Digital Signal Processing

Beat Tracking And Tempo Estimation

Using Onset Time Period Estimation

Introduction

Automatic tempo estimation is a useful tool for musicians for the purposes of transcription and also for audio researchers as we can use the tempo of a piece of music to inform other types of analysis such as pitch detection or chord detection. Since musically significant events such as chord changes tend to occur on the beat, knowing when the beats occur should provide greater accuracy in detection algorithms.

Tempo

In musical terminology, **tempo** [ˈtɛmpo] ("time" in Italian; plural: *tempi* [ˈtɛmpi]) is the speed or pace of a given piece.

In classical music, tempo is usually indicated with an instruction at the start of a piece (often using conventional Italian terms). Tempo is usually measured in beats per minute (bpm). In modern classical compositions a "metronome mark" in beats per minute may supplement or replace the normal tempo marking, while in modern genres like electronic dance music, tempo will typically simply be stated in bpm.

Autocorrelation

Autocorrelation, also known as serial correlation, is the correlation of a signal with a delayed copy of itself as a function of delay. Informally, it is the similarity between observations as a function of the time lag between them. Autocorrelation can be used as a mathematical tool for finding repeating patterns, such as the presence of a periodic signal obscured by noise, or identifying the missing fundamental frequency in a signal implied by its harmonic frequencies. It is often used in signal processing for analyzing functions or series of values, such as time domain signals.

$$R_{yy}(l) = \sum_{n \in Z} y(n)\, \overline{y}(n-l).$$

Differentiation

A signal is differentiated to determine the rate at which it changes. That is, if $x(t)$ is the continuous-time signal, then its differentiation yields the output signal $y(t)$, given by

$$y(t) = \frac{\mathrm{d}}{\mathrm{d}t}\{x(t)\}.$$

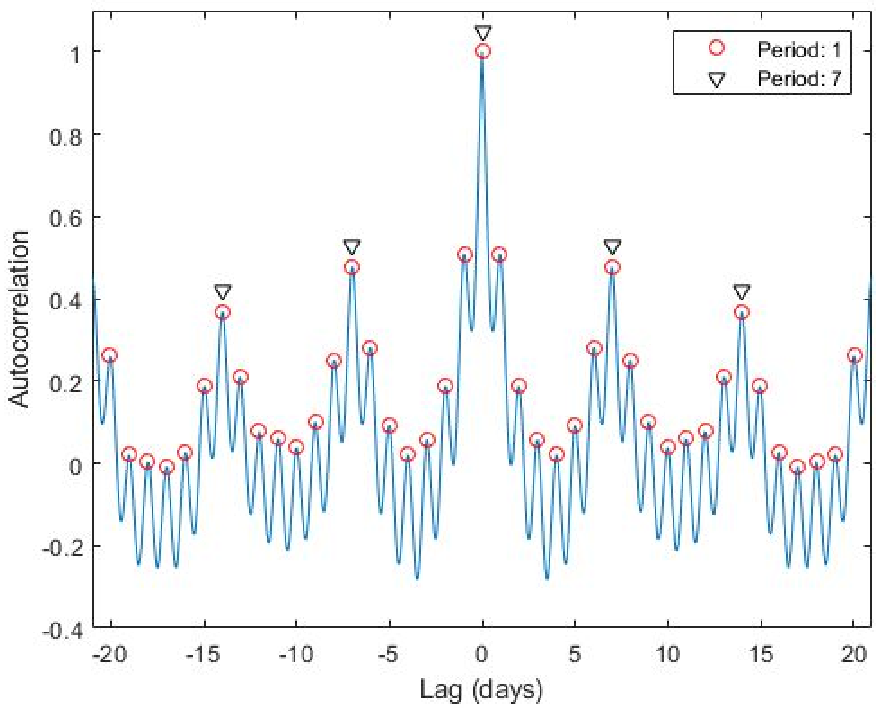
Onset Strength Signal

The first stage of processing is to convert the audio into a one-dimensional function of time at a lower sampling rate (but since this is an experimental system and we want max accuracy we have retained the original sampling rate of the .wav file) that reflects the strength of onsets (beats) at each time. The first-order difference along time in each frequency channel is half-wave rectified (to leave only onset information) then summed across frequency.

