda_1 \ \lambda_2 \ \cdots \ \lambda_L]$ are, respectively, orthogonal and diagonal matrices. The orthonormal vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_L$ are the eigenvectors corresponding, respectively, to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_L$ of the matrix \mathbf{R}_{xx} .

The vector $\mathbf{x}(k)$ can be written as a combination (expansion) of the eigenvectors of the correlation matrix \mathbf{R}_{xx} as follows:

$$\mathbf{x}(k) = \sum_{l=1}^{L} a_{x,l}(k) \mathbf{q}_l,$$
(5)

where

$$a_{x,l}(k) = \mathbf{q}_l^T \mathbf{x}(k), \ l = 1, 2, \dots, L,$$
(6)

are the coefficients of the expansion. The representation of the random vector $\mathbf{x}(k)$ described by (5) and (6) is the KLE where (5) is the synthesis part and (6) represents the analysis part [4].

From (6), we can easily verify that

$$E[a_{x,l}(k)] = 0, \ l = 1, 2, \dots, L \tag{7}$$

and

$$E[a_{x,i}(k)a_{x,j}(k)] = \begin{cases} \lambda_i, & i=j\\ 0, & i\neq j \end{cases}, \ i,j=1,2,\dots,L.$$
(8)

It can also be checked from (5) that

$$\sum_{l=1}^{L} a_{x,l}^2(k) = \|\mathbf{x}(k)\|_2^2, \qquad (9)$$

where $\|\mathbf{x}(k)\|_2$ is the Euclidean norm of $\mathbf{x}(k)$. The previous expression shows the energy conservation through the KLE process.

One of the most important aspects of the KLE is its potential to reduce the dimensionality of the vector $\mathbf{x}(k)$ for low-rank signals. This idea has been extensively exploited in different manners for noise reduction where the signal of interest (speech) is assumed to be a low-rank signal [2], [3]. In what follows, we will take a different approach by working directly in the KLE domain.

Let us assume that the correlation matrix, \mathbf{R}_{vv} , of the noise is known or can be estimated from the noisy speech. Since the correlation matrix, \mathbf{R}_{yy} , of the noisy signal can be computed from the observations, an estimate of the correlation matrix \mathbf{R}_{xx} can be easily computed according to $\mathbf{R}_{xx} = \mathbf{R}_{yy} - \mathbf{R}_{vv}$. As a result, the orthogonal matrix \mathbf{Q} and diagonal matrix $\mathbf{\Lambda}$ can be determined. Now, a quick look at (5) tells us that in order to estimate the desired signal vector $\mathbf{x}(k)$ we only need to estimate the coefficients $a_{x,l}(k)$ since the eigenvectors \mathbf{q}_l are known. Left-multiplying (2) by \mathbf{q}_l^T , we get

$$a_{y,l}(k) = \mathbf{q}_l^T \mathbf{y}(k) = \mathbf{q}_l^T \mathbf{x}(k) + \mathbf{q}_l^T \mathbf{v}(k)$$

= $a_{x,l}(k) + a_{v,l}(k), \ l = 1, 2, \dots, L.$ (10)

Again, we see that

$$\sum_{l=1}^{L} a_{y,l}^{2}(k) = \|\mathbf{y}(k)\|_{2}^{2}, \quad \sum_{l=1}^{L} a_{v,l}^{2}(k) = \|\mathbf{v}(k)\|_{2}^{2}.$$
(11)

We also have (for i, j = 1, 2, ..., L)

$$E\left[a_{y,i}(k)a_{y,j}(k)\right] = \begin{cases} \lambda_i + \mathbf{q}_i^T \mathbf{R}_{vv} \mathbf{q}_i, & i = j \\ \mathbf{q}_i^T \mathbf{R}_{vv} \mathbf{q}_j, & i \neq j \end{cases}$$
(12)

