Global Expectation-Violation as Fitness Function in Evolutionary Composition

Tim Murray Browne[1] and Charles Fox[2]

[1] Computing Laboratory
University of Oxford*
tim.murraybrowne@elec.qmul.ac.uk
[2] Adaptive Behaviour Research Group
University of Sheffield
charles.fox@gmail.com

Abstract. Previous approaches to Common Practice Period style automated composition – such as Markov models and Context-Free Grammars (CFGs) – do not well characterise global, context-sensitive structure of musical tension and release. Using local musical expectation violation as a measure of tension, we show how global tension structure may be extracted from a source composition and used in a fitness function. We demonstrate the use of such a fitness function in an evolutionary algorithm for a highly constrained task of composition from pre-determined musical fragments. Evaluation shows an automated composition to be effectively indistinguishable from a similarly constrained composition by an experienced composer.

1 Introduction

Mechanical composition in the Common Practice Period (CPP) style dates at least to Mozart's dice game, which presented 170 bar-length musical fragments and a first-order Markov transition matrix for assembling them into sequences [20]. The Markovian approach to automated composition has continued to the present day, and n-gram Markov models may be found in both generative [6] and evaluative [12] stages of composition. An alternative class of automated composition methods is the use of context-free grammars (CFGs) [4, 9].

Both Markovian and CFG approaches can produce 'correct' sounding compositions, but on extended listening can still sound dull and formulaic. Both approaches tend to lay out notes without considering the global context which is, of course, everything that is heard before *and after* a given note. The need for such a global structure in algorithmic composition is often identified as an open research question [8, 13]. A recent state-of-the-art evolutionary model for rhythm generation [7] identified but circumvented this issue by composing only rhythms and taking their cues from global structures provided in human-composed tracks of the composition.

* Tim Murray Browne is now a research student at Queen Mary University of London.

A new idea in automated composition is to provide a global structure by systematically violating listeners' expectations. Minsky [15] presented the idea of listening to music as exploring a scene, where we absorb different ideas and infer their structural relations. To maintain interest we must always be learning new ideas and relating them to previous ones, which gives some clue as to why Markov and rule-based processes that imitate what has already been heard produce dull results. The problem is that, whilst Markov chains and grammatical parsing are effective models of the *local* listening process, they are not so of the generative process and global perception. In this view, composing is about *manipulating* the listening processes and so whilst the composer must be able to evaluate their composition by listening, there are distinct mechanisms for creating music [18]. The listener needs to be taught ideas then surprised when they are broken or extended. Abdallah and Plumbley [1] recently modelled listening with information measures of a model-based observer. The Kullback-Leibler (KL) divergence, between before and after a note is heard, measured the average rate at which information arrives about the future. Applied to minimalist music, this showed that the listener is constantly receiving *new* information about the rules the music is following, and so learning its structure. This raises questions about the completeness of reductionist approaches to composing music such as CFGs. A composition cannot be broken down into ever smaller sections as ideas that are presented early on must develop and interact, and any part of a composition is always heard in the context of its absolute position within the piece. Analogously to linguistic generation, we may use knowledge of grammar to produce grammatically correct sentences, and with some advanced linguistics perhaps even make them coherent, but by themselves they does not produce interesting or moving prose, let alone poetry.

We present a new method for aesthetic analysis of harmony based on emotional tension, and its use as a fitness function in an evolutionary algorithm for a simplified compositional task. We combine Abdallah and Plumbley's notion of expectation-violation with a classic neural-network listening model [2]. Both of these models were intended for analysis and are here novelly redeployed for use in creative composition. Our compositional task is deliberately constrained, and the present work is intended to illustrate the use of the fitness function and suggest its incorporation into more advanced existing composition systems – we do not intend the simplified compositional task to be a full compositional system in itself.

