Noise Reduction Algorithms in a Generalized Transform Domain

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

***Abstract*—Noise reduction for speech applications is often formulated as a digital filtering problem, where the clean speech estimate is obtained by passing the noisy speech through a linear filter/transform. With such a formulation, the core issue of noise reduction becomes how to design an optimal filter (based on the statistics of the speech and noise signals) that can significantly suppress noise without introducing perceptually noticeable speech distortion. The optimal filters can be designed either in the time or in a transform domain. The advantage of working in a transform space is that, if the transform is selected properly, the speech and noise signals may be better separated in that space, thereby enabling better filter estimation and noise reduction performance. Although many different transforms exist, most efforts in the field of noise reduction have been focused only on the Fourier and Karhunen–Loève transforms. Even with these two, no formal study has been carried out to investigate which transform can outperform the other. In this paper, we reformulate the noise reduction problem into a more generalized transform domain. We will show some of the advantages of working in this generalized domain, such as 1) different transforms can be used to replace each other without any requirement to change the algorithm (optimal filter) formulation, and 2) it is easier to fairly compare different transforms for their noise reduction performance. We will also address how to design different optimal and suboptimal filters in such a generalized transform domain.**

***Index Terms*—cosine transform, Fourier transform, Hadamard transform, Karhunen–Loève expansion (KLE), noise reduction, speech enhancement, tradeoff filter, Wiener filter.**

I. Introduction

NOISE is ubiquitous in almost all acoustic environments. In applications related to speech, sound recording, telecommunications, voice over IP (VoIP), teleconferencing, telecollaboration, and human–machine interfaces, the signal of interest (usually speech) that is picked up by a microphone is generally contaminated by noise originating from various sources. Such contamination can dramatically change the characteristics of the speech signals and degrade the speech quality and intelligibility, thereby causing significant harm to human-to-human and human-to-machine communication systems. In order to mitigate the detrimental effect of noise on speech processing and communication, it is desirable to develop digital signal processing techniques to "clean" the noisy speech before it is stored, transmitted, or played out. This cleaning process, which is often referred to as noise reduction, has been a major challenge for many researchers and engineers for more than four decades.

Generally speaking, noise is a term used to signify any unwanted signal that interferes with the measurement and processing of the desired speech signal. This broad-sense definition, however, makes the problem too complicated to deal with, and as a result, research is focused on coping with one category of noise at once. In the area of speech processing, we normally divide noise into four categories: additive noise (from various ambient sound sources), interference (from concurrent competing speakers), reverberation (caused by multipath propagation), and echo (resulting from coupling between loudspeakers and microphones). Combating these four types of noise has led to the developments of four broad classes of acoustic signal processing techniques: noise reduction/speech enhancement, source separation, speech dereverberation, and echo cancellation/suppression. Now in the context of noise reduction, the term noise is widely accepted as additive noise that is statistically independent of the desired speech signal. In this situation, the problem of noise reduction becomes one of restoring the clean speech from the microphone signal, which is basically a superposition of the clean speech and noise.

The complexity of this problem depends on many factors such as the noise characteristics, the number of microphones, the performance measure, etc. In a given noise condition and with a specified performance measure, the problem is generally easier as the number of microphones increases [1]–[5]. However, most of today's speech communication devices are equipped with only one microphone. In such a situation, the estimation of the clean speech has to be based on manipulation of the single microphone output. This has made noise reduction a very difficult problem since no reference is accessible for the estimation of the noise. Fortunately, speech and noise usually have very different statistics. By taking advantage of this difference, we can design some filter where the desired signal can pass through while the additive noise can be attenuated. Note, however, that this filtering process will inevitably modify the clean speech while reducing the level of noise [6]. Therefore, the core problem in noise reduction becomes one of how to design an optimal filter that can significantly suppress noise without introducing perceptually noticeable speech distortion.

The design of optimal noise reduction filters can be achieved directly in the time domain by optimizing the expected value

Manuscript received October 20, 2008; revised March 22, 2009. Current version published June 26, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Nakatani Tomohiro.

J. Benesty is with INRS-EMT, University of Quebec, Montreal, QC H5A 1K6, Canada.

J. Chen is with Bell Labs, Alcatel-Lucent, Murray Hill, NJ 07974 USA (e-mail: jingdong@research.bell-labs.com).

Y. A. Huang is with WeVoice, Inc., Bridgewater, NJ 08807 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASL.2009.2020415

of some distortion measure using the clean and estimated signals. For example, the well known Wiener filter is obtained by minimizing the mean-squared error (MSE) between the clean speech and its estimate [5]–[8]. However, most developed noise reduction approaches so far prefer to consider the optimal filters in a transform space. This is due to the fact that if the transform is properly selected, the speech and noise signals can be better separated in that space, making it easier to estimate the noise statistics. A typical example is the well-studied subspace method [9]–[15]. This approach projects the noisy signal vector into a different domain either via the Karhunen–Loève (KL) transform through eigenvalue decomposition of an estimate of the correlation matrix of the noisy signal [9]–[14] or by using the singular value decomposition of a data matrix constructed from the noisy signal vector [15]. Once transformed, the speech signal only spans a portion of the entire space, and as a result, the entire vector space can be divided into two subspaces: the signal-plus-noise and the noise only. The noise statistics can then be estimated from the noise only subspace. These statistics can subsequently be used to remove the noise subspace and clean the signal-plus-noise subspace, thereby restoring the desired clean speech. Another advantage of working in a transform domain is that the noise reduction filter on each base space (or subband) can be manipulated individually, which provides us with more flexibility in controlling the compromise between the amount of noise reduction and the degree of speech distortion.

Remarkably, there are many transforms that can be used; however, we do not know which transform would be best suited for the application of noise reduction. In the literature, most efforts have been focused on the use of the Fourier and KL transforms, but even with these two transforms, no formal study has been carried out to investigate which one can outperform the other (with the same experimental configuration). In this paper, we attempt to provide a new framework that can be used not only for deriving different noise reduction filters but also for fairly comparing different transforms for their noise reduction performance. Our major contributions include the following. 1) We reformulate the noise reduction problem into a more generalized transform domain, where any unitary (or orthogonal) matrix can be used to serve as a transform. 2) We address how to design different optimal and suboptimal filters in the generalized transform domain. 3) We demonstrate some advantages of working in the generalized transform domain, such as: different transforms can be used to replace each other without any requirement to change the algorithm (optimal filter) formulation; and it is easier to fairly compare different transforms for their noise reduction performance. 4) We compare several popularly used transforms (including the Fourier, KL, cosine, Hadamard, and identity transforms) for their performance in noise reduction.

The rest of this paper is organized as follows. In Section II, we briefly describe the signal model used in this paper. We then discuss the principle of noise reduction in the KL expansion (KLE) domain in Section III. In Section IV, we present a new generalized transform domain, where any given unitary (or orthogonal) matrix can be used to serve as the transform. Some performance measures will then be provided in Section V. These measures are critical for designing as well as evaluating noise reduction filters. Detailed discussions on how to design different optimal and suboptimal filters will be given in Section VI. In Section VII, we present some experimental results. Finally, some conclusions will be drawn in Section VIII.

II. Problem Formulation

The noise reduction problem considered in this paper is one of recovering the 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 $x(k)$.

The signal model given in (1) can be written in a vector form if we process the data on a per block basis with a block size of L

$$\mathbf{y}(k) = \mathbf{x}(k) + \mathbf{v}(k) \tag{2}$$

where

$$\mathbf{y}(k) \triangleq [y(k) \quad y(k-1) \quad \cdots \quad y(k-L+1)]^T.$$

Superscript ^{T} denotes transpose of a vector or a matrix, 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 desired and noise signals, i.e.,

$$\mathbf{R}_{yy}(k) = \mathbf{R}_{xx}(k) + \mathbf{R}_{vv}(k) \tag{3}$$

where $\mathbf{R}_{yy}(k) \triangleq E[\mathbf{y}(k)\mathbf{y}^T(k)]$, $\mathbf{R}_{xx}(k) \triangleq E[\mathbf{x}(k)\mathbf{x}^T(k)]$, and $\mathbf{R}_{vv}(k) \triangleq E[\mathbf{v}(k)\mathbf{v}^T(k)]$ are, respectively, the correlation matrices of the signals $y(k)$, $x(k)$, and $v(k)$ at time instant k, with $E(\cdot)$ denoting mathematical expectation. Note that the correlation matrices for nonstationary signals like speech are in general time-varying, and hence a time index is used here, but for convenience of presentation, in the rest of this paper, we will drop the time index k and assume that all signals are quasi-stationary.

Our objective in this paper is to estimate either $\mathbf{x}(k)$ or $x(k)$ from the observation vector $\mathbf{y}(k)$, which is normally achieved by applying a linear transformation to the microphone signal [3], [5], [16], i.e.,

$$\mathbf{z}(k) = \mathbf{H}\mathbf{y}(k) = \mathbf{x}_{\mathrm{f}}(k) + \mathbf{v}_{\mathrm{rn}}(k) \tag{4}$$

where $\mathbf{H}$ is a filtering matrix of size $L \times L$, $\mathbf{z}(k)$ is supposed to be an estimate of $\mathbf{x}(k)$, and $\mathbf{x}_{\mathrm{f}}(k) \triangleq \mathbf{H}\mathbf{x}(k)$ and $\mathbf{v}_{\mathrm{rn}}(k) \triangleq \mathbf{H}\mathbf{v}(k)$ are, respectively, the filtered speech and residual noise after noise reduction. With this formulation, the noise reduction problem becomes one of finding an optimal filter that would attenuate the noise as much as possible while keeping the speech from being dramatically distorted. One of the most used solu-

tions to this is the classical Wiener filter derived from the MSE criterion $E\{[\mathbf{x}(k)-\mathbf{z}(k)]^T[\mathbf{x}(k)-\mathbf{z}(k)]\}$. This optimal filter is [17], [18]

$$\mathbf{H}_{\mathrm{W}} = \mathbf{R}_{xx}\mathbf{R}_{yy}^{-1} \tag{5}$$

and most known filters, in the time and frequency (or other) domains, are somehow related to this one as will be discussed 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 $\mathbf{y}(k)$ denote a data sequence drawn from a zero-mean stationary process with the correlation matrix $\mathbf{R}_{yy}$. This matrix can be diagonalized as follows [19]:

$$\mathbf{Q}^T\mathbf{R}_{yy}\mathbf{Q} = \mathbf{\Lambda} \tag{6}$$

where

$$\mathbf{Q} = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \cdots \quad \mathbf{q}_L]$$

and

$$\mathbf{\Lambda} = \mathrm{diag}[\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_L]$$

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 $\mathbf{R}_{yy}$.

The vector $\mathbf{y}(k)$ can be written as a combination (expansion) of the eigenvectors of the correlation matrix $\mathbf{R}_{yy}$ as follows [20]:

$$\mathbf{y}(k) = \sum_{l=1}^{L} a_y(k,\mathbf{q}_l)\mathbf{q}_l \tag{7}$$

where

$$a_y(k,\mathbf{q}_l) = \mathbf{q}_l^T\mathbf{y}(k), \quad l = 1,2,\ldots,L \tag{8}$$

are the coefficients of the expansion.

The representation of the random vector $\mathbf{y}(k)$ described by (7) and (8) is the KLE [20], where (7) is the synthesis part and (8) represents the analysis part.

It can be verified from (8) that

$$E[a_y(k,\mathbf{q}_l)] = 0, \quad l = 1,2,\ldots,L \tag{9}$$

and

$$E[a_y(k,\mathbf{q}_i)a_y(k,\mathbf{q}_j)] = \begin{cases} \lambda_i, & i = j \\ 0, & i \neq j. \end{cases} \tag{10}$$

It can also be checked from (8) that the Parseval's theorem holds, i.e.,

$$\sum_{l=1}^{L} E\left[a_y^2(k,\mathbf{q}_l)\right] = \mathrm{tr}(\mathbf{R}_{yy}) \tag{11}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix.

Note that the extension of the KLE to nonstationary signals like speech is straightforward.

