Aalborg Universitet

Noise Reduction with Optimal Variable Span Linear Filters

Jensen, Jesper Rindom; Benesty, Jacob; Christensen, Mads Græsbøll

Published in:
I E E E Transactions on Audio, Speech and Language Processing

DOI (link to publication from Publisher):
10.1109/TASLP.2015.2505416

Publication date:
2016

Document Version
Accepted author manuscript, peer reviewed version

Link to publication from Aalborg University

Citation for published version (APA):
Jensen, J. R., Benesty, J., & Christensen, M. G. (2016). Noise Reduction with Optimal Variable Span Linear Filters. *I E E E Transactions on Audio, Speech and Language Processing*, *24*(4), 631-644. Advance online publication. https://doi.org/10.1109/TASLP.2015.2505416

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: April 25, 2024

Noise Reduction with Optimal Variable Span Linear Filters

Jesper Rindom Jensen*, *Member, IEEE,* Jacob Benesty, and Mads Græsbøll Christensen, *Senior Member, IEEE.*

***Abstract*—In this paper, the problem of noise reduction is addressed as a linear filtering problem in a novel way by using concepts from subspace-based enhancement methods, resulting in variable span linear filters. This is done by forming the filter coefficients as linear combinations of a number of eigenvectors stemming from a joint diagonalization of the covariance matrices of the signal of interest and the noise. The resulting filters are flexible in that it is possible to trade off distortion of the desired signal for improved noise reduction. This tradeoff is controlled by the number of eigenvectors included in forming the filter. Using these concepts, a number of different filter designs are considered, like minimum distortion, Wiener, maximum SNR, and tradeoff filters. Interestingly, all these can be expressed as special cases of variable span filters. We also derive expressions for the speech distortion and noise reduction of the various filter designs. Moreover, we consider an alternative approach, wherein the filter is designed for extracting an estimate of the noise signal, which can then be extracted from the observed signals, which is referred to as the indirect approach. Simulations demonstrate the advantages and properties of the variable span filter designs, and their potential performance gain compared to widely used speech enhancement methods.**

***Index Terms*—Noise reduction, speech enhancement, joint diagonalization, span, subspace, optimal filters.**

I. Introduction

Noise reduction, or speech enhancement as it is commonly called in speech processing, is the art of reducing the influence of additive noise on a signal of interest. Such additive noise occurs naturally in many important applications, including hearing aids, teleconferencing, and mobile telephony. Examples of commonly occurring noises are babble, car, traffic, and air conditioning noise. The additive noise should be attenuated as much as possible, which is measured in terms of noise reduction, but, at the same time, the noise reduction process may distort the signal of interest, which, in the case of speech signals, is measured in terms of speech distortion. These two criteria are conflicting, something that can easily be seen by observing the trivial (and uninteresting) extremes: by multiplying the observed signal by zero, all noise is removed, but a maximum of speech distortion is also obtained. Conversely, by doing nothing at all, no speech distortion is incurred but no noise is reduced either. A challenge for current noise reduction research is thus to design noise reduction algorithms in which the tradeoff between the amount of noise reduction and speech distortion can be controlled and quantified, and most current efforts fail to offer any explicit control over, especially, the speech distortion.

Many different methods for noise reduction have been proposed, including linear filtering methods [1], spectral subtraction methods [2], statistical methods [3]–[5], and subspace methods [6], [7]. Most recent research in the field has focused its attention to noise power spectral density estimation [8], [10], [11], [40], extensions to multiple channels [12]–[15], and various improvements to linear filtering techniques, e.g., [1], [16]–[18], [20], while very little progress have been made on subspace methods. We refer the interested reader to [1], [19], [20] for overviews of recent advances in noise reduction. On the matter of subspace methods and their application to noise reduction, we refer to [21] and the references therein.

Of the aforementioned methods, there are two that are particularly relevant to the present paper, namely the linear filtering methods and the subspace methods. In the methods based on linear filtering, noise reduction is obtained by convolution of the observed signal, which comprises both the signal of interest and the additive noise, with the impulse response of a filter. The noise reduction problem then amounts to designing this filter so that it meets some requirements, in terms of, for example, noise reduction and speech distortion. For example, when the mean-square error (MSE) is used as a performance measure and the filter is optimized so as to minimize the MSE, the classical Wiener filter is obtained. In subspace methods, a diagonalization of the involved correlation matrices is obtained by means of, for example, the Karhunen-Loève transform, the eigenvalue decomposition, or the singular value decomposition, and this is then used as a basis for noise reduction by identifying bases for the speech-plus-noise subspace (also sometimes simply called the signal subspace) and the noise subspace, respectively. In the past couple of decades, there has been some attempts to combine the filtering and subspace-based approach, and a few early examples of this can be found in [22], [23]. Of particular relevance to the present work, is the prior use of joint diagonalization for noise reduction, something that has previously been done in [7] and later in [24] to account for colored noise. In this connection, it should be noted that it only served as a computational tool in [12].

In this paper, we express the (single-channel) noise reduction problem as a linear filtering problem using joint diagonalization of the correlation matrix of the signal of interest and the noise. More specifically, we consider filter

*J. R. Jensen is the corresponding author.

This research was funded by the Villum Foundation and the Danish Council for Independent Research, grant ID: DFF 1337-00084.

J. R. Jensen and M. G. Christensen are with the Audio Analysis Lab, AD:MT, Aalborg University, DK-9000 Aalborg, Denmark e-mail: {jrj, mgc}@create.aau.dk.

J. Benesty is with the INRS-EMT, University of Quebec, 800 de la Gauchetiere Ouest, Suite 6900, Montreal, QC H5A 1K6, Canada e-mail: benesty@emt.inrs.ca and with the Audio Analysis Lab, AD:MT, Aalborg University, Denmark.

designs, wherein the filter coefficients are formed as linear combinations of a desired number of eigenvectors. This way, speech distortion can be traded for more noise reduction by changing the number of eigenvectors. We also proposed noise reduction filters based on the joint diagonalization in [25], [26]. As opposed to the filters proposed herein, these considered only an indirect approach where the noise is estimated first and subtracted from the observation to obtain the enhanced signal. Moreover, we propose a wider range of filter designs herein and clearly show how all the different filter designs are related. In the proposed framework, a number of noise reduction filters, which are referred to as variable span filters, are derived. Our naming of the filters stems from the fact that the filters are linear combinations of a varying number of eigenvectors, which therefore span a varying space. The derived filters include minimum distortion, Wiener, and tradeoff filters. Interestingly, it is also possible to express various well-known filters in this framework. We acknowledge that a link between linear filtering and subspace-based enhancement has been established before [27], [28]. However, in these papers, the relationship between traditional filtering methods for enhancement such as Wiener filtering and subspace-based methods has not been identified explicitly. Moreover, these papers regard the enhancement problem as a subspace fitting problem, but this gives no explicit control over relevant performance measures. Both of these issues are addressed with the proposed variable span filtering framework. We also introduce a so-called indirect approach, wherein the noise reduction is performed in two stages. In the first stage, an estimate of the noise signal is found using the variable span filter framework, which is then used, in the second stage, to obtain an estimate of the desired signal by subtracting the estimate from the observed signal. Again, several different filter designs are proposed. We show that this alternative way of formulating the enhancement problem resembles the direct approach in a few special cases, but, generally, it leads to other filter designs which gives more flexibility when tackling the enhancement problem.

The rest of the paper is organized as described here: In the following section, namely Section II, the signal model and problem formulation used throughout the paper are presented. Moreover, in this section, the joint diagonalization is presented along with some useful notation. In Section III, we then introduce the notion of variable span linear filtering, which forms the basis of the proposed filters designs. This is followed by, in Section IV, the performance measures used to quantify the performance of the various filter designs, namely noise reduction, speech distortion, and mean-square error measures. In Section V, some optimal filters that belong to the class of variable span filters are derived. Then, in Section VI, an alternative approach is investigated, wherein the noise reduction is performed in two steps: a first step, wherein the variable span filters are used to form an estimate of the noise, and a second step, where this estimate is then subtracted from the observed signal. Finally, we present some experimental results in Section VII and conclude on the work in Section VIII.

II. Signal Model and Problem Formulation

We consider the very general signal model:

$$\mathbf{y} = \mathbf{x} + \mathbf{v}, \tag{1}$$

where $\mathbf{y}$ is the observation or noisy signal vector of length M, $\mathbf{x}$ is the speech signal vector, and $\mathbf{v}$ is the noise signal vector. This general model can be applied in different domains, e.g., for both time- and frequency-domain processing. We assume that the components of the two vectors $\mathbf{x}$ and $\mathbf{v}$ are zero mean and stationary. Moreover, in cases with complex signals, i.e., in frequency-domain processing, the vector components are assumed to be circular [29], [30]. We further assume that these two vectors are uncorrelated, i.e., $E\left(\mathbf{x}\mathbf{v}^H\right) = E\left(\mathbf{v}\mathbf{x}^H\right) = \mathbf{0}_{M\times M}$, where $E(\cdot)$ denotes mathematical expectation, the superscript ^{H} is the conjugate-transpose operator, and $\mathbf{0}_{M\times M}$ is a matrix of size $M \times M$ with all its elements equal to 0. In this context, the correlation matrix (of size $M \times M$) of the observations is

$$\begin{aligned} \mathbf{\Phi}_{\mathbf{y}} &= E\left(\mathbf{y}\mathbf{y}^H\right) \\ &= \mathbf{\Phi}_{\mathbf{x}} + \mathbf{\Phi}_{\mathbf{v}}, \end{aligned} \tag{2}$$

where $\mathbf{\Phi}_{\mathbf{x}} = E\left(\mathbf{x}\mathbf{x}^H\right)$ and $\mathbf{\Phi}_{\mathbf{v}} = E\left(\mathbf{v}\mathbf{v}^H\right)$ are the correlation matrices of $\mathbf{x}$ and $\mathbf{v}$, respectively. In the rest of this paper, we assume that the rank of the speech correlation matrix, $\mathbf{\Phi}_{\mathbf{x}}$, is equal to $P \leq M$ and the rank of the noise correlation matrix, $\mathbf{\Phi}_{\mathbf{v}}$, is equal to M. It should be emphasized, though, that the filters presented later can still be implemented in practice even if these assumptions do not hold. If the rank of $\mathbf{\Phi}_{\mathbf{x}}$ equals M, the proposed filters can still be implemented if speech distortion is allowed. Moreover, if $\mathbf{\Phi}_{\mathbf{v}}$ is rank deficient, although this is rather unrealistic due to inevitable microphone self noise, we can always add a small degree of matrix regularization to ensure invertibility.

Let x_1 be the first element of $\mathbf{x}$. It is assumed that x_1 is the desired speech signal. Then, we here consider the objective of speech enhancement (or noise reduction) as that of estimating x_1 from $\mathbf{y}$. This should be done in such a way that the noise is reduced as much as possible with no or little distortion of the desired signal sample [20], [19], [31].

The use of the joint diagonalization in noise reduction was first proposed in [7] and then in [24]. In this paper, we give a different perspective, as will be shown later.

