Don't Just Play it, Grow it! : Breeding Sound Synthesis and Performance Mappings

James Mandelis
Computational Creativity Group
Dept. Informatics,University of Sussex
United Kingdom
+44 7968 005203
jamesm@cogs.susx.ac.uk

Phil Husbands
Computational Creativity Group
Dept. Informatics, University of Sussex
United Kingdom
+44 1273 678 556
philh@cogs.susx.ac.uk

ABSTRACT
This paper describes the use of evolutionary and artificial life techniques in sound design and the development of performance mapping to facilitate the real-time manipulation of such sounds through some input device controlled by the performer. A concrete example of such a system is described which allows musicians without detailed knowledge and experience of sound synthesis techniques to interactively develop new sounds and performance manipulation mappings according to their own aesthetic judgements. Experiences with the system are discussed.

Keywords
Performance mapping, musical interaction, sound synthesis, artificial life.

1. INTRODUCTION
The basic physical aspects of music performance, the generation and manipulation of sounds, have always fed off our technological capabilities. It is no surprise, then that there has been growing interest over the last decade or so in using recently developed technologies from Artificial Life and the more avant-garde areas of Artificial Intelligence to provide new ways of generating and manipulating sounds [Miranda 2002, Griffith and Todd 1998]. This has opened up very interesting musical avenues where various processes for generating sounds, or whole pieces of music, or for controlling aspects of musical performance, can be thought of in terms of interaction with evolving Artificial Life forms.

Musical interaction with artificial life forms can be separated into two broad categories: interaction at note level and interaction at sound level. Interactions at note level usually produce complete musical pieces or music fragments made up of notes that comply with accepted musical and harmonic rules described in modern music theory. The interaction at sound level is concerned with the manipulation of parameters that define a sound using a particular sound synthesis technique (SST), or with parameters that define a particular deformation on an input stream (sound effects).

In the first case the end result is usually constrained by expectations of adherence to a large number of rules that include considerations of structural coherence. In Artificial Life implementations this is achieved either by limiting the music formed by the generation process to a set of legal or valid forms or by using a "judge" subsystem that checks for such validity and rejects pieces that do not conform. In evolutionary terms this can be likened to natural selection. Additional user feedback can be used to steer the course of the evolution and this can be likened to sexual selection where certain characteristics are transmitted to the next generation by being preferentially chosen by prospective mates.

In the second case of sound design such rules tend to be either non-explicit or non-existent. This is partly because of the complexity and lack of transparency of SSTs. In this domain the subjective usually rules over the objective with personal aesthetics acting as the only guide. In Artificial Life implementations this can be achieved by employing the user's aesthetic judgement to power the evolutionary processes that are used to develop the sound generating forms [Dhalstedt 2001, Mandelis 2001, Woolf 1999]. This allows for more 'purist' approaches in terms of artificial evolutionary paradigms -- that is, it is not necessary to encode domain specific knowledge (especially aesthetics-based knowledge) to constrain and guide the process. This is not to say that embedding formalised knowledge of this kind is a bad thing, but in an area such as sound design, where aesthetics are very difficult to formalise, the less constrained approach allows for a powerful exploration of sound space, turning up interesting and unexpected new forms that can be put to good artistic use.

As well as applying Artificial Life techniques to the generation of sounds which are later used in a performance, it is possible to employ them in the closely related area of developing real-time sound parameter manipulation devices for use in performance. This paper concentrates on the unconstrained, exploratory, use of artificial life evolutionary techniques in these two areas.

2. INSTRUMENT EVOLUTION AND PERFORMANCE POSSIBILITIES
A very useful framework for thinking about the core themes of this paper is that introduced by Mulder (1994) to describe the classification and development of musical instruments and their performance interfaces. The first step in instrument development, according to Mulder, involves traditional acoustic instruments that are manipulated in a certain way in order to produce their sounds. The next development is the use of electronics in order to

apply sound effects on acoustic instruments (electroacoustic). The manipulations remain essentially the same. The next step suggested by Mulder is that of Electronic Musical Instruments, where the essential manipulations of a piano (or other MIDI controllers i.e. wind, drums, guitar etc) produce sounds that mimic other acoustic or electronic instruments. His comments on the characteristics of these types of instruments are: "Expanded timbral control, though hardly accessible in real-time and discretized; gesture set adaptivity still limited. Sound emission can be displaced."[Mulder 94]

