A New Approach to Cooperative Performance: A Preliminary Experiment

Ben Curry and Geraint Wiggins
School of Artificial Intelligence
Division of Informatics, University of Edinburgh
80 South Bridge, Edinburgh EH1 1HN, Scotland
Email: {benc,geraint}@dai.ed.ac.uk
Fax: +44-131-650-6516

Abstract

This paper describes a preliminary investigation into the relationship between the high-level structure of music and expressive performance. The experiment aims to determine whether it is possible in principle to develop a system which uses both the structure of a piece and real-time performance data from a human musician to provide an expressive performance.

Keywords: Expression, Performance, Structure, GTTM, Music

1 Introduction

The purpose of this experiment is to investigate the relationship between the high-level structure of a piece of music and the expressive performance of that piece and whether it can be used in automated accompaniment.

This investigation forms part of our research into creating a cooperative performer that, instead of merely tracking a human performer and playing along, has some notion of the structure of a piece and will adjust its performance both according to this structure and the human's performance.

The work presented below has been done with the intention of building a computer system to do the same task. We have been adopting the role of the system whilst we performed this preliminary experiment. The long term aim of our research is to construct a system that will perform all these tasks and then be capable of using the information gathered to produce an expressive performance in the manner described above.

The structure of this paper is as follows: first, we discuss some work related to this experiment. Then we describe the various decisions that were made when we designed the experiment. A description of how we analysed the raw data is given and then we go on to present the results. Finally, we present our conclusions from this preliminary experiment.

2 Related Work

2.1 Score Tracking

Score tracking tries to match events in the performance with events in the score. Roger Dannenberg has made significant contributions to this field (e.g. Dannenberg and Mukaino, 1988 ; Dannenberg, 1993) and has recently developed a system along with Lorin Grubb (1997) that tracks and accompanies a vocal performer using statistical methods. Another recent body of work by Desain et al. (1997) has focussed on using properties of music, such as the relatively strict order of events, to assist with the matching process. None of these techniques use any information about the structure of the music they are matching.

2.2 Expressive Performance

Generating expressive performances is another important field of research. We will mention three recent noteworthy attempts to solve this problem. Arcos et al. (1997) used case based reasoning to generate expressive performances from mechanical ones. The SaxEx system, which they have developed, takes a mechanical performance of a piece of music along with a MIDI file containing the harmonic and melodic information and attempts to find a similar phrase amongst its stored cases. If it finds a match, it applies the same deviations which were used in an expressive performance of the stored case to this new phrase.

Friberg et al. (1997) use a technique which they call "Analysis by Synthesis" to create a set of rules to perform musical "punctuation". This technique involves giving the system a rule which it is believed is used in performing this punctuation task and then gradually improving the rule through a process of using, evaluating and then adjusting the rule.

Gerhard Widmer (1995) used machine learning techniques to produce a set of rules which encapsulated the expressive deviations introduced by a performer. He presents his system with a score and an expressive performance of that score and lets the system discover the rules behind that expressive performance.

2.3 Musical Understanding

Human listeners are capable of deriving more information about a piece of music than is present in the score. They attribute various different structures and relationships to the music which are rarely explicit in the score. Here, we briefly mention two theories and then describe in more detail a third, which we use in the current research.

Eugene Narmour's (1992) theory of analysis is based upon expectation. His theory attempts to encompass "the specific, note-to-note principles by which listeners perceive, structure and comprehend the vast world of melody". The model defines style in terms of bottom-up shapes and top-down structures. The bottom-up shapes are based upon the perception of events and the top-down structures are based upon “complex hierarchical interrelations”.

Recent work by Emilios Cambouropoulos (1998) has highlighted the need for a theory which is based upon “general cognitive and logical principles”. His theory attempts to derive a high-level structure of a musical piece independent of musical style or idiom. Although attractive because it is a very formal theory, it is still in the process of being implemented and so cannot be used for our research.

The authors of The Generative Theory of Tonal Music (GTTM) state its goal to be a “*formal description of the musical intuitions of a listener who is experienced in a musical idiom*” (Lerdahl and Jackendoff, 1983). GTTM attempts to formalise the process of how an “experienced listener” perceives the structure of the piece being heard. The theory itself is split up into four sections which each have three different types of rules within them. The three different types of rules are:

Well-formedness Rules which state what sort of structural descriptions are possible.

Preference Rules which try to select from the possible structures the ones that correspond to what an experienced listener would hear.

Transformational Rules that allow certain “distortions” of the strict structures prescribed by the well-formedness rules.

The theory is broadly divided into two groups of rules. The first describes rhythmic structure which encompasses both the grouping structure and the metrical structure of a piece. The second group is based upon the notion of reduction in which some musical events are structurally less important than others. These three types of rule are stated for each of the four different, but related, structural properties of a piece which are dealt with by GTTM.

Grouping Structure is the segmentation of musical events into groups of similar or related events. These rules try to encapsulate the notion of “chunking” in which a listener groups certain events together whilst hearing the piece.

Metrical Structure models the relative strength and weakness of events at various levels in a metrical hierarchy. It captures the notion of strong and weak beats.

Time-span Reduction identifies pitch events which are perceived as being of greater structural importance at various levels.

Prolongational Reduction identifies events which represent the harmonic movement of the piece. The prolongational reduction deals with issues such as tension, relaxation, continuity and progression.

Although the authors attempt to be thorough and formal throughout the theory, they do not resolve much of the ambiguity that exists through the application of the preference rules. There is little or no ranking of these rules to say which should be preferred over others and this detracts from what was presented as a formal theory.

Criticisms aside, GTTM does provide a very useful and important framework from which we can begin to build autonomous systems that should be able to analyse a piece and construct the appropriate high-level structures, even if multiple possible solutions may be generated.

