Estimation of Perceptual Qualities of Percussive Sounds Inspired by Schaefferian Criteria: Attack Profile, Mass, and Harmonic Timbre

Sérgio Freire, José Henrique Padovani and Caio Campos

Federal University of Minas Gerais
`sfreire@musica.ufmg.br`
`jhp@ufmg.br`
`costacaiocampos@gmail.com`

Abstract. Pierre Schaeffer's typomorphology (1966) proposes seven criteria of musical perception for the qualification of sound objects under reduced listening; these criteria form the basis of a theory (solfège) of musical objects fitted to musical contexts where pitch is not the most relevant feature. We developed a real-time setup that uses low-level audio descriptors to identify and classify percussive sounds as bundles of features related to Schaefferian concepts. The paper describes a segmentation method and the tools and strategies used for addressing three of these criteria: attack profiles (as genres of the criterion dynamic) and mass (which closely relates to the criterion harmonic timbre). The examples depict quantitative results and discuss their correlation with perceptual qualities.

Keywords: typomorphology; Pierre Schaeffer; percussion; audio descriptors

1 Introduction

In recent years, many papers have sought to bring the tools and procedures related to audio descriptors and Music Information Retrieval closer to the theoretical and methodological contributions of Pierre Schaeffer. Such works seek to correlate Schaefferian morphological descriptions and the quantitative data extracted through various techniques of digital signal processing, aiming to automatically index sounds [1], associate "subjective labels" and acoustic features [2, 3], or conjugate descriptors data and perceptual criteria in analytical contexts by employing statistical methods [4], among other approaches [5, 6].

In this paper, we aim to approximate the Schaefferian *solfège* criteria and low-level audio descriptors by implementing real-time audio analysis processes using the computer music language/environment *Max*[1]. In this study, we have chosen the universe of percussive sounds since their sonic qualities are represented and qualified in a very rough manner when using the concepts and parameters of traditional music theory (notes and durations). Bringing into account concepts like complex sound, mass profile, grain, among others, we hope to develop a more efficient tool for the qualification of these sounds to be used in interactive contexts. In this article, we analyze a limited but

---

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

varied selection of percussive sounds, and the results show the suitability of the chosen descriptors for the qualification and differentiation tasks.

The text is organized as follows. Firstly we summarize the seven criteria of musical perception as defined by Schaeffer. After that comes the description of the dataset and procedures used in audio pre-processing and segmentation. A central section deals with the chosen audio descriptors and their correlations with Schaefferian concepts. A set of selected examples comes next. Finally, we present the next steps of the project.

2 Schaeffer's Criteria of Musical Perception

In the sixth book of his *Traité des Objets Musicaux* (TOM), Pierre Schaeffer presents one of the most remarkable contributions of his research: a proposal of a generalized *solfège*, dedicated not only to the traditional musical notes, but also to any sound considered "potentially musical". This *solfège* entails seven typomorphological criteria of musical perception that have the purpose to guide the listening process that consciously attempts to detach the sonic characteristics from any referential or causal events that may generate sound objects themselves: a method that Schaeffer, borrowing the Husserlian concept of *epoché*, named *reduced listening*. In section 34.3, the TOM includes a Summary Diagram (*tableau général*), offering an overview of the whole method, where types, classes, genres, and species of sound objects are described according to seven criteria — mass, dynamic, harmonic timbre, melodic profile, mass profile, grain, and *allure* [7, 8][2]. The criteria help to locate the position and thickness (*site/calibre*) of sound object attributes in the three-dimensional space of a perceptual field formed by *pitches*, *durations*, and *intensities*.

In section 88 of the *Guide des Objets Sonores* [9], Michel Chion outlines the distinctive features of sound objects that each of the morphological criteria proposed by Schaeffer aims to evaluate.

