Aalborg Universitet

Joint DOA and Fundamental Frequency Estimation Methods based on 2-D Filtering

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

JOINT DOA AND FUNDAMENTAL FREQUENCY ESTIMATION METHODS BASED ON 2-D FILTERING

Jesper Rindom Jensen[†], Mads Græsbøll Christensen[‡] and Søren Holdt Jensen[†]

[†] Dept. of Electronic Systems, Aalborg University
Fredrik Bajers Vej 7, 9220, Aalborg, Denmark
email: {jrj,shj}@es.aau.dk

[‡] Dept. of Media Technology, Aalborg University
Niels Jernes Vej 14, 9220, Aalborg, Denmark
email: mgc@imi.aau.dk

ABSTRACT

It is well-known that filtering methods can be used for processing of signals in both time and space. This comprises, for example, fundamental frequency estimation and direction-of-arrival (DOA) estimation. In this paper, we propose two novel 2-D filtering methods for joint estimation of the fundamental frequency and the DOA of spatio-temporarily sampled periodic signals. The first and simplest method is based on the 2-D periodogram, whereas the second method is a generalization of the 2-D Capon method. In the experimental part, both qualitative and quantitative measurements show that the proposed methods are well-suited for solving the joint estimation problem. Furthermore, it is shown that the methods are able to resolve signals separated sufficiently in only one dimension. In the case of closely spaced sources, however, the 2-D Capon-based method shows the best performance.

1. INTRODUCTION

In the last couple of decades, filtering methods have been used for processing of both spatial and temporal signals [1, 2]. Processing of spatial signals is also known as array signal processing and within this field, filtering methods are better known as beamformers [1]. One application of this, is processing of audio and speech signals recorded using a multi-microphone setup as considered in [3]. A common task is to estimate the direction-of-arrival (DOA) of such signals. However, often it is also desired to estimate the fundamental frequency of the self-same signals. Lately, it has been shown [4] that filtering methods are useful in this context as well. In the rest of the paper will refer to the fundamental frequency as the pitch. It is essential to estimate both the DOA and the pitch, since these features are useful for both separation and enhancement [5]. Furthermore, the pitch is also relevant regarding compression [6].

In many applications, such as hands-free communication, teleconferencing, surveillance systems and hearing-aids, it is necessary to know both the DOA and the pitch of speech and audio signals. This is needed for, e.g., tracking, separation and enhancement purposes. Therefore, joint pitch and DOA estimation is a relevant problem. We can formulate the joint estimation problem as follows: consider a periodic source $s(n_t)$ impinging on an array containing N_s sensors. On the n_sth sensor, the periodic source is corrupted by the noise source $w_{n_s}(n_t)$. The signal sampled by the n_sth sensor, for $n_t = 0, \ldots, N_t - 1$ and $n_s = 0, \ldots, N_s - 1$, can then be written as

$$x_{n_s}(n_t) = s(n_t - \tau_{n_s}) + w_{n_s}(n_t) , \qquad (1)$$

where τ_{n_s} is the time delay of the signal on sensor n_s compared to a reference point. Note that the DOA can be estimated by realizing that there is a relationship between the DOA and the time delay. The relation depend on the array structure. In this paper, we assume a uniform linear array (ULA) and that the signals of interest are located in the so-called far-field. This implies a simple relationship between the DOA and the time delay. The problem considered in this paper, is join estimation of the DOA θ and the pitch ω_t of the periodic source $s(n_t)$ which can also be modeled as

$$s(n_t) = \sum_{l=1}^{L} \alpha_l e^{j\omega_t l n_t} , \qquad (2)$$

where L is the model order and $\alpha_l = \Upsilon_l e^{j\phi_l}$ is the complex amplitude of the lth sinusoid with $\Upsilon_l > 0$ and ϕ_l being the amplitude and the phase, respectively. In this paper, we consider the model order as being known. The model in (2) allows us to consider the signal of interest as several narrowband sources. The narrowband assumption is a key assumption in many array processing methods and we also employ in this paper.

