

COMPOSING WITH INTERACTIVE GENETIC ALGORITHMS

Artemis Moroni

*Automation Institute - IA
Technological Center for
Informatics - CTI
CP 6162
Campinas, SP, Brazil – 13081/970*

Jônatas Manzolli

*Interdisciplinary Nucleus for Sound
Communication - NICS
State University of Campinas –
UNICAMP
Campinas, SP, Brazil - 13081/970*

Fernando Von Zuben

*Faculty of Electrical and Computer
Engineering - FEEC
State University of Campinas -
UNICAMP
Campinas, SP, Brazil - 13081/970*

Abstract

This paper addresses the use of IGA in composition. Starting upon theoretical issues and examples of IGA applications in computer graphics it moves to computer music. We concentrate our discussion only on interactive strategies used in Vox Populi to provide dynamic control of the fitness function to evaluate melodic, harmonic and voice range musical features. The fitness is tuned in real time through control curves in a graphic pad.

1 INTRODUCTION

Genetic Algorithms (GA) has been used to generate music structures based on differing approaches. Biles (1994) developed GenJam, a system to simulate a novice jazz musician learning to improvise. Horowitz (1994) generated rhythmic sequences and percussive material based on fitness criteria derived from the user's musical evaluation. Jacob (1995) implemented motive variations using a computer ear based on chromosome like structures to construct musical phrases. In the same direction of these previous system, this article presents a theoretical discussion on Vox Populi; a system based on Evolutionary Computation techniques for composing music in real time (Moroni et al., 1999).

Despite all these research projects have been implemented to reproduce different musical views, they do not dispense human intervention as part of the musical fitness criteria. Applications of GA in artistic domains are often hampered by the lack of an algorithm for determining fitness. A genetic algorithm that uses human judgement to provide fitness is called an interactive genetic algorithm (IGA), a reference to its interactive training cycle. In Computer Graphics, IGAs are increasingly employed to create scenes and animations (Foley et al., 1996; Sims, 1993).

Inside this context, we address the use of IGA in composition. Starting upon theoretical issues such as concept derived from the Darwin's Evolution, we move to applications in computer graphics and computer music. It follows a brief presentation of features of the Vox Populi system (see fully presentation in Moroni et. all, 1999; Manzolli et all, 1999). Here we concentrate our discussion only on interactive strategies used in Vox Populi to provide dynamic control of the fitness function to evaluate melodic, harmonic and voice range musical features. The fitness is tuned in real time through control curves in a graphic pad.

This paper starts with a theoretical point, a brief discussion on complexity and evolution bringing basic concepts of Evolutionary Computation. It follows a non-musical example where IGAs produce interactive evolving graphics, showing IGA as a powerful tool to manipulate images. From this point, we derive a discussion on how IGAs linked to interactive musical strategies can be useful to generate interactively evolving music. All these ideas are elucidated with examples founded in literature and in the Vox Populi system. Finally, we reach to the point of shortly discussing the

relationship between complexity and creativity expanding the use of IGA in computer music applications.

2 SIMULATING EVOLUTION

In “The origin of species” Darwin proposed the theory that all species came about via the process of natural evolution. Evolution is now considered not only powerful enough to bring about the biological entities as complex as humans and consciousness, but also useful in simulation to create algorithms and structures of higher levels of complexity than could easily be built by design. Both biological and simulated evolutions involve the basic concepts of a genotype and a phenotype, and the processes of expression, selection and reproduction with variations. The *genotype* is the genetic encoded information for the creation of an individual. The *phenotype* is the individual itself, or the form that results from the developmental rules and the genotypes. *Expression* is the process by which a phenotype is generated from a genotype. *Selection* depends on the process by which the fitness of phenotypes is determined. *Fitness* is simply a numerical index expressing the ability of an organism to survive and reproduce. In simulation, it can be evaluated by an explicitly defined mathematical function, or a human observer can provide it. *Reproduction* is the process by which new genotypes are generated from an existing genotype.

For the evolution to progress there must be *variations* or mutations in new genotypes having some frequency of occurrence. Mutations are usually probabilistic, as opposed to deterministic. Selection is, in general, non-random and is performed on phenotypes, while variation is usually at random and is performed on the corresponding genotypes. The repeated cycle of reproduction with variations and selections of the fittest individuals drives the evolution of a population towards higher and higher level of fitness. *Sexual combination* allows genetic material of more than one parent to be mixed together in some way to create new genotypes. This permits the features to evolve independently and later on be combined into an individual genotype.

