An Optimal Spatio-Temporal Filter for Extraction and Enhancement of Multi-Channel Periodic Signals

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

Published in:
Asilomar Conference on Signals, Systems and Computers. Conference Record

DOI (link to publication from Publisher):
10.1109/ACSSC.2010.5757857

Publication date:
2010

Document Version
Accepted author manuscript, peer reviewed version

AN OPTIMAL SPATIO-TEMPORAL FILTER FOR EXTRACTION AND ENHANCEMENT OF MULTI-CHANNEL PERIODIC SIGNALS

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 Architecture, Design and Media Technology, Aalborg University
Niels Jernes Vej 14, 9220, Aalborg, Denmark
email: mgc@imi.aau.dk

ABSTRACT

Filtering methods have been widely used for extraction of signals in both time and space. Recently, two multi-channel filters have been proposed which can be applied for extraction of multi-channel periodic signals. In these filters, the harmonic structure of periodic signals is exploited. The filters were based on the periodogram and the LCMV beamformer, respectively. The periodogram-based filter is unsuitable for multi-source scenarios whereas the LCMV-based filter has an erratic filter response behaviour at high SNRs. We propose an optimal filtering method which is useful in multi-source scenarios and which has a nicely behaving filter response for a larger range of SNRs compared to the LCMV-based filter. Our simulations show that our method solves the high SNR issue of the LCMV-based filter and that the proposed filter is applicable to real-life signals.

***Index Terms*—** Signal extraction, optimal filtering, microphone arrays, beamforming.

1. INTRODUCTION

In many applications, it is beneficial to separate or extract one or more desired signals from a mixture. A few examples of such applications are teleconferencing, surveillance systems and hearing-aids. Often, the signal of interest (SOI) in these applications is speech and/or musical instrument signals. These types of signals are known to be quasi-periodic. Thus, for short data segmentents, we can model such signals as

$$s(n_t) = \sum_{l=1}^{L} \alpha_l e^{jl\omega_t n_t} \text{ , for } n_t = 0, \ldots, N_t - 1 \text{ ,} \qquad (1)$$

where N_t is the number of temporal samples, L is the model order and $\alpha_l = A_l e^{j\phi_l}$ with $A_l > 0$ and ϕ_l being the real amplitude and the phase of the lth harmonic, respectively. Previously, it has been investigated how a single-channel signal as in (1) can be extracted from a noisy mixture using, for example, algebraic separation [1] and comb filtering [2]. More recently, optimal filter designs for fundamental frequency estimation were proposed [3–5], and these filters can be seen as either generalizations of the MVDR beamformer [6] or special cases of the LCMV beamformer [7]. While the filtering methods have a good parameter estimation performance in settings with multiple interfering sources, they perform poorly for extraction purposes. This is particularly true for the high signal-to-noise ratio (SNR) settings. In these settings, the filter design problem becomes ill-conditioned, hence, the poor extraction performance. In [8, 9], a set of optimal filters for extractions and enchancement purposes were derived. As opposed to the MVDR/LCMV-like optimal filters, these filters have a good performance regarding extraction of synthetic as well as real-life periodic sources.

Sometimes, however, the signal is recorded by an array of microphones in the aforementioned applications. The mentioned single-channel methods are therefore inappropriate in such cases. In a multi microphone scenario, we can write the signal observed at the n_sth microphone as

$$x_{n_s}(n_t) = s_{n_s}(n_t) + w_{n_s}(n_t) \text{ , for } n_t = 0, \ldots, N_t - 1 \text{ ,} \qquad (2)$$

where $s_{n_s}(n_t) = s(n_t - \tau_{n_s})$ is the SOI, $w_{n_s}(n_t)$ is the noise on the n_sth sensor, N_s is the number of microphones and τ_{n_s} is the time delay of the sound wave from sensor n_s to a reference point. We assume a uniform linear array (ULA) structure so the time delay is given by $\tau_{n_s} = n_s \frac{d \sin\theta}{c}$ for $\theta \in [-90°; 90°]$ where d is the microphone spacing, θ is the DOA and c is the wave propagation velocity. Combining (1) and (2) leads to a multi-channel harmonic model

$$\mathbf{x}_{n_s}(n_t) = \sum_{l=1}^{L} \alpha_l e^{jl\omega_t n_t} e^{-jl\omega_s n_s} \text{ ,} \qquad (3)$$

where $\omega_s = \omega_t f_s d c^{-1} \sin\theta$. Due to the harmonic structure, the voiced speech and audio extraction problem can be considered as extraction of L narrowband sources while, traditionally, voiced speech and audio have been considered broadband in multi-channel extraction methods. The narrowband simplification enables us to derive much simpler extraction algorithms which is evident from the following sections.

