Music Generation from Statistical Models *

Darrell Conklin
Department of Computing
City University
Northampton Square
London EC1V 0HB

Abstract

This paper discusses the use of statistical models for the problem of musical style imitation. Statistical models are created from extant pieces in a stylistic corpus, and have an objective goal which is to accurately classify new pieces. The process of music generation is equated with the problem of sampling from a statistical model. In principle there is no need to make the classical distinction between analytic and synthetic models of music. This paper presents several methods for sampling from an analytic statistical model, and proposes a new approach that maintains the *intra opus* pattern repetition within an extant piece. A major component of creativity is the adaptation of extant art works, and this is also an efficient way to sample pieces from complex statistical models.

1 Introduction

In his *Syntactic Structures* Chomsky (1957) used the term *creativity* to refer to the unique capacity of humans to understand and produce an indefinitely large number of sentences in a language, most of which have never been encountered or spoken before. The goal of a linguistic theory was the formulation of general principles for the evaluation of alternative grammars that could account for human creativity.

The construction of computational methods for musical style imitation has been far more difficult than initially imagined. Pioneering activity and early hope in the 1950s (Pinkerton, 1956; Brooks et al., 1956; Hiller and Isaacson, 1959), driven by advances in machine learning and information theoretic musicology (Meyer, 1956; Cohen, 1962), was soon replaced by frustration as the models used proved incapable of generating even simple well-formed melodies. This naturally led to the shift of research to hand-crafted algorithms for style imitation that had only small explicit statistical and empirical components (Lidov and Gabura, 1973; Baroni and Jacobini, 1978; Cope, 1991; Sundberg and Lindblom, 1991). While unquestionably successful, these models are heavily biased by their developers, and to the extent that they constrain the set of possible musical productions, are not robust and do not exhibit creativity. The promise of robust, creative, style-independent empirical learning methods for musical style imitation has not yet been fulfilled.

The study of grammars for music has a long history that has in many ways paralleled that of natural language. Chomsky's (1957) criticism of Markov (or finite-state, or n-gram) models for language proved to be damaging to both areas. Chomsky's main position was that a) finite-state methods could not capture, in a compact fashion, non-adjacent dependencies between words in sentences; b) they could not model the recursive embedding of phrase structure found in natural language; and therefore c) grammatical sentence production cannot be achieved by a left-to-right process. Though the relative impact of these points can be debated with respect to music, the effect in the 1970s and 1980s was to initiate the development of more powerful grammars for music generation, while at the same time suppressing the high empirical component of earlier work. In the 1990s work in computational language modeling, motivated by the applied task of speech recognition, made a dramatic shift back to statistical models (Jelinek, 1997). Improved methods for language modeling were developed that extended the basic n-gram model while maintaining its tractability and simplicity. Research in computational music modeling, partly driven by the growth of on-line music databases, also made this shift (Conklin and Witten, 1995; Ponsford et al., 1999), and also expanded into polyphonic music.

In other ways, computational modeling of music and natural language have taken different directions in their intended application. Natural language modeling was focused from the beginning of the rationalist program of the 1960s on developing *analytic* models — those intended to classify sentences as grammatical or non-grammatical. The main goal was thereby to demonstrate linguistic *competence*, while operational issues of sentence generation and production were relegated to the topics of *performance* and *pragmatics*. Motivated by astounding successes at speech recognition, computer scientists have recently been called upon to develop the best analytic models possible for natural language; the topic of sentence

*In *Proceedings of the AISB 2003 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences*, Aberystwyth, Wales, 30–35, 2003.

production has been confined to a small dedicated group of natural language generation researchers.

By contrast, in music, the situation has been nearly reversed. The focus from the early years was on *synthetic* models — those intended to generate well-formed sentences from a grammar. This is probably because the practical application of analytic models has not, until recently, been so clear as it has for statistical language models. The generation of music, yielding immediate results that can be listened to and evaluated by the researcher, is an exciting and compelling research project. The wholly subjective evaluation of results, however, is fraught with methodological problems (Pearce and Wiggins, 2001). Only recently have researchers in music informatics turned their attention to the analytic topics of music prediction (Conklin and Witten, 1995; Reis, 1999; Triviño-Rodriguez and Morales-Bueno, 2001; Pearce, 2003), phrase structure analysis (Bod, 2002), and music classification (Westhead and Smaill, 1993; Sawada and Satoh, 2000; Cruz-Alcázar and Vidal-Ruiz, 1997; Dubnov et al., 1998; Chai and Vercoe, 2001). A common thread to these analytic studies is the importance given to objective measures with which alternative models of music can be compared and ranked.

