The Impact of the Digital Interface and the Score on Controlled Improvisation When Using Acoustic Instruments in an Electroacoustic Context

Dr Marc Estibeiro
Staffordshire University
m.l.estibeiro@staffs.ac.uk

ABSTRACT

This paper considers various elements that influence the decisions taken when designing performance environments for compositions featuring acoustic instruments with electronics. The implications of choosing different models for the control of the electronics are discussed. Models lying along a continuum between fully automated systems at one end and passive environments controlled by a human performer at the other are considered. The ways in which choices made when mapping control parameters to the sound engine affect the affordances available to the human performer are explored, and the advantages of using a score to define parameters for improvisation are discussed. Following from this, different models for notating the electronic part in mixed compositions are presented and the implication of using the computer as the principal agent for controlling the electronic part are considered. This leads into a discussion as to what constitutes a score for the electronic part. Different paradigms are presented and their implications considered. The paper then discusses the ease of encoding and decoding musical information at different representational levels into and out of a computer. This has implications for machine learning and score following systems. The paper concludes by presenting an example of a digital performance environment from the author's own work, together with a score for controlled improvisation using the interface.

1. MODELS AND PRECEDENTS FOR CONTROLLING ELECTRONICS WHEN USED WITH ACOUSTIC INSTRUMENTS IN REAL-TIME PERFORMANCE

Figure 1 shows some of the various elements that influence the decisions taken when designing the control systems for electronics used with acoustic instruments in real time performance.

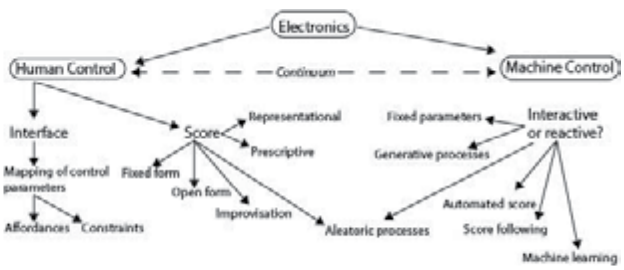

Figure 1: Relationships that influence the design of the digital performance environments

Although the focus in this paper is on mixed compositions for acoustic instruments with electronics, these models are applicable to a wide range of composition and performance situations. It is important to remember that human computer relationships exists on a continuum; there are a range of models for control available, with fully automated autonomous computer systems on one end and passive systems controlled by human performers on the other. When designing a performance environment, the first decision, therefore, is whether the electronics should be controlled by a human performer, by the computer or by some combination lying on the continuum between the two. We begin by considering issues raised if the electronics are to be controlled by a human performer.

2. HUMAN CONTROL OF ELECTRONICS

A human performer will need a means of interfacing with the sound engine of the computer and this raises issues of parameter mapping. The designer of the digital instrument makes explicit or implicit judgements as to which affordances to offer the user of the instrument and what constraints to impose. The performer of the electronics may or may not play from a score. Compositional decisions will determine if there is a need for a score and the nature of the score that may be used. Broadly speaking, and accepting that various combinations are possible, the material will be either improvised, carefully notated, or it will be some combination of an improvised approach and notation. An advantage of this latter approach is that it allows parameters to be clearly defined within which the improvisation can develop. If the performer is encouraged to improvise freely with the material using the performance interface, then, although the results may be aesthetically and artistically satisfying, there may not be clear boundaries set for the improvisation and the performance could depend too much on the improvisational abilities of the particular performer. One result of this could be that the composition is not repeatable in any sense where different performances could still be considered to be of the same piece. (This raises interesting questions as to when a performance of a composition ceases to be "the same" but it is outside the scope of this paper to address these issues). Also, when the composition is repeated with other performers, the lack of guidelines for the improvisations could mean that the results are inconsistent and unsatisfactory, especially if the performers are not already experienced improvisers. We now consider some different models for notating the electronic part.

The score could be in a fixed or an open form (Stockhausen's *Klavierstücke XI* (1956) [9] is one example of an open form composition, there are many other models). The score may also contain any number of other aleatoric processes. Questions then arise concerning the nature of the notation that is to be used and how the score is to be presented. For example, should the score be representative of the sounds produced or should it be purely prescriptive? If representational material is included or excluded in the score, will this influence the performance or the act of composition in any way? The solutions offered to these issues will also have an impact on the design of the electronic interface and the choice of features presented to the performer.

