UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Engenharia Elétrica e de Computação

Christhian Henrique Gomes Fonseca

Low-Latency f_0 Estimation for the Finger Plucked Electric Bass Guitar Using Absolute Difference Function

Estimativa de f0 de Baixa Latência para o Baixo Elétrico Tocado com os Dedos Usando a Função Diferença Absoluta

Campinas

2020

UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Engenharia Elétrica e de Computação

Christhian Henrique Gomes Fonseca

Low-Latency f_0 Estimation for the Finger Plucked Electric Bass Guitar Using Absolute Difference Function

Estimativa de f0 de Baixa Latência para o Baixo Elétrico Tocado com os Dedos Usando a Função Diferença Absoluta

Dissertation presented to the School of Electrical and Computer Engineering of the University of Campinas in partial fulfillment of the requirements for the degree of Master in Electrical Engineering, in the area of computer engineering.

Dissertação apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica, na Área de Engenharia de Computação.

Supervisor: Prof. Dr. Tiago Tavares

Este exemplar corresponde à versão final da dissertação defendida pelo aluno Christhian Henrique Gomes Fonseca, e orientada pelo Prof. Dr. Tiago Tavares

Campinas

2020

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Luciana Pietrosanto Milla - CRB 8/8129

F733e

Fonseca, Christhian Henrique Gomes, 1988-
Low-latency f0 estimation for the finger plucked electric bass guitar using absolute difference function / Christhian Henrique Gomes Fonseca. – Campinas, SP : [s.n.], 2020.

Orientador: Tiago Fernandes Tavares.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. MIDI (Normas). I. Tavares, Tiago Fernandes, 1984-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Estimativa de f0 de baixa latência para o baixo elétrico tocado com os dedos usando a função diferença absoluta
Palavras-chave em inglês:
MIDI (Standard)
Área de concentração: Engenharia de Computação
Titulação: Mestre em Engenharia Elétrica
Banca examinadora:
Tiago Fernandes Tavares [Orientador]
Bruno Sanches Masiero
Stéphan Olivier Schaub
Data de defesa: 12-05-2020
Programa de Pós-Graduação: Engenharia Elétrica

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-1758-1695
- Currículo Lattes do autor: http://lattes.cnpq.br/0548825598476595

COMISSÃO JULGADORA - DISSERTAÇÃO DE MESTRADO

Candidato: Christhian Henrique Gomes Fonseca RA: 138183

Data da Defesa: 12 de maio de 2020

Título da Tese: "Low-Latency f0 Estimation for the Finger Plucked Electric Bass Guitar Using Absolute DifferenceFunction" (Estimativa de f0 de baixa latência para o baixo elétrico tocado com os dedos usando a função diferença absoluta)

Prof. Dr. Tiago Fernandes Tavares (Presidente)

Prof. Dr. Bruno Sanches Masiero

Prof. Dr. Stéphan Olivier Schaub

A ata de defesa, com as respectivas assinaturas dos membros da Comissão Julgadora, encontra-se no SIGA (Sistema de Fluxo de Dissertação/Tese) e na Secretaria de PósGraduação da Faculdade de Engenharia Elétrica e de Computação.

Dedico esta dissertação a minha família e amigos.

Acknowledgements

Agradeço

A meus pais (in memorian), por existir e pela minha criação.

A minha esposa, Heloisa, pelo carinho e compreensão.

A minha filha, Aurora, pela motivação.

A minha família, por acreditarem em mim.

Ao meu orientador, Tiago, pela paciência e ensinamentos.

Aos amigos, Giovani, Junior e Marcio, pelas gravações.

Aos professores Bruno e Tuti pelas sugestões que enriqueceram este trabalho.

"Music is the hidden arithmetical exercise of a mind
unconscious that it is calculating"
(Gottfried Wilhelm Leibniz)

Abstract

Audio-to-MIDI conversion can be used to allow digital musical control through an analog instrument. Audio-to-MIDI converters rely on fundamental frequency estimators that are usually restricted to a minimum delay of two fundamental periods. This delay is perceptible for the case of bass notes. In this dissertation, we propose a low-latency fundamental frequency estimation method that relies on specific characteristics of the electric bass guitar. By means of physical modeling and signal acquisition, we show that the assumptions of this method are based on the generalization of all electric basses. We evaluated our method in a dataset with musical notes played by diverse bassists. Results show that our method outperforms the Yin method in low-latency settings, which indicates its suitability for low-latency audio-to-MIDI conversion of the electric bass sound.

Keywords: f_0 estimation; low latency, Audio-to-MIDI converter; Music information retrieval; MIDI-bass.

Resumo

A conversão de áudio para MIDI pode ser usada para permitir o controle musical digital por meio de um instrumento analógico. Os conversores de áudio para MIDI dependem de estimadores de frequência fundamental que são frequentemente restritos a um atraso mínimo de dois períodos da frequência fundamental. Este atraso é perceptível no caso de notas graves, pois as frequências fundamentais tem períodos mais longos. Nesta dissertação, propõe-se um método de estimativa da frequência fundamental de baixa latência que se baseia em características específicas do baixo elétrico. Por meio de modelagem física e aquisição de sinais, mostramos que o método se baseia na generalização para todos os baixos elétricos. Avaliamos nosso método em um conjunto de dados com notas musicais tocadas por diversos baixistas. Os resultados mostram que nosso método supera o método Yin em configurações de baixa latência, o que indica sua adequação à conversão de baixa latência de áudio em MIDI do som de baixo elétrico.

Palavras-chaves: Estimador de f_0; Conversor de áudio para MIDI de baixa latência; Recuperação de informações musicais; Baixo MIDI.

List of Figures

List of Tables

Contents

1 Introduction

Digital instruments and controllers commonly use communication protocols such as the MIDI (Musical Instrument Digital Interface) standard to communicate with each other. This allows combining different digital synthesizers, controllers, and effect racks, which expands the expressive possibilities related to timbres, musical performances, musical recordings, and notations (GIBSON J., accessed 05/9/2019). This toolchain can also include analog instruments by means of audio-to-MIDI converters (DERRIEN, 2014).

Audio-to-MIDI converters are devices that aim at identifying the notes played by an instrument in real-time or retrieving them from an audio file. For such, they use a perceptual model that relates the fundamental frequency (f_0) of an audio signal of a tonal sound to its pitch (P., 2003). Many well-known algorithms aim at estimating f_0, such as the autocorrelation (Rabiner, 1977) and the Yin method (CHEVEIGNé; KAWAHARA, 2002).

f_0 estimators commonly aim at finding periodicity in a signal s_j. The periodicity is based on the model

$$s_t = s_{t+kT_0}, \tag{1.1}$$

where T_0 is the fundamental period of s_t and $k \in \mathbb{Z}$. Methods that rely on this property commonly require analyzing at least two fundamental periods of the signal. This incurs in a lower-bound for the latency of Audio-to-MIDI conversion that can be close to 50 ms for the lowest notes (41.2 Hz) in standard 4-string electric basses. These long delays are perceptually detectable and this can impair the use of basses as a MIDI controller.

In this work, we aimed at attenuating this problem using an f_0 estimation method especially crafted for the electric bass guitar. The method exploits specific properties of the electric bass guitar waveform. Our method allows f_0 estimation with an algorithmic latency of 1.1 times the fundamental period of the signal, which is about 27 ms for the lowest frequency note of the four-string traditional bass guitar.

Experimental results show that this method is effective with an error rate of 15%. This is half the error rate of Yin, the baseline method, when an equal latency is considered. The method was tested for the frequency range from 41.2 Hz to 392 Hz, that is, from the lowest to the highest note of the standard four-string electric bass guitar.

2 Theoretical Background

2.1 Pitch theory

Pitch is a psychoacoustical attribute of the sound related to the perception of a repetition rate of a waveform (HELLER, 2012) above 20 Hz, in which it is perceived not as rhythm but as tone. The lowest regular repetition rate is called Fundamental Frequency (f0) and can be used to decompose harmonic complex tones into sinusoidal harmonic components whose frequencies are multiple integers of the fundamental frequency f_0, that is:

$$s_t = \sum_{m=1}^{M} a_m \cos(2\pi m f_0 t + \phi_m). \tag{2.1}$$

The relative harmonic amplitudes a_m, among other attributes, are commonly associated with timbre differences, and the fundamental frequency f_0 is closely related to the sensation of pitch (OXENHAM, 26 September 2012). In this study, we assume that the fundamental frequency is the physical counterpart of the psychological sensation of tonality, commonly named as pitch, hence estimating the fundamental frequency is equivalent to finding the pitch of a signal.

Moreover, perfectly periodic waveforms are rare, because in the real world the signals differ between each repetition, even if small. Thus it is interesting to extend the concept of the pitch to quasi-periodic signals, that is, waveforms that are not perfectly identical in each cycle but have reasonable similarities between them to the point where they can be identified as repetitions. Within this concept, the signals can be modulated, turned off and on, and yet have a pitch. Still, there are exceptions to pitch determination by fundamental frequency such as non-periodic but pitch-evoking signals (CHEVEIGNé; KAWAHARA, 2002).

The human ability to detect the pitch of a sound, that is, human tonal perception, has been linked to biological traits such as the periodicity of neural patterns (CARIANI; DELGUTTE, 1996) and the harmonic partial pattern present by the cochlea (TERHARDT, 1974). Tonal perception allows us to perceive the amount of repetition of events that are too fast to be counted (FORNARI, 2010).

In music, several standards define the tuning frequency for each note. The most commonly used nowadays is called Pitch International Standard, which defines the fundamental frequency of the note A above middle C should be 440 Hz (ISO16:1975-ACOUSTICS, 1975).

For the western music, in the equal tempered chromatic system, the frequency variation between one note and the next is $2^{\frac{1}{12}}$ and the variation given an interval Δ_{notes} of notes is given by the equation (IAZZETTA, accessed 04/25/2019):

$$\Delta_{freq} = f_1 2^{\frac{\Delta_{notes}}{12}} \tag{2.2}$$

where f_1 is the frequency of the lower note in the interval.

Using this rule, the fundamental frequency is set for every note. Some examples are shown in Table 2.1

Table 2.1 – Some notes and their respectives frequencies e periods.

Note	**Frequency (Hz)**	**Period (ms)**
A 2	220.	4.545
A# 2	233.08	4.29
B 2	246.94	4.05
C 3	261.62	3.822
C# 3	277.18	3.608
D 3	293.66	3.405
D# 3	311.12	3.214
E 3	329.62	3.034
F 3	349.22	2.863
F# 3	369.99	2.703
G 3	391.99	2.551
G# 3	415.30	2.408
A 3	440.	2.273

2.2 Pitch detection

Human beings are able to hear frequencies from 20 Hz to 20 kHz and our ability to interpret them comes from the cochlea, more specifically from the organ of Corti, which has thousands of ciliated cells. Depending on the components present in the sound, different regions of this organ resonate and the ciliated cells connected to that region generate electrical impulses that will be transported by the auditory nerve for later interpretation as sound. So we can interpret it from its various components, detecting the pitch of a quasiperiodic sound, usually by its Lower component (FORNARI, 2010).

Based on various mathematical principles, various methods have been developed for determining the pitch of a signal. Assuming pitch is an auditory sensation related to the fundamental frequency of a periodic or pseudo-periodic waveform, it is sufficient to determine the interval T_0 between its repetitions and then find f0 through its inverse. This can be simple for perfectly sinusoidal signals, but natural waveforms are diverse and can have many harmonic components that make the task difficult (GERHARD, 2003).

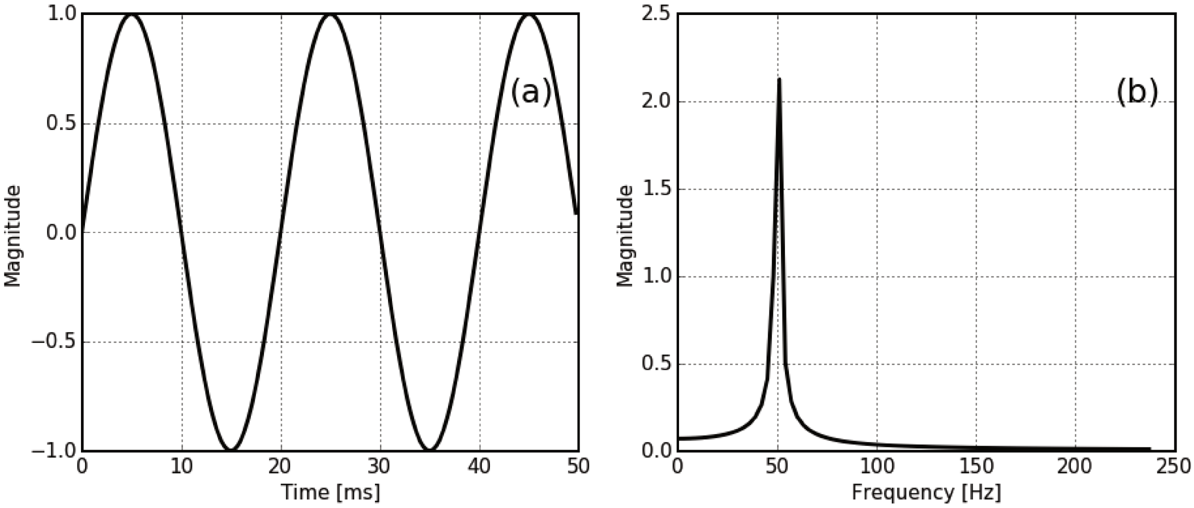

Figure 2.1 – Waveform with no upper harmonics.

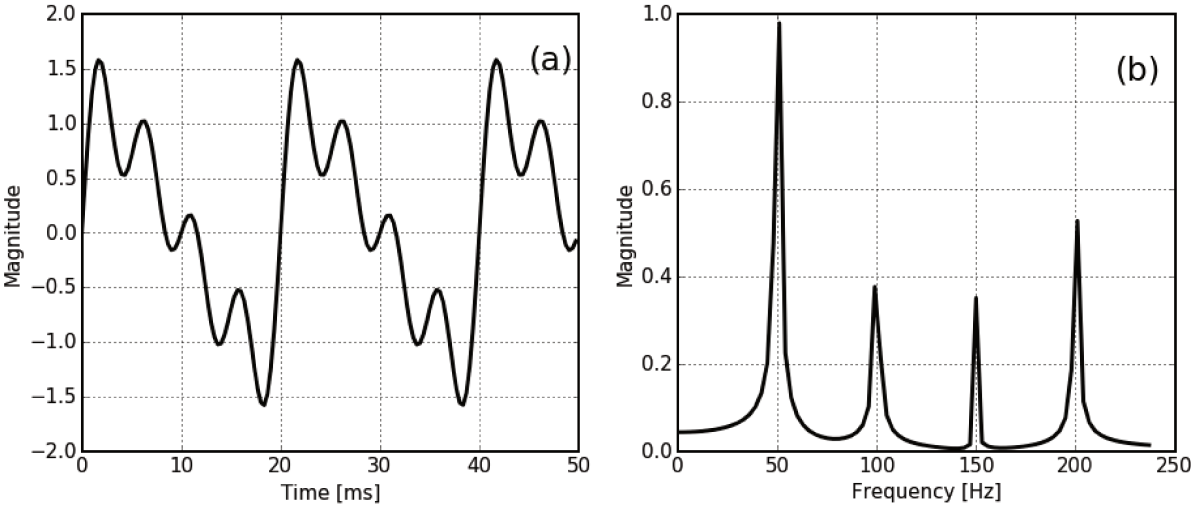

Figure 2.2 – Waveform with lower power upper harmonics.

Following (GERHARD, 2003), Figures 2.1, 2.2, 2.3, and 2.4 are some examples of waveforms and their corresponding Fourier transforms, that represent the different harmonics that compose them, in the frequency domain. These figures illustrate different difficulties in estimating their respective fundamental frequencies.

Waveforms with low harmonics or where harmonics have low power, as in Figures 2.1 and 2.2, have their periods more easily detectable because the component referring to fundamental frequency is easily detectable in the spectrum because it is unique or has an intensity higher than the upper harmonics, preventing it from being ignored by some minimum intensity limit that may be used in the detection method.

Consequently, waveforms in which harmonic power is greater than the fundamental, as in Figure 2.3, or even when the fundamental frequency is missing, as shown in Figure 2.4 make the task of estimating the fundamental frequency more complex.

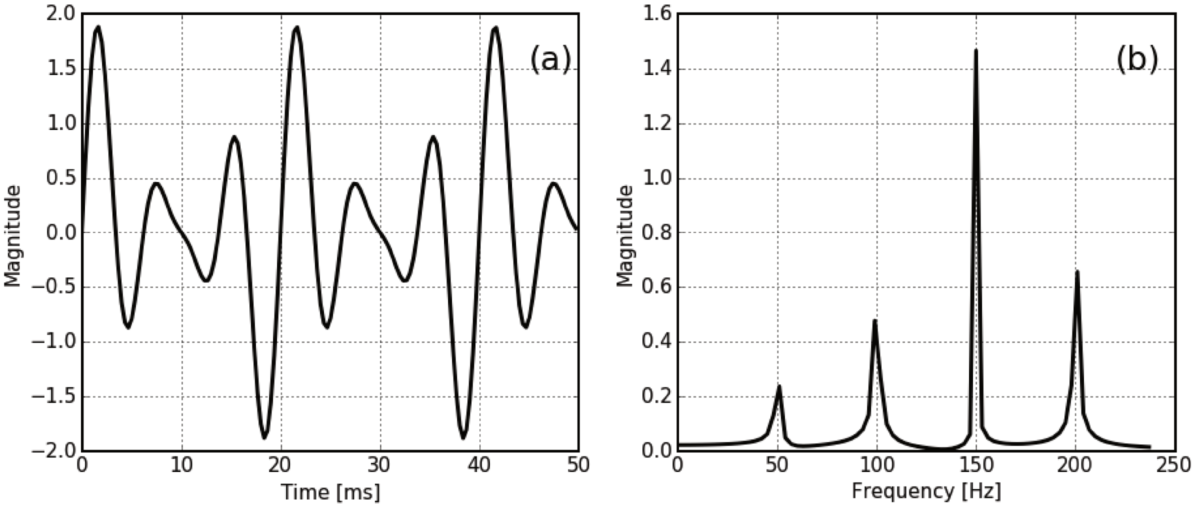

Figure 2.3 – Waveform with higher power upper harmonics.

