Score Following of Orchestral Music Using Acoustic Pressure Peak-Tracking and Linear Stretch Matching

Takefumi Miura¹, Ayumu Akabane¹, Makoto Sato¹, Takao Tsuda² and Seiki Inoue²
¹Precision and Intelligence Laboratory, Tokyo Institute of Technology
tmiura@hi.pi.titech.ac.jp
²Science & Technical Research Laboratories, Japan Broadcasting Corporation

Abstract

In this research, we present a new real-time score following system for orchestral music. A score of orchestral music includes portions where we can follow the score by adopting the existing method using the short-time spectral analysis and Dynamic Programming (DP) matching. Also, it includes portions where the fundamental tones and overtones of many instruments are complicatedly overlapped; hence, the existing method can not be used. But we can observe distinct acoustic pressure peaks in the latter portions. So we propose a new score following system using these pressure peaks, which should be used in combination with a system using the existing method. In this research, first, we originate a new method with which we can presume pressure peaks from an orchestral score. And then, in order to follow the score, we match these presumed peaks with the corresponding peaks extracted from the performance sound. When we use DP in the process of the matching, it is often occurred that the time values of notes are deviated from the correct values by the nonlinear characteristic of DP. Finally, we propose a new linear stretch matching method to avoid these deviations of time values. We implement a system using this new method and compare its results with those obtain by using DP.

1 Introduction

In recent years, several algorithms and techniques have been developed in the field of the score following (Grubb and Dannenberg 1997), (Orio et al. 2003) and in its related fields, such as the score alignment (Müller, Kurth, and Röder 2003) and the beat tracking (Goto 2001). As a result, it has become possible to follow a score automatically in monophonic music cases and step by step in polyphonic music cases.

On the other hand, an automatic score-based camerawork planning has begun to be proposed for orchestra shooting (Shiba et al. 2003). So we think that it is the time to take a first step toward the full-automatic TV-shooting of orchestral performance program, in which score following techniques are to be adopted.

However, the existing score following systems using short-time spectral analysis and DP matching (or Hidden Markov Model) can follow instruments with low polyphony such as a violin, but cannot track highly polyphonic instruments until now (Orio et al. 2003). These methods analyse the frequency spectrums of the performance sound, and use these spectrums to detect pitches of notes in the corresponding score.

In orchestral performances, many instruments often make sounds of different pitches at a time and it is very difficult to detect pitches at these moments. But we can observe distinct pressure peaks (corresponding to onset of notes) in these portions. So we propose a score following system using these peaks instead of pitch-detection, which is to be used in combination with the existing system of pitch-detection.

2 The pre-matching processes

In our research, we use standard transcriptions (CDs of Orchestral performance) as our sound sources and extract acoustic pressure peaks \( p_a(t) \) from them. In this process, first, we rectify the performance sound, and then low-pass filter it. By using this process, we can get the required pressure peaks. (Fig. 1)

2.1 Presumed pressure peaks from a score

To compare the performing sound with the corresponding score, we originate a method to presume pressure peaks from the score. Fig. 2 represents the method with which we get the presumed pressure peaks. First, we change an orchestral score to a standard MIDI file. In the process, we make each part of the orchestral score corresponds to each channel of MIDI, and the dynamics of each note on the orchestral score corresponds to the Velocity byte of Note-On status. Next, we define a relative weight that is proportional to the acoustic pressure from each instrument. We obtain the Total Velocity...
by multiplying the Velocity bytes with the relative weights and then adding the results over all channels. We specify the positions on the score-time $T$ (i.e. the time axis of status’ time in the standard MIDI file) when large peaks are expected by using the peak-finding techniques with a threshold on this Total Velocity. We then put a unit of pressure peak on each of these positions. After these processes, we can get the desired presumed pressure peaks $p_s(T)$.