Expression (10) is equivalent to (2) but in the KLE domain. In the rest of this paper, we assume that $|\mathbf{q}_i^T \mathbf{R}_{vv} \mathbf{q}_j| \ll \lambda_i + \mathbf{q}_i^T \mathbf{R}_{vv} \mathbf{q}_i$ or $|\mathbf{q}_i^T \mathbf{R}_{vv} \mathbf{q}_j| \approx 0$, for $i \neq j$, so that we can estimate the $a_{x,l}$, $l = 1, 2, \ldots, L$, independently of each others. Clearly, our problem this time is to find an estimate of $a_{x,l}(k)$ by passing $a_{y,l}(k)$ through a linear filter, i.e.,

$$a_{z,l}(k) = \mathbf{h}_l^T \mathbf{a}_{y,l}(k) = \mathbf{h}_l^T \left[\mathbf{a}_{x,l}(k) + \mathbf{a}_{v,l}(k) \right], \qquad (13)$$

where l = 1, 2, ..., L, $\mathbf{h}_l = [h_{l,0} \ h_{l,1} \ \cdots \ h_{l,L_l-1}]^T$ is a finite-impulse-response (FIR) filter of length L_l ,

 $\mathbf{a}_{y,l}(k) = \begin{bmatrix} a_{y,l}(k) & a_{y,l}(k-1) & \cdots & a_{y,l}(k-L_l+1) \end{bmatrix}^T$, and $\mathbf{a}_{x,l}(k)$ and $\mathbf{a}_{v,l}(k)$ are defined similarly to $\mathbf{a}_{y,l}(k)$. We see that the filters \mathbf{h}_l can take different lengths for the different spectral bands. Finally, an estimate of the vector $\mathbf{x}(k)$ would be

$$\mathbf{z}(k) = \sum_{l=1}^{L} a_{z,l}(k) \mathbf{q}_l.$$
 (14)

Later in this paper, we will show some filter design examples for noise reduction, but we first give some important definitions.

3. PERFORMANCE MEASURES

In this paper, we use the signal-to-noise ratio (SNR) and its improvement to evaluate the noise-reduction performance and the speechdistortion index to assess the amount of speech distortion due to the filter \mathbf{h}_l .

3.1 Signal-to-Noise Ratio

With the time-domain signal model given in (1), the SNR of the noisy signal can be written as,

$$\operatorname{SNR} = \frac{E\left[x^2(k)\right]}{E\left[v^2(k)\right]} = \frac{\sigma_x^2}{\sigma_v^2} = \frac{\sum_{l=1}^L \lambda_l}{\sum_{l=1}^L \mathbf{q}_l^T \mathbf{R}_{vv} \mathbf{q}_l}.$$
 (15)

In the KLE domain, it is also very useful to study the SNR in each spectral subband. With the KLE-domain model shown in (10), we define the subband input SNR as

$$\operatorname{SNR}_{l} = \frac{E\left[a_{x,l}^{2}(k)\right]}{E\left[a_{v,l}^{2}(k)\right]} = \frac{\lambda_{l}}{\mathbf{q}_{l}^{T}\mathbf{R}_{vv}\mathbf{q}_{l}}, \ l = 1, 2, \dots, L.$$
(16)

After noise reduction with the model given in (13), the subband output SNR is

$$\operatorname{oSNR}(\mathbf{h}_l) = \frac{\mathbf{h}_l^T \mathbf{R}_{a_x a_x, l} \mathbf{h}_l}{\mathbf{h}_l^T \mathbf{R}_{a_v a_v, l} \mathbf{h}_l}, \ l = 1, 2, \dots, L$$
(17)

and the fullband output SNR is

$$\operatorname{oSNR}(\mathbf{h}_{1:L}) = \frac{\sum_{l=1}^{L} \mathbf{h}_{l}^{T} \mathbf{R}_{a_{x}a_{x},l} \mathbf{h}_{l}}{\sum_{l=1}^{L} \mathbf{h}_{l}^{T} \mathbf{R}_{a_{v}a_{v},l} \mathbf{h}_{l}},$$
(18)

where $\mathbf{R}_{a_x a_x,l} = E[\mathbf{a}_{x,l}(k)\mathbf{a}_{x,l}^T(k)]$ and $\mathbf{R}_{a_v a_v,l} = E[\mathbf{a}_{v,l}(k)]$ $\mathbf{a}_{v,l}^T(k)]$ are the correlation matrices of the sequences $a_{x,l}(k)$ and $a_{v,l}(k)$, respectively.

It can be easily checked that $\sum_{l=1}^{L} \text{SNR}_l \geq \text{SNR}$, and $\sum_{l=1}^{L} \text{oSNR}(\mathbf{h}_l) \geq \text{oSNR}(\mathbf{h}_{1:L})$, which means that the aggregation of the subband SNRs is greater than or equal to the real fullband SNR.

