Tracking Musical Beats in Real Time[1]

Paul E. Allen and Roger B. Dannenberg
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213, USA
email: Paul.Allen@cs.cmu.edu, Roger.Dannenberg@cs.cmu.edu

ABSTRACT: Identifying the temporal location of downbeats is a fundamental musical skill. Observing that previous attempts to automate this process are constrained to hold a single current notion of beat timing and placement, we find that they will fail to predict beats and not recover beyond the point at which the first mistake is made. We propose a new model that uses beam search to consider multiple interpretations of the performance. At any time, predictions of beat timing and placement are made according to the most credible of many interpretations under consideration.

Introduction. Identifying the temporal location of downbeats is a fundamental musical skill. Even musically untrained humans can tap their foot with the beat when they hear a musical performance. Humans generally perform this task with ease and precision even in the face of unusual rhythms, musical expressiveness, and imprecise performances. A fully general, automatic beat tracker would be of great value in many tasks such as human-computer improvisation, music transcription, music editing and synchronization, and (ethno-) musicological studies.

Beat tracking is differentiated from the score following of Dannenberg [1] and Vercoe [2] by the absence of a score. In addition, most beat trackers disregard pitch, limiting input to the timing of note onsets and an initial tempo.

Early Models of Beat Tracking. Much of the previous work on beat tracking has come about as a by-product of research directed at different music understanding problems such as transcription and the perception of rhythm, meter, and melody. Longuet-Higgins [3, 4, 5] describes what is probably the first attempt to follow the beat of a performance in the face of changing tempo. He uses a static tolerance limit on the variability of downbeats and adjusts perceived tempo according to whether a note onset is later or earlier than expected. This work is part of a computational psychological model of the perception of Western classical melody proposed by Longuet-Higgins.

Chafe, Mont-Reynaud, and Rush [6] use methods similar to Longuet-Higgins in a transcription oriented task. Desain and Honing [7] report on a method that uses relaxation techniques for quantizing intervals into musical note values, though that approach does not seem amenable to real-time applications. As part of a real-time jazz improvisation following task, Mont-Reynaud [8] describes a new method for real-time beat tracking that incorporates the concept of confidence and a history mechanism that uses a weighted average of previous perceived tempos to compute the current perceived tempo. The method relies on a small amount of state to represent the status of the beat tracker with a simple method to update the state when

---

[1]Published as: Allen and Dannenberg, ‘‘Tracking Musical Beats in Real Time,’’ in ***1990 International Computer Music Conference***, International Computer Music Association (September 1990), pp. 140-143.

new note data is received.

We chose to begin our experiments in beat tracking by implementing Mont-Reynaud's real-time method. We were dissatisfied with its performance; in particular, it was sluggish and not sensitive enough to tempo changes, causing errors. In an attempt to make this method more responsive to tempo changes we parameterized the history and confidence mechanism with the expectation of being able to find parameter values that work well for a given style of music or performance. The original history mechanism uses an exponential decay for the weighting of previous tempos. The decay rate has a direct affect on the responsiveness and stability of the beat tracker. At one extreme, 100% decay, the tracker will ignore all history and respond only with respect to the latest note data, allowing rapid tempo fluctuations, but making the tracker very unstable. At the other extreme, 0% decay, the tracker will consider each bit of history equally with the present state so that consideration of the latest note data is slight and response is slow. The decay rate is one parameter we studied.

The second parameterization applies to the confidence calculation. Confidence serves to indicate the certainty with which the beat tracker believes a given note onset occurred on a downbeat and thus, indirectly, the amount the tempo should change in response to the onset. The original confidence function is a sawtooth function with a maximum value of 1.0 at predicted downbeats and minimum value of 0.0 at points exactly halfway between two adjacent predicted downbeats; its slope is either 1.0 or -1.0. We took this original confidence calculation and raised it to an exponent parameter (thus with a parameter value of 1.0 we attain the same confidence calculation as the original method). This allows us to regulate the impact a note onset will have on the new tempo as a function of the distance of the onset from the predicted downbeat. Here we are searching for a compromise between tolerance for imprecise downbeat notes and rejection of correctly played non-downbeat notes.

Our hope was that we could find, by searching the parameter space, parameter values that achieve good performance. In practice, we found that while we were often able to find large ranges of parameter values that were effective for tracking the more straightforward performances, those same parameter value ranges gave rather haphazard results for performances with reasonable tempo changes and/or more complex musical structure.

