MIREX-09 “MUSIC MOOD, MIXED-GENRE, LATIN-GENRE AND CLASSICAL COMPOSER CLASSIFICATION” TASKS: IRCAMCLASSIFICATION08 SUBMISSION

Geoffroy Peeters
Ircam Sound Analysis/ Synthesis Team - CNRS STMS
1, pl. Igor Stravinsky - 75004 Paris - France
peeters@ircam.fr

ABSTRACT

This extended abstract details a submission to the Music Information Retrieval Evaluation eXchange (MIREX) 2009 for the training and classification tasks “Music Mood, Mixed-Genre, Latin-Genre and Classical Composer classification” tasks. Ircam has submitted two systems: ircamclassification08 (GP) which is the same system as the one submitted for MIREX-08 and ircamclassification09 (BP) which is a new version using a larger set of audio features and a full binarization, optimization, SVM classifier. We have submitted ircamclassification08 (GP) to have a baseline performance measure to test the improvement of our system. We review here the system ircamclassification08. The system ircamclassification09 is presented in a separated extended abstract and described in details into [1].

The same system ircamclassification08 has been submitted for the various tasks without any adaptations to the specific problems. The system named ircamclassification08 is a generic system which performs batch feature extraction, models training (using various classifiers) and file indexing (or file segmentation) into classes. The features extracted are generic in order to be applicable to many different audio and music indexing problems. The features are not specific to the above mentioned MIREX09 tasks.

1. SYSTEM DESCRIPTION

Ircamclassification08 is an extension of a system initially developed for instrument-samples indexing described in [2] using the features described in [3]. Only the subset of features applicable to polyphonic audio signals (music) has been used here. In [4] the system has been extended for speech/music segmentation. It is this system that has been used for MIREX09 tasks. We briefly review it in the following.

2. FEATURE EXTRACTION

In the present submission, only three sets of audio features are extracted from the signal.

MFCC: The first set aims at describing the shape of the spectrum at each time. Mel Frequency Cepstral Coefficients (40 Mel bands, 13 coefficients including DC component) are extracted every 20ms using a Blackman window of length 40ms.

SFM/ SCM: MFCCs only describes the shape of the spectrum whatever the content of the signal is noise or sinusoidal (harmonic) components. In order to describe this noise/ sinusoidal content, we also compute height Spectral Flatness [5] and Spectral Crest Measure coefficients. This is done using the same analysis parameters.

Chroma/ PCP: The third set of features gives rough information about the meaning of the harmonic content of the signal. For this, twelve Chroma [6]/ Pitch Class Profiles (PCP) [7] coefficients are computed using a Blackman window of length 100ms synchronized in time with the two other feature sets.

Delta and acceleration coefficients of the above mentioned features are also computed.

Finally, a simple temporal modelling (mean and standard deviation) of each feature is performed using a sliding window of length 500ms and a hop size of 250ms.

3. MODELS TRAINING

Training of the class-models is performed using the following steps:

Feature processing: Features are first normalized and outliers are removed (based on IQR).

Feature selection: The Inertia Ratio Maximization with Feature Space Projection (IRMFSP) algorithm [2] is used to select independently the best 40 features (independently means that we don’t take into account the set the features belong to).

Feature space transform: Linear Discriminant Analysis is then applied to the reduced feature space.

Class modelling

Class modelling is done in two stages

First stage: frame-statistical-model We first model the belonging of each frame to each class using a simple Gaussian Mixture Models (8 Gaussians, full matrix). For this we use all the feature vectors $\underline{f}(t)$ for all the time $t \in J_k$ where J_k is the set of tracks labelled as belonging to class k. We call this model a frame-statistical model: it gives the probability to observe class k given the feature vector at time t: $p(t \in c_k|\underline{f}(t))$. As explained in [4], the labels are assigned to the tracks (a collection of frames) and not independently to the frames. A track of a given class may in fact include frames from another class: a track labelled as rock may contain frames belonging to the blues class. It is the succession of the frame-belongings that makes the track being rock. We model this in the second stage of the classifier.

Second stage: track-statistical-model In the second stage we model the probability that the whole track belong to a class given the set of probability-vectors of its frame: $p(J \in c_k|\underline{p}(t_J \in c)$, where J is a track, t_J is the set of frames belonging to track J and $\underline{p}(t \in c)$ is the probability-vector coming from the frame-statistical model. For this, the whole training set is first classified using the frame-statistical-model. For each track belonging to class c_k we then study the belonging of its frames over time. This allows creating a track-statistical model.

4. CLASSIFICATION

The classification of an unknown track is also performed in two stages:

- first at the frame level using $p(t \in c_k|\underline{f}(t))$,
- then at the track level using $p(J \in c_k|\underline{p}(t_J \in c)$.

The training and classification process is represented in Figure 1.

