Fast, Accurate, and Robust Pitch Estimation

NordicSMC Winter School 2019

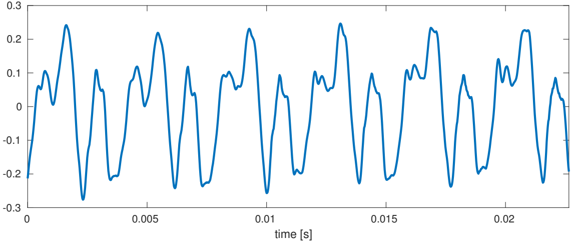
March 7, 2019

Jesper Kjær Nielsen
jkn@create.aau.dk

Audio Analysis Lab, CREATE
Aalborg University, Denmark
Website: http://audio.create.aau.dk
YouTube: http://tinyurl.com/yd8mo55z

AALBORG UNIVERSITY
DENMARK

Motivation



Periodic signals

A periodic signal repeats itself after some period τ or, equivalently, with some frequency ω_0.

- We refer to ω_0 as either the pitch (perceptual) or the fundamental frequency (physical).
- How do we estimate this value from possibly noisy and non-stationary data?

Motivation

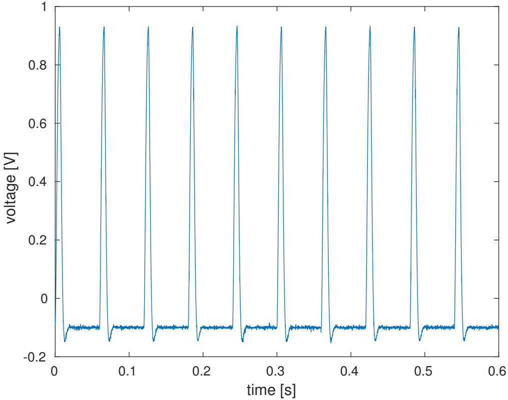
Some examples of periodic signals and applications:

- Voiced speech and singing
  - Are people singing on-key?
  - Diagnosis of the Parkinson's disease
- Many musical instruments (e.g., guitar, violin, flute, trumpet, piano)
  - Tuning of instruments
  - Music transcription
- Electrocardiographic (ECG) signals
  - Measure your heart rate or heart rate variability
  - Heart defect diagnosis
- Rotating machines
  - Vibration analysis
  - Rotation speed

Motivation

Example: RPM estimation from tachometer signal

SNR: 40 dB



Motivation

Example: RPM estimation from tachometer signal

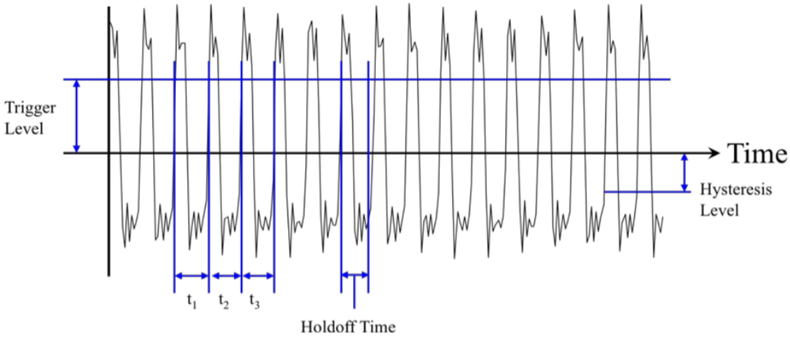


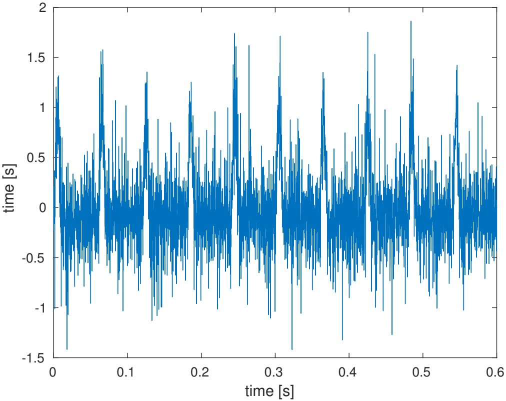
Figure 1: Example tachometer signal with processing parameters labeled.

Figure courtesy of *A. Brandt, Noise and vibration analysis: signal analysis and experimental procedures. John Wiley & Sons, 2011.*

Motivation

Example: RPM estimation from tachometer signal

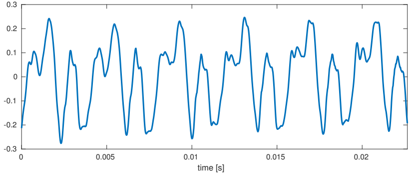
SNR: 0 dB



Outline

Outline

Correlation-based Methods



For a periodic signal $x(n)$ with a period $\tau = 2\pi/\omega_0$, we have that

$$x(n) = x(n - \tau) = x(n - 2\pi/\omega_0) \ . \tag{1}$$

- Unfortunately, τ is unknown so we have to try out different τ's (or ω_0's) to find one that satisfies the above equation.
- Real-world signals are not perfectly periodic so we might never find one.
- Instead, the estimate of τ is the value which minimises some objective function.

Correlation-based Methods

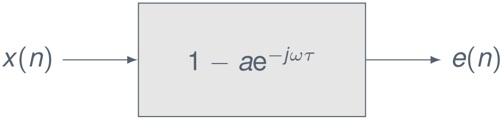
Consider the objective function

$$J(a,\tau) = \sum_{n=\tau_{\text{MAX}}}^{N-1} |e(n)|^2 \tag{2}$$

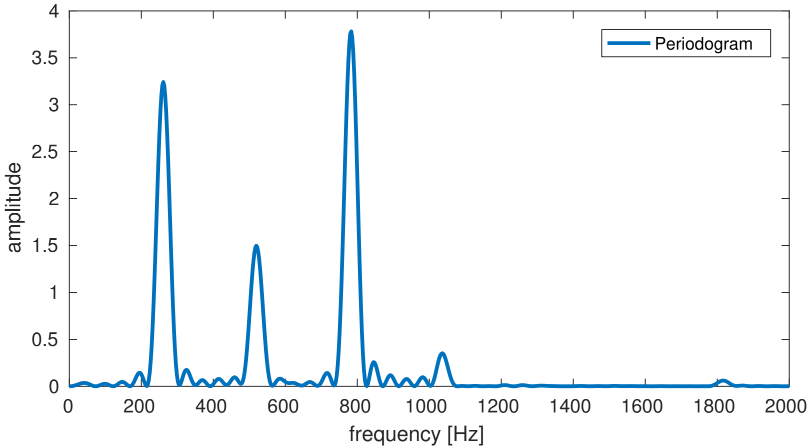
for a segment of data $\{x(n)\}_{n=0}^{N-1}$ where

$$e(n) = x(n) - ax(n-\tau)\ , \quad a > 0\ \wedge\ \tau \in [\tau_{\text{MIN}}, \tau_{\text{MAX}}] \tag{3}$$

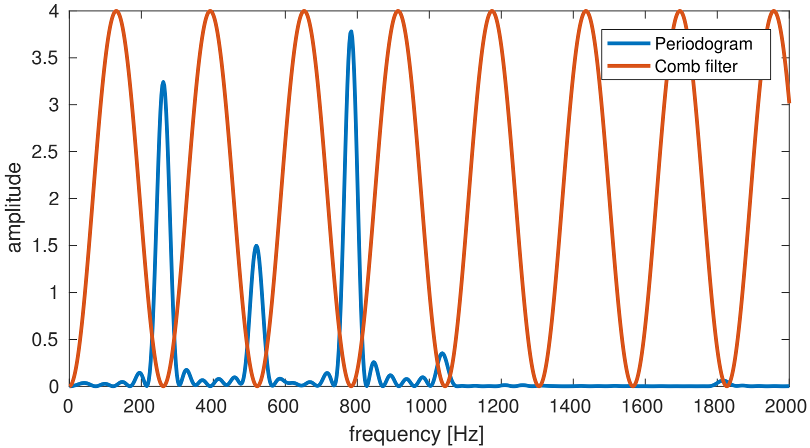
Often referred to as comb-filtering.



Correlation-based Methods



Correlation-based Methods



Correlation-based Methods

Conditioned on τ, the optimal value for a is

$$\hat{a}(\tau) = \max\left(\frac{\sum_{n=\tau_{\text{MAX}}}^{N-1} x(n)x(n-\tau)}{\sum_{n=\tau_{\text{MAX}}}^{N-1} x^2(n-\tau)}, 0\right) \quad (4)$$

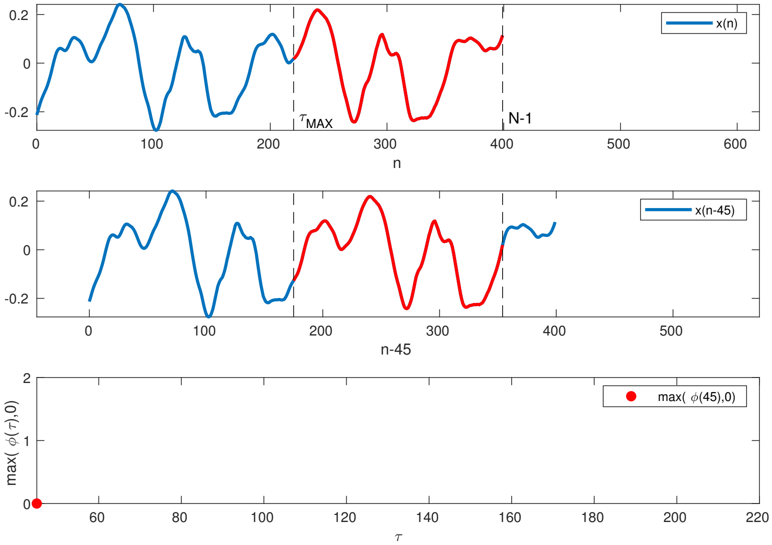
Inserting this into the objective $J(a, \tau)$ yields the estimator

$$\hat{\tau} = \underset{\tau \in [\tau_{\text{MIN}}, \tau_{\text{MAX}}]}{\text{argmax}} \max(\phi(\tau), 0) \quad (5)$$