Recently, the problem of joint pitch and DOA estimation has attracted considerable attention. Some of the first approaches to solve the problem only considered how the frequencies of single 2-D sinusoids can be estimated. A few examples of such methods are [7], where a state-space realization technique is used, [8] which is based on the 2-D Capon method, [9] which is based on the ES-PRIT method, and [10] where a signal-dependent multistage wiener filter (MWF) is used. The DOA and the pitch should, when possible, be estimated jointly for several reasons. For example, by estimating the parameters jointly we can process signals separated sufficiently in only one dimension as opposed to 1-D methods. Recently, a few methods have been proposed for joint DOA and pitch estimation; in [11] a ML-based method is proposed; in [11, 12] subspace-based methods are proposed; in [13] a correlation-based method is proposed; and in [14] a spatio-temporal filtering method based on the LCMV beamformer is proposed. In this paper, we present two novel methods for DOA and pitch estimation. Both methods are 2-D filtering methods based on a filter-bank interpretation of the periodogram and a generalization of the 2-D Capon method, respectively. As opposed to the method in [14], we do not require any prior knowledge on the spatial or temporal characteristics, other than a harmonic structure, since we propose to estimate the DOA and pitch jointly. In cases with colored noise, the 2-D Capon-based method is preferred because of its excellent performance in a multi-source scenario. However, the 2-D periodogram-based method may in some cases be preferred because of its lower computational complexity.

The rest of the paper is organized as follows: in Section 2 we introduce the joint estimation problem and present the proposed methods. Section 3 contains the experimental part of the paper and, finally, Section 4 concludes our work.

2. PROPOSED METHODS

In this section, we briefly review the concept of 2-D filtering methods for spectral estimation and we present the proposed methods. In 2-D filtering methods for spectral estimation, it is desired to design a filter which passes a signal component with a given frequency pair undistorted. At the same time, the filter should attenuate signal components at all other frequency pairs. The two different filter design procedures used in the two proposed methods are described following.

Assume that we have a matrix $\mathbf{X}_D$ of dimension $N_s \times N_t$ containing our spatio-temporarily sampled data. Note that N_s and N_t

are the numbers of spatial and temporal samples, respectively. The input data is then to be filtered by a $M_s \times M_t$ order 2-D finite impulse response (FIR) filter

$$\mathbf{H}_{\omega_t,\omega_s} = \begin{bmatrix} H_{\omega_t,\omega_s}(0,0) & \cdots & H_{\omega_t,\omega_s}(0,M'_t) \\ \vdots & \ddots & \vdots \\ H_{\omega_t,\omega_s}(M'_s,0) & \cdots & H_{\omega_t,\omega_s}(M'_s,M'_t) \end{bmatrix}, \quad (3)$$

where $M'_s = M_s - 1$, $M'_t = M_t - 1$ and the filter is designed for the temporal and spatial frequencies ω_t and ω_s. The filter is then applied on sub-blocks $\mathbf{X}_{n_s}(n_t)$ of the data matrix. One sub-block is defined as

$$\mathbf{X}_{n_s}(n_t) = \begin{bmatrix} x_{n_s}(n_t) & \cdots & x_{n_s}(n_t - M'_t) \\ \vdots & \ddots & \vdots \\ x_{n_s+M'_s}(n_t) & \cdots & x_{n_s+M'_s}(n_t - M'_t) \end{bmatrix}. \quad (4)$$

Due to the ULA and far-field assumptions, the spatial frequency is given by $\omega_s = \omega_t f_s \frac{d \sin\theta}{c}$, where f_s is the sampling frequency, d is the inter-sensor spacing, θ is the DOA in radians, and c is the wave propagation velocity. Following, we stack the filter response and the sub-blocks in (3) and (4), i.e.,

$$\mathbf{h}_{\omega_t,\omega_s} = \text{vec}\{\mathbf{H}_{\omega_t,\omega_s}\} \quad (5)$$
$$\mathbf{x}_{n_s}(n_t) = \text{vec}\{\mathbf{X}_{n_s}(n_t)\}, \quad (6)$$

with $\text{vec}\{\cdot\}$ denoting the column-wise stacking operator. Since we have mapped the filtering operation from 2-D to 1-D, it can be seen that the filter design somehow resembles that of 1-D filtering methods. As the first step in the design procedure we need to find an expression for the filter output power

$$E\{|y_{n_s}(n_t)|^2\} = E\{\mathbf{h}^H_{\omega_t,\omega_s}\mathbf{x}_{n_s}(n_t)\mathbf{x}^H_{n_s}(n_t)\mathbf{h}_{\omega_t,\omega_s}\} \quad (7)$$
$$= \mathbf{h}^H_{\omega_t,\omega_s}\mathbf{R}\mathbf{h}_{\omega_t,\omega_s}, \quad (8)$$

