Evolutionary Sound Matching: A Test Methodology and Comparative Study

Thomas J. Mitchell and David P. Creasey.
Bristol Institute of Technology,
Faculty of Environment and Technology,
University of the West of England, Bristol, UK.
Tom.Mitchell@uwe.ac.uk

Abstract

With the ever-increasing complexity of sound synthesisers, there is a growing demand for automated parameter estimation and sound space navigation techniques. Recent research in this domain has focused on the application of general-purpose evolutionary algorithms to match specific types of target sounds. However, it is difficult to establish whether success or failure of a particular match is due to the inefficiency of the optimisation engine, or the limitations of the matching synthesiser. In this paper the distinction between optimiser inefficiency and synthesiser limitations is elucidated with a contrived target test methodology that enables the performance of different optimisation techniques to be measured and compared. The methodology is applied to a Frequency Modulation synthesiser, in order to compare the performance of different Evolution Strategy-based algorithms. The algorithm producing the best results with contrived targets is then used to match a non-contrived acoustic instrument tone.

1. Introduction

Contemporary audio synthesisers provide composers with the musical freedom to play any known instrument electronically, or to explore beyond the realms of the known, to create sounds previously unheard. Controlled and efficient navigation of the sound space of a particular synthesiser requires expert knowledge of the underlying synthesis form, which may stem from an understanding of synthesis theory or experiential knowledge. It is often the case that composers must defer traditional notions of musicianship to concentrate on the task of synthesiser programming (manipulating parameters to produce the desired effect).

The synthesiser interface often presents an obstacle between artistic ideas and their expression. Each different synthesis technique is capable of creating a considerable range of timbres (tonal characters). The parameters which are used to shape the sound character are specific to the particular synthesis architecture being employed, and rarely relate to sound in human terms. Consequently, there is a complex mapping between the dimensions of a synthesis parameter (or control) space, and the perceived sound (or timbre) space. This can often result in a synthesiser control being unintuitive and more concerned with scientific process than artistic creativity.

If it were possible to relate the parameters of a synthesiser more directly to the user's timbral requirements, synthesis control could become more transparently about sound creation than computer programming. The first step to achieving this is the development of a process which is able to map known sound qualities onto sound synthesis parameters. This requires a matching technique that can efficiently search a synthesis parameter space to achieve specific aural requirements.

2. Background

Many efforts have been made to automatically derive synthesis parameters that match given target sounds. The most notable and advanced methods have employed the optimisation algorithms of evolutionary computation. The earliest evolutionary sound matching systems were presented by Horner [1], [2] for reproducing sounds produced by real acoustic instruments with Frequency Modulation (FM) and wavetable synthesis. More recent studies have extended this application to physical modelling synthesis [3], and additive and granular synthesis [4].

The difficulty in locating accurate target sound reproductions can be directly attributed to 3 components of the sound matching process:

1. the mechanisms of the underlying synthesis model,
2. the method by which match quality is quantified,
3. the characteristics of the target sound.

In many previous matching studies these aspects of the problem are not considered in isolation, but together as single problem component. That is, a synthesiser is chosen along with an appropriate similarity metric and optimiser. The resulting system is then applied to match arbitrary target sounds with results presented in the form of spectrum error plots. Consequently, it becomes difficult examine the pathology of inaccurate matches. It is not always clear if the matching synthesiser is incapable of reproducing the target, or if the optimisation engine is unable to negotiate the problem domain. It is desirable that the characteristics of the target sound are not circumscribed in the design of the matching system, so our attention is drawn to the latter issue.

3. Contrived matching method

The key aim is the development of a process which can efficiently negotiate the synthesiser's sound space in its entirety. By ensuring that the target sounds can be exactly matched by the synthesiser, it is possible to measure the performance of the optimiser within the search environment. This requirement leads naturally to contrived sound matching, originally introduced by Justice [5] in his early analytical matching work with FM.

