% \tableofcontents % enable to generate .toc file used in listings.tex

\section{Introduction}\label{sec:level1}

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~\cite{Roederer}.

Since the speed of sound is $\approx 343.2 m/s$,

such frequency limits correspond 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~\cite{Roederer}.

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~\ref{sec:discNote} and~\ref{sec:internalVar}),

and higher level musical structures (such as tunings, scales and chords, in Section~\ref{sec:notesMusic}).

\begin{figure\*}[!h]

\centering

\includegraphics[width=.7\textwidth]{figures/pcm\_}

\caption{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.}

\label{fig:PCM}

\end{figure\*}

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 imply quality loss (to yield smaller files, ease storage and transfer)

and are called \emph{lossy}~\cite{loss}.

This is the case e.g. of MP3 and Ogg Vorbis.

Non-lossy representations of digital audio, called \emph{lossless} protocols or formats,

on the other hand, assure 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 $\delta\_s$, 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~\ref{fig:PCM} 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~\cite{Openheim} 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~\footnote{There

are well known incidences of such characteristics in ethnic music,

such as in Pygmy music, but western theory assimilated them

only in the last century~\cite{Wisnick}.}.

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~\ref{sec:notesMusic})

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 \massa\ toolbox.

Hopefully, the interested reader or programmer will be able to use

this framework to synthesize music beyond traditional conceptualizations when intended.

% \subsection{Contributions and paper organization}

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~\cite{dissertacao}.

Articles on the topic appear to be lacking, to the best of our knowledge.

Although current music software uses 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:

\begin{enumerate}

\item Present a concise set of mathematical and algorithmic relations between musical basic elements and sequences of PCM audio samples.

\item 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~\ref{cap:conclusao}).

\item Provide a powerful theoretical framework which can be used to synthesize musical pieces and albums.

\item Provide approachability to the developed framework~\footnote{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 \massa\ toolbox, available in an open source Git repository~\cite{gitBook}.

These scripts are written in Python and make use of Numpy,

which performs numerical routines efficiently (e.g. through LAPACK).

Part of the scripts has been transcribed to JavaScript (which favors their use in Web browsers such as Firefox and Chromium) and native Python~\cite{numpy, audiolab, tutpython, python}.

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~\cite{Raymond,Lessig}.}.

\item Provide a didactic presentation of the content, which is highly multidisciplinary,

involving signal processing, music, psychoacoustics and programming.

\end{enumerate}

The reminder of this article is organized as follows:

Section~\ref{sec:discNote} characterizes the basic musical note;

Section~\ref{sec:internalVar} develops internal dynamics of musical notes;

Section~\ref{sec:notesMusic} tackles the organization of musical notes in

higher level musical structures~\cite{Wisnick,Webern,Lerdahl,Cook,Lacerda,Zamacois,Schoenberg,microsound}.

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~\ref{cap:conclusao} is dedicated to final considerations and further work.

\subsection{Additional material}

The Supporting Information document~\cite{SI} holds commented listings of all the equations, figures, tables and sections in this document

and the scripts in the \massa\ toolbox\footnote{The toolbox contains a collection of Python scripts which:

\begin{itemize}

\item implement each of the equations;

\item render music and illustrate the concepts;

\item render each of the figures used in this article.

\end{itemize}

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

The musical pieces are not very traditional which facilitates the understanding of specific techniques

and the extrapolation of the note concept.

\subsection{Synonymy, polysemy and theoretical frames (disclaimer)}\label{sec:discl}

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:

\begin{itemize}

\item 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.

\item The sampling rate (discussed above) is also called the sampling frequency or sample rate.

\item A 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).

\item 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.

\item 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.

\end{itemize}

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.

\section{Characterization of the musical note in discrete-time audio} \label{sec:discNote}\label{sec:notaDisc}

In diverse artistic and theoretical contexts, music is conceived as constituted by fundamental units referred to as notes, ``atoms'' that constitute music itself~\cite{Wisnick, Lovelock, Webern}.

In a cognitive perspective, notes are seen as discernible elements that facilitate and enrich the transmission of information through music~\cite{Roederer, Lacerda}.

Canonically, the basic characteristics of a musical note are duration, loudness, pitch and timbre~\cite{Lacerda}.

All relations described in this section are implemented in the file \texttt{src/sections/1.py}.

The musical pieces \emph{5 sonic portraits} and \emph{reduced-fi} are also available online to corroborate the concepts.

\subsection{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 $\Delta$ seconds can be expressed as a sequence $T^{\Delta}$ with $\Lambda = \lfloor \Delta . f\_s \rfloor$ samples.

That is, the integer part of the multiplication is considered, and an error of at most $\delta\_s$ missing seconds is admitted, which is usually fine for musical purposes. Thus:

\begin{equation}\label{eq:dur}

T^{\Delta}={\{t\_i\}}\_{i=0}^{\lfloor \Delta . f\_s \rfloor -1} = \{t\_i\}\_0^{\Lambda-1}

\end{equation}

\subsection{Loudness}\label{subsec:volume}

Loudness\footnote{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.} is a sensation that depends on reverberation,

spectrum and other characteristics described in Section~\ref{sec:varInternas}.

One can achieve loudness variations through the power of the wave~\cite{Chowning}:

\begin{equation}\label{eq:potencia}

pow(T)=\frac{\sum\_{i=0}^{\Lambda -1} t\_i^2}{\Lambda}

\end{equation}

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:

\begin{equation}\label{decibels}

V\_{dB}=10log\_{10}\frac{pow(T^{'})}{pow(T)}

\end{equation}

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: \emph{pianissimo}, \emph{piano}, \emph{mezzoforte}, \emph{forte} and \emph{fortissimo}. Other useful references are $dB$ values related to double amplitude or power:

\begin{align}

t\_i^{'}=2 t\_i \Rightarrow pow(T^{'})=4 pow(T) \Rightarrow V^{'}\_{dB}=10log\_{10} 4 \approx 6 dB \label{eq:ampVol}\\

t\_i^{'}=\sqrt{2} t\_i \Rightarrow pow(T^{'})=2 pow(T) \Rightarrow V^{'}\_{dB}=10log\_{10} 2 \approx 3 dB\label{eq:potVol}

\end{align}

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

\begin{equation}\label{eq:dobraVol}

\begin{split}

10log\_{10}\frac{pot(T^{'})}{pot(T)} = 10 \quad \Rightarrow \\ \Rightarrow \quad \sum\_{i=0}^{\lfloor \Delta.f\_s \rfloor -1}t^{'2}\_i=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{split}

\end{equation}

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:

\begin{equation}\label{ampDec}

A = 10^{\frac{V\_{dB}}{20}}

\end{equation}

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

\subsection{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 the number of samples per cycle: $\lambda=f\_s . \delta =f\_s/f$.

For didactic reasons, let $f$ divide $f\_s$ and result $\lambda$ integer. If $T^f$ is a sonic sequence with fundamental frequency $f$, then:

\begin{equation}\label{periodicidade}

T^f=\left\{ t\_i^f \right\}=\left\{ t^f\_{i+\lambda} \right\}= \left\{ t^f\_{i+\frac{f\_s}{f}} \right\}

\end{equation}

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.

\subsection{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 \emph{waveform} and defines the harmonic spectrum.

The waveform thus entails a timbre\footnote{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.~\cite{Roederer}}. Sonic spectra with minimum differences can result in timbres with crucial differences and, consequently, distinct timbres can be produced using different spectra~\cite{Roederer}.

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 \emph{simple harmonic motion}).

Let $S^f$ be a sequence whose samples $s\_i^f$ describe a sinusoid with frequency $f$:

\begin{equation}\label{sinusoid}

S^f=\{ s^f\_i \}=\Bigl\{ \sin\bigl(2\pi \frac{i}{\lambda\_f} \bigr) \Bigr\} = \Bigl\{ \sin\bigl(2\pi f \frac{i}{f\_s}\bigr) \Bigr\}

\end{equation}

\noindent 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 presents a succession of harmonic components.

The most basic waveforms are specified in Equations~\ref{sinusoid},~\ref{sawTooth},~\ref{triangular} and~\ref{square}, and are illustrated in Figure~\ref{fig:formasDeOnda}.

These artificial waveforms are traditionally used in music for synthesis and oscillatory control of variables. They are also useful outside musical contexts~\cite{Openheim}.

\begin{figure\*}

\centering

\includegraphics[width=.7\textwidth]{figures/waveForms\_\_}

\caption{Basic musical waveforms: (a) the basic synthetic waveforms; (b) real waveforms.}

\label{fig:formasDeOnda}

\end{figure\*}

The sawtooth presents all components of the harmonic series with decreasing energy of $-6dB/octave$\footnote{In

musical jargon, an ``octave'' means a frequency $f$ and twice such frequency $2f$ or the bandwidth $\{f,2f\}$.}. The sequence of temporal samples can be described as:

\begin{equation}\label{sawTooth}

D^f=\left\{ d^f\_i \right\}=\left\{ 2\frac{i\,\%\lambda\_f}{\lambda\_f} -1 \right\}

\end{equation}

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

\begin{equation}\label{triangular}

T^f=\left\{ t^f\_i \right\}=\left\{1- \left| 2 - 4\frac{i\,\%\lambda\_f}{\lambda\_f} \right| \right\}

\end{equation}

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

\begin{equation}\label{square}

Q^f=\left\{ q^f\_i \right\}= \left\{

\begin{array}{l l}

1 & \quad \text{for } \; \; (i\,\%\lambda\_f) < \lambda\_f /2 \\

-1 & \quad \text{otherwise}\\

\end{array} \right.

\end{equation}

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 make 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~\cite{Roederer}.

Figure~\ref{fig:formasDeOnda} presents the waveforms described in Equations~\ref{sinusoid},~\ref{sawTooth},~\ref{triangular} and~\ref{square} 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~\ref{fig:espectroDeOndas}. 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.

\begin{figure\*}

\centering

\includegraphics[width=.7\textwidth]{figures/waveSpectrum\_}

\caption{Spectra of basic artificial musical waveforms.}

\label{fig:espectroDeOndas}

\end{figure\*}

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~\ref{eq:serieHarmonica}). 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 processes (this separation is done by diverse living organisms by mechanisms similar to the human cochlea~\cite{Roederer}).