The aim of this preliminary experiment is to explore the relationship between GTTM and an expressive performance of a piece. If there does appear to be such a relationship, further, more extensive, work is needed to try and discover its mathematical properties. In the meantime, this experiment is expected merely to point towards the existence of a relationship and not expected to provide any sound understanding of what exactly the relationship is.

3 Design

The first stage in the experiment was selecting an appropriate piece of music for analysis. We needed a piece that was both relatively simple but was not too trivial. The piece needed to be a duet which could be performed on a MIDI enabled instrument.

We chose Gabriel Fauré's *Dolly Suite: Berceuse* (1894). It is a relatively slow piece for two pianos which has two mostly distinct parts. The first part is largely responsible for the melody and the second part provides the bass and the harmony. The piece is intended to be performed on one piano with the two parts being played mostly in non-intersecting ranges of the keyboard.

Two pianists were chosen to perform the piece: one a professional pianist with more than ten years of experience, the other an experienced amateur pianist.

After a brief period of rehearsal to get used to the piece and the MIDI enabled piano, three separate performances were recorded. The performances were recorded using an Apple Macintosh SE/30 with EZVision software recording the MIDI data.

4 Analysis

The MIDI data files were transferred to a PC and the two parts were separated using Evolution Audio Lite by Evolution Electronics (UK). A C program modified from Watkinson (1997) was used to extract the relevant information from the midi files (i.e., pitch, onset time, duration and velocity).

For the purposes of this preliminary experiment we chose to analyse just the first 22 bars of the piece. This contained five complete occurrences of the 4 bar theme with some repetition and variation and with the first two bars being silent.

In the case of the first voice, the melody is played by both hands with the notes an octave apart. This meant that we could average both the velocity and timing information of the notes which are played simultaneously. In most cases there was no difference between the onset times or the velocity, but if there was a difference it tended to be very small.

The next step was to calculate the average velocity and the deviations from that during the performance. This was a simple calculation which involved adding all the velocity information together and dividing it by the number of note positions. The relative deviation for each note position was then calculated by subtracting the average from the actual velocity at that point.

The timing information is harder to derive. To investigate the timing information, we need to take into account the fact that we are most interested in the onset time relative to the previous note rather than the absolute onset time within the piece. If we chose the latter option we would find that an initial difference in timing would then put all of the subsequent notes out of synch with our timing scale and give us distorted information. Instead we measured the deviation in relation to when the previous note was performed. This meant that we effectively had to re-calibrate the system at each note.

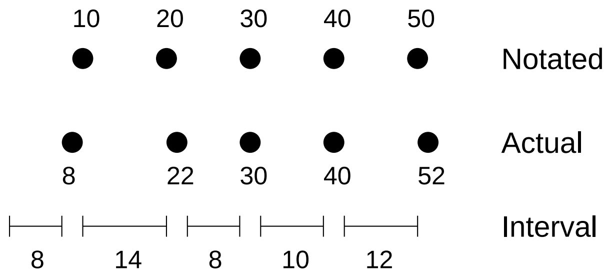

Figure 1: Illustration of re-calibration in action

Shift	0	-2	2	0	0
Notated Time	10	20	30	40	50
Calibrated Time	10	18	32	40	50
Actual Time	8	22	30	40	52
Difference	-2	+4	-2	0	2

Table 1: Re-calibration in action

If we take the example of the performance given in Figure 1 the first row of dots shows when the events should ideally occur given the information from the score. The second row of dots show when the notes were performed. Finally, the interval between the performed events is shown in the last row. The algorithm works as follows (using data from Table 1), we initially set the *shift* to be 0. We then calculate the difference between the time of the event that was performed and the time that was prescribed in the score. In the case of the first event the difference is -2 meaning that the event was performed 2 units of time early. We need to use this difference to calculate when we should expect the next event.

Since the first event was played 2 units early, we should then expect the second event to be performed with the same deviation. So we add -2 to the *notated* time of the next event to get our *calibrated* time. Now when we calculate the difference between the notated time and the performed time we get the correct deviation of +4.

Once we had timing and velocity data for the three performances, we averaged them to try and reduce some of the noise which might be present in the performances. The averaging was done using the deviations not the original data.

In order to investigate the data fully we investigated not only the overall timing and velocity deviations for the first 22 bars as a whole, but also what happened when we took each phrase (4 bars) in turn.

5 Results

The following section is divided into two parts; the first part describes some of the results of the GTTM analysis of the piece, the second presents the data obtained from the experiment and highlights some significant relationships.

5.1 GTTM Analysis

Shown in Table 2 are some of the more interesting features of the piece that helped to decide on the shape of the final structures selected. The first stage in the analysis is to examine the grouping and metrical structures of the piece. Figure 2 shows how these look for the first 4 bars of interest (the first 2 bars are silent). We can see that the tie in bar 4 causes the segmentation at the lowest level. The metrical structure, i.e. the strong and weak beats, were chosen partly due to the relatively long notes that occur at those points and partly to the location of the dynamics. It is this grouping and metrical structure that appears in most of the subsequent phrases.

