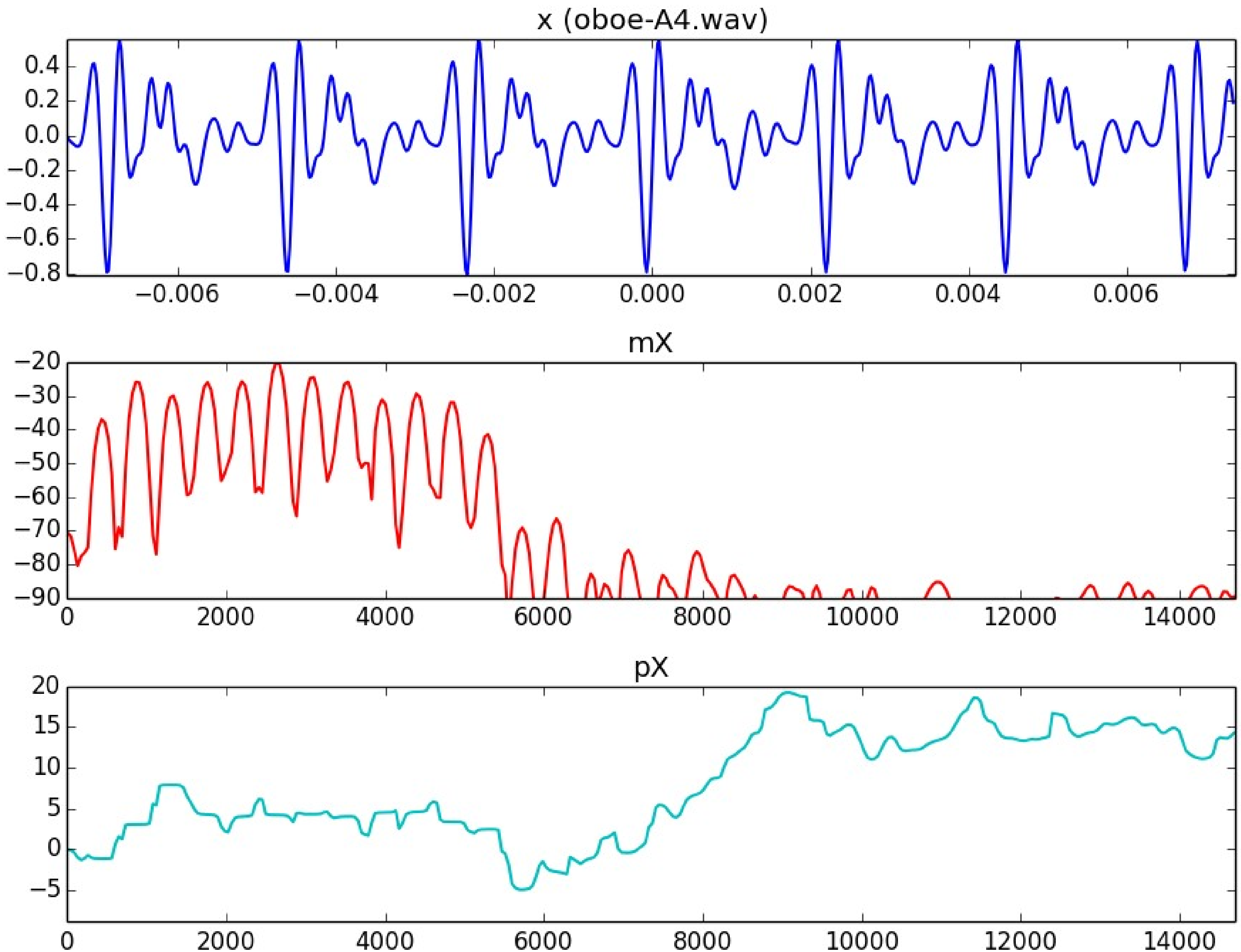
# **6T2:** Fundamental frequency detection

