Study of the Noise-Reduction Problem in the Karhunen–Loève Expansion Domain

Jingdong Chen, Member, IEEE, Jacob Benesty, Senior Member, IEEE, and Yiteng (Arden) Huang, Member, IEEE

Abstract—Noise reduction, which aims at estimating a clean speech from a noisy observation, has long been an active research area. The standard approach to this problem is to obtain the clean speech estimate by linearly filtering the noisy signal. The core issue, then, becomes how to design an optimal linear filter that can significantly suppress noise without introducing perceptually noticeable speech distortion. Traditionally, the optimal noise-reduction filters are formulated in either the time or the frequency domains. This paper studies the problem in the Karhunen–Loève expansion domain. We develop two classes of optimal filters. The first class achieves a frame of speech estimate by filtering the corresponding frame of the noisy speech. We will show that many existing methods such as the widely used Wiener filter and subspace technique are closely related to this category. The second class 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, through both theoretical analysis and experiments, the properties of the deduced optimal filters.

Index Terms—Karhunen–Loève expansion (KLE), maximum signal-to-noise ratio (SNR) filter, noise reduction, Pearson correlation coefficient, speech enhancement, subspace approach, Wiener filter.

I. INTRODUCTION

In practice, speech signals can seldom be recorded and processed in pure form and they are generally contaminated by background noise originating from various noise sources. Noise contamination can dramatically change the characteristics of speech signals and degrade speech quality and intelligibility, thereby causing significant harm to human-to-human and human-to-machine communication systems. As a result, digital signal processing techniques have to be developed to “clean” the noisy speech before it is stored, transmitted, processed, or played out. This problem, often referred to as either noise reduction or speech enhancement, has been a major challenge for many researchers and engineers for decades.

Mathematically, the microphone signal can be modeled as a superposition of the clean speech and noise. With this signal model, a normal practice for reducing noise is to pass the microphone signal through a filter/transformation. Usually, we only consider linear filters/transformations since the nonlinear ones are much more difficult to design and analyze. So, the problem of noise reduction becomes one of finding an optimal linear filter/transformation such that, after the filtering process, the signal-to-noise ratio (SNR) can be improved, or in other words, the processed signal would become cleaner. However, since the filtering operation will not only attenuate the noise, but also affect the speech signal, careful attention has to be paid to the speech distortion while deriving the optimal filter.

Traditionally, the optimal noise-reduction filters/transformations are considered in either the time or the frequency domains. In the time domain, the optimal filters/transformations are often obtained by minimizing the mean-square error (MSE) between the clean speech and its estimate. These approaches can be sample based in the sense that they make an estimate of one speech sample at a time [1]–[4]. They can also be frame based, applying a transformation matrix to a frame of the noisy speech to produce an estimate of a frame of the clean speech [5]–[15]. In comparison, the frequency-domain methods are often formulated on a frame basis where a block of the noisy speech signal is transformed into the frequency domain using the discrete Fourier transform (DFT); a gain filter is then estimated and applied to filter the frame spectrum; and the filtered spectrum is finally converted back into the time domain using the inverse DFT (IDFT), thereby producing a block of clean speech estimate [16]–[29]. Both the time- and frequency-domain algorithms have their own advantages and drawbacks. In general, the frequency-domain algorithms have more flexibility in controlling the noise-reduction performance versus speech distortion since the gain filter is estimated and operated independently in each subband. However, special attention has to be paid to the aliasing distortion as well as to other artifacts such as the musical residual noise. In comparison, the time-domain formulation does not have aliasing problems and the resulting filters are usually causal, but they are less flexible in terms of performance management and computational complexity.

In this paper, we formulate the noise-reduction problem in the Karhunen–Loève expansion (KLE) domain. Similar to the frequency-domain approaches, this new formulation achieves noise reduction on a subband basis. It first transforms a block of the noisy speech into the KLE domain. An optimal (or suboptimal for a better compromise between noise reduction and speech distortion) filter is then estimated and applied to the KLE coefficients in each subband (here the term subband refers to the
signal component along each base vector of the KLE). The filtered KLE coefficients are finally transformed back to the signal original (time) domain, giving an estimate of a frame of the clean speech. There are many differences between this new approach and the frequency-domain methods. The major one is that this new method employs the Karhunen–Loève transform (KLT) while the frequency-domain technique uses the DFT. Since the KLT can exactly diagonalize the signal correlation matrix, the signal components from different subbands in this new formulation are uncorrelated and can be processed independently. In comparison, the Fourier matrix can only approximately diagonalize the noisy covariance matrix (since this matrix is Toeplitz and its elements are usually absolutely summable [30]). This approximation may cause much distortion to the clean speech when noise reduction is performed separately in each subband. Note that the KLT has been used in the well-known subspace method [5]–[15]. The difference between the subspace method and our new formulation is that the former achieves noise reduction by diagonalizing the noisy covariance matrix, removing the noise eigenvalues, and cleaning the signal-plus-noise eigenvalues, but our new formulation approaches noise reduction by diagonalizing an estimate of the clean speech correlation matrix and estimating the KLE coefficients of the clean speech in the KLE domain via a filtering process.

We will address how to design the optimal and suboptimal filters in the KLE domain. Particularly, we will discuss two classes of filters. The first class achieves a frame of speech estimate by filtering the corresponding frame of the noisy speech. We will show the close relationship between this class of optimal filters and many existing methods such as the widely used Wiener filter and subspace technique. The second category does noise reduction by filtering not only the current frame, but also a number of previous consecutive frames of the noisy speech components. We will demonstrate that, when the algorithmic parameters are properly chosen, the optimal filters in the second class can achieve better noise-reduction performance than those in the first category.

II. PROBLEM FORMULATION

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

\[
y(k) = x(k) + v(k)
\]

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 \(x(k)\).

The signal model given in (1) can be written in a vector form if we process the data on a frame-by-frame basis

\[
y(k) = x(k) + v(k)
\]

where

\[
y(k) = [y(k) \ y(k-1) \ \cdots \ y(k-L+1)]^T
\]

superscript \(^T\) denotes transpose of a vector or a matrix. \(L\) is the frame length, and \(x(k)\) and \(v(k)\) are defined in a similar way to \(y(k)\). 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.,

\[
R_{yy}(k) = R_{xx}(k) + R_{vv}(k)
\]

where

\[
R_{yy}(k) \triangleq E[y(k)y^T(k)] \quad (4a)
\]

\[
R_{xx}(k) \triangleq E[x(k)x^T(k)] \quad (4b)
\]

\[
R_{vv}(k) \triangleq E[v(k)v^T(k)] \quad (4c)
\]

are, respectively, the correlation (also covariance since \(y(k), x(k), \text{and } v(k)\) are assumed to be zero mean) matrices of the signals \(y(k), x(k), \text{and } v(k)\) at time instant \(k\), and \(E[\cdot]\) denotes mathematical expectation. Note that the correlation matrices for nonstationary speech signals are in general time-varying, and hence a time index is used here, but for convenience and exposition simplicity, in the rest of this paper we will drop the time index \(k\) and assume that all signals are quasi-stationary (meaning that their statistics stay the same within a frame, but can change over frames). With this vector form of signal model, the noise-reduction problem becomes one of estimating \(x(k)\) from the observation vector \(y(k)\).

In this paper, we will mainly use the signal model given in (2) and focus on estimating \(x(k)\) [estimating \(x(k)\) can be viewed as a special case of estimating \(x(k)\)]. Generally, \(x(k)\) can be estimated by applying a linear transformation to \(y(k)\) [3]–[15], i.e.,

\[
z(k) = H y(k) = H[x(k) + v(k)] = x(k) + v_{\text{res}}(k)
\]

where \(H\) is a filtering matrix of size \(L \times L\), and \(x(k) = Hx(k)\) and \(v_{\text{res}}(k) = Hv(k)\) are, respectively, the filtered speech and residual noise after noise reduction. With this time-domain formulation, the noise-reduction problem becomes one of finding an optimal \(H\) that would attenuate the noise as much as possible while keeping the clean speech from being dramatically distorted. One of the most used algorithms for noise reduction is the classical Wiener filter derived from the MSE criterion

\[
E[(x(k) - z(k))^T (x(k) - z(k))] 
\]

This optimal filter is

\[
H_W = R_{xx} R_{yy}^{-1}
\]

and most of the existing noise-reduction filters, in either the time or the frequency domains, are related to this one in one way or another, as will be shown later on.

III. 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 \(x(k)\) denote a data sequence drawn from a zero-mean stationary process with the correlation matrix \(R_{xx}\). This matrix can be diagonalized as follows [31]:

\[
Q^T R_{xx} Q = \Lambda
\]
where
\[ Q = [q_1 \, q_2 \, \cdots \, q_L] \]
and
\[ \Lambda = \text{diag}[\lambda_1 \, \lambda_2 \, \cdots \, \lambda_L] \]
are, respectively, orthogonal and diagonal matrices. The orthonormal vectors \( q_1, q_2, \ldots, q_L \) are the eigenvectors corresponding, respectively, to the eigenvalues \( \lambda_1, \lambda_2, \ldots, \lambda_L \) of the matrix \( R_{xx} \).

The vector \( x(k) \) can be written as a combination (expansion) of the eigenvectors of the correlation matrix \( R_{xx} \) as follows:
\[ x(k) = \sum_{l=1}^{L} a_{x,l}(k) q_l \]  
(8)

where
\[ a_{x,l}(k) = q_l^T x(k), \quad l = 1, 2, \ldots, L \]  
(9)
are the coefficients of the expansion. The representation of the random vector \( x(k) \) described by (8) and (9) is the KLE where (8) is the synthesis part and (9) represents the analysis part [31].

From (9), we can verify that
\[ E[a_{x,l}(k)] = 0, \quad l = 1, 2, \ldots, L \]  
(10)
and
\[ E[a_{x,i}(k) a_{x,j}(k)] = \begin{cases} \lambda_i, & i = j \\ 0, & i \neq j \end{cases} \quad i, j = 1, 2, \ldots, L. \]  
(11)

It can also be checked from (9) that
\[ \sum_{l=1}^{L} a_{x,l}^2(k) = ||x(k)||_2^2 \]  
(12)
where \( ||x(k)||_2 \) is the Euclidean norm of \( x(k) \). The previous expression shows the energy conservation through the KLE process.

The KLE is originally introduced for analyzing stationary signals, but we will extend its use in this study to processing non-stationary signals like speech. So, in our context, a \( Q \) matrix will be estimated at time \( k \) by diagonalizing the correlation matrix \( R_{xx} \). The KLE expression for nonstationary speech may look the same as that for stationary signals. However, it should be easy to tell the difference from the context.