Recently, two methods for joint DOA and fundamental frequency esimation were proposed [10]. One of them was signal independent since it was based on the periodogram while the other was based on the LCMV beamformer and therefore signal dependent. Although these filtering methods could also be used for extraction of multi-channel periodic sources, they suffer from the same issues as the corresponding single-channel methods. In this paper, we therefore derive a new joint spatio-temporal optimal filter for extraction of (quasi-)periodic sources from multi microphone recordings. We will term the filter design method as the filtering-based multi-channel periodic signal extraction (FIMPSIX) method. The filter is designed optimally from the observed signal and is therefore signal adaptive. Like the filters in [8, 9], the proposed filter is inspired by the well known amplitude and phase estimation (APES) method [11]. We expect that the proposed filter will outperform the filtering methods in [10] regarding extraction, since this is the case for the analogous single-channel filtering methods. The main application of the

proposed method is extraction and enhancement, however, it can also be used for joint DOA and fundamental frequency estimation, model order selection and amplitude estimation of the individual harmonics.

The rest of the paper is organized as follows. In Section 2, we state the filter design problem and introduce the notation. We solve the filter design problem in Section 3. In Section 4, we describe the experimental evaluation of the proposed filter design, and, finally, we conclude on our work in Section 5.

2. JOINT SPATIO-TEMPORAL FILTER DESIGN PROBLEM

We consider the problem of designing a joint optimal spatio-temporal filter for extraction of periodic sources. Generally speaking, the output $y(n_t)$ of an FIR filter with the coefficients $h(n_s, m_t)$ from the input $x_{n_s}(n_t - m_t)$ can be written as

$$y(n_t) = \sum_{n_s=0}^{N_s-1} \sum_{m_t=0}^{M_t-1} h(n_s, m_t) x_{n_s}(n_t - m_t) \ , \tag{4}$$

for $n_{\{s,t\}} = 0, \ldots, N_{\{s,t\}-1}$. Our goal is to design the filter such that its output resembles a desired signal $\hat{y}(n_t)$ as much as possible in the mean squared error (MSE) sense. The MSE P is given by

$$P = \frac{1}{N_t - M_t + 1} \sum_{n_t=M_t-1}^{N_t-1} |y(n_t) - \hat{y}(n_t)|^2 \ . \tag{5}$$

In this filter design, the desired signal is defined as the noise-free signal given by the signal model in (1). If we insert (1) and (4) into (5) we get

$$P = \frac{1}{N_t - M_t + 1} \sum_{n_t=M_t-1}^{N_t-1} \left| \sum_{n_s=0}^{N_s-1} \sum_{m_t=0}^{M_t-1} h(n_s, m_t) x_{n_s}(n_t - m_t) - \sum_{l=1}^{L} \alpha_l e^{jl\omega_t n_t} \right|^2 \ . \tag{6}$$

Whereas we initially assume that the fundamental frequency ω_0 and the model order L are known, it is shown later how the proposed filter can also estimate these parameters. The expression in (6) can be simplified by introducing matrix/vector notation. Consider, for example, the filter and signal matrices, $\mathbf{H}$ and $\mathbf{X}(n_t)$, defined as

$$\mathbf{H} = \begin{bmatrix} h^*(0,0) & \cdots & h^*(0, M_t - 1) \\ \vdots & \ddots & \vdots \\ h^*(N_s - 1, 0) & \cdots & h^*(N_s - 1, M_t - 1) \end{bmatrix} \tag{7}$$

