

Optical Digital Theremin with Audio Synthesis and Graphic Interface

Marcos Ramos dos Santos and Andrés Eduardo Coca Salazar,
Computer engineering coordination (COENC),
Federal University of Technology - Paraná (UTFPR-TD),
Toledo-PR, Brazil

Received: April 19, 2021. Received: September 28, 2021. Accepted: October 23, 2021. Published: November 3, 2021.

Abstract- The theremin is one of the first electronic musical instruments and one of the few played without physical contact since it only requires hand and finger movements to control the amplitude and frequency of the musical note. However, the capacitive functioning of the antennas increases the sensitivity to electrical interference, its timbre is fixed, and the frequency antenna's vertical arrangement could limit the use of people with amputated fingers. Furthermore, it does not contain any help to guide the execution, which makes it a very difficult instrument to play. In this paper, we present the development of a digital optical theremin with an audio synthesis process, intuitive graphical interface, frequency antenna in the horizontal position, and linearization of the frequency-distance relationship. These features are intended to aid learning and interpretation of the instrument and extend access to people with finger limitations. In order to validate the instrument's behavior and characteristics, we conducted three experiments: 1) accuracy analysis of the linearization through the mean absolute error in units of cents and the Kruskal-Wallis statistical inference test, 2) validation of the steps of the audio synthesis module, and 3) checking of the timbral diversity, both through the Fourier spectrum. This prototype could be used as an auxiliary tool in musical initiation and the development of musical perception.

Keywords- Electronic musical instruments, Optical sensors, Subtractive Synthesis.

I. INTRODUCTION

MUSICAL instruments are the fundamental basis for the creation and interpretation of music since without them it would not be possible to produce sound in a controlled manner. Hence, throughout the history of music new instruments have appeared to create innovative timbres and, in some cases, to facilitate interpretation [1].

Electronics plays a fundamental role in this evolution through the development of new instruments, capable

of breaking the physical limitations of conventional mechanical instruments [2]. In this context, the theremin is one of the most significant examples, especially because it is played without physical contact. This instrument was patented in 1928, but created in 1920 by Leon Theremin, while he was working on methods for measuring high frequencies. In one of his experiments, he realized that the approach of the hands to the antennas interfered with the capacitance of the circuits and used it to control the oscillators frequency. This phenomenon was used by him to create a revolutionary and innovative instrument as its form of interpretation [3]. This was one of the milestones for the evolution of the music scene, being considered by many to be the father of analog synthesizers, which in turn shaped the construction of electronic music [4].

However, the operating mode of the antennas creates some problems. Musicians report that the instrument suffers interference from objects within a radius of up to three meters away, which modifies its tuning and makes interpretation difficult, a problem that is commonly observed in situations where other musicians are moving around during a performance, especially in small spaces [5]. Besides, it doesn't allow changing the timbre, the frequency-distance relationship contains a non-linear region, and it doesn't contain aids that guide the interpretation such as keyboard or fretboard, therefore, it is considered one of the hardest instruments to play [6]. Generally, due to the vertical arrangement of the frequency antenna, notes are played with finger movements of the right hand, which restricts use to people who cannot perform these movements.

Over the years, many studies have been carried out to solve the interference problem, so it is possible to find different types of research on the subject, as well as some reinterpretations of the instrument. Bearing this in mind, Gomes et al. proposed the TheremUS, which detects hand movements using ultrasonic sensors and, through a DSP, maps the distance with the frequency of musical notes and applies an FM synthesis for sound generation [7]. In addition, Louis Bertrand Castel's Harpsichord Ocular concept was applied to relate the frequency range of notes to visible light, using a pulse width modulation (PWM) signal that controls the light intensity of an array of RGB LEDs, whose col-

ors can be visualized through the translucent box that encloses the instrument. Another prototype that uses ultrasonic sensors was proposed by Hanindhito et al., but in this one, the mapping is performed in an FPGA and the sound generation with a CORDIC (Coordinate Rotation Digital Computer) digital circuit [8]. Moreover, there are the T-Voks developed by Xiao et al., where a theremin is used to control a speech synthesizer. The frequency antenna controls the tone of the expression while the other controls, in addition to the volume, the voice quality and vocal effort. The voice utterances are pre-recorded and the syllabic sequencing is done through an additional pressure sensor connected to the interpreter's hand [9].

In this paper, we propose a low-cost optical theremin with audio synthesis and a graphical interface. In our prototype, the movements are detected by optical sensors in a horizontal position and the distance is linearly mapped by the Arduino platform microcontroller. The obtained data are sent to a computer responsible for generating the subtractive synthesis and running the graphical interface. Consequently, timbre can be modified and the interface contains visual aids that guide interpretation and learning and shows in a didactic way the steps of audio synthesis. In the experimental stage, we conducted tests to analyze the accuracy of the frequency sensor signal mapping, to validate the synthesis module's operating mode, and to check its timbrical diversity. The first was performed using the mean absolute error (MAE) of the frequency divergence in cents and the Kruskal-Wallis statistical inference test, while the last two followed the Fourier spectrum analysis.

The main contributions of this project are: 1) the inclusion of a user-friendly software interface that facilitates the interpretation of the instruments and that can be used in the auditory training of beginning musicians, 2) the presentation of the steps of the process of subtractive audio synthesis, useful for the understanding and studio of signals by engineering students, 3) the horizontal arrangement of the frequency antenna together with the linearization of the distance-frequency relationship, which was selected to increase accessibility for people with physical limitations, especially with amputation of fingers.

The rest of the paper is organized as follows. Section II describes the theoretical framework, which includes the concepts needed to understand the electronic functioning of the instrument, and the related works. Section III describes the development and implementation of the instrument. Section IV presents the experimental results, and Section V concludes the paper.

II. THEORETICAL BACKGROUND

In this section, we will present some important concepts about the mode of operation of the original theremin, as well as a state-of-the-art review.

