Psychophysics of musical elements in the discrete-time representation of sound

Renato Fabbri,[a] Vilson Vieira da Silva Jr.,[b] Débora Corrêa,[c] Luciano da Fontoura Costa,[d] and Osvaldo N. Oliveira Jr.[e]

Instituto de Física de São Carlos, Universidade de São Paulo (IFSC/USP)

(Dated: 1 October 2013)

Notes, ornaments and intervals are examples of basic elements of music, their representation as discrete-time digital audio signal plays a central role in software for music creation and design. Nevertheless, there is no systematic relation, in analytical terms, of these musical elements to the sound samples. Such a compendium is important since it enables scientific experiments in precise and trustful ways. In this paper, a comprehensive description of the musical elements within a unified approach is presented. Musical elements, like pitch, duration and timbre are expressed by equations on sample level. This quantitatively relates characteristics of the discrete-time signal to musical qualities. Internal variations, e.g., tremolos, vibratos and spectral fluctuations, are also considered as means to achieve variation within a note. Moreover, the generation of musical structures such as rhythmic meter, pitch intervals and cycles, are reached canonically with notes. The availability of these resources in scripts (MASSA - Music and Audio in Sequences and Samples) is also provided in public domain. Authors observe that the implementation of sample-domain analytical results as open source can encourage concise research. As further illustrated in the paper, MASSA has already been employed by users for diverse purposes, including acoustics experimentation, art and education. The efficacy of these physical descriptions of basic musical elements was confirmed by the synthesis of small musical pieces. As shown, it is possible to synthesize whole albums through collage of scripts and parametrization.

PACS numbers: *43.66.-x,43.66.+y,05.10.-a
Keywords: psychophysics, acoustics, statistics, signal processing, digital audio, music

I. INTRODUCTION

Music is commonly defined as the art made by sounds and silences. Sound is the longitudinal wave of mechanical pressure. The frequency bandwidth between $20Hz$ and $20kHz$ is appreciated by the human hearing system with boundaries dependent on the person, climate conditions and sonic characteristics themselves[?] . Considered the speed of sound as $\approx 343.2m/s$, these limits imply wavelengths of $\frac{343.2}{20} = 17.16\,m$ and $\frac{343.2}{20000} = 17.16\,mm$.

Human perception of sound involves captivation by bones, stomach, ears, transfer functions of head and torso, and processing by the nervous system. Besides that, the ear is a dedicated organ for the appreciation of these waves. Its mechanism decomposes sound into its sinusoidal spectrum and delivers to the nervous system. The sinusoidal components are crucial to musical phenomena, as one can perceive in the composition of sounds of musical interest (such as harmonic sounds and noises, seen in sections II and III), and sonic structures of musical interest (such as tunings and scales, in section IV).

The representation of sound is commonly called audio (although these terms are often used without distinction). Audio expresses waves prominent from the capture of microphones or from direct synthesis, although these sources entangle as captured sounds are processed to generate new sonorities. Digital audio specified by protocols that eases file transfering and storage often imply a quality loss. Standard representation of digital audio, on the other hand, assures perfect reconstruction of the analog wave, within any convenient precision. This paradigm consists of equally spaced samples, by λ_s durations, each specified by a fixed number of bits. This is the Pulse Code Modulation (PCM) representation of sound. A sound in PCM audio is characterized by a sampling frequency $f_s = \frac{1}{\lambda_s}$ (also called the sampling rate), which is the number of samples used for representing a second of sound; and by a bit depth, which is the number of bits used for representing the amplitude of each sample. Figure 1 shows 25 samples of a PCM audio with 4 bits each. The fixed $2^4 = 16$ values for the amplitude of each sample, with the regular spacing λ_s, yields the quantization error or noise. This noise diminishes as bit depth increases.

The Nyquist theorem states that the sampling frequency is twice the maximum frequency of the represented signal. Thus, for general musical purposes, it is necessary to have samples in a rate at least twice the highest frequency heard by humans, that is, $f_s \geq$

[a] `http://www.ifsc.usp.br/~fabbri`; Electronic mail: renato.fabbri@gmail.com
[b] `http://automata.cc/`; Electronic mail: vilson@void.cc; Also at IFSC-USP
[c]; Electronic mail: debcris.cor@gmail.com; Instituto de Ciências Matemáticas e de Computação ICMC-USP
[d] `http://cyvision.ifsc.usp.br/~luciano/`; Electronic mail: ldfcosta@gmail.com; Also at IFSC-USP
[e] `www.polimeros.ifsc.usp.br/professors/professor.php?id=4`; Electronic mail: chu@ifsc.usp.br; Also at IFSC-USP

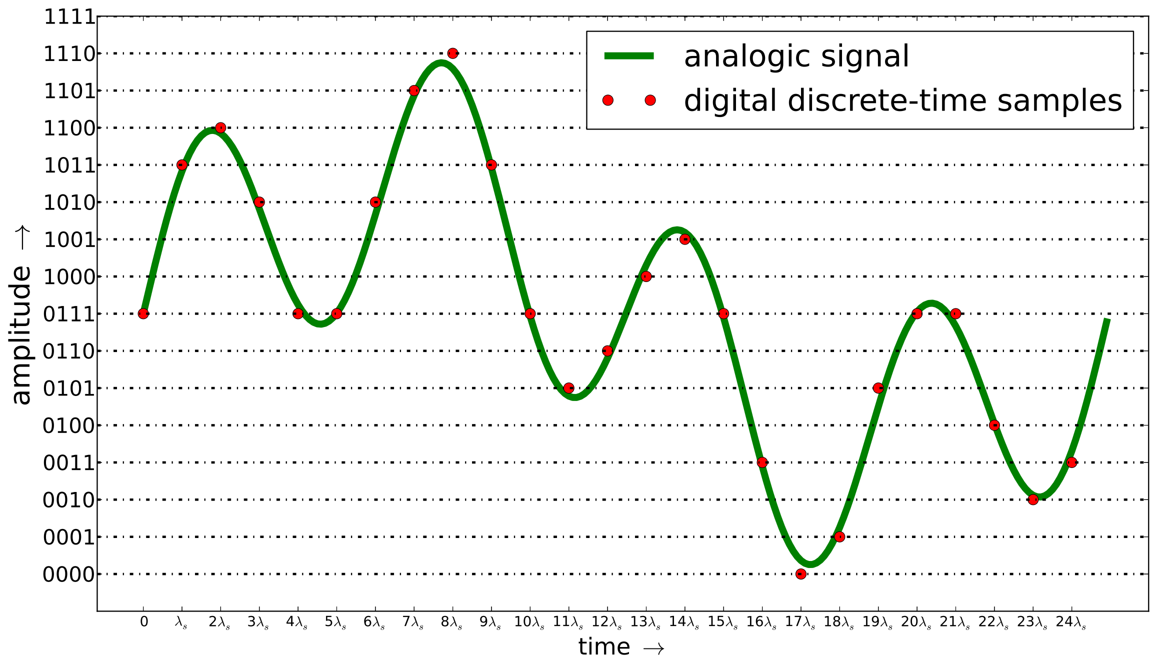

FIG. 1. Pulse Code Modulation (PCM) audio: an analogical signal is represented by 25 samples with 4 bits each.

$2 \times 20kHz = 40kHz$. This is the fundamental reason for the adoption of sampling frequencies like $44.1kHz$ and $48kHz$, standards in Compact Disks (CD) and broadcast systems (radio and television), respectively.

Within this framework, musical notes can be characterized. The note still stands paradigmatic as the 'fundamental unit' of musical structures and, in practice, it can unfold into sounds that uphold other approaches[3]. Notes are also convenient for another reason: the average listener – and considerable part of the specialists – presupposes rhythmic and pitch organization (made explicit in section **??**) as fundamental musical properties, and these are developed in traditional musical theory in terms of notes.

A. Contributions and paper organization

This article aims at representing musical structures and artifices by their discrete-time sonic characteristics. Results include mathematical relations, usually in terms of samples, and their direct computer program implementations. Despite the general interests involved, there are few books and computer implementations that tackle the subject. These mainly focus on computer implementations and ways to mimic traditional instruments, with scattered mathematical formalisms. A compilation of the works and their contributions is in the bibliography[?] . To the best of the author's knowledge, there is a lack of articles on the topic. Moreover, although current computer implementations use the analytical descriptions presented in this study in a implicit manner, it seems that there has been no concise and mathematical description of the processes implemented.

In order to address this concise description of musical elements and structures, in terms of the digitilized sound, the objectives of this paper are:

1. Present a concise set of relations among musical basic elements and sequences of PCM audio samples.

2. Introduce a framework of sound synthesis with control at sample level and with potential uses in psycho-acoustical experiments and high-fidelity synthesis.

3. Provide the accessibility of the developed framework. The analytic relations presented in this article are implemented as public domain scripts, i.e. small computer programs using accessible technologies for better distribution and validation. This constitute the MASSA toolbox, available in an open Git repository[?] . These scripts are written in Python and make use of external libraries, such as Numpy that performs calls to Fortran routines for better computational efficiency. Part of the scripts have been transcribed to JavaScript and native Python with readiness, what favors their use in Web browsers such as Firefox and Chromium[? ? ? ?] . Furthermore, these are all open technologies, that is, published using licenses that

allow copy, distribution and use of any part for research and derivatives. This way, the work presented here is embedded in recommended practices for availability and eases co-authorship processes[? ?] .

4. To provide a didactic presentation of the content to favor its apprehension and usage. It is worthwhile to mention that this subject comprises diverse topics on signal processing, music, psycho-acoustics and programming.

The remaining parts of this work are organized as follows: section II characterizes the basic musical note; internals of the musical note is further developed in section III; section IV tackles the organization of musical notes in higher levels of musical structure[1? ? ? ?] . As these descriptions embody topics such as psycho-acoustics, cultural traditions, formalisms and protocols, the text points to external complements as needed[? ? ?] .

The next section is a minimum text in which musical elements are presented side-by-side with the discrete-time samples they result. In order to account for validation and sharing, implementations on computer code of each one of these relations are gathered in the MASSA toolbox together with little musical montage resulting from them.

II. CHARACTERIZATION OF THE DISCRETE-TIME MUSICAL NOTE

In diverse artistic and theoretical contexts, music is conceived as constituted by fundamental units called notes, "atoms" that constitute music itself[1? ,2]. Nowadays, these notes are understood as central elements of certain musical paradigms. In a cognitive perspective, the notes are seen as discretized elements that easy and enrich the flow of information through music[? ?] . Canonically, the principal properties of a musical note are duration, volume, pitch and timbre[?] . These qualities can be managed quantitatively, dictated by the evenly time spaced sonic samples.

All the relations described in this section are implemented at the file *eqs2.1.py* of the MASSA toolbox. Musical pieces *5 sonic portraits* and *reduced-fi* are also available online in order to corroborate the concepts.

A. Duration

The sample frequency f_s is defined as the number of samples in each second of the discrete-time signal. Let $T_i = \{t_i\}$ be an ordered set of real samples separated by $\delta_s = 1/f_s$ seconds. A musical note of duration Δ seconds is presented as a sequence T_i^Δ of $\lfloor \Delta.f_s \rfloor$ samples. That is, the integer part of the multiplication is considered, and an error of at most δ_s missing seconds is admitted, which is usually fine for musical purposes. As an example, if $f_s = 44.1kHz \quad \Rightarrow \quad \lambda_s = \frac{1}{44100} \approx 23$ microseconds. Thus:

$$T_i^\Delta = \{t_i\}_{i=0}^{\lfloor \Delta.f_a \rfloor - 1} \tag{1}$$

With $\Lambda = \lfloor \Delta.f_a \rfloor$, the number of samples in the sequence, the more condensed notation is $T_i^\Delta = \{t_i\}_0^{\Lambda-1}$.

B. Volume

The sensation of sound volume depends on reverberation and harmonic distribution, among other characteristics described on section **??**. One can get volume variations through the potency of the wave[?] :

$$pot(T_i) = \frac{\sum_{i=0}^{\Lambda-1} t_i^2}{\Lambda} \tag{2}$$

The final volume is dependent on the speakers amplification of the signal. Thus, what matters is the relative potency of a note in relation to the other ones around it, or the potency of a music section in relation to the rest. Differences in volume are measured in decibels, calculated directly from the amplitudes through energy or potency:

$$V_{dB} = 10log_{10}\frac{pot(T_i^{'})}{pot(T_i)} \tag{3}$$

The quantity V_{dB} has the decibel unit (dB). For each $10dB$ it is associated a "doubled volume". A handy reference is $10dB$ for each step in the intensity scale: *pianissimo*, *piano*, *mezzoforte*, *forte* e *fortissimo*. Other useful references are dB values related to double amplitude or potency:

$$\begin{aligned} t_i^{'} = 2.t_i \Rightarrow pot(T_i^{'}) = 4.pot(T_i) \Rightarrow \\ \Rightarrow V_{dB}^{'} = 10log_{10}4 \approx 6dB \end{aligned} \tag{4}$$

$$\begin{aligned} pot(T_i^{'}) = 2pot(T_i) \Rightarrow \\ \Rightarrow V_{dB}^{'} = 10log_{10}2 \approx 3dB \end{aligned} \tag{5}$$

and the amplitude gain for a sequence whose volume has been doubled ($10dB$):

$$\begin{aligned} 10log_{10}\frac{pot(T_i^{'})}{pot(T_i)} = 10 \quad \Rightarrow \\ \Rightarrow \quad \sum_{i=0}^{\lfloor \Delta.f_a \rfloor -1} t_i^{'2} = 10\sum_{i=0}^{\Lambda-1} t_i^2 = \sum_{i=0}^{\Lambda-1}(\sqrt{10}.t_i)^2 \\ \therefore \quad t_i^{'} = \sqrt{10}t_i \quad \Rightarrow \quad t_i^{'} \approx 3.16t_i \end{aligned} \tag{6}$$

As shown, it is necessary an amplitude increase by a bit more than a factor of 3 for a doubled volume. These values are guides for increasing or decreasing the absolute values on the sample sequences with musical purposes. The conversion from decibels to amplitude gain (or attenuation) is straightforward:

$$A = 10^{\frac{V_{dB}}{20}} \tag{7}$$

where A is the multiplicative factor that relates the amplitudes before and after the amplification.

C. Pitch

As exposed in the previous subsections, to a note corresponds a sequence T_i in which duration and volume are directly related to the size of the sequence and the amplitude of its samples, respectively. The pitch is specified by the fundamental frequency f_0 whose cycle has duration $\delta_{f_0} = 1/f_0$. This duration, multiplied by the sampling frequency f_s results on the number of samples per cycle: $\lambda_{f_0} = f_s.\delta_{f_0} = f_a/f_0$.

For didactic reasons, be f_0 such that it divides f_s and λ_{f_0} results integer. If T_i^f is a sonic sequence of fundamental frequency f, then:

$$T_i^f = \left\{t_i^f\right\} = \left\{t_{i+\lambda_f}^f\right\} = \left\{t_{i+\frac{f_a}{f}}^f\right\} \tag{8}$$

In the next section, frequencies f that does not divide f_s will be considered. This restriction does not imply in lost of generality of this section's content.

D. Timbre

In a sound with harmonic spectrum, the (wave) period corresponds to the duration of the cycle, given by the inverse of the fundamental frequency. The trajectory of the wave inside the period - called the waveform - defines a harmonic spectrum and, thus, a timbre[4]. Musically, it matters that sonic spectra with minimum differences can result in timbres with crucial expressive differences and, consequently, different timbres can be produced by using different spectra[?] .

The simplest case is the spectrum that consists of only of the fundamental frequency f. This defines the sinusoid: a frequency in pure oscillatory movement called 'simple harmonic movement'. Be S_i^f a sequence whose samples s_i^f describes a sinusoid of frequency f:

$$S_i^f = \{s_i^f\} = \left\{\sin\left(2\pi\frac{i}{\lambda_f}\right)\right\} = \left\{\sin\left(2\pi f\frac{i}{f_s}\right)\right\} \tag{9}$$

where $\lambda_f = \frac{f_s}{f} = \frac{\delta_f}{\lambda_s}$ is the number of samples in the period.

In a similar fashion, other waveforms are applied in music for their spectral qualities and simplicity. While the sinusoid is an isolated point in the spectrum, these waves present a succession of harmonic components. These waveforms are specified in equations 9, 10, 11 and 12 and illustrated in Figure 2. These artificial waveforms are traditionally used in music for synthesis and oscillatory control of variables. They are also useful outside musical contexts[?] .

The sawtooth presents all components of the harmonic series with decreasing energy of $-6dB/octave$. The sequence of temporal samples can be described as:

$$D_i^f = \left\{d_i^f\right\} = \left\{2\frac{i\,\%\lambda_f}{\lambda_f} - 1\right\} \tag{10}$$

The triangular waveform present only odd harmonics falling with $-12dB/octave$:

$$T_i^f = \left\{t_i^f\right\} = \left\{1 - \left|2 - 4\frac{i\,\%\lambda_f}{\lambda_f}\right|\right\} \tag{11}$$

The square wave preservers only odd harmonics falling at $-6dB/octave$:

$$Q_i^f = \left\{q_i^f\right\} = \begin{cases} 1 & \text{for} \quad (i\,\%\lambda_f) < \lambda_f/2 \\ -1 & \text{otherwise} \end{cases} \tag{12}$$

The square wave can be used in a subtractive synthesis with the purposes of mimicking a clarinet. This instrument has only the odd harmonic components and the square wave is convenient with its abundant energy in high frequencies. The sawtooth is a common starting point for a subtractive synthesis, because it has both odd and even harmonics with high energy. In general, these waveforms are appreciated as excessively rich in sharp harmonics, and attenuator filtering on treble and middle parts of the spectrum is specially useful for reaching a more natural and pleasant sound. The relatively attenuated harmonics of the triangle wave makes it the more functional - among the listed ones - to be used in the synthesis of musical notes without any treatment. The sinusoid is often a nice choice, but a problematic one. While pleasant if not loud in a very high pitch (above 500Hz, it requires careful dosage), the pitch is not accuratly detected, specially in low frequencies. It requires a big amplitude to give sinusoid volume, if compared to other waveforms. Both particularities are seen as a consequence of the non-existence of pure sinusoidal sounds in nature **??**.

Figure 2 presents the waveforms described in equations **??**, **??**, 11 and **??** for $\lambda_f = 100$ (period of 100 samples). If $f_s = 44.1kHz$, the PCM standard in Compact Disks, the wave has fundamental frequency $f = \frac{f_a}{\lambda_f} = \frac{44100}{100} = 441\ Herz$, around A4, just above the central "C", whatever the waveform is.

The spectrum of each basic waveform is in Figure 3. The isolated and exactly harmonic components of the

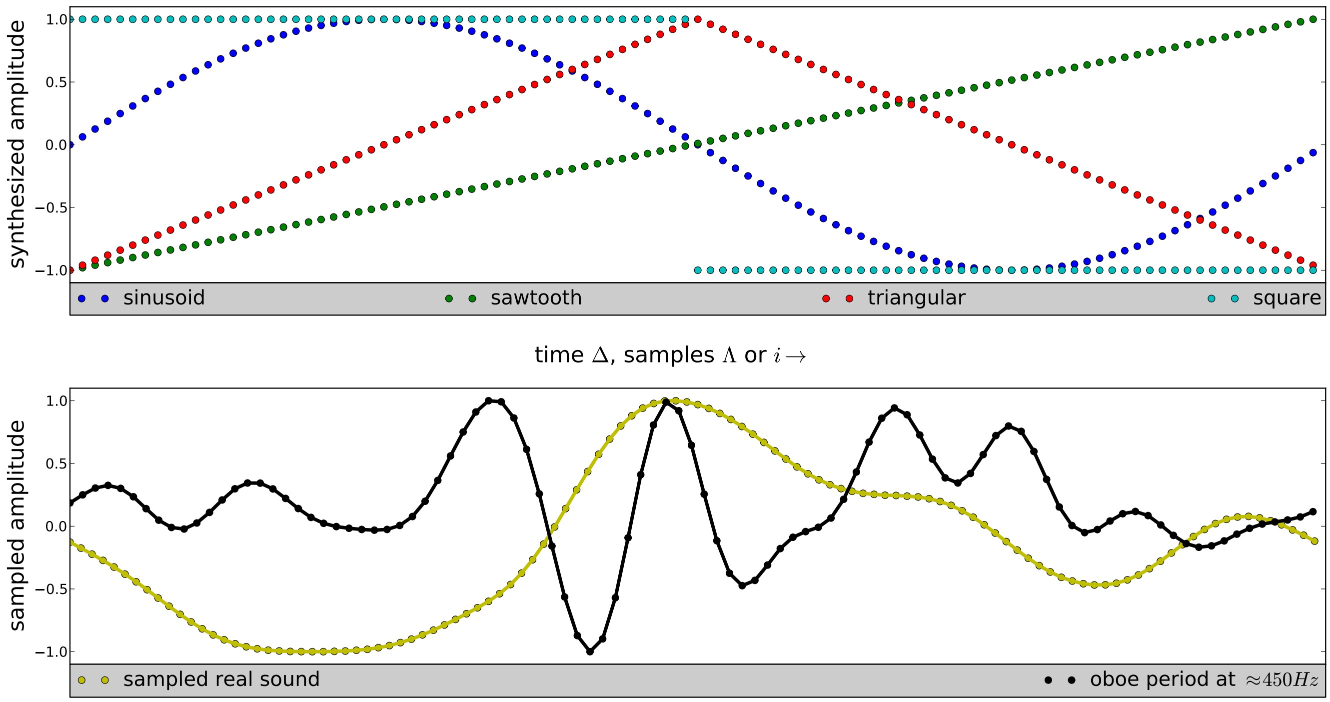

FIG. 2. Basic musical waveforms: (a) the synthetic waveforms; (b) the real waveforms.

spectrum is a consequence of the fixed period usage. The sinusoid consists of an one and only node in the spectrum, pure frequency. The figure exibit the spectra described: the sawtooth is the only waveform with a complete harmonic series (odd and even components); triangular and square waves has the same components (odd harmonics), decaying at $-12dB/octave$ and $-6dB/octave$, respectively.