The two Hermitian matrices $\mathbf{\Phi}_{\mathbf{x}}$ and $\mathbf{\Phi}_{\mathbf{v}}$ can be jointly diagonalized as follows [32]:

$$\mathbf{B}^H\mathbf{\Phi}_{\mathbf{x}}\mathbf{B} = \mathbf{\Lambda}, \tag{3}$$

$$\mathbf{B}^H\mathbf{\Phi}_{\mathbf{v}}\mathbf{B} = \mathbf{I}_M, \tag{4}$$

where $\mathbf{B}$ is a full-rank square matrix (of size $M \times M$), $\mathbf{\Lambda}$ is a diagonal matrix whose main elements are real and nonnegative, and $\mathbf{I}_M$ is the $M\times M$ identity matrix. Furthermore, $\mathbf{\Lambda}$ and $\mathbf{B}$ are the eigenvalue and eigenvector matrices, respectively, of $\mathbf{\Phi}_{\mathbf{v}}^{-1}\mathbf{\Phi}_{\mathbf{x}}$, i.e.,

$$\mathbf{\Phi}_{\mathbf{v}}^{-1}\mathbf{\Phi}_{\mathbf{x}}\mathbf{B} = \mathbf{B}\mathbf{\Lambda}. \tag{5}$$

Since the rank of the matrix $\mathbf{\Phi}_{\mathbf{x}}$ is equal to P, the eigenvalues of $\mathbf{\Phi}_{\mathbf{v}}^{-1}\mathbf{\Phi}_{\mathbf{x}}$ can be ordered as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_P > \lambda_{P+1} = \cdots = \lambda_M = 0$. In other words, the last $M - P$ eigenvalues

of the matrix product $\mathbf{\Phi}_{\mathbf{v}}^{-1}\mathbf{\Phi}_{\mathbf{x}}$ are exactly zero, while its first P eigenvalues are positive, with λ_1 being the maximum eigenvalue. We also denote by $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_P, \mathbf{b}_{P+1}, \ldots, \mathbf{b}_M$, the corresponding eigenvectors. A consequence of this joint diagonalization is that the noisy signal correlation matrix can also be diagonalized as

$$\mathbf{B}^H\mathbf{\Phi}_{\mathbf{y}}\mathbf{B} = \mathbf{\Lambda} + \mathbf{I}_M. \tag{6}$$

We can decompose the matrix $\mathbf{B}$ as

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}'_Q & \mathbf{B}''_Q \end{bmatrix}, \tag{7}$$

where

$$\mathbf{B}'_Q = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_Q \end{bmatrix} \tag{8}$$

is an $M \times Q$ matrix,

$$\mathbf{B}''_Q = \begin{bmatrix} \mathbf{b}_{Q+1} & \mathbf{b}_{Q+2} & \cdots & \mathbf{b}_M \end{bmatrix} \tag{9}$$

is an $M \times (M-Q)$ matrix, and $1 \leq Q \leq M$. For the particular case $Q = P$, the matrices $\mathbf{B}'_P$ and $\mathbf{B}''_P$ span the speech-plus-noise subspace and the noise subspace, respectively. It can be verified from (3) that

$$\mathbf{B}''^H_P\mathbf{x} = \mathbf{0}_{P\times 1}. \tag{10}$$

To show this, we first deduce from (3) that

$$\mathbf{B}''^H_P\mathbf{\Phi}_{\mathbf{x}}\mathbf{B}''_P = \mathbf{0}. \tag{11}$$

If $\mathbf{b}''_P$ is any column vector of $\mathbf{B}''_P$, we can define a signal s as $s = \mathbf{b}''^H_P\mathbf{x}$. The signal variance is $\phi_s = E\left(|s|^2\right) = E\left(|\mathbf{b}''^H_P\mathbf{x}|^2\right)$. From (11), we have

$$\mathbf{b}''^H_P\mathbf{\Phi}_{\mathbf{x}}\mathbf{b}''_P = E(|\mathbf{b}''^H_P\mathbf{x}|^2) = \phi_s = 0. \tag{12}$$

The above equation can only be 0 if $\mathbf{b}''^H_P\mathbf{x} = 0$ which proves the observation in (10). Moreover, it can be verified from (4) that

$$\mathbf{\Phi}_{\mathbf{v}}^{-1} = \mathbf{B}'_P\mathbf{B}'^H_P + \mathbf{B}''_P\mathbf{B}''^H_P. \tag{13}$$

The joint diagonalization is a very natural tool to use if we want to fully exploit the speech-plus-noise and noise subspaces in noise reduction and fully optimize the linear filtering process.

III. Variable Span Linear Filtering

One of the most convenient ways to estimate the desired signal, x_1, from the observation signal vector, $\mathbf{y}$, is through a filtering operation, i.e.,

$$z = \mathbf{h}^H\mathbf{y}, \tag{14}$$

where z is the estimate of x_1 and

$$\mathbf{h} = \begin{bmatrix} h_1 & h_2 & \cdots & h_M \end{bmatrix}^T \tag{15}$$

is a complex-valued filter of length M. While the considered filtering operation is only estimating x_1, the performance measures and filter designs derived in the following sections can easily be generalized for estimation of x_m for $m = 2, \ldots, M$. Alternatively, in time-domain processing, for example, we only need the filter estimating x_1 since we can just be update it at every time instance to obtain the most recent estimate of the desired signal. It is always possible to write $\mathbf{h}$ in a basis formed from the vectors $\mathbf{b}_m,\ m = 1, 2, \ldots, M$, i.e.,

$$\begin{aligned} \mathbf{h} &= \mathbf{B}\mathbf{a} \\ &= \mathbf{B}'_Q\mathbf{a}'_Q + \mathbf{B}''_Q\mathbf{a}''_Q, \end{aligned} \tag{16}$$

where the components of

$$\begin{aligned} \mathbf{a} &= \begin{bmatrix} a_1 & \cdots & a_Q & a_{Q+1} & \cdots & a_M \end{bmatrix}^T \\ &= \begin{bmatrix} \mathbf{a}'^T_Q & \mathbf{a}''^T_Q \end{bmatrix}^T \end{aligned} \tag{17}$$

are the coordinates of $\mathbf{h}$ in the new basis, and $\mathbf{a}'_Q$ and $\mathbf{a}''_Q$ are vectors of length Q and $M - Q$, respectively. Now, instead of estimating the coefficients of $\mathbf{h}$ as in conventional approaches, we can estimate, equivalently, the coordinates $a_m,\ m = 1, 2, \ldots, M$. When $\mathbf{a}$ is estimated, it is then easy to determine $\mathbf{h}$ from (16). Furthermore, for $Q = P$, several optimal noise reduction filters with at most P constraints will lead to $\mathbf{a}''_P = \mathbf{0}_{(M-P)\times 1}$ since there is no speech in the directions $\mathbf{B}''_P$. Therefore, we can sometimes simplify our problem and force $\mathbf{a}''_P = \mathbf{0}_{(M-P)\times 1}$; as a result, the filter and the estimate are, respectively, $\mathbf{h}' = \mathbf{B}'_P\mathbf{a}'_P$ and $z' = \mathbf{a}'^H_P\mathbf{B}'^H_P\mathbf{y}$.

From the previous discussion and from (16), we see that we can build a more flexible linear filter. We define our variable span filter of length M as

$$\mathbf{h}'(Q) = \mathbf{B}'_Q\mathbf{a}'_Q. \tag{18}$$

Obviously, $\mathbf{h}'(Q) \in \mathrm{Span}\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_Q\}$. As a consequence, the estimate of x_1 is

$$\begin{aligned} z' &= \mathbf{a}'^H_Q\mathbf{B}'^H_Q\mathbf{x} + \mathbf{a}'^H_Q\mathbf{B}'^H_Q\mathbf{v} \\ &= x'_{\mathrm{fd}} + v'_{\mathrm{rn}}, \end{aligned} \tag{19}$$

where $x'_{\mathrm{fd}} = \mathbf{a}'^H_Q\mathbf{B}'^H_Q\mathbf{x}$ is the filtered desired signal and $v'_{\mathrm{rn}} = \mathbf{a}'^H_Q\mathbf{B}'^H_Q\mathbf{v}$ is the residual noise. We deduce that the variance of z' is

$$\phi_{z'} = \mathbf{a}'^H_Q\mathbf{\Lambda}'_Q\mathbf{a}'_Q + \mathbf{a}'^H_Q\mathbf{a}'_Q, \tag{20}$$

where

$$\mathbf{\Lambda}'_Q = \mathrm{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_Q\right) \tag{21}$$

is a diagonal matrix containing the first Q eigenvalues of $\mathbf{\Phi}_{\mathbf{v}}^{-1}\mathbf{\Phi}_{\mathbf{x}}$.

Notice that the proposed linear processing implies implicitly that we force the last $M - Q$ components of $\mathbf{a}$ to 0.

IV. Performance Measures

In this section, we briefly define the most useful performance measures for noise reduction with variable span filters. We can divide these measures into two categories. The first category evaluates the noise reduction performance while the second one evaluates speech distortion. We also discuss the very convenient mean-square error (MSE) criterion, which we tailor for variable span filters, and show how it is related to the performance measures.

A. Noise Reduction

One of the most fundamental measures in all aspects of speech enhancement is the signal-to-noise ratio (SNR). Since x_1 is the desired signal, we define the input SNR as

$$\text{iSNR} = \frac{\phi_{x_1}}{\phi_{v_1}}, \tag{22}$$

where $\phi_{x_1} = E\left(|x_1|^2\right)$ is the variance of x_1 and $\phi_{v_1} = E\left(|v_1|^2\right)$ is the variance of the first component of $\mathbf{v}$, i.e., v_1.

From (20), it is easy to find that the output SNR is

$$\begin{aligned}\text{oSNR}\left(\mathbf{a}'_Q\right) &= \frac{\mathbf{a}'^H_Q \mathbf{\Lambda}'_Q \mathbf{a}'_Q}{\mathbf{a}'^H_Q \mathbf{a}'_Q} \\ &= \frac{\sum_{q=1}^{Q} \lambda_q \left|a_q\right|^2}{\sum_{q=1}^{Q} \left|a_q\right|^2},\end{aligned} \tag{23}$$

and it can be shown that

$$\text{oSNR}\left(\mathbf{a}'_Q\right) \leq \lambda_1, \tag{24}$$

which means that the output SNR can never exceed the maximum eigenvalue, λ_1. The filters should be derived in such a way that $\text{oSNR}\left(\mathbf{a}'_Q\right) \geq \text{iSNR}$.

The noise reduction factor, which quantifies the amount of noise which is rejected by the complex filter, is given by

$$\xi_{\text{nr}}\left(\mathbf{a}'_Q\right) = \frac{\phi_{v_1}}{\mathbf{a}'^H_Q \mathbf{a}'_Q}. \tag{25}$$

For optimal noise reduction filters, we should have $\xi_{\text{nr}}\left(\mathbf{a}'_Q\right) \geq 1$.

B. Speech Distortion

In practice, the complex filter may distort the desired signal. In order to evaluate the level of this distortion, we define the speech reduction factor:

$$\xi_{\text{sr}}\left(\mathbf{a}'_Q\right) = \frac{\phi_{x_1}}{\mathbf{a}'^H_Q \mathbf{\Lambda}'_Q \mathbf{a}'_Q}. \tag{26}$$

For optimal filters, we should have $\xi_{\text{sr}}\left(\mathbf{a}'_Q\right) \geq 1$. The larger the value of $\xi_{\text{sr}}\left(\mathbf{a}'_Q\right)$ is, the more the desired speech signal is distorted.

By making the appropriate substitutions, one can derive the relationship:

$$\frac{\text{oSNR}\left(\mathbf{a}'_Q\right)}{\text{iSNR}} = \frac{\xi_{\text{nr}}\left(\mathbf{a}'_Q\right)}{\xi_{\text{sr}}\left(\mathbf{a}'_Q\right)}. \tag{27}$$

This expression indicates the equivalence between gain/loss in SNR and distortion (for both speech and noise).

Another way to measure the distortion of the desired speech signal due to the filter is the speech distortion index, which is defined as the mean-square error between the desired signal and the filtered desired signal, normalized by the variance of the desired signal, i.e.,

$$\upsilon_{\text{sd}}\left(\mathbf{a}'_Q\right) = \frac{E\left(\left|x_1 - \mathbf{a}'^H_Q \mathbf{B}'^H_Q \mathbf{x}\right|^2\right)}{\phi_{x_1}}. \tag{28}$$

The speech distortion index is usually upper bounded by 1 for optimal filters. Moreover, as opposed to the speech reduction factor in (26), the speech distortion index also takes phase distortion into account. The distortion could, of course, also be measured using other conventional measures such as log spectral distortion and cepstral distance [33].