2 The Global Structure of Expectation-Violation

To form an aesthetic judgement we require tools that work at a human level, emotionally and semantically. Meyer describes the aesthetic experience of listening to music as the *manipulation* of expectation involving the buildup of tension and subsequent release [14]. We will follow this approach and adopt a formal measure of tension based on Abdallah and Plumbley's unmet expectation of resolution. The importance of addressing tension and resolution in composition

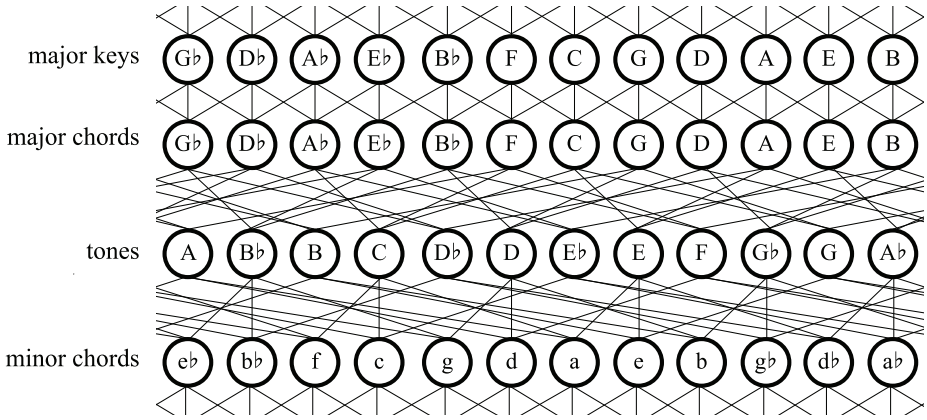


Fig. 1. Bharucha's network, which we redeploy for tension measurement. Links at the sides wrap around.

systems has been noted in recent research as well as the analysis of such qualities with respect to their context [17].

Throughout CPP music many common chord progressions occur. If we play the beginning of a common chord sequence the listener will experience anticipation of which chord is coming next. While we maintain this expectation we build up tension. To prevent our listener from becoming bored we require consistent build up and release of tension and also, to prevent our piece from becoming segmented, we require an increase in the intensity of these build-ups throughout the piece. As demonstrated in [1], if there is no recognisable ground laid, no expectation is produced in the listener and the music is just noise. Likewise if every expected event is granted, the music is predictable and boring. It is the *manipulation* of this expectation that allows us to build up tension.

Bharucha [2] presented the undirected, spreading-activation neural network of fig. 1 as a model of how *listeners* of CPP music infer structural associations between keys by using only the triad chord as foundation. Given a group of simultaneously sounded tones, activation around the network stabilised around related keys that would have been suggested by traditional music theory. Further studies established that people listening to a group of tones do indeed 'prime' related chords in a similar fashion to the network, suggesting expectation [3]. Note that, although based around the triad, the network does indeed capture the relationship between modified chords through how far they are connected. We will here consider Bharucha's past converged network states as specifying a transition prior for the forthcoming key state.

Bharucha's network uses a hierarchical structure of keys, chords and tones, with tones linked to chords they occur in and chords linked to keys they occur in (see fig. 1). At each time step, nodes corresponding to groups of the latest 'heard' tones are activated, then activation spreads through the network in update cycles until an equilibrium is reached. For example, if the tones of a C Major chord {C,

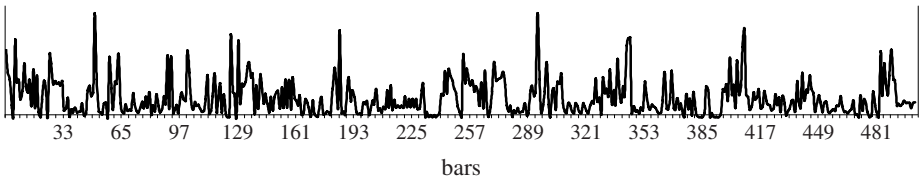


Fig. 2. Graph showing the value of V given in (3) over the entire first movement of Beethoven's fifth symphony.