One of the most important aspects of the KLE is its potential to reduce the dimensionality of the vector $\mathbf{y}(k)$. This idea has been extensively investigated in the so-called subspace method for noise reduction, where the signal of interest (speech) is assumed to be low-rank, and noise reduction is achieved by diagonalizing the noisy covariance matrix, removing the noise eigenvalues, and cleaning the signal-plus-noise eigenvalues [9], [11]–[13], [15], [21], [22]. In the following, we will take an approach different from the subspace method. Instead of manipulating the eigenvalues of the noisy correlation matrix, we will work directly in the KLE domain and achieve noise reduction by estimating the KLE coefficients of the clean speech in each KLE subband. Indeed, substituting (2) into (8), we get

$$\begin{aligned} a_y(k,\mathbf{q}_l) &= \mathbf{q}_l^T\mathbf{y}(k) \\ &= \mathbf{q}_l^T\mathbf{x}(k) + \mathbf{q}_l^T\mathbf{v}(k) \\ &= a_x(k,\mathbf{q}_l) + a_v(k,\mathbf{q}_l), \quad l = 1,2,\ldots,L. \end{aligned} \tag{12}$$

This expression is equivalent to (2) but in the KLE domain. We also have

$$E[a_y(k,\mathbf{q}_i)a_y(k,\mathbf{q}_j)] = \begin{cases} \mathbf{q}_i^T\mathbf{R}_{xx}\mathbf{q}_i + \mathbf{q}_i^T\mathbf{R}_{vv}\mathbf{q}_i, & i = j \\ 0, & i \neq j. \end{cases} \tag{13}$$

Therefore, the KLE coefficients of the noisy speech from one subband (here the term subband refers to the signal component along the base vector $\mathbf{q}_l$) are uncorrelated with those from other subbands, and as a result, we can estimate the KLE coefficients of the clean speech in each subband independently without considering the contribution from other subbands. Clearly, our problem this time is to find an estimate of $a_x(k,\mathbf{q}_l)$ by multiplying $a_y(k,\mathbf{q}_l)$ with a scalar filter h_l, i.e.,

$$\begin{aligned} a_z(k,\mathbf{q}_l) &= h_l a_y(k,\mathbf{q}_l) \\ &= h_l[a_x(k,\mathbf{q}_l) + a_v(k,\mathbf{q}_l)], \\ & \qquad l = 1,2,\ldots,L. \end{aligned} \tag{14}$$

We see that

$$E[a_z(k,\mathbf{q}_i)a_z(k,\mathbf{q}_j)] = \begin{cases} h_i^2\lambda_i, & i = j \\ 0, & i \neq j. \end{cases} \tag{15}$$

Finally, an estimate of the vector $\mathbf{x}(k)$ would be

$$\begin{aligned} \mathbf{z}(k) &= \sum_{l=1}^{L} a_z(k,\mathbf{q}_l)\mathbf{q}_l = \left(\sum_{l=1}^{L} h_l\mathbf{q}_l\mathbf{q}_l^T\right)\mathbf{y}(k) \\ &= \mathbf{H}(\mathbf{Q})[\mathbf{x}(k) + \mathbf{v}(k)] \end{aligned} \tag{16}$$

where

$$\mathbf{H}(\mathbf{Q}) \triangleq \mathbf{Q}\mathrm{diag}[h_1 \quad h_2 \quad \cdots \quad h_L]\mathbf{Q}^T \tag{17}$$

is an $L\times L$ (time-domain) filtering matrix which depends on the orthogonal matrix $\mathbf{Q}$ and is equivalent to the KLE-domain filter $\mathbf{h} = [h_1 \quad h_2 \quad \cdots \quad h_L]^T$. Moreover, it is easy to check

that the correlation matrix $\mathbf{R}_{zz} = E[\mathbf{z}(k)\mathbf{z}^T(k)]$ can be diagonalized as follows:

$$\mathbf{Q}^T\mathbf{R}_{zz}\mathbf{Q} = \text{diag}\left[h_1^2\lambda_1 \quad h_2^2\lambda_2 \quad \cdots \quad h_L^2\lambda_L\right]. \tag{18}$$

We see from the previous expression how the coefficients $h_l, l = 1, 2, \ldots, L$, affect the spectrum of the estimated signal $z(k)$, depending on how they are optimized.

IV. Generalization of the KLE

In this section, we are going to generalize the principle of the KLE to any given unitary transform $\mathbf{U}$. In order to do so, we need to use some of the concepts presented in [23]–[26]. The basic idea behind this generalization is to find other ways to exactly diagonalize the correlation matrix $\mathbf{R}_{yy}$. The Fourier matrix, for example, diagonalizes approximately $\mathbf{R}_{yy}$ (since this matrix is Toeplitz and its elements are usually absolutely summable [27]). However, this approximation may cause more distortion to the clean speech when noise reduction is performed in the frequency domain.

We define the square root of the positive definite matrix $\mathbf{R}_{yy}$ as

$$\mathbf{R}_{yy}^{1/2} \triangleq \mathbf{Q}\mathbf{\Lambda}^{1/2}\mathbf{Q}^T. \tag{19}$$

This definition is very useful in the derivation of a generalized form of the KLE.

Consider the $L \times L$ unitary matrix

$$\mathbf{U} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_L]$$

where $\mathbf{U}^H\mathbf{U} = \mathbf{U}\mathbf{U}^H = \mathbf{I}$, superscript ^{H} denotes transpose conjugate of a vector or a matrix, and $\mathbf{I}$ is the identity matrix. We would like to minimize the positive quantity $\mathbf{g}^H(\mathbf{u}_l)\mathbf{R}_{yy}^{1/2}\mathbf{g}(\mathbf{u}_l)$ subject to the constraint

$$\mathbf{g}^H(\mathbf{u}_l)\mathbf{u}_l = \mathbf{u}_l^H\mathbf{g}(\mathbf{u}_l) = 1. \tag{20}$$

Under this constraint, the process $y(k)$ is passed through the filter

$$\mathbf{g}(\mathbf{u}_l) \triangleq [g_0(\mathbf{u}_l) \quad g_1(\mathbf{u}_l) \quad \cdots \quad g_{L-1}(\mathbf{u}_l)]^T$$

with no distortion along $\mathbf{u}_l$ and signals along other vectors than $\mathbf{u}_l$ tend to be attenuated. Mathematically, this is equivalent to minimizing the following cost function:

$$J_S[\mathbf{g}(\mathbf{u}_l)] \triangleq \mathbf{g}^H(\mathbf{u}_l)\mathbf{R}_{yy}^{1/2}\mathbf{g}(\mathbf{u}_l) + \mu[1 - \mathbf{g}^H(\mathbf{u}_l)\mathbf{u}_l] \tag{21}$$

where μ is a Lagrange multiplier. The minimization of (21) leads to the following solution:

$$\mathbf{g}(\mathbf{u}_l) = \frac{\mathbf{R}_{yy}^{-1/2}\mathbf{u}_l}{\mathbf{u}_l^H\mathbf{R}_{yy}^{-1/2}\mathbf{u}_l}. \tag{22}$$

We define the spectrum of $y(k)$ along $\mathbf{u}_l$ as

$$\phi_{yy}(\mathbf{u}_l) \triangleq \mathbf{g}^H(\mathbf{u}_l)\mathbf{R}_{yy}\mathbf{g}(\mathbf{u}_l). \tag{23}$$

Substituting (22) into (23) gives

$$\phi_{yy}(\mathbf{u}_l) = \frac{1}{\left(\mathbf{u}_l^H\mathbf{R}_{yy}^{-1/2}\mathbf{u}_l\right)^2}. \tag{24}$$

Expression (24) is a general definition of the spectrum of the signal $y(k)$, which depends on the unitary matrix $\mathbf{U}$. Using (22) and (24), we get

$$\mathbf{R}_{yy}^{1/2}\mathbf{g}(\mathbf{u}_l) = \phi_{yy}^{1/2}(\mathbf{u}_l)\mathbf{u}_l. \tag{25}$$

By taking into account all vectors $\mathbf{u}_l, l = 1, 2, \ldots, L$, (25) can be written into the following general form

$$\mathbf{R}_{yy}^{1/2}\mathbf{G}(\mathbf{U}) = \mathbf{U}\mathbf{\Phi}_{yy}^{1/2}(\mathbf{U}) \tag{26}$$

where

$$\mathbf{G}(\mathbf{U}) \triangleq [\mathbf{g}(\mathbf{u}_1) \quad \mathbf{g}(\mathbf{u}_2) \quad \cdots \quad \mathbf{g}(\mathbf{u}_L)]$$

and

$$\mathbf{\Phi}_{yy}(\mathbf{U}) \triangleq \text{diag}[\phi_{yy}(\mathbf{u}_1) \quad \phi_{yy}(\mathbf{u}_2) \quad \cdots \quad \phi_{yy}(\mathbf{u}_L)]$$

is a diagonal matrix.

Property 1: The correlation matrix $\mathbf{R}_{yy}$ can be diagonalized as follows:

$$\mathbf{G}^H(\mathbf{U})\mathbf{R}_{yy}\mathbf{G}(\mathbf{U}) = \mathbf{\Phi}_{yy}(\mathbf{U}). \tag{27}$$

Proof: This form follows immediately from (26).

Property 1 shows that there are an infinite number of ways to diagonalize the matrix $\mathbf{R}_{yy}$, depending on how we choose the unitary matrix $\mathbf{U}$. Each one of these diagonalizations gives a representation of the spectrum of the signal $y(k)$ in the subspace $\mathbf{U}$. Expression (27) is a generalization of the KLT; the only major difference is that $\mathbf{G}(\mathbf{U})$ is not a unitary matrix except for the case where $\mathbf{U} = \mathbf{Q}$. For this special case, it is easy to verify that $\mathbf{G}(\mathbf{Q}) = \mathbf{Q}$ and $\mathbf{\Phi}_{yy}(\mathbf{Q}) = \mathbf{\Lambda}$, which is the KLT formulation.

Property 2: The vector $\mathbf{y}(k)$ can be written as a combination (expansion) of the vectors of the matrix $\mathbf{U}' = \mathbf{R}_{yy}^{1/2}\mathbf{U}\mathbf{\Phi}_{yy}^{-1/2}(\mathbf{U})$ as follows:

$$\begin{aligned}\mathbf{y}(k) &= \sum_{l=1}^{L} a_y(k, \mathbf{u}_l)\frac{\mathbf{R}_{yy}^{1/2}}{\phi_{yy}^{1/2}(\mathbf{u}_l)}\mathbf{u}_l \\ &= \sum_{l=1}^{L} a_y(k, \mathbf{u}_l)\mathbf{u}_l'\end{aligned} \tag{28}$$

where

$$a_y(k, \mathbf{u}_l) = \mathbf{g}^H(\mathbf{u}_l)\mathbf{y}(k), \quad l = 1, 2, \ldots, L \tag{29}$$

are the coefficients of the expansion. The two previous expressions are the time- and transform-domain representations of the vector signal $\mathbf{y}(k)$.

Proof: Expressions (28) and (29) can be shown by substituting one into the other.

Property 3: We always have

$$E[a_y(k,\mathbf{u}_l)] = 0, \quad l = 1,2,\ldots,L \tag{30}$$

$$\text{and} \quad E[a_y(k,\mathbf{u}_i)a_y^*(k,\mathbf{u}_j)] = \begin{cases} \phi_{yy}(\mathbf{u}_i), & i = j \\ 0, & i \neq j \end{cases} \tag{31}$$

where the superscript * is the complex conjugate operator.

Proof: These properties can be verified from (29).

It can be checked that the Parseval's theorem does not hold anymore if $\mathbf{U} \neq \mathbf{Q}$. This is due to the fact that the matrix $\mathbf{G}(\mathbf{U})$ is not unitary. Indeed

$$\begin{aligned} \sum_{l=1}^{L} E[|a_y(k,\mathbf{u}_l)|^2] &= \sum_{l=1}^{L} \phi_{yy}(\mathbf{u}_l) \\ &= \text{tr}[\mathbf{R}_{yy}\mathbf{G}(\mathbf{U})\mathbf{G}^H(\mathbf{U})] \neq \text{tr}(\mathbf{R}_{yy}). \end{aligned} \tag{32}$$

This is the main difference between the KLT and the generalization proposed here for $\mathbf{U} \neq \mathbf{Q}$. This difference, however, should have no impact on the noise reduction applications and Properties 1, 2, and 3 are certainly the most important ones.

