REAL-TIME INTERACTIVE MUSICAL SYSTEMS: AN OVERVIEW

A.N.Robertson
Centre for Digital Music
Department of Electronic Engineering
Queen Mary University of London

M. D. Plumbley
Centre for Digital Music
Department of Electronic Engineering
Queen Mary University of London

ABSTRACT

We present an overview of developments towards interactive musical systems. A description of an interactive system is given and we consider potential uses for the automation of creative processes within live performance. We then look at the history of research into the problem of automatic accompaniment, discuss a variety of current interactive systems and present some ideas for future research.
Keywords – Score following, automatic accompaniment, musical interactive system, machine learning.

1. INTRODUCTION

It has been a long-standing goal to empower computers with the ability to interpret sounds in a musically meaningful way and use this information to integrate themselves into a performance by human musicians [2]. In the last forty years, huge developments have been made in the creation of new electronic technology for music. However, almost all of this technology requires a human agent to write the musical rules that the computer follows when creating sounds.

An interactive system uses information from the actions of the user, perhaps via an interface, to formulate its response or output. The term, 'musical interactive system', encompasses all systems which depend on the users actions to generate sound and also includes systems which interact musically with the user.

Any musical interface, such as a keyboard, might form part of an interactive system in the first sense by being connected to a sound generator. Also, technology that allows the user to alter the parameters used in creating the sound offers a form of musical interaction for the user, but here we are primarily concerned with systems that interact musically by 'listening' to external input from the user and generate their responses in real-time. The difference is that the actual interaction is musical as opposed to a system giving rise to an interaction that has a musical context. Musically interactive systems include automatic accompaniment and systems capable of improvising with musicians.

There are significant constraints on such an interactive system. It must negotiate errors within detection and performance. It must also be efficient so that it is able to respond to new input in real-time. In the case of score followers they must be robust enough to relate the performance to the score despite the fact that no two performances are ever the same, whilst in the case of improvising systems, they must be able to make sense of unexpected musical events and use some form of knowledge database when making musical decisions.

2. TECHNOLOGY IN LIVE PERFORMANCE

The most significant step in the use of computers for music has been in the recording studio, where the sound quality available from digital hardware now rivals that of analogue tape. Computers have an advantage over analogue recording in that they allow non-destructive over-dubs and can act as powerful sequencers of audio samples. The most widely used software packages for digital recording are ProTools and Logic, which base themselves on multitrack tape. Some software packages are also capable of some interesting processing, such as Ableton Live's ability to offer easy time-stretching and pitch-shifting so that audio sample manipulation is comparatively easy.

It seems that existing software is capable of performing many of the creative tasks that musicians can formulate, provided that this is done in the studio and hence, not in real-time. Recordings are made through thousands of small tasks so that gradually a single audio track is edited together for each channel, effects are applied and the mix is performed with automated volume, panning and effects changes.

It seems natural to ask what functions the computer can play within the context of live performance. There is already considerable use of technology and automation within contemporary musical performance. Some of the main uses are as follows:

- Electronic instruments - keyboards and other interfaces to electronic synthesizers which make the sounds. Sounds will be triggered by the interface - sometimes pitched as is the case with keyboards , sometimes triggered from a pad. Considerable research into designing new interfaces is regularly presented at NIME conference [19] with some interesting innovations.

- Audio processing - vocals and drum signals might be gated. Compression and other effects are ap-

Email: andrew.robertson@elec.qmul.ac.uk

plied. A sound engineer mixes the signals for the audience and provides on-stage monitoring.

- Lighting - various methods are used. Often a set of pre-written lighting settings can be triggered at a lighting console. Also, back projections are sometimes used to considerable effect.

- Live coding [17] - using audio programming languages such as SuperCollider [12], Max/MSP [13], ChucK [14], it is possible to create sounds ‘on the fly’. Performances making use of laptops are increasingly common - for instance, Polar Bear, a jazz band nominated for the Mercury Music Prize 2005, have performed with one musician creating sounds within Max using a joystick interface.