E, G} are 'heard' together, the key tones reach equilibrium with energy levels monotonically decreasing as we move away in fifths from C.[3] The converged state of the network nodes, when appropriately normalised, may then be considered to represent a prediction on the next time step's key, chords and notes.

We define *key state* X_t at time step t as the normalised activation vector over keys k once the network is stable, and this is interpreted as a probability density function over the current key. Key state alone is used for making and testing predictions. During convergence, the activation x_j of node j is updated via weights w_{ij} as follows:

$$x_j = \sum_{i:i\to j} w_{i,j} x_i \tag{1}$$

where $\{i : i \to j\}$ are the nodes connected to j. $w_{i,j}$ are as in [2]: 0.244 between major chords and keys, 0.22 between minor chords and keys and 0.0122 between tones and chords.

In order to use this network as a model of harmonic expectation the analyser divides a score into temporal units and works through the piece in order, activating the nodes of the tones in each unit as it is heard, and allowing the network to stabilise at each stage. At time t we consider the key state X_t as our observation. We produce a prediction E_{t+1} of the next state by smoothing over all previous observations. We wish predictions to decay by a factor of γ over each beat if not reinforced giving:

$$E_{t+1} \propto \gamma^{1/n} E_t + X_t \tag{2}$$

where we have n units of time per beat. This is a two-variable, first-order Markov process (and is similar to the energy form of a Hidden Markov Model update).

This notion of expectation is meant at a physiological level rather than as a result of logical reasoning and is equivalent to cognitive 'priming' in [3]. We gain a measure of how far our harmonic expectations were violated by considering the KL-divergence of our prediction and observation of the key state X_t:

$$V(t) = \sum_k X_t(k) \log \frac{X_t(k)}{E_t(k)} \tag{3}$$

[3] Whilst we talk of notes in terms of absolute pitch rather than harmonic function (e.g. tonic, dominant), the network's symmetry makes these concepts equivalent within a fixed key.

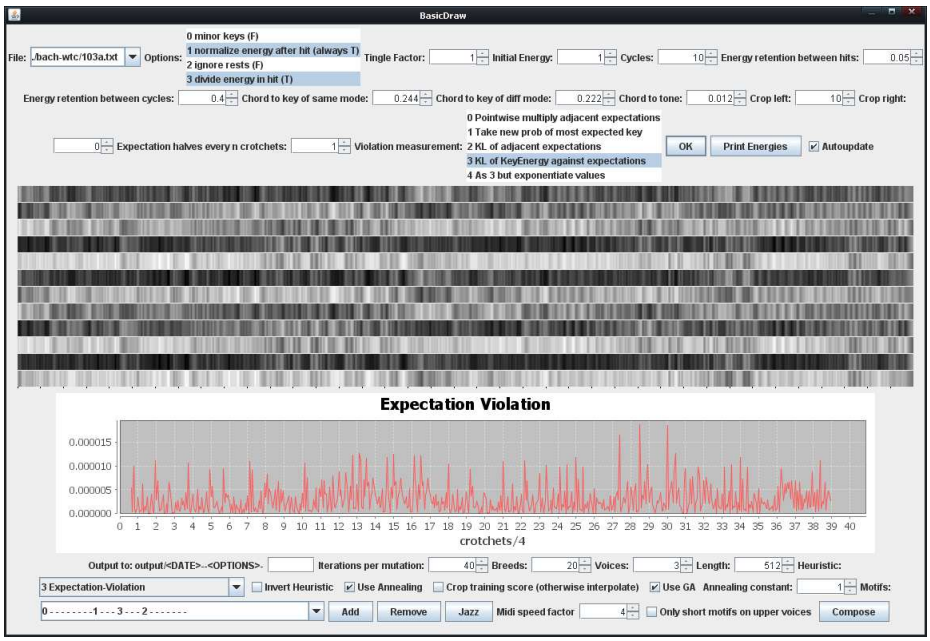


Fig. 3. The interface of the fragment arranger. The central graphic displays the normalised activation vector across the 12 key nodes after the network has stabilised after each unit of time. White is the maximum value and black the minimum.

