Vox Populi: Evolutionary Computation for Music Evolution

Artemis Moroni; Jônatas Manzolli; Fernando Von Zuben; Ricardo Gudwin
CTI - Technological Centre for Informatics;
NICS/UNICAMP – Interdisciplinary Nucleus for Sound Studies;
FEEC/UNICAMP - Electrical and Computer Engineering School
artemis@ia.cti.br; jonatas@nics.unicamp.br; vonzuben@dca.fee.unicamp.br;
gudwin@dca.fee.unicamp.br

Abstract

In this article we describe an application of Evolutionary Computation to Algorithmic Composition. The individuals of the population were defined as groups of four voices: *soprano, contralto, tenor and bass*, or "chords"; and they are potential solutions for a selection process. Each chord was evaluated under three criteria: melodic, harmonic and voice range. Based on the ordering of consonance of musical intervals we use the notion of approximating a sequence of notes to its harmonically compatible note or *tonal centre*. Tonal centres can be thought of as an approximation of the melody describing its flow. This method uses fuzzy formalism and is posed as an optimisation problem based on the physiological factors relevant to hearing music.

1 Introduction

Western music is based on harmony; hence a general theory of music has to cope deeply with formal theories on this matter. However, this is a very subjective issue; the judgement of harmony does not seem to have a natural basis, but appears to be a common response acquired by people in a certain cultural area. Therefore opinions on the subject may vary widely depending on social and cultural backgrounds, and many attempts to formalise the concept have been inadequate.

Starting from this point, we developed another approach for algorithmic composition based on three musical aspects: melody, harmony and voice range. The specification of the melodic, harmonic and voice range criteria define the *fitness* of a group to the selection function applied. This function returns the "*best individual*", or "best chord", according to the aspects measured. The selected group is treated as a set of MIDI notes and played. The resulting system, Vox Populi, allows the user to modify the fitness function through four controls, the first for the melodic criterion; the second for the duration of the genetic cycle and music rhythm; the third for the set of octave range to be considered, and the fourth for the time segment for each selected orchestra. All these controls are available for real time performance allowing the user to play and interact with Vox Populi's music evolution.

2 Definitions

Some concepts are fundamental do the understanding of the article.

2.1 Population as MIDI data

An auditory event can be characterised for our purposes by four parameters: pitch, timbre, loudness and duration.

a) Pitch can be defined as the auditory property of a note that is conditioned by its frequency relative to the other notes. The range of musical pitch has been defined as the range within which the interval of an octave can be perceived. This has been found to correspond roughly to the range of the piano. From this continuum of frequencies, a set of discrete frequencies is selected in such a way that the frequencies bear a definite interval relationship to another. So, pitch in the musical sense corresponds to a frequency that is selected from a predefined repertoire. In this scheme, two discrete frequencies are chosen in the interval of an octave such that the ratio between any two adjacent frequencies is $2^{1/12}$. This interval ratio in music terminology is termed a semitone.

b) Timbre is the individuality of sound acquired by the addition of harmonics to the fundamental pitch. This is characteristic of a given musical instrument and the mode of playing it.

c) Loudness is that aspect of an auditory event related to its intensity.

d) Duration is characterised by the period of time for which the event is perceivable.

Using the notions above, a melody is defined as a fixed temporal ordering of auditory events. So, a melody in conventional occidental notation resembles a system of

Cartesian co-ordinates. The pitch and duration are carefully marked; timbre is decided by the instrument for which it is written, and loudness is marked more crudely (Vidyamurthy, 1992).

2.2 Rhythmic Genetic Cycle

The general architecture of the rhythmic genetic cycle is shown in Fig. 1. The individuals of the population are defined as groups of four voices. These voices are randomly generated in the interval [0..127], which correspond to 7 bits values for a MIDI event. In each era, 30 chords are generated. The chord is internally represented as a chromosome of 28 bits, or 4 words of 7 bits, one for each voice.
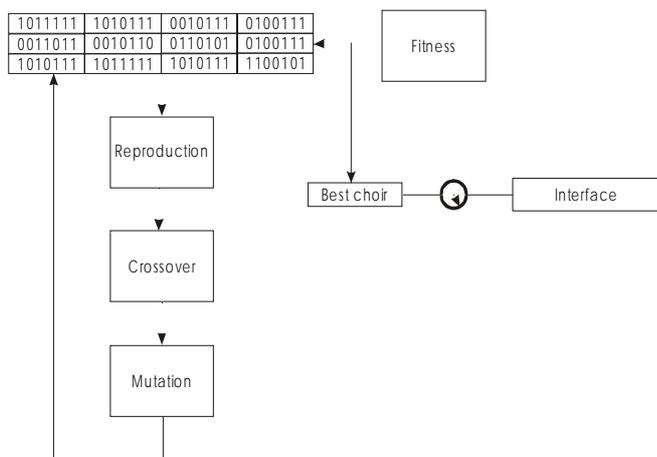

