EL BONGOSERO: A CROWD-SOURCED SYMBOLIC DATASET OF IMPROVISED HAND PERCUSSION RHYTHMS PAIRED WITH DRUM PATTERNS

Nicholas Evans* **Behzad Haki*** **Daniel Gómez-Marín** **Sergi Jordà**

Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain

nicholas.evans@upf.edu, behzad.haki@upf.edu, daniel.gomez@upf.edu, sergi.jorda@upf.edu

ABSTRACT

We present El Bongosero, a large-scale, open-source symbolic dataset comprising expressive, improvised drum performances crowd-sourced from a pool of individuals with varying levels of musical expertise. Originating from an interactive installation hosted at Centre de Cultura Contemporània de Barcelona, our dataset consists of 6,035 unique tapped sequences performed by 3,184 participants. To our knowledge, this is the only symbolic dataset of its size and type that includes expressive timing and dynamics information as well as each participant's level of expertise. These unique characteristics could prove to be valuable to future research, particularly in the areas of music generation and music education. Preliminary analysis, including a step-wise Jaccard similarity analysis on a subset of the data, demonstrate that this dataset is a diverse, non-random, and musically meaningful collection. To facilitate prompt exploration and understanding of the data, we have also prepared a dedicated website and an open-source API in order to interact with the data.

1. INTRODUCTION

Symbolic drum datasets derived from live performance typically feature a select number of experienced drummers improvising or playing a composed piece. However, given that rhythm perception is a fundamental human trait [1], we contend that an expressive crowd-sourced drum dataset representing a diverse range of musical expertise could offer unique research utility not fulfilled by existing datasets. Although it may be possible to compile a dataset of this nature by scraping the web for recorded performances, it would be unlikely that a web-scraped dataset would include expressive performance information along with the level of expertise of each performer.

This past year, our research lab participated in an exhibition at Centre de Cultura Contemporània de Barcelona (CCCB) centered around the history, ethics, and creative possibilities of Artificial Intelligence. More specifically, we were tasked with preparing a 6-month installation that would be included in the "Data Worlds" section of the exhibition, the purpose of which was to examine the role of data in generative systems and the methods employed to gather data. We addressed both of these aspects with a two-part installation. In the first activity, participants used a "bongo-like" two-voice MIDI pad to interact with a Variational Auto-Encoder (VAE) model. This model had the capability to transform the participant's tapped rhythmic sequences into symbolic multi-voice, expressive drum patterns [2,3], which were subsequently synthesized to audio. In the second activity, which serves as the focus of this paper, participants were given the opportunity to contribute to a crowd-sourced dataset that may later be used to improve the generative model they had just interacted with. They were invited to use the MIDI pad to tap along to a multi-voice drum pattern in a genre and tempo of their choosing. Providing minimal instructions, this task serves as an examination of how participants freely improvise alongside another rhythm. Upon completing the task, the participant could choose to contribute their tapped, improvised sequence to our public dataset or to submit nothing and delete their data.

In this paper, we present El Bongosero, a crowd-sourced expressive symbolic dataset consisting of 6,035 improvised tapped sequences performed by 3,184 participants with varying levels of musical expertise. Each sample contains expressive timing and dynamics information and is annotated with the participant's level of musical expertise, the genre of the selected pattern, the chosen tempo, the total duration to complete the activity, and a user-rating for their performance and how much they enjoyed the exhibit. We anticipate that this dataset can promote further research in the following areas:

- Advancing the development of more nuanced generative models capable of accommodating a range of skill levels.
- Facilitating music education studies focused on music understanding and rhythm expertise.

* Equal contribution

- Evaluating the proficiency, diversity, and creativity with which humans improvise rhythms.

Furthermore, the collection of data in this study adheres to rigorous ethical standards. Unlike other collection methods such as web-scraping, which may involve utilizing data in ways unintended by the original providers, our approach prioritizes clarity and consent with the participants throughout the entire process.

2. RELATED WORK

In this section, we will review other notable datasets consisting of human-performed recordings or synthesized web-scraped symbolic sequences. Reviewing these datasets aims to underscore the various applications and constraints associated with each approach.

