An analysis of *Desdobramentos do continuo* for violoncello and live electronics performed by audio descriptors

Danilo Rossetti[1], William Teixeira[2], Jônatas Manzolli[1]

[1] First Laboratory – Interdisciplinary Nucleus for Sound Communication of UNICAMP
Rua da Reitoria, 165 - "Cidade Universitária Zeferino Vaz" – 13083-872, Campinas, SP

[2] Second Laboratory – Federal University of Mato Grosso do Sul
Avenida Costa e Silva, 79070-900, Campo Grande, MS

danilo.rossetti@nics.unicamp.br, william.teixeira@ufms.br, jotamanzo@gmail.com

Abstract

This article proposes an audio descriptors model for the analysis of live electroacoustic music. In this context, an analysis of the work *Desdobramentos do contínuo* for violoncello and live electronics is addressed, concerning both tape (deferred time) sounds and live electronics (instrument sound and real-time processing). For this analysis, audio descriptors such as spectral flux, energy mean, centroid, and loudness were employed. The objective was to define which events produce huge timbre variations and to identify timbre subtle nuances which are not perceptible on a first listen of the work. We conclude comparing the analysis results to the compositional hypotheses presented in sections 2 and 3.

1. Introduction

Audio parametric descriptors are tools which extract different information from audio recordings. The objective of this procedure is to analyze these data aiming to understand some features related to human auditory perception and to perform a classification of the evaluated pieces and musical styles. This research area is known as MIR (Music Information Retrieval) and the obtained analyses results until now are available in the MIREX web page[1] (Music Information Retrieval Evaluation eXchange).

The use of audio descriptors for musical classification was already employed in previous researches, such as in Pereira [1], Peeters [2] and Peeters *et Al.* [3]. The Interdisciplinary Nucleus for Sound Communication of UNICAMP (NICS) developed similar research in the past few years, resulting in the works of Monteiro [4], and Simurra; Manzolli [5,6]. In relation to the use of audio descriptors specifically for the analysis of contemporary music, we mention the work of Malt and Jourdan [7].

The general objective of this article is to contribute in the area of analysis of live electroacoustic music by audio descriptors. Specifically, an analysis of Rossetti's work *Desdobramentos do contínuo* (2016), for violoncello and live electronics is performed. We first present the work contextualization and a discussion about the instrumental extended techniques employed in the instrumental writing. For the analysis, parametric audio descriptors are applied to the audio recording. These audio descriptors form an analysis model which can possibly be used to analyze other live electronics works. Our analysis will be centered on the behavior of the spectral flux, energy mean, centroid and loudness descriptors whose definitions will be detailed further up.

The reason for the use of audio descriptors in this analysis is that the employed instrumental extended techniques produce new sounds and transients. In this sense, an analysis based only on the score do not encapsulate the entire possibilities of the work. Furthermore, the interaction between the fixed tape sounds and the dynamic part of the sound processing produce dynamic elements which are beyond the score notation. Thus, to contemplate these aspects related to the work, a new methodology of analysis is necessary and will be developed next.

[1] http://www.music-ir.org/mirex/wiki/MIREX_HOME

2. Work Contextualization

Desdobramentos do contínuo is a work for violoncello and live electronics composed in 2016 by Danilo Rossetti. It is the last work included in his doctoral thesis [8], which investigates interaction and convergence possibilities between acoustical instruments and electroacoustic treatments [9]. This work is dedicated to the cellist William Teixeira, who participated in its development, which involved rehearsals, cello recordings, and audio analyses.

The general form of the work contents two parts that differ from each other mainly concerning the employed electroacoustic treatments. These treatments can be implemented in real time (morphological transformations of the cello sound captured live along the performance) or in deferred time [10] (audio manipulations involving phase vocoder and convolution processes from pre-recorded cello phrases).

Among the real-time treatments, granulation, microtemporal decorrelation, and dephasing are employed. An ambisonic spatialization of the electroacoustic sounds is conceived, creating a diffused sound field that surrounds the listener in performance. This spatialization is planned for an octophonic model, however quadraphonic and stereo versions of the piece also can be performed. The integration of real-time electroacoustic treatments with an ambisonics spatialization is achieved by the utilization of the *process~* object, belonging to the High Order Ambisonics Library (HOA). This library was developed by the CICM of *Université* Paris 8 [11]. In *Desdobramentos do contínuo*, the architecture of the patch was implemented in Max MSP.