where $\mathbf{R}$ is the covariance matrix

$$\mathbf{R} = E\{\mathbf{x}_{n_s}(n_t)\mathbf{x}^H_{n_s}(n_t)\}. \quad (9)$$

Note that $E\{\cdot\}$ and $(\cdot)^H$ denotes the expectation operator and the complex transpose, respectively. Often we do not have access to the true covariance matrix, which we therefore will replace with the sample covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{(N_s - M'_s)(N_t - M'_t)} \sum_{p=0}^{N_s-M_s} \sum_{q=0}^{N_t-M_t} \mathbf{x}_p(n_t-q)\mathbf{x}^H_p(n_t-q). \quad (10)$$

The next task is to design the filter such that the output power is minimized subject to a distortionless constraint at desired frequencies. The filter design procedure is what differs between the two proposed method.

2.1 2-D Periodogram-based Method

In the 2-D periodogram-based method, we use a filter-bank structure constituted by stacked 2-D filter responses. We assume that there is no cross-talk between the filters in the filter-bank, which allows us to write the total filter-bank output power as

$$\sum_{l=1}^{L} E\{|y_{n_s,l}(n_t)|^2\} = \sum_{l=1}^{L} \mathbf{h}^H_{l\omega_t,l\omega_s}\mathbf{R}\mathbf{h}_{l\omega_t,l\omega_s}. \quad (11)$$

Note that $\mathbf{h}_{l\omega_t,l\omega_s}$ is the impulse response of the lth filter in the filter-bank. We then define the filter-bank matrix as

$$\mathbf{H}_{\text{fb},(\omega_t,\omega_s)} = [\mathbf{h}_{\omega_t,\omega_s} \quad \cdots \quad \mathbf{h}_{L\omega_t,L\omega_s}]. \quad (12)$$

Having introduced the filter-bank matrix, we can rewrite the total filter-bank output power in 11 as

$$\sum_{l=1}^{L} E\{|y_{n_s,l}(n_t)|^2\} = \text{Tr}\left\{\mathbf{H}^H_{\text{fb},(\omega_t,\omega_s)}\mathbf{R}\mathbf{H}_{\text{fb},(\omega_t,\omega_s)}\right\}, \quad (13)$$

with $\text{Tr}\{\cdot\}$ denoting the trace operator. In the 2-D periodogram-based method, it is assumed that the input signal is white Gaussian noise, hence, the method is independent on the signal statistics. Using this assumption, it can be shown that the individual filters in the filter-bank are constituted by Fourier vectors which will ensure a unit gain at the desired frequencies. However, the attenuation of other frequency components will not be optimal since the signal statistics are not used in the design procedure. The lth filter response is then given by

$$\mathbf{h}_{l,(\omega_t,\omega_s)} = \mathbf{a}_{l\omega_t,\omega_s}, \quad (14)$$

where

$$\mathbf{a}_{\omega_t,\omega_s} = \mathbf{a}_{\omega_t} \otimes \mathbf{a}_{\omega_s} \quad (15)$$
$$\mathbf{a}_{\omega_k} = \begin{bmatrix} 1 & e^{-j\omega_k} & \cdots & e^{-jM'_k\omega_k} \end{bmatrix}^T. \quad (16)$$

By inserting (14) into (13) we get that

$$\sum_{l=1}^{L} E\{|y_{n_s,l}(n_t)|^2\} = \text{Tr}\{\mathbf{A}^H_{\omega_t,\omega_s}\mathbf{R}\mathbf{A}_{\omega_t,\omega_s}\} \quad (17)$$
$$= J_{\text{P},(\omega_t,\omega_s)}, \quad (18)$$

with $\mathbf{A}_{\omega_t,\omega_s} = [\mathbf{a}_{\omega_t,\omega_s} \quad \cdots \quad \mathbf{a}_{L\omega_t,\omega_s}]$. We can then obtain joint estimates of the pitch and the DOA by maximizing the total filter-bank output power over sets of candidate DOAs Θ and pitch frequencies Ω, i.e.,

$$(\hat{\theta}, \hat{\omega}) = \underset{(\theta,\omega)\in\Theta\times\Omega}{\text{argmax}} J_{\text{P},(\omega_t,\omega_s)}. \quad (19)$$

