Analysis and Comparison of Multichannel Noise Reduction Methods in a Common Framework

Article *in* IEEE Transactions on Audio Speech and Language Processing · August 2008

DOI: 10.1109/TASL.2008.921754 · Source: IEEE Xplore

3 authors, including:

Jacob Benesty
Institut National de la Recherche Scientifique
636 PUBLICATIONS **14,950** CITATIONS

SEE PROFILE

Jingdong Chen
Institute of Electrical and Electronics Engineers
323 PUBLICATIONS **5,852** CITATIONS

SEE PROFILE

Analysis and Comparison of Multichannel Noise Reduction Methods in a Common Framework

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

***Abstract*—Noise reduction for speech enhancement is a useful technique, but in general it is a challenging problem. While a single-channel algorithm is easy to use in practice, it inevitably introduces speech distortion to the desired speech signal while reducing noise. Today, the explosive growth in computational power and the continuous drop in the cost and size of acoustic electric transducers are driving the interest of employing multiple microphones in speech processing systems. This opens new opportunities for noise reduction. In this paper, we present an analysis of three multichannel noise reduction algorithms, namely Wiener filter, subspace, and spatial-temporal prediction, in a common framework. We intend to investigate whether it is possible for the multichannel noise reduction algorithms to reduce noise without speech distortion. Finally, we justify what we learn via theoretical analyses by simulations using real impulse responses measured in the varechoic chamber at Bell Labs.**

***Index Terms*—Microphone array signal processing, multichannel subspace method, multichannel Wiener filter, noise reduction, spatial prediction, speech enhancement.**

I. Introduction

WHEREVER we are, noise (originating from various ambient sound sources) is permanently present. As a result, speech signals cannot be acquired and processed, in general, in pure form. It has been known for a long time that noise can profoundly affect human-to-human and human-to-machine communications, including changing a talker's speaking pattern, modifying the characteristics of the speech signal, degrading speech quality and intelligibility, and affecting the listener's perception and machine's processing of the recorded speech. In order to make voice communication feasible, natural, and comfortable in the presence of noise regardless of the noise level, it is desirable to develop digital signal processing techniques to "clean" the microphone signal before it is stored, transmitted, or played out. This problem has been a major challenge for many researchers and engineers for more than four decades [1].

The signal picked up by the microphone can be modeled as a superposition of the clean speech and noise. The objective of noise reduction then becomes to restore the original clean speech from the mixed signal. The first single-channel noise reduction algorithm was developed more than 40 years ago by Schroeder [2], [3]. He proposed an analog implementation of the spectral magnitude subtraction. This work, however, has not received much public attention, probably because it was never published in journals or conferences. About 15 years later, Boll, in his informative paper [4], reinvented the spectral subtraction method but in the digital domain. Almost at the same time, Lim and Oppenheim, in their landmark work [5], systematically formulated the noise-reduction problem and studied and compared the different algorithms known at that time. Since then many algorithms have been derived in the time and frequency domains [1], [6]–[8]. The main drawback of single-channel speech enhancement algorithms is that they distort the desired speech signal. So researchers have proposed to use multiple microphones or microphone arrays in order to better deal with this fundamental problem.

In this paper, we present a common framework to study the most important noise reduction algorithms in the multichannel case. The main desire is to see whether, indeed, the use of multiple microphones can help in minimizing speech distortion while having a good amount of noise reduction at the same time. This paper is organized as follows. Section II describes the problem and the signal model while Section III gives some very useful definitions that will help the reader understand how noise reduction algorithms work. Section IV explains the multichannel Wiener filter. Section V develops the subspace method with multiple microphones. In Section VI, the spatial-temporal prediction approach is derived. In Section VII, we present some simulations. Finally, we give our conclusions in Section VIII.

II. Signal Model and Problem Description

In this section, we explain the problem that we are going to tackle. We consider the general situation where we have N microphone signals whose outputs, at the discrete time k, are

$$\begin{aligned} y_{\mathrm{n}}(k) &= s(k) * g_{\mathrm{n}} + v_{\mathrm{n}}(k) \\ &= x_{\mathrm{n}}(k) + v_{\mathrm{n}}(k), \quad n = 1, 2, \ldots, N \end{aligned} \tag{1}$$

where $*$ stands for convolution, g_{n} is the impulse response of length L_g from the unknown source to the nth microphone, and $v_{\mathrm{n}}(k)$ is the additive background noise at microphone n. We assume that the noise signals $v_{\mathrm{n}}(k)$ and $x_{\mathrm{n}}(k)$ are uncorrelated and zero-mean. Moreover, we further assume that the noise signals are *not* perfectly coherent. Without loss of generality, we consider the first microphone signal $y_1(k)$ as the reference. Our main objective in this paper is noise reduction [1], [7]; hence, we will try to recover $x_1(k)$ the best way we can by using not

Manuscript received July 30, 2007; revised February 27, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hiroshi Sawada.

Y. Huang is with WeVoice, Inc., Bridgewater, NJ 08807 USA (e-mail: arden_huang@ieee.org).

J. Benesty is with the Université du Québec, INRS-EMT, Montréal, QC H5A 1K6, Canada (e-mail: benesty@emt.inrs.ca).

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

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.2008.921754

just one microphone signal but rather N signals. We do not attempt here to recover $s(k)$ (i.e., speech dereverberation). This problem, although very important, is difficult and requires other techniques to solve it [9]–[11]. Contrary to most beamforming techniques, the knowledge about the geometry of the microphone array is not required in the algorithms presented in this paper. Therefore, measurement or estimation errors in the locations of the microphones have little or no impact on these multichannel noise reduction algorithms, and a calibration with respect to microphone positions is not necessary.

The signal model given in (1) can be written in a vector/matrix form if we process the data by blocks of L samples

$$\begin{aligned}\mathbf{y}_{\mathrm{n}}(k) &= \mathbf{G}_{\mathrm{n}}^T\mathbf{s}_{L'}(k) + \mathbf{v}_{\mathrm{n}}(k)\\ &= \mathbf{x}_{\mathrm{n}}(k) + \mathbf{v}_{\mathrm{n}}(k), \quad n = 1, 2, \ldots, N\end{aligned} \tag{2}$$

where

$$\mathbf{y}_{\mathrm{n}}(k) = [\, y_{\mathrm{n}}(k) \quad y_{\mathrm{n}}(k-1) \quad \ldots \quad y_{\mathrm{n}}(k-L+1)\,]^T$$

is a vector containing the L most recent samples of the noisy speech signal $y_{\mathrm{n}}(k)$,

$$\mathbf{G}_{\mathrm{n}} = \begin{bmatrix} g_{n,0} & \cdots & g_{n,L_g-1} & 0 & \cdots & 0 \\ 0 & g_{n,0} & \cdots & g_{n,L_g-1} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & g_{n,0} & \cdots & g_{n,L_g-1} \end{bmatrix}^T$$

is a Sylvester matrix of size $L' \times L$ with $L' = L + L_g - 1$,

$$\mathbf{s}_{L'}(k) = [\, s(k) \quad s(k-1) \quad \ldots \quad s(k-L'+1)\,]^T$$

is a vector containing the L' most recent samples of the source signal, $(\cdot)^T$ denotes a vector/matrix transpose, and $\mathbf{x}_{\mathrm{n}}(k)$ and $\mathbf{v}_{\mathrm{n}}(k)$ are defined in a similar way to $\mathbf{y}_{\mathrm{n}}(k)$. Again, our objective is to estimate $\mathbf{x}_1(k)$ from the observations $\mathbf{y}_{\mathrm{n}}(k)$, $n = 1, 2, \ldots, N$.

Usually, we estimate the noise-free speech $\mathbf{x}_1(k)$ by applying a linear transformation to the microphone signals, i.e.,

$$\begin{aligned}\mathbf{z}(k) &\triangleq \sum_{n=1}^{N} \mathbf{H}_{\mathrm{n}}\mathbf{y}_{\mathrm{n}}(k)\\ &= \mathbf{H}\mathbf{y}(k)\\ &= \mathbf{H}\,[\mathbf{x}(k) + \mathbf{v}(k)]\end{aligned} \tag{3}$$

where

$$\begin{aligned}\mathbf{y}(k) &= [\, \mathbf{y}_1^T(k) \quad \mathbf{y}_2^T(k) \quad \ldots \quad \mathbf{y}_N^T(k)\,]^T\\ \mathbf{x}(k) &= [\, \mathbf{x}_1^T(k) \quad \mathbf{x}_2^T(k) \quad \ldots \quad \mathbf{x}_N^T(k)\,]^T\\ \mathbf{v}(k) &= [\, \mathbf{v}_1^T(k) \quad \mathbf{v}_2^T(k) \quad \ldots \quad \mathbf{v}_N^T(k)\,]^T\\ \mathbf{H} &= [\, \mathbf{H}_1 \quad \mathbf{H}_2 \quad \ldots \quad \mathbf{H}_N\,]\end{aligned}$$

and $\mathbf{H}_{\mathrm{n}}, n = 1, 2, \ldots, N$ are the filtering matrices of size $L \times L$. From this estimate, we define the error signal vector as

$$\begin{aligned}\mathbf{e}(k) &\triangleq \mathbf{z}(k) - \mathbf{x}_1(k)\\ &= (\mathbf{H} - \mathbf{U})\,\mathbf{x}(k) + \mathbf{H}\mathbf{v}(k)\\ &= \mathbf{e}_x(k) + \mathbf{e}_v(k)\end{aligned} \tag{4}$$

where

$$\mathbf{U} \triangleq [\, \mathbf{I}_{L\times L} \quad \mathbf{0}_{L\times L} \quad \ldots \quad \mathbf{0}_{L\times L}\,]$$

is an $L \times NL$ matrix with $\mathbf{I}_{L\times L}$ being the identity matrix of size $L \times L$

$$\mathbf{e}_x(k) \triangleq (\mathbf{H} - \mathbf{U})\,\mathbf{x}(k) \tag{5}$$

is the speech distortion due to the linear transformation, and

$$\mathbf{e}_v(k) \triangleq \mathbf{H}\mathbf{v}(k) \tag{6}$$

represents the residual noise.