- The *mass* details how the sound occupies the pitch perceptive dimension.
- *Harmonic timbre* describes the "diffuse halos" and "related qualities" that seem to be related to the *mass* and allow its qualification.
- *Grain*, in its turn, is related to the "micro-structure" of sound matter and is associated with rapid variations or reiterations of constituent sounds.
- While *grain* outlines the link between form and matter as one of the sustainment criteria, *allure* expresses the dynamism (mechanical, living, or natural) of what could be defined as a "generalized type of vibrato".
- *Dynamic* expresses the evolution of a sound in the perceptive dimension of *intensities*.
- *Melodic profile* describes the general contour of a sound in the perceptive dimension of pitches, a sort of trajectory in the tessitura.
- *Mass profile*, on the other hand, describes the "internal" variations of a sound in this same perceptive dimension: these changing shapes are responsible for "sculpting" the mass, making it to be more or less thick or thin, having thus a more or less complex or tonic quality, for instance [9].

---

[2] See, particularly, pp. 584-587 of the original edition; pp. 464-467 of the English translation.

As a general method, *reduced listening* involves a conscious attitude that refrains the habitual curiosity towards the sound sources and their meanings in favor of addressing intrinsic features of the sonic phenomena. While authors like Di Scipio [10] remark that the concept of *reduced listening* is ideologically and technologically circumscribed, ignoring the very audible traces of electroacoustic tools that enable us to focus on the 'sound itself', its relevance, since the Schaefferian seminal contributions, lies in the fact that the project of a "generalized *solfège*" has been successful in providing a rich theoretical framework that makes possible to describe different features, behaviors, and qualities of sound objects according to morphological criteria and perceptual dimensions.

Considering Pierre 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, it is relevant to underline the experimental nature of the present work. Thus, despite the differences between perceptual and signal-based evaluation, description, and categorization of sounds, our work is motivated by a common trait of low-level audio descriptors and the Schaefferian *solfège*: both focus on intrinsic qualities of sound phenomena, seeking to discriminate particular characteristics based on certain criteria, dimensions, or parameters. Indeed, Schaeffer himself, even warning to the differences between perceptual processes and what signals can represent, also recognized the usefulness of real-time visualization of signals using bathygraphs and sonagraphs. [7, 8] [3].

3 Selection of Sounds, Pre-processing and Segmentation

In our program, the expected inputs are audio streams delivered by microphones, pickups, or mixers, featuring different background noise levels and dynamic ranges. In the current phase, we have chosen to use a set of pre-recorded sounds. This procedure offers not only variety but also repeatability, two relevant factors for building and improving tools. The sound selection, depicted in Table 1, was based on Schaefferian types. These recorded sounds function as live inputs to the setup[4], which runs with a sampling frequency of 48 kHz.

The sounds are segmented between onset and offset points. In some situations, a new segmentation clue may occur before the offset; in these cases, this clue determines the offset of the previous event and the beginning of the current one, characterized as "slurred". The detection of onsets and offsets occurs by comparing an RMS envelope (expressed in dBFS) with two thresholds, 6 dB and 3 dB, respectively, above the background noise level. This envelope uses a very short window for its estimation — 256

---

[3] In the pp. 556-557 of the English translation of the TOM [8]: "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.(...) 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."

[4] The soundfiles used in this study are available in the following repository: `https://github.com/lapis-ufmg/2021_CMMR_arquivos`

Table 1. Selected sounds.

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

points and hop size of 64 —, which will be referred to as *rms256:4*. In order to obtain more efficiency and precision, we implemented this curve with the Max `[gen~]` object, which uses native audio signal processing routines. Its output is smoothed with a low-pass filter (a single one pole filter, with a –6 dB per octave attenuation), and different cutoff frequencies are employed, depending on the purposes of its use. We use a cutoff frequency of 4 Hz in the estimation of onsets, offsets, and attacks. In the latter case, the audio stream may pass through a filter before the calculations. The estimation of attack profiles and iterative grains uses this same signal with a cutoff frequency of 30 Hz. A control-rate version of this envelope builds the attack profile. Onsets and offsets also function as gates for other processing tools in time and frequency domains. These processes explore the data delivered by a *rms2048:4* curve and by spectral peaks values estimated by the `[sigmund~]` object [11], using the same window and hop size.

4 Perceptual Attributes and Algorithms

Due to text size restrictions, we will concentrate on a subset of Schaefferian perceptual criteria (or sub-criteria) and their correlated audio tools, namely the *attack profile*, *mass*, and *harmonic timbre*.

