A COVER SONG IDENTIFICATION SYSTEM BASED ON SEQUENCES OF TONAL DESCRIPTORS

Joan Serrà and Emilia Gómez
Music Technology Group
Universitat Pompeu Fabra
{jserra,egomez}@iua.upf.edu

ABSTRACT

The present paper corresponds to the extended abstract of a system for cover version identification submitted to the Audio Cover Song task in the context of the Music Information Retrieval Evaluation eXchange (MIREX) 2007. The proposed algorithm extracts sequences of tonal descriptors from audio recordings and uses them to compute a similarity measure between two musical pieces.

1 INTRODUCTION

Nowadays, in popular music, the term *cover song* (also named version, or simply cover) has come to mean any new rendition, performance or recording, of a previously recorded song [1] .

Cover song identification has become a very active topic of research within the last few years in the Music Information Retrieval (MIR) community [3, 4, 1] (for citing only some), as it provides a direct and objective way of evaluating music similarity algorithms. Some efforts are then being devoted to compare and evaluate different alternatives for this purpose. This is the case of MIREX, an international effort to develop formal, common evaluation standads for MIR, where, for the first time in 2006, there was an evaluation for cover song identification [2] .

Tonal sequences are useful descriptors for cover song identification. In popular music, the main purpose of recording a version can be to investigate a radically different performance of the original song. Then, important changes at different musical facets are involved. These might be: timbre, tempo, rhythm, structure, key (or main tonality), harmonization (e.g., adding or deleting chords, substituting them by relatives, adding tensions), lyrics and language. Thus, it seems that one of the features that is mostly preserved (apart from the main melody) is the overall tonal sequence.

[1] http://en.wikipedia.org/wiki/Cover_version
[2] http://www.music-ir.org/mirex2006/index.php/Audio_Cover_Song

2 OVERVIEW

In figure 1 we show a general block diagram of the system. Each pair of songs is pre-processed independently, and a similarity matrix between them is computed. Then, this is used by a Dynamic programming method [5], which aligns the two musical pieces. At the end, some post-processing steps for obtaining the final distance measure are followed.

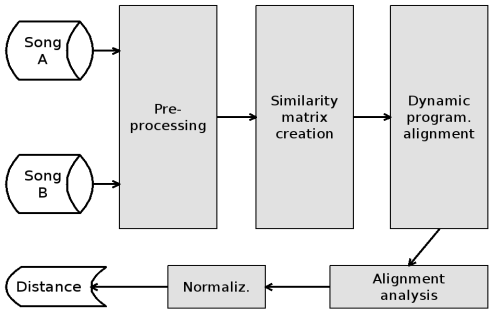

Figure 1. General block diagram of the cover song identification system.

Pre-processing basically consists in obtaining an enhanced version of chroma features: the Harmonic Pitch Class Profile (HPCP). The extraction is done as explained in [2]. We first start by cutting the song into short overlapping and windowed frames. For that, we use a Blackman-Harris window of 93 ms length with 50% overlap. We consider a whitened frequency spectrum that ranges from 40 Hz up to 5 KHz, which is processed in order to extract the main spectral peaks. Peak frequencies are mapped into 36 pitch-class values considering the presence of harmonic frequencies. In addition, a global average HPCP vector is calculated. This global HPCP is used to transpose each pair of songs to a common key or tonality.

The next step is to compute a similarity matrix between two transposed HPCP sequences. Each element of this similarity matrix represents the resemblance between two elements of the respective HPCP sequences. Resemblance between HPCPs is assessed with a new binary similarity measure to get a highly contrasted matrix. This matrix is the only input of a Dynamic programming local alignment algorithm based on a constrained recurrent relation,

Measure	Range	EC	JB	JEC	KL1	KL2	KP	IM	**SG**
$TNCI_{10}$	[0-3300]	1207	869	762	425	291	190	34	**1653**
$MNCI_{10}$	[0-10]	3.658	2.633	2.309	1.288	0.882	0.576	0.103	**5.009**
MAP	[0-1]	0.330	0.267	0.238	0.13	0.086	0.061	0.017	**0.521**
$Rank_1$	[0-1000]	13.994	29.527	22.209	57.542	51.094	46.539	97.470	**9.367**

Table 1. Results for MIREX07 Audio Cover Song task. $TNCI_{10}$ corresponds to the total number of covers identified in top 10, $MNCI_{10}$ to 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. The results for the algorithm presented are shown in the last column (SG).

which computes all possible subsequence alignments between the given tracks. At the end, only the 'best' path is considered to output a final alignment score. Finally, to yield a distance measure, we compute the ratio between a normalizing factor (which depends on the song lengths), and the obtained similarity score.

A detailed explanation on this and other cover song identification approaches can be found in [5], as well as some tests of the affection to the final performance results of some specific parts of these.

3 EVALUATION

3.1 Test material and methodology

The MIREX 2007 test data was composed of 30 *cover sets*, each one having 11 different versions. Therefore, the total cover song collection contained 30x11 = 330 songs. These were embedded in a database summing up a total of 1000 tracks. The test collection included a wide diversity of genres (e.g., classical, jazz, gospel, rock, folk-rock, etc.), and the variations spanned a variety of styles and orchestrations.

Each of the 330 cover songs were used as queries and the systems were required to return 10 results for each query. Systems were evaluated on the number of the songs from the same class/set as the query that were retrieved.

3.2 Results and discussion

A total of 8 different methods were presented to the Audio Cover Song task. Table 1 shows the overall summary results obtained [3] . Our algorithm (SG, last column) performed the best in all evaluation measures considered, reaching an average accuracy of 5.009 of correctly identified covers within the 10 first retrieved elements ($MNCI_{10}$) and a Mean Average Precision (*MAP*) of 0.521. Furthermore, the next best performing system reached and $MNCI_{10}$ of 3.658 and a *MAP* of 0.330, which represents a substantial difference to ours. In addition, significance tests showed that the results for our system were significantly better than the 6 other systems. For that, the Friedman test was run against the Average Precision summary data over the 30 song groups.

[3] http://www.music-ir.org/mirex2007/index.php/Audio_Cover_Song_Identification_Results

4 CONCLUSIONS

We submitted a system that identifies cover songs using Dynamic programming and an harmonic representation from the raw audio as a feature set. Our method obtained the highest values for all the evaluation measures considered, being substantially superior to all the other algorithms presented.

5 ACKNOWLEDGEMENTS

The authors wish to thank their colleagues at the MTG (UPF), specially Perfecto Herrera for his constant support and helpful ideas. They also want to mention all the IS-MIRSEL team for the organization and running of this evaluation, specially Stephen Downie and Cameron Jones.

This research has been partially funded by the EU-IP project PHAROS [4] .

6 REFERENCES

[1] D. P. W. Ellis and G. E. Polliner. Identifying cover songs with chroma features and dynamic programming beat tracking. *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, April 2007.

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

[3] E. Gómez, B. S. Ong and P. Herrera. Automatic tonal analysis from music summaries for version identification. *Conv. of the Audio Engineering Society (AES)*, October 2006.

[4] M. Marolt. A mid-level melody-based representation for calculating audio similarity. *Proc. Int. Symposium on Music Information Retrieval (ISMIR)*, 2006.

[5] J. Serrà. *Music similarity based on sequences of descriptors: tonal features applied to audio cover song identification*. Master's thesis, MTG, Universitat Pompeu Fabra, Barcelona, Spain, 2007.

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