

AN INSTANCE BASED MODEL FOR GENERATING EXPRESSIVE PERFORMANCE DURING COMPOSITION

Alexis Kirke

Interdisciplinary Centre for
Computer Music Research,
School of Computing,
Communications and
Electronics,
University of Plymouth,
Plymouth, UK

ABSTRACT

We detail our work on a multi-agent model for the generation of structurally expressive music performance. It is a combined performance and composition system which attempts to implement the common observation in expressive music performance studies that performers slow down at boundaries in a musical piece, with the amount of slowing down being larger when the boundary is more significant. Because the model combines expressive performance with composition at a low enough level it is able to express group boundaries in an accurate and efficient way with no need for complex musical analysis; and this even if the music that is non-tonal and does not have fixed time signatures. This is demonstrated using an experimental example.

1. INTRODUCTION

Recently the Artificial Life approach of Multi-Agent Systems (MAS) [1] has been applied to music composition and expressive music performance. Our particular research goal is the creative combination of expressive performance and music performance - which, as will be seen, has both creative and efficiency advantages. For this purpose we will focus on one particular area of MAS, where the agents are provided with Biologically-inspired constraints – often called Instance-based Models (IBM) [2]. A significant number of multi-agent systems are listed in [3] - the earliest being Hudak and Berger's Game Theory modeling [4] and the most recent Zhang and Miranda's [5]. Similarly to [5], we utilize M.A.S. to generate expressive performances.

2. EXPRESSIVE PERFORMANCE

Research into computer composition of classical/art music had been around since the 1950s, and there have been many successful systems produced for automated computer composition [6]. However a computer will perform such generated tunes in perfect metronomic time, which can sound inhuman and unattractive. This is because *human* performers adjust the microstructure of the music. They perform expressively - speeding up and slowing down while playing, and changing how loudly they play. The performer's changes in tempo and dynamics allow them to express a fixed score, hence the term "Expressive" Performance. A Computer System for

Eduardo R. Miranda

Interdisciplinary Centre for
Computer Music Research,
School of Computing,
Communications and
Electronics,
University of Plymouth,
Plymouth, UK

Expressive Music Performance (CSEMP) is a system for generating expressive performances of music. There are a number of motivations for researching CSEMPs [7] as well as playing computer-generated music expressively, including: investigating human expressive performance; realistic playback on a scoring or composing tool; playing MIDI files; and computer accompaniment tasks.

3. HIERARCHICAL EXPRESSION

A number of CSEMPs have been produced - thirty are detailed in [7]. These have some common threads running through them. To clarify those threads, we must first introduce the concept of musical hierarchical structure. A piece of music often has a number of levels of meaning – a hierarchy. For example: notes make up motifs, motifs make up phrases, phrases make up sections, sections make up movements, and movements make up a piece. Each element - note, motif, etc - plays a role in higher elements. Research into human performers has suggested that – although there are a number of factors involved - the main consistent contributing factor to the expressive deviations during a performance is the hierarchical structure of the music [8][9]. Human performers have been shown to express the hierarchical structure in their performances.

As mentioned, a survey of previous CSEMPs shows two frequent commonalities:

1. Formal musical analysis - CSEMPs often base their performances on the hierarchical structure of the music requiring an analysis, sometimes by a musicologist.
2. Hierarchical combination of microstructure deviations – Most hierarchical systems generate the final tempo and dynamics expressive deviations by combining separate multipliers calculated for each level of the hierarchy.

A significant amount of CSEMP effort is often in analysing the musical structure of the score/audio. However doing musical analysis reliably is nowhere near an automatic process [10]. But the problem of musical structure analysis for expressive performance has new potential solutions when considered in the context of computer music. Most computer composition systems generate a piece based on some structure that can be made explicitly available. For example consider a constraint-based algorithmic composition system being used by a composer. Suppose the composer constrains

the system to generate a piece of the form ABCA where the main theme (A) is required to be made up of 3 motifs x, y and z. This information is then explicit in the composition system. Once the piece is generated, and if it is to be performed by computer using a CSEMP, then digitally communicating that structure information (concerning ABCA and x,y,z) will allow the CSEMP to more automatically and accurately analyse the hierarchical structure of the music.