3.2 Speech-Distortion Index

The speech-distortion index was introduced in [1], [5] to evaluate the amount of speech distortion. Here we extend the original definition to the model given in (13) and define the subband speechdistortion index as

$$\upsilon_{\rm sd}\left(\mathbf{h}_{l}\right) = \frac{E\left\{\left[a_{x,l}(k) - \mathbf{h}_{l}^{T}\mathbf{a}_{x,l}(k)\right]^{2}\right\}}{\lambda_{l}}.$$
(19)

At the subband l, the higher the value of $v_{sd}(\mathbf{h}_l)$, the more the speech distortion. The fullband speech-distortion index is

$$\nu_{\rm sd}\left(\mathbf{h}_{1:L}\right) = \frac{\sum_{l=1}^{L} E\left\{\left[a_{x,l}(k) - \mathbf{h}_{l}^{T} \mathbf{a}_{x,l}(k)\right]^{2}\right\}}{\sum_{l=1}^{L} \lambda_{l}}.$$
 (20)

4. OPTIMAL FILTERS IN THE KLE DOMAIN

In this section, we are going to derive two classes of optimal filters in the KLE domain depending on the length, L_l , of the filters \mathbf{h}_l .

4.1 Class I

ı

In this first category, we consider the particular case where $L_1 = L_2 = \cdots = L_L = 1$. Hence $\mathbf{h}_l = h_{l,0}$, $l = 1, 2, \ldots, L$, are simply scalars. For this class of filters, we always have $\operatorname{oSNR}(\mathbf{h}_l) = \operatorname{SNR}_l, \forall l$. Therefore, the subband SNR cannot be improved. But the fullband output SNR can be improved with respect to the input SNR. From the previous section we know that $\operatorname{oSNR}(\mathbf{h}_{l:L}) \leq \sum_{l=1}^L \operatorname{oSNR}(\mathbf{h}_l)$, which gives the upper bound of the output SNR.

4.1.1 Wiener Filter

Let us define the error signal between the clean speech and its estimate in the KLE domain

$$e_l(k) = a_{x,l}(k) - a_{z,l}(k) = a_{x,l}(k) - h_{l,0}a_{y,l}(k),$$
(21)

the corresponding mean-square error (MSE) can be written as

$$J(h_{l,0}) = E\left[e_l^2(k)\right], \ l = 1, 2, \dots, L.$$
(22)

The Wiener filter is easily found by taking the gradient of $J(h_{l,0})$ with respect to $h_{l,0}$ and equating the result to zero:

$$h_{\mathrm{W},l,0} = \frac{E\left[a_{x,l}^{2}(k)\right]}{E\left[a_{y,l}^{2}(k)\right]} = \frac{\lambda_{l}}{\lambda_{l} + \mathbf{q}_{l}^{T}\mathbf{R}_{vv}\mathbf{q}_{l}}$$
$$= \frac{\mathrm{SNR}_{l}}{1 + \mathrm{SNR}_{l}}, \ l = 1, 2, \dots, L.$$
(23)

This optimal filter is the equivalent form of the frequency-domain Wiener filter [6]. The estimator of the vector $\mathbf{x}(k)$ can be written as

$$\mathbf{z}_{\mathrm{KLE,W}}(k) = \sum_{l=1}^{L} h_{\mathrm{W},l,0} a_{y,l} \mathbf{q}_{l} = \mathbf{H}_{\mathrm{KLE,W}} \mathbf{y}(k), \qquad (24)$$

where $\mathbf{H}_{\text{KLE,W}} = \sum_{l=1}^{L} h_{\text{W},l,0} \mathbf{q}_l \mathbf{q}_l^T$ is the time-domain version of the KLE-domain filters $h_{\text{W},l,0}$ (l = 1, 2, ..., L). We easily find that

$$\mathbf{H}_{\mathrm{KLE,W}} = \mathbf{Q}\mathbf{\Lambda} \left[\mathbf{\Lambda} + \mathrm{diag}\left(\mathbf{Q}^{T}\mathbf{R}_{vv}\mathbf{Q}\right)\right]^{-1}\mathbf{Q}^{T}.$$
 (25)

Property 1. $oSNR(H_{KLE,W}) \ge SNR$. (The proof is not shown here due to space limitation.) Therefore, this method improves the (fullband) SNR.

