Optimal Filter Designs for Separating and Enhancing Periodic Signals

Christensen, Mads Græsbøll; Jakobsson, Andreas

Published in:
I E E E Transactions on Signal Processing

DOI (link to publication from Publisher):
10.1109/TSP.2010.2070497

Publication date:
2010

Document Version
Accepted author manuscript, peer reviewed version

Optimal Filter Designs for Separating and Enhancing Periodic Signals

Mads Græsbøll Christensen* and Andreas Jakobsson

***Abstract*— In this paper, we consider the problem of separating and enhancing periodic signals from single-channel noisy mixtures. More specifically, the problem of designing filters for such tasks is treated. We propose a number of novel filter designs that 1) are specifically aimed at periodic signals, 2) are optimal given the observed signal and thus signal-adaptive, 3) offer full parametrizations of periodic signals, and 4) reduce to well-known designs in special cases. The found filters can be used for a multitude of applications including processing of speech and audio signals. Some illustrative signal examples demonstrating its superior properties as compared to other related filters are given and the properties of the various designs are analyzed using synthetic signals in Monte Carlo simulations.**

I. Introduction

Many natural signals that are of interest to mankind are periodic by nature or approximately so. In mathematics and engineering sciences, such periodic signals are often described by Fourier series, i.e., a sum of sinusoids, each described by an amplitude and a phase, having frequencies that are integer multiples of a fundamental frequency. In mathematical descriptions of periodic functions, the period which is inversely proportional to the fundamental frequency is assumed to be known and the function is observed over a single period over which the sinusoids form an orthogonal basis. When periodic signals are observed over arbitrary intervals, generally have unknown fundamental frequencies, and are corrupted by some form of observation noise, the problem of parametrizing the signals is a different and much more difficult one. The problem of estimating the fundamental frequency from such an observed signal is referred to as fundamental frequency or pitch estimation. Additionally, some signals contain many such periodic signals, in which case the problem is referred to as multi-pitch estimation. Strictly speaking, the word pitch originates in the perception of acoustical signals and is defined as "that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale" [1], but since this attribute in most cases is the same as the fundamental frequency of a Fourier series, these terms are often used synonymously. Some pathological examples do exist, however, where it is not quite that simple. The pitch estimation problem has received much attention in the fields of speech and audio processing, not just because it is an interesting and challenging problem, but also because it is the key, or, perhaps more correctly, a key to many fundamental problems such as separation of periodic sources [2], enhancement, and compression of periodic sources [3] as Fourier series constitute naturally compact descriptions of such signals. A fundamental problem in signal processing is the source separation problem, as many other problems are trivially, or at least more easily, solved once a complicated mixture has been broken into its basic parts (for examples of this, see [4], [5]). We remark that for periodic signals, this problem is different from that of blind source separation, as assumptions have been made as to the nature of the sources (for an overview of classical methods for blind source separation, see, e.g., [6], [7]). For periodic signals, once the fundamental frequencies of the periodic sources have been found, it is comparably easy to estimate either the individual periodic signals directly [8]–[11] or their remaining unknown parameters, i.e., the amplitudes, using methods like those in [12]. With amplitudes and the fundamental frequency found, the signal parametrization is complete. Some representative methodologies that have been employed in fundamental frequency estimators are: linear prediction [13], correlation [14], subspace methods [15]–[17], harmonic fitting [18], maximum likelihood [19], [20], cepstral methods [21], Bayesian estimation [22]–[24], and comb filtering [8], [25], [26]. Several of these methodologies can be interpreted in several ways and one should therefore not read too much into this rather arbitrary grouping of methods. For an overview of pitch estimation methods and their relation to source separation, we refer the interested reader to [27]. It should also be noted that separation based on parametric models of the sources is closely related to source separation using sparse decompositions (for an example of such an approach, see [28]).

The scope of this paper is filtering methods with application to periodic signals in noise. We propose a number of novel filter design methods, which are aimed specifically at the processing of noisy observations of periodic signals or from single-channel mixtures of periodic signals. These filter design methods result in filters that are optimal given the observed signal, i.e., they are signal-adaptive, and contain as special cases several well-known designs. The proposed filter designs are inspired by the principle used in the Amplitude and Phase EStimation (APES) method [29], [30], a method which is well-known to have several advantages over the Capon-based estimators. The obtained filters can be used for a number of tasks involving periodic signals, including separation, enhancement, and parameter estimation. In other words, the filtering approaches proposed herein provide full

Part of this work was presented at the 43rd Annual Asilomar Conference on Signals, Systems, and Computers.

M. G. Christensen is with the Dept. of Architecture, Design and Media Technology, Aalborg University, Niels Jernes Vej 14, DK-9220 Aalborg, Denmark (phone: 99 40 97 93, email: mgc@imi.aau.)

A. Jakobsson is with the Dept. of Mathematical Statistics, Lund University, SE-221 00 Lund, Sweden (phone: +46 46 222 4520, email: andreas.jakobsson@ieee.org).

parametrizations of periodic signals through the use of filters. We will, however, focus on the application of such filters to extraction, separation, and enhancement of periodic signals. A desirable feature of the filters is that they do not require prior knowledge of the noise or interfering source but are able to automatically reject these.

The paper is organized as follows. In Section II, we introduce the fundamentals and proceed to derive the initial design methodology leading to single filter that is optimal given the observed signal in Section III. We then derive an alternative design using a filterbank in Section IV, after which, in Section V, we first illustrate the properties of the proposed design and compare the resulting filters to those obtained using previously published methods. Moreover, we demonstrate its application for the extraction of real quasi-periodic signals from mixtures of interfering periodic signals and noise, i.e., for separation and enhancement. Finally, we conclude on the work in Section VI.

II. Fundamentals

We define a model of a signal containing a single periodic component, termed a source, consisting of a weighted sum of complex sinusoids having frequencies that are integer multiples of a fundamental frequency[1] ω_k, and additive noise. Such a signal can, for $n = 0, \ldots, N-1$, be written as

$$x_k(n) = \sum_{l=1}^{L_k} a_{k,l} e^{j\omega_k l n} + e_k(n) \tag{1}$$

where $a_{k,l} = A_{k,l} e^{j\phi_{k,l}}$ is the complex amplitude of the lth harmonic of the source (indexed by k) and $e_k(n)$ is the noise which is assumed to be zero-mean and complex. The complex amplitude is composed of a real, non-zero amplitude $A_{k,l} > 0$ and a phase $\phi_{k,l}$ distributed uniformly on the interval $\{-\pi, \pi]$. The number of sinusoids, L_k, is referred to as the order of the model and is often considered known in the literature. We note that this assumption is generally not consistent with the behavior of speech and audio signals, where the number of harmonics can be observed to vary over time. In most recordings of music, the observed signal consists of many periodic signals, in which case the signal model is

$$x(n) = \sum_{k=1}^{K} x_k(n) = \sum_{k=1}^{K} \sum_{l=1}^{L_k} a_{k,l} e^{j\omega_k l n} + e(n). \tag{2}$$

Note that all noise sources $e_k(n)$ are here modeled by a single noise source $e(n)$. We refer to signals of the form (2) as multi-pitch signals and the model as the multi-pitch model. Even if a recording is only of a single instrument, the signal may be multi-pitch as only some instruments are monophonic. Even in that case, room reverberation may cause the observed signal to consist of several different tones at a particular time, i.e., the signal is effectively a multi-pitch signal.

The algorithms under consideration operate on vectors consisting of M time-reversed samples of the observed signal, defined as $\mathbf{x}(n) = [\, x(n)\ x(n-1)\ \cdots\ x(n-M+1)\,]^T$, where $M \leq N$ and $(\cdot)^T$ denotes the transpose, and similarly for the sources $x_k(n)$ and the noise $e(n)$. Defining the filter output $y_k(n)$ as

$$y_k(n) = \sum_{m=0}^{M-1} h_k(m) x(n-m), \tag{3}$$

and introducing $\mathbf{h}_k = [\, h_k(0)\ \cdots\ h_k(M-1)\,]^H$, we can express the output of the filter as $y_k(n) = \mathbf{h}_k^H \mathbf{x}(n)$, with $(\cdot)^H$ being the Hermitian transpose operator. The expected output power can thus be expressed as

$$\mathrm{E}\left\{|y_k(n)|^2\right\} = \mathrm{E}\left\{\mathbf{h}_k^H \mathbf{x}(n) \mathbf{x}^H(n) \mathbf{h}_k\right\} \tag{4}$$

$$= \mathbf{h}_k^H \mathbf{R} \mathbf{h}_k, \tag{5}$$

where $\mathrm{E}\{\cdot\}$ denotes the statistical expectation. The above expression can be seen to involve the covariance matrix defined as $\mathbf{R} = \mathrm{E}\left\{\mathbf{x}(n)\mathbf{x}^H(n)\right\}$. We will now analyze the covariance matrix a bit more in detail.

The signal model in (2) can now be written using the above definitions as

$$\mathbf{x}(n) = \sum_{k=1}^{K} \mathbf{Z}_k \begin{bmatrix} e^{-j\omega_k 1 n} & & 0 \\ & \ddots & \\ 0 & & e^{-j\omega_k L_k n} \end{bmatrix} \mathbf{a}_k^* + \mathbf{e}(n) \tag{6}$$

$$\triangleq \sum_{k=1}^{K} \mathbf{Z}_k \mathbf{a}_k^*(n) + \mathbf{e}(n) \tag{7}$$

or, alternatively, as $\mathbf{x}(n) \triangleq \sum_{k=1}^{K} \mathbf{Z}_k(n) \mathbf{a}_k^* + \mathbf{e}(n)$. Here, $\mathbf{Z}_k \in \mathbb{C}^{M \times L_k}$ is a Vandermonde matrix, being constructed from L_k harmonically related complex sinusoidal vectors as $\mathbf{Z}_k = [\, \mathbf{z}(\omega_k)\ \cdots\ \mathbf{z}(\omega_k L_k)\,]$, with $\mathbf{z}(\omega) = [\, 1\ e^{-j\omega}\ \cdots\ e^{-j\omega(M-1)}\,]^T$, and $\mathbf{a}_k = [\, a_{k,1}\ \cdots\ a_{k,L_k}\,]^H$ is a vector containing the complex amplitudes. Introducing $z_k = e^{-j\omega_k}$, the structure of the matrix $\mathbf{Z}_k$ can be seen to be

$$\mathbf{Z}_k = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_k^1 & z_k^2 & \cdots & z_k^{L_k} \\ \vdots & \vdots & \ddots & \vdots \\ z_k^{(M-1)} & z_k^{(M-1)2} & \cdots & z_k^{(M-1)L_k} \end{bmatrix}. \tag{8}$$

From this, it can be observed that either the complex amplitude vector or the Vandermonde matrix can be thought of as time-varying quantities, i.e., $\mathbf{a}_k^*(n) = \mathbf{D}^n \mathbf{a}_k^*$ and $\mathbf{Z}_k(n) = \mathbf{Z}_k \mathbf{D}^n$ with

$$\mathbf{D}^n = \begin{bmatrix} e^{-j\omega_k 1 n} & & 0 \\ & \ddots & \\ 0 & & e^{-j\omega_k L_k n} \end{bmatrix}, \tag{9}$$

meaning that the time index n can be seen as either changing the sinusoidal basis or, equivalently, the phases of the sinusoids. Depending on the context, one perspective may be more appropriate or convenient than the other.

For statistically independent sources, the covariance matrix of the observed signal can be written as $\mathbf{R} = \sum_{k=1}^{K} \mathbf{R}_k = \sum_{k=1}^{K} \mathrm{E}\left\{\mathbf{x}_k(n)\mathbf{x}_k^H(n)\right\}$, i.e., as a summation of the covariance matrices of the individual sources. By inserting the

[1] For many signals, the frequencies of the harmonics will not be exact integer multiples of the fundamental. This can be handled in several ways by modifying the signal model (see, e.g., [27] for more on this), but this is beyond the scope of this paper and will not be discussed any further.

single-pitch signal model in this expression, we can express the covariance matrix of the multi-pitch signal $\mathbf{x}(n)$ as

$$\mathbf{R} = \sum_{k=1}^{K} \mathbf{Z}_k \mathrm{E}\left\{\mathbf{a}_k^*(n)\mathbf{a}_k^T(n)\right\}\mathbf{Z}_k^H + \mathrm{E}\left\{\mathbf{e}_k(n)\mathbf{e}_k^H(n)\right\} \quad (10)$$

$$= \sum_{k=1}^{K} \mathbf{Z}_k \mathbf{P}_k \mathbf{Z}_k^H + \mathbf{Q}, \quad (11)$$

where the matrix $\mathbf{P}_k$ is the covariance matrix of the amplitudes, i.e., $\mathbf{P}_k = \mathrm{E}\left\{\mathbf{a}_k^*(n)\mathbf{a}_k^T(n)\right\}$. For statistically independent and uniformly distributed stochastic phases (on the interval $(-\pi, \pi]$), this matrix reduces to a diagonal matrix having the power of the sinusoidal components on the diagonal, i.e., $\mathbf{P}_k = \mathrm{diag}\left(\left[\begin{array}{ccc} A_{k,1}^2 & \cdots & A_{k,L}^2 \end{array}\right]\right)$. We note, however, that one can also arrive at the same result by considering the complex amplitudes deterministic as in (6). Moreover, the matrix $\mathbf{Q}$ is the covariance matrix of the combined noise source $e(n)$, i.e., $\mathbf{Q} = \mathrm{E}\left\{\mathbf{e}(n)\mathbf{e}^H(n)\right\} = \sum_{k=1}^{K}\mathbf{Q}_k$ also referred to as the noise covariance matrix.

