Learning Microrhythm in Uruguayan Candombe using Transformers

Anmol Mishra[1] **Satyajeet Prabhu**[1] **Behzad Haki**[1] **Martín Rocamora**[1]

[1]Music Technology Group, Universitat Pompeu Fabra, Barcelona

anmol.mishra01@estudiant.upf.edu, martin.rocamora@upf.edu

ABSTRACT

Musicians rely on nuanced microrhythm, small variations in timing, dynamics, and other aspects, to create an expressive rhythmic feel in music performance. Electronic music production often attempts to replicate these qualities through algorithmic manipulations to achieve similar effects. In this work, we address the generation of microrhythm using a method that learns microtiming and dynamics from onset timing and strength annotations of drum performances. We frame microrhythm learning as a sequence modeling task, leveraging a Transformer-based model. Our focus is on Uruguayan candombe drumming, where we explore its rhythmic patterns at both the beat and rhythmic cycle levels. To evaluate the model's effectiveness in replicating the original microrhythm, we compare the mean, standard deviation and histogram intersection of timing deviations and dynamics values at each subdivision for the original and the generated data. The model is deployed as a VST enabling artists to incorporate candombe grooves into drum scores. With this work, we aim to bridge the gap between algorithmic rhythm creation and the expressive qualities of live performance, striving to produce music with the authentic grooves of various Latin American genres.

1. INTRODUCTION

Microtiming refers to subtle deviations from strict, grid-like timing in musical performance, where notes or beats are slightly ahead of or behind their expected positions. These minor, often almost imperceptible timing variations, along with other rhythmic aspects such as the dynamic envelope, contribute to the expressive quality of the music, creating a sense of feel or groove [1].

While microtiming occurs in many musical traditions, it plays a particularly significant role in genres emphasising rhythm and embodiment, such as jazz, funk, and various folk traditions [2]. For instance, swing ratios have been characterized in the context of jazz, where consecutive eighth notes are performed as long-short patterns [3]. Irish fiddle music also displays timing deviations at the eighth-note level [4], while Viennese Waltz at the quarter-note level; each third quarter-note in a bar is shorter [5].

Our focus in this work is candombe drumming, which is a defining element of Uruguayan popular culture [6]. It is performed using three drums of varying sizes and pitches - *chico*, *repique*, and *piano* - each playing a distinct rhythmic pattern. An additional pattern, shared by all three drums, is the *madera* pattern or *clave*, with functions similar to the timeline in Afro-Cuban and sub-Saharan African music traditions; serving as a means of temporal organization and synchronization. The candombe rhythm emerges from the interplay of these patterns, and its metric structure, a cycle of four beats with sixteen pulses, bears similarities to other Afro-Atlantic music traditions. The *chico* drum is the *timekeeper* of the ensemble. It maintains a repetitive pattern throughout the performance, establishing the foundational layer of the rhythm. Some prior works have investigated the microtiming properties of the rhythmic patterns in candombe music [7–9].

In this work, we aim to model microrhythm in candombe drumming. To do that, we use a Transformer-based model to learn the microtiming and dynamics from onset timing and strength annotations of real performances. We evaluate the model by comparing the mean, standard deviation and histogram intersection of the original and the generated onset data. To sample the learned groove, we export the model for inference inside a Virtual Studio Technology (VST) wrapper. Our VST is available here. [1]

2. METHODS

2.1 Dataset Preparation

The models were trained on the candombe dataset derived from Rocamora et al. [10]. This corpus has been compiled as a part of the Interpersonal Entrainment in Music Performance (IEMP) Project [11]. It consists of 12 performances, 9 trios and 3 quartets. The trios have three channels corresponding to chico (C), piano (P) and repique (R1) drums. The quartets have an additional channel for repique (R2) drum. For modeling all the performances together, we reduce the quartets to three channels by discarding one repique (R2) drum.

Data	Matrix	Values
Hits	$H_{32\times3}$	$h_{ij} \in \{0,1\}$
Velocities	$V_{32\times3}$	$v_{ij} \in [0,1]$
Offsets	$O_{32\times3}$	$v_{ij} \in [-0.5, 0.5)$

Table 1. Input/output sequence representation for 2-bar beats in 4/4 with 16th note resolution for a total of 32 time steps (i), and 3 drum voices (j).

Following [12], we represent our data as three matrices corresponding to hits (H), velocities (V) and offsets (O).

