IMPROVING BINARY SIMILARITY AND LOCAL ALIGNMENT FOR COVER SONG DETECTION

Joan Serrà, Emilia Gómez & Perfecto Herrera
Music Technology Group
Universitat Pompeu Fabra, Barcelona, Spain
{joan.serraj,emilia.gomez,perfecto.herrera}@upf.edu

ABSTRACT

This is an extended abstract that overviews our cover song detection system as submitted to the MIREX 2008 audio cover song identification task. The system is developed starting from our 2007 submission but includes some important modifications and parameter tuning. Our system obtains the best results in all evaluation measures considered and its accuracy is proved to be statistically significant in comparison to the other systems in this task.

1 INTRODUCTION

Cover song identification has been a very active topic within the last few years in the music information retrieval (MIR) community. From a research perspective, it is a task where the relationship between songs is context-independent and can be qualitatively defined and objectively measured. In addition, it expands the notions of music similarity beyond acoustic resemblance to include the important idea that musical works retain their identity notwithstanding variations in many musical dimensions [1]. From an industrial perspective, music similarity plays a central role in searching and organizing music collections. Besides, finding cover songs in a database has a direct implication to musical rights management and licenses. Finally, from a user's perspective, finding all versions of a particular song can be valuable and fun.

In popular music, the main purpose of recording a cover (or version) might be to perform a radically different interpretation of the original song. Then, important changes at different musical facets (timbre, tempo, rhythm, song structure, main key, harmonization, lyrics, language and so on) are involved. A robust mid-level characteristic that is largely preserved under the mentioned musical variations is the tonal sequence, which can be understood as series of different note combinations played sequentially. These notes can be unique for each time slot (a melody) or can be played jointly with others (chord or harmonic progressions). Systems for cover song identification usually exploit these aspects and attempt to be robust against changes in other musical facets. In general, they either try to extract the predominant melody, a chord progression, or a chroma sequence from the raw audio signal and normalize it in respect to the key. Then, for obtaining a similarity measure between songs, these descriptor sequences are usually compared by means of alignment techniques like dynamic time warping (DTW), edit-distance variants, string matching algorithms, or by a simple correlation function. A more extensive introduction to audio cover song identification and an overview of its state-of-the-art can be found in [5].

2 ALGORITHM OVERVIEW

The general schema for the submitted system is very similar to the one presented in [3, 5]. The basic system comprises the same modules as [5] (figure 1) but with substantial differences in *song transposition* and *binary similarity* modules. We have also exhaustively tuned the parameters of the *dynamic programming local alignment* module.

We now enumerate the basic modules of the system and provide a short description of them while highlighting important modifications. Further references to previous work are given through the text.

2.1 Chroma feature extraction

We use *harmonic pitch class profile* (HPCP) features [2]. The extraction procedure and parameters are the same as specified in [5]: HPCPs are extracted in a frame basis (62 dB Blackman-Harris, 93 ms, 50% overlap) and consecutive frames are averaged each 250 ms (we do not use any kind of beat information). Finally, due to modifications in the binary similarity matrix computation, instead of 36-bin HPCP features, we use 12-bin feature vectors. Thus, we also reduce computational costs.

2.2 Chroma transposition

The objective of chroma transposition (*global HPCP*, *optimal transposition index* and *song transposition* modules) is to normalize chroma representations with respect to the employed musical key. These modules have been modified

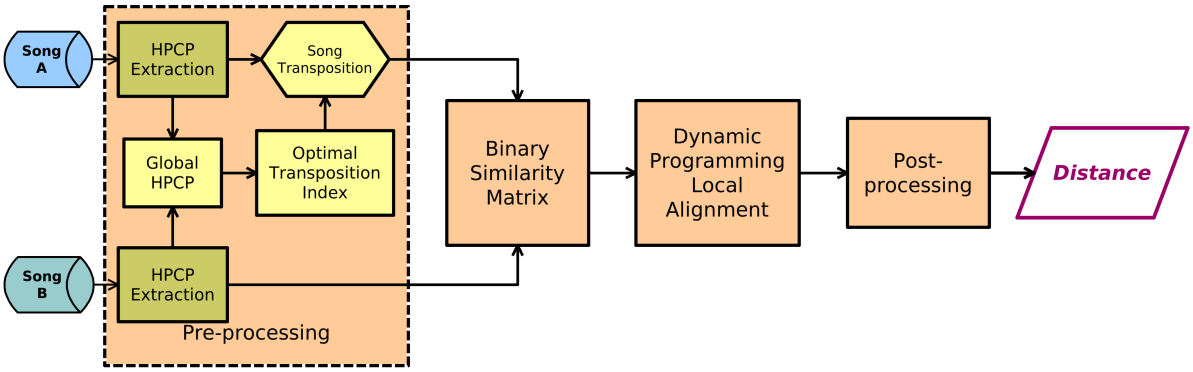

Figure 1. General block diagram of the cover song identification system as depicted in [5].

as described in [4] in order to account for multiple transposition options while avoiding to compute all possible ones. With a global average chroma feature vector (*global HPCP*) we obtain the two most optimal transpositions (*optimal transposition index*, OTI). We compute the cover song similarity measure for both and keep the highest one as the final decision.