A contrived target is a sound or tone that originates from within the search space of, and is generated by, the matching synthesiser. Contrived target sounds provide 2 significant advantages over experimentation with non-contrived alternatives:

- Firstly, and most importantly, it is simple to determine when an optimal solution has been found, as it will match the target sound exactly and achieve a relative spectral error of zero. If non-contrived target sounds are chosen as test specimens, confirmation of optimal convergence is not so easy. For example, the matching synthesiser may not be capable of exactly reproducing a particular target sound recorded from a real acoustic instrument, in which case a match delivering a relative error of zero cannot be achieved. In these circumstances an optimal match may only be confirmed when an exhaustive search yields no better result. In a high dimensional synthesis space this is not a feasible approach.
- Secondly, producing targets by randomly-generating points within the synthesis space ensures that the test set constitutes a diversity of search space positions, and thus assesses performance on a variety of search space landscapes (as the topology of the search space is dependent on the properties of the target sound). Moreover, repeated successful matching of random contrived targets demonstrates that it is possible to access all regions of the search space.

The contrived sound matching method represents a retrieval problem; the target is known to exist within the search space, and the ability of a search algorithm to derive its location is quantified. It may then be postulated that if it is possible to consistently and accurately match contrived targets, the system can then be applied to the problem of matching arbitrary target sounds. Match inaccuracy may then be attributed to the limitations of the synthesiser, as the optimisation algorithm is known to be well-suited to the problem domain. In section 7 the performance of 4 Evolution Strategy-based algorithms are examined in application to 3 different FM search environments. In section 8 the most effective algorithm is then used to match a tone produced by an acoustic instrument.

4. Test domain: FM synthesis

In this paper the chosen test domain of the contrived matching method is FM audio synthesis as originally defined by John Chowning [6]. FM audio synthesis enables complex spectra to be created simply and efficiently. In what is termed "simple FM", the instantaneous frequency of 1 sinusoidal oscillator is modulated by another. A diagram of the simple FM model is shown in figure 1.

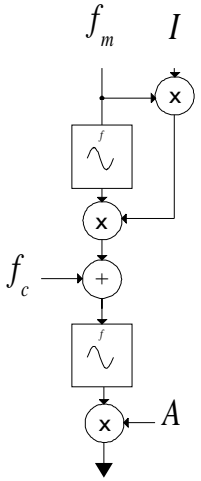

Figure 1. Simple FM model

In this model, the amplitude of the modulating oscillator controls the peak deviation of the carrier oscillator frequency from that specified by the parameter f_c. The amplitude function for simple FM is given by the formula:

$$e = A\sin\left(2\pi f_c t + I\sin 2\pi f_m t\right) \quad (1)$$

in which e is the modulated carrier output, A is the peak amplitude of the carrier, f_c and f_m are the carrier and modulator frequencies respectively, and I is the modulation index.

When I is assigned a value of zero there is no modulation, and the generated signal is a sine wave at frequency f_c. However, when $I > 0$, frequency partials are generated around the carrier at integer multiples of the modulating frequency (side-bands). The amplitudes of these partials are governed by the Bessel functions of the first kind and order n. This non-linear relationship between the synthesis parameters and the spectral form of the modulated signal can often complicate the process of sound design with FM. When parameters are altered by hand it can be difficult to find specific combinations of partials to produce a particular timbre. A process which is complicated further by the unintuitive effects of reflected side frequencies; partials synthesised with negative frequencies are directly mapped onto their positive values with negative phase. With so few parameters with which to access such a large range of output forms (and so sound characters), combined with non-linear effects outlined above, FM has become widely regarded as a difficult synthesis type to control [7], [8], [9] and [10].

In the work presented here the synthesis model in figure 1 (referred to as single simple FM) is examined alongside 2 parallel expansions, in which multiple instances of the simple FM model are summed at the synthesiser output, referred to as double and triple parallel simple FM respectively (figure 2).