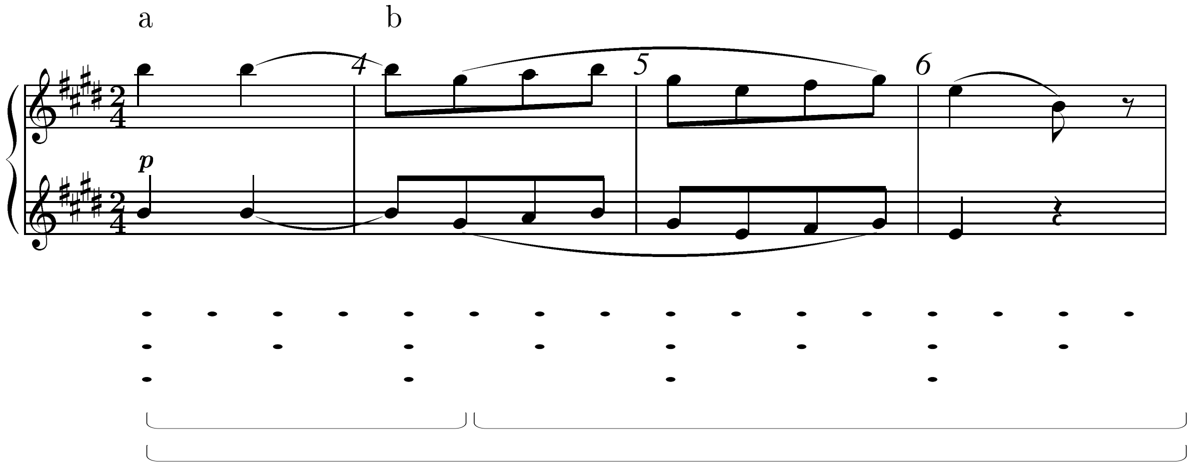

Figure 2: Metrical and Grouping structure for bars 3–6

Point of interest	Location (Bar:Beat)	Indicated by (Figure:Symbol)
Piano	3:1	Fig. 10:*a*
Tie across barline	4:1, 8:1	Fig. 10:*b*,*c*
Crescendo	9:1,2	Fig. 10:*d*
Diminuendo	10:1,2	Fig. 10:*e*
Piano	11:1	Fig. 12:*f*
Modulation	11:1	Fig. 12:*g*
Tie across barline	12:1	Fig. 12:*h*
Crescendo	15:1 – 17:2	Fig. 12:*i*
Tie across barline	16:1	Fig. 12:*j*
Diminuendo	18:1,2	Fig. 12:*k*
Piano	19:1	Fig. 14:*l*
Modulation	19:1	Fig. 14:*m*
Tie across barline	20:1	Fig. 14:*n*

Table 2: Points of interest in the piece

Bars 7–10 (see Figure 3) provide a slightly different grouping due to the modulation that begins in the middle of bar 9. This causes the extra division within the grouping structure which appears at the lowest level. One final point of interest occurs at the cadence at the end of bar 18 and beginning of bar 19. This cadence creates what is referred to as an "overlap" in the grouping structure which means that the first event of bar 19 is shared both with the previous phrase (to provide the cadence) and as the start of the next phrase.

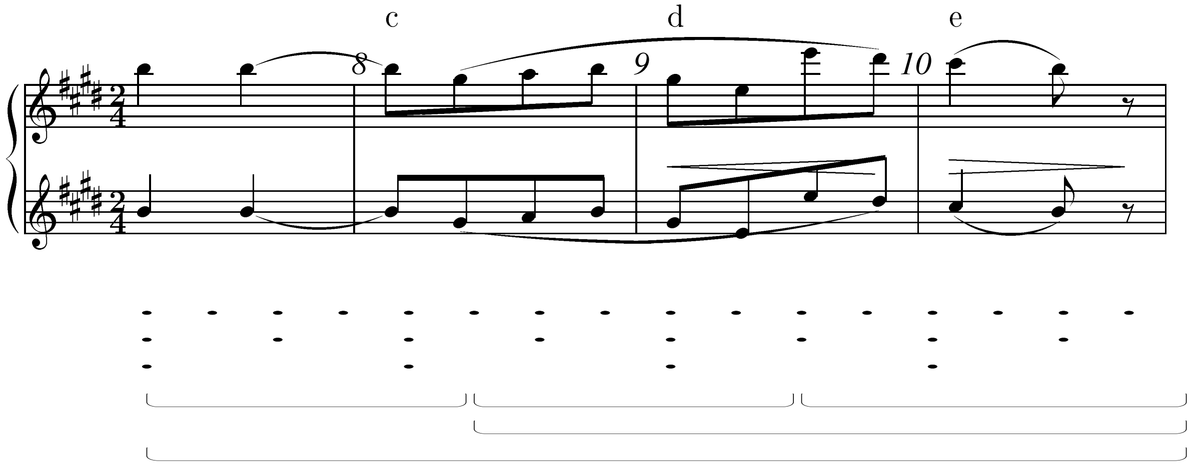

Figure 3: Metrical and Grouping structure for bars 7–10

The next two analyses that were performed are the time-span and prolongational reductions. The time-span reduction is strongly influenced by the metrical and grouping structures that occur at that point in the piece. Figure 4 shows the shape of the time-span analysis that occurs for the first, third and final phrases.

