Convergence Synthesis of Dynamic Frequency Modulation Tones Using an Evolution Strategy

Thomas J. Mitchell and Anthony G. Pipe

Faculty of Computing, Engineering & Mathematical Sciences, The University of the West of England, Frenchay, Bristol, BS16 1QY, UK
`thomas.mitchell@uwe.ac.uk` `anthony.pipe@uwe.ac.uk`

Abstract. This paper reports on steps that have been taken to enhance previously presented evolutionary sound matching work. In doing so, the convergence characteristics are shown to provide a synthesis method that produces interesting sounds. The method implements an Evolution Strategy to optimise a set of real-valued Frequency Modulation parameters. The development of the evolution is synthesised as optimisation takes place, and the corresponding dynamic sound can be observed developing from initial disorder, into a stable, static tone.

1 Background

Horner et al. have presented a collection of evolutionary dynamic-tone matching systems applied to a variety of synthesis techniques: Frequency Modulation Synthesis (FM) [1], [2], Wavetable synthesis [3], and Group Synthesis [4]. Throughout Horner's work, a Genetic Algorithm (GA) is employed to optimise a set of static wavetable *basis-spectra*[1], which are, as in wavetable synthesis, combined in time-varying quantities to match dynamic target tones.

The Evolution of modular synthesis arrangements and interconnections has formed the subject of two independent studies. Originally, Wehn applied a GA to 'grow' synthesis 'circuits' [5]. Later, Garcia used Genetic Programming to evolve tree structures that represent the synthesis topology [6].

A further matching technique has been put forward by Manzolli [7]. The Evolutionary Sound Synthesis Method (ESSynth) evolves waveforms directly applying the principles of Evolutionary Computation to recombine and mutate waveform segments. At each generation, ESSynth passes the 'best' waveform to an output buffer for playback allowing users to experience the evolution as it takes place. Wehn also observed this phenomenon, noting that sounds produced throughout the search phase can be 'quite entertaining', and occasionally 'rich and strange' [5].

The convergence synthesis technique, presented here, is built upon the work introduced above, and is concerned with the dynamic sounds that are produced as an Evolutionary Algorithm converges upon an optimal target match. The synthesis model at the heart of this process is based upon Horner's early FM

---

[1] Where *basis-spectra* refers to the harmonic content of a wavetable

matching experiments and provides initial steps towards a real-valued extension to his work.

2 Tone Matching with Frequency Modulation Synthesis

FM is an effective musical synthesis tool [8]. Despite the efficiency with which complex tones can be synthesised, FM is often regarded as unintuitive and cumbersome. Consequently, parameter estimation, specifically for the reproduction of acoustic musical instrument tones, has formed the subject of numerous academic studies. Recent advances in sound matching with FM have been presented by Horner [2]. Horner's early work utilised a unique FM arrangement referred to as *formant* FM. The *formant* FM model is ideal for matching harmonic instrument tones, as the carrier frequency can only be set to integer multiples of the modulating frequency (which is tied to f_0). Restriction of the synthesis parameters in this way significantly reduces the search complexity, as non-harmonic solutions are excluded from the model space. However, with the omission of non-integer variables, the algorithm is only able to match a discrete number of harmonic sounds, and could not be applied directly to regular FM models, as the majority of the sound space would be inaccessible.

Throughout Horner's FM matching experiments, a GA was applied to optimise a set of FM synthesis variables. GAs' perform their genetic operations on bit-strings, which are naturally suited to integer based combinatorial search problems. With the intention of expanding the search into the full parameter space, a reduced static-tone version of Horner's *formant* FM matching model was developed without limiting the synthesis parameters to integer numbers. Whilst GAs' can be modified to represent real-valued numbers, with specialised operators that permit arithmetic crossover and ordinal mutation, an Evolution Strategy was chosen, for the work presented here, as it provides a powerful optimisation paradigm that is naturally real-valued [9].

Objective Function. In previous sound matching work, objective calculations are often performed in the frequency domain using the *Squared Spectral Error* (SSE) (1), or a variant thereof [1] - [7]. The target spectrum for the evolutionary algorithm is obtained via the spectral analysis of the target waveform, prior to the execution of the search. A complete run of the objective function can be summarised as follows:

1. Insert candidate solution into the FM model,
2. Subject the corresponding synthesised waveform to spectrum analysis,
3. Calculate error between target and synthesised candidate spectra.

The SSE is given by the equation:

$$SSE = \sqrt{\sum_{b=0}^{N_{bin}} (T_b - S_b)^2} \ . \tag{1}$$

T = The target spectrum amplitude coefficients
S = The synthesised candidate spectrum amplitude coefficients
N_{bin} = The number of frequency bins produced by spectrum analysis
As will be demonstrated in section 3, a modified SSE metric is sufficient for this work.

