

# 6T2: Fundamental frequency detection

***Xavier Serra***

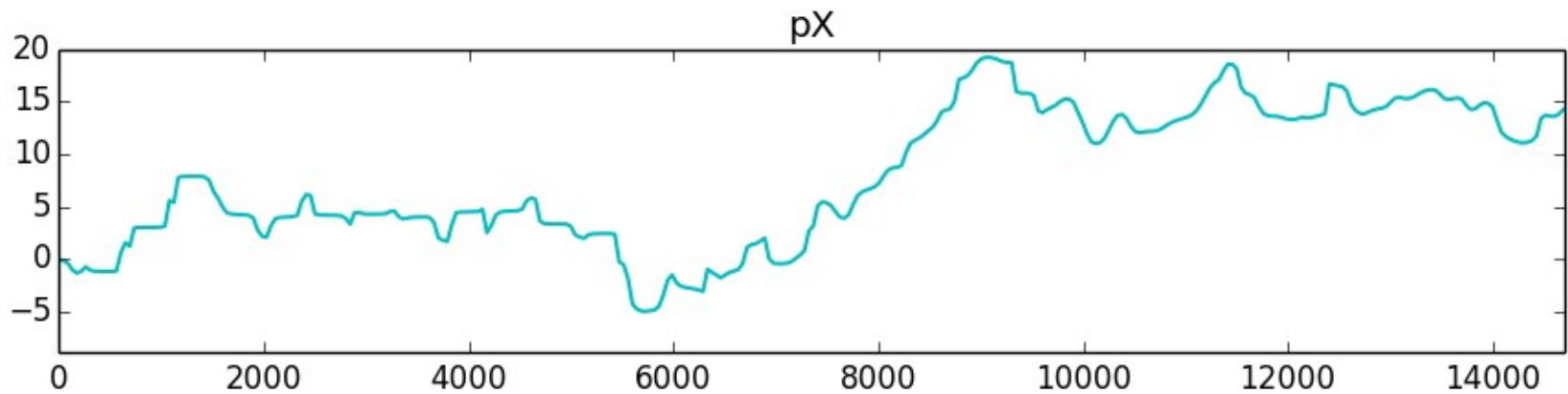
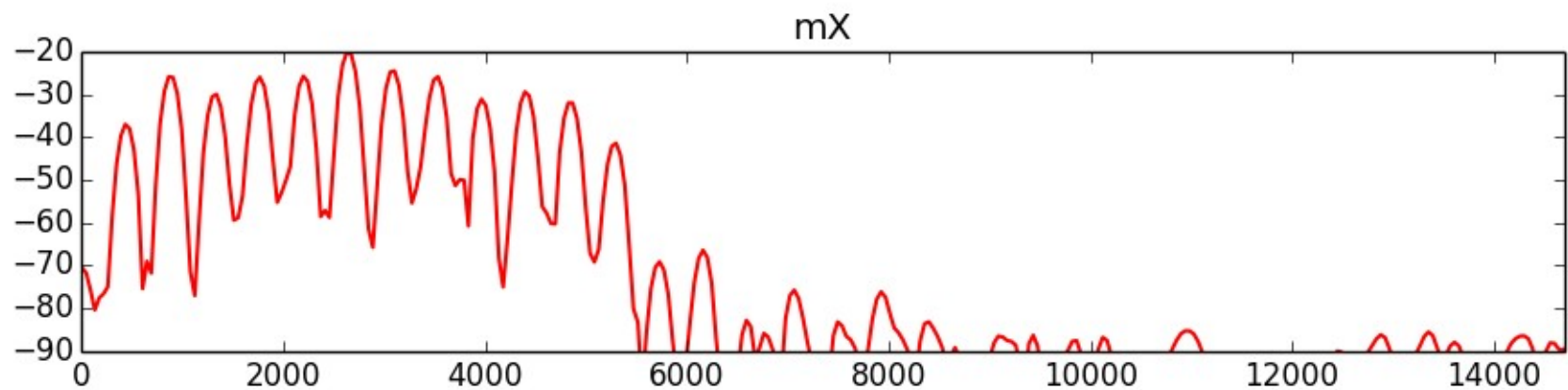
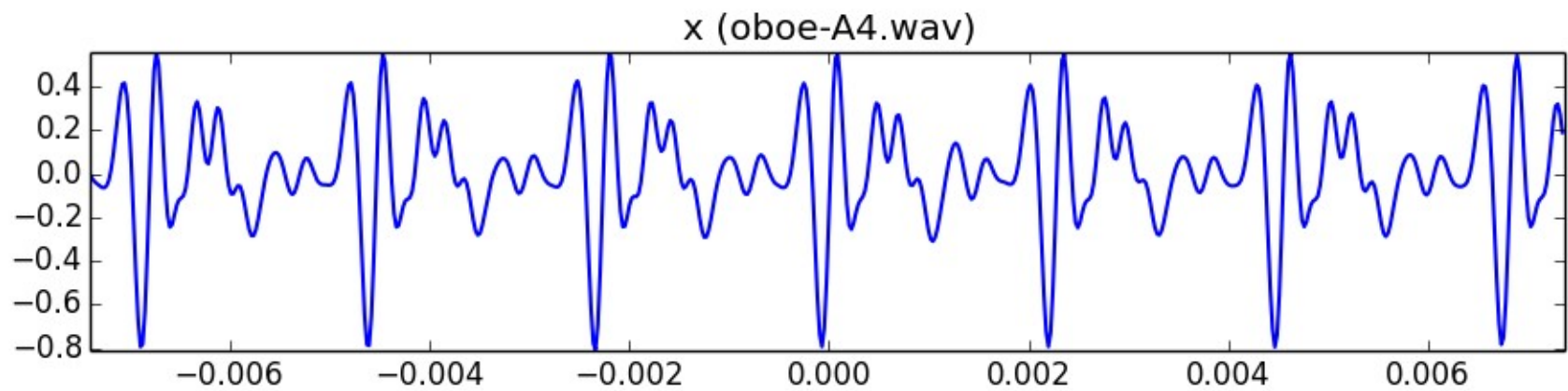
Universitat Pompeu Fabra, Barcelona

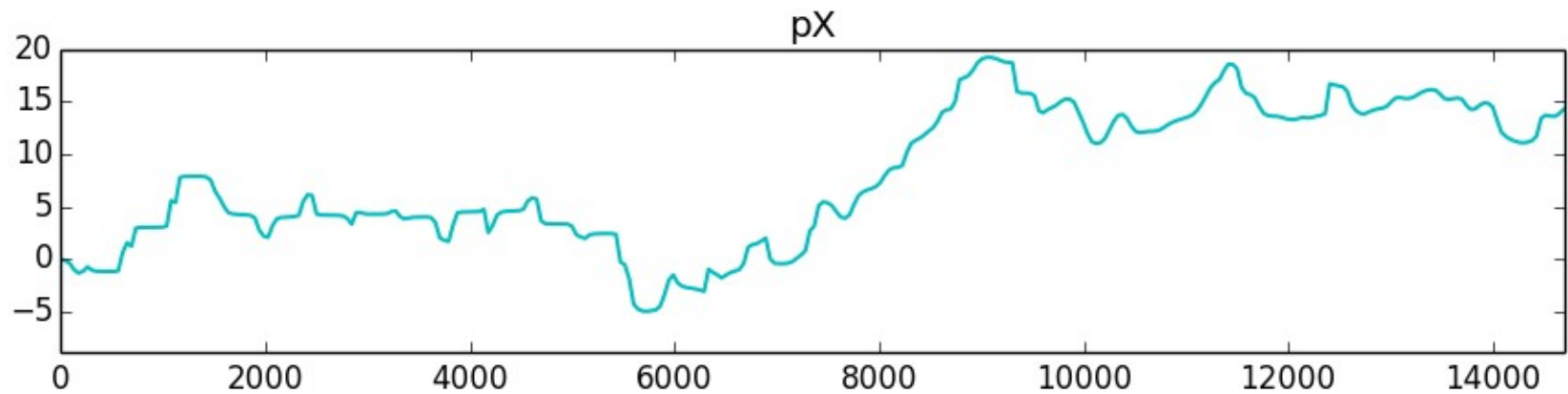
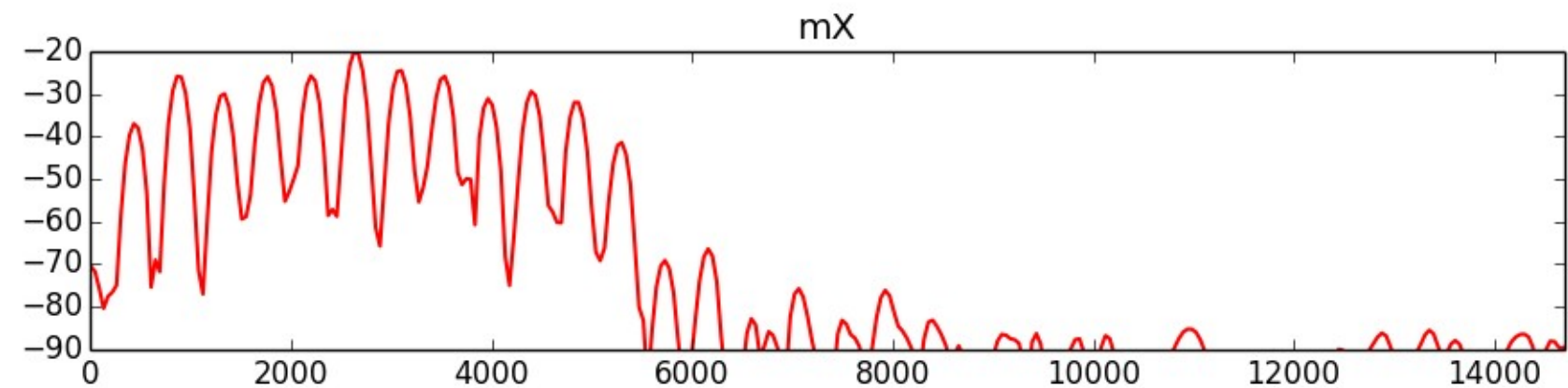
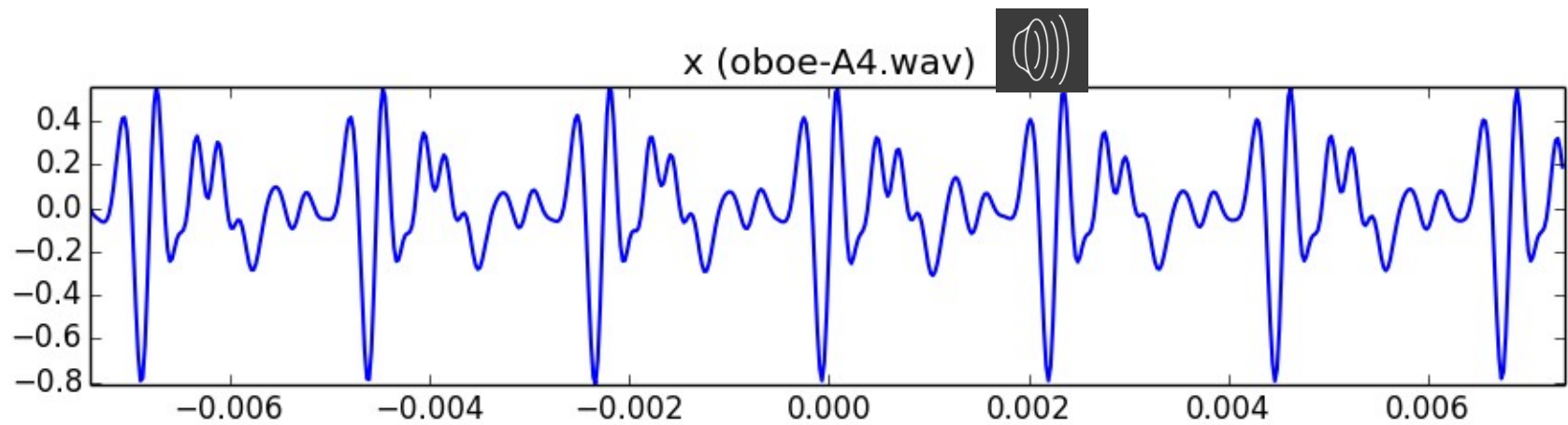
&

