Revista de Informática Teórica e Aplicada - RITA - ISSN 2175-2745
Vol. 27, Num. 04 (2020) 79-94

RESEARCH ARTICLE

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

Estimador de f_0 de Baixa Latência para o Contrabaixo Elétrico Tocado com os Dedos Usando a Função de Diferença Absoluta

Christhian Fonseca[1,2]*, Tiago Tavares[1,2]

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: Fundamental frequency 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-Chave: Estimador de frequência fundamental — Conversor de áudio para MIDI de baixa latência — Recuperação de informações musicais — Baixo MIDI

[1] *School of Electric and Computer Engineering (FEEC), University of Campinas, Brazil*
[2] *Interdisciplinary Nucleus for Sound Studies (NICS), University of Campinas, Brazil*
***Corresponding author**: christian_h@hotmail.com
DOI: http://dx.doi.org/10.22456/2175-2745.103182 • **Received:** 19/05/2020 • **Accepted:** 05/09/2020

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 [1]. This toolchain can also include analog instruments by means of audio-to-MIDI converters [2].

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 [3]. Many well-known algorithms aim at estimating f_0, such as the autocorrelation [4] and the Yin method [5].

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

$$s_t = s_{t+kT_0}, \tag{1}$$

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.

1.1 Pitch theory

Pitch is a psychoacoustical atribute of the sound given by the auditory sensation often related to the perception of a repetition rate of a waveform [6] above 20 Hz, where it is perceived not as rhythm but as tone. The lowest regular repetition rate is called Fundamental Frequency (f_0) 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:

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

The relative harmonic amplitudes a_m, among other atributes, are commonly associated to timbre differences and the fundamental frequency f_0 is closely related to the sensation of pitch [7]. 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 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 [5].

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 [8] and the harmonic partial pattern present by the cochlea [9]. Tonal perception allows us to perceive the amount of repetition of events that are too fast to be counted [10].

In music there are several standards that 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 [11].

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 [12]:

$$\Delta_{freq} = f_1 2^{\frac{\Delta_{notes}}{12}} \quad (3)$$

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

2. Related work

There are several methods that aim at finding the pitch of periodic signals such as Maximum likelihood [13], Spectral peak picking [14], Cepstrum [15], Harmonic Product e Sum Spectrum [13]. Two of them, Autocorrelation and Yin Method, were implemented and applied to a reference signal, which is shown in Figure 1, of an excerpt from a recording of an electric bass playing the note E0, which has approximately a fundamental frequency of 41, 2Hz and a fundamental period of 24.3 milliseconds.

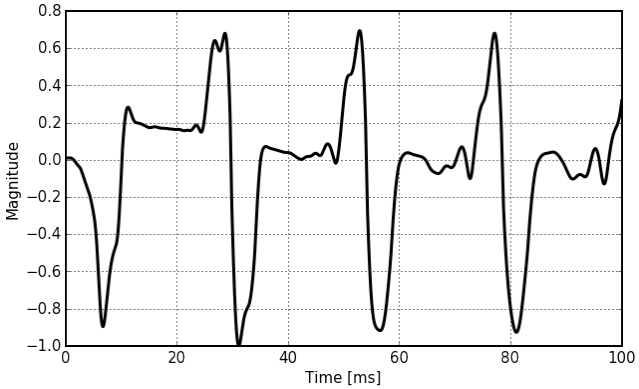

Figure 1. Waveform from a electric bass guitar's recorded signal. $f_0 = 44.1Hz$ and $T_0 = 24.2ms$. Its used as reference signal for the application of the following pitch detection methods.

2.1 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:

$$r_t(\tau) = \sum_{n=t+1}^{t+W_L} s_n s_{n+\tau} \tag{4}$$

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.

A common method for estimating pitch of periodic signals is by detecting the greatest positive peak of the autocorrelation function r_t [4], 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:

$$f_0 = \frac{1}{\tau_{max}}, \tau > 0, \tag{5}$$

tal que:

$$r_t(\tau_{max}) = max r_t(\tau) \tag{6}$$

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, which shows the autocorrelation function obtained from the reference signal, shown in Figure 1.

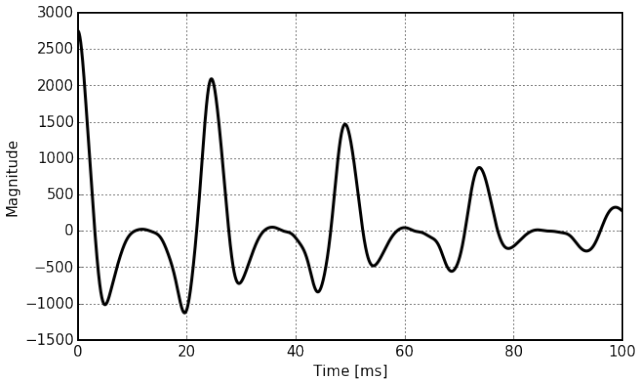

Figure 2. Autocorrelation function calculated from the waveform in Figure 1. First peak after the initial one occurs near 24.2*ms*, as expected.

2.2 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 [5]. Next, we briefly describe the improvements applied to each step according to the authors' study information.

2.2.1 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 f_0 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.

2.2.2 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 3 presents the difference function calculated from the waveform of Figure 1. The difference function is defined by the equation:

$$d_t(\tau) = \sum_{n=1}^{W_L} (s_n - s_{n+\tau})^2. \tag{7}$$

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.