Mulder's next step, illustrated in Figure 1, involves Virtual Musical Instruments (VMI) where gestures from motion capture devices are used to drive sound engines. His comments on the characteristics of these types of instruments are: "Expanded real-time, continuous timbral control; gesture-set user selectable and adaptive. Any gestures or movements can be mapped to any class of sounds."[Mulder 94]. As a development of the last step, and as an extension to the overall classification, we suggest a new class. It involves VMIs produced by an Artificial Life based framework for adaptive generation of sounds and their gesture mappings.

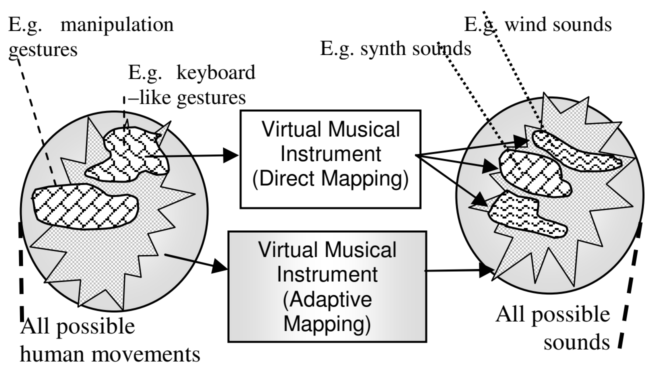

Fig. 1. Steps 4 [Mulder 94] & 5 [Mandelis 01, 02, 03] of instrument development

Genophone [Mandelis 01, 02], which is described in more detail in Section 4, is an example of a system belonging to this new class of Adaptive VMIs. It exhibits the following characteristics: expanded real-time, continuous timbral control; gesture-set and sounds are user designed via an interactive artificial evolution based exploratory search process. Any gestures or movements can be mapped to any class of sounds where both the mappings and the sounds are subject to the same evolutionary forces applied by the user.

3. SOUND SYNTHESIS & PERFORMANCE

Music performed with traditional instruments is the production of sounds whose fundamental frequency corresponds to the note played in a given scale. As such it is normally encoded in a musical score that describes mainly the notes to be played and when they should be played, together with some encoded information describing how these notes are played e.g. legato, fortissimo etc. Identical scores can be interpreted in various ways giving rise to unique performances that are separated by the aesthetic values and the skills of the performer. Some of these differences are temporal, as in micro-fluctuations of the note timing [Longuet-Higgins 82,84], others are qualitative as in modulations of intensity or timbre characteristics affected by skilful manipulations of the instrument. Today, with the widespread availability of music sequencers, the differences between the 'execution' and 'performance' of a piece are more evident than ever. We are all familiar with the mechanical sterile way a musical score can be executed by a computer with 'perfect' timing and 'perfect' pitch. Various commercially available systems have been developed that address this problem by intelligently modulating the timing and the intensity of the notes in accordance with a particular musical style, therefore making a more live-sounding and pleasing musical performance.

This paper focuses on those aspects of musical performance differences that are not encodable in a traditional score, especially in the possibilities of novel expressivities provided by synthesizers and their exploration with Artificial Life paradigms. For a long time synthesisers have been used to emulate traditional instruments and as such they sport pitch-bend and modulation wheels that aid in the expressivity of the instrument. Other parameters of the Sound Synthesis Technique (SST) employed can be modulated by knobs and sliders, giving rise to the now widely accepted practice of 'knob-twiddling' (especially in recent generations of musicians). Music makers have discovered, through trial-and-error, aesthetic values that can be expressed in a way that wasn't possible before: through the modulation of SST parameters. These new expressivities are circumscribed by the SST parameters available for real-time manipulation. Although individual SST parameters are often used for expressivity purposes, it is possible to manipulate multiple values simultaneously. Thus by varying an input parameter (i.e. knob, slider or other control device) a number of SST parameters can be simultaneously controlled, thus defining a 'meta-SST parameter'. At these low-level strata of performance possibilities there is no accepted way or model of how parameter changes can be implemented, as opposed to at the note level where well established theories, models and rules are established.

