Vox Populi: An Interactive Evolutionary System for Algorithmic Music Composition

Artemis Moroni, Jônatas Manzolli, Fernando Von Zuben and Ricardo Gudwin

In Darwin's time, most geologists subscribed to "catastrophe theory": that the Earth would be punished many times over by floods, earthquakes and other catastrophes, able to destroy all forms of life. On his voyage on board the Beagle, Darwin verified that the diverse animal species of a region differed from each other in minimal details, but he did not understand how this could result from a "natural" selection. In October 1838, he learned from a small book, Essay on Population Origin by Thomas Malthus, about the factors influencing evolution. Malthus, in turn, was inspired by Benjamin Franklin (the same person who had invented the lightning rod). Franklin had noted the fact that in nature there must be locally limiting factors, or a unique plant or animal would spread all over the Earth; it was only the existence of different kinds of animals that maintained them in equilibrium. This was the universal mechanism that Darwin was looking for. The factor responsible for the way evolution happens is natural selection in the fight for life, i.e. those who are better adapted to the environment survive and assure species continuity. Furthermore, the fight for survival among members of a species is more obstinate, since they must fight over shared resources; small differences, or positive deviations from the typical, are most valuable. The more obstinate the fight is, the faster the evolution; in this context only those better adapted themselves survive. However, characteristics that are positive in a specific environment may have no value in another.

D. Hofstadter, in Metamagical Themas [1], discusses the arbitrariness of the genetic code. According to him, the first moral of this development is: Efficiency matters. A second moral, more implicit, is: Having variants matters. The ratchet of evolution will advance toward ever more efficient variants. If, however, there is no mechanism for producing variants, then the individual will live or die simply on the basis of its own qualities vis-à-vis the rest of the world.

ALGORITHMIC COMPOSITION AND EVOLUTION

R. Dawkins demonstrated the power of Darwinism in The Blind Watchmaker, using a simulated evolution of two-dimensional (2D) branching structures made from sets of genetic parameters. The user selects the "biomorphs" that survive and reproduce to create a new generation [2]. S. Todd and W. Latham applied these concepts to help generate computer sculptures using constructive solid geometry techniques [3,4]. K. Sims used evolutionary mechanisms of creating variations and making selections to "evolve" complex equations to be used in procedural models for computer graphics and animation [5].

A new generation of algorithmic composition researchers has discovered that it is easy to obtain new musical material by using simulated-evolution techniques to create new approaches for composition. These techniques have been useful for searching large spaces using simulated systems of variation and selection. J.A. Biles has described an application of genetic algorithms to generate jazz solos [6] that has also been studied by D. Horovitz as a way of controlling rhythmic structures [7]. On the other hand, it is difficult to drive the results in a desired direction. The challenge faced by the designers of evolutionary composition systems is how to bring more structures and knowledge into the compositional loop. This loop, in an evolutionary system, is a rather simple one; it generates, tests and repeats. Such systems maintain a population of potential solutions; they have a selection process and some "genetic operators," typically mathematical functions that simulate crossover and mutation. Basically, a population is generated; the individuals of the population are tested according to certain criteria, and the best are kept. The process is...
repeated by generating a new population of individuals—or things or solutions—based on the old ones [8]. This loop continues until the results are satisfactory according to the criteria being used. The effective challenge is to specify what “to generate” and “to test” mean.

All evolutionary approaches do, however, share many features. They are all based, like the diagram in Fig. 1, on the general framework provided by J.H. Holland’s original genetic algorithm (GA) [9] or, indirectly, by the genetic programming paradigm of J.R. Koza, who proposed a system based on evolution to search for the computer program most fit for solving a particular problem [10]. In nearly every case, new populations of potential solutions to problems (here, the problem of music composition) are created, generation after generation, through three main processes:

1. By making sure that better solutions to the problem will prevail over time, more copies of currently better solutions are put into the next generation.
2. By introducing new solutions into the population; that is, a low level of mutation operates on all acts of reproduction, so that some offspring will have randomly changed characteristics.
3. By employing sexual crossover to combine good components between solutions; that is, the “genes” of the parents are mixed to form offspring with aspects of both.

With these three processes taking place, the evolutionary loop can efficiently explore many points of the solution space in parallel, and good solutions can often be found quite quickly. In creative processes such as music composition, however, the goal is rarely to find a single good solution and then stop; an ongoing process of innovation and refinement is usually more appropriate.

