On the Importance of the Pearson Correlation Coefficient in Noise Reduction

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

***Abstract*—Noise reduction, which aims at estimating a clean speech from noisy observations, has attracted a considerable amount of research and engineering attention over the past few decades. In the single-channel scenario, an estimate of the clean speech can be obtained by passing the noisy signal picked up by the microphone through a linear filter/transformation. The core issue, then, is how to find an optimal filter/transformation such that, after the filtering process, the signal-to-noise ratio (SNR) is improved but the desired speech signal is not noticeably distorted. Most of the existing optimal filters (such as the Wiener filter and subspace transformation) are formulated from the mean-square error (MSE) criterion. However, with the MSE formulation, many desired properties of the optimal noise-reduction filters such as the SNR behavior cannot be seen. In this paper, we present a new criterion based on the Pearson correlation coefficient (PCC). We show that in the context of noise reduction the squared PCC (SPCC) has many appealing properties and can be used as an optimization cost function to derive many optimal and suboptimal noise-reduction filters. The clear advantage of using the SPCC over the MSE is that the noise-reduction performance (in terms of the SNR improvement and speech distortion) of the resulting optimal filters can be easily analyzed. This shows that, as far as noise reduction is concerned, the SPCC-based cost function serves as a more natural criterion to optimize as compared to the MSE.**

***Index Terms*—Mean-square error (MSE), noise reduction, Pearson correlation coefficient, speech distortion, speech enhancement, Wiener filter.**

I. Introduction

THE PROBLEM of noise reduction, as its name indicates, is to suppress the unwanted noise, thereby enhancing a desired speech signal in terms of speech quality and intelligibility. The difficulty of this problem varies from application to application, and also depends on the system configuration. In general, if more microphone channels are available, the problem would be easier to solve, at least theoretically. For example, when an array of microphones can be used, a beam can be formed and steered to a desired direction. As a result, the signal propagating from the desired direction will be passed through without degradation, while signals originating from all other directions will either suffer a certain amount of attenuation or be completely rejected [1]–[3]. In the two-microphone case, with one microphone picking up the noisy signal and the other measuring the noise field, we can use the latter microphone signal as a noise reference and eliminate the noise in the former microphone by means of adaptive cancellation [4]–[8]. However, most of today's communication devices are equipped with only one microphone. In this case, noise reduction becomes a very difficult problem since no reference of the noise is accessible and the clean speech cannot be preprocessed prior to being corrupted by noise. Due to its wide range of potential applications, however, this single-channel noise-reduction problem has attracted a significant amount of research attention, and is also the focus of this paper.

In the single-channel scenario, the observed microphone signal can be modeled as a superposition of the clean speech and noise. In order to reduce the level of noise, we need to pass the microphone signal through a filter/transformation [9]–[24]. Normally, we only consider linear filters/transformations since the nonlinear ones are much more difficult to design and analyze. So, the problem of noise reduction is to find an optimal linear filter such that, after the filtering process, the signal-to-noise ratio (SNR) would be improved; in other words, the microphone 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 forming the optimal filter. Currently, most existing approaches, such as the Wiener filter and the subspace method, are derived from the mean-square-error (MSE) criterion. Although the MSE criterion has been well studied and understood, the application of this principle to noise reduction often masks many important properties of the noise-reduction filter. For example, it is difficult to see the SNR behavior since there is no direct connection between the SNR and the MSE criterion.

In this paper, we present a new principle based on the Pearson correlation coefficient (PCC). PCC is a statistical metric that measures the strength and direction of a linear relationship between two random variables [25]–[27]. It has been widely used in many applications such as time-delay estimation [28], pattern recognition [29], data analysis, to name a few. In this paper, we discuss the use of the PCC for noise reduction, which has not been addressed before. We will introduce the concept of the squared PCC (SPCC) and discuss many interesting properties of the SPCC associated with the clean speech, noise, noisy speech, and clean speech estimate. We show that in the context of noise reduction, minimizing the MSE is equivalent to maximizing the SPCC between the clean speech and its estimate. Similar to the MSE, we can derive the Wiener filter and many other optimal filters with the SPCC-based criterion. The clear advantage of using this new criterion over the MSE is that the noise-reduction performance (in terms of the SNR improvement and speech

Manuscript received July 12, 2007; revised January 23, 2008.The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hiroshi Sawada.

J. Benesty is with the Université du Québec, INRS-EMT, Montréal, QC H5A 1K6, Canada.

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

Y. Huang is with the 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.2008.919072

distortion) of the resulting optimal filters can be easily analyzed. This shows that, for the problem of noise reduction, the SPCC serves as a more natural criterion to optimize as compared to the MSE.

II. Problem Formulation and Definitions

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

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

where $v(k)$ is the unwanted additive noise, which is assumed to be a zero-mean random process (white or colored) and uncorrelated with $x(k)$. An estimate of $x(k)$ can be obtained by passing $y(k)$ through a linear filter, i.e.,

$$\begin{aligned} z(k) &= \mathbf{h}^T \mathbf{y}(k) \\ &= \mathbf{h}^T [\mathbf{x}(k) + \mathbf{v}(k)] \end{aligned} \tag{2}$$

where

$$\mathbf{h} = [h_0 \quad h_1 \quad \cdots \quad h_{L-1}]^T$$

is a finite-impulse response (FIR) filter of length L, superscript ^{T} denotes transpose of a vector or a matrix, and

$$\begin{aligned} \mathbf{y}(k) &= [y(k) \quad y(k-1) \quad \cdots \quad y(k-L+1)]^T \\ \mathbf{x}(k) &= [x(k) \quad x(k-1) \quad \cdots \quad x(k-L+1)]^T \\ \mathbf{v}(k) &= [v(k) \quad v(k-1) \quad \cdots \quad v(k-L+1)]^T. \end{aligned}$$

With this formulation, the objective of noise reduction is to find an optimal filter that would attenuate the noise as much as possible while keeping the distortion of the clean speech low. This apparently requires the filter length to satisfy $L \geq 2$ since there will be no noise reduction, but just a volume adjustment if $L = 1$. In addition, the filter $\mathbf{h}$ cannot be $\mathbf{0}$, otherwise the output would be muted off.

Note that all three random variables x, y, and z as given in (1) and (2) are linearly related. This observation is key in order to be able to use the PCC as a criterion to estimate the optimal noise-reduction filter. However, before we discuss this criterion, let us first give some important measures that will be extensively used later on.

One of the most frequently used measures in noise reduction is the SNR. With the signal model given in (1), the input SNR is defined as the intensity of the signal of interest relatively to the intensity of the background noise, i.e.,

$$\text{SNR} \triangleq \frac{\sigma_x^2}{\sigma_v^2} \tag{3}$$

where $\sigma_x^2 = E[x^2(k)]$ and $\sigma_v^2 = E[v^2(k)]$ are the variances of the signals $x(k)$ and $v(k)$, respectively, with $E[\cdot]$ denoting mathematical expectation.

After noise reduction, the output SNR can be written as [30], [31]

$$\text{SNR}(\mathbf{h}) \triangleq \frac{\mathbf{h}^T \mathbf{R}_{xx} \mathbf{h}}{\mathbf{h}^T \mathbf{R}_{vv} \mathbf{h}} \tag{4}$$

where

$$\begin{aligned} \mathbf{R}_{xx} &= E\left[\mathbf{x}(k)\mathbf{x}^T(k)\right] \\ \mathbf{R}_{vv} &= E\left[\mathbf{v}(k)\mathbf{v}^T(k)\right] \end{aligned}$$

are the covariance matrices of the signals $x(k)$ and $v(k)$, respectively. The most important goal of noise reduction is to improve the SNR. Therefore, we must have $\text{SNR}(\mathbf{h}) > \text{SNR}$.