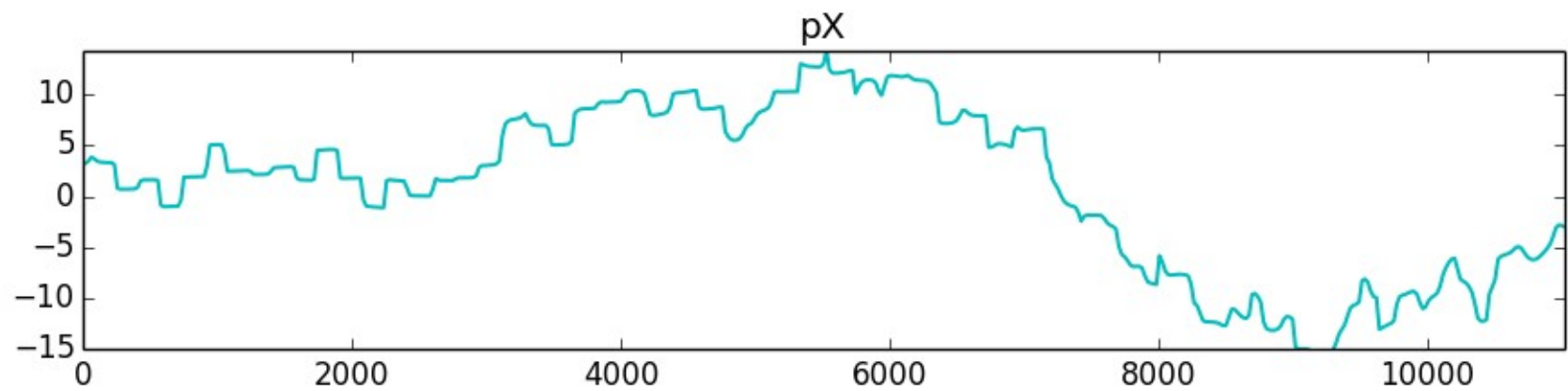
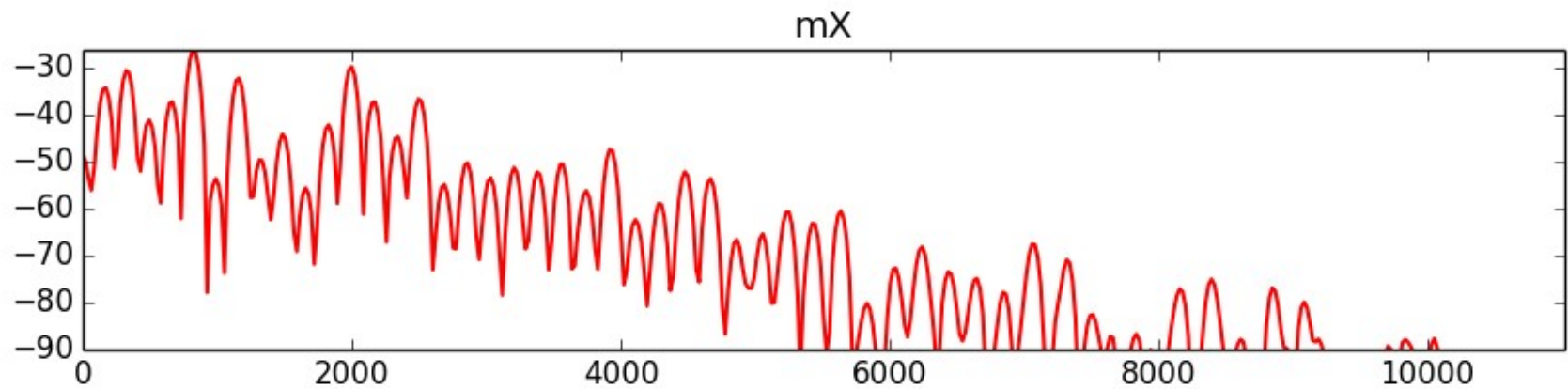
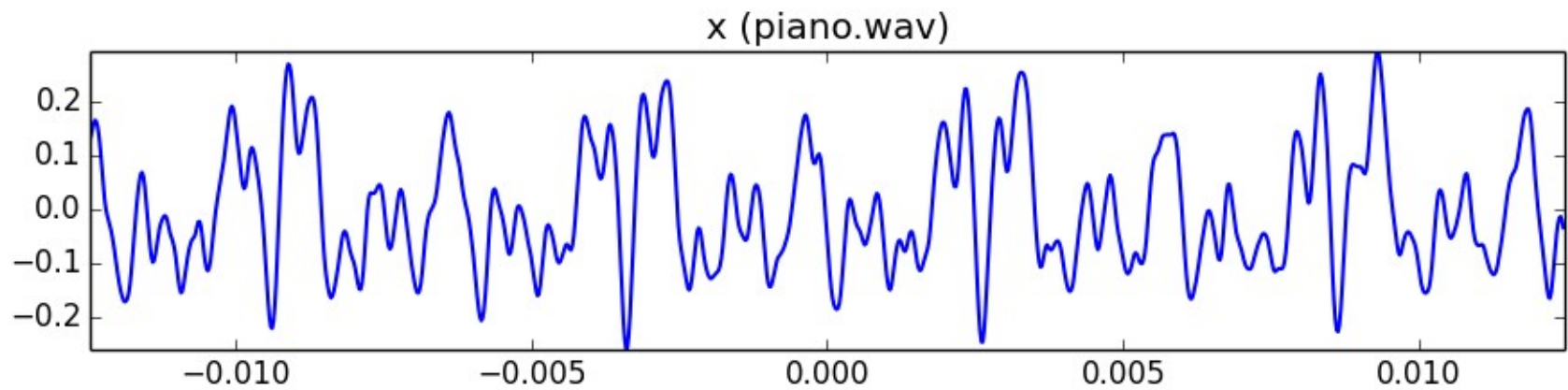
Stanford University

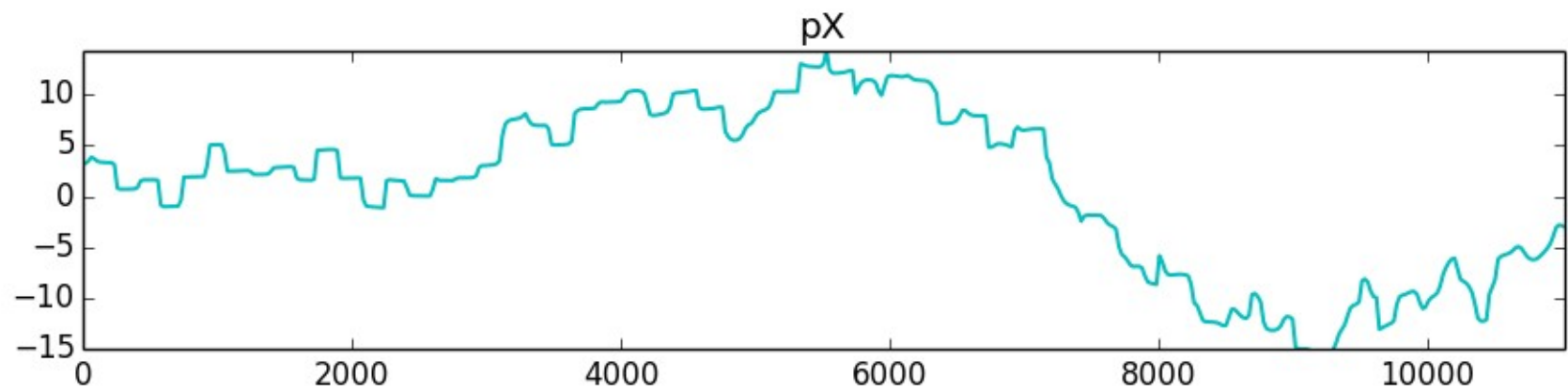
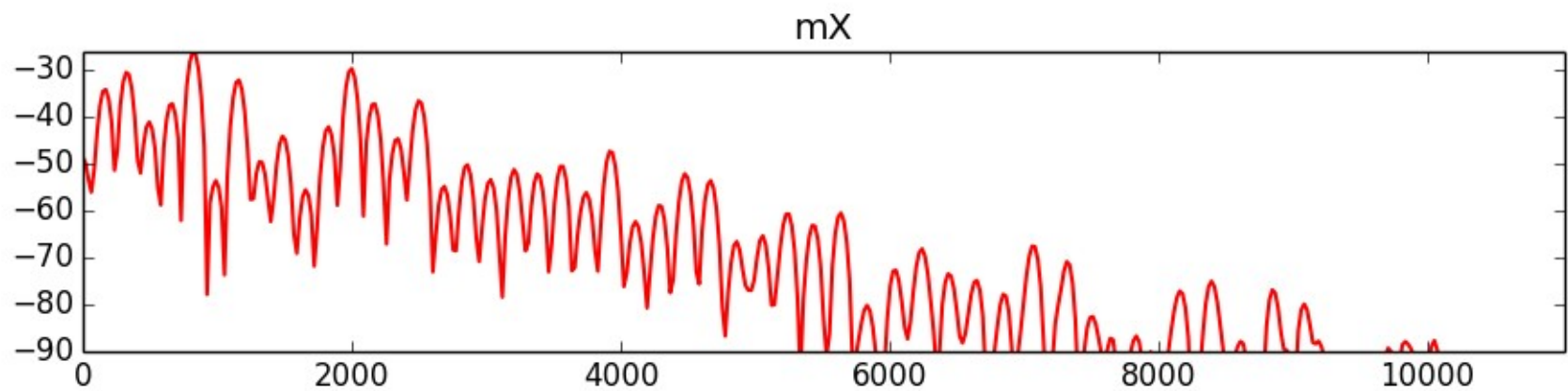
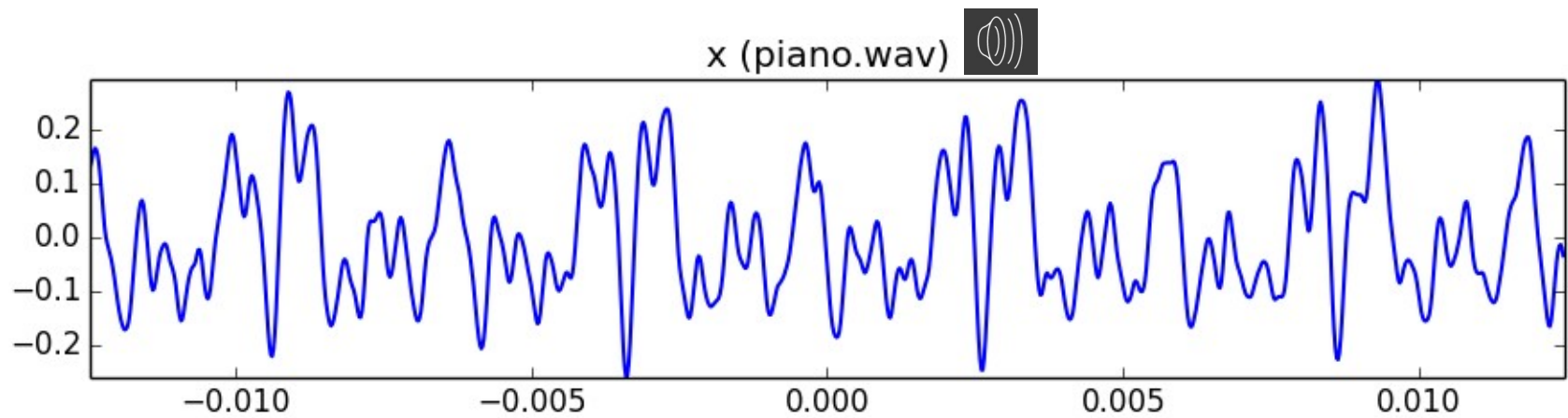
# Index

