Reviewer #1

1. Tab. 1: Consider replacing "+" by a checkmark.
    - **[r] Updated in manuscript**

2. Sec. 2: Consider replacing "consider" by "review".
    - **[r] Updated in manuscript**

3. Sec. 2: Highlight more clearly, which datasets were originally intended to be used for drum transcription
    - **[r] The content here is not focused on a specific task. As such, we did not specify which datasets were intended for drum transcription.**

4. Sec. 2: Why is the MAESTRO dataset listed here as it is a piano transcription dataset
    - **[r] Added clarification in the text that MAESTRO is relevant: “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.”**

5. Sec. 3.1: Liine 208: singular ("participant") vs. line 209: plural ("they")
    - **[r] We are using they for the sake of gender neutrality**

6. Fig. 2: "|" not clear in the caption (is this a typo?)
    - **[r] We did not find “|” in the caption**

7. Fig. 9: The text in the colored plot is a bit hard to read (in a colored printout, consider increasing the fontsize slightly)
    - **[r] Font size increased**

8. references: Add page range for references 4-11
    - **[r] Page ranges were updated in manuscript**

Reviewer #2

1. Section 2: I'm unsure about referencing MAESTRO here amid mostly drumming datasets. Moving it to the end, as a related dataset, could improve the cohesion here.
    - **[r] See response to reviewer #1, comment 4.**

2. Section 3.3, lines 290-295: Please include here the number of tracks in the "large in-house MIDI collection".
    - **[r] Updated the paragraph to: “... collection of over 200,000 MIDI files...”**

3. Lines 296-303: If I understood this correctly, this k-means procedure is carried out only once. After that, all genres have 100 patterns each (except for Electronic, which has 16, and possibly Bossa Nova, which has 97 in the ZIP file). I understand that the clustering procedure followed by selection is meant to diversify the quality of the selected patterns, but why was a pattern selected at random from each cluster? […]
    - **[r] We agree about this comment, in future iterations of the work, we will certainly consider this comment.**

4. Moreover, it would be interesting to add a note on how patterns are selected for the user after he selects the genre..

- **[r] We have this specified in Figure 3 but the text is made more clear now in the figure.**

5. Section 4 and Section 5 (All discrepancies)
    - **[r] Computations, analyses and figures for sections 4 and 5 were rechecked. We noticed that there was an error in our calculation script which resulted in showing only a subset of Jaccard calculations. We have fixed this issue, and the figures are now fully updated. This implied updating figures 7, 8 and 9 and removing line 441.**
    - **Detailed explanation of the error:**
        - **Fix 1: The list used for compiling the pairwise comparisons within each experience group/pattern was incorrectly placed inside the second loop, essentially implying that not all the pairwise selections were used in the calculation.**
        - **Fix 2: As shown in the following code snippets, when conducting pairwise calculations, we missed that the diagonal of the matrix should be excluded. In cases where p1 and p2 are the same pattern, the Jaccard would be 1 and hence boost up the averages shown in the submitted draft.**

Original Incorrect Version

```
for j, p_list in enumerate(temp_e.keys()):
  for p1 in temp_e[p_list]:
    temp_jaccards = [] # Incorrect, move before prev line
    for p2 in temp_e[p_list]:
      jaccard_p1p2 = jaccard(p1,p2)
      temp_jaccards.append(jaccard_p1p2)
```

Correct Version

```
for j, p_list in enumerate(temp_e.keys()):
  temp_jaccards = []            # - - - - - - - -  Fix 1
  for p1 in temp_e[p_list]:
    for p2 in temp_e[p_list]:
      if not np.all(p1 == p2):  # - - - - - - - -  Fix 2
        jaccard_p1p2 = jaccard(p1,p2)
        temp_jaccards.append(jaccard_p1p2)
```

6. Section 3, footnote 7: "... confirm this speculation." (remove "s", add ".")

    - **[r] Updated in manuscript**

7. Section 4, lines 311f: There is no need for this mention of section 5 here ("and in Section 5... the Electronic genre")...

    - **[r] All the suggested changes related with sections 4 and 5 were corrected accordingly. As mentioned, the graphs (7, 9 and 9) were updated and some lines were removed as suggested (lines 312 and 395).**

8. Figures 4 and 5: Consider presenting the genres in the same order, to facilitate comparing the two figures.

    - **[r] The graphs in correct order are updated in manuscript**

9. Lines 443-447: Maybe include here an interpretation of the results…

    - **[r] This paper is intended as an introduction to the dataset, so we believe this and other additional analyses that can be derived from the El Bongosero dataset should be reported in further manuscripts. Therefore we decided not to include this suggested interpretation.**

Metareviewer

Major Remarks

1. Some numbers are different in the text and in the figures, it is mandatory to thoroughly check the correspondance between both and to correct the paper for the final version of the paper.

    - **[r] As mentioned above, the relevant information has been updated.**

2. Figure 5 also does not correspond to the equivalent figure given in the dataset, probably due to a version problem. This has to be fixed.

    - **[r] This change has been made to the website.**

Other remarks:

3. A focus is done on electronic genre, limiting it to 16 original patterns. More details should be given in the paper on the genres represented in the dataset and their distribution in number of original patterns…

    - **[r] In line 301, we mention that for each genre (with the exception of Electronic) we have selected 100 original drum patterns. However, for better clarity, in the camera-ready version, this information has been restated in the opening paragraph of the conclusion.**

4. On the Electronic genre focus, more exploration of the similarity with the original pattern through time could certainly give interesting results…

    - **[r] We certainly agree about this comment. At this point, we have not conducted this analysis and given the space limitations, we will defer this analysis to future in depth explorations of the dataset.**

5. At this step, the dataset contains only 4/4 metric, but that makes sense for electronic music. It would be interesting to extend it to other metrics.

    - **[r] If a similar opportunity arises in the future, we will certainly improve the diversity of the tested drum patterns.**

Typos:

6. 155. select -> selected

    - **[r] Updated in manuscript**

7. 276 and l 280. the “the”

    - **Updated in manuscript**