A. The electronics behind the instrument

Theremins commonly have one or more beat frequency oscillator (BFO) modules to control the fre-

quency and amplitude of the signal. The BFO module contains two oscillators that generate independent signals with angular frequencies ω_1 and ω_2 , whose outputs are expressed mathematically as: $A_1 \sin \omega_1 t$ and $A_2 \sin \omega_2 t$, respectively [10]. The BFO module acts as a heterodyne mixer, which multiplies the two signals and provides an output (V_{out}) equal to:

$$V_{out} = A_1 A_2 \cdot \sin(\omega_1 t) \cdot \sin(\omega_2 t), \quad (1)$$

which by means of a trigonometric identity is expanded as:

$$V_{out} = \frac{A}{2} [\cos(\omega_1 - \omega_2)t - \cos(\omega_1 + \omega_2)t], \quad (2)$$

where $A = A_1 + A_2$. Therefore, the BFO produces two signals, that corresponding, respectively, with the sum and difference of the frequencies of the original signal. Using a low-pass filter, the part of the signal corresponding to the sum of the frequencies is eliminated, leaving only the difference component, like this: $V_{out} = \frac{A}{2} \cos(\omega_1 - \omega_2)t$. Then, the frequency of one of the two oscillators must be changed so that the beat frequency also changes [4].

The instrument is designed so that one of its oscillators is adjusted through the capacitive effect when moving the hand close to the antenna. As the oscillator operates on the order of MHz, a small displacement is capable of providing a variation of 1 kHz, which represents a significant fraction of the musical scale [10].

The capacitance effect can be understood through a Colpitts oscillator, whose operating frequency (f_o) is given by:

$$f_o = \frac{1}{2\pi\sqrt{L}} \left(\frac{1}{C_1} + \frac{1}{C_2} \right)^{\frac{1}{2}}, \quad (3)$$

where L represents the inductance of the circuit, C_2 the parallel association between antenna capacitance (C), dispersive capacitance (C_S) and all the other capacitances existing between the circuit and the antenna (C_A), that is, $C_2 = C + C_S + C_A$. The value of capacitance C_A varies according to the movement of the player's hands and modifies the instrument's tone; and C_1 is a preset value capacitor placed in series with C_2 . Initially the value of C_2 is designed so that the oscillator responsible for tuning has a frequency equal to zero for a given initial distance x_0 . Thus, the instrumentalist's hand is considered as an infinite ground plane approaching the antenna, and the presence of this plane can be modeled as the presence of an image antenna [10].

According to the relationship between the instrument's tonal range and the distance from the antenna, it has a non-linear relation, but in the central region, this non-linearity decreases. Nonetheless, the non-linearity that occurs at very small and very large distances allows a wide excursion between notes, which facilitates the performance of vibratos. Generally, when the instrument is turned on, it is necessary to adjust its tuning. This tuning is performed electromechanically using a rotary axis

that moves the disks of a capacitor connected in parallel with the pitch antenna [11].

The instrument's amplitude control uses two BFO oscillators, one with a fixed frequency, while the other with a variable frequency depending on the antenna's capacitance. Both oscillator outputs are connected to a mixer circuit, whose output is connected to a low-pass filter, which results in a signal whose amplitude varies according to disturbances in the antenna capacitance. The obtained signal is amplified and rectified through an integration filter, which provides a control signal DC (Direct Current) that varies directly as a function of the amplitude, which in turn varies according to the distance of the hand to the antenna. The control signal voltage increases as the hand move away from the antenna, typically the signal has a voltage of 0 volts when the hand is close to the antenna and 24 volts for the greatest projected distance. In this way, the volume is attenuated with the approach of the hand to the amplitude antenna [10].

B. Related works

With technological advances, different theremin prototypes have emerged, for example, Geiger et al. proposed a set of four different 3D interfaces (VRemin I to IV) for a virtual theremin, as follows: interface I use the Wii and Nunchuck controllers; the first to control pitch and volume, and the second to add effects; II uses a sensor coupled to a glove, a camera, and an image analysis process to perform the mapping; III uses an ultra-red sensor for both hands simultaneously; IV uses a sensor on the musician's shirt to map the position of the body [12]. In [13] the authors performed a formal evaluation of variants I and II according to hedonic and pragmatic qualities through the AttrakDiff questionnaire, and a comparative classification between different variants according to interface complexity and interaction space.

On the other hand, the Quad-Theremin variant was proposed by Liu et al., which consists of an expanded theremin with four antennas to control 4 MIDI channels: pitch, volume, timbre, and the tremolo effect. Theremins were implemented using an analog circuit, DSB modulation, and low-pass filtering [14]. Another variant called radarTHEREMIN was proposed by Wöldecke et al., in which the musician is in front of a virtual screen that contains a melodic table and interacts with a multi-touch device based on an infrared laser [15].

The RFID-Vox variant was proposed by Nikitin et al., and it uses UHF RFID technology through the commercial Gen2 reader. Hand movements are transmitted by two markers, and the RFID reader can send analog or discrete signals to the music controller. A differential of this variant is that it can be run over long distances, according to the authors, up to more than 20 meters [16].

The variant proposed by Sharma et al. uses photodiodes to map incident light intensity to the frequency and volume of the musical note, and these parameters are used to generate true random number sequences (TRNG), which can be used in encryption algorithms,

such as SSH/SSL. The signal from the photodiodes is pre-amplified with an analog circuit and then connected to a NI myDAQ device that continuously samples and produces the audio signal. The device has a LabVIEW interface for configuring signal intervals and activating an autotuner [17].

In turn, Cervera proposed a theremin based on the Leap Motion device, which has two infrared cameras and recognizes hand and arm gestures. However, the author detected and explained several disadvantages found in this prototype [18]. Later, Johnson et al. created VRMin by adapting Mixed Reality (MR) to the theremin [13] and García et al. created the hybrid digital theremin by using digital gates and a data acquisition board [19].

Also, the theremin has been used in education [20] [21], specifically, for teaching basic electronics and circuits through the construction of an analogue theremin [22], and for learning geometry [23]. Other researches are focused on the development of algorithms to teach robots to play the theremin, for example, HRP-2 and Asimo robots [24], and the iCub robot [13].

Note that the original theremin and most new prototypes do not cover accessibility criteria. With that in mind, in our prototype, the pitch antenna is in a horizontal position aiming at accessibility to people who are unable to perform right hand finger movements.

III. PROTOTYPE DEVELOPMENT

In this section, we will present and explain the operating mode and the elements of the proposed digital theremin prototype. Figure 1 presents a block diagram that depicts the steps and processes of the developed prototype.

The prototype has a pair of laser distance sensors responsible for detecting the movement of the musician's hands, whose collected data are processed by the microcontroller of an Arduino Nano. Data is sent to PC via serial port, and this is processed by synthesis module, whose parameters can be modified through a graphical interface. Finally, the audio signal is generated and amplified. Figure 2 presents the prototype of the optical digital theremin developed, specifically, Fig. 2(a) its internal view and Fig. 2(b) its external view.