4.1.2 Tradeoff Filter

The error signal defined in (21) can be rewritten as follows,

$$e_l(k) = e_{x,l}(k) - e_{v,l}(k), \ l = 1, 2, \dots, L,$$
 (26)

where $e_{x,l}(k) = (1 - h_{l,0})a_{x,l}(k)$ is the speech distortion due to the linear filter, and $e_{v,l}(k) = h_{l,0}a_{v,l}(k)$ represents the residual noise.

An important filter can be designed by minimizing the speech distortion with the constraint that the residual noise is smaller than a positive threshold level. This optimization problem can be translated mathematically as

$$\min_{h_{l,0}} J_x(h_{l,0}) \quad \text{subject to} \quad J_v(h_{l,0}) \le \beta \mathbf{q}_l^T \mathbf{R}_{vv} \mathbf{q}_l, \tag{27}$$

where $J_x(h_{l,0}) = E\left[e_{x,l}^2(k)\right]$, $J_v(h_{l,0}) = E\left[e_{v,l}^2(k)\right]$, $l = 1, 2, \ldots, L$, and $0 < \beta < 1$ in order to have some noise reduction. If we use a Lagrange multiplier, $\mu(\geq 0)$, to adjoin the constraint to the cost function, we easily find the optimal filter:

$$h_{\mathrm{T},l,0} = \frac{\lambda_l}{\lambda_l + \mu \mathbf{q}_l^T \mathbf{R}_{vv} \mathbf{q}_l}, \ l = 1, 2, \dots, L,$$
(28)

where the Lagrange multiplier satisfies $J_v(h_{l,0}) = \beta \mathbf{q}_l^T \mathbf{R}_{vv} \mathbf{q}_l$, l = 1, 2, ..., L. Hence, $h_{T,l,0}$ is a Wiener filter with adjustable input noise level $\mu \mathbf{q}_l^T \mathbf{R}_{vv} \mathbf{q}_l$. It can be shown that oSNR ($\mathbf{h}_{T,1:L}$) \geq SNR. Therefore, this method improves the (fullband) SNR.

Similar to the Wiener filter, the KLE-domain tradeoff filter can be written equivalently into the time domain:

$$\mathbf{H}_{\text{KLE,T}} = \mathbf{Q}\mathbf{\Lambda} \left[\mathbf{\Lambda} + \mu \text{diag}\left(\mathbf{Q}^T \mathbf{R}_{vv} \mathbf{Q}\right)\right]^{-1} \mathbf{Q}^T.$$
(29)

This optimal filter is the equivalent form of the subspace approach [2] and strictly equivalent to the filter proposed in [7].

Many other optimal and suboptimal filters such as the parametric Wiener filter can be easily developed.

4.2 Class II

Although they can improve the fullband SNR, the optimal filters derived in Class I have no impact on the subband SNR. In this subsection, we consider another category of filters \mathbf{h}_l with length $L_l > 1$. In this case, it is possible to improve both the subband and fullband SNRs at the same time.

4.2.1 Wiener Filter

From the MSE

$$J(\mathbf{h}_l) = E\left\{\left[a_{x,l}(k) - \mathbf{h}_l^T \mathbf{a}_{y,l}(k)\right]^2\right\}, \ l = 1, 2, \dots, L, \quad (30)$$

we easily deduce the KLE-domain Wiener filter:

$$\mathbf{h}_{\mathrm{W},l} = \mathbf{R}_{a_y a_y,l}^{-1} \mathbf{R}_{a_x a_x,l} \mathbf{u}_l = \left(\mathbf{I}_l - \mathbf{R}_{a_y a_y,l}^{-1} \mathbf{R}_{a_v a_v,l} \right) \mathbf{u}_l, \quad (31)$$

where $\mathbf{R}_{a_y a_y,l}$ is the covariance matrix of the signal $a_{y,l}(k)$, $\mathbf{u}_l = [1 \ 0 \ \cdots \ 0]^T$ is a vector of length L_l , and \mathbf{I}_l is the identity matrix of size $L_l \times L_l$.

Property 2. With the optimal KLE-domain Wiener filter given in (31), the subband output SNR is always greater than or equal to the subband input SNR, i.e., $oSNR(\mathbf{h}_{W,l}) \ge SNR_l, \forall l$.

