Beat Tracking and Tempo Estimation in Music Signals Using DSP Techniques

DSP Lab Course WS 2024/25

Authors:
Mozhgan Rezakhani (mozhgan.rezakhani@fau.de),
Quang Hoang Nguyen Vo (quang.nguyen.vo@fau.de)
Supervisor: Jeremy Bai

Introduction

In music theory, the concept of beat plays a vital role in defining the temporal and structural framework of a musical piece. When listening to music, we are naturally inclined to tap our feet along with the beat without much difficulty.[1]. However, when tempo changes (tempo rubato) or complex rhythms are involved in the music, it becomes challenging to accurately track the beat. Therefore, beat tracking and tempo estimation are fundamental tasks in music information retrieval (MIR) that aim to automatically detect the underlying rhythmic structure of music signals. These techniques are crucial for various applications such as automatic music analysis, beat-driven effects in music production, and synchronization in multimedia applications[2]. The goal of this project is to implement a system that can detect the rhythmic structure of music signals, estimate the predominant tempo, and evaluate its performance across various music genres.

Algorithms

This project incorporates various techniques for beat tracking and tempo estimation, with a primary focus on onset detection and tempo tracking, as well as advanced methods to handle tempo variations.

Short-Time Fourier Transform (STFT)

Audio signals can consist of a mixtures of multitude of different sound components. Hence, Fourier Transform is used to decompose the audio signal into its constituent frequencies. One variaton that was utilized in this project is the Short-Time Fourier

Transform (STFT), which is a Fourier Transform that is applied to small segments of the signal. This allows us to analyze the frequency content of the signal over time, which is essential for detecting onsets and tracking tempo variations.4

We consider a window function $w(t)$ of length N and a hop size of H. The STFT of the signal $x(t)$ is defined as:

$$X(n,m) = \sum_{k=0}^{N-1} x(nH+k)w(k)e^{-j2\pi mk/N} \tag{1}$$

where $X(m,n)$ is the STFT of the signal at frame n and frequency bin m. The STFT provides a time-frequency representation of the signal, which is essential for analyzing the spectral content of the audio signal over time.

Onset Detection

Onset detection was implemented using **spectral flux**. Spectral flux is a measure of the rate of change in spectral energy, which is a good indicator of musical events (e.g., note or beat onset). Mathematically, spectral flux is defined as:

$$\text{Flux}(t) = \sum_f \Big| |X(f,t)| \; - \; |X(f,t-1)| \Big| \tag{2}$$

where $X(f,t)$ represents the spectral content of the signal at time t and fr“equency f. The absolute difference between consecutive time frames provides a measure of the change in spectral energy, and peaks in this flux correspond to onsets *(Müller et al., 2015).*

Half-Wave Rectification

We can further enhance the onset detection by applying **half-wave rectification** to the spectral flux, as we only focus on the increase of spectral energy. The half-wave rectified spectral flux is defined as:

$$|\text{Flux}(t)|_{\geq 0} = \max\Big(0, \text{Flux}(t)\Big) \tag{3}$$

This operation ensures that only positive changes in spectral energy are considered, which helps in detecting the onset of musical events more accurately.

Tempo Estimation using Tempogram representation

Tempo estimation was performed by detecting the periodicity of detected onsets using **autocorrelation** of inter-onset intervals (IOIs). The autocorrelation function is defined as:

$$r(\tau) = \sum_k p(k+\tau)\, p(k) \tag{4}$$

where $p(k)$ is the onset signal and τ is the time lag. The peak in the autocorrelation function corresponds to the beat period, which can then be converted into beats per minute (BPM):

$$\text{BPM} = \frac{60}{\text{Beat Period (seconds)}} \tag{5}$$

Another method to estimate tempo is using Fourier-based method. Where the tempo is estimated by finding the peak in the **tempogram** representation of the onset signal. The tempogram is computed by applying a short-time Fourier transform (STFT) to the onset envelope and measuring periodicity at different time scales. Given the STFT from equation 1, we can define the Fourier tempogram as:

$$\mathcal{F}(\tau, n) = |X(\tau/60, n)|^2 \tag{6}$$

Where τ is the tempo in beats per minute (BPM) and n is the frame index.

Global Tempo Estimation

Given a tempogram, we can estimate the global tempo by averaging the tempo values across time frames, obtaining a function$T_{avg}(\tau)$ that represents the tempo at each time frame[1]. The average tempo can be computed as:

$$T_{avg} = \frac{1}{T}\sum_{t=1}^{T} T(n, \tau) \tag{7}$$

The maximum of T_{avg} corresponds to the global tempo of the track.

$$\text{Tempo}_{\text{global}} = \arg\max_{\tau} T_{avg}(\tau) \tag{8}$$

Predominant Local Pulse (PLP) Estimation

PLP estimation is used to dynamically track the predominant beat position over time. It is achieved by accumulating **sinusoidal kernels** $\kappa_n(m)$ at each time frame n, where the frequency of each sinusoidal kernel corresponds to the detected tempo at that time. The PLP curve can be defined as:

$$\Gamma(t) = \Big|\sum_{i=1}^{N} \kappa_n\Big(m\Big)\Big|_{\geq 0} \tag{9}$$

where f_i and ϕ_i are the frequency and phase of each kernel at time t. The accumulated kernels form the PLP curve, which provides a robust representation of the beat throughout the track [3].