5. RESULTS

5.1 Audio Music Mood Classification

The results for Audio Music Mood classification are indicated into Table 1. Ircamclassification08 scored exactly the same as last year (same system, same test-set): 63.67%. Although it ranked first last year, a new system (CL1,CL2) has now run over it 65.67%. It is still the second best system for Music Mood classification. Surprisingly, for this specific task, ircamclassification08 (GP) performed better than ircamclassification09 (BP2). It is also worth mentioning the fact that the performances of all systems have increased since last year.

5.2 Audio Mixed and Latin Genre Classification

The results for Audio Music Mixed-Genre and Latin-Genre are indicated into Table 2 and Table 3. Ircamclassification08 scored pretty close to the performance obtained last year: 64.24% (63.90% in 2008) for mixed-genre and 62.63% (62.72% in 2008) for Latin-genre. However, because the performances of all systems have increased a lot since last year, ircamclassification08 (GP) is not anymore in the top ranking of Mixed Genre. It is still in the top ranking of Latin Genre but not in the second place as last year.

The new version of it, ircamclassification09 (BP) ranked third for mixed-genre (70.63%) and second for Latin-genre (67.31%).

Audio Music Mood	
Participant	**Mean Accuracy**
CL1	65,67%
CL2	65,50%
GP	**63,67%**
MTG5	62,83%
HW2	61,67%
LZG	61,67%
HW1	61,33%
GLR1	60,83%
FCY1	60,33%
VA2	60,17%
XZZ	60,00%
MTG3	59,83%
BP2	**59,67%**
MTG6	59,50%
GT1	59,33%
MTG4	59,33%
VA1	59,33%
SS	58,83%
HNOS1	58,67%
HNOS3	58,67%
FCY2	58,33%
BP1	**58,17%**
MTG1	57,67%
MTG2	57,50%
XLZZG	57,00%
GT2	56,83%
TAOS	56,83%
RK1	53,17%
GLR2	53,00%
HNOS4	51,17%
ANO	50,67%
RK2	41,33%
HNOS2	34,67%

Table 1. Audio Music Mood Classification results

Audio Genre Classification (Mixed Set)		
Participant	**Mean Accuracy**	**Mean Discounted**
CL2	73,33%	80,61%
CL1	73,23%	80,48%
GLR1	71,23%	78,51%
BP1	**70,63%**	**77,61%**
MTG5	70,44%	77,69%
XZZ	69,36%	77,25%
XLZZG	68,93%	76,54%
VA1	68,84%	76,53%
BP2	**68,51%**	**76,21%**
LZG	68,29%	76,29%
TTOS	67,89%	76,47%
GT2	67,87%	76,21%
VA2	67,39%	75,56%
SS	66,60%	74,54%
HW1	65,99%	74,33%
HW2	65,31%	73,68%
GT1	65,10%	73,68%
MTG1	64,79%	73,05%
HNOS1	64,47%	72,96%
HNOS3	64,34%	72,97%
GP	**64,24%**	**72,97%**
MTG3	64,06%	71,95%
MTG4	64,00%	71,69%
RK1	61,41%	70,20%
ANO	60,50%	70,60%
GLR2	60,14%	69,11%
RCJ4	50,99%	61,20%
HNOS4	45,16%	55,09%
RCJ3	37,71%	49,46%
RCJ1	32,50%	44,08%
HNOS2	20,90%	23,22%

Figure 2. Audio Mixed Genre Classification results

5.3 Audio Classical Composer Identification

The results for Audio Classical Composer are indicated into Table 2. Ircamclassification08 scored pretty close to the performance obtained last year: 48.85% (48.99% in 2008). However, because the performances of all systems have increased a lot since last year, ircamclassification08 (GP) is not anymore in the top ranking.

The new version of it, ircamclassification09 (BP) ranked fifth in Classical Composer (55.66%).

5.4 Audio Tag Classification

As last year, ircamclassification08 (GP) was also submitted for the task of Audio Tag Classification (MajorMiner

