ACCESSING STRUCTURE OF SAMBA RHYTHMS THROUGH CULTURAL PRACTICES OF VOCAL PERCUSSION

Luiz Naveda
IPEM – Ghent University
luiznaveda@gmail.com

Marc Leman
IPEM – Ghent University
marc.leman@ugent.be

ABSTRACT

In the field of computer music, melodic based forms of vocalizations have often been used as channels to access subject's queries and retrieve information from music databases. In this study, we look at percussive forms of vocalizations in order to retrieve rhythmic models entrained by subjects in Samba culture. By analyzing recordings of vocal percussions collected from randomly selected Brazilian subjects, we aim at comparing emergent rhythmic structures with the current knowledge about Samba music forms. The database of recordings was processed using a psychoacoustically inspired auditory model and further displayed on loudness and onset images. The analyses of emergent rhythmic patterns show intriguing similarities with the findings in previous studies in the field and put different perspectives on the use of vocal forms in music information retrieval and musicology.

1. INTRODUCTION

Beatboxing, Puirt-a-beul or *bols* are some of the examples of vocal percussion forms found in different cultural backgrounds. These practices generally make use of non-meaningful phonemes, which imitate instruments and often rely on onomatopoeia [1 , 2]. The information available on forms of vocal percussion account for examples distributed over different musical cultures and practices, which range from simple devices for musical learning to elaborated forms of performing art. Vocal percussion may have differentiated from melodic singing due to the necessity for more freedom in expressing rhythmical ideas. It is also possible that it originated from the combination of a vocal apparatus in human species and the necessity of rhythmical expression in all human cultures. However, what makes these forms of vocal percussion relevant to our study is the use of the voice as a seamless link between musical intentionality and acoustic energy. In this study, we use this link to access rhythmic intentionality and analyze rhythmic structure within Brazilian samba.

SMC 2009, July 23-25, Porto, Portugal

1.1. Vocal percussion

Cultural forms of vocal percussion have been rarely mentioned in traditional musicological research. More recently, the emergence of *hip-hop* in the cultural sector has shed light on the *beatboxing* form*,* which is only a modern and localized form of vocal percussion. Other examples found in the bibliography describe music forms in India (*bols*), song genres in Ireland (*Puirt-a-beul)*, pedagogic devices for conga teaching in Cuba [1] and verbal art in Africa [3]. So far, it seems that vocal percussion assumes diverse socio-cultural roles and importance, although only a small number of dispersed scholarly and non-scholarly records have showed this in detail.

In computer music some attention has been devoted to the potential use of vocalizations, in special pitch based vocalizations in western music contexts. The easy assessment of user's musical intentionality for music retrieval applications appears to be a central element in the "query-by-humming" approach. During the last years, a large number of publications and implementations have been produced in the field [see 4, for a review of algorithms]. Less noticeably, a small number of studies approached vocal percussion from this perspective. Kapur et al. [5] used the *beatboxing* as a mechanism to retrieve and analyze drum loops and their rhythmical structures. Nakano et al. [6] developed a similar approach by using native Japanese speakers as subjects, which demonstrates the application of the approach in phonetically different backgrounds. Kang & Kim [7] used vocal percussion for real-time animation of motion clips (dance animations). Although most of these studies aim at understanding how vocal queries relate with musical databases, in very few academic work the opposite is shown, namely, how vocal queries are related with subjects' conceptions of musical forms.

Heylen et al. [8] provides a study that uses spontaneous vocalizations to access subjects' musical conceptions or models. In this study, subjects were asked to sing along to several music pieces in different tonal contexts. The results show emergent major and minor tonal structures that resulted from spontaneous vocalizations. The use of vocal apparatus to retrieve the tonal evaluation from subjects is understood as a *corporeal articulation* in response to different tonal stimulus. The insightful turn of vocalizations into corporeal articulations is also a crucial concept in our study. It opens channels to less formalized responses to music, which include not only vocalizations but also body movement. The framework of embodied music cognition [see 9], in which this concept is developed, seem to adapt more naturally to the problems of musicological investigation developed in less-formalized music cultures. The music and dance traditions of the samba culture are examples of cultural artifacts dominated by informal learning and practices, in which our universe of study is delimited.

1.2. Vocal percussion in Samba

Samba music is generally described as having a binary meter music form accentuated in the second beat, and a rhythmic texture that is characterized by syncopated rhythms [10-13]. The music is only one component of an intricate complex, in which forms of dance, music, poetry, rituals and social relations develop mostly through in an informal context [14, 15].