Another important measure is the noise-reduction factor, which is defined as the ratio between the power of the original noise and that of the residual noise after filtering [30], [31]

$$\xi_{\text{nr}}(\mathbf{h}) \triangleq \frac{\sigma_v^2}{\mathbf{h}^T \mathbf{R}_{vv} \mathbf{h}}. \tag{5}$$

The larger the value of $\xi_{\text{nr}}(\mathbf{h})$, 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. So, this factor should be lower bounded by 1.

The filtering operation will distort the speech signal. To measure the amount of speech distortion, we use the concept of speech-distortion index, which is defined as the attenuation in speech power relatively to the power of the original clean speech [30], [31]

$$\begin{aligned} \upsilon_{\text{sd}}(\mathbf{h}) &\triangleq \frac{E\left\{\left[x(k) - \mathbf{h}^T \mathbf{x}(k)\right]^2\right\}}{\sigma_x^2} \\ &= \frac{(\mathbf{h}_1 - \mathbf{h})^T \mathbf{R}_{xx} (\mathbf{h}_1 - \mathbf{h})}{\mathbf{h}_1^T \mathbf{R}_{xx} \mathbf{h}_1} \end{aligned} \tag{6}$$

where

$$\mathbf{h}_1 = [1 \quad 0 \quad \cdots \quad 0]^T. \tag{7}$$

This speech-distortion index is lower bounded by 0 and expected to be upper bounded by 1. The larger the value of $\upsilon_{\text{sd}}(\mathbf{h})$, the more the speech is distorted. The last two definitions (noise-reduction factor and speech-distortion index) are directly derived from the MSE criterion.

III. Pearson Correlation Coefficient

Let x and y be two zero-mean real-valued random variables. The Pearson correlation coefficient (PCC) is defined as [25]–[27]

$$\rho(x, y) = \frac{E[xy]}{\sigma_x \sigma_y} \tag{8}$$

where $E[xy]$ is the cross-correlation between x and y, and $\sigma_x^2 = E[x^2]$ and $\sigma_y^2 = E[y^2]$ are the variances of the signals x and y, respectively. In the context of noise reduction, it will be more convenient to work with the squared Pearson correlation coefficient (SPCC)

$$\rho^2(x, y) = \frac{E^2[xy]}{\sigma_x^2 \sigma_y^2}. \tag{9}$$

One of the most important properties of the SPCC is that

$$0 \leq \rho^2(x, y) \leq 1. \tag{10}$$

The SPCC gives an indication on the strength of the linear relationship between the two random variables x and y. If $\rho^2(x,y) = 0$, then x and y are said to be uncorrelated. The closer the value of $\rho^2(x,y)$ is to 1, the stronger the correlation between the two variables. If the two variables are independent, then $\rho^2(x,y) = 0$. However, the converse is not true because the SPCC detects only *linear* dependencies between the two variables x and y. For a nonlinear dependency, the SPCC may be equal to zero. However, in the special case when x and y are jointly normal, "independent" is equivalent to "uncorrelated."

IV. Relations Between Noise Reduction and SPCC

In this section, we discuss many interesting properties regarding the SPCCs among the four signals x, v, y, and z: some of these properties are related to noise reduction, and others may provide an indication of the degree of speech distortion.

The SPCC between $x(k)$ and $y(k)$ [as defined in (1)] is

$$\rho^2(x,y) = \frac{\sigma_x^2}{\sigma_y^2} = \frac{\text{SNR}}{1+\text{SNR}} \tag{11}$$

where $\sigma_y^2 = E[y^2(k)] = \sigma_x^2 + \sigma_v^2$ is the variance of the signal $y(k)$.

The SPCC between $x(k)$ and $z(k)$ [as defined in (2)] is

$$\begin{aligned}\rho^2(x,z) &= \frac{\left(\mathbf{h}_1^T\mathbf{R}_{xx}\mathbf{h}\right)^2}{\sigma_x^2(\mathbf{h}^T\mathbf{R}_{yy}\mathbf{h})}\\ &= \frac{\left(\mathbf{h}_1^T\mathbf{R}_{xx}\mathbf{h}\right)^2}{\sigma_x^2(\mathbf{h}^T\mathbf{R}_{xx}\mathbf{h})}\cdot\frac{\text{SNR}(\mathbf{h})}{1+\text{SNR}(\mathbf{h})}\end{aligned} \tag{12}$$

where $\mathbf{R}_{yy} = E[\mathbf{y}(k)\mathbf{y}^T(k)] = \mathbf{R}_{xx} + \mathbf{R}_{vv}$ is the covariance matrix of the signal $y(k)$.

Property 1: We have

$$\rho^2(x,z) = \rho^2(x,\mathbf{h}^T\mathbf{y}) = \rho^2(x,\mathbf{h}^T\mathbf{x})\cdot\rho^2(\mathbf{h}^T\mathbf{x},\mathbf{h}^T\mathbf{y}) \tag{13}$$

where

$$\rho^2(x,\mathbf{h}^T\mathbf{x}) = \frac{\left(\mathbf{h}_1^T\mathbf{R}_{xx}\mathbf{h}\right)^2}{\sigma_x^2(\mathbf{h}^T\mathbf{R}_{xx}\mathbf{h})} \tag{14}$$

and

$$\rho^2(\mathbf{h}^T\mathbf{x},\mathbf{h}^T\mathbf{y}) = \frac{\text{SNR}(\mathbf{h})}{1+\text{SNR}(\mathbf{h})}. \tag{15}$$

The SPCC $\rho^2(x,\mathbf{h}^T\mathbf{x})$ can be viewed as a speech-distortion index. If $\mathbf{h} = \mathbf{h}_1$ (no speech distortion) then $\rho^2(x,\mathbf{h}^T\mathbf{x}) = 1$. The closer the value of $\rho^2(x,\mathbf{h}^T\mathbf{x})$ is to 0, the more the speech signal is distorted (except for the simple delay filter). The SPCC $\rho^2(\mathbf{h}^T\mathbf{x},\mathbf{h}^T\mathbf{y})$ shows the SNR improvement, so it can be viewed as a noise-reduction index that reaches its maximum when $\text{SNR}(\mathbf{h})$ is maximized.

Property 1 is fundamental in the noise-reduction problem. It shows that the SPCC $\rho^2(x,\mathbf{h}^T\mathbf{y})$, which can be a cost function as explained later, is simply the product of two important indices reflecting noise reduction and speech distortion. In contrast, the MSE (see next section) has a much more complex form with no real physical meaning in the context of noise reduction.

Property 2: We have

$$\rho^2(x,z) \le \frac{\text{SNR}(\mathbf{h})}{1+\text{SNR}(\mathbf{h})} \tag{16}$$

with equality when $\mathbf{h} = \mathbf{h}_1$.

Proof: This property follows immediately from (12) since $\rho^2(x,\mathbf{h}^T\mathbf{x}) \le 1$.

Property 3: We have

$$\rho^2(\mathbf{h}^T\mathbf{x},y) = \rho^2(x,\mathbf{h}^T\mathbf{x})\cdot\rho^2(x,y). \tag{17}$$

The SPCC $\rho^2(\mathbf{h}^T\mathbf{x},y)$ can be viewed as a speech-distortion measure since $\rho^2(x,\mathbf{h}^T\mathbf{x})$ evaluates the degree of speech distortion due to the filtering process.

Property 4: We have

$$\rho^2(\mathbf{h}^T\mathbf{x},y) \le \frac{\text{SNR}}{1+\text{SNR}} \tag{18}$$

with equality when $\mathbf{h} = \mathbf{h}_1$.

