Understanding Expressive Music Performance Using Genetic Algorithms

Rafael Ramirez, Amaury Hazan

Music Technology Group
Pompeu Fabra University
Ocata 1, 08003 Barcelona, Spain
Tel:+34 935422165, Fax:+34 935422202
rafael@iua.upf.es, ahazan@iua.upf.es

Abstract. In this paper, we describe an approach to learning expressive performance rules from monophonic Jazz standards recordings by a skilled saxophonist. We use a melodic transcription system which extracts a set of acoustic features from the recordings producing a melodic representation of the expressive performance played by the musician. We apply genetic algorithms to this representation in order to induce rules of expressive music performance. The rules collected during different runs of our system are of musical interest and have a good prediction accuracy.

1 Introduction

Expressive performance is an important issue in music which has been studied from different perspectives (e.g. [8]). The main approaches to empirically study expressive performance have been based on statistical analysis (e.g. [26]), mathematical modelling (e.g. [27]), and analysis-by-synthesis (e.g. [7]). In all these approaches, it is a person who is responsible for devising a theory or mathematical model which captures different aspects of musical expressive performance. The theory or model is later tested on real performance data in order to determine its accuracy.

In this paper we describe an approach to investigate musical expressive performance based on evolutionary computation. Instead of manually modelling expressive performance and testing the model on real musical data, we let a computer use a genetic algorithm [14] to automatically discover regularities and performance principles from real performance data (i.e. Jazz standards example performances).

The rest of the paper is organized as follows: Section 2 describes how the acoustic features are extracted from the monophonic recordings. In Section 3 our approach for learning rules of expressive music performance is described. Section 4 reports on related work, and finally Section 5 presents some conclusions and indicates some areas of future research.

2 Melodic Description

In this section, we summarize how the melodic description is extracted from the monophonic recordings. This melodic description has already been used to characterize monophonic recordings for expressive tempo transformations using CBR [12]. We refer to this paper for a more detailed explanation.

We compute descriptors related to two different temporal scopes: some of them related to an analysis frame, and some other features related to a note segment. All the descriptors are stored into a XML document. A detailed explanation about the description scheme can be found in [11].

The procedure for description computation is the following one. First, the audio signal is divided into analysis frames, and a set of low-level descriptors are computed for each analysis frame. Then, we perform a note segmentation using low-level descriptor values. Once the note boundaries are known, the note descriptors are computed from the low-level and the fundamental frequency values. We refer to [10, 12] for details about the algorithms.

2.1 Low-level descriptors computation

The main low-level descriptors used to characterize expressive performance are instantaneous energy and fundamental frequency. Energy is computed on the spectral domain, using the values of the amplitude spectrum. For the estimation of the instantaneous fundamental frequency we use a harmonic matching model, the Two-Way Mismatch procedure (TWM) [18]. First of all, we perform a spectral analysis of a portion of sound, called analysis frame. Secondly, the prominent spectral peaks of the spectrum are detected from the spectrum magnitude. These spectral peaks of the spectrum are defined as the local maxima of the spectrum which magnitude is greater than a threshold. These spectral peaks are compared to a harmonic series and an TWM error is computed for each fundamental frequency candidates. The candidate with the minimum error is chosen to be the fundamental frequency estimate. After a first test of this implementation, some improvements to the original algorithm where implemented and reported in [10].

2.2 Note segmentation

Note segmentation is performed using a set of frame descriptors, which are energy computation in different frequency bands and fundamental frequency. Energy onsets are first detected following a band-wise algorithm that uses some psychoacoustical knowledge [17]. In a second step, fundamental frequency transitions are also detected. Finally, both results are merged to find the note boundaries.

2.3 Note descriptor computation

We compute note descriptors using the note boundaries and the low-level descriptors values. The low-level descriptors associated to a note segment are computed

by averaging the frame values within this note segment. Pitch histograms have been used to compute the pitch note and the fundamental frequency that represents each note segment, as found in [19]. This is done to avoid taking into account mistaken frames in the fundamental frequency mean computation.

