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

Renato Fabbri,[1, a] Vilson Vieira da Silva Junior,[b] Antônio Carlos Silvano Pessotti,[c] Débora Cristina Corrêa,[d] and Osvaldo N. Oliveira Jr.[e]
Instituto de Física de São Carlos, Universidade de São Paulo (IFSC/USP)

(Dated: 24 April 2017)

Notes, ornaments and intervals are examples of basic musical elements. Their representation as discrete-time digital audio 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 sonic samples. Such a compendium enables scientific experiments in precise and trustful ways, among educational and artistic uses. This paper presents a comprehensive description of music in digital audio, within an unified approach. 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 to operate within a note. Moreover, the generation of musical structures such as rhythmic meter, pitch intervals and cycles, are attained canonically with notes. The availability of these resources in scripts is provided in public domain within the MASSA toolbox - Music and Audio in Sequences and Samples. 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 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, where sound corresponds to the longitudinal wave of mechanical pressure. The human hearing system perceives sounds within the frequency bandwidth between $20Hz$ and $20kHz$, with the actual limits depending on the person, climate conditions and the sonic characteristics themselves[1]. Since the speed of sound is $\approx 343.2m/s$, these limits imply wavelengths of $\frac{343.2}{20} = 17.16\,m$ and $\frac{343.2}{20000} = 17.16\,mm$. Such perception involves stimuli in bones, stomach, ears, transfer functions of head and torso, and processing by the nervous system[1]. Obviously, the ear is a dedicated organ for the appreciation of these waves, which decomposes them into their sinusoidal spectra 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, discussed 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 referred to as audio, although these terms are often used without distinction. Audio expresses waves from the capture by microphones or from direct synthesis, although these sources are not neatly distinguishable as captured sounds are processed to generate new sonorities. Digital audio specified by protocols that facilitate file transferring and storage often implies 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 representing the audio with equally spaced samples, of 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 the bit depth increases.

The Nyquist theorem[2] 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, $f_s \geq 2 \times$

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

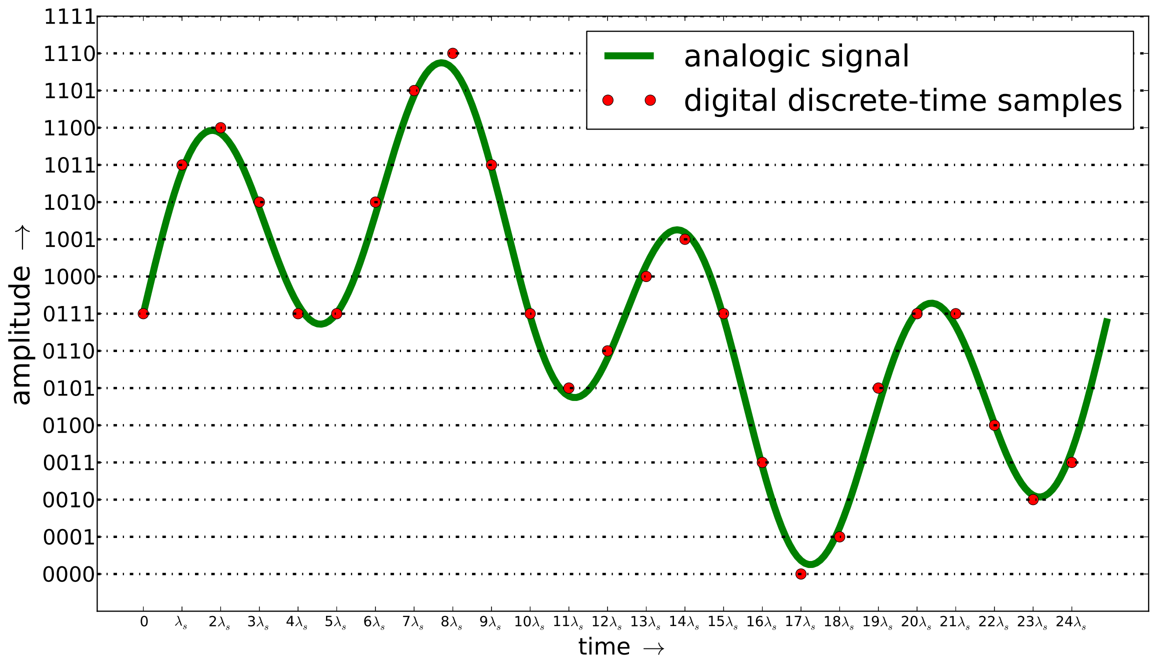

FIG. 1. Pulse Code Modulation (PCM) audio: an analogical signal is represented by 25 samples with 4 bits each.

$20kHz = 40kHz$. This is the fundamental reason for the adoption of sampling frequencies such as $44.1kHz$ 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 IV) 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[4]. 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 digitalized 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[5]. 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[6–9]. 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[10,11].

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; section III further develops internals of the musical note; section IV tackles the organization of musical notes in higher levels of musical structure[12–16]. As these descriptions embody topics such as psycho-acoustics, cultural traditions, formalisms and protocols, the text points to external complements as needed[17–19].

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 comprising fundamental units referred to as notes, "atoms" that constitute music itself[12,13,20]. These notes are now understood as central elements of certain musical paradigms. In a cognitive perspective, the notes are seen as discretized elements that facilitate and enrich the flow of information through music[1,16]. Canonically, the principal properties of a musical note are duration, volume, pitch and timbre[16], which can be quantified using the evenly time spaced sonic samples.

All the relations described in this section are implemented in the file *eqs2.1.py* of the MASSA toolbox. The musical pieces *5 sonic portraits* and *reduced-fi* are also available online 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:

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

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 in section III. One can get volume variations through the power of the wave[21]:

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

The final volume is dependent on the speakers amplification of the signal. Thus, what matters is the relative power of a note in relation to the other ones around it, or the power 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 ($10dB$):

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

An amplitude increase by a factor slightly above of 3 is required for yielding a doubled volume. These values are guides for increasing or decreasing the absolute values in 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 amplification.

C. Pitch

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 in the number of samples per cycle: $\lambda_{f_0} = f_s.\delta_{f_0} = f_a/f_0$.

For didactic reasons, let f_0 be 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 do not divide f_s will be considered. This restriction does not imply in loss of generality of this current section's content.

D. Timbre

In a sound with harmonic spectrum, the (wave) period corresponds to the cycle duration, 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[22]. Sonic spectra with minimum differences can result in timbres with crucial differences and, consequently, distinct timbres can be produced using different spectra[1].

The simplest case is the spectrum with only the fundamental frequency f, which is a sinusoid as in "simple harmonic oscillation". Let S_i^f be a sequence whose samples s_i^f describe 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, 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[2].

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 has 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 preserves 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 at 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 especially 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 accurately detected, particularly at low frequencies. It requires a large 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[1].

Figure 2 presents the waveforms described in equations 9, 10, 11 and 12 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 spectrum is a consequence of the use of a fixed period. The sinusoid consists of a unique node in the spectrum, pure frequency. The figure exhibits the spectra described: the sawtooth is the only waveform with a complete harmonic series (odd and even components); triangular and

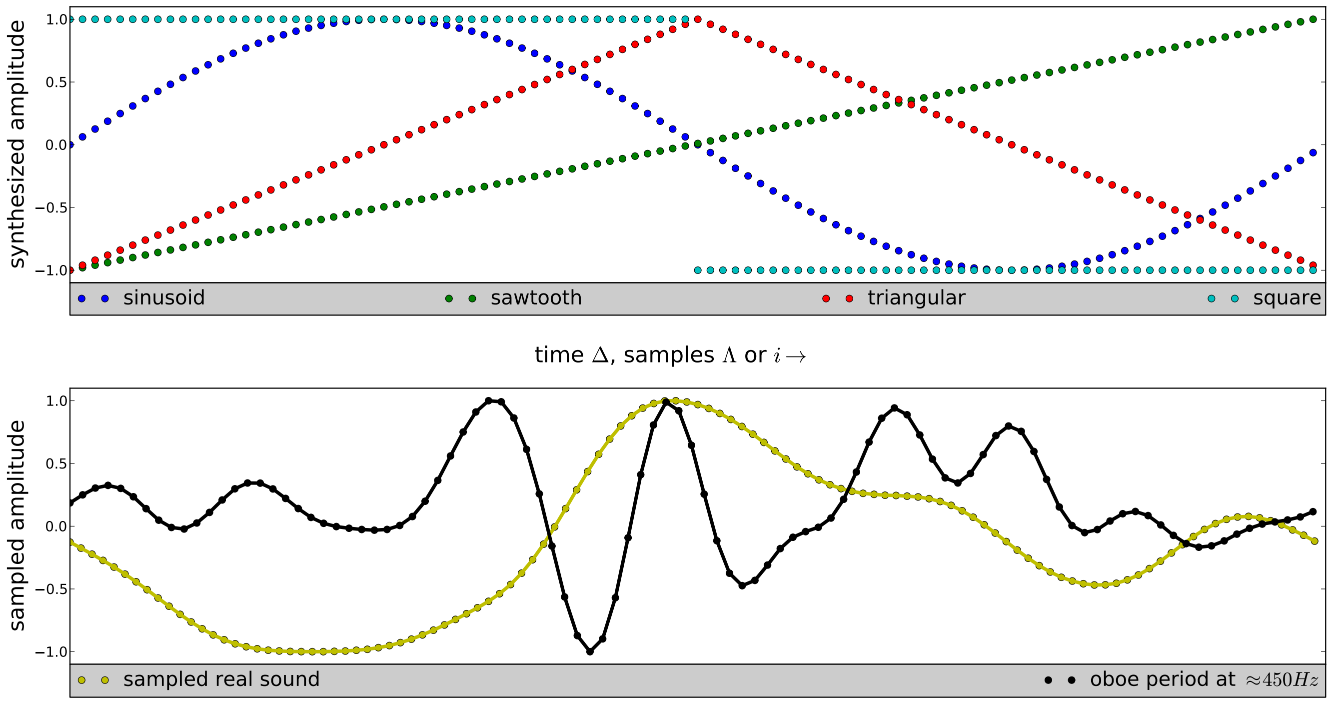

FIG. 2. Basic musical waveforms: (a) the synthetic waveforms; (b) the real waveforms.

square waves have 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 several mechanisms similar to the human cochlea[1]).

The sinusoidal components are usually responsible 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[1]. For a fixed length period and waveform, the spectrum is perfectly harmonic and static, and each waveform is comprised of specific proportions of harmonic components. High curvatures are a 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 at $44.1kHz$. The chosen period for sampling was relatively short, with 98 samples, and corresponds to the frequency $\frac{44100}{98} = 450Hz$, which is associated with a slightly out-of-tune A4 pitch. One can notice from the curvatures: the oboe's rich spectrum at 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 spectrum of the original waveform. As a consequence of its repetition in an identical form, the spectrum is perfectly harmonic, without noise or 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, whose samples consists of the repetition of cycle of Figure 2. Summing up, the natural spectrum exhibits variations in the frequencies of the harmonics, in their intensities and some noise, while the note made from the sampled period has a perfectly harmonic spectrum.

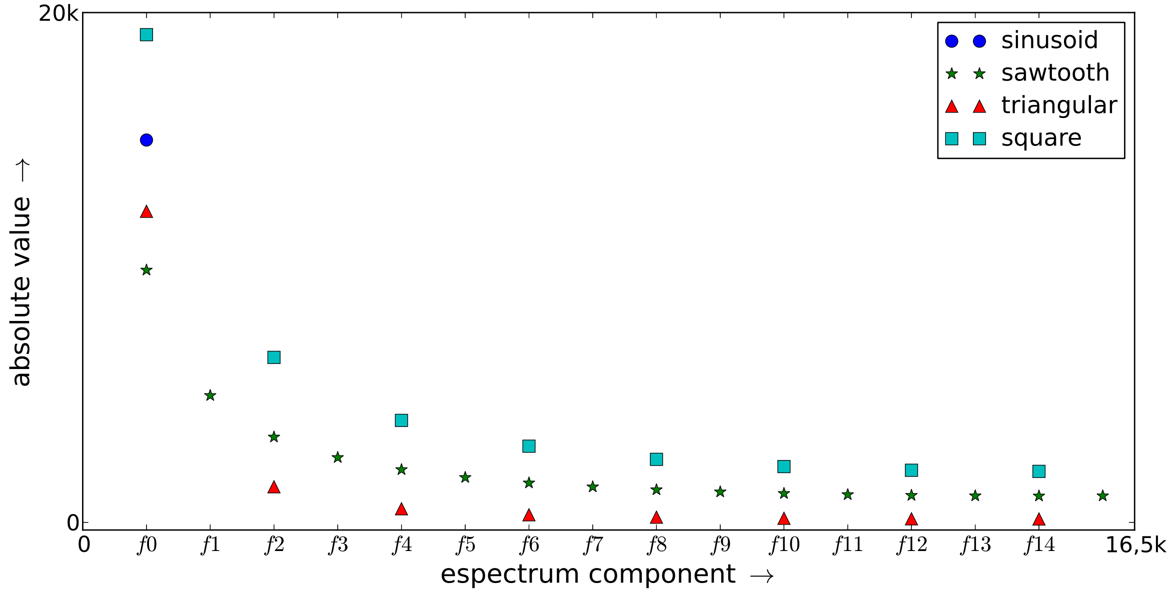

FIG. 3. Spectrum of basic artificial musical waveforms.

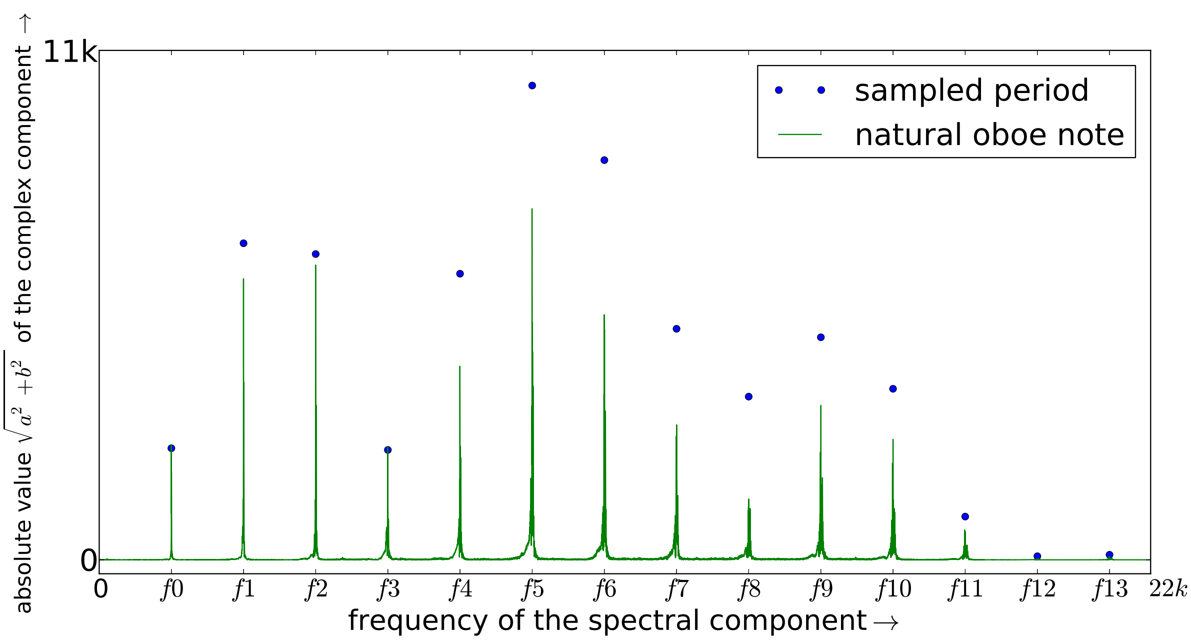

FIG. 4. Spectrum of the sonic waves of a natural oboe note and from a sampled period. The natural sound has fluctuations in the harmonics and in its noise, while the sampled period note has a perfectly harmonic spectrum.