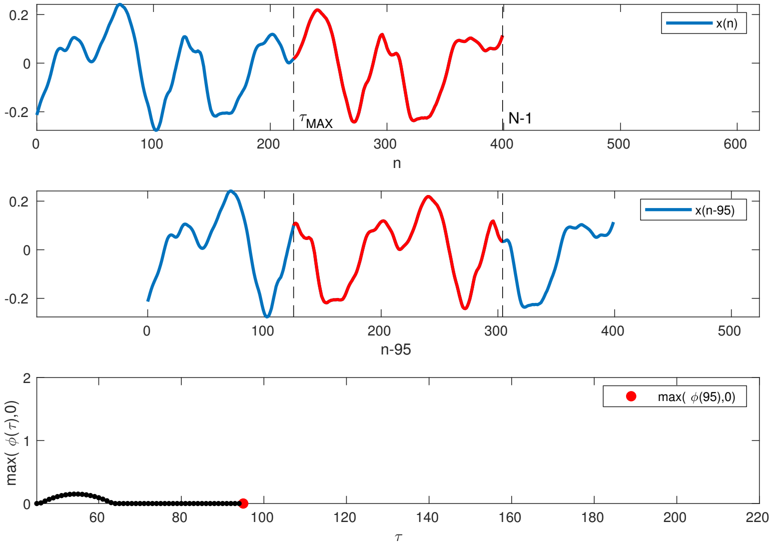
where $\phi(\tau) \in [-1, 1]$ is the normalised cross correlation function given by

$$\phi(\tau) = \frac{\sum_{n=\tau_{\text{MAX}}}^{N-1} x(n)x(n-\tau)}{\sqrt{\sum_{n=\tau_{\text{MAX}}}^{N-1} x^2(n) \sum_{n=\tau_{\text{MAX}}}^{N-1} x^2(n-\tau)}} \quad (6)$$

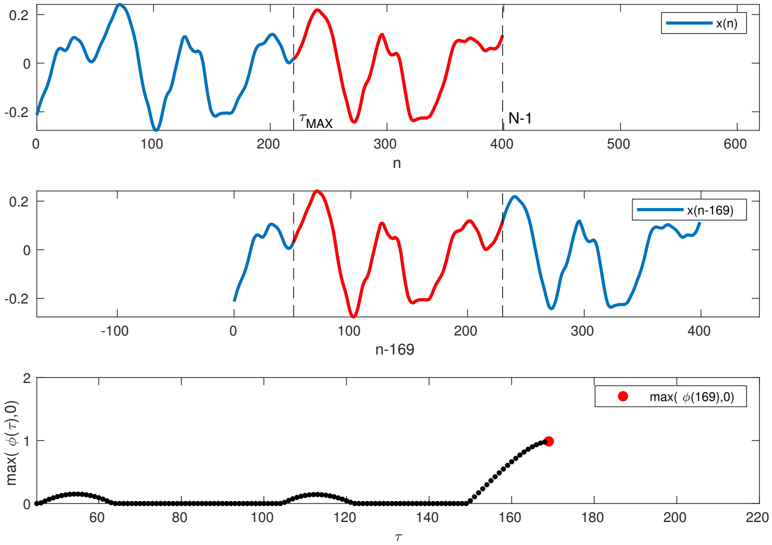
Correlation-based Methods



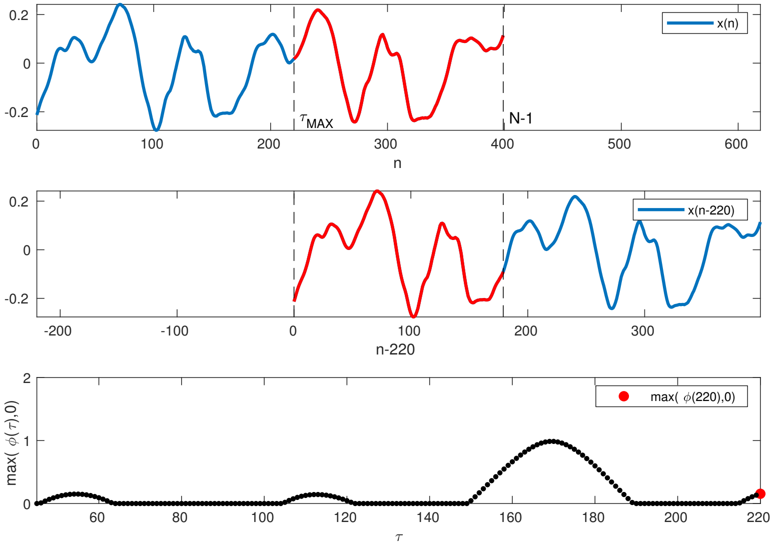
Correlation-based Methods



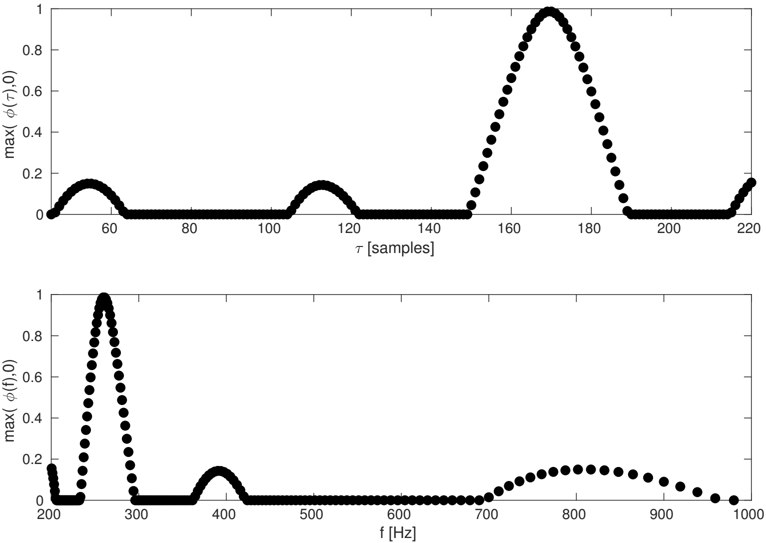
Correlation-based Methods



Correlation-based Methods



Correlation-based Methods



Correlation-based Methods

.... but is anyone actually using the comb filtering method?

PRAAT: (Boersma, 1993), well over 1000 citations (Google Scholar)
Maximises a windowed normalised cross-correlation function

RAPT: (Talkin, 1995), nearly 1000 citations (Google Scholar)
Maximises a normalised cross-correlation function

YIN: (Cheveigné, 2002), nearly 2000 citations (Google Scholar)
Minimises the comb filtering error for $a = 1$

Kaldi: (Ghahremani et al., 2014), nearly 150 citations (Google Scholar)
Maximises a normalised cross-correlation function

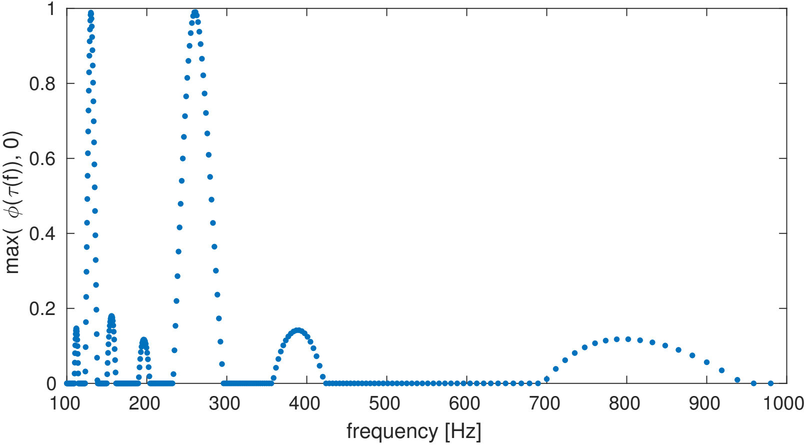
Correlation-based Methods

Was that really everything?

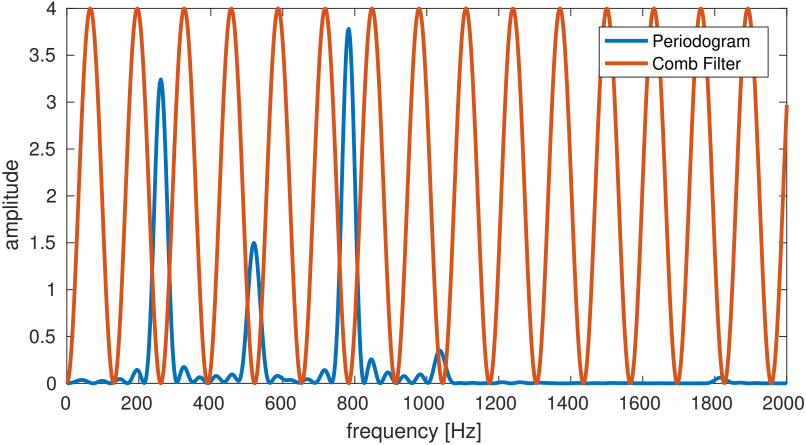
No! Four problems with the correlation-based methods:

1. is prone to producing subharmonic errors,
2. has a sub-optimal time-frequency resolution,
3. is not robust to noise, and
4. not statistically efficient.

Correlation-based Methods
Subharmonic error



Correlation-based Methods
Subharmonic error



Correlation-based Methods

What can we do about these problems?

- Hundreds of published pitch estimators trying to solve these problems using various heuristics.
- A fundamental flaw of the comb-filtering principle?