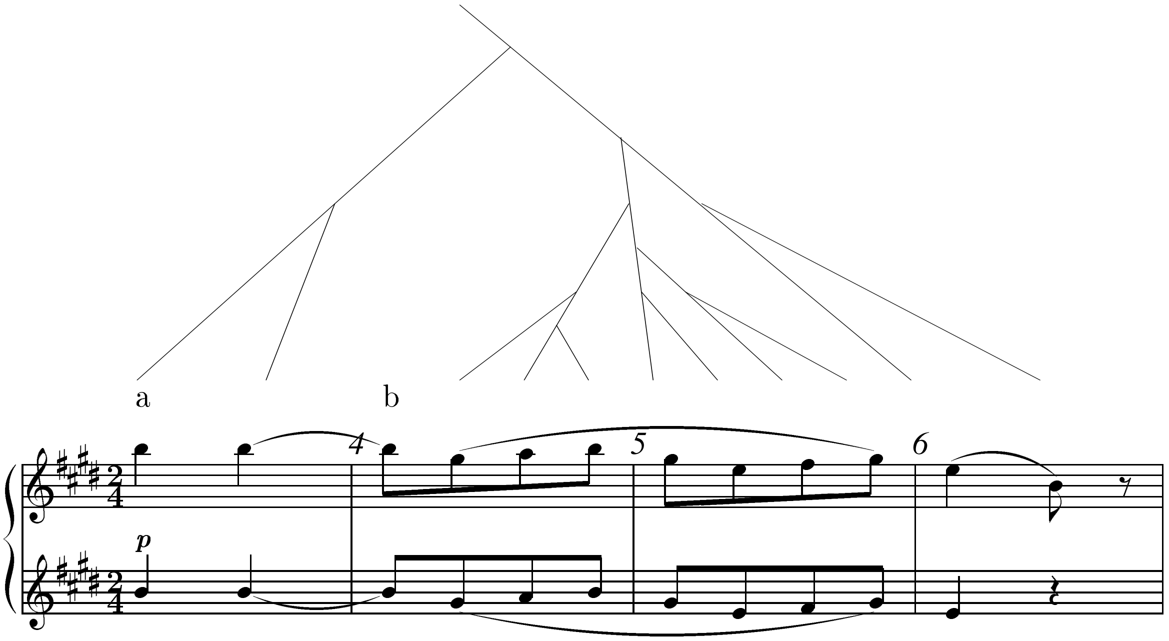

Figure 4: Time-span reduction for bars 3–6

Since the grouping structure is different for the second phrase (due to the beginning modulation) the resulting time-span analysis is also affected (see Figure 5). We can see that the last two notes in the third bar of the phrase (i.e., bar 9) have moved from being elaborations of the first note in that bar to elaborations of the first note in the last bar.

The fourth phrase provides a more complicated time-span analysis due to the cadence which links the fourth and fifth phrases (see Figure 6). This creates a significantly different structure for the final three events (18:1,2 & 19:1). Whereas in the previous time-span reductions the penultimate note of the phrase was highlighted as the most significant, in this case the C# at the end of bar 18 is highlighted.

The prolongational reduction in Figure 7 shows the shape of the analysis that covers all but the fourth phrase. The structure indicates that the phrase moves towards resolution at the penultimate note in the phrase. We can also see that the notes at 4:2,3,4 act to create a minor amount of tension that is resolved at 5:1, similarly notes 5:2,3,4 lead to the stable note at 6:1.

The structure shown in Figure 8 shows how the cadence affects the location of the focus. Again the focus has moved to the final event in the phrase and the event before has become less important.