In practice, the covariance matrix is unknown and is replaced by an estimate, namely the sample covariance matrix defined as $\widehat{\mathbf{R}} = \frac{1}{G}\sum_{n=M-1}^{N-1}\mathbf{x}(n)\mathbf{x}^H(n)$ where $G = N - M + 1$ is the number of samples over which we average. For the sample covariance matrix $\widehat{\mathbf{R}}$ to be invertible, we require that $M < \frac{N}{2} + 1$ so that the averaging consists of at least M rank 1 vectors (see, e.g., [31] for details). In the rest of the paper, we will assume that M is chosen proportionally to N such that when N grows, so does M. This is important for the consistency of the methods under consideration.

III. Optimal Single Filter Designs

A. Basic Principle

We will now proceed with the first design. We seek to find an optimal set of coefficients, $\{h_k(m)\}$, such that the mean square error (MSE) between the filter output, $y_k(n)$, and a desired output, a signal model if you will, $\hat{y}_k(n)$, is minimized in the following sense:

$$P = \frac{1}{G}\sum_{n=M-1}^{N-1}\left|y_k(n) - \hat{y}_k(n)\right|^2, \quad (12)$$

Since we are here concerned with periodic signals, this should be reflected in the choice of the signal model $\hat{y}_k(n)$. In fact, this should be chosen as the sum of sinusoids having frequencies that are integer multiples of a fundamental frequency ω_k weighted by their respective complex amplitudes $a_{k,l}$, i.e., $\hat{y}_k(n) = \sum_{l=1}^{L_k} a_{k,l}e^{j\omega_k l n}$. This leaves us with the following expression for the MSE:

$$P = \frac{1}{G}\sum_{n=M-1}^{N-1}\left|\sum_{m=0}^{M-1}h_k(m)x(n-m) - \sum_{l=1}^{L_k}a_{k,l}e^{j\omega_k l n}\right|^2. \quad (13)$$

In the following derivations, we assume the fundamental frequency ω_0 and the number of harmonics L_k to be known (with $L_k < M$), although the so-obtained filters can later be used for finding these quantities. Next, we proceed to find not only the filter coefficients but also the complex amplitudes $a_{k,l}$. We now introduce a vector containing the complex sinusoids at time n, i.e.,

$$\mathbf{w}_k(n) = \left[\begin{array}{ccc} e^{j\omega_k 1 n} & \cdots & e^{j\omega_k L_k n} \end{array}\right]^T. \quad (14)$$

With this, we can express (12) as

$$P = \frac{1}{G}\sum_{n=M-1}^{N-1}\left|\mathbf{h}_k^H\mathbf{x}(n) - \mathbf{a}_k^H\mathbf{w}_k(n)\right|^2, \quad (15)$$

which in turn can be expanded into

$$P = \mathbf{h}_k^H\widehat{\mathbf{R}}\mathbf{h}_k - \mathbf{a}_k^H\mathbf{G}_k\mathbf{h}_k - \mathbf{h}_k^H\mathbf{G}_k^H\mathbf{a}_k + \mathbf{a}_k^H\mathbf{W}_k\mathbf{a}_k, \quad (16)$$

where the new quantities are defined as

$$\mathbf{G}_k = \frac{1}{G}\sum_{n=M-1}^{N-1}\mathbf{w}_k(n)\mathbf{x}^H(n) \quad (17)$$

and

$$\mathbf{W}_k = \frac{1}{G}\sum_{n=M-1}^{N-1}\mathbf{w}_k(n)\mathbf{w}_k^H(n). \quad (18)$$

B. Solution

Solving for the complex amplitudes in (16) yields the following expression [31]

$$\hat{\mathbf{a}}_k = \mathbf{W}_k^{-1}\mathbf{G}_k\mathbf{h}_k, \quad (19)$$

which depends on the yet unknown filter $\mathbf{h}_k$. For $\mathbf{W}_k$ to be invertible, we require that $G \geq L_k$, but to ensure that also the covariance matrix is invertible (as already noted), we will further assume that $G \geq M$. By substituting the expression above back into (16), we get

$$P = \mathbf{h}_k^H\widehat{\mathbf{R}}_k\mathbf{h}_k - \mathbf{h}_k^H\mathbf{G}_k^H\mathbf{W}_k^{-1}\mathbf{G}_k\mathbf{h}_k. \quad (20)$$

By some simple manipulation, we see that this can be simplified somewhat as

$$P = \mathbf{h}_k^H\left(\widehat{\mathbf{R}}_k - \mathbf{G}_k^H\mathbf{W}_k^{-1}\mathbf{G}_k\right)\mathbf{h}_k \triangleq \mathbf{h}_k^H\widehat{\mathbf{Q}}_k\mathbf{h}_k \quad (21)$$

where

$$\widehat{\mathbf{Q}}_k = \widehat{\mathbf{R}}_k - \mathbf{G}_k^H\mathbf{W}_k^{-1}\mathbf{G}_k \quad (22)$$

can be thought of as a *modified* covariance matrix estimate that is formed by subtracting the contribution of the harmonics from the covariance matrix given the fundamental frequency. It must be stressed, though, that for multi-pitch signals, this estimate will differ from $\mathbf{Q}_k$ in the sense that $\widehat{\mathbf{Q}}_k$ will then also contain the contribution of the other sources. Therefore, $\widehat{\mathbf{Q}}_k$ is only truly an estimate of $\mathbf{Q}_k$ for single-pitch signals. Note also that similar observations apply to the usual use of APES [29], [30].

Solving for the unknown filter in (21) directly results in a trivial and useless result, namely the zero vector. To fix this, we will introduce some additional constraints. Not only should the output of the filter be periodic, i.e., resemble a sum of harmonically related sinusoids, the filter should also have unit gain for all the harmonic frequencies of that particular source, i.e., $\sum_{m=0}^{M-1}h_k(m)e^{-j\omega_k l m} = 1$ for $l = 1, \ldots, L_k$, or, equivalently, as $\mathbf{h}_k^H\mathbf{z}(\omega_k l) = 1$. We can now state the

filter design problem as the following constrained optimization problem:

$$\min_{\mathbf{h}_k} \mathbf{h}_k^H \widehat{\mathbf{Q}}_k \mathbf{h}_k \quad \text{s.t.} \quad \mathbf{h}_k^H \mathbf{z}(\omega_k l) = 1, \quad \text{for} \quad l = 1, \ldots, L_k. \tag{23}$$

The constraints for the L_k harmonics can also be expressed as $\mathbf{h}_k^H \mathbf{Z}_k = \mathbf{1}$, where $\mathbf{1} = [\ 1\ \cdots\ 1\ ]^T$. The problem in (23) is a quadratic optimization problem with equality constraints that can be solved using the Lagrange multiplier method. Introducing the Lagrange multiplier vector

$$\boldsymbol{\lambda} = [\ \lambda_1\ \cdots \lambda_{L_k}\ ]^T, \tag{24}$$

the Lagrangian dual function of the problem stated above can be expressed as

$$\mathcal{L}(\mathbf{h}_k, \boldsymbol{\lambda}) = \mathbf{h}_k^H \widehat{\mathbf{Q}}_k \mathbf{h}_k - \left(\mathbf{h}_k^H \mathbf{Z}_k - \mathbf{1}^T\right) \boldsymbol{\lambda}. \tag{25}$$

By taking the derivative with respect to the unknown filter vector and the Lagrange multiplier vector, we get

$$\nabla \mathcal{L}(\mathbf{h}_k, \boldsymbol{\lambda}) = \begin{bmatrix} \widehat{\mathbf{Q}}_k & -\mathbf{Z}_k \\ -\mathbf{Z}_k^H & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{h}_k \\ \boldsymbol{\lambda} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}. \tag{26}$$

Equaling this to zero, i.e., $\nabla \mathcal{L}(\mathbf{h}_k, \boldsymbol{\lambda}) = \mathbf{0}$, we obtain

$$\boldsymbol{\lambda} = \left(\mathbf{Z}_k^H \widehat{\mathbf{Q}}_k^{-1} \mathbf{Z}_k\right)^{-1} \mathbf{1} \tag{27}$$

and

$$\mathbf{h}_k = \widehat{\mathbf{Q}}_k^{-1} \mathbf{Z}_k \boldsymbol{\lambda}, \tag{28}$$

which combine to yield the following optimal filters:

$$\hat{\mathbf{h}}_k = \widehat{\mathbf{Q}}_k^{-1} \mathbf{Z}_k \left(\mathbf{Z}_k^H \widehat{\mathbf{Q}}_k^{-1} \mathbf{Z}_k\right)^{-1} \mathbf{1}. \tag{29}$$

We will refer to this filter as SF-APES (single filter APES-like design). This filter is optimal in the sense that it has unit gain at the harmonic frequencies and an output that resembles a sum of harmonically related sinusoids while everything else is suppressed maximally. It can readily be used for determining the amplitudes of those sinusoids by inserting (29) into (19), which yields the following estimate:

$$\hat{\mathbf{a}}_k = \mathbf{W}_k^{-1} \mathbf{G}_k \widehat{\mathbf{Q}}_k^{-1} \mathbf{Z}_k \left(\mathbf{Z}_k^H \widehat{\mathbf{Q}}_k^{-1} \mathbf{Z}_k\right)^{-1} \mathbf{1} \tag{30}$$

$$= \mathbf{W}_k^{-1} \mathbf{G}_k \left(\widehat{\mathbf{R}} - \mathbf{G}_k^H \mathbf{W}_k^{-1} \mathbf{G}_k\right)^{-1} \mathbf{Z}_k \tag{31}$$

$$\times \left(\mathbf{Z}_k^H \left(\widehat{\mathbf{R}} - \mathbf{G}_k^H \mathbf{W}_k^{-1} \mathbf{G}_k\right)^{-1} \mathbf{Z}_k\right)^{-1} \mathbf{1}. \tag{32}$$

The output power of the filter, when this is applied to the original signal, can be expressed as $\hat{\mathbf{h}}_k^H \widehat{\mathbf{R}} \hat{\mathbf{h}}_k$, which may be used for determining the fundamental frequency by treating ω_k in $\mathbf{Z}_k$, $\mathbf{G}_k$, $\mathbf{W}_k$ as an unknown parameter and then pick as an estimate the value for which the output power is maximized, i.e.,

$$\hat{\omega}_k = \arg\max_{\omega_k} \hat{\mathbf{h}}_k^H \widehat{\mathbf{R}} \hat{\mathbf{h}}_k. \tag{33}$$

In practice, this is done in the following manner: For a segment of data, the optimal filters are found for each candidate fundamental frequency. The filters are then applied to the signal and the output power is measured. This shows how much power is passed by the filters as a function of the fundamental frequency, and the fundamental frequency estimate is then picked as the fundamental frequency for which the most power is passed. One can also obtain an estimate of the number of harmonics L by estimating the noise variance by filtering out the harmonics and applying one of the many statistical model order estimation tools, like, e.g., the MAP-rule of [32], as shown in [33]. From the optimal filter, it is thus possible to obtain a full parametrization of periodic signals as was claimed in the introduction.

The proposed filter design leads to filters that are generally also much well-behaved for high SNRs, where Capon-like filters are well-known to perform poorly and require that diagonal loading or similar techniques be applied [31]. The proposed filter also holds several advantages over traditional methods, like the comb filtering approach or sinusoidal filters (also known as FFT filters), namely that it is 1) optimal given the observed signal, and 2) optimized for periodic filter output. To quantify further what exactly is meant by the filter being optimal, one has to take a look back at (12). The found filter is optimal in the sense that it minimizes the difference in (12), the exact time interval being determined by the summation limits, under the constraint that it should pass the content at specific frequencies undistorted and the output should to the extent possible resemble a periodic signal.

We will now discuss some simplified designs that are all special cases of the optimal single filter design.

