Pattern Anchoring, Identity and Signatures in Sound, Time and Space

Özgür Izmirli
Ammerman Center for Arts and Technology
Connecticut College
270 Mohegan Avenue
New London, Connecticut 06320
oizm@conncoll.edu

Noel Zahler
School of Music
University of Minnesota
200 Ferguson Hall
South Fourth Street
Minneapolis, Minnesota 55118
nbz@umn.edu

ABSTRACT

The process of score following requires that the automated follower be capable of making robust decisions during a performance, especially in light of the number of errors that need to be accounted for between the live performer and the follower. . Based on the idea of variable-length signatures for monophonic input we define a causal method for tracking location in score following. The signatures and their lengths are calculated prior to performance based on their novelty in a predetermined locality within the score. Results are shown using a graphical display that depicts the associations, determined by a score follower, between performed notes and the score. This display aids in visual evaluation of the performance of the score following process and specifically its capability of recovery following any of the performance aberrations.

1. INTRODUCTION

For some time now, our work has been focused on score following [1][2]. Most recently, we have come to believe that methods derived from "traditional" approaches to music analysis used for parsing scores may not be the most efficient or cognitively correct models for mapping a given composition's data input to a machine oriented performance program[2]. The need to have an unambiguous trajectory that can be easily traversed by a score follower is of great general interest. Further, we hope to shed some light on the needs of any automatic, real-time, computer assisted performance program and how this division of territory may be of use when contemplating issues related to performance, signal interpretation and fragmented data reception. Special cases, including accompaniment and spatialization, will also be considered, but we also make a special effort to show relevance to other areas where the division of a work into domains and regions may be of interest. Dance, theater, and installation art all work with texts that must be analyzed and traced. In an effort to offer a solution to such tracking, we have reappraised our own methods and those of others. We offer a possible solution as it relates to our work in score following. The method is based on finding signatures of different sizes for all notes of a score and incorporating them into a score follower capable of using such information.

2. PERFORMANCE

In score following, performance considerations have taken the form of tempo and beat tracking, heuristic definition of common performance mistakes, and incomplete data assessments for intelligent accompaniment (improvisation). What has been excluded from this discussion, until recently, has been consideration of the kind of formal and analytical process with which performers overlay a score before they perform the work. Machine pre-performance score consultation has been used by one of the authors in previous work [3] but the intelligent division of a score into structurally significant areas that can be used by a score following program in much the same manner as a performer "chunks" a musical composition in order to measure his/her progress through the piece has, to our knowledge, not been used. This ability to "know" the territory through which a performer will pass has only been attempted by programs seeking to have the ability to recall and learn from each performance and the type of information extracted by these programs is quite different from what we are trying to do [4]. An intelligent score following program needs to have the ability to recognize signposts, in much the same way as their human counterpart. These flagged areas, or as we have termed them, signatures, divide the score according to unique patterns whose infrequency of occurrence distinguish them from the more common "musical analysis" we have witnessed. They act as "markers" and they fly in the face of traditional methods that divide a composition into phrase, section and movement; the traditional divisions calculating similarity and repetition as markers of coherence. Why then have we found this traditional approach to marking a score poor for the purposes of score following? Why is it that we have concluded that a new morphology is needed for machine driven performance?

3. ANALYSIS

Just about all approaches to "musical analysis" rely on symmetry and repetition as syntactical elements in building morphological models of the compositions they examine. Symmetry has proven itself a useful tool in the definition of pattern mapping for score following but repetition, in most cases, confuses the issue by making it impossible to identify where a particular motivic signature appears in a work without a specific "time stamp." Dannenberg [5], in his early work on score following, while preparing the machine copy of the score, chose to eliminate all consecutive repetitions of the same pitch because it was impossible for the machine to differentiate between repeated instances of the same note. When dealing with larger patterns (motives), some types of compositions, say Bach fugues, the motivic material may be so similar that it is more fruitful to consider what is not "motivic" rather than what is motivic. In our own work [2], we have shown how we calculate signatures as "anchor" points. The salient issues regarding signatures pertain to the composition of the identifying mark. In other words, what are the elements that comprise the signature and how may they be generalized? In this instance, there are a number of similarities that might be compared to data mining. In our own work, it has become clear that signatures must be of variable length to be effective markers of a body of composition. This might be of further help in generalizing the use of signatures in both gesture (theater) and space (dance). Therefore, our investigation takes on an interest in calculating what is the optimal technique of traversing a score through the unique division of its topology by markers that point to a unique location. We envision the application of these methods to continuous domains such as theater and dance through some symbolic means of representation. That is to say, for example, a score in the context of dance consists of a sequence of characteristic key frames (whether in the manner of film, video or Laban notation) defining movement in space from which the performance has ensued.