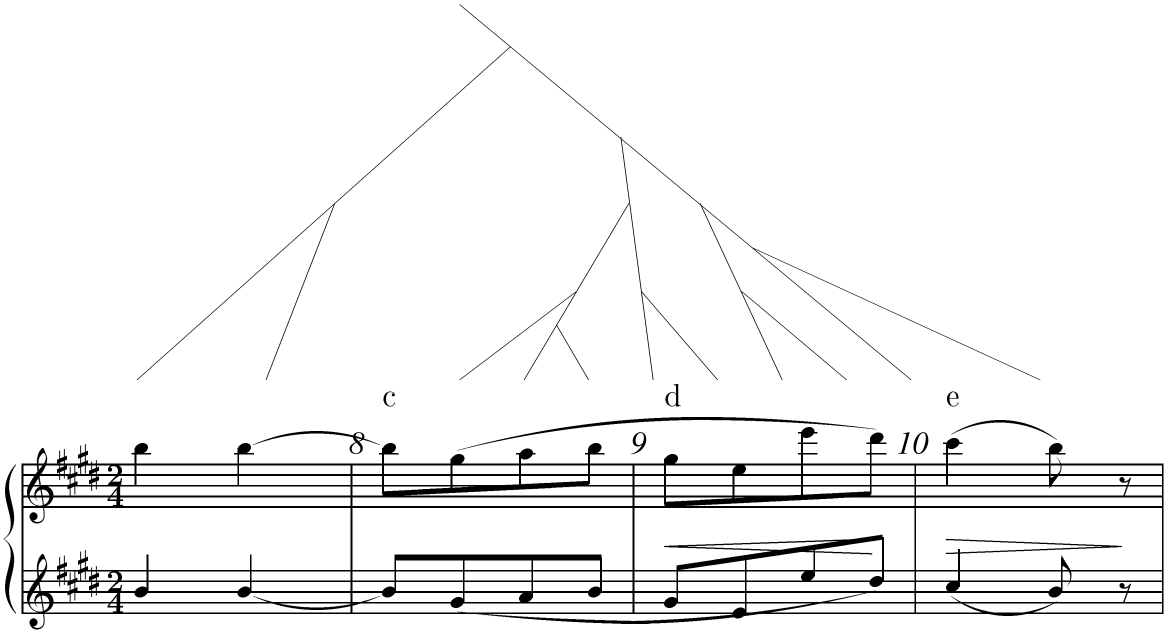

Figure 5: Time-span reduction for bars 7–10

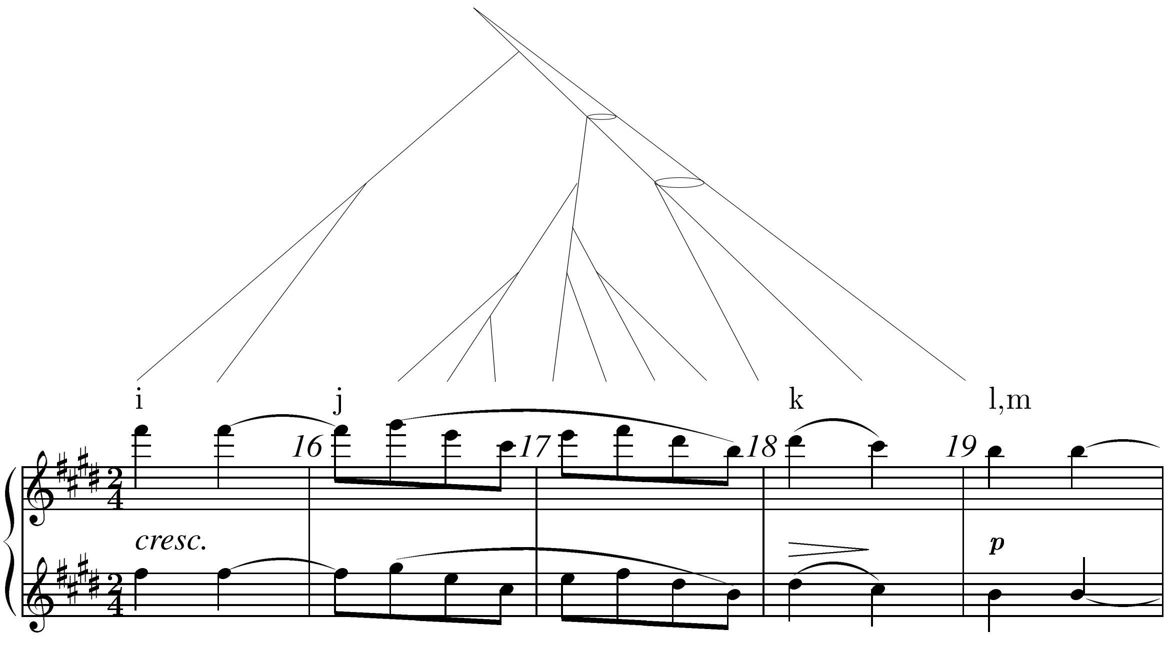

Figure 6: Time-span reduction for bars 15–19

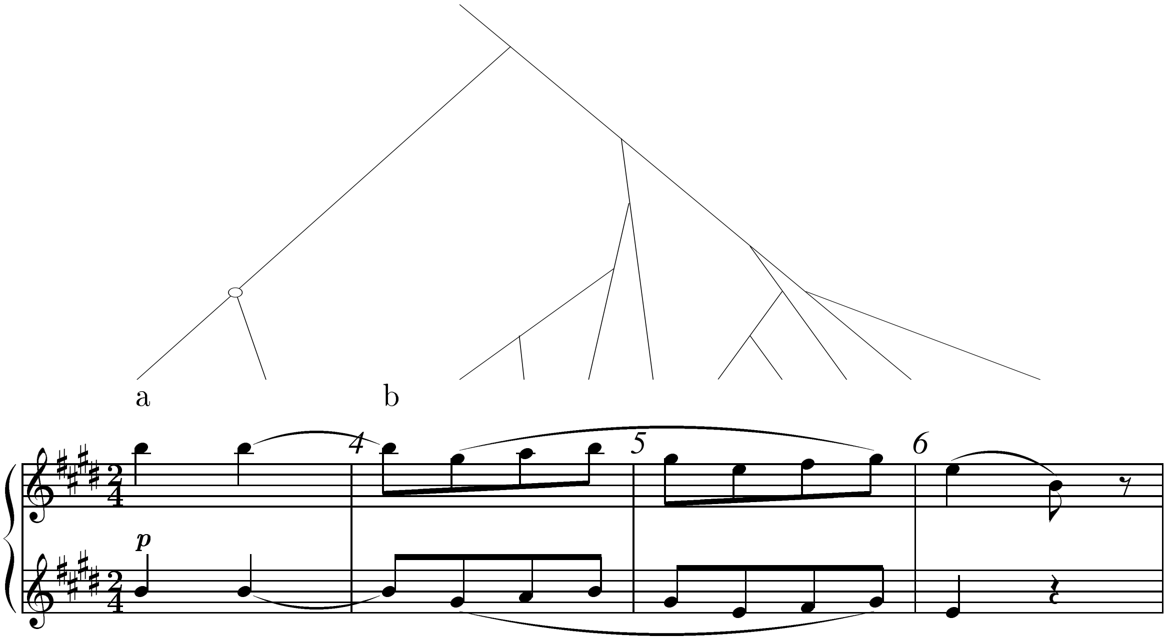

Figure 7: Prolongational reduction for bars 3–6

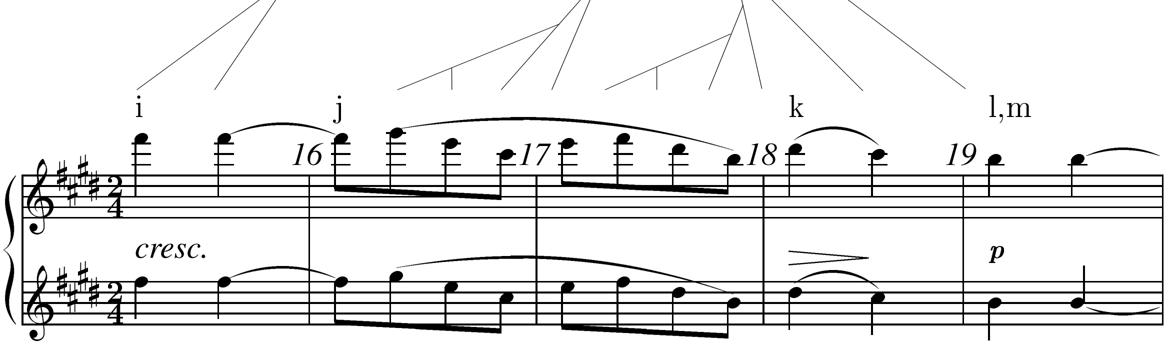

Figure 8: Prolongational reduction for bars 15–19

5.2 Timing Deviations

Before we begin discussing how the structures and timing data interact, we should first point out an interesting, but undesirable, feature of the data collected. Figure 9 shows the timing deviations which took place during all three performances. Although the overall pattern appears to be quite similar, when we inspect the graph more closely we can see that there is quite a lot of deviation between the performances. This deviation is not just in magnitude but in direction meaning that in some performances the same notes were delayed and in others they were played early. This most likely due to the small data set we used and the fact that the piece was not extensively rehearsed by either of the pianists. However, inspection shows that the data is at least usable for a preliminary experiment such as this and we avoid drawing strong conclusions from those parts of the data which are contradictory.