We define the spectra of the clean speech $x(k)$ and noise $v(k)$ in the subspace $\mathbf{U}$ as

$$\phi_{xx}(\mathbf{u}_l) \triangleq \mathbf{g}^H(\mathbf{u}_l)\mathbf{R}_{xx}\mathbf{g}(\mathbf{u}_l), \quad l = 1,2,\ldots,L \tag{33}$$

$$\phi_{vv}(\mathbf{u}_l) \triangleq \mathbf{g}^H(\mathbf{u}_l)\mathbf{R}_{vv}\mathbf{g}(\mathbf{u}_l), \quad l = 1,2,\ldots,L. \tag{34}$$

Of course, $\phi_{xx}(\mathbf{u}_l)$ and $\phi_{vv}(\mathbf{u}_l)$ are always positive real numbers.

We can now apply the three previous properties to our noise reduction problem. Indeed, with the help of Property 2 and substituting (2) into (29), we get

$$\begin{aligned} a_y(k,\mathbf{u}_l) &= \mathbf{g}^H(\mathbf{u}_l)\mathbf{y}(k) \\ &= \mathbf{g}^H(\mathbf{u}_l)\mathbf{x}(k) + \mathbf{g}^H(\mathbf{u}_l)\mathbf{v}(k) \\ &= a_x(k,\mathbf{u}_l) + a_v(k,\mathbf{u}_l), \quad l = 1,2,\ldots,L. \end{aligned} \tag{35}$$

We also have from Property 3 that

$$E[a_y(k,\mathbf{u}_i)a_y^*(k,\mathbf{u}_j)] = \begin{cases} \phi_{xx}(\mathbf{u}_i) + \phi_{vv}(\mathbf{u}_i), & i = j \\ 0, & i \neq j. \end{cases} \tag{36}$$

Expression (35) is equivalent to (2) but in the transform domain. Similar to the KLE case, our problem becomes one of finding an estimate of $a_x(k,\mathbf{u}_l)$ by multiplying $a_y(k,\mathbf{u}_l)$ with a (complex) scalar filter h_l, i.e.,

$$\begin{aligned} a_z(k,\mathbf{u}_l) &= h_l a_y(k,\mathbf{u}_l) \\ &= h_l[a_x(k,\mathbf{u}_l) + a_v(k,\mathbf{u}_l)], \quad l = 1,2,\ldots,L. \end{aligned} \tag{37}$$

From Property 3, we have

$$E[a_z(k,\mathbf{u}_i)a_z^*(k,\mathbf{u}_j)] = \begin{cases} |h_i|^2\phi_{yy}(\mathbf{u}_i), & i = j \\ 0, & i \neq j. \end{cases} \tag{38}$$

Finally by using Property 2 again, we see that an estimate of the vector $\mathbf{x}(k)$ would be

$$\begin{aligned} \mathbf{z}(k) &= \sum_{l=1}^{L} a_z(k,\mathbf{u}_l)\mathbf{u}_l' \\ &= \mathbf{R}_{yy}^{1/2}\left(\sum_{l=1}^{L} h_l\mathbf{u}_l\mathbf{u}_l^H\right)\mathbf{R}_{yy}^{-1/2}\mathbf{y}(k) \\ &= \mathbf{H}(\mathbf{U})[\mathbf{x}(k) + \mathbf{v}(k)] \end{aligned} \tag{39}$$

where

$$\mathbf{H}(\mathbf{U}) \triangleq \mathbf{R}_{yy}^{1/2}\mathbf{U}\,\text{diag}[h_1 \quad h_2 \quad \cdots \quad h_L]\mathbf{U}^H\mathbf{R}_{yy}^{-1/2} \tag{40}$$

is an $L \times L$ (time-domain) filtering matrix, which depends on the unitary matrix $\mathbf{U}$ and is equivalent to the transform-domain filter $\mathbf{h} = [h_1 \quad h_2 \quad \cdots \quad h_L]^T$. Moreover, it can be checked, with the help of Property 1, that the correlation matrix $\mathbf{R}_{zz} = E[\mathbf{z}(k)\mathbf{z}^H(k)]$ can be diagonalized as follows:

$$\begin{aligned} &\mathbf{G}^H(\mathbf{U})\mathbf{R}_{zz}\mathbf{G}(\mathbf{U}) = \text{diag} \\ &\quad [|h_1|^2\phi_{yy}(\mathbf{u}_1) \quad |h_2|^2\phi_{yy}(\mathbf{u}_2) \quad \cdots \quad |h_L|^2\phi_{yy}(\mathbf{u}_L)]. \end{aligned} \tag{41}$$

We see from the previous expression how the coefficients h_l, $l = 1,2,\ldots,L$, affect the spectrum of the estimated signal $z(k)$ in the subspace $\mathbf{U}$, depending on how they are optimized.

V. Performance Measures

In this section, we present some very useful measures that are necessary for designing properly the filters $\mathbf{H}$, $\mathbf{H}(\mathbf{U})$, or $\mathbf{h}$. These definitions will also help us better understand how noise reduction works in the transform 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 desired signal over the intensity of the background noise, i.e.,

$$\text{iSNR} \triangleq \frac{\sigma_x^2}{\sigma_v^2} \tag{42}$$

where $\sigma_x^2 \triangleq E[x^2(k)]$ and $\sigma_v^2 \triangleq E[v^2(k)]$ are the variances of the signals $x(k)$ and $v(k)$, respectively.

With the transform-domain model shown in (35), we define the subband and fullband input SNRs, respectively, as

$$\begin{aligned} \text{iSNR}(\mathbf{u}_l) &\triangleq \frac{E[|a_x(k,\mathbf{u}_l)|^2]}{E[|a_v(k,\mathbf{u}_l)|^2]} \\ &= \frac{\phi_{xx}(\mathbf{u}_l)}{\phi_{vv}(\mathbf{u}_l)}, \quad l = 1,2,\ldots,L \end{aligned} \tag{43}$$

$$\text{iSNR}(\mathbf{U}) \triangleq \frac{\sum_{l=1}^{L}\phi_{xx}(\mathbf{u}_l)}{\sum_{l=1}^{L}\phi_{vv}(\mathbf{u}_l)}. \tag{44}$$

In general, iSNR$(\mathbf{U}) \neq$ iSNR, but for $\mathbf{U} = \mathbf{Q}$, iSNR$(\mathbf{Q}) =$ iSNR.

After noise reduction with the (time-domain) model given in (4), the output SNR can be written as

$$\text{oSNR}(\mathbf{H}) \triangleq \frac{\sigma_{x_{\text{f}}}^2}{\sigma_{v_{\text{rn}}}^2} = \frac{\text{tr}(\mathbf{H}\mathbf{R}_{xx}\mathbf{H}^T)}{\text{tr}(\mathbf{H}\mathbf{R}_{vv}\mathbf{H}^T)}. \tag{45}$$

One of the most important objectives of noise reduction is to improve the SNR after filtering [8], [6]. Therefore, we must design a filter, $\mathbf{H}$, in such a way that oSNR$(\mathbf{H}) \geq \text{SNR}$. For example, with the time-domain Wiener filter [given in (5)], $\mathbf{H}_{\text{W}}$, it was shown that oSNR$(\mathbf{H}_{\text{W}}) \geq \text{SNR}, \quad \forall \text{SNR}$ [8], [6], [18], [28], [29].

After noise reduction with the model given in (39), the output SNR is

$$\text{oSNR}[\mathbf{H}(\mathbf{U})] \triangleq \frac{\text{tr}[\mathbf{H}(\mathbf{U})\mathbf{R}_{xx}\mathbf{H}^H(\mathbf{U})]}{\text{tr}[\mathbf{H}(\mathbf{U})\mathbf{R}_{vv}\mathbf{H}^H(\mathbf{U})]}. \tag{46}$$

Note that this definition is identical to (45). In (46), we only make the output SNR dependent on the unitary matrix $\mathbf{U}$ since the filtering matrix depends on it.

With the transform-domain model shown in (37) and after noise reduction, the subband output SNR is

$$\begin{aligned}\text{oSNR}(\mathbf{u}_l) &\triangleq \frac{|h_l|^2\phi_{xx}(\mathbf{u}_l)}{|h_l|^2\phi_{vv}(\mathbf{u}_l)} \\ &= \text{iSNR}(\mathbf{u}_l), \quad l = 1, 2, \ldots, L\end{aligned} \tag{47}$$

and the fullband output SNR is

$$\text{oSNR}(\mathbf{h}, \mathbf{U}) \triangleq \frac{\sum_{l=1}^{L} |h_l|^2\phi_{xx}(\mathbf{u}_l)}{\sum_{l=1}^{L} |h_l|^2\phi_{vv}(\mathbf{u}_l)}. \tag{48}$$

In general, oSNR$(\mathbf{h}, \mathbf{U}) \neq$ oSNR$[\mathbf{H}(\mathbf{U})]$, but in the special case where $\mathbf{U} = \mathbf{Q}$, we have oSNR$(\mathbf{h}, \mathbf{Q}) =$ oSNR$[\mathbf{H}(\mathbf{Q})]$.

Let c_l and d_l denote two positive real series, it can be shown that

$$\frac{\sum_l c_l}{\sum_l d_l} = \sum_l \left(\frac{c_l}{d_l} \cdot \frac{d_l}{\sum_i d_i} \right) \leq \sum_l \frac{c_l}{d_l}. \tag{49}$$

Using the above inequality, we can verify that

$$\sum_{l=1}^{L} \text{iSNR}(\mathbf{u}_l) \geq \text{iSNR}(\mathbf{U}) \tag{50}$$

$$\sum_{l=1}^{L} \text{oSNR}(\mathbf{u}_l) \geq \text{oSNR}(\mathbf{h}, \mathbf{U}). \tag{51}$$

This means that the aggregation of the subband (input or output) SNRs is greater than or equal to the fullband (input or output) SNR.

Another important measure in noise reduction is the noise-reduction factor, which quantifies the amount of noise being attenuated with the noise reduction filter. With the time-domain formulation in (4), this factor is defined as [8], [6]

$$\xi_{\text{nr}}(\mathbf{H}) \triangleq \frac{\text{tr}(\mathbf{R}_{vv})}{\text{tr}(\mathbf{H}\mathbf{R}_{vv}\mathbf{H}^T)}. \tag{52}$$

By analogy to the previous definition, we define the noise reduction-factor for the model in (39) as

$$\xi_{\text{nr}}[\mathbf{H}(\mathbf{U})] \triangleq \frac{\text{tr}(\mathbf{R}_{vv})}{\text{tr}[\mathbf{H}(\mathbf{U})\mathbf{R}_{vv}\mathbf{H}^H(\mathbf{U})]}. \tag{53}$$

The larger the value of $\xi_{\text{nr}}[\mathbf{H}(\mathbf{U})]$, the more the noise is reduced. After the filtering operation, the residual noise level is expected to be lower than that of the original noise level, therefore this factor should be lower bounded by 1.

In the transform domain with the formulation given in (37), the subband noise-reduction factor can be defined as

$$\xi_{\text{nr}}(h_l) \triangleq \frac{1}{|h_l|^2}, \quad l = 1, 2, \ldots, L \tag{54}$$

and the corresponding fullband noise-reduction factor is

$$\xi_{\text{nr}}(\mathbf{h}, \mathbf{U}) \triangleq \frac{\sum_{l=1}^{L} \phi_{vv}(\mathbf{u}_l)}{\sum_{l=1}^{L} |h_l|^2\phi_{vv}(\mathbf{u}_l)}. \tag{55}$$

In general, $\xi_{\text{nr}}(\mathbf{h}, \mathbf{U}) \neq \xi_{\text{nr}}[\mathbf{H}(\mathbf{U})]$, but for $\mathbf{U} = \mathbf{Q}$, $\xi_{\text{nr}}(\mathbf{h}, \mathbf{Q}) = \xi_{\text{nr}}[\mathbf{H}(\mathbf{Q})]$.

