

# VISUALIZATION TOOLS FOR MUSICAL TIMING APPLIED TO AFRO-CUBAN PERCUSSION

Matthew Wright<sup>1,2</sup>

W. Andrew Schloss<sup>1</sup>

George Tzanetakis<sup>2</sup>

University of Victoria, School of Music<sup>1</sup> and Department of Computer Science<sup>2</sup>  
Engineering/Computer Science Building (ECS), Room 504,  
PO Box 3055, STN CSC, Victoria, BC, Canada V8W 3P6

## ABSTRACT

Timing is perhaps the most fundamental aspect of music, and visualization tools can help in formulating hypotheses and exploring questions regarding musical timing. We present a series of novel graphical representations of musical timing, generated by computer from signal analysis of audio recordings and from listeners' annotations collected in real time. We have tested our methods on recordings of Afro-Cuban percussion with particular emphasis on the "clave" rhythmic patterns used for temporal organization. The proposed visualizations are based on the idea of *Bar Wrapping*, which is the breaking and stacking of a linear time axis at a fixed metric location.

## 1. INTRODUCTION

We propose a set of visualization techniques [16] that can assist researchers in understanding the complexities of timing in metric music. We would like to explore questions such as how do expert players differ from each other, and also from competent musicians who are not familiar with a given style; are there consistent timing deviations for notes at different metric positions; how does tempo change over the course of a recording, etc. Our goal is to assist exploratory data analysis using visualizations that can reveal interesting patterns and information about performance. Using these techniques, which are part of what we call *Computational Ethnomusicology* [17], we can ask and answer questions that were previously impossible or too tedious to explore without computers. Microtemporal analysis also can have pedagogical applications, e.g., see [8] for a review of visualization software designed to assist singers.

## 2. DATA SETS

So far we have focused on the Afro-Cuban *Rumba* [2, 4, 15], a musical and dance form in which an instrument named the *clave* (a pair of short sticks hit together) plays a repeating syncopated pattern (also called *clave*) to establish the ensemble's metric framework. *Rumba* generally uses *Rumba Clave*, shown in Figure 1.



Figure 1: One way to notate *Rumba Clave* (left) and *Son Clave* (right), two very common patterns in Cuba.

We gathered data by having subjects tap [11] directly on a laptop's built-in microphone via custom sample-accurate tap detection/logging software [20]. For most examples the subjects tapped *Rumba Clave* (not just the beats) while listening to a recording of a full ensemble.<sup>1</sup>

<sup>1</sup> We used *Cantar Bueno* from the 1993 recording *El Callejon De Los Rumberos* by *Yoruba Andabo* (DiscMedi DM203 CD), and *La*

## 3. BAR WRAPPING

In this genre, a typical performance is about 5 minutes long with tempos around 100-160 BPM, equivalent to 125-200 four-beat measures. With five notes per bar, a *clave* performance therefore typically consists of about 625-1000 time points. Simply plotting each point along a linear time axis would require either excessive width, or making the figure too small to see anything; this motivates bar wrapping. Conceptually, we start by marking each note of *clave* on a linear time axis. If we imagine this time axis as a strip of magnetic tape holding our recording, then metaphorically we cut the tape just before<sup>2</sup> each downbeat, so that we have 200 short pieces of tape, which we then stack *vertically*,<sup>3</sup> so that time reads from left to right along each row and then down to the next row, like text in languages such as English.

Figure 2 displays the times of the second author's taps to the *clave* part of *Cantar Bueno*, with the "tape" in light grey. The ragged right edge of the light grey "tape strips" comes from the fact that the tempo is not constant; in particular, the first few bars are significantly shorter in duration (i.e., faster) than the rest. (This particular recording begins sparsely with just the *clave* before drums and vocals layer in.) We can in fact interpret the ragged right edge as a kind of tempo curve,<sup>4</sup> with time running vertically (from zero at the top to the end of the piece at the bottom), and the horizontal axis displaying tempo (with faster tempi to the *left*, since they correspond to shorter bar durations). The right hand plot in Figure 2 is a tempo curve; we see that it is parallel to the ragged right edge of the "tape splice" representation.