1) Simplification No. 1: We remark that it can be shown that $\mathbf{W}_k$ is asymptotically identical to the identity matrix. By replacing $\mathbf{W}_k$ by $\mathbf{I}$ in (21), one obtains the usual noise covariance matrix estimate, used, for example, in [12]. As before, the optimal filters are

$$\hat{\mathbf{h}}_k = \widehat{\mathbf{Q}}_k^{-1} \mathbf{Z}_k \left(\mathbf{Z}_k^H \widehat{\mathbf{Q}}_k^{-1} \mathbf{Z}_k\right)^{-1} \mathbf{1}, \tag{34}$$

but the modified covariance matrix estimate is now determined as

$$\widehat{\mathbf{Q}}_k = \widehat{\mathbf{R}} - \mathbf{G}_k^H \mathbf{G}_k, \tag{35}$$

which is computationally simpler as it does not require the inversion of the matrix $\mathbf{W}_k$ for each candidate frequency. We refer to this design as SF-APES (appx). It must be stressed that for finite N, this is only an approximation that, nonetheless, may still be useful for practical reasons as it is much simpler. This approximation is actually equivalent to estimating the noise covariance matrix by subtracting from $\widehat{\mathbf{R}}_k$ an estimate of the covariance matrix model (for a single source) in (11) based on periodogram-like amplitude estimates.

2) Simplification No. 2: Interestingly, the Capon-like filters of [19], [34] can be obtained as a special case of the solution presented here by setting the modified covariance matrix equal to the sample covariance matrix of the observed signal, i.e., $\widehat{\mathbf{Q}}_k = \widehat{\mathbf{R}}$. More specifically, the optimal filter is then

$$\hat{\mathbf{h}}_k = \widehat{\mathbf{R}}^{-1} \mathbf{Z}_k \left(\mathbf{Z}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{Z}_k\right)^{-1} \mathbf{1}, \tag{36}$$

which is the design that we will refer to as Capon in the experiments. The main difference between the design proposed here and the Capon-like designs previously proposed is that the modified covariance matrix $\widehat{\mathbf{Q}}_k$ is used in (23) in place of

$\widehat{\mathbf{R}}$, i.e., the difference is essentially in terms of the output of the filter being periodic.

3) Simplification No. 3: A simpler set of filters yet are obtained from (36) by assuming that the input signal is white, i.e., $\widehat{\mathbf{R}} = \sigma^2\mathbf{I}$. These filters are then no longer signal adaptive, but they also only have to be calculated once. The optimal filters are then given by

$$\hat{\mathbf{h}}_k = \mathbf{Z}_k \left(\mathbf{Z}_k^H \mathbf{Z}_k\right)^{-1} \mathbf{1}, \tag{37}$$

which is thus fully specified by the pseudo-inverse of $\mathbf{Z}_k$.

4) Simplification No. 4: Curiously, the filters defined in (37) can be further simplified as follows: complex sinusoids are asymptotically orthogonal for any set of distinct frequencies, which means that the pseudo-inverse of $\mathbf{Z}_k$ can be approximated as

$$\lim_{M\to\infty} M\mathbf{Z}_k \left(\mathbf{Z}_k^H \mathbf{Z}_k\right)^{-1} = \mathbf{Z}_k \lim_{M\to\infty} \left(\frac{1}{M}\mathbf{Z}_k^H \mathbf{Z}_k\right)^{-1} \tag{38}$$

$$= \mathbf{Z}_k. \tag{39}$$

This means that the filter becomes particularly simple. In fact, it is just

$$\hat{\mathbf{h}}_k = \frac{1}{M}\mathbf{Z}_k \mathbf{1}, \tag{40}$$

i.e., the normalized sum over a set of filters defined by Fourier vectors.

IV. Optimal Filterbank Designs

A. Basic Principle

We will now consider a different approach to designing optimal filters for periodic signals. Suppose that we design a filter not for the entire periodic signal, but one for each of the harmonics of the signal. In that case, we seek to find a set of filter coefficients that depend on the harmonic number l, i.e., $\{h_{k,l}(m)\}$. The corresponding output of such a filter, we denote $y_{k,l}(n)$. The output of each filter should resemble a signal model $\hat{y}_{k,l}(n)$ exhibiting certain characteristics. As was the case with the single filter, we propose a cost function defined as

$$P_l = \frac{1}{G}\sum_{n=M-1}^{N-1} \left|y_{k,l}(n) - \hat{y}_{k,l}(n)\right|^2, \tag{41}$$

which measures the extent to which the filter output $y_{k,l}(n)$ resembles $\hat{y}_{k,l}(n)$. Adding this cost up across all harmonics of the kth source, we obtain an estimate of the discrepancy as

$$P = \sum_{l=1}^{L_k} P_l = \frac{1}{G}\sum_{l=1}^{L_k}\sum_{n=M-1}^{N-1} \left|y_{k,l}(n) - \hat{y}_{k,l}(n)\right|^2. \tag{42}$$

For the single filter design, the output of each filter should resemble a periodic function having possibly a number of harmonics. In the present case, however, the output of the filter should be just a single sinusoid, i.e., $\hat{y}_{k,l}(n) = a_{k,l}e^{j\omega_k l n}$. Defining

$$y_{k,l}(n) = \sum_{m=0}^{M-1} h_{k,l}(m)x(n-m) \triangleq \mathbf{h}_{k,l}^H \mathbf{x}(n), \tag{43}$$

we can express (42) as

$$P = \frac{1}{G}\sum_{l=1}^{L_k}\sum_{n=M-1}^{N-1} \left|\mathbf{h}_{k,l}^H \mathbf{x}(n) - a_{k,l}e^{j\omega_k l n}\right|^2. \tag{44}$$

To form an estimate of the kth source from the output of the filterbank, we simply sum over all the outputs of the individual filters, as each output is an estimate of the lth harmonic, i.e.,

$$y_k(n) = \sum_{l=1}^{L_k} y_{k,l}(n) = \sum_{m=0}^{M-1}\sum_{l=1}^{L_k} \mathbf{h}_{k,l}^H \mathbf{x}(n), \tag{45}$$

which shows that the filters of the filterbank can be combined to yield the single filter needed to extract the source. As before, we proceed in our derivation of the optimal filters by expanding this expression

$$P = \sum_{l=1}^{L_k} \mathbf{h}_{k,l}^H \widehat{\mathbf{R}} \mathbf{h}_{k,l}^H + \sum_{l=1}^{L_k} |a_{k,l}|^2 \tag{46}$$

$$- \sum_{l=1}^{L_k} \mathbf{h}_{k,l}^H \mathbf{g}(\omega_k l) a_{k,l}^* - \sum_{l=1}^{L_k} a_{k,l}\mathbf{g}^H(\omega_k l)\mathbf{h}_{k,l}, \tag{47}$$

where the $\widehat{\mathbf{R}}$ is defined as before and the only new quantity is

$$\mathbf{g}(\omega) = \frac{1}{G}\sum_{n=M-1}^{N-1} \mathbf{x}(n)e^{-j\omega n}. \tag{48}$$

B. Solution

With all the basic definitions in place, we can now derive the optimal filterbank. First, however, we must solve for the amplitudes. Differentiating (47) by $\hat{a}_{k,l}$ and setting the result equal to zero, we obtain

$$\hat{a}_{k,l} = \mathbf{h}_{k,l}^H \mathbf{g}(\omega_k l) \quad \text{for} \quad l = 1, \ldots, L_k. \tag{49}$$

Inserting this back into (47), we are left with an expression that depends only on the filters $\{\mathbf{h}_{k,l}\}$:

$$P = \sum_{l=1}^{L_k} \mathbf{h}_{k,l}^H \widehat{\mathbf{R}} \mathbf{h}_{k,l} - \sum_{l=1}^{L_k} \mathbf{h}_{k,l}^H \mathbf{g}(\omega_k l)\mathbf{g}^H(\omega_k l)\mathbf{h}_{k,l} \tag{50}$$

$$= \sum_{l=1}^{L_k} \mathbf{h}_{k,l}^H \left(\widehat{\mathbf{R}} - \mathbf{g}(\omega_k l)\mathbf{g}^H(\omega_k l)\right)\mathbf{h}_{k,l} \tag{51}$$

$$\triangleq \sum_{l=1}^{L_k} \mathbf{h}_{k,l}^H \widehat{\mathbf{Q}}_{k,l} \mathbf{h}_{k,l}, \tag{52}$$

where $\widehat{\mathbf{Q}}_{k,l}$ is a modified covariance matrix estimate as before, only it now depends on the individual harmonics. We can now move on to the problem of solving for the filters. As before, we must introduce some constraints to solve this problem. It is natural to impose that each filter $\mathbf{h}_{k,l}$ should have unit gains for the lth harmonic. However, one can take additional knowledge into account in the design by also requiring that the other harmonics are canceled by the filter. Mathematically, we can state this as

$$\mathbf{h}_{k,l}^H \mathbf{Z}_k = \mathbf{b}_l, \tag{53}$$

where

$$\mathbf{b}_l = [\ \underbrace{0\ \cdots\ 0}_{l-1}\ 1\ \underbrace{0\ \cdots\ 0}_{L_k-l}\ ]. \tag{54}$$

We can now state the design problem for the lth filter of the filterbank as

$$\min_{\mathbf{h}_{k,l}} \mathbf{h}_{k,l}^H \widehat{\mathbf{Q}}_{k,l} \mathbf{h}_{k,l} \quad \text{s.t.} \quad \mathbf{h}_{k,l}^H \mathbf{Z}_k = \mathbf{b}_l. \tag{55}$$

For this problem, the Lagrangian dual function is

$$\mathcal{L}(\mathbf{h}_{k,l}, \boldsymbol{\lambda}) = \mathbf{h}_{k,l}^H \widehat{\mathbf{Q}}_{k,l} \mathbf{h}_{k,l} - \left(\mathbf{h}_{k,l}^H \mathbf{Z}_k - \mathbf{b}_l^T\right) \boldsymbol{\lambda}. \tag{56}$$

By taking the derivative with respect to the unknown filter vector and the Lagrange multiplier vector, we get

$$\nabla \mathcal{L}(\mathbf{h}_{k,l}, \boldsymbol{\lambda}) = \begin{bmatrix} \widehat{\mathbf{Q}}_{k,l} & -\mathbf{Z}_k \\ -\mathbf{Z}_k^H & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{h}_{k,l} \\ \boldsymbol{\lambda} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{b}_l \end{bmatrix}. \tag{57}$$

By the usual method, we obtain

$$\boldsymbol{\lambda} = \left(\mathbf{Z}_k^H \widehat{\mathbf{Q}}_{k,l}^{-1} \mathbf{Z}_k\right)^{-1} \mathbf{b}_l \tag{58}$$

and

$$\mathbf{h}_{k,l} = \widehat{\mathbf{Q}}_{k,l}^{-1} \mathbf{Z}_k \boldsymbol{\lambda}. \tag{59}$$

This, finally, results in the following optimal filters for $l = 1, \ldots, L_k$

$$\hat{\mathbf{h}}_{k,l} = \widehat{\mathbf{Q}}_{k,l}^{-1} \mathbf{Z}_k \left(\mathbf{Z}_k^H \widehat{\mathbf{Q}}_{k,l}^{-1} \mathbf{Z}_k\right)^{-1} \mathbf{b}_l. \tag{60}$$

We will refer to this design as FB-APES (filterbank APES-like design). The individual filters can now be applied to obtain amplitude estimates as

$$\hat{a}_{k,l} = \hat{\mathbf{h}}_{k,l}^H \mathbf{g}(\omega_k l) \tag{61}$$

$$= \mathbf{b}_l^H \left(\mathbf{Z}_k^H \widehat{\mathbf{Q}}_{k,l}^{-1} \mathbf{Z}_k\right)^{-1} \mathbf{Z}_k^H \widehat{\mathbf{Q}}_{k,l}^{-1} \mathbf{g}(\omega_k l). \tag{62}$$

Organizing all the filters for the kth source in a matrix, we get

$$\widehat{\mathbf{H}}_k = [\ \mathbf{h}_{k,1}\ \cdots\ \mathbf{h}_{k,L_k}\ ]. \tag{63}$$

The optimal filters in (60) can also be rewritten using the matrix inversion lemma to obtain an expression that does not require direct inversion of $\widehat{\mathbf{Q}}_{k,l}$ of each l:

$$\widehat{\mathbf{Q}}_{k,l}^{-1} = \left(\widehat{\mathbf{R}} - \mathbf{g}(\omega_k l)\mathbf{g}^H(\omega_k l)\right)^{-1} \tag{64}$$

$$= \widehat{\mathbf{R}}^{-1} + \frac{\widehat{\mathbf{R}}^{-1}\mathbf{g}(\omega_k l)\mathbf{g}^H(\omega_k l)\widehat{\mathbf{R}}^{-1}}{1 - \mathbf{g}^H(\omega_k l)\widehat{\mathbf{R}}^{-1}\mathbf{g}(\omega_k l)}, \tag{65}$$

which can then be inserted into (60). As with the single filter approach, this design can also be used for estimating the fundamental frequency by summing over the output powers of all the filters, i.e.,

$$\hat{\omega}_k = \arg\max_{\omega_k} \sum_{l=1}^{L_k} \hat{\mathbf{h}}_{k,l}^H \widehat{\mathbf{R}} \hat{\mathbf{h}}_{k,l} \tag{66}$$

$$= \arg\max_{\omega_k} \sum_{l=1}^{L_k} \mathrm{Tr}\left\{\widehat{\mathbf{H}}_k^H \widehat{\mathbf{R}} \widehat{\mathbf{H}}_k\right\}. \tag{67}$$

Note that the filters can also be applied in a different way, or, rather, the output power can be measured differently. In (66), the output power is determined as the sum of output power of the individual filters. If, instead, the output power is measured on the estimated source obtained as in (45), one obtains

$$\mathrm{E}\left\{|y_k(n)|^2\right\} = \left(\sum_{l=1}^{L_k} \hat{\mathbf{h}}_{k,l}^H\right) \widehat{\mathbf{R}} \left(\sum_{l=1}^{L_k} \hat{\mathbf{h}}_{k,l}\right). \tag{68}$$

However, assuming that the output of the individual filters is uncorrelated, the two estimates will be identical (see [34] for more details about this).