Proof. Let us evaluate the squared Pearson correlation coefficient (SPCC) between $a_{y,l}(k)$ and $\mathbf{h}_{W,l}^{T} \mathbf{a}_{y,l}(k)$, which is defined as

$$\rho^{2}\left(a_{y,l}, \mathbf{h}_{\mathrm{W},l}^{T} \mathbf{a}_{y,l}\right) = \frac{E^{2}\left[a_{y,l} \cdot \mathbf{h}_{\mathrm{W},l}^{T} \mathbf{a}_{y,l}\right]}{E[a_{y,l}^{2}] \cdot E\left[(\mathbf{h}_{\mathrm{W},l}^{T} \mathbf{a}_{y,l})^{2}\right]}.$$
(32)

2

With some simple manipulation, we obtain

$$\rho^{2}\left(a_{y,l}, \mathbf{h}_{W,l}^{T} \mathbf{a}_{y,l}\right) = \frac{\left(\mathbf{u}_{l}^{T} \mathbf{R}_{a_{y}a_{y},l} \mathbf{h}_{W,l}\right)^{2}}{\left(\lambda_{l} + \mathbf{q}_{l}^{T} \mathbf{R}_{vv} \mathbf{q}_{l}\right) \left(\mathbf{h}_{W,l}^{T} \mathbf{R}_{a_{y}a_{y},l} \mathbf{h}_{W,l}\right)}$$
$$= \frac{\lambda_{l}}{\lambda_{l} + \mathbf{q}_{l}^{T} \mathbf{R}_{vv} \mathbf{q}_{l}} \cdot \frac{\lambda_{l}}{\mathbf{u}_{l}^{T} \mathbf{R}_{a_{x}a_{x},l} \mathbf{h}_{W,l}} = \frac{\rho^{2}\left(a_{x,l}, a_{y,l}\right)}{\rho^{2}\left(a_{x,l}, \mathbf{h}_{W,l}^{T} \mathbf{a}_{y,l}\right)}.$$
(33)

Therefore $\rho^2(a_{x,l}, a_{y,l}) = \rho^2(a_{y,l}, \mathbf{h}_{W,l}^T \mathbf{a}_{y,l}) \cdot \rho^2(a_{x,l}, \mathbf{h}_{W,l}^T \mathbf{a}_{y,l}) \le \rho^2(a_{x,l}, \mathbf{h}_{W,l}^T \mathbf{a}_{y,l})$. It is easy to check that

$$\rho^2(a_{x,l}, a_{y,l}) = \frac{\mathrm{SNR}_l}{1 + \mathrm{SNR}_l}, \tag{34}$$

$$\rho^{2} \left(\mathbf{h}_{\mathrm{W},l}^{T} \mathbf{a}_{x,l}, \mathbf{h}_{\mathrm{W},l}^{T} \mathbf{a}_{y,l} \right) = \frac{\mathrm{oSNR} \left(\mathbf{h}_{\mathrm{W},l} \right)}{1 + \mathrm{oSNR} \left(\mathbf{h}_{\mathrm{W},l} \right)}, \quad (35)$$

and

$$\rho^{2}\left(a_{x,l},\mathbf{h}_{\mathrm{W},l}^{T}\mathbf{a}_{y,l}\right) = \rho^{2}\left(a_{x,l},\mathbf{h}_{\mathrm{W},l}^{T}\mathbf{a}_{x,l}\right) \cdot \rho^{2}\left(\mathbf{h}_{\mathrm{W},l}^{T}\mathbf{a}_{x,l},\mathbf{h}_{\mathrm{W},l}^{T}\mathbf{a}_{y,l}\right)$$
$$\leq \rho^{2}\left(\mathbf{h}_{\mathrm{W},l}^{T}\mathbf{a}_{x,l},\mathbf{h}_{\mathrm{W},l}^{T}\mathbf{a}_{y,l}\right).$$

Hence

$$\frac{\mathrm{SNR}_{l}}{1 + \mathrm{SNR}_{l}} \leq \frac{\mathrm{oSNR}\left(\mathbf{h}_{\mathrm{W},l}\right)}{1 + \mathrm{oSNR}\left(\mathbf{h}_{\mathrm{W},l}\right)}.$$
(36)

It is immediately clear that

$$\operatorname{oSNR}(\mathbf{h}_{\mathrm{W},l}) \ge \operatorname{SNR}_{l}, \ l = 1, 2, \dots, L,$$
(37)

which completes the proof.

Property 3. With the optimal KLE-domain Wiener filter given in (31), the fullband output SNR is always greater than or equal to the input SNR, i.e., $oSNR(\mathbf{h}_{W,1:L}) \ge SNR$.