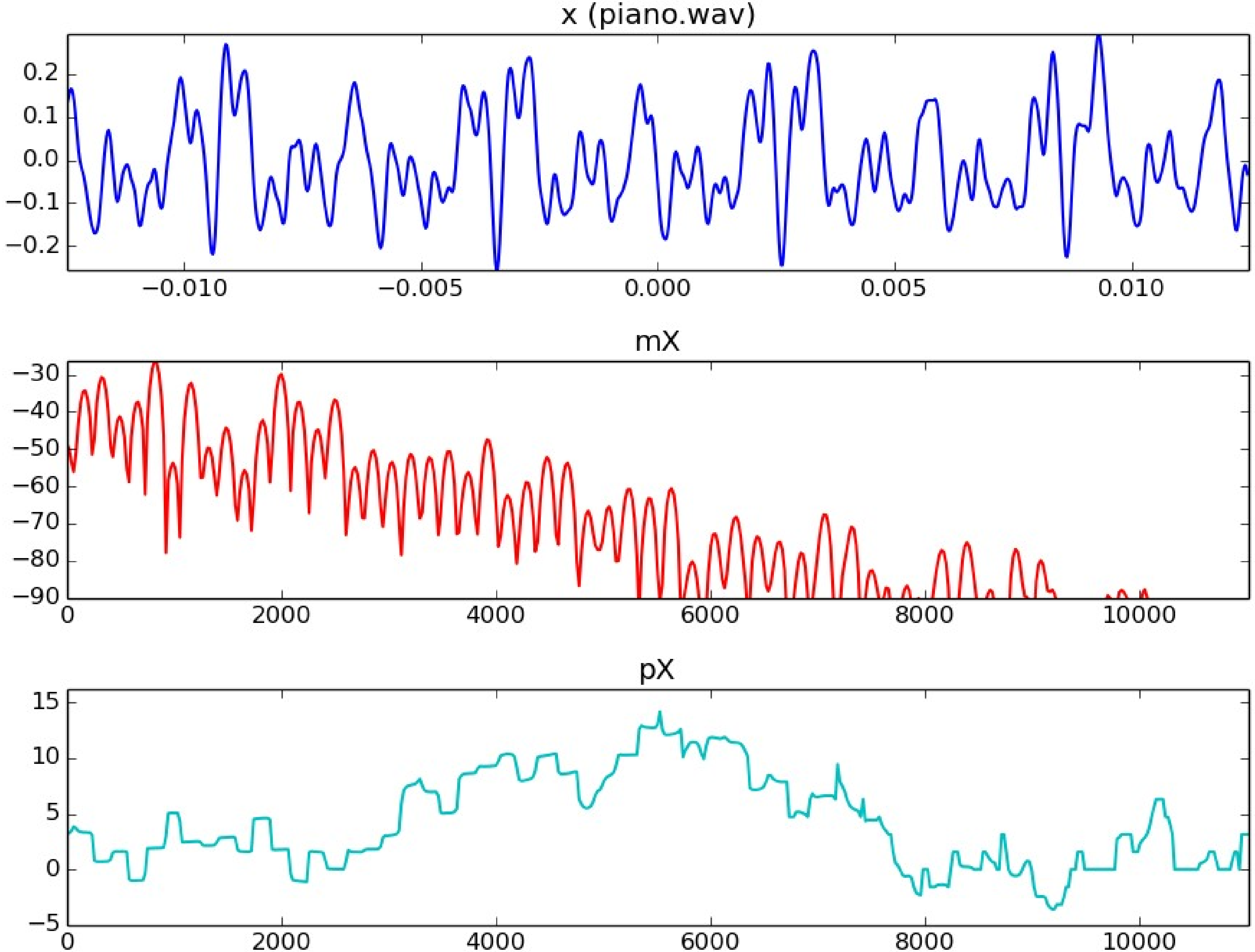
***Xavier Serra***

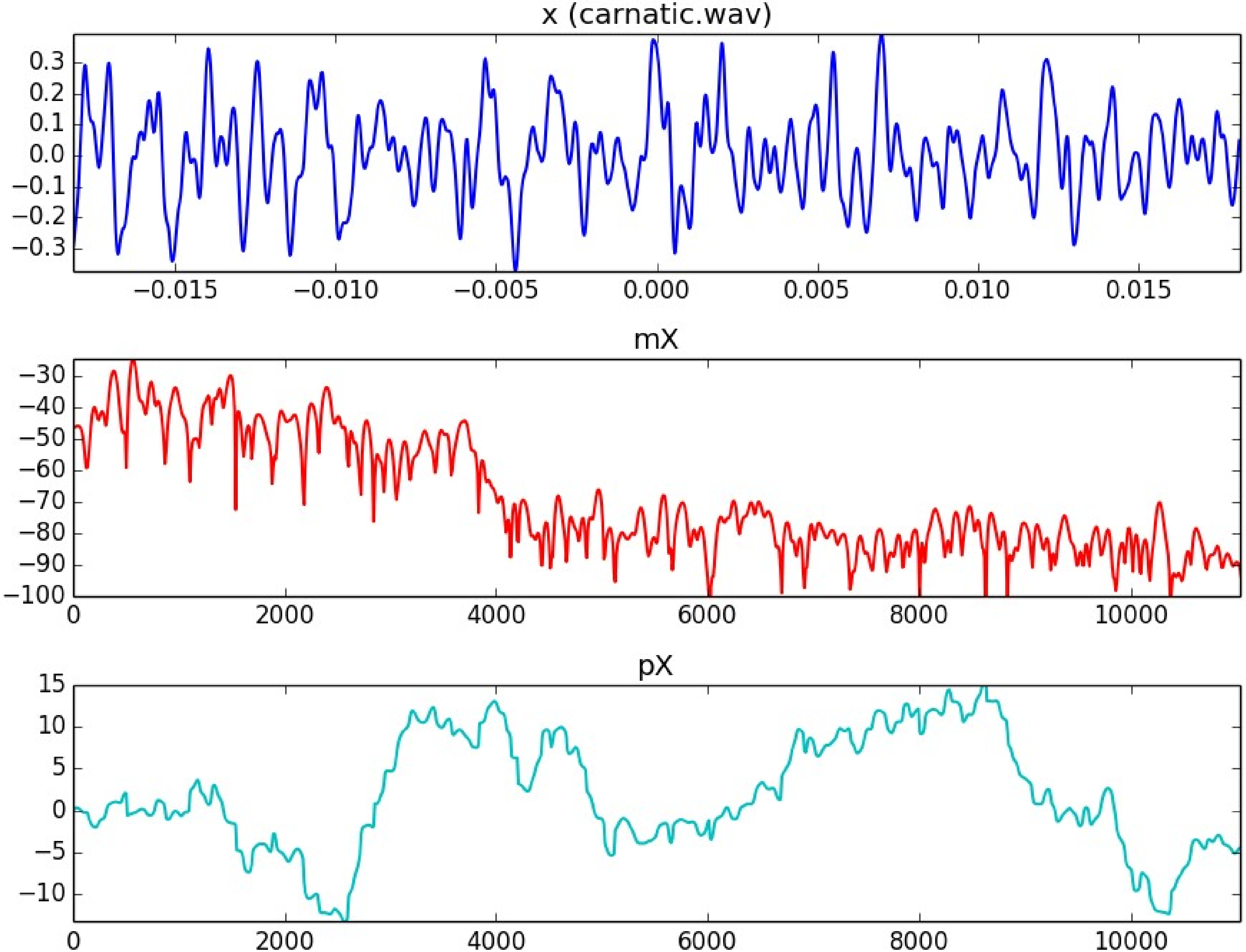
Universitat Pompeu Fabra, Barcelona

# Index

- F0 detection in time domain
  - monophonic signals
- F0 detection in frequency domain
  - monophonic and polyphonic signals


x (oboe-A4.wav)
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−0.006
−0.004
−0.002
0.000
0.002
0.004
0.006
mX
−20
−30
−40
−50
−60
−70
−80
−90
0
2000
4000
6000
8000
10000
12000
14000
pX
20
15
10
5
0
−5
0
2000
4000
6000
8000
10000
12000
14000


x (piano.wav)
0.2
0.1
0.0
−0.1
−0.2
−0.010
−0.005
0.000
0.005
0.010
mX
−30
−40
−50
−60
−70
−80
−90
0
2000
4000
6000
8000
10000
pX
15
10
5
0
−5
0
2000
4000
6000
8000
10000

