Interactive Control of Evolution Applied to Sound Synthesis

Caetano, M.F.[1,2] , Manzolli, J.[2,3] , Von Zuben, F.J.[1]

1 Laboratory of Bioinformatics and Bioinspired Computing (LBiC)/DCA/FEEC – PO Box 6101
2 Interdisciplinary Nucleus for Sound Communication (NICS) – PO Box 6166
3 Music Department at the Arts Institute (DM/IA)– PO Box 6159
University of Campinas – Campinas – SP – 13083-852 – Brazil
{caetano,vonzuben}@dca.fee.unicamp.br
{caetano,jonatas}@nics.unicamp.br

Abstract

In this paper, we present a sound synthesis method that utilizes evolution as generative paradigm. Such sounds will be thereon referred to as evolutionary sounds. Upon defining a population of complex sounds, i.e. sound segments sampled from acoustical instruments and speech; we generated sounds that resulted from evolution applied to those populations. The methodology presented here is an extension to the Evolutionary Sound Synthesis Method (ESSynth) created recently. In ESSynth, a set of waveforms, the Population, is evolved towards another set, the Target, through the application of a Genetic Algorithm (GA). Fitness evaluation is a mathematical distance metric. We enhance features of the previous implementation herein and present the codification. The genetic operators and selection criterion applied are depicted together with the relevant genetic parameters involved in the process. To evaluate the results we present a sound taxonomy based on an objective and a subjective criterion. Those criteria are discussed, the experimental procedure is explained and the results are depicted and evaluated.

Introduction

Motivation

Music composition is a creative process that, here, can be described in terms of an aesthetical search in the space of possible structures which satisfy the requirements of the process (Moroni 2002); in this case, to generate interesting music. In a broad sense, our view of sound synthesis is a digitally controlled process that produces signals that can be used in musical applications. The main objective of our research is to verify the musical potential of a specific set of mathematical tools used to implement objective functions and search operators in evolutionary computation processes.

Complex sounds are remarkably difficult to generate for they pertain to a distinctive class of sounds that present certain characteristics. Such sounds usually have dynamic spectra, i.e. each partial has a unique temporal envelope evolution. They are slightly inharmonic and the partials possess a certain stochastic, low-amplitude, high-frequency deviation. The partials have onset asynchrony, i.e. higher partials attack later than the lower ones. Our ears are highly selective and often reject sounds that are too mathematically perfect and stable.

Dawkins (1986) describes lucidly how natural selection can lead to a building up of complexity, beauty and efficiency of design. Dawkins also extends this notion to a form of "Universal Darwinism" (Dawkins 1989) in which the creative generation of intellectual ideas itself derives from such process of iterative refining of ideas competing with one another, or "memes". Compositions tend to exhibit various structural degrees, where the composer sculpts the initial idea to transform it into satisfactory final products. In this work, we are viewing music composition as a process that can be adequately modeled by the evolutionary paradigm manipulating the generation of complex sounds, driving the sonic process to potentially the same diversity found in nature.

Evolution

The first attempt at representing Darwin's theory by means of a mathematical model appeared in the book "*The Genetic Theory of Natural Selection*" (Fisher 1930). Later on, Holland (1975) devoted himself to the study of adaptive natural systems with the objective of formally studying the phenomenon of adaptation as it occurs in nature. Genetic Algorithms (GAs), were proposed by Holland (1975) to indicate that adaptation mechanisms could be used in computers.

Success in the employment of evolutionary techniques is best found in the work of William Latham and his systems *Mutator* and *Form Grow* to create sculptures in 3D on the computer (Latham and Todd 1992). Thywissen (1993), inspired by the works of Dawkins and Latham, describes a successful attempt at transferring these evolutionary concepts to the domain of music composition guiding the composer through the sample space of possibilities. Evolutionary mechanisms, not restricted to GAs, have demonstrated to be extremely efficient means of "blindly" searching an acceptable structural candidate in large sample spaces.

