# EDM tempo estimation submission for MIREX 2018


**First author**
Affiliation1
author1@ismir.edu

**Second author**
**Retain these fake authors in submission to preserve the formatting**

**Third author**
Affiliation3
author3@ismir.edu



## ABSTRACT

In this study, we propose a high-accuracy method for tempo estimation for Electronic Dance Music. The algorithm is based on peak-picking of autocorrelation on a novelty curve, and exploiting temporal elements of this popular music genre. Preliminary tests on two self-annotated datasets show that the model achieved 100% accuracy with minimum error tolerance, while ignoring octave differences.


## 1. INTRODUCTION

Tempo estimation has been studied extensively for various purposes, yet past studies generally allowed large error tolerance levels [4], which could fall short when more accurate tempo is required. In Electronic Dance Music (EDM), Disc Jockeys (DJs) often mix tracks by accurately matching their tempo, before connecting and overlapping segments for a live performance. As a result, most tracks are made to be mixed with others easily, and their music structure often shares common properties such as constant tempo, and 4/4 meter throughout a track [2]. In this study, we propose a tempo estimation algorithm for EDM that takes advantage of these properties.

## 2. METHODOLOGY

We assume all music tracks follow a constant tempo and metre from start to finish, therefore only one tempo is estimated for each track. We follow the approach proposed by Davies and Plumbley [3] in four steps:

a. Compute the onset detection function of each audio frame at a hop size of 512 samples, based on Mel-spectrum with a maximum frequency of 400 Hz in 8 bins. Instead of the complex spectral difference onset detection function, we apply the spectral flux function introduced by Böck and Widmer [1]

b. Compute the adaptive mean of the onset strength with a window size of 16 frames, and set all frames with onset strength below the mean to zero. This creates an onset novelty function

c. Compute the frame-level autocorrelation of the novelty function at a range of hypothetical tempi, averaged by the number of repetitions for each tempo, and we get a collection of autocorrelation value, shown in Figure 1

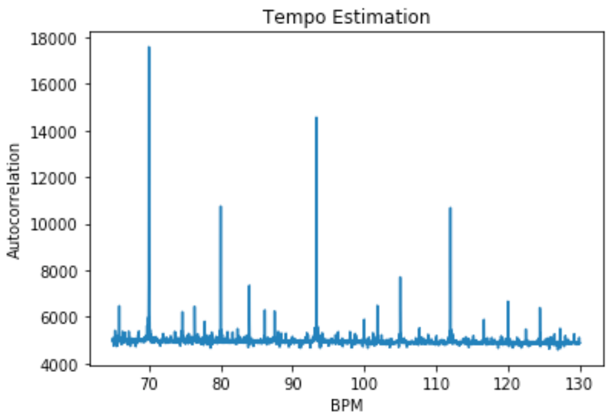


**Figure 1**. Autocorrelation at hypothetical tempo

d. The tempi with the highest peaks in the collection are chosen as the output candidates

To improve accuracy, we restrict the hypothetical tempi to be within one tempo octave (thus ignoring so-called octave differences, which allows for errors that correspond to a factor of 2, 3, 1/2, or 1/3). In addition, in some tracks the tempo at 3/2 of the annotated tempo has a higher autocorrelation value than the annotated tempo. To correct this, we pick the second highest peak as the detected tempo if it is 2/3 of the tempo with the highest peak, and its autocorrelation value is more than 90% of the highest peak.

## 3. EVALUATION

Two datasets are employed for the evaluation of this study. One of the authors has an in-house dataset of 214 popular EDM tracks from multiple live performances. In this dataset, there are 115 Tropical House, 50 Trap, and 49 Dubstep tracks. The second dataset is collected by Rocha et al. [6], which is comprised of 35 popular EDM songs of various EDM genres. For the second dataset, we found the audio files of 32 annotated songs online, and determined that they match the original songs used in the dataset. We then used the Serato DJ[1] software to annotate the tempo of all tracks manually.

For consistency, we set the range of tempo to 65-130 beats per minute (BPM) for both annotation and model output, with BPM increment of 0.01. In the in-house dataset, there are a handful of tracks in which the tempo

[1] Serato DJ: https://serato.com/dj

changes mid-way, and in this case the initial tempo is annotated. Test results show that the Accuracy2 of the tempi reported (allowing octave differences) for both datasets are 100% with a tolerance of 0.1 BPM.

## 4. CONCLUSION

Through novelty-based methods and exploiting the music structure of EDM, we have developed a process for tempo estimation that yields very high accuracy with small tolerance level. Although the method is developed mainly for EDM, it can be readily applied to other genres of music with a constant tempo.

## 5. MIREX SUBMISSION

For the purpose of MIREX submission, we include 3 outputs from the model, a slower tempo, a faster tempo, and a salience factor of the two tempi. Considering "preferred" tempo zone reported by Moelants and McKinney [5], we make octave adjustments to the estimated tempo, and report the slower tempo within the range of 55-110, and the faster tempo within 110-220. The salience factor is set to 0.5, which bears no useful information and is purely to meet MIREX submission standard.

## 6. REFERENCES


[1] Sebastian Böck and Gerhard Widmer. Maximum filter vibrato suppression for onset detection. In *Proc. of the 16th Int. Conf. on Dig. Audio Effects (DAFx-13)*, Maynooth, Ireland, 2013.

[2] Mark J. Butler. *Unlocking the groove: Rhythm, meter, and musical design in electronic dance music*. Bloomington, IN: Indiana Univ. Press, 2006, pp. 113–116, 197.

[3] Matthew E. P. Davies and Mark D. Plumbley. Context-dependent beat tracking of musical audio. *IEEE Tran. on Audio, Speech & Language Processing*, vol. 15, no. 3, pp. 1009–1020, 2007.

[4] Peter Knees, Ángel Faraldo, Perfecto Herrera, Richard Vogl, Sebastian Böck, Florian Hörschläger, and Mickael Le Goff. Two data sets for tempo estimation and key detection in electronic dance music annotated from user corrections. In *Proc. of the 16th Int. Society for Music Information Retrieval Conf. (ISMIR)*, Málaga, Spain, 2015.

[5] Dirk Moelants and Martin F. McKinney. Tempo perception and musical content: What makes a piece fast, slow or temporally ambiguous? In *Proc. of the 8th Int. Conf. Music Perception & Cognition*, Evanston, IL, 2004.

[6] Bruno Rocha, Niels Bogaards, and Aline Honingh. Segmentation and timbre similarity in electronic dance music. *In Proc. of the Int. Sound & Music Computing Conf.*, Stockholm, Sweden, 2013.