Fundamental frequency estimation

summer 2006 lecture on analysis,
modeling and transformation of audio signals

Axel Röbel
Institute of communication science TU-Berlin
IRCAM Analysis/Synthesis Team

25th August 2006

Contents

1 Pitch and fundamental frequency

Pitch

- perceived tone frequency of a sound in comparison with the perceptively best match with a pure sinusoid.

fundamental frequency

- inverse of the signal period of a periodic or quasi periodic signal,
- spectrum contains energy mostly at integer multiples of the fundamental frequency.
- the sinusoids above the fundamental are called the harmonics of the fundamental frequency.

1.1 Relation between pitch and fundamental frequency

- For many signals both terms are closely related, if the higher harmonics are strongly amplified due to formants the pitch may be located not at th first but a higher harmonic.
- if a sound signal contains in-harmonically related sinusoids more then one pitch may be perceived while the signal is actually not periodic.

Applications and importance

- perceptual feature, qualifies an important perceived tone quality
- signal processing, qualifies necessary frequency resolution to resolve and process the harmonically related partials of a quasi periodic sound.
- pitch estimation is often necessary if perceptually relevant processing is desired

1.2 Harmonic sound sources

Sound sources can be roughly divided into two main groups according their capability to create a clear pitch perception.

harmonic sound sources :

- spectrum consists of peaks that create a regular grid at frequencies being an integer multiple of the fundamental frequency
- create a unique and clear pitch perception
- A real sound source can never be ideally periodic because it already time limitation.
- harmonic sound sources are locally periodic

quasi harmonic sound source :

- slight deviations from the harmonic case are denoted quasi harmonic.
- deviations can be due to
  - noise
  - slight deviation from partial position (vibrating strings, bars)

Examples:

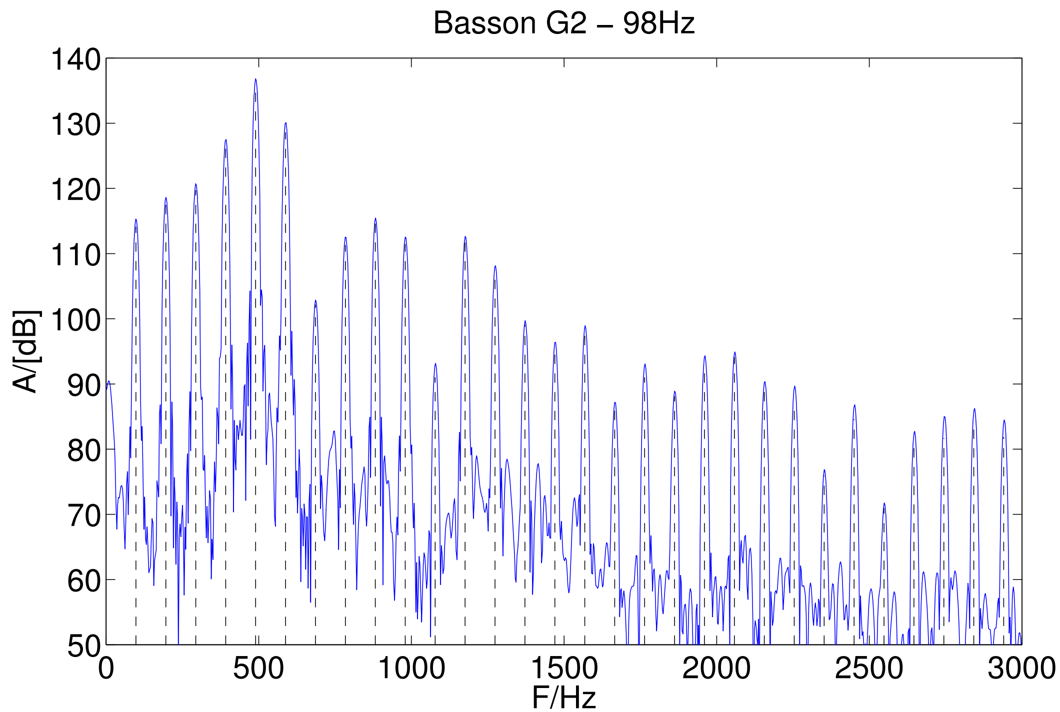

Figure 1: Harmonic spectrum of a Bassoon, Note G5 - 98Hz.