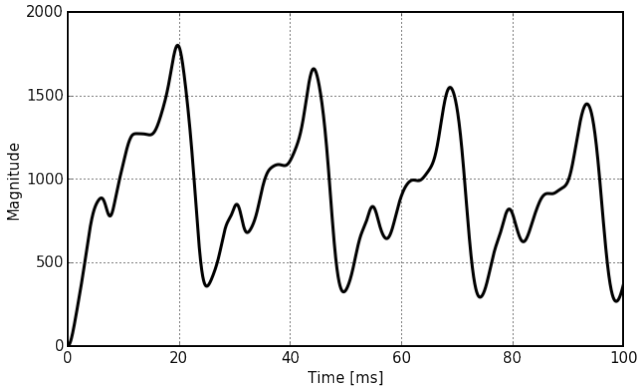

Figure 3. Difference function calculated from the waveform in Figure 1. First big dip after the initial one occurs near 24.2 ms, as expected.

2.2.3 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 4, 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 frequency dip. The (CMNDF) is defined by:

$$d_t^{'}(\tau) = \left\{ \begin{array}{ll} 1 & ,\text{if } \tau = 0 \\ \dfrac{d_t(\tau)}{(1/\tau)\sum_{n=1}^{\tau} d_t(n)} & ,\text{otherwise} \end{array} \right\} \quad (8)$$

where $d_t(\tau)$ is the diference function defined in equation (7) and τ the lag in samples between the signal and the shifted version of itself in diference function.

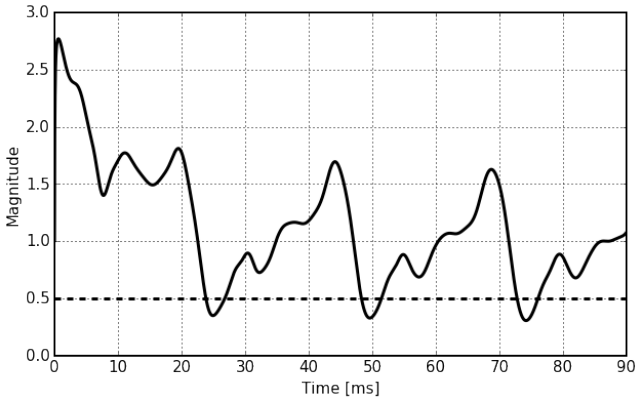

Figure 4. Cumulative mean normalized difference function calculated (CMNDF) from the waveform in Figure 1 and an absolute threshold on dashed line. First big dip occurs near 24.2 ms, as expected.

2.2.4 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 4. Using this threshold, dips above this value are disregarded, avoiding the selection of harmonic generated dips.

2.2.5 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.

2.2.6 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 1. 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 1.

2.3 Discussion about pitch estimation methods

Many methods, such as those discussed in the previous section, directly rely on the periodicity property as stated in Equation 1 or the harmonic series model shown in Equation 2. 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 [2] 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 [5] study, it is commonly used as a reference method, as in the study by [2], addressed in this work. It was also chosen because it is a well-known and cited method, as in the works of [16] and [17], also cited in this

work, counting more than 1300 citations according to the portal [18].

2.4 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^oC, 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 [19].

2.5 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 [19].

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 [20].

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.

2.6 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 [21].

A previous study [21] 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.

2.6.1 Latency Discussion

Table 2 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.

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 3.2, then the physical model discussed in Section 3.3 was used to generalize these results, as shown in Section 3.4.

3.2 Plucking an Electric Bass String

The traditional electric bass guitar is an electro-acoustic instrument with a body and neck made of wood and four metal string tuned to E, A, G and D, which are fixed in a metal bridge on the body and in the nuts of the neck. The neck has a fingerboard with 20 to 24 frets which divides it in tonal areas. The index and middle fingers of the right hand are used to pluck the strings and the fingertips of the left hand are used to hold the strings against the fretted fingerboard. This changes the free length of the string, which modulates its natural oscillation frequency.

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 5.

The waveform of the voltage signal at the pickups, as shown in Figure 8, 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 instru-

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. Tolerable latency - Instruments comparison using in ear monitoring (IEM) and wedge monitoring [21]

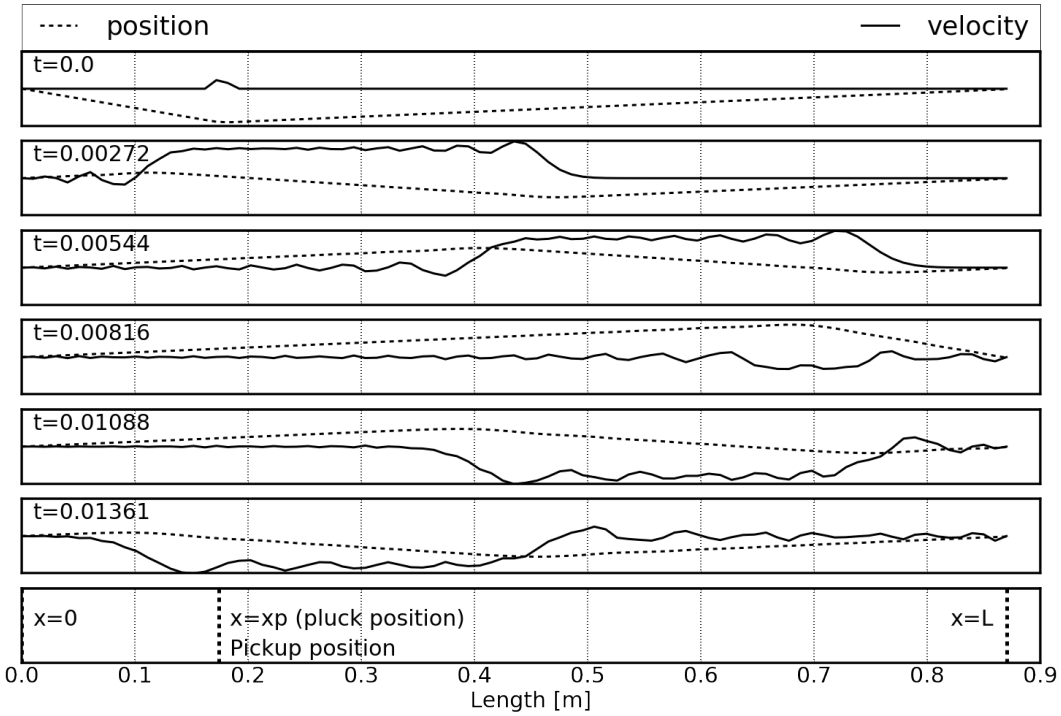

Figure 5. Position and velocity of a string along the x axis at different times t with fixed ends at $x = 0$ and $x = L$

ment), the behavior of its string was mathematically modeled, as discussed in the next section.