Methodology and Result

The system was implemented in Python using the **Librosa** library for audio processing and analysis. The system consists of the following steps:

1. **Preprocessing**: The audio signal is loaded and preprocessed using a short-time Fourier transform (STFT) to obtain the spectral content of the signal.
2. **Onset Detection**: The spectral flux of the signal is computed to detect onsets, which are further enhanced using half-wave rectification.
3. **Tempo Estimation**: The inter-onset intervals (IOIs) are computed from the detected onsets, and the tempo is estimated using both autocorrelation and Fourier-based methods.
4. **Global Tempo Estimation**: The tempo is estimated across time frames to obtain the global tempo of the track.
5. **PLP Estimation**: The PLP curve is computed using sinusoidal kernels at each time frame to track the predominant beat position over time.

Three audio tracks are used for evaluation spanning different genres (pop, electronic, classical) to assess the system's performance across various music styles. For illustration, we present the results for audio track 3 in this section. Figure 1 shows the novelty curve of the audio. The tempogram representation using Fourier-based method and autocorrelation of the track are shown in Figure 2 respectively, with detected tempos annotated. Table 1 shows the tempo estimation results using both methods, along with the global tempo estimation and PLP curve.

Challenges and Solutions

Several challenges were encountered during the implementation:

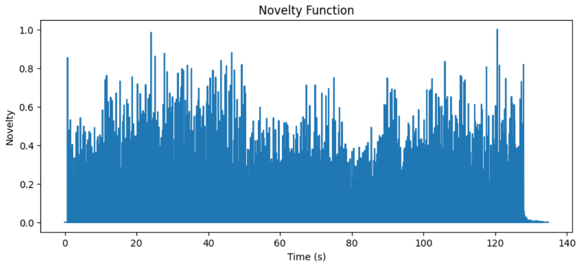
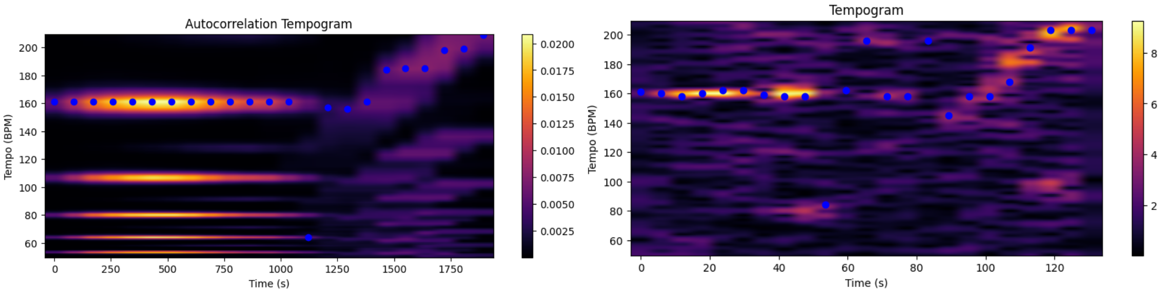


Figure 1: Novelty curve of the audio track



(a) Autocorrelation

(b) Fourier-based

Figure 2: Tempogram representation of the audio track

Track	Tempo (BPM) (ACF)	Tempo range (BPM)
1 (pop)	140	[100, 180]
2 (electronic)	100	[70, 180]
3 (classical)	161	[50, 210]

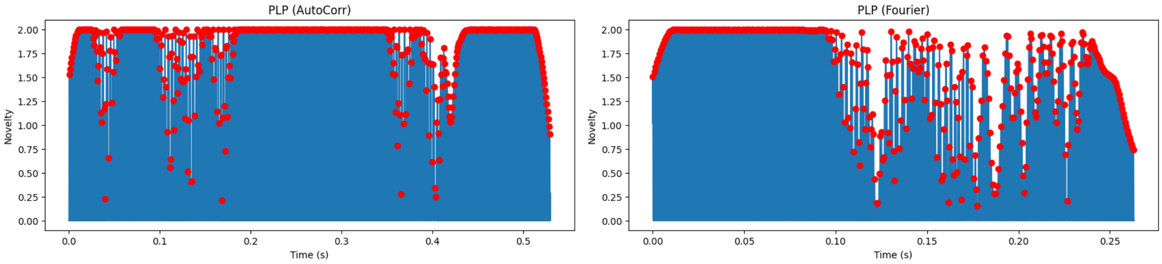
Table 1: Tempo estimation results for the audio track

Difficulties Encountered

1. **Detecting Onsets in Non-Percussive Music**: Tracks with subtle or blurred onsets (e.g., classical or jazz music) posed difficulties for onset detection using spectral flux.

2. **Tempo Variations**: Handling tempo changes (such as *accelerando* or *ritardando*) was challenging, particularly for tracks with drastic tempo fluctuations.

3. **PLP Estimation**: Accurately estimating the *Predominant Local Pulse* in highly dynamic and expressive music was challenging, as small variations in tempo or rhythm can lead to instability in the PLP curve.



(a) Autocorrelation

(b) Fourier-based

Figure 3: PLP curve of the audio track

Possible Improvements

1. **Enhanced Onset Detection**: Future improvements could focus on using more advanced onset detection algorithms that better handle non-percussive music, such as those relying on **machine learning**.
2. **Adaptive Tempo Tracking**: Introducing more advanced models, such as **dynamic time warping (DTW)**, would help improve accuracy in tracking tempo changes over time.
3. **Noise Reduction in PLP Estimation**: Smoothing the PLP curve using more sophisticated filters, or employing **machine learning** to model the predominant pulse, could enhance accuracy in challenging tracks.

Bibliography

[1] M. Müller, *Fundamentals of Music Processing*, 01 2015.

[2] M. Alonso, B. David, and G. Richard, "Tempo and beat estimation of musical signals," in *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR)*, 2004.

[3] M. Müller *et al.*, "Beat tracking and tempo estimation of musical signals," *Journal of the Acoustical Society of America*, vol. 138, no. 4, pp. 2015–2024, 2015.