The objective of overlapping both deferred time sounds and real-time treatments of violoncello sound was to explore different possibilities of the electroacoustic universe. The adopted compositional hypothesis was that this combination would be complementary in terms of sound morphology. So, the overlapped sounds and would merge together into a single timbre. In this process, the tape sounds have a continuous and similar development, on the other hand, real time treatments generate sounds with discontinuous granular characteristics. In the analysis performed further on, these questions will be verified.

Next, the instrumental part of *Desdobramentos do continuo* will be discussed, focusing on extended techniques and the resultant sound morphology.

3. Instrumental extended techniques

The role of instrumental writing within the discourse[2] [12] of the piece is immense, but perhaps not in the way expected from a piece regularly written for an acoustic instrument with electronic support. The concept of the piece started in an attempt to escape from two extremes usually noticed in pieces written within the genre.

On the one hand, a compositional tendency is identified where the instrument functions simply as a signal generator, with the electronic synthesis being the most prominent character of the development of musical discourse. The instrumental gesture functions almost in subordination to the electronic gesture and the later function is to constitute continuity to the insertions, almost disparate, of the former.

On the other hand, it is possible to notice another extreme, where the instrumental gesture alone assumes the role of the generator of musical material, and the electronic support working just like that, a support, a kind of effects box that only ornaments the musical discourse, almost autonomous, executed by the acoustic instrument. In this case, electronics create only small inserts of effects, and sometimes act like a tape, executing another set of materials without any interaction with the instrumental gesture.

Desdobramentos comes from this attempt to overcome such extremes, starting from the interaction between the two sound sources as basic writing material. It is important to state that, because, in this sense, the instrumental writing works in a way not so much "soloistic", but much more like chamber music, since musical materials are generated by both and often, from the interaction of both. At all times there is a feeding of new gestures, where it is up

[2] Musical discourse is a term adopted here to mean the whole relationship between musical agents and not as a synonym of musical work and even less to notation.

to the instrumentalist to be able to respond instantaneously to the stimuli produced by the electronic source, including the sort of sound produced by extended techniques; like chamber music, these stimuli never repeat themselves, because they are also responses, in turn, of events previously produced by the instrument and which are never identical. This is the great beauty and difficulty presented by the proposal and that ends up giving a great dynamism to musical discourse. Understood in this totality, the discourse assumes more fully its vocation of interacting with, for and through its agents.

Even so, the instrumental writing brings difficulties of a very advanced technical level that needs to be overcome for the effectiveness of the mentioned interactions. One of the first questions that rise when the musical score is performed is the presence of three levels of bow pressure, as in the following passage, Figure 1:

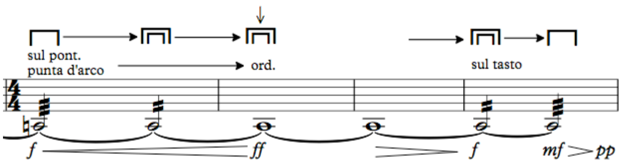

Figure 1: Measures 31 to 34 of the score

Although the performance of this kind of sonority is already very settled within the writing for strings in the contemporary repertoire, the piece brings a new issue that is the execution of long passages in *legato* with these different levels. This requires more than just changing the 90° angle of the bow in relation to the string, which usually generates a distorted sound, but it requires the extra weight of the interpreter. Making the three levels sound distinctive and at the same time homogeneous throughout the piece results in the fact that the piece requires a lot of the musician's physique when played in its entirety.

Another very demanding gesture it is in the “rebounds section”, so to speak, where different kinds of bow-strokes are prescribed in different rhythmical structures and with different numbers of notes per stroke, as in Figure 2:

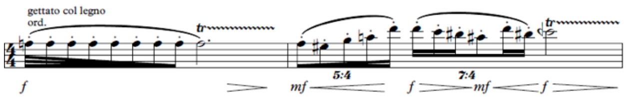

Figure 2: Measures 43 and 44 of the score

The aim here is to make the sounds always respond to the granular sound in the electronic synthesis, so the duration of each note must be at the same time proportional to the duration written, fluent in the gestural flow and as short as a granular sound must be to put the sound inside the bigger sonority.