A. Motion detection

To detect hand movements, two VL53L0X laser distance sensors were used. The distance measured by the first sensor is used to map the amplitude, while the second one to map the frequency of the musical note. This sensor model operates in a tension voltage range of 3 to 5 volts and is capable of accurately measuring distances between 30 and 1000 mm regardless of the detected object's reflectance. For detection, this sensor emits a laser with a wavelength of 940 nm, which is reflected in the presence of an obstacle, and then, through the reflection time, the distance between the object and the sensor is determined. Its communication with the microcontroller is carried out through the I2C protocol.

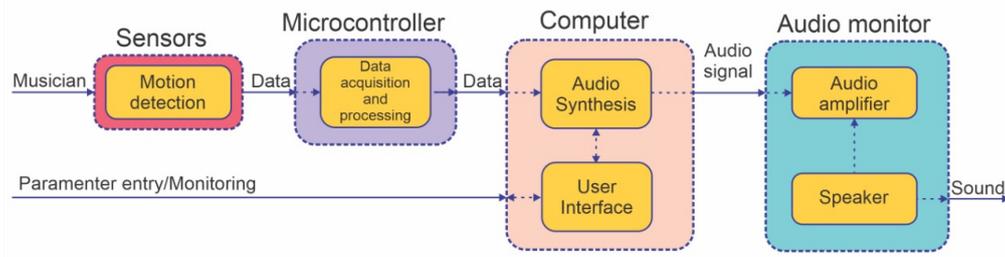
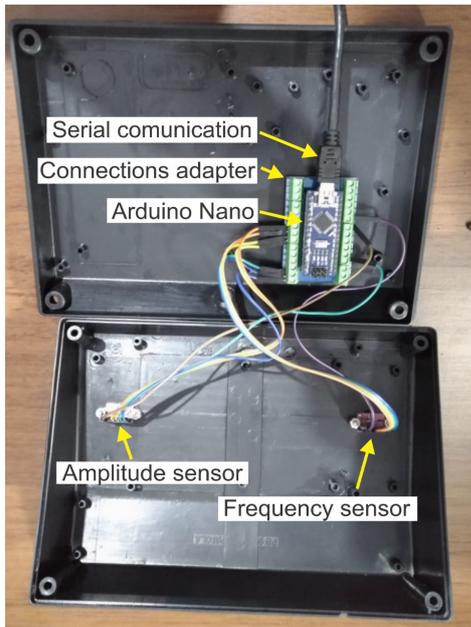
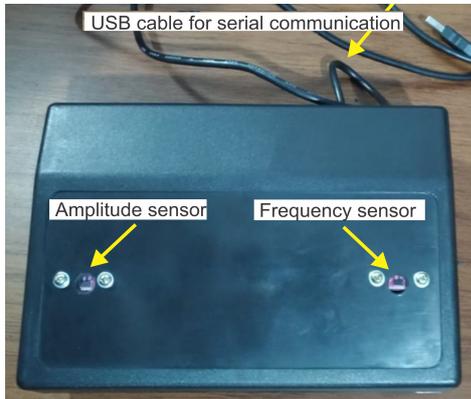


Fig. 1: Theremin block diagram developed.



(a) Internal view



(b) External view

Fig. 2: Prototype of the optical digital theremin developed.

B. Data acquisition and processing

For acquisition and processing of sensor data, we use the digital platform Arduino Nano V3.0, which is equipped with an ATmega328 microcontroller, 2kB SRAM memory, 32kB flash memory, 14 digital ports, 8 analog ports, and 5V operating voltage. This platform is responsible for reading the signals from the sensors and

mapping them at a suitable interval to later be sent to the computer.

The distance (x_1) read by the first sensor is limited to a range of 0 to 200 mm, which is mapped to intensity values (I) from 0 to 100 units, so: $I = x_1/2$. While the distance (x_2) read by the second sensor is limited in a range from 57 to 777 mm and mapped inversely proportional to semitones values (S) in a range from 60 to 0 units, which corresponds with five octaves and its calculation is given by:

$$S = \frac{1}{12} (777 - x_2). \quad (4)$$

The intervals obtained through the previous mappings are used by the synthesis module to generate the musical note and its timbre.

C. Audio synthesis

The synthesis step is responsible for creating the texture sound, that is, it is responsible for shaping the generated sound, imprinting the desired characteristics on the timbre according to the musician's preference. This process is executed by the PC from data received via Arduino's serial port, and it was implemented in Python language, together with the help of the dedicated module for digital signal processing called PYO [6]. The synthesis method adopted was the subtractive synthesis, which consists of generating a primitive signal rich in harmonics and then filtering it to remove specific components and sculpting the sound [25]. Figure 3 presents a block diagram to understand the synthesis process implemented, whose main steps will be described below.

Frequency data handling: To determine the frequency of the audio signal, the synthesis algorithm determines the frequency (f) of the musical notes as:

$$f = 2^{\left(\frac{S}{12}\right)} \times f_0, \quad (5)$$

where f_0 is the reference frequency, and S the number of semitones between the reference note f_0 and the note played by the musician (calculated with the Eq. (4)). Here, we designate the reference frequency as 55 Hz, which corresponds with the A note from octave one (A1). To control the signal amplitude, the algorithm determines the amplitude A of the musical note from the equation:

$$A = I \times 0.01, \quad (6)$$

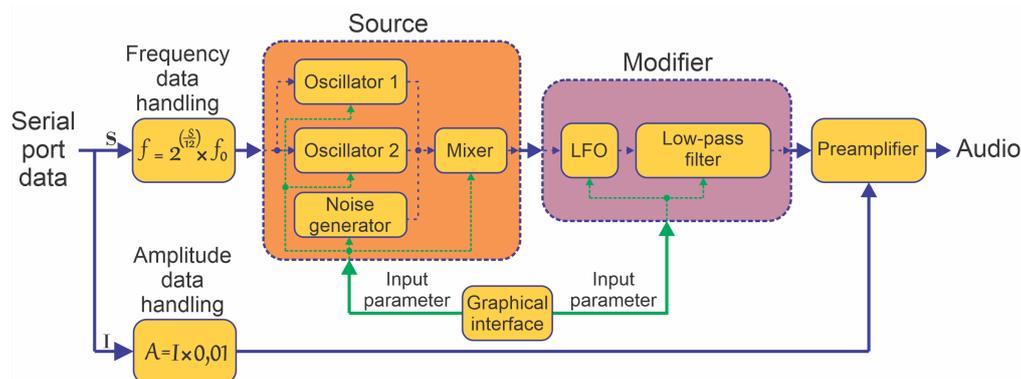


Fig. 3: Synthesis module block diagram.

where I is the intensity value received from the second sensor, which varies between 0 and 100, and the value 0.01 multiplies the received values causing A to vary in the range from 0 to 1, which corresponds to the range of values used by the pre-amplifier block.