So for computer music it is often inefficient to produce separate computer composition and expressive performance systems. Another relevant observation to be made from previous CSEMP research is that much of it is based on classical and tonal musicological analysis. These CSEMPs would not necessarily work so well with less traditional art music – the theories available for analysing such music are far more limited [11]. Furthermore a significant amount of modern music actually generates timing microstructure as well, questioning the traditional performance/composition dichotomy of microstructure/macrostructure.

4. AN M.A.S. FOR COMBINED COMPOSITION AND EXPRESSIVE PERFORMANCE

In past multi-agent systems used in music performance and composition the agents themselves and their interaction protocols are often inspired by algorithmic programming techniques (the exception being Miranda's Mimetic system where agents have a simple biological model of a vocal system and auditory perception system). As has been mentioned, this biologically-inspired type of multi-agent systems is more common in ecological Instance-based Modelling. Biologically inspired models have two main advantages: biological evolution has provided us with templates for systems which work; and biologically-inspired systems can be useful as modelling tools for the area they are inspired by (e.g. neurobiology, ecology, evolution of music, etc). These advantages of biological-inspiration can be combined with the emergent capabilities of multi-agent modelling to provide powerful modelling, creative and problem-solving tools. We now present a system that is an IBM for combined generation of composition and expressive performance.

Our system is based on the MIMACS system (Memetics-Inspired Multi-agent Composition System), which was designed to solve a specific compositional problem: taking advantage of a multi-agent system to solve the positioning problem in a multi-speaker field. Agents in this more recent version of MIMACS differ from their older cousins by being constrained to communicate by a biologically inspired model of music and language. [9] talks about "the notion that music performance and perception have their origins in the kinematic and dynamic characteristics of typical motor actions" and how "models of music performance in

which the auditory system interacts directly with the motor system". These ideas are also discussed in [12][13] and supported by the fMRI studies [14]. [15] suggests that our perception and performance of music are governed by models of motion in our auditory system, modelling this using kinematic equations whose solutions are parabolas which cause an acceleration and slowing at the start and end of musical phrases. We use this model of the auditory/motor system link and constrain the agents to an implicit kinematic physiological model – leading to a parabola curve being imposed over the tempo of any agent communication.

Extending the theme of implicit physiologically-inspired constraints, each agent has random imperfections in their hearing and in their speech actions. Humans make errors when they memorize music [16], and the agents will make memory errors when they play the tunes they learned from other agents. These errors can be quite large, involving getting a note pitch class or timing wrong. Similarly humans do not perform music in perfect time or exactly the same loudness due to small random motor errors [17]. These errors are simulated in our agents through random micro-changes in timing and dynamics of notes.

The only previous multi-agent system dedicated to expressive performance - Zhang/Miranda[5] - designs agents to imitate each others' expression performances of the same piece of music, whereas in our system each agent communicates with a different piece of music; and rather than agents imitating each others' pieces of music, they add music they hear (and "like") to the end or beginning of the music they already have learned. So each agent will be learning an ever-growing piece of music, based on what it hears from the other agents. If a limited number of agents are used and they communicate back and forth to each other, this will work as a form of hierarchical composition system. This hierarchical approach is used in other algorithmic composition systems [6] and fits in well with western approaches to music [9]. On top of this, the learned music will be affected by the agents implicit physiological constraints already described. So to summarise, our agents have the following interaction protocol:

1. Initialisation: each agent has a motif, phrase, or theme stored in its memory
2. A random agent (A) is picked to sing its tune.
3. The singing agent is constrained to sing with a timing parabola, and with random motor micro-timing errors.
4. The agent whose tune is most similar to A (say B) attempts to memorize A's tune, but in so doing may make errors in memorizing pitch or timing. B adds A's tune to the end or start (randomly picked) of its own tune, creating 1 larger tune from concatenating the two.
5. B sings its tune: go back to step 3.

The results of an experiment will now be shown, to demonstrate how a piece of music can be composed whose performance follows one of the most common observations in music performance studies [9][11][18]: performers slow down at boundaries in a musical piece, with the amount of slowing down being greater for the more significant the boundary; and this will be shown to occur despite the music being non-tonal and not having any fixed time signature.

5. EXPERIMENT AND RESULTS

To demonstrate this we set up a system of 3 agents. The 3 agents are given one unique melody each. The melodies' pitches are generated by an alleatoric algorithm[6] based on a random walk with jumps. Three rhythm palettes were designed, from which one was selected randomly for each melody. Then rhythms for the pitches were selected randomly from the palette. As a result, agents (A, B and C) started with quite rapid atonal phrases (a, b and c) of lengths 27, 43 and 39 notes.