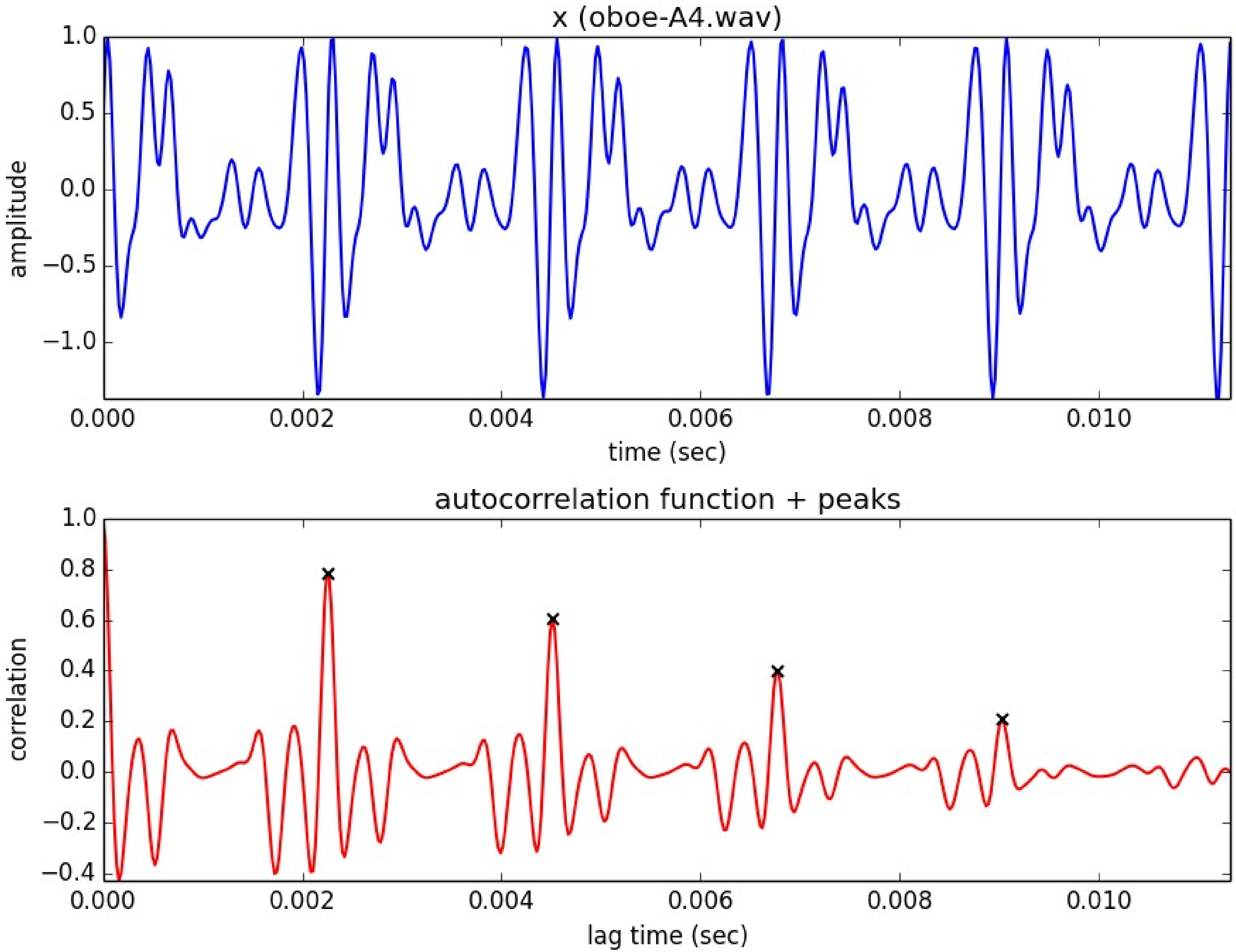

x (carnatic.wav)
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.015
−0.010
−0.005
0.000
0.005
0.010
0.015
mX
−30
−40
−50
−60
−70
−80
−90
−100
0
2000
4000
6000
8000
10000
pX
15
10
5
0
−5
−10
0
2000
4000
6000
8000
10000

# F0 detection in time domain

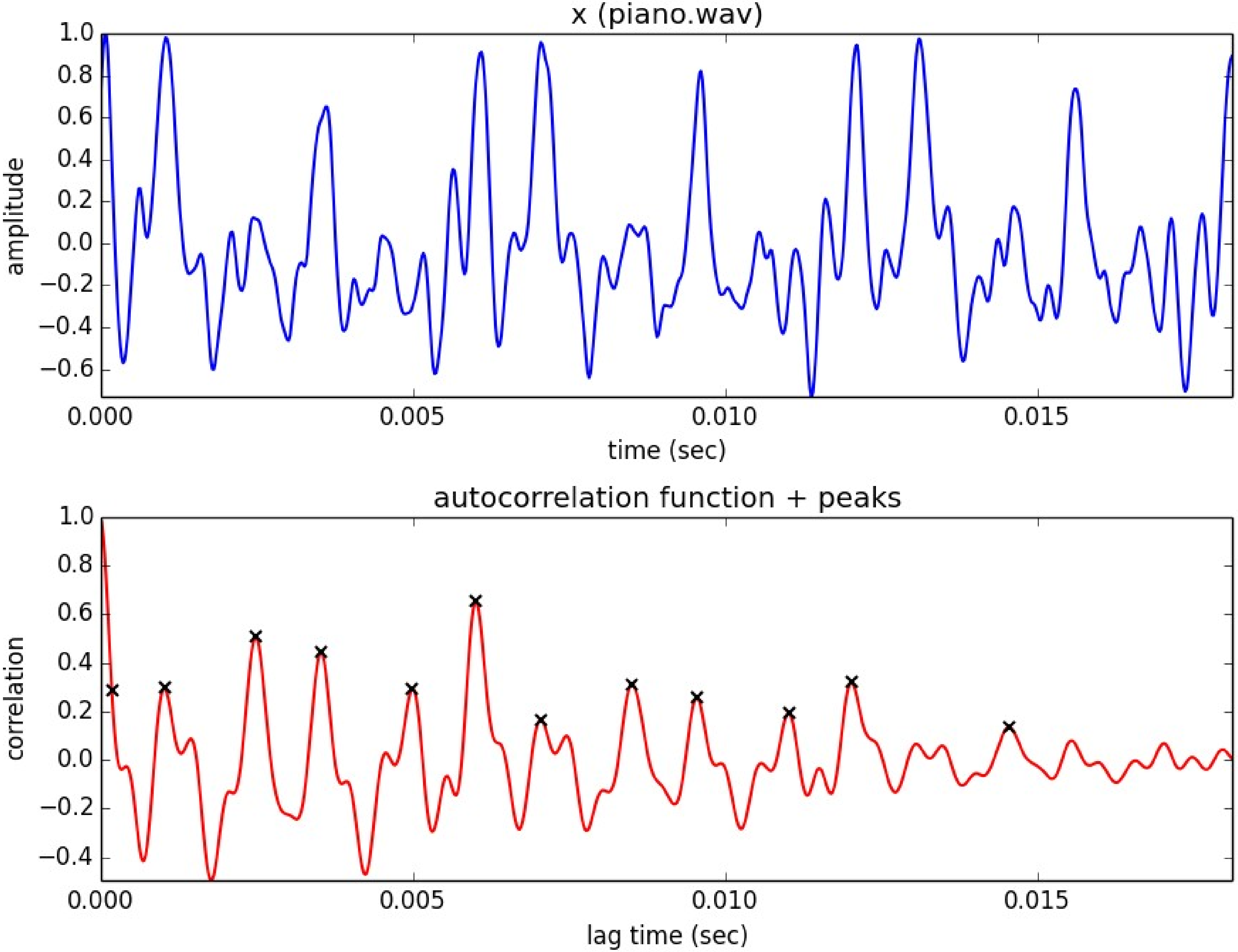
Autocorrelation function (with tapering)