4. PATTERN MATCHING AND SIGNATURES

The concept of using previously calculated unique patterns as anchors in score following was introduced as a means of replacing the process of 'chunking' the score prior to performance [2]. The signature matching mechanism was proposed as a process that had global information (within a piece), and operated above the score follower that concentrated on local information. The detection of an anchor during score following either would affirm the position of the score follower or would put it back on track if it were lost. The signatures utilized in our previous work were fixed in length and global within a piece. That is to say, the anchors were calculated for the entire piece and as the performance unfolded the current input pattern was compared to all of the remaining (uncommitted) anchors. Global anchors have shown to be useful in this context but due to the real time processing constraints that limit the size and number of anchors and the information content of the score that imposes a particular distribution of anchors over time, their use has been limited to relatively long pieces that have few fragments of repetition. This led us to explore a more flexible form of signature matching in which the anchors were of varying size and more in number. The method presented here treats each musical onset as a local anchor with an associated degree of uniqueness. Depending on the size of the pattern, appropriate values are calculated that distinguish one anchor from another. These uniqueness values are then consolidated to determine the shortest length of a pattern that best describes this point in the score.

Given a sequence of musical events, it is straightforward to calculate the similarity matrix and determine either repetitions or find unique patterns within a piece using a predetermined window size. In this case, the window size needs to be determined prior to processing and the optimal size of the window will vary depending on the music content. The choice of window size is a tradeoff between a short window, that favors real time implementation due to the reduced computational load, and a longer window for more reliable pattern matching. The aim, in the context of score following, is to find novel patterns in a work that are either locally distinct or, preferably, globally unique. These patterns, called signatures, are significant markers that act as placeholders and enable a follower to move to these points in the score without ambiguity. In the following, we describe a method for determining the minimum size signature while locally maintaining pattern novelty. As this signature mechanism will only have meaning in reference to a score follower, we also briefly discuss their integration and present the proposed algorithm for score following.

The initial problem is one of finding the shortest pattern that is unique for each note. One fundamental problem in determining signatures by analyzing the score is that of representation and pattern similarity. Similarity is pertinent to score following for the reason that any number of changes to the score can be encountered during performance due to errors made by the performer or the signal processing front-end. There has been considerable work on melodic similarity [e.g. 6, 7, 8, 9] emphasizing different approaches, and comparison of musical sequences independent of key, mode or tempo [10]. In our work, we are also interested in melodic similarity but in a more isolated and information theoretic way. Hence, our model does not deal with finding melodic descriptors such as phrase structure, cadence, implied harmony etc.

We define score following to be real-time. This assumes that even though the entire score is known to the follower at all times, the information about performance is only available up to the current time during the score following process. This is in contrast to non-real-time score-performance matching in which the matcher has all the information at its disposal (e.g. [11]). Heijink et al. [12] summarize approaches to this problem and discuss methods of exploration of multiple alternative matches in order to find the optimal solution in matching a score to a performance. Our approach is different in that we calculate signatures and their lengths prior to performance, perform local searches during performance and make real-time decisions based on the available information at whatever point in time we happen to be at in the rendering of the performance.

5. FINDING SIGNATURES AND LENGTHS

Each note in the score is processed in relation to its neighboring notes to find its minimum size signature. Signatures are found within a locality, L, given in number of notes, around the single note under consideration. The choice of this locality is not very critical and its size can vary from a small number of notes to the length of the entire piece. Notes are represented with 2-tuples, (a,b), the first element holding the onset time value and the second holding the midi pitch value. The 'score' data structure is a list of these 2-tuples.

The time values in the score serve to express the relative durations of the notes and have no direct correspondence to the tempo of the performance in follow mode. The midi note values are absolute. The dissimilarity of two patterns in the score is calculated using the Levenshtein distance, or so called edit distance. This is calculated for each note in the score within its locality L. Overlapping of the search patterns is not allowed.

The edit distance is calculated with a dynamic programming algorithm using the following recurrence equations:

$$d_{i,j} = \begin{cases} d_{i-1,j} + C_{del} \\ d_{i-1,j-1} + C_{rep} \\ d_{i,j-1} + C_{ins} \end{cases} \quad (1)$$

where C_{del}, C_{rep} and C_{ins} are the costs associated with deletion, replacement and insertion respectively. The initial conditions are $d_{0,0} = 0$, $d_{i,0} = d_{i-1,0} + C_{del}$ for all i and $d_{0,j} = d_{0,j-1} + C_{ins}$ for all j.