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~\cite{Roederer}. 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~\ref{fig:formasDeOnda} 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~\ref{fig:formasDeOnda} can be taken as basis for a sound $T^f$ in the following way:

\begin{equation}\label{sampleandoFormaDeOnda}

T^f=\{ t\_i^f \}=\Bigl\{ r\_{(i\,\%\lambda\_{f})} \Bigr\}

\end{equation}

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~\ref{fig:espectroOboe}, 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~\ref{fig:formasDeOnda}. 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.

\begin{figure\*}

\centering

\includegraphics[width=.7\textwidth]{figures/oboeNaturalSampledSpectrum\_}

\caption{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.}

\label{fig:espectroOboe}

\end{figure\*}

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

\subsection{Spectrum of sampled sound}\label{sec:spectrum}

These sinusoidal components in the discretized sound have some particularities. Considering a signal $T$ and its corresponding Fourier decomposition $\mathcal{F}\langle T\rangle=C=\{c\_i\}\_0^{\Lambda-1}=\left\{ \sum\_{k=0}^{\Lambda-1}t\_ke^{-j2\pi ki/\Lambda} \right\}\_0^{\Lambda-1}$, the recomposition is the sum of the frequency components to yield the temporal samples\footnote{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 $j^2=-1$.}:

\begin{equation}\label{recomposicaoFourier}

\begin{split}

t\_i = & \frac{1}{\Lambda}\sum\_{k=0}^{\Lambda-1}c\_ke^{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{split}

\end{equation}

\noindent 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 $\pi$ 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~\ref{recomposicaoFourier}:

\begin{equation}\label{moduloEfase}

\begin{split}

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{split}

\end{equation}

Equation~\ref{moduloEfase} 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 $a\_k$ specifies the right or left side of the trigonometric circle, which completes the phase domain: $\left[-\frac{\pi}{2},+\frac{\pi}{2}\right] \cup \left[\frac{\pi}{2},\frac{3\pi}{2}\right]\equiv [2\pi]$.

\begin{figure\*}

\centering

\includegraphics[width=.7\textwidth]{figures/amostras2c\_\_\_}

\caption{Oscillation of 2 samples (maximum frequency for any $f\_s$). The first coefficient reflects a constant detachment (called \emph{offset} or \emph{bias}) and the second coefficient specifies the oscillation amplitude.}

\label{fig:amostras2}

\end{figure\*}

Figure~\ref{fig:amostras2} shows two samples and their spectral component. When there are 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 $\{f\_k\}\_0^1=\left\{w\_k\frac{f\_s}{2\pi}\right\}\_0^1=\left\{k\frac{f\_s}{\Lambda=2}\right\}\_0^1=\left\{0,\frac{f\_s}{2}=f\_{\text{max}}\right\}$

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 \emph{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 $f\_{\text{max}}=\frac{f\_s}{2}$, 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 $f=f\_s/3$.

\begin{figure\*}

\centering

\includegraphics[width=.7\textwidth]{figures/amostras3b\_}

\caption{Three fixed samples present only one non-null frequency. $c\_1=c\_2^\*$ and $w\_1 \equiv w\_2$.}

\label{fig:amostras3}

\end{figure\*}

$\Lambda$ real samples $t\_i$ result in $\Lambda$ 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 moduli are equal and phases have opposite signs. Recalling that $f\_k = k\frac{f\_s}{\Lambda}, \; k \in \left\{0, ..., \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: $f\_k=\frac{f\_s}{2} - (f\_k-\frac{f\_s}{2})=f\_s-f\_k=f\_s - k\frac{f\_s}{\Lambda}=(\Lambda-k)\frac{f\_s}{\Lambda} \;\;\;\; \Rightarrow \;\;\;\; f\_k\equiv f\_{\Lambda-k} \; ,\;\; \forall \;\; k<\Lambda$.

The same applies to $w\_k=f\_k.\frac{2\pi}{f\_s}$ and the periodicity $2\pi$: 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 $\Lambda$ samples, the $\Lambda$ frequency components $\{c\_i\}\_0^{\Lambda-1}$ are equivalent in pairs,

except for $f\_0$, and, when $\Lambda$ is even, for $f\_{\Lambda/2}=f\_{\text{max}}=\frac{f\_s}{2}$. Both components are isolated, i.e.\ there is one and only one component at frequency $f\_0$ or $f\_{\Lambda/2}$ (if $\Lambda$ 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 $\tau$ of equivalent coefficient pairs is:

\begin{equation}\label{coefsPareados}

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

\end{equation}

This discussion makes it clear the equivalence in Equations~\ref{equivalenciasFreqs},~\ref{equivalenciasModulos} and~\ref{equivalenciasFases}:

\begin{equation}\label{equivalenciasFreqs}

f\_{k}\equiv f\_{\Lambda-k}\;, \;\; w\_{k}\equiv-w\_{\Lambda-k}\;\;\;, \quad \;\; \forall \quad 1 \leq k \leq \tau

\end{equation}

\begin{figure\*}

\centering

\includegraphics[width=.7\textwidth]{figures/amostras4\_\_\_\_}

\caption{Frequency components for 4 samples.}

\label{fig:amostras4}

\end{figure\*}

Real $T \; \Rightarrow \; a\_k = a\_{\Lambda -k}\;\;$ and $\;\;b\_k = - b\_{\Lambda -k}$, and thus:

\begin{equation}\label{equivalenciasModulos}

\sqrt{a\_k^2 + b\_k^2} = \sqrt{a\_{\Lambda - k}^2 + b\_{\Lambda -k}^2} \;\;, \quad \;\; \forall \quad 1 \leq k \leq \tau \\

\end{equation}

\begin{equation}\label{equivalenciasFases}

tg^{-1}\left(\frac{b\_k}{a\_k}\right)=-tg^{-1}\left(\frac{b\_{\Lambda -k}}{a\_{\Lambda - k}}\right)\;\;,\quad \;\; \forall \quad 1 \leq k \leq \tau

\end{equation}

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

To discuss the general case for components combination in each sample $t\_i$, one can gather relations in Equation~\ref{moduloEfase} for the real signal reconstruction, equivalences of moduli (Equation~\ref{equivalenciasModulos}) and phases (Equation~\ref{equivalenciasFases}), the number of paired coefficients (Equation~\ref{coefsPareados}), and the equivalence of paired frequencies (Equation~\ref{equivalenciasFreqs}):

\begin{equation}\label{eq:reconsCompleta}

\begin{split}

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]

\end{split}

\end{equation}

\noindent with $a\_{\Lambda/2}=0$ if $\Lambda$ odd.

\begin{figure\*}

\centering

\includegraphics[width=.9\textwidth]{figures/amostras4formas\_\_\_}

\caption{Basic wave forms with 4 samples.}

\label{fig:formas4}

\end{figure\*}

\begin{figure\*}[!h!]

\centering

\includegraphics[width=.7\textwidth]{figures/amostras6\_\_}

\caption{Frequency components for 6 samples: 4 sinusoids, one of them is the \emph{bias} with zero frequency.}

\label{fig:amostras6}

\end{figure\*}

With 4 samples it is possible to represent 1 or 2 frequencies in any proportions (i.e. with independence). Figure~\ref{fig:amostras4} 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 $f\_0=0$.

\begin{figure\*}[!h!]

\centering

\includegraphics[width=.9\textwidth]{figures/amostras6formas\_\_\_}

\caption{Basic waveforms with 6 samples: triangular and square waveforms have odd harmonics, with different proportions and phases; the sawtooth has even harmonics.}

\label{fig:formas6}

\end{figure\*}

Figure~\ref{fig:formas4} shows the harmonics for the basic waveforms of Equations~\ref{sinusoid},~\ref{sawTooth},~\ref{triangular} and~\ref{square} 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~\ref{fig:amostras6} shows the sinusoidal decomposition for 6 samples, while Figure~\ref{fig:formas6} 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.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

\subsection{The basic note}\label{notaBasica}

In a nutshell, let $f$ be such that it divides $f\_s$\footnote{As mentioned before, this limitation simplifies the explanation for now and will be overcome in the next section.}. A sequence $T$ of sonic samples separated by $\delta\_a=1/f\_s$ expresses a musical note with a frequency of $f$ Hertz and $\Delta$ seconds of duration if, and only if, it has the periodicity $\lambda\_f=f\_s/f$ and size $\Lambda=\lfloor f\_s . \Delta \rfloor$:

\begin{equation}\label{eq:notaBasica}

T^{f,\; \Delta}=\{t\_{i \, \% \lambda\_f} \}\_0^{\Lambda-1}= \left \{t^f\_{i \; \% \left( \frac{f\_s}{f} \right) } \right \}\_0^{\Lambda-1}

\end{equation}

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~\ref{sinusoid},~\ref{sawTooth},~\ref{triangular} and~\ref{square} and let $R\_i^f$ be a sampled real waveform):

\begin{equation}\label{periodoUnico}

L^{f , \delta\_f } = \left\{ l\_i^f \right\}\_0^{\delta\_f . f\_s -1}=\left\{ l\_i^f \right\}\_0^{\lambda\_f-1}

\end{equation}

Therefore, the sequence $T\_i$ will consist in a note of duration $\Delta$ and frequency $f$ if:

\begin{equation}\label{eq:notaBasicaTimbre}

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}

\end{equation}

\subsection{Spatialization: localization and reverberation}\label{subsec:spac}

A sound is always produced within the three dimensional physical space.

The consideration of this fact is the subject of spatialization\footnote{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 second item are also tackled.}.

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~\cite{floEsp}.

\subsubsection{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, especially the chest, head and ears~\cite{Roederer, hrtf, Heeger}.

\begin{figure}[h!]

\centering

\includegraphics[width=.5\textwidth]{figures/espacializacao\_\_\_}

\caption{Detection of sound source localization: schema used to calculate Interaural Time Difference (ITD) and Interaural Intensity Difference (IID).}

\label{fig:spac}

\end{figure}

Considering only the direct incidences in each ear, the equations are quite simple. An object placed at $(x,y)$, as in Figure~\ref{fig:spac}, is distant of each ear by:

\begin{equation}\label{eq:distOuvidos}

\begin{split}

d & =\sqrt{\left (x-\frac{\zeta}{2} \right )^2+y^2} \\

d' & =\sqrt{\left (x+\frac{\zeta}{2} \right )^2 + y^2}

\end{split}

\end{equation}

\noindent where $\zeta$ 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:

\begin{equation}\label{eq:dti}

ITD=\frac{d'-d}{v\_{sound\;at\;air}\approx 343.2 }\quad \text{seconds}

\end{equation}

\noindent and in the Interaural Intensity Difference:

\begin{equation}\label{eq:dii}

IID=20\log\_{10}\left (\frac{d}{d'}\right) \quad decibels

\end{equation}

$IID\_a=\frac{d}{d'}$ 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~\cite{Heeger}.

With $\Lambda\_{ITD}=\lfloor ITD . f\_s \rfloor$:

\begin{equation}\label{eq:locImpl}

\begin{split}

\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} & =\left\{IID\_a . t\_i\right\}\_0^{\Lambda-1} \\

\left\{t\_i'\right\}\_0^{\Lambda\_{ITD}-1} & = 0

\end{split}

\end{equation}

\noindent 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 $\theta$ given by:

\begin{equation}\label{eq:angulo}

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

\end{equation}

\noindent with $x,y$ as presented in Figure~\ref{fig:spac}.

Furthermore, there are problems when $\theta$ 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~\footnote{The cone itself is not shown in Figure~\ref{fig:spac} 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~\cite{hrtf}.}.

On those points the inference of the azimuthal angle depends especially on the filtering of the frequencies by the head and torso~\cite{Heeger,hrtf}.

Figure~\ref{fig:spac} 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 $\approx 0.7ms$~\cite{floEsp}.

The complete localization, including height and distance of sound source, is given by the Head Related Transfer Function (HRTF)~\cite{hrtf}. There are well known open databases of HRTF, such as CIPIC, and it is possible to apply those transfer functions in a sonic signal by convolution (see Equation~\ref{eq:conv})~\cite{CIPIC}. Each human body has its filtering and there are techniques to generate HRTFs to be universally used~\cite{lazaSPA}.

\subsubsection{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 on each independent reflection and that explore valuable cues to the auditory system. In fact, a reverberation can be modeled with a set of two temporal and two spectral regions~\cite{JOSPhy}:

\begin{itemize}

\item First period: 'first reflections' are more intense and scattered.

\item Second period: 'late reverberation' is practically a dense succession of indistinct delays with exponential decay and statistical occurrences.

\item First band: the bass has some resonance bandwidths relatively spaced.

\item Second band: mid and treble have a progressive decay and smooth statistical fluctuations.

\end{itemize}

Smith III states that usual concert rooms have total reverberation time of $\approx 1.9$ 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~\ref{eq:dti} and~\ref{eq:dii}. Assuming that each one of the first reflections, before reaching the ear, will propagate at least $3-30m$, depending on the room dimensions, the separation between 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~\ref{subsec:mus2}, particularly in the paragraphs about reverberation.