4.1 Attack Profile

The importance of the attack portion of sounds has been stressed clearly in the work of Pierre Schaeffer, deserving special attention in the TOM and *Solfège*. At that time, he complained that physical measurements were far from representing an accurate picture of the perceived sonic dimensions. About the initial transient portion of sounds (ca. 50 ms), he observed that: “A more spectacular experiment involved asking a very good trumpet player to play a staccato with an accuracy appreciable to the ear: none of this sound's eight impulses gave an oscillogram similar to the others (fig. 5)" [8], (p. 164).

It is not our purpose to argue about how different oscillograms associate with similar perceptions of the beginning of sounds. On the other hand, we intend to approach the initial portion of sounds (ca. 400 ms) with a descriptor that allows for an association between audio features and perceived qualities, using a tool similar to the "bathygraphic traces" depicted by Schaeffer [7, 8][5]. This approach is similar but not equal to the estimation of log-attack-times and attack slopes [12]. We prefer to analyze the entire profile, which may surpass 300 ms, instead of stopping at the point usually named the end of attack. For the same reason, this point will be called the *attack first plateau*. The difference between the levels of the first plateau and the onset is named *attack size*, and the time interval between them *attack duration*. The slope of the first plateau (**FPSlope**) is the ratio between these two values.

The first plateau is estimated as the instant when the derivative of the low-pass filtered audio-rate *rms256:4* curve from the (possibly filtered) input audio stream crosses (or comes near to) zero, just after having surpassed a predetermined positive threshold (a *sharpness* parameter). A value of 200 ms is set as the default reattack threshold since we prefer to consider multiple fast strokes (such as flans, drags, ricochets) as belonging to the same profile. Depending on the settings (filtering and thresholds), this estimation may not produce results for soft attacks[6]. We also prefer to consider some iterative sustainments as a single object displaying *allures*, even when the distance between peaks exceeds the reattack threshold. The onset of a slurred event also marks the offset of the previous one and is defined as the instant when the already mentioned derivative surpasses the given threshold.

The attack profiles depend on their context of production, mainly on the dynamic level and duration (which is also related to excitation and sustainment types). Schaeffer defines seven genres: *abrupt*, *steep*, *soft*, *flat*, *gentle*, *sforzando*, and *nil*. The first three genres relate to different attack-resonance types; a sudden burst of energy characterizes the flat profile; the *gentle* genre has no apparent attack; *sforzando* generally associates with short sounds with a characteristic crescendo; *nil* points to the very progressive emergency of a profile. For the sake of comparison, we have kept all the parameters, except the background noise threshold, fixed for all sounds selected for the present study. The first 300 points after the onset, corresponding to 400 ms, are plotted on a user screen, and stored in a buffer, for further analysis. Figure 1 depicts the attack profiles of nine sounds.

4.2 Mass / Harmonic Timbre

Schaeffer defines seven classes of mass: pure sound, *tonic*, *tonic group*, *channeled*, *nodal group*, *node*, *white noise*. Pure and tonic sounds present a clear pitch, while tonic

---

[5] See p. 533 of the original edition; p. 425 of the English translation.

[6] These parameters (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 contexts with different segmentation clues. The quotations are from p. 402 of the English translation [8].

