Review of "Automatic Extraction of Tempo and Beat from Expressive Performances"

Abhijit Bhattacharjee

IN THIS PAPER, author Simon Dixon proposes a new, robust method for offline beat tracking of expressively performed music. The task of tracking a beat in a piece of music is intuitive for most humans, but very difficult for machines to perform. This is because, traditionally, the problem has been challenging to define and even more so to implement as an algorithm or machine language. Dixon has addressed the issues in a new manner that provides a high rate of beat recognition, with information automatically derived from data and not specific to any musical style or genre.

Beat is thought of as a fundamental, low-level property of music. That is, it does not typically require specialized knowledge to understand a musical beat, although musical training can improve one's rate of rhythm perception considerably. Here, Dixon takes advantage of this property to further enhance computer rhythm perception through a multiple hypothesis, multiple agent search method.

First, a musical file must be chosen. This file can be a symbolic representation of musical data, such as MIDI, or it can be a digital audio waveform, such as in WAV or MP3 files. The processing method differs slightly with the choice of format. In either case, the next step is to determine the times when rhythmic events—abstractions of beat candidates such as peak amplitudes and chord changes—occur. In audio waveforms, the rhythmic events are detected by smoothing the amplitude envelope and searching for peaks in a localized window. The highest peaks in each window are labeled as rhythmic onsets. In symbolic data, onset times are available directly from MIDI information, but their relative importance is not so clear. Thus, onset times derived from MIDI data are chosen according to a weighting function that uses some high-level musical knowledge to determine their saliency as a possible rhythmic event.

After finding a timeline of rhythmic events, the next step is to calculate the different metrical levels of tempo that are possibly present in the music. This is *tempo induction*, where a clustering algorithm gathers and ranks different levels of time separation for every pair of events to create levels of metrical hierarchy for further analysis. After tempo induction, the last step is actual *beat tracking*, where the programmed clusters of beat intervals are analyzed by multiple automated agents that search and interpolate between candidate beats in the rhythmic events timeline. When each agent is finished creating a possible beat map, each path is ranked by a score. The tempo map with the highest score is taken as the calculated beat timeline.

Dixon compared the ability of his system to find beats with human-annotated musical samples. Taking these annotations as ground truth, Dixon then computed a percentage based on how many beats were successfully captured by the beat tracking algorithm. The final average number of beats successfully tracked came to 91.1%, an impressive number given the variety of genres and musical samples used in the experiment.

The drawbacks to the system are that it does not run in real time, which means that synchronization for live performances cannot be achieved through this method, and that rhythmic event calculation cannot yet be performed without ad hoc or genre-specific knowledge. These are areas which will require further investigation for the improvement of such a beat-tracking system.