*Cachamba* from the 1993 recording *Tambores Cubanos* by *Los Papines* (EGREM CD 0037).

<sup>2</sup> We need to wrap *before* the downbeat so that if a note on the downbeat is slightly early it will appear on the correct row to the left of its expected position, rather than at the right edge of the previous bar.

<sup>3</sup> This method of displaying vertically stacked musical material may appear similar to Ruwet's paradigmatic analysis found in musicological research [13], but in reality they are totally different: we are comparing repetitions of actual performed temporal patterns stacked according to metric structure, whereas paradigmatic analysis arranges fragments of symbolic notation by melodic or thematic resemblance. Winfree also charted data (sleep/wake cycles) with a linear time axis broken and stacked vertically [19], but in his case with the breaks coming at regular intervals, not irregular intervals like our successive bars of performed music.

<sup>4</sup> Although many researchers have plotted local tempo as a function of time (a "tempo curve") in the aid of musical analysis (for example, [2, 3, 14]), and many systems for synchronizing computer performers [9] or automatically generating expressive-sounding musical performances (see [18] for a review) are based on these tempo curves, there is considerable controversy about whether this concept "has a musical and psychological reality" [5]. An inextricably related question is the relationship between per-note expressive timing deviations and tempo, e.g., do the deviations scale linearly with tempo [7, 12]? Without taking a stance in this debate, our goal is simply to develop visualization tools that can help shed light on these questions.

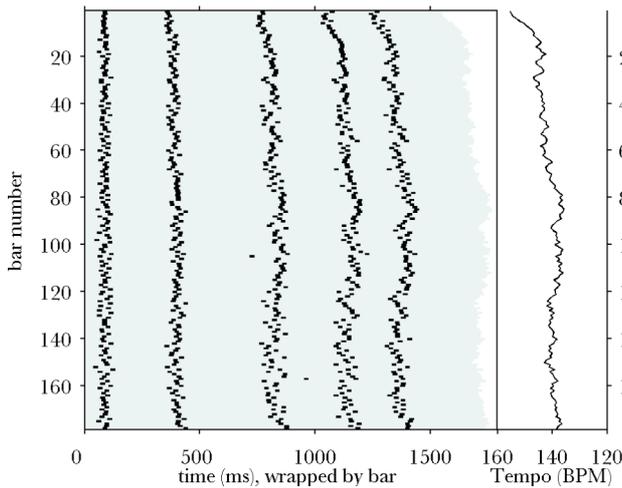


Figure 2: “Cantar Bueno” clave taps by Andy.

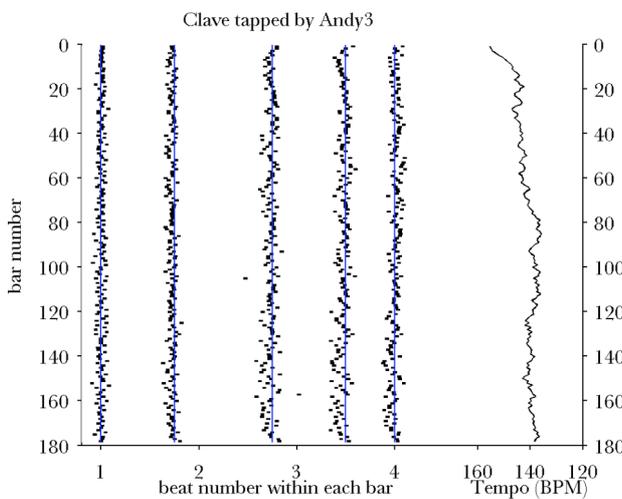


Figure 3: Same data as Figure 2, stretched. Vertical lines are ideal/quantized *Rumba clave*.