2.2 Definition of the matching segments

To define the matching segments on the score-time $T$, we set two parameters for $n$-th segment which are the length of the segment ($L = T_{\text{end}}(n) - T_{\text{begin}}(n)$) and the advance of the segment ($L_A = T_{\text{begin}}(n) - T_{\text{begin}}(n - 1)$), as is shown in Fig 3. The parameters ($L$, $L_A$) keep constant throughout all segments.

We divide the segment into $N$ intervals by introducing a resolution $\Delta T$ of the score-time; hence, the length of segment $L$ becomes $L = N \Delta T$. By dividing $T$ into a width of $\Delta T$, we can get the discrete time $T_i = i\Delta T + T_{\text{begin}}(n)$ ($i = 0, \ldots , N$) on the $n$-th segment and the $p_s(T_i)$ as the value on the presumed pressure peaks. This $p_s(T_i)$ constitutes a $(N + 1)$-dimensional vector.

3 Linear stretch matching method

The score alignment of the performing sound is totally non-linear. For realizing the non-linearity of the alignment, DP matching is frequently used. However, in our case of the score following, we match the score with the performing sound on the relatively short-time segments. And we could regard the matching on the short-time segments as linear. So we could use linear-time matching method instead of DP (Müller, Kurth, and Röder 2003), because, in DP matching, a partial stretch (i.e. performance tempo changes drastically) occurs frequently even on short-time segments. This
partial stretch occurs under the influence of the neighboring large peaks. (Fig. 4)

We think that this partial stretch could be avoided if we use a linear stretch matching. So we originate a linear stretch matching which is based on the template matching. We regard the presumed pressure peaks vector \( p_s(T_i) \) as the standard pattern (the template). We specify a segment as \( l = t_{\text{end}} - t_{\text{begin}} + \Delta l \) on the performance current time, where \( t_{\text{end}}, t_{\text{begin}} \) are the estimated current time corresponding to \( T_{\text{end}}, T_{\text{begin}} \) respectively before the matching, and \( \Delta l \) is a margin on the current time. Half of this margin is put on both sides of \( l \) in order that \( t_{\text{end}}, t_{\text{begin}} \) are surely included in \( l \) for any matching result. As input patterns of the template matching, we transform \( p_a(t_i) \) on the segment \( l \) to \( 2M + 1 \) patterns using different performance tempo \( V(m) \) (diff. of score-time / diff. of performance time), as:

\[
V(m) = \tilde{V} \times \text{(maximum stretch rate)}^{m/M} \tag{1}
\]

where \( \tilde{V} \) is the accumulated mean tempo, and \( m \) is a stretch parameter which is an integer from \(-M \) to \( M \). The correspondence between the performance time \( t_i \) and the score-time \( T_i \) is given by:

\[
t_i = T_i - T_{\text{base}} \frac{V(m)}{V(m)} + t_{\text{base}}(m, k) \tag{2}
\]

where \( T_{\text{base}} \) is defined as the begin time \( T_{\text{begin}}(n) \) of the \( n \)-th segment and \( t_{\text{base}}(m, k) \) is the current time corresponding to \( T_{\text{base}} \). The optimal value of this parameter \( t_{\text{base}}(m, k) \) is given as the matching result.

The matching distance is defined as:

\[
D(m, k) = \sum_{i=0}^{N} \left( p_a(t_i) - \alpha p_s(T_i) \right)^2 \tag{3}
\]

where \( p_a(t_i) \) is the acoustic pressure peak value over the segment \( l \). The dimension \( N + 1 \) of the pressure vector \( p_a(T_i) \) is identified with the dimension of the characteristic space in the template matching and \( \alpha \) is a normalization factor between \( p_a(t_i) \) and \( p_s(T_i) \) on \( n \)-th segment.