$$\mathbf{X}(n_t) = \begin{bmatrix} x_0(n_t) & \cdots & x_0(n_t - M_t + 1) \\ \vdots & \ddots & \vdots \\ x_{N_s-1}(n_t) & \cdots & x_{N_s-1}(n_t - M_t + 1) \end{bmatrix} \ , \tag{8}$$

where $(\cdot)^*$ denotes the complex conjugate. We define two new vectors $\mathbf{h} = \text{vec}\{\mathbf{H}\}$ and $\mathbf{x}(n_t) = \text{vec}\{\mathbf{X}(n_t)\}$ with $\text{vec}\{\cdot\}$ denoting the column-wise matrix stacking operator. This enable us to obtain a much more convenient MSE expression as

$$P = \frac{1}{N_t - M_t + 1} \sum_{n_t=M_t-1}^{N_t-1} |\mathbf{h}^H \mathbf{x}(n_t) - \boldsymbol{\alpha}^H \mathbf{e}(n_t)|^2 \ , \tag{9}$$

where

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \cdots & \alpha_L \end{bmatrix}^H \tag{10}$$

$$\mathbf{e}(n_t) = \begin{bmatrix} e^{j\omega_t n_t} & \cdots & e^{jL\omega_t n_t} \end{bmatrix}^T \ . \tag{11}$$

It turns out that we can expand the MSE expression in (9) as

$$P = \mathbf{h}^H \hat{\mathbf{R}} \mathbf{h} - \boldsymbol{\alpha}^H \mathbf{G} \mathbf{h} - \mathbf{h}^H \mathbf{G}^H \boldsymbol{\alpha} + \boldsymbol{\alpha}^H \mathbf{E} \boldsymbol{\alpha} \ , \tag{12}$$

with

$$\hat{\mathbf{R}} = \frac{1}{N_t - M_t + 1} \sum_{n_t=M_t-1}^{N_t-1} \mathbf{x}(n_t) \mathbf{x}^H(n_t) \tag{13}$$

$$\mathbf{G} = \frac{1}{N_t - M_t + 1} \sum_{n_t=M_t-1}^{N_t-1} \mathbf{e}(n_t) \mathbf{x}^H(n_t) \tag{14}$$

$$\mathbf{E} = \frac{1}{N_t - M_t + 1} \sum_{n_t=M_t-1}^{N_t-1} \mathbf{e}(n_t) \mathbf{e}^H(n_t) \ . \tag{15}$$

We recognize that $\hat{\mathbf{R}}$ is the spatio-temporal sample covariance matrix [10].

3. DERIVATION OF THE OPTIMAL FILTER

Following, we derive the optimal spatio-temporal filter by solving the filter design problem introduced in Section 2. First, if we differentiate and solve with respect to $\boldsymbol{\alpha}$ in (12) we get that

$$\hat{\boldsymbol{\alpha}} = \mathbf{E}^{-H} \mathbf{G} \mathbf{h} \ . \tag{16}$$

Inserting the amplitude estimate $\hat{\boldsymbol{\alpha}}$ into (12) yields

$$P = \mathbf{h}^H (\hat{\mathbf{R}} - \mathbf{G}^H \mathbf{E}^{-1} \mathbf{G}) \mathbf{h} \tag{17}$$

$$= \mathbf{h}^H \hat{\mathbf{Q}} \mathbf{h} \ , \tag{18}$$

where $\hat{\mathbf{Q}} = \hat{\mathbf{R}} - \mathbf{G}^H \mathbf{E}^{-1} \mathbf{G}$ can be interpreted as an estimate of the noise covariance matrix [12]. Note that asymptotically, the matrix $\mathbf{E}$ equals $\mathbf{I}$ which can be exploited to obtain a computationally simpler algorithm [8].

