Musical elements in the discrete-time representation of sound

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

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

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

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

ACM Reference format:
Renato Fabbri, Vilson Vieira da Silva Junior, Antônio Carlos Silvano Pessotti, Débora Cristina Corrêa, and Osvaldo N. Oliveira Jr.. 0. Musical elements in the discrete-time representation of sound. *ACM Comput. Surv.* 0, 0, Article 0 (0), 5 pages.
DOI: 0000001.0000001

This work is supported by FAPESP and CNPq.

DOI: 0000001.0000001

1 SPECTRA OF SAMPLED SOUNDS

The sinusoidal components in the discretized sound have some particularities. Considering a signal T and its corresponding Fourier decomposition $\mathcal{F}\langle T\rangle = C = \{c_k\}_0^{\Lambda-1} = \left\{\sum_{i=0}^{\Lambda-1} t_i e^{-ji(k\frac{2\pi}{\Lambda})}\right\}_0^{\Lambda-1}$, the recomposition is the sum of the frequency components to yield the temporal samples[1]:

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

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, and are limited by $w_k \leq \pi$ and $f_k \leq \frac{f_s}{2}$ as given by the Nyquist Theorem.

For a sonic signal, samples t_i are real and are given by the real part of Equation 1:

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

where $arctan(x, y) \in [0, 2\pi]$ is the inverse tangent with the right choice of the quadrant in the imaginary plane.

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

The same applies to $w_k = f_k\frac{2\pi}{f_s}$ and the periodicity 2π: it follows that $w_k = -w_{\Lambda-k}\ ,\ \ \forall\ \ k < \Lambda$. Given the cosine (an even function) and the inverse tangent (an odd function), the components in w_k and $w_{\Lambda-k}$ contribute with coefficients $c_k = c^*_{\Lambda-k}$ in the reconstruction of the real samples. In summary, in a decomposition of Λ samples, the Λ frequency components $\{c_i\}_0^{\Lambda-1}$ are equivalent in pairs, except for f_0, and, when Λ is even, for $f_{\Lambda/2} = f_{\max} = \frac{f_s}{2}$. Both these components are isolated, i.e. there is one and only one component at frequency f_0 or $f_{\Lambda/2}$ (if Λ is even). In fact, when $k = 0$ or $k = \Lambda/2$ the mirror of the frequencies are themselves: $f_{\Lambda/2} = f_{(\Lambda-\Lambda/2)=\Lambda/2}$ and $f_0 = f_{(\Lambda-0)=\Lambda} = f_0$. Furthermore, these two frequencies (zero and Nyquist frequency) do not have a phase offset: their coefficients are strictly real. Therefore, the number τ_Λ of equivalent coefficient pairs in a decomposition of Λ samples is:

$$
\tau_\Lambda = \frac{\Lambda - \Lambda\%2}{2} - 2 + \Lambda\%2 = \left\lceil\frac{\Lambda}{2}\right\rceil - 2 \tag{3}
$$

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

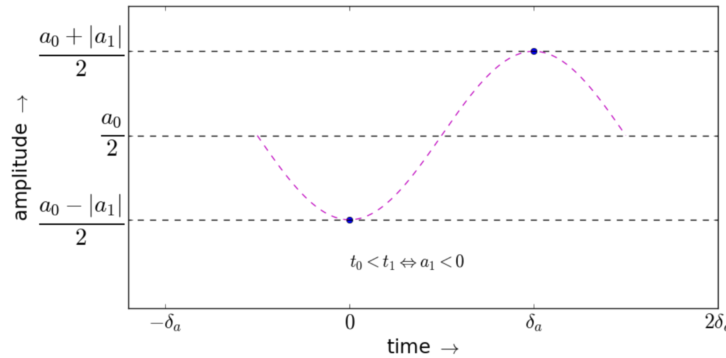

Fig. 1. Oscillation of 2 samples (maximum frequency for any f_s). The first coefficient determines a constant detachment (called *offset*, *bias* or *DC component*) and the second coefficient specifies the oscillation amplitude.

This discussion can be summarized in the following equivalences:

$$f_k \equiv f_{\Lambda-k} \quad , \quad w_k \equiv -w_{\Lambda-k} \tag{4}$$

$$a_k = a_{\Lambda-k} \quad , \quad b_k = -b_{\Lambda-k} \tag{5}$$

$$\sqrt{a_k^2 + b_k^2} = \sqrt{a_{\Lambda-k}^2 + b_{\Lambda-k}^2} \tag{6}$$

$$arctan(b_k, a_k) = -arctan(b_{\Lambda-k}, a_{\Lambda-k}) \tag{7}$$

with $\forall\ 1 \leq k \leq \tau_\Lambda, k \in \mathbb{N}$.