The thesis of this paper is that the topics of creative music generation and analysis are in fact highly interconnected, and that in principle there is no need to make the classical distinction (Ruwet, 1966) between analytic and synthetic models of music. Analytic statistical models have an objective goal which is to assign high probability to new pieces in a style. These models can guide the generation process by evaluating candidate generations and ruling out those with low probabilities. The generation of music is thereby equated with the problem of sampling from a statistical model, or equivalently, exploring a search space with the statistical model used for evaluation. This paper presents several ways that this sampling may be performed. Perhaps surprisingly, only a few of the proposed sampling methods have been explored in the music generation literature. For example, sampling may be performed in a way that generates incremental modifications to extant pieces in a style, retaining their overall repetition structure during the process. In this way, an obvious feature of human creativity — the use and adaptation of existing pieces — may be computationally explored.

This paper is structured as follows. In Section 2 a review of statistical models of music is provided. Particular emphasis is given to the prevalent class of *history-based* models, which condition events in music based on features computable from preceding events. In Section 3 methods for music generation from statistical models are outlined. The music generation process is equated with the problem of sampling high probability pieces from a statistical model. The prevalent random walk methods are discussed and shown to be flawed in that they do not guarantee that pieces with high overall probability will be produced. For certain simple types of statistical model, it is possible to compute the optimal pieces from the model. For complex models, computing the optimal pieces can be infeasible, and heuristic search or stochastic sampling is required. Modifications to extant pieces in a style is a very efficient way to sample high probability pieces from a model. The section concludes with an outline of a sampling approach that maintains the *intra opus* pattern repetition within an extant piece in a style.

2 Statistical models of music

A piece of music is represented by a sequence of *events*, which are music objects together with a duration and an onset time after instantiation into a piece. Notes are music objects, as are (recursively) sequences or simultaneities of music objects (Conklin, 2002). In addition to linear melody, this scheme allows for the representation of phrase structure and polyphony.

A statistical model of music assigns to every possible piece of music a probability. A statistical model captures regularities in a *class* of music, be this a genre, a style, a composer's style, or otherwise. A good model assigns high probabilities to pieces in the class, and low probabilities to pieces that are not in class. The probability of a piece p according to a model m is denoted $P(p \mid m)$.

In the Bayesian classification framework, there are several statistical models m_i, each of a different class i. A piece p is classified into the class i for which $P(p \mid m_i)$ is the highest (this assumes equal prior probabilities on all models, a detail which is not of concern in this paper). One of these models may be a *null model* which assigns equal probabilities to all valid pieces and is not permitted to assign a piece a higher probability than its true class model.

Statistical models of music are created empirically by induction. A corpus of training pieces in a class is used to instantiate the parameters of a statistical model. Competing models for the same class are evaluated according to their performance on a blind test set, or alternatively, by simulation using a cross-validation technique.

2.1 Context models

The most prevalent type of statistical model encountered for music, both for analysis and synthesis, are models which assign probabilities to events conditioned only on earlier events in the sequence. By the general term *context model*, we intend to include Markov, hidden Markov, n-gram, and finite state models. There are several reasons for the prevalence of context models in the music generation literature:

- events in a piece can be sometimes be predicted from preceding events;
- context models are easy to induce from examples: this typically involves storing subsequences in a data

structure providing rapid access (such as a suffix tree);

- context models are usually very fast: efficient data structures can be used to rapidly match contexts. This allows their use in real-time algorithmic composition systems;
- the probability of a piece according to a context model is easily computed, being simply the product of probabilities of events in the sequence;
- it is straightforward to generate new music with a context model.

Models which employ highly specific contexts to make predictions are bound to fail when applied to new music, because training corpora are always limited in size and few sequences of events will be encountered *inter opus*. Solutions to this *sparse data* problem range from simple data preprocessing steps such as transposition to a common key, to *smoothing* of short and long contexts (Jelinek, 1997) and the related PPM technique (Conklin and Witten, 1995).

2.2 Complex statistical models

The sparse data problem recalls the nativist position that children cannot possibly learn grammars from the limited amount of primary data encountered. Chomsky (1957) argued for innate human capacity and cognitive structures, with learning being more a form of abductive grammar selection rather than inductive inference from examples. In a piece of music, events can be ascribed properties, encoded into and derived from background knowledge, and these properties will be far less numerous than actual events. The method of *viewpoints* (Conklin and Witten, 1995; Conklin and Anagnostopoulou, 2001; Conklin, 2002) therefore deals with the sparse data problem by applying knowledge of music. Events are predicted by an interpolation of the predictions of multiple viewpoints. Similar ideas have been employed in statistical natural language modeling (Brown et al., 1992), where context models over lexical categories are developed.