The optimal filter is derived from (18). However, solving directly for the unknown filter leads to the zero vector solution. We circumvent this by introducing some additional constraints. The constraints are formulated such that the filter has a unit gain at all of the harmonic frequencies and DOA pairs of the SOI. This leaves us with the following constrained optimization problem

$$\min_{\mathbf{h}} \mathbf{h}^H \hat{\mathbf{Q}} \mathbf{h} \quad \text{s.t.} \quad \mathbf{h}^H \mathbf{z}_{l\omega_t, l\omega_s} = 1 \ , \quad \text{for} \quad l = 1, \ldots, L \ , \tag{19}$$

where

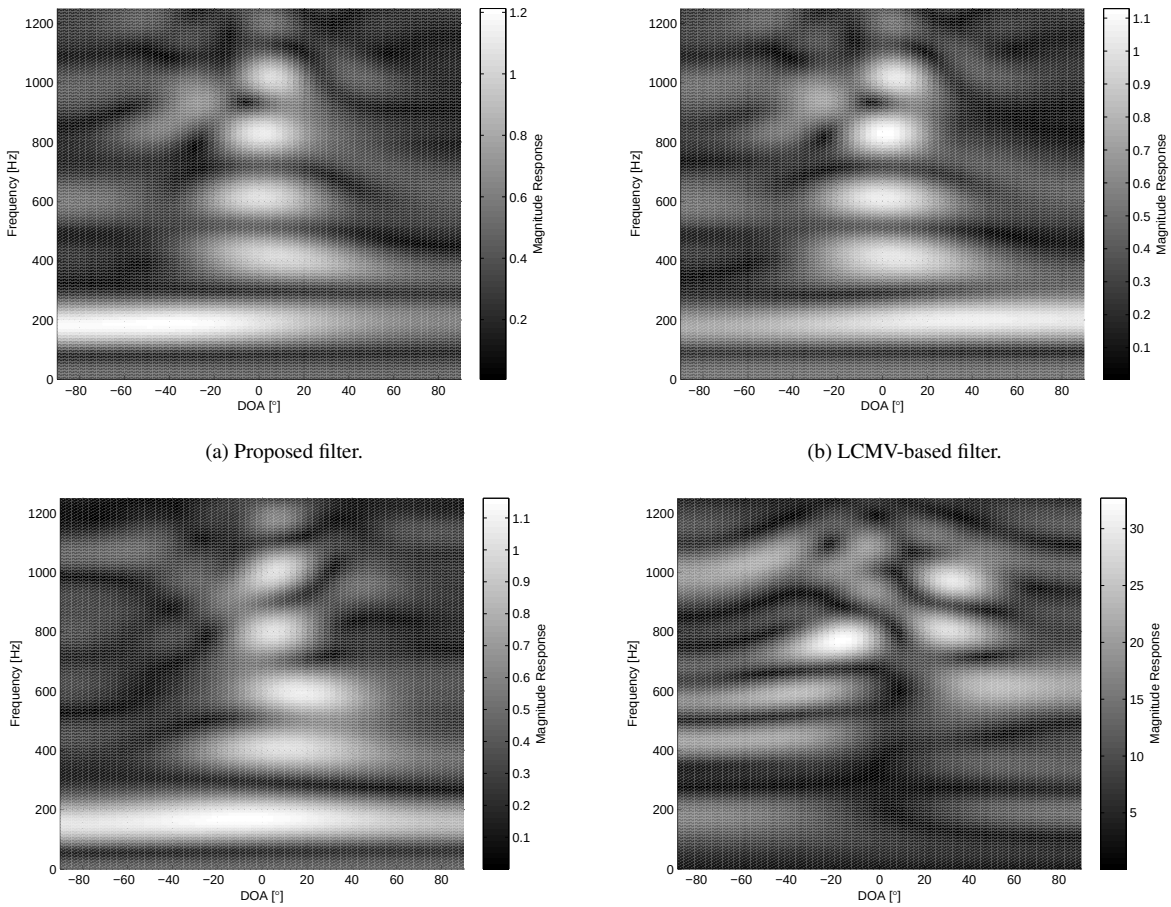
$$\mathbf{z}_{l\omega_t, l\omega_s} = \mathbf{z}_{l\omega_t} \otimes \mathbf{z}_{l\omega_s} \tag{20}$$

$$\mathbf{z}_{l\omega_t} = \begin{bmatrix} 1 & e^{-jl\omega_t} & \cdots & e^{-jl\omega_t(M_t-1)} \end{bmatrix}^T \tag{21}$$

$$\mathbf{z}_{l\omega_s} = \begin{bmatrix} 1 & e^{-jl\omega_s} & \cdots & e^{-jl\omega_s(N_s-1)} \end{bmatrix}^T \ , \tag{22}$$

with $\otimes$ denoting the Kronecker product operator. Note that all constraints can be written as a singel matrix-vector product as

$$\mathbf{h}^H \mathbf{Z}_{\omega_t, \omega_s} = \mathbf{1} \ , \tag{23}$$



(a) Proposed filter.

