Enhancement of Single-Channel Periodic Signals in the Time-Domain

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

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

DOI (link to publication from Publisher):
10.1109/TASL.2012.2191957

Publication date:
2012

Document Version
Accepted author manuscript, peer reviewed version

Enhancement of Single-Channel Periodic Signals in the Time-Domain

Jesper Rindom Jensen*, *Student Member, IEEE,* Jacob Benesty, Mads Græsbøll Christensen, *Senior Member, IEEE,* and Søren Holdt Jensen, *Senior Member, IEEE,*

***Abstract*—Most state-of-the-art filtering methods for speech enhancement require an estimate of the noise statistics, but the noise statistics are difficult to estimate in practice when speech is present. Thus, non-stationary noise will have a detrimental impact on the performance of most speech enhancement filters. The impact of such noise can be reduced by using the signal statistics rather than the noise statistics in the filter design. For example, this is possible by assuming a harmonic model for the desired signal; while this model fits well for voiced speech, it will not be appropriate for unvoiced speech. That is, signal-dependent methods based on the signal statistics will introduce undesired distortion for some parts of speech compared to signal-independent methods based on the noise statistics. Since both the signal-independent and signal-dependent approaches to speech enhancement have advantages, it is relevant to combine them to reduce the impact of their individual disadvantages. In this paper, we give theoretical insights into the relationship between these different approaches, and these reveal a close relationship between the two approaches. This justifies joint use of such filtering methods which can be beneficial from a practical point of view. Our experimental results confirm that both signal-independent and signal-dependent approaches have advantages and that they are closely-related. Moreover, as a part of our experiments, we illustrate the practical usefulness of combining signal-independent and signal-dependent enhancement methods by applying such methods jointly on real-life speech.**

***Index Terms*—Single-channel speech enhancement, time-domain filtering, MVDR filter, LCMV filter, non-stationary noise, orthogonal decomposition, harmonic decomposition, pitch, performance measures.**

I. Introduction

HUMAN speech is frequently encountered in several signal processing applications such as telecommunications, teleconferencing, hearing-aids, and human-machine interfaces. Before the speech can be utilized in such applications, it must be picked up by one or more microphones. Unfortunately, the desired signal (in this case speech) will always, to a certain degree, be corrupted by noise which is present when sampling the signal. The noise will most likely have a detrimental impact on speech applications since it may degrade the speech quality and intelligibility. In hearing-aids, for example, a decreased speech quality (i.e., a high noise level) can cause listener fatigue. Therefore, it is of great importance to develop methods for reducing the noise of speech recordings before the speech is utilized in any relevant application. Such methods are typically termed noise reduction methods or enhancement methods. In the past few decades, developing such methods have been a major challenge. For an overview of existing enhancement methods, we refer to, e.g., [1], [2]. In general, we can divide speech enhancement methods into three groups, i.e., spectral-subtractive algorithms [3], statistical-model-based algorithms [4], [5], and subspace algorithms [6]–[8]. The references, [3]–[8], refer to some of the pioneering work within each of the groups.

A common approach used in speech enhancement is linear filtering. In this approach, the speech enhancement problem is formulated as a filter design problem. That is, a filter should be designed such that it reduces the noise level of the observed signal as much as possible while not introducing any noticeable distortion of the speech. The design of such a filter can be performed either directly in the time domain or in some transform domain. This could for example be in the frequency [3], [7], [9] or in the Karhunen-Loève expansion (KLE) domains [10]. The advantage of filtering in transform domains can, for example, be a reduced computational complexity. Filters derived in transform domains, however, can also be derived equivalently in other domains and vice versa. In this paper, we consider time-domain filters for single-channel recordings which can also be extended to other domains according to the previous discussion. Typically, time-domain filters are designed by minimizing some error function like in the classical Wiener filter design [11]. The first step in the design is therefore to define the error function.

In the vast majority of filtering methods for speech enhancement, the filter is designed from the statistics of the observed signal and the noise. We term this the signal-independent filter design approach. In practice, however, the noise signal is not directly available, and the noise statistics could, for example, be estimated during silence periods where only the noise is present. The main advantage of this approach is that it is completely independent of the statistics of the desired speech signal since it only uses the observed signal and the noise statistics, and it is well-known that the speech structure changes drastically over time. However, the signal-independent filter approach will not be influenced by this, since it does not rely on the statistics of the desired signal. Non-stationary noise, on the other hand, will have a detrimental impact on

Manuscript received June 07, 2011; revised September 25, 2011 and December 27, 2011; accepted March 12, 2012.

J. R. Jensen and S. H. Jensen are with the Dept. of Electronic Systems, Aalborg University, DK-9220 Aalborg, e-mail: {jrj,shj}@es.aau.dk.

J. Benesty is with INRS-EMT, University of Quebec, Montreal, QC H5A 1K6, Canada, e-mail: benesty@emt.inrs.ca

M. G. Christensen is with the Dept. of Architecture, Design & Media Technology, Aalborg University, DK-9220 Aalborg, e-mail: mgc@create.aau.dk

Digital Object Identifier XXXX

this filter design approach since the noise statistics are difficult to estimate when speech is present.

Recently, a signal-dependent filter design approach has been proposed [12]. By signal-dependent, we mean that the filter is calculated using the statistics of the desired signal and without using the statistics of the noise. The desired signal is assumed to be periodic in this approach and is therefore well-modeled by a sum of harmonically related sinusoids. This type of harmonic modeling has been used extensively within speech processing. Due to the periodicity assumption, the filter in [12] ends up being driven only by the pitch, the harmonic model order, and the statistics of the observed signal. In this paper, the pitch and the number of harmonics will be treated as known parameters, and we refer the interested reader to [13]–[22] and the references therein for an overview of methods for estimation of these parameters when they are unknown. Since the signal-dependent approach does not depend directly on the noise statistics, it will be robust against non-stationary noise as opposed to the signal-independent filter design approach. However, the harmonic model will only be appropriate for voiced speech segments. For unvoiced speech segments, the signal-dependent approach will therefore introduce some distortion of the speech signal due to model mismatch.

As highlighted in the previous discussion that the signal-independent and signal-dependent filter design approaches have complementary advantages and disadvantages. Therefore, it is highly relevant to investigate if these approaches can be combined to obtain the advantages of both while reducing the impact of their disadvantages. As a first step in this direction, we here provide further insight into the relationship between the signal-independent and signal-dependent filter design approaches in this paper. More specifically, we consider the relationship between two recently proposed filter designs, namely the orthogonal decomposition based minimum variance distortionless response (ODMVDR) filter [23], and the harmonic decomposition linearly constrained minimum variance (HDLCMV) filter [12], [21]. The ODMVDR filter is signal-independent whereas the HDLCMV filter is signal-dependent. Moreover, we present some closed-form performance measures for filters designed using both the signal-independent and signal-dependent design approaches when the desired signal is periodic. A new performance measure for the harmonic distortion is also proposed. The closed-form expressions for the performance measures enable easy comparison of the filters. Finally, in the experimental part of the paper, we propose a filtering scheme where the ODMVDR and HDLCMV filters are used jointly. By doing this, we can, to some extend, have the individual advantages of both a signal-independent and a signal-dependent filtering approach.

The remainder of the paper is organized as follows. In Section II, we define the signal model which forms the basis of the paper. Then, in Section III, we introduce the notion of using filtering for enhancement purposes for different signal decompositions. Based on this, we briefly introduce two recently proposed optimal filter designs for enhancement in Section IV. In Section V, we perform a theoretical study of the two filters, and we show that there is a clear link between them. When the desired signal is periodic, we can obtain closed-form expression for the filter performance measures which we describe in Section VI. In the experimental part of the paper, in Section VII, we compare the ODMVDR and HDLCMV filters through simulations, and we propose and evaluate a scheme where the ODMVDR and HDLCMV filters are used jointly for speech enhancement. Finally, we conclude on the paper in Section VIII.

II. Signal Model

In this paper, we consider the performance and the relationship of recent optimal filter designs for enhancement of a zero-mean desired signal, $x(n) \in \mathbb{R}^{1\times 1}$, buried in additive noise, $v(n) \in \mathbb{R}^{1\times 1}$, where n denotes the discrete-time index. That is, the objective is to recover $x(n)$ from a mixture signal given by

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

The mixture signal, $y(n) \in \mathbb{R}^{1\times 1}$, could be a microphone recording and the desired signal could be a speech signal. We assume that the noise, $v(n)$, is a zero-mean random process uncorrelated with the desired signal, $x(n)$. Specifically, we consider the special scenario where $x(n)$ is quasi-periodic which is a reasonable assumption for voiced speech segments. Considering this special scenario enables us to provide closed-form solutions for the enhancement performance measures, and it enables us to investigate the relationship between different optimal filter designs. These observations will become clear from the later sections.

By assuming quasi-periodicity, we can rewrite the signal model in (1) as

$$y(n) = \sum_{l=1}^{L} A_l \cos(l\omega_0 n + \phi_l) + v(n) \ , \tag{2}$$

where ω_0 is the pitch, L is the number of harmonics, A_l is the amplitude of the lth harmonic, and ϕ_l is the phase of the lth harmonic. For many signals, the harmonic model does not fit exactly due to inharmonicity, but we can cope with this by modifying the signal model in several ways (see, e.g., [21] and the references therein). However, inharmonicity is out of the scope of this paper, and it will not be discussed any further. Without loss of generality, we can also write the signal model in (2) as

$$y(n) = \sum_{l=1}^{L} \left(a_l e^{jl\omega_0 n} + a_l^* e^{-jl\omega_0 n}\right) + v(n) \ , \tag{3}$$

with $a_l = \frac{A_l}{2} e^{j\phi_l}$ being the complex amplitude of the lth harmonic, and $(\cdot)^*$ denotes the element-wise complex conjugate of a matrix/vector.

The observed data can be stacked into a vector, $\mathbf{y}(n) \in \mathbb{R}^{M\times 1}$, which enables us to do block processing. The vector signal model is given by

$$\mathbf{y}(n) = \mathbf{x}(n) + \mathbf{v}(n) \ , \tag{4}$$

where

$$\mathbf{y}(n) = \begin{bmatrix} y(n) & y(n-1) & \cdots & y(n-M+1) \end{bmatrix}^T \ , \tag{5}$$

with $(\cdot)^T$ denoting the matrix/vector transpose, and the definitions of $\mathbf{x}(n) \in \mathbb{R}^{M\times 1}$ and $\mathbf{v}(n) \in \mathbb{R}^{M\times 1}$ resemble the definition of $\mathbf{y}(n)$. Since we have assumed that $x(n)$ and $v(n)$ are uncorrelated, we can obtain the following simple expression for the covariance matrix, $\mathbf{R}_{\mathbf{y}} \in \mathbb{R}^{M\times M}$, of the observed signal

$$\mathbf{R}_{\mathbf{y}} = \mathrm{E}[\mathbf{y}(n)\mathbf{y}^T(n)] = \mathbf{R}_{\mathbf{x}} + \mathbf{R}_{\mathbf{v}} \ , \tag{6}$$

where $\mathrm{E}[\cdot]$ is the expectation operator, $\mathbf{R}_{\mathbf{x}} \in \mathbb{R}^{M\times M}$ is the covariance matrix of $\mathbf{x}(n)$ and $\mathbf{R}_{\mathbf{v}} \in \mathbb{R}^{M\times M}$ is the covariance matrix of $\mathbf{v}(n)$. Under the assumption of $x(n)$ being quasi-periodic, we know that $\mathbf{R}_{\mathbf{x}}$ can be modeled by [24]

$$\mathbf{R}_{\mathbf{x}} \approx \mathbf{Z}(\omega_0)\mathbf{P}\mathbf{Z}^H(\omega_0) \ , \tag{7}$$

where $(\cdot)^H$ denotes the complex conjugate transpose operator, and

$$\mathbf{P} = \mathrm{diag}\left\{\begin{bmatrix}|a_1|^2 & |a_1^*|^2 & \cdots & |a_L|^2 & |a_L^*|^2\end{bmatrix}\right\} \ , \tag{8}$$

$$\mathbf{Z}(\omega_0) = \begin{bmatrix}\mathbf{z}(\omega_0) & \mathbf{z}^*(\omega_0) & \cdots & \mathbf{z}(L\omega_0) & \mathbf{z}^*(L\omega_0)\end{bmatrix} \ , \tag{9}$$

$$\mathbf{z}(l\omega_0) = \begin{bmatrix}1 & e^{-jl\omega_0} & \cdots & e^{-jl\omega_0(M-1)}\end{bmatrix}^T \ , \tag{10}$$

with $\mathrm{diag}\{\cdot\}$ denoting the construction of a diagonal matrix from a vector. In the remainder of the paper, we denote $\mathbf{Z}(\omega_0)$ as $\mathbf{Z}$ to get a simpler notation.

A common goal in different enhancement algorithms is then to find a "good" estimate of $x(n)$ or $\mathbf{x}(n)$. Often, in enhancement problems, "good" means that the noise reduction should be significant while the desired signal remains nearly undistorted. In this paper, we focus on two recently proposed filtering methods which estimate $x(n)$ from an observation vector, $\mathbf{y}(n)$, of length M.