The technique of viewpoints is an instance of the general class of *history-based* models, wherein any features computable from preceding events can be used to condition the probability of the current event. Furthermore, the current event can be predicted by an interpolation of two separate models: one which captures short-term phenomena specific to the current piece, and a long-term model which captures regularities of the general style (Conklin and Witten, 1995).

Statistical models are not restricted to general history-based models. More powerful grammars in the Chomsky hierarchy, such as context-free grammars, can have a statistical interpretation. Learning the parameters of these statistical models, however, presupposes an existing grammar for a musical style. A type of context-free grammar, called the *dependency grammar* (Chelba et al., 1997), may have promise for music. This type of model could elegantly capture dependencies between, for example, chord and non-chord tones in tonal music.

In a final analysis, however, music requires features even beyond the capabilities of context-free grammars. For example, the simple phenomenon of a repeated phrase in a piece cannot adequately be expressed using a context-free grammar. The sampling techniques discussed below for the generation of music from statistical models are applicable regardless of the power and complexity of the statistical model. The next section will outline a way to preserve pattern repetition in a piece.

3 Generation of music from statistical models

As outlined in the previous section, the goal of an analytic statistical model of a corpus of music is to assign high probabilities to pieces in the class, and lower probabilities to all other pieces. In the Bayesian framework, given multiple class models, a piece is classified by the model which assigns it the highest probability.

To generate a piece from an analytic model is therefore to sample a piece which has a high probability according to the model; presumably higher than its probability according to competing models. It is important to note that a high probability piece need not comprise only high probability events.

3.1 Random walk method

The simplest way to generate music from a history-based model is to sample, at each stage, a random event from the distribution of events at that stage. After an event is sampled, it is added to the piece, and the process continues until a specified piece duration is exceeded. This method has been employed by many, if not most, synthetic models of music (Brooks et al., 1956; Mozer, 1994; Conklin and Witten, 1995; Ponsford et al., 1999).

The random walk method, while applicable for real-time music improvisation systems that require fast and immediate system response (Assayag et al., 1999; Pachet, 2002), is flawed for generating complete pieces because it is "greedy" and cannot guarantee that pieces with high overall probability will be produced. The method may generate high probability events but may at some stage find that subsequently only low probability events are possible, or equivalently, that the distribution at subsequent stages have high entropy.

Allan (2002) provides a demonstration of this effect in music. The probability of the optimal sequence of harmonic symbols generated from a hidden Markov model (see next section) is compared to a sequence produced by the random walk method. The probability of the sequence

produced can be significantly lower with the random walk method.

3.2 Hidden Markov models and Viterbi decoding

A *hidden Markov model* (HMM) is a statistical model in which *observed events* are generated from underlying *hidden states*. After generating an event from a state, a model moves into a new state based on its transition probabilities. The term *hidden* derives from the fact that many different state sequences can generate the same observed sequence of events. Different state sequences can therefore have different probabilities.

A key concept of HMM theory is the *decoding* step: given a sequence of observed events, find the most probable hidden state sequence. For a first-order HMM, this decoding can be computed by a dynamic programming algorithm (called the Viterbi algorithm) in time proportional to the length of the observed event sequence times the number of states in the model. A similar dynamic programming algorithm can be applied to non-hidden first-order Markov models to find the most probable sequence.

HMMs have recently been applied with success to music generation. Farbood and Schoner (2001) describe a system for producing a counterpoint line to a cantus firmus in the style of Palestrina. For a given cantus firmus, the most probable counterpoint line is found by Viterbi decoding. Allan (2002) describes an HMM approach for chorale melody harmonization, where Viterbi decoding is used to produce the most probable underlying sequence of harmonic symbols for a given melody line.

3.3 Stochastic sampling

A drawback of Viterbi decoding is that its computation time increases exponentially with the context length of the underlying Markov model on states. Furthermore, more complex and powerful statistical models are not readily transformed into a form in which Viterbi decoding is applicable. Producing the highest probability pieces from complex statistical models is therefore a computationally expensive task, and heuristic search and control strategies must be applied. With an increase in computation time, techniques such as A^* search can be used to compute the n best state sequences (Jiménez et al., 1995). However, a system that produces only a few high-probability pieces obviously cannot be called creative, and more robust methods are required.

Sampling from complex statistical models can be performed in various ways (Rosenfeld et al., 2001). To generate a piece using *Gibbs sampling* from a model m, we start with some initial piece p. The following process is iterated: a random event of the piece is chosen and all valid events are substituted into that position, each producing a new piece p'. One such piece p', having the probability $P(p' \mid m)$, is chosen at random for the next iteration.