Proof. The proof is similar to the previous one by using a more general version of the SPCC.

4.2.2 Tradeoff Filter

The filter for this approach is obtained by solving the following optimization problem:

$$\min_{\mathbf{h}_l} J_x(\mathbf{h}_l) \text{ subject to } J_v(\mathbf{h}_l) \le \beta \mathbf{q}_l^T \mathbf{R}_{vv} \mathbf{q}_l, \ l = 1, 2, \dots, L,$$
(38)

where

$$J_x(\mathbf{h}_l) = E\left\{\left[\left(\mathbf{u}_l - \mathbf{h}_l\right)^T \mathbf{a}_{x,l}\right]^2\right\},\tag{39}$$

$$J_{v}(\mathbf{h}_{l}) = E\left[\left(\mathbf{h}_{l}^{T}\mathbf{a}_{v,l}\right)^{2}\right],\tag{40}$$

and $0 < \beta < 1$ in order to have some noise reduction. If we use a Lagrange multiplier, $\mu (\geq 0)$, to adjoin the constraint to the cost function, we easily find the optimal filter:

$$\mathbf{h}_{\mathrm{T},l} = (\mathbf{R}_{a_x a_x, l} + \mu \mathbf{R}_{a_v a_v, l})^{-1} \mathbf{R}_{a_x a_x, l} \mathbf{u}_l$$
$$= \left[\mathbf{I}_l + (\mu - 1) \mathbf{R}_{a_y a_y, l}^{-1} \mathbf{R}_{a_v a_v, l} \right]^{-1} \mathbf{h}_{\mathrm{W}, l}, \tag{41}$$

where the Lagrange multiplier satisfies $J_v(\mathbf{h}_l) = \beta \mathbf{q}_l^T \mathbf{R}_{vv} \mathbf{q}_l$, l = 1, 2, ..., L. In practice it is not easy to determine μ . Therefore, when this parameter is chosen in an ad-hoc way, we have four cases. 1) $\mu = 1$: in this case, $\mathbf{h}_{T,l} = \mathbf{h}_{W,l}$. 2) $\mu = 0$: in this circumstance, $\mathbf{h}_{T,l} = \mathbf{u}_l$. With this filter, there will be no speech distortion, but no noise reduction as well. 3) $\mu > 1$: this corresponds to more aggressive noise reduction (compared with Wiener). So the residual noise level would be lower, but it is achieved at the expense of higher speech distortion. 4) $\mu < 1$: this corresponds to less aggressive noise reduction (compared with Wiener). In this situation, we get less speech distortion but not so much noise reduction.



Fig. 1. Class I filters: performance versus L in white Gaussian noise with: SNR = 10 dB, $\alpha_y = 0.985$, and $\alpha_v = 0.995$.



Fig. 2. Class II filters: performance versus L_l in white Gaussian noise with: SNR = 10 dB, $\alpha_{a_{y}} = 0.8$, $\alpha_{a_{y}} = 0.91$, and L = 20.

5. EXPERIMENTS

We present two sets of experimental results to demonstrate the performance of the developed KLE-domain noise-reduction filters.

The speech signal used in our experiments was recorded from a female speaker in a quiet office environment. It was sampled at 8 kHz. The overall length of the signal is 30 s. The noise used is a computer generated white Gaussian random process. The noisy speech is obtained by adding noise to the clean speech.

To implement the optimal noise-reduction filters developed in Section 4, we need to know the statistics of both the noisy and noise signals. Specifically, the Class I filters require the knowledge of the covariance matrices \mathbf{R}_{yy} and \mathbf{R}_{vv} , while the Class II filters need to know the matrices $\mathbf{R}_{a_y a_y, l}$ and $\mathbf{R}_{a_v a_v, l}$ in addition to \mathbf{R}_{yy} and \mathbf{R}_{vv} . In this study, for the Class I filters, we estimate \mathbf{R}_{yy} through the widely used recursive approach,

$$\mathbf{R}_{yy}(k) = \alpha_y \mathbf{R}_{yy}(k-1) + (1-\alpha_y)\mathbf{y}(k)\mathbf{y}^T(k), \qquad (42)$$

where α_{y} is a forgetting factor. The matrix \mathbf{R}_{vv} is directly computed from the noisy signal (to avoid the influence of the noise estimation error on the parameter optimization) using the same recursive method as in (42) but with a different forgetting factor α_v . We have performed a set of experiments to study the impact of the forgetting factors α_{y} and α_{v} on the noise-reduction performance. It turned out that the best performance was obtained when $\alpha_v = 0.995$,

 $\alpha_y = 0.985$. Another important parameter in Class I is the filter length L. Figure 1 plots the performance as a function of filter length L. As seen, good performance for all the studied algorithms is achieved when the filter length L is around 20.

