A Model for Musical Rhythm

Jeff Bilmes
<bilmes@amt.mit.edu>

The MIT Media Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139

ABSTRACT

At least four elements characterize musical rhythm: 1) metric content, a quantized attribute; 2) ametric phrases, those unrelated to any tactus; 3) tempo variation, change in the number of beats per second; and 4) event shifting, time deviations from a steady beat. This deconstruction provides a means to represent timing features from different musical styles. Both tempo variation and event shift information can operate at different levels of the metric hierarchy found in music. This representation will facilitate the analysis and production of musical style to the extent that it is rhythmically expressed.

1. Introduction

This paper introduces a model for representing the expressive timing in musical rhythm. The motivation for the model is to facilitate the accurate prediction and production of human-like rhythmic phrases by computer. In the model, rhythmic style is a conspicuous parameter; that is, analyses of rhythmic phrases from differing musical styles should produce different model parameters.

2. Rhythm

At least four elements characterize musical rhythm: metric content, ametric characteristics, tempo variation, and event shifting. Each of these elements can be modeled separately.

The metric content of musical rhythm is the perceptual relation of successive rhythmic events to an evenly-spaced time grid. The time grid determines the "beat" within which all musical events are heard. Our perception of rhythm, sometimes referred to as *subjective rhythmization*(Fraisse, 1983), is a psychological linking of sequential event stimuli. For the effect to occur, the event inter-onset times must be bounded in time, ranging from about 120 msec to about 2 sec. Traditional Western musical notation is one example of a model which represents the metric content. Other examples include the output of a music quantizer and representations used by computer-music sequencers.

Ametric models of rhythm represent those phrases that do not have an associated beat. A musician produces these phrases without using a constant beat as a reference for note placement. There are two musical styles where these phrases occur: 1) Certain music has no perceptible beat and rhythmic phrases are completely unanchored from any fixed time grid. Due to the nature of these phrases, no beat-based representation is suitable. 2) When there is a clearly defined beat, sometimes a performer will only coincide with that beat once every phrase (e.g., the performer will coincide with a beat every eight bars, but in the interim, the performer's notes occur independently). In this case, the model must account for the very low frequency beats defined by the occasional coincidence. However, we will not consider ametric rhythmic models in this paper since many types of music can be represented using the three other models.

Tempo variation models of rhythm have recently been given some attention. Models in this category are usually functions that map time durations to deformed time durations or that map a beat number (essentially a time position) to a beat duration. By appropriately varying the tempo of a musical piece

– that is, by varying the number of grid marks that pass by in a given amount of time, or, by varying the beat duration – we create what is perceived to contain more expression. Functions describing tempo variation have been discovered that correspond closely to real musical performances. Some examples are time maps (Jaffe, 1985), time deformations (Anderson & Kuivila, 1989), sentic curves (Clines, 1977), and force model constructs (Feldman, Epstein, & Richards, pres).

One important difference between Western classical music and ethnic music (such as African) or modern music (such as African-American jazz) is that expressive timing in the former can often be described using only tempo variation functions. To describe expressive timing in the latter (and to better describe timing in classical music) requires the development of new timing representations called *event shift models*. Event shift models of musical rhythm are, like tempo variation models, functions of a time position (or metric position); but, in addition, they are functions of instantaneous tempo. These functions, when given a metric position and an instantaneous tempo, provide a duration. That duration is used to shift, ahead or behind in time, a musical event occurring at that metric position. Thus, such a function can be used to model African and jazz (or swing) music where performers deviate in time from a relatively uniform beat, and can model (for example) piano music in the Romantic style where individual voices are shifted.

It is commonly believed that music with appreciable deviations from a stable beat (event shifting music), such as jazz, has an implicit fixed time grid. But how is a fixed time grid implicitly defined, especially in music where all performers deviate from the grid? In jazz, people are said to play "behind the beat," "right on the beat," or "in front (or on top) of the beat" where note events occur respectively later than, right on, or earlier than the grid marks. In cases where the majority of performers play on the beat, the time grid is explicitly defined – although the music might be considered "lifeless". However, in cases where the majority of performers play both behind and in front of the beat, how do we know where the fixed time grid lies? If all performers in an ensemble play behind the beat, why do we not perceive the time grid as being shifted forward in time, and thus perceive all performers as playing on the beat? Several reasons are possible: 1) Performers do not deviate by the same amount. In ensembles, certain instruments, whose role is to define the back-beat, typically play more on the beat than others instruments; for example, the bass in a jazz ensemble and a support drum in an African ensemble tend to play more on the beat. 2) The amount a performer plays behind or ahead of the beat is time-varying and can be represented as a function that maps a metric position and a tempo[1] to a time duration that is used to adjust a note event. For example, at a section beginning, a performer might play right on the beat and then deviate appreciably at the section middle. Event shift functions can easily model both of these situations, thus demonstrating the sufficiency of representing event shifting music with models that are based on a fixed time grid.

3. Hierarchical Rhythmic Deconstruction

Music, and its rhythm, can be hierarchically deconstructed based on the metric content. One simple example of a metric hierarchy can be obtained from the 32 bar jazz AABA form; the form repeats indefinitely. Each repetition is divided into four sections, each section into eight measures, each measure into four beats, and each beat into three pulses. There are many forms that can be disassembled in this way; the point is that each may be deconstructed according to its own implicit hierarchical structure.

The arrangement of the hierarchy may change over time. For example, each beat might for a time be divided into four, rather than three, pulses to produce a "straight eights" feel rather than a "swing" feel. Also, a jazz tune might alternatively move from a structured jazz form (e.g., AABA) to free form. During free form, the hierarchy still contains the measure level on down, but there is no need for higher levels. A time-varying hierarchical structure thus varies due to both changes in the musician and changes in the

[1] And perhaps even musical intent, emotional mood, etc.

piece; as will be discussed below, it provides a means to describe tempo variation and event shift functions that occur in rhythmic figures.