3. MACHINE CONTROL OF ELECTRONICS

If the computer is to be used as the principal agent for controlling the electronics, then models of control afforded to the machine become the issue. Is the computer system passive, reactive, or interactive? Can a computer system, incapable of making aesthetic judgments but arguably capable of mimicking aesthetic judgments, ever be considered truly interactive? Or are there just degrees of reactivity? If a passive model is chosen, which we could refer to as the "effects processor" model, then the source material will be processed by the computer but there will be no updating of the parameters as the performance unfolds. Such a system could, of course, still include parameters which change state over time; a low frequency oscillator, for example. This model may appear at first to be somewhat limited but of course the output of the system will reflect the input. If the input is coming from an acoustic instrument, then all the expressiveness and morphing spectral content of the acoustic performance could still be present in the output. There could also be generative processes built into the software to control the electronics in the composition. These could be triggered at the beginning or at any stage during the performance and allowed to develop over the course of the piece. For these models to work parameters will need to be defined and set, and generative processes will need to be designed and initiated. There is, however, no need for a score in any traditional sense. As soon as models are introduced which do rely on scores then issues such as repeatability and interactivity arise. These are discussed in the next section.

4. THE AUTOMATED SCORE: RECORDING CONTROLLER DATA FOR PLAYBACK DURING PERFORMANCE

If we define the score as some sequence of instructions defining the instrument's state at different points during the composition, then a very simple model for scoring the electronic part for machine performance would be to record controller data during the composition process and to use this as an automated score to be played back during performance. One implication of this approach is that some form of human computer interface, however rudimentary, would need to exist and this again raises the issues of interface design referred to in the previous section. Recording and playing back controller data has an obvious limitation in that unless the source material used during the performance is identical to the source material used during the composition process then the automated controller movements will no longer be relevant. The only way to ensure that the material is identical would be to use a pre-recorded sound file. Furthermore, the movement of the controller data would need to be perfectly synchronised to the playback of the source material. Any performers of acoustic instruments would also have to synchronise exactly to the electronic part. This is all clearly impractical if live acoustic instruments are to be used as the source material.

One possible way to overcome the limitations of such a model would be to use a machine listening and learning system or to use a score following system such as IRCAM's Antescofo [1].

5. MACHINE LISTENING

In machine listening and learning models, the computer is "trained" to recognise certain features of the input material, for example pitch, timbre, amplitude etc., and these are then used as triggers to implement different sets of parameters or generative processes in the software during the performance. This model eliminates or reduces the need for human agency during the performance stage.

A machine listening learning device has to recognise music on a number of levels and the way in which the computer responds when the material has been recognised will determine how interactive or reactive a system will be. Music exists on different representational levels [2] some of which are easier to encode into a computer than others. These are summarised in Figure 2 (adapted from [2]).

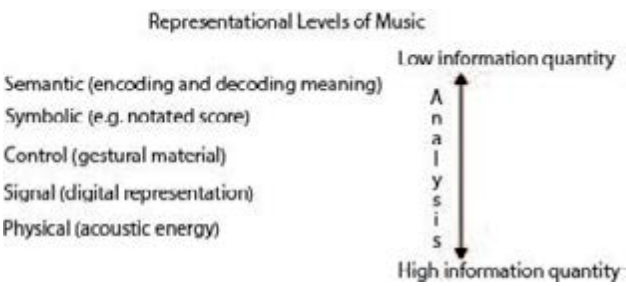

Figure 2: Different representational levels of music

At the top of Figure 2 is the semantic level of representation. This is the hardest level for a computer to encode or decode as it consists of embedded meaning, which a human listener is able to interpret only through a complicated network of implicit knowledge on social, cultural, and psychoacoustic levels. An example of music on the symbolic level would be a notated score in which the semantic level of music is encoded into representational symbols to be interpreted by a performer. Musical gesture (in the sense of performance gesture) exists on the control level. Music at the signal level is perhaps the easiest to represent as this refers to acoustic energy which is converted into voltage with a transducer, sampled using an analogue to digital convertor, and then stored and manipulated. Music on the physical level refers to the physical excitation of a body to produce acoustic energy. All of these different levels of representation will have different data rates on a computer and all will arrive for analysis asynchronously requiring complex sets of relationships in order to analyze the music. There is also a directional hierarchy to the levels. In order to synthesize a musical sound the direction is from the top to the bottom (semantic to physical) while to analyze a musical sound the direction is from bottom to top (physical to semantic). A machine listening learning system, which aims to be at least partly interactive, would need to be able to discriminate between these levels to at least some degree.

An example of a machine listening system is the analyzer~ object developed by the Center for New Musical Research at Berkeley [3]. Analyzer~ is an FFT based Max external which can output values for pitch (as a raw frequency in Hertz and as a MIDI note value), loudness, brightness, noisiness, and Bark scale values. The user can set values to limit amplitude thresholds, pitch deviations (vibrato), re-attack times and harmonic partial weighting. The user can also define window size, hop size and windowing function. It is stable and easy to implement although it is more effective when used to track fundamental frequencies and amplitude variations rather than timbral information.

ll~ is a machine listening learning device (also available as a Max external) developed by Nick Collins [4]. and used in the composition *schismatics II* by Sam Hayden [5]. ll~ was initially developed as a tool for machine-human improvisation and is able to analyse an input stream in order to index different timbral and rhythmic clusters. In *schismatics II* Hayden preloads the ll~ object with preset data files so that it will recognise different timbral clusters produced by an electric violin during the performance. These are then used to trigger different scenes in the electronics, which correspond to different movements of the composition. Hayden reports that the object made the computer's behaviour less predictable, which lead to a greater sense of computer agency and hence increased interactivity between the computer and the performer.