To express the general case for components combination in each sample t_i, one can gather the relations for the reconstruction of a real signal (Equation 2), for the number of paired coefficients (Equation 3), and for the equivalences of modules (Equation 6) and phases (Equation 7):

$$t_i = \frac{a_0}{\Lambda} + \frac{a_{\Lambda/2}}{\Lambda}(1 - \Lambda\%2) + \frac{2}{\Lambda}\sum_{k=1}^{\tau_\Lambda}\sqrt{a_k^2 + b_k^2}\ cos\left[w_k i - arctan(b_k, a_k)\right] \tag{8}$$

Figure 1 shows two samples and their spectral component. When there is only two samples, the Fourier decomposition has only one pair of coefficients $\{c_k = a_k - j.b_k\}_0^{\Lambda-1=1}$ relative to frequencies $\{f_k\}_0^1 = \left\{w_k \frac{f_s}{2\pi}\right\}_0^1 = \left\{k \frac{f_s}{\Lambda=2}\right\}_0^1 = \left\{0, \frac{f_s}{2} = f_{\max}\right\}$ with energies $e_k = \frac{(c_k)^2}{\Lambda=2}$. The role of amplitudes a_k is clearly observed with $\frac{a_0}{2}$, the fixed offset (also called *bias* or *DC component*), and $\frac{a_1}{2}$ for the oscillation with frequency $f_1 = \frac{f_s}{\Lambda=2}$. This case has special relevance: at least 2 samples are necessary to represent an oscillation and it yields the Nyquist frequency $f_{\max} = \frac{f_s}{2}$, which is the maximum frequency in a sound sampled with f_s samples per second.

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 $f = f_s/3$. This case is illustrated in Figure 2.

With 4 samples it is possible to represent 1 or 2 frequencies with independence of magnitude and phase. Figure 3 depicts the contribution of each of the two (possible) components. The individual

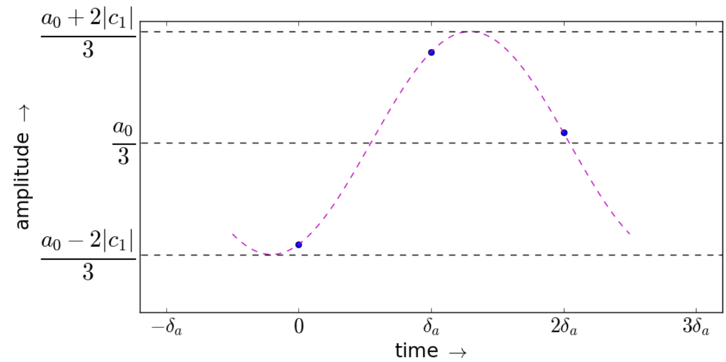

Fig. 2. Three fixed samples present only one non-null frequency. $c_1 = c_2^*$ and $w_1 \equiv w_2$.

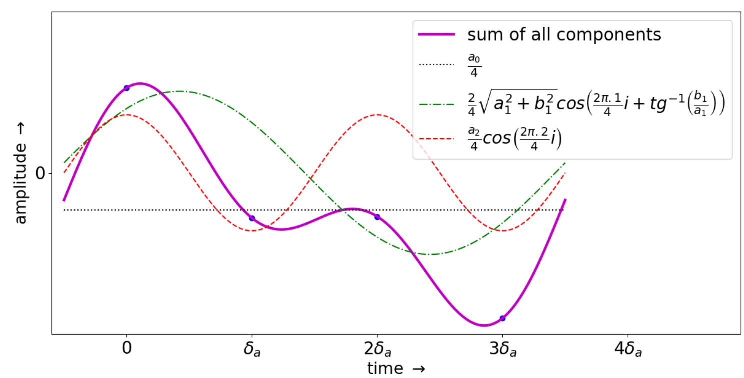

Fig. 3. Frequency components for 4 samples.