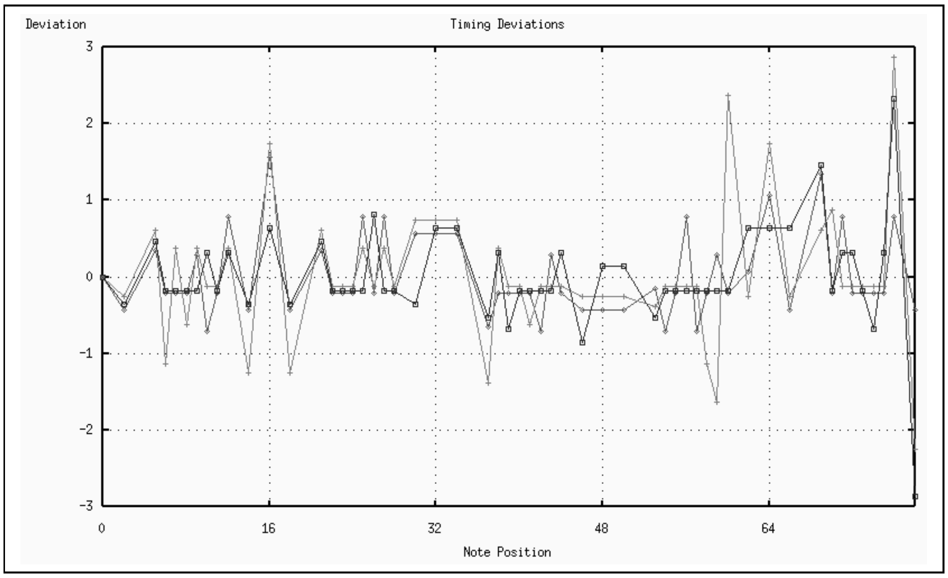

Figure 9: Timing Deviations for all three performances (based upon the duration of the first 22 bars)

5.2.1 Bars 3–10

The score and timing information for bars 3–10 are shown in Figures 10 and 11. We can see that the tie across the barline (marked by b) causes the third note of the phrase to be played significantly earlier than intended. There are then some smaller deviations fluctuating around the melodic rises in bars 4 and 6. The E in bar 6 receives emphasis by the pianist playing it early and then the B is slowed to compensate for this variation. The significance of the E is highlighted both by the time-span analysis and by the prolongational analysis given in Figures 4 and 7.