The earliest open-source drum dataset we identified is the ENST-Drums dataset [4]. This is a fairly comprehensive dataset, consisting of around 225 minutes of annotated audio and video recordings of 3 live drummers. While still useful, this is significantly smaller than other datasets compiled via web-scraping or crowd-sourcing.

The TMIDT (Towards Multi-Instrument Drum Transcription) dataset, consisting of 259 hours worth of synthesized audio, was created via web-scraping every MIDI track from a freely available online collection [1] [5]. In a similar manner, the ADTOF (Automatic Drums Transcription On Fire) dataset, containing over 114 hours of annotated music, is constructed of openly shared [2] crowd-sourced symbolic annotations, typically a MIDI file, of real songs for use in rhythm games [6]. As such, this data does not contain detailed expressive information unlike Magenta's MAESTRO (MIDI and Audio Edited for Synchronous TRacks and Organization) [7]. Although MAESTRO consists of ten years of International Piano-e-Competition performances on a Yamaha Disklavier, it is relevant to include here as it is a large-scale, crowd-sourced dataset. This dataset, which has been used effectively in generative models, is comprised of over 172 hours of finely aligned ($\sim$3 ms) audio waveforms and expressive MIDI information.

Similar to MAESTRO, Magenta's Groove MIDI Dataset (GMD) is composed of 13.6 hours of aligned MIDI and (synthesized) audio of human-performed, expressive drumming [8]. The nature of this dataset has proven to be useful for predictive generative models such as GrooVAE [8] as well as for perceptual experiments such as TapTamDrum [9].

The TapTamDrum dataset was the result of an experiment in which 4 experienced drummers were given the task of reducing expressive, multi-voice drum patterns from Magenta's GMD to dual-voice representations. The resultant dataset includes 1,116 total dualizations annotated with expressive timing and velocity from 345 unique patterns.

[1] http://www.midiworld.com

[2] https://rhythmgamingworld.com/

	Format			Annotations		
Dataset	Audio	Symbolic	Human-Performed	Velocity	Genre	Level of Expertise
ENST	✓	✓	✓			
TMIDT	✓	✓				
ADTOF	✓	✓				
GMD	✓	✓	✓	✓	✓	
MAESTRO	✓	✓	✓	✓		
TapTamDrum	✓	✓	✓	✓		
MAST	✓		✓			
El Bongosero		✓	✓	✓	✓	✓

Table 1. Comparison of datasets.

Lastly, there is the MAST (Musical Aptitude Standard Test) Rhythmic Dataset, sourced from university examinations in which candidates were expected to reproduce a tapped rhythmic pattern after it had been played two times by a member of the jury [10]. Therefore, this audio dataset includes 2,681 recordings of jury members performing the target rhythm, along with 1,040 recordings of student attempts annotated with their grade (pass or fail).

Table 1 offers a comparison of the datasets based on two key attributes: format and annotations. Format indicates whether the dataset comprises audio or symbolic samples and whether these samples are derived from recorded human performances. Annotations, on the other hand, encompass details such as the presence of velocity annotations for each onset, the genre of the sample, and the level of expertise of the performer. As shown in the table, El Bongosero is the only dataset that annotates the performer's level of expertise.

3. METHODOLOGY

As mentioned above, the installation consisted of two parts. In the first part, participants were to engage with a generative model. In the second part, they were asked to contribute to a dataset that may be used for training future iterations of the generative model used in the first part. The focus of this paper is on the latter part of the installation, specifically, the collection of a symbolic dataset of rhythmic improvisations played alongside a selected number of drum patterns.

As the installation was to be used in a public exhibition space, it was imperative to design an interface that could accommodate a broad spectrum of participants without assuming specific technical or musical expertise. To this end, we made several decisions in designing the data collection stage of the installation. First, we ensured that the interactive elements in the system were nearly identical to the first stage of the installation. This strategy was aimed at eliminating any need for participants to acquaint themselves with the mechanics of the system. Second, we minimized the instructions provided to participants. The intention here was to encourage the participants to improvise freely using their personal intuition and creativity, rather than adhering to a very specific procedure. Lastly, before initiating the second part of the installation, we informed the participants that we would ask for their consent to contribute to our dataset at the conclusion of this activity. The purpose of this approach was to ensure that participants

were fully aware of this aspect of the activity prior to deciding if they wished to interact. While the primary motivation behind implementing this level of transparency was to adhere to ethical principles, it was also our aim to foster more open and genuine interaction with the installation by making participants feel valued and secure.