E. Spectrum at sampled sound

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[23]:

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

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. 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 enable 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 specify 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 their spectral components, where 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[24] and $\frac{a_1}{2}$, oscillation amplitude with frequency given by $f_k = k\frac{f_a}{\Lambda=2}$. This case has special relevance. At least 2 samples are necessary 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[25].

All fixed sequences T_i of only 3 samples also have just 1 frequency, since the first harmonic would have 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 reconstructing a 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 the 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. Recalling 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 applies to $w_k = f_k.\frac{2\pi}{f_a}$ 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 one component at frequency f_0 or $f_{\Lambda/2}$ (if Λ is even). This assertion can be verified with $k = 0$ and $k = \Lambda/2$ like this: $f_{\Lambda/2} = f_{(\Lambda-\Lambda/2)=\Lambda/2}$ and $f_0 = f_{(\Lambda-0)=\Lambda} = f_0$. Furthermore, these two frequencies (zero and Nyquist frequency) do not have phase variation, their coefficients being strictly real. Therefore, the number τ of equivalent coefficient pairs is:

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

This discussion makes it clear the equivalence 17, 18 and 19:

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

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

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

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

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

To discuss 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 phase 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

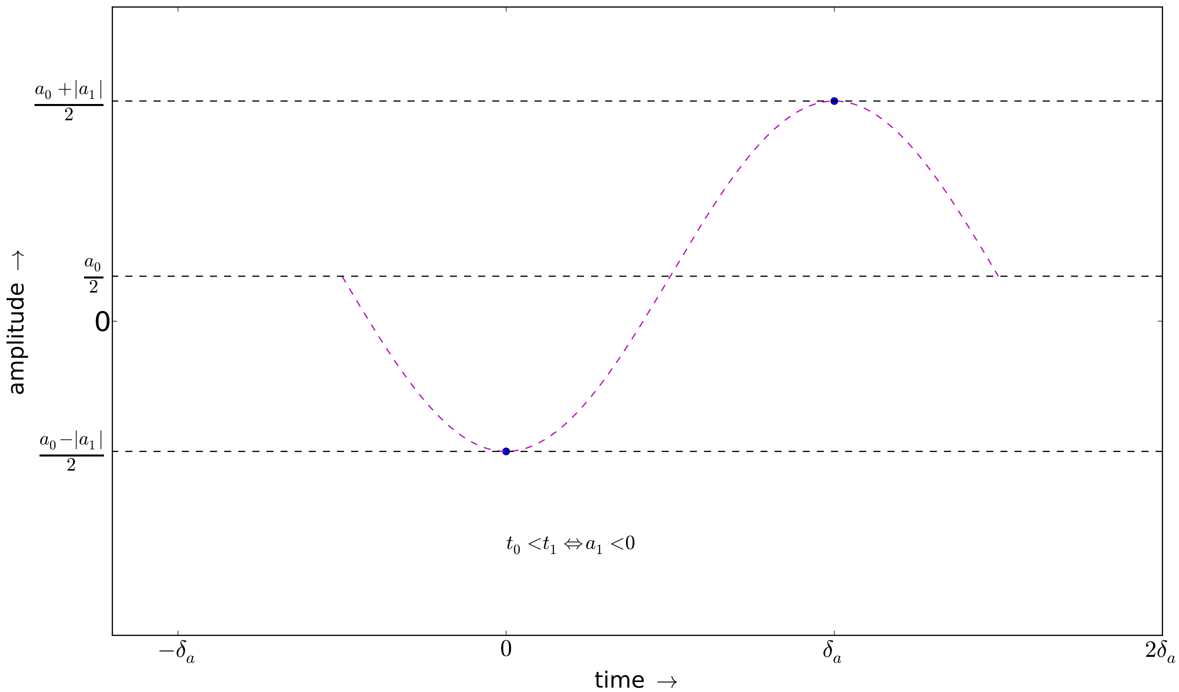

FIG. 5. Oscillation of 2 samples (maximum frequency for any f_a). The first coefficient reflects a detachment (*offset* or *bias*) and the second coefficient specifies the oscillation amplitude.

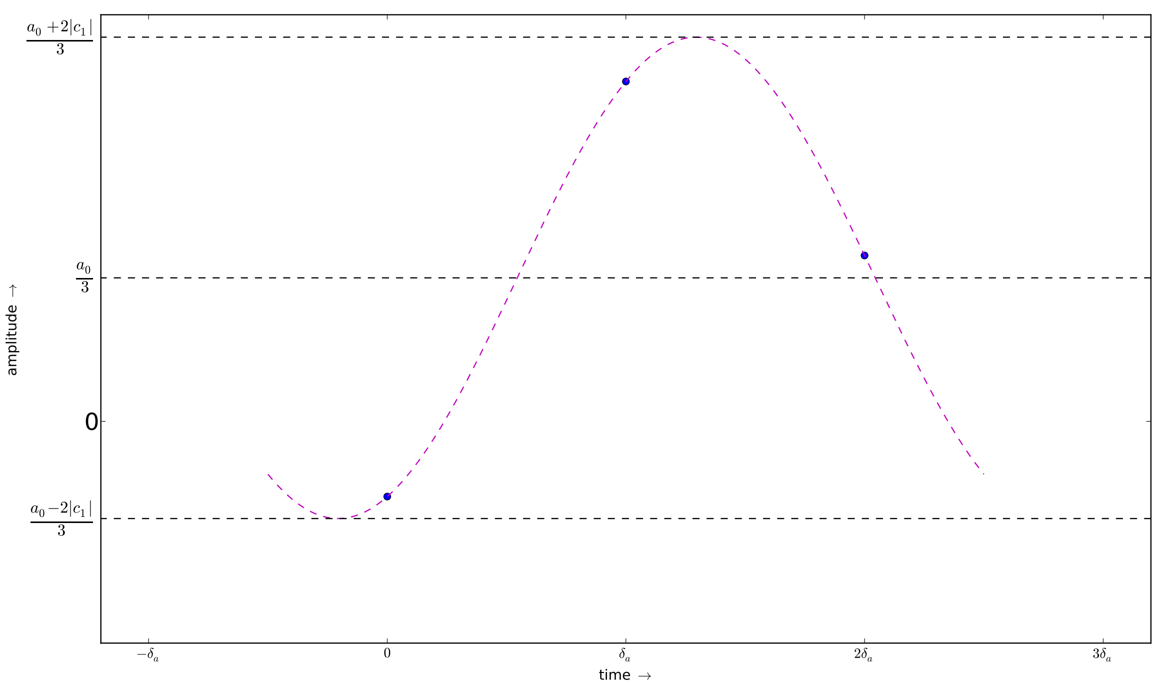

FIG. 6. Three fixed samples present only one non-null frequency. $c_1 = c_2^*$ and $w_1 \equiv w_2$.

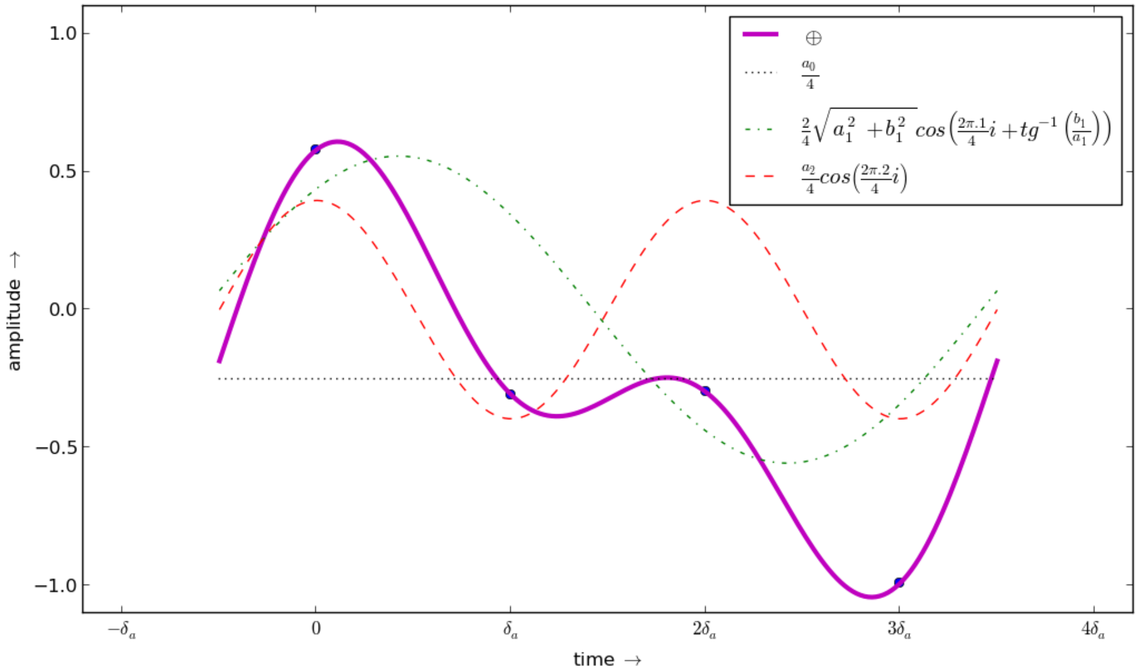

FIG. 7. Frequency components for 4 samples.

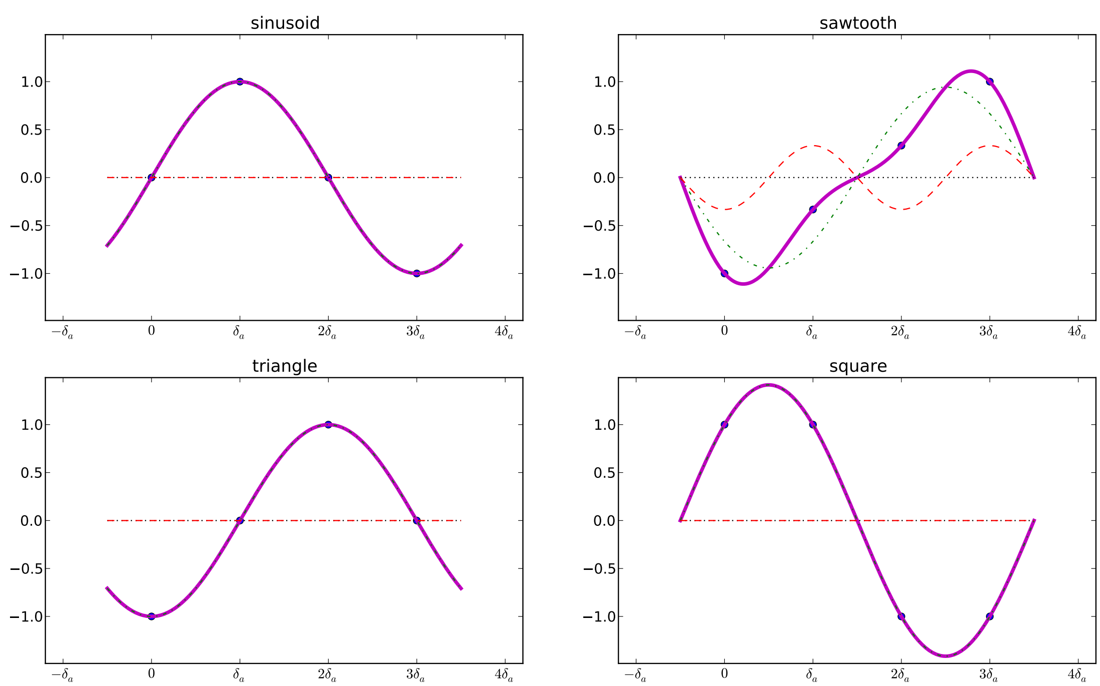

FIG. 8. Basic wave forms with 4 samples.

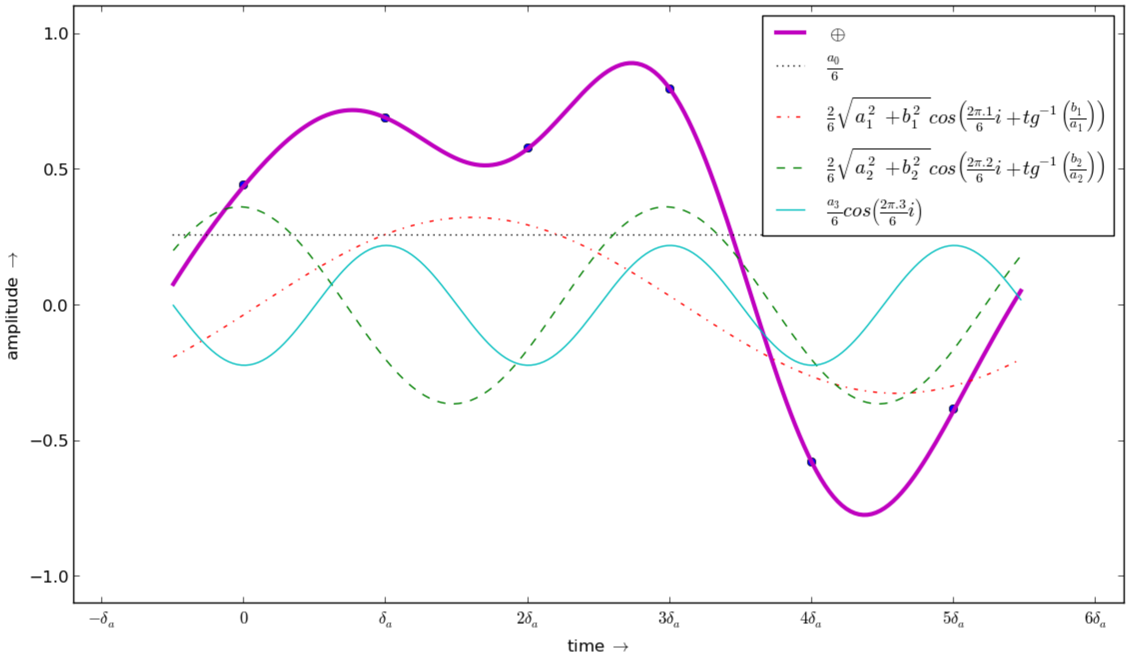

FIG. 9. Frequency components for 6 samples: 4 sinusoids, one of them is the *bias* with zero frequency.

their 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 9, 10, 11 and 12 for the case of 4 samples. There is only 1 sinusoid for each waveform, with the exception of the sawtooth, which has even harmonics.

Figure 9 shows the sinusoidal decomposition for 6 samples, while 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 has an extra component.

F. The basic note

Let f be such that it divides f_a[26]. 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. A 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 9, 10, 11 and 12) 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

A musical note always has a spatial localization, even though this is not one of its four basic properties: the note source position at the ordinary physical space. The reverberation in the environment in which sound occurs is the matter of *spatialization*. Both fields, spatialization and spatial localization, are widely valued by audiophiles and the music industry[27].