$$r_x[l] = \sum_{n=0}^{n=N-1-l} x[n]\,x[n+l] \qquad l = 0, 1, \ldots, N-1$$

where $l = \text{lag}$



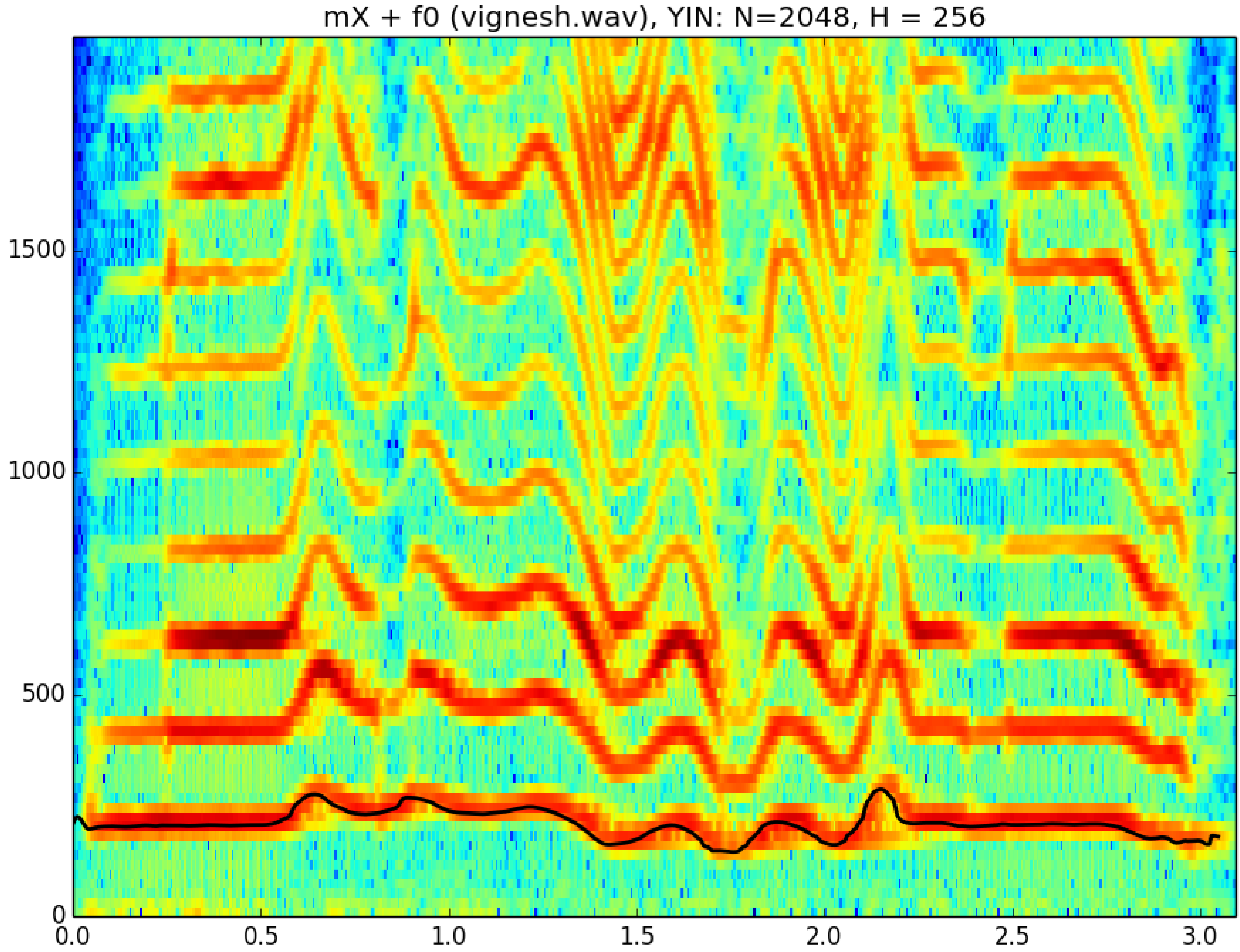
x (oboe-A4.wav)
amplitude
time (sec)
autocorrelation function + peaks
correlation
lag time (sec)


x (piano.wav)
amplitude
time (sec)
autocorrelation function + peaks
correlation
lag time (sec)

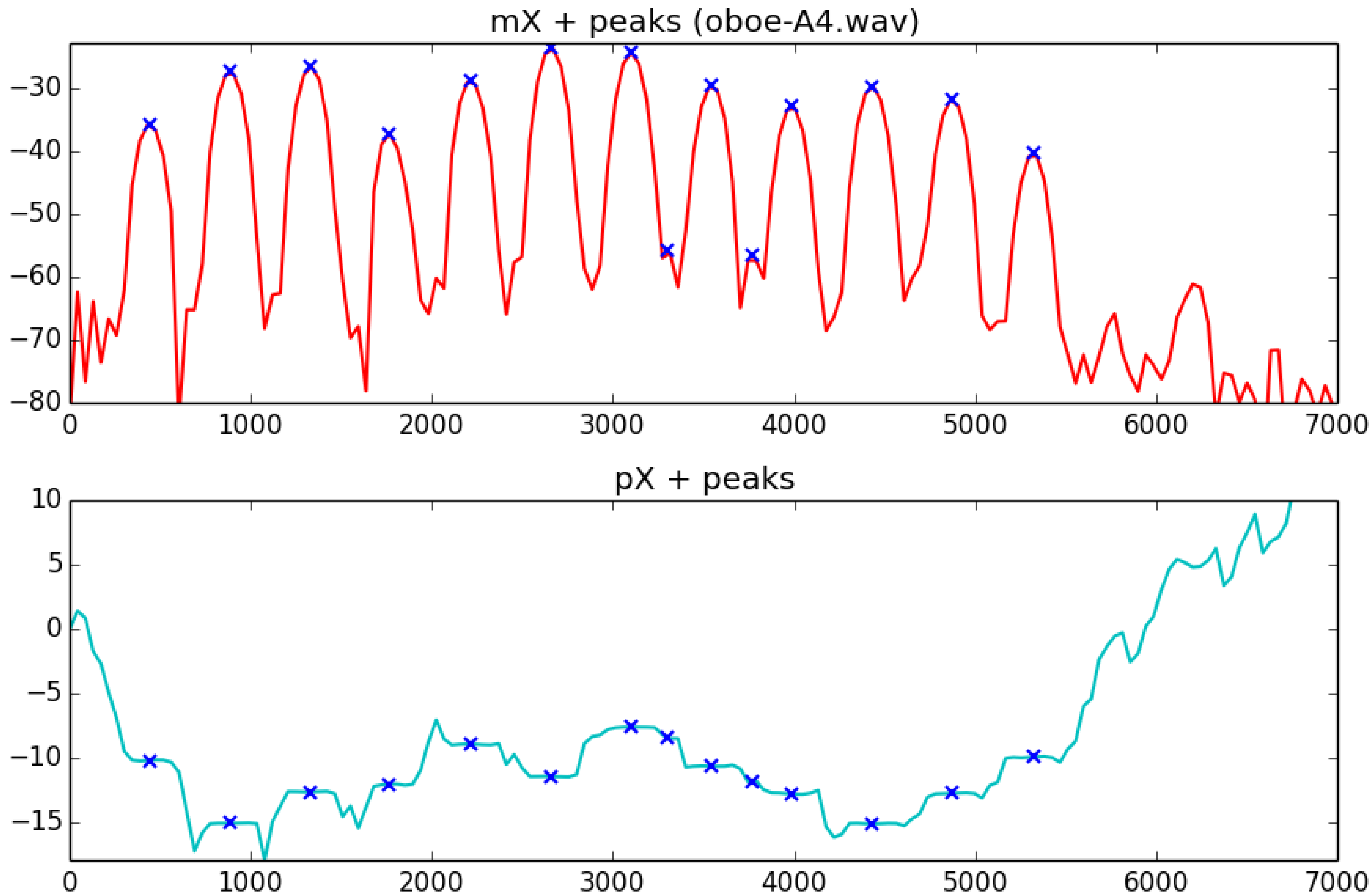
# YIN Algorithm (Cheveigné and Kawahara, 2002)