2.2 2-D Capon-based Method

The proposed 2-D Capon-based method is a generalization of the 2-D Capon method [15]. The generalization is obtained by introducing multiple harmonic constraints in the filter design. That is, the proposed 2-D Capon-based method relies on a single 2-D filter. We design the 2-D filter by minimizing the output power subject to distortionless constraints on the harmonics, i.e.,

$$\min_{\mathbf{h}} \mathbf{h}^H_{\omega_t,\omega_s}\mathbf{R}\mathbf{h}_{\omega_t,\omega_s} \text{ s.t. } \mathbf{h}^H_{\omega_t,\omega_s}\mathbf{a}_{l\omega_t,\omega_s} = 1, \quad (20)$$
$$\text{for } l = 1,\ldots,L. \quad (21)$$

The optimization problem is easily solved by using the Lagrange multiplier method which leads to the following result

$$\mathbf{h}_{\omega_t,\omega_s} = \mathbf{R}^{-1}\mathbf{A}_{\omega_t,\omega_s}(\mathbf{A}^H_{\omega_t,\omega_s}\mathbf{R}^{-1}\mathbf{A}_{\omega_t,\omega_s})^{-1}\mathbf{1}, \quad (22)$$

with $\mathbf{1}$ being a $L \times 1$ vector containing ones. Inserting the optimal filter expression into the filter output power expression leads to

$$E\{|y_{n_s}(n_t)|^2\} = \mathbf{1}^H(\mathbf{A}^H_{\omega_t,\omega_s}\mathbf{R}^{-1}\mathbf{A}_{\omega_t,\omega_s})^{-1}\mathbf{1} \quad (23)$$
$$= J_{\text{C},(\omega_t,\omega_s)}. \quad (24)$$

Finally, we can jointly estimate the DOA and the pitch by maximizing the filter output power for a sets of candidate DOAs Θ and pitch frequencies Ω

$$(\hat{\theta}, \hat{\omega}) = \underset{(\theta,\omega)\in\Theta\times\Omega}{\text{argmax}} J_{\text{C},(\omega_t,\omega_s)}. \quad (25)$$
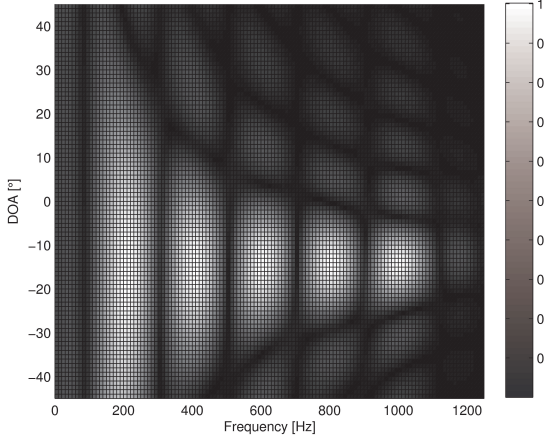
Figure 1: Contour plot in dB of a 2-D periodogram-based filter-bank response designed for a signal constituted by five harmonics having a pitch of 200 Hz and a DOA of -15°.
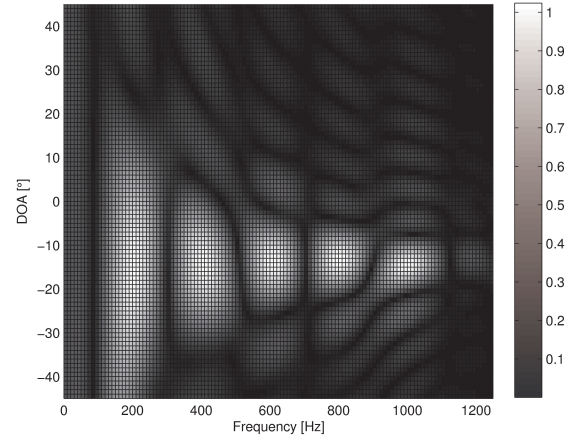

Figure 2: Contour plot in dB of a 2-D Capon-based filter response designed for a signal constituted by five harmonics having a pitch of 200 Hz and a DOA of -15°. The signal was corrupted by white Gaussian noise with an SNR of -40 dB.