\subsection{Musical usages}\label{subsec:basMus}

The sum of the respective $ith$ samples of $N$ sequences with the same size $\Lambda$

results in the overlapped spectral contents of each sequence, in a process called mixing:

\begin{equation}\label{eq:mixagem}

\{t\_i\}\_0^{\Lambda-1}=\left \{ \sum\_{k=0}^{N-1}t\_{k,i} \right \}\_0^{\Lambda-1}

\end{equation}

\begin{figure}[h!]

{\centering

\includegraphics[width=.5\columnwidth]{figures/mixagem\_\_}}

\caption{Mixing of three sound sequences. The amplitudes are directly summed sample-by-sample.}

\label{fig:mixagem}

\end{figure}

Figure~\ref{fig:mixagem} 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~\cite{Roederer}.

This process of mixing musical notes results in musical harmony and counterpoint which guide

subjective aspects of music appreciation~\cite{Harmonia} and are addressed in Section~\ref{sec:notesMusic}.

\begin{figure}[h!]

{ \centering

\includegraphics[width=.5\columnwidth]{figures/concatenacao\_\_}}

\caption{Concatenation of three sound sequences by temporal overlap of their samples.}

\label{fig:concatenacao}

\end{figure}

Sequences can be concatenated in time. If the sequences $\{t\_{k,i}\}\_0^{\Lambda\_k-1}$ of size $\Lambda\_k$ represent $k$ musical notes, their concatenation in a unique sequence $T$ results in a simple melodic sequence:

\begin{equation}\label{eq:concatenacao}

\begin{split}

\{t\_i\}\_0^{\sum\Delta\_k-1}= \{t\_{l,i}\}\_0^{\sum\Delta\_k-1}, \;\; l\text{ smallest integer}: \;\; \sum\_{j=0}^{l}\Lambda\_j > i

\end{split}

\end{equation}

This mechanism is demonstrated in Figure~\ref{fig:concatenacao} with the same sequences of Figure~\ref{fig:mixagem}.

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 \emph{reduced-fi} explores the temporal juxtaposition of notes, resulting in a homophonic piece.

The vertical principle is demonstrated in the \emph{sonic portraits} which are static sounds with peculiar spectrum.\cite{MASSA}

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

\section{Variation in the basic note}\label{sec:internalVar}\label{sec:varInternas}

The basic digital music note defined in Section~\ref{sec:notaDisc} 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~\cite{Chowning}.

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~\cite{Roederer}.

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 \emph{Transit para meter}, \emph{Vibrates and shakes}, \emph{Tremolos, vibratos and the frequency}, \emph{Little train of impulsive hillbillies}, \emph{Noisy band}, \emph{Bela rugosi}, \emph{Children's chorus}, \emph{ADa and SaRa} were made to validate and illustrate concepts of this section.~\cite{MASSA}

\subsection{Lookup table}\label{subsec:lookup}

The \emph{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 of a single wave period to synthesize sounds in the whole audible spectrum, with any waveform.

\begin{figure\*}

\centering

\includegraphics[width=.7\textwidth]{figures/lut\_}

\caption{Search in the \emph{lookup table} to synthesize sounds at different frequencies using a unique waveform with high resolution.

Each \emph{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.}

\label{fig:lut}

\end{figure\*}

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 $\Delta$ can be obtained by means of $\widetilde{L}$:

\begin{equation}\label{eq:lut}

\begin{split}

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

\end{split}

\end{equation}

In other words, with the right LUT indexes ($\gamma\_i\%\widetilde{\Lambda}$) it is possible to synthesize sounds at an arbitrary frequency. Figure~\ref{fig:lut} 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 $\gamma\_i$ introduces noise which decreases as $\widetilde{\Lambda}$ increases.

In order to use this calculation in sound synthesis, with $f\_s=44.1 kHz$, 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~\cite{Geiger}.

The calculation of $\gamma\_i$ 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 $i f \frac{\widetilde{\Lambda}}{f\_s}$ 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 \emph{Wavetable Synthesis}, which consists of many LUTs used together to generate a quasi-periodic musical note~\cite{Cook,Wavetable}.

\subsection{Incremental variations of frequency and intensity}\label{subsec:vars}

As stated by the Weber and Fechner~\cite{Weber-Fechner} 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~\ref{subsec:tvaf}), 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 $f\_{\Lambda -1}$:

\begin{equation}\label{freqLinear}

F=\{f\_i\}\_0^{\Lambda-1}=\left\{f\_0 + (f\_{\Lambda-1}-f\_0)\frac{i}{\Lambda-1} \right\}\_0^{\Lambda-1}

\end{equation}

\begin{equation}\label{indiceLinear}

\begin{split}

\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{split}

\end{equation}

\begin{equation}\label{serieAmostralLin}

\left\{t\_i^{\;\overline{f\_0,\, f\_{\Lambda-1}}}\right\}\_0^{\Lambda-1}=\left\{\,\widetilde{l}\_{\gamma\_i \% \widetilde{\Lambda}}\,\right\}\_0^{\Lambda-1}

\end{equation}

\noindent 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~\ref{freqLinear},~\ref{indiceLinear} and~\ref{serieAmostralLin} are related with the linear progression of the frequency. As stated above, the frequency progression \emph{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:

\begin{equation}\label{freqExponencial}

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}

\end{equation}

\begin{equation}\label{indiceExponencial}

\begin{split}

\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{split}

\end{equation}

\begin{equation}\label{serieAmostralLog}

\left\{t\_i^{\;\overline{f\_0,\,f\_{\Lambda-1}}}\right\}\_0^{\Lambda-1}=\left\{\,\widetilde{l}\_{\gamma\_i \% \widetilde{\Lambda}}\,\right\}\_0^{\Lambda-1}

\end{equation}

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\footnote{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~\ref{subsec:volume}). The increase of energy without changing amplitude ambit is the \emph{sound compression}, quite popular nowadays for music production~\cite{guillaume}.}. It is sufficient to multiply the original sequence element-wise by the sequence $a\_{\Lambda-1}^{\left( \frac{i}{\Lambda-1} \right )^\alpha}$, where $\alpha$ 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:

\begin{equation}\label{seqAmp}

\begin{split}

\{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 \{ \left ( {a\_{\Lambda-1}} \right )^{\left ( \frac{i}{\Lambda-1} \right )^\alpha} \right \}\_0^{\Lambda-1}

\end{split}

\end{equation}

where $a\_0=1$. And, therefore:

\begin{equation}\label{transAmp}

\begin{split}

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{split}

\end{equation}

\begin{figure}[h!]

\centering

\includegraphics[width=.7\columnwidth]{figures/transicao\_}

\caption{Intensity transitions for different values of $\alpha$ (see Equations~\ref{seqAmp} and~\ref{transAmp}).}

\label{fig:transicao}

\end{figure}

It is often convenient to have $a\_0=1$ to start a new sequence with the original amplitude and then progressively change it.

If $\alpha=1$, the amplitude variation follows the geometric progression that implies the linear variation of loudness. Figure~\ref{fig:transicao} depicts transitions for different values of $\alpha$, a gain of $\approx 6dB$ as exposed by Equation~\ref{eq:ampVol}.

Special attention should be dedicated while considering $a=0$.

In Equation~\ref{seqAmp}, $a\_0=0$ 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 \emph{fade in} ($a\_0=10^{-4}$) or for a \emph{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~\cite{Cook}.

For linear amplification -- but not linear perception -- it is sufficient to use an appropriate sequence $\{a\_i\}$:

\begin{equation}\label{seqAmpLin}

a\_i=a\_0 + (a\_{\Lambda-1}-a\_0)\frac{i}{\Lambda-1}

\end{equation}

Here, the conversion between decibels and amplitude is convenient. Given Equations~\ref{ampDec} and~\ref{transAmp}, a transition of $V\_{dB}$ decibels is:

\begin{equation}\label{seqAmpDB}

T^{'}=\left\{ t\_i 10^{\frac{V\_{dB}}{20}\left( \frac{i}{\Lambda-1} \right)^\alpha} \right\}\_0^{\Lambda-1}

\end{equation}

\noindent for the general case of amplitude variations following a geometric progression. The greater the value of $\alpha$, the smoother the sound introduction and more intense its end. $\alpha>1$ results in loudness transitions commonly called \emph{slow fade}, while $\alpha<1$ results in \emph{fast fade}~\cite{guillaume}.

The linear transitions will be used for AM and FM synthesis, while logarithmic transitions are proper tremolos and vibratos, as developed in Section~\ref{subsec:tvaf}. A non-oscillatory exploration of these variations is in the music piece \emph{Transit para meter}~\cite{MASSA}.

\subsection{Application of digital filters}\label{subsec:filtros}

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\footnote{The

design and implementation of filters is a field of recognized complexity,

with dedicated literature and software~\cite{Openheim,smith}.}. Filtering can be part of the sound synthesis or made subsequently as part of processes commonly referred to as ``acoustic/sound treatment''.

\subsubsection{Convolution and finite impulse response (FIR) filters}\label{subsec:conv}

\begin{figure\*}

\centering

\includegraphics[width=\textwidth]{figures/convolucao\_}

\caption{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.}

\label{fig:conv}

\end{figure\*}

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\footnote{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)~\cite{Openheim}.}. For the purposes of this work, convolution $T\*H$ of $T$ with $H$ is defined as:

\begin{equation}\label{eq:conv}

\begin{split}

\left\{t\_i'\right\}\_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{split}

\end{equation}

\noindent 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 $\Lambda\_h$ 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~\ref{fig:conv}, 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 $\Lambda\_{t'}=\Lambda\_t+\Lambda\_h -1$.

With this procedure it is possible to achieve reverberators, equalizers, \emph{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~\cite{Cook}. 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\footnote{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~\ref{sec:spectrum}.}.

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 \emph{delay lines} by means of convolution with an impulse response that has an impulse for each reincidence of the sound. Figure~\ref{fig:delays} 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.

\begin{figure\*}

\centering

\includegraphics[width=\textwidth]{figures/delays\_}

\caption{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.}

\label{fig:delays}

\end{figure\*}

\subsubsection{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:

\begin{equation}\label{eq:diferencas}

t\_i' = \frac{1}{b\_0}\left ( \sum\_{j=0}^Ja\_j . t\_{i-j} + \sum\_{k=1}^Kb\_k . t\_{i-k}' \right )

\end{equation}

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~\ref{eq:diferencas} 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 responses are illustrated in Figure~\ref{fig:iir}.

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'\footnote{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: $\mathbb{H}(z)=\frac{a\_0+a\_1.z^{-1}+a\_2.x^{-2}}{1- b\_1.z^{-1} -b\_2 . z^{-2}}$.}

or a Chebichev filter\footnote{Butterworth and Elliptical filters can be considered as special cases of Chebichev filters~\cite{Openheim,smith}.}.

Both alternatives are explored by~\cite{JOSFM,smith}, and by the collection of filters maintained by the \emph{Music-DSP} community of the Columbia University~\cite{music-dsp,Openheim}.

\begin{figure\*}

\centering

\includegraphics[width=\textwidth]{figures/iir\_\_}

\caption{Moduli for the frequency response (a), (b), (c) and (d) for IIR filters of Equations~\ref{eq:passa-baixas},~\ref{eq:passa-altas},~\ref{eq:passa-banda} and~\ref{eq:rejeita-banda} respectively, considering different cutoff frequencies, center frequencies and bandwidths.}

\label{fig:iir}

\end{figure\*}

\begin{enumerate}

\item Low-pass with a simple pole and modulus of the frequency response in the upper left corner of Figure~\ref{fig:iir}. 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{equation}\label{eq:passa-baixas}

\begin{split}

a\_0 & = 1-b\_1 \\

b\_1 & = e^{-2\pi f\_c}

\end{split}

\end{equation}

\item High-pass filter with a simple pole and modulus of the frequency responses at the upper right corner of Figure~\ref{fig:iir}.