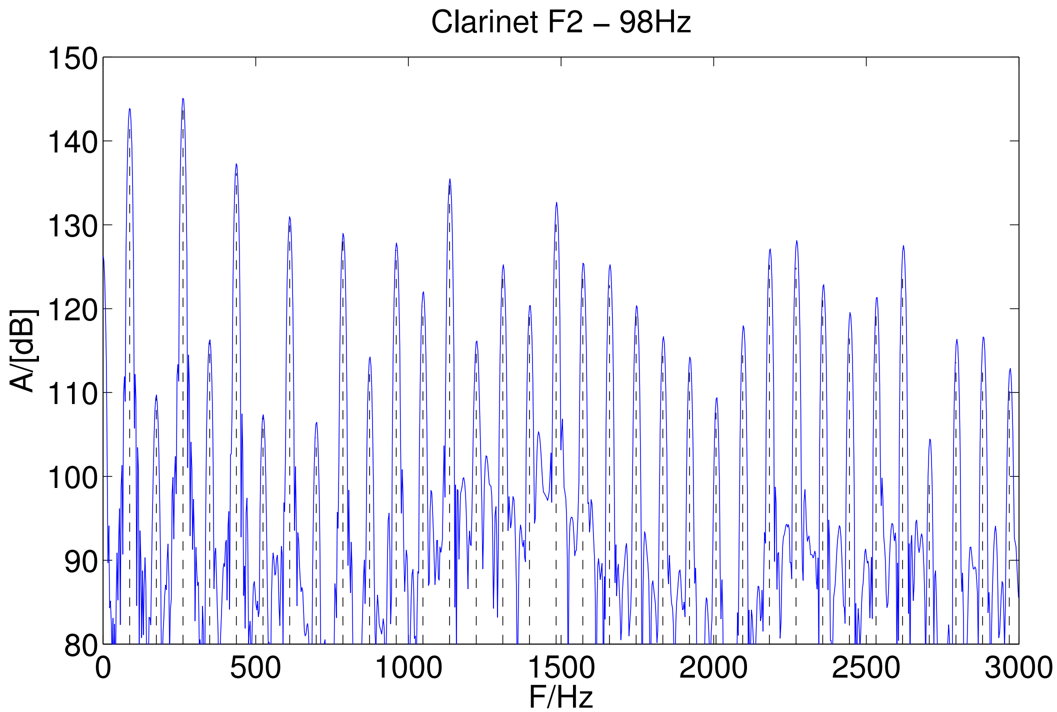

Figure 2: Harmonic spectrum of a Clarinet, Note F2 - 87.3Hz. The first harmonics with odd frequency factors are rather weak.