The harmonic spectrum is composed of frequencies f_n that are multiples of the fundamental frequency: $f_n = (n+1)f_0$. As the human linear perception of pitch follows a geometric progression of frequencies, the spectrum has notes different from the fundamental frequency (see equation 84). Additionally, the number of harmonics will be limited by the Nyquist frequency $f_s/2$.

From a musical perspective, it is critical to internalize that energy in a component of frequency f_n means an oscillation in the constitution of the sound, purely harmonic and in that frequency f_n. This energy, specifically concentrated on the frequency f_n, is separated by the ear for further cognitive processes (this separation is done in many species with mechanisms similar to the human cochlea[?]).

The sinusoidal components are usually the main responsibles for timbre qualities. If they are not presented in harmonic proportions (small number relations), the sound is perceived as noisy or dissonant, in opposition to sonorities with an unequivocally established fundamental. Furthermore, the notion of absolute pitch relies on the similarity of the spectrum to the harmonic series[?] .

In the case of a fixed length period and waveform, the spectrum is perfectly harmonic and static and each waveform is compound of specific proportions of harmonic components. High curvatures are sign of high harmonics in the wave. Figure 2 depicts a wave, labeled as "sampled real sound", with a period of $\Lambda_f = 114$ samples, extracted from a relatively well behaved recorded sound. The oboe wave was also sampled in $44.1kHz$. The chosen period for sampling was relatively short, with 98 samples, and corresponds to the frequency $\frac{44100}{98} = 450Hz$, which corresponds to a slightly out-of-tune A4 pitch. One can notice from the curvatures: the oboe's rich spectrum in high frequencies and the lower spectrum of the real sound.

The sequence $R_i = \{r_i\}_0^{\lambda_f - 1}$ of samples in the real sound of Figure 2 can be taken as basis for a sound T_i^f in the following way:

$$T_i^f = \{t_i^f\} = \left\{ r_{(i\,\%\,\lambda_f)} \right\} \tag{13}$$

The resulting sound has the momentary spectrum of the original waveform. As a consequence of its repetition in an identical form, the spectrum is perfectly harmonic, without noise and with variations typical of the natural phenomenon. This can be observed in Figure 4, that shows the spectrum of the original oboe note and a note with same duration and whose samples consists of the repetition of cycle of Figure 2. Summing up, the natural spectrum exhibits variations in the frequencies of

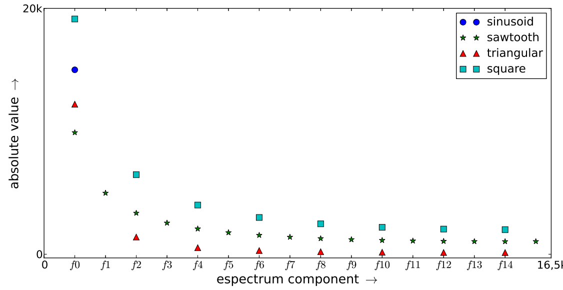

FIG. 3. Spectrum of basic artificial musical waveforms.

the harmonics, in their intensities and some noise, while the note made from the sampled period has a perfectly harmonic spectrum.

E. Spectrum at sampled sound

The presence and behavior of these sinusoidal components in the discretized sound have some particularities. Considering a signal T_i and its corresponding Fourier decomposition $\mathcal{F}\langle T_i \rangle = C_i = \{c_i\}_0^{\Lambda-1}$, the recomposition is the sum of the frequency components as time samples[5]:

$$
\begin{aligned}
t_i =& \frac{1}{\Lambda} \sum_{k=0}^{\Lambda-1} c_k e^{j\frac{2\pi k}{\Lambda}i} \\
=& \frac{1}{\Lambda} \sum_{k=0}^{\Lambda-1} (a_k + j.b_k)\left[cos(w_k i) + j.sen(w_k i)\right]
\end{aligned}
\tag{14}
$$

where $c_k = a_k + j.b_k$ defines the amplitude and phase of each frequency: $w_k = \frac{2\pi}{\Lambda}k$ in radians or $f_k = w_k \frac{f_a}{2\pi} = \frac{f_a}{\Lambda}k$ in Hertz, taking into account the respective limits in π and in $\frac{f_a}{2}$ given by the Nyquist Theorem. As can be inferred, j is the complex number, with $j^2 = -1$.

For a sound signal, samples t_i are real and are given by the real part of equation 14:

$$
\begin{aligned}
t_i =& \frac{1}{\Lambda} \sum_{k=0}^{\Lambda-1} \left[a_k cos(w_k i) - b_k sen(w_k i)\right] \\
=& \frac{1}{\Lambda} \sum_{k=0}^{\Lambda-1} \sqrt{a_k^2 + b_k^2}\, cos\left[w_k i - tg^{-1}\left(\frac{b_k}{a_k}\right)\right]
\end{aligned}
\tag{15}
$$

Equation 15 shows how the imaginary term of c_k adds a phase to the real sinusoid: the terms b_k enables the phase sweep $\left[-\frac{\pi}{2}, +\frac{\pi}{2}\right]$ given by $tg^{-1}\left(\frac{b_k}{a_k}\right)$ which has this image. The terms a_k specifies the right or left side of the trigonometric circle, which completes the phase domain: $\left[-\frac{\pi}{2}, +\frac{\pi}{2}\right] \cup \left[\frac{\pi}{2}, \frac{3\pi}{2}\right] \equiv [2\pi]$.

Figure 5 shows two samples and its spectral components. In this case, the Fourier decomposition has one unique pair of coefficients $\{c_k = a_k - j.b_k\}_0^{\Lambda-1=1}$ relative to frequencies $\{f_k\}_0^1 = \left\{w_k \frac{f_a}{2\pi}\right\}_0^1 = \left\{k \frac{f_a}{\Lambda=2}\right\}_0^1 = \left\{0, \frac{f_a}{2} = f_{\max}\right\}$ with energies $e_k = \frac{(c_k)^2}{\Lambda=2}$. The role of amplitudes a_k is clearly observed with $\frac{a_0}{2}$, the fixed offset[6] and $\frac{a_1}{2}$, oscillation amplitude with frequency given by the relation $f_k = k\frac{f_a}{\Lambda=2}$. This case have special relevance. It is necessary at least 2 samples to represent an oscillation and it yields the Nyquist frequency $f_{\max} = \frac{f_a}{2}$, which is the maximum frequency in a sound sampled with f_a samples per second[7].

All fixed sequences T_i of only 3 samples also have just 1 frequency, since their first harmonic has $1,5$ samples and exceeds the bottom limit of 2 samples, i.e. the frequency of the harmonic would exceed the Nyquist frequency: $\frac{2.f_a}{3} > \frac{f_a}{2}$. The coefficients $\{c_k\}_0^{\Lambda-1=2}$ are present in 3 frequency components. One is relative to zero frequency (c_0), and the other two (c_1 and c_2) have the same role for the reconstruction of sinusoid with $f = f_a/3$.

Λ real samples t_i result in Λ complex coefficients $c_k = a_k + j.b_k$. The coefficients c_k are equivalent two by two, corresponding to the same frequencies and with same contribution to its reconstruction. They are complex conjugates: $a_{k1} = a_{k2}$ and $b_{k1} = -b_{k2}$ and, as a consequence, the modules are equal and phases have opposite signs. Remembering that $f_k = k\frac{f_a}{\Lambda},\ k \in \left\{0, ..., \left\lfloor \frac{\Lambda}{2} \right\rfloor\right\}$. When $k > \frac{\Lambda}{2}$, the frequency f_k is mirrored by $\frac{f_a}{2}$ in this way: $f_k = \frac{f_a}{2} - (f_k - \frac{f_a}{2}) = f_a - f_k = f_a - k\frac{f_a}{\Lambda} = (\Lambda - k)\frac{f_a}{\Lambda} \quad \Rightarrow \quad f_k \equiv f_{\Lambda-k}\ ,\ \forall\ k < \Lambda$.

The same can be observed with $w_k = f_k.\frac{2\pi}{f_a}$ and the periodicity 2π: it follows that $w_k = -w_{\Lambda-k}\ ,\ \forall\ k < \Lambda$. Given the cosine (an even function) and the inverse tangent (an odd function), the components in w_k and $w_{\Lambda-k}$ contribute with coefficients c_k and $c_{\Lambda-k}$ in the reconstruction of the real samples.

In other words, in a decomposition of Λ samples, the Λ frequency components $\{c_i\}_0^{\Lambda-1}$ are equivalents in pairs, except for f_0, and, when Λ is even, for $f_{\Lambda/2} = f_{\max} = \frac{f_a}{2}$. Both components are isolated, i.e. there is one and only component in frequency f_0 or $f_{\Lambda/2}$ (if even Λ). One can verify this assertion considering $k = 0$ and $k = \Lambda/2$. Uncoiled: $f_{\Lambda/2} = f_{(\Lambda-\Lambda/2)=\Lambda/2}$ and $f_0 = f_{(\Lambda-0)=\Lambda} = f_0$.

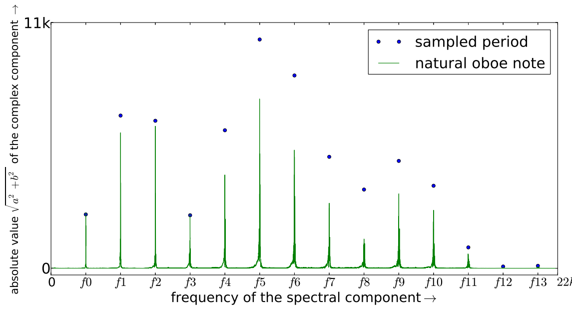

FIG. 4. Spectrum of the sonic waves of a natural oboe note and from a sampled period. The natural sound has fluctuations in the harmonics and in its noise, while the sampled period note has a perfectly harmonic spectrum.

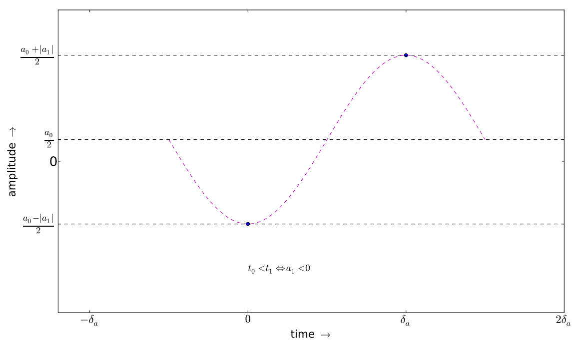

FIG. 5. Oscillation of 2 samples (maximum frequency for any f_a). The first coefficient reflects a detachment (*offset* or *bias*) and the second coefficient specifies the oscillation amplitude.

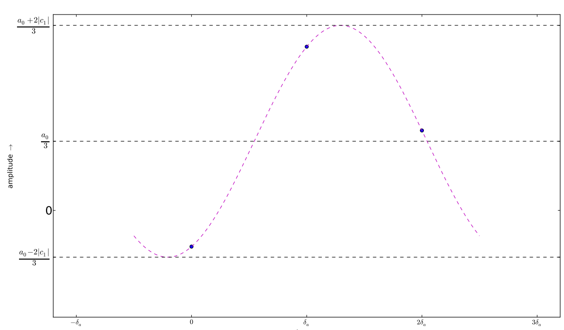

FIG. 6. Three fixed samples presents only one non-null frequency. $c_1 = c_2^*$ and $w_1 \equiv w_2$.

Furthermore, these two frequencies (zero and Nyquist frequency) does not have phase variation, being their coefficients strictly real. In this way, it is possible to conclude the number τ of equivalent coefficient pairs:

$$\tau = \frac{\Lambda - \Lambda\%2}{2} + \Lambda\%2 - 1 \tag{16}$$

This discussion makes clear the equivalences 17, 18 and 19:

$$f_k \equiv f_{\Lambda-k}\ ,\ \ w_k \equiv -w_{\Lambda-k}\quad ,\qquad \forall \quad 1 \leq k \leq \tau \tag{17}$$

$T_i \;\Rightarrow\; a_k = a_{\Lambda-k}$ and $b_k = -b_{\Lambda-k}$, and thus:

$$\sqrt{a_k^2 + b_k^2} = \sqrt{a_{\Lambda-k}^2 + b_{\Lambda-k}^2}\quad ,\qquad \forall \quad 1 \leq k \leq \tau \tag{18}$$

$$tg^{-1}\left(\frac{b_k}{a_k}\right) = -tg^{-1}\left(\frac{b_{\Lambda-k}}{a_{\Lambda-k}}\right)\quad ,\qquad \forall \quad 1 \leq k \leq \tau \tag{19}$$

with $k \in \mathbb{N}$.

To expose the general case for components combination in each sample t_i, one can gather relations in equation 15 for the real signal reconstruction, relations of modules 18 and phases equivalences 19, the number of paired coefficients 16, and equivalence of paired frequencies 17:

$$t_i = \frac{a_0}{\Lambda} + \frac{2}{\Lambda}\sum_{k=1}^{\tau}\sqrt{a_k^2+b_k^2}\ cos\left[w_k i - tg^{-1}\left(\frac{b_k}{a_k}\right)\right] + \frac{a_{\Lambda/2}}{\Lambda}.(1-\Lambda\%2) \tag{20}$$

with $a_{\Lambda/2} = 0$ if Λ odd.

With 4 samples it is possible to represent 1 or 2 frequencies in any proportions (i.e. with independence). Figure 7 depicts the basic waveforms with 4 samples and its two (possible) components. The individual contributions sum to the original waveform and a brief inspection reveals the major curvatures resulting from the higher frequency, while the fixed offset is captured in the zero frequency component.

Figure 8 shows the harmonics for the basic waveforms of equations **??**, **??**, 11 and **??** for the case of 4 samples. There is only 1 sinusoid for each waveform, with the exception of the sawtooth, which has even harmonics.

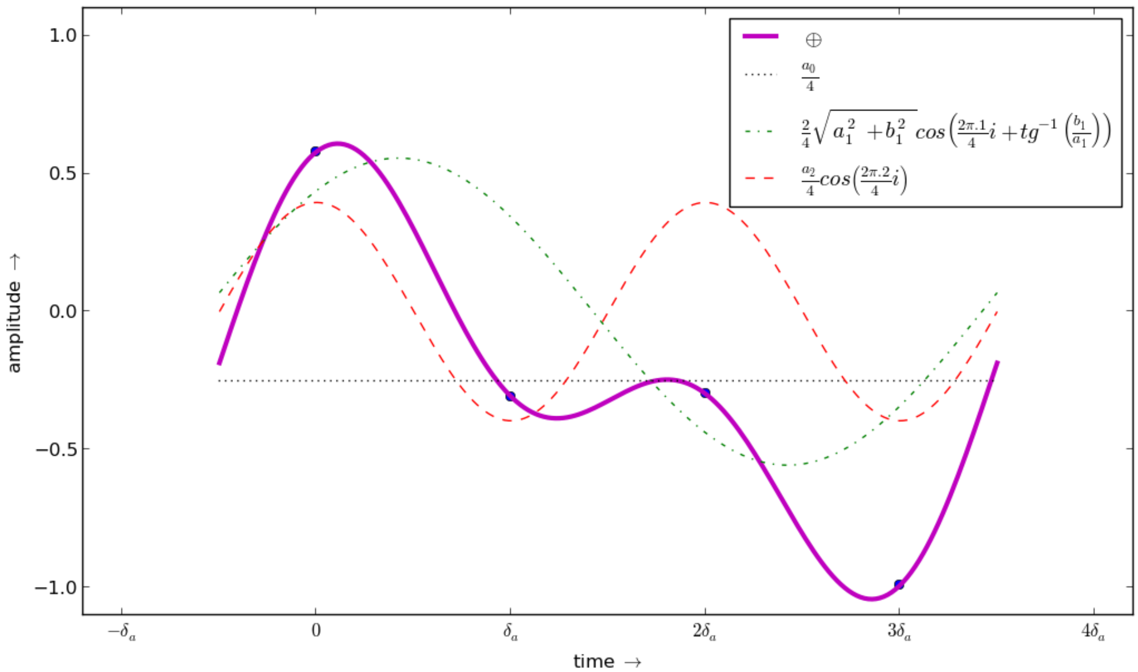

FIG. 7. Frequency components for 4 samples.

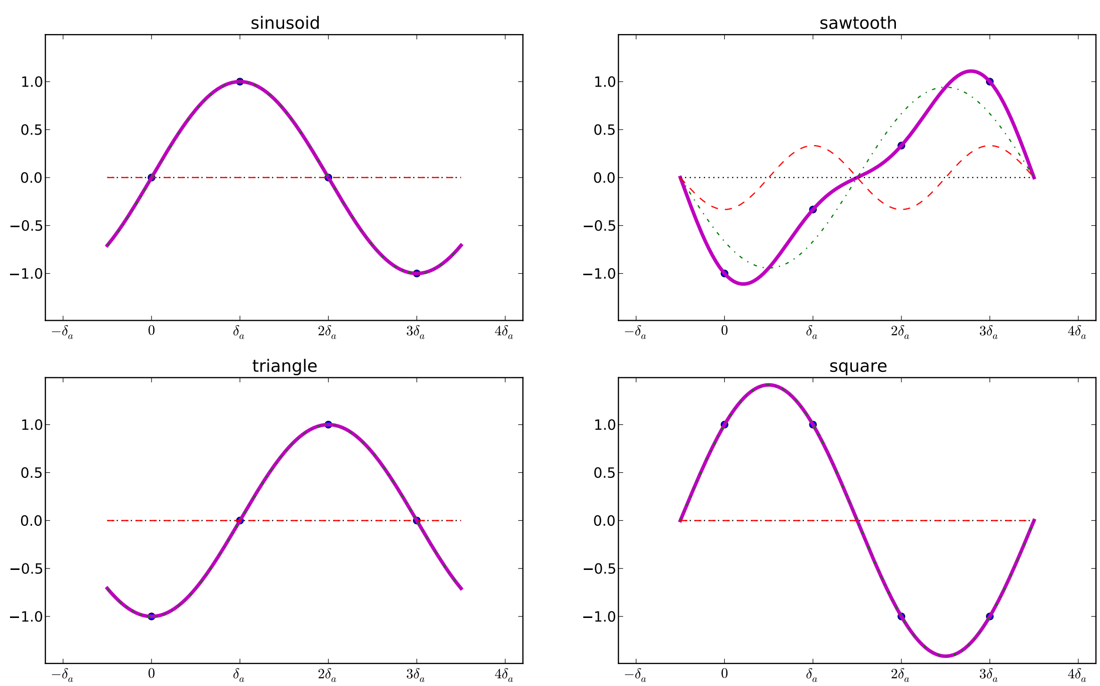

FIG. 8. Basic wave forms with 4 samples.

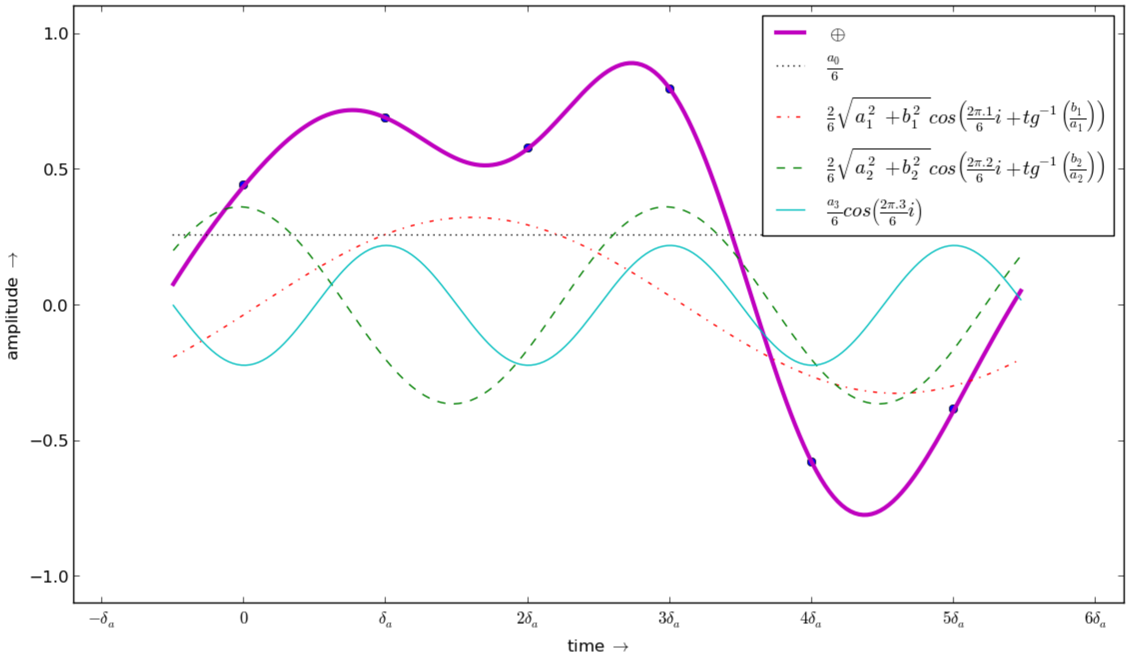

FIG. 9. Frequency components for 6 samples: 4 sinusoids, one of them is the *bias* with zero frequency.

