AALBORG UNIVERSITY

DENMARK

Joint Fundamental Frequency and Order Estimation using Optimal Filtering

Christensen, Mads Græsbøll; Jensen, Jesper Højvang; Jakobsson, Andreas; Jensen, Søren Holdt

Link to publication from Aalborg University

RESEARCH   **Open Access**

Joint fundamental frequency and order estimation using optimal filtering

Mads Græsbøll Christensen[1*], Jesper Lisby Højvang[3], Andreas Jakobsson[2] and Søren Holdt Jensen[3]

Abstract

In this paper, the problem of jointly estimating the number of harmonics and the fundamental frequency of periodic signals is considered. We show how this problem can be solved using a number of methods that either are or can be interpreted as filtering methods in combination with a statistical model selection criterion. The methods in question are the classical comb filtering method, a maximum likelihood method, and some filtering methods based on optimal filtering that have recently been proposed, while the model selection criterion is derived herein from the maximum a posteriori principle. The asymptotic properties of the optimal filtering methods are analyzed and an order-recursive efficient implementation is derived. Finally, the estimators have been compared in computer simulations that show that the optimal filtering methods perform well under various conditions. It has previously been demonstrated that the optimal filtering methods perform extremely well with respect to fundamental frequency estimation under adverse conditions, and this fact, combined with the new results on model order estimation and efficient implementation, suggests that these methods form an appealing alternative to classical methods for analyzing multi-pitch signals.

Introduction

Periodic signals can be characterized by a sum of sinusoids, each parametrized by an amplitude, a phase, and a frequency. The frequency of each of these sinusoids, sometimes referred to as harmonics, is an integer multiple of a fundamental frequency. When observed, such signals are commonly corrupted by observation noise, and the problem of estimating the fundamental frequency from such observed signals is referred to as fundamental frequency, or pitch, estimation. Some signals contain many such periodic signals, in which case the problem is referred to as multi-pitch estimation, although this is somewhat of an abuse of terminology, albeit a common one, as the word pitch is a perceptual quality, defined more specifically for acoustical signals as "that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale" [1]. In most cases, the fundamental frequency and pitch are related in a simple manner and the terms are, therefore, often used synonymously. The problem under investigation here is that of estimating the fundamental

frequencies of periodic signals in noise. It occurs in many speech and audio applications, where it plays an important role in the characterization of such signals, but also in radar and sonar. Many different methods have been invented throughout the years to solve this problem, with some examples being the following: linear prediction [2], correlation [3-7], subspace methods [8-10], frequency fitting [11], maximum likelihood [12-16], cepstral methods [17], Bayesian estimation [18-20], and comb filtering [21-23]. Note that several of the listed methods can be interpreted in several ways, as we will also see examples of in this paper. For a general overview of pitch estimation methods, we refer the interested reader to [24].

The scope of this paper is filtering methods with application to estimation of the fundamental frequencies of multiple periodic signals in noise. First, we state the problem mathematically in Sect. II and introduce some useful notation and results after which we present, in Sect. III, some classical methods for solving the aforementioned problem. These are intimately related to the methods under consideration in this paper. Then, we present our optimal filter designs in Sect. IV. This work has recently been published by the authors [16,25]. These designs are generalizations of Capon's classical

* Correspondence: mgc@create.aau.dk
[1]Department of Architecture, Design and Media Technology, Aalborg University, Aalborg, Denmark
Full list of author information is available at the end of the article

optimal beamformer and are not novel to this paper, but the key aspects of this paper are based on them. The resulting filters are signal-adaptive and optimal in the sense that they minimize the output power while passing the harmonics of a periodic signal undistorted, and they have been demonstrated to have excellent performance for parameter estimation under adverse conditions [16]. Especially their ability to reject interfering periodic components is remarkable and important as it leads to a natural decoupling of the fundamental frequency estimation problem for multiple sources, a problem that otherwise involves multi-dimensional non-linear optimization. However, also the resulting filters' ability to adapt to the noise statistics without prior knowledge of these is worth noting. We also note that the filter designs along with related methods have been proven to work well for enhancement and separation of periodic signals [26]. After the presentation of the filters, an analysis of the properties of the optimal filtering methods follows in Sect. V which reveals some valuable insights. It should be noted that the first part of this analysis appeared also in [25], in a very brief form, but we here repeat it for completeness along with some additional details and information. It was shown in [9] that it is not only necessary for a fundamental frequency estimator to be optimal to also estimate the number of harmonics, but it is in fact also necessary to avoid ambiguities in the cost functions, something that is often the cause of spurious estimates at rational values of the fundamental frequency for single pitch estimation. In Sect. VI, we derive an order estimation criterion specifically for the signal model used through this paper, and, in Section VII, we show how to use this criterion in combination with the filtering methods. This order estimation criterion is based on the maximum a posteriori principle following [27]. Compared to traditional methods such as the comb filtering method [23] and maximum likelihood methods [12,16], the optimal filtering methods suffer from a high complexity, requiring that operations of cubic complexity be performed for each candidate fundamental frequency and order. Indeed, this complexity may be prohibitive for many applications and to address this, we derive an exact order-recursive fast implementation of the optimal filtering methods in Sect. VIII. Finally, we present some numerical results in Sect. IX, comparing the performance of the estimators to other state-of-the-art estimators before concluding on our work in Sect. X.

Preliminaries

A signal containing a number of periodic components, termed sources, consists of multiple sets of complex sinusoids having frequencies that are integer multiples of a set of fundamental frequencies, $\{\omega_k\}$, and additive noise. Such a signal can be written for $n = 0, ..., N - 1$ as

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

where $a_{k,l} = A_{k,l} e^{j\phi_{k,l}}$ is the complex-valued amplitude of the lth harmonic of the source, indexed by k, and $e_k(n)$ is the noise associated with the kth source which is assumed to be zero-mean and complex. The complex-valued amplitude is composed of a real, non-zero amplitude $A_{k,l} > 0$ and a phase $\varphi_{k,l}$. The number of sinusoids, L_k, is referred to as the order of the model and is often considered known in the literature. However, this is often not the case for speech and audio signals, where the number of harmonics can be observed to vary over time. Furthermore, for some signals, the frequencies of the harmonics will not be exact integer multiples of the fundamental. There exists several modified signal models for dealing with this (e.g., [24,28-32]), but this is beyond the scope of this paper and we will defer from any further discussion of this.

We refer to signals of the form (1) as multi-pitch signals and the model as the multi-pitch model. The special case with $K = 1$ is referred to as a single-pitch signal. The methods under consideration can generally be applied to multi-pitch signals (and will be in the experiments), but when we wish to emphasize that the derivations strictly speaking only hold for single-pitch signal, those will be based on $x_k(n)$ and related quantities. It should be noted that even if a recording is only of a single instrument, the signal may still be multi-pitch as only a few instruments are monophonic. Room reverberation may also cause the observed signal to consist of several different tones at a particular time instance.

We define a sub-vector as consisting of M, with $M \leq N$ (we will introduce more strict bounds later), time-reversed samples of the observed signal, as

$$\mathbf{x}(n) = [\, x(n) \; x(n-1) \; \cdots \; x(n-M+1) \,]^{\mathrm{T}}, \tag{2}$$

where $(\cdot)^{\mathrm{T}}$ denotes the transpose, and similarly for the sources $x_k(n)$ and the noise $e_k(n)$. Next, we define a Vandermonde matrix $\mathbf{Z}_k \in \mathbb{C}^{M \times L_k}$, which is constructed from a set of L_k harmonics, each defined as

$$\mathbf{z}(\omega) = [\, 1 \; e^{-j\omega} \; \cdots \; e^{-j\omega(M-1)} \,]^{\mathrm{T}}, \tag{3}$$

leading to the matrix

$$\mathbf{Z}_k = [\mathbf{z}(\omega_k) \; \cdots \; \mathbf{z}(\omega_k L_k)], \tag{4}$$

and a vector containing the corresponding complex amplitudes as $\mathbf{a}_k = [a_{k,1} \; \cdots \; a_{k,L_k}]^{\mathrm{T}}$. Introducing the following matrix

$$\mathbf{D}_n = \begin{bmatrix} e^{-j\omega_k 1 n} & & 0 \\ & \ddots & \\ 0 & & e^{-j\omega_k L_k n} \end{bmatrix}, \tag{5}$$

the vectorized model in (1) can be expressed as

$$\mathbf{x}(n) = \sum_{k=1}^{K} \mathbf{Z}_k \mathbf{D}_n \mathbf{a}_k + \mathbf{e}_k(n) \tag{6}$$

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

It can be seen that the complex amplitudes can be thought of as being time-varying, i.e., $\mathbf{a}_k(n) = \mathbf{D}_n \mathbf{a}_k$. Note that it is also possible to define the signal model such that the Vandermonde matrix is time-varying.

In the remainder of the text, we will make extensive use of the covariance matrix of the sub-vectors. Let E {·} and $(\cdot)^H$ denote the statistical expectation operator and the conjugate transpose, respectively. The covariance matrix is then defined as

$$\mathbf{R} = \mathrm{E}\{\mathbf{x}(n)\mathbf{x}^H(n)\}, \tag{8}$$

and similarly we define $\mathbf{R}_k$ for $\mathbf{x}_k(n)$. Assuming that the various sources are statistically independent, the covariance matrix of the observed signal can be written as

$$\mathbf{R} = \sum_{k=1}^{K} \mathbf{Z}_k \mathrm{E}\left\{\mathbf{a}_k(n)\mathbf{a}_k^H(n)\right\}\mathbf{Z}_k^H + \mathrm{E}\left\{\mathbf{e}_k(n)\mathbf{e}_k^H(n)\right\} \tag{9}$$

$$= \sum_{k=1}^{K} \mathbf{Z}_k \mathbf{P}_k \mathbf{Z}_k^H + \mathbf{Q}_k, \tag{10}$$

where the matrix $\mathbf{P}_k$ is the covariance matrix of the amplitudes, which is defined as

$$\mathbf{P}_k = \mathrm{E}\left\{\mathbf{a}_k(n)\mathbf{a}_k^H(n)\right\}. \tag{11}$$

For statistically independent and uniformly distributed phases (on the interval $(-\pi, \pi]$), this matrix reduces to the following (see [33]):

$$\mathbf{P}_k = \mathrm{diag}\left([A_{k,1}^2 \ \cdots \ A_{k,L}^2]\right), \tag{12}$$

with diag (·) being an operator that generates a diagonal matrix from a vector. Furthermore, $\mathbf{Q}_k$ is the covariance matrix of the noise $\boldsymbol{e}_k(\boldsymbol{n})$, i.e., $\mathbf{Q}_k = \mathrm{E}\left\{e_k(n)e_k^H(n)\right\}$. The sample covariance matrix, defined as

$$\widehat{\mathbf{R}} = \frac{1}{N-M+1} \sum_{n=0}^{N-M} \mathbf{x}(n)\mathbf{x}^H(n), \tag{13}$$

is used as an estimate of the covariance matrix. It should be stressed that for $\widehat{\mathbf{R}}$ to be invertible, we require that $M < \dfrac{N}{2} + 1$. Throughout the text, we generally assume that M is chosen proportionally to N, something that is essential to the consistency of the proposed estimators.