(b) LCMV-based filter.

(c) Proposed filter.

(d) LCMV-based filter

Fig. 1: Frequency responses of the filters at SNRs of (a),(b) -20 dB and (c),(d) 20 dB, respectively.

where

$$\mathbf{Z}_{\omega_t,\omega_s} = \begin{bmatrix}\mathbf{z}_{\omega_t,\omega_s} & \cdots & \mathbf{z}_{L\omega_t,L\omega_s}\end{bmatrix}. \tag{24}$$

We recognize that the problem in (19) is a quadratic optimization problem which is solvable using the Lagrange multiplier method. If we introduce the Lagrange multiplier vector $\boldsymbol{\lambda} = \begin{bmatrix}\lambda_1 & \cdots & \lambda_L\end{bmatrix}$, the Lagrangian dual function is given by

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

By differentiating the Lagrange dual function with respect to the unknown Lagrange multiplier $\boldsymbol{\lambda}$ and the unknown filter $\mathbf{h}$, by equating with $\mathbf{0}$, and by inserting the so-obtained expressions into each other, we get that the optimal filter $\hat{\mathbf{h}}$ is given by

$$\hat{\mathbf{h}} = \hat{\mathbf{Q}}^{-1}\mathbf{Z}_{\omega_t,\omega_s}\left(\mathbf{Z}^H_{\omega_t,\omega_s}\hat{\mathbf{Q}}^{-1}\mathbf{Z}_{\omega_t,\omega_s}\right)^{-1}\mathbf{1}. \tag{26}$$

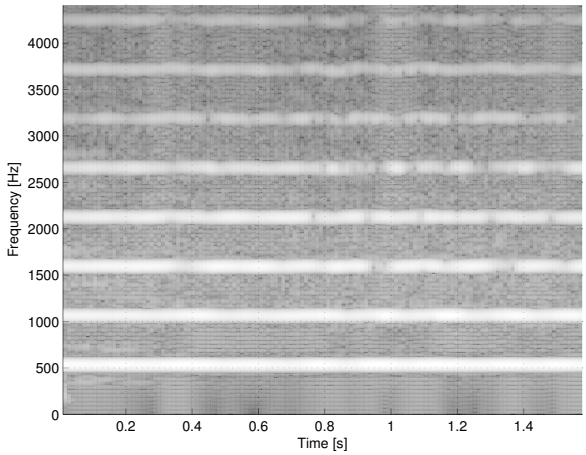
Note that the optimality criterion for the filter design is twofold: 1) the filter gain should be one at all harmonic frequencies and DOA pairs while the filter minimizes all other frequency/DOA components, and 2) the filter output should resemble a sum of sinusoids as much as possible under the given constraints. If we insert the optimal filter response in (26) into (16), we can obtain estimates of the amplitudes of the harmonics

$$\hat{\boldsymbol{\alpha}} = \mathbf{E}^{-1}\mathbf{G}\hat{\mathbf{Q}}^{-1}\mathbf{Z}_{\omega_t,\omega_s}(\mathbf{Z}^H_{\omega_t,\omega_s}\hat{\mathbf{Q}}^{-1}\mathbf{Z}_{\omega_t,\omega_s})^{-1}\mathbf{1}. \tag{27}$$