Source block: Although some forensic applications [26] aim to discriminate and separate noise from acoustic signals to highlight information from the audio signal, in our prototype the addition of noise plays an important role in enriching the time-frequency information of the acoustic signal. So, this block contains a noise generator and two oscillators, both connected to a mixer.

The two implemented oscillators can generate eight different wave types, namely: rising sawtooth, falling sawtooth, square, triangular, pulse, bipolar pulse, sample & hold, and sine. Both have an octave control to modify the region of the produced notes and tuning control. The noise generator can alternate among pink, white, and brown noise; and it allows to add variability to the pure signal. Finally, the mixer combines the signals from both oscillators and adds the noise given by the generator.

Modifier block: This block consists of an LFO (Low Frequency Oscillator) and a low-pass filter. The frequency of the LFO's internal signal can be changed between 0 and 20 kHz. The LFO is applied over the sharpness parameter of the signal emitted by the source block. The parameter in question is responsible for incrementing and decreasing the number of harmonics, adding dynamics and movement to the sound. Also, to alter the spectral components and further expand the instrument's tonal diversity, the modifier block has a low-pass filter with modifiable cut-off frequency and resonance parameters.

Pre-amplifier block: Finally, in this block, the amplitude of the low-pass filter output signal is modified according to the intensity value (I) coming from the volume sensor and adjusted by Eq. (6).

D. User interface

To facilitate the visualization of the signals, as well as the configuration of the synthesis module parameters, a user interface was implemented. Through this interface, the user can check in real-time the musical note being

played, its amplitude on a scale from 0 to 100, and modify the timbre according to his taste. This screen also contains three buttons to adjust the delay effect over time, feedback, or amplitude, which causes an "echo" in the instrument timbre and creates an ambiance in the sound generated. The interface was developed using Python 3.6 language through Visual Studio Code editor, PyQT 5 library [27], QT Designer [28] and Corel Draw 2020. The main screen of the user interface is shown in Fig. 4(a).

The audio synthesis module interface screen, shown in Fig. 4(b), was developed so that the user can visually perceive how the instrument's synthesis process chain is organized. Here the user can directly change the waveforms generated by the oscillators, these being: octave and volume of each oscillator, pitch, type of noise, noise level, LFO waveform, amplitude, frequency, and sharpness; and the cutoff frequency of the low-pass filter and its resonance values.

E. Audio amplifier and sound projection

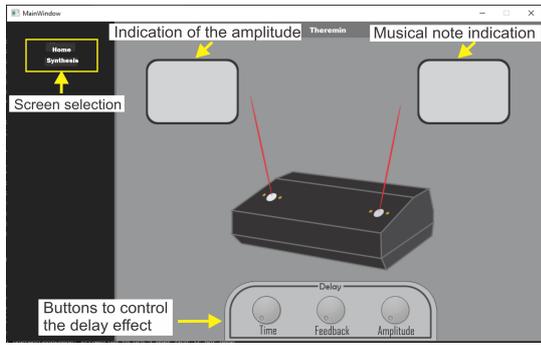
The signal generated by the synthesis module can be heard through the PC speakers, but to obtain a greater sound projection needed in live performances, it was connected an external audio amplifier model Edifier R1100, which consists of a pair of speakers, one of them passive and the other active, an embedded amplifier circuit, a power of 42 W RMS, input sensitivity 700 ± 50 mV, signal-to-noise ratio 25 dB, and frequency response 55 Hz - 20 kHz.

IV. EXPERIMENTAL RESULTS

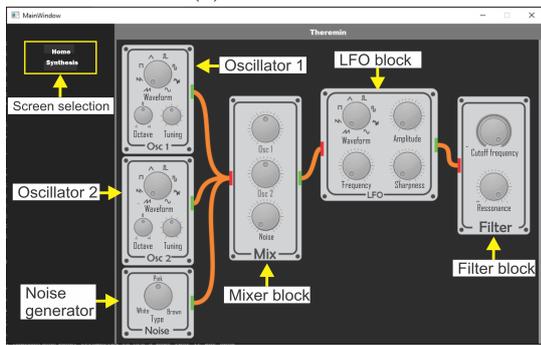
In order to test the technical and functional behavior of the prototype, three types of experiments were carried out: 1) analysis of the frequency-distance relationship accuracy, 2) validation of the audio synthesis operation mode, and 3) checking the diversity of timbres produced by the synthesis module. These experiments and their results are described below.

1 - Frequency-distance relationship accuracy:

A feature of the original theremin is the non-linearity in the frequency-distance relationship [10]. However, in the proposed prototype a linear equation (Eq. (4)) was used to map these two variables. Then, in this experiment,



(a) Main screen



(b) Synthesis module screen

Fig. 4: User interface to modify and adjust theremin parameters.

the accuracy of this linear mapping is objectively and statistically analyzed.

As explained in Section III-B., the operating distance range of the sensor is 720 mm, which has been divided into 60 equidistant parts, therefore, the separation between semitones is 12 mm and, consequently, each note is separated from its consecutive octave by a distance 144 mm. This can be expressed mathematically as:

$$d_{i,s} = 57 + 12S + 144(i - 1), \quad (7)$$

where $d_{i,s}$ is the distance in mm of a note separated from A1 (55 Hz) S semitones ($0 \geq S < 12$) and i octaves ($1 \leq i \leq 6$). Figure 5 shows the base 2 semilog plot of the frequency in Hz versus the distance in mm, for the reference frequency (f_{ref}) and for the frequency measured (f_m) of A note in different octaves at the preset distance, which was calculated with Eq. (7). It can be seen that even with the distance being fixed, some frequencies deviate from the expected value, for example, at a distance of 633 mm the frequency should be 110 Hz, but 114.87 Hz (highlighted with zoom in mini-chart) was measured. This error varies in each measurement.

For a better understanding, the error is converted to cents (C), like this: $C = 1200 \cdot \log_2(f_m/f_{ref})$ [29]. Figure 6 shows the error in cents for different A3 ($f_{ref} = 220$ Hz) measurements. It can be seen that there are measurements with an error of up to 40 cents, which is considerable since a trained ear can perceive frequency differences between 5 and 10 cents [30].

