Music Performance Generation as Time Series Prediction

Maarten Grachten and Josep Lluís Arcos

IIIA, Artificial Intelligence Research Institute,
CSIC, Spanish Council for Scientific Research,
Campus UAB, 08193 Bellaterra, Catalonia, Spain,
{maarten,arcos,mantaras}@iiia.csic.es,
http://www.iiia.csic.es

Abstract. In this paper, we present a case-based reasoning system for rendering expressive music performances in the context of popular jazz themes. Music performance by skilled musicians is far from being a literal translation process from notes written at the score to sounds with the nominally described attributes. These score deviations can vary greatly from one musician to another musician and represent the musician style. The case-based reasoning system presented deals with this problem as a time series prediction problem.

1 Introduction

In this paper, we present a case-based reasoning system for rendering expressive music performances in the context of popular jazz themes. We propose to model the problem as a time series prediction problem and describe how a CBR approach can handle the prediction problem. Music performance by skilled musicians is far from being a literal translation process from from notes written at the score to sounds with the nominally described attributes (such as onset time and duration) [9,1]. The attributes of performed notes typically differ from their nominal values to varying degrees. These score deviations are certainly not random by nature, since musicians are well able to reproduce a particular pattern of deviations [8]. On the other hand, within the range of musically acceptability, musicians can vary greatly the nature of their performance (i.e. the pattern of deviations), depending on the factors like the affective mood that is being expressed (e.g. sad vs. joyful) [3,5] or the tempo (e.g. slow vs. fast) [2].

In this paper we focus on the problem of predicting the performance continuation of a musical score, given a performance of the initial part of the score. The given performance has a characteristic that is appropriate for a particular (but unknown) musical tempo. The problem is to predict the performance continuation in such a way that the predicted performance is also suitable for that tempo.

This work is part of the ProMusic project [4], a research project that has as a goal the use of case-based reasoning techniques for the automatic generation of expressive music performances. In that context, the tracking of a given performance and a good prediction model has an important role.

2 A CBR Approach to Music Performance Prediction

We have applied a CBR approach to the problem of music performance prediction. The input problem is a musical score (i.e. a symbolic representation of the melody), and a partial performance, represented as a sequence of events with onset and duration attributes. The goal of the CBR system is to predict the sequence of events in a segment-wise manner, by predicting a segment of performance events, using the performance events of the previous segment as a guide. The case base consists of a set of musical phrases, represented by their score. Each case contains a number performances (as played by a professional musician) at different tempos (about 11 tempos for each phrase). The inference process can be summarized with the following steps:

1. retrieve the cases from the case base whose phrases are most similar (comparing analyses the musical scores) to the input phrase
2. segment the input phrase into melodic segments
3. segment the retrieved phrases into melodic segments
4. for two consecutive segments $\langle s_i, s_{i+1} \rangle$ from the input performance:
   (a) find the two most similar consecutive segments $\langle r_j, r_{j+1} \rangle$ from the retrieved phrases
   (b) for the performance $P(s_i)$ of s_i, find the tempo T, such that the performance of $P^T(r_j)$ of r_j at T is closest to $P(s_i)$
   (c) use the performance values of $P^T(r_{j+1})$ to estimate the values of $P(s_{i+1})$
5. repeat step 4 for each consecutive pair of segments from the input phrase

Step 1 is realized by comparing analyses of the musical scores according the Implication/Realization (I/R) Model [6], a model for the structure of melodies, based on gestalt principles. These analyses describe the melodies as a sequence of archetypical patterns (usually called *structures*), that capture the degree to which the various parts of the melody realize the listener's expectations. Such analyses are supposed to present a more abstract and cognitively relevant description of the melody than sequences of notes. The distance between melodies is assessed by calculating the edit-distance between the corresponding sequences of (possibly overlapping) I/R structures.

The segmentation (steps 2 and 3) is also based on the I/R analysis, since this analysis implicitly captures boundaries between perceptual note groups (such boundaries are not spanned by an I/R structure).