The general equation with cutoff frequency $f\_c \in (0,\frac{1}{2})$ has coefficients:

\begin{equation}\label{eq:passa-altas}

\begin{split}

a\_0 & = \frac{b\_1+1}{2} \\

a\_1 & = -\frac{b\_1+1}{2} \\

b\_1 & = e^{-2\pi f\_c}

\end{split}

\end{equation}

\item 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 $\approx 0.707$ of the amplitude, i.e. an attenuation of $3dB$. The auxiliary variables $K$ and $R$ are defined as:

\begin{equation}\label{eq:varAux}

\begin{split}

R & = 1 - 3bw \\

K & = \frac{1-2R\cos(2\pi f\_c) + R^2}{2 - 2 \cos (2 \pi f\_c)}

\end{split}

\end{equation}

The band-pass filter in the lower left corner of Figure~\ref{fig:iir} has the following coefficients:

\begin{equation}\label{eq:passa-banda}

\begin{split}

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{split}

\end{equation}

The coefficients of band-reject filter, depicted in the lower right of Figure~\ref{fig:iir}, are:

\begin{equation}\label{eq:rejeita-banda}

\begin{split}

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{split}

\end{equation}

\end{enumerate}

\subsection{Noise}\label{subsec:ruidos}

Sounds without an easily recognizable pitch are generally called noise~\cite{Lacerda}. 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 do not exhibit an unequivocal pitch and their sounds are generally regarded as noise~\cite{Roederer}. In electronic music, including electro-acoustic and dance genres, noise has diverse uses and frequently characterizes the music style~\cite{Cook}.

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~\cite{Cook}. 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.

\begin{figure\*}

\hspace\*{-.75cm}

\includegraphics[width=1.\textwidth]{figures/ruidos\_}

\caption{Colors of noise generated by Equations~\ref{eq:branco},~\ref{eq:rosa},~\ref{eq:marrom},~\ref{eq:azul} and~\ref{eq:violeta}: spectrum and example waveforms.}

\label{fig:ruidos}

\end{figure\*}

Some noises with static spectrum are listed below. They are called \emph{colored noises} since they are associated with colors for many reasons.

Figure~\ref{fig:ruidos} 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.

\begin{itemize}

\item 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{equation}\label{eq:branco}

\begin{split}

c\_0 & =0 \;\text{,\quad 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 \; \text{, \; if $\Lambda$ even}\\

c\_i & = c\_{\Lambda - i}^\*\;,\;\; \text{for}\; i \; > \; \frac{\Lambda}{2}

\end{split}

\end{equation}

The exponential $e^{j.x}$ is a way to obtain unitary modulus and random phase for the value of $c\_i$. In addition, $c\_{\Lambda/2}$ is always real (as discussed in Section~\ref{sec:spectrum}).

\item 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~\cite{Roederer}.

%\begin{split}

\begin{equation}\label{eq:rosa}

\begin{split}

f\_{\text{min}} & \approx 15 Hz \\

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\_{\text{min}}} \right )} \\

c\_i & =0\;,\;\; \forall \; i \; : f\_i<f\_{\text{min}} \\

c\_i & =e^{j.x} \alpha\_i\;, \; x \; \text{random} \; \in \; [0,2\pi]\;,\;\; \forall \; i \; : f\_{\text{min}} \le f\_i < f\_{\lceil \Lambda/2-1 \rceil} \\

c\_{\Lambda/2} & = \alpha\_{\Lambda/2}\;, \; \text{if $\Lambda$ even} \\

c\_i & = c\_{\Lambda - i}^\*\;,\;\; \text{for}\; i \; > \; \Lambda/2 \qquad \qquad

\end{split}

\end{equation}

%\end{split}

The minimum frequency $f\_{\text{min}}$ 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 $\alpha\_i$.

\item The brown noise received this name after Robert Brown, who described the Brownian movement\footnote{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~\cite{Cook,guillaume}.}. What characterizes a brown noise is the decrease of $6dB$ per octave, with $\alpha\_i$ in Equations~\ref{eq:rosa} being:

\begin{equation}\label{eq:marrom}

\alpha\_i=(10^{-\frac{6}{20}})^{\log \_2 \left( \frac{f\_i}{f\_{\text{min}}} \right )}

\end{equation}

\item In the blue noise there is a gain of $3dB$ per octave in a band limited by the minimum frequency $f\_{\text{min}}$ and the maximum frequency $f\_{\text{max}}$.

Therefore, using Equations~\ref{eq:rosa} again:

\begin{equation}\label{eq:azul}

\begin{split}

\alpha\_i & = (10^{\frac{3}{20}})^{\log \_2 \left ( \frac{f\_i}{f\_{\text{min}}} \right )} \\

c\_i & =0\;,\;\; \forall \; i \; : f\_i<f\_{\text{min}} \;\; \text{or} \;\; f\_i>f\_{\text{max}} \\

\end{split}

\end{equation}

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

\begin{equation}\label{eq:violeta}

\alpha\_i = (10^{\frac{6}{20}})^{\log \_2 \left ( \frac{f\_i}{f\_{\text{min}}} \right )}

\end{equation}

\item The black noise has higher losses than $6dB$ per octave:

\begin{equation}\label{eq:preto}

\alpha\_i=(10^{-\frac{\beta}{20}})^{\log \_2 \left( \frac{f\_i}{f\_{\text{min}}} \right )}\;\;, \quad \beta > 6

\end{equation}

\item 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 $\alpha\_i$. An implementation of ISO 226, which is the last revision of these curves, is in the file \texttt{src/aux/iso226.py}~\cite{MASSA}.

\end{itemize}

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~\cite{Cook}.

\subsection{Tremolo and vibrato, AM and FM}\label{subsec:tvaf}

Vibrato is a periodic variation in pitch (frequency) and tremolo is a periodic variation in loudness (intensity)\footnote{The

jargon may be different in other contexts as stated in Section~\ref{sec:discl}~\cite{Lacerda,Harmonia}.}.

A vibrato with frequency $f'$ can be achieved by:

\begin{equation}\label{vbrGamma}

\gamma\_i'=\left \lfloor i f' \frac{\widetilde{\Lambda}\_M}{f\_s} \right \rfloor

\end{equation}

\begin{equation}\label{vbrAux}

t\_i'=\widetilde{m}\_{\gamma\_i' \;\% \widetilde{\Lambda}\_M}

\end{equation}

\begin{equation}\label{vbrF}

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

\end{equation}

\begin{equation}\label{vbrGamma2}

\begin{split}

\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{split}

\end{equation}

\begin{equation}\label{vbrT}

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}

\end{equation}

\begin{figure}[h!]

\centering

\includegraphics[width=\columnwidth]{figures/vibrato\_}

\caption{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.}

\label{fig:vibrato}

\end{figure}

It is important to 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 $\gamma\_i'$ yield the sequence $\{t\_i'\}$

which is the oscillation pattern in the frequency while table $\widetilde{L}$ and the indices $\gamma\_i$

yield $\{t\_i\}$ which is the sound itself.

Variables are useful to quantify the vibrato amplitude (or depth):

\begin{itemize}

\item let $\mu$ be frequency difference (in $Hz$) between the main frequency and the maximum frequency, and

\item let $\nu$ be the the number of semitones (or half steps) added and subtracted by the oscillation ($2\nu$ is the number of semitones between the upper and lower peaks of frequency in the sound).

\end{itemize}

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~\ref{fig:vibrato} is the spectrogram of an artificial vibrato for a note with $1000Hz$ (between a \emph{B} and a \emph{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\footnote{The pitch deviation is called 'vibrato depth' and is generally given in semitones or cents (one cent $= \frac{1}{100}$ of a semitone).}.

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: $f'$, $\gamma\_i'$ and $t\_i'$ remains the same.

The amplitude sequence to be multiplied by the original sequence $t\_i$ is:

\begin{equation}\label{trA}

a\_i=10^{\frac{V\_{dB}}{20}t\_i' } = a\_{\text{max}}^{t\_i'}

\end{equation}

\noindent and, finally:

\begin{equation}\label{trT}

\begin{split}

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\_{\text{max}}^{t\_i'} \right\}\_0^{\Lambda-1}

\end{split}

\end{equation}

\noindent where $V\_{dB}$ is the oscillation depth in decibels of the tremolo and $a\_{\text{max}}=10^{\frac{V\_{dB}}{20}}$ 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~\ref{fig:tremolo} 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 $\text{duration} = \frac{i\_{\text{max}}=82000}{f\_s}= 2s \; \Rightarrow \; \frac{3\text{oscillations}}{2s}=1.5$ oscillations per second.

The music piece \emph{Vibrates and shakes} explores these possibilities given by tremolos and vibratos

used simultaneously and independently, with different frequencies $f'$, depths ($\nu$ 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.~\cite{MASSA}

\begin{figure\*}

\centering

\includegraphics[width=\textwidth]{figures/tremolo\_}

\caption{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 $f\_s=44.1kHz$).}

\label{fig:tremolo}

\end{figure\*}

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 \emph{Dubstep} genre. Roughness is also generated by spectral content that produces beats\footnote{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~\cite{Porres,porres2009}.}. The sequence \emph{Bela Rugosi} explores this roughness threshold with concomitant tremolos and vibratos at the same voice, with different intensities and waveforms.~\cite{MASSA}

As the frequency increases further, these oscillations no longer remain noticeable individually. In this case, the oscillations are audible as pitch. Then, $f'$, $\mu$ and the waveform together change the audible spectrum of original sound $T$ in different ways for both tremolos and vibratos. They are called AM (\emph{Amplitude Modulation}) and FM (\emph{Frequency Modulation}) synthesis,

respectively. These techniques are well known, with applications in

synthesizers like \emph{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 $\mu$

in a sinusoidal sound $T$ with frequency $f$ generates bands centered around $f$ and far from each other by $f'$:

\begin{equation}\label{eq:fmEsp}

\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}

\end{equation}

\noindent where

\begin{equation}\label{eq:Bessel}

J\_k(\mu) = \frac{2}{\pi} \int\_0^{\frac{\pi}{2}}\left [ cos \left (\overline{k}\;\frac{\pi}{2} + \mu . \sin w \right ) . cos \left ( \overline{k}\;\frac{\pi}{2} + k . w \right ) \right ] dw \;,\; \overline{k} = k \% 2 \;,\; k \in \mathbb{N}

\end{equation}

\noindent is the Bessel function~\cite{BesselCCRMA,JOSFM} 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~\ref{freqLinear}.

The result of using Equations~\ref{vbrF} for FM is described in the Appendix D of~\cite{dissertacao}, 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~\ref{eq:fmEsp}, which yield less strident and less noisy sounds.

For the amplitude modulation (AM):

\begin{equation}\label{eq:specAM}

\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}

\end{equation}

% \newcommand{\OneColEqu}[1]{%

% \end{multicols}%

% \begin{twocolequfloat}%

% \begin{equation}

% \{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{equation}

% \end{twocolequfloat}%

% \begin{multicols}{2}%

% }

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~\cite{dissertacao}.

In FM and AM jargon,

the sequence $T$ is called `carrier' and is modulated by

$T'$, called 'modulator'; $\mu$ 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 $\{t\_i'\}$:

\begin{equation}\label{fmGammaAux}

\gamma\_i'=\left \lfloor i f' \frac{\widetilde{\Lambda}\_M}{f\_s} \right \rfloor

\end{equation}

\begin{equation}\label{fmAux}

t\_i'=\widetilde{m}\_{\gamma\_i' \;\% \widetilde{\Lambda}\_M}

\end{equation}

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

\begin{equation}\label{fmF}

f\_i=f + \mu . t\_i'

\end{equation}

\begin{equation}\label{fmGamma}

\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

\end{equation}

\begin{equation}\label{fmT}

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}

\end{equation}

\noindent 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:

\begin{equation}\label{amA}

a\_i=1 + \alpha . t\_i'

\end{equation}

\begin{equation}\label{amT}

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}

\end{equation}

\subsection{Musical usages}\label{subsec:mus2}

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.