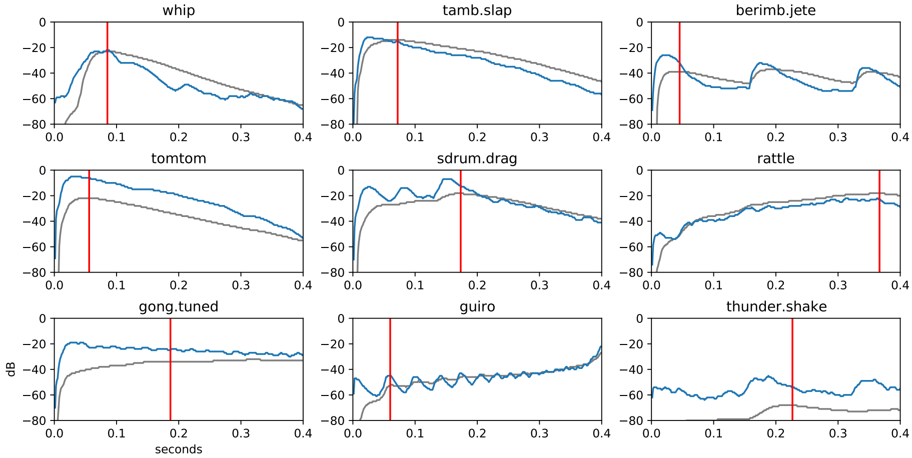

Fig. 1. Attack profiles and first plateau (red mark) of nine percussive sounds. The gray curve is a *rms256:4* low-pass filtered (4 Hz) envelope of the low-pass filtered input signal, and is used for the estimation of the plateau. The blue curve is a *rms256:4* low-pass filtered (30 Hz) envelope of the non-filtered input signal, and is used for the qualification of the attack profile. Filters and threshold parameters remained unchanged for all sounds.

groups are chord-like sounds. Node is a filtered noise occupying a definite spectral region, and nodal group a combination of nodes. Channeled sounds are an ambiguous class between the pitched and unpitched sounds. Within percussion instruments, the most common classes are tonic, channeled, node, and nodal group. However, the latter occurs more like a combination of single nodes than as a single sound object. Harmonic timbre is a complementary criterion to mass, being almost inseparable in some situations. In Schaeffer's words: "Consequently we intend to use the two criteria of mass and harmonic timbre in conjunction with each other, considering them rather as connecting vessels, with the exception of certain specific examples(...)" [8][7]. These specific examples are the pure and tonic sounds. Therefore, we have opted to use the same audio descriptors for both criteria. Typical harmonic timbre attributes are expressed by terms like *full/hollow/narrow*, *rich/poor*, and *bright/matt*. The last attribute pair may overlap with the judgment of the spectral region occupied by the mass of a given sound.

Our strategy to deal with the main classes of mass, and the associated harmonic timbres in the percussive realm, relies on the analysis of spectral peaks and (monophonic) pitches estimated by the `[sigmund~]` object. As we are presently not interested in isolating individual notes inside chords, these two outputs are sufficient for our purposes. Since our context does not presuppose the existence of a harmonic series of spectral peaks, we can not use traditional descriptors as noisiness, inharmonicity, odd-to-even harmonic ratio, tristimulus, among others.

The following descriptors use the spectral peaks (up to 20) estimated for each analysis frame and the energy equivalence given by Parseval's theorem. Their output is a

---

[7] Quoted from p. 412.

curve with a refresh rate of 10.67 ms. Percentile 50 (**pct50**) and percentile 80 (**pct80**) are the number of spectral peaks needed to obtain 50% (-3dB) and 80% (-1dB) of the total energy present in the signal. For sounds with a broad spectral distribution, 20 peaks may not reach the chosen percentiles; in these cases, the output will be 20 peaks, and further information is to be delivered by the next descriptor. The percentage of sound energy represented by up to 20 peaks (**20P/total**) is calculated for each frame and expressed in values between 0 and 1. The frequency of the most prominent peak (**MPP**) in each frame is expressed in Midicents[8]. The object `[sigmund~]` outputs a value in Midicents for the pitched frames, and the value -1500 for unpitched frames. Our descriptor outputs a scalar (percentage of unpitched to total frames: **unpitched/total**) and a curve with all values. In this curve, the unpitched values are represented by the number 1. The estimation of the intrinsic **dissonance** uses the algorithm developed by Sethares [13].

The spectral centroid (**SC**), or the center of gravity of a spectrum, is estimated with a `[gen~]` routine delivered with the Max program since its version 6. 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. The difference between the lowest and highest peak frequencies (Δ **peaks**) is also expressed in Midicents. The spectral **region** is estimated from the contribution of each peak to different spectral 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. If none of these ranges carry 40% or more of the total energy, the sound frame is classified as wideband, labeled as (7). Otherwise, any range with more than 40% of the total energy contributes to qualify one the six spectral combinations: (1) Low, (2) Low/Medium, (3) Medium, (4) Low/High, (5) Medium/High, (6) High.

4.3 Time Series Statistics

Most of the descriptors detailed above are represented by time series, which are subjected to simple statistical analysis just after the offset. Our implementation adapts the algorithms given in [12], and we have chosen the following scalar descriptors: mean value and standard deviation; temporal centroid and spread (normalized by the total duration); skewness; kurtosis; crest; flatness. These values will support the correlations with the Schaefferian perceptual attributes.