### 3.1. Duration of Each Bar

There are many ways to find the duration of each bar. Here we know *a priori* that each bar must have exactly five clave notes, so we can find each bar’s instantaneous tempo as simply the reciprocal of the time between each successive 5 notes. Computing tempo like this, however, would guarantee that the leftmost column, representing the time of the first clave note played in each bar, would always be a perfect straight line. (In other words, any deviation in the first note’s timing would become the reference point for the bar, time shifting all the other notes.) To get around this problem we compute each bar’s instantaneous tempo as described, but then smooth the resulting series with a 3-point (i.e., 3-bar) moving average.

### 3.2. Stretched Bar Wrapping

An advantage of the form of plot shown in Figure 2 is that the X axis is absolute time in milliseconds. A disadvantage is that while the downbeat will appear as a fairly straight vertical line, the subsequent notes of the bar will display both per-note deviations and the tempo

curve. This motivates “stretched” bar wrapping, in which the metaphoric one-bar tape segments are each stretched horizontally to make them equal in length, as shown in Figure 3. Now the displayed tempo curve is no longer redundant. The X axis no longer displays absolute time, but rather relative time within each bar regardless of tempo, so we have labelled it with beat number.

### 3.3. Bar Wrapped Energy Display

All of the plots so far have displayed tap time data, which contain the original expressive timing plus noise added by subjects’ tapping inaccuracies. One way to remove this added noise would be to display the time points output by an automatic onset detector [1], though the inaccuracies of every onset detector would then add a different source of noise. Another approach is to avoid the technological and epistemological challenges of finding discrete instants in music [20], and instead plot continuous functions of time with bar wrapping.

We created a matched filter from one isolated note of the clave, which has the effect of increasing the relative volume of the clave in the polyphonic mix. Figure 4 shows the bar-wrapped energy in each 512-sample (11.6 ms) frame of the output of this filter, with the greyscale shading displaying the top 60 decibels of energy.

Note that although this figure still relies on manual tapping to determine the downbeats, the data displayed are computed objectively from the signal, without any confounding noise from our imperfect tapping abilities. (The fact that we see substantially vertical lines indicates that our taps can’t have been too far off.) We could entirely eliminate manual tapping by using one of many algorithms for automatic beat and tempo extraction [6, 10] or bypass the general purpose tempo induction and directly find the tempo that makes the clave (whether detected instants or continuous features) best fit a (learned microtiming or ideal) clave template [21].

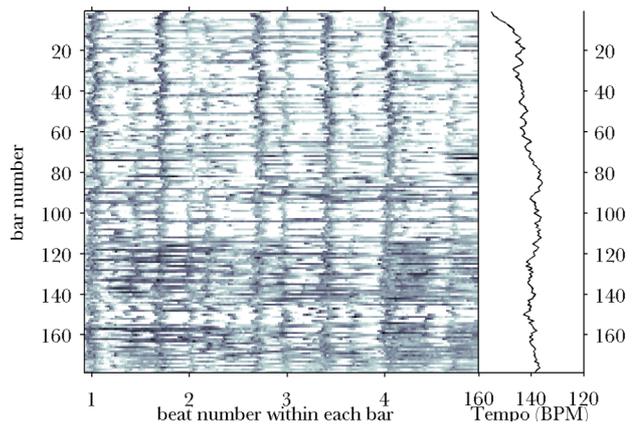


Figure 4: “Cantar Bueno” energy from matched filter.

## 4. COMPARING SERIES OF TIME POINTS

How can we compare the results of two or more people attempting the same rhythmic task? The first and second authors each independently tapped rumba clave along to *La Cachamba*, a recording in the rumba style but which does not include anyone playing the clave pattern.

First we matched each tap time with the closest one in the other dataset. It is well known that subjects tend to tap early (“negative mean asynchrony”) in these kinds of tasks by a subject-dependent average amount [11]; the mean difference was for Matt to tap 14 ms later than Andy. Figure 5 shows an average shifted histogram (ASH)<sup>1</sup> of the timing differences, a generally Gaussian-looking shape slightly skewed to the right.