Figure 9 presents the sinusoidal decomposition for 6 samples and figure 10 presents the decomposition of the basic wave forms. In this case, the waveforms have spectra with fundamental differences: square and triangular have the same components but with different proportions, while the sawtooth have an extra component.

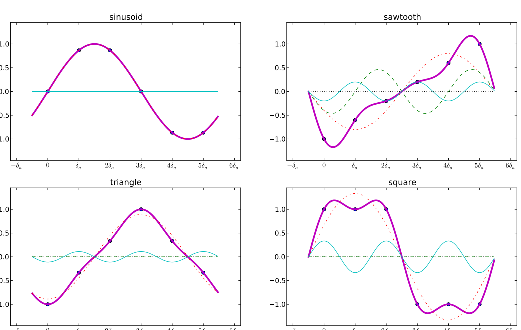

FIG. 10. Basic waveforms with 6 samples: triangular and square waveforms have odd harmonics, with different proportions and phases; the sawtooth has even harmonics.

F. The basic note

Be f such that it divides f_a[8]. A sequence T_i of sonic samples separated by $\delta_a = 1/f_a$ describes a musical note with a frequency of f Hertz and Δ seconds of duration if, and only if, it has the periodicity $\lambda_f = f_a/f$ and size $\Lambda = \lfloor f_a.\Delta \rfloor$:

$$T_i^{f,\,\Delta} = \{t_{i\,\%\lambda_f}\}_0^{\Lambda-1} = \left\{t^f_{i\,\%\left(\frac{f_a}{f}\right)}\right\}_0^{\Lambda-1} \tag{21}$$

The note by itself does not specify a timbre. Nevertheless, it is necessary to choose a waveform for the samples t_i to have a value. An unique period from the basic waveforms can be used to specify the note, where $\lambda_f = \frac{f_a}{f}$ is the number of samples at the period. Here, $L_i^{f,\,\delta_f}$ is the sequence that describes a period of the waveform $L_i^f \in \{S_i^f, Q_i^f, T_i^f, D_i^f, R_i^f\}$ with duration $\delta_f = 1/f$ (as given by equations **??**, **??**, 11 and **??**) and R_i^f is a sampled real waveform:

$$L_i^{f,\delta_f} = \left\{l_i^f\right\}_0^{\delta_f.f_a-1} = \left\{l_i^f\right\}_0^{\lambda_f-1} \tag{22}$$

Therefore, the sequence T_i will consist in a note of duration Δ and frequency f if:

$$T_i^{f,\,\Delta} = \left\{t_i^f\right\}_0^{\lfloor f_a.\Delta\rfloor-1} = \left\{l^f_{i\,\%\left(\frac{f_a}{f}\right)}\right\}_0^{\Lambda-1} \tag{23}$$

G. Spatial localization and spatialization

Although it is not one of its four basic properties, a musical note always has a spatial localization: the note source position at the ordinary physical space. The existence of an environment that reverberates the note is the matter of 'spatialization'. Both fields, spatialization and spatial localization, are widely valued by audiophiles and music industry[?] .

1. Spatial localization

It is understood that the perception of sound localization occurs in our nervous system by three informations: the delay of incoming sound between both ears, the difference of sound intensity at each ear and the filtering realized by the human body, including its chest, head and ears[? ? ?] .

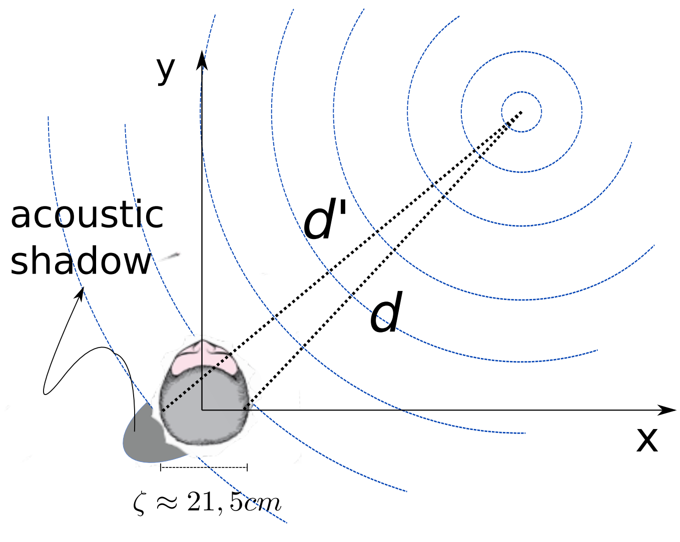

FIG. 11. Detection of sound source localization: schema used to calculate Interaural Time Difference (ITD) and Interaural Intensity Difference (IID).

Considering only the direct incidences in each ear, the equations are quite simple. An object placed at (x, y), as in figure 11, is distant of each ear by:

$$\begin{aligned} d &= \sqrt{\left(x - \frac{\zeta}{2}\right)^2 + y^2} \\ d' &= \sqrt{\left(x + \frac{\zeta}{2}\right)^2 + y^2} \end{aligned} \tag{24}$$

Where ζ is the distance between ears ζ, known to be $\zeta \approx 21,5cm$ in an adult human. Immediate calculations result in the ITD:

$$ITD = \frac{d' - d}{v_{sound\ at\ air} \approx 343.2} \quad \text{seconds} \tag{25}$$

and in the Interaural Intensity Difference:

$$IID = 20 \log_{10}\left(\frac{d}{d'}\right) \quad decibels \tag{26}$$

which, converted to amplitude, yields $IID_a = \frac{d}{d'}$. The IID_a can be used as a multiplicative constant to the right channel of a stereo sound signal: $\{t'_i\}_0^{\Lambda-1} = \{IID_a.t_i\}_0^{\Lambda-1}$, where $\{t'_i\}$ are samples of the wave incident in the left ear. It is possible to use IID together with ITD as a time advance for the right channel. It is a crucial vestige to localization perception of bass sounds and percussive sonorities[?] . With $\Lambda_{ITD} = \lfloor ITD.f_a \rfloor$:

$$\begin{aligned} \Lambda_{ITD} &= \left\lfloor \frac{d' - d}{343,2} f_a \right\rfloor \\ IID_a &= \frac{d}{d'} \\ \left\{t'_{(i+\Lambda_{ITD})}\right\}_{\Lambda_{ITD}}^{\Lambda+\Lambda_{ITD}-1} &= \{IID_a.t_i\}_0^{\Lambda-1} \\ \{t'_i\}_0^{\Lambda_{ITD}-1} &= 0 \end{aligned} \tag{27}$$

with t_i as the right channel and t'_i the left channel. If $\Lambda_{ITD} < 0$, it is only needed to change t_i by t'_i and to use $\Lambda'_{ITD} = |\Lambda_{ITD}|$ and $IID'_a = 1/IID_a$.

Although simple until here, spatial localization depends considerably of other cues. By using only ITD and IID it is possible to specify solely the horizontal angle (azimuthal) θ given by:

$$\theta = \tan^{-1}\left(\frac{y}{x}\right) \tag{28}$$

with x, y as presented in Figure 11. Henceforth, there are problems when θ fall upon the so called "cone of confusion": the same pair of ITD and IID results in a large number of points inside the cone. On those points the inference of the azimuthal angle depends specially of the attenuative filtering for high frequencies, since the head interferes much more in the treble than in bass waves[? ?] . Also relevant to the hearing of lateral sources is that low frequencies diffracts and the wave arrive to the opposite ear with a delay of $\approx 0,7ms$.[?]

Figure 11 depicts the acoustic shadow of the cranium, an important phenomenon to perception of source azimuthal angle in the cone of confusion. The cone itself is not shown in figure 11 because it is not exactly a cone and its precise dimensions were not encountered in the literature. Given the filtering and diffraction dependent of the sound spectrum, it is hard, if not impossible, to correctly draw the confusion cone. Even so, the cone of confusion can be understood as a cone with its top placed in the middle of head and growing out in the direction of each ear? .

On the other hand, the complete localization, including height and distance of sound source, is given by the Head Related Transfer Function (HRTF)? . There are well known open databases of HRTF, such as CIPIC, and it is possible to apply those transfer functions in a sonic signal by convolution (see equation 42)? . Each human body has its filtering and there are techniques to generate HRTFs to be universally used? .

2. Spatialization

Spatialization results from sound reflections and absorptions by room/environment surface where the note is played. The sound propagates through the air with a speed of $\approx 343,2m/s$ and can be emitted from a source with any directionality pattern. When a sound pulse encounters a surface there is reflection. In this reflection occurs: 1) the inversion of propagation speed component that is normal to the surface; 2) energy absorption, specially in trebles. The sonic waves propagate until they reaches inaudible levels (and even further). As a sonic front reaches the human ear, it can be described as the original sound, with the last reflection point as the source, and the absorption filters of each surface it has reached. It is possible to simulate reverberations that are impossible in real systems. For example, it is possible to use asymmetric reflections with relation to the axis perpendicular to the surface, or to increase specific frequency bands (known as 'resonances'), both characteristics are not found in real systems.

There are some reverberation models that are less related to each independent reflection, exploring valuable information to the auditory system. In fact, reverberation can be modeled with a set of 2 temporal and spectral sections:

- First period: 'first reflections' are more intense and scattered.
- Second period: 'late reverberation' is practically a dense succession of indistinct delays with exponential decay and statistical occurrences.
- First band: the bass has some resonance bandwidths relatively spaced.
- Second band: mid and treble have progressive decay and smooth statistical fluctuations.

Smith III points that reasonable concert rooms have total reverberation time of ≈ 1.9 seconds, and that the period of first reflections is around 0.1 seconds. These values suggest that, in the given conditions, there are perceived wave pulses which propagates for 652.08 meters long ($83.79k$ samples in $f_a = 44.1kHz$) before reaching the ear. In addition, sound reflections made after propagation for 34.32 meters long ($4.41k$ samples in $f_a = 44.1kHz$) have incidences less distinct by hearing. These first reflections are particularly important to spacial sensation. The first incidence is the direct sound, described by ITD and IID of equations 25 and 26. Admitting that each one of the first reflections, before reaching the ear, will propagate at least $3-30m$, dependent of room dimension, the separation between the first reflections is $8-90ms$ ($\approx 350-4000$ samples in $f_a = 44.1kHz$). It is possible to experimentally verify that the number of reflections increases with squared proportion $\approx k.n^2$. A discussion about the use of convolutions and filtering to favor implementation of these phenomena is provided in subsection III F, specially at the paragraphs about reverberation.

H. Musical use

Once defined the basic note, it is convenient to build musical structures with sequences based on these particles. The sum of elements with same index of N sequences $T_{k,i} = \{t_{k,i}\}_{k=0}^{N-1}$ with same size Λ results in the overlapped spectral contents of each sequence, in a process called mixing:

$$\{t_i\}_0^{\Lambda-1} = \left\{ \sum_{k=0}^{N-1} t_{k,i} \right\}_0^{\Lambda-1} \tag{29}$$

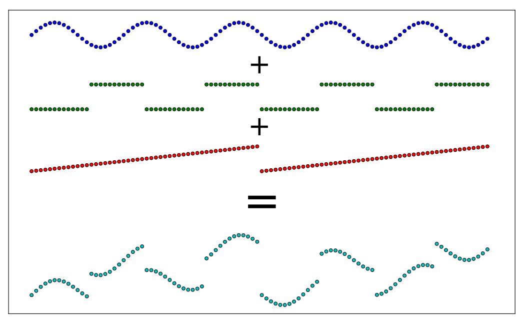

FIG. 12. Mixing of three sound sequences. The amplitudes are directly overlapped.

Figure 12 illustrates this overlapping process of discretized sound waves, each with 100 samples. It is possible to conclude that, if $f_a = 44.1kHz$, the frequencies of the sawtooth, square and sine wave are, respectively: $\frac{f_a}{100/2} = 882Hz$, $\frac{f_a}{100/4} = 1764Hz$ and $\frac{f_a}{100/5} = 2205Hz$. The samples duration are very short $\frac{f_a = 44.1kHz}{100} \approx 2$ milliseconds. One can complete the sequence with zeros to sum (mix) sequences with different sizes.

The mixed notes are generally separated by the ear according to the physical laws of resonance and by the nervous system? . This musical notes mixing process results in musical harmony whose intervals between frequencies and chords of simultaneous notes guide subjective and abstract aspects of music and its appreciation**?** , which is tackled in section IV.

Sequences can be concatenated in time. If the sequences $\{t_{k,i}\}_0^{\Lambda_k-1}$ of size Λ_k represent k musical notes, their concatenation in a unique sequence T_i is a simple melodic sequence, or melody of its own:

$$\begin{aligned} \{t_i\}_0^{\sum \Delta_k - 1} &= \{t_{l,i}\}_0^{\sum \Delta_k - 1}, \\ l \text{ smallest integer } &: \quad \Lambda_l > i - \sum_{j=0}^{l-1} \Lambda_j \end{aligned} \tag{30}$$

This mechanism is demonstrated in Figure 13 with same sequences of Figure 12. Although the sequences are short for the usual sample rates, it is easy to observe the concatenation of sound sequences. Besides that, each note has a duration larger than $100ms$ if $f_a < 1kHz$.

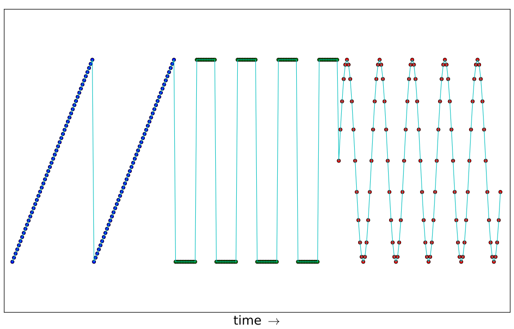

FIG. 13. Concatenation of three sound sequences by temporal overlap of its samples.

The musical piece *reduced-fi* explores the temporal juxtaposition of notes, resulting in a homophobic piece. The vertical principle is demonstrated at the *sounic pictures*, static sounds with peculiar spectrum. Both pieces were written in Python and are available as part of the MASSA *toolbox.***?**

With the basic musical digital note carefully described, the next section develops the temporal evolution of its contents as in *glissandi* and volume envelopes. Filtering of spectral components and noise generation complements the musical note as a self contained unity. Section IV is dedicated to the organization of these notes as music by using metrics and trajectories, with regards to traditional music theory.

III. VARIATION IN THE BASIC MUSICAL NOTE

The basic digital music note defined in section **??** has following parameters: duration, pitch, intensity (volume) and timbre. This is a useful and paradigmatic model, but does not exhaust all the aspects of a musical note.

First of all, characteristics of the note modifie along the note itself**?** . For example, a 3 seconds piano note has intensity with abrupt rise at beginning and progressive decay, has spectrum variations with harmonics decaying some others that emerge along time. These variations are not mandatory but sound synthesis guidance for musical uses, as it reflects how sounds appear in nature. This is so considered to be true, that there is a rule of thumb: to make a sound that incites interest by itself, do internal variations on it**?** . To explore all the ways in which variations occur within a note is out of the scope of any work, given the sensibility of the human ear and the complexity of human sound cognition. In the following, primary resources are presented to produce variations in the basic note. It is worthwhile to recall that all the relations in this and other sections are implemented in Python and published as the public domain MASSA toolbox. The musical pieces *Transita para metro*, *Vibra e treme*, *Tremolos, vibratos e a frequncia*, *Trenzinho de caipiras impulsivos*, *Ruidosa faixa*, *Bela rugosi*, *Chorus infantil*, *ADa e SaRa* were made to validate and illustrate concepts of this section. The code that synthesizes these pieces are also part of the toolbox**?** .

A. Lookup Table

The *Lookup Table* (or simply LUT), is an array for indexed operations which substitutes continuous and repetitive calculation. It is used to reduce computational complexity and to use functions without direct calculation, as data sampled from nature. In music its usage transcends these applications, as it simplifies many operations and makes possible the use a single wave period to synthesize sounds in the whole audible spectrum, with any waveform.

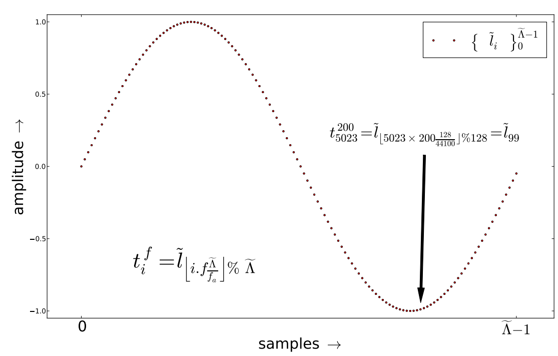

FIG. 14. Search in the *lookup table* to synthesize sounds in different frequencies using an unique waveform with high resolution.

Be $\widetilde{\Lambda}$ the wave period in samples and $\widetilde{L_i} = \left\{ \widetilde{l_i} \right\}_0^{\widetilde{\Lambda}-1}$

the samples $\widetilde{l}_i$ (refer to Equation 22), a sequence $T_i^{f,\Delta}$ with samples of a sound with frequency f and duration Δ can be obtained by means of $\widetilde{L}_i$:

$$T_i^{f,\Delta} = \left\{t_i^f\right\}_0^{\lfloor f_a.\Delta \rfloor - 1} = \left\{\widetilde{l}_{\gamma_i \% \widetilde{\Lambda}}\right\}_0^{\Lambda-1}, \quad \text{where } \gamma_i = \left\lfloor i.f\frac{\widetilde{\Lambda}}{f_a}\right\rfloor \tag{31}$$

In other words, with the right LUT indexes ($\gamma_i \% \widetilde{\Lambda}$) it is possible to synthesize sounds at any frequency. Figure 14 illustrates the calculation of $\{t_i\}$ sample from $\left\{\widetilde{l}_i\right\}$ for $f = 200Hz$, $\widetilde{\Lambda} = 128$ and adopting the sample rate of $f_s = 44.1kHz$. Though this is not a practical configuration (as assigned below), it makes possible a graphical visualization of the procedure.

The calculation of the integer γ_i introduces noise which decreases as $\widetilde{\Lambda}$ increases. In order to use this calculation in sound synthesis, with $f_s = 44.1kHz$, the standard guidelines suggests the use of $\widetilde{\Lambda} = 1024$ samples, since it does not produce relevant noise on the audible spectrum. The rounding or interpolation method are not decisive in this process[?] .

The expression defining the variable γ_i can be understood as f_s being added to i at each second. If i is divided by the sample frequency, $\frac{i}{f_a}$ is incremented in 1 at each second. Multiplied by the period, it results in $i\frac{\widetilde{\Lambda}}{f_a}$, which covers the period in each second. Finally, with frequency f it results in $i.f\frac{\widetilde{\Lambda}}{f_a}$ which completes f periods $\widetilde{\Lambda}$ in 1 second, i.e. the resulting sequence presents the fundamental frequency f.

There are important considerations here: it is possible to use practically any frequency f. Limits exist only in low frequencies when the size of the table $\widetilde{\Lambda}$ is not sufficient for the sample rate f_a. The lookup procedure is virtually costless and replaces calculations by simple indexed searches (what is generally understood as an optimization process). Unless otherwise stated, this procedure will be used along all the text for every applicable case. LUTs are broadly used in computational implementations for music. A classical usage of LUTs is known as *Wavetable Synthesis*, which consists of many LUTs used together to generate a quasi-periodic music note[? ?] .

B. Incremental Variations of Frequency and Intensity

As stated by the Weber and Fechner[?] law, the human perception has a logarithmic relation with the stimulus. In other words, the exponential progression of a stimulues is perceived as linear. For didactic reasons, and given its use in AM and FM synthesis (subsection III E), linear variation is discussed first.

Consider a note with duration $\Delta = \frac{\Lambda}{f_a}$, in which the frequency $f = f_i$ varies linearly from f_0 to $f_{\Lambda-1}$. Thus:

$$F_i = \{f_i\}_0^{\Lambda-1} = \left\{f_0 + (f_{\Lambda-1} - f_0)\frac{i}{\Lambda-1}\right\}_0^{\Lambda-1} \tag{32}$$

$$\Delta_{\gamma_i} = f_i\frac{\widetilde{\Lambda}}{f_a} \quad \Rightarrow \quad \gamma_i = \left\lfloor \sum_{j=0}^{i} f_j\frac{\widetilde{\Lambda}}{f_a}\right\rfloor$$
$$\gamma_i = \left\lfloor \sum_{j=0}^{i} \frac{\widetilde{\Lambda}}{f_a}\left[f_0 + (f_{\Lambda-1} - f_0)\frac{j}{\Lambda-1}\right]\right\rfloor \tag{33}$$

$$\left\{t_i^{\overline{f_0, f_{\Lambda-1}}}\right\}_0^{\Lambda-1} = \left\{\widetilde{l}_{\gamma_i \% \widetilde{\Lambda}}\right\}_0^{\Lambda-1} \tag{34}$$

where $\Delta_{\gamma_i} = f_i\frac{\widetilde{\Lambda}}{f_a}$ is the LUT increment between two samples given the sound frequency of the first sample. Therefore, it is handy to calculate the elements $t_i^{\overline{f_0, f_{\Lambda-1}}}$ by means of the period $\left\{\widetilde{l}_i\right\}_0^{\Lambda-1}$.