III. Some Useful Definitions

The simplest and most intuitive way to quantify the amount of noise from an observed signal is the signal-to-noise ratio (SNR). Since our reference microphone is the first one, we define the input SNR as

$$\begin{aligned}\mathrm{SNR} &\triangleq \frac{\sigma_{x_1}^2}{\sigma_{v_1}^2} = \frac{E\left[\mathbf{x}_1^T(k)\mathbf{x}_1(k)\right]}{E\left[\mathbf{v}_1^T(k)\mathbf{v}_1(k)\right]}\\ &= \frac{\mathrm{tr}\left\{E\left[\mathbf{U}\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{U}^T\right]\right\}}{\mathrm{tr}\left\{E\left[\mathbf{U}\mathbf{v}(k)\mathbf{v}^T(k)\mathbf{U}^T\right]\right\}}\end{aligned} \tag{7}$$

where $E[\cdot]$ and $\mathrm{tr}[\cdot]$ denote mathematical expectation and the trace of a matrix, respectively.

The primary issue that we must determine with noise reduction is how much noise is actually attenuated. The noise-reduction factor is a measure of this and its mathematical definition is

$$\begin{aligned}\xi_{\mathrm{nr}}(\mathbf{H}) &\triangleq \frac{E\left[\mathbf{v}_1^T(k)\mathbf{v}_1(k)\right]}{E\left[\mathbf{e}_v^T(k)\mathbf{e}_v(k)\right]}\\ &= \frac{\mathrm{tr}\left\{E\left[\mathbf{U}\mathbf{v}(k)\mathbf{v}^T(k)\mathbf{U}^T\right]\right\}}{\mathrm{tr}\left\{E\left[\mathbf{H}\mathbf{v}(k)\mathbf{v}^T(k)\mathbf{H}^T\right]\right\}}.\end{aligned} \tag{8}$$

This factor should be lower bounded by 1. The larger the value of $\xi_{\mathrm{nr}}(\mathbf{H})$, the more is the noise reduction.

Most, if not all, of the known methods achieve noise reduction at the price of distorting the speech signal. Therefore, it is extremely useful to quantify this distortion. The speech-distortion index is defined as follows:

$$\upsilon_{\mathrm{sd}}(\mathbf{H}) \triangleq \frac{E\left[\mathbf{e}_x^T(k)\mathbf{e}_x(k)\right]}{E\left[\mathbf{x}_1^T(k)\mathbf{x}_1(k)\right]}. \tag{9}$$

This parameter is lower bounded by 0 and expected to be upper bounded by 1. The higher the value of $\upsilon_{\mathrm{sd}}(\mathbf{H})$, the more the speech signal $x_1(k)$ is distorted.

Noise reduction is done at the expense of speech reduction. Similar to the noise-reduction factor, we give the definition of the speech-reduction factor

$$\xi_{\mathrm{sr}}(\mathbf{H}) = \frac{\mathrm{tr}\left\{E\left[\mathbf{U}\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{U}^T\right]\right\}}{\mathrm{tr}\left\{E\left[\mathbf{H}\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{H}^T\right]\right\}}. \tag{10}$$

This factor is also lower bounded by 1.

In order to know if the filtering matrix $(\mathbf{H})$ can improve the SNR, we define the output SNR after noise reduction as

$$\text{SNR}(\mathbf{H}) \triangleq \frac{\text{tr}\left\{E\left[\mathbf{H}\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{H}^T\right]\right\}}{\text{tr}\left\{E\left[\mathbf{H}\mathbf{v}(k)\mathbf{v}^T(k)\mathbf{H}^T\right]\right\}}. \tag{11}$$

It is nice to find a filter $\mathbf{H}$ in such a way that $\text{SNR}(\mathbf{H}) > \text{SNR}$. SNR is a reliable and easily analyzed objective measure for the evaluation of speech enhancement algorithms. In addition, it is also reasonable to assume, to some extent (a properly frequency-dependent weighted SNR may correlate better with speech intelligibility [12]), some correlation between SNR and subjective listening. However, maximizing $\text{SNR}(\mathbf{H})$ is certainly not the best thing to do since the distortion of the speech signal will likely be maximized as well.

Using expressions (7), (8), (10), and (11), it is easy to see that we always have

$$\frac{\text{SNR}(\mathbf{H})}{\text{SNR}} = \frac{\xi_{\text{nr}}(\mathbf{H})}{\xi_{\text{sr}}(\mathbf{H})}. \tag{12}$$

Hence, $\text{SNR}(\mathbf{H}) > \text{SNR}$ if and only if $\xi_{\text{nr}}(\mathbf{H}) > \xi_{\text{sr}}(\mathbf{H})$. So is it possible that with a judicious choice of the filtering matrix $\mathbf{H}$ we can have $\xi_{\text{nr}}(\mathbf{H}) > \xi_{\text{sr}}(\mathbf{H})$? The answer is yes. A generally rough and intuitive justification to this answer is quite simple: improvement of the output SNR is due to the fact that speech signals are partly predictable. In this situation, $\mathbf{H}$ is a kind of a complex predictor or interpolator matrix and as a result, $\xi_{\text{sr}}(\mathbf{H})$ can be close to 1 while $\xi_{\text{nr}}(\mathbf{H})$ can be much larger than 1. This fact is very important for the single-microphone case and even more important in the multichannel case where we can exploit not only the temporal prediction of the speech signal but also the spatial prediction of the observed signals from different microphones in order to improve the output SNR and minimize the speech distortion.

IV. Wiener Filter

In this section, we derive the classical optimal Wiener filter for noise reduction yet in the multichannel case. Let us first write the mean-square error (MSE) criterion

$$\begin{aligned} J(\mathbf{H}) &= \text{tr}\left\{E\left[\mathbf{e}(k)\mathbf{e}^T(k)\right]\right\} \\ &= E\left[\mathbf{x}_1^T(k)\mathbf{x}_1(k)\right] + \text{tr}\left[\mathbf{H}\mathbf{R}_{yy}\mathbf{H}^T\right] - 2\text{tr}\left[\mathbf{H}\mathbf{R}_{yx_1}\right] \end{aligned} \tag{13}$$

where $\mathbf{R}_{yy} = E\left[\mathbf{y}(k)\mathbf{y}^T(k)\right]$ is the $NL \times NL$ correlation matrix of the observation signals and $\mathbf{R}_{yx_1} = E\left[\mathbf{y}(k)\mathbf{x}_1^T(k)\right]$ is the $NL \times L$ cross-correlation matrix between the observation and speech signals. Differentiating the MSE criterion with respect to $\mathbf{H}$ and setting the result to zero, we find the Wiener filter matrix [13], [14]

$$\mathbf{H}_{\text{W}}^T = \mathbf{R}_{yy}^{-1}\mathbf{R}_{yx_1}. \tag{14}$$

The previous equation is of little help in practice since the vector $\mathbf{x}_1(k)$ is unobservable. However, it is easy to check that

$$\mathbf{R}_{yx_1} = \left(\mathbf{R}_{yy} - \mathbf{R}_{vv}\right)\mathbf{U}^T \tag{15}$$

with $\mathbf{R}_{vv} = E\left[\mathbf{v}(k)\mathbf{v}^T(k)\right]$ being the $NL \times NL$ correlation matrix of the noise signals. Now $\mathbf{R}_{yx_1}$ depends on the correlation matrices $\mathbf{R}_{yy}$ and $\mathbf{R}_{vv}$: the first one can be easily estimated during speech-and-noise periods while the second one can be estimated during noise-only intervals assuming that the statistics of the noise change slowly over time. Substituting (15) into (14), we get

$$\mathbf{H}_{\text{W}}^T = \left(\mathbf{I}_{NL\times NL} - \mathbf{R}_{yy}^{-1}\mathbf{R}_{vv}\right)\mathbf{U}^T. \tag{16}$$

The minimum MSE (MMSE) is obtained by replacing $\mathbf{H}$ with $\mathbf{H}_{\text{W}}$ in (13), i.e., $J(\mathbf{H}_{\text{W}})$. There are different ways to express this MMSE. One useful expression is

$$J(\mathbf{H}_{\text{W}}) = \text{tr}\left[\mathbf{U}\mathbf{R}_{vv}\mathbf{U}^T\right] - \text{tr}\left[\mathbf{U}\mathbf{R}_{vv}\mathbf{R}_{yy}^{-1}\mathbf{R}_{vv}\mathbf{U}^T\right]. \tag{17}$$

Now we can define the normalized MMSE (NMMSE)

$$\tilde{J}(\mathbf{H}_{\text{W}}) = \frac{J(\mathbf{H}_{\text{W}})}{J(\mathbf{U})} = \frac{J(\mathbf{H}_{\text{W}})}{E\left[\mathbf{v}_1^T(k)\mathbf{v}_1(k)\right]} \tag{18}$$

where $0 \leq \tilde{J}(\mathbf{H}_{\text{W}}) \leq 1$. This definition is related to the speech-distortion index and the noise-reduction factor by the formula

$$\tilde{J}(\mathbf{H}_{\text{W}}) = \text{SNR} \cdot \upsilon_{\text{sd}}(\mathbf{H}_{\text{W}}) + \frac{1}{\xi_{\text{nr}}(\mathbf{H}_{\text{W}})}. \tag{19}$$

As a matter of fact, (19) is valid for any filter $\mathbf{H}$, i.e.,

$$\tilde{J}(\mathbf{H}) = \text{SNR} \cdot \upsilon_{\text{sd}}(\mathbf{H}) + \frac{1}{\xi_{\text{nr}}(\mathbf{H})}. \tag{20}$$

We deduce the two inequalities

$$\upsilon_{\text{sd}}(\mathbf{H}) \leq \frac{1}{\text{SNR}}\left[1 - \frac{1}{\xi_{\text{nr}}(\mathbf{H})}\right] \tag{21}$$

$$\xi_{\text{nr}}(\mathbf{H}) \geq \frac{1}{1 - \text{SNR} \cdot \upsilon_{\text{sd}}(\mathbf{H})}. \tag{22}$$

It can be shown that $\text{SNR}(\mathbf{H}_{\text{W}}) \geq \text{SNR}$ for any filter matrix dimension and for all possible speech and noise correlation matrices [1], [15], [16]. Using this property and expression (12), we deduce that

$$\text{SNR} \leq \text{SNR}(\mathbf{H}_{\text{W}}) \leq \text{SNR} \cdot \xi_{\text{nr}}(\mathbf{H}_{\text{W}}). \tag{23}$$

In the Wiener formulation, we do not explicitly exploit the spatial information. From (19) and (23), we can get this upper bound for $\text{SNR}(\mathbf{H}_{\text{W}})$

$$\text{SNR}(\mathbf{H}_{\text{W}}) \leq \frac{1}{\frac{\tilde{J}(\mathbf{H}_{\text{W}})}{\text{SNR}} - \upsilon_{\text{sd}}(\mathbf{H}_{\text{W}})} \tag{24}$$

which shows that the output SNR is improved at the expense of speech distortion.