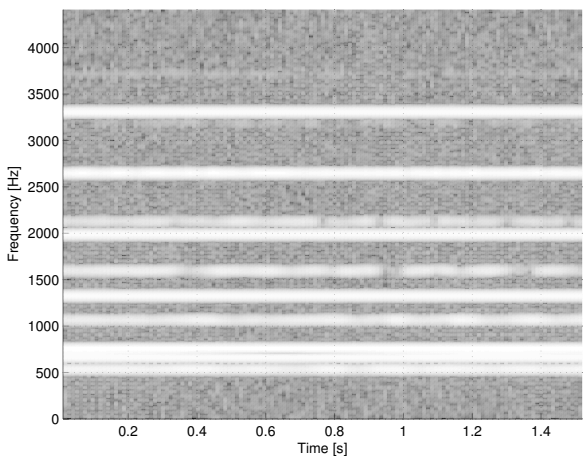
Introductory, we asummed that the fundamental frequency ω_0 was known. If this is not the case we could either estimate it using another method or using the just proposed optimal filter. To estimate it using the proposed filter, the optimal filter is applied on the input signal and the output power is then estimated. This procedure is repeated for a two-dimensional grid of candidate fundamental frequencies and DOAs. The fundamental frequency estimated is obtained by taking the argument of the maximizing fundamental frequency and DOA pair as

$$\{\hat{\omega}_t,\hat{\theta}\} = \arg\max_{(\omega_t,\theta)\in\Omega_t\times\Theta}\hat{\mathbf{h}}^H\hat{\mathbf{R}}\hat{\mathbf{h}}, \tag{28}$$

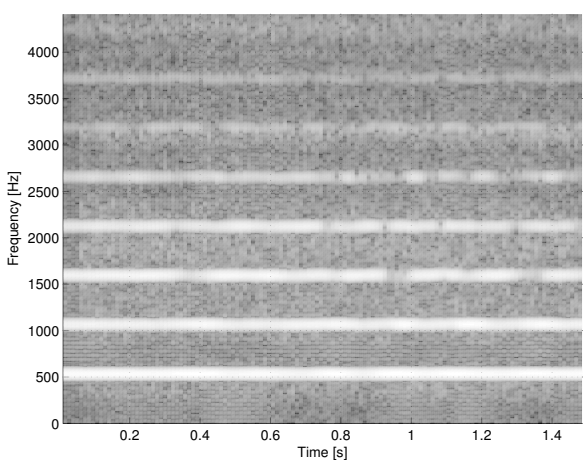
with Ω_t and Θ being sets of candidate fundamental frequencies and DOAs, respectively. Likewise, the optimal filtering method can be used for model order L estimation according to [5].



(a)



(b)



(c)

Fig. 2: Spectrograms of (a) a trumpet signal, (b) a trumpet signal in noise consisting of interfering periodic sources and complex Gaussian noise at a 30 dB SNR and (c) a signal extracted using the optimal filter.

4. EXPERIMENTAL RESULTS

Following, we describe the experimental evaluation of the proposed filter design. As mentioned previously, filtering methods based on the minimum variance principle suffers from a bad performance regarding signal extraction. This is well known, and the main reason is their unfortunate behavior at high SNRs. Several books and papers (e.g., [13, 14]) have dealt with this issue and a common fix is to, for instance, use diagonal loading techniques. The erratic high SNR behavior is also apparent from our first experiment. In this experiment, we investigate the frequency response of the proposed filter and the LCMV-based filter proposed in [10]. The filters were designed to extract a multi-channel periodic signal with $N_t = 250$, $f_s = 2,500$ Hz, $f_t = 200$ Hz, $L = 5$, $\theta = 6^\circ$, and unit amplitudes of the harmonics. Moreover, the signal was corrupted by complex Gaussian noise, the array was specified by $N_s = 6$, $c = 343.2$ m/s and $d = c/f_s$, and the filter was of orders $M_t = 20$ and $M_s = 6$. We designed the filters for SNRs of -20 dB (depicted in Fig. 1a and 1b) and 20 dB (depicted in Fig. 1c and 1d), respectively. From the plots, we observe that both filter types behave nicely at low SNRs. At high SNRs, the proposed filter still seems to perform nicely in terms of emphasizing the harmonics while the LCMV-based filter has huge side lobes.