In the following subsections we discuss the tasks presented to the participants (3.1), the installation setup (3.2), and the drum pattern curation process (3.3).

3.1 Overview of Tasks

Figure 1 provides an overview of the tasks involved in the installation.

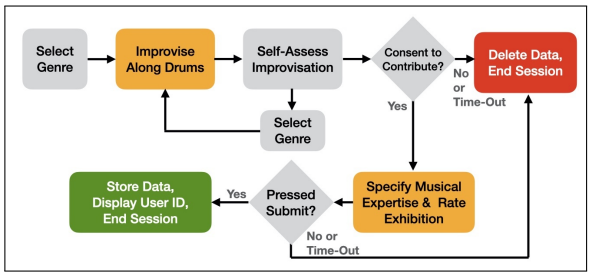

Figure 1. Flow-chart of the installation steps

The main objective of the data collection part of the installation was to present the participants with a randomly selected drum pattern and ask them to improvise alongside the pattern. To accommodate the diverse musical tastes and backgrounds of the participants, we allowed the participants to select the genre of the drum pattern.

Once the genre was selected, the improvisation environment was initiated. In this environment, the participant was presented with a looping 2-bar drum pattern and was asked to improvise alongside it using a provided two-voice MIDI pad.

The starting tempo of the session would be associated with the selected drum pattern, however, to accommodate participants of various skill levels, we allowed them to modify the tempo of the session. Each tap on the MIDI pad was recorded in a real-time looping 2-bar buffer, allowing participants to listen to previous taps and overdub additional taps. Lastly, participants were given as much time as needed.

Once the participant stopped the session, we asked them to self-assess their performance using a 5-level Likert scale. Once the assessment was provided, the participant was given two choices. They could select a new pattern to improvise alongside or they could finish the session. Once the participant finished the session, they were asked if they wish to contribute to the dataset. If they decided not to contribute, the session would end. Otherwise, they were presented with a brief questionnaire and then subsequently asked to press a button to explicitly submit their data.

Given that this was a public installation, we presented consenting users with only 2 questions: (1) "How would you assess your level of musical expertise?", and (2) "How much did you enjoy this exhibit?". In order to assure the participants that the only aim of the installation was to collect improvisations, as opposed to metadata related to the participants, we avoided any demographic questions on gender, age, and occupation. Recognizing the vast spectrum of musical proficiency among participants, from novices to experienced musicians, the question on "Musical Expertise" aimed to contextualize the improvisational outcomes within a broader narrative of skill and experience. Furthermore, we recognize that the term "expertise" in this context may be subject to interpretation, with participants not necessarily associating it solely with musical proficiency. Our intention was to allow participants to define "expertise" based on their own understanding within the context of their improvisations.

Once the final questions were answered, the submission button would be enabled to finalize the contribution. Note that participant data was only added to the dataset if the "Submit" button was pressed. That is, we wanted to ensure that participants explicitly consented to the contribution. In any case that the participants left the session mid-experiment, explicitly chose not to contribute, or forgot to press the submission button, their data was immediately deleted.

3.2 Installation Setup

The installation, shown in Figure 2, consisted of a touch screen application and an *Embodme*'s *ERAE Touch MPE* controller [3] for registering the improvisations.

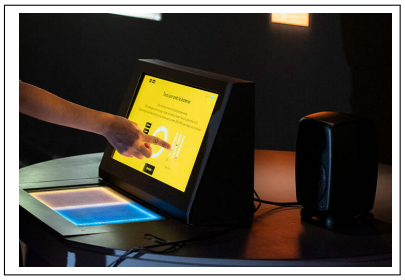

Figure 2. A photo of the installation (photo credit: CCCB)

The touchscreen application was used to prompt the participant and to allow the user to navigate through different stages of the installation. During the improvisation sessions, the graphical interface visualized the “bongo” performance using a circular representation of the 2-bar looping buffer filled with each tap onset registered on the MIDI pad (refer to the center of Figure 3 for the actual graphic representation). In this section, the participant was allowed to remove a specific onset by double tapping its location in the buffer, or to remove all of the onsets using a dedicated button; however, they were not allowed to reposition any registered onsets. In other words, the timing of the onsets were only to be associated with the timing registered from the performed taps on the MIDI pad.

