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

Link to publication from Aalborg University

Citation for published version (APA):
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

\[ y_m(t) = \sum_{n=1}^{N} \sum_{m=1}^{L_n} \alpha_{mn} e^{j\omega_m t + \varphi_{mn}} e^{-j\omega_m \Delta \tau_m} + v_m(t) \] (1)

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

where \( \omega_m \) and \( L_n \) are fundamental frequency and order of the \( n \)th signal source, i.e., \( x_m(t) = \sum_{m=1}^{L_n} \alpha_{mn} e^{j\omega_m t + \varphi_{mn}} \), the time difference of arrival between the \( m \)th and a reference microphone is \( \Delta \tau_m \), depending on direction of arrivals \( \theta_m \), and \( v_m \) is Gaussian noise.

Using a frequency-domain vector notation, the received signals are

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

where \( Y(\omega) = [Y_1(\omega) Y_2(\omega) \cdots Y_M(\omega)]^T \), and \( d(\theta_n, \omega) = [1 e^{-j\omega \Delta \tau_1} \cdots e^{-j\omega \Delta \tau_L}]^T \), for \( \omega \in [0, \pi] \) and \( \theta_n \in [0, \pi] \).

Spatial Filtering

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

\[ Z(\omega, \theta) = H^H(\theta, \omega) Y(\omega). \] (3)

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

\[ J_Z(\theta_1, \omega) = E\{Z(\theta_1, \omega)Z^H(\theta_1, \omega)\} = J_X(\omega) + \psi(\theta_1, \omega), \] (4)

where \( J_X(\omega) = E\{|X_1(\omega)|^2\} \), and \( \psi(\theta_1, \omega) = H^H(\theta_1, \omega) R_V(\omega) H(\theta_1, \omega) + \sum_{n=2}^{N} H^H(\theta_1, \omega) \mathbf{d}(\theta_n, \omega) J_X(\omega) H(\theta_1, \omega) \).

Proposed Method

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

\[ J_Z(\theta_1, \omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} J_Z(\theta_1, \omega) d\omega, \]

\[ \psi(\theta_1, \omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \psi(\theta_1, \omega) d\omega = J_Z(\theta_1) - J_X(\theta_1). \] (5)

With the constrained (C) harmonic-model

\[ X_C^N(\omega_1) = [X_C(\omega_1) X_C(2\omega_1) \cdots X_C(L^C_{\omega_1})]^T, \] (6)

\[ \psi_C(\theta_1) = J_Z(\theta_1) - J_X^C(\omega_1) = J_Z(\theta_1) - 2 \|X_C^N(\omega_1)\|_2^2. \] (7)

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

\[ (\hat{L}_1, \hat{\omega}_1^C) \approx \arg \min_{L_1, \omega_1^C} N \ln[p_C(\omega_1)] + \frac{3}{2} \ln N + \frac{1}{4} \hat{L}_1^2 \ln N, \] (8)

With the unconstrained (UC) model

\[ X_{UC}^N(\Omega_n) = [X_n(\omega_{n1}) X_n(\omega_{n2}) \cdots X_n(\omega_{nL^U_n})]^T, \] (9)

where \( \Omega_n = [\omega_{n1} \omega_{n2} \cdots \omega_{nL^U_n}]^T \), we have

\[ \psi_{UC}(\theta_1) = J_Z(\theta_1) - J_{UC}^n(\Omega_n) = J_Z(\theta_1) - 2 \|X_{UC}^N(\Omega_n)\|_2^2. \] (10)

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

\[ (\hat{L}_{UC}, \hat{\Omega}_n) \approx \arg \min_{L_{UC}, \Omega_n} N \ln[p_{UC}(\omega_1)] + \frac{5}{4} \hat{L}_{UC}^2 \ln N, \] (11)

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: \( \theta_1 = 60^\circ, \omega_1 = 0.0450 \pi, L_1 = 5 \), and \( \theta_2 = 40^\circ, \omega_2 = 0.0550 \pi, L_2 = 7 \), with unit amplitudes. Harmonic frequencies were perturbed by a normal distribution (\( \Delta \omega_{n1} = 0.0005\pi \)), 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


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