3.3 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:

$$y(x = 0, t) = 0. \tag{9}$$

$$y(x = L, t) = 0. \tag{10}$$

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 6. In this situation, the initial transverse displacement $y(x, 0)$ can be expressed by

$$y(x, t = 0) = \left\{ \begin{array}{cc} A(\frac{x}{x_p}) & , \text{if } x < x_p \\ A(1 - \frac{x - x_p}{L - x_p}) & , \text{otherwise} \end{array} \right\} \tag{11}$$

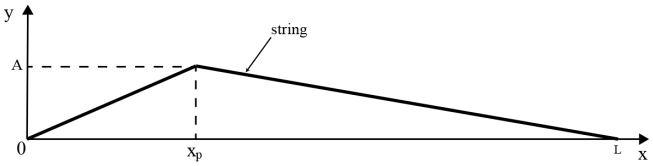

Figure 6. String with fixed ends at $x = 0$ and $x = L$ being plucked at $x = x_p$ with transversal displacement $y(x_p) = A$.

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

$$y'(x, t = 0) = 0. \tag{12}$$

As depicted in Figure 7, 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:

$$F = F_T \sin(\theta(x + \Delta x)) - F_T \sin(\theta(x)) \tag{13}$$

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 (13) as:

$$F = F_T (\frac{\partial y}{\partial x}(x + \Delta x) - \frac{\partial y}{\partial x}(x)) \tag{14}$$

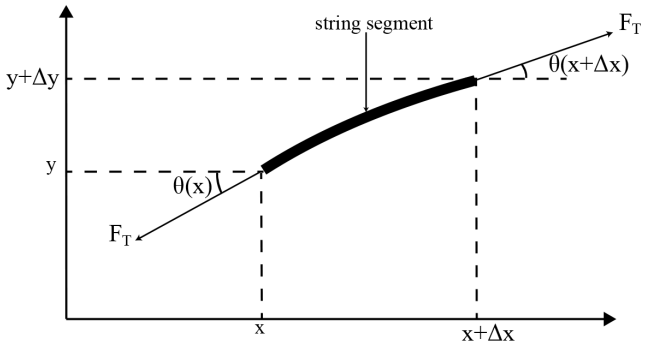

Figure 7. Short segment of a string between (x,y) and $(x+\Delta x, y+\Delta y)$ where a tension F_T is applied.

Using the Newton's second law:

$$F = m\frac{\partial^2 y}{\partial t^2} \tag{15}$$

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

$$F_T\left(\frac{\partial y}{\partial x}(x+\Delta x) - \frac{\partial y}{\partial x}(x)\right) = \mu\Delta x\frac{\partial^2 y}{\partial t^2} \tag{16}$$

dividing both sides of Equation (16) by Δx, applying the second derivative definition with $\Delta x \to 0$ and making $c = \sqrt{F_T/\mu}$, it becomes the wave equation:

$$\frac{\partial^2 y}{\partial t^2} = c^2\frac{\partial^2 y}{\partial x^2}, \qquad x \in (0,L), t \in (0,t_f] \tag{17}$$

This model was used to simulate plucked strings and the resulting waveforms were compared to measured waveforms, as discussed in Section 3.4.

3.4 Plucked string simulation

Equation 17 was numerically solved using the finite difference method [22] and the algorithmic steps used by Langtangen [23]. The Taylor series expansion was used to approximate it as:

$$\frac{y(x+\partial x,t) - 2y(x,t) + y(x-\partial x,t)}{\partial x^2} = \frac{1}{c^2}\frac{y(x,t+\partial t) - 2y(x,t) + y(x,t+\partial t)}{\partial t^2} \tag{18}$$

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 18 yields:

$$y_{i,j+1} = C^2 y_{i-1,j} + 2(1-C^2)y_{i,j} + C^2 y_{i+1,j} - y_{i,j-1}. \tag{19}$$

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 12 and Tailor's series as follows:

$$\frac{y(x,t+\partial t) - y(x,t-\partial t)}{2\partial t} = 0. \tag{20}$$

Rearranging equation 20 and rewriting in the i,j notation, we find that:

$$y_{i,j-1} = y_{i,j+1}. \tag{21}$$

Finally, replacing $y_{i,j-1}$ by $y_{i,j+1}$ in Equation 19, isolating $y_{i,j-1}$ and dividing both sides by 2, we have:

$$y_{i,j+1} = \frac{C^2}{2}y_{i-1,j} + (1-C^2)y_{i,j} + \frac{C^2}{2}y_{i+1,j}, \tag{22}$$

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 $c = f/(2L)$ [12] 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 11 and $y_{i,1}$ using Equation 22 and applying the boundary conditions from Equations 9 and 10. Then, for each element j from temporal domain, apply Equation 19 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 [24]. The resulting signals were compared to the recorded signals, as shown in Figure 8.

Figure 8 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.

