

# HYBRID MUSIC SIMILARITY MEASURE

Dmitry Bogdanov, Joan Serrà, Nicolas Wack, and Perfecto Herrera  
Music Technology Group  
Universitat Pompeu Fabra  
Roc Boronat, 138, 08018 Barcelona, Spain  
{dmitry.bogdanov,joan.serraj,nicolas.wack,perfecto.herrera}@upf.edu

## ABSTRACT

This paper describes our submission for the MIREX 2010 audio music similarity and retrieval task. The submission is similar to the one of our previous systems, sent to MIREX 2009 [1]. This task of audio music similarity can be regarded as obtaining a suitable distance measurement between songs defined on a certain feature space. We propose a hybrid measure defined as a combination of a number of existing state-of-the-art measures, which include three low-level measures, together with one high-level semantic measure, based on regression by support vector machines of different groups of musical dimensions such as genre and culture, moods and instruments, or rhythm and tempo.

## 1. ALGORITHM OVERVIEW

### 1.1 Descriptors

We characterize each song using an in-house audio analysis tool<sup>1</sup>. We used over 60 descriptor classes in total, characterizing global properties of songs. Extracted descriptor classes include inharmonicity, odd to even harmonic energy ratio, tristimulus, spectral centroid, spread, skewness, kurtosis, decrease, flatness, crest, and roll-off factors, MFCCs, spectral energy bands, zero-crossing rate, spectral and tonal complexities, transposed and untransposed harmonic pitch class profiles, key strength, tuning, chords, BPM, and onsets. Detailed references according used descriptors can be found in [3].

### 1.2 Hybrid distance

We use the following distance measures to form our hybrid distance: (i) an Euclidean distance based on principal component analysis [4], (ii) a Kullback-Leibler divergence based on single Gaussian MFCC modeling [5], (iii) a tempo-based distance [3], and (iv) a semantic classifier-based distance [3]. Comparing to our MIREX 2009 sub-

mission [1], we updated the classifier-based distance. To this extent, we employed 4 new collections, covering genre, moods, timbre color, and voice gender. We define our hybrid distance as a linear combination of these measures, preliminary normalized. We use the results, obtained from our previous subjective evaluation, and select human graders' average ratings for each measure as the weights. We refer you to [2] for the detailed information about the components of the proposed hybrid distance, the normalization approach, and the weights used.

## 2. EVALUATION

The evaluation of the systems submitted to the MIREX 2010 audio music similarity and retrieval task was conducted as follows. Each system was given 7000 songs (30 sec. excerpts) chosen from IMIRSEL's<sup>2</sup> collections. Each system then calculated a  $7000 \times 7000$  distance matrix. A query set of 100 songs was randomly selected from the 10 genre groups (10 per genre). For each query in the set, the closest 5 songs out of the 7000 (after filtering out the query itself and all songs of the same artist) were chosen as candidates. Then, for each query, the candidates from all systems were grouped and subsequently evaluated by human graders using the Evalutron 6000 grading system. Each individual query/candidate set was evaluated by a single grader. For each query/candidate pair, graders provided (i) a categorical broad score in the set  $\{0, 1, 2\}$  (corresponding to "not similar", "somewhat similar", and "very similar" categories) and (ii) a fine score in the range from 0 (failure) to 100 (perfection).

## 3. RESULTS

A listening experiment with 50 human graders was conducted. According to the MIREX 2010 wiki<sup>3</sup> our measure is noted as BWL1. The overall evaluation results are presented in Table 1. The results of the Friedman test against the summary data of fine scores are presented in Fig. 1. As expected, all submitted systems outperformed

<sup>1</sup> <http://mtg.upf.edu/technologies/essentia>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2010 International Society for Music Information Retrieval.