The filtering operation adds distortion to the speech signal; so a measure needs to be introduced to quantify the amount of speech distortion. With the time-domain model in (4), the speech-distortion index is defined as [8], [6]

$$\upsilon_{\text{sd}}(\mathbf{H}) \triangleq \frac{E\{[\mathbf{x}(k) - \mathbf{H}\mathbf{x}(k)]^T[\mathbf{x}(k) - \mathbf{H}\mathbf{x}(k)]\}}{\text{tr}(\mathbf{R}_{xx})}. \tag{56}$$

With the model given in (39), we define the speech-distortion index as

$$\upsilon_{\text{sd}}[\mathbf{H}(\mathbf{U})] \triangleq \frac{E\{[\mathbf{x}(k) - \mathbf{H}(\mathbf{U})\mathbf{x}(k)]^H[\mathbf{x}(k) - \mathbf{H}(\mathbf{U})\mathbf{x}(k)]\}}{\text{tr}(\mathbf{R}_{xx})}. \tag{57}$$

This index is lower bounded by 0 and expected to be upper bounded by 1 for optimal filters. The higher the value of $\upsilon_{\text{sd}}[\mathbf{H}(\mathbf{U})]$, the more the speech is distorted.

Following the same line of ideas, in the transform domain with the formulation given in (37), we define the subband and fullband speech-distortion indices, respectively, as

$$\upsilon_{\text{sd}}(h_l) \triangleq |1 - h_l|^2, \quad l = 1, 2, \ldots, L \tag{58}$$

and

$$\upsilon_{\text{sd}}(\mathbf{h}, \mathbf{U}) \triangleq \frac{\sum_{l=1}^{L} |1 - h_l|^2\phi_{xx}(\mathbf{u}_l)}{\sum_{l=1}^{L} \phi_{xx}(\mathbf{u}_l)}. \tag{59}$$

In general, $\upsilon_{\text{sd}}(\mathbf{h}, \mathbf{U}) \neq \upsilon_{\text{sd}}[\mathbf{H}(\mathbf{U})]$, but for the special case of $\mathbf{U} = \mathbf{Q}$, $\upsilon_{\text{sd}}(\mathbf{h}, \mathbf{Q}) = \upsilon_{\text{sd}}[\mathbf{H}(\mathbf{Q})]$.

We always have

$$\sum_{l=1}^{L} \xi_{\text{nr}}(h_l) \geq \xi_{\text{nr}}(\mathbf{h}, \mathbf{U}) \tag{60}$$

$$\sum_{l=1}^{L} \upsilon_{\text{sd}}(h_l) \geq \upsilon_{\text{sd}}(\mathbf{h}, \mathbf{U}). \tag{61}$$

The two previous inequalities show that the fullband noise-reduction factor and speech-distortion index are upper bounded by values independent of the spectra of the noise and desired speech. It is also interesting to notice that the subband noise-reduction factor and speech-distortion index depend only explicitly on the scalars $h_l,\ l = 1, 2, \ldots, L$, but the corresponding fullband variables depend also on the unitary matrix; this implies that the choice of $\mathbf{U}$ can affect noise reduction and speech distortion.

Although there are many more measures available in the literature, the four measures (input and output SNRs, noise-reduction factor, and speech-distortion index) explained in this section will be primarily used to study, evaluate, or derive optimal or suboptimal filters for noise reduction in the following sections.

VI. Examples of Filter Design in the Transform Domain

In this section, we are going to develop and study the most important single-channel noise reduction filters in the transform domain.

A. Wiener Filter

Let us define the transform-domain error signal between the clean speech and its estimate as follows:

$$\begin{aligned} e(k, \mathbf{u}_l) &\triangleq a_x(k, \mathbf{u}_l) - a_z(k, \mathbf{u}_l) \\ &= a_x(k, \mathbf{u}_l) - h_l a_y(k, \mathbf{u}_l), \quad l = 1, 2, \ldots, L. \end{aligned} \tag{62}$$

The transform-domain MSE is

$$J(\mathbf{u}_l) \triangleq E[|e(k, \mathbf{u}_l)|^2], \quad l = 1, 2, \ldots, L. \tag{63}$$

Taking the gradient of $J(\mathbf{u}_l)$ with respect to h_l^* and equating the result to 0 leads to

$$-E\{a_y^*(k, \mathbf{u}_l)[a_x(k, \mathbf{u}_l) - h_{\mathrm{W},l} a_y(k, \mathbf{u}_l)]\} = 0. \tag{64}$$

Hence

$$\phi_{yy}(\mathbf{u}_l) h_{\mathrm{W},l} = \phi_{xy}(\mathbf{u}_l), \quad l = 1, 2, \ldots, L. \tag{65}$$

The cross-spectrum on the right-hand side of (65) can be written as

$$\begin{aligned} \phi_{xy}(\mathbf{u}_l) &= E[a_x(k, \mathbf{u}_l) a_y^*(k, \mathbf{u}_l)] \\ &= \phi_{xx}(\mathbf{u}_l), \quad l = 1, 2, \ldots, L. \end{aligned} \tag{66}$$

Therefore, the optimal filter can be put into the following forms:

$$h_{\mathrm{W},l} = \frac{\phi_{xx}(\mathbf{u}_l)}{\phi_{yy}(\mathbf{u}_l)} = 1 - \frac{\phi_{vv}(\mathbf{u}_l)}{\phi_{yy}(\mathbf{u}_l)}, \quad l = 1, 2, \ldots, L. \tag{67}$$

We note that the optimal Wiener filter in the transform domain is always real and positive and its form is similar to that of the frequency-domain Wiener filter [4], [30].

Property 4: We have

$$\begin{aligned} &|\rho[a_x(k, \mathbf{u}_l), a_y(k, \mathbf{u}_l)]|^2 + |\rho[a_v(k, \mathbf{u}_l), \\ &\qquad a_y(k, \mathbf{u}_l)]|^2 = 1, \quad l = 1, 2, \ldots, L \end{aligned} \tag{68}$$

where

$$\begin{aligned} &|\rho[a_x(k, \mathbf{u}_l), a_y(k, \mathbf{u}_l)]|^2 \\ &\quad = \frac{|E[a_x(k, \mathbf{u}_l) a_y^*(k, \mathbf{u}_l)]|^2}{E[|a_x(k, \mathbf{u}_l)|^2] E[|a_y(k, \mathbf{u}_l)|^2]} \end{aligned} \tag{69}$$

and

$$\begin{aligned} &|\rho[a_v(k, \mathbf{u}_l), a_y(k, \mathbf{u}_l)]|^2 \\ &\quad = \frac{|E[a_v(k, \mathbf{u}_l) a_y^*(k, \mathbf{u}_l)]|^2}{E[|a_v(k, \mathbf{u}_l)|^2] E[|a_y(k, \mathbf{u}_l)|^2]} \end{aligned} \tag{70}$$

are, respectively, the squared Pearson correlation coefficients (SPCCs) between $a_x(k, \mathbf{u}_l)$ and $a_y(k, \mathbf{u}_l)$, and $a_v(k, \mathbf{u}_l)$ and $a_y(k, \mathbf{u}_l)$.

Proof: From (69) and (70), we have

$$\begin{aligned} &|\rho[a_x(k, \mathbf{u}_l), a_y(k, \mathbf{u}_l)]|^2 \\ &\quad = \frac{\phi_{xx}(\mathbf{u}_l)}{\phi_{yy}(\mathbf{u}_l)} \\ &\quad = \frac{\mathrm{iSNR}(\mathbf{u}_l)}{1 + \mathrm{iSNR}(\mathbf{u}_l)}, \quad l = 1, 2, \ldots, L \end{aligned} \tag{71}$$

and

$$\begin{aligned} &|\rho[a_v(k, \mathbf{u}_l), a_y(k, \mathbf{u}_l)]|^2 \\ &\quad = \frac{\phi_{vv}(\mathbf{u}_l)}{\phi_{yy}(\mathbf{u}_l)} \\ &\quad = \frac{1}{1 + \mathrm{iSNR}(\mathbf{u}_l)}, \quad l = 1, 2, \ldots, L. \end{aligned} \tag{72}$$

Adding (71) and (72) together, we find (68).

Property 4 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.

Property 5: We have

$$h_{\mathrm{W},l} = |\rho[a_x(k, \mathbf{u}_l), a_y(k, \mathbf{u}_l)]|^2 \tag{73}$$

$$= 1 - |\rho[a_v(k, \mathbf{u}_l), a_y(k, \mathbf{u}_l)]|^2, \quad l = 1, 2, \ldots, L. \tag{74}$$

These fundamental forms of the transform-domain Wiener filter, although obvious, do not seem to be known in the literature. They show that they are simply related to two SPCCs. Since $0 \leq |\rho[a_x(k, \mathbf{u}_l), a_y(k, \mathbf{u}_l)]|^2 \leq 1$, then $0 \leq h_{\mathrm{W},l} \leq 1, \forall l$. The Wiener filter acts like a gain function. When the level of noise along $\mathbf{u}_l$ is high $[|\rho[a_v(k, \mathbf{u}_l), a_y(k, \mathbf{u}_l)]|^2 \approx 1]$, then $h_{\mathrm{W},l}$ is close to 0 since there is a large amount of noise that has to be removed. When the level of noise along $\mathbf{u}_l$ is low $[|\rho[a_v(k, \mathbf{u}_l), a_y(k, \mathbf{u}_l)]|^2 \approx 0]$, then $h_{\mathrm{W},l}$ is close to 1 and is not going to affect much the signal since there is little noise that needs to be removed.

We deduce the subband noise-reduction factor and speech-distortion index

$$\xi_{\mathrm{nr}}(h_{\mathrm{W},l}) = \frac{1}{|\rho[a_x(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^4} \geq 1, \quad l = 1,2,\ldots,L, \tag{75}$$

$$\upsilon_{\mathrm{sd}}(h_{\mathrm{W},l}) = |\rho[a_v(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^4 \leq 1, \quad l = 1,2,\ldots,L \tag{76}$$

and the fullband noise-reduction factor and speech-distortion index

$$\xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{W}},\mathbf{U}) = \frac{\sum_{l=1}^{L}\phi_{vv}(\mathbf{u}_l)}{\sum_{l=1}^{L}|\rho[a_x(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^4\phi_{vv}(\mathbf{u}_l)} \geq 1 \tag{77}$$

$$\upsilon_{\mathrm{sd}}(\mathbf{h}_{\mathrm{W}},\mathbf{U}) = \frac{\sum_{l=1}^{L}|\rho[a_v(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^4\phi_{xx}(\mathbf{u}_l)}{\sum_{l=1}^{L}\phi_{xx}(\mathbf{u}_l)} \leq 1. \tag{78}$$

The subband speech-distortion index and noise-reduction factor are related by the formula

$$\upsilon_{\mathrm{sd}}(h_{\mathrm{W},l}) = 1 - \frac{2}{\sqrt{\xi_{\mathrm{nr}}(h_{\mathrm{W},l})}} + \frac{1}{\xi_{\mathrm{nr}}(h_{\mathrm{W},l})}, \quad l = 1,2,\ldots,L. \tag{79}$$

We see clearly how noise reduction and speech distortion depend on the two SPCCs $|\rho[a_x(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^2$ and $|\rho[a_v(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^2$ in the transform-domain Wiener filter. When $|\rho[a_x(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^2$ increases, $\xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{W}},\mathbf{U})$ decreases; at the same time $|\rho[a_v(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^2$ decreases and so does $\upsilon_{\mathrm{sd}}(\mathbf{h}_{\mathrm{W}},\mathbf{U})$.

Property 6: With the optimal transform-domain Wiener filter $\mathbf{h}_{\mathrm{W}}$, the (fullband) output SNR is always greater than or equal to the (fullband) input SNR, i.e., $\mathrm{oSNR}(\mathbf{h}_{\mathrm{W}},\mathbf{U}) \geq \mathrm{iSNR}(\mathbf{U})$.