One of the most important aspects of the KLE is its potential to reduce the dimensionality of the vector \( x(k) \) for low-rank signals. This idea has been extensively exploited, by way of sub-space separation and cleaning, for noise reduction where the signal of interest (speech) is assumed to be a low-rank signal [5]--[15]. In the following, we will take a approach that is different from what used in the subspace method [5], [14]. Instead of manipulating the eigenvalues of the noisy correlation matrix, we will attempt to estimate the KLE coefficients of the clean speech by filtering the KLE coefficients of the noisy speech.

Let us assume that the correlation matrix \( R_{xx} \) of the noise is known or can be estimated from the noisy speech. Since the correlation matrix \( R_{yy} \) of the noisy signal can be estimated from the observations, then an estimate of the correlation matrix \( R_{xx} \) can be computed according to \( R_{xx} = R_{yy} - R_{ev} \). As a result, the orthogonal matrix \( Q \) and diagonal matrix \( \Lambda \) can be determined. Now, a quick look at (8) tells us that in order to estimate the desired signal vector \( x(k) \) we only need to estimate the coefficients \( a_{x,l}(k) \) since the eigenvectors \( q_l \) \( l = 1, 2, \ldots, L \) are known. Substituting (2) into (9), we get
\[ a_{y,l}(k) = q_l^T y(k) = q_l^T x(k) + q_l^T v(k) \]
\[ = a_{x,l}(k) + a_{v,l}(k), \quad l = 1, 2, \ldots, L. \]  
(13)
Again, we see that
\[ \sum_{l=1}^{L} a_{p,q,l}^2(k) = ||y(k)||_2^2, \]  
(14)
\[ \sum_{l=1}^{L} a_{v,l}^2(k) = ||v(k)||_2^2. \]  
(15)

We also have
\[ E[a_{y,i}(k) a_{y,j}(k)] = \begin{cases} \lambda_i + q_l^T R_{ev} q_i, & i = j \\ q_l^T R_{ev} q_i, & i \neq j \end{cases} \quad i, j = 1, 2, \ldots, L. \]  
(16)
Expression (13) is equivalent to (2) but in the KLE domain. In the rest of this paper, we assume that \( q_l^T R_{ev} q_i \approx q_l^T R_{ev} q_j \) or \( q_l^T R_{ev} q_i \approx 0 \), for \( i \neq j \) (if the noise is white, \( q_l^T R_{ev} q_i = 0 \), \( \forall i \neq j \)). In this case, both the speech and noise KLE coefficients in one subband are uncorrelated with those from all the other subbands. As a result, we can estimate \( a_{x,l} \) from the KLE coefficients of the noisy speech in the \( l \)th subband without need to consider signal components from all the other \( L - 1 \) subbands. So, 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.,

\[ a_{x,l}(k) = h_l^T a_{y,l}(k) \]
\[ = h_l^T [a_{x,l}(k) + a_{v,l}(k)], \quad l = 1, 2, \ldots, L \]  
(17)
where
\[ h_l \triangleq [h_{t,0} \, h_{t,1} \, \cdots \, h_{t,L-1}]^T \]
is a finite-impulse-response (FIR) filter of length \( L_t \), and
\[ a_{y,l}(k) \triangleq [a_{y,l}(k) \, a_{y,l}(k-1) \, \cdots \, a_{y,l}(k-L_t+1)]^T \]  
(18a)
\[ a_{x,l}(k) \triangleq [a_{x,l}(k) \, a_{x,l}(k-1) \, \cdots \, a_{x,l}(k-L_t+1)]^T \]  
(18b)
\[ a_{v,l}(k) \triangleq [a_{v,l}(k) \, a_{v,l}(k-1) \, \cdots \, a_{v,l}(k-L_t+1)]^T. \]  
(18c)
We see that the filters \( h_l, l = 1, 2, \ldots, L \), can take different lengths in the different subbands. Finally, an estimate of the vector \( x(k) \) would be

\[
\hat{x}(k) = \sum_{l=1}^{L} a_{x,l}(k) q_l. \tag{19}
\]

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

IV. PERFORMANCE MEASURES

In this section, we present some very useful measures that are necessary for properly designing the filters \( \mathbf{H} \) and \( h_l \). These definitions will also help us better understand how noise reduction works in the KLE domain.

The most important measure in noise reduction is the signal-to-noise ratio (SNR). With the time-domain signal model given in (1), the input SNR is defined as the ratio of the intensity of the signal of interest over the intensity of the background noise, i.e.,

\[
\text{SNR} = \frac{\sigma^2_s}{\sigma^2_n} = \frac{\sum_{l=1}^{L} \lambda_l}{\sum_{l=1}^{L} \frac{1}{q_l^2} \mathbf{R}_{v,n} q_l}, \tag{20}
\]

where \( \sigma^2_s = \mathbb{E}[x^2(k)] \) and \( \sigma^2_n = \mathbb{E}[n^2(k)] \) are the variances of the signals \( x(k) \) and \( n(k) \), respectively.

After noise reduction with the model given in (5), the output SNR can be written as

\[
\text{oSNR}(\mathbf{H}) = \frac{\sigma^2_s}{\sigma^2_n} = \frac{\text{tr}(\mathbb{E}[x_t x_t^T])}{\text{tr}(\mathbb{E}[v_t v_t^T])} = \frac{\text{tr}(\mathbf{H} \mathbf{R}_{x,e} \mathbf{H}^T)}{\text{tr}(\mathbf{H} \mathbf{R}_{v,n} \mathbf{H}^T)}, \tag{21}
\]

where \( \text{tr}(\cdot) \) denotes the trace of a matrix. One of the most important goals of noise reduction is to improve the SNR after filtering [1]. Therefore, we must design a filter, \( \mathbf{H} \), in such a way that \( \text{oSNR}(\mathbf{H}) \geq \text{SNR} \). For example, with the time-domain Wiener filter, \( \mathbf{H}_W \), it was shown that \( \text{oSNR}(\mathbf{H}_W) \geq \text{SNR}, \forall \text{SNR} \) [2], [3], [25], [32], [33].

In the KLE domain, it is also very useful to study the SNR in each subband. With the KLE-domain model shown in (13), we define the subband input SNR as

\[
\text{SNR}_l = \frac{\mathbb{E}[a_{x,l}(k)]}{\mathbb{E}[a_{v,l}(k)]} = \frac{\lambda_l}{q_l^2 \mathbf{R}_{v,n} q_l}, \quad l = 1, 2, \ldots, L. \tag{22}
\]

After noise reduction with the model given in (17), the subband output SNR is

\[
\text{oSNR}(h_l) = \frac{h_l^T \mathbf{R}_{a_{x,l} a_{x,l}} h_l}{h_l^T \mathbf{R}_{a_{v,l} a_{v,l}} h_l}, \quad l = 1, 2, \ldots, L \tag{23}
\]

and the fullband output SNR is

\[
\text{oSNR}(h_{1:L}) = \frac{\sum_{l=1}^{L} h_l^T \mathbf{R}_{a_{x,l} a_{x,l}} h_l}{\sum_{l=1}^{L} h_l^T \mathbf{R}_{a_{v,l} a_{v,l}} h_l}. \tag{24}
\]

where

\[
\mathbf{R}_{a_{x,l} a_{x,l}} \triangleq \mathbb{E}[a_{x,l}(k) a_{x,l}^T(k)] \tag{25a}
\]

\[
\mathbf{R}_{a_{v,l} a_{v,l}} \triangleq \mathbb{E}[a_{v,l}(k) a_{v,l}^T(k)] \tag{25b}
\]

are the correlation matrices of the sequences \( a_{x,l}(k) \) and \( a_{v,l}(k) \), respectively.

It can be checked that

\[
\sum_{l=1}^{L} \text{SNR}_l \geq \text{SNR} \tag{26}
\]

\[
\sum_{l=1}^{L} \text{oSNR}(h_l) \geq \text{oSNR}(h_{1:L}). \tag{27}
\]

This means that the aggregation of the subband SNRs is greater than or equal to the real fullband SNR. The proof of (26) and (27) can be shown by using the following inequality:

\[
\sum_{l=1}^{L} \frac{c_l}{d_l} \leq \frac{\sum_{l=1}^{L} c_l}{\sum_{l=1}^{L} d_l} \leq \sum_{l=1}^{L} \frac{c_l}{d_l} \tag{28}
\]

where \( c_l \) and \( d_l \) are two positive series.

Another important measure in noise reduction is the noise-reduction factor, which quantifies the amount of noise being attenuated by the noise-reduction filter. With the time-domain formulation, this factor is defined as [1], [2]

\[
\xi_{\text{tnr}}(\mathbf{H}) \triangleq \frac{\sigma^2_n}{\sigma^2_n} = \frac{\text{tr}(\mathbf{R}_{v,n})}{\text{tr}(\mathbf{H} \mathbf{R}_{v,n} \mathbf{H}^T)}. \tag{29}
\]

By analogy to the above time-domain definition, we define the subband noise-reduction factor as

\[
\xi_{\text{tnr}}(h_l) \triangleq \frac{q_l^2 \mathbf{R}_{v,n} q_l}{h_l^T \mathbf{R}_{a_{v,l} a_{v,l}} h_l}, \quad l = 1, 2, \ldots, L \tag{30}
\]

The larger the value of \( \xi_{\text{tnr}}(h_l) \), the more the noise is reduced at the subband \( l \). After the filtering operation, the residual noise level at the subband \( l \) is expected to be lower than that of the original noise level, and therefore this factor should have a lower bound of 1. The fullband noise-reduction factor is

\[
\xi_{\text{tnr}}(h_{1:L}) = \frac{L \sigma^2_n}{\sum_{l=1}^{L} h_l^T \mathbf{R}_{a_{v,l} a_{v,l}} h_l}. \tag{31}
\]

The filtering operation adds distortion to the speech signal. In order to evaluate the amount of speech distortion, the concept of speech-distortion index has been introduced [1], [2]. With the time-domain model, the speech-distortion index is defined as

\[
\upsilon_{\text{sd}}(\mathbf{H}) \triangleq \frac{\mathbb{E}[(x(k) - x_t(k))^T(x(k) - x_t(k))]}{\text{tr}(\mathbf{R}_{x,x})} = \frac{\mathbb{E}[(x(k) - \mathbf{H} x(k))^T(x(k) - \mathbf{H} x(k))]}{\text{tr}(\mathbf{R}_{x,x})}. \tag{32}
\]

Extending this definition to the model given in (17), we introduce the subband speech-distortion index as

\[
\upsilon_{\text{sd}}(h_l) \triangleq \frac{\mathbb{E}[[a_{x,l}(k) - h_l^T a_{x,l}(k)]^2]}{\lambda_l}. \tag{33}
\]
This index has a lower bound of 0 and should have an upper bound of 1 for optimal filters. The higher the value of \( v_{sd}(h_L) \), the more the speech distortion. The fullband speech-distortion index is

\[
v_{sd}(h_{1:L}) \triangleq \frac{\sum_{l=1}^{L} E \left\{ \left[ a_{a,x}(k) - h_{L}^T a_{a,x}(k) \right]^2 \right\}}{\sum_{l=1}^{L} \lambda_l}.
\] (34)

We always have

\[
\sum_{l=1}^{L} \xi_{in}(h_L) \geq \xi_{in}(h_{1:L}) \tag{35}
\]

\[
\sum_{l=1}^{L} v_{sd}(h_L) \geq v_{sd}(h_{1:L}). \tag{36}
\]

Although there are many more measures available in the literature, the aforementioned ones (input and output SNRs, noise-reduction factors, and speech-distortion indices) will be primarily used to study, evaluate, and derive optimal or suboptimal filters for noise reduction in the following sections.

