MIREX 2016 AUDIO DOWNBEAT TRACKING SUBMISSIONS: KB1, KB2

Florian Krebs
Department of Computational Perception
Johannes Kepler University, Linz, Austria

Sebastian Böck
Department of Computational Perception
Johannes Kepler University, Linz, Austria

ABSTRACT

In this submission we use the hybrid RNN-HMM model [13] to identify the downbeat positions within an audio signal. First, two feature streams are extracted: A spectral flux with 45 logarithmically spaced frequency bins, and the 12 bin CLP chroma feature [16]. These audio features are synchronized to the beat, which is extracted by a beat trackler [3, 15]. The beat-synchronized features are then fed into two parallel RNNs, which output a probability that indicates wether a beat is a downbeat or not. This output activation is then further post processed with an HMM that finally yields the most probable downbeat-beat sequence.

1. MODEL DESCRIPTION

The model structure is identical to the one published in [13]. Please see the paper for further details.

2. SUBMISSIONS

We submitted two versions of the model (KB1 and KB2). The two submissions use the same number of parameters, but differ in the data they were trained on. The used training sets for each submission are listed in Table 2. Note that we train the beat tracking RNN separately from the downbeat tracking RNN. As training the beat tracking RNN takes approximately one week, we trained it with only two configurations.

3. ACKNOWLEDGMENTS

This work is supported by the European Union Seventh Framework Programme FP7 / 2007-2013 through the GiantSteps project (grant agreement no. 610591), and the Austrian Science Fund (FWF) project Z159. For this research, we have made extensive use of free software, in particular Python, Lasagne, Theano and GNU/Linux. The Tesla K40 used for this research was donated by the NVIDIA corporation.

4. REFERENCES

[1] J.P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M.B. Sandler. A tutorial on onset detection in music signals. *IEEE Transactions on Speech and Audio Processing*, 13(5):1035–1047, 2005.

[2] S. Böck, F. Krebs, and M. Schedl. Evaluating the online capabilities of onset detection methods. In *Proceedings of the 14th International Conference on Music Information Retrieval (ISMIR)*, Porto, 2012.

[3] S. Böck, F. Krebs, and G. Widmer. A multi-model approach to beat tracking considering heterogeneous music styles. In *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR)*, Taipei, 2014.

[4] M.E.P. Davies, N. Degara, and M.D. Plumbley. Evaluation methods for musical audio beat tracking algorithms. *Queen Mary University of London, Tech. Rep. C4DM-09-06*, 2009.

[5] T. De Clercq and D. Temperley. A corpus analysis of rock harmony. *Popular Music*, 30(01):47–70, 2011.

[6] B. D. Giorgi, M. Zanoni, A. Sarti, and S. Tubaro. Automatic chord recognition based on the probabilistic modeling of diatonic modal harmony. In *Proceedings of the 8th International Workshop on Multidimensional Systems*, 2013.

[7] M. Goto. AIST annotation for the RWC music database. In *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR)*, pages 359–360, Victoria, 2006.

[8] F. Gouyon, A. Klapuri, S. Dixon, M. Alonso, G. Tzanetakis, C. Uhle, and P. Cano. An experimental comparison of audio tempo induction algorithms. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(5):1832–1844, 2006.

[9] S. Hainsworth and M. Macleod. Particle filtering applied to musical tempo tracking. *EURASIP Journal on Applied Signal Processing*, 2004:2385–2395, 2004.

[10] J. Hockman, M. Davies, and I. Fujinaga. One in the jungle: Downbeat detection in hardcore, jungle, and drum and bass. In *Proceedings of the 13th International Society for Music Information Retrieval (ISMIR)*, Porto, 2012.

[11] A. Holzapfel, M.E.P. Davies, J. Zapata, J. Oliveira, and F. Gouyon. Selective sampling for beat tracking evaluation. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(9):2539–2548, 2012.

Submission	beat train set	downbeat train set	supported meters	min_tempo	max_tempo
KB1	beat1	db1	3/4, 4/4	55	215
KB2	beat2	db2	3/4, 4/4	55	215

Table 1. Overview of submitted systems and their parameters

Identifier	# files	meters	contents
beat1	1124	2/4, 3/4, 4/4, 6/8	Ballroom [8, 14], Hainsworth [9], SMC [11]
beat2	3472	2/4, 3/4, 4/4, 6/8	Ballroom [8, 14], Beatles [4], Hainsworth [9], RWC [7], Rock [5], Böck [1, 2], Turkish [17], SMC [11], Simac [1] , Klapuri [12], Hjdb [10], Cuidado [19], Carnatic [18]
db1	1371	3/4, 4/4	Ballroom [8, 14], Hainsworth [9], RWC pop [7], Robbie Williams [6], Rock [5], Böck [1, 2]
db2	1328	3/4, 4/4	Ballroom [8, 14], Beatles [4], RWC pop [7],Robbie Williams [6], Rock [5], Böck [1, 2]

Table 2. Training datasets

.

[12] A. Klapuri, A. Eronen, and J. Astola. Analysis of the meter of acoustic musical signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1):342–355, 2006.

[13] F. Krebs, S. Böck, M. Dorfer, and G. Widmer. Downbeat tracking using beat-synchronous features and recurrent neural networks. In *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR)*, New York, 2016.

[14] F. Krebs, S. Böck, and G. Widmer. Rhythmic pattern modeling for beat and downbeat tracking in musical audio. In *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR)*, Curitiba, 2013.

[15] F. Krebs, S. Böck, and G. Widmer. An efficient state space model for joint tempo and meter tracking. In *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR)*, Malaga, 2015.

[16] Meinard Müller and Sebastian Ewert. Chroma Toolbox: MATLAB implementations for extracting variants of chroma-based audio features. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR)*, Miami, 2011.

[17] A. Srinivasamurthy, A. Holzapfel, and X. Serra. In search of automatic rhythm analysis methods for Turkish and Indian art music. *Journal for New Music Research*, 43(1):94–114, 2014.

[18] A. Srinivasamurthy and X. Serra. A supervised approach to hierarchical metrical cycle tracking from audio music recordings. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5217–5221, Florence, 2014.

[19] H. Vinet, P. Herrera, and F. Pachet. The cuidado project. In *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR)*, pages 197–203, Paris, 2002.