\subsubsection{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~\ref{vbrGamma},~\ref{vbrF} and~\ref{trA}, it is possible to set:

\begin{equation}\label{eq:vinculos}

\begin{split}

f^{vbr} = f^{tr} & = func\_a(f) \\

\nu & = func\_b(f) \\

V\_{dB} & = func\_c(f)

\end{split}

\end{equation}

\noindent with both $f^{vbr}$ and $f^{tr}$ as $f'$ in the equations of Section~\ref{subsec:tvaf}.

$\nu$ 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 \emph{Tremolos, vibratos and the

frequency} explores such bonds and exhibits variations in the oscillation waveform with the purpose of building a \emph{musical

language} (details in the next section).~\cite{MASSA}

\subsubsection{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~\ref{subsec:conv}.

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, yield 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~\cite{Gramani}. 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 \emph{Little train of impulsive hillbillies}.

The piece also encompasses

the creation of `sound amalgams' e.g. by granular synthesis (see Figure~\ref{fig:pulsoSubAgl}).

This piece is a link to the contents of next section because of the explicit organization of notes in higher level structures.~\cite{MASSA}.

\subsubsection{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:

\begin{equation}\label{eq:fDoppler}

f=\left(\frac{s\_{sound}+s\_r}{s\_{sound}+s\_s}\right)f\_0

\end{equation}

With both frequencies $f$ and $f\_0$, and the IID and ITD from the new source

position (see Section~\ref{subsec:spac}), 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~\ref{subsec:volume} for potency to amplitude conversion). Thus:

\begin{equation}\label{eq:aDoppler}

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}

\end{equation}

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:

\begin{equation}\label{eq:ffDoppler}

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}

\end{equation}

There is an implementation of the Doppler effect in \texttt{src/sections/3.py}~\cite{MASSA}.

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~\cite{floEsp}.

\subsubsection{Filters and noises (Sections~\ref{subsec:ruidos} and~\ref{subsec:filtros})}

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~\cite{Roederer}. Chains of these filters can perform complex and more accurate filtering routines. The music piece \emph{Noisy band} explores filters and many kinds and noise synthesis.~\cite{MASSA}

A sound can be altered through different filtering and then mixed to create an effect known as \emph{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~\ref{eq:mixagem}).

The music piece \emph{Child chorus} implements a chorus in many ways with different sounds.~\cite{MASSA}

\subsubsection{Reverberation}\label{subsubsec:reverb}

Using the same terms of Section~\ref{subsec:spac}, 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 \emph{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 $\Delta\_1$ as the duration of the first section and $\Delta\_R$ 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~\ref{subsec:spac}, the samples $R^1$ of the impulse response of the first period can be described as:

\begin{equation}\label{eq:p1rev}

\begin{split}

R^1 =\left\{r\_i^1\right\}\_0^{\Lambda\_1-1}\;,\; \text{where}\;\;

r\_i^1 =\left\{

\begin{array}{l l}

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{array} \right.

\end{split}

\end{equation}

\noindent 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:

\begin{equation}\label{eq:p2rev}

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}

\end{equation}

The reverberation impulse response is then:

\begin{equation}\label{eq:rev}

R=\left\{r\_i\right\}\_0^{\Lambda\_R-1}\;,\; \text{where} \; r\_i=\left\{

\begin{array}{l l}

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{array} \right.

\end{equation}

A sound with simulated reverberation can be achieved by a simple convolution of $R$ with the sound sequence $T$, as described in Section~\ref{subsec:filtros}.

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 \textit{clich\'{e}}) to surprise and attract the listener. An implementation of the reverb recipe described here is in the file \texttt{src/sections/3.py}.~\cite{MASSA}.

\subsubsection{ADSR envelopes}

The variation of loudness along the duration of sound is crucial to our timbre perception. The intensity envelope known as ADSR (\emph{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~\cite{ADSR}. 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.

\begin{figure}[htp!]

\centering

\includegraphics[width=\columnwidth]{figures/adsr\_}

\caption{An ADSR envelope (\emph{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.}

\label{fig:adsr}

\end{figure}

The ADSR envelope with durations $\Delta\_A$, $\Delta\_D$ and $\Delta\_R$, with total duration $\Delta$ 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{equation}\label{eq:adsr}

\begin{split}

\{a\_i\}\_0^{\Lambda\_A-1} = & \left\{\xi\left(\frac{1}{\xi}\right)^{\frac{i}{\Lambda\_A-1}}\right\}\_0^{\Lambda\_A-1} \quad \text{ or }\\ = & \left\{\frac{i}{\Lambda\_A-1}\right\}\_0^{\Lambda\_A}\\

\{a\_i\}\_{\Lambda\_A}^{\Lambda\_A+\Lambda\_D-1} = & \left\{a\_S^{\frac{i-\Lambda\_A}{\Lambda\_D-1}} \right\}\_{\Lambda\_A}^{\Lambda\_A+\Lambda\_D-1} \quad \text{ or } \\ = & \left\{1-(1-a\_S)\frac{i-\Lambda\_A}{\Lambda\_D-1}\right\}\_{\Lambda\_A}^{\Lambda\_A+\Lambda\_D-1}\\

\{ a\_i \}\_{\Lambda\_A+\Lambda\_D}^{\Lambda-\Lambda\_R-1} = & \left\{ a\_S \right\}\_{\Lambda\_A+\Lambda\_D}^{\Lambda-\Lambda\_R-1} \\

\{ a\_i \}\_{\Lambda-\Lambda\_R}^{\Lambda-1} = & \left\{ a\_S\left(\frac{\xi}{a\_S} \right)^{\frac{i-(\Lambda-\Lambda\_R)}{\Lambda\_R-1}} \right\}\_{\Lambda-\Lambda\_R}^{\Lambda-1} \quad \text{ or } \\ = & \left\{ a\_S - a\_S\frac{i+\Lambda\_R-\Lambda}{\Lambda\_R-1}\right\}\_{\Lambda-\Lambda\_R}^{\Lambda-1}

\end{split}

\end{equation}

\noindent with $\Lambda\_X=\lfloor \Delta\_X . f\_s \rfloor\;\;\forall\;\; X \; \in

(A,D,R\;)$ and $\xi$ being a small value that provides a satisfactory \emph{fade in} and \emph{fade out}, e.g.\ $\xi=10^{\frac{-80}{20}}=10^{-4}\;$. The lower the $\xi$, the slower the \emph{fade}, like the $\alpha$ illustrated in Figure~\ref{fig:transicao}. The right side of Equations~\ref{eq:adsr} can cater for both introduction and ending of sound from zero intensity because they are linear. Schematically, Figure~\ref{fig:adsr} 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~\ref{eq:adsr} by the samples of the note or musical except:

\begin{equation}\label{eq:adsrApl}

\left\{t\_i^{ADSR}\right\}\_0^{\Lambda-1} =\{t\_i . a\_i\}\_0^{\Lambda-1}

\end{equation}

The music piece \emph{ADa and SaRa} explores many configurations of the ADSR envelope.~\cite{MASSA}

\section{Organization of notes in music}\label{notasMusica} \label{sec:notesMusic}

Let $S=\left\{ s\_j=T^j=\{t\_i^j\}\_{i=0}^{\Lambda\_j-1} \right\}\_{j=0}^{H-1}$ be a sequence of $H$ musical events $s\_j$.

Consider $S$ as a `musical structure', composed by events which are also musical structures themselves, e.g.\ notes.

This section is dedicated to techniques that make $S$ aesthetically appealing 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~\cite{MASSA} and can be further formalized~\cite{topos},

although at a cost of intelligibility which we chose to avoid in this exposition.

The elements of $S$ may be overlapped by mixing them together, as in

Equation~\ref{eq:mixagem} and Figure~\ref{fig:mixagem}, for building intervals and chords.

This reflects the `vertical thought' in music. On the other hand, the concatenation of events

in $S$, as in Equation~\ref{eq:concatenacao} and in Figure~\ref{fig:concatenacao}, builds melodic sequences and rhythms, which are associated with the `horizontal thought'. The fundamental frequency $f$ and the starting moment

(attack or onset) are generally the most important characteristics of the elements $s\_j$ in $S$.

This makes it possible to create music that relies heavily on pitches (both for harmony and melody) and on rhythms.

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

One will verify in the scripts that illustrate these concepts that it is considerably straightforward to

reach sonic materials of some artistic significance by

extrapolating musical paradigms established on traditional notions of melody, harmony and rhythm.

\subsection{Tuning, intervals, scales and chords}\label{subsec:afinacao}

\subsubsection{Tuning}

Doubling the frequency is equivalent to ascending one octave\footnote{A pitch

is both considered equivalent and different from any of its octaves.

It is different because one or more octaves higher or lower corresponds to

a higher or lower pitch.

It is equivalent in the sense that a pitch is very consonant to its octaves

and that any other pitch holds a similar relation to a pitch and its octaves.} ($f=2f\_0$).

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~\cite{Wisnick}.

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 $\varepsilon$. Twelve successive semitones (or half steps), equidistants to the human ear, yield an

octave. 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 $2^{\frac{1}{12}}$

for two main reasons: 1) the harmonic series does not match multiples of $2^{\frac{1}{12}}$,

and 2) because the harmonics of real sounds 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 (in the scientific literature) date from around two thousand years before the advent of the equal temperament~\cite{Roederer}.

Two emblematic tunings are:

\begin{itemize}

\item The {\bf 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.

\item The {\bf 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.

\end{itemize}

In order to account for microtonality\footnote{Microtonality 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 \emph{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 $\varepsilon=2^{\frac{1}{12}}$~\cite{Wisnick}.} 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 $\varepsilon$. For example, a tuning that approximates the harmonic series

is proposed with the equal division of the octave in $53$ notes:

$\varepsilon=2^{\frac{1}{53}}$~\cite{microtonalidade}.

Notice that if $S$ is a pitch sequence related by means of $\varepsilon\_1$,

the sequence with the same notes, but related by $\varepsilon\_2$, is

$\left\{s\_i'\right\}=\left\{

s\_i \frac{\varepsilon\_1}{\varepsilon\_2}\right\}$. The music piece \emph{Micro

tone} uses microtonal features.~\cite{MASSA}

\subsubsection{Intervals}\label{subsec:intervalos}

The frequency gap between two pitches is called an "interval" in musical jargon.

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~\ref{eq:intervalos} summarizes the intervals: traditional notation, qualifications of consonance and

dissonances, and number of semitones.

\begin{table\*}[htp!]

\centering

\caption{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 strong dissonances.

Major seconds and minor sevenths are the weak 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 the dominant (chord or harmonic field, see Section~\ref{subsec:harmonia}) and seeks urgent resolution into a third or sixth. Due to this instability it is considered a dissonant interval.}

\begin{tabular}{ c | c | c }\hline

\multicolumn{3}{c}{\bf consonances} \\\hline

& traditional notation & number of semitones \\

perfect: & P1, P5, P8 & 0, 7, 12 \\

imperfect: & m3, M3, m6, M6 & 3, 4, 8, 9 \\\hline\hline

\multicolumn{3}{c}{\bf dissonances} \\\hline

& traditional notation & number of semitones \\

strong: & m2, M7 & 1, 11 \\

weak: & M2, m7 & 2, 10 \\\hline\hline

\multicolumn{3}{c}{\bf special cases} \\\hline

& traditional notation & number of semitones \\

consonance or dissonance: & P4 & 5 \\

dissonance in Western tradition: & tritone, aug4, dim5 & 6 \\\hline

\end{tabular}\label{eq:intervalos}

\end{table\*}

The nomenclature, based on conveniences for tonal music and practical aspects of manipulating notes, can be specified

as follows~\cite{Roederer,Wisnick}:

\begin{itemize}

\item Intervals are inspected first by the number of steps between notes in some scale.