V. OPTIMAL FILTERS IN THE KLE DOMAIN

In this section, we are going to derive two classes of optimal and suboptimal filters in the KLE domain depending on the length \( L_4 \) of the filters \( h_L \).

A. Class 1

In this first category, we consider the particular case where 

\[ L_1 = L_2 = \cdots = L_L = 1. \]

Hence \( h_L = h_{L,0} \), i.e., \( l = 1, 2, \ldots, L \), are simply scalars. For this class of filters, we have

\[
o_{SNR}(h_L) = SNR_L, \quad \forall L. \tag{37}
\]

In this situation, the subband SNR cannot be improved. (Note that speech signals are nonstationary in nature, so \( h_L \) may change from one frame to another. Therefore, if we compute the subband SNR by averaging the signal and noise powers across frames, then the cross-frame, long-term subband SNR can still be improved.) Unlike the subband SNR, the fullband output SNR can be improved with respect to the input SNR. From the previous section we deduce that it is upper bounded (for all filters) as follows:

\[
o_{SNR}(h_{1:L}) \leq \sum_{l=1}^{L} SNR_l, \tag{38}
\]

1) Wiener Filter: Let us define the error signal in the KLE domain between the clean speech and its estimate

\[
e_k(k) \triangleq a_{x,\ell}(k) - a_{\ell}(k) = a_{x,\ell}(k) - h_{L,0}a_{y,\ell}(k), \quad l = 1, 2, \ldots, L. \tag{39}
\]

The KLE-domain MSE is

\[
J(h_{L,0}) \triangleq E \left[ \hat{e}_x^2(k) \right], \quad l = 1, 2, \ldots, L. \tag{40}
\]

Taking the gradient of \( J(h_{L,0}) \) with respect to \( h_{L,0} \) and equating the result to zero, we obtain the Wiener filter:

\[
h_{W,L,0} = \frac{E \left[ a_{x,\ell}^2(k) \right]}{E \left[ a_{y,\ell}^2(k) \right]} = \frac{\lambda_l}{SNR_L + \lambda_l}, \quad l = 1, 2, \ldots, L. \tag{41}
\]

It is seen that the form of this optimal filter is the same as that of the frequency-domain Wiener filter developed in [26], [34].

Property 1: We have

\[
\rho^2(a_{x,\ell}, a_{y,\ell}) + \rho^2(a_{\ell}, a_{y,\ell}) = 1, \quad l = 1, 2, \ldots, L \tag{42}
\]

where

\[
\rho^2(a_{x,\ell}, a_{y,\ell}) \triangleq \frac{E^2[a_{x,\ell}(k)a_{y,\ell}(k)]}{E[a_{x,\ell}(k)] E[a_{y,\ell}(k)]} \tag{43}
\]

and

\[
\rho^2(a_{\ell}, a_{y,\ell}) \triangleq \frac{E^2[a_{\ell}(k)a_{y,\ell}(k)]}{E[a_{\ell}(k)] E[a_{y,\ell}(k)]} \tag{44}
\]

are, respectively, the squared Pearson correlation coefficients (SPCCs) between \( a_{x,\ell}(k) \) and \( a_{y,\ell}(k) \), and \( a_{\ell}(k) \) and \( a_{y,\ell}(k) \).

Proof: It can be checked that

\[
\rho^2(a_{x,\ell}, a_{y,\ell}) = \frac{SNR_L}{1 + SNR_L}, \quad l = 1, 2, \ldots, L \tag{45}
\]

and

\[
\rho^2(a_{\ell}, a_{y,\ell}) = \frac{1}{1 + SNR_L}, \quad l = 1, 2, \ldots, L. \tag{46}
\]

Adding (45) and (46) together, we find (42).

Property 1 shows that the sum of the two SPCCs is always constant and equal to 1. So if one increases the other decreases. In comparison, the definition and properties of the SPCC in the KLE domain are similar to those of the magnitude squared coherence function defined in the frequency domain [34].

Property 2: We have

\[
h_{W,L,0} = \rho^2(a_{x,\ell}, a_{y,\ell}) = 1 - \rho^2(a_{\ell}, a_{y,\ell}), \quad l = 1, 2, \ldots, L. \tag{47}
\]

These fundamental forms of the KLE-domain Wiener filter, although obvious, do not seem to be known in the literature. They show that the Wiener filter is simply related to two SPCCs. Since

\[
0 \leq \rho^2(a_{x,\ell}, a_{y,\ell}) \leq 1, \quad 0 \leq h_{W,L,0} \leq 1, \quad \forall L.
\]

The Wiener filter acts like a gain function. When the level of noise at the subband \( l \) is high \( \left[ \rho^2(a_{x,\ell}, a_{y,\ell}) \to 1 \right] \), then \( h_{W,L,0} \) is close to 0 since there is a large amount of noise that needs to be removed. When the level of noise at the subband \( l \) is low \( \left[ \rho^2(a_{x,\ell}, a_{y,\ell}) \to 0 \right] \), then \( h_{W,L,0} \) is close to 1 and is not going to affect much the signals since there is little noise that needs to be removed.
We deduce the subband noise-reduction factor and speech-distortion index as:

\[
\xi_{\text{fr}}(h_{\text{W},l}) = \frac{1}{\rho^l(a_{x,l}, a_{y,l})} \geq 1, \quad l = 1, 2, \ldots, L
\]

(49)

\[
u_{\text{sd}}(h_{\text{W},l}) = \rho^l(a_{v,l}, a_{y,l}) \leq 1, \quad l = 1, 2, \ldots, L.
\]

(50)

It can be checked that these two measures are related by the formula

\[
u_{\text{sd}}(h_{\text{W},l}) = 1 - \frac{2}{\sqrt{\xi_{\text{fr}}(h_{\text{W},l})}} + \frac{1}{\xi_{\text{fr}}(h_{\text{W},l})}, \quad l = 1, 2, \ldots, L.
\]

(51)

At the fullband level, the noise-reduction factor and speech-distortion index due to the Wiener filter can be written as

\[
\xi_{\text{fr}}(h_{\text{W},1:L}) = \sum_{l=1}^{L} \frac{q_l^T R_{v,l} q_l}{\sum_{l=1}^{L} \rho^l(a_{x,l}, a_{y,l}) q_l^T R_{v,l} q_l} \geq 1
\]

(52)

\[
u_{\text{sd}}(h_{\text{W},1:L}) = \sum_{l=1}^{L} \frac{\rho^l(a_{v,l}, a_{y,l})}{\sum_{l=1}^{L} \lambda_l} \leq 1.
\]

(53)

We see clearly how noise reduction and speech distortion depend on the two SPCCs \(\rho^2(a_{x,l}, a_{y,l})\) and \(\rho^2(a_{v,l}, a_{y,l})\) in the KLE-domain Wiener filter. When \(\rho^2(a_{x,l}, a_{y,l})\) increases, \(\xi_{\text{fr}}(h_{\text{W},1:L})\) decreases; at the same time \(\rho^2(a_{v,l}, a_{y,l})\) decreases and so does \(\nu_{\text{sd}}(h_{\text{W},1:L})\).

**Property 3:** With the optimal KLE-domain Wiener filter given in (41), the fullband output SNR is always greater than or equal to the input SNR, i.e., \(\text{oSNR}(h_{\text{W},1:L}) \geq \text{SNR}\).

**Proof:** The fullband output SNR with the Wiener filter given in (41) can be written as

\[
\text{oSNR}(h_{\text{W},1:L}) = \sum_{l=1}^{L} \frac{\lambda_l}{\sum_{l=1}^{L} \lambda_l + q_l^T R_{v,l} q_l}.
\]

(54)

Since \(\lambda_l \geq 0\) and \(q_l^T R_{v,l} q_l \geq 0\) \(\forall l\), we always have the following inequality (it can be shown by induction):

\[
\sum_{l=1}^{L} \frac{\lambda_l^2}{\sum_{l=1}^{L} \lambda_l + q_l^T R_{v,l} q_l} \geq \frac{\sum_{l=1}^{L} \lambda_l}{\sum_{l=1}^{L} q_l^T R_{v,l} q_l},
\]

(55)

with equality if and only if \(\lambda_l / (\lambda_l + q_l^T R_{v,l} q_l)\) is a constant \(\forall l\). Using the above inequality, together with (20) and (54), we obtain

\[
\text{oSNR}(h_{\text{W},1:L}) \geq \text{SNR}.
\]

(56)

Property 3 is fundamental. It shows that the KLE-domain Wiener filter is able to improve the (fullband) SNR of an observed noisy signal.