3 INTERACTIVELY EVOLVING GRAPHICS

Procedural models have been increasingly employed in Computer Graphics to create scenes and animations having high degrees of complexity. Procedural models describe objects that can interact with external events to modify themselves. Thus, a model of a sphere that generates a polygonal representation of the sphere at a requested fineness of subdivision is procedural; the actual model is determined by the fineness parameters. A model that determines the origin of its co-ordinate systems, by requesting information from nearby entities, is also procedural. The price paid for this complexity is that the user often loses the ability to maintain sufficient control over the results. A collection of polygons specified by their vertices is *not* a procedural model [Foley et al., 1996].

Procedural models can also have limitations because the details of the procedure must be conceived, understood, and designed by humans. The techniques introduced by [Sims, 1993] contributes towards the solutions to these problems by enabling the “evolution” of procedural models using interactive “perceptual selection”. Evolutionary mechanisms of variation and selection are used to “evolve” complex equations used in procedural models for computer graphics and animation. An interactive process between the user and the computer allows the user to guide evolving equations by observing results and providing aesthetic information at each step of the process. The computer automatically generates random mutations of equations and combinations between equations to create new generations of results. This repeated interaction between user and computer allows the user to search hyperspaces of possible equations, without being required to design the equations by hand or even understand them. Three examples of these techniques were implemented and described: procedurally generated pictures and textures, three-dimensional shapes represented by

parametric equations, and two-dimensional dynamical systems described by sets of differential equations. These methods have been used as powerful tools for exploring procedural models and achieving a flexible complexity with a minimum of user input and detailed knowledge.

A genetic algorithm that uses human judgement to provide fitness is called an interactive genetic algorithm (IGA), a reference to its interactive training cycle. This cycle typically begins with the presentation of the individuals in the current population for the human mentor to experience them. In visual domains, where each individual typically decodes to an image, all the individuals are usually presented at once, often in reduced size so that the entire population can be viewed at once. The mentor can compare and contrast the images concurrently and determine the fitness of each individual in the context of all the others.

4 INTERACTIVELY EVOLVING MUSIC

Research on computer generation of musical material has also received extensive attention. In the musical domain, the temporal evolution of musical events prevents the compressed, parallel presentation of individuals as in computer graphics. Most of the applications of GA to music found in literature, population is seen as an evolving trajectory of music material such as chords, motives, phrases represented as MIDI events.

Biles (1994) presented a genetic algorithm-based model of a student learning to improvise jazz solos under the guidance of a human mentor. The populations represent a co-operating knowledge base of melodic ideas that serve as building blocks for improvised jazz solos. Each individual represents a measure or phrase of music and has to be heard both in an arbitrary harmonic context and in the context of other individuals. GenJam improvises on the tune by building choruses of MIDI events decoded from members of the measure and phrase populations. Since a phrase is implemented as a sequence of four measures, these two populations form a mutually dependent hierarchy of melodic structures. One may think of measure and phrase populations as being its store of melodic ideas or licks. In breeding successive generations of these populations, then, and better collection of licks from which it can construct solos on arbitrary tunes is built.

To improvise a tune, GenJam read a progression file, which provided it with the tempo, rhythmic style (swing or even eighth notes) and the chord progression of the tune being played. It also read a MIDI sequence for the rhythm section, and a MIDI sequence for the head (pre-generated melody and harmony parts). The choruses file told when it should solo, trade fours or eights, rest for a human soloist, or play the head or a pre-written riff. While listening to a solo, the mentor can type one or more `g's if a portion is judged to be good, or one or more `b's if a portion is judged to be bad. The fitness for a given measure or phrase is accumulated by incrementing counters for the currently playing measure and phrase every time a `g' is typed, and decrementing them every time a `b' is typed. The modified fitness values are written back to the population files after the solo terminates.

In training interactively, mentors tend to lose concentration, particularly in early generations when most of the melodic ideas are literally random. At the early stage, mentors tend to reward anything remotely musical and often have a difficult time recognising melodic fragments that might have promise. Sooner or later a few pleasant licks begin to emerge, and one or two solid phrases tend to appear. Typically, at around the eighth or tenth generation, a "golden" generation occurs where almost all the newly hatched phrases sound musical. At this point, the mentor's standards can shift from rewarding anything that sounds vaguely musical to rewarding only what really sounds nice.