Proof: This property follows immediately from (17) since $\rho^2(x,\mathbf{h}^T\mathbf{x}) \le 1$.

Property 4 shows that speech distortion is unavoidable when we achieve noise reduction and the amount of speech distortion depends on the input SNR. The lower the input SNR, the more the speech distortion.

The SPCC between $v(k)$ and $y(k)$ [as defined in (1)], which, like the SNR, measures how noisy the observation signal, is

$$\rho^2(v,y) = \frac{\sigma_v^2}{\sigma_y^2} = \frac{1}{1+\text{SNR}}. \tag{19}$$

The SPCC between $v(k)$ and $z(k)$ [as defined in (2)], which measures the amount of noise reduction, is

$$\begin{aligned}\rho^2(v,z) &= \frac{\left(\mathbf{h}_1^T\mathbf{R}_{vv}\mathbf{h}\right)^2}{\sigma_v^2(\mathbf{h}^T\mathbf{R}_{yy}\mathbf{h})}\\ &= \frac{\left(\mathbf{h}_1^T\mathbf{R}_{vv}\mathbf{h}\right)^2}{\sigma_v^2(\mathbf{h}^T\mathbf{R}_{vv}\mathbf{h})}\cdot\frac{1}{1+\text{SNR}(\mathbf{h})}.\end{aligned} \tag{20}$$

Property 5: We have

$$\rho^2(v,z) = \rho^2(v,\mathbf{h}^T\mathbf{y}) = \rho^2(v,\mathbf{h}^T\mathbf{v})\cdot\rho^2(\mathbf{h}^T\mathbf{v},\mathbf{h}^T\mathbf{y}) \tag{21}$$

where

$$\rho^2(v,\mathbf{h}^T\mathbf{v}) = \frac{\left(\mathbf{h}_1^T\mathbf{R}_{vv}\mathbf{h}\right)^2}{\sigma_v^2(\mathbf{h}^T\mathbf{R}_{vv}\mathbf{h})} \tag{22}$$

and

$$\rho^2(\mathbf{h}^T\mathbf{v},\mathbf{h}^T\mathbf{y}) = \frac{1}{1+\text{SNR}(\mathbf{h})}. \tag{23}$$

Property 6: We have

$$\rho^2(v,z) \le \frac{1}{1+\text{SNR}(\mathbf{h})} \tag{24}$$

with equality when $\mathbf{h} = \mathbf{h}_1$. The proof of this property follows immediately from (21). The SPCC $\rho^2(v,z)$ measures the amount of noise reduction.

Property 7: We have

$$\rho^2(\mathbf{h}^T\mathbf{v}, y) = \rho^2(v, \mathbf{h}^T\mathbf{v}) \cdot \rho^2(v, y). \tag{25}$$

Property 8: We have

$$\rho^2(\mathbf{h}^T\mathbf{v}, y) \le \frac{1}{1+\text{SNR}} \tag{26}$$

with equality when $\mathbf{h} = \mathbf{h}_1$. The proof of this property follows immediately from (25). The SPCC $\rho^2(\mathbf{h}^T\mathbf{v}, y)$ measures the amount of noise reduction. The lower the value of $\rho^2(\mathbf{h}^T\mathbf{v}, y)$, the more the noise reduction.

Property 9: We have

$$\text{SNR}(\mathbf{h}) = \frac{\rho^2(\mathbf{h}^T\mathbf{x}, \mathbf{h}^T\mathbf{y})}{\rho^2(\mathbf{h}^T\mathbf{v}, \mathbf{h}^T\mathbf{y})}. \tag{27}$$

We have discussed different forms of the SPCC and its properties. In the next section, we will show that the SPCC can be used as a criterion to derive different optimal filters for noise reduction. Many of the properties shown here are relevant and will help us better understand the fundamental role of the SPCC in the application of noise reduction.

V. Examples of Optimal Filters Derived From the SPCC

Intuitively, the noise-reduction problem can be formulated as one of finding the filter that maximizes the SPCC $\rho^2(x, z)$ in order to make the clean speech signal $x(k)$ and the filter output signal $z(k)$ correlated as much as possible. Furthermore, since the SPCC $\rho^2(x, z)$ is the product of two other SPCCs $\rho^2(x, \mathbf{h}^T\mathbf{x})$ and $\rho^2(\mathbf{h}^T\mathbf{x}, \mathbf{h}^T\mathbf{y})$ (see Property 1), we can find other forms of optimal filters that maximize either one of these two SPCCs with or without constraints.

A. Speech Distortionless Filter

Speech distortion is always of great concern in noise reduction. Let us try to find an optimal filter that can pass the desired speech signal without creating distortion. As explained in Section IV, the SPCC $\rho^2(x, \mathbf{h}^T\mathbf{x})$ is a speech-distortion index, so the problem becomes one of finding a filter that maximizes $\rho^2(x, \mathbf{h}^T\mathbf{x})$.

Let us assume that the covariance matrix $\mathbf{R}_{xx}$ is either full rank or it is rank deficient but $\mathbf{h}$ is not in its null space (if $\mathbf{h}$ is in the null space of $\mathbf{R}_{xx}$, we have $\rho^2(x, \mathbf{h}^T\mathbf{x}) = 0$, which is minimized instead of being maximized). We know that

$$\rho^2(x, \mathbf{h}^T\mathbf{x}) \le 1 \tag{28}$$

where equality holds if and only if $\mathbf{h} = \alpha\mathbf{h}_1$, and α is a nonzero (real) constant. Therefore, $\mathbf{h} = \alpha\mathbf{h}_1$ is the optimal speech-distortionless filter. In order to illustrate this solution, we consider a simple example where we have a speech signal corrupted by some noise recorded in a small office environment and we use a filter with only two taps, i.e., $L = 2$, to obtain noise reduction. The sampling rate is 16 kHz. The signal covariance matrix is computed as

$$\mathbf{R}_{xx} = \begin{bmatrix} 1.0 & 0.6 \\ 0.6 & 1.0 \end{bmatrix}. \tag{29}$$

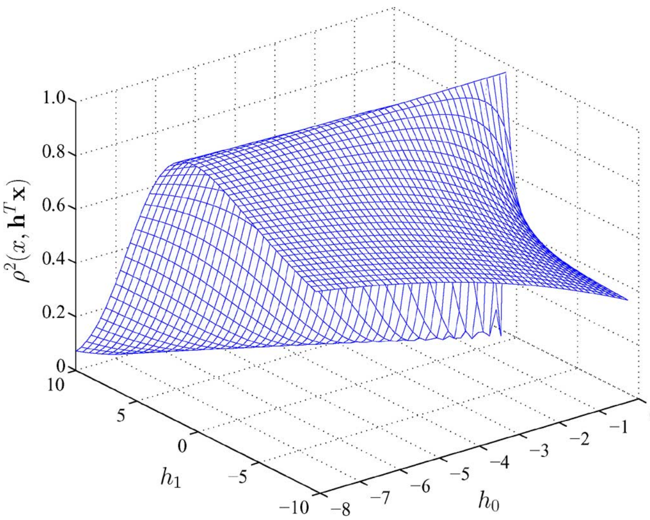

Fig. 1. The SPCC $\rho^2(x, \mathbf{h}^T\mathbf{x})$ versus a two-tap filter $\mathbf{h} = [h_0\ h_1]^T$.

Fig. 1 shows a three-dimensional plot of the SPCC $\rho^2(x, \mathbf{h}^T\mathbf{x})$ as a function of h_0 and h_1. It is clearly seen that $\rho^2(x, \mathbf{h}^T\mathbf{x})$ reaches its maximum when $h_1 = 0$ (h_0 can take any value but 0).

