Revista Vórtex | Vortex Music Journal | ISSN 2317–9937 | http://vortex.unespar.edu.br/
D.O.I.: https://doi.org/10.33871/23179937.2022.10.1.5

Schaeffer's *Solfège*, Percussion, Audio Descriptors: Towards an Interactive Musical System

Sérgio Freire, José Henrique Padovani, Caio Costa Campos

Universidade Federal de Minas Gerais | Brazil

Resumo: A tipo-morfologia de Pierre Schaeffer (1966) propõe sete critérios de percepção musical para a identificação e qualificação de objetos sonoros, que formam a base de seu solfejo. Este solfejo se aplica bem a contextos nos quais a altura não é a dimensão musical dominante. Baseado em similaridades entre a prática da escuta reduzida e a utilização de descritores de áudio de baixo nível, apresentamos a versão inicial de um programa em tempo real no qual esses descritores são aplicados na qualificação de sons percussivos. O artigo descreve as ferramentas e estratégias usadas para abordar os diferentes critérios: extratores de envoltórias com diferentes tamanhos de janelas e filtragens; detecção de transientes e modulação de amplitude; extração e contagem de componentes espectrais; estimação dissonância intrínseca e distribuição espectral; dentre outras. Os dados extraídos passam por uma análise estatística simples, produzindo valores escalares associados a cada objeto identificado. Finalmente, apresentamos uma variedade de exemplos.

Palavras-chave: solfejo de Schaeffer, percussão, descritores de áudio, sistemas interativos

Abstract: Pierre Schaeffer's typomorphology (1966) proposes seven criteria of musical perception for the identification and qualification of sound objects, which form the basis of his musical theory. This Solfège fits well into contexts where pitch is not the dominant dimension. Relying on similarities between the practice of reduced listening and the utilization of low-level audio descriptors, we present the first version of a real-time setup in which these descriptors are applied to qualify percussive sounds. The paper describes the tools and strategies used for addressing different criteria: envelope followers with different window sizes and filtering; detection of transients and amplitude modulations; extraction and counting of spectral components; estimation of intrinsic dissonance and spectral distribution; among others. The extracted data is subjected to simple statistical analysis, producing scalar values associated with each segmented object. Finally, we present a variety of examples.

Keywords: Schaeffer's Solfège, Percussion, Audio Descriptors, Interactive Systems.

Received on: September 29th, 2021. Approved on: October 27th, 2021. Online available on: April 30th, 2022. Presented at the 18th Brazilian Symposium on Computer Music (2021). Guest editors: Dr. Ivan Eiji Simurra (UFAC), Dr. Flávio Schiavoni (UFSJ), Dr. Cristiano Figueiró (UFBA), Dra. Isabel Nogueira (UFRGS) e Dr. Tiago Tavares (Unicamp). Creative Commons CC-BY-NC.

The association of audio descriptors with the Schaefferian theory is not rare in recent literature. One can find it in analytical contexts (VALLE, 2015), in the association of acoustic data and subjective labels (GODØY, 2021; BERNARDES; DAVIES; GUEDES, 2015; RICARD, 2004), and automatic sound indexing (PEETERS; DERUTY, 2010). Solomon (SOLOMON, 2016), although not using audio descriptors and not explicitly using this theory, presents a conceptual approach for the classification and association of percussive sounds similar to the one intended here.

In the last years, we developed a real-time setup in *Max*[1] to qualify percussive sounds using the Solfège's perceptual criteria, aiming at its application in interactive situations. Our approach uses time and frequency domain representations to tackle the diversity of criteria and sounds. In this paper, we present its first version, tested with diverse sound types. The intention was to build a general framework upon which we could later refine some specific tools. The background assumption is that the practice of reduced listening has similarities with the application of low-level audio descriptors.

The text organizes as follows. First, we present the criteria for musical perception that constitute the Schaefferian *solfège* and discuss the relationship of *reduced listening* and audio descriptors. In the second section, we present the segmentation and preprocessing procedures used in the setup. The description of the implemented tools comes next. Then we discuss the intended correspondences between the criteria and the descriptors. Two sections with examples follow. One set of examples is dedicated to each criterium separately, using a varied database. Two further examples focus on performances on specific instruments, exploring a limited set of pertinent criteria able to identify and classify the different sounds explored in each case. Finally, we present a discussion of the results already achieved and the plans for future work.

1. Schaeffer's Solfège and Perceptive Criteria

Technological mediation tools and practices used since the mid-twentieth century in listening, manipulation, and sound creation processes demanded composers, theorists, and analysts to

[1] https://cycling74.com/products/max

elaborate new concepts and methodologies to describe sonic processes. Pierre Schaeffer's *Treatise on Musical Objects* (TOM) (SCHAEFFER, 1966, 2017) is a theoretical landmark in this context. By inverting the traditional notion of *solfège* – associated with the practice of singing intervals and scalar excerpts through the *solmization* of musical notes – Schaeffer proposed, in the sixth chapter of the TOM, a "generalized *solfège*" that could address the complexity of sonic phenomena that only became evident through the technical tools of the electroacoustic studio.

TABLE 1 – Schaeffer's descriptions of typomorphological criteria and the related perceptual fields

criterion	**description**	**perceptual fields**
mass	"…quality through which sound installs itself (in a somewhat a priori fashion) in the pitch field." (p. 412)	pitches
dynamic	"…the variation in intensity of this sound in the course of its duration" (p. 33); "…its *energetic development*" [*évolution énergétique*] (p. 174)	intensities durations
harmonic timbre	"…more or less diffuse halo and more generally the secondary qualities that seem to be associated with mass and enable us to describe it." (p. 412)	pitches
melodic profile	"Neumes, although intended to represent variations in a specific source (the voice), can provide us with a model. (…) Type of sounds deliberately varied in tessitura" (p. 458)	pitches intensities durations
mass profile	"The *mass profile* is made up of all the (perceived) intensities of the various components of the spectrum of a sound." (p. 433)	pitches intensities durations
grain	"…a microstructure, generally due to sustainment from a bow, a reed, or even a drum roll. This property of *sound matter* reminds us of the *grain* of a textile or a mineral. (…) We find ourselves in a zone where two sensations from the same phenomenon [bassoon reed] merge: the perception of pitch from the beats, and the perception of beats from differentiation of the impacts" (p. 437)	pitches intensities durations
allure	"…the more or less regular oscillations that are its hallmark also cause variations in pitch (vibrato in stringed instruments, singers, etc.) and harmonic timbre. We could say that allure is made up of many factors (…), the most important of which are associated with the dynamic and pitch of sounds." (p. 438)	pitches intensities durations

Source: SCHAEFFER (2017, p. 412; 33; 174; 412; 458; 433; 437; 438)

To furnish means to his proposal of describing perceptual aspects of sound phenomena – despite any referential or causal events such as physical or instrumental factors that may have generated them – Schaeffer proposes, borrowing the Husserlian concept of *epochè*, a *reduced listening*. By this term, the author refers to a conscious attempt to scrutinize the attributes of sound objects by taking into account the three-dimensional perceptual space of *pitches*, *durations*, and *intensities*. To enable a more detailed description of sounds in these dimensions, Schaeffer proposes seven typomorphological criteria: *mass*, *dynamic*, *harmonic timbre*, *melodic profile*, *mass profile*, *grain*, and *allure*.

The seven typomorphological criteria proposed by Schaeffer allow him to build a methodological framework to evaluate audible characteristics of sound objects with the help of categories like *types*, *classes*, *genres*, and *species*. The TOM summarizes this method in the TARSOM (SCHAEFFER, 1966, p. 584–587, 2017, p. 464–467), a *Summary Diagram* that outlines the analytical concepts developed by the author. Thus, while the first three columns of this table – *types*, *classes*, and *genres* – relate the criteria (7 rows), respectively, to *typology*, *morphology*, and *characterology*, the following six seek to draw connections between these seven morphological criteria and the perceptual dimensions of *pitches*, *intensities*, and *durations*, employing the *site/calibre* binomial (two columns per perceptual field).

As Michel Chion clarifies (section 25, on the perceptual field, in CHION, 1983), each criterion has a more evident relationship with one or more of these perceptual fields. In Table 1, we present the morphological criteria proposed by Schaeffer. In the description column of the table, we have included excerpts from the English translation of TOM where Pierre Schaeffer outlines general observations about which aspects of sonic phenomena are addressed by each criterion.

