A MULTI-MODEL APPROACH TO BEAT TRACKING CONSIDERING HETEROGENEOUS MUSIC STYLES

Department of Computational Perception

JOHANNES KEPLER UNIVERSITY LINZ

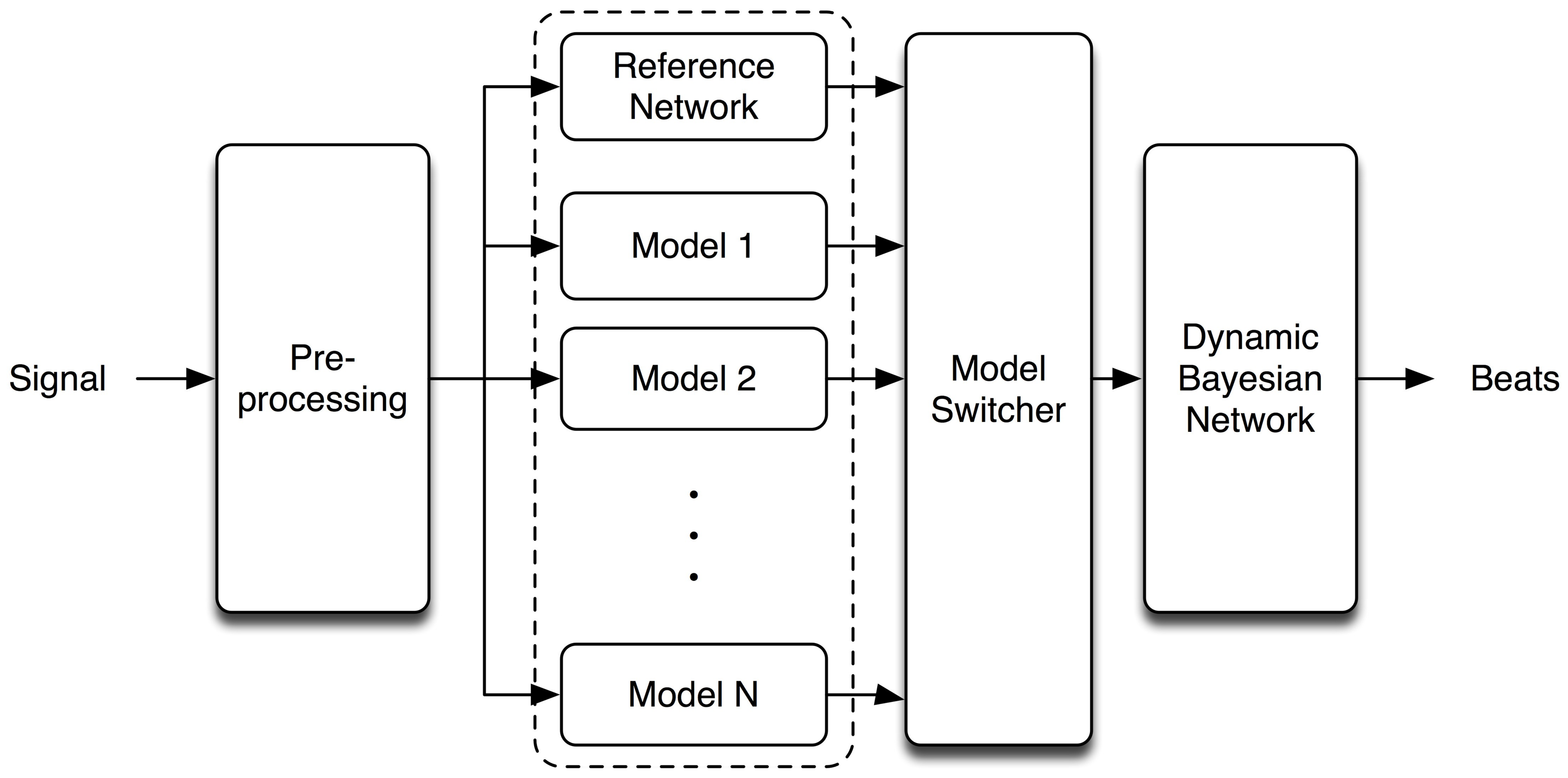
Sebastian Böck, Florian Krebs, Gerhard Widmer

sebastian.boeck@jku.at | http://www.cp.jku.at

Abstract

In this paper we present a new beat tracking algorithm which extends an existing state-of-the-art system with a multi-model approach to represent different music styles. The system uses multiple recurrent neural networks, which are specialised on certain musical styles, to estimate possible beat positions. It chooses the model with the most appropriate beat activation function and jointly models the tempo and phase of the beats from this activation function with a dynamic Bayesian network. We test our system on three huge datasets of various styles and report performance gains of up to 27% over existing state- of-the-art methods. Under certain conditions the system is even able to match human tapping performance.

System Overview



Pre-processing

Short Time Fourier Transform (STFT) of the input signal

- 3 parallel STFT with different window sizes
- logarithmically filtered magnitude spectrograms
- positive 1st order differences

Neural Networks

Bidirectional Recurrent Neural Networks

- linear input layer
- 3 fully connected hidden layers à 25 LSTM units
- sigmoid output layer
- reference network trained on all data
- specialised networks re-trained on specific subsets (different musical styles)

Model Switcher

Chooses the most appropriate model by

- comparing the predictions of each specialised model to the reference model
- selecting the one with the lowest mean squared difference

Dynamic Bayesian Network

Adapted the bar pointer model [2] to a beat pointer model

Transition model:

- model every beat with 640 equidistant beat states
- allow 46 discrete tempo states (≈ 56 – 215 bpm)
- set tempo change probability to 0.002

Observation model:

- use the network's output directly as state-conditional observation distribution
- Consider the first 16th of the beat period as beat location, the remainder as non-beat

Initial state distribution:

- uniform distribution

Inference of beats:

- use the Viterbi algorithm to infer the most probable state sequence
- use the frame with the highest NN activation within the beat portion of the interval as the final beat position

Results

	F-measure	Cemgil	AMLc	AMLt
Ballroom				
BeatTracker [1]	0.887	0.855	0.748	0.831
– Multi-Model	0.897	**0.866**	0.759	0.841
– DBN	0.903	0.838	0.873	0.915
– Multi-Model + DBN	**0.910**	0.845	**0.885**	**0.924**
Krebs et al. [2]	0.855	0.772	0.818	0.865
Zapata et al. [3]	0.767	0.672	0.824	0.860
SMC				
BeatTracker [1]	0.497	0.402	0.279	0.436
– Multi-Model	0.514	**0.415**	0.296	0.467
– DBN	**0.516**	0.404	0.378	0.550
– Multi-Model + DBN	**0.529**	**0.415**	**0.383**	**0.567**
Zapata et al. [3]	0.369	0.285	0.239	0.397

Acknowledgements

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.

References

[1] S. Böck and M. Schedl. *Enhanced Beat Tracking with Context-Aware Neural Networks*. Proc. of the 14th International Conference on Digital Audio Effects (DAFx), 2011.

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

[3] J. Zapata, M. Davies and E. Gómez. *Multi-feature beat tracking*. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(4), 2014.

The 15th International Society for Music Information Retrieval Conference (ISMIR), 27th – 31st October 2014, Taipei, Taiwan