Laboratório VISGRAF
Instituto de Matemática Pura e Aplicada

A Survey of Real-Time Interactive Music Systems

Sergio Krakowski
Luiz Velho (orientador)

Technical Report TR-06-05 Relatório Técnico

December - 06 - Dezembro

Sergio Krakowski
skrako@gmail.com

Rio de Janeiro
31.07.2006

Lecture Course - Techincal Report

1 Interactive Systems

In this section we present some of the most important musical interactive systems found in the Computer Music literature. This domain is sometimes named Interactive Music Companion, or just IMC.

1.1 BoB

The first of them is the work done by Belinda Thom ([thom00]). The author builds an agent named *Band-out-of-the-Box* (BoB) that is supposed to trade fours with a musician. Trade four is a Jazz music term to express an improvising scene where two musicians exchange ideas by alternating their improvisation. Each musician plays during a four bar period and then listens to the other's play trying to build up a musical conversation. The BoB agent uses unsupervised machine learning techniques to configure its aesthetic musical sense to each user/musician. The approach she uses to deal with this question comes from the Belivable Agents domain. The system must be previously trained and then works in the real-time mode where it should trade fours. It is not specified if the system deals with MIDI or audio data.

1.1.1 Believable Agents

The inital purpose of the BA was to create interactive virtual worlds whose characteristics and story lines were believable. This approach is useful because adapts quite well to each user/improviser instead of trying to use rule based approaches as many AI systems do. This system don't try to capture a general theory of Jazz but the notable aspects of an improvisor.

1.1.2 Representation

The representation used are the variable length trees. This tree is generated in a per-bar basis. Each node can be divided in two or three depending on the rhythm of the melody played and the pitch class and octave are stored in each tree leave.

Then three per-bar histograms are created: Pitch Class, Intervallic Motion and Melodic Direction. The first histogram stores the frequency of a pitch on a bar, the second, which intervals are between each note, and the third stores the direction of each interval, if

acendent or decendent. Although it seems the rhythmic content is lost in these histogram representation, the author argues it isn't.

With these histograms, the author builds clusters that are supposed to classify in diferent subjective sets, the musical contents of each bar in the four-bar solo. This way the computer can classify, after being trained, what is the sequence of subjective classes played by the user and then can create a four-bar phrase that repeats this sequence of classes.

1.2 The Continuator

In [pachet02], François Pachet divides the musical systems domain into two cathegories: the interactive musical systems and the music imitation systems. The main goal of this work is to present a system that combine both worlds creating an interactive system that produces stylistically consistent music learning during the real-time mode. The base used is the the Markov model augmented to be able to deal with musical issues such as harmony, rhythm, beat and precision. This system is based on MIDI data.

The Markov-based model are powerful to build musical phrases in a specific style learned. Its drawback comes from the fact that their generative power is limited because it does not have long-term memory. Because of that the system will not be able to create phrases that respect longer period musical issues such as form. This will be left to the user. The system just "fill in the gaps" left by the musician.

The phrases received as input are indexed in a way that all the subsequences can be easily acessed. This procedure builds up a database of all that has been played.

1.2.1 Trees

A tree is built by parsing each learnt phrase. This is done from right to left. Each node is an element of the general phrase played by the user. And this node indicates what are the possible elements in the general sequence that have followed this element.

To build the continuation of a sequence s the system chooses the biggest sequence that matches, from right to left, the sequence s. Then it chooses the next element from the list of possible elements contained in the node of the final element of s. The procedure is again iterated. This way the system builds a phrase which a Markovian distribution equal to the input phrase.

1.2.2 Reduction functions

The sequences above mentioned are completely abstract. They are not built from the raw midi data. This data contain too much information. Instead the elements of the sequences can be many caracteristics of the input notes. For example the pitch, or the

pitch and duration, or the pitch region, which is another new concept to deal with Jazz music. The reduction function extracts one these characteristics which will allow to build the continuations.