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

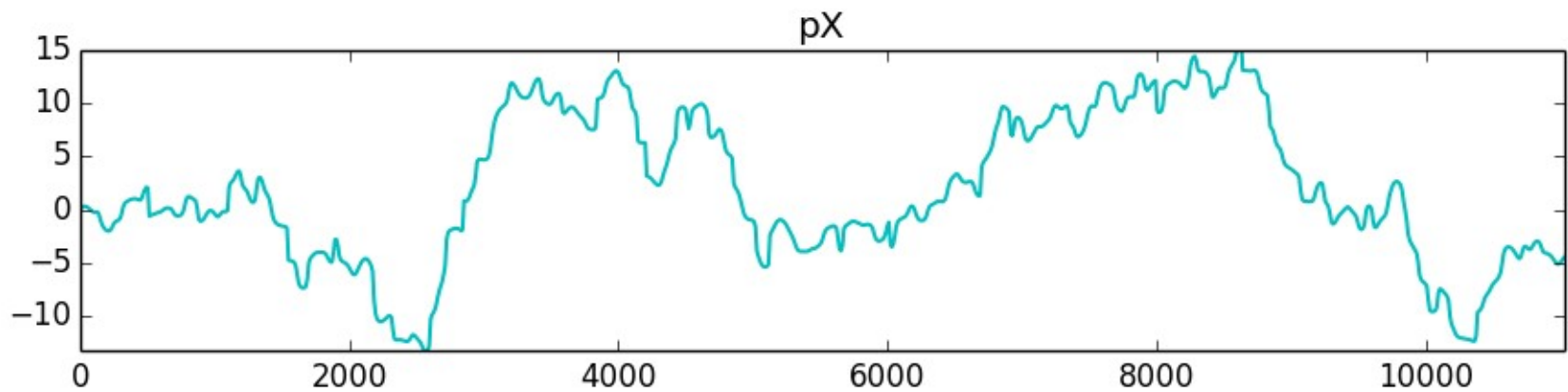
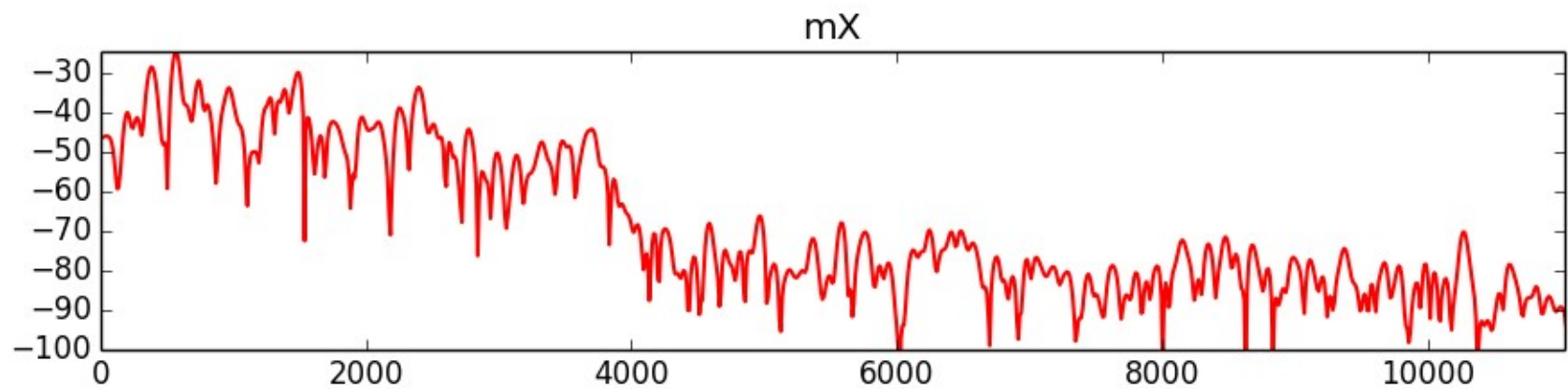
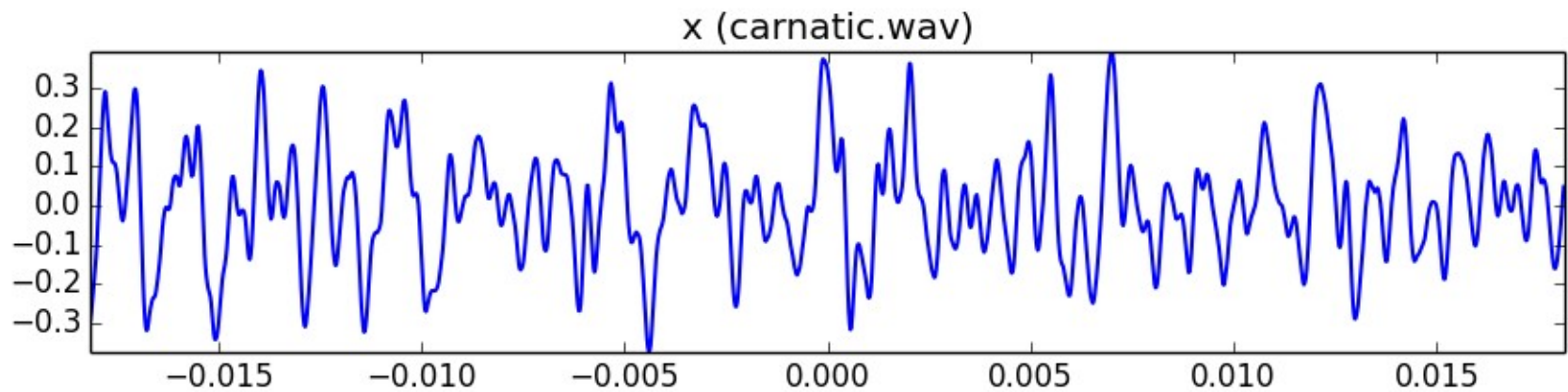






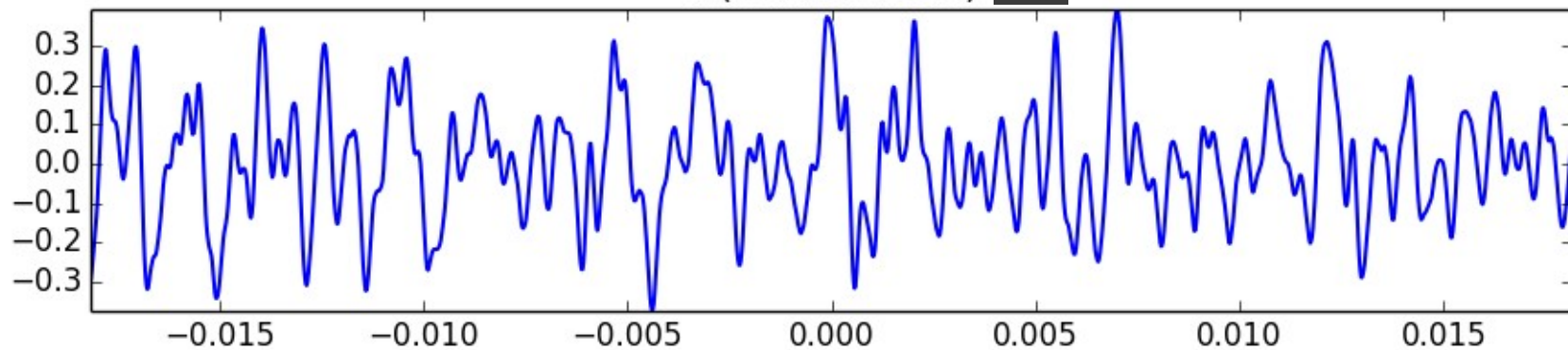




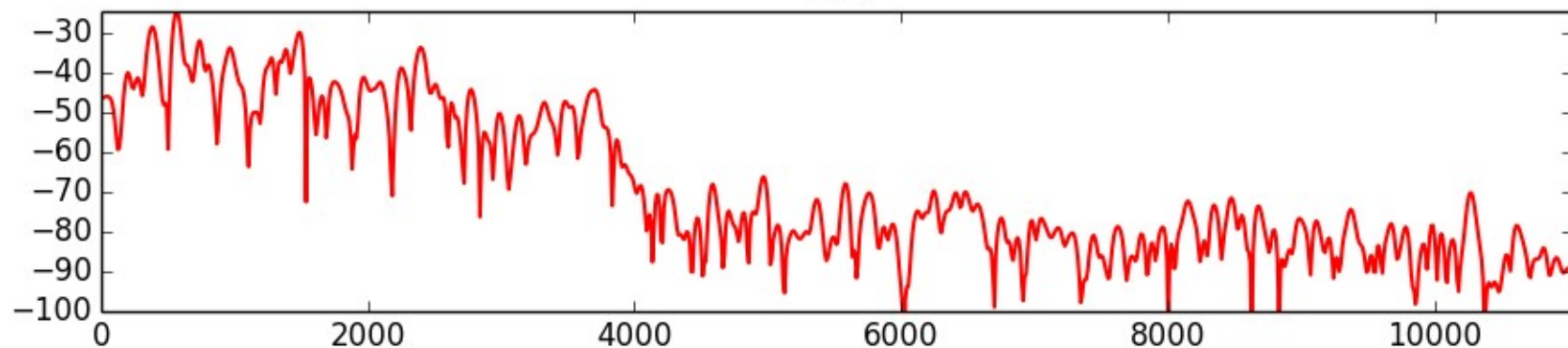




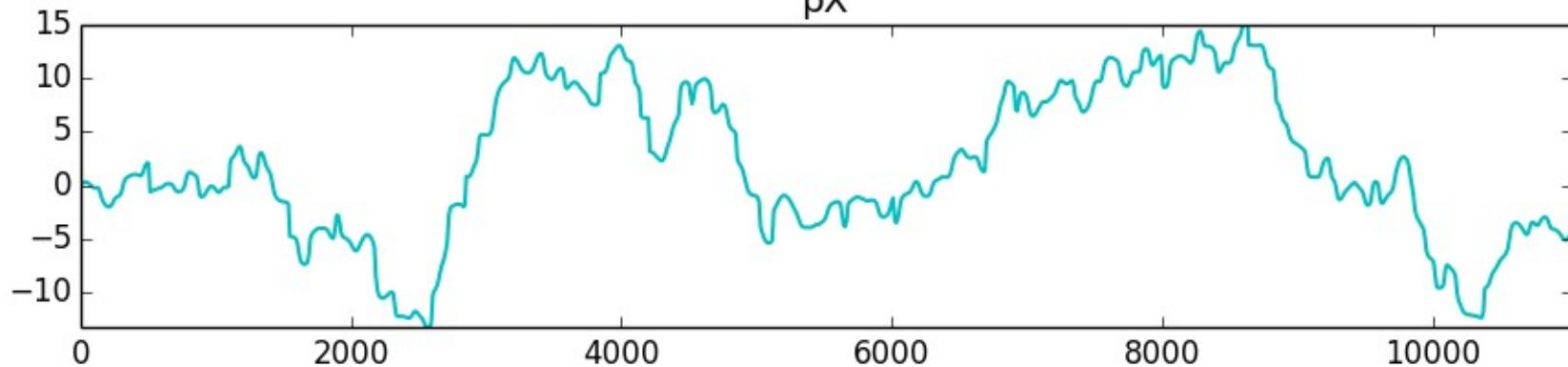
x (carnatic.wav)



mX



pX





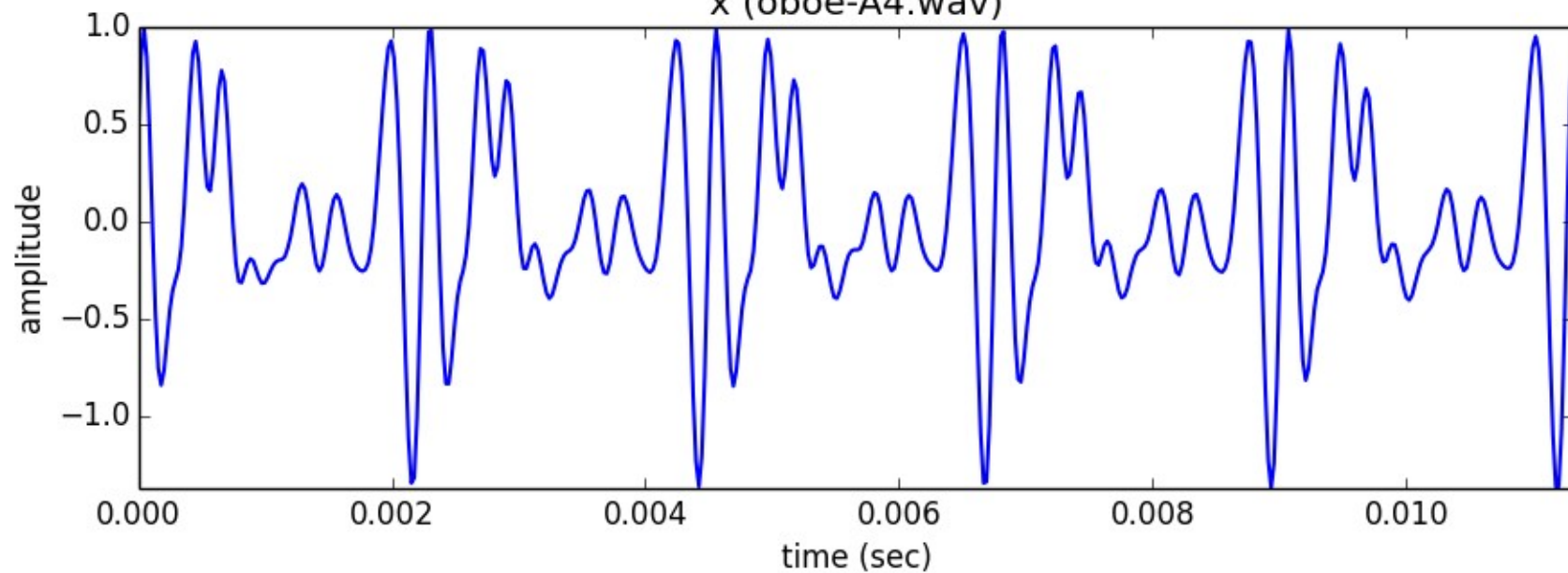
# F0 detection in time domain

Autocorrelation function (with tapering)