Substituting Equation 2 into Equation 3, we get Equation 4.

\[
D(m, k) = \sum_{i=0}^{N} \left\{ p_a \left( \frac{T_i - T_{\text{base}}}{V(m)} + t_{\text{base}}(m, k) \right) - \alpha p_s(T_i) \right\}^2 \tag{4}
\]

The template matching is executed by shifting the template \( p_a(T_i) \) (length \( L \)) over \( l \times V(m) \). The shifting step \( k \) of the matching is an integer from \( 0 \) to \( k_{\text{max}}(m) \) where \( k_{\text{max}}(m) \) is the largest integer that is not beyond \( \{ l \times V(m) - L \} / \Delta T \). We choose the length \( \Delta l \) of the margin long enough to make \( k_{\text{max}}(m) \) considerably large for any stretch parameter \( m \).
The matching result is decided as follows; First, for each stretch parameter \(m\), the optimal value \(k_{\text{opt}}(m)\) of the shifting step \(k\) is decided by \(k_{\text{opt}}(m) = \arg\min_k D(m, k)\). Next, the optimal value \(m_{\text{opt}}\) of the stretch parameter \(m\) is decided by \(m_{\text{opt}} = \arg\min_mD(m, k_{\text{opt}}(m))\). Then we get the optimal values \((V(m_{\text{opt}}), T_{\text{base}}(m_{\text{opt}}, k_{\text{opt}}(m_{\text{opt}))))\) as the matching result.

4 Real-time score following system

In our system, we need to recognize instantaneously the performing position on the score at any instance. However, matching process needs to use a performed segment on current time. So we cannot instantaneously recognize the performing position from the matching results, directly. Therefore, we propose a system in which the recognition is instantaneously executed by using the subsystem as given in Section 4.1. The inner-states are always renewed with a delay of finite time needed in a matching process. The time needed in the matching process is too short to influence the real-time character of our system.

4.1 Subsystem determining the performing position on the score

If any current time \(t_{\text{cur}}\) of the performance is inputted into this subsystem, the corresponding position \(T_{\text{cur}}\) on the score is outputted instantaneously from it. This subsystem uses a linear estimation as:

\[
T_{\text{cur}} = V \times (t_{\text{cur}} - t_{\text{base}}) + T_{\text{base}}
\]

where \(t_{\text{cur}}\) is a current time (at the performance) inputted into this subsystem, \(T_{\text{cur}}\) is the outputted-corresponding position on the score, and \(V\), \(t_{\text{base}}\), \(T_{\text{base}}\) are the inner-states of this subsystem.

\(V\), \(t_{\text{base}}\) and \(T_{\text{base}}\) are given in Section 3. These inner-states are obtained as the results of the matching which is finished just before the running matching.

4.2 Subsystem of the matching process

The following three data are inputted into this subsystem:

- The acoustic pressure peaks \(p_a(t)\) described in Section 2 (Fig. 1)
- The presumed pressure peaks \(p_s(T)\) described in Section 2.1 (Fig. 2)
- The matching segments defined in Section 2.2 (Fig. 3)

Then the next inner-states are outputted from it. This subsystem uses the linear stretch matching method described in Section 3 and we get the values \((V(m_{\text{opt}}), T_{\text{base}}(m_{\text{opt}}, k_{\text{opt}}(m_{\text{opt}))))\) as the matching results. We then update the inner-states of the subsystem described in the previous Section 4.1, by using these three values.

The two subsystems, which are described in this and previous section, are combined to form a real-time score following system.

5 Conclusion

In this research, we proposed a real-time score following system for orchestral musics. In order to make an orchestral score comparable with the corresponding performance sound, first we extracted acoustic pressure peaks from the performance sound and then devised a new method to presume pressure peaks from the score. We showed that our linear stretch matching method is suitable as the matching method of the above two pressure peaks. Based on these methods, we implemented a real-time score following system for orchestral musics. Our system is to be used only in those portions of orchestral musics as a system using the existing pitch-detecting method cannot be applied. We are evaluating this system now. But as long as we attained (e.g. Fig. 4), this system is robust enough to replace DP.

Acknowledgment

This research is sponsored by the assistance grant of Hosobunka Foundation, Inc. (2004/2005 Grant Cycle)

References