2) **Parametric Wiener Filtering:** Some applications may need more aggressive (as compared to the Wiener filter) noise reduction, while others on the contrary may require less speech distortion (so less aggressive noise reduction). An easy way to control the compromise between noise reduction and speech distortion is via the parametric Wiener filtering [19], [27]. The equivalent approach in the KLE domain is

\[
h_{\text{G},l,0} = [1 - \rho(a_{v,l}, a_{y,l})]^{\beta_2} = \frac{\text{SNR}}{1 + \text{SNR}}, \quad l = 1, 2, \ldots, L
\]

(57)

where \(\beta_1\) and \(\beta_2\) are two positive parameters that allow the control of this compromise. For \((\beta_1, \beta_2) = (2, 1)\), we get the KLE-domain Wiener filter developed in the previous section. Taking \((\beta_1, \beta_2) = (2, 1/2)\) leads to

\[
h_{\text{P},l,0} = \sqrt{1 - \rho^2(a_{v,l}, a_{y,l})} = \frac{\text{SNR}}{\sqrt{1 + \text{SNR}}}, \quad l = 1, 2, \ldots, L
\]

(58)

which is the equivalent form of the power subtraction method studied in [19], [22], [24], [27], and [35]. The pair \((\beta_1, \beta_2) = (1, 1)\) gives the equivalent form of the magnitude subtraction method [16]–[18], [36], [37].

\[
h_{\text{M},l,0} = 1 - \rho^2(a_{v,l}, a_{y,l}) = \frac{1}{\sqrt{1 + \text{SNR}}}, \quad l = 1, 2, \ldots, L.
\]

(59)

We can verify that the subband noise-reduction factors for the power subtraction and magnitude subtraction methods are

\[
\xi_{\text{fr}}(h_{\text{P},l}) = \frac{1}{\rho^2(a_{x,l}, a_{y,l})}, \quad l = 1, 2, \ldots, L
\]

(60)

\[
\xi_{\text{fr}}(h_{\text{M},l}) = \frac{1}{[1 - \sqrt{1 - \rho^2(a_{x,l}, a_{y,l})}]^2}, \quad l = 1, 2, \ldots, L
\]

(61)

and the corresponding subband speech-distortion indices are

\[
u_{\text{sd}}(h_{\text{P},l}) = [1 - \rho^2(a_{v,l}, a_{y,l})]^2, \quad l = 1, 2, \ldots, L.
\]

(62)

\[
u_{\text{sd}}(h_{\text{M},l}) = \rho^2(a_{v,l}, a_{y,l}), \quad l = 1, 2, \ldots, L.
\]

(63)

It can also be checked that

\[
\xi_{\text{fr}}(h_{\text{M},l}) \geq \xi_{\text{fr}}(h_{\text{P},l}) \geq \xi_{\text{fr}}(h_{\text{W},l}), \quad l = 1, 2, \ldots, L
\]

(64)

\[
u_{\text{sd}}(h_{\text{P},l}) \leq \nu_{\text{sd}}(h_{\text{W},l}) \leq \nu_{\text{sd}}(h_{\text{M},l}), \quad l = 1, 2, \ldots, L.
\]

(65)

The two previous inequalities are very important from a practical point of view. They show that, among the three methods, the magnitude subtraction is the most aggressive one as far as noise reduction is concerned, a very well-known fact in the literature [26], but at the same time it is the one that will likely add most distortion to the speech signal. The smoother approach is the power subtraction while the Wiener filter is between the two others in terms of speech distortion and noise reduction. Since \(0 \leq h_{\text{P},l,0} \leq 1\), then \(\text{oSNR}(h_{\text{G},1:L}) \geq \text{SNR}\). Therefore, all
three methods improve the (fullband) SNR. Many other variants of these algorithms can be found in [28] and [29].

3) Subspace Approach: The error signal defined in (39) can be rewritten as follows:

\[ e_l(k) = e_{x,l}(k) - e_{v,l}(k), \quad l = 1, 2, \ldots, L \]  \hspace{1cm} (66)

where

\[ e_{x,l}(k) \triangleq (1 - h_{L,0})a_{x,l}(k), \quad l = 1, 2, \ldots, L \]  \hspace{1cm} (67)

is the speech distortion due to the linear transformation, and

\[ e_{v,l}(k) \triangleq h_{L,0}a_{v,l}(k), \quad l = 1, 2, \ldots, L \]  \hspace{1cm} (68)

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

\[
\min_{h_{L,0}} J_x(h_{L,0}) \quad \text{subject to} \quad J_v(h_{L,0}) \leq \beta_l q_l^T R_{xy} q_l, \quad l = 1, 2, \ldots, L
\]  \hspace{1cm} (69)

where

\[ J_x(h_{L,0}) \triangleq E \left[ e_{x,l}(k)^2 \right], \quad l = 1, 2, \ldots, L \]  \hspace{1cm} (70)

\[ J_v(h_{L,0}) \triangleq E \left[ e_{v,l}(k)^2 \right], \quad l = 1, 2, \ldots, L \]  \hspace{1cm} (71)

and \( 0 < \beta_l < 1 \) in order to have some noise reduction. If we use a Lagrange multiplier \( \mu_l (\geq 0) \) to adjoint the constraint to the cost function, we find the optimal filter

\[ h_{S,L,0} = \frac{\lambda_l}{\lambda_l + \mu_l q_l^T R_{xy} q_l}, \quad l = 1, 2, \ldots, L \]  \hspace{1cm} (72)

Hence, \( h_{S,L,0} \) is a Wiener filter with adjustable input noise level \( \mu_l q_l^T R_{xy} q_l \). This optimal filter is equivalent to the subspace approach \([5], [11], [12], [15]\), but in the KLE domain. Since \( 0 \leq h_{S,L,0} \leq 1, \forall \mu \geq 0 \), then \( \delta \text{SNR}(h_{S,1:L}) \geq \text{SNR} \). Therefore, this method improves the (fullband) SNR.

4) Relationship Between the Time- and KLE-Domain Filters: We now discuss the relationship between the time-domain (given in (6)) and KLE-domain (given in (41)) Wiener filters. As a matter of fact, if we substitute the KLE-domain Wiener filter into (19), the estimator of the vector \( x(k) \) can be written as

\[
z_{\text{KLE},W}(k) = \sum_{l=1}^{L} h_{W,L,0} a_{y,l} q_l
\]

\[
= \left( \sum_{l=1}^{L} h_{W,L,0} q_l^T q_l \right) y(k)
\]

\[
= H_{\text{KLE},W} y(k).
\]  \hspace{1cm} (73)

Therefore, the time-domain version of the KLE-domain filters \( h_{W,L,0}(l = 1, 2, \ldots, L) \) can be expressed as

\[
H_{\text{KLE},W} = \sum_{l=1}^{L} h_{W,L,0} q_l q_l^T.
\]  \hspace{1cm} (74)

Substituting (41) into (74) leads to

\[
H_{\text{KLE},W} = QA[A + \text{diag}(Q^T R_{xy} Q)]^{-1} Q^T.
\]  \hspace{1cm} (75)

Now, substituting (7) into (6), we get another form of the time-domain Wiener filter

\[
H_W = QA[A + Q^T R_{xy} Q]^{-1} Q^T.
\]  \hspace{1cm} (76)

Clearly, the two filters \( H_W \) and \( H_{\text{KLE},W} \) are very close. For example if the noise is white, then \( H_W = H_{\text{KLE},W} \). Also the orthogonal matrix \( Q \) tends to diagonalize the Toeplitz matrix \( R_{xy} \). In this case, \( Q^T R_{xy} Q = \text{diag}(Q^T R_{xy} Q) \) and as a result, \( H_W \approx H_{\text{KLE},W} \).

Following the same line of analysis, all KLE-domain filters derived in the previous sections can be rewritten, equivalently, into the time domain.

- Power subtraction:

\[
H_{\text{KLE},P} = QA^{1/2}[A + \text{diag}(Q^T R_{xy} Q)]^{-1/2} Q^T.
\]  \hspace{1cm} (77)

- Magnitude subtraction:

\[
H_{\text{KLE},M} = I - Q[I + \Lambda \text{diag}^{-1}(Q^T R_{xy} Q)]^{-1/2} Q^T
\]  \hspace{1cm} (78)

- Subspace:

\[
H_{\text{KLE},S} = QA[A + \Omega \text{diag}(Q^T R_{xy} Q)]^{-1} Q^T
\]  \hspace{1cm} (79)

where \( \Omega = \text{diag} \{ \mu_1, \mu_2, \ldots, \mu_L \} \).

It is worth noticing that, if \( \mu_1 = \cdots = \mu_L = \mu \), the filter \( H_{\text{KLE},S} \) is identical to the filter proposed in [11]. The above short analysis has shown in a very simple manner how the most well-known filters are linked in the time and transformed domains.

B. Class II

In this section, we consider another category of filters \( h_W \) with length \( L > 1 \) (of course, we now have to assume that the \( Q \) matrix is the same over different frames, which is different from the Class I, where each frame can have a different \( Q \)). In this case, it is possible to improve both the subband and fullband SNRs at the same time.

1) Wiener Filter: From the MSE

\[
J(h_l) \triangleq E \left\{ [a_{x,l}(k) - h_l^T a_y(k)]^2 \right\}, \quad l = 1, 2, \ldots, L
\]  \hspace{1cm} (80)
we deduce the KLE-domain Wiener filter

\[
\begin{align*}
\mathbf{h}_{W,l} &= \mathbf{R}_{a_{x}a_{y},l}^{-1} \mathbf{R}_{a_{x}a_{y},l} \mathbf{u}_{l} = \left( \mathbf{I}_{l} - \mathbf{R}_{a_{y}a_{y},l}^{-1} \mathbf{R}_{a_{x}a_{y},l} \right) \mathbf{u}_{l} \\
&= \text{SNR}_{l} \left( \mathbf{I}_{l} + \text{SNR}_{l} \mathbf{R}_{a_{x}a_{y},l}^{-1} \mathbf{R}_{a_{x}a_{y},l} \right)^{-1} \times \mathbf{\hat{R}}_{a_{x}a_{y},l}^{-1} \mathbf{\hat{R}}_{a_{x}a_{y},l} \mathbf{u}_{l},
\end{align*}
\]

\[l = 1, 2, \ldots, L \tag{81}\]

where \(\mathbf{R}_{a_{x}a_{y},l}\) is defined in a similar way to \(\mathbf{R}_{a_{x}a_{y},l}\) given in (25)

\[\mathbf{u}_{l} = [1 \ 0 \ \cdots \ 0]^{T}\]

is a vector of length \(L_{d}\), \(\mathbf{I}_{l}\) is the identity matrix of size \(L_{d} \times L_{d}\), and

\[
\begin{align*}
\mathbf{\hat{R}}_{a_{x}a_{y},l} &\triangleq \frac{\mathbf{R}_{a_{x}a_{y},l}}{\lambda_{l}} \\
\mathbf{\hat{R}}_{a_{y}a_{y},l} &\triangleq \frac{\mathbf{R}_{a_{y}a_{y},l}}{q_{l}^{2} \mathbf{R}_{w}\mathbf{q}_{l}^{T}}.
\end{align*}
\]

Property 4: With the optimal KLE-domain Wiener filter given in (81), the subband output SNR is always greater than or equal to the subband input SNR, i.e., \(\text{SNR}_{W,l} \geq \text{SNR}_{l}\), \(\forall l\).