Ziriguidum, balacobaco and *telecoteco* are some of the very common onomatopoeic expressions used in Brazil. They are not easily found in dictionaries, but they are intuitively linked with samba concepts, behaviors and culture. Expressions like these appear in thousands of internet references (mostly in Brazilian Portuguese): blog posts, magazines, books, little enterprises, dance clubs, restaurants and others. Surprisingly, there are almost no references on the use of vocal percussion as a common practice. To what extent is this practice common in Brazilian society? Which characteristics of this form of vocal percussion are consistently aligned with the musical repertoire? How do they reflect the conceptions of samba within the acculturated population?

In this preliminary study, we concentrate on the last two questions. By analyzing a database of vocal percussions recorded from randomly selected Brazilian subjects, we aim at providing the first images about this aspect of samba culture. The database of audio recordings was analyzed using a psychoacoustically inspired auditory model and further processed into loudness and onset images as described in the next sections. Section 2 describes the procedures used in the recordings and dataset. Section 3 explains the methods used to analyze and produce images of rhythmical content. In the Section 4, we show and discuss the results, rhythmical structures derived from the analysis and compare them with results from previous studies about samba music.

2. DATASET

The dataset used in this study is a growing database of vocal percussions recorded in Brazil, between 2008 and 2009. This database will be further complemented with questionnaires in order to provide better information about socio-cultural profiles of the subjects (not analyzed in this study).

In order to create conditions to access practical measures of how the practice is present in the universe of study, we opted to randomly choose subjects (passers-bys) in public spaces. The sessions took place in four different locations in Belo Horizonte (Brazil), in relatively quiet spaces (classrooms). The recordings were done with a professional digital recorder and high-quality microphones, using a sample definition of 44100 b/s at 16 bits (stereo), stored in SD cards.

First, the subjects filled out a brief contact form. No information about the study was provided before the recordings. During the second part of the experiment, the subjects were invited to perform samba rhythms, using their voice in spontaneously organized sequences. These sequences were registered in one single take. If the subject refused to perform or declared her/himself unable to perform the task, this was registered in a form. No training sessions or repeated takes were used in this experimental model.

3. METHODS AND ANALYSIS

3.1. Segmentation and normalization

The audio excerpts were segmented manually. The criteria for segmentation were selecting and extracting homogeneous excerpts that last a minimum of 4 or 8 beats. Each excerpt was then normalized at 0 dB (amplitude) and the channels merged into mono aural WAV files. Although female and male voice differences may have an influence on the overall auditory images, we opted to avoid any kind of spectral normalization or further processing aimed at normalizing these differences.

3.2. Analysis

3.2.1. Loudness images

During the first stage of our analysis, the excerpts were processed with a psychoacoustic inspired auditory model based on [16] and implemented as a windows executable. The auditory model simulates the auditory decomposition in the periphery of the auditory system, which results in 40 channels of loudness patterns obtained from a simulation of neural rate activity distributed over the audible spectrum [for more details see 16, p. 3514]. The loudness patterns were processed by the auditory model implementation at sample rate of 100 frames/second. All images were further resampled and normalized to a sample rate of 98 frames per musical beat. This procedure allows comparing and accumulating groups of auditory images and onset images, which originally had different absolute time durations.

In this study, we display auditory spectrum images with 4 or 8 beats lengths, depending on which illustration was used. When the auditory images are displayed with 8-beat length, 4 beat images will duplicate. When 4-beat images are displayed, only the first half of the 8-beat images is displayed.

3.2.2. Onset images

To avoid imprecise onset accumulations derived from a sum or a mean of loudness patterns, we applied a method for onset detection to each auditory image using an integrate-and-fire neural net based on an approach developed in [17] and implemented in IPEMToolbox [18][1]. In the sequence, onset images were quantized in 256 sample images (128 samples for 4-beat images), which means that onset attacks were integrated in 32 sample "slots" per musical beat. This procedure helps to slightly integrate deviated onsets in a single position when concatenating onsets in mean images. Mean onset images were produced from the integration of onsets in the same "slot" position (1/32 beat). Figure 2a show mean images of loudness, Figure 2b show mean image of onsets.