Classical methods
Comb filter
One of the oldest methods for pitch estimation is the comb filtering method [21,22], which is based on the following ideas. Mathematically, we can express periodicity as $x(n) \approx x(n - D)$ where D is the repetition or pitch period. From this observation it follows that we can measure the extent to which a certain waveform is periodic using a metric on the error $e(n)$, defined as $e(n) = x(n) - \alpha x(n - D)$. The Z-transform of this is $E(z) = X(z)(1 - \alpha z^{-D})$. This shows that the matching of a signal with a delayed version of itself can be seen as a filtering process, where the output of the filter is the modeling error $e(n)$. This can of course also be seen as a prediction problem, only the unknowns are not just the filter coefficient α but also the lag D. If the pitch period is exactly D, the output error is just the observation noise. Usually, however, the comb filter is not used in this form as it is restricted to integer pitch periods and is rather inefficient in several ways. Instead, one can derive more efficient methods based on notch filters [23]. Notch filters are filters that cancel out, or, more correctly, attenuate signal components at certain frequencies. Periodic signals can be comprised of a number of harmonics, for which reason we use L_k such filters having notches at frequencies $\{\psi_i\}$. Such a filter can be factorized into the following form

$$P(z) = \prod_{i=1}^{L_k} \left(1 - e^{j\psi_i}z^{-1}\right), \tag{14}$$

i.e., consisting of a polynomial that has zeros on the unit circle at angles corresponding to the desired frequencies. From this, one can define a polynomial $P(\rho^{-1}z) = \prod_{i=1}^{L_k} \left(1 - \rho e^{j\psi_i}z^{-1}\right)$ where $0 < \rho < 1$ is a parameter that leads to poles located inside the unit circle at the same angles as the zeros of $P(z)$ but at a distance of ρ from the origin. ρ is typically in the range 0.95-0.995 [23]. For our purposes, the desired frequencies are given by $\psi_l = \omega_k l$, where ω_k is considered an unknown parameter. As a consequence, the zeros of (14) are distributed uniformly on the unit circle in the z-plane. By combining $P(z)$ and $P(\rho^{-1}z)$, we obtain the following filter:

$$H(z) = \frac{P(z)}{P(\rho^{-1}z)} = \frac{1 + \beta_1 z^{-1} + \ldots + \beta_{L_k} z^{-L_k}}{1 + \rho \beta_1 z^{-1} + \ldots + \rho^{L_k} \beta_{L_k} z^{-L_k}}, (15)$$

where $\{\beta_l\}$ are the complex filter coefficients that result from expanding (14). This filter can be used by filtering the observed signal $x(n)$ for various candidate fundamental frequencies to obtain the filtered signal $e(n)$ where the harmonics have been attenuated. This can also be expressed as $E(z) = X(z)H(z)$ which results in the following difference function:

$$\begin{aligned} e(n) &= x(n) + \beta_1 x(n-1) + \ldots + \beta_{L_k} x(n - L_k) \\ &\quad - \rho \beta_1 e(n-1) - \ldots - \rho^{L_k} \beta_{L_k} e(n - L_k). \end{aligned} \quad (16)$$

By imposing a metric on $e(n)$ and considering the fundamental frequency to be an unknown parameter, we obtain the estimator

$$\hat{\omega}_k = \arg \min_{\omega} \sum_{n=1}^{N} |e(n)|^2, \quad (17)$$

from which ρ can also be found in a similar manner as done in [23], if desired. In [23], this is performed in a recursive manner given an initial fundamental frequency estimate, leading to a computationally efficient scheme that can be used for either suppressing or extracting the periodic signal from the noisy signal.

Maximum likelihood estimator
Perhaps the most commonly used methodology in estimators is maximum likelihood. Interestingly, the maximum likelihood estimator for white Gaussian noise can also be interpreted as a filtering method when applied to the pitch estimation problem. First, we will briefly present the maximum likelihood pitch estimator. For an observed signal x_k with M=N (note that we have omitted the dependency on n for this special case) consisting of white Gaussian noise and one source, the log-likelihood function $\ln f(\mathbf{x}_k|\omega_k, \mathbf{a}_k, \sigma_k^2)$ is given by

$$\ln f(\mathbf{x}|\omega_k, \mathbf{a}_k, \sigma_k^2) = -N \ln \pi - N \ln \sigma_k^2 - \frac{1}{\sigma_k^2} \|\mathbf{x}_k - \mathbf{Z}_k \mathbf{a}_k\|_2^2, \ (18)$$

By maximizing (18), the maximum likelihood estimates of ω_k, $\mathbf{a}_k$, and, σ_k^2 are obtained. The expression can be seen to depend on the unknown noise variance σ_k^2 and the amplitudes $\mathbf{a}_k$, both of which are of no interest to us here. To eliminate this dependency, we proceed as follows. Given ω_k and L_k, the maximum likelihood estimate of the amplitudes is obtained as

$$\widehat{\mathbf{a}}_k = (\mathbf{Z}_k^H \mathbf{Z}_k)^{-1} \mathbf{Z}_k^H \mathbf{x}_k \quad (19)$$

and the noise variance as

$$\hat{\sigma}_k^2 = \frac{1}{N} \|\mathbf{x}_k - \mathbf{\Pi}_Z \mathbf{x}_k\|_2^2. \quad (20)$$

The matrix $\mathbf{\Pi}_Z$ in (20) is the projection matrix which can be approximated as

$$\lim_{N \to \infty} N \mathbf{\Pi}_Z = \lim_{N \to \infty} N \mathbf{Z}_k (\mathbf{Z}_k^H \mathbf{Z}_k)^{-1} \mathbf{Z}_k^H = \mathbf{Z}_k \mathbf{Z}_k^H. \quad (21)$$

This is essentially because the columns of $\mathbf{Z}_k$ are complex sinusoids that are asymptotically orthogonal. Using this approximation, the noise variance estimate can be simplified significantly, i.e.,

$$\hat{\sigma}_k^2 \approx \frac{1}{N} \|\mathbf{x}_k - \frac{1}{N} \mathbf{Z}_k \mathbf{Z}_k^H \mathbf{x}_k\|_2^2, \quad (22)$$

which leaves us with a log-likelihood function that depends only on the fundamental frequency. We can now express the maximum likelihood pitch estimator as

$$\hat{\omega}_k = \arg \max_{\omega_k} \ln f(\mathbf{x}_k|\omega_k, \hat{\mathbf{a}}_k, \hat{\sigma}_k^2) \quad (23)$$

$$= \arg \max_{\omega_k} \|\mathbf{Z}_k^H \mathbf{x}_k\|_2^2. \quad (24)$$

Curiously, the last expression can be rewritten into a different form that leads to a familiar estimator:

$$\|\mathbf{Z}_k^H \mathbf{x}_k\|_2^2 = \sum_{l=1}^{L_k} |\sum_{n=0}^{N-1} x_k(n) e^{-j\omega_k l n}|^2 \quad (25)$$

$$\triangleq \sum_{l=1}^{L_k} |X_k(\omega_k l)|^2, \quad (26)$$

which shows that harmonic summation methods [12,34] are in fact approximate maximum likelihood methods under certain conditions. We note that it can be seen from these derivations that, under the aforementioned conditions, the minimization of the 2-norm leads to the maximum likelihood estimates. Since the fundamental frequency is a nonlinear parameter, this approach is sometimes referred to as the nonlinear least-squares (NLS) method.

Next, we will show that the approximate maximum likelihood estimator can also be seen as a filtering method. First, we introduce the output signal $y_{k,l}(n)$ of the lth filter for the kth source having coefficients $h_{k,l}(n)$ as

$$y_{k,l}(n) = \sum_{m=0}^{M-1} h_{k,l}(m) x(n-m) = \mathbf{h}_{k,l}^H \mathbf{x}_k(n), \quad (27)$$

with $\mathbf{h}_{k,l}$ being a vector containing the filter coefficients of the lth filter, i.e.,

$$\mathbf{h}_{k,l} = [h_{k,l}(0) \ \cdots \ h_{k,l}(M-1)]^{\mathrm{H}}. \tag{28}$$

The output power of the lth filter can be expressed in terms of the covariance matrix $\mathbf{R}_k$ as

$$\mathrm{E}\left\{|\gamma_{k,l}(n)|^2\right\} = \mathrm{E}\left\{\mathbf{h}_{k,l}^{\mathrm{H}}\mathbf{x}_k(n)\mathbf{x}_k^{\mathrm{H}}(n)\mathbf{h}_{k,l}\right\} \tag{29}$$

$$= \mathbf{h}_{k,l}^{\mathrm{H}}\mathbf{R}_k\mathbf{h}_{k,l}. \tag{30}$$

The total output power of all the filters is thus given by

$$\sum_{l=1}^{L_k} \mathrm{E}\{|\gamma_{k,l}(n)|^2\} = \sum_{l=1}^{L_k} \mathbf{h}_{k,l}^{\mathrm{H}}\mathbf{R}_k\mathbf{h}_{k,l} \tag{31}$$

$$= \mathrm{Tr}\left[\mathbf{H}_k^{\mathrm{H}}\mathbf{R}_k\mathbf{H}_k\right], \tag{32}$$

where $\mathbf{H}_k$ is a filterbank matrix containing the individual filters, i.e., $\mathbf{H}_k = [\mathbf{h}_{k,1} \ \cdots \ \mathbf{h}_{k,L_k}]$ and $\mathrm{Tr}[\cdot]$ denotes the trace. The problem at hand is then to choose or design a filter or a filterbank. Suppose we construct the filters from finite length complex sinusoids as

$$\mathbf{h}_{k,l} = \left[e^{-j\omega_k l 0} \cdots e^{-j\omega_k l(M-1)}\right]^T, \tag{33}$$

which is the same as the vector $\mathbf{z}(\omega_k l)$ defined earlier. The matrix $\mathbf{H}_k$ is therefore also identical to the Vandermonde matrix $\mathbf{Z}_k$. Then, we may express the total output power of the filterbank as

$$\mathrm{Tr}\left[\mathbf{H}_k^{\mathrm{H}}\mathbf{R}_k\mathbf{H}_k\right] = \mathrm{Tr}\left[\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}\mathbf{Z}_k\right] \tag{34}$$

$$= \mathrm{E}\left\{\left\|\mathbf{Z}_k^{\mathrm{H}}\mathbf{x}_k(n)\right\|_2^2\right\}. \tag{35}$$

This shows that by replacing the expectation operator by a finite sum over the realizations $\mathbf{x}_k(n)$, we get the approximate maximum likelihood estimator, only we average over the sub-vectors $\mathbf{x}_k(n)$. By using only one sub-vector of length N, leaving us with just a single observed sub-vector, the method becomes asymptotically equivalent (in N) to the NLS method and, therefore, the maximum likelihood method for white Gaussian noise. For more on the relation between various spectral estimators and filterbank methods, we refer the interested reader to [33,35].

Optimal filter designs

We will now delve further into signal-adaptive and optimal filters and in doing so we will make use of the notation and definitions of the previous section. Two desirable properties of a filterbank for our application is that the individual filters pass power undistorted at specific frequencies, here integer multiples of the fundamental frequency, while minimizing the power at all other frequencies. This problem can be stated

mathematically as the following quadratic constrained optimization problem:

$$\min_{\mathbf{H}_k} \mathrm{Tr}\left[\mathbf{H}_k^{\mathrm{H}}\mathbf{R}\mathbf{H}_k\right] \text{ s.t. } \mathbf{H}_k^{\mathrm{H}}\mathbf{Z}_k = \mathbf{I}. \tag{36}$$

Here, $\mathbf{I}$ is the $L_k \times L_k$ identity matrix. The matrix constraints specify that the Fourier transforms of the filterbank should have unit gain at the lth harmonic frequency and zero for the others. Using the method of Lagrange multipliers, we obtain that the filter bank matrix $\mathbf{H}_k$ solving (36) is (see [16] for details)

$$\mathbf{H}_k = \mathbf{R}^{-1}\mathbf{Z}_k(\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}_k^{-1}\mathbf{Z}_k)^{-1}, \tag{37}$$

which is a data and fundamental frequency dependent filter bank. It can be used to estimate the fundamental frequency by evaluating the output power of the filterbank for a set of candidate fundamental frequencies, i.e.,

$$\hat{\omega}_k = \arg\max_{\omega_k} \mathrm{Tr}\left[\left(\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k\right)^{-1}\right]. \tag{38}$$

Suppose that instead of designing a filterbank, we design a single filter for the kth source, $\mathbf{h}_k$ that passes the signal undistorted at the harmonic frequencies while otherwise minimizing the output power. This problem can be stated mathematically as

$$\min_{\mathbf{h}_k} \mathbf{h}_k^{\mathrm{H}}\mathbf{R}\mathbf{h}_k \text{ s.t. } \mathbf{h}_k^{\mathrm{H}}\mathbf{z}(\omega_k l) = 1 \tag{39}$$

for $l = 1, \ldots, L_k$.