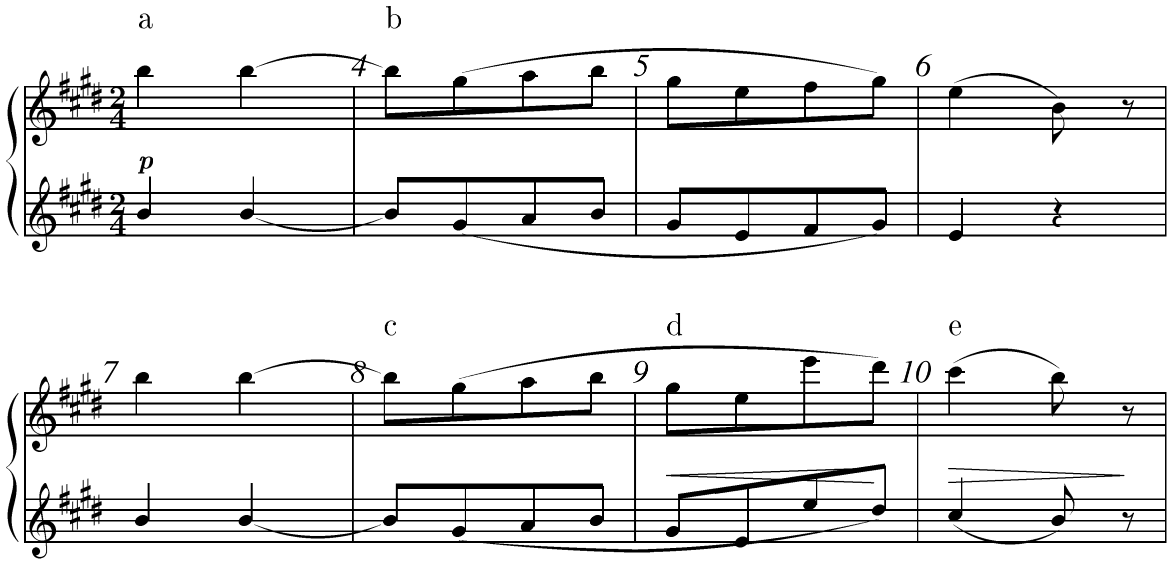

Figure 10: Prima part for bars 3 – 10

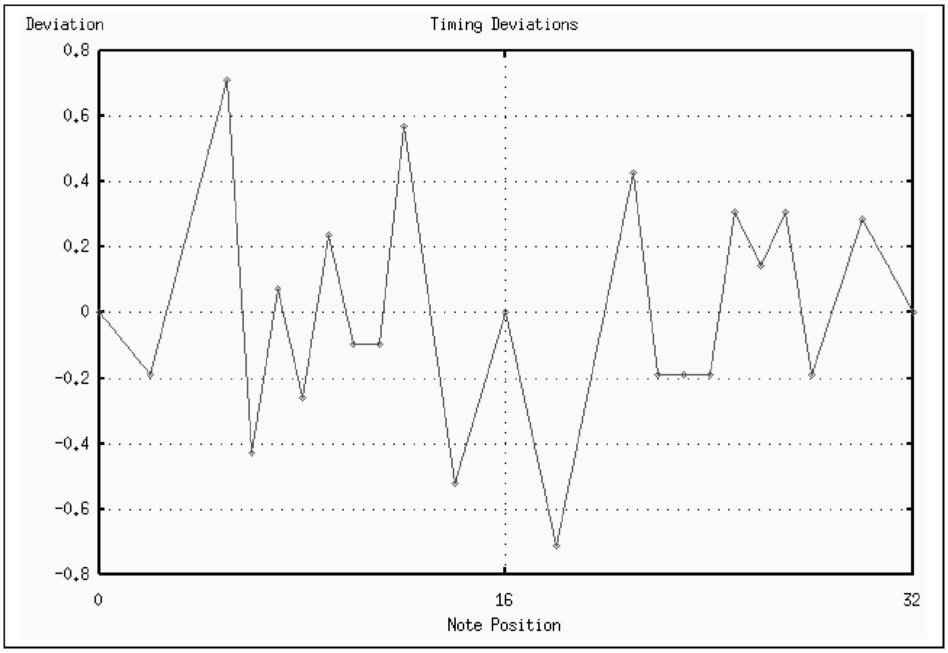

Figure 11: Timing Deviations for bars 3–10

Bars 7 and 8 show a similar pattern of timing deviations leading up to and including the tie (marked by c). However, the second half of the phrase shows a different timing profile to the first. This is probably due to a number of things; the extra grouping boundary caused by the start of the modulation (see Figure 3), the crescendo and the diminuendo (d & e). The results of these differences are clearly visible when the time-span structure in Figure 4 is compared with Figure 5.

5.2.2 Bars 11–18

The first feature of interest in the timing deviation in Figure 13 is the fact that the timing deviations happen in mostly the opposite direction to the deviations in the first phrase. Our tentative explanation for this surprising feature is that the performer is emphasising the results of the modulation and is asserting, particularly with the second event, that we are now in the dominant. This deviation then affects the rest of the phrase.

Figure 12: Prima part for bars 11 – 18

The start of bar 15 marks the beginning of a crescendo that leads towards the cadence at the end the phrase. We also have a reversal of the slope in the centre of the phrase so that the melodic rises have become melodic falls. If we look at the time-span reduction for this part (Figure 6) we can see that the middle two bars are viewed as elaborations of the cadence. The events in the middle bars gradually slow down until, when we reach the start of the cadence, the speed then returns to “normal” to help emphasise the cadence. Thus, again, the performance may be said to correspond with the GTTM tree.

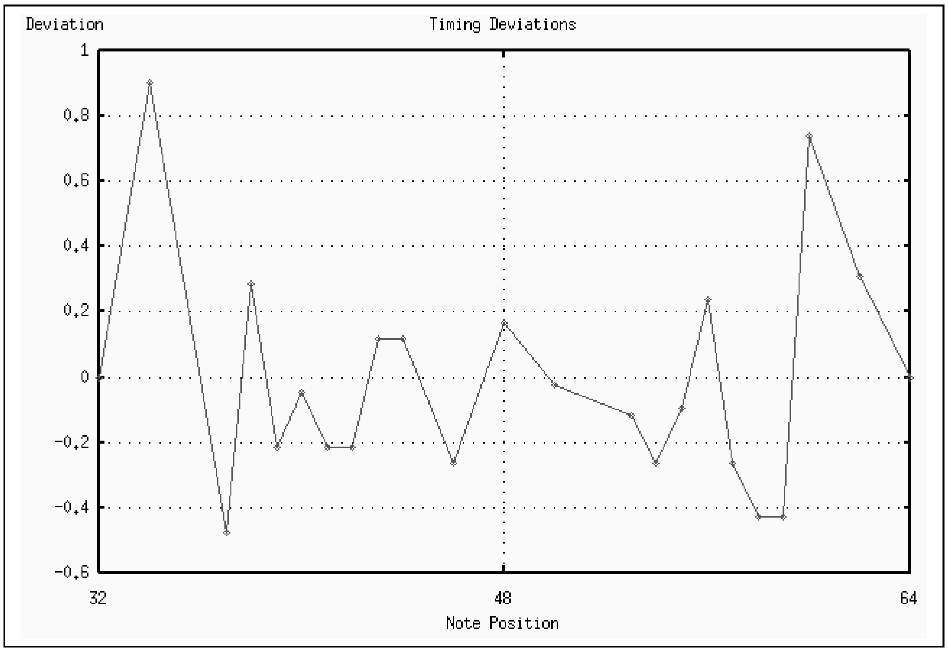

Figure 13: Timing Deviations for bars 11–18

5.2.3 Bars 19–22

The phrase in this section of the piece is an exact copy of the first phrase. We can see the familiar delay of the second note and then the early performance of the third note after the tie (shown as n in Figure 14). The third bar of the phrase continuously slows down until we reach the fourth bar at which point the first note is played very early. Finally, we have a large delay near the end of the phrase which was probably due to the fact that the page needs to be turned at this point. Despite the fact that the first event of this phrase forms part of the cadence of the previous phrase, the structure of this phrase is unaltered as we view the event as being shared between the two phrases.