Figure 7 shows the box diagram of the error in cents

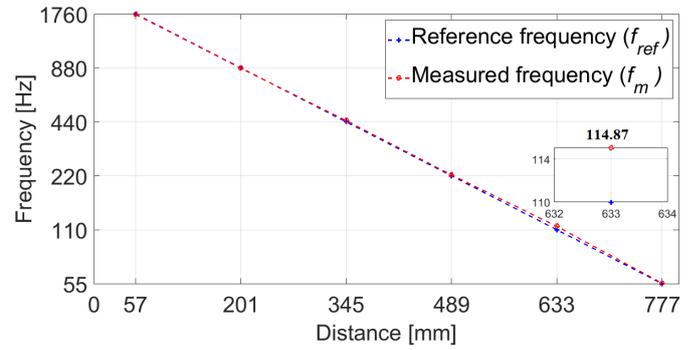


Fig. 5: Relation between reference frequency (f_{ref}) and measured frequency (f_m) and preset distance of A note in different octaves. The mini chart highlights the frequencies in 633 mm (A2).

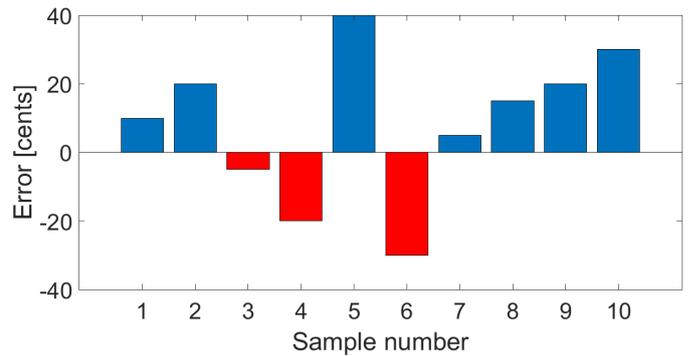


Fig. 6: Bar chart of the error in cents of the ten samples of A3 note (220 Hz) in Table 1.

of the 10 samples of A note in different octaves shown in Table 1. Note that the median and variability are greater in A2 samples. This divergence was quantified and summarized through the mean absolute error (MAE) and is presented in the last column and last line of Table 1. Note that samples from the second octave produce the highest MAE, 42.03 cents, and that the total MAE is 23.09 cents, which corresponds approximately to an octave-of-tone (1/8 the size of the whole-tone diatonic and equal to 25 cents).

In order to objectively check if the error divergence between octaves shown in Fig. 7 is statistically significant, we chose Kruskal-Wallis statistical inference test over the parametric equivalent ANOVA because the data rejected the null hypothesis of the Kolmogorov-Smirnov normality test [31]. For this purpose, we measured 60 frequencies from the A note in five octaves, from octave 1 to 6, in the predefined distances given by the Eq. (7), for a total of $N = 360$ samples. To apply the test we calculate the error in cents and set the level of significance in 1% ($\alpha = 0.01$). The result of Kruskal-Wallis test is shown in Table 2. The last column shows that for 5 degrees of freedom ($df = 5$) we obtained a theoretical value of χ^2 equal to 15.09, and with the sample data the statistical variable H of Kruskal-Wallis is equal to 13.64. Therefore, as H (13.64) is smaller than the critical value (15.09), we can infer at the 1% level of significance that

Table 1: Error in cents of 10 samples of A note in different octaves and Mean Absolute Error (MAE) for each octave (last column) and for each sample number (last row).

Note	Sample										MAE
	1	2	3	4	5	6	7	8	9	10	
A1	29.96	10.04	79.96	20.03	39.83	34.90	5.03	64.87	25.00	39.83	34.95
A2	15.04	75.00	45.05	20.03	10.04	60.01	49.95	35.05	65.02	45.05	42.03
A3	9.97	20.03	-4.96	-20.02	39.98	-30.00	5.03	14.97	20.03	30.03	19.50
A4	-5.00	25.00	45.01	30.00	19.99	0	-19.98	-5.00	35.02	0	18.50
A5	-10.00	-5.00	-25.01	-15.00	-30.00	-20.00	-35.00	0	-5.00	-10.60	15.56
A6	-5.00	-10.00	0	-15.00	-10.00	0	-5.00	-20.00	-15.00	0	8.00
MAE	7.50	14.51	20.00	12.01	14.99	14.49	12.00	13.99	16.51	12.55	23.09

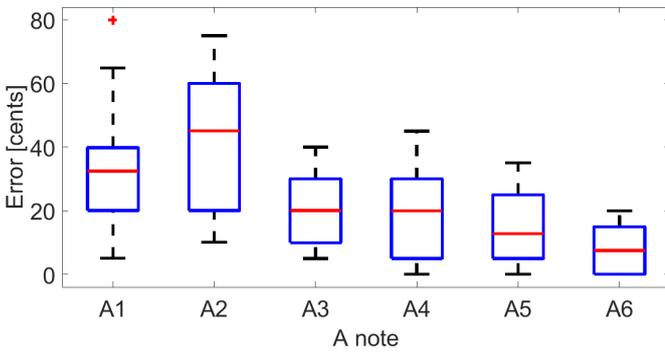


Fig. 7: Box diagram of the error in cents of the 10 samples of the A note in different octaves shown in the Table 1.

Table 2: Kruskal-Wallis test result.

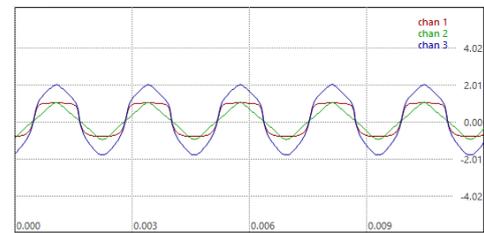
Octave	Samples	Sum of ranks	Parameter	Value
1	60	11025	df	5
2	60	12421.5	α	0.01
3	60	8775	χ^2	15.09
4	60	11934	H	13.64
5	60	10912.5		
6	60	9912		
N	360			

there is no significant evidence to reject the null hypothesis that the media of the octaves are equal. Therefore, the error divergence between octaves is not statistically significant.

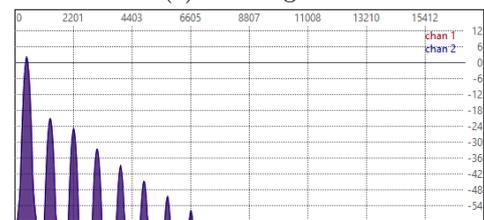
2 - Audio synthesis validation: To validate the functioning of the synthesis module, we analyze the behavior and characteristics of the output signal of each block of the diagram of Fig. 3.