As mentioned earlier, score following requires that midi pitches be absolute and time information be relative. For example, a transposition of a phrase with the same timing is not considered similar to the original, whereas, a time-dilated segment with the same sequence of pitches is deemed similar. The equality operator for this algorithm uses midi pitch equality and the ratio of the duration of the current note to the duration of the previous note. That is to say, two notes are equal if their pitches are equal and their duration ratios are within a percentage range determined by δ:

$$\frac{1}{1+\delta} \le \frac{\left(a_m / a_{m-1}\right)}{\left(a_n / a_{n-1}\right)} \le (1+\delta) \quad (2)$$

The edit distance between a pattern at position m and a pattern at position n is calculated for various lengths, r:

$$S(m,n,r) = E\{\text{score}(m), \text{score}(n), r\} \quad (3)$$

where E is the r-element edit distance operator using the source sequence ending with note m, target sequence ending with note n in the score. That is to say, for a three element window the notes would be score(m-2), score(m-1) and score(m) with the latter having the largest time value. The minimum size pattern is then given by R(m)=r that satisfies the following:

$$M(m) = \max_{n,r}\left[\frac{S(m,n,r)}{r}\right]; \quad \begin{matrix} n \ge (m+r) \text{ or } n \le (m-r) \\ r = 1 + 2^k, \quad k = 0..K \end{matrix} \quad (4)$$

M(m) is the minimum dissimilarity and K+1 is the number of different window sizes. R(m) is the shortest pattern that is least similar to all other patterns in the locality L for note m. Here, r is chosen to take exponentially increasing values, but, the choice can be changed to a linear relationship which will in fact give more accurate minimal pattern lengths at the expense of increased computation times.

6. THE SCORE FOLLOWER

The score follower has been designed keeping the real-time aspect at the central focus. Although the model has access to future score information, naturally, it does not have insight into the future values of the performance. The score follower algorithm given below combines the properties of our previous score follower algorithm [1] and a mechanism that incorporates signature length information on a note by note basis. The score and the performance are sequences of musical events represented by 2-tuples. They can have entirely different time scales as long as their starting tempos are comparable.

The algorithm for the score follower is given below. Some abbreviations used in the algorithm are as follows: Score Tempo (**STE**): holds the value of score tempo in score time units; Score Time (**STI**): interpolated position of the score in score time units; Score Pointer (**SP**): index to note in score; Performance Pointer (**PP**): index to performed note in the performance list; **R**(m) : length of signature at note m; α : similarity threshold, $P(m,n) = E\{\text{score}(m), \text{performance}(n), R(m)\}$.

```
SP =1;   PP =1;  STI = 0;
do
    interpolate STI to reach SP+1 with current STE;
    if a note onset is encountered play the note;
    if new performance event encountered, update PP
        find the index I=i that has the smallest P(SP, i),
            with SP – R(SP) <= i <=  SP + R(SP);
        if P(SP, I ) < α and score(I) = = performance(PP)
            SP = I and play note if not played;
            associate score(SP) with performance(PP);
            calculate STE;
        end
    end
while( STI <= STI_max )    // real-time loop
```

7. RESULTS

The operation of the model is demonstrated by means of a causal simulation. Results are shown by use of an association plot in which notes of the score and the performance are displayed. This plot enables visualization of the associations of notes in the two streams. The top subplot, in each figure, shows the midi pitch value versus onset time of the notes in the score. The lower subplot shows the notes for the performance. Associations that have been established as a result of the score following are shown as lines connecting the notes of the score to the ones that have been performed. An association between a note in the score and a performance note means that the score follower is locked onto the input note stream and is following with confidence. A note in the score that is not associated however, is still played when its time is crossed by the score time variable but is an indication of ambiguity.

The parameters chosen for the following examples are as follows: L=80, C_{del}=0.75, C_{rep}=0.65, C_{ins}=1 and K=3. Figure 1 shows an initial fragment from Bach's Partita BWV1013 Allemande. The performance is the copy of the score with identical midi pitch values and onset times. The plot at the top shows the associations and the plot below it shows the corresponding onset times of the associated notes. In this case, both streams have the exact same tempo leading to a straight line with an angle of 45 degrees. All notes have been associated in this figure.

