Pitch Estimation and Tracking with Harmonic Emphasis on the Acoustic Spectrum

April 23, 2015

Sam Karimian-Azari\textsuperscript{1}, Nasser Mohammadiha\textsuperscript{2}, Jesper R. Jensen\textsuperscript{1}, and Mads G. Christensen\textsuperscript{1}

ska@create.aau.dk

\textsuperscript{1}Audio Analysis Lab, AD:MT, Aalborg University
\textsuperscript{2}University of Oldenburg
Agenda

- Introduction
  - Noisy Harmonic Signal Approximation
  - ML Pitch Estimate from UFE
- Bayesian Methods
  - Motivation
  - HMM
  - Kalman Filter
- Numerical Results
- Conclusion
Introduction
Harmonic Signal Model
Introduction
Harmonic Signal Model

Pitch Estimation and Tracking
Sam Karimian-Azari et al.

Introduction
Harmonic Signal Model
Noisy Signal Approx.
ML Pitch Estimation

Proposed Method
Motivation-Hypothesis
1- Discrete state-space: HMM
2- Continuous state-space: Kalman Filter

Numerical Results
Conclusion

Audio Analysis Lab, AD:MT,
Aalborg University, Denmark
Pitch Estimation and Tracking
Sam Karimian-Azari et al.

Introduction
Harmonic Signal Model

Noisy Signal Approx.
ML Pitch Estimation

Proposed Method
Motivation-Hypothesis
1- Discrete state-space: HMM
2- Continuous state-space: Kalman Filter

Numerical Results
Conclusion
Harmonic Signal Model:

\[
s(n) = \sum_{l=1}^{L(n)} \alpha_l e^{i(\omega_l(n)n + \varphi_l)},
\]

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

- \(L(n)\) : number of sinusoids
- \(\alpha_l\) : real magnitudes
- \(\omega_0\) : fundamental frequency
- \(\varphi_l\) : phases of harmonics
The observed signal can be written as a sum of a desired signal $s(n)$ and a noise signal $v(n)$, i.e.,

$$x(n) = s(n) + v(n)$$

$$= \sum_{l=1}^{L} \alpha_l e^{j(\omega_l n + \phi_l)} + v(n).$$
At a high narrowband SNR, the harmonic frequency $\omega_l$ is perturbed with a real-valued phase-noise [S.Tretter 1985], which has a normal distribution with zero mean and the variance

$$E\{\Delta\omega_l^2(n)\} = \frac{\sigma^2}{2\alpha_l^2}$$  \hspace{1cm} (3)

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

$$x(n) \approx \sum_{l=1}^{L} \alpha_l e^{j(\omega_l n + \Delta\omega_l(n) + \varphi_l)}$$  \hspace{1cm} (4)
Unconstrained frequency estimates (UFE) of the constrained frequencies:

\[ \hat{\Omega}(n) = \left[ \hat{\omega}_1(n), \hat{\omega}_2(n), \ldots, \hat{\omega}_L(n) \right]^T \]  
\[ = d_L(n) \omega_0(n) + \Delta \Omega(n), \]  
where

\[ d_L(n) = \left[ 1, 2, \ldots, L(n) \right]^T \]  
\[ \Delta \Omega(n), = \left[ \Delta \omega_1(n), \Delta \omega_2(n), \ldots, \Delta \omega_L(n) \right]^T, \]

and

\[ R_{\Delta \Omega}(n) = \mathbb{E}\{\Delta \Omega(n) \Delta \Omega^T(n)\} \]
\[ = \frac{\sigma^2}{2} \text{diag}\left\{ \frac{1}{\alpha_1^2}, \frac{1}{\alpha_2^2}, \ldots, \frac{1}{\alpha_L^2} \right\}. \]
Max. Likelihood (ML) Pitch Estimator

For the time-frame $x(n) = \left[ x(n), x(n-1), \ldots, x(n-M-1) \right]^T$, the PDF of the UFE is

$$P(\hat{\Omega}(n) | \omega_0(n)) \sim \mathcal{N}(d_L(n) \omega_0(n), R_{\Delta\Omega}(n)). \quad (10)$$

The ML pitch estimator:

$$\hat{\omega}_0(n) = \arg \max_{\omega_0(n)} \log P(\hat{\Omega}(n) | \omega_0(n)) \quad (11)$$

$$= \left[ d_L^T(n) R_{\Delta\Omega}^{-1}(n) d_L(n) \right]^{-1} d_L^T(n) R_{\Delta\Omega}^{-1}(n) \hat{\Omega}(n) \quad (12)$$
The ML Estimators are statistically efficient, e.g., the non-linear least-squares (NLS), and the weighted least squares (WLS) [H.Li, et al. 2000], but the minimum variance is limited by the number of samples.

Consecutive pitch values are estimated independently.

Motivation:
- Consecutive pitch values are estimated independently.
Bayesian Pitch Estimator
Motivation

- Pitch values are usually correlated in a sequence, i.e.,

\[ P(\omega_0(n)|\omega_0(n-1), \omega_0(n-2), \ldots), \]  

(13)

that motivate Bayesian methods to minimize an error incorporating prior distributions.