At this point some remarks are in order. For the Capon-like filters of [19], [34], the single filter and the filterbank approaches are closely related. This is, however, not the case for the designs considered here in that they operate on different covariance matrix estimates, $\widehat{\mathbf{Q}}_k$ and $\widehat{\mathbf{Q}}_{k,l}$, respectively. While it is more complicated to compute the former than the latter, the latter must be computed a number of times, once for each harmonic l. This suggests that, in fact, the single filter should be preferable from a complexity point of view if the number of harmonics is high.

As with the single filter design, it is possible to obtain some simplified versions of the optimal design. Next, we will look more into some of these.

1) Simplification No. 1: By posing the optimization problem in (55) in a slightly different way, we obtain an important special case. More specifically, by changing the constraints of (55) such that each filter only has to have unit gain for the corresponding harmonic, we obtain the following problem:

$$\min_{\mathbf{h}_{k,l}} \mathbf{h}_{k,l}^H \widehat{\mathbf{Q}}_{k,l} \mathbf{h}_{k,l} \quad \text{s.t.} \quad \mathbf{h}_{k,l}^H \mathbf{z}(\omega_k l) = 1, \tag{69}$$

where, as before, $\widehat{\mathbf{Q}}_{k,l} = \widehat{\mathbf{R}} - \mathbf{g}(\omega_k l)\mathbf{g}(\omega_k l)^H$. The solution to this problem is, in fact, the usual single sinusoid APES filter [29], [30], which is

$$\hat{\mathbf{h}}_{k,l} = \frac{\widehat{\mathbf{Q}}_{k,l}^{-1}\mathbf{z}(\omega_k l)}{\mathbf{z}^H(\omega_k l)\widehat{\mathbf{Q}}_{k,l}^{-1}\mathbf{z}(\omega_k l)}. \tag{70}$$

This design takes only the individual harmonics into account in the design of the individual filters. Essentially, the filter that is obtained from (55) takes the presence of all the harmonics of the kth source into account, while the present one does not.

2) Simplification No. 2: Taking this one step further and replacing $\widehat{\mathbf{Q}}_{k,l}$ by $\widehat{\mathbf{R}}$, one obtains the well-known single sinusoid Capon filter [35]

$$\hat{\mathbf{h}}_{k,l} = \frac{\widehat{\mathbf{R}}^{-1}\mathbf{z}(\omega_k l)}{\mathbf{z}^H(\omega_k l)\widehat{\mathbf{R}}^{-1}\mathbf{z}(\omega_k l)}. \tag{71}$$

As with the prior simplification, this design leaves it for the algorithm to automatically cancel out the contribution of the other harmonics.

3) Simplification No. 3: Similarly, replacing $\widehat{\mathbf{Q}}_{k,l}$ by $\widehat{\mathbf{R}}$ in (60) results in the filters

$$\hat{\mathbf{h}}_{k,l} = \widehat{\mathbf{R}}^{-1}\mathbf{Z}_k \left(\mathbf{Z}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{Z}_k\right)^{-1} \mathbf{b}_l, \tag{72}$$

which are identical to the filters of the optimal Capon-like filterbank of [19]. Interestingly, when summed, it result in the

optimal single Capon-like filter as

$$\sum_{l=1}^{L_k} \hat{\mathbf{h}}_{k,l} = \hat{\mathbf{h}}_k. \tag{73}$$

4) Simplification No. 4: The previous design can, of course, be simplified further by assuming that the covariance matrix is white, i.e., $\widehat{\mathbf{R}} = \sigma^2\mathbf{I}$, which results in static filters that have to be calculated only once. The filters are then given by

$$\hat{\mathbf{h}}_{k,l} = \mathbf{Z}_k \left(\mathbf{Z}_k^H \mathbf{Z}_k\right)^{-1} \mathbf{b}_l, \tag{74}$$

which when organized in a filterbank matrix can be written as

$$\widehat{\mathbf{H}}_k = \mathbf{Z}_k \left(\mathbf{Z}_k^H \mathbf{Z}_k\right)^{-1}. \tag{75}$$

Source estimates obtained using this filterbank, as described in (45), will be exactly the same estimates as one would get using (37)–this can easily be verified by inserting the right hand side of (74) in (45). The resulting fundamental frequency estimators are, however, generally different, but are equivalent under certain conditions. In the experimental part of this paper, we will refer to this method as the FB-WNC design (filterbank white noise Capon-like design).

5) Simplification No. 5: Applying the asymptotic approximation in (39) to the filters in (74), we obtain even simpler filters. More specifically, (74) reduces to

$$\hat{\mathbf{h}}_{k,l} = \frac{1}{M}\mathbf{Z}_k\mathbf{b}_l, \tag{76}$$

and the filterbank matrix is then simply given by

$$\widehat{\mathbf{H}}_k = \frac{1}{M}\mathbf{Z}_k. \tag{77}$$

When applied to the problem of fundamental frequency estimation, as in (66), this leads to the familiar approximate non-linear least squares (NLS) method–it is nonlinear in the fundamental frequency, hence the name; it is also sometimes referred to as the harmonic summation method [27]. Note that when source estimates are obtain using this filterbank as described in (45), one will obtain exactly the same estimate as with (40). We will refer to this method as FB-WNC (appx) in the experiments, where it will serve as a method representative of the usual way filters are designed. A large class of methods exist for enhancement and separation of signals that operate on the coefficients of the short-time Fourier Transform (STFT) (see, e.g., [36], [37]). The individual bases of the STFT are the same as the individual filters of the filterbank (76), in fact, this will be the case for all methods that operate directly on the coefficients of the STFT, including mask-based methods like [38] and non-negative matrix factorization-based methods like [39].

6) Simplification No. 6: We will close this section by introducing one final simplification. If in lieu of $\widehat{\mathbf{Q}}_{k,l}$ we use $\widehat{\mathbf{Q}}_k$ as obtained for the single filter approach in (22) in (60), the optimal filters of the filterbank are then given by

$$\hat{\mathbf{h}}_{k,l} = \widehat{\mathbf{Q}}_k^{-1}\mathbf{Z}_k \left(\mathbf{Z}_k^H \widehat{\mathbf{Q}}_k^{-1}\mathbf{Z}_k\right)^{-1} \mathbf{b}_l. \tag{78}$$

It can be seen that the only difference between the different filters of the filterbank is then the vector $\mathbf{b}_l$, which serves to extract the filter for the individual harmonics. The filterbank matrix containing these filters can then be expressed as

$$\widehat{\mathbf{H}}_k = \widehat{\mathbf{Q}}_k^{-1}\mathbf{Z}_k \left(\mathbf{Z}_k^H \widehat{\mathbf{Q}}_k^{-1}\mathbf{Z}_k\right)^{-1}. \tag{79}$$

It is then also easy to see that these filters are related to the optimal single filter in (29) in a trivial way as

$$\hat{\mathbf{h}}_k = \widehat{\mathbf{H}}_k\mathbf{1}. \tag{80}$$

A similar relationship exists for the corresponding Capon-like filters [34]. Curiously, one would also obtain these filters by modifying (42) by moving the summation over the harmonics inside the absolute value, which would also be consistent with the formation of the source estimates according to (45).

V. Results

A. Practical Considerations

Before moving on to the experimental parts of the present paper, we will now go a bit more into details of how to apply the proposed filters and what issues one has to consider in doing so.

Given a segment of new data $\{x(n)\}$, the procedure is as follows:

1) Estimate the fundamental frequencies $\{\omega_k\}$ of all sources of interest for the data $\{x(n)\}$.
2) Determine or update recursively the sample covariance matrix $\widehat{\mathbf{R}}$.
3) Compute a noise covariance matrix estimate $\mathbf{Q}_k$ for each source (or for its harmonics $\mathbf{Q}_{k,l}$) and the inverse.
4) Compute the optimal single filter $\mathbf{h}_k$ or filterbank $\mathbf{H}_k$ for each source of interest k using one of the proposed designs.
5) Perform block filtering on the data $\{x(n)\}$ to obtain source estimates $y_k(n)$ for each source of interest k (using the observed signal from the previous segment as filter states as appropriate).

In performing the above, there are a number of user parameters that must be chosen. The following may serve as a basis for choosing these. Generally speaking, the higher the filter length M, the better the filter will be in attenuating noise and canceling interference from other sources as the filter has more degrees of freedom. This also means that the higher the model order, the more interfering sources the filter can deal with. However, there are several concerns that limit the filter length. First of all, the validity of the signal model. If the signal is not approximately stationary over the duration of the segment, the filters cannot possibly capture the signal of interest, neither can it deal with noise and other sources. On a related issue, the filter length M must be chosen, as mentioned, with $M < N/2 + 1$ to yield a well-conditioned problem. This means that the signal should be stationary over N and not just M. It should of course also be taken into account that the higher the filter order, the more computationally complex the design will also be. Regarding how often one should compute the optimal filters, i.e., how high the update-rate should be relative to M and N, it should be noted that for the filter outputs to be well-behaved, the filters must not

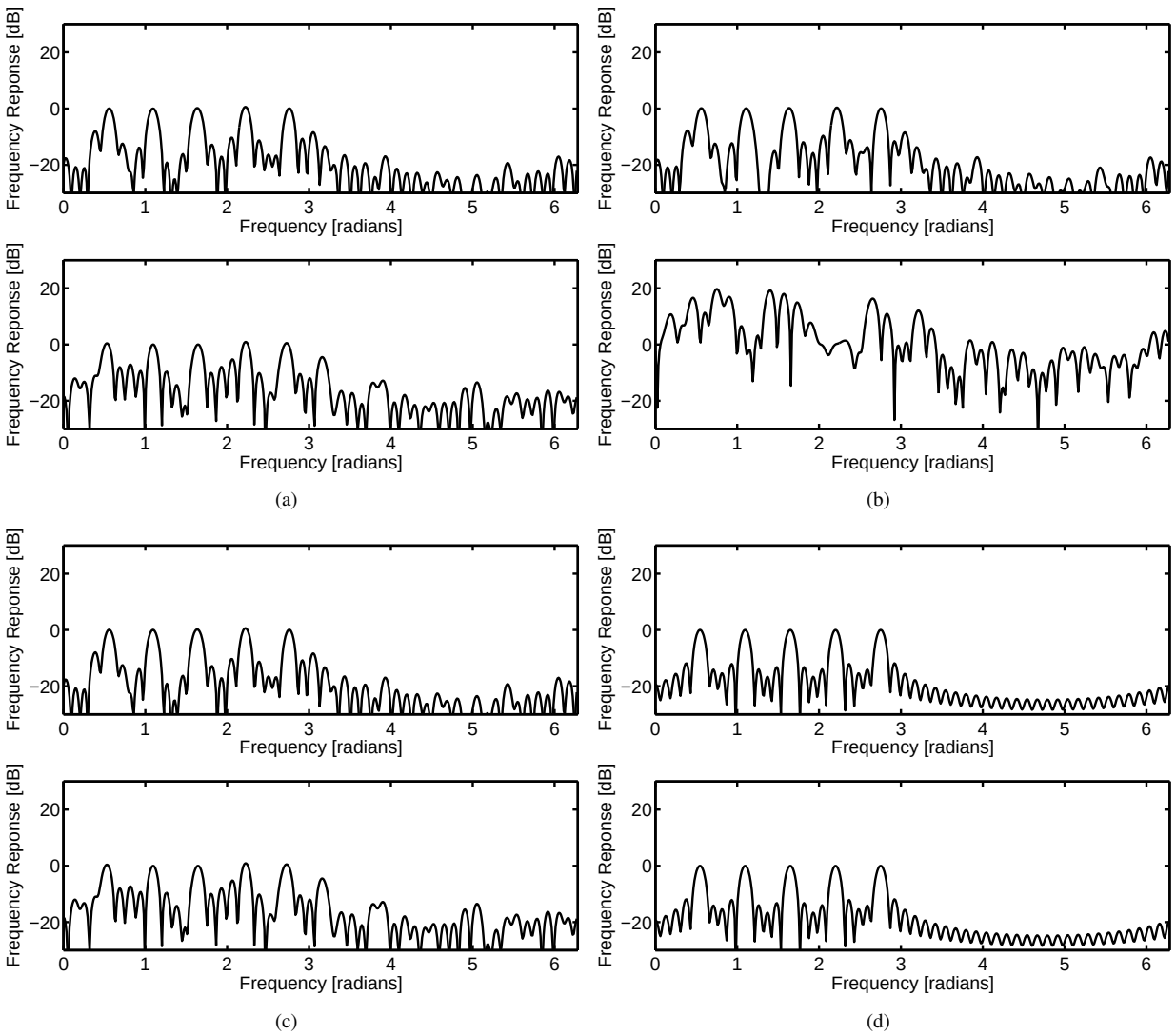


Fig. 1. Frequency responses of the various filters for a set of harmonically related sinusoids in white Gaussian noise at an SNR of -20 dB (top panels) and 20 dB (bottom panels). The designs shown here are (a) SF-APES, (b) SF-Capon, (c) FB-APES, (d) FB-WNC.