In order to visualize a single rhythm profile in the high portion of the auditory images, we summed up the high channels of the auditory images, as displayed in the Figure 3 (basically half of the channel distribution: 21:40 for high-frequency channels), which seem to be sufficient enough to demonstrate the propositions of section 4.2.3.

4. RESULTS AND DISCUSSION

The results consist of information regarding the overall profile of the database and a discussion about the selected images of the dataset, analyzed with the methods mentioned above.

4.1. Overview

We collected recordings from 55 subjects, which produced a database of 80 excerpts (1.5 excerpts/subject). A percentage of 5.4% subjects (3 subjects) was unable to perform the task. A percentage of 41% of the recordings was performed by female subjects and 59% by male ones. The actual database is composed of 45 excerpts with 4-beat length and 35 excerpts with 8-beat excerpts. We extracted the mean BPM values for all excerpts manually. The BPM list has a normal distribution (Kolmogorov-Sminov test, alfa = 0.05) with mean 102.09 and standard deviation 18.5. The minimum BPM found was 59.6 while the maximum was 184.4.

4.2. Analysis of auditory and onset images

4.2.1. First images

Figure 1 (a, b, c), displays 3 phases of the generation of loudness/onset images. Firstly, (1a) the first loudness images are extracted and resampled. They are followed by an (1b) onset image derived from the loudness images and (1c) the same onset image quantized to 32 samples/beat. The quantization of onset attacks provides a better integration of onsets for each 1/32nd segment of the musical beat.

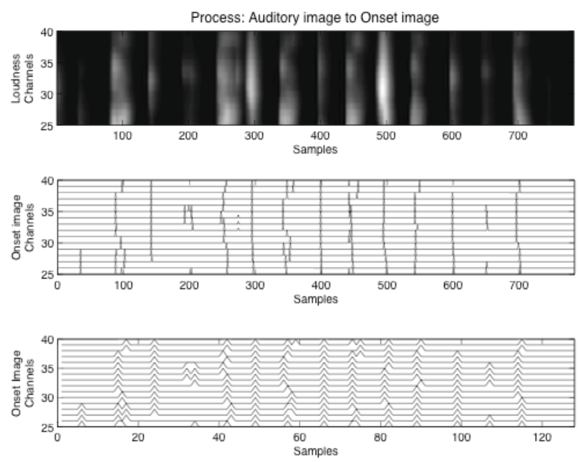


Figure 1(a) - First loudness image of an excerpt of vocal percussion performed by a female subject. Degrees of gray represent normalized loudness (white:black=0:1); **(b)** - onset image at 98 samples/beat; **(c)** - quantized onset image at 32 samples/beat.

4.2.2. Mean loudness and onset images

We calculated the mean loudness image and a mean onset image for all excerpts. Figure 2a and 2b summarize loudness and onset channels for all vocalizations.

[1] IPEM toolbox is available at http://www.ipem.ugent.be

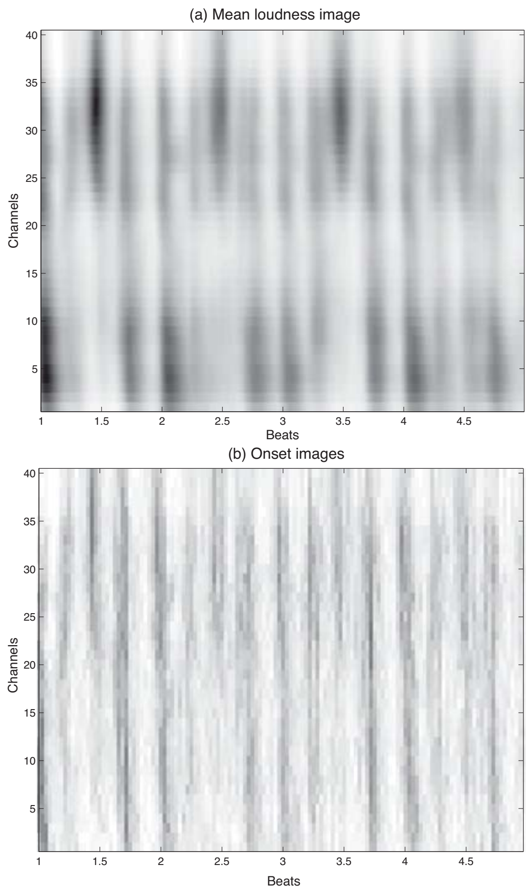
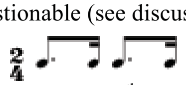


Figure 2. 4-beat mean **(a)** loudness image and **(b)** onset image for all excerpts.