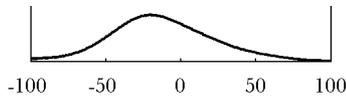


Figure 5: ASH of difference between tap times (ms). Positive means Matt tapped before Andy.

Figure 6 shows a separate ASH for each note of clave; we see that mean difference varies greatly for the various notes of clave, and the distribution of differences for the 4<sup>th</sup> note is much flatter than for the others.

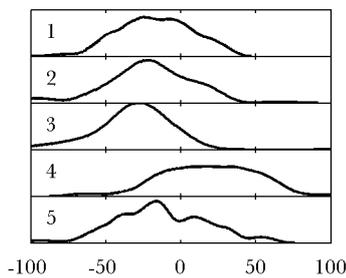


Figure 6: Same as Figure 5, but with a separate ASH for each note of clave.

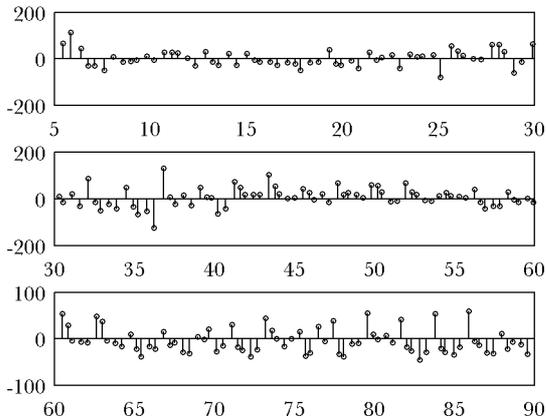


Figure 7: Matched timing difference stems. X is time (sec) of Andy’s tap, wrapped every 30 sec. Y is delay (ms) between Matt’s and Andy’s matched tap times, with positive meaning Matt before Andy. (Only the first 90 sec are shown due to space restrictions.)

Figure 7 represents another style of plot that displays the time difference between each pair of taps. We see that the timing differences have different behaviours in different segments. One important factor in this performance is the two bass solos at 0:30-0:40 and 1:30-1:42: the tempo drops substantially and all other instruments stop, making it much more difficult to “keep

<sup>1</sup> To make an average shifted histogram from a set of data points, first make a series of histograms from the same data points and with the same bin widths, but steadily varying the absolute starting position of all bins, then add them all together.

the beat.” Another option is to plot the two time series together on the same bar-wrapped display (Figure 8).