The visual interface was implemented using the *PyQt5* [4] Graphical User Interface (GUI) toolkit. To ensure

[3] https://www.embodme.com/erae-touch
[4] https://www.qt.io/

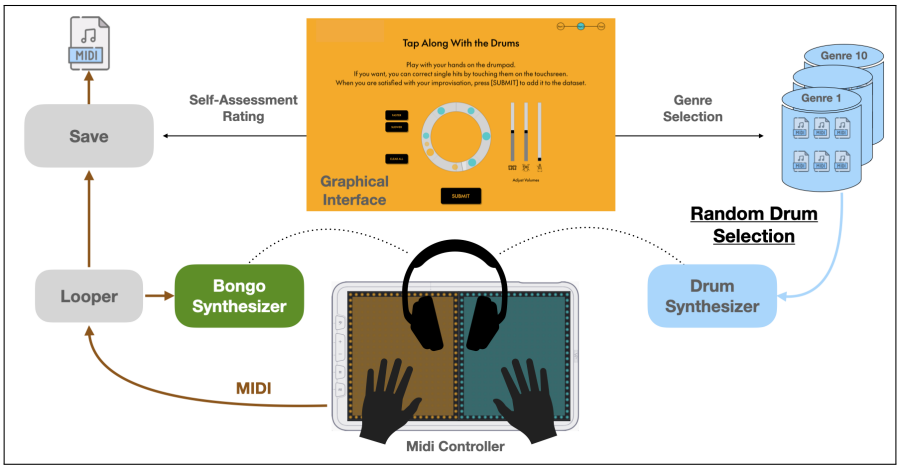

Figure 3. Installation setup

precise synchronization between the synthesis and playback of the source drum pattern and the MIDI recording of each improvisation, we developed a C++ backend using *JUCE* framework [5] . The recorded performances are provided in two formats: (1) a linear sequence preserving the timing of each tap onset throughout the activity's entire duration, and (2) a 2-bar loop preserving the timing of each overdubbed onset within the 2-bar recording buffer.

As shown in Figure 3, the touch MIDI pad was customized into two distinct pitch regions, with the right region pitched lower than the left. Each "bongo" tap was displayed on the graphical interface using a circle located within the 2-bar looping buffer. Each circle was color-coded to match the region from which the tap onset was registered and the radius of the circle was correlated with the velocity of the registered tap onsets.

In order to ensure that participants could properly listen to the sounds of the installation and to reduce the the possibility of a participant feeling that their performance was being judged by other visitors, we decided to only provide headphones to the participants rather than loudspeakers [6] .

Lastly, dedicated sliders were provided on the interface to allow the participant to adjust the volumes of the bongo sounds and the drum sounds as needed. Moreover, a dedicated slider (initially muted) was also provided to participants to utilize a synchronized metronome track if needed.

3.3 Drum Pattern Selections

The source drum patterns were in a 4/4 metric, selected from a large in-house collection of over 200,000 MIDI files, which included both open-source and proprietary MIDI collections. The MIDI files in this collection were divided into 10 genres: Afrobeat, Afrocuban, Bossanova, Disco, Electronic, Funk, Hiphop, Jazz, Rock, and Soul.

For each pattern in the collection, we extracted the rhythmic features provided in *GrooveToolbox* [11] and *Rhythm Toolbox* [12]. The extracted features were normalized and subsequently mapped to a two-dimensional space using Principal Component Analysis (PCA). For each genre, the mapped values were grouped into 100 clusters using k-means clustering method, and subsequently, a single pattern was randomly selected from each cluster.

In order to ensure a small subset of patterns with a sizeable collection of varied responses per pattern, we opted to limit the Electronic genre to 16 patterns. We selected this genre as we suspected it would be the most popular choice among participants. [7]

4. DATASET

In this section the contents of the collected dataset are described. A total of 4 variables were recorded per participant (ID, number of attempts, level of musical expertise, and exhibition rating). The ID was assigned in sequential order and each participant could attempt multiple improvisations. For each attempt 8 variables were collected (attempt duration, assessment time, attempt tempo, drum pattern, genre, improvisation pattern, level of expertise, and exhibition rating).