Other important work on timbral recognition has been carried out by Bill Hsu [6].

6. SCORE FOLLOWING

Score following involves the real time synchronisation of an acoustic performance with the score of that performance. Antescofo [1] is an example of a system which analyses a real time audio stream and identifies the score position and tempo from the performance. This information can then be used to synchronise computer generated performance parameters with the acoustic performances [7]. Aligning a real time performance with a symbolic score involves embedding a representation of the score within the program. Antescofo uses its own score format but can also import scores formatted in MIDI or Music XML, both of which are available as export options in most popular music processing packages. Antescofo can follow polyphonic audio streams as well as fundamental pitch (in Hertz). It can also be adapted to accept a MIDI performance as an input. Antescofo can also communicate directly in real time with the score editor Note Ability Pro [8].

There are, however, some potentially serious limitations when dealing with score followers. Real time score following systems can encounter problems when dealing with performance errors (the signal input deviating from the embedded notation: i.e. missing notes, extra notes or wrong notes). There can also be serious problems if synchronisation is lost during a performance. Although Antescofo would appear to be a very stable and reliable system with regard to these issues, the possibility of performance errors is still quite high.

7. DESIGNING AN INTERACTIVE ENVIRONMENT

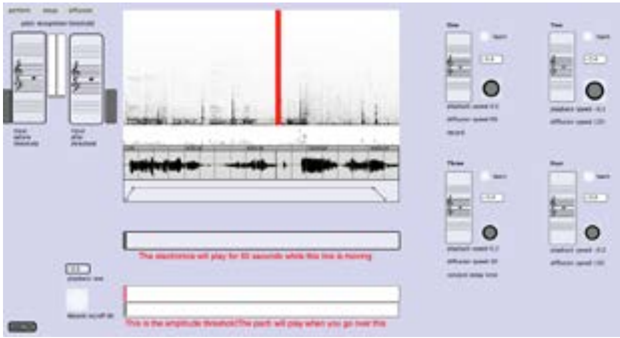

Figure 3: An example of an interactive environment used by the author

Figure 3 shows the environment used for some interactive compositions written by the author. The environment

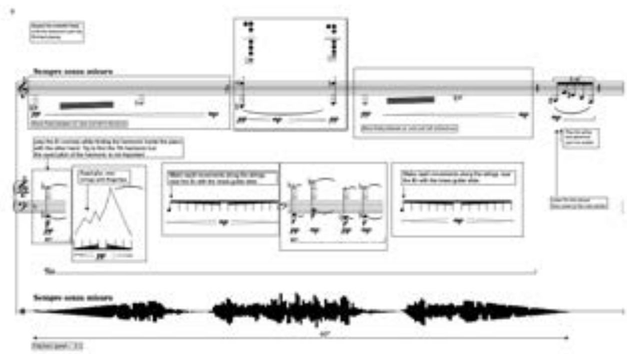

Figure 4: An example of notation for controlled improvisation from the author's own work

was mapped to an FFT based spectral processing patch. The version of the software shown also includes an automated four-channel diffusion system and a delay line with a randomised delay time.

After some experimentation informed by discussions with performers, and after considering their feedback carefully, the following features were incorporated onto the instrument. First of all, the environment was designed to be easy to see in a performance context, taking into consideration the fact that the performer would also be reading from a score. Necessary instructions were kept to a minimum and presented in a bright colour to draw the user's attention to them. A very prominent red scroll bar was also included in order to make it easy for the musician to see the position in the audio buffer from which the computer was reading. More detailed information and instructions were included in a "read me" patch positioned in the bottom corner of the screen. The environment, which featured a machine listening algorithm, was controlled with four programmable trigger notes mapped to initiate events when the amplitude of the input signal exceeded a user-determined setting. It was not possible for the performer to re-map the trigger notes to different parameters. It was a relatively easy task, however, for the composer to reprogram the patch in order to remap the trigger notes to different parameters. This made it easy to reuse the interface in different contexts for different compositions. The number of trigger notes used was kept deliberately low. The following parameters are examples of some typically mappings:

- Trigger note 1: playback speed (0.5), diffusion speed (60), start recording
- Trigger note 2: playback speed (-0.1), diffusion speed (120)
- Trigger note 3: playback speed (0.2), diffusion speed (20)
- Trigger note 4: Playback speed (-0.2) diffusion speed (150)

The environment was designed so that after the amplitude threshold was exceeded, an audio buffer would play for 60 seconds, during which time the performer would be able to change the speed and direction of playback, randomise the delay time, or initiate a new recording. The levels of freedom given to the performer, however, were carefully delineated in the notated score.