Note that the constant α does not affect the noise-reduction performance (it merely changes the volume of the output signal). If we require the filtered and clean speech signals have the same power (in other words, set $\alpha = 1$), the distortionless filter becomes $\mathbf{h} = \mathbf{h}_1$. In this case, we have

$$\text{SNR}(\mathbf{h}_1) = \text{SNR} \tag{30}$$

$$\rho^2\left(x, \mathbf{h}_1^T\mathbf{x}\right) = 1 \tag{31}$$

$$z_1(k) = y(k). \tag{32}$$

The degenerated filter $\mathbf{h}_1$ has no impact on either the clean speech or the noise. In other words, the speech-distortionless filter $\mathbf{h} = \mathbf{h}_1$ will not distort the clean speech signal but will not improve the output SNR either.

B. Maximum SNR Filter

The major objective of noise reduction is to reduce noise, thereby improving the SNR. Now let us find a filter that can maximize the output SNR. It is easy to see from (15) that maximizing the output SNR, $\text{SNR}(\mathbf{h})$, is equivalent to maximizing $\rho^2(\mathbf{h}^T\mathbf{x}, \mathbf{h}^T\mathbf{y})$. It is easy to check that

$$\rho^2(\mathbf{h}^T\mathbf{x}, \mathbf{h}^T\mathbf{y}) = \frac{\mathbf{h}^T\mathbf{R}_{xx}\mathbf{h}}{\mathbf{h}^T\mathbf{R}_{yy}\mathbf{h}} = \frac{\mathbf{h}^T\mathbf{R}_{xx}\mathbf{h}}{\mathbf{h}^T\mathbf{R}_{xx}\mathbf{h} + \mathbf{h}^T\mathbf{R}_{vv}\mathbf{h}}. \tag{33}$$

In practice, the covariance matrices $\mathbf{R}_{yy}$ and $\mathbf{R}_{vv}$ are in general full rank (which is assumed to be true in this study). Therefore, maximizing $\rho^2(\mathbf{h}^T\mathbf{x}, \mathbf{h}^T\mathbf{y})$ is equivalent to solving the generalized eigenvalue problem

$$\mathbf{R}_{xx}\mathbf{h} = \lambda\mathbf{R}_{vv}\mathbf{h}. \tag{34}$$

The optimal solution to this problem is $\alpha\mathbf{h}_{\max}$, where α, again, is a nonzero (real) constant, and $\mathbf{h}_{\max}$ is the eigenvector corresponding to the maximum eigenvalue $\lambda_{\max}$ of $\mathbf{R}_{vv}^{-1}\mathbf{R}_{xx}$.

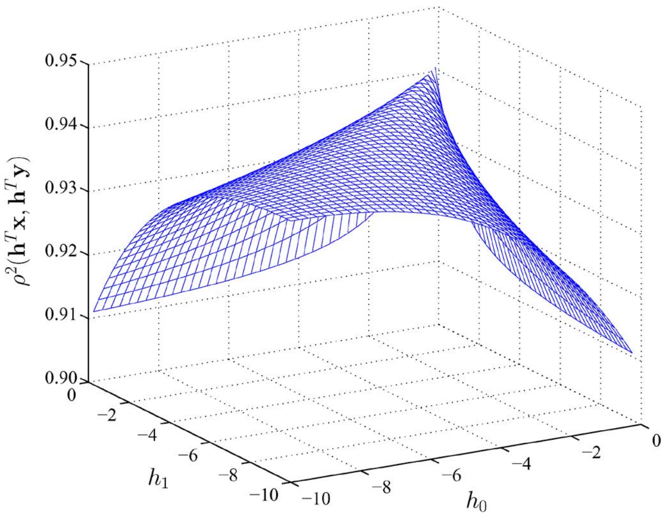

Fig. 2. The SPCC $\rho^2(\mathbf{h}^T\mathbf{x}, \mathbf{h}^T\mathbf{y})$ versus a two-tap filter $\mathbf{h} = [h_0\ h_1]^T$.

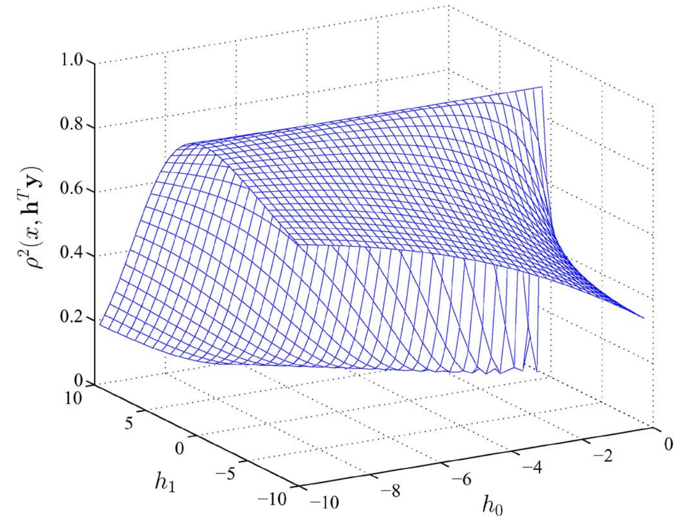

Fig. 3. The SPCC $\rho^2(x, \mathbf{h}^T\mathbf{y})$ versus a two-tap filter $\mathbf{h} = [h_0\ h_1]^T$.

To illustrate the result, let us consider the same example as used in Section V-A, where SNR $= 10$ dB, the filter length $L = 2$, and the signal and noise covariance matrices are computed respectively as

$$\mathbf{R}_{xx} = \begin{bmatrix} 1.0 & 0.6 \\ 0.6 & 1.0 \end{bmatrix} \quad (35)$$

and

$$\mathbf{R}_{vv} = \begin{bmatrix} 0.1 & 0.003 \\ 0.003 & 0.1 \end{bmatrix}. \quad (36)$$

The corresponding generalized eigenvector is $\mathbf{h}_{\max} = [0.7071\ 0.7071]^T$. Fig. 2 plots the performance surface as a function of h_0 and h_1. It is seen that the SPCC $\rho^2(\mathbf{h}^T\mathbf{x}, \mathbf{h}^T\mathbf{y})$ achieves its maximum when $h_0 = h_1$. As we mentioned earlier, the constant α does not affect the noise-reduction performance. Let us choose $\alpha = 1$, we have

$$\mathrm{SNR}(\mathbf{h}_{\max}) = \lambda_{\max} \quad (37)$$

$$\rho^2\left(\mathbf{h}_{\max}^T\mathbf{x}, \mathbf{h}_{\max}^T\mathbf{y}\right) = \frac{\lambda_{\max}}{1+\lambda_{\max}} \quad (38)$$

$$z_{\max}(k) = \mathbf{h}_{\max}^T\mathbf{y}(k). \quad (39)$$

From this filter, we can deduce another interesting property of the SPCC.

Property 10: We have

$$\rho^2\left(x, \mathbf{h}_{\max}^T\mathbf{x}\right) = \frac{\mathrm{SNR}(\mathbf{h}_{\max})}{\mathrm{SNR}} \cdot \rho^2\left(v, \mathbf{h}_{\max}^T\mathbf{v}\right). \quad (40)$$

Since $\mathrm{SNR}(\mathbf{h}_{\max}) \geq \mathrm{SNR}(\mathbf{h}_1) = \mathrm{SNR}$, this implies that

$$\rho^2\left(x, \mathbf{h}_{\max}^T\mathbf{x}\right) \geq \rho^2\left(v, \mathbf{h}_{\max}^T\mathbf{v}\right) \quad (41)$$

which means that the filter $\mathbf{h}_{\max}$ yields less distortion to the clean speech signal $x(k)$ than to the noise signal $v(k)$.