Proof: Let us evaluate the SPCC between \(a_{y,l}(k)\) and \(\mathbf{h}_{W,l}^{T} \mathbf{a}_{y,l}(k)\)

\[
\rho^{2}(a_{y,l}, \mathbf{h}_{W,l}^{T} \mathbf{a}_{y,l}) = \frac{(\mathbf{u}_{l}^{T} \mathbf{R}_{a_{y}a_{y},l} \mathbf{h}_{W,l})^{2}}{\lambda_{l} + q_{l}^{2} \mathbf{R}_{w}\mathbf{q}_{l}^{T}} \left( \mathbf{h}_{W,l}^{T} \mathbf{R}_{a_{y}a_{y},l} \mathbf{h}_{W,l} \right)
\]

\[
= \frac{\lambda_{l}}{\lambda_{l} + \frac{q_{l}^{2} \mathbf{R}_{w}\mathbf{q}_{l}^{T}}{q_{l}^{2} \mathbf{R}_{w}\mathbf{q}_{l}^{T}}} \cdot \frac{\mathbf{u}_{l}^{T} \mathbf{R}_{a_{y}a_{y},l} \mathbf{h}_{W,l}}{\mathbf{u}_{l}^{T} \mathbf{h}_{W,l} \mathbf{a}_{y,l}}
\]

\[= \rho^{2}(a_{x,l}^{T} \mathbf{h}_{W,l} \mathbf{a}_{y,l}). \tag{82}\]

Therefore

\[
\rho^{2}(a_{x,l}^{T} \mathbf{h}_{W,l} \mathbf{a}_{y,l}) = \rho^{2}(a_{y,l}^{T} \mathbf{h}_{W,l} \mathbf{a}_{y,l}) \cdot \rho^{2}(a_{x,l}^{T} \mathbf{h}_{W,l} \mathbf{a}_{y,l}) \\
\leq \rho^{2}(a_{y,l}^{T} \mathbf{h}_{W,l} \mathbf{a}_{y,l}). \tag{83}\]

Using the fact that

\[
\rho^{2}(a_{x,l}^{T} \mathbf{a}_{y,l}) = \frac{\text{SNR}_{l}}{1 + \text{SNR}_{l}} \tag{84}\]

\[
\rho^{2}(\mathbf{h}_{W,l}^{T} \mathbf{a}_{x,l}, \mathbf{h}_{W,l}^{T} \mathbf{a}_{y,l}) = \frac{\text{SNR}_{l}}{1 + \text{SNR}_{l}} \tag{85}\]

and

\[
\rho^{2}(a_{x,l}^{T} \mathbf{h}_{W,l} \mathbf{a}_{y,l}) = \rho^{2}(a_{x,l}^{T} \mathbf{h}_{W,l} \mathbf{a}_{x,l}) \cdot \rho^{2}(\mathbf{h}_{W,l}^{T} \mathbf{a}_{x,l}, \mathbf{h}_{W,l}^{T} \mathbf{a}_{y,l}) \\
\leq \rho^{2}(a_{x,l}^{T} \mathbf{h}_{W,l} \mathbf{a}_{x,l}). \tag{86}\]

we obtain

\[
\frac{\text{SNR}_{l}}{1 + \text{SNR}_{l}} \leq \frac{\text{oSNR}(\mathbf{h}_{W,l})}{1 + \text{oSNR}(\mathbf{h}_{W,l})}. \tag{87}\]

It is immediately clear that

\[
\text{oSNR}(\mathbf{h}_{W,l}) \geq \text{SNR}_{l}, \quad l = 1, 2, \ldots, L \tag{88}\]

which completes the proof.

Property 5: With the optimal KLE-domain Wiener filter given in (81), we always have

\[
\xi_{\text{err}}(\mathbf{h}_{W,l}) \geq \frac{\lambda_{l}}{\mathbf{h}_{W,l}^{T} \mathbf{R}_{a_{y}a_{y},l} \mathbf{h}_{W,l}} \geq 1, \quad l = 1, 2, \ldots, L \tag{89}\]

and

\[
\xi_{\text{err}}(\mathbf{h}_{W,1:L}) \geq 1. \tag{90}\]

Proof: Let us first show that \(\lambda_{l} \geq \mathbf{h}_{W,l}^{T} \mathbf{R}_{a_{y}a_{y},l} \mathbf{h}_{W,l}, \quad \forall l\).

Indeed

\[
\mathbf{h}_{W,l}^{T} \mathbf{R}_{a_{y}a_{y},l} \mathbf{h}_{W,l} \leq \mathbf{h}_{W,l}^{T} \mathbf{R}_{a_{y}a_{y},l} \mathbf{h}_{W,l} = \mathbf{h}_{W,l}^{T} \mathbf{R}_{a_{y}a_{y},l} \mathbf{u}_{l}. \tag{91}\]

Using the Cauchy–Schwarz inequality

\[
\mathbf{h}_{W,l}^{T} \mathbf{R}_{a_{y}a_{y},l} \mathbf{u}_{l} \leq \left( \mathbf{h}_{W,l}^{T} \mathbf{R}_{a_{y}a_{y},l} \mathbf{h}_{W,l} \right)^{1/2} \lambda_{l}^{1/2} \tag{92}\]

we deduce that

\[
\mathbf{h}_{W,l}^{T} \mathbf{R}_{a_{y}a_{y},l} \mathbf{h}_{W,l} \leq \lambda_{l}, \quad l = 1, 2, \ldots, L. \tag{93}\]

We can write the subband output SNR as

\[
\text{oSNR}(\mathbf{h}_{W,l}) = \text{SNR}_{l} \cdot \xi_{\text{err}}(\mathbf{h}_{W,l}) \\
= \frac{\mathbf{h}_{W,l}^{T} \mathbf{R}_{a_{y}a_{y},l} \mathbf{h}_{W,l}}{\lambda_{l}} \geq \text{SNR}_{l}, \quad l = 1, 2, \ldots, L \tag{94}\]

and hence

\[
\xi_{\text{err}}(\mathbf{h}_{W,l}) \geq \frac{\lambda_{l}}{\mathbf{h}_{W,l}^{T} \mathbf{R}_{a_{y}a_{y},l} \mathbf{h}_{W,l}} \geq 1, \quad l = 1, 2, \ldots, L. \tag{95}\]

It can be proved that

\[
\mathbf{h}_{W,l}^{T} \mathbf{R}_{a_{y}a_{y},l} \mathbf{h}_{W,l} \leq q_{l}^{2} \mathbf{R}_{w}\mathbf{q}_{l}, \quad l = 1, 2, \ldots, L. \tag{96}\]

It follows immediately that

\[
\sum_{l=1}^{L} \mathbf{h}_{W,l}^{T} \mathbf{R}_{a_{y}a_{y},l} \mathbf{h}_{W,l} \leq \sum_{l=1}^{L} q_{l}^{2} \mathbf{R}_{w}\mathbf{q}_{l}. \tag{97}\]

Therefore

\[
\xi_{\text{err}}(\mathbf{h}_{W,1:L}) \geq 1. \tag{98}\]

We can write the fullband output SNR as

\[
\text{oSNR}(\mathbf{h}_{W,1:L}) = \text{SNR} \cdot \xi_{\text{err}}(\mathbf{h}_{W,1:L}) \sum_{l=1}^{L} \lambda_{l} \tag{99}\]
We see from the previous expression that the fullband SNR is improved if
\[
\xi_{\text{opt}}(h_{W,l}) \geq \frac{\sum_{l=1}^{L} \lambda_l}{\sum_{l=1}^{L} h_{W,l}^T R_{a_l a_l} h_{W,l}}. \tag{100}
\]

2) Maximum SNR Filter: The minimization of the MSE criterion \[\{80\}\] leads to the Wiener filter. Another straightforward criterion is to maximize the subband output SNR, \(\sigma_{\text{SNR}}(h_l)\), defined in \[\{23\}\] since SNR improvement is one of the major concerns in noise reduction.

Maximizing \(\sigma_{\text{SNR}}(h_l)\) is equivalent to solving the following generalized eigenvalue problem:
\[
R_{a_l a_l} h_l = \lambda h_{l} R_{a_l a_l} h_l, \quad l = 1, 2, \ldots, L. \tag{101}
\]
The optimal solution to this well-known problem is \(h_{\text{max},l}\), the eigenvector corresponding to the maximum eigenvalue, \(\lambda_{l,\text{max},l}\), of the matrix \(R_{-1} a_l a_l R_{a_l a_l} a_l a_l d\). In this case we have
\[
\sigma_{\text{SNR}}(h_{\text{max}}) = \lambda_{l,\text{max},l}, \quad l = 1, 2, \ldots, L. \tag{102}
\]
It is clear that \(d h_{\text{max},l}\), for any scalar \(c\), is also a solution of \[\{101\}\]. Usually we choose the eigenvector that has the unit norm, i.e., \(h_{\text{max},l}^T h_{\text{max},l} = 1\).

It is important to observe that the maximum SNR filter does not exist in Class I.

3) Subspace Approach: The filter for this approach is obtained by solving the following optimization problem:
\[
\min_{h_l} J_c(h_l) \quad \text{subject to} \quad J_c(h_l) \leq \beta \xi_{c}^2 R_{a_l a_l} q_l, \quad l = 1, 2, \ldots, L. \tag{103}
\]
where
\[
J_c(h_l) \triangleq E\left\{[u_l - h_l]^T a_{x,l}^2\right\}, \quad l = 1, 2, \ldots, L \tag{104}
\]
\[
J_c(h_l) \triangleq E\left\{h_l^T a_{x,l}^2\right\}, \quad l = 1, 2, \ldots, L \tag{105}
\]
and \(0 < \beta \leq 1\) in order to have some noise reduction. If we use a Lagrange multiplier \(\mu_l (\geq 0)\) to adjoin the constraint to the cost function, we find the optimal filter
\[
h_{l}s,l = (R_{a_l a_l} a_l + \mu_l R_{a_l a_l} a_l)^{-1} R_{a_l a_l} a_l u_l
\]
\[
= [I_l + (\mu_l - 1) R_{a_l a_l} a_l]^{-1} (R_{a_l a_l} a_l - R_{a_l a_l} a_l u_l)
\]
\[
= [I_l + (\mu_l - 1) R_{a_l a_l} a_l]^{-1} h_{W,l} a_l, \quad l = 1, 2, \ldots, L. \tag{106}
\]

where the Lagrange multiplier satisfies \(J_c(h_l) = \beta \xi_{c}^2 R_{a_l a_l} q_l\), \(l = 1, 2, \ldots, L\). In practice it is not easy to determine \(\mu_l\). Therefore, when this parameter is chosen in an ad-hoc way, we can see the following:

• \(\mu_l > 1\). This corresponds to more aggressive noise reduction (compared with the Wiener filter). So the residual noise level would be lower, but it is achieved at the expense of higher speech distortion.

