Multi-Agent Simulation for Generating Expressive Music Performance

Qijun Zhang, Eduardo R Miranda

Interdisciplinary Centre of Computer Music Research,
Smeaton 206, University of Plymouth,
Drake Circus, Plymouth, PL4 8AA, UK
{qijun.zhang, eduardo.miranda}@plymouth.ac.uk

Abstract. This paper presents a system that implements a society of agent performers that evolve expressive music performances (EMP) through their interactions. Each agent performer evaluates a performance using a fitness function derived from the structure of the performing piece. Agents are born with different fitness functions, representing preferences for different types of performance. A new-born agent firstly applies a Genetic Algorithm to generate a refined performance with its fitness function. Through an agent's life, it listens to the performances of others' that it connects to in the social network. Given specific conditions, it may try imitate some performances and modify its own preference. An agent performer dies of age or inactivity. New agent performers are occasionally created. With this model, we aim to generate expressive performances through simulating the effect of social pressure and culture transmission.

Keywords: Expressive Music Performance (EMP), Multi-Agent Simulation

1 Introduction

Expressive music performance (EMP) describes the performance that brings music to life. In the context of western tonal music, expressions in a performance are commonly agreed to be delivered by delicate deviations from the notated musical score, associating the structures of the music piece being played. To build a computational model of EMP is to connect the properties of a musical score and performance context with the physical parameters of a performance, such as timing, loudness, tempo, articulation, etc. Comprehensive surveys of the state-of-the-art strategies of computational modelling of music performance can be found in [11][15]. Overall speaking, most computational models of EMP are devoted to discover and understand the commonality as well as diversity within or between performances. In order to explain how our approach fits into the field, we briefly introduce a few representative strategies for modelling EMPs as follows.

Analysis-by-measurement: Deviations in human performances are measured and analysed. These studies apply mathematical, statistical models or neural networks to describe the characteristics or regularities in the deviation patterns. [12][14] [2][8]

Analysis-by-synthesis: Performances are synthesized based on the expertise or hypothesis about human performances. In this process, the synthesis instructions are adjusted according to the listeners' (usually expert musicians) opinions.[7] ***Machine learning:*** Machine learning and data mining techniques are employed to discover regularities in large amounts of human performances data using computer.[16] Some systems use Case-Based Reasoning to generate human-like interpretation based on human performance data.[13] ***Model from intuition:*** A representative example is Manfred Clynes' pulse set theory.[5] Clynes proposed that a particular microstructure exists for a composer. Such a 'microstructure' defines a deviation pattern to perform each note's duration and loudness in this composer's compositions. Thus by discovering this pattern and performing each note accordingly, it is possible to achieve a rather ideal interpretation.

Given a music piece, among all its possible interpretations, which performances are acceptable is confined by musical psychology and perception of human listeners. Social factors, including effects of historical practice, interactions among performers, the practice of virtuosi etc, shape the evolving performance traditions.[6] From musicological point of view, to interpret a musical composition, 'work-specific traditions stand between period style and individual innovation.'[10] As far as we know, there hasn't been any computational model of EMPs built from the social point of view, although some statistic studies briefly appeared. [12] We consider social simulation a feasible methodology that can be used to generate performances artificially (i.e., from scratch). The idea is to place agent musicians in a simulated social context. In addition to a social network, an agent's interactions with others are guided and influenced by its preference of performance as well as its own performance. By simulating with this model, we expect to get some expressive performances by computer through agents' interactions. Another research goal is to test the hypothesis about social factors' effects on emerging expressive music performance.

Prior to this model, we have developed two systems to evolve music performances: one uses a Genetic Algorithm and the other learns by imitation. As reported earlier, the experiments show that given a piece of music, both systems could evolve performances from random initialization that sound realistic and interesting to listeners.[17][18] The design of these two systems were to mimic the mechanisms of (1) generational transmission and (2) social learning of performing profiles. The current model integrates these mechanisms. Furthermore, each agent is discriminated from others with its lifespan, preference of performances, social affiliates, in addition to its performance (interpretation) for the music piece.

