

Jonathan Mooser

Review of

*An Audio-based Real-time Beat Tracking System for Music With Or Without Drum-sounds*

By Masataka Goto

2/3/2005

In *An Audio-based Real-time Beat Tracking System for Music With Or Without Drum-sounds*, Masataka Goto examines a novel method for detecting and predicting beats from real-time audio signals, a notoriously challenging problem. He contrasts his algorithm with previous attempts to solve similar problems and describes how his approach is an improvement. The details of the algorithm are laid out, including a theoretical justification for each component. Finally the author analyzes the results and briefly proposes a few potential applications.

The author's proposed beat tracking system has several key features. It works with real-world acoustic signals rather than a note-based format such as MIDI, which is significant because a MIDI representation is not readily available for most music. It is also not only finds beats but is able to discover higher level musical structures, grouping the beats into half notes and measures. Finally, an essential advantage is the system's ability to analyze signals in real time, making it useful for real time animation or improvisation. The author cites a wide variety of previously published beat tracking methods. Some of them contain a few of these features, but none have all of them.

The algorithm is intended analyze music by conceptually inverting the process of creating music. Goto describes music creation as a top-down process, with creating beat structure at the top level and creating individual sounds at the bottom level. The analysis, therefore, will look to first extract individual musical sounds, then logically group them into higher order structures.

In order to discover individual beats and beat structure, the algorithm tries to extract three pieces of information from an input signal: onset times, chord changes, and drum beats. Onset times are derived from a frequency spectrum. If the amplitude of a given band increases significantly, and is significantly larger than the amplitude of neighboring bands, an onset is declared. Those onset times are then labeled "provisional beats," meaning that they should be considered beat time candidates to be further examined. The next step is to look for chord changes. If there is a significant change in the dominant frequencies of the signal at a provisional beat, then the confidence in that beat is increased, and it is likely to represent a chord change. Finally the algorithm looks at drum sounds, which may or may not be present in the music. Using reasonable knowledge of the sound characteristics of snare drums and base drums, the presence or absence of drums is determined. If present, the sequence of drum sounds is compared to a library of likely patterns, and the results are used to further improve confidence about provisional beats.

All of this information is then further analyzed to group beats into half-notes and measures. A set of assumptions, such as the assertion that chord changes will most likely occur at the beginning of a measure, is used to label certain beats as the beginning of a new measure.

The reported results are impressive. With or without drums, the system correctly identified the quarter-notes over 85% of the time. Of those, it was able to group beats at the measure level 94% of the time without drums and 87% of the time with drums. The author does a particularly good job of explaining what went wrong when his algorithm was unsuccessful. In most cases, the was a problem with one of the fundamental assumptions, like the assumption that chord changes come at the beginning of a measure.

A few potential applications are cited, most interestingly the idea of a virtual dancer moving to a piece of music as it is played in real time. It is easy to imagine the value of being able to create an animation that plays in time with a live performance. There are also potential applications to automated composition by creating what the author calls an "intelligent drum machine."

Overall, the value and usefulness of the proposed method is clear. The results are strong enough that wide variety of applications could immediately benefit.