[1] https://github.com/dhunstack/candombe-groove-vst

This HVO matrix representation is a useful way to represent expressive percussion performances for training machine learning models. The matrices have size T time steps (one step per 16th note) and M instruments per time step. We quantize the onset annotations to get the hits matrix, scale the microtiming values between [-0.5, 0.5) for offsets, and normalize the onset annotation strengths between [0, 1] for velocities. Same as [12] and [13], we choose to learn these sequences over 2 bars, thus getting matrices of size T=32 and M=3. We obtain a total of 1070 bars (cycles) of onset data across 12 performances, which then forms the ground truth for our experiments. A summary of the HVO representation is provided in Table 1.

2.2 Model Architecture

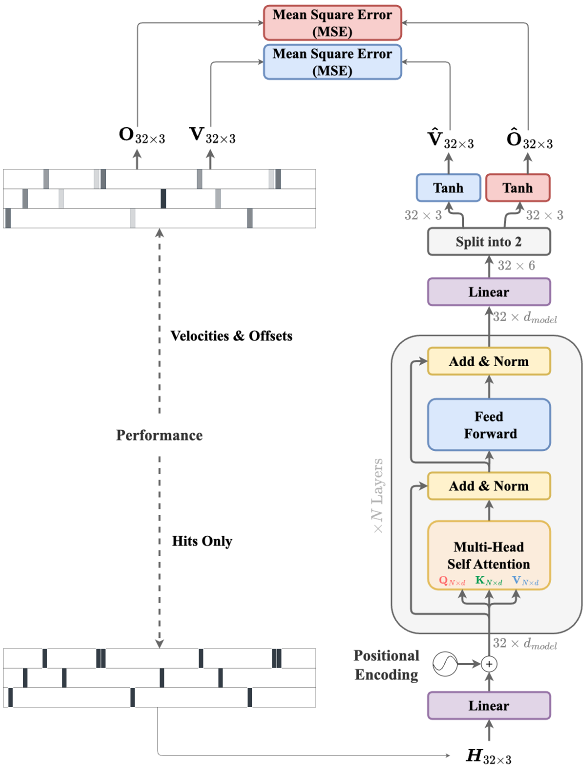

Figure 1. Model architecture

We attempt to model the problem of learning velocities and offsets from hits as a sequence-to-sequence problem. To be able to learn the microrhythm representation, we train transformer models that take the hits matrix as input and output velocity and offset matrices for the hits matrices.

The transformer architecture we use is based on the encoder section of the transformer architecture and is depicted in Figure 1. The drum patterns are processed over T=32 time steps, and a transformer encoder encodes the hit patterns into performance outputs. The encoder uses multi-head self-attention with 4 heads and a model dimension of 128. The feed forward layer also has a dimension of 128 and the number of encoder blocks is 11. The model jointly predicts velocities ($\hat{V}$), and timing offsets ($\hat{O}$). The outputs at each time step are split into (1) tanh for velocities $\hat{v}_t$ and (2) tanh for offsets $\hat{o}_t$. A square error loss is computed at each time step t for drum channel k as follows:

$$L_{t,k} = (v_{t,k} - \hat{v}_{t,k})^2 + (o_{t,k} - \hat{o}_{t,k})^2.$$

and mean is computed across all time steps and channels to obtain the final loss. During inference, the output velocities are scaled to [0, 1] and offsets are scaled to [-0.5, 0.5).

We generate train and validation splits for selecting models during hyperparameter optimization. We finally train the model on the entire dataset for deployment inside a VST plugin. The VST plugin takes input drum hits pattern as MIDI and adds groove to it. The humanized MIDI can be played within the plugin using default sounds, or dragged out to be used in DAWs.

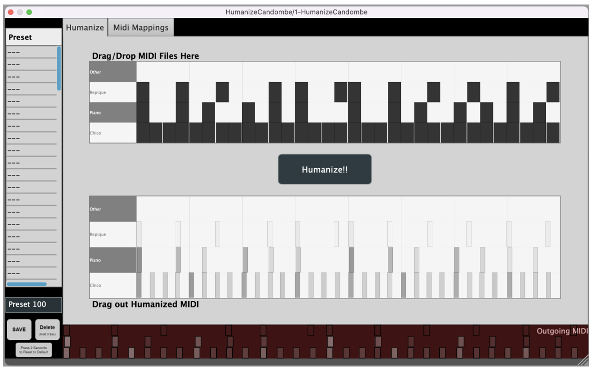

Figure 2. VST interface

3. EXPERIMENTS

For evaluation of our model's learned groove, we leverage the musicological understanding of rhythmic performance in candombe drumming. We select one representative performance from the dataset, extract its onset data (i.e. the hits) and infer the velocities and offsets using our model. Candombe drumming has repeating rhythmic structures at two different metric levels: the beat and the rhythmic cycle. The inference of the model is then analysed at these two temporal scales.