The rest of the paper is organized as follows. Firstly in section 2, we will focus on introducing the representation and evaluation of a music performance. Then in section 3 we will present the system design, the architecture of agent and the agent society. Next, we will present experimental results of some typical simulations, followed by discussion of future works.

2 Representation and Evaluation of a Performance

An autonomous agent performer in the system has the ability of performing and evaluating a performance for the given piece of music. In this section, we are introducing how an agent represents its own performance and the criteria it uses to evaluate performances.

2.1 The Representation of An Agent's Performance

An agent's performance profile includes deviation values for all the notes in the music piece. To define precisely, a musical note's 'deviation values' are calculated as its duration and amplitude played in a performance, relative to its normalized values in score: i.e., the notated note length and flat volume. Figure 1 gives an example. A note that is lengthened when its duration deviation $Dev_{Dur}>1$, or shortened if $Dev_{Dur}<1$, or played exactly as notated in the score if $Dev_{Dur}=1$. Similarly, the higher value that a note has for Dev_{Amp}, the higher amplitude (loudness) it is played with.

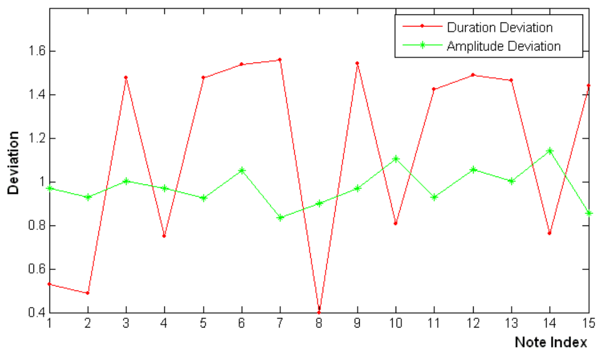

Figure 1. Representation of performance.

2.2 Structural Fitness for Evaluating a Performance

Expressions in performance are commonly agreed to primarily highlight music structures. A system that evolves EMP without human supervision hence demands structural information of the performing piece and explicit rules used for evaluating a performance. As a matter of fact, there is no agreement on a universal understanding of music structures or a single set of criteria that an EMP should satisfy. This is the main idea behind our building multi-agent system to generate expressive performances. In this concern, each agent performer has its 'preference' for a certain type of performance. An agent's preference is determined by its several attributes. The preference affects firstly its understanding of the musical structure, and secondly its evaluation about a given performance.

- Structural Analysis

In the system, all agent performers are informed with a few aspects of structural information of the performing piece, such as rhythmic, melodic, and harmonic information. An agent integrates those aspects of structural information into its own structural analysis, using a number of its parameters P_{str}={P_{group1}, P_{group2}, P_{group3}, P_{acc1}, P_{acc2}, P_{acc3}}. Agents vary in their values of P_{str}.

Grouping and accentuation are the most significant structural features that performers dedicate to highlight with expressions in performance. For the present system, each agent uses a modified Local Boundary Detection Model (LBDM) [3] to conduct a grouping analysis. It combines the rhythmic, melodic and harmonic information of the piece, and outputs the discontinuities between every two consecutive notes. This process involves the use of P_{group1}, P_{group2}, P_{group3}. In a similar manner, an agent combines metrical accent, melody accent and harmonic hierarchies in an analysis of accentuation. Every note is assigned a value of accentuation as a weighted sum of above aspects. P_{acc1}, P_{acc2}, P_{acc3} are used in the calculation. Due to space limit, we cannot include detailed algorithms. Please refer to earlier report if interested.

- Fitness Evaluation

To evaluate a performance, an agent judges how well the deviation values are mapped from its conducted structural analysis. We design the evaluation rules based on some generative principles for expressive music performance[4]. These principles describe the characteristics broadly found in EMPs, relating to a music piece's grouping and accentuation structure. To get a numerical evaluation for a performance, named 'structural fitness', each agent requires sets of parameters P_r={p_{r1}, p_{r2}, p_{r3}, p_{r4}, p_{r5}, p_{r6}} and P_{fit}={P_{dur}, P_{amp}} to add together the results of following rules. Similar to afore-mentioned P_{str}, each agent has different values for P_r and P_{fit}. We are not describing the equations for calculating following rules due to space limit.