3 Real Valued Static Formant FM Matching

An initial step towards the creation of a real-valued extension to Horner's FM matching algorithm has been developed for matching exclusively static tones. The matching synthesiser employs a single carrier/modulator FM arrangement, similar to Horner's *formant* model. The synthesis parameters, *Modulation Index* and *Carrier Frequency Multiple*, can be set to any value within the range [0, 15]. The object parameters are permitted to search regions of the sound space that would be unavailable with an integer restricted model. An ES, with strategy parameters[2] (5/5,25) is employed to optimise the synthesis parameters. To ensure that globally-optimal matches are consistently achieved, target tones are generated by the matching synthesiser. *Contrived* target tones are useful for testing purposes as they are known to exist within the matching sound space. Successful matches, therefore, yield parameters identical to those with which the target tone was produced.

Landscapes. The ES was found to become trapped at locally-optimal points of the SSE object landscape, see figure 1(a). Very rugged, Multi-modal landscapes, such as this, are problematic for any optimisation engine (including EC). To overcome this problem, the spectrum of the target and synthesised tones (T and S), are modified according to (2), to produce *windowed* spectra TW and SW which then replace T and S in (1) to provide the *Windowed Squared Spectral Error* (WSSE) metric.

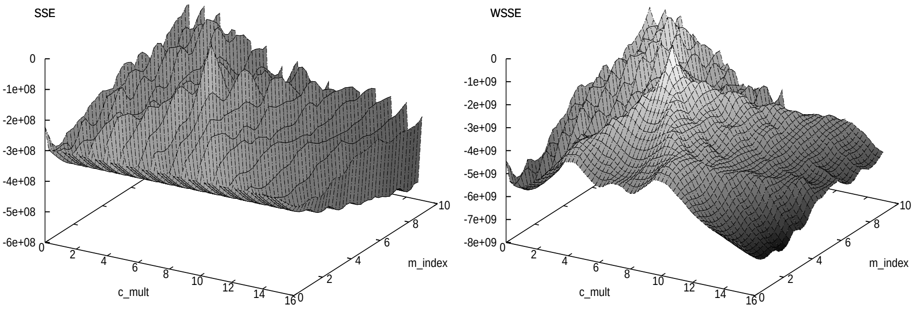
$$XW_b = \sum_{n=0}^{N_{bin}} \sum_{b=0}^{w} \left(\frac{w-b}{w} \right)^2 (X_{n+b} + X_{n-b}) \ . \tag{2}$$

X = The spectrum amplitude coefficients
XW = The 'windowed' spectrum amplitude coefficients
w = The bandwidth of the window

Windowing, allows spectrum error to be measured across a band, which has a smoothing effect on the object landscape. The landscapes of SSE and WSSE are plotted in Fig. 1(a) and (b) respectively. The location of the optimum is immediately obvious in the WSSE plot as the surrounding landscape slopes downwards isotropically. With a (5/5, 25) evolution strategy using the WSSE

---

[2] In the standard ES format (μ/ρ,λ) where μ is the parent size, ρ is the mixing number and λ is the offspring size



(a) (b)

Fig. 1. Exhaustive SSE & WSSE landscape plots *Carrier Frequency Multiple*/c_mult = 7.00 and *Modulation Index*/m_index = 4.00

objective function, a globally optimal solution is located, for any target parameter setting, within 20 generations. A typical convergence plot is provided in Fig. 2.

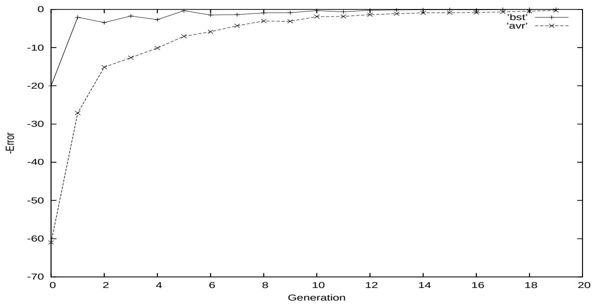


Fig. 2. Typical convergence plot for the evolutionary matching process

4 Convergence Synthesis

Running an FM synthesiser program in parallel with the ES allows convergence to be monitored aurally in real time. At the turn of each generation, the strongest offspring (or, in fact, any other) is passed to the synthesiser and a corresponding tone can be heard. Prior to the evolutionary process, the user is required to select a target tone by adjusting the synthesis parameters, *Modulation Index* and *Carrier Frequency Multiple*, via their respective sliders on the program interface. The target provides the final static tone upon which the ES will converge. The application responds to midi messages, enabling the user to set the modulator frequency (f_0) and inspect the target tone before the evolutionary synthesis process begins. To ensure that there are no audible clicks as the synthesiser is played, a short fade-in (*attack*) function is applied to the carrier amplitude on receipt of a MIDI *note-on* message, and equivalent fade-out (*release*) function on corresponding *note-off* messages. Following the selection of a satisfactory target tone, the evolutionary synthesis process begins. The *note-on* message, in addition to initiating the carrier fade-in function, triggers the initialisation of the ES. Individuals are randomly seeded and optimisation begins. The synthesis parameters are controlled by the ES as they progress towards their target values. Random reseeding ensures that the evolutionary path to the optimum is never the same twice.