The net result for music, then, is that each individual in a population must be presented individually and in real time. This leads to a severe fitness bottleneck, which often limits the population size and the number of generations that realistically can be bred in a musical IGA. These limits are necessary not only to cut down on the length of time it takes to train a musical IGA, but

also to help reduce the unreliability of human mentors as they attempt to sort through the individuals in a population, listening to only one sample at a time.

The human mentor runs into a fitness bottleneck in a musical IGA, because the mentor's task is especially challenging. The ideal mentor would be able to reliably rank the individual members of each population according to their musical merit; however, this is clearly an unaccomplishable goal, given the size of the populations and the inability of mentors to compare individuals easily. Another issue is that individuals can be experienced realistically only in a harmonic context, since the melodic templates only become instantiated to actual notes when played over the chords of a specific tune. Another training issue arises from the tendency of the GA machinery to converge when one highly fit individual emerges early and dominates a population. The set of musically meaningful mutation operators includes mutations that thin out overused measures and reintroduce under-used measures in the phrase population in an effort to promote diversity. Nevertheless, mentors often become tired of an overused lick and start punishing individuals in later generations that had been rewarded heavily in earlier generations.

5 TUNING THE FITNESS FUNCTION

A subclass of the field of algorithmic composition includes those applications which use the computer as a "cross" between an instrument, in which a user "plays" through the application's interface, and as a compositional aid, which a user experiments with in order to generate stimulating and varying musical material. Much of the work that has been done in this field has been based on the idea of determining a set of rules (constraints) which guide the generation of material, rules which are either coded explicitly, or are "learned" by the system through its interaction with the user. Horowitz's development falls into this latter category: given a set of constraining assumptions from which a large number of rhythms can be generated. The system uses an interactive genetic algorithm to learn the user's criteria for distinguishing amongst rhythms. As the system learns (develops an increasingly accurate model of the function which represents the user's choices), the quality of the rhythms it produces improves to suit the user's taste. Interactive genetic algorithms are well suited to solving this problem because they allow to a user to simply execute a fitness function (that is, to choose which rhythms he likes), without necessarily understanding the details or parameters of this function, all that a user needs to be able to do is to evaluate the rhythms.

In a similar approach, but in a completely different way, Vox Populi (Manzolini et al., 1999) is a hybrid made up of an instrument and a compositional environment. The population is made up of groups of four notes, and they are potential solutions for a selection process. Genetic algorithms are used to generate and evaluate a sequence of chords. This sequence produces a sound result resembling a chord cadence or a fast counterpoint of note blocks. These are sent to the MIDI port and can be heard as sound events in real time. Melodic, harmonic, and voice range fitness is used to control musical features. Based on the ordering of consonance of musical intervals, the notion of approximating a sequence of notes to its harmonically compatible note or tonal centre is used.

Tonal centres can be thought of as an approximation of the melody describing its flow. This method employs fuzzy formalism and is posed as an optimisation approach based on factors relevant to hearing music, technically detailed in (Moroni et al., 1999). In the selection process, the group of voices with the highest musical fitness is selected and played. The musical fitness of each chord is a conjunction of three partial fitness functions: *melody*, *harmony* and *voice range*.

$$\text{Musical Fitness} = \text{Melodic Fitness} + \text{Harmonic Fitness} + \text{Voice Range Fitness}$$

Differently from other systems found in genetic algorithms or evolutionary computation, in which people have to listen to and judge the musical items, Vox Populi uses the computer and the mouse as real-time music controllers, acting as a new interactive computer-based musical instrument. The interface is designed to be flexible for the user to modify the music being generated. It explores Evolutionary Computation in the context of Algorithmic Composition and provides a graphical interface that allows modifying the tonal centre and the voice range, changing the evolution of the music by using the mouse. These results reflect current concerns at the forefront of Computer Music in interactive composition and in the development of new controller interfaces.

