Robust Pitch Estimation Using an Optimal Filter on Frequency Estimates

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

Published in:
2014 Proceedings of the 22nd European Signal Processing Conference (EUSIPCO 2014)

Publication date:
2014

Document Version
Early version, also known as pre-print

Robust Pitch Estimation Using an Optimal Filter on Frequency Estimates

Sep. 4, 2014

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

Audio Analysis Lab, AD:MT,
Aalborg University, Denmark

Funded by the Villum foundation and DFF–1337-00084

AALBORG UNIVERSITY
DENMARK

Agenda

Introduction
- Harmonic Signal Model
- Pitch Estimation

The WLS Pitch Estimator

Proposed Method
- Problem Formulation
- Filtering UFEs
- Cramér-Rao Lower Bound

Experimental Results

Conclusion

Introduction

In speech and audio signal processing applications such as:

- hearing-aids
- teleconference systems

we can design filters based on parametric models.

Pitch estimate is often necessary in some applications such as:

- Coding
- Enhancement
- Separation
- Compression

Harmonic Signal Model:

$$s(n, \boldsymbol{\theta}) = \sum_{l=1}^{L} \alpha_l \, e^{j(\omega_l n + \varphi_l)}, \qquad (1)$$

$$\boldsymbol{\theta} = [\,\alpha_1,\, \varphi_1,\, \omega_1,\, \ldots, \alpha_L,\, \varphi_L,\, \omega_L\,]^T$$

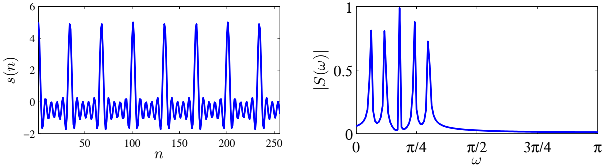
where $\omega_l = l\omega_0$ (for $l = 1, \ldots, L$),

L : number of sinusoids

α_l : real magnitudes

ω_0 : fundamental frequency

φ_l : phases of harmonics



Existing approaches*:

- **Filtering Methods:** Based on filtering the observed signal that pass power undistorted frequency of harmonics, e.g., optimal adaptive designs.
- **Subspace Methods:** Based on the principles of subspace orthogonality, e.g., MUSIC method
- **Statistical Methods:** Based on maximizing the likelihood function of pitch, e.g., the non-linear least-squares (NLS), and weighted least squares (WLS) [H. Li, P. Stoica, and J. Li 2000].

* [M.G.Christensen and A.Jakobsson 2009]

The Weighted Least Squares (WLS) Parameter Estimator
[H. Li, P. Stoica, and J. Li 2000]

The observed signal is contaminated by Gaussian noise $v(n)$ with variance of σ^2

$$x(n) = \sum_{l=1}^{L} \alpha_l \, e^{j\,(\omega_l n + \varphi_l)} + v(n). \tag{2}$$

With the assumption of white Gaussian noise, the WLS pitch estimator yields

$$\hat{\omega}_0 = \frac{1}{\sum_{l=1}^{L} (l\,\hat{\alpha}_l)^2} \sum_{l=1}^{L} l\,\hat{\alpha}_l^2\,\hat{\omega}_l, \tag{3}$$

where $\hat{\omega}_l$ and $\hat{\alpha}_l$ can be estimated using different methods, e.g., the DFT, MUSIC, ESPRIT, NLS, Capon, and APES.

The WLS Pitch Estimator
Drawbacks

The WLS pitch estimator is **computationally efficient** with good **statistical performance**,

BUT! It is not optimal for cases:

- nonidentical noise variances among harmonics, e.g., colored noise.
- spurious frequency estimates.
- missing harmonics.

Proposed Method
Assumption

At a high narrowband SNR, i.e.,

$$\mathrm{SNR}(\omega_l) = \frac{\alpha_l^2}{\sigma^2} \gg 1, \tag{4}$$

the harmonic frequency ω_l is perturbed with a real angular noise [S. Tretter 1985]

$$\Delta\omega_l(n) = \frac{v(n)}{\alpha_l} \sin(l\omega_0 n + \varphi_l), \tag{5}$$

which has a normal distribution with zero mean and variance

$$\mathrm{E}\{(\Delta\omega_l)^2\} = \frac{1}{2\,\mathrm{SNR}(\omega_l)}, \tag{6}$$

where the white Gaussian noise has a homogeneous power spectrum $\Phi(\omega) = \sigma^2$ across frequencies $\omega \in [0, \pi]$.
In general,

$$\mathrm{E}\{(\Delta\omega_l)^2\} = \frac{\Phi(\omega_l)}{2\,\alpha_l^2}. \tag{7}$$

We can approximate $x(n) = \sum_{l=1}^{L} \alpha_l \, e^{j\,(\omega_l n + \varphi_l)} + v(n)$ like,

$$x(n) \approx \sum_{l=1}^{L} \alpha_l \, e^{j\,(\omega_l n + \Delta\omega_l(n) + \varphi_l)}. \tag{8}$$

Harmonic frequencies:

$$\mathbf{\Omega} = [\, \omega_1, \; \omega_2, \; \ldots, \omega_L \,]^T = \mathbf{d}_L \, \omega_0, \tag{9}$$

where $\mathbf{d}_L = [\, 1, 2, \; \cdots \; , L \,]^T$.