More than half a century after the publication of the TOM, one can question the theoretical consistency and the pragmatic objectivity of reduced listening. For instance, Di Scipio remarks that the concept of *reduced listening* is technologically circumscribed, ignoring the very audible traces of electroacoustic tools that enable us to focus on the ‘sound itself’ (DI SCIPIO, 2015). Soddel, on the other hand, emphasizes the subjective processes of listening that cannot be ignored by any alleged “suspension of judgment” advocated by the Husserlian *époche* and by the Schaefferian *reduced listening* (SODDELL, 2020):

> ...my compositional practice instead considers how the reduced space of acousmatic sound might become a mirror into the mind, where the psychological perception of the abstracted sound material can reflect something about the listening self, and its interaction with the external world. (SODDELL, 2020, p. 347)

Despite these legitimate objections, the relevance of the listening attitude proposed by Schaeffer lies in the fact that his project of a "*generalized solfège*" has been successful in providing a rich theoretical framework that makes it possible to describe different features, behaviors, and qualities of sound objects according to morphological criteria and perceptual dimensions. By providing conceptual tools and a methodological outline to understand aspects of sound phenomena until then little systematized, it can be said that Schaeffer makes possible the formulation of a new approach to the understanding of the sound universe that overcomes the insufficiencies of both the traditional music theory and of a purely physical and acoustic approach to sound processes.

In this sense, it is important to underline the nature of the work proposed here, especially considering Schaeffer's well-founded warnings regarding the differences between the study of sound objects using perceptual-sensory criteria, on the one hand, and physical-acoustic analysis of audio signals, on the other hand. Despite the differences between perceptual and signal-based evaluation, description, and categorization of sounds, our work is motivated by a common trait of audio descriptors and the Schaefferian solfège: both focus on intrinsic qualities of sound phenomena, seeking to discriminate particular characteristics based on specific criteria, dimensions, or parameters.

It is possible to characterize a large set of computational audio tools as responses to the demands and challenges of a research/artistic field named *machine listening*. In this work, we associate the core intention of *reduced listening* (the perception of the form and matter of sounds, regardless of their meaning or source) with the employment of low-level audio descriptors, which intend to highlight similar features. In different sections of the TOM, Schaeffer expresses his skepticism concerning a physical-acoustic explanation of the auditory experience. Even with the development of new tools in the last 60 years, we believe this distance still prevails. Nevertheless, Schaeffer even recommends using technological aids for some specific tasks:

> It is perhaps disconcerting to see us, after so many warnings, recommending the use of the bathygraph and the Sonagraph to describe a piece of music. We have taken the precaution of pointing out in the third part of this work the usefulness and the limitations of this: while machines work in accordance with their scientific logic, the musician's activity is divided between sound and the musical. As soon as notation and, even more, the score come into the equation in electronic musics, these three levels of analysis are very clearly needed. (...) On the physical level the bathygraph and the Sonagraph give two graphs of the *signal* in *real-time*: its projection on the dynamic and the harmonic plane. Of course, these lines are not very intelligible because perceptions of sound differ so much (by

> anamorphosis) from the signal on the printout. But for giving a rough organization of events these two printouts will save hundreds of hours of painstaking, often impossible and probably premature, work. It is a silent but precise temporal map. (SCHAEFFER, 2017, p. 556–557)

Thus, our research does not aim to make a rough equivalence between the Schaefferian criteria and digital signal analysis processes related to audio descriptors. Rather, it has the practical purpose of bringing together Schaeffer's elaborate theoretical-methodological contributions and the real-time digital signal processing and analysis technologies that have only recently become available in artistic and academic practices related to computer music and interactive music systems.

2. Pre-processing and Segmentation

Defining the adequate sound portion to be analyzed in a real-time setup is crucial for working with reliable data. In our program, the inputs are audio streams delivered by microphones, pickups, or mixers, presenting diverse background noise and dynamic ranges. It is not difficult to control background noise by observing the input amplitude during a “silent” period. On the other hand, the dynamic range poses some challenges, like the leakage from other sources (including sounds coming from loudspeakers). We decided to set a maximal dynamic range of 40 dB, which can cover the range of the majority of instruments but may cut off earlier some long resonances (GIESELER; LOMBARDI; WEYER, 1985). In the development phase, we decided to use a set of pre-recorded sounds, which offers not only variety but also repeatability, two relevant factors for building and refining tools. The sound selection depicted in Table 3 was based on Schaefferian typology. These sounds become real-time inputs to the setup, running with a sampling frequency of 48 kHz. We also present examples with specific instruments: a series of contrasting strokes on a tom-tom, and a rhythmic pattern played on a pandeiro[2].

Portions to be analyzed are segmented between onset and offset points. A new segmentation clue may occur before the offset; in these cases, this clue determines the offset of the last event and the onset of the present one, which is qualified as “slurred”. We can also "force" an offset, in cases

[2] The sound files used in this study are available in the following link: https://musica.ufmg.br/lapis/?page_id=1188

where the most relevant information is concentred in the attack phase, such as percussion-resonance types. For this task, we may set a delay interval after the onset, or a decay threshold.

The detection of onsets and offsets relies on the comparison of a dynamic envelope with two thresholds, 6 dB and 3 dB, respectively, above the background noise level. This envelope is an RMS curve estimated with a short window size – 256 points – and a hop size of 64. We will refer to this curve as *rms256:4,* and use a similar notation for other envelopes. The implementation uses [gen~] routines, which employ native audio signal processing and offer more efficiency and precision. A low-pass filter (a single one-pole filter, with a -6 dB per octave attenuation) smoothens these envelopes, and we use different cutoff frequencies depending on the purpose. The detection of onsets and offsets employs a cutoff frequency of 4 Hz. The same setting applies to the detection of the end of an attack (in this case, the input signal may pass through a filter before the estimation of the envelope). Routines dedicated to attack profiles and iterative grains use the same envelope with a cutoff frequency of 30 Hz. The attack profiles are expressed by a control-rate version of this curve. We also use an *rms2048:4* curve, low-pass filtered at 10Hz, for the global dynamic envelope and the detection of allures. Other processes use spectral peaks values estimated by the [sigmund~] object (PUCKETTE; APEL; ZICARELLI, 1998), using the same window and hop sizes.

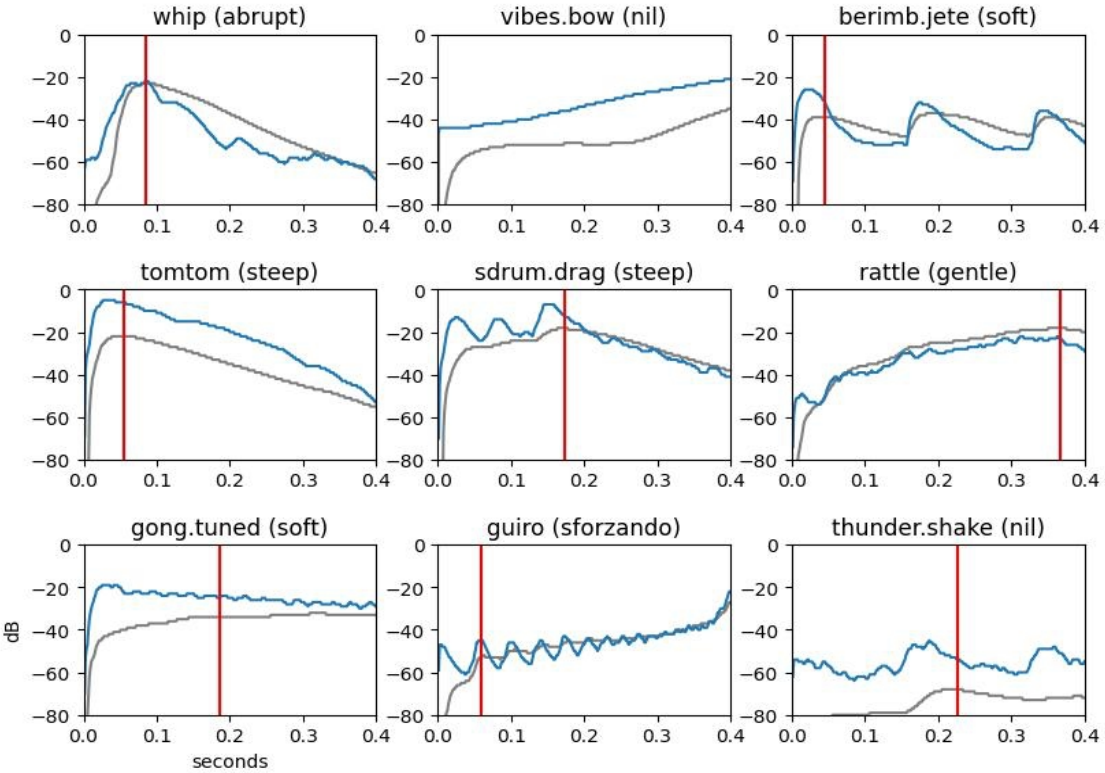
3. Implemented Tools

This section presents the tools developed so far, divided into different sub-sections. We use the usual division between time and frequency domain representations for the tools' classification. Most of them generate time-varying curves; a few deliver instantaneous information (onset, offset, end of the first plateau, peak of an allure, presence of an iterative grain). The curves go through simple statistical analysis just after the offset. There is also a dedicated sub-section for profiles and another for the implemented visualization tools.