It seems natural that some of these processes could be automated in a manner which takes advantage of the computers powerful ability to schedule events. Synchronisation of these processes with the performance could be achieved to a level that would not be possible manually. This would require that the computer can relate inputs derived from the performance to an abstract representation, or score, in order to schedule events such as electronic parts, audio or lighting effects.

At present, performers wishing to use pre-recorded parts synchronise with the computer by having the drummer listening to a click track. This forces the tempo to be constant, immediately imposing a constraint on potential dynamics for the piece. By having the computer follow the musicians, automation of this technology could result in powerful, new creative effects. Research into this area began with the problem of score following of live performance since the knowledge provided by the score simplified the wider problem of automatic transcription.

3. HISTORY OF AUTOMATIC ACCOMPANIMENT

At the 1984 ICMC, Barry Vercoe [1] and Roger Dannenberg [3] both independently presented working systems that could perform the task of automatic accompaniment. In Vercoe’s case he used optical sensors on a flute and some pitch tracking to peform score following.

Over the subsequent year, Vercoe began working with Miller Puckette [2] on the problem of training the Synthetic Performer to learn from rehearsals. This was a significant step as it acknowledged the important role that rehearsals have in improving conventional performances by enabling the musicians to learn when to expect rhythmic deviation, such variations in onset timing and tempo, which are not predicted by the score. Their research suggested that machine learning techniques would need to be used if the computer is to be able to do more than effectively sight-read.

Importantly, the distinction was made between rhythmic aberrations, where a note is played early or late, and tempo variation. The fraction by which a note is early or late is learnt from rehearsals and written into the performance record. The system then expects a variation in the onsets of notes from that implied by a literal reading of the score and when using a note onset time to determine the tempo, can factor into account how likely this timing is to vary between performances.

Vercoe also formulated automatic accompaniment as consisting of three processes: Listen, Perform and Learn. The role of Listen is to extract information from the performance which Perform then uses to match against the score and synchronise the accompaniment. The Learn process involves extracting information from performances which aids the system to follow subsequent renditions of the piece. This formulation of the problem has been influential in subsequent research in the area.

Dannenberg used a form of dynamic programming to calculate the least cost match between the observed performance, translated into Midi notes, and the score representation. He formalised the problem of score following as finding the longest common subsequence of two strings, one representing the score and the other the observed performance. Where notes are skipped, they must be removed from the score string. Where they are inserted they are removed from the performance string and where they are wrongly played or detected, they must be removed from both. The matcher is an algorithm designed to solve this problem with use of heuristics in order that it is successful in practice, as was demonstrated at the 1984 ICMC.

Puckette subsequently worked on the design of the Max/MSP language at IRCAM and re-implemented the ideas in an open source version called Pure Data [6]. Both languages aim to allow real-time manipulation of audio and Midi data. He also continued to work on automatic accompaniment, introducing the idea of concurrent matchers, using a slow but reliable matcher to follow the global position within the piece and a faster matcher used to trigger the computer’s response to notes [5].

4. STATE OF THE ART

Raphael’s Music Plus One system [4] is capable of providing automatic accompaniment to solo oboe. This system uses a hidden Markov model (HMM) to model the note transitions within a piece. He follows Vercoe [2] in delineating listen, perform and learn processes for the system. Previous score followers had used pitch-to-Midi conversion where as the listen component of Raphael’s system processes monophonic audio to form a vector of features, including the energy of the signal and the presence of individual notes computed via a Fourier transform.

The HMM works by assuming that the observed states, derived from the processed audio of the performance, are created by a hidden sequence of states, resulting from the player’s movement through the score. This assumption in the architecture of an HMM has made it a popular method of tackling the problem of score following. In Raphael’s model, the states used in the hidden layer represent the attack and steady-state parts of each pitched note of the scale and a rest state is included for when the soloist is not playing. Each note within the score is modeled as a series of states which are then chained together.

Raphael then models the tempo and duration of each note as two random variables. The system trains on rehearsals to learn how the tempo tends to fluctuate within the piece and to what extent each note has an early or late onset relative to this underlying tempo. By training the system on previous renditions of the piece, the Music Plus One system is sensitive to the rhythmic variations of the musician. Raphael has extended the system to use real audio accompaniment of a recorded orchestra by time-stretching the accompaniment to synchronise with the soloist.

