THE 2009 LABROSA PRETRAINED AUDIO CHORD RECOGNITION SYSTEM

Daniel P.W. Ellis
LabROSA, Columbia University
New York NY USA
dpwe@ee.columbia.edu

1 OVERVIEW

Our pre-trained audio chord recognition system relies on labeled data to train Gaussian models of each chord class, based on a beat-synchronous chroma representation developed for cover song detection [1]. All chord models are based on two prototype models, one for major chords and one for minor, which are trained on all available examples, suitably transposed to align their tonality. Chord recognition is accomplished with a hidden Markov model using the per-chord Gaussians to calculate observation likelihoods [4].

The system estimates a single key for each track, based on a single Gaussian full-covariance model of the chroma distribution, then transposing the track's chroma representation to maximize its likelihood under this model. This key model was trained using ground-truth labels for each track as reported in the Allan Pollack annotations [3]. Recognition then uses a transition matrix derived from the training labels which is normalized relative to the estimated key.

Our submission is a complete model that we trained in-house on all 180 tracks corresponding to the set of chord transcriptions released by Chris Harte [2]. Testing this model on the same 180 songs gives a time-weighted accuracy of around 74% (over a set of 25 classes consisting of 12 Major, 12 minor, and one "no chord" class). Four-way cross-validated training of the same approach, in which testing albums were disjoint from the training set, yielded performance less than 2% lower absolute.

This system is essentially the same as the one we submitted to the "train-test" chord evaluation in MIREX 2008, except for the key-estimation and key-relative transition matrices. We are also using maximal gamma values (from forward-backward calculation) instead of Viterbi path to compute the final reported class. Both of these changes contributed improvements of under 2% absolute each.

The main change is that we have correctly aligned Harte's labels to our local versions of the audio, based on a forced alignment of earlier models, followed by manual tuning of the results. We found timing skews of the order of 1 second or more, and overall tempo differences of up to 1%, in almost all files, with no clear systematic pattern. Fixing these alignments for both train and test data improved reported system performance by close to 20% absolute, and made it worthwhile to submit a pre-trained system.

The entire codebase and all the data used in training the model is available at `http://labrosa.ee.columbia.edu/projects/chords/`.

2 ACKNOWLEDGMENTS

Thanks to Jesper Højvang Jensen for the accelerated calculation of the instantaneous-frequency chromagram.

This work was supported by the National Science Foundation (NSF) under Grant No. IIS-0713334. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

3 REFERENCES

[1] D. P. W. Ellis and G. Poliner. Identifying cover songs with chroma features and dynamic programming beat tracking. In *Proc. ICASSP*, pages IV–1429–1432, Hawai'i, 2007.

[2] C.A. Harte and M.B. Sandler. Automatic chord identification using a quantised chromagram. In *Proceedings of the Audio Engineering Society 118th Convention*, Barcelona, 2005.

[3] Alan W. Pollack. Notes on ... series. `http://www.recmusicbeatles.com/public/files/awp/awp.html`, 2001.

[4] Alexander Sheh and Daniel P. W. Ellis. Chord segmentation and recognition using EM-trained hidden markov models. In *Proc. Int. Conf. on Music Info. Retrieval ISMIR-03*, pages 185–191, 2003.