2.3 Refined Estimates

In cases where there are high resolution requirements, the proposed methods may imply a huge computational burden since they rely on a grid search. However, to circumvent this issue we can instead use a smaller grid to obtain an initial estimate of the parameters which can then be refined by making a gradient search. Note that in the following derivations we leave out the dependencies on ω_s and ω_t for a simpler notation. The first order derivatives of the filter output powers are readily obtained as

$$\begin{bmatrix} \frac{\partial J_P}{\partial \theta} \\ \frac{\partial J_P}{\partial \omega_t} \end{bmatrix} = \frac{2}{M^2} \text{Re} \left\{ \begin{bmatrix} \text{Tr}\{\mathbf{A}^H \mathbf{R} \mathbf{B}_\theta\} \\ \text{Tr}\{\mathbf{A}^H \mathbf{R} \mathbf{B}_{\omega_t}\} \end{bmatrix} \right\} \tag{26}$$

$$\begin{bmatrix} \frac{\partial J_C}{\partial \theta} \\ \frac{\partial J_C}{\partial \omega_t} \end{bmatrix} = -2\text{Re} \left\{ \begin{bmatrix} \mathbf{1}^H \mathbf{Q} \mathbf{A}^H \mathbf{R}^{-1} \mathbf{B}_\theta \mathbf{Q} \mathbf{1} \\ \mathbf{1}^H \mathbf{Q} \mathbf{A}^H \mathbf{R}^{-1} \mathbf{B}_{\omega_t} \mathbf{Q} \mathbf{1} \end{bmatrix} \right\} , \tag{27}$$

where $\mathbf{Q} = (\mathbf{A}^H \mathbf{R}^{-1} \mathbf{A})^{-1}$, and $\text{Re}\{\cdot\}$ denotes taking the real part. The entries in the $\mathbf{B}$ matrices are given by

$$[\mathbf{B}_\theta]il = -j\omega_t f_s l \frac{d\cos\theta}{c} k_{s,i} e^{-j\omega_t l \left(f_s \frac{d\sin\theta}{c} k_{s,i} + k_{t,i} \right)} \tag{28}$$

$$[\mathbf{B}_{\omega_t}]_{il} = -jl \left(f_s \frac{d\sin\theta}{c} k_{s,i} + k_{t,i} \right) e^{-j\omega_t l \left(f_s \frac{d\sin\theta}{c} k_{s,i} + k_{t,i} \right)} . \tag{29}$$

The intermediate k variables are defined as

$$k_{s,i} = (i-1) \mod M_s \tag{30}$$

$$k_{t,i} = \left\lfloor \frac{i-1}{M_s} \right\rfloor , \tag{31}$$

with $(\cdot \mod \cdot)$ and $\lfloor \cdot \rfloor$ denoting the modulus and the flooring operators, respectively. We can then obtain refined parameter estimates by using an iterative procedure

$$\begin{bmatrix} \hat{\theta}^{(i+1)} \\ \hat{\omega}_t^{(i+1)} \end{bmatrix} = \begin{bmatrix} \hat{\theta}^{(i)} \\ \hat{\omega}_t^{(i)} \end{bmatrix} + \delta \nabla J , \tag{32}$$

where i is the iteration number, δ is a small positive constant found through line search and $\nabla J = \begin{bmatrix} \frac{\partial J}{\partial \theta} & \frac{\partial J}{\partial \omega_t} \end{bmatrix}^T$.
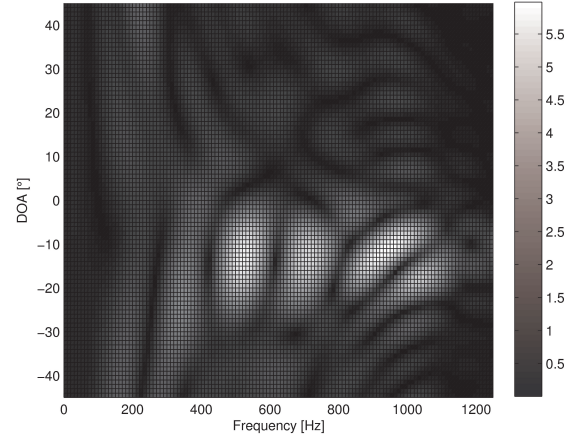

Figure 3: Contour plot in dB of a 2-D Capon-based filter response designed for a signal constituted by five harmonics having a pitch of 200 Hz and a DOA of -15°. The signal was corrupted by white Gaussian noise with an SNR of 10 dB.