C. Wiener Filter

We are going to maximize the SPCC $\rho^2(x, \mathbf{h}^T\mathbf{y})$. Indeed, if we differentiate this term with respect to $\mathbf{h}$ and equate the result to zero, we easily obtain

$$\frac{2\mathbf{h}_1^T\mathbf{R}_{xx}\mathbf{h}}{\sigma_x^2(\mathbf{h}^T\mathbf{R}_{yy}\mathbf{h})^2}\left[\mathbf{R}_{xx}\mathbf{h}_1(\mathbf{h}^T\mathbf{R}_{yy}\mathbf{h}) - \left(\mathbf{h}_1^T\mathbf{R}_{xx}\mathbf{h}\right)\mathbf{R}_{yy}\mathbf{h}\right] = \mathbf{0}_{L\times 1}. \quad (42)$$

Depending on the rank of $\mathbf{R}_{xx}$ ($\mathbf{R}_{yy}$ is supposed to be full rank), we have at least two cases:

Case 1) $\mathbf{R}_{xx}$ is not full rank and $\mathbf{h}$ is in the null space of $\mathbf{R}_{xx}$. In this situation, although $\mathbf{h}$ satisfies the solution of (42), we can easily check that $\rho^2(x, \mathbf{h}^T\mathbf{y}) = 0$, which is minimized instead of being maximized.

Case 2) $\mathbf{R}_{xx}$ is full rank or $\mathbf{h}$ does not belong to the null space of $\mathbf{R}_{xx}$ if $\mathbf{R}_{xx}$ is rank deficient. In this situation, another possible solution to (42) is

$$\mathbf{h} = \alpha\mathbf{R}_{yy}^{-1}\mathbf{R}_{xx}\mathbf{h}_1. \quad (43)$$

To illustrate this result, we consider the same example used previously. The three-dimensional surface is shown in Fig. 3. Although there are multiple filters that can maximize the SPCC $\rho^2(x, \mathbf{h}^T\mathbf{y})$, they are only different by a constant factor and the resulting noise-reduction performance is the same. Now let us set $\alpha = 1$, we obtain

$$\mathbf{h}_{\mathrm{W}} = \mathbf{R}_{yy}^{-1}\mathbf{R}_{xx}\mathbf{h}_1 \quad (44)$$

which is the classical Wiener filter [31], [32].

From the above derivation, we see that the Wiener filter is obtained by solving the equation $\mathbf{R}_{xx}\mathbf{h}_1(\mathbf{h}^T\mathbf{R}_{yy}\mathbf{h}) = (\mathbf{h}_1^T\mathbf{R}_{yy}\mathbf{h})\mathbf{R}_{yy}\mathbf{h}$, which is equivalent to finding a filter that satisfies the relation $\mathbf{h}^T\mathbf{R}_{yy}\mathbf{h} = \mathbf{h}_1^T\mathbf{R}_{xx}\mathbf{h}$. With this relation, it can be easily checked that $\rho^2(x, \mathbf{h}^T\mathbf{y}) = E[z^2(k)]/\sigma_x^2$. Therefore, for the Wiener filter, we have the following properties.

Property 11: Maximizing the SPCC $\rho^2(x, \mathbf{h}^T\mathbf{y})$ is equivalent to maximizing the variance, $E[z^2(k)]$, of the filter output signal, $z(k)$, subject to the constraint $\mathbf{h}^T\mathbf{R}_{yy}\mathbf{h} = \mathbf{h}_1^T\mathbf{R}_{xx}\mathbf{h}$.

Property 12: We have

$$\rho^2\left(x, \mathbf{h}_{\mathrm{W}}^T\mathbf{y}\right) = \frac{1}{\xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{W}})} \cdot \frac{1+\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})}{\mathrm{SNR}}. \quad (45)$$

This implies that

$$\xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{W}}) \geq \frac{1+\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})}{\mathrm{SNR}}. \quad (46)$$

However, using Properties 2 and 12, we deduce a better lower bound:

$$\xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{W}}) \geq \frac{[1+\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})]^2}{\mathrm{SNR}\cdot\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})} \geq \frac{1+\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})}{\mathrm{SNR}}. \quad (47)$$

This lower bound shows that the amount of noise reduction using the Wiener filter depends on the SNR. The lower the input SNR, the more the noise is reduced. This is understandable since a lower input SNR means that the observation signal is more noisy, so there is more noise to attenuate.

Property 13: With the optimal Wiener filter given in (44), the output SNR is always greater than or equal to the input SNR.

Proof: Let us evaluate the SPCC between $y(k)$ and $\mathbf{h}_{\mathrm{W}}^T\mathbf{y}(k)$

$$\begin{aligned}\rho^2\left(y, \mathbf{h}_{\mathrm{W}}^T\mathbf{y}\right) &= \frac{\left(\mathbf{h}_1^T\mathbf{R}_{yy}\mathbf{h}_{\mathrm{W}}\right)^2}{\sigma_y^2\left(\mathbf{h}_{\mathrm{W}}^T\mathbf{R}_{yy}\mathbf{h}_{\mathrm{W}}\right)} \\ &= \frac{\sigma_x^2}{\sigma_y^2}\cdot\frac{\sigma_x^2}{\mathbf{h}_{\mathrm{W}}^T\mathbf{R}_{xx}\mathbf{h}_1} \\ &= \frac{\rho^2(x,y)}{\rho^2\left(x, \mathbf{h}_{\mathrm{W}}^T\mathbf{y}\right)}. \quad (48)\end{aligned}$$

Therefore

$$\rho^2(x,y) = \rho^2\left(y, \mathbf{h}_{\mathrm{W}}^T\mathbf{y}\right)\cdot\rho^2\left(x, \mathbf{h}_{\mathrm{W}}^T\mathbf{y}\right) \leq \rho^2\left(x, \mathbf{h}_{\mathrm{W}}^T\mathbf{y}\right). \quad (49)$$

Using (11) and Property 2 in the previous expression, we get

$$\frac{\mathrm{SNR}}{1+\mathrm{SNR}} \leq \frac{\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})}{1+\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})}. \quad (50)$$

Slightly reorganizing (50) gives

$$\frac{1}{1+\frac{1}{\mathrm{SNR}}} \leq \frac{1}{1+\frac{1}{\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})}} \quad (51)$$

which implies that

$$\frac{1}{\mathrm{SNR}} \geq \frac{1}{\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})}. \quad (52)$$

As a result

$$\mathrm{SNR}(\mathbf{h}_{\mathrm{W}}) \geq \mathrm{SNR}. \quad (53)$$

That completes the proof.

Note that the relationship between the input and output SNRs of the Wiener filter has been given in [31] and [33]. However, the proof shown there is carried out in a transformed domain, which is very complicated. In comparison, the proof shown here is amazingly simpler and much easier to follow.

Property 14: We have

$$\frac{[1+\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})]^2}{\mathrm{SNR}\cdot\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})} \leq \xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{W}}) \leq \frac{(1+\mathrm{SNR})\,[1+\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})]}{\mathrm{SNR}^2} \quad (54)$$

or

$$\begin{aligned}&\frac{1}{\rho^2\left(\mathbf{h}_{\mathrm{W}}^T\mathbf{v}, \mathbf{h}_{\mathrm{W}}^T\mathbf{y}\right)\cdot\rho^2\left(\mathbf{h}_{\mathrm{W}}^T\mathbf{x}, \mathbf{h}_{\mathrm{W}}^T\mathbf{y}\right)} \\ &\quad\leq \mathrm{SNR}\cdot\xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{W}}) \\ &\quad\leq \frac{1}{\rho^2(x,y)\cdot\rho^2\left(\mathbf{h}_{\mathrm{W}}^T\mathbf{v}, \mathbf{h}_{\mathrm{W}}^T\mathbf{y}\right)}. \quad (55)\end{aligned}$$

Proof: For the lower bound, see (47). The upper bound is easy to show by using Property 12 and (49).