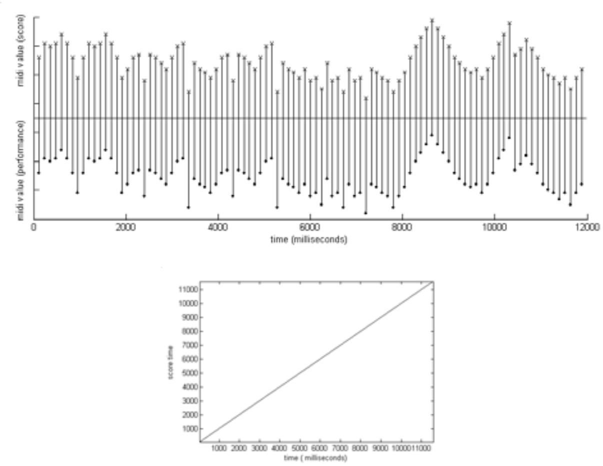

Figure 1. Fragment of Bach Partita BWV1013 with identical performance and score streams.

Figure 2 shows the same piece with initially a decrease in tempo followed by an increase. Again all notes have been associated. It can be seen from the plot that the score is leading in terms of its own time scale. Nevertheless, the follower has stayed locked to the performance. Figure 3 shows a 3% slower performance compared to the score. In addition to this, for each note, a note error is introduced with a probability of 0.1. The note error can be one of note insertion, note deletion or pitch replacement with equal probabilities. If no note error is introduced then the time onset is altered with a uniform distribution of up to half the distance between the current note and its neighbors. It can be seen from the figure that some associations are missing, mainly due to deletions in the performance. This however, has not interrupted the course of the score following and as can be seen from the lower plot, the path is straight with a slope less than 45 degrees to reflect the slower performance. Figure 4 shows a similar performance using Telemann's Fantasia No.1 with 4% slower tempo, note errors and onset alterations as in the previous figure.

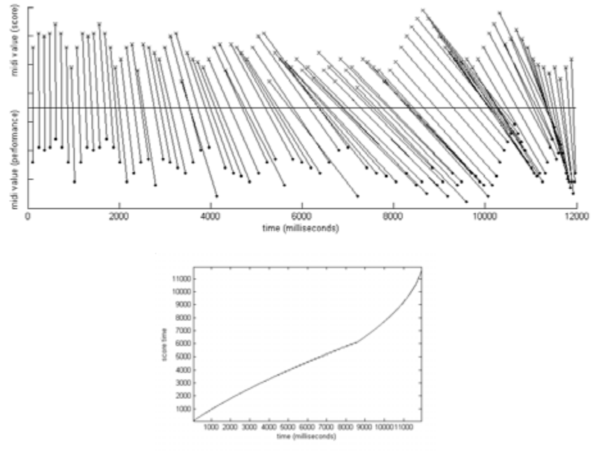

Figure 2. Fragment of Partita with a decrease followed by an increase in tempo of performance.

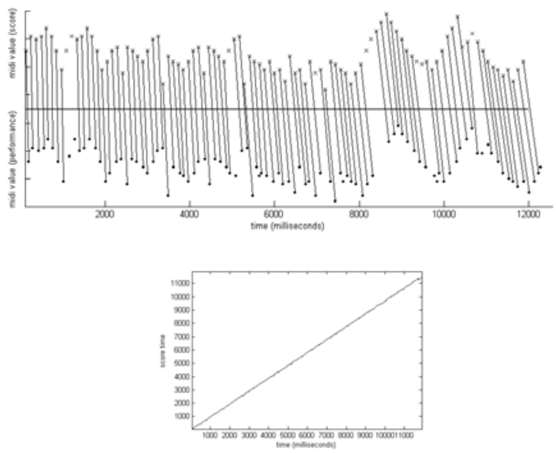

Figure 3. Fragment of Partita with 3% slower performance, note errors and onset alterations.

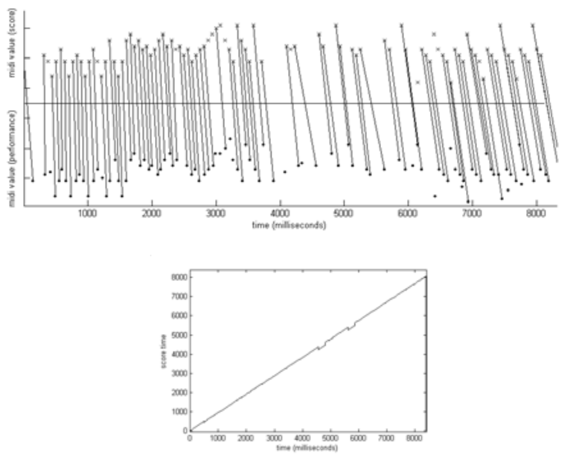

Figure 4. Fragment of Telemann Fantasia with 4% slower performance, and errors as in Figure 3.