III. Enhancement by Linear Filtering

Linear filters have been widely used for enhancement purposes. For example, enhancement performed by applying a finite impulse response (FIR) filter to the observed signal vector, $\mathbf{y}(n)$. The filtering operation can be written as

$$\hat{x}(n) = \sum_{m=0}^{M-1} h_m y(n-m) = \mathbf{h}^T\mathbf{y}(n) \ , \tag{11}$$

where

$$\mathbf{h} = \begin{bmatrix}h_0 & h_1 & \cdots & h_{M-1}\end{bmatrix}^T \tag{12}$$

and $\hat{x}(n)$ should be an estimate of $x(n)$. The output of the filter is often decomposed into a filtered desired signal part and a filtered noise part to facilitate the filter design. We here describe three different decompositions of the filter output: the classical, the orthogonal, and the harmonic decompositions.

A. Classical Decomposition

In most classical filtering methods for signal enhancement, the filter output is decomposed as

$$\hat{x}(n) = \mathbf{h}^T\mathbf{x}(n) + \mathbf{h}^T\mathbf{v}(n) = x_{\mathrm{f}}(n) + v_{\mathrm{rn}}(n), \tag{13}$$

where $x_{\mathrm{f}}(n) \triangleq \mathbf{h}^T\mathbf{x}(n)$ is the signal after filtering and $v_{\mathrm{rn}}(n) \triangleq \mathbf{h}^T\mathbf{v}(n)$ is the residual noise. The goal in the filter design is then two-fold. First, the noise should be attenuated significantly by filtering. Second, the distortion of the desired signal introduced by the filter should be low. Numerous filter designs have been proposed according to these design criteria. A common approach is to minimize the mean-square error (MSE) between the desired signal and the enhanced signal, where the error is defined as

$$e(n) = x(n) - \hat{x}(n) \ . \tag{14}$$

In [23], however, it was claimed and shown that this approach can be inappropriate since only some of the information embedded in $\mathbf{x}(n)$ is useful for the estimation of $x(n)$.

B. Orthogonal Decomposition

Recently, it has been proposed to design an enhancement filter based on an orthogonal decomposition of the desired signal since some components of $\mathbf{x}(n)$ interfere with the estimation of the desired signal $x(n)$ [23]. Using the orthogonal decomposition, the clean signal can be rewritten as

$$\mathbf{x}(n) = x(n)\boldsymbol{\rho}_{\mathbf{x}x} + \mathbf{x}_{\mathrm{i}}(n) = \mathbf{x}_{\mathrm{d}}(n) + \mathbf{x}_{\mathrm{i}}(n) \ , \tag{15}$$

where

$$\begin{aligned}\boldsymbol{\rho}_{\mathbf{x}x} &= \frac{\mathrm{E}[\mathbf{x}(n)x(n)]}{\mathrm{E}[x^2(n)]} \\ &= \begin{bmatrix}1 & \rho_x(1) & \cdots & \rho_x(M-1)\end{bmatrix}^T \ ,\end{aligned} \tag{16}$$

$$\rho_x(m) = \frac{\mathrm{E}[x(n-m)x(n)]}{\mathrm{E}[x^2(n)]} \ . \tag{17}$$

Note that $\mathbf{x}_{\mathrm{d}}(n)$ is the part of $\mathbf{x}(n)$ being proportional to the desired signal $x(n)$ and $\mathbf{x}_{\mathrm{i}}(n)$ is the "interference" being orthogonal to $\mathbf{x}_{\mathrm{d}}(n)$. Inserting (15) into (13) yields

$$\hat{x}(n) = \mathbf{h}^T\mathbf{x}_{\mathrm{d}}(n) + \mathbf{h}^T\mathbf{x}_{\mathrm{i}}(n) + \mathbf{h}^T\mathbf{v}(n) \ . \tag{18}$$

It can be shown that the variance of $\hat{x}(n)$ is given by [23]

$$\sigma_{\hat{x}}^2 = \sigma_{x_{\mathrm{fd}}}^2 + \sigma_{x_{\mathrm{ri}}}^2 + \sigma_{v_{\mathrm{rn}}}^2 \ , \tag{19}$$

where

$$\sigma_{x_{\mathrm{fd}}}^2 = \mathbf{h}^T\mathbf{R}_{\mathbf{x}_{\mathrm{d}}}\mathbf{h} = \sigma_x^2(\mathbf{h}^T\boldsymbol{\rho}_{\mathbf{x}x})^2 \ , \tag{20}$$

$$\sigma_{x_{\mathrm{ri}}}^2 = \mathbf{h}^T\mathbf{R}_{\mathbf{x}_{\mathrm{i}}}\mathbf{h} \ , \tag{21}$$

$$\sigma_{v_{\mathrm{rn}}}^2 = \mathbf{h}^T\mathbf{R}_{\mathbf{v}}\mathbf{h} \ , \tag{22}$$

$\mathbf{R}_{\mathbf{x}_{\mathrm{d}}} = \sigma_x^2\boldsymbol{\rho}_{\mathbf{x}x}\boldsymbol{\rho}_{\mathbf{x}x}^T$ is the covariance matrix of $\mathbf{x}_{\mathrm{d}}(n)$, $\sigma_x^2 = \mathrm{E}[x^2(n)]$ is the variance of the desired signal, and $\mathbf{R}_{\mathbf{x}_{\mathrm{i}}} = \mathrm{E}[\mathbf{x}_{\mathrm{i}}(n)\mathbf{x}_{\mathrm{i}}^T(n)]$ is the covariance matrix of the interference, $\mathbf{x}_{\mathrm{i}}(n)$.

The main difference between the classical approach and this approach is that we have two noise terms to minimize in this approach, namely $\sigma_{x_{\mathrm{ri}}}^2$ and $\sigma_{v_{\mathrm{rn}}}^2$. Moreover, the filtered desired signal is different in this approach since it does not include the interfering part of $\mathbf{x}(n)$ which is here considered as noise. Like in the previous approach, the filter should be designed such that the error in (14) is small (e.g., in the MSE sense) while there is no or only a little distortion of the desired signal.

C. Harmonic Decomposition

The harmonic model in (2) has been used in many pitch estimation methods [21]. In general, the model can be used for describing periodic signals as

$$\mathbf{x}(n) = \mathbf{Z}\mathbf{a}(n) = \mathbf{x}'_{\text{d}}(n) \ , \tag{23}$$

where

$$\mathbf{a}(n) = \begin{bmatrix} a_1 e^{j\omega_0 n} & a_1^* e^{-j\omega_0 n} & \cdots \\ & & a_L e^{jL\omega_0 n} \quad a_L^* e^{-jL\omega_0 n} \end{bmatrix}^T \ . \tag{24}$$

Note that in this approach there is no interference as opposed to in the orthogonal decomposition approach since all samples in $\mathbf{x}(n)$ can be fully used for describing the desired signal $x(n)$. This is due to the underlying harmonic signal model. Therefore, the vector, $\mathbf{x}'_{\text{d}}(n)$, describing the desired signal, $x(n)$, is simply equal to the signal vector, $\mathbf{x}(n)$, in this approach. The desired signal, $x(n)$, is equal to the first entry of the vector $\mathbf{Z}\mathbf{a}(n)$, i.e.,

$$x(n) = \mathbf{1}^T \mathbf{a}(n) \ , \tag{25}$$

where $\mathbf{1} = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}^T$. Like in the orthogonal decomposition approach, we can insert (23) into (13) which yields the following estimate of $x(n)$

$$\hat{x}'(n) = \mathbf{h}^T \mathbf{x}'_{\text{d}}(n) + \mathbf{h}^T \mathbf{v}(n) \ . \tag{26}$$

If we exploit the orthogonality between $\mathbf{x}'_{\text{d}}(n)$ and $\mathbf{v}(n)$ in (26), we can write the variance of $\hat{x}'(n)$ as

$$\sigma^2_{\hat{x}'} = \sigma^2_{x'_{\text{fd}}} + \sigma^2_{v_{\text{rn}}} \ , \tag{27}$$

where

$$\sigma^2_{x'_{\text{fd}}} = \mathbf{h}^T \mathbf{R}_{\mathbf{x}'_{\text{d}}} \mathbf{h} = \mathbf{h}^T \mathbf{Z}\mathbf{P}\mathbf{Z}^H \mathbf{h} \ , \tag{28}$$

and $\sigma^2_{v_{\text{rn}}}$ is defined as in (22). Moreover, $\mathbf{R}_{\mathbf{x}'_{\text{d}}} = \text{E}\left[\mathbf{x}'_{\text{d}}(n)\mathbf{x}'^T_{\text{d}}(n)\right] = \mathbf{Z}\mathbf{P}\mathbf{Z}^H$ is the covariance matrix of $\mathbf{x}'_{\text{d}}(n)$.

Compared to the orthogonal decomposition approach, this approach only has one noise term, $\sigma^2_{v_{\text{rn}}}$. When this approach is used, the filter, $\mathbf{h}$, should therefore be designed such that it minimizes $\sigma^2_{v_{\text{rn}}}$ without distorting the $x(n)$ too much.

IV. Optimal Filters for Enhancement

We consider two recently proposed filter designs for enhancement of single-channel signals: 1) the orthogonal decomposition MVDR filter [23] and 2) the harmonic decomposition LCMV filter [20]. Following, we will revisit the two filter designs.

A. Orthogonal Decomposition MVDR

Traditionally, the minimum variance distortionless response (MVDR) filter proposed by Capon [25], [26] has been derived and applied in the context of multichannel signals. Recently, however, the MVDR filter has also been applied for single-channel speech enhancement [23]. Here, we term the MVDR filter proposed in [23] as the orthogonal decomposition MVDR (ODMVDR) filter. The ODMVDR filter design is based on an orthogonal decomposition of the desired signal as described in Section III-B. The filter is designed to minimize the sum of the residual interference variance, $\sigma^2_{x_{\text{ri}}}$, and the residual noise variance, $\sigma^2_{v_{\text{rn}}}$, while it should not distort the desired signal. That is,

$$\min_{\mathbf{h}} \mathbf{h}^T \mathbf{R}_{\text{in}} \mathbf{h} \quad \text{s.t.} \quad \mathbf{h}^T \boldsymbol{\rho}_{\mathbf{x}x} = 1, \tag{29}$$

where $\mathbf{R}_{\text{in}} = \mathbf{R}_{\mathbf{x}_{\text{i}}} + \mathbf{R}_{\mathbf{v}}$ is the interference-plus-noise covariance matrix. The constraint comes from the measure of desired signal reduction (aka. speech reduction) for the orthogonal decomposition introduced in [23]

$$\xi_{\text{dsr}}(\mathbf{h}) = \frac{\sigma_x^2}{\sigma^2_{x_{\text{fd}}}} = \frac{1}{(\mathbf{h}^T \boldsymbol{\rho}_{\mathbf{x}x})^2} \ . \tag{30}$$

When $\xi_{\text{dsr}}(\mathbf{h}) = 1$ there is no desired signal reduction (or distortion if you will) while it is expected to be greater than 1 when there is a reduction. That is, to make the filter distortionless according to this measure, we must require that $\mathbf{h}^T \boldsymbol{\rho}_{\mathbf{x}x} = 1$ which exactly corresponds to the constraint in (29).

The well-known solution to the quadratic optimization problem in (29) is given by

$$\mathbf{h}_{\text{ODMVDR}} = \frac{\mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x}}{\boldsymbol{\rho}_{\mathbf{x}x}^T \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x}} = \frac{\mathbf{R}_{\mathbf{y}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x}}{\boldsymbol{\rho}_{\mathbf{x}x}^T \mathbf{R}_{\mathbf{y}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x}} \ . \tag{31}$$

In practice, the correlation vector, $\boldsymbol{\rho}_{\mathbf{x}x}$, in (31) is replaced by

$$\begin{aligned} \boldsymbol{\rho}_{\mathbf{x}x} &= \frac{\text{E}[\mathbf{y}(n)y(n)] - \text{E}[\mathbf{v}(n)v(n)]}{\sigma_y^2 - \sigma_v^2} \\ &= \frac{\sigma_y^2 \boldsymbol{\rho}_{\mathbf{y}y} - \sigma_v^2 \boldsymbol{\rho}_{\mathbf{v}v}}{\sigma_y^2 - \sigma_v^2} \ , \end{aligned} \tag{32}$$

where σ_y^2 is the variance of $y(n)$, σ_v^2 is the variance of $v(n)$, and $\boldsymbol{\rho}_{\mathbf{y}y}$ and $\boldsymbol{\rho}_{\mathbf{v}v}$ are defined similarly to $\boldsymbol{\rho}_{\mathbf{x}x}$ in (16). The evaluation of the performance of the ODMVDR filter follows from later sections.

B. Harmonic Decomposition LCMV

Like the MVDR filter, the linearly constrained minimum variance (LCMV) filter proposed by Frost [27] has mainly been used in multichannel settings. Recently, however, an LCMV filtering method for enhancement of periodic signals was proposed which is applicable on single-channel signals [12], [20]. Following, we recast the LCMV design procedure from [20] such that it is more general and compliant with the harmonic decomposition in Section III-C. This design procedure is somewhat similar to that of the ODMVDR filter.