The Gibbs sampling method can be slow because the term $P(p' \mid m)$, which might itself be expensive to compute, must be computed many times. In *Metropolis sampling* from a model m, a random event in the piece p is chosen, and a *single* event is substituted into that position, producing a new piece p'. This piece is accepted for the next iteration if $P(p' \mid m) > P(p \mid m)$, and otherwise rejected with a rejection probability that increases with each iteration.

Both sampling methods discussed above can fall into local valleys, where valid substitutions at all locations cannot improve the probability of the current piece. Both methods can be assisted by providing a high probability extant piece as a starting point. This brings up the interesting fact that modifications to extant pieces in a style is a very efficient way to generate high probability pieces in a style.

3.4 Pattern-based sampling

The application of statistical sampling techniques to music faces two main challenges. First, it is unlikely that single-event substitutions can provide the necessary diversity to exhibit creativity. Motif or phrase-level substitutions seem to be required, and this again raises the issue of the high computation time needed to compute all valid motifs for substitution. This limitation is beyond the scope of this paper. Second, musical cohesion (Anagnostopoulou, 1997) is not preserved by the sampling procedures. For example, in a piece in which repetition of a certain pattern should be preserved a substitution of an event within one occurrence of the pattern should also be reflected in the other.

An early attempt to handle *intra opus* repetition with a complex statistical model involved the use of a short-term model that adapted to the current piece being generated (Conklin and Witten, 1995). A flaw with this approach is that while pattern continuation can easily be handled, there is no way to specify at the outset of generation where repeated patterns should begin and end. One way to handle this effect is to apply modern pattern discovery algorithms (Cambouropoulos, 1998; Rolland and Ganascia, 2000; Conklin and Anagnostopoulou, 2001) which can discover *intra opus* repetition, often at deeper levels than the basic musical surface.

A pattern-based sampling approach can first apply a pattern discovery algorithm to an extant piece to reveal patterns at various levels of abstraction. During stochastic sampling, the discovered pattern structure can be conserved. This is one way to overcome the limitations with n-gram models of music discussed earlier.

4 Conclusions

Analytic statistical models have an objective goal which is to construct models for a stylistic corpus that assign high probability to new pieces in the style. This essay

has argued for increased attention to analytic models of music. Operational issues of music generation, expressed as the problem of sampling from a statistical model, can be separated from the issues of model selection and training. The generation of music should not be confined to the prevalent random walk method for sampling from a history-based model. With music generation rephrased as a classical search and sampling problem, generation algorithms are free to apply deep knowledge and draw from extant pieces to reduce the search space and more rapidly focus on high probability pieces. The generation of music can use modern pattern discovery methods applied to extant pieces to reveal their repetition structure and provide musical cohesion to new productions. A major component of creativity is the adaptation of extant art works, and this is also an efficient way to generate music from complex statistical models.

References

M. Allan. Harmonising Chorales in the Style of J.S. Bach. Master's thesis, School of Informatics, University of Edinburgh, 2002.

C. Anagnostopoulou. Lexical cohesion in linguistic and musical discourse. In *Proc. 3rd ESCOM*, 1997.

G. Assayag, S. Dubnov, and O. Delerue. Guessing the composer's mind: applying universal prediction to musical style. In *Proceedings of the International Computer Music Conference*, Bejing, China, 1999. International Computer Music Association.

M. Baroni and C. Jacobini. *Proposal for a Grammar of Melody*. Les Presses de l'Université de Montréal, 1978.

R. Bod. Memory-based models of melodic analysis: Challenging the gestalt principles. *Journal of New Music Research*, 31(1):27–37, 2002.

F. P. Brooks, A. L. Hopkins Jr., P. G. Neumann, and W. V. Wright. An experiment in musical composition. *IRE Transactions on Electronic Computers*, EC–5:175–182, 1956.

P. Brown, V. Della Pietra, P. deSouza, J. Lai, and R. Mercer. Class-based n-gram models of natural language. *Computational Linguistics*, 18(4):467–479, 1992.

E. Cambouropoulos. *Towards a General Computational Theory of Musical Structure*. PhD thesis, University of Edinburgh, 1998.

W. Chai and B. Vercoe. Folk music classification using hidden Markov models. In *Proc International Conference on Artificial Intelligence*. 2001.

C. Chelba et al. Dependency language modeling. Technical Report Research Note 24, Center for Language and Speech Processing, The Johns Hopkins University, 1997.