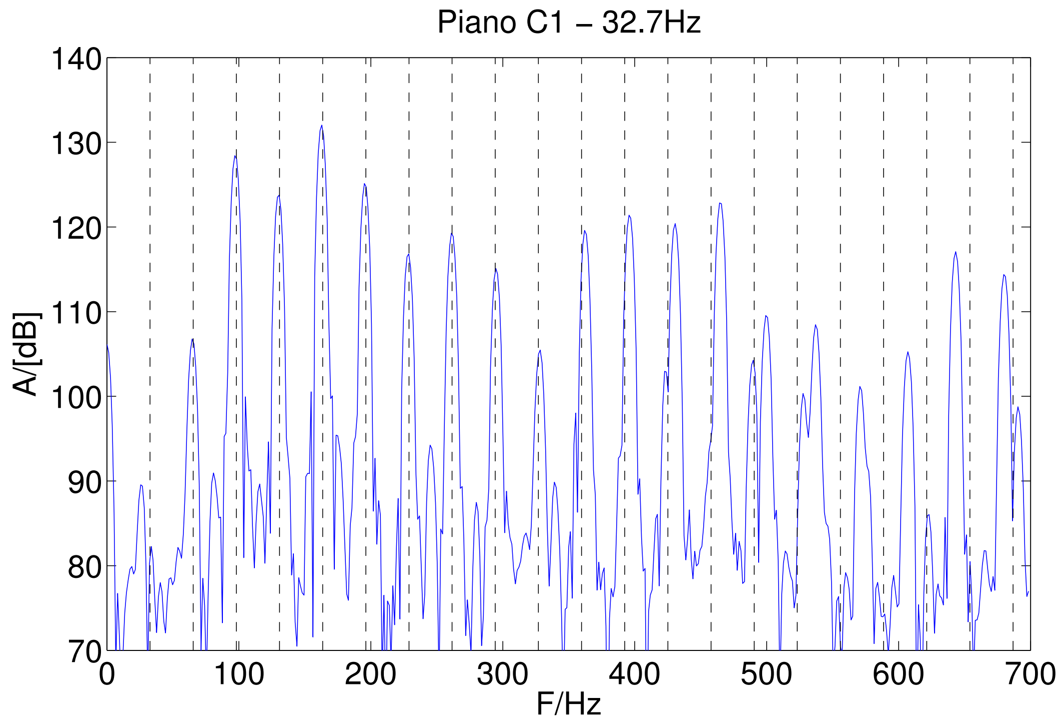

Figure 3: Inharmonic spectrum of a Piano, Note C1 - 32.7Hz. The higher frequency partials clearly deviate from the harmonic positions. Inharmonicity is such that 25nd harmonic will generally be at the $26F_0$.

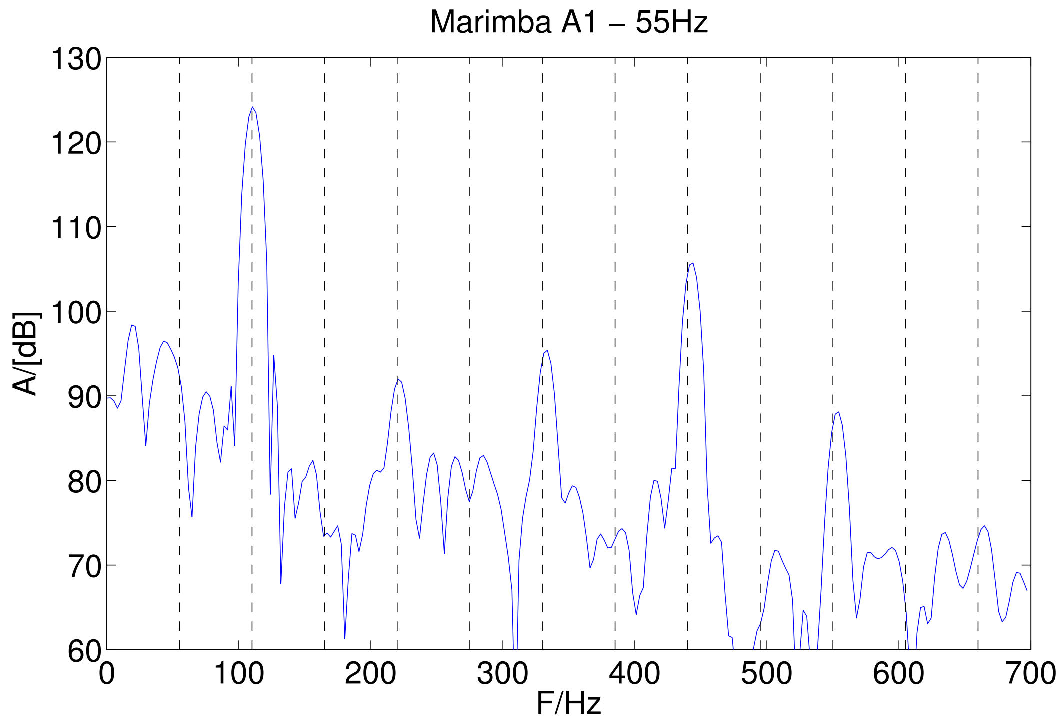

Figure 4: Inharmonic spectrum of a Marimba, Note A1 - 55Hz. The Marimba is normally tuned to the 2nd harmonic of the musical note. There are only a few partials which are slightly inharmonic.

2 Fundamental frequency estimation problem

The basic practical problem of monophonic fundamental frequency estimation the problem is the following:

estimate the time varying fundamental frequency from a given signal!

We start the approach to the problem by discussing the constraints that we may impose

- Validity :
  We assume that the problem is solvable, that is, the signal is quasi harmonic such that a pitch will be perceived,
- F0 range :
  for each practical problem there is a more or less clearly defined range of possible f0 values. The more can the range of possible values the less we have the chance to make an error,
- F0 variability :
  Besides the range of possible F0 values also the expected rate of change will affect the choice of the algorithm. We assume that the rate of change is sufficiently slow.