Correlation-based Methods

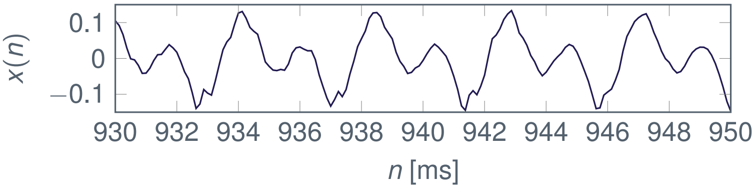
Five minutes active break

Please complete the SMCNordic pitch survey.

- Go to `http://tinyurl.com/y3ny4n4n`
- Fill out the form to the best of your ability

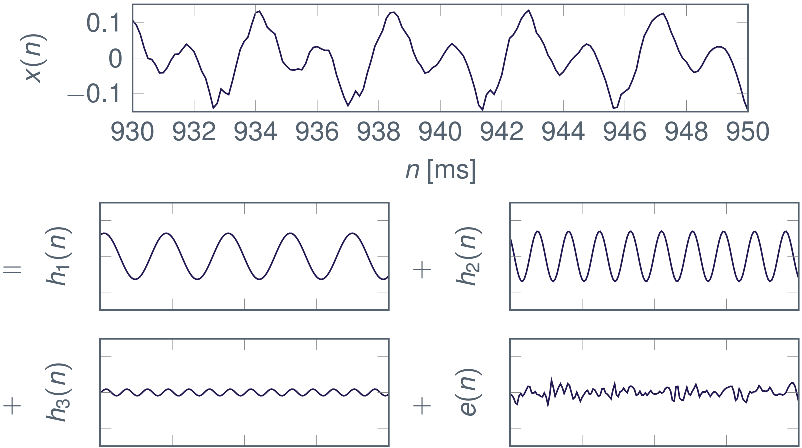
Outline

Nonlinear Least Squares Methods
Harmonic Model



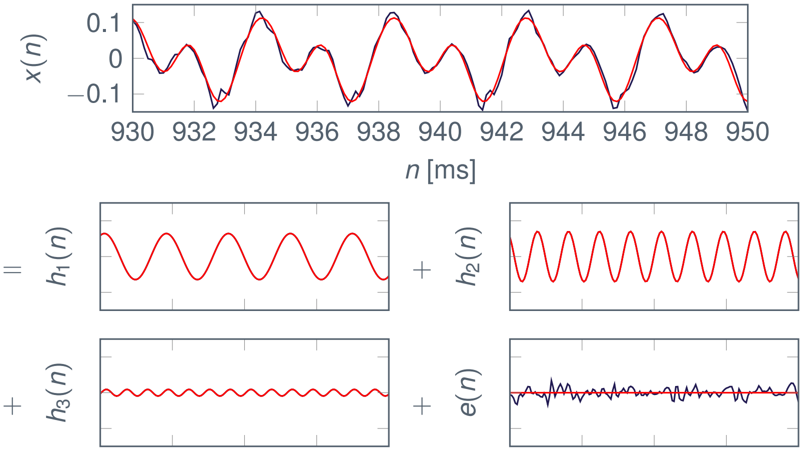
Nonlinear Least Squares Methods

Harmonic Model



Nonlinear Least Squares Methods

Harmonic Model



Nonlinear Least Squares Methods

Harmonic Model

Mathematical Model

The signal model for any periodic signal is

$$s(n) = \sum_{l=1}^{L} h_l(n) = \sum_{l=1}^{L} A_l \cos(\omega_0 l n + \phi_l) \quad (7)$$

where

- A_l real amplitude of the lth harmonic
- ϕ_l initial phase of the lth harmonic
- ω_0 fundamental frequency in radians/sample
- L the number of harmonics/model order

Nonlinear Least Squares Methods

Harmonic Model

Can we actually use models?

In 1987, G. E. P. Box (a British statistician) wrote

> *Essentially, all models are wrong, but some are useful.*

Nonlinear Least Squares Methods
Harmonic Model

Can we actually use models?

In 1987, G. E. P. Box (a British statistician) wrote

Essentially, all models are wrong, but some are useful.

- Do NOT think about models as exact physical representations of a phenomenon in the real world.
- Instead, think of models as an explicit way of stating your assumptions about the phenomenon.
- Models can be critisised (and improved on) since the assumptions are explicit.
- Models allow us to assert under which conditions a problem is optimally solved .

Nonlinear Least Squares Methods

Method of Least Squares

Instead of considering the comb-filtering error

$$e(n) = x(n) - ax(n-\tau)\ , \qquad (8)$$

we consider the least-squares error

$$e(n) = x(n) - s(n, \boldsymbol{\theta})\ , \qquad n = 0, 1, \ldots, N-1 \qquad (9)$$

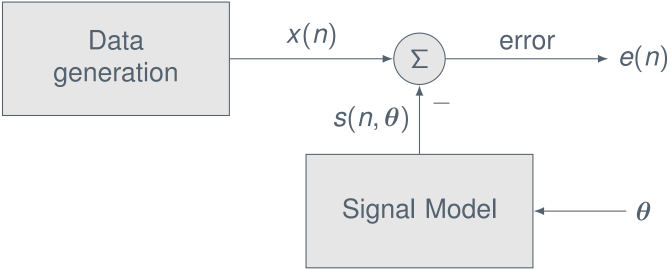
where $s(n, \boldsymbol{\theta})$ is a harmonic model given by

$$s(n, \boldsymbol{\theta}) = \sum_{l=1}^{L} A_l \cos(l\omega_0 n + \phi_l) \qquad (10)$$

$$\boldsymbol{\theta} = \begin{bmatrix} A_1 & \cdots & A_L & \phi_1 & \cdots & \phi_L & \omega_0 \end{bmatrix}^T \qquad (11)$$

Nonlinear Least Squares Methods
Method of Least Squares

The method of least-squares



- The vector θ contains the model parameters
- The signal $s(n, \theta)$ is produced by the signal model
- The signal $x(n)$ is the observed data
- The error consists of noise and model inaccuracies

Nonlinear Least Squares Methods

Method of Least Squares

The nonlinear least squares (NLS) method is that of solving

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, J(\theta) \tag{12}$$

where $J(\theta)$ measures the squared error

$$J(\theta) = \sum_{n=0}^{N-1} |e(n)|^2 = \sum_{n=0}^{N-1} |x(n) - s(n,\theta)|^2 \tag{13}$$

- Solving this problem naïvely is very computationally demanding since the fundamental frequency is a nonlinear parameter.
- Asymptotically, however, an efficient solution exists which for historical reasons is called harmonic summation (Noll, 1969).

Outline

The NLS Estimator

The harmonic model

$$x(n) = \sum_{l=1}^{L} \left[a_l \cos(l\omega_0 n) - b_l \sin(l\omega_0 n)\right] + e(n) \tag{14}$$

for $n = n_0, n_0 + 1, \ldots, n_0 + N - 1$ can be written as

$$\boldsymbol{x} = \boldsymbol{Z}_L(\omega_0)\boldsymbol{\alpha}_L + \boldsymbol{e} \tag{15}$$

where

$$\boldsymbol{Z}_L(\omega) = \begin{bmatrix}\boldsymbol{c}(\omega) & \boldsymbol{c}(2\omega) & \cdots & \boldsymbol{c}(L\omega) & \boldsymbol{s}(\omega) & \boldsymbol{s}(2\omega) & \cdots & \boldsymbol{s}(L\omega)\end{bmatrix}$$

$$\boldsymbol{c}(\omega) = \begin{bmatrix}\cos(\omega n_0) & \cdots & \cos(\omega(n_0 + N - 1))\end{bmatrix}^T$$

$$\boldsymbol{s}(\omega) = \begin{bmatrix}\sin(\omega n_0) & \cdots & \sin(\omega(n_0 + N - 1))\end{bmatrix}^T$$

$$\boldsymbol{\alpha}_l = \begin{bmatrix}\boldsymbol{a}_L^T & -\boldsymbol{b}_L^T\end{bmatrix}^T, \ \boldsymbol{a}_L = \begin{bmatrix}a_1 & \cdots & a_L\end{bmatrix}^T, \ \boldsymbol{b}_L = \begin{bmatrix}b_1 & \cdots & b_L\end{bmatrix}^T$$

The NLS Estimator

The least squares error is

$$\sum_{n=0}^{N-1} e^2(n) = \boldsymbol{e}^T\boldsymbol{e} = [\boldsymbol{x} - \boldsymbol{Z}_L(\omega_0)\boldsymbol{\alpha}_L]^T [\boldsymbol{x} - \boldsymbol{Z}_L(\omega_0)\boldsymbol{\alpha}_L] \quad (16)$$

Conditioned on ω_0, the estimate of $\boldsymbol{\alpha}_L$ is

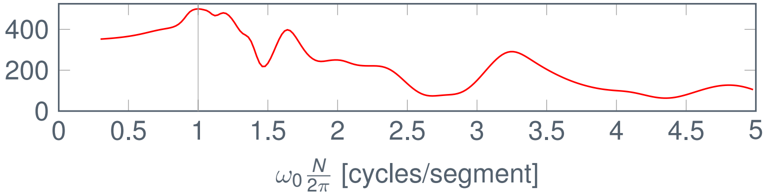
$$\hat{\boldsymbol{\alpha}}_L(\omega_0) = \left[\boldsymbol{Z}_L^T(\omega_0)\boldsymbol{Z}_L(\omega_0)\right]^{-1} \boldsymbol{Z}_L^T(\omega_0)\boldsymbol{x} \quad (17)$$

Inserting this back into the objective yields the NLS estimator

$$\hat{\omega}_{0,L} = \underset{\omega_0 \in [\omega_{\text{MIN}}, \omega_{\text{MAX}}]}{\text{argmax}} \; \boldsymbol{x}^T\boldsymbol{Z}_L(\omega_0)\left[\boldsymbol{Z}_L^T(\omega_0)\boldsymbol{Z}_L(\omega_0)\right]^{-1} \boldsymbol{Z}_L^T(\omega_0)\boldsymbol{x} \quad (18)$$

The NLS estimator has been known since (Quinn and Thomson, 1991), but is costly to compute.