FIGURE 1. Vox Populi Reproduction and MIDI Cycles: The Reproduction Cycle is an evolving process that generates chords by using genetic operators and selecting individuals and is based on the general framework provided by J.H. Holland’s original genetic algorithm. The MIDI Cycle refers to the interface’s search for notes to be played by the computer. When selected, a chord is put in a critical area that is continually verified by the interface. These notes are played until the next group is selected. (© Artemis Moroni)

INFORMATION SEEN AS GENOTYPES AND PHENOTYPES

Both biological and simulated evolution involve the basic concepts of genotype and phenotype, and the processes of selection and reproduction with variation. The genotype is the genetic code for the creation of an individual. In biological systems, genotypes are normally composed of DNA. In simulated evolutions there are many possible representations of genotypes, such as strings of binary digits, sets of procedural parameters or symbolic expressions. The phenotype is the individual itself or the form that results from the developing rules and genotypes. Selection depends on the process by which the fitness of phenotypes is determined. The likelihood of survival and the number of new offspring that an individual generates are proportional to its fitness measure. Fitness simply expresses the ability of an organism to survive and reproduce. In simulation, it can be evaluated by an explicitly defined mathematical function or it can be provided by a human observer. Reproduction is the process by which new genotypes are generated from an existing genotype. For 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.

Note that selection is, in general, nonrandom and operates on phenotypes, while variation is usually random and operates on the corresponding genotypes. The repeated cycle of reproduction with variations and selections of the fittest individuals drives the evolution of a population toward a 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 forces to evolve independently and later to combine into an individual genotype. Although it is not necessary for evolution to occur, it is a valuable achievement that may enhance progress in both biological and simulated evolutions.

If the mechanics of an evolutionary system are well understood and the chain of causation is properly represented, the process of evolution can be stated in rather simple terms and can be simulated for engineering and art purposes. Given the complexity of evolved structures, it may be somewhat surprising that evolution here appears reduced to rather few rules [11]. In our approach, the population is made up of four note groups, or chords, as potential survivors of a selection process. Melodic, harmonic and vocal-range fitnesses are 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 center, is used. The selected notes are sent to the MIDI port and can be heard as sound events in real time. This sequence produces a sound resembling a chord cadence or fast counterpoint of note blocks.

Individuals of the population are defined as groups of four voices, or notes. (Henceforth, voices and notes will be used interchangeably.) These voices are randomly generated in the interval 0–127, with each value representing a MIDI event, described by a string of 7 bits. In each iteration, 30 groups are generated. Figure 2 shows an example of a group—the genotype—internally represented as a chromosome of 28 bits, or 4 words of 7 bits, one word for each voice. The phenotype is the corresponding chord.

Two processes are integrated: (1) Reproduction Cycle, an evolving process that generates chords using genetic operators and selecting individuals; (2) MIDI Cycle: the interface looking for notes to be played by the computer. When a chord is selected, the program puts it in a critical area that is continually verified by the interface. These notes are played until the next group is selected.

The timing of these two processes determines the rhythm of the music being heard. In any case, a graphic interface allows the user to interfere with the rhythm by modifying the cycles. Figure 1 depicts the Reproduction Cycle and the MIDI Cycle.

FITNESS EVALUATION

Traditionally, Western music is based on harmony; hence, a general theory of music has to engage deeply with formal theories on this matter. The term “harmony” is inherently ambiguous, since it refers to a lower level where smoothness and roughness are evaluated and, at the same time, to a higher aesthetic level where harmony is functional to a given style. However, harmony is a very subjective concept; the perception of harmony does not seem to have a natural basis, but appears to be a common response acquired by people in
specific cultural settings. Nevertheless, while there is a difference of opinion on what constitutes harmony, there is a general agreement on the relative order of music interval consonance. Numerical theories of consonance have tried to capture this aspect, but here again, a lot is left to the imagination, as theory does not clearly define what constitutes the order of simplicity of musical intervals.

In our case, we have applied, as a fitness function, a numerical theory of consonance from a physical point of view. Based on a relative ordering of consonance of musical intervals, a sequence of notes is approximated to its most harmonically compatible note or tonal center. Tonal centers can be thought of as an approximation of the melody, describing its flow. This method uses fuzzy formalism, or fuzzy sets, which are classes of objects with a continuum of membership grades. Such a set is characterized by a function that assigns to each object a grade of membership ranging between 0 and 1 [12]. In Vox Populi, harmony is treated as a function of the commonality, or overlap, between the harmonic series of notes. The overlap measurement is then scaled to be a value between 0 and 1, with 1 denoting complete overlap (i.e., the two notes are the same) and 0 denoting no overlap at all [13].