2.4 Implementation

All the algorithms for melodic description have been implemented within the CLAM framework [1]. They have been integrated within a tool for melodic description, *Melodia*. This tool is available under GPL license. of the melodic description tool.

3 Learning expressive performance rules in Jazz

In this section, we describe our inductive approach for learning expressive performance rules from Jazz standard performances by a skilled saxophone player. Our aim is to find note-level rules which predict, for a significant number of cases, how a particular note in a particular context should be played (e.g. longer than its nominal duration). We are aware of the fact that not all the expressive transformations regarding tempo (or any other aspect) performed by a musician can be predicted at a local note level. Musicians perform music considering a number of abstract structures (e.g. musical phrases) which makes of expressive performance a multi-level phenomenon. In this context, our ultimate aim is to obtain an integrated model of expressive performance which combines note-level rules with structure-level rules. Thus, the work presented in this paper may be seen as a starting point towards this ultimate aim.

The training data used in our experimental investigations are monophonic recordings of four Jazz standards (*Body and Soul, Once I loved, Like Someone in Love* and *Up Jumped Spring*) performed by a professional musician at 5 different tempos around the nominal one (i.e. the nominal, 2 slightly faster and 2 slightly slower).

In this paper, we are concerned with note-level expressive transformations, in particular transformations of note duration, onset and energy. The note-level performance classes which interest us are *lengthen, shorten, advance, delay, louder* and *softer*. A note is considered to belong to class *lengthen* if its performed duration is 20% or more longer that its nominal duration, e.g. its duration according to the score. Class *shorten* is defined analogously. A note is considered to be in class *advance* if its performed onset is 5% of a bar earlier (or more) than its nominal onset. Class *delay* is defined analogously. A note is considered to be in class *louder* if it is played louder than its predecesor and louder then the average level of the piece. Class *softer* is defined analogously.

Each note in the training data is annotated with its corresponding class and a number of attributes representing both properties of the note itself and

[1] http://www.iua.upf.es/mtg/clam

some aspects of the local context in which the note appears. Information about intrinsic properties of the note includes the note duration and the note's metrical position, while information about its context includes the duration of previous and following notes, and extension and direction of the intervals between the note and both the previous and the subsequent note.

Using this data, we applied a genetic algorithm to automatically discover regularities and music performance principles. A genetic algorithm can be seen as a general optimization method that searches a large space of candidate hypothesis seeking one that performs best according to a fitness function. The genetic algorithm we used for this investigation is the standard algorithm (reported in [6]) with parameters r, m and p respectively determining the fraction of the parent population replaced by crossover, the mutation rate, and population size. We set these parameters as follws: $r = 0.8$, $m = 0.05$ and $p = 200$. During the evolution of the population, we collected the rules with best the fitness for the classes of interest (i.e. shorten, same and lengthen). It is worth mentioning that although the test was running over 40 generations, the fittest rules were obtained around the 20th generation.

Hypothesis representation. The hypothesis space of rule preconditions consists of a conjunction of a fixed set of attributes. Each rule is represented as a bit-string as follows: the previous and next note duration are represented each by five bits (i.e. much shorter, shorter, same, longer and much longer), previous and next note pitch are represented each by five bits (i.e. much lower, lower, same, higher and much higher), metrical strength by five beats (i.e. very weak, weak, medium, strong and very strong), and tempo by three bits (i.e. slow, nominal and fast). For example in our representation the rule