Fig. 1: The rhythmic genetic cycle

We can see from the diagram that there are two co-operative processes in the genetic cycle, one producing notes and the other - the interface - using notes. Once the initial population of individuals has been created, the fitness of each chord is evaluated. The fitness function is defined as a composition of three sub-functions: the harmonic fitness, the melodic fitness and the voice range fitness. For the evaluation of the harmonic fitness and the melodic fitness the consonance criterion is considered. For the voice range evaluation, an exclusion criterion is used. After the fitness evaluation, typical operations of genetic programming like crossover and mutation are applied to the individuals, according to probabilistic rates. Once the best chord is selected, it is put available to be played. The second process, which is looking for new available notes, plays the notes. The following steps are realised in the genetic cycle:

STEP 1: Create an initial population randomly;

STEP 2: Until the termination criterion has been satisfied, perform the following:
- Evaluate the fitness of each individual in the population;
- Apply the genetic operators to individual chromosomes, or groups of voices, chosen with a probability based on fitness, to create a new population. That is:
- **Reproduction:** Copy existing individual strings to the new population;
- **Crossover:** Create two new chromosomes by crossing over randomly chosen sub-lists (sub-strings) from two existing chromosomes;
- **Mutation:** Create new chromosomes from an existing one by randomly mutating the character in the list.

STEP 3: Designate the best individual that appeared in any generation as the result (Pedrycz & Gomide, 1998).

The steps above were detailed in order to make visible the many operations realised in each cycle. Despite the fact that there is a medium time cycle to designate the best individual in each generation, these small variations in each time cycle determine the *genetic rhythm.* The time interval between the selection of the best chords in two successive cycles is different. In the other side, the interface is regularly "asking for new notes". When the best chord is selected, and consequently is available, the notes that constitute it are played until the next chord is selected. Once the next chord is selected the new notes are played. The different times of the notes being played define the rhythm and the melody of the genetic cycle.

3   Fitness Evaluation

This section presents the criteria used to develop a set of function to evaluate the system musical fitness.

3.1 Consonance Criterion

The consonance among the four voices is evaluated as a function of the voices attributes. Using this function, the tonal centre cognition can be formulated as an optimisation problem. Consonance is a combination of two simultaneous notes judged pleasing to listen and is defined as a function of the commonality or overlap between the spectral components of notes (Vidyamurthy, 1992). This overlap measurement is then scaled to a value between 0 and 1 with 1 denoting complete overlap (i. e., the two notes being the same) and 0 denoting no overlap at all. This notion of overlap can be succinctly captured in the fuzzy set formalism (Pedrycz & Gomide, 1998). Fig. 2 presents the weighting of the partials of a note versus the relative pitch.

A musical note is a compound tone consisting of its primary tone and upper harmonic partials. Represented graphically as a spectrum, a musical note is a plot of frequency against amplitude. The figure above repre-

sents the weighting of the partials of a note versus the relative pitch. Note that n denotes the nth key on the piano, and that (n + k) denotes the key k semitones above key n. In Fig. 3 the upper partials for notes 60 are evaluated.
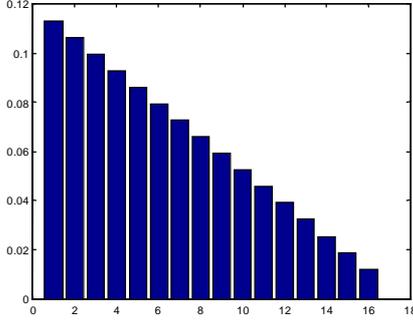

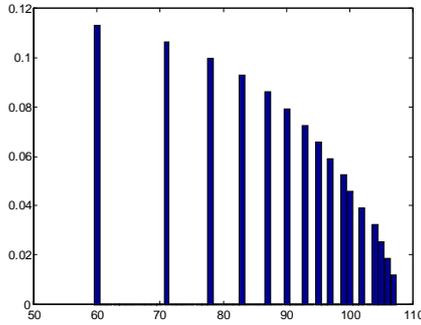Fig. 2: The weighting of the partials of a note versus the relative pitch

Fig. 3: The upper partials of notes 60

Formally defining, each note is a fuzzy set on a countable universe of discourse U; that is,

$$S_N / f_N(x) = \{(x, y) / x \in U, y \in [0, 1]\}$$

such that S_N is normalised:

$$f_N : U \rightarrow [0, 1] \text{ such that } \Sigma f_N(x) = 1$$

and

$$S^i_N(x) = \{(x, y^*) / y^* = inf(y^j), (x, y^j) \in f_N, i = 1, 2\}$$

Consequently, the **consonance** or overlap between the notes is defined as

$$Co(x) = C(S^i_N(x), i = 1, 2) = \Sigma y^*.$$

3.1.1 Melodic Fitness

We recall our description of a melody as an ordered sequence of notes with their corresponding start times. In formal terms, a melody can be defined as a string in which each character is an ordered pair (f_N, t), f_N being the note and t the time duration for playing it. Given an ordered sequence of notes it seems intuitively appealing to call the note that is most consonant with all other