3.1 Distribution of Chico onsets in beat

In candombe, the chico assumes the role of the timekeeper playing repeating patterns at the level of the beat throughout the whole performance (see Figure 3) [8]. Thus our first experiment compares the distributions of chico onsets at the beat level in the actual and predicted data, to evaluate if we are able to learn its characteristic microtiming.

In Figure 4, we plot the chico onsets in actual (green) and predicted (red) data at the level of a beat. The pattern articulates the four sixteenth-note subdivisions of the beat (notated as .1, .2, .3 and .4). The means of offset values at each subdivision are also computed and displayed as a percentage of the beat duration, in a manner that represents average microtiming around that particular subdivision. We see that the model is able to learn the characteristic timing deviations of the chico onsets at each subdivision.

Table 2 provides the mean, standard deviation and histogram intersection values of the actual and predicted onset distributions at each subdivision computed across the entire dataset, showing the model is correctly capturing the microtiming information from the original data.

Accents form a major part of expressing groove. Since our model also captures velocity distributions, we compare actual and predicted velocities of chico onsets in the span of a beat. Figure 5 shows velocity values of chico at each subdivision of the beat for the actual (green) and predicted (red) data. The model appears to learn the velocity distribution in the ground truth. However, there is a discrepancy between the ground truth velocity data and the theoretical pattern shown in Figure 3, which clearly shows an accent at the second beat subdivision that is not reflected in the ground truth data and calls for further investigation.

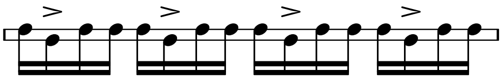

Figure 3. Chico pattern of the performance shown in Figure 4 in music notation (the lower line represents the hand, and the upper line represents the stick). The pattern is repeated for each of the four beats of the rhythmic cycle.

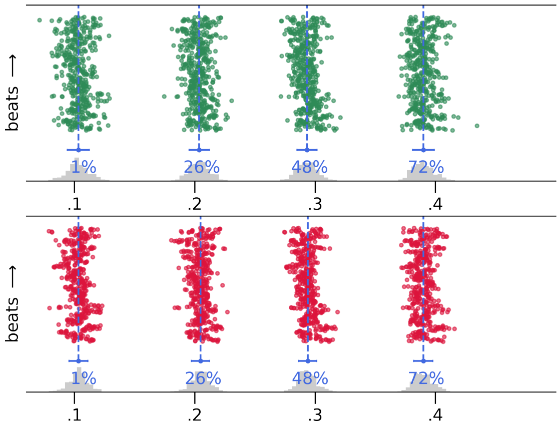

Figure 4. Chico actual (top) vs predicted onsets (bottom) for all beats in one of the performances of the dataset.

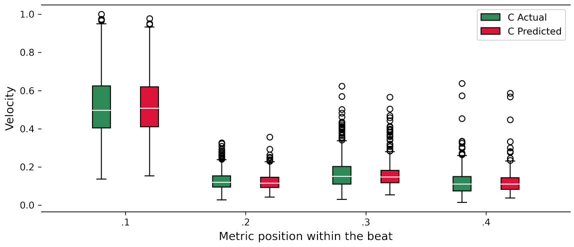

Figure 5. Predicted and actual velocities of chico onsets for all beats of the same performance of Figure 4.

Sub Div	Mean		Std		Hist Int
	Actual	*Pred.*	*Actual*	*Pred.*	
.1	0.01	0.01	0.02	0.02	0.84
.2	0.25	0.26	0.03	0.03	0.94
.3	0.48	0.49	0.02	0.02	0.81
.4	0.72	0.73	0.02	0.02	0.84

Table 2. Chico actual and predicted mean, standard deviation and histogram intersection of offset distribution across beats computed for the entire dataset.