Table 2 presents a summary segmented by participants' level of musical expertise. The mean level of musical expertise is 2.95 (std = 1.25). The most common level of musical expertise is level 2 (915 participants) and the least common is level 1 (392 participants). The mean number of attempts per participant is 1.89 (std = 1.32). The highest amount of attempts were carried out by participants of musical expertise level 2 (1692) and the lowest amount of attempts were carried out by participants of musical expertise level 1 (691). The mean number of unique patterns presented per level is 591.8 (std = 103.8). Participants with musical expertise of level 2 were exposed to the largest number of unique musical patterns (720) while participants with musical expertise of level 1 were exposed to the least number (433). The mean number of attempts increases with the level of musical expertise, as participants with more expertise made more attempts on average.

Figure 4 presents the number of attempts per genre. The mean number of attempts per genre is 603.5 (std =

[5] `https://juce.com/`

[6] A loudspeaker was available in the setup (as in Figure2), however, it was only used in special occasions decided by CCCB organizers.

[7] The results discussed in next section confirm this speculation.

	Level of Musical Expertise				
	1	2	3	4	5
No. participants	392	915	805	594	478
Attempt count	691	1692	1536	1074	1042
Mean no. attempts	1.76	1.85	1.91	1.81	2.18
Unique patterns	433	720	691	562	553

Table 2. Summary of participants, attempts, patterns, and musical expertise.

169.67). The Electronic and Rock genres represent the highest amount of attempts (919 and 850 respectively) while the Soul genre represents the lowest amount (371).

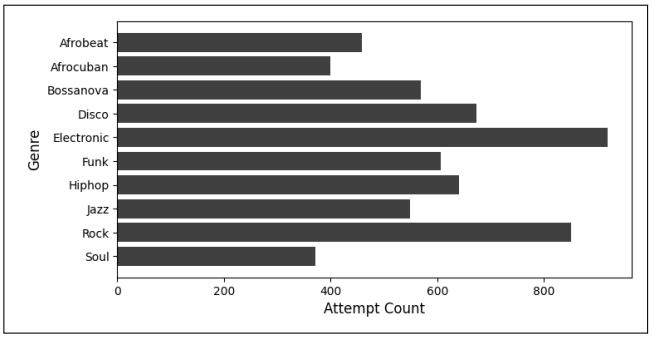

Figure 4. Histogram of attempts per genre.

Figure 5 presents an overview of the number of attempts per level of expertise and genre. The combination with the highest number of attempts is the Rock and Electronic genres combined with expertise levels 2 and 3. The combination with the least attempts is the Afrobeat genre and expertise level 1.

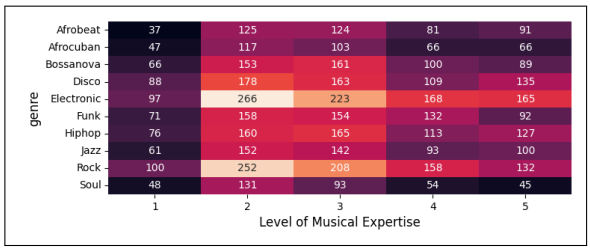

Figure 5. Attempts per genre and level of expertise.

As described in Section 3.3, all drum patterns used in the installation are in the 4/4 metric. Figure 6 presents the step densities of both the original patterns and the recorded improvisations, obtained by adding onsets at each step and dividing by the step with most onsets. Patterns are wrapped to 16 steps for convenience and onsets quantized to the closest 16th note. In order to establish a comparison, the normalized theoretical metrical weight is displayed. Notice how densities at each inter-pulse group of steps (0-3, 4-7, 8-11, 12-15) complies with the "high, low, mid, low" contour expressed in the theoretical metrical weights for a 4/4 rhythmic pattern. This suggests that participants consistently induced a meter from the source drum patterns. The improvised rhythms by the participants (Figure 6 below) showcase the same general intra-pulse contour with two differences. First, the low contours are higher (uneven steps), and second, the first step contains less onset density than its intra-pulse set (steps 1, 2 and 3). However steps 1, 2, and 3 comply with the "low-mid-low" contour observed in the original pattern's intra-pulse density. We believe many of the participants were slightly inaccurate at the beginning of the loop, thus causing onsets intended for the first step to be played early, registering in the last step of the previous bar, or played late, registering in the second step of the current bar.