A. Particular Case: Single Microphone and White Noise

We assume here that only one microphone signal is available (i.e., $N = 1$) and the noise picked up by this microphone is white (i.e., $\mathbf{R}_{v_1v_1} = \sigma_{v_1}^2\mathbf{I}_{L\times L}$). In this situation, the Wiener filter matrix becomes

$$\mathbf{H}_{\text{W}} = \mathbf{I}_{L\times L} - \sigma_{v_1}^2\mathbf{R}_{y_1y_1}^{-1} \tag{25}$$

where

$$\mathbf{R}_{y_1y_1} = \mathbf{R}_{x_1x_1} + \sigma_{v_1}^2\mathbf{I}_{L\times L}.$$

It is well known that the inverse of the Toeplitz matrix $\mathbf{R}_{y_1y_1}$ can be factorized as follows [17], [18]:

$$\mathbf{R}_{y_1y_1}^{-1} = \begin{bmatrix} 1 & -c_{10} & \cdots & -c_{(L-1)0} \\ -c_{01} & 1 & \cdots & -c_{(L-1)1} \\ \vdots & \vdots & \ddots & \vdots \\ -c_{0(L-1)} & -c_{1(L-1)} & \cdots & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{E_0} & 0 & \cdots & 0 \\ 0 & \frac{1}{E_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{E_{L-1}} \end{bmatrix} \quad (26)$$

where the columns of the first matrix on the right-hand side of (26) are the linear interpolators of the signal $y_1(k)$ and the elements E_l in the diagonal matrix are the respective interpolation-error powers.

Using the factorization of $\mathbf{R}_{y_1y_1}^{-1}$ in (17), the MMSE and NMMSE can be rewritten, respectively, as

$$J(\mathbf{H}_{\mathrm{W}}) = L\sigma_{v_1}^2 - \left(\sigma_{v_1}^2\right)^2 \sum_{l=0}^{L-1} \frac{1}{E_l} \quad (27)$$

$$\tilde{J}(\mathbf{H}_{\mathrm{W}}) = 1 - \frac{\sigma_{v_1}^2}{L} \sum_{l=0}^{L-1} \frac{1}{E_l}. \quad (28)$$

Assume that the noise-free speech signal $x_1(k)$ is very well predictable. In this scenario, $E_l \approx \sigma_{v_1}^2$, $\forall\ l$, and replacing this value in (28) we find that $\tilde{J}(\mathbf{H}_{\mathrm{W}}) \approx 0$. From (19), we then deduce that $\upsilon_{\mathrm{sd}}(\mathbf{H}_{\mathrm{W}}) \approx 0$ (almost no speech distortion) and $\xi_{\mathrm{nr}}(\mathbf{H}_{\mathrm{W}}) \approx \infty$ (almost infinite noise reduction). Notice that this result seems independent of the SNR. Also, since $\mathbf{H}_{\mathrm{W}}\mathbf{x}(k) \approx \mathbf{x}_1(k)$, this means that $\xi_{\mathrm{sr}}(\mathbf{H}_{\mathrm{W}}) \approx 1$; as a result $\mathrm{SNR}(\mathbf{H}_{\mathrm{W}}) \approx \infty$ and we can almost perfectly recover the signal $x_1(k)$.

In the other extreme case, let us see now what happens when the source signal $x_1(k)$ is not predictable at all. In this situation, $E_l \approx \sigma_{y_1}^2, \forall\, l$ and $c_{ij} \approx 0, \forall\, i, j$. Using these values, we get

$$\mathbf{H}_{\mathrm{W}} \approx \frac{\mathrm{SNR}}{1+\mathrm{SNR}} \mathbf{I}_{L\times L} \quad (29)$$

$$\tilde{J}(\mathbf{H}_{\mathrm{W}}) \approx \frac{\mathrm{SNR}}{1+\mathrm{SNR}}. \quad (30)$$

With the help of the two previous equations, it's straightforward to obtain

$$\xi_{\mathrm{nr}}(\mathbf{H}_{\mathrm{W}}) \approx \left(1 + \frac{1}{\mathrm{SNR}}\right)^2 \quad (31)$$

$$\upsilon_{\mathrm{sd}}(\mathbf{H}_{\mathrm{W}}) \approx \frac{1}{(1+\mathrm{SNR})^2} \quad (32)$$

$$\mathrm{SNR}(\mathbf{H}_{\mathrm{W}}) \approx \mathrm{SNR}. \quad (33)$$

While some noise reduction is achieved (at the price of speech distortion), there is no improvement in the output SNR, meaning that the Wiener filter has no positive effect on the microphone signal $y_1(k)$.

This analysis, even though simple, is quite insightful. It shows that the Wiener filter can really help achieve noise reduction as long as the source signal is somewhat predictable. However, in practice some discontinuities could be heard from a voiced signal to an unvoiced one, since for the former the noise will be mostly removed while it will not for the latter.

A surprising consequence of this analysis is the effect of reverberation. Indeed, even if the source signal $s(k)$ is white, thanks to the effect of the impulse response g_1, the signal $x_1(k)$ is not and becomes more "predictable." Hence, we make the following claim: for *white* source signal, reverberation helps make it more predictable and hence helps the Wiener filter for better noise reduction. We can draw the same kind of conclusion for any number of microphones.

V. Subspace Method

In the Wiener filter, we cannot control the compromise between noise reduction and speech distortion. So this filter derived from the classical MSE criterion may be limited in practice because of its lack of flexibility. Ephraim and Van Trees proposed, in the single-channel case, a more meaningful criterion which consists of minimizing the speech distortion while keeping the residual noise power below some given threshold [19]. The deduced optimal estimator is shown to be a Wiener filter with adjustable input noise level. This filter was developed in the white noise case. Since then, many algorithms have been proposed to deal with the general colored noise [20], [22]–[25]. We think that the most elegant algorithm is the one proposed by Hu and Loizou [22], [26] using the generalized eigenvalue decomposition.

Using the same signal model described in Section II, the optimal filter with the subspace technique can be mathematically derived from the optimization problem

$$\mathbf{H}_{\mathrm{S}} = \arg\min_{\mathbf{H}} J_x(\mathbf{H}) \ \text{ subject to } J_v(\mathbf{H}) \le L\sigma^2 \quad (34)$$

where

$$J_x(\mathbf{H}) \triangleq \mathrm{tr}\left\{E\left[\mathbf{e}_x(k)\mathbf{e}_x^T(k)\right]\right\} \quad (35)$$

$$J_v(\mathbf{H}) \triangleq \mathrm{tr}\left\{E\left[\mathbf{e}_v(k)\mathbf{e}_v^T(k)\right]\right\} \quad (36)$$

and $\sigma^2 < \sigma_{v_1}^2$ in order to have some noise reduction. If we use a Lagrange multiplier μ to adjoin the constraint to the cost function, (34) can be rewritten as

$$\mathbf{H}_{\mathrm{S}} = \arg\min_{\mathbf{H}} \mathcal{L}(\mathbf{H}, \mu) \quad (37)$$

where

$$\mathcal{L}(\mathbf{H}, \mu) = J_x(\mathbf{H}) + \mu\left[J_v(\mathbf{H}) - L\sigma^2\right] \quad (38)$$

and $\mu \ge 0$. We can easily prove from (37) that the optimal filter is

$$\begin{aligned} \mathbf{H}_{\mathrm{S}}^T &= \left(\mathbf{R}_{xx} + \mu\mathbf{R}_{vv}\right)^{-1}\mathbf{R}_{xx}\mathbf{U}^T \\ &= \left[\mathbf{R}_{yy} + (\mu-1)\mathbf{R}_{vv}\right]^{-1}\left[\mathbf{R}_{yy} - \mathbf{R}_{vv}\right]\mathbf{U}^T \\ &= \left[\mathbf{I}_{NL\times NL} + (\mu-1)\mathbf{R}_{yy}^{-1}\mathbf{R}_{vv}\right]^{-1}\mathbf{H}_{\mathrm{W}}^T \end{aligned} \quad (39)$$

where $\mathbf{R}_{xx} = E\left[\mathbf{x}(k)\mathbf{x}^T(k)\right]$ is the $NL \times NL$ correlation matrix of the speech signal at the different microphones and the

Lagrange multiplier satisfies $J_v(\mathbf{H}_{\mathrm{S}}) = L\sigma^2$, which implies that

$$\xi_{\mathrm{nr}}(\mathbf{H}_{\mathrm{S}}) = \frac{\sigma_{v_1}^2}{\sigma^2} > 1. \tag{40}$$

From (21), we get

$$\upsilon_{\mathrm{sd}}(\mathbf{H}_{\mathrm{S}}) \leq \frac{\sigma_{v_1}^2 - \sigma^2}{\sigma_{x_1}^2}. \tag{41}$$

Since $\tilde{J}(\mathbf{H}_{\mathrm{W}}) \leq \tilde{J}(\mathbf{H}_{\mathrm{S}}), \forall\, \mu$, we also have

$$\upsilon_{\mathrm{sd}}(\mathbf{H}_{\mathrm{S}}) \geq \upsilon_{\mathrm{sd}}(\mathbf{H}_{\mathrm{W}}) + \frac{1}{\mathrm{SNR}}\left[\frac{1}{\xi_{\mathrm{nr}}(\mathbf{H}_{\mathrm{W}})} - \frac{1}{\xi_{\mathrm{nr}}(\mathbf{H}_{\mathrm{S}})}\right]. \tag{42}$$

Therefore, $\xi_{\mathrm{nr}}(\mathbf{H}_{\mathrm{S}}) \geq \xi_{\mathrm{nr}}(\mathbf{H}_{\mathrm{W}})$ implies that $\upsilon_{\mathrm{sd}}(\mathbf{H}_{\mathrm{S}}) \geq \upsilon_{\mathrm{sd}}(\mathbf{H}_{\mathrm{W}})$. However, $\xi_{\mathrm{nr}}(\mathbf{H}_{\mathrm{S}}) \leq \xi_{\mathrm{nr}}(\mathbf{H}_{\mathrm{W}})$ does not imply that $\upsilon_{\mathrm{sd}}(\mathbf{H}_{\mathrm{S}}) \leq \upsilon_{\mathrm{sd}}(\mathbf{H}_{\mathrm{W}})$.