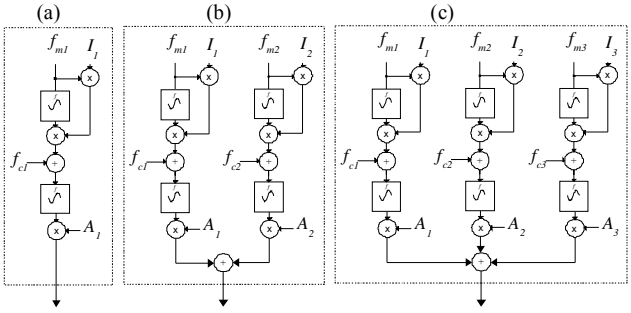

Figure 2. Single, double and triple parallel simple FM

5. Evolutionary Algorithms

The Evolutionary Algorithms (EAs) included in the experimentation are all based upon the Evolution Strategy (ES) developed originally by Rechenberg [13]. Due to space limitations, there is only room for a brief description of each EA, for a more comprehensive treatment the reader is directed to the references cited in each section.

5.1. Evolution Strategy (ES)

The traditional (μ, λ) ES is defined in [11], in which μ parent individuals are selected from λ offspring at each generation. Individuals are comprised from vectors of object and strategy parameters which are varied together in a process which has been termed self-adaptation [11]. For all experiments, the extinctive (*comma*) selection operator is employed and individuals are varied by discrete recombination [11] and derandomised mutation [12].

5.2. Multi-Start Evolution Strategy (MSES)

The MSES is a variant of the basic 2-membered (1+1) ES as defined originally by Reichenberg [11]. Multiple instances of the algorithm are evolved concurrently without recombination. This algorithm is also referred to as a multi-start hill-climber. Object parameters are mutated isotropically according to a single mutation step size, which is adapted by the $1/5^{th}$ rule.

5.3. Fuzzy Clustering Evolution Strategy (FCES)

The FCES [14] is a global optimisation EA designed to reduce the likelihood of preconvergence, by combining the local search characteristics of the ES with the strengths of fuzzy-cluster analysis. Evolution proceeds with the alternate application of optimisation and clustering, in which the fuzzy membership information is used to reduce the disruptive effects of variation, by limiting recombination between individuals belonging to different clusters. As the standard ES selection mechanism is employed, the entire FCES population is driven towards search space regions that offer the highest payoff, preventing the maintenance of multiple search space optima.

5.3. Clustering Evolution Strategy (CES)

The CES is a development of the FCES algorithm which has been designed by the first author to improve niching capabilities in rugged multimodal search space environments, such as the FM matching problem. The algorithm follows the same procedure as the FCES, however species are preserved with a local selection operator (termed *restricted cluster selection*), which ensures that a finite number of individuals from each cluster are carried into subsequent generations. Cluster boundaries are also reinforced with the use of *K*-means cluster analysis in place of the c-means fuzzy cluster analysis. The CES algorithm is represented by the pseudo code provided in figure 3, in which μ and λ represent the parent and offspring populations respectively, and t is the generational counter.

```
t = 0
initialise( μ(t) );
loop begin
        cluster( μ(t) )
        λ(t) = recombine( μ(t) );
        λ(t) = mutate( λ(t) );
        evaluate( λ(t) );
        μ(t+1) = select( λ(t) );
        t = t + 1;
loop end;
```

Figure 3. CES pseudo code

6. Experimental setup

6.1. Fitness measurement

Unsupervised sound matching requires a method for automatically determining the quality of a sound simulation. A fitness function is then required that enables strong individuals to be identified and selected for breeding.

In this experimentation sound similarity is measured by computing the relative spectral error between spectra of the target and candidate sounds. Recent studies, performed by Beauchamp *et al* [15], have established that the relative spectral error delivers the best correspondence to average discrimination data, extracted from human listeners, when compared with alternative spectral error metrics.