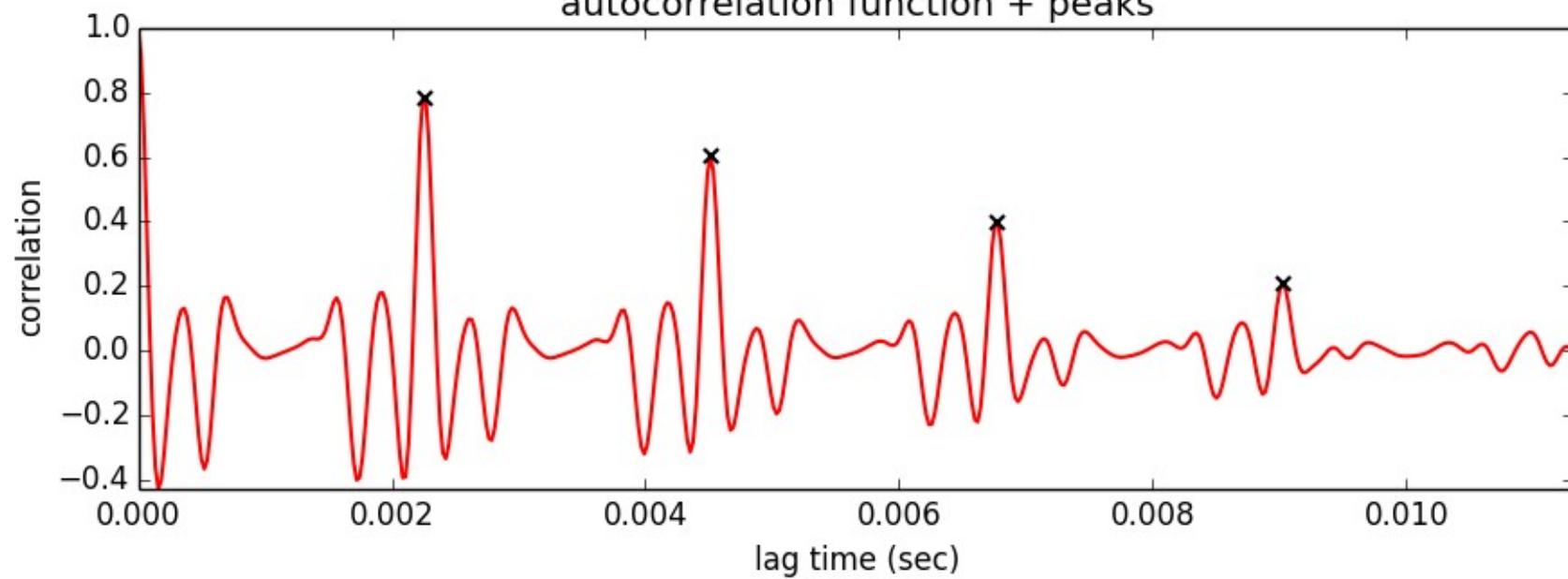
$$r_x[l] = \sum_{n=0}^{N-1-l} x[n]x[n+l] \quad l=0,1,\dots,N-1$$

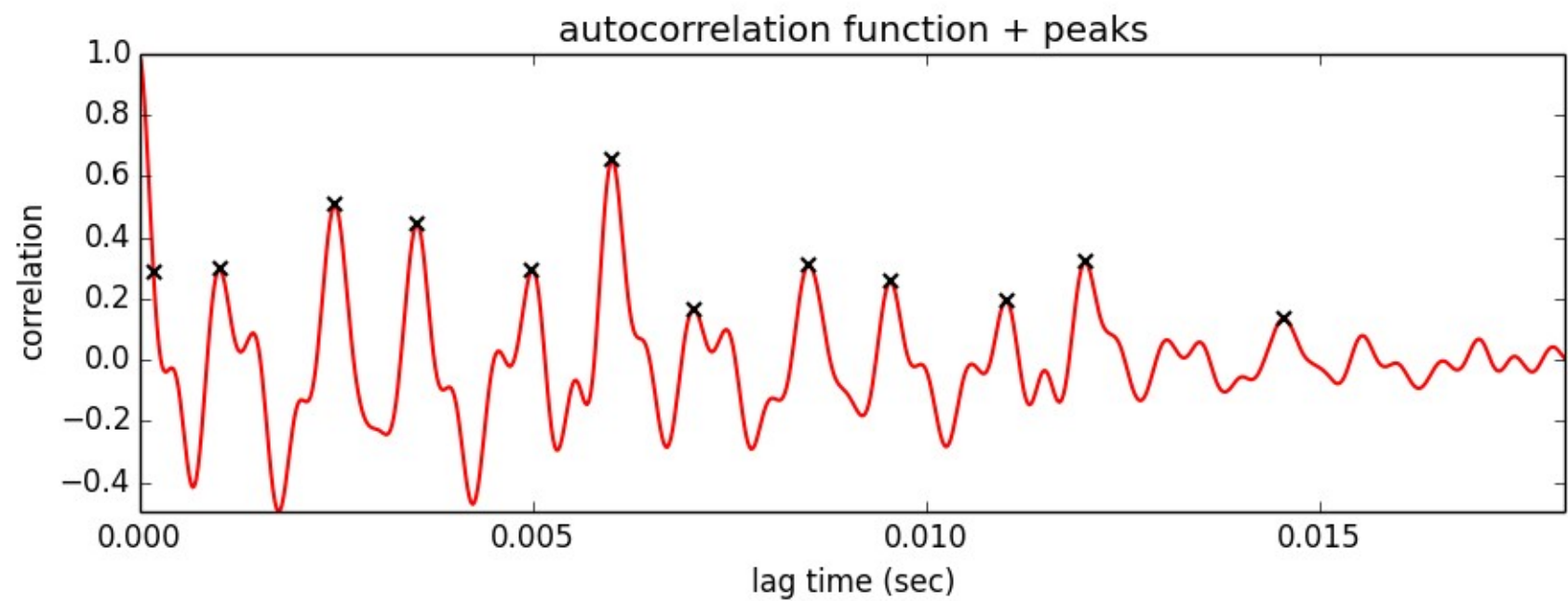
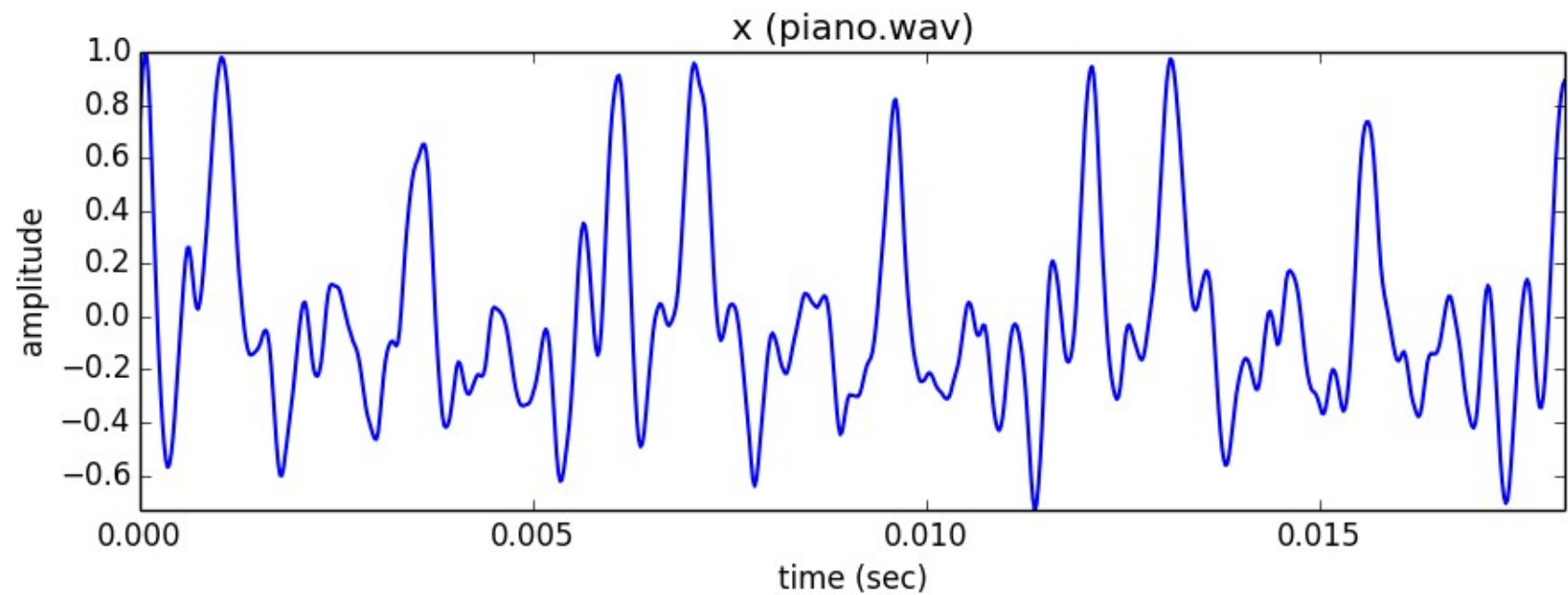
where  $l = \text{lag}$

x (oboe-A4.wav)