Time Period/Tempo Estimation

The onset strength for the entire signal is autocorrelated out to a maximum lag of 4 s (i.e. 4*44100 samples). The autocorrelation sequence of a periodic signal has the same cyclic characteristics as the signal itself. Thus, autocorrelation can help verify the presence of cycles and determine their durations.A graph is projected for the autocorrelation data and we have to manually adjust the minimum hieght of the peaks for peak detection for now in some cases.A value of 35 seems to work the best for all cases tested.Determine the short and long periods by finding the peak locations and determining the average time differences between them.

To find the long period, we restrict findpeaks to look for peaks separated by more than the short period and with a minimum height of some h.



For example the triangles represent the long peaks and the circles represent the short peaks in the auto correlated data above.

Results

4 different audio samples were tested to check the accuracy of the tempo estimation in accordance with the actual tempo of the song.

Dsp60.wav Metronome of 60hz

https://www.youtube.com/watch?v=NHwt-IOQ1nE

Output = 60 BPM

Dsp80.wav Metronome of 80hz

Output = 80 BPM

Dsp100.wav Metronome of 100hz

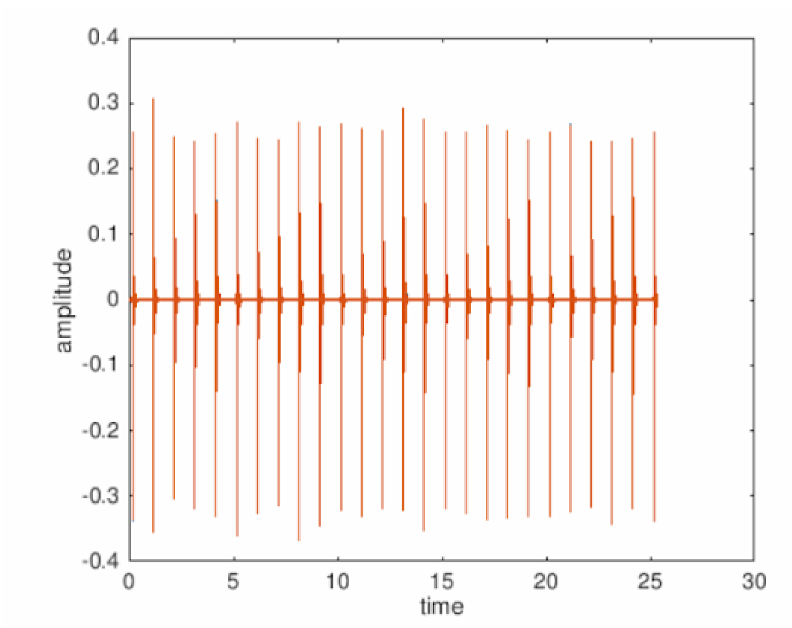
Output = 100 BPM

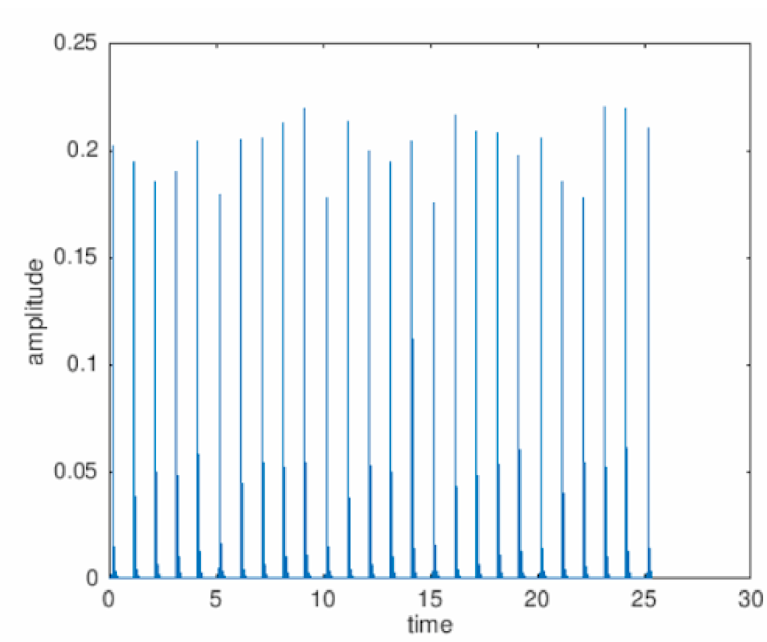
Dsp_test1.wav EDM song (Do I Wanna Know) of length 30secs at 84 BPM

