MIREX 2013: DISCOVERING MUSICAL PATTERNS USING AUDIO STRUCTURAL SEGMENTATION TECHNIQUES

Oriol Nieto, Morwaread Farbood
Music and Audio Research Lab
New York University
{oriol, mfarbood}@nyu.edu

ABSTRACT

This extended abstract discusses our pattern discovery algorithm submitted to the MIREX 2013 Discovery of Repeated Themes & Sections task. This algorithm estimates the musical patterns by finding specific repetitions within a piece and applying certain perceptually inspired rules. Four different versions of the algorithm were submitted: two that take an audio track as an input (monophonic and polyphonic) and two more that take a symbolic representation (monophonic and polyphonic). Each version follows a similar implementation, which is common to the task of audio structural segmentation: convert the music representation into a *chromagram* (or pitch class profiles), compute the key-invariant self-similarity matrix, and then extract the most prominent repeated segments by analyzing the matrix diagonally. Once the segments have been extracted, they are split into smaller segments if repetition is found within the segments, following perceptual rules regarding pattern length and number of rests. Once these segments meet these requirements, they are considered patterns, and their occurrences are matched using the self-similarity matrix.

1. DESCRIPTION OF THE ALGORITHM

For this year's MIREX submission, we decided to implement an algorithm generic enough that it would be able to run with audio or symbolic music, either monophonic or polyphonic. For evaluation purposes, we submitted the following versions:

- NF1: Symbolic, monophonic music.
- NF2: Symbolic, polyphonic music.
- NF3: Audio, monophonic music.
- NF4: Audio, polyphonic music.

In the four versions we transform the input into a Chromagram (or Pitch Class Profile vector), a common harmonic representation often used in the task of audio structure segmentation [2, 5] (see Figure 1 for an example of a Chromagram extracted from an audio-polyphonic input). While in the audio versions we use a fixed frame size for the chroma vectors, in the symbolic versions we use the shortest note or rest duration (i.e. the tatum) to determine the frame size.

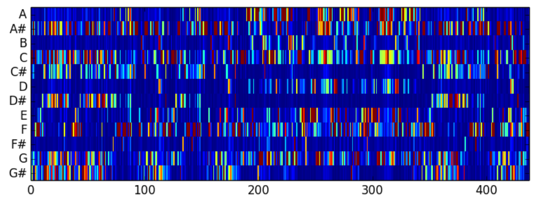

Figure 1. Chromagram of the audio-polyphonic version of Beethoven Op. 2, No. 1, Movement 3.

In order to find repetition across a piece, we take the key-invariant Self-Similarity Matrix (SSM) of the Chromagram using the Euclidean distance [3] and find the most similar segments using a modified version of the techniques described in [2] such that we detect the smallest repetition or musical pattern instead of longer musical segments. To do so, we initially detect longer segments and keep splitting them based on the repetition of other similar segments. We set a threshold τ perceptually inspired, as in [4], that is automatically adapted in case no music patterns are found. See Figure 2 for an example of the potential occurrences found in a key-invariance SSM.

Once we have found a set of potential patterns and its occurrences, we apply a set of perceptual rules [4] to filter out the patterns that are unlikely to constitute a musical motive (e.g. the amount of silence or rests dominate the pattern). Given this final set of patterns, we use the self-similarity matrix in order to find the occurrences which will also be part of the output of the algorithm.

2. RESULTS

We use the JKU Patterns Development Dataset [1] to evaluate the four different versions of our algorithm. This dataset contains annotations of five pieces:

- Bach BWV 889
- Beethoven Op. 2, No. 1, Movement 3

[1] https://dl.dropbox.com/u/11997856/JKU/JKUPDD-Aug2013.zip

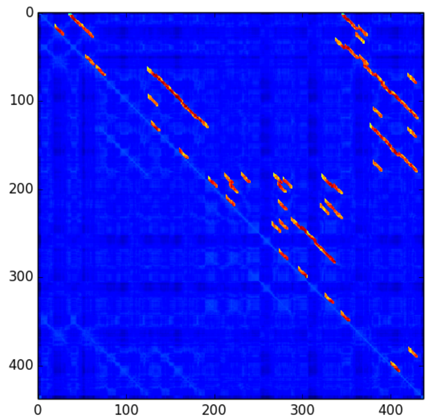

Figure 2. Key-invariance Self-Similarity Matrix of Beethoven Op. 2, No. 1, Movement 3, using polyphonic audio as input, with the possible occurrences found marked on red.