In practice, it is not easy to determine an optimal value of μ. Therefore, when this parameter is chosen in an ad hoc way, we can see that for

- $\mu = 1$, $\mathbf{H}_{\mathrm{S}} = \mathbf{H}_{\mathrm{W}}$;
- $\mu = 0$, $\mathbf{H}_{\mathrm{S}} = \mathbf{U}$;
- $\mu > 1$, results in low residual noise at the expense of high speech distortion;
- $\mu < 1$, we get little speech distortion but not so much noise reduction.

In the single-channel case, it can be shown that $\mathrm{SNR}(\mathbf{H}_{\mathrm{S}}) \geq \mathrm{SNR}$ [27]. The same kind of proof holds for any number of microphones [16].

As shown in [28], the two symmetric matrices $\mathbf{R}_{xx}$ and $\mathbf{R}_{vv}$ can be jointly diagonalized if $\mathbf{R}_{vv}$ is positive definite. This joint diagonalization was first used [21], [22], [26], [29] in the single-channel case. In our multichannel context as shown in [30], [31], we have

$$\mathbf{R}_{xx} = \mathbf{B}^T \mathbf{\Lambda} \mathbf{B} \tag{43}$$

$$\mathbf{R}_{vv} = \mathbf{B}^T \mathbf{B} \tag{44}$$

$$\mathbf{R}_{yy} = \mathbf{B}^T \left[\mathbf{I}_{NL\times NL} + \mathbf{\Lambda}\right] \mathbf{B} \tag{45}$$

where $\mathbf{B}$ is a full rank square matrix but not necessarily orthogonal, and the diagonal matrix

$$\mathbf{\Lambda} = \mathrm{diag}\left[\lambda_0 \quad \lambda_1 \quad \ldots \quad \lambda_{NL-1}\right] \tag{46}$$

are the eigenvalues of the matrix $\mathbf{R}_{vv}^{-1}\mathbf{R}_{xx}$ with $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{NL-1} \geq 0$.

Applying the decompositions (43)–(45) in (39), the optimal estimator becomes

$$\mathbf{H}_{\mathrm{S}} = \mathbf{U}\mathbf{B}^T \mathbf{\Lambda} \left(\mathbf{\Lambda} + \mu \mathbf{I}_{NL\times NL}\right)^{-1} \mathbf{B}^{-T}. \tag{47}$$

Therefore, the estimation of the speech signal $\mathbf{x}_1(k)$ is done in three steps: first we apply the transform $\mathbf{B}^{-T}$ to the noisy signal; second the transformed signal is modified by the gain function $\mathbf{\Lambda}\left(\mathbf{\Lambda} + \mu \mathbf{I}_{NL\times NL}\right)^{-1}$; and finally we modify back the signal to the time domain by applying the transform $\mathbf{U}\mathbf{B}^T$.

Usually, a speech signal can be modeled as a linear combination of a number of some (linearly independent) basis vectors smaller than the dimension of these vectors. As a result, the vector space of the noisy signal can be decomposed in two subspaces: the signal-plus-noise subspace of length L_{s} and the noise subspace of length L_{n}, with $NL = L_{\mathrm{s}} + L_{\mathrm{n}}$. This implies that the last L_{n} eigenvalues of the matrix $\mathbf{R}_{vv}^{-1}\mathbf{R}_{xx}$ are equal to zero. Therefore, we can rewrite (47) as

$$\mathbf{H}_{\mathrm{S}} = \mathbf{U}\mathbf{B}^T \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0}_{L_{\mathrm{s}}\times L_{\mathrm{n}}} \\ \mathbf{0}_{L_{\mathrm{n}}\times L_{\mathrm{s}}} & \mathbf{0}_{L_{\mathrm{n}}\times L_{\mathrm{n}}} \end{bmatrix} \mathbf{B}^{-T} \tag{48}$$

where

$$\mathbf{\Sigma} = \mathrm{diag}\left[\frac{\lambda_0}{\lambda_0 + \mu}, \frac{\lambda_1}{\lambda_1 + \mu}, \ldots, \frac{\lambda_{L_{\mathrm{s}}-1}}{\lambda_{L_{\mathrm{s}}-1} + \mu}\right] \tag{49}$$

is an $L_{\mathrm{s}} \times L_{\mathrm{s}}$ diagonal matrix. We now clearly see that noise reduction with the subspace method is achieved by nulling the noise subspace and cleaning the speech-plus-noise subspace via a reweighted reconstruction [32].

Like the Wiener filter, the optimal filter based on the subspace approach does not take fully advantage of the spatial information in order to minimize the distortion of the speech signal.

VI. Spatial–Temporal Prediction Approach

As explained in the previous sections, the fact that speech is partially predictable helps all algorithms in reducing the level of noise in the microphone signal $y_1(k)$. Implicitly, temporal prediction of the signal of interest plays a fundamental role in speech enhancement. What about spatial prediction? Is its role as important as temporal prediction? Since the speech signals picked up by the microphones come from a unique source, the same signals at microphones $2, \ldots, N$ can be predicted from the first microphone signal. Can this help?

In an earlier study of the authors [33], we proposed a novel algorithm in which both spatial and temporal prediction are explicitly exploited. We assume that we can find an $L \times L$ filter matrix $\mathbf{W}_{\mathrm{n}}$ such that

$$\mathbf{x}_{\mathrm{n}}(k) = \mathbf{W}_{\mathrm{n}}^T \mathbf{x}_1(k), \quad n = 2, \ldots, N. \tag{50}$$

We will see later on how to determine the optimal matrix $\mathbf{W}_{n,\mathrm{o}}$. Expression (50) can be seen as spatial–temporal prediction (STP), where we try to predict the microphone signal samples $\mathbf{x}_{\mathrm{n}}(k)$ from $\mathbf{x}_1(k)$.

Substituting (50) into (5), we find that

$$\mathbf{e}_x(k) = \left(\mathbf{H}\mathbf{W}^T - \mathbf{I}_{L\times L}\right) \mathbf{x}_1(k) \tag{51}$$

where

$$\mathbf{W} = \left[\mathbf{I}_{L\times L} \quad \mathbf{W}_2 \quad \ldots \quad \mathbf{W}_N\right]$$

is a matrix of size $L \times NL$.

In the single-channel case, there is no way we can reduce the level of the background noise without distorting the speech signal. In the Wiener filter (with one or more microphones), we minimize the classical MSE without much concern on the residual noise and speech distortion. In the subspace approach, we minimize the speech distortion while keeping the residual noise power below a threshold. However, from the STP approach, we see clearly that by using at least two microphones it is possible to have noise reduction with no speech distortion

[if (50) is met] by simply minimizing $J_v(\mathbf{H})$ with the constraint that $\mathbf{H}\mathbf{W}^T = \mathbf{I}_{L\times L}$. Therefore, our optimization problem is

$$\min_{\mathbf{H}} J_v(\mathbf{H}) \text{ subject to } \mathbf{I}_{L\times L} = \mathbf{H}\mathbf{W}^T. \tag{52}$$

By using Lagrange multipliers, we easily find the optimal solution

$$\mathbf{H}_{\mathrm{P}} = \left(\mathbf{W}\mathbf{R}_{vv}^{-1}\mathbf{W}^T\right)^{-1}\mathbf{W}\mathbf{R}_{vv}^{-1} \tag{53}$$

where we assumed that the noise signals $v_{\mathrm{n}}(k)$, $n = 1,2,\ldots,N$, are not perfectly coherent so that $\mathbf{R}_{vv}$ is not singular. Equation (53) has the same form as the linearly constrained minimum variance (LCMV) beamformer [34], [35]; however, the STP based-approach and the LCMV beamformer deal with two different problems.

The second step is to determine the filter matrix $\mathbf{W}$ for spatial-temporal prediction. An optimal estimator, in the Wiener sense, can be obtained by minimizing the following cost function:

$$J_{\mathrm{f}}(\mathbf{W}_{\mathrm{n}}) = E\left\{\left[\mathbf{x}_{\mathrm{n}}(k) - \mathbf{W}_{\mathrm{n}}^T\mathbf{x}_1(k)\right]^T\left[\mathbf{x}_{\mathrm{n}}(k) - \mathbf{W}_{\mathrm{n}}^T\mathbf{x}_1(k)\right]\right\}. \tag{54}$$

We easily find the optimal STP filter

$$\mathbf{W}_{n,\mathrm{o}}^T = \mathbf{R}_{x_{\mathrm{n}}x_1}\mathbf{R}_{x_1x_1}^{-1} \tag{55}$$

where $\mathbf{R}_{x_{\mathrm{n}}x_1}=E\left\{\mathbf{x}_{\mathrm{n}}(k)\mathbf{x}_1^T(k)\right\}$ and $\mathbf{R}_{x_1x_1}=E\left\{\mathbf{x}_1(k)\mathbf{x}_1^T(k)\right\}$ are the cross-correlation and autocorrelation matrices of the microphone signals, respectively. Using (2), we know that

$$\mathbf{R}_{x_{\mathrm{n}}x_1} = \mathbf{G}_{\mathrm{n}}^T\mathbf{R}_{ss}\mathbf{G}_1, \quad n = 1,2,\ldots,N \tag{56}$$

where $\mathbf{R}_{ss} = E\left\{\mathbf{s}_{L'}(k)\mathbf{s}_{L'}^T(k)\right\}$ is the autocorrelation matrix of the source signal. Substituting (56) into (55) produces

$$\mathbf{W}_{n,\mathrm{o}}^T = \mathbf{G}_{\mathrm{n}}^T\mathbf{R}_{ss}\mathbf{G}_1\left[\mathbf{G}_1^T\mathbf{R}_{ss}\mathbf{G}_1\right]^{-1}. \tag{57}$$

If the source signal is white, then $\mathbf{R}_{ss} = \sigma_{\mathrm{s}}^2\mathbf{I}_{L'\times L'}$, where σ_{s}^2 is the variance of the source signal, and (57) becomes

$$\mathbf{W}_{n,\mathrm{o}}^T = \mathbf{G}_{\mathrm{n}}^T\mathbf{G}_1\left[\mathbf{G}_1^T\mathbf{G}_1\right]^{-1} \tag{58}$$

which is a function merely of the channel impulse responses. In this particular case, the STP matrix $\mathbf{W}_{n,\mathrm{o}}$ can be seen as the time-domain counterpart of the transfer-function ratio (TFR), and hence the STP solution (53) is in principle equivalent to the TF-GSC approach [36]. However, in the real world, speech is not white, and so $\mathbf{W}_{n,\mathrm{o}}$ depends not only on the channel impulse responses, but also on the second-order statistics of the speech source. This indicates that the prediction matrix $\mathbf{W}_{n,\mathrm{o}}$ has exploited both spatial correlation between the channels and short-term temporal audio-correlation of speech.

