Recent Work In Music Understanding[1]

Roger B. Dannenberg

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213 USA
Email: dannenberg@cs.cmu.edu

Abstract

Interaction with computers in musical performances is very much limited by a lack of music understanding by computers. If computers do not understand musical structures such as rhythmic units, chords, keys, and phrases, then interaction with computers will necessarily be difficult and cumbersome. Research into Music Understanding by computer aims to raise the level of human computer interaction in musical tasks including live music performance.

1. Introduction

Music Understanding is the recognition or identification of structure and pattern in musical information. Music Understanding is important because it opens the doors to high-level interaction between musicians and computers. Research at Carnegie Mellon University has led to a number of interesting developments which are summarized here.

The following sections briefly describe five projects related to Music Understanding. The first two describe computer accompaniment systems, which listen to live musical performances and synchronize pre-stored computer music accompaniments. The third project is a system for analyzing the harmonic and rhythmic content of an improvised solo in order to follow a jazz improvisation. This work led to further investigations of the ‘‘foot-tapping’’ or beat detection problem. Music understanding and intelligence have been applied in the Piano Tutor, which is the fifth project described.

One of the results of this work is a better appreciation of the difficulty of these tasks. Some directions for future research and conclusions are given following the descriptions of Music Understanding projects.

2. Computer Accompaniment

When human musicians perform together, they typically listen to one another and synchronize their music according to a musical score. In contrast, most performances with computers require that humans follow a computer-based ‘‘sequencer’’ that has no listening abilities whatsoever. An alternative approach is to build a computer system that can listen to the human musician and synchronize its performance. I call this task *computer*

---

[1]Dannenberg, ‘‘Recent Work in Music Understanding,’’ in ***Proceedings of the 11th Annual Symposium on Small Computers in the Arts***, Philadelphia: SCAN, (November 1991), pp. 9-14.

accompaniment.

Computer accompaniment has been described in the literature [3, 2, 5, 9, 7], so this section will only sketch our approach to the problem. As shown in Figure 1, a computer accompaniment system consists of several stages. In the first stage, the Listening Task detects note onsets and pitches in the Solo Performance. This data is sent to the Matcher, which compares the live human performance with a stored score. It is assumed that the intended performance is completely notated. This assumption (and the score) gives the computer a great deal of information. (Sections 4 and 5 discuss systems where this assumption is not made.)

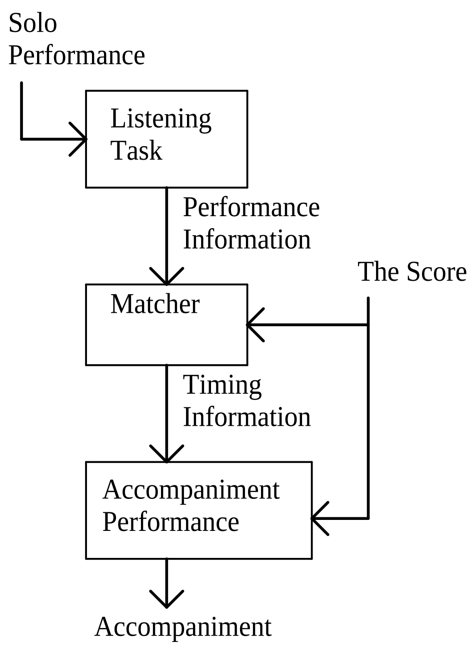

Figure 1: The correspondence between a score and a performance.

The result of the Matcher is an indication of where and when the performance corresponds to the score. This information can be used to estimate the performer's tempo and current location in the score. This in turn can be used to perform an accompaniment, which is also stored in a score (no attempt is made to compose or improvise accompaniment in these systems). The output consists of real-time control information for a synthesizer.

One of the problems of this Accompaniment Performance stage (see Figure 1) is performing the accompaniment in a musical manner even when the performer is missing notes and changing tempo. The Accompaniment Performance stage uses a number of rules about musical performance that help it to respond appropriately and musically when new information is received from the matcher. As a result, the computer accompaniment system performs in a fairly autonomous manner; it is guided by the human performance, but quite capable of performing on its own as well.