The relative spectral error is computed by accumulating the normalised difference between each frequency component of the candidate spectrum from their corresponding components in the target spectrum, both of which are extracted by Fast Fourier Transform (FFT). The error metric is defined by:

$$relitive\ error = \sqrt{\frac{\sum_{b=0}^{N_{bin}} (T_b - S_b)^2}{\sum_{b=0}^{N_{bin}} T_b^{\,2}}} \qquad (2)$$

Where T is a vector of the target spectrum amplitude coefficients, S a vector of synthesised candidate spectrum amplitude coefficients and N_{bin} the number of frequency bins produced by spectrum analysis. An exact match will result in a relative squared error of zero.

6.2. Synthesis parameter ranges

Within the matching system, the frequency parameters f_c and f_m are expressed as multiples (range 0-8) of 440Hz. The parameter A controls the carrier amplitude and thus the overall amplitude level of the synthesiser output (range 0-1). The amplitude of the modulating oscillator specifies the deviation of carrier frequency around f_c, and is controlled by the modulation index I (range 0-8).

No *a priori* information of the problem is employed in order to tune the system for matching any particular types of sounds (harmonic or otherwise). Each synthesis parameter is represented by a real number, which may take any value within the specified range without bias.

6.3. Algorithmic parameters

To ensure parity across all experiments, consistent algorithmic parameters and operators are fixed for all test-cases. Indicated results are calculated by the mean average and standard deviation of the final relative error produced by 50 separate runs when matching a set of 50 randomly generated contrived targets. Also indicated is the number of successful matches of the entire set, where a successful match produces a relative error of less than 0.05 (i.e. 95% of the target spectrum is matched). Each algorithm is tested when matching the same target set and populations are initialised with the same random data points, enabling performance differences to be attributed to the search properties of the EAs.

For the multimembered ESs, selection pressure is maintained at a fixed ratio of $\mu/\lambda = 1/7$, as indicated to be optimal by Schwefel [16]. Population sizes are set to (400, 2800) for all experiments and run for 100 generations. Algorithms that include cluster analysis partition the population into 80 clusters, corresponding to a cluster cardinality of 5. The parameters of the MSES have been chosen such that 1000 separate (1+1) ESs compute an equivalent number of fitness evaluations in 280 generations as the multimenbered ESs compute in 100 generations. The objective for each algorithm is to minimise the relative spectral error (equation 2).

7. Contrived matching results

For each algorithm type, the results of the contrived matching experiments for each synthesis model with each Evolutionary Algorithm are provided in table 1. S indicates the number of successful contrived matches out of 50, E is the mean, and σ is the standard deviation of the error for all 50 matches.

	ES			MSES		
Model	**S**	**E**	**σ**	**S**	**E**	**σ**
Single Simple FM	30	0.223	0.280	6	0.269	0.170
Double Simple FM	4	0.408	0.211	0	0.501	0.124
Triple Simple FM	2	0.428	0.193	0	0.578	0.106
	FCES			**CES**		
Model	**S**	**E**	**σ**	**S**	**E**	**σ**
Single Simple FM	34	0.157	0.241	50	0	0
Double Simple FM	6	0.385	0.220	10	0.246	0.151
Triple Simple FM	0	0.507	0.187	2	0.356	0.141

Table 1. Contrived matching results

From the result provided in table 1 it is apparent that the problem space becomes less tractable as the number

of parallel simple FM elements in the matching synthesiser is increased. This result is expected, as all algorithmic parameters remain constant while the dimensionality of the problem is increased. All tested EAs struggle to produce successful matches when optimising parameters for the triple simple FM model. The contrived matching method enables this deficiency to be attributed to the optimisation algorithms, since the target is known to exist within the search space. In all tests, CES produces the lowest average error with all 50 matches classed as successful in the experimentation with the single simple FM model. This result indicates that the CES is effective at navigating all regions of the simple FM search space. The performance advantage is attributed to the improved maintenance of population species due to the *K*-means cluster analysis method and the *restricted cluster selection* operator implemented in this algorithm. Of the other algorithms tested, the MSES appears to perform least well on all problems, while the FCES appears to outperform the ES on the smaller single and double FM models, while the reverse is true for the larger triple FM model.

