# Microtiming in “Samba de Roda” —Preliminary experiments with polyphonic audio

**Fabien Gouyon**[1]

[1]INESC Porto (Institute for Systems and Computer Engineering of Porto),
Campus da FEUP, Rua Dr. Roberto Frias, 378
4200-465 Porto, Portugal

fgouyon@inescporto.pt

***Abstract.*** *In this paper we report on preliminary experiments in the study of microtiming features in Samba music. Focusing on polyphonic audio music with almost constant tempo, we propose an algorithm to discover from data in a bottom-up manner systematic timing deviations at the 16th-note level in quarter-note-long temporal patterns.*
*Our analysis of the data highlights a systematic shift of third and fourth 16th-note beats, slightly ahead of their corresponding quantized positions.*

## 1. Introduction

Music styles can often be characterized by specific repeating rhythmic patterns. There are two important aspects to such patterns. First, the relative time instants of musical notes as measured on a discrete, quantized time grid (specified by integer multiples or divisions of a basic pulse). This is how a pattern would be noted on a score. And second, the small and yet systematic timing deviations between instants where notes are actually played and their corresponding quantized positions. In some cases, these deviations can be represented as series of tempo changes, while in other cases they are better represented as *event shifts* at a *constant tempo* [Desain and Honing, 1991, Bilmes, 1993]. In this paper, we focus on the latter type of deviations, which we will refer to as “microtiming.”

Such deviations occur in many different musical styles and depend on the position in the metrical structure. For instance, Jazz is characterized by a particular pattern of deviations: the “swing,” where “consecutive eighth-notes are performed as long-short patterns” [Friberg and Sundström, 2002]. Traditional Irish fiddle music also shows comparable timing deviations at the eighth-note level [Rosinach and Traube, 2006]. One characteristic of Viennese Waltz is also a certain amount of swing, but this time at the quarter-note level (each third quarter-note in a bar is shorter). In this paper, we are interested in a particular type of Samba music (see Section 2), and focus on systematic deviations at the 16th-note level. That is, we consider quarter-note-long patterns and seek whether systematic timing deviations occur around each of the four 16th-note beats.

There is a number of computational approaches to the study of performers’ timing expressiveness [Widmer, 2002]. The work of [Bilmes, 1993] is of special interest here. He analyzes percussion-based audio data,[1] introduces the notion of “tatum” (fastest metrical pulse) and focuses on deviations with respect to this level of the metrical hierarchy. He proposes a semi-automatic[2] transcription system relying on onset detection and stroke classification, then, the metrical position of each stroke is determined and timing deviations are computed. Similar phrases are then clustered and systematic deviations can be

[1]with a separate audio track for each instrument
[2]i.e. with complete knowledge of the metrical structure

induced via machine learning algorithms. The learned microtiming deviations can then be applied to quantized phrases in order to generate expressive musical phrases. Synthesis is done via the triggering of isolated percussion samples.

[Wright and Berdahl, 2006] propose a system for percussion-based MIDI data. Their system learns deviations from quantized positions for 9 different Brazilian rhythms (none of them being Samba de Roda) via diverse machine learning algorithms and then apply these learned patterns to quantized data, with satisfactory results.

Finally, some authors proposes algorithms for determining the swing of audio signals, e.g. [Laroche, 2001]. Additionally, the swing of such signals can be modified by time-scaling techniques [Gouyon et al., 2003, Janer et al., 2006].

The paper is structured as follows. First we provide details of the data used for experiments. We then propose an algorithm to highlight patterns of microtiming deviations in quarter-note-long segments. We then discuss some findings and propose lines of future work.