3.1 Time Domain Low-level Descriptors

Duration is a simple attribute, whose value is the time interval (in ms) between onset and offset. The estimation of onsets and offsets was described in Section 2. The dynamic profile is the portion of the *rms2048:4* curve between onset and offset. The dynamic level is a simple feature, represented by the mean value and standard deviation of the same curve. For most sounds produced by a sharp attack followed by a resonance, the perception of loudness is defined by its initial portion, regardless of the duration of the resonance. Therefore, a simple mean value of the envelope of the entire sound would be meaningless. In such cases, it is advisable to use a forced offset. Even if we try to improve the correspondence between dynamic levels and perceptual attributes by observing the total duration and spectral region, it is necessary to remember Schaeffer's words: "For sounds with unremarkable mass and profile this dynamic field is almost unknown." (SCHAEFFER, 2017, p. 432)

FIGURE 1 – Attack profiles of nine sounds. The blue and gray curves derive from rms256:4 estimations; the latter is calculated from a filtered audio. The red line indicates the first plateau. The gender of the attack profiles are indicated between parentheses. For comparison purposes, all curves use the same set of parameters.



Schaeffer has stressed the importance of the attack for some sound typologies and the perception of durations. In the case of percussive sounds, this is a primordial characteristic. Therefore, we prefer to analyze the entire attack profile, which may last up to 400 ms, instead of stopping at the point usually called the "end of attack" (PEETERS, G. *et al.*, 2011). For the same reason, this point will be called the "attack first plateau". We define the "first plateau" as the moment when the derivative of the low-pass filtered audio-rate *rms256:4* curve from the (possibly filtered) input audio stream comes near (or cross) to zero, just after having surpassed a predetermined positive threshold. We call the ratio between the amplitude difference and the time interval that occurs between the onset and the first plateau as **FPSlope**. Since we prefer to consider multiple fast strokes (such as flans, drags, ricochets) as belonging to the same profile, a value of 150 ms stays as the reattack threshold. Depending on the settings (filtering and thresholds), the algorithm may not detect soft attacks[3]. On the other hand, some iterative sustainments are considered single allured sound objects, even when the distance between peaks exceeds the given threshold.

The attack profile curve also provides worthy data to complete analysis of short sounds (up to 400 ms) – or sounds characterized mainly by this portion – since it has a time resolution eight times greater than the remaining descriptors. Figure 1 depicts the attack profiles of nine percussive sounds with varied characteristics.

3.1.1 *Allures*

As practically any sound event with mechanical origins presents small fluctuations in its envelope, it is necessary to define a minimum threshold for the occurrence of a noticeable allure. In our setup, this threshold is a value expressed in dB (3 dB by default). The detection of allures and related features uses a control-rate version of the *rms2048:4* curve of the input stream, expressed in dBFS. The sign change of the derivative of the signal (crossing a tiny region around zero) is indicative of a possible peak or trough, which may be validated if the difference between the present peak and the last trough (and vice-versa) exceeds the chosen threshold. The estimated descriptors are: (1)

[3] These parameters (filtering, reattack time, and sharpness) help to redefine the fluids limits between the *context* ("whether the criteria are artificially put into a structure...") and the *contexture* ("...or naturally form a structure") of percussive sounds in diverse situations. (SCHAEFFER, 2017, p. 402)

number of occurrences; (2) mean value and standard deviation of the (a) difference of intensity between peaks and troughs; (b) time interval between successive peaks; (c) proportion between peak/trough and trough/peak intervals (symmetry); and (d) maximal value of the derivative in each inflection (spikiness). For a rough estimation of the distribution of peaks through the duration of the entire sound, their temporal centroid and spread are also calculated (in a time series filled with zeros, we insert the value 1 for each detected peak).

We have chosen to detect allures by inspection of the temporal envelope, since it is mostly related to the sustainment of the form of the sound (on the other hand, the grain criterium relates to the concrete pole of the sustainment of mass). But a similar analysis should be also effected on other descriptors curves, following Schaeffer's argument: "Allures, as we can see, are more like dynamic than mass criteria. Nevertheless, allure is not only a dynamic criterion; the more or less regular oscillations that are its hallmark also cause variations in pitch (vibrato in stringed instruments, singers, etc.) and harmonic timbre." (SCHAEFFER, 2017, p. 438) Allure is thought to be the main perceptive criteria when deciding over "the energetic agent's way of being" (*idem*): mechanical, living, natural.

3.1.2 Grains

We use two different algorithms for the estimation of the presence and quality of grains in the audio stream, one linked to iteration, the other to resonance and friction types. These last two types are treated jointly under the term tiny grains. For iterative grains, we use the derivative of an *rms256:4* audio-rate curve low-pass filtered at 30 Hz. This signal goes through a Schmitt trigger (with thresholds -0.01 and 0.01). When its value goes below the lower limit, there is an indication of a possible grain, which is confirmed if the time interval between two occurrences has a value below 75 ms. The grain amplitude correlates with the difference between the peaks and troughs of the RMS curve in the same time interval. The estimated descriptors are (1) number of iterative grains; (2) instants of occurrence; (3) mean value and standard deviation of (a) grain amplitudes; (b) time interval between successive peaks.

The estimation of tiny grains is done on the proper audio stream, or better, on its derivative. We assume that a granular signal will change directions more often than a smooth one. Although it may be objected that signals with higher frequencies will also bend more often–and with a larger difference between samples, due to the sampling resolution–, our experience has shown that in the realm of percussive sounds the granular influence is stronger than the register[4]. This descriptor generates two curves, one with the number of grains in every 512 samples, another with the mean value of the amplitudes of the detected bends[5]. An alternative would be the use of the spectral modeling synthesis (SMS), an algorithm proposed by Serra and Smith (SERRA; SMITH, 1990) to separate between deterministic and residual parts of a signal, using a frequency domain representation. However, as it presupposes the existence of a fundamental frequency, it does not adequately apply to most percussive sounds.

3.2 Frequency Domain Low-level Descriptors

This set of descriptors employs the spectral peaks estimation algorithm used by the [sigmund~] object, with the following parameters: analysis size of 2048 points; hop size of 512 points; 20 peaks; outputs (peaks, envelope, and pitch). These outputs are also controlled by the onset and offset descriptors described above and have a refresh rate of 10.67 ms with a sampling frequency of 48 kHz. For each frame analysis, we calculate the following descriptors.

Pct50 and **pct80** (number of peaks for energy percentiles 50% and 80%). According to Parseval's theorem, the energy of a time signal equals the sum of the energy of the absolute values of its frequency components. Applying this principle to [sigmund~] outputs (peaks and envelope), it is possible to estimate the number of peaks needed to obtain the 50% (-3 dB) and the 80% (-1 dB) energy percentiles. The descriptor outputs two curves with the estimated number of peaks. For sounds with a more continuous spectral distribution, 20 peaks may not reach the chosen percentiles. In such cases, the output will be 20 peaks, and further information is obtained through the next descriptor.

[4] A few examples: sine waves with frequency of 100 Hz (2 grains), 1000 Hz (21 grains), 10 kHz (213 grains), pink noise (ca. 300 grains) and white noise (ca. 350 grains).

[5] In Figure 3 it is possible to observe that this descriptor may work well as an auxiliary means for onset detection in the pandeiro case.

