Tempo Induction by Stream-Based Evaluation of Musical Events

Frank Seifert
Department of Computer Science
University of Technology
Chemnitz, 09107 Germany
fsei@cs.tu-chemnitz.de

Katharina Rasch
Department of Computer Science
University of Technology
Chemnitz, 09107 Germany

Michael Rentzsch
Department of Computer Science
University of Technology
Chemnitz, 09107 Germany
mren@cs.tu-chemnitz.de

Abstract

We present an approach for tempo induction that is based on a more perception-oriented analysis of inter-onset intervals. Therefore we utilize auditory grouping concepts and define some rules for their formation. Finally, we show preliminary results that confirm our aim of improving the quality of tempo induction by reducing the amount of perceptually irrelevant data.

Keywords: tempo induction, stream segregation.

1. Introduction

Most beat detection algorithms of symbolical music such as MIDI rely on either a stochastic evaluation of inter-onset intervals (IOIs), e.g. [1], or oscillatory models, e.g. [2]. To improve both approaches sometimes several methods like beam search [3] and weighting of IOIs are used in order to select more consistent beat hypotheses. Weighting of IOIs [4] is based on the assumption that listeners place long notes on strong beats. However, although we also suggest a more perception-oriented IOI-analysis we would like to do this in a more generic way by evaluating only IOIs. Thus we are able to handle both life-performed symbolical music without restrictions (e.g. staccato) and prepare the foundations for an audio analysis.

Foremost, let us illustrate drawbacks of a pure IOI-analysis. Although time between onsets is essential for estimating beat, not every possible IOI contributes to the generation of tempo-hypotheses to the same degree:

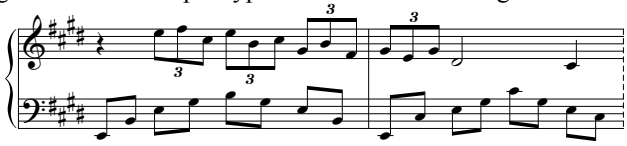

Figure 1: Triplets versus eights

Figure 1 shows a fragment of Debussy's Arabesque in E-Major, presenting parallel eighths and triplets. The resulting distribution of IOIs within a real performance is given in Figure 2. For simplification only distances that are shorter than a quarter note are presented. Quarter notes are the supposed beat, thus we denote them by a distance value of 1, triplets by 1/3 and eights by 1/2.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.
© 2006 University of Victoria

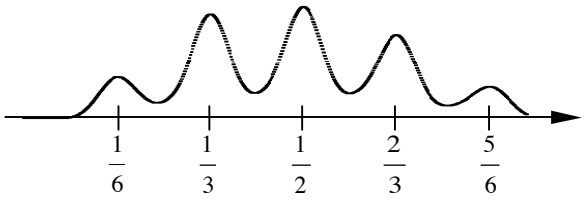

Figure 2. Inter-Onset Intervals

Which intervals contribute mostly to the perceived beat? Using the maximum does not seem very satisfying due to the existence of similar dominant neighbors. Additionally, more complex rhythms, tuplets, parallelism and performance deviations can cause even further complications.

2. A perceptual approach

Obviously not each interval influences the perceived beat to the same degree. Intuitively, we should restrict irrelevant hypotheses by an individual evaluation of melody and bass line. Basis of a separate analysis is the human auditory system, which groups musical events [5]. Grouped musical events form streams, which are perceived independently. We hypothesize that evaluating only IOIs of streams respective distances between connected tones should improve the quality of beat-detection-algorithms.

2.1 Stream segregation

We consider a stream to be a path through successive events. Whether two tones are interconnected within a stream depends on their presentation rate *PR* and pitch difference Δp. The inverse PR^{-1} describes the average time between two events. It typically ranges from 0.1 to 0.8s. The greater *PR* is the more streams are likely to be segregated [6]. Van Noorden [7] formalized the maximal pitch and volume difference Δp_{max} respective Δv_{max} (in dB):

$$\Delta p_{\max} = \begin{cases} \text{not defined} & PR^{-1} < 0.1s \vee PR^{-1} > 0.8s \\ 1 & 0.1s \le PR^{-1} \le 0.4s \\ 10PR^{-1} - 3 & 0.4s < PR^{-1} \le 0.8s \end{cases}$$

$$\Delta v_{\max} = \begin{cases} \text{not defined} & PR^{-1} < 0.1s \vee PR^{-1} > 0.8s \\ 3 & 0.1s \le PR^{-1} \le 0.4s \\ 25PR^{-1} - 5 & 0.4s < PR^{-1} \le 0.8s \end{cases}$$

To find an optimal allocation of all tones *N(e)* to all possible streams *S* for each event *e* requires a rating value

$R(s,n)$, which describes how good n fits into s for each $n \in N$ and each $s \in S$. Non-allocatable tones form a new stream. Finally, by evaluating all ratings we can determine the ideal allocation of tones to streams. $R(s,n)$ delivers values between 0 and 100 or less than 0. Is $R(s,n)$ negative, no allocation happens. Otherwise, $R(s,n)$ describes how good n fits into s. To determine $R(s,n)$, we have to compute single ratings of component parameters first.

