Fundamental Frequency and Model Order Estimation Using Spatial Filtering

Karimian-Azari, Sam; Jensen, Jesper Rindom; Christensen, Mads Græsbøll

Published in:
2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)

DOI (link to publication from Publisher):
10.1109/ICASSP.2014.6854748

Publication date:
2014

Document Version
Early version, also known as pre-print

AUDIO ANALYSIS LAB

Fundamental Frequency and Model Order Estimation Using Spatial Filtering

Sam Karimian-Azari, Jesper Rindom Jensen, and Mads Græsbøll Christensen
{ska, jrj, mgc}@create.aau.dk

AALBORG UNIVERSITY DENMARK

Introduction

In real scenarios, a desired signal is contaminated by different levels of noise and interferers, which complicate the estimation of the signal parameters.
In most of the state-of-the-art methods for fundamental frequency and number of harmonics estimation, the desired signal

- is assumed to be degraded by additive white Gaussian noise,
- has higher power than the interferers,
- and does not have spectral overlapping with interferers.

We present an estimation procedure for harmonic-structured signals in situations with strong interference using spatial filtering featuring:

- Joint estimates of the fundamental frequency and the constrained model order
- A procedure to account for inharmonicity using an unconstrained model

Signal Model

$$y_m(t) = \sum_{n=1}^{N}\sum_{l=1}^{L_n} \alpha_{n,l}\, e^{j(l\omega_n t+\varphi_{n,l})}\, e^{-jl\omega_n\Delta\tau_{m,n}} + v_m(t) \qquad (1)$$

(for $m = 1, \ldots, M$ microphones),

where ω_n and L_n are fundamental frequency and order of the n^{th} signal source, i.e., $x_n(t) = \sum_{l=1}^{L_n} \alpha_{n,l}\, e^{j(l\omega_n t+\varphi_{n,l})}$, the time difference of arrival between the m^{th} and a reference microphone is $\Delta\tau_{m,n}$, depending on direction of arrivals θ_n, and v_m is Gaussian noise.
Using a frequency-domain vector notation, the received signals are

$$\mathbf{Y}(\omega) = \sum_{n=1}^{N} \mathbf{d}(\theta_n,\omega)X_n(\omega) + \mathbf{V}(\omega), \qquad (2)$$

where $\mathbf{Y}(\omega) = [Y_1(\omega)\ Y_2(\omega) \cdots\ Y_M(\omega)]^T$, and $\mathbf{d}(\theta_n,\omega) = [1\ \ e^{-j\omega\Delta\tau_{2,n}}\ \cdots\ e^{-j\omega\Delta\tau_{M,n}}]^T$, for $\omega \in [0,\pi]$ and $\theta_n \in [0,\pi]$.

Spatial Filtering

A complex-valued spatial filter $\mathbf{H}(\theta,\omega)$ is applied on the microphone outputs subject to $\mathbf{H}^H(\theta,\omega)\mathbf{d}(\theta,\omega) = 1$ like

$$Z(\theta,\omega) = \mathbf{H}^H(\theta,\omega)\mathbf{Y}(\omega). \qquad (3)$$

Assuming uncorrelated signal sources and noise, the output power corresponding to the direction of the desired signal, i.e., $\theta = \theta_1$, is

$$J_Z(\theta_1,\omega) = \mathrm{E}\{Z(\theta_1,\omega)Z^H(\theta_1,\omega)\} = J_{X_1}(\omega) + \Psi(\theta_1,\omega), \qquad (4)$$

where $J_{X_n}(\omega) = \mathrm{E}\{|X_n(\omega)|^2\}$, and $\Psi(\theta_1,\omega) = \mathbf{H}^H(\theta_1,\omega)\mathbf{R}_{\mathbf{V}}(\omega)\mathbf{H}(\theta_1,\omega) + \sum_{n=2}^{N}\mathbf{H}^H(\theta_1,\omega)\mathbf{d}(\theta_n,\omega)J_{X_n}(\omega)\mathbf{d}^H(\theta_n,\omega)\mathbf{H}(\theta_1,\omega)$.

Proposed Method

The broadband power of the output signal and the output noise-plus-interference are, respectively,

$$J_Z(\theta_1) = \frac{1}{2\pi}\int_0^{2\pi} J_Z(\theta_1,\omega)\mathrm{d}\omega, \qquad (5)$$

$$\Psi(\theta_1) = \frac{1}{2\pi}\int_0^{2\pi} \Psi(\theta_1,\omega)\mathrm{d}\omega = J_Z(\theta_1) - J_{X_1}. \qquad (6)$$

- With the constrained (C) harmonic-model

$$\mathbb{X}_n^{\mathrm{C}}(\omega_n) = [\ X_n(\omega_n)\ X_n(2\omega_n) \ldots X_n(L_n^{\mathrm{C}}\omega_n)\ ]^T, \qquad (7)$$

$$\Psi^{\mathrm{C}}(\theta_1) = J_Z(\theta_1) - J_{X_1}^{\mathrm{C}}(\omega_1) = J_Z(\theta_1) - 2\,\|\mathbb{X}_1^{\mathrm{C}}(\omega_1)\|_2^2. \qquad (8)$$