The harmonic series of notes 60 and 64 (do and mi, in the center of the piano, according to the MIDI protocol) are depicted in Fig. 3, while Fig. 4 depicts their overlap, or consonance measure. According to our approach, approximation to the tonal center is posed as an optimization problem based on physical factors relevant to hearing music. This approach is technically detailed in Moroni et al. [14]. In the selection process, the group of voices with the highest musical fitness is selected and played. The musical fitness for each chord is a conjunction of three partial fitness functions:

\[
\text{Musical Fitness} = \text{Melodic Fitness} + \text{Harmonic Fitness} + \text{Vocal Range Fitness}
\]

Melodic fitness is evaluated by comparing the notes that compose a chord to a value \( I_d \) (identity), which can be modified by the composer in real time using the melodic control of the interface. This control "forces" the notes of the selected chord to be close to (or distant from) the \( I_d \) value, which acts as a tonal center and is treated as an attractor. Harmonic fitness is a function of the consonance among the components of the chords. Vocal range fitness verifies which notes of the chord are in the range desired by the composer, who may modify it through the octave control.

The melodic control and the octave control allow the composer to conduct the music that is being created, interfering directly in the musical fitness, while other controls simply modify attributes of the chord that has been selected. Also, the biological and rhythmic controls allow the user to modify the duration of the genetic cycle by modifying the duration of the evolution eras. Eras can be thought as the number of iterations necessary to generate a new population. The combined use of the controls gives birth to sound orbits, which can be perceived through intermittent cycles.

**FITNESS TUNING**

Part of the reason why evolution in nature is very slow is that the forces of selection can be imperfect and at times ineffectual. Non-privileged individual organisms may still succeed in finding mates, having offspring and passing on...
The sum of heights of the components of

the chord and note 64 as the tonal center.

Note 60 can be thought of as one of the notes of

harmonic series of notes 60 and 64. Note

Fig. 4. Vox Populi: Overlap between the

harmonic series of notes 60 and 64. Note 60 can be thought of as one of the notes of

their genes, while organisms with a new

advantageous trait may not manage to

live long enough to find a mate and in-

fluence the next generation. Todd and

Werner have made a charming compari-

son with the Frankenstein tale; Franken-

stein hoped for much more than the cre-

ation of a single superior living

being—he intended his creature to be-

get a whole new race that would grow in

number and goodness, generation after

generation. Later he worried that this

process might not go exactly as he

planned, with the children becoming

more monstrous than their parents, a re-

alization that led him to abandon his ef-

forts to create a female progenitor. But,

suppose, like Frankenstein, one wants to

enter the “workshop of filthy creation”

[15] and replace the human composer

with an artificial composition system—

due to a wish to ease a composer’s

workload, an intellectual interest in un-

derstanding the composition process,

the desire to explore unknown musical

styles or mere curiosity about the possi-

bilities. Maybe Vox Populi could have

been initially included only in the last

group as inspired by a “mere curiosity

about the possibilities” but given Vox

Populi’s surprising results, it can now be

included in the first two.

Two main approaches have been tried to

express the fitness evaluation, both

presenting interesting effects. The first

one, derived from a composer’s musical

experience, provided a faster fitness

evaluation. This method allows the use

of a large population, 100–200 chords,

producing greater diversification and

resulting in a slower convergence to the

best chord sequence. In the second ap-

proach, the consonance criterion is

used, and a longer calculation is needed

to evaluate musical fitness. In order to

assure quick enough real-time perfor-

mance by the system, the population was

limited to 30 chords. The advantage of

this approach is that it formalizes math-

eematically the concept of consonance. It

can be easily described and flexibly pro-

grammed and modified. Since the musi-

cal fitness criterion used was stricter in

the second example (using 30 chords

instead of 100–200), the resulting sound

output was less diversified; it was pos-

sible to hear the musical sequence con-

verging to unison. This fact highlighted

the notion that, in musical composition,

not only consonance but also disso-

nance is desirable. Figure 5 depicts a

Vox Populi musical output.