autocorrelation function + peaks



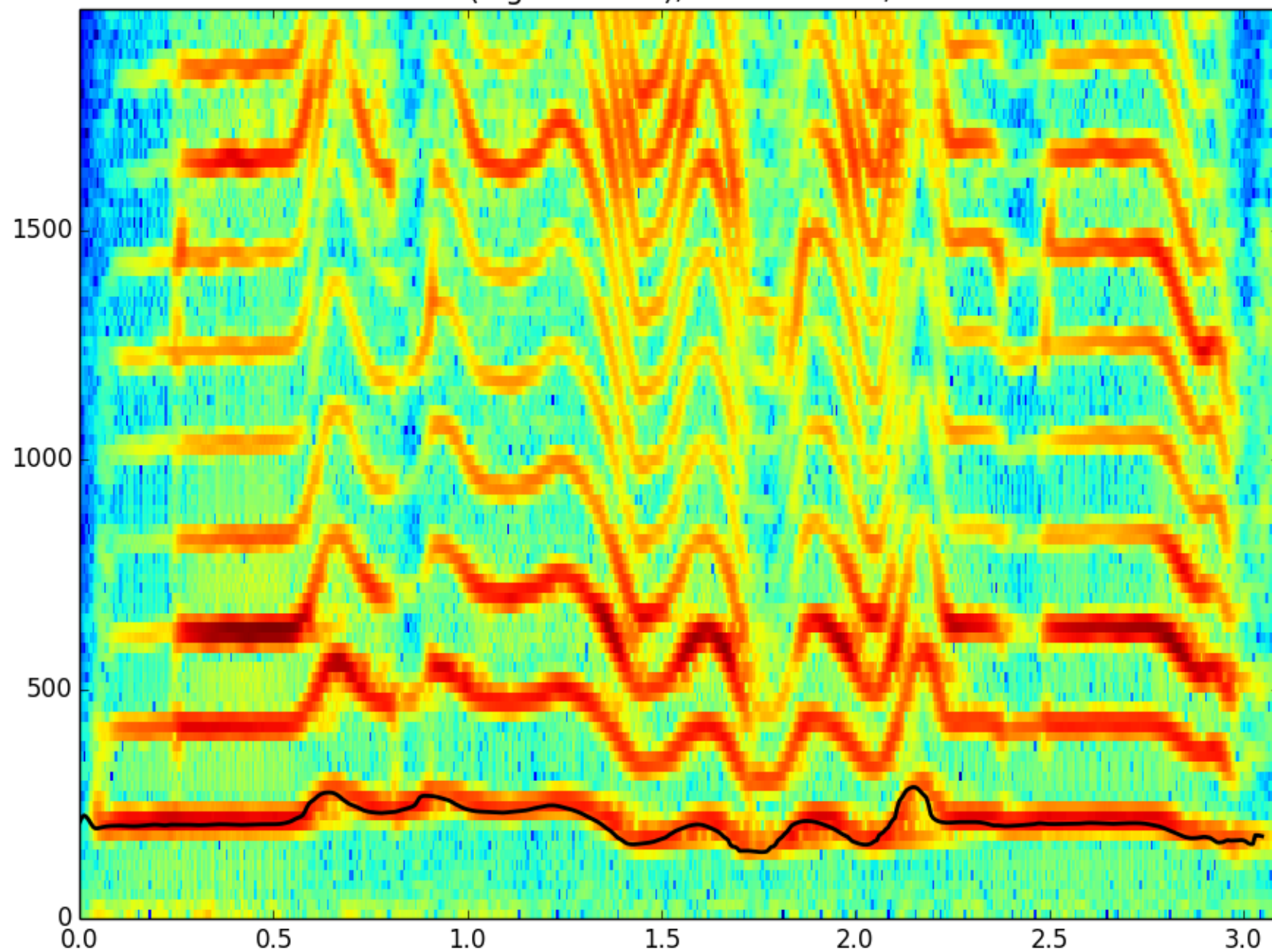


# YIN Algorithm (Cheveigné and Kawahara, 2002)

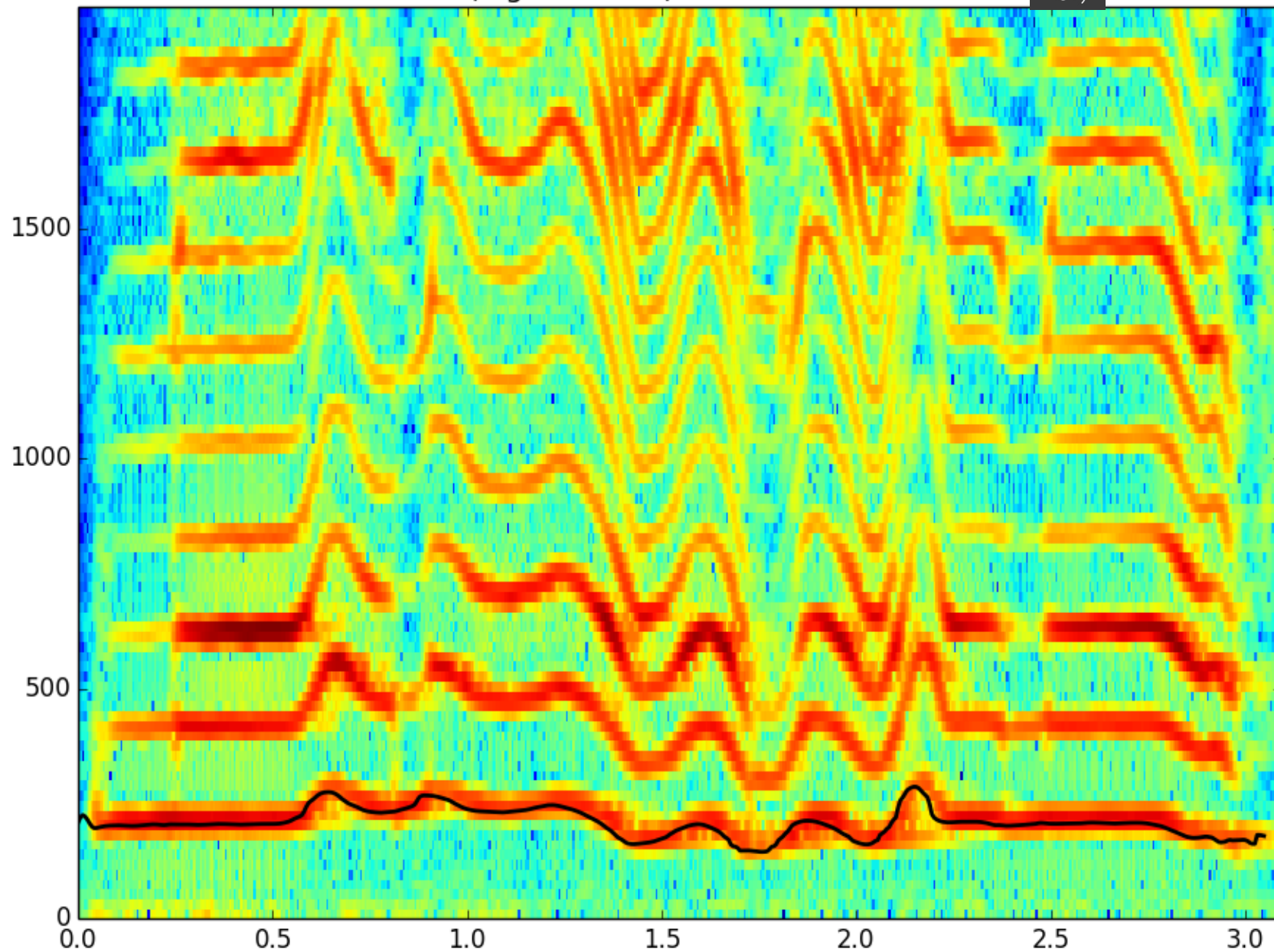
- Based on the difference function

$$d[l] = \sum_{n=0}^{n=N-1-l} (x[n] - x[n+l])^2 \quad l=0,1,\dots,N-1$$

mX + f0 (vignesh.wav), YIN: N=2048, H = 256

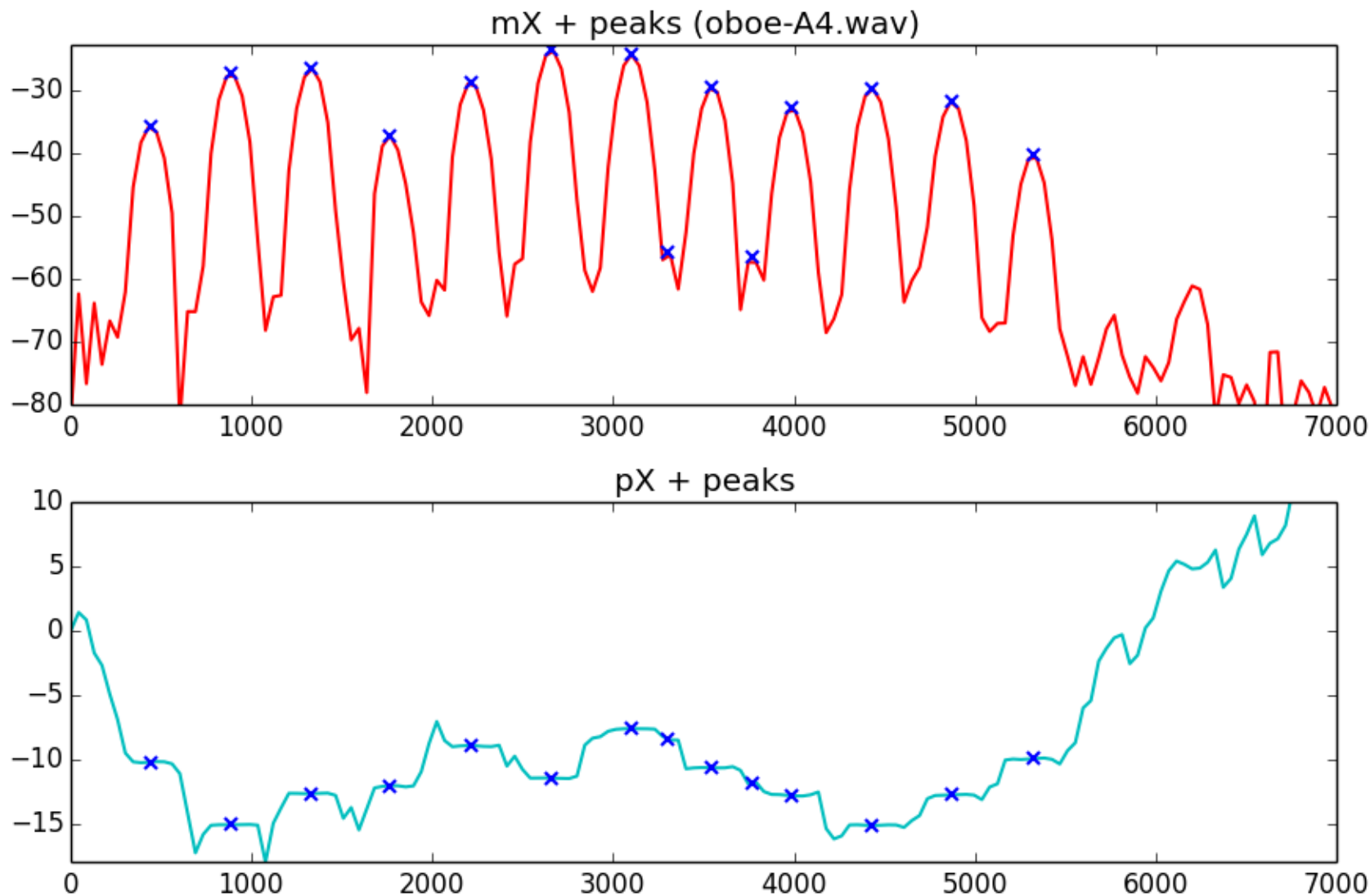


mX + f0 (vignesh.wav), YIN: N=2048, H = 256



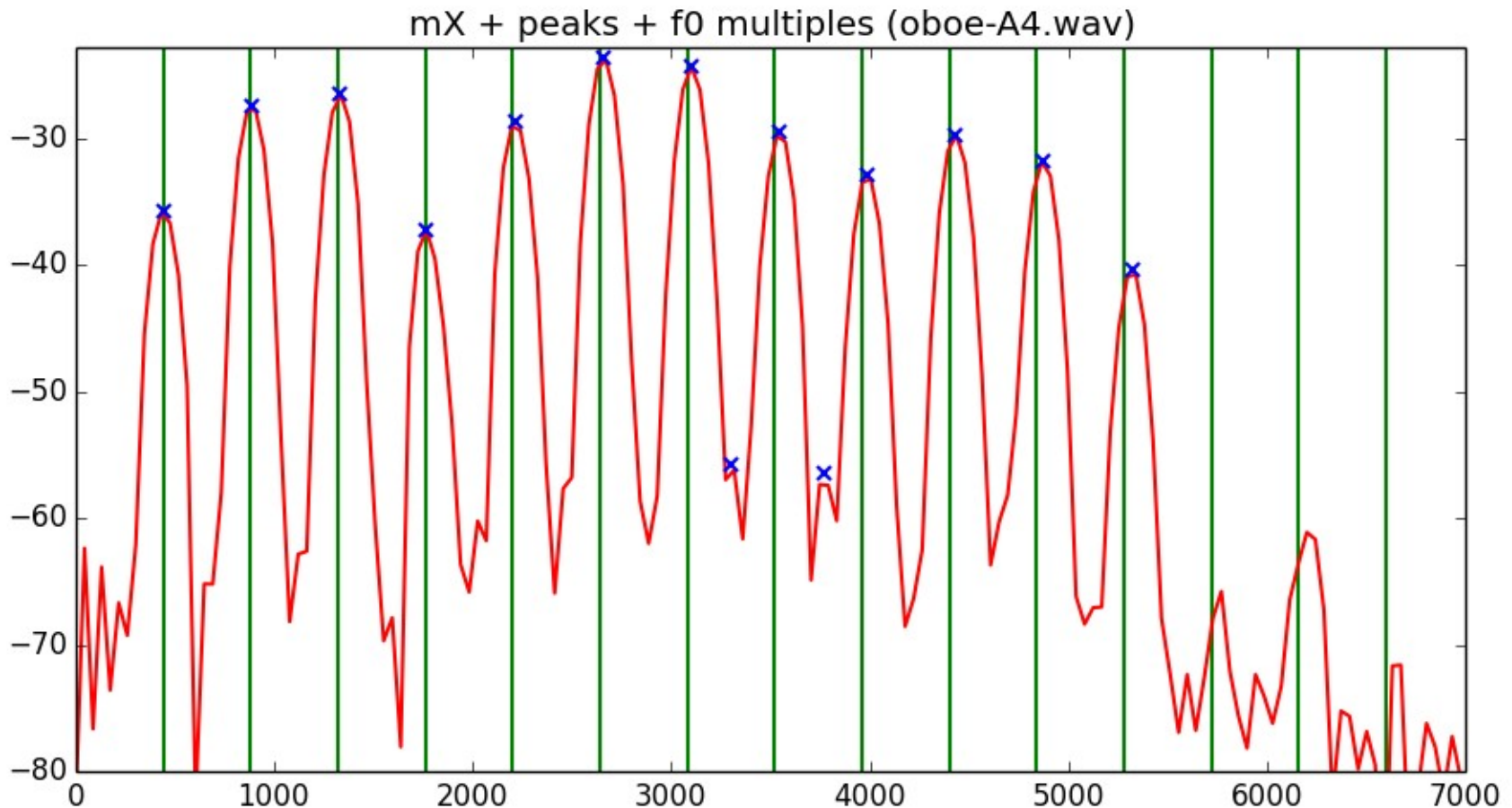


# F0 detection in frequency domain



# 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

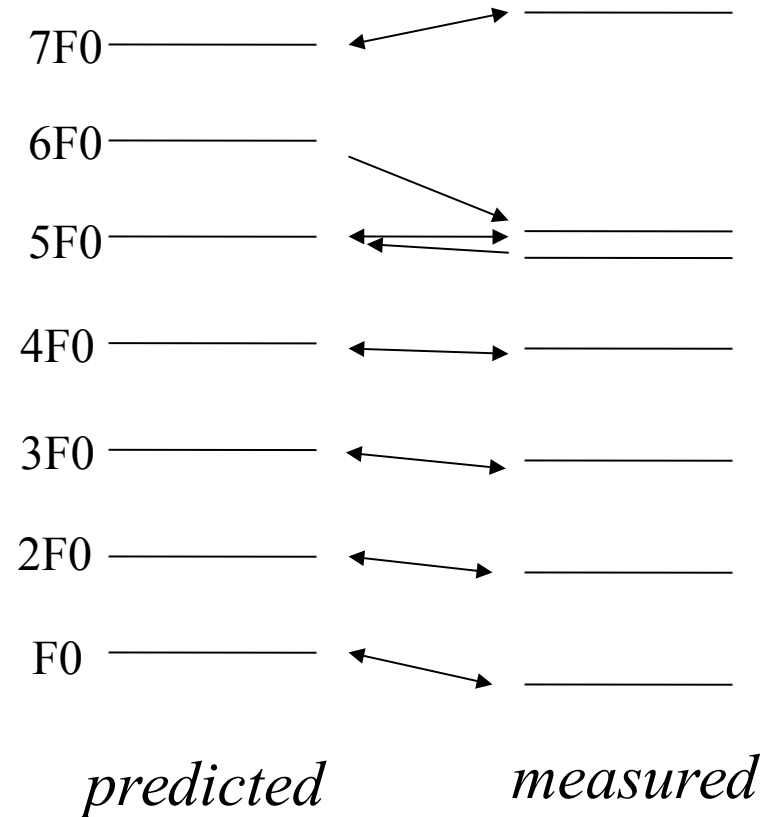
Two-way mismatch algorithm (Maher and Beauchamp, 1994)

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

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