As can be noticed in Table~\ref{eq:intervalos}, one step is not one semitone.

A step, in this sense, is yielded 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 simple intervals are: first (unison), second, third, fourth, fifth, sixth, seventh and eighth (octave).

Each of these intervals are related to one step less then their names suggest: a third is an interval with two steps.

\item The intervals are represented by numeric digits, e.g. 1, 3, 5 are a unison, a third and a fifth, respectively.

\item 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' or 'minor'.

The tritone is an exception to this rule because it is a dissonant interval and cannot be major or minor.

\item The perfect fourth can be a perfect consonance or a dissonance according to the context and theoretical background.

As a general rule, it is considered a consonance except when it is followed by a third or a fifth

by the movement of one of the notes.

\item The tritone is a dissonance in Western music because

it characterizes the ``dominant'' chord in the tonal system (see Section~\ref{subsec:harmonia}) and represents (or yields) instability. Some cultures consider the interval a consonance, using it in melodies and chants as a stable interval.

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

\item 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 or TT).

\item 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 or TT).

\item 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.

\item Notes played simultaneously yield a harmonic interval.

\item Notes played as a sequence in time yield a

melodic interval. When the lowest note comes first there is an ascending interval,

while in a descending interval the highest note comes first.

\item 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 doubly-augmented results in a

doubly-diminished and vice-versa, etc).

An inverted perfect interval is a perfect interval as well.

\item 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

7 to the corresponding simple interval, e.g. P11 is an octave plus a forth ($7 + P4 = P11$), M9 is an octave plus a major second ($7 + M2 = M9$).

\end{itemize}

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 \textit{C flat} major scale, the tonic -- first degree -- is \textit{C flat}, not \textit{B}, and the leading tone -- seventh degree -- is \textit{B flat}, not \textit{A sharp} or \textit{C double flat}.

To grasp what this entails for intervals, let the second degree (second note) of a scale be one semitone from the first degree.

Consider 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~\cite{Lacerda}.

Notice that the dim3 is only two semitones wide, as is the major second (or e.g. a doubly-augmented unison!).

This description summarizes the traditional theory of musical intervals~\cite{Lacerda}.

The music piece \emph{Intervals between pitches} explores these intervals in both independent and related ways.~\cite{MASSA}

\subsubsection{Scales}\label{subsec:escalas}

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 mostly 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 admit.

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{equation}\label{escSim}

\begin{aligned}

\text{chromatic} & = E^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^{wt} & = & \{e\_i^{wt}\}\_0^{5} & = & \{0,2,4,6,8,10\} & = & \{2i\}\_0^{5} \\

\text{minor thirds} & = E^{mt} & = & \{e\_i^{mt}\}\_0^{3} & = & \{0,3,6,9\} & = & \{3i\}\_0^3 \\

\text{major thirds} & = E^{Mt} & = & \{e\_i^{Mt}\}\_0^{2} & = & \{0,4,8\} & = & \{4i\}\_0^2\\

\text{tritones} & = E^{tt} & = & \{e\_i^{tt}\}\_0^{1} & = & \{ 0, 6 \} & = & \{6i\}\_0^1

\end{aligned}

\end{equation}

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 \approx 251.98

Hz$. These `scales', or patterns, generate stable structures by their internal symmetries and can be repeated in a sustained way which are musically effective. Section~\ref{estCic} discusses symmetries.

The music piece \emph{Cristais} uses each one of these scales both melodically and harmonically.~\cite{MASSA}

The \emph{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{equation}\label{eq:escalas}

\begin{aligned}

\text{Aeolian} & = \text{natural minor scale} & = & \\

& = E^m = \{e\_i^m\}\_0^6 & = & \{0,2,3,5,7,8,10\} \\

\text{Locrian} & = E^{lo} = \{e\_i^{lo}\}\_0^6 & = & \{0,1,3,5,6,8,10\} \\

\text{Ionian} & = \text{major scale} & = & \\

& = E^M = \{e\_i^M\}\_0^6 & = & \{0,2,4,5,7,9,11\} \\

\text{Dorian} & = E^{d} = \{e\_i^{d}\}\_0^6 & = & \{0,2,3,5,7,9,10\} \\

\text{Phrygian} & = E^{p} = \{e\_i^{p}\}\_0^6 & = & \{0,1,3,5,7,8,10\} \\

\text{Lydian} & = E^{l}=\{e\_i^{l}\}\_0^6 & = & \{0,2,4,6,7,9,11\} \\

\text{Mixolydian} & = E^{mi} = \{e\_i^{mi}\}\_0^6 & = & \{0,2,4,5,7,9,10\}

\end{aligned}

\end{equation}

\noindent They have only major, minor and perfect intervals.

The unique exception is the tritone found as an augmented fourth or a diminished fifth.

% It is useful to name the diatonic scale degrees as: tonic, supertonic, mediant, subdominant, dominant, submediant, leading note (and then the tonic again).

% Accordingly, the first degree is the tonic, the fifth degree is the dominant.

% These names are unanimously used for the major scale and within tonal harmony (as presented in Section~\ref{subsec:harmonia}.

% In other scales these names might be used by extension.

% One critical exception is the dominant in the locrian scale: it is not a perfect fifth above the tonic, so it does not play the same role as in the other scales.

% This is the main reason why the locrian is sometimes not considered a diatonic scale.

Diatonic scales follow a circular pattern of successive intervals \textit{tone, tone, semitone, tone, tone, tone, semitone}. Thus, it is possible to write:

\begin{equation}\label{eq:relacaoDia}

\begin{split}

\{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{split}

\end{equation}

\noindent with $\kappa \in \mathbb{N}$. For each mode there is only one value for $\kappa \in [0,6]$ by which $\{e\_i\}$ matches.

For example, a brief inspection reveals that

$e\_i^{l}=d\_{(i+2)\%7}+e\_{i-1}^{l}$. Then, $\kappa=2$ for the Lydian scale.

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

\begin{equation}\label{eq:escalasMenores}

\begin{split}

\text{natural minor}& = E^m = \{e\_i^m\}\_0^6 = \\

& = \{0,2,3,5,7,8,10\} \\

\text{harmonic minor} & = E^{mh} = \{e\_i^{mh}\}\_0^6 = \\

& = \{0,2,3,5,7,8,11\} \\

\text{melodic minor} & = E^{mm} = \{e\_i^{mm}\}\_0^{14} = \\

& = \{0,2,3,5,7,9,11,12,10,8,7,5,3,2,0\} \\

\end{split}

\end{equation}

The different ascending and descending contours of the melodic minor scale is required in tonal harmony contexts.

The minor scale has one whole tone between the seventh and eighth (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~\cite{Harmonia}.

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

\begin{equation}\label{eq:serieHarmonica}

\begin{split}

H = & \{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{split}

\end{equation}

In this scale, the frequency of the $i$th note $h\_i$ is the frequency of $i$th harmonic $f=\varepsilon^{h\_i} f\_0$ from the spectrum generated by $f\_0$. Natural sounds have such frequencies (as discussed in Section~\ref{sec:discNote}) 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~\cite{Messiaen}.

One last observation: the words \emph{scale} and \emph{mode} are often used as synonyms both in the literature and in colloquial discussions.

The word \emph{mode} can also be used to mean two other things:

\begin{itemize}

\item an unordered set of pitches (i.e. an unordered scale).

\item A scale used in the context of modal harmony, in the sense presented in Section~\ref{subsec:harmonia}.

\end{itemize}

\subsubsection{Chords}\label{subsec:acordes}

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~\cite{Lacerda}, 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{equation}\label{triades}

\begin{split}

\text{major triad} = A^M= \{a\_i^M\}\_0^2=\{0,4,7\} \\

\text{minor triad} = A^m = \{a\_i^m\}\_0^2=\{0,3,7\} \\

\text{diminished triad} = A^d = \{a\_i^d\}\_0^2=\{0,3,6\} \\

\text{augmented triad} = A^a = \{a\_i^a\}\_0^2=\{0,4,8\}

\end{split}

\end{equation}

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 by the

addition of $\pm 12$ to the selected pitch. 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:

\begin{itemize}

\item A fifth confirms the root (fundamental).

There are theoretical discussions 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 enhances the sensation that

the root of a chord is the root.

\item Major or minor thirds from the root entail major or minor chord qualities.

\item Every tritone, especially if built between a major third and a minor seventh, tends to resolve into a third or a sixth.

\item Note duplication is avoided. If duplication is needed, the preference is, in descending order: the root, fifth, third and seventh.

\item Note omission is avoided in the triad. If needed, the fifth is first considered for omission, then third and then the fundamental.

\item 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.

\item Chords built by successive intervals different from thirds -- such as fourths and seconds -- are recurrent in compositions of advanced tonalism or experimental character.

\item 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.

\end{itemize}

\subsection{Tonal, modal and atonal music, harmonies, harmonic expansion and modulation}\label{subsec:harmonia}

The omission of basic tonal structures is the key to achieving modal and atonal harmonies\footnote{To avoid very complicated discussions,

the term 'harmony' is used in this article in a very loose sense:

considerations, grouped by music theory as harmony, about how sounds are analyzed by hearing.

These considerations stem substantially from chords and scales and the absence of them.}. In the absence of minimal tonal organization,

harmony is (usually) considered modal if the notes match with some diatonic scale (see Equations~\ref{eq:escalas}) 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 diverse enough and dissonant (between themselves) to avoid reduction of the notes by polarization\footnote{By polarization we mean the presence of notes that are way more important than others and to which the other notes are ornaments or subordinates.}, 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.

Atonal harmony is hard to be fulfilled and, indeed, no matter how dissonant and diverse

a set of notes is, tonal harmonies arise very easily if not avoided~\cite{harmEXT}.

\subsubsection{Atonal harmony}\label{sec:atonal}

In fact, atonal music techniques aim at avoiding that a direct relation of the notes with modes and tonality be established by the listener.

Manifesting such atonal structures is of such difficulty that the dodecafonism emerged.

Dodecafonism consists in using a sequence of pitches (ideally the 12 pitches of the chromatic scale),

and performing each pitch, 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:

\begin{itemize}

\item Repeat pitches. By considering an immediate repetition as an extension of the previous incidence, the use of the same pitch in sequence does not add relevant information.

\item To play adjacent pitches (of the dodecafonic sequence) at the same time, making harmonic intervals and chords.

\item Use durations and pauses with freedom while respecting the order of incidence of the pitches.

\item Vary note sequences by temporal magnification and translation or pitch transposition and by pitch sequence inversion, retrograde and retrograde inversion. See Sections~\ref{subsec:motivos}

and~\ref{subsec:usosmusicais3} for the meaning of these terms.

\item Make variations in timbre (e.g. orchestration), articulation, spatialization, among other possibilities in the presentation of the same pitches.

\end{itemize}

The atonal harmony can be observed, paradigmatically, within these presented conditions (which is a simple dodecaphonic model).

Part 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 as stated above, based on a sequence of pitches (called the series or row), but not within a 12 tone sequence and using sequences of other parameters (e.g. duration, timbre).

Serialism and pos-serialism is briefly considered in Secion~\ref{sec:ser}.