The last passage that worths to be mentioned for its odd instrumental technique is this in Figure 3, but that occurs in other sections in the end of the piece:

Figure 3: Measures 43 and 44 of the score

This is a good example where traditional technique must be expanded not because a different timbre is required, but due to a new musical context; here a regular rush passage is full of notes written in hard string skipping and in the same direction of the bow, everything in a *crescendo* gesture. The result of such requirements among the microtonal pitches is an only and single sonority, almost like the Mannheim rockets in the Classical period, but that reviews the prominence of sound instead of only notes.

4. Audio Descriptors Model

To analyze *Desdobramentos do contínuo*, a model formed by different types of audio descriptors was determined. This model included descriptors which provide temporal, spectral, energy, and psychoacoustic features. The selected descriptors (detailed below) were spectral flux [1,4], energy mean (RMS) [1,4], spectral centroid [1,2,3,4], and loudness [1,2,3,4]. The computational environment used for the descriptors calculation was the *Pdescriptors Library*, designed by Adriano Monteiro in PureData software [4], and revised by Gabriel Rimoldi.

Spectral flux descriptor is defined as the magnitude difference between two successive analysis windows (frames). This descriptor provides lower values when the spectrum remains relatively invariable, on the other hand, it provides higher values when huge variations between successive frames are found. It is calculated from the expression below:

$$fe_i = \sum_{K=1}^{K} \{\log_{10}[X_i(k)] - \log_{10}[X_{i-1}(k)]\}^2$$

Energy mean or RMS (root mean square) is the root mean square of amplitude values in a window analysis. These values describe the energy envelope profile of a sound and are defined by the following equation:

$$RMS_i = \sqrt{\frac{\sum_{n=0}^{N-1} x_i[n]}{N}}$$

The spectral centroid is the barycenter of the energy distribution belonging to the spectral envelope of a sound. Perceptively it is related to the sound brightness perception. Higher values characterize the predominance of high frequencies in the signal (in Hertz) and lower values characterize the predominance of lower frequencies, in terms of energy. The spectral centroid is calculated from the following expression:

$$ce(i) = \frac{\sum_{k=1}^{k} k * \left|X_i(k)\right|^2}{\sum_{k=1}^{k} \left|X_i(k)\right|^2}$$

Loudness is a psychoacoustic measure related to the perception of sound amplitude. It is variable according to different frequency bands (as demonstrated by the Fletcher and Munson curves, in 1933) and describes the auditory sensation of amplitude variation of a given sound. The loudness of a spectral analysis window is determined by this equation, according to Pereira's model [1]:

$$L_i = \sum_{k=1}^{k-1} \left|X_i[k]\right|^2 10^{W[k]/20}$$

These chosen audio descriptors will be applied to the audio recording of the piece whose analysis will be presented next.

5. Analysis by Audio Descriptors

In the audio of the performance used for this analysis[3], the entire piece lasts 10'35''. Its first part goes from the beginning to 4'47'', and the second part from 4'47'' to its ending. In the first part, the deferred time process corresponds to a phase vocoder that stretches the spectrum of a given sound and repeats it continuously as a loop. During this part, the sound is sent to a granulator which has six different presets containing a sort of parameter values (such as grain size and rarefaction). These presets determine the direction of the sound mass evolution, whose perception changes gradually from a continuous timbre to a grainy sound cloud considerably rarefied.

In the second part, the applied deferred time process to generate the tape sounds was the convolution between different pre-recorded cello sounds. In total five sounds of different durations were generated by this process (which have respectively 35'', 26'', 50'', 62'' and 78'' of duration). As common perceptual features among them, all these sounds have continuous spectral evolutions in time. It is important to remark that during the entire piece, besides the tape sounds, the cello sound is granulated in real time (its parameters are constantly modified), and the electroacoustic timbre (formed by these layers) is spatialized through high-order ambisonics models.

