

An Audio-based Real-time Beat Tracking System for Music With or Without Drum-sounds is to recognize a hierarchical beat structure from music with quarter-note, half-note, or measure levels in real time.

Previous work on such issue either lacks of power to process the recognition in real time, or not consider the higher-level beat structure above quarter-note level. The author claim his work is able to track beats both in real time and up to measure level. The main issues in beat tracking include (i) detecting beat-tracking cues from music, (ii) interpreting the cues to infer beat, (iii) and ambiguity solving. The author mentions that music knowledge-based inference might be helpful in tracking beat, which he adopts it in his work. The author also address the biggest problem in beat tracking is that the original beat structure does not being expressed explicitly from the performer's brain to audio signals (inverse problem).

To solve this problem, the author proposes a beat-tracking model with a component extracting musical elements from audio signal. Another is the inverse model, which includes detection on onset times, chord changes, and drum patterns. The author use FFT to extract these features. I am quite interested what if apply Wavelet Transform instead. Since the sample rate is exponent of 2, Wavelet is able to be performed. But FFT is fast enough to be processed in real time. I am not sure whether wavelet could be.

Onset-time vector is used to transform onset times of sever frequency ranges. I am not sure the criteria of split into these frequency ranges. The frequency spectrum is also sliced into strips at the provisional beat times, which is to obtain chord-change possibility. For the drum pattern, onset component is used when to detect onset time of bass drum and noise component is used when to detect onset time of snare drum.

All musical knowledge can be adopt in detecting beats from onset times, chord changes, and drum patterns also be mentioned in the paper. A noticeable point is that different musical knowledge will be applied for inference whether the audio signal contains drum sounds. Thus an algorithm detecting whether drum sounds in the music is needed. The author assumes that snare drum is typically played on the second and fourth quarter nodes in a measure. Seems this is not a complete solution, some future work can be done at this point. Also multiple-agent model is used to solve the problem interpretation ambiguity.

The author develops a system using the inverse model. Extensive experiments are performed on this system. Out of my expectation is that the cases without drum have better tracking accuracy and time compare to with drum. I think with the help of drum, the system should have better beat tracking accuracy, but only holds true in the cases of half-note level. The author does not address the reason for this.

The author also presents potential applications of this research, like virtual dancers, and stage light control. Overall the system is robust in experiments. Future work could be done to solve the problem with tempo changes and work with other musical genres.