• \(\mu_l < 1\). This corresponds to less aggressive noise reduction (compared with the Wiener filter). In this situation, we get less speech distortion but not so much noise reduction.

VI. SIMULATIONS

We have formulated the noise reduction problem in the KLE domain and developed two classes of optimal noise-reduction filters in Section V. In this section, we study their performance through experiments.

A. Estimation of Correlation Matrices

The clean speech signal used in our experiments was recorded from a female speaker in a quiet office environment. It was sampled at 8 kHz and quantized with 16 bits (2 B). The overall length of the signal is 30 s. The noisy speech is obtained by adding noise to the clean speech (the noise signal is properly scaled to control the SNR). We considered two types of noise: one is a computer generated white Gaussian random process and the other is a noise signal recorded in a New York Stock Exchange (NYSE) room. The NYSE noise is also digitized with a sampling rate of 8 kHz and quantized with 16 bits. Compared with the Gaussian random noise which is stationary and white, the NYSE noise tends to be nonstationary and colored. It consists of sound from various sources such as electric fans, telephone rings, and even speakers. See [39] for some statistics of this babbling noise.

To implement the optimal noise-reduction filters developed in Section V, we need to know the statistics of both the noisy and noise signals. Specifically, the Class I filters require to know the correlation matrices \(R_{yy}\) and \(R_{y,}\), while the Class II filters need to know the matrices \(R_{a_l y,}\) and \(R_{a_l y,}\) in addition to \(R_{yy}\) and \(R_{y,}\). Since the noisy signal is accessible, the correlation matrix \(R_{yy}\) can be estimated from its definition in \(\{4a\}\) by approximating the mathematical expectation with a sample average. However, due to the fact that speech is nonstationary, the sample average has to be performed on a short-term basis so that the estimated correlation matrix can follow the short-term variations of the speech signal. Alternatively, we can estimate \(R_{yy}\) through the widely used recursive approach, where an estimate of \(R_{yy}\) at time instant \(k\) is obtained as
\[
R_{yy}(k) = \alpha_{y} R_{yy}(k-1) + \left(1 - \alpha_{y}\right) y(k) y^T(k) \tag{107}
\]
where \(\alpha_{y}\) is a forgetting factor that controls the influence of the previous data samples on the current estimate of the noisy correlation matrix.

In our formulation, signals are processed on a frame-by-frame basis. Therefore, we can also combine the short-term sample average and the recursive method to estimate the correlation matrix, where the frame correlation matrix is calculated based on the current frame of the signal, and an estimate of \(R_{yy}\) is then obtained by smoothing the frame correlation matrix, i.e.,
\[
R_{yy}(k) = \alpha_{y} R_{yy}(k-1) + \left(1 - \alpha_{y}\right) R_{yy}(k) \tag{108}
\]
where $\alpha_r^0$ is a forgetting factor, and $\hat{\mathbf{R}}_{yy}(k) = (1/L) \sum_{i=k-L+1}^k \mathbf{y}(k)\mathbf{y}^T(k)$ is the frame correlation matrix at time instant $k$.

We compared the above three estimation approaches [the short-term sample average, the recursive method given in (107), and combination of the short-term average and recursive method given in (108)] using experiments and found that they all can lead to similar noise reduction performance if the parameters associated with each method are properly optimized, but in general the recursive approach given in (107) is easier to tune up, and as a result, this method will be used in our experiments.

The noise statistics can be estimated in many different ways. The most straightforward approach is to estimate them during the periods where the speech signal is absent. Such a method relies on a voice activity detector (VAD), and assumes that the background noise is stationary so that the estimated noise statistics during the absence of speech can represent the noise characteristics in the presence of speech. In our study, we have developed a sequential algorithm, which estimates the noise signal in the time–frequency domain [38]. This method has been shown to be able to produce reasonably accurate estimate of the statistics of the noise in practical environments. However, for most experiments that will be presented in this section, we intend not to use any noise estimator, but compute the noise correlation matrix directly from the noise signal (in a similar way to $\hat{\mathbf{R}}_{yy}$ in (107), but with a different forgetting factor $\alpha_n$). The reason behind this is that we want to study the optimal values of the parameters used in different noise-reduction filters. To find the optimal values of those parameters, it is better to simplify the experiments and avoid the influence of the noise estimation errors.

B. Experimental Results of the Class I Filters

Now let us investigate the performance of the Class I optimal noise-reduction filters. We will focus on the Wiener filter [either (41) or (75)], the power subtraction method [either (58) or (77)], and the subspace approach [either (72) or (79)]. During implementation, we first estimate the matrices $\mathbf{R}_{yy}(k)$ and $\mathbf{R}_{vn}(k)$. The KLT matrix $\mathbf{Q}$ is then obtained by eigenvalue decomposition of $\mathbf{R}_{xx}(k) = \mathbf{R}_{yy}(k) - \mathbf{R}_{vn}(k)$. In order to compute the filters $\mathbf{H}_{\text{KLE},W}$, $\mathbf{H}_{\text{KLE},P}$, and $\mathbf{H}_{\text{KLE},S}$, we have to compute, respectively, the inverse of the diagonal matrices $[\mathbf{A} + \text{diag}(\mathbf{Q}^T \mathbf{R}_{vn}\mathbf{Q})]$, $[\mathbf{A} + \text{diag}(\mathbf{Q}^T \mathbf{R}_{vn}\mathbf{Q})]$, and $[\mathbf{A} + \mu^{\lambda} \text{diag}(\mathbf{Q}^T \mathbf{R}_{vn}\mathbf{Q})]$ (note that in the subspace method we only consider the case where $\mu_k = \cdots = \mu_L = \mu$ for simplicity). However, considering the numerical stability issue, we computed the Moore–Penrose pseudoinverse of these matrices instead of their direct inverse in our implementation.

The first experiment studies the effect of the forgetting factor on the performance of noise reduction. As we have explained in the previous subsection, the forgetting factor plays a critical role in the estimation accuracy of the correlation matrices, which in turn may significantly affect the noise-reduction performance. For computing $\mathbf{R}_{yy}(k)$, the forgetting factor $\alpha_y$ cannot be too large. If it is too large (close to 1), the recursive estimate will essentially be a long-term average and will not be able to follow the short-term variations of the speech signal. As a result, the nature of the speech signal is not fully taken advantage of, which limits the noise-reduction performance. Conversely, if $\alpha_y$ is too small, the estimation variance of $\mathbf{R}_{yy}(k)$ will be large, which, again, may lead to performance degradation in noise reduction. Furthermore, $\mathbf{R}_{yy}(k)$ may tend to be rank deficient, causing numerical stability problems. Therefore, a proper value of the forgetting factor is very important. Unfortunately, it is very difficult to determine the optimal value of the forgetting factor using analytical methods. So, in this experiment, we attempt to find the optimal forgetting factor by directly examining the noise reduction performance. White noise is used in this experiment and $\text{SNR} = 10$ dB. The noise correlation $\mathbf{R}_{vn}(k)$ is directly computed from the noise signal using a recursive method. Since this noise is stationary, we can use a large forgetting factor. We set $\alpha_y = 0.995$. Fig. 1 plots both the output SNR and speech distortion index as a function of $\alpha_y$ (in the evaluation, the noise reduction filter is directly applied to the clean speech $\mathbf{x}(k)$ and the noise signal $\mathbf{v}(k)$ to obtain the filtered speech $\mathbf{x}_f(k)$ and residual noise $\mathbf{v}_n(k)$, and the output SNR and speech distortion index are then computed according to (21) and (32), respectively). It is seen that, for all the investigated algorithms, both the output SNR and speech distortion index bear a non-monotonic relationship with $\alpha_y$. Specifically, the output SNR first increases as $\alpha_y$ and then decreases, but the speech distortion index first decreases with $\alpha_y$ and then increases. The optimal noise-reduction performance (highest output SNR and lowest speech distortion) appears when $\alpha_y$ is in the range between 0.985 and 0.99. So, in the subsequent experiments, we will set $\alpha_y$ to 0.985.

It is also seen from Fig. 1 that the power subtraction method yielded the least SNR gain, but it also has the lowest speech distortion as compared to the Wiener filter and subspace method. The performance of the subspace technique depends on the value of $\mu$. When $\mu = 2$, this method achieved higher output SNR than the Wiener filter, but at the cost of higher speech distortion as seen in Fig. 1(b). When $\mu = 0.8$, the subspace method yielded less SNR improvement as compared to the Wiener filter. All these agreed very well with the theoretical analysis given in Section V. It seems from Fig. 1 that when $\mu$ is small (e.g., $\mu = 0.8$), the performance of the subspace method is more sensitive (compared to the case where $\mu$ is large) to the value of the forgetting factor $\alpha_y$. This can be explained from (79). Slightly rearranging (79) gives

$$
\mathbf{H}_{\text{KLE},S} = \mathbf{Q} \mathbf{A} [\mathbf{A} + \text{diag}(\mathbf{Q}^T \mathbf{R}_{vn}\mathbf{Q})]^{-1} \mathbf{Q}^T \mathbf{y} - (\mu - 1)\text{diag}(\mathbf{Q}^T \mathbf{R}_{vn}\mathbf{Q})^{-1} \mathbf{Q}^T \mathbf{y}. \tag{109}
$$

The summation of the first two terms in the brackets is the eigenvalue matrix of $\hat{\mathbf{R}}_{yy}$. This sum matrix is supposed to be positive definite. If $\mu < 1$, then $\mu - 1$ becomes negative, which means that we are subtracting a positive definite matrix (the matrix $\text{diag}(\mathbf{Q}^T \mathbf{R}_{vn}\mathbf{Q})$ is supposed to be positive definite) from the sum matrix, which may cause the overall summation matrix in the brackets to be no longer positive definite. Although with the use of the pseudoinverse we do not experience any numerical problem, the subtraction operation can significantly affect

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 3, 2009 at 14:00 from IEEE Xplore. Restrictions apply.
the signal subspace, particularly when \( \alpha_y \) is small and the estimation variance of \( R_{yy} \) is large. Therefore, for the subspace method with \( \mu < 1 \), we should make \( \alpha_y \) reasonably large.