The single filter in (39) is designed subject to L_k constraints, whereas the filterbank design problem in (36) is stated using number of constraints for each filter. In solving for the optimal filter, we proceed as before by using the Lagrange multiplier method, whereby we get the optimal filter expressed in terms of the covariance matrix and the (unknown) Vandermonde matrix $\mathbf{Z}_k$, i.e.,

$$\mathbf{h}_k = \mathbf{R}^{-1}\mathbf{Z}_k(\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k)^{-1}\mathbf{1}, \tag{40}$$

where $\mathbf{1} = [1 \ldots 1]^{\mathrm{T}}$. The output power of this filter can then be expressed as

$$\mathbf{h}_k^{\mathrm{H}}\mathbf{R}\mathbf{h}_k = \mathbf{1}^{\mathrm{H}}(\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k)^{-1}\mathbf{1}, \tag{41}$$

By maximizing the output power, we can obtain an estimate of the fundamental frequency as

$$\hat{\omega}_k = \arg\max_{\omega_k} \mathbf{1}^{\mathrm{H}}(\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k)^{-1}\mathbf{1}. \tag{42}$$

Properties

We will now relate the two filter design methods and the associated estimators in (38) and (42). Comparing

the optimal filters in (37) and (40), two facts can be established. First, the two cost functions are generally different as

$$\mathbf{1}^{\mathrm{H}}(\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k)^{-1}\mathbf{1} \neq \mathrm{Tr}\left[(\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k)^{-1}\right] \qquad (43)$$

$$\mathbf{h}_k^{\mathrm{H}}\mathbf{R}\mathbf{h}_k \neq \mathrm{Tr}[\mathbf{H}_k^{\mathrm{H}}\mathbf{R}\mathbf{H}_k], \qquad (44)$$

with equality only $(\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k)^{-1}$ when is diagonal. Second, the two methods are clearly related in some way as the single filter can be expressed in terms of the filterbank, i.e., $\mathbf{h}_k = \mathbf{H}_k\mathbf{1}$. To quantify under which circumstances $(\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k)^{-1}$ is diagonal and thus when the methods are equivalent, we will analyze the properties of $(\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k)^{-1}$, which figures in both estimators. More specifically, we analyze the asymptotic properties of the expression, i.e.,

$$\lim_{M\to\infty} M(\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k)^{-1}, \qquad (45)$$

where M has been introduced to ensure convergence. We here assume M to be chosen proportional to N, so asymptotic analysis based on M going towards infinity simply means that we let the number of observations tend to infinity. For simplicity, we will in the following derivations assume that the power spectral density of x (n) is finite and non-zero. Although this is strictly speaking not the case for our signal model, the analysis will nonetheless provide some insights into the properties of the filtering methods. The limit in (45) can be rewritten as (see [25] for more details on this subtlety)

$$\lim_{M\to\infty} M(\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k)^{-1} = \left(\lim_{M\to\infty} \frac{1}{M}\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k\right)^{-1}, \quad (46)$$

which leads to the problem of determining the inner limit. To do this, we make use of the asymptotic equivalence of Toeplitz and circulant matrices. For a given Toeplitz matrix, here $\mathbf{R}$, we can construct an asymptotically equivalent circulant $M \times M$ matrix $\mathbf{C}$ in the sense that [36]

$$\lim_{M\to\infty} \frac{1}{\sqrt{M}}\|\mathbf{C} - \mathbf{R}\|_{\mathrm{F}} = 0, \qquad (47)$$

where $||\cdot||_{\mathrm{F}}$ is the Frobenius norm and the limit is taken over the dimensions of $\mathbf{C}$ and $\mathbf{R}$. The conditions under which this was derived in [36] apply to the noise covariance matrix when the stochastic components are generated by a moving average or a stable auto-regressive process. More specifically, the auto-correlation sequence has to be absolutely summable. The result also applies to the deterministic signal components as $\mathbf{Z}_k\mathbf{P}_k\mathbf{Z}_k$ is asymptotically the EVD of the covariance

matrix of $\mathbf{Z}_k\mathbf{a}_k$ (except for a scaling) and circulant. A circulant matrix $\mathbf{C}$ has the eigenvalue decomposition $\mathbf{C} = \mathbf{U}\mathbf{\Gamma}\mathbf{U}^{\mathrm{H}}$ where $\mathbf{U}$ is the Fourier matrix. Thus, the complex sinusoids in $\mathbf{Z}_k$ are asymptotically eigenvectors of $\mathbf{R}$. Therefore, the limit is (see [36,37])

$$\lim_{M\to\infty} \frac{1}{M}\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}\mathbf{Z}_k = \begin{bmatrix} \Phi_x(\omega_k) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Phi_x(\omega_k L_k) \end{bmatrix}, \qquad (48)$$

with $\Phi_x(\omega)$ being the power spectral density of $x(n)$. Similarly, an expression for the inverse of $\mathbf{R}$ can be obtained as $\mathbf{C}^{-1} = \mathbf{U}\mathbf{\Gamma}^{-1}\mathbf{U}^{\mathrm{H}}$ (again, see [36] for details). We now arrive at the following (see also [37] and [38]):

$$\lim_{M\to\infty} \frac{1}{M}\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k = \begin{bmatrix} \Phi_x^{-1}(\omega_k) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Phi_x^{-1}(\omega_k L_k) \end{bmatrix}. \qquad (49)$$

This shows that the expression in (42) asymptotically tends to the following:

$$\lim_{M\to\infty} M\mathbf{1}^{\mathrm{H}}(\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k)^{-1}\mathbf{1} = \sum_{l=1}^{L_k} \Phi_x(\omega_k l), \qquad (50)$$

and similarly for the filterbank formulation:

$$\lim_{M\to\infty} M\,\mathrm{Tr}\left[(\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k)^{-1}\right] = \sum_{l=1}^{L_k} \Phi_x(\omega_k l). \qquad (51)$$

We conclude that the methods are asymptotically equivalent, but may be different for finite M and N. In [25], the two approaches were also reported to have similar performance although the output power estimates deviate. An interesting consequence of the analysis in this section is that the methods based on optimal filtering yield results that are asymptotically equivalent to those obtained using the NLS method.

The two methods based on optimal filtering involve the inverse covariance matrix and we will now analyze the properties of the estimators further by first finding a closed-form expression for the inverse of the covariance matrix based on the covariance matrix model. For the single-pitch case, the covariance matrix model is

$$\mathbf{R}_k = \mathrm{E}\{\mathbf{x}_k(n)\mathbf{x}_k^{\mathrm{H}}(n)\} \qquad (52)$$

$$= \mathbf{Z}_k\mathbf{P}_k\mathbf{Z}_k^{\mathrm{H}} + \mathbf{Q}_k, \qquad (53)$$

and, for simplicity, we will use this model in the following. A variation of the matrix inversion lemma provides us with a useful closed-form expression of the inverse covariance matrix model, i.e.,

$$\mathbf{R}_k^{-1} = (\mathbf{Z}_k \mathbf{P}_k \mathbf{Z}_k^H + \mathbf{Q}_k)^{-1}$$
$$= \mathbf{Q}_k^{-1} - \mathbf{Q}_k^{-1} \mathbf{Z}_k (\mathbf{P}_k^{-1} + \mathbf{Z}_k^H \mathbf{Q}_k^{-1} \mathbf{Z}_k)^{-1} \mathbf{Z}_k^H \mathbf{Q}_k^{-1}. \tag{54}$$

Note that $\mathbf{P}_k^{-1}$ exists for a set of sinusoids having distinct frequencies and non-zero amplitudes and so does the inverse noise covariance matrix $\mathbf{Q}_k^{-1}$ as long as the noise has non-zero variance.

Proceeding in our analysis, we evaluate the expression for a candidate fundamental frequency resulting in a Vandermonde matrix that we denote $\widehat{\mathbf{Z}}_k$. Based on this definition, we get the following expression:

$$\widehat{\mathbf{Z}}_k^H \mathbf{R}_k^{-1} \widehat{\mathbf{Z}}_k = \widehat{\mathbf{Z}}_k^H \mathbf{Q}_k^{-1} \widehat{\mathbf{Z}}_k \tag{55}$$

$$-\widehat{\mathbf{Z}}_k^H \mathbf{Q}_k^{-1} \mathbf{Z}_k (\mathbf{P}_k^{-1} + \mathbf{Z}_k^H \mathbf{Q}_k^{-1} \mathbf{Z}_k)^{-1} \mathbf{Z}_k^H \mathbf{Q}_k^{-1} \widehat{\mathbf{Z}}_k. \tag{56}$$

As before, we normalize this matrix to analyze its behavior as M grows, i.e.,

$$\lim_{M\to\infty} \frac{\widehat{\mathbf{Z}}_k^H \mathbf{R}_k^{-1} \widehat{\mathbf{Z}}_k}{M} = \lim_{M\to\infty} \frac{\widehat{\mathbf{Z}}_k^H \mathbf{Q}_k^{-1} \widehat{\mathbf{Z}}_k}{M} - \lim_{M\to\infty} \frac{\widehat{\mathbf{Z}}_k^H \mathbf{Q}_k^{-1} \mathbf{Z}_k}{M}$$
$$\times \left(\lim_{M\to\infty} \frac{\mathbf{P}_k^{-1}}{M} + \lim_{M\to\infty} \frac{\mathbf{Z}_k^H \mathbf{Q}_k^{-1} \mathbf{Z}_k}{M} \right)^{-1} \lim_{M\to\infty} \frac{\mathbf{Z}_k^H \mathbf{Q}_k^{-1} \widehat{\mathbf{Z}}_k}{M}.$$

Noting that $\lim_{M\to\infty} \frac{1}{M} \mathbf{P}_k^{-1} = 0$, we obtain

$$\lim_{M\to\infty} \frac{\widehat{\mathbf{Z}}_k^H \mathbf{R}_k^{-1} \widehat{\mathbf{Z}}_k}{M} = \lim_{M\to\infty} \frac{\widehat{\mathbf{Z}}_k^H \mathbf{Q}_k^{-1} \widehat{\mathbf{Z}}_k}{M} - \lim_{M\to\infty} \frac{\widehat{\mathbf{Z}}_k^H \mathbf{Q}_k^{-1} \mathbf{Z}_k}{M}$$
$$\times \left(\lim_{M\to\infty} \frac{\mathbf{Z}_k^H \mathbf{Q}_k^{-1} \mathbf{Z}_k}{M} \right)^{-1} \lim_{M\to\infty} \frac{\mathbf{Z}_k^H \mathbf{Q}_k^{-1} \widehat{\mathbf{Z}}_k}{M}. \tag{57}$$

Furthermore, by substituting $\widehat{\mathbf{Z}}_k$ by $\mathbf{Z}_k$, i.e., by evaluating the expression for the true fundamental frequency, we get

$$\lim_{M\to\infty} \frac{1}{M} \mathbf{Z}_k^H \mathbf{R}_k^{-1} \mathbf{Z}_k = \mathbf{0}. \tag{58}$$

This shows that the expression tends to the zero matrix as M approaches infinity for the true fundamental frequency. The cost functions of the two optimal filtering approaches in (50) and (51) therefore can be thought of as tending towards infinity.

Because the autocorrelation sequence of the noise processes $e_k(n)$ can safely be assumed to be absolutely summable and have a smooth and non-zero power spectral density $\Phi_{e_k}(\omega)$ the results of [36,38] can be applied directly to determine following limit:

$$\lim_{M\to\infty} \frac{1}{M} \widehat{\mathbf{Z}}_k^H \mathbf{Q}_k^{-1} \widehat{\mathbf{Z}}_k = \begin{bmatrix} \Phi_{e_k}^{-1}(\hat\omega_k) & & 0 \\ & \ddots & \\ 0 & & \Phi_{e_k}^{-1}(\hat\omega_k L_k) \end{bmatrix}. \tag{59}$$

For the white noise case, the noise covariance matrix is diagonal, i.e., $\mathbf{Q}_k = \sigma_k^2 \mathbf{I}$. The inverse of the covariance matrix model is then

$$\mathbf{R}_k^{-1} = \mathbf{Q}_k^{-1} - \mathbf{Q}_k^{-1} \mathbf{Z}_k (\mathbf{P}_k^{-1} + \mathbf{Z}_k^H \mathbf{Q}_k^{-1} \mathbf{Z}_k)^{-1} \mathbf{Z}_k^H \mathbf{Q}_k^{-1}$$
$$= \frac{1}{\sigma_k^2} \left(\mathbf{I} - \mathbf{Z}_k (\sigma_k^2 \mathbf{P}_k^{-1} + \mathbf{Z}_k^H \mathbf{Z}_k)^{-1} \mathbf{Z}_k^H \right). \tag{60}$$

Next, we note that asymptotically, the complex sinusoids in the columns of $\mathbf{Z}_k$ are orthogonal, i.e.,

$$\lim_{M\to\infty} \frac{1}{M} \mathbf{Z}_k^H \mathbf{Z}_k = \mathbf{I}. \tag{61}$$

Therefore, for large M (and thus N), the inverse covariance matrix can be approximated as

$$\mathbf{R}_k^{-1} \approx \frac{1}{\sigma_k^2} \left(\mathbf{I} - \mathbf{Z}_k (\sigma_k^2 \mathbf{P}_k^{-1} + M\mathbf{I}_k)^{-1} \mathbf{Z}_k^H \right). \tag{62}$$

It can be observed that the remaining inverse matrix involves two diagonal matrices that can be rewritten as

$$\sigma_k^2 \mathbf{P}_k^{-1} + M\mathbf{I} \triangleq \Gamma_k^{-1} \tag{63}$$

$$= \text{diag} \left(\begin{bmatrix} \frac{\sigma_k^2}{A_{k,1}^2} + M & \cdots & \frac{\sigma_k^2}{A_{k,L_k}^2} + M \end{bmatrix} \right), \tag{64}$$

which leads to the inverse

$$\Gamma_k = \text{diag} \left(\begin{bmatrix} \frac{A_{k,1}^2}{\sigma_k^2 + MA_{k,1}^2} & \cdots & \frac{A_{k,L_k}^2}{\sigma_k^2 + MA_{k,L_k}^2} \end{bmatrix} \right). \tag{65}$$

Finally, we arrive at the following expression, which is an asymptotic approximation of the inverse of the matrix covariance model:

$$\mathbf{R}_k^{-1} \approx \frac{1}{\sigma_k^2} (\mathbf{I} - \mathbf{Z}_k \Gamma_k \mathbf{Z}_k^H). \tag{66}$$

Interestingly, it can be seen that the inverse covariance matrix asymptotically exhibits a similar structure to that of the covariance matrix model.