3.2 Distribution of microtiming in cycle

We focus on the *madera* pattern whose duration spans over one rhythmic cycle, as depicted in Figure 7. The pattern is played by all the drums as an introduction to and preparation for the rhythm, but once the performance starts it is only played by the *repique* drum in between phrases [7]. The IEMP candombe dataset provides annotations for sections containing the madera pattern. So we consider the same performance used in Section 3.1 but now we focus on the cycles in which the repique drum plays the madera pattern. We infer on the 59 cycles of madera repique hits to identify whether such cycle level microtiming patterns are learned by the model. Figure 6 shows the distribution of repique onsets for the ground truth and the model inference on these sections. The model can learn the actual microtiming of the madera pattern. Interestingly, we observe that the onsets at the 4th subdivision of the first and fourth beats (1.4 and 4.4) are clearly ahead of the isochronous grid and are consistent with the microtiming patterns observed for the repique drum.

4. DISCUSSION AND CONCLUSION

In this work, we attempted to learn the microrhythm characteristics of Uruguayan candombe drumming. The onset timing and strength data were represented as hits, velocity, and offset (HVO) matrices and used to train a transformer model on 2-bar length sequences. The trained model was integrated into a VST for artistic use, allowing users to incorporate microrhythms to quantized drum hit patterns. We qualitatively evaluate the model at two temporal scales, beat and rhythm cycle, using the chico drum and the madera pattern played by the repique drum. The results obtained are promising, and we plan to conduct listening studies in future for comprehensive evaluation. We found the mean and standard deviation values of our model's learned microtiming distributions to resemble the original distribution in the dataset. The model also managed to learn the velocity distributions at the beat subdivisions.

The model architecture is capable of learning the groove (offset and velocity) distribution of any given dataset in the HVO representation, irrespective of music style. In future work, we plan to extend our model to learn the microtiming profiles of other Latin American music genres, seeking to contribute to more diverse tools for algorithmic rhythm creation in electronic music production.

5. ETHICS STATEMENT

Our utmost priority in this study is to honor the cultural heritage and safeguard the privacy of the communities represented in the IEMP collection. We acknowledge that each culture and tradition possesses unique nuances that cannot be fully captured through computational methods

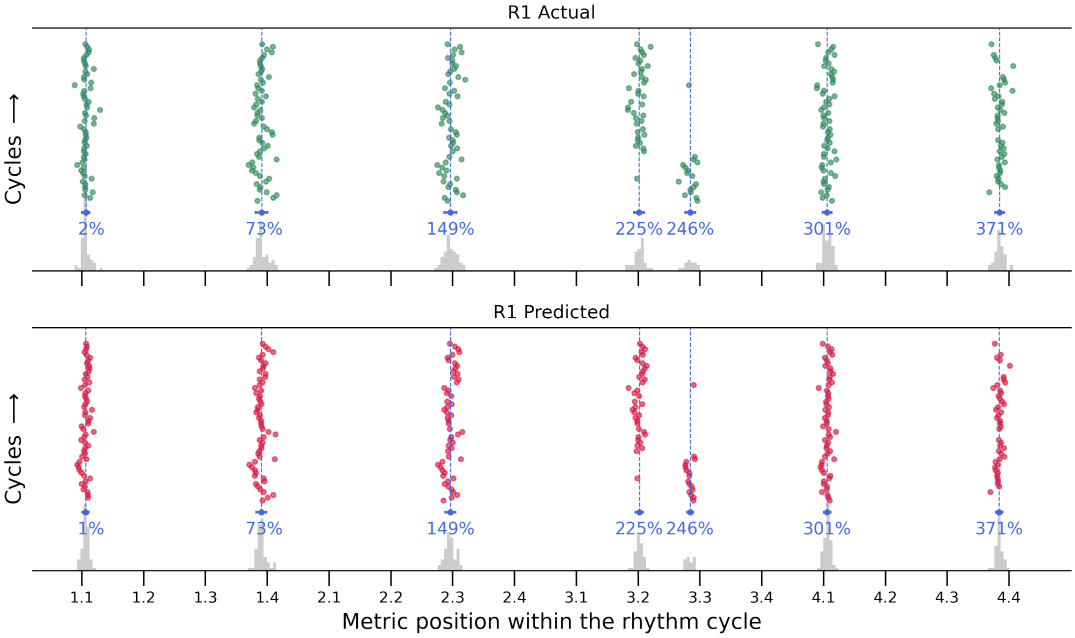

Figure 6. Example of madera patterns actual vs predicted onsets

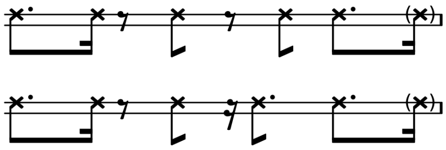

Figure 7. The two madera patterns played by the repique drum in the performance of Figure 6 shown in music notation (× symbol for madera sound). The performance starts with the top pattern and then switches to the bottom one.

