MIREX-2012 "AUDIO TEMPO ESTIMATION" TASK: IRCAMTEMPO-GMMREG SUBMISSION

Geoffroy Peeters
STMS IRCAM-CNRS-UPMC 1, pl. Igor Stravinsky - 75004 Paris - France

ABSTRACT

This extended abstract details a submission to the Music Information Retrieval Evaluation eXchange (MIREX) 2012 for the Audio Tempo Estimation task. We submitted a prototype software [1] corresponding to the GMM-Regression method we recently proposed in [2] for the estimation of perceptual tempo. We briefly summarize this new method here.

1. GMM-REGRESSION FOR PERCEPTUAL TEMPO

For the estimation of the perceptual tempo, we start from a set of assumptions. For each of these assumptions we create a related audio feature. We assume that the perception of tempo is related to the rate of variation of four musical attributes: $-1-$ the rate of variation of energy (as did the previous works) but also $-2-$ the rate of variation of harmonic content, $-3-$ the rate of variation of spectral balance (the distribution in high or low frequency of the energy) and $-4-$ the rate of short-term-event-repetitions. We assume that a track with a rapid chord changes, rapid spectral-balance changes or rapid short-term repetitions will be perceived as fast even if the tempo of the sequencer was set to slow.

We then create a model to find the relationship between the perceptual tempo, the perceptual tempo class and the four feature-sets. This model is then used to predict the perceptual tempo given the audio features. The model and the prediction is done using a technique borrowed from speech processing: GMM-Regression [1]. In [1], a GMM model is trained to learn the relationship between pitch and spectral envelope. The model is then used to predict the most-likely pitch given an observed spectral envelope. This is done using a regression over the values of the most-likely components of the GMM. The same method is applied here to predict the most-likely perceptual tempo given the audio features.

The feature extraction, training and regression processes of our method are illustrated in Figure 1. Details of the method can be found in [2].

[1] For the MATLAB©envirronment

This document is licensed under the Creative Commons Attribution-Noncommercial-Share Alike 3.0 License.
http://creativecommons.org/licenses/by-nc-sa/3.0/
© 2010 The Authors.

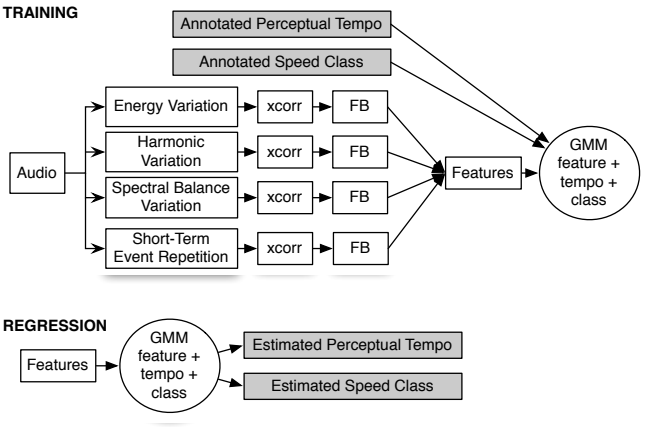

Figure 1. Overall schema of the proposed GMM training and GMM-Regression method

1.1 Configuration used:

For the MIREX-2012, the GMM used for the regression has been trained using audio extracts from the following data-sets: LastFM-Percepual-Tempo (see [2]), Ballroom, Songs, Klapuri, Davies, Hainsworth, Beatles and PopRock. The GMM has 16 components.

1.2 Estimating two tempi

The main tempo is estimated using the GMM-Regression method we proposed in [2], we denote it by τ_1.

We then test the likelihoods of the following candidates for the second tempo: $\tau_2 = [1/4, 1/3, 1/2, 1.5, 2, 3, 4] \cdot \tau_1$. To estimate these likelihoods we compute the probability of observing simultaneously τ_2 and the feature vector. We then choose the τ_2 with the largest likelihood.

The method used to estimate τ_2 is therefore close to the method proposed by Xiao in [3]. While Xiao only used timbre-based features (MFCC) we use here our four-feature-sets to compute this likelihood.

2. MIREX-2012 RESULTS AND DISCUSSIONS

TO DO

3. CONCLUSION

TODO

4. ACKNOWLEDGMENTS

This work was partly supported by the Quaero Program funded by Oseo French agency and by the MIReS project funded by EU-FP7-ICT-2011.1.5-287711.

5. REFERENCES

[1] T. En-Najjary, O. Rosec, and T. Chonavel. A new method for pitch prediction from spectral envelope and its application in voice conversion. In *Proc. of Eurospeech*, Geneva, 2003.

[2] Geoffroy Peeters and Joachim Flocon-Cholet. Perceptual tempo estimation using gmm regression. In *ACM Multimedia, MIRUM Workshop*, Nara, Japan, October 2012.

[3] Linxing Xiao, Aibo Tian, Wen Li, and Jie Zhou. Using a stastic model to capture the association between timbre and perceived tempo. In *Proc. of ISMIR*, Philadeplhia, USA, 2008.