- Based on the difference function

$$d[l] = \sum_{n=0}^{n=N-1-l} \left( x[n] - x[n+l] \right)^2 \qquad l = 0, 1, \ldots, N-1$$

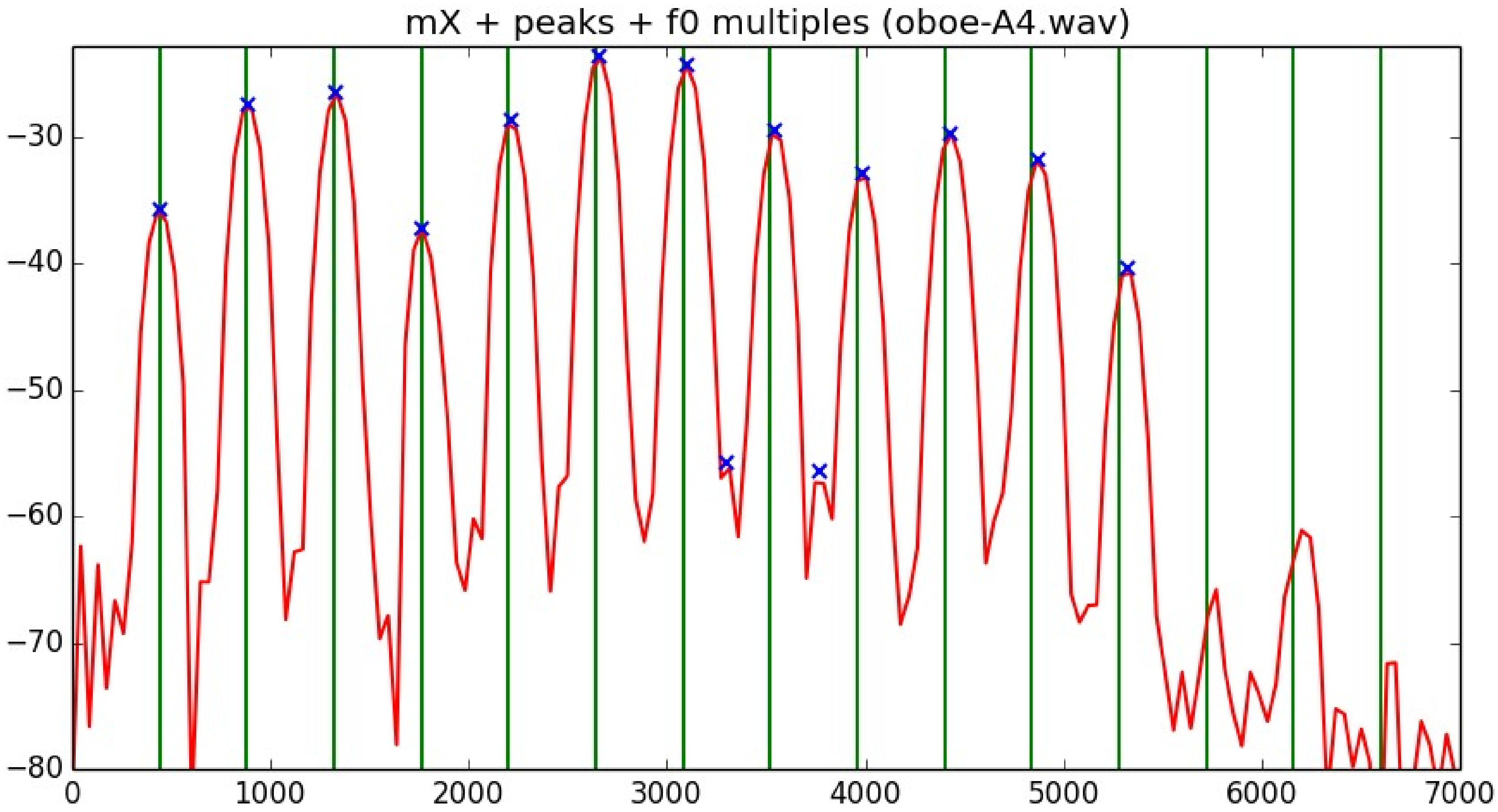

mX + f0 (vignesh.wav), YIN: N=2048, H = 256
0
500
1000
1500
0.0
0.5
1.0
1.5
2.0
2.5
3.0

# F0 detection in frequency domain



# F0 in the spectrum

The F0 can be defined as the common divisor of the harmonic series that best explains the spectral peaks.