The NLS Estimator



1. Compute NLS cost function

$$\hat{\omega}_{0,L} = \underset{\omega_0 \in [\omega_{\text{MIN}}, \omega_{\text{MAX}}]}{\text{argmax}} \; \boldsymbol{x}^T \boldsymbol{Z}_L(\omega_0) \left[\boldsymbol{Z}_L^T(\omega_0) \boldsymbol{Z}_L(\omega_0) \right]^{-1} \boldsymbol{Z}_L^T(\omega_0) \boldsymbol{x} \quad (19)$$

   on an F/L-point uniform grid for all model orders $L \in \{1, \dots, L_{\text{MAX}}\}$.
2. Optionally refine the L_{MAX} grid estimates.
3. Do model comparison.

The NLS Estimator

Fast NLS Algorithm

A MATLAB implementation of the NLS estimator

```
% create an estimator object (the data independent step is computed)
f0Estimator = fastF0Nls(nData, maxNoHarmonics, f0Bounds);
% analyse a segment of data
[f0Estimate, estimatedNoHarmonics, estimatedLinParam] = ...
    f0Estimator.estimate(data);
```

- The algorithm also includes model comparison.
- The algorithm can also be set-up to work for a model with a non-zero DC-value.
- A C++-implementation is also available (although not as refined as the MATLAB implementation).
- Can be downloaded from `https://github.com/jkjaer/fastF0Nls`.

Outline

The Harmonic Summation (HS) estimator

Harmonic summation (HS) estimator

Asymptotically,

$$\lim_{N\to\infty} \frac{2}{N} \boldsymbol{Z}_L^T(\omega_0)\boldsymbol{Z}_L(\omega_0) = \boldsymbol{I}_L \ . \tag{20}$$

Using this limit as an approximation gives the harmonic summation estimator (Noll, 1969)

$$\hat{\omega}_{0,L} = \underset{\omega_0\in[\omega_{\text{MIN}},\omega_{\text{MAX}}]}{\text{argmax}} \ \boldsymbol{x}^T\boldsymbol{Z}_L(\omega_0)\boldsymbol{Z}_L^T(\omega_0)\boldsymbol{x} = \underset{\omega_0\in[\omega_{\text{MIN}},\omega_{\text{MAX}}]}{\text{argmax}} \sum_{l=1}^{L} |X(\omega_0 l)|^2$$

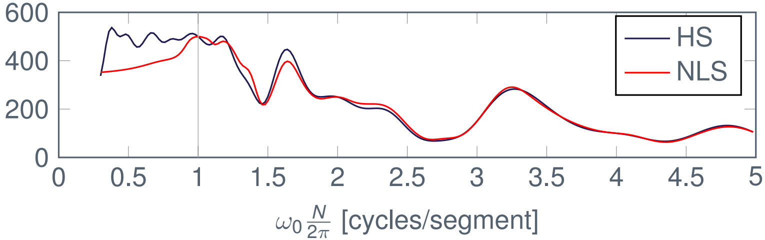
The HS estimator is also referred to as approximate NLS (aNLS).

Harmonic summation (HS) estimator

NLS vs. HS

Some remarks:

- The HS method works very well, unless the fundamental frequency is low or the maximum harmonic component is close to the Nyquist frequency.
- The HS method can be implemented very efficiently using a single FFT.
- The order of complexity for NLS has recently been decreased to that of HS (Nielsen et al., 2017).



Outline

Comparison of Methods

What could be evaluated?

1. Estimation accuracy
2. Robustness to noise
3. Time-frequency resolution
4. Computational complexity

Outline

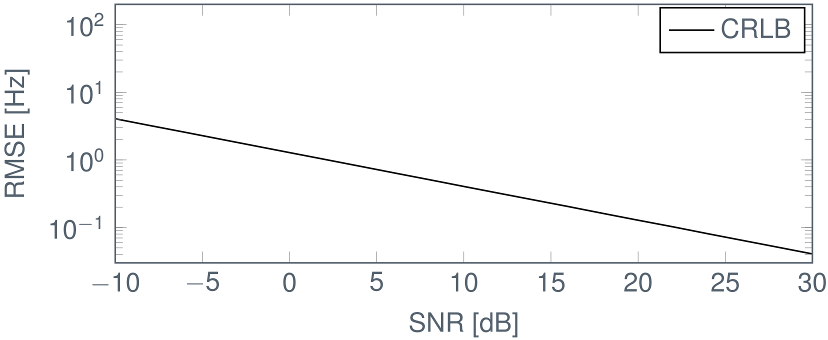
Comparison of Methods

Robustness to noise

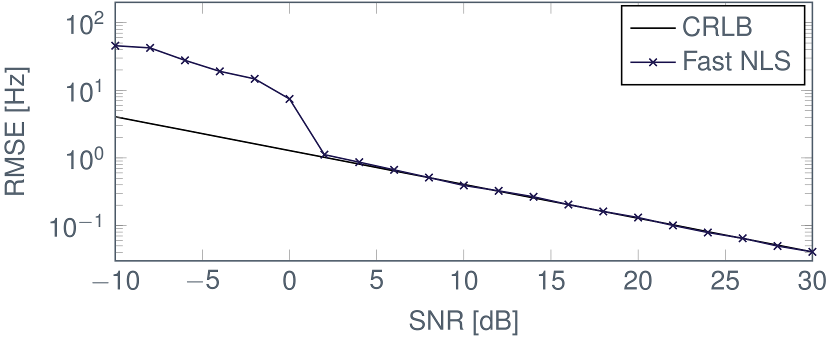
Simulation setup

- Segment size of 25 ms at a sampling frequency of 8000 Hz.
- Estimate the pitch from 1000 Monte Carlo runs for every SNR.
- In each run, the true pitch is randomly selected from $[90, 380]$ Hz and the true phases are also generated at random.
- The true amplitudes are exponentially decreasing.
- The true model order is 7.
- Each method searches for a pitch in the range $[80, 400]$ Hz.
- The maximum model order in NLS is set to 15.
- The noise is white and Gaussian.
- No pitch tracking used in any method.

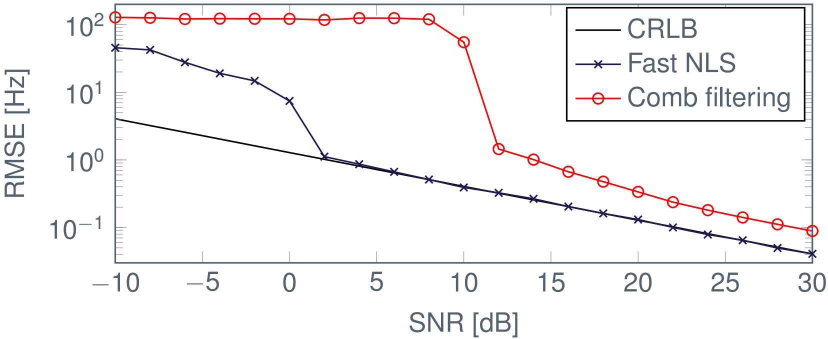
Comparison of Methods
Robustness to noise