Proof: Since $\phi_{yy}(\mathbf{u}_l) \geq \phi_{xx}(\mathbf{u}_l) \geq 0$ and $0 \leq |\rho[a_x(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^4 \leq 1, \forall l$, we always have

$$\frac{\sum_{l=1}^{L}|\rho[a_x(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^4\phi_{xx}(\mathbf{u}_l)}{\sum_{l=1}^{L}\phi_{xx}(\mathbf{u}_l)} \geq \frac{\sum_{l=1}^{L}|\rho[a_x(k,\mathbf{q}_l), a_y(k,\mathbf{q}_l)]|^4\phi_{yy}(\mathbf{u}_l)}{\sum_{l=1}^{L}\phi_{yy}(\mathbf{u}_l)} \tag{80}$$

with equality if and only if $|\rho[a_x(k,\mathbf{q}_l), a_y(k,\mathbf{q}_l)]|^4$ is a constant $\forall l$. Substituting $\phi_{yy}(\mathbf{u}_l) = \phi_{xx}(\mathbf{u}_l) + \phi_{vv}(\mathbf{u}_l)$ into the previous expression, we readily obtain

$$\frac{\sum_{l=1}^{L}|\rho[a_x(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^4\phi_{xx}(\mathbf{u}_l)}{\sum_{l=1}^{L}|\rho[a_x(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^4\phi_{vv}(\mathbf{u}_l)} \geq \frac{\sum_{l=1}^{L}\phi_{xx}(\mathbf{u}_l)}{\sum_{l=1}^{L}\phi_{vv}(\mathbf{u}_l)} \tag{81}$$

which means that

$$\mathrm{oSNR}(\mathbf{h}_{\mathrm{W}},\mathbf{U}) \geq \mathrm{iSNR}(\mathbf{U}). \tag{82}$$

Property 6 is fundamental. It shows that the transform-domain Wiener filter is able to improve the (fullband) output SNR of a noisy observed signal for any unitary matrix $\mathbf{U}$.

To finish this study, let us show how the time- and transform-domain Wiener filters are related. With (40) and (67) we can rewrite, equivalently, the transform-domain Wiener filter into the time domain

$$\mathbf{H}_{\mathrm{W}}(\mathbf{U}) = \mathbf{R}_{yy}^{1/2}\mathbf{U}\left[\mathbf{I} - \mathbf{\Phi}_{yy}^{-1}(\mathbf{U})\mathbf{\Phi}_{vv}(\mathbf{U})\right]\mathbf{U}^H\mathbf{R}_{yy}^{-1/2} \tag{83}$$

where

$$\mathbf{\Phi}_{vv}(\mathbf{U}) = \mathrm{diag}[\mathbf{G}^H(\mathbf{U})\mathbf{R}_{vv}\mathbf{G}(\mathbf{U})] \tag{84}$$

is a diagonal matrix whose nonzero elements are the elements of the diagonal of the matrix $\mathbf{G}^H(\mathbf{U})\mathbf{R}_{vv}\mathbf{G}(\mathbf{U})$. Now if we substitute (27) into (5), the time-domain Wiener filter [given in (5)] can be written as

$$\mathbf{H}_{\mathrm{W}} = \mathbf{R}_{yy}^{1/2}\mathbf{U}\left\{\mathbf{I} - \mathbf{\Phi}_{yy}^{-1/2}(\mathbf{U})[\mathbf{G}^H(\mathbf{U})\mathbf{R}_{vv}\mathbf{G}(\mathbf{U})]\mathbf{\Phi}_{yy}^{-1/2}(\mathbf{U})\right\}\mathbf{U}^H\mathbf{R}_{yy}^{-1/2}. \tag{85}$$

It is clearly seen that if the matrix $\mathbf{G}^H(\mathbf{U})\mathbf{R}_{vv}\mathbf{G}(\mathbf{U})$ is diagonal, the two filters $\mathbf{H}_{\mathrm{W}}$ and $\mathbf{H}_{\mathrm{W}}(\mathbf{U})$ are identical. In this scenario, it would not matter which unitary matrix we choose.

B. Parametric Wiener Filtering

Some applications may need aggressive noise reduction, while others on the contrary may require little 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 [31], [32]. The equivalent approach in the transform domain is

$$h_{\mathrm{G},l} = [1 - |\rho[a_v(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^{\beta_1}]^{\beta_2}, \quad l = 1,2,\ldots,L, \tag{86}$$

where β_1 and β_2 are two positive parameters that allow the control of this compromise. For $(\beta_1,\beta_2) = (2,1)$, we get the transform-domain Wiener filter developed in the previous section. Taking $(\beta_1,\beta_2) = (2,1/2)$ leads to

$$h_{\mathrm{P},l} = \sqrt{1 - |\rho[a_v(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^2} = |\rho[a_x(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]| = \sqrt{\frac{\mathrm{iSNR}(\mathbf{u}_l)}{1+\mathrm{iSNR}(\mathbf{u}_l)}}, \quad l = 1,2,\ldots,L \tag{87}$$

which is the equivalent form of the power subtraction method studied in [31]–[35]. The pair $(\beta_1,\beta_2) = (1,1)$ gives the equivalent form of the magnitude subtraction method [36]–[40]

$$h_{\mathrm{M},l} = 1 - |\rho[a_v(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|$$

$$= 1 - \sqrt{1 - |\rho[a_x(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^2}$$
$$= 1 - \frac{1}{\sqrt{1 + \text{iSNR}(\mathbf{u}_l)}}, \quad l = 1, 2, \ldots, L. \quad (88)$$

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

$$\xi_{\text{nr}}(h_{\text{P},l}) = \frac{1}{|\rho[a_x(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^2}, \quad l = 1, 2, \ldots, L, \quad (89)$$
$$\xi_{\text{nr}}(h_{\text{M},l}) = \frac{1}{[1 - \sqrt{1 - |\rho[a_x(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^2}]^2}, \quad l = 1, 2, \ldots, L \quad (90)$$

and the corresponding subband speech-distortion indices are

$$\upsilon_{\text{sd}}(h_{\text{P},l}) = [1 - \sqrt{1 - |\rho[a_v(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^2}]^2, \quad l = 1, 2, \ldots, L \quad (91)$$
$$\upsilon_{\text{sd}}(h_{\text{M},l}) = |\rho[a_v(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^2, \quad l = 1, 2, \ldots, L. \quad (92)$$

We can also show that

$$\xi_{\text{nr}}(h_{\text{M},l}) \geq \xi_{\text{nr}}(h_{\text{W},l}) \geq \xi_{\text{nr}}(h_{\text{P},l}), \quad l = 1, 2, \ldots, L \quad (93)$$
$$\upsilon_{\text{sd}}(h_{\text{P},l}) \leq \upsilon_{\text{sd}}(h_{\text{W},l}) \leq \upsilon_{\text{sd}}(h_{\text{M},l}), \quad l = 1, 2, \ldots, L. \quad (94)$$

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 [30], but at the same time it's the one that will likely adds most distortion to the speech signal. The least aggressive approach is the power subtraction while the Wiener filter is between the two others in terms of speech distortion and noise reduction. Since $0 \leq h_{\text{G},l} \leq 1$, then $\text{oSNR}(\mathbf{h}_{\text{G}}, \mathbf{U}) \geq \text{iSNR}(\mathbf{U})$. Therefore, all three methods improve the (fullband) output SNR. Other variants of these algorithms can be found in [41], [42].

The two particular transform-domain filters derived above can be rewritten, equivalently, into the time domain.

- Power subtraction:

$$\mathbf{H}_{\text{P}}(\mathbf{U}) = \mathbf{R}_{yy}^{1/2}\mathbf{U}\left[\mathbf{I} - \mathbf{\Phi}_{yy}^{-1}(\mathbf{U})\mathbf{\Phi}_{vv}(\mathbf{U})\right]^{1/2}\mathbf{U}^H\mathbf{R}_{yy}^{-1/2}. \quad (95)$$

- Magnitude subtraction:

$$\mathbf{H}_{\text{M}}(\mathbf{U}) = \mathbf{R}_{yy}^{1/2}\mathbf{U}\left[\mathbf{I} - \mathbf{\Phi}_{yy}^{-1/2}(\mathbf{U})\mathbf{\Phi}_{vv}^{1/2}(\mathbf{U})\right]\mathbf{U}^H\mathbf{R}_{yy}^{-1/2}. \quad (96)$$

These two filters are, of course, not optimal in any sense but they can be very practical.

C. Tradeoff Filter

The error signal defined in (62) can be rewritten as follows:

$$e(k,\mathbf{u}_l) = e_x(k,\mathbf{u}_l) - e_v(k,\mathbf{u}_l), \quad l = 1, 2, \ldots, L \quad (97)$$

where

$$e_x(k,\mathbf{u}_l) \triangleq (1 - h_l)a_x(k,\mathbf{u}_l), \quad l = 1, 2, \ldots, L \quad (98)$$

is the speech distortion due to the linear transformation, and

$$e_v(k,\mathbf{u}_l) \triangleq h_l a_v(k,\mathbf{u}_l), \quad l = 1, 2, \ldots, L \quad (99)$$

represents the residual noise [9].

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

$$\min_{h_l} J_x(h_l,\mathbf{u}_l) \quad \text{subject to} \quad J_v(h_l,\mathbf{u}_l) = \beta_l\phi_{vv}(\mathbf{u}_l), \quad l = 1, 2, \ldots, L \quad (100)$$

where

$$J_x(h_l,\mathbf{u}_l) \triangleq E[|e_x(k,\mathbf{u}_l)|^2], \quad l = 1, 2, \ldots, L \quad (101)$$
$$J_v(h_l,\mathbf{u}_l) \triangleq E[|e_v(k,\mathbf{u}_l)|^2], \quad l = 1, 2, \ldots, L \quad (102)$$

and $0 < \beta_l < 1$ in order to have some noise reduction. Using a Lagrange multiplier, $\mu_l (\geq 0)$, to adjoin the constraint to the cost function, we can derive the optimal filter:

$$h_{\text{T},l} = \frac{\phi_{xx}(\mathbf{u}_l)}{\phi_{xx}(\mathbf{u}_l) + \mu_l\phi_{vv}(\mathbf{u}_l)} = \frac{\phi_{yy}(\mathbf{u}_l) - \phi_{vv}(\mathbf{u}_l)}{\phi_{yy}(\mathbf{u}_l) + (\mu_l - 1)\phi_{vv}(\mathbf{u}_l)} = \frac{1 - |\rho[a_v(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^2}{1 + (\mu_l - 1)|\rho[a_v(k,\mathbf{u}_l), a_y(k,\mathbf{u}_l)]|^2}, \quad l = 1, 2, \ldots, L. \quad (103)$$

Hence, $h_{\text{T},l}$ is a Wiener filter with adjustable input noise level $\mu_l\phi_{vv}(\mathbf{u}_l)$. It can be shown that this optimal filter is closely related to the subspace approach [9], [14], [15], [43], [44]. Since $0 \leq h_{\text{T},l} \leq 1$, $\forall \mu_l \geq 0$, then $\text{oSNR}(\mathbf{h}_{\text{T}}, \mathbf{U}) \geq \text{iSNR}(\mathbf{U})$. Therefore, this method improves the (fullband) output SNR.

The Lagrange multiplier must satisfy

$$J_v(h_{\text{T},l},\mathbf{u}_l) = \beta_l\phi_{vv}(\mathbf{u}_l) = h_{\text{T},l}^2\phi_{vv}(\mathbf{u}_l), \quad l = 1, 2, \ldots, L. \quad (104)$$

Substituting (103) into (104), we can find

$$\mu_l = \text{iSNR}(\mathbf{u}_l)\left(\frac{1}{\sqrt{\beta_l}} - 1\right), \quad l = 1, 2, \ldots, L \quad (105)$$

and from (104), we also have

$$h_{\text{T},l} = \sqrt{\beta_l}, \quad l = 1, 2, \ldots, L. \quad (106)$$

The Lagrange multiplier μ_l can always be chosen in an ad-hoc way if we prefer. Then, we can see from (103) that there are four cases.

- $\mu_l = 1$; in this case, the tradeoff and Wiener filters are the same, i.e., $h_{\mathrm{T},l} = h_{\mathrm{W},l}$.
- $\mu_l = 0$; in this circumstance, we have $h_{\mathrm{T},l} = 1$ and there will be no noise reduction and no speech distortion as well.
- $\mu_l > 1$; this situation corresponds to a more aggressive (as compared to the Wiener filter) noise reduction, at the expense of higher speech distortion.
- $\mu_l < 1$; this case corresponds to a more conservative noise reduction (as compared to the Wiener filter) with less noise reduction and also less speech distortion.