- State-of-the-art methods mostly track pitch estimates in a sequential process without concerning noise statistics.

\[
\cdots \rightarrow \omega_0(n-2) \rightarrow \omega_0(n-1) \rightarrow \omega_0(n)
\]
Bayesian Pitch Estimator
Hypothesis

1- Jointly estimate and track pitch incorporating both the harmonic constraints and noise characteristics.
2- Estimate the state $\omega_0(n)$ through a series of noisy observations:

$$P(\omega_0(n)|\hat{\Omega}(n), \hat{\Omega}(n-1), \cdots) \quad (14)$$

3- Recursively update the prior distribution of the pitch value.
Bayesian Pitch Estimator
Discrete state-space (HMM)

\( \omega_0(n) \) : Discrete random variable (Hidden states)
\( P(\omega_0(n)|\omega_0(n-1)) \) : Transition probability in a 1st-order Markov model,
i.e., \( \sum_{\omega_0(n)} P(\omega_0(n)|\omega_0(n-1)) = 1 \)

\[ \hat{\omega}_0(n) = \arg \max_{\omega_0(n)} \log P(\omega_0(n)|\hat{\Omega}(n), \hat{\Omega}(n-1), \cdots) \]
\[ = \arg \max_{\omega_0(n)} \log P(\hat{\Omega}(n)|\omega_0(n)) + \log P(\omega_0(n)|\hat{\Omega}(n-1), \cdots). \] (15)

The priori distribution is defined recursively like
\[ P(\omega_0(n)|\hat{\Omega}(n-1), \hat{\Omega}(n-2), \cdots) = \]
\[ \sum_{\omega_0(n-1)} P(\omega_0(n)|\omega_0(n-1)) P(\omega_0(n-1)|\hat{\Omega}(n-1), \cdots), \] (16)

where \( P(\omega_0(n-1)|\hat{\Omega}(n-1), \cdots) \) is the past estimate.
Bayesian Pitch Estimator
state-space representation of the pitch continuity

Continuous state-space:

\[
\omega_0(n) = \omega_0(n-1) + \delta(n) \\
\hat{\Omega}(n) = d_L(n) \omega_0(n) + \Delta\Omega(n),
\]

where \(\delta(n) \sim \mathcal{N}(0, \sigma^2_t)\) and \(\Delta\Omega(n) \sim \mathcal{N}(0, R_{\Delta\Omega}(n))\) are the state evolution and observation noise, respectively.
Bayesian Pitch Estimator
Continuous state-space (Kalman filter)

First, a pitch estimate is predicted using the past estimates as

\[ \hat{\omega}_0(n|n-1) = \hat{\omega}_0(n-1|n-1) \]  
(17)

with the variance

\[ \sigma^2_K(n|n-1) = \sigma^2_K(n-1|n-1) + \sigma_t^2. \]  
(18)

Second, the pitch estimate is updated with the error of

\[ e(n) = \hat{\Omega}(n) - d_L(n)\hat{\omega}_0(n|n-1). \]  
(19)

Then, the predicted estimate is updated:

\[ \hat{\omega}_0(n|n) = \hat{\omega}_0(n|n-1) + h_K(n)e(n) \]  
(20)

\[ h_K(n) = \sigma^2_K(n|n-1)d_L^T(n)\left[ \Pi_L(n)\sigma^2_K(n|n-1) + R_{\Delta\Omega}(n) \right]^{-1}, \]  
(21)

where \( \Pi_L(n) = d_L(n)d_L^T(n) \), and update

\[ \sigma^2_K(n|n) = \left[ 1 - h_K(n)d_L(n) \right] \sigma^2_K(n|n-1). \]  
(22)
The ML estimator of the covariance matrix among $N$ estimates:

$$R_{\Delta \Omega}(n) = E\{\Delta \Omega(n)\Delta \Omega^T(n)\}$$

$$= \frac{1}{N} \sum_{i=n-N+1}^{n} \Delta \Omega(i)\Delta \Omega^T(i),$$  \hspace{1cm} (23)

where $\Delta \Omega(n) = \hat{\Omega}(n) - \hat{\mu}(n)$, and $\mu(n) = E\{\hat{\Omega}(n)\}$.

Exponential moving average:

$$\hat{\mu}(n) = \lambda \hat{\Omega}(n) + (1 - \lambda) \hat{\mu}(n-1)$$  \hspace{1cm} (24)

The forgetting factor $0 < \lambda < 1$ recursively updates the time-varying mean value.
A linear chirp signal \( (r = 100 \text{ Hz/s}) \) with \( L = 5 \) harmonics, random phases, and identical amplitudes during 0.1 s.

\[
M = 80, \quad \omega_0(1) = 400\pi / f_s, \quad f_s = 8.0 \text{ kHz}, \quad \sigma_t = \sqrt{2\pi r / f_s^2}, \quad \text{and for the HMM-based pitch estimator, the frequency range} \quad \omega \in [150, 280] \times (2\pi / f_s) \quad \text{was discretized into} \quad N_d = 1000 \text{ samples.}
\]
Speech signal + Car noise at SNR= 5 dB.

The MAP order estimation [Djuric 1998], $M = 240$, $\lambda = 0.9$, and $N = 150$. 
Conclusion

- For pitch estimation, we have formulated the ML estimate from the UFE.
- For pitch estimation and tracking, we have proposed HMM- and KF-based methods.
- Experimental results showed that both HMM- and KF-based methods outperform the corresponding ML pitch estimators.
- The KF-based method statistically performs better than the HMM-based method, while the it tracks pitch changes more accurate than the KF-based method.
Thank you!