A New Beat Tracking Method. We extend the musical concept of ''beat'' to include two aspects, period and phase. In a musical performance, we will say that the period of a beat is the time duration between two successive beats (the reciprocal of the tempo) and that the phase of a beat determines where a beat occurs with respect to performance time. A beat tracker must contain state information that, at a minimum, includes the current perceived beat phase and period.

One of the difficulties with the methods previously discussed is that the beat tracker is required to decide upon a single new state that represents its belief of the current beat period and phase. If it makes a mistake at some point it most likely will never be able to recover and will fail to predict beats beyond the point at which it made the mistake.

We have built a method that uses real-time beam search to allow the beat tracker to consider several possible states at once. The method uses a history mechanism similar to the previous method but no longer relies solely on a confidence measure to guide it. Instead, a credibility measure is used so that at any given time there is a set of active states that represent the most credible interpretations for the performance encountered so far. Each state has a credibility value determined by a heuristic evaluation function that measures the credibility of the particular interpretation that the state represents. The state that describes the most credible interpretation is used as the basis for making beat phase and period predictions until a new most credible state is

found. When a new note onset is encountered in the performance, each active state is expanded into several new states that represent a subset of the multiple possible interpretations that arise from the onset and the interpretation described by the active state (see Figure 1). Those interpretations that do not make musical sense are discarded immediately. The set of states that remain become the new active states for the next expansion.

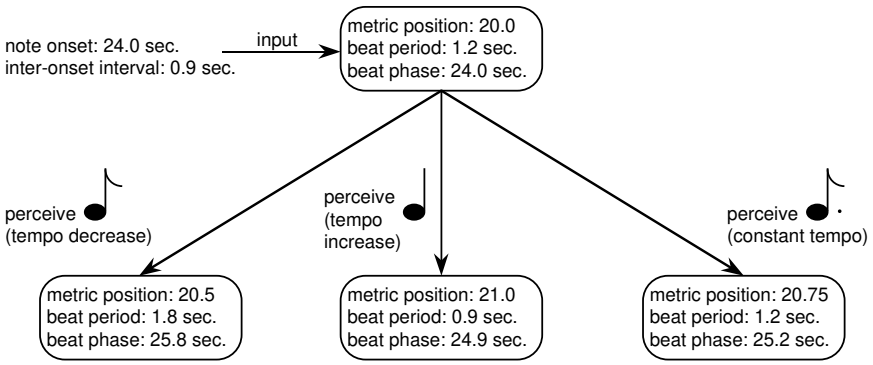

Figure 1: *Simplified example of state expansion.*

Because this is a real-time method and the number of interpretations is endless, we must liberally apply pruning techniques to decrease the magnitude of the search (see Figure 2). The first pruning method was already mentioned — we discard those states whose interpretations make no musical sense. A second technique involves the credibility of the state and the order in which expansions are made: since we want to perform expansions only until new input is ready, we expand states in order according to their credibility, those with highest credibility being expanded first. Thus during expansion, if we find that new input is ready to be processed, we can immediately stop expanding the active states knowing that we have already expanded the most credible ones. (Of course, we assume there is enough time between inputs to expand at least a few of the most credible states.) The third pruning method employed combines multiple states that represent similar interpretations into a single state representing a single interpretation. The fourth and final technique for pruning the search limits the branching factor by bounding the number of new interpretations that may arise from the expansion of a single active state. Some pruning already occurs at this point from the first pruning method, but since there are an infinite number of interpretations possible given a state and a new note onset, the number of expansions derived from a single state is limited by an arbitrary static cutoff. At this point, expansions from a single state are performed in an order such that the resulting interpretations are those that have the smallest change of tempo.

This concept of tempo coherence — that tempo does not change drasticly from note to note (and if it does change significantly it should always either increase or decrease for short periods) — is also used in the heuristic evaluation function. The heuristic credibility value of a state is a function of three parameters: (1) the magnitude of the tempo change between the state and its parent's state (favoring small changes over large changes), (2) the heuristic value of its parent state (favorable credibility values for the parent lead to favorable credibility values for the child), and (3) static penalties for representing musically possible but unlikely interpretations (favoring simple, common musical structures over complex structures).

