Musical elements in the discrete-time representation of sound

RENATO FABBRI, University of São Paulo
VILSON VIEIRA DA SILVA JUNIOR, The Grid
ANTÔNIO CARLOS SILVANO PESSOTTI, Universidade Metodista de Piracicaba
DÉBORA CRISTINA CORRÊA, ??
OSVALDO N. OLIVEIRA JR., University of São Paulo

The representation of basic elements of music - such as notes, ornaments and intervals - in terms of discrete audio signals is often used in software for music creation and design. Nevertheless, there is no unified approach that relates these elements to the discrete samples of digitized sound. In this article, each musical element is related by equations and algorithms to the discrete-time samples of sounds, and each of these relations are implemented in scripts within a software toolbox, referred to as MASS (Music and Audio in Sample Sequences). The toolbox also includes this article and every script necessary to render the images and musical examples, pieces and albums, which adds to the educational and artistic uses. The fundamental element, the musical note with duration, volume, pitch and timbre, is related quantitatively to characteristics of the digital signal. Internal variations of a note, such as tremolos, vibratos and spectral fluctuations, are also considered, which enables the synthesis of notes inspired by real instruments and new sonorities. With this representation of notes, resources are provided for the generation of higher scale musical structures, such as rhythmic meter, pitch intervals and cycles. This framework enables precise and trustful scientific experiments and is useful for education and art. The efficacy of MASS is confirmed by the synthesis of small musical pieces using basic notes, elaborated notes and notes in music, which reflects the organization of the toolbox and of this article. It is possible to synthesize whole albums through collage of the scripts and settings specified by the user. With the open source paradigm, the toolbox can be promptly scrutinized, expanded in co-authorship processes and used with freedom by musicians, engineers and other interested parties. In fact, MASS has already been employed by external users for diverse purposes which include music production, artistic presentations, psychoacoustic experiments and computer language diffusion where the appeal of audiovisual artifacts is exploited for education.

CCS Concepts: •**Applied computing** →**Sound and music computing;** •**Computing methodologies** →*Modeling methodologies;* •**General and reference** →*Surveys and overviews; Reference works;*

Additional Key Words and Phrases: music, acoustics, psychophysics, digital audio, signal processing

ACM Reference format:
Renato Fabbri, Vilson Vieira da Silva Junior, Antônio Carlos Silvano Pessotti, Débora Cristina Corrêa, and Osvaldo N. Oliveira Jr.. 0. Musical elements in the discrete-time representation of sound. *ACM Comput. Surv.* 0, 0, Article 0 (0), 56 pages.
DOI: 0000001.0000001

This work is supported by FAPESP and CNPq.

DOI: 0000001.0000001

1 INTRODUCTION

Music is usually defined as the art whose medium is sound. The definition might also state that the medium includes silences and temporal organization of structures, or that music is also a cultural activity or product. Sound is a physical phenomenon which corresponds to the longitudinal wave of mechanical pressure. The human auditory system perceives sounds in the frequency bandwidth between $20Hz$ and $20kHz$, with the actual boundaries depending on the person, climate conditions and the sonic characteristics themselves [54]. Since the speed of sound is $\approx 343.2m/s$, such frequency limits corresponds to wavelengths of $\frac{343.2}{20} \approx 17.16\,m$ and $\frac{343.2}{20000} \approx 17.16\,mm$. Hearing involves stimuli in bones, stomach, ears, transfer functions of head and torso, and processing by the nervous system [54]. 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 recognize in the constitution of sounds of musical interest (such as harmonic sounds and noises, discussed in Sections 2 and 3), and higher level musical structures (such as tunings, scales and chords, in Section 4).

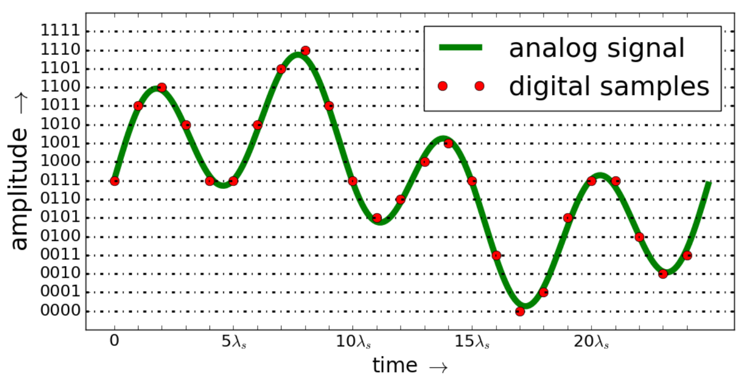

Fig. 1. Example of PCM audio: a sound wave is represented by 25 samples equally spaced in time where each sample has an amplitude specified with 4 bits.

The representation of sound can take many forms, from musical scores and texts in a phonetic language to electric analog signals and binary data. It includes arbitrary sets of features and spectral decompositions such as obtained through wavelets. Although the terms 'audio' and 'sound' are often used without distinction and 'audio' has many definitions which depend on the context and the author, audio most often means a representation of sound by means of the wave amplitude through time. In this sense, audio expresses sonic waves from direct synthesis or input by microphones, although these sources are not always neatly distinguishable e.g. as captured sounds are processed to generate new sonorities. Digital audio protocols often implies quality loss (to yield smaller files, ease storage and transfer) and are called *lossy* [46]. This is the case e.g. of MP3 and Ogg Vorbis. Non-lossy representations of digital audio, called *lossless* protocols or formats, on the other hand, assures perfect reconstruction of the analog wave within any convenient precision. The standard paradigm of lossless audio consists of representing the sound with samples equally spaced by a duration δ_s, and specifying the amplitude of each sample by a fixed number of bits. This is the Pulse Code Modulation (PCM) representation of sound. A PCM audio format has two essential attributes: a sampling frequency $f_s = \frac{1}{\delta_s}$ (also called e.g. sampling rate or sample rate), which is the number of samples used for representing a second of sound; and a bit depth, which is the number of bits for the amplitude of each sample. Figure 1 shows 25 samples of a PCM audio with a bit depth of 4, which yields $2^4 = 16$ possible values for the amplitude of each sample and a total of $4 \times 25 = 100$ bits for representing the whole sound.

The fixed sampling frequency and bit depth yields the quantization error or quantization noise. This noise diminishes as the bit depth increases while greater sampling frequency allows higher frequencies to be represented. The Nyquist theorem [48] asserts that the sampling frequency is twice the maximum frequency that the represented signal can contain. Thus, for general musical purposes, it is necessary to use a sample rate of at least twice the highest frequency heard by humans, that is, $f_s \geq 2 \times 20kHz = 40kHz$. This is the basic reason for the adoption of sampling frequencies such as $44.1kHz$ and $48kHz$, which are the standards in Compact Disks (CD) and broadcast systems (radio and television), respectively.

Within this framework for representing sounds, musical notes can be characterized. The note often stands as the 'fundamental unit' of musical structures (such as atoms in matter or cells in macroscopic organisms) and, in practice, it can unfold into sounds that uphold other approaches to music. This is of capital importance because science and scholastic artists widened the traditional comprehension of music in the twentieth century to encompass discourse without explicit rhythm, melody or harmony. This is very evident in e.g. in the concrete, electronic, electroacoustic, and spectral musical styles. In the 1990s, it became evident that popular (commercial) music had also incorporated sound amalgams and abstract discursive arcs [1]. Notes are also convenient for another reason: the average listener – and a considerable part of the specialists – presupposes rhythmic and pitch organization (made explicit in Section 4) as fundamental musical properties, and these are developed in traditional musical theory in terms of notes. Thereafter, in this article we describe musical notes in PCM audio through equations and then indicate mechanisms for deriving higher level musical structures. We understand that this is not the unique approach to mathematically express music in digital audio, but musical theory and practical issues suggest that this is a proper framework for understanding and making computer music, as should become evident in the reminder of this text and be verifiable by usage of the MASS toolbox. Hopefully, the interested reader or programmer will be able to use this framework to synthesize music beyond traditional conceptualizations when intended.

This article aims at representing musical structures and artifices in discrete-time audio. The results include mathematical relations, usually in terms of musical characteristics and samples, concise musical theory considerations, and their implementations as software routines both as very raw algorithms and in the context of rendering musical pieces. Despite the general interests involved, there are only a 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 [31]. Articles on the topic appear to be lacking, to the best of our knowledge. Although current music software use the analytical descriptions presented here, there is no concise mathematical description of them, and it is far from trivial to achieve the equations by analyzing the available software implementations.

The objectives of this paper are:

(1) Present a concise set of mathematical and algorithmic relations between musical basic elements and sequences of PCM audio samples.
(2) Introduce a framework of sound and musical synthesis with control at sample level which entails potential uses in psychoacoustic experiments and synthesis with extreme precision (recap in Section 5).

[1]There are well known incidence of such characteristics in ethnic music, such as in Pygmy music, but western theory assimilated them only in the last century [72].

(3) Provide a powerful theoretical framework which can be used to synthesize musical pieces and albums.
(4) Provide approachability to the developed framework [2].
(5) Provide a didactic presentation of the content, which is highly multidisciplinary, involving signal processing, music, psychoacoustics and programming.

The reminder of this article is organized as follows: Section 2 characterizes the basic musical note; Section 3 develops internal dynamics of musical notes; Section 4 tackles the organization of musical notes in higher level musical structures [23, 40, 41, 53, 61, 70, 72, 74]. As these descriptions require knowledge on topics such as psychoacoustics, cultural traditions, and mathematical formalisms, the text points to external complements as needed and presents methods, results and discussions altogether. Section 5 is dedicated to final considerations and further work.

1.1 Additional material

The Supporting Information document [6] holds commented listings of all the equations, figures, tables and sections in this document and the scripts in the MASS toolbox[3]. The musical pieces are not very traditional which facilitates the understanding of specific techniques and the extrapolation of the note concept.

1.2 Synonymy, polysemy and theoretical frames (disclaimer)

Given that the main topic of this article (the expression of musical elements in PCM audio) is multidisciplinary and involves art, the reader should be aware that much of the vocabulary admit different choices of terms and definitions. More specifically, it is often the case where many words can express the same concept and where one word can carry different meanings. This is a very deep issue which might receive a dedicated manuscript. The reader might need to read the rest of this document to understand this small selection of synonymy and polysemy in the literature, but it is important to illustrate the point before the more dense sections:

- A "note" can mean a pitch or an abstract construct with pitch and duration or a sound emitted from a musical instrument or a specific note in a score or a music.
- The sampling rate (discussed above) is also called the sampling frequency or sample rate.
- An harmonic in a sound is most often a sinusoidal component which is in the harmonic series of the fundamental frequency. Many times, however, the terms harmonic and component are not distinguished. A harmonic can also be a note performed in an instrument by preventing certain overtones (components).
- Harmony can refer to chords or to note sets related to chords or even to "harmony" in a more general sense, as a kind of balance and consistency.

[2] All the analytic relations presented in this article are implemented as public domain scripts, i.e. small computer programs for prompt distribution and validation. This constitutes the MASS toolbox, available in an open source Git repository [18]. These scripts are written in Python and make use of Numpy, which performs numerical routines efficiently (e.g. through LAPACK). Part of the scripts have been transcribed to JavaScript (which favors their use in Web browsers such as Firefox and Chromium) and native Python [2, 47, 55, 68]. These are all open technologies, published using licenses that allow for copy, distribution and use for research and generating derivatives. Hence, the work presented here aims at being compliant with recommended practices for availability and validation and should ease co-authorship processes [42, 51].

[3] The toolbox contains a collection of Python scripts which:

- implement each of the equations;
- render music and illustrate the concepts;
- render each of the figures used in this article.

The documentation of the toolbox consists of this article, the Supporting Information and the scripts themselves.

- A “tremolo” can mean different things, e.g. in a piano score, a tremolo is a fast alternation of two notes (pitches) while in computer music theory it is (most often) an oscillation of intensity.

We strived to avoid nomenclature clashes and the use of more terms than needed. Also, there are many theoretical standpoints for the understanding of musical phenomena, which is an evidence that most often there is not a single way to express or characterize musical structures. Therefore, in this article, adjectives such as ”often”, ”commonly” and ”frequently” are abundant and they would probably be even more numerous if we wanted to be pedantically precise. Some of these issues are exposed when the context is convenient, such as in the first considerations of timbre.

2 CHARACTERIZATION OF THE MUSICAL NOTE IN DISCRETE-TIME AUDIO

In diverse artistic and theoretical contexts, music is conceived as constituted by fundamental units referred to as notes, “atoms” that constitute music itself [43, 70, 72]. In a cognitive perspective, notes are seen as discernible elements that facilitate and enrich the transmission of information through music [40, 54]. Canonically, the basic characteristics of a musical note are duration, loudness, pitch and timbre [40]. All relations described in this section are implemented in the file `src/sections/1.py`. The musical pieces *5 sonic portraits* and *reduced-fi* are also available online to corroborate the concepts.

2.1 Duration

The sample frequency f_s is defined as the number of samples in each second of the discrete-time signal. Let $T = \{t_i\}$ be an ordered set of real samples separated by $\delta_s = 1/f_s$ seconds ($f_s = 44.1kHz \Rightarrow \delta_s = 1/44100 \approx 23ms$). A musical note of duration Δ seconds can be expressed as a sequence T^Δ with $\Lambda = \lfloor \Delta.f_s \rfloor$ 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. Thus:

$$T^\Delta = \{t_i\}_{i=0}^{\lfloor \Delta.f_s \rfloor - 1} = \{t_i\}_0^{\Lambda-1} \tag{1}$$

2.2 Loudness

Loudness[4] is a sensation that depends on reverberation, spectrum and other characteristics described in Section 3. One can achieve loudness variations through the power of the wave [20]:

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

The final loudness is dependent on the amplification of the signal by the speakers. Thus, what matters is the relative power of a note in relation to the others around it, or the power of a music section in relation to the rest. Differences in loudness are the result of complex psychophysical phenomena but can often be reasoned about in terms of decibels, calculated directly from the amplitudes through energy or power:

$$V_{dB} = 10log_{10}\frac{pow(T^{'})}{pow(T)} \tag{3}$$

[4]Loudness and “volume” are often used indistinctly. In technical contexts, loudness is used for the subjective sensation of sound intensity while volume might be used for some measurement of loudness or to a change in the signal by equipment. Accordingly, one can perceive a sound as loud or soft and change the volume by turning a knob. We will use the term loudness and avoid the more ambiguous term volume.

The quantity V_{dB} has the decibel unit (dB). By standard, a "doubled loudness" is associated to a gain of $10dB$ (10 violins yield the double of the loudness of a violin). A handy reference is $10dB$ for each step in the musical intensity scale: *pianissimo*, *piano*, *mezzoforte*, *forte* and *fortissimo*. Other useful references are dB values related to double amplitude or power:

$$t_i^{'} = 2t_i \Rightarrow pow(T^{'}) = 4pow(T) \Rightarrow V_{dB}^{'} = 10log_{10}4 \approx 6dB \tag{4}$$

$$t_i^{'} = \sqrt{2}t_i \Rightarrow pow(T^{'}) = 2pow(T) \Rightarrow V_{dB}^{'} = 10log_{10}2 \approx 3dB \tag{5}$$

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

$$\begin{aligned} 10log_{10}\frac{pot(T^{'})}{pot(T)} = 10 \quad \Rightarrow \\ \Rightarrow \quad \sum_{i=0}^{\lfloor \Delta.f_s \rfloor -1} t_i^{'2} = 10\sum_{i=0}^{\Lambda-1} t_i^2 = \sum_{i=0}^{\Lambda-1}(\sqrt{10}.t_i)^2 \\ \therefore \quad t_i^{'} = \sqrt{10}t_i \quad \Rightarrow \quad t_i^{'} \approx 3.16t_i \end{aligned} \tag{6}$$

Thus, an amplitude increase by a factor slightly above 3 is required for achieving a doubled loudness. These values are guides for increasing or decreasing the absolute values in sample sequences. 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.

2.3 Pitch

The pitch is specified by the fundamental frequency f whose cycle has duration $\delta = 1/f$. This duration, multiplied by the sampling frequency f_s, yields in the number of samples per cycle: $\lambda = f_s.\delta = f_s/f$. For didactic reasons, let f divide f_s and result λ integer. If T^f is a sonic sequence with fundamental frequency f, then:

$$T^f = \left\{t_i^f\right\} = \left\{t_{i+\lambda}^f\right\} = \left\{t_{i+\frac{f_s}{f}}^f\right\} \tag{8}$$

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.

2.4 Timbre

A spectrum is said harmonic if all the frequencies it contains are (whole number) multiples of the fundamental frequency. A sound with harmonic spectrum has a wave period which corresponds to the wave cycle duration and is given by the inverse of the fundamental frequency. The trajectory of the wave inside the period is the *waveform* and defines the harmonic spectrum. The waveform

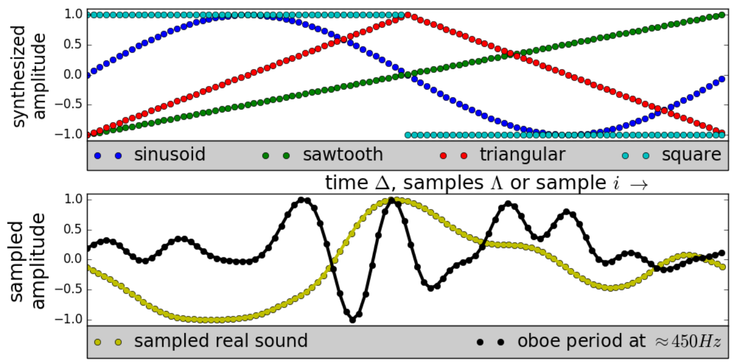

Fig. 2. Basic musical waveforms: (a) the basic synthetic waveforms; (b) real waveforms.

thus entails a timbre[5]. Sonic spectra with minimum differences can result in timbres with crucial differences and, consequently, distinct timbres can be produced using different spectra [54].