Unlike the Class I filters where each frame may have a different transformation Q, the Class II algorithms assume that all the frames share the same Q transformation. So, the estimation of Q is relatively easier than that for the Class I case. We simply use the longterm sample average to compute the covariance matrices \mathbf{R}_{uy} and \mathbf{R}_{vv} , thereby obtaining an estimate of \mathbf{R}_{xx} and \mathbf{Q} . This \mathbf{Q} matrix is then applied to each frame of the signal to compute the KLE coefficients $a_{y,l}$, and $a_{v,l}$. We then estimate the matrix $\mathbf{R}_{a_y a_y,l}$ using the recursive method similar to (42), i.e., 1)

$$\mathbf{R}_{a_y a_y, l}(k) = \alpha_{a_y, l} \mathbf{R}_{a_y a_y, l}(k - k)$$

$$+(1-\alpha_{a_y,l})\mathbf{a}_{y,l}(k)\mathbf{a}_{y,l}^T(k),\qquad(43)$$

where $\alpha_{a_{y},l}$, same as α_{y} in (42), is a forgetting factor. The $\mathbf{R}_{a_{y}a_{y},l}$ matrix is estimated using the same recursion given in (43), but with a different forgetting factor $\alpha_{a_v,l}$. Experiments were conducted to study the impact of $\alpha_{a_v,l}$ and $\alpha_{a_u,l}$ on the noise-reduction performance, and good performance was achieved when $\alpha_{a_v,l} = 0.91$, $\alpha_{a_y,l} = 0.8, l = 1, \dots, L.$ Another important parameter in Class II is the filter length L_l . Figure 2 shows the noise-reduction performance as a function of the filter length L_l . It is seen that as L_l increases, the output SNR increases first to its maximum, and then decreases slightly. In comparison, the speech distortion index with both methods increases monotonically with L_l . Taking into account of both SNR improvement and speech distortion, we would suggest to use L_l between 5 and 10.

Comparing Figs. 2 and 1, one can see that, with the same L, the optimal filters in Class II can achieve much higher SNR gain than the filters of Class I. The Class II filters also have slightly more speech distortion. But the additional amount of distortion compared to that of the Class I filters is not significant when $L_l \leq 10$. This indicates that the Class II filters may have a great potential in practice.

6. CONCLUSIONS

In this paper, we have studied the noise-reduction problem in the Karhunen-Loève expansion domain. We have discussed two classes of optimal noise-reduction filters in that domain. While the first class of filters estimates a frame of speech by filtering only the corresponding frame of the noisy speech, the second class of filters are inter-frame techniques, which obtain noise reduction by filtering not only the current frame, but also a number of previous consecutive frames of the noisy speech. Through experiments, we demonstrated that better noise reduction performance can be achieved with the Class II filters when the parameters associated with this class are properly chosen, which demonstrated the great potential of the filters in this category for noise reduction.

7. REFERENCES

- J. Benesty, S. Makino, and J. Chen, Eds., Speech Enhancement. Berlin, Germany: [1] Springer-Verlag, 2005
- Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhance-ment," *IEEE Trans. Speech Audio Process.*, vol. 3, pp. 251–266, July 1995. [2]
- K. Hermus, P. Wambacq, and H. Van hamme, "A review of signal subspace speech enhancement and its application to noise robust speech recognition," EURASIP J. Appl. Signal Process., vol. 2007, Article ID 45821, 15 pages, 2007.
- S. Haykin, Adaptive Filter Theory. Fourth Edition, Upper Saddle River, NJ: [4] Prentice-Hall, 2002
- J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction [5] Wiener filter," IEEE Trans. Audio, Speech, Lang. Process., vol. 14, pp. 1218-1234, July 2006.
- J. Benesty, J. Chen, and Y. Huang, Microphone Array Signal Processing. Berlin, [6] Germany: Springer-Verlag, 2008.
- A. Rezayee and S. Gazor, "An adpative KLT approach for speech enhancement," IEEE Trans. Speech Audio Process., vol. 9, pp. 87-95, Feb. 2001.