

SUBMISSION TO MIREX 2009 AUDIO SIMILARITY TASK

Tim Pohle¹, Dominik Schnitzer^{1,2}

¹Dept. of Computational Perception

Johannes Kepler University, Linz, Austria

²Austrian Research Institute for Artificial Intelligence (OFAI)
Vienna, Austria

ABSTRACT

This submission to the MIREX 2009 Audio Similarity Task is a variant of the algorithm described in [1]. For comparison to the MIREX 2007 Audio Similarity Task, also the 2007 algorithm is re-submitted, which ranked first in 2007.

1 BRIEF DESCRIPTION

This abstract only contains a brief description of the algorithm components. A more detailed description is planned in the final version of this abstract. The algorithm has two major components which are weighted equally (i.e., 1 : 1), a *rhythm* component and a *timbral* component.

1.1 Rhythm Component

The rhythm component is based on a modification of the Fluctuation Patterns [2], using a *cent/sones* representation (sones values s are estimated from the amplitudes a by $s = 2^{\log_{10} a}$). The rhythm features are described in detail in [1].

1.2 “Timbre” Component

The “timbre” component consists of the well-known MFCCs [3] (coefficients 0..15), Spectral Contrast Feature [4] using the “2N” method [5], and for each frame, two feature values estimating the amount of harmonic and percussive elements in the current audio frame (cf. [6]). Feature values are represented by a single Gaussian.

1.3 Distance Computation

The rhythm and “timbre” distances are calculated separately. Before they are combined, each of the two distance measures is normalized by mean removal and division by standard deviation (based on a track’s distance to all other tracks in the music collection). Symmetry is re-created by subsequently summing up the distances in both directions for each pair of tracks.

2 USAGE

To extract the features:

```
extract_features_ps09(in_file, tmp_dir),
```

where `in_file` is a text file containing the full path to a wav file in each line. `tmp_dir` is a directory where the algorithm has write access. In this directory, the extracted feature data is stored.

To calculate the distances:

```
calc_closest_ps09(tmp_dir, out_file),
```

where `tmp_dir` is the directory where the feature data was stored (i.e., the same directory as in previous call to feature extraction routine), and `out_file` is the name of the file into which the distance matrix should be written. Most entries will be *inf*, except about 100 entries per line.

3 ACKNOWLEDGMENTS

This work is supported by the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung under project number L511-N15.

4 REFERENCES

- [1] Tim Pohle, Dominik Schnitzer, Markus Schedl, Peter Knees, and Gerhard Widmer, “On rhythm and general music similarity,” in *Proceedings of the 10th International Conference on Music Information Retrieval (ISMIR’09)*, 2009.
- [2] E. Pampalk, *Computational Models of Music Similarity and their Application in Music Information Retrieval*, Doctoral dissertation, Vienna University of Technology, Austria, March 2006.
- [3] Beth Logan, “Mel frequency cepstral coefficients for music modeling,” in *Proceedings of the First International Symposium on Music Information Retrieval (ISMIR)*, Plymouth, Massachusetts, oct 2000.
- [4] Dan-Ning Jiang Jiang, Lie Lu, Hong-Jiang Zhang, Jian-Hua Tao, and Lian-Hong Cai, “Music type classification by spectral contrast feature,” in *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, 2002.

- [5] J.-J. Aucouturier and F. Pachet, "Improving timbre similarity: How high is the sky?," *Journal of Negative Results in Speech and Audio Sciences*, vol. 1, no. 1, 2004.
- [6] Nobutaka Ono, Kenichi Miyamoto, Hirokazu Kameoka, and Shigeki Sagayama, "A real-time equalizer of harmonic and percussive components in music signals," in *Proc. International Conference on Music Information Retrieval (ISMIR'08)*, 2008.