The most advanced accompaniment system to date can also handle a small degree of improvisation in the form of trills, grace notes, and glissandi, which do not match up note-for-note with the score [5]. This system has a number of features for controlling the accompaniment, for example limiting the range of tempo adjustment or ignoring input during a steady-rhythm passage. This system has been used by professional musicians in performances.

3. Polyphonic Accompaniment

The first accompaniment systems only worked with monophonic input, that is, input without chords or overlapping notes. This is of course a major drawback for keyboard performers, so a new system was developed for polyphonic accompaniment. Referring again to Figure 1, it can be seen that the challenge in making a polyphonic system is developing a matcher that can match polyphonic performances to polyphonic scores. Two matchers were developed for this purpose [2]. The resulting matchers work quite well and allow accurate following even in the presence of many performance errors.

4. Following Improvisations

Computer accompaniment is based upon traditional (Western) music making in which a composition governs what notes are played by the performers. In jazz and other improvisational styles, this information is not available. However, improvisations are not without structure. In particular, jazz improvisations usually have a tempo, measures, and chord sequences among other features.

Working with Bernard Mont-Reynaud, I developed a computer system that listens to a blues improvisation played on trumpet. The goal of the program is to accompany the trumpet with a rhythm section of synthesized bass, drums, and piano. This requires that the program understand a fair amount of structure in the solo part. Notice that in this system, the rhythm section will listen to and follow the soloist rather than the other way around.

‘‘The Blues’’ is a sequence of chords that repeats every 12 measures, where a measure is a musical unit of 4 downbeats. The solo harmonizes with the chord sequence, so the solo part gives some indication of the underlying chord progression. Certain pitches are more likely to occur in combination with one chord than another. However, any note in isolation could occur almost anywhere in the 12 measures, so the problem is to infer a precise location from a large number of ambiguous indicators.

Our tempo-tracking software is based on the idea that note onsets typically occur on either an upbeat or a downbeat. Once a tempo is established, the system predicts where note onsets will occur and compares these predictions to the actual performance data [8]. When performed notes correspond closely to predictions, the current notion of tempo and where beats occur in time is adjusted slightly to make the predictions even better. In this way, a slowly wandering performance tempo will tend to ‘‘pull’’ the estimated tempo, and tracking occurs.

Once the tempo and beat locations are determined, a statistical approach is used to find how chord changes relate to the solo. A table of probabilities is used. The table lists the likelihood that a given pitch will occur on a given downbeat or upbeat (these will be called simply ‘‘beats’’). Each column of the table corresponds to one of 96 downbeats and upbeats, and each row corresponds to one of 12 possible pitches (octaves are ignored), so the table has 1152 entries. Suppose the first note of the solo is an F. Then the F row of the table gives an estimate of the likelihood that the solo started at each of 96 possible beats in the 12-measure blues. Now suppose the next beat of the solo is a G. The G row of the table gives the probability that the G was played at each possible location. Now, taking the F probability of the first column times the G probability of the second column gives a combined likelihood estimate of the combination F-G being played at the beginning of the 12 measures. Similarly, taking the F-G probabilities from some other pair of columns gives a likelihood estimate that the F-G combination was played elsewhere.

Extending this process, we can get likelihood estimates for the entire solo starting on any beat. Figure 2 shows a graph of this likelihood estimate. (See [4] for more details.) It is interesting that the curve shows a 4-measure periodicity that reflects the fact that 12-measure blues has 3 somewhat similar 4-measure phrases. The peak at zero and at 96 is the same peak (the graph repeats itself every 12 measures or 96 beats). This peak correctly indicates the most likely starting point for the 12-measure blues chord progression.

A real-time implementation of this system demonstrates some ability to recognize and follow a blues improvisation. However, the program requires a full 12 measures of performance data before a location estimate can be made, and the system is not nearly as good as humans at recognition. This experience led to further research on the beat tracking, or ‘‘foot tapping’’ problem, which is described in the next section.