change abruptly. Consequently, it is advantageous to update the filters as often as possible by computing a new covariance matrix and subsequently new filters at the cost of increased computational complexity. In this process, one may also just as well update the fundamental frequency. In fact, it may also be advantageous to estimate a new fundamental frequency frequently relative to M and N to track changes in the signal of interest. This all suggests that it should be preferable in most situations to update the fundamental frequency, the covariance matrix and filters frequently.

Regarding numerical issues, as we have seen, the Capon-design suffers from bad conditioning of the covariance matrix for high SNRs, and it may thus be reasonable to use a regularized estimate of the covariance matrix, like $\widetilde{\mathbf{R}} = \widehat{\mathbf{R}} + \delta \mathbf{I}$, where δ is a small positive constant, before computing inverses. It is also possible that the APES-like designs may benefit from such modified estimates under extreme conditions.

B. Tested Designs

In the tests to follow, we will compare the proposed design methods to a number of existing FIR design methods. More specifically, we will compare the following:

- SF-APES which is the optimal single filter design given by (29).
- SF-Capon, i.e., the single filter design proposed in [19], [34], which is based on a generalization of the Capon principle. The optimal filter is given by (36).
- SF-APES (appx) is an approximation of SF-APES based on the simpler modified covariance matrix estimate in (35). It is thus a computationally simpler approximation to SF-APES.
- FB-APES is the optimal filterbank design given by (60).
- FB-WNC is a static single filter design based on Fourier vectors. The filter is given by (74). It serves as reference method as such filters are often used for processing of periodic signals.
- FB-WNC (appx) is an approximation of the FB-WNC

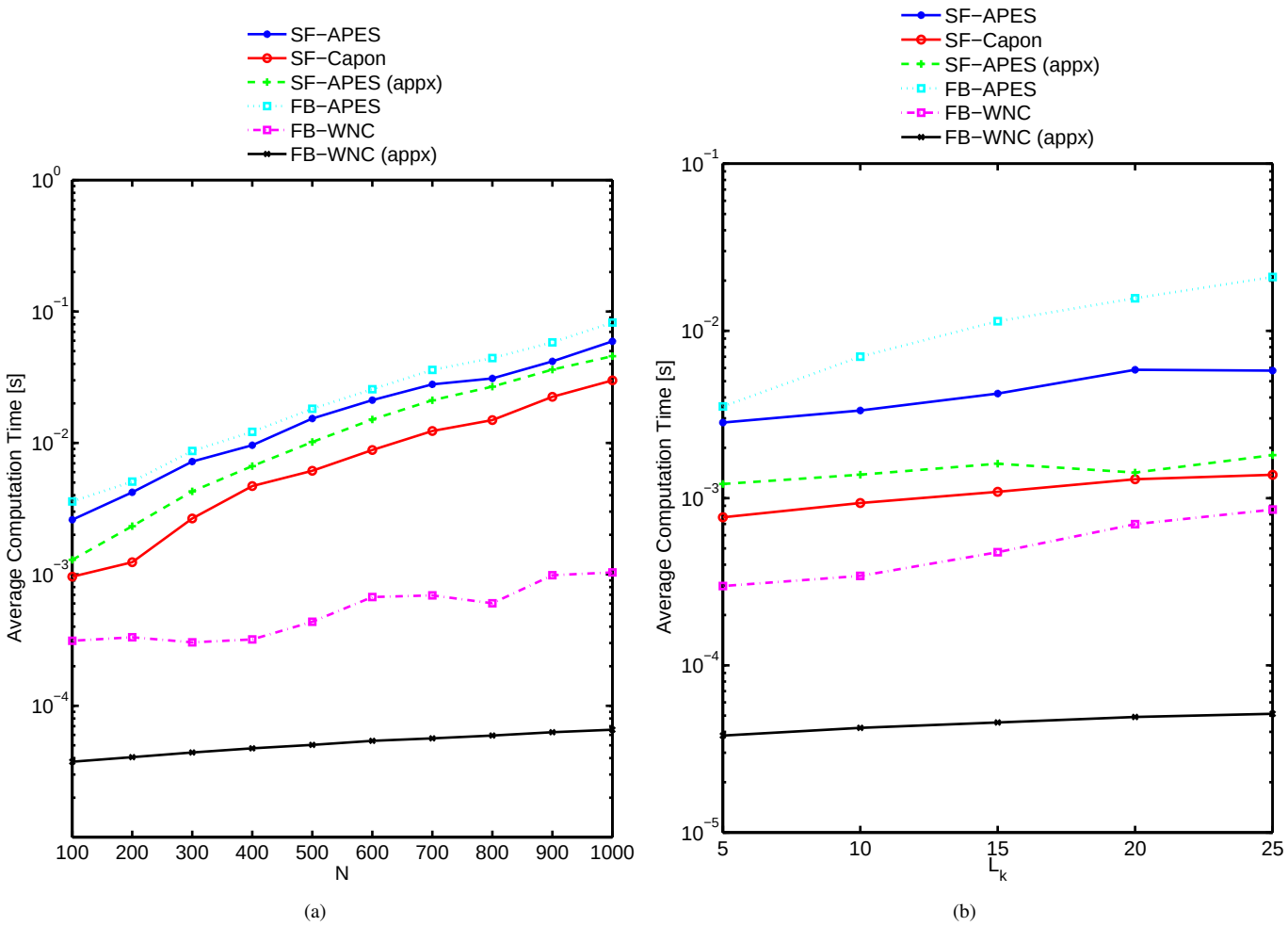


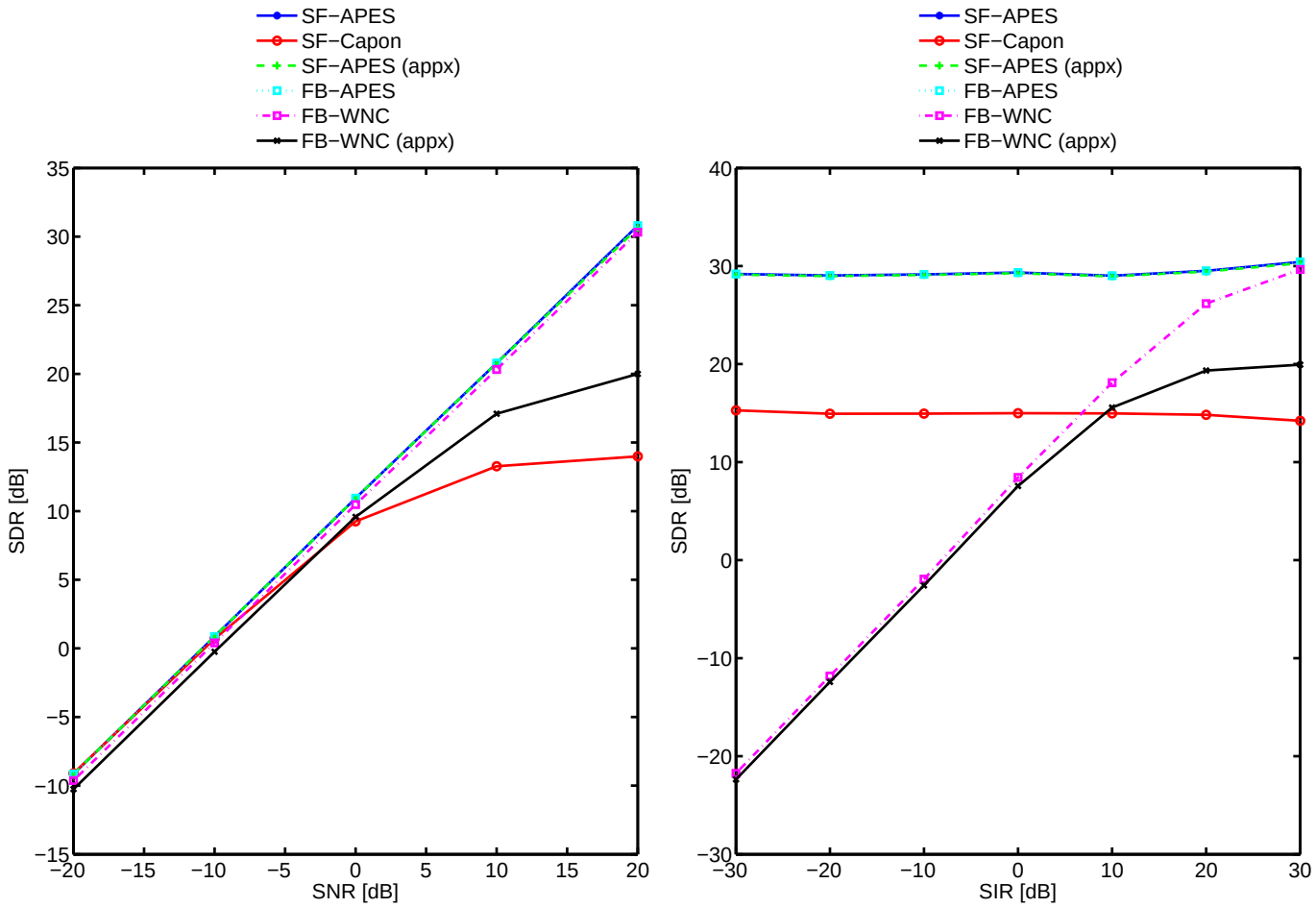
Fig. 2. Estimated computation times for the various filters designs (a) as a function of the number of observations N with $M = N/4$ and $L_k = 5$, and (b) as a function of the number of harmonics L_k with $N = 100$ and $M = 25$. For each data point, each filter was computed 1000 times and the average was computed.

filters with the filters being defined in (76). It is based on the asymptotic orthogonality of complex sinusoids. It is perhaps the most commonly used filter design method for processing periodic signals and is sometimes also referred to as the frequency sampling design method or the resulting filters as FFT filters.

Note that we do not include all the simplifications of Sections III and IV as some of them are trivially related.