Order estimation
We will now consider the problem of finding the model order L_k. This problem is a special case of the general model selection problem, where the models under consideration are nested as the simple models are special cases of the more complicated models. Many methods

for dealing with this problem have been investigated over the years, but the most common ones for order selection are the Akaike information criterion (AIC) [39] and the minimum description length criterion (MDL) [40] (see also [41]). Herein, we derive a model order selection criterion using the asymptotic MAP approach of [27,42] (see also [43]), a method that penalizes linear and nonlinear parameters differently. We will do this for the single-pitch case, but the principles can be used for multi-pitch signals too. First, we introduce a candidate model index set

$$\mathbb{Z}_q = \{0, 1, \ldots, q - 1\} \tag{67}$$

and the candidate models $\mathcal{M}_m$ with $m \in \mathbb{Z}_q$. We will here consider the problem of estimating the number of harmonics for a single source from a single-pitch signal $\mathbf{x}_k$. In the following, $f(\cdot)$ denotes probability density function (PDF) of the argument (with the usual abuse of notation).

The principle of MAP-based model selection can be explained as follows: Choose the model that maximizes the a posteriori probability $f(\mathcal{M}_m|\mathbf{x}_k)$ of the model given the observation $\mathbf{x}_k$. This can be stated mathematically as

$$\widehat{\mathcal{M}}_k = \arg \max_{\mathcal{M}_m, m \in \mathbb{Z}_q} f(\mathcal{M}_m|\mathbf{x}_k) \tag{68}$$

$$= \arg \max_{\mathcal{M}_m, m \in \mathbb{Z}_q} \frac{f(\mathbf{x}_k|\mathcal{M}_m)f(\mathcal{M}_m)}{f(\mathbf{x}_k)}, \tag{69}$$

Noting that the probability of $\mathbf{x}_k$, i.e., $f(\mathbf{x}_k)$, is constant once $\mathbf{x}_k$ has been observed and assuming that all the models are equally probable $f(\mathcal{M}_m) = \frac{1}{q}$, the MAP model selection criterion reduces to

$$\widehat{\mathcal{M}}_k = \arg \max_{\mathcal{M}_m, m \in \mathbb{Z}_q} f(\mathbf{x}_k|\mathcal{M}_m), \tag{70}$$

which is the likelihood function when seen as a function of $\mathcal{M}_k$. The various candidate models depend on a number of unknown parameters, in our case amplitudes, phases and fundamental frequency, that we here denote θ_k. To eliminate this dependency, we seek to integrate those parameters out, i.e.,

$$f(\mathbf{x}_k|\mathcal{M}_m) = \int f(\mathbf{x}_k|\theta_k, \mathcal{M}_m)f(\theta_k|\mathcal{M}_m)d\theta_k. \tag{71}$$

However, simple analytic expression for this integral does not generally exist, especially so for complicated nonlinear models such as the one used here. We must therefore seek another, possibly approximate, way of evaluating this integral. One such way is numerical

integration, but we will here instead follow the Laplace integration method as proposed in [27,42].

The first step is as follows. Define $g(\theta_k)$ as the integrand in (71), i.e., $g(\theta_k) = f(\mathbf{x}_k|\theta_k, \mathcal{M}_m)f(\theta_k|\mathcal{M}_m)$. Next, let $\widehat{\theta}_k$ be the mode of $g(\theta_k)$, i.e., the MAP estimate. Using a Taylor expansion of $g(\theta_k)$ in $\widehat{\theta}_k$, the integrand in (71) can be approximated as

$$g(\theta_k) \approx g(\widehat{\theta}_k)e^{-\frac{1}{2}(\theta_k - \widehat{\theta}_k)^{\mathrm{T}}\widehat{\mathbf{G}}_k(\theta_k - \widehat{\theta}_k)}, \tag{72}$$

where $\widehat{\mathbf{G}}_k$ the Hessian of the logarithm of $g(\theta_k)$ evaluated in $\widehat{\theta}_k$, i.e.,

$$\widehat{\mathbf{G}}_k = -\frac{\partial^2 \ln g(\theta_k)}{\partial \theta_k \partial \theta_k^{\mathrm{T}}}\bigg|_{\theta_k = \widehat{\theta}_k}. \tag{73}$$

Note that the Taylor expansion of the function in (72) is of a real function in real parameters, even if the likelihood function is for complex quantities. The above results in the following simplified expression for (71):

$$f(\mathbf{x}_k|\mathcal{M}_m) \approx g(\widehat{\theta}_k)\int e^{-\frac{1}{2}(\theta_k - \widehat{\theta}_k)^{\mathrm{T}}\widehat{\mathbf{G}}_k(\theta_k - \widehat{\theta}_k)_{d\theta_k}}. \tag{74}$$

The integral involved in this expression involves a quadratic expression that is much simpler than the highly nonlinear one in (71). It can be shown to be

$$\int e^{-\frac{1}{2}(\theta_k - \widehat{\theta}_k)^{\mathrm{T}}\widehat{\mathbf{G}}_k(\theta_k - \widehat{\theta}_k)}d\theta_k = (2\pi)^{D_k/2}|\widehat{\mathbf{G}}_k|^{-\frac{1}{2}}, \tag{75}$$

where $|\cdot|$ is the matrix determinant and D_k the number of parameters in θ_k. The expression in (71) can now be written as [27,42] (see also [43])

$$f(\mathbf{x}_k|\mathcal{M}_m) \approx (2\pi)^{D_k/2}|\widehat{\mathbf{G}}_k|^{-\frac{1}{2}}g(\widehat{\theta}_k). \tag{76}$$

Next, assuming a vague prior on the parameters given the model, i.e., on $f(\theta_k|\mathcal{M}_m)$, $g(\theta_k)$ reduces to a likelihood function and $\widehat{\theta}_k$ to the maximum likelihood estimate. Note that this will also be the case for large N, as the MAP estimate will then converge to the maximum likelihood estimate. In that case, the Hessian matrix reduces to

$$\widehat{\mathbf{G}}_k = -\frac{\partial^2 \ln f(\mathbf{x}_k|\theta_k, \mathcal{M}_m)}{\partial \theta_k \partial \theta_k^{\mathrm{T}}}\bigg|_{\theta_k = \widehat{\theta}_k}, \tag{77}$$

which is sometimes referred to as the observed information matrix. This matrix is related to the Fisher information matrix in the following way: it is evaluated in $\widehat{\theta}_k$ instead of the true parameters and no expectation is

taken. However, it was shown in [43] that (77) can be used as an approximation (for large N) of the Fisher information matrix, and, hence, also vice versa, leading to the following approximation:

$$\widehat{\mathbf{G}}_k \approx -\mathrm{E}\left\{ \frac{\partial^2 \ln f(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathcal{M}_m)}{\partial \boldsymbol{\theta}_k \partial \boldsymbol{\theta}_k^T} \right\}\Bigg|_{\boldsymbol{\theta}_k = \widehat{\boldsymbol{\theta}}_k} \qquad (78)$$

The benefit of using (78) over (77) is that the former is readily available in the literature for many models, something that also is the case for our model [9].

Taking the logarithm of the right-hand side of (76) and sticking to the tradition of ignoring terms of order $\mathcal{O}(1)$ and $\frac{D_k}{2} \ln 2\pi$ (which are negligible for large N), we get that under the aforementioned conditions, (70) can be written as

$$\widehat{\mathcal{M}}_k = \arg \min_{\mathcal{M}_m, m \in \mathbb{Z}_q} - \ln f(\mathbf{x}_k | \widehat{\boldsymbol{\theta}}_k, \mathcal{M}_m) + \frac{1}{2} \ln |\widehat{\mathbf{G}}_k|, \quad (79)$$

which can be used for determining which models is the most likely explanation for the observed signal. Now we will derive a criterion for selecting the model order of the single-pitch model and detecting the presence of a periodic source.

Based on the Fisher information matrix as derived in [9], we introduce the normalization matrix (see [43])

$$\mathbf{K}_N = \begin{bmatrix} N^{-3/2} & \mathbf{0} \\ \mathbf{O} & N^{-1/2}\mathbf{I} \end{bmatrix}, \qquad (80)$$

where $\mathbf{I}$ is an $2L_k \times 2L_k$ identity matrix. The diagonal terms are due to the fundamental frequency and the L_k amplitudes and phases, respectively. The determinant of the Hessian in (79) can be written as

$$|\widehat{\mathbf{G}}_k| = |\mathbf{K}_N^{-2}| \, |\mathbf{K}_N \widehat{\mathbf{G}}_k \mathbf{K}_N|. \qquad (81)$$

By observing that $\mathbf{K}_N \widehat{\mathbf{G}}_k \mathbf{K}_N = \mathcal{O}(1)$ and taking the logarithm, we obtain the following expression:

$$\ln |\widehat{\mathbf{G}}_k| = \ln |\mathbf{K}_N^{-2}| + \ln |\mathbf{K}_N \widehat{\mathbf{G}}_k \mathbf{K}_N| \qquad (82)$$

$$= \ln |\mathbf{K}_N^{-2}| + \mathcal{O}(1) \qquad (83)$$

$$= 3 \ln N + 2L_k \ln N + \mathcal{O}(1). \qquad (84)$$

Assuming that the observation noise is white and Gaussian distributed, the log-likelihood function in (79) depends only on the term $N \ln \sigma_k^2$ where σ_k^2 is replaced by an estimate for each candidate order L_k. We denote this estimate as $\hat{\sigma}_k^2 (L_k)$. Finally, substituting (84) into (79), the following simple and useful expression for selecting the model order is obtained:

$$\hat{L}_k = \arg \min_{L_k} N \ln \hat{\sigma}_k^2(L_k) + \frac{3}{2} \ln N + L_k \ln N \qquad (85)$$

Note that for low N, the inclusion of the term $\frac{D_k}{2} \ln 2\pi = (L_k + \frac{1}{2}) \ln 2\pi$ may lead to more accurate results. To determine whether any harmonics are present at all, i.e., performing pitch detection, the above cost function should be compared to the log-likelihood of the zero order model, meaning that no harmonics are present if

$$N \ln \hat{\sigma}_k^2(0) < N \ln \hat{\sigma}_k^2(\hat{L}_k) + \frac{3}{2} \ln N + \hat{L}_k \ln N, \qquad (86)$$

where, in this case, $\hat{\sigma}_k^2(0)$ is simply the variance of the observed signal. The rule in (86) is essentially a *pitch detection* rule as it detects the presence of a pitch. It can be seen that both (85) and (86) require the determination of the noise variance for each candidate model order. The criterion in (85) reflects the tradeoff between the variance of the residual and the complexity of the model. For example, for a high model order, the estimated variance will be low, but the number of parameters will be high. Conversely, for a low model order, there are only few parameters but a high variance residual.