An application of GAs to generate Jazz solos is described by Biles (1994) and this technique has also been studied as a means of controlling rhythmic structures (Horowits 1994). There is also a description of an algorithmic composition procedure in *Vox Populi* (Moroni et al. 2000) based on the controlled production of a set of chords. It consists in defining a fitness criterion to indicate the best chord in each generation. *Vox Populi* was capable of

producing sounds varying from clusters to sustained chords, from pointillist sequences to *arpeggios*, depending upon the number of chords in the original population, the duration of the generation cycle and interactive drawings made by the user on a graphic pad.
Fornari et al. (2001) worked on a model of Evolutionary Sound Synthesis using the psychoacoustic curves of loudness, pitch and spectrum, extracted from each waveform that represents an individual of the population. The psychoacoustic curves map genotypic into phenotypic characteristics. That is, they relate physical aspects of the chromosomes to the corresponding psychoacoustic attributes. Reproduction and selection are done on the phenotypic level, similarly to what is done in nature, instead of in genotype.

Paper Structure

In the next section we focus on the generation of evolutionary sounds. There is a brief overview of GAs and the biological concepts that inspire this approach. Interactive Genetic Algorithms (IGAs) are briefly explained and their association with exploratory creative processes is discussed, along with the possibility that our system also be regarded as a creative process. The genetic parameters and operators are also presented, along with the codification. Then, we emphasize the different aspects to be considered when evaluating evolutionary sounds. We propose a tentative sound taxonomy as a means of classifying the results. Finally, the results are shown and analyzed according to the criteria adopted along the development of the method.

Generating Evolutionary Sounds

Genetic Algorithms

GAs are the most commonly used paradigm of Evolutionary Computation (EC) due to the robustness with which they explore complex search spaces. GAs are techniques of computational intelligence that imitate nature in accordance with Darwin's survival of the fittest principle. They (qualitative or quantitative) codify physical variables via digital DNA on computers. The resulting search space contains the candidate solutions, and the evolutionary operators will implement exploration and exploitation of the search space aiming at finding quasi-global optima. The evolutionary process combines survival of the fittest with the exchange of information in a structured yet random way. The better its performance in the solution of a determined problem, the more efficient a GA is considered, no matter the degree of fidelity to biological concepts. In fact, the majority of the algorithms that follow this approach are extremely simple on the biological point of view, though they are associated with extremely powerful and efficient search tools.
The GA iteratively manipulates populations of individuals at a given generation by means of the simple genetic operations of selection, crossover and mutation. Taking the reproduction rate of the individuals directly proportional to performance, the fittest individuals tend to eventually dominate the population. Therefore its superior genetic content is allowed to disseminate in time.
We understand applications of GAs in computer music as a means of developing aesthetically pleasant musical structures. The user will be responsible for the subjective evaluation of the degree of adaptability, thus implementing a kind of fitness function.

Existence of a Target Set

The method consists of the generation of two distinct sets of waveforms, Population and Target (Manzolli et al. 2001b). Previously, these sets were initialized at random (Manzolli et al. 2001a). Presently, the user is allowed to load up to 5 waveforms to each of these sets. Each individual in these sets is codified as a chromosome composed of 1024 samples of a given waveform at a sampling frequency of 44100 samples per second. This is equivalent to a wave-format sound segment of approximately 0.0232s.
Evolution drives the individuals in the Population towards the individuals in the Target set. In ESSynth, the waveform is the genetic code that carries all the information regarding the sound and can be manipulated. The resultant timbre, or the way the sound distinguishes from others, is the characteristic that can be perceived. In this sense, the genotype (i.e. the waveforms in the populations) is changed, but the phenotype (i.e. the overall timbre) is preserved producing a variant. Therefore, these two elements are integrated as in biological evolution, which uses genetic information to generate new individuals.