8. Acoustic tone matching results

In the previous section, the CES was found to be the most advantageous for matching contrived FM target tones. In this section the CES is used to derive FM synthesis parameters that match a real acoustic instrument tone. The target tone originates from a muted trumpet recorded by Opolko and Warpnick [17]. The algorithm is set up as it was for the contrived experiments and the mean relative error and standard deviation are tabulated for 5 runs on each of the 3 FM synthesis models, with the population initialised randomly anew for each test case.

	CES	
Model	**E**	**σ**
Single Simple FM	0.199	0.003
Double Simple FM	0.137	0.0152
Triple Simple FM	0.119	0.008

Table 2. Trumpet tone matching results

Interestingly, the results (table 2) exhibit the opposite trend to those produced in the contrived matching experiments. Previously the relative error rates were shown to increase when using the larger synthesis models, whereas here, the error rates decrease with the larger model. These results illustrate the opposing limitations of the matching process. As established in section 7, the CES is well suited to the problem domain of the single simple FM model. The small standard deviation for this model suggests that all 5 runs have converged at the same fitness level - the optimum for this synthesis model. In attempting to match the trumpet target sound the CES has reached the limitations of the matching synthesiser. This error result cannot be improved upon unless a more elaborate synthesis model is employed. The introduction of additional oscillators to the model directly results in a more accurate match. While the results in table 1 suggested that the CES is less effective at exploring the double and triple simple FM synthesis spaces, when approaching the limitations of the matching synthesiser (table 2), the larger space is beneficial.

9. Spectrum plots

In the previous sections the accuracy of each match has been indicated in terms of the relative squared error. Figure 4 provides a spectrum plot of the original muted trumpet sound (top), and a corresponding match synthesised by the triple simple FM synthesiser (bottom) (this particular match achieves a relative squared error of 0.105).

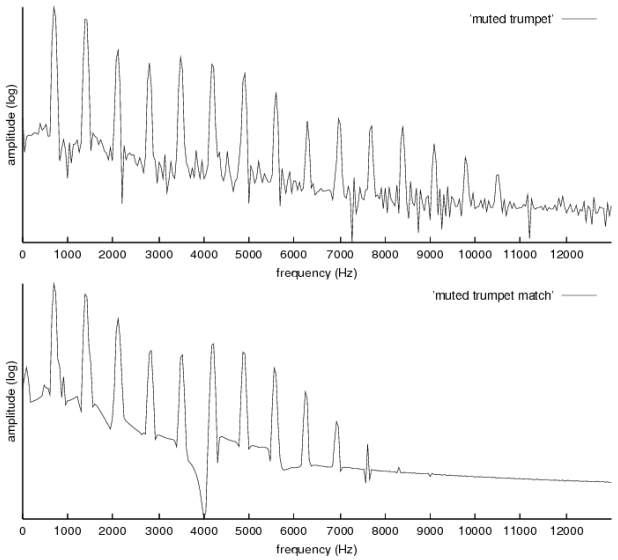

Figure 4. Muted trumpet tone (top), triple simple FM match (bottom).

The successes and limitations of the match are apparent in the figure. All of the peaks in the synthesised tone coincide accurately with the frequencies of partials in the target spectrum, with a good match in terms of amplitude (spectral envelope). However, several partials at the upper end of the frequency spectrum (beyond the 11^{th} harmonic) have limited amplitude or are absent from the synthesised sound. The matching synthesiser is not sufficiently complex to accurately recreate those details.

