David Li
ISE 575
2/3/2005

Review of Simon Dixon's paper "Automatic Extraction of Tempo and Beat From Expressive Performances"

Simon Dixon's paper "Automatic Extraction of Tempo and Beat From Expressive Performances" is about estimating the tempo from digital audio. It uses a multiple hypothesis search to find the beat that best fits the rhythm. He does a very thorough job with his algorithm to support syncopation and tempo fluctuations in music. His method is broken down into 3 stages: Finding the onsets, finding beat rate by Tempo Induction, and finding the phase of the beat by Beat Tracking.

Lerdahl and Jackendoff (1983) suggested a few rules: "beat coincide with note onsets; strong beats coincide with onsets of long notes; parallel groups receive parallel metrical structure; and the strongest beat occurs early in the group." This means that by following the onset we can tract the potential beats.

The onset detection method uses a high pass filter to remove slower beats, and smoothener to produce an amplitude envelope. Then extract the slope from the amplitude envelope, and selection the beat with the steepest slope. This applies for most musical instruments, but Dixon says beats would slow rise times would filtered out and not record the correct onset.

The Tempo Induction is calculated using the onsets to generate a list of tempo hypotheses. This method only needs the first few seconds of a music piece to do the calculation. The inter-onset interval is defined by two successive onset events, and then ranked using a relationship factor. The score increases by the relationship factor and the cluster size. Each ranking score is saved as a hypothesis beat rate.

The Beat Tracking system uses a prediction to view how accurate the rate is. An agent is created for each hypothesis. Using a small segment of music, the agents are used to track possibilities of where the beat should start. If some of the agents start and end at the same time, then the agent with the most prediction is kept.

To improve on this method, Dixon introduces the use of salience factors. Lee (1991) suggests that "every meter has a canonical accent pattern of strong and weak beats, and that listeners induce meter by matching the natural accent patterns occurring in the music to the canonical accent pattern of possibly rhythmic interpretations." The saliences help determine dynamics and distinguish certain pitches. There are linear and non-linear salience factors which can affect the beat tracking. The linear one can be used through the MIDI pitch and velocity and add to its salience function. The non-linear one uses the same MIDI pitch and velocity and is multiplied by the log of its velocity. The salience factors can then be weighted into the possible beat tracking.

The results from his experiments proved to be quite accurate. Dixon use classical music, jazz, rock, country, and hip-hop for his analysis. All the songs, except maybe Mozart's Piano Sonata in C managed achieve the proper beat rate due to the changing tempo. For the phase, the beat tracking the algorithm managed a decent level of success except for Bob Dylan's *On A Night Like This* and James Morrison's *Jitterbug Waltz*. They had syncopation, so the beat track would have difficult tracking the first few phases, but would recover eventually.

I was skeptical at first, because I felt it would have been easier to use the Discrete Wave Transform, but there are many different types and DWT will not always work. Dixon compares the human perception and cognition often with his algorithm. He analyzed a lot of previous methods and finds his algorithm matches the beat as well as a regular person can pick up the beat.

References

Dixon, Simon. "Automatic Extraction of Tempo and Beat from Expressive Performances." Austrian Research Institute for Artificial Intelligence. Austria. 2001.

Lee, C. "The Perception of Metrical Structure: Experimental Evidence and a model." Representing Musical Structures. Academic Press. 1991. pp 59-127.

Lerdal, F. and Jackendoff, R. "A Generative Theory of Tonal Music." MIT Press. 1983.