"if the previous note duration is much longer and its pitch is the same and it is in a very strong metrical position then lengthen the duration of the current note"

is coded as the binary string

00001 11111 00100 11111 00001 111 001

The exact meaning of the adjectives which the particular bits represent are as follows: previous and next note durations are considered much shorter if the duration is less than half of the current note, shorter if it is shorter than the current note but longer than its half, and same if the duration is the same as the current note. Much longer and longer are defined analogously. Previous and next note pitches are considered much lower if the pitch is lower by a minor third or more, lower if the pitch is within a minor third, and same if it has same pitch. Higher and much higher are defined analogously. The note's metrical position is very strong, strong, medium, weak, and very weak if it is on the first beat of the bar, on the third beat of the bar, on the second or fourth beat, offbeat, and in none of the previous, respectively. The piece was played at slow, nominal, and fast tempos if it was performed at a speed slower of more than 15% of the

nominal tempo (i.e. the tempo identified as the most natural by the performer), within 15% of the nominal tempo, and faster than 15% of the nominal tempo, respectively.

Genetic operators. We use the standard single-point crossover and mutation operators with two restrictions. In order to perform a crossover operation of two parents the crossover points are chosen at random as long as they are on the attributes substring boundaries. Similarly the mutation points are chosen randomly as long as they do not generate inconsistent rule strings, e.g. only one class can be predicted so exactly one 1 can appear in the last three bit substring.

Fitness function. The fitness of each hypothesized rule is based on its classification accurracy over the training data. In particular, the function used to measure fitness is

$$tp^{1.15}/(tp + fp)$$

where tp is the number of true positives and fp is the number of false positives.

Despite the relatively small amount of training data some of the rules generated by the learning algorithms have proved to be of musical interest and correspond to intuitive musical knowledge. In order to illustrate the types of rules found let us consider some examples of duration rules:

RULE1: 01000 11100 01111 01110 00111 111 010

"If the previous note is slightly shorter and not much lower in pitch, and the next note is not longer and has a similar pitch (within a minor third), and the current note is not on a weak metrical position, then the duration of the current note remains the same (i.e. no lengthening or shortening)."

RULE2: 11111 01110 11110 00110 00011 010 001

"In nominal tempo, if the duration of the next note is similar and the note is in a strong metrical position then lengthen the current note."

RULE3: 00111 00111 00011 01101 10101 111 100

"If the previous and next notes durations are longer (or equal) than the duration of the current note and the pitch of the previous note is higher then shorten the current note."

These simple rules turn out to be very accurate: the first rule predicts 90%, the second rule predicts 92% and the third rule predicts 100% of the relevant cases. The rules were collected during 10 independent runs of the genetic algorithm. The mean accuracy of the 10 best rules collected (one for each run of the algorithm) for "shorten", "same" and "lengthen" was 81%, 99% and 64%, respectively. We implemented our system using the evolutionary computation framework GAlib [9].

4 Related work

4.1 Evolutionary computation

Evolutionary computation has been considered with growing interest in musical applications. Since [15], it has often been used in a compositional perspective, either to generate melodies ([4]) or rhythms ([28]). In [22] the harmonization subtask of composition is addressed, and a comparison between a rule-based system and a genetic algorithm is presented.

Evolutionary computation has also been considered for improvisation applications such as [3], where a genetic algorithm-based model of a novice Jazz musician learning to improvise was developed. The system evolves a set of melodic ideas that are mapped into notes considering the chord progression being played. The fitness function can be altered by the feedback of the human playing with the system.

Nevertheless, few works focusing on the use of evolutionary computation for expressive performance analysis have been done. The issue of annotating correctly a human Jazz performance regarding the score is addressed in [13], where the weights of the edit distance operations are optimized with genetic algorithm techniques.

4.2 Other machine learning techniques

Previous research in learning sets of rules in a musical context has included a broad spectrum of music domains. The most related work to the research presented in this paper is the work by Widmer [29, 30]. Widmer has focused on the task of discovering general rules of expressive classical piano performance from real performance data via inductive machine learning. The performance data used for the study are MIDI recordings of 13 piano sonatas by W.A. Mozart performed by a skilled pianist. In addition to these data, the music score was also coded. The resulting substantial data consists of information about the nominal note onsets, duration, metrical information and annotations. When trained on the data, the inductive rule learning algorithm named PLCG [31] discovered a small set of 17 quite simple classification rules [29] that predict a large number of the note-level choices of the pianist.In the recordings the tempo of a performed piece is not constant (as it is in our case). In fact, of special interest to them are the tempo transformations throughout a musical piece.