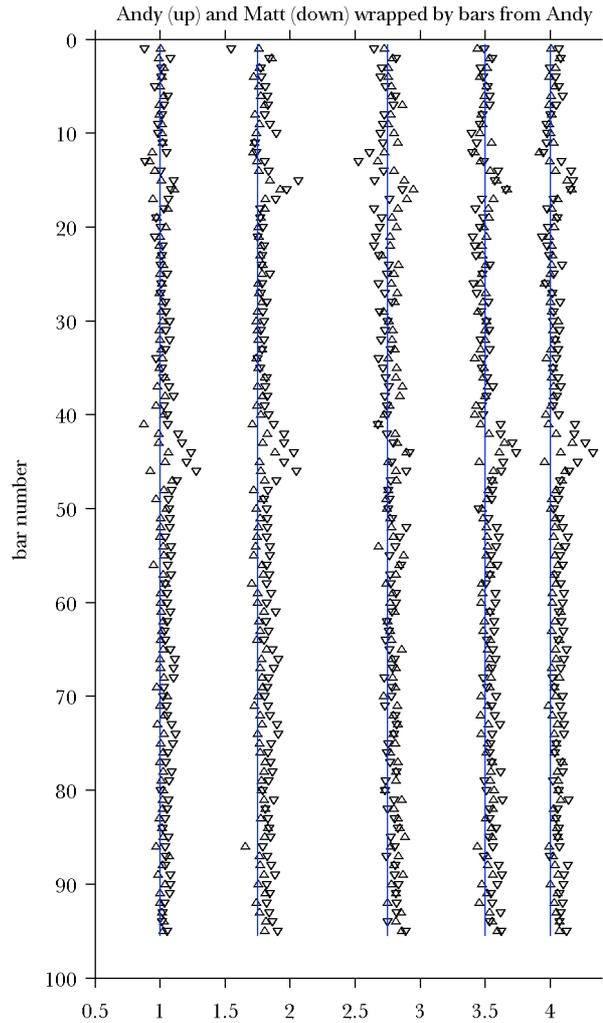


Figure 8: Comparison of two tapping “performances” for *La Cachamba*. X is beat number.

## 5. PEDAGOGICAL TOOL

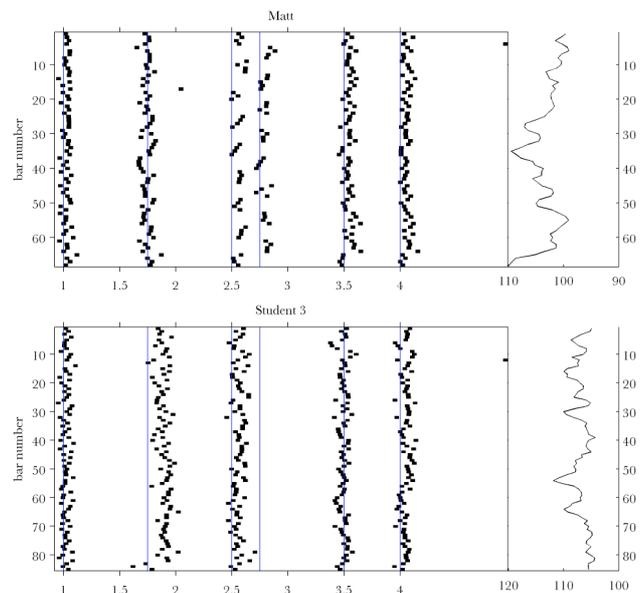


Figure 9: Trials alternating *rumba* and *son* clave.

We had some students (musicians but not experienced in Cuban music) alternate 4 bars each of *rumba* and *son* claves. For this tapping continuation task [11] they heard only 4 clicks at 100 BPM to set the tempo. We see from the tempo curves in Figure 9 that both subjects tended to rush, and then cyclically slowed back down and sped back up. Student 3 was usually quite late on the second note. In general the syncopated 2<sup>nd</sup> and 3<sup>rd</sup> notes are harder to play accurately and hence have higher variance. The student didn't understand the alternating clave task or couldn't do it. Matt tended to rush the second note in rumba clave (the wrong way to increase the gap in time between the 2<sup>nd</sup> and 3<sup>rd</sup> notes.)

## 6. CONCLUSIONS AND FUTURE WORK

We have presented a set of visualization tools and concepts that can be used to explore questions related to microtiming, with specific examples from the study of Afro-Cuban clave, including an expert tapping along with the clave of a recording, multiple subjects tapping the clave part along to a recording that doesn't include any instruments playing it, and students spontaneously generating the clave without any reference other than a 4-beat count-in. We recently collected data in Cuba from some of Cuba's top musicians, including Enrique Pla and Jorge Reyes, that we intend to analyze with these techniques. In the future we plan to collect more tapping data and for more pieces of music, and to combine our tools with techniques for extracting timing information directly from recordings. The tapping data collection and plotting software is freely available and can be requested by emailing the authors.

We hope that visual study of these plots will help us understand better how music is performed and what characterizes an expert performance, as well as shed light on the difference between students' and experts' performance for pedagogical purposes. The resulting knowledge about the intricacies of microtiming in human performance would be difficult if not impossible to obtain without computer assistance.