- Chopin Op. 24, No. 4
- Gibbons Silver Swan
- Mozart k.282, Movement 2

The evaluation metrics are the ones defined by Collins and Meredith, explained in [1]. In Table 1 the results for each version submitted to MIREX are shown, averaged for all the pieces of the dataset.

NF2 obtains the best scores overall compared with the other three versions. NF4 yields the second best results (except in $F_o(c = .5)$, where it is slightly less than in NF1). These results suggest that our method performs best using polyphonic inputs.

We obtain better results for the occurrence scores than for the pattern retrieval scores. Therefore, our algorithm better retrieves the occurrences of a given pattern once it has discovered this pattern, than finding all the patterns of a given piece.

The standard F-measure (not shown in the Table) is always 0 for all the four versions. This means that our method is not precise at finding the exact start and end points of the patterns and its occurrences.

Except in NF1, where R_{est} is 10 points greater than P_{est}, the rest of the algorithms have a similar P_{est} and R_{est}. Therefore, our method does not find too many non-existent patterns (tendency of higher R_{est}), nor too few (tendency of higher P_{est}), which is usually desired.

Finally, the fastest version of our method is NF4. Audio-based implementations are faster due to longer frame lengths. Algorithms using polyphonic inputs are faster because it is more likely to find possible occurrences in the SSM when having more information on it. Note that the SSM in the monophonic cases tends to be more sparse.

3. CONCLUSIONS

We have submitted four versions of a novel pattern discovery algorithm to the first edition of the MIREX task of Discovery of Repeated Themes & Sections. This algorithm is generic enough such that it can run with different input data (symbolic/audio and monophonic/symbolic). This generalization speeds up the process of finding motives but it can decrease its performance, especially when using monophonic data. This new MIREX task presents an exceptional opportunity for the research community to compare their pattern discovery algorithms, and we hope to see more submissions in the upcoming years for all the four different versions of the task.

4. ACKNOWLEDGMENTS

Thanks to Fundación Caja Madrid for the funding.

5. REFERENCES

[1] Tom Collins. MIREX Task: Discovery of Repeated Themes & Sections, 2013.

[2] Meinard Müller. *Information Retrieval for Music and Motion*. Springer, 2007.

[3] Meinard Müller and Michael Clausen. Transposition-Invariant Self-Similarity Matrices. In *Proc. of the 8th International Conference on Music Information Retrieval*, pages 47–50, Vienna, Austria, 2007.

[4] Oriol Nieto and Morwaread Mary Farbood. Perceptual Evaluation of Automatically Extracted Musical Motives. In *Proc. of the 12th International Conference on Music Perception and Cognition*, pages 723–727, Thessaloniki, Greece, 2012.

[5] Jouni Paulus, Meinard Müller, and Anssi Klapuri. Audio-Based Music Structure Analysis. In *Proc of the 11th International Society of Music Information Retrieval*, pages 625–636, Utrecht, Netherlands, 2010.

Algo	P_{est}	R_{est}	F_{est}	$P_{\text{o}(c=.75)}$	$R_{\text{o}(c=.75)}$	$F_{\text{o}(c=.75)}$	P_3	R_3	F_3	$P_{\text{o}(c=.5)}$	$R_{\text{o}(c=.5)}$	$F_{\text{o}(c=.5)}$	Time (s)
NF1	22.78	32.25	26.57	0.00	0.00	0.00	18.62	32.84	23.48	28.12	22.22	24.81	590.12
NF2	54.77	53.41	**48.06**	65.50	51.23	**57.09**	43.51	47.51	**38.78**	60.30	50.45	**53.38**	475.79
NF3	21.01	25.94	22.55	13.33	3.33	5.33	11.02	16.92	13.05	33.30	12.66	18.20	236.36
NF4	40.83	46.43	41.43	32.08	21.24	24.87	30.43	31.92	28.23	26.60	20.94	23.18	**196.29**

Table 1. Results of the four different versions of our method submitted to MIREX on the JKU Patterns Development Dataset, averaged across pieces.