Unconstrained frequency estimates (UFEs):

$$\begin{aligned} \hat{\mathbf{\Omega}} &= [\, \hat{\omega}_1, \; \hat{\omega}_2, \; \ldots, \; \hat{\omega}_L \,]^T \\ &= \mathbf{\Omega} + \Delta\mathbf{\Omega} = \mathbf{d}_L \, \omega_0 + \Delta\mathbf{\Omega}, \end{aligned} \tag{10}$$

where $\Delta\mathbf{\Omega} = [\, \Delta\omega_1, \; \Delta\omega_2, \; \ldots, \; \Delta\omega_L \,]^T$.

Proposed Method
Filtering UFEs

Apply a filter $\mathbf{h} \in \mathbb{R}^L$

$$\hat{\omega}_0 = \mathbf{h}^T \hat{\boldsymbol{\Omega}} = \mathbf{h}^T \mathbf{d}_L \, \omega_0 + \mathbf{h}^T \Delta\boldsymbol{\Omega}. \tag{11}$$

With the distortionless constraint that $\mathbf{h}^T \mathbf{d}_L = 1$, the mean squared error (MSE) of the unbiased estimator is given by

$$\begin{aligned} \mathrm{MSE}[\,\hat{\omega}_0\,] &= \mathrm{E}\{(\hat{\omega}_0 - \omega_0)^2\} \\ &= \mathrm{E}\{(\mathbf{h}^T \Delta\boldsymbol{\Omega})(\Delta\boldsymbol{\Omega}^T \mathbf{h})\} = \mathbf{h}^T \boldsymbol{\Phi}_{\Delta\Omega} \mathbf{h}. \end{aligned} \tag{12}$$

Independent UFEs would be implicitly uncorrelated, i.e., $\mathrm{E}\{\Delta\omega_i \Delta\omega_k\} = 0$ for $i \neq k$, when harmonics are well-separated and the narrowband SNRs are high enough;

$$\boldsymbol{\Phi}_{\Delta\boldsymbol{\Omega}} = \mathrm{E}\{\Delta\boldsymbol{\Omega}\,\Delta\boldsymbol{\Omega}^T\}$$
$$= \mathrm{diag}\left\{\left[\,\frac{\Phi(\omega_1)}{2\,\alpha_1^2},\ \frac{\Phi(\omega_2)}{2\,\alpha_2^2},\ \ldots,\ \frac{\Phi(\omega_L)}{2\,\alpha_L^2}\,\right]\right\}. \quad (13)$$

In practice:

$$\boldsymbol{\Phi}_{\Delta\boldsymbol{\Omega}} = \mathrm{E}\{(\hat{\boldsymbol{\Omega}} - \mathrm{E}\{\hat{\boldsymbol{\Omega}}\})(\hat{\boldsymbol{\Omega}} - \mathrm{E}\{\hat{\boldsymbol{\Omega}}\})^T\}, \quad (14)$$

where $\mathrm{E}\{\cdot\}$ is the mathematical expectation, and

$$\mathrm{E}\{\hat{\boldsymbol{\Omega}}(n)\} = \frac{1}{M}\sum_{m=0}^{M-1} \hat{\boldsymbol{\Omega}}(n-m). \quad (15)$$

Proposed Method
Filtering UFEs

Optimal filter is given as the solution to

$$\min_{\mathbf{h}} \quad \mathbf{h}^T \mathbf{\Phi}_{\Delta\Omega} \mathbf{h}$$
$$\text{subject to} \quad \mathbf{h}^T \mathbf{d}_L = 1$$