The simplest case is the spectrum with only one frequency, which is a sinusoid, often regarded as a "pure" oscillation (e.g. in terms of the *simple harmonic motion*). Let S^f be a sequence whose samples s_i^f describe a sinusoid with frequency f:

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

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

In a similar fashion, other waveforms are used in music for their spectral qualities and simplicity. While the sinusoid is an isolated node in the spectrum, any other waveform present a succession of harmonic components. The most basic waveforms are specified in Equations 9, 10, 11 and 12, and are 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 [48].

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

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

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

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

The square wave also has only odd harmonics but falling at $-6dB/octave$:

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

[5]The timbre of a sound is a subjective and complex characteristic. The timbre can be considered by a temporal evolution of energy in the spectral components that are harmonic or noisy (and by deviations of the harmonics from the ideal harmonic spectrum). 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. Timbre is not only about spectrum: culture and context alter our perception of timbre. [54]

[6]In musical jargon, an "octave" means a frequency f and twice such frequency $2f$ or the bandwidth $\{f, 2f\}$.

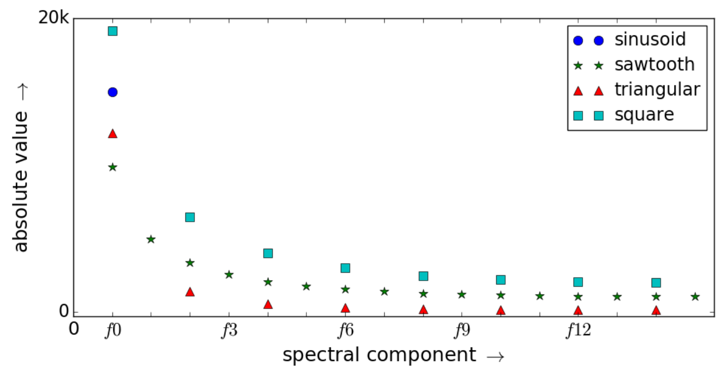

Fig. 3. Spectra of basic artificial musical waveforms.

The square wave can be used in a subtractive synthesis with the purpose of mimicking a clarinet. This instrument has only the odd harmonics and the square wave is convenient with its abundant energy at high frequencies. The sawtooth is a common starting point for 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 attenuation by 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 further processing. The sinusoid is often a nice choice, but a problematic one. While pleasant if not loud in a very high pitch (above $\approx 500Hz$ it requires careful dosage), the pitch of a pure sinusoid is not accurately detected by the human auditory system, particularly at low frequencies. Also, it requires a great amplitude gain for an increase in loudness of a sinusoid if compared to other waveforms. Both particularities are understood in the scientific literature as a consequence of the nonexistence of pure sinusoidal sounds in nature [54].

Figure 2 presents the waveforms described in Equations 9, 10, 11 and 12 for $\lambda_f = 100$ (i.e. period of 100 samples). If $f_s = 44.1kHz$, the PCM standard for Compact Disks, the wave has fundamental frequency $f = \frac{f_s}{\lambda_f} = \frac{44100}{100} = 441\ Hz$, 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. The figure exhibits the spectra described in the last paragraphs: the sawtooth is the only waveform with a complete harmonic series (odd and even components); triangular and 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 harmonic spectrum has notes different from the fundamental frequency (see Equation 85). 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 with frequency f_n is a sinusoidal oscillation in the constitution of the sound in that frequency f_n. This energy, specifically concentrated on f_n, is separated from other frequencies by the ear for further cognitive

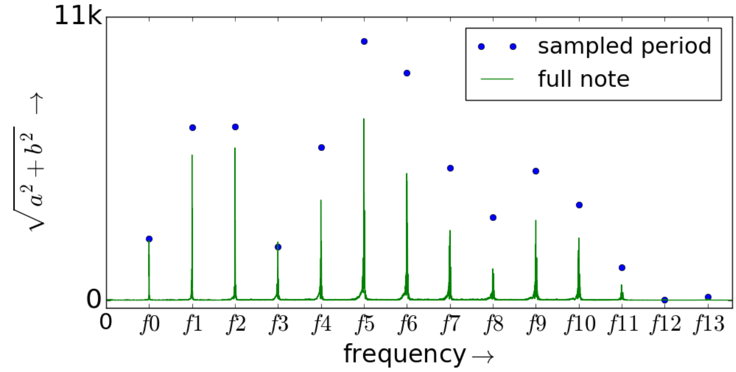

Fig. 4. Spectra of the sonic waves of a natural oboe note and obtained through a sampled period. The natural sound has fluctuations in the harmonics and in its noise, while the sampled period note has a perfectly harmonic (and static) spectrum.

processes (this separation is done by diverse living organisms by mechanisms similar to the human cochlea [54]).

The sinusoidal components are responsible for timbre qualities. If their frequencies do not relate by small integers, 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 [54]. For a fixed length period and waveform, the spectrum is perfectly harmonic and static. Each waveform has a different distribution of energy in the (same) harmonic components. High curvatures are a sign of energy in the high harmonics. Figure 2 depicts a wave, labeled as "sampled real sound", with a period of $\Lambda_f = 114$ samples. The oboe waveform was also sampled at $44.1kHz$. The period is 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 $R = \{r_i\}_0^{\lambda_f - 1}$ of samples in the real sound of Figure 2 can be taken as basis for a sound T^f in the following way:

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

The resulting sound has the spectrum of the original waveform. As a consequence of the identical repetitions, the spectrum is perfectly harmonic, without noise or variations typical of natural phenomena. This can be observed in Figure 4, which shows the spectrum of the original oboe note and a note with the same duration, whose samples consist of the repetition of the cycle on Figure 2. Summing up, the spectrum of a naturally produced note (e.g. by a musical instrument) exhibits variations in the frequencies of the harmonics, in their intensities and noise, while a note made from a sampled period has a perfectly harmonic and static spectrum.

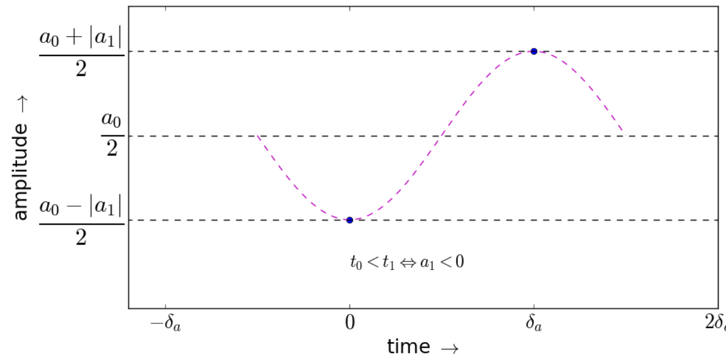

Fig. 5. Oscillation of 2 samples (maximum frequency for any f_s). The first coefficient reflects a constant detachment (called *offset* or *bias*) and the second coefficient specifies the oscillation amplitude.

2.5 Spectrum of sampled sound

These sinusoidal components in the discretized sound have some particularities. Considering a signal T and its corresponding Fourier decomposition $\mathcal{F}\langle T\rangle = C = \{c_i\}_0^{\Lambda-1} = \left\{\sum_{k=0}^{\Lambda-1} t_k e^{-j2\pi k i/\Lambda}\right\}_0^{\Lambda-1}$, the recomposition is the sum of the frequency components to yield the temporal samples[7]:

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

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 $f_k = w_k\frac{f_s}{2\pi} = \frac{f_s}{\Lambda}k$ in Hertz, taking into account the respective limits of π and $\frac{f_s}{2}$ given by the Nyquist Theorem.

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

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

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 $arctan\left(\frac{b_k}{a_k}\right)$ which has this image. The sign of 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 component. When there is only two samples, the Fourier decomposition has only one pair of coefficients $\{c_k = a_k - j.b_k\}_0^{\Lambda-1=1}$ relative to frequencies

[7]The factor $\frac{1}{\Lambda}$ can be distributed among the Fourier transform and its reconstruction, as preferred. Note that j here is the imaginary unit $j^2 = -1$.

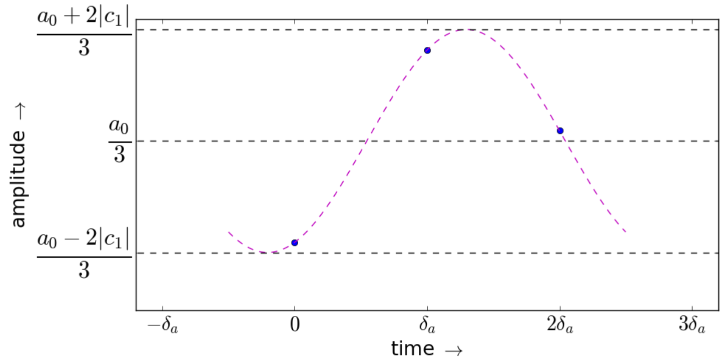

Fig. 6. Three fixed samples present only one non-null frequency. $c_1 = c_2^*$ and $w_1 \equiv w_2$.

$\{f_k\}_0^1 = \left\{w_k \frac{f_s}{2\pi}\right\}_0^1 = \left\{k\frac{f_s}{\Lambda=2}\right\}_0^1 = \left\{0, \frac{f_s}{2} = f_{\max}\right\}$ 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 (also called *bias*), and $\frac{a_1}{2}$ for the oscillation with frequency $f_1 = \frac{f_s}{\Lambda=2}$. This case has special relevance. At least 2 samples are necessary to represent an oscillation and it yields the Nyquist frequency $f_{\max} = \frac{f_s}{2}$, which is the maximum frequency in a sound sampled with f_s samples per second. In fact, any discrete-time signal has this property, not only digitized sound.

All fixed sequences T 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: $\frac{2.f_s}{3} > \frac{f_s}{2}$. The coefficients $\{c_k\}_0^{\Lambda-1=2}$ are present in 3 frequency components. One is relative to frequency zero (c_0), and the other two (c_1 and c_2) have the same role for reconstructing a sinusoid with $f = f_s/3$.

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 $f_k = k\frac{f_s}{\Lambda},\ k \in \left\{0, \ldots, \left\lfloor\frac{\Lambda}{2}\right\rfloor\right\}$. When $k > \frac{\Lambda}{2}$, the frequency f_k is mirrored through $\frac{f_s}{2}$ in this way: $f_k = \frac{f_s}{2} - (f_k - \frac{f_s}{2}) = f_s - f_k = f_s - k\frac{f_s}{\Lambda} = (\Lambda - k)\frac{f_s}{\Lambda} \quad \Rightarrow \quad f_k \equiv f_{\Lambda-k}\ ,\ \ \forall\ k < \Lambda$.

The same applies to $w_k = f_k.\frac{2\pi}{f_s}$ 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 equivalent in pairs, except for f_0, and, when Λ is even, for $f_{\Lambda/2} = f_{\max} = \frac{f_s}{2}$. 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). In fact when $k = 0$ or $k = \Lambda/2$ the mirror of the frequencies are themselves: $f_{\Lambda/2} = f_{(\Lambda-\Lambda/2)=\Lambda/2}$ and $f_0 = f_{(\Lambda-0)=\Lambda} = f_0$. Furthermore, these two frequencies (zero and Nyquist frequency) do not have phase offset: their coefficients are 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 in Equations 17, 18 and 19:

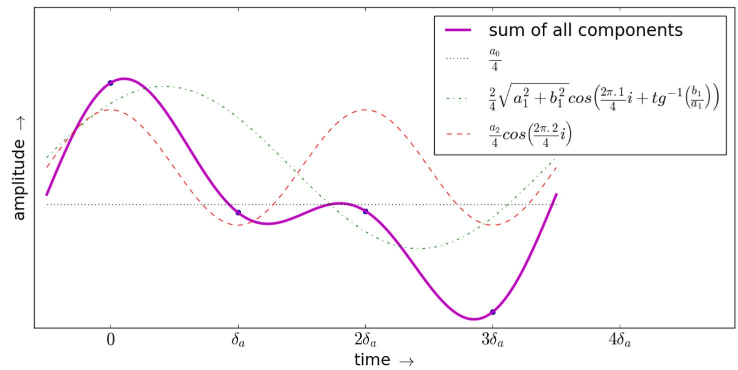

Fig. 7. Frequency components for 4 samples.

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

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

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

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

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

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

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

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 their two (possible) components. The individual components 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 component with frequency $f_0 = 0$.

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.

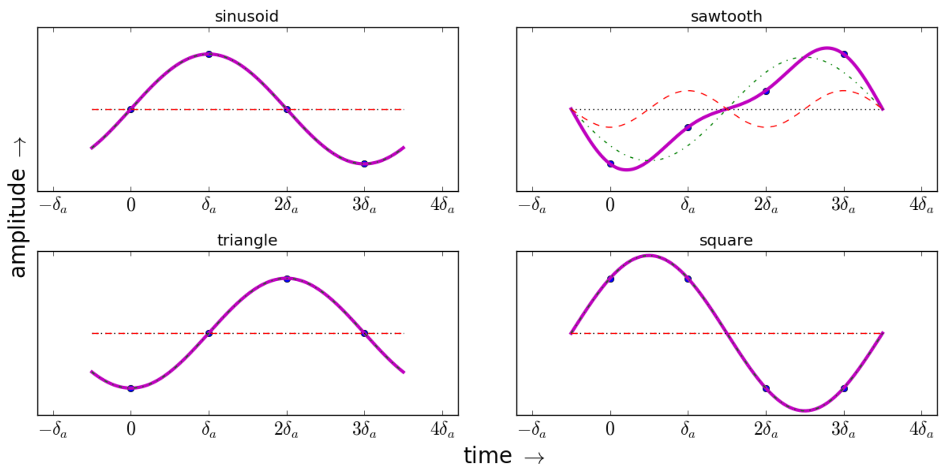

Fig. 8. Basic wave forms with 4 samples.

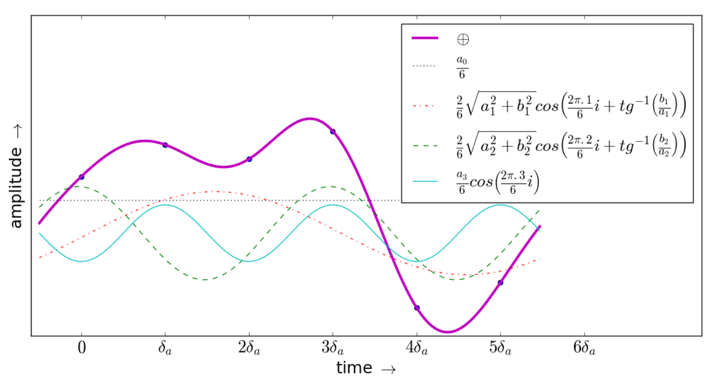

Fig. 9. Frequency components for 6 samples: 4 sinusoids, one of them is the *bias* with zero frequency.

2.6 The basic note

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

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

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 one of the basic waveforms can be used to specify the note, where $\lambda_f = \frac{f_s}{f}$ is the number of samples at the period. Let $L^{f,\,\delta_f}$ be the sequence that describes a period of the waveform $L^f \in \{S^f, Q^f, T^f, D^f, R^f\}$ with duration $\delta_f = 1/f$ (as given by Equations 9, 10, 11 and 12 and let R_i^f be a sampled real waveform):

[8]As mentioned before, this limitation simplifies the explanation for now and will be overcome in the next section.

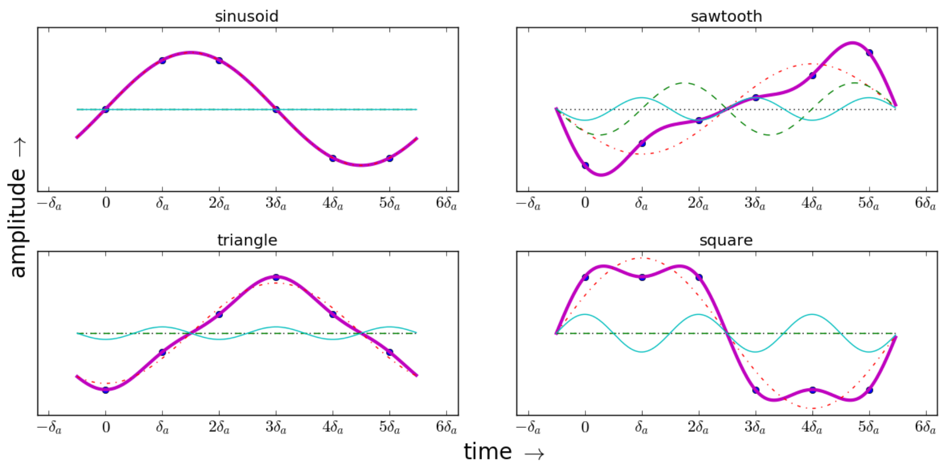

Fig. 10. Basic waveforms with 6 samples: triangular and square waveforms have odd harmonics, with different proportions and phases; the sawtooth has even harmonics.