5.1 Analysis of Deferred Time Sounds

In this section, the looped phase vocoder sound of the first part (which changes gradually in time) and the five tape sounds of the second part, generated by convolution processes, will be analyzed and discussed.

In the first part, the phase vocoder generated sound evolves directionally from a continuous texture to a grainy sound mass which gradually becomes more discontinuous in perception. Our audio descriptors model was applied to the audio and, the resultant graphics are presented in Figure 4.

[3] The performance took place in NICS UNICAMP at 16th December 2016, by Pedro Bortolin (cello) and Danilo Rossetti (live electronics).

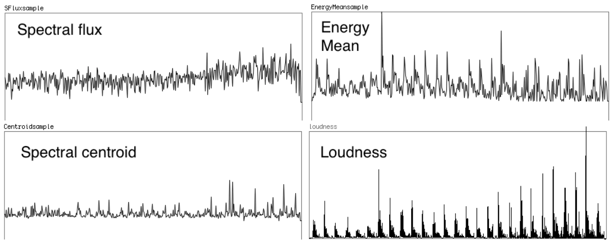

Figure 4: Descriptor model applied to phase vocoder sound

As it is showed in the graphics above, energy mean evolution (RMS) presents few peaks of energy which appear periodically. Otherwise, the centroid curve has also fewer and weaker peaks in the beginning. It just has relatively stronger peaks in the end of sound evolution. From these observations, we assume that, in average, both of these curves have no considerable variations.

On the other hand, despite the strong peaks belonging to the loudness curve and the smaller peaks belonging to the spectral flux curve, both of them raise in intensity at their end. At the same point (around 3'10"), spectral flux and loudness curves both start to have higher values. Here, the growth of the spectral flux curve is more consistent, since it is more constant, meaning that there are more intensity variations between successive frames. In loudness curve, the perceived increase (from 3'10") is mainly related to the periodically existent peaks. After them, the intensity sensation falls to lower levels.

We assume that the variations found in spectral flux and loudness curves are related to the granulation parameters applied to the phase vocoder sound. In the first part of the work, six different granulation preset values were applied. On them, while the feedback rate and grain delay remain constant, the grain size decreases from 400 to 75ms, while rarefaction rate increases from 0 (a totally continuous sound in perception) to 0,8 (indicating 80% of silence in the diffused sound mass).

In relation to the grain cloud perception, it is important to emphasize that bigger grains generate sonorities that privilege the sustained parts of the sounds (normally characterized by the presence of a fundamental frequency and upper partials). Smaller grains have a prominent presence of attack transients. For this reason, from a sound morphology standpoint, grain clouds formed by smaller grains sizes (of less than 100ms) have a noisier auditory perception [8,9].

On the second part of *Desdobramentos* five tape sounds were addressed (Seq. 1 to 5). As a common feature of all these tape sounds they all have a continuous evolution. However, it is desired to investigate if they have different evolution characteristics. In this sense, audio descriptors can help us to evaluate some timbre features of these sounds, in order to describe their behavior. We applied the presented descriptors model to each sound and extracted the normalized (from 0 to 1) arithmetic average of each descriptor value (Table 1). This strategy was adopted to obtain significant data, in order to compare the evolution of the descriptors applied to the sounds.

Sound/ Descriptor	Seq. 1	Seq. 2	Seq. 3	Seq. 4	Seq. 5
Flux	0,32	0,4074	0,3111	0,2733	0,3051
RMS	0,183	0,3612	0,2962	0,4797	0,4094
Centroid	0,084	0,3515	0,2388	0,2581	0,3816
Loudness	0,63	0,6051	0,7636	0,7291	0,7113

Table 1: Normalized averages of the audio descriptors of the second section

From Table 1, it is possible to verify that the five sequences of the deferred time sounds show a gradual increase of energy means. This behavior is more prominent in Seq. 4 and 5. Therefore there is more spectral energy at the end of the piece. These spectral changes act in the perception as an increase of intensity and sound density.

Regarding spectral centroid values, we observe that the five tape sounds are organized in three brightness levels: low, middle and high. The low brightness level is assigned to the Seq. 1, the middle brightness level is assigned to Seq. 3 and Seq. 4, and the high brightness level is related to Seq. 2 and 5.