In step 4a, a melodic similarity assessment is done for a fragment of the input phrase, and the various fragments of retrieved phrases. This assessment is realized by calculating the edit-distance between the sequences of notes within the fragments. A by-product of this assessment is an alignment between the notes of the input fragment and the retrieved fragments, which is used in step 4b, to match the expressive deviations of the input phrase against the expressive deviations of the retrieved phrase performances. The sum of the pairwise differences between onset-deviations and duration-deviations is used as a measure of similarity between performances. The retrieved phrase segments $\langle r_j, r_{j+1} \rangle$ with the

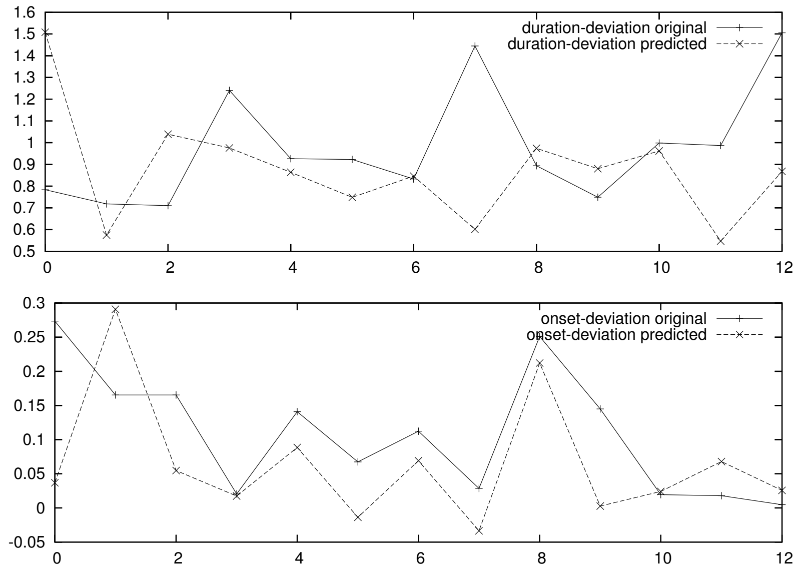

Fig. 1. Prediction of duration deviations (upper plot; measured as the ratio of nominal note duration) and onset deviations (lower plot; measures in metrical beats) for BodyAndSoul-B1 played at 60 bpm

highest melodic and performance similarity are used to predict the performance of the latter (s_{i+1}) of the two input segments. This is done in step 4c, by applying the expressive deviations of the elements of ($P(r_{j+1})$) to s_{i+1} through the alignment between s_{i+1} and r_{j+1}. For the elements of s_{i+1} that have no corresponding element in r_{j+1} (and hence no expressive deviation values), the onset and timing deviation values are interpolated, using the values of neighboring elements.

The adaptation phase is realized by Constructive Adaptation [7], using the quality of segment and performance matches as a guide for searching the space of partial solutions (i.e. best first search). When a complete solution is encountered, this solution is returned, and the adaptation phase is finished.

3 Results

Although we are still in the phase of testing the system, we have some preliminary prediction results. The results were realized using a case base that contained 9 phrases (phrases from the popular jazz songs *Body and Soul*, *Once I Loved*, and *Like Someone in Love*). In turn each phrase is decomposed on the order of 10 segments, i.e. holds 10 cases. Figures 1, 2, and 3 show the predictions for three phrases. Some clear correspondences can be observed between the original and predicted deviations, especially for the onset deviations. Also the predictions

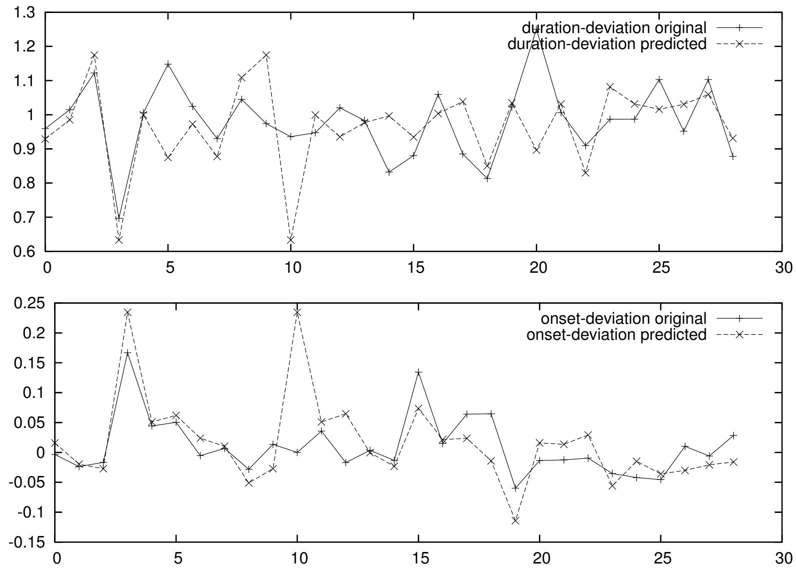

Fig. 2. Prediction of duration deviations (upper plot; measured as the ratio of nominal note duration) and onset deviations (lower plot; measures in metrical beats) for BodyAndSoul-A1 played at 50 bpm

in figures 1, and 2 seem to be more accurate, than the predictions shown in figure 3. We suspect this is due to the unavailability of sufficiently similar melodic fragments in the case base (for the other two phrases, there *are* segments that are rather similar). This can hopefully be solved by increasing the size and diversity of the case base.