1.3 Jam'aa

The authors of [jamaa06] present a musical ensemble of two Darbuka (Middle East percussion instrument) players and a robot. The robot named *Haile* receives audio input, analyses it in real-time to extract the musical content, and plays back in an improvisational manner. The aspects dealt by the analysis are, among others, note onset, pitch, amplitude, beat and rhythmic density of the input information. The playing is done by six modes of interaction that are consistent to the specific Middle Eastern percussion ensembles improvisatory aesthetics. The robot achieves different timbre sounds by hitting a drum in different positions.

1.3.1 Low level analysis

It is used MAX/MSP and the object bonk to detect and classify the Darbuka hits. Also a beat detector and a density descriptor were implemented.

1.3.2 Modes of interaction

There were implemented six modes of interaction:*Imitation*, *Stochastic Transformation*, *Algorhythm Morphing*, *Beat Detection*, *Syncronized Sequencying* and *Perceptual Accompaniment*. *Imitation* does what it says, imitates after two seconds of pause. In *Stochastic Transformation* the robot divides, multiplies or skips hits that were present in the input signal. The author says this is consistent to the *Middle-Eastern* percussion style.

In the *Algorhythm Morphing* mode, the system gets pre-recorded playings and morphs them together creating also style consistent phrasing. *Beat Detection* serves to let the system adapts itself to the current tempo. The last two modes *Syncronized Sequencying* and *Perceptual Accompaniment* work in parallel with the input. In the first of them the robot plays just a Midi file working as a sinthetizer, in the second, it listen to what is being played and adapts itself. If the density of the input notes is high, the robot play sparser notes, and vice versa. Also the robot can play some call and response games during this mode.

1.4 Voyager from [lewis]

The [lewis] article presents interesting concerns about ethnomusicological questions although does not elucidate much about the underlying methods found in Voyager's system. Clearly the author aims to create a system not hierarchically subordinate to the

musician. Also the aesthetics of the musical output is predefined and programmed prior to the performance. This aesthetics is chosen to be what the author calls "the aesthetic of multidominance" which he claims to come from the African-American culture.

1.5 CHIME

The work of [franklin] is about Jazz improvisation interactive system named *CHIME* which is supposed, like *BoB*, to trade fours with a musician. In this case the author uses a neural networks approach to this problem. It divides the training phase two. First the system begins by learning through a supervised technique. Then in the second part of this phase the system uses reinforcement learning.

The supervised technique consists of using three songs. Four bars of one of these songs are given as input together with the underlying chords. The systems output is compared to the next four bars of this song. This comparison is quantified by an error which is used to train the network's weight during this phase. The result seems to be satisfactory to the author.

The reinforcement phase is done by using as input also the chord information, and a musicians improvisation over the current chord. The system generates a probability distribution to each note choice and with that it generates an output. This is analysed by theoretical based rules that generate a reinforcement value which is used to adjust the next's iterate probability distribution.

It is clear that in this case the musical context is defined previously. In this case even the chords must be given. Also the trade four approach is quite specific to Jazz music.

References

[pachet02] Pachet F.,*"The Continuator: Musical Interaction With Style"*, The Journal of New Music Research, vol.31, No.1, 2002.

[thom00] Thom, B., , in the Proceedings of The 4th International Conference of Autonomous Agents (Agents - 2000), Barcelona, Spain 2000.

[weinberg06] Weinberg, G.; Driscoll, S.; and Thatcher, T.,*"Jam'aa - A Percussion Ensemble for Human and Robotic Players"*, accepted in SIGGRAPH '06, Boston, MA.

[jamaa06] Weinberg, G.; Driscoll, S.; and Thatcher, T.,*"Jam'aa - A Percussion Ensemble for Human and Robotic Players"*, accepted in SIGGRAPH '06, Boston, MA.

[lewis] Lewis, G. E., *"Too Many Notes: Computers, Complexity and Culture in Voyager"*

[franklin] Franklin, J. A., *"Multi-Phase Learning for Jazz Improvisation and Interaction"*