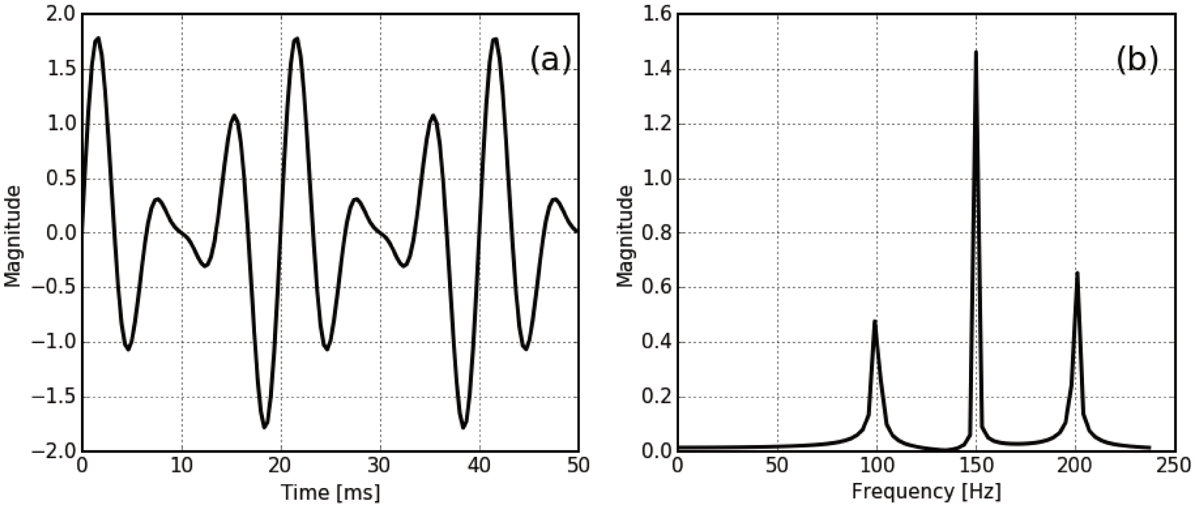

Figure 2.4 – Waveform with high power upper harmonics and no fundamental.

2.3 Related work

The signal presented in equation (2.1) is a continuous-time signal. Its digital processing is only possible after sampling (RAMA; THEODORIDIS, 2014). Representing the continuous-time signal by s_t, and the corresponding discrete-time signal by s_n, we have:

$$s_n = s_t(nT) \tag{2.3}$$

where $n \in \mathbb{Z}$ and T is a time interval called sampling period which allows us to define another important parameter for processing and analyzing a uniform discrete signal, the sampling frequency, f_s, as:

$$f_s = \frac{1}{T} \tag{2.4}$$

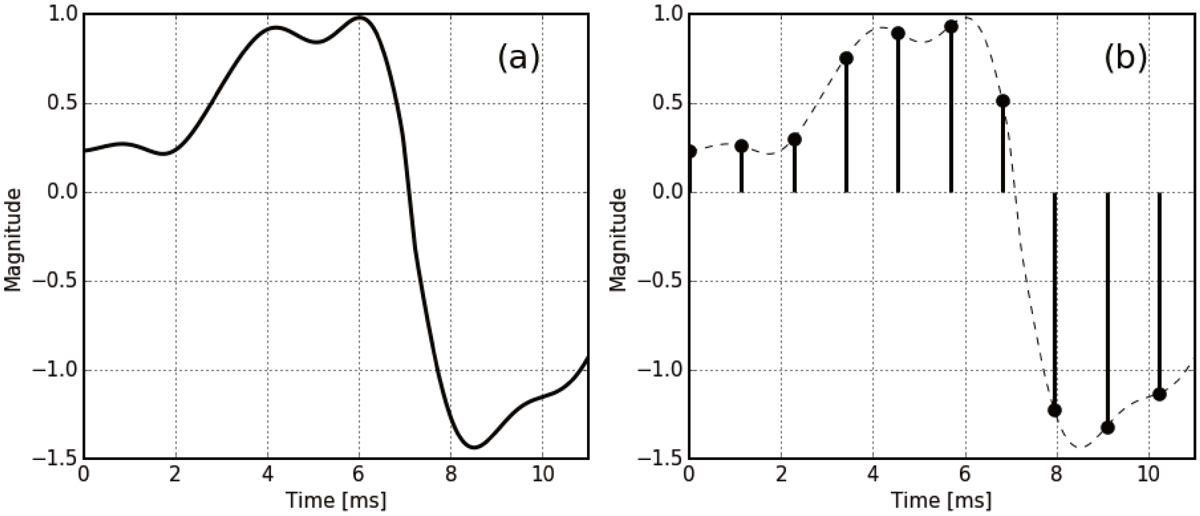

Figure 2.5 – (a) Continuous-time signal s_t. (b) The corresponding discrete-time signal s_n.

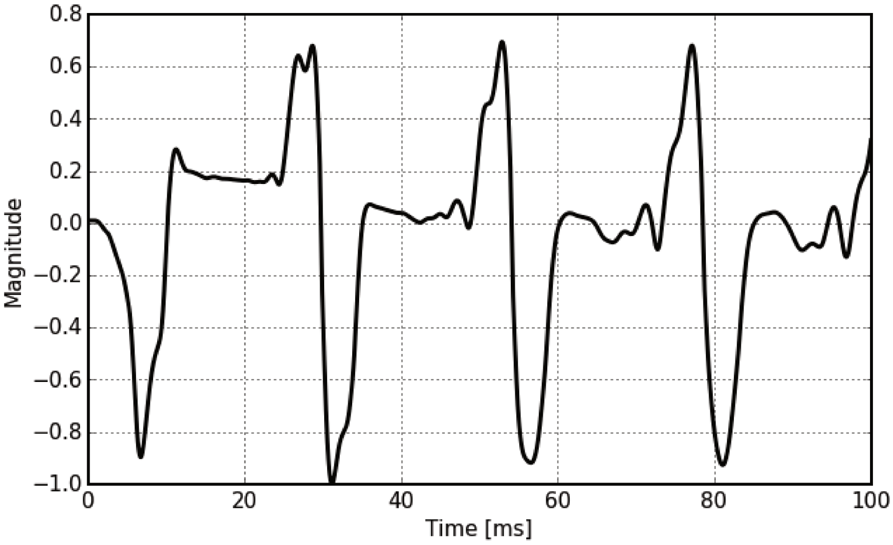

Figure 2.6 – Waveform from a electric bass guitar's recorded signal.f_0 = 44.1 Hz and T_0 = 24.2 ms. Its used as reference signal for the application of the following pitch detection methods.

As exemplified in Figure 2.5, sampling occurs at the time instances $t = nT$.

Several methods aim at finding the pitch of periodic signals, as discussed next. Some of them were implemented and applied to a reference signal, which is shown in Figure 2.6, of an excerpt from a recording of an electric bass playing the note E0, which has approximately a fundamental frequency of 41,2 Hz and a fundamental period of 24.3 milliseconds.

2.3.1 Zero-Crossing Rate (ZCR)

The Zero-Crossing rate is a low-level property of audio signals that is often used in audio classification. It can be interpreted as a measure of signal noisiness. It is very popular to discriminate whether audio is a speech or a song. However, it is widely used in a wide range of other audio applications such as musical genre classification, speech analysis, and environmental sound recognition (MITROVIć, 2010).

ZCR of a signal is the rate at which that signal varies its sign over a given time interval or frame. Thus, this determines how many times the signal has varied from positive to negative and negative to positive, divided by the total frame size (GIANNAKOPOULOS T., 2014). The following equation defines the zero-crossing rate:

$$Z_n = \frac{1}{2W_L} \sum_{n=1}^{W_L} |\operatorname{sign}[s_n] - \operatorname{sign}[s_{n-1}]|, \tag{2.5}$$

where W_L is the frame lenght, s_n is the input signal to be analyzed and sign[] is the sign function, i.e.

$$\operatorname{sign}[s_n] = \left\{ \begin{array}{ll} 1 & \text{, if } s_n \geq 0, \\ -1 & \text{, if } s_n < 0, \end{array} \right\} \tag{2.6}$$

The ZCR is also used for determining the fundamental frequency. The basic idea is that the zero-crossing rate should be directly related to the number of waveform repetitions, but problems arise in determining f0. For a waveform without or with few higher harmonics, as in Figure 2.7 (a), the signal crosses zero only twice. In the case of the waveform of Figure 2.7 (b), due to the strong presence of higher frequency harmonics, the signal crosses zero several times. Thus it is difficult to determine a zero-crossings pattern that allows us to estimate f0. To try to avoid this problem, the ZCR fundamental frequency detector can use a low pass filter, but this naturally implies a delay in the signal, affecting the overall process latency (GERHARD, 2003).

2.3.2 Autocorrelation

It is possible to measure the similarity between two signals using the correlation function, which compares and determines the similarity of two waveforms at different intervals. It presents a function that shows how similar two similar signals are for different intervals between the start of the two waveforms. Autocorrelation is the application of the correlation between a waveform and itself and is defined by the following equation:

$$r_t(\tau) = \sum_{n=t+1}^{t+W_L} s_n s_{n+\tau} \tag{2.7}$$

The autocorrelation $r_t(\tau)$ is a measure of the similarity between the signal s_n and a temporally shifted version $s_{n+\tau}$ of itself analyzed over a window with length W_L.

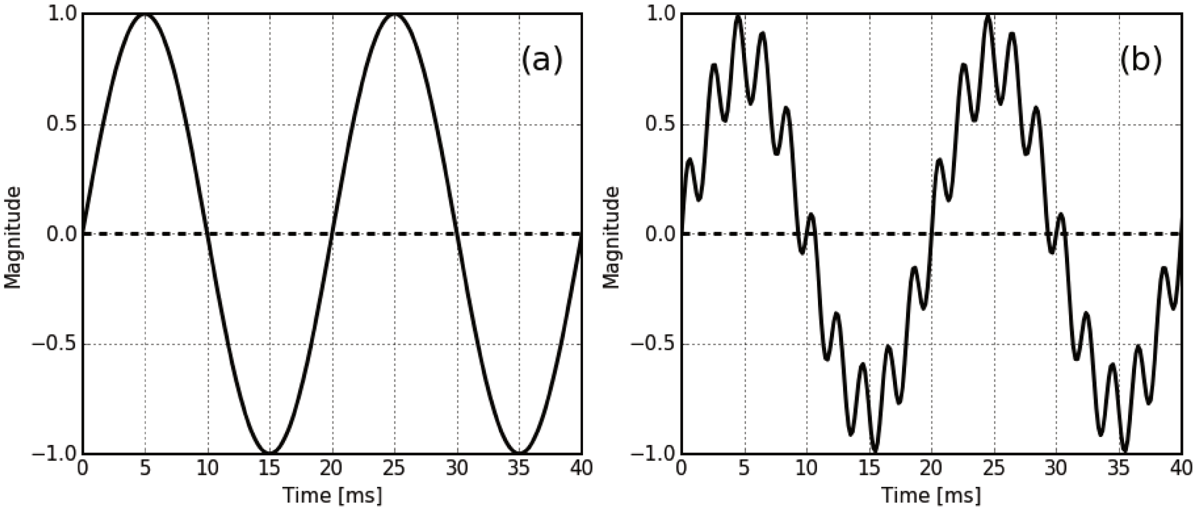

Figure 2.7 – Influence of higher harmonics - (a) Waveform whitout harmonics. (b) Waveform with higher harmonics

A common method for estimating pitch of periodic signals is by detecting the greatest positive peak of the autocorrelation function r_t (Rabiner, 1977), as it presents peaks in values of τ that correspond to the fundamental periods of s_n. The fundamental frequency f_0 is calculated by:

$$f_0 = \frac{1}{\tau_{max}}, \tau > 0, \tag{2.8}$$

tal que:

$$r_t(\tau_{max}) = max r_t(\tau) \tag{2.9}$$

The autocorrelation function, when applied to a periodic waveform, is also periodic, showing maxima when the time lag τ is equal to or multiple of the fundamental signal period and minimums when it is close to half of the period, as can be seen in Figure 2.8, which shows the autocorrelation function obtained from the reference signal, shown in Figure 2.6.

2.3.3 Maximum likelihood in time domain

According to Noll (1970), maximum likelihood is the optimal way to detect the fundamental frequency in the time domain. The main idea is to find the interval in which an input signal can be divided repeatedly so that each of these sections is as similar as possible (NOLL, 1970).

The periodic signal to be analyzed s_n, which has length T_s, is divided into N_s segments of length τ ($\tau < T_s$), as shown in Figure 2.9. Then these segments are added

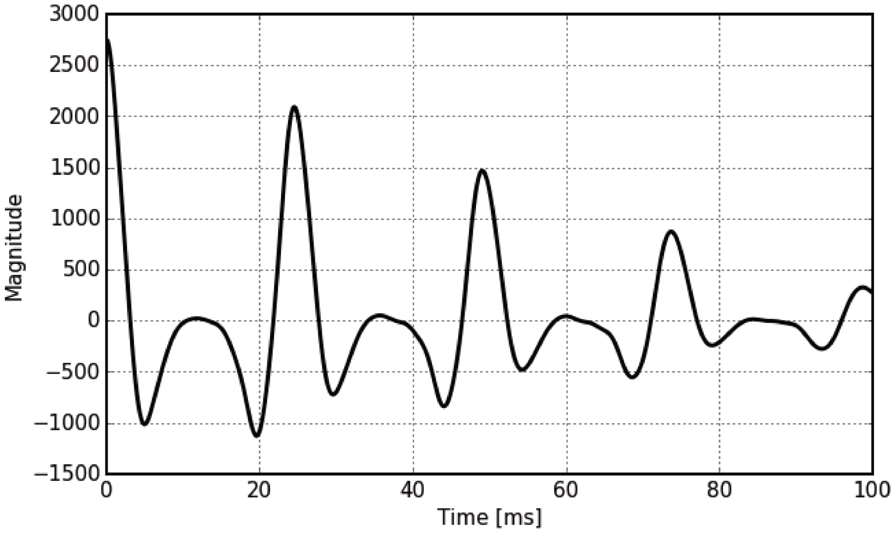

Figure 2.8 – Autocorrelation function calculated from the waveform in Figure 2.6. First peak after the initial one occurs near $24.2ms$, as expected.

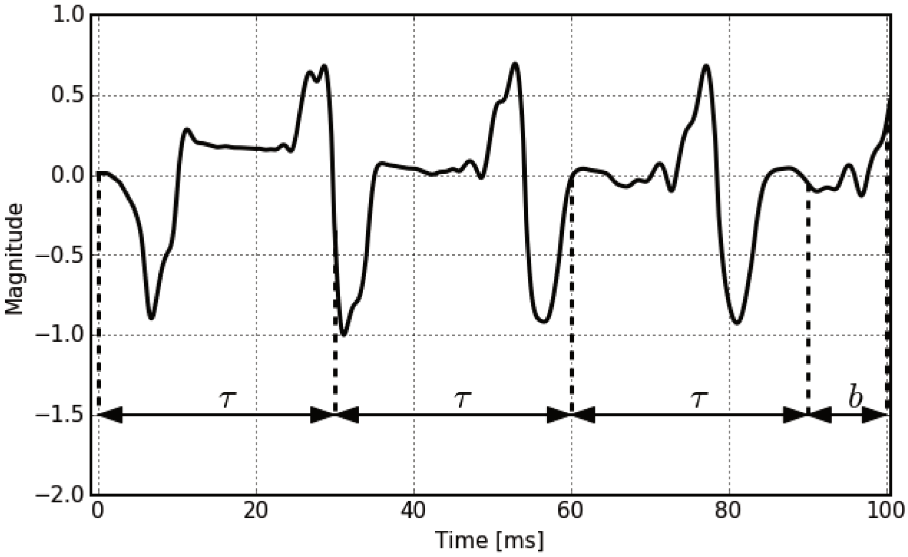

Figure 2.9 – Finite lenght T_n of a quasi-periodic signal splited in $N_s = 3$ intervals of length τ so that $T_n = N_s\tau + b$.

together according to the following equation:

$$Sml_{n,\tau} = \left\{ \begin{array}{ll} \dfrac{1}{N_s+1} \sum\limits_{n_s=0}^{N_s} s_{n+n_s\tau} & \text{, if } 0 \leq n < b \\ \dfrac{1}{N_s} \sum\limits_{n_s=0}^{N_s-1} s_{n+n_s\tau} & \text{, if } b \leq n < \tau \end{array} \right\} \tag{2.10}$$

Finally, considering the different ranges in proportion to the number in the

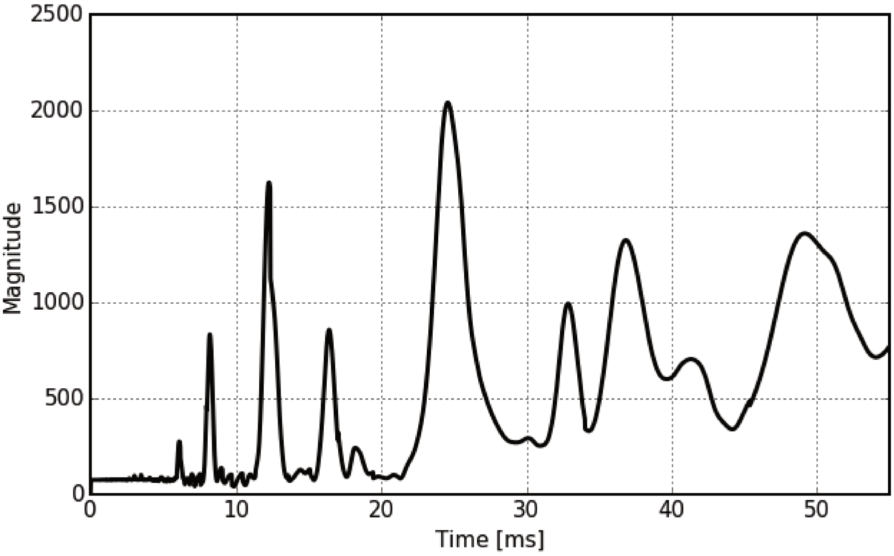

Figure 2.10 – Maximum likelihood function J_τ calculated from the waveform in Figure 2.6. Bigger peak occurs near 24.2 ms, as expected.

sum, is calculated the following integral of the square of $Sml_{n,\tau}$:

$$J_\tau = (N_s + 1) \int_0^b Sml^2_{n,\tau} dn + N_s \int_b^\tau Sml^2_{n,\tau} dn. \tag{2.11}$$

When τ is equal to the fundamental period T_0, all small intervals will add up coherently so that J_τ will be maximum when $\tau = T_0$. Thus, the maximum likelihood estimate for the fundamental period is the value of τ that maximizes the J_τ function, as shown in Figure 2.10.

2.3.4 Spectral peak picking

The fundamental frequency of a digital signal can also be determined in the frequency domain applying the Discrete Fourier Transform (DFT) to this input signal s_n, which is defined by the following equation:

$$S_k = \sum_{n=0}^{W_L-1} s_n e^{-i\frac{2\pi}{W_L}kn}, \tag{2.12}$$

where $k = 0, ..., W_L - 1$. Thus, we will obtain its spectrum S_k that will present peaks in the frequencies corresponding to the harmonics present in the input signal. Selecting the lowest frequency peak, we find the fundamental frequency of this signal.