20P/total (percentage of sound energy represented by up to 20 peaks) estimates the amount of energy represented by the set of peaks calculated for each analysis frame and ranges between 0 and 1. The descriptor, along with the last one, helps differentiate sounds between the classes tonic and node (see Section 5).

MPP (most prominent peak) is a very simple descriptor, represented by the curve formed by the values (expressed in Midicents) of the peaks with the largest amplitude in each analysis frame.

Estimated fundamental frequency. The object [sigmund˜] outputs a value in Midicents for frames considered to bear a fundamental frequency and the value -1500 for unpitched frames. Our descriptor outputs a scalar (ratio of pitched to total frames– unpitched/total), a curve with all numbers, in which -1500 is substituted by 1, another curve with only the pitched values, and a third with 1 for pitched, and 0 for unpitched frames[6].

Intrinsic dissonance. The estimation of the intrinsic dissonance uses an implementation of the algorithm developed by Sethares (2005), using the frequency and amplitude values delivered by the [sigmund~] analysis. As he prescribes the use of SPL pressure values, we used 0.00001 for the minimal audible reference.

SC (spectral centroid). We use a [gen˜] routine delivered with the *Max* program since its version 6 for the estimation of the spectral centroid. Instead of using a nominal value in Hz, we use values in Midicents, which define a scale ranging from 15.5 to 155 in the audible range.

Δ **peaks** (interval between the lowest and highest peaks). For each frame, we calculate the difference between the highest and the lowest estimated peaks, which is also expressed in Midicents.

Spectral region. The starting point for the definition of a region is a rough division of the audible spectrum in three ranges. The first three octaves (20–160 Hz) define the low range, the four intermediate octaves (160–2,560 Hz) the medium range, and the last three octaves (2,560–20,000 Hz) the high range. We estimate the energy carried by the peaks in each analysis frame for each of these ranges. If none of these ranges hold 40% or more of the total energy, the sound frame is classified as wide-band, labeled as (7). Otherwise, any range with more than 40% of the total energy contributes to qualifying one of the six spectral combinations: (1) Low, (2) Low/Medium, (3) Medium, (4)

[6] To complement the sinusoidal model used by [sigmund˜], we intend to implement a pitchness descriptor using the autocorrelation function, following the advice given by Heller: "Beware of any theory of pitch perception that entirely leaves out autocorrelation." (HELLER, 2013, p. 450)

Medium/High, (5) High, (6) Low/High. When exploring a more restricted spectral distribution, it is advisable–and easy–to set different frequencies for the limits between each region. It is also possible to visualize the amplitude curves of the three spectral ranges.

3.3 Profiles

Schaeffer uses the term profile in many different contexts and situations, including the name of two of the main perceptual criteria. With a few exceptions, this term is associated with a temporal variation of at least one sonic feature. He usually connects this term with a sound object bearing an optimal duration for the memorization, as in the case of dynamic profiles. Moreover, this temporal variation is expected not to be perceived as cyclic, in which case he prefers the term allure.

Although any percussive sound presents some degree of temporal variation in diverse aspects, the seven Schaefferian criteria relate to first-order perceptions and presuppose pregnant features when referring to melodic or mass profiles. The melodic profile is a clear trajectory in the tessitura, pitched or not; the mass profile is a clear transition between different mass classes, such as starting with a node and finishing with a clear tone. It is easier to find instances of melodic than mass profiles in the percussive practice, as different types of glissando demonstrate. Although we could consider, f. i., the bowed cymbal as a mass profile, the node phase may represent just an unavoidable initial transient phase to many listeners. Second-order profiles, such as the speeding up of an allure, the gradual granulation of a sound, or even a "profiled" profile, would count as internal variations of the corresponding criterium.

Schaeffer proposes the crossing of two features for the analysis of variations: density of information (weak, medium, strong) and type of facture (fluctuation, evolution, modulation). Evolution, "a progressive development", is the most common type we have found. There may happen some confusion between the idea of melodic (or mass) fluctuation – "an imperfection in a desired stability" (SCHAEFFER, 2017, p. 453) – and the occurrence of an irregular allure. In such cases, the following remark would be helpful:

> We preferred, however, to keep the theory of allure and grain for a chapter of their own, thinking that if mass and dynamic profile come from the abstract pole of the object—that is, its effects—then grain and allure on the contrary are perceptions that reveal the *concrete* pole of objects, closely linked to the energetic history that relates the origin of *every moment of the sound*. (SCHAEFFER, 2017, p. 436–437)

3.4 Statistical Analysis

The time series generated by most descriptors are subjected to simple statistical analysis just after the offset. We adapted the algorithms given in (PEETERS, G. *et al.*, 2011), and have chosen the following scalar descriptors: mean value and standard deviation; temporal centroid and spread (normalized by the total duration); skewness; kurtosis; crest; flatness. Additionally, the curves may be searched for the presence of allures, or some kind of cyclical behaviour.

The flowchart in Figure 2 depicts the main processes involved in the estimation of all low-level descriptors. The correlations with the Schaefferian perceptual attributes rely on these values.

3.5 Visualization

Our program in *Max* offers the possibility of flexibly displaying particular selections of curves (and markers for attacks, allures, and iterative grains). It permits observing different sets of descriptors for specific cases. There is also a dedicated display for attack profiles, and horizontal and vertical dimensions allow rescaling. The setup has been used not only with percussive sounds but also explored by other instrumentalists in the practice of non-usual sonorities. Figure 3 depicts part of the analysis of the pandeiro excerpt discussed in Section 6.

FIGURE 2 – Flowchart showing the main routes and procedures for the estimation of low-level audio descriptors.

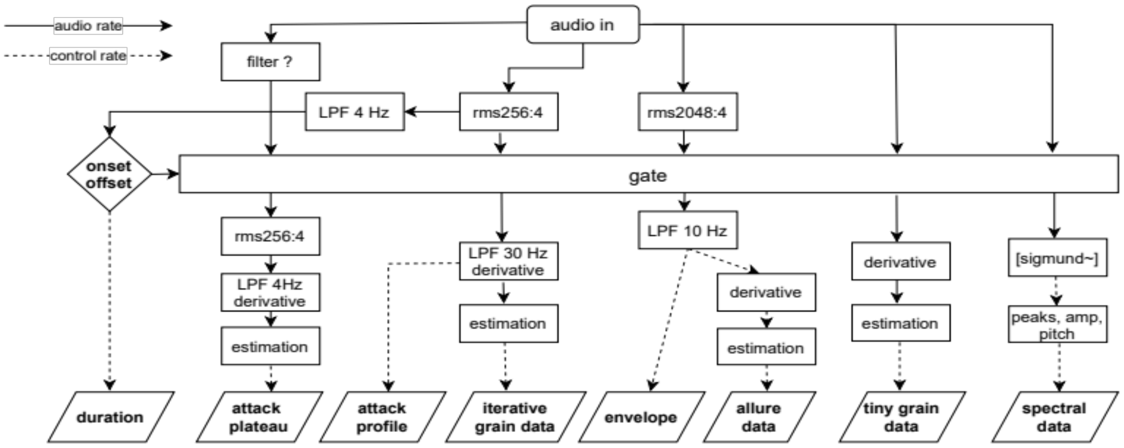
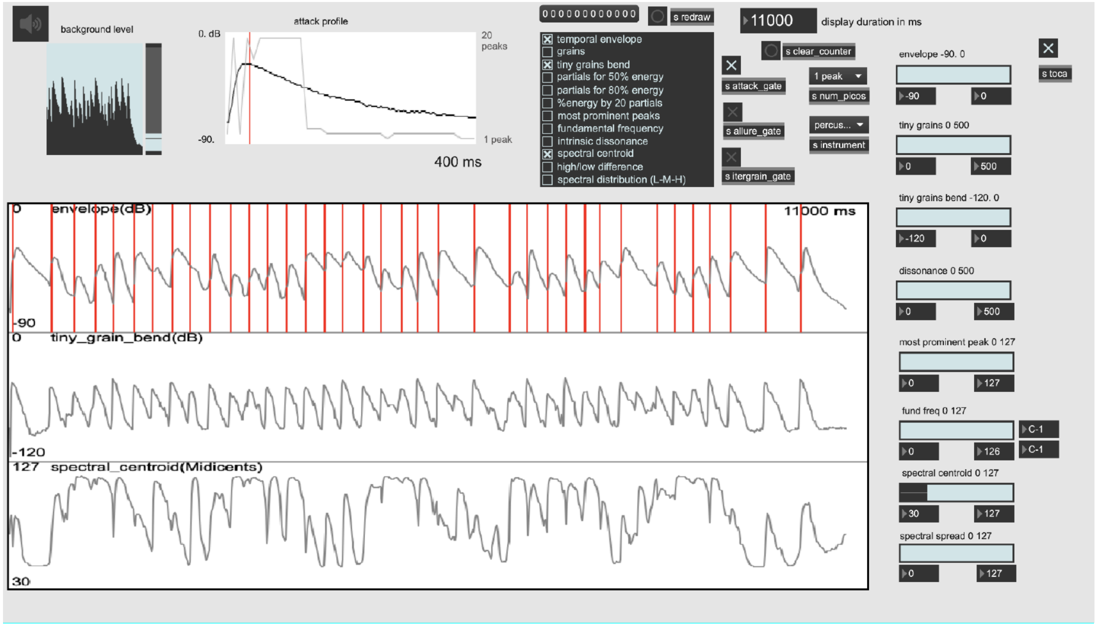