As t increases, the model's expectation decays if not reinforced so it is not possible to maintain high tension with random notes. Fig. 2 is an example of V over time.

We now use this *local* measure of violation to construct a *global* fitness function R by considering its temporal structure. R compares the temporal tension profile of a candidate score c with that of a training score s:

$$R(c) = \left(\sum_t |V_s(t) - V_c(t)| \right)^{-1} \tag{4}$$

R thus assigns high fitness to new compositions having similar tension profiles to a known training composition, chosen for its pleasing tension profile.

3 The Fragment Arranging Task

As our performance measure is mostly concerned with harmony and less so with rhythm or melodic development we decided to reduce the problem of composing into the easier fragment arranging task. In this task, the program is presented with a number of pre-written motifs and is required to arrange them into a score in a musically desirable fashion. (More advanced motif-arranging systems may

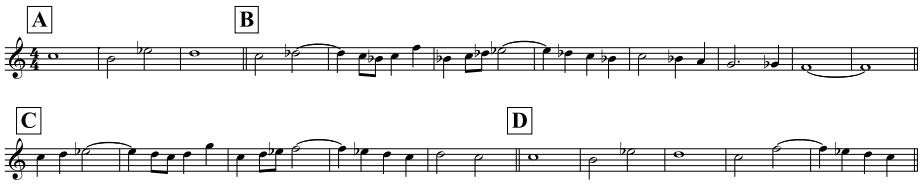


Fig. 4. Fragments used in the evaluation, taken from Fugue 4, Book I of Bach's *Well-Tempered Clavier*.

be found in [10] and the work of Cope [22].) A motif is defined as a sequence of notes and is typically a fragment of a melody that might be used several times by a composer. Arranging motifs into a score means deciding when and at what starting pitch they should be played (where motifs may be used any number of times, and may begin at any multiple of halfbeats through a measure). A score is made up of a number of voices and a voice may only be playing one motif at any one time. Far from being a 'toy problem' this demands an ingenious approach to harmony and produces interesting results quite different from the training material.

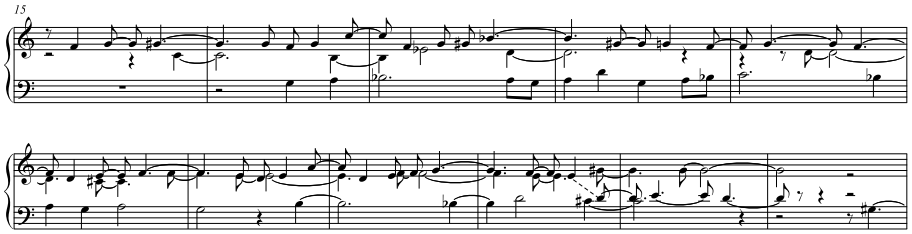
The task is therefore a state-space searching problem where we can move about the space by adding or removing motifs from the score. We implemented evolutionary search using a Metropolis-Hastings (MH) algorithm with simulated annealing [19]. In the MH algorithm the state space is randomly sampled for a proposed alternative to the current state (in our case sampling only involves adding or removing motifs). We accept any proposal that is better than the current state. We also accept some that are worse probabilistically based on how much worse and our *annealing schedule*, which reduces the chance of acceptance as the algorithm progresses. The probability of proposal p being accepted over the current score s was:

$$\Pr(accept) = e^{100(R'(p)-R'(s)).\beta} \tag{5}$$

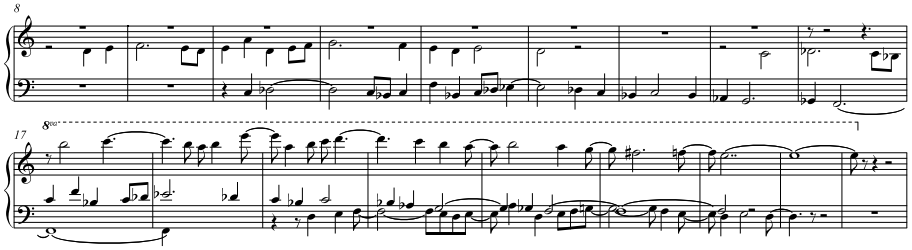
where β decreases linearly from 1 to 0 over the course of the algorithm. The fitness function $R' = RS$ used eqn. (4) together with a density evaluator S, which penalised scores with more than three quarters or less than a quarter of their units as rests. As the fitness function does not differentiate between notes of the same tone in different octaves, the octave that each fragment was placed in was decided by hand afterwards. (The MH algorithm may be viewed as a particular case of Genetic Algorithms: maintaining a population of one; using mutation but no crossover; and always generating two children, one of which is identical to the parent [11].)

4 Evaluation

Although the potential for computers in composition extends well beyond imitating human composers, the task of composing lends itself naturally to a Turing



(a) Computer



(b) Human

Fig. 5. Excerpts from the evaluation scores. The full performances may be listened to at `www.elec.qmul.ac.uk/digitalmusic/papers/2009/MurrayBrowne09evo-data`.

Test. The nature of the fragments task is reminiscent of a fugue so we decided for the evaluation to extract four short (2-8 measure) motifs from a Bach fugue (see fig. 4). The output piece was specified to be in 4/4 time, 32 measures long and in three voices for piano.

We also wished to demonstrate that the performance measure is quite abstracted from the notes and details of the training score, so we are actually modelling the listener's emotional response rather than comparing how similar the candidate score is to our training score. To do so, the same fugue that provided the fragments was used as the training score, to see if the arranger would try to reproduce it.

An experienced composer, Zac Gvirtzman, was given two days to write a piece to the same restrictions with the same input, having heard no program output. Both pieces were recorded by a pianist. We randomly approached 22 Oxford music students in their faculty, told them the composition brief, played them both pieces and then asked them to decide which was written by a human and which by computer. Presentation order was alternated. Excerpts from the two scores are shown in fig. 5. Having heard both pieces, nine answered correctly, 12 answered incorrectly and one refused to answer. Assuming a flat prior on

the population's probability of making a correct decision, the observations yield a Beta posterior having mean 0.42 and standard deviation 0.10, showing the human and machine to be effectively indistinguishable.

5 Discussion

The fragment task is by no means trivial, as confirmed by the composer. However, fragment-arranging is just one of many compositional tools that may be used by humans and larger composing programs, and the human composer commented that he felt very limited without having control over the dynamics and tempo. Even with such a small number of building blocks the program output sounded less repetitive than many Markovian and rule based approaches. Some outputs show original and resolving harmonies and are notable for their tolerance of dissonance that enhance the resolution. Despite not utilising any grammar the outputs have a strong sense of structure, which has been learnt from existing compositions. Interestingly, both computer and human scores used in the evaluation followed an ABA structure, with A drawing exclusively on fragment **A** (see fig. 4). By deliberately violating the listener's harmonic expectation in a controlled way they build up and release global tension, in a similar way to Abdallah and Plumbley's analysis of expectation in notes. They do still however lack a sense of purpose, which will be in part due to the model not having any learning capabilities. Future work may expand on the expectation provided by the listening model by incorporating it alongside Markov chains applied to melody, harmonic progressions and rhythm.

The abstraction of the fitness function from the training material is demonstrated by how, even with both the training material and the set of motifs to be arranged sourced from the same fugue, the output has not attempted to imitate it in any recognisable sense. Further work is required to determine how the function V given in (3) may be used without any training material. Experimenting with the program showed that obvious ways of analysing V such as considering its average over a large corpus of material or attempting to maintain some proportional cumulative value (e.g. $\sum_{i=1}^{t} V_c(i) = \lambda t$ for some λ) does not produce good results. Averaging V over a large corpus of work such as the *Well-Tempered Clavier* results in a fairly flat graph, dissimilar to any of the individual pieces in the work. This was the reason we used just a single composition for our fitness function rather than a set.