4. Fundamental Frequency Estimation

The simulated and measured waveforms in Figure 8 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

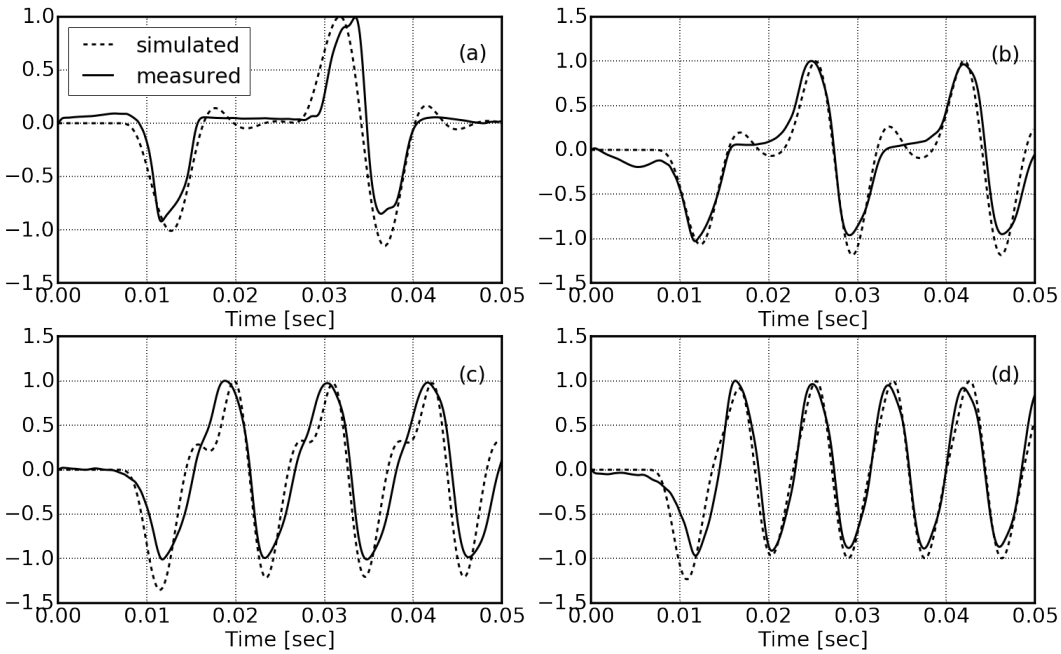

Figure 8. Simulated and measured notes played on string E of the electric bass guitar (a) E0 (b) A♯0 (c) F1 (d) A1

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 9.

4.1 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 [25]. As in the case of the electric bass guitar, there is usually a rapid and considerable increase in relative energy when a note is played, it is proposed to use the method based on the energy function:

$$En_n = \frac{1}{N_{en}} \sum_{n=1}^{N_{en}} |s_n|^2, \tag{23}$$

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.

4.2 Determine starting peak

When onset is detected, the algorithm will seek to determine the initial peak in the buffer, as expected according to figure 10. 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 10, which will be used in the following steps in the application of the absolute difference function.

4.3 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.

4.4 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 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:

$$d(\tau) = \sum_{n=1}^{W} |s_n - s_{n+\tau}|, \tag{24}$$

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 11.

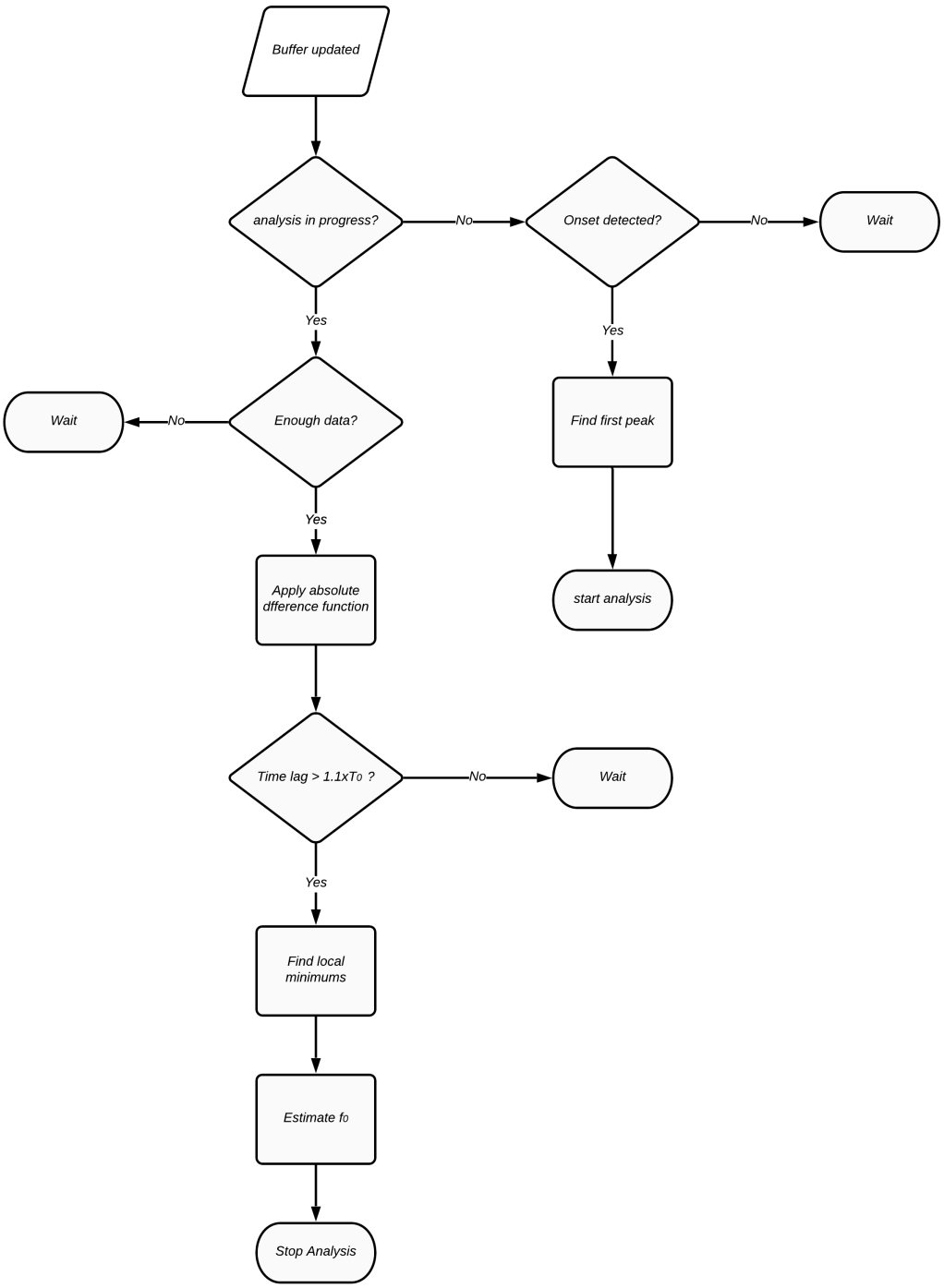

Figure 9. Proposed method flowchart process.