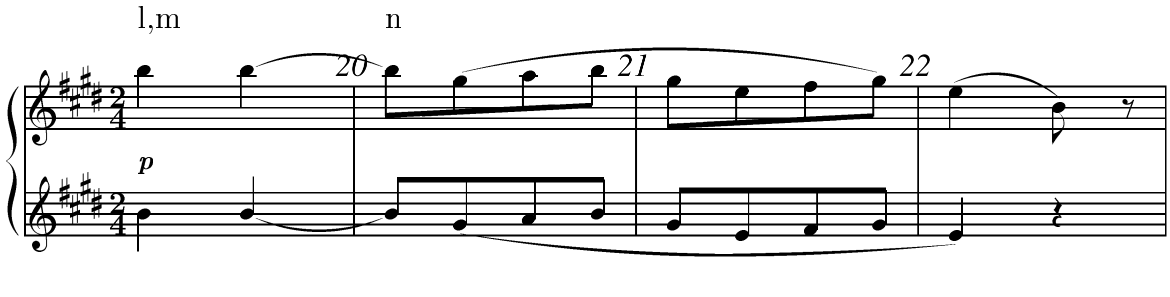

Figure 14: Prima part for bars 19 – 22

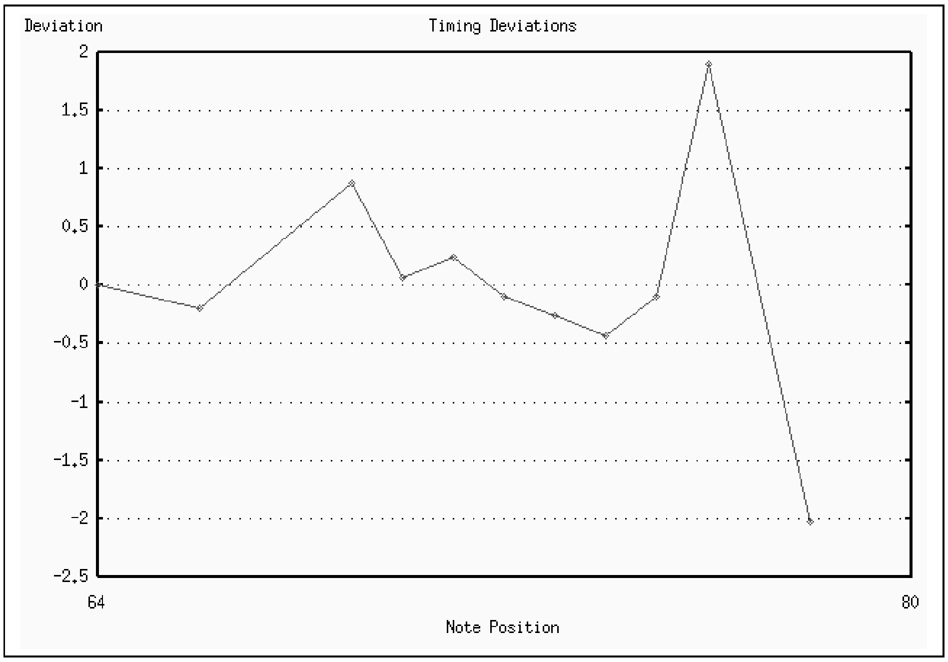

Figure 15: Timing Deviations for bars 19–22

6 Conclusions

The aim of this preliminary experiment was to investigate the relationship between the structure of a piece of music and an expressive performance of that piece. The results suggest that there is indeed a relationship there to be explored. However, looking at our small sample, we can see that there is also quite a lot of “expression” that is not directly accounted for in the structure. Perhaps this is a feature of the data set used, or perhaps we will need to model some aspects of the performers (Parncutt 1997) as well as the piece.

The results of this preliminary experiment do encourage further work to be done in this area. However, when it comes to trying to quantify the relationship between the performance and the structure it will be important to have a larger sample of data to work from. With a large data set, we expect to be able to identify which aspects of the performance are consistent in the majority of the performances and which aspects cannot be accounted for by our research.

Acknowledgements

Ben Curry is supported by UK EPSRC postgraduate studentship 97305827.

References

Arcos, J. L., de Mántaras, R. L., and Serra, X. (1997). Saxex: a case-based reasoning system for generating expressive musical performances. In Cook, P. R., editor, *Proceedings of the International Computer Music Conference*, pages 329–336. Computer Music Association.

Cambouropoulos, E. (1998). *Towards a General Theory of Musical Structure*. PhD thesis, Faculty of Music, University of Edinburgh.

Dannenberg, R. B. (1993). Music understanding by computer. In *IAKTA/LIST International Workshop on Knowledge Technology in the Arts Proceedings*, pages 41–56.

Dannenberg, R. B. and Mukaino, H. (1988). New techniques for enhanced quality of computer accompaniment. In *Proceedings of the International Computer Music Conference*, pages 243–249.

Desain, P., Honing, H., and Heijink, H. (1997). Robust score-performance matching: Taking advantage of structural information. In Cook, P. R., editor, *Proceedings of the International Computer Music Conference*, pages 337–340. Computer Music Association.

Fauré, G. (1894). *Dolly Pour Quatre Mains (op. 56)*. J. Hamelle (Editor), Boulevard Malesherbes, Paris.

Friberg, A., Frydén, L., and Sundberg, J. (1997). A rule for automatic musical punctuation of melodies. In Behne, K.-E., Deliège, I., Gabrielsson, A., and Sloboda, J., editors, *Proceedings of the European Society for the Cognitive sciences Of Music Conference*, pages 719–723. ESCOM.

Grubb, L. and Dannenberg, R. B. (1997). A stochastic method of tracking a vocal performer. In Cook, P. R., editor, *Proceedings of the International Computer Music Conference*, pages 301–308. Computer Music Association.

Lerdahl, F. and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. MIT Press.

Narmour, E. (1992). *The analysis and cognition of melodic complexity: the implication-realization model*. University of Chicago Press.

Parncutt, R. (1997). Modeling piano performance: Physics and cognition of a virtual pianist. In Cook, P. R., editor, *Proceedings of the International Computer Music Conference*, pages 15–18. Computer Music Association.

Watkinson, S. P. (1997). Induction of musical syntax. Master's thesis, Department of Artificial Intelligence, University of Edinburgh.

Widmer, G. (1995). Modeling the rational basis of musical expression. *Computer Music Journal*, 19(2):76–96.