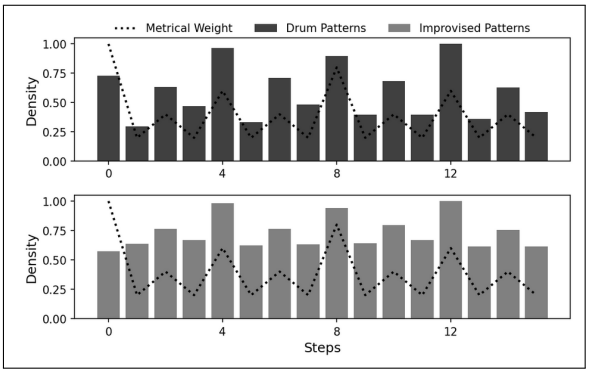

Figure 6. Onset density and metrical weight per step.

Typical of crowd-sourced datasets, these general observations led us to identify some instances of crowd-driven bias. Specifically, we observed a preference for two genres (Rock and Electronic) out of a possible ten, along with a distribution of musical expertise leaning towards mid-low levels. On the other hand, the general compliance of participants' patterns with metrical expectations suggest that their improvisations were carried out under expected pulse-entrainment conditions. Thus, in general, it seems that the data gathered corresponds to sensory-motor activities and not a random collection of taps.

5. PRELIMINARY INSIGHTS

As explained in Section 3.3, we limited the number of Electronic patterns to 16 in order to increase the number of reproductions per pattern for different levels of musical expertise. The brief preliminary analysis presented here focuses solely on the Electronic genre and explores the patterns used and assesses the similarity between the reproduced drum patterns and the participants' improvisations.

The number of attempts per Electronic pattern is presented in Figure 7. The range of attempts fluctuates between 46 (pattern 1) and 65 (pattern 9). The mean is 57.44 attempts and the standard deviation is 5.33.

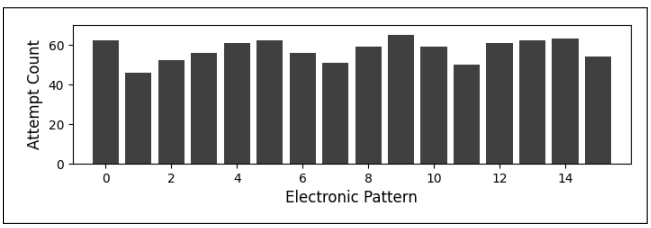

Figure 7. Number of attempts per Electronic pattern.

In order to establish a first metric that can account for comparing the multi-voice drum patterns with the participants' tapped improvisations, the Jaccard similarity metric is used. Jaccard is a common similarity metric used in data analysis, especially suited for comparing two sets of elements. The simplest implementation of the metric is the quotient of the sum of the intersection elements with the sum of the union elements. The more elements in common between the intersection and the union, the closer the Jaccard similarity gets to 1.

We implemented Jaccard similarity comparing a step-wise flattened version of the drum pattern and participants' improvisations. The rationale of this metric is: in a step where (at least) one onset in the drum pattern is observed, (at least) one onset is expected in the participant's improvisation. The intersection is composed of steps with onsets in the drum pattern that coincide with steps with onsets in the improvised pattern. The union comprises all steps from the pattern and the improvisation containing an onset. If a participant produces an onset every time the drum pattern produces an onset, Jaccard similarity is equal to one. If all of a participant's onsets are on steps where the drum pattern is silent, Jaccard similarity is 0.

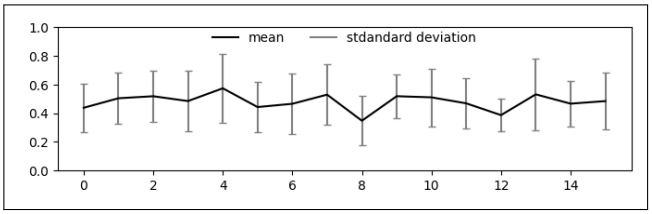

Figure 8. Jaccard similarity means and standard deviation for every pattern in the Electronic music genre.