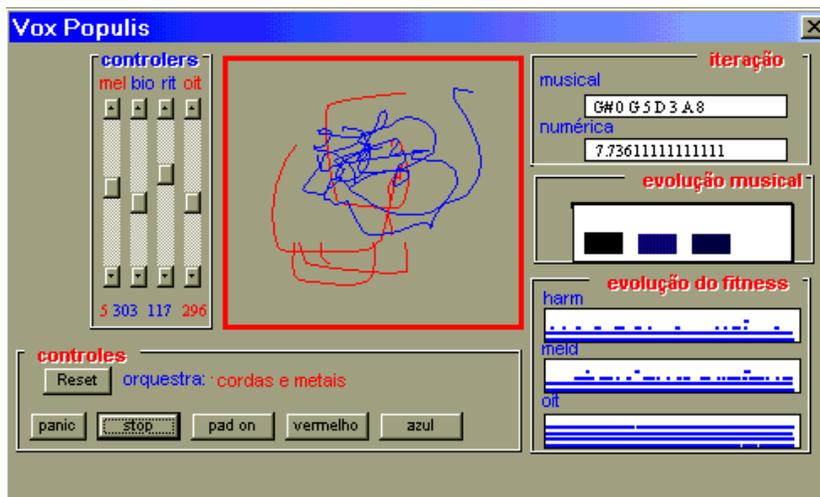


Fig. 1 – Vox Populi Interface

In Vox Populi, the resultant music moves from very pointillist sounds to sustained chords; it depends upon the duration of the genetic cycle and the number of individuals of the original population. The octave fitness forces the notes to be in the range reached by human voices and associated with the centre region of the notes in the piano, but since several orchestras of instruments are used, this range is too limited for some of them.

The interactive pad control supplies a graphical area in which two-dimensional curves can be drawn. These curves, a blue and a red one, links to them other controls of the interface, as depicted in **Fig. 1**. The red curve links the melodic and octave range control; and the blue curve links the biological and rhythmic controls. When the interactive pad is active, the four other linked controls are disabled. Each curve describes a phase space between the linked variables. They are traversed in the order they were created; their horizontal and vertical components are used for fitness evaluation and to modify the duration of the genetic cycles, interfering directly in the rhythm of the composition. The pad control allows the composer to conduct the music through drawings, suggesting an elementary conductor gesture. By different drawings, the composer can experience the generated music and conduct it, trying different trajectories or sound orbits.



Fig. 2 – A score of a MIDI file raw material produced by Vox Populi

The pad control may be considered as one more feature to the mentor for evaluating the music. The composer does not need to “hear” the music to guess how it occurs; he is able of “mentally” hearing by just reading a score. Even an interpreter creates a mental schema to memorise the music for a performance. Fulvia Escobar, well-known erudite Brazilian pianist uses a “gate” metaphor when a musical sequence changes to another musical sequence, and memorises the order of the gates. In Vox Populi, pad drawings can be associated to musical sequences allowing the mentor to use them as an additional evaluating feature.

The idea of producing two-dimensional pictures from a one-dimensional sound source is not new. Pronovost and colleagues describe a real-time process using analogue circuits which produces two-dimensional images (Pronovost, 1968). Pickover describes a different technique for obtaining two-dimensional images (Pickover, 1986). These works were aimed at helping deaf people to speak. Pellegrino describes a number of techniques for mapping sound into visual images (Pellegrino, 1983). More recently, Monro and Pressing examined the adaptation of a standard technique of mathematical analysis for the representation of sound (Monro & Pressing, 1998). This technique is commonly referred to as *embedding*, or the method of delays, or the pseudo-phase space method, and shows that this method provides a compact way to represent multi-dimensional correlation of a musical signal; it also readily produces intricate colour plots and movies that, in their view, are striking displays of sonic visual art.

Vox Populi uses the computer and the mouse as real-time music controllers, acting as a new interactive computer-based musical instrument. The interface is designed to be flexible for the user to modify the music being generated. To further remove the necessity of human interaction in the algorithmic composition process, the criticism used in evolving artificial composers can be trained using easy-to-collect musical examples, rather than constructed using difficult-to-determine musical

rules. Baluja, Pomerleau and Jochem (Baluja et al., 1994), working in the visual domain, have trained a neural network to replace the human critic in an interactive image evolution system similar to that created by Sims. The network “watches” the choices that a human user makes when selecting two-dimensional images from one generation to reproduce in the next generation, and over time learns to make the same kind of aesthetic evaluations as those made by a human user. Since Vox Populi is a hybrid made up of an instrument and a compositional environment, the network can accurately follow human choices during the process of composition.

Vox Populi has been used in different ways. As an autonomous or demonstrative system generating music. As a sound laboratory, where people can try and experience sound. In a studio, being manipulated and generating samples which were used in compositions and in soundness landscapes. Another way that is being considered is coupled with sensors, allowing describing orbits in the space. This system received at the Dream Centenary Computer Graphics Grand Prix 99 award in Aizu, Japan, as an Interactive Application (see www.ia.cti.br/~artemis/voxpathuli).