Comparison of Methods
Robustness to noise

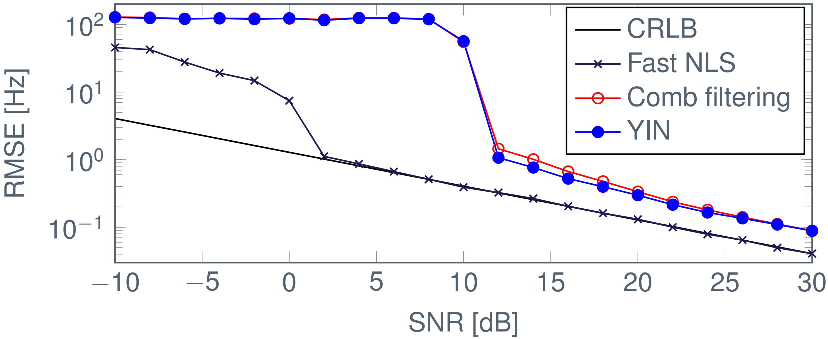


Comparison of Methods
Robustness to noise



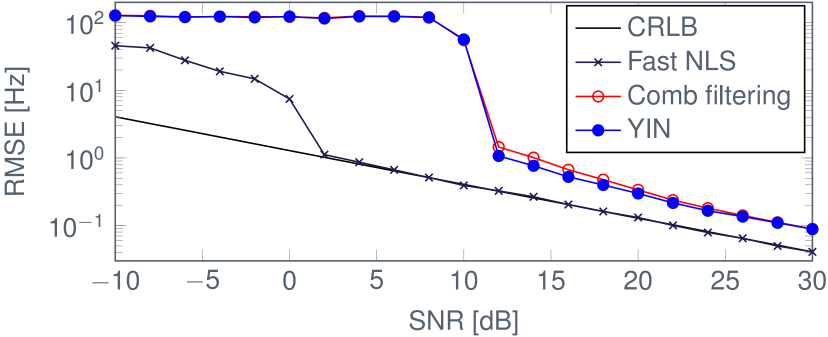
Comparison of Methods

Robustness to noise



Comparison of Methods

Robustness to noise

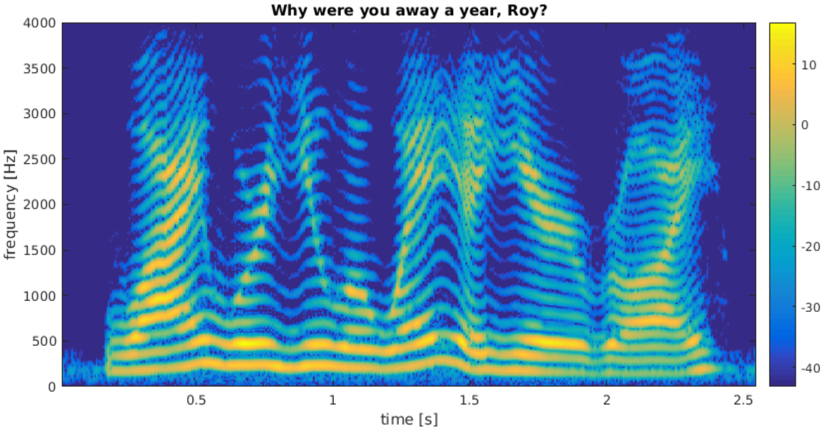


Average computation times in MATLAB

Fast NLS: 7.6 ms, Comb filter: 2.4 ms, YIN: 0.7 ms

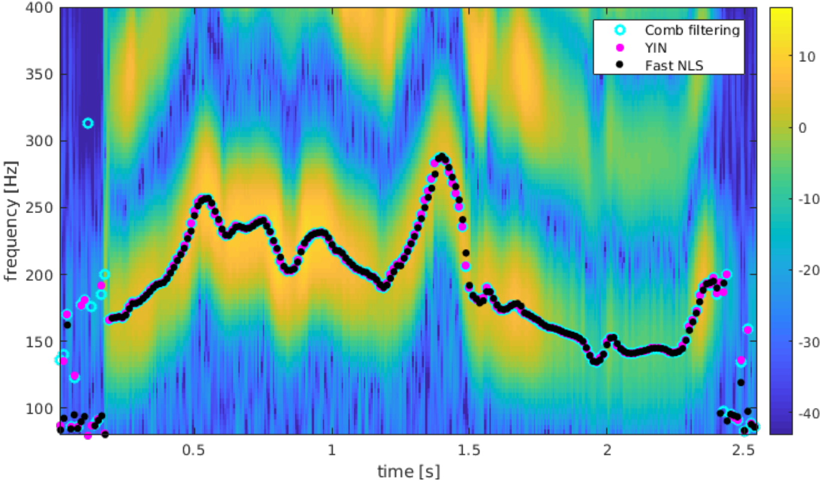
Comparison of Methods

Robustness to noise



Comparison of Methods
Robustness to noise

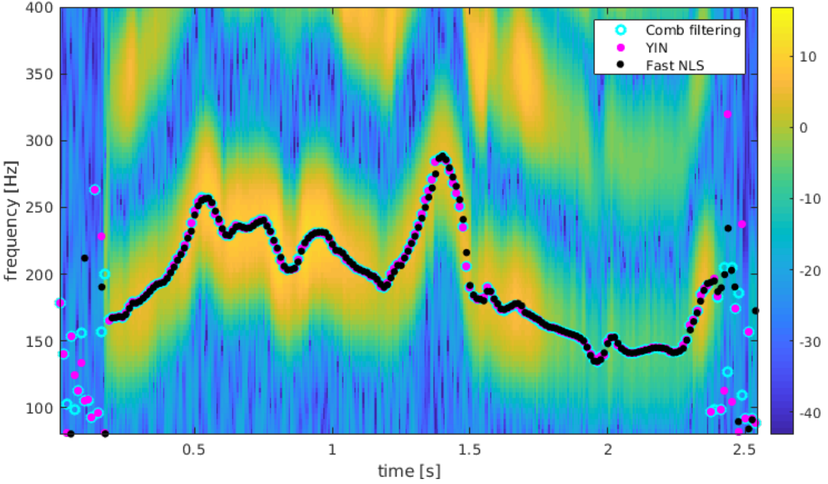
No noise and window size of 25 ms.



Comparison of Methods

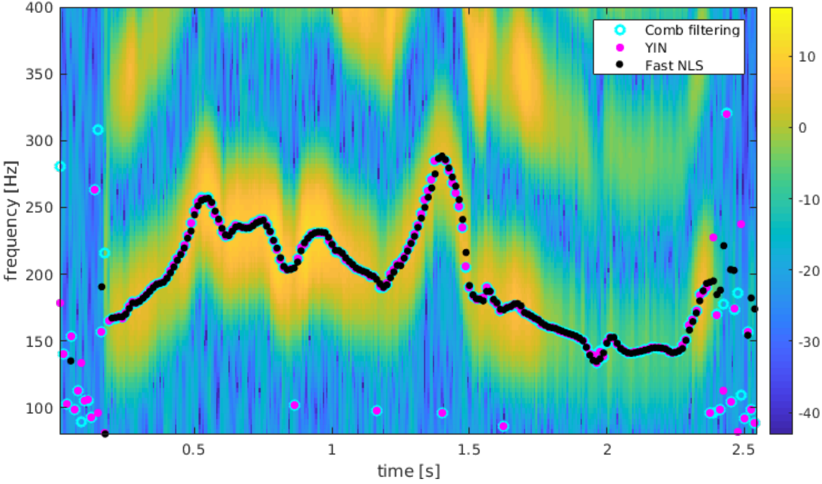
Robustness to noise

20 dB SNR and window size of 25 ms.



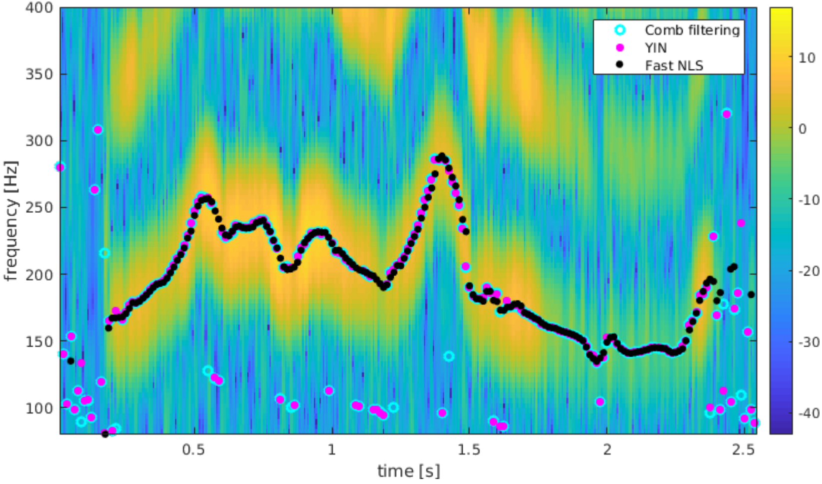
Comparison of Methods
Robustness to noise

15 dB SNR and window size of 25 ms.



Comparison of Methods
Robustness to noise

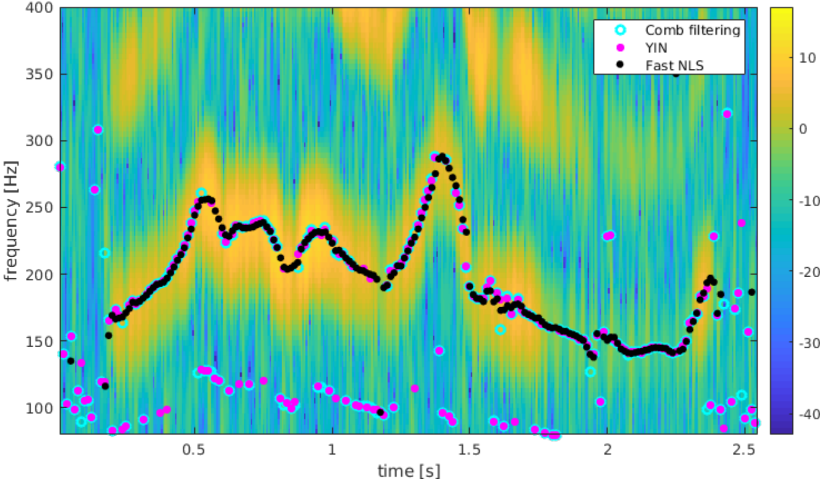
10 dB SNR and window size of 25 ms.



Comparison of Methods

Robustness to noise

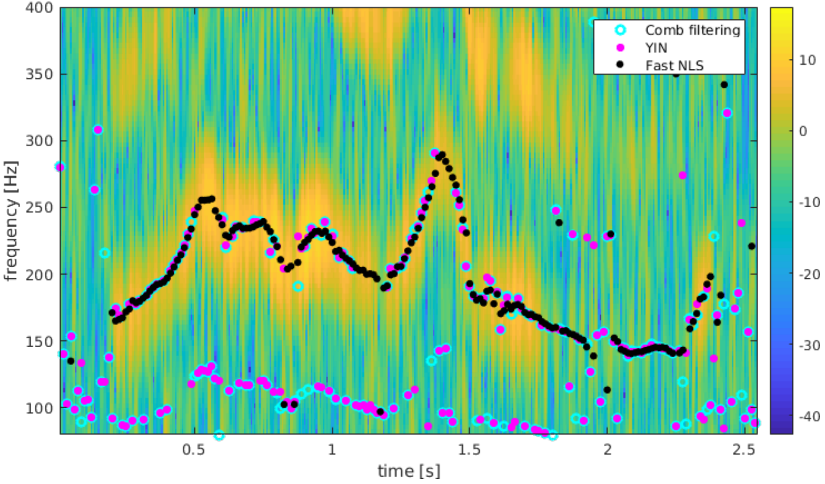
5 dB SNR and window size of 25 ms.