Finally, the normalized loudness average values are concentered in a middle-high level. Seq. 1 are nearer to the middle level, while Seq. 3, Seq. 4 and Seq. 5 have higher intensity perception. Next, in Figure 5, a histogram graphic is presented, showing the descriptors average values related to each tape sound, in complementation to Table 1.

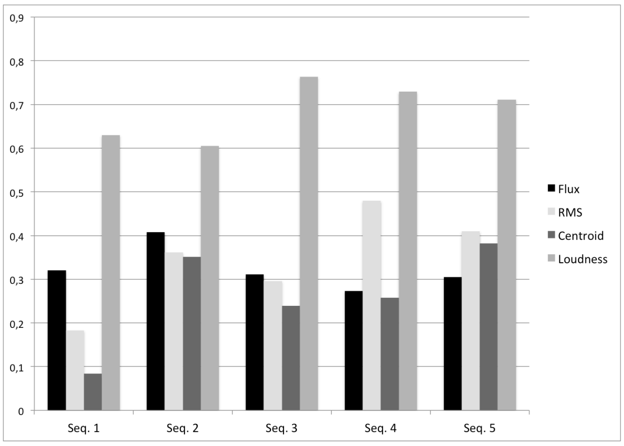

Figure 5: Descriptor values for each sound

Taking into account this figure, some observations can be made, in consideration to a global view of the descriptor values of each sound. Seq. 1 has the lower energy, lower centroid and one of the lowest loudness values. Seq. 2 has the higher flux, high centroid, and the lower loudness. Seq. 3 has the higher loudness, average centroid, and average-low energy. Seq. 4 has the higher energy, high loudness, average centroid and the lower flux. Seq. 5 has the higher centroid value, high energy, and low flux.

5.1 Analysis of Real-Time Processing

In this analysis, we focus on the real-time granulation of the violoncello sounds merged together with its acoustical sound. For this purpose, two different excerpts of the piece will be addressed. These excerpts were chosen from a spectral flux analysis of the entire piece. They correspond to moments where the flux descriptor reaches high values (over the average).

As we can see in Figure 6[4] (sonogram and spectral flux curve of the entire piece), there are more than two moments in the graphic presenting high flux values. Then, to decide which excerpts would be addressed, we opted to take those in which we have distinct instrumental techniques performed. It is known that different instrumental techniques generate different timbre morphologies. So, it becomes possible to compare the average descriptor values correspondent to these morphologies (instrument and real-time electronic sounds).

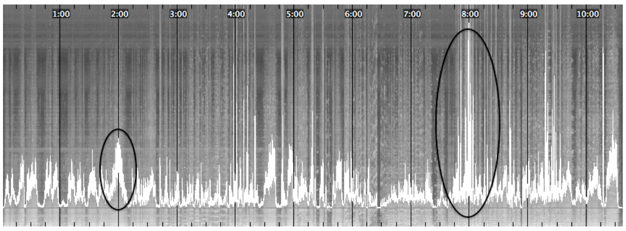

Figure 6: *Desdobramentos do contínuo* sonogram and spectral flux curve: two circulated excerpts to be analyzed

The first excerpt corresponds to measures 31 to 34 of the score, shown in Figure 1 (1'55'' to 2'10'' of the audio recording). As it is observed in this figure, there is a C2 pitch which is sustained by four measures, having as features different speeds of *tremolo*, different bow positions (*ordinario*, *sul ponticello* and *sul tasto*), and different bow pressures over the string (normal pressure, exaggerated and creaking noise). As the combination result of these instrumental possibilities, different timbres were generated.

In addition, the descriptors model (spectral flux, energy mean – RMS –, centroid and loudness) is applied to the correspondent audio of this part. The arithmetic average of the achieved normalized values was also calculated, in order to make possible the comparison with the average values of the second example.

The second excerpt refers to measures 119 to 123 of the score (7'42'' to 8'05'' of the audio recording), shown in Figure 7. In this example, the main instrumental techniques played by the violoncello are closer to percussive gestures (*pizzicatti*, *pizzicatti Bartók*, and *gettato col legno*), generating discontinuous sonorities. *Pizzicatti tremolo* and bow positions from *ordinario* to *sul tasto* are also requested in the score in this excerpt.