3. EXPERIMENTAL RESULTS

We will now consider the evaluation of the proposed methods. In all of the experiments described in the rest of this section, an ULA was assumed having an inter-sensor spacing of $d = \frac{c}{f_s}$ where $c = 343.2$ m/s is the speed of sound in air at 20° C. Furthermore, in all experiments the sampling frequency was $f_s = 2.5$ kHz. First, we evaluate the functionality of the filters in terms of investigating the filter response. In Fig. 1, a filter response of a 2-D periodogram-based filter-bank is shown. The filter orders were in this case set to $M_t = M_s = 20$. The filter-bank was designed for a signal constituted by five harmonics with a pitch of 200 Hz and a DOA of -15°. As mentioned, this filter is independent of the signal statistics and will therefore have a rather similar attenuation of signal components not having both one of the target harmonic frequencies and the target DOA. At the target harmonic frequencies and DOA, the filter will have a unit gain. These observations can also be made from the experimental result which verifies the filter design. Likewise, we also conducted experiments on the 2-D Capon-based filter response. In
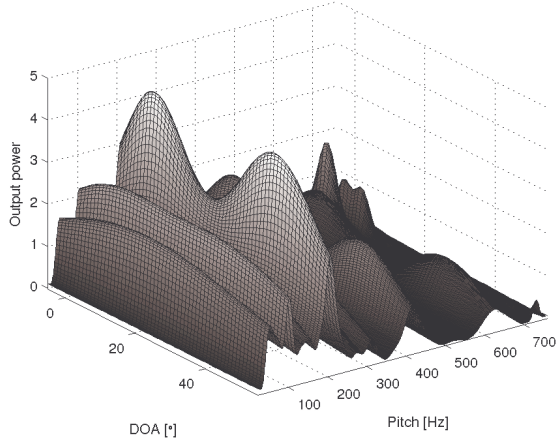
Figure 4: Output power of the 2-D periodogram-based filter-bank of orders $M_t = 30$ and $M_s = 10$ applied on a mixture of two signals with DOAs of 4° and 40°, respectively, and both with a pitch of 213 Hz. The SNR with respect to each signal was 10 dB.
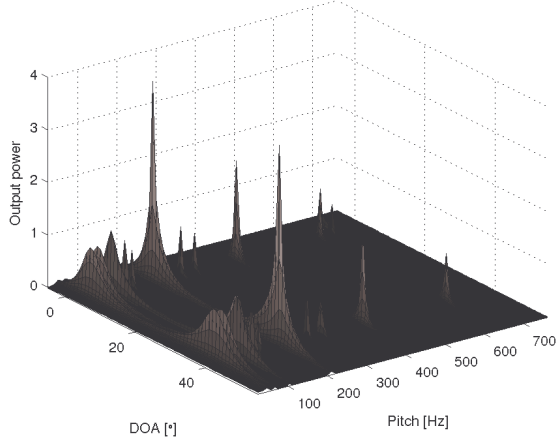

Figure 5: Output power of the 2-D Capon-based filter of orders $M_t = 30$ and $M_s = 10$ applied on a mixture of two signals with DOAs of 4° and 40°, respectively, and both with a pitch of 213 Hz. The SNR with respect to each signal was 10 dB.
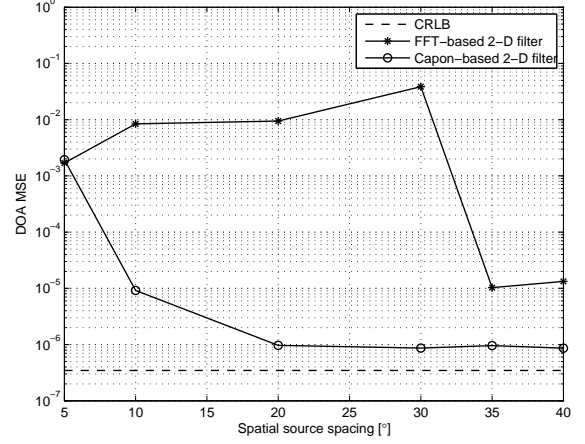

Figure 6: MSE in a two-source scenario as a function of the source spacing in degrees, and the CRLB for the single-source scenario.