The equations 32, 33 and 34 are related with the linear progression of the frequency. As stated above, the frequency progression *perceived* as linear follows an exponential progression, i.e. a geometric progression of frequency is perceived as an arithmetic progression of pitch. It is possible to write: $f_i = f_0.2^{\frac{i}{\Lambda-1}n_8}$ where $n_8 = \log_2\frac{f_{\Lambda-1}}{f_0}$ is the number of octaves between f_0 and $f_{\Lambda-1}$. Therefore, $f_i = f_0.2^{\frac{i}{\Lambda-1}\log_2\frac{f_{\Lambda-1}}{f_0}} = f_0.2^{\log_2\left(\frac{f_{\Lambda-1}}{f_0}\right)^{\frac{i}{\Lambda-1}}} = f_0\left(\frac{f_{\Lambda-1}}{f_0}\right)^{\frac{i}{\Lambda-1}}$. Accordingly, the equations for linear pitch transition are:

$$F_i = \{f_i\}_0^{\Lambda-1} = \left\{f_0\left(\frac{f_{\Lambda-1}}{f_0}\right)^{\frac{i}{\Lambda-1}}\right\}_0^{\Lambda-1} \tag{35}$$

$$\Delta_{\gamma_i} = f_i\frac{\widetilde{\Lambda}}{f_a} \quad \Rightarrow \quad \gamma_i = \left\lfloor \sum_{j=0}^{i} f_j\frac{\widetilde{\Lambda}}{f_a}\right\rfloor$$
$$\gamma_i = \left\lfloor \sum_{j=0}^{i} f_0\frac{\widetilde{\Lambda}}{f_a}\left(\frac{f_{\Lambda-1}}{f_0}\right)^{\frac{j}{\Lambda-1}}\right\rfloor \tag{36}$$

$$\left\{t_i^{\overline{f_0, f_{\Lambda-1}}}\right\}_0^{\Lambda-1} = \left\{\widetilde{l}_{\gamma_i \% \widetilde{\Lambda}}\right\}_0^{\Lambda-1} \tag{37}$$

The term $\frac{i}{\Lambda-1}$ convers the interval $[0,1]$ and it is possible to power it to a value in such a way that the beginning of the transition will be smoother or steeper. This procedure is useful for energy variations with the purpose of changing the volume[9]. It is sufficient to multiply the original sequence by the sequence $a_{\Lambda-1}^{\left(\frac{i}{\Lambda-1}\right)^\alpha}$, where α is

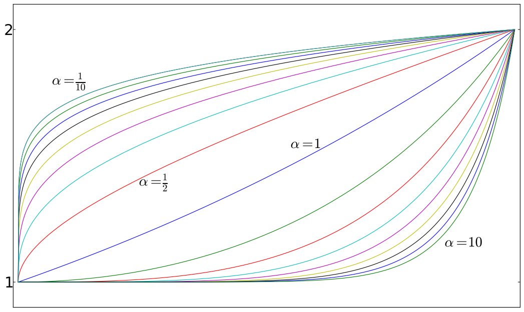

FIG. 15. Intensity transitions to different values of α (see equations 38 and 39).

the given coefficient and $a_{\Lambda-1}$ is a fraction of the original amplitude, which is the value to be reached at the end of transition.

Thus, for amplitude variations:

$$
\begin{aligned}
\{a_i\}_0^{\Lambda-1} &= \left\{ a_0 \left(\frac{a_{\Lambda-1}}{a_0} \right)^{\left(\frac{i}{\Lambda-1}\right)^\alpha} \right\}_0^{\Lambda-1} = \\
&= \left\{ (a_{\Lambda-1})^{\left(\frac{i}{\Lambda-1}\right)^\alpha} \right\}_0^{\Lambda-1} \quad (\text{with } a_0 = 1)
\end{aligned} \tag{38}
$$

and

$$
\begin{aligned}
T_i^{'} = T_i \odot A_i =& \{t_i.a_i\}_0^{\Lambda-1} \\
=& \left\{ t_i.(a_{\Lambda-1})^{\left(\frac{i}{\Lambda-1}\right)^\alpha} \right\}_0^{\Lambda-1}
\end{aligned} \tag{39}
$$

It is often convenient to have $a_0 = 1$ in order to start a new sequence with the original amplitude and then progressively change it. If $\alpha = 1$, the amplitude variation follows the geometric progression whose defines the linear variation of volume. Figure 15 depicts transitions between values 1 and 2 and for different values of α, a gain of $\approx 6dB$ as given by equation 4.

Special attention should be dedicated while considering $a = 0$. In equation 38, $a_0 = 0$ results in a division by zero and if $a_{\Lambda-1} = 0$, there will be multiplication by zero. Both cases make the procedure useless, once no number different from zero can be represented as a ratio in relation of zero. It is possible to solve this dilemma choosing a number that is small enough like $-80dB \;\Rightarrow a = 10^{\frac{-80}{20}} = 10^{-4}$ as the minimum volume for a *fade in* ($a_0 = 10^{-4}$) or for a *fade out* ($a_{\Lambda-1} = 10^{-4}$). A linear fade can be used to reach zero amplitude. Another common solution is the use of the quartic polynomial term x^4, as it reaches zero without these difficulties and gets reasonably close to the curve with $\alpha = 1$ as it withdraws from zero**?** .

For linear amplification – but not linear perception – it is sufficient to use an appropriate sequence $\{a_i\}$:

$$
a_i = a_0 + (a_{\Lambda-1} - a_0)\frac{i}{\Lambda - 1} \tag{40}
$$

Here the convertion between decibels and amplitude is convenient. In this way, the equations 7 and 39 specify a transition of V_{dB} decibels:

$$
T_i^{'} = \left\{ t_i 10^{\frac{V_{dB}}{20}\left(\frac{i}{\Lambda-1}\right)^\alpha} \right\}_0^{\Lambda-1} \tag{41}
$$

for the general case of amplitude variations following a geometric progression. The greater the value of α, the smoother the sound introduction and more intense its end. $\alpha > 1$ results in volume transitions commonly called *slow fade*, while $\alpha < 1$ results in *fast fade***?** .

The linear transitions will be used for AM and FM synthesis, while logarithmic transitions are proper tremolos and vibratos, as developed in subsection III E. A non-oscillatory exploration of these variations is in the music piece *Transita para metro*, which code is online as part of the MASSA toolbox**?** .

C. Application of Digital Filters

This subsection is limited to a description of the procedure in sequences by convolution and differential equations, and the immediate applications. Its complexity exceeds to the scope of this study[10]. The filters application can be part of the synthesis process or made individually as part of processes commonly called sound treatment.

- Convolution and finite impulse response (FIR) filters

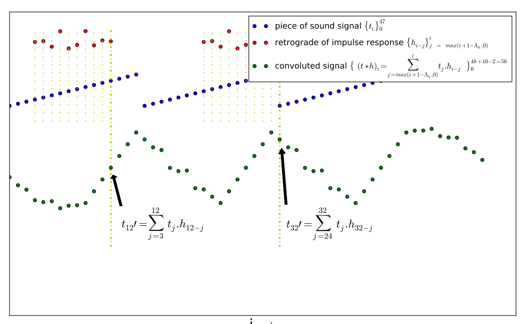

FIG. 16. Graphical interpretation of convolution. Each resulting sample is the sum of the previous samples of a signal, with each one multiplied by the retrograde sequence of samples from the other signal.

Filters applied by means of convolution are known by acronym FIR (Finite Impulse Response) and are characterized by having a time finite sample representation. This sample representation is called 'impulse response' $\{h_i\}$. The FIR filters are applied to the time domain of the digital sound by means of convolution with the respective impulse response of the filter[11]. For this study, the convolution is defined as:

$$\{t'_i\}_0^{\Lambda_t+\Lambda_h-2\,=\,\Lambda_{t'}-1} = \{(T_j * H_j)_i\}_0^{\Lambda_{t'}-1} = \left\{\sum_{j=0}^{min(\Lambda_h-1,i)} h_j.t_{i-j}\right\}_0^{\Lambda_{t'}-1} = \left\{\sum_{j=max(i+1-\Lambda_h,0)}^{i} t_j.h_{i-j}\right\}_0^{\Lambda_{t'}-1} \tag{42}$$

where $t_i = 0$ for the samples not given beforehand. In other words, the sound $\{t'_i\}$ resulting from the convolution of $\{t_i\}$ with the given impulse response $\{h_i\}$ has each i-th sample t_i overwritten by the sum of its last Λ_h samples $\{t_{i-j}\}_{j=0}^{\Lambda_h-1}$ multiplied one-by-one by samples of the another impulse response $\{h_i\}_0^{\Lambda_h-1}$. This procedure is illustrated in Figure 16, where the impulse response $\{h_i\}$ is traveled by its retrograde form, and t'_{12} and t'_{32} are two calculated samples using the convolution given by $(T_j * H_j)_i = t'_i$. The resulting signal always has the length of $\Lambda_t + \Lambda_h - 1 = \Lambda_{t'}$.

With this procedure it is possible to apply reverbs, equalizers, *delays* to name a few of a variety of other filters for sound processing, as well as musical/artistic effects.

The impulse response can be provided by physical measures or by pure synthesis. An impulse response for reverb application can be obtained by sound recording of the environment when someone triggers a snap which resembles an impulse or a sinusoidal lookup whose Fourier transform approximates its frequency response. Both are impulse responses which, properly convoluted with the sound sequence, result in the own sequence with a reverberation that resembles the physical one in the original environment where the measure happened[?] .

The inverse Fourier transform of an envelope even and real is equivalent to a impulse response of a FIR filter. It performs a frequency filtering with the envelope. The greater the number of samples, the higher the envelope resolution and the computational complexity as well. The convolution is recognized as computationally expensive.

An important property is the time shift caused by convolution with a shifted impulse. Despite computationally expensive, it is possible to create *delay lines* by means of sound convolution with an impulse response that has an impulse to each relapse of sound. In Figure 17 it is possible to observe the shifting caused by convolution with each impulse. Depending on the intensity of the impulses, the result is perceived in rhythm (around 20 impulses by second) or in pitch (around 20-40 impulses by second). In the last case, the processes resemble granular synthesis, delays, reverbs and equalization.

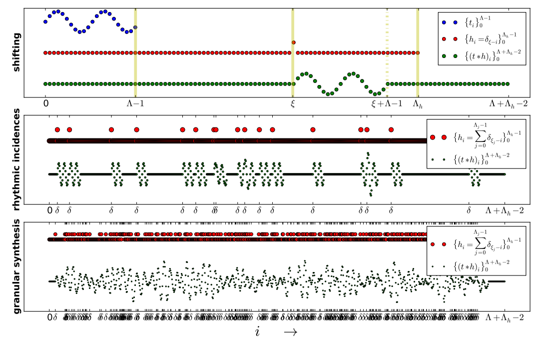

FIG. 17. Convolution with impulse: shifting (a), delay lines (b) and granular synthesis (c). Disposed in increasing order of its pulse density.

- Infinite impulse response (IIR) filters

  This class of filters is known by the acronym IIR. They are characterized by having an infinite time representation, i.e. the impulse response does not converges to zero. Its application is usually made by the following equation:

$$t'_i = \frac{1}{b_0}\left(\sum_{j=0}^{J} a_j.t_{i-j} + \sum_{k=1}^{K} b_k.t'_{i-k}\right) \tag{43}$$

with $b_0 = 1$ in the most cases it is possible to normalize the variables: $a'_j = \frac{a_j}{b_0}$ and $b'_k = \frac{b_k}{b_0} \Rightarrow b'_0 = 1$. Equation 43 is called 'difference equation' because the resulting samples $\{t'_i\}$ are given by differences between the original samples $\{t_i\}$ and the previous resulting ones $\left\{t'_{i-k}\right\}$.

There are many methods and tools to prepare IIR filters. The following texts list a selection of them for didactic purpose and to make easy future consulting. They are well behaved filters whose aspects are described in Figure 18.

Considering filters with simple order, the cutoff frequency f_c is the position where the filter performs an attenuation of $-3dB \approx 0.707$ in the original amplitude. For band-pass and band-reject (or 'notch') filters, this attenuation results on two specifications: f_c (in this case, referred as 'center frequency') band width bw, in both frequencies $f_c \pm bw$ there is an attenuation of ≈ 0.707 in the original amplitude. There is sound amplification in band-pass and band-reject filters when the cutoff frequency is low and the band width is large enough.

In trebles those filters present only a shift in the expected waveform, expanding the envelope to the bass side of the band.

For filters whose frequency responses have other envelopes (for the module), it is possible to create chains and apply them successively. Another possibility is to use some biquad 'filter receipt'[12] or procedures for the calculation of coefficients of Chebichev filters[13]. Both possibilities are explored by the referenced studies, in special[?, ?], and the collection of filters maintained by the *Music-DSP* community from University of Columbia.[?, ?]

1. Low-pass with simple pole and module of the frequency response in the upper left corner of Figure 18. The general equation has as reference by the cutoff frequency $f_c \in (0, \frac{1}{2})$, fraction of the sample frequency f_a that there is approximately an attenuation of $3dB$. The coefficients a_0 and b_1 of the IIR filter are given by means of the intermediate variable $x \in [e^{-\pi}, 1]$:

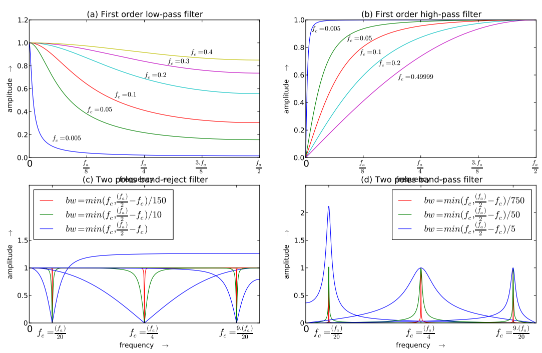

FIG. 18. Modules for the frequency response (a), (b), (c) and (d) for IIR filters of equations 44, 45, 47 and 48 respectively, considering different cutoff frequencies, center frequencies and band width.

$$
\begin{aligned}
x &= e^{-2\pi f_c} \\
a_0 &= 1 - x \\
b_1 &= x
\end{aligned} \tag{44}
$$

2. High-pass filter with simple pole and modules of its frequency responses at the upper right corner of Figure 18. The general equation with cutoff frequency $f_c \in (0, \frac{1}{2})$ is calculated by means of the intermediate variable $x \in [e^{-\pi}, 1]$:

$$
\begin{aligned}
x &= e^{-2\pi f_c} \\
a_0 &= \frac{x+1}{2} \\
a_1 &= -\frac{x+1}{2} \\
b_1 &= x
\end{aligned} \tag{45}
$$

3. Notch filter. This filter is parametrized by a center frequency[14] f_c and band width bw – $f_c \pm bw$ that results in 0.707 of the amplitude, i.e. attenuation of $3dB$ – both given as fractions of f_a, therefore $f,\ bw \in (0, 0.5)$.

   In order to simplify, consider the auxiliary variables K and R defined as:

$$
\begin{aligned}
R &= 1 - 3bw \\
K &= \frac{1 - 2R\cos(2\pi f_c) + R^2}{2 - 2\cos(2\pi f_c)}
\end{aligned} \tag{46}
$$

   The band-pass filter in the lower left corner of Figure 18 has the following coefficients in equation 43:

$$
\begin{aligned}
a_0 &= 1 - K \\
a_1 &= 2(K - R)\cos(2\pi f_c) \\
a_2 &= R^2 - K \\
b_1 &= 2R\cos(2\pi f_c) \\
b_2 &= -R^2
\end{aligned} \tag{47}
$$

   The coefficients of band-reject filter are:

$$
\begin{aligned}
a_0 &= K \\
a_1 &= -2K\cos(2\pi f_c) \\
a_2 &= K \\
b_1 &= 2R\cos(2\pi f_c) \\
b_2 &= -R^2
\end{aligned} \tag{48}
$$

   with the module of its frequency response represented in the lower left corner of the Figure 18.

D. Noise

Generally, sounds without a defined pitch are called noise[?]. They are important parts of the musical sounds with undefined pitch like the noise present in piano notes, violin, etc. Besides that, the majority of percussion instruments do not have defined pitch and their sounds are generally comprehended as noise[?]. In electronic music, including electro-acoustic and dance genres, noise has diversified use, and commonly characterizes the music style[?].

The absence of a defined pitch is due to the absence of a harmonic organization in the sinusoidal components that compose the sound. In this way, there are many ways to generate noise. The use of random values to generate the sound sequence T_i is an attractive method but the result is generally not useful because it tends to the white noise[?].

Another way is to generate noise using the desired spectrum, from which it is possible to perform the inverse Fourier transform. The spectral distribution should be done with care because, if it uses the same phase or phases with strong correlation, the synthesized sound will include considerably amount of concentrated energy in some periods of its duration.

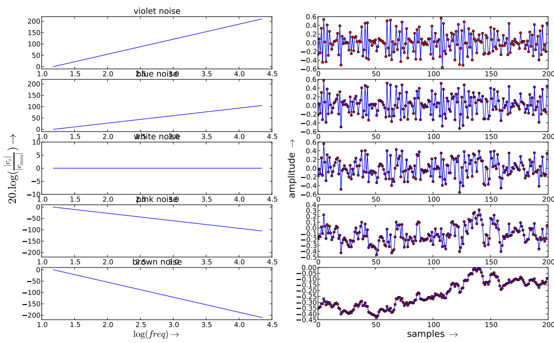

FIG. 19. Colors of noise generated by equations 49, 50, 51, 52 and 53: resulting spectrum and waveforms.

Some noise with static spectrum are listed below. They are called *colors of noise* since they are associated with colors. Figure 19 shows the spectrum profile and the corresponding sound sequence side-by-side. All the noise were generated with a same phase, making it possible to observe the contributions of different regions of the spectrum.

- The white note has its name because its energy are distributed equally between all the frequencies. It is possible to realize the white noise with the inverse transform of the following coefficients:

$$\begin{aligned}
&c_0 = 0 \quad \text{avoiding bias}\\
&c_i = e^{j.x}\ ,\ \ j^2 = -1\ ,\ \ x \text{ random } \in\ [0, 2\pi]\ ,\ \ i\ \in\ \left[1,\ \frac{\Lambda}{2} - 1\right]\\
&c_{\Lambda/2} = 1 \qquad \text{(if even } \Lambda)\\
&c_i = c^*_{\Lambda-i}\ ,\ \ \text{for } i\ >\ \frac{\Lambda}{2}
\end{aligned} \tag{49}$$

  The value c_i, calculated by the exponential, is only an artifice to obtain unitary module and random phase. Besides that, $c_{\Lambda/2}$ is always real (as discussed in the previous section).

- The pink noise has a decrease of $3dB$ for octave. This noise is useful for testing electronic devices, and it has prominent presence in nature[?] .

$$\begin{aligned}
&f_{\text{min}} \approx 15Hz\\
&f_i = i\frac{f_a}{\Lambda}\ ,\qquad i\ \leq\ \frac{\Lambda}{2},\ \ i\ \in\ \mathbb{N}\\
&\alpha_i = \left(10^{-\frac{3}{20}}\right)^{\log_2\left(\frac{f_i}{f_{\text{min}}}\right)}\\
&c_i = 0\ ,\ \ \forall\, i\ :f_i < f_{\text{min}}\\
&c_i = e^{j.x}.\alpha_i\ ,\ \ j^2 = -1\ ,\\
&x \text{ random } \in\ [0, 2\pi]\ ,\ \ \forall\, i\ :f_{\text{min}} \leq f_i < f_{\lceil\Lambda/2-1\rceil}\\
&c_{\Lambda/2} = \alpha_{\Lambda/2} \quad (\text{se } \Lambda \text{ par})\\
&c_i = c^*_{\Lambda-i}\ ,\ \ \text{for } i\ >\ \Lambda/2
\end{aligned} \tag{50}$$

  The minimum frequency f_{min} can be chosen based on the human hearing limit, since no one listens to a pitch with a sound component which frequency is bellow $\approx\ 20Hz$.

  Other noises can be made from the pink noise procedure by simply modifying some details, specially the equation that defines α_i.