C. Mean-Square Error (MSE) Criterion

The error signal between the estimated and desired signals is

$$\begin{aligned}e' &= z' - x_1 \\ &= \mathbf{a}'^H_Q \mathbf{B}'^H_Q \mathbf{y} - x_1,\end{aligned} \tag{29}$$

which can also be written as the sum of two uncorrelated error signals:

$$e' = e'_{\text{ds}} + e'_{\text{rs}}, \tag{30}$$

where

$$e'_{\text{ds}} = \mathbf{a}'^H_Q \mathbf{B}'^H_Q \mathbf{x} - x_1 \tag{31}$$

is the speech distortion due to the filter and

$$e'_{\text{rs}} = \mathbf{a}'^H_Q \mathbf{B}'^H_Q \mathbf{v} \tag{32}$$

represents the residual noise. The mean-square error (MSE) criterion is then

$$\begin{aligned}J\left(\mathbf{a}'_Q\right) &= E\left(\left|e'\right|^2\right) \\ &= \phi_{x_1} - \mathbf{i}^T \mathbf{\Phi}_{\mathbf{x}} \mathbf{B}'_Q \mathbf{a}'_Q - \mathbf{a}'^H_Q \mathbf{B}'^H_Q \mathbf{\Phi}_{\mathbf{x}} \mathbf{i} \\ &\quad + \mathbf{a}'^H_Q \left(\mathbf{\Lambda}'_Q + \mathbf{I}_Q\right) \mathbf{a}'_Q \\ &= J_{\text{ds}}\left(\mathbf{a}'_Q\right) + J_{\text{rs}}\left(\mathbf{a}'_Q\right),\end{aligned} \tag{33}$$

where $\mathbf{i}$ is the first column of $\mathbf{I}_M$, $\mathbf{I}_Q$ is the $Q \times Q$ identity matrix, with

$$\begin{aligned}J_{\text{ds}}\left(\mathbf{a}'_Q\right) &= E\left(\left|e'_{\text{ds}}\right|^2\right) \\ &= \upsilon_{\text{sd}}\left(\mathbf{a}'_Q\right) \phi_{x_1},\end{aligned} \tag{34}$$

and

$$\begin{aligned}J_{\text{rs}}\left(\mathbf{a}'_Q\right) &= E\left(\left|e'_{\text{rs}}\right|^2\right) \\ &= \frac{\phi_{v_1}}{\xi_{\text{nr}}\left(\mathbf{a}'_Q\right)}.\end{aligned} \tag{35}$$

We deduce that

$$\begin{aligned}\frac{J_{\text{ds}}\left(\mathbf{a}'_Q\right)}{J_{\text{rs}}\left(\mathbf{a}'_Q\right)} &= \text{iSNR} \times \xi_{\text{nr}}\left(\mathbf{a}'_Q\right) \times \upsilon_{\text{sd}}\left(\mathbf{a}'_Q\right) \\ &= \text{oSNR}\left(\mathbf{a}'_Q\right) \times \xi_{\text{sr}}\left(\mathbf{a}'_Q\right) \times \upsilon_{\text{sd}}\left(\mathbf{a}'_Q\right).\end{aligned} \tag{36}$$

This shows how the different performances measures are related to the MSEs.

V. Optimal Variable Span (VS) Filters

In this section, we derive a large class of variable span (VS) filters for noise reduction from the different MSEs developed in the previous section. We will see how all these filters, with different objectives, are strongly connected.

A. VS Minimum Distortion

The VS minimum distortion filter is obtained by minimizing the distortion-based MSE, $J_{\rm ds}\left(\mathbf{a}'_Q\right)$. We get

$$\mathbf{a}'_{Q,\rm MD} = \mathbf{\Lambda}'^{-1}_Q \mathbf{B}'^H_Q \mathbf{\Phi}_{\mathbf{x}}\mathbf{i}, \tag{37}$$

where $Q \leq P$. Therefore, the VS minimum distortion filter is

$$\begin{aligned}\mathbf{h}'_{\rm MD}(Q) &= \mathbf{B}'_Q \mathbf{a}'_{Q,\rm MD} \\ &= \sum_{q=1}^{Q} \frac{\mathbf{b}_q \mathbf{b}_q^H}{\lambda_q} \mathbf{\Phi}_{\mathbf{x}}\mathbf{i}. \end{aligned}\tag{38}$$

We note that the idea of variable rank subspace filters have been considered before [28], but previous approaches have been derived from a subspace fitting perspective rather than an enhancement perspective as mentioned in the introduction. Therefore, the previous approaches do not provide any explicit control over or analysis of the amounts of noise reduction and signal distortion.

One important particular case of (38) is $Q = P$. In this situation, we obtain the celebrated minimum variance distortionless response (MVDR) filter:

$$\begin{aligned}\mathbf{h}'_{\rm MVDR} &= \mathbf{h}'_{\rm MD}(P) = \mathbf{B}'_P \mathbf{a}'_{P,\rm MD} \\ &= \sum_{p=1}^{P} \frac{\mathbf{b}_p \mathbf{b}_p^H}{\lambda_p} \mathbf{\Phi}_{\mathbf{x}}\mathbf{i} \\ &= \sum_{p=1}^{P} \mathbf{b}_p \mathbf{b}_p^H \mathbf{\Phi}_{\mathbf{v}}\mathbf{i}. \end{aligned}\tag{39}$$

Let us show why (39) corresponds to the MVDR filter. With $\mathbf{h}'_{\rm MVDR}$, the filtered desired signal is

$$\begin{aligned}x'_{\rm fd} &= \left(\mathbf{B}'_P \mathbf{B}'^H_P \mathbf{\Phi}_{\mathbf{v}}\mathbf{i}\right)^H \mathbf{x} \\ &= \left(\mathbf{i} - \mathbf{B}''_P \mathbf{B}''^H_P \mathbf{\Phi}_{\mathbf{v}}\mathbf{i}\right)^H \mathbf{x} \\ &= x_1 - \mathbf{i}^T \mathbf{\Phi}_{\mathbf{v}} \mathbf{B}''_P \mathbf{B}''^H_P \mathbf{x} \\ &= x_1, \end{aligned}\tag{40}$$

where we have used (10) and (13) in the previous expression. Then, it is clear that

$$\upsilon_{\rm sd}\left(\mathbf{a}'_{P,\rm MD}\right) = 0, \tag{41}$$

proving that, indeed, $\mathbf{h}'_{\rm MVDR}$ is the MVDR filter.

Another interesting case of (38) is $Q = 1$. In this scenario, we obtain the maximum SNR filter:

$$\begin{aligned}\mathbf{h}'_{\rm max,0} &= \mathbf{h}'_{\rm MD}(1) = \mathbf{b}_1 a_{1,\rm MD} \\ &= \frac{\mathbf{b}_1 \mathbf{b}_1^H}{\lambda_1} \mathbf{\Phi}_{\mathbf{x}}\mathbf{i}. \end{aligned}\tag{42}$$

Indeed, it can be verified that

$$\text{oSNR}\left(a_{1,\rm MD}\right) = \lambda_1. \tag{43}$$

We should always have

$$\begin{aligned}\text{oSNR}\left(\mathbf{a}'_{P,\rm MD}\right) \leq \text{oSNR}\left(\mathbf{a}'_{P-1,\rm MD}\right) \leq \cdots \\ \leq \text{oSNR}\left(a_{1,\rm MD}\right) = \lambda_1 \end{aligned}\tag{44}$$

and

$$\upsilon_{\rm sd}\left(\mathbf{a}'_{P,\rm MD}\right) \leq \upsilon_{\rm sd}\left(\mathbf{a}'_{P-1,\rm MD}\right) \leq \cdots \leq \upsilon_{\rm sd}\left(a_{1,\rm MD}\right) \leq 1. \tag{45}$$

If $\mathbf{\Phi}_{\mathbf{x}}$ is a full-rank matrix, i.e., $P = M$, then

$$\mathbf{h}'_{\rm MD}(M) = \mathbf{i}, \tag{46}$$

which is the identity filter. Assume that the rank of $\mathbf{\Phi}_{\mathbf{x}}$ is $P = 1$. In this case, $Q = 1$, the filter is $\mathbf{h}'_{\rm MD}(1) = \mathbf{h}'_{\rm MVDR}$, and the output SNR is maximized, i.e., equal to λ_1. Also, we can write the speech correlation matrix as

$$\mathbf{\Phi}_{\mathbf{x}} = \phi_{x_1} \mathbf{d}\mathbf{d}^H, \tag{47}$$

where $\mathbf{d}$ is a vector of length M, whose first element is equal to 1. As a consequence,

$$\begin{aligned}\lambda_1 &= \phi_{x_1} \mathbf{d}^H \mathbf{\Phi}_{\mathbf{v}}^{-1} \mathbf{d} \\ &= \phi_{x_1} \left|\mathbf{d}^H \mathbf{b}_1\right|^2 \end{aligned}\tag{48}$$

and

$$\mathbf{h}'_{\rm MVDR} = \frac{\mathbf{\Phi}_{\mathbf{v}}^{-1}\mathbf{d}}{\mathbf{d}^H \mathbf{\Phi}_{\mathbf{v}}^{-1}\mathbf{d}}. \tag{49}$$

B. VS Wiener

The VS Wiener filter is obtained from the optimization of the MSE criterion, $J\left(\mathbf{a}'_Q\right)$. The minimization of $J\left(\mathbf{a}'_Q\right)$ leads to

$$\mathbf{a}'_{Q,\rm W} = \left(\mathbf{\Lambda}'_Q + \mathbf{I}_Q\right)^{-1} \mathbf{B}'^H_Q \mathbf{\Phi}_{\mathbf{x}}\mathbf{i}, \tag{50}$$

where $Q \leq M$. We deduce that the VS Wiener filter is

$$\begin{aligned}\mathbf{h}'_{\rm W}(Q) &= \mathbf{B}'_Q \mathbf{a}'_{Q,\rm W} \\ &= \sum_{q=1}^{Q} \frac{\mathbf{b}_q \mathbf{b}_q^H}{1+\lambda_q} \mathbf{\Phi}_{\mathbf{x}}\mathbf{i}. \end{aligned}\tag{51}$$

It is interesting to compare $\mathbf{h}'_{\rm W}(Q)$ to $\mathbf{h}'_{\rm MD}(Q)$. The two VS filters are very close to each other; they differ by the weighting function, which strongly depends on the eigenvalues of the joint diagonalization. For the VS Wiener filter, this function is equal to $(1+\lambda_q)^{-1}$ while it is equal to λ_q^{-1} for the VS minimum distortion filter. Also, in the latter filter, Q must be smaller than or equal to P, while Q can be greater than P in the former one.