8. DISCUSSION

The choice of locality L is of particular interest. Larger locality will enable larger spans of follower awareness, enabling it to speed up and move to distant positions ahead of its current position. The disadvantage of choosing a large L is that signature patterns tend to be longer in size due to the increased probability of complete or partial repetitions of shorter patterns in the larger span. This, of course, is with the assumption that there are no exact repetitions in the piece. On the lower end, L needs to be significantly larger than the maximum signature length in order for the method to evaluate the similarities in its neighborhood.

The score following algorithm described above assumes errors are rare and are not encountered in bursts that are much longer than the maximum signature length. If such a case is encountered, the performance pointer might fall outside of the scope of the score pointer and the follower might get lost. For this particular case, a solution might be to increase the size of the maximum signature length (determined by K). This comes at the expense of increased computational cost and hence latency, both in real time and off-line (pre-processing,) and longer delays in recovering from minor errors. This also applies to repeated notes which have been problematic for score followers in general. In the presence of repeated notes and especially with the lack of rhythmic variation our follower might lose track. In this case, it will have to regain focus when new musical patterns are encountered following the sequence of repeated notes.

9. CONCLUSIONS

We have begun an investigation wherein the events of a performance can be unambiguously matched to a score. Accurate tracking can be facilitated with an acceptable rate of error for a score follower regardless of change in tempo because the score follower now has the ability to reference events in the score in a more precise and better suited manner. This is facilitated by a method that allows for a variable size pattern that characterizes each signature. The mechanism that utilizes this information is integrated into a score follower that has been shown to have robust qualities. This is a method that can be generalized to scores other than those used for music and will greatly facilitate the automation of text or symbol based performances.

10. REFERENCES

[1] Izmirli, O., Seward, R and Zahler, N. “Compositional Imperatives for Implementing an Audio Alignment Program in MAX/MAP.” *Proceedings of the International Computer Music Conference (ICMC),* Göteborg, Sweden, pp. 266-269, 2002.

[2] Izmirli, O., Seward, R and Zahler, N. “Melodic Pattern Anchoring for Score Following Using Score Analysis.” *Proceedings of the International Computer Music Conference (ICMC)*, Singapore, pp. 411-414, 2003.

[3] Baird, B., Blevins, D. and Zahler, N. "Implementing an Interactive Computer Performer." *Computer Music Journal*, Summer, M.I.T. Press, Cambridge, Massachusetts, 1993.

[4] Vercoe, B. and Puckette, M.. “Synthetic rehearsal: Training the synthetic performer.” *Proceedings of the International Computer Music Conference (ICMC)*, pp.275-278, 1985.

[5] Dannenberg, R. and Mukaino, H.. “New Techniques for Enhanced Quality of Computer Accompaniment,” *Proceedings of the International Computer Music Conference (ICMC)*, pp. 243-249, 1988.

[6] Crawford, T., Iliopoulos, C. and Raman, R. “String-Matching Techniques for Musical Similarity and Melodic Recognition,”. In *Melodic Similarity, Concepts, Procedures, and Applications*, ed. Hewlett, W. and Selfridge-Field, E., Computing in Musicology 11. MIT Press, 1998.

[7] Typke, R., Giannopoulos, P., Veltkamp, R.C., Wiering, R., van Oostrum, R., “Using Transportation Distances for Measuring Melodic Similarity,” *Proceedings of the International Symposium on Music Information Retrieval (ISMIR),* Baltimore, MaryLand, USA, 2003.

[8] Hofmann-Engl, L.. “Towards a cognitive model of melodic similarity,” *Proceedings of the International Symposium on Music Information Retrieval (ISMIR),* Bloomington, Indiana, USA, 2001.

[9] Cambouropoulos, E., Crochemore, M., Iliopoulos, C.S., Mouchard, L. and Pinzon, Y.J., ”Algorithms for Computing Approximate Repetitions in Musical Sequences,” *International Journal of Computer Mathematics*, 79(11):1135-1148, 2002.

[10] Mongeau, M. and Sankoff, D., “Comparison of Musical Sequences,” *Computers and the Humanities,* 24:161-175, 1990.

[11] Large, E. W., “Dynamic Programming for the Analysis of Serial Behaviors,” *Behavior Research Methods, Instruments and Computers,* 25(2):238-241, 1993.

[12] Heijink, H, Desain, P., Honing, H. and Windsor, L., “Make Me a Match: An Evaluation of Different Approaches to Score-Performance Matching,” Computer Music Journal, 24(1): 43-56, 2000.