In the harmonic decomposition LCMV (HDLCMV) filter, it is assumed that the desired signal is periodic. When the desired signal is periodic and modeled by (3), all information in $\mathbf{x}(n)$ can be used in the estimation of $x(n)$ which, in general, is not the case in the orthogonal decomposition approach where there will be some interference, $\mathbf{x}_{\text{i}}(n)$. Therefore, we only need to care about minimizing the residual noise power, $\sigma^2_{v_{\text{rn}}}$, in the harmonic decomposition approach without introducing too much desired signal distortion. The HDLCMV filter, in particular, is designed such that the residual noise variance, $\sigma^2_{v_{\text{rn}}}$, is minimized while the desired signal, $x(n)$, is passed

undistorted. This can also be casted as the following optimization problem

$$\min_{\mathbf{h}} \mathbf{h}^T\mathbf{R}_{\mathbf{v}}\mathbf{h} \quad \text{s.t.} \quad \mathbf{Z}^H\mathbf{h} = \mathbf{1} \ . \tag{33}$$

To verify that the constraint in (33) makes the filter distortionless, we consider the desired signal reduction measure for the harmonic decomposition approach which is given by

$$\xi'_{\text{dsr}}(\mathbf{h}) = \frac{\sigma_x^2}{\sigma_{x'_{\text{fd}}}^2} = \frac{\sigma_x^2}{\mathbf{h}^T\mathbf{Z}\mathbf{P}\mathbf{Z}^H\mathbf{h}} \ . \tag{34}$$

It can be seen that when the signal is periodic, the desired signal variance is given by $\sigma_x^2 = \mathbf{1}^T\mathbf{P}\mathbf{1}$. That is, the filter will indeed be distortionless with respect to the distortion measure in (34) if it is designed such that $\mathbf{Z}\mathbf{h} = \mathbf{1}$. It can also be shown that the constraint in (33) ensures that the individual harmonics are not distorted [24].

If we solve the quadratic optimization problem with multiple constraints in (33), we get

$$\mathbf{h}_{\text{HDLCMV}} = \mathbf{R}_{\mathbf{v}}^{-1}\mathbf{Z}\left(\mathbf{Z}^H\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{Z}\right)^{-1}\mathbf{1} \ . \tag{35}$$

In the Appendix, we have shown that replacing $\mathbf{R}_{\mathbf{v}}$ by $\mathbf{R}_{\mathbf{y}}$ does not change the filter response. If we utilize this, we can also write the HDLCMV filter as

$$\mathbf{h}_{\text{HDLCMV}} = \mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\left(\mathbf{Z}^H\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\right)^{-1}\mathbf{1} \ . \tag{36}$$

We can see from this expression that if $x(n)$ is periodic, the pitch, ω_0, is known, and the number of harmonics, L, is known, we only need the statistics, $\mathbf{R}_{\mathbf{y}}$, of the observed signal to design the HDLCMV filter. This is a key difference from the design of the ODMVDR filter for which we also need to know either the statistics of the desired signal, $\boldsymbol{\rho}_{\mathbf{x}x}$, or of the noise, $\boldsymbol{\rho}_{\mathbf{v}v}$.

V. Relation between the ODMVDR and HDLCMV Filters

Although the ODMVDR and HDLCMV filters were derived under different constraints, we show in this section that there is a clear link between the filters. For this analysis, we assume that the noise is a sum of interfering sinusoids and white Gaussian noise such that

$$\mathbf{R}_{\mathbf{v}} = \mathbf{Z}_{\text{sn}}\mathbf{P}_{\text{sn}}\mathbf{Z}_{\text{sn}}^H + \sigma_{\text{wn}}^2\mathbf{I} \ , \tag{37}$$

where $\mathbf{Z}_{\text{sn}}$ and $\mathbf{P}_{\text{sn}}$ are the steering and power matrices of the sinusoidal noise source, and σ_{wn}^2 is the variance of the white Gaussian noise. The matrices are defined similarly to (8) and (9).

It is clear from (16) that $\boldsymbol{\rho}_{\mathbf{x}x}$ corresponds to the first column of $\mathbf{R}_{\mathbf{x}}$ normalized with respect to the signal variance, σ_x^2. That is, without loss of generality, we can also write $\boldsymbol{\rho}_{\mathbf{x}x}$ as

$$\boldsymbol{\rho}_{\mathbf{x}x} = \frac{\mathbf{R}_{\mathbf{x}}\mathbf{i}}{\mathbf{i}^T\mathbf{R}_{\mathbf{x}}\mathbf{i}} \ , \tag{38}$$

where $\mathbf{i} = \begin{bmatrix}1 & 0 & \cdots & 0\end{bmatrix}^T \in \mathbb{R}^{M\times 1}$. Under the periodicity assumption, we can rewrite this expression by inserting (7) into (38)

$$\boldsymbol{\rho}_{\mathbf{x}x} = \frac{\mathbf{Z}\mathbf{P}\mathbf{Z}^H\mathbf{i}}{\mathbf{i}^T\mathbf{Z}\mathbf{P}\mathbf{Z}^H\mathbf{i}} = \frac{\mathbf{Z}\mathbf{P}\mathbf{1}}{\mathbf{1}^T\mathbf{P}\mathbf{1}} = \frac{\mathbf{Z}\mathbf{P}\mathbf{1}}{\sigma_x^2} \ . \tag{39}$$

If we substitute this expression for $\boldsymbol{\rho}_{\mathbf{x}x}$ back into the expression for the ODMVDR filter in (31), we get that

$$\begin{aligned}\mathbf{h}_{\text{ODMVDR}} &= \mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\mathbf{P}\mathbf{1}\left(\mathbf{1}^T\mathbf{P}\mathbf{Z}^H\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\mathbf{P}\mathbf{1}\right)^{-1}\sigma_x^2 \\ &= \sigma_x^2\mathbf{B}\mathbf{P}\left(\mathbf{1}^T\mathbf{P}\mathbf{C}\mathbf{P}\mathbf{1}\right)^{-1}\mathbf{1} \ ,\end{aligned} \tag{40}$$

where $\mathbf{B} = \mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}$ and $\mathbf{C} = \mathbf{Z}^H\mathbf{B}$. Note that using the same notation, the HDLCMV filter can be written as

$$\mathbf{h}_{\text{HDLCMV}} = \mathbf{B}\mathbf{C}^{-1}\mathbf{1} \ . \tag{41}$$

At a first glance, the filters in (40) and (41) do not look similar. However, by using the matrix inversion lemma on $\mathbf{C}$, we see that it can be rewritten as

$$\begin{aligned}\mathbf{C} &= \mathbf{Z}^H\left(\mathbf{Z}\mathbf{P}\mathbf{Z}^H + \mathbf{R}_{\mathbf{v}}\right)^{-1}\mathbf{Z} \\ &= \mathbf{Z}^H\left[\mathbf{R}_{\mathbf{v}}^{-1} - \mathbf{R}_{\mathbf{v}}^{-1}\mathbf{Z}\left(\mathbf{P}^{-1} + \mathbf{Z}^H\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}^H\mathbf{R}_{\mathbf{v}}^{-1}\right]\mathbf{Z} \\ &= \mathbf{D} - \mathbf{D}\left(\mathbf{P}^{-1} + \mathbf{D}\right)^{-1}\mathbf{D} \ ,\end{aligned} \tag{42}$$

where $\mathbf{D} = \mathbf{Z}^H\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{Z}$. If we also use the matrix inversion lemma on $\mathbf{D}$, we get that

$$\mathbf{D} = \frac{1}{\sigma_{\text{wn}}^2}\mathbf{Z}^H\mathbf{Z} - \frac{1}{\sigma_{\text{wn}}^2}\mathbf{Z}^H\mathbf{Z}_{\text{sn}}\left(\mathbf{P}_{\text{sn}}^{-1} + \frac{1}{\sigma_{\text{wn}}^2}\mathbf{Z}_{\text{sn}}^H\mathbf{Z}_{\text{sn}}\right)^{-1}\mathbf{Z}_{\text{sn}}^H\mathbf{Z} \ . \tag{43}$$

Moreover, if we then assume that the frequencies of the sinusoidal noise sources are different from the harmonic frequencies, and if we let $M \to \infty$, we can write [21]

$$\lim_{M\to\infty}\frac{1}{M}\mathbf{Z}^H\mathbf{Z} = \mathbf{I} \ , \tag{44}$$

$$\lim_{M\to\infty}\frac{1}{M}\mathbf{Z}^H\mathbf{Z}_{\text{sn}} = \mathbf{0} \ . \tag{45}$$

Thus, for large M, we can approximate $\mathbf{C}$ as

$$\mathbf{C} \approx \sigma_v^{-2}\left[M\mathbf{I} - M^2\left(\sigma_v^2\mathbf{P}^{-1} + M\mathbf{I}\right)^{-1}\right] \ . \tag{46}$$

Furthermore, it turns out that we can approximate the (p,q)th element of $\mathbf{C}$ as

$$[\mathbf{C}]_{pq} \approx \begin{cases}\dfrac{M}{\sigma_v^2 + P_q M} \ , & \text{for } p = q \\ 0 \ , & \text{for } p \neq q\end{cases} \ . \tag{47}$$

When M is large and $P_q M \gg \sigma_v^2$, the expression for the qth diagonal element of $\mathbf{C}$ can be further simplified as $[\mathbf{C}]_{qq} \approx P_q^{-1}$. In this case, we can write

$$\mathbf{C} \approx \mathbf{P}^{-1} \ . \tag{48}$$

If we insert this approximation for $\mathbf{C}$ in (40), we readily obtain that

$$\begin{aligned}\lim_{M\to\infty}\mathbf{h}_{\text{ODMVDR}} &= \mathbf{B}\mathbf{P}\mathbf{1} \\ &= \mathbf{h}_{\text{HDLCMV}} \ .\end{aligned} \tag{49}$$

Thus, when the desired signal is periodic, the noise is a summation of interfering sinusoids and white Gaussian noise, and the filter order M is large, then the ODMVDR and HDLCMV filters are approximately identical. This observation is important since it justifies the joint use of the two filters

for enhancement of quasi-periodic signals. The two different filters are based on different knowledge, i.e., the noise and signal statistics, respectively. Depending on which statistics are available, the appropriate filter can be applied. In the experimental part of the paper, we also investigate the relation between the filters for small Ms.

VI. Performance Measures

In [23], a number of performance measures for enhancement methods were introduced. In this section, we exploit the periodicity of the desired signal to derive closed-form expressions for the performance measures for each of the filters described in Section IV.

A. Noise Reduction

The most fundamental measure of the performance of enhancement algorithms is the signal-to-noise ratio (SNR). In general, we can consider two SNRs, namely the input SNR (iSNR) and the output SNR (oSNR). The iSNR is defined as the SNR of the observed signal before filtering, i.e.,

$$\text{iSNR} = \frac{\sigma_x^2}{\sigma_v^2} \ . \tag{50}$$

The oSNR, on the other hand, is the SNR after noise reduction. That is, when using the orthogonal decomposition, it is obtained as

$$\text{oSNR}^{\text{OD}}(\mathbf{h}) = \frac{\sigma_{x_{\text{fd}}}^2}{\sigma_{x_{\text{ri}}}^2 + \sigma_{v_{\text{rn}}}^2} = \frac{\sigma_x^2 \left(\mathbf{h}^T \boldsymbol{\rho}_{\mathbf{x}x}\right)^2}{\mathbf{h}^T \mathbf{R}_{\text{in}} \mathbf{h}} \ . \tag{51}$$

where $(\cdot)^{\text{OD}}$ denotes that the measure is applicable when using the orthogonal decomposition. We can then obtain a closed-form expression for the oSNR of the ODMVDR filter when the desired signal is periodic by inserting (39) and (40) into (51). This yields

$$\text{oSNR}^{\text{OD}}(\mathbf{h}_{\text{ODMVDR}}) = \frac{\mathbf{1}^T \mathbf{P} \mathbf{Z}^H \mathbf{R}_{\text{in}}^{-1} \mathbf{Z} \mathbf{P} \mathbf{1}}{\sigma_x^2} \ . \tag{52}$$

When the harmonic decomposition is utilized, the oSNR is given as

$$\text{oSNR}^{\text{HD}}(\mathbf{h}) = \frac{\sigma_{x'_{\text{fd}}}^2}{\sigma_{v_{\text{rn}}}^2} = \frac{\mathbf{h}^T \mathbf{Z} \mathbf{P} \mathbf{Z}^H \mathbf{h}}{\mathbf{h}^T \mathbf{R}_{\mathbf{v}} \mathbf{h}} \ , \tag{53}$$

where $(\cdot)^{\text{HD}}$ denotes that the measure is applicable when using the harmonic decomposition. A closed-form expression for the oSNR of the HDLCMV filter is then found by inserting (41) into (53), which yields

$$\text{oSNR}^{\text{HD}}(\mathbf{h}_{\text{HDLCMV}}) = \frac{\sigma_x^2}{\mathbf{1}^T \left(\mathbf{Z}^H \mathbf{R}_{\mathbf{v}}^{-1} \mathbf{Z}\right)^{-1} \mathbf{1}} \ . \tag{54}$$