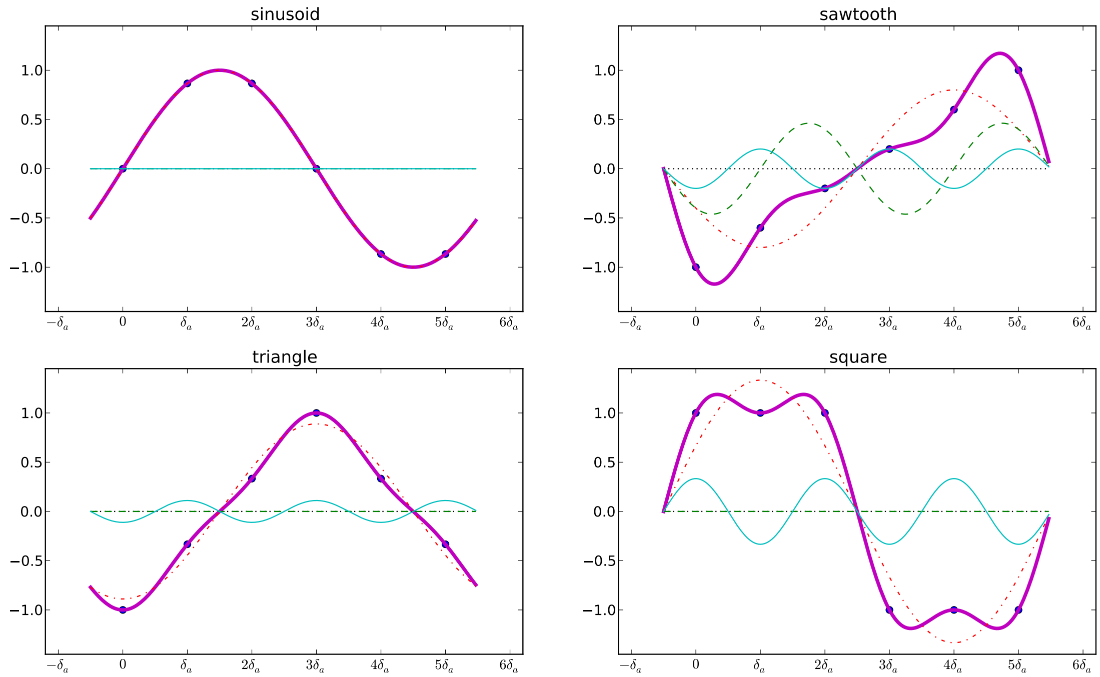

FIG. 10. Basic waveforms with 6 samples: triangular and square waveforms have odd harmonics, with different proportions and phases; the sawtooth has even harmonics.

1. Spatial localization

It is understood that the perception of sound localization occurs in our nervous system by three pieces of information: the delay of incoming sound between both ears, the difference of sound intensity at each ear and the filtering performed by the human body, including its chest, head and ears[1,28,29].

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. Straightforward calculations result in ITD:

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

and in the Interaural Intensity Difference:

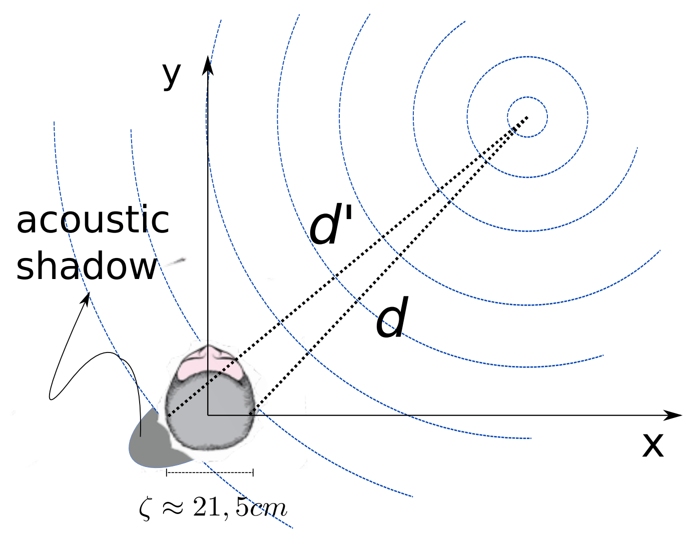

FIG. 11. Detection of sound source localization: schema used to calculate Interaural Time Difference (ITD) and Interaural Intensity Difference (IID).

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

which, converted to amplitude, yields $IID_a = \frac{d}{d'}$. 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[29]. 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 $IID'_a = 1/IID_a$.

Spatial localization depends considerably on 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 falls within 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 especially on the attenuation filtering for high frequencies, since the head interferes much more in the treble than in bass waves[28,29]. Also relevant to the hearing of lateral sources is that low frequencies diffract and the wave arrives to the opposite ear with a delay of $\approx 0.7ms$.[27]

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 on 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 the head and growing out in the direction of each ear[28].

On the other hand, the complete localization, including height and distance of sound source, is given by the Head Related Transfer Function (HRTF)[28]. 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)[30]. Each human body has its filtering and there are techniques to generate HRTFs to be universally used[31].

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, and there are: 1) inversion of the propagation speed component normal to the surface; 2) energy absorption, especially in trebles. The sonic waves propagate until they reach 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'); neither of these characteristics are found in real systems.

There are reverberation models 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. With these values, under the given conditions, there are perceived wave pulses which propagate for 652.08 m ($83.79k$ samples in $f_a = 44.1kHz$) before reaching the ear. In addition, sound reflections made after propagation for 34.32 m ($4.41k$ samples in $f_a = 44.1kHz$) have incidences less distinct by hearing. These first reflections are particularly important to spatial sensation. The first incidence is the direct sound, described by ITD and IID of equations 25 and 26. Assuming that each one of the first reflections, before reaching the ear, will propagate at least $3-30m$, depending on the room dimensions, the separation between the first reflections is $8-90ms$ ($\approx 350-4000$ samples in $f_a = 44.1kHz$). It is experimentally verifiable that the number of reflections increases with the square of $\approx k.n^2$. A discussion about the use of convolutions and filtering to favor implementation of these phenomena is provided in subsection III F, particularly in the paragraphs about reverberation.

H. Musical use

Once the basic note is defined, 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 referred to as mixing:

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

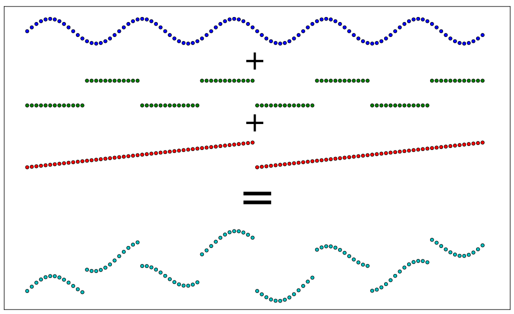

FIG. 12. Mixing of three sound sequences. The amplitudes are directly overlapped.

Figure 12 illustrates this overlapping process of discretized sound waves, each with 100 samples. 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 is very short $\frac{f_a=44.1kHz}{100} \approx 2$ milliseconds. One can complete the sequence with zeroes 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[1]. This process of mixing musical notes results in musical harmony whose intervals between frequencies and chords of simultaneous notes guide subjective and abstract aspects of music and its appreciation[32], which is addressed 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 the 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. In addition, each note has a duration larger than $100ms$ if $f_a < 1kHz$.

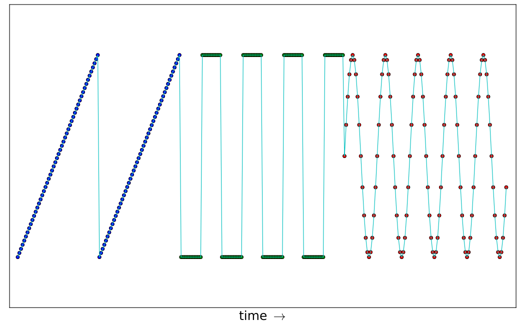

FIG. 13. Concatenation of three sound sequences by temporal overlap of their samples.

The musical piece *reduced-fi* explores the temporal juxtaposition of notes, resulting in a homophonic 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*.[33]

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 II has the following parameters: duration, pitch, intensity (volume) and timbre. This is a useful and paradigmatic model, but it does not exhaust all the aspects of a musical note. First of all, characteristics of the note modify along the note itself[21]. For example, a 3 s piano note has intensity with abrupt rise at the beginning and progressive decay, has spectrum variations with harmonics decaying and some others emerging along time. These variations are not mandatory, but they are used in sound synthesis for music because they reflect how sounds appear in nature. This is considered to be so true that there is a rule of thumb: to make a sound that incites interest by itself, do internal variations on it[1]. 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[33].

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 employ functions without direct calculation, as data sampled from nature. In music its usage transcends these applications, as it simplifies many operations and makes it possible to use a single wave period to synthesize sounds in the whole audible spectrum, with any waveform.

Let $\widetilde{\Lambda}$ be 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 $\widetilde{L_i}$:

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

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 discussed below), it allows for 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 suggest the use of $\widetilde{\Lambda} = 1024$ samples, since it does not produce relevant noise on the audible spectrum. The rounding or interpolation method is not decisive in this process[34].

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 at low frequencies when the size of 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[15,35].

B. Incremental Variations of Frequency and Intensity

As stated by the Weber and Fechner[36] law, the human perception has a logarithmic relation with the stimulus. That is to say, the exponential progression of a stimulus 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:

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

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

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

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 $\left\{\widetilde{l_i}\right\}_0^{\Lambda-1}$. 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:

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

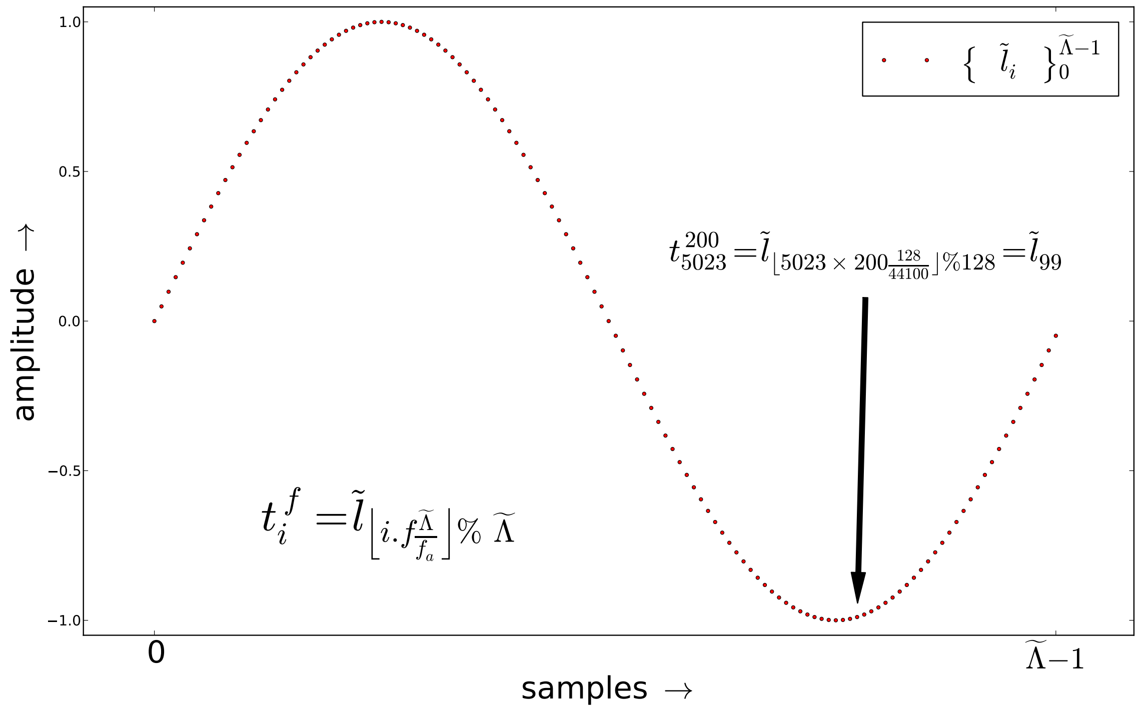

FIG. 14. Search in the *lookup table* to synthesize sounds at different frequencies using a unique waveform with high resolution.

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

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

The term $\frac{i}{\Lambda-1}$ covers the interval $[0,1]$ and it is possible raise it to a power 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[37]. It is sufficient to multiply the original sequence by the sequence $a_{\Lambda-1}^{\left(\frac{i}{\Lambda-1}\right)^\alpha}$, where α is 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 the transition.

Thus, for amplitude variations:

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

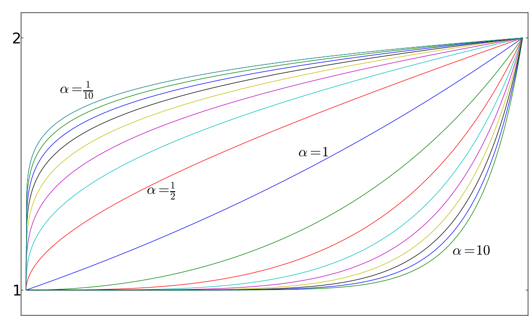

FIG. 15. Intensity transitions for different values of α (see equations 38 and 39).

and

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

It is often convenient to have $a_0 = 1$ to start a new sequence with the original amplitude and then progressively change it. If $\alpha = 1$, the amplitude variation follows the geometric progression that 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 to 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[15].

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 conversion between decibels and amplitude is convenient, with equations 7 and 39 specifying 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*[38].

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*, whose code is online as part of the MASSA toolbox[33].

C. Application of Digital Filters

This subsection is limited to a description of sequences processing, by convolution and differential equations, and immediate applications, as its complexity escapes the scope of this study[39]. Filter applications can be part of the synthesis process or made subsequently as part of processes commonly referred to as "sound treatment".

1. Convolution and finite impulse response (FIR) filters

Filters applied by means of convolution are known by the acronym FIR (Finite Impulse Response) and are characterized by having a finite sample representation. This sample representation is called 'impulse response' $\{h_i\}$. FIR filters are applied in the time domain of digital sound by means of convolution with the respective impulse response of the filter[40]. For the purposes of this work, convolution is defined as:

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

where $t_i = 0$ for the samples not given. In other words, the sound $\{t'_i\}$, resulting from the convolution of $\{t_i\}$, with the 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 impulse response $\{h_i\}_0^{\Lambda_h-1}$. This procedure is illustrated in Figure 16, where the impulse response $\{h_i\}$ is in 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 final signal always has the length of $\Lambda_t + \Lambda_h - 1 = \Lambda_{t'}$.

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

The impulse response can be provided by physical measures or by pure synthesis. An impulse response for a reverberation application, for example, can be obtained by sound recording of the environment when someone triggers a snap which resembles an impulse, or obtained by a sinusoidal sweep whose Fourier transform approximates its frequency response. Both are impulse responses which, properly convoluted with the sound sequence, result in the own sound with a reverberation that resembles the original environment where the measure was made[15]. The inverse Fourier transform of an even and real envelope is an impulse response of a FIR filter. Convoluted with a sound, it performs the frequency filtering specified by the envelope. The greater the number of samples, the higher the envelope resolution and the computational complexity, which should often be weighted, for convolution is expensive.

An important property is the time shift caused by convolution with a shifted impulse. Despite being computationally expensive, it is possible to create *delay lines* by means of sound convolution with an impulse response that has an impulse for each reincidence of the sound. Figure 17 shows the shift caused by convolution with an impulse. Depending on the intensity of the impulses, the result is perceived as rhythm (from an impulse for each couple of seconds to about 20 impulses per second) or as pitch (from about 20 impulses per second and higher frequencies). In the latter case, the process resembles

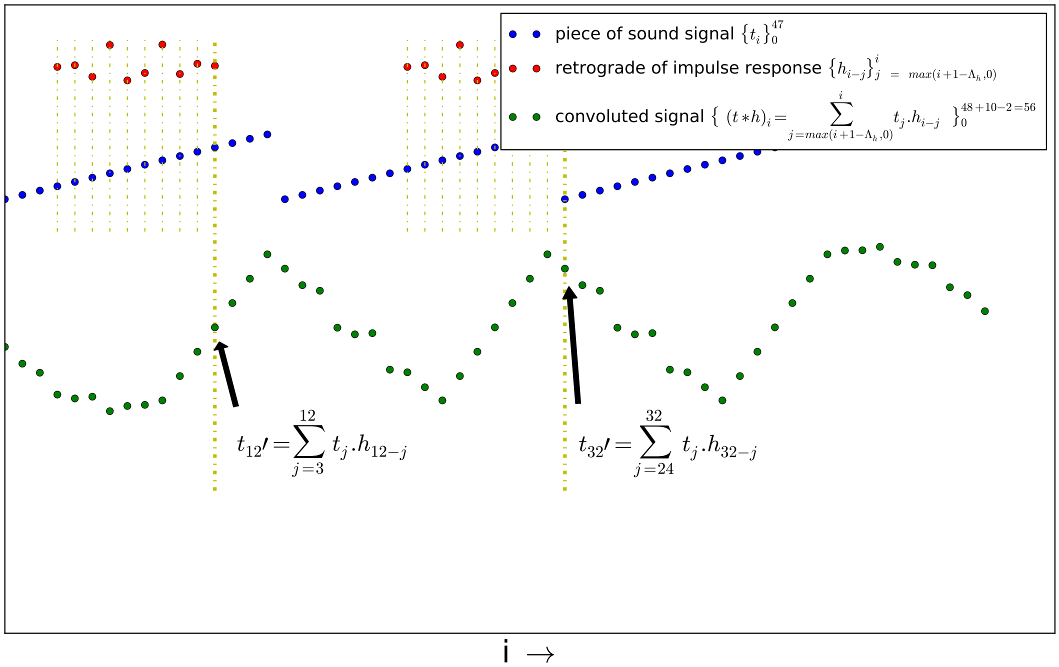

FIG. 16. Graphical interpretation of convolution. Each resulting sample is the sum of the previous samples of a signal, with each one multiplied by the retrograde of the other sequence.

granular synthesis, reverbs and equalization.