A few additions were made to the system building upon the description in the last section. When an agent sings, there is some randomness in the picking of the agent which listens. The main factor affecting the picking is similarity of an agent's tune, but a random element is added to this to allow for some unexpectedness in the agent system behaviour. So there is always a chance that the agent that listens will not be the agent with the most similar tune, though it will be statistically the most likely agent. This prevents the system from getting permanently locked into a pair of agents swapping. The actual measure of difference used is a simple one – it is based on comparing the mean pitches and rhythms of the notes in a tune.

The multi-agent system was allowed to run through 4 cycles of interaction. For these cycles the chance of a macro-error in pitch and of a macro-error in rhythm was set to 10% per interaction. Furthermore there was a 50% chance of an agent adding new parts to its tune by appending and a 50% chance of the addition being prepended. The amount of random micro-error added during interaction was between 0 and 0.1 beats for tempo and 0 to 1 MIDI velocity units for dynamics. After all the cycles were run, the agent with the longest theme was chosen to be examined – this turned out to be agent B. Agent B's theme as B would sing it is shown in piano roll notation in Figure 1.

This is a composition of 206 notes and 6 phrases: abacab (a is A's original tune, b is B's, and c is C's). In fact this is more precisely written as (ab)(a(c(a'b'))') where each prime represents a transformation on a phrase. This also shows the hierarchical structure of B's piece of music – the greater the number of brackets around a subphrase, the lower the subphrase is in the

hierarchy. Figure 2 plots the performance curves generated for B's tune. As applied to tempo it can be seen that there are series of accelerations and decelerations – 7 in total.

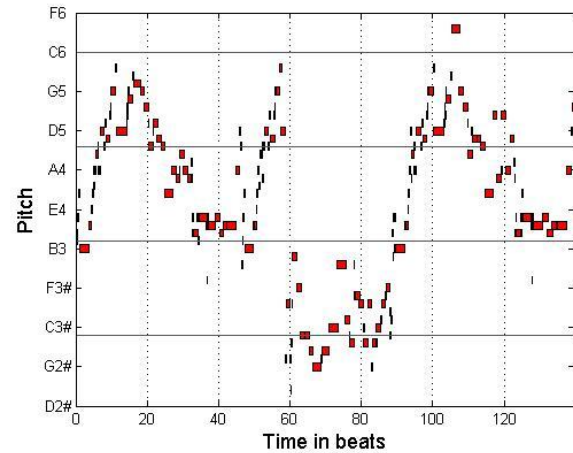


Figure 1: Agent B's theme as sung - in piano roll notation - phrase form abacab

It can be shown that this performance curve is consistent with performers slowing down at boundaries in a musical piece, with the amount of slowing down being greater for the more significant the boundary. The largest phrase boundaries in any composition are the beginning and end. And we can see in Figure 2 that the slowest tempo performance is found at the start and ends of the curve, as expected. To analyse the rest of the phrase boundaries we will order the deceleration minima, with 6 representing the most decelerated and 1 represented the least decelerated. This gives (in temporal order): 6,3,5,4,2,1,6. If we now fit these into the abc representation of B's theme building blocks, things become clearer. We place each deceleration number in the middle of its relevant phrase/subphrase: 6(a3b)5(a4(c2(a'1b'))')'6.

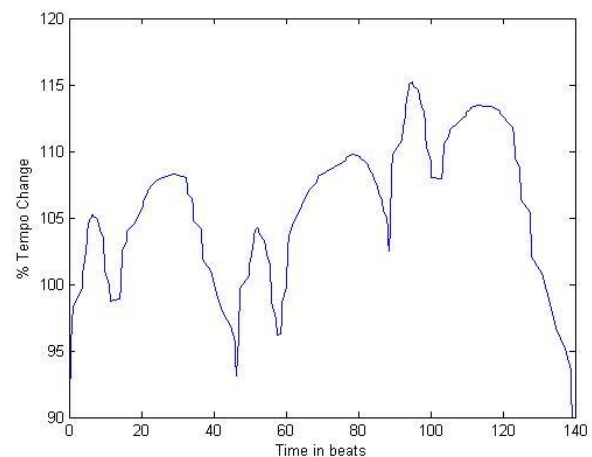


Figure 2: Performance Curves for Agent B's final tune

It can be seen that in all cases the greatest decelerations occur where there are the least number of brackets around them. Now the larger the number of brackets

around a phrase, the lower the “order” of the built phrase in the hierarchy, since if it is contained in a lot of brackets it is a sub-phrase of a sub-phrase and so forth. So the location of the numbers in the brackets supports the fact that the greatest slowing down is at the most significant phrase boundaries. Though, it is important to note that the importance of a phrase boundary is usually but not always decided by its location in the hierarchy of a composition.