Yet another performance measure related to the noise reduction, is the so-called noise reduction factor, $\xi_{\text{nr}}(\mathbf{h})$. This factor is defined as the ratio between the noise in the observed signal and the noise remaining in the signal after filter. That is, when the orthogonal decomposition is used, the noise reduction factor is given by

$$\begin{aligned} \xi_{\text{nr}}^{\text{OD}}(\mathbf{h}) &= \frac{\sigma_v^2}{\sigma_{x_{\text{ri}}}^2 + \sigma_{v_{\text{rn}}}^2} \\ &= \frac{\sigma_v^2}{\mathbf{h}^T \mathbf{R}_{\text{in}} \mathbf{h}} \ . \end{aligned} \tag{55}$$

The noise reduction factor is expected to be larger than or equal to 1, since $\xi_{\text{nr}}(\mathbf{h}) < 1$ would imply that the noise is amplified through the filtering. If we insert the expression for the ODMVDR filter into (40), we get that

$$\xi_{\text{nr}}^{\text{OD}}(\mathbf{h}_{\text{ODMVDR}}) = \frac{\sigma_v^2 \mathbf{1}^T \mathbf{P} \mathbf{Z}^H \mathbf{R}_{\text{in}}^{-1} \mathbf{Z} \mathbf{P} \mathbf{1}}{\sigma_x^4} \ . \tag{56}$$

If the harmonic decomposition is used instead, the noise reduction factor is obtained as

$$\xi_{\text{nr}}^{\text{HD}}(\mathbf{h}) = \frac{\sigma_v^2}{\sigma_{v_{\text{rn}}}^2} = \frac{\sigma_v^2}{\mathbf{h}^T \mathbf{R}_{\mathbf{v}} \mathbf{h}} \ . \tag{57}$$

This gives the following noise reduction factor for the HDL-CMV filter

$$\xi_{\text{nr}}^{\text{HD}}(\mathbf{h}_{\text{HDLCMV}}) = \frac{\sigma_v^2}{\mathbf{1}^T \left(\mathbf{Z}^H \mathbf{R}_{\mathbf{v}}^{-1} \mathbf{Z}\right)^{-1} \mathbf{1}} \ . \tag{58}$$

Note that if we know the pitch, ω_0, the number of harmonics, L, the powers of the harmonics, P_l, and the noise statistics, $\mathbf{R}_{\mathbf{v}}$, we can calculate the output SNRs and the noise reduction factors for the two filters.

B. Signal Distortion

A common and unwanted side-effect of most enhancement procedures is that they also attenuate the desired signal in the process of attenuating the noise. The desired signal attenuation can also be considered as distortion. The amount of distortion can be quantified by the speech reduction factor measure [23]. Here, the measure will be termed the desired signal reduction factor since we do not consider speech only. The reduction factor is defined as the ratio between the variance of the desired signal and the variance of the desired signal after filtering. That is, when the orthogonal decomposition is used, the factor is given by

$$\begin{aligned} \xi_{\text{dsr}}^{\text{OD}}(\mathbf{h}) &= \frac{\sigma_x^2}{\sigma_{x_{\text{fd}}}^2} \\ &= \frac{1}{\left(\mathbf{h}^T \boldsymbol{\rho}_{\mathbf{x}x}\right)^2} \ . \end{aligned} \tag{59}$$

If distortion occurs, the noise reduction factor will be greater or less than one (expectedly greater than one) and it will equal 1 otherwise. Therefore, if a filter should be distortionless, we must require that

$$\mathbf{h}^T \boldsymbol{\rho}_{\mathbf{x}x} = 1 \ . \tag{60}$$

The ODMVDR filter was derived exactly under this constraint, i.e.,

$$\xi_{\text{dsr}}^{\text{OD}}(\mathbf{h}_{\text{ODMVDR}}) = 1 \ , \tag{61}$$

which can also be easily verified. Similarly, for the harmonic decomposition approach, the desired signal distortion is defined as

$$\xi_{\text{dsr}}^{\text{HD}}(\mathbf{h}) = \frac{\sigma_x^2}{\sigma_{x'_{\text{fd}}}^2} = \frac{\sigma_x^2}{\mathbf{h}^T\mathbf{Z}\mathbf{P}\mathbf{Z}^H\mathbf{h}} \ . \tag{62}$$

The HDLCMV filter is designed to be distortionless when the desired signal is periodic, i.e.,

$$\xi_{\text{dsr}}^{\text{HD}}(\mathbf{h}_{\text{HDLCMV}}) = 1 \ . \tag{63}$$

This result can easily be verified. On a side note, it can be seen that the HDLCMV filter is also distortionless with respect to the desired signal reduction measure for the orthogonal decomposition approach since

$$\begin{aligned}\mathbf{h}_{\text{HDLCMV}}^T\boldsymbol{\rho}_{\mathbf{x}x} &= \frac{\mathbf{1}^T\left(\mathbf{Z}^H\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}^H\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\mathbf{P}\mathbf{1}}{\sigma_x^2} \\ &= \frac{\mathbf{1}^T\mathbf{P}\mathbf{1}}{\sigma_x^2} = 1 \ . \end{aligned} \tag{64}$$

This emphasizes the strong link between the two filters.

We also propose a new distortion measure, namely the harmonic distortion. The harmonic distortion is the sum of the differences between the powers of the harmonics before and after filtering which can also be written as

$$\begin{aligned}\xi_{\text{hd}}(\mathbf{h}) &= 2\sum_{l=1}^{L}|P_l - P_{\text{f},l}| \\ &= 2\sum_{l=1}^{L}P_l|1 - \mathbf{h}^T\mathbf{z}(l\omega_0)\mathbf{z}^H(l\omega_0)\mathbf{h}| \ , \end{aligned} \tag{65}$$

where $P_l = |a_l|^2$ and $P_{\text{f},l}$ is the power of the lth harmonic after filtering. This performance measure is defined in exactly the same way for both the orthogonal decomposition approach and the harmonic decomposition approach. The harmonic distortion will be equal to 0 when there is no distortion of the harmonics while it will be greater than 0 otherwise. A closed-form expression for the harmonic distortion of the ODMVDR filter can be obtained by inserting (40) into (65) which yields

$$\xi_{\text{hd}}(\mathbf{h}_{\text{ODMVDR}}) = 2\sum_{l=1}^{L}P_l\left|1 - \frac{\sigma_x^4\left|\mathbf{1}^T\mathbf{P}\mathbf{Z}^H\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{z}(l\omega_0)\right|^2}{\left(\mathbf{1}^T\mathbf{P}\mathbf{Z}^H\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\mathbf{P}\mathbf{1}\right)^2}\right| \ . \tag{66}$$

It is clear from the above expression that the harmonic distortion of the ODMVDR filter will be close to 0 when M is large. The HDLCMV filter is derived under the constraints that the harmonics should not be distorted, i.e.,

$$\xi_{\text{hd}}(\mathbf{h}_{\text{HDLCMV}}) = 0 \ , \tag{67}$$

which is readily verified by inserting (41) into (65).

VII. Experimental Results

In the previous sections, we presented two single-channel filtering methods which can be used for extraction of periodic sources. These are the ODMVDR and HDLCMV filters. We showed that there is a clear link between the filters and that they are even equivalent in some special scenarios. To illustrate the link, we compare the responses of the filters in this section. The link between the filters suggests that they can be used jointly which can be useful in practice as we illustrate and account for in the application example later in this section. Furthermore, we defined some performance measures for both of the methods given that the underlying desired signal is periodic and modeled by (3). In this section, we will also study these measures through theoretical simulations.

A. Qualitative Comparison of Filter Responses

In this theoretical experiment, we compared the ODMVDR and HDLCMV filters in terms of their filter responses in different scenarios. The signal and noise statistics were assumed to be known in this experiment, i.e., we assumed that the desired signal was constituted by a sum of $L = 6$ harmonic sinusoids with a pitch of $\omega_0 = 0.245$. Each of the sinusoids was assumed to have a unit amplitude ($A_l = 1$).

In the first part of the experiment, we compared the ODMVDR and HDLCMV filters in (31) and (36), respectively, when white Gaussian noise, $v_{\text{wn}}(n)$, was added to the desired signal, $x(n)$, at an iSNR of 10 dB. When the filter length was set to $M = 20$, we obtained the filter responses depicted in Fig. 1. We observe from the plot that the filters have poor noise reduction capabilities due to the relatively short filter length. Furthermore, we can see that the filters have different magnitude responses. By careful inspection, we note that the HDLCMV filter has unit gains at the harmonic frequencies as a result of its constraints which is not the case for the ODMVDR filter. When we increase the filter length to $M = 40$, we get the responses in Fig. 2. In accordance with the theoretical discussion in Section V, we observe that the filters become equivalent when the filter order becomes large.

In the second part of the experiment, the noise was a summation of white Gaussian noise, $v_{\text{wn}}(n)$, and sinusoidal noise, $v_{\text{sn}}(n)$, containing 6 harmonics with unit amplitudes. The pitch of the sinusoidal noise source was 0.247. The ratio between the desired signal and the white Gaussian noise was 10 dB resulting in an iSNR of -0.41 dB. First, we designed ODMVDR and HDLCMV filters of length $M = 50$, and the resulting responses are shown in Fig. 3. The filter responses are close, and they both seem to extract the desired signal while attenuating both the sinusoidal noise, $v_{\text{sn}}(n)$, and the white noise, $v_{\text{wn}}(n)$. When we increase the filter order, the filters become almost equivalent, as can be seen from Fig. 4. This was also expected in the sinusoidal noise scenario according to Section V.

B. Evaluation of the Filter Performances

The second experiment was about evaluation of the performance of the ODMVDR and HDLCMV filters in different scenarios. The performance measures considered in this section were the output SNR and the harmonic distortion. As in the first experiment, this experiment was conducted with exact statistics, i.e., without synthetic data samples. In all simulations, the desired signal, $x(n)$, was a periodic signal containing $L = 6$ harmonic sinusoids. We conducted

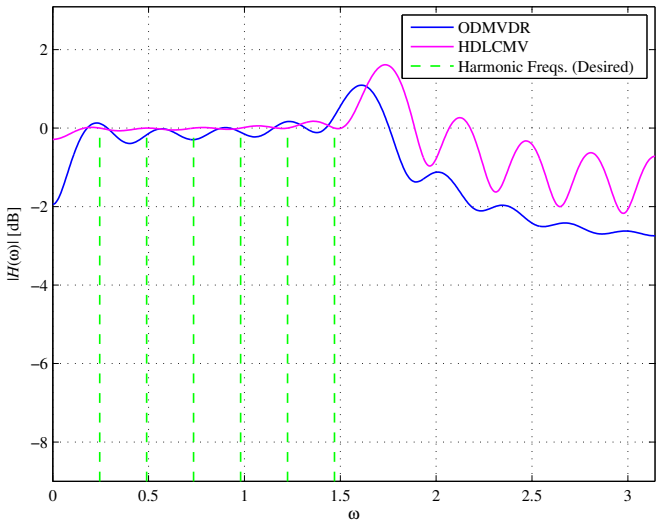


Fig. 1. Magnitude responses of the ODMVDR and HDLCMV filters of order $M = 20$ designed for a periodic signal corrupted by white Gaussian noise.

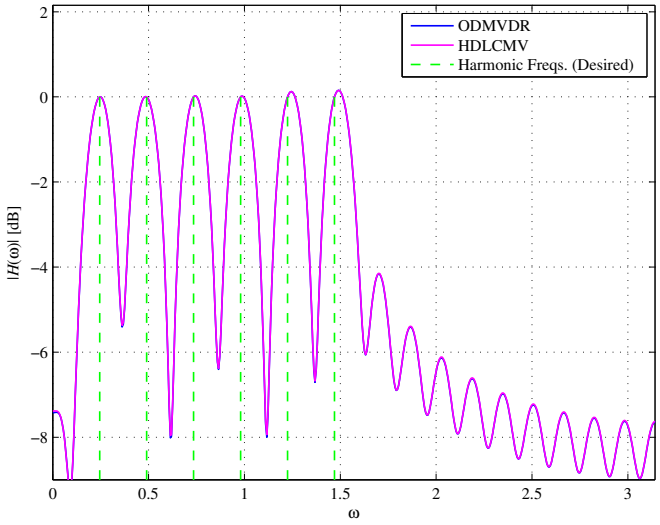


Fig. 2. Magnitude responses of the ODMVDR and HDLCMV filters of order $M = 40$ designed for a periodic signal corrupted by white Gaussian noise.

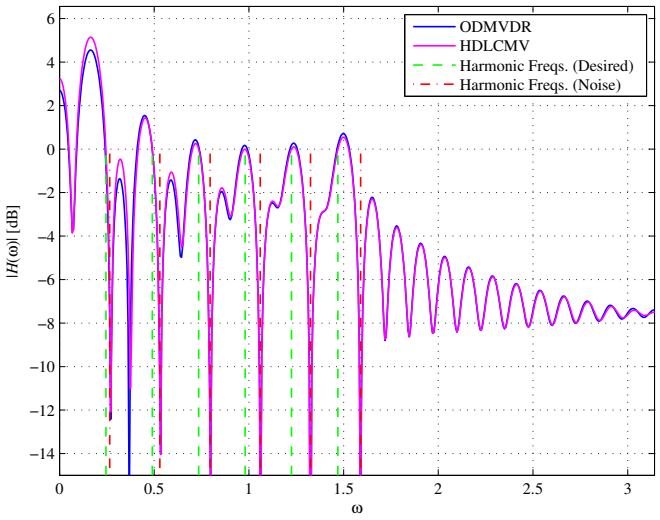


Fig. 3. Magnitude responses of the ODMVDR and HDLCMV filters of order $M = 50$ designed for a periodic signal corrupted by sinusoidal noise and white Gaussian noise.