alone and should not be simplified or generalized. Any tools developed using materials from this corpus are intended solely for academic use, aiming to enhance cultural diversity in music information research.

6. ACKNOWLEDGEMENTS

This work was supported by “IA y Música: Cátedra en Inteligencia Artificial y Música” (TSI-100929-2023-1), funded by the Secretaría de Estado de Digitalización e Inteligencia Artificial, and the European Union-Next Generation EU, under the program “Cátedras ENIA 2022 para la creación de cátedras universidad-empresa en IA”.

7. REFERENCES

[1] M. Davies, G. Madison, P. Silva, and F. Gouyon, “The Effect of Microtiming Deviations on the Perception of Groove in Short Rhythms,” *Music Perception*, vol. 30, no. 5, pp. 497–510, Jun. 2013.

[2] V. Iyer, “Embodied Mind, Situated Cognition, and Expressive Microtiming in African-American Music,” *Music Perception*, vol. 19, no. 3, pp. 387–414, Mar. 2002.

[3] A. Friberg and A. Sundström, “Swing Ratios and Ensemble Timing in Jazz Performance: Evidence for a Common Rhythmic Pattern,” *Music Perception*, vol. 19, no. 3, pp. 333–349, Mar. 2002.

[4] V. Rosinach and T. Caroline, “Measuring Swing in Irish Traditional Fiddle Music,” in *International Conference on Music Perception and Cognition*, 2006.

[5] A. Gabrielsson, “Interplay between Analysis and Synthesis in Studies of Music Performance and Music Experience,” *Music Perception*, vol. 3, no. 1, pp. 59–86, Oct. 1985.

[6] L. Ferreira, “An Afrocentric Approach to Musical Performance in South Black Atlantic: The Candombe Drumming,” *Trans : Transcultural Music Review = Revista Transcultural de Música, ISSN 1697-0101, Nº. 11, 2007*, Jan. 2007.

[7] L. Jure and M.Rocamora, “Microtiming in the Rhythmic Structure of Candombe Drumming Patterns,” in *Fourth International Conference on Analytical Approaches to World Music*, New York, USA, Jun. 2016.

[8] M. Fuentes, L. S. Maia, M. Rocamora, L. Biscainho, H. Crayencour, S. Essid, and J. Bello, “Tracking Beats and Microtiming in Afro-Latin American Music Using Conditional Random Fields and Deep Learning,” in *International Society for Music Information Retrieval Conference*, 2019.

[9] L. Jure and M. Rocamora, “Subir La Llamada: Negotiating Tempo and Dynamics in Uruguayan Candombe Drumming,” in *International Workshop on Folk Music Analysis*, Jun. 2018.

[10] M. Rocamora, L. Jure, B. Marenco, M. Fuentes, F. Lanzaro, and A. Gomez, “An Audio-Visual Database

of Candombe Performances for Computational Musicological Studies," in *Congreso Internacional de Ciencia y Tecnología Musical*, Sep. 2015.

[11] M. Clayton, K. Jakubowski, T. Eerola, P. E. Keller, A. Camurri, G. Volpe, and P. Alborno, "Interpersonal Entrainment in Music Performance: Theory, Method, and Model," *Music Perception*, vol. 38, no. 2, pp. 136–194, Nov. 2020.

[12] J. Gillick, A. Roberts, J. Engel, D. Eck, and D. Bamman, "Learning to Groove with Inverse Sequence Transformations," in *Proceedings of the 36th International Conference on Machine Learning*. PMLR, May 2019, pp. 2269–2279.

[13] B. Haki, M. Nieto, T. Pelinski, and S. Jordà, "Real-Time Drum Accompaniment Using Transformer Architecture," in *Proceedings of the 3rd International Conference on on AI and Musical Creativity*. AIMC, Sep. 2022.