Variance estimation

As we have seen, the order selection criterion requires that the noise variance is estimated, and we will now show how to use these filters for estimating the variance of the signal by filtering out the harmonics. We will do this based on the filterbank design. First, we define an estimate of the noise obtained from $x(n)$ as

$$\hat{e}(n) = x(n) - y_k(n), \qquad (87)$$

which we will refer to as the residual. Moreover, $y_k(n)$ is the sum of the input signal filtered by the filterbank, i.e.,

$$y_k(n) = \sum_{m=0}^{M-1} \sum_{l=1}^{L_k} h_{k,l}(m) x(n - m) \qquad (88)$$

$$= \sum_{m=0}^{M-1} h_k(m) x(n - m), \qquad (89)$$

where $h_k(m)$ is the sum over the impulse responses of the filters of the filterbank. From the relation between the single filter design and the filterbank design, it is now clear that when used this way, the two approaches lead to the same output signal $y_k(n)$. This also offers some insights into the difference between the designs in (36) and (39). More specifically, the difference is in the

way the output power is measured, where (36) is based on the assumption that the power is additive over the filters, i.e., that the output signals are uncorrelated. We can now write the noise estimate as

$$\hat{e}(n) = x(n) - \sum_{m=0}^{M-1} h_k(m)x(n-m) \tag{90}$$

$$\triangleq \mathbf{g}_k^{\mathrm{H}} \mathbf{x}(n) \tag{91}$$

where $\mathbf{g}_k = [(1 - h_k(0)) - h_k(1) \ldots - h_k(M-1)]^{\mathrm{H}}$ is the modified filter. From the noise estimate, we can then estimate the noise variance for the L_kth order model as

$$\hat{\sigma}^2(L_k) = \mathrm{E}\{|\hat{e}(n)|^2\} = \mathrm{E}\{\mathbf{g}_k^{\mathrm{H}}\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n)\mathbf{g}_k\} \tag{92}$$

$$= \mathbf{g}_k^{\mathrm{H}} \mathbf{R} \mathbf{g}_k. \tag{93}$$

This expression is, however, not very convenient for a number of reasons: A notable property of the estimator in (42) is that it does not require the calculation of the filter and that the output power expression in (41) is simpler than the expression for the optimal filter in (40). To use (93) directly, we would first have to calculate the optimal filter using (40), then calculate the modified filter $\mathbf{g}_k$, before evaluating (93). Instead, we simplify the evaluation of (93) by defining the modified filter as $\mathbf{g}_k = \mathbf{b}_1 - \mathbf{h}_k$ where, as defined earlier,

$$\mathbf{b}_1 = [1 \ 0 \cdots 0]^{\mathrm{H}}. \tag{94}$$

Next, we use this definition to rewrite the variance estimate as

$$\hat{\sigma}^2(L_k) = \mathbf{g}_k^{\mathrm{H}}\mathbf{R}\mathbf{g}_k = (\mathbf{b}_1 - \mathbf{h}_k)^{\mathrm{H}}\mathbf{R}(\mathbf{b}_1 - \mathbf{h}_k) \tag{95}$$

$$= \mathbf{b}_1^{\mathrm{H}}\mathbf{R}\mathbf{b}_1 - \mathbf{b}_1^{\mathrm{H}}\mathbf{R}\mathbf{h}_k - \mathbf{h}_k^{\mathrm{H}}\mathbf{R}\mathbf{b}_1 + \mathbf{h}_k^{\mathrm{H}}\mathbf{R}\mathbf{h}_k. \tag{96}$$

The first term can be identified to equal the variance of the observed signal $x(n)$, i.e., $\mathbf{b}_1^{\mathrm{H}}\mathbf{R}\mathbf{b}_1 = \mathrm{E}\{|x(n)|^2\}$, and $\mathbf{h}_k^{\mathrm{H}}\mathbf{R}\mathbf{h}_k$ we know from (41). Writing out the cross-terms $\mathbf{b}_1^{\mathrm{H}}\mathbf{R}\mathbf{h}_k$ using (40) yields

$$\mathbf{b}_1^{\mathrm{H}}\mathbf{R}\mathbf{h}_k = \mathbf{b}_1^{\mathrm{H}}\mathbf{R}\mathbf{R}^{-1}\mathbf{Z}_k(\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k)^{-1}\mathbf{1} \tag{97}$$

$$= \mathbf{b}_1^{\mathrm{H}}\mathbf{Z}_k(\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k)^{-1}\mathbf{1}. \tag{98}$$

Furthermore, it can easily be verified that $\mathbf{b}_1^{\mathrm{H}}\mathbf{Z}_k = \mathbf{1}^{\mathrm{H}}$, from which it can be concluded that

$$\mathbf{b}_1^{\mathrm{H}}\mathbf{R}\mathbf{h}_k = \mathbf{1}^{\mathrm{H}}(\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k)^{-1}\mathbf{1} \tag{99}$$

$$= \mathbf{h}_k^{\mathrm{H}}\mathbf{R}\mathbf{h}_k. \tag{100}$$

Therefore, the variance estimate can be expressed as

$$\hat{\sigma}^2(L_k) = \hat{\sigma}^2(0) - \mathbf{1}^{\mathrm{H}}(\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k)^{-1}\mathbf{1}, \tag{101}$$

where $\hat{\sigma}^2(0) = \mathrm{E}\{|x(n)|^2\}$ is simply the variance of the observed signal. The variance estimate in (101) involves the same expression as in the fundamental frequency estimation criterion in (42), which means that the same expression can be used for estimating the model order and the fundamental frequency, i.e., the approach allows for joint estimation of the model order and the fundamental frequency. The variance estimate in (101) also shows that the same filter that maximizes the output power minimizes the variance of the residual. A more conventional variance estimate could be formed by first finding the frequency using, e.g., (42) and then finding the amplitudes of the signal model using (weighted) least-squares [38] to obtain a noise variance estimate. Since the discussed procedure uses the same information in finding the fundamental frequency and the noise variance, it is superior to the least-squares approach in terms of computational complexity. Note that for finite filter lengths, the output of the filters considered here are generally "power levels" and not power spectral densities (see [44]), which is consistent with our use of the filters for estimating the variance. Asymptotically, the filters do comprise power spectral density estimates [25].

By inserting (101) in (85), the model order can be determined using the MAP criterion for a given fundamental frequency. By combining the variance estimate in (101) with (85), we obtain the following fundamental frequency estimator for the case of unknown model orders (for $L_k > 0$):

$$\hat{\omega}_k = \arg\min_{\omega_k} \min_{L_k} N \ \ln \hat{\sigma}^2(L_k) + \frac{3}{2}\ln N + L_k \ln N, \tag{102}$$

where the model order is also estimated in the process. To determine whether any harmonics are present at all, the criterion in (86) can be used.

Order-recursive implementation
Both the filterbank method and the single filter method require the calculation of the following matrix for every combination of candidate fundamental frequencies and orders:

$$\mathbf{\Xi}_{L_k} \triangleq (\mathbf{Z}_k^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{Z}_k)^{-1}, \tag{103}$$

where $\mathbf{\Xi}_{L_k}$ denotes the inverse matrix for an order L_k model. The respective cost function are formed from (103) as either the trace or the sum of all elements of this matrix. Since this requires a matrix inversion of cubic complexity for each pair, there is a considerable

computational burden of using these methods. We will now present an efficient implementation of the matrix inversion in (103). The methods also require that the inverse covariance matrix be calculated, but this is less of a concern for two reasons. Firstly, it is calculated only once per segment, and, secondly, many standard methods exist for updating the matrix inverse over time (see, e.g., [45]).

The fast implementation that we will now proceed to derive, is based on the matrix inversion lemma, and is basically a recursion over the model order. To apply the matrix inversion lemma to the calculation of (103), we first define the matrix composed of vectors corresponding to the L_k - 1 first harmonics of the full matrix $\mathbf{Z}_k$ as

$$\mathbf{Z}_k^{(L_k-1)} = [\mathbf{z}(\omega_k) \; \cdots \; \mathbf{z}(\omega_k(L_k-1))], \tag{104}$$

and a vector containing the last harmonic L_k as

$$\mathbf{z}_k^{(L_k)} = \left[e^{-j\omega_k L_k 0} \; \cdots \; e^{-j\omega_k L_k(M-1)} \right]^T. \tag{105}$$

Using these definitions, we can now rewrite (103) as

$$\mathbf{\Xi}_{L_k} = (\mathbf{Z}_k^H \mathbf{R}^{-1} \mathbf{Z}_k)^{-1} \tag{106}$$

$$= \begin{bmatrix} \mathbf{Z}_k^{(L_k-1)H} \mathbf{R}^{-1} \mathbf{Z}_k^{(L_k-1)} & \mathbf{Z}_k^{(L_k-1)H} \mathbf{R}^{-1} \mathbf{z}_k^{(L_k)} \\ \mathbf{z}_k^{(L_k)H} \mathbf{R}^{-1} \mathbf{Z}_k^{(L_k-1)} & \mathbf{z}_k^{(L_k)H} \mathbf{R}^{-1} \mathbf{z}_k^{(L_k)} \end{bmatrix}^{-1}. \tag{107}$$

Next, define the scalar quantity

$$\xi_{L_k} = \mathbf{z}_k^{(L_k)H} \mathbf{R}^{-1} \mathbf{z}_k^{(L_k)}, \tag{108}$$

which is real and positive since $\mathbf{R}^{-1}$ is positive-definite and Hermitian, and the vector

$$\boldsymbol{\eta}_{L_k} = \mathbf{Z}_k^{(L_k-1)H} \mathbf{R}^{-1} \mathbf{z}_k^{(L_k)}. \tag{109}$$

We can now express the matrix in (103) in terms of the order $(L_k$ - 1) matrix $\mathbf{\Xi}_{L_k-1}$, ξ_{L_k}, and $\boldsymbol{\eta}_{L_k}$ as

$$\mathbf{\Xi}_{L_k} = \begin{bmatrix} \mathbf{\Xi}_{L_k-1} & \mathbf{0} \\ \mathbf{O} & 0 \end{bmatrix} + \begin{bmatrix} -\mathbf{\Xi}_{L_k-1}\boldsymbol{\eta}_{L_k} \\ 1 \end{bmatrix}$$
$$\times \frac{1}{\xi_{L_k} - \boldsymbol{\eta}_{L_k}^H \mathbf{\Xi}_{L_k-1} \boldsymbol{\eta}_{L_k}} \left[-\boldsymbol{\eta}_{L_k}^H \mathbf{\Xi}_{L_k-1} \; 1 \right]. \tag{110}$$

This can be rewritten as follows:

$$\mathbf{\Xi}_{L_k} = \begin{bmatrix} \mathbf{\Xi}_{L_k-1} & \mathbf{0} \\ \mathbf{O} & 0 \end{bmatrix}$$
$$+ \begin{bmatrix} \mathbf{\Xi}_{L_k-1}\boldsymbol{\eta}_{L_k}\boldsymbol{\eta}_{L_k}^H \mathbf{\Xi}_{L_k-1} & -\mathbf{\Xi}_{L_k-1}\boldsymbol{\eta}_{L_k} \\ -\boldsymbol{\eta}_{L_k}^H \mathbf{\Xi}_{L_k-1} & 1 \end{bmatrix} \tag{111}$$

$$\times \frac{1}{\xi_{L_k} - \boldsymbol{\eta}_{L_k}^H \mathbf{\Xi}_{L_k-1} \boldsymbol{\eta}_{L_k}}$$
$$\triangleq \begin{bmatrix} \mathbf{\Xi}_{L_k-1} & \mathbf{0} \\ \mathbf{O} & 0 \end{bmatrix} + \frac{1}{\beta_{L_k}} \begin{bmatrix} \boldsymbol{\zeta}_{L_k}\boldsymbol{\zeta}_{L_k}^H & -\boldsymbol{\zeta}_{L_k} \\ -\boldsymbol{\zeta}_{L_k}^H & 1 \end{bmatrix}. \tag{112}$$

This shows that once $\mathbf{\Xi}_{L_k-1}$ is known, $\mathbf{\Xi}_{L_k}$ can be obtained in a simple way. To use this result to calculate the cost functions for the estimators (38) and (42) for a model order L_k, we proceed as follows. For a given ω_k, calculate the order 1 inverse matrix as

$$\mathbf{\Xi}_1 = \frac{1}{\xi_1}, \tag{113}$$

and then for l = 2, ..., L_k calculate the quantities needed to update the inverse matrix, i.e.,

$$\boldsymbol{\kappa}_l = \mathbf{R}^{-1} \mathbf{z}_k^{(l)} \tag{114}$$

$$\xi_l = \mathbf{z}_k^{(l)H} \boldsymbol{\kappa}_l \tag{115}$$

$$\boldsymbol{\eta}_l = \mathbf{Z}_k^{(l-1)H} \boldsymbol{\kappa}_l \tag{116}$$

$$\boldsymbol{\zeta}_l = \mathbf{\Xi}_{l-1} \boldsymbol{\eta}_l \tag{117}$$

$$\beta_l = \xi_l - \boldsymbol{\eta}_l^H \boldsymbol{\zeta}_l \tag{118}$$

$$\mathbf{\Xi}_l = \begin{bmatrix} \mathbf{\Xi}_{l-1} & \mathbf{0} \\ \mathbf{O} & 0 \end{bmatrix} + \frac{1}{\beta_l} \begin{bmatrix} \boldsymbol{\zeta}_l\boldsymbol{\zeta}_l^H & -\boldsymbol{\zeta}_l \\ -\boldsymbol{\zeta}_l^H & 1 \end{bmatrix}, \tag{119}$$

using which the estimators in (38) and (42) along with the variance estimate in (101) can easily be implemented. In assessing the efficiency of the proposed recursion, we will use the following assumptions:

- Matrix-vector product: the computation of $\mathbf{Ax}$ where $\mathbf{A} \in \mathbb{C}^{m \times n}$ and $\mathbf{x} \in \mathbb{C}^n$ requires $\mathcal{O}(mn)$ operations.
- Matrix-matrix product: the computation of $\mathbf{AB}$ where $\mathbf{A} \in \mathbb{C}^{m \times n}$ and $\mathbf{B} \in \mathbb{C}^{n \times p}$ requires $\mathcal{O}(mpn)$ operations.
- Matrix inversion: the computation of $\mathbf{A}^{-1}$ where $\mathbf{A} \in \mathbb{C}^{m \times m}$ requires $\mathcal{O}(m^3)$ operations.