8. DEFINING THE BOUNDARIES FOR IMPROVISATION IN THE SCORE

If a composition is to include improvised material, the boundaries for the improvisation must be considered and, where appropriate, included in a score. If the score contains little information as to the nature of the improvisation, then this is mostly left to the discretion of the performer. A performer could find the experience of improvising with an electronic interface empowering and fully engage with the software. Another performer, however, could find the experience intimidating and fail to engage with the environment in any meaningful way. For this reason, as well as other reasons discussed earlier, the boundaries for the improvisatory elements could be carefully defined and embedded in the score.

Error! Reference source not found. shows an example from a composition by the author where a bass clarinet player and a pianist are presented with material for improvisation while an electronic event unfolds. A graphic at the bottom of the page shows a representation of the audio buffer. Also shown is an indication of the length of time for which the buffer will be playing. The performers are presented with a series of gestures in boxes to show that they are to be played independently and not necessarily in order. Verbal instructions are also included which tell the performers to repeat the material freely until the electronic part has finished playing. The marking "sempre senza misura" is included to encourage the musicians to play in free time. As only the bass clarinet player is able to view the interface of the digital instrument, and is therefore the only performer who will be able to see when the electronic part will have finished, an aural cue is included which acts as an instruction to the piano player to move to the next section.

Figure 5: A trigger note notated in the score

The uses of the trigger notes and amplitude thresholds are also very carefully controlled in the composition. They are notated into the score and presented with verbal instructions in order to make their function clear. An example of this is shown in **Error! Reference source not found.**. There are, of course, many precedents for the controlled approach to improvisation described above. Lutosławski, for example, in his composition *Jeux Vénetiens* (1961) makes use of controlled aleatoric techniques where chance processes are allowed to develop within tightly controlled formal and harmonic boundaries.

9. CONCLUSION

This paper has considered various factors that influence the decisions taken when designing digital interfaces for use in compositions that mix acoustic instruments with electronics. Different models for computer control were presented and discussed and the choices made when mapping control parameters to a sound engine were considered. Various methods for notating and controlling the electronic part were presented and their implications discussed. Finally, examples from the author’s own work were given, including a digital interface designed for controlled improvisation in mixed compositions and a score with boundaries for improvisation embedded into the notation. The design of the GUI for the digital instrument used by the author had a significant impact on the experience of the performer and, by extension, the quality of the performance. Improvisation was most successful when clear boundaries were embedded into both the performance environment and the score. Keeping the design of the GUI clear and simple also encouraged the performer to engage more fully with the digital instrument. Presenting the performer with a limited but focused set of affordances encouraged creativity and experimentation. Hybrid human/machine control of the electronic part generally produced the most satisfying results with automated processes being triggered by the performer. Using pitch recognition algorithms to trigger events in the electronic part was an effective strategy which allowed the existing skillset of the performer to be exploited without the need for external controllers.

10. REFERENCES

[1] Cont, A. (2008). Antescofo: Anticipatory synchronisation and control of interactive parameters in computer music. *Proceedings of the International Computer Music Conference.* Belfast: ICMA.

[2] Cont, A. (2010). *Musical Synchrony.* Retrieved 2016 22-April from IRCAM-MRT: http://repmus.ircam.fr/antescofo/videos

[3] CNMAT. (2012). *Center for New Music and Audio Technologies*. Retrieved 2016 24-August from University of California, Berkeley: http://cnmat.berkeley.edu/downloads

[4] Collins, N. (2011). *Listening and Learning in an Interactive Improvisation System.* Retrieved 2016 7-April from www.sussex.ac.uk: http://www.sussex.ac.uk/Users/nc81/code.html

[5] Hayden, S., & Kanno, M. (2011). Towards musical interaction; Sam Hayden's schismatics for e-violin and computer. *Proceedings of the International Computer Music Conference.* ICMA.

[6] Hsu, W. (2010). Strategies for Managing Timbre and Interaction in Automatic Improvisation Systems. *Leonardo Music Journal , 20*, 33-39.

[7] IRCAM. (2007-2016). *Antescofo*. Retrieved 2016 21-April from IRCAM-Musical Representations Team: http://repmus.ircam.fr/antescofo

[8] Hamel, K. (2000-2007). *Note Ability Pro Music Notation Software*. Retrieved 2016 21-April from Debussymusic.ubc.ca: http://debussy.music.ubc.ca/NoteAbility/index.html

[9] Stockhausen, K (1956). *Klavierstücke XI.* Universal Edition UE12654