Minimum variance distortionless response (MVDR):

$$\mathbf{h}_{\text{MVDR}} = \mathbf{\Phi}_{\Delta\Omega}^{-1} \mathbf{d}_L (\mathbf{d}_L^T \mathbf{\Phi}_{\Delta\Omega}^{-1} \mathbf{d}_L)^{-1}. \tag{16}$$

$$\Downarrow$$

Maximum likelihood (ML):
Assuming white Gaussian noise with the variance σ^2

$$\mathbf{\Phi}_{\Delta\Omega}^{-1} = \frac{2}{\sigma^2} \operatorname{diag}\left\{ [\,\alpha_1^2,\, \alpha_2^2,\, \ldots,\, \alpha_L^2\,] \right\}, \tag{17}$$

$$\mathbf{h}_{\text{ML}} = \frac{1}{\sum_{l=1}^{L} (l\alpha_l)^2} [\,\alpha_1^2,\, 2\alpha_2^2,\, \ldots,\, L\alpha_L^2\,]^T. \tag{18}$$

Asymptotic phase estimates of harmonics at $n \in [0, N-1]$:

$$\hat{\boldsymbol{\Psi}}(n) = \mathbf{d}_L \omega_0 n + \boldsymbol{\Xi}_L + \Delta\boldsymbol{\Omega}(n) \tag{19}$$

Likelihood Function:

$$P(\hat{\boldsymbol{\Psi}}, \xi) \sim \mathcal{N}(\mathbf{d}_L \omega_0 + \boldsymbol{\Xi}_L, \boldsymbol{\Phi}_{\Delta\Omega}) \tag{20}$$

where $\xi = [\omega_o, \boldsymbol{\Xi}_L]^T$ and $\boldsymbol{\Xi}_L = [\varphi_1, \varphi_2, \ldots, \varphi_L]^T$.

CRLB of pitch estimate:

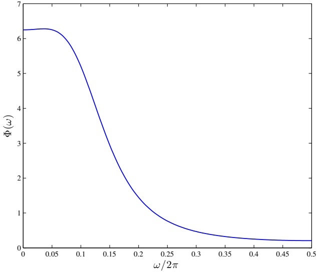
$$[\mathbf{I}(\xi)^{-1}]_{1,1} = \frac{12}{N(N^2-1)} (\mathbf{d}_L^T \boldsymbol{\Phi}_{\Delta\Omega}^{-1} \mathbf{d}_L)^{-1}, \tag{21}$$

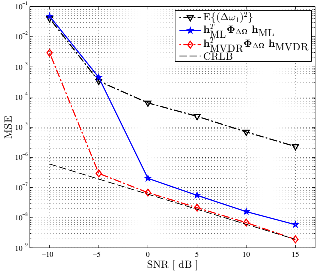
where $\mathbf{I}(\xi) = -\mathrm{E}\{\frac{\partial^2 \ln p(\boldsymbol{\Psi}, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}\, \partial \boldsymbol{\xi}^T}\}$ is the Fisher information matrix.

Experimental Results
Experiment 1 (synthetic signal)

A synthetic harmonic signal $\omega_0 = 0.15\pi$, $L = 5$ sinusoids with identical amplitudes. We estimated UFEs from $N = 60$ samples using the subspace orthogonality method based on the MUSIC method.

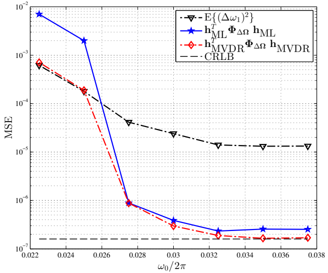
- Additive colored Gaussian noise, different SNR:





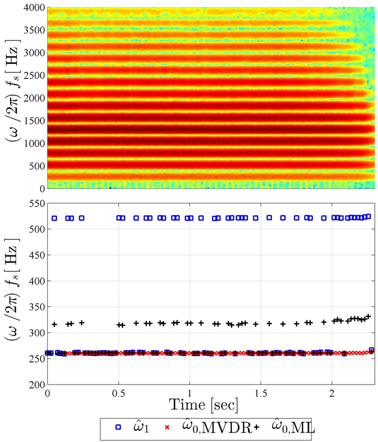
Experimental Results
Experiment 2 (synthetic signal)

- White Gaussian noise in SNR$=$ 5 dB, different ω_0:



Experimental Results
Experiment 3 (real signal)

- A trumpet signal contaminated by colored Gaussian noise, $f_s = 8.0$ kHz, $N = 128$
- We estimated L using the subspace orthogonality method [M.G.Christensen et al. 2007]



Summery and Conclusion

- We estimated pitch from unconstrained frequency estimates (UFEs).
- We designed an optimal filter based on noise statistics which is estimated from UFEs.
- The optimal filter design is a general method that we can derive the WLS pitch estimator.
- Pitch estimation using the optimal filter is robust against different noise situations, e.g., colored noise.
- Pitch estimation using the optimal filter is robust against missing harmonics in some time-frames and outperforms the WLS pitch estimator.

Thank you!

AALBORG UNIVERSITY
DENMARK