Property 15: We have

$$\upsilon_{\mathrm{sd}}(\mathbf{h}_{\mathrm{W}}) = 1 - \rho^2\left(x, \mathbf{h}_{\mathrm{W}}^T\mathbf{x}\right)\cdot\left\{1 - \frac{1}{[1+\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})]^2}\right\}. \quad (56)$$

This expression shows the link between the speech-distortion index, $\upsilon_{\mathrm{sd}}(\mathbf{h}_{\mathrm{W}})$, and the SPCC $\rho^2(x, \mathbf{h}_{\mathrm{W}}^T\mathbf{x})$. When $\rho^2(x, \mathbf{h}_{\mathrm{W}}^T\mathbf{x})$ is high (resp. low), $\upsilon_{\mathrm{sd}}(\mathbf{h}_{\mathrm{W}})$ is small (resp. large) and, as a result, the clean speech signal is lowly (resp. highly) distorted. We also have

$$\rho^2\left(x, \mathbf{h}_{\mathrm{W}}^T\mathbf{x}\right) \geq \frac{\mathrm{SNR}}{1+\mathrm{SNR}}\cdot\frac{1+\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})}{\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})} \quad (57)$$

so when the output SNR increases, the lower bound of the SPCC $\rho^2(x, \mathbf{h}_{\mathrm{W}}^T\mathbf{x})$ decreases; as a consequence, the distortion of the clean speech likely increases.

Link With the MMSE: Now we discuss the connection between maximizing the SPCC and minimizing the MSE. Define the error signal between the clean speech sample and its estimate at time k:

$$\begin{aligned}e(k) &= x(k) - z(k) \\ &= x(k) - \mathbf{h}^T\mathbf{y}(k). \quad (58)\end{aligned}$$

The MSE is

$$\begin{aligned}J(\mathbf{h}) &= E\left[e^2(k)\right] \\ &= \sigma_x^2 + \mathbf{h}^T\mathbf{R}_{yy}\mathbf{h} - 2\mathbf{h}_1^T\mathbf{R}_{xx}\mathbf{h} \\ &= \sigma_x^2\left[1 + \frac{1}{\xi_{\mathrm{nr}}(\mathbf{h})}\cdot\frac{1+\mathrm{SNR}(\mathbf{h})}{\mathrm{SNR}}\right. \\ &\qquad\left. -2\frac{\mathbf{h}^T\mathbf{R}_{xx}\mathbf{h}}{\mathbf{h}_1^T\mathbf{R}_{xx}\mathbf{h}}\cdot\rho^2(x, \mathbf{h}^T\mathbf{x})\right]. \quad (59)\end{aligned}$$

We define the normalized MSE as

$$\tilde{J}(\mathbf{h}) = \frac{J(\mathbf{h})}{\sigma_v^2}. \quad (60)$$

The normalized minimum MSE (MMSE) is obtained by replacing the filter $\mathbf{h}$ in (60) by the Wiener filter $\mathbf{h}_{\mathrm{W}}$.

Property 16: We have

$$\tilde{J}(\mathbf{h}_{\mathrm{W}}) = \mathrm{SNR}\left[1 - \rho^2\left(x, \mathbf{h}_{\mathrm{W}}^T\mathbf{y}\right)\right]. \tag{61}$$

Therefore, as expected, the MSE is minimized when the SPCC is maximized.

Proof: Equation (61) can be easily verified by using Properties 1 and 12, the relation $\mathbf{h}_{\mathrm{W}}^T\mathbf{R}_{yy}\mathbf{h}_{\mathrm{W}} = \mathbf{h}_1^T\mathbf{R}_{xx}\mathbf{h}_{\mathrm{W}}$, and (59).

Property 17: We have

$$\frac{\mathrm{SNR}}{1+\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})} \le \tilde{J}(\mathbf{h}_{\mathrm{W}}) \le \frac{\mathrm{SNR}}{1+\mathrm{SNR}} \tag{62}$$

or

$$\rho^2\left(\mathbf{h}_{\mathrm{W}}^T\mathbf{v}, \mathbf{h}_{\mathrm{W}}^T\mathbf{y}\right) \le \frac{\tilde{J}(\mathbf{h}_{\mathrm{W}})}{\mathrm{SNR}} \le \rho^2(v,y). \tag{63}$$

Proof: These bounds can be proven by using the bounds of $\rho^2(x, \mathbf{h}_{\mathrm{W}}^T\mathbf{y})$ and (61).

Property 18: We have

$$\upsilon_{\mathrm{sd}}(\mathbf{h}_{\mathrm{W}}) = \frac{1}{\mathrm{SNR}}\left[\tilde{J}(\mathbf{h}_{\mathrm{W}}) - \frac{1}{\xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{W}})}\right]. \tag{64}$$

Proof: See [31].

Property 19: We have

$$\frac{1}{[1+\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})]^2} \le \upsilon_{\mathrm{sd}}(\mathbf{h}_{\mathrm{W}}) \le \frac{1+\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})-\mathrm{SNR}}{(1+\mathrm{SNR})[1+\mathrm{SNR}(\mathbf{h}_{\mathrm{W}})]} \tag{65}$$

or

$$\begin{aligned}\rho^4\left(\mathbf{h}_{\mathrm{W}}^T\mathbf{v}, \mathbf{h}_{\mathrm{W}}^T\mathbf{y}\right) &\le \upsilon_{\mathrm{sd}}(\mathbf{h}_{\mathrm{W}}) \\ &\le \rho^2(v,y)\cdot\rho^2\left(\mathbf{h}_{\mathrm{W}}^T\mathbf{v}, \mathbf{h}_{\mathrm{W}}^T\mathbf{y}\right) + \rho^2(v,y) \\ &\quad - \rho^2\left(\mathbf{h}_{\mathrm{W}}^T\mathbf{v}, \mathbf{h}_{\mathrm{W}}^T\mathbf{y}\right).\end{aligned} \tag{66}$$

Proof: These bounds can be proven by using Properties 14, 17, and 18.

Property 20: From the MSE perspective, with the Wiener filter

$$\mathrm{SNR}(\mathbf{h}_{\mathrm{W}}) \ge \mathrm{SNR} \Longleftrightarrow \xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{W}}) > 1, \quad \upsilon_{\mathrm{sd}}(\mathbf{h}_{\mathrm{W}}) < 1. \tag{67}$$

Therefore, the measures $\xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{W}})$ and $\upsilon_{\mathrm{sd}}(\mathbf{h}_{\mathrm{W}})$ may be good indicators of the behavior of the Wiener filter except for the case where both the signal and noise are white random processes, indicating that they are not self predictable. In this scenario, both the signal and noise covariance matrices are diagonal, and we have

$$\mathbf{h}_{\mathrm{W}} = \frac{\mathrm{SNR}}{1+\mathrm{SNR}}\mathbf{h}_1 \tag{68}$$

$$\xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{W}}) = \frac{(1+\mathrm{SNR})^2}{\mathrm{SNR}^2} > 1 \tag{69}$$

$$\upsilon_{\mathrm{sd}}(\mathbf{h}_{\mathrm{W}}) = \frac{1}{(1+\mathrm{SNR})^2} > 0 \tag{70}$$

$$\mathrm{SNR}(\mathbf{h}_{\mathrm{W}}) = \mathrm{SNR}. \tag{71}$$

This particular case shows a slight anomaly in the definitions (5) and (6) since noise reduction and speech distortion are possible while the output SNR is not improved at all. This is due to the fact that

$$\xi_{\mathrm{nr}}(c\cdot\mathbf{h}_{\mathrm{W}}) \ne \xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{W}}) \tag{72}$$

$$\upsilon_{\mathrm{sd}}(c\cdot\mathbf{h}_{\mathrm{W}}) \ne \upsilon_{\mathrm{sd}}(\mathbf{h}_{\mathrm{W}}) \tag{73}$$

for a nonzero constant c and $c \ne 1$.