In practice, the signals $x_{\mathrm{n}}(k)$, $n = 1,2,\ldots,N$ are not observable. So the Wiener filter matrix, as given in (55), cannot be estimated. However, using $\mathbf{x}_{\mathrm{n}}(k) = \mathbf{y}_{\mathrm{n}}(k) - \mathbf{v}_{\mathrm{n}}(k)$, we can verify that

$$\mathbf{R}_{x_{\mathrm{n}}x_1} = \mathbf{R}_{y_{\mathrm{n}}y_1} - \mathbf{R}_{v_{\mathrm{n}}v_1}, \quad n = 1,2,\ldots,N \tag{59}$$

where $\mathbf{R}_{y_{\mathrm{n}}y_1}=E\left\{\mathbf{y}_{\mathrm{n}}(k)\mathbf{y}_1^T(k)\right\}$ and $\mathbf{R}_{v_{\mathrm{n}}v_1}=E\left\{\mathbf{v}_{\mathrm{n}}(k)\mathbf{v}_1^T(k)\right\}$. As a result

$$\mathbf{W}_{n,\mathrm{o}}^T = \left(\mathbf{R}_{y_{\mathrm{n}}y_1} - \mathbf{R}_{v_{\mathrm{n}}v_1}\right)\left(\mathbf{R}_{y_1y_1} - \mathbf{R}_{v_1v_1}\right)^{-1}. \tag{60}$$

The optimal filter matrix depends now only on the second order statistics of the observation and noise signals. The statistics of the noise signals can be estimated during silences [when $s(k) = 0$] if we assume that the noise is stationary so that its statistics can be used for a next frame when the speech is active. We also assume that a voice activity detector (VAD) is available so that the optimal STP filter matrix is estimated only when the speech source is active. Finally, the optimal filter matrix based on STP is given by

$$\mathbf{H}_{\mathrm{P}} = \left(\mathbf{W}_{\mathrm{o}}\mathbf{R}_{vv}^{-1}\mathbf{W}_{\mathrm{o}}^T\right)^{-1}\mathbf{W}_{\mathrm{o}}\mathbf{R}_{vv}^{-1} \tag{61}$$

where

$$\mathbf{W}_{\mathrm{o}} = \left[\mathbf{I}_{L\times L} \quad \mathbf{W}_{2,\mathrm{o}} \quad \ldots \quad \mathbf{W}_{N,\mathrm{o}}\right].$$

In general, we do not have exactly $\mathbf{x}_{\mathrm{n}}(k) = \mathbf{W}_{n,\mathrm{o}}^T\mathbf{x}_1(k)$, so that some speech distortion is expected. However, for large filter matrices, we can approach this equality so that this distortion can be kept low. In this case, it can be verified that

$$\upsilon_{\mathrm{sd}}(\mathbf{H}_{\mathrm{P}}) \approx 0 \tag{62}$$

$$\xi_{\mathrm{sr}}(\mathbf{H}_{\mathrm{P}}) \approx 1 \tag{63}$$

$$\xi_{\mathrm{nr}}(\mathbf{H}_{\mathrm{P}}) = \frac{L\sigma_{v_1}^2}{\mathrm{tr}\left[\left(\mathbf{W}_{\mathrm{o}}\mathbf{R}_{vv}^{-1}\mathbf{W}_{\mathrm{o}}^T\right)^{-1}\right]} \approx \frac{1}{\tilde{J}(\mathbf{H}_{\mathrm{P}})} \geq 1 \tag{64}$$

which implies that

$$\mathrm{SNR}(\mathbf{H}_{\mathrm{P}}) \approx \mathrm{SNR}\cdot\xi_{\mathrm{nr}}(\mathbf{H}_{\mathrm{P}}) \geq \mathrm{SNR}. \tag{65}$$

Also, since $\tilde{J}(\mathbf{H}_{\mathrm{W}}) \leq \tilde{J}(\mathbf{H}_{\mathrm{P}})$, we have $\xi_{\mathrm{nr}}(\mathbf{H}_{\mathrm{P}}) \leq \xi_{\mathrm{nr}}(\mathbf{H}_{\mathrm{W}})$.

Clearly, we see that this approach has the potential to introduce minimum distortion to the speech signal thanks to the fact that the microphone observations of the source signal are spatially predictable.

VII. Simulations

We have carried out a number of simulations to experimentally study the three main multichannel noise reduction algorithms (Wiener filter, subspace, and spatial-temporal prediction) in real acoustic environments under different operation conditions. In this section, we will present the results, which highlight the merits and limitations inherent in these techniques, and justify what we learned through theoretical analyses in the previous sections. In these experiments, we use the output SNR and speech-distortion index defined in Section III as the performance measures.

A. Acoustic Environments and Experimental Setup

The simulations were conducted with the impulse responses measured in the varechoic chamber at Bell Labs [37]. A diagram of the floor plan layout is shown in Fig. 1. For convenience, positions in the floor plan are designated by (x, y) coordinates with reference to the southwest corner and corresponding to

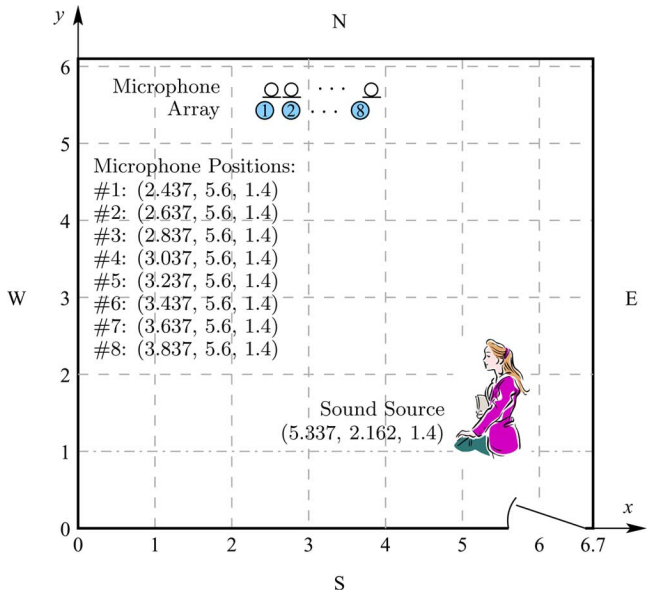


Fig. 1. Floor plan of the varechoic chamber at Bell Labs (coordinate values measured in meters).

meters along the (South, West) walls. The chamber measures $x = 6.7$ m wide by $y = 6.1$ m deep by $z = 2.9$ m high. It is a rectangular room with 368 electronically controlled panels that vary the acoustic absorption of the walls, floor, and ceiling [38]. Each panel consists of two perforated sheets whose holes, if aligned, expose sound absorbing material (fiberglass) behind, but if shifted to misalign, form a highly reflective surface. The panels are individually controlled so that the holes on one particular panel are either fully open (absorbing state) or fully closed (reflective state). Therefore, by varying the binary state of each panel in any combination, 2^{238} different room characteristics can be simulated. In the database of channel impulse responses from [37], there are four panel configurations with 89%, 75%, 30%, and 0% of panels open, respectively corresponding to approximately 240, 310, 380, and 580 ms 60-dB reverberation time T_{60} in the 20–4000 Hz band. In our study, all four configurations were used to evaluate the performance of the noise-reduction algorithms. However, for conciseness and also due to space limitations, we present here only the results for the least and the most reverberant environments, i.e., $T_{60} = 240$ ms and 580 ms, respectively.

A linear microphone array which consists of 22 omnidirectional microphones was employed in the measurement and the spacing between adjacent microphones is about 10 cm. The array was mounted 1.4 m above the floor and parallel to the North wall at a distance of 50 cm. A loudspeaker was placed at 31 different prespecified positions to measure the impulse response to each microphone. In the simulations, no more than eight microphones will be chosen, and the sound source is fixed at one loudspeaker position. The positions of the microphones and the sound source are shown in Fig. 1.

Signals were sampled at 8 kHz, and the length of the measured impulse responses is of 4096 samples. We have tried different source signals. However, again due to space limitations, we present in this paper only the results using a female speech

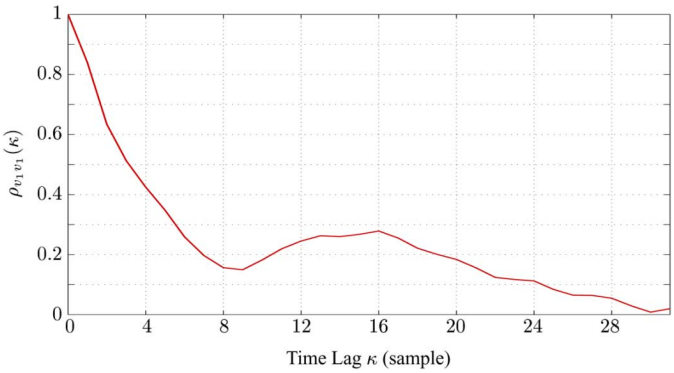


Fig. 2. Autocorrelation coefficient $\rho_{v_1v_1}(\kappa)$ of the additive NYSE babbling noise at the first microphone.