A first problem encountered in this approach is the computational cost. Solving equation (2.12) presents computational complexity O(N^2), where N is the size of the data set.

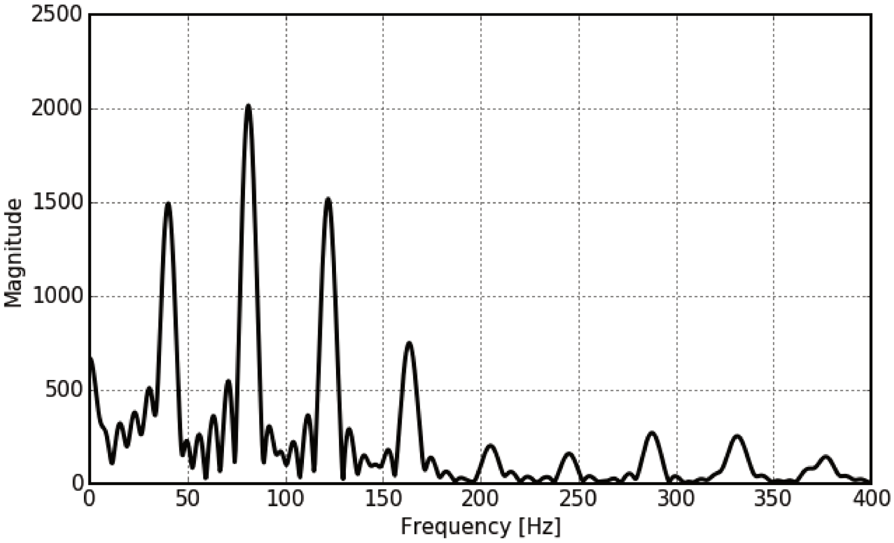

Figure 2.11 – FFT calculated from the waveform in Figure 2.6. First bigger peak occurs near 41.2 Hz, as expected.

In practice, in order to reduce this complexity to O(NlogN), fast algorithms called Fast Fourier Transform (FFT) are used. Some best-known examples are the Cooley-Tukey, Danielson-Lancsoz, and Winograd algorithms (OPPENHEIM; SCHAFER, 1999). Figure 2.11 present the resulting spectrum from FFT application to reference signal presented in Figure 2.6.

2.3.4.1 Increasing DFT resolution

Simply applying the discrete Fourier transform as a pitch estimator will present a resolution problem for low frequencies (KNESEBECK; ZOLZER, 2010). The resolution of the DFT is given by:

$$\Delta_f = \frac{f_s}{W_L}, \tag{2.13}$$

where W_L is the frame size of the transform and f_s is the sampling frequency of the input signal s_n. Thus, the DFT points are linearly distributed in the frequency domain with an interval of Δf. In a common example, where we have $f_s = 44.1kHz$ and $W_L = 2048$, there will have a resolution of approximately 21.53Hz, which shows that the method would be unable to correctly determine if a played note was $E0(41.20Hz)$ or $F0(43.6Hz)$, as the difference between the fundamental frequency of both is approximately $2.4Hz$.

(KNESEBECK; ZOLZER, 2010) demonstrates how to use the phase information to increase DFT resolution as follows:

Each harmonic component, represented by a point in the DFT, can be defined as:

$$s_{nh} = \cos(\Omega_0 n + \varphi_0) = \cos(\phi_n), \tag{2.14}$$

where the fundamental frequency is:

$$\Omega_0 = k_0 \frac{2\pi}{W_L}. \tag{2.15}$$

Therefore the fundamental frequency also can be represented as the derivative of the cosine argument ϕ_n by n,

$$\Omega_0 = \frac{d\phi_n}{dn}, \tag{2.16}$$

Using the phase difference $\Delta\phi_n$ between two discrete Fourier transforms from the same signal shifted of R samples, this derivation can be approximated as:

$$\hat{\Omega}_0 = \frac{\Delta\phi_n}{R}. \tag{2.17}$$

To estimate the corrected fundamental frequency of an input signal that contains harmonics is to select, for a given frame, the pitch candidate with the lowest corrected frequency. The following equation is used to calculate the corrected frequency for pitch candidates:

$$\hat{f}_0 = \frac{1}{2\pi}\hat{\Omega}_0 f_s = \frac{1}{2\pi}\frac{\varphi_{2u} - \varphi_1}{R} f_s. \tag{2.18}$$

where φ_1 is the phase angle at bin k_0 of the first DFT.

The real unwrapped phase φ_{2u} is calculated as sum of the expected phase and the phase error:

$$\varphi_{2u} = \varphi_{2t} + \varphi_{2err} \tag{2.19}$$

The expected phase φ_{2t} after a progression of R samples can be calculated with:

$$\varphi_{2t} = \varphi_1 + \frac{2\pi}{W_L} k_0 R \tag{2.20}$$

and the phase error φ_{2err} is given by:

$$\varphi_{2err} = \varphi_2 - \varphi_{2t}, \tag{2.21}$$

where φ_2 is the phase angle at bin k_0 of the second DFT.

Therefore, by calculating the expected phase after a progression of R samples, the frequency resolution of a pitch candidate at frequency $k0$ can be improved.

2.3.5 Cepstrum method

Basically, Cepstrum is the spectrum of the log spectrum and its name comes from the inversion of the first four letters of the word "spectrum" referring to the new transformation applied to the spectrum of an original signal.

The cepstrum of a signal is defined by:

$$c_n = F^{-1}\{\log F\{s_n\}\} \tag{2.22}$$

Where s is the input signal, F is the Fourier transform and F^-1 is the inverse Fourier transform.

For the case where the input signal is a windowed frame of sound signal with size W_L , cepstrum is:

$$c_n = \sum_{n=1}^{W_L} \log\left(\sum_{n=1}^{W_L} s_n e^{-i\frac{2\pi}{W_L}kj}\right) e^{i\frac{2\pi}{W_L}kn}. \tag{2.23}$$

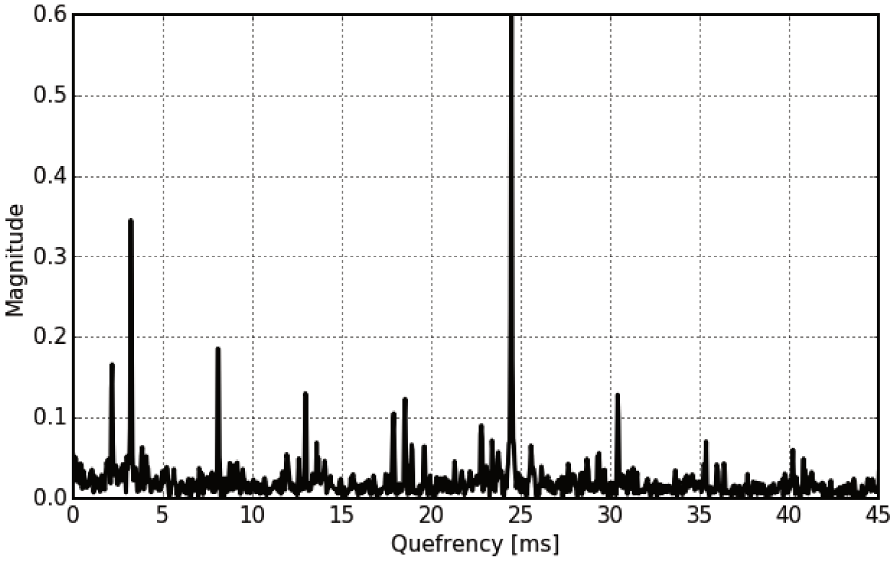

Figure 2.12 – Cepstrum calculated from the waveform in Figure 2.6. Bigger peak occurs near 24.3 ms, as expected.

Initially, the Fourier transform is applied to the input signal in the time domain, obtaining its spectrum in the frequency domain. In the second step, this spectrum is converted to a logarithmic scale and then the Fourier transform is applied again, obtaining the power cepstrum, a function in a domain very related to the time domain, called quefrency. It has peak values that correspond to the period of the input signal, as shown in Figure 2.12. Briefly, an attempt is made to determine the frequency of the input signal spectrum (Singh; Kumar, 2014).

However, the method may present incorrect results if the input signals have an inharmonic spectrum or the single-partial spectrum of a sinusoid, as it assumes that the signal has regularly-spaced harmonics.

2.3.6 Harmonic product spectrum

The determination of the fundamental frequency of a signal can also be done by using the method called Harmonic product spectra, widely-used when there is a considerable presence of noise.

The main idea of this method is that the pitch peaks in the log spectrum are multiples of the fundamental frequency. Thus, adding some compressed versions of the spectrum would result in a large peak at the point of coincidence of these harmonics, as shown in Figure 2.13. This is because the pitch peaks in the log spectrum are coherently added while the rest of the log spectrum is uncorrelated and add non coherently (NOLL, 1970). Figure 2.13(a) shows the spectrum from waveform presented in Figure 2.6, Figure 2.13(b), (c) and (d) shows the compressed versions of this spectrum and Figure 2.13(e) present the resulting Harmonic Product Spectrum.

The Harmonic Product Spectrum is defined as:

$$\pi(\omega) = \prod_{m=1}^{M} S_k(m\omega), \tag{2.24}$$

where $S_k(m\omega)$ is the input signal S_n spectrum compressed by m and M is the number of compressed versions of the spectrum. After this processing, the argument of the global maximum of σ corresponds to the f0.

2.3.7 Harmonic Sum Spectrum

Frequency-compressed versions of power Spectrum also present the same coherent sum of the log spectrum, making it possible to use the Harmonic Sum Spectrum method to determine the fundamental frequency.

The Harmonic Sum Spectrum is obtained by adding the Spectrum of the input signal to its sub-sampled versions of itself (NOLL, 1970). This corresponds to calculating:

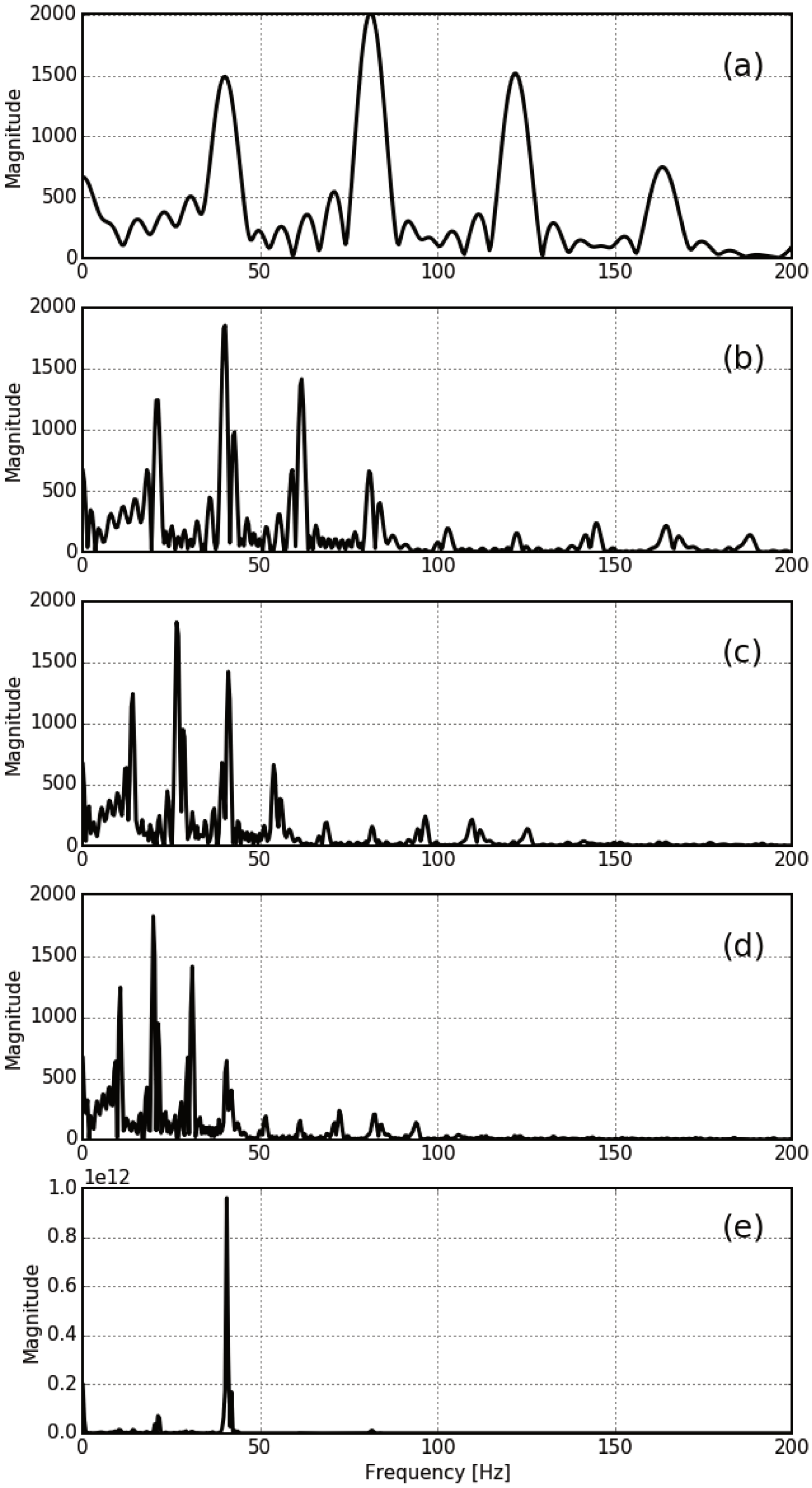

Figure 2.13 – (a): spectrum from waveform presented in Figure 2.6. (b), (c) and (d): compressed versions of the spectrum in (a). (e): Harmonic Product Spectrum calculated from the waveform in Figure 2.6. Bigger peak occurs near 41.2 Hz, as expected.Only the first 200 samples of the waveforms are represented in this figure.

$$\sigma(\omega) = \sum_{m=1}^{M} S_k(m\omega), \tag{2.25}$$

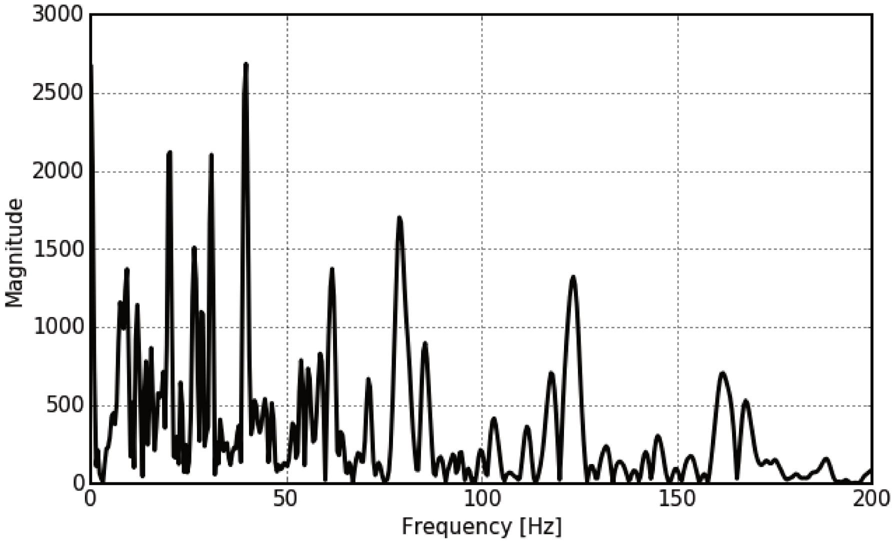

Figure 2.14 – Harmonic Sum Spectrum calculated from the waveform in Figure 2.6. Bigger peak occurs near 41.2 Hz, as expected.

where, as previously, $S_k(m\omega)$ is the input signal S_n spectrum compressed by m and M is the number of compressed versions of the spectrum. Finally, the argument of the global maximum of σ corresponds to the f0.

Although the peak generated by the harmonic product spectrum is more acute and the secondary peaks are negligible in size compared to the harmonic sum spectrum method, which has larger secondary peaks, as shown in Figure 2.14, both will show a higher peak at the fundamental frequency, which it must be detected above a fixed threshold value.

2.3.8 Darrien low latency method

Darrien proposes a low latency parametric method for pitch estimation based on the Exponentially Damped Sinusoidal (EDS) model and a statistical model.

The method consists of two stages: initially, the most significant harmonic components are selected in the input signal according to the Exponentially Damped Sinusoidal (EDS) model. In the second stage, a statistical model is used to determine the most likely fundamental frequency (DERRIEN, 2014).

Using the Exponentially Damped Sinusoidal (EDS) model, the input signal is described by:

$$s_n = d_n + w_n, \tag{2.26}$$

with d_n being the deterministic part and w_n the stochastic part. d_s is a sum of K damped sinusoids described by:

$$d_n = \sum_{k=0}^{K-1} \alpha_k z_k^n. \tag{2.27}$$

The complex amplitude α_k is:

$$\alpha_k = a_k e^{i\phi_k} \tag{2.28}$$

where a_k is the initial amplitude and ϕ_k is the phase.

The poles z_k is:

$$z_k = e^{-d_k + 2i\pi\nu_k} \tag{2.29}$$

where d_k is the damping and ν_k is the normalized frequency.

The stochastic part, w_n, is white Gaussian noise.

The model parameters are estimated with a method derived from the ESPRIT algorithm (BADEAU; DAVIDWANG, 2002). This phase is similar to spectral analysis and a peak-picking stage, presented earlier in this work.

The estimation algorithm consists in finding the best values of k, α_k and z_k for a given signal in the least square sense, using an estimation algorithm proposed by Badeau and Wang (2002).

Then a statistical model is used to determine the most likely fundamental frequency within the set of partials obtained in the previous step. This is done by calculating the likelihood function of the fundamental frequency based on the probabilistic model of the partials found so that the global maximum of this function is the best estimate for the fundamental frequency. The probabilistic model used is a modification of that proposed by B. Doval and X. Rodet (1991), with fewer parameters, not depending on the learning database, as the original.

2.3.9 Yin method

Autocorrelation, presented earlier, commonly peaks not only with each waveform repetition but also due to the harmonics present in the signal. This creates difficulties for fundamental frequency estimators that use autocorrelation, as they are eventually unable to determine if a peak is relative to the fundamental frequency or signal harmonics.

The Yin method was proposed by Cheveigné and Kawahara. It is based on the same assumptions as of the autocorrelation method, with the addition of a series of modifications that reduce errors. The name of the method (Yin) alludes to the Yin and

Yang of Eastern philosophy, alluding to the search for the balance between autocorrelation and cancellation proposed by the method to reduce errors.