3 Algorithms

The main cue for fundamental frequency estimation is the definition of the fundamental frequency as the period of the quasi harmonic signal.

Therefore, all fundamental frequency estimation algorithms try to evaluate the periodicity hypothesis related to each f0 in the search range.

A huge number of different algorithms have been developed.

Some fundamental strategies will be presented below.

3.1 direct evaluation of periodicity

- the simplest approach to periodicity evaluation is based on the investigation of the time domain waveform.
- We may test each F0 hypothesis by testing how well the signal will resemble a delayed version of itself.
- evaluation criteria can be either the correlation or the difference between the signal and its delayed version.

both criteria are closely related. The auto correlation function at time position n_0 is

$$C_x(d) = \sum_{n=n_0}^{n_0+W} x(n)x(n-d) \tag{1}$$

with W being the length of the summation window and d being the time lag.

For the difference function using a squared difference and the same window we get

$$D_x(d) = \frac{1}{W} \sum_{n=n_0}^{n_0+W} (x(n) - x(n-d))^2 \tag{2}$$

which can be transformed into

$$\begin{aligned} D_x(d) &= \sum_{n=n_0}^{n_0+W} (x(n)^2 + x(n-d)^2 - 2(x(n)x(n-d)) && (3) \\ &= R_W(n)^2 + R_W(n-d)^2 - 2C_x(d). && (4) \end{aligned}$$

The latter formula allows us to see the relation between both measures.

For periodic signals:

- the auto correlation function $C_x(d)$ will start with its maximum value determined by the signal energy
- the difference function $D_x(d)$ starts with 0.
- Both functions are periodic with period P equal to the period of the analyzed signal.

- the first position d where $C_x(d) = C_x(0)$ or $D_x(d) = D_x(0)$ will be the period.

The basic problem is that the position $d = 0$ always achieves optimal performance with respect to the criterion and due to noise and F0 evolution the correct period $d = P$ will not achieve this value.

A nice solution has been proposed in the algorithm **YIN** [1] [dCK02].

The normalized difference function is defined as

$$Dn_x(d) = \begin{cases} 1 & \text{for} \quad d = 0 \\ \frac{D_x(d)}{\frac{1}{d}\sum_{k=1}^{d} D_x(d)} & \text{else.} \end{cases} \tag{5}$$

It starts at 1 instead of 0 and achieves its minimum value for the first time at $d = P$.

[1] matlab code for YIN can be downloaded from `http://www.ircam.fr/pcm/cheveign/sw/yin.zip`