One important particular case of (51) is $Q = M$. In this situation, we obtain the classical Wiener filter:

$$\begin{aligned}\mathbf{h}'_{\rm W} &= \mathbf{h}'_{\rm W}(M) = \mathbf{B}'_M \mathbf{a}'_{M,\rm W} \\ &= \sum_{m=1}^{M} \frac{\mathbf{b}_m \mathbf{b}_m^H}{1+\lambda_m} \mathbf{\Phi}_{\mathbf{x}}\mathbf{i} \\ &= \mathbf{\Phi}_{\mathbf{y}}^{-1} \mathbf{\Phi}_{\mathbf{x}}\mathbf{i}. \end{aligned}\tag{52}$$

For $Q = 1$, we obtain another form of the maximum SNR filter:

$$\mathbf{h}'_{\max,1} = \mathbf{h}'_{\mathrm{W}}(1) = \mathbf{b}_1 a_{1,\mathrm{W}} = \frac{\mathbf{b}_1\mathbf{b}_1^H}{1+\lambda_1}\mathbf{\Phi}_{\mathbf{x}}\mathbf{i}, \tag{53}$$

since

$$\mathrm{oSNR}\left(a_{1,\mathrm{W}}\right) = \lambda_1. \tag{54}$$

We should always have

$$\begin{aligned}\mathrm{oSNR}\left(\mathbf{a}'_{M,\mathrm{W}}\right) \leq \mathrm{oSNR}\left(\mathbf{a}'_{M-1,\mathrm{W}}\right) \leq \cdots \\ \leq \mathrm{oSNR}\left(a_{1,\mathrm{W}}\right) = \lambda_1\end{aligned} \tag{55}$$

and

$$\upsilon_{\mathrm{sd}}\left(\mathbf{a}'_{M,\mathrm{W}}\right) \leq \upsilon_{\mathrm{sd}}\left(\mathbf{a}'_{M-1,\mathrm{W}}\right) \leq \cdots \leq \upsilon_{\mathrm{sd}}\left(a_{1,\mathrm{W}}\right) \leq 1. \tag{56}$$

C. VS Tradeoff

Another interesting approach that can compromise between noise reduction and speech distortion is the VS tradeoff filter obtained by

$$\min_{\mathbf{a}'_Q} J_{\mathrm{ds}}\left(\mathbf{a}'_Q\right) \quad \text{subject to} \quad J_{\mathrm{rs}}\left(\mathbf{a}'_Q\right) = \beta'\phi_{v_1}, \tag{57}$$

where $0 \leq \beta' \leq 1$, to ensure that filtering achieves some degree of noise reduction. We easily find that the optimal filter is

$$\mathbf{h}'_{\mathrm{T},\mu'}(Q) = \sum_{q=1}^{Q}\frac{\mathbf{b}_q\mathbf{b}_q^H}{\mu'+\lambda_q}\mathbf{\Phi}_{\mathbf{x}}\mathbf{i}, \tag{58}$$

where $\mu' \geq 0$ is a Lagrange multiplier[1]. By changing Q and μ', we are able to choose between a wide range of noise reduction and signal distortion trade offs. Clearly, for $\mu' = 0$ and $\mu' = 1$, we get the VS minimum distortion and VS Wiener filters, respectively.

For $Q = M$, we obtain the classical tradeoff filter:

$$\begin{aligned}\mathbf{h}'_{\mathrm{T},\mu'} &= \mathbf{h}'_{\mathrm{T},\mu'}(M) \\ &= \sum_{m=1}^{M}\frac{\mathbf{b}_m\mathbf{b}_m^H}{\mu'+\lambda_m}\mathbf{\Phi}_{\mathbf{x}}\mathbf{i} \\ &= \left(\mathbf{\Phi}_{\mathbf{x}}+\mu'\mathbf{\Phi}_{\mathbf{v}}\right)^{-1}\mathbf{\Phi}_{\mathbf{x}}\mathbf{i}\end{aligned} \tag{59}$$

and for $Q = 1$, we obtain the maximum SNR filter:

$$\mathbf{h}'_{\max,\mu'} = \mathbf{h}'_{\mathrm{T},\mu'}(1) = \frac{\mathbf{b}_1\mathbf{b}_1^H}{\mu'+\lambda_1}\mathbf{\Phi}_{\mathbf{x}}\mathbf{i}. \tag{60}$$

This shows us that all these traditional enhancement filter types are clearly related. In Table I, we summarize all optimal VS filters developed in this section, showing how they are strongly related in the proposed framework. As we can see, these filter designs are simple to compute, so the main computational complexity of the proposed method lies in the computation of the joint diagonalization. However, fast and recursive implementations exist, which can reduce its complexity from order $\mathcal{O}(M^3)$ to order $\mathcal{O}(M^2)$ [34].

[1]For $\mu' = 0$, Q must be smaller than or equal to P.

TABLE I
OPTIMAL VS FILTERS FOR NOISE REDUCTION.

VS Minimum Distortion:	$\mathbf{h}'_{\mathrm{MD}}(Q) = \sum_{q=1}^{Q}\frac{\mathbf{b}_q\mathbf{b}_q^H}{\lambda_q}\mathbf{\Phi}_{\mathbf{x}}\mathbf{i},\ Q \leq P$
MVDR:	$\mathbf{h}'_{\mathrm{MVDR}} = \sum_{p=1}^{P}\frac{\mathbf{b}_p\mathbf{b}_p^H}{\lambda_p}\mathbf{\Phi}_{\mathbf{x}}\mathbf{i}$
VS Wiener:	$\mathbf{h}'_{\mathrm{W}}(Q) = \sum_{q=1}^{Q}\frac{\mathbf{b}_q\mathbf{b}_q^H}{1+\lambda_q}\mathbf{\Phi}_{\mathbf{x}}\mathbf{i},\ Q \leq M$
Wiener:	$\mathbf{h}'_{\mathrm{W}} = \sum_{m=1}^{M}\frac{\mathbf{b}_m\mathbf{b}_m^H}{1+\lambda_m}\mathbf{\Phi}_{\mathbf{x}}\mathbf{i}$
VS Tradeoff:	$\mathbf{h}'_{\mathrm{T},\mu'}(Q) = \sum_{q=1}^{Q}\frac{\mathbf{b}_q\mathbf{b}_q^H}{\mu'+\lambda_q}\mathbf{\Phi}_{\mathbf{x}}\mathbf{i},\ \mu' \geq 0$
Tradeoff:	$\mathbf{h}'_{\mathrm{T},\mu'} = \sum_{m=1}^{M}\frac{\mathbf{b}_m\mathbf{b}_m^H}{\mu'+\lambda_m}\mathbf{\Phi}_{\mathbf{x}}\mathbf{i}$
Maximum SNR:	$\mathbf{h}'_{\max,\mu'} = \frac{\mathbf{b}_1\mathbf{b}_1^H}{\mu'+\lambda_1}\mathbf{\Phi}_{\mathbf{x}}\mathbf{i}$

VI. INDIRECT OPTIMAL VARIABLE SPAN (VS) FILTERS

The indirect approach is based on two successive stages. In the first stage, we find an estimate of the noise signal. This estimate is then used in the second stage by subtracting it from the observation signal. This will lead to an estimate of the desired signal. Interestingly, as it appears later, there is a strong link between the indirect and direct filters, i.e., the optimal variable span filters can be derived and understood in different ways. This also means that the same filter can be implemented in many different ways, which may be relevant from a numerical point of view. Moreover, we show that the direct and indirect filter types are only equivalent in some special cases, which means we can obtain different filters with the indirect approach. This gives us even more flexibility for solving the enhancement problem.

A. Indirect Approach

Let $\mathbf{h}'' = \begin{bmatrix} h''_1 & h''_2 & \cdots & h''_M \end{bmatrix}^T$ be a complex-valued filter of length M. By applying this filter to the observation signal vector, we obtain

$$\widehat{v} = \mathbf{h}''^H\mathbf{x} + \mathbf{h}''^H\mathbf{v} \tag{61}$$

and the corresponding output SNR is

$$\mathrm{oSNR}_{\widehat{v}}\left(\mathbf{h}''\right) = \frac{\mathbf{h}''^H\mathbf{\Phi}_{\mathbf{x}}\mathbf{h}''}{\mathbf{h}''^H\mathbf{\Phi}_{\mathbf{v}}\mathbf{h}''}. \tag{62}$$

Then, we find $\mathbf{h}''$ that minimizes $\mathrm{oSNR}_{\widehat{v}}\left(\mathbf{h}''\right)$. It is easy to check that the solution is

$$\mathbf{h}'' = \mathbf{B}''_P\mathbf{a}''_P, \tag{63}$$

where $\mathbf{B}''_P$ and $\mathbf{a}''_P$ are defined in the previous sections. With (63), $\mathrm{oSNR}_{\widehat{v}}\left(\mathbf{h}''\right) = 0$ and $\widehat{v}$ can be seen as the estimate of the noise.

We consider the more general scenario:

$$\mathbf{h}''(Q) = \mathbf{B}''_Q\mathbf{a}''_Q, \tag{64}$$

where $0 \leq Q < M$, $\mathbf{B}''_Q = \begin{bmatrix} \mathbf{b}_{Q+1} & \mathbf{b}_{Q+2} & \cdots & \mathbf{b}_M \end{bmatrix}$ is a matrix of size $M \times (M-Q)$, $\mathbf{a}''_Q = \begin{bmatrix} a_{Q+1} & a_{Q+2} & \cdots & a_M \end{bmatrix}^T$ is a vector of length $M-Q$, and $\mathbf{h}''(Q) \in \text{Span}\{\mathbf{b}_{Q+1}, \mathbf{b}_{Q+2}, \ldots, \mathbf{b}_M\}$. As a consequence,

$$\widehat{v} = \mathbf{h}''^H(Q)\mathbf{x} + \mathbf{h}''^H(Q)\mathbf{v}. \tag{65}$$

Now, in general, $\text{oSNR}_{\widehat{v}}[\mathbf{h}''(Q)] \neq 0$, and this implies distortion as it will become clearer soon. This concludes the first stage.

In the second stage, we estimate the desired signal, x_1, as the difference between the observation, y_1, and the estimate of the noise obtained from the first stage, i.e.,

$$\begin{aligned} z'' &= y_1 - \mathbf{h}''^H(Q)\mathbf{x} - \mathbf{h}''^H(Q)\mathbf{v} \\ &= x_1 - \mathbf{a}''^H_Q\mathbf{B}''^H_Q\mathbf{x} + v_1 - \mathbf{a}''^H_Q\mathbf{B}''^H_Q\mathbf{v} \\ &= \overline{\mathbf{h}}''^H(Q)\mathbf{y}, \end{aligned} \tag{66}$$

where

$$\overline{\mathbf{h}}''(Q) = \mathbf{i} - \mathbf{B}''_Q\mathbf{a}''_Q \tag{67}$$

is the equivalent filter applied to the observation signal vector.

B. MSE Criterion and Performance Measures

We define the error signal between the estimated and desired signals as

$$e'' = z'' - x_1 = -\mathbf{a}''^H_Q\mathbf{B}''^H_Q\mathbf{x} + v_1 - \mathbf{a}''^H_Q\mathbf{B}''^H_Q\mathbf{v}. \tag{68}$$

This error can be written as the sum of two uncorrelated error signals, i.e.,

$$e'' = e''_{\text{ds}} + e''_{\text{rs}}, \tag{69}$$

where

$$e''_{\text{ds}} = -\mathbf{a}''^H_Q\mathbf{B}''^H_Q\mathbf{x}, \tag{70}$$

$$e''_{\text{rs}} = v_1 - \mathbf{a}''^H_Q\mathbf{B}''^H_Q\mathbf{v} \tag{71}$$

are the speech distortion due to the filter and the residual noise, respectively. Then, the MSE criterion is

$$\begin{aligned} J\left(\mathbf{a}''_Q\right) &= E\left(|e''|^2\right) \\ &= \phi_{v_1} - \mathbf{i}^T\mathbf{\Phi}_{\mathbf{v}}\mathbf{B}''_Q\mathbf{a}''_Q - \mathbf{a}''^H_Q\mathbf{B}''^H_Q\mathbf{\Phi}_{\mathbf{v}}\mathbf{i} \\ &\quad + \mathbf{a}''^H_Q\left(\mathbf{\Lambda}''_Q + \mathbf{I}_{M-Q}\right)\mathbf{a}''_Q \\ &= J_{\text{ds}}\left(\mathbf{a}''_Q\right) + J_{\text{rs}}\left(\mathbf{a}''_Q\right), \end{aligned} \tag{72}$$

where $\mathbf{I}_{M-Q}$ is the $(M-Q)\times(M-Q)$ identity matrix,

$$\mathbf{\Lambda}''_Q = \text{diag}\left(\lambda_{Q+1}, \lambda_{Q+2}, \ldots, \lambda_M\right) \tag{73}$$

is a diagonal matrix containing the last $M-Q$ eigenvalues of $\mathbf{\Phi}_{\mathbf{v}}^{-1}\mathbf{\Phi}_{\mathbf{x}}$,