Other inductive machine learning approaches to rule learning in music and musical analysis include [5], [2], [21] and [16]. In [5], Dovey analyzes piano performances of Rachmaniloff pieces using inductive logic programming and extracts rules underlying them. In [2], Van Baelen extended Dovey's work and attempted to discover regularities that could be used to generate MIDI information derived from the musical analysis of the piece. In [21], Morales reports research on learning counterpoint rules. The goal of the reported system is to obtain standard counterpoint rules from examples of counterpoint music pieces and basic musical knowledge from traditional music. In [16], Igarashi et al. describe the analysis of

respiration during musical performance by inductive logic programming. Using a respiration sensor, respiration during cello performance was measured and rules were extracted from the data together with musical/performance knowledge such as harmonic progression and bowing direction.

5 Conclusion

This paper describes an evolutionary computation approach for learning expressive performance rules from Jazz standards recordings by a skilled saxophone player. Our objective has been to find note-level rules which predict, for a significant number of cases, how a particular note in a particular context should be played (e.g. longer or shorter than its nominal duration). In order to induce expressive performance rules, we have extracted a set of acoustic features from the recordings resulting in a symbolic representation of the performed pieces and then applied a genetic algorithm to the symbolic data and information about the context in which the data appear.

Future work: This paper presents work in progress so there is future work in different directions. We plan to increase the amount of training data as well as experiment with different information encoded in it. Increasing the training data, extending the information in it and combining it with background musical knowledge will certainly generate a more complete set of rules. Another short-term research objective is to compare expressive performance rules induced from recordings at substantially different tempos. This would give us an indication of how the musician note-level choices vary according to the tempo. We also intend to incorporate structure-level information to obtain an integrated model of expressive performance which combines note-level rules with structure-level rules. A more ambitious goal of this research is to be able not only to obtain interpretable rules about expressive transformations in musical performances, but also to generate expressive performances. With this aim we intend to use genetic programming to evolve an initial population of rule trees and interpret these trees as regression trees.

Acknowledgments: This work is supported by the Spanish TIC project PromMusic (TIC 2003-07776-C02-01). We would like to thank Emilia Gomez, Esteban Maestre and Maarten Grachten for processing the data, as well as the anonymous reviewers for their insightful comments and pointers to related work.

References

1. Agrawal, R.T. (1993). Mining association rules between sets of items in large databases. International Conference on Management of Data, ACM, 207,216.
2. Van Baelen, E. and De Raedt, L. (1996). Analysis and Prediction of Piano Performances Using Inductive Logic Programming. International Conference in Inductive Logic Programming, 55-71.
3. Biles, J. A. (1994). GenJam: A genetic algorithm for generating Jazz solos. In ICMC Proceedings 1994.