5 Examples

For each live input sound, our program generates real-time curves (or markers) for all descriptors and calculates the scalar values described in section 4.3. The results of the analysis of the attack profiles shown in Figure 1 appear in Table 2. We will focus firstly on the sounds produced by one single stroke (percussion-resonance type). Perceptually, the whip sound presents an abrupt profile. This attribute correlates with its short duration, low values for temporal centroid and spread, a positive skewness, and a high crest.

---

[8] Midicent is the unit of a logarithmic scale for frequencies, in which the value 69 represents 440 Hz (note A4), and each integer step is an equal-tempered semitone.

The tomtom stroke presents a steep profile. Although it has a temporal centroid similar to the whip, its spread is much larger, and the skewness and crest are less pronounced. The profile of the tambourine slap lies in between these two sounds. The tuned gong stroke presents a soft profile; in this case, we have a “reinforcement of the resonator”. A longer duration, a slightly positive skewness, and a high value for flatness correspond to this attribute. The rattle profile, produced by a shake, is perceived as gentle since an initial shock is absent. The sound production mixes iterative and continuous aspects. The negative skewness, a small crest, and medium flatness point to this attribute. There are three sounds with a clear iterative or granular profile: the guiro rub, the berimbau jeté, and the drag on a snare drum. The parameters of the two latter are halfway between the steep and the soft profiles; in them, a new stroke prolongs the short resonance. On the other hand, the guiro has a sforzando profile: a high temporal centroid, a considerable negative skewness, a high crest. We believe that the iterative character could integrate the basic profiles as a second-order qualifier. Finally, we have the long thunder shake, whose profile approaches the genre nil: its medium temporal centroid, a low value for crest, and a high value for flatness corroborate this qualification. In this case, the dynamic level (the average value of the entire object) indicates that a marked crescendo will happen during its course. In general, long sounds will rely less on their attack portion for their characterization. It seems to us that the use of the slope of the first plateau (which may vary significantly according to specific parameter settings) and kurtosis could be more significant with more homogenous sounds. In the present selection, they are not as meaningful as the other parameters.

Table 2. Attack parameters for nine selected percussive sounds (the same from Figure 1), plus total duration and dynamic level.

sound	**FPSlope (dB/ms)**	**temp. centroid**	**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
tamb.slap	0.85	0.22	0.74	0.31	0.25	4.07	0.42	461	−33.5
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
guiro	0.63	0.72	0.39	-0.66	1.36	11.09	0.67	631	−42
rattle	0.13	0.67	0.67	-0.21	0.28	2.18	0.67	1151	−42
berimb.jete	0.81	0.35	0.53	0.33	0.66	4.77	0.60	2019	−55.6
gong.tuned	0.24	0.42	1.13	0.08	0.11	1.96	0.92	9219	−42.6
thunder.shake	0.18	0.53	0.20	-0.19	3.82	2.80	0.86	15541	−33

The discussion about mass and harmonic timbre relies on data displayed in Tables 3 and 4. A significant presence of pitched frames points to a tonic sound; the opposite indicates a node or a channeled sound. In addition, the concentration of energy in a few spectral components helps differentiating between tonic and channeled sounds on one side and nodal sounds on the other. The combination of these two descriptors can discriminate between *tonic* (bass drum, whistle, friction of a tambourine skin, and tabla), *channeled* (tomtom, tuned gong, and snare-drum without snares), and *nodal* (cymbal, guiro, ratched, and rattle, among other) sounds. The stability of the most prominent peak may also favor *tonic* and *channeled* sounds, but one must be careful with the presence of melodic profiles, as in the case of the tabla sample. Although intrinsic dissonance values may also point to tonic sounds, this is not a univocal association since spectral peaks not presenting a simple harmonic ratio only increment this value if they are close enough in frequency (see details in [13]). The interpretation of spectral centroid (and region) values is straightforward; however, a high value for the standard deviation indicates the presence of some profile (glissando, undulation, filtering, etc.). We can analyze the shape of these profiles with tools similar to those used for the attack profiles.