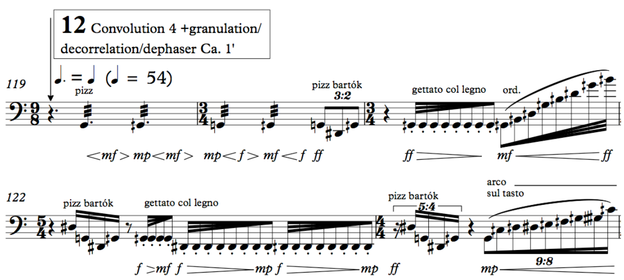

Figure 7: Measures 119 to 123 of the score

[4] In order to have both the sonogram and the spectral flux curve in the same figure, this audio analysis was performed in Sonic Visualiser software.

In the score (Figure 7), the number “12” above measure 119 indicates the Max patch preset which corresponds to specific values of the cello granulation. It also indicates the beginning of the fourth tape sound belonging the second part of the piece, which was previously analyzed. The same audio descriptors model was applied to this excerpt and their arithmetical averages are shown below in Figure 8 graphic. In this graphic, a comparison between values of excerpts 1 and 2 is performed.

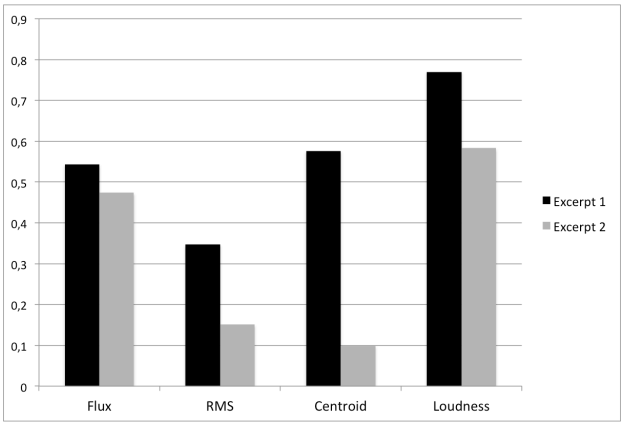

Figure 8: Descriptor average values of excerpts 1 and 2

Regarding the Figure 8 graphic, the average values of spectral flux descriptor for both excerpts are relatively close. Loudness level is a little higher for the first excerpt, in comparison to the second.

Higher differences were found in RMS and centroid descriptors, which may be attributed to the morphological characteristics of the generated timbre. The first excerpt contains continuous timbre variations of a C2 pitch, according to the bow pressure and position, and also to the presence of the *tremolo* effect. We observed that the intensity has a constant behavior in this example, due to a variation from *forte* to *fortissimo* in the score notation. In relation to the real-time granulation response, it is correlated in most aspects to the acoustical instrument sound morphology.

Average RMS values of both excerpts are very different. It can be attributed to the timbre prominent discontinuity of the second excerpt, characterized by instrumental techniques which produce “percussive” sonorities. Moreover, the average values corresponding to centroid have a huge difference. It means that the spectrum barycenter (brightness sensation) of the first excerpt is located in an average-high area, on the other hand in the second excerpt it is located in a low-frequency area.

Thus, from these two excerpts, we observed that different instrumental techniques combined with real-time granulation (having both high spectral flux values) differ each other mainly from centroid and RMS average values. Otherwise, the average loudness values were not considerably different, reflecting on similar auditory sensation levels.

6. Conclusion

In this article, we intended to establish an audio descriptors model which is useful for the analysis of live electroacoustic music. This is an attempt to contribute in the field of computer-aided music analysis. For the analysis of *Desdobramentos do contínuo*, we found relevant the use of spectral flux, energy mean, centroid and loudness descriptors. For further analyses or even to a more detailed analysis of this work, it is possible to increase our model with new descriptors.

