Isaac Schankler
2/1/2010
ISE575/EE675
Prof. Elaine Chew

Synopsis and Commentary on "Automatic Extraction of Tempo and Beat from Expressive Performances" by Simon Dixon

This article describes a computer program which purports to estimate the tempo and beat timings of a given piece of music. After some introductory comments and definitions of terms, it begins with a review of previous research in three categories—tempo induction, beat tracking, and determining musical salience. Tempo induction is defined as the process of estimating tempo from event onsets, while beat tracking simultaneously tests the various tempo hypotheses given by the method of tempo induction. Calculating musical salience determines how to weight the importance of various musical events, and is crucial for capturing rhythmic hierarchy. These techniques can be used on both audio data and symbolic performance data (e.g. MIDI).

The tempo induction algorithm works by tracking onsets of musical events and calculating the inter-onset intervals (IOIs) between events. (This includes both adjacent and non-adjacent pairs of onsets.) Similar intervals are then clustered together, and the clusters are then ranked according to how many elements are in a cluster, and how many related clusters there are (defined as a cluster with an interval that is a multiple or divisor of the first cluster's interval). This creates a ranked list of tempo hypotheses.

These hypotheses are then passed to the beat tracking algorithm. (Tempo induction approximates the time between beats but not the beat times themselves.) The beat tracking algorithm uses multiple agents testing out various tempo hypotheses simultaneously. If two agents are found to be duplicating one another, the one with the less accurate history is terminated. If a beat time falls in a particular agent's outer tolerance window, the agent will provisionally accept that beat time, but fork off a new agent to track the possibility that the beat time was a false positive.

For symbolic data, an event's musical salience may also affect an agent's particular score. Greater duration, density, amplitude and pitch are all correlated to greater salience of a rhythmic event. Two models for calculating salience were tested, one additive and one multiplicative.

The program was evaluated by testing it with nine songs, a sample size that seems far too small to determine the program's general efficacy. The songs are in a variety of styles (rock/pop, soul/funk, folk, classical, bossa nova and jazz), making the sample size even smaller for particular styles. Also, while most of the songs were in audio form, the only classical piece was tested as symbolic (MIDI) data, without any comments on how this might affect the results.

I tried out the BeatRoot software by the author, which I assume uses a similar algorithm, using "Au Suivant" by Jacques Brel, a singer-songwriter in the French cabaret tradition. "Au Suivant" is characterized by a persistent beat with strong accents on the downbeat, and long exaggerated pauses between certain phrases. I thought that the accents would make it easy to track, but BeatRoot was thrown off by the cadential pauses, and spent much of the time ticking on off-beats. I'm not yet convinced of BeatRoot's robustness.