A DEMONSTRATOR FOR AUTOMATIC MUSIC MOOD ESTIMATION

Janto Skowronek, Martin McKinney, Steven van de Par
Philips Research Laboratories
Hightech Campus 36, 5656 AE Eindhoven, The Netherlands

ABSTRACT

Interest in automatic music mood classification is increasing because it could enable people to browse and manage their music collections by means of the music's emotional expression complementary to the widely used music genres. We continue our work on designing a well defined ground-truth database for music mood classification and show a demonstrator of automatic mood estimation. While a subjective evaluation of this algorithm on arbitrary music is ongoing, the initial classification results are encouraging and suggest that an automatic predicition of music mood is possible.

1 INTRODUCTION

In the development of an automatic music mood classifier, we must treat the high degree of subjectivity associated with mood. One can do this by developing personalized models requiring comprehensive user feedback for training, or one can develop generalized models by minimizing the subjectivity involved in the evaluation of music mood. Here we apply the latter approach. A first step in minimizing subjectivity is to deliberately define the goal of the classification algorithm: it should model how people would describe the mood expressed in the music (*affect attribution*) and not how they would actually feel when they are listening to the music (*affect induction*). See also [1] for a detailed discussion on affect attribution and induction. A second step is to define mood classes on which users show relative agreement when applying them to music. In addition - from an application point of view - the mood classes should be *easy to use* and *important* to the users. Our previous work [8] reported on the identification of those consistent-, important- and easy-mood classes; some issues will be recapitulated in Section 2. As a third step the ground-truth database used for training and testing the desired classifier should contain only music that people can clearly and consistently assign to or exclude from a mood category. This issue will be discussed in more detail in Sections 3 and 4.

2 CLASS DEFINITION

In previous work [8] we searched for mood classes that enable the development of the above mentioned generalized mood classification models. The basic idea was to ask subjects to evaluate music excerpts using 33 candidate labels used in literature (e.g. [1, 6]), followed by a data analysis in order to identify those mood classes on which subjects show a certain agreement and which subjects assessed as *important* and *easy to use*. As a measure for across subject consistency we computed the Cronbach's coefficient α and as a measure for the importance and easiness of the classes we used questionnaires. In [8] we identified 8 mood classes that fulfilled our criteria ($\alpha \geq 0.7$, at least *important*, at least *easy to use*).

A more detailed analysis suggested a slight modification to that selection of mood classes. First, some mood classes that were highly appreciated in the questionnaires got an α slightly below 0.7. Therefore we lowered our threshold to 0.65. Second, there were very high correlations between some classes, whose meaning of the class names (adjectives) were quite close. Thus we decided to merge them. Third, we loosened the criteria *importance* and *easiness* due to the small number of subjects (10): a mood class was selected if at least one of the two criteria was fulfilled. Hence the new set of mood classes comprised 12 categories: arousing-awakening, angry-furious-agressive, calming-soothing, carefree-lighthearted-light-playful, cheerful-festive, emotional-passionate-touching-moving, sad, loving-romantic, powerful-strong, restless-jittery-nervous, peaceful, tender-soft.

3 MATERIAL COLLECTION

In a follow-up experiment subjects labelled a large number of music excerpts using the 12 identified mood classes. The idea was to select those excerpts that got a clear and consistent rating by the subjects to compile a ground-truth database for training and testing a mood classification algorithm. In our previous experiment [8] many excerpts did not fulfill our criteria for conveying a clear mood. Hence the new set-up aimed at obtaining a large number of labelled excerpts to ensure a sufficient number of training material, resulting in compromises with respect to the number of involved subjects and an unbalanced distribution of excerpts across music genres.

As in the first experiment, we selected the excerpts such that their moods are likely to be constant by avoiding drastic changes in the music (structure, tempo, instrumentation etc.) We distributed 1059 excerpts from 12 music genres across 12 subjects such that each excerpt was rated by 6 subjects; each subject had to rate 530 excerpts and the sets per subject were mutually overlapping. The subjects were students and employees working at our laboratory(age between 23 and 40, 8 different nationalities, musical practise from 0 to 21 years). Three of them participated already in the previous experiment.

We collected the subjective judgments on a 4-point scale: *not*, *slightly*, *moderately*, *definitively that mood*.