Rule1: A performance is given reward if its notes' duration and amplitude deviation values in a phrase have a parabolic shape. The positions where a phrase starts, turns and ends are read from grouping analysis.

Rule2: A performance is given a penalty if the ending note of a phrase is not lengthened.

Rule3: The deviation value of a note's duration or amplitude has to be within a range, otherwise, a performance is punished.

Rule4: A performance is rewarded for having amplitude dynamics that follows the accentuation analysis curve.

Rule5: A performance is given a penalty if any note at significant accentuated positions has neither lengthened inter-onset value nor the local maximum amplitude.

Rule6: A means to highlight the discontinuity at grouping boundaries is to contrast the related notes’ deviation of duration. A performance is punished if it fails to do so.

3 The Society and Interaction of Agents

In a simulation, performances are dependent properties of the agent performers. During an agent's lifetime, it interacts with those agents that it is associated in the social network. It may modify both its performance and preference as a result of the interaction, by imitating others. In this section, we will firstly introduce the evolving social structure that agents' interactions are based upon. Then we will explain how an interaction affects related agents.

3.1 Structure of Agent Society

In one run of our simulation, 20 randomly generated agent performers are connected with 150 random links among them. As explained in last section, each agent is born with a number of weight parameters determining their preference (fitness evaluation) for performances.

Figure 2 depicts the flowchart of population dynamics. An agent dies at some point and is removed from the network. There are two possible ways to create a new agent (A_n) and add to the population. The first method is to generate A_n randomly, and to connect it with others based on preferential attachment principle.[1] The second is to make a copy of an existing agent (A_q) (randomly selected) and modify its parameters as A_n's. A_n is then linked with A_q and all the agents that A_q is connected with. For both methods, A_n evolves an initial performance $P_{n(ini)}$ using a Genetic Algorithm with its fitness function. The stopping criteria of Genetic Algorithm is if $P_{n(ini)}$ remains the fittest for a certain number of generations. An agent will build connections with others if all of its original affiliates have died, according to preferential attachment principle. We use the Java library of Universal Network/Graph (JUNG) for network and visualization. [9]

```
Randomly initialize a population of q agents
Run in parallel
    An agent interacts with others until it dies;
    when an agent dies
        Check the number (n_r) of agents alive
        if n_r<r1*q  (r1<1, e.g. 0.9)
            randomly select a number n,1≤ n ≤2*(1-r1)*q;
        else generate a random number s
            if s<r2, n=0;
            else n=1;  (e.g. r2=0.2)
    Create n agents and add them to the population.
```

Figure 2. Control of population dynamic

3.2 Interaction of agent performers

Table 1 groups the attributes of an agent performer respectively relating to its age, its performance, preference, and social situation. Those age parameters with '*' are threshold values for an agent to change its status. At every step (age), an agent listens to all or some of its affiliates' performances, adding 1 to its CurrentAge. Whenever an agent attempts to imitate a preferable performance, its value of 'learning' (i.e. the number of times it has learned) increases by 1. An agent's attributes with '++' are modified if by imitating another performance its current performance gets improved. Its popularity rises if other (unrepeated) agent finds its performance preferable. An agent updates its network because of other agents' death, birth or reconnection. The complete flowchart of an agent's activities and changes of its attributes is as shown in Figure 3.

Table 1. Attributes of an agent

Age	AdultAge*, DieAge*
	CurrentAge
Performance	Deviation Values^{++}
Preference	Parameters for fitness function^{++}
Social Interaction	Learning^{+}, Popularity$^{\wedge}$
	Network$^{\wedge}$