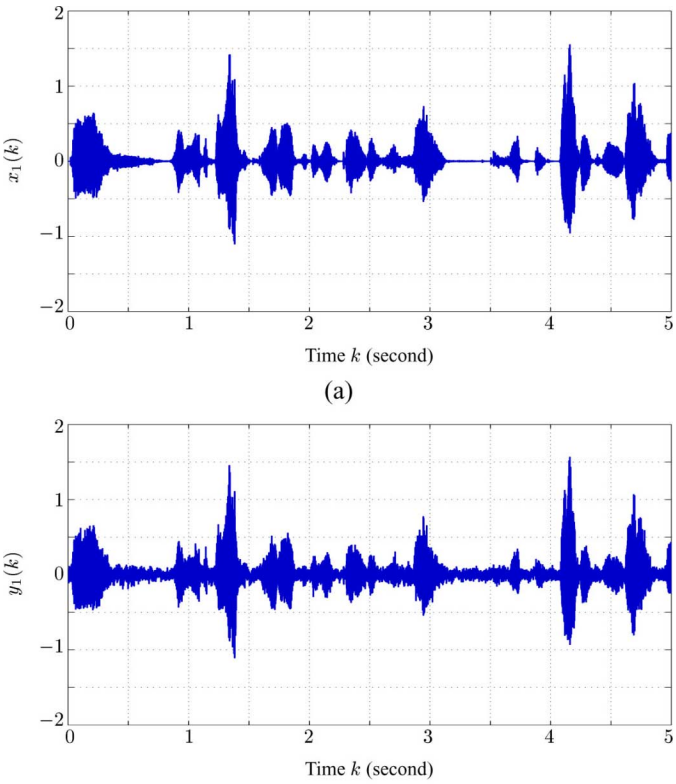


Fig. 3. Waveforms of (a) $x_1(k)$ and (b) $y_1(k)$ for SNR $= 10$dB and $T_{60} = 240$ ms.

signal as the source. We compute the microphone outputs by convolving the source signal and the corresponding channel impulse responses. The additive noise is babbling noise recorded in the New York Stock Exchange (NYSE). Different segments of the recorded babbling noise were used at different microphones to avoid perfectly coherent additive noise. The NYSE babbling noise is colored as seen by its autocorrelation coefficient

$$\rho_{v_1v_1}(\kappa) = E\{v_1(k)v_1(k-\kappa)\} \tag{66}$$

which is computed with a batch method and plotted in Fig. 2. The SNR at the microphones is fixed at 10 dB. Fig. 3 shows the first 5 s of $x_1(k)$ and $y_1(k)$ obtained in the environment with $T_{60} = 240$ ms.

The source signal is 30 s long. The first 5 s of the microphone outputs are used to compute the initial estimates of $\mathbf{R}_{yy}$ and $\mathbf{R}_{vv}$. The last 21 s are then used for performance evaluation of the noise-reduction algorithms. In this procedure, the estimates of $\mathbf{R}_{yy}$ and $\mathbf{R}_{vv}$ are recursively updated according to

$$\mathbf{R}_{yy}(k) = \lambda_y \mathbf{R}_{yy}(k-1) + (1-\lambda_y)\mathbf{y}(k)\mathbf{y}^T(k) \quad (67)$$

$$\mathbf{R}_{vv}(k) = \lambda_v \mathbf{R}_{vv}(k-1) + (1-\lambda_v)\mathbf{v}(k)\mathbf{v}^T(k) \quad (68)$$

where $0 < \lambda_y < 1$ and $0 < \lambda_v < 1$ are the forgetting factors. Note that in these simulations, VAD was not implemented and the additive noise sequence was directly used to estimate $\mathbf{R}_{vv}(k)$.

B. Experimental Results

1) Wiener Filter With Various Numbers of Microphones and Filter Lengths: Let us first investigate the Wiener filter algorithm for noise reduction using various numbers of microphones and filter lengths. The performance of the optimal Wiener filter obtained here will be used as a benchmark for comparison with other noise-reduction algorithms in the following experiments. We take $\lambda_y = \lambda_v = 0.9975$. The output SNR and speech-distortion index are plotted in Fig. 4.

It is clearly demonstrated that by using moderately more microphones and longer filters, the Wiener filter can effectively boost the output SNR at the price of introducing more speech distortion. However, this trend is not monotonic over N and L. We see that for $N = 8$, the output SNR drops after L exceeds 20. Note that the relative time delay of arrival between the first and the eighth microphones is about 18 samples.

2) Effect of the Forgetting Factor on the Performance of the Wiener Filter: In the development of the Wiener filter as well as other algorithms for noise reduction, we assume the knowledge of $\mathbf{R}_{yy}$ and $\mathbf{R}_{vv}$. As a result, one may unfortunately overlook the importance and under-evaluate the difficulty of accurately estimating these statistics (though they are only second order) in practice. Actually the forgetting factor plays a critical role in tuning a noise-reduction algorithm. On one hand, if the forgetting factor is too large (close to 1), the recursive estimate of $\mathbf{R}_{yy}(k)$ according to (67) is essentially a long-term average and cannot follow the short-term variation of speech signals. Consequently, the potential for greater noise reduction is not fully taken advantage of. On the other hand, if the forgetting factor is too small (much less than 1), then the recursive estimate of $\mathbf{R}_{yy}(k)$ is more likely rank deficient. This leads to the numerical stability problem when computing the inverse of $\mathbf{R}_{yy}(k)$, and hence causes performance degradation. Therefore, a proper forgetting factor is the one that helps achieve the balance between tracking capability and numerical stability. In this experiment, we would like to study this effect of the forgetting factor. We consider the Wiener filter again in the environment of $T_{60} = 240$ ms.

Fig. 5 depicts the results of six systems under investigation. These curves visibly justify the tradeoff effect mentioned above. Note that the size of $\mathbf{R}_{yy}(k)$ is $NL \times NL$. It is clear from Fig. 5 that the greater NL and the larger the size of $\mathbf{R}_{yy}(k)$, the greater is the optimal forgetting factor. The Wiener filters with the same value of NL perform almost identically against the forgetting factor regardless of the combination of N and L.

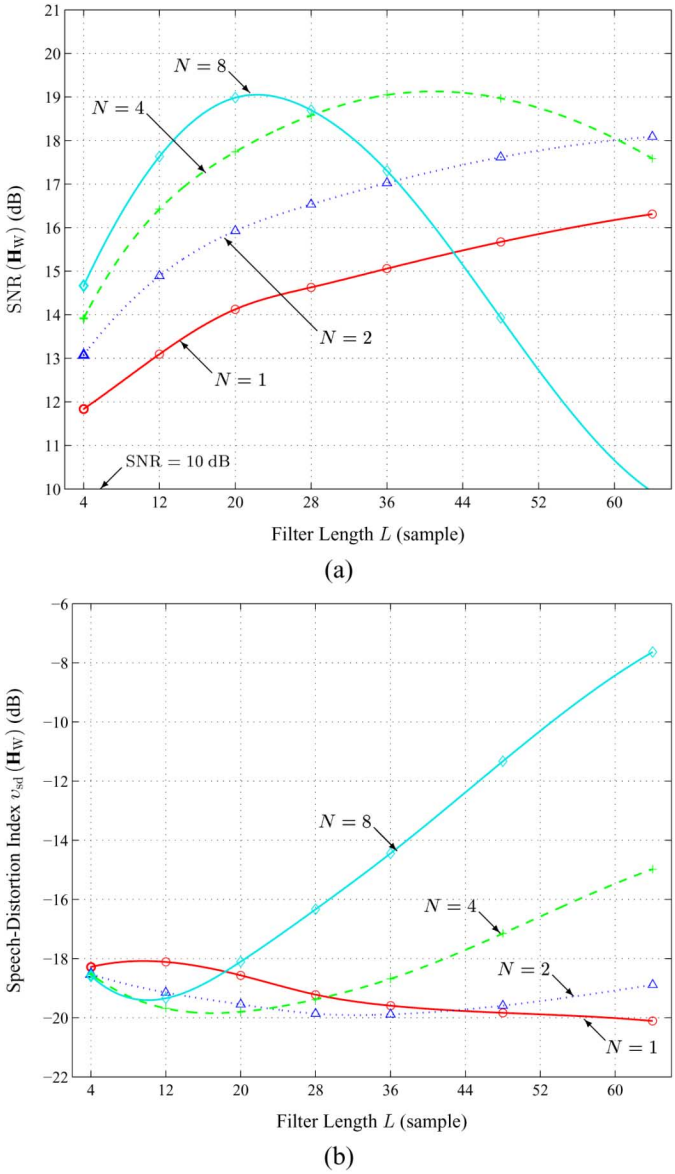


Fig. 4. Performance of the Wiener filter for noise reduction using various numbers of microphones $N = 1$, 2, 4, and 8, respectively. (a) Output SNR, and (b) speech-distortion index. Input SNR = 10 dB, room reverberation time $T_{60} = 240$ ms, and the forgetting factor $\lambda_y = \lambda_v = 0.9975$.

3) Performance Evaluation of the Subspace Method: In the first two experiments, we studied the Wiener filter for noise reduction under various operation conditions. Now we turn to the subspace method. Again, we take $\lambda_y = \lambda_v = 0.9975$ and $N = 2$. This experiment was carried out in two acoustic environments with $T_{60} = 240$ ms and 580 ms, respectively, and μ varies from 0.5, 1.0, to 2.0. Note that when $\mu = 1$, the subspace method is theoretically equivalent to the Wiener filter. The results are plotted in Fig. 6. It is evident that by decreasing μ, speech distortion is reduced but we gain little noise reduction. In the opposite direction, increasing μ results in low residual noise at the expense of high speech distortion.