With the assumption of white Gaussian noise and using N_f frequency samples, we can jointly estimate the fundamental frequency and the number of harmonics using maximum a posteriori (MAP) [1] like

$$(\hat{L}_1^{\mathrm{C}}, \hat{\omega}_1^{\mathrm{C}}) \approx \arg\min_{L_1^{\mathrm{C}},\omega_1} N_f \ln[\Psi^{\mathrm{C}}(\theta_1)] + \frac{3}{2}\ln N_f + L_1^{\mathrm{C}}\ln N_f, \qquad (9)$$

- With the unconstrained (UC) model

$$\mathbb{X}_n^{\mathrm{UC}}(\boldsymbol{\Omega}_n) = [\ X_n(\omega_{n,1})\ X_n(\omega_{n,2}) \ldots X_n(\omega_{n,L_n^{\mathrm{UC}}})\ ]^T, \qquad (10)$$

where $\boldsymbol{\Omega}_n = [\ \omega_{n,1}\ \omega_{n,2} \ldots \omega_{n,L_n^{\mathrm{UC}}}\ ]^T$, we have

$$\Psi^{\mathrm{UC}}(\theta_1) = J_Z(\theta_1) - J_{X_1}^{\mathrm{UC}}(\boldsymbol{\Omega}_1) = J_Z(\theta_1) - 2\,\|\mathbb{X}_1^{\mathrm{UC}}(\boldsymbol{\Omega}_1)\|_2^2. \qquad (11)$$

We can extend the MAP model order estimation method for estimating the number of independent sinusoids like

$$(\hat{L}_1^{\mathrm{UC}}, \hat{\boldsymbol{\Omega}}_1) \approx \arg\min_{L_1^{\mathrm{UC}},\boldsymbol{\Omega}_1} N_f \ln[\Psi^{\mathrm{UC}}(\theta_1)] + \frac{5}{2}L_1^{\mathrm{UC}}\ln N_f, \qquad (12)$$

and apply the Markov-like weighted least-squares (WLS) method [2] to estimate the fundamental frequency.