# Pattern matching

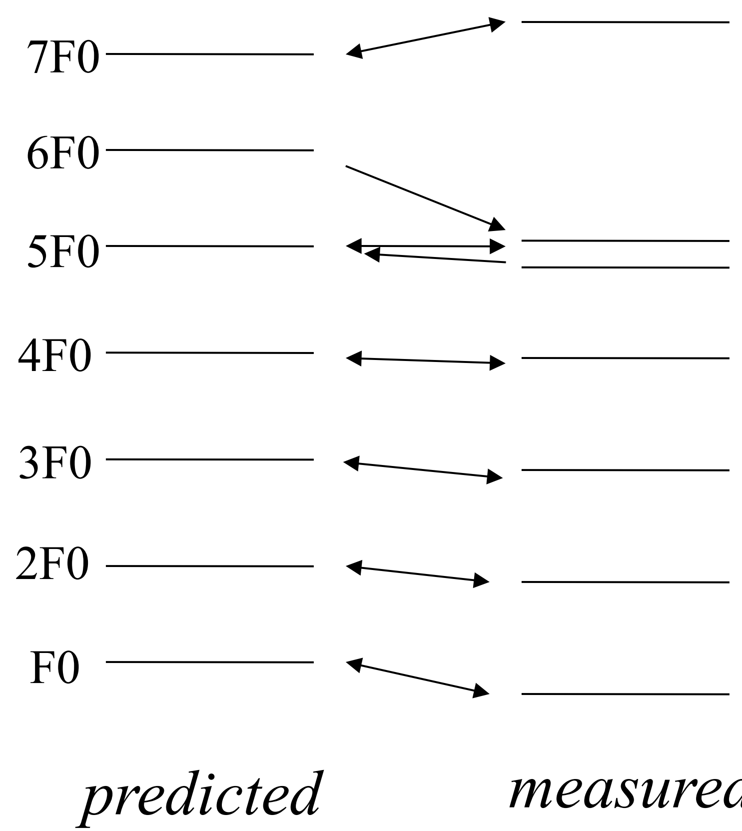
Two-way mismatch algorithm (Maher and Beauchamp, 1994)

$$\begin{aligned}\mathrm{Err}_{p\rightarrow m} &= \sum_{n=1}^{N} E_{\omega}(\Delta f_n, f_n, a_n, A_{\max}) \\ &= \sum_{n=1}^{N} \Delta f_n \cdot (f_n)^{-p} \\ &\quad + \left(\frac{a_n}{A_{\max}}\right) \times \left[ q \, \Delta f_n \cdot (f_n)^{-p} - r \right]\end{aligned}$$

$\Delta f_n$: diff. between predicted and the closest measured peaks

$f_n, a_n$: frequency and magnitude of predicted peaks

$A_{\max}$: maximum peak magnitude



$$\mathrm{Err}_{m\rightarrow p} = \sum_{k=1}^{K} E_{\omega}(\Delta f_k, f_k, a_k, A_{\max})$$
$$= \sum_{k=1}^{K} \Delta f_k \cdot (f_k)^{-p} + \left(\frac{a_k}{A_{\max}}\right) \times [q \Delta f_k \cdot (f_k)^{-p} - r]$$

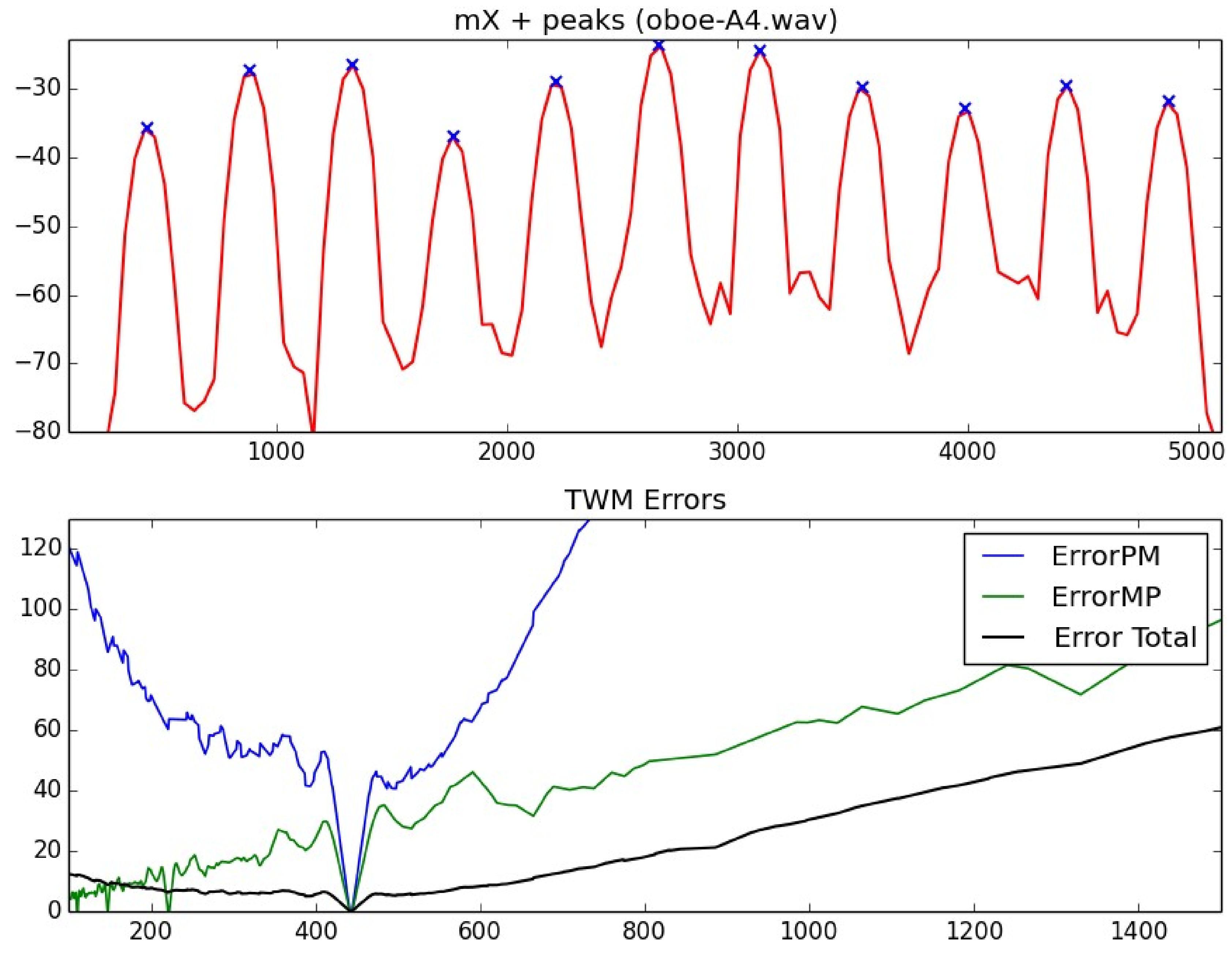
$\Delta f_k$ : diff. between predicted and its closest measured peaks
$f_k, a_k$ : frequency and magnitude of predicted peaks
$A_{\max}$ : maximum peak magnitude

**Total error:** $\mathrm{Err}_{\mathrm{total}} = \mathrm{Err}_{p\rightarrow m}/N + \rho\, \mathrm{Err}_{m\rightarrow p}/K$