10. Conclusions

A contrived matching methodology has been developed and tested that enables practitioners to further understand the limitations of the sound matching process. It has been possible to clearly identify that the important

limiting factors in complex FM matching systems are the ability of the underlying evolutionary algorithm to navigate the complex synthesis space, and the ability of the synthesis form to recreate the target sounds. A selection of ES-based EAs have been tested and it has been possible, through the contrived matching method, to accurately compare their results. The proposed CES was shown to consistently outperform all of the other algorithms tested, as it is the only algorithm to enable species to be maintained. The CES-driven matching synthesisers were then applied to accurately match a real acoustic tone produced by a muted trumpet.

12. References

[1] A. Horner, J. Beauchamp, and L. Haken, "Machine Tongues XVI: Genetic Algorithms and Their Application to FM Matching Synthesis", *Computer Music Journal*, 1993, 17(4), 17-29

[2] A. Horner, *Spectral Matching of Musical Instrument Tones*, PhD dissertation, University of Illinois Computer Science Department, 1993.

[3] J. Riionheimo, and V. Välimäki, "Parameter Estimation of a Plucked String Synthesis Model Using a Genetic Algorithm with Perceptual Fitness Calculation", EURASIP Journal on Applied Signal Processing, Vol 2003 (8) pp 791-805.

[4] J. McDermott, N.J.L. Griffith, and M. O'Neill, "Toward User-Directed Evolution of Sound Synthesis Parameters", *EvoWorkshops*, Springer, 2005, pp 517-526.

[5] J. H. Justice , "Analytic Signal Processing in Music Computation", *IEEE Transactions on Acoustics, Speech, and Signal Processing*,Volume ASSP-27, Number 6, 1979.

[6] J. M. Chowning, "The synthesis of complex audio spectra by means of frequency modulation", *J. Audio Eng. Soc.*, 1973, v. 21, pp 526-534.

[7] R. Kronland-Martinet, P. Gillemain, and S. Ystad, "From sound modeling to analysis-synthesis of sounds", *Proceedings of Workshop on Current Research Directions in Computer Music*, Barcelona, Pompeu Fabra University, 2001, pp. 217-224.

[8] A. Horner, "Auto-Programmable FM and Wavetable Synthesizers", *Contemporary Music Review*, 2003, 22(3), pp 21-29.

[9] N. Delprat, "Global Frequency Modulation Laws Extraction from the Gabor Transform of a Signal: A First study of the Interacting Component Case", *IEEE Transactions of Speech and Audio Processing*, 1997, vol 5 (1), pp 64-71.

[10] R. Payne, "A microcomputer based analysis/ resynthesis scheme for processing sampled sounds using FM", *Proc. Int. Computer Music Conf.*, San Francisco, CA, 1987, pp. 282 289.

[11] H. P. Schwefel, *Numerical Optimisation of Computer Models*, New York: John Wiley & Sons, Inc. 1981.

[12] A. Ostermeier, A. Gawelczyk, and N. Hansen, "A derandomised approach to self-adaptation of evolution strategies". *Evolutionary Computation*, 1994, 2(4), 369-380.

[13] I. Rechenberg, Cybernetic solution path of an experimental problem, Tech. report RAE Translation 1122. Farnborough, Hants. 1965.

[14] J. C. W. Sullivan, 2001, *An Evolutionary Computing Approach to Motor Learning with an Application to Robot Manipulators*, PhD Thesis, University of the West of England, Bristol, 2001

[15] J. Beauchamp, and A. Horner, "Error metrics for predicting discrimination of original and spectrally altered musical instrument sounds", Journal of the Acoustical Society of America, 2003, Vol. 114(4), Pt. 2, p. 2325.

[16] H. P. Schwefel, "Collective phenomena in evolutionary systems". Pages 1025-1032 of: *Preprints of the 31st Annual Meeting of the International Society for General Systems Research*, 1987.

[17] F. Opolko and J. Wapnick, *McGill University Master Samples. (MUMS)*, Faculty of Music, McGill Univ., Montreal, QC, Canada. CD-ROM set, 1989.