

# Machine Rhythm

David Rosenthal

International Media Research Foundation  
TOHMA Building B1, 2-14-1 Nishi-Waseda  
Shinjuku-ku, Tokyo, 169, JAPAN  
dfr@media.mit.edu

The video discusses *Machine Rhythm*, a program which emulates human rhythm perception. Given a musical performance represented as a MIDI stream, the program determines the program's rhythm — that is, it decides the meter of the performance, the rhythmic value of each note, and the location of barlines. The basic orientation of the video is to demonstrate applications of the program; more theoretical aspects are treated in (Rosenthal 1992).

Machine rhythm-finding is a difficult problem for several reasons. First, the timing of downbeats, which are heard by human listeners as a regular pulse, actually varies quite wildly; in certain cases the period can vary by a factor of two in adjacent beats. This can happen even when tempo variation is not being used as an expressive device. Second, not all downbeats correspond to actual musical events — in other words, one sometimes taps to a beat that doesn't correspond to a note in the music. Syncopations are particularly difficult instances of this, but the problem occurs even in unsyncopated music. Third, there is no straightforward method by which the location of downbeats can be extracted from a performance; downbeats are not reliably louder than other notes, for example. Finally, the rhythm of a piece of music is not a single pulse, but rather an interlocking hierarchy or pulses with periods of different sizes; hence the nomenclature of "measure," "half-note," "quarter-note," and so on.

Despite these apparent difficulties, human listeners have little trouble arriving at an unambiguous interpretation; in this sense rhythm finding is as well-defined a problem as, say, speech understanding. Humans apparently compensate for the problems mentioned in the last paragraph by integrating a variety of acoustic and musical cues, such as texture, melodic pattern, relative time between onsets, length of notes, and others. Humans also have the ability to fluently change their interpretations, either because of an initial interpretation is incorrect or because of a change in the music. Humans apparently track the various-size periods simultaneously, and use them to confirm each other.

In an effort to build a machine rhythm tracker with performance approaching that of a human, we integrated many of these methods into our system: The system integrates evidence from a variety of cues, some of which involve subtle analyses of the music from the MIDI data. The system achieves the flexibility to handle difficult situations by maintaining a hierarchy of hypotheses. The

system tracks several "levels," or different-sized periods simultaneously, using them as checks on each other.

The first section of the video uses an animation to explain the problem of rhythm parsing, and briefly outlines our approach.

The next section shows the program parsing various live performances, and demonstrates an application — automatic transcription. This section also demonstrates the program's ability to handle somewhat more challenging situations, such as varying tempo and triplets.

The final section of the video shows another kind application, which we call automatic synchronization. In this section we take a piece written for four hands and record each pianist's part separately. We then demonstrate that the two parts need to be synchronized in order to be played together. Furthermore, they cannot be synchronized by globally altering their tempos; local adjustment of the tempo is necessary. To do this, the program must make use of the rhythmic parsing that the *Machine Rhythm* program produces. When the two parts are synchronized in this way, the piece sounds as it should.

## References

Rosenthal, D. 1992. *Machine Rhythm: Computer Emulation of Human Rhythm Perception*. Ph.D. diss., The Media Laboratory, Massachusetts Institute of Technology.

## Acknowledgments

This video was produced while I was a doctoral student at the MIT Media Lab. I'd like to thank Stuart Cody and Greg Tucker, both of the Media Lab, for their help.