Expressive Performance of Polyphonic Piano Music using Hierarchical Structure

Bastiaan J. van der Weij
5922151

Bachelor thesis
Credits: 15 EC

Bachelor Opleiding Kunstmatige Intelligentie

University of Amsterdam
Faculty of Science
Science Park 904
1098 XH Amsterdam

Supervisor
Prof. dr. ir. R.J.H. Scha

Institute for Language and Logic
Faculty of Science
University of Amsterdam
Science Park 904
1098 XH Amsterdam

June 24th, 2010

Abstract

Both machine learning and rule based techniques have been extensively applied to performance rendering. However, relatively few systems make explicit use of musical structure, combined with machine learning. This paper introduces a performance rendering approach that tries to learn expression exclusively at a structural level. The approach attempts to to be complementary to approaches that learn expression at note level. Musical structure is extracted automatically. A dataset containing performances aligned to scores is then used to learn how structure relates to expression. Finally, the model can be applied to a a new score to produce an expressive performance of the score.

Keywords: performance rendering, musical structure, constituent structure, polyphonic piano music

1 Introduction

In Western art music tradition it is customary that composers write down their compositions in scores. The task of a performer is to some extend to accurately reproduce the score, however, a perfect reproduction of a score generally sounds robotic and unpleasant. What makes a performance appealing is that the performer deviates from the score, altering timing, articulation and loudness, creating an expressive performance.

Given a hypothetical perfect synthesizer, performing music with computers is a trivial task. However *expressively* performing music is not and much research has focussed on this issue. Computers are widely used in music production. Since support for automatic expression is mostly absent or very poor in music editing software, some genres of music have evolved to sound acceptable even without expression. If automatic expression better in this software, computer music production could greatly widen its field of application.

Some music transcription Performance rendering

YQX is state of the art

Rencon

1.1 Motivation

The YQX system, as widmer admitted, tended to sometimes produce nervous sounding changes in expression. He presents two extensions that seek to generate smoother performances. The problem with these extensions is, as widmer admits himself, that the increased smoothness comes at the expense of expressitivity. To counter this, he adds three explicit rules that he uses to postprocess the performances. We think the reason that Widmer stumbled upon this tradeoff between expressiveness and nervousness because the nervousness is inherent to the way performances are generated at note-level. The system simply misses one component of expression, namely structure level expression. (that Widmer had to artificially introduce using rules.) ?

Structure seems to be directly related to expression and it seems logical that a performance rendering system should have some notion of usical structure. The system presented here will use structure exclusively to generate performances and will completely ignore note level expression. The expressiveness of the performances it generates is therefore limited. In section XX we will describe

how this system could be integrated with a note level performance rendering system to hopefully produce some variant of YQX that doesn't need explicit rules to produce expressive performances.

2 Approach

2.1 Musical structure

When listening to music, the listener's musical intuition assigns a certain hierarchical structure to the music: Notes make up phrases, phrases make up themes and themes make up a piece. In a performance, this structure may be accentuated in different ways. Accentuation happens at different levels, at note level performers may slow down at the end a phrase or introduce small pauses in the performance at phrase transitions. At constituent level one phrase may be played very loud, fast or staccato, while the next may be played slow, soft and legato.

To formally describe musical structure, we can look at music in a way similar to the way we look at natural language processing(NLP). In this analogy we see a piece of music as a sentence, which consists of constituents that individually can be made of constituents as well. We can recursively subdivide constituents in more constituents until we reach a terminal symbol. In NLP this may be a word, in music, this may be a note. We could represent musical structure as a parse tree. This paradigm corresponds to the intuition that a melody is not simply a sequence of notes but that notes form phrases and we can look at phrases at different levels.

It must be noted that the resulting parse tree can be highly ambiguous and even experienced listeners may not agree on the correct parsetree of a piece. Quite often there may simply be more than one parse tree that makes musical sense. This should not be a problem for a performance rendering system: different expressive interpretations of one piece can be very diverse and still be accepted as sensible interpretations of the piece. As long as the parse tree does make at least some musical sense, a performance rendering system should be able to use it.

Although the YQX does have some notion of structure[1], it still only looks at note level expression. The authors admit that the first simple version of the system "tended to sometimes produce unstable, 'nervous' sounding performances". We see this as a symptom of a note level expression based system.

In this thesis, we propose a structure based performance rendering (SBPR) system. The system presented here ignores note level expression. Instead we will only try to predict constituent level expression. The assumption is that this kind of expression really exists in performances and that is different and independend from note level structure. We think that a constituent level system also corresponds better to how actual human performers play music. A performance redering system that only predicts note level expression would have rather meaningless fluctuations in tempo and dynamics as it does not have a notion of constituent level expression.