The method consists in the application of 6 steps that reduce the error rate in the fundamental frequency estimation (CHEVEIGNé; KAWAHARA, 2002). Next, we briefly describe the improvements applied to each step according to the authors' study information.

2.3.9.1 Step 1: The autocorrelation method

In the first step, the method uses autocorrelation, presented in the previous subsection, obtaining an error rate of 10 % in the estimate of f0 when applied to the database presented in the study of its authors. As shown in the next steps, autocorrelation will no longer be used by the method.

2.3.9.2 Step 2: Difference function

In the second step of the method, the autocorrelation function is replaced by the difference function, reducing the error rate to 1.95%. Here the period is no longer defined by the largest peak, but by the largest dip in the function. A possible cause for this reduction would be the high sensitivity of autocorrelation to amplitude changes, so that, increases in signal amplitude lead the method to choose correlation function peaks from harmonics rather than fundamental ones. Figure 2.15 presents the difference function calculated from the waveform of Figure 2.6. The difference function is defined by the equation:

$$d_t(\tau) = \sum_{n=1}^{W_L} (s_n - s_{n+\tau})^2. \tag{2.30}$$

where s_n is the input signal and $s_{n+\tau}$ a τ samples shifted version of itself analyzed over a window with length W_L.

2.3.9.3 Step 3: Cumulative mean normalized difference function (CMNDF)

In the third step, the difference function is replaced by the cumulative mean normalized difference function reducing the error rate a little more. As can be seen in Figure 2.16, unlike the difference function, which starts at 0, the (CMNDF) starts at 1, eliminating the need for an upper frequency limit. This limit is required when the difference function is used, so that the first dip does not be selected as the fundamental

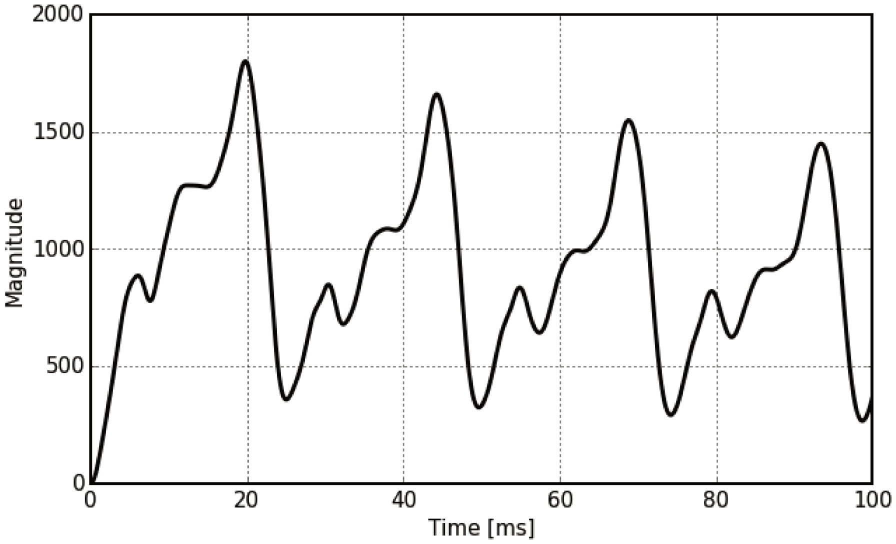

Figure 2.15 – Difference function calculated from the waveform in Figure 2.6. First big dip after the initial one occurs near $24.2ms$, as expected.

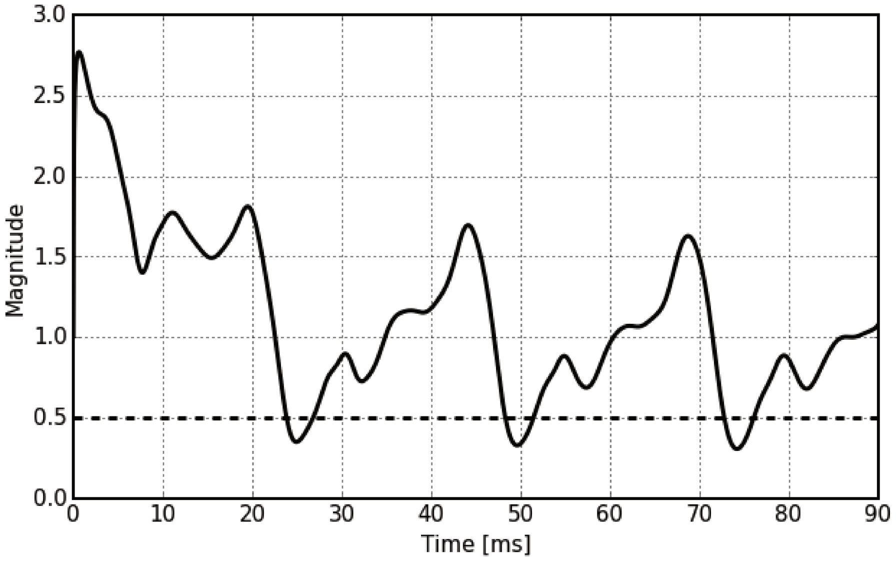

Figure 2.16 – Cumulative mean normalized difference function calculated (CMNDF) from the waveform in Figure 2.6 and an absolute threshold on dashed line. First big dip occurs near $24.2ms$, as expected.

frequency dip. The (CMNDF) is defined by:

$$d_t^{'}(\tau) = \left\{ \begin{array}{cc} 1 & \text{, if } \tau = 0 \\ \dfrac{d_t(\tau)}{(1/\tau)\sum\limits_{n=1}^{\tau} d_t(n)} & \text{, otherwise} \end{array} \right\} \tag{2.31}$$

where $d_t(\tau)$ is the diference function defined in equation (2.30) and τ the lag in samples between the signal and the shifted version of itself in diference function.

2.3.9.4 Step 4: Absolute threshold

The fourth step is the use of an absolute threshold that decreases by approximately half the error rate obtained in the previous step, which generates a normalized function. This Absolute threshold is represented by the dashed horizontal line in Figure 2.16. Using this threshold, dips above this value are disregarded, avoiding the selection of harmonic generated dips.

2.3.9.5 Step 5: Parabolic interpolation

In the fifth step, a parabolic interpolation of the minimum location is included, but the reduction in the error rate is minimal. The idea is that this reduces the error when the period is not a multiple of the sampling period which could lead to an error of up to half of the sampling period.

2.3.9.6 Step 6: Best local estimate

In the sixth step a new estimate is made, but now only in the vicinity of the location indicated by the first estimate in order to find the best local estimate. Seeking around 20% variation around the initial estimate, we obtained a reduction of approximately 1/3 in the error rate compared to the previous step.

Version	**Error rate** (%)
Step 1	10
Step 2	1.95
Step 3	1.69
Step 4	0.78
Step 5	0.77
Step 6	0.50

Table 2.2 – Error rates after application of each step of Yin method.

According to the study of Cheveigné and Kawahara (2002), the error rates obtained by the Yin method are about one-third times lower than the best competing methods, as evaluated over a database of speech recorded together with a laryngograph signal. The error rates at each step are shown in table 2.2.

2.4 Discussion about pitch estimation methods

All the methods discussed in the previous section directly rely on the periodicity property as stated in Equation (1.1) or the harmonic series model shown in Equation (2.1). This allows them to be applicable for the general case of finding pitch in periodic signals but bounds them to a minimum delay of twice the fundamental period.

In this work, we propose a pitch detection method that relies on specific characteristics of the plucked electric bass string. This restricts our method to signals generated by this specific instrument. However, it allows reducing the delay to 1.1 times the fundamental period, which is very close to the theoretical minimum latency.

This reduction is critical for real-time pitch detection in lower-pitch notes. In this range of notes, general-purpose methods require a delay of around 50ms to work properly. Our method allows detecting the same pitch with a delay of around 30ms.

The method proposed by (DERRIEN, 2014) also indicates to estimate f0 close to the theoretical minimum latency, i.e. the fundamental period of the lowest observable pitch, but with higher computational complexity, which can be problematic for embedded real-time applications, which can lead to an increase in delay due to computational cost.

The proposed method is based on specific properties of the plucked electric bass signal. These properties are analyzed using a physical model, which guides its generalization possibilities. Then, the proposed model is compared to the Yin method using a dataset containing recordings from electric bass guitars.

For comparison purposes, the Yin method was chosen as the reference method. In addition to presenting excellent performance as shown in (CHEVEIGNé; KAWAHARA, 2002) study, it is commonly used as a reference method, as in the study by (DERRIEN, 2014), addressed in this work. It was also chosen because it is a well-known and cited method, as in the works of (GERHARD, 2003) and (KNESEBECK; ZOLZER, 2010), also cited in this work, counting more than 1300 citations according to the portal (RESEARCHGATE, accessed 06/02/2020).

2.5 Latency

The human perception of sounds is very sensitive to its temporal characteristics. Therefore, audio delays are experienced in many different scenarios and for many different reasons and is called latency. In the context of this work, the sound delay refers to the time elapsed between an initial event, such as playing a note on the electric bass guitar, for example, and a second event, such as the moment when the sound is perceived by a specific listener.

When you hear the sound from a sound source a few meters away, there is a delay due to the amount of time it takes for this sound to travel through space over that distance. For example, in a room with a temperature of 20°C, the speed of sound is approximately 323.3 meters per second, which causes a delay of 2.91 milliseconds per meter of distance between the sound source and the listener. This delay or the delay between two sound events can be large enough to be noticed and often causing several negative effects.

In music applications latency can be a very serious problem as it directly impacts musicians' performance in many ways, making it difficult to maintain steady tempo, rhythmic synchronism between musicians and even tuning depending on the instrument (GREEF, 2016).

2.5.1 Causes for Latency

There are many causes of unwanted delays. In orchestras, for example, musicians on opposite sides can experience latencies of up to 80 milliseconds due to the time it takes the sound to propagate through the distance between one musician and another. Nowadays in most current performances, musicians use close speakers and headsets as feedback, most of the latency comes from processing audio signals (GREEF, 2016).

Digital processing of an instrument's audio signal implies a series of delays, starting with converting the analog to a digital signal at the system input and from digital to analog at the output. Buffering digital samples and phase delay of digital filters also add latency. Finally, the time required for processing the audio samples according to the applications used (WANG, 2017).

In the case of audio-to-midi converters, besides the time spent performing the algorithm operations to determine the fundamental frequency of the signal, there is still the necessary interval from the onset of a note played on the instrument for the algorithm to estimate what is the fundamental frequency. Most f0 estimators need at least two periods to accomplish its task.

2.5.2 Tolerable Latency

The perception of how much a certain amount of latency bothers, hinders, or even precludes the correct use of the instrument by the musician depends on the type of instrument played and also on the musician's listening skills. For example, musicians such as professional saxophonists are more affected by latency and need more immediate feedback, considering a latency of up to 10 milliseconds as acceptable, while keyboard players have a higher latency tolerance, considering latencies of up to 40.5 milliseconds as acceptable (LESTER; BOLEY, 2007).

A previous study (LESTER; BOLEY, 2007) has investigated the acceptable latency in live sound applications for different professional musicians using in-ear monitoring (IEM) or wedge monitoring. The results of this study are presented in Table 2.3.

Latency (ms)	Sax	Vocals	Guitar	Drums	Bass	Keys
IEM Good	0	1	4.5	8	4.5	27
Wedge Good	1.5	10	6.5	9	8	22
IEM Fair	3	6.5	14.5	54.5	25.5	46
Wedge fair	10	26	16	25	30	40.5

Table 2.3 – Tolerable latency - Instruments comparison using in ear monitoring (IEM) and wedge monitoring (LESTER; BOLEY, 2007)

2.5.3 Latency Discussion

Table 2.3 shows that professional bassists consider a latency of up to 30 milliseconds acceptable when using wedge monitoring. However, as already seen, most algorithms require the use of at least two periods to estimate f0, and the lowest note of a traditional four-string bass, E0, has a period of 24.27 milliseconds. That is, only the algorithmic delay for these methods is at least $2 \times 24.27 = 48.54$ milliseconds.

The method proposed in this study estimates the fundamental frequency using a time interval of 1.1 times the period, starting from the note onset. For the same note E0, the algorithmic delay is $1.1 \times 24.57 = 26.697$ milliseconds, within the latency considered acceptable by professional bassists.

2.6 The electric bass guitar

The electric double bass, also known as electric bass, or simply the bass, is an electroacoustic musical instrument and can be described as a plucked box-chordophone with a solid body instead of a hollow resonator. It receives the classification 513 - Electroacoustic chordophones in the Revision of the Hornbostel-Sachs Classification of Musical Instruments (MIMO, 2011).

In both its appearance and construction, the electric bass guitar is similar to the electric guitar, but has a larger neck and the most common models have four strings, not six like the guitars and are tuned one octave lower than the four lower strings of the guitar (STRINGVIBE, accessed 12/10/2019b).

The double bass, acoustic musical instruments with approximately 180 cm and usually tuned the same way as the electric bass, have been replaced by those since the 1960s in most popular musical productions, taking on the role of the bass instrument in the rhythm section of musical ensembles.

Many musical styles rely on the presence of the electric bass, such as blues, jazz, rock, metal, pop, reggae, and many others. Although bassists play very varied melodic lines, electric basses perform a similar function in most of these musical styles.

This function is to set the rhythm and set the harmonic framework (VETTER, accessed 12/10/2019).

2.6.1 History

The first records of electric bass date back to the 1930s, when musician and inventor Paul Tutmarc created the first electric string bass, a fretted instrument designed to be held and played horizontally. His company, Audiovox, launched a first model in the year 1935 and the second one in 1947, but none of them achieved success in the market (VARIOUS, accessed 12/10/2019). In 1951, Leo Fender invented Fender Precision, the first electric bass guitar that used the technologies contained in his electric guitars. This model brings to the bass the magnetic pickups and the guitar-shaped body with fretted fingerboard features. The Fender precision is considered a revolutionary instrument. After all, it was easy to play as it contained frets, easier to carry because they were much smaller than a double bass, and could have their volume easily adjusted because they were directly connected to the amplifiers. From that moment on there were no drastic changes in the design of the electric bass guitars. The changes were restricted to increasing the number of pickups, adding more strings and subtle changes in body shape, as well as the natural evolution of the electronics contained in these instruments (VETTER, accessed 12/10/2019).

2.6.2 Construction

The Bass guitar is made up of several parts assembled that affect the sound produced by the instrument or how to use it. Basically, it is composed of a body, a neck, and a headstock, but will be discussed below in more detail each of its parts for a complete understanding of its operation. The following descriptions are based on articles by Porter (PORTER, accessed 12/10/2019) and Stringvibe (STRINGVIBE, accessed 12/10/2019a)

2.6.2.1 Headstock

The headstock is at one end of the bass (Figure 2.17 (l)). It is attached to the tuning machine (Figure 2.17 (g)), also called machine head, which decreases or increases the traction of the strings (Figure 2.17 (d)) by rotating the tuning pegs. Changing the traction of a string changes its pitch.

2.6.2.2 Nut

The nut (Figure 2.17 (f)) is a rectangular block, most commonly made of plastic or synthetic ’bone’, positioned between the headstock and the neck (Figure 2.17

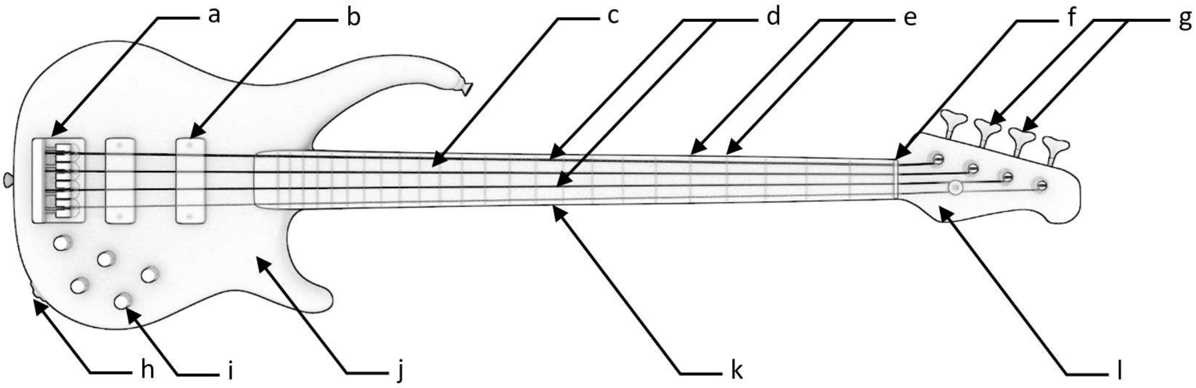

Figure 2.17 – The traditional 4 strings electric bass guitar: *a* body, *b* neck, *c* head

(k)). It keeps the strings evenly spaced and also helps to adjust their height relative to the rest of the instrument.

2.6.2.3 Neck

The neck (Figure 2.17 (k)) is made of wood, most commonly maple, and connects the headstock to the body (Figure 2.17 (j)) of the instrument. It is usually attached to the body using four screws, but there are bass models in which the body and neck are made of a single piece of wood, which helps increase the sustain of the played notes. It is very common for basses to have a solid bolt running through the neck, which helps to prevent warping due to string traction.

2.6.2.4 Fretboard

The fretboard (Figure 2.17 (c)) is a layer of wood, commonly rosewood, attached to the front surface of the neck. Mounted on it are several small metal bars called frets (Figure 2.17 (e)) that divide it into different tonal areas along its length. When a string is pressed over an area of the fretboard, it will touch at least one of these frets and the one closest to the instrument body will set the active width of the string. This ensures that the notes you play are in tune.

Most basses have between 20 and 24 frets, although there are models with up to 32 frets, usually used by soloists who need a longer range of notes. There are also models without frets called fretless that feature the more muffled sound.

2.6.2.5 Body

The Electric bass guitar body is a solid piece of wood, usually Maple or Alder. Its most traditional format is rounded along the outside with two curved horns on either side of the neck. Its main function is to serve as a point of attachment for other parts and pieces of the musical instrument.

In it are installed all the electronics of the instrument such as pickups (Figure 2.17 (b)), switch knobs, control knobs and volume pots (Figure 2.17 (i)) and output jack (Figure 2.17 (h)). Also, other structural mechanical components are attached to the body, such as the bridge (Figure 2.17 (a)) and strap pins, which make it possible to use a strap to support the bass, making it possible to play it while standing.

2.6.2.6 Pickups

Pickups (Figure 2.17 (b)) are bars, usually with plastic surfaces, fixed to the center of the body, and under the strings. Inside are magnets surrounded by a coil of copper wire. String vibration by the magnetic field induces an electrical voltage at the coil terminals, which is sent to the amplifier via cables connected to the output jack.