## 2. Data

For these preliminary experiments, we collected a relatively small number of audio excerpts, namely 49, of length ranging between around 10 to 30 s. Audio data were ripped from commercial CDs to 44.1 kHz mono. These excerpts are representative of traditional Samba music in the particular style of Rio de Janeiro's "Samba de Roda," with acoustic guitar, four-stringed small Brazilian guitar (i.e. "cavaquinho"), and a percussion section (tambourine —i.e. "pandeiro"—, friction drums, etc.), following a characteristic duple rhythm with second and fourth beats in a bar often marked by a low-frequency percussion sound. In this style of music, it is very common that the tambourine and "cavaquinho" follow a rhythmic pattern at the 16th-note level. Artists and bands are Teresa Cristina & Grupo Semente (albums "A vida me fez assim", "A música de Paulinho da Viola" vol. 1 & 2 and "O mundo é o meu lugar"), Renascença Clube (album "Samba do trabalhador"), Velha guarda da Portela (album "Tudo azul"), Grupo Fundo do Quintal (album "Seja sambista também"), Elton Medeiros, Nelson Sargento and Galo Preto (album "Só Cartola") and Paulinho da Viola and Elton Medeiros (album "Samba na madrugada").

There are between 15 to 73 beats per excerpt (e.g. around 5 to 18 bars per excerpt), reaching a total number of 1803 beats.

## 3. Algorithm

In order to discover *systematic* patterns of deviations with respect to quantized positions, we make use of an algorithm to compute rhythmic patterns inspired from [Dixon et al., 2004]. Among other differences, detailed below, we focus on patterns at a different level of the metrical hierarchy (quarter-notes instead of bars), use a different beat tracking algorithm and use a different signal representation (complex spectral difference instead of amplitude envelope).

### 3.1. Beat tracking

Studying timing patterns in quarter-note segments requires knowledge of individual beats at this level. We segmented the data with the use of the semi-automatic software described in [Gouyon et al., 2004]. When necessary, we manually oriented tracking towards quarter-note level.

### 3.2. Rhythmic patterns

#### 3.2.1. Onset detection function

Audio data is processed into a representation of lower dimensionality highlighting note onsets. Instead of using the signal amplitude envelope, as in [Dixon et al., 2004], we chose to use one of the onset detection functions proposed in [Bello et al., 2004]: the “complex spectral difference” (the spectral difference between consecutive signal frames computed in the complex domain, i.e., accounting for magnitude and phase). Frames are 23.2 ms-long and hop size is set to 11.6 ms. The resulting sampling frequency is $44100/512 = 86.1$ Hz. See an example in Figure 1.

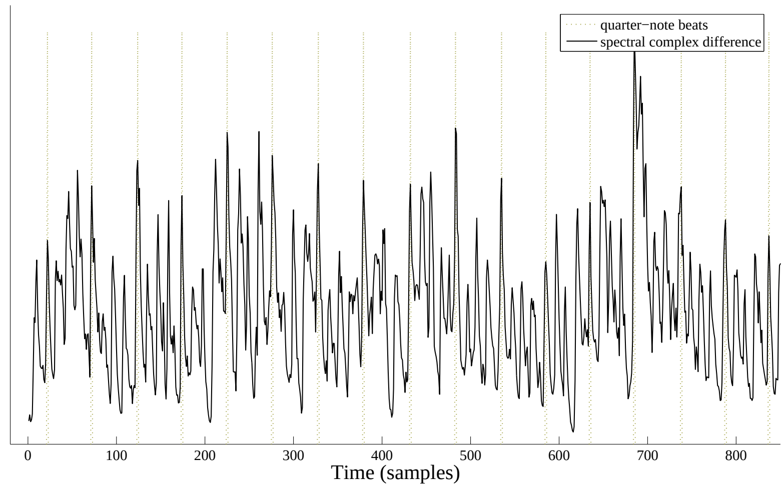


**Figure 1: Example of complex spectral difference, and quarter-note beats (beginning of the excerpt “Tive Sim” from the album “Só Cartola” by Elton Medeiros, Nelson Sargento and Galo Preto).**

#### 3.2.2. Beat recentering

Beat positions are then slightly corrected (automatically) so that they would correspond precisely to note onsets. For this, we use a tolerance window of 50 ms around beats and reset beats to the closest maximal onset in this window.[3]

#### 3.2.3. Resampling and normalization

As written above, in the data we use, tempi are roughly constant. However, some slight differences can appear in Inter-Beat Intervals (IBIs). In order to be able to accurately compare patterns in quarter-note segments of slightly different lengths, we must resample data in each segment so that they would all have the exact same length. We chose to resample to 40 points per quarter-note segment, using a polyphase implementation.

Segment amplitudes are then normalized to unity.

[3] The length of the tolerance window is not critical.