With (40) and (106) we can rewrite, equivalently, the transform-domain tradeoff filter into the time domain:

$$\mathbf{H}_{\mathrm{T}}(\mathbf{U}) = \mathbf{R}_{yy}^{1/2}\mathbf{U}\,\mathrm{diag}^{1/2}[\beta_1 \quad \beta_2 \quad \cdots \quad \beta_L]\mathbf{U}^H\mathbf{R}_{yy}^{-1/2}. \tag{107}$$

D. Examples of Unitary Matrices

There are perhaps a very large number of unitary (or orthogonal) matrices that can be used in tandem with the different noise reduction filters presented in this section, but does a transformation exist in such a way that an optimal filter maximizes noise reduction while minimizing speech distortion at the same time? The answer to this question is not straightforward. However, intuitively we believe that some unitary matrices will be more effective than others for a given noise reduction filter.

The first obvious choice is the KLT developed in Section III. In this case, $\mathbf{U} = \mathbf{Q}$ where $\mathbf{Q}$ contains the eigenvectors of the correlation matrix $\mathbf{R}_{yy}$ of the noisy signal $y(k)$ for which the spectral representation are the eigenvalues of $\mathbf{R}_{yy}$. This choice seems to be the most natural one since the Parseval's theorem is verified.

Another choice for $\mathbf{U}$ is the Fourier matrix

$$\mathbf{F} = [\mathbf{f}_1 \quad \mathbf{f}_2 \quad \cdots \quad \mathbf{f}_L] \tag{108}$$

where

$$\mathbf{f}_l = \frac{1}{\sqrt{L}}[1 \quad \exp(j\omega_l) \quad \cdots \quad \exp[j\omega_l(L-1)]]^T \tag{109}$$

and $\omega_l = 2\pi(l-1)/L, l = 1, \ldots, L$. Even though $\mathbf{F}$ is unitary, the matrix $\mathbf{G}(\mathbf{F})$ constructed from $\mathbf{F}$ is not; as a result, the Parseval's theorem does not hold but the transform signals at the different frequencies are uncorrelated. Filters in this new Fourier domain will probably perform differently as compared to the classical frequency-domain filters.

In our application, the signal $y(k)$ is real and it may be more convenient to select an orthogonal matrix instead of a unitary one. So another choice close to the previous one is the discrete cosine transform

$$\mathbf{C} = [\mathbf{c}_1 \quad \mathbf{c}_2 \quad \cdots \quad \mathbf{c}_L] \tag{110}$$

where

$$\mathbf{c}_l = \left[c(1) \;\; c(2)\cos\frac{\pi(2l-1)}{2L} \quad \cdots \quad c(L)\cos\frac{\pi(2l-1)(L-1)}{2L}\right]^T \tag{111}$$

with $c(1) = \sqrt{1/L}$ and $c(l) = \sqrt{2/L}$ for $l \neq 1$. We can verify that $\mathbf{C}^T\mathbf{C} = \mathbf{C}\mathbf{C}^T = \mathbf{I}$.

One other important option is to take $\mathbf{U} = \mathbf{I}$ (the identity matrix). The matrix $\mathbf{G}(\mathbf{I})$ derived from this choice is a kind of an interpolation matrix [4] of the noisy signal and the spectrum

$$\phi_{yy}(\mathbf{i}_l) = \frac{1}{\left(\mathbf{i}_l^T\mathbf{R}_{yy}^{-1/2}\mathbf{i}_l\right)^2} \tag{112}$$

is the interpolation error power (with $\mathbf{i}_l$ being the lth column of $\mathbf{I}$). Therefore, if the signal is predictable along $\mathbf{i}_l$ (meaning that speech is dominant), $\phi_{yy}(\mathbf{i}_l)$ will be small and h_l should be chosen close to 1. On the other hand, if the signal is not predictable along $\mathbf{i}_l$ (meaning that noise is dominant), $\phi_{yy}(\mathbf{i}_l)$ will be large and h_l should be chosen close to 0.

Other possible choices for $\mathbf{U}$ are Hadamard and Haar transforms.

VII. Simulations

We have formulated the noise reduction problem in a generalized transform domain and discussed the design of different optimal and tradeoff noise reduction filters in that domain. In this section, we study different filters through experiments and compare different transforms and their impact on noise reduction performance.

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. The overall length of the signal is 30 seconds. The noisy speech is obtained by adding noise to the clean speech (the noise signal is properly scaled to control the input SNR level). We considered two types of noise: a computer generated white Gaussian random process and a babbling noise signal recorded in a New York Stock Exchange (NYSE) room. The NYSE noise is also digitized with a sampling rate of 8 kHz. Compared with the Gaussian random noise which is stationary and white, the NYSE noise tends to be nonstationary and colored. It consists of sounds from various sources such as electrical fans, telephone rings, and even some speech from background speakers. See [45] for some statistics of this babbling noise.

A. Estimation of the Correlation Matrices

The most critical information that we need to estimate are the correlation matrices $\mathbf{R}_{yy}$ and $\mathbf{R}_{vv}$. Since the noisy signal is accessible, $\mathbf{R}_{yy}$ can be estimated from its definition in Section II by approximating the mathematical expectation with the 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. Another widely used way to estimate $\mathbf{R}_{yy}$ is through the recursive approach, where an estimate of $\mathbf{R}_{yy}$ at time k is obtained as

$$\mathbf{R}_{yy}(k) = \alpha_y\mathbf{R}_{yy}(k-1) + (1-\alpha_y)\mathbf{y}(k)\mathbf{y}^T(k) \tag{113}$$

where α_y is a forgetting factor that controls the influence of the previous data samples on the current estimate of the noisy correlation matrix.

We have learned, through experimental study, that the short-term average and the recursive methods can produce similar noise reduction performance if the parameters associated with each approach are properly optimized, but in general the recursive approach given in (113) is easier to tune up. Therefore, this method will be adopted in our experiments. In order to obtain an initial estimate of $\mathbf{R}_{yy}$, we separate the 30-s-long noisy signal into two parts. The first part lasts 5 s and a long-term average is applied to this to compute an initial estimate of $\mathbf{R}_{yy}$. The second part lasts 25 s and is used for performance evaluation.

The noise statistics can be estimated in many different ways using a noise estimator [2], [6], [46]–[50]. In this study, however, we intend not to use any noise estimator, but compute the noise correlation matrix directly from the noise signal using either a long-term average (for stationary noise) or a recursive method [similar to the estimation of $\mathbf{R}_{yy}$ in (113), but with a different forgetting factor α_v]. The reason for this is that we want to study the optimal values of the parameters used in the different noise reduction filters and the effect of different transforms on noise reduction performance. To find the optimal values of those parameters and the transform most suited for noise reduction, it is better to simplify the experiments and avoid the influence due to noise estimation error.

B. Performance of the Wiener Filter in Stationary Noise

With the recursive estimation of the correlation matrices, the performance of the Wiener filter given in (83) is mainly affected by three major elements: forgetting factors (α_y and α_v), frame length L, and transform matrix $\mathbf{U}$. In the first experiment, we study the effect of the forgetting factors with different transforms. White noise is used in this experiment and the input SNR is 10 dB. Since this noise is stationary, we computed the noise correlation matrix using a long-term average. We also fixed the frame length to $L = 32$. With this setup, the noise reduction performance is only affected by the transform matrix $\mathbf{U}$ and the forgetting factor α_y. For the matrix $\mathbf{U}$, we choose to compare five widely used transforms: KL, Fourier, cosine, Hadamard, and identity. The value of α_y should be in the range between 0 and 1. Within this range, α_y should not be too small, otherwise, a large error would occur in the $\mathbf{R}_{yy}(k)$ estimate, causing performance degradation. In addition, a small α_y may make the estimated $\mathbf{R}_{yy}(k)$ matrix ill-conditioned (with a large condition number), thereby causing numerical problems when we attempt to compute the inverse of this matrix. To circumvent this problem, we computed the Moore-Penrose pseudoinverse of this matrix instead of its direct inverse in our implementation. Of course, α_y cannot be set too large (close to its upper bound 1) either. Otherwise, the recursive estimation will essentially be a long-term average and will not be able to follow the short-term variations of the speech signal, which limits the noise reduction performance. The optimal value α_y will be determined through experiments. Fig. 1 plots the output SNR and the speech-distortion index for different transforms as a function of the forgetting factor α_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 $\mathbf{x}_{\mathrm{f}}(k)$ and residual noise $\mathbf{v}_{\mathrm{rn}}(k)$, and the output SNR and speech distortion index are then computed according to (46) and (57), respectively].

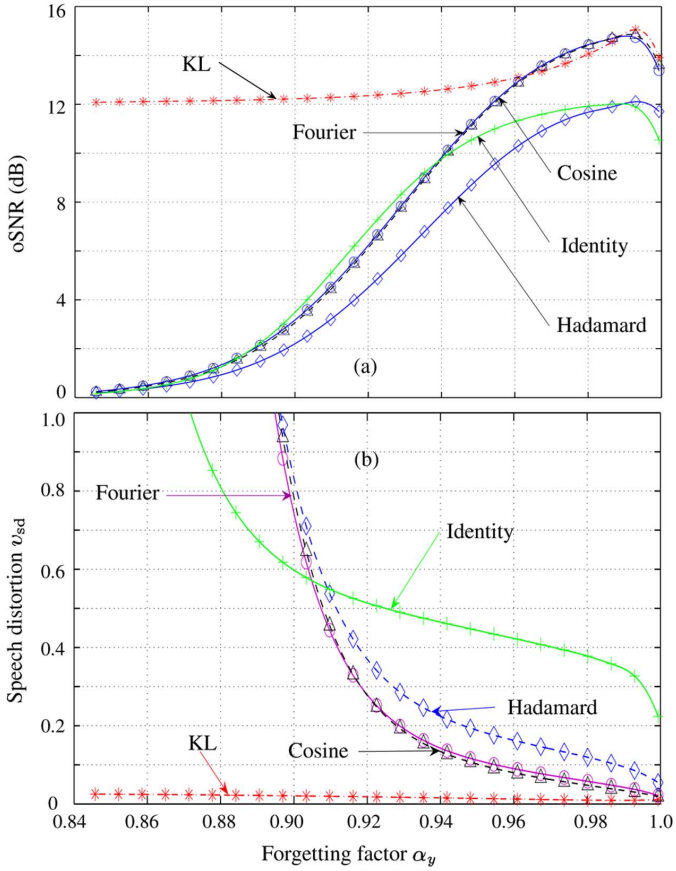


Fig. 1. Noise reduction performance of the Wiener filter versus α_y in white Gaussian noise: iSNR = 10 dB and $L = 32$.

It is seen from Fig. 1 that the output SNR for all the studied transforms first increases with α_y, and then decreases. The highest output SNR is obtained when α_y is between 0.985 and 0.995. This coincides with our intuition that α_y has to be large enough for accurate estimation of $\mathbf{R}_{yy}$, but meanwhile it cannot be too close to 1 so that the correlation estimate can follow the variation of the speech signal. Unlike the output SNR, the speech-distortion index υ_{sd} for all the five transforms bears a monotonic relationship with the parameter α_y. The larger the value of α_y, the smaller the speech distortion index. This can be explained by the following fact: as α_y increases, the estimation variation of the matrix $\mathbf{R}_{yy}$ decreases, thereby leading to less speech distortion.