The number of operations required to calculate the matrix inverse in (103) is determined by the filter length M and the number of harmonics L_k. This leads to a complexity of $\mathcal{O}(L_k^3 + M^2 L_k + ML_k^2)$ for the direction implementation of (103). On the other hand, an update (from the $(l$ - 1)th order model to the lth order model)

in the order-recursive implementation requires $\mathcal{O}(M^2 + l^2 + M + Ml)$ operations with $l = 1, ..., L_k$. For the case where only the cost for an order L_k model needs to be calculated, the saving is negligible. On the other hand, if the order is unknown and the cost function has to be calculated for a wide range of orders, e.g., $L_k = 1, 2, ...$, only a single update is required as the information from prior iterations are simply reused. At this point it should also be noted that the filter length and thus the covariance matrix size is generally much larger than the model order, i.e., $M \gg L_k$. In Figure 1, the approximate complexities of the respective implementations are depicted for $M = 50$.

It should be stressed that this order recursive implementation is exact as no approximations are involved, meaning that it implements exactly the matrix inversion in (103). We note in passing that, as usual, all the inner products involving complex sinusoids of different frequencies can be calculated efficiently using FFTs.

Experimental results

First, we will provide an illustrative example of what the derived optimal filters may look like. In Figure 2, an example of such filters are given with the magnitude response of the optimal filterbank and the single filter being shown for white Gaussian noise with $\omega_1 = 1.2566$ and $L_1 = 3$. It should be stressed that for a non-diagonal **R**, i.e., when the observed signal contains sinusoids and/or colored noise, the resulting filters can look radically different.

We will now evaluate the statistical performance of the proposed scheme. In doing so, we will compare to some other methods based on well-established estimation theoretical approaches that are able to jointly
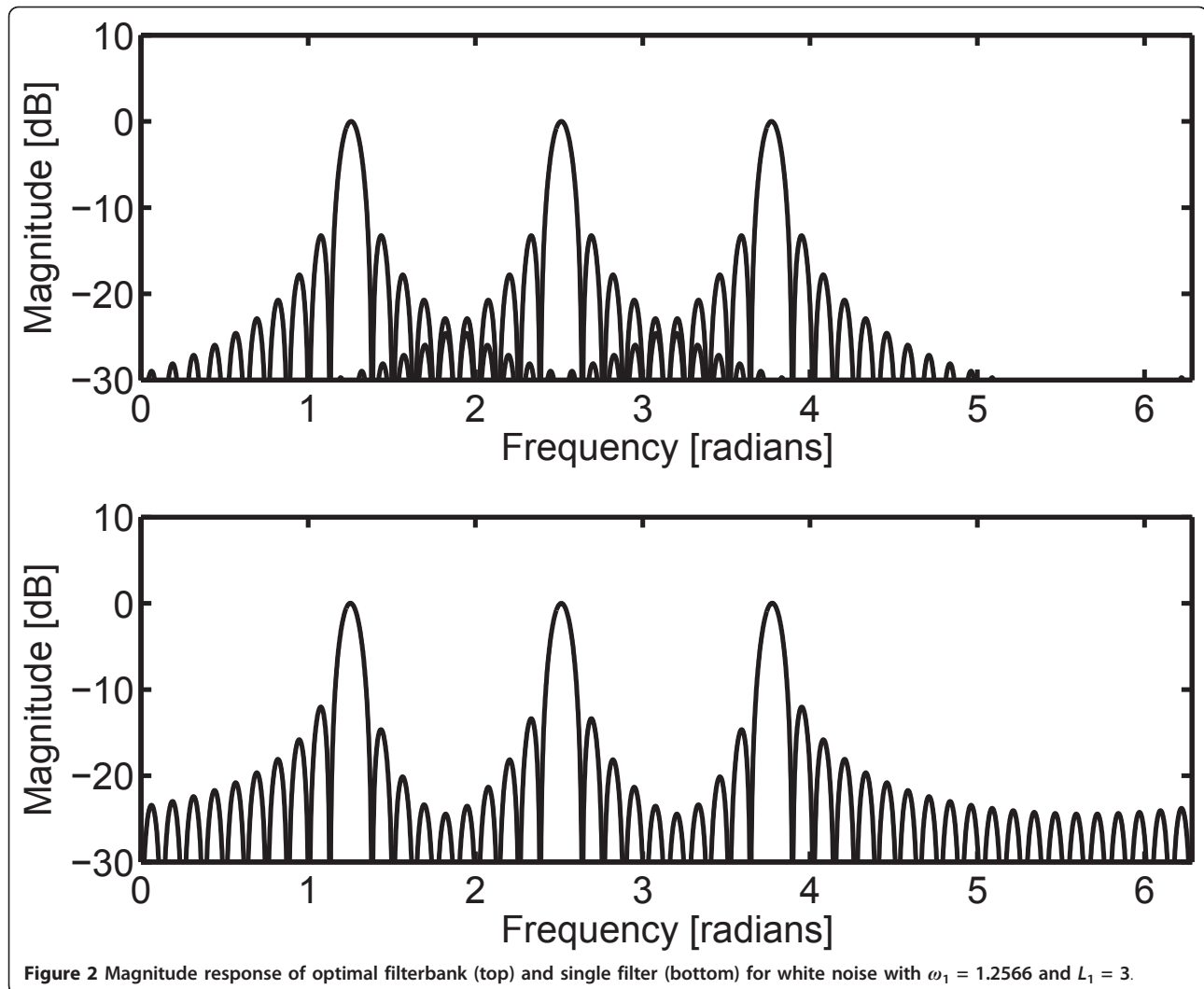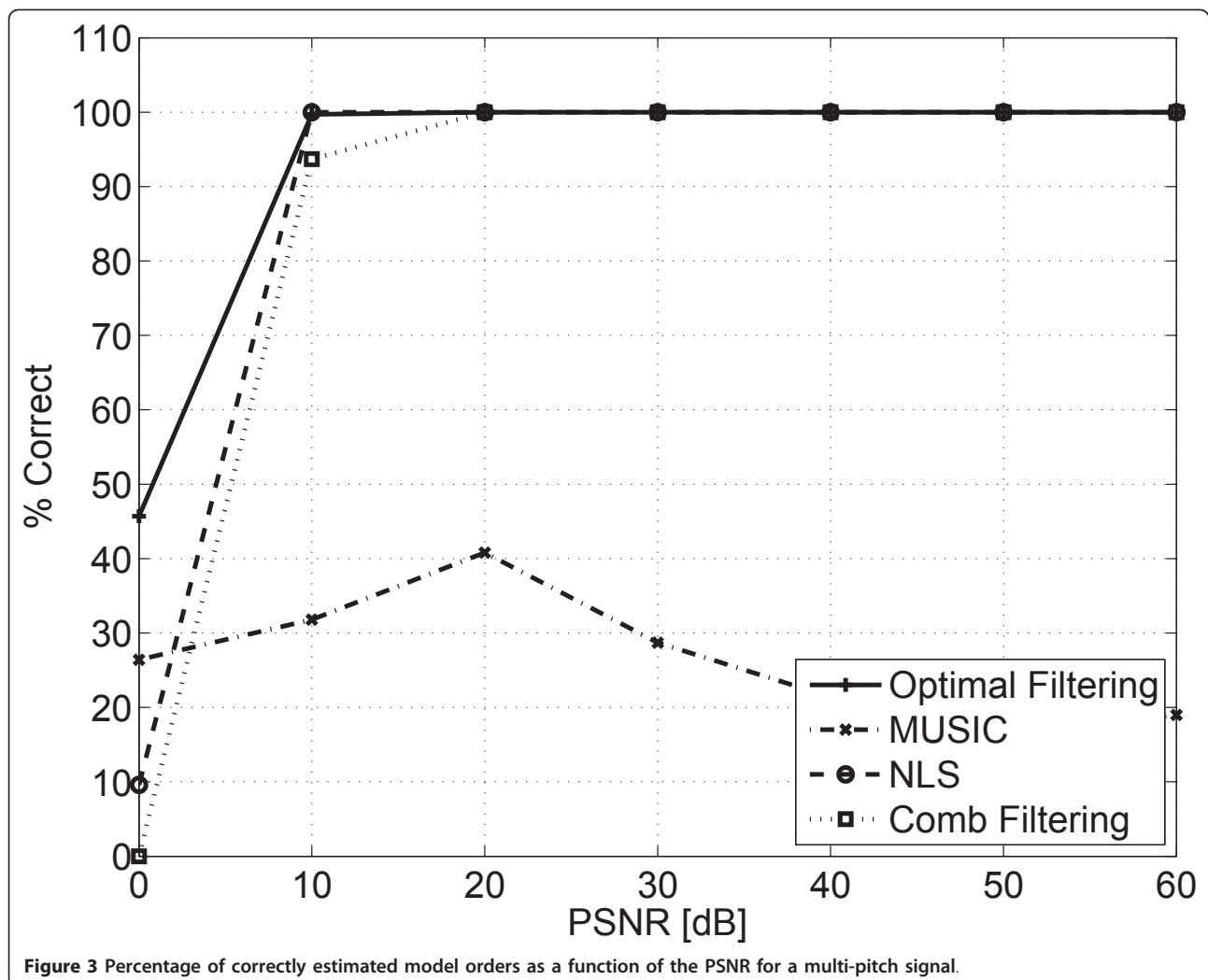