$$J_{\text{ds}}\left(\mathbf{a}''_Q\right) = E\left(|e''_{\text{ds}}|^2\right) = \upsilon_{\text{sd}}\left(\mathbf{a}''_Q\right)\phi_{x_1} \tag{74}$$

is the distortion-based MSE,

$$\upsilon_{\text{sd}}\left(\mathbf{a}''_Q\right) = \frac{\mathbf{a}''^H_Q\mathbf{\Lambda}''_Q\mathbf{a}''_Q}{\phi_{x_1}} \tag{75}$$

is the speech distortion index,

$$J_{\text{rs}}\left(\mathbf{a}''_Q\right) = E\left(|e''_{\text{rs}}|^2\right) = \frac{\phi_{v_1}}{\xi_{\text{nr}}\left(\mathbf{a}''_Q\right)} \tag{76}$$

is the MSE corresponding to the residual noise, and

$$\xi_{\text{nr}}\left(\mathbf{a}''_Q\right) = \frac{\phi_{v_1}}{\phi_{v_1} - 2\text{Re}\{\mathbf{a}''^H_Q\mathbf{B}''^H_Q\mathbf{\Phi}_{\mathbf{v}}\mathbf{i}\} + \mathbf{a}''^H_Q\mathbf{a}''_Q} \tag{77}$$

is the noise reduction factor, with $\text{Re}\{\cdot\}$ denoting the real part of a complex number. We deduce that

$$\begin{aligned} \frac{J_{\text{ds}}\left(\mathbf{a}''_Q\right)}{J_{\text{rs}}\left(\mathbf{a}''_Q\right)} &= \text{iSNR} \times \xi_{\text{nr}}\left(\mathbf{a}''_Q\right) \times \upsilon_{\text{sd}}\left(\mathbf{a}''_Q\right) \\ &= \text{oSNR}\left(\mathbf{a}''_Q\right) \times \xi_{\text{sr}}\left(\mathbf{a}''_Q\right) \times \upsilon_{\text{sd}}\left(\mathbf{a}''_Q\right), \end{aligned} \tag{78}$$

where

$$\text{oSNR}\left(\mathbf{a}''_Q\right) = \frac{\phi_{x_1} - 2\text{Re}\{\mathbf{a}''^H_Q\mathbf{B}''^H_Q\mathbf{\Phi}_{\mathbf{x}}\mathbf{i}\} + \mathbf{a}''^H_Q\mathbf{\Lambda}''_Q\mathbf{a}''_Q}{\phi_{v_1} - 2\text{Re}\{\mathbf{a}''^H_Q\mathbf{B}''^H_Q\mathbf{\Phi}_{\mathbf{v}}\mathbf{i}\} + \mathbf{a}''^H_Q\mathbf{a}''_Q} \tag{79}$$

is the output SNR, and

$$\xi_{\text{sr}}\left(\mathbf{a}''_Q\right) = \frac{\phi_{x_1}}{\phi_{x_1} - 2\text{Re}\{\mathbf{a}''^H_Q\mathbf{B}''^H_Q\mathbf{\Phi}_{\mathbf{x}}\mathbf{i}\} + \mathbf{a}''^H_Q\mathbf{\Lambda}''_Q\mathbf{a}''_Q} \tag{80}$$

is the speech reduction factor.

C. Optimal Filters

1) Indirect VS Minimum Residual Noise: The indirect VS minimum residual noise filter is derived from $J_{\text{rs}}\left(\mathbf{a}''_Q\right)$. Indeed, by minimizing $J_{\text{rs}}\left(\mathbf{a}''_Q\right)$, we easily get

$$\mathbf{a}''_{Q,\text{MR}} = \mathbf{B}''^H_Q\mathbf{\Phi}_{\mathbf{v}}\mathbf{i}. \tag{81}$$

Therefore, the indirect VS minimum residual noise filter is

$$\overline{\mathbf{h}}''_{\text{MR}}(Q) = \mathbf{i} - \mathbf{B}''_Q\mathbf{B}''^H_Q\mathbf{\Phi}_{\mathbf{v}}\mathbf{i} = \mathbf{B}'_Q\mathbf{B}'^H_Q\mathbf{\Phi}_{\mathbf{v}}\mathbf{i} \tag{82}$$

for $Q \geq 1$ and $\overline{\mathbf{h}}''_{\text{MR}}(0) = \mathbf{0}_{M\times 1}$. We can express the previous filter as

$$\begin{aligned} \overline{\mathbf{h}}''_{\text{MR}}(Q) &= \sum_{q=1}^{Q}\mathbf{b}_q\mathbf{b}_q^H\mathbf{\Phi}_{\mathbf{v}}\mathbf{i} \\ &= \sum_{q=1}^{Q}\frac{\mathbf{b}_q\mathbf{b}_q^H}{\lambda_q}\mathbf{\Phi}_{\mathbf{x}}\mathbf{i} + \sum_{i=Q+1}^{Q}\mathbf{b}_i\mathbf{b}_i^H\mathbf{\Phi}_{\mathbf{v}}\mathbf{i} \\ &= \mathbf{h}'_{\text{MD}}(Q) + \sum_{i=Q+1}^{Q}\mathbf{b}_i\mathbf{b}_i^H\mathbf{\Phi}_{\mathbf{v}}\mathbf{i} \\ &= \overline{\mathbf{h}}''_{\text{MR}}(Q) + \sum_{i=Q+1}^{Q}\mathbf{b}_i\mathbf{b}_i^H\mathbf{\Phi}_{\mathbf{v}}\mathbf{i} \end{aligned} \tag{83}$$

for $Q < M$ [and $\overline{\mathbf{h}}''_{\text{MR}}(M) = \mathbf{i}$]. We observe that for $Q \leq Q \leq P$, $\overline{\mathbf{h}}''_{\text{MR}}(Q) = \mathbf{h}'_{\text{MD}}(Q)$. But for $Q > P$, the two filters are different since $\mathbf{h}'_{\text{MD}}(Q)$ is not defined in this context.

We have at least two interesting particular cases:

- $\overline{\mathbf{h}}''_{\text{MR}}(1) = \mathbf{h}'_{\text{max},0}$, which corresponds to the maximum SNR filter; and
- $\overline{\mathbf{h}}''_{\text{MR}}(P) = \mathbf{h}'_{\text{MVDR}}$, which corresponds to the MVDR filter.

2) Indirect VS Wiener: The indirect VS Wiener filter is obtained from the optimization of the MSE criterion, $J\left(\mathbf{a}_Q''\right)$. The minimization of $J\left(\mathbf{a}_Q''\right)$ leads to

$$\mathbf{a}_{Q,\mathrm{W}}'' = \left(\mathbf{\Lambda}_Q'' + \mathbf{I}_{M-Q}\right)^{-1} \mathbf{B}_Q''^H \mathbf{\Phi}_\mathbf{v}\mathbf{i}. \tag{84}$$

We deduce that the indirect VS Wiener filter is

$$\overline{\mathbf{h}}_{\mathrm{W}}''(Q) = \mathbf{i} - \mathbf{B}_Q''\left(\mathbf{\Lambda}_Q'' + \mathbf{I}_{M-Q}\right)^{-1} \mathbf{B}_Q''^H \mathbf{\Phi}_\mathbf{v}\mathbf{i} \tag{85}$$

for $Q \geq 1$ and $\overline{\mathbf{h}}_{\mathrm{W}}''(0) = \mathbf{h}_{\mathrm{W}}' = \mathbf{\Phi}_\mathbf{y}^{-1}\mathbf{\Phi}_\mathbf{x}\mathbf{i}$, which is the classical Wiener filter. Expression (85) can be rewritten as

$$\begin{aligned}
\overline{\mathbf{h}}_{\mathrm{W}}''(Q) &= \mathbf{i} - \mathbf{B}_Q''\left(\mathbf{\Lambda}_Q'' + \mathbf{I}_{M-Q}\right)^{-1} \mathbf{B}_Q''^H \mathbf{\Phi}_\mathbf{y}\mathbf{i} \\
&\quad + \mathbf{B}_Q''\left(\mathbf{\Lambda}_Q'' + \mathbf{I}_{M-Q}\right)^{-1} \mathbf{B}_Q''^H \mathbf{\Phi}_\mathbf{x}\mathbf{i} \\
&= \mathbf{B}_Q'\left(\mathbf{\Lambda}_Q' + \mathbf{I}_Q\right)^{-1} \mathbf{B}_Q'^H \mathbf{\Phi}_\mathbf{y}\mathbf{i} \\
&\quad + \mathbf{B}_Q''\left(\mathbf{\Lambda}_Q'' + \mathbf{I}_{M-Q}\right)^{-1} \mathbf{B}_Q''^H \mathbf{\Phi}_\mathbf{x}\mathbf{i} \\
&= \mathbf{\Phi}_\mathbf{y}^{-1}\mathbf{\Phi}_\mathbf{x}\mathbf{i} + \mathbf{B}_Q'\left(\mathbf{\Lambda}_Q' + \mathbf{I}_Q\right)^{-1} \mathbf{B}_Q'^H \mathbf{\Phi}_\mathbf{v}\mathbf{i} \\
&= \mathbf{h}_{\mathrm{W}}' + \sum_{q=1}^{Q} \frac{\mathbf{b}_q\mathbf{b}_q^H}{1+\lambda_q}\mathbf{\Phi}_\mathbf{v}\mathbf{i} \\
&= \overline{\mathbf{h}}_{\mathrm{W}}''(0) + \sum_{q=1}^{Q} \frac{\mathbf{b}_q\mathbf{b}_q^H}{1+\lambda_q}\mathbf{\Phi}_\mathbf{v}\mathbf{i}
\end{aligned} \tag{86}$$

for $Q < M$ [and $\overline{\mathbf{h}}_{\mathrm{W}}''(M) = \mathbf{i}$]. It is of interest to observe that $\overline{\mathbf{h}}_{\mathrm{W}}''(P) = \mathbf{h}_{\mathrm{MVDR}}'$.

3) Indirect VS Tradeoff: The indirect VS tradeoff filter is obtained from the optimization problem:

$$\min_{\mathbf{a}_Q''} J_{\mathrm{rs}}\left(\mathbf{a}_Q''\right) \quad \text{subject to} \quad J_{\mathrm{ds}}\left(\mathbf{a}_Q''\right) = \beta''\phi_{x_1}, \tag{87}$$

where $0 \leq \beta'' \leq 1$. We find that

$$\mathbf{a}_{Q,\mathrm{T},\mu''}'' = \left(\mu''\mathbf{\Lambda}_Q'' + \mathbf{I}_{M-Q}\right)^{-1} \mathbf{B}_Q''^H \mathbf{\Phi}_\mathbf{v}\mathbf{i}, \tag{88}$$

where $\mu'' \geq 0$ is a Lagrange multiplier. As a result, the indirect VS tradeoff filter is

$$\overline{\mathbf{h}}_{\mathrm{T},\mu''}''(Q) = \mathbf{i} - \mathbf{B}_Q''\left(\mu''\mathbf{\Lambda}_Q'' + \mathbf{I}_{M-Q}\right)^{-1} \mathbf{B}_Q''^H \mathbf{\Phi}_\mathbf{v}\mathbf{i} \tag{89}$$

for $Q \geq 1$ and

$$\begin{aligned}
\overline{\mathbf{h}}_{\mathrm{T},\mu''}''(0) &= \mathbf{i} - \left(\mu''\mathbf{\Phi}_\mathbf{x} + \mathbf{\Phi}_\mathbf{v}\right)^{-1}\mathbf{\Phi}_\mathbf{v}\mathbf{i} \\
&= \left(\mathbf{\Phi}_\mathbf{x} + \mu''^{-1}\mathbf{\Phi}_\mathbf{v}\right)^{-1}\mathbf{\Phi}_\mathbf{x}\mathbf{i} = \mathbf{h}_{\mathrm{T},1/\mu''}'.
\end{aligned} \tag{90}$$

Obviously, $\overline{\mathbf{h}}_{\mathrm{T},0}''(Q) = \overline{\mathbf{h}}_{\mathrm{MR}}''(Q)$ and $\overline{\mathbf{h}}_{\mathrm{T},1}''(Q) = \overline{\mathbf{h}}_{\mathrm{W}}''(Q)$. Also, we have $\overline{\mathbf{h}}_{\mathrm{T},\mu''}''(P) = \mathbf{h}_{\mathrm{MVDR}}'$.