A particular timbre can be defined as a point in a P-dimensional parametric space, where P is the number of parameters used by the SST engine that produces that timbre. A musical performance can be thought of as an aesthetically pleasing trajectory (or set of trajectories) within that parametric space. For instance, if one of the parameters is the main oscillator frequency, then playing a monophonic melody can be thought of as moving the timbre's point back and forth along that parameter dimension in intervals defined by the scale used. This particular parameter would normally be controlled by the keyboard key position (or equivalent), other parameters do not have such usage expectations associated with them but they can also be used to aid expressivity. Essentially the problem is one of mapping a number of input parameters (I) (i.e. sliders, knobs etc) to a subset (S) of the total number of SST parameters (P), where $I \leq S \leq P$ [Pressing 90, Rovan et al. 97, Wessel 2000]. If each controlled SST parameter has a unique relationship to an input (performance) parameter then a performance subspace is circumscribed within the parametric space, within which an I-dimensional trajectory can be defined as a performance if it satisfies some arbitrary aesthetic sensibilities. This mapping in effect defines an instrument with unique timbral characteristics and expressive manipulative behaviour -- a virtual musical instrument [Machover 89, Mulder 97, Mulder 94, Wessel 2000] .

4. GENOPHONE

To design the kinds of control gesture mappings and timbres described in the previous section is a complex and lengthy affair [Dahlstedt 01]; it involves an intimate knowledge of the SST involved that can be gained usually only after years of experience with the particular SST. Genophone [Mandelis 01, 02] is a system that has been developed to facilitate the design and exploration of such virtual instruments without the need for such detailed knowledge and experience. It uses an Artificial Life paradigm in order to “breed” VMIs and their control mappings. In the current implementation a data-glove is used as an additional control device that provides 5 independent input control parameters that can be modulated simultaneously. The traditional performance control parameters are used also i.e. keyboard, velocity, after-touch, pitch-bend & modulation wheels. Most of these input control parameters can affect multiple synthesis parameters. For instance, the glove input control parameters can control up to 4 synthesis parameters each. The kind of mapping between the input (control) and the output (synthesis) parameters is defined by; lower & upper values and a function of linear, exponential or logarithmic mapping. The mapping of all input parameters in its entirety defines the reactive behaviour of the instrument, in other words its performance characteristics. These performance characteristics and the rest of instruments timbral qualities are subject to breeding. The system is shown in Figure 2.

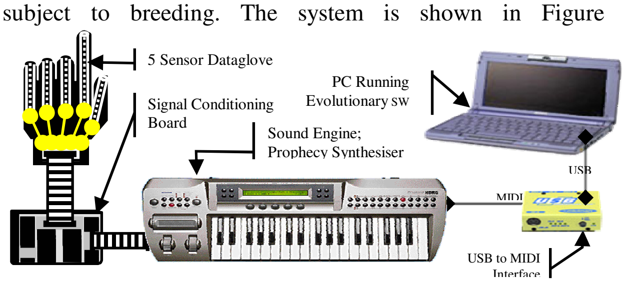

Fig. 2. Genophone; System Setup [Mandelis 01, 02]

During a typical run of Genophone, two (or more) hand designed VMIs are used as seeding parents, these then create a generation of offspring through the application of one (of several) genetic operators. Crossover operators mix parameter values from the two parents to create new individuals, mutation operators randomly change the value of one or more parameters encoded on an individual. After being previewed by the user, the offspring are assigned a relative fitness reflecting how much they are liked by the user. The previewing process involves a fragment of performance so that the user can experiment with the sounds, and the (glove) gesture mapping for manipulating them, that are encoded on the offspring in question. This fitness is used by some of the genetic operators to bias the resulting offspring towards the fitter members of the population. The new generation of offspring is then previewed by the user and a number of them are again selected as parents to create the next generation. Additionally it is possible to allow some other hand-designed parents to enter into the breeding process and contribute towards the next generation. This cycle continues until one or more individuals are deemed satisfactory as VMIs. Genophone has demonstrated that this technique is relatively quick and painless compared to any hand-design method, and that the breeding paradigm is a simple and intuitive one to grasp, while being very powerful. In practice aesthetically interesting and useable VMIs are generated after a few algorithm cycles.