6 CONCLUSION

Psychologically, one of the great powers of computer programming is the ability to define new compound operations in terms of old ones, and to do this over and over again, thus building up a vast repertoire of ever more complex operations. It is quite reminiscent of evolution, in whichever more complex molecules evolve out of fewer complexes ones, in an ever-upward spiral of complexity and creativity. It is also quite reminiscent of the industrial revolution, in which people used very simple early machines to help them build more complex machines, then used those in turn to build even more complex machines, and so on, once again in an ever upward spiral of complexity and creativity. At each stage, whether in evolution or revolution, the products get more flexible and more intricate, more “intelligent” and yet, more vulnerable to delicate “bugs” or breakdowns (Hofstadter, 1985).

Evolution is a method for creating and exploring complexity that does not require human understanding of the specific process involved. This process of interactive evolution could be considered a system for helping the user with creative explorations, or it might be considered a system that attempts to “learn” about human aesthetics from the user. In either case, it allows the user and computer to work together interactively in a new way to produce results that neither could easily produce alone.

From the discussion of on the use of a graphic pad as a tool for controlling the evolution of a Fitness Function, it is possible to draw some conclusions. Vox Populi interface, designed to be flexible for the user to modify the music being generated, might be appropriate to study human musical perception. The choices interactively made by a composer in response to the evolving music can be stored as a parametric control file and recorded as a musical signal as well. So these data can be compared and analysed. This also can be applied to training neural networks, which in turn may be used as fitness functions. The mappings are easy to produce and give detailed pictures of musical sequences. Finally, the use of IGA in compositional systems, is a powerful tool to control the complexity of the music material in the flow. IGA associated to other interactive strategies, such as the graphic environment used in Vox Populi, could be a strategy to avoid the musical fitness bottleneck. High dimensional curves would lead the musician to have overall view of the musical evolution.

Acknowledgements

Part of this project is possible through the support of FAPESP to the Gesture Interface Laboratory in which Vox Populi was developed. Fernando J. Von Zuben is supported by the CNPq grant 300910/96-7. Artemis Moroni is supported by CTI.

References

- Atmar, W. (1994).** Notes on the Simulation of Evolution, *IEEE Transactions on Neural Networks* Vol. 5, No. 1, 130 --147.
- Baluja, S., D. Pomerleau and T. Jochem (1994).** Towards automated artificial evolution for computer-generated images, *Connection Science*, 6(2-3), 325 -- 354.
- Biles, J. A. (1994).** GenJam: A Genetic Algorithm for Generating Jazz Solos, *Proceedings of the 1994 International Computer Music Conference*, (ICMC'94), 131 -- 137, 1994.
- Fogel, B. (1995).** Evolutionary Computation - Toward a New Philosophy of Machine Intelligence, USA: *IEEE Press*, USA, 46 -- 47.
- Foley, J. D., Andries van Dam, Steven K. Feirner and John F. Hughes. (1996).** Computer Graphics Principles and Practice, *Addison-Wesley Publishing Company*, p. 1018.
- Hofstadter, D. R. (1985).** Methamagical Themas, USA: *Basic Books*, p. 693.
- Horowitz, D. (1994).** Generating rhythms with genetic algorithms, *Proceedings of the 1994 International Computer Music Conference*, 142 -- 143.
- Moroni, A., J. Manzolli, F. Von Zuben, R. Gudwin (1999).** Evolutionary Computation applied to Algorithmic Composition, *Proceedings of CEC99 - IEEE International Conference on Evolutionary Computation*, Washington D. C., p. 807 – 810.
- Manzolli, J., A. Moroni, F. Von Zuben, R. Gudwin (1999).** An Evolutionary Approach Applied to Algorithmic Composition, *Proceedings of SBC'99 - XIX National Congress of the Computation Brazilian Society*, Rio de Janeiro, Vol. 3, 201 -- 210 (1999).
- Sims, K. (1993).** Interactive evolution of equations for procedural models, *The Visual Computer* Vol. 9, No. 9, 466--476.
- Pellegrino, R. (1993).** The Electronic Arts of Sound and Light, *New York: Van Nostrand*.
- Pickover, C. (1986).** On the Use of Symmetrized Dot Patterns for the Visual Characterization of Speech Waveforms and Other Sampled Data, *Journal of the Acoustical Society of America* 80(3): 955 – 960.
- Pronovost, W., L. Yenkin, D. Anderson and R. Lerner (1968).** The Voice Visualizer, *American Annals of the Deaf*, 113: 230 – 238.