Parametric Control of Sound Generation

In traditional GAs fitness can be encoded in an algorithm, but in artistic applications, fitness is an aesthetic judgment that must be made by a human, usually the artist. The idea of using EC with a human loop first occurs in the works of Dawkins (1989) and Latham and Todd (1992). This approach was entitled Interactive Genetic Algorithm (IGA), where a human mentor must experience the individuals in the population and provide feedback that either directly or indirectly determines their fitness values. Like Biles (1994) observed, this is frequently the bottleneck in a GA based system because they typically operate in relatively great populations of candidate chromosomes, where the listener must evaluate each individual.
When one considers the implementation of a GA, the challenge is to find an interesting representation that maps the characteristics of the chromosome in musical features such that music can be gradually evolved. The fitness function, previously objective, is to be reinterpreted as subjective; the taste and judgment of the composer start to dictate the relative success of musical structures competing among themselves.
In ESSynth, the fitness function is given by a mathematical metric to avoid the burden of evaluating each individual

separately in each generation. However, the user is free to explore the search space by choosing the Population and Target sets. This can be used both to search the space of possible structures in an exploratory way, and to search the space for a particular solution. The former considers an evolvable Target set, and the latter makes use of a fixed Target set. The idea of adopting static or dynamic templates has since been applied in music, in computer-aided design (Bentley 1999) and in knowledge extraction from large data sets (Venturini et al. 1997).

It is the genetic operators that transform the population along successive generations, extending the search until a satisfactory result is reached. A standard genetic algorithm evolves, in its successive generations, by means of three basic operators, described as follows.

Selection: The main idea of the selection operator is to allow the fittest individuals to pass its characteristics to the next generations (Davis 1991). In ESSynth, fitness is given by the Hausdorff (multidimensional Euclidean) distance between each individual and the Target set. The individual in the Population with the smallest distance is selected as the best individual in that generation (Manzolli et al. 2001a).

Crossover: It represents the mating between individuals (Holland 1975). The central idea of crossover is the propagation of the characteristics of the fittest individuals in the population by means of the exchange of information segments between them, which will give rise to new individuals. In ESSynth, crossover operation exchanges chromosomes, i.e. a certain number of samples, between the best individual in each generation and each individual in the Population. The segments are windowed to avoid glitch noise. After the operation of crossover, each individual in the Population has sound segments from the best individual.

Mutation: It introduces random modifications and is responsible for the introduction and maintenance of genetic diversity in the population (Holland 1975). Thus, mutation assures that the probability of arriving at any point of the search space will never be zero. In ESSynth, mutation is performed by adding a perturbation vector to each individual of the population. The amplitude of this vector is given by the coefficient of mutation. This operator introduces a certain noisy distortion to the original waveform.

Genetic Parameters

The user will adjust the search according to predefined requirements achieved by the manipulation of the parameters that follow.

Size of the Population: The size of the population directly affects the efficiency of the GA (Davis 1991). A small population supplies a small covering of the search space of the problem. A vast population generally prevents premature convergences to local solutions. However, greater computational resources are necessary (Davis 1991). We used five individuals in the Target and Population sets.

Coefficient of Crossover: The higher this coefficient, the more quickly new structures will be introduced into the population. But if it is very high, most of the population will be replaced, and loss of structures of high fitness can occur. With a low value, the algorithm can become very slow (Davis 1991). In this implementation of ESSynth, the coefficient of crossover is an internal parameter and cannot be affected by the user. It defines how much of the best individual will be introduced in each individual in the next generation.

Coefficient of Mutation: It determines the probability of mutation. A properly defined coefficient of mutation prevents a given position from stagnating in a particular value besides, making it possible for the candidate solutions to explore the search space. A very high coefficient of mutation causes the search to become essentially random and increases the possibility of destroying a good solution (Davis 1991). In ESSynth, the coefficient of mutation ranges from 0 to 1.

