State-of-the-art Beat Tracking

MUMT 621 Presentation 5. March 30, 2021. Sevag Hanssian, 260398537

Summary

McFee and Ellis (2014) state that onset detection is a critical first stage of most beat tracking algorithms. Works by Goto (2002), Ellis (2007), and Stark, Davies, and Plumbley (2009) are some examples among many of onset-based beat tracking. Bello et al. (2005) provide multiple definitions for onsets, but the simplest is that of the beginning of musical events. A typical onset-based beat tracker, as described by McFee and Ellis (2014), operates as follows:

> First, the audio signal is processed by an onset strength function, which measures the likelihood that a musically salient change (e.g., note onset) has occurred at each time point. The tracking algorithm then selects the beat times from among the peaks of the onset strength profile

Eyben et al. (2010) used a recurrent neural network architecture to achieve state-of-the-art results in onset detection. This was first adapted by Böck and Schedl (2011) to create the "RNNBeatProcessor," so named in the open-source Python madmom library (Böck et al. 2016). The network architecture is based on Bi-directional Long Short-Term Memory recurrent neural networks. This network architecture was chosen for the following reasons:

- In the most basic form of feed-forward neural network, the relationship of inputs to outputs is strictly causal, i.e., inputs at the current time compute outputs at the current time.
- To introduce using inputs from the past to influence current outputs, which is important in beat tracking, cycles are created in the network, leading to recurrent neural networks. However, these suffer from the vanishing gradient problem, which causes inputs to decay or blow up exponentially over time.
- The Long Short-Term Memory (LSTM) network solves the problem of the vanishing gradient by introducing memory gates within the recurrent unit.
- Finally, to consider inputs from the future in the output, a Bi-directional LSTM (BLSTM) network introduces a second hidden layer which introduces the inputs to the network in a reverse temporal order.

As past, present, and future inputs are all useful in a beat tracking algorithm, the BLSTM network from Eyben et al. (2010) was considered to be an appropriate choice. An additional modification was to introduce a peak picking stage for beat selection.

In their next paper, Böck, Krebs, and Widmer (2014) adapted the RNNBeatProcessor to introduce multiple RNN models, each of which was trained on a different musical style, and added a dynamic Bayesian network in the front-end for the beat estimation. The choice of dynamic Bayesian network was based on prior work by Whiteley, Cemgil, and Godsill (2006), which achieved robust results in joint tempo and beat estimation.

In yet another follow-up paper, Böck, Krebs, and Widmer (2015) modified the probabilistic tempo and beat sequence model of Whiteley, Cemgil, and Godsill 2006 to make it more computationally efficient (both in CPU cycles and memory) while maintaining the high accuracy of their solution in Böck, Krebs, and Widmer 2014. The original joint tempo and beat model of Whiteley, Cemgil, and Godsill (2006) is referred to as the *tempo-bar model*, and it has two main deficiencies which were adjusted:

1. The tempo-bar model assumes that human tempo discrimination is consistent across all tempi and use linearly-spaced tempi, while Böck, Krebs, and Widmer (2015) describe that in fact humans have finer tempo resolution at lower tempos. To achieve a tempo spacing consistent with the just-noticeable-difference (JND) limits of human tempo resolution with the linearly-spaced model requires a huge number of tempi, while using a nonlinear spacing of more tempi in the low tempo range and fewer tempi in the high tempo range significantly reduces the total number of possible tempi and matches human perception better.
2. The tempo-bar model allows tempo transitions at any possible time, while Böck, Krebs, and Widmer (2015) limit their more efficient model to only allow tempo transitions on beat locations, which is consistent with tempo transitions in music.

The final algorithm for state-of-the-art beat tracking is therefore presented by Böck, Krebs, and Widmer 2015, incorporating the algorithms from Böck and Schedl 2011 and Böck, Krebs, and Widmer 2014, and achieving the best results in the last **M**usic **I**nformation **R**etrieval **E**valuation e**X**change audio beat tracking challenge.[1]

1. https://www.music-ir.org/mirex/wiki/2019:MIREX2019_Results

Bibliography

Bello, Juan, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark Sandler. 2005. "A Tutorial on Onset Detection in Music Signals." *IEE Transactions on Speech and Audio Processing* 13:1035–1047. https://doi.org/10.1109/TSA.2005.851998.

Böck, Sebastian, Filip Korzeniowski, Jan Schlüter, Florian Krebs, and Gerhard Widmer. 2016. "madmom: a new Python Audio and Music Signal Processing Library." In *Proceedings of the 24th ACM International Conference on Multimedia,* 1174–1178. Amsterdam, The Netherlands. https://doi.org/10.1145/2964284.2973795. https://github.com/CPJKU/madmom.

Böck, Sebastian, Florian Krebs, and Gerhard Widmer. 2014. "A Multi-model Approach to Beat Tracking Considering Heterogeneous Music Styles." In *Conference: 15th International Society for Music Information Retrieval Conference (ISMIR).*

———. 2015. "Accurate Tempo Estimation Based on Recurrent Neural Networks and Resonating Comb Filters." In *Proceedings of the 16th International Society for Music Information Retrieval Conference,* 625–631. Málaga, Spain: ISMIR. https://doi.org/10.5281/zenodo.1416026.

Böck, Sebastian, and Markus Schedl. 2011. "Enhanced Beat Tracking with Context-Aware Neural Networks." In *Proceedings of the 14th International Conference on Digital Audio Effects, DAFx 2011.*

Ellis, Daniel. 2007. "Beat Tracking by Dynamic Programming." *Journal of New Music Research* 36:51–60. https://doi.org/10.1080/09298210701653344.

Eyben, Florian, Sebastian Böck, Björn W. Schuller, and Alex Graves. 2010. "Universal Onset Detection with Bidirectional Long Short-Term Memory Neural Networks." In *Proceedings of the 11th International Society for Music Information Retrieval Conference,* 589–594. Utrecht, Netherlands: ISMIR. https://doi.org/10.5281/zenodo.1417131.

Goto, Masataka. 2002. "An Audio-based Real-time Beat Tracking System for Music With or Without Drum-sounds." *Journal of New Music Research* 30. https://doi.org/10.1076/jnmr.30.2.159.7114.

McFee, Brian, and Daniel Ellis. 2014. "Better beat tracking through robust onset aggregation," 2154–2158. ISBN: 978-1-4799-2893-4. https://doi.org/10.1109/ICASSP.2014.6853980.

Stark, Adam, Matthew Davies, and Mark Plumbley. 2009. "Real-time beat-synchronous analysis of musical audio." In *Proceedings of the 12th International Conference on Digital Audio Effects (DAFx-09), Como, Italy.*

Whiteley, Nick, Ali Taylan Cemgil, and Simon J. Godsill. 2006. "Bayesian Modelling of Temporal Structure in Musical Audio." In *Proceedings of the 7th International Conference on Music Information Retrieval,* 29–34. Victoria, Canada: ISMIR. https://doi.org/10.5281/zenodo.1415138.