4. Event Shift and Tempo Variation Functions

It is possible to define both tempo variation and event shift functions on domains corresponding to levels in the metric hierarchy. A function defined on a domain for a given level operates on unit durations equal to the unit durations of that hierarchical level (e.g., measure, beat). Mathematically, it is sufficient to add together the functions of different levels (or, essentially, of different time resolutions) and produce one function that is the sum of functions with different frequencies. However, it is more stylistically lucid to have separate functions defined on domains provided by the hierarchical levels.

Tempo variation should be separated into functions defined for different levels in the hierarchy. Consequently, they will reflect both long and short-range tempo variation. For example, a gradual increase in tempo is better modeled by a tempo variation function higher in the hierarchy, whereas a local tempo fluctuation is better represented by a tempo variation function lower in the hierarchy.

A tempo variation function's domain is defined by a particular level in the hierarchy. For instance, a measure-level tempo variation function operates on units of a measure. In the case where intra-measure domain values arise (e.g., one and a half measures), output values are computed for non-integer inputs. A tempo variation function's range, or output, is a tempo-scaling factor. The function itself is used for 1) a time deformation-style transformation (Anderson & Kuivila, 1989) that takes a time duration as input and produces as output a deformed time duration. This deformed time duration corresponds to the integral of the tempo variation function over the interval defined by the input time duration; and 2) instantaneous tempo that, for a given metric position, provides a tempo value; it is this latter value that the event shift functions use.

Similarly, event shift functions should be separated into functions defined for different levels in the hierarchy, rather than grouped into one function. The output of an event shift function is a percentage of the time duration unit defined by the level corresponding to a beat. A note onset occurring at a particular metric position is thus shifted ahead or behind in time by a percentage of that unit. For example, given an event shift function operating at a particular level and whose value for the current metric position is one percent, any event occurring at that moment will be shifted forward in time by one percent of a beat duration. In general, the higher the hierarchical level on which the function operates, the more we control the overall feel of the music; the lower the level, the more we control the local time variations of note onsets. It is thus possible to produce what is perceived as "playing behind the beat" by having a small percentage of forward shift high in the hierarchy; it is also possible to produce phrase-dependent timing variation for each rhythmic figure by having larger percentages of forward and backward shift lower in the hierarchy.

A performer does not use the same time shift values when playing a piece at different tempos. Thus, event shift functions must also, at times, be functions of instantaneous tempo. An event shift function of two variables (Desain & Honing, 1991) sufficiently models this situation.

Most metrical rhythmic figures can be represented using these models. However, we need to gather timing data from real performances by quality musicians to discover the functions operating at each level in the metric hierarchy. Additionally, we require an algorithm that will factor out from a performance tempo variation and event shift information for each level in an inferred hierarchy; that is, given a hierarchy for a particular rhythmic phrase, the algorithm must produce tempo variation and event shift functions describing the phrase for each hierarchical level. However, these functions often might not be continuous since an event shift function is undefined during musical periods without note events. Time sequence learning algorithms, including connectionist approaches(Jordan, 1989) and statistical clustering analysis, look promising for learning and representing such functions.

5. Current Work/Conclusion

It is possible to disassemble musical rhythm into four separate components: a metric hierarchy, two forms of timing functions that operate in cooperation with the hierarchy, and an amorphous structure. We are currently adding constructs to specify the metric hierarchy into a C++ based music system(Anderson & Bilmes, 1991). We are also adding to the system primitives that specify tempo variation and event shift functions and primitives that situate the functions at domains specified by hierarchical levels. Once this work is complete, we can verify that the model accurately describes expressive timing by having trained musicians evaluate nondeterministically-computed music. We need, however, more psychophysical experiments to discover the functions that describe tempo variation and event shifting, and to learn how to use these functions for classification.

Ultimately, as noted by Richards (Richards, 1988, pages 307–308), discovering functions such as those described above might also provide insight into the more general issue of temporal perception. Musical rhythm is created to stimulate our time sensing processes, and it is often easiest to extract useful information about perception from that which most saliently contains a stimulus.

6. Acknowledgements

I thank the following people who contributed to this work, wittingly or not, through our conversations: David P. Anderson, Dan Ellis, Whitman Richards, David Rosenthal, Alan Ruttenburg, Kathryn Vaughn, Barry Vercoe, and David Wessel.

7. References

Anderson, D. P., & Bilmes, J. A. (1991). Concurrent real-time music in C++. In *USENIX C++ Conference Proceedings* Washington D.C.

Anderson, D. P., & Kuivila, R. (1989). Continuous abstractions for discrete event languages. *Computer Music Journal 13(3).*

Clines, M. (1977). *Sentics, The Touch of Emotion.* Doubleday Anchor, New York.

Desain, P., & Honing, H. (1991). Tempo curves considered harmful. In *ICMC Proceedings* Montreal, Canada.

Feldman, J., Epstein, D., & Richards, W. (in pres.). Force dynamics of tempo change in music. *Music Perception, forthcoming.*

Fraisse, P. (1983). Rhythm and tempo. In Deutch, D. (Ed.), *The Psychology of Music.* Academic Press.

Jaffe, D. (1985). Ensemble timing in computer music. *CMJ*, *9*(4), 38–48.

Jordan, M. (1989). Serial order: A parallel, distributed processing approach. In Elman, J. L. & Rumelhart D. E. (Ed.), *Advances in Connectionist Theory: Speech.* Hillsdale: Erlbaum.

Richards, W. (1988). *Natural Computation.* MIT Press.