In both Figures 2a and 2b we observe the isolation of low and high-pitched onsets, denoted by a separation between high and low frequency patterns. These two distinct layers seem to exhibit periodic rhythmic patterns along time. The high portion of the spectra shows a characteristic tatum layer of samba, composed of ¼-beat onset profiles (16th notes in a 2/4 bar). The low portion of the spectra exhibits a punctuated rhythm (1) that stresses each beat mark, which can also be found in several samba recordings (such as the one detailed in Figure 3). An auditory inspection of the sound database confirms the relevance of these observations and that low-frequency patterns are usually performed with consonants *t* or *d*, followed by a very low sound. Of course, transients of these consonants spread over the high spectrum, which makes the separation of onsets between spectral regions questionable (see discussion in section 4.2.4).

(1) 

The third 16th-note onsets of each beat (at the 0.5 position of the beat) are often accentuated in the high-spectra. The second 16th-note onsets are often less clearly defined in both high and low portions of the auditory spectra. Double onset peaks verified some of the 2nd, 3rd and 4th 16th-notes in the onset image (Figure 2b) seem to indicate the presence of different groups of microtiming behaviors, not evident in the loudness images.

4.2.3. Tatum and microtiming

The tatum layer is defined as the lowest level of musical metrical hierarchy, and normally detected through the fastest rhythmical figures observed in the rhythmical texture. This musical layer is found in high-pitched patterns of instrumental samba ensembles and commonly represented in musical notation as isochronous 16th note figures. Figure 3, extracted from [19], displays a similar onset analysis applied to a commercial recording of samba music (*Ela veio do lado de lá* – Benito Di Paula, 1975). In this case, our results seem to confirm the same structure observed in commercial samba recordings.

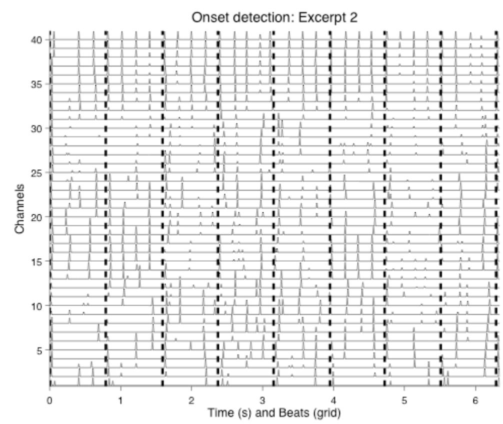


Figure 3. Onset analysis of a commercial samba recording. Traced lines indicate beat points.

Recent studies have examined the role of microtiming in the induction of "groove" in Brazilian music. Gouyon [20] studied the microtiming in the samba using a database of commercial recordings. Lindsay & Nordquist [21] analyzed microtiming using commercial recordings and recording of individual samba performers. Gerischer [22] studied the groove and microtiming using field recordings of Brazilian percussion groups. All these studies demonstrated the existence of a systematic anticipation of the 4th and 3rd onsets in groups of four 16th-notes (1 beat).

Figure 4 displays the analysis of microtiming between the peaks of the onset images (high auditory spectrum, channels 21:40). The top graph shows the accumulated onset image (1/32 beat) and the bottom graph displays the sample difference between each onset and the following one.

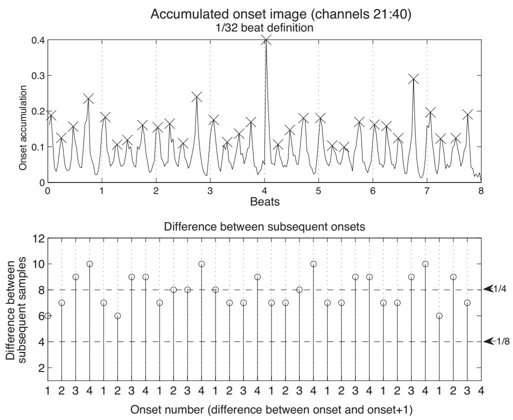


Figure 4. Microtiming analysis of the higher onset image (1/32 beat samples). The bottom graph shows the distance (in samples) between onsets. Traced lines represent 1/8 and 1/4 isochronous sample divisions of the beat.