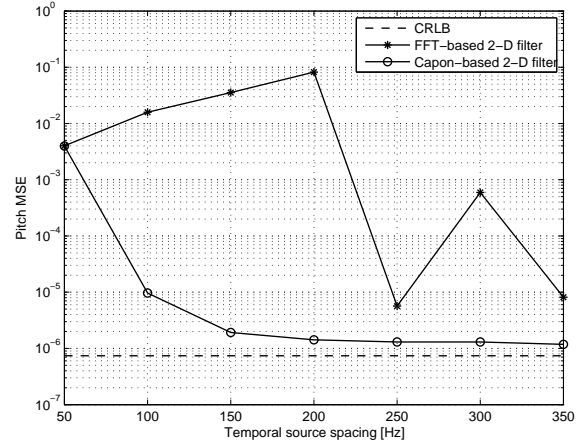Figure 7: MSE in a two-source scenario as a function of the source spacing in Hz, and the CRLB for the single-source scenario.

these experiments the sample lengths were set to $N_t = N_s = 80$ and the filter orders were the same as in the previous experiment. The filter was designed for a signal having five harmonics with a pitch of 200 Hz and a DOA of -15°. Also, the signal was corrupted by white Gaussian noise and the experiment was conducted for SNRs of -40 dB and 10 dB. The results are depicted in Fig. 2 and 3, respectively. For the low SNR of -40 dB the filter response somehow resembles that of the 2-D periodogram-based filter which can also be verified mathematically. At high SNRs the 2-D Capon-based method seems to suffer from leakage, since it has a relatively high gain at off pitch frequencies and DOAs. This is, however, characteristic for the minimum variance distortionless response (MVDR) principle.

Following, we evaluate the proposed methods ability to jointly estimate the pitch and DOA in a multi-source scenario. First, we have calculated the output power of both the 2-D periodogram-based filter and the 2-D Capon-based filter for several candidate pitch frequencies and DOAs. In these experiments, the number of sensors was $N_s = 30$, the sample length was $N_t = 100$ and the filter orders were $M_t = 30$ and $M_s = 10$. The filters were applied on a

mixture of two signals both constituted by four sinusoids. The two signals had DOAs of 4° and 40°, respectively, and they both had a pitch of 213 Hz. The results from the experiments are depicted in Fig. 4 and 5, respectively. The first observation from these experiments are, that the pitch frequencies and DOAs of the two sources can be estimated correctly using both methods by taking the arguments of the two largest peaks of the filter output powers. Also, we observe that the peaks in the filter output power are much narrower for the 2-D Capon-based method which indicates that the 2-D Capon-based method will be superior when it comes to resolving closely-spaced sources.

We will now evaluate further on the proposed methods ability to resolve closely spaced sources. For this purpose we have conducted some Monte-Carlo simulations on the estimation error as a function of the source spacing in a two-source scenario. Due to a high computational complexity, we assumed that the pitch and DOA estimates were close to the true pitch and DOA which allowed us to just doing a gradient search from the true pitch and DOA instead of doing a fine grid search. In the first series of Monte-Carlo simulations we measured the MSE of the DOA estimates as a function of the source spacing in degrees. The two sources were each constituted by four harmonics having a pitch of 236 Hz. The DOA was fixed to -20° for one of the sources while the other was varied.

White Gaussian noise was added to the signal such that the SNR with respect to each source was 10 dB. The sample lengths were $N_t = N_s = 30$ while the filter orders were $M_t = M_s = 10$. We conducted 500 Monte-Carlo simulations for each of the different source spacings and the results are depicted in Fig. 6. From the results, it is clearly seen that the 2-D Capon-based method are superior in multi-source scenarios. The 2-D periodogram-based method shows some thresholding behavior around a source spacing of $35°$ while the 2-D Capon-based method does not show any thresholding behavior before a source spacing of only $10°$. Note that the MSE for the 2-D periodogram-based method is not necessarily decreasing when the source spacing is increased. This is due to the fact that this method is signal independent and the MSE will therefore depend heavily on if the filter response occasionally has a dip at the pitch and DOA of the interfering source or not. We also conducted a series of Monte-Carlo simulations where the MSE was measured as a function of the source spacing in Hz. In these simulations, the two sources were constituted by one 2-D sinusoid. Both signals were having a DOA of $7°$. The frequency of one of the sources was fixed to 200 Hz while the frequency of the other source was varied. The noise conditions, sample lengths and filter orders were the same as in the other series of Monte-Carlo simulations. Again, we conducted 500 Monte-Carlo simulations for each source spacing and the results are depicted in Fig. 7. We see from the results that the 2-D Capon-based method is better for resolving closely-spaced sources. The dips in MSE at certain spacings for the 2-D periodogram-based method can be explained in the same way as in the previous series of Monte-Carlo simulations. Note that the 2-D periodogram-based method shows thresholding behavior below a spacing of 250 Hz whereas the 2-D Capon-based method does not show any thresholding behavior until below a spacing of 100 Hz.