2.3 Binary similarity matrix

This module has been completely modified from the original implementation in [3, 5]. The objective was to reduce computational costs and, at the same time, find a more general binary similarity matrix construction method that was not restricted to chroma features and that could be applied to any kind of descriptor. In addition, our tests confirm some increase in performance due to this modification. However, further experiments need to be done. Method and results will be extensively explained in a forthcoming document.

2.4 Dynamic programming local alignment

The binary similarity measure computed in previous module is used as a local cost function for a dynamic programming local alignment (DPLA) algorithm, which finds the best subsequence match between all possible ones while considering tempo deviations and sequence gaps. This makes our method independent of song structure, tempo changes and tonal sequence insertions and/or deletions. The DPLA algorithm was extensively described in [5]. Fine tuning of this algorithm's parameters for the new binary similarity matrix module has been done, obtaining an extra performance improvement with respect to the old ones.

2.5 Post-processing

2.5.1 Distance normalization

Only the best local alignment for the two chroma representations is finally used to obtain a dissimilarity measure between them. We use the same formulation as in [5].

2.5.2 Distance matrix refinement

All the previous modules are shared by the two submitted systems (1 and 2), but the second one (2) incorporates an extra distance refinement module at the end of the block chain. This is a very preliminary algorithm that is still currently being developed.

3 EVALUATION

3.1 Test material and methodology

The MIREX 2008 test data is composed of 30 cover groups, each one having 11 different versions. Therefore, the total cover song collection contains 30x11 = 330 songs. These are embedded in a database summing up a total of 1000 tracks, which includes a wide diversity of genres (e.g., classical, jazz, gospel, rock, folk-rock, etc.), and the variations span a variety of styles and orchestrations. This music collection is the same as used in previous MIREX editions (2006 and 2007).

Each of the 330 cover songs were used as queries and the systems were required to return a 330x1000 distance matrix (one row for each query). From this distance matrix, several evaluation measures were computed. These were the same as the ones employed in 2007, including the mean average precision (MAP) measure. More information on the audio cover song identification task can be found in the MIREX wiki [1] or in [1].

[1] `http://www.music-ir.org/mirex/2008/index.php/Audio_Cover_Song_Identification`

Participant	*Evaluation measures*			
	$TNCI_{10}$ [3300..0]	$MNCI_{10}$ [10..0]	MAP [1..0]	$Rank_1$ [1..989]
Jensen, Christensen & Jensen	763	2.31	0.23	23.98
Cao & Li (1)	1056	3.20	0.34	18.83
Cao & Li (2)	1073	3.25	0.34	18.78
Egorov & Linetsky (1)	1762	5.34	0.55	11.19
Egorov & Linetsky (3)	1778	5.39	0.56	11.55
Egorov & Linetsky (2)	1781	5.40	0.56	11.17
Serrà, Gómez & Herrera (1)	**2116**	**6.41**	**0.66**	**4.55**
Serrà, Gómez & Herrera (2)	**2422**	**7.34**	**0.75**	**4.80**

Table 1. Results for MIREX08 Audio Cover Song task. $TNCI_{10}$ corresponds to the total number of covers identified in top 10, $MNCI_{10}$ is the mean number of covers identified in top 10 (average performance), *MAP* is the arithmetic mean of Average Precision, and $Rank_1$ is the rank of the first correctly identified cover.

3.2 Results

Table 1 shows the overall summary results obtained. Our algorithms (bottom rows) performed the best in all evaluation measures considered, reaching a MAP of 0.66 and 0.75 respectively. This is the best score reached in the MIREX audio cover song identification task since it was first run in 2006. Furthermore, our second submission (2) seems to be significantly better than the other systems presented (Friedman test, figure 2), and therefore, considering previous MIREX evaluations, the difference is statistically significant for all the methods evaluated for this task until now.

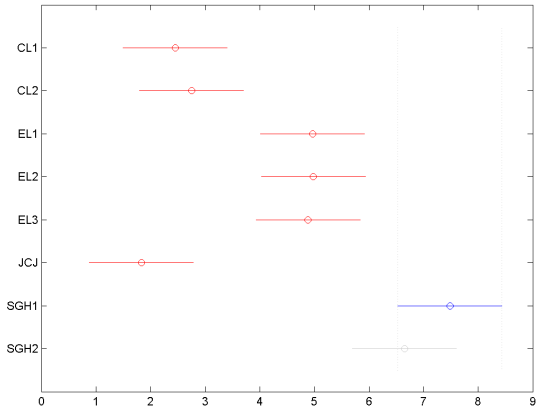

Figure 2. MIREX08 Friedman's test for significant differences.