Figure 8 shows that all participant improvisations to Electronic patterns exhibit a very similar mean (from 0.348 for pattern 8 to 0.574 for pattern 4) and spread standard deviation (from 0.115 for pattern 12 to 0.249 for pattern 13). There is no apparent agreement (there are no high mean values) towards any of the Electronic patterns, suggesting diversity in the improvisations for all patterns of this genre.

For more detail, Figure 9 presents a spread of similarity by Electronic pattern and level of musical expertise. The expertise level with the highest Jaccard similarity sum for all patterns (5.98) is level 1 while level 2 has the lowest Jaccard similarity sum for all patterns (5.63). The most diverse case, signified by a low mean Jaccard similarity (0.18), is observed in improvisations for pattern 7 performed by participants of expertise level 5. On the other hand, improvisations with the most average agreement with the reference, signified by a high mean Jaccard similarity (0.51), is observed in improvisations for pattern 10 performed by participants of expertise level 1.

The consistent mid agreement presented in Figure 8 and Figure 9 suggests improvisations were not exhibiting an automatic onset-for-onset behavior. On the contrary, there seems to be rich musical behavior to be explored within the Electronic genre.

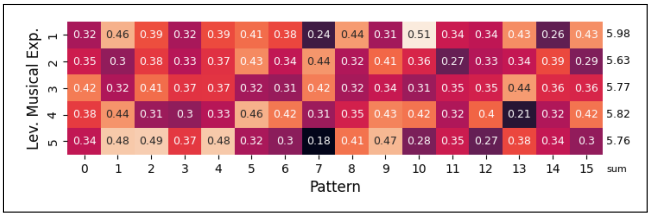

Figure 9. Jaccard similarity among all improvisations for different musical levels and pattern.

6. DISCUSSION AND CONCLUSION

In this paper, we introduced El Bongosero, a crowd-sourced dataset consisting of 6,035 tapped improvised rhythms performed by a total of 3,184 participants with varying levels of musical expertise. The improvisations are collected across 10 genres, each corresponding to a set of 100 unique drum patterns selected for that genre, except for the Electronic genre, which includes 16 samples. The main focus of this work was the collection, curation, and organization of this data from an interactive public exhibit.

Preliminary analysis, including a step-wise Jaccard similarity analysis on the Electronic genre data subset, demonstrate that this dataset is a diverse, non-random, and musically meaningful collection of improvised rhythms. Through our review of existing datasets, we identified the unique qualities of El Bongosero that could make it particularly useful for music generation and music education research. More specifically, it is the combination of the sheer number of participants, the diverse range of participants' level of musical expertise, and the inclusion of expressive performance information that distinguishes this dataset.

For example, in the context of a model that generates music based on a rhythmic input, a skilled musician may have different expectations than a novice musician regarding how a model should interpret an input rhythm or how it should respond to subtle variations in timing or dynamics. Integrating a diverse crowd-sourced dataset, such as El Bongosero, with the development of generative models could prove to be an effective approach to constructing more nuanced models that are capable of adjusting to individuals with varying skill levels.

Similarly, in the context of music education, deep analysis of El Bongosero may allow educators to gain insights into the learning trajectory of percussion students and help them to better develop a curriculum that supports skill development. As an evaluation tool, this dataset could serve as a valuable resource for developing assessments and criteria for drumming proficiency. Furthermore, researchers may be able to identify key indicators of musical growth and proficiency by comparing performances across different levels of expertise.

To conclude this work and facilitate prompt exploration of the collected data, we have prepared a dedicated website and an open-source API available at:

`https://elbongosero.github.io/`

7. ETHICS STATEMENT

Conscientious consideration of ethical principles has been central throughout this project. We recognize that as researchers it is our responsibility to ensure that there is complete transparency of the collection process and that participants have a full understanding of their involvement. Accordingly, this study attempts to uphold ethical standards at every stage of the data life cycle, from collection to utilization.

Firstly, the installation was crafted so that prior to starting the activity, participants were explicitly notified that we would later request their permission to store the data they generate while interacting with the exhibit. At the end of the activity, participants had to explicitly consent once more in order to be included in the dataset. If they declined, or took no action, their data was not stored.