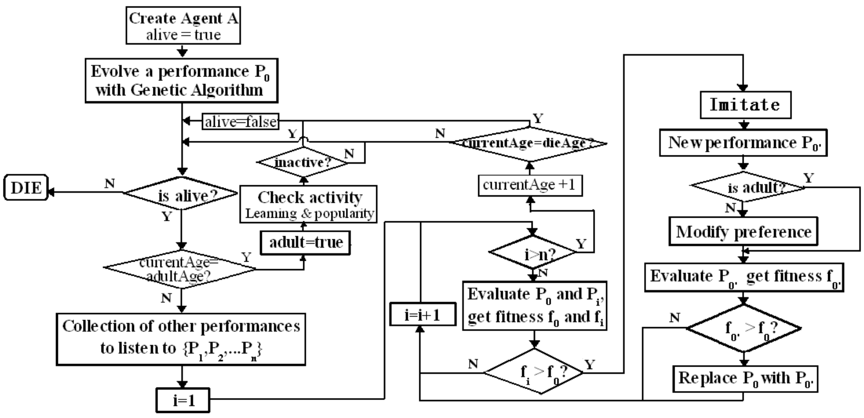

Figure 3. Flowchart of an agent's activities in its life time

4 Experiments

To demonstrate our system, we present below the results of the experiments with Chopin's 'Etude E major'. This piece is chosen because there is a body of work about human performances for it.[12] These studies nicely illustrate some commonality and diversity among real performances, related to the piece's music structure. Though we have clarified that it is not our goal to reproduce human performances, a comparison is somehow suggestive.

Figure 4 depicts the average timing profiles of human and agents' performances. The raw data was the Inter-Onset-Intervals (the time interval between the starting of successive notes) in the first 5 measures from the selected performances. The musical score is shown underneath for guidance. Our current system processes only monophonic melodies (i.e. no simultaneous notes). The related notes are represented as filled circles on the curve of human performance, with open circles for the accompaniment tones. Before calculating the average, data from each performance was standardized firstly. This is to remove the tempo difference between performances. In this way, the result reflect accurately the average 'within-performance deviation pattern' of the chosen performances. The human data is taken from paper by Repp [12], based on 117 available performance recordings. The result of agents' is by averaging the performances of 212 agents in five simulations. Each simulation was started with 20 agents, and finished when 800 agents have died. Those chosen have the highest popularity (they were imitated by at least 20 different agents) regardless which simulation it is from. Considering the fact that the average human performance is from samples of top international concert pianists, while our agent performers started with randomly initiated performances, we do not expect the agents to evolve performances that exactly follows the human performance. A possible cause of the difference is that our system only concerns melodic notes in the structural fitness. Even so, we can find that for many notes, the agents' performance is comparable with human performance. And by listening to the evolved performances, we found some agents had developed very interesting interpretations for the piece.

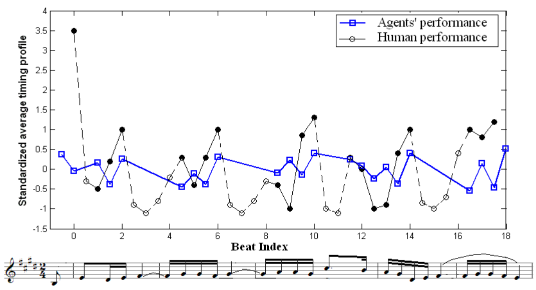

Figure 4. Comparison between averaged human performances and agents' performance

Figure 5 shows the effect of social interactions on pressuring agents to evolve bounded preference while maintaining diversity. To explain it further, the three graphs are taken from a single run of simulation. Each graph shows the values of a chosen parameter that determines an agent's fitness function. They are respectively P_{dur} , P_{group1} and P_{acc1}, as explained in section 2.2. The x-axis corresponds to the time when the agents died. Since agents interact with each other in parallel, we use the index of an agent's death to pinpoint it within the simulation. An agent is more popular if it is imitated by more different other agents. An agent's 'learning' attributes shows how many different other agents it has imitated during its lifetime. The most popular agents in this simulation are notated with squares. In each graph, a pair of '+' and 'o' is for the same agent. Such an agent's 'learning' attribute is among the highest in the simulation. '+' locates the value that the agent was born with, and 'o' is where it finalized.

It is quite clear that for each pair of '+' and 'o', 'o' is always closer than '+' to the nearest popular agent (i.e., a square) along x-axis. It shows that through interaction (by imitating others), an agent initially far from the main stream has been adapting itself. Its preference is modified towards those popular stars at that period. The fact that those learning agents may not have directly interacted with the popular ones demonstrates the effect of social pressure in the system.