Evaluating Evolutionary Sounds

Sound Taxonomy

The process of sound perception is remarkably non-trivial. Schaeffer (1966) introduced the idea of timbre classification distinguishing sounds between form and matter in the context of concrete music. Later, Risset (1991) associated the concept of form to the loudness curve and matter to the magnitude of the frequency spectrum of the sound. Risset (1991) stated that Schaeffer’s concept of form is the amplitude envelope of the sound and matter the contents of the frequency spectrum. This has perhaps been the first attempt at describing the timbre nature of sound. Nowadays it is known that the frequency spectrum of sound varies dynamically with time (Risset 1966), and cannot be adequately defined by such a static concept as matter. The dynamic changes of the frequency spectrum carry important information about the sound itself. Smalley (1990) declared that the information contained in the frequency spectrum cannot be separated from the time domain once “*spectrum is perceived through time and time is perceived as spectral motion*”. Risset (1991) declared that sound variants produced by changes in the synthesis control parameters are intriguing in the sense that usually there is not an intuitive relation between parametric control and sound variation. We feel it is extremely important for the user to be able to relate subjective characteristics of sound to the input parameters of the method in order to better explore the sound-space towards a desired result. For such, an Objective Criterion and a Subjective Criterion were chosen to classify the results. Finally, the outcome of both experiments was cross-correlated. This analysis shall serve as the basis for future applications of ESSynth in musical composition. Thus we defined:

Objective Criterion: evolution of the partials in time (spectrogram) and energy displacements.
Subjective Criterion: classification made by trained listeners in accordance with a scale of values that relates sounds with qualitative aspects.
The scale of values was inspired by the works of Gabrielsson (1981) and Plomp (1970) and represents some timbre dimensions that are commonly adopted.

Results

The output sound set resulting from a run of the program will be shown and discussed. The result of both the Objective and Subjective analyses will be presented individually and then cross-correlated.
We expected the output sound to be a timbral merger of the individuals in the Population and Target sets. So, the results of neither criterion alone shall suffice the classification purposes. Only by combining the analyses will one be able to decide whether the output sound actually presents characteristics from both Population and Target sets, representing a variant.
The genetic parameters adopted were 20 interactions (generations), 5 individuals in both the Population and Target sets and coefficient of mutation of 0.87. The coefficient of mutation is rather high and was chosen so as to reinforce the transformations induced by the method. All these values were obtained empirically by running the program a number of times and analyzing the results.

Objective Criterion

A case study will be presented with quite a significant result that is thought to represent the transformations generated by the method. The Population waveforms utilized are tenor sax sounds shown in Figure 1 a. The Target waveforms are cicada sounds shown in Figure 1 b.

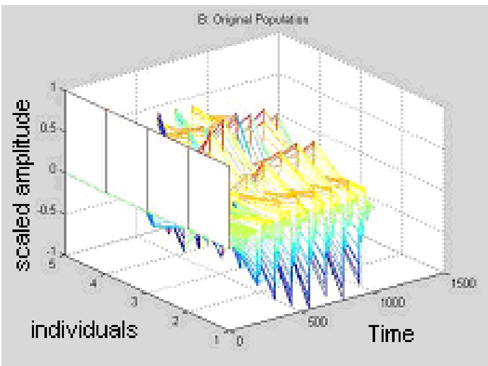

Figure 1a: Population waveforms. Z-axis represents amplitude scaled in the interval [-1,1], Y-axis is the number of individuals (5 in each population) and X-axis is the number of samples (time scale)

Notice that although the individuals in Figure 1 appear as a surface they are actually separate entities along the axis labeled individuals.

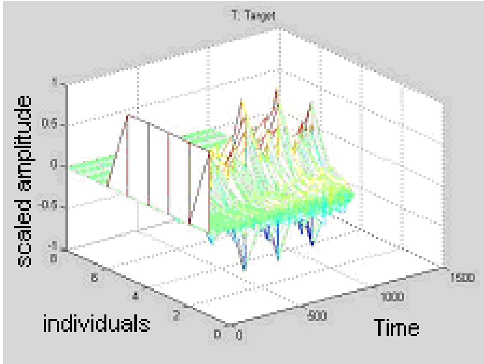

Figure 1b: Target waveforms. Z-axis represents amplitude scaled in the interval [-1,1], Y-axis is the number of individuals (5 in each population) and X-axis is the number of samples (time scale)

These particular sounds were chosen so as to highlight the transformations once the characteristics of both the Population and Target sets are thought of as distinctively different. The Population is sonorous and has spectral contents harmonically distributed, because they are derived from an acoustic musical instrument, while the individuals in the Target set are rather noisy and inharmonic.
Next, the spectrograms of one of the individuals of the Population (B_orig5) and Target (T1) sets are shown in Figure 2.