4. Dahlstedt, P., and Nordahl, M. Living Melodies: Coevolution of Sonic Communication First Iteration Conference on Generative Processes in the Electronic Arts, Melbourne, Australia, December 1-3 1999.
5. Dovey, M.J. (1995). Analysis of Rachmaninoff's Piano Performances Using Inductive Logic Programming. European Conference on Machine Learning, Springer-Verlag.
6. De Jong, K.A. et al. (1993). Using Genetic Algorithms for Concept Learning. Machine Learning, 13, 161-188.
7. Friberg, A. (1995). A Quantitative Rule System for Musical Performance. PhD Thesis, KTH, Sweden.
8. Gabrielsson, A. (1999). The performance of Music. In D.Deutsch (Ed.), The Psychology of Music (2nd ed.) Academic Press.
9. The GAlib system. lancet.mit.edu/ga
10. Gómez, E. (2002). Melodic Description of Audio Signals for Music Content Processing. Doctoral Pre-Thesis Work, UPF, Barcelona.
11. Gómez, E., Gouyon, F., Herrera, P. and Amatriain, X. (2003). Using and enhancing the current MPEG-7 standard for a music content processing tool, Proceedings of the 114th Audio Engineering Society Convention.
12. Gómez, E. Grachten, M. Amatriain, X. Arcos, J. (2003). Melodic characterization of monophonic recordings for expressive tempo transformations. Stockholm Music Acoustics Conference.
13. Grachten, M., Luis Arcos, J., and Lopez de Mantaras, R. (2004). Evolutionary Optimization of Music Performance Annotation.
14. Holland, J.H. (1975). Adaptation in Natural and Srtificial Systems. University of Michigan Press.
15. Horner, A., and Goldberg, 1991. Genetic Algorithms and Computer-Assisted Music Composition, in proceedings of the 1991 International Computer Music Conference, pp. 479-482.
16. Igarashi, S., Ozaki, T. and Furukawa, K. (2002). Respiration Reflecting Musical Expression: Analysis of Respiration during Musical Performance by Inductive Logic Programming. Proceedings of Second International Conference on Music and Artificial Intelligence, Springer-Verlag.
17. Klapuri, A. (1999). Sound Onset Detection by Applying Psychoacoustic Knowledge, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP.
18. Maher, R.C. and Beauchamp, J.W. (1994). Fundamental frequency estimation of musical signals using a two-way mismatch procedure, Journal of the Acoustic Society of America, vol. 95 pp. 2254-2263.
19. McNab, R.J., Smith Ll. A. and Witten I.H., (1996). Signal Processing for Melody Transcription,SIG working paper, vol. 95-22.
20. Mitchell, T.M. (1997). Machine Learning. McGraw-Hill.
21. Morales, E. (1997). PAL: A Pattern-Based First-Order Inductive System. Machine Learning, 26, 227-252.
22. Phon-Amnuaisuk, S., and A. Wiggins, G. (1999?) The Four-Part Harmonisation Problem: A comparison between Genetic Algorithms and a Rule-Based System.
23. Quinlan, J.R. (1993). C4.5: Programs for Machine Learning, San Francisco, Morgan Kaufmann.
24. Ramirez, R. Hazan, A. Gómez, E. Maestre, E. (2004). Understanding Expressive Transformations in Saxophone Jazz Performances Using Inductive Machine Learning. Sound and Music Computing '04, IRCAM, Paris.

25. Ramirez, R. Hazan, A. Gómez, E. Maestre, E. (2004). A Machine Learning Approach to Expressive Performance in Jazz Standards MDM/KDD'04, Seattle, WA, USA.
26. Repp, B.H. (1992). Diversity and Commonality in Music Performance: an Analysis of Timing Microstructure in Schumann's 'Traumerei'. Journal of the Acoustical Society of America 104.
27. Todd, N. (1992). The Dynamics of Dynamics: a Model of Musical Expression. Journal of the Acoustical Society of America 91.
28. Tokui, N., and Iba, H. (2000). Music Composition with Interactive Evolutionary Computation.
29. Widmer, G. (2002). Machine Discoveries: A Few Simple, Robust Local Expression Principles. Journal of New Music Research 31(1), 37-50.
30. Widmer, G. (2002). In Search of the Horowitz Factor: Interim Report on a Musical Discovery Project. Invited paper. In Proceedings of the 5th International Conference on Discovery Science (DS'02), Lbeck, Germany. Berlin: Springer-Verlag.
31. Widmer, G. (2001). Discovering Strong Principles of Expressive Music Performance with the PLCG Rule Learning Strategy. Proceedings of the 12th European Conference on Machine Learning (ECML'01), Freiburg, Germany. Berlin: Springer Verlag.