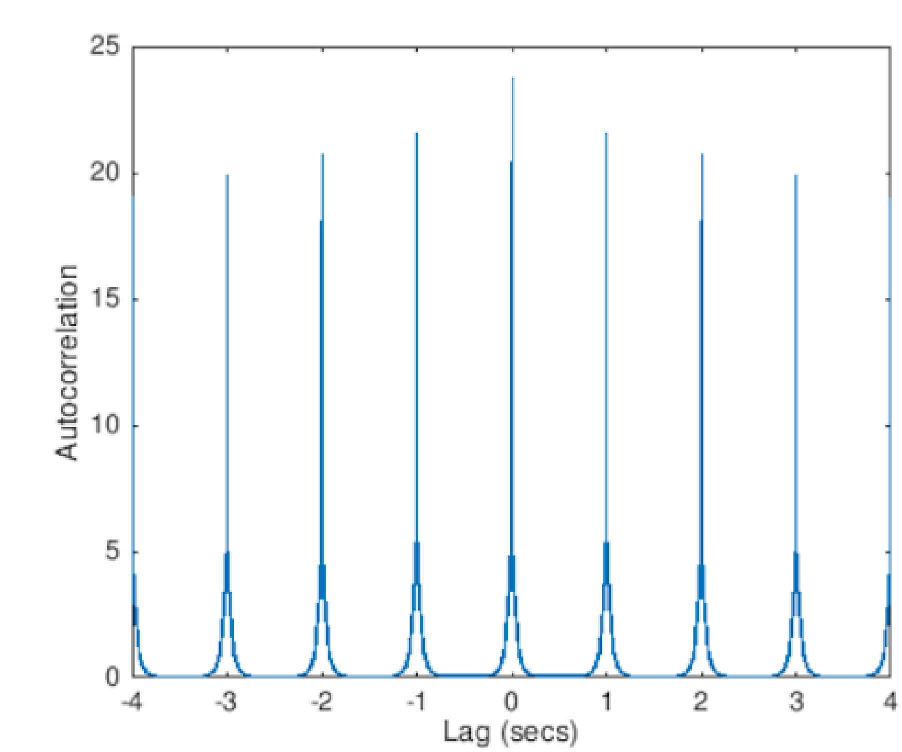
https://www.youtube.com/watch?v=bpOSxM0rNPM

Output = 85.145 BPM

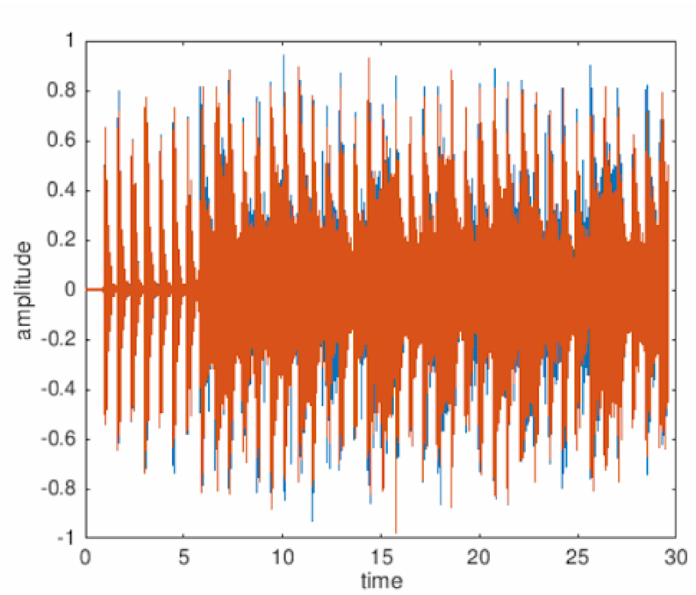
Dsp60 output graphs (audio signal,diffrect,autocorr)