4 Conclusions

We have proposed a CBR system for predicting expressive musical performances using a time series prediction approach. The system receives as input a partial performance of a jazz theme, played by a professional musician, and predicts the score deviations for the following incoming notes. Some initial promising results have been produced. Nevertheless, the case base is rather small, which limits the applicability of the system to a wide range of different melodies.

We are now improving the system by incorporating new jazz themes in the case base, as well as extending the adaptation phase to be able to apply not only timing and duration changes, but also apply more sophisticated forms of musical expression, like note ornamentations and consolidation of notes.

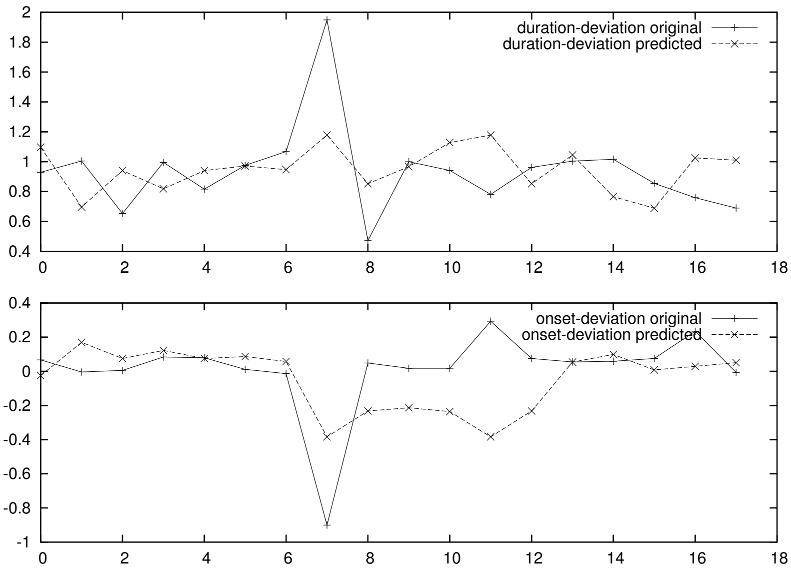

Fig. 3. Prediction of duration deviations (upper plot; measured as the ratio of nominal note duration) and onset deviations (lower plot; measures in metrical beats) for LikeSomeoneInLove-A played at 120 bpm

Acknowledgments

This research has been partially supported by the Spanish Ministry of Science and Technology under the project TIC 2003-07776-C2-02 "CBR-ProMusic: Content-based Music Processing using CBR" and EU-FEDER funds.

References

1. I. Bengtsson and A. Gabrielsson. Methods for analyzing performance of musical rhythm. *Scandinavian Journal of Psychology*, 21:257–268, 1980.
2. P. Desain and H. Honing. Tempo curves considered harmful. In "Time in contemporary musical thought" J. D. Kramer (ed.), Contemporary Music Review. 7(2), 1993.
3. A. Gabrielsson. Expressive intention and performance. In R. Steinberg, editor, *Music and the Mind Machine*, pages 35–47. Springer-Verlag, Berlin, 1995.
4. M. Grachten, J. Ll. Arcos, and R. Lopez de Mantaras. Tempo-express: tempo transformations preserving musical expression. In *Working notes of the IJCAI03 Rencon Workshop*, 2003. Online: *http://shouchan.ei.tuat.ac.jp/~rencon/IJCAI-03/IJCAI_Rencon2003_FINALLAST.pdf*.
5. P.N. Juslin. Communicating emotion in music performance: a review and a theoretical framework. In P.N. Juslin and J.A. Sloboda, editors, *Music and emotion: theory and research*, pages 309–337. Oxford University Press, New York, 2001.

6. E. Narmour. *The Analysis and cognition of basic melodic structures : the implication-realization model*. University of Chicago Press, 1990.
7. E. Plaza and J. Ll. Arcos. Constructive adaptation. In Susan Craw and Alun Preece, editors, *Advances in Case-Based Reasoning*, number 2416 in Lecture Notes in Artificial Intelligence, pages 306–320. Springer-Verlag, 2002.
8. B. H. Repp. Expresive timing in Schumann's "Trämerei": An analysis of performances by graduate student pianists. *Journal of the Acoustical Society of America*, 98(5):2413–2427, 1995.
9. C. E. Seashore. *Psychology of Music*. McGraw-Hill, New York, 1938. (Reprinted 1967 by Dover Publications New York).