By changing the position of the pickups it is possible to change the tone of the instrument, so many basses have more than one pickup, making it possible to adjust the tone by mixing and controlling the contribution of each pickup to the output signal.

There are also non-magnetic pickups, such as piezoelectric and optical pickups that allow the use of non-ferrous strings, such as nylon. However, they are not very usual.

2.6.2.7 Bridge

The bass guitar bridge (Figure 2.17 (a)) is made of metal and is fixed to the base of the body by screws. Its function is to anchor the strings to the body through holes in its base through which each string is strung through. You can adjust the string height using the saddle height adjustments. Each string goes through one of these saddles, which contains a groove in the center where the string rests. They are movable and adjustable by screws

2.6.2.8 Strings

Each electric bass string (Figure 2.17 (d)) is named by the note they produce when they are played open, that is, without being pressed against the fretboard. Starting from the top of the fretboard and moving downwards, the strings are called E, A, D, and G. The strings can be made of different materials like Nickel-plated steel, Stainless Steel, and others. They are usually medium-caliber, so in a four-string counter bass pattern, they are between 0.045 to 0.105 inches thick (STRINGVIBE, accessed 12/10/2019b).

Frets and its respective notes

string	0	1	2	3	4	5	6	7	8	9	10	11
	12	13	14	15	16	17	18	19	20	21	22	23
G	G	G#	A	A#	B	C	C#	D	D#	E	F	F#
D	D	D#	E	F	F#	G	G#	A	A#	B	C	C#
A	A	A#	B	C	C#	D	D#	E	F	F#	G	G#
E	E	F	F#	G	G#	A	A#	B	C	C#	D	D#

Table 2.4 – Frets and its repective notes for each string on the bass guitar in standard EAGD tuning

2.6.3 Tuning

The standard tuning for the strings of the four strings electric bass guitar is $E0$ (41.20 Hz) for the E string, $A0(55Hz)$ for the A string, $D1(73.41Hz)$ for the D string and $G1(97.99Hz)$ for the G string. The frequency notes for each fret of these string are following the rule presented in equation (2.2) and its respective notes are presented in table 2.4.

2.6.4 Playing techniques

Usually, bassists play the electric bass guitar while standing and the instrument is played horizontally across the body. The fingers of the left hand press the string against the fretboard so that it touches the fret, defining the note that will be played. The right hand will excite the string, i.e., it will start its vibration, either using the fingers or a pick.

The pick is held between the thumb and index finger and used to downstroke or upstroke the string with the motion supplied by the wrist. It is used to achieve greater speed, a more articulate attack, or by personal preference and is more used among bassists who play rock, punk rock, and metal.

The most traditional and characteristic way of playing the bass is using the fingers of the right hand, a technique known as pizzicato. Usually, the bassist uses the index and middle fingers alternately to pluck the strings, although the other fingers are also sometimes used. The thumb can be used to support the right hand, resting over the edges of the pickup. The strings can be plucked from the bridge to the fret where it is pressed. Near the bridge, the sound has more brilliance, with a greater presence of harmonics while in medium positions a mellow tone is obtained.

3 Methodology

3.1 Time-domain Behavior of a Plucked String

This section discusses the properties of the plucked string signal that were used as a basis for our f0 estimation method. These properties were inferred by analyzing the audio signal of an electric bass string, as shown in Section 3.1.1, then the physical model discussed in Section 3.1.2 was used to generalize these results, as shown in Section 3.1.3.

3.1.1 Plucking an Electric Bass String

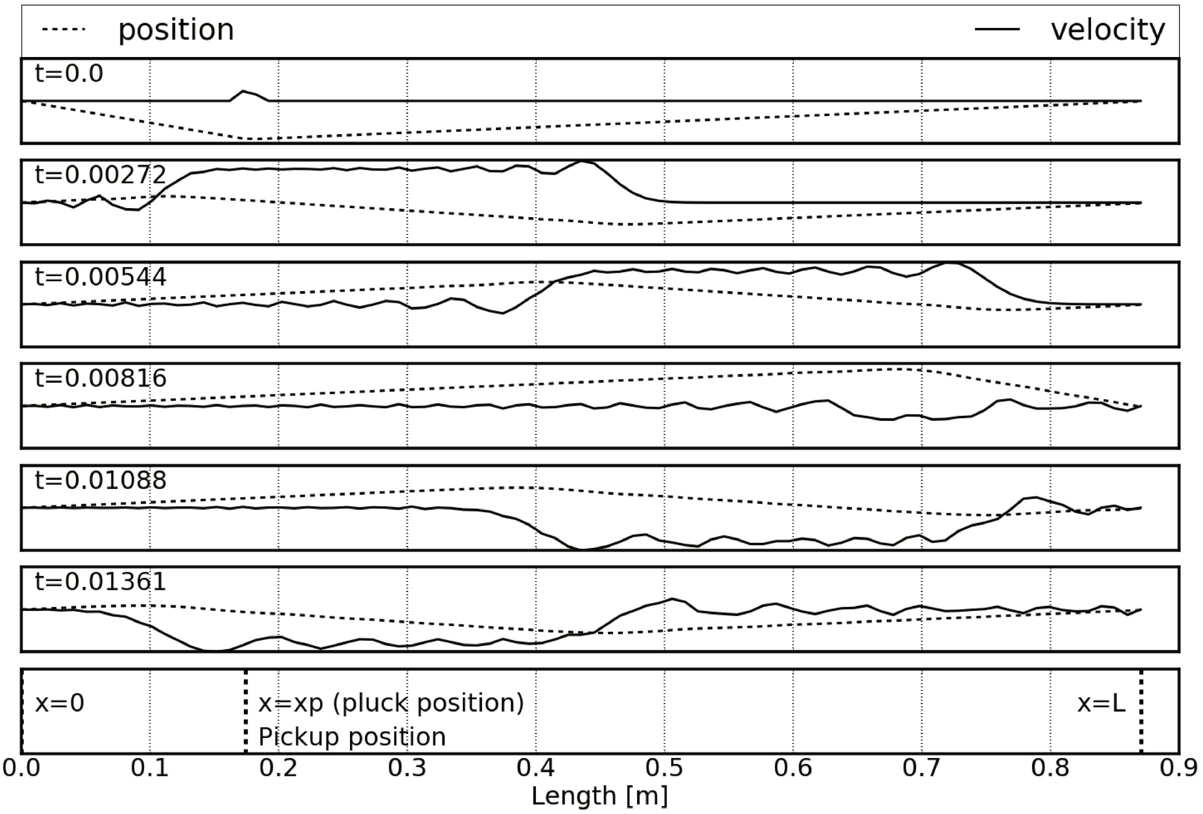

Figure 3.1 – Position and velocity of a string along the x axis at different times t with fixed ends at $x = 0$ and $x = L$

There are magnetic pickups placed on the body of the instrument, under the strings. They convert the string transverse velocity at its position into an electric voltage. The string transverse velocity can be seen as a wave that propagates from the pluck position along the string length, reflecting and inverting when reaching the string end, as shown in Figure 3.1.

The waveform of the voltage signal at the pickups, as shown in Figure 3.4, indicates repetitions of a peak (positive or negative) at the beginning of each cycle. In order to confirm that this characteristic is maintained for all the electric bass guitars (instead of being a characteristic of the specific instrument), the behavior of its string was mathematically modeled, as discussed in the next section.

3.1.2 Physical model

The behavior of the bass string can be modelled using an ideal string along the coordinate x with fixed ends at $x = 0$ and $x = L$ with a transversal displacement along the coordinate y, which give us the following boundary conditions:

$$y(x = 0, t) = 0. \tag{3.1}$$

$$y(x = L, t) = 0. \tag{3.2}$$

The string has linear density μ and is stretched with a force F_T. It is initially at rest and is plucked in the position $x = x_p$ with amplitude $y(x_p, 0) = A$ as depicted in Figure 3.2. In this situation, the initial transverse displacement $y(x, 0)$ can be expressed by

$$y(x, t = 0) = \left\{ \begin{array}{cc} A(\frac{x}{x_p}) & \text{, if } x < x_p \\ A(1 - \frac{x - x_p}{L - x_p}) & \text{, otherwise} \end{array} \right\} \tag{3.3}$$

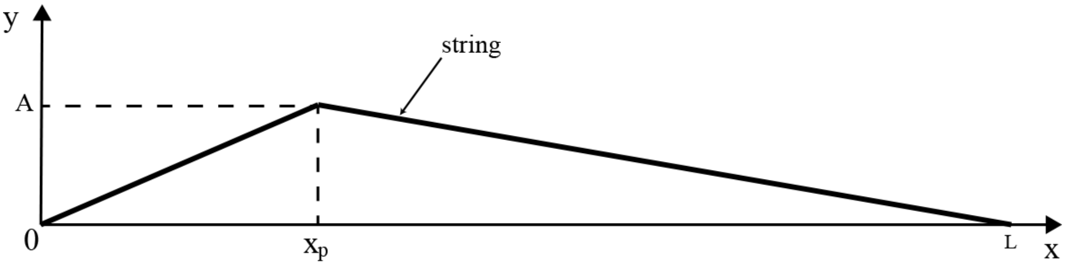

Figure 3.2 – String with fixed ends at $x = 0$ and $x = L$ being plucked at $x = x_p$ with transversal displacement $y(x_p) = A$.

Initially, the velocity distribution $y'(0, x)$ is:

$$y'(x, t = 0) = 0. \tag{3.4}$$

As depicted in Figure 3.3, for a short segment of this string between x and Δx there is a slope $\delta y/\delta x = \tan(\theta)$ and a vertical force F defined by:

$$F = F_T \sin(\theta(x + \Delta x)) - F_T \sin(\theta(x)) \tag{3.5}$$

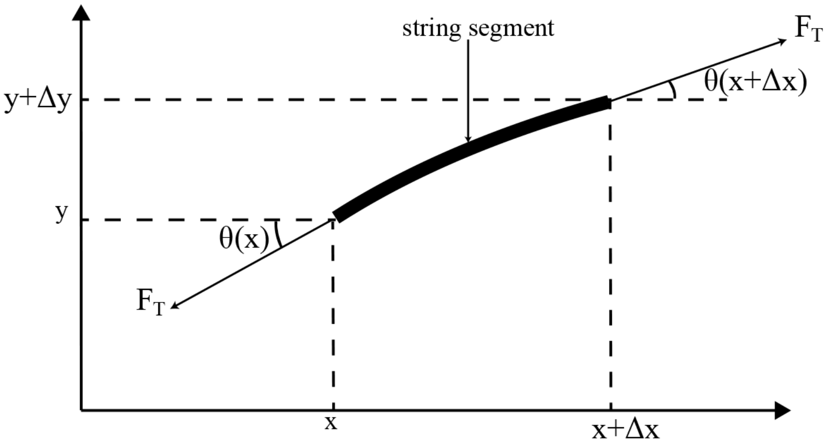

Figure 3.3 – Short segment of a string between (x, y) and $(x+\Delta x, y+\Delta y)$ where a tension F_T is applied.

If y corresponds to a small displacement, θ is also small and can be approximated using $\sin(\theta) \approx \tan(\theta)$ and $\tan(\theta) = \frac{\partial y}{\partial x}$. This allows re-writing Equation (3.5) as:

$$F = F_T(\frac{\partial y}{\partial x}(x+\Delta x) - \frac{\partial y}{\partial x}(x)) \tag{3.6}$$

Using the Newton's second law:

$$F = m\frac{\partial^2 y}{\partial t^2} \tag{3.7}$$

and knowing that the mass for this string segment is $m = \mu\Delta x$, we have:

$$F_T(\frac{\partial y}{\partial x}(x+\Delta x) - \frac{\partial y}{\partial x}(x)) = \mu\Delta x\frac{\partial^2 y}{\partial t^2} \tag{3.8}$$

dividing both sides of Equation (3.8) by Δx, applying the second derivative definition with $\Delta x \to 0$ and making $c = \sqrt{F_T/\mu}$, it becomes the wave equation:

$$\frac{\partial^2 y}{\partial t^2} = c^2\frac{\partial^2 y}{\partial x^2}, \qquad x \in (0, L), t \in (0, t_f] \tag{3.9}$$

This model was used to simulate plucked strings and the resulting waveforms were compared to measured waveforms, as discussed in Section 3.1.3.

3.1.3 Plucked string simulation

Equation 3.9 was numerically solved using the finite difference method (JAIN, 2003) and the algorithmic steps used by Langtangen (LANGTANGEN, 2016). The Taylor series expansion was used to approximate it as:

$$\begin{gathered}\frac{y(x+\partial x,t) - 2y(x,t) + y(x-\partial x,t)}{\partial x^2} = \\ \frac{1}{c^2}\frac{y(x,t+\partial t) - 2y(x,t) + y(x,t+\partial t)}{\partial t^2}\end{gathered} \tag{3.10}$$

Using the i, j notation such that $y(x,t) = y_{ij}$, inserting the wave number $C = \frac{c\partial t}{\partial x}$ and rearranging Equation 3.10 yields:

$$y_{i,j+1} = C^2 y_{i-1,j} + 2(1 - C^2) y_{i,j} + C^2 y_{i+1,j} - y_{i,j-1}. \tag{3.11}$$

To calculate the value of this function in the first time step, $y_{i,j-1}$ must be determined. This can be done using the initial velocity in Equation 3.4 and Tailor's series as follows:

$$\frac{y(x, t + \partial t) - y(x, t - \partial t)}{2\partial t} = 0. \tag{3.12}$$

Rearranging equation 3.12 and rewriting in the i, j notation, we find that:

$$y_{i,j-1} = y_{i,j+1}. \tag{3.13}$$

Finally, replacing $y_{i,j-1}$ by $y_{i,j+1}$ in Equation 3.11, isolating $y_{i,j-1}$ and dividing both sides by 2, we have:

$$y_{i,j+1} = \frac{C^2}{2} y_{i-1,j} + (1 - C^2) y_{i,j} + \frac{C^2}{2} y_{i+1,j}, \tag{3.14}$$

which is the finite difference scheme. The numerical simulation was executed over the discrete spatial domain [0,L] equally spaced by ∂x and over the discrete temporal domain $[0, T]$ equally spaced by ∂t.

The model's pluck position $x_p = L/5$ and the string length $L = 0.87m$ were directly measured from the strings of an electric bass. The wave velocity c was calculated using $c = f/(2L)$ (IAZZETTA, accessed 04/25/2019) related to note E0. The simulation time was define as $t_f = 0.05s$.

Over the spatial domain, the algorithm computes $y_{i,0}$ using Equation 3.3 and $y_{i,1}$ using Equation 3.14 and applying the boundary conditions from Equations 3.1 and 3.2. Then, for each element j from temporal domain, apply Equation 3.11 to find $y_{i,j+1}$ for each element i from the spatial domain, applying the boundary conditions again.

The output simulated signal was retrieved from the string velocity in the position $x = L/5$, approximately the pickup position, and was yielded to a 5th order low-pass Butterworth filter with a $150Hz$ cutoff frequency. This simulates the smoother bend of the string due to its stiffness and the soft touch from the fingertip, which are responsible for generating tones with weaker high-frequency components (JANSSON, 2002). The resulting signals were compared to the recorded signals, as shown in Figure 3.4.

Figure 3.4 shows that the physical model generates shapes that are similar to those found in the acquired signals. This means that the peak behavior is not a particular behavior of the specific electric basses that were used in our acquisitions. Rather, this behavior can be expected to appear in electric basses in general, hence it can be used for further steps in fundamental frequency estimation.

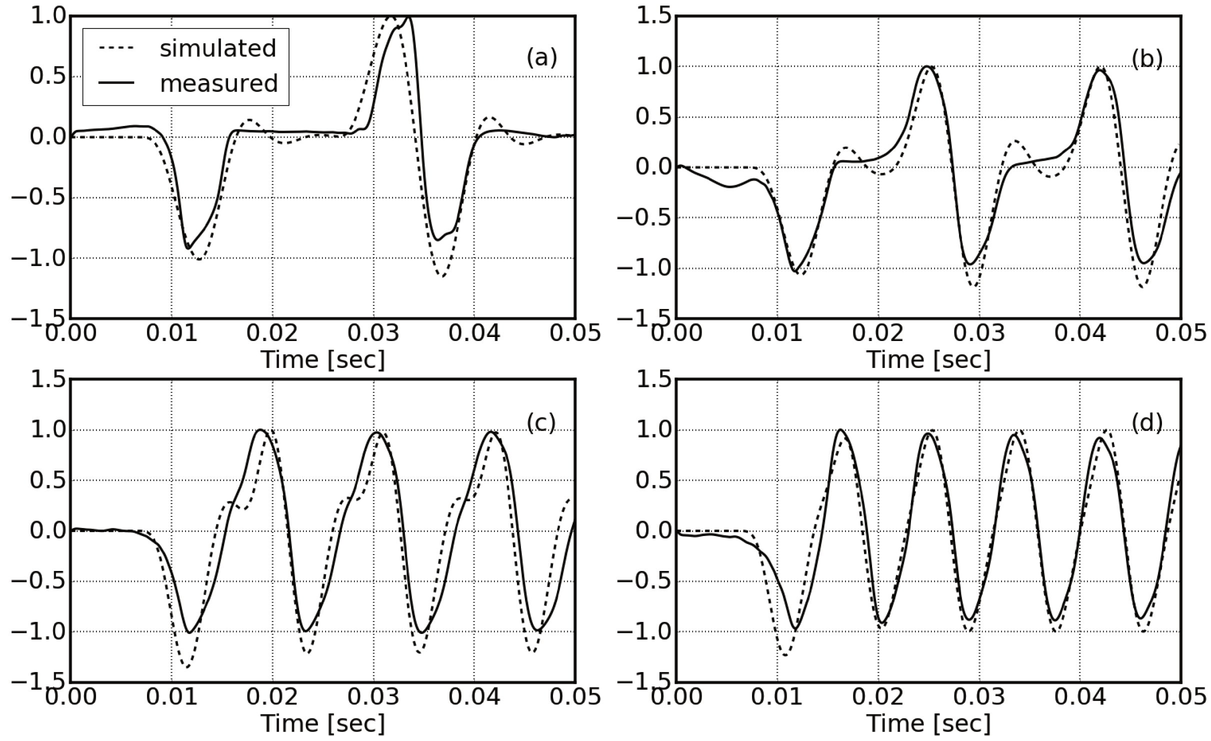

Figure 3.4 – Simulated and measured notes played on string E of the electric bass guitar (a) E0 (b) A♯0 (c) F1 (d) A1

3.2 Fundamental Frequency Estimation

The simulated and measured waveforms in Figure 3.4 show that there is a peak at the onset of the note and at the beginning of each cycle after it. These peaks have approximately the same width, regardless of the note's frequency, and the note's fundamental frequency occurs due to the rate in which peaks appear in the signal. The proposed method is based on these two characteristics, as follows:

As it is a proposal for analysis in real-time, the signal coming from the electric bass guitar must be analyzed continuously, that is, the analog electrical signal must be converted to digital and the samples saved in a buffer for analysis. For each new sample obtained, the buffer must be updated, eliminating the "oldest" sample from it. A flowchart of the entire algorithm process is presented in Figure 3.5.