- The brown noise received this name after Robert Brown, who described the brownian movement[15]

  What characterizes this noise is the decreasing of $6dB$ by octave. In this way, α_i in the set 50 is defined as:

$$\alpha_i = (10^{-\frac{6}{20}})^{\log_2\left(\frac{f_i}{f_{\text{min}}}\right)} \tag{51}$$

- In the blue noise there is a gain of $3dB$ by octave in a band limited by the minimum frequency f_{min} and the maximum frequency f_{max}. Therefore, the corresponding equation is also based on the equations set 50:

$$\begin{aligned}
&\alpha_i = (10^{\frac{3}{20}})^{\log_2\left(\frac{f_i}{f_{\text{min}}}\right)}\\
&c_i = 0\ ,\ \ \forall\, i\ :f_i < f_{\text{min}} \ \text{ ou }\ f_i > f_{\text{max}}
\end{aligned} \tag{52}$$

- The violet noise is similar to the blue noise, but its gain is $6dB$ by octave:

$$\alpha_i = (10^{\frac{6}{20}})^{\log_2\left(\frac{f_i}{f_{\text{min}}}\right)}\ ,\qquad f_{\text{min}} \approx 15Hz \tag{53}$$

- The black noise has higher losses than $6dB$ for octave:

$$\alpha_i = (10^{-\frac{\beta}{20}})^{\log_2\left(\frac{f_i}{f_{\text{min}}}\right)}\ ,\quad \beta > 6 \tag{54}$$

- The gray noise is defined as a white noise subject to one of the ISO-audible curves. Those curves are obtained by experiments and imperative to obtain α_i. An implementation of ISO 226, that is the last revision of those curves, is in the toolbox MASSA.[?]

This subsection exposed only noises with static spectrum. There are also characterizations for noises with dynamic spectrum during the time, or noises which are fundamentally transient, like clicks and chirps. The former are easily modeled by an impulse relatively isolated, while chirps are not in fact a noise, but a fast scan of some given frequency band[?] .

The noise from equations 49, 50, 51, 52 and 53 are presented in Figure 19. The spectrum were build with the same phase and frequency for each coefficient, making it straightforward to observe the contribution of treble harmonics and bass frequencies.

E. Tremolo and vibrato, AM and FM

Vibrato is a period variation in pitch (frequency) and tremolo is a period variation in volume (intensity)[16] For the general case, the vibrato is described as follows:

$$\gamma'_i = \left\lfloor i f' \frac{\widetilde{\Lambda}_M}{f_a} \right\rfloor \tag{55}$$

$$t'_i = \widetilde{m}_{\gamma'_i \, \% \widetilde{\Lambda}_M} \tag{56}$$

$$f_i = f \left(\frac{f+\mu}{f} \right)^{t'_i} = f.2^{t'_i \frac{\nu}{12}} \tag{57}$$

$$\Delta_{\gamma_i} = f_i \frac{\widetilde{\Lambda}}{f_a} \quad \Rightarrow \quad \gamma_i = \left\lfloor \sum_{j=0}^{i} f_j \frac{\widetilde{\Lambda}}{f_a} \right\rfloor = \\ = \left\lfloor \sum_{j=0}^{i} \frac{\widetilde{\Lambda}}{f_a} f \left(\frac{f+\mu}{f} \right)^{t'_j} \right\rfloor = \left\lfloor \sum_{j=0}^{i} \frac{\widetilde{\Lambda}}{f_a} f.2^{t'_j \frac{\nu}{12}} \right\rfloor \tag{58}$$

$$T_i^{f,vbr(f',\nu)} = \left\{ t_i^{f,vbr(f',\nu)} \right\}_0^{\Lambda-1} = \left\{ \widetilde{l}_{\gamma_i \% \widetilde{\Lambda}} \right\}_0^{\Lambda-1} \tag{59}$$

For the properly realization of vibrato, it is important to pay attention on the following tables (quais tabelas? seriam termos?) and sequences. Table $\widetilde{M_i}$ with length $\widetilde{\Lambda}_M$ and the sequence with indices γ'_i make the sequence t'_i which is the oscillation pattern in the frequency while table $\widetilde{L}_i$ with length $\widetilde{\Lambda}$ and the sequence with indices γ_i make t_i which is the sound itself. Variables μ and ν quantify vibrato intensity: μ is a direct measure of how many Hertz are involved in the upper limit of the oscillation while ν is the direct measure of how many semitones (or half steps) are involved in the oscillation (2ν is the number of semitones between the upper and lower pikes of frequency oscillations of the sound $\{t_i\}$ caused by the vibrato). It is convenient to use $\nu = \log_2 \frac{f+\mu}{f}$ in this case because the maximum frequency increase is not equivalent to the maximum diminish, but the semitones variations remains.

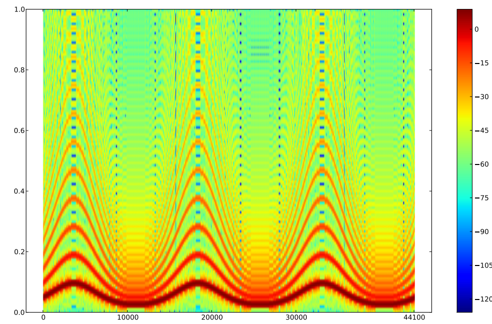

FIG. 20. Spectrogram of a sound with sinusoidal vibrato of $3Hz$ and one octave depth in a $1000Hz$ sawtooth wave (considering $f_a = 44.1kHz$).

Figure 20 illustrates the spectrogram of an artificial vibrato for a note with $1000Hz$ (between a B and a C), and which frequency deviation reaches one octave above and one below. Any waveform can be used to generate a sound and a vibrato oscillation pattern given any oscillation frequency and pitch deviation[17]. Those oscillations with precise forms and arbitrary amplitudes are not possible in traditional music instruments and, in this way, it introduces novelty in the artistic possibilities.

Tremolo is similar: f', γ'_i and t'_i remains the same. The amplitude sequence to be multiplied by the original sequence t_i turns in:

$$a_i = 10^{\frac{V_{dB}}{20} t'_i} = a_{\max}^{t'_i} \tag{60}$$

$$T_i^{tr(f')} = \left\{ t_i^{tr(f')} \right\}_0^{\Lambda-1} = \{t_i.a_i\}_0^{\Lambda-1} = \\ = \left\{ t_i.10^{t'_i \frac{V_{dB}}{20}} \right\}_0^{\Lambda-1} = \left\{ t_i.a_{\max}^{t'_i} \right\}_0^{\Lambda-1} \tag{61}$$

where V_{dB} is the oscillation depth in decibels of tremolo and $a_{\max} = 10^{\frac{V_{dB}}{20}}$ is the maximum amplitude gain. The measurement in decibels is pertinent because the maximum increase in amplitude is not equivalent to the related maximum decrease, while the difference in decibels remains.

Figure 21 shows the amplitude of sequences $\{a_i\}_0^{\Lambda-1}$ and $\{t'_i\}_0^{\Lambda-1}$ for three oscillations of a tremolo with a sawtooth waveform. The curvature is due to the logarithmic progression of the intensity. The tremolo frequency is $1,5Hz$ because $f_a = 44,1kHz \Rightarrow$ duration $= \frac{i_{\max}=82000}{f_a} = 2s \Rightarrow \frac{3\text{oscillations}}{2s} = 1,5$ oscillations for second (Hz).

The music piece *Vibra e treme* explores those possibilities given by tremolos and vibratos, both used in conjunction and independently, with frequencies f', different depths (ν and V_{dB}), and progressive parameters variations[18]. Aiming a qualitative appreciation, the piece also develops a comparison between vibratos and tremolos in logarithmic and linear scales. Its source code is in the Appendix **??** and available online as part of the MASSA*toolbox*.

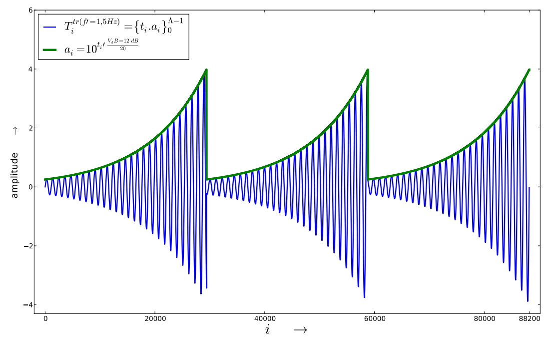

FIG. 21. Tremolo with a depth of $V_{dB} = 12dB$ with a sawtooth waveform as its oscillatory pattern with $f' = 1.5Hz$ in a sine of $f = 40Hz$ (considering a sample frequency of $f_a = 44,1kHz$).

With progressive increasing of f' for a linear proximity of frequency aiming to hear the phenomena as pitch ($\approx 20Hz$) it (o que?) generates roughness to both tremolos and vibratos. Those roughness are largely appreciated both in traditional classical music and current electronic music, specially in the *Dubstep* genre. Roughness is also generated by spectral content that produce beating[? ?]. The sequence *Bela Rugosi* explores the threshold with concomitants of tremolos and vibratos at the same voice, with different intensity and waveforms. The respective code is presented in Appendix **??** and available online in MASSA.

Increasing even more the frequency makes those oscillations no longer remains noticeable. In this case, the oscillations are audible as pitch. Then, f', μ and the waveform changes together the frequency of original sound T_i in different ways for both tremolos and vibratos. They are called AM (*Amplitude Modulation*) and FM (*Frequency Modulation*) synthesis, respectively. These techniques are well known, with applications in synthesizers like *Yamaha DX7*, and even with applications outside music, like in telecommunications to data transferring by means of electromagnetic waves (e.g. AM and FM radios).

For musical ends, it is possible to understand FM based on the case of sines and to decompose the signals into their respective Fourier components (i.e. sinusoidal) to more complex cases. In this way, the FM synthesis performed with a sinusoidal vibrato with frequency f' and depth μ in a sinusoidal sound T_i with frequency f generates bands centered in f and far from each other with a distance of f':

$$\begin{aligned}\{t'_i\} &= \left\{\cos\left[f.2\pi\frac{i}{f_a-1} + \mu.sen\left(f'.2\pi\frac{i}{f_a-1}\right)\right]\right\} = \\ &= \left\{\sum_{k=-\infty}^{+\infty} J_k(\mu)\cos\left[f.2\pi\frac{i}{f_a-1} + k.f'.2\pi\frac{i}{f_a-1}\right]\right\} = \\ &= \left\{\sum_{k=-\infty}^{+\infty} J_k(\mu)\cos\left[(f+k.f').2\pi\frac{i}{f_a-1}\right]\right\}\end{aligned} \tag{62}$$

where

$$\begin{aligned}&J_k(\mu) = \\ &= \frac{2}{\pi}\int_0^{\frac{\pi}{2}}\left[cos\left(\overline{k}\,\frac{\pi}{2} + \mu.\sin w\right).cos\left(\overline{k}\,\frac{\pi}{2} + k.w\right)\right]dw \\ &\overline{k} = k\%2 \;\;,\;\; k \in \mathbb{N}\end{aligned} \tag{63}$$

is the Bessel function[? ?] which specifies the amplitude of each component in FM synthesis.

On those equations, the frequency variation introduced by $\{t'_i\}$ does not respect the geometric progression that follows the pitch perfection but reflects the equation 32. Equations 57 are used for FM and are described in Appendix **??**, where the spectral content of the FM synthesis is calculated and obtained with oscillations in the logarithmic scale. In fact, what is attractive in FM is its simple behavior (only with linear variations in 62).

For the amplitude modulation (AM):

The resulting sound is the original one together with the reproduction of its spectral content below and above the original frequency, with a distance of f' from f. Again, this is obtained by variations in the linear scale of the amplitude. Appendix **??** has an exposition of the spectrum considering an AM performed with an oscillation in the amplitude logarithmic scale. It also loses the simple behavior.

The sequence T_i with frequency f, called 'carrier signal', is modulated by f', called 'modulation signal'. Assuming the FM and AM jargon, μ and $\alpha = 10^{\frac{V_{dB}}{20}}$ are called 'modulation indexes'. The following equations are defined for the vibration pattern of the modulation signal $\{t'_i\}$:

$$\gamma'_i = \left\lfloor i f' \frac{\widetilde{\Lambda}_M}{f_a} \right\rfloor \tag{64}$$

$$t'_i = \widetilde{m}_{\gamma'_i \, \% \widetilde{\Lambda}_M} \tag{65}$$

The modulation signal $\{t'_i\}$ into the carrier signal $\{t_i\}$ for FM is applied as:

$$f_i = f + \mu . t'_i \tag{66}$$

$$\Delta_{\gamma_i} = f_i \frac{\widetilde{\Lambda}}{f_a} \quad \Rightarrow \quad \gamma_i = \left\lfloor \sum_{j=0}^{i} f_j \frac{\widetilde{\Lambda}}{f_a} \right\rfloor = \left\lfloor \sum_{j=0}^{i} \frac{\widetilde{\Lambda}}{f_a} (f + \mu . t'_j) \right\rfloor \tag{67}$$

$$T_i^{f,\, FM(f',\mu)} = \left\{ t_i^{f,\, FM(f',\mu)} \right\}_0^{\Lambda-1} = \left\{ \widetilde{l}_{\gamma_i \% \, \widetilde{\Lambda}} \right\}_0^{\Lambda-1} \tag{68}$$

where $\widetilde{l}$ is the waveform period with a length $\widetilde{\Lambda}$ for the carrier signal.

To perform AM someone just needs to modulate the signal $\{t_i\}$ with $\{t'_i\}$ using the following equations:

$$a_i = 1 + \alpha . t'_i \tag{69}$$

$$T_i^{f,\, AM(f',\alpha)} = \left\{ t_i^{f,\, AM(f',\alpha)} \right\}_0^{\Lambda-1} = \{t_i . a_i\}_0^{\Lambda-1} = \{t_i . (1 + \alpha . t'_i)\}_0^{\Lambda-1} \tag{70}$$

F. Musical usage

At this point, the musical possibilities have been exploded. All characteristics like pitch (given by frequency), timbre (given by the waveform and filters), volume (given by intensity) and duration (given by the number of samples) can be considered by themselves or treated during their duration (with the exception of the duration itself).

The following musical usages comprehend a collection of possibilities with the purpose of exemplifying types of sound manipulation that results in musical material. Some of them are deeply discussed in the next section.

1. Relations between characteristics

Another interesting possibility is to use the relations between parameters of tremolo and vibrato, and some parameters of the basic note like frequency. In this way, it is possible to say that vibrato frequency is proportionally different from pitch, or the tremolo depth is inversely proportional to pitch. Therefore, with equations 55, 57 and 60, it is possible to define:

$$\begin{aligned} f^{vbr} = f^{tr} &= func_a(f) \\ \nu &= func_b(f) \\ V_{dB} &= func_c(f) \end{aligned} \tag{71}$$

with f^{vbr} and f^{tr} as f' in the referenced equations. They can also be associated with the vibrato and tremolo oscillation frequency in equation 55.

Besides that, ν and V_{dB} are the respective depth values of vibrato and tremolo. Functions $func_a$, $func_b$ and $func_c$ are arbitrary and dependents on the musical intentions. The music piece *Tremolos, vibratos e a frequncia* explores such characteristics and shows variations in the oscillation waveform with specific relations with the purpose of building a *musical language*[19]. The corresponding code is on Appendix **??** and it is also available online as part of the MASSAtoolbox.

About convolution it is possible to use a musical pulse as duration – like a BPM pulse – and to distribute impulses during this original pulse, aiming at establishing metrics and rhythms[20]. For example, two impulses equally spaced builds a binary division into the pulse. Two signals, one with 2 pulses and another with 3 pulses, both with equally spaced impulses in the pulse duration, results in the pulse maintenance with a rhythmic mark that is possible to use as binary or ternary divisions in many ethnic or traditions music styles[?] . The absolute values of the impulses results in proportions among the amplitudes of the convoluted signals. Using convolution with impulses as a metric is explored in the music piece *Trenzinho de caipiras impulsivos.* The features embrace the creation of 'sound amalgam' based on granular synthesis and this piece provides a link to the next section. Refer specially to the Figure 24. The source code of the music piece is included in the section **??** and online as well (MASSAtoolbox).

2. Moving audio source and receptor, Doppler effect

Resuming the exposition in subsection II G: when an audio source and a receptor are moving, their characteristics are ideally updated at each sample of the digital signal. The speeds are decomposed in relation of each ear direction. In this way, given the audio source speed (or velocity) v_s, that is positive if the source moves away from receptor, and receptor speed v_r, that is positive when it gets closer of the audio source, the frequency is given by the well known Doppler effect:

$$f = \left(\frac{v_{sound} + v_r}{v_{sound} + v_s} \right) f_0 \tag{72}$$

With this frequency and relations given by the new IID from the new source position, it is possible to create the Doppler effect. There is an addendum to improve the fidelity of the physical phenomena: to increase the received potency. It is possible to understand this potency gain as being proportional to the relative speed that, in each second, adds the traversed period, by means of a waveform with potency: $\Delta P = P_0 \left(\frac{v_r - v_s}{343{,}2}\right)$, where P_0 represents the signal potency.

In this way it is possible to obtain both amplitude and frequency of a moving audio source. Being this audio source in front of receptor with y_0 meters of horizontal distance and z_0 meters of height, the distance is given by $D_i = \left\{d_i = \sqrt{y_i^2 + z_0^2}\right\}_0^{\Lambda-1}$, where $y_i = y_0 + v_s - v_r$ if considering v_s and v_r both in horizon. Amplitude changes with the distance and with the potency factor cited above (see subsection II B for potency to amplitude conversion).

$$A_i = \left\{\frac{z_0}{d_i} A_{\Delta P}\right\}_0^{\Lambda-1} = \left\{\frac{z_0}{\sqrt{y_i^2 + z_0^2}} \sqrt{\frac{v_r - v_s}{343,2} + 1}\right\}_0^{\Lambda-1} \tag{73}$$

Observe that the factor which changes the amplitude (caused by the distance) is even, while the factor caused by the potency variation is antisymmetric in relation to the crossing of source with receptor. The frequency has a symmetric progression in relation to pitch. In other words, the same semitones (or fractions) added during the approach are decreased during the departure. Besides that, the transition becomes steep if both source and receptor intersects each other exactly at the same point, otherwise, there is a monotonic progression. In the given case, where there is a static height z_0, it is necessary to observe the speed component in the direction between the observer and audio source:

$$F_i = \{f_i\}_0^{\Lambda-1} = \left\{\frac{v_{sound} + v_r \frac{y_i}{\sqrt{z_0^2 + y_i^2}}}{v_{sound} + v_s \frac{y_i}{\sqrt{z_0^2 + y_i^2}}} f_0\right\}_0^{\Lambda-1} \tag{74}$$

In the Appendix **??** there is a Python implementation of the Doppler effect as describe above, also considering the intersection between audio source and receptor.

3. Filter and noises (subsections III D and III C)

Using filters the possibilities are many. It is possible to convolve a signal to reverberate it, to remove its noise, to generate distortions or to treat the audio with aesthetically manipulation in mind. For example, it is possible to simulate sounds from and old television if a band-pass filter is applied to accept just sound between $1kHz$ and $3kHz$. Or if someone removes (with some precision) just the frequency of electric oscillation (usually $50Hz$ or $60Hz$) and the harmonics, it will remove noises caused by audio devices. A more musical application is to perform filtering in specific bands and to use those bands as an additional parameter to the notes.

Inspired by traditional music instruments, it is possible to apply a time-dependent filter[?] . Chains of those filters can perform complex and more accurate filtering routines. The music piece *Ruidosa faixa* explores those features by using filters and many kinds of noise synthesis. The source code is in Appendix **??** and is available online as part of MASSA.

When used together the features can create an effect known as *chorus*. Based on what happens in a singers choir, in this effect the sound is performed using small and potentially arbitrary modifications of parameters like center frequency, presence (or absence) of vibrato or tremolo and its characteristics, equalization, volume, etc. As a final result, those versions of the original sound are mixed together (see equation 29). The music piece *Chorus infantil* implements a chorus in many ways with different sounds and its source code is listed in Appendix **??**. The script is also available in the MASSAtoolbox.

4. Reverberation

Using the same spatialization terms of subsection II G, the late reverberation can be modeled as a convolution with a period of pink, brown or black noise, with exponential decay of amplitude and time related. In this way, the treble attenuation and irregular smooth in the frequency response are contemplated with great success. Delay lines can be added as prefix to noise with the decay, and this contemplates both time parts of the reverberation: the first reflections and the late reverberation. It is possible to improve the ?? quality by calculating the geometric localization of the last surface where each wave front reflected before reach the ear in the fairest $100-200$ milliseconds. Then, one could apply a low-pass filter as described in subsection III C. The colored noise can be gradually introduced since the initial moment gives direct incidence (i.e. without any reflection and given by ITD and IID), with *fade-in*, reaching its maximum at the beginning of the 'late reverberation', when the geometric incidences loses their importance to the statistics properties of the noise decay.