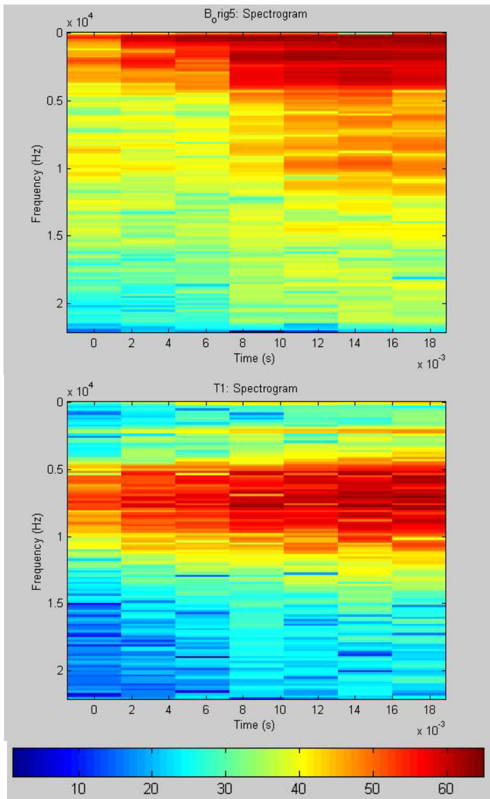

Figure 2: Spectrogram of B_orig5 (top), and of T1 (bottom). X-axis is time in seconds, Y-axis is frequency in Hertz (increasing from top to bottom) and intensity is represented in the scale shown below the figure

The resultant waveforms obtained after 20 generations are shown below in Figure 3, as well as a representative spectrogram.

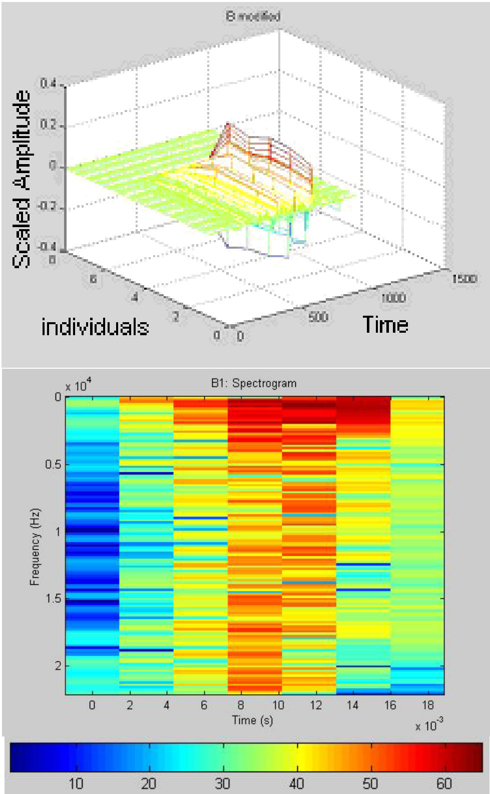

Figure 3: Result after 20 generations. All the individuals of the Modified Population, i.e. output sound (top), thereon denominated B_modified, are shown. The frequency spectrogram of the first individual of B_modified, denominated B1, is shown (bottom)

It is interesting to notice that the waveforms were sculpted by the program due to the fact that crossover is applied, in average, in the middle of the waveform. The spectrogram features characteristics from the Population and from the Target. The result is a merger of aspects of similar nature to B_orig5 and T1, shown in Figures 1 and 2. One individual from B_modified (B1) was chosen to represent the spectral transformations.

Subjective Criterion

The results of the Subjective experiment are presented concerning the scale of qualitative values. Although subjective, the chosen scale is part of the perceptional context of the individuals used in the experiment. The estimation of a subject is shown in Table 1 and was chosen to represent the overall result. The individuals taken into consideration throughout the text are highlighted in the table. The result of this analysis is only considered for B1, i.e. the individual whose spectrogram is shown at the bottom of Figure 3.