Oscillators: The source block allows to create and combine different waveforms. For example, the Fig. 8(a) shows two signals, a square signal generated by oscillator 1 (red signal) and a triangular signal generated by oscillator 2 (green signal), both with a frequency equal

to 440 Hz, and the resulting signal from mixing these two signals (blue signal). Figure 8(b) shows the spectrum of the mixed-signal, which contains, in addition to the fundamental component, the odd harmonics, from the 3rd to the 15th harmonic. According to the Fourier transform superposition property, these harmonics come from the addition of the harmonics of each signal, and, in this case, the odd harmonics from the 1st to the 5th are highlighted, while the others (odd from 7th to 15th) are added with less intensity by the spectrum of the square signal.



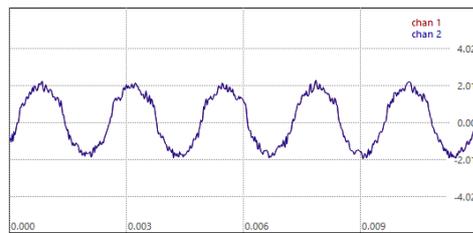
(a) Time-signal.



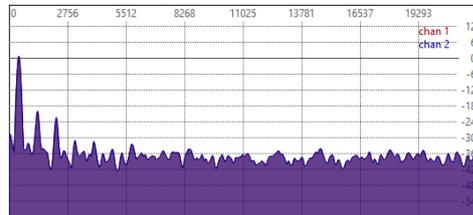
(b) Spectrum.

Fig. 8: Test signals selected for source block: (a) signal in time of each oscillator (square wave (red) and triangle wave (green)), both with 440 Hz, and mixed signal (blue); (b) spectrum of the mixed signal.

Noise generator: The signals from each oscillator (Fig. 8(a)) were mixed, and the resulting signal was altered with the addition of noise, which is shown in the Fig. 9(a), and its respective Fourier spectrum is shown in Fig. 9(b). This spectrum has components at all frequencies due to white noise, but the odd components 1st to 5th are still highlighted for having greater intensity

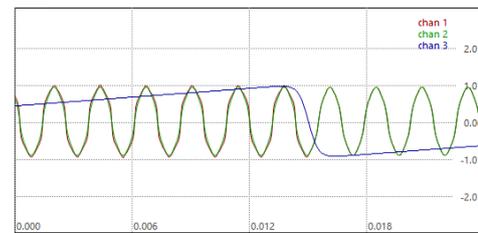


(a) Time-signal.

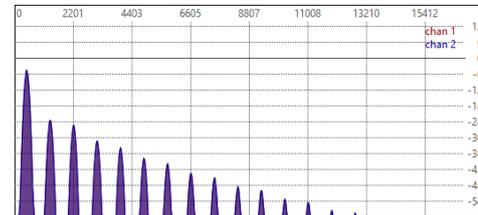


(b) Spectrum.

Fig. 9: Test signal chosen for the noise generator block: (a) Mixed signal (blue signal from Fig. 8(a)) and with added noise; (b) mixed signal spectrum.



(a) Time-signal.



(b) Spectrum.

Fig. 10: Test signal chosen for the LFO block: (a) input (green signal, which is the same blue signal from Fig. 8(a)), internal (blue) and output (red) signals from the LFO block; (b) LFO block output signal spectrum.

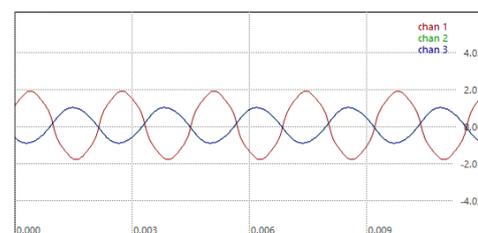
than the harmonics of white noise.

LFO block: For the mixed signal and without noise (blue signal from Fig. 8(a)) an LFO was applied, with a crescent sawtooth type signal and frequency equal to 20 Hz. The result is shown in Fig. 10(a). The green signal is the input of the LFO block. The application of the LFO (blue signal) causes the peaks of the output signal (red signal) to increase according to the amplitude of the LFO signal, so for positive amplitudes, the peaks are sharper and for negative ones more rounded.

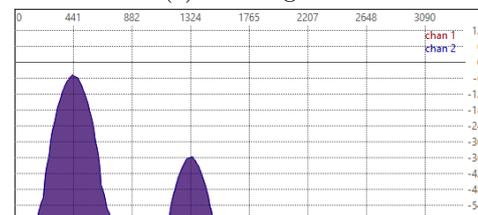
In Fig. 10(b) we can see the spectrum of the green signal. Comparing with the input signal spectrum (Fig. 8(b)), in this new spectrum, the number of harmonics increased and the intensity of the harmonics gradually decreased according to the frequency of the LFO internal signal. The sound obtained corresponds to the increase in the “brightness” of the timbre as a function of time, that is, the timbre goes from an opaque sound to a brighter one, which generates dynamics and movement sensation.

Low-pass filter: Responsible for the final treatment of the generated signal, the filter is the basis of the concept of subtractive synthesis because it precisely subtracts frequency components from the signal. To demonstrate its effect, a filter without resonance was set to a cutoff frequency of 440 Hz and applied to the mixed-signal (blue signal from Fig. 8(a)). The resulting signal and its spectrum are shown in Fig. 11(a) and 11(b), respectively. It is observed that in the time domain the amplitude is attenuated, the phase increases almost 180 degrees, the output signal approaches a sinusoidal signal and, in the frequency domain, most harmonics are fully attenuated, leaving only the fundamental component and the third harmonic.

3 - Timbral diversity: Due to the huge range of possible parameter combinations offered by the audio synthesis module, it is impossible to count the total num-



(a) Time-signal.



(b) Spectrum.

Fig. 11: Test signal selected for the low-pass filter: (a) input (red signal, which is the same mixed signal from Fig. 8(a)) and output (blue) signals; (b) spectrum of the output signal.

ber of timbres, since the smallest variation in a parameter causes a change in the acoustic characteristics of the timbre, which makes it be evaluated as a new one. However, to check the timbral diversity of the prototype, we present three examples, with the following settings:

- **Basic:** This timbre consists of a mixture of two sine signals of equal frequency (440 Hz) and amplitude (1); no added noise, neither LFO nor filtered.
- **Ghost:** This timbre consists of mixing two sine waves with equal amplitude (1), but with different frequencies, 440 Hz and 200 Hz. Likewise, no added noise, neither LFO nor filtered.