The frequency plots, of the sound produced as a (20/20,25) evolution strategy converges upon it's FM target tone, are illustrated in Fig. 3. An initial stochastic period can be observed giving rise to the smooth harmonic partials of the target tone. After approximately 3.5 seconds the strategy has converged and the tone is stable.

Evolutionary Synthesis Parameters. The temporal characteristics of the sounds produced by convergence synthesis are closely coupled with the exogenous strategy parameters μ, ρ and λ. The variables that control the ES are, themselves, synthesis parameters that now control the sound directly. As μ and λ are varied, their values and respective ratio affect the dynamic characteristics of the synthesised tone. The former adjusts the selection pressure and, thus, the generational rate of convergence. The latter controls the period of each generation, as the objective function is called λ times. As the ES progresses, the offspring become increasingly similar until each λ is identical. This homogenising of the population is apparent in the convergence plot of Fig. 2. By passing the n^{th} *fittest* λ to the synthesiser, the rate at which the tone stabilises can again be prolonged, increasing the duration of the initial transient.

5 Results & Future Work

Matching can be carried out on any tone within the synthesis parameter space. The matching process itself provides time-variant parameter control to produce dynamic sounds. From one run of the ES to the next, the path to the target is different, providing a tone evolution that is varied yet predictable. This effect can be observed in Fig. 3, where two differing convergence sonograms can be seen converging upon the same target. There is considerably more high frequency content visible between 0.0s and 0.1s on the Fig. 3(a), which takes slightly longer to converge than Fig. 3(b). It would be desirable to expand the FM synthesis model with the use of multiple carrier/modulator, nested modulator or even feedback arrangements. The ES would then be required to optimise a significantly more complicated multi-dimensional landscape. Optimisation of parameter envelopes may also enable the matching of dynamic target sounds, both harmonic and non-harmonic.

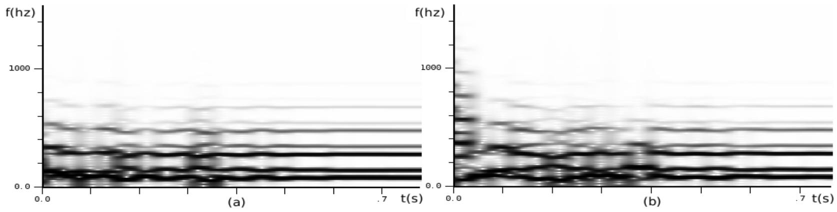


Fig. 3. Sonograms providing differing convergence paths for the same target tone

6 Conclusions

An evolutionary synthesis method has been presented that produces interesting dynamic sounds whilst approaching the match of a static target tone. The technique has emerged from the early developmental stages of a real-valued FM synthesis parametric optimisation process. The algorithm itself uses a basic Evolution Strategy, to optimise a set of *formant* FM synthesis parameters that most closely match a given static target tone. The sounds produced can be observed evolving from a stochastic initial transient, as a result of the random seeding of the initial population, into the static target tone that is chosen by the user at the beginning of the process. This convergence process also provides an alternative means by which an object landscape and convergence can be observed aurally.

References

1. Horner, A.: Spectral Matching of Musical Instrument Tones, PhD dissertation, University of Illinois Computer Science Department, Report No. UIUCDCS-R-93-1805 UILU-ENG-93-1720. (1993)
2. Horner, A.: Nested Modulator and Feedback FM Matching of Instrument Tones, IEEE Transactions on Speech and Audio Processing, 6(6), 398-409. (1998)
3. Wavetable Matching Synthesis of Dynamic Instruments, Audio Engineering Society Preprint No. 4389, pp. 1-23. (presented at the 101st Audio Engineering Society Convention). (1996)
4. Cheung, Ngai-Man, and Horner, A.: Group Synthesis with Genetic Algorithms, Journal of the Audio Engineering Society, 44(3), 130-147. Abstract on JAES Homepage. (1996)
5. Wehn, K.: Using Ideas from Natural Selection to Evolve Synthesized Sounds. Proc.Digital Audio Effects (DAFX98), Barcelona. (1998)
6. Garcia, R.: Growing Sound Synthesizers using Evolutionary Methods. European Conference in Artificial Life ECAL2001. Artificial Life Models for Musical Applications. University of Economics, Prague, Czech Republic. (2001).
7. Manzolli, J., Maia, A., Fornari, J., Damiani, F.: The evolutionary sound synthesis method. Proceedings of the ninth ACM international conference on Multimedia, Ottawa, Canada (2001)
8. Chowning, J.: The synthesis of complex audio spectra by means of frequency modulation. Journal of the Audio Engineering Society, vol. 21, pp. 526-534. (1973)
9. Beyer, H., Schwefel, H.: Evolution Strategies A Comprehensive Introduction. Natural Computing: an international journal, v.1 n.1, p.3-52, (2002)