In addition to ensuring explicit consent from participants, we also gave careful consideration to exactly which data we collected. To this end, we opted to collect no personal or demographic information from the participants. The collected data from consenting participants included only their interactions with the installation, resulting in a symbolic representation of their tapped improvised pattern, along with their responses to two questions: "How would you assess your level of musical expertise?" and "How much did you enjoy this exhibit?".

Moreover, ethical considerations extend beyond the initial data collection phase to encompass the subsequent use and application of the data. As stewards of this dataset, we are committed to employing the collected data solely for academic research purposes, ensuring that it is used in a manner consistent with what was communicated to participants.

8. ACKNOWLEDGMENTS

This research was partly funded by the Secretaría de Estado de Digitalización e Inteligencia Artificial, and the European Union-Next Generation EU, under the program Cátedras ENIA 2022. "IA y Música: Cátedra en Inteligencia Artificial y Música" (Reference: TSI-100929-2023-1).

9. REFERENCES

[1] H. Honing, "Without it no music: beat induction as a fundamental musical trait," in *Annals of the New York Academy of Sciences*, vol. 1252, 2012, pp. 85–91.

[2] B. Haki, M. Nieto, T. Pelinski, and S. Jordà, "Real-Time Drum Accompaniment Using Transformer Architecture," in *Proceedings of the 3rd Conference on AI Music Creativity (AIMC)*, September 2022.

[3] N. Evans, B. Haki, and S. Jorda, "GrooveTransformer: A Generative Drum Sequencer Eurorack Module," in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, September 2024.

[4] O. Gillet and G. Richard, "ENST-Drums: an extensive audio-visual database for drum signals processing," in *Proceedings of 7th International Society for Music Information Retrieval Conference (ISMIR 2006), Victoria, BC, Canada*, October 2006, pp. 156–159.

[5] R. Vogl, G. Widmer, and P. Knees, "Towards multi-instrument drum transcription," in *Proceedings of the 21th International Conference on Digital Audio Effects (DAFx18), Aveiro, Portugal*, September 2018, pp. 57–64.

[6] M. Zehren, M. Alunno, and P. Bientinesi, "ADTOF: A large dataset of non-synthetic music for automatic drum transcription," in *Proceedings of the 22nd International Society for Music Information Retrieval Conference (ISMIR 2021), Online*, November 2021, pp. 818–824.

[7] C. Hawthorne, A. Stasyuk, A. Roberts, I. Simon, C.-Z. A. Huang, S. Dieleman, E. Elsen, J. Engel, and D. Eck, "Enabling factorized piano music modeling and generation with the maestro dataset," in *Proceedings of the International Conference on Learning Representations (ICLR 2019), New Orleans, Louisiana, USA*, May 2019, pp. 9092–9103.

[8] J. Gillick, A. Roberts, J. H. Engel, D. Eck, and D. Bamman, "Learning to groove with inverse sequence transformations," in *Proceedings of the 36th International Conference on Machine Learning (ICML 2019), Long Beach, California, USA*, vol. 97, June 2019, pp. 2269–2279.

[9] B. Haki, B. Kotowski, C. L. I. Lee, and S. Jordà Puig, "TapTamDrum: a dataset for dualized drum patterns," in *Proceedings of the 24th Conference of the International Society for Music Information Retrieval (ISMIR 2023), Milan, Italy*, November 2023, pp. 114–120.

[10] F. Falcao, B. Bozkurt, X. Serra, N. Andrade, and O. Baysal, "A dataset of rhythmic pattern reproductions and baseline automatic assessment system," in *Proceedings of the 20th Conference of the International Society for Music Information Retrieval (ISMIR 2019), Delft, The Netherlands*. International Society for Music Information Retrieval (ISMIR), November 2019, pp. 439–445.

[11] F. Bruford, O. Lartillot, S. McDonald, and M. B. Sandler, "Multidimensional similarity modelling of complex drum loops using the GrooveToolbox," in *Proceedings of the 21th International Society for Music Information Retrieval Conference (ISMIR 2020), Montreal, Canada*, October 2020, pp. 263–270.

[12] D. Gómez-Marín, S. Jordà, and P. Herrera, "Drum rhythm spaces: From polyphonic similarity to generative maps," in *Journal of New Music Research*, vol. 49, no. 5. Taylor & Francis, 2020, pp. 438–456.