- **Helicopter:** This timbre was generated by mixing two random waves of equal amplitude (1), but with different frequencies, 440 Hz and 200 Hz. Also added white noise (5%), an LFO signal with 10 Hz, and a low-pass filter with a cutoff frequency equal to 2 kHz.

Figure 12(a) shows the signal of the “Basic” timbre and Fig. 12(b) its spectrum. This signal has a sinusoidal shape with a frequency equal to 440 Hz, but with an amplitude twice the amplitude of the original signal, and its frequency component is concentrated at 441 Hz.

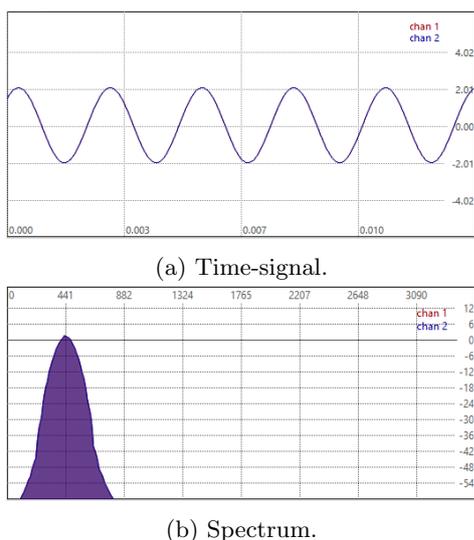


Fig. 12: Graphic representation of the “Basic” timbre.

Figure 13(a) shows the signal of the “Ghost” timbre and Fig. 13(b) its spectrum. This signal has an undefined shape in time, with two peaks, one larger than the other, and an amplitude twice the amplitude of the oscillator signals. Regarding frequency, the components are concentrated at 220 Hz and 440 Hz. From the subjective point, its sound is similar to two instruments being played at the same time, but with opposite registers, high and low.

For the last timbre, Fig. 14(a) shows the generated signal, whose format is undefined, and Fig. 14(b) its spectrum. It is possible to observe small distortions on the signal, resulting from the addition of white noise. Regarding the frequency domain, the spectrum contains more components and is in a wide frequency range, but with decreasing intensity, possibly due to the joint effect of the LFO and the low-pass filter.

V. DISCUSSION

As seen in the state-of-the-art (Section II-B.), the conventional theremin and most new prototypes, the frequency antenna is located vertically for the performer’s right hand, and orthogonally for the amplitude antenna. Due to this arrangement, the interpretation of the instrument requires several movements of the fingers, and some joint movements of the shoulder and elbow, which

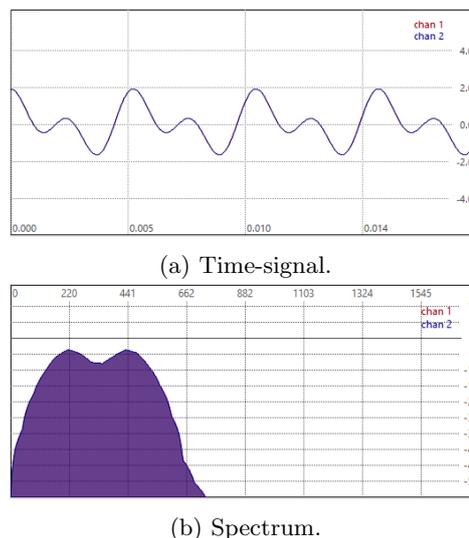


Fig. 13: Graphic representation of the “Ghost” timbre.

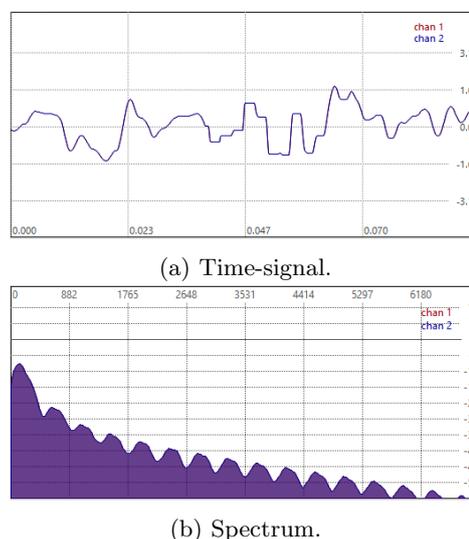


Fig. 14: Graphic representation of the “Helicopter” timbre.

makes it difficult to use by people with amputated fingers. In the vertical disposition, only few arm movements are necessary, mainly the movements of flexion and extension of the elbow.

Furthermore, the linearization of the distance-frequency relation makes it even easier to interpret the instrument, because the equidistant distribution of notes could help estimate the elbow angle needed to hit a particular note. As demonstrated in the experiments, the inherent error found in linearization is not significant, so it should not be considered a hindrance when interpreting the instrument.

On the other hand, the wide variety of tones is both beneficial and counterproductive as it could saturate the user’s choice, so incorporating a parameter recording mode and adding a list of predefined timbres would be of great help to the interpreter.

VI. CONCLUSIONS AND FUTURE WORK

The development of a low-cost digital theremin was developed and presented. Based on the results obtained, the prototype revealed positive results about the linearization of the frequency-distance relationship, since the divergence found is not statistically significant according to a non-parametric statistical test. Also, the developed audio synthesis module proved to be capable of producing different timbres, through the adjustment of parameters, and the development of the software interface contributes to an intuitive use of the instrument, in addition to helping the user to understand the elements of sound synthesis.

However, some limitations of this studio were found, such as the lack of development of a more specialized analysis to assess and determine the advantages of the vertical arrangement of the frequency antenna and the linearization of its distance-frequency relationship in the accessibility of people with amputation of fingers, as well as the expansion of this objective to the software interface through the modification and selection of timbres by voice commands.

As future works, we propose to incorporate the synthesis module into the instrument, which would allow it to be used without a PC, adapt virtual frets with a tempered and microtonal tuning selector, and add a guided learning mode to the interface, in which the student can train the ear through listening, repetition, and correction of musical excerpts (isolated notes, intervals, and small melodies) at different levels of difficulty. Furthermore, it is proposed to adjust preset intervals to the frequency sensor to play known scales and arpeggios and to include a third sensor to map the distance with chords.

VII. ACKNOWLEDGMENT

The authors would like to thank the UTFPR's research and post-graduate directorate at the Toledo campus (DIRPPG-TD) for the financial support.