3.2.1 Detect onset

Initially, it is necessary to detect the onset of the note that will be played on the instrument. There are several methods for detecting onsets that can be applied in this case, according to the study by (PORCIDES; TAVARES, 2014). As in the case of the electric bass guitar, there is usually a rapid and considerable increase in relative energy

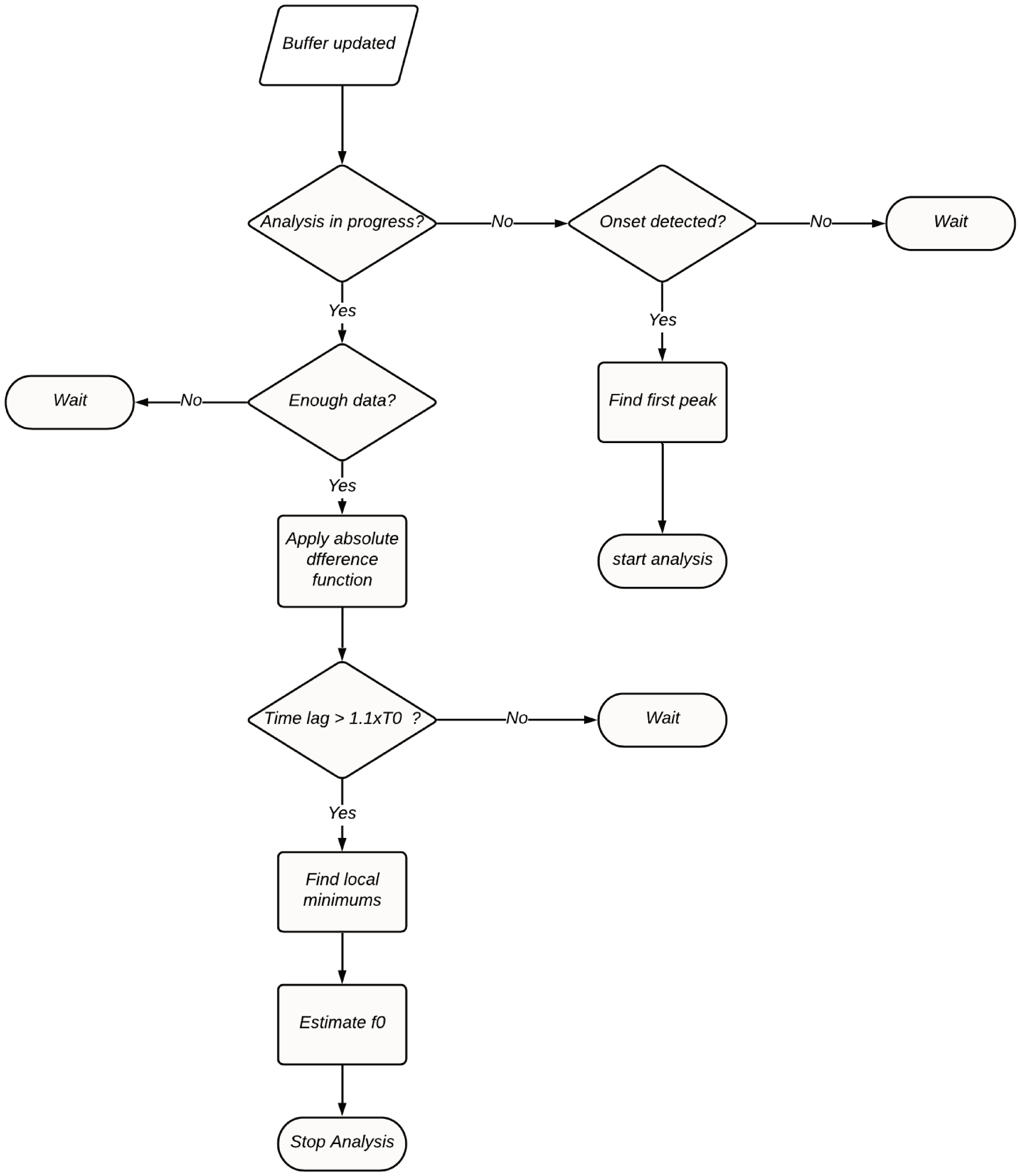

Figure 3.5 – Proposed method flowchart process.

when a note is played, it is proposed to use the method based on the energy function:

$$En_n = \frac{1}{N_{en}} \sum_{n=1}^{N_{en}} |s_n|^2, \tag{3.15}$$

where N_{en} is the length of the analysis window. A note onset is detected when the energy variation is positive and bigger than a threshold value.

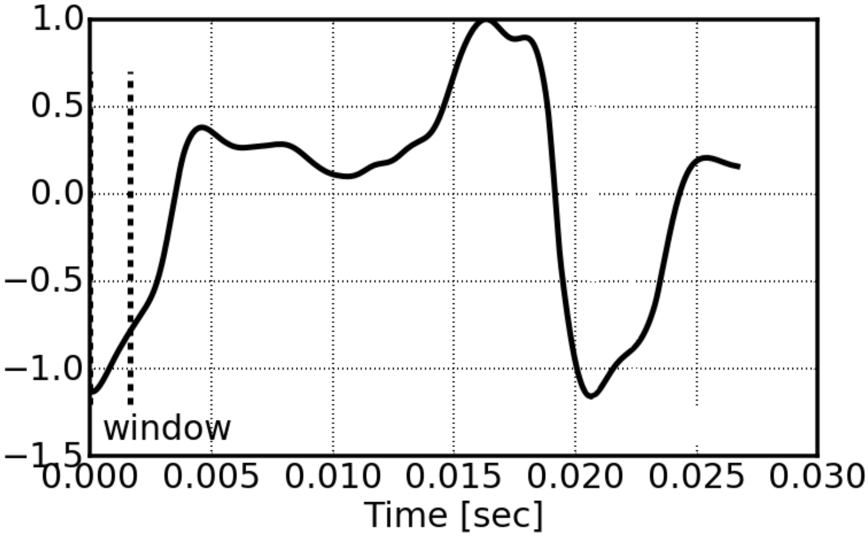

Figure 3.6 – Analyzed signal s_n and integration short window with size W. This signal is a recording of the G0 note played on the E string of an electric bass guitar

3.2.2 Determine starting peak

When onset is detected, the algorithm will seek to determine the initial peak in the buffer, as expected according to figure 3.6. The instant of occurrence of this peak is used to define the start time of the short W-size integration window, also shown in Figure 3.6, which will be used in the following steps in the application of the absolute difference function.

3.2.3 Detect if there is enough data

The W size of the short integration window is one of the input parameters of the algorithm and must be less than half the width of the initial peak. Bearing in mind that for the same string, the width of this peak remains approximately constant, regardless of the note.

To perform the next step, it is necessary to check if the number of samples available in the buffer generated after the initial peak is greater than W, as an onset can be detected so quickly that the analyzed signal has not yet toured enough for the generation of the samples necessary for the application of the absolute difference function.

3.2.4 Absolute difference function

The next step is to apply the absolute difference function to the W length section of the signal available in the buffer. In the initial instant, this signal will be exactly the same as the short integration window itself, that is, the result will be zero. For each new sample that becomes available in the buffer, the absolute difference function

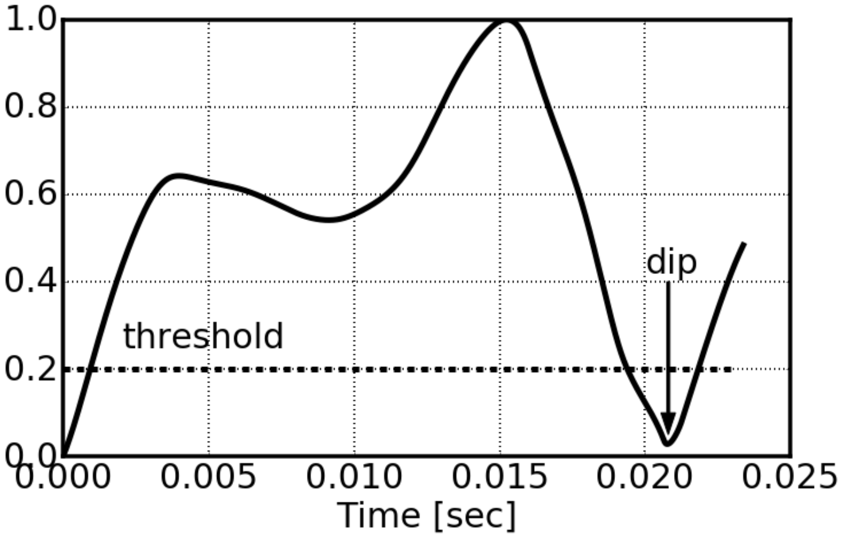

Figure 3.7 – Absolute difference function from the analyzed signal s_n from figure 3.6 and threshold value represented as the horizontal dotted line

is applied again, keeping the same short integration window samples, but comparing it to a new signal segment, which contains the new sample made available and does not contain the sample from the "oldest" instant.

The absolute difference function is defined as:

$$d(\tau) = \sum_{n=1}^{W} |s_n - s_{n+\tau}|, \tag{3.16}$$

where τ is the temporal lag between the initial peak and the analized section from the audio signal s_n. So we are measuring the absolute difference between the first moments of the signal after the initial peak in relation to the following sections of this same signal, resulting in a function like the one illustrated in Figure 3.7.

The absolute difference function must be applied to each buffer update until it has passed, from the initial peak, an interval of 1.1 times the fundamental period T_0 of the lowest frequency to be detected. Theoretically, this interval could be $T_0 + W$, but the first cycle from the onset is subject to harmonics that can vary the interval between the first two peaks of the signal. Thus, $1.1 \times T_0$ gives a margin of tolerance.

3.2.5 Find local minima

In sequence, the algorithm searches for local minima in the absolute difference function, referenced as dips in Figure 3.7. For the lowest notes, there will be only a local minimum as depicted in Figure 3.7, from which we will obtain the τ_0 interval. For the highest notes, as exemplified in Figure 3.8, there may be 2 or 3 local minimums, as

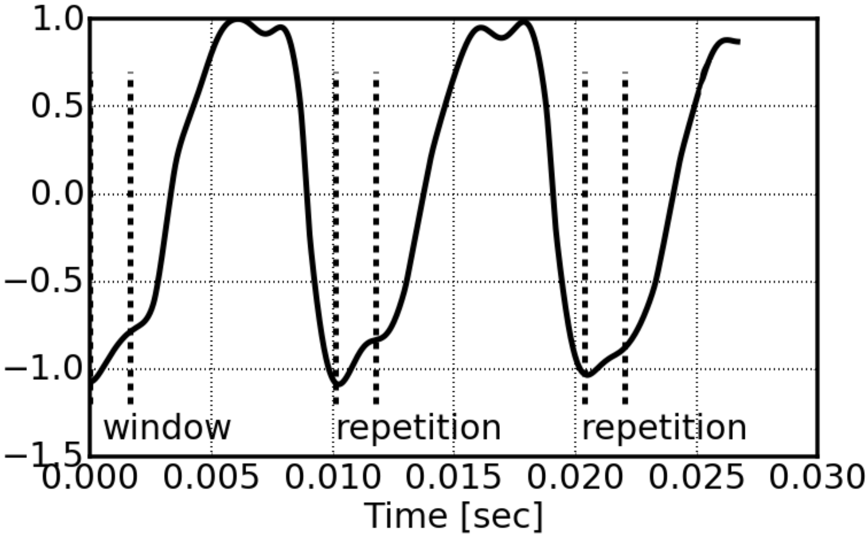

Figure 3.8 – Analyzed signal s_n and integration short window with size W. This signal is a recording of the G1 note played on the E string of an electric bass guitar

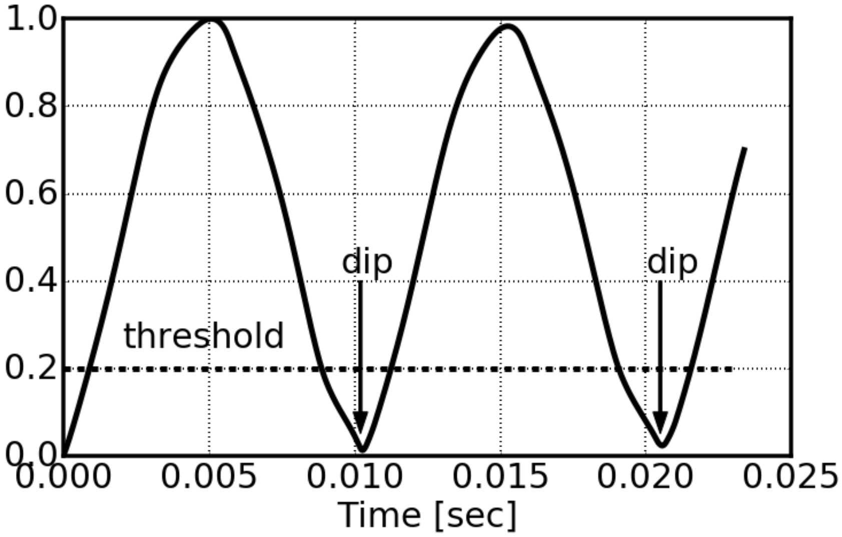

Figure 3.9 – Absolute difference function from the analyzed signal s_n from figure 3.8 and threshold value represented as the horizontal dotted line

depicture in Figure 3.9, depending on how many frets the bass has. In this case, τ_0 is obtained by:

$$\tau_0 = \sum_{n=1}^{N_\tau} \frac{\tau_n}{n}, \tag{3.17}$$

where N_τ is the number of local minimums and τ_n is the temporal lag between the initial peak and nTH local minimums.

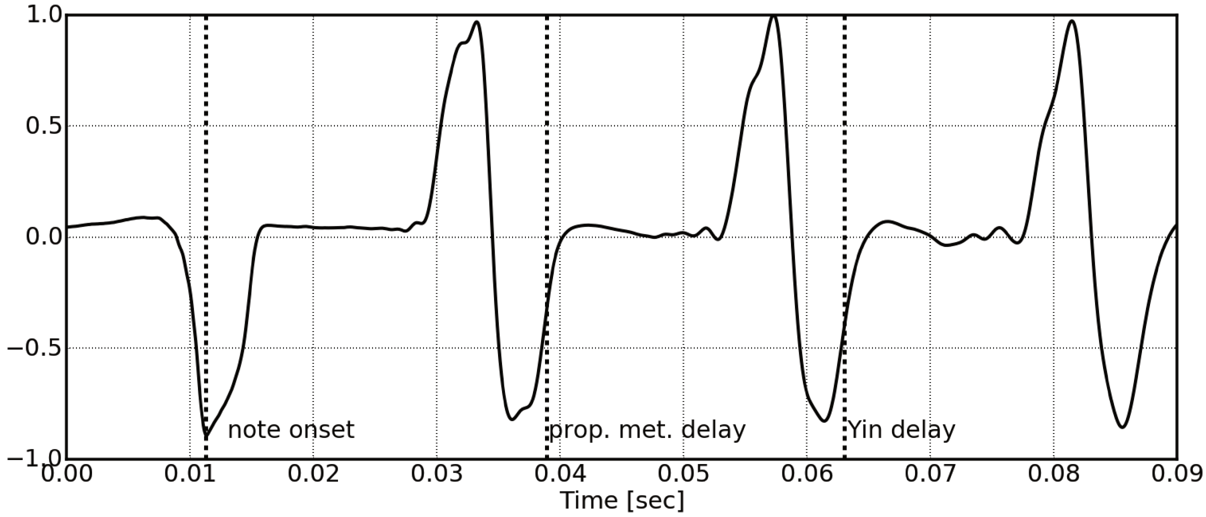

Figure 3.10 – Algorithmic delay for the proposed method and for the Yin method.

3.2.6 Determine f0

Since τ_0 represents the interval in which the signal most seems to repeat its initial stretch, we define that the fundamental period of the T_0 signal is equal to τ_0. Thus we determine the fundamental frequency f_0 of the signal as:

$$f_0 = \frac{1}{T_0} \tag{3.18}$$

Therefore, briefly, the proposed method consists of the application of the signal to an absolute difference with a window size W shorter than half-width of this first peak as shown in Figure 3.6. This thin window plays an important role to make it possible for the method to find f0 after 1.1 times the fundamental period, whereas the Yin method needs more than two fundamental periods (CHEVEIGNé; KAWAHARA, 2002), as shown in Figure 3.10.

The next section discusses experiments regarding the proposed method.

4 Experiments and results

4.1 Dataset

The proposed method was tested using a set of audio recordings acquired from 3 different electric bass guitars. Each of them was played by a different musician, and all of them used the finger-plucking technique. All notes within the instrument's range were recorded from each of the guitars, using two different instrument equalizations (full bass and full treble). This yielded 528 recordings, which were all manually cropped to start at the note onset since the proposed method does not have a note onset detector.

4.2 Experiments

This section describes experiments that compare the proposed method to the Yin method (CHEVEIGNé; KAWAHARA, 2002), as implemented by Guyot (GUYOT, 2018). The experiments comprised executing both the proposed method and the Yin method to estimate the f0 in the dataset samples.

4.2.1 Test 1 - sample length for note

In this first test, the sample length provided as input parameters for the algorithms is equal to $1.1 \times T_{t1} \times fs$, being T_{t1} the fundamental period of the expected note and fs the sampling frequency of the digital audio signal. To serve as a reference, the test was repeated for the Yin method with a sample length equal to $2.1 \times T_{t1} \times fs$ and is referenced as "Yin2" in Figure 4.2 (a).

4.2.2 Test 2 - sample length for string

This second test is a more common application for a pitch detector in a string instrument, where the fundamental frequency should be estimated from a range of approximately 2 octaves. So, the sample length provided as input parameters for the algorithms is equal to $1.1 \times T_{t2} \times fs$, being T_{t2} the fundamental period of the lower note from the specific string to which the recorded note belongs. Also, in this case, the test was repeated for the Yin method with a sample length equal to $2.1 \times T_{t2} \times fs$ and is referenced as "Yin2" in Figure 4.2 (b).

Figure 4.1 compares the length of the samples used in test 1, shown in the first column, and test 2, shown in the second column of the figure.