\subsubsection{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 common 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~\ref{triades}-- built on the root note of each degree of a scale -- such as listed in Equations~\ref{eq:escalas} -- form the pillars of harmony.

Tonal (and modal) harmony theory aims at understanding the incidence of chord progressions and chaining rules.

Even a monophonic melody entails harmonic fields, making it possible to perceive a 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~\ref{eq:escalasMenores}). 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~\cite{Harmonia}.

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 \textbf{tonic} chord (\textit{T} or \textit{t} for a major or minor tonic, respectively) and its function (role) consists in maintaining a center, usually referred to as a ``ground'' for the music. The chord built on the fifth degree is the \textbf{dominant} (\textit{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 \textbf{subdominant} (\textit{S} or \textit{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 \textit{Tr}, relative

subdominant \textit{Sr} and relative dominant \textit{Dr}) is the triad built a third below, and the associated counter-relative (counter-relative tonic \textit{Tc}, counter-relative subdominant \textit{Sc} and the counter-relative dominant \textit{Dc}) 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 (tC, sC) 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~\ref{tab:harmonia} shows relations between the triads built in each degree of the major scale.

\begin{table}[htp!]

\centering

\caption{Summary of tonal harmonic functions on the major scale.

The 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 relatives or counter-relatives.}

\begin{tabular}{l | c | r}

relative & main chord of the function & counter-relative \\\hline\hline

$\hat{6},\hat{1},\hat{3}$ & 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: $\hat{4},\hat{6},\hat{1}$ & $\hat{6},\hat{1}, \hat{3}$

\end{tabular}

\label{tab:harmonia}

\end{table}

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'~\cite{Koellheuteur}.

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.

\subsubsection{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, also have subdominants and 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, with the root at the third of the original chord;

and the lower mediant, with the fifth at the third of original chord.

Both chords also are separated by a third from the original chord, like the relatives and counter-relatives,

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 the third on the fifth of the original triad;

and the lower form, with the third on 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.)

Mediants are considered of 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 late 19th century by Wagner, Lizt, Richard Strauss, among others~\cite{Harmonia,Salzer}.

\subsubsection{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:

\begin{itemize}

\item Transposing the discourse to a new key, without any preparation.

It is a common Baroque procedure although incident in other periods. Sometimes it is

called phrasal modulation or unprepared modulation.

\item Careful use of an individual dominant, and perhaps also the individual

subdominant, to confirm a change in key.

\item Use of chromatic alterations to reach a chord in the new key by starting from a chord in the previous key. Called chromatic modulation.

\item 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.

\item Changing the function, without changing the notes, of a chord to

reach a new key. This procedure is called enharmony.

\item 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.

\end{itemize}

The dominant has great importance and is a natural pivot in modulations,

a fact that emphasizes the relevance of the 'circle of fifths'~\cite{Harmonia,Salzer,Koellheuteur,Harmony}.

Other inventive ways to modulate are possible, to point but one common example, the minor thirds tetrad ($E\_i^{tm}$ in Equations~\ref{escSim}) can be sustained as a bridge between tonalities, with the possible use of its both tritones.

The music piece \emph{Acorde cedo} explores these chord relations.~\cite{MASSA}.

\subsection{Counterpoint}\label{subsec:contraponto}

Counterpoint is a set of techniques for 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~\cite{Fux,SchoenbergContra}. It is possible to

summarize counterpoint rules, and it is known that Beethoven --

among others -- also outlined such a digest of counterpoint.

\begin{figure}[h!]

\centering

\includegraphics[width=.8\columnwidth]{figures/movContraponto\_}

\caption{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.}

\label{fig:movContraponto}

\end{figure}

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~\ref{fig:movContraponto}.

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

relevance lies mainly in the highest and then in the lowest of the voices~\cite{Fux,Tragtenberg,SchoenbergContra}.

These rules were used in the music piece \emph{Conta ponto}.~\cite{MASSA}

\subsection{Rhythm}\label{subsec:ritmo}

Rhythm is dependent on events separated by durations~\cite{Lacerda}.

For the temporal separation between them to be perceived as duration,

the period should be at least around $75ms$~\cite{microsound}.

It is possible to summarize the durations heard as rhythm or pitch, and its transition,

as in Table~\ref{tab:duracoes}~\cite{Alfaix, microsound}.

\begin{table\*}[htp!]

\caption{Durations heard as rhythm, as pitch and transition.}

\begin{tabular}{ l | r r r r r r r r r r || r r } \hline

& \multicolumn{10}{c}{\bf perception of durations as rhythm} & \multicolumn{2}{c}{} \\

duration (s) & \multirow{2}{\*}{\bf ...} & {\bf 32,} & {\bf 16,} & {\bf 8,} & {\bf 4,} & {\bf 2,} & {\bf 1,} & {\bf 1/2,} & {\bf 1/4,} & {\bf 1/8,} & \multirow{2}{\*}{ ... transition } & \\

frequency (Hz) & & {\color{gray} 1/32,} & {\color{gray} 1/16,} & {\color{gray} 1/8,} & {\color{gray} 1/4,} & {\color{gray} 1/2,} & {\color{gray} 1,} & {\color{gray} 2,} & {\color{gray} 4,} & {\color{gray} 8,} & & \\

& \multicolumn{10}{c}{ - } & \multicolumn{2}{c}{} \\ \hline

\end{tabular}

\vspace{.2cm}

\begin{tabular}{ l | p{1.7cm} || r r || r } \hline

& \multicolumn{1}{c}{} & \multicolumn{2}{c}{transition} & \\

duration (s) & \multirow{2}{\*}{\bf rhythm ...} & $\frac{1}{16}=62.5ms$ , & $\frac{1}{20}=50ms$ & \multirow{2}{\*}{\bf ... pitch} \\

frequency (Hz) & & 16, & 20 & \\

& \multicolumn{1}{c}{} & \multicolumn{2}{c}{transition} & \\ \hline

\end{tabular}

\hspace{3.36cm}

\vspace{.2cm}

\begin{tabular}{ l | p{1.7cm} || r r r r r r }\hline

& \multicolumn{1}{c}{} & \\

duration (s) & \multirow{2}{\*}{transition ...} & {\color{gray} 1/40} & {\color{gray} 1/80 } & {\color{gray} 1/160 } & {\color{gray} 1/320 } & {\color{gray} 1/640 } & \multirow{2}{\*}{\bf ... } \\

frequency (Hz) & & {\bf 40} & {\bf 80} & {\bf 160} & {\bf 320} & {\bf 640} & \\

& \multicolumn{1}{c}{} & \multicolumn{6}{c}{\bf perception of durations as pitch} \\ \hline

\end{tabular}

\hspace{2.70cm}

\label{tab:duracoes}

\end{table\*}

\begin{figure\*}

\centering

\includegraphics[width=.9\textwidth]{figures/metricaMusical\_\_}

\caption{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~\protect\footnotemark.}

\label{fig:pulsoSubAgl}

\end{figure\*}

\footnotetext{The

definition of the terms 'quarter note' and 'time signature' are beyond the scope of this work and might be found in elementary texts about musical theory~\cite{Lacerda}.

Anyway, what is important for the reader to grasp is that: 1) successive temporal elements of equal duration

are both grouped and subdivided into other elements of equal duration;

2) these groupings and subdivisions are usually made in terms of small integers and are discerned into

simple, compound and complex.}

The transition span in Table~\ref{tab:duracoes} is minimized because the limits

are not well defined,

i.e. the exact duration that is perceived as a

frequency or a separation between occurrences depends on the

listener and the sonic characteristics~\cite{microsound,Roederer}.

The rhythm is commonly based on a basic duration called pulse, whose typical duration ranges between $0.25$ and $1.5s$ ($240$

and $40 BPM$, respectively~\footnote{BPM stands for Beats Per Minute and is just a frequency measure like Herz, but is the number of incidences per minute instead of second. BPM is often used as a measure of musical tempo and of heart rate.}). In music education and cognitive studies, this range of frequencies is often associated with the duration of the heart beat, movements of respiration and steps of a walking or running person~\cite{Lacerda,Roederer}.

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, popular and ethnic

music, are: 2, 4 and 8; 3, 6 (2 groups of 3 or 3 groups of 2), 9 and 12 (3 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.

A schematic illustration is shown in Figure~\ref{fig:pulsoSubAgl}.~\cite{Gramani,Roederer}

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 through a ternary division 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 and 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 \emph{'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~\cite{Roederer}.

Small variations along the temporal grid are crucial for musical interpretation and

styles~\cite{Cook}.

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 $j$: $P^0\_{10}$ is the tenth pulse, $P^{1}\_3$ is the third grouped unit,

$P^{-1}\_2$ 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 $P^{min(j)}\_i$, $\forall \; i$, should be greater than $50 ms$ and the durations

summed together $\sum\_{\forall i}P^{\text{max}(j)}\_i$ 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~\ref{fig:pulsoSubAgl}. Any unit (note) of a given musical sequence with a time metric can be unequivocally

specified as:

\begin{equation}\label{eq:rhythmicUnit}

P^{ \{ j\_k \} }\_{ \{ i\_{k} \}}

\end{equation}

\noindent where $j\_k$ is the grouping level and $i\_k$ is the unit itself.

As an example, consider $P^{-1,0,1}\_{3,2,2}$ as the third subdivision $P^{-1}\_3$ of the

second pulse $P^0\_2$ and of the second pulse group $P^1\_2$.

Each unit $P\_i^j$ can be associated with a sequence of temporal samples $T$ that constitutes a

note. The music piece \emph{Poli Hit Mia} uses different rhythmic metrics.~\cite{MASSA}

\subsection{Repetition and variation: motifs and larger units}\label{subsec:motivos}

Given the basic musical structures of the previous sections, it is

natural to present these structures in a coherent and meaningful way~\cite{Boulez}.

The concept of arc 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 music and

section, constitutes an arc.

Music in which the arcs do not present consistency with one another can be understood as music with little or 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

$S=\left\{s\_j=T^j=\{t\_i^{j}\}\_0^{\Lambda\_j-1}\right\}\_0^{H-1}$ be a sequence

of $H$ musical events $s\_j$, each event with its $\Lambda\_j$ samples $t\_i^j$

(refer to the beginning of this Section~\ref{notasMusica} if needed). Below is a list of basic techniques

for variation.

\begin{itemize}

\item Temporal translation is a displacement

$\delta$ 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 $\Gamma$ is the duration between the beginning of the piece (or considered

section) and the first event $s\_0$ of the original structure $S$, and

$\delta$ is the temporal offset of the displacement.

\item 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.

\item Temporal reversion consists on generating a sequence with elements

in the reverse order of the original sequence $S$, thus:

$S'=\left\{s\_j'\right\}\_0^{H-1}=\left\{s\_{(H-j-1)}\right\}\_0^{H-1}$.

\item Pitch translation, or transposition, is a displacement $\tau$ of the material to a

different pitch $\Xi'=\Xi + \tau$. It is a variation that changes pitch

localization of material:

$\left\{s\_j'\right\}=\left\{s\_j^{\Xi'}\right\}=\left\{s\_j^{\Xi+\tau}\right\}$.

If $\tau$ is given in semitones, new frequencies are given by $f'=f2^{\frac{\tau}{12}}$;

and the new pitches: $\Xi'=\Xi +12 \log\_2\left(\frac{f'}{f}\right)=\Xi+\tau$.\footnote{In

the MIDI protocol, $\Xi\_{f}=55Hz$ when pitch $\Xi=33$

(an \textit{A1} note). Another good MIDI reference is $\Xi\_{f}=440Hz$ and

$\Xi=69$ (\textit{A4}). The difference ($\tau=\Xi\_1 - \Xi\_2$) is in semitones.

