Cheng Zhi Anna Huang
ISE (CS) 575
Week 5

Fabien Gouyon and Simon Dixon. (2005). "A review of automatic rhythm description systems." CMJ 29:1

This paper defines rhythm in its most generic sense, referring it to “all of the temporal aspects of a musical work”, which broadly includes the onset times, metrical structure, tempo and timing. The authors proposed “a unifying framework for automatic rhythm description systems, and review existing systems with respect to the functional units of the proposed framework.

Before we can start analyzing rhythm, we must first parse acoustic events into more abstract digital representation. However, there has been no consensus on the representation. This is due to the fact that rhythm takes different meanings for composers, performers and listeners, and rhythm transmission suffers a trade-off between the level of abstraction and the comprehensiveness of the representation. The choice of a suitable representation is thus dependent on the level of details of the various aspects of rhythm needed.

Once we have decided on a rhythmic representation, we can begin to develop an analytical framework. The author divides the analytical procedure linearly into the Highest Common Factorial functional units, including feature list creation (e.g. onset detection), followed by pulse induction (periodicity computation), and pulse tracking (phase detection). The analytical results of these sub-procedures are than used for more specialized tasks, such as time signature detection and quantised duration determination etc.

The feature list creation functional unit attempts to parse the input data into sequences of features, including onsets, durations, relative amplitudes, pitch, chord etc. Pulse tracking and pulse induction occur as complementary processes. The pulse-induction unit is a bottom-up model that looks for periodic behaviors assuming “the pulse period is stable over the data used for its computation”. Timing deviations (if any) are assumed to be short term. After evaluation, the unit chooses the pulse of the highest salience out of a number of possible periodicities. On the contrary, the pulse-tracking unit “handles short-term time deviations and attempts to determine changes in the pulse period and phase, without assuming that the tempo remains constant.” It is a top-down model. Pulse period and phase from the previous data are used as predictions propagated onto incoming data, and tracking is then “a process of reconciliation between these predictions and the observed data.”

This paper gives an informative survey of the methodologies and algorithms being developed, categorized under the functional units. However, it concludes that a systematic evaluation of competing models in rhythm description is not possible. Two of the reasons being the models are usually not described completely enough, and that there is no common database of test music labeled with the “ground truth”. The same phenomena exist in music cognition and perception models in general. These models are often a fusion of diverse fields, from music itself, to psychology, to signal processing and statistical mathematics. Disclosure of all backstage details might detour the goal of the papers. Due to this “necessity of ambiguity”, it is crucial to explain the choice of test music and its imposed limitations to avoid the illusion that a model can accomplish more than it actually has.