The space of scores searched by our evolutionary algorithm were intensionally constrained, and we do not claim to have constructed a full automated composition system. However the performance of the expectation-violation fitness function on the constrained task is encouraging, and suggests that it could be used to enhance most existing composition systems, if weighted appropriately and incorporated into their fitness functions.

References

1. Abdallah, S. and Plumbley, M.: "Information dynamics and the perception of temporal structure in music", Connection Science Journal, *in press.*
2. Bharucha, J. "MUSACT: A Connectionist Model of Musical Harmony", *Proc. Ninth Annual Meeting of the Cognitive Science Society*, Hillsdale, New Jersey, 1989.
3. Bharucha, J., Bigand, E. and Tillmann, B. "Implicit Learning of Tonality: A Self-Organising Approach", *Psychological Review*, 2000.
4. Fox, C.W. "Genetic Hierarchical Music Structures", *Proceedings of the International Florida Artificial Research Society Conference*, 2006.
5. Gill, S. "A Technique for the Composition of Music in a Computer", *The Computer Journal*, 1963.
6. Hiller, L. and Isaacson, L. "Musical Composition with a High-Speed Digital Computer", *J. Audio Eng. Soc.*, 1958.
7. Hoover, A.K., Rosario, M.P. and Stanley, K.O. "Scaffolding for Interactively Evolving Novel Drum Tracks for Existing Songs" *Proceedings of the Sixth European Workshop on Evolutionary and Biologically Inspired Music, Sound, Art and Design (EvoMUSART)*, Springer, 2008.
8. P. Husbands and P. Copley and A. Eldridge and J. Mandelis. "An Introduction to Evolutionary Computing for Musicians", *Evolutionary Computer Music, 1-27, Springer, 2007.*
9. Keller, R.M. and Morrison, D. R. "A Grammatical Approach to Automatic Improvisation", *Proceedings of the Fourth Sound and Music Conference*, 2007.
10. Keller, R., Hunt, M., Jones, S., Morrison, D., Wolin, A., and Gomez, S. "Blues for Gary: Design Abstractions for a Jazz Improvisation Assistant", *Electronic Notes in Theoretical Computer Science)*, 193:47-60, 2007.
11. Laskey, K. and Myers, J. "Population Markov Chain Monte Carlo", *Machine Learning*, 2003.
12. Lo, M.Y. and Lucas, S.M. "Evolving Musical Sequences with N-Gram Based Trainable Fitness Functions", *IEEE Congress on Evolutionary Computation*, 2006.
13. McCormack, J. "Open problems in evolutionary music and art.", *Proceedings of the Sixth European Workshop on Evolutionary and Biologically Inspired Music, Sound, Art and Design (EvoMUSART)*, Springer, 2005.
14. Meyer, L. *Emotion and Meaning in Music*, University of Chicago Press, 1956.
15. Minsky, M. "Music, Mind and Meaning", *CMJ*, 1981.
16. Moorer, J. "Music and Computer Composition", *Communications of the ACM*, 1972.
17. Papadopoulos, G and Wiggins, G. "AI Methods for Algorithmic Composition: A Survey, a Critical View and Future Prospects", *AISB Symp. Musical Creativity*, 1999.
18. Pearce, M. and Wiggins, G. "Aspects of a Cognitive Theory of Creativity in Musical Composition", *Proceedings of the ECAI02 Workshop on Creative Systems*, 2002.
19. Rao, R. P. "Bayesian computation in recurrent neural circuits", *Neural Computation*, 2004.
20. Schwanauer, S. and Levitt, D. *Machine Models of Music*, MIT Press, 1993.
21. Sundberg, J. and Lindblom, B. "Generative Theories in Language and Music Description", *Cognition*, 1976.
22. Wiggins, G. "Review of Computer Models of Musical Creativity by David Cope" *Literary and Linguistic Computing*, 2007.