Vox Populi differs from other systems

found in genetic algorithms or evolu-

tionary computation in which people

have to listen to and judge musical

items; instead, Vox Populi uses the key-

board and mouse as real-time music

controllers, acting as an interactive com-

puter-based musical instrument. It ex-

plores evolutionary computation in the

context of algorithmic composition and

provides a graphical interface that al-

lows the composer to change the evolu-

tion of the music by using the mouse.

These results reflect current concerns at

the forefront of interactive composition

computer music and in the develop-

ment of new control interfaces.

Interface controls use nonlinear itera-

tive mappings. They can give rise to

attractors, defined as geometric figures

that represent the set of stationary states

during the composition process, of a
dynamic system or simply trajec-

tories to which the system is attracted.

A piece of music consists of several sets of

typically material manipulated and

exposed to the listener, such as pitches,

harmonies, rhythms, timbres, etc. These

sets are composed of a finite number of

elements, and the basic aim of a com-

poser is to organize them in an aesthetic

way. Modeling a piece as a dynamic sys-

tem implies a view in which the com-

poser draws trajectories or orbits using

the elements of each set [16].

The interactive pad control supplies a

graphical area in which 2D curves can

be drawn. These curves, a blue one and

a red one, are linked to the controls of the

interface. The red curve links to the

melodic and octave range controls; and

the blue curve links to the biological

and rhythmic controls. When the inter-

active pad is active, the four other linked

controls are disabled. Each curve de-

scribes a relation between the linked

variables. They are traversed in the or-

der in which they were created; their

horizontal and vertical components are

used for fitness evaluation and to modify

the duration of the genetic cycles, inter-

fering directly in the rhythm of the com-

position. The pad control allows the

composer to conduct the music through

drawings, suggesting metaphorical “con-

ductor gestures” used when conducting

an orchestra. Using different drawings,

the composer can experience the gener-

ated music and conduct it, trying differ-

ent trajectories or sound orbits. The tra-

jectories then affect the reproduction

cycle and musical fitness evaluation.

INTERFACE AND PARAMETER

CONTROL

The resulting music moves from very

pointillistic sounds to sustained chords,

Fig. 5. Score of MIDI raw material produced by Vox Populi. This material was produced by Vox Populi in an interactive session by Jónatas Manzolli, composer. In the latest Vox Populi version, the user is able to record a piece that is composed during performance.
depending upon the duration of the genetic cycle and the number of individuals of the original population. The interface is designed to be flexible enough for the user to modify the music being generated. Below is a short description of the controls available to the user interacting with Vox Populi. The melodic, biological, rhythmic and octave controls allow the composer to modify the fitness function in real time and are associated with attractors. Vox Populi’s interface is depicted in Fig. 6 and in Color Plate A No. 2.

**Melodic Control**

The mel scroll bar allows one to modify the value \( Id \), which is the tonal center in the evaluation of melodic fitness. Given an ordered sequence of notes, it seems intuitively appealing to call the note that is most consonant with all the other notes the coloring, or tonal, center. Hence, the extraction of the tonal center of a sequence of notes would involve finding an optimally harmonically compatible note. As mentioned before, in Vox Populi, the consonance is measured according to the \( Id \) value. This value is obtained from the interface control and can be changed by the user.

**Biological Control**

The bio scroll bar allows interference in the duration of the genetic cycle, modifying the time between genetic iterations. Since the music is being generated in real time, this artifice is necessary for the timing of the different processes that are running. This value determines the slice of time necessary to apply the genetic operators, such as crossover and mutation, and may also be interpreted as the reproduction time for each generation.

**Rhythmic Control**

The rhy scroll bar changes the time between evaluations of musical fitness. It determines the “time to produce a new generation” or the slice of time necessary to evaluate the musical fitness of the population. It interferes directly in the rhythm of the music; any change makes the rhythm faster or slower.

**Octave Control**

The oct scroll bar allows enlarging or diminishing the interval of voices considered in the vocal range criterion. The octave fitness forces the notes to be in range \( H \), assuming that \( H \) is the range of the human voice and associated with the central keys on the piano; but since several orchestras of instruments are used, this range is too limited for some instruments. We originally intended to restrict the generated voices to specific ranges in order to make those voices resemble the human voice. Nevertheless, a user can now enlarge these ranges by using the octave control.

**Orchestra Control**

Six MIDI orchestras are used to play the sounds: (1) keyboards; (2) strings and brasses; (5) keyboards, strings and percussion; (4) percussion; (5) sound effects and (6) random orchestral parts, by taking an instrument from the general MIDI list. Using the order above, these orchestras are sequentially changed into time segments controlled by the seg scroll bar.

