A Genetically Generated Drone A/V Composition Using Video Analysis as a 'Disturbance Factor' to the Fitness Function

Tristan McGuire and Giuseppe Torre
Digital Media & Research Centre
Dept. of Computer Science & Information Systems
University of Limerick, Ireland

Abstract

This paper discusses the development of an audio-visual composition based on genetic algorithms strategies. The genetic algorithm's fitness function dynamically adjusts the optimisation targets linked to the mechanisms responsible for the generating of drone soundscapes. The fitness function continuously changes based on the results of an analysis of the visual elements of the artwork thus acting as disturbance factor. In doing so, the audio material never achieves full optimisation and constantly shapes itself. The paper offers both a technical and aesthetic analysis of the development of the composition.

An algorithm is simply a series of instructions that, when followed, achieve a specific result (Cipriani and Giri 2013) (Dodge and Jerse 1997). In this very general sense, all artworks, as well as many other aspect of our lives, could be referred to as 'algorithmic'. In reality, when the topic of algorithmic composition is discussed it is limited to the application of repeating instructions and mathematical processes which create or control a set of sounds in a score (Supper 2001). The tradition of this particular approach has a long history that goes beyond the advent of computers. While algorithmic composition today is directly associated with computer music, composers began using the idea of algorithmic processes in music long before electronic computing was widely available (Supper 2001)(Edwards 2011).

As computers became more widely available and easier to use, the study of algorithmic composition was accelerated. Many processes have been experimented with and applied to the task of composition and more to the many sub-tasks involved in the creation of music. Some of the areas that have been thoroughly, although not exhaustively, explored are the application of stochastic systems to create randomly or semi-randomly controlled processes within a composition, the use of self-similar processes, and chaotic systems. Often these are distributions weighted in various ways and mapped to characteristics of the sounds. Maps of chaotic mathematical systems and l-systems have also been applied to compositional parameters.

In the late 20th century, algorithms that mimic natural selection processes have been developed. These genetic algorithms were initially proposed as a way to find solutions for problems that encompassed very large search spaces. A genetic algorithm tries to create solutions that achieve a target fitness by evolving through a process of crossing parent solutions that display a high fitness level. Exploration into the application of genetic algorithms to musical tasks began in the early to mid nineties. An early example of genetic algorithms to music was Biles' creation GenJam (Biles 1994) which generated jazz solos over an accompanying set of chords. Since then, genetic algorithms have been used to control granular synthesis, generate melodies, control timbral changes, and compose pieces of music to match the style of classical composers (Burton and Vladimorova 1999)(Fels and Manzolli 2002)(Fujinaga and Vantonne 1994)(Moroni et al. 2000).

Most of these efforts at applying genetic algorithms have focused on achieving a specific result. The effort has been made to reach maximal optimization which is measured by achieving a close match to the criteria set forth in the fitness function of the algorithm.

Genetic Algorithms are well suited for finding solutions or sets of solutions in problems for which the potential solutions are in a large search space or where there is not one single optimal solution (Tzimeas and Mangina 2009). The algorithms mimic the process of evolution by testing for fitness in a population, 'breeding' the fittest candidates in a generation's population, mutating a small percentage of the population and creating a new population before repeating the fitness evaluation (Fujinaga and Vantonne 1994).

The fitness function can be described as using a deterministic approach, a formalistic approach, or a user-defined approach (Burton and Vladimorova 1999). The deterministic approach describes a method of measuring fitness based on pattern matching. In this type of system the fitness is determined by how closely the results match a set of samples. In musical examples the genetic algorithms are attempting to match the patterns of samples scores which have been entered as targets (Alfonseca, Cebrian, and Ortega 2006).

The formalistic approach attempts to find solutions that satisfy a set of rules. In this type of algorithm a set of rules like the rules of counterpoint are entered as mathematical functions and the fitness is determined by how closely the results follow these rules. In the user-defined approach, human feedback is used to determine fitness. This could take the form of listeners pressing buttons when they hear a phrase that they have determined to be pleasing.