IRCAM have developed a set of Max patches, under the name Suivi [8], which perform score following from audio input. A two-level HMM is used [7]: a low level HMM which uses features relating to the attack, sustain and decay of a note and a higher-level HMM in which transitions are made between states corresponding to the notes in the score. The system analyses the score to generate a HMM for note transitions within the score, using ghost states to account for possible errors in performance. The HMM is dynamic in the sense that the set of states used changes as one progresses through the score. For a particular point, approximately twenty notes might contribute information to the transition matrix and the number of states might be around 200. Suivi tends to be used on monophonic instruments at present since pitch detection is much more reliable when there is only one note to detect. A 4096 sample frame FFT, with a hop size of 512 samples, is used as input to the system which extract features from the Fourier transform of the audio. A note is more likely to be detected once it is present in the central section of the window and so the resulting latency is up to 2000 samples or 45ms. This delay is analogous to the kind of delay one finds naturally in a large room and the system has been sucessfully used to provide automated electronic parts for several classical pieces which blend natural and electronic sounds such as *En Echo* by Philippe Manoury and *Explosante-fixe* by Pierre Boulez.

The Continuator by Francois Pachet [9] uses Markovian techniques to interact with a pianist in a novel way. Pachet was influenced in this by David Cope [10], who used Markov models to discover new musical phrases that were coherent with his own style. It builds a database of patterns played by the musician and indexes all subsequences of the input. When the musician stops playing, the system continues the phrase by using the transitions from the longest subsequence matching the input to continue the phrase. Its effect has surprised musicians who have played with the system. The Continuator has several modes. The first, 'Autarcy', the system has no memory and progressively trains on input to learn the musical style of the performer. In the 'Virtual Duo' mode, the 'memory' of another musician is used as the transition matrix for the system, with the result that phrases are completed in the style of the previous player. Pachet's system is novel in that it is suited to improvisation and the learning takes place in real-time.

Another improvisation system is GenJam by John Biles [11]. GenJam uses genetic algorithms to modify a population of individuals, in this case musical phrases or 'licks', which are played over a set sequence of jazz chords. The human classifies each individual according to its suitability as a solution a problem, which is the aesthetic appeal of the phrase as a musical improvisation. This suitability rating is used by the algorithm to determine the individual's evolution, both its own survival in the population and how it is combined with other individuals to create new individuals. By the fifth generation, the improvisation is fairly pleasing with regards to tonality and rhythm. Biles then modifies the generator of the sound to bend into the notes and introduces variations within onset and duration time. This has the effect of humanizing the part. A live demonstration of GenJam was given at the ICMC'98.

Blackwell and Young created an interactive system, Swarm Granulator [15], which creates musical events from the movement of swarm-like particles in a virtual space. There exist attractors within this space which are produced by human performers, whilst the particles repel those close by them, so that globally, as a swarm, they gravitate towards the attractor but do not coalesce into one another. The particles' behaviour around the attractors has a musical influence over the actions of the performer, hence bringing about an interaction between the human improviser and the computer generated events.

Nick Collins has developed a a real-time event detection system [18] in SuperCollider, making use of the onset detection function developed by Bello and co-authors [16], which is implemented with low latency. By detecting note onsets, this system is suited to performing automated creative tasks based on note onsets within the signal.

5. A RESEARCH PROPOSAL

The first stage of the project is the development of a system that is capable of predictive score following and sequencing. We consider this a sub-problem of our eventual aim of an interactive system capable of autonomous generation of output. Given an appropriate automatic accompaniment system, compositional techniques could be employed whereby the computer is given a set of rules or makes use of machine learning techniques in creating a musical part. Both neural networks and genetic algorithms could allow the user to modify the system so that their aesthetic is imposed through positive reinforcement of appealing contributions from the computer.