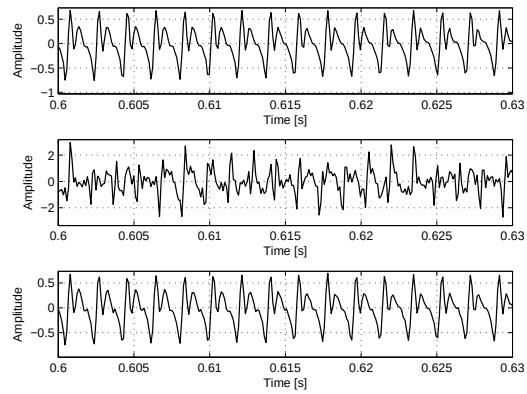


Fig. 3: Segments of (top) the trumpet signal, (middle) the observed signal with complex Gaussian noise at a SNR of 30 dB and interfering periodic sources, and (bottom) the signal extracted using the optimal filter.

In the second experiment, we applied the proposed filter for extraction of a trumpet signal. The utilized trumpet signal was originally single-channel and sampled at $f_s = 8,820$ Hz. Therefore, we resynthesized it spatially as if it was impinging on a 4-element ULA with a DOA of $\theta = 17^\circ$. We corrupted the trumpet signal with additional synthetic periodic sources and complex Gaussian noise at an SNR of 30 dB. The FIMPSIX filter was designed for 60 ms segments with a filter length of 40 and it was updated every 30 ms. For each input segment, we estimated the fundamental frequency of the trumpet signal using an MVDR-based method. The filter was designed for the estimated fundamental frequency and for a fixed model order of $L = 8$. The spectrograms of the original trumpet signal, the noisy signal, observed on the first sensor, and the extracted signal are depicted in Fig. 2(a)-(c), respectively. Furthermore, short segments of the different signals are shown in Fig. 3. It is clear from these figures that the proposed filter design are useful for extraction of periodic signals (or nearly periodic signals such as the trumpet signal).

5. CONCLUSION

In this paper, we proposed a novel optimal joint spatio-temporal filtering method for extraction of periodic signal recorded in time and space using a uniform linear microphone array. The proposed filter is based on a harmonic model which makes it suitable for all signals being (quasi-)periodic of nature such as audio and speech. Specifically, the proposed filter is inspired by the amplitude and phase estimation (APES) method. By using the APES principle rather than the minimum variance principle in the filter design, we obtain a filter with a less erratic filter response at high SNRs compared to minimum variance based filters. This is also evident from the experimental results. Due to the better filter response behaviour, the proposed filter is better suited for signal extraction. Our simulation results showed that the proposed filter is also applicable for extraction of real-life signals such as a trumpet signal. From the results, it is clear that the filter is useful for suppressing both random noise and interferering sources.

6. REFERENCES

[1] Z. Mouyan, C. Zhenming, and R. Unbehauen, "Separation of periodic signals by using an algebraic method," in *Proc. IEEE Int. Symp. Circuits and Systems*, Jun. 1991, pp. 2427–2430.

[2] A. Nehorai and B. Porat, "Adaptive comb filtering for harmonic signal enhancement," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1124–1138, Oct 1986.

[3] M. G. Christensen, J. H. Jensen, A. Jakobsson, and S. H. Jensen, "On optimal filter designs for fundamental frequency estimation," *IEEE Signal Process. Lett.*, vol. 15, pp. 745–748, 2008.

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

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

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

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

[8] M. G. Christensen and A. Jakobsson, "Optimal filters for extraction and separation of periodic sources," in *Rec. Asilomar Conf. Signals, Systems, and Computers*, 2009, Accepted.

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

[10] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Joint doa and fundamental frequency estimation methods based on 2-d filtering," in *Proc. European Signal Processing Conf.*, 2010, accepted.

[11] J. Li and P. Stoica, "An adaptive filtering approach to spectral estimation and SAR imaging," *IEEE Trans. Signal Process.*, vol. 44, no. 6, pp. 1469–1484, jun 1996.

[12] P. Stoica, H. Li, and J. Li, "Amplitude estimation of sinusoidal signals: survey, new results, and an application," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 338–352, Feb. 2000.

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

[14] L. Du, T. Yardibi, J. Li, and P. Stoica, "Review of user parameter-free robust adaptive beamforming algorithms," *Digital Signal Processing*, vol. 19, no. 4, pp. 567 – 582, 2009.