**Interactive Pad Control**

The “Pad On” button enables and disables the pad change on the controls defined above. They may be grouped into two pairs, which may be interpreted as variables of a 2D phase space. This allows a user to draw and orient the curve to determine the evolution of the music.

**Fitness Displays**

Three other displays allow the user to follow the evolution of fitness. The upper display, at the right side of Fig. 6, shows the notes and the fitness of the chord that is being played. In the middle display, a bar graph shows the four voices (bass, tenor, contralto, soprano) and their values. It is equivalent to the membership function values related to the range of the voices. The bottom display shows the melodic, harmonic and octave fitness bars.

**CONCLUSION**

Despite the fact that Vox Populi works at the level of sound events controlled by MIDI protocols, or notes, in a macrostructural context, we learned two lessons. First, an evolutionary computational approach was successfully applied to generate complex sound structures with a perceptual and efficient control in real time. Second, applications of evolutionary computation may be foreseen to prospect sound synthesis. Complex behavior systems have been used for sound synthesis, like Chaosynth, which uses cellular automata to control structures [17]. In Chaosynth, the generation occurs via granular synthesis. In another approach, Fracwave [18] uses the dynamics generated by complex systems to synthesize sounds using complex dynamics.

We may say that varying the fitness controls in Vox Populi promotes a “sound catastrophe,” in which the previous winner may no longer be the best. Conditions for survival have changed, as they do in nature.

The question we pose is how does an idea, or concept, survive? Vox Populi is simple, efficient and has been used in different ways, which may be considered variants: as an autonomous or demonstrative system generating music; as a sound laboratory, where people can try and experience the sound produced; as a studio, manipulating and generating samples that have been used in compositions and...
in sound landscapes. Another use currently being considered is to couple the system with sensors, allowing the user to describe orbits in space that would be treated like the 2D curves supplied by the interactive pad. Will Vox Populi survive?

Vox Populi means “voice of the people.” Since the individuals in the population are defined as groups of four voices, we can think of them as “choirs,” fighting to survive and to be present in the next generation, while the environment and survival conditions are changing dynamically.

One of the first noted proposals to formalize composition was made by the Italian monk Guido d’Arezzo in 1026, who resorted to using a number of simple rules to map liturgical texts in Gregorian chants [19] due to the overwhelming number of orders he received for his compositions. The text below is attributed to d’Arezzo. His compositional approach has survived for several centuries, and even today, we still seek strategies for constructing the unknown melody.

As I cannot come to you at present, I am in the meantime addressing you using a most excellent method of finding an unknown melody, recently given to us by God and I found it most useful in practice. . . .

To find an unknown melody, most blessed brother, the first and common procedure is this. You sound on the monochord the letters belonging to each neume, and by listening you will be able to learn the melody as if you were hearing it sung by a teacher. But this procedure is childish, good indeed for beginners, but very bad for pupils who have made some progress. For I have seen many keen witted philosophers who had sought out not merely Italian, but French, German, and even Greek teachers for the study of this art, but who, because they relied on this procedure alone, could never become, I should not say, skilled musicians, but even choristers, nor could they duplicate the performance of our choir boys [20].

References
20. Strunk [19].

Manuscript received 18 January 1999.

Artemis Moroni is a technologist at the Automation Institute of the Technological Center for Informatics in Campinas, São Paulo, Brazil. The main topics of her research are multimedia devices applied to automation environments, evolutionary computation and technology applied to art and music.

Jônatas Manzolli is composer and head of the Interdisciplinary Nucleus of Sound Communication at the State University of Campinas, São Paulo, Brazil. He teaches in the department of music, and the main topics of his research are algorithmic composition, gesture interfaces and multimedia devices for sound environments.

F.J. Von Zuben is a member of the department of computer engineering and industrial automation at the State University of Campinas, São Paulo, Brazil. The main topics of his research are artificial neural networks, evolutionary computation, nonlinear control systems, nonlinear optimization and multivariable data analysis.

Ricardo Gadewin is a faculty member of the electrical and computer engineering department at the State University of Campinas, São Paulo, Brazil, where he develops research into intelligence and intelligent systems, intelligent agents, semiotics and computational semiotics. His topics of interest also include fuzzy systems, neural networks, evolving systems and artificial life.