

Jonathan Mooser

Review of:

Analysis of the meter of acoustic musical signals

By A.P. Klapuri et al.

February 22, 2005

A.P. Klapuri's paper Analysis of the meter of acoustic musical signals present a novel method for detecting metrical structure of audio signals specifically intended to work across musical genres. The proposed approach divides the metrical hierarchy into three levels labeled tatum, tactus, and measure, with the success of the algorithm ultimately measured by how often it can correctly identify those three levels compared to human listeners. In unusually deep mathematical detail, the authors break down four stages of processing to transform an input signal in to an estimate of metrical structure. The results are compared with several previously published methods.

One of the most important features of the algorithm is its ability to handle raw audio data. Other similar systems, like the one proposed by Douglas Eck, assume that musical onsets have already been found. Because most music exists in the form of wav samples, this feature vastly expands the number of possible applications, and eliminates the need to perform and kind of pre-processing to generate suitable input. In particular, the authors describe using 16-bit digital samples at a rate of 44100 samples per second, the format typically used on a compact disk recording. It is also noteworthy that the system is designed to handle changing meter with a given piece, something that most other systems are not.

Tatum is defined as the lowest level of metric hierarchy, the shortest repeating pulse that one hears in a piece of music. Tactus is the next level up, and represents the rate at which a human listener would typically tap. At the highest level is a measure, meaning the grouping of several tactus pulses into a rhythmic phrase.

The process consists of four basic stages. First, a series of accent are derived from the input signal using a compressed DFT transform and finding the times at which the amplitude jumps for a given band. This expands on the work of E.D. Scheirer, who discovered that recognizing beats depends on breaking signals into frequency bands. Next, the accent signals are passed through a comb filter, which, similar to an autocorrelation filter, finds their intrinsic periodicities. Then a Bayesian probability model is used to find the most likely period of the tatum, tactus, and measure pulses. Finally the phases of the pattern, meaning the point in the pattern to be labeled the beginning, is estimated separately at each of those levels.

The level of mathematical detail presented in the paper is somewhat unusual, a style that has advantages and disadvantages. On the one hand, it makes the paper more difficult to follow. On the other hand, however, if one were to try to actually implement the proposed method, those details would surely be very useful. It is likely, in fact, that other authors' systems, in their real-world implementations, are just as complicated, but they omit mathematical details for the sake of readability.

The results applied to a number a music samples are impressive compared to Dixon and Scheirer, both of whom proposed similar methods. The results are best at the tactus level, which the author points out is also the easiest for humans to recognize. The improvement over previous methods is ultimately good enough that work definitely represents an advancement in the field of automatic meter analysis.