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

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

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

2.7 Spatialization: localization and reverberation

A sound is always produced within the three dimensional physical space. The consideration of this fact is the subject of spatialization[9]. A note is always spatialized even though it is not one of its four basic properties in canonical musical theory (duration, loudness, pitch and timbre). A note has a source which has a position at the ordinary three dimensional physical space. This position is the spatial localization of the sound. It is often (modeled as) a single point but can be a surface or a volume. The reverberation in the environment in which a sound occurs is another one of the main topics of spatialization. Both concepts, spatial localization and reverberation, are widely valued by composers, audiophiles and the music industry [45].

2.7.1 Spatial localization. It is understood that the perception of sound localization occurs in our nervous system mainly by three cues: the delay of incoming sound between both ears, the difference of sound intensity at each ear and the filtering performed by the human body, specially its chest, head and ears [8, 19, 54].

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:

[9]By spatialization one might find both: 1) the consideration of cues in sound that derive from the environment, including the localization of the listener and the sound source; 2) techniques to produce sound through various sources, such as loudspeakers, singers and traditional musical instruments, for musical purposes. We focus in the first item although issues of the item are also tackled.

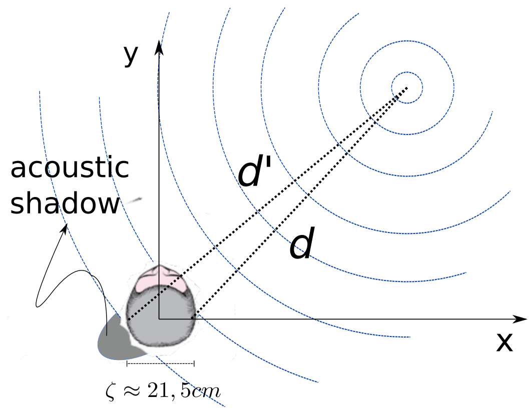

Fig. 11. Detection of sound source localization: schema used to calculate Interaural Time Difference (ITD) and Interaural Intensity Difference (IID).

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

where ζ is the distance between the ears, known to be $\zeta \approx 21.5cm$ in an adult human. Straightforward calculations result in the Interaural Time Difference:

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

and in the Interaural Intensity Difference:

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

$IID_a = \frac{d}{d'}$ 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 temporal advance for the right channel which is a crucial cue to localization perception of bass sounds and percussive sonorities [8]. With $\Lambda_{ITD} = \lfloor ITD.f_s \rfloor$:

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

where t_i is a right channel sample and t_i' is a left channel sample. If $\Lambda_{ITD} < 0$, it is necessary 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 (azimuthal) angle θ given by:

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

with x, y as presented in Figure 11. Furthermore, there are problems when θ falls within the so-called "cone of confusion": the same pair of ITD and IID are related to a large number of points inside the cone [10]. On those points the inference of the azimuthal angle depends especially on the filtering of the frequencies by the head and torso [8, 19]. Figure 11 depicts the acoustic shadow of the cranium, an important phenomenon to perception of source azimuthal angle in the cone of confusion. 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 $\approx 0.7ms$ [45]. The complete localization, including height and distance of sound source, is given by the Head Related Transfer Function (HRTF) [19]. 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) [12]. Each human body has its filtering and there are techniques to generate HRTFs to be universally used [17].

2.7.2 Reverberation. The reverberation results from sound reflections and absorption by the environment (e.g. a room) surface where a sound occurs. 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 wave encounters a surface there are: 1) inversion of the propagation speed component normal to the surface; 2) energy absorption, especially in high frequencies. The sonic waves propagate until they reach inaudible levels (and further but then can often be neglected). As a sonic front reaches the human ear, it can be described as an ordinary 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, one can use asymmetric reflections with relation to the axis perpendicular to the surface, or increase the intensity of the last reflections.

There are reverberation models less based in each independent reflection and that explores valuable cues to the auditory system. In fact, a reverberation can be modeled with a set of two temporal and two spectral regions [65]:

- 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 a progressive decay and smooth statistical fluctuations.

Smith III states that usual concert rooms have total reverberation time of ≈ 1.9 seconds, and that the period of first reflections is around $0.1s$. Under these conditions, there are perceived wave pulses which propagate for $652.08m$ ($83.79k$ samples in $f_s = 44.1kHz$) before reaching the ear. In addition, reflections that reach the ear after propagating for $34.32m$ ($4.41k$ samples in $f_s = 44.1kHz$) or more have incidences less distinct by hearing. The 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

[10]The cone itself is not shown in Figure 11 because it is not exactly a cone and its precise dimensions were not found in the literature. Given the filtering and diffraction dependent on the sound spectrum, it is hard, if not impossible, to correctly draw the cone. Even so, the cone of confusion can be pictured as a cone with its apex placed in the middle of the head and growing out in the direction of the ear [19].

at least $3-30m$, depending on the room dimensions, the separation between each reflection is $8-90ms$ ($\approx 350-4000$ samples in $f_s = 44.1kHz$). Also, it is experimentally verifiable that the number of reflections increases with the square of time. A discussion about the use of convolutions and filtering to favor implementation of these phenomena is provided in Section 3.6, particularly in the paragraphs about reverberation.

2.8 Musical usages

The sum of the respective *ith* samples of N sequences with the same size Λ results in the overlapped spectral contents of each sequence, in a process called mixing:

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

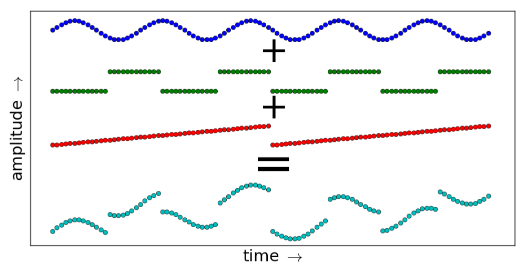

Fig. 12. Mixing of three sound sequences. The amplitudes are directly summed sample-by-sample.

Figure 12 illustrates this overlapping of discretized sound waves, each with 100 samples. The duration is very short $\frac{f_s=44.1kHz}{100} \approx 2ms$. If $f_s = 44.1kHz$, the frequencies of the sawtooth, square and sine wave are, respectively: $\frac{f_s}{100/2} = 882Hz$, $\frac{f_s}{100/4} = 1764Hz$ and $\frac{f_s}{100/5} = 2205Hz$. One can complete the sequence with zeroes to sum (mix) sequences of different sizes. The mixed notes are generally separated by the ear according to the physical laws of resonance and by the nervous system [54]. This process of mixing musical notes results in musical harmony and counterpoint which guide subjective aspects of music appreciation [59] and are addressed in Section 4.

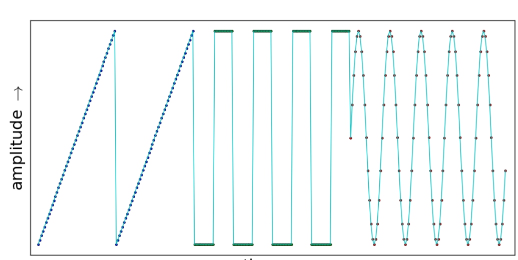

Fig. 13. Concatenation of three sound sequences by temporal overlap of their samples.

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 results in a simple melodic sequence:

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

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 visually observe the concatenation of sound sequences. Moreover, each note has a duration larger than $100ms$ if $f_s < 1kHz$.

The musical piece *reduced-fi* explores the temporal juxtaposition of notes, resulting in a homophonic piece. The vertical principle is demonstrated in the *sonic portraits* which are static sounds with peculiar spectrum.[9]

3 VARIATION IN THE BASIC NOTE

The basic digital music note defined in Section 2 has the following characteristics: duration, pitch, loudness and timbre. We also stated that a note is always spatialized and considered both localization and reverberation. This is a useful and paradigmatic model, but it does not exhaust the aspects of a musical note. First of all, a real note is not static [20]. For example, a $3s$ piano note has an abrupt rise of intensity at the beginning and a progressive decline, has harmonic components decaying and emerging. These variations are not mandatory, but they are used in sound synthesis for music because they reflect how sounds appear in nature and because they yield interesting sounds [54]. To explore all the ways in which variations occur within a note is out of the scope of any work given the complexity and sensibility of the human cognition of sounds. Resources essential to produce variations in the basic note are presented in this section. The musical pieces *Transit para meter*, *Vibrates and shakes*, *Tremolos, vibratos and the frequency*, *Little train of impulsive hillbillies*, *Noisy band*, *Bela rugosi*, *Children's chorus*, *ADa and SaRa* were made to validate and illustrate concepts of this section. [9]

3.1 Lookup table

The *Lookup Table* (LUT) is an array for indexed operations which substitutes continuous and repetitive calculation. It is used to reduce computational complexity and for employing functions without calculating them directly, e.g. from sampled data or hand picked values. In music its usage transcends these applications: it simplifies many operations and enables the 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 $\widetilde{L} = \left\{\widetilde{l_i}\right\}_0^{\widetilde{\Lambda}-1}$ the sequence of samples $\widetilde{l_i}$. A sequence $T^{f,\Delta}$ with samples of a sound with frequency f and duration Δ can be obtained by means of $\widetilde{L}$:

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

In other words, with the right LUT indexes $(\gamma_i \% \widetilde{\Lambda})$ it is possible to synthesize sounds at an arbitrary 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 [33].

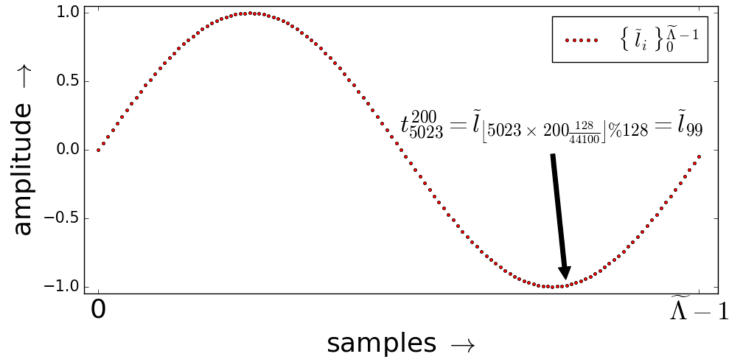

Fig. 14. Search in the *lookup table* to synthesize sounds at different frequencies using a unique waveform with high resolution. Each *i-th* sample t_i^f of a sound with frequency f is related to the samples in the table $\widetilde{l}$ by the relation $t_i^f = \widetilde{l}_{\left\lfloor i.f\frac{\widetilde{\Lambda}}{f_s}\right\rfloor \% \widetilde{\Lambda}}$ where $\widetilde{\Lambda}$ is the number of sample in $\widetilde{L}$ and f_s is the sampling rate.

The calculation of γ_i can be understood as follows: $\frac{i}{f_s}$ is incremented by 1 at each second. Multiplied by the period, $i\frac{\widetilde{\Lambda}}{f_s}$ covers $\widetilde{\Lambda}$ in one second. Finally, with frequency f it results in $if\frac{\widetilde{\Lambda}}{f_s}$ which completes f cycles on $\widetilde{L}$ in 1 second, i.e. the resulting sequence has 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_s. The lookup procedure 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 musical note [14, 23].

3.2 Incremental variations of frequency and intensity

As stated by the Weber and Fechner [25] 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 (Section 3.5), linear variation is discussed first.

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

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

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

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

where $\Delta_{\gamma_i} = f_i \frac{\widetilde{\Lambda}}{f_s}$ is the LUT index increment between two samples given the frequency of the first sample. Therefore, it is handy to calculate the amplitudes $t_i^{\overline{f_0, f_{\Lambda-1}}}$ 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:

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

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

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

The term $\frac{i}{\Lambda-1}$ covers the interval $[0, 1]$ and it is possible to 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 loudness[11]. It is sufficient to multiply the original sequence element-wise by the sequence $a_{\Lambda-1}^{\left(\frac{i}{\Lambda-1}\right)^\alpha}$, where α is an arbitrary coefficient and $t_{\Lambda-1}.a_{\Lambda-1}$ is the value to be reached at the end of the transition.

Thus, for amplitude variations:

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

where $a_0 = 1$. And, therefore:

$$\begin{aligned} T' = T \odot A &= \{t_i a_i\}_0^{\Lambda-1} \\ &= \left\{ t_i (a_{\Lambda-1})^{\left(\frac{i}{\Lambda-1}\right)^\alpha} \right\}_0^{\Lambda-1} \end{aligned} \tag{39}$$

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

[11]The loudness (a psychophysic quality) is a consequence of different sound characteristics, like reverberation and concentration of high harmonics, among which is the wave total energy. Wave energy can be varied in many ways. The simplest way consists of modifying the amplitude by multiplying the whole sequence by a real number (see Section 2.2). The increase of energy without changing amplitude ambit is the *sound compression*, quite popular nowadays for music production [37].

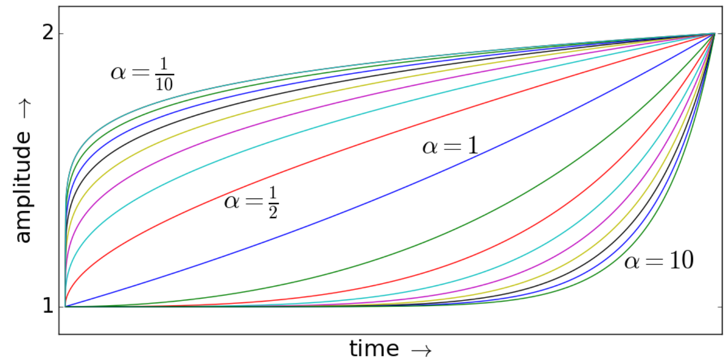

Fig. 15. Intensity transitions for different values of α (see Equations 38 and 39).

that implies the linear variation of loudness. Figure 15 depicts transitions for different values of α, a gain of $\approx 6dB$ as exposed 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 a ratio of any number in relation to zero is not well defined for our purposes. It is possible to solve this dilemma by choosing a number that is small enough such as $-80dB \Rightarrow a = 10^{\frac{-80}{20}} = 10^{-4}$ as the minimum amplitude for a *fade in* ($a_0 = 10^{-4}$) or for a *fade out* ($a_{\Lambda-1} = 10^{-4}$). A linear fade can be used then to reach zero amplitude, if needed. 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 departs from zero [23].

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. Given Equations 7 and 39, a transition of V_{dB} decibels is:

$$T' = \left\{ t_i 10^{\frac{V_{dB}}{20}\left(\frac{i}{\Lambda-1}\right)^{\alpha}} \right\}_0^{\Lambda-1} \tag{41}$$

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 loudness transitions commonly called *slow fade*, while $\alpha < 1$ results in *fast fade* [37].

The linear transitions will be used for AM and FM synthesis, while logarithmic transitions are proper tremolos and vibratos, as developed in Section 3.5. A non-oscillatory exploration of these variations is in the music piece *Transit para meter* [9].

3.3 Application of digital filters

This subsection is limited to a description of processing by convolution and difference equations, and immediate applications, as a thorough discussion of filtering is beyond the scope of this study[12]. Filtering can be part of the sound synthesis or made subsequently as part of processes commonly referred to as "acoustic/sound treatment".

[12] The design and implementation of filters is field of recognized complexity, with dedicated literature and software [48, 64].

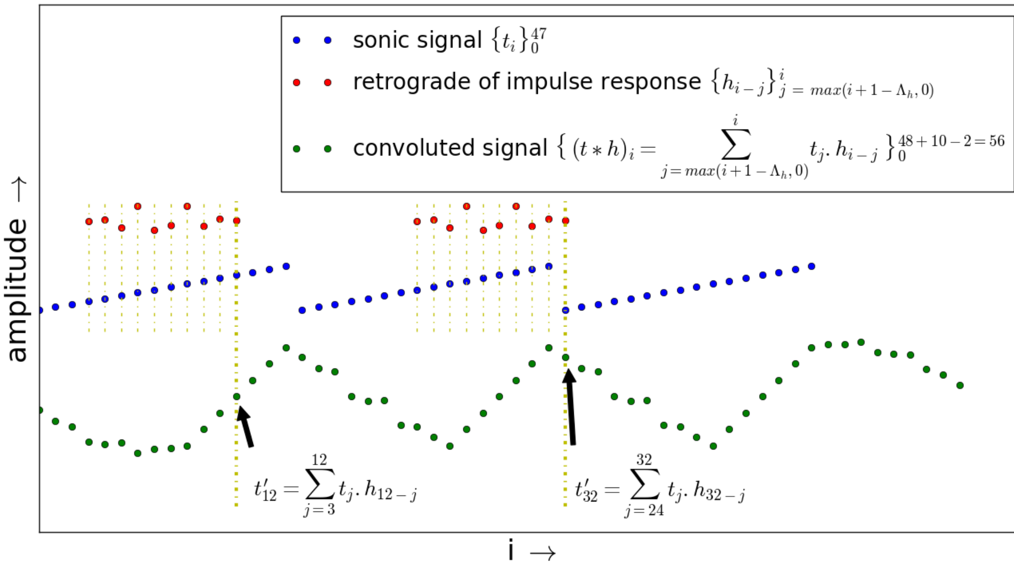

Fig. 16. Graphical interpretation of convolution. Each resulting sample is the sum of the previous samples of a signal, with each one multiplied by the retrograde of the other sequence.