Comparison of Methods

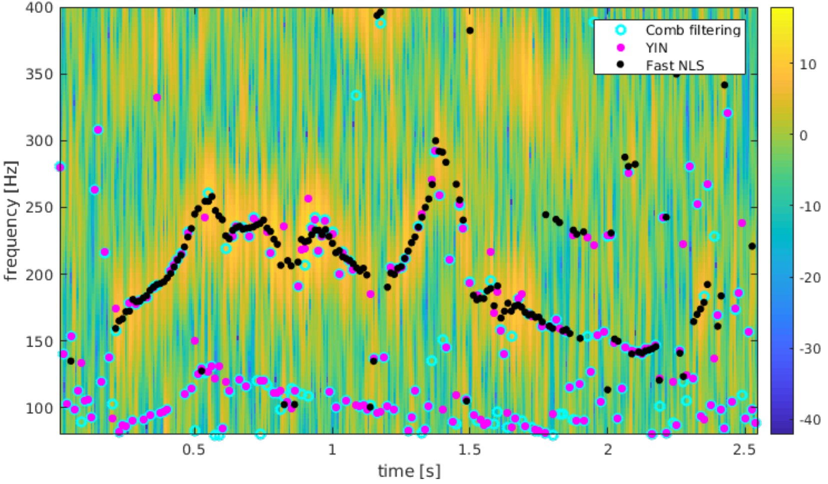
Robustness to noise

0 dB SNR and window size of 25 ms.



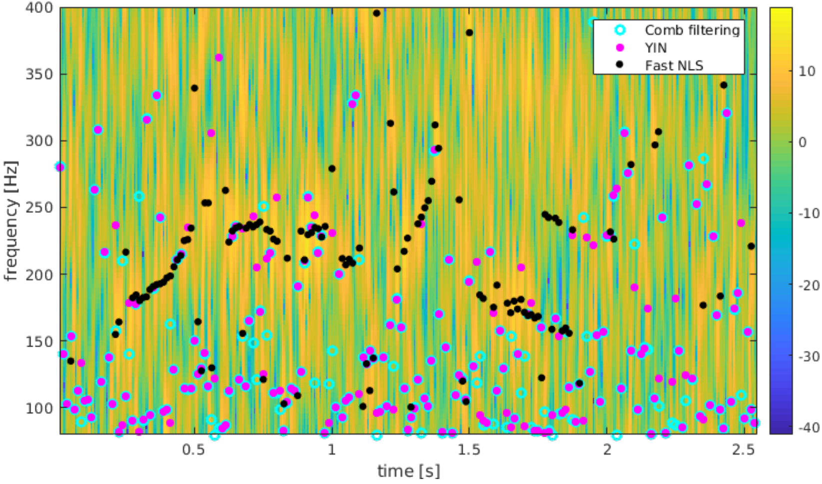
Comparison of Methods
Robustness to noise

-5 dB SNR and window size of 25 ms.



Comparison of Methods
Robustness to noise

-10 dB SNR and window size of 25 ms.



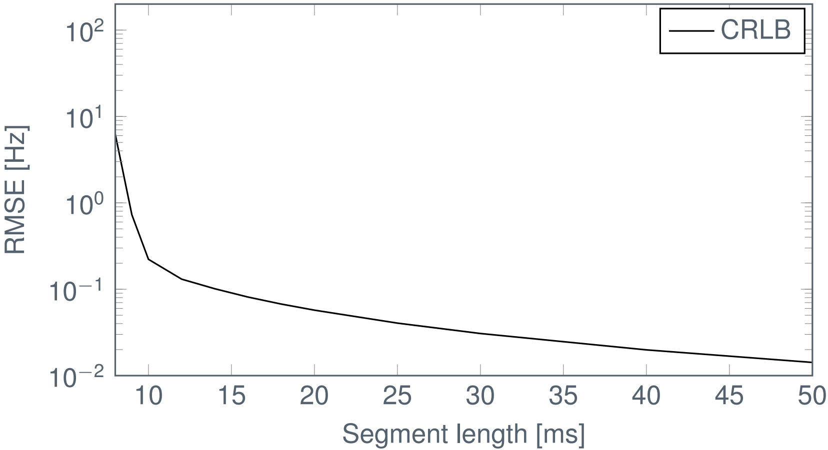
Outline

Comparison of Methods
Time-frequency resolution

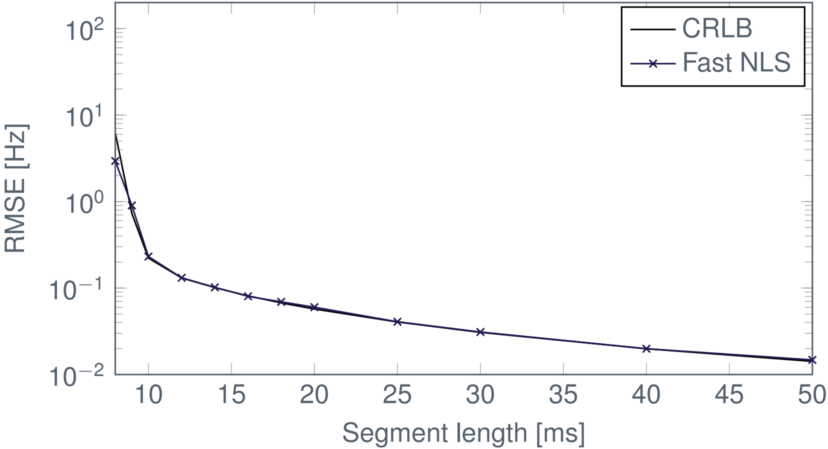
Simulation setup

- SNR of 30 dB at a sampling frequency of 8000 Hz.
- Estimate the pitch from 1000 Monte Carlo runs for every segment time.
- In each run, the true pitch is randomly selected from $[90, 380]$ Hz and the true phases are also generated at random.
- The true amplitudes are exponentially decreasing.
- The true model order is 7.
- Each method searches for a pitch in the range $[80, 400]$ Hz.
- The maximum model order in NLS is set to 15.
- The noise is white and Gaussian.
- No pitch tracking used in any method.

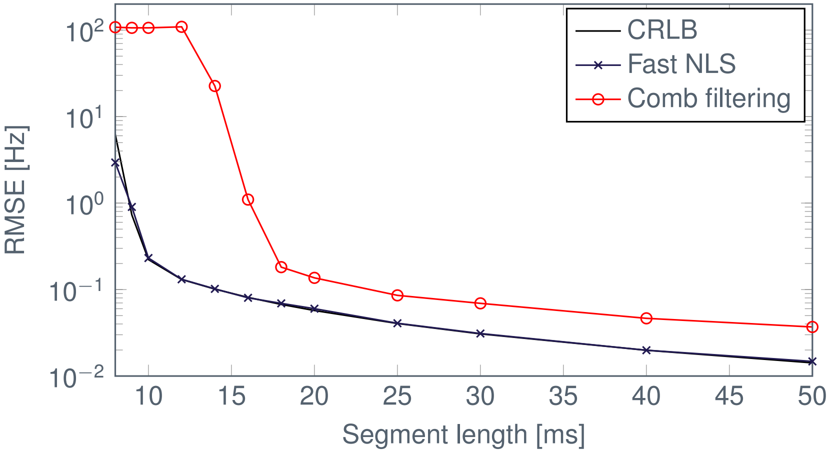
Comparison of Methods
Time-frequency resolution



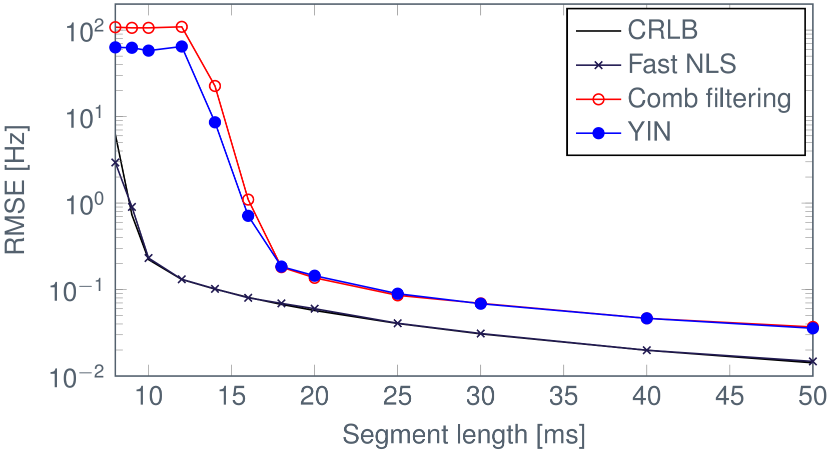
Comparison of Methods
Time-frequency resolution



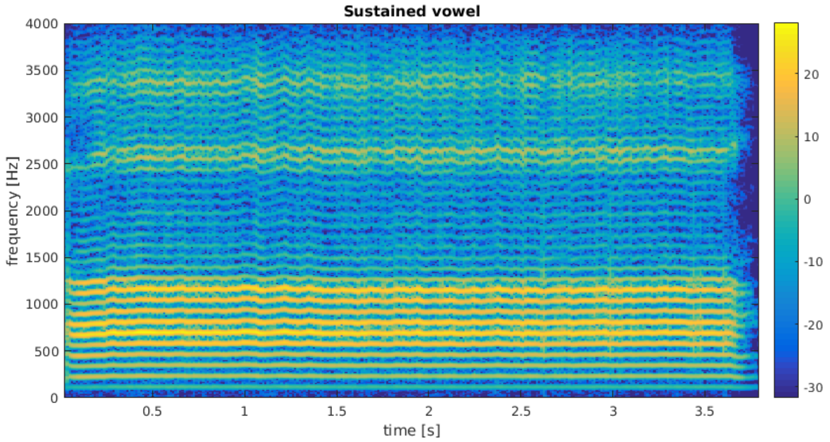
Comparison of Methods
Time-frequency resolution