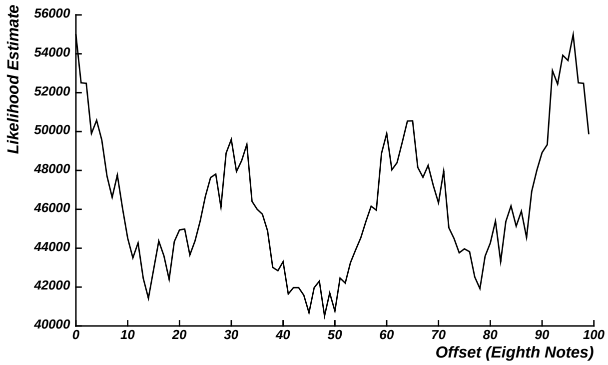

Figure 2: Likelihood estimates (arbitrary units) of the solo starting at different offsets in a 12-bar blues progression.

5. Rhythm Understanding

A fundamental musical skill is the ability to recognize a pulse or beat in a music performance. One approach to this ‘‘foot tapping’’ problem was describe in the previous section. In this approach, an estimate of beat location and tempo is adjusted to fit incoming data points. Unfortunately, this approach fails quite often, especially if there are sudden tempo changes. Once a failure occurs, it is hard to recover because adjustments tend to be made due to random coincidences between anticipated beats and actual note onsets. The more responsive the tracker is, the more likely it is to become confused.

Paul Allen and I developed a new approach to solve problems with earlier trackers [1]. Our basic new idea is that trackers are thrown off by input that is misleading and should be ignored. Suppose at each decision point, the tracker made two choices instead of one. For several beats, the consequences of each decision could be carried out, and then the result that seems best could be retained and the other one discarded. We implemented a system that considers two or three rhythmic interpretations of each incoming note onset. At any time, the beat tracker is considering tens or hundreds of alternative ways of ‘‘tapping its foot’’ to the music.

Some interpretations make more musical sense than others. For example, an interpretation where the beat is steady is generally preferred as is an interpretation that requires only simple rhythms. Based on a musical evaluation, interpretations are constantly pruned from the search to make room for better ones. Figure 3 illustrates how interpretations, represented by circles, give rise to multiple interpretations of new events (each new event corresponds to a new row) or are pruned from the search (represented by crosses).

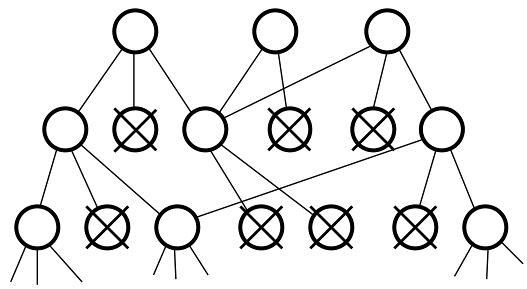

Figure 3: Three levels of search.

A real-time implementation of this algorithm shows that the technique is successful in recovering from bad decisions that would have thrown off earlier trackers. However, the system still has difficulty in separating good interpretations from bad ones, and by considering so many interpretations, it is hard not to prune the correct one occasionally. The performance is very dependent upon the musical input, making evaluation difficult, and more evaluation is needed to better characterize the limitations of this approach.

6. The Piano Tutor

The Piano Tutor is an intelligent computer system for teaching beginners to play the piano [6]. The Piano Tutor makes extensive use of music understanding to support an instructional dialog with the student. In a typical interaction, the Piano Tutor delivers a multimedia presentation to the student and asks the student to perform an exercise. The student performs, but usually makes a mistake. The Piano Tutor corrects the student and asks for another attempt. This continues until the student masters the exercise. Then the Piano Tutor selects new material for the student and the interaction cycle repeats.

Music understanding takes place on three levels. First, the Piano Tutor uses computer accompaniment technology to follow student performances. The Piano Tutor can play musical accompaniments to student performances and also turn pages of music on a computer graphics display. Second, student performances are analyzed by the Piano Tutor to determine if the student is having problems, and if so, the likely cause of the problem. For example, a duration error might be accounted for by a misunderstanding of ties, a problem keeping a steady tempo, forgetting to release a note at the beginning of a rest, or any number of other possibilities. The Piano Tutor looks for the most significant error (in a pedagogical sense), finds an explanation for the error, and then computes an appropriate remediation for the student. The third level of understanding has to do with modeling the student as a developing musician. The Piano Tutor keeps track of what skills the student has mastered and uses its model of student progress to select appropriate lessons for the student. As the student masters new material, the model is updated. As a consequence, lessons are tailored to the individual needs of the student.