Another important parameter for all the Class I filters is the filter length \( L \). So, in the second experiment, we study the impact of the filter length (also the frame size) \( L \) on the performance of noise reduction. Again, white noise is used, \( \text{SNR} = 10 \text{ dB} \), and the noise correlation matrix is directly computed from the noise signal using a recursive method. Based on the previous experiment, we set \( \alpha_y = 0.995 \) and \( \alpha_p = 0.995 \). Fig. 2 depicts the results. It is clear that the length \( L \) should be reasonably large enough to achieve good noise reduction performance. When \( L \) increases from 1 to 20, the output SNR improves while speech distortion decreases, but if we continue to increase \( L \), there is either marginal additional SNR improvement (for the subspace method with \( \mu = 2 \)), or even slight SNR degradation (for the Wiener filter, the power subtraction, and the subspace with \( \mu = 0.8 \)), and there is also some increase in speech distortion. In general, good performance for all the studied algorithms is achieved when the filter length \( L \) is around 20. This result coincides with what was observed with the frequency-domain Wiener filter [2]. The reason behind this is that a speech sample can be predicted from its neighboring values. It is this predictability that helps us achieve noise reduction without noticeably distorting the desired speech signal. In order to fully take advantage of the speech predictability, the filter length needs to be larger than the order of speech prediction, which is in the range between 10 and 20 for 8-kHz sampling rate. But if we continue to increase \( L \), the additional performance improvement will be limited. In theory, there should not be performance degradation for large \( L \). However, in practice, the estimation variance of the correlation matrix \( R_{yy} \) increases with \( L \), which generally leads to performance degradation.

In the third experiment, we test the performance of the Class I filters with different SNRs and noise conditions. We consider two types of noise: white Gaussian and NYSE. Based on the previous experiments, we choose \( L = 20 \), \( \alpha_y = 0.995 \). Again, the noise correlation matrix is directly computed from the noise signal using the recursive method. For white noise, \( \alpha_p \) is set to 0.995. But for the NYSE noise, \( \alpha_p \) is set to 0.99. (This value is obtained from experiments. Similar to the first experiment, we fixed \( \alpha_y \) to 0.995 and \( L = 20 \), but changed \( \alpha_p \) from 0 to 1. We found that the best noise-reduction performance is achieved when \( \alpha_y = 0.99 \) for the NYSE noise.) The experimental results are shown in Fig. 3, where we only plotted the results of the Wiener filter and subspace method with \( \mu = 2 \) to simplify the presentation.

It is seen from Fig. 3 that both the Wiener filter and subspace method perform better in white Gaussian noise environments than in NYSE noise conditions. This is due to the fact

---

**Fig. 1.** Noise-reduction performance versus \( \alpha_y \) in white Gaussian noise with: \( \text{SNR} = 10 \text{ dB}, \alpha_y = 0.995, \) and \( L = 20 \). Note that in the subspace method we set \( \mu_1 = \cdots = \mu_L = \mu \).

**Fig. 2.** Noise-reduction performance versus \( L \) in white Gaussian noise with: \( \text{SNR} = 10 \text{ dB}, \alpha_y = 0.985, \) and \( \alpha_p = 0.995 \). Note that in the subspace method we set \( \mu_1 = \cdots = \mu_L = \mu \).
that NYSE noise is nonstationary, and therefore it is more difficult to deal with. In general, with the same type of noise, the lower is the SNR, the more the noise reduction (higher SNR improvement) is achieved. But speech distortion increases almost exponentially as SNR decreases. This also agrees with what was observed with the frequency-domain Wiener filter. When SNR is too low (below 0 dB), the optimal noise-reduction filters may cause negative impact to the speech quality (instead of improving the speech quality, it may degrade it due to large speech distortion). To circumvent this problem in practical noise reduction systems, we suggest to use grace degradation, where when SNR is above a certain threshold (around 10 dB), the optimal filters can be directly applied to the noisy speech; but when SNR is below some lower threshold (around or below 0 dB), we should leave the noisy speech unchanged; if SNR is between the two thresholds (we call it the grace degradation range), we can use some suboptimal filter so that there is a smooth transition from low SNR to high SNR environments.

C. Experimental Results of the Class II Filters

The fourth experiment pertains to the Class II noise-reduction filters. 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 (otherwise, filtering the KLE coefficients across different frames would not make much sense). In this situation, the estimation of $Q$ is relatively easier than that for the Class I case. We can simply use a long-term sample average to compute the correlation matrices $R_{yy}$ and $R_{yw}$, thereby obtaining an estimate of $R_{yw}$. The KLT matrix $Q$ can then be computed using eigenvalue decomposition. In the course of our study, we found that the estimation accuracy of the $Q$ matrix plays a less important role in noise-reduction performance of the Class II methods than it does in performance of the Class I filters. We can even replace the $Q$ matrix with the Fourier matrix $F$ used in DFT, or the coefficient matrix in the discrete cosine transform (DCT) without degrading noise reduction performance (this indicates that the idea of the Class II filters can also be used in the frequency-domain approaches). However, strictly following the theoretical development in Section V-B, we still use the transformation matrix $Q$ in our experiments, with the correlation matrices $R_{yy}$ and $R_{yw}$ being estimated using a long-term average and the $R_{yw}$ matrix being computed as $R_{yy} - R_{yw}$. This $Q$ matrix is then applied to each frame of the signals to compute the KLE coefficients $a_{yi,k}$ and $a_{li,k}$.

The construction of the Class II optimal filters requires the knowledge of the correlation matrices $R_{a_{yi}a_{yi},k}$, $R_{a_{yi}a_{li},k}$, and $R_{a_{li}a_{yi},k}$. Since the noisy signal $y(k)$ is accessible, applying the $Q$ matrix to $y(k)$ would give us the KLE coefficients $a_{yi,k}$. We can then estimate $R_{a_{yi}a_{yi},k}$ using the recursive method similar to (107), i.e.,

$$R_{a_{yi}a_{yi},k} = \alpha_{a_{yi}} R_{a_{yi}a_{yi},k-1} + (1 - \alpha_{a_{yi}}) a_{yi,k} a_{yi,k}^T$$

where $\alpha_{a_{yi}}$, same as $\alpha_y$ in (107), is a forgetting factor, which will be optimized through experiments.

In order to estimate $R_{a_{yi}a_{yi},k}$, we need to have an estimate of the noise signal $v(k)$. Although we have developed a noise detector, we compute the noise statistics directly from the noise signal in this experiment to avoid the influence of the noise estimation error on the parameter optimization. Specifically, same as the way the $R_{a_{yi}a_{yi},k}$ matrix is computed, the KLT $Q$ is applied to the noise signal $v(k)$ to obtain the KLE coefficients $a_{vi,k}$. The $R_{a_{yi}a_{yi},k}$ matrix is then estimated using the same recursion given in (110), but with a different forgetting factor $\alpha_{a_{yi}}$.

The forgetting factors $\alpha_{a_{yi}}$ and $\alpha_{a_{li}}$ ($l = 1, \ldots, L$) play an important role in noise reduction performance of the Class II filters. In principle, each subband $l$ may take a different forgetting factor, but for simplicity, in this study, we assume the same forgetting factor for all the subbands, i.e., $\alpha_{a_{yi,1}} = \alpha_{a_{yi,2}} = \cdots = \alpha_{a_{yi,L}} = \alpha_{a_{yi}}$, and $\alpha_{a_{li,1}} = \alpha_{a_{li,2}} = \cdots = \alpha_{a_{li,L}} = \alpha_{a_{li}}$. Again, white noise is used. Since we already know an appropriate value of $\alpha_v$ for this noise, we can simply determine $\alpha_{a_{yi}}$ by forcing the two single-pole filters that are used to compute $R_{wv}(k)$ and $R_{v_{l}a_{yi}(k)}$ to have the same time constant. In our experimental setup, the sampling rate is 8 kHz, and the frame length $L = 20$. For $\alpha_v = 0.995$, it can be easily checked that the corresponding value of $\alpha_{a_{yi}}$ is approximately 0.91. Experiments also verified that this value can give reasonably accurate estimation of the noise statistics. So, in this experiment, we set $\alpha_{a_{yi}}$ to 0.91 and examine the noise-reduction performance for different values of $\alpha_{a_{yi}}$. The result of this experiment is plotted.
in Fig. 4. Note, again, that for the subspace method we only considered the case where $\mu_1 = \cdots = \mu_L = \mu$. It is observed that, for all the three studied algorithms, the performance first increases, and then decreases as $\alpha_{\text{SNR}}$ increases. The best performance is obtained with $\alpha_{\text{SNR}}$ being in the range between 0.8 and 0.9. We also see that, compared with the Wiener filter and subspace method ($\mu = 2$), the maximum SNR approach achieved much higher SNR improvement. However, the speech-distortion index with this method is also significantly higher than that of the Wiener filter and subspace method, which makes the method almost unusable.

In the next experiment, we study the impact of the filter length $L_d$ on the noise-reduction performance. Here we only consider the Wiener filter and subspace approach since the maximum SNR method introduces too much speech distortion. Again, the background noise is white and no noise estimator is used. The parameters used in this experiment are: $\alpha_{\text{SNR}} = 0.8$, $\alpha_{\text{SNR}} = 0.91$, and $L = 20$. The result is depicted in Fig. 5. It is seen from Fig. 5(a) that as $L_d$ 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 $L_d$. Taking into account both SNR improvement and speech distortion, we would suggest to use $L_d$ between 5 and 10.

Comparing Figs. 5 and 2, 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. This indicates that the Class II filters may have a great potential in practice.