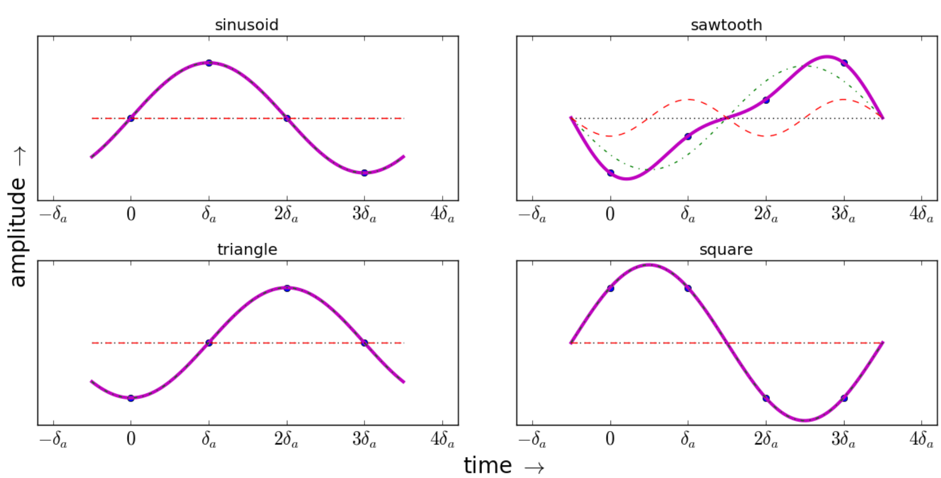

Fig. 4. Basic waveforms with 4 samples.

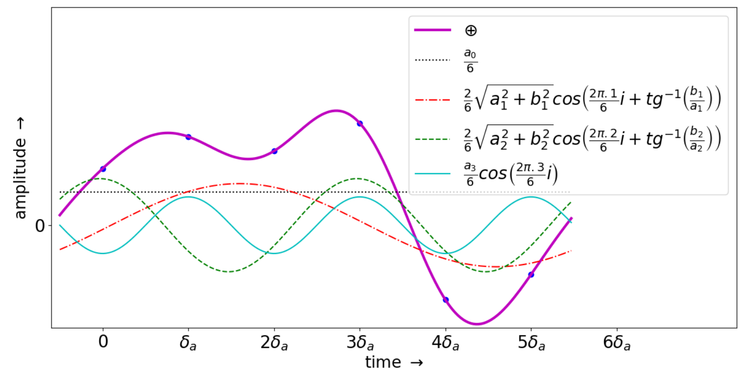

Fig. 5. Frequency components for 6 samples: 4 sinusoids, one of them is the *bias* with zero frequency.

components sum to the original waveform and a brief inspection reveals the major curvatures resulting from the higher frequency, while the fixed offset is captured in the component with frequency $f_0 = 0$. Figure 4 shows the harmonics for the basic waveforms of Equations 10, 11, 12 and 13 in the case of 4 samples. There is only 1 sinusoid for each waveform, with the exception of the sawtooth, which has the even harmonics.

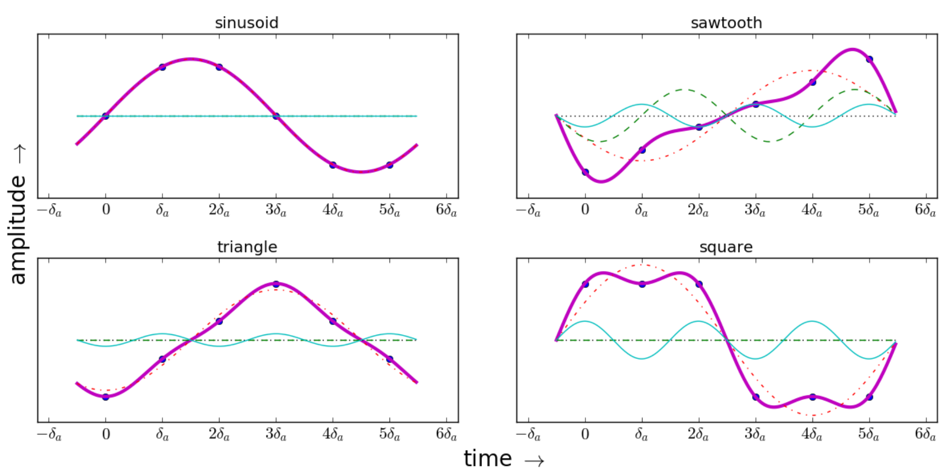

Fig. 6. Basic waveforms with 6 samples: triangular and square waveforms have odd harmonics, with different proportions and phases; the sawtooth has even harmonics.

Figure 5 exposes the sinusoidal components within 6 samples, while Figure 6 presents the decomposition of the basic waveforms: square and triangular have the same components but in different proportions, while the sawtooth has an extra component.

Received XXXXX; revised XXXXX; accepted XXXXX