3.3.1 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[13]. For the purposes of this work, convolution $T * H$ of T with H is defined as:

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

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 samples from the convolution. The final signal always has the length

[13]It is possible to apply the filter in the frequency domain multiplying the Fourier coefficients of both the sound and the impulse response, and then performing the inverse Fourier transform of the resulting spectrum. This is usually computationally cheaper than performing the convolution in the time domain but might imply greater complexity especially if the sequence of samples is not fully available (e.g. in real-time applications) [48].

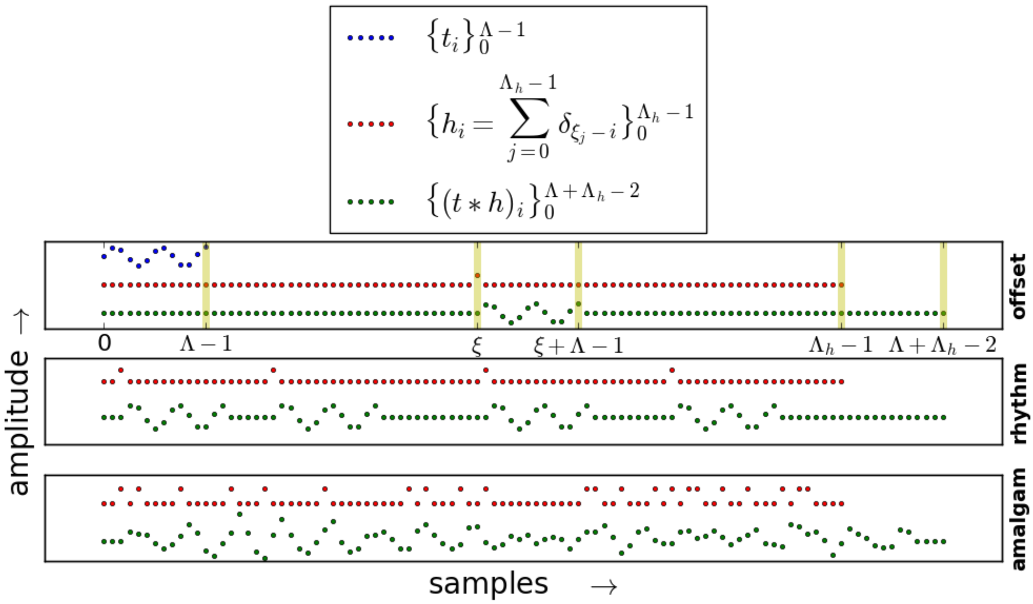

Fig. 17. Convolution with the impulse: shifting (a), delay lines (b) and granular synthesis (c). Represented in increasing order of its pulse density. The vertical axis is related to amplitude although one should keep in mind that each subplot has two or three displaced signals.

$\Lambda_{t'} = \Lambda_t + \Lambda_h - 1$. With this procedure it is possible to achieve reverberators, equalizers, *delays*, to name a few of a variety of filters for sound processing used 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 recording the sound in a room when someone triggers a click which resembles an impulse, or be obtained by a sinusoidal sweep whose Fourier transform approximates its frequency response. Both are impulse responses which, when convoluted with a sequence of sound samples, result in the own sound with a reverberation that resembles the original environment where the measure was made [23]. The inverse Fourier transform of a sequence of coefficients might be used as an impulse response of a FIR filter. Convoluted with a sound, it performs the frequency filtering specified by coefficients[14]. The greater the number of samples, the higher the spectral 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 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 density 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 a sound amalgam (more than $\approx$ 20 impulses per second). In the latter case, the process resembles granular synthesis, reverberation and equalization.

[14]Given that the sound is described by a sequence of real numbers, the coefficients should follow the rule $c_k = c^*_{\Lambda-k}$ as discussed in Section 2.5.

3.3.2 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 $\left\{t'_i\right\}$ are given by differences between original samples $\{t_i\}$ and previous ones in the resulting signal $\left\{t'_{i-k}\right\}$.

There are many methods and tools to obtain IIR filters. The text below lists a selection for didactic and reference purposes. They are well behaved filters whose frequency response are illustrated in Figure 18. For the low-pass and high-pass filters, 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 $-3dB \approx 0.707$ 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.

It is possible to apply these filters successively in order to obtain filters with other frequency responses. Other possibilities include the use of a biquad 'filter recipe'[15] or a Chebichev filter[16]. Both alternatives are explored by [64, 66], and by the collection of filters maintained by the *Music-DSP* community of the Columbia University [21, 48].

(1) Low-pass with a simple pole and module of the frequency response in the upper left corner of Figure 18. The general equation has the cutoff frequency $f_c \in (0, \frac{1}{2})$, which is the 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:

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

(2) High-pass filter with a simple pole and module of the frequency responses at the upper right corner of Figure 18. The general equation with cutoff frequency $f_c \in (0, \frac{1}{2})$ has coefficients:

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

(3) Notch filter. This filter is parametrized by a center frequency f_c and bandwidth bw, both given as a fraction of f_s, therefore $f,\ bw \in (0, 0.5)$. Both frequencies $f_c \pm bw$ have ≈ 0.707

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

[16]Butterworth and Elliptical filters can be considered as special cases of Chebichev filters [48, 64].

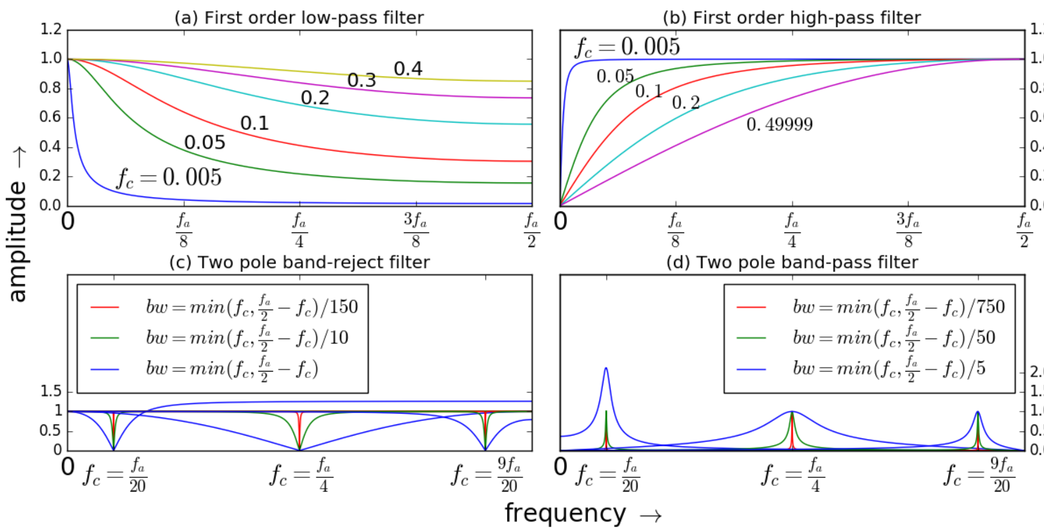

Fig. 18. Modules for the frequency response (a), (b), (c) and (d) for IIR filters of Equations 44, 45, 47 and 48 respectively, considering different cutoff frequencies, center frequencies and bandwidths.

of the amplitude, i.e. an 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:

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

3.4 Noise

Sounds without an easily recognizable pitch are generally called noise [40]. They are important musical sounds, as noise is present in real notes, e.g. emitted by a violin or a piano. Furthermore, many percussion instruments does not exhibit an unequivocal pitch and their sounds are generally

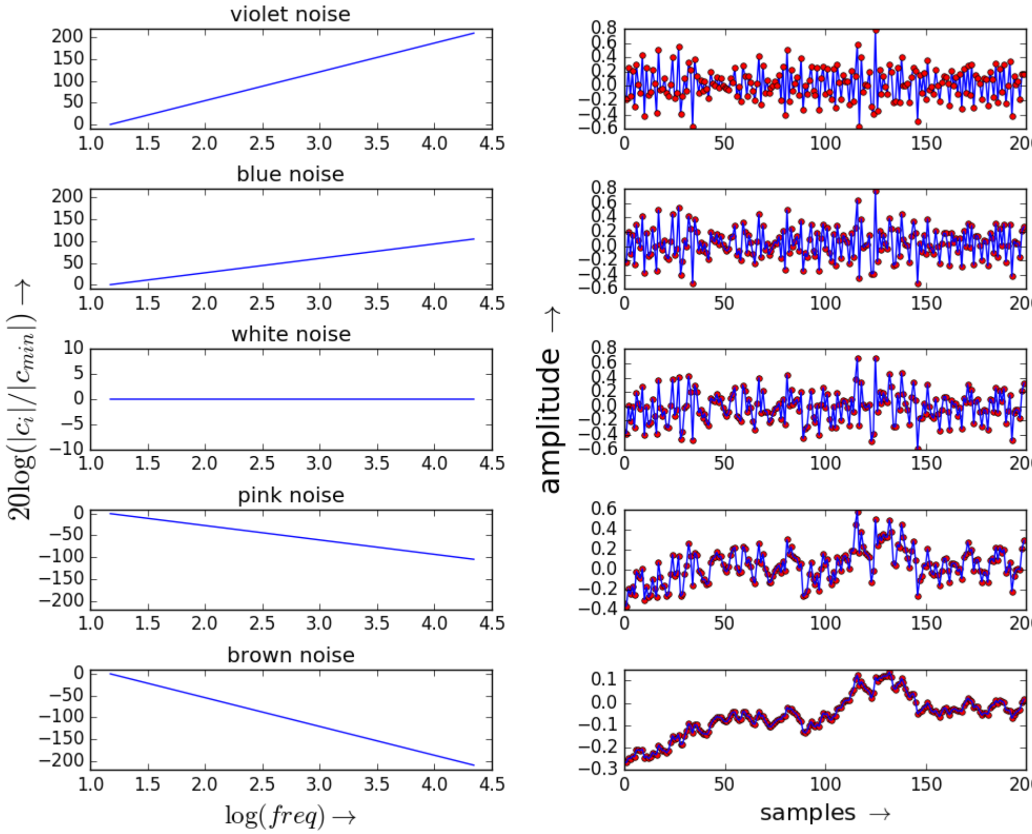

Fig. 19. Colors of noise generated by Equations 49, 50, 51, 52 and 53: spectrum and example waveforms.

regarded as noise [54]. In electronic music, including electro-acoustic and dance genres, noise has diverse uses and frequently characterizes the music style [23].

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 [23]. 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 noises* since they are associated with colors for many reasons. Figure 19 shows the spectrum profile and the corresponding sonic samples sequence side-by-side. All five noises were generated with the same phase for each component, making it straightforward 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, such as the white color. It is possible to obtain white noise with the inverse transform of the following coefficients:

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

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 Section 2.5).

- The pink noise is characterized by a decrease of $3dB$ per octave. This noise is useful for testing electronic devices and are prominent in nature [54].

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

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[17]. What characterizes a brown noise is the decrease of $6dB$ per octave, with α_i in Equations 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, using Equations 50 again:

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

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

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

[17] 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 vivid colors [23, 37].

- The black noise has higher losses than $6dB$ per 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 file `src/aux/iso226.py` [9].

This subsection discussed only noise with static spectrum. There are also characterizations for noise with dynamic spectrum, 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 [23].

3.5 Tremolo and vibrato, AM and FM

Vibrato is a periodic variation in pitch (frequency) and tremolo is a periodic variation in loudness (intensity)[18]. A vibrato with frequency f' can be achieved by:

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

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

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

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

$$T^{f,vbr(f',\nu)} = \left\{ t_i^{f,vbr(f',\nu)} \right\}_0^{\Lambda-1} = \left\{ \widetilde{l}_{\gamma_i \% \widetilde{\Lambda}} \right\}_0^{\Lambda-1} \tag{59}$$

It is important correctly distinguish both tables ($\widetilde{M} = \{\widetilde{m}\}_0^{\widetilde{\Lambda}_M - 1}$ and $\widetilde{L} = \{\widetilde{l}\}_0^{\widetilde{\Lambda}-1}$) and sequences ($T' = \{t'_i\}_0^{\Lambda-1}$ and $T = \{t_i\}_0^{\Lambda-1}$). Table $\widetilde{M}$ and the indices γ'_i yield the sequence $\{t'_i\}$ which is the oscillation pattern in the frequency while table $\widetilde{L}$ and the indices γ_i yield $\{t_i\}$ which is the sound itself. Variables are useful to quantify the vibrato amplitude (or depth):

- let μ be frequency difference (in Hz) between the main frequency and the maximum frequency, and
- let ν be the the number of semitones (or half steps) added and subtracted by the oscillation (2ν is the number of semitones between the upper and lower peaks of frequency in the sound).

[18]The jargon may be different in other contexts as stated in Section 1.2 [40, 59].

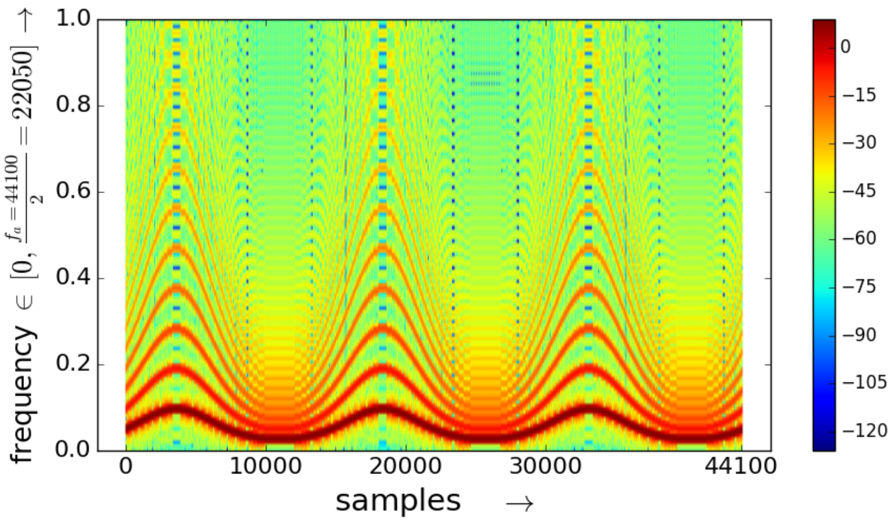

Fig. 20. Spectrogram of a sound with a sinusoidal vibrato of $3Hz$ and one octave of depth in a $1000Hz$ sawtooth wave, with $f_s = 44.1kHz$. The color bar is in decibels.

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. The semitone/pitch variation is the invariant quantity.

Figure 20 is the spectrogram of an artificial vibrato for a note with $1000Hz$ (between a B and a C), in which the 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[19]. Oscillations with precise waveforms and arbitrary amplitudes are not possible in traditional music instruments, and thus it introduces novelty in the artistic possibilities.

The 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:

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

where V_{dB} is the oscillation depth in decibels of the 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 is invariant.

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 $1.5Hz$ if $f_s = 44.1kHz$ because duration $= \frac{i_{\max}=82000}{f_s} = 2s \Rightarrow \frac{3 oscillations}{2s} = 1.5$ oscillations per second.

[19]The pitch deviation is called 'vibrato depth' and is generally given in semitones or cents (one cent $= \frac{1}{100}$ of a semitone).

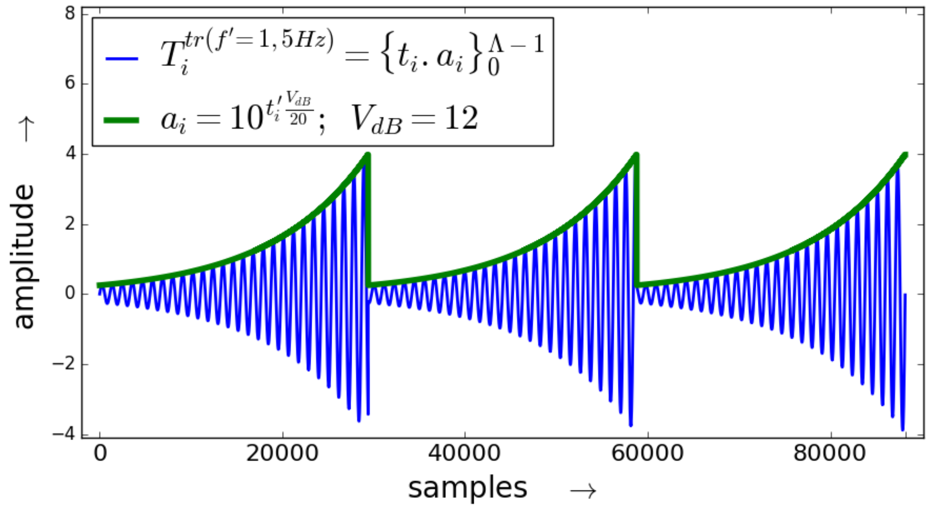

Fig. 21. Tremolo with a depth of $V_{dB} = 12dB$, with a sawtooth waveform as its oscillatory pattern, with $f' = 1.5Hz$ in a sine of $f = 40Hz$ (if sample frequency is $f_s = 44.1kHz$).