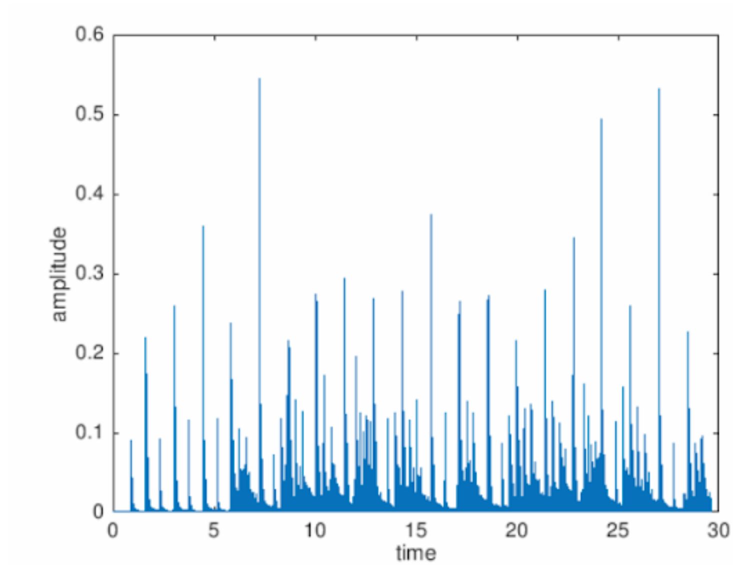


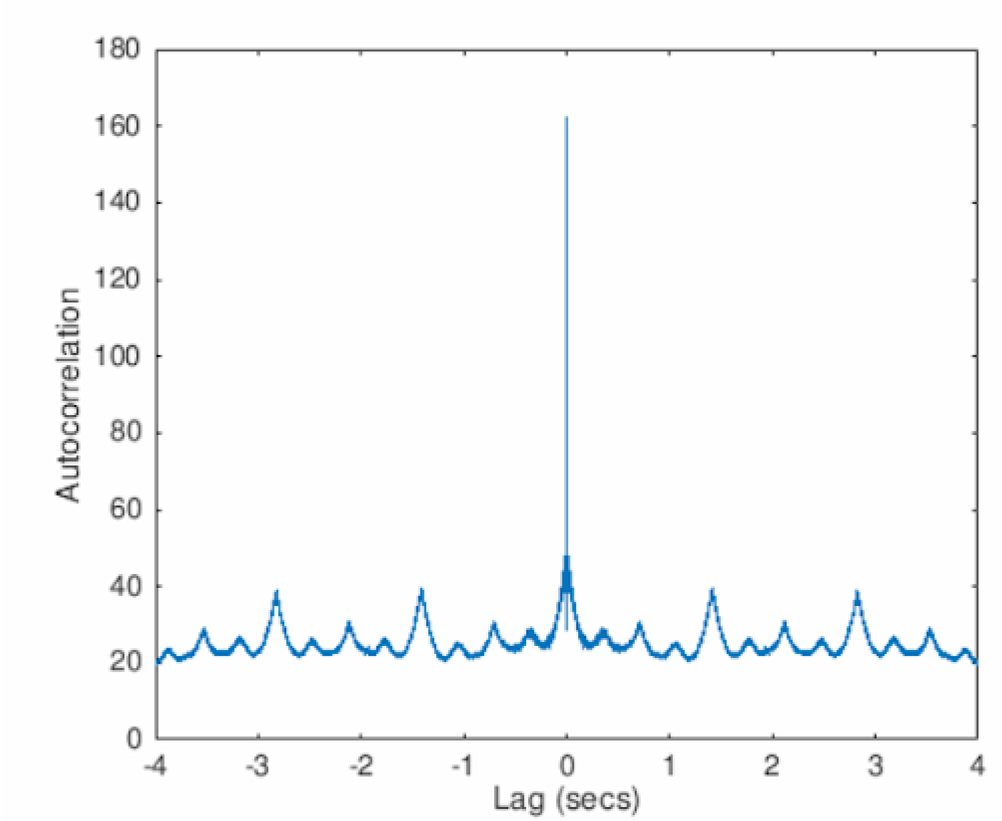




Dsp_test1 output graphs(audio signal,diffrect,autocorr)







Conclusion

Auto-correlation therefore is a cross-correlation of a signal with itself at a given lag or delay. By carrying out the autocorrelation of a signal we can identify at what lag the signal is most similar to itself. For most periodic music with a strong beat, the highest correlation will occur on the beat and we can therefore perform a quick calculation to convert the lag index (number of samples delayed) of the greatest correlation to a bpm value. This value is simply the reciprocal of the lag index divided by the sampling frequency with the result multiplied by 60 i.e

tempo = fs x 60 / lag_index_max

But these lag peaks are periodic hence we can get the time period of the short peaks then get the time periods of the long peaks.

Therefore

BPM = 1/Tp_longpeak

References

https://in.mathworks.com/help/signal/ug/find-periodicity-using-autocorrelation.html

https://labrosa.ee.columbia.edu/~dpwe/pubs/Ellis06-beattrack.pdf

http://signalprocessingsociety.org/sites/default/files/uploads/publications_resources/news/ICASSP_2017_SPCup.pdf