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

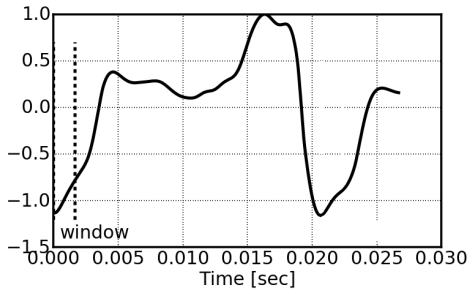

Figure 10. Analyzed signal s_n and integration short window with size W. This signal is a recording of the G0 note played on the E string of an electric bass guitar

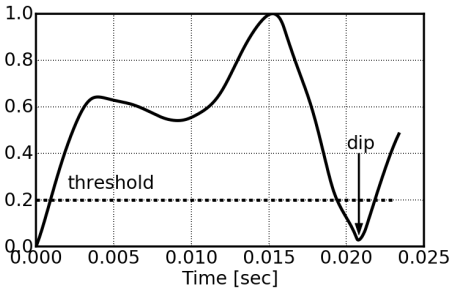

Figure 11. Absolute difference function from the analyzed signal s_n from figure 10 and threshold value represented as the horizontal dotted line

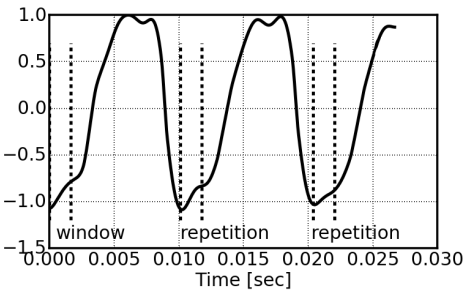

Figure 12. Analyzed signal s_n and integration short window with size W. This signal is a recording of the G1 note played on the E string of an electric bass guitar

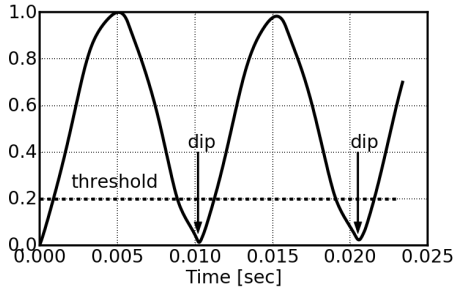

Figure 13. Absolute difference function from the analyzed signal s_n from figure 12 and threshold value represented as the horizontal dotted line

peaks of the signal. Thus, $1.1 \times T_0$ gives a margin of tolerance.

4.5 Find local minima

In sequence, the algorithm searches for local minima in the absolute difference function, referenced as dips in Figure 11. For the lowest notes, there will be only a local minimum as depicted in Figure 11, from which we will obtain the τ_0 interval. For the highest notes, as exemplified in Figure 12, there may be 2 or 3 local minimums, as depicture in Figure 13, depending on how many frets the bass has. In this case, τ_0 is obtained by:

$$\tau_0 = \sum_{n=1}^{N_\tau} \frac{\tau_n}{n}, \tag{25}$$

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

4.6 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:

$$f_0 = \frac{1}{T_0} \tag{26}$$

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 10. 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 [5], as shown in Figure 14.

5. Experiments and results

5.1 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.

5.2 Experiments

This section describes experiments that compare the proposed method to the Yin method [5], as implemented by Guyot [26]. The experiments comprised executing both the proposed method and the Yin method to estimate the f0 in the dataset samples.

5.2.1 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

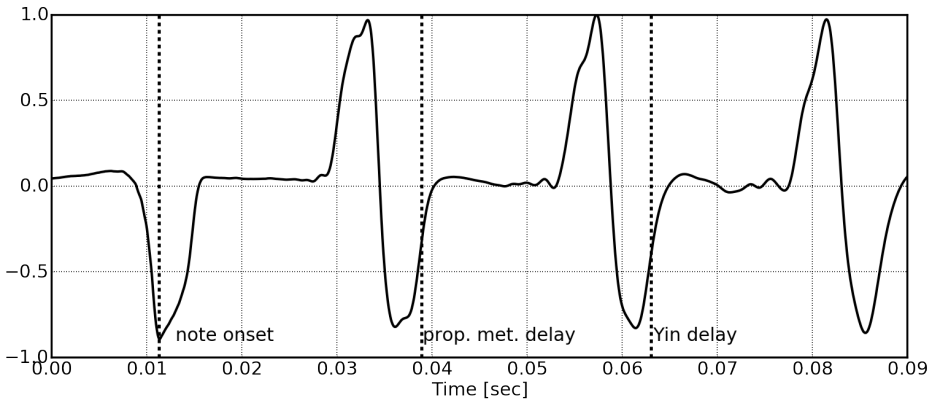

Figure 14. Algorithmic delay for the proposed method and for the Yin method.

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 16 (a).

5.2.2 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 16 (b).

Figure 15 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.

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

$$M_{note} = 12\log\left(\frac{f_0}{16.351597}\right)\frac{1}{\log(2)}, \tag{27}$$

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.