The music piece *Vibrates and shakes* explores these possibilities given by tremolos and vibratos used simultaneously and independently, with different frequencies f', depths (ν and V_{dB}), waveforms, and incremental modifications of parameters (tremolos and vibratos occur often together in traditional music instruments and voices). Aiming at a qualitative appreciation, the piece also develops a comparison between vibratos and tremolos in logarithmic and linear scales. [9]

The proximity of f' to $20Hz$ generates roughness in both tremolos and vibratos. This roughness is largely appreciated both in traditional classical music and current electronic music, especially in the *Dubstep* genre. Roughness is also generated by spectral content that produces beats[20]. The sequence *Bela Rugosi* explores this roughness threshold with concomitant tremolos and vibratos at the same voice, with different intensities and waveforms. [9]

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 audible spectrum of original sound T 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 convenient to understand FM when only pure sines are used and, when other waveforms are employed, to decompose the signals into their respective Fourier components (i.e. sines as well). The FM synthesis performed with a sinusoidal vibrato of frequency f' and depth μ in a sinusoidal sound T with frequency f generates bands centered around f and far from each other by f':

[20] A beat is an interference pattern between two sounds which is perceived as a periodic variation of loudness. If the sounds are sinusoids, the variation is also sinusoidal with a frequency that is the difference between the frequencies of the sinusoids [49, 50].

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

where

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

is the Bessel function [5, 66] 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 result of using Equations 57 for FM is described in the Appendix D of [31], where the spectral content of the FM synthesis is calculated for oscillations in the logarithmic scale. In fact, the simple and attractive FM behavior is usually observed within linear oscillations, such as in Equation 62, which yield less strident and less noisy sounds.

For the amplitude modulation (AM):

$$\begin{split}\{t'_i\}_0^{\Lambda-1} = \{(1+a_i).t_i\}_0^{\Lambda-1} = \left\{\left[1 + M.\sin\left(f'.2\pi\frac{i}{f_s-1}\right)\right].P.\sin\left(f.2\pi\frac{i}{f_s-1}\right)\right\}_0^{\Lambda-1} = \\ = \left\{P.\sin\left(f.2\pi\frac{i}{f_s-1}\right) + \frac{P.M}{2}\left[\sin\left((f-f').2\pi\frac{i}{f_s-1}\right) + \sin\left((f+f').2\pi\frac{i}{f_s-1}\right)\right]\right\}_0^{\Lambda-1}\end{split} \tag{64}$$

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 description of the spectrum of an AM performed with oscillations in the logarithmic amplitude scale is also available in the Appendix D of [31]. In FM and AM jargon, the sequence T is called 'carrier' and is modulated by T', called 'modulator'; μ and $\alpha = 10^{\frac{V_{dB}}{20}}$ are the 'modulation indexes' and f' is the modulation frequency. The following equations are useful for the achievement of the oscillatory pattern of the modulator sequence $\{t'_i\}$:

$$\gamma'_i = \left\lfloor i f'\frac{\widetilde{\Lambda}_M}{f_s}\right\rfloor \tag{65}$$

$$t'_i = \widetilde{m}_{\gamma'_i\,\%\widetilde{\Lambda}_M} \tag{66}$$

The modulator $\{t'_i\}$ is applied to the carrier $\{t_i\}$ in a FM by:

$$f_i = f + \mu.t'_i \tag{67}$$

$$\Delta_{\gamma_i} = \frac{\widetilde{\Lambda}}{f_s}f_i \quad \Rightarrow \quad \gamma_i = \left\lfloor\sum_{j=0}^{i}\frac{\widetilde{\Lambda}}{f_s}f_j\right\rfloor = \left\lfloor\sum_{j=0}^{i}\frac{\widetilde{\Lambda}}{f_s}(f+\mu.t'_j)\right\rfloor \tag{68}$$

$$T^{f,\, FM(f',\, \mu)} = \left\{ t_i^{f,\, FM(f',\, \mu)} \right\}_0^{\Lambda-1} = \left\{ \widetilde{l}_{\gamma_i \% \widetilde{\Lambda}} \right\}_0^{\Lambda-1} \tag{69}$$

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

To perform AM, the signal $\{t_i\}$ needs to be modulated by $\{t'_i\}$ using the following equations:

$$a_i = 1 + \alpha.t'_i \tag{70}$$

$$T_i^{f,\, AM(f',\, \alpha)} = \left\{ t_i^{f,\, AM(f',\, \alpha)} \right\}_0^{\Lambda-1} = \{t_i.a_i\}_0^{\Lambda-1} = \{t_i.(1 + \alpha.t'_i)\}_0^{\Lambda-1} \tag{71}$$

3.6 Musical usages

At this point the musical possibilities are very wide. The following musical usages comprehend a collection of possibilities with the purpose of exemplifying types of sound manipulations that result in musical material. Some of them are discussed in depth in the next section.

3.6.1 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 set:

$$\begin{aligned} f^{vbr} = f^{tr} &= func_a(f) \\ \nu &= func_b(f) \\ V_{dB} &= func_c(f) \end{aligned} \tag{72}$$

with both f^{vbr} and f^{tr} as f' in the equations of Section 3.5. 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 and the frequency* explores such bonds and exhibits variations in the oscillation waveform with the purpose of building a *musical language* (details in the next section). [9]

3.6.2 Convolution for rhythm and meter. With the purpose of establishing metric and rhythms, the convolution with an impulse shifts the sound to the position of the impulse, as stated in Section 3.3.1. A musical pulse - such as specified by a BPM tempo - can be implied by an impulse at the start of each beat. For example, two impulses equally spaced build a binary division in 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 which eases both binary or ternary divisions of the pulse. This is found in many ethnic and traditional musical styles [34]. Absolute values of the impulses entail proportions among the amplitudes of the sonic reincidences. The use of convolution with impulses in this context is explored in the music piece *Little train of impulsive hillbillies*. The piece also encompass the creation of 'sound amalgams' e.g. by granular synthesis (see Figure 24). This piece is a link to the contents of next section because of the explicit organization of notes in higher level structures. [9].

3.6.3 Moving source and receptor, Doppler effect. Given the audio source speed s_s, with positive values if the source moves away from receptor, and receptor speed s_r, positive when it gets closer to audio source, the frequency is given by the well-known shift in the Doppler effect:

$$f = \left(\frac{s_{sound} + s_r}{s_{sound} + s_s}\right) f_0 \tag{73}$$

With both frequencies f and f_0, and the IID and ITD from the new source position (see Section 2.7), it is possible to simulate 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 the receptor. Both amplitude and frequency of a moving audio source can be modeled. If the audio source is in front of the receptor with y_0 meters of horizontal distance and z_0 meters of height, the distance is given by $D = \left\{d_i = \sqrt{y_i^2 + z_0^2}\right\}_0^{\Lambda-1}$, 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 Section 2.2 for potency to amplitude conversion). Thus:

$$A = \left\{\frac{z_0}{d_i} A_{\Delta P}\right\}_0^{\Lambda-1} = \left\{\frac{z_0}{\sqrt{y_i^2 + z_0^2}} \sqrt{\frac{v_r - v_s}{343,2} + 1}\right\}_0^{\Lambda-1} \tag{74}$$

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 smooth progression. In the given case, where there is a static height z_0, the frequencies F_i at the observer are given by:

$$F = \{f_i\}_0^{\Lambda-1} = \left\{\frac{v_{sound} + v_r \frac{y_i}{\sqrt{z_0^2+y_i^2}}}{v_{sound} + v_s \frac{y_i}{\sqrt{z_0^2+y_i^2}}} f_0\right\}_0^{\Lambda-1} \tag{75}$$

There is an implementation of the Doppler effect in `src/sections/3.py` [9]. We considered this a musical usage given that it is not a basic method to introduce variation into a note and because it is a valued effect in audio plugins and electro-acoustic music [45].

3.6.4 Filters and noises (Sections 3.4 and 3.3). With the use of filters, the possibilities are even wider. Convolve a signal to have a reverberated version of it, to remove noise, to distort or to handle 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 $300Hz$ and $2kHz$. By rejecting the frequency of an electric oscillation (usually $50Hz$ or $60Hz$) and the harmonics, one can remove noises caused by audio devices connected to the power supply (called hiss noise). Another 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 [54]. Chains of these filters can perform complex and more accurate filtering routines. The music piece *Noisy band* explores filters and many kinds and noise synthesis. [9]

A sound can be altered through different filtering and then mixed to create an effect known as *chorus*. Inspired by what happens in a choir of singers, the same note is synthesized many times using small and potentially arbitrary modifications of parameters like center frequency, presence (or absence) of vibrato or tremolo and its characteristics, equalization, loudness, duration, etc. The versions of the sound are then mixed together (see Equation 29). The music piece *Child chorus* implements a chorus in many ways with different sounds. [9]

3.6.5 Reverberation. Using the same terms of Section 2.7, 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 accounts for both temporal parts of the reverberation: the early reflections and the late reverberation. Quality can be improved by varying the geometric trajectory and filtering by each surface where wavefront reflected before reaching the ear in the first $100-200ms$ (mainly with a LP). The colored noise can be gradually introduced with a *fade-in*: the initial moment given by the 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 a decaying noise. As an example, consider Δ_1 as the duration of the first section and Δ_R as the duration of total reverberation (let also $\Lambda_1 = \Delta_1 f_s$, $\Lambda_R = \Delta_R f_s$). Let p_i be the probability of a sound being repeated in the i-th sample. The amplitude decreases exponentially. Following Section 2.7, the samples R^1 of the impulse response of the first period can be described as:

$$R^1 = \left\{r_i^1\right\}_0^{\Lambda_1-1} \quad , \text{ where } \quad r_i^1 = \begin{cases} 10^{\frac{V_{dB}}{20}\frac{i}{\Lambda_R-1}} & \text{with probability} \quad p_i = \left(\frac{i}{\Lambda_1}\right)^2 \\ 0 & \text{with probability} \quad 1-p_i \end{cases} \tag{76}$$

where V_{dB} is the total decay in decibels, typically $-80dB$ or $-120dB$. The samples R^2 of the impulse response of the second period can be taken from a brown noise (or by a pink noise) with an exponential amplitude decay of the waveform:

$$R^2 = \left\{r_i^2\right\}_{\Lambda_1}^{\Lambda_R-1} = \left\{10^{\frac{V_{dB}}{20}\frac{i}{\Lambda_R-1}} . r_i^m\right\}_{\Lambda_1}^{\Lambda_R-1} \tag{77}$$

The reverberation impulse response is then:

$$R = \{r_i\}_0^{\Lambda_R-1} \quad , \text{ where } r_i = \begin{cases} 1 & \text{if} \quad i = 0 \\ r_i^1 & \text{if} \quad 1 \leq i < \Lambda_1 - 1 \\ r_i^2 & \text{se} \quad \Lambda_1 \leq i < \Lambda_R - 1 \end{cases} \tag{78}$$

A sound with simulated reverberation can be achieved by a simple convolution of R with the sound sequence T, as described in Section 3.3. Reverberation is well known for causing great interest in listeners and to provide sonorities that are more enjoyable than the sound without any reverberation. Furthermore, modifications in the reverberation space consist in a common resource (almost a *cliché*) to surprise and attract the listener. An implementation of the reverb recipe described here is in the file `src/sections/3.py`. [9].

3.6.6 ADSR envelopes. The variation of loudness along the duration of sound is crucial to our timbre perception. The intensity 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 [52]. The canonical 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 intensity (in which the intensity remains stable after the decay) and release duration (after the 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 sum of the attack, decay and sustain durations.

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 expressed as:

$$\begin{aligned}
\{a_i\}_0^{\Lambda_A-1} &= \left\{ \xi \left(\frac{1}{\xi}\right)^{\frac{i}{\Lambda_A-1}} \right\}_0^{\Lambda_A-1} \qquad \text{or} \\
&= \left\{ \frac{i}{\Lambda_A - 1} \right\}_0^{\Lambda_A} \\
\{a_i\}_{\Lambda_A}^{\Lambda_A+\Lambda_D-1} &= \left\{ a_S^{\frac{i-\Lambda_A}{\Lambda_D-1}} \right\}_{\Lambda_A}^{\Lambda_A+\Lambda_D-1} \qquad \text{or} \\
&= \left\{ 1-(1-a_S)\frac{i-\Lambda_A}{\Lambda_D-1} \right\}_{\Lambda_A}^{\Lambda_A+\Lambda_D-1} \\
\{a_i\}_{\Lambda_A+\Lambda_D}^{\Lambda-\Lambda_R-1} &= \{a_S\}_{\Lambda_A+\Lambda_D}^{\Lambda-\Lambda_R-1} \\
\{a_i\}_{\Lambda-\Lambda_R}^{\Lambda-1} &= \left\{ a_S \left(\frac{\xi}{a_S}\right)^{\frac{i-(\Lambda-\Lambda_R)}{\Lambda_R-1}} \right\}_{\Lambda-\Lambda_R}^{\Lambda-1} \qquad \text{or} \\
&= \left\{ a_S - a_S \frac{i+\Lambda_R-\Lambda}{\Lambda_R-1} \right\}_{\Lambda-\Lambda_R}^{\Lambda-1}
\end{aligned} \tag{79}$$

with $\Lambda_X = \lfloor \Delta_X . f_s \rfloor \ \forall \ X \in (A, D, R)$ and ξ being a small value that provides a satisfactory *fade in* and *fade out*, e.g. $\xi = 10^{\frac{-80}{20}} = 10^{-4}$. The lower the ξ, the slower the *fade*, like the α illustrated in Figure 15. The right side of Equations 79 can cater for 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 outlines of attack or decay. To aply the ADSR envelope, simply multiply the amplitudes of Equation 79 by the samples of the note or musical except:

$$\left\{ t_i^{ADSR} \right\}_0^{\Lambda-1} = \{t_i . a_i\}_0^{\Lambda-1} \tag{80}$$

The music piece *ADa and SaRa* explores many configurations of the ADSR envelope. [9]

4 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. More specifically, what follows is a summary of academic music composition theory and praxis. This section does not benefit from equations that dictate the amplitude of each sample as deeply as the previous sections. Even so, we understand that this content is very useful for synthesizing music given all the content above. The concepts are given algorithmic implementations in MASS [6] and can be further formalized [44], although at a cost of intelligibility which we chose to avoid in this exposition.

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,

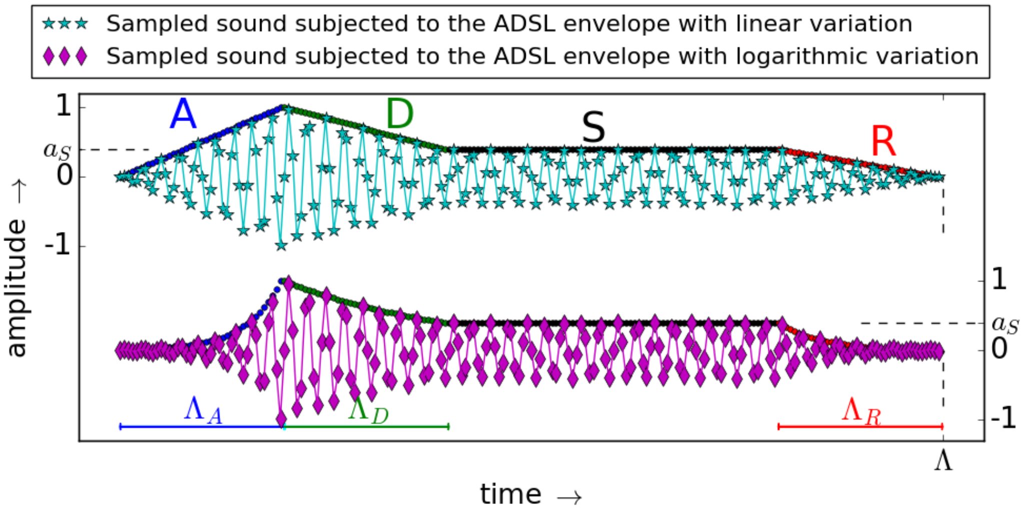

Fig. 22. An ADSR envelope (*Attack, Decay, Sustain, Release*) applied to an arbitrary sound sequence. The linear variation of the amplitude is above, in blue. Below the amplitude variation is exponential.

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 'horizontal thought'. The fundamental frequency 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. We will start by considering these aspects of musical organization as they are more traditional in music theory and usually easier to understand.

4.1 Tuning, intervals, scales and chords

4.1.1 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, e.g. in ceremonial/religious and ethnic contexts [72]. The semitones need not to be equivalent, as will become clear in the next paragraphs, but, roughly, the factor given by $\varepsilon = 2^{\frac{1}{12}}$ defines a semitone. This entails 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 successive semitones (or half steps), equidistants to the human ear, yield an octave, equivalent to the human ear. In other words, if $f = 2^{\frac{1}{12}} f_0$, there is a semitone between f_0 and f.

This absolute accuracy is usual in computational implementations. Real music instruments, however, often present semitones that are not exactly $2^{\frac{1}{12}}$ because the (real) harmonics of the notes are not perfect multiples of the fundamental frequency. The fixed interval $\varepsilon = 2^{\frac{1}{12}}$ characterizes an equally tempered tuning but there are other tunings. The first formalizations of tunings (that the scientific tradition has repported) date from around two thousand years before the advent of the equal temperament [54].). Two emblematic tunings are:

- The **just intonation**, defined by association of intervals with ratios of low-order integers, as found in the harmonic series. E.g. the white piano keys from C to C are achieved by the ratios of frequency: 1, 9/8, 5/4 4/3, 3/2, 5/3, 15/8, 2/1. The following intervals are also often 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 'white piano keys' become: 1, 9/8, 81/64, 4/3, 3/2, 27/16, 243/128, 2/1. Also often 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[21] non-integer real values can be used as factors of $\varepsilon = 2^{\frac{1}{12}}$ between frequencies, or one can maintain integer values and change ε. For example, a tuning that approximates the harmonic series is proposed with the equal division of the octave in 53 notes: $\varepsilon = 2^{\frac{1}{53}}$. [71] Note that if S_i is a pitch sequence related by means of ε_1, the sequence with the same notes, but related by ε_2, is $S_i' = \left\{s_i'\right\} = \left\{s_i \frac{\varepsilon_1}{\varepsilon_2}\right\}$. The music piece *Micro tone* uses microtonal features and its code is part of MASS toolbox.