The inputs to the system will initially be audio signals from drums in the case of rhythmic information and guitar input for harmonic content. The score will be specified by chord transitions as used in rock, blues or jazz, rather than note-to-note transitions as used in classical scores. Whilst there already exist techniques for beat induction and chord transcription, these are not sufficiently reliable at present for real-time score following of polyphonic instruments. Most previous score following systems [1] [3] [4][7] have concentrated on monophonic signals from solo instruments, so the use of multiple instruments and the type of signals to be processed are novel directions in the development of automatic accompaniment systems for contemporary music.

6. ACKNOWLEDGEMENTS

I would like to acknowledge the EPSRC for their support of this research and the Multimedia Knowledge Management Network for their support of my attending the NIME 2006 Conference.

7. REFERENCES

[1] B. Vercoe, "The Synthetic Performer in the Context of Live Performance", in *Proceedings of the International Computer Music Conference (ICMC)*, pp. 199-200, 1984

[2] B. Vercoe and M. Puckette, "Synthetic Rehearsal, Training the Synthetic Performer", in *Proceedings of the International Computer Music Conference (ICMC)*, pp. 275-278, Simon Frasier University, Canada, 1985.

[3] R. B. Dannenberg, "An On-line Algorithm for Real-Time Accompaniment", in *Proceedings of the International Computer Music Conference (ICMC)*, pp. 193-198, San Francisco. 1984

[4] C. Raphael, "Automatic Segmentation of Acoustic Musical Signals Using Hidden Markov Models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 21, No. 4, pp. 360-370, April, 1999.

[5] M. Puckette, "Score Following Using the Sung Voice", in *Proceedings of the International Computer Music Conference (ICMC)*, pp. 199-200, Banff Centre for the Arts, Canada, 1995.

[6] M. Puckette, "Pure Data: another integrated computer music environment", in *Proceedings of the Second Intercollege Computer Music Concerts, Tachikawa.*, pp. 37-41, 1996.

[7] N. Orio and F. Dechelle, "Score Following Using Spectral Analysis and Hidden Markov Models", in *Proceedings of the International Computer Music Conference (ICMC)*, Havana, 2001.

[8] N. Orio and S. Leemouton and D. Schwarz, "Score Following: State of the Art and New Developments", in *Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME)*, Montreal, Canada ,pp. 36-41, 2003.

[9] F. Pachet, "Interacting with a Musical System: The Continuator", in *C. Anagnostopoulou, M. Ferrand, and A. Smaill, editors, Music and Artificial Intelligence: Lecture Notes in Artificial Intelligence 2445*, pages 119132. Springer-Verlag, 2002.

[10] D. Cope, "Experiments in Musical Intelligence". *Madison, WI: A-R Editions*, 1996.

[11] , J. Biles, "Interactive GenJam: Integrating real-time performance with a genetic algorithm", in *Proceedings of the International Computer Music Conference (ICMC)*, pp. 232-235, Ann Arbor, MI., USA, 1998.

[12] J. McCartney, "Supercollider: A new real-time synthesis language", in *Proceedings of the International Computer Music Conference (ICMC)*, Hong Kong University of Science and Technology, China, 1996.

[13] D. Zicarelli, "An Extensible Real-Time Signal Processing Environment for Max", in *Proceedings of the International Computer Music Conference (ICMC)*, Univ. of Michigan, Ann Arbor, MI., 1998.

[14] G. Wang and P. Cook, "ChucK: A Programming Language for On-the-fly, real-time Audio Synthesis and Multimedia", in *Proceedings of the International Computer Music Conference (ICMC)*, Singapore, 2003.

[15] T. Blackwell and M. Young. "Swarm granulator". In *EvoWorkshops 2004*, pp. 399408, Berlin, 2004. Springer-Verlag.

[16] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and S. B. Sandler. "A tutorial on onset detection in music signals". *IEEE Transactions on Speech and Audio Processing*, 2004.

[17] N. Collins , A. McLean , J. Rohrhuber and A. Ward, "Live coding in laptop performance", *Organised Sound*, Vol. 8 No.3, pp. 321-330, December, 2003.

[18] N. Collins, "An Automated Event Analysis System with Compositional Applications", in *Proceedings of the International Computer Music Conference (ICMC)*, Barcelona, 2005.

[19] Proceedings viewable at `http://www.nime.org`

[20] `http://www.ableton.com`