notes the colouring or tonal centre. Hence the extraction of the tonal centre of a sequence of notes would involve finding a single harmonically compatible note such that the time weighted dissonance between the note and each note in the sequence is minimised. In the Vox Populi system, the consonance is measured according to the value Id. This value is obtained from an interface control and can be changed by the user. The melodic fitness is evaluated as:

$$M(x_1, x_2, x_3, x_4) = Max[C_{Id}(x_j)], j = 1, 4;$$

where

$$C_{Id}(x_j) = C(S_N(x_j), S_N(Id))$$

3.1.2 Harmonic Fitness

Consonance is a combination of two simultaneous notes judged pleasing to listen. Harmony is defined here as a function of the commonality or overlap between the spectral components of the notes, and the sum of the consonance of the notes of the chord. Therefore, the harmonic fitness is defined as:

$$H(x_1, x_2, x_3, x_4) = Max(Co(S^i_N \, i = 1, 2)$$

$$Co(S^i_N \, i = 2, 3), Co(S^i_N \, i = 3, 4), Co(S^i_N \, i = 1, 4))$$

3.1.3 Voice Range Fitness

Given a voice note x_i, we define its range as an integer interval $R_i = [m_i...M_i]$ where m is the minimal and M is the maximal value for each voice. The function $f(x_i)$ called Voice Range Fitness is evaluated as follows:

$$\text{if } x_i \in R_i \text{ then } f(x_i) = 1 \text{ otherwise } f(x_i) = 0$$

The overall fitness is written as

$$O(x_1, x_2, x_3, x_4) = \Sigma f(x_i), i = 1..4$$

3.2 Musical Fitness

The resulting musical fitness is a conjunction of the previous functions and is defined as:

$$F(O, M, H) = O(x_1, x_2, x_3, x_4) + M(x_1, x_2, x_3, x_4) + H(x_1, x_2, x_3, x_4)$$

In the selection process, the chord with biggest fitness is chosen and played.

4 Vox Populi System

We implemented a real time system to perform a series of sound experiment and eventually to develop a tool for Algorithmic Composition.

4.1 Interface and Parametric Control

The user can interfere in the fitness function through four controls. The first control is associated to the Id value, used to evaluate the melodic fitness. Moving this control the user changes the tonal centre of the music. The second control interferes in the duration of the genetic cycle. Changing this control, the user modifies the rhythmic, making it faster or slower. The third control is associated to the selected octaves. The range of valid voices can be enlarged or diminished. The fourth control sets the time segment for each orchestra. When the time segment is reached the orchestra changes.

Fig. 4: Vox Populi interface. The controls used by the user to change music evolution are shown on the left

4.2 Results

The music results move from very pontilistic sounds to sustained chords. This depends upon the duration of the genetic cycles and the number of individuals on the original population. Variations in the fitness function were tried, allowing for example the whole MIDI interval [0..127] for the voices generation, resulting in more diversified chords. The original decision to restrict the generated voices to specific ranges was just to resemble human's voices; nevertheless a composer can enlarge that ranges. The real time performance is satisfactory, allowing a sound interaction and possible use of Vox Populi in live music. The program was developed for MS Windows environment and it runs on any IBM PC with most of the commercial soundboards. Pentium machines will work better starting upon 166 MHz.

5 Conclusion

In this article we describe an application of Evolutionary Computation to Algorithmic Composition. The individuals of the population were defined as groups of four voices: *soprano, contralto, tenor and bass*, or *"chord"*; and they are potential solutions for a selection process. Each chord was evaluated under three criteria: melodic, harmonic and voice range. Based on the ordering of consonance of musical intervals we use

the notion of approximating a sequence of notes to its harmonically compatible note or *tonal centre*. This method uses fuzzy formalism and is posed as an optimisation problem based on the physiological factors relevant to hearing music. The Graphic Interface allows the user to change parameters, interfere in the fitness function and, consequently, on the music evolution. Further, we are developing an integration of this system with Gesture Interface such as gloves devices to enhance the real time man/machine interaction, allowing a human gesture real time control. This approach was applied in the previous project ActContAct (Manzolli et all. 1998). In this way, Vox Populi, designed to reflect the natural evolution on the sound domain, will close its own life cycle returning the control to the human.

References

Manzolli, J., Moroni, A. and Matallo, C., "ActContAct: new media performance for video and interactive tap shoes music", 6[Th] ACM International Multimedia Conference, pg. 3, 1998.

Vidyamurthy, G. and Chakrapani, J., "Cognition of Tonal Centres: A Fuzzy Approach", Computer Music Journal, 16:2, 1992.

Pedrycz, W. and Gomide, F., *An Introduction to Fuzzy Sets Analysis and Design.* The MIT Press, Cambridge, Massachusetts, 1998.