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

$A_{\max}$ : maximum peak magnitude



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

$\Delta f_k$ : diff. between predicted and its closest measured peaks

$f_k, a_k$ : frequency and magnitude of predicted peaks

$A_{\max}$ : maximum peak magnitude

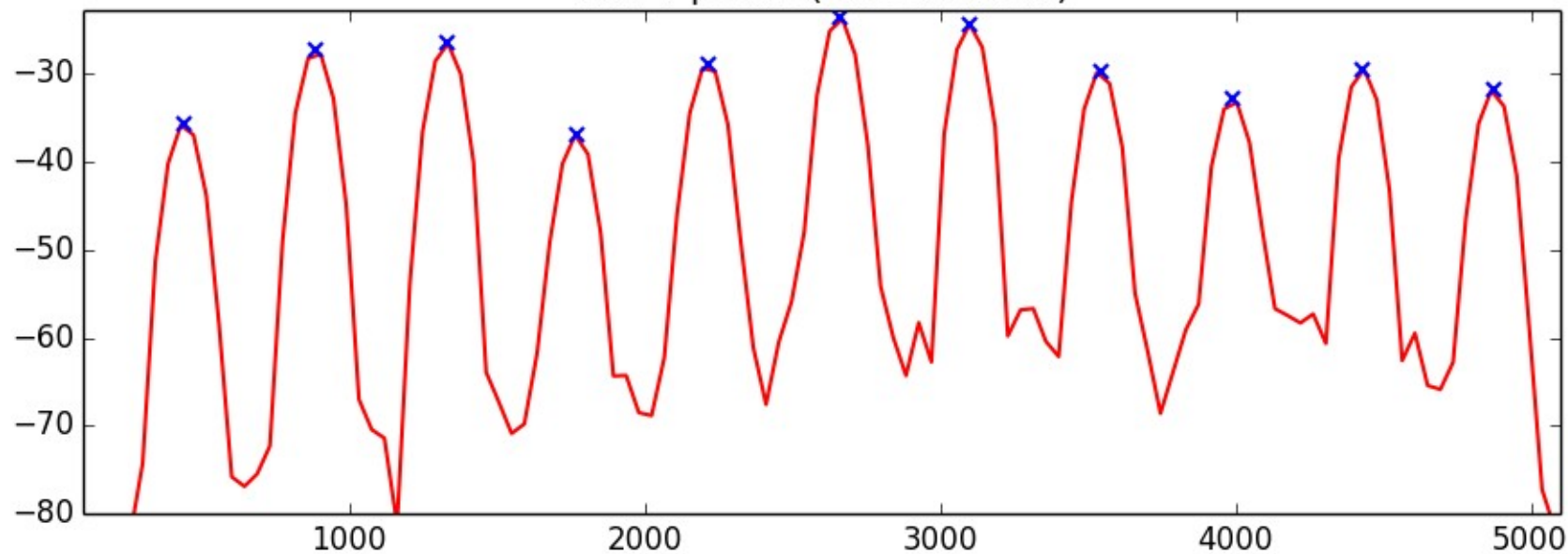
**Total error:**  $\text{Err}_{\text{total}} = \text{Err}_{p \rightarrow m} / N + \rho \text{Err}_{m \rightarrow p} / K$

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

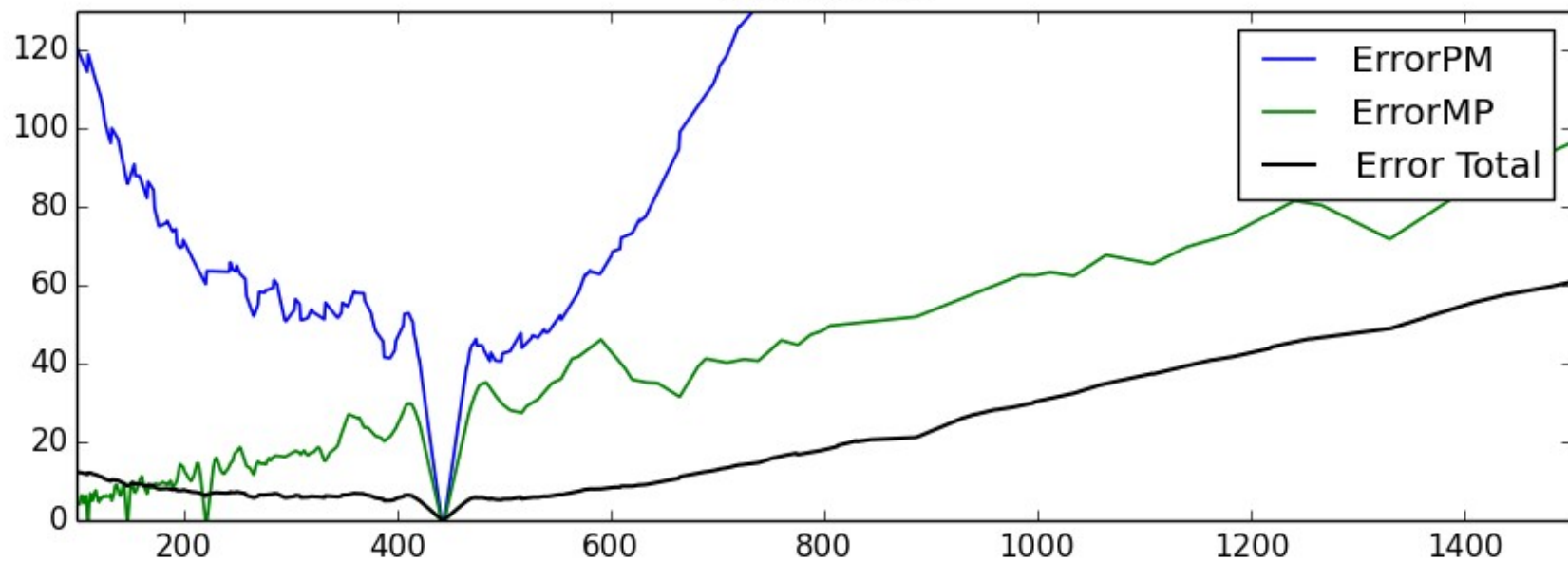
	Err <sub>p-&gt;m</sub>	Err <sub>m-&gt;p</sub>	Err <sub>total</sub>
50Hz	122.58	-3.0	7.49
100Hz	32.0	-3.0	3.83
200Hz	10.0	30.66	4.2

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

mX + peaks (oboe-A4.wav)