2. Infinite impulse response (IIR) filters

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

$$t_i' = \frac{1}{b_0}\left(\sum_{j=0}^{J} a_j.t_{i-j} + \sum_{k=1}^{K} b_k.t_{i-k}'\right) \tag{43}$$

In most cases the variables may be normalized: $a_j' = \frac{a_j}{b_0}$ 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 original samples $\{t_i\}$ and previous resulting ones $\{t_{i-k}'\}$.

There are many methods and tools to obtain IIR filters. The text below lists a selection for didactic purposes and as a reference. They are well behaved filters whose aspects are described in Figure 18. For filters of simple order, the cutoff frequency f_c is where the filter performs an attenuation of $-3dB \approx 0.707$ of the original amplitude. For band-pass and band-reject (or 'notch') filters, this attenuation has two specifications: f_c (in this case, the 'center frequency') and bandwidth bw. In both frequencies $f_c \pm bw$ there is an attenuation of ≈ 0.707 of 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 deviation of the expected profile, expanding the envelope to the bass.

For filters with other frequency responses, it is possible to apply them successively. Another possibility is to use a biquad 'filter receipt'[41] or the calculation of Chebichev filter coefficients[42]. Both alternatives are explored by [43,44], and by the collection of filters maintained by the *Music-DSP* community of the Columbia University[2,45].

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

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

2. High-pass filter with a simple pole, module of its

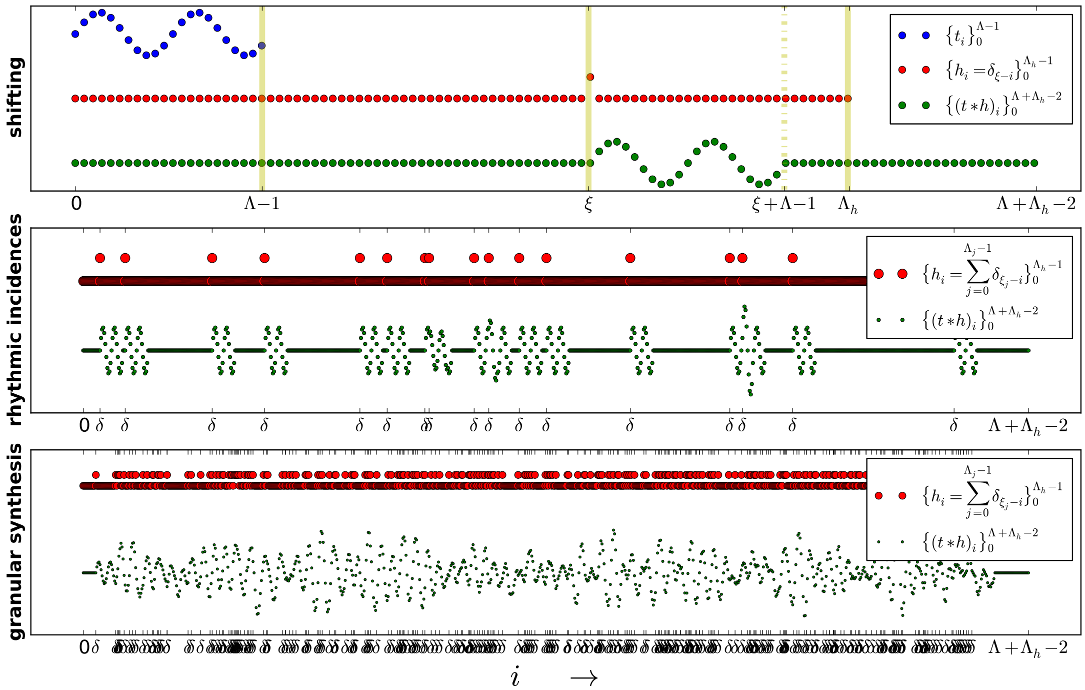

FIG. 17. Convolution with the impulse: shifting (a), delay lines (b) and granular synthesis (c). Represented in increasing order of its pulse density.

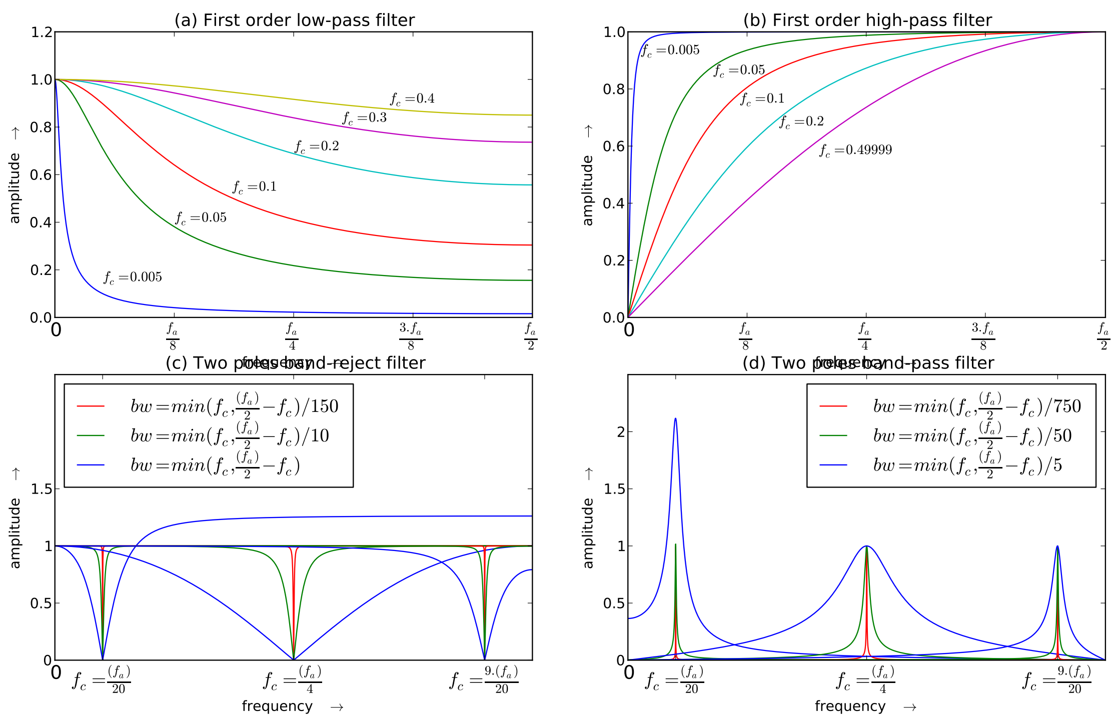

FIG. 18. Modules for the frequency response (a), (b), (c) and (d) for IIR filters of equations 44, 45, 47 and 48 respectively, considering different cutoff frequencies, center frequencies and band width.

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 f_c and bandwidth bw, both given as fraction of f_s, therefore $f,\ bw \in (0, 0.5)$. Both frequencies $f_c \pm bw$ have ≈ 0.707 of the amplitude, i.e. attenuation of $3dB$. The auxiliary variables K and R are 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:

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

The coefficients of band-reject filter, depicted in the lower right of Figure 18, are:

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

D. Noise

Sounds without an explicit pitch are generally called noise[16]. They are important musical sounds, as noise is present in piano notes, violin, etc. Furthermore, the majority of percussion instruments does not exhibit an unequivocal pitch and their sounds are generally regarded as noise[1]. In electronic music, including electro-acoustic and dance genres, noise has diverse uses and frequently characterizes the music style[15].

The absence of a definite pitch is due to the lack of a perceptible harmonic organization in the sinusoidal components of the sound. Hence, there are many ways to generate noise. The use of random values to generate the sound sequence T_i is an attractive method but not outstandingly useful because it tends to produce white noise with little or no variations[15]. Another possibility to generate noise is by using the desired spectrum, from which it is possible to perform the inverse Fourier transform. The spectral distribution should be done with care: if phases of components exhibit prominent correlation, the synthesized sound will concentrate energy in some periods of its duration.

Some noises with static spectrum are listed below. They are called *colored noise* since they are associated with colors. Figure 19 shows the spectrum profile and the corresponding sonic sequence side-by-side. All five noises were generated with the same phase for each component, making it possible to observe the contributions of different parts of the spectrum.

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

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

  The exponential $e^{j.x}$ is a way to obtain unitary module and random phase for the value of c_i. In addition, $c_{\Lambda/2}$ is always real (as discussed in the previous section).

- The pink noise is characterized by a decrease of $3dB$ per octave. This noise is useful for testing electronic devices, being prominent in nature[1].

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

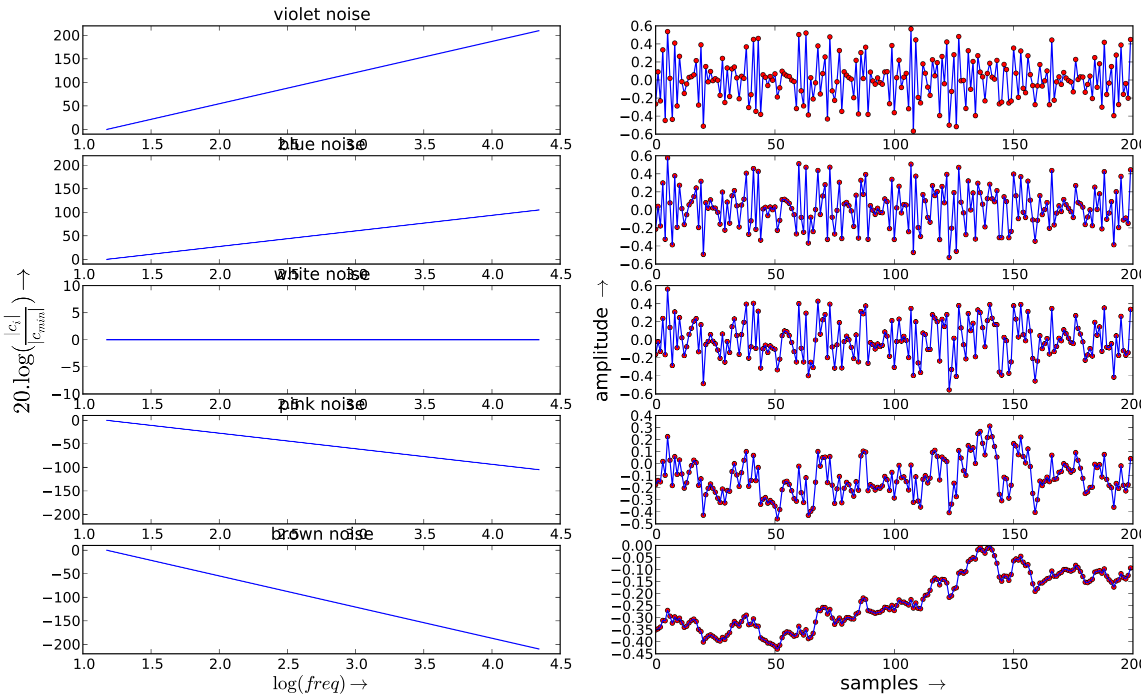

FIG. 19. Colors of noise generated by equations 49, 50, 51, 52 and 53: spectrum and waveforms.

The minimum frequency $f_{\min}$ is chosen with regard to the human hearing, since a sound component with frequency below $\approx 20Hz$ is virtually inaudible.

Other noises can be made by similar procedures. Simple modifications are needed, especially in the equation that defines α_i.

- The brown noise received this name after Robert Brown, who described the Brownian movement[46] What characterizes brown noise is the decrease of $6dB$ per octave, with α_i in equation set 50 being:

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

- In the blue noise there is a gain of $3dB$ per 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:

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

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

$$\alpha_i = (10^{\frac{6}{20}})^{\log_2\left(\frac{f_i}{f_{\min}}\right)} \tag{53}$$

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

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

- The gray noise is defined as a white noise subject to one of the ISO-audible curves. Those curves are obtained by experiments and are imperative to obtain α_i. An implementation of ISO 226, which is the last revision of these curves, is in the MASSA toolbox[33].

This subsection discussed only noise with static spectrum. There are also characterizations for noise with dynamic spectrum during the time, and 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[15].

The noise from equations 49, 50, 51, 52 and 53 are presented in Figure 19. The spectra were built 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 periodic variation in pitch (frequency) and tremolo is a period variation in volume (intensity)[47] For the general case, vibrato is described as:

$$\gamma_i' = \left\lfloor i f' \frac{\widetilde{\Lambda}_M}{f_s} \right\rfloor \tag{55}$$

$$t_i' = \widetilde{m}_{\gamma_i' \, \% \widetilde{\Lambda}_M} \tag{56}$$

$$f_i = f \left(\frac{f+\mu}{f} \right)^{t_i'} = f.2^{t_i' \frac{\nu}{12}} \tag{57}$$

$$\Delta_{\gamma_i} = f_i \frac{\widetilde{\Lambda}}{f_s} \quad \Rightarrow \quad \gamma_i = \left\lfloor \sum_{j=0}^{i} f_j \frac{\widetilde{\Lambda}}{f_s} \right\rfloor =$$

$$= \left\lfloor \sum_{j=0}^{i} \frac{\widetilde{\Lambda}}{f_s} f \left(\frac{f+\mu}{f} \right)^{t_j'} \right\rfloor = \left\lfloor \sum_{j=0}^{i} \frac{\widetilde{\Lambda}}{f_s} f.2^{t_j' \frac{\nu}{12}} \right\rfloor \tag{58}$$

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

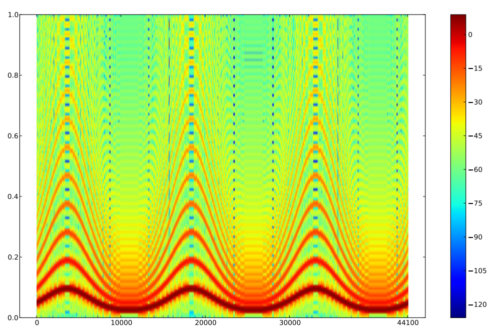

FIG. 20. Spectrogram of a sound with a sinusoidal vibrato of $3Hz$ and one octave of depth in a $1000Hz$ sawtooth wave, with $f_s = 44.1kHz$.

For the proper realization of the vibrato, it is important to pay attention to both tables 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 the 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 peaks 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 frequency decrease, but the semitone variation remains.

Figure 20 is the spectrogram of an artificial vibrato for a note with $1000Hz$ (between a B and a C), in which pitch deviation reaches one octave above and one below. Practically any waveform can be used to generate a sound and the vibrato oscillation pattern, with virtually any oscillation frequency and pitch deviation[48]. Those oscillations with precise waveforms and arbitrary amplitudes are not possible in traditional music instruments, and thus 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 is:

$$a_i = 10^{\frac{V_{dB}}{20} t_i'} = a_{\max}^{t_i'} \tag{60}$$

and finally:

$$T_i^{tr(f')} = \left\{ t_i^{tr(f')} \right\}_0^{\Lambda-1} = \{t_i.a_i\}_0^{\Lambda-1} =$$

$$= \left\{ t_i.10^{t_i' \frac{V_{dB}}{20}} \right\}_0^{\Lambda-1} = \left\{ t_i.a_{\max}^{t_i'} \right\}_0^{\Lambda-1} \tag{61}$$

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 suitable 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 per second (Hz).