Most examples of fitness functions are static in the sense that they are trying to achieve a close match to fixed criteria. A more recent development is to design the fitness function with dynamic criteria to create a system that never reaches an ideal state (Fels and Manzolli 2002) (Freitas and Guimaraes 2011).These dynamic fitness functions may also use multiple objectives to return new results based on balancing weighted criteria (Miranda, Kirke, and Zhang 2010).

The approach to the development of the a/v composition described here is based on the use of an ever changing fitness function. The fitness function is constantly being 'disrupted' by data received from a video analysis module. The video along with the resulting composition, comprise a single piece of work to be presented as an installation. The dynamic nature of the fitness function should ensure that the music created is evolving continuously while the video plays and by linking the sound to the video in this manner a meaningful relationship between visible and audible parts of the work. The following sections present the technical and aesthetic considerations of the piece followed by a method for implementing the system and producing the desired result.

Description of the Algorithm

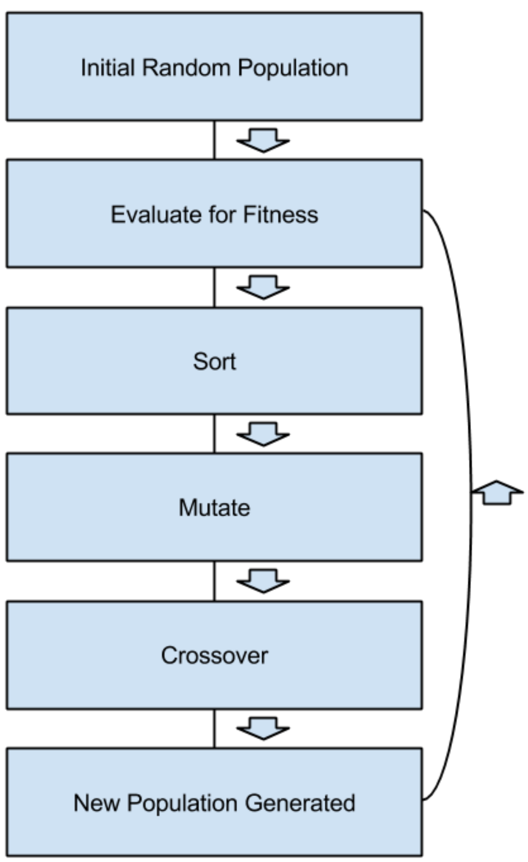

Figure 1: Map of the genetic algorithm.

An initial population is created randomly or semi-randomly. How random the initial population is can be weighted with distributions favouring certain characteristics of the initial genotypes. In the system presented here, the initial population is randomly generated. This initialisation phase is not repeated as the genetic algorithm progresses through its iterations. The population is tested for fitness. The fitness function measures the characteristics of each individual in the population and ranks them according to how closely they match the target parameters.

The fitness function deployed for the composition is formalistic and dynamic. A set of rules is entered which maps the RGB value analysis from the video to sound parameters which have been shown to affect the spectral and emotional content of the generated soundscape. These rules can be varied according to the design of the sound generating module and tailored to the composition that the system is being applied to. On initialisation, the target value for fitness is set to a median value (127), and the initial random population is evaluated in the same process used in subsequent iterations based on this figure.

The next step is the crossover stage. It is a point in the process in which parent individuals with a higher fitness value are selected from the evaluated population, split at a crossover point, and recombined with each other. The result is a new population which has a higher average fitness than the prior generation.

Mutation is a stage, sometimes left out, which randomly changes a small percentage of the new population. This random change helps avoid the problem of a genetic algorithm settling into a less than ideal solution by reaching a local maximum. These local maxima act as 'traps' in genetic algorithms which can cause false termination of the process as the algorithm seems to have achieved the highest possible fitness. In the system presented here, the mutation stage occurs after the evaluation but prior to crossover which can be seen in Figure 1. Placing the mutation stage at this point in the algorithm causes an altered member of the population to crossover regardless of its own fitness, provided the individual selected for mutation would have been selected as a parent initially. The result of placing the mutation at this point is that mutation will have an affect on the population rather than dying out immediately due to the sorting process.