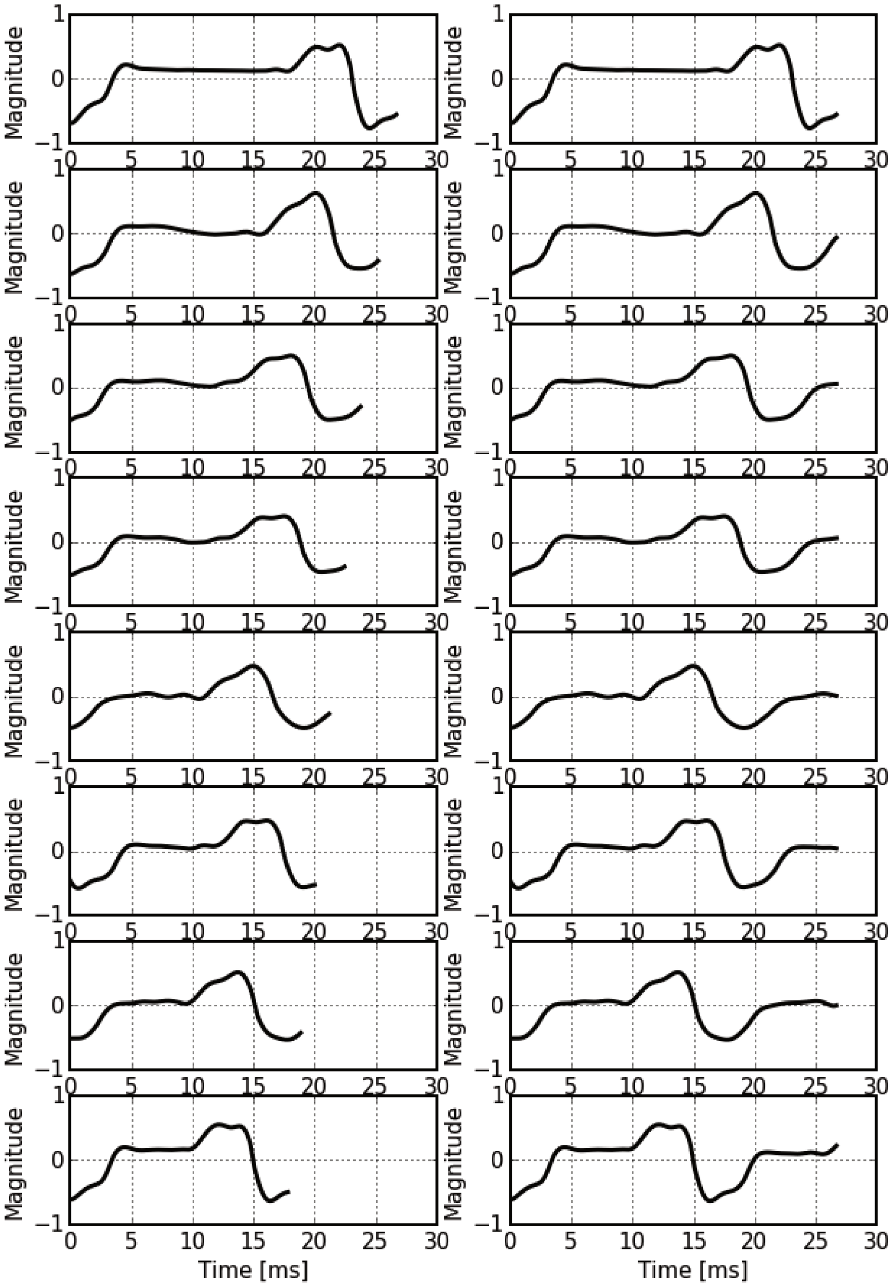

Figure 4.1 – sample lengths for test 1 in the first column and for test 2 in the second column

To determined if the method fails, the MIDI note correspondent to the fundamental frequency estimated is calculated as:

$$M_{note} = 12 \log \left(\frac{f_0}{16.351597} \right) \frac{1}{\log(2)}, \tag{4.1}$$

where f_0 is the estimated fundamental frequency and 16.351597 is the f_0 for the MIDI note = 0. The result is rounded to the nearest integer. If the calculated MIDI note differs from the expected one, it is counted as one error.

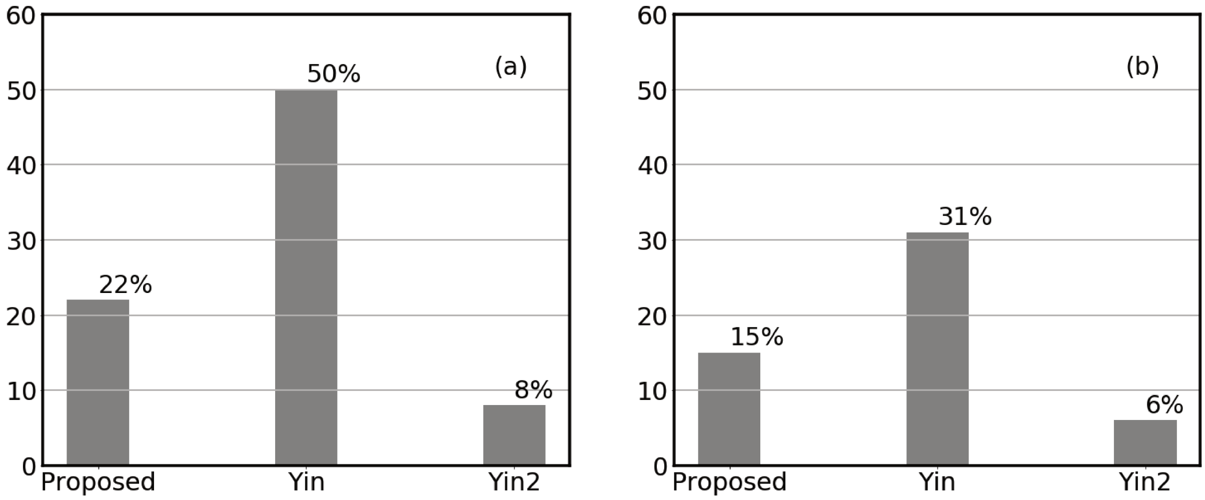

Figure 4.2 – (a) Test 1 error rates . (b) Test 2 error rates

4.3 Proposed method applied to other musical instruments

The proposed method was developed based on specific characteristics of the electric bass waveform when played using the finger plucking technique. These characteristics were observed in samples of recordings made with the referred instrument and mathematically modeled to guarantee that they will be present in the waveforms generated by electric basses in general. As these characteristics may be also present in waveforms generated by other instruments, this section presents the results of applying the method to audio samples of some other instruments in order to indicate promising paths for future work in the expansion of the method application.

The samples of musical instruments analyzed below were obtained from the soundbank of the FreePats project (FREEPATS, accessed 09/02/2020).

4.3.1 Electric Guitar

The waveforms illustrated in Figure 4.3 (a) and (c) were obtained from recorded samples from the Fender Telecaster Electric Guitar, direct from its bridge pickup output.

From the analyzed waveforms we can find the main characteristics for the application of the proposed method. The signal generated by the electric guitar shows sharp peaks at the beginning of each cycle and that varies little in relation to the note played.

The proposed method was able to detect the fundamental frequency of the analyzed signals, presenting local minimums at the waveform repetition points, as can be seen in Figure 4.3 (b) and (d).

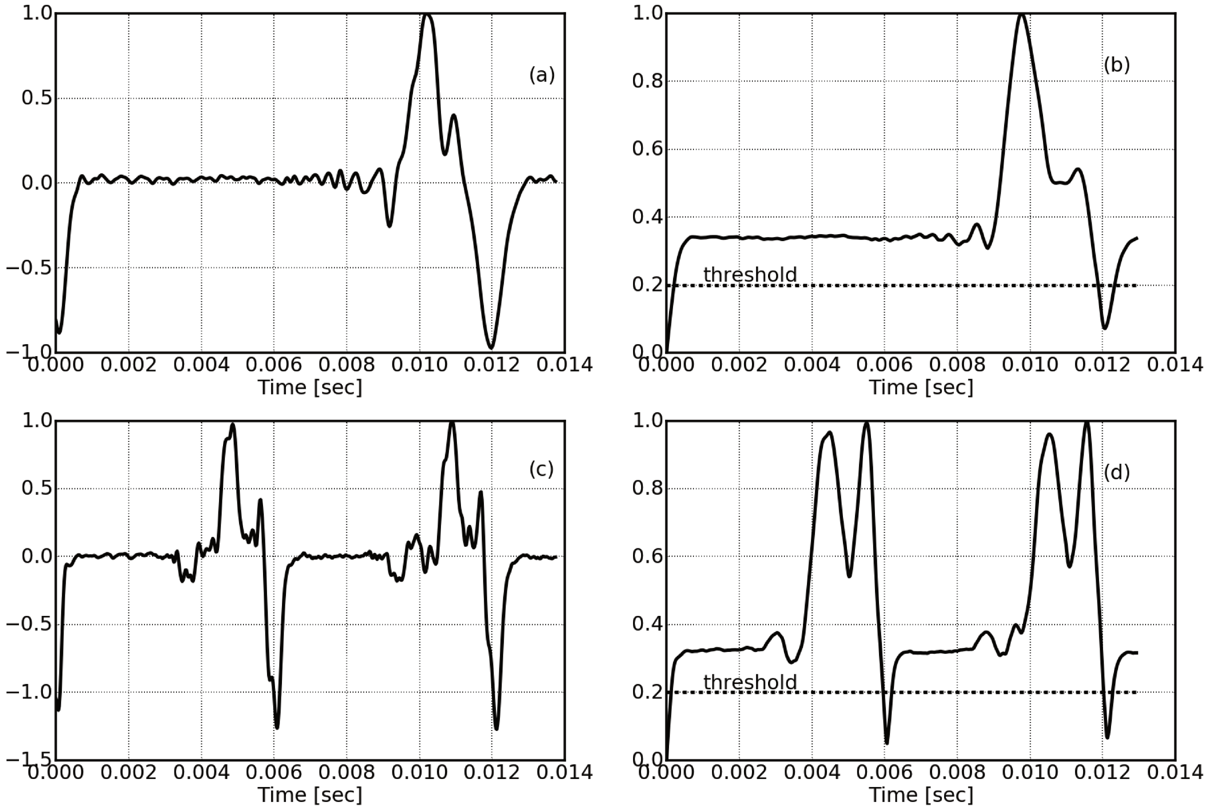

Figure 4.3 – Analyzed signals from an electric guitar: (a) note $E1$; (c) note $E2$. Absolute difference function: (b) from signal in Figure(a); (d) from signal in Figure (c).

4.3.2 Acoustic Guitar

The waveforms illustrated in Figure 4.4 (a) and (c) were obtained from samples recorded from a Spanish classical guitar through a microphone.

The analyzed waveforms do not have the necessary characteristics for the application of the proposed method. Consequently, the application of the absolute difference function does not have a function that allows us to determine the fundamental frequency, as shown in the figure 4.4 (b) and (d).

Observing the waveform of the figure 4.5 it is possible to notice that in the first moments after the note onset there is a strong presence of harmonics, probably due to the contact between the musician's nail and the instrument's nylon string, which hinder the use of the proposed method. Another factor responsible for the big difference in the waveform is the capture method, which was made by a microphone, adding to the signal the effects of room reverberation.

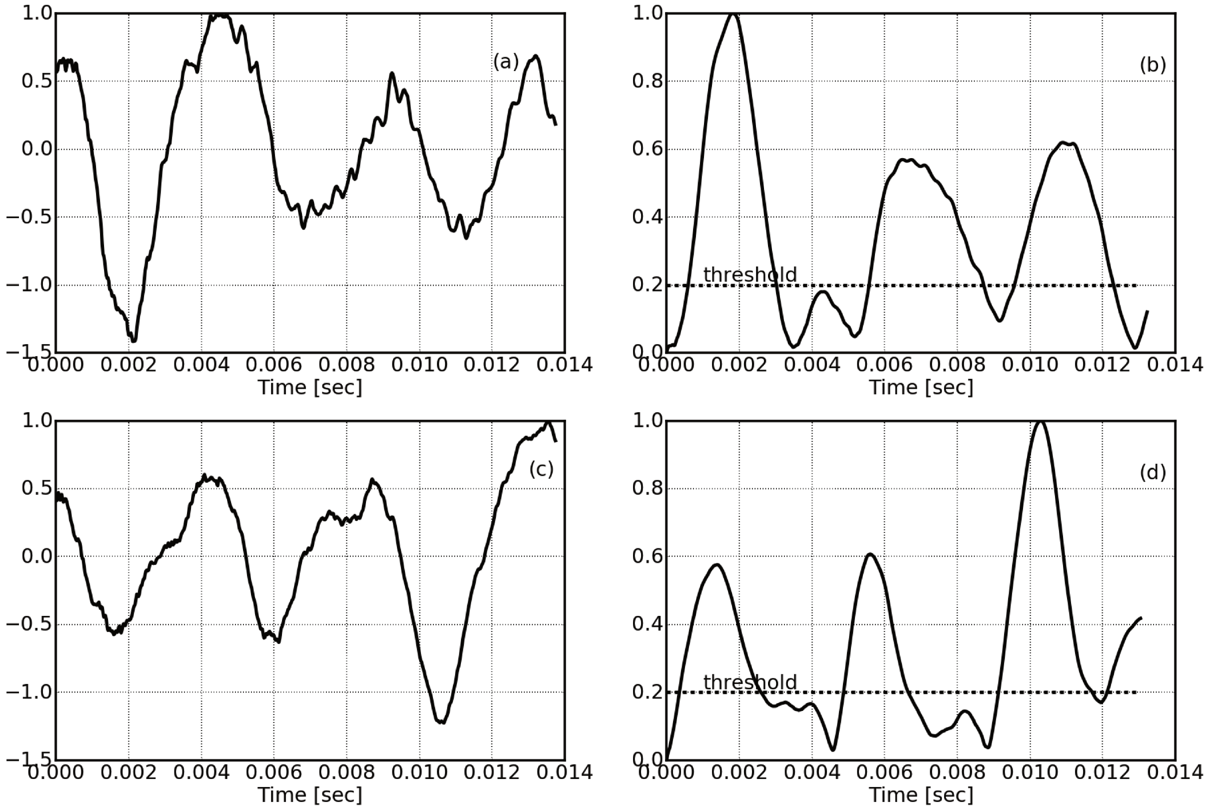

Figure 4.4 – Analyzed signals from an acoustic guitar: (a) note $E1$; (c) note $E2$. Absolute difference function: (b) from signal in Figure(a); (d) from signal in Figure (c).

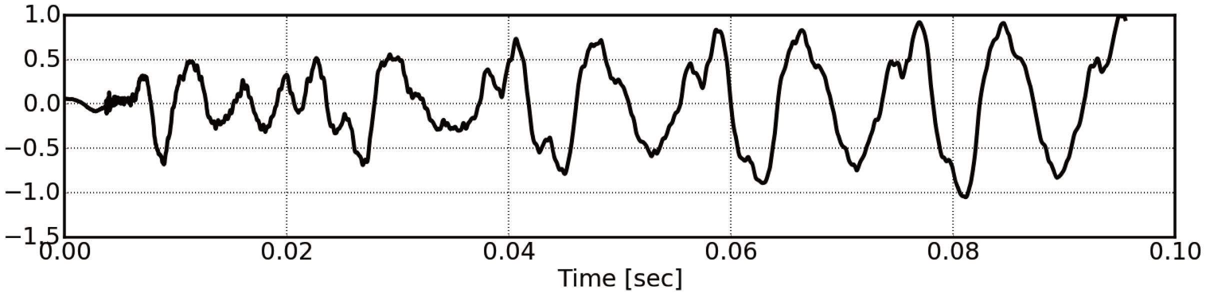

Figure 4.5 – waveform from an acoustic guitar attack and first cycles.

4.3.3 Upright Piano

The waveforms illustrated in Figure 4.6 (a) and (c) were obtained from samples recorded from a Kawai upright piano, located in a living room through a microphone positioned in front of the piano, approximately at the place where the head of a piano player would be.

Again, the analyzed waveforms do not have the necessary characteristics for the application of the proposed method. Consequently, the application of the absolute difference function does not have a function that allows us to determine the fundamental frequency, as shown in the figure 4.4 (b) and (d).

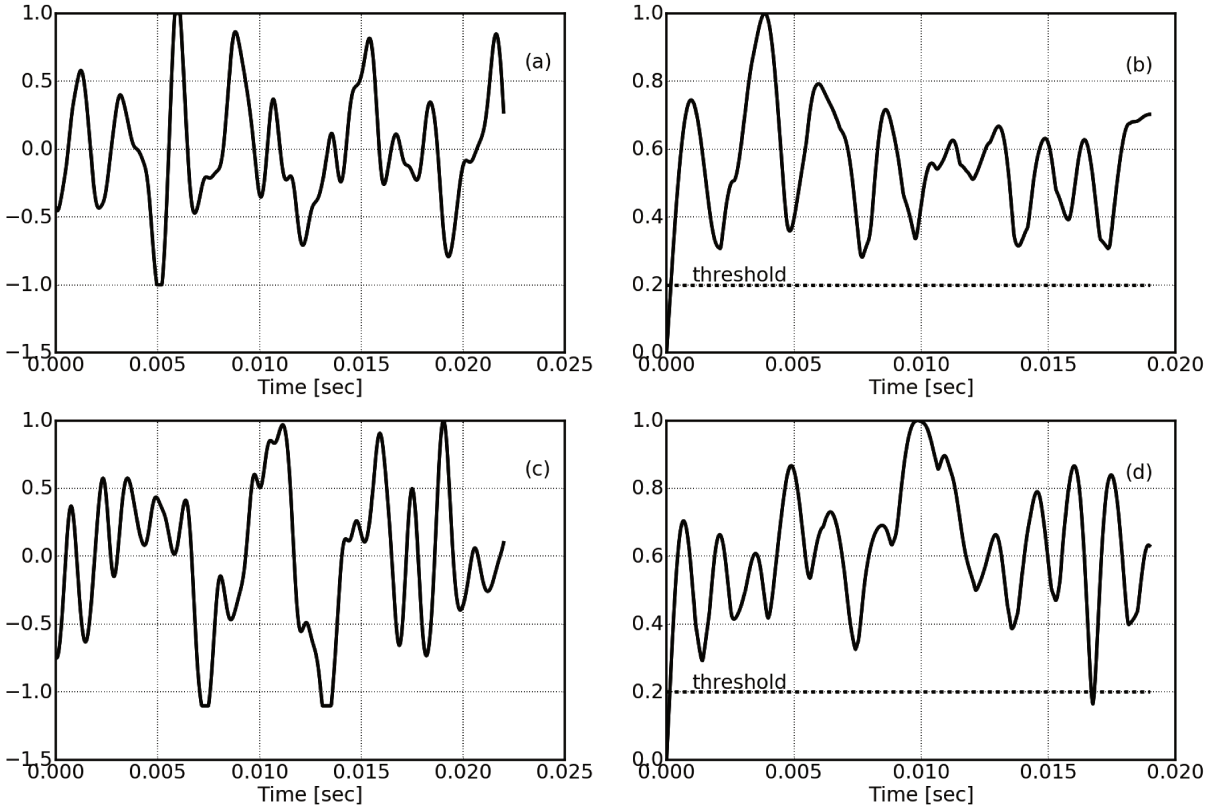

Figure 4.6 – Analyzed signals from an Upright Piano: (a) note $A0$; (c) note $A1$. Absolute difference function: (b) from signal in Figure(a); (d) from signal in Figure (c).