FIGURE 3 – Max window showing the visualization options and an example: the analysis of the pandeiro excerpt commented in Section 6. The red vertical lines indicate the estimated onsets.



☒ temporal envelope
□ grains
☒ tiny grains bend
□ partials for 50% energy
□ partials for 80% energy
□ %energy by 20 partials
□ most prominent peaks
□ fundamental frequency
□ intrinsic dissonance
☒ spectral centroid
□ high/low difference
□ spectral distribution (L-M-H)

4. Intended Correspondences

As stated earlier, our purpose is to find correlations between low-level descriptors and the criteria of musical perception defined by Schaeffer. In this study, we reduced the seven criteria to six, rearranging them as follows. Since in the realm of percussion (and of everyday sounds) tonic sounds are not the rule, we prefer to unify mass and harmonic timbre under one single category, relying on an observation made by Schaeffer, "considering them rather as connecting vessels, with the exception of certain specific examples..." (SCHAEFFER, 2017, p. 412). Melodic and mass profiles are joined by the same reasons. Although the same set of descriptors supports the qualification of sounds under the four mentioned criteria, we prefer to put the profiles in a dedicated category since they use different statistical values. On the other hand, due to the great variety of attack types, we chose to treat them not as genres belonging to the criterion dynamic but as a separated criterion.

Table 2 depicts the intended correlations between Schaefferian criteria (and their attributes expressed in *types*, *classes*, *genres*, or *species*) and the selected low-level descriptors. Note that a few estimated attributes function as descriptors for other categories, like the spectral regions for the dynamic levels or allures for the melodic/mass profiles. It is important to note that any sound object enacts at least three perceptual criteria: mass, dynamic (including the attack), and harmonic timbre. The other criteria may be relevant to characterize specific sounds, and their presence may affect the others. For example, allures may affect the melodic profile, grains may affect the mass and the mass profile, and percussion-resonance types condition the harmonic profile.

TABLE 2 – Intended correspondences between Schaeffer criteria and low-level audio descriptors.

Criterion	**Attributes**	**Audio Descriptors**
dynamic	duration: short, formed, long dynamic level: pp to ff dynamic forms: shock, resonance, profiles (5 classes), flat, nil	onset–offset rms curve statistics skewness of spectral centroid spectral region, attack genre
attack profile	genre: abrupt, solid, soft, gentle, stressed, nil	attack profile statistics iterative grain / allure data
mass / harmonic timbre	class: tonic, channeled, nodal group, node region: low, medium, high genre/species: full/hollow/narrow, rich/poor	spectral peaks data / statistics spectral centroid
melodic / mass profiles	density of information: weak, medium, strong type: fluctuation, evolution, modulation	spectral peaks data / statistics spectral centroid / allure data
grain	type: iterative, tiny (friction or resonant) density: rough, matt, smooth	iterative / tiny grain statistics
allure	intensity: weak, medium, strong genre: regular, progressive, irregular, etc.	allure data

5. Examples and Discussion

TABLE 3 – Selected sounds.

sound	description
tabla.gliss	single tabla stroke with glissando
whip	single whip attack
tamb.slap	single tambourine hand slap
sdrum.nosnare	single snare drum stroke, without snare
chin.opera.gong	single chinese opera gong stroke
bassdrum	single bassdrum stroke
cymbal	single cymbal stroke
gong.tuned	single tuned gong stroke
guiro	single directional guiro rub
tomtom	single tom-tom stroke
sdrum.drag	snare drum drag, withsnare
ratchet	single ratchet swing
rattle	single directional rattle shake
tamb.tremolo	tambourine tremolo
berimb.jete	berimabau jete, multiple strokes
berimb.vib	single berimbau stroke, with vibrato
pand.rim.frict	pandeiro tremolo-like rim friction
sleighbells	multiple sleighbells shakes
thunder.shake	multiple thunder sheet shakes
rainstick	rainstick tip
slidewhistle	slide whistle blow with glissando
pand.skin.frict	single pandeiro skin friction
vibes.bow	single vibraphone key bow
cymbal.bow	single cymbal bow

For every input sound, our program generates real-time curves (or markers) for all descriptors and calculates the scalar values described in Section 3. These results are relatively numerous and probably present some degree of redundancy, not yet analyzed. For the sake of clarity, we discuss these results separated by criterion (or sub-criterion), using subsets of sounds and descriptors.

Our first example depicts the quantitative results (Table 4) for the different attack profiles illustrated in Figure 1. These shapes depend on the facture of the sound objects (single stroke, iteration, continuous excitation) and dynamic levels. Schaeffer defines seven genres of attack: *abrupt*, *steep*, *soft*, *flat*, *gentle*, *sforzando*, and *nil*. The whip sound has a short duration, low values for temporal centroid and spread, a positive skewness, and a high crest. All this data corresponds to an abrupt genre. The stroke on a tom-tom presents a steep profile. It has a short resonance following the attack. Its temporal centroid is similar to the whip, but with a larger spread and less pronounced values for skewness and crest. A reinforcement of the resonance follows a soft profile, such as with the tuned gong. Here, a longer duration, a slightly positive skewness, and a high value for flatness contribute to the characterization of the genre. We perceive the rattle profile as gentle, given the absence of an initial shock. Its sound production combines iterative and continuous energies, which correlates with a negative skewness, a small crest, and medium flatness.

TABLE 4 – Attack parameters for nine selected percussive sounds (the same from Figure 1), plus total duration and dynamic level.

sound	**FPSlope (dB/ms)**	**temp. centr.**	**temp. spread**	**skewness**	**kurtosis**	**crest**	**flatness**	**dur (ms)**	**DL (dB)**
whip	0.42	0.24	0.26	1.25	2.87	6.08	0.30	396	-46
vibes.bow	-	0.71	0.68	-0.35	0.40	2.84	0.70	2177	-26.5
berimb.jete	0.81	0.35	0.53	0.33	0.66	4.77	0.60	2019	-55.6
tomtom	1.03	0.24	1.18	0.17	0.09	3.60	0.43	487	-28.5
sdrum.drag	0.33	0.35	1.06	0.08	0.10	4.28	0.61	616	-31.8
rattle	0.13	0.67	0.67	-0.21	0.28	2.18	0.67	1151	-42
gong.tuned	0.24	0.42	1.13	0.08	0.11	1.96	0.92	9219	-42.6
guiro	0.63	0.72	0.39	-0.66	1.36	11.09	0.67	631	-42
thunder.shake	0.18	0.53	0.20	-0.19	3.82	2.80	0.86	15541	-33

Two sounds have a clear iterative or granular profile, the guiro rub, and the berimbau jeté. This last has a profile between steep and soft, due to the reinforcement of the resonance by repeated strokes. The guiro profile is sforzando, demonstrated by a high temporal centroid, a negative skewness, a high crest. Two sounds bear a nil profile, the bowed vibes, and the thunder shake. This fact is reflected by the high temporal centroid, negative skewness and small crest. There was no estimation of the first plateau during the first 400 ms for the bowed sound. Long sounds will rely less on their attack profile for their qualification. We also believe that the iterative/granular character should be a second-order qualifier for the attack profiles.