#### 3.2.4. Extraction of typical quarter-note segment patterns

One way of computing the typical quarter-note pattern is to compute the average, for each point in the segments, over all $1803 - 49 = 1754$ segments. Figure 2 shows an illustration of this average pattern together with some individual patterns (randomly selected).

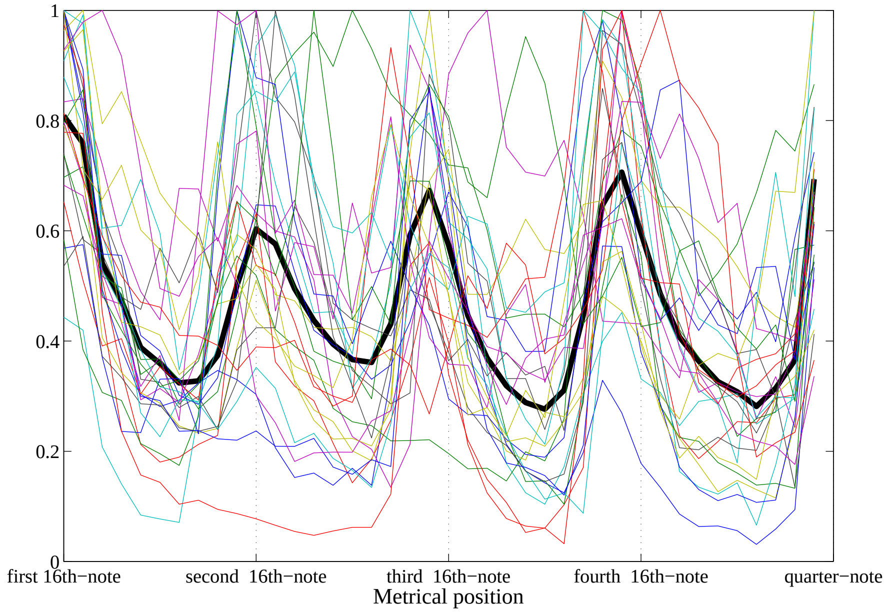


**Figure 2: 27 randomly selected patterns, and average pattern (thick line).**

However, we can expect some patterns to be outliers (as e.g. those corresponding to quarter-note segments in an excerpt's introduction, or fill-ins). We therefore remove outliers by clustering patterns with a $k$-means algorithm, as in [Dixon et al., 2004].[4] Figure 3 illustrates three of the typical patterns in our data, obtained by $k$-means clustering. They account for 35%, 31% and 34% of the patterns, respectively.

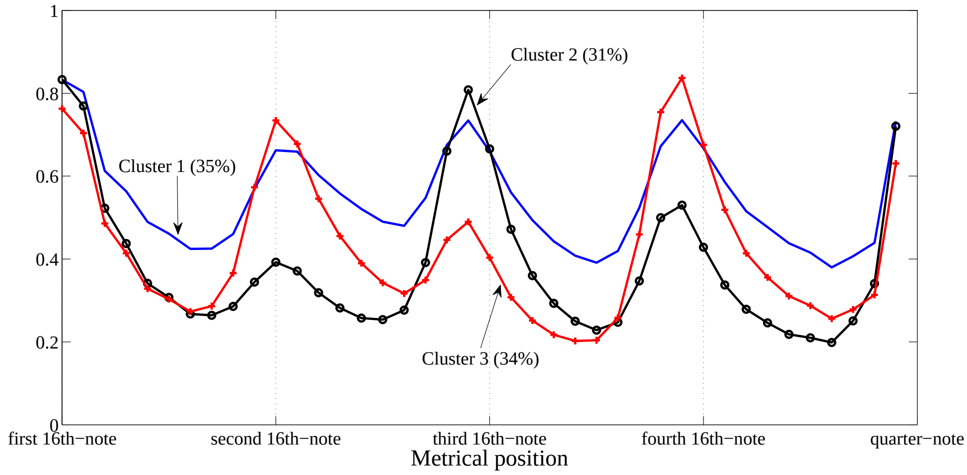


**Figure 3: Three patterns obtained via $k$-means clustering (with $k = 3$).**

[4]with $k = 3$, we used Weka for this, see `http://www.cs.waikato.ac.nz/ml/weka` and [Witten and Frank, 2000].