4.1.2 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 1 summarizes the intervals: traditional notation, qualifications of consonance and dissonances, and number of semitones.

Table 1. Music intervals together with their traditional notation, basic classification for dissonances and consonances, 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. Another special case is the tritone (A4 or aug4, d5 or dim5, tri or TT). This interval is consonant in some cultures. For tonal music, the tritone indicates dominant (chord or harmonic field, see Section 4.2) and seeks urgent resolution into a third or sixth. Due to this instability it is considered a dissonant interval.

consonances		
	traditional notation	number of semitones
perfect:	P1, P5, P8	0, 7, 12
imperfect:	m3, M3, m6, M6	3, 4, 8, 9
dissonances		
	traditional notation	number of semitones
strong:	m2, M7	1, 11
weak:	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 nomenclature, based on conveniences for tonal music and practical aspects of manipulating notes, can be specified as follows [54, 72]:

[21] Micro-tonality is the use of intervals smaller than one semitone and has ornamental and structuring functionalities in music. The division of the octave in 12 notes has physical grounds but is still a *convention* adopted by western classical music. Other tunings are incident, e.g. a traditional Thai music style uses an octave division in seven notes equally spaced ($\varepsilon = 2^{\frac{1}{7}}$), which allows intervals quite different than those found when $\varepsilon = 2^{\frac{1}{12}}$ [72].

- Intervals are inspected first by the number of steps between notes. The simple intervals are: first (unison), second, third, fourth, fifth, sixth, seventh and eighth (octave). Each of these intervals are related to one step less the their names suggest: a third is an interval with two steps. As can be noticed in Table 1, one step is not one semitone. A step, in this sense, is yield by two consecutive notes in a musical scale. A scale for now can be regarded as any arbitrary monotonic sequence of pitches and will be discussed in the next section.
- 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, fourth, 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. The tritone is an exception to this rule because it is a dissonant interval and cannot be major or minor.
- 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 is followed by a third or a fifth by the movement one of the notes.
- Tritone is a dissonance in Western music because it characterizes the "dominant" chord in the tonal system (see Section 4.2) and represents (or yields) instability. Some cultures considers the interval 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 (P1, P4, P5, or P8), or a major interval (M2, M3, M6 or M7), increased by one semitone results in an augmented interval (e.g. aug3 has five semitones). The augmented forth is also called tritone (aug4 tri TT).
- A perfect interval or a minor interval (m2, m3, m6 or 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 yields a harmonic interval.
- Notes played as a sequence in time yields a melodic interval. When the lowest note comes first there is an ascending interval, while a descending interval is observed when the highest comes note first.
- A simple interval is inverted if the lower pitch is raised one octave, or if the upper pitch is lowered one octave. 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-versa, etc). An inverted perfect interval is a perfect interval as well.
- An interval wider than an octave (e.g. ninth, tenth, eleventh) is called a 'compound interval' and is classified in terms of the interval between the same notes but in the same octave. Their notation can be achieved 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 have the same number of semitones of other intervals (e.g. minor, major or perfect) and are consequences of the tonal system. Scale notes 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*. To grasp what this entails for intervals, let the second degree (second note) of a scale to be one semitone from the first degree. Let also the leading tone (i.e. the seventh degree at one ascending semitone from the first degree). There is a diminished third between the seventh and second scale degrees [40]. Notice that the dim3 is only two semitones wide, as is the major second (or e.g. an doubly-augmented unisson!).

This description summarizes the traditional theory of musical intervals [40]. The music piece *Intervalos entre alturas* explores these intervals in both independent and related manners. The source code is available online in the MASS toolbox [9].

4.1.3 Scales. A scale is an ordered set of pitches. Strictly speaking, any (ordered) set of pitches can be considered a scale. The complexity about musical scales lean on the tradition, i.e. on the scales and their uses which result from practice throughout history. Usually, scales repeat at each octave. The ascending sequence with all notes from the octave division in 12 equal intervals ($\varepsilon = 2^{\frac{1}{12}}$) is known as the chromatic scale within the equal temperament. There are 5 perfectly symmetric divisions of the octave within the chromatic scale. These divisions are often regarded as scales themselves owing to the easy and peculiar uses they entail.

Let e_i be integers indexed by i such that $f = \varepsilon^{e_i} f_0$, where f_0 is any fixed frequency. The symmetric scales mentioned above can be expressed as:

$$\begin{aligned}
\text{chromatic} = E_i^c &= \{e_i^c\}_0^{11} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\} &&= \{i\}_0^{11} \\
\text{whole tones} = E_i^{wt} &= \{e_i^{wt}\}_0^5 = \{0, 2, 4, 6, 8, 10\} &&= \{2.i\}_0^5 \\
\text{minor thirds} = E_i^{mt} &= \{e_i^{mt}\}_0^3 = \{0, 3, 6, 9\} &&= \{3.i\}_0^3 \\
\text{major thirds} = E_i^{Mt} &= \{e_i^{Mt}\}_0^2 = \{0, 4, 8\} &&= \{4.i\}_0^2 \\
\text{tritones} = E_i^{tt} &= \{e_i^{tt}\}_0^1 = \{0, 6\} &&= \{6.i\}_0^1
\end{aligned} \tag{81}$$

For example, the third note of the whole tone scale with $f_0 = 200Hz$ is $f_3 = \varepsilon^{e_2^{wt}}.f_0 = 2^{\frac{4}{12}}.200 \approxeq 251.98Hz$. These 'scales', or patterns, generate stable structures by their internal symmetries and can be repeated in a sustained way which are musically effective. Section 4.7 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 MASS toolbox.

The *diatonic scales* are scales with seven notes in which the sequence consists of five whole tones and two semitones (in each octave). There are seven of them:

$$\begin{aligned}
\text{aeolian} = \text{natural minor scale} &= \\
= E_i^m = \{e_i^m\}_0^6 &= \{0, 2, 3, 5, 7, 8, 10\} \\
\text{locrian} = E_i^{lo} = \{e_i^{lo}\}_0^6 &= \{0, 1, 3, 5, 6, 8, 10\} \\
\text{ionian} = \text{major scale} &= \\
= E_i^M = \{e_i^M\}_0^6 &= \{0, 2, 4, 5, 7, 9, 11\} \\
\text{dorian} = E_i^d = \{e_i^d\}_0^6 &= \{0, 2, 3, 5, 7, 9, 10\} \\
\text{phrygian} = E_i^p = \{e_i^p\}_0^6 &= \{0, 1, 3, 5, 7, 8, 10\} \\
\text{lydian} = E_i^l = \{e_i^l\}_0^6 &= \{0, 2, 4, 6, 7, 9, 11\} \\
\text{mixolydian} = E_i^{mi} = \{e_i^{mi}\}_0^6 &= \{0, 2, 4, 5, 7, 9, 10\}
\end{aligned} \tag{82}$$

They have only major, minor and perfect intervals. The unique exception is the tritone found as an augmented fourth 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:

$$\begin{aligned} \{d_i\} &= \{2, 2, 1, 2, 2, 2, 1\} \\ e_0 &= 0 \\ e_i &= d_{(i+\kappa)\%7} + e_{i-1} \quad for \;\; i > 0 \end{aligned} \tag{83}$$

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 $e^l_i = d_{(i+2)\%7} + e^l_{i-1}$. Then, $\kappa = 2$ for the lydian scale.

The minor scale have two additional forms, named harmonic and melodic:

$$\begin{aligned} \text{natural minor} &= E^m_i = \{e^m_i\}_0^6 = \\ &= \{0, 2, 3, 5, 7, 8, 10\} \\ \text{harmonic minor} &= E^{mh}_i = \{e^{mh}_i\}_0^6 = \\ &= \{0, 2, 3, 5, 7, 8, 11\} \\ \text{melodic minor} &= E^{mm}_i = \{e^{mm}_i\}_0^{14} = \\ &= \{0, 2, 3, 5, 7, 9, 11, 12, 10, 8, 7, 5, 3, 2, 0\} \end{aligned} \tag{84}$$

The different ascending and descending contour of the melodic minor scale is required in tonal harmony contexts. The minor scale has one whole tone between the seventh and eigth (or first) degrees but the separation by one semitone is critical to the polarization of the first degree. This is not necessary in the descending trajectory, and therefore the scale recovers the standard form. The harmonic scale presents the modified seventh degree but does not avoid the augmented second between the sixth and seventh degrees; it does not consider the melodic trajectory and thus does not need to avoid the aug2 [59].

Although it is not a traditional scale, the harmonic series is often used as such:

$$\begin{aligned} H_i =&\{h_i\}_0^{19} = \\ =&\{0, 12, 19 + 0.02, 24, 28 - 0.14, 31 + 0.2, 34 - 0.31, \\ &36, 38 + 0.04, 40 - 0.14, 42 - 0.49, 43 + 0.02, \\ &44 + 0.41, 46 - 0.31, 47 - 0.12, \\ &48, 49 + 0.05, 50 + 0.04, 51 - 0.02, 52 - 0.14\} \end{aligned} \tag{85}$$

In this scale, the frequency of the ith note h_i is the frequency of ith harmonic $f = \varepsilon^{h_i} f_0$ from the spectrum generated by f_0. Natural sounds have such frequencies (as discussed in Section 2) usually with deviations from the expected values and with noise.

Many other scales can be expressed using the framework exposed in this section, e.g. the pentatonic scales and the modes of limited transposition of Messiaen [22].

One last observation: the words *scale* and *mode* are often used as synonyms both in the literature and in colloquial discussions. The word *mode* can also be used to mean two other things:

- an unordered set of pitches (i.e. an unordered scale).
- A scale used in the context of modal harmony, in the sense presented in Section 4.2.

4.1.4 Chords. A musical chord is implied by the simultaneous occurrence of three or more notes. Chords are often 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 the root, the chord is in the root position, otherwise it is an inverted chord. A closed position is any in which no chord note fits between two consecutive notes [40], any non-closed position is an open position. In closed and fundamental positions, and with the fundamental in 0, triads can be expressed as:

$$\begin{aligned}
\text{major triad} = A_i^M &= \{a_i^M\}_0^2 = \{0, 4, 7\} \\
\text{minor triad} = A_i^m &= \{a_i^m\}_0^2 = \{0, 3, 7\} \\
\text{diminished triad} = A_i^d &= \{a_i^d\}_0^2 = \{0, 3, 6\} \\
\text{augmented triad} = A_i^a &= \{a_i^a\}_0^2 = \{0, 4, 8\}
\end{aligned} \tag{86}$$

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. These are often interpreted as the result of further extending the succession of thirds. E.g. $\{0, 2, 4, 7\}$ will often be understood as a major chord with a major ninth (a major ninth has 14 semitones and $14 - 12 = 2$).

For general guidance:

- A fifth confirms the root (fundamental). There are theoretical discussion about why this happens, and the most usual arguments are that the fifth is the first (non-octave) harmonic of a note and that the harmonics of the fifth are in the harmonics of the fundamental. Important here is to grasp the fact that musical theory and practice assures that the fifth establishes the fundamental as the root of a chord.
- Major or minor thirds from the root entails 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 preference is, in descending order: the root, fifth, third and seventh.
- Note omission is avoided in the triad. If needed, the fifth is first considered for omission, then third and then 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 – such as fourths and seconds – are recurrent in compositions of advanced tonalism or experimental character.
- The repetition of chord successions (or of characteristics they hold) fixes a trajectory and makes it possible to introduce exotic arrangements without implying musical incoherence.

4.2 Atonal and tonal harmonies, harmonic expansion and modulation

Omission of basic tonal structures is the key to achieving modal and atonal harmonies. In the absence of minimal tonal organization, harmony is (usually) considered modal if the notes match with some diatonic scale (see Equations 82) or if there is only a small number of notes. If basic tonal progressions are absent and notes do not match any diatonic scale and are enough diverse and dissonant (between themselves) to avoid reduction of the notes by polarization[22], the harmony is atonal. In this classification, the modal harmony is not tonal or atonal and is reduced to the incidence of notes within a (most often diatonic) scale 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 very easily if not avoided [39].

[22] By polarization we mean having some notes that are way more important than others and to which the other notes are ornaments or subordinates.

4.2.1 Atonal harmony. In fact, atonal music techniques avoid that a direct relation of the notes with modes and tonality be established. Manifesting such atonal structures is of such difficulty that the dodecafonism emerged. 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 automatically searches for tonal elements in music and obstinately 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 musical piece, it is possible to:

- Repeat notes. By considering immediate repetition as an extension of the previous incidence, the use of the same note in sequence does not add relevant information.
- To play adjacent notes (e.g. of a dodecafonic progression) at the same time, making harmonic intervals and chords.
- Use durations and pauses with liberty, respecting notes order.
- Vary note sequences by temporal magnification and translation or pitch transposition and by pitch sequence inversion, retrograde and retrograde inversion. See Sections 4.5 and 4.10 for what these terms mean.
- Make variations in orchestration, articulation, spatialization, among other possibilities in presenting the same notes.

The atonal harmony can be observed, paradigmatically, within these presented conditions (which is 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. Most frequently, atonal music is not strictly dodecafonic, but "serial", i.e. they use the same kind of techniques based in a sequence of notes (called the series or row).

4.2.2 Tonal harmony. In the XX century, music with emphasis on sonorities/timbres and rhythm, extended the concepts of tonality and harmony. Even so, tonal harmony is very often in artistic movements and commercial venues. In addition, dodecafonism itself is sometimes considered of tonal nature because it was conceived to deny tonal characteristics of polarization. In tonal or modal music, chords – like the ones listed in Equations 86– built on the root note of each degree of a scale – such as listed in Equations 82 – form the pillars of harmony. Tonal (and modal) harmony aims at understanding the incidence of chord progressions and chaining rules. Even a monophonic melody entails harmonic fields, making it possible to perceive the chord progression even in unaccompanied melodies.

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, as in Equations 84). 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 the 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 in a minor scale) and of the forth degree $\hat{4}, \hat{6}, \hat{1}$ are also important. The triads build on the other degrees less important then these and are usually understood in relation to them. The 'traditional harmony' comprises conventions and stylistic techniques to create progressions with such chords [59].

The 'functional harmony' ascribes functions to the three main 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 on 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 fourth 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 tonal discourse aims at confirming the tonic using tonic-dominant-tonic chains which are expanded by using other chords in numerous 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 but are basically the same as the chords they are associated to. Table 2 shows relations between the triads built in each degree of the major scale.

Table 2. Summary of tonal harmonic functions on the major scale. Tonic is the musical center, the dominant goes to the tonic and the subdominant moves the music away from the tonic. The three chords can, in principle, be freely replaced by their respective relative or 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 $\hat{4}\#$. The diminished chord $\hat{7}, \hat{2}, \hat{4}$, is generally considered a 'dominant seventh chord with the root omitted' [38]. In the minor mode, there is a change in $\hat{7}$ by an ascending semitone to achieve a separation between $\hat{7}$ and $\hat{1}$ of a semitone. This is important for the dominant function (which should lean 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 the counter-relative remains a third above.

4.2.3 Tonal expansion: individual functions and chromatic mediants. Each chord can be stressed 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 most basic chords and from the notes of the scale. The unique condition is that the occurrence presents a coherent trajectory of dominants and subdominants (or their relatives and counter-relatives) to the original tonality.

There are two mediants, or 'chromatic mediants', for each chord: the upper mediant, formed with 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. 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 as an expansion and dissolution of tonalism, with strong and impressive effects although they are simple, consonant triads. Chromatic mediants are used since the end of Romanticism by Wagner, Lizt, Richard Strauss, among others [57, 59].

4.2.4 Modulation. Modulation is the change of key (tonic, or tonal center) in music, being characterized by start and end keys, and transition artifacts. Keys are always (thought of as) related by fifths and their relatives 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 in 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 by 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-versa) is a 'parallel modulation'. Keys with same tonic but different (major/minor) qualities are known as homonyms.

The dominant has great importance and is a natural pivot in modulations, a fact that leads to the circle of fifths [11, 38, 57, 59]. Other inventive ways to modulate are possible, to point but one common example, the minor thirds tetrad (E_i^{tm} in Equations 81) 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 implemented online as part of MASS toolbox [9].

4.3 Counterpoint

Counterpoint is 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 [32, 60]. It is possible to summarize counterpoint rules, and it is known that Beethoven – among others – also outlined such a digest of counterpoint.

The purpose of (scholastic) 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, as depicted in Figure 23. The parallel motion is a direct motion in which the starting and final intervals are the same. The golden rule here is to take care of the direct motions, avoiding them when ending in a perfect consonance. The parallel motion should occur only between imperfect consonances and no more than three consecutive times. Dissonances can be forbidden or used when followed and preceded by consonances of neighbor degrees, i.e. adjacent notes in a scale. The motions that lead to a neighbor note in the scale sound coherent and are prioritized. When having 3 or more voices, the melodic importance lies mainly in the highest and then in the lowest of the voices [32, 60, 67].