Schaeffer's dynamic classes consist of *shocks* (very short sounds), *anamorphoses* (forms determined by the attack and the following resonance), *profiles* (cresc., decresc., delta, hollow, mordent), and *lifeless*. In real-time applications, the setting of background noise levels may considerably alter some profiles. Short sounds with a sudden offset (cresc. and hollow) may also be mistaken for a delta class since we normally expect a decay phase (be it from the proper instrument, be it from the ambiance). The envelopes linked to the basic profile classes may be influenced by internal modulations due to the presence of allures, without losing their main shape. Table 5 shows quantitative data from the envelopes of seven selected sounds. Duration, temporal centroid and spread, skewness, and flatness values can differentiate between the classes described above. Since shock sounds are already fully characterized by the attack profiles (like the whip), we will concentrate on the remaining classes. Attack-resonant classes (gong) have a low value for temporal centroid and a positive skewness. The delta class (crescendo followed by decrescendo) has a medium temporal centroid (bowed vibes), while the crescendo sounds (snare drum roll) present a higher temporal centroid and negative skewness. Lifeless envelopes correlate with large flatness values, and low standard deviation in sounds with medium or long durations (cf. ratchet and tympanum roll). The crest value may indicate either a sharp attack or the presence of salient inflections in the middle of long sounds, as in the thunder sheet shake.

TABLE 5 – Values of low-level descriptors for the envelopes of seven selected percussive sounds.

sound	**dur (ms)**	**mean SD (dB)**	**TC / (spread)**	**skewness**	**kurtosis**	**crest**	**flatness**
whip	396	-46 ± 13	0.29 /0.1	2.37	14.84	4.64	0.37
ratchet	753	-14.5 ±7	0.51 / 1.05	-0.02	0.12	1.33	0.87
sdrum.roll.cresc	2568	-29.5 ± 8	0.59 / 0.71	-0.08	0.22	2.08	0.76
vibes.bow	4561	-27.5±12	0.37 / 1	0.1	0.1	2.44	0.52
gong.tuned	7408	-42.6±11	0.23 / 0.65	0.38	0.4	6.13	0.5
thunder.shake	15541	-33 ± 11	0.35 / 1.47	0.16	0.07	9.38	0.49
timp.roll	19667	-43 ± 4	0.41 / 1.03	0.08	0.13	2.27	0.91

Table 6 depicts allure data for six sounds, each of them with a different excitation pattern. As already exposed, we have chosen to interpret as allures (and not as new attacks) iterative sustainments with clearly differentiable impulses, as in the cases of the berimbau and sleigh bells. Time intervals with small standard deviation values are related to ordered or regular instances (like the berimbau), while the opposite points to higher irregularity (rainstick). The amplitudes indicate the depth of variation; for example, the tuned gong resonant allures are much softer than the iterative allures of the sleigh bells. Symmetry indicates the regularity of transitions between peaks and troughs (and vice-versa), and spikiness values point to the suddenness of variation.

TABLE 6– *Allure* values for six selected percussive sounds.

sound	dur (ms)	number	amp (dB)	Δ t (ms)	symmetry	spikiness
berimb.jete	2019	11	5.1 ± 0.8	162 ± 32	1.99 ± 0.94	4.93 ± 3.2
berimb.vib	4687	12	7.3 ± 2.4	349 ± 68	0.99 ± 0.48	1.91 ± 0.79
gong.tuned	9219	6	4.2 ± 1.2	1476 ± 485	1.05 ± 0.56	0.26 ± 11
sleighbells	10295	30	17.1 ± 4.6	333 ± 55	1.47 ± 1	5.34 ± 3.3
thunder.shake	15541	57	7.2 ± 3.8	261 ± 95	1.46 ± 1.2	2.97 ± 2.2
rainstick	17295	27	5.6 ± 3.1	647 ± 448	1.23 ± 1.1	1.77 ± 1.65

The estimation of grains for 10 sounds is depicted in Table 7. As expected, sounds with iterative sustainment present a large number of these grains. The exception is the bass drum, whose resonant grains, due to the slow rate, also fit into this category. The diverse values for size and duration also helps differentiating between iterative grains. Our tiny grain descriptor is dedicated to resonant and friction grains, and their mixture. A closer inspection of the number of tiny grains and their standard deviation furnish information about their temporal behavior. The large standard deviation, along with a high value of temporal centroid in the bass drum, indicates an increase of background noise at the end of the resonance. The bowed cymbal also displays a considerable standard deviation, but a temporal centroid below 0.5; in this case, the granular characteristic is more present at the beginning. Friction grains tend to have higher bend values than resonant ones, as displayed by the ratchet and cymbal data. On the other hand, the resonant characteristic seems to prevail in the bowed cymbal and vibraphone, despite the excitation mode.

TABLE 7 – Grain values for 10 selected percussive sounds.

sound	dur (ms)	iterative grains			tiny grains		
		number	size (dB)	dur (ms)	number	TC	bend (dB)
guiro	631	14	5 ± 3	22.5 ± 13	140 ± 25	0.49	−60.5
ratchet	753	12	15 ± 5	54.5 ± 8	173 ± 18	0.49	−31.0
rattle	1103	2	4.8	73.3	129 ± 18	0.48	−59
pand.rim.frict	1365	3	5.2 ± 1.5	27	218 ± 17	0.5	−29
whistle	1463	30	4.6 ± 2.7	37.5 ± 12	56 ± 11	0.48	−58
bassdrum	3671	96	3.4 ± 1.4	37.4 ± 17	190 ± 122	0.67	−86
cymbal.bow	4565	3	2.7 ± 0.4	34.7 ± 8	65 ± 34	0.44	−76
cymbal	4963	-	-	-	116 ± 48	0.57	−71
tamb.tremolo	7884	59	5.6 ± 1.8	49.5 ± 17	171 ± 11	0.49	−41.5
rainstick	17245	110	6.8 ± 3.3	44.3 ± 16.5	191 ± 32	0.49	−54

Seven classes of mass are defined by Schaeffer: *pure sound*, *tonic*, *tonic group*, *channeled*, *nodal group*, *node*, *white noise*. *Pure* and *tonic* sounds bear a clear pitch, and tonic groups indicate chord-like sonorities. *Node* is a filtered noise (or a dense spectral region), while *nodal group* is a set of *nodes*. In the middle of this classification we encounter the ambiguous *channeled* sound, sharing properties for pitched and unpitched classes. The most common classes in the percussive realm are the *tonic*, *channeled*, *node*, and *nodal group*, although this latter occurs more in combination than in a single object[7]. As stated above, the harmonic timbre is a complementary characteristic of the spectral perception, and Schaeffer points to oppositions like *hollow/full*, *rich/poor*, and *bright/matt*. Data related to mass and harmonic timbre is displayed in Tables 8 and 9. The ratio between the unpitched and total frames estimated for the entire sound object points to its tonic character. Low values indicate tonic sounds, like the bass drum, whistle, and the friction of a pandeiro's skin. Higher values of this descriptor, combined with small values of percentiles 50% and 80%, can qualify channeled sounds, like the snare drum (with no snare) and the tuned gong. A high unpitched/total ratio, along with a low value for the energy carried by the 20 first spectral peaks, characterizes nodes, as in the cases of the ratchet and rattle. Spectral centroid and region values are self-explaining.

The *intrinsic dissonance* values are more ambiguous. Although they may help differentiating between *tonic* and *non-tonic* sounds, this is not a univocal association since inharmonic spectral peaks only increment this value if they are close enough in frequency (see SETHARES, 2005). We try to approximate the perceptive attributes *full/hollow/narrow* with Δ **peaks**, **pct50**, **pct80** and **region**, and the *rich/poor* with the values estimated for **pct80** and **20P/total**. For instance, two sounds classified in the medium region may be contrasted through the attributes *hollow* (snare drum) and *narrow* (rattle).

[7] It is advisable to increase the number of spectral regions–or redefine their limits–for dealing with *nodal* groups.