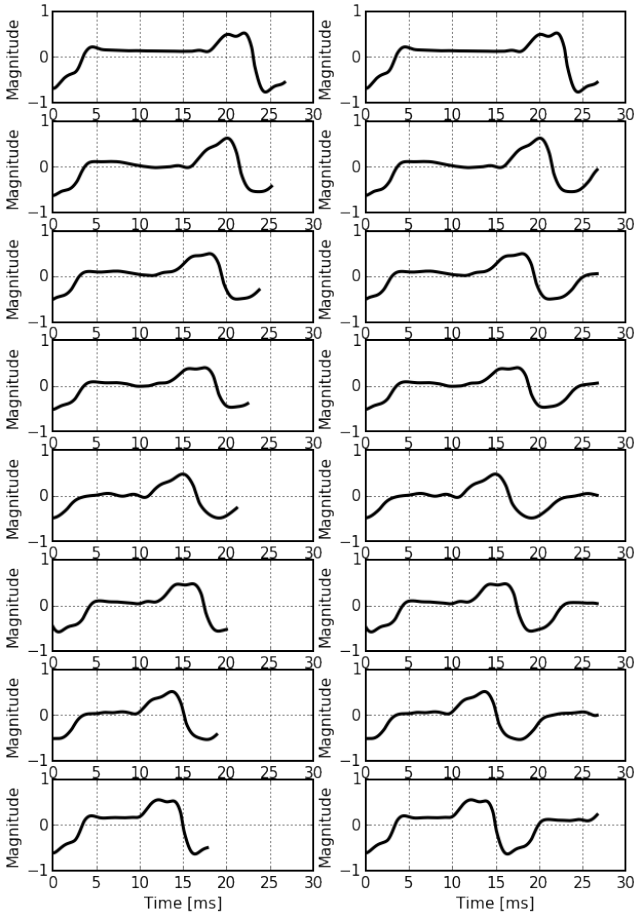

Figure 15. sample lengths for test 1 in the first column and for test 2 in the second column

5.3 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 [27].

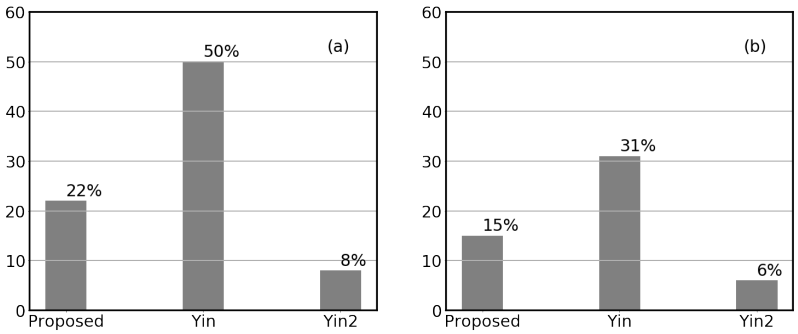

Figure 16. (a) Test 1 error rates . (b) Test 2 error rates

5.3.1 Electric Guitar

The waveforms illustrated in Figure 17 (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 17 (b) and (d).

5.3.2 Acoustic Guitar

The waveforms illustrated in Figure 18 (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 18 (b) and (d).

Observing the waveform of the figure 19 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.

5.3.3 Upright Piano

The waveforms illustrated in Figure 20 (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 18 (b) and (d).

From the waveform shown in figure 21 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.

5.3.4 Wooden Recorder

The waveforms illustrated in Figure 22 (a) and (c) were obtained from samples recorded from a "Venus" wooden recorder through a microphone.

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 22 (b) and (d), making it possible to determine the fundamental frequency.

As may be observed in figure 23, 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.

5.4 Discussion

The error rates presented in Figure 16 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.

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 ($2/f0 = 2,1/41.20Hz \approx 0,051s$),

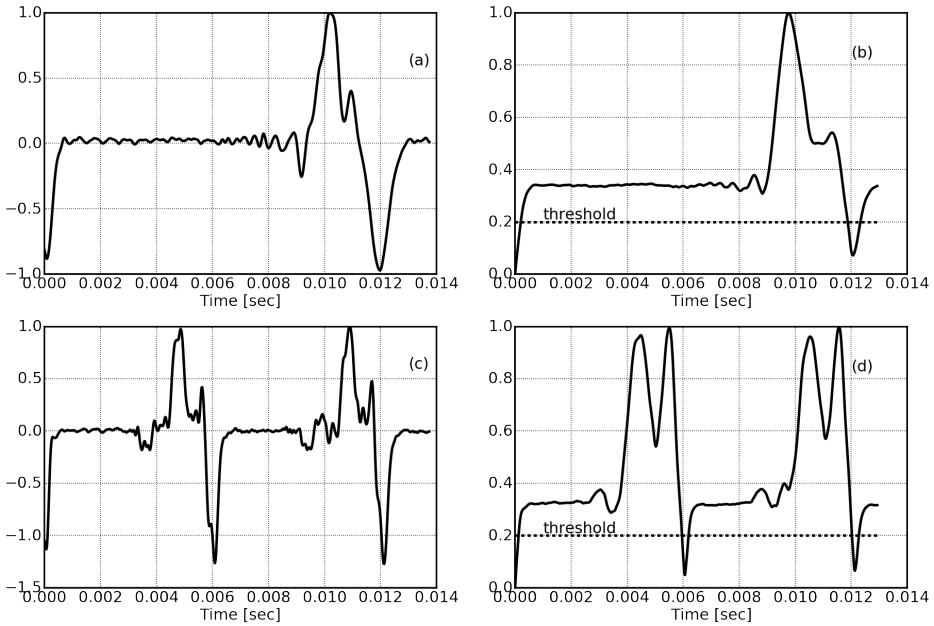

Figure 17. Analyzed signals from an electric guitar: (a) note $E1$; (c) note $E2$. Absolute difference function: (b) from signal in Figure (a); (d) from signal in Figure (c).