## 7. ACKNOWLEDGEMENTS

Ed Large, Michelle Logan, Graham Percival, Michael Spiro. We thank the Social Sciences and Humanities Research Council of Canada (SSHRC) for financial support.

## 8. REFERENCES

- [1] J. P. Bello, *et al.*, "A Tutorial on Onset Detection in Music Signals," *IEEE Transactions on Speech and Audio Processing*, 13(5) pp. 1035-1047, 2005.
- [2] J. Bilmes, "Timing is of the Essence: Perceptual and Computational Techniques for Representing, Learning, and Reproducing Timing in Percussive Rhythm," Master's thesis, MIT, 1993.
- [3] M. Clayton, *Time in Indian Music: Rhythm, Metre, and Form in North Indian Râg Performance*. Oxford: Oxford University Press, 2000.
- [4] Y. Daniel, *Rumba: Dance and Social Change in Contemporary Cuba*. Bloomington, IN: Indiana University Press, 1995.
- [5] P. Desain and H. Honing, "Tempo Curves Considered Harmful," in *Music, Mind, and Machine*, P. Desain and H. Honing, Eds. Amsterdam: Thesis Publishers, pp. 25-40, 1992.
- [6] F. Gouyon, *et al.*, "An experimental comparison of audio tempo induction algorithms," *IEEE Transactions on Audio, Speech and Language Processing*, 14(5), pp. 1832-1844, 2006.
- [7] H. Honing, "Evidence for tempo-specific timing in music using a web-based experimental setup," *J. Exp. Psych.: Human Perception and Performance*, 32(3), pp. 780-786, 2006.
- [8] D. Hoppe, *et al.*, "Development of real-time visual feedback assistance in singing training: a review," *J. Comp. Assisted Learning*, 22, pp. 308-316, 2006.
- [9] D. Jaffe, "Ensemble Timing in Computer Music," *CMJ*, 9(4), pp. 38-48, 1986.
- [10] M. F. McKinney, *et al.*, "Evaluation of Audio Beat Tracking and Music Tempo Extraction Algorithms," *JNMR*, 36(1), pp. 1-16, 2007.
- [11] B. H. Repp, "Sensorimotor synchronization: A review of the tapping literature," *Psychonomic Bulletin & Review*, 12(6) pp. 969-992, 2005.
- [12] B. H. Repp, *et al.*, "Effects of Tempo on the Timing of Simple Musical Rhythms," *Music Perception*, 19(4), pp. 565-593, 2002.
- [13] N. Ruwet and M. Everist, "Methods of Analysis in Musicology," *Music Anal.*, 6(1/2) pp. 3-36, 1987.
- [14] W. A. Schloss, "On the Automatic Transcription of Percussive Music: From Acoustic Signal to High-Level Analysis," Ph.D. dissertation, Stanford, 1985.
- [15] M. Spiro, *The Conga Drummer's Guidebook*. Petaluma, CA: Sher Music Co., 2006.
- [16] E. R. Tufte, *The Visual Display of Quantitative Information*, 2 ed. Cheshire, CT: Graphics Press, 2001.
- [17] G. Tzanetakis, *et al.*, "Computational Ethnomusicology," *Journal of Interdisciplinary Music Studies*, 1(2) pp. 1-24, 2007.
- [18] G. Widmer and W. Goebel, "Computational Models of Expressive Music Performance: The State of the Art," *JNMR*, 33(3) pp. 203-216, 2004.
- [19] A. T. Winfree, "Circadian Timing of Sleepiness in Man and Woman," *Am. J. Physiology: Regulatory, Integrative and Comparative Physiology*, 243(3) pp. R193-R204, 1982.
- [20] M. Wright, "The Shape Of an Instant: Measuring and Modelling Perceptual Attack Time with Probability Density Functions," Ph.D. Diss, Stanford, 2008.
- [21] M. Wright, *et al.*, "Analyzing Afro-Cuban Rhythm Using Rotation-Aware Dynamic Programming," submitted to *ISMIR*, Philadelphia, 2008.