As an example, consider Δ_1 as the duration of the first period and Δ_R as the duration of total reverberation ($\Lambda_1 = \Delta_1 f_a$, $\Lambda_R = \Delta_R f_a$). It is possible to add a probability p_i of a sound reincidencing in the i-th sample with amplitude with exponential decay. Following subsection II G, the reverberation R_i^1 of the first period can be described as:

$$\begin{aligned} R_i^1 &= \{r_i^1\}_0^{\Lambda_1-1} \quad : \\ r_i^1 &= \begin{cases} 10^{\frac{V_{dB}}{20}\frac{i}{\Lambda_R-1}} & \text{with probability} \quad p_i = \left(\frac{i}{\Lambda_1}\right)^2 \\ 0 & \text{with probability} \quad 1 - p_i \end{cases} \end{aligned} \tag{75}$$

where V_{dB} is the total decay in decibels, typically $-80dB$ or $-120dB$. Reverberation R_i^2 of the second period can be emulated by a brown noise R_i^m (or by a pink noise R_i^r) with exponential decay:

$$R_i^2 = \left\{r_i^2\right\}_{\Lambda_1}^{\Lambda_R-1} = \left\{10^{\frac{V_{dB}}{20}\frac{i}{\Lambda_R-1}} . r_i^m\right\}_{\Lambda_1}^{\Lambda_R-1} \tag{76}$$

like:

$$R_i = \{r_i\}_0^{\Lambda_R-1} \ : \ r_i = \begin{cases} 1 & \text{if} \quad i = 0 \\ r_i^1 & \text{if} \quad 1 \leq i < \Lambda_1 - 1 \\ r_i^2 & \text{se} \quad \Lambda_1 \leq i < \Lambda_R - 1 \end{cases} \tag{77}$$

the reverberation represented by R_i can be applied as simple convolution of R_i (called reverberation 'impulse response') with the sound sequence T_i as described in subsection III C.

Reverberation is well known by causing great interest in the listeners and to provide more enjoyable sonorities. Besides that, modifications in the reverb space consists of a tricking (almost a *cliché*) to induce surprise and interest in the listener.

5. ADSR Envelopes

The volume variation along the duration of sound is crucial to our timbre perception. The volume envelope, known as ADSR (*Attack-Decay-Sustain-Release*), has many implementations in both hardware and software synthesizers. A pioneer implementation can be found in the Hammond Novachord synthesizer of 1938 and some variants are cited below**?** .

The stochastic ADSR envelope is characterized by 4 parameters: attack duration (time to sound reaches its maximum volume), decay duration (follows the attack immediately), level of sustain volume (in which the volume remains stable after the decay) and release duration (after sustain, this is the duration until the volume decays to zero). Note that the sustain duration is not specified because it is the difference between the duration itself and the durations of attack, decay and sustain.

It is possible to apply the ADSR envelope with durations Δ_A, Δ_D and Δ_R, with total duration Δ and sustain level a_S, given by the fraction of the maximum amplitude, as a sound sequence t_i defined as:

$$\begin{gathered}
\{a_i\}_0^{\Lambda_A-1} = \left\{\xi\left(\frac{1}{\xi}\right)^{\frac{i}{\Lambda_A-1}}\right\}_0^{\Lambda_A-1} \quad \text{or} \quad \left\{\frac{i}{\Lambda_A-1}\right\}_0^{\Lambda_A} \\
\{a_i\}_{\Lambda_A}^{\Lambda_A+\Lambda_D-1} = \left\{a_S^{\frac{i-\Lambda_A}{\Lambda_D-1}}\right\}_{\Lambda_A}^{\Lambda_A+\Lambda_D-1} \\
\text{or} \quad \left\{1-(1-a_S)\frac{i-\Lambda_A}{\Lambda_D-1}\right\}_{\Lambda_A}^{\Lambda_A+\Lambda_D-1} \\
\{a_i\}_{\Lambda_A+\Lambda_D}^{\Lambda-\Lambda_R-1} = \{a_S\}_{\Lambda_A+\Lambda_D}^{\Lambda-\Lambda_R-1} \\
\{a_i\}_{\Lambda-\Lambda_R}^{\Lambda-1} = \left\{a_S\left(\frac{\xi}{a_S}\right)^{\frac{i-(\Lambda-\Lambda_R)}{\Lambda_R-1}}\right\}_{\Lambda-\Lambda_R}^{\Lambda-1} \\
\text{or} \quad \left\{a_S - a_S\frac{i+\Lambda_R-\Lambda}{\Lambda_R-1}\right\}_{\Lambda-\Lambda_R}^{\Lambda-1}
\end{gathered} \tag{78}$$

with $\Lambda_X = \lfloor \Delta . f_a \rfloor \ \ \forall \ \ X \ \in (A, D, R, \ )$ and being ξ a small value that provides a satisfactory *fade in* and *fade out*, e.g. $\xi = 10^{\frac{-80}{20}} = 10^{-4}$ or $\xi = 10^{\frac{-40}{20}} = 10^{-2}$. The lower the ξ, the slower the *fade*, like the α illustrated in Figure 15. The terms in the right side of the equation 78 attend for both introduction and ending of sound from the zero intensity because they are linear. Schematically, Figure 22 shows the ADSR envelope, a classical implementation that supports many variations. For example, between attack and decay it is possible to add an extra partition where the maximum amplitude stays. Another common example is the use of more elaborated tracings to attack or decay. The music piece *ADa e SaRa*, available both in Appendix **??** and in MASSA, explores many configurations of the ADSL envelope.

$$\left\{t_i^{ADSR}\right\}_0^{\Lambda-1} = \{t_i . a_i\}_0^{\Lambda-1} \tag{79}$$

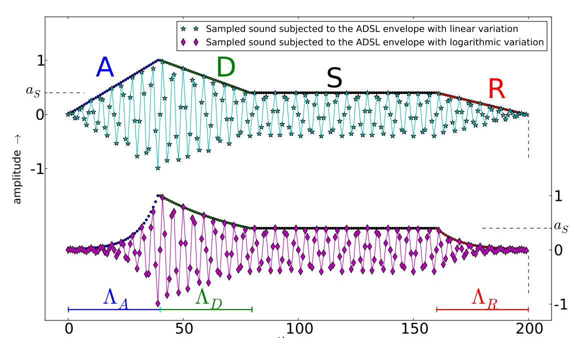

FIG. 22. An ADSR envelope (*Attack, Decay, Sustain, Release*) applied to an arbitrary sound sequence. The linear variation of the amplitude is above. Below the amplitude variation is exponential.

IV. NOTES ORGANIZATION IN MUSIC

Consider $S_j = \left\{ s_j = T_i^j = \{t_i^j\}_{i=0}^{\Lambda_j - 1} \right\}_{j=0}^{H-1}$ as the sequence S_j of H musical events s_j. In addition, S_j, called a 'musical structure', as composed by events s_j which are musical structure themselves, e.g. notes. This section is dedicated to techniques that make S_j interesting and enjoyable for audition.

The elements of S_j can be overlapped by mixing them together as in equation 29 and Figure 12, building intervals and chords. This scenery reflects the 'vertical thought' in music. On the other hand, the concatenation of S_j as in equation 30 and in Figure 13, builds melodic sequences and rhythms, which are associated with the musical 'horizontal thought' in music. The fundamental frequency f and the beginning moment (attack) are generally the most important characteristics of the elements in S_j. Such characteristics make possible the creation of music made by pitches (both harmony and melody) and by the presence of temporal metrics and rhythms.

A. Tuning, intervals, scales and chords

Here we describe characteristics in music related to harmony and melody.

1. Tuning

Doubling the frequency is equivalent to ascending one octave ($f = 2f_0$). The octave division in twelve notes is the canon of classical western music, although its usage has also been observed outside western tradition, like ceremonial/religious and ethnic context[?] .

The factor given by $\varepsilon = 2^{\frac{1}{12}}$ defines a semitone. It builds a note grid along the hearing spectrum. Fixing a frequency f, all the possible fundamental frequencies stay separated by intervals that are multiples of ε. Twelve semitones (or half steps), equidistant to the human ear, make an octave. Therefore, if $f = 2^{\frac{1}{12}} f_0$, there is a semitone between f_0 and f.

This absolute accuracy is common in computational implementations. For the real music instruments, however, there are deviations of those frequencies in order to make the harmonics of corresponding notes more compatible. Besides that, the fixed reference $\varepsilon = 2^{\frac{1}{12}}$ characterizes an equally tempered tuning. There are tunings which respective intervals are ratios of low-order integers, originally based on observations of physical behaviors. First of these tunings were formalized around two thousand years before the existence of an equal temperament[?] .

Two iconic tunings:

- The **just intonation** is defined as notes of the diatonic scale that are associated with ratios of low-order integers that defines the harmonic series. The basic ratios are the same in the Ionian mode (white piano keys from to (see below in subsection IV A 3): 1, 9/8, 5/4, 4/3, 3/2, 5/3, 15/8, 2/1). The intervals are considered as notes in a scale together with the following: the semitone 16/15, the 'minor tone' 10/9, and the 'major tone' 9/8. There are many ways to perform the division of the 12 notes.
- The **Pythagorean tuning** is based on the interval 3/2 (perfect fifth). The Ionian mode becomes: 1, 9/8, 81/64, 4/3, 3/2, 27/16, 243/128, 2/1. The intervals are also considered as notes in a scale. Besides the intervals in the mode, it is also used the minor second 256/243, the minor third 32/27, the augmented fourth 729/512, the diminished fifth 1024/729, the minor sixth 128/81 and the minor seventh 16/9.

In order to attends for micro-tonality[21]. It is possible to use non-integer real values as pitch sequences or even to maintain the integer values and change the factor $\varepsilon = 2^{\frac{1}{12}}$. For example, a tuning really near of the harmonic series itself is proposed with the division of the octave in 53 notes: $\varepsilon_2 = 2^{\frac{1}{53}}$[?] . In that division the notes are related by means of integers with ε_2. Note that if S_i is a pitch sequence related by means of ε_1, a mapping for all the notes related by ε_2 defines a new sequence $S'_i = \{s'_i\} = \left\{ s_i \frac{\varepsilon_1}{\varepsilon_2} \right\}$. The music piece *Micro tom* explores this micro-tonal features and its code is in Appendix **??** and is part of MASSA.

2. Intervals

Using the ratio $\varepsilon = 2^{\frac{1}{12}}$ between the note frequencies (i.e. one semitone) the intervals in the twelve note system can be represented by integers. Table I summarizes the characteristics of each interval: its traditional nomenclature, consonance and dissonance properties, and number of semitones per interval.

The nomenclature is based both in impositions and for convenience of the tonal system and practical aspects of note manipulation and can be specified as[?] [?] :

- Intervals as number of ratios between notes: first (unison), second, third, forth, fifth, sixth, seventh and eighth. The ninth, tenth, eleventh, etc, are compound intervals made by one or more octaves + a interval inside the octave, which characterizes the compound interval. The intervals are represented by the numeric digits, e.g. 1, 3, 5 are a unison, a third and a fifth, respectively.
- Qualities of each interval: the perfect consonances – i.e. unison, forth, fifth and octave – are 'perfect'. The imperfect consonances – i.e. thirds and sixths – and dissonances – i.e. seconds and sevenths – can be major and minor. Exception for the tritone.

TABLE I. The music intervals together with their traditional notation, basic classification for dissonance and consonance, and number of semitones. Perfect (P) unison, fifths and octaves are the perfect consonances. Major (M) and minor (m) thirds and sixths are the imperfect consonances. Minor seconds and major sevenths are the harsh dissonances. Major seconds and minor sevenths are the mild dissonances. Perfect fourth is the first special case, being a perfect consonance when considering it as the inversion of the perfect fifth and a dissonance or a imperfect consonance otherwise. The second special case is the tritone (A4 or aug4, d5 or dim5, tri,TT). This interval is consonant in some cultures. For tonal music, the tritone indicates dominant and searches for its resolution in a third or sixth. Due to this instability it is considered a dissonance interval.

consonances		
	traditional notation	number of semitones
perfects:	P1, P5, P8	0, 7, 12
imperfects:	m3, M3, m6, M6	3, 4, 8, 9

dissonances		
	traditional notation	number of semitones
fortes:	m2, M7	1, 11
brandas:	M2, m7	2, 10

special cases		
	traditional notation	number of semitones
consonance or dissonance:	P4	P5
dissonance in Western tradition:	tritone, aug4, dim5	6

- The perfect fourth is both perfect consonance or dissonance according to the context and theoretical background. As general rule, it can be considered as consonance except when it is used as passage to resolve a fifth or a third.

- Tritone is dissonance in Western music because it characterizes the "dominant" in the tonal system (see subsection IV B) and represents the instability. Some cultures chant the interval as consonance.

- A major interval decreased by one semitone results in a minor interval. A minor interval increased by one semitone results in a major interval.

- A perfect interval (unison, perfect forth, perfect fifth, perfect octave), or a major interval (major second M2, major third M3, major sixth M6 or major seventh M7), increased by one semitone results in an augmented interval (e.g. augmented third aug3 with five semitones). The augmented forth is also called tritone (aug4 tri TT).

- A perfect interval or a minor interval (minor second m2, minor third m3, minor sixth m6 or minor seventh m7), decreased by one semitone results in a diminished interval. The diminished fifth is also called tritone (dim5 tri TT).

- An augmented interval increased by one semitone results in a 'doubly augmented' interval and a diminished interval decreased by one semitone results in a 'doubly diminished' interval.

- Notes played simultaneously configure a harmonic interval.

- Notes played as a sequence in time configure a melodic interval. The order of the notes: considering first the lowest note or the highest note, results in an ascending or descending interval, respectively.

- If the lower pitch is raised one octave, or if the upper pitch is lowered one octave below, the interval is inverted. When summed to its inversion an interval results in 9 (e.g. m7 is inverted to M2: $m7 + M2 = 9$). An inverted major interval results in a minor interval and vice-versa. An inverted augmented interval results in a diminished interval and vice-versa (like a doubly augmented results in a doubly diminished and vice-verse). An inverted perfect interval results in a perfect interval as well.

- An interval higher than an octave is called a 'compound interval' and is classified like the interval between the same notes but in the same octave. They are also specified adding a 7 to the interval: P11 is an octave plus a forth ($7 + P4 = P11$), M9 is an octave plus a major second ($7 + M2 = M9$).

The augmented/diminished intervals and the doubly augmented/doubly diminished intervals are consequences of the tonal system. The scale degrees (subsection IV A 3) generally determine different pitches, with specific uses and functions. Henceforth, in a *C flat* major scale, the tonic – first degree – is *C flat*, not *B*, and the leading tone – seventh degree – is *B flat*, not *A sharp* or *C double flat*. In a similar way, the second degree of a scale can be one semitone far from the first degree like the leading tone (seventh degree at one ascending semitone from the first degree), where there is a diminished third between the two semitones of the seventh and second scale degrees, consequence of the first degree be between the near two degrees: the second and the leading note[?] .

The descriptions presented here summarizes the traditional theory of the music intervals[?] . The music piece *Intervalos entre alturas* explores these intervals in a independently and conjunction manner. The reader is invited to see the source code in Appendix **??** and available online in the MASSAtoolbox[?] .

3. Scales

Basically speaking, scale is a set of ordered pitches. Usually, scales repeat at each octave. The ascending sequence with all notes from the octave division in 12 equal intervals, separated by ratio $\varepsilon = 2^{\frac{1}{12}}$, is known as the chromatic scale with equal temperament. There are 5 perfect symmetric divisions of the octave within the chromatic scale. These divisions are considered as scales by the easily and peculiar use they provide. Considering the integers that the factor $\varepsilon = 2^{\frac{1}{12}}$ is powered to multiply f_0, the scales are the following:

$$\begin{aligned}
&\text{chromatic} = E_i^c = \\
&= \{e_i^c\}_0^{11} = \{0,1,2,3,4,5,6,7,8,9,10,11\} = \{i\}_0^{11} \\
&\text{whole tones} = E_i^t = \{e_i^t\}_0^5 = \{0,2,4,6,8,10\} = \{2.i\}_0^5 \\
&\text{minor thirds} = E_i^{tm} = \{e_i^{tm}\}_0^3 = \{0,3,6,9\} = \{3.i\}_0^3 \\
&\text{major thirds} = E_i^{tM} = \{e_i^{tM}\}_0^2 = \{0,4,8\} = \{4.i\}_0^2 \\
&\text{tritones} = E_i^{tt} = \{e_i^{tt}\}_0^1 = \{0,6\} = \{6.i\}_0^1
\end{aligned} \tag{80}$$

Therefore, the third note of the whole tone scale with $f_0 = 200Hz$ is $f_3 = \varepsilon^{e_3^t}.f_0 = 2^{\frac{6}{12}}.200 \approx 282.843Hz$. These 'scales' or patterns generate stable structures by their intern symmetries and can be repeated in a efficient and sustained way. Section IV G specially discusses the symmetries. The music piece *Cristais* exposes each one of these scales, in both melodic and harmonic counterpart and the corresponding source code is presented in Appendix **??** and as part of MASSA.

The diatonic scales are:

$$\begin{aligned}
&\text{natural minor} = \text{aeolian mode} = \\
&= E_i^m = \{e_i^m\}_0^6 = \{0,2,3,5,7,8,10\} \\
&\text{locrian mode} = E_i^{mlo} = \{e_i^{mlo}\}_0^6 = \{0,1,3,5,6,8,10\} \\
&\text{major} = \text{ionian mode} = E_i^M = \{e_i^M\}_0^6 = \{0,2,4,5,7,9,11\} \\
&\text{dorian mode} = E_i^{md} = \{e_i^{md}\}_0^6 = \{0,2,3,5,7,9,10\} \\
&\text{phrygian mode} = E_i^{mf} = \{e_i^{mf}\}_0^6 = \{0,1,3,5,7,8,10\} \\
&\text{lydian mode} = E_i^{ml} = \{e_i^{ml}\}_0^6 = \{0,2,4,6,7,9,11\} \\
&\text{mixolydian mode} = E_i^{mmi} = \{e_i^{mmi}\}_0^6 = \{0,2,4,5,7,9,10\}
\end{aligned} \tag{81}$$

They have only the major, minor and perfect intervals. The unique exception is the tritone that is presented as a augmented forth or diminished fifth.

All the diatonic scales follow a pattern of successive intervals *tone, tone, semitone, tone, tone, tone, semitone*. Thus, it is possible to write:

$$\begin{aligned}
&\{d_i\} = \{2,2,1,2,2,2,1\} \\
&e_0 = 0 \\
&e_i = d_{(i+\kappa)\%7} + e_{i-1} \quad for \;\; i > 0
\end{aligned} \tag{82}$$

with $\kappa \in \mathbb{N}$. For each mode there is only one value for $\kappa \in [0,6]$. For example, a brief inspection reveals that $e_i^{ml} = d_{(i+2)\%7} + e_{i-1}^{ml}$. Then, $\kappa = 2$ for the lydian mode.

The minor scales has two additional forms, named melodic and harmonic:

$$\begin{aligned}
&\text{natural minor (same as above)} = \\
&= E_i^m = \{e_i^m\}_0^6 = \{0,2,3,5,7,8,10\} \\
&\text{harmonic minor} = E_i^{mh} = \{e_i^{mh}\}_0^6 = \{0,2,3,5,7,8,11\} \\
&\text{melodic minor} = \\
&= E_i^{mm} = \{e_i^{mm}\}_0^{14} = \{0,2,3,5,7,9,11,12,10,8,7,5,3,2,0\}
\end{aligned} \tag{83}$$

The ascending and descending contour of the melodic minor scale is necessary for the existence of the leading tone (seventh and last degree, separated by one semitone of the octave, enhances the tonic polarization) in the ascending trajectory, which is not necessary in the descending mode, since it recovers the normal form. The harmonic scale presents the leading tone but does not avoid the augmented second interval between the sixth and seventh degrees. In addition, it does not need to cover a melodic trajectory, just having to present the leading tone, crucial to the tonal system (the leading tone tends to the tonic, asserting it)[?] . Other scales can be represented in a similar form, like the pentatonic and the modes of limited transposition by Messiaen[?] .

Although not a scale, the harmonic series is often used as such. Besides, this point is a convenient moment to put it plain simple:

$$\begin{aligned}
H_i =&\{h_i\}_0^{19} = \\
&\{0,12,19+0.02,24,28-0.14,31+0.2,34-0.31, \\
&36,38+0.04,40-0.14,42-0.49,43+0.02, \\
&44+0.41,46-0.31,47-0.12, \\
&48,49+0.05,50+0.04,51-0.02,52-0.14\}
\end{aligned} \tag{84}$$

This scale is composed of the frequencies on harmonic spectra. Nature exhibits this frequencies within sounds and with all kinds of distortions.

4. Chords

The simultaneous occurrence of three or more notes is observed by means of chords. Their base of tonal music are the triads. Triads are built by two successive thirds within 3 notes: root, third and fifth. If the lower note of a chord is different from the root, this chord is referred as inverted. A close position is where any note of chord fits between two consecutive notes[?] . All the non-inverted triads in the close position form and with root in 0 are described as following:

$$
\begin{aligned}
\text{major triad} &= A_i^M = \{a_i^M\}_0^2 = \{0,4,7\} \\
\text{minor triad} &= A_i^m = \{a_i^m\}_0^2 = \{0,3,7\} \\
\text{diminished triad} &= A_i^d = \{a_i^d\}_0^2 = \{0,3,6\} \\
\text{augmented triad} &= A_i^a = \{a_i^a\}_0^2 = \{0,4,8\}
\end{aligned}
\tag{85}
$$

To consider another third superimposed to the fifth, is sufficient to add 10 as the highest note in order to form a tetrad with minor seventh, or add 11 in order to form a tetrad with major seventh. The inversions and open positions can be obtained with a selective addition of 12 to the components.

Incomplete triadic chords, with extra notes ('dirty' chords), and non-triadic are also common.