TABLE 8 – Mass and harmonic timbre parameters (1) for 10 selected percussive sounds.

sound	dur	pct50	pct80	20P/total (ratio)	unpitched/total (ratio)
tabla.gliss	227	1.3 ± 0.5	6.6 ± 8.5	0.85 ± 0.12	0.24
sdrum.nosnare	560	1.1 ± 0.4	2.4 ± 3.7	0.96 ± 0.1	0.32
ratchet	753	19.8 ± 0.6	20 ± 0	0.42 ± 0.1	0.86
rattle	1103	8.4 ± 2.2	19.7 ± 1.3	0.71 ± 0.1	1.0
pand.skin.frict	1915	1 ± 0.2	1.5 ± 2.1	0.98 ± 0	0.07
slidewhistle	641	6.3 ± 7.7	14.4 ± 8.4	0.66 ± 0.5	0.13
chin.opera.gong	1911	2.5 ± 1.9	8.5 ± 5.9	0.9 ± 0.1	0.75
berimb.jete	2019	4.6 ± 5.4	10 ± 6.7	0.82 ± 0.2	0.94
bassdrum	3671	1.1 ± 1	1.2 ± 1.5	0.98 ± 0.1	0.04
gong.tuned	9219	1.3 ± 0.8	2.2 ± 1.6	0.98 ± 0.05	0.42

TABLE 9 – Mass and harmonic timbre parameters (2) for 10 selected percussive sounds.

sound	diss	MPP (mc)	Δ peaks (mc)	SC (mc)	region
tabla.gliss	37 ± 22.5	46.2 ± 10.4	69.9 ± 19.3	54 ± 7.6	1.8 ± 1
sdrum.nosnare	45.3 ± 21.9	61.2 ± 7.7	59.4 ± 15.6	65.5 ± 13.3	3
ratchet	122.5 ± 30.9	97.6 ± 9.5	30.7 ± 4.5	110.2 ± 3.6	6.5 ± 1.3
rattle	138.8 ± 42.6	95.5 ± 1.9	17.5 ± 9.1	101.2 ± 2.7	3
pand.skin.frict	43.2 ± 30.5	48 ± 1.6	70.2 ± 6.7	51.7 ± 6.7	1
slidewhistle	141.1 ± 66.8	89.8 ± 14.8	79.8 ± 20.5	95.3 ± 8.1	4.4 ± 1.8
chin.opera.gong	69.1 ± 15.1	73.2 ± 5.7	40.9 ± 12.6	82.8 ± 6.8	3
berimb.jete	24.8 ± 18.6	65.5 ± 13.9	80.8 ± 13.2	86.5 ± 8.5	3.5 ± 1.3
bassdrum	22.6 ± 23	27.8 ± 3.8	99.3 ± 21.6	30.5 ± 8.7	1 ± 0.3
gong.tuned	24.2 ± 17.8	61 ± 2.9	76.4 ± 31	68.2 ± 7.6	3 ± 0.2

Although our sound selection does not focus on melodic or mass variations, we can depict a few examples from there. A good starting point is to look for significant standard deviations in the estimated curves for the different descriptors, taking into account two warnings. The first is related to percussion-resonance sound types, in which temporal and harmonic envelopes closely correlate, and the high-frequency content decay is perceived as natural, not as an intended profile. The shape of the curve also deserves observation since it is possible to have substantial variation in the absence of a clear-cut profile (or a soft contour): the variations may also be stochastic, periodic, or concentrated in a short segment. The three melodic profile examples are the tabla glissando, the friction of a pandeiro skin, and the slide whistle. Observing the tabla data, we find considerable standard deviations for the MPP, spectral centroid, and fundamental frequency curves. Although derived from a percussion-resonance type, these curves are non-oscillating ascending ones–with negative skewness – in opposition to the temporal envelope. Similar reasoning may apply to the other two sounds.

6. Preliminary Applications

Here we discuss the results obtained for the differentiation of diverse sonorities played on two individual percussion instruments. We have recorded six different single strokes on a standard 12″ tom-tom, using diverse materials (soft and regular drumsticks, finger rimshot) and stroke positions (from standard to the rim). The musician played with the intention of a transition from heavy/dark to light/bright sounds. In this case, the main criteria used to differentiate between them are the attack profile, amplitude, and harmonic timbre.

Table 10 depicts values from the analysis of the attack profile and of the spectral centroid curve (with a forced offset of 400 ms after the onset). Although some of the values do not vary only in one direction, their combination allows a good differentiation between the six instances.

TABLE 10 – Selection of estimated descriptors for 6 different strokes on a tom-tom.

sound	**start (ms)**	**attack profile**						**spec. centr. (MC)**
		attack size (dB)	**flatness**	**peakness**	**kurtosis**	**temp. centr.**	**FPSlope**	
1	156	39	0.81	2.96	0.47	0.61	3.25	48.02
2	2044	47	0.77	2.79	0.36	0.71	1.86	49.52
3	4360	43	0.82	2.28	0.38	0.67	1.54	55.13
4	7013	36	0.90	2.20	0.71	0.48	1.08	58.58
5	9736	34	0.91	2.30	0.96	0.42	1.11	59.19
6	12005	28	0.93	2.06	1.59	0.32	0.90	64.48

We also recorded a rhythmic accompaniment pattern (from the Brazilian instrumental music choro) played on a pandeiro. It uses five different strokes, as depicted in the transcribed score (Figure 4). It is possible to find direct strokes (the fingers attack the membrane, types d and e) and passive ones (the instrument reaches the fingers or the wrist, types b and c). The employment of the thumb (type

a) adds low-frequency components to the omnipresent jingle sounds. Accents are also present. The criteria used to identify these diverse sonorities are attack profile, duration, mass (presence or absence of low-frequency components) and harmonic timbre. It is important to note that there is no univocal correspondence between a performing gesture and the produced sound since the latter depends also on the stroke position and dynamics. It is possible to identify typical sonorities linked to specific gestures, but it is also usual to find more ambiguous sonorities, especially in musical contexts–when we analyze phrases instead of isolated sounds.

FIGURE 4 – Transcription of a choro accompaniment pattern played on a pandeiro.

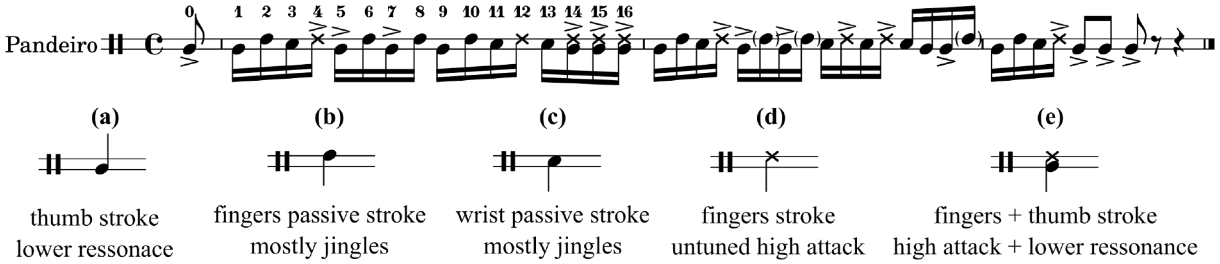


Table 11 depicts the first 17 (from 0 to 16) strokes in this pattern. The attack size values characterize well the performed accents, and the most prominent spectral peak values indicate sounds with a significant low-frequency component. The thumb generally performs these sounds, but not exclusively. They may also be produced by other fingers or by the resonance of the last stroke. The pure high sounds always count with large values for spectral centroid and the most prominent peak. Some sounds with low-frequency components have a considerable high value for the spectral centroid; in these cases, the resonance-causing stroke combines with a more intense excitation of the jingles. A machine learning algorithm applied to typical instances of each stroke should help their differentiation, besides indicating more ambiguous situations. It would also be easy to infer the rhythmic expression by calculating the inter-onset intervals (IOIs) and observing the accents.