We also see from Fig. 1 that the Fourier and cosine transforms yielded almost the same performance. When α_y is reasonably large (e.g., ≥ 0.96), the KL, Fourier, and cosine transforms produced similar output SNRs. Comparatively, however, the KL transform has a much lower speech-distortion index. In addition, the KL transform can improve the SNR while maintaining a lower level of speech distortion even when α_y is small (e.g., ≤ 0.94), but when α_y is small, both the Fourier and cosine transforms yielded negative SNR gain with tremendous speech distortion. This result indicates that the KL transform is more immune to the estimation error of $\mathbf{R}_{yy}$. When the value of α_y is in a reasonable range (e.g., $\in [0.96, 1]$), the Hadamard and

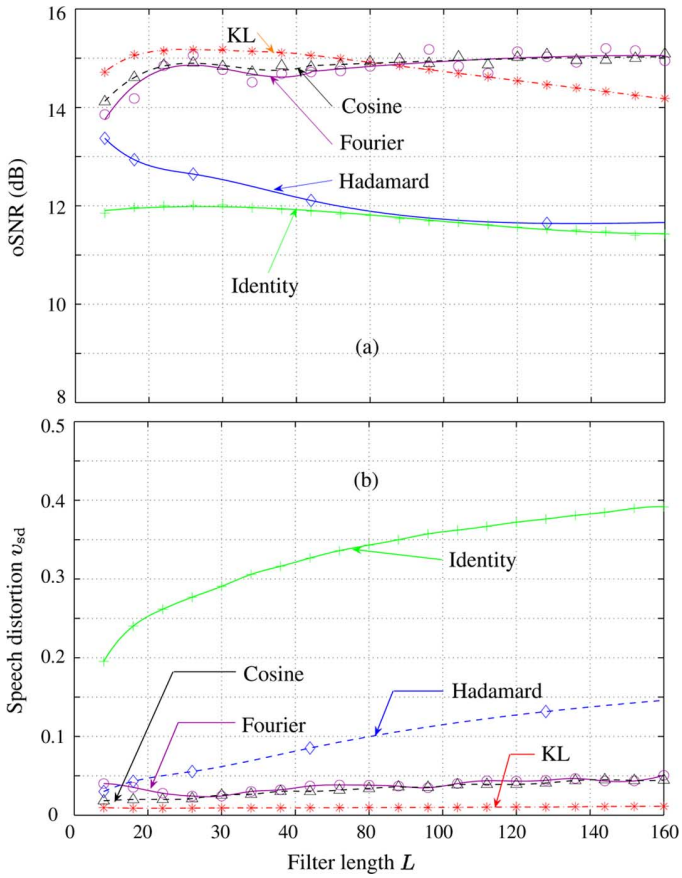


Fig. 2. Noise reduction performance of the Wiener filter versus L in white Gaussian noise: iSNR $= 10$ dB and $\alpha_y = 0.99$.

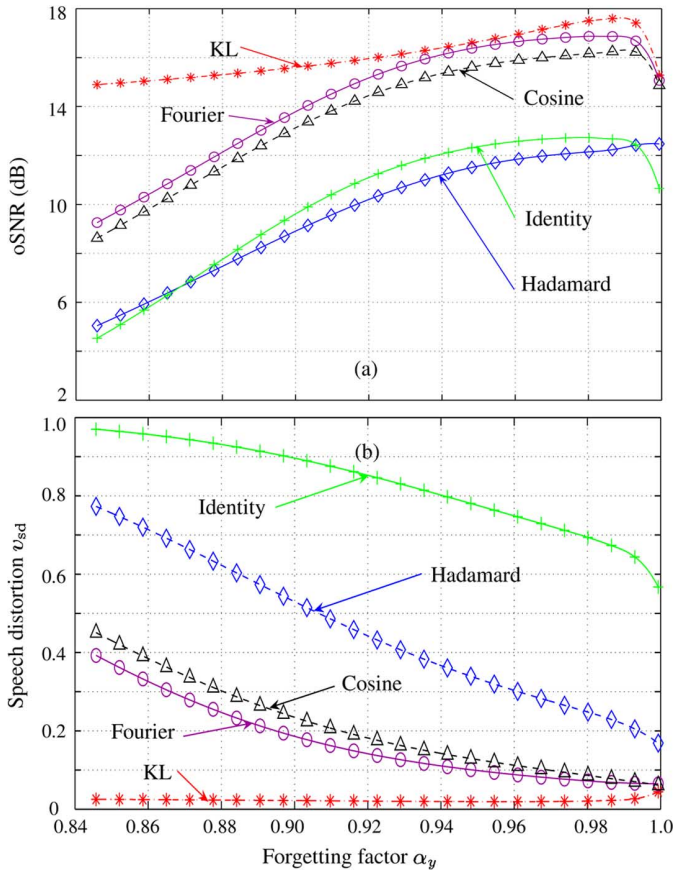


Fig. 3. Noise reduction performance of the tradeoff filter versus α_y in white Gaussian noise: iSNR $= 10$ dB, $L = 32$, and $\mu = 4$.

identity transforms can also improve the SNR, but their performance is generally inferior to that of the other three transforms.

In the second experiment, we study the effect of the frame length L on the noise reduction performance. Same as the previous experiment, white noise is used and iSNR $= 10$ dB. Again, the noise correlation matrix is computed using a long-term average. Based on the previous results, we set $\alpha_y = 0.99$. Fig. 2 depicts the output SNR and speech-distortion index, both as a function of L. It is seen that, as L increases, the output SNR of the Wiener filter using the KL transform first increases, and then decreases. Good performance with this transform is obtained when L is in the range of 20–40. This result agrees with what we observed in our previous studies [6]. The reason for this can be explained in terms of speech predictability. It is widely known, from speech production and analysis theory, that a speech signal can be well modeled with a low-rank prediction (or generally an interpolation) model, which is especially true for the quasi-steady voiced regions of speech in which a prediction model of order 10–20 provides a good approximation to the vocal tract spectral envelope. During unvoiced and transit regions of speech, the prediction model is less effective than for voiced regions, but it still provides an acceptable model for speech if the model order is increased. Usually, a prediction model between 20–40 is sufficient to model a speech signal. Therefore, we see that good performance is achieved when L is in that range. Further increasing L does not improve modeling accuracy, but leading to a larger error in the $\mathbf{R}_{yy}$ estimate, which causes performance degradation.

The Fourier and cosine transforms yielded similar performance, particularly when $L > 20$. Both the output SNR and speech-distortion index with these two transforms slightly increase with L (up to 160). For $L > 100$, these two transforms even produced a higher output SNR than the KL transform with the same L value. However, the speech-distortion index with these two transforms are also higher than that of the KL transform. In addition, the largest SNR gain with these two transforms (achieved when L is around 160) is similar to that of the KL transform achieved with a smaller L.

While the output SNR of the identity transform is almost invariant with respect to L, the speech-distortion index increases significantly with L. For the Hadamard transform, a larger L corresponds to a less SNR gain and a larger speech-distortion index, which indicates that a small frame length L should be preferred if Hadamard transform is used. Generally, however, both the identity and Hadamard transforms are much inferior to the KL, Fourier, and cosine transforms in performance.

C. *Performance of the Tradeoff Filter in Stationary Noise*

In the next experiment, we evaluate the performance of the transform-domain tradeoff filter given in (107) in different conditions. From the analysis shown in Section VI-C, we already see that if $\mu = 1$, the tradeoff filter is the Wiener filter. Increasing the value of μ will give more noise reduction, but will also lead to more speech distortion. In this experiment, we set $\mu = 4$. Again, the noise used is a white Gaussian random process and iSNR $= 10$ dB. The noise correlation matrix is

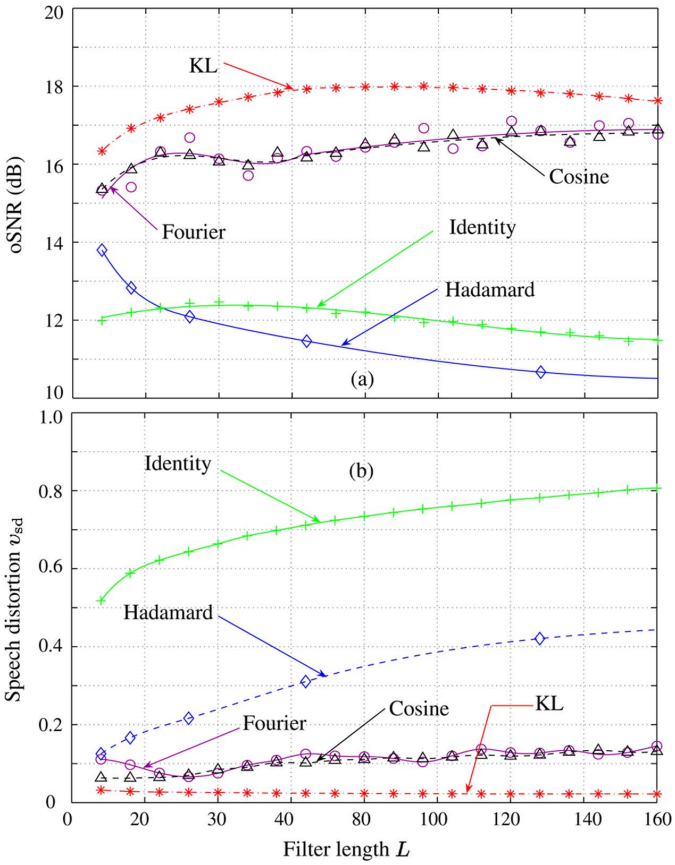


Fig. 4. Noise reduction performance of the tradeoff filter versus L in white Gaussian noise: iSNR $= 10$ dB, $\alpha_y = 0.99$, and $\mu = 4$.

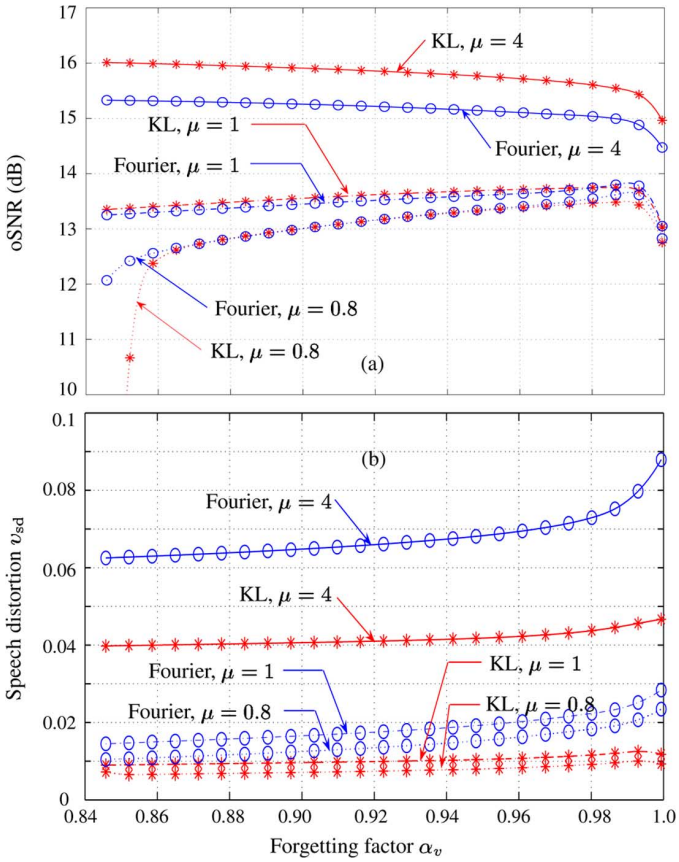


Fig. 5. Noise reduction performance of the tradeoff filter versus α_v in NYSE noise: iSNR $= 10$ dB, $\alpha_y = 0.99$, and $L = 32$.

computed using a long-term average. We first fix the frame length L to 32 and investigate the effect of α_y and different transforms on the performance. Fig. 3 portrays the output SNR and speech-distortion index as a function α_y.

Similar to the Wiener filter case, the output SNR (for all the studied transforms) first increases and then drops as α_y increases. The largest SNR gain for each transform is obtained when α_y is between 0.985 and 0.995. The KL transform yielded the best performance (with the highest output SNR and lowest speech-distortion index). The Fourier and cosine transforms behave similarly. When α_y is in the range between 0.93 and 1, these two transforms can achieve an output SNR similar to that of the KL transform, but their speech-distortion index is higher than that of the KL transform. The identity and Hadamard transforms produce similar output SNR, but the former has a much higher speech-distortion index. In general, the performance of these two transforms is relatively poor as compared to the other three transforms, again, indicating that these two transforms are less effective for the purpose of noise reduction.

Comparing Figs. 1 and 3, one can see that the output SNR of the tradeoff filter is boosted with a large μ, but this is achieved at the price of adding more speech distortion, which confirms the analysis presented in Section VI-C.