The patterns displayed in the bottom graph of Figure 4 show that the intervals between 1st and 2nd, 2nd and 3rd 16th-notes (second graph, numbers 1 and 2, horizontal axis) are normally shorter than others. 3rd and especially 4th intervals (numbers 3 and 4, horizontal axis) are often larger than the mathematical isochronous ¼ rule of the tatum layer. The pattern of deviations verified by our database of vocal percussions is strikingly consistent with studies mentioned above, especially with the results displayed in [22, p. 105], [20, p. 200] and [21, p. 26]. Although such an observation seems to be a trivial confirmation of findings of previous studies, the fact that untrained subjects could re-enact precise microtiming structures verified in musical stimuli is worth paying attention to.

4.2.4. *Rhythmical hypotheses*

By subsuming the hypothetical voicing possibilities of our observations, we can create a collection of rhythmical motives that may provide explanations for the rhythmic possibilities found in the results. Figure 5 demonstrates these possibilities in traditional music notation:

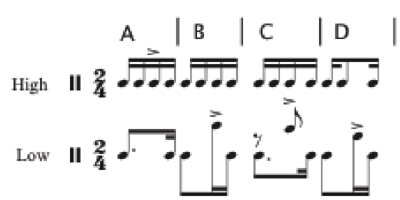


Figure 5: Hypothetical rhythmic motives (A, B, C, D) extracted from our observations. This collection of motives is not exhaustive and other variations may occur.

The motive A is the most obvious structure found in the loudness/onset images because it directly mirrors the high-low spectral layers. If we accept that the third 16th-note is linked with the low voice, a low-high-low pattern (B) will appear in the low voice, while tatum layer is still maintained. However, this strong accent may also represent a third mid-frequency voice (C) due its slight distribution over the mid frequencies. In this case, the mixture of channels is supposed to be an inevitable consequence of the monophonic condition of human voice. Finally, the same condition in B may be exclusively pertinent to the low voice, which gives rise to a syncopated pattern (D) in the high spectra. The last hypothesis of course implies that the existence of a constant tatum layer in the higher portion of the auditory spectrum must be rejected. After all, these hypotheses seem to indicate that syncopated or contrametric lines are not strongly represented by the models that lie underneath the vocal percussion practices. It is also possible that syncopation onsets may exist, but they are so dispersed and distributed over the time span that their presence is masked by the commetric forces or by the average procedure in the calculation of mean images.

The results raise very intriguing questions. Samba musicians, composers dancers, researchers and expert listeners seem to agree that syncopation, polyrhythmic and contrametric content are the most salient characteristic of Samba music [14, 23, 24]. It was expected that subjects would likely perceive and store traces of highly syncopated stimuli and, therefore, be able to perform entrained patterns with the same characteristics. However, what we observed is that vocal percussion patterns do not show observable syncopation at relatively accessible macro time and mid spectrum, but demonstrate a surprising consistency at microtiming level. In which part of the chain stimuli-subject-vocalization was the syncope filtered?

The answer to this question will take us further than this preliminary study. Improved methodologies and the incorporation of other modalities may be necessary to better represent the samba complex as a systemic structure. Nevertheless, the results seem to demonstrate the richness of the vocal practices in providing elements for both information retrieval and cognitive modeling investigation.

5. CONCLUSION

In this paper we analyzed a database of vocal percussion in samba culture. We aimed at understanding how rhythmic models emerged from the vocalizations and how they articulate with models of the samba music. The study shows that the ability to perform this practice seem to be spread over the group of subjects of this study, and that rhythmic models are relatively consistent throughout mean images of the loudness and onset patterns. Rhythmic patterns derived from these images show 3 important observations: that (1) rhythmic patterns are similar with overall samba music characteristics and (2) strikingly entrained at microtiming level. The syncopation priority of samba music (3) seems to be absent in vocal percussion representations but a better methodology may clarify the presence of syncopated onsets.

In terms of perspectives for the SMC and MIR fields, this study shows that vocal queries are not simple copies of models of musical intentionality, but may emerge from complex interactions between acoustic stimuli/environment and other modalities such as corporeal engagement with music (dance). However, the precision of vocal queries may reach levels of performance comparable with professional musicians, with the advantage that it situated within the context of the ubiquitous and normalized medium of human vocal emissions. Future work must include a better pattern detection methodology, a robust statistical verification, analyses of cross-modal interactions and a more systematic investigation over the profile of subjects, metrical content and microtiming.