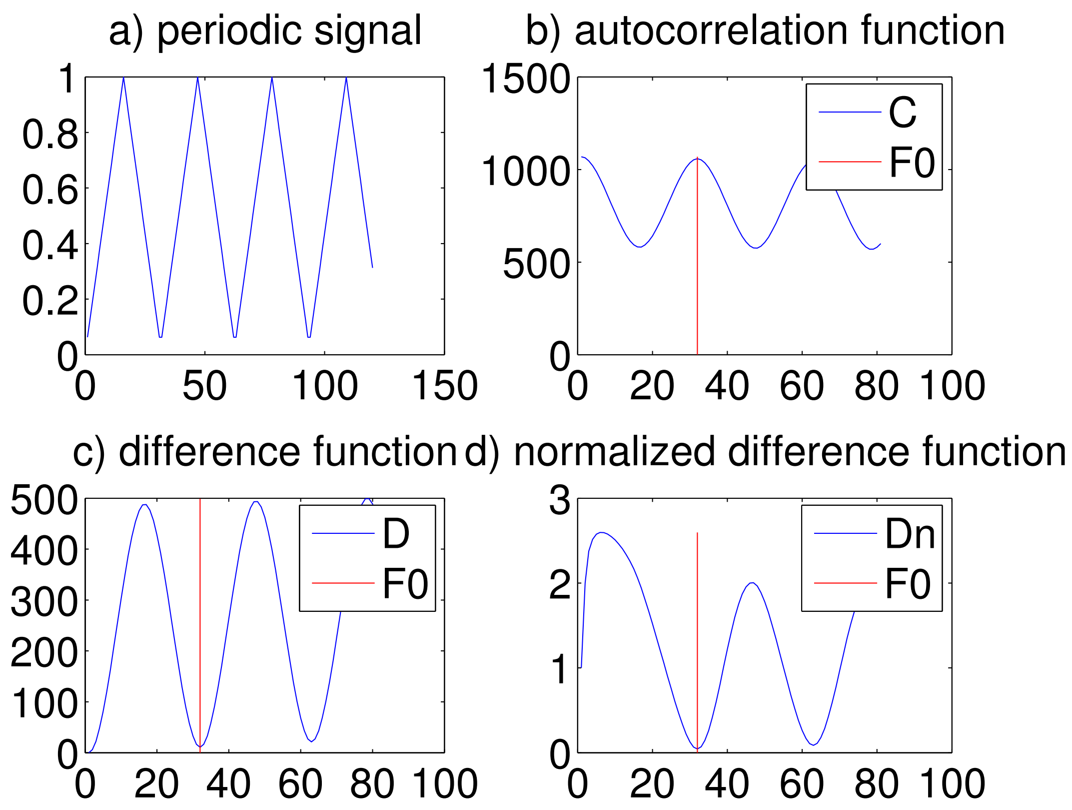

Figure 5: periodicity evaluation in the time domain. a) signal, b) autocorrelation function, c) difference function, d) normalized difference function.

General problems with the time domain approach:

Ambiguity of the evaluation function with respect to sub harmonics, if P is the correct path then $d = lP$ with $l = 1 \ldots$ will all achieve an optimal result.

Further means to select the optimal solution are required.

3.2 frequency domain harmonic matching

periodicity hypothesis can be evaluated in the spectral domain.

- The auto correlation function can be calculated from the DFT of the signal. Due to the symmetry of the magnitude spectrum, only cosine terms contribute which yields

$$C_x(d) = \sum_{k=0}^{N-1} |X(k)|^2 \cos(\frac{2\pi}{N}kd) \tag{6}$$

The correct d gives the best match between the observed spectral peaks and the cosine basis function.

The generalized form of the acf based evaluation is

$$V(d) = \sum_{k=0}^{N-1} \mathsf{f}(|X(k)|) \cos(\frac{2\pi}{N}kd) \tag{7}$$

with f representing an arbitrary function that may be used to transform the spectrum.

f$(|X|) = \log(|X|)$ yields a cepstrum based fundamental frequency estimator.

Conclusion:

- period evaluation is related to testing for spectral energy at a regular grid.
- Arbitrary evaluation functions may be used to test these positions : comb filters, probabilistic models with Gaussian functions at the expected positions [Got01].
- all methods tend to make sub harmonic errors and need further means to compensate these errors.
- time domain processing does not allow to weight the impact of different spectral regions
- the spectral domain processing allows for more flexible counter balancing of subharmonic errors: often used are physically motivated envelope smoothness criteria.
- the method does not work very well for inharmonic sounds.

3.3 spectral period evaluation

a periodic waveform will create a regular grid of peaks in the spectral domain

instead of a test for spectral position an evaluation for the major period of the amplitude spectrum is possible as well.

Simplest approach: periodicity analysis with the part of the spectrum at positive frequencies

$$C_X(\lambda) = \sum_{k=0}^{n_0+W} |X(k)||X(k-\lambda)| \quad (8)$$

Differences compared to the periodicity evaluation in the time domain are:

- less sensitive to absolute position of partials, therefore better suited for inharmonic spectra (Piano),
- ambiguity with respect to super harmonics, for a quasi harmonic spectrum with fundamental frequency F_0 the frequencies lF_0 with $l \in 1, 2, \ldots$ will all get similar score,
- spectral periodicity and temporal periodicity give rise to opposite errors,

- spectral periodicity analysis can be used to achieve counter balance of ambiguity of temporal periodicity analysis (similar as envelope smoothness).

3.4 psycho acoustic approaches

Inspired by auditory processing chain in the human ear and brain:

- bandpass filtering in the cochlea,
- encoding of bandpass signals via the hair cells into neural activation, usually modeled by means of half-wave rectification and lowpass filtering.
- periodicity analysis within individual channels
- combination of individual results from the different channels into a single result.

A recent interesting approach to translate the auditory processing chain into a polyphonic fundamental frequency detector has been presented in [Kla05].