4. CONCLUSION

In this paper, we proposed two new 2-D filtering methods for joint estimation of the pitch and the DOA of periodic signals recorded in space and time by a ULA. Since the proposed methods are based on a harmonic model, they are relevant for all signals being periodic of nature such as audio and speech signals. The first proposed method is based on the 2-D periodogram, and it can thereby be implemented easily using the 2-D fast Fourier transform (FFT). By doing this, the computational complexity can be reduced significantly. The second proposed method is the 2-D Capon-based method, which is a generalization of the 2-D Capon method. Our experiments showed that both proposed methods can be used for jointly estimating the pitch frequencies and DOAs of multiple sources that are only separated sufficiently in one dimension, i.e., either space or time. We evaluated the proposed methods with respect to the sufficiency condition in regard to separation and these experiments showed that the 2-D Capon-based method outperforms the 2-D periodogram-based method in multi-source scenarios under adverse conditions.

REFERENCES

[1] H. L. Van Trees. *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. John Wiley & Sons, Inc., 2002.

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

[3] M. Brandstein and D. Ward. *Microphone Arrays - Signal Processing Techniques and Applications*. Springer-Verlag, 2001.

[4] M. G. Christensen and A. Jakobsson. Multi-pitch estimation. *Synthesis Lectures on Speech and Audio Processing*, 5(1):1–160, 2009.

[5] D. Wang and G. J. Brown. *Computational Auditory Scene Analysis - Principles, Algorithm, and Applications*. John Wiley & Sons, Inc., 2006.

[6] B. Edler and H. Purnhagen. Parametric audio coding. In *Proc. Conf. Signal Process.*, volume 1, pages 21–24, 2000.

[7] M. Viberg and P. Stoica. A computationally efficient method for joint direction finding and frequency estimation in colored noise. In *Rec. Asilomar Conf. Signals, Systems, and Computers*, 1998.

[8] A. Jakobsson, S. L. Jr. Marple, and P. Stoica. Computationally efficient two-dimensional Capon spectrum analysis. *IEEE Trans. Signal Process.*, 48(9):2651–2661, Sep. 2000.

[9] A. N. Lemma, A. J. van der Veen, and E. F. Deprettere. Analysis of joint angle-frequency estimation using ESPRIT. *IEEE Trans. Signal Process.*, 51(5):1264–1283, May 2003.

[10] T. Shu and X. Liu. Robust and computationally efficient signal-dependent method for joint DOA and frequency estimation. *EURASIP J. on Advances in Signal Processing*, 2008:16 pages, 2008.

[11] X. Qian and R. Kumaresan. Joint estimation of time delay and pitch of voiced speech signals. *Rec. Asilomar Conf. Signals, Systems, and Computers*, 1996.

[12] L. Y. Ngan, Y. Wu, H. C. So, P. C. Ching, and S. W. Lee. Joint time delay and pitch estimation for speaker localization. In *Proc. IEEE Int. Symp. Circuits and Systems*, volume 3, pages 722–725, May 2003.

[13] M. Wohlmayr and M. Képesi. Joint position-pitch extraction from multichannel audio. In *Proc. Interspeech*, 2007.

[14] J. Dmochowski, J. Benesty, and S. Affes. Linearly constrained minimum variance source localization and spectral estimation. *IEEE Trans. Audio, Speech, and Language Process.*, 16(8):1490–1502, 2008.

[15] J. Capon. High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE*, 57(8):1408–1418, Aug. 1969.