VII. Simulations

We now proceed by presenting the experimental evaluation of the proposed filter designs for enhancement in the time domain. That is, we consider the problem of extracting a desired speech signal, $x(n)$, at the time instance n from a vector of time-consecutive observations: $\mathbf{y}(n) = \begin{bmatrix} y(n) & \cdots & y(n-M+1) \end{bmatrix}^T$. Moreover, $\mathbf{y}(n) = \mathbf{x}(n) + \mathbf{v}(n)$ where $\mathbf{x}(n)$ and $\mathbf{v}(n)$ are vectors containing the clean speech and the noise, respectively. The evaluation is twofold: first, we have conducted a validation of the theory by using closed-form expressions for the necessary statistics in the filter design under the assumption of the desired signal being either periodic or generated by an autoregressive (AR) process, which emulates well voiced and unvoiced speech, respectively. By using such models and closed-form expressions, the performance of the filters can be evaluated without the need of estimating any statistics, which strengthens the clarity and interpretability of the results. Moreover, we use these evaluations to validate the theoretical findings regarding the relationship between the filters performances. Second, we have conducted experiments on real speech data in different noise scenarios using the proposed method and three well-known enhancement methods. The outcome of these experiments emphasizes the practical usefulness of the proposed methods.

A. Validation of Theory

In the theory validation, we first considered a scenario where the desired signal is periodic with six real harmonics, and that it is corrupted by colored noise generated by a second order AR [AR(2)] process. For this, and all the following theoretical scenarios, the filter performances were measured and averaged over 500 Monte-Carlo simulations for each setting. In each Monte-Carlo simulation, the following quantities were randomized: the fundamental frequency of the desired, periodic signal, was sampled from the uniform distribution $\mathcal{U}(0.1, 0.4)$, each harmonic amplitude was sampled from $\mathcal{U}(0.5, 1.5)$, the AR coefficients of the colored noise were found by fitting an AR(2) process [with MATLAB's `lpc()` function] to a periodic signal of length 500 with six harmonics with amplitudes sampled from $\mathcal{U}(0.5, 1.5)$ and a fundamental frequency[2] sampled from $\mathcal{U}(0.15, 0.45)$. With this setup, we first measured the output SNRs and speech reduction factors for the minimum distortion (MD), Wiener (W), and Tradeoff (T) (with $\mu = 0.2$) filters presented in Section V for $Q = 6$ (-·-·-) and $Q = 12$ (——) for different filter lengths, M, with an input SNR of 15 dB as depicted in Figure 1. We see that the differences between the filters vanish as M grows, however, we clearly see that the MD filter always has a significantly lower distortion than the other filters, especially for $Q = 12$, where the MD filter is in fact distortionless. This is expected since $Q = 12 = 2 \cdot (\#\text{ of harmonics})$. Moreover, we see that we can only expect the speech reduction factor to decrease when Q is chosen correctly when M is increased, and that there is a significant gap between the maximum output SNR and the output SNRs of the different filtering methods. A similar experiment was carried out, where the performances were instead evaluated versus the input SNR for $M = 30$, and these results are found in Figure 2. Again, we see that the output SNR is generally higher for the filters designed with $Q = 6$ than for the $Q = 12$ filters. However, these filters also have a higher distortion, and never become distortionless. At low input SNRs, the MD filters yield significantly lower output SNRs than the Wiener filters, but they also have much less distortion. As expected, the T filters can be used to obtain performances in between.

[2]The fundamental frequency is measured in radians per sample in Section VII.

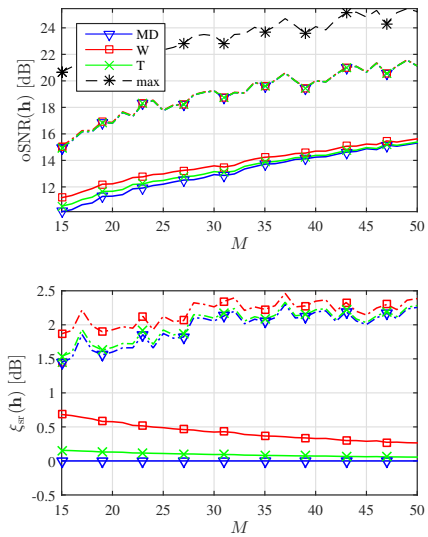

Fig. 1. Plots of (upper) the output SNR and (lower) the speech reduction factor of the minimum distortion (MD), Wiener (W), Tradeoff (T) (with $\mu = 0.2$), and maximum SNR (max) filters for $Q = 6$ (-·-·-) and $Q = 12$ (——) versus M in a scenario with a periodic desired signal corrupted by colored noise.

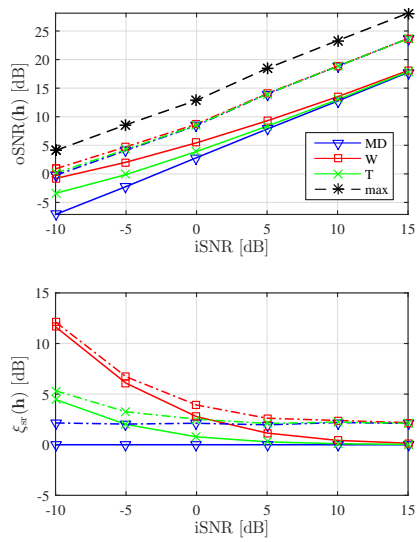

Fig. 2. Plots of (upper) the output SNR and (lower) the speech reduction factor of the minimum distortion (MD), Wiener (W), Tradeoff (T) (with $\mu = 0.2$), and maximum SNR (max) filters for $Q = 6$ (-·-·-) and $Q = 12$ (——) versus the input SNR in a scenario with a periodic desired signal corrupted by colored noise.

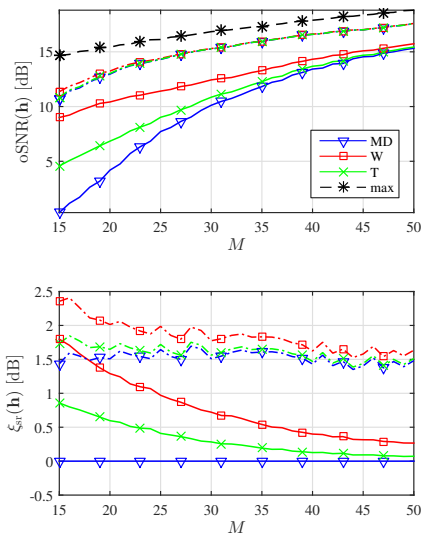

Fig. 3. Plots of (upper) the output SNR and (lower) the speech reduction factor of the minimum distortion (MD), Wiener (W), Tradeoff (T) (with $\mu = 0.2$), and maximum SNR (max) filters for $Q = 6$ (-·-·-) and $Q = 12$ (——) versus M in a scenario with a periodic desired signal corrupted by white noise and a periodic interferer.

We then considered another scenario, where the desired signal was again periodic, but the noise was instead a sum of white noise and a periodic interferer. The interfering source was a single, real sinusoid with frequency $\omega_{\mathrm{i}} = \omega_0 + 0.05$, with ω_0 being the fundamental frequency of the desired signal. The ratio between the powers of the desired signal and white noise was 10 dB, and we then fixed the input signal-to-interference (SIR) ratio to 0 dB, and evaluated the performances versus different M's. The outcomes of this evaluation are found in Figure 3. We observe that, in this scenario, the differences between the different filters are larger for low M's, but that they are generally following the same trend and approach each others performance for increasing M's. Moreover, we observe that the filters provide output SNRs closer to the maximum output SNR. A similar experiment was carried out, where M was fixed to 30, and the performances were evaluated versus the input SIR. The results are provided in Figure 4. From these results, we see that, for $Q = 6$, the filters have nearly identical output SNRs for all input SNRs, but, on the other hand, there are notable differences in their signal reduction factors, with the MD filters having the lowest. For $Q = 12$, there are significant differences in both output SNRs and signal reduction factors between the filters, though. Moreover, we note that the filters do not converge to the same output SNR in this case.

Furthermore, we considered a scenario where the desired signal was generated by an AR(6) process and the noise was generated by and AR(2). That is, in this case, we cannot avoid distorting the desired signal, as the covariance matrix of the desired signal is full rank. The colored noise was generated as previously described, whereas the desired, noisy signal was generated with an AR(6) process of which the coefficients were found by fitting to a periodic signal (as before) with the same parameters as the desired, periodic signal in the previous experiments. With this setup, we first fixed the input SNR again to 10 dB and varied the filter length, M, to obtain the results in Figure 5. First of all, we observe that the filters show equal performance in this scenario for high M's and for low M's we see the same relationship between the filter performances as in the previous experiments. Generally, the filters have an increasing output SNR as we increase M, but we also note that the distortion of the filters increase for an increasing M. In another experiment, M was fixed to 30, and the performance was measured versus different input SNRs as depicted in Figure 6. Here, the filters again show similar performance in terms of output SNR for input SNRs larger than 10 dB. For lower input SNRs, the MD and tradeoff filters have much less distortion than the Wiener filter, but they also have lower output SNRs. Generally, $Q = 6$ filters have higher output SNRs but also yield more distortion of the desired signal.

We then conducted a final evaluation on synthetic data, where we compared the proposed direct and indirect filters in Section V and VI, respectively. As described in these sections, they are only equivalent in some special cases, and will thus lead to different filters in general. This is also shown in this evaluation. We considered a scenario where the desired signal is periodic with six real harmonics and a fundamental frequency of 0.25 rad/sample. The harmonics had unit amplitudes and uniformly distributed phases between 0 and 2π. Colored noise was added, which was generated by an AR(2) process as described before. Using this setup, we then evaluated the direct and indirect tradeoff filters of length $M = 60$. The evaluation was conducted over a grid of different ranks ($1 \leq Q \leq M$ and $1 \leq \boldsymbol{Q} \leq M$), and different tradeoff parameters ($\mu' \in [10^{-2}; 10^{4}]$ and $\mu'' = 1/\mu'$). Both

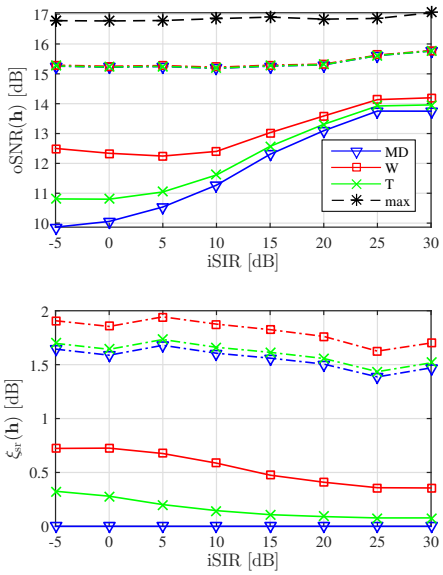

Fig. 4. Plots of (upper) the output SNR and (lower) the speech reduction factor of the minimum distortion (MD), Wiener (W), Tradeoff (T) (with $\mu = 0.2$), and maximum SNR (max) filters for $Q = 6$ (-·-·-) and $Q = 12$ (——) versus the input SIR in a scenario with a periodic desired signal corrupted by white noise and a periodic interferer.