N. Chomsky. *Syntactic Structures*. Mouton, 1957.

J. E. Cohen. Information theory and music. *Behavioral Science*, 7:137, 1962.

D. Conklin. Representation and discovery of vertical patterns in music. In C. Anagnostopoulou, M. Ferrand, and A. Smaill, editors, *Music and Artificial Intelligence: Lecture Notes in Artificial Intelligence 2445*, pages 32–42. Springer-Verlag, 2002.

D. Conklin and C. Anagnostopoulou. Representation and discovery of multiple viewpoint patterns. In *Proceedings of the International Computer Music Conference*, pages 479–485, Havana, Cuba, 2001. International Computer Music Association.

D. Conklin and I. Witten. Multiple viewpoint systems for music prediction. *Journal of New Music Research*, 24 (1):51–73, 1995.

D. Cope. *Computers and Musical Style*. A-R Editions, 1991.

P. Cruz-Alcázar and E. Vidal-Ruiz. A study of grammatical inference algorithms in automatic music composition and musical style recognition. In *Proc. ICML-97 Workshop on Automata induction, grammatical inference, and language acquisition*, 1997.

S. Dubnov, G. Assayag, and R. El-Yaniv. Universal classification applied to musical sequences. In *Proceedings of the International Computer Music Conference*, pages 322–340, 1998.

M. Farbood and B. Schoner. Analysis and synthesis of Palestrina-style counterpoint using Markov chains. In *Proceedings of the International Computer Music Conference*, pages 471–474, Havana, Cuba, 2001. International Computer Music Association.

L. Hiller and L. M. Isaacson. *Experimental Music*. McGraw–Hill, 1959.

F. Jelinek. *Statistical Methods for Speech Recognition*. MIT Press, 1997.

V. Jiménez, A. Marzal, and J. Monné. A comparison of two exact algorithms for finding the N-best sentence hypotheses in continuous speech recognition. In *Proc. 4th EUROSPEECH*, pages 1071–1074, 1995.

D. Lidov and J. Gabura. A melody writing algorithm using a formal language model. *Computer Studies in the Humanities*, 4(3-4):138–148, 1973.

L. B. Meyer. *Emotion and Meaning in Music*. University of Chicago Press, 1956.

M. Mozer. Neural network music composition by prediction. *Connection Science*, 6(2-3):247–280, 1994.

F. Pachet. Interacting with a musical learning system: the Continuator. In C. Anagnostopoulou, M. Ferrand, and A. Smaill, editors, *Music and Artificial Intelligence: Lecture Notes in Artificial Intelligence 2445*, pages 119–132. Springer-Verlag, 2002.

M. Pearce. Improving the prediction performance of PPM variants with monophonic music. In *Proc AISB 2003 Symposium on AI and Creativity in the Arts and Sciences*, 2003.

M. Pearce and G. Wiggins. Towards a framework for the evaluation of machine compositions. In *Proc AISB 2001 Symposium on AI and Creativity in the Arts and Sciences*, pages 22–32, 2001.

R. Pinkerton. Information theory and melody. *Scientific American*, 194:77–86, 1956.

D. Ponsford, G. Wiggins, and C. Mellish. Statistical learning of harmonic movement. *Journal of New Music Research*, 28(2), 1999.

B. Reis. Simulating music learning: on-line perceptually guided pattern induction of context models for multiple-horizon prediction of melodies. In *Proc AISB 1999 Symposium on Musical Creativity*, pages 58–63, 1999.

P-Y. Rolland and J-G. Ganascia. Musical pattern extraction and similarity assessment. In E. Miranda, editor, *Readings in Music and Artificial Intelligence*, chapter 7, pages 115–144. Harwood Academic Publishers, 2000.

R. Rosenfeld, S. Chen, and X. Zhu. Whole-sentence exponential language models: a vehicle for linguistic-statistical integration. *Computers, Speech and Language*, 15(1), 2001.

N. Ruwet. Méthodes d'analyse en musicologie. *Revue belge de musicologie*, 20:65–90, 1966.

T. Sawada and K Satoh. Composer classification based on patterns of short note sequences. In *Proc AAAI-2000 Workshop on AI and Music*, pages 24–27, 2000.

J. Sundberg and B. Lindblom. Generative theories in language and music descriptions. In P. Howell, R. West, and I. Cross, editors, *Representing musical structure*, pages 245–272. Academic Press, 1991.

J. Triviño-Rodriguez and R. Morales-Bueno. Using multiattribute prediction suffix graphs to predict and generate music. *Computer Music Journal*, 25(3), 2001.

M. Westhead and A. Smaill. Automatic characterisation of musical style. In *Music Education: An AI Approach*, pages 157–170. Springer-Verlag, 1993.