In the following we present general recommendations:

- A fifth is a root (fundamental) confirmed by the interval.
- The major or minor third points to the major or minor quality of the chord.
- Every tritone, specially if built between a major third and a minor seventh, tends to resolve in a third or sixth.
- Note duplication is avoided. If duplication is needed, the preferred order is: the root, fifth, third and seventh.
- It is possible to build chords with notes different from triads, specially if they respect to a recurrent logic or musical chain that justifies these different notes.
- Chords built by successive intervals different of thirds – like fourths and seconds – are recurrent in compositions with advanced tonalism or experimental character.
- The repetition of chords successions (or their characteristics) fixes a trajectory by means of the recurrence and make possible the introduction of exotic formations without incoherence.

B. Atonal and tonal harmonies, harmonic expansion and modulation

Omission of the basic chaining is the key to obtain modal and atonal harmonies. In case of absence of these minimal tonal structures, we say that the harmony is modal if the notes match with some diatonic scale (see equations 81) or if the notes are presented in a small number. If basic tonal chaining are absent, the notes do not match any diatonic scale and are diverse and dissonant (by relation with each other) enough to avoid reduction by polarization, the harmony is atonal. In this classification, the modal harmony are not tonal neither atonal. The modal harmony is reduced to the incidence of notes within the diatonic scale (or simplifications) and to the absence of tonal structures. It is possible to note, following this concept, that atonal harmony is hard to be realized**?** .

1. Atonal harmony

In fact, the techniques around atonal music comprehend structures for avoiding the hearing relation with modes and tonality. This difficulty imposes the use of dodecafonism. The purpose of dedocafonism is to use a set of notes (ideally 12 notes), and to perform each note, one by one, in the same order (no sei se ficou mt claro). In this context, the tonic becomes difficult to settle. Nevertheless, the Western hearing searches for tonal traces in musics and easily finds them by unexpected paths and even tortuous.

The use of dissonant intervals (specially tritones without resolution), together with ninths, seconds and sevenths, reinforces the absence of tonality. Considering this context, it is possible to account for the following considerations while creating a music piece:

- Repeating notes. Considering the immediate repetition as an extension of the previous incidence, the use of same notes in sequence does not add relevant information.
- It is interesting to play adjacent notes at the same time, making harmonic intervals and chords.
- Present durations with liberty, respecting the notes order.
- In order to gain variation, one can use extension, transposition, translation, inversion, retrograde and retrograde inversion. See subsections IV E and IV I for more details.
- Account for the presence of note structures, it is possible to use variations in orchestration, articulation, spatialization, among others.

The atonal harmony can be observed, paradigmatically, within these presented conditions. Most of them were written by great dodecafonic composers like Alban Berg and even Schoenberg. Many pieces create purposeful links between tonal and atonal techniques.

2. Tonal harmony

In the XX century, rhythmic music with emphasis on sonorities/timbres, extended the concepts about tonality and harmony. Nevertheless, the tonal harmony has strong influence in the artistic movements and commercial venues. In addition, the dodecafonism is considered having tonal nature because it was created to deny the tonal characteristics of polarization.

In tonal or modal music, the base of the harmonic filed are chords – like the ones listed in equations 85– built with a root note for each degree of a scale – like the ones in equation 81. In this context, music harmony aims at studying the observation of incident chord progressions and chaining rules. Even a monophonic melody generates harmonic fields, making it possible to observe the suggested chords at each passage.

In the 'traditional tonal harmony', a scale has its corresponding tonic (first scale degree) and can be major (with the same notes of the Ionian mode) or minor (notes from Eolian mode form the 'natural minor', which has both a harmonic and a melodic version, see equation 83). The scale is base on triads, each one with its root and representing a degree into the scale: $\hat{1}, \hat{2}, \hat{3}, \hat{4}, \hat{5}, \hat{6}, \hat{7}$. To build triads, the third and the fifth note above the root is considered together with the root (or fundamental). It is possible to write $\hat{1}, \hat{3}, \hat{5}$ as the first degree chord, built by the first degree of scale and the center for a tonal music. The chords of the fifth degree $\hat{5}, \hat{7}, \hat{2}$ ($\hat{7}$ sharp when a minor scale) and of the forth degree $\hat{4}, \hat{6}, \hat{1}$ are secondary. After that, other degrees are considered. The traditional harmony comprises conventions and stylistic techniques to create chord chains, made in each scale degree[?] .

The 'functional harmony' ascribes functions to these three central chords and tries to understand their use by means of these functions. The chord built under the first degree is named the **tonic** chord (T or t for a major or minor tonic, respectively) and its (role) function consisting of preserving a center, usually referred as a "ground" for the music. The chord built under the fifth degree is the **dominant** (D, the dominant is always major) and its function is call for the tonic (we say that the dominant chord asks for a conclusion and this conclusion is usually the tonic chord). Thus, the dominant chord guides the music to the tonic. The triad built under the forth degree is the **subdominant** (S or s for a major or minor subdominant, respectively) and its basically function is to distance the music from the tonic. The process aims at confirming the tonic using tonic-subdominant-tonic chains which are expanded by using other chords in different ways.

The remaining triads are associated to these three most important chords. In the major scale, the associated relative (relative tonic Tr, relative subdominant Sr and relative dominant Dr) is the triad built with a third below, and the associated counter-relative (counter-relative tonic Ta, counter-relative subdominant Sa and the counter-relative dominant Da) is the triad built with a third above. In the minor scale the same happens, but the triad with a third below is called counter-relative (tA, sA) and the triad with one third above is called relative (tR, sR). The precise functions and musical effects of these chords are controversial. Table II shows the relation between the triads built in each degree of the major scale.

TABLE II. Summary of tonal harmonic functions of the major scale. Tonic is the music center, the dominant goes to the tonic and the subdominant moves the music away from the tonic. The three chords can, at first, be freely replaced by their respective relative and counter-relatives.

relative	main chord of the function	counter-relative
$\hat{6}, \hat{1}, \hat{3}$	tonic: $\hat{1}, \hat{3}, \hat{5}$	$\hat{3}, \hat{5}, \hat{7}$
$\hat{3}, \hat{5}, \hat{7}$	dominant: $\hat{5}, \hat{7}, \hat{2}$	[$\hat{7}, \hat{2}, \hat{4}\#$]
$\hat{2}, \hat{4}, \hat{6}$	subdominant: $\hat{4}, \hat{6}, \hat{1}$	$\hat{6}, \hat{1}, \hat{3}$

The dominant counter-relative should form a minor chord. It explains the change in the forth degree by a semitone above $\hat{7}\#$. The diminished chord $\hat{7}, \hat{2}, \hat{4}$, is generally considered a 'dominant seventh chord with no root'[?] . In the minor mode, there is the change in $\hat{7}$ by an ascending semitone to make possible the existence of an unique semitone separating the tonic, and making also possible the dominant (that should be major and goes to the tonic). In this way, the dominant is always major, for both major and minor scale and, because of that, even as a minor tone, the relative dominant remains a third below and in the counter-relative, a third above.

3. Tonal expansion: individual functions and chromatic mediants

Each one of these chords can be confirmed and developed by performing their dominant or individual subdominant, which is the chord based in the third formed by a fifth above or by a fifth below, respectively. These dominants and individual subdominants, in the same way, have also subdominants and individual dominants that can be used. Given a tonality, any chord can occur, no matter how distant it is from the harmonic field and from the notes within the scale. The unique on condition is that the occurrence presents a coherent trajectory of dominants and subdominants to the original tonality.

Mediants, or 'chromatic mediants', are present in each chord in two ways: the upper chromatic mediant, formed by the root at the third of original chord; and the lower chromatic mediant, formed by the fifth at the third of original chord. Both chords also are formed by a third, but with a chromatic alteration regarding the original chord. If two chromatic alterations exist, i.e. two notes altered by one semitone each regarding the original chord, the mediant is called 'doubly chromatic mediant'. Again, there are two forms for each chord: the upper form, with a third in the fifth of the original chord; and the lower form, with a third in the root of the original triad. It is interesting to observe that a major chord has upper chromatic mediants and upper doubly chromatic mediants. A minor chord has lower chromatic mediants and lower doubly chromatic mediants. This relation between chords is considered as advanced tonalism, sometimes even considered as expansion and dissolution of tonalism, and, as consequence, it has strong and outstanding effects although perfectly consonant. The chromatic mediants are used since the end of Romanticism by Wagner, Lizt, Richard Strauss, among others, and are quite simple to be realized[?] [?] .

4. Modulation

Modulation is the change of the key (tonic, or tonal center) in a music. It can be characterized by observing the key at the start and at end of the chord transitions. Keys are always conceived as related by fifths and their relative and counter-relatives. Forms to perform modulation include:

- Transposing the discourse to a new key, without any preparation. It is a common Baroque procedure although incident at other periods as well. Sometimes it is called phrasal modulation or unprepared modulation.
- Using carefully an individual dominant and perhaps the individual subdominant, to confirm the change in key and harmonic field.
- Using chromatic alterations to reach a chord in the new key starting from a chord in the previous key. It is called chromatic modulation.
- Featuring a unique note, possibly repeated or suspended with no accompaniment, common to start and end a key, constitute a peculiar way to introduce the new harmonic field.
- Changing the function, without changing the notes, of a chord to contemplate a new key. This procedure is called enharmony.
- Maintaining the tonal center and changing the key quality from major to minor (or vice-verse) comprehends a parallel modulation. Key with same tonic and different quality is known as homonyms.

The dominant has great importance and is a natural pivot into modulations, which leads to the circle of fifths[?] [?] [?] [?] . The music piece *Acorde cedo* explores these chord relations. Its code is available both in Appendix **??** and online as part of MASSA[?] .

C. Counterpoint

Counterpoint is defined as the conduction of simultaneous melodic lines, called voices. The bibliography covers systematic ways to conduct voices, leading to scholastic genres like canons, inventions and fugues. It is possible to summarize the main counterpoint rules, and it is known that Beethoven – among others – outlined such similar synthesis.

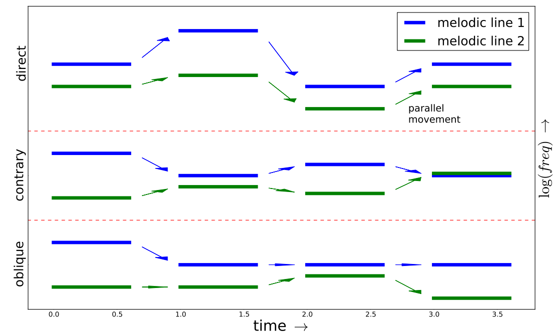

FIG. 23. Different motions of counterpoint aiming to preserve independence between voices. There are 3 types of motions: direct, contrary and oblique. The parallel motion is a type of direct motion.

The purpose of counterpoint is to conduct voices in a way they sound independent. In order to do that, the relative motion of voices (in pairs), is crucial and categorized as: direct, oblique and contrary motion (see Figure 23). The parallel motion is an oblique motion. The gold rule here is to take care with the direct motions, avoiding making them lead to a perfect consonance. The parallel motion should occur only between imperfect consonances and no more than three consecutive times. The dissonances can be unadmitted or used when followed and preceded by consonances of joint degrees, i.e. neighbor notes in a scale. The motions that lead a note to a neighbor sound more coherent. When having 3 or more voices, the melodic importance lies on the higher and lower voices, in this particular order[?] [?] [?] .

These rules were used in the music piece *Conta ponto*, the source code is in Appendix **??** and available online in MASSA.

D. Rhythm

Basically speaking, the rhythm notion is dependent on events separated by durations[?] . These events can be heard individually if spaced by at least $50 - 63ms$. To make the temporal separation between them be perceived as duration, the period should be large enough, around $100ms$[?] . It is possible to summarize the duration transitions heard as rhythm and pitch, in the following way[?] [?] .

The duration region marked as transition is presented minimized because the limits are not well defined. In fact, the duration where someone begins to perceive a fundamental frequency or a separation between occurrences, depends on the listener and sound characteristics[?] [?] .

The rhythmic metric is commonly based on a basic duration called pulse. The pulse typically comprehends durations between $0.25 - 1.5s$ (240 and $40BPM$, respectively). In music education and cognitively studies, it is common to associate these range of pulse frequencies to

TABLE III. Transition of durations heard individually until it turns into pitch.

	perception zone of duration as rhythm ←									transition		-					
duration (s)	**...**	**32,**	**16,**	**8,**	**4,**	**2,**	**1,**	**1/2,**	**1/4,**	**1/8,**	$\frac{1}{16} = 62,5ms$,	$\frac{1}{20} = 50ms$	1/40	1/80	1/160	1/320	1/640	...
frequency (Hz)	...	1/32,	1/16,	1/8,	1/4,	1/2,	1,	2,	4,	8,	16,	20	**40**	**80**	**160**	**320**	**640**	**...**
	-									transition		**perception zone of duration as pitch** →					

the durations of the heart beat, movements of inspiration and expiration and walk steps[?] [?] .

The pulse is subdivided into equal parts and repeated in sequence. These relations (division and concatenation) usually follow relations of small order integers[22]. A schematic exposition is shown in Figure 24.

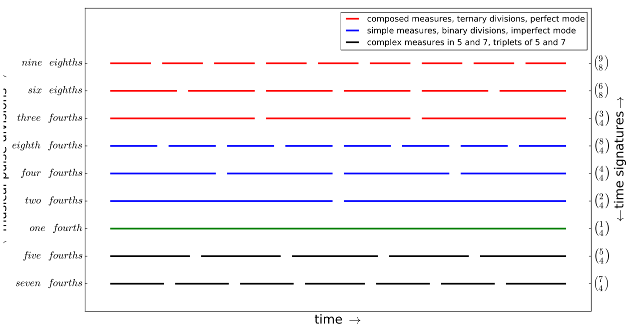

FIG. 24. Divisions and grouping in bars for the music pulse to establish a rhythmic metric. Divisions of the quarter note, established as the pulse, is presented in the left. The time signature whose specifies the same metrics but having as scale the groupings of the music pulse is presented in the right.

Dual relations (common time and binary divisions) are frequent in dance rhythms and celebrations. Ternary relations are common in ritualistic music and is related to the sacred. Dual relations are called imperfect, while and the ternary relations are called perfect.

The strong unities (accents) are the ones that fall in the 'head' of bars, the downbeats. The head of a unity is the first part of the subdivision. In binary divisions (2, 4 and 8, for example), the unities originally strong turn into weak (no sei se isso ficou claro) (e.g. division in 4 is: strong, weak, strong medium, weak). In ternary divisions (3, 6 and 9) two weak unities succeeds the downbeat (e.g. division in 3 is: strong, weak, weak)[23]

The accent in the weak beat is known as the backbeat, whereas a note starting in a weak beat which duration coming across the strong beat is known as syncope.

Notes can occur inside and outside of these *'music metric'* divisions. In most behaved cases, notes occur exactly on these divisions, with broader incidence on attacks of strong beats. In extreme cases, the time metric cannot be perceived[?] . Small grid variations help to build musical interpretation and differences between styles[?] .

Being pulse the grouping level $j = 0$, level $j = -1$ the first pulse subdivision, level $j = 1$ the first pulse agglomeration and so on. In this way, P_i^j is the i-th unit of pulses at grouping level j: P_{10}^0 is the tenth pulse, P_3^1 is the third pulse grouping unit (and, possibly, the third measure), P_2^{-1} is the second unit of pulse subdivision.

Special attention should be given to the limits of j: pulse divisions are durations sensible as rhythm. Furthermore, the pulse joints sums, at its maximum, a music or a cohesive set of musics. In other words, a duration given by $P_i^{min(j)}$, $\forall\ i$, is greater than $50ms$ and the durations summed together $\sum_{\forall i} P_i^{\max(j)}$ are less than a few minutes or at most, a few hours.

Each level j has some indexes i. A index i always has three different values (or multiple of three) that occur in a perfect relation. When i has only multiple of two, four or eight different values, than an imperfect relation occurs, as shown in Figure 24.

Any unit (note) of a given musical sequence which has a time metric can be unequivocally specified as:

$$P_{\{i_k\}}^{\{j_k\}} \tag{86}$$

where j_k is the grouping level andi_k is the unit order itself.

As example, consider $P_{3,2,2}^{-1,0,1}$ as the third subdivision P_3^{-1} of the second pulse P_2^0 and of the second pulse group P_2^1. Each unit or unit set P_i^j can be associated with a sequence of temporal samples T_i that forms a music note.

The music piece *Poli Hit Mia* explorers different metrics (avaleble in Appendix **??** and as part of MASSA).

E. Repetition and variation: motifs and large units

Given the basic music structures, either frequential (chords and scales) or rhythmic (beat divisions and simple, compound and complex grouping), it is natural to organize these structures in a cohesive way[?] . To this end, the concept of arcs are fundamental: we make an arc by departing from a place (music unit) and going to another one. The audition of melodic and harmonic lines is pervaded as musical arcs thanks to the cognitive nature of the musical hearing. It is possible to consider the note as the smaller arc, and each motif (definir motif aqui? nao foi falado ainda) and melody as another arc. Each time, beat subdivision, measure and music section constitutes an arc. A music in which the arcs do not present consistency by themselves, can be understood as a music with no cohesion. The coherence feeling comes, mostly, from the skilled manipulation of arcs in a music piece.

Musical arcs are abstract structures and liable of basic operations. An spectral arc, like a chord, can be inverted, enlarged and permuted, to cite some examples. Temporal arcs, like a melody, a motif, a measure or a note are also capable of variations. Remembering that $S_j = \left\{s_j = T_i^j = \{t_i^j\}_0^{\Lambda_j - 1}\right\}_0^{H-1}$ is a sequence of H musical events s_j, each event with its Λ_j samples t_i^j (refer to the beginning of this section IV), the basic techniques can be described as:

- Temporal translation is a displacement δ of a specific material to another instant $\Gamma' = \Gamma + \delta$ of the music. It is a variation that changes temporal localization in a music: $\left\{s_j'\right\} = \left\{s_j^{\Gamma'}\right\} = \left\{s_j^{\Gamma+\delta}\right\}$ where Γ is the duration between the beginning of the piece (or considered period) and the first event s_0 of original structure S_j. Observe that δ is the time offset of the displacement.

- Temporal expansion or contraction is a change in duration of each arc by a factor μ : $s_j'^{\Delta} = s_j^{\mu_j . \Delta}$. Possibly, $\mu_j = \mu$ is constant.

- Temporal reversion consists of generating a sequence with elements in reverse order of the original sequence S_j, thus: $S_j' = \left\{s_j'\right\}_0^{H-1} = \left\{s_{(H-j-1)}\right\}_0^{H-1}$.

- Pitch translation is a displacement τ of a material for a different pitch $\Xi' = \Xi + \tau$. It is a variation that changes pitch localization of material: $\left\{s_j'\right\} = \left\{s_j^{\Xi'}\right\} = \left\{s_j^{\Xi+\tau}\right\}$ where Ξ is the pitch of a period S_j or of an event s_0 of the original structure S_j. Observe that τ is the pitch offset of the displacement. If τ is given in semitones, the displacement in frequency is $\tau_f = f_0 . 2^{\frac{\tau}{12}}$ where f_0 results from the adoption of some reference: $f_0 = \Xi_{f_0} Hz \sim \Xi_0$ absolute value of pitch. For the frequency of any pitch value: $\Xi_f = \Xi_{f_0} . 2^{\frac{\Xi - \Xi_0}{12}}$. For the pitch of any frequency value: $\Xi = \Xi_0 + 12. \log_2 \left(\frac{\Xi_f}{\Xi_{f_0}}\right)$. In the MIDI protocol, $\Xi_{f_0} = 55Hz$ while pitch $\Xi_0 = 33$ marks a *A 1*, another reference point is $\Xi_{f_0} = 440Hz$ which corresponds to $\Xi_0 = 69$. The unit difference is equal for semitones. It is not possible to say that the unit is the semitone because $\Xi = 1$ is not a semitone, but a note with an audible frequency as rhythm, with less than 9 occurrences each second (see Table III).

- Interval inversion is the inversion traversed by the material. The inversion is strict if the number of semitones is being used as a reference for the operation: $S_j' = \{s_j'\}_0^{H-1} = \left\{s_j^{-\varepsilon_j . f_0}\right\}$, where ε_j is the factor between the event frequency s_j and the frequency of s_0. The inversion is said tonal if the distances are considered in terms of degree number of the scale E_k: $S_j' = \{s_j'\}_0^{H-1} = \left\{s_j^{-\varepsilon^{\left(e_{(j_e)}\right)} . f_0}\right\}_0^{H-1}$ where $j_e = e_j^{inv}$ is the index $k = j_e$ in E_k of the note of event s_j.

- Rotation of musical elements is the transference of the last element to the position of the first element, and the displacement of that to the penult, one position ahead. It is possible to define rotation of $\tilde{n}$ positions by $S_n' = S_{(n+\tilde{n})\%H}$. If $\tilde{n} < 0$, it is enough to use $\tilde{n}' = H - \tilde{n}$. It is reasonable to associate $\tilde{n} > 0$ with the clockwise rotation and $\tilde{n} < 0$ with tne anti clockwise rotation. More information about rotations in presented in subsection IV G.