Property 21: From the SPCC perspective, with the Wiener filter

$$\begin{aligned}\mathrm{SNR}(\mathbf{h}_{\mathrm{W}}) \ge \mathrm{SNR} \Longleftrightarrow\ & \rho^2\left(\mathbf{h}_{\mathrm{W}}^T\mathbf{x}, \mathbf{h}_{\mathrm{W}}^T\mathbf{y}\right) \\ &\ge \rho^2(x,y), \rho^2\left(x, \mathbf{h}_{\mathrm{W}}^T\mathbf{x}\right) \le 1.\end{aligned} \tag{74}$$

When $\mathrm{SNR}(\mathbf{h}_{\mathrm{W}}) = \mathrm{SNR}$, then

$$\rho^2\left(\mathbf{h}_{\mathrm{W}}^T\mathbf{x}, \mathbf{h}_{\mathrm{W}}^T\mathbf{y}\right) = \rho^2(x,y) \tag{75}$$

$$\rho^2\left(x, \mathbf{h}_{\mathrm{W}}^T\mathbf{x}\right) = 1. \tag{76}$$

This time, the measures based on the SPCCs $\rho^2(\mathbf{h}_{\mathrm{W}}^T\mathbf{x}, \mathbf{h}_{\mathrm{W}}^T\mathbf{y})$ and $\rho^2(x, \mathbf{h}_{\mathrm{W}}^T\mathbf{x})$ reflect accurately the output SNR, since when the latter is not improved the speech-distortion index $\rho^2(x, \mathbf{h}_{\mathrm{W}}^T\mathbf{x})$ says that there is no speech distortion and the noise-reduction index $\rho^2(\mathbf{h}_{\mathrm{W}}^T\mathbf{x}, \mathbf{h}_{\mathrm{W}}^T\mathbf{y})$ says that there is no noise reduction indeed. The anomaly discussed above no longer exists in the context of the SPCC thanks to the properties:

$$\rho^2\left(c\cdot\mathbf{h}_{\mathrm{W}}^T\mathbf{x}, c\cdot\mathbf{h}_{\mathrm{W}}^T\mathbf{y}\right) = \rho^2\left(\mathbf{h}_{\mathrm{W}}^T\mathbf{x}, \mathbf{h}_{\mathrm{W}}^T\mathbf{y}\right) \tag{77}$$

$$\rho^2\left(x, c\cdot\mathbf{h}_{\mathrm{W}}^T\mathbf{x}\right) = \rho^2\left(x, \mathbf{h}_{\mathrm{W}}^T\mathbf{x}\right) \tag{78}$$

for a constant $c \ne 0$.

Properties 20 and 21 show basically that the noise-reduction factor, $\xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{W}})$, and the speech-distortion index, $\upsilon_{\mathrm{sd}}(\mathbf{h}_{\mathrm{W}})$, derived from the MSE formulation present a slight anomaly compared to the equivalent measures based on the SPCCs and derived from an SPCC criterion.

D. Tradeoff Filters

It is also possible to derive other optimal filters that can control the tradeoff between speech distortion and SNR improvement. For example, it can be more attractive to find a filter that minimizes the speech distortion while it guaranties a certain level of SNR improvement. Mathematically, this optimization problem can be written as follows:

$$\max_{\mathbf{h}} \rho^2(x, \mathbf{h}^T\mathbf{x}) \text{ subject to } \mathrm{SNR}(\mathbf{h}) = \beta\cdot\mathrm{SNR} \tag{79}$$

where $\beta > 1$. If we use a Lagrange multiplier, μ, to adjoin the constraint to the cost function, (79) can be rewritten as

$$\max_{\mathbf{h}} \mathcal{L}(\mathbf{h}, \mu) \tag{80}$$

with

$$\mathcal{L}(\mathbf{h}, \mu) = \frac{\left(\mathbf{h}_1^T\mathbf{R}_{xx}\mathbf{h}\right)^2}{\sigma_x^2(\mathbf{h}^T\mathbf{R}_{xx}\mathbf{h})} + \mu\left(\frac{\mathbf{h}^T\mathbf{R}_{xx}\mathbf{h}}{\mathbf{h}^T\mathbf{R}_{vv}\mathbf{h}} - \beta\cdot\mathrm{SNR}\right). \tag{81}$$

Taking the gradient of $\mathcal{L}(\mathbf{h}, \mu)$ with respect to $\mathbf{h}$ and equating the result to zero, we get

$$\frac{2\sigma_x^2\left(\mathbf{h}_1^T\mathbf{R}_{xx}\mathbf{h}\right)(\mathbf{h}^T\mathbf{R}_{xx}\mathbf{h})\mathbf{R}_{xx}\mathbf{h}_1 - 2\sigma_x^2\left(\mathbf{h}_1^T\mathbf{R}_{xx}\mathbf{h}\right)^2\mathbf{R}_{xx}\mathbf{h}}{(\sigma_x^2\cdot\mathbf{h}^T\mathbf{R}_{xx}\mathbf{h})^2} + \mu\frac{2(\mathbf{h}^T\mathbf{R}_{vv}\mathbf{h})\mathbf{R}_{xx}\mathbf{h} - 2(\mathbf{h}^T\mathbf{R}_{xx}\mathbf{h})\mathbf{R}_{vv}\mathbf{h}}{(\mathbf{h}^T\mathbf{R}_{vv}\mathbf{h})^2} = \mathbf{0}_{L\times 1}. \quad (82)$$

Now let us look for the optimal filter, $\mathbf{h}_\mathrm{T}$, that satisfies the relation

$$\mathbf{h}_1^T\mathbf{R}_{xx}\mathbf{h}_\mathrm{T} = \mathbf{h}_\mathrm{T}^T\mathbf{R}_{xx}\mathbf{h}_\mathrm{T}. \quad (83)$$

In this case, (82) becomes

$$\frac{\mathbf{R}_{xx}\mathbf{h}_1}{\sigma_x^2} - \frac{\mathbf{R}_{xx}\mathbf{h}_\mathrm{T}}{\sigma_x^2} + \mu\frac{\left(\mathbf{h}_\mathrm{T}^T\mathbf{R}_{vv}\mathbf{h}_\mathrm{T}\right)\mathbf{R}_{xx}\mathbf{h}_\mathrm{T} - \left(\mathbf{h}_\mathrm{T}^T\mathbf{R}_{xx}\mathbf{h}_\mathrm{T}\right)\mathbf{R}_{vv}\mathbf{h}_\mathrm{T}}{\left(\mathbf{h}_\mathrm{T}^T\mathbf{R}_{vv}\mathbf{h}_\mathrm{T}\right)^2} = \mathbf{0}_{L\times 1}. \quad (84)$$

Left-multiplying both sides of (84) by $\mathbf{h}_\mathrm{T}^T$, we can check that, indeed, the filter $\mathbf{h}_\mathrm{T}$ satisfies the relation (83). After some simple manipulations on (84), we find that

$$\mathbf{R}_{xx}\mathbf{h}_1 - \mathbf{R}_{xx}\mathbf{h}_\mathrm{T} + \mu\mathrm{SNR}\xi_{\mathrm{nr}}(\mathbf{h}_\mathrm{T})\mathbf{R}_{xx}\mathbf{h}_\mathrm{T} - \mu\beta\mathrm{SNR}^2\xi_{\mathrm{nr}}(\mathbf{h}_\mathrm{T})\mathbf{R}_{vv}\mathbf{h}_\mathrm{T} = \mathbf{0}_{L\times 1}. \quad (85)$$