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

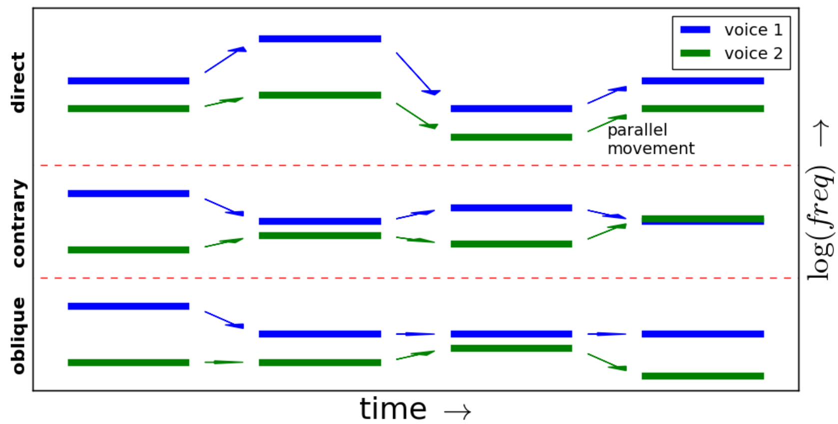

Fig. 23. Different motions of counterpoint aiming to preserve independence between voices. There are 3 types of motion: direct, contrary and oblique. The parallel motion is a type of direct motion. A voice is often called a 'melodic line' or a melody.

4.4 Rhythm

Rhythmic notion is dependent on events separated by durations [40], 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 $100ms$ [53]. It is possible to summarize the durations heard as rhythm or pitch, and its transition, as in Table 3 [24, 53].

Table 3. Durations heard as rhythm, as pitch and transition.

		perception of durations as rhythm										
duration (s)	...	**32,**	**16,**	**8,**	**4,**	**2,**	**1,**	**1/2,**	**1/4,**	**1/8,**	‖	... transition
frequency (Hz)		1/32,	1/16,	1/8,	1/4,	1/2,	1,	2,	4,	8,		

			transition			
duration (s)	**rhythm ...**	‖	$\frac{1}{16} = 62.5ms$,	$\frac{1}{20} = 50ms$	‖	**... pitch**
frequency (Hz)			16,	20		
			transition			

duration (s)	transition ...	‖	1/40	1/80	1/160	1/320	1/640	...
frequency (Hz)			**40**	**80**	**160**	**320**	**640**	
			perception of durations as pitch					

The transition span in Table 3 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 [53, 54]. The rhythmic metric is commonly based on a basic duration called pulse, whose typical durations range between $0.25and1.5s$ (240 and $40BPM$, respectively [23]). In music education and cognitive studies, it is

[23] BPM stands for Beats Per Minute and is just a frequency measure like Herz, but is the number of incidences per minute intead of second. BPM is often used as a measure of musical tempo and of heart rate.

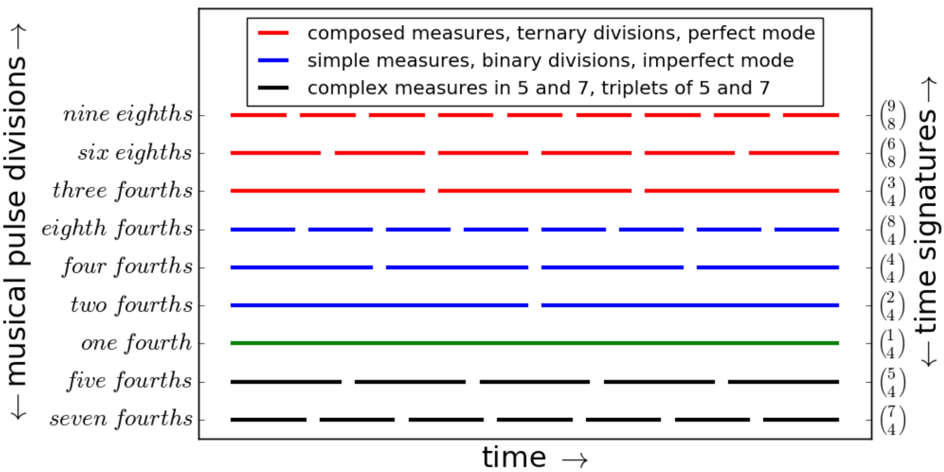

Fig. 24. Divisions and groupings of the musical pulse for establishing a metric. Divisions of the quarter note, regarded as the pulse, is presented on the left. The time signature yielded by groupings of the music pulse is presented on the right.

common to associate this range of frequencies with the durations of the heart beat, movements of respiration and steps of a walking or running person [40, 54].

The pulse is subdivided into equal parts and is also repeated in sequence. These relations (division and concatenation) usually follow relations of small integers. By far, the most often musical pulse divisions (and their sequential groupings), in written and ethnic music, are: 2, 4 and 8; 3, 6 (two groups of 3 or 3 groups of 2), 9 and 12 (three and 4 groups of 3). At last, the prime numbers 5 and 7, completing 1-9 and 12. Other metrics are less common, like division or grouping in 13, 17, etc, and are mainly used in experimental music or classical music of the XX and XXI centuries. No matter how complex they seem, metrics are almost always compositions and decompositions of 1-9 equal parts [34, 54]. A schematic illustration is shown in Figure 24.

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 the units (the first subdivision) and are called downbeats. 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 suffer a ternary division yields two units divided into three units each: strong (subdivided in strong, weak, weak) and weak (also subdivided in strong, weak, weak). Another way to perform the division in 6 is ternary division units that subdivides as binary, resulting in: a strong unit (subdivided in strong and weak) and two weak units (subdivided in strong and weak each).

An accent in the weak beat is the ‘backbeat’, whereas a note starting in a weak beat persisting across the strong beat is a ‘syncope’. These are often found in ethnic and popular music and was used with parsimony in classical music before the XX century.

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 beats. In extreme cases, rythmic metric cannot be perceived [54]. Small variations along the temporal grid are crucial for musical interpretation and styles [23].

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. Accordingly, let P_i^j be the i-th unit at grouping level j: P_{10}^0 is the tenth pulse, P_3^1 is the third grouped unit (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 musical pieces. In other words, a duration given by $P_i^{min(j)}$, $\forall\ i$, should be greater than $50ms$ and the durations summed together $\sum_{\forall i} P_i^{\max(j)}$ should be less than a few minutes or, at most, a few hours.

Each level j has some parts i. When i has three different values (or multiple of three) there is a perfect (i.e. ternary) relation. When i has only two, four or eight possible 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:

$$P_{\{i_k\}}^{\{j_k\}} \tag{87}$$

where j_k is the grouping level and i_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 constitutes a note. The music piece *Poli Hit Mia* uses different metrics (available as part of MASS toolbox).

4.5 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 [13]. The concept of arcs is essential in this context: by departing from a place and returning, an arc is made. One important case is the arc from and to the absence of a unit: from the beginning to the end. The audition of melodic and harmonic lines is permeated by arcs due to the cognitive nature of the musical hearing: as the mind divides an excerpt, and groups excerpts, each of the units yields an arc. Accordingly, the note can be considered the smallest (relevant) arc, and each motif and melody as an arc as well. Each beat and subdivision, each measure and music section, constitutes an arc. Music in which the arcs do not present consistency with one another can be understood as 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, magnified and permuted, to mention just a few possibilities. Temporal arcs, like a melody, a motif, a measure or a note, are also prone to variations. Let $S_j = \left\{ s_j = T_i^j = \{t_i^j\}_0^{\Lambda_j - 1} \right\}_0^{H-1}$ be a sequence of H musical events s_j, each event with its Λ_j samples t_i^j (refer to the beginning of this Section 4 if needed). Bellow is a list of basic techniques for variation.

- Temporal translation is a displacement δ of a specific material to another instant $\Gamma' = \Gamma + \delta$ of the music. It is a variation that changes temporal localization in a music: $\left\{ s_j' \right\} = \left\{ s_j^{\Gamma'} \right\} = \left\{ s_j^{\Gamma + \delta} \right\}$ where Γ is the duration between the beginning of the piece (or considered section) and the first event s_0 of original structure S_j, and δ is the time offset of the displacement.
- Temporal expansion or contraction is a change in duration of each arc by a factor $\mu\ :\ s_j'^{\Delta} = s_j^{\mu_j . \Delta}$. Possibly, $\mu_j = \mu$ is constant.

- Temporal reversion consists on generating a sequence with elements in the reverse order of the original sequence S_j, thus: $S'_j = \left\{s'_j\right\}_0^{H-1} = \left\{s_{(H-j-1)}\right\}_0^{H-1}$.
- Pitch translation, 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: $\left\{s'_j\right\} = \left\{s_j^{\Xi'}\right\} = \left\{s_j^{\Xi+\tau}\right\}$ 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: $\Xi = \Xi_0 + 12\log_2\left(\frac{f}{f_0}\right)$.[24]
- Interval inversion is either: 1) the inversion of note pitch order, within octave equivalence, such as described in Section 4.1.2, and $S'_j = \{s'_j\}_0^{H-1} = \left\{s_j^{\varepsilon.f_0}\right\}$ with selective $s_j = 2,\ 1/2,\ \text{or}\ 1$; or 2) the inversion of interval orientation. In the former case, the number of semitones is preserved in the "strict invertion": $S'_j = \{s'_j\}_0^{H-1} = \left\{s_j^{-\varepsilon_j.f_0}\right\}$, 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 which are out of the 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 usual to associate $\tilde{n} < 0$ (events advance) with the clockwise rotation and $\tilde{n} < 0$ (elements delay) with the anti-clockwise rotation. Additional information about rotations is given in Section 4.7.
- 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, or $s'_j = s_j^{*_j}$, where $*_j$ is the new characteristic incorporated by element s'_j.
- Accompaniment. Both orchestration and melodic lines present when S_j occurs can suffer modifications and be considered as a variation of S_j itself.

From these processes, many others are derived, such as the inverted retrograde, the temporal contraction with an external suffix, etc. Variations are often thought about in the terms above but are also often very loose, such as an arbitrary shuffle of the notes in a melody which the composer finds interesting. As a result, a whole process of mental and neurological activity is unleashed for relating the arcs, 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. Also, it is known that regions of the human brain responsible for sonic processing are also used for other activities, such as for performing verbal discourse and mathematics [54, 56].

[24]In the MIDI protocol, $\Xi_{f_0} = 55Hz$ when pitch $\Xi_0 = 33$ (an *A1* note). Another good MIDI reference is $\Xi_{f_0} = 440Hz$ and $\Xi_0 = 69$ (*A4*). The difference $(\Xi_1 - \Xi_2)$ is 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 3).

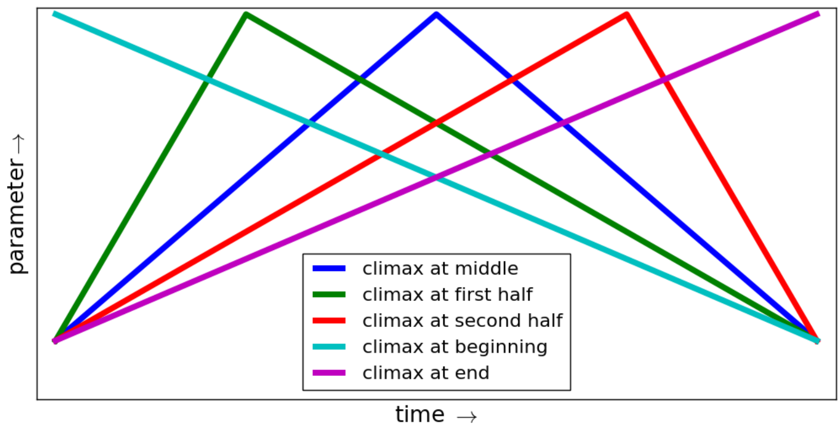

Fig. 25. Canonical distinctions of musical climax in a given melody and other arcs. The possibilities considered are: climax at the beginning, at the first half, in the middle, in the second half and at the end. The x and y-axis are not properly specified since the parameters can be non-existent, as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

Paradigmatic structures guide the creation of new musical material. One of the most central is the tension/relaxation dipole. Traditional dipoles are tonic/dominant, repetition/variation, consonance/dissonance, coherence/rupture, symmetry/asymmetry, equivalence/difference, arrival/departure, near/far, stationary/moving, etc. All these dipoles are often thought of as parallel or even as equivalent. 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 a description dedicated to directional and cyclic arcs.

For a scholastic discussion on the composition of motives, phrases, melodies, themes and music form (such as rondo, ternary, theme and variations) see [61].

4.6 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 traditional music theory. It is usual to distinguish between arcs whose climax is 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 [61][25].

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 [36, 61]. 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 numerous distinct arcs, with different sizes and climax. This is an interesting and useful resource, and the correlation of arcs yields coherence [57].

[25] A case which resembles a note without the fundamental frequency. Brass instruments can emit such notes and the listener identifies the fundamental frequency even though it is absent.

In practice, and historically, there is special incidence and use of the golden ratio. The Fibonacci sequence might be generalized as follows in order for any two numbers be used to approach the golden ratio. Given any two numbers x_0 and x_1, define the elements of the sequence $\{x_n\}$ 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 discrepant initial values. E.g. let $x_0 = 1$, $x_1 = 100$ and $y_n = \frac{x_n}{x_{n+1}}$, the error for the 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 the same error progression, but starts at the second step of a more discrepant initial setting ($\frac{1}{1} \approx \frac{100+1=101}{100}$). One might benefit from the On-Line Encyclopedia of Integer Sequences (OEIS [62]) for exploring various sequences.

The music piece *Dirracional* uses arcs into directional structures. Its source code is available online as part of MASS toolbox [9].

4.7 Cyclic structures

The philosophical understanding that human thought is founded on the concept of similarities and differences (e.g. perceived in stimuli), places symmetries at the core of cognition [26]. 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 can express any symmetry in a finite system [16]. In music, permutations are ubiquitous in scholastic techniques, which confirms their central role. The successive application of permutations generates cyclic arcs [15, 27, 74]. Two academic studies were dedicated to generating musical structures using permutation groups [28, 29]. Any permutation set can be used as a generator of algebraic groups [15]. The properties defining a group G are:

$$\begin{aligned}
\forall\ p_1, p_2 \in G \Rightarrow \quad & p_1 \bullet p_2 = p_3 \in G \\
& \text{(closure property)} \\
\forall\ p_1, p_2, p_3 \in G \Rightarrow \quad & (p_1 \bullet p_2) \bullet p_3 = p_1 \bullet (p_2 \bullet p_3) \\
& \text{(associativity property)} \\
\exists\ \ e \in G: \quad & p \bullet e = e \bullet p\ , \quad \forall\, p \in G \\
& \text{(existence of the identity element)} \\
\forall\ \ p \in G,\ \exists\, p^{-1}: \quad & p \bullet p^{-1} i = p^{-1} \bullet p = e \\
& \text{(existence of the inverse element)}
\end{aligned} \tag{88}$$

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 generated by p would be infinite). 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 notes or other musical structures, yields a cyclic arc.

These arcs can be established by using a set of 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 is repeated until it reaches the initial ordering. The sequence of different orderings is a *peal*. Table 4 presents a traditional *peal*, named "Plain Change" [27], for 3 bells (1, 2 and 3), which explores all possible orderings. Each line indicates one bell ordering to be played. Permutations occur between each line. In this case, the musical structure consists of permutations and some different permutations add into a cyclic behavior.

Table 4. 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 2.6, 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 MASS toolbox presents a computational implementation that isolates permutation application to sonic characteristics, in order to deliver musical structures [9, 28, 29].

4.8 Serialism and post-serial techniques

Expanding the concepts in Section 4.2.1, sequences of characteristics can be predefined and used throughout a musical piece. These sequences can be of intensities, timbre, durations, density of events, etc. The sequences can be used very strictly or loose, such as by skipping some elements. The sequences can be of different sizes, yielding arcs until the initial condition is reached again (i.e. cycles). One paradigmatic case is the "total serialism" where all the musical characteristics are serialized. Although the use of sequences is inherent to music (e.g. scales, metric pulses), their use with greater emphasis than tonal (or modal) elements in western music, as an artistic trend, took place only in the first half of the twentieth century and is called "serialism". Post-serial techniques are numerous, but here is a description of important concepts:

- Spectralism: consists on use the (Fourier) spectrum of a sound or the harmonic series for music composition, such as to derive harmonies, sequences of pitches or a temporal evolution of the overall spectrum [35]. In other words, the most prominent frequencies can be used as pitches, real notes can be used to mimic an original spectrum (e.g. use piano notes to mimic the spectrum of a spoken sentence) or portions of the spectrum made to vary [?].
- Spectromorphology: can be considered a spectral music (spectralism) theoretical framework [58, 63] that examines the relation between sound spectra and their temporal evolution. The theory poses e.g. different onsets, continuations and terminations; characteristics of (sonic) "motion"; and spectral density.
- Stochastic music: the use of random variables to describe musical elements are largely considered in stochastic music [73]. In summary, one can use probability distributions for the synthesis of sounds and larger scale musical elements. Changes in these distributions or in other characteristics yield the discourse.
- Textures: sounds can be assembled in terms of a "sonic texture". Here, a texture is often thought about very abstractly as a sonic counterpart of visual texture. Parameters that can be used for such are: range between highest and lowest note, density of notes, durations of notes, motives, number of voices, etc. If the sounds are small enough (typically ¡ $100ms$) the process can be though of in terms of *granular synthesis* [53].