C. Frequency Response

We will start out the experimental part of this paper by showing an example of the optimal filters obtained using some of the proposed methods and their various simplifications and the Capon-like filters of [19], [34]. More specifically, we will show the frequency response of the filters obtained using some of the various designs for a synthetic signal. In Figure 1, these are shown for a synthetic signal having $\omega_0 = 0.6283$, $L = 5$, Rayleigh distributed amplitudes and uniformly distributed phases with white Gaussian noise added at a -20 dB SNR (top panels) and 20 dB (bottom panels). The filters all have length 50 in these examples and were estimated from 200 samples. All the filters can be seen to exhibit the expected response for -20 dB SNR following the harmonic structure of the signal having 0 dB gain for the harmonic frequencies, and several of them are also quite similar. For an SNR of 20 dB, however, it can clearly be seen that the proposed filters still exhibit the desired response emphasizing the harmonics of the signal. The Capon-like design, SF-Capon, however, behaves erratically for 20 dB SNR, and this is typical of the Capon-like filters. Comparing the response of this method to the proposed ones, namely SF-APES, and FB-APES, it can be seen that this problem is overcome by the new design methodology. The erratic behavior of the Capon-like filter can be understood by noting that for high SNR, the Capon method will generally suffer from poor conditioning of the sample covariance matrix (as the eigenvalues only due to the noise tending toward zero), explaining the low accuracy of the resulting filter, and as the SNR increases, the filters obtained using the SF-Capon design will get progressively worse. We also remark that for the example considered here, SF-APES (appx) will be quite similar to SF-APES and FB-WNC (appx) to FB-WNC, for which reason these designs are not shown. This is because the asymptotic approximations that these derivative methods are based on are quite accurate in this case. This is also the likely



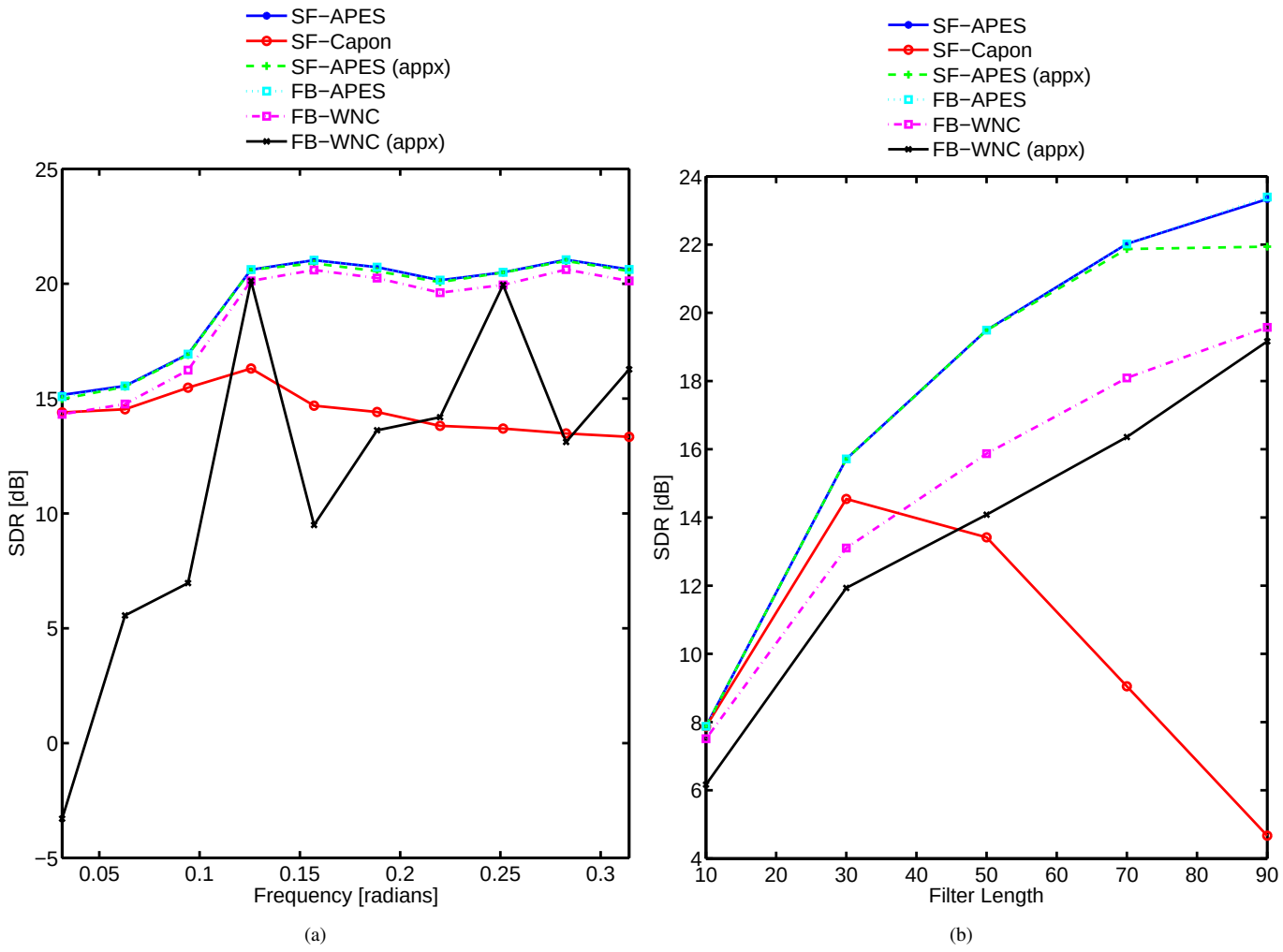
(a) (b)

Fig. 3. Performance of the various filters in SDR (a) as a function of the SNR and (b) the SIR with an interfering source present (with noise added at a fixed SNR of 20 dB).

explanation for the frequency responses of SF-APES and FB-APES looking extremely similar for both SNRs. We remark that while the adaptive designs will change with the observed signal, FB-WNC and its simplification will remain the same.

D. Computational Complexity and Computation Times

In comparing the performance of the various methods, it is of course also important to keep the computational complexity of the various methods in mind. All the tested methods, except the FB-WNC (appx) design, have cubic complexities involving operations of complexity $\mathcal{O}(M^3)$, $\mathcal{O}(L_k^3)$, $\mathcal{O}(M^2 L_k)$, and $\mathcal{O}(M L_k^2)$, as they involve matrix inversions and matrix-matrix multiplications. Some of the designs avoid some matrix inversions, like the SF-APES (appx) design, but such details cannot be differentiated with these asymptotic complexities. We therefore have measured average computation times of the various designs in MATLAB. More specifically, we have computed the average computation times over 1000 trials as a function of L_k and N as M is assumed to be chosen proportionally to N. The measurements were obtained on an Intel(R) Core(TM)2 CPU 6300 @ 1.86GHz with 2GB of RAM running MATLAB 7.6.0 (R2008a) and Linux 2.6.31-17 (Ubuntu). Note that the current implementations do not take into account the structure of the various matrices like, e.g., Toeplitz structure of the covariance matrix. The obtained results are shown in Figure 2(a) as a function of N with $M = N/4$ and $L_k = 5$ and as a function of L_k with $N = 100$ and $M = 25$ in Figure 2(b) for typical ranges of these quantities. From Figure 2(a) it can be observed that the computational complexity of the the designs SF-APES, SF-APES (appx), FB-APES, and SF-Capon indeed are cubic in M (and thus N), the difference essentially being a scaling. It can be observed that the FB-APES design is the most complex, owing to the different noisy covariance matrix estimates that must be determined for each harmonic. Note that for a very low number of harmonics, this design is less complex than SF-APES and SF-APES (appx). It can also be seen that, as expected, the SF-Capon design is the least complex of the adaptive designs, as it does not require the computation of a noise covariance matrix estimate. The general picture is the same in Figure 2(b), although it can be observed that the difference in computation time between the FB-APES method and the others appear to increase on the logarithmic scale as the number of harmonics is increased, the reason again being that the higher the number of harmonics, the more noise covariance matrices (and their inverses) must be determined.



(a) (b)

Fig. 4. Performance of the various filters in SDR (a) as function of the fundamental frequency, (b) and the filter length with an interfering source present.

E. Enhancement and Separation

Next, we will consider the application of the various filter designs to extracting periodic signals from noisy mixtures containing other periodic signals and noise or just noise. We will test the performance under various conditions by generating synthetic signals and then use the filters for extracting the desired signal. More specifically, the signals are generated in the following manner: A desired signal $\mathbf{s}_1(n)$ that we seek to extract from an observed signal $\mathbf{x}(n)$ is buried in a stochastic signal, i.e., noise $\mathbf{e}(n)$; additionally, an interfering source $\mathbf{s}_2(n)$ is also present, here in the form of a single sinusoid. The observed signal is thus constructed as

$$\mathbf{x}(n) = \mathbf{s}_1(n) + \mathbf{s}_2(n) + \mathbf{e}(n). \tag{81}$$

We will measure the extent to which the various filter designs are able to extract $\mathbf{s}_1(n)$ from $\mathbf{x}(n)$ using the signal-to-distortion ratio (SDR) defined as:

$$SDR = 20\log_{10}\frac{\|\mathbf{s}_1(n)\|_2}{\|\mathbf{s}_1(n) - \mathbf{y}_1(n)\|_2} \quad [\text{dB}]\,, \tag{82}$$

where $\mathbf{y}_1(n)$ is the signal extracted by applying the obtained filters to $\mathbf{x}(n)$. The ultimate goal is of course to reconstruct $\mathbf{s}_1(n)$ as closely as possible and, therefore, to maximize the SDR.

As a measure of the power of the interfering signal $\mathbf{s}_2(n)$ relative to the desired signal $\mathbf{s}_1(n)$, we use the following measure:

$$SIR = 20\log_{10}\frac{\|\mathbf{s}_1(n)\|_2}{\|\mathbf{s}_2(n)\|_2} \quad [\text{dB}]\,, \tag{83}$$

which we refer to as the signal-to-interference ratio (SIR) (for a discussion of performance measures for assessment of separation algorithms see, e.g., [38], [40]). It is expected that the higher the SIR, the worse the SDR will be. Finally, we measure how noisy the signal is using the signal-to-noise ratio (SNR) defined as

$$SNR = 20\log_{10}\frac{\|\mathbf{s}_1(n)\|_2}{\|\mathbf{e}(n)\|_2} \quad [\text{dB}]\,. \tag{84}$$

The reader should be aware that our definitions of SDR and SIR are consistent with those of [40], but also that our definition of SNR differs but is consistent with its use in estimation theory. In the experiments reported next, unless otherwise stated, the conditions were as follows; the above quantities were calculated by applying the found filters to the observed

signal and the SDR was then measured. This was then repeated 100 times for each test condition, i.e., the quantities are determined using Monte Carlo simulations. In doing this, the zero-state responses of the filters were ignored. Segments of length $N = 200$ were used with filter lengths of $M = N/4$ (for all designs) and an SNR of 20 dB was used. The desired signal was generated with a fundamental frequency of 0.5498 and five harmonics. The real and imaginary values of the complex amplitudes were generated as realizations of i.i.d. Gaussian random variables, leading to Rayleigh distributed amplitudes and uniformly distributed phases. The interfering source was a periodic signal having a fundamental frequency of 0.5890, five harmonics and with Rayleigh distributed amplitudes and uniformly distributed phases. Its amplitudes were then scaled to match the desired SIR in each realization. In these experiments, we will assume that the fundamental frequency of the desired signal is known while the fundamental frequency of the interference is unknown. As has already been mentioned, it is possible to estimate the fundamental frequency using the proposed filters, but this is beyond the scope of this paper, and we will just assume that the fundamental frequency has been estimated a priori using one of the methods of [27].

In the first experiment, only the desired signal and the noise are present, i.e., no interfering source was added, and the performance of the filters is observed as a function of the SNR. The resulting measurements are plotted in Figure 3(a). It can be seen that the Capon-like filter design, SF-Capon, that was the starting point of this work, performs poorly in this task. In fact, it is worse than the static designs FB-WNC and FB-WNC (appx). It can also be observed that the APES-like filters, SF-APES, SF-APES (appx) and FS-APES, all perform well, achieving the highest SDR. In [19], it was shown that the Capon-like filters perform well in terms of multi-pitch estimation under adverse conditions compared to the alternatives. This was especially true when multiple periodic sources were present at the same time as the signal-adaptive optimal designs were able to cancel out the interference without prior knowledge of it. It appears that with this particular setup, there is a 10 dB reduction in the noise regardless of the SNR for the proposed filters, and, interestingly, all the filter designs seem to tend perform similarly for low SNRs. This means that there appears to be no reason to prefer one method over the others for low SNRs, in which case the simplest design then should be chosen.

The next experiment is, therefore, concerned with the performance of the filters when interference is present. Here, the noise level, i.e., the SNR, is kept constant at 20 dB while the SIR is varied. The results are depicted in Figure 3(b). This figure clearly shows the advantage that the adaptive designs, SF-APES, SF-APES (appx), FB-APES, and SF-Capon, hold over the static ones, FB-WNC and FB-WNC (appx) in that the former perform well even when the interference is very strong, while the latter does not. The advantages of the designs proposed herein are also evident as the APES-like filters, SF-APES, SF-APES (appx), and FS-APES, outperform all others for the entire tested range of SIR values. We remark that in several of these figures, it may be hard to distinguish the performance of SF-APES, SF-APES (appx), and FB-APES as the curves are very close; indeed they appear to have similar performance in terms of SDR.