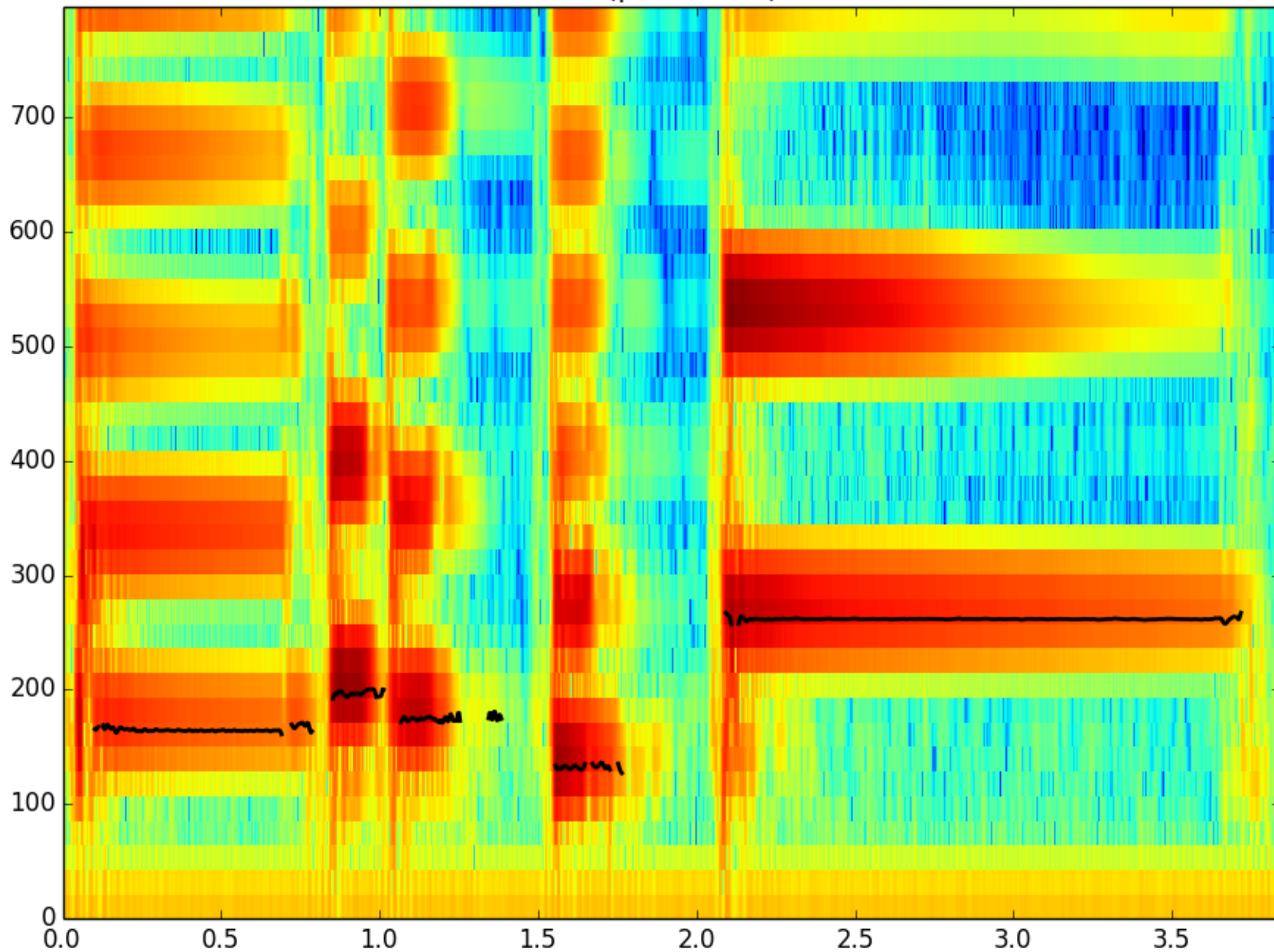


TWM Errors

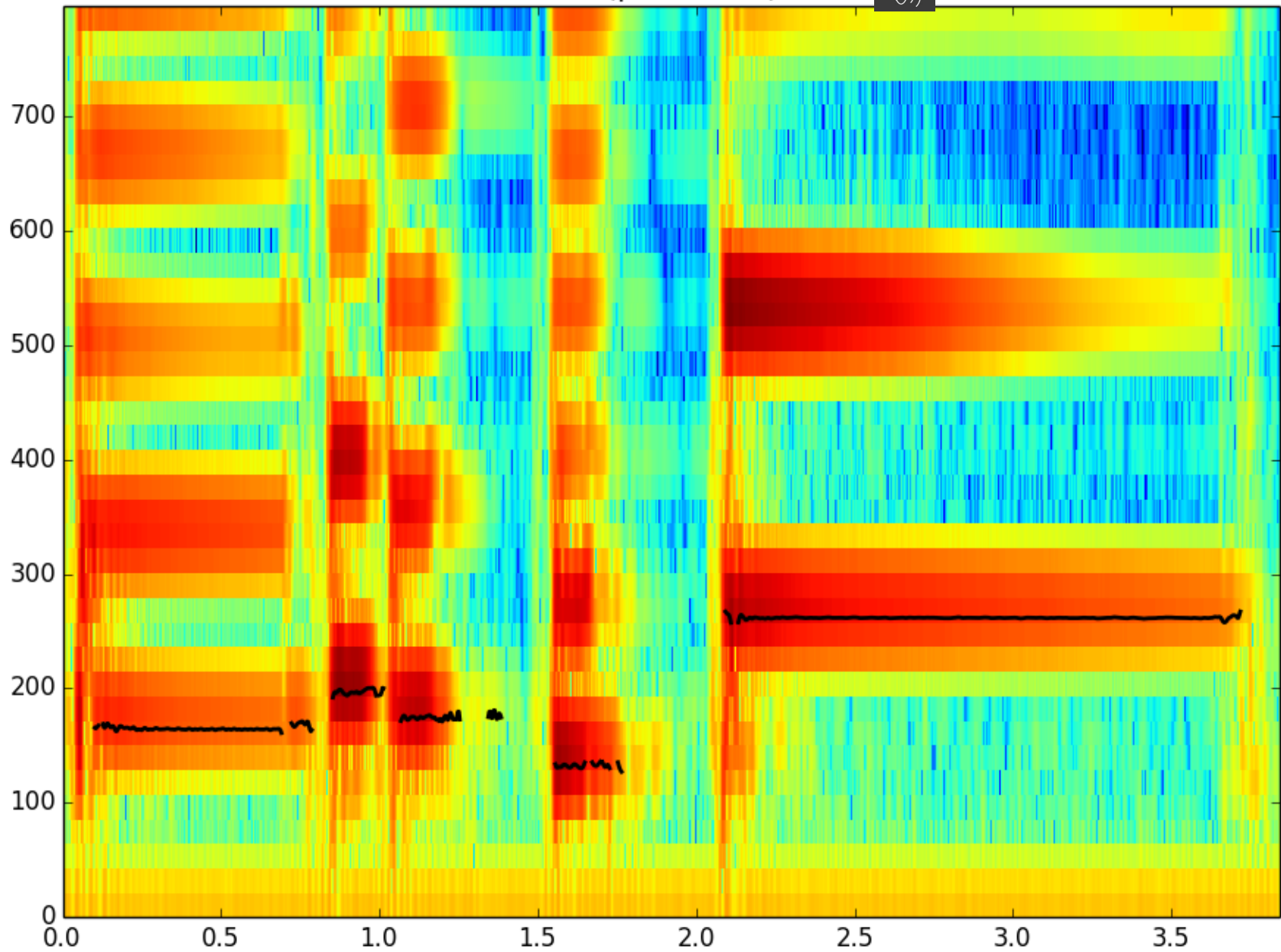




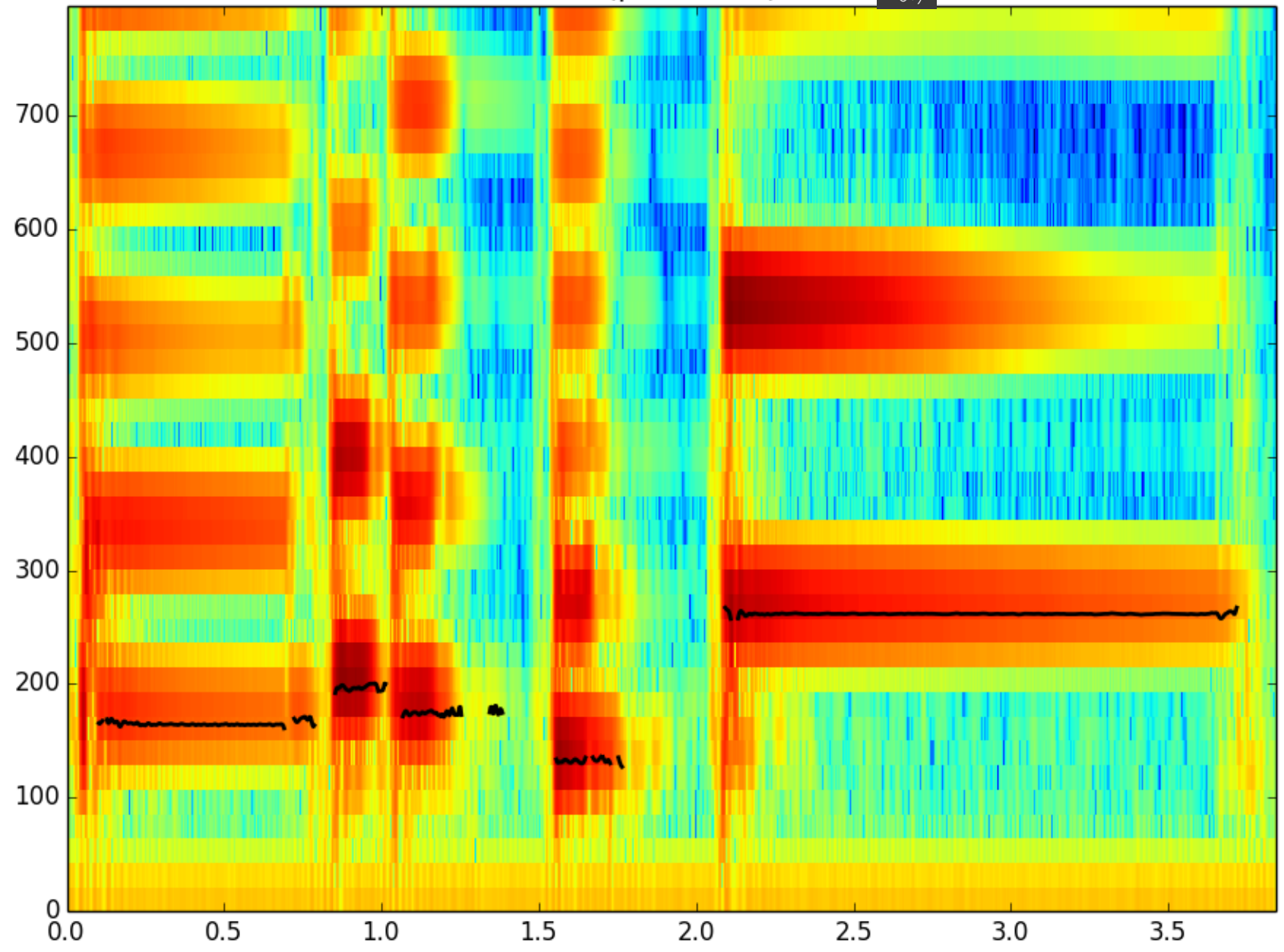
mX + f0 (piano.wav), TWM



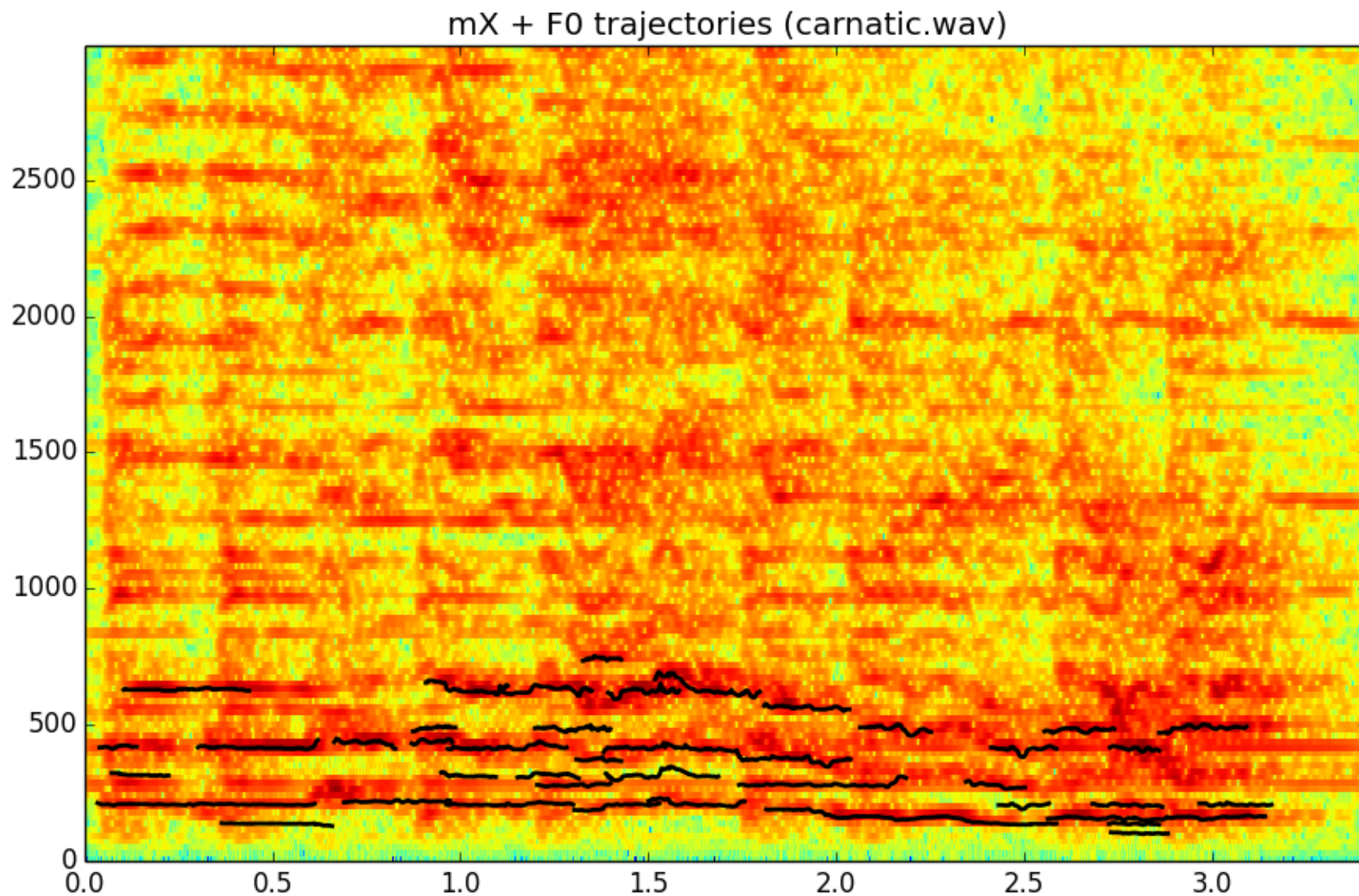
mX + f0 (piano.wav), TWM  $\phi\phi\phi$