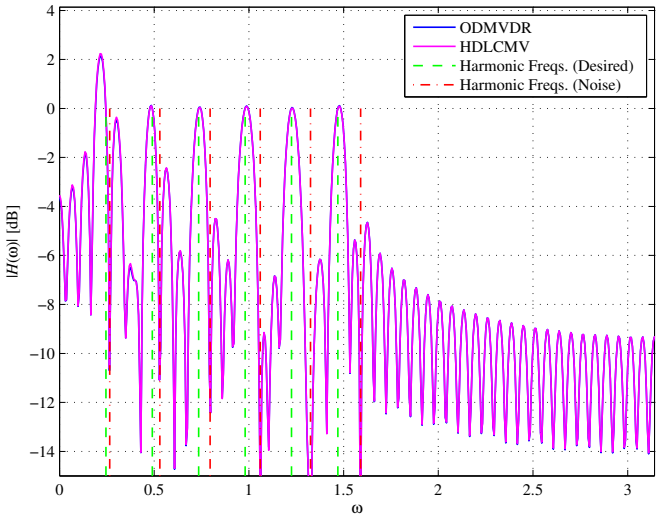


Fig. 4. Magnitude responses of the ODMVDR and HDLCMV filters of order $M = 100$ designed for a periodic signal corrupted by sinusoidal noise and white Gaussian noise.

simulations with both unit amplitude harmonics ($A_l = 1$) and harmonics with decreasing amplitudes

$$\begin{bmatrix} A_1 & \cdots & A_6 \end{bmatrix}^T = \begin{bmatrix} 1 & 0.8 & 0.5 & 0.35 & 0.2 & 0.1 \end{bmatrix}^T . \quad (68)$$

By using decreasing amplitudes, we believe that we get a slightly better insight into the performance of the filters when the desired signal is speech which often has decreasing harmonic amplitudes. In all of the simulations in this experiment, the pitch of the desired signal was $\omega_0 = 0.245$.

First, we measured the performance of the two filters as a function of the iSNR. In this simulation, the filter length was $M = 30$, and the desired signal, $x(n)$, was corrupted by white Gaussian noise. For the scenario with unit amplitude harmonics, we obtained the results depicted in Fig. 5. Both filters improved the SNR by approximately 6 dB for all iSNRs. However, the ODMVDR filter had a little distortion of the harmonics at low iSNRs. For decreasing harmonic amplitudes, we got the results in Fig. 6. Note that in this scenario, the ODMVDR filter has a slightly higher oSNR than the HDLCMV filter at low iSNRs. However, the higher oSNR comes at the cost of distortion of the harmonics.

Next, we compared the performance of the filters as a function of the filter length. In these simulations, the desired signal, $x(n)$, was corrupted by white Gaussian noise at an iSNR of 10 dB. First, the performance comparison was conducted for unit harmonic amplitudes resulting in the plot in Fig. 7. While the oSNRs of the filters are close, the ODMVDR filter has a little harmonic distortion. We also conducted the comparison for decreasing harmonic amplitudes as seen in Fig. 8. Here we see a larger difference in performance. For all filter lengths, the oSNR of the ODMVDR filter is greater than that of the HDLCMV filter. However, there is also some

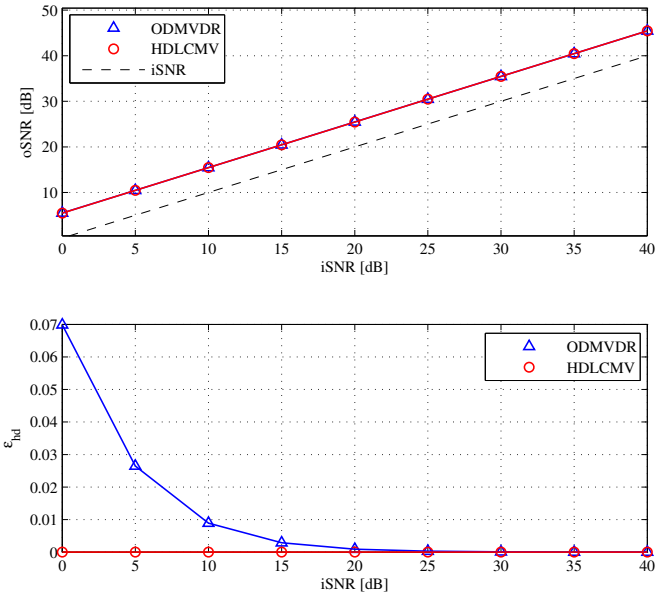


Fig. 5. Performance of the filters for $M = 30$ as a function of the iSNR when the harmonics has unit amplitudes and the noise is white Gaussian.

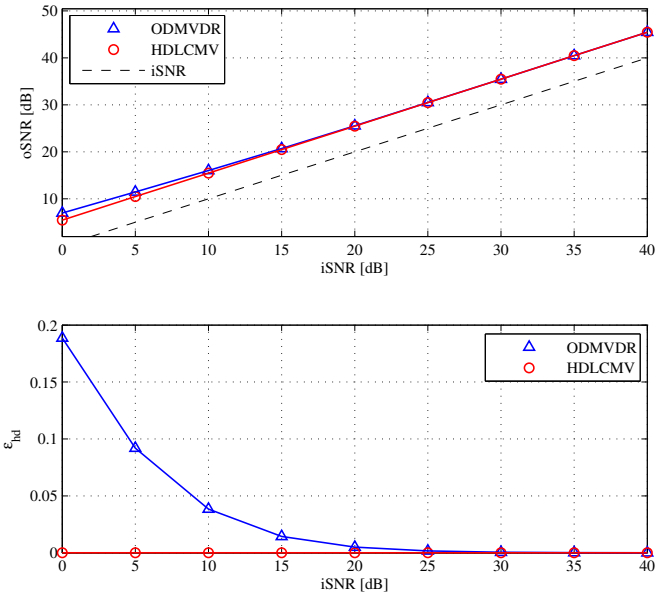


Fig. 6. Performance of the filters for $M = 30$ as a function of the iSNR when the harmonics has decreasing amplitudes and the noise is white Gaussian.

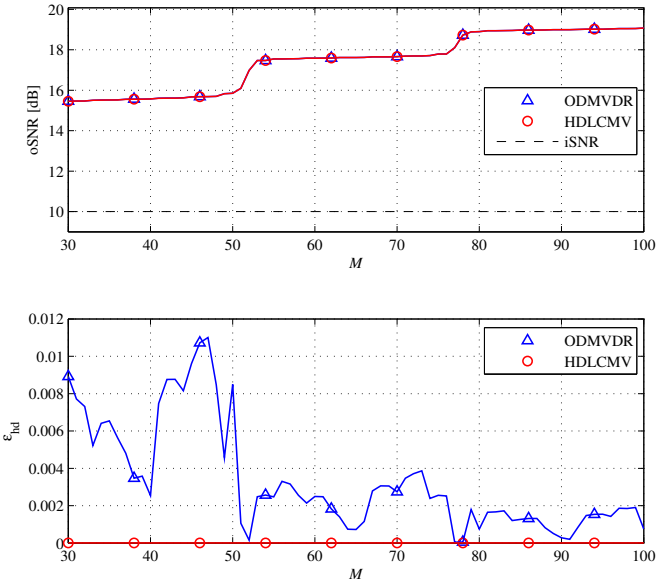


Fig. 7. Performance of the filters as a function of M when the harmonics has unit amplitudes and the noise is white Gaussian.

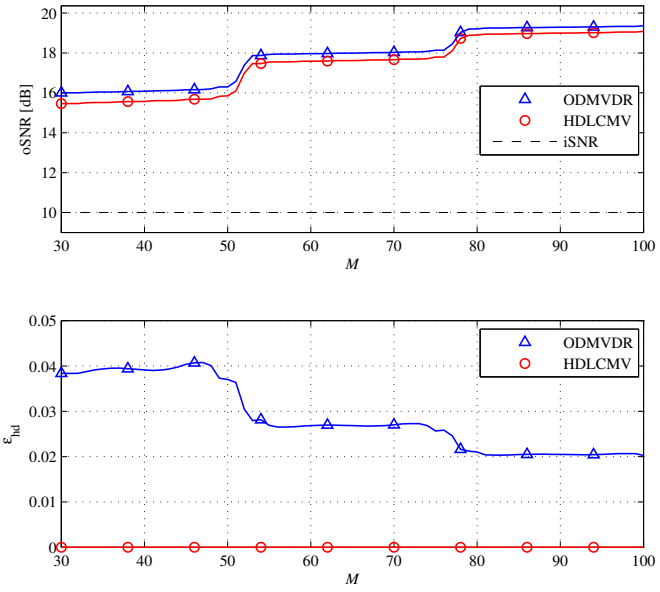


Fig. 8. Performance of the filters as a function of M when the harmonics has decreasing amplitudes and the noise is white Gaussian.

harmonic distortion introduced by the ODMVDR filter. Note that the step-wise increase in the oSNR in Fig. 7 and Fig. 8 is caused by the orthogonality (or the lack thereof) between the harmonics which is evident from (54) when the noise is white Gaussian.

Furthermore, we conducted simulations where the noise was a sum of white Gaussian noise, $v_{\text{wn}}(n)$, and sinusoidal noise, $v_{\text{sn}}(n)$. The variance, $\sigma^2_{v_{\text{sn}}}$, of the sinusoidal noise source was normalized with respect to the variance, σ^2_x, of the desired signal such that they had the same power. White Gaussian noise was also added to the desired signal resulting in the following iSNR

$$\text{iSNR} = \frac{\sigma^2_x}{\sigma^2_{v_{\text{sn}}} + \sigma^2_{v_{\text{wn}}}} . \tag{69}$$

Note that since the sinusoidal noise source has the same variance as the desired signal, the iSNR will always be smaller than or equal to zero (in dB) in these simulations according to the above equation. First, for the sinusoidal noise scenario, we compared the filter performances as a function of the iSNR when the filter order was $M = 50$. The result for unit harmonic amplitudes are given in Fig. 9. The oSNRs of the filters are relatively close, but with the largest difference when the white noise variance, $\sigma^2_{v_{\text{wn}}}$, is largest. For all iSNRs, the ODMVDR filter has more harmonic distortion compared to the scenario

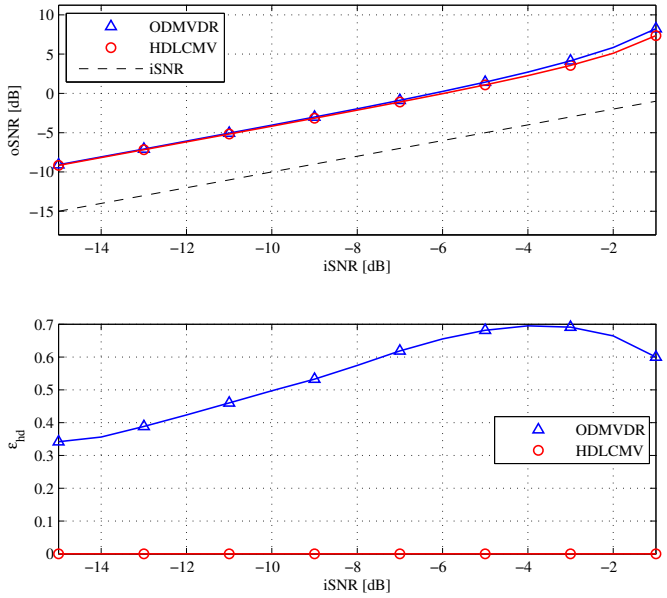


Fig. 9. Performance of the filters for $M = 50$ as a function of the iSNR when the harmonics has unit amplitudes and the noise is a sum of sinusoidal noise and white Gaussian noise.

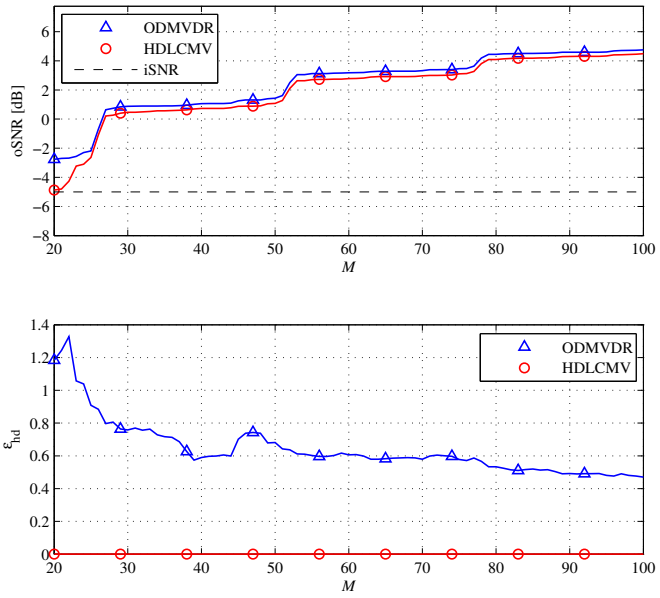


Fig. 11. Performance of the filters as a function of M when the harmonics has unit amplitudes and the noise is a sum of sinusoidal noise and white Gaussian noise.