4) Performance Evaluation of the Spatial–Temporal Prediction Approach: In the last but probably the most interesting ex-

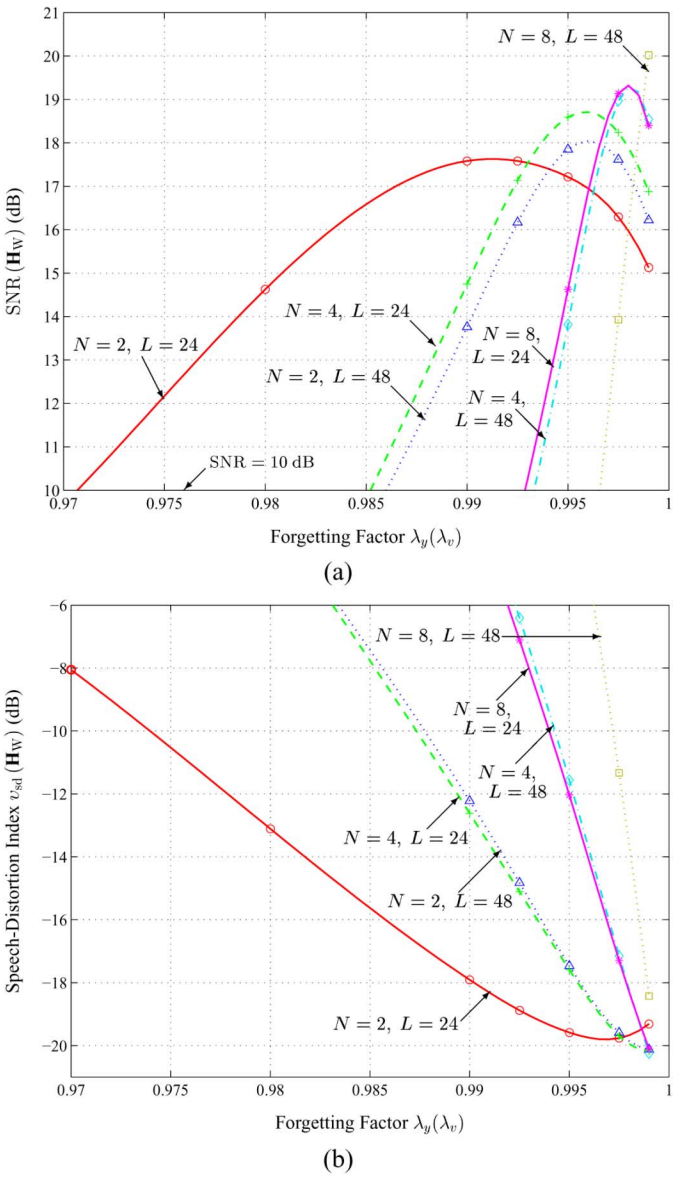


Fig. 5. Effect of the forgetting factors ($\lambda_y = \lambda_v$) on the performance of the Wiener filter for noise reduction. (a) Output SNR. (b) Speech-distortion index. Input SNR $= 10$ dB and room reverberation time $T_{60} = 240$ ms.

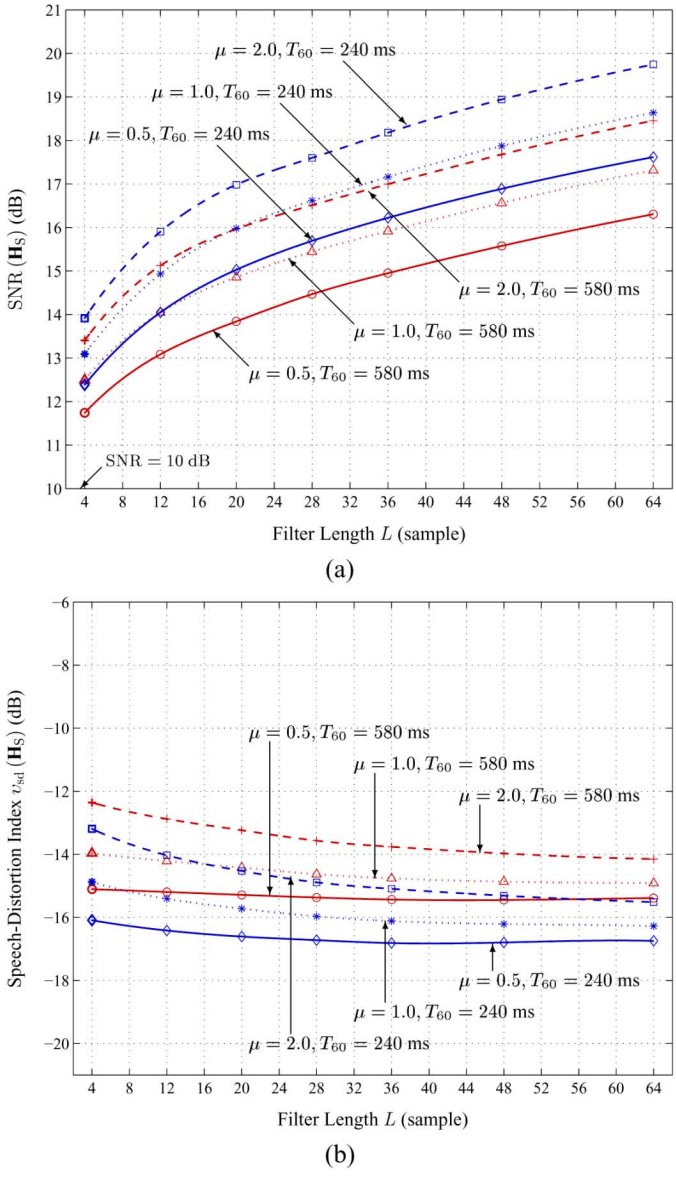


Fig. 6. Performance of the subspace algorithm for noise reduction using different values for μ in two different acoustic environments with $T_{60} = 240$ ms and 580 ms, respectively. (a) Output SNR. (b) Speech-distortion index. Input SNR $= 10$ dB, $N = 2$, and $\lambda_y = \lambda_v = 0.9975$.

periment, we tested the STP approach to noise reduction in comparison with the Wiener filter.

In our study, we learned that the performance of the Wiener filter and the subspace method is limited by the aforementioned numerical stability problem. By inspecting (16) and (39), we know that in the Wiener filter and subspace algorithms, we need to compute the inverse of $\mathbf{R}_{yy}$, which is of dimension $NL \times NL$. When we intend to use more microphones and longer filters (i.e., larger N and L) for a greater output SNR as well as less speech distortion, the covariance matrix $\mathbf{R}_{yy}$ becomes larger in size, which leads to the following two drawbacks:

- Using a short-term average, a larger error can be expected in the estimate $\mathbf{R}_{yy}(k)$. However, with a long-term average, the variation of speech statistics cannot be well followed. Both cause performance degradation. The larger $\mathbf{R}_{yy}$, the more prominent is the dilemma.
- The estimate of the covariance matrix $\mathbf{R}_{yy}(k)$ becomes more ill-conditioned (with a larger condition number) when NL gets larger. As a result, it is more problematic to find its inverse.

Therefore, as revealed by the results in the previous experiments, we do not gain what we expect from the Wiener filter and subspace algorithms by increasing N and L.

Alternatively, the STP approach utilizes the spatial correlation among the outputs of a microphone array with respect to a speech source separately only in the first step of determining an STP matrix. If we look closer at (60), we can recognize that the STP is proceeded on a pair-by-pair basis. In this procedure, only $\mathbf{R}_{x_1x_1}$ or equivalently $(\mathbf{R}_{y_1y_1} - \mathbf{R}_{v_1v_1})$ needs to be inverted. This matrix is $L \times L$ and does not grow in size with the number of microphones that we use. In addition, from (53), we know that $\mathbf{R}_{vv}$ rather than $\mathbf{R}_{yy}$ needs to be inverted in computing $\mathbf{H}_{\mathrm{P}}$.

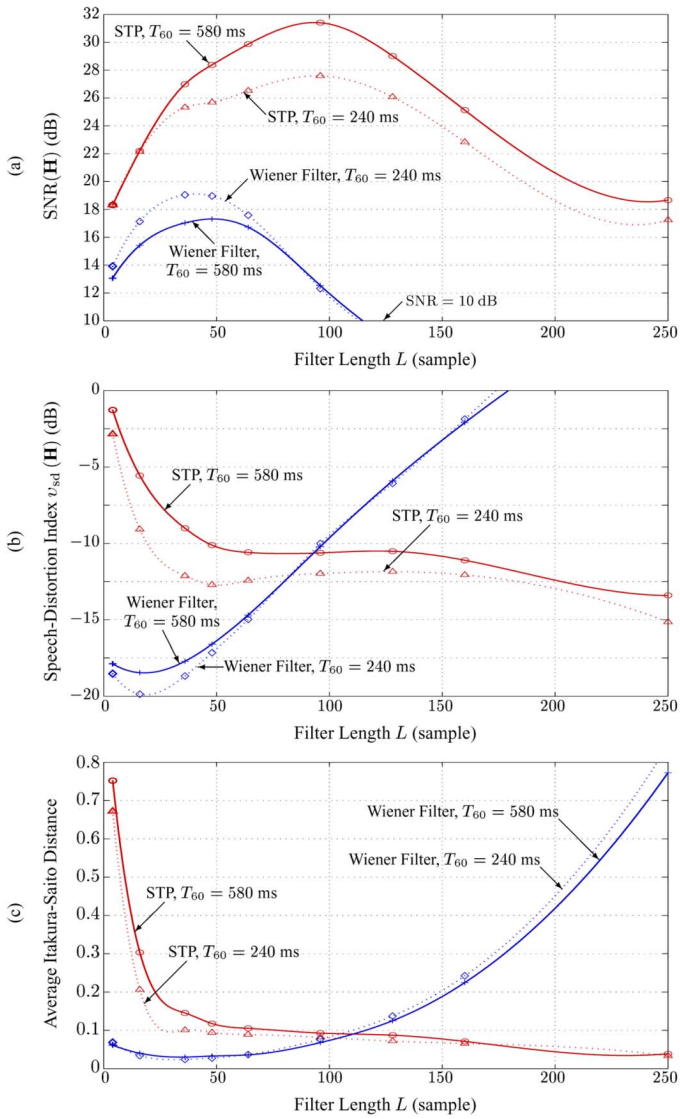


Fig. 7. Performance comparison between the STP and the Wiener filter algorithms for noise reduction in two different acoustic environments with $T_{60} = 240$ ms and 580 ms, respectively. (a) Output SNR. (b) Speech-distortion index. (c) Average Itakura-Saito distance. Input SNR $= 10$ dB and $N = 4$. For the Wiener filter $\lambda_y = \lambda_v = 0.9975$ and for the STP $\lambda_y = \lambda_v = 0.98$.