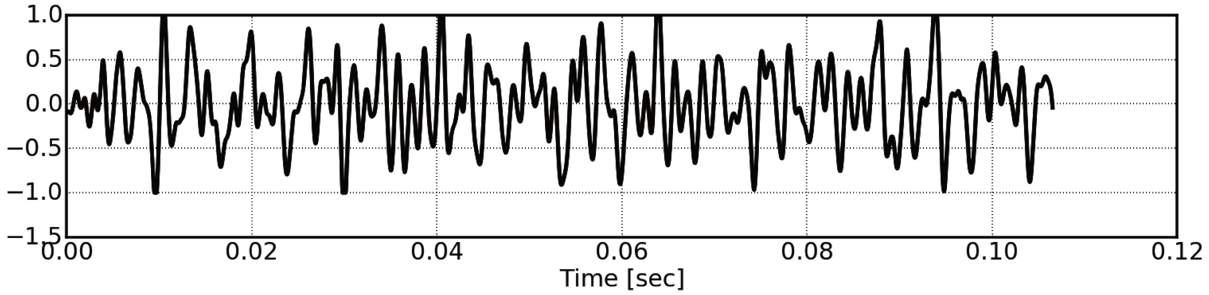

Figure 4.7 – waveform from an Upright Piano attack and first cycles.

From the waveform shown in figure 4.7 it is possible to notice that there is a strong presence of harmonics along with the signal, probably due to the impact of the hammer on the string and the construction of the instrument that differs greatly from the electric bass. Therefore, the use of the proposed method for this type of instrument also seems unfeasible.

4.3.4 Wooden Recorder

The waveforms illustrated in Figure 4.8 (a) and (c) were obtained from samples recorded from a "Venus" wooden recorder through a microphone.

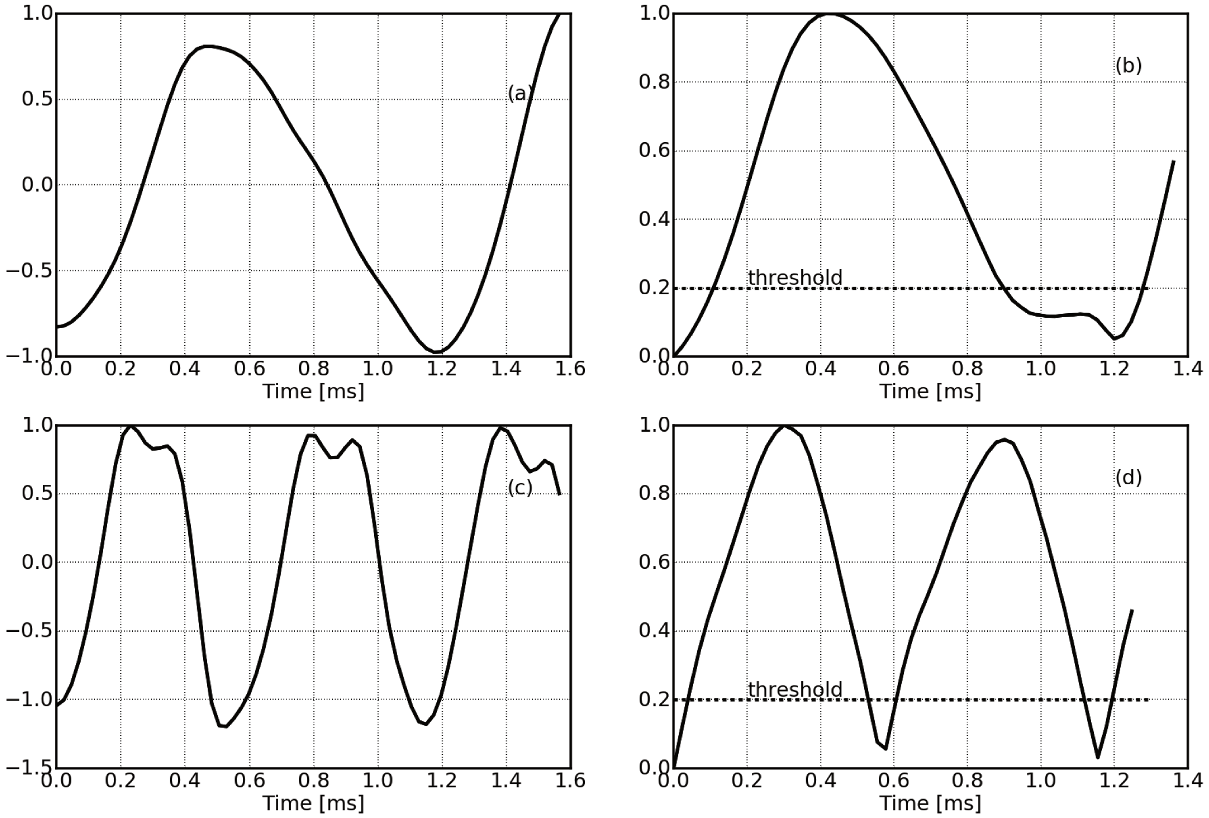

Figure 4.8 – Analyzed signals from a wooden recorder: (a) note $A5$; (c) note $A6$. Absolute difference function: (b) from signal in Figure(a); (d) from signal in Figure (c).

This is yet another case where the analyzed waveforms do not have the characteristics that supported the development of the proposed method. However, as the signal generated by this instrument has few harmonics, approaching a sinusoid, the application of the absolute difference function generated a signal with local minimums at the beginning of each cycle, as shown in the figure 4.8 (b) and (d), making it possible to determine the fundamental frequency.

As may be observed in figure 4.9, this instrument has a relatively slow attack so that the initial peak detected in the onset has a lower amplitude than the following peaks. Depending on how big this difference is, the point of occurrence of the local minima of the absolute difference function can be changed enough to cause the error of the note determined by the method.

4.4 Discussion

The error rates presented in Figure 4.2 show that the proposed method had less than half of the Yin method's error rate, so having a better performance estimating f0 on both tests.

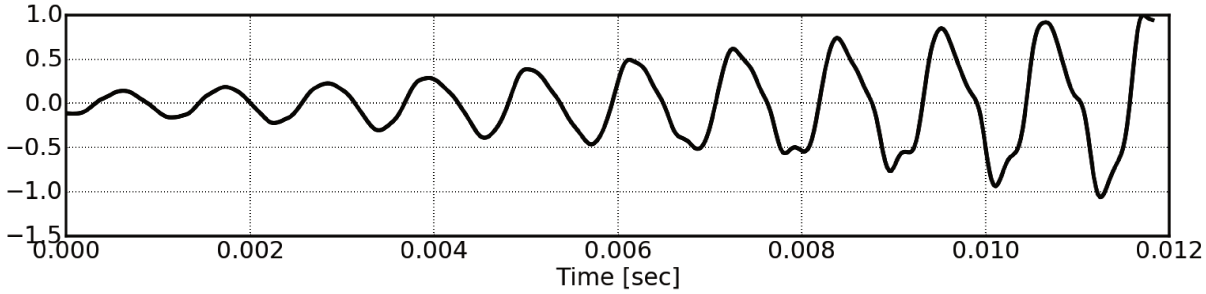

Figure 4.9 – waveform from an wooden recorder attack and first cycles.

It is important to note that the tests refer to a very specific condition, as they aim to verify the performance of the methods to determine the fundamental frequency of notes played on a specific musical instrument, the electric bass, right after its first oscillation cycle. In addition, the method was tested for the frequency range from 41.2 Hz to 392 Hz, that is, from the lowest to the highest note of the standard four-string electric bass.

As expected, the Yin method is a better solution when sample length is longer than 2 cycles of the fundamental period, but for the string E of an electric bass guitar, the algorithmic delay is higher than 50 ms ($2/f0 = 2, 1/41.20Hz \approx 0, 051s$), which is perceptible for a bass player, making it harder to play the bass guitar with real-time MIDI outputs, as shown in the (LESTER; BOLEY, 2007) study, where professional bassists deemed acceptable latencies of up to 30 ms.

The study on the application of the proposed method to other musical instruments indicated that there is a possibility of obtaining good results with the electric guitar. This is due to the fact that the instruments share many constructive characteristics, such as metallic strings and capture by electromagnetic pickups. For the acoustic guitar and upright piano, the results were not promising. The waveforms generated by these instruments are quite different from those generated by the electric bass, mainly because the sound generated is not a simple capture of the vibrating string, but rather the vibration of its entire structure. Finally, the method even proved to be reasonably applicable to the Wooden Recorder, but as this instrument reproduces high notes, more accurate methods that use more than two cycles for the detection of the pitch will not present great latencies.

5 Conclusion

A method based on the absolute difference function and the waveforms from a finger plucked strings of an electric bass guitar was presented. It was tested over 528 notes recorded from three different bass guitars and it shows to be capable to estimate these notes from samples with length equal to 1.1 times their fundamental periods, while our reference method, Yin, under the same conditions, had double the error rate. This shorter algorithmic delay, near the minimal theoretical delay (one fundamental period) and low computational complexity, makes the proposed method suitable for real-time applications for the electric bass guitar, such as a MIDI bass guitar.

However the method missed 15% of the notes on test 2, which is a similar application, so future studies should be made to improve these results. An approach to reduce errors, unrelated to improvements in the method, would be to adopt a specific way of playing the musical instrument. If the bass player always plucks the string smoothly, in order to keep the first cycles of the signal similar to the modeled ones, error rates can be drastically improved. It can be a useful alternative way to a MIDI bass guitar, where the way you pluck the strings will not affect the sound timber. But, clearly, this imposes a limited way to play in exchange for a more precise note detection and lower latency

Also, the method was not tested for notes played on top of an already vibrating string which certainly should make it harder to estimate the correct f0. However, it is possible that contact with the plucking finger, at the moment of playing the new note, dampens the string enough to not interfere with the performance of the method. This case will be approached in future work.

The method is applicable for pitch determination for monophonic electric bass signals, so in a real application, it would be necessary to use individual pickups per string, so that each generated signal can be analyzed individually and ensuring that there will be no more than one note simultaneously for each signal. In addition, the method requires a quick onsets detector, which provides the information that a note has been played to begin the analysis process.

A promising path for future work would be the development of a hybrid method, which uses the proposed method for rapid pitch detection in low notes and another more accurate method using at least two cycles, such as Yin, for higher notes. Thus, adjusting the proposed method to provide an estimate after an analysis window of 1.1 times the period of the lowest fundamental frequency, and the second method to provide an estimate as soon as it is obtained, that is, after two cycles of the analyzed frequency, we will have the following process: if the note is high, the second method will offer the estimate before the end of the analysis of the proposed method, otherwise the proposed method will provide its estimate, avoiding greater latencies.

Finally, future works can study how the use of the reed to play the strings affects the error rates, which could allow the application of the method for the electric guitar, an instrument that indicated to have similar characteristics in the waveforms, from those used in the analysis by the proposed method.

Bibliography

BADEAU, R. B. R.; DAVIDWANG, Y. B. Eds parametric modeling and tracking of audio signals. *Proc. DAFx'02, Hamburg, Germany*, 2002. Citado na página 31.

CARIANI, P. A.; DELGUTTE, B. Neural correlates of the pitch of complex tones. i. pitch and pitch salience. *J. Neurophysiol. 76*, 1996. Citado na página 16.

CHEVEIGNé, A. de; KAWAHARA, H. Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, v. 111, n. 4, p. 1917–1930, 2002. Citado 6 vezes nas páginas 15, 16, 32, 35, 51, and 52.

DERRIEN, O. A very low latency pitch tracker for audio to midi conversion. *17th International Conference on Digital Audio Effects (DAFx-14)*, 2014. Citado 3 vezes nas páginas 15, 30, and 35.

FORNARI, J. Percepção, cognição e afeto musical. In: _____. [S.l.: s.n.], 2010. ISBN 978-85-63046-01-7. Citado 2 vezes nas páginas 16 and 17.

FREEPATS. *Sound Banks.* [S.l.]: http://freepats.zenvoid.org/index.html, accessed 09/02/2020. Citado na página 54.

GERHARD, D. Pitch extraction and fundamental frequency: History and current techniques. *Technical Report TR-CS 2003-06*, 2003. Citado 4 vezes nas páginas 17, 18, 21, and 35.

GIANNAKOPOULOS T., P. A. *Introduction to Audio Analysis.* [S.l.]: Elsevier, 2014. ISBN 978-0-08-099388-1. Citado na página 21.

GIBSON J., W. A. *The MIDI Standard.* [S.l.]: http://www.indiana.edu/ emusic/361/midi.htm, accessed 05/9/2019. Citado na página 15.

GREEF, W. The influence of perception latency on the quality of musical performance during a simulated delay scenario. *University of Pretoria, Department of Music*, 2016. Citado na página 36.

GUYOT, P. Fast python implementation of the yin algorithm. *http://doi.org/10.5281/zenodo.1220947*, 2018. "accessed 01/02/2018". Citado na página 52.

HELLER, E. J. *Why You Hear What You Hear.* [S.l.]: Princeton University Press, 2012. (Chapter 23; pp. 437-504). Citado na página 16.

IAZZETTA, F. *Tutoriais de Audio e Acustica.* [S.l.]: http://www2.eca.usp.br/prof/iazzetta/tutor/acustica, accessed 04/25/2019. Citado 2 vezes nas páginas 17 and 45.

ISO16:1975-ACOUSTICS. *Standard tuning frequency.* [S.l.]: International Organization for Standardization, 1975. Citado na página 16.

JAIN, M. *Numerical Methods for Scientific and Engineering Computation.* 1st ed.. ed. [S.l.]: New Age International, 2003. ISBN-10: 8122414613. pp: 844. Citado na página 44.

JANSSON, E. *Acoustics for Violin and Guitar Makers*. 4th ed.. ed. [S.l.]: http://www.speech.kth.se/music/acviguit4/, 2002. (Chapter 4; pp. 16-18). Citado na página 45.

KNESEBECK, A.; ZOLZER, U. Comparison of pitch trackers for real-time guitar effects. *13th Int. Conference on Digital Audio Effects (DAFx-10)*, 2010. Citado 2 vezes nas páginas 25 and 35.

LANGTANGEN, H. *Finite difference methods for wave motion*. preliminary version. [S.l.]: Department of Informatics, University of Oslo, 2016. Citado na página 44.

LESTER, M.; BOLEY, J. The effects of latency on live sound monitoring. *Journal of the Audio Engineering Society*, 2007. Citado 4 vezes nas páginas 11, 36, 37, and 59.

MIMO. *Revision of the Hornbostel-Sachs Classification of Musical Instruments by the MIMO Consortium*. [S.l.]: http://www.mimo-international.com/documents/Hornbostel/20Sachs.pdf, 2011. Citado na página 37.

MITROVIć, D. *Advances in Computers*. [S.l.]: Elsevier, 2010. ISSN 0065-2458. Citado na página 21.

NOLL, A. M. Pitch determination of human speech by the harmonic product spectrum, the harmonic surn spectrum, and a maximum likelihood estimate. *Symposium on Computer Processing in Communication, ed.*, University of Broodlyn Press, New York, v. 19, p. 779–797, 1970. Disponível em: <https://ci.nii.ac.jp/naid/10000045637/en/>. Citado 2 vezes nas páginas 22 and 28.

OPPENHEIM, A.; SCHAFER, R. *Discrete-Time Signal Processing*. [S.l.]: Prentice Hall, 1999. ISBN-10: 0137549202. Citado na página 25.

OXENHAM, A. J. Pitch perception. *Journal of Neuroscience*, v. 32, n. 39, p. 13335–13338, 26 September 2012. Citado na página 16.

P., K. A. Multiple fundamental frequency estimation based on harmonicity and spectral smoothness. *IEEE Trans. Speech and Audio Proc*, 2003. Citado na página 15.

PORCIDES, C.; TAVARES, L. Resultados preliminares de um estudo comparativo de métodos de detecção de onsets em sinais de Áudio. *Anais do Simpósio de Processamento de Sinais da UNICAMP, Vol. 1*, 2014. Citado na página 46.

PORTER, J. *Parts of the Bass*. [S.l.]: https://www.liveabout.com/bass-guitar-parts-1711849, accessed 12/10/2019. Citado na página 38.

Rabiner, L. On the use of autocorrelation analysis for pitch detection. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 25, n. 1, p. 24–33, February 1977. ISSN 0096-3518. Citado 2 vezes nas páginas 15 and 22.

RAMA, C.; THEODORIDIS, S. *Academic Press Library in Signal Processing*. [S.l.]: Elsevier, 2014. ISSN 2351-9819. Citado na página 19.

RESEARCHGATE. *YIN, A fundamental frequency estimator for speech and music*. [S.l.]: https://www.researchgate.net/publication/11367890$_{Y}IN_{Af}undamental_{f}requency_{e}stimator_{f}or_{s}peec$

Singh, C. P.; Kumar, T. K. Efficient pitch detection algorithms for pitched musical instrument sounds: A comparative performance evaluation. In: *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. [S.l.: s.n.], 2014. p. 1876–1880. ISSN null. Citado na página 28.

STRINGVIBE. *Bass Guitar Parts 101: The Parts That Make the Music*. [S.l.]: https://www.stringvibe.com/bass-guitar-building/, accessed 12/10/2019. Citado na página 38.

STRINGVIBE. *The Only Bass Guitar Strings Guide You'll Ever Need*. [S.l.]: https://www.stringvibe.com/bass-guitar-strings-guide/, accessed 12/10/2019. Citado 2 vezes nas páginas 37 and 40.

TERHARDT, E. Pitch, consonance and harmony. *J. Acoust. Soc. Am. 55*, 1974. Citado na página 16.

VARIOUS. *The string bass instrument According to Wikipedia*. [S.l.]: https://basgitarr.files.wordpress.com/2012/09/the-string-bass-instrument.pdf, accessed 12/10/2019. Citado na página 38.

VETTER, J. *ELECTRIC BASS GUITAR*. [S.l.]: https://omeka1.grinnell.edu/MusicalInstruments/items/show/248, accessed 12/10/2019. Citado na página 38.

WANG, Y. Low latency audio processing. *Queen Mary University of London, School of Electronic Engineering and Computer Science*, 2017. Citado na página 36.