Comparison of Methods
Time-frequency resolution



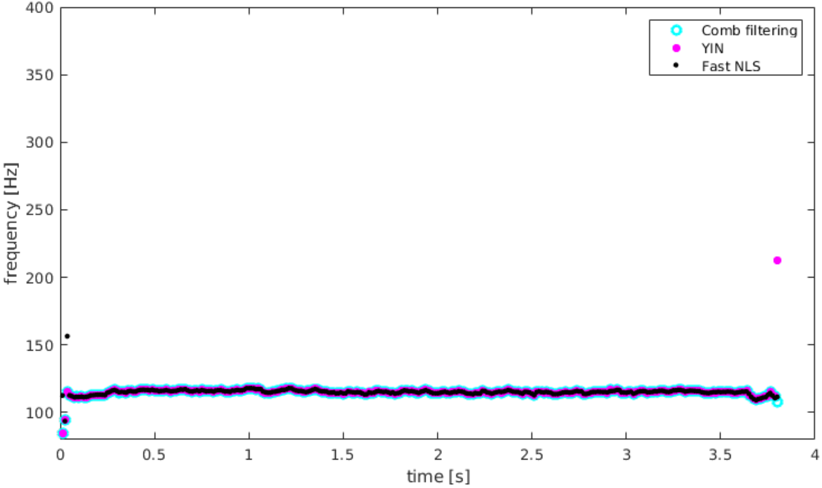
Comparison of Methods
Time-frequency resolution



Comparison of Methods

Time-frequency resolution

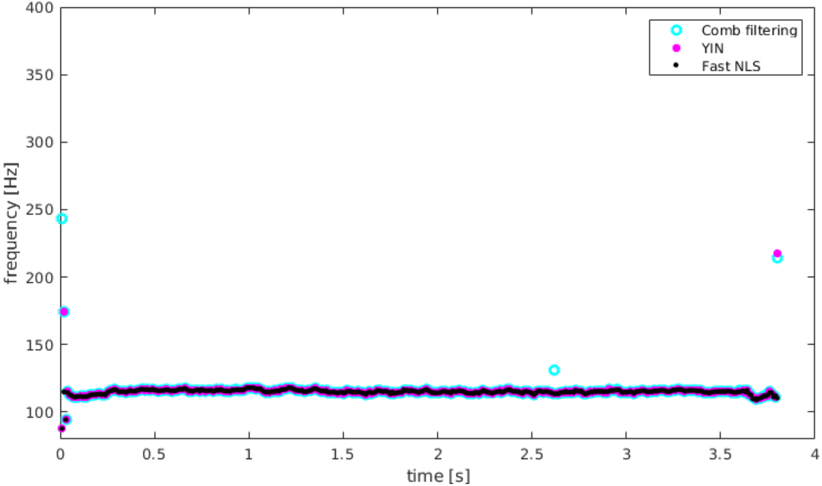
Window size of 25 ms and no noise.



Comparison of Methods

Time-frequency resolution

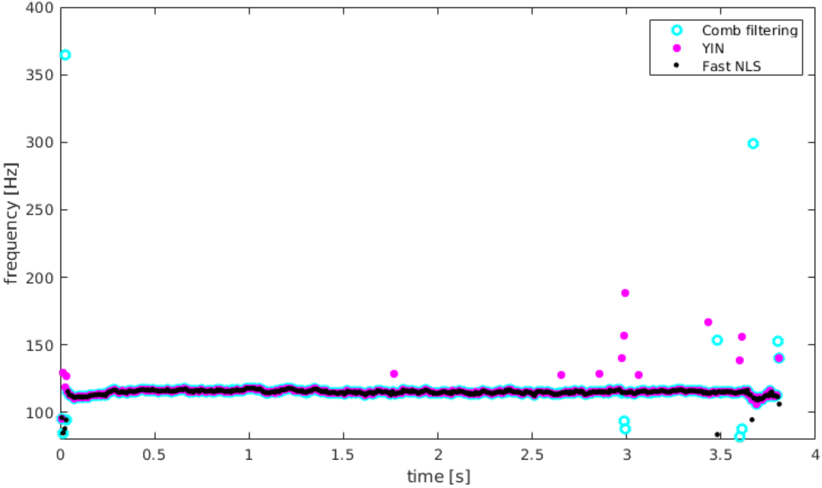
Window size of 20 ms and no noise.



Comparison of Methods

Time-frequency resolution

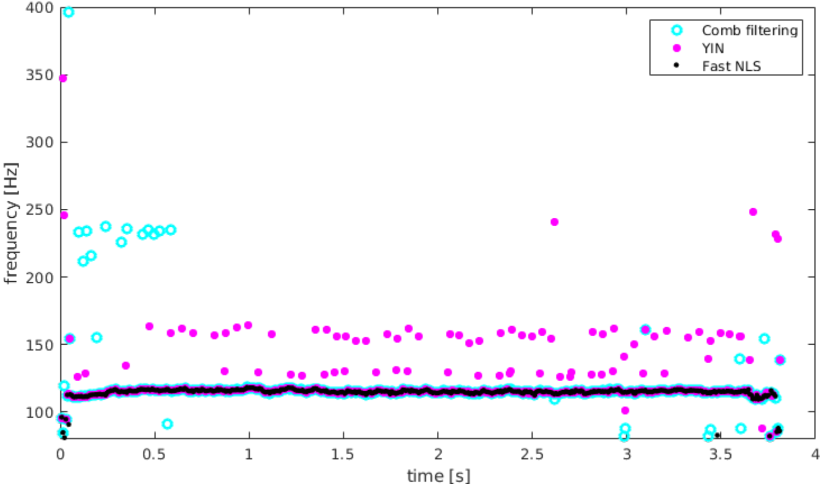
Window size of 16 ms and no noise.



Comparison of Methods

Time-frequency resolution

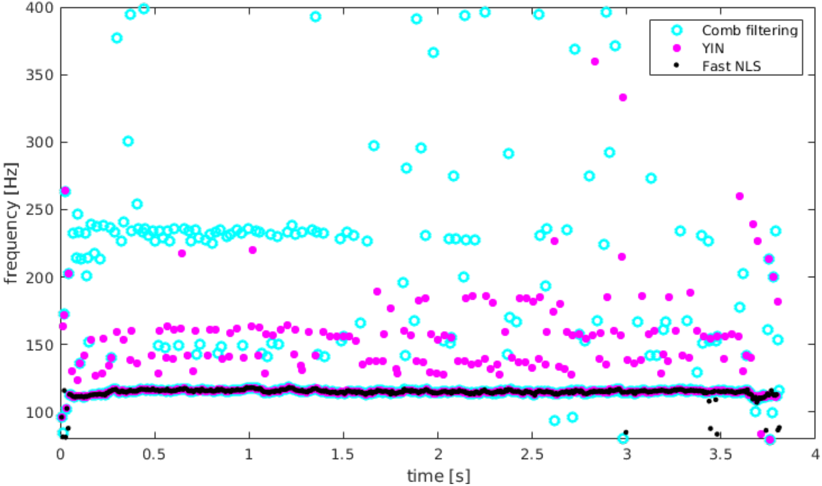
Window size of 15 ms and no noise.



Comparison of Methods

Time-frequency resolution

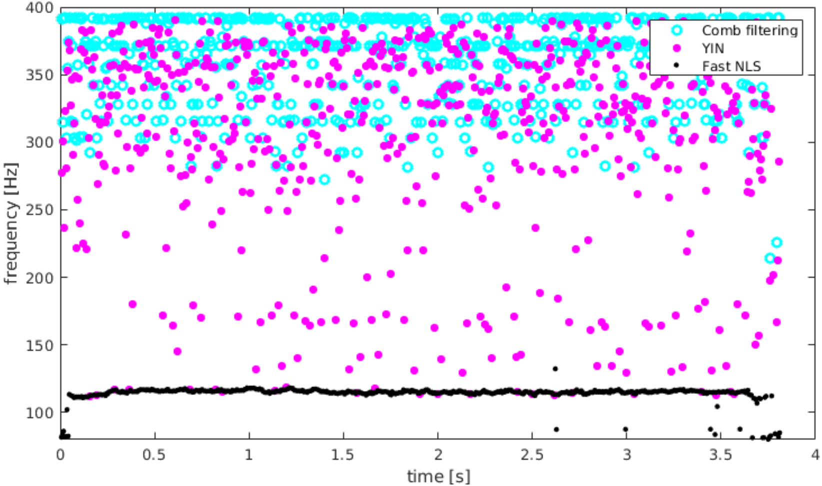
Window size of 14 ms and no noise.



Comparison of Methods

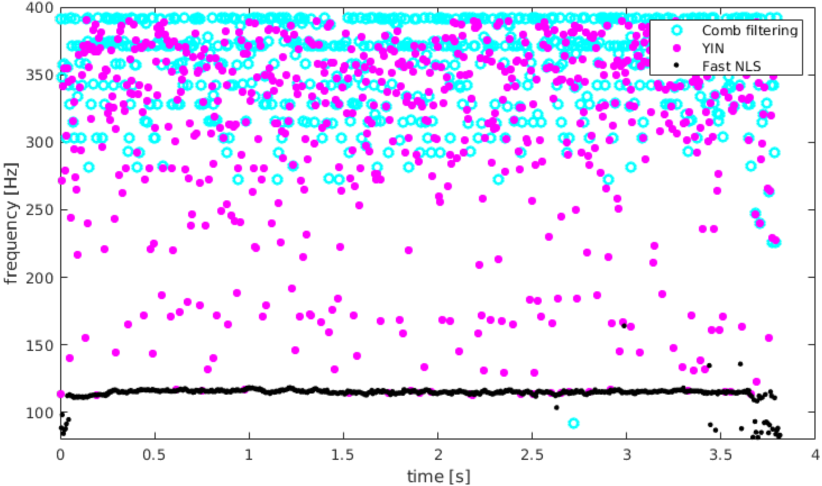
Time-frequency resolution

Window size of 12 ms and no noise.