5. Experiences with Genophone

The preliminary results from this project are encouraging and will be followed by system enhancements that will allow more complex experiments to be performed and move into the next phase. Most of the initial aims for this pilot phase have been satisfied and can be summarized as:

5.1 Usage Modes

The system has been used in three distinct modes of operation:

As a Solo Instrument; where the right hand plays a melody on the keyboard while the left hand is changing the sound via the glove. This mode was generally found to be the most difficult to operate for users lacking keyboard virtuosity. Results were often much better for this group of users when they used a sequencer to play familiar melodies via MIDI, while at the same time changing the sound with the glove. This way it is possible take a “sterile” sounding MIDI file and breath a lot of life into it.

As a Single Event / Sound Effect Generator: a single note is played either as drone or until the sound expires. The glove is not often used in this mode with the exception of drones. The Sound Effect Generator mode was much easier to use, it produced a lot of single event sounds that were very rich and dynamic. Quite often they did not play well musically on the keyboard, but had enough structure and complexity to be satisfying, often providing their own melodic or rhythmic framework. Drone sounds where also produced in this way, sometimes using the glove to change their characteristics. Several of these sounds were successfully incorporated into a range of musical settings including some involving traditional instruments as well as the genophone.

As a Pattern Arpeggiator: where an arpeggiated pattern is played with the left hand as a chord, whilst the sound is changed via the glove with the other hand. The Pattern Arpeggiator mode is the most fun to use; it has an instant appeal due to the responsiveness of the glove and the fact that rhythmic structures can be created in a very intuitive way. Also the repetition of the phrase facilitates the perception and prediction of the sound changes within a rhythmic framework.

It is obvious from the above that there is an appropriate mode of operation for creating each part of a track, whether these parts are rhythmic, melodic, drones or single events.

5.2 Hand Rearing vs. Hand Design

The ease of use of the interface was a surprising outcome. The selective breeding paradigm is an accessible one and users were able to breed complex sounds after only a brief introduction. The sounds produced were of such quality that would take someone with quite a bit of experience in the SST involved if they were to be programmed manually, which would be much slower. The overall process tends to be exploratory rather than goal orientated; it is not designed to satisfy a priori sound specifications, i.e. “I would like to produce a bell sound”. It does not preclude the possibility of doing so in indirect ways, though. For instance, if bell or bell-like sounds are used for seeding, then is conceivable that a satisfactory bell sound will be produced within a few generations of selective breeding and variable mutation.

5.3 Meta-SST

Different SSTs can be used without the use of Specific Domain Knowledge. It was an initial requirement that no specific domain knowledge should be used in the system. That is, the parameters are treated as going into a black box, no knowledge of their function is kept in the system. As a result a new SST can be added by just specifying the System Exclusive Implementation Chart of the new synthesiser. As a down side, when sounds are produced that are interesting but very quiet then the there is no evolutionary way in the current framework to address the problem. The only solution is to either selectively breed part of the genotype that is suspected of being responsible, or manually tweak individual values until the desired result is achieved.

5.4 Recombination vs. Mutation

The Evolutionary Paradigm can be successfully applied for the creation of novel sounds often with surprising complexity. It seems that viable (fit) parameter sections are preserved through the genetic recombination, as it is also the case with Genetic Algorithm optimisation. In other words, if the starting sounds are professionally designed ones, then the offspring are likely to be of comparable quality. This is also shown by the observation that genetic recombination produces higher quality results than if mutation is used alone. In implementations [Yee-King 00] where no genetic recombination is used, and mutation or a type of "genetic space crawling" is used instead, it is much harder to produce sounds that are complex and of high (subjective) quality.

6. CONCLUSIONS

This paper has discussed the use of evolutionary artificial life techniques for the interactive exploration of sound-space and its extension to virtual musical instrument space. A concrete example of a system that has successfully demonstrated the efficacy of the approach has been briefly described. It has been argued that artificial life techniques can open up new creative and aesthetic possibilities.