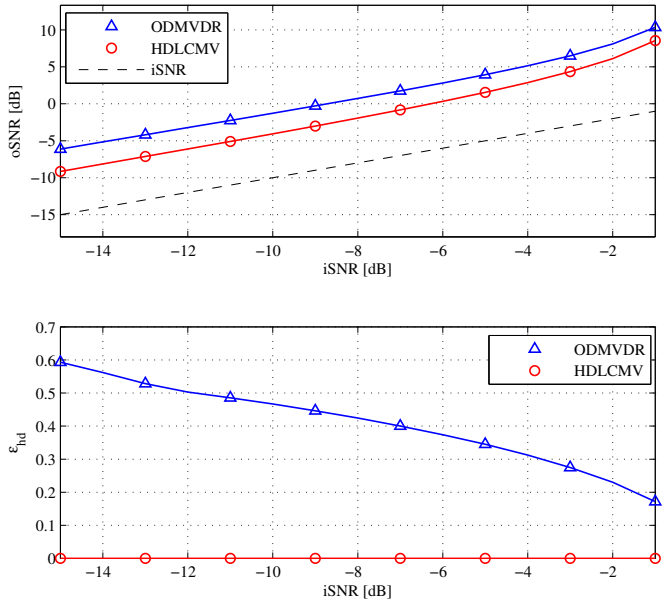


Fig. 10. Performance of the filters for $M = 50$ as a function of the iSNR when the harmonics has decreasing amplitudes and the noise is a sum of sinusoidal noise and white Gaussian noise.

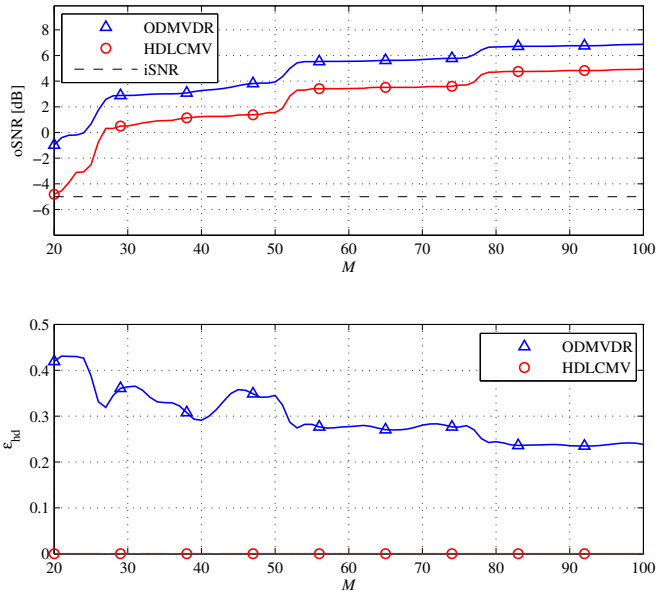


Fig. 12. Performance of the filters as a function of M when the harmonics has decreasing amplitudes and the noise is a sum of sinusoidal noise and white Gaussian noise.

with white Gaussian noise only. When decreasing harmonic amplitudes were considered (see Fig. 10), the difference in oSNRs between the filters was more pronounced with the ODMVDR having the highest oSNR for all iSNRs. The ODMVDR filter, however, also had more harmonic distortion in this case.

In the sinusoidal noise scenario, we also compared the performances as a function of the filter length, and the results are depicted in Fig. 11 and Fig. 12, respectively. As in the previous simulations, we observe that the oSNR of the ODMVDR filter is in general higher than the oSNR of the HDLCMV filter. However, the difference between the filters decreases when M increases. The harmonic distortion of the ODMVDR filter is more significant in this simulation compared to the white Gaussian noise only scenario, but it decreases as we increase M.

Finally, we compared the filter performances as a function of the pitch spacing $\Delta\omega_0$ between the desired signal and the sinusoidal noise source. In this simulation, the filter order was $M = 100$. The results are given in Fig. 13 and Fig. 14, respectively. For both unit and decreasing amplitudes, the oSNRs of the two filters are not much different for all source spacings, but with the ODMVDR having a slightly better

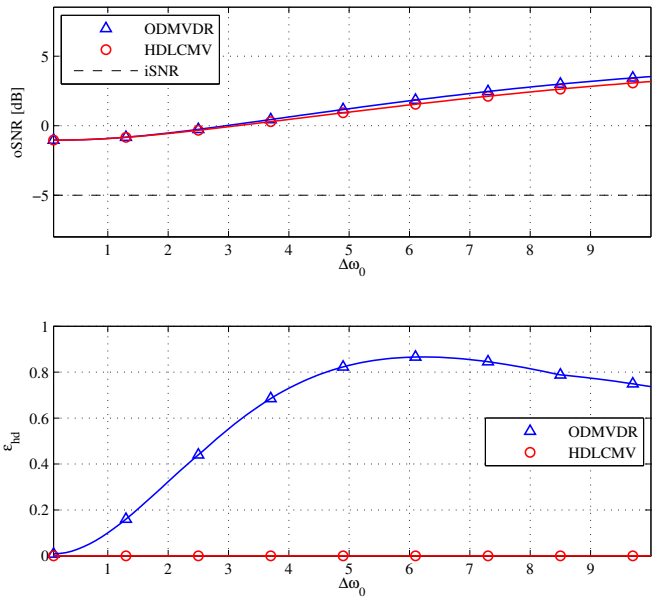


Fig. 13. Performance of the filters for $M = 100$ as a function of the source spacing $\Delta\omega_0$ when the harmonics has unit amplitudes and the noise is a sum of sinusoidal noise and white Gaussian noise.

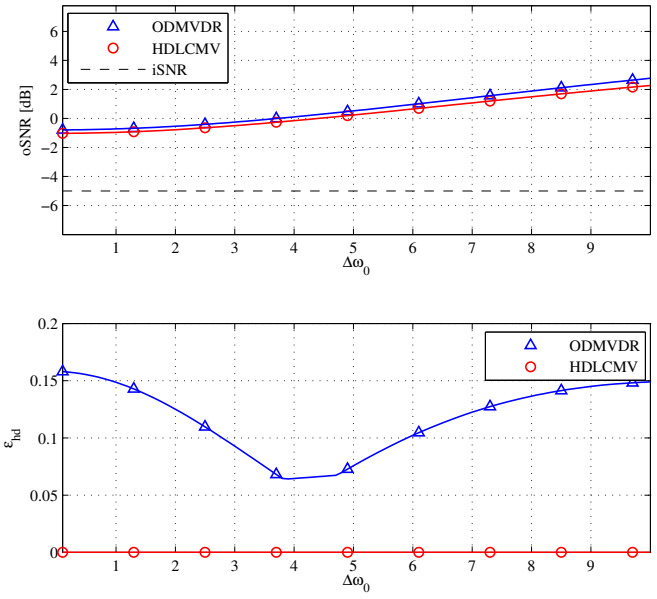


Fig. 14. Performance of the filters for $M = 100$ as a function of the source spacing $\Delta\omega_0$ when the harmonics has decreasing amplitudes and the noise is a sum of sinusoidal noise and white Gaussian noise.

oSNR. Moreover, for both filters the oSNR increases as we increase the spacing of the harmonic sinusoidal sources. We also observe that for both types of amplitudes, the ODMVDR has much harmonic distortion in this case compared to the other simulations.

C. Application Example: Using the ODMVDR and HDLCMV Filters Jointly for Speech Enhancement

In this experimental example, we show how the ODMVDR and HDLCMV can be applied jointly for enhancement of speech signals. For the experiment, we used a 2.2 second long

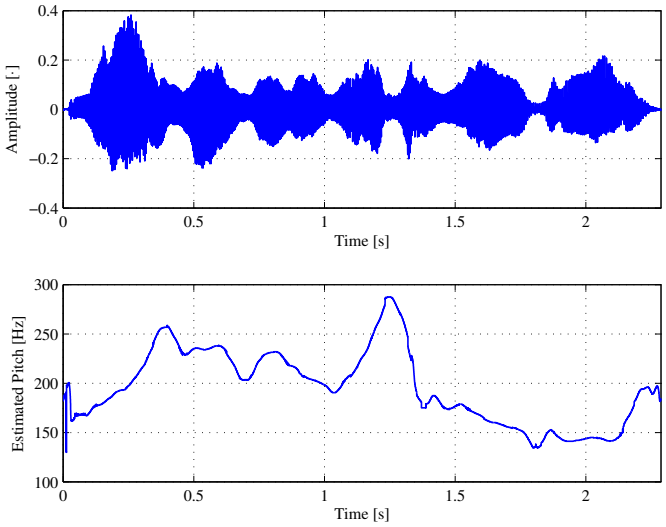


Fig. 15. A plot of a female speech signal (top) and the pitch estimates associated with it (bottom).

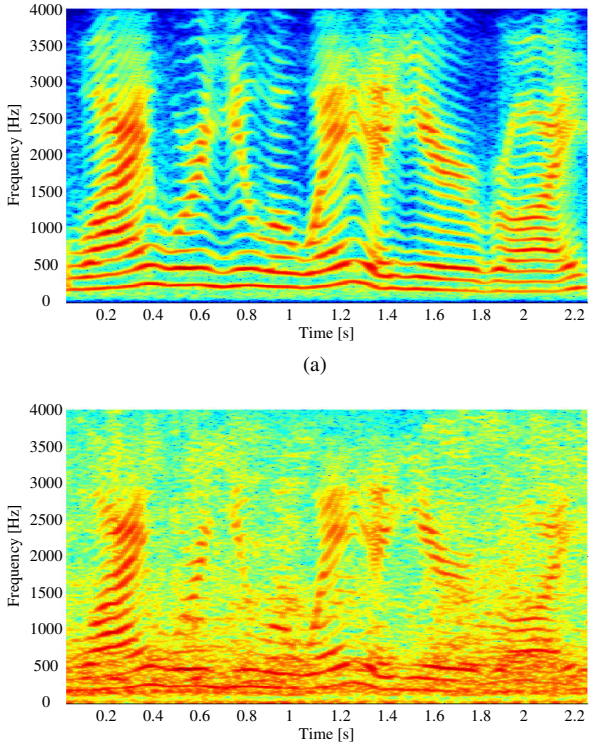


Fig. 16. Spectrograms of (a) the clean speech signal in Fig. 15 and (b) the speech signal in Fig. 15 corrupted by babble noise at an iSNR of 5 dB.

speech segment sampled at 8 kHz. The segment contains a female speaker reading aloud the sentence "Why where you away a year Roy?" and it is plotted in Fig. 15. Since the pitch is needed in the HDLCMV filter design, we estimated the pitch of the speech signal at all time instances using an orthogonality based subspace method [19], [21]. The pitch estimator is available from an online toolbox[1]. The pitch track resulting from the pitch estimation is also depicted in Fig. 15,

[1]http://www.morganclaypool.com/page/multi-pitch

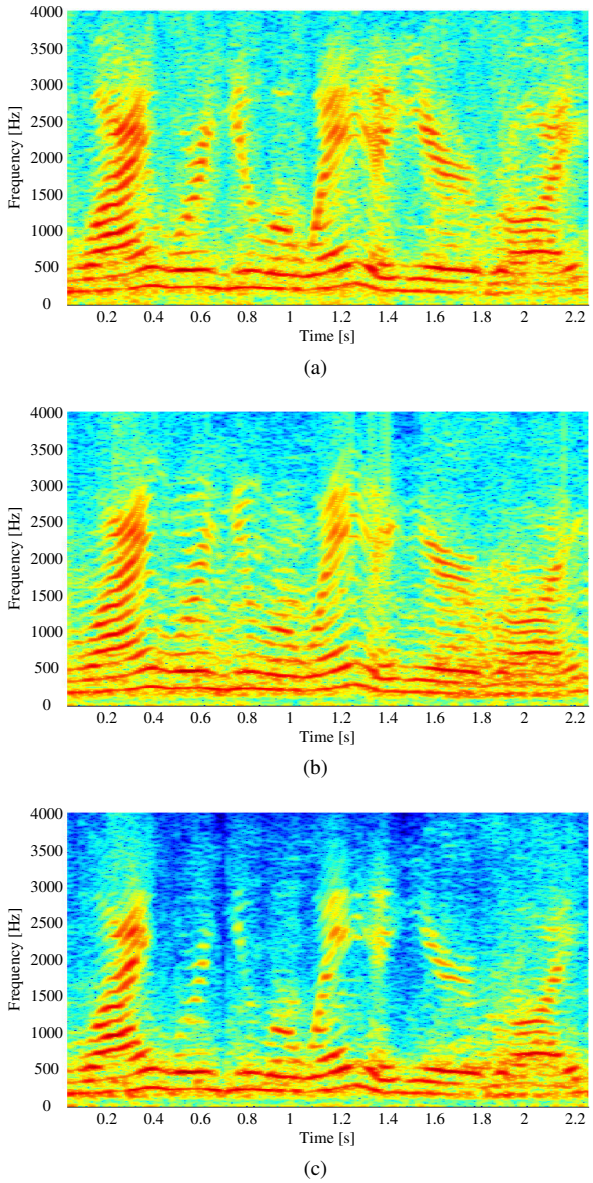


Fig. 17. Spectrograms of enhanced versions of the noisy signal in Fig. 16b. The enhanced signals are obtained using (a) the ODMVDR filter only, (b) the HDLCMV filter only, and (c) the joint HDLCMV and ODMVDR filtering setup, respectively.

and it is used for later filter designs. Note that since we focus on speech enhancement rather than pitch estimation in this paper, we estimated the pitch directly from the clean speech signal, $x(n)$. The spectrogram of the speech signal, $x(n)$, is shown in Fig. 16a.