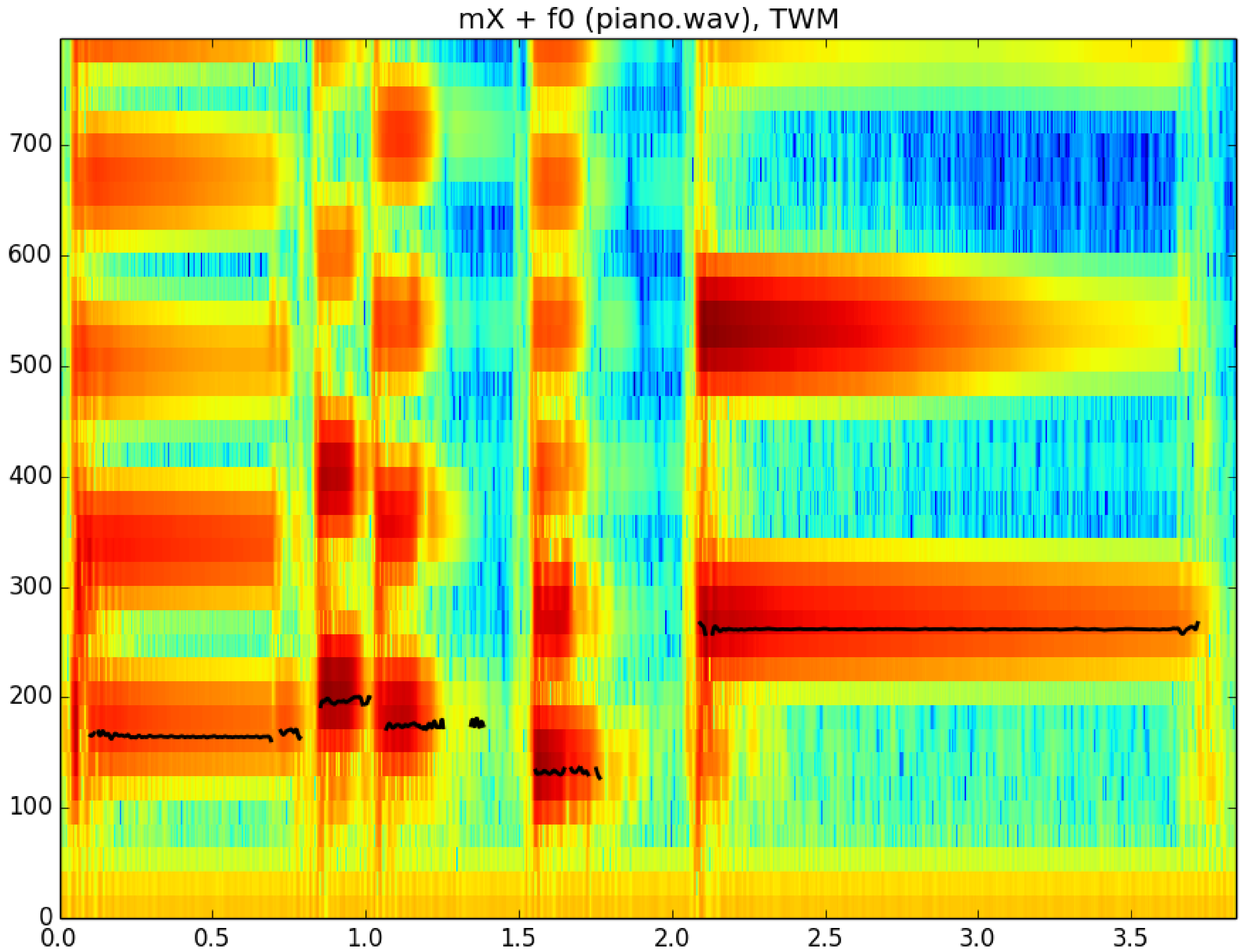
Maher and Beauchamp propose: $p = 0.5, q = 1.4, r = 0.5, \rho = 0.33$

| | $Err_{p->m}$ | $Err_{m->p}$ | $Err_{total}$ |
|---|---|---|---|
| 50Hz | 122.58 | -3.0 | 7.49 |
| 100Hz | 32.0 | -3.0 | 3.83 |
| 200Hz | 10.0 | 30.66 | 4.2 |

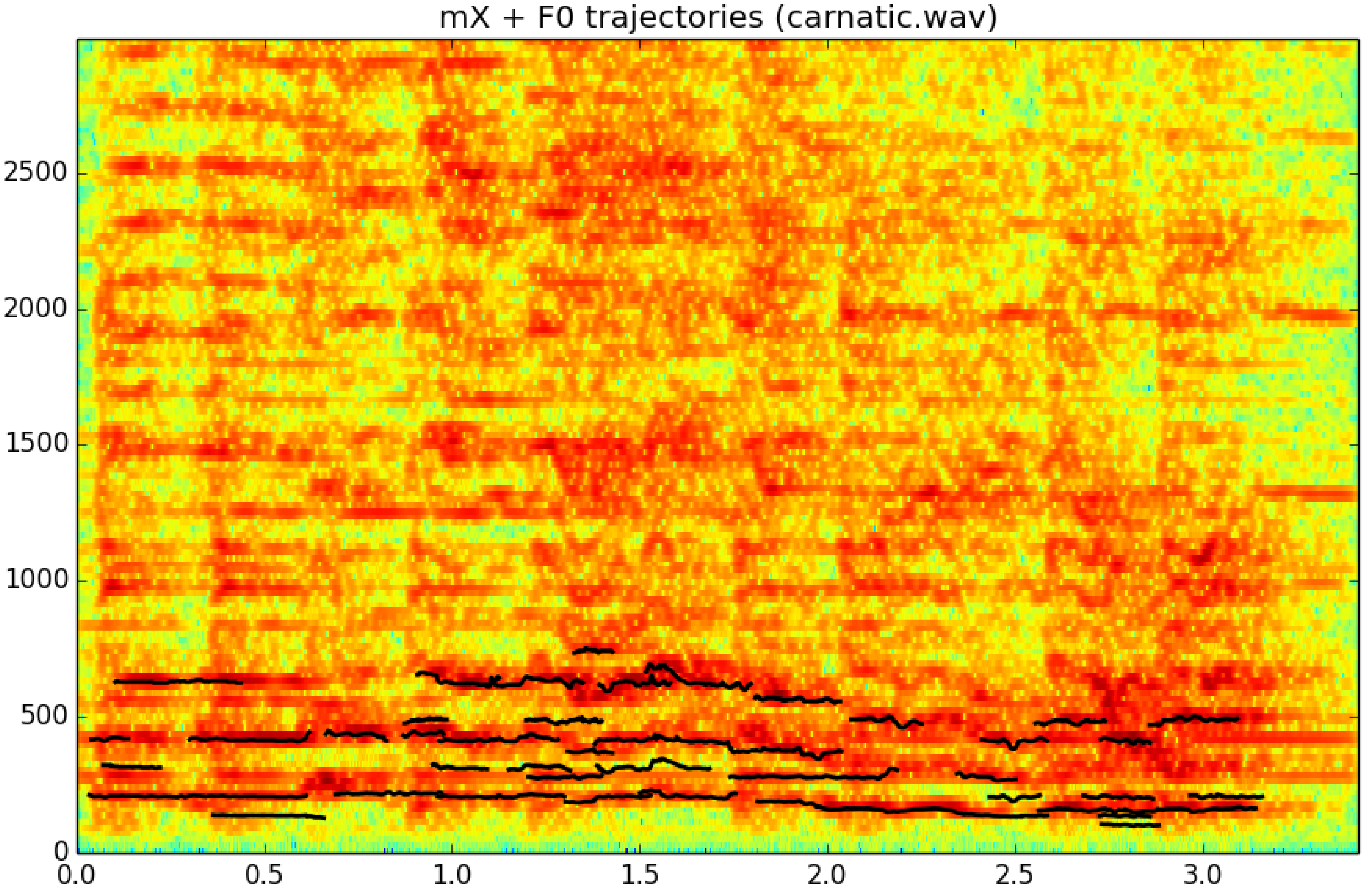
TWM error calculation from the frequencies: 200, 300, 500, 600, 700, 800.