The music piece *Vibra e treme* explores these 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 (tremolos and vibratos occur many times together in a traditional music instrument and voices). Aiming at a qualitative appreciation, the piece also develops a comparison between vibratos and tremolos in logarithmic and linear scales. Its source code is available online as part of the MASSA toolbox.

The proximity of f' to $20Hz$ generates roughness in both tremolos and vibratos. This roughness is largely

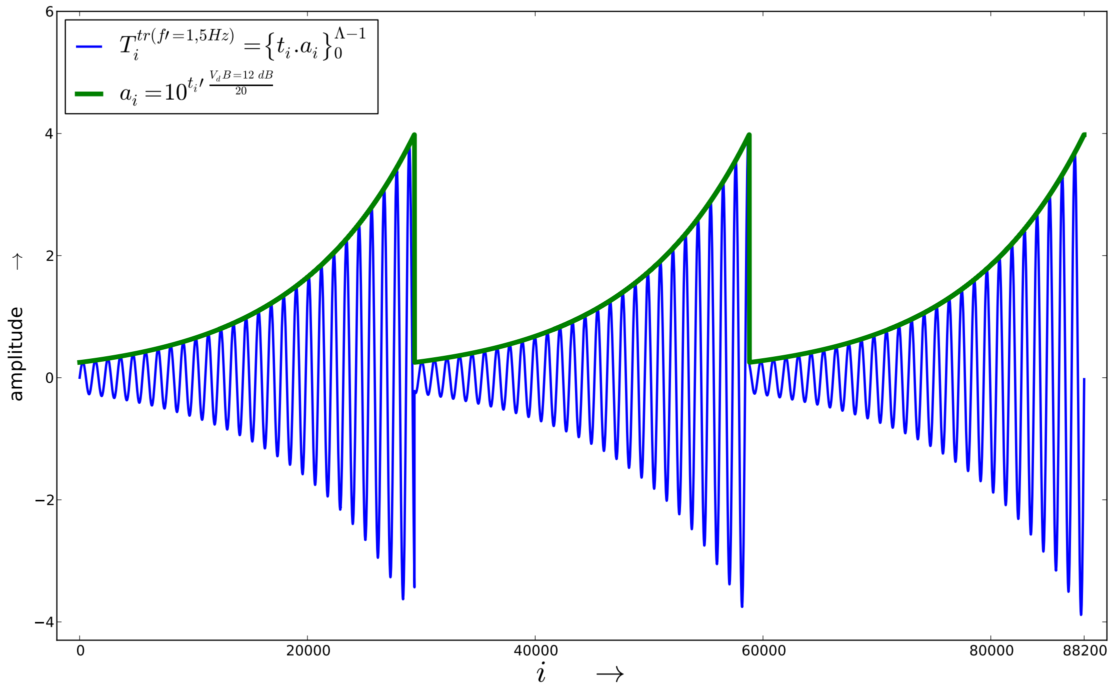

FIG. 21. Tremolo with a depth of $V_{dB} = 12dB$, with a sawtooth waveform as its oscillatory pattern, with $f' = 1.5Hz$ in a sine of $f = 40Hz$ (sample frequency $f_s = 44.1kHz$).

appreciated both in traditional classical music and current electronic music, especially in the *Dubstep* genre. Roughness is also generated by spectral content that produces beating[49,50]. The sequence *Bela Rugosi* explores this roughness threshold with concomitant tremolos and vibratos at the same voice, with different intensity and waveforms. The corresponding code is available online in the MASSA toolbox.

As the frequency increases further, these oscillations no longer remain noticeable individually. In this case, the oscillations are audible as pitch. Then, f', μ and the waveform together change the spectrum 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, as in telecommunications for data transfer by means of electromagnetic waves (e.g. AM and FM radios).

For musical goals, it is possible to understand FM based on the case of sines and, in occurrences of greater complexity, to decompose the signals into their respective Fourier components (i.e. sines as well). 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':

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

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[43,51] which specifies the amplitude of each component in FM synthesis.

In these equations, the frequency variation introduced by $\{t'_i\}$ does not follow the geometric progression that yields linear pitch variation, but reflects equation 32. The use of equations 57 for FM is described in[4], where the spectral content of the FM synthesis is calculated for oscillations in the logarithmic scale. In fact, the simple and attractive FM behavior are observed only with linear variations, such as 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 the distance f' from f. Again, this is obtained by variations in the linear scale of the amplitude. A discussion of the spectrum of an AM performed with oscillations in the logarithmic amplitude scale is available in[4]. The sequence T_i, with frequency f, called 'carrier', is modulated by f', called 'modulator'. In FM and AM jargon, μ and $\alpha = 10^{\frac{V_{dB}}{20}}$ are 'modulation indexes'. The following equations are defined for the vibration pattern of the modulator sequence $\{t'_i\}$:

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

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

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

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

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

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

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

To perform AM, the signal $\{t_i\}$ needs to be modulated 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 usages

At this point, the musical possibilities are very wide. All characteristics, like pitch (given by frequency), timbre (given by the waveform, filters and noise), volume (manipulated by intensity) and duration (given by the number of samples), can be considered in an absolute form or treated during duration of sound (clearly, with the exception of duration itself). The following musical usages comprehend a collection of possibilities with the purpose of exemplifying types of sound manipulation that result in musical material. Some of them are discussed in depth in the next section.

1. Relations between characteristics

An interesting possibility is to use relations between parameters of a tremolo or vibrato, and some parameters of the basic note like frequency. It is possible to have a vibrato frequency proportional to note pitch, or a tremolo depth inversely proportional to pitch. Therefore, with equations 55, 57 and 60, it is possible to choose:

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

with f^{vbr} and f^{tr} as f' in the referenced equations. They can also be associated with vibrato and tremolo oscillation frequency of equation 55. ν and V_{dB} are the respective depth values of vibrato and tremolo. Functions $func_a$, $func_b$ and $func_c$ are arbitrary and dependent on musical intentions. The music piece *Tremolos, vibratos e a frequncia* explores such bonds and exhibits variations in the oscillation waveform with the purpose of building a *musical language* (details in the next section). The corresponding code is also available online as part of the MASSA toolbox.

2. Convolution for rhythm and meter

A musical pulse - such as specified by a BPM tempo - can be used with an impulse at the start of each beat, with the purpose of establishing metric and rhythms: the convolution with an impulse shifts the sound to impulse position. For example, two impulses equally spaced build a binary division into the pulse. Two signals, one with 2 impulses and the other with 3 impulses, both equally spaced in the pulse duration, yields a pulse maintenance with a rhythm in which eases both binary or ternary divisions of the pulse. This is found in many ethnic and traditional musical styles[52]. Absolute values of the impulses implies in proportions among the amplitudes of the sonic reincidences. The use of convolution with impulses in this context is explored in the music piece *Trenzinho de caipiras impulsivos*. These features embrace the creation of 'sound amalgams' based on granular synthesis, refer to Figure 24. This piece is a link to next section contents. The source code of the music piece is online, as part of the MASSA toolbox[33].

3. Moving source and receptor, Doppler effect

Following the exposition in Subsection II G, when an audio source or receptor is moving, its characteristics are ideally updated at each sample of the digital signal. Speed components should be found for each ear. In this way, given the audio source speed (or velocity) v_s, with positive values if the source moves away from receptor, and receptor speed v_r, positive when it gets closer to audio source, the frequency is given by the well-known Doppler shift:

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

With both frequencies f and f_0, and the 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 power, which may be understood as being proportional to the wave shrinking or expansion: $\Delta P = P_0 \left(\frac{v_r - v_s}{343.2}\right)$, where P_0 is signal power and P the potency at receptor. Both amplitude and frequency of a moving audio source can be obtained. If this audio source is in front of the receptor with y_0 m of horizontal distance and z_0 m of height, the distance is given by $D_i = \left\{ d_i = \sqrt{y_i^2 + z_0^2} \right\}_0^{\Lambda-1}$, where $y_i = y_0 + (v_s - v_r)\frac{i}{f_s}$ with v_s and v_r both horizontal, having null non-y components. Amplitude changes with the distance and with the potency factor mentioned above (see subsection II B for potency to amplitude conversion).

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

The amplitude change caused by the distance is even, while the change 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. Moreover, the transition is abrupt if source and receptor intersect with zero distance, otherwise, there is a monotonic progression. In the given case, where there is a static height z_0, the speed component in the direction given by the observer and source positions, the frequencies F_i at the observer is given by:

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

There is a Python implementation of the Doppler effect, also considering the intersection between audio source and receptor, in MASSA toolbox[33].

4. Filters and noises (subsections III D and III C)

With the use of filters, the possibilities are even wider. Convolve a signal to have a reverberated version of it, to remove its noise, to distort or to treat the audio aesthetically in other ways. For example, sounds originated from an old television or telephone can be simulated with a band-pass filter, allowing only frequencies between $1kHz$ and $3kHz$. By rejecting the frequency of electric oscillation (usually $50Hz$ or $60Hz$) and harmonics, one can remove noises caused by audio devices connected to a usual power suply. 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[1]. Chains of these filters can perform complex and more accurate filtering routines. The music piece *Ruidosa faixa* explores filters and many kinds and noise synthesis. The source code is available online as part of MASSA toolbox.

A sound can be altered through different filtering and then mixed to create an effect known as *chorus*. Based on what happens in a choir of singers, 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. Its source code is available in MASSA toolbox.

5. Reverberation

Using the same terms of subsection II G, the late reverberation can be achieved by a convolution with a section of pink, brown or black noise, with exponential decay of amplitude along time. Delay lines can be added as a prefix to the noise with the decay, and this contemplates both time parts of the reverberation: the first reflections and the late reverberation. Quality can be improved by varying the geometric trajectory and (mainly low-pass, subsection III C) filtering by each surface where wavefront reflected before reaching the ear in the first $100-200$ ms. The colored noise can be gradually introduced with a *fade-in*: the initial moment given by direct incidence of sound (i.e. without any reflection and given by ITD and IID), reaching its maximum at the beginning of the 'late reverberation', when the geometric incidences lose their relevance to the statistical properties of noise decay. As an example, consider Δ_1 as the duration of the first section and Δ_R as the duration of total reverberation ($\Lambda_1 = \Delta_1 f_s$, $\Lambda_R = \Delta_R f_s$). Let p_i be the probability of a sound to repeat in the i-th sample. The amplitude decreases exponentially. Following subsection II G, reverberation R_i^1 of the first period can be described as:

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

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 amplitude decay of the waveform:

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

With

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

, a reverberated auralization of a sound can be achieved by 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 for causing great interest in listeners and to provide more enjoyable sonorities. Furthermore, modifications in the reverberation space consists on a common resource (almost a *cliché*) to surprise and attract the listener. An implementation of the reverb recipe described here is in MASSA toolbox[33]

6. ADSR Envelopes

The variation of volume 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 pioneering implementation can be found in the Hammond Novachord synthesizer of 1938 and some variants are mentioned below[53]. The scholastic ADSR envelope is characterized by 4 parameters: attack duration (time in which sound reaches its maximum amplitude), decay duration (follows the attack immediately), level of sustained volume (in which the volume remains stable after the decay) and release duration (after sustained section, this is the duration needed for amplitude to reach zero or final value). 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.

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, to be applied to any sound sequence $T_i = \{t_i\}$, can be presented as:

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

with $\Lambda_X = \lfloor \Delta_X . f_s \rfloor \ \forall \ X \ \in (A, D, R)$ with ξ being 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 right side of Equations 78 can attend both introduction and ending of sound from zero intensity because they are linear. Schematically, Figure 22 shows the ADSR envelope in 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 remains for more than a peak. Another common example is the use of more elaborated tracings of attack or decay. The music piece *ADa e SaRa*, available in MASSA, explores many configurations of the ADSL envelope.

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

IV. ORGANIZATION OF NOTES IN MUSIC

Consider $S_j = \left\{s_j = T_i^j = \{t_i^j\}_{i=0}^{\Lambda_j - 1}\right\}_{j=0}^{H-1}$, the sequence S_j of H musical events s_j. Let S_j be a 'musical structure', composed by events s_j which are musical structures 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, for building intervals and chords. This reflects the 'vertical thought' in music. On the other hand, the concatenation of events in S_j, as in equation 30 and in Figure 13, builds melodic sequences and rhythms, which are associated with the musical 'horizontal thought'. The fundamental frequency

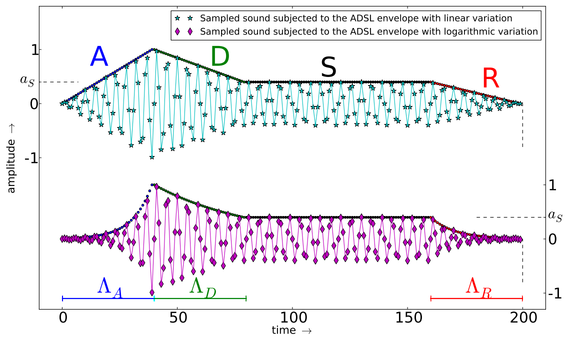

FIG. 22. An ADSR envelope (*Attack, Decay, Sustain, Release*) applied to an arbitrary sound sequence. The linear variation of the amplitude is above, in blue. Below the amplitude variation is exponential.

f and the starting moment (attack) are generally the most important characteristics of the elements s_j in S_j. This makes it possible to create music made by pitches (both harmony and melody) and by temporal metrics and rhythms.

A. Tuning, intervals, scales and chords

1. Tuning

Doubling the frequency is equivalent to ascending one octave ($f = 2f_0$). The octave division in twelve steps is the canon of classical western music. Its usage has also been observed outside western tradition, as in ceremonial/religious and ethnic context[12]. In this system, ideally the factor given by $\varepsilon = 2^{\frac{1}{12}}$ defines a semitone. This constitutes a note grid along the spectrum in which, given the frequency f, all the possible fundamental frequencies are separated by intervals which are multiples of ε. Twelve semitones (or half steps), equidistant to the human ear, reach an octave, equivalent to the human year. 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. Real music instruments, however, exhibit deviations in order to make compatible deviations from the harmonics of notes. In addition, the fixed reference $\varepsilon = 2^{\frac{1}{12}}$ characterizes an equally tempered tuning. There are tunings in which intervals are ratios of low-order integers, based on observations of physical phenomena. These tunings were first formalized around two thousand years before the arrival of the equal temperament[1].). Two emblematic tunings are:

- The **just intonation**, defined by association of diatonic scale notes with ratios of low-order integers, as found in the harmonic series. Considered the ionian mode (see Section IV A 3), we achieve notes correspondent to the white piano keys from C to C, i.e. intervals 1, 2M, 3M, 4J, 5J, 6M, 7M, 8, are considered ratios of frequency: 1, 9/8, 5/4 4/3, 3/2, 5/3, 15/8, 2/1. The following intervals are also considered: 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 in the just intonation.

- The **Pythagorean tuning**, based on the interval 3/2 (perfect fifth). The Ionian mode ratios become: 1, 9/8, 81/64, 4/3, 3/2, 27/16, 243/128, 2/1. These intervals are also considered to yield notes in a scale. Besides the intervals in the mode, also used are 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 account for micro-tonality[54]. It is possible to use non-integer real values as factors of $\varepsilon = 2^{\frac{1}{12}}$ between frequencies, or to maintain integer values and change ε. For example, a tuning really near the harmonic series itself is proposed with the equal division of the octave in 53 notes: $\varepsilon_2 = 2^{\frac{1}{53}}$.[55] In this division, notes are related by means of integers, factors of ε_2 relating frequencies by a valid interval. Note that if S_i is a pitch sequence related by means of ε_1, a direct mapping for notes related by ε_2 is a new sequence $S'_i = \{s'_i\} = \left\{ s_i \frac{\varepsilon_1}{\varepsilon_2} \right\}$. The music piece *Micro tom* uses microtonal features and its code is part of MASSA toolbox.