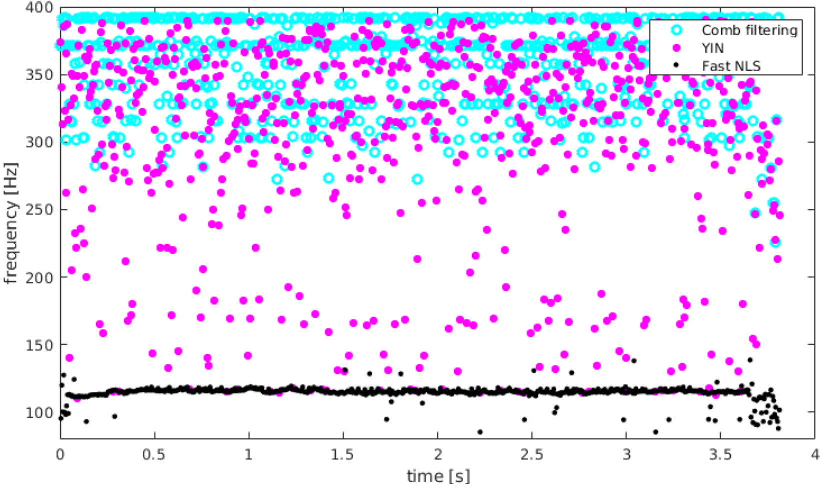
Comparison of Methods
Time-frequency resolution

Window size of 11 ms and no noise.



Comparison of Methods
Time-frequency resolution

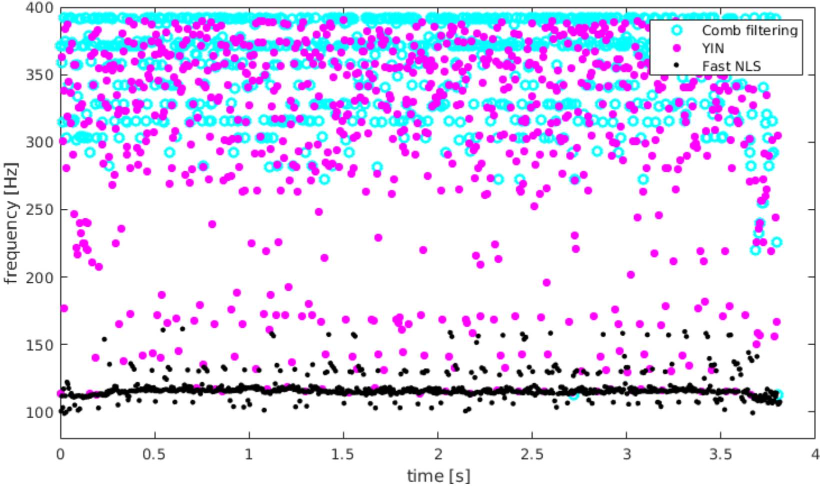
Window size of 10 ms and no noise.



Comparison of Methods

Time-frequency resolution

Window size of 9 ms and no noise.



Outline

Comparison of Methods
Summary

Correlation-based Methods
A periodic signal satisfies that

$$x(n) = x(n - \tau) \quad (21)$$

where $\tau = 2\pi/\omega_0$ is the period.

- \+ Intuitive and simple
- \+ Low computational complexity
- \+ Mature and refined set of methods
- +/- No need to estimate the model order
- \- Interpolation needed for fractional delay estimation
- \- Poor time-frequency resolution
- \- Are sensitive to noise

Comparison of Methods
Summary

Parametric Methods
Estimate the parameters in

$$x(n) = \sum_{l=1}^{L} A_l \cos(l\omega_0 n + \phi_l) + e(n) \quad (22)$$

- \+ High estimation accuracy
- \+ Work very well in even noisy conditions
- \+ Good time-frequency resolution
- +/- The model order has to be estimated
- \- Higher computational complexity
- \- Early stage methods without fine tuning (yet)
- \- Might produce over-optimistic results (e.g., due to non-stationarity)

Outline

Model Improvements

What is wrong with the harmonic model?

The harmonic model

So far, we have used the model

$$x(n) = s(n) + e(n) = \sum_{l=1}^{L} A_l \cos(\omega_0 ln + \phi_l) + e(n) \qquad (23)$$

What could be improved?

Noise model Noise is typically not white, but coloured.

Pitch tracking The pitch is typically smoothly evolving between successive frames.

Inharmonic pitch For, e.g., stiff-stringed instruments, the frequencies of the harmonics $\{\omega_l\}$ deviate (slightly) from whole multiples of the pitch ($\omega_l = \omega_0 l\sqrt{1 + Bl^2}$).

Non-stationary pitch Within a segment, the pitch is typically not stationary, but time-varying.

Model Improvements
Non-stationary pitch estimation

Non-stationary pitch estimation

- Real-world signals are non-stationary since the fundamental frequency is continuously changing.
- The harmonic model assumes that the the fundamental frequency is constant in a segment of data
- We can extend the model of the phase of the lth harmonic component to
$$\theta_l(n) \approx \phi_l + l\omega_0 n + l\beta_0 n^2/2 \quad (24)$$
where β_0 is the fundamental chirp rate.
- We refer to this model as the harmonic chirp model

$$s(n) = \sum_{l=1}^{L} A_l \cos(l\beta_0 n^2/2 + l\omega_0 n + \phi_l) \quad (25)$$

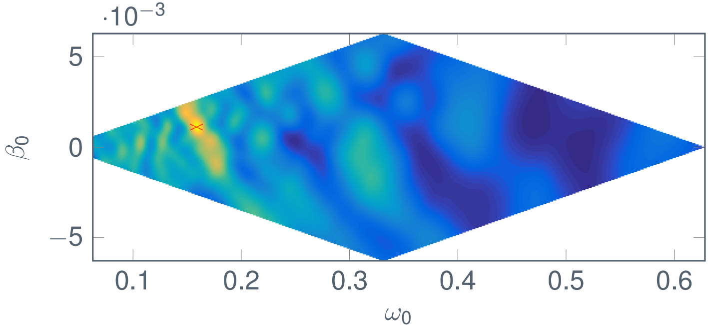
Model Improvements
Non-stationary Pitch Estimation

Nonlinear least squares (NLS) objective

$$J_L(\omega_0, \beta_0) = \mathbf{x}^T \mathbf{Z}_L(\omega_0, \beta_0) \left[\mathbf{Z}_L^T(\omega_0, \beta_0)\mathbf{Z}_L(\omega_0, \beta_0)\right]^{-1} \mathbf{Z}_L^T(\omega_0, \beta_0)\mathbf{x} \quad (26)$$

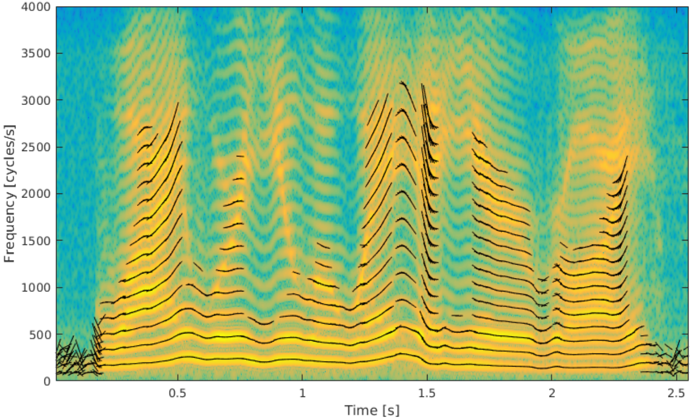
Harmonic chirp summation objective:

$$J_L(\omega_0, \beta_0) = \mathbf{x}^T \mathbf{Z}_L(\omega_0, \beta_0)\mathbf{Z}_L^T(\omega_0, \beta_0)\mathbf{x} \quad (27)$$



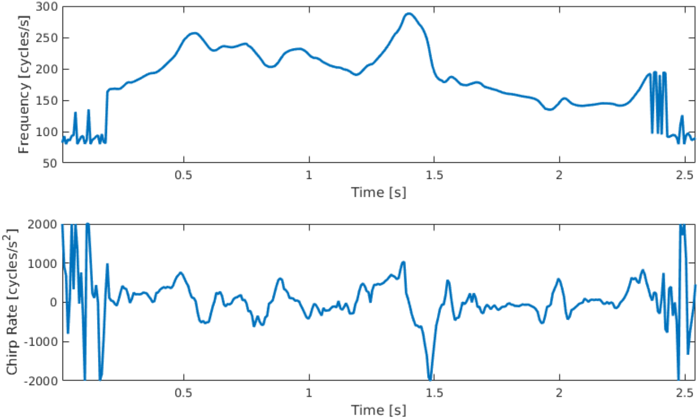
Model Improvements
Non-stationary Pitch Estimation

Window size of 30 ms, 75 % overlap, and no noise



Model Improvements
Non-stationary Pitch Estimation

Window size of 30 ms, 75 % overlap, and no noise



Outline

Summary

- Published correlation-based methods are more mature than published parametric methods in that they tend to include everything (pitch detection, estimation, and tracking) and are less computationally costly.
- However, parametric pitch estimation methods typically outperform correlation-based methods in terms of estimation accuracy, noise robustness, and time-frequency resolution.
- The modelling assumptions are explicit in parametric methods.
- Consequently, we can easily extend the model to take more complex phenomena into account.
- Besides NLS, examples of other parametric methods are subspace and filtering methods (Christensen and Jakobsson, 2009).

Resources

- Audio Analysis Lab: `https://audio.create.aau.dk/`
- Pitch Estimation for Dummies: `http://madsgc.blog.aau.dk/resources/`
- MATLAB code: `https://github.com/jkjaer/fastF0Nls`
- YouTube videos: `http://tinyurl.com/yd8mo55z`

[1] J. K. Nielsen, T. L. Jensen, J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Fast fundamental frequency estimation: Making a statistically efficient estimator computationally efficient," *Elsevier Signal Processing*, vol. 135, pp. 188–197, 2017.

[2] J. K. Nielsen, M. G. Christensen, and S. H. Jensen, "Default Bayesian estimation of the fundamental frequency," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 3, pp. 598–610, Mar. 2013.

[3] M. G. Christensen and A. Jakobsson, *Multi-Pitch Estimation*, San Rafael, CA, USA: Morgan & Claypool, 2009.