The process then repeats. The new population is evaluated for fitness and ranked accordingly. Crossover between the best candidates is performed, mutation occurs, and the new population is evaluated. This repetition creates a 'generational clock'. Initial versions of the system used a timed loop but this was eliminated later and the clock is determined by the length of time it takes to process each iteration. This process can be terminated when a certain number of iterations have been performed.

However, the parameters that the algorithm described here is attempting to match are shifting according to the results offered by the RGB video analysis. Thus, by constantly 'moving the goalpost', the process can continue indefinitely provided it hasn't been instructed to terminate after a certain time or number of iterations have passed. This termination condition is not necessary to the system as presented, but has been included in order to facilitate presentation of the work within varying time constraints.

System Design

The system is entirely built in Max/MSP and Jitter. It is divided into conceptual modules which work with each other and each perform a specific part of the task. The modules are Sound Synthesis, Video Analysis, and Genetic Algorithm. These modules and their relationships are depicted in Figure 2.

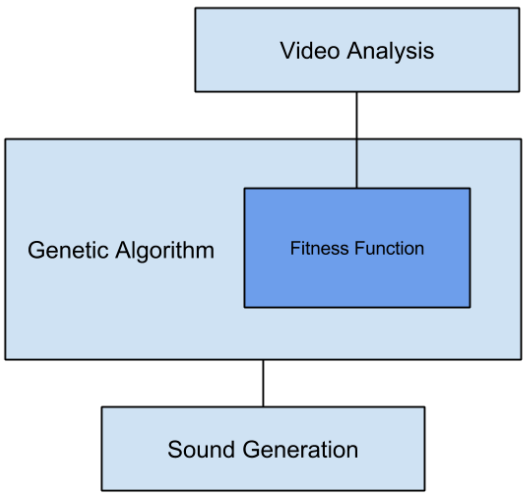

Figure 2: Modules as they are arranged in the system.

Video analysis

The video analysis is conducted in Jitter, the graphical toolbox for Max/MSP. Its duty is to return RGB values from a video created and imported specifically for this composition. RGB values were chosen for their uniform nature, i.e. - each value is a range of 0 - 255, as opposed to HSL for example, which are two values of 0 - 100 percent and one value of 360 degrees. Using uniform values allowed them to reconfigured as a spectrum as depicted in Figure 3. A mean value across the whole frame for each colour was used. The RGB values are mapped to the parameters used in the genetic algorithm to measure fitness. As the video plays, these values change which means that the fitness evaluation can never reach an ideal state. The video analysis has been kept simple in order to retain clarity of purpose. The focus of the work is on the evolution of the sound as the video plays, not on the video analysis itself. The video playback itself will also be kept relatively simple. While the aesthetics of evolution and genetic techniques are being taken into account, the video itself will not evolve. It would, of course, be possible to use the video analysis module to feedback data from the genetic algorithm to create an evolving visual display, however it has been determined that this is beyond the scope of this work. A more sophisticated approach to the development of the video content would distract from the focus of the work, which is the evolution of the sound and composition.

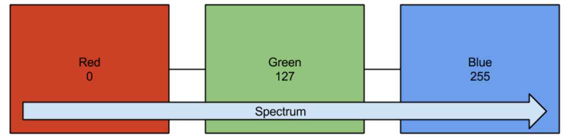

Figure 3: Colour value spectrum.

Audio-visual and genetic algorithm mappings

The first step in the process is to create a 'vocabulary' to translate colour into emotional description and translate that description into auditory parameters. Emotional content is, of course, subjective so this vocabulary is, to some extent, a personal interpretation. This subjectivity should not affect the results however, as the criteria and parameters are set in advance which gives a set of rules against which the results of the process can be measured and quantified.