First, we consider a scenario in which the speech signal is corrupted by babble noise at an average iSNR of 5 dB. The babble noise was taken from the AURORA database [28]. The spectrogram of the noisy signal is depicted in Fig. 16b. We then enhanced the noisy signal using three different filtering setups, i.e., using the ODMVDR filter only, using the HDLCMV filter only, and using the ODMVDR and HDLCMV filters jointly. The joint filtering method is proposed since using only either the ODMVDR or the HDLCMV filter has drawbacks. For example, the ODMVDR method is sensitive

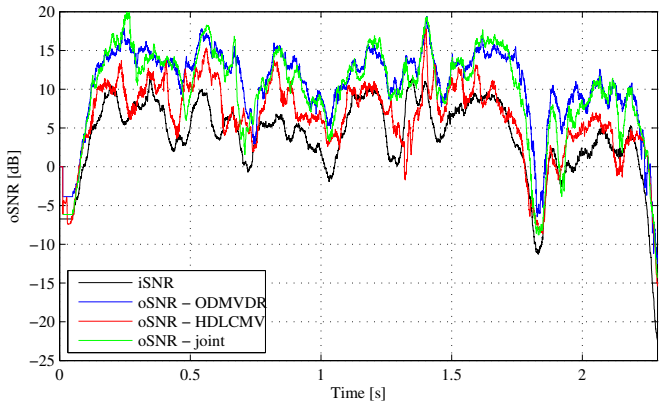


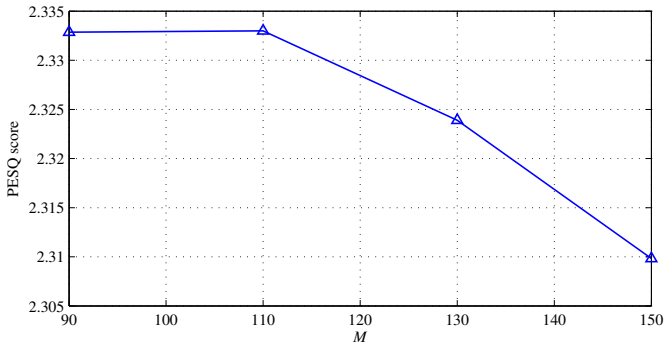
Fig. 18. The estimated iSNR and oSNRs over time for the enhanced signals in Fig. 17.

to non-stationary noise, since it requires that knowledge about the noise statistics which we do not always have access to in practice. This is not an issue for the HDLCMV filter, but, on the other hand, it will introduce some distortion of speech signals because the harmonic model does not hold exactly. Furthermore, the HDLCMV filter has, in general, more constraints than the ODMVDR filter, and it will therefore most likely have a lower oSNR compared to the ODMVDR filter. The joint use of the filters can be justified by their close relationship described in Section V. In the joint filtering scheme, we first use the HDLCMV filter to obtain a rough estimate of the speech signal. The rough speech estimate is then subtracted from the observed signal to obtain an estimate of the noise signal. We estimate the noise statistics from the estimated noise signal, and the noise statistics are used for designing the ODMVDR filter. Finally, the ODMVDR filter is applied for enhancement of the observed signal. By using the ODMVDR filter for the enhancement rather than the HDLCMV filter, we expect to remove some of the distortion introduced by the HDLCMV filter in practice. Moreover, we expect to obtain more noise reduction, since the ODMVDR filter is less constrained compared to the HDLCMV filter.

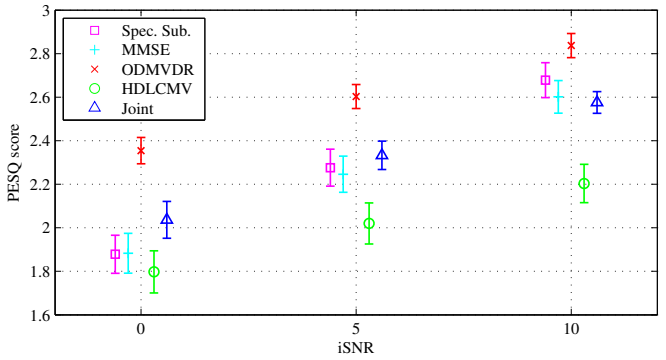
In all the filtering setups, the filters were updated for each time instance. The update was conducted by recalculating the filters from the signal and noise statistics ($\hat{\mathbf{R}}_{\mathbf{y}}$ and $\hat{\mathbf{R}}_{\mathbf{v}}$) estimated from the previous 400 samples (≈ 50 ms). Both $\hat{\mathbf{R}}_{\mathbf{y}}$ and $\hat{\mathbf{R}}_{\mathbf{v}}$ were used to calculate the ODMVDR filter. That is, we assumed that the noise signal was available in this simulation, albeit it is not the case in practice. The HDLCMV filter was updated using $\hat{\mathbf{R}}_{\mathbf{y}}$, the pitch estimates in Fig. 15, and a model order of $L = 13$. The model order was chosen by inspecting the spectrogram in Fig. 16a since we do not consider model order estimation in this paper. Furthermore, in the calculations of the HDLCMV filter and the filters in the joint filtering setup, we regularized the covariance matrix using [29]

$$\hat{\mathbf{R}}_{\mathbf{y},\text{reg}} = (1-\gamma)\hat{\mathbf{R}}_{\mathbf{y}} + \gamma\frac{\text{Tr}\left\{\hat{\mathbf{R}}_{\mathbf{y}}\right\}}{M}\mathbf{I}\,, \tag{70}$$

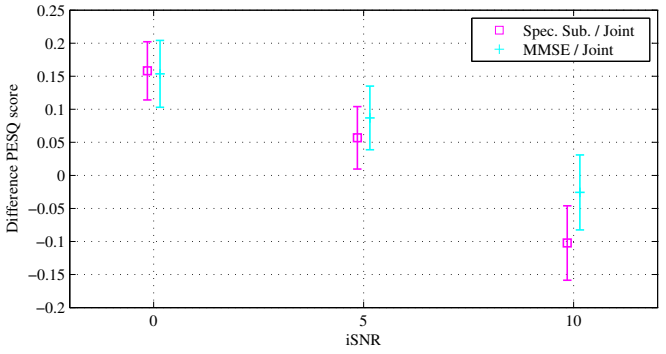
where $\text{Tr}\{\cdot\}$ denotes the trace operator. The regularization is used to compensate for, e.g., numerical stability, model



(a)



(b)



(c)

Fig. 19. Average PESQ scores (a) for the joint filtering scheme as a function of M for an iSNR of 5 dB, and (b) for several enhancement methods as a function of the iSNR for $M = 110$ with 95% confidence intervals. In (c), the average differences in PESQ scores between the joint filtering scheme and the spectral subtraction and MMSE-based methods, respectively, are plotted with 95% confidence intervals.

mismatch, and noisy statistics. Choosing $\gamma = 0.7$ was found to give the best results in terms of oSNR and perceptual scores. All filters were chosen to be of order $M = 100$.

The observed signal containing the speech signal and babble noise was then enhanced using the three filtering setups, and the spectrograms of the resulting enhanced signals are shown in Fig. 17. The spectrograms indicate that the joint filtering method has better noise reduction abilities than when using either the ODMVDR or the HDLCMV filter only. Regarding distortion, the ODMVDR filter seems to outperform the joint filtering method. However, it is important to remember that the ODMVDR filter was designed using the noise signal, and it will therefore most likely have a worse performance in practice. To confirm the observations on the performances of the filters, we also measured the oSNRs associated with the enhanced signals in Fig. 17 using

$$\text{oSNR}(\mathbf{h}) = \frac{\sigma_{x_{\text{fl}}}^2}{\sigma_{v_{\text{rn}}}} = \frac{\mathbf{h}^T\mathbf{R}_{\mathbf{x}}\mathbf{h}}{\mathbf{h}^T\mathbf{R}_{\mathbf{v}}\mathbf{h}} \ . \tag{71}$$

Note that we here use the traditional oSNR measure, since, in practice, the interference term of the ODMVDR approach is relatively large which complicates the comparison of the oSNR measures in (51) and (53), respectively. The measured oSNRs are shown in Fig. 18. These measurements show that both the ODMVDR and the joint filtering methods outperform the HDLCMV filtering method in terms of noise reduction. The ODMVDR and joint filtering methods have comparable noise reduction performance even though the joint filtering method is implemented without access to the noise signal directly. This justifies the use of the joint filtering method in practice as it is more tractable than the ODMVDR filtering method when the noise signal is not available.

The oSNR measure, however, does not quantify how much the filtering methods distort the desired signal. Therefore, we also evaluated the filtering methods in terms of "Perceptual Evaluation of Speech Quality" (PESQ) scores [30]. The PESQ score is an objective measure which reflects the perceptual quality of a speech signal. That is, the PESQ scores give a more complete picture of the performance of the filtering methods since the perceptual quality is affected both by noise reduction and distortion. We compared the PESQ scores of noisy speech signal enhanced using the joint filtering method, the ODMVDR filtering method, the HDLCMV filtering method, a spectral subtraction based method [31], and a method using MMSE estimates of the spectral amplitudes [32]. Note that, in these simulations, we design the ODMVDR filter from the true noise signal, and it therefore only serves as a bound to the proposed joint filtering scheme.

Followingly, we describe how the different enhancement methods were set up for the PESQ score evaluations. In all of the filtering methods, i.e., the joint method, the ODMVDR method, and the HDLCMV method, the observed signal and noise statistics were calculated as in the previous experiment. The noise statistics were calculated directly from the noise signal, and they were only used for designing the ODMVDR filter. In the joint and HDLCMV filtering methods, the observed signal statistics were regularized as in the previous experiment. The model order was set to $L = \min([15, \lfloor \pi/\omega_0 \rfloor - 1])$ at each time instance when designing the HDLCMV filters. The speech signals used in these evaluations contained both voiced and unvoiced speech segments. However, the HDLCMV filter used in both the joint and HDLCMV filtering methods are designed for voiced speech segments only. Therefore, we updated the HDLCMV filter in these evaluations as follows; for voiced speech segments, the HDLCMV filter was designed as in (36), and for unvoiced speech segments, the filter was updated as

$$\mathbf{h}(n) = (1-\gamma)\mathbf{0} + \gamma\mathbf{h}(n-1) \ , \tag{72}$$

when $\|\mathbf{h}(n-1)\|_2 > 0.1$ with $\gamma = 0.95$ and $\mathbf{0}$ is a vector of zeros. The norm conditional update was introduced to avoid abrupt changes when transitioning between unvoiced/no

speech and voiced speech. Both the spectral subtraction and the MMSE-based methods are available in the VOICEBOX toolbox[2] for MATLAB, in which they are implemented using noise power spectral density estimates based on optimal smoothing and minimum statistics [33]. We used the default settings given by the VOICEBOX toolbox for the spectral subtractions and MMSE methods.

For the PESQ score evaluations of the aforementioned enhancement methods, we used two female and two male speech excerpts each of length 4-6 seconds taken from the Keele database [34]. Since pitch estimation is not the main topic of this paper, we used the pitch estimates of the voiced parts of the speech excerpts from the Keele database for the design of the HDLCMV filters. Moreover, the pitch estimates in the Keele database are 0 when the speech is unvoiced or no voice is present. We exploited this to distinguish between voiced and unvoiced speech since the unvoiced/voiced speech detection problem is not considered here. The chosen speech excerpts were then buried in white Gaussian noise, car noise, babble noise, exhibition hall noise, and street noise. All noise sources except the white noise were taken from the AURORA database [28]. First, we applied the proposed joint filtering method on all four speech excerpts in all five noise scenarios for different filtering lengths when the iSNR was 5 dB. The PESQ scores averaged across the different noisy speech excerpts are shown in Fig. 19a. We can see that the perceptual performance of the proposed joint filtering method peaks around $M = 110$. We then applied all of the enhancement methods of the comparison on all the speech excerpts in all of the different noise scenarios for different iSNRs. For these simulations, the filter length of the filtering-based enhancements methods was set to 110, and the PESQ results averaged over the different speech excerpts and noise scenarios are shown in Fig. 19b with 95% confidence intervals. From these results, it seems that the joint filtering method outperforms the spectral subtraction and MMSE-based methods on average for relative low iSNRs (≤ 5 dB) and vice versa for a higher iSNR (10 dB). However, from these results, we cannot say this with 95% confidence due to overlapping confidence intervals, but it does not preclude that the observations are statistically significant since we can also consider the difference in PESQ scores. To investigate this further, we measured the average of the difference in PESQ scores between the proposed joint filtering scheme and the spectral subtraction and MMSE-based methods, respectively; the results from this investigation is plotted in 19c with 95% confidence intervals. From these results, we can conclude with 95% confidence that the proposed joint filtering method outperforms the spectral subtraction and MMSE-based methods on average for iSNRs of 0 dB and 5 dB in terms of PESQ scores since the confidence intervals do not include 0. In practice, it is expected that the proposed joint filtering method only outperforms the other methods for relatively low iSNRs since the harmonic model assumption embedded in the proposed joint filtering design introduces a small amount of distortion due to model mismatch.