3.5 Properties of the auditory processing model

- The auditory process seems to date to perform a combined spectral domain and time domain approach,
- cochlea filter splits signal into individual channels with frequency dependent bandwidth,
- half wave rectifier and lowpass filter

  evaluates:
  - spectral position for low frequency partials that are individually resolved
  - spectral period for high frequency partials that fall together into a single band.
- periodic analysis in the individual bands create a fundamental period hypothesis, due to nonlinear procedure the fundamental partial is always present,
- weighted summation of all fundamental period hypothesis give the final result.

3.6 half wave rectification

half wave rectification is a nonlinear process that maps a signal $s(n)$ into

$$s_r(n) = \max(0, s(n)) \tag{9}$$

Because the process is nonlinear an analytic investigation is difficult. A complete closed form solutions can be obtained only for the rectification of individual sinusoids.

Suppose $s(n)$ is a real sinusoid with frequency Ω and $v(n)$ is the rectangular square wave with $\max(v(n)) = 1$, $\min(v(n)) = 0$, and period $\frac{2\pi}{\omega}$ and the same phase as the sinusoid $s(n)$. In this case the result of half-wave rectification is equivalent to multiplication between he sinusoid and the square wave:

$$s_r(n) = \max(0, s(n)) = s(n)v(n) \tag{10}$$

Because the spectrum of the square wave is known the spectrum of $s_r(n)$ is simply the superposition of two frequency shifted versions of the spectrum $V(k)$ of the square wave function $v(n)$ shifted to the positions of the sinusoids $V(k-w)$ and $V(k+w)$.

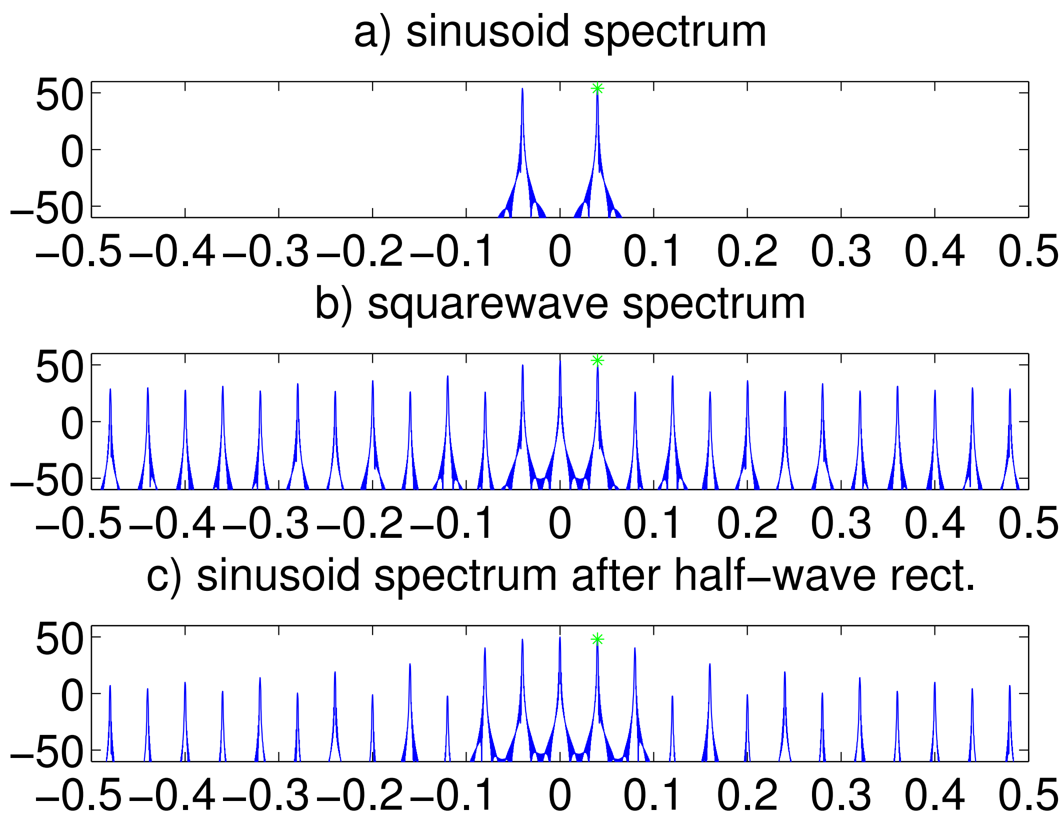

Figure 6: Half-wave rectification of an individual sinusoid (a) using the square-wave with the spectrum (b) creates an harmonic overtone spectrum (c) with fundamental given by the original sinusoid.