4.9 Musical idiom?

In numerous studies, there are models, discussions and exploitation of a 'musical language'. Some of them are linguistic theories applied to music and some discern different 'musical idioms' [24, 41, 57, 59]. Simply put, a musical idiom is the result of chosen materials together with variation techniques and relations established between elements along a music piece. In these matters, dichotomies are prominent, as explained in Section 4.5: repetition and variation, relaxation and tension, stability and instability, consonance and dissonance, etc. A thorough discussion of what can be considered a musical language is out of the scope of this article, but this brief consideration of the subject is useful as a onvergence of all the previous content.

4.10 Musical usages

First, the basic note was defined and characterized in quantitative terms in Section 2. Next, the internal note composition was addressed within both internal transitions and elementary sonic treatment (Section 3). Finally, this section aims at organizing these notes in music. The numerous resources and consequent infinitude of praxis possibilities is typical and highly relevant for artistic contexts [59, 70].

There are studies and further developments 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 superimposition of rhythms in different metrics, constituting what is called *polyrhythm*. The music piece *Poli-hit mia* in MASS toolbox [9] explores these simultaneous metrics by impulse trains convolved with notes.

Microtonal scales are important for 20th century music [71] and yielded diverse remarkable results throughout history, e.g. fourths of a tone ($\epsilon = 2^{\frac{1}{24}}$) is often used in some genres of Indian and Arabic music. The musical sequence *MicroTom* in MASS toolbox [9] explores these resources, including microtonal melodies and harmonies with many notes in a very reduced note scope.

As in Section 3.6, relations between parameters are powerful ways to achieve musical pieces. The number of permuted notes can vary during the music, revealing relationship with piece duration. Harmonies can be made from triads (Equations 86) with duplicated notes at each octave and more numerous duplication when the depth and frequency of vibratos are lower (Equations 55, 56, 57, 58, 59). Incontestably, the possibilities are very wide, which is evidenced in the wide range of musical pieces and styles.

The symmetries at octave divisions (Equation 81) and the symmetries presented as permutations (Table 4 and Equations 88) can be used together. In the music piece *3 trios* this association is done in a systematic way in order to achieve a specific style. This is an instrumental piece, not included as a source code but available online [1].

PPEPPS (Pure Python EP: Project Solvent) is an EP (Extended Play) synthesized using resources presented in this study. With minimal parametrization, the scripts generate complete musical pieces, allowing easy composition of sets of music. A simple script of a few lines specifies music delivered as 16 bit 44.1kHz PCM files (WAVE). This facility and technological arrangement creates aesthetic possibilities for both sharing and education.

5 CONCLUSIONS AND FURTHER DEVELOPMENTS

In our understanding, this article is effective in relating musical elements to digital audio. We aimed at achieving a concise presentation of the subject because it involves many knowledge fields, and therefore can very easily blast into thousands of pages. Some readers might benefit from the text alone, but the *Scripts* in the MASS toolbox, where all the equations and concepts are directly and simply implemented as software (in Python), are very helpful for one to achieve elaborated implementations and deeper understandings. The scripts include routines that render musical pieces that illustrate the concepts in practical contexts. This is valuable since art (music) can involve many non-trivial processes and is often deeply glamorized, which results in a nearly unmanageable terrain for a newcomer. Moreover, this didactic report and the supplied open source scripts should facilitate the use of the proposed framework. The Supporting Information document holds listings of sections, equations, figures, tables, scripts and other documents.

The possibilities provided by this exposition pour from both the organization of knowledge and the ability to achieve sounds which are extremely true to the models. For example, one can produce noises with arbitrary resolution of the spectrum and a musical note can be synthesized with the parameters (e.g. of a vibrato) updated sample-by-sample. Furthermore, software for synthesis and processing of sounds for musical purposes by standard restrict the bit depth to 16 or 24. This is achievable in this framework but by standard Python uses more bits per floating point number. These “higher fidelity” characteristics can be crucial e.g. for psychoacoustic experiments or to generate high quality musical sounds or pieces. Simply put, it is compelling for many scientific and artistic purposes. The didactic potential of the framework is evident when noticed that:

- the integrals and derivatives, ubiquitous in continuous signal processing, are all replaced, in discrete signals, by summations, which are more intuitive and does not require fluency in calculus.
- The equations and concepts are implemented as software which can be easily assembled and inspected.

In fact, this framework was used in a number of contexts, including courses, software implementations and making music [3, 4, 31, 69]. Such detailed analytical descriptions, together with the computational implementations, have not been covered before in the literature, as far as the authors know, such as testified in the literature review (Appendix G of [31]), where books, articles and open software are related to this framework.

The free software license and online availability of the content, with the respective codes and sonic examples, facilitate collaborations and the generation of sub-products in a co-authorship fashion, easing new implementations and development of musical pieces. The scripts can be divided in three groups: implementation of all the equations and topics of music theory; routines for rendering musical pieces that illustrate the concepts; scripts that render the figures of this article and the article itself.

This framework favored the formation of interest groups in topics such as musical creativity and computer music. In particular, the project labMacambira.sourceforge.net groups Brazilian and foreign co-workers in diverse areas that range from digital direct democracy and georeferencing to art and education. This was only possible because of the usefulness of audiovisual abilities in many contexts, in particular because of the knowledge and mastery condensed in the MASS framework.[26]

Future work might include application of these results in artificial intelligence for the generation of appealing artistic materials. Some psychoacoustic effects were detected, which needs validation and should be reported, specially with [30].[27] Other foreseen advances are: a JavaScript version of the toolbox, better hypermedia deliverables of this framework, user guides for different goals (e.g. musical composition, psychophysical experiments, sound synthesis, education), creation of musical pieces and open experiments to be studied with EEG recordings, and further analytical specification of musical elements in the discrete-time representation of sound as feedback is received from the community.

ACKNOWLEDGMENTS

This work was supported by CNPq and FAPESP (Brazil).

REFERENCES

[1] [n. d.]. 3 Trios para obo, flauta e fagote. ([n. d.]).
[2] [n. d.]. Audiolab, a python package to make noise with numpy arrays. ([n. d.]). ”http://www.ar.media.kyoto-u.ac.jp/members/david/softwares/audiolab/”
[3] [n. d.]. Canal Vimeo do Lab Macambira (mais de 700 videos). ([n. d.]). https://vimeo.com/channels/labmacambira
[4] [n. d.]. Dissertation about Music in digital audio: psychophysics description and software toolbox. ([n. d.]). http://labmacambira.git.sourceforge.net/git/gitweb.cgi?p=labmacambira/dissertacao; a=tree; h=refs/heads/msc; hb=refs/heads/msc
[5] [n. d.]. An Introduction To FM (Snd Manual). ([n. d.]). https://ccrma.stanford.edu/software/snd/snd/fm.html
[6] [n. d.]. Listings of equations, figures, tables and sections of the article ’Musical elements in the discrete-time representation of sound’ and of the scripts in the MASS toolbox. ([n. d.]). https://github.com/ttm/dissertacao/
[7] [n. d.]. Página principal do Lab Macambira. ([n. d.]). http://labmacambira.sourceforge.net
[8] [n. d.]. Perception Lecture Notes: Auditory Pathways and Sound Localization. ([n. d.]). http://www.cns.nyu.edu/~david/courses/perception/lecturenotes/localization/localization.html
[9] [n. d.]. MASSA: música e áudio em sequências e séries amostrais. ([n. d.]). http://labmacambira.git.sourceforge.net/git/gitweb.cgi?p=labmacambira/massa; a=blob; f=README.txt;
[10] [n. d.]. Wiki do Lab Macambira. ([n. d.]). http://wiki.nosdigitais.teia.org.br/Lab_Macambira
[11] E. Aldwell, C. Schachter, and A. Cadwallader. 2010. *Harmony & voice leading*. Wadsworth Publishing Company.
[12] V.R. Algazi, R.O. Duda, D.M. Thompson, and C. Avendano. 2001. The cipic hrtf database. In *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*. IEEE, 99–102.

[26]There are more than 700 videos, written documents, original software applications and contributions in well-known external software (such as Firefox, Scilab, LibreOffice, GEM/Puredata, to name just a few) [3, 7, 10]. Some of these efforts are available online [31]. It is evident that all these contributions are a consequence of more that just MASS, but it is also evident to the authors that MASS had a primary role in converging interests and attracting collaborators.
[27]The sonic portraits where sent to a public mailing list and the fifth piece was reported by some individuals to induce a state in which noises from the own tongue, teeth and jaw of the individual echoed for some seconds [**?**].

[13] Pierre Boulez. 1972. *A música hoje.*
[14] R. Bristow-Johnson. 1996. Wavetable synthesis 101, a fundamental perspective. In *Proc. AES Convention*, Vol. 101.
[15] F.J. Budden and FJ Budden. 1972. *The fascination of groups.* Cambridge University Press London.
[16] Francis James Budden. 1972. *The fascination of groups.* Cambridge Univ. Press.
[17] B. carty and V. lazzarini. 2009. 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,, 1–6.
[18] S. Chacon, J.C. Hamano, and S. Pearce. 2009. *Pro Git.* Vol. 288. Apress.
[19] C.I. Cheng and G.H. Wakefield. 2012. Introduction to head-related transfer functions (HRTFfis): Representations of HRTFfis in time, frequency, and space. *Watermark* 1 (2012).
[20] John M. Chowning. 2000. Digital sound synthesis, acoustics and perception: A rich intersection. *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)* (Dec 2000).
[21] B.D. Class, FFT Java, A. Wah, L.F. Crusher, P. Waveshaper, S. Enhancer, and S.F.R.V.T. Matrix. [n. d.]. musicdsp. org source code archive. ([n. d.]).
[22] J. Clough, N. Engebretsen, and J. Kochavi. 1999. Scales, sets, and interval cycles: A taxonomy. *Music Theory Spectrum* (1999), 74–104.
[23] Perry R. Cook. 2002. *Real sound synthesis for interactive applications.* A K Peters, Natick, Massachusetts.
[24] Gustavo Oliveira Alfaix de Assis. [n. d.]. *Em busca do som.* Editora UNESP.
[25] S. Dehaene. 2003. The neural basis of the Weber–Fechner law: A logarithmic mental number line. *Trends in cognitive sciences* 7, 4 (2003), 145–147.
[26] G. Deleuze. 1968. *Difference and Repetition.* Continuum.
[27] R. Duckworth and F. Stedman. 2007. *Tintinnalogia, or, the Art of Ringing.* Echo Library.
[28] Renato Fabbri e Adolfo Maia Jr. 2007. Applications of Group Theory on Granular Synthesis. (2007).
[29] Renato Fabbri e Adolfo Maia Jr. 2008. Applications of Group Theory on Sequencing and Spatialization of Granular Sounds. (2008).
[30] R. Fabbri et al. [n. d.]. Quadros sonoros. ([n. d.]). https://github.com/ttm/dissertacao/blob/master/article/scripts/pecas2.1/quadrosSonoros.py.
[31] R. FABBRI. 2013. *Msica no udio digital : descrio psicofsica e caixa de ferramentas [online].* Ph.D. Dissertation. Instituto de Fsica de So Carlos, Universidade de So Paulo.
[32] J.J. Fux and A. Mann. 1965. *The study of counterpoint from Johann Joseph Fux's Gradus ad Parnassum.* Vol. 277. WW Norton & Company.
[33] Gnter Geiger. 2006. Table lookup oscillators using generic integrated wavetables. *Proc. of the 9th Int. Conference on Digital Audio Effects (DAFx-06)* (September 2006).
[34] J.E. Gramani. 1996. *Rítmica viva: a consciência musical do ritmo.* UNICAMP.
[35] Gérard Grisey and Joshua Fineberg. 2000. Did you say spectral? *Contemporary music review* 19, 3 (2000), 1–3.
[36] H.L. Guidorizzi. 2001. *Um curso de cálculo.* Livros Técnicos e Científicos Editora.
[37] P. Guillaume. 2006. *Music and acoustics: from instrument to computer.* Iste.
[38] H.J. Koellreutter. 1986. Harmonia funcional. (1986).
[39] S.M. Kostka, J.P. Clendinning, R. Ottman, and J. Phillips. 1995. *Tonal Harmony: With an Introduction to Twentieth.* McGraw-Hill.
[40] Osvaldo Lacerda. 1966. *Compndio de Teoria Elementar da Msica* (9.a edifjflo ed.). Ricordi Brasileira.
[41] Fred Lerdahl and Ray Jackendoff. 1983. *A Generative Theory of Tonal Music.* MIT Press.
[42] L. Lessig. 2002. Free culture. *Retrieved February* 5 (2002), 2006.
[43] William LOVELOCK. 1972. *A concise history of music. Reprinted with revised record list.*
[44] Guerino Mazzola. 2012. *The topos of music: geometric logic of concepts, theory, and performance.* Birkhäuser.
[45] Florivaldo Menezes. 2004. *A Acústica Musical em Palavras e Sons.* Ateliê Editorial.
[46] Jan Newmarch. 2017. Sound Codecs and File Formats. In *Linux Sound Programming.* Springer, 11–14.
[47] T.E. Oliphant. 2006. *A Guide to NumPy.* Vol. 1. Trelgol Publishing USA.
[48] A.V. Oppenheim, R.W. Schafer, J.R. Buck, et al. 1989. *Discrete-time signal processing.* Vol. 2. Prentice hall Englewood Cliffs, NJ:.
[49] A.T. Porres and A.S. Pires. 2009. Um External de Aspereza para Puredata & MAX/MSP. In *Proceedings of the 12th Brazilian Symposium on Computer Music.*
[50] TA Porres, J. Manzolli, and F. Furlanete. 2006. Análise de Dissonância Sensorial de Espectros Sonoros. In *Congresso da ANPPOM*, Vol. 16.
[51] E.S. Raymond. 2004. *The art of Unix programming.* Addison-Wesley Professional.
[52] C. Roads. 1996. *The computer music tutorial.* MIT press.
[53] C. Roads. 2004. *Microsound.* MIT press.

[54] Juan G. Roederer. 2008. *The Physics and Psychophysics of Music: An Introduction* (fourth edition ed.). Springer.
[55] G. Van Rossum and F. L. Drake Jr. 1995. *Python tutorial.* Odense Universitet, Institut for Matematik og Datalogi.
[56] O. Sacks. 2008. *Musicophilia: Tales of Music and the Brain, Revised and Expanded Edition.* Knopf Doubleday Publishing Group.
[57] F. Salzer. 1962. *Structural hearing: Tonal coherence in music.* Vol. 1. Dover publications.
[58] Pierre Schaeffer. 2017. *Treatise on Musical Objects: Essays Across Disciplines.* Vol. 20. Univ of California Press.
[59] A. Schoenberg and M. Maluf. 1999. *Harmonia.* Ed. UNESP.
[60] A. Schoenberg and L. Stein. 1963. *Preliminary exercises in counterpoint.* Faber & Faber.
[61] Arnold Schoenberg and Leonard Stein. 1967. *Fundamentals of musical composition.* London: Faber.
[62] Neil JA Sloane et al. 2010. The on-line encyclopedia of integer sequences. (2010). https://oeis.org/
[63] Denis Smalley. 1997. Spectromorphology: explaining sound-shapes. *Organised sound* 2, 2 (1997), 107–126.
[64] S.W. Smith. 2009. The Scientist and Engineerfis Guide to Digital Signal Processing, 1999. (2009).
[65] Julious O. Smith III. 2006. *Physical Audio Signal Processing (for virtual musical instruments and audio effects).* https://ccrma.stanford.edu/~jos/pasp/
[66] Julious O. Smith III. 2012. *Mathematics of the discrete fourier transform (dft) with audio applications* (second ed.). https://ccrma.stanford.edu/~jos/log/FM_Spectra.html
[67] L. Tragtenberg. 2002. *Contraponto: uma arte de compor.* Edusp.
[68] G. Van Rossum and F.L. Drake Jr. 1995. *Python reference manual.* Centrum voor Wiskunde en Informatica.
[69] V. Vieira, G. Lunhani, G.M.C. Rocha Junior, C.M. Luporini, D. Penalva, R. Fabbri, and R. Fabbri. 2017. Vivace: a collaborative live coding language and platform. In *Proceedings of the 16th Brazilian Symposium on Computer Music.*
[70] Anton Webern. 1963. *The Path To The New Music.* Theodore Presser Company.
[71] S.R. Wilkinson, K. Laubach, R.S. Schiff, and J. Eiche. 1988. *Tuning in: Microtonality in Electronic Music: a Basic Guide to Alternate Scales, Temperaments, and Microtuning Using Synthesizers.* H. Leonard Books.
[72] Jos Miguel Wisnik. 1999. *O som e o sentido.* Companhia das Letras.
[73] Iannis Xenakis. 1992. *Formalized music: thought and mathematics in composition.* Number 6. Pendragon Press.
[74] Joaqum Zamacois. 2002. *Curso de formas musicales: Con numerosos ejemplos musicales.* Barcelona : Idea Books.

Received February 2007; revised March 2009; accepted June 2009