Notice that $\Xi=1$ is not a semitone, it is a note with an audible frequency as rhythm, with

less than 9 occurrences each second (see Table~\ref{tab:duracoes}).}

\item Interval inversion is either: 1) the inversion of note pitch order, within octave equivalence, such as described in Section~\ref{subsec:intervalos}; or 2) the inversion of interval orientation. In the former case, the number of semitones

is preserved in the ``strict invertion'':

$S'=\{s\_j'\}\_0^{H-1}=\left\{s\_j^{-\varepsilon\_j . f\_0}\right\}$, where

$\varepsilon\_j$ 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'=\{s\_j'\}\_0^{H-1}=\left\{s\_j^{2^{\left(\frac{12-e\_{\left(7-j\_e\right)}}{12}\right)}

. f\_0}\right\}\_0^{H-1}$ where $j\_e$ is the index in

$E\_k$ of the note $s\_j$.

\item 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 $\tilde{n}'=H-\tilde{n}$.

One might associate $\tilde{n}>0$ (events advance) with the anti-clockwise rotation and

$\tilde{n}<0$ (elements delay) with the clockwise rotation.

Additional information about rotations is given in Section~\ref{estCic}.

\item The insertion and removal of material in $S$ can be

ornamental or structural: $S'=\{s\_j'\}=\{s\_j \text{ if condition A,

otherwise } r\_j\}$, for any music material $r\_j$, including silence.

Elements of $R$ can be inserted at the beginning, as a prefix;

at the end, as a suffix; or in the middle, dividing $S$ and making

it the prefix and suffix of $R$. Both materials can be mixed in a variety of ways.

\item Changes in articulation, orchestration and spatialization, or

$s\_j'=s\_j^{\*\_j}$, where $\*\_j$ is the new characteristic incorporated by

element $s\_j'$.

\item Accompaniment.

E.g. orchestration and melodic lines present when $S$ occurs can

be modified and be considered a variation of $S$ itself.

\end{itemize}

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.~\cite{Sacks,Roederer}

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 constructions 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~\cite{Schoenberg}.

% melody, answer and reply, motives, etc

\subsection{Directional structures}\label{subsec:dir}

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~\ref{fig:climax}. The varying parameter can be non-existent, a case in which

the arc consists only of a reference structure~\cite{Schoenberg}\footnote{A

case which resembles a note without the fundamental frequency. In some cases, the listener is able to identify

the fundamental frequency even though it is absent.}.

\begin{figure}

\centering

\includegraphics[width=.8\columnwidth]{figures/climax\_}

\caption{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.}

\label{fig:climax}

\end{figure}

Consider the $S=\{s\_i\}\_0^{H-1}$ increasing sequence. The sequence

$R=\{r\_i\}\_0^{2H -2}=\left\{s\_{(H-1-|H-1-i|)}\right\}\_0^{2H-2}$

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~\cite{Guidorizzo,Schoenberg}.

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~\cite{Salzer}.

In practice, and historically, there is special use of the golden ratio.

The Fibonacci sequence might be generalized as follows in order to use any two numbers

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~\cite{oeis})

for exploring various sequences.

The musical piece \emph{Dirracional} explores directional arcs,~\cite{MASSA}

\subsection{Cyclic structures}\label{estCic}

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~\cite{Deleuze}. 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~\cite{gruposFascination}.

In music, permutations are ubiquitous in scholastic techniques,

which confirms their central role.

The successive application of permutations generates cyclic arcs~\cite{change,Zamacois,permMusic}. Two academic studies were dedicated to generating musical structures using permutation groups~\cite{figgusOriginal, figgusEspacializacao}. Any permutation set can be used as a generator of algebraic groups~\cite{permMusic}.

The properties defining a group $G$ are:

\begin{equation}\label{eq:groups}

\begin{split}

\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 \;,\;\;\; \forall\; p \in G \\

& \text{(existence of the identity element)} \\

\forall \;\; p \in G, \;\exists\; p^{-1} :\quad & p\bullet p^{-1} =p^{-1}\bullet p = e \\

& \text{(existence of the inverse element)}

\end{split}

\end{equation}

From the first property follows that two permutations act as one permutation. In fact, it is possible to apply a

permutation $p\_1$ and then 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.

As a historic

example, the \emph{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 \emph{peal}. Table~\ref{tab:change}

presents a traditional \emph{peal}, named ``Plain Change''~\cite{change}, for 3 bells (\textcolor{red}{1}, \textcolor{blue}{2} and \textcolor{green}{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.

\begin{table}[htp!]

\centering

\caption{A traditional change ringing \emph{peal} for 3 bells. Permutations

occur between each line. Each line is a bell ordering to be played before next line.}

\begin{tabular}{l c r}

\textcolor{red}{1} & \textcolor{blue}{2} & \textcolor{green}{3} \\

\textcolor{blue}{2} & \textcolor{red}{1} & \textcolor{green}{3} \\

\textcolor{blue}{2} & \textcolor{green}{3} & \textcolor{red}{1} \\

\textcolor{green}{3} & \textcolor{blue}{2} & \textcolor{red}{1} \\

\textcolor{green}{3} & \textcolor{red}{1} & \textcolor{blue}{2} \\

\textcolor{red}{1} & \textcolor{green}{3} & \textcolor{blue}{2} \\

\textcolor{red}{1} & \textcolor{blue}{2} & \textcolor{green}{3}

\end{tabular}

\label{tab:change}

\end{table}

The use of permutations in music can be summarized in the following way:

with $S=\{s\_i\}$ a sequence of musical events $s\_i$ (e.g.\ notes), and a

permutation $p$. $S'=p(S)$ comprises the same elements of $S$ 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 between elements and its successors, and between the last element and the first.

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 $S$,

but only some characteristic. Thus, if $p$ is a permutation and $S$ is a sequence of basic notes as in the end of Section~\ref{notaBasica}, the

sequence $S'=p^f(S)=\left\{s\_i^{p(f)}\right\}$ 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.e.\ it can operate in a subset of $S$.

Second, not all elements $s\_i$ need to be executed at each access to $S$. To exemplify, let $S$ 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 $S$ can occur in other measures where permutations

allocate those notes to the first four events $s'\_i$ of $S'$.

To each permutation $p\_i$ described above, we have to determine:

1) note characteristics where it operates (frequency, duration, \emph{fades},

intensity, timbre, etc); and

2) the period of incidence (how many consults before a permutation is

applied).

For a coherent realization of the notes in $S$ is to

execute the first $n$ notes given that the execution of disjoint sets of $S$

is the same as modifying the permutation and executing the first $n$ notes.

PPEPPS is a set of musical pieces that uses permutations.~\cite{MASSA,figgusOriginal,figgusEspacializacao}.

\subsection{Serialism and post-serial techniques}\label{sec:ser}

Expanding the concepts exposed in Section~\ref{sec:atonal},

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 loosely,

such as by arbitrarily 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:

\begin{itemize}

\item Spectralism:

consists in using 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.

Examples: the most prominent frequencies of a sound 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.~\cite{grisey}

\item Spectromorphology:

can be considered a spectral music (spectralism) theoretical framework

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.~\cite{smalley,schaeffer}

\item Stochastic music:

the use of random variables to describe musical elements are largely considered in stochastic music.

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.~\cite{formalized}

\item 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 \emph{granular synthesis}~\cite{microsound}.

\end{itemize}

\subsection{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'~\cite{Lerdahl, Harmonia, Salzer,Alfaix}. 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~\ref{subsec:motivos}:

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 convergence of all the previous content.

\subsection{Musical usages}\label{subsec:usosmusicais3}

We strived to supply the reader with a very useful collection of

fundamental concepts to elaborate musical structures

with the content of Sections~\ref{sec:discNote} and~\ref{sec:internalVar}.

The numerous resources and consequent infinitude

of praxis possibilities are typical and highly relevant for artistic contexts~\cite{Harmonia,Webern}.

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 \emph{polyrhythm}. The music piece \emph{Poli-hit mia} explores these simultaneous metrics by impulse trains convolved with notes.~\cite{MASSA}

Microtonal scales are important for 20th

century music~\cite{microtonalidade} and yielded diverse remarkable results throughout history. For instance, fourths of a tone ($\epsilon=2^{\frac{1}{24}}$) are often used in some genres of Indian and Arabic music.

The musical sequence \emph{MicroTom} explores microtonal melodies and harmonies.~\cite{MASSA}

As stated in Section~\ref{subsec:mus2}, relations between

parameters are powerful ways to achieve musical pieces.

The number of permuted

notes can vary during the music piece, revealing relationship with piece

duration. Harmonies can be made from triads (Equations~\ref{triades}) with duplicated

notes at each octave and more numerous duplications entail lower depth and frequency of

vibratos (Equations~\ref{vbrGamma},~\ref{vbrAux},~\ref{vbrF},~\ref{vbrGamma2},~\ref{vbrT}).

For sure, the possibilities are very wide, which is evidenced in the wide range of musical pieces and styles.

The symmetries at octave divisions (Equation~\ref{escSim}) and the

symmetries presented as permutations (Table~\ref{tab:change} and

Equations~\ref{eq:groups}) can be used together. In the music piece \emph{3 trios}

this association is done in a systematic way in order to achieve a specific style (or idiom).

This is an instrumental piece, not included as a source

code but available online~\cite{3Trios}.

\emph{PPEPPS} (Pure Python EP: Project Solvent) is an EP (Extended Play) synthesized using

resources presented in this article. 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 with potential to both

sharing and education.

\section{Conclusions and further developments}

\label{cap:conclusao}

In this article, we hope to have effectively related musical elements to digital audio.

We aimed at achieving a concise presentation of the subject that involves many knowledge fields. Some readers might benefit from the text alone, but the \emph{scripts} in the \massa\ 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 insights.

The scripts include routines that render musical pieces to illustrate the concepts in practical contexts

even when we did not use equations in the descriptions (mainly in Section~\ref{sec:notesMusic}).

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 article and the open source scripts supplied 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 in this article stem from the organization of knowledge and the ability to achieve sounds that are 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, standard software for synthesis and processing of sounds for musical purposes restricts 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 by noticing that:

\begin{itemize}

\item the integrals and derivatives, ubiquitous in continuous signal processing,

are all replaced, in discrete signals, by summations,

which are generally more intuitive and do not require fluency in calculus.

\item The equations and concepts are implemented as software which can be easily assembled and inspected.

\end{itemize}

\noindent Significantly, this framework was used in a number of contexts, including courses, software implementations and making music~\cite{vimeoLM,vivace,repoDissertacao,dissertacao}.

Such detailed analytical descriptions, together with the computational implementations, have not been covered before in the literature, as far as the authors know (see Appendix G of~\cite{dissertacao}), 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 generation of sub-products in a co-authorship fashion,

favoring new implementations and 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.

There are also a few auxiliary scripts which did not fit into these categories.~\cite{SI}

% This framework favored the formation of interest groups in topics such as musical creativity and computer music.

% In particular, the project \url{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 \massa\ framework.\footnote{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)~\cite{siteLM,wikiLM,vimeoLM}.

% Some of these efforts are available online~\cite{dissertacao}.

% It is evident that all these contributions are a consequence of more that just \massa, but it is also evident to the authors that \massa\ had a primary role in converging interests and attracting collaborators.}

Future work might include application of these results in artificial intelligence

for the generation of appealing artistic materials. Some psychoacoustic effects were detected,

which need validation and should be reported, especially with~\cite{quadrosSonoros}.\footnote{The sonic portraits were 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~\cite{quadrosMetarec}.} 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.

\begin{acks}

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

\end{acks}

% Bibliography

\bibliographystyle{ACM-Reference-Format}

% \bibliography{sample-bibliography}

\bibliography{../article2}