The case of an higher number of partials is difficult to calculate, however, a simple example provides some insights into the results

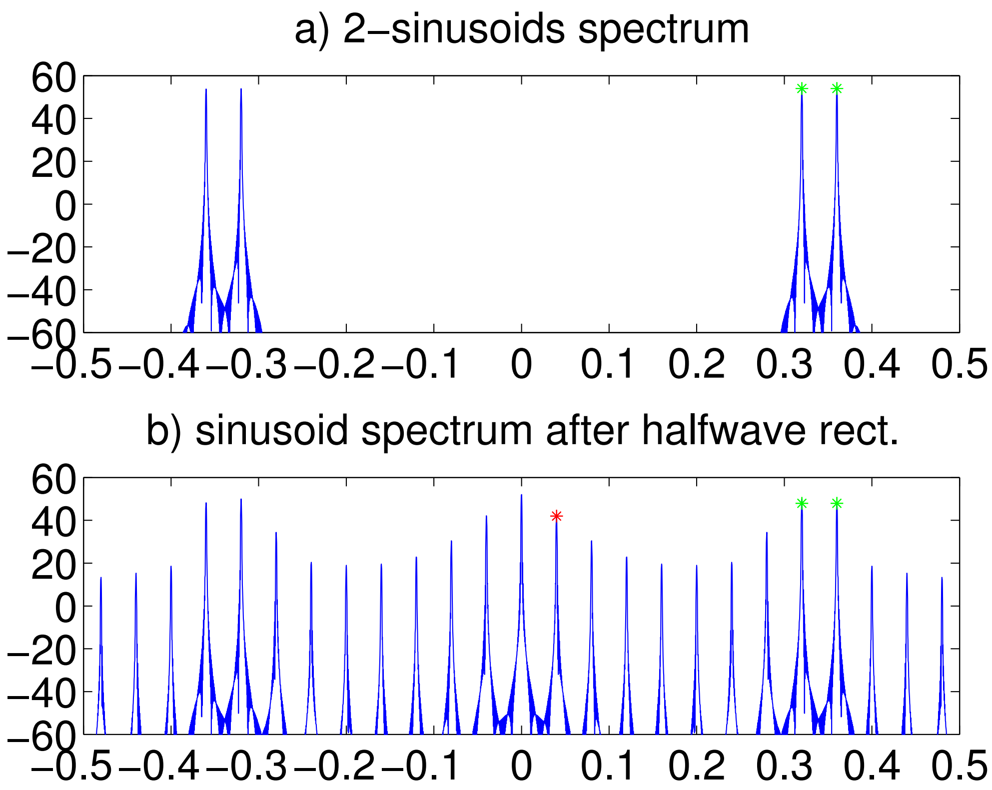

Figure 7: Half-wave rectification of a sum of 2 sinusoids (a) with frequency offset $\Delta_w = 0.04$ creates the overtone spectrum (b) with fundamental frequency given by the frequency difference between the two sinusoids. The overtone spectrum is harmonic if the position of the original sinusoids and the frequency difference are compatible.

3.7 Performance

The existing algorithms for fundamental frequency estimation from monophonic signals are rather effective. Due to the fact that it requires not more than two time the period of the minimum fundamental to be detected The algorithm yin is especially suitable for highly nonstationary speech signals. For musical instruments the problem of reverberation due to slow decay of previous notes renders the analysis more or less polyphonic. Due to the possibilities to restrict the analysis to limited frequency ranges and to threshold spectral amplitudes it appears that for musical instruments spectral domain preprocessing is helpful such that spectral domain evaluation of the acf is more appropriate. For the harmonic instruments the error is below 1% if less than a quarter note of deviation is excepted. For speech signals evaluation usually allows a larger tolerance of about $\pm 20\%$ of the pitch. With this performance criterion yin operates with less then 1% error on speech signals as well.

References

[dCK02] A. de Cheveigné and H. Kawahara. Yin, a fundamental frequency estimator for speech and music. *Journal Acoust. Soc. Am.*, 111(4), 2002. 14

[Got01] M. Goto. A predominant-F0 estimation method for CD recordings: MAP estimation using an EM algorithm for adaptive tone models. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, pages 3365–3368 (Vol. V), 2001. 18

[Kla05] A. Klapuri. A perceptually motivated multiple-f0 estimation method. In *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2005. 21