In addition to the symbiotic relation between mass and timbre, Schaeffer states that "there is no classified index for perceptions of harmonic timbre" (TOM, p. 420). In the present case, we try to approximate the perceptive attributes *full/hollow/narrow* with the descriptors Δ peaks, pct50, pct80, and region (helped by the intrinsic dissonance), and the opposition rich/poor with the values estimated for pct80 and 20P/total. For instance, observing different sounds classified in the medium region, it is possible to split them between *hollow* (bass drum, tomtom) and *narrow* (whistle, rattle) according to these parameters.

Table 3. 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
tomtom	487	1 ± 0.2	1.96 ± 3.1	0.96 ± 0.1	0.38
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
tamb.tremolo	1365	12.5 ± 4.2	20 ± 0	0.62 ± 0.1	0.77
whistle	1463	1.4 ± 0.6	5.2 ± 6	0.89 ± 0.1	0.13
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

6 Final remarks

The results obtained so far have demonstrated that our setup can qualify and differentiate diverse types of percussive sounds with a good approximation to the Schaefferian criteria. We believe that we have shown the importance of attack profiles for percussive sounds and the pertinence of the implemented tools for the qualification of mass and harmonic timbre. The next planned steps are the work with performers in real-time interactive contexts (when the pandemic allows), the choice of the most efficient descriptors for each intended perceptive feature, the training of a machine learning algorithm, and the development of interactive musical works.

Table 4. 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
tomtom	46.2 ± 17.5	55.8 ± 0.2	64.9 ± 15.3	60.8 ± 10.2	3
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
tamb.tremolo	235.9 ± 91.7	111.9 ± 23.8	35.5 ± 31.7	118.4 ± 2.7	6 ± 0.6
whistle	132.8 ± 29.5	98.2 ± 0.4	31.9 ± 13.3	98.4 ± 2.4	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

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

References

1. Peeters, G., Deruty, E.: Sound Indexing Using Morphological Description. IEEE Transactions on Audio, Speech, and Language Processing 18(3), 675–687 (Mar 2010)
2. Bernardes, G., Davies, M., Guedes, C.: A Pure Data Spectro-Morphological Analysis Toolkit for Sound-Based Composition. pp. 31–38. Proceedings of the eaw2015, Aveiro (2015)
3. Godøy, R.I.: Perceiving Sound Objects in the Musique Concrète. Frontiers in Psychology 12 (2021), `https://www.frontiersin.org/articles/10.3389/fpsyg.2021.672949/full`
4. Martins, F., Padovani, J.H.: Analysis Informed by Audio Features Associated With Statistical Methods: a Case Study on 'Imago' (2002) by Trevor Wishart. In: Proceedings of the 5º Encontro Internacional de Teoria e Análise Musical. vol. 1, pp. 219–234. Campinas (2020)
5. Ricard, J.: Towards Computational Morphological Description of Sound. Doctorate, Universitat Pompeu Fabra, Barcelona (Sep 2004)
6. Valle, A.: Schaeffer Reconsidered: a Typological Space and its Analytical Applications. Analitica 8(1), 1–15 (2015)
7. Schaeffer, P.: Traité des Objets Musicaux. Éditions du Seuil, Paris (1966)
8. Schaeffer, P.: Treatise on Musical Objects: Essays Across Disciplines. No. 20 in California studies in 20th-century music, University of California Press, Oakland, California (2017)
9. Chion, M.: Guide des Objets Sonores. Buchet/Chastel, Paris (1983)
10. Di Scipio, A.: The Politics of Sound and the Biopolitics of Music: Weaving Together Sound-Making, Irreducible Listening, and the Physical and Cultural Environment. Organised Sound 20(3), 278–289 (Dec 2015)
11. Puckette, M.S., Apel, T., Zicarelli, D.D.: Real-time Audio Analysis Tools for Pd and MSP. In: Proceedings of the International Computer Music Conference. pp. 109–112. San Francisco (1998)
12. Peeters, G., Giordano, B.L., Susini, P., Misdariis, N., McAdams, S.: The Timbre Toolbox: Extracting Audio Descriptors from Musical Signals. The Journal of the Acoustical Society of America 130(5), 2902–2916 (Nov 2011)
13. Sethares, W.A.: Tuning, Timbre, Spectrum, Scale. Springer, London (2005), `http://public.eblib.com/EBLPublic/PublicView.do?ptiID=303730`