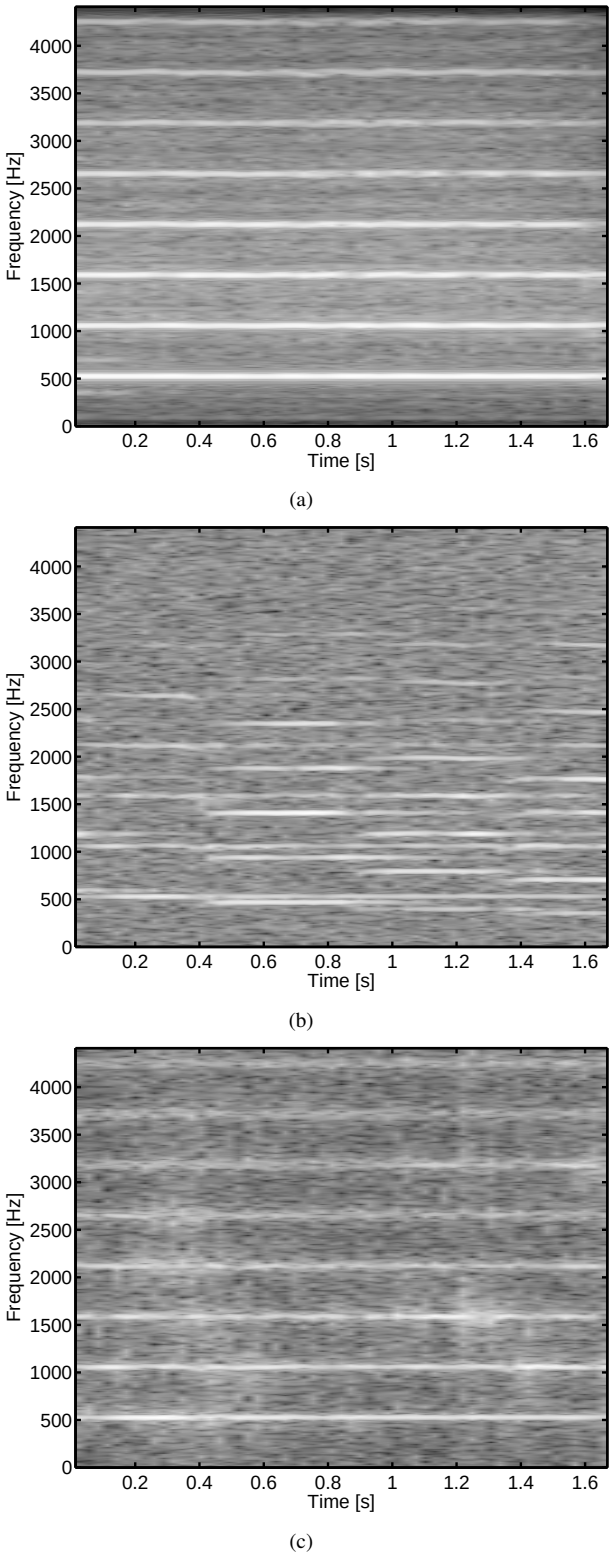


Fig. 5. Shown are: (a) the spectrogram of the original signal, (b) with noise and interference added with $SNR = 0$ dB and $SIR = -10$ dB (c) and the signal extracted using the FS-APES design.

As some of the simpler designs are based on sinusoids being asymptotically orthogonal, namely SF-APES (appx) and FB-

WNC (appx), it is interesting to see how the various filters perform when this is not the case. We do this by lowering the fundamental frequency for a given N, as for a given N, the fundamental frequency has to be high, relatively speaking, for the asymptotic approximation to hold. In this case, only noise is added to the desired signal at an SNR of 10 dB. The results are shown in Figure 4(a). As could be expected, the aforementioned approximate designs perform poorly (as does the Capon-like filters SF-Capon), but, generally, the performance of all the methods degrades as the fundamental frequency is lowered. This is, however, to be expected. Note that the reason FB-WNC (appx) performs well for certain fundamental frequencies is that the harmonics may be close to (or exactly) orthogonal, but this would merely be a coincidence in all practical situations.

Now we will investigate the influence of the filter length by varying M while keeping N fixed at 200, here in the presence of an interfering source. In this case, noise is added at an SNR of 10 dB while the SIR was 10 dB. In Figure 4(b), the results are shown. The conclusions are essentially the same as for the other experiments; the proposed filter designs perform the best, the SF-Capon filters behave erratically, and the static designs FB-WNC and FB-WNC (appx) perform poorly when interference is present. We note that for the respective matrices to be invertible, the filter lengths cannot be too long. On the other hand, one would expect that the longer the filters, the better the performance as the filters have more degrees of freedom to capture the desired signal while canceling noise and interference, and this indeed seems to be the case.

These experiments generally show that the proposed filter designs have a number of advantages over previous designs and static designs alike when applied to the problem of separating periodic signals. Among the proposed designs, SF-APES and FB-APES appear to perform the best and equally well while SF-APES (appx) is sometimes slightly worse.

F. Some Speech and Audio Examples

We will now demonstrate the applicability of the proposed methods to real signals. In the experiments to follow, we will use the SF-APES design. In the first such experiment, we will use the filters obtained using the said method to extract a real trumpet signal, a single tone sampled at ~ 8 kHz using 50 ms segments and a filter length of 100 and the filter is updated every 5 ms. Note that both the signal and the filters are complex by mapping the input signal to its analytic counterpart using the Hilbert transform. For each segment the fundamental frequency and the model order was found using the approximate non-linear least squares method of [27] and the optimal filter was updated every 1 ms. The single tone has been buried in noise at an SNR of 0 dB and interfering tones, which were also trumpet tones (both signals are from the SQAM database [41]), have been added with an SIR of -10 dB. The spectrogram of the original signal is shown in Figure 5(a) and the same signal with noise and interference added is depicted in Figure 5(b). The spectrogram of the extracted signal is shown in Figure 5(c). These figures clearly demonstrate the ability of the APES-like designs to extract the signal while rejecting not only noise, but also strong periodic interference even when these are fairly close to the harmonics of the desired signal. Note that for this particular example, because the SIR and SNR are quite low, the FS-Capon method would also perform quite well.

Regarding the application of the proposed filters to speech signals, an interesting question is whether the filters are suitable for such signals, as they exhibit non-stationarity. To address this question, we apply the SF-APES method to a voiced speech signal, this particular signal being from the SQAM database [41] and sampled at 11025 Hz. As with the prior example, we estimate the pitch for each segment, which are here of size 30 ms (corresponding to 165 complex samples), a size commonly used in speech processing and coding. From these segments, the optimal filterbank is then also determined using the estimated pitch. In this example, the complex filters of length 40 are updated every 2.5 ms. The signal is depicted in Figure 6(a) and the extracted signal is shown in Figure 6(b). The difference between the original signal and the extracted one is shown in Figure 6(c) and the estimated pitch is shown in Figure 6(d). A number of observations can be made regarding the original signal. Firstly, it is non-stationary at the beginning and the end with a time-varying envelope, and the pitch can be observed to vary too. It can, however, be observed from the extracted signal and the corresponding error signal that the filters are indeed able to track this signal, resulting in an SDR of 20 dB. This demonstrates that the filters may be useful even if the signal is not completely stationary.

Our final example involves the separation of two speech signals, more specifically two quasi-stationary segments of voiced speech mixed at an SIR of 0 dB. These signals are sampled at 8 kHz and are from the EUROM.1 corpus [42]. As before 30 ms segments are used for determining the pitch and the optimal filters resulting in segments consisting of 120 complex samples along with filters of length 30. We here update the filters every 2.5 ms. In Figures 7(a) and 7(b), the two signals are shown along with their mixture in Figure 7(c). As before, the fundamental frequencies of the two sources are estimated with the approximate non-linear least squares method [27], and the resulting estimates are shown in Figure 7(d). It can be seen that one source has an average pitch of approximately 162 Hz while that of the other is about 200 Hz. The two extracted signals are shown in Figures 7(e) and 7(f), respectively. As can be seen, the filters are able to separate the signals achieving SDRs of 14 and 12 dB, respectively. Of course, some errors occur, as can also clearly be seen, as parts of the other interfering source will be passed by the filters.

VI. Conclusion

In this paper, new filter designs for extracting and separating periodic signals have been proposed, a problem occurring frequently in, for example, speech and audio processing. The proposed filters are designed such that they have unit gain at the frequencies of the harmonics of the desired signal and suppress everything else. The novel part of the present designs is that they are optimized for having an output that is

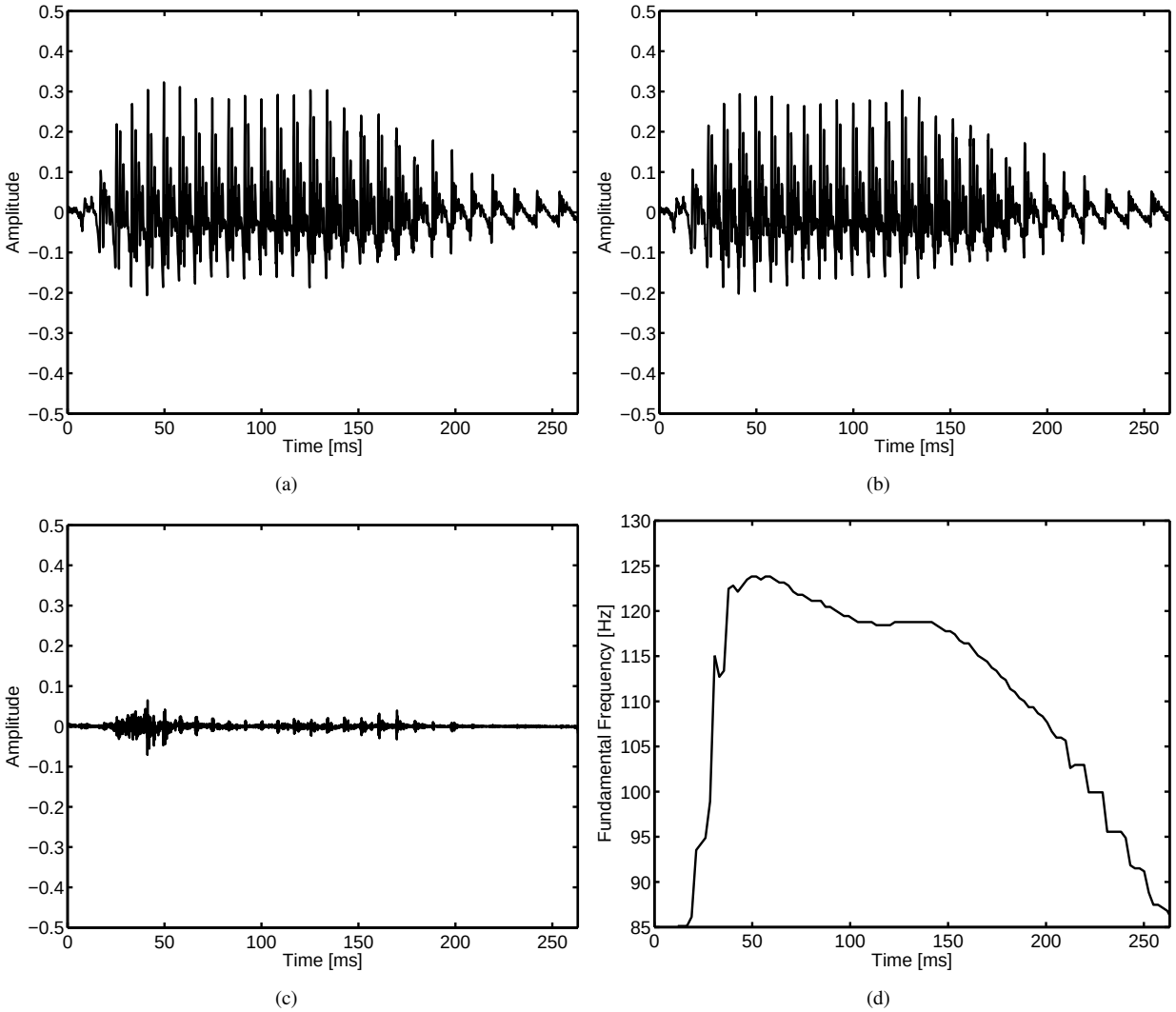


Fig. 6. Shown are: (a) the original voiced speech signal, (b) the extracted signal, (c) the difference between the two signals, i.e., the part of the signal that was not extracted, and (d) the estimated pitch used in the filters.

approximately periodic too. Additionally, the obtained filters are optimal for a segment of the observed signal and are thus signal-adaptive. The filter designs can be used not only for the aforementioned applications but also for estimating the parameters of periodic signals. The designs have been demonstrated to overcome the shortcomings of previous designs while retaining their desirable properties, like the ability to cancel out interfering signals. We have shown how the new designs reduce to a number of well-known designs under certain conditions and they can thus be seen as generalizations of previous methods. In simulations, we have demonstrated the superior performance of the obtained filters in enhancement and separation applications.

REFERENCES

[1] American Standards Association (ASA), "Acoustical Terminology, SI, 1-1960," 1960.

