

Edward Stein  
February 2, 2005  
EE675 (ISE575)

An Audio-based Real-time Beat Tracking System for Music  
With or Without Drum-Sounds  
Author: Masataka Goto

Masataka Goto has released numerous papers on the topics of audio-data beat detection with drums and some without. In this paper he outlines a merger of the two. In his paper he describes a real-time system that can track the beat in music of either of these forms with the restriction that the piece is in 4/4 time and a relatively constant tempo between 61 and 185 quarter-notes per minute. In addition to the traditional beat detection format Goto goes farther to establish the beat marks at multiple metric levels. (Measures, Half-note and Quarter-note level) Unlike Dixon, Goto uses a significant amount of high level musical knowledge to generate algorithm rules and logic. The result of this is the fixed 4/4 time signature.

Goto's strategy breaks the beat-tracking process into three parts. The first is detecting beat tracking cues in the audio data. The second step is the interpretation or translation of cues into a beat structure. Finally, correction for interpretation ambiguities are done with existing templates for expected nature of the beat. To handle this, his process implements a bottom up subsystem to perform the first and half the second stage. The other half of the second stage and the removal of ambiguities are done with his model of the inverse problem in which known musical properties, rules and templates are applied to the given data.

It is obvious that drummed music is easier to find a beat in because of the dynamic impulses of the drum. One of the things Goto does is to determine if drums are used or not. To do this he checks for the frequency signature of the snare drum on the 2<sup>nd</sup> or 4<sup>th</sup> quarter note of the provisional beat. This is a strong detection but also has a limiting effect on the amount of play styles that are compatible. If the 2<sup>nd</sup> and 4<sup>th</sup> quarter note snare onset times do not yield a high enough autocorrelation the music is assumed to have no drums.

If drum patterns are detected, the system follows the low frequency bass drum and noise generation of the snare. As onset times are discovered it compares the detected pattern to 8 preset drum templates to infer the quarter-note level by assuming two conditions. One, the frequent inter-onset interval is likely to be the inter-beat interval. Two, Onset times tend to coincide with the beat times. This information is fed to main algorithm for consideration.

If the drums are not detected, there are still two other systems working. An onset detection and a chord change detector. The onset detection works much like you would expect, looking for strong changes in the level at a particular frequency from the noise floor. The chord change system however is much more sophisticated.

Goto realizes that chords or notes may be run together or materialize at off times, blurring the onset. He also feels it is unnecessary to classify chords, because classification could be wrong and lead to unstable results. Instead he allows a significant change in the frequency spectrum to indicate a chord change. The provisional beat is

used to infer whether the chord change was likely at the  $\frac{1}{2}$  or  $\frac{1}{8}$ <sup>th</sup> note level and subsequent information is again fed to the main algorithm.

Similar to the Dixon system agents are used to generate multiple predictions which are fed to a manager that decides which beat to transmit as the end result. The ending data is not in beats per minute but actually in the metric form of measure bars, half note and quarter-note levels of beat.

The results of this method are pretty good tracking most of the songs correctly in less than 45 seconds. A majority of the songs were tracked at the  $\frac{1}{2}$  note level in under 10s but took longer to decide which quarter was a half or a measure level beat. I feel that the system holds interesting potential as a concept but could be improved upon significantly. Items he has described such as the computer driven dancer has made it to the market under various forms or plug-ins in applications such as Winamp or Oozic, but these systems allow you to play any type of music at any beat and time signature. As a result either the library of rules he has constructed should be expanded or modified to support other time signatures and be more robust under larger playing fields. His system also seems to lack the ability to adjust quickly taking 10-20seconds to adjust for a change. So for example a beat drop in the middle of a song such as Love and Peace(U2) would give it just enough time to drop down the tempo before returning for the last part of the song. This would most likely leave his system out of phase for the remainder of the track. We should also keep in mind 10-20seconds could be 10-12 measures of a reasonably passed song.

In addition I think if the system has already detected chord changes (although we do not need to know what they are) we would benefit for more information to be presented about the nature of the audio within the measure. If you were trying to utilize this data for light show or fire-works etc a building or falling action in the energy of the chords and overall energy might be useful information to time significance of events.