2. Intervals

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

The nomenclature, based on conveniences for tonal music and practical aspects of note manipulation, can be specified as follows[1,12]:

- Intervals are seen first by the number of steps between notes: first (unison), second, third, forth, fifth, sixth, seventh and eighth (octave). The ninth, tenth, eleventh, etc, are compound intervals made by one or more octaves plus a "simple" interval inside the octave, which characterizes the compound interval. The intervals are represented by numeric digits, e.g. 1, 3, 5 are a unison, a third and a fifth, respectively.

- Qualities of each interval: 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. Except for the tritone.

TABLE I. Music intervals together with their traditional notation, basic classification for dissonance and consonance, and number of semitones. Unison, fifth and octave are the perfect (P) 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 a special case, as it is a perfect consonance when considered as an inversion of the perfect fifth and a dissonance or an 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 seeks urgent resolution in a third or sixth. Due to this instability it is considered a dissonant 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	5
dissonance in Western tradition:	tritone, aug4, dim5	6

- The perfect fourth can be a perfect consonance or a dissonance according to the context and theoretical background. As a general rule, it can be considered a consonance except when it resolves into a fifth or a third with movimentation of just one note.

- Tritone is a dissonance in Western music because it characterizes the "dominant" chord in the tonal system (see subsection IV B) and represents the instability. Some cultures have the interval as a consonance, using it in melodies and chants as a stable interval.

- 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: 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, the interval is inverted. The sum of an interval and its inversion is 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 (inverting a doublyi-augmented results in a doubly-diminished and vice-verse, etc.). 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 reached by 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. Scale degrees (subsection IV A 3) are in fact 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 fashion, the second degree of a scale can be one semitone from 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[16].

The descriptions presented summarize the traditional theory of music intervals[16]. The music piece *Intervalos entre alturas* explores these intervals in both independent and related manners. The source code is available online in MASSA toolbox[33].

3. Scales

A scale is an ordered set of pitches. Usually, scales repeat at each octave. The ascending sequence with all notes from the octave division in 12 equal intervals, separated by the ratio $\varepsilon = 2^{\frac{1}{12}}$, is known as the chromatic scale with equal temperament. There are 5 perfectly symmetric divisions of the octave within the chromatic scale. These divisions are considered scales owing to the easy and peculiar uses they provide. Considering integers e_i used as powers of the factor $\varepsilon = 2^{\frac{1}{12}}$ to multiply the lower note frequency ($f = \varepsilon^{e_i} f_0$), the scales are as follows:

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

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

Diatonic modes are:

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

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

Diatonic scales follow a circular 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 $\kappa \in [0,6]$ by which $\{e_i\}$ matches. 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 have two additional forms, named melodic and harmonic:

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

The different ascending and descending contour of the melodic minor scale is required by the leading tone (seventh and last degree, separated by one semitone from the octave, enhances tonic polarization) in the ascending trajectory, which is not necessary in the descending mode, and therefore it recovers the normal form. The harmonic scale presents the leading tone but does not avoid the augmented second between the sixth and seventh degrees; it does not consider the melodic trajectory[32]. Other scales can be represented in a similar form, like the pentatonic and the modes of limited transposition by Messiaen[56].

Although it is not a scale, the harmonic series is often used as such. It is a convenient to write the notes related to each harmonic:

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

In this scale, the frequency of the $i-th$ note h_i is the frequency of $i-th$ harmonic $f = \varepsilon^{h_i} f_0$ from the spectra generated by f_0. Nature exhibits these 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. Chords are based on triads, especially in tonal music. 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 an inverted chord as opposed to the chord in its root position. A closed position is any in which no chord note fits between two consecutive notes[16], any non-closed position is an open position. In closed and fundamental position, and with fundamental in 0, triads are described as:

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

To have another third superimposed to the fifth, it 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. Inversions and open positions can be obtained with the addition of ± 12 to the selected component. Incomplete triadic chords, with extra notes ('dirty' chords), and non-triadic are also common.

For general guidance:

- A fifth confirms the root (fundamental).
- Major or minor thirds suggest major or minor chord qualities.
- Every tritone, especially if built between a major third and a minor seventh, tends to resolve into a third or a sixth.
- Note duplication is avoided. If duplication is needed, the preferred order is: the root, fifth, third and seventh.
- Note omission is avoided in the triad. If needed, the fifth is first considered for omission, than third and, if needed, the fundamental.
- It is possible to build chords with notes different from triads, particularly if they obey a recurrent logic or musical concatenation that justifies these different notes.
- Chords built by successive intervals different from thirds like fourths and seconds – are recurrent in compositions of advanced tonalism or experimental character.
- The repetition of chord successions (or their characteristics) fixes a trajectory and makes it possible to Introduce exotic formations without incoherence.

B. Atonal and tonal harmonies, harmonic expansion and modulation

Omission of basic tonal chaining is the key to obtain modal and atonal harmonies. In the absence of these minimal tonal structures, harmony is considered modal if the notes match with some diatonic scale (see equations 81) or if notes are presented in a small number. If basic tonal chainings are absent and notes do not match any diatonic scale and are diverse and dissonant (by relation with each other) enough to avoid reduction by polarization, harmony is atonal. In this classification, the modal harmony is not tonal or atonal and is reduced to the incidence of notes within the diatonic scale (or simplifications) and to the absence of tonal structures. Following this concept, one observes that atonal harmony is hard to be realized and, indeed, no matter how dissonant and diverse a set of notes is, tonal harmonies arise if not avoided[57].

1. Atonal harmony

In fact, the techniques around atonal music aim at structures for avoiding direct relation with modes and tonality. Manifesting such structures is of such difficulty that the dodecafonism arouse. The purpose of dodecafonism is to use a set of notes (ideally 12 notes), and to perform each note, one by one, in the same order. In this context, the tonic becomes difficult to be established. Nevertheless, the western listener searches for tonal traces in music and easily finds them by unexpected and tortuous paths. The use of dissonant intervals (especially tritones) without resolution, reinforces the absence of tonality. In this context, while creating a music piece, it is possible to:

- Repeat notes. By considering immediate repetition as an extension of previous incidence, the use of the same note in sequence does not add relevant information.
- To play adjacent notes at the same time, making harmonic intervals and chords.
- Present durations and pauses with liberty, respecting notes order.
- Vary by extension, transposition, translation, inversion, retrograde and retrograde inversion. See subsections IV E and IV I for more details.
- Accounting 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 (a simple dodecaphonic model). Most of what was written by great dodecafonic composers, e.g. Alban Berg and even Schoenberg, had the purpose of mixing tonal and atonal techniques.

2. Tonal harmony

In the XX century, rhythmic music and with emphasis on sonorities/timbres, extended the concepts of tonality and harmony. Hence, tonal harmony has strong influence on artistic movements and commercial venues. In addition, dodecafonism itself is considered of tonal nature because it denies tonal characteristics of polarization. In tonal or modal music, chords – like the ones listed in equations 85– built on the root note of each degree of a scale – such as listed in equations 81 – form the pilars of the harmonic field. Music harmony aims at studying the incidence of 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 music, a scale has its tonic (first degree) on any note, and can be major (with the same notes of Ionian mode) or minor (same notes as Eolian mode, 'natural minor', which has both harmonic and melodic versions, see equations 83). The scale is the base for triads, each with its root in a degree: $\hat{1}, \hat{2}, \hat{3}, \hat{4}, \hat{5}, \hat{6}, \hat{7}$. To build triads, the third and the fifth notes above the root are taken together with the root (or fundamental). $\hat{1}, \hat{3}, \hat{5}$ is the first degree chord, built on top of scale's first degree and central for 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 these, other degrees are considered. The 'traditional harmony' comprises conventions and stylistic techniques to create chainings with such chords[32].

The 'functional harmony' ascribes functions to these three central chords and tries to understand their use by means of these functions. The chord built on top of the first degree is the **tonic** chord (T or t for a major or minor tonic, respectively) and its function (role) consists of maintaining a center, usually referred to as a "ground" for the music. The chord built on the fifth degree is the **dominant** (D, the dominant is always major) and its function is to lean for the tonic (the dominant chord asks for a conclusion and this conclusion is the tonic). 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 function is to deviate the music from the tonic. The process aims at confirming the tonic using tonic-dominant-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 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 in a third above. In the minor scale the same happens, but the triad a third below is called counter-relative (tA, sA) and the triad a third above is called relative (tR, sR). The precise functions and musical effects of these chords are controversial. Table II shows relations 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 omitted root'[58]. In the minor mode, there is a change in $\hat{7}$ by an ascending semitone to yield a unique semitone separating $\hat{7}$ and $\hat{1}$, and making the dominant possible (that should be major and goes to the tonic). In this way, the dominant is always major, for both major and minor scale and, therefore, even in 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 individual dominant or subdominant, which are the triads based on a fifth above or a fifth below, respectively. These individual dominants and subdominants, in the same way, have also subdominants and individual dominants of their own. 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 condition is that the occurrence presents a coherent trajectory of dominants and subdominants to the original tonality.

Mediants, or 'chromatic mediants', are two for each chord: the upper mediant, formed by the root at the third of original chord; and the lower 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, it is a 'doubly chromatic mediant'. Again, there are two forms for each chord: the upper form, with a third in the fifth of the original triad; and the lower form, with a third in the root of the original triad. It is worth observing that a major chord has major chromatic and doubly chromatic mediants. A minor chord has minor chromatic and doubly chromatic mediants. (Recall that relatives and counter-relatives have opposite major-minor quality.) This relation between chords is considered advanced tonalism, sometimes even considered expansion and dissolution of tonalism, with strong and impressing effects although they are perfectly consonant triads. Chromatic mediants are used since the end of Romanticism by Wagner, Lizt, Richard Strauss, among others, and are quite simple to be realized[32,59].

4. Modulation

Modulation is the change of key (tonic, or tonal center) in a music, being characterized by start and end keys, and transition artifacts. Keys are always conceived as related by fifths and their relative and counter-relatives. Some ways 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.
- Careful use of an individual dominant, and perhaps also the individual subdominant, to confirm change in key and harmonic field.
- Use of chromatic alterations to reach a chord in the new key starting from a chord in the previous key. Called chromatic modulation.
- Featuring a unique note, possibly repeated or suspended with no accompaniment, common to start and end keys, it constitutes 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. Keys with same tonic and different quality are known as homonyms.

The dominant has great importance and is a natural pivot into modulations, which leads to the circle of fifths[32,58–60]. Other inventive ways to modulate are possible, to point but one common example, the minor thirds tetrad (E_i^{tm} in equations 80) can be sustained to bridge to other tonalities, with the ease of its both tritones. The music piece *Acorde cedo* explores these chord relations, and is implmemented online as part of MASSA toolbox[33].

C. Counterpoint

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

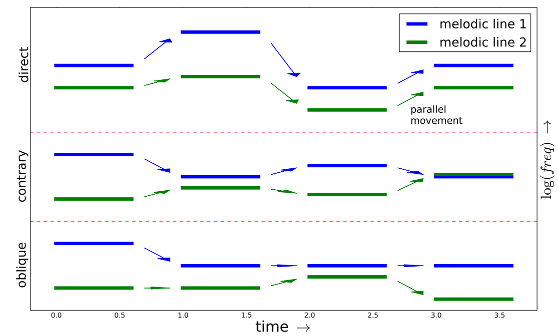

FIG. 23. Different motions of counterpoint aiming to preserve independence between voices. There are 3 types of motion: direct, contrary and oblique. The parallel motion is a type of direct motion.

The purpose of counterpoint is to conduct voices in a way that they sound independent. In order to do that, the relative motion of voices (in pairs) is crucial and categorized as: direct, oblique and contrary (see Figure 23). The parallel motion is a direct motion. The gold rule here is to take care of the direct motions, avoiding them when leading to a perfect consonance. The parallel motion should occur only between imperfect consonances and no more than three consecutive times. Disonances 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 to a neighbor note in the scale sound coherent. When having 3 or more voices, the melodic importance lies in the higher and lower voices, in this order[61–63].

These rules were used in the music piece *Conta ponto*, whose source code is available online in MASSA toolbox.

D. Rhythm

Rhythmic notion is dependent on events separated by durations[16], which can be heard individually if spaced by at least $50 - 63ms$. For the temporal separation between them to be perceived as duration, the period should even a bit larger, around $100ms$[19]. It is possible to summarize duration heard as rhythm or pitch, and its transition, as in table III[19,64].

The duration transition span is 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 sonic characteristics[1,19].

The rhythmic metric is commonly based on a basic duration called pulse, whose durations range between $0.25 and 1.5s$ (240 and $40BPM$, respectively). In music

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 →

education and cognitive studies, it is common to associate these range of frequencies to the durations of the heart beat, movements of inspiration and expiration and steps of a walking or running person[1,16].

The pulse is subdivided into equal parts and is also repeated in sequence. These relations (division and concatenation) usually follow relations of small order integers. The occurrences, in written and ethnic music, of musical pulse division (and their sequential grouping at time), are, in ascending order: 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 division 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 common compositions and decompositions of 1-9 equal parts[1,52]. A schematic illustration is shown in Figure 24.

Dual relations (simple measures and binary divisions) are frequent in dance rhythms and celebrations, and are called "imperfect". Ternary relations are common in ritualistic music and is related to the sacred, and are called "perfect". Strong units (accents) fall in the 'head' of bars, the downbeats, of which the first in every bar is the greatest. The head of a unit is the first part of the subdivision. In binary divisions (2, 4 and 8), strong units alternate with weak units (e.g. division in 4 is: strong, weak, average strong, weak). In ternary divisions (3, 6 and 9) two weak units succeed the downbeat (e.g. division in 3 is: strong, weak, weak). Division in 6 is considered compound but can also occur as a binary division. Binary division units which suffers a ternary division yields two units divided into 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 ternary division units that divides as binary, resulting in: a strong unit (subdivided in strong and weak) and two weak units (subdivided in strong and weak for each).

The accent in the weak beat is the backbeat, whereas a note starting in a weak beat persisting across the strong beat is the syncope.

Notes can occur inside and outside of these *'musical metric'* divisions. In most well-behaved cases, notes occur exactly on these divisions, with greater incidence on strong beat attacks. In extreme cases, time metric cannot be perceived[1]. Small variations along the grid are crucial for musical interpretation and different styles[15].

Let the pulse be the grouping level $j = 0$, the first pulse subdivision be level $j = -1$, the first pulse agglomeration be level $j = 1$ and so on. Hence, P_i^j is the i-th unit at grouping level j: P_{10}^0 is the tenth pulse, P_3^1 is the third grouped unit (and, possibly, the third measure), P_2^{-1} is the second part of pulse subdivision. The limits of j are of special interest: pulse divisions are durations perceivable as rhythm; furthermore, the pulses sum, 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 sections i. When i has three different values (or multiple of three) there is a perfect relation. When i has only two, four or eight different values, than there is an imperfect relation, as shown in Figure 24. Any unit (note) of a given musical sequence with 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 itself.

As an 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 (possibly second measure). Each unit P_i^j can be associated with a sequence of temporal samples T_i that comprehends a note. The music piece *Poli Hit Mia* uses different metrics (available as part of MASSA toolbox).

E. Repetition and variation: motifs and larger units

Given the basic musical structures, both frequential (chords and scales) and rhythmic (simple, compound and complex beat divisions and agglomerations), it is natural to present these structures in a coherent and meaningful way[65]. The concept of arcs is essential in this context: by departing from a place and returning, an arc is made. The audition of melodic and harmonic lines is permeated by musical arcs due to the cognitive nature of the musical hearing. The note can be considered the smaller arc, and each motif and melody as an arc as well. Each beat and subdivision, each measure and music section, constitutes an arc. A music in which the arcs do not present consistency with one another, can be understood as a music with no coherence. Coherence impression comes, mostly, from the skilled handling of arcs in a music piece.