The emotional descriptions for colour are set as shown in Figure 3. These descriptions are mapped to parameters based on the findings discussed in Balkwill and Thompson's paper on cross cultural emotional cues - anger/tension created through 'rougher' timbres and complexity of texture, sadness through greater range of harmonics along with slower changes to the microtonal qualities of the drone, and joy/happiness through consonance and simplicity (Balkwill and Thompson 1999).

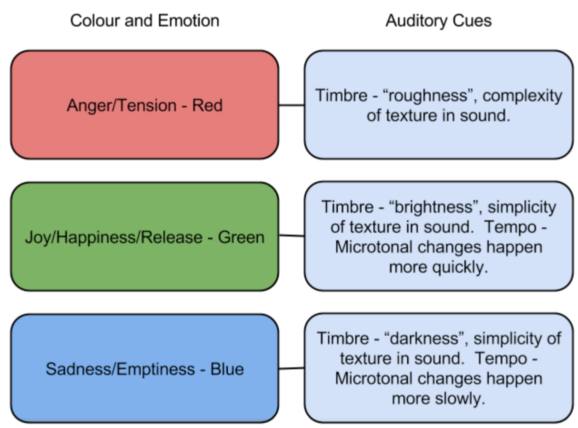

Figure 4: Mapping strategies devised.

These mappings of conceptual cues links the video analysis module to the fitness function of the genetic algorithm. The genetic algorithm will then have its output mapped to the a list auditory parameters in the sound synthesis module. The parameters being controlled in the synthesis module vary the range of partial tones to affect pitch range and timbre of a complex waveform. The population's minimum, maximum, median, and deviation from target values are returned from the genetic algorithm. These values are used to control delay and filtering parameters along with frequency and granulation parameters to affect spatial and timbral cues. Combining these factors will produce a drone which has evolving microtonal qualities mapped to a predetermined set of rules based on the emotional designations given to the RGB values. Short demonstrations of the resulting a/v composition can be viewed at https://vimeo.com/album/2961623.

Aesthetic Considerations

The idea of evolution in art has been explored for a relatively long time. Evolutionary computational techniques have been applied to visual media and auditory media in many ways. As correctly pointed out by McCormack (2014), 'the basis of all generative art resides in its engagement with process'. This, however, does not allow for an exhaustive aesthetic analysis of the artwork per se. Indeed, if the generative process is at the core of the artistic intent, the modalities with which the process is executed and interpreted can probably provide better means for an aesthetic analysis. These two elements are 'intimately intertwined' (ibidem). In light of these considerations, it is useful to justify the choices that led to the implementation of the genetic algorithm and its specific modalities of deployment alongside other artistic practises it touches upon.

A musical analysis of the presented artwork provides a first element of discussion. The concept at the heart of the proposed compositional work explores the timelessness of change just as evolution is infinite and constantly changing. One of the primary features of the resulting acoustic output is an ever changing drone-like soundscape. By using a single sound over an extended period of time, the sound becomes a container for subtle variations, a canvas for small timbral changes to be heard against. The sound becomes a backdrop to its own details. The duration of the sound forces its audience to listen differently, to notice the smaller variations that wouldn't normally be apparent. This extended duration can test the limits of concentration and can suspend the perception of time (Demers 2010) (Voegelin 2010). In this composition, the concept of evolutionary time is explored through this extended duration. In the words of Joanna Demers, drone music becomes a 'maximal sound objects' that avoids development 'according to the standards of non electronic western music' and soundscape practices. Yet, through an aesthetic of excess that materialises over long stretches of time testing our listening concentration and perceiving of time, the composition develops via the 'tension between stasis and action' (ibidem). The proposed algorithm becomes then the technical means to the achievement of this aesthetic goal.