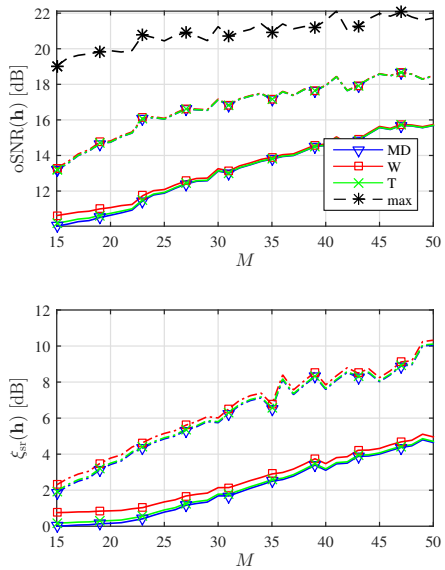

Fig. 5. Plots of (upper) the output SNR and (lower) the speech reduction factor of the minimum distortion (MD), Wiener (W), Tradeoff (T) (with $\mu = 0.2$), and maximum SNR (max) filters for $Q = 6$ (-·-·-) and $Q = 12$ (——) versus M in a scenario with a desired signal generated by an AR(6) process corrupted by colored noise.

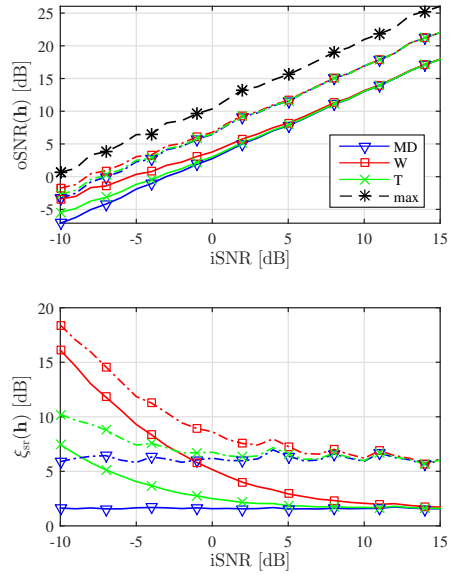

Fig. 6. Plots of (upper) the output SNR and (lower) the speech reduction factor of the minimum distortion (MD), Wiener (W), Tradeoff (T) (with $\mu = 0.2$), and maximum SNR (max) filters for $Q = 6$ (-·-·-) and $Q = 12$ (——) versus the input SNR in a scenario with a desired signal generated by an AR(6) process corrupted by colored noise.

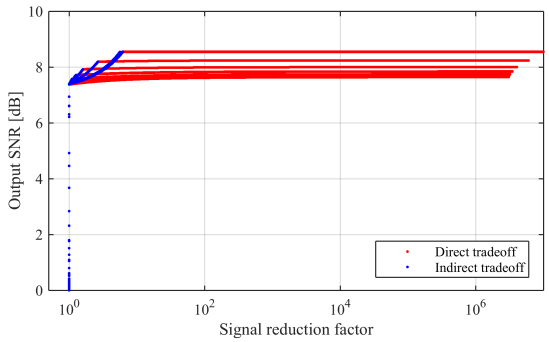

Fig. 7. Plot of the output SNR versus the signal reduction factor for the direct and indirect tradeoff filters.

the output SNRs and signal reduction factors of the filters were measured in the evaluation as depicted in Figure 7. It is clearly seen, that the filters yield the same performances in a few special cases as claimed previously. Moreover, we can see that there are different benefits of the two approaches. If, in a given application, the amount of distortion must be as low as possible, the indirect tradeoff filter generally yields lower distortion levels when the tradeoff parameter and rank is varied. On the other hand, the indirect tradeoff filter yields a relatively low output SNR for some parameter choices, as opposed to the direct tradeoff filter.

B. Real-Data Experiments

The second part of the experimental evaluation of the filters is on real-life speech. For these experiments, we used four speech excerpts (two females and two males) each of length 4–6 seconds from the Keele database [35]. In the experiments, the speech was corrupted by different kinds of noise (white, car, babble, exhibition hall, and street). All noise signals except the white noise were taken from the AURORA database [36]. The proposed filtering methods were then evaluated on these signals in different scenarios. To apply the filters on real speech we did two things: first, we assumed that the noise statistics could be estimated directly from the noise signal only since noise estimation is not the main topic of this paper. More specifically, the noise statistics were found from the past 200 samples at each time instance, and the statistics of the desired signal was found as $\mathbf{\Phi_x} = \mathbf{\Phi_y} - \mathbf{\Phi_v}$. Likewise, the observed signal statistics were estimated from the past 200 samples of the observed signal using the sample covariance estimator. Second, since the dimensionality of the signal subspace is not known in practice, we estimated Q using a percentage-of-variance (PoV) principle. That is, for each segment, Q was chosen such that

$$\gamma < \frac{\sum_{q=1}^{Q} \lambda_q}{\sum_{m=1}^{M} \lambda_m}, \tag{91}$$

is satisfied with the smallest possible Q. Besides the proposed filtering methods, two speech enhancement methods based on spectral subtraction from the VOICEBOX[3] toolbox for MATLAB, were included in the evaluation for comparison. The first of these method (SS) uses the gain function proposed in [37], whereas the other (SSm) uses the one proposed in [38]. In addition to this, the optimally modified LSA (OMLSA) speech estimator [39], [40] by Cohen & Berdugo was included in the evaluation[4]. The methods were applied with the default settings provided with the toolbox.

[3]http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html
[4]http://webee.technion.ac.il/people/IsraelCohen/Info/Software.html

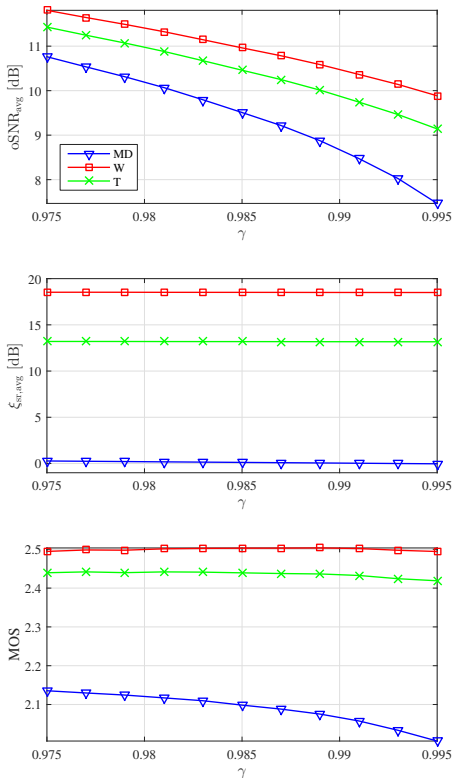

Fig. 8. Speech enhancement performance of the proposed and comparison methods in terms of (top) output SNR, (mid) speech reduction factor, and (bottom) MOS (PESQ) scores versus the PoV used for choosing Q.

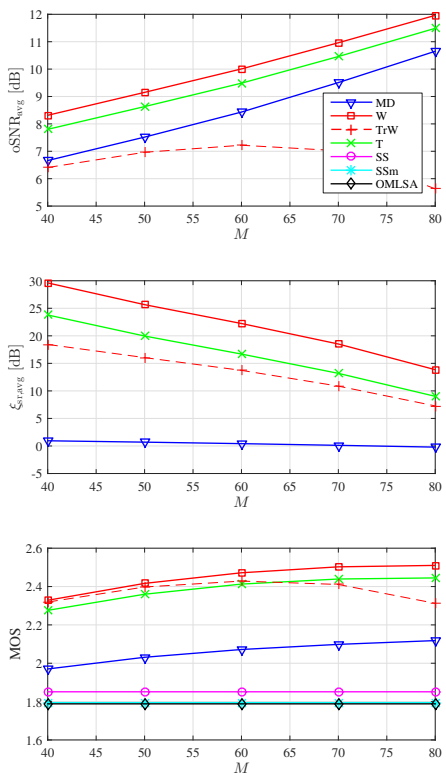

Fig. 9. Speech enhancement performance of the proposed and comparison methods in terms of (top) output SNR, (mid) speech reduction factor, and (bottom) MOS (PESQ) scores versus the filter length M.

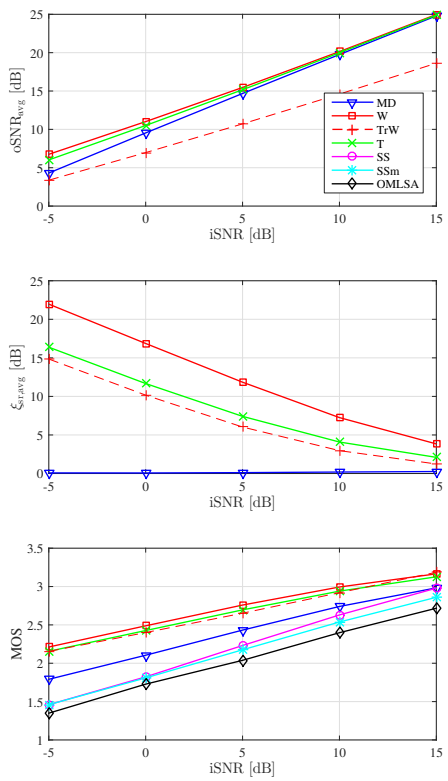

Fig. 10. Speech enhancement performance of the proposed and comparison methods in terms of (top) output SNR, (mid) speech reduction factor, and (bottom) MOS (PESQ) scores versus the input SNR.

With the above-mentioned setup, we first evaluated the performance of the proposed methods versus the PoV threshold, γ, in terms of the output SNR, the signal reduction factor, and an objective perceptual mean opinion score (MOS) obtained using a "Perceptual Evaluation of Speech Quality" (PESQ) [41], [42] MATLAB implementation[5]. In this simulation, the input SNR was 0 dB, and the filter length was $M = 70$, and the results are provided in Figure 8. Clearly, the choice of the PoV threshold have an impact on the enhancement performance, and it affects the methods differently. All methods have decreasing output SNRs and slightly decreasing signal reduction factors for an increasing γ. Regarding the perceptual performance, the MOS score for the MD methods is highest for low γ's, whereas higher γ's are needed to maximize the MOS scores of the Tradeoff and Wiener filter ($\gamma \approx 0.981$ and $\gamma \approx 0.987$, respectively). As a tradeoff, we used $\gamma = 0.985$ in the following experiments.

In the next experiment, the performances of the aforementioned methods were evaluated versus the filter length, M, when the input SNR was fixed to 0 dB. The results from this experiment are provided in Figure 9, where (- - -) denotes the performance of the traditional Wiener (TrW) filter. The first thing to note is that the variable span Wiener filter can indeed outperform the traditional Wiener filter in terms of output SNR and perceptual performance, especially for $M > 60$. Additionally, the variable span Wiener filter outperforms the traditional Wiener filter in terms of both output SNR and the speech reduction factor. This is very interesting since increase noise reduction is often considered to only be achievable by increasing the signal distortion, which is not the case here. Besides that we see the same relation between the filters in terms of output SNR and speech reduction as in the theoretical experiments. For all M's, the proposed methods clearly outperforms the SS, SSm, and OMLSA methods. Note, however, that the difference might be smaller in practice, where the noise estimation will be more difficult, but these results show the potential performance improvement compared to those methods. Moreover, it should be noted that the SS, SSm, and OMLSA methods are only evaluated using the perceptual scores, as the output SNRs and speech reduction factor cannot be easily measured for those methods. This would require that these methods are rewritten as linear transformations of the observations. Finally, the performances were also evaluated versus different input SNRs for a fixed filter length of $M = 70$, where the performance of the proposed methods peaks in the previous experiment. Again, similar relations between the proposed methods compared to the theoretical simulations are observed, regarding the output SNRs and signal reduction factors (see Figure 10). Furthermore, the proposed methods generally seem to outperform the SS, SSm, and OMLSA methods for all input SNRs, with the largest performance improvement being obtained for low input SNRs. Finally, the

[5]http://ecs.utdallas.edu/loizou/speech/software.htm

variable span Wiener filter is again shown to outperform the traditional Wiener filter in terms of both noise reduction and speech distortion performance. Aside from the aforementioned objective measurements, we have also conducted informal listening tests. From our experience, the MOS's obtained using PESQ reflects very well the actual perceptual quality of the evaluated enhancement methods.