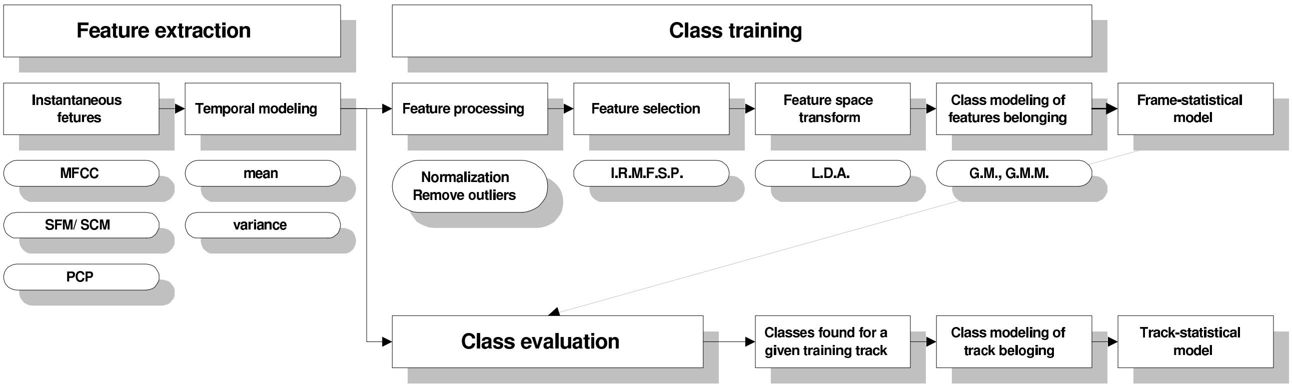

Figure 1. Flowchart of the two stages training and classification system

Audio Genre Classification (Latin Set)		
Participant	**Mean Accuracy**	**Mean Discounted**
CL1	74,66%	83,17%
CL2	73,58%	82,54%
BP1	**67,31%**	**77,47%**
SS	64,69%	75,92%
BP2	**63,52%**	**76,13%**
MTG6	63,16%	74,95%
GLR1	62,79%	75,70%
GP	**62,63%**	**73,63%**
MTG2	62,39%	74,47%
MTG1	61,68%	73,81%
MTG5	61,14%	73,36%
VA1	58,37%	72,76%
RK1	57,11%	68,56%
HNOS1	56,32%	68,26%
HNOS3	56,22%	67,85%
LZG	55,96%	69,38%
XLZZG	55,25%	68,61%
XZZ	55,25%	69,29%
RCJ4	55,22%	67,28%
HW1	54,72%	68,97%
VA2	54,49%	69,96%
TTOS	53,70%	67,01%
GT2	52,82%	65,05%
RCJ2	52,43%	66,90%
HW2	52,28%	67,21%
GLR2	49,84%	65,33%
GT1	49,75%	62,90%
MTG4	47,79%	63,61%
RCJ3	46,78%	61,72%
MTG3	45,80%	62,31%
RCJ1	38,93%	55,26%
ANO	38,87%	55,78%
HNOS4	30,05%	51,43%

Figure 3. Audio Latin Genre Classification results

set and Mood set). For the same reasons as last year, very unbalanced training-sets, the algorithm failed to learn the characteristics of the classes. This is actually the prime reason for starting the development of ircamclassification09 (BP).

6. CONCLUSION

This extended abstract reviewed the results obtained on the classification task with Ircam 2008 submission, ircamclassifiction08 (GP). The goal was to have a baseline performance measure to test the improvement of the new version of the system, ircamclassification09 (BP). As expected, the performances of ircamclassification09 were better during MIREX-09 than the old system, except for the task Music Mood were ircamclassification08 still behaves as an excellent system by ranking second.

Audio Classical Composer	
Participant	**Mean Accuracy**
MTG2	62,05%
CL1	60,97%
CL2	60,03%
XZZ	57,18%
HW1	56,35%
BP1	**55,66%**
GLR1	55,34%
BP2	**54,76%**
MTG1	54,73%
LZG	54,40%
VA1	53,57%
VA2	53,57%
XLZZG	53,54%
HW2	53,10%
SS	52,56%
GT2	51,48%
MTG6	50,36%
MTG5	49,75%
GP	**48,85%**
RK1	48,41%
MTG4	48,20%
MTG3	48,12%
GLR2	45,92%
TTOS	44,37%
GT1	43,69%
HNOS1	43,33%
HNOS3	42,24%
ANO	41,77%
HNOS4	29,04%
HNOS2	15,84%

Table 2. Audio Classical Composer Identification results

7. ACKNOWLEDGEMENTS

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

8. REFERENCES

[1] J-J. Burred and G. Peeters. An adaptive system for music classification and tagging. In *3rd Int. Workshop on Learning the Semantics of Audio Signals (LSAS)*, Graz, Austria, 2009.

[2] G. Peeters. Automatic classification of large musical instrument databases using hierarchical classifiers with inertia ratio maximization. In *Proc. of AES 115th Convention*, New York, USA, 2003.

[3] G. Peeters. A large set of audio features for sound description (similarity and classification) in the cuidado project. Cuidado project report, Ircam, 2004.

[4] G. Peeters. A generic system for audio indexing: application to speech/ music segmentation and music genre. In *Proc. of DAFX*, Bordeaux, France, 2007.

[5] O. Izmirli. Using a spectral flatness based feature for audio segmentation and retrieval. In *Proc. of ISMIR*, Pymouth, Massachusetts, USA, 2000.

[6] G. Wakefield. Mathematical representation of joint time-chroma distributions. In *Proc. of SPIE conference on Advanced Signal Processing Algorithms, Architecture and Implementations*, pages 637–645, Denver, Colorado, USA, 1999.

[7] T. Fujishima. Realtime chord recognition of musical sound: a system using common lisp music. In *Proc. of ICMC*, pages 464–467, Bejing, China, 1999.