Define the quantities

$$\tilde{\mathbf{R}}_{xx} = \frac{\mathbf{R}_{xx}}{\sigma_x^2} \quad (86)$$

$$\tilde{\mathbf{R}}_{vv} = \frac{\mathbf{R}_{vv}}{\sigma_v^2} \quad (87)$$

$$\mu' = \mu\beta\mathrm{SNR}^2\xi_{\mathrm{nr}}(\mathbf{h}_\mathrm{T}). \quad (88)$$

We find the optimal tradeoff filter

$$\mathbf{h}_\mathrm{T} = \left[\frac{\mu'}{\mathrm{SNR}}\mathbf{I}_{L\times L} + \left(\mathbf{I}_{L\times L} - \frac{\mu'}{\beta\mathrm{SNR}}\right)\tilde{\mathbf{R}}_{vv}^{-1}\tilde{\mathbf{R}}_{xx}\right]^{-1} \times\tilde{\mathbf{R}}_{vv}^{-1}\tilde{\mathbf{R}}_{xx}\mathbf{h}_1 \quad (89)$$

which can be compared to the Wiener filter [31]

$$\mathbf{h}_\mathrm{W} = \left[\frac{1}{\mathrm{SNR}}\mathbf{I}_{L\times L} + \tilde{\mathbf{R}}_{vv}^{-1}\tilde{\mathbf{R}}_{xx}\right]^{-1}\tilde{\mathbf{R}}_{vv}^{-1}\tilde{\mathbf{R}}_{xx}\mathbf{h}_1. \quad (90)$$

The purpose of the filter $\mathbf{h}_\mathrm{T}$ is the same as the filters derived in [12], [16]. We can play with the parameters μ' and β to get different forms of the tradeoff filter. For examples, for $\mu' = 0$ we have the speech distortionless filter, $\mathbf{h}_\mathrm{T} = \mathbf{h}_1$, and for $\mu' = 1$ and $\beta \to \infty$, we get the Wiener filter, $\mathbf{h}_\mathrm{T} = \mathbf{h}_\mathrm{W}$.

Another example of a tradeoff filter can be derived by maximizing the output SNR while setting the speech distortion to a certain level. Mathematically, this optimization problem can be formulated as follows:

$$\max_{\mathbf{h}}\mathrm{SNR}(\mathbf{h}) \text{ subject to } \rho^2(x, \mathbf{h}^T\mathbf{x}) = \gamma \quad (91)$$

where $\gamma < 1$. Following the same steps developed for the optimization problem of (79), it can be shown that the optimal tradeoff filter derived from (91) is

$$\mathbf{h}_{\mathrm{T},2} = \left[\frac{\mu''}{\mathrm{SNR}}\mathbf{I}_{L\times L} + \left(\mathbf{I}_{L\times L} - \frac{\mu''}{\gamma'\mathrm{SNR}}\right)\tilde{\mathbf{R}}_{vv}^{-1}\tilde{\mathbf{R}}_{xx}\right]^{-1} \times\tilde{\mathbf{R}}_{vv}^{-1}\tilde{\mathbf{R}}_{xx}\mathbf{h}_1 \quad (92)$$

where

$$\gamma' = \gamma\xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{T},2}) \quad (93)$$

$$\mu'' = \gamma\frac{\left[\mathrm{SNR}\xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{T},2})\right]^2}{\mu}. \quad (94)$$

The two optimal tradeoff filters $\mathbf{h}_\mathrm{T}$ and $\mathbf{h}_{\mathrm{T},2}$ are in the same form even though the latter is rarely used in practice because the level of speech distortion is very difficult to control.

VI. Conclusion

Traditionally, optimal noise-reduction filters are typically formulated from the MSE criterion. With such a formulation, however, many important properties such as the SNR behavior of the resulting optimal filters are masked. In this paper, we discussed the squared Pearson correlation coefficient (SPCC) and showed that this coefficient has many appealing properties. Similar to the MSE criterion, the SPCC can be used to derive optimal noise-reduction filters. For example, we illustrated how to deduce the widely used Wiener filter and some other optimal filters that have a better control on the tradeoff between noise reduction and speech distortion. The major advantage of using the SPCC as a cost function is that the noise-reduction performance (in terms of SNR improvement and speech distortion) of the resulting optimal filters can be easily analyzed. So, for the problem of noise reduction, the SPCC-based criterion is a more natural objective function to optimize as compared to the MSE.

References

[1] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Upper Saddle River, NJ: Prentice-Hall, 1993.

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

[3] Y. Huang and J. Benesty, Eds., *Audio Signal Processing for Next-Generation Multimedia Communication Systems*. Boston, MA: Kluwer, 2004.

[4] B. Widrow, J. R. Glover, J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, and R. C. Goodwin, "Adaptive noise canceling: Principles and applications," *Proc. IEEE*, vol. 63, no. 12, pp. 1692–1716, Dec. 1975.

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

[6] M. M. Goulding and J. S. Bird, "Speech enhancement for mobile telephony," *IEEE Trans. Veh. Technol.*, vol. 39, no. 4, pp. 316–326, Nov. 1990.

[7] H. J. Kushner, "On closed-loop adaptive noise cancellation," *IEEE Trans. Automat. Contr.*, vol. 43, no. 8, pp. 1103–1107, Aug. 1998.

[8] A. S. Abutaled, "An adaptive filter for noise canceling," *IEEE Trans. Circuits Syst.*, vol. 35, no. 10, pp. 1201–1209, Oct. 1998.

[9] 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.

[10] 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.

[11] J. S. Lim, Ed., *Speech Enhancement*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[12] 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.

[13] 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.

[14] 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.

[15] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 2, pp. 443–445, Apr. 1985.

[16] 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.

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

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

[19] 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.

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

[21] 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.

[22] 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.

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

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

[25] K. Pearson, "Mathematical contributions to the theory of evolution.—III. Regression, heredity and panmixia," *Philos. Trans. R. Soc. London*, ser. A, vol. 187, pp. 253–318, 1896.

[26] O. J. Dunn and V. A. Clark, *Applied Statistics: Analysis of Variance and Regression*. New York: Wiley, 1974.

[27] J. L. Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient," *Amer. Statistician*, vol. 42, pp. 59–66, Feb. 1988.

[28] G. C. Carter, "Coherence and time delay estimation," in *Signal Processing Handbook*, C. H. Chen, Ed. New York: Marcel Dekker, 1988.

[29] R. O. Duda, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[30] 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.

[31] 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.

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

[33] 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.

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 the University of Quebec, INRS-EMT, Montreal, QC, Canada, as an Associate Professor. His research interests are in signal processing, acoustic signal processing, and multimedia communications. He coauthored the books *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* (Spinger-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 and was the cochair of the 1999 International Workshop on Acoustic Echo and Noise Control.

Jingdong Chen (M'99) received the B.S. degree in electrical engineering and the M.S. degree in array signal processing 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 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, signal separation, and source localization. He coauthored the books *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 helped organize the 2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustisc (WASPAA), and he will serve as a technical cochair of the 2009 WASPAA. He is the recipient of a 1998–1999 research grant 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. He is currently an Associated Editor of the *EURASIP Journal on Applied Signal Processing*. He is a coeditor/coauthor of the books *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).

Dr. Huang served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS from 2002 to 2005. He served as a technical cochair 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.