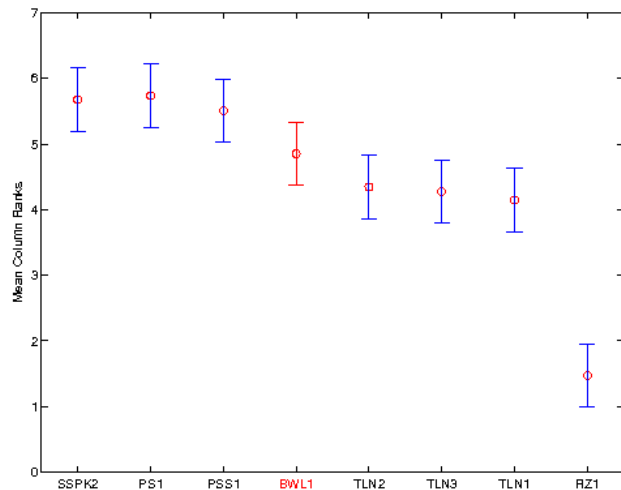
<sup>2</sup> <http://www.music-ir.org/evaluation/>

<sup>3</sup> [http://www.music-ir.org/mirex/wiki/2010:Audio\\_Music\\_Similarity\\_and\\_Retrieval\\_Results](http://www.music-ir.org/mirex/wiki/2010:Audio_Music_Similarity_and_Retrieval_Results)

Acronym	Authors (measure)	Average fine score	Average broad score
SSPK2	Klaus Seyerlehner, Markus Schedl, Tim Pohle, and Peter Knees	56.642	1.248
PS1	Tim Pohle, Dominik Schnitzer	55.080	1.228
PSS1	Tim Pohle, Klaus Seyerlehner, and Dominik Schnitzer	54.984	1.212
BWL1	Dmitry Bogdanov, Joan Serrà, Nicolas Wack, and Perfecto Herrera	49.704	1.078
TLN3	George Tzanetakis, Mathieu Lagrange, and Steven Ness	46.604	0.968
TLN2	George Tzanetakis, Mathieu Lagrange, and Steven Ness	46.544	0.970
TLN1	George Tzanetakis, Steven Ness, and Mathieu Lagrange	45.842	0.940
RZ1	Rainer Zufall (random baseline)	16.668	0.240

**Table 1.** MIREX 2010 overall summary results sorted by average fine score.

a random baseline (RZ1). We found our distance to perform adequately well comparing to other submitted systems. Though the SSPK2, PS1, PSS1 systems outperform our approach according to the average fine and broad scores, no statistically significant difference between these approaches and our approach was found in the Friedman test. In general, it is interesting to note that none of the submitted systems performed with the user satisfaction greater than average.



**Figure 1.** MIREX 2010 Friedman's test (fine scores).

#### 4. CONCLUSION

We review our submissions to the MIREX 2010 audio music similarity and retrieval task. The hybrid measure takes advantage of different low-level approaches to music similarity together with the semantic classifier-based distance measure. The human evaluations conducted show that our measures have much potential.

#### 5. ACKNOWLEDGMENTS

The authors would like to thank the colleagues from Music Technology Group and all participants of the evaluation. This work was partially funded by the FI Grant of Generalitat de Catalunya (AGAUR).

#### 6. REFERENCES

[1] D. Bogdanov, J. Serrà, N. Wack, and P. Herrera. Hybrid similarity measures for music recommendation.

In *Music Information Retrieval Evaluation Exchange (MIREX'09)*, 2009.

[2] D. Bogdanov, J. Serrà, N. Wack, P. Herrera, and X. Serra. Unifying low-level and high-level music similarity measures. Submitted.

[3] Dmitry Bogdanov, Joan Serrà, Nicolas Wack, and Perfecto Herrera. From low-level to high-level: Comparative study of music similarity measures. In *International Workshop on Advances in Music Information Research (AdMIRE'09)*, 2009.

[4] P. Cano, M. Koppenberger, and N. Wack. Content-based music audio recommendation. In *ACM International Conference on Multimedia (ACMMM'05)*, pages 211–212, 2005.

[5] Michael I. Mandel and Daniel P.W. Ellis. Song-level features and support vector machines for music classification. In *International Conference on Music Information Retrieval (ISMIR'05)*, pages 594–599, 2005.