From the analysis of tape (deferred time) sounds by these descriptors, we could extract important information that clarifies some characteristics of the timbre of these sounds. On a first listen, we tend to consider them similar to each other, due to their continuous time evolution. However, after the application of our descriptors model, subtle variations become noticeable and our perception becomes more attentive to these nuances. It is also desirable during the performance that the interpreters are aware of these nuances. Thus, they can interact with them on a higher level, in order to produce a more convergent performance, considering acoustic and electroacoustic parts.

These subtle timbre variations, in a certain way, complement the previously presented compositional hypothesis. The tape sounds have a global continuous evolution. However, for the phase vocoder sound, after a certain point, there is a discontinuity perception demonstrated by higher flux and loudness values. In relation to the five tape sounds of the second part, variable values of RMS and centroid indicate different features of their global timbre perception.

In addition, despite the tape sounds nuances, the main timbre differences in the global perception of the work (defined by the variations of the spectral flux descriptor applied to the entire piece) are related to the employed instrumental techniques and their real-time granulation. Considering other audio descriptors, these timbre variations reflect mostly on RMS and centroid differences.

Thus, from this analysis, we verified that the change in the perceived timbre morphology is mostly guided by the instrumental writing. However, in real-time electroacoustic music (in general) this causality can be broken by the employment of tape sounds with unexpected effects and events.

7. Acknowledgments

Danilo Rossetti is supported by FAPESP in his post-doctoral research at NICS-UNICAMP (process 2016/23433-8). Jônatas Manzolli is supported by CNPq under a Pq fellowship, process 305065/2014-9.

References

[1] Erica M. Pereira. Estudos sobre uma ferramenta de classificação musical. Dissertação (Mestrado). FEEC Universidade Estadual de Campinas, Campinas, 2009.

[2] Geoffroy Peeters. A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project. Cuidado Project Report. Institut de Recherche et de Coordination Acoustique-Musique (IRCAM), 2004.

[3] Geoffroy Peeters *et Al*. The Timbre Toolbox: Extracting audio descriptors from musical signals. In J. Acoust. Soc. Am. 130 (5), November, 2011, pages 2902-2916.

[4] Adriano C. Monteiro. Criação e performance musical no contexto dos instrumentos musicais digitais. Dissertação (Mestrado). Instituto de Artes, Universidade Estadual de Campinas, Campinas, 2012.

[5] Ivan Simurra; Jônatas Manzolli. Sound Shizuku Composition: a Computer-Aided Composition Systems for Extended Music Techniques. In MusMat Brazilian Journal of Music and Mathematics 1 (1), Rio de Janeiro, 2016, pages 86-101.

[6] Ivan Simurra; Jônatas Manzolli. O azeite, a lua e o rio: o segundo diário de bordo de uma composição a partir de descritores de áudio. In Revista Música Hodie 16 (1), Goiânia, 2016, pages 101-123

[7] Mikhail Malt; Emmanuel Jourdan. Zsa.Descriptors: A Library for Real-Time Descriptors Analysis. In SMC 08 Sound and Music Computing Conference, Berlin, 2008.

[8] Danilo Rossetti. Processos microtemporais de criação sonora, percepção e modulação da forma: uma abordagem analítica e composicional. Tese de Doutorado. Instituto de Artes, Universidade Estadual de Campinas, Campinas, 2016.

[9] Danilo Rossetti. Interaction, Convergence and Instrumental Synthesis in Live Electronic Music. In: M. Aramaki, R. Kronland-Martinet, S. Ystad, editors, *Proceedings of the 12th CMMR*, pages 209-216, The Laboratory of Mechanics and Acoustics, 2016, pages 209-216.

[10] Flo Menezes. Fusão e contraste entre a escritura instrumental e as estruturas eletroacústicas. In Flo Menezes, *Atualidade estética da música eletroacústica*, Editora UNESP, pages 13-20, 1999.

[11] Pierre Gillot. Documentation de la bibliothèque HOA: Les champs sonores. CICM, Université Paris 8, Labex Arts H2H. http://www.mshparisnord.fr/hoalibrary/ambisonie/les-champs-sonores/

[12] William Teixeira. O discurso musical. In: F. Presgrave; L. Noda; J. J. Mendes, editors, *Ensaios sobre a música do século XX e XXI*: composição, performance e projetos colaborativos, EDUFRN, pages 226-243, 2016.