7. Summary and Conclusions

Music understanding requires the recognition of pattern and structure in music. I have presented an overview of five projects in music understanding, all conducted at Carnegie Mellon University. Some of the projects, such as computer accompaniment and the Piano Tutor have produced surprisingly good results. Others, such as following improvisations and ‘‘foot tapping’’ show how difficult more general problems of music understanding can be.

Many other tasks are still open problems to be tackled. These include following the performance of an ensemble, with or without a score, integrating pitch information in the rhythm understanding task, following vocal music, and the use of learning to improve the performance of various music understanding systems.

There are many reasons for continuing research in music understanding. Computer music system interfaces can be improved if they can deal with musical structures and commands at a high level. Music understanding can shed light on human cognition, and there is much related research in the area of Music Psychology. Finally, music understanding can have a great impact on music theory and music formalisms. Music understanding is still a new field, and one can expect many breakthroughs in the future.

8. Acknowledgments

This paper is based on a talk prepared for the International Wenner-Gren Symposium on Music, Language, Speech, and Brain. Many colleagues have contributed to the projects described here, including Joshua Bloch, Bernard Mont-Reynaud, and Paul Allen. Marta Sanchez and Annabelle Joseph conceived of the Piano Tutor and are co-principal investigators on the project, which also benefited from the work of Peter Capell, Ron Saul, Robert Joseph, John Maloney, and Hal Mukaino.

This work has been made possible largely through the Carnegie Mellon University School of Computer Science and was partially supported by Yamaha (computer accompaniment) and the Markle Foundation (the Piano Tutor).

References

[1] Allen, P. E. and R. B. Dannenberg. Tracking Musical Beats in Real Time. In S. Arnold and G. Hair (editor), *ICMC Glasgow 1990 Proceedings*, pages 140-143. International Computer Music Association, 1990.

[2] Bloch, J. J. and R. B. Dannenberg. Real-Time Computer Accompaniment of Keyboard Performances. In B. Truax (editor), *Proceedings of the International Computer Music Conference 1985*, pages 279-290. International Computer Music Association, 1985.

[3] Dannenberg, R. B. An On-Line Algorithm for Real-Time Accompaniment. In W. Buxton (editor), *Proceedings of the International Computer Music Conference 1984*, pages 193-198. International Computer Music Association, 1984.

[4] Dannenberg, R. B. and B. Mont-Reynaud. Following an Improvisation in Real Time. In J. Beauchamp (editor), *Proceedings of the 1987 International Computer Music Conference,* pages 241-248. International Computer Music Association, San Francisco, 1987.

[5] Dannenberg, R. B. and H. Mukaino. New Techniques for Enhanced Quality of Computer Accompaniment. In C. Lischka and J. Fritsch (editor), *Proceedings of the 14th International Computer Music Conference*, pages 243-249. International Computer Music Association, San Francisco, 1988.

[6] Dannenberg, R. B., M. Sanchez, A. Joseph, P. Capell, R. Joseph, and R. Saul. A Computer-Based Multi-Media Tutor for Beginning Piano Students. *Interface* 19(2-3):155-73, 1990.

[7] Lifton, J. Some Technical and Aesthetic Considerations in Software for Live Interactive Performance. In B. Truax (editor), *Proceedings of the International Computer Music Conference 1985*, pages 303-306. International Computer Music Association, 1985.

[8] Longuet-Higgins, H. C. and C. S. Lee. The Perception of Musical Rhythms. *Perception* 11:115-128, 1982.

[9] Vercoe, B. and M. Puckette. Synthetic Rehearsal: Training the Synthetic Performer. In B. Truax (editor), *Proceedings of the International Computer Music Conference 1985*, pages 275-278. International Computer Music Association, 1985.