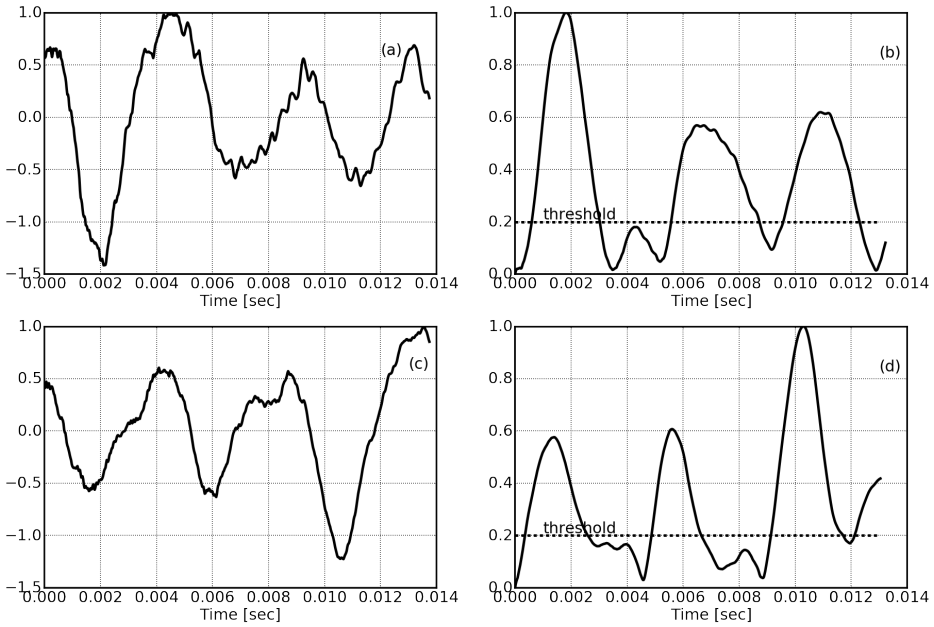

Figure 18. Analyzed signals from an acoustic guitar: (a) note $E1$; (c) note $E2$. Absolute difference function: (b) from signal in Figure(a); (d) from signal in Figure (c).

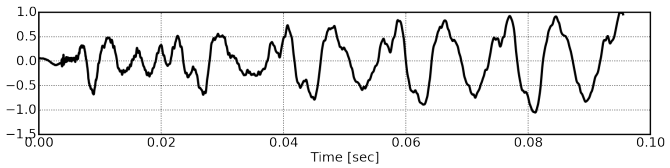

Figure 19. Waveform from an acoustic guitar attack and first cycles.

which is perceptible for a bass player, making it harder to play the bass guitar with real-time MIDI outputs, as shown in the [21] study, where professional bassists deemed acceptable latencies of up to 30 ms.

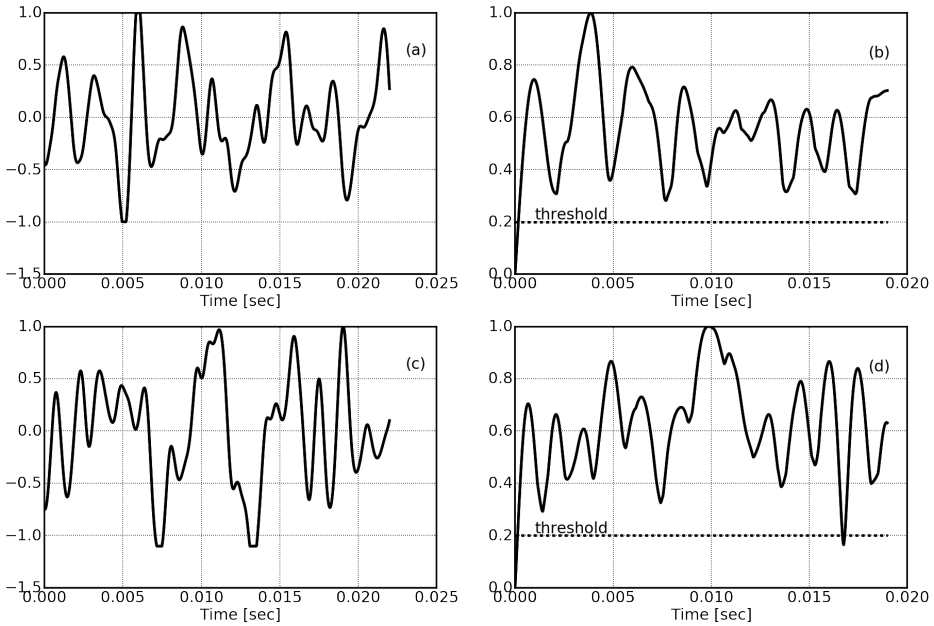

Figure 20. Analyzed signals from an Upright Piano: (a) note $A0$; (c) note $A1$. Absolute difference function: (b) from signal in Figure(a); (d) from signal in Figure (c).

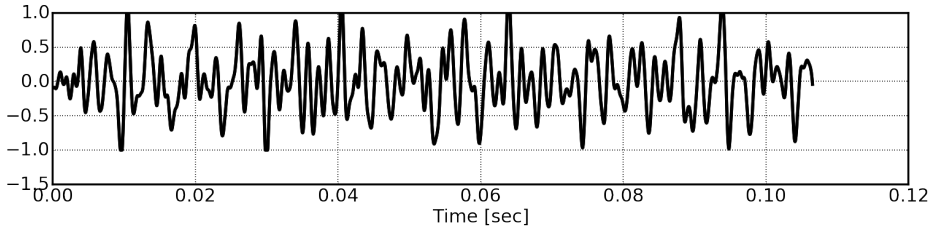

Figure 21. Waveform from an Upright Piano attack and first cycles.