7. Future Directions

It would be interesting to see if the ease of internalising mappings is retained when input devices of more channels are used i.e. more than five. In the future, when different synthesisers and input devices (with more degrees of freedom) are used, the issue of a mapping formalisation will have to be readdressed. Also the two processes for *sound* evolution and *motion-to-sound-mapping* evolution will have to be separated from the same genotype. More operators are currently being developed and tested. Since this is an exploration system each operator has unique properties that can be used appropriately to guide the search.

Acknowledgements: Special thanks to Jon Bird, Andrew Gartland-Jones, Jon McCormack, Alan Montgomery and Sam Woolf for their incisive feedback and stimulating discussions; to Prof. Margaret Boden and the late Prof. Longuet-Higgins for inspiring me on computers and creativity during my most impressionable years, and to Alison Husbands for her tips in glove making.

8. REFERENCES

[1] Dahlstedt, P. (2001), Creating and Exploring Huge Parameter Spaces: Interactive Evolution as a Tool for Sound Generation, Proceedings of International Computer Music Conference, Habana, Cuba.

[2] Griffith, N., and Todd, P.M. (Eds.) (1998). Musical networks: Parallel distributed perception and performance. Cambridge, MA: MIT Press/Bradford Books.

[3] Longuet-Higgins, H.C. and Lee, C.S (1982). The Perception of Musical Rhythms. Perception 11 (1982): 115-128.

[4] Machover, T. & Chung, J. (1989), Hyperinstruments: Musically intelligent and interactive performance and creativity systems, Proceedings International Computer Music Conference, Columbus, Ohio, USA. San Fransisco CA, USA: International Computer Music Association.

[5] Mandelis, J. (2001), Genophone: An Evolutionary Approach to Sound Synthesis and Performance, E. Bilotta, E. R. Miranda, P. Pantano and P. Todd (Eds.) Proceedings ALMMA 2001: Artificial Life Models for Musical Applications Workshop, ECAL 2001, Editoriale Bios, pp. 37-50. http://www.cogs.susx.ac.uk/users/jamesm/Papers/ECAL(2001)ALMMAMandelis.ps

[6] Mandelis, J. (2002), "Adaptive Hyperinstruments: Applying Evolutionary Techniques to Sound Synthesis and Performance," Proceedings NIME 2002: New Interfaces for Musical Expression, Dublin, Ireland, pp. 192-193, 2002. http://www.cogs.susx.ac.uk/users/jamesm/Papers/NIME(2002)Mandelis.pdf

[7] Miranda, E. (2002), Computer Sound Design: synthesis techniques and programming (2nd ed), Oxford: Focal Press.

[8] Miranda, E. (Ed.) (2000), Readings in Music and Artificial Intelligence. Amsterdam: Harwood Academic Publishers.

[9] Mulder, A.G.E., (1994). Virtual Musical Instruments: Accessing the Sound Synthesis Universe as a Performer," Proceedings of the First Brazilian Symposium on Computer Music, pp. 243-250.

[10] Mulder, A.G.E. Fels, S.S. & Mase, K. (1997), Mapping virtual object manipulation to sound variation, IPSJ SIG notes Vol. 97, No. 122, 97-MUS-23 (USA/Japan intercollege computer music festival), pp. 63-68.

[11] Pressing, J. (1990), Cybernetic issues in interactive performance systems, Computer Music Journal, 14 (1), pp. 12-25.

[12] Rovan, J.B. Wanderley, M.M. Dubnov, S. & Depalle, P. (1997), Instrumental Gestural Mapping Strategies as Expressivity Determinants in Computer Music Performance, presented at "Kansei - The Technology of Emotion" workshop.

[13] Wessel, D. and Wright, M. (2000), Problems and Prospects for Intimate Musical Control of Computers, ACM SIGCHI, CHI '01 Workshop New Interfaces for Musical Expression (NIME'01) .

[14] Woolf, S. (1999), Sound Gallery: An Interactive Artificial Life Artwork, MSc Thesis, School of Cognitive and Computing Sciences, University of Sussex