The submitted systems have 71 (1) and 116 (2) queries with an average precision higher than 0.95. These results are quite remarkable because [1], given the dataset size (1000) and the number of relevant items per query (10), the probability of randomly returning the entire relevant set within the top 10 list once in 330 queries is only 1.26×10^{-21}!

The complete 2008 evaluation results can be seen at the MIREX wiki [2] .

3.3 Comparison with MIREX 2007 algorithm

In table 2 we show the results for our submissions in 2007 and 2008. We can see that our 2008 submissions improve substantially (28 and 44 % relative on MAP) the previous version of the algorithm, which, in its turn, was the best performing one among 2007 submissions. This confirms that the highlighted upgrades have yielded much higher accuracies.

Algorithm	*Evaluation measures*			
	$TNCI_{10}$ [3300..0]	$MNCI_{10}$ [10..0]	MAP [1..0]	$Rank_1$ [1..989]
2007	1653	5.00	0.52	9.37
2008 (1)	2116	6.41	0.66	4.55
2008 (2)	2422	7.34	0.75	4.80

Table 2. Comparison for MIREX07 and MIREX08 submitted algorithms.

If we compare errors between 2007 and 2008 (table 3), we can see that there are some query groups which had already (near) perfect retrieval that our improved version still manages to maintain. Examples of this would be query groups 4, 7, 14 and 17. Greatest improvement has been achieved in poor scoring query groups like 3, 9, 12, 20, 21, 24, 25, 27 and 29. Specifically, we obtain a huge improvement with queries 3, 20, 24 and 27. Furthermore, we cannot find any query group where precision decreases in a significant manner. Again, this denotes the goodness of the

[2] `http://www.music-ir.org/mirex/2008/index.php/Audio_Cover_Song_Identification_Results`

introduced upgrades. However, some few query groups still resist to be retrieved with high precision: 6, 19 and 22.

Query group	*Algorithms*		
	2007	2008 (1)	2008 (2)
1	0.86	0.77	1.00
2	0.40	0.56	0.69
3	0.20	0.53	0.70
4	1.00	1.00	1.00
5	0.83	0.85	0.88
6	0.12	0.21	0.21
7	0.96	0.98	1.00
8	0.44	0.54	0.67
9	0.34	0.73	0.89
10	0.70	0.92	1.00
11	0.86	0.85	0.85
12	0.19	0.53	0.99
13	0.41	0.48	0.47
14	1.00	1.00	1.00
15	0.77	0.83	0.83
16	0.53	0.77	0.90
17	0.91	0.98	0.98
18	0.70	0.93	0.95
19	0.05	0.10	0.12
20	0.17	0.49	0.61
21	0.39	0.52	0.72
22	0.05	0.04	0.03
23	0.76	0.80	0.82
24	0.09	0.44	0.57
25	0.47	0.68	0.84
26	0.84	1.00	1.00
27	0.23	0.47	0.72
28	0.45	0.46	0.47
29	0.24	0.52	0.65
30	0.70	0.95	0.97
MAP	0.52	0.66	0.75

Table 3. Comparison for MIREX07 and MIREX08 submitted algorithms. Arithmetic means of the average precisions within each of the 30 query groups.

4 CONCLUSION

We have submitted a cover song detection system that has much better accuracy than last year's submission. The improvement can be due to the extended chroma transposition module, the new binary similarity matrix formulation and the tuning of DPLA algorithm's parameters. Furthermore, our method obtains the highest values for all the evaluation measures considered, being substantially superior to all the other algorithms presented in this and previous years.

5 ACKNOWLEDGEMENTS

The authors wish to thank their colleagues at the MTG (UPF). They also want to acknowledge all the IMIRSEL team for the organization and running of this evaluation, specially Mert Bay.

This research has been partially funded by the EU-IP project PHAROS [3] IST-2006-045035.

6 REFERENCES

[1] J. S. Downie, M. Bay, A. F. Ehmann and M. C. Jones. Audio cover song identification: Mirex 2006-2007 results and analyses. *Int. Symp. on Music Information Retrieval (ISMIR)*, pp. 468–473, September 2008.

[2] E. Gómez. *Tonal description of music audio signals*. Ph.D. thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2006.

[3] J. Serrà and E. Gómez. Audio cover song identification based on sequences of tonal descriptors. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 61–64, April 2008.

[4] J. Serrà, E. Gómez and P. Herrera. Transposing chroma representations to a common key. *IEEE CS Conference on The Use of Symbols to Represent Music and Multimedia Objects*, October 2008. In Press.

[5] J. Serrà, E. Gómez, P. Herrera and X. Serra. Chroma binary similarity and local alignment applied to cover song identification. *IEEE Trans. on Audio, Speech and Language Processing*, vol. 16 (6), pp. 1138–1152, August 2008.

[3] `http://www.pharos-audiovisual-search.eu`