- The insertion and removal of materials into structure S_j can be ornamental or structural: $S_j' = \{s_j'\} = \{s_j \text{ if condition A, otherwise } r_j\}$, for any music material r_j, including the empty instant. Elements of R_j can be inserted at the beginning, like a prefix in S_j; at the end, as a suffix; or at the middle, dividing S_j or making it the prefix or suffix. Both materials can be mixed in a variety of ways.

- Changes in articulation, orchestration and spatialization, $s_j' = s_j^{*_j}$, where $*_j$ is the new characteristic incorporated by the element s_j'.

- Accompaniment. Both orchestration and melodic lines presented when S_j occurs can suffer modifications and be considered as a variation of S_j itself.

Other proceesings can be derived together with the presented ones, as, for example, the retrograde inversion, temporal contraction with an external suffix, etc. The music structures resound in the human cognitive system due to the own nature of thought. In its many veneers, an idea read with the same number of elements and connective aspects between them. The music, when tunned with those mental structures, raises impressions. In this way, a whole process of mental and neurological resonance is unleashed, responsible by the feelings, memories and imaginations, typical to a mindful music listening. This cortical activity helps the musical therapy, known by its utility in cases of depression and neurological injury. It is known that regions of the human brain responsible by hearing processing, are also used for other activities, even linguistic and mathematical[?, ?] .

The paradigmatic structures guide the creation of new music material. One of the most central of them is the dipole tension/relax. Traditional dipoles are relate to tonic/dominant, repetition/variation, consonance/dissonance, coherence/rupture, symmetry/asymmetry, equality/difference, arrival/departure, near/far, stationary/moving, etc. The ternary relations tend to relate to the circular and unification. The lucid ternary communion, 'modus perfectus', opposes to the passionate dichotomic, 'modus imperfectus'. Hereafter, there is an exposition dedicated to directional and cyclic arcs.

F. Directional structures

The arcs can be decomposed in two convergent sequences: the first arc reaches the apex and the second arc returns from the apex to the start region. This apex is called climax by traditional music theory. It is possible to differ between arcs whose climax are placed at the beginning, middle, end, and at the first and second half of the considered duration. These structures are shown in Figure 25. The parameter of variation does not exist, thus the arc consists only of the reference structure[?] .

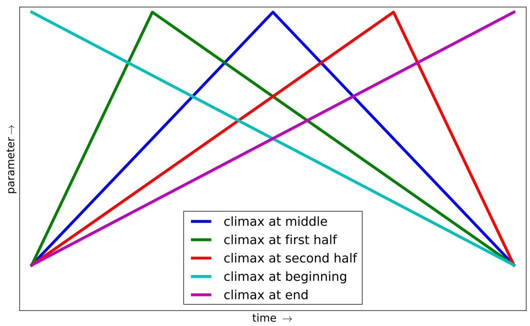

FIG. 25. Canonical distinctions of musical climax in a given melody and other domains. The considered possibilities are: climax at beginning, climax at first half, climax at middle, climax at second half and climax at the end. The y-axis is not properly specified since there exist no real parametric variation and, therefore, the structure is a reference.

Consider $S_i = \{s_i\}_0^{H-1}$ as an increasing sequence. The sequence $R_i = \{r_i\}_0^{2H-2} = \left\{s_{(H-1-|H-1-i|)}\right\}_0^{2H-2}$ is a sequence that presents a perfect profiteer symmetry, i.e. the second half is the mirrored version of the former. Following the musical concepts, the climax is localized in the middle of the sequence. It is possible to modify this fact using sequences with different sizes. All the mathematical theory of sequences, already well established and taught routinely in calculus courses, can be used to generate those arcs[? ?] . Theoretically, when applied to any characteristic of music events, these sequences produce arcs, since they imply the deviation and return of an initial parametrization. Henceforth, it is possible to have a number of distinct arcs, with different sizes and climax for a same sequence of events. This is an interesting and useful resource, and the correlation of arcs results in coherence[?] .

Historically (and practically nowadays), the golden ratio and the Fibonacci sequence have special importance. The Lucas sequence allows the generalization of Fibonacci sequence, making it easy to understand. Given any two numbers x_0 and x_1, the Lucas sequence can be obtained as: $x_n = x_{n-1} + x_{n-2}$. The greater n is, the more $\frac{x_n}{x_{n_1}}$ approaches the golden ratio: 1.61803398875.... The sequence converges fast even with high discrepant initial values. Being $x_0 = 1$ and $x_1 = 100$, and $y_n = \frac{x_n}{x_{n+1}}$ an auxiliary sequence, the error occurring in the first values with relation to the golden ratio is, approximately, $\{e_n\} = \left\{100\frac{y_n}{1.61803398875} - 100\right\}_1^{10} = \{6080.33, -37.57, 23, -7.14, 2.937, -1.09, 0.42, -0.1601, 0.06125, -0.02338\}$. The Fibonacci sequence presents about the same error progression but starts right at the second step since $\frac{1}{1} \approx \frac{101}{100}$.

The music piece *Dirracional* exposes the use of arcs into directional structures. Its source code is included in Appendix **??** and available online as part of MASSA[?] .

G. Cyclic structures

The philosophical understanding that the human thought is based on the perception of similarities, differences, stimuli and objects, places the symmetries at the core of the cognitive process[27]. Mathematically, symmetries are algebraic finite groups, and a finite group is always isomorphic to a permutation group. It is possible to say that permutations represent any symmetry in a finite system (precisa de uma referncia aqui). In music, permutations are ubiquitous and considerably present in techniques, which confirms their central role. The successive application of permutations generates cyclic arcs[? ? ?] . Two academic studies were dedicated to this end, aiming at generating music structures[? ?] .

Any permutation set can be used as a generator of algebraic groups[?] . The properties defining a group G are:

$$\begin{aligned}
\forall \;\; p_1, p_2 \in G \Rightarrow \quad & p_1 \bullet p_2 = p_3 \in G \\
& \text{(closure property)} \\
\forall \;\; p_1, p_2, p_3 \in G \Rightarrow \quad & (p_1 \bullet p_2) \bullet p_3 = p_1 \bullet (p_2 \bullet p_3) \\
& \text{(associativity property)} \\
\exists \;\; e \in G: \;\; & p \bullet e = e \bullet p \quad \forall p \in G \\
& \text{(identity element property)} \\
\forall \;\; p \in G, \; \exists \, p^{-1}: \quad & p \bullet p^{-1} = p^{-1} \bullet p = e \\
& \text{(inverse element property)}
\end{aligned} \tag{87}$$

From the first property it is possible to conclude that any permutation can be operated with another permutation. In fact, it is possible to apply a permutation p_1 and another permutation p_2, and, comparing both initial and final orderings, results in another permutation p_3.

Every element p operated with itself a sufficient number of times n reaches the identity element $p^n = e$ (otherwise the group would be infinite, generated by p). The lower $n \;:\; p^n = e$ is called the element order. Thus, a finite permutation p, successively applied, reaches the initial ordering of its elements, making a cycle. This cycle, if used as parameter of music notes, implies in a cyclic musical arc.

These arcs can be established by using a set of different permutations. As a historic example, the tradition called *change ringing* conceives the music through bells played one after other and then played again, but in a different order. This process repeats until it reaches the initial ordering. The set of different traversed orderings is a *peal*. Table IV represents one traditional *peal* for 3 bells (1, 2 and 3) which explores all possible orderings. Each line indicates one bell ordering to be played. Permutations occur between each line. In this case, the music structure consists of permutations by itself and some different permutations operate due to the cyclic behavior.

TABLE IV. Change Ringing: Standard *peal* for 3 bells. Permutations occur between each line. Each line is a bell ordering and each ordering is played a line a time.

1	2	3
2	1	3
2	3	1
3	2	1
3	1	2
1	3	2
1	2	3

The use of permutations in music can be summarized in the following way: consider $S_i = \{s_i\}$ as a sequence of musical events s_i (e.g. notes), and a permutation p. $S_i' = p(S_i)$ comprises the same elements of S_i but in a different order. The permutations can be written based on two notations: cyclic or natural. The natural notation basically comprehends the order of the resulting indexes from the permutation. Thus, agreed the original ordering given the sequence of its indexes [0 1 2 3 4 5 ...], the permutation is noted by the sequence it produces (ex. [1 3 7 0 ...]). In the cyclic notation, permutation is the changing of one element by the posterior one, and the last element by the first one.

It is not necessary to permute elements of S_i, but only some characteristics. In this way, being p^f a permutation in frequency and S_i a sequence of basic notes as exposed in the end of section II F, the new sequence $S_i' = p^f(S_i) = \left\{ s_i^{p(f)} \right\}$ consists of the same music notes, respecting same order and maintaining the same characteristics, but with the fundamental frequencies permuted following the pattern that p^f presents.

It is worthwhile to mention two subtleties of this procedure. First, the permutation p does not have to involve all elements of S_i, i.e. it can operate in subsets of S_i. Second, the elements s_i do not have to be entirely executed at each state access. To exemplify, consider S_i as a sequence of music notes s_i. If i goes from 0 to n, and $n > 4$, at each measure of 4 notes it is possible to execute the first 4 notes. The other notes of S_i can occur in measures where permutations set aside those notes to the first four notes of S_i.

Each one of the permutations p_i described above relates the note dimensions where it operates (frequency, duration, *fades*, intensity, etc) and the period of incidence (at how much access a permutation is applied). During realization of notes in S_i, an easy and coherent form is to execute the first n notes[24].

Appendix **??** and MASSApresent the computational implementation about permutations[? ? ?] .

H. Musical idiom?

There are many studies intending to model and explore a 'musical language', or 'linguistics applied to music' or even to discern between different 'musical idioms'[? ? ? ?] . In a simple way, a musical idiom is the result of a chosen material together with repetition of elements and repetition of relations between elements present along the music. The related dichotomies are outstanding, as explained at subsection IV E: repetition and variation, relax and tension, equilibrium and instability, consonance and dissonance, etc.

I. Musical usages

The basic note was defined and characterized in quantitative terms (section **??**). Next, the internal note composition was addressed and both internal transitions and immediate treatment were understood (section **??**). Finally, this section aims at organizing these notes in music. The gamma of resources and consequent infinitude of possibilities is a typical situation and highly relevant for art contexts[1?] .

There are studies dedicated to each one of the presented resource. For example, it is possible to obtain the 'dirty' triadic harmonies (with notes without the triad) by superpositions of dirty fourths. Another interesting example is the simultaneous presence of rhythms in different metrics, constituting what is called *polyrhythms*. The music piece *Poli-hit mia* explores these simultaneous metrics by impulse train (??) convolved with notes that compound the line. Its source code is included in Appendix **??** and available online as part of MASSA.

The microtonal scales are important for the music of the 20th century[?] and present diverse outstanding results, like the fourths of tone in the Indian music. The musical sequence *MicroTom* explores these resources, including microtonal melodies and microtonal harmonies with many notes in a very reduced note scope. Its code is presented in Appendix **??** and available online in MASSA.

As pointed in subsection III F, the relations between parameters are powerful ways to acquire music pieces. The number of permuted notes can vary during the music, revealing relationship with piece duration. Harmonies can be made of triads (eqs. 85) with replicated notes at each octaves and more numerous as minor the depth and frequency of vibratos (eqs. 55, 56, 57, 58, 59), among other uncountable possibilities.

The presented symmetries at octave divisions (eqs. 80) and the symmetries presented as permutations (Table IV and eqs. 87) can be used together. In the music piece *3* *trios* this association is done in a systematic way in order to enable the required listening. This is a instrumental piece, not included as a source code[?] .

The *PPEPPS* (Pure Python EP: Projeto Solvente) is an EP synthesized using the resources presented in this study. With minimal parametrization, the program generates complete musics, allowing the easy composition of musics and sets of musics. Using few lines of code it is possible to obtain a directory with the generated musics. This facility and technological demystification create aesthetically possibilities for both sharing and education.

V. CONCLUSIONS AND FUTURE WORKS

We have described a concise system that explores the musical elements and relate them to the digital signal. The reader is invited to access *Scripts* and MASSA, where these relations are computationally implemented. We hope the didactic report along the paper and the supplied scripts facilitates and encourages the use of the proposed framework.

The open possibilities provided by the techniques and results discussed in this paper involve psychoacoustic experiments and the creation of interfaces for the generation of music, noise and other music applications in a high fidelity (*hi-fi*). It is worthwhile to mention the benefit of these results for artistic and didactic purposes. The incorporation of programming skills is facilitated by the provided visual aids. Initial *live-coding* practices and guided courses based on this work have already been realized with successful. Examples are Puredata and Chuck.

This work systematically investigates the parameterization issues (like the tremolo, ADSR, etc.) in a high fidelity, which has significant artistic utility. Such detailed analytical descriptions, together with the computational implementations, have not been covered before in the literature (as shown in Appendix **??**).

Besides, the free software license and online availability of the exposed content as hypertext, in conjunction with the respective codes and sound examples, strongly facilitates future collaborations and the generation of subproducts in a co-authorship fashion. As consequence, the expansion of MASSAis feasible and straightforward, making it possible new developments and implementations of musical context.

In addition, this work permitted the formation of groups with common interests, like music creativity and computer music. Specially, the project `labMacambira.sf.net` groups Brazilian and foreign co-workers in order to offer relevant contributions in diverse areas like Digital Direct Democracy (no devo ter traduzido certo), georeferencing techniques, artistic and educational activities. Some of these reports are available in Appendix **??** and made online. There are more than 700 videos, written documentations, original softwares and contributions in well-known external softwares such as Firefox, Scilab, LibreOffice, GEM/Puredata, to name a few[? ? ?] .

Future works include the application of the reported results in machine learning and artificial intelligence methods for the generation of appealing artistic materials.

[1] Anton Webern, *The Path To The New Music* (Theodore Presser Company, 1963).

[2] William Lovelock, *A Concise History of Music* (Hammond Textbooks, 1962).

[3] Timbre is a subjective and complex characteristic. Physically, the timbre is multidimensional and given by the temporal dynamics behavior of energy in the spectral components that are harmonic or noisy. Beyond that, the word timbre is used to designate different things: one same note has different timbres, a same instrument has different timbres, two instruments of the same family have, at the same time, the same timbre that blends them in the same family, and different timbres as they are different instruments. It is worth to mention that timbre is not always manifested in spectral traces, since cultural or circumstantial aspects alter our perception of timbre.

[4] It is important to note that the factor $\frac{1}{\Lambda}$ could be distributed among the Fourier transform and its reconstruction, as preferred.

[5] Also called *bias* or *offset*.

[6] Any sampled signal has this property, not only the digitalized sound.

[7] Equal real part and imaginary with inverse order: $a_{k1} = a_{k2}$ and $b_{k1} = -b_{k2}$. As consequence the modules are equal and phases have inverse order.

[8] As pointed before, this limitation simplify the explanation without losing generality, and will be overcome in the next section.

[9] It is known that $\zeta \approx 21,5cm$ to an adult human.

[10] The gold tip here is: to make a sound that incites interest by itself, do internal variations on it[?] .

[11] In other words, a geometric progression of frequency is perceived as an arithmetic progression of pitch.

[12] The volume change (psychophisical quality) is consequence of different sound characteristics like reverberation and concentration of treble harmonics, among which is wave energy. Wave energy is the easiest one to modify (see Equation 2) and it can vary in many ways. The most simple way consists of modifying the amplitude by multiplying the whole sequence by a real number. The increased energy without amplitude variation is the *sound compression*, useful nowadays for music production[?] .

[13] The implementation of filters comprehends an area of recognized complexity with dedicated literature and software implementations. We recommended the reader to consult the bibliography[? ?] .

[14] It is possible to apply the filter in the frequency domain multiplying the Fourier coefficients of both sound and the impulse response, and then performing the inverse Fourier transform in the resulting spectrum.[?] .

[15] Short for 'biquadratic': its transfer function has two poles and two zeros, i.e. its first direct form consists of two quadratic polynomials forming the fraction: $\mathbb{H}(z) = \frac{a_0+a_1.z^{-1}+a_2.x^{-2}}{1-b_1.z^{-1}-b_2.z^{-2}}$.

[16] Butterworth and Elliptical filters can be considered as specific cases of Chebichev filters.[? ?] .

[17] Attention with the cutoff frequency f_c in low-pass and high-pass filters.

[18] Although its origin is disparate given its association with the brown color, this noise become established with this specific name. Anyway, this association is satisfactory once the white and pink noises are more strident and associated with intense colors[? ?] .

[19] Some musical instruments and contexts use different terms. For example, in piano, the called tremolo is a vibrato and a tremolo is the classification used here. The presented definitions are common in contexts regarding music theory and electronic music. In addition, they are based on a broader literature than the one used for an specific instrument, practice or musical tradition[? ?] .

[20]The pitch deviation is called 'vibrato depth' and is generally given as semitones or cents, as convenience.

[21]Tremolos and vibratos occur many times together in a traditional music instrument and voices.

[22]Details in the next section.

[23]It is important to remember that the convolution with a impulse results in a sound shifted to the moment that the impulse occurred.

[24]Micro-tonality is the use of intervals less than one semitone and have ornamental and structuring functionalities in music. The division of the octave in 12 notes is motivated by physical fundamentals but it is also a *convention* adopted even by western classical music. Other tunings are present. To cite one example, the traditional Thai music uses a octave division in seven notes equally spaced ($\varepsilon = 2^{\frac{1}{7}}$), resulting in intervals that have little in common with the intervals in the diatonic scale[?] .

[25]In ascending order of occurrence in written and ethnic music, these indicate the musical pulse divisions and their sequential grouping at time: 2, 4 and 8, after that 3, 6 (two groups of 3 or 3 groups of 2) and 9 and 12 (three and 4 groups of 3). At last, the prime numbers 5 and 7, complementing 1-9 and 12. Other metrics are less common, like divisions and grouping in 13, 17, etc, and are mainly used in contexts of experimental music and classical music of XX and XXI. No matter how complex they seem, metrics are commonly compositions and decompositions of 1-9 equal parts[?,?] .

[26]Division in 6 is considered compound but can also occur as a binary division. A binary division which suffers a ternary division results in two units divided in tree units each: strong (subdivided in strong, weak, weak) and weak (subdivided in strong, weak, weak). Another way to perform the division in 6 is applying ternary division that suffers a binary division, resulting in: a strong unit (subdivided in strong and weak) and two weak units (subdivided in strong and weak for each).

[27]G. Deleuze, *Difference and Repetition* (Continuum, 1968).

[28]The execution of disjoint notes of S_i equals to modify the permutation and to execute the first notes.

[29]F. de Saussure, *Course in General Linguistics* (Books LLC, 1916).

[30]A. Papoulis S. U. Pillai, *Probability, Random Variables and Stochastic Processes* (McGraw Hill Higher Education, 2002).

[31]R. A. Johnson D. W. Wichern, *Applied Multivariate Statistical Analysis* (Prentice Hall, 2007).

[32]C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing* (Prentice Hall, 1992).

[33]R. O. Duda P. E. Hart D. G. Stork, *Pattern Classification* (Wiley-Interscience, 2000).

[34]L. da F. Costa R. M. C. Jr., *Shape Analysis and Classification: Theory and Practice (Image Processing Series)* (CRC Press, 2000).

[35]D. Papineau, *Philosophy* (Oxford University Press, 2009).

[36]Bertrand Russel, *A History of Western Philosophy* (Simon and Schuster Touchstone, 1967).

[37]F. G. G. Deleuze, *What Is Philosophy?* (Simon and Schuster Touchstone, 1991).

[38]Alan Walker, *Hans von Bülow: a life and times* (Oxford University Press, 2010).

[39]Roy Bennett, *History of Music* (Cambridge University Press, 1982).

[40]Aaron Kozbelt, "Performance time productivity and versatility estimates for 102 classical composers," Psychology of Music **37**, 25–46 (2009).

[41]Aaron Kozbelt, "A quantitative analysis of Beethoven as self-critic: implications for psychological theories of musical creativity," Psychology of Music **35**, 144–168 (2007).

[42]Dean Keith Simonton, "Emergence and realization of genius: The lives and works of 120 classical composers," Journal of Personality and Social Psychology **61**, 829 – 840 (1991), ISSN 0022-3514.

[43]Dean K. Simonton, "Creative productivity, age, and stress: A biographical time-series analysis of 10 classical composers," Journal of Personality and Social Psychology **35**, 791 – 804 (1977), ISSN 0022-3514.

[44]Peter van Kranenburg, "Musical style recognition – a quantitative approach," in *Proceedings of the Conference on Interdisciplinary Musicology (CIM04)* (2004).

[45]Peter van Kranenburg, "On measuring musical style – the case of some disputed organ fugues in the J.S. Bach (BWV) catalogue," Computing In Musicology **15** (2007-8).

[46]Joe H Ward Jr, "Hierarchical grouping to optimize an objective function," Journal of the American Statistical Association **58**, 236–244 (1963).

[47]Hal Varian, "Bootstrap tutorial," The Mathematica Journal **9** (2005).

[48]Christian P. Robert, "Simulation in statistics," in *Proceedings of the 2011 Winter Simulation Conference* (2011) arXiv:1105.4823.

[49]J. Cohen *et al.*, "A coefficient of agreement for nominal scales," Educational and psychological measurement **20**, 37–46 (1960).

[50]D.D. Boos, "Introduction to the bootstrap world," Statistical Science **18**, 168–174 (2003).