# MIREX 2014: OPTIMIZING THE FLUCTUATION PATTERN EXTRACTION PROCESS

**Klaus Seyerlehner, Markus Schedl**
Dept. of Computational Perception, Johannes Kepler University, Linz, Austria

## ABSTRACT

Similar to the submissions of the previous years our submissions are based on the so-called block-level features (BLF). The main change in our 2014 submission was the modification of the feature extraction steps for the *Logarithmic Fluctuation Pattern*. Beside this modification the structure of all three algorithms did not changed. This abstract gives an overview on the feature set and presents some specific details of the submitted algorithms.

## 1. INTRODUCTION

In our system the same set of features is extracted for all three tasks. The feature extraction is implemented in MATLAB. All submitted algorithms also contain a classification part, which is based on the WEKA machine learning toolbox [2]. In the following subsection we first discuss the audio features set used in our submissions. Then in the subsequent sections we discuss the most important algorithmic details of our submissions and point out the differences to the last year's submission.

## 2. AUDIO FEATURES

In all our submissions we extract the same set of block-level features (BLF), as we did in our last year's submission. Altogether, the extracted feature set consists of the following BLF:

- Spectral Pattern (SP)
- Delta Spectral Pattern (DSP)
- Variance Delta Spectral Pattern (VDSP)
- Logarithmic Fluctuation Pattern (LFP)
- Correlation Pattern (CP)
- Spectral Contrast Pattern (SCP)
- Local Single Gaussian Model (LSG)
- George Tzanetakis Model (GT)



For a detailed description of these features and their extraction process we refer to [7, 9, 10]. Here we only report on this year's optimization of the LFP feature extraction steps. In a series of genre classification experiments we have re-evaluated the individual feature extraction steps of the LFP. The findings of these experiments were that *fluctuation strength weighting* [5] step and the *blurring step* [5] do have a minor negative influence on the classification accuracy. This can be explained by the fact that the FP were initially designed as a feature for music similarity estimation. In this context smoothing the extracted patterns is an important step to increase the similarity of patterns that exhibit similar structure, but at slightly different periodicities. In the context of classification this smoothing step could potentially also blur out discriminant information. While these experiments indicated that we have achieved an improvement in classification accuracy, further experiments based on the combination of all extracted features did only reveal a marginal improvement. This findings were also confirmed by the MIREX 2014 evaluation results. Both algorithms returned almost identical results except for one strange outlier result - *K-POP Genre Classification (Annotated by American Annotators) - Fold 3*.

## 3. GENRE CLASSIFICATION

The genre classification approach itself is rather straight forward. The presented block-level features are combined into a single feature vector that forms the input to the classification stage. To train and predict genre labels the WEKA support vector machine implementation (SMO) is used. Compare to our last year's submission, the genre classification part has not changed at all.

## 4. AUTOMATIC TAG PREDICTION

In general tag prediction can be viewed as a simple extension of the genre classification approach from single to multi-label classification. In tag classification there is, instead of a single classifier like in genre classification, one classifier per tag. A random forest classifier is trained on the full high dimensional feature set and under-sampling is used to balance positive and negative training samples. The tag affinity estimates are then binarized based on a dynamic thresholding approach similar to [4].

## 5. MUSIC SIMILARITY ESTIMATION

Our music similarity estimation approach is based on two distinct components: *Block-Level Feature Similarity* and *Tag Affinity Based Similarity*. The following two subsections present the algorithmic details of these two components. Both components are basically identical to our last year's submission. However, any improvements on the tag prediction part should — according to our assumptions [8] — indirectly help to further improve our similarity algorithm.

### 5.1 Block-Level Feature Similarity

To directly estimate music similarity based on the presented block-level features we follow the approach presented in [9]. First, pairwise song similarities are estimated by computing the Manhattan distance for each of the presented block-level features separately (expect for the LSG pattern which is not used in this task). Then in a second step the individual distance matrices resulting from the individual patterns are combined into a single distance matrix. This is realized by by first normalizing the individual distance matrices using a distance space normalization approach (DSN) [6, 9] and then combining the individual matrices by summing up the corresponding pairwise distances over all matrices. The weights for the contribution of the individual patterns to the overall similarity are the same as last year.

### 5.2 Tag Affinity Based Similarity

For the tag affinity based music similarity as proposed in [1,11] we use a set of about 1500 classifiers pretrained on 4 different tag collections yielding a probabilistic tag affinity vector per song. The training data contained the *Magnatagatune* [3] dataset and three additional datasets. Then for each dataset separately a similarity estimate is derived using the Manhattan distance between the auto-tag vectors of each pair of songs. The similarity estimates resulting from each dataset are then once more combined using the DSN approach.

Finally, to generate the overall similarity matrix the matrices of both components (Block-Level Similarity and Tag Affinity Based Similarity) are simply added to combine them.

## 6. REFERENCES


[1] T. Bertin-Mahieux, D. Eck, F. Maillet, and P. Lamere. Autotagger: A model for predicting social tags from acoustic features on large music databases. *Journal of New Music Research*, 2008.

[2] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: An update. *SIGKDD Explorations*, 2009.

[3] E. Law and L. Ahn. Input-agreement: A new mechanism for collecting data using human computation games. In *Proc. of the 27th Int. Conf. on Human Factors in Computing Systems (CHI-09)*, 2009.

[4] S. R. Ness, A. Theocharis, G. Tzanetakis, and L.G. Martins. Improving automatic music tag annotation using stacked generalization of probabilistic svm outputs. In *Proc. of the 17th ACM Int. Conf. on Multimedia (MM -09)*, 2009.

[5] E. Pampalk, A. Rauber, and D. Merkl. Content-based organization and visualization of music archives. In *Proc. of the 10th ACM Int. Conf. on Multimedia*, 2002.

[6] T. Pohle and D. Schnitzer. Striving for an improved audio similarity measure. In *online Proc. of the 4th Annual Music Information Retrieval eXchange (MIREX-07)*, 2007.

[7] K. Seyerlehner and M. Schedl. Block-level audio feature for music genre classification. In *online Proc. of the 5th Annual Music Information Retrieval Evaluation eXchange (MIREX-09)*, 2009.

[8] K. Seyerlehner, M. Schedl, R. Sonnleitner, D. Hauger, and B. Ionescu. From improved auto-taggers to improved music similarity measures. In *Proc. of the 10th International Workshop on Adaptive Multimedia Retrieval (AMR-12)*, 2012.

[9] K. Seyerlehner, G. Widmer, and T. Pohle. Fusing block-level features for music similarity estimation. In *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, 2010.

[10] K. Seyerlehner, G. Widmer, M. Schedl, and P. Knees. Automatic music tag classification based on block-level features. In *Proc. of the 7th Sound and Music Computing Conference*, 2010.

[11] K. West, S. Cox, and P. Lamere. Incorporating machine-learning into music similarity estimation. In *Proc. of the 1st ACM Workshop on Audio and Music Computing Multimedia (AMCMM-06)*, 2006.