Figure 1 Approximate complexities of the order-recursive implementation and the direct implementation of the matrix inverse in (103).
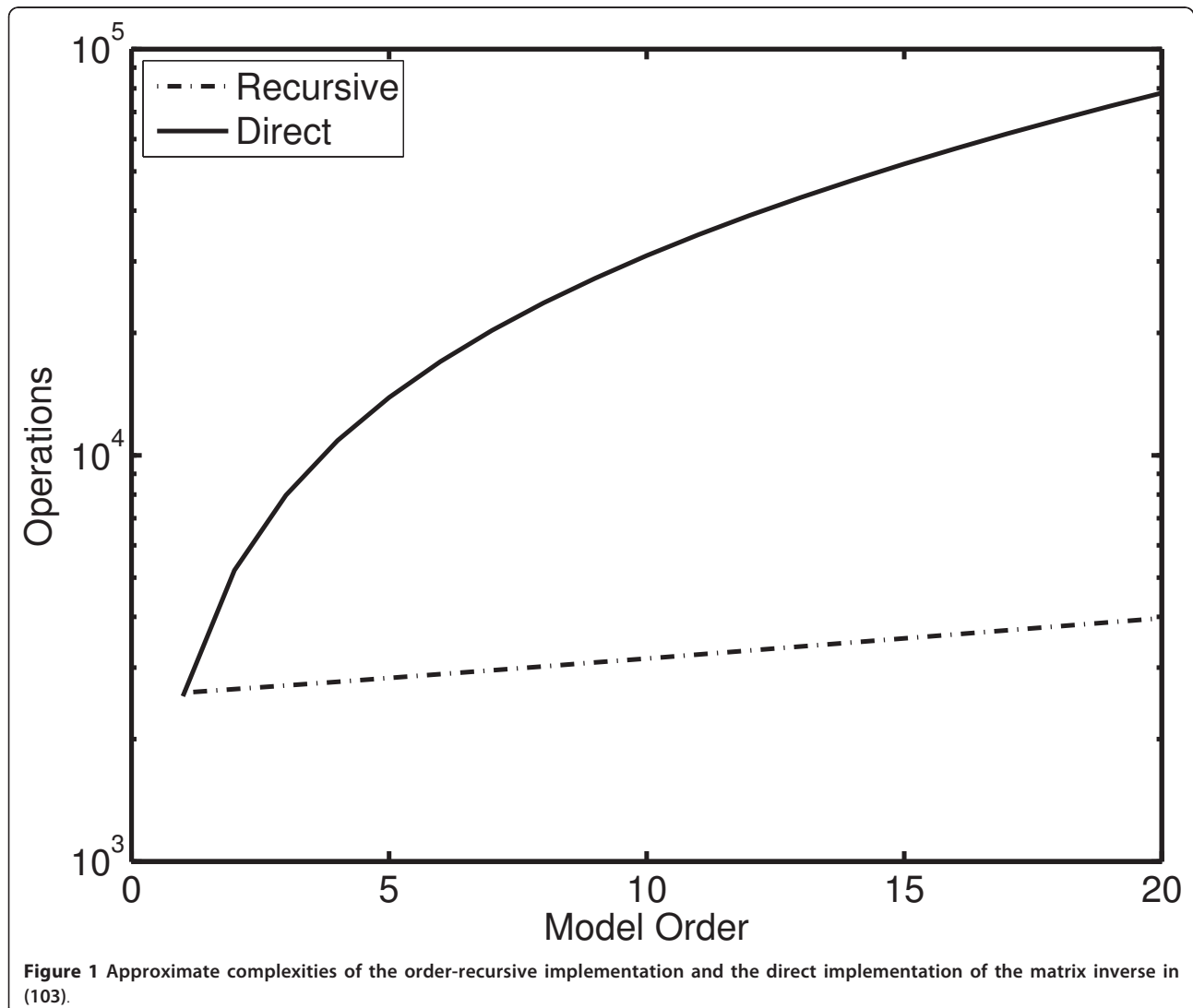
Figure 2 Magnitude response of optimal filterbank (top) and single filter (bottom) for white noise with ω_1 = 1.2566 and L_1 = 3.

estimate the fundamental frequency and the order, namely a subspace method, the MUSIC method of [9], and the NLS method [16]. The NLS method in combination with the criterion (85) yields both a maximum likelihood fundamental frequency estimate and a MAP order estimate (see [27] for details) and it is asymptotically a filtering method as described in Sect. III. More specifically, noise variance estimates are obtained using (20) after which the order estimation criterion is applied. For a finite number of samples, however, the exact NLS used here is superior to the filtering method of Sect. III. We remark that the NLS, MUSIC and optimal filtering methods under consideration here are comparable in terms of computational efficiency as they all have cubic complexity, involving either inverses of matrices, matrix-matrix products or eigenvalue decompositions of matrices. Additionally, we also compare to the performance of the comb filtering method described in Sect. III combined with the criterion (85). We will

here focus on their application to order estimation, investigating the performance of the estimators given the fundamental frequency. The reason for this is simply that the high-resolution estimation capabilities of the proposed method, MUSIC and NLS for the fundamental frequency estimation problem for both single- and multi-pitch signals are already well-documented in [9,16,25], and there is little reason to repeat those experiments here. We note that the NLS method reduces to a linear least-squares method when the fundamental frequency is given, but the joint estimator is still nonlinear. The statistical order estimation criterion in (85) was derived based on $\mathbf{x}_k$, i.e., a single-pitch model, and none of the methods considered in this paper take additional sources into account in an explicit manner. Since one cannot generally assume that only a single periodic source is present, we will test the methods for a multi-pitch signal containing two sources, namely the signal of interest and an interfering periodic

source that is considered to be of no interest to us. However, we will use all the methods as if only one source is present, thereby testing the robustness of the respective methods. In the experiments the following additional conditions were used: a periodic signal was generated using (1) with a fundamental frequency of ω_1 = 0.8170, L_1 = 5 and A_l = 1 $\forall l$ along with an interfering periodic source having the same number of harmonics and amplitude distribution but with ω_2 = 1.2. For each test condition, 1000 Monte Carlo iterations were run. In the first experiment, we will investigate the performance as a function of the pseudo signal-to-noise (PSNR) as defined in [9]. Note that this PSNR is higher than the usual SNR, meaning that the conditions are more noisy than they may appear at first sight. The performance of the estimators has been evaluated for N = 200 observed samples with a covariance matrix size/filter length of M = 50. The results are shown in Figure 3 in terms of the percentage of correctly estimated orders. Similarly, the

performance is investigated as a function of N with M = $N/4$ in the second experiment for PSNR = 40 dB, i.e., the filter length is set proportionally to the number of samples. Note that the NLS method operates on the entire length N signal and thus does not depend on M. This experiment thus reveals not only the dependency of the performance on the number of observed samples but also on the filter length. The results are shown in Figure 4. Similarly, it is interesting to investigate the importance of the number of unknown parameters relative to the number of samples. Consequently, an experiment has been carried out to do exactly this with the results being shown in Figure 5. The signals were generated as before with N = 400 for PSNR = 40 dB with M = $N/4$ while L_k was varied from 1 to 10. In the final experiment, the N is kept fixed while the filter length M is varied with PSNR = 40 dB. In the process, the covariance matrix of MUSIC is varied too. The results can be seen in Figure 6.
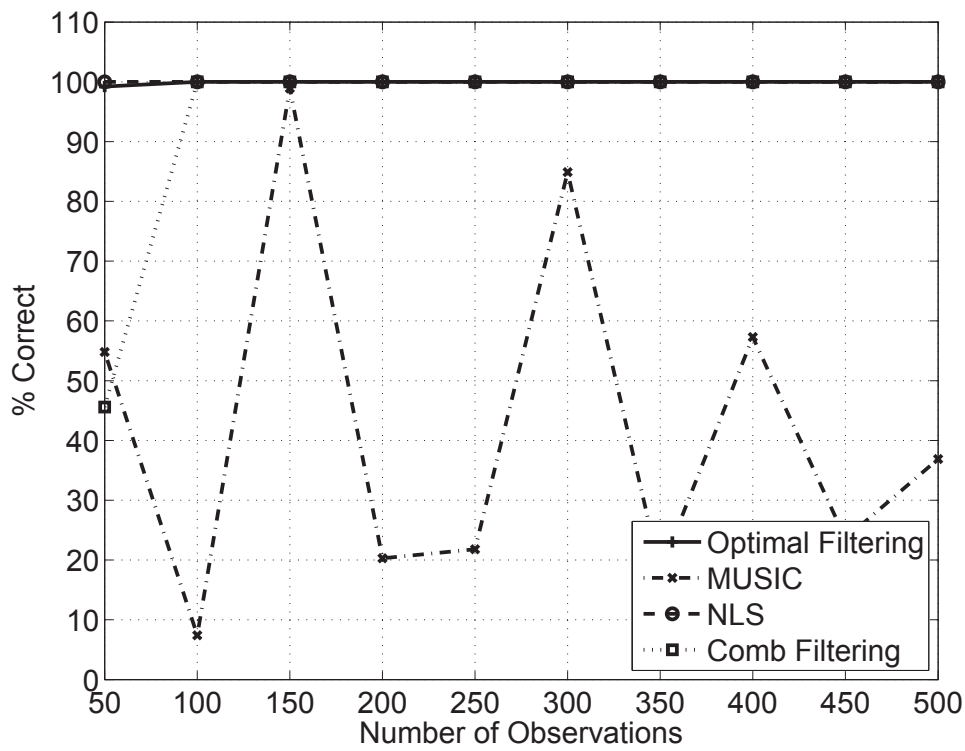

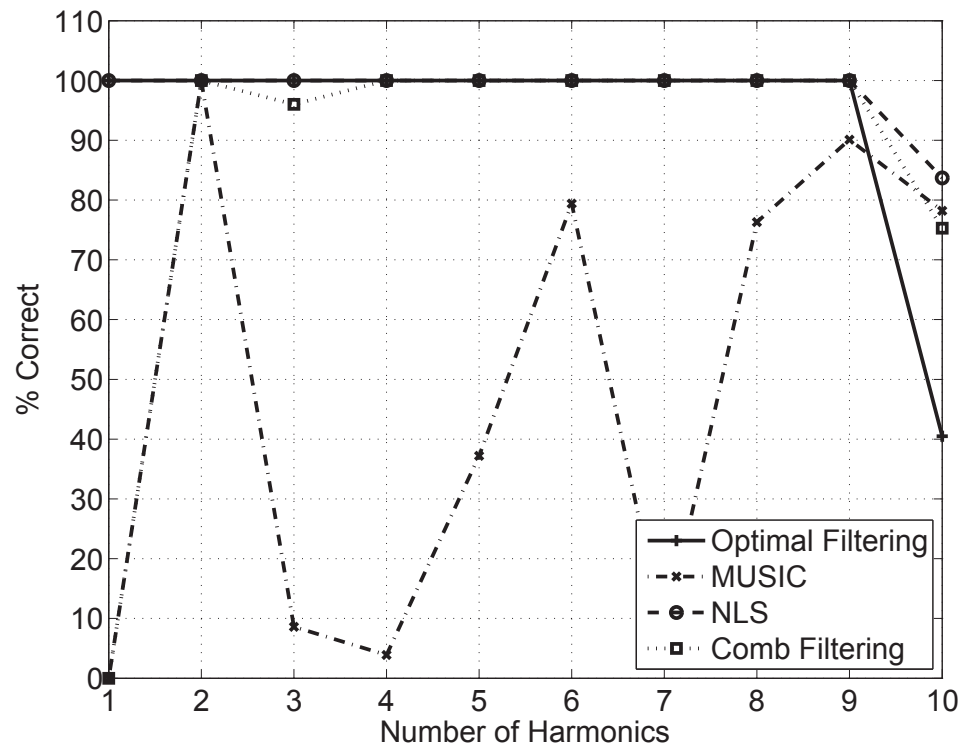**Figure 3 Percentage of correctly estimated model orders as a function of the PSNR for a multi-pitch signal**.

Figure 4 Percentage of correctly estimated model orders as a function of the number of samples *N* for a multi-pitch signal.

Figure 5 Percentage of correctly estimated model orders as a function of the number of harmonics *L_k* for a multi-pitch signal.