## 4. Discussions

We can see in Figures 2 and 3 that most of the patterns show local maxima around each of the four 16th-note beats in a quarter-note. This corresponds to the fact that in Samba, there is usually an explicit metrical grid at the 16th-note level, mostly set by the tambourine and other percussive instruments, or harmonic-percussive instruments, as the "cavaquinho."

We can also see in Figure 3 that there are three typical patterns in our data: one with accents (by "accent," we mean local maxima in the onset detection function) on all 16th-notes (cluster 1, where the onset detection function has similar amplitudes around all 16th-notes), one with accents on the first and third 16th-notes (cluster 2), and one with accents on the first, second and fourth 16th-notes (i.e. cluster 3).

We can retrieve representative instances of these clusters by correlating clusters with all quarter-note segments in the data. For instance, the waveform and spectrogram[5] of the most representative instance of cluster 3 (with accents mostly on the first, second and fourth 16th-note, this is especially clear on the waveform) is shown in Figure 4.

### 4.1. Systematic deviations

More interestingly, we can also see, in the average pattern, and even more so in the typical patterns obtained by $k$-means clustering, that both the third and fourth 16th-note beats are slightly *ahead* of their corresponding quantized positions. Notes seem to be played typically (on average) slightly before their quantized positions, with an advance of around 1/40 of the IBI. This corresponds to almost 20 ms at a tempo of 90 BPM (typical for this kind of music).

We can see in Figure 4 and example of such a pattern, where the third and fourth 16th-notes are played almost 30 ms ahead of their quantized positions.

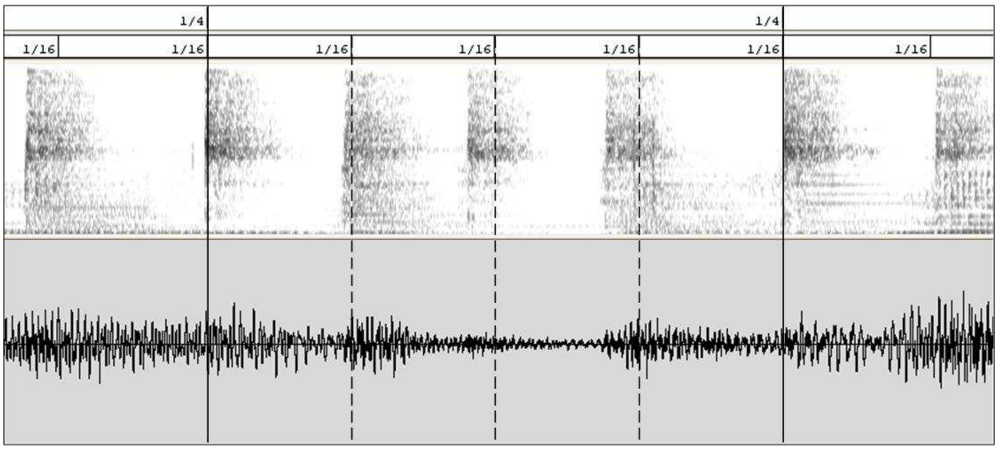


**Figure 4: Illustration, from bottom to top, of the audio waveform, spectrogram and metrical structure at the 16th-note and quarter-note levels for one of the 1754 quarter-note-long patterns in our data (i.e. one of the patterns in the excerpt "Alvorada" from the album "Só Cartola" by Elton Medeiros, Nelson Sargento and Galo Preto). This pattern is typical of the** *strong-strong-weak-strong* **form found in our data (i.e. cluster 3). Note also the timing shift of the third and fourth 16th-notes, slightly ahead of their corresponding quantized positions.**

## 5. Future work

The work reported here should be continued and cross-validated by further experiments with a larger dataset. A complementary avenue for future work would be to purchase

[5] frequencies are represented on a linear scale from 0 to 22050 Hz

MIDI-matched audio data in the same music style, this would open the way to a finer precision in the analysis which would be necessary to capture subtle microtiming differences in different quarter-note segments, or different measures [Wright and Berdahl, 2006].