6. ACKNOWLEDGEMENTS

The authors wish to thank the subjects who contributed to our study and the anonymous reviewers for the suggestions. This research is supported by a grant from Ghent University (Belgium), and partially support by CAPES (Brazil).

7. REFERENCES

[1] M. Atherton, "Rhythm-Speak: Mnemonic, Language play or Song?," in *International Conference on Music Communication Science*, Sidney, 2007, p. 15.

[2] D. Stowell and M. Plumbley, "Characteristics of the beatboxing vocal style," Tech. Rep. C4DM-TR-08-01, Dept. of Electronic Engineering, Queen Mary, University of London, 2008.

[3] L. Hunter, "Transformation in African Verbal Art: Voice, Speech, Language," *Journal of American Folklore,* vol. 109, pp. 178-192, 1996.

[4] R. Typke, F. Wiering, and R. Veltkamp, "MIREX symbolic melodic similarity and query by singing/humming," *MIREX 2006,* p. 19, 2006.

[5] A. Kapur, M. Benning, and G. Tzanetakis, "Query-by-beat-boxing: Music retrieval for the DJ," 2004, pp. 170–177.

[6] T. Nakano, J. Ogata, M. Goto, and Y. Hiraga, "A Drum Pattern Retrieval Method by Voice Percussion," *Joho Shori Gakkai Kenkyu Hokoku,* vol. 2004, pp. 45-50, 2004.

[7] K. Kang and D. Kim, "Synthesis of Dancing Character Motion from Beatboxing Sounds," 2007, pp. 216-219.

[8] E. Heylen, D. Moelants, and M. Leman, "Singing along with music to explore tonality," in *9th International Conference on Music Perception and Cognition*, Bologna, 2006.

[9] M. Leman, *Embodied Music Cognition and Mediation Technology*: Mit Press, 2007.

[10] J. Chasteen, "The prehistory of Samba: Carnival Dancing in Rio de Janeiro, 1840-1917," *Journal of Latin American Studies,* vol. 28, pp. 29-47, Feb. 1996.

[11] M. E. Mariani, "African influences in Brazilian Dance," in *African Dance*, K. W. Asante, Ed. Asmara: Africa World Press, 1998, pp. 79-98.

[12] M. Salazar, "Batucadas de Samba," *Rio de Janeiro: Lumiar Editora,* 1991.

[13] C. Sandroni, *Feitiço decente: transformações do samba no Rio de Janeiro, 1917-1933*: Jorge Zahar Editor: Editora UFRJ, 2001.

[14] M. Sodré, "Samba, O Dono do Corpo," *Rio de Janeiro: Codecri,* 1979.

[15] P. Fryer, *Rhythms of Resistance: African Musical Heritage in Brazil*: Pluto, 2000.

[16] L. Van Immerseel and J. Martens, "Pitch and voiced/unvoiced determination with an auditory model," *The Journal of the Acoustical Society of America,* vol. 91, p. 3511, 1992.

[17] L. Smith, "Onset-based sound segmentation," *Advances in neural information processing systems,* pp. 729-735, 1996.

[18] M. Leman, M. Lesaffre, and K. Tanghe, "A toolbox for perception-based music analyses," in *IPEM-Dept. Of Musicology, Ghent University,* Ghent, 2001.

[19] L. Naveda and M. Leman, " A cross-modal heuristic for periodic pattern analysis of Samba music and dance," *Journal of New Music Research (in press),* 2009.

[20] F. Gouyon, "Microtiming in "Samba de Roda"— Preliminary experiments with polyphonic audio," in *XII Simpósio da Sociedade Brasileira de Computação Musical* São Paulo, 2007.

[21] K. Lindsay and P. Nordquist, "Pulse and swing: Quantitative analysis of hierarchical structure in swing rhythm," *The Journal of the Acoustical Society of America,* vol. 122, p. 2945, 2007.

[22] C. Gerischer, "O suingue baiano: Rhythmic feeling and microrhythmic phenomena in Brazilian percussion," *Ethnomusicology,* vol. 50, pp. 99-119, 2006.

[23] M. E. Mariani, "A Portrayal of the Brazilian Samba Dance with the Use of Lab Analysis as a Tool for Movement Analysis," University of Wisconsin, Madison, 1986.

[24] B. Browning, *Samba: Resistance in Motion*: Indiana University Press, 1995.