2011 IRCAM AUDIO MUSIC SIMILARITY SYSTEM #2

Christophe Charbuillet,Damien Tardieu, Frédéric Cornu, Geoffroy Peeters
Ircam Sound Analysis/ Synthesis Team - CNRS STMS
1, pl. Igor Stravinsky - 75004 Paris - France
{charbuillet,dtardieu,cornu,peeters}@ircam.fr

1. INTRODUCTION

This extended abstract describes the system submitted by IRCAM to the MIREX Evaluation 2011 for the Audio Music Similarity task. This system is similar to the primary system IrcamSimilarity1. The difference is the use of the MIREX database for training the GMM-UBM model. Details about these processing steps are given above.

2. FEATURE EXTRACTION

2.1 Low Level Features

2.1.1 Timbral features

We extract two timbral low level features, Mel Frequency Cepstral Coefficients (MFCC) and Spectral Flatness Measure (SFM). They are extracted using a 40 ms Blackmann window with a 20 ms overlap.

2.1.2 Rythmic features

We extracted the Onset Coefficients as described in [4].

2.2 GMM Supervector

2.2.1 Universal Background Model

The Universal Background Model (UBM) aims at modeling the overall data distribution. It consists of a classical Gaussian Mixture Model. The UBM is usually composed of Gaussian models with diagonal covariance matrix. The loss of modeling ability due the diagonal covariance matrix can be compensated by increasing the number of Gaussian in the mixture The UBM is trained using a large and representative set of data by using the Expectation Maximization (EM) algorithm. The system is provided with a pre trained UBM.

2.2.2 UBM adaptation

The UBM adaptation is the process of modifying the UBM parameters in order to fit a particular data distribution. In our application, this subset is the data extracted from a track to model. This adaptation is made using the Maximum A Prosteriori (MAP) approach [5].

2.2.3 GMM supervector

To summarize, a music track model is directly derived from a generic GMM, estimated using a large set of representative data (the so called UBM). During the adaptation process, only the mean vectors of the Gaussians are modified to fit the particular music track distribution. Consequently, all the the music track models have both the same covariance matrix and weight. Knowing the parameter of the UBM, a particular music model can be summarized by the mean vectors of its Gaussian mixture components. The mean vectors are thus stacked in a one dimensional vector.

2.2.4 GMM supervector for rythm and timbral modeling

We use this model for modeling timbral and rythmic features. UBM were trained on the MIREX database by selecting 1000 audio files for the timbral model and 7000 files for the rhythmic model.

2.3 Autoregressive Vector Model

An autoregressive vector model (ARV) is also computed on the timbra low level feature vectors [1].

2.4 GMM supervector of the residual

A gmmsupervector is built using the reidual of the ARV applied on timbral features.

3. SIMILARITY COMPUTATION

First, all model parameters are normalized by the MCS-norm [2]. Then, similarity matrixes are constructed by a dot product between the model parameters. The obtained matrixes are normalized using the P-norm [3, 2]. The final similarity matrix is obtained by a weithed sum of the four matrixes.

4. ACKNOWLEDGEMENTS

This work was partly supported by “Quaero” Programme, funded by OSEO, French State agency for innovation

5. REFERENCES

[1] F. Bimbot, L. Mathan, A. De Lima, and G. Chollet. Standard and target driven ar- vector models for speech analysis and speaker recognition. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 5–8, 1992.

[2] C. Charbuillet, D. Tardieu, and G. Peeters. Gmm supervector for content based music similarity. In *Proceedings of the 14th Int. Conference on Digital Audio Effects (DAFx-11), year = 2011,.*

[3] T. Pohle and D. Schnitzer. Striving for an improved audio similarity measure. In *4th Annual Music Information Retrieval Evaluation Exchange*, 2007.

[4] T. Pohle, D. Schnitzer, M. Schedl, P. Knees, and G. Widmer. On rhythm and general music similarity. In *Proceedings of the 10th International Conference on Music Information Retrieval (ISMIR'09)*, 2009.

[5] DA Reynolds, TF Quatieri, and RB Dunn. Speaker verification using adapted Gaussian mixture models. *Digital signal processing*, 2000.