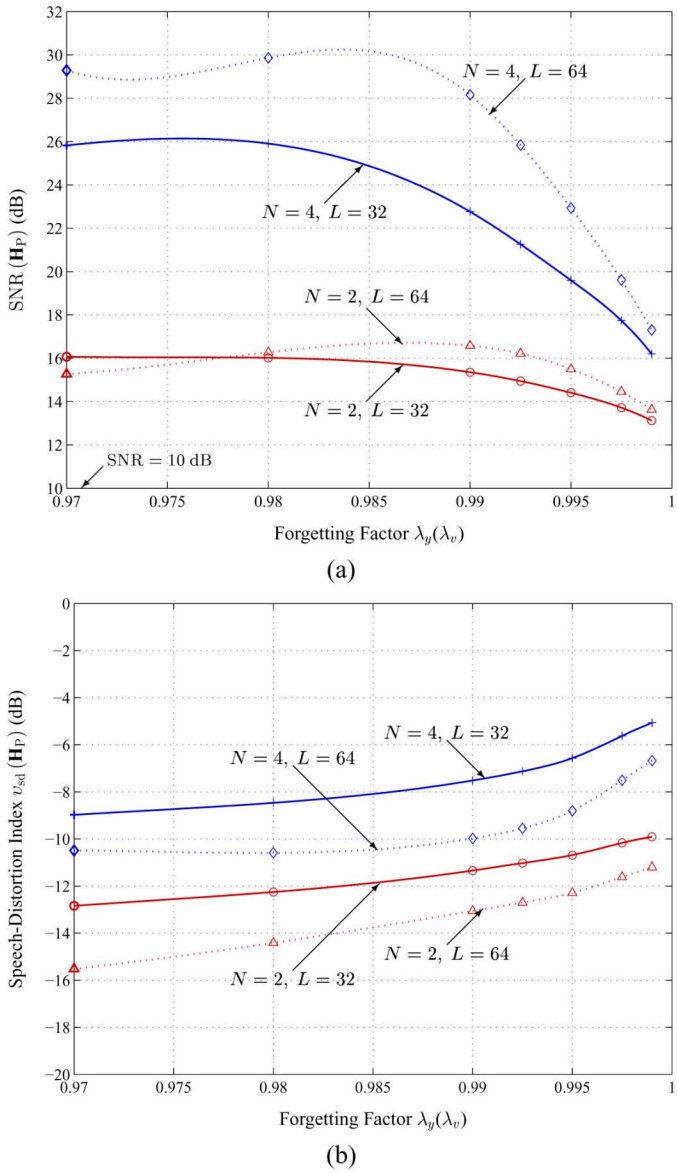


Fig. 8. Effect of the forgetting factors ($\lambda_y = \lambda_v$) on the performance of the STP algorithm for multichannel noise reduction. (a) Output SNR. (b) Speech-distortion index. Input SNR $= 10$ dB and room reverberation time $T_{60} = 580$ ms.

In most applications, the noise signals have flatter spectra and are relatively more stationary. Consequently, $\mathbf{R}_{vv}$ usually has a low condition number and can be accurately estimated with a long-term average. Therefore, with the STP algorithm, we can use a larger system with more microphones and longer filters for better performance.

Fig. 7 shows the results of the performance comparison between the STP and Wiener filter algorithms. We specified $N = 4$, $\lambda_y = \lambda_v = 0.9975$ for the Wiener filter, and $\lambda_y = \lambda_v = 0.98$ for the STP. The results for $T_{60} = 240$ ms and 580 ms are presented. In addition to the output SNR and the speech-distortion index, the average Itakura–Saito (IS) distance between $x_1(k)$ and $\mathbf{Hx}(k)$ is also plotted. The IS distance [39] exhibits a high correlation (0.59) with subjective judgments [40]. Many experiments in speech recognition show that if the IS measure is less than about 0.1, the two examined spectra are perceptually nearly identical. We see that in this experiment, when $v_{\mathrm{sd}}(\mathbf{H})$ is lower than -10 dB, the IS distance is approximately less than 0.1 for both the Wiener filter and the STP algorithms. Given the constraint that the IS distance is less than 0.1, the STP apparently can yield a much higher output SNR than the Wiener filter. For the STP algorithm, if L is too small, the prediction given by (50) cannot be accurate, and therefore the speech distortion as expressed by (51) would be significantly strong. We see that in this simulation, only when L is greater than 48, the speech distortion reaches an acceptable level.

In Fig. 8, we visualize the performance sensitivity of the STP algorithm to the change of the forgetting factors. We see that the performance of the STP algorithm is not sensitive to the forgetting factors. This implies that the STP algorithm is very easy to tune and is very robust to different acoustic conditions, which are very appealing features in practice.

VIII. Conclusion

Noise reduction is a very difficult problem and still remains a challenge today even after 40 years of tremendous progress. While some useful and interesting solutions exist in the single-microphone case at the price of distorting the desired speech signal, we will not draw the same conclusion with multiple microphones. From a theoretical point of view, though, it is possible to reduce noise with no speech distortion with a microphone array. However, the derivation of a practical solution is still an open area of research. In this paper, we studied three important multichannel noise reduction algorithms, namely, the classical Wiener filter, subspace, and the novel spatial–temporal prediction approaches. We showed their potentials and limitations via theoretical analysis and numerical simulations. The simulation results indicate that the spatial–temporal prediction approach is a very promising technique. It can achieve a much higher gain in the output SNR while keeping the speech distortion at a reasonable low level. In addition, it is easy to tune and is very robust to various acoustic conditions in practice.

References

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

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

[3] M. R. Schroeder, "Processing of communication signals to reduce effects of noise," U.S. Patent 3 403 224, Sep. 24, 1965.

[4] 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.

[5] 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.

[6] J. Chen, J. Benesty, Y. Huang, and E. J. Diethorn, "Fundamentals of noise reduction," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Berlin, Germany: Springer-Verlag, 2008.

[7] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment*. Chichester, U.K.: Wiley, 2006.

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

[9] J. Benesty, J. Chen, Y. Huang, and J. Dmochowski, "On microphone-array beamforming from a MIMO acoustic signal processing perspective," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 3, pp. 1053–1065, Mar. 2007.

[10] Y. Huang, J. Benesty, and J. Chen, "A blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 882–895, Sep. 2005.

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

[12] J. E. Greenberg, P. M. Peterson, and P. M. Zurek, "Intelligibility-weighted measures of speech-to-interference ratio and speech system performance," *J. Acoust. Soc. Amer.*, vol. 94, pp. 3009–3010, Nov. 1993.

[13] 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.

[14] S. Doclo, "Multi-microphone noise reduction and dereverberation techniques for speech applications," Ph.D dissertation, Katholieke Universiteit Leuven, Leuven, Belgium, 2003.

[15] 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.

[16] 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.

[17] J. Benesty and T. Gaensler, "New insights into the RLS algorithm," *EURASIP J. Appl. Signal Process.*, vol. 2004, pp. 331–339, Mar. 2004.

[18] S. Kay, "Some results in linear interpolation theory," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, no. 3, pp. 746–749, Jun. 1983.

[19] 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.

[20] 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. Adv. Signal Process.*, vol. 2007, pp. 15–15, 2007.

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

[22] Y. Hu and P. C. Loizou, "A subspace approach for enhancing speech corrupted by colored noise," *IEEE Signal Process. Lett.*, vol. 9, no. 6, pp. 204–206, Jul. 2002.

[23] 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.

[24] 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.

[25] 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.

[26] Y. Hu and P. C. Loizou, "A subspace approach for enhancing speech corrupted by colored noise," in *Proc. IEEE ICASSP*, 2002, pp. I-573–I-576.

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

[28] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. San Diego, CA: Academic, 1990.

[29] 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.

[30] F. Asano, S. Hayamizu, T. Yamada, and S. Nakamura, "Speech enhancement based on the subspace method," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 497–507, Sep. 2000.

[31] S. Doclo and M. Moonen, "Combined frequency-domain dereverberation and noise reduction technique for multi-microphone speech enhancement," in *Proc. Int. Workshop Acoust. Echo and Noise Control*, 2001, pp. 31–34.

[32] P. C. Hansen and S. H. Jensen, "FIR filter representations of reduced-rank noise reduction," *IEEE Trans. Signal Process.*, vol. 46, no. 6, pp. 1737–1741, Jun. 1998.

[33] J. Chen, J. Benesty, and Y. Huang, "A minimum distortion noise reduction algorithm with multiple microphones," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 3, pp. 481–493, Mar. 2008.

[34] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 10, pp. 1365–1376, Oct. 1987.

[35] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[36] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[37] A. Härmä, "Acoustic measurement data from the varechoic chamber," Agere Systems, Tech. Memo., 2001.

[38] W. C. Ward, G. W. Elko, R. A. Kubli, and W. C. McDougald, "The new varechoic chamber at AT&T Bell Labs," in *Proc. Wallance Clement Sabine Centennial Symp.*, 1994, pp. 343–346.

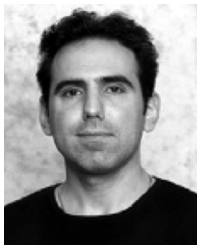
[39] L. R. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[40] S. R. Quackenbush, T. P. Barnwell, and M. A. Clements, *Objective Measures of Speech Quality*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

Yiteng (Arden) Huang (S'97–M'01) received the B.S. degree from 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 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. He served as an Associate Editor 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.



Jacob Benesty (M'92–SM'04) was born in 1963. He received the Masters degree in microwaves from Pierre & Marie Curie University, France, in 1987, and the Ph.D. degree in control and signal processing from Orsay University, France, in April 1991. During his Ph.D. program (from November 1989 to April 1991), he worked on adaptive filters and fast algorithms at the Centre National d'Etudes des Telecommunications (CNET), Paris, France.

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.

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 Co-Chair of the 1999 International Workshop on Acoustic Echo and Noise Control. 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).

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 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, Alcatel-Lucent, 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 is the recipient of 1998–1999 research grant from the Japan Key Technology Center, and the 1996–1998 President's Award from the Chinese Academy of Sciences. He helped organize the 2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), and he will serve as a technical Co-Chair of the 2009 WASPAA.