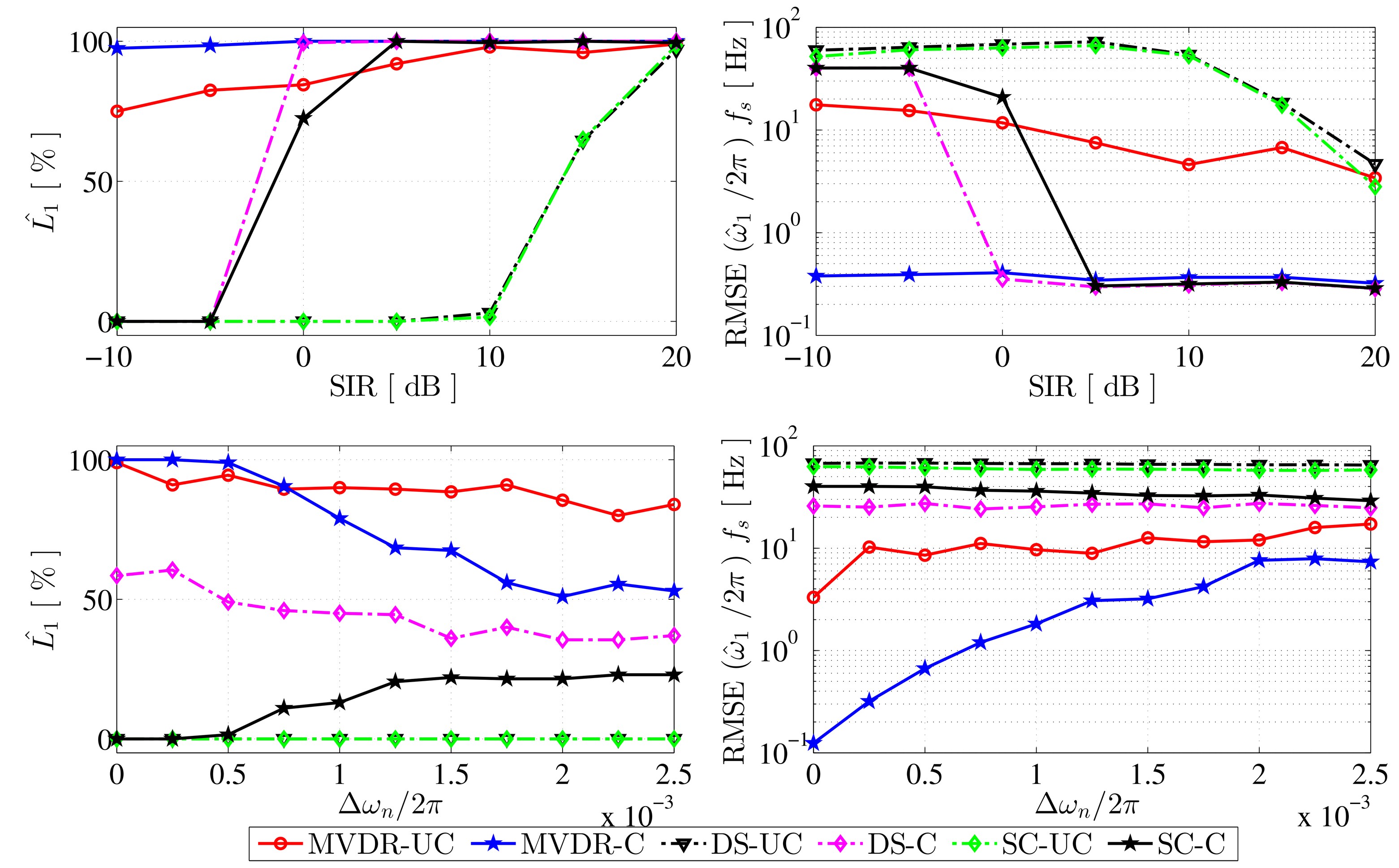
Conclusion

- In situations with spatially separated interference signals with low SIRs, the joint fundamental frequency and model order estimation can be facilitated using spatial filters.
- Simulations indicate that the UC model order estimates are more accurate than the C model. However, the fundamental frequency estimates via the C model are more accurate than the UC based estimator.

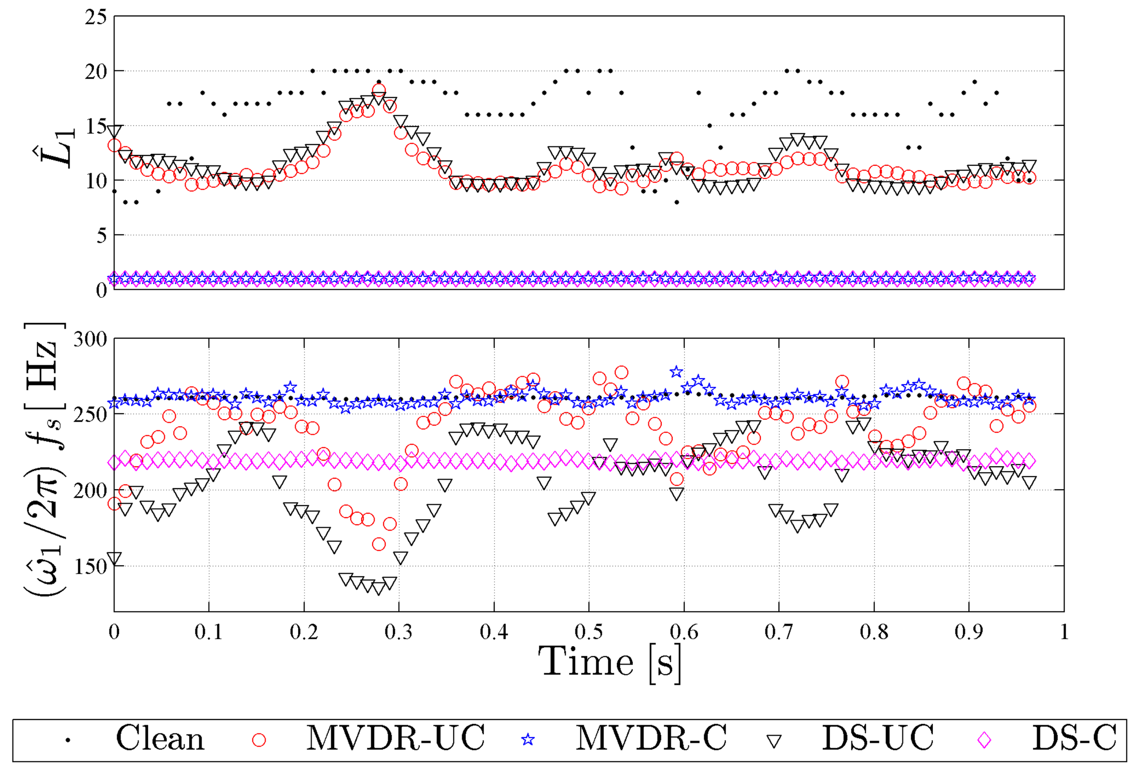
Experimental Results

We compared the results of single-channel (SC) parameter estimators with the proposed method, using the delay-and-sum (DS) and the minimum variance distortionless response (MVDR) beamformers.

- Two synthetic signals: $\theta_1 = 60°$, $\omega_1 = 0.0450\pi$, $L_1 = 5$, and $\theta_2 = 40°$, $\omega_2 = 0.0550\pi$, $L_2 = 7$, with unit amplitudes. Harmonic frequencies were perturbed by a normal distribution ($\Delta\omega_{n,l} = 0.0005\pi$), and the received signals were distorted by white Gaussian noise (20 dB SNR).



- A real trumpet signal with vibrato (SIR = −1.5 dB and SNR = 10 dB), and estimates of order and pitch.



References

[1] P. M. Djuric. Asymptotic MAP criteria for model selection. *IEEE Trans. Signal Process.*, 46(10):2726 –2735, Oct. 1998.

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

Thanks to the Villum Foundation, and the Danish Council for Independent Research grant DFF-1337-00084