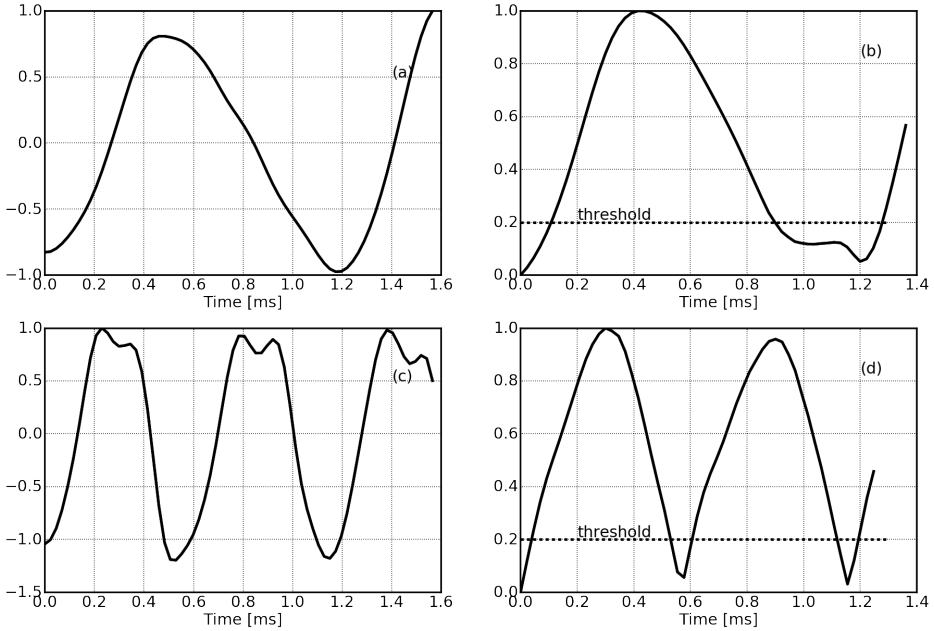

Figure 22. Analyzed signals from a wooden recorder: (a) note $A5$; (c) note $A6$. Absolute difference function: (b) from signal in Figure(a); (d) from signal in Figure (c).

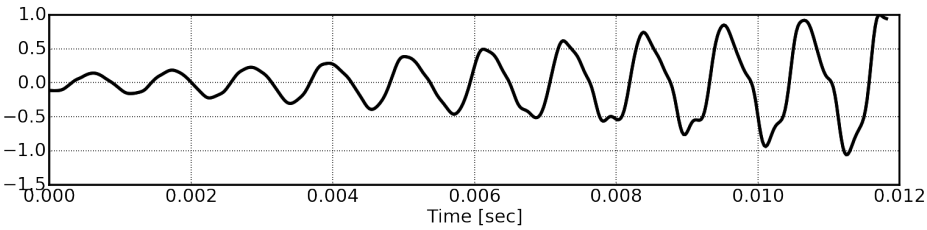

Figure 23. Waveform from an wooden recorder attack and first cycles.

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.

6. 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.

7. Acknowledgements

Thanks are due to Mario Junior Patreze from ”Escola de Música de Piracicaba Maestro Ernst Mahle”, Marcio H. Goldschmidt and Giovani Guerra from musical studio ”Esgoto” for the records from their own electric bass guitars which compose the database of this work. This study was financed in part by the ”Coordenação de Aperfeiçoamento de Pessoal de Nível Superior” - Brasil (CAPES) - Finance Code 001.

Author contributions

Christhian Fonseca developed the study as his master’s research, supervised by Tiago Tavares.

References

[1] GIBSON, J.; WARREN, A. *The MIDI Standard*. [S.l.]: http://www.indiana.edu/ emusic/361/midi.htm, accessed 05/9/2019.

[2] DERRIEN, O. A very low latency pitch tracker for audio to midi conversion. *17th International Conference on Digital Audio Effects (DAFx-14)*, 2014.

[3] KLAPURI, A. P. Multiple fundamental frequency estimation based on harmonicity and spectral smoothness. *IEEE Trans. Speech and Audio Proc*, 2003.

[4] 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.

[5] 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.

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

[7] OXENHAM, A. J. Pitch perception. *Journal of Neuroscience*, v. 32, n. 39, p. 13335–13338, 26 September 2012.

[8] CARIANI, P. A.; DELGUTTE, B. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J. Neurophysiol. 76*, 1996.

[9] TERHARDT, E. Pitch, consonance and harmony. *J. Acoust. Soc. Am. 55*, 1974.

[10] FORNARI, J. *Percepção, Cognição e Afeto Musical*. [S.l.: s.n.], 2010.

[11] ISO16:1975-ACOUSTICS. *Standard tuning frequency*. [S.l.]: International Organization for Standardization, 1975.

[12] IAZZETTA, F. *Tutoriais de Áudio e Acústica*. [S.l.]: http://www2.eca.usp.br/prof/iazzetta/tutor/acustica, accessed 04/25/2019.

[13] 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/⟩.

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

[15] 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.

[16] GERHARD, D. Pitch extraction and fundamental frequency: History and current techniques. *Technical Report TR-CS 2003-06*, 2003.

[17] KNESEBECK, A.; ZOLZER, U. Comparison of pitch trackers for real-time guitar effects. *13th Int. Conference on Digital Audio Effects (DAFx-10)*, 2010.

[18] RESEARCHGATE. *YIN, A fundamental frequency estimator for speech and music*. [S.l.]: https://www.researchgate.net/publication/11367890 YIN A fundamental frequency estimator for speech and music, accessed 06/02/2020.

[19] 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.

[20] WANG, Y. Low latency audio processing. *Queen Mary University of London, School of Electronic Engineering and Computer Science*, 2017.

[21] LESTER, M.; BOLEY, J. The effects of latency on live sound monitoring. *Journal of the Audio Engineering Society*, 2007.

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

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

[24] 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).

[25] 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.

[26] GUYOT, P. Fast python implementation of the yin algorithm. *http://doi.org/10.5281/zenodo.1220947*, 2018. "accessed 01/02/2018".

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