The system will be similar to YQX in a number of ways, but instead of predicting expression per note, it will predict expression per constituent. Every

[1] One of the note features is distance to nearest point of closure

constituent will be played with consistend expression. A number of score features defined per constituent will be used to predict epressive parameters of the constituent.

The simplification of ignoring note level expression is of course unjustified and severly limits the expressiveness of the performances that can be generated. However, if succesful, the resulting performances will clearly demonstrate the phenomenon of constituent level expression. It should be possible to extend the system to incorporate note level expression, section **??** will further explore this possibility.

The succes of a structure based performance rendering (SBPR) system depends largely on two factors. The ability to generate musically meaningful parse trees of a piece and the ability to accurately characterize the individual constituents and their relations with other constituents. The following two sections address these issues.

3 The Delta Framework

In his Phd thesis **??**, Markwin van der Berg introduces a formal way of parsing music into parsetrees: the delta framework. He relates his work to the work of Lerdahl and Jackendoff **??** but claims to have found a more general approach.

The delta framework is based on the intuition that differences between notes indicate splits between constituents. The higher the difference, the higher the level of split in the parse tree (where the root note is at the highest level). Van der Berg proposes a delta rule that converts a set of differences, or deltas, between notes into a parsetree following this intuition.

The differences between notes are defined as the difference in value of a certain note feature. More formally, we can look at a piece of music as a sequence of notes, ordered by onset time:

$$M = n_i, n_{i+1}, \cdots, n_j$$

We can assign features to these notes.

3.1 Extracting structure

- We only need a segmentation for now
- Using the deltaframework to define segmentation

4 Constituent Features

We can now convert a piece of music into a serie of constituents. These constituents will be used to predict expression so we must be able to charaterize them in a way correlates with the way they are performed. Analogous to YQX we are looking for the *context* of the constituent.

(Clearly, expressive markings in the score play a role in how the constituent should be played. If the segmentation is specific enough there will hopefully be at most expressive marking within each constituent.)?

5 Targets

- Use average expression parameters per constituent

6 Performance Rendering

- Extract melody
- Extract scorefeatures from melody
- lookup notes in deviation, ignore missing notes
- Extract expressive and non-expressive melody and extract expression features
- Perform notes other than melody notes the same as the nearest melody note

6.1 Dataset and representation

- attack release
- tempo curves

6.2 Dealing with polyphony

6.3 Discretization

Distinction of expressive tempo and attack and release.

The delta framework

6.4 Performance Model

6.4.1 Expressive timing

Honing and Desain, Honing and timmers, Honing.

6.5 Learning

7 Method

8 Evaluation and Results

The proposed system actually requires more training data to do meaningful statistics. However, since expressive interpretation can be

8.1 Subjective Listening

8.2 Correlation

9 Integration with Note Level Performance Rendering

The YQX system defines expressive tempo implicitly by predicting the logarithmic ratio of the IOI in a performance and the IOI in a score. Timing alterations of notes are always defined relative to the base tempo. This does not correspond to the intuition that the *the tempo itself* is altered during the performance and that rhythmic changes should be seen relative to the local tempo. The same applies to the way YQX looks at loudness. This is specified as the logarithmic ratio between the notes loudness and the average loudness of the performance.

An intergration of constituent level expression and note level expression can provide a solution to this problem. We can define expressive tempo relative and dynamics relative to the expression parameters of the constituent

10 Conclusion

[1]

11 Discussion

Unfortunately we do not have acces to the large dataset that YQX uses. The dataset we use is smaller. Since we do not learn per note but per group of notes, the impact of a smaller dataset is even larger. We have discretize into rather large bins for this reason, resulting in cartoonistisc performances.

We think we can afford to do this because:...

- Top notes is not always the melody, musical attention
- Bass and harmony shouldn't be played with the same expression as melodynotes

12 Future Work

- Use loudness and tempo direction instead of averages (requires more data)
- Generalize approach to incorporate hierarchical structure, let structure extend to note level, so note level expression and structurelevel expression become integrated

12.1 Repetition

A notion of repetition and similarity would certainly improve the system. Repetition is a very good indicator of constituent breaks. Repetition and similarity could also be used to improve expressiveness of performances. It is probably telling when a phrase is repeated three times and then slightly altered the fourth

time. Although finding similarity and musically significant repetition is a subject of its own the delta framework could help to define repetition arbitrarily of transposition or rhythm. A list of pitch deltas can for example be used to detect repetition independend of transposition and a list of duration deltas can be used to detect repetition of rhythm independend of the notes used.

References

[1] M.J. van den Berg. Aspects of a formal theory of music cognition. 1996.