The everlasting feature is provided by the combination of the seeming unpredictability of the genetic algorithm and the dynamic character of the fitness function which avoids the reaching of a prefixed target. The video source is the element enabling for the dynamic attribute of the fitness function. In that regard, the video artwork and the musical output are not necessarily linked by any predetermined idea. The system was indeed tested with several video sources, from abstract to real footage. This approach, however, depicts an aesthetic which aims at providing a link, if even speculative, between visual content and sonic output. This link, if supported by a knowledge of the algorithm in place, is therefore exclusively subjective and created in the mind of the observer. However, provided the aesthetic of utopian excess, extended duration, perception of time and phenomenology, the authors have experimented with randomly generated colour pixel noise. It is believed, however, that a post-modern psychedelic approach to the development of the video material, such as kaleidoscopic fractals, is also a valid approach that will be investigated in the near future.

Conclusions

This paper has presented and discussed the idea for a genetically generated a/v drone composition in which the analysis of the video source influenced the fitness function causing it to change endlessly.

The combination of video art with sound art is not a new one, but is a tradition that has often been revisited as new avenues of study and research become available. The paper offered both a technical and aesthetic discussion on the premises of the presented work. As the appropriation of the work predominantly comes via a refined aesthetic approach, future works will aim towards a more clear definition of the modalities in which the artwork (or a series of the pieces in which the same system is deployed) can be presented. Currently, these may include digital format distribution over the internet, public multimedia installations, a/v concert performances. From a technical perspective, future work may include creating a closed loop in which the video not only drives the sound but the sound feeds back into the system affecting the video output. Additionally, the use of colourspaces which more accurately reflect human visual perception and video analysis using shape or edge detection are being explored.

References

Alfonseca, M.; Cebrian, M.; and Ortega, A. 2006. A Fitness Function for Computer-Generated Music Using Genetic Algorithms. *WSEAS Transactions on Information Science and Applications,* 3.3:518–525.

Balkwill, L. L., and Thompson, W. F. 1999. A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues. *Music Perception* 43–64.

Biles, J. A. 1994. GenJam: A Genetic Algorithm for Generating Jazz Solos. *Proceedings of the 1994 International Computer Music Conference*.

Burton, R., and Vladimorova, T. 1999. Generation of Musical Sequences with Genetic Techniques. *Computer Music Journal* 23(4):59–73.

Cipriani, A., and Giri, M. 2013. *Electronic Music and Sound Design - Theory and Practice with Max and MSP*, volume 1. Rome: ConTempoNet s. a. s., 2nd edition.

Demers, J. 2010. *Listening Through the Noise - The Aesthetics of Experimental Music*. New York: Oxford University Press.

Dodge, C., and Jerse, T. A. 1997. *Computer Music - Synthesis, Composition, and Performance*. Boston: Schirmer.

Edwards, M. 2011. Algorithmic composition: Computational thinking in music. *Communications of the ACM* 54(7):58–67.

Fels, S., and Manzolli, J. 2002. *Interactive, evolutionary textured sound composition*. Vienna: Springer.

Freitas, A. R. R., and Guimaraes, F. G. 2011. Melody Harmonization in Evolutionary Music Using Multiobjective Genetic Algorithms. *Proceedings of the Sound and Music Computing Conference*.

Fujinaga, I., and Vantonne, J. 1994. Genetic Algorithms as a Method for Granular Synthesis Regulation. *Proceedings of the International Computer Music Conference* 138–138.

McCormack, J.; Bown, O.; Dorin, A.; McCabe, J.; Monro, G.; and Whitelaw, M. 2014. Ten Questions Concerning Generative Computer Art. *LEONARDO* 47(2):135–141.

Miranda, E. R.; Kirke, A.; and Zhang, Q. 2010. Aritificial Evolution of Expressive Performance of Music: An Imitative Multi-agent Systems Approach. *Computer Music Journal* 34(1):80–96.

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.

Supper, M. 2001. A Few Remarks on Algorithmic Composition. *Computer Music Journal* 25(1):48–53.

Tzimeas, D., and Mangina, E. 2009. Dynamic Techniques for Genetic Algorithm-based Music Systems. *Computer Music Journal* 33(3):45–60.

Voegelin, S. 2010. *Listening to Noise and Silence - Towards a Philosophy of Sound Art*. New York: Continuum International.