Musical arcs are abstract structures and amenable to basic operations. A spectral arc, like a chord, can be inverted, enlarged and permuted, to mention a few possibilities. Temporal arcs, like a melody, a motif, a measure or a note, are also capable of variations. Recalling 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

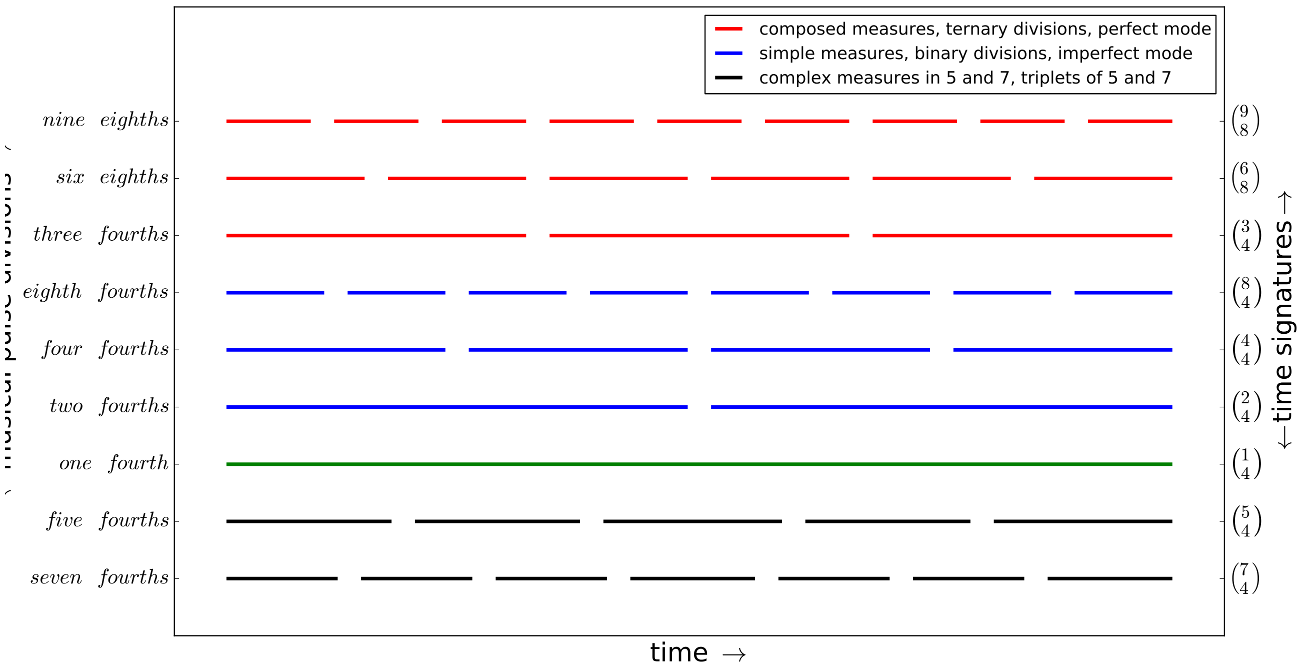

FIG. 24. Divisions and groupings of the musical pulse for establishing a metric. Divisions of the quarter note, regarded as the pulse, is presented on the left. The time signature yielded by groupings of the music pulse is presented on the right.

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: $\{s'_j\} = \left\{s_j^{\Gamma'}\right\} = \{s_j^{\Gamma+\delta}\}$ where Γ is the duration between the beginning of the piece (or considered section) and the first event s_0 of original structure S_j. One should 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 : $s'^{\Delta}_j = s_j^{\mu_j.\Delta}$. 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: $S'_j = \{s'_j\}_0^{H-1} = \{s_{(H-j-1)}\}_0^{H-1}$.

- Pitch translation, or transposition, is a displacement τ of the material to a different pitch $\Xi = \Xi_0 + \tau$. It is a variation that changes pitch localization of material: $\{s'_j\} = \left\{s_j^{\Xi'}\right\} = \{s_j^{\Xi+\tau}\}$ where Ξ_0 is a reference value, such as the pitch of a section S_j or of the first event s_0. If τ is given in semitones, the displacement in frequency is $\tau_f = f_0.2^{\frac{\tau}{12}}$ where f_0 is the reference frequency value: $f_0 = \Xi_{f_0} Hz \sim \Xi$ absolute value of pitch. The frequency of any pitch value is $f = f_0.2^{\frac{\tau}{12}}$s; and the pitch of any frequency is: $\Xi = \Xi_0 + 12\log_2\left(\frac{f}{f_0}\right)$. In the MIDI protocol, $\Xi_{f_0} = 55Hz$ while pitch $\Xi_0 = 33$ marks a *A 1*. Another good MIDI reference is $\Xi_{f_0} = 440Hz$ and $\Xi_0 = 69$. The difference $(\Xi_1 - \Xi_2)$ is measured in semitones. Ξ is not a measure in semitones: $\Xi = 1$ is not a semitone, it is a note with an audible frequency as rhythm, with less than 9 occurrences each second (see table III).

- Interval inversion is: 1) the inversion of note pitch order, within octave equivalence, such as described in IV A 2, and $S'_j = \{s'_j\}_0^{H-1} = \left\{s_j^{\varepsilon.f_0}\right\}$ with selective $s_j = 2,\ 1/2,$ or 1; or 2) the inversion of interval orientation. In the former case, the number of semitones is preserved in the "strict inversion": $S'_j = \{s'_j\}_0^{H-1} = \left\{s_j^{-\varepsilon_j.f_0}\right\}$, where ε_j is the factor between the frequency of event s_j and the frequency of s_0. The inversion is said tonal if the distances are considered in terms of the diaonic scale E_k: $S'_j = \{s'_j\}_0^{H-1} = \left\{ s_j^{2^{\left(\frac{12-e_{(7-j_e)}}{12}\right)}.f_0} \right\}_0^{H-1}$

where j_e is the index in E_k of the note s_j.

- Rotation of musical elements is the translation of all elements a number of positions ahead or behind, with the care to fill empty positions with events out of convenient slots. Thus, a rotation of $\tilde{n}$ positions is $S'_n = S_{(n+\tilde{n})\%H}$. If $\tilde{n} < 0$, it is sufficient to use $\tilde{n}' = H - \tilde{n}$. It is reasonable to associate $\tilde{n} < 0$ (events advance) with the clockwise rotation and $\tilde{n} < 0$ (elements delay) with the anti-clockwise rotation. More information about rotations in presented in subsection IV G.

- The insertion and removal of material in 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 silence. Elements of R_j can be inserted at the beginning, like a prefix in S_j; at the end, as a suffix; or in the middle, dividing S_j and making it the prefix and 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 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.

From these presented processes, others are derived, as the inverted retrograde, the temporal contraction with an external suffix, etc. A whole process of mental and neurological activity is unleashed, responsible for feelings, memories and imaginations, typical of a diligent music listening. This cortical activity is critical to musical therapy, known by its utility in cases of depression and neurological injury. It is known that regions of the human brain responsible by sonic processing are also used for other activities, such as language-related and mathematics[1,66].

Paradigmatic structures guide the creation of new musical material. One of the most central is the tension/relaxation dipole. Traditional dipoles are relate to tonic/dominant, repetition/variation, consonance/dissonance, coherence/rupture, symmetry/asymmetry, equivalence/difference, arrival/departure, near/far, stationary/moving, etc. Ternary relations tend to relate to the circle and to 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 reaches the apex and the second returns from apex to start region. This apex is called climax by

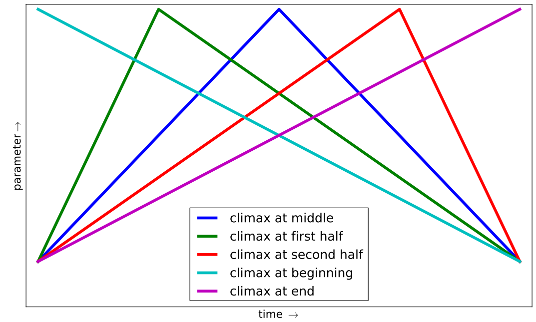

FIG. 25. Canonical distinctions of musical climax in a given melody and other domains. The possibilities considered are: climax at the beginning, at the first half, in the middle, in the second half and at the end. The x and y-axis is not properly specified since the parameters can be non-existent, as in a reference structure.

traditional music theory. It is possible to distinguish between arcs whose climax is placed at the beginning, middle, end, and at the first and second half of the duration. These structures are shown in Figure 25. The varying parameter can be non-existent, a case in which the arc consists only of a reference structure[18].

Consider the $S_i = \{s_i\}_0^{H-1}$ increasing sequence. The sequence $R_i = \{r_i\}_0^{2H-2} = \left\{s_{(H-1-|H-1-i|)}\right\}_0^{2H-2}$ presents perfect specular symmetry, i.e. the second half is the mirrored version of the first. In musical terms, the climax is in the middle of the sequence. It is possible to modify this by using sequences with different sizes. All the mathematics of sequences, already well established and taught routinely in calculus courses, can be used to generate these arcs[18,67]. Theoretically, when applied to any characteristic of musical events, these sequences produce arcs, since they imply a deviation and return of an initial parametrization. Henceforth, it is possible for a given sequence to have numberous distinct arcs, with different sizes and climax. This is an interesting and useful resource, and the correlation of arcs yields coherence[59].

In practice, and historically, there is special incidence and use of the golden ratio. The Lucas sequence allows the generalization of Fibonacci sequence, making it easy to follow. 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. If $x_0 = 1$ and $x_1 = 100$, and $y_n = \frac{x_n}{x_{n+1}}$ an auxiliary sequence, the error for the fraction of first values with respect 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 exactly the same error progression, but starts at the second step of the most discrepant case ($\frac{1}{1} \approx \frac{100+1=101}{100}$).

The music piece *Dirracional* uses arcs into directional structures. Its source code is available online as part of MASSA toolbox[33].

G. Cyclic structures

The philosophical understanding that human thought is founded on the concept of similarities and differences (e.g. perceived in stimuli and objects), places symmetries at the core of cognitive processes[68]. Mathematically, symmetries are algebraic groups, and a finite group is always isomorphic to a permutation group (by Cayley's theorem). In a way, this states that permutations represent any symmetry in a finite system[69]. In music, permutations are ubiquitous in scholastic techniques, which confirms their central role. The successive application of permutations generates cyclic arcs[17,70,71]. Two academic studies were dedicated to generating musical structures[72,73]. Any permutation set can be used as a generator of algebraic groups[71]. The properties defining a group G are:

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

From the first property follows that two permutations act as one permutation. In fact, it is possible to apply a permutation p_1 and another permutation p_2, and, comparing both initial and final orderings, observe 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 the element order. Thus, a finite permutation p, successively applied, reaches the initial ordering of its elements, and makes a cycle. This cycle, if used for parameters of musical notes or structures, yields a cyclic musical arc.

These arcs can be established by using a set of different permutations. As a historic example, the *change ringing* tradition conceives music through bells played one after another 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 presents a traditional *peal*, named "Hunt", 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, music structure consists of permutations itself and some different permutations operate in the cyclic behavior.

TABLE IV. Change Ringing: Traditional *peal* for 3 bells. Permutations occur between each line. Each line is a bell ordering and each ordering is played at 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: with $S_i = \{s_i\}$ 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. Permutations have two notations: cyclic and natural. The natural notation basically indicates the order of the resulting indexes from the permutation. Thus, given the original ordering of the sequence by its indexes [0 1 2 3 4 5 ...], the permutation is noted by the sequence of indexes it produces (ex. [1 3 7 0 ...]). In the cyclic notation, a permutation consists of swaps by elements and its successors, and the last element by the first one. E.g. $(1, 2, 5)(3, 4)$ in cyclic notation is equivalent to $[0, 2, 5, 4, 3, 1]$ in natural notation.

In the auralization of a permutation, it is not necessary to permute elements of S_i, but only some characteristic. Thus, if p is a permutation and S_i is a sequence of basic notes as in the end of section II F, the sequence $S_i' = p^f(S_i) = \left\{s_i^{p(f)}\right\}$ consists of the same music notes, following the same order and maintaining the same characteristics, but with the fundamental frequencies permuted according to the pattern of p.

Two subtleties of this procedure should be commented upon. First, a permutation p is not restricted to involve all elements of S_i, i.e. it can operate in a subset of S_i. Second, not all elements s_i need to be executed at each access to S_i. To exemplify, let S_i be 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 other measures where permutations allocate those notes to the first four events s_i' of S_i'.

To each permutation p_i described above, we have to determine: 1) note dimensions where it operates (frequency, duration, *fades*, intensity, etc); and 2) the period of incidence (how much consults before a permutation is applied). During realization of notes in S_i, an easy and coherent form is to execute the first n notes, the execution of disjoint sets of S_i is the same as modifying the permutation and executing the first n notes.

The MASSA toolbox presents a computational implementation that isolates permutation aplication to sonic characteristics, in order to deliver musical structures[33,72,73].

H. Musical idiom?

In numerous studies, there are models and explorations of a ‘musical language’, many can somewhat be considered ‘linguistics applied to music’ and some even discern different ‘musical idioms’[14,32,59,64]. In a simple way, a musical idiom is the result of chosen material together with repetition of elements and of relations between elements, along a music piece. In these matters, dichotomies are prominent, as explained in subsection IV E: repetition and variation, relaxation and tension, equilibrium and instability, consonance and disonance, etc.

I. Musical usages

First, the basic note was defined and characterized in quantitative terms (section II). Next, the internal note composition was addressed and both internal transitions and immediate sonic treatment were understood (section ??). Finally, this section aims at organizing these notes in music. The gamma of resources and consequent infinitude of praxis possibilities is a typical and highly relevant situation for artistic contexts[13,32].

There are studies for each of the presented resources. For example, it is possible to obtain ‘dirty’ triadic harmonies (with notes out of the triad) by superposition of perfect fourths. Another interesting example is the simultaneous presence of rhythms in different metrics, constituting what is called *polyrhythm*. The music piece *Poli-hit mia* in MASSA toolbox[33] explores these simultaneous metrics by impulse trains convolved with notes.

Microtonal scales are important for 20th century music[55] and present diverse remarkable results, e.g. fourths of tone in the Indian music. The musical sequence *MicroTom* in MASSA toolbox[33] explores these resources, including microtonal melodies and microtonal harmonies with many notes in a very reduced note scope.

As in subsection III F, relations between parameters are powerful ways to acquire musical pieces. The number of permuted notes can vary during the music, revealing relationship with piece duration. Harmonies can be made from 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 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 a dedicated listening. This is an instrumental piece, not included as a source code but available online[74].

PPEPPS (Pure Python EP: Projeto Solvente) is an EP synthesized using resources presented in this study. With minimal parametrization, the scripts generate complete musici pieces, allowing easy composition of sets of musics. A simple script of a few lines specifies music delivered as 16 bit 44.1kHz PCM files (WAVE). This facility and technological demystification create aesthetical possibilities for both sharing and education.

V. CONCLUSIONS AND FURTHER DEVELOPMENTS

This concise presentation relates basic musical elements to digital sound. The reader is invited to access *Scripts* and MASSA, where these relations are computationally implemented. The didactic report along the paper and the supplied scripts eases and encourages the use of the proposed framework.

The possibilities provided by these results include psychoacoustic experiments, interfaces for the generation of noise and other high fidelity sounds and musical structures, and for artistic and didactic purposes. The incorporation of programming skills is facilitated by visual aids, which was explored by *live-coding* practices and courses based on this framework, with success.

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, such as reviewed in [4].

The free software license and online availability of the exposed content as hypertext, with the respective codes and sonic examples, strongly facilitates future collaborations and the generation of sub-products in a co-authorship fashion. As consequence, the expansion of MASSA is favoured and straightforward, easing new implementarions and musical pieces development.

In addition, this framework permitted the formation of interests groups, with topics such as 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, georeferencing techniques, artistic and educational activities. Some of these reports are available online[4]. There are more than 700 videos, written documentats, original software applications and contributions in well-known external softwares such as Firefox, Scilab, LibreOffice, GEM/Puredata, to name a few[75–77].