For the current event e_i at timestamp $t(e_i)$ the presentation rate PR^{-1} consists of the average distance of k predecessors:

$$PR^{-1}(t(e_i)) = k^{-1}(t(e_i) - t(e_{i-1})) + \ldots + (t(e_{i-k+1}) - t(e_{i-k}))$$

$\Delta p(s,n)$ denotes the pitch difference and $\Delta v(s,n)$ the volume difference between n and the last tone of s. Both are rated by $R_p(s,n)$ respective $R_v(s,n)$:

$$R_p(s,n) = \begin{cases} -1 & \Delta p(s,n) > 2\Delta p_{\max} \\ \dfrac{-50}{\Delta p_{\max}} \Delta p(s,n) + 100 & otherwise \end{cases}$$

$$R_v(s,n) = \begin{cases} -1 & \text{if } \Delta v(s,n) > 2\Delta v_{\max} \\ \dfrac{-50}{\Delta v_{\max}} \Delta v(s,n) + 100 & \text{otherwise} \end{cases}$$

$R_d(s,n)$ describes whether n continues an ascending or descending pitch sequence. $R_c(s,n)$ describes whether n continues a sequence of tones with similar IOIs. In both cases *count()* denotes the length of a found sequence.

$$R_d(s,n) = \min(100, 20 count(d))$$
$$R_c(s,n) = \min(100, 20 count(d))$$

$R_t(s,n)$ rates the temporal distance from n to last tone of s:

$$R_t(s,n) = \begin{cases} 0.1^{\Delta t(s,n)} & \Delta t(s,n) \le 2 \\ -1 & \Delta t(s,n) > 2 \end{cases}$$

All single ratings have to be integrated into $R(s,n)$. However, pitch is the most important criterion for stream segregation. The other criteria affect stream segregation only if at least two of them indicate the same result. Therefore we combine the additional criteria first:

$$R_{v,d,c}(s,n) = 0.3R_v(s,n) + 0.55R_d(s,n) + 0.15R_c(s,n)$$

These coefficients have been determined empirically. $R_{v,d,c}$ is useful to confirm or attenuate a pitch-based rating. We have to check, if $R_{v,d,c}$ indicates a different stream allocation than R_p. If $R_p < 50$ and $R_{v,d,c} \ge 50$ or $R_p \ge 50$ and $R_{v,d,c} < 50$ then influence of pitch is diminished: Thus

$R(s,n) = R_t(s,n)(0.3R_p + 0.7R_{v,d,c}(s,n))$, otherwise

$R(s,n) = R_t(s,n)(0.7R_p + 0.3R_{v,d,c}(s,n))$

2.2 Beat estimation

For a perceptually plausible time window of a few seconds we construct a histogram to show the frequency of directly connected IOIs of each stream within the window. Then, we attempt to generate a beat hypothesis that is consistent with the most frequent intervals and which lies within a plausible time frame from 60 to 240 beats per minute. By stepwise moving of this window, we can evaluate the beat-evolution over an entire song and are able to discover typical performance deviations, such as ritartando and rubato.

3. Results

Based on a real performance of Debussy's Arabesque Figure 3 shows the resulting stream segregation of the first time window:

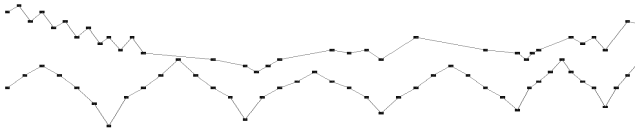

Figure 3. Stream segregation of the Debussy-fragment

One can recognize that both left and right hand score form their individual streams.

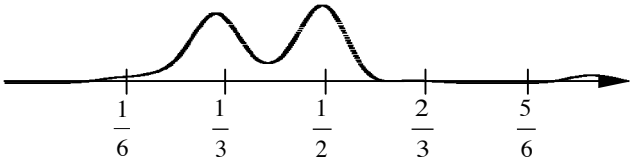

Figure 4. Stream-based IOI-distribution

Figure 4 contrasts the results of stream-based interval estimation to the pure IOI-distribution of figure 2. As predicted, the stream-based algorithm only finds eights and triplets. Integrating these IOIs in a consistent hypothesis results in the supposed quarter note beat.

A short subjective study confirmed our findings especially for romantic and impressionistic music, which shows a high parallelism of complex tuplets.

4. Conclusions & further work

We have presented a stream-based approach for improving IOI-based symbolic tempo detection systems. By implementing a perception-oriented evaluation of events we could reduce the amount of perceptually irrelevant data considerably and improve the quality of beat estimation substantially. Furthermore, our approach should enable a much more reliable detection of difficult rhythmic situations, such as rubato or ritartando.

Our future research will focus on a quantitative and qualitative comparison of our system with existing ones.

References

[1] S. Dixon. "A lightweight multi-agent musical beat tracking system," *In PRICAI 2000 Proc. Of the Pacific Rim Int. Conf. on Artificial Intelligence*, Springer, 2000.

[2] B. Pardo: "Tempo Tracking with a Single Oscillator," in *ISMIR2004 Fifth Int. Conf. on Music Inf. Retr. Proc.*, 2004.

[3] P. Allen, R. Dannenberg: "Tracking Musical Beats in Real Time," in *ICMC 1990 Int. Comp. Music Conf. Proc.*, 1990.

[4] J.C. Brown, "Determination of the meter of musical scores by autocorrelation," J. Acoust. Soc. Am. 94 (4), Oct. 1993.

[5] A.S. Bregman. *Auditory Scene Analysis*. Cambridge, MA: Bradford/MIT Press, 1990.

[6] S. Handel. *Listening*. MIT Press, 1989.

[7] Van Noorden. "Temporal coherence in the perception of tone sequences, Institute of Perception Research, Diploma thesis, 1975.