To investigate the effect of the frame length L on performance, we set $\alpha_y = 0.99$ and change L from 4 to 160. All other conditions are the same as used in the previous experiment. The results are shown in Fig. 4. Similar to the Wiener-filter case, we observe that the output SNR for the KL transform first increases to its maximum and then drops as L increases. However, there are two major differences as compared to the Wiener-filter case: 1) the near-optimal performance with the tradeoff filter appears when L is in the range of 40–120, while such performance occurs when L in the range of 20–40 for the Wiener filter; 2) although the performance with the KL transform decreases if we keep increasing L after the optimal performance is achieved, the performance degradation with L is almost negligible. The reason for these two differences can be explained as follows. In our experiment, we set $\mu = 4$, and all the β_l in the diagonal matrix $\text{diag}^{1/2}[\beta_1\beta_2\cdots\beta_L]$ that are less than 0 are forced to 0. After a certain threshold, if we further increase L, the dimension of the signal subspace that consists of all the positive β_l value does not increase much. In other words, even though we increases L, which results in a larger size for $\mathbf{R}_{yy}$, we are still dealing with a signal subspace of similar order. As a result, the performance does not change much. Again, the Fourier and cosine transforms have similar performance. Comparatively, the effect of L on the Fourier, cosine, Hadamard, and identity transforms in the tradeoff-filter case is almost the same as that in the Wiener-filter situation. The only difference is that now we have achieved a higher SNR gain, but the speech distortion is also higher.

D. Performance of the Tradeoff Filter in Nonstationary Noise

In the last experiment, we examine the tradeoff filter in the NYSE noise conditions. Since this noise is nonstationary, the

recursive method is used to estimate the noise correlation matrix. From the previous study, we set $\alpha_y = 0.99$, $L = 32$, and iSNR $= 10$ dB. The results of this experiment are depicted in Fig. 5. For a clear presentation, we excluded the results using the identity, Hardamard, cosine transforms since the former two yielded much poorer performance, and the cosine transform delivered a performance similar to that of the Fourier transform. It is seen that when μ is small (1 and 0.8), the KL and Fourier transforms yielded a similar SNR gain, but when μ is increased to 4, the KL transform achieves a higher output SNR. However, the speech-distortion index with the Fourier transform is always higher than that of the KL transform. In addition, for $\mu = 1$ and 0.8, the output SNR bears a nonmonotonic relationship with α_v, with the highest SNR is obtained when α_v is approximately 0.993. It is also seen that when $\mu = 4$, a small α_v is preferred.

VIII. Conclusion

This paper has focused on the noise reduction problem for speech applications. We have formulated the problem as one of optimal filtering in a generalized transform domain, where any unitary (or orthogonal) matrix can be used to construct the forward (for analysis) and inverse (for synthesis) transforms. We have demonstrated some advantages of working in this generalized domain, including different transforms can be used to replace each other without any requirement to change the algorithm (optimal filter) formulation, and it is easier to fairly compare different transforms for their noise reduction performance. We have addressed the design of different optimal and suboptimal filters in such a generalized transform domain, including the Wiener filter, the parametric Wiener filter, tradeoff filter, etc. We have also compared, through experiments, five different transforms (KL, Fourier, cosine, Hadamard, and identity) for their noise reduction performance. In general, the KL transform yielded the best performance. The Fourier and cosine transforms have quite similar performance, which is slightly inferior to that of the KL transform. While Hadamard and identity transforms can improve the SNR, their speech distortion is very high as compared to the other three studied transforms.

References

[1] *Microphone Arrays*, M. Brandstein and D. Ward, Eds.. Berlin, Germany: Springer, 2001.

[2] J. Chen, Y. Huang, and J. Benesty, "Filtering techniques for noise reduction and speech enhancement," in *Adaptive Signal Processing: Applications to Real-World Problems*, J. Benesty and Y. Huang, Eds. Berlin, Germany: Springer, 2003, pp. 129–154.

[3] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO Signal Processing*. Berlin, Germany: Springer, 2006.

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

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

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

[7] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[8] J. Benesty, J. Chen, Y. Huang, and S. Doclo, "Study of the Wiener filter for noise reduction," in *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds. Berlin, Germany: Springer-Verlag, 2005, pp. 9–41.

[9] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 4, pp. 251–266, Jul. 1995.

[10] M. Dendrinos, S. Bakamidis, and G. Garayannis, "Speech enhancement from noise: A regenerative approach," *Speech Commun.*, vol. 10, pp. 45–57, Feb. 1991.

[11] H. Lev-Ari and Y. Ephraim, "Extension of the signal subspace speech enhancement approach to colored noise," *IEEE Signal Process. Lett.*, vol. 10, no. 4, pp. 104–106, Apr. 2003.

[12] A. Rezayee and S. Gazor, "An adaptive KLT approach for speech enhancement," *IEEE Trans. Speech Audio Process*, vol. 9, no. 2, pp. 87–95, Feb. 2001.

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

[14] Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 4, pp. 334–341, Jul. 2003.

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

[16] P. Loizou, *Speech Enhancement: Theory and Practice*. Boca Raton, FL: CRC, 2007.

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

[18] J. Chen, J. Benesty, and Y. Huang, "On the optimal linear filtering techniques for noise reduction," *Speech Commun.*, vol. 49, pp. 305–316, 2007.

[19] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[20] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2002.

[21] J. Huang and Y. Zhao, "Energy-constrained signal subspace method for speech enhancement and recognition," *IEEE Signal Process. Lett.*, vol. 4, no. 10, pp. 283–285, Oct. 1997.

[22] F. Jabloun and B. Champagne, "Signal subspace techniques for speech enhancement," in *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds. Berlin, Germany: Springer-Verlag, 2005, pp. 135–159.

[23] J. Benesty, J. Chen, and Y. Huang, "A generalized MVDR spectrum," *IEEE Signal Process. Lett.*, vol. 12, no. 12, pp. 827–830, Dec. 2005.

[24] I. Santamaría and J. Vía, "Estimation of the magnitude squared coherence spectrum based on reduced-rank canonical coordinates," in *Proc. IEEE ICASSP*, 2007, pp. III-985–III-988.

[25] L. L. Scharf and J. T. Thomas, "Wiener filters in canonical coordinates for transform coding, filtering, and quantizing," *IEEE Trans. Signal Process.*, vol. 46, pp. 647–654, Mar. 1998.

[26] C. Zheng, M. Zhou, and X. Li, "On the relationship of non-parametric methods for coherence function estimation," *Elsevier Signal Process.*, vol. 11, pp. 2863–2867, Nov. 2008.

[27] R. M. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends in Commun. Inf. Theory*, vol. 2, pp. 155–239, 2006.

[28] S. Doclo and M. Moonen, "On the output SNR of the speech-distortion weighted multichannel Wiener filter," *IEEE Signal Process. Lett.*, vol. 12, no. 12, pp. 809–811, Dec. 2005.

[29] J. Benesty, J. Chen, and Y. Huang, "On the importance of the Pearson correlation coefficient in noise reduction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 4, pp. 575–765, May 2008.

[30] E. J. Diethorn, , Y. Huang and J. Benesty, Eds., "Subband noise reduction methods for speech enhancement," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*. Boston, MA: Kluwer, 2004, pp. 91–115.

[31] W. Etter and G. S. Moschytz, "Noise reduction by noise-adaptive spectral magnitude expansion," *J. Audio Eng. Soc.*, vol. 42, pp. 341–349, May 1994.

[32] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, no. 12, pp. 1586–1604, Dec. 1979.

[33] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. ASSP-28, no. 2, pp. 137–145, Apr. 1980.

[34] M. M. Sondhi, C. E. Schmidt, and L. R. Rabiner, "Improving the quality of a noisy speech signal," *Bell Syst. Tech. J.*, vol. 60, pp. 1847–1859, Oct. 1981.

[35] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.

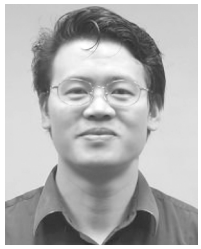
[36] M. R. Schroeder, "Apparatus for suppressing noise and distortion in communication signals," U.S. patent 3,180,936, Dec. 1, 1960, issued Apr. 27, 1965.

[37] M. R. Schroeder, "Processing of communication signals to reduce effects of noise," U.S. patent 3,403,224, May 28, 1965, issued Sep. 24, 1968.
[38] M. R. Weiss, E. Aschkenasy, and T. W. Parsons, "Processing speech signals to attenuate interference," in *Proc. IEEE Symp. Speech Recognition*, 1974, pp. 292–295.
[39] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. IEEE ICASSP*, 1979, pp. 208–211.
[40] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. ASSP-27, no. 2, pp. 113–120, Apr. 1979.
[41] J. H. L. Hansen, "Speech enhancement employing adaptive boundary detection and morphological based spectral constraints," in *Proc. IEEE ICASSP*, 1991, pp. 901–904.
[42] B. L. Sim, Y. C. Tong, J. S. Chang, and C. T. Tan, "A parametric formulation of the generalized spectral subtraction method," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 4, pp. 328–337, Jul. 1998.
[43] Y. Hu and P. C. Loizou, "A subspace approach for enhancing speech corrupted by colored noise," *IEEE Signal Process. Lett.*, vol. 9, no. 7, pp. 204–206, Jul. 2002.
[44] 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, pp. 195–195, 2007.
[45] Y. Huang, J. Benesty, and J. Chen, "Analysis and comparison of multichannel noise reduction methods in a common framework," *IEEE Trans. Audio, Speech. Lang. Process.*, vol. 16, no. 5, pp. 957–968, Jul. 2008.
[46] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process*, vol. 9, no. 5, pp. 504–512, Jul. 2001.
[47] H. G. Hirsch and C. Ehrlicher, "Noise estimation techniques for robust speech recognition," in *Proc. IEEE ICASSP*, 1995, vol. 1, pp. 153–156.
[48] V. Stahl, A. Fischer, and R. Bippus, "Quantile based noise estimation for spectral subtraction and Wiener filtering," in *Proc. IEEE ICASSP*, 2000, vol. 3, pp. 1875–1878.
[49] N. W. D. Evans and J. S. Mason, "Noise estimation without explicit speech, non-speech detection: A comparison of mean, modal and median based approaches," in *Proc. Eurospeech*, 2001, vol. 2, pp. 893–896.
[50] E. J. Diethorn, "A subband noise-reduction method for enhancing speech in telephony and teleconferencing," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust.*, 1997.

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, France, in 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 was with Telecom Paris University, working 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 and 2008 Best Paper Awards 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 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).



Jingdong Chen (M'99) received the B.S. and M.S. degrees in electrical engineering from the Northwestern Polytechnic University, Xiaan, 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 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, signal 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 coeditor/coauthor of the book *Speech Enhancement* (Springer-Verlag, 2005) and a section editor of the reference *Springer Handbook of Speech Processing* (Springer-Verlag, 2007).

Dr. Chen is currently an Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, a member of the IEEE Audio and Electroacoustics Technical Committee, and 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 is the technical Co-Chair of the 2009 WASPAA. He received the 2008 Best Paper Award from the IEEE Signal Processing Society, the 1998–1999 Research Grant Award from the Japan Key Technology Center, and the 1996–1998 President's Award from the Chinese Academy of Sciences.

Yiteng (Arden) Huang (S'97–M'01) received the B.S. degree from the Tsinghua University, Beijing, China, in 1994 and 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.

From March 2001 to January 2008, he was a Member of Technical Staff at Bell Laboratories, Murray Hill, NJ. In January 2008, he joined the WeVoice, Inc., Bridgewater, NJ and served as its CTO. His current research interests are in acoustic signal processing and multimedia communications.

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 is a coeditor/coauthor of the books *Noise Reduction in Speech Processing* (Springer-Verlag, 2009) *Microphone Array Signal Processing*(Springer-Verlag, 2008), *Springer Handbook of Speech Processing* (Springer-Verlag, 2007), *Acoustic MIMO Signal Processing* (Springer-Verlag, 2006), *Audio Signal Processing for Next-Generation Multimedia Communication Systems* (Kluwer, 2004) and *Adaptive Signal Processing: Applications to Real-World Problems* (Springer-Verlag, 2003). He received the 2008 Best Paper Award and 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.