6. CONCLUSIONS

We have introduced a multi-agent model for the generation of expressive music performance during composition. The system attempts to implement the common observation that performers slow down at boundaries in a musical piece, with the amount of slowing down being greater for the more significant the boundary. Because it combines expressive performance with composition at a low enough level it is able to express group boundaries in an accurate and efficient way with no need for complex musical analysis; and this even if the music that is non-tonal and does not have fixed time signatures. This has been demonstrated using an experimental example. Despite the limitations of our system and the large scope for future study, it is hoped that this work has highlighted the advantages of combining expressive performance and algorithmic composition into a common system in computer music by demonstrating one possible approach to do this.

This work has been funded by the UK EPSRC project “Learning the Structure of Music” (EPD063612-1).

7. REFERENCES

- [1] Kirke, A., *Learning and Co-operation in Multi-Robot Systems*. PhD Thesis, University of Plymouth 1997
- [2] Grimm, V. and Railsback, S.F. Agent-Based Models in Ecology: Patterns and Alternative Theories of Adaptive Behaviour. In *Agent-Based Computational Modelling: Applications in Demography, Social, Economic and Environmental Sciences*, Billari, F.C., Fent, T., Prskawetz, A., and Scheffran, J., Eds. Physica-Verlag HD, 139-152, 1995
- [3] Kirke, A. and Miranda, E. “Using a Biophysically-constrained Multi-agent System to combine Expressive Performance with Algorithmic Composition”, In Print, 2008
- [4] Hudak, P. and Berger, J. A Model of Performance, Interaction, and Improvisation. In *Proceedings of ICMC, ICMA 1995*
- [5] 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 (MusicAL 2007)*, Lisbon, 2007.
- [6] Miranda, E. *Composing Music with Computers*. Focal Press, Oxford, 2001
- [7] Kirke, A. and Miranda, E. A Survey of Computer Systems for Expressive Music Performance, *ACM Computing Surveys*, In Print, 2008
- [8] Clarke, E.F. Generative Principles in Music Performance. In *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition*, Sloboda, J.A., Ed. Clarendon Press, Oxford, 1-26, 1998
- [9] Palmer, C. Music performance. *Annual Review of Psychology* 48, 115-138, 1997
- [10] Masatoshi, H., Keiji, H., and Satoshi, T. ATTA: Automatic Time-span Tree Analyzer based on extended GTTM. *IPSJ SIG Technical Reports*, 19-26, 2005.
- [11] Clarke, E.F. Expression and communication in musical performance. In Sundberg, J., Nord, L., and Carlson, R., Eds. Macmillan Press, London, 1991
- [12] Clarke, E.F. Generativity, mimesis and the human body in music performance. *Contemporary Music Review* 9, 207-219, 1993.
- [13] De Poli, G. Expressiveness in music performance. In *Sound to Sense: Sense to Sound: A State-of-the-Art in Sound and Music Computing*, Polotti, P. and Rocchesso, D., Eds. Logos Verlag, 2006.
- [14] Durrant, S., Miranda, E.R., Haroon, D., Shawe-Taylor, J., Brechmann, A., and Scheich, H. Neural Correlates of Tonality in Music. In *Proceedings of Music, Brain & Cognition Workshop - NIPS Conference*, 2007
- [15] Todd, N.P. The kinematics of musical expression. *Journal of Acoustical Society of America* 97, 1940-1949, 1995
- [16] Gfeller, K.E., Olszewski, C., Turner, C., Bruce, G., AND Oleson, J. Music Perception with Cochlear Implants and Residual Hearing. *Audiology and Neurotology* 11, 12-15, 2006
- [17] Cemgil, A. and Kappen, B. Monte Carlo Methods for Tempo Tracking and Rhythm Quantization. *Journal of Artificial Intelligence Research* 18, 45-81, 2003
- [18] Shaffer L.H., Todd N.P. “The interpretive component in musical performance” in *Action and Perception in Rhythm and Music*. Stockholm: R.Swed. Acad. Music, 139-52, 1987