[2] D. Chazan, Y. Stettiner, and D. Malah, "Optimal multi-pitch estimation using the EM algorithm for co-channel speech separation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 27–30 April 1993, vol. 2, pp. 728–731.

[3] R. J. McAulay and T. F. Quatieri, "Sinusoidal coding," in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds., chapter 4, pp. 121–174. Elsevier Science B.V., 1995.

[4] B. Kostek, "Musical instrument classification and duet analysis employing music information retrieval techniques," *Proc. IEEE*, vol. 92, no. 4, pp. 712–729, Apr 2004.

[5] J. H. Jensen, M. G. Christensen, D. P. W. Ellis, and S. H. Jensen, "Quantitative analysis of a common audio similarity measure," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 17(4), pp. 693–703, May 2009.

[6] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 9(10), pp. 2009–2025, Oct. 1998.

[7] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley, 2001.

[8] A. Nehorai and B. Porat, "Adaptive comb filtering for harmonic signal enhancement," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34(5), pp. 1124–1138, Oct. 1986.

[9] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "Energy Separation in Signal Modulations with Application to Speech Analysis," *IEEE Trans. Signal Process.*, vol. 41(10), pp. 3024–3051, Oct. 1993.

[10] M.-Y. Zou, C. Zhenming, and R. Unbehauen, "Separation of periodic signals by using an algebraic method," in *Proc. IEEE Int. Symp. Circuits and Systems*, 1991, vol. 5, pp. 2427–2430.

[11] B. Santhanam and P. Maragos, "Demodulation of Discrete Multicomponent AM-FM Signals using Periodic Algebraic Separation and Energy Demodulation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1997.

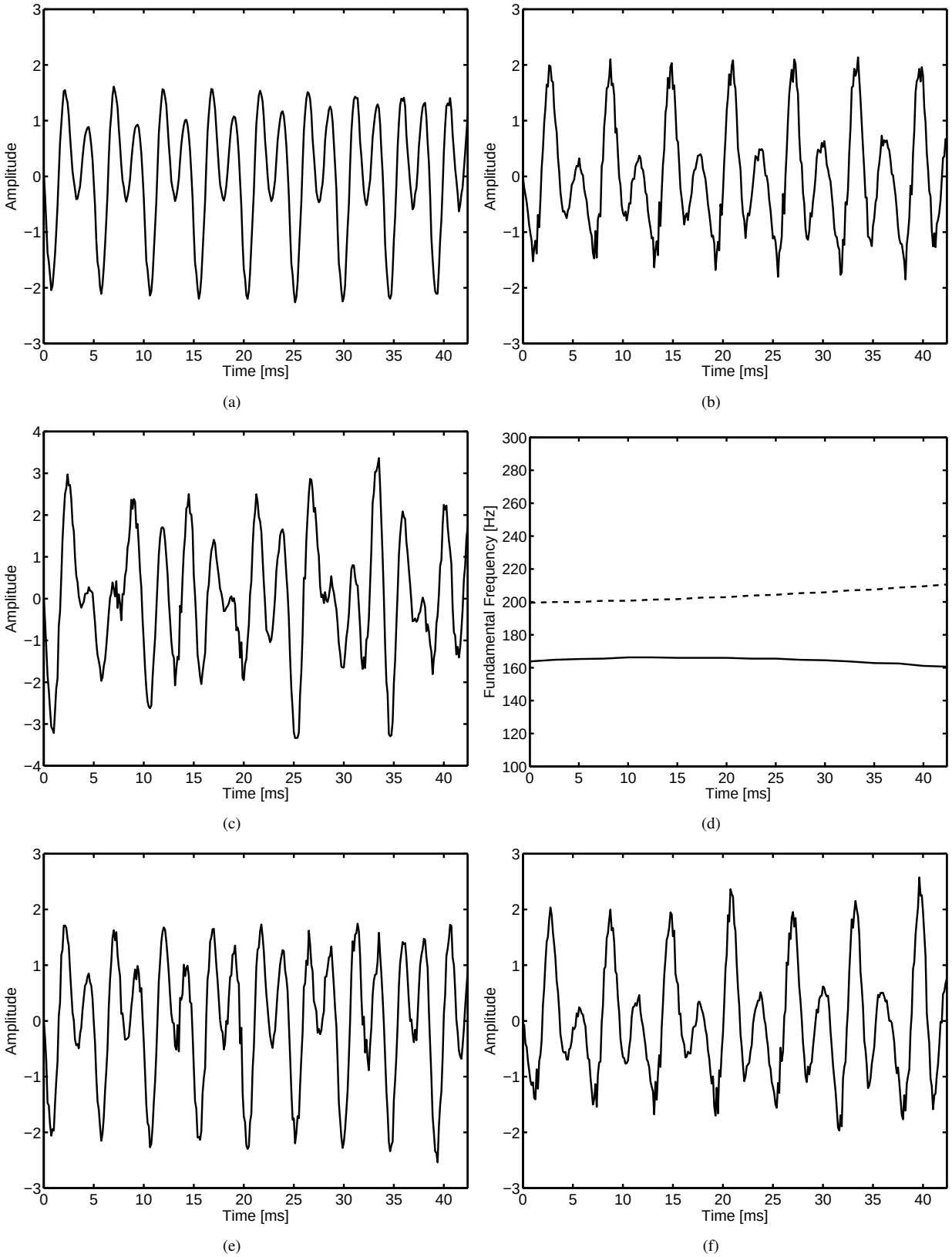


Fig. 7. Shown are the following signals: (a) voiced speech signal of source 1, (b) voiced speech signal of source 2, (c) the mixture of the two signals, (d) the estimated pitch tracks for source 1 (dashed) and 2 (solid), (e) the estimate of source 1 obtained from the mixture, and (f) the estimate of source 2 extracted from the mixture.

[12] P. Stoica, H. Li, and J. Li, "Amplitude estimation of sinusoidal signals: Survey, new results and an application," *IEEE Trans. Signal Process.*, vol. 48(2), pp. 338–352, Feb. 2000.

[13] K. W. Chan and H. C. So, "Accurate frequency estimation for real

harmonic sinusoids," *IEEE Signal Process. Lett.*, vol. 11(7), pp. 609–612, July 2004.
[14] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Am.*, vol. 111(4), pp. 1917–1930, Apr. 2002.
[15] M. G. Christensen, S. H. Jensen, S. V. Andersen, and A. Jakobsson, "Subspace-based fundamental frequency estimation," in *Proc. European Signal Processing Conf.*, 2004, pp. 637–640.
[16] 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(5), pp. 1635–1644, July 2007.
[17] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Fundamental frequency estimation using the shift-invariance property," in *Rec. Asilomar Conf. Signals, Systems, and Computers*, 2007, pp. 631–635.
[18] H. Li, P. Stoica, and J. Li, "Computationally efficient parameter estimation for harmonic sinusoidal signals," *Signal Processing*, vol. 80, pp. 1937–1944, 2000.
[19] M. G. Christensen, P. Stoica, A. Jakobsson, and S. H. Jensen, "Multi-pitch estimation," *Elsevier Signal Processing*, vol. 88(4), pp. 972–983, Apr. 2008.
[20] M. Noll, "Pitch determination of human speech by harmonic product spectrum, the harmonic sum, and a maximum likelihood estimate," in *Proc. Symposium on Computer Processing Communications*, 1969, pp. 779–797.
[21] A. M. Noll, "Cepstrum pitch determination," *J. Acoust. Soc. Am.*, vol. 41(2), pp. 293–309, 1967.
[22] A. T. Cemgil, H. J. Kappen, and D. Barber, "A generative model for music transcription," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 2, pp. 679–694, March 2006.
[23] A. T. Cemgil, *Bayesian Music Transcription*, Ph.D. thesis, Nijmegen University, 2004.
[24] S. Godsill and M. Davy, "Bayesian harmonic models for musical pitch estimation and analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2002, vol. 2, pp. 1769–1772.
[25] J. Moorer, "The optimum comb method of pitch period analysis of continuous digitized speech," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 22, no. 5, pp. 330–338, Oct 1974.
[26] J. Lim, A. Oppenheim, and L. Braida, "Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 26, no. 4, pp. 354–358, Aug 1978.
[27] M. G. Christensen and A. Jakobsson, *Multi-Pitch Estimation*, vol. 5 of *Synthesis Lectures on Speech & Audio Processing*, Morgan & Claypool Publishers, 2009.
[28] C. Fevotte and S. J. Godsill, "A Bayesian approach for blind separation of sparse sources," *IEEE Trans. Signal Process.*, vol. 14(6), pp. 2174–2188, Nov. 2006.
[29] J. Li and P. Stoica, "An adaptive filtering approach to spectral estimation and SAR imaging," *IEEE Trans. Signal Process.*, vol. 44(6), pp. 1469–1484, June 1996.
[30] P. Stoica, H. Li, and J. Li, "A new derivation of the APES filter," *IEEE Signal Process. Lett.*, vol. 6(8), pp. 205–206, Aug. 1999.
[31] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Pearson Prentice Hall, 2005.
[32] P. M. Djuric, "Asymptotic MAP criteria for model selection," *IEEE Trans. Signal Process.*, vol. 46, pp. 2726–2735, Oct. 1998.
[33] M. G. Christensen, J. H. Jensen, A. Jakobsson, and S. H. Jensen, "Joint fundamental frequency and order estimation using optimal filtering," in *Proc. European Signal Processing Conf.*, 2009, pp. 1334–1338.
[34] M. G. Christensen, J. H. Jensen, A. Jakobsson, and S. H. Jensen, "On optimal filter designs for fundamental frequency estimation," *IEEE Signal Process. Lett.*, vol. 15, pp. 745–748, 2008.
[35] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57(8), pp. 1408–1418, 1969.
[36] P. C. Loizou, *Speech Enhancement: Theory and Practice*, CRC Press, 2007.
[37] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 9(5), pp. 504–512, June 2001.
[38] O. Yilmaz and S. Reckard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Signal Process. Lett.*, vol. 52(7), July 2004.
[39] N. Bertin, R. Badeau, and E. Vincent, "Enforcing harmonicity and smoothness in Bayesian non-negative matrix factorization applied to polyphonic music transcription," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 18(3), pp. 538–549, Mar. 2010.
[40] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Signal Process.*, vol. 14(4), pp. 1462–1469, July 2006.
[41] European Broadcasting Union, *Sound Quality Assessment Material Recordings for Subjective Tests*, EBU, Apr. 1988, Tech. 3253.
[42] Center for PersonKommunikation, *Documentation of the Danish EU-ROM.1 Database*, Institute of Electronic Systems, Aalborg University, 1995.

Mads Græsbøll Christensen Mads Græsbøll Christensen was born in Copenhagen, Denmark in March 1977. He received the M.Sc. and Ph.D. degrees in 2002 and 2005, respectively, from Aalborg University in Denmark, where he is also currently employed at the Department of Architecture, Design and Media Technology as Associate Professor. He was previously with the Department of Electronic Systems, Aalborg University and has been a visiting researcher at Philips Research Labs, Ecole Nationale Supérieure des Télécommunications (ENST), and Columbia University. Dr. Christensen has received several awards, namely an IEEE Int. Conf. Acoust. Speech, and Signal Proc. Student Paper Contest Award, the Spar Nord Foundation's Research Prize awarded annually for an excellent Ph.D. thesis, and a Danish Independent Research Council's Young Researcher's Award. He has published more than 75 papers in peer-reviewed conference proceedings and journals is author (with A. Jakobsson) of the book "Multi-Pitch Estimation", Morgan & Claypool Publishers, 2009. He is a Member of the IEEE and an Associate Editor for the IEEE Signal Processing Letters. His research interests include digital signal processing theory and methods with application to speech and audio, in particular parametric analysis, modeling, and coding.

Andreas Jakobsson received his M.Sc. from Lund Institute of Technology and his Ph.D. in Signal Processing from Uppsala University in 1993 and 2000, respectively. Since, he has held positions with Global IP Sound AB, the Swedish Royal Institute of Technology, King's College London, and Karlstad University, as well as an Honorary Research Fellowship at Cardiff University. He is currently Professor of Mathematical Statistics at Lund University, Sweden. He has published his research findings in over 100 refereed journal and conference papers, and has filed four patents. He has also co-authored (together with M. G. Christensen) a recent book on multi-pitch estimation (Morgan & Claypool, 2009). He is a Senior Member of IEEE, a member of the IEEE Sensor Array and Multichannel (SAM) Signal Processing Technical Committee, and an Associate Editor for the IEEE Transactions on Signal Processing, the IEEE Signal Processing Letters and the Research Letters in Signal Processing. His research interests include statistical and array signal processing, detection and estimation theory, and related application in remote sensing, telecommunication and biomedicine.