REFERENCES

- [1] G. Cunha and M. Martins, "Tecnologia, produção & educação musical. descompassos e desafinos," in *Proc. Congresso da Rede Iberoamericana de Informática Educativa (RIBIE)*, 1998.
- [2] F. Iazzetta, "Reveno o papel do instrumento na música eletroacústica," in *Proc. Encontro de Música Eletroacústica*, 1997.
- [3] P. Nikitin, "Leon Theremin (Lev Termen)," *IEEE Antennas and Propagation Magazine*, vol. 54, no. 5, p. 252–257, 2012.
- [4] A. Glinsky, *Theremin: Music and Espionage*. University of Illinois Press, 2000.
- [5] C. Martín, J. Martínez, A. Ricchiuti, H. González, and C. Franco, "Study of the interference affecting the performance of the theremin," *International Journal of Antennas and Propagation*, no. 348151, pp. 1–9, 2012.
- [6] O. Bélanger, "Ajax sound studio. Pyo 1.0.4 documentation. [computer software]," 2021. [Online].

Available: <http://ajaxsoundstudio.com/pyodoc/index.html>

- [7] A. Gomes, D. Albuquerque, G. Campos, and J. Vieira, "TheremUS: The Ultrasonic Theremin," in *Proc. Audio Engineering Society Convention (AES)*, 2009, pp. 1–7.
- [8] B. Hanindhito, H. Hogantara, A. Rahmah, N. Ahmadi, and T. Adiono, "Ultrasonic sensor based contactless theremin using pipeline CORDIC as tone-generator," in *Proc. International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, 2015, pp. 268–273.
- [9] X. Xiao, G. Locqueville, C. d'Alessandro, and B. Doval, "T-Voks: the singing and speaking theremin," in *Proc. International Conference on New Interfaces for Musical Expression (NIME)*, 2019, p. 110–115.
- [10] K. Skeldon, L. Reid, V. McNally, B. Dougan, and C. Fulton, "Physics of the theremin," *American Journal of Physics*, vol. 66, no. 11, pp. 945–955, 2012.
- [11] M. Kasey, "Strange vibrations. the evolution of the theremin," Master's thesis, California State University, Monterey Bay, 2019.
- [12] C. Geiger, H. Reckter, D. Paschke, and F. Schulz, "Poster: Evolution of a theremin-based 3D-interface for music synthesis," in *Proc. IEEE Symposium on 3D User Interfaces (3DUI)*, 2008, pp. 163–164.
- [13] C. Geiger, H. Reckter, D. Paschke, F. Schulz, and C. Poepel, "Towards participatory design and evaluation of theremin-based musical interfaces," in *Proc. International Conference on New Interfaces for Musical Expression (NIME)*, 2008, pp. 303–306.
- [14] T. Liu, S. Chang, and C. Hsiao, "A modified quad-theremin for interactive computer music control," in *Proc. International Conference on Multimedia Technology (ICMT)*, 2011, pp. 6179–6182.
- [15] B. Wöldecke, D. Marinos, P. Pogscheba, C. Geiger, J. Herder, and T. Schwirten, "radarTHEREMIN – creating musical expressions in a virtual studio environment," in *Proc. IEEE International Symposium on Virtual Reality Innovation (ISVRI)*, 2011, pp. 345–346.
- [16] P. Nikitin, A. Parks, and J. Smith, "RFID-Vox: a tribute to Leon Theremin," in *Wirelessly Powered Sensor Networks and Computational RFID*, J. Smith, Ed. Springer, 2013, pp. 259–268.
- [17] R. Sharma, R. Ullagaddimath, A. Baran, A. Halder, and V. Hegde, "Optical theremin based true random number generation (TRNG) system," in *Proc. International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2015, pp. 571–575.
- [18] J. Cervera, "Implementación de un theremin utilizando *Leap Motion*," Bachelor Thesis. Universitat Politècnica de València, 2016.
- [19] J. García, D. Salazar, G. Celis, and J. Vélez, "Design and experimental prototype testing of a hybrid digital theremin system," in *Proc. International*

- Conference on Consumer Electronics (ICCE)*, 2018, pp. 1–4.
- [20] C. Quimelli, “Theremin: música e eletrônica no ensino da arte-ciência,” Master’s thesis, Universidade Tecnológica Federal Do Paraná (UTFPR), Brasil, 2019.
- [21] D. Johnson, D. Damian, and G. Tzanetakis, “Evaluating the effectiveness of mixed reality music instrument learning with the theremin,” *Virtual Reality*, vol. 24, no. 2, p. 303–317, 2019.
- [22] B. Campbell, “Inspiring future engineers: Teaching basic electronics to create theremin-based musical instruments,” in *Proc. Annual Conference & Exposition of American Society For Engineering Education (ASEE)*, 2017, pp. 1–20.
- [23] D. Čoko, L. Rodić, T. Perković, and P. Šolić, “Geometry from thin air: Theremin as a playful learning device,” in *Proc. International Conference on Telecommunications (ConTEL)*, 2021, pp. 89–96.
- [24] T. Mizumoto, H. Tsujino, T. Takahashi, T. Ogata, and H. Okuno, “Thereminist robot: Development of a robot theremin player with feedforward and feedback arm control based on a theremin’s pitch model,” in *Proc. International Conference on Intelligent Robots and Systems (IROS)*, 2009, pp. 2297–2302.
- [25] M. Russ, *Sound Synthesis and Sampling*. Routledge, 2008.
- [26] L. Caligiuri and D. Giordano, “The use of time-scale analysis in forensic acoustical investigations,” *WSEAS Transactions on Acoustics and Music*, vol. 7, pp. 6–11, 2020.
- [27] C. Riverbank, “PyQt5 version 12.8.1 [computer software],” 2020. [Online]. Available: <https://pypi.org/project/PyQt5>
- [28] Q. Company, “QT designer [computer software],” 2018. [Online]. Available: <https://doc.qt.io/qt-5/qt designer-manual.html>
- [29] J. Beauchamp, *Analysis, Synthesis, and Perception of Musical Sounds. The Sound of Music*. Springer, 2007, ch. Analysis and Synthesis of Musical Instrument Sounds.
- [30] J. Meyer, *Acoustics and the Performance of Music*. Springer, 2009.
- [31] L. Chao, *Estadística para las ciencias administrativas*. Mc. Graw Hill, 1975.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0

https://creativecommons.org/licenses/by/4.0/deed.en_US