Future work includes application of these results in machine learning and artificial intelligence methods for the generation of appealing artistic materials. Some psychoacoustical effects were detected, which needs validation and should be reported, specially with[78]. Other foreseen advances are: JavaScript version of MASSA toolbox, better hypermedia deliverables of this framework, use guides for different ends (musical composition, psychophysical experiments, sound synthesis, education, etc), creation of musical pieces and open experiments to be studied with EEG recordings and further analitical specification of musical elements in the discrete-time representation of sound as community feedback corroborate. .

[1] J. G. Roederer, *The Physics and Psychophysics of Music: An Introduction*, fourth edition ed. (Springer, 2008).

[2] A. Oppenheim, R. Schafer, J. Buck, *et al.*, *Discrete-time signal processing*, Vol. 2 (Prentice hall Englewood Cliffs, NJ:, 1989).

[3] Music from the twentieth century enlarged this traditional comprehension of music. This occurred in the concert music, especially in the concrete, electronic and electroacoustic styles. In the 1990s, it became evident that popular music had also incorporated sounds without defined pitch and temporal organization, to name a few of simple characteristics.

[4] R. FABBRI, *Msica no udio digital : descrio psicofsica e caixa de ferramentas [online]*, Ph.D. thesis, Instituto de Fsica de So Carlos, Universidade de So Paulo (2013).

[5] S. Chacon, J. Hamano, and S. Pearce, *Pro Git*, Vol. 288 (Apress, 2009).

[6] T. Oliphant, *A Guide to NumPy*, Vol. 1 (Trelgol Publishing USA, 2006).

[7] "Audiolab, a python package to make noise with numpy arrays," .

[8] G. V. Rossum and F. L. Drake Jr, *Python tutorial* (Odense Universitet, Institut for Matematik og Datalogi, 1995).

[9] G. Van Rossum and F. Drake Jr, *Python reference manual* (Centrum voor Wiskunde en Informatica, 1995).

[10] E. Raymond, *The art of Unix programming* (Addison-Wesley Professional, 2004).

[11] L. Lessig, "Free culture," Retrieved February **5**, 2006 (2002).

[12] J. M. Wisnik, *O som e o sentido* (Companhia das Letras, 1999).

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

[14] F. Lerdahl and R. Jackendoff, *A Generative Theory of Tonal Music* (MIT Press, 1983).

[15] P. R. Cook, *Real sound synthesis for interactive applications* (A K Peters, Natick, Massachusetts, 2002).

[16] O. Lacerda, *Compndio de Teoria Elementar da Msica*, 9th ed. (Ricordi Brasileira, 1966).

[17] J. Zamacois, *Curso de formas musicales: Con numerosos ejemplos musicales* (Barcelona : Idea Books, 2002).

[18] A. Schoenberg and E. Seincman, *Fundamentos da composição musical* (Edusp, 1990).

[19] C. Roads, *Microsound* (MIT press, 2004).

[20] W. LOVELOCK, *A concise history of music. Reprinted with revised record list* (1972).

[21] J. M. Chowning, "Digital sound synthesis, acoustics and perception: A rich intersection," Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00) (2000).

[22] Timbre is a subjective and complex characteristic. Physically, the timbre is multidimensional and given by the temporal dynamics of energy in the spectral components that are harmonic or noisy. In addition, the word timbre is used to designate different things: one same note can have (be produced with) different timbres, an 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 mentioning that timbre is not always manifested in spectral traces, since cultural or circumstantial aspects alter our perception of timbre.

[23] The factor $\frac{1}{\Lambda}$ could be distributed among the Fourier transform and its reconstruction, as preferred.

[24] Also called *bias*.

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

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

[27] F. Menezes, *A Acústica Musical em Palavras e Sons* (Ateliê Editorial, 2004).

[28] C. Cheng and G. Wakefield, "Introduction to head-related transfer functions (hrtfs): Representations of hrtfs in time, frequency, and space," Watermark **1** (2012).

[29] "Perception lecture notes: Auditory pathways and sound localization," .

[30] V. Algazi, R. Duda, D. Thompson, and C. Avendano, "The cipic hrtf database," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the* (IEEE, 2001) pp. 99–102.

[31] B. carty and V. lazzarini, "binaural hrtf based spatialisation: new approaches and implementation," in *dafx 09 proceedings of the 12th international conference on digital audio effects, politecnico di milano, como campus, sept. 1-4, como, italy* (Dept. of Electronic Engineering, Queen Mary Univ. of London,, 2009) pp. 1–6.

[32] A. Schoenberg and M. Maluf, *Harmonia* (Ed. UNESP, 1999).

[33] "`MASSA`: música e áudio em sequências e séries amostrais," .

[34] G. Geiger, "Table lookup oscillators using generic integrated wavetables," Proc. of the 9th Int. Conference on Digital Audio Effects (DAFx-06) (2006).

[35] R. Bristow-Johnson, "Wavetable synthesis 101, a fundamental perspective," in *Proc. AES Convention*, Vol. 101 (1996).

[36] S. Dehaene, "The neural basis of the weber–fechner law: A logarithmic mental number line," Trends in cognitive sciences **7**, 145–147 (2003).

[37] The volume (psychophysical quality) is a consequence of different sound characteristics, like reverberation and concentration of high harmonics, among which is the wave total energy. Wave energy is easy to modify (see Equation 2) and can be varied in many ways. The simplest way consists of modifying the amplitude by multiplying the whole sequence by a real number. The increase of energy without amplitude variation is the *sound compression*, quite popular nowadays for music production[38].

[38] P. Guillaume, *Music and acoustics: from instrument to computer* (Iste, 2006).

[39] The implementation of filters encompasses an area of recognized complexity, with dedicated literature and software implementations. The reader is welcome to visit the bibliography for further discussion[2,44].

[40] 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.[2].

[41] 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: $\mathbb{H}(z) = \frac{a_0+a_1.z^{-1}+a_2.x^{-2}}{1-b_1.z^{-1}-b_2.z^{-2}}$.

[42] Butterworth and Elliptical filters can be considered as special cases of Chebichev filters[2,44].

[43] J. O. Smith III, *Mathematics of the discrete fourier transform (dft) with audio applications*, 2nd ed. (2012).

[44] S. Smith, "The scientist and engineers guide to digital signal processing, 1999," (2009).

[45] B. Class, F. Java, A. Wah, L. Crusher, P. Waveshaper, S. Enhancer, and S. Matrix, "musicdsp. org source code archive," .

[46] Although its origin is disparate with its color association, this noise became established with this specific name, in musical contexts. Anyway, this association can be considered satisfactory once violet, blue, white and pink noises are more strident and associated with more intense colors[15,38].

[47] The jargon may be different in other contexts. For example, in piano, tremolo is a vibrato in the classification used here. The definitions given here are common in contexts regarding music theory and electronic music. In addition, they are based on a broader literature than the one used for a specific instrument, practice or musical tradition[16,32].

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

[49] T. Porres, J. Manzolli, and F. Furlanete, "Análise de dissonância sensorial de espectros sonoros," in *Congresso da ANPPOM*, Vol. 16 (2006).

[50] A. Porres and A. Pires, "Um external de aspereza para puredata & max/msp," in *Proceedings of the 12th Brazilian Symposium on Computer Music 2009* (2009).

[51] "An introduction to fm (snd manual)," .
[52] J. Gramani, *Rítmica viva: a consciência musical do ritmo* (UNICAMP, 1996).
[53] C. Roads, *The computer music tutorial* (MIT press, 1996).
[54] Micro-tonality is the use of intervals less than one semitone and has ornamental and structuring functionalities in music. The division of the octave in 12 notes is motivated by physical fundaments but also a *convention* adopted by western classical music. Other tunings are incident. By way of illustration, the traditional Thai music uses an octave division in seven notes equally spaced ($\varepsilon = 2^{\frac{1}{7}}$), which allows intervals that have little in common with the intervals in the diatonic scale[12].
[55] S. Wilkinson, K. Laubach, R. Schiff, and J. Eiche, *Tuning in: Microtonality in Electronic Music: a Basic Guide to Alternate Scales, Temperaments, and Microtuning Using Synthesizers* (H. Leonard Books, 1988).
[56] J. Clough, N. Engebretsen, and J. Kochavi, "Scales, sets, and interval cycles: A taxonomy," Music Theory Spectrum , 74–104 (1999).
[57] S. Kostka, J. Clendinning, R. Ottman, and J. Phillips, *Tonal Harmony: With an Introduction to Twentieth* (McGraw-Hill, 1995).
[58] H. Koellreutter, "Harmonia funcional," (1986).
[59] F. Salzer, *Structural hearing: Tonal coherence in music*, Vol. 1 (Dover publications, 1962).
[60] E. Aldwell, C. Schachter, and A. Cadwallader, *Harmony & voice leading* (Wadsworth Publishing Company, 2010).
[61] J. Fux and A. Mann, *The study of counterpoint from Johann Joseph Fux's Gradus ad Parnassum*, Vol. 277 (WW Norton & Company, 1965).
[62] L. Tragtenberg, *Contraponto: uma arte de compor* (Edusp, 2002).
[63] A. Schoenberg and L. Stein, *Preliminary exercises in counterpoint* (Faber & Faber, 1963).
[64] G. O. A. de Assis, *Em busca do som* (Editora UNESP).
[65] P. Boulez, *A música hoje* (1972).
[66] O. Sacks, *Musicophilia: Tales of Music and the Brain, Revised and Expanded Edition.* (Knopf Doubleday Publishing Group, 2008).
[67] H. Guidorizzi, *Um curso de cálculo* (Livros Técnicos e Científicos Editora, 2001).
[68] G. Deleuze, *Difference and Repetition* (Continuum, 1968).
[69] F. J. Budden, *The fascination of groups* (Cambridge Univ. Press, 1972).
[70] R. Duckworth and F. Stedman, *Tintinnalogia, or, the Art of Ringing* (Echo Library, 2007).
[71] F. Budden and F. Budden, *The fascination of groups* (Cambridge University Press London, 1972).
[72] R. F. e Adolfo Maia Jr., "Applications of group theory on granular synthesis," *SBCM 2007 Proceedings*, (2007).
[73] R. F. e Adolfo Maia Jr., "Applications of group theory on sequencing and spatialization of granular sounds," *Anais do 6o Congresso de Engenharia de udio da AES Brasil*, (2008).
[74] "3 trios para obo, flauta e fagote,".
[75] "Página principal do lab macambira," .
[76] "Wiki do lab macambira," ().
[77] "Canal vimeo do lab macambira (mais de 700 videos)," .
[78] R. F. et al., "Quadros sonoros," `https://github.com/ttm/dissertacao/blob/master/article/scripts/pecas2.1/quadrosSonoros.py`.
[79] F. de Saussure, *Course in General Linguistics* (Books LLC, 1916).
[80] A. P. S. U. Pillai, *Probability, Random Variables and Stochastic Processes* (McGraw Hill Higher Education, 2002).
[81] R. A. J. D. W. Wichern, *Applied Multivariate Statistical Analysis* (Prentice Hall, 2007).
[82] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing* (Prentice Hall, 1992).
[83] R. O. D. P. E. Hart D. G. Stork, *Pattern Classification* (Wiley-Interscience, 2000).
[84] L. da F. Costa R. M. C. Jr., *Shape Analysis and Classification: Theory and Practice (Image Processing Series)* (CRC Press, 2000).
[85] D. Papineau, *Philosophy* (Oxford University Press, 2009).
[86] B. Russel, *A History of Western Philosophy* (Simon and Schuster Touchstone, 1967).
[87] F. G. G. Deleuze, *What Is Philosophy?* (Simon and Schuster Touchstone, 1991).
[88] A. Walker, *Hans von Bülow: a life and times* (Oxford University Press, 2010).
[89] R. Bennett, *History of Music* (Cambridge University Press, 1982).
[90] A. Kozbelt, "Performance time productivity and versatility estimates for 102 classical composers," Psychology of Music **37**, 25–46 (2009).
[91] A. Kozbelt, "A quantitative analysis of Beethoven as self-critic: implications for psychological theories of musical creativity," Psychology of Music **35**, 144–168 (2007).
[92] D. K. Simonton, "Emergence and realization of genius: The lives and works of 120 classical composers," Journal of Personality and Social Psychology **61**, 829 – 840 (1991).
[93] D. 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).
[94] P. van Kranenburg, "Musical style recognition – a quantitative approach," in *Proceedings of the Conference on Interdisciplinary Musicology (CIM04)* (2004).
[95] P. 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).
[96] J. H. W. Jr, "Hierarchical grouping to optimize an objective function," Journal of the American Statistical Association **58**, 236–244 (1963).
[97] H. Varian, "Bootstrap tutorial," The Mathematica Journal **9** (2005).
[98] C. P. Robert, "Simulation in statistics," in *Proceedings of the 2011 Winter Simulation Conference* (2011) arXiv:1105.4823.
[99] J. Cohen *et al.*, "A coefficient of agreement for nominal scales," Educational and psychological measurement **20**, 37–46 (1960).
[100] D. Boos, "Introduction to the bootstrap world," Statistical Science **18**, 168–174 (2003).
[101] R. Fabbri and R. S. Mendes, "Compressão de áudio via wavelets, aproximações polinomiais e permutações," *Anais EADCA 2009*, (2009).
[102] V. Vieira, R. Fabbri, G. Travieso, O. Oliveira Jr, and L. da Fontoura Costa, "A quantitative approach to evolution of music and philosophy," Journal of Statistical Mechanics: Theory and Experiment **2012**, P08010 (2012).
[103] D. Amancio, R. Fabbri, O. Oliveira Jr, M. Nunes, and L. da F Costa, "Distinguishing between positive and negative opinions with complex network features," in *Proceedings of the 2010 Workshop on Graph-based Methods for Natural Language Processing* (Association for Computational Linguistics, 2010) pp. 83–87.
[104] R. Fabbri, D. Amancio, L. da F. Costa, and O. N. O. O. Jr., "Speech polarity detection using complex networks measures: First explorations," in *Complex Networks: Second International Workshop, CompleNet 2010, Rio de Janeiro, Brazil, October 13-15, 2010, Revised Selected Papers*.
[105] D. Amancio, R. Fabbri, O. Oliveira, M. Nunes, and L. Fontoura Costa, "Opinion discrimination using complex network features," Complex Networks , 154–162 (2011).
[106] C. Newman, C. Frisbie, D. da Silva, J. Bredas, P. Ewbank, and K. Mann, "Introduction to organic thin film transistors and design of n-channel organic semiconductors," Chemistry of Materials **v.16**, 4436–4451 (2004).
[107] C. Newman, C. Frisbie, and D. da SILVA, "Introduction to organic thin film transistors and design of n-channel organic semiconductors," Chemistry of Materials **v. 16**, 4436–4451 (2004).
[108] C. Dimitrakopoulos and D. Mascaro, "Organic thin-film tran-

sistors: a review of recent advances," IBM Journal of Research and Development **v. 45**, 11–27 (2001).
[109]C. Kittel, *Introduction to solid state physics*, 8th ed. (John Willey & Songs, Hoboken, NJ, 2005).
[110]S. R. Alten, *Audio in media* (Wadsworth Publishing, 2010).
[111]F. A. Everest, *Master Handbook of Acoustics* (McGraw-Hill, 2001).
[112]E. Zampronha, *Notação, Representação e Composição* (Annablume, 2000).
[113]S. Emmerson, *Living electronic music* (Ashgate Publishing Company, 2007).
[114]J. O. Smith III, *Physical Audio Signal Processing (for virtual musical instruments and audio effects)* (2006).
[115]J. O. Smith III, *Introduction to Digital Filters (with audio applications)* (2007).
[116]J. O. Smith III, *Spectral Audio Signal Processing* (W3K Publishing, 2011).
[117]J. O. Smith III, *Audio Signal Processing in FAUST* (2012).
[118]"Book announcement 13 may 1997," .
[119]"Timeline of operating systems on wikipedia," `http://tinyurl.com/rcuxwj`.
[120]"Quadros sonoros," .
[121]"Audioexperiments," .
[122]"Dissertation about music in digital audio: psychophysics description and software toolbox," .
[123]"reduced-fi," .
[124]"Sound localization," ().