Figure 6 Percentage of correctly estimated model orders as a function of the filter length *M* for a multi-pitch signal.
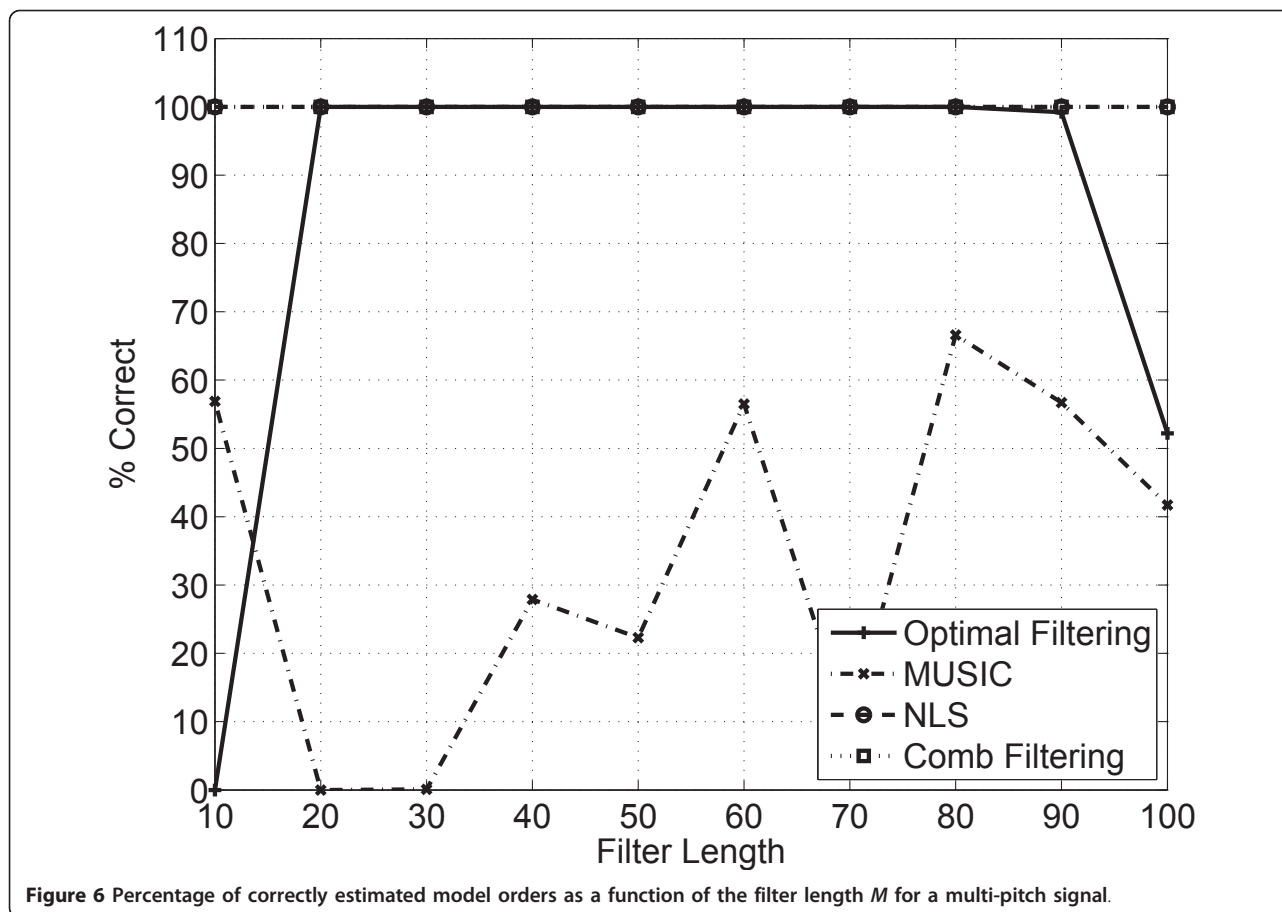
From the figures, it can be observed that the proposed method shows good performance for high PSNRs and N with the percentage approaching 100%. Furthermore, it can also be observed that the filter length should not be chosen too low or too close to $N/2$. As expected, the proposed method generally also exhibits better performance than the comb filtering approach. That the NLS, comb filtering, and optimal filtering methods all perform well for high PSNRs and N confirms that the MAP order estimation criterion indeed works well, as all of them are based on this criterion. The subspace method, MUSIC, appears not to work at all, and there is a simple explanation for this: the presence of a second interfering source has not been taken into account in any of the methods, and for MUSIC, this turns out to be a critical problem; indeed, the one source assumption leads to an incorrect noise subspace estimate. At this point, it should be stressed that it is possible to take multiple sources into account in MUSIC at the cost of an increased computational complexity [10,16]. While the method based on optimal filtering appears to sometimes exhibit slightly worse performance than NLS in terms of estimating the model order, it generally outperforms both MUSIC and the NLS with respect to fundamental frequency estimation under adverse conditions, in particular

when multiple periodic sources are present at the same time [16], something that happens frequently in audio signals. That is, unless the NLS approach is modified to either iteratively estimate the parameters of the individual sources using expectation maximization (EM) like iterations or is modified to incorporate the presence of multiple sources in the cost function [16]. The latter approach is to be avoided as it requires multi-dimensional nonlinear optimization. Overall, it can be concluded that the optimal filtering methods form an intriguing alternative for joint fundamental frequency and order estimation, especially so for multi-pitch signals.

Summary

A number of filtering methods for fundamental frequency estimation have been considered in this paper, namely the classical comb filtering and maximum likelihood methods along with some more recent methods based on optimal filtering. The latter approaches are generalizations of Capon's classical optimal beamformer. These methods have recently been demonstrated to show great potential for high-resolution pitch estimation. In this paper, we have extended these methods to account for an unknown number of harmonics, a

quantity also known as the model order, by deriving a model-specific order estimation criterion based on the maximum a posteriori principle. This has led to joint fundamental frequency and order estimators that can be applied in situations where the model order cannot be known a priori or may change over time, as is the case in speech and audio signals. Additionally, some new analyses of the optimal filtering methods and their properties have been provided. Moreover, a computationally efficient order-recursive implementation that is much faster than a direct implementation has been proposed. Finally, the optimal filtering methods have been demonstrated, in computer simulations, to have good performance in terms of the percentage of correctly estimated model orders when multiple sources are present.

Abbreviations

AIC: Akaike information criterion; EM: expectation maximization; MDL: minimum description length criterion; NLS: nonlinear least-squares; PDF: probability density function; PSNR: pseudo signal-to-noise.

Acknowledgements

Author details

[1]Department of Architecture, Design and Media Technology, Aalborg University, Aalborg, Denmark [2]Department of Mathematical Statistics, Lund University, Lund, Sweden [3]Department of Electronic Systems, Aalborg University, Aalborg, Denmark

Competing interests

The authors declare that they have no competing interests.

References

1. American Standards Association (ASA), Acoustical Terminology, SI, 1-1960, New York, 1960.
2. KW Chan, HC So, Accurate frequency estimation for real harmonic sinusoids. IEEE Signal Process Lett. **11**(7), 609–612 (2004)
3. M Ross, H Shaffer, A Cohen, R Freudberg, H Manley, Average magnitude difference function pitch extractor. IEEE Trans Acoust Speech Signal Process. **22**(5), 353–362 (1974)
4. L Rabiner, On the use of autocorrelation analysis for pitch detection. IEEE Trans Acoust Speech Signal Process. **25**(1), 24–33 (1977)
5. Y Medan, E Yair, D Chazan, Super resolution pitch determination of speech signals. IEEE Trans Signal Process. **39**(1), 40–48 (1991)
6. A de Cheveigné, H Kawahara, YIN, a fundamental frequency estimator for speech and music. J Acoust Soc Am. **111**(4), 1917–1930 (2002)
7. D Talkin, A robust algorithm for pitch tracking (RAPT). in *Speech Coding and Synthesis,* Chap. 5, ed. by Kleijn WB, Paliwal KK (Elsevier Science B.V., New York, 1995), pp. 495–518
8. MG Christensen, SH Jensen, SV Andersen, A Jakobsson, Subspace-based fundamental frequency estimation. in *Proceedings of the European Signal Processing Conference*, 2004, pp. 637–640
9. MG Christensen, A Jakobsson, SH Jensen, Joint high-resolution fundamental frequency and order estimation. IEEE Trans Audio Speech Lang Process. **15**(5), 1635–1644 (2007)
10. MG Christensen, A Jakobsson, SH Jensen, Fundamental frequency estimation using the shift-invariance property. in *Record of the Asilomar Conference on Signals, Systems, and Computers*, 2007, pp. 631–635
11. H Li, P Stoica, J Li, Computationally efficient parameter estimation for harmonic sinusoidal signals. Signal Process. **80**, 1937–1944 (2000)
12. M Noll, Pitch determination of human speech by harmonic product spectrum, the harmonic sum, and a maximum likelihood estimate. in *Proceedings of the Symposium on Computer Processing Communications*, 1969, pp. 779–797
13. BG Quinn, PJ Thomson, Estimating the frequency of a periodic function. Biometrika **78**(1), 65–74 (1991)
14. D Kundu, S Nandi, A note on estimating the fundamental frequency of a periodic function. Elsevier Signal Process. **84**, 653–661 (2004)
15. M Lavielle, C Lévy-Leduc, Semiparametric estimation of the frequency of unknown periodic functions and its application to laser vibrometry signals. IEEE Trans Signal Process. **53**(7), 2306–2314 (2005)
16. MG Christensen, P Stoica, A Jakobsson, SH Jensen, Multi-pitch estimation. Signal Process. **88**(4), 972–983 (2008)
17. AM Noll, Cepstrum pitch determination. J Acoust Soc Am. **41**(2), 293–309 (1967)
18. AT Cemgil, HJ Kappen, D Barber, A generative model for music transcription. IEEE Trans Audio Speech Lang Process. **14**(2), 679–694 (2006)
19. AT Cemgil, Bayesian music transcription. Ph.D. dissertation, Nijmegen University, 2004.
20. S Godsill, M Davy, Bayesian harmonic models for musical pitch estimation and analysis. in *Proceedings of the IEEE International Conference on Acoust., Speech, Signal Processing*, **2**, 1769–1772 2002
21. J Moorer, The optimum comb method of pitch period analysis of continuous digitized speech. IEEE Trans Acoust Speech Signal Process. **22**(5), 330–338 (1974)
22. J Lim, A Oppenheim, L Braida, Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition. IEEE Trans Acoust Speech Signal Process. **26**(4), 354–358 (1978)
23. A Nehorai, B Porat, Adaptive comb filtering for harmonic signal enhancement. IEEE Trans Acoust Speech Signal Process. **34**(5), 1124–1138 (1986)
24. MG Christensen, A Jakobsson, in *Multi-Pitch Estimation*. Synthesis Lectures on Speech & Audio Processing, vol. 5 (Morgan & Claypool Publishers, San Rafael, CA, 2009)
25. MG Christensen, JH Jensen, A Jakobsson, SH Jensen, On optimal filter designs for fundamental frequency estimation. IEEE Signal Process Lett. **15**, 745–748 (2008)
26. MG Christensen, A Jakobsson, Optimal filter designs for separating and enhancing periodic signals. IEEE Trans Signal Process. **58**(12), 5969–5983 (2010)
27. PM Djuric, Asymptotic MAP criteria for model selection. IEEE Trans Signal Process. **46**, 2726–2735 (1998)
28. S Godsill, M Davy, Bayesian computational models for inharmonicity in musical instruments. in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2005, pp. 283–286
29. M Davy, S Godsill, J Idier, Bayesian analysis of western tonal music. J Acoust Soc Am. **119**(4), 2498–2517 (2006)
30. MG Christensen, P Vera-Candeas, SD Somasundaram, A Jakobsson, Robust subspace-based fundamental frequency estimation. in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2008, pp. 101–104
31. EB George, MJT Smith, Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model. IEEE Trans Speech Audio Process. **5**(5), 389–406 (1997)
32. EB George, MJT Smith, Generalized overlap-add sinusoidal modeling applied to quasi-harmonic tone synthesis. in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1993, pp. 165–168
33. P Stoica, R Moses, *Spectral Analysis of Signals*. (Pearson Prentice Hall, Upper Saddle River, NJ, 2005)
34. DJ Hermes, Measurement of pitch by subharmonic summation. J Acoust Soc Am. **83**(1), 257–264 (1988)
35. P Stoica, A Jakobsson, J Li, Matched-filterbank interpretation of some spectral estimators. Signal Process. **66**(1), 45–59 (1998)
36. RM Gray, Toeplitz and circulant matrices: a review. Found. Trends Commun Inf Theory. **2**(3), 155–239 (2006)
37. EJ Hannan, B Wahlberg, Convergence rates for inverse Toeplitz matrix forms. J Multivar Anal. **31**, 127–135 (1989)
38. P Stoica, H Li, J Li, Amplitude estimation of sinusoidal signals: survey, new results and an application. IEEE Trans Signal Process. **48**(2), 338–352 (2000)

39. H Akaike, A new look at the statistical model identification. IEEE Trans
    Autom Control **19**, 716–723 (1974)
40. J Rissanen, Modeling by shortest data description. *Automatica* **14**, 468–478
    (1978)
41. G Schwarz, Estimating the dimension of a model. Ann Stat. **6**, 461–464
    (1978)
42. PM Djuric, A model selection rule for sinusoids in white Gaussian noise.
    IEEE Trans Signal Process. **44**(7), 1744–1751 (1996)
43. P Stoica, Y Selen, Model-order selection: a review of information criterion
    rules. IEEE Signal Process Mag. **21**(4), 36–47 (2004)
44. MA Lagunas, ME Santamaria, A Gasull, A Moreno, Maximum likelihood filters
    in spectral estimation problems. Elsevier Signal Process. **10**(1), 19–34 (1986)
45. S Haykin, *Adaptive Filter Theory*, 3rd edn. (Prentice-Hall, Upper Saddle River,
    NJ, 1996)