Figure 4 shows the different levels each dimension was quantified into. Figure 5 depicts the result of the subjective experiment highlighting the characteristics passed on to B1 and from which of the sets, Population or Target, it probably inherited them.

Cross-correlating the information of the spectral analysis with the information from the qualitative analysis, it can be inferred that the output wave presents an intermediate spectrogram with an intermediate subjective evaluation. This can be considered as a form of spectral crossover resulting from ESSynth. Cross-correlating the results of the Objective and Subjective analyses one can infer that the presence of concentrated harmonic spectral components in low frequencies can be associated with the subjective classification of bright given to the sax sounds. The presence of inharmonic spectral components, due to the lack of defined pitch and great spectral power density, can be associated with the quality of noisy given to the cicada sound. The final sound acquired noisy characteristics, maintaining, however, its brightness, probably due to the presence of harmonic frequencies between 0Hz and 5KHz that preserved some of the subjective characteristics of the sax sounds. Intuitively, it can be stated that the resultant sound is a spectral mixture that, using a biological terminology, is a crossover process between the two populations.

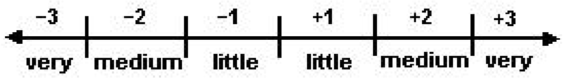

Figure 4: Scale of values for crossing the subjective terms

	Brightness×Dullness	Sharpness×Softness	Fullness×Thinness	Clearness×Noisiness
	Very Medium Little	Very Medium Little	Very Medium Little	Very Medium Little
B_orig1	Medium Dull	Medium Sharp	Little Thin	Little Clear
B_orig2	Medium Dull	Little Soft	Medium Thin	Medium Noisy
B_orig3	Little Dull	Very Sharp	Medium Full	Medium Noisy
B_orig4	Medium Dull	Very Sharp	Little Full	Little Clear
B_orig5	Medium Dull	Little Soft	Little Full	Little Clear
T1	Very Bright	Medium Sharp	Medium Full	Medium Noisy
T2	Very Bright	Very Sharp	Very Full	Medium Noisy
T3	Very Bright	Medium Sharp	Medium Full	Medium Noisy
T4	Very Bright	Very Sharp	Little Full	Medium Noisy
T5	Very Bright	Little Sharp	Little Full	Medium Noisy
B1(Output)	Very Bright	Little Soft	Medium Full	Medium Noisy

Table 1: Result of the fourth subject's estimation of the presented samples

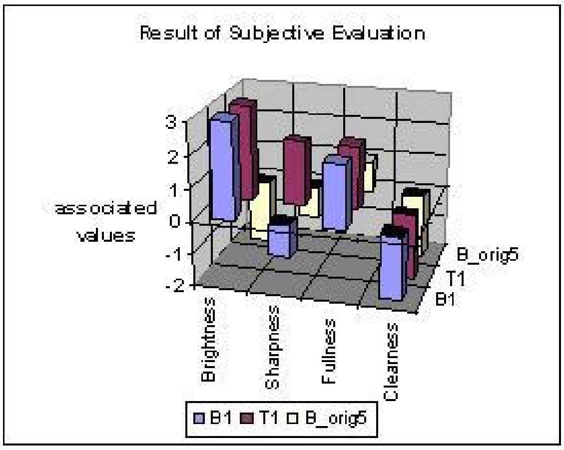

Figure 5: Graphic depiction of the result of the subjective evaluation

Conclusion

An extension to the original Evolutionary Sound Synthesis Method for complex sounds was presented and the results obtained were shown and evaluated. These results were analyzed observing waveform and spectral transformations caused by the method. The method can be regarded as a novel framework for timbre design. It is a new paradigm for Evolutionary Sound Synthesis for it incorporates subjectivity by means of interaction with the user.

Many extensions can still be envisioned and tested. It can be used to compose soundscapes, as a timbre design tool or in live electroacoustic presentations where an evolutionary timbre is generated, which evolves in real time along with the evolution of other music materials.

Future trends of this research include using co-evolution as generative paradigm, experimenting other distance metrics (fitness) and even other bio-inspired approaches applied to sound synthesis.