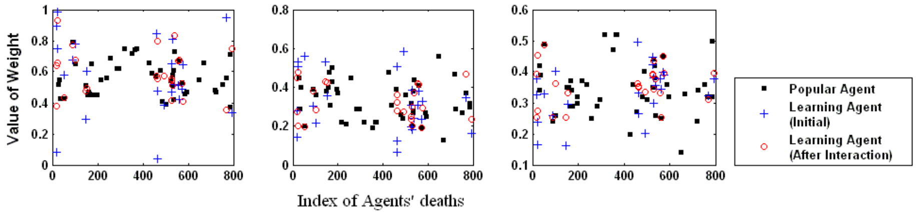

Figure 5. Illustration of the changes of agents' preference

5 Discussions and Future Work

This paper presented a system that evolves music performances through interactions in an artificial society of agent-performers. We have introduced the significance and feasibility of multi-agent simulation in generating expressive music performances. While the results are encouraging, several perspectives are interesting for further investigation. For example, to study the relationship between an agent's performance and its preference, and to run simulations with more music pieces of selective features to test the system. We are currently conducting formal human listening test to assess the generated performances.

References

1. A.-L. Barabási and R. Albert. Emergence of Scaling in random networks. Science, 286:509- 512, October, (1999).
2. Bresin, R.: Artificial neural networks based models for automatic performance of musical scores. Journal Of New Music Research 27 239-270 (1998).
3. Cambouropoulos E. The Local Boundary Detection Model (LBDM) and its Application in the Study of Expressive Timing. Proceedings of the International Computer Music Conference, Havana, Cuba, (2001).
4. Clarke, E. F. Generative Processes in Music. The Psychology of Performance, Improvisation, and Composition. Oxford Science Publications, (1988).
5. Clynes, M. Generative principles of musical thought integration of microstructure with structure. Journal For The Integrated Study Of Artificial Intelligence, Cognitive Science And Applied Epistemology, Vol 3, No.3, (1986).
6. Davidson, J. W. and North, A. C. The Social Psychology of Music. Oxford University Press, (1997).

7. Friberg A. Matching the rule parameters of Phrase Arch to performances of ‘Traumerei’: a preliminary study. Proceedings of the KTH symposium of Grammars for Music Performance, (1995)
8. Jan Beran, Guerino Mazzola, Analyzing Musical Structure and Performance- A statistical Approach, Statistical Science, Vol.14, 1, 47-79, (1999)
9. Java Universal Network/Graph Framework, http://jung.sourceforge.net/doc/api/index.html.
10. José A.Bowen. Finding the Music in Musicology: Performance History and Musical Works, in Cook and Everist (eds.) Rethinking Music. Oxford University Press, (1999).
11. Poli, G. D. Methodologies for expressiveness modelling of and for music performance. Journal Of New Music Research Vol 33, No. 3, (2004).
12. Repp, B. H. A microcosm of musical expression. I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major. Journal of the Acoustical Society of America, (1998).
13. Ramon Lopez de Mantaras and Josep Lluis Arcos, AI and music from composition to expressive performance Source, AI Magazine,Vol.23, No.3 ,(2002).
14. Todd, N.: The dynamics of dynamics: A model of musical expression. Journal Of The Acoustical Society Of America 91, 3540-3550, (1992)
15. Widmer, G. and Goebl, W. “Computational models of expressive music performance: The state of the art”. Journal of New Music Research, Vol 33, No.3, (2004).
16. Widmer, G. Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries, Artificial Intelligence, 146, (2003).
17. Zhang, Q. and Miranda, E. R. Evolving Musical Performance Profiles using Genetic Algorithms with Structural Fitness, Proceedings of Genetic and Evolutionary Computation Conference, Seattle USA, (2006).
18. Zhang, Q. and Miranda, E. R. Evolving Expressive Music Performance through Interaction of Artificial Agent Performers, Proceedings of ECAL 2007 Workshop on Music and Artificial Life, Lisbon, Portugal, (2007).