TABLE 11 – Selection of descriptors for the 17 strokes of the pandeiro pattern.

stroke number	**stroke type**	**attack profile**				**spec. centr.**	**MPP**	**spec. region**	**Δ peaks (mc)**
		attack size(dB)	**flatness**	**temp. centr.**	**temp. spread**				
0	**a**	48	0.67	0.49	0.29	58.84	47.9	1.26 ± 1.22	83.40 ± 11.78
1	**a**	42	0.37	0.18	0.22	92.67	62	2.52 ± 2.50	81.13 ± 11.72
2	**b**	27	0.62	0.13	0.28	119.21	107.8	6.60 ± 0.80	45.88 ± 34.88
3	**c**	36	0.49	0.15	0.24	118.34	73.1	6.55 ± 1.34	83.07 ± 9.88
4	**d**	44	0.27	0.17	0.17	109.19	75.1	5.00 ± 2.77	75.63 ± 21.07
5	**a**	45	0.73	0.34	0.33	81.77	46.4	1.27 ± 1.25	80.07 ± 5.18
6	**b**	33	0.77	0.19	0.30	104.12	46.2	3.45 ± 2.95	82.00 ± 4.14
7	**a**	47	0.82	0.49	0.35	71.59	46.3	1.50 ± 1.66	80.22 ± 6.53
8	**b**	30	0.66	0.16	0.27	85.06	46.3	3.18 ± 2.89	87.67 ± 13.46
9	**a**	44	0.40	0.19	0.21	95.79	49.6	1.00 ± 0.00	79.20 ± 8.31
10	**b**	28	0.71	0.15	0.32	116.01	87.6	6.57 ± 1.31	72.45 ± 24.70
11	**c**	31	0.59	0.16	0.28	118.38	90.4	6.7 ± 0.71	47.75 ± 36.09
12	**d**	35	0.43	0.14	0.20	103.92	66.9	4.17 ± 2.94	77.26 ± 21.04
13	**c**	34	0.50	0.14	0.27	111.82	90.9	5.91 ± 2.31	79.02 ± 16.63
14	**e**	44	0.83	0.41	0.35	70.32	46.4	1.00 ± 0.00	79.00 ± 4.74
15	**e**	45	0.85	0.42	0.35	67.77	46.3	1.00 ± 0.00	79.04 ± 9.16
16	**e**	45	0.76	0.39	0.32	72.37	46.4	1.00 ± 0.00	79.96 ± 9.00

7. Final Remarks

We presented a general overview of the setup, including the theoretical background and the implemented tools. The examples showed that this set of audio descriptors correlates with the Schaefferian criteria applied to percussive sounds and also pointed to the possibility of exploring this analytical framework in real-time interactive situations. Using Rowe's (ROWE, 1993) concept of three stages for the processing chain in interactive musical systems, we have planned how to deal with the first two: sensing and processing. Concerning the sensing stage, we intend to use close miking and contact pickups for audio acquisition. This option minimizes two common problems in signal processing and interactive performance: the blurring effect of reverberation in analyses based on time slices, the capture of loudspeaker sounds as not desired inputs. A set of pedals and some portable movement sensors complete the planned hardware. The processing phase has to be completed by two additional tools: a more flexible way to select and map data from the estimated descriptors (and statistical results) to different routines in the program and the training of a machine learning process (at present we use Wekinator[8]) to identify contrasting sound criteria or types.

Due to the pandemic, the response phase–as in most interactive systems, founded on electroacoustic and digital technologies–is still waiting for the opportunity for a more continuous and consistent practice with musicians (performers, composers, sound artists), from which many different paths may surge: use of distinct sounds as triggers, use of phrases as musical material to be analyzed and varied, homogeneity (or heterogeneity) of the sound palette, among many others. We hope that soon we will be able to put this setup to use and then refine it according to different usage demands.

ACKNOWLEDGMENT

This work has been supported by the Brazilian research agency CNPq (National Council for Scientific and Technological Development).

REFERENCES

BERNARDES, G.; DAVIES, M.; GUEDES, C. A Pure Data Spectro-Morphological Analysis Toolkit for Sound-Based Composition. In: EAW2015 - INTERNATIONAL CONGRESS FOR ELECTROACOUSTIC MUSIC - ELECTROACOUSTIC WINDS 2015, 2015, Aveiro. *Proceedings*... Aveiro: Proceedings of the eaw2015, 2015. p. 31–38.

CHION, Michel. *Guide des Objets Sonores*. Paris: Buchet/Chastel, 1983.

DI SCIPIO, Agostino. The Politics of Sound and the Biopolitics of Music: Weaving Together Sound-Making, Irreducible Listening, and the Physical and Cultural Environment. *Organised Sound*, v. 20, n. 3, p. 278–289, dec. 2015.

[8] http://www.wekinator.org/

GIESELER, W.; LOMBARDI, L.; WEYER, R. *Instrumentation in der Musik des 20. Jahrhunderts: Akustik, Instrumente, Zusammenwirken*. Celle: Moeck Verlag, 1985.

GODØY, Rolf Inge. Perceiving Sound Objects in the Musique Concrète. *Frontiers in Psychology*, v. 12, 2021. Available at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.672949/full>. Access: 7 jun. 2021.

HELLER, Eric Johnson. *Why you hear what you hear: an experiential approach to sound, music, and psychoacoustics*. 2013.

PEETERS, G. *et al.* The Timbre Toolbox: Extracting Audio Descriptors from Musical Signals. *The Journal of the Acoustical Society of America*, v. 130, n. 5, p. 2902–2916, nov. 2011.

PEETERS, Geoffroy; DERUTY, Emmanuel. Sound Indexing Using Morphological Description. *IEEE Transactions on Audio, Speech, and Language Processing*, v. 18, n. 3, p. 675–687, mar. 2010.

PUCKETTE, Miller S; APEL, Theodore; ZICARELLI, David D. Real-time Audio Analysis Tools for Pd and MSP. In: INTERNATIONAL COMPUTER MUSIC CONFERENCE, 1998, San Francisco. *Proceedings*... San Francisco, 1998. p. 109–112.

RICARD, Julien. *Towards Computational Morphological Description of Sound*. 2004. Doctorate – Universitat Pompeu Fabra, Barcelona, 2004.

ROWE, Robert. *Interactive music systems: machine listening and composing*. Cambridge: MIT press, 1993. Available at: <http://dl.acm.org/citation.cfm?id=530519>. Access: 16 jun. 2014.

SCHAEFFER, Pierre. *Traité des Objets Musicaux*. Paris: Éditions du Seuil, 1966.

SCHAEFFER, Pierre. *Treatise on Musical Objects: Essays Across Disciplines*. Translated by Christine North; John Dack. Oakland, California: University of California Press, 2017.

SERRA, Xavier; SMITH, Julius. Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic Plus Stochastic Decomposition. *Computer Music Journal*, v. 14, n. 4, p. 12–24, 1990.

SETHARES, William A. *Tuning, Timbre, Spectrum, Scale*. London: Springer, 2005. Available at: <http://public.eblib.com/EBLPublic/PublicView.do?ptiID=303730>. Access: 15 may. 2013.

SODDELL, Thembi. The Acousmatic Gap as a Flexile Path to Self-Understanding: A case for experiential listening. *Organised Sound*, v. 25, n. 3, p. 344–352, dec. 2020.

SOLOMON, Samuel Z. *How to Write for Percussion: a Comprehensive Guide to Percussion Composition*. 2nd. ed. New York: Oxford University Press, 2016.

VALLE, Andrea. Schaeffer Reconsidered: a Typological Space and its Analytical Applications. *Analitica*, v. 8, n. 1, p. 1–15, 2015.

ABOUT THE AUTHORS

Sérgio Freire is Associate Professor at the School of Music, Federal University of Minas Gerais (UFMG, Brazil), teaching composition, orchestration and sonology. Since 1998, coordinates the Laboratory for Performance with Interactive Systems (LaPIS). Graduated in Music (composition) at UFMG. Holds a master degree on Sonology (1993), from the Institute of Sonology, The Hague, Holland, and a PhD degree on Communication and Semiotics (2004) from PUC-SP, Brazil. Collaborator researcher at IDMIL (McGill University) from 2017. His main academic and artistic interests are focused on different forms of interaction between the acoustic musical practice and the new technological means. ORCID: https://orcid.org/0000-0002-5072-3114. E-mail: sfreire@musica.ufmg.br

José Henrique Padovani is Assistant Professor at the School of Music of the Federal University of Minas Gerais (UFMG), teaching Composition, Computer Music, and Music Theory. Padovani is a member of the Music Graduate Programs at the School of Music at UFMG and the Institute of Arts of the State University of Campinas (UNICAMP). More information at: http://josehenriquepadovani.com and http://musica.ufmg.br/padovani/. ORCID: https://orcid.org/0000-0002-8919-7393. E-mail: jhp@ufmg.br

Caio Campos is informatics technician by CEFET-MG, graduated in Music Teaching and currently graduating in Music Composition at UFMG. Caio Campos has his work and research focused on art and technology, being part of composition, free improvisation and sound art groups and practices, many times in interaction with multiple arts. ORCID: https://orcid.org/0000-0002-6702-7799. E-mail: costacaiocampos@gmail.com