[2]http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

VIII. Conclusion

In this paper, we considered two recent filter designs for speech enhancement, namely the ODMVDR and HDLCMV filters. The ODMVDR filter is not explicitly dependent of the desired signal since it is calculated from the observed signal and noise statistics. This makes it a general filtering method which is appropriate for enhancement of all types of speech (e.g., both voiced and unvoiced). However, the ODMVDR filter is vulnerable to non-stationary noise since the noise statistics are typically estimated during periods of silence. On the other hand, the HDLCMV filter is signal-dependent since it is designed using the observed signal and the desired signal statistics. In this filter, a harmonic model is assumed which enables the estimation of the signal statistics if the pitch and the number of harmonics are known. While this filter is robust against non-stationary noise, it will only be appropriate for voiced speech due to the harmonic model assumption. Since both filters have complementary advantages and disadvantages, we investigated the relationship between them in this paper. Our theoretical studies confirmed that the filters are indeed closely related. We also proposed some performance measures for both filters which are available in closed-form when the desired signal is periodic. We compared the performance measures in theoretical simulations. From these simulations, it was again clear that the methods are closely related, but each filter had its own advantages. For example, the ODMVDR filter has, in general, a slightly higher oSNR than the HDLCMV while the HDLCMV filter does not distort the harmonics as opposed to the ODMVDR filter. The close relationship between the filters inspired us to propose a filtering scheme where the ODMVDR and HDLCMV filters are used jointly. This scheme was applied on real speech signals in different noise scenarios. The results of these experiments showed that, for relatively low iSNRs (i.e., < 10 dB) , the joint filtering scheme outperforms some existing enhancement techniques in terms of average PESQ scores with 95% confidence.

Appendix
On Rewriting the HDLCMV Filter in Terms of the Observed Signal Covariance Matrix

In this appendix, we show that it makes no difference whether we use the noise covariance matrix, $\mathbf{R_v}$, or use the observed signal covariance matrix, $\mathbf{R_y}$, in (35). First, recall that the HDLCMV filter is given by

$$\mathbf{h}_{\text{HDLCMV}} = \mathbf{R_v}^{-1}\mathbf{Z}\left(\mathbf{Z}^H\mathbf{R_v}^{-1}\mathbf{Z}\right)^{-1}\mathbf{1} \ . \tag{73}$$

Note that in the following derivations we denote the HDLCMV filter as $\mathbf{h}$. If we use the covariance matrix model on $\mathbf{R_y}$, the noise covariance matrix can also be written as [24]

$$\mathbf{R_v} = \mathbf{R_y} - \mathbf{ZPZ}^H \ . \tag{74}$$

If we substitute (74) back into (73), we get that

$$\begin{aligned}\mathbf{h} &= \left(\mathbf{R_y} - \mathbf{ZPZ}^H\right)^{-1}\mathbf{Z}\left[\mathbf{Z}^H\left(\mathbf{R_y} - \mathbf{ZPZ}^H\right)^{-1}\mathbf{Z}\right]^{-1}\mathbf{1} \\ &= \mathbf{AZB1} \ ,\end{aligned} \tag{75}$$

where

$$\mathbf{A} = \left(\mathbf{R}_{\mathbf{y}} - \mathbf{Z}\mathbf{P}\mathbf{Z}^H\right)^{-1} , \tag{76}$$

$$\begin{aligned}\mathbf{B} &= \left[\mathbf{Z}^H\left(\mathbf{R}_{\mathbf{y}} - \mathbf{Z}\mathbf{P}\mathbf{Z}^H\right)^{-1}\mathbf{Z}\right]^{-1} \\ &= \left(\mathbf{Z}^H\mathbf{A}\mathbf{Z}\right)^{-1} .\end{aligned} \tag{77}$$

Applying the matrix inversion lemma on $\mathbf{A}$ yields

$$\mathbf{A} = \mathbf{R}_{\mathbf{y}}^{-1} + \mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\left(\mathbf{P}^{-1} - \mathbf{Z}^H\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}^H\mathbf{R}_{\mathbf{y}}^{-1} . \tag{78}$$

If we insert this expression for $\mathbf{A}$ back into (77), we get

$$\mathbf{B} = \left(\mathbf{Z}^H\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\right)^{-1} - \mathbf{P} . \tag{79}$$

We can then rewrite the HDLCMV filter expression by inserting (78) and (79) into (75) which yields

$$\begin{aligned}\mathbf{h} &= \mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\left(\mathbf{Z}^H\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\right)^{-1}\mathbf{1} - \mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\mathbf{P}\mathbf{1} \\ &+\mathbf{R}_{\mathbf{y}}^{-1}\left(\mathbf{P}^{-1} - \mathbf{Z}^H\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\right)^{-1}\mathbf{1} \\ &-\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\left(\mathbf{P}^{-1} - \mathbf{Z}^H\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}^H\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\mathbf{P}\mathbf{1} .\end{aligned} \tag{80}$$

After some algebra, it turns out that the somewhat complex expression for the filter in (80) can be reduced to

$$\mathbf{h} = \mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\left(\mathbf{Z}^H\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{Z}\right)^{-1}\mathbf{1} . \tag{81}$$

That is, there is no difference between using the noise covariance matrix, $\mathbf{R}_{\mathbf{v}}$, and the observed signal covariance matrix, $\mathbf{R}_{\mathbf{y}}$, in (73).

REFERENCES

[1] J. Benesty, S. Makino, and J. Chen, *Speech Enhancement*, ser. Signals and Communication Technology. Springer, 2005.

[2] P. Loizou, *Speech Enhancement: Theory and Practice*. CRC Press, 2007.

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

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

[5] 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, 1985.

[6] M. Dendrinos, S. Bakamidis, and G. Carayannis, "Speech enhancement from noise: A regenerative approach," *Speech Communication*, vol. 10, no. 1, pp. 45 – 57, 1991.

[7] 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.

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

[9] J. S. Lim, Ed., *Speech Enhancement*. Prentice-Hall, 1983.

[10] J. Chen, J. Benesty, and Y. Huang, "Study of the noise-reduction problem in the Karhunen-Loève expansion domain," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 17, no. 4, pp. 787–802, 2009.

[11] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*. M.I.T. Press, 1949.

[12] 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, Dec. 2010.

[13] H. Li, P. Stoica, and J. Li, "Computationally efficient parameter estimation for harmonic sinusoidal signals," *Elsevier Signal Process.*, vol. 80(9), pp. 1937–1944, 2000.

[14] K. W. Chan and H. C. So, "Accurate frequency estimation for real harmonic sinusoids," *IEEE Signal Process. Lett.*, vol. 11, no. 7, pp. 609–612, 2004.

[15] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1917–1930, 2002.

[16] V. Emiya, B. David, and R. Badeau, "A parametric method for pitch estimation of piano tones," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, 2007, pp. 249–252.

[17] S. Godsill and M. Davy, "Bayesian harmonic models for musical pitch estimation and analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 2, 13–17 May 2002, pp. 1769–1772.

[18] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, 2004.

[19] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Joint high-resolution fundamental frequency and order estimation," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 15, no. 5, pp. 1635–1644, 2007.

[20] M. G. Christensen, P. Stoica, A. Jakobsson, and S. H. Jensen, "Multi-pitch estimation," *Elsevier Signal Process.*, vol. 88(4), pp. 972–983, 2008.

[21] M. G. Christensen and A. Jakobsson, "Multi-pitch estimation," *Synthesis Lectures on Speech and Audio Processing*, vol. 5, no. 1, pp. 1–160, 2009.

[22] M. G. Christensen, J. L. Højvang, A. Jakobsson, and S. H. Jensen, "Joint fundamental frequency and order estimation using optimal filtering," *EURASIP J. on Advances in Signal Processing*, vol. 2011, no. 1, p. 13, 2011.

[23] J. Benesty and J. Chen, *Optimal Time-Domain Noise Reduction Filters – A Theoretical Study*, 1st ed., ser. SpringerBriefs in Electrical and Computer Engineering. Springer, 2011, no. VII.

[24] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Pearson Education, Inc., 2005.

[25] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[26] ——, *Nonlinear Methods of Spectral Analysis*. Springer-Verlag, 1983, ch. Maximum-Likelihood Spectral Estimation.

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

[28] 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.

[29] F. van der Heijden, R. P. W. Duin, D. de Ridder, and D. M. J. Tax, *Classification, Parameter Estimation and State Estimation - An Engineering Approach using MATLAB®*. John Wiley & Sons Ltd, 2004.

[30] "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.

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

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

[33] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, 2001.

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

Jesper Rindom Jensen (S'09) was born in Ringkøbing, Denmark in August 1984. He received the B.Sc. degree in 2007 from Aalborg University in Denmark. In the fall 2007, he was enrolled in the elite candidate programme in wireless communications at Aalborg University, and he received the M.Sc. degree cum laude for completing the elite candidate education at Aalborg University in 2009.

Currently, he is a Ph.D. student at the Department of Electronic Systems at Aalborg University in Denmark. He has been a Visiting Researcher at University of Quebec, INRS-EMT, in Montreal, Quebec, Canada. He has published several conference papers and some journal papers on the topics of parameter estimation, coding, and enhancement of single- and multichannel audio and speech signals. Among others, his research interests are digital signal processing theory, parametric analysis, modeling and coding of audio and speech signals, and microphone array processing.

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. 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 and Gaensler conceived and designed the world-first PC-based multi-party hands-free full-duplex stereo conferencing system over IP networks.

He is the editor of the book series: "Springer Topics in Signal Processing. 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 was a member of the IEEE Signal Processing Society Technical Committee on Audio and Electroacoustics and a member of the editorial board of the EURASIP Journal on Applied Signal Processing. 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 published with Chen. He has co-authored and co-edited/co-authored many books in the area of acoustic signal processing. He is also the lead editor-in-chief of the reference "Springer Handbook of Speech Processing (Berlin: Springer-Verlag, 2007).

Mads Græsbøll Christensen (S'00–M'05–SM'11) was born in Copenhagen, Denmark, in March 1977. He received the M.Sc. and Ph.D. degrees from Aalborg University, Denmark, in 2002 and 2005, respectively.

He was formerly with the Department of Electronic Systems, Aalborg University, and is currently an Associate Professor in the Department of Architecture, Design and Media Technology. He has been a Visiting Researcher at Philips Research Labs, Ecole Nationale Supérieure des Télécommunications (ENST), University of California, Santa Barbara (UCSB), and Columbia University. He has published about 100 papers in peer-reviewed conference proceedings and journals and is coauthor (with A. Jakobsson) of the book *Multi-Pitch Estimation* (Morgan & Claypool Publishers, 2009). His research interests include digital signal processing theory and methods with application to speech and audio, in particular parametric analysis, modeling, enhancement, separation, and coding.

Dr. Christensen has received several awards and prestigious grants, including an ICASSP Student Paper Award, the Spar Nord Foundation's Research Prize for his Ph.D. thesis, a Danish Independent Research Council postdoc grant and Young Researcher's Award, and a Villum Foundation Young Investigator Programme grant. He is an Associate Editor for the IEEE Signal Processing Letters.

Søren Holdt Jensen (S'87–M'88–SM'00) received the M.Sc. degree in electrical engineering from Aalborg University, Aalborg, Denmark, in 1988, and the Ph.D. degree in signal processing from the Technical University of Denmark, Lyngby, Denmark, in 1995. Before joining the Department of Electronic Systems of Aalborg University, he was with the Telecommunications Laboratory of Telecom Denmark, Ltd, Copenhagen, Denmark; the Electronics Institute of the Technical University of Denmark; the Scientific Computing Group of Danish Computing Center for Research and Education (UNI•C), Lyngby; the Electrical Engineering Department of Katholieke Universiteit Leuven, Leuven, Belgium; and the Center for PersonKommunikation (CPK) of Aalborg University. He is Full Professor and is currently heading a research team working in the area of numerical algorithms, optimization, and signal processing for speech and audio processing, image and video processing, multimedia technologies, and digital communications. Prof. Jensen was an Associate Editor for the IEEE Transactions on Signal Processing and Elsevier Signal Processing, and is currently Associate Editor for the IEEE Transactions on Audio, Speech and Language Processing and EURASIP Journal on Advances in Signal Processing. He is a recipient of an European Community Marie Curie Fellowship, former Chairman of the IEEE Denmark Section, and Founder and Chairman of the IEEE Denmark Sections Signal Processing Chapter. He is member of the Danish Academy of Technical Sciences and was in January 2011 appointed as member of the Danish Council for Independent Research—Technology and Production Sciences by the Danish Minister for Science, Technology and Innovation.