In real applications, the noise statistics have to be estimated based on a noise estimator. So, in the last experiment, we evaluate the Class I and II filters for their performance when noise is estimated using the sequential algorithm developed in [38]. Briefly, this algorithm obtains an estimate of noise using the overlap-add technique on a frame-by-frame basis. The noisy speech signal $y(k)$ is segmented into frames with a frame width of 8 ms and an overlapping factor of 75%. Each frame is then
transformed via a DFT into a block of spectral samples. Successive blocks of spectral samples form a two-dimensional time-frequency matrix denoted by $\tilde{Y}_k(\omega)$, where subscript $k$ denotes the frame index, and $\omega$ is the angular frequency. Then an estimate of the magnitude of the noise spectrum is formulated as in (111), show at the bottom of the page, where $\omega_{\alpha}$ and $\omega_{\beta}$ are the “attack” and “decay” coefficients respectively. Meanwhile, to reduce its temporal fluctuation, the magnitude of the noisy speech spectrum is smoothed according to the following recursions, as shown in (112) at the bottom of the page, where again $\alpha_{\omega} = 0.985$ and $\alpha_{\omega} = 0.9$. For the Class II filters, only one $R_{yy}$ and $R_{xe}$ are computed for all the frames (using a long-term average), and $\alpha_{\omega} = 0.8$ and $\alpha_{\omega} = 0.91$ are used to compute the $R_{e_{\omega} a_{\omega} e_{\omega}}$ and $R_{e_{\omega} a_{\omega} e_{\omega}}$ matrices. In the subspace method we set $\mu_1 = \cdots = \mu_L = \mu$.

for the ideal case where noise statistics are directly computed from the noise signal are also provided in the table. It is seen that the noise estimator does not affect much the performance of the Class I filters. For the Class II filters, there is approximately a 3-dB sacrifice in SNR gain for both the Wiener filter and subspace method when $L_1$ is small (e.g., 4, 8), but when $L_1$ is large enough (e.g., 16, 20), the Class II filters can achieve a performance close to the ideal case. This indicates the feasibility of the developed algorithms for noise reduction in real applications.

VII. CONCLUSION

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 filters achieve a frame of speech estimate by filtering only the corresponding frame of the noisy speech, the second class 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. We have also discussed some implementation issues with the

<table>
<thead>
<tr>
<th>Class I filters</th>
<th>Subspace ($\mu = 2$)</th>
<th>$L_1 = 4$</th>
<th>15.7</th>
<th>0.022</th>
<th>12.5</th>
<th>0.020</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$L_1 = 8$</td>
<td>16.2</td>
<td>0.032</td>
<td>13.3</td>
<td>0.027</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$L_1 = 12$</td>
<td>15.9</td>
<td>0.043</td>
<td>13.8</td>
<td>0.040</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$L_1 = 16$</td>
<td>15.0</td>
<td>0.055</td>
<td>14.1</td>
<td>0.051</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$L_1 = 20$</td>
<td>12.6</td>
<td>0.085</td>
<td>13.2</td>
<td>0.088</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class II filters</th>
<th>Subspace ($\mu = 2$)</th>
<th>$L_1 = 4$</th>
<th>16.8</th>
<th>0.027</th>
<th>14.1</th>
<th>0.028</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$L_1 = 8$</td>
<td>18.6</td>
<td>0.030</td>
<td>15.6</td>
<td>0.029</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$L_1 = 12$</td>
<td>19.8</td>
<td>0.037</td>
<td>17.1</td>
<td>0.043</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$L_1 = 16$</td>
<td>20.4</td>
<td>0.045</td>
<td>18.2</td>
<td>0.069</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$L_1 = 20$</td>
<td>20.2</td>
<td>0.055</td>
<td>18.9</td>
<td>0.102</td>
<td></td>
</tr>
</tbody>
</table>

NOTES: SNR = 10 dB, $L = 20$. For the Class I filters, the correlation matrices $R_{yy}(k)$ and $R_{xe}(k)$ for each frame are computed using the recursive approach with $\alpha_y = 0.985$ and $\alpha_x = 0.9$. For the Class II filters, only one $R_{yy}$ and $R_{xe}$ are computed for all the frames (using a long-term average), and $\alpha_{\omega} = 0.8$ and $\alpha_{\omega} = 0.91$ are used to compute the $R_{e_{\omega} a_{\omega} e_{\omega}}$ and $R_{e_{\omega} a_{\omega} e_{\omega}}$ matrices. In the subspace method we set $\mu_1 = \cdots = \mu_L = \mu$.

\[
\hat{Y}_k(\omega) = \begin{cases} 
\alpha_y \hat{Y}_{k-1}(\omega) + (1 - \alpha_y)\hat{Y}_k(j\omega), & \text{if } |\hat{Y}_k(j\omega)| \geq |\hat{Y}_{k-1}(\omega)| \\
\alpha_y \hat{Y}_{k-1}(\omega) + (1 - \alpha_y)\hat{Y}_k(j\omega), & \text{if } |\hat{Y}_k(j\omega)| < |\hat{Y}_{k-1}(\omega)|
\end{cases}
\]

(111)

\[
\tilde{Y}_k(\omega) = \begin{cases} 
\beta_y \tilde{Y}_{k-1}(\omega) + (1 - \beta_y)\tilde{Y}_k(j\omega), & \text{if } |\tilde{Y}_k(j\omega)| \geq |\tilde{Y}_{k-1}(\omega)| \\
\beta_y \tilde{Y}_{k-1}(\omega) + (1 - \beta_y)\tilde{Y}_k(j\omega), & \text{if } |\tilde{Y}_k(j\omega)| < |\tilde{Y}_{k-1}(\omega)|
\end{cases}
\]

(112)

<table>
<thead>
<tr>
<th>The noise statistics are computed directly from the noise signal</th>
<th>The noise statistics are computed based on the sequential noise estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{SNR} (\text{dB})$</td>
<td>$\mu_{\text{ad}}$</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Class I filters</td>
<td>Subspace ($\mu = 2$)</td>
</tr>
<tr>
<td></td>
<td>$L_1 = 8$</td>
</tr>
<tr>
<td></td>
<td>$L_1 = 12$</td>
</tr>
<tr>
<td></td>
<td>$L_1 = 20$</td>
</tr>
<tr>
<td>Class II filters</td>
<td>Subspace ($\mu = 2$)</td>
</tr>
<tr>
<td></td>
<td>$L_1 = 8$</td>
</tr>
<tr>
<td></td>
<td>$L_1 = 12$</td>
</tr>
<tr>
<td></td>
<td>$L_1 = 16$</td>
</tr>
<tr>
<td></td>
<td>$L_1 = 20$</td>
</tr>
</tbody>
</table>
KLE domain optimal filters. Through experiments, we have investigated the optimal values of the forgetting factors and the length of the optimal filters. We also 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.

REFERENCES


Jingdong Chen (M’99) received the B.S. degree and the M.S. degree in electrical engineering from the Northwestern Polytechnic University, Xi’an, China, in 1993 and 1995 respectively, and the Ph.D. degree in pattern recognition and intelligence control from the Chinese Academy of Sciences, Beijing, in 1998. From 1998 to 1999, he was with ATR Interpreting Telecommunications Research Laboratories, Kyoto, Japan, where he conducted research on speech synthesis, speech analysis, as well as objective measurements for evaluating speech synthesis. He then joined the Griffith University, Brisbane, Australia, as a Research Fellow, where he engaged in research in robust speech recognition, signal processing, and discriminative feature representation. From 2000 to 2001, he was with ATR Spoken Language Translation Research Laboratories, Kyoto, where he conducted research in robust speech recognition and speech enhancement. He joined Bell Laboratories, Murray Hill, NJ, as a Member of Technical Staff in July 2001. His current research interests include adaptive signal processing, speech enhancement, adaptive noise/echo cancellation, microphone array signal processing, speech separation, and source localization. He coauthored the books Noise Reduction in Speech Processing (Springer-Verlag, 2009), Microphone Array Signal Processing (Springer-Verlag, 2008), and Acoustic MIMO Signal Processing (Springer-Verlag, 2006). He is a co-editor/co-author of the book Speech Enhancement (Springer-Verlag, 2005) and a section editor.

Dr. Chen is currently an Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING. He is also a member of the editorial board of the Open Signal Processing Journal. He helped organize the 2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), and he is the technical co-chair of the 2009 WASPAA. He is the recipient of 1998–1999 research grant from the Japan Key Technology Center, and the 1996–1998 President’s Award from the Chinese Academy of Sciences.

Jacob Benesty (M’92–SM’04) was born in 1963. He received the M.S. degree in microwaves from Pierre and Marie Curie University, Paris, France, in 1987 and the Ph.D. degree in control and signal processing from Orsay University, Paris, in April 1991.

During the Ph.D. degree (from November 1989 to April 1991), he worked on adaptive filters and fast algorithms at the Centre National d’Etudes des Telecommunications (CNET), Paris. From January 1994 to July 1995, he worked at Telecom Paris University on multichannel adaptive filters and acoustic echo cancellation. From October 1995 to May 2003, he was first a Consultant and then a Member of the Technical Staff at Bell Laboratories, Murray Hill, NJ. In May 2003, he joined INRS-EMT, University of Quebec, Montreal, QC, Canada, as a Professor. His research interests are in signal processing, acoustic signal processing, and multimedia communications. He coauthored the books Noise Reduction in Speech Processing (Springer-Verlag, 2009), Microphone Array Signal Processing (Springer-Verlag, 2008), Acoustic MIMO Signal Processing (Springer-Verlag, 2006), and Advances in Network and Acoustic Echo Cancellation (Springer-Verlag, 2001). He is the Editor-In-Chief of the reference Springer Handbook of Speech Processing (Springer-Verlag, 2007). He is also a coeditor/coauthor of the books Speech Enhancement (Springer-Verlag, 2005), Audio Signal Processing for Next Generation Multimedia Communication Systems (Kluwer, 2004), Adaptive Signal Processing: Applications to Real-World Problems (Springer-Verlag, 2003), and Acoustic Signal Processing for Telecommunication (Kluwer, 2000).

Dr. Benesty received the 2001 Best Paper Award from the IEEE Signal Processing Society. He was a member of the editorial board of the EURASIP Journal on Applied Signal Processing, a member of the IEEE Audio and Electroacoustics Technical Committee, and was the co-chair of the 1999 International Workshop on Acoustic Echo and Noise Control (IWAENC). He is the general co-chair of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA).

Yiteng (Arden) Huang (S’97–M’01) received the B.S. degree from the Tsinghua University, Beijing, China, in 1994 the M.S. and Ph.D. degrees from the Georgia Institute of Technology (Georgia Tech), Atlanta, in 1998 and 2001, respectively, all in electrical and computer engineering.


Dr. Huang served as an Associate Editor for the EURASIP Journal on Applied Signal Processing from 2004 and 2008 and for the IEEE SIGNAL PROCESSING LETTERS from 2002 to 2005. He served as a technical co-chair of the 2005 Joint Workshop on Hands-Free Speech Communication and Microphone Array and the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. He received the 2002 Young Author Best Paper Award from the IEEE Signal Processing Society, the 2000–2001 Outstanding Graduate Teaching Assistant Award from the School Electrical and Computer Engineering, Georgia Tech, the 2000 Outstanding Research Award from the Center of Signal and Image Processing, Georgia Tech, and the 1997–1998 Colonel Oscar P. Cleaver Outstanding Graduate Student Award from the School of Electrical and Computer Engineering, Georgia Tech.