Acknowledgements

The authors wish to thank FAPESP (process 03/11122-8) and CNPq for their financial support to this research.

References

Bentley, P.J. 1999. *Evolutionary Design by Computers*. San Francisco, CA: Morgan Kaufmann.

Biles, J. A. 1994 GenJam: A Genetic Algorithm for Generating Jazz Solos. In *Proc. of the 1994 International Computer Music Conference, (ICMC'94)*, 131-137.

Davis, L. 1991. *Hanbook of Genetic Algorithms*. New York:Van Nostraud Reinhold..

Dawkins, R. 1989. *The Evolution of Evolvability*. Reading, Mass.: Addison-Wesley, 201-220.

Dawkins, R. 1986 *The Blind Watchmaker,* Penguin Books.

Fischer, R.A. 1930 *The Genetic Theory of Natural Selection*. Oxford: Clarendon Press.

Fornari, J.E. 2001. Síntese Evolutiva de Segmentos Sonoros. Ph.D. diss, Dept. of Semiconductors, Instruments and Photonic, University of Campinas.

Gabrielsson, A. 1981. Music Psychology – A Survey of Problems and Current Research Activities. In *Basic Musical Functions and Musical Ability.* Papers given at a seminar arranged by the Royal Swedish Academy of Music, Stockholm, February 1981, 7–80. Publications issued by the Royal Swedish Academy of Music No. 32. Stockholm: Kungliga Musikhögskolan.

Goldberg, D. E. 1989. *Genetic Algorithms in Search, Optimization and machine Learning*. Addison Wesley.

Holland, J H. 1975. *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.

Horowits, D. 1994. Generating Rhythms with Genetic Algorithms. In Proc. of the 1994 International Computer Music Conference (ICMC'94), 142-143.

Latham, W. and Todd, S. 1992 *Evolutionary Art and Computers*. Academic Press.

Manzolli, J, A. Maia Jr., Fornari J.E. and Damiani,F. 2001a. The Evolutionary Sound Synthesis Method. ACM Multimedia, Ottawa, Ontario, Canada.

Manzolli, J, A. Maia Jr., Fornari J.E. and Damiani,F. 2001b. Waveform Synthesis using Evolutionary Computation. Proc. of the VII Brazilian Symposium on Computer Music.

Moroni, A., Manzolli, J., Von Zuben, F. and Gudwin, R. 2000. Vox Populi: An Interactive Evolutionary System for Algorithmic Music Composition. *Leonardo Music Journal* 10: 49-54.

Moroni, A. 2002. ArTEbitrariedade: Uma reflexão sobre a Natureza da Criatividade e sua Possível Realização em Ambientes Computacionais. Ph.D. diss., Dept. of Computer Engineering and Industrial Automation, State University of Campinas.

Plomp, R. 1970. Timbre as a multidimensional attribute of complex tones. In Frequency Analysis and Periodicity Detection in Hearing. Eds R. Plomp and G.F. Smoorenburg, Sijthoff, Leiden.

Risset, J. C. 1966. *Computer Study of Trumpet Tones*. Murray Hill, N.J.: Bell Telephone Laboratories.

Risset, J. C. 1991. Timbre Analysis by Synthesis: Representations, Imitations and Variants for Musical Composition. In Representations of Musical Signals, 7-43. ed. De Poli, Piccialli & Road, Cambrige, Massachussets: The MIT Press.

Schaeffer, P. 1966. *Traité des objects musicaux*. Paris: Editions du Seuil.

Smalley, D. 1990. Spectro-morphology and Structuring Processes. In *The Language of Electroacoustic Music*, 61-93.London: Macmillan.

Thywissen, K. 1993 GeNotator: An Environment for Investigating the Application of Genetic Algorithms in Computer Assisted Composition. M. Sc. Thesis, Univ. of York.

Venturini G., Slimane M., Morin F. and Asselin de Beauville J.-P. 1997. On using Interactive Genetic Algorithms for Knowledge Discovery in Databases. In Proceedings of the Seventh International Conference on Genetic Algorithms, 696-703. San Francisco, CA: Morgan Kaufmann. T. Bäck, ed.