Conclusion. We have been quite pleased with the initial results of our beam search method. A real-time prototype of our method is running and has given good results during initial testing. It follows tempo changes in simple melodies played by an amateur keyboard player with very little training. We have taken actual, constant tempo human performances and artificially

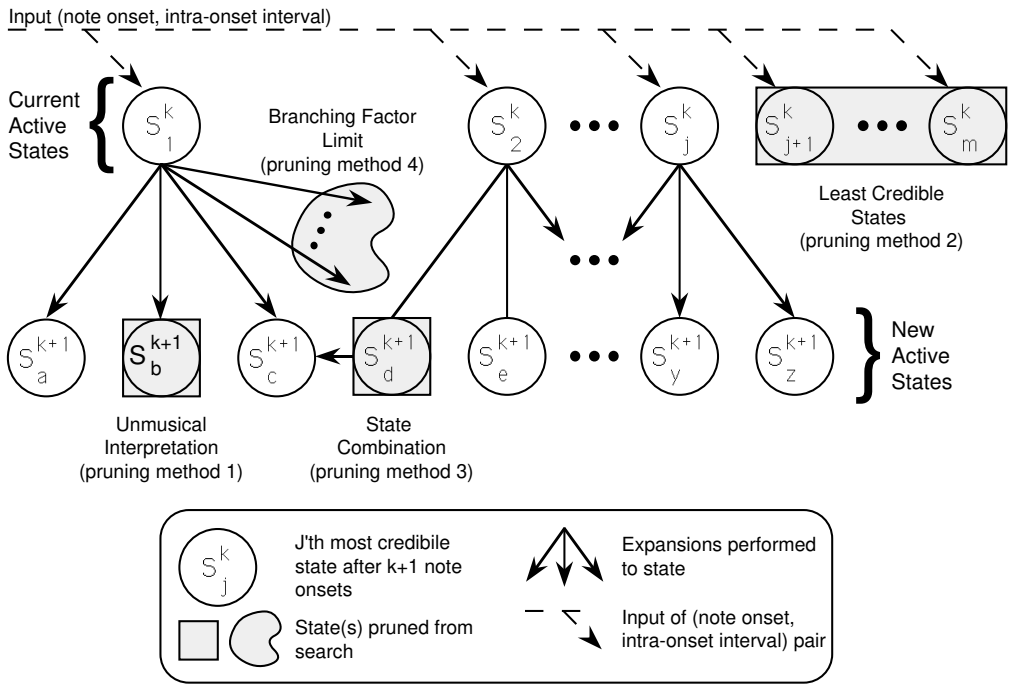

Figure 2: *One cycle of the beam search with example pruning applications.*

added tempo changes of such magnitude (over 50% change per beat) as to be completely unmusical, yet an earlier non-real-time prototype successfully follows such changes.

An added value of our method is that after successful beat tracking of a musical performance, a transcription of the performance can be recovered from the history of the final most credible state. This transcription is derived from the note values that produced each expansion along the path from the root of the search tree to the final most credible note.

References

1. Dannenberg, R. B. An On-Line Algorithm for Real-Time Accompaniment. Proceedings of the International Computer Music Conference 1984, International Computer Music Association, 1984, pp. 193-198.

2. Vercoe, B. and M. Puckette. Synthetic Rehearsal: Training the Synthetic Performer. Proceedings of the International Computer Music Conference 1985, International Computer Music Association, 1985, pp. 275-278.

3. Longuet-Higgins, H. C. "Perception of Melodies". *Nature 263* (1976), 646-653.

4. Longuet-Higgins, H. C. "The Perception of Music". *Interdisciplinary Science Reviews 3*, 2 (1978), 148-156.

5. Longuet-Higgins, H. C. and C. S. Lee. "The Perception of Musical Rhythms". *Perception 11* (1982), 115-128.

6. Chafe, Chris, Bernard Mont-Reynaud, and Loren Rush. "Toward an Intelligent Editor of Digital Audio: Recognition of Musical Constructs". *Computer Music Journal 6*, 1 (Spring 1982), 30-41.

7. Desain, Peter and Henkjan Honing. "Quantization of Musical Time: a connectionist approach". *Computer Music Journal 13*, 3 (1989), 56-66.

8. Dannenberg, R. B. and B. Mont-Reynaud. Following an Improvisation in Real Time. Proceedings of the 1987 International Computer Music Conference, International Computer Music Association, San Francisco, 1987, pp. 241-248.