4 CLASSIFICATION ALGORITHM

In contrast to the mood classes defined by Lu et al. [2], our mood classes are not mutually exclusive. Consequently, we implemented an individual detector for each mood class (binary decision: 1 = *that class*, 2 = *not that class*). For that purpose we had to decide for each excerpt and each mood category, whether the excerpt belongs to class 1 or class 2 or whether it should be excluded. An excerpt was accepted if it was relatively consistently rated across subjects (standard deviation $\leq$ 1 point on the scale) and if its mean rating was not in an area of ambiguity, which we defined between *not that mood* and *slightly that mood*. Depending on the mood label between 5% (sad) and 20% (restless) of the 1059 excerpts were accepted for class 1, between 27% (emotional) and 66% (angry) for class 2. While the consistency criterion was often fulfilled, the unambiguity criterion excluded most of the excerpts.

The used feature extraction algorithm computes every 743ms a feature vector comprising 4 general feature types: basic signal describing features and their temporal modulation [3], perceptually relevant tempo and rhythm based features [4], features based on chroma and key information [5] as well as features that evaluate the occurences of percussive sound events in the music [7].

The classification stage is based on quadratic discriminant analysis and used a randomized 80/20 split of the training and test data in combination with bootstrap repetitions in order to estimated the classification performance. Results are shown in Table 1. While some classes show about 75-80 % correct classifciation (e.g. carefree, loving), the performance of other classes (e.g. angry, calming) is with about 90% rather good. In general these results show that an automatic estimation of music mood, as we defined it in Section 1, is possible.

Mood Class	Performance
arousing-awakening	85.1 $\pm$ 1.7 %
angry-furious-agressive	90.1 $\pm$ 1.6 %
calming-soothing	90.9 $\pm$ 2.4 %
carefree-lighthearted-light-playful	77.1 $\pm$ 1.7 %
cheerful-festive	79.8 $\pm$ 1.9 %
emotional-passionate-touching-moving	82.2 $\pm$ 1.1 %
loving-romantic	80.2 $\pm$ 1.7 %
peaceful	88.5 $\pm$ 1.7 %
powerful-strong	80.7 $\pm$ 1.8 %
sad	85.0 $\pm$ 1.8 %
restless-jittery-nervous	90.6 $\pm$ 1.5 %
tender-soft	89.5 $\pm$ 0.6 %

Table 1. Classification performance (mean $\pm$ standard error across bootstrap repetitions) of the individual mood detectors.

5 CONCLUSIONS

The aim of this work is the development of a music mood classification algorithm. The applied procedure of groundtruth database design was intended to minimize the subjectivity that is involved in the perception of music mood. Under these constraints the achieved classification results as well as the behavior of our real-time demonstrator show that music mood estimation is to a certain extend possible. While a subjective evaluation of the classification algorithm on abitrary music is currently being performed, open issues for future research are to further investigate why so many excerpts seem to be ambiguous and how to deal with such music pieces.

6 REFERENCES

[1] M. Leman, V. Vermeulen, L. De Voogdt, D. Moelants, M. Lesaffre, *Prediction of Musical Affect Using a Combination of Acoustic Structural Cues*, J. of New Music Research, Vol. 34(1), 39-67, 2005.

[2] L. Lu, D. Liu, H. Zhang, *Automatic Mood Detection and Tracking of Music Audio Signals*, IEEE transactions on audio, speech, and language processing, Vol. 14(1), 5-18, 2006.

[3] M. McKinney, J. Breebart, *Features for Audio Music Classification*, Proceedings of 4th International Conference on Music Information Retrieval, 2003.

[4] M. McKinney, D. Moelants, *Extracting the perceptual tempo from music audio*, Proceedings of 5th International Conference on Music Information Retrieval, Barcelona, 2004.

[5] S. van de Par, M. McKinney, A. Redert, *Musical Key Extraction from Audio using Profile Training*, Proceedings of 7th International Conference on Music Information Retrieval, Victoria, 2006.

[6] J. Russell, *A circumplex model of affect*, J. Personality & Social Psychology, Vol. 39, 1161-1178, 1980.

[7] J. Skowronek, M. McKinney, *Features for audio classification: Percussiveness of sounds*, Intelligent Algorithms in Ambient and Biomedical Computing, Philips research Book Series Vol. 7, Springer, 2006.

[8] J. Skowronek, S. van de Par, M. McKinney, *Groundtruth for automatic music mood classification*, Proceedings of 7th International Conference on Music Information Retrieval, Victoria, 2006.