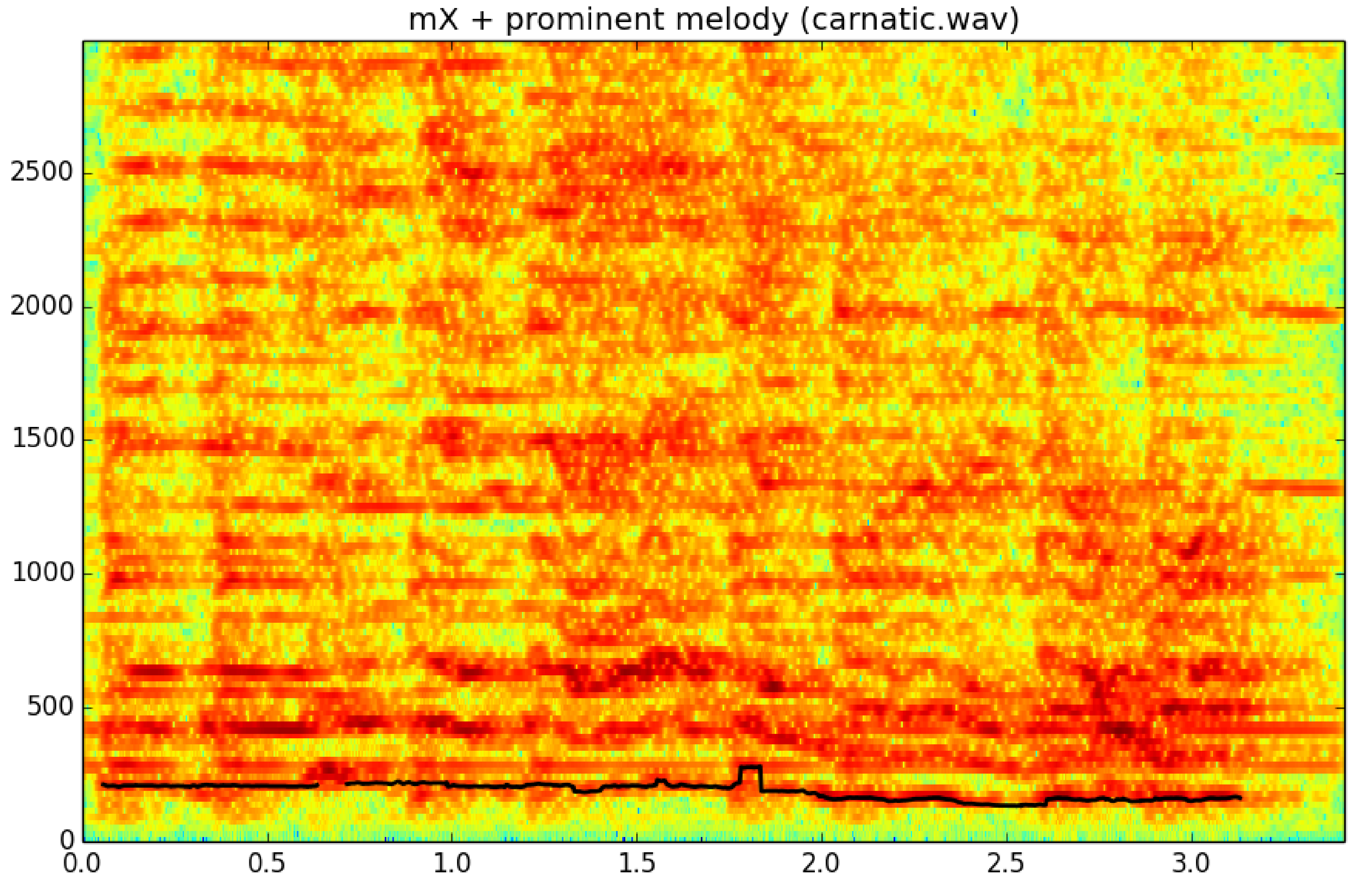
mX + peaks (oboe-A4.wav)
−30
−40
−50
−60
−70
−80
1000
2000
3000
4000
5000
TWM Errors
120
100
80
60
40
20
0
200
400
600
800
1000
1200
1400
ErrorPM
ErrorMP
Error Total


mX + f0 (piano.wav), TWM
0
100
200
300
400
500
600
700
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5

# F0 in polyphonic signals



# Prominent pitch in polyphonic signals
(Salamon and Gómez, 2012)



# References and credits

- More information in:
  http://en.wikipedia.org/wiki/Fundamental_frequency
  http://en.wikipedia.org/wiki/Pitch_detection_algorithm
  http://en.wikipedia.org/wiki/Autocorrelation
- F0 detection algorithms:

  - A. de Cheveigné and H. Kawahara. “YIN, a fundamental frequency estimator for speech and music,” J. Acoust. Soc. Am. 111, 1917 (2002).
  - R. C. Maher and J. W. Beauchamp, “Fundamental frequency estimation of musical signals using a Two-Way Mismatch procedure,” J. Acoust. Soc. Am., vol. 95., no. 4, pp. 2254-2263 (1994).
  - J. Salamon and E. Gómez, "Melody extraction from polyphonic music signals using pitch contour characteristics," IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 6, pp. 1759–1770 (2012).

- Sounds from:  http://www.freesound.org/people/xserra/packs/13038/

# **6T2:** Fundamental frequency detection

***Xavier Serra***
Universitat Pompeu Fabra, Barcelona