It would also be interesting to study the sensitivity of these findings to the number of clusters. Further experimentations with different machine-learning algorithms, or different front-ends (e.g. spectral centroid normalized by energy as suggested in [Paulus and Klapuri, 2002]), could also be of interest. In complement, one could also explore outliers properties.

In these experiments, we purposely normalized the length of each quarter-note pattern. It would be interesting in further experiments to study a possible dependency of the found deviations with respect to tempo (as has been demonstrated in the particular case of swing [Friberg and Sundström, 2002]).

A cross-disciplinary effort could also be done in comparing findings reported here to Samba-specific musicological literature, e.g. [Sandroni, 1996].

There are diverse applications to the findings reported here. First, one might think of taking advantage of these characteristic features for music similarity and music genre classification [Dixon et al., 2004]. Second, it would also be interesting to develop a system for music transformation where such deviations could be explored (either augmented or subtracted). This would permit to study their perceptual relevance. This would also open the way to change the expressiveness of quantized data, and make it sound "more human" [Bilmes, 1993, Gouyon et al., 2003, Janer et al., 2006, Wright and Berdahl, 2006, Widmer, 2002, Ramirez and Hazan, 2006]. Finally, further studies could be dedicated to the relation between microtiming deviations and perception of "Groove" in Samba, as well as in other music styles.

## Acknowledgments

Carioca da Gema, Claudia Vargas, Guy Madison, Simon Dixon and Kalle Hörnström. Thanks to two anonymous reviewers for their insightful comments.

## References

Bello, J., Duxbury, C., Davies, M., and Sandler, M. (2004). On the use of phase and energy for musical onset detection in the complex domain. *IEEE Signal Processing Letters*, 11(6):553–556.

Bilmes, J. (1993). *Timing is of the Essence: Perceptual and Computational Techniques for Representing, Learning, and Reproducing Expressive Timing in Percussive Rhythm*. Master Thesis, MIT, Cambridge.

Desain, P. and Honing, H. (1991). Tempo curves considered harmful. A critical review of the representation of timing in computer music. In *Proc. International Computer Music Conference*.

Dixon, S., Gouyon, F., and Widmer, G. (2004). Towards characterisation of music via rhythmic patterns. In *Proc. International Conference on Music Information Retrieval*, pages 509–516.

Friberg, A. and Sundström, J. (2002). Swing ratios and ensemble timing in jazz performances: Evidence for a common rhythmic pattern. *Music Perception*, 19(3):333–349.

Gouyon, F., Fabig, L., and Bonada, J. (2003). Rhythmic expressiveness transformations of audio recordings: Swing modifications. In *Proc. Digital Audio Effects Conference*.

Gouyon, F., Wack, N., and Dixon, S. (2004). An open-source tool for semi-automatic rhythmic annotation. In *Proc. International Conference on Digital Audio Effects*, pages 193–196.

Janer, J., Bonada, J., and Jordà, S. (2006). Groovator — An implementation of real-time rhythm transformations. In *Proc. 121st Convention of the Audio Engineering Society*.

Laroche, J. (2001). Estimating tempo, swing and beat locations in audio recordings. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 135–138.

Paulus, J. and Klapuri, A. (2002). Measuring the similarity of rhythmic patterns. In *Proc. International Conference on Music Information Retrieval*.

Ramirez, R. and Hazan, A. (2006). A tool for generating and explaining expressive music performances of monophonic jazz melodies. *International Journal on Artificial Intelligence Tools*, 15(4):673–691.

Rosinach, V. and Traube, C. (2006). Measuring swing in irish traditional fiddle music. In *Proc. International Conference on Music Perception and Cognition*, pages 1168–1171.

Sandroni, C. (1996). Mudanças de padrão rítmico no samba carioca, 1917-1937. *Revista Transcultura de Música*, 2.

Widmer, G. (2002). Machine discoveries: A few simple, robust local expression principles. *Journal of New Music Research*, 31(1):37–50.

Witten, I. and Frank, E. (2000). *Data Mining: Practical machine learning tools with Java implementations*. Morgan Kaufmann, San Francisco.

Wright, M. and Berdahl, E. (2006). Towards machine learning of expressive microtiming in Brazilian drumming. In *Proc. International Computer Music Conference*.