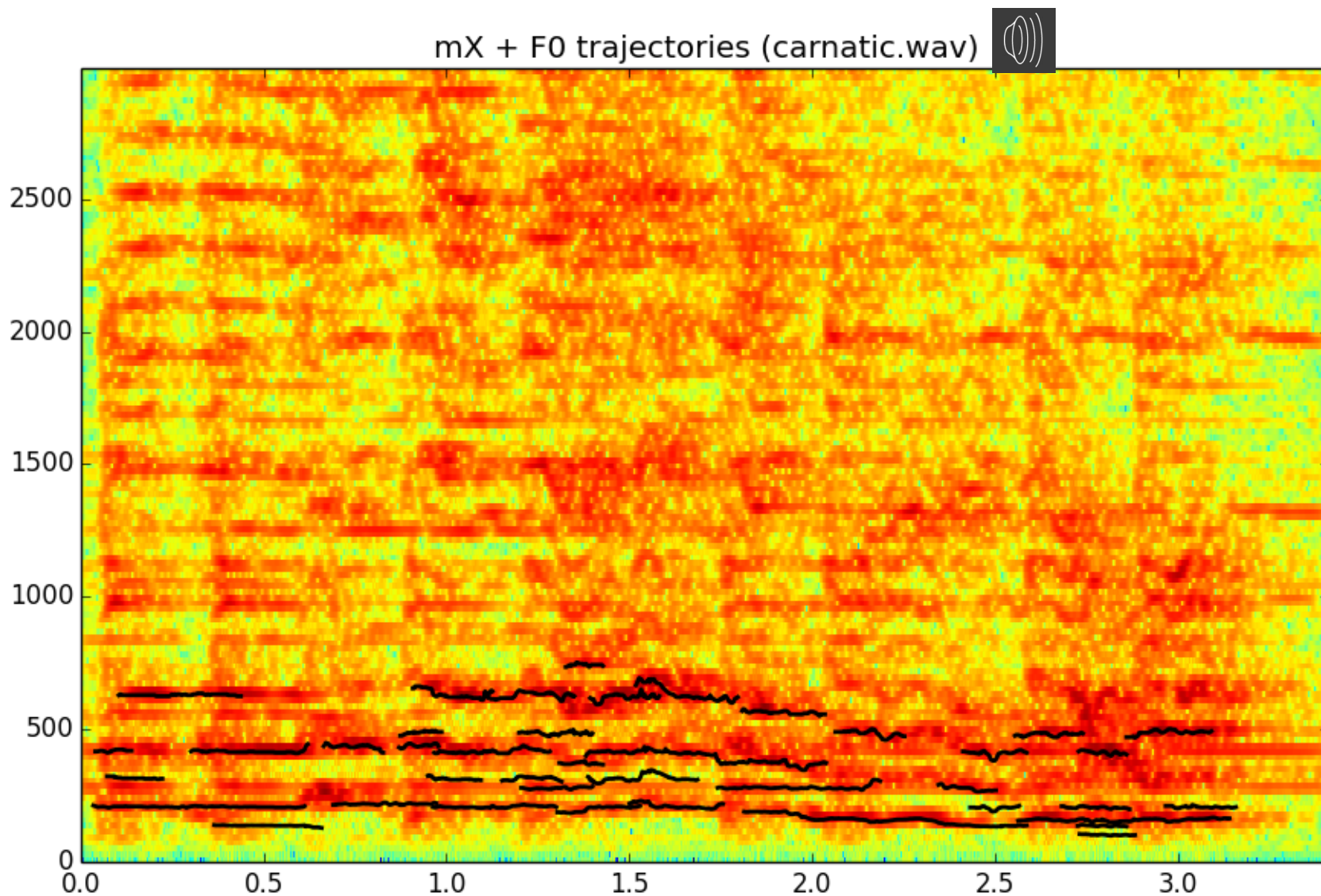
mX + f0 (piano.wav), TWM



# F0 in polyphonic signals

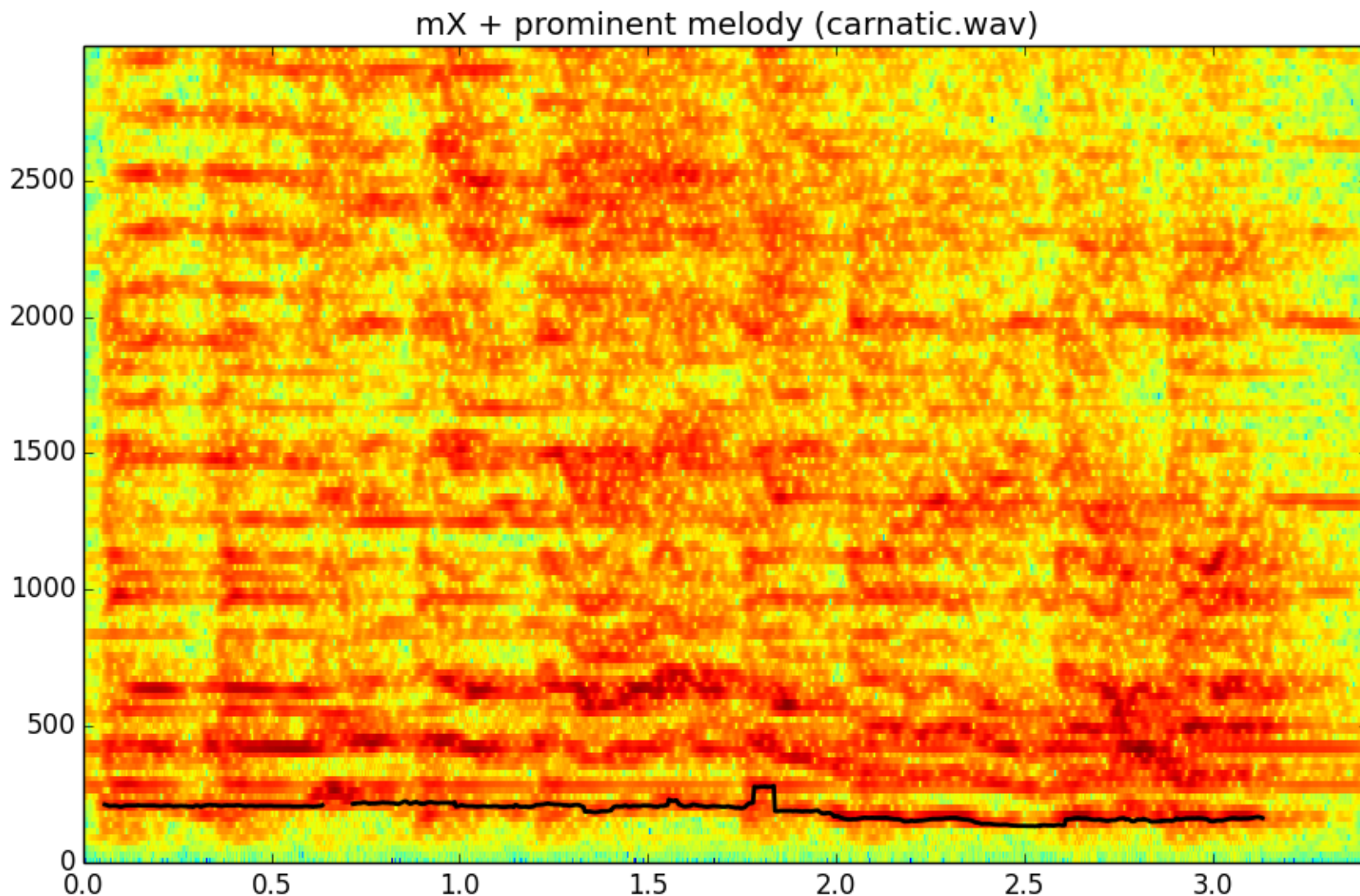


# F0 in polyphonic signals



# Prominent pitch in polyphonic signals

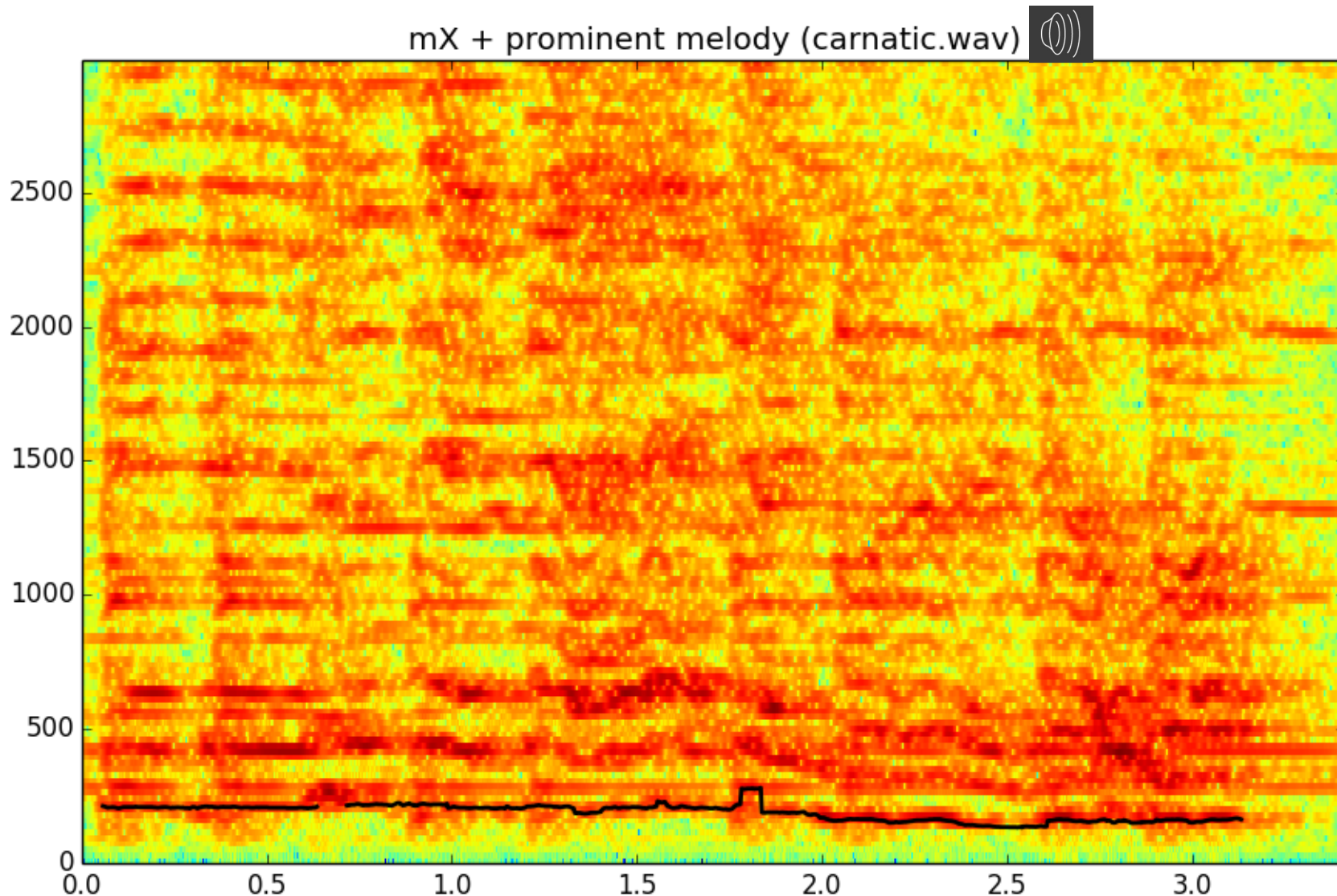
(Salamon and Gómez, 2012)





# 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/Fundamental_frequency)
  - [http://en.wikipedia.org/wiki/Pitch\\_detection\\_algorithm](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/>
- Slides and code released using the CC Attribution-Noncommercial-Share Alike license or the Affero GPL license and available from <https://github.com/MTG/sms-tools>

# 6T2: Fundamental frequency detection

***Xavier Serra***

Universitat Pompeu Fabra, Barcelona

&

Stanford University