VIII. Conclusions

We have introduced the novel concept of variable span filters, wherein noise reduction filters are formed by linear combinations of the eigenvectors from the joint diagonalization of the covariance matrices of the signal of interest and the noise. We have shown how the resulting filters reduce to well-known filter designs such as MVDR, Wiener, maximum SNR, and tradeoff filters in some special cases. Moreover, the variable span filters include also generalizations of these filter designs, resulting in new optimal filters for noise reduction. An interesting aspect of the variable span filters is that it is possible to trade signal distortion for noise reduction in an explicit way, meaning that via two simple user parameter, it is possible to achieve higher output SNRs by allowing some distortion on the signal of interest. Tradeoff filters with this capability have been considered before, but the proposed variable span tradeoff filters can control the distortion and noise reduction levels over a much wider range since an additional tradeoff parameter, namely the number of eigenvectors used in building the filter, has been introduced. Simulations have demonstrated the properties of the various designs and emphasized the flexibility of the variable span filter in terms of being able to trade off noise reduction for less signal distortion. Moreover, the simulations indicated that there is a potential perceptual advantage of using the proposed filters compared to traditional, widely used speech enhancement methods. Another key observation from the simulations is that the proposed variable span Wiener filter can outperform its traditional counterpart in terms of both output SNR and speech reduction factor levels. In other words, we can exploit speech subspaces with a low rank to obtain more noise reduction without increasing the amount of distortion compared to traditional filter designs.

References

[1] J. Benesty and J. Chen, *Optimal Time-Domain Noise Reduction Filters – A Theoretical Study*, 1st ed. Springer, 2011, no. VII.

[2] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 27, no. 2, pp. 113–120, Apr. 1979.

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

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

[5] S. Srinivasan, J. Samuelsson, and W. B. Kleijn, "Codebook-based bayesian speech enhancement for nonstationary environments," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 15, no. 2, pp. 441–452, Feb. 2007.

[6] 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, 1995.

[7] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sørensen, "Reduction of broad-band noise in speech by truncated QSVD," *IEEE Trans. Speech Audio Process.*, vol. 3(6), pp. 439–448, 1995.

[8] S. Rangachari and P. Loizou, "A noise estimation algorithm for highly nonstationary environments," *Speech Communication*, vol. 28, pp. 220–231, 2006.

[9] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 466–475, 2003.

[10] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 20, no. 4, pp. 1383–1393, 2012.

[11] R. C. Hendriks, R. Heusdens, J. Jensen, and U. Kjems, "Low complexity DFT-domain noise PSD tracking using high-resolution periodograms," *EURASIP J. on Advances in Signal Process.*, vol. 2009, no. 1, p. 15, 2009.

[12] 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.

[13] M. Souden, J. Benesty, and S. Affes, "On optimal frequency-domain multichannel linear filtering for noise reduction," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 18, no. 2, pp. 260–276, Feb. 2010.

[14] J. Benesty, M. Souden, and J. Chen, "A perspective on multichannel noise reduction in the time domain," *Applied Acoustics*, vol. 74, no. 3, pp. 343–355, Mar. 2013.

[15] R. C. Hendriks and T. Gerkmann, "Noise correlation matrix estimation for multi-microphone speech enhancement," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 20, no. 1, pp. 223–233, Jan. 2012.

[16] M. G. Christensen and A. Jakobsson, "Optimal filter designs for separating and enhancing periodic signals," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 5969–5983, 2010.

[17] J. R. Jensen, J. Benesty, M. G. Christensen, and S. H. Jensen, "Enhancement of single-channel periodic signals in the time-domain," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 7, pp. 1948–1963, 2012.

[18] ——, "Non-causal time-domain filters for single-channel noise reduction," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 5, pp. 1526–1541, 2012.

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

[20] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Noise Reduction in Speech Processing*. Berlin, Germany: Springer-Verlag, 2009.

[21] P. C. Hansen and S. H. Jensen, "Subspace-based noise reduction for speech signals via diagonal and triangular matrix decompositions: Survey and analysis," *EURASIP J. on Advances in Signal Process.*, vol. 2007, no. 1, p. 24, 2007.

[22] L. L. Scharf, "The SVD and reduced rank signal processing," *Signal Process.*, vol. 25, no. 2, pp. 113–133, Nov. 1991.

[23] P. Strobach, "Low-rank adaptive filters," *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 2932–2947, Dec. 1996.

[24] Y. Hu and P. C. Loizou, "A subspace approach for enhancing speech corrupted by colored noise," *IEEE Signal Process. Lett.*, vol. 9, no. 7 pp. 204–206, Jul. 2002.

[25] S. M. Nørholm, J. Benesty, J. R. Jensen, and M. G. Christensen, "Single-channel noise reduction using unified joint diagonalization and optimal filtering," *EURASIP J. on Applied Signal Process.*, vol. 2014, no. 1, p. 37, Mar. 2014.

[26] S. M. Nørholm, J. Benesty, J. R. Jensen, and M. G. Christensen, "Noise reduction in the time domain using joint diagonalization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2014, pp. 7058–7062.

[27] 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.

[28] K. Hermus, P. Wambacq, and HV. hamme, "A review of signal subspace speech enhancement and its application to noise robust speech recognition," *EURASIP J. on Advances in Signal Process.*, vol. 2007, no. 1, p. 15, 2007.

[29] B. Picinbono, "On circularity," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3473–3482, Dec. 1994.

[30] J. Benesty, J. Chen, and Y. A. Huang, "A widely linear distortionless filter for single-channel noise reduction," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 469–472, May 2010.

[31] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment*. Chichester, England: John Wiley & Sons Ltd, 2006. 439–448, Nov. 1995.

[32] J. N. Franklin, *Matrix Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1968.

[33] A. H. Gray, and J. D. Markel, "Distance measures for speech processing," *IEEE Trans. Acoustics, Speech, and Signal Process.*, vol. ASSP-24, no. 5, Oct. 1976.

[34] J. Yang, H. Xi, F. Yang, and Y. Zhao, "RLS-based adaptive algorithms for generalized eigen-decomposition," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1177–1188, Apr. 2006.

[35] F. Plante, G. F. Meyer, and W. A. Ainsworth, "A pitch extraction reference database," in *Proc. Eurospeech*, Sep. 1995, pp. 837–840.

[36] D. Pearce, and H. G. Hirsch, "The AURORA experimental framework for the performance evaluation of speech recognition systems under noisy conditions," in *Proc. Int. Conf. Spoken Language Process.*, Oct. 2000.

[37] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, 1979, pp. 208–211.

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

[39] I. Cohen, and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Process.*, vol. 81, no. 11, pp. 2403–2418, Nov. 2001.

[40] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," *IEEE Trans. Speech and Audio Process.*, vol. 11, no. 5, pp. 466–475, Sep. 2003.

[41] "Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," ITU-T Rec. P.862, 02/2001.

[42] Y. Hu, and P. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. on Speech and Audio Process.*, vol. 16 no. 1, pp. 229–238, Jan. 2008.

Jesper Rindom Jensen (S'09–M'12) was born in Ringkøbing, Denmark in August 1984. He received the M.Sc. degree *cum laude* for completing the elite candidate education in 2009 from Aalborg University in Denmark. In 2012, he received the Ph.D. degree from Aalborg University. Currently, he is a Postdoctoral Researcher at the Department of Architecture, Design & Media Technology at Aalborg University in Denmark, where he is also a member of the Audio Analysis Lab. He has been a Visiting Researcher at the University of Quebec, INRS-EMT, in Montreal, Quebec, Canada, and at the Friedrich-Alexander Universität Erlangen-Nürnberg in Erlangen, Germany.

His research interests include signal processing theory and methods for, e.g., microphone array and joint audio-visual signal processing. Examples of more specific research interests within this scope are enhancement, separation, localization, tracking, parametric analysis, and modeling. He has published nearly more than 50 papers on these topics in top-tier, peer-reviewed conference proceedings and journals. Moreover, he has published 2 research monographs including the book Speech Enhancement A Signal Subspace Perspective which is co-authored with Prof. Jacob Benesty, Prof. Mads Grsbll Christensen, and Prof. Jingdong Chen.

He has received a highly competitive postdoc grant from the Danish Independent Research Council, as well as several travel grants from private foundations. Furthermore, he is an affiliate member of the IEEE Signal Processing Theory and Methods Technical Committee, and is Member of the IEEE.

Jacob Benesty was born in 1963. He received a Master degree in microwaves from Pierre & Marie Curie University, France, in 1987, and a Ph.D. degree in control and signal processing from Orsay University, France, in April 1991. During his Ph.D. (from Nov. 1989 to Apr. 1991), he worked on adaptive filters and fast algorithms at the Centre National d'Etudes des Telecomunications (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, USA. In May 2003, he joined the University of Quebec, INRS-EMT, in Montreal, Quebec, Canada, as a Professor. He is also a Visiting Professor at the Technion, in Haifa, Israel, an Adjunct Professor at Aalborg University, in Denmark, and a Guest Professor at Northwestern Polytechnical University, in Xi'an, Shaanxi, China.

His research interests are in signal processing, acoustic signal processing, and multimedia communications. He is the inventor of many important technologies. In particular, he was the lead researcher at Bell Labs who conceived and designed the world-first real-time hands-free full-duplex stereophonic teleconferencing system. Also, he conceived and designed the world-first PC-based multi-party hands-free full-duplex stereo conferencing system over IP networks.

He was the co-chair of the 1999 International Workshop on Acoustic Echo and Noise Control and the general co-chair of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. He is the recipient, with Morgan and Sondhi, of the IEEE Signal Processing Society 2001 Best Paper Award. He is the recipient, with Chen, Huang, and Doclo, of the IEEE Signal Processing Society 2008 Best Paper Award. He is also the co-author of a paper for which Huang received the IEEE Signal Processing Society 2002 Young Author Best Paper Award. In 2010, he received the "Gheorghe Cartianu Award" from the Romanian Academy. In 2011, he received the Best Paper Award from the IEEE WASPAA for a paper that he co-authored with Chen.

Mads Græsbøll Christensen (S'00–M'05–SM'11) received the M.Sc. and Ph.D. degrees in 2002 and 2005, respectively, from Aalborg University (AAU) in Denmark, where he is also currently employed at the Dept. of Architecture, Design & Media Technology as Professor in Audio Processing and is head and founder of the Audio Analysis Lab.

He was formerly with the Dept. of Electronic Systems at AAU and has been held visiting positions at Philips Research Labs, ENST, UCSB, and Columbia University. He has published 3 books and more than 150 papers in peer-reviewed conference proceedings and journals, and he has given tutorials at EUSIPCO and INTERSPEECH. His research interests include signal processing theory and methods with application to speech and audio, in particular parametric analysis, modeling, enhancement, separation, and coding.

Prof. Christensen has received several awards, including an ICASSP Student Paper Contest Award, the Spar Nord Foundations Research Prize, a Danish Independent Research Council Young Researchers Award, and the Statoil Prize, and he is also co-author of the paper *Sparse Linear Prediction and Its Application to Speech Processing* that received an IEEE Signal Processing Society Young Author Best Paper Award. Moreover, he is a beneficiary of major grants from the Danish Independent Research Council, the Villum Foundation, and Innovation Fund Denmark. He is an Associate Editor for IEEE/ACM Trans. on Audio, Speech, and Language Processing, a former Associate Editor of IEEE Signal Processing Letters, a member of the IEEE Audio and Acoustic Signal Processing Technical Committee, and a Senior Member of the IEEE.