Melody transformation with semiotic patterns

Izaro Goienetxea[1] Darrell Conklin[1,2]

[1] Department of Computer Science and Artificial Intelligence
University of the Basque Country UPV/EHU, San Sebastian, Spain
[2] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
{izaro.goienetxea,darrell.conklin}@ehu.eus

Abstract. This paper presents a music generation method based on the extraction of a semiotic structure from a template piece followed by generation into this semiotic structure using a statistical model of a corpus. To describe the semiotic structure of a template piece, a pattern discovery method is applied, covering the template piece with significant patterns using melodic viewpoints at varying levels of abstraction. Melodies are generated into this structure using a stochastic optimization method. A selection of melodies was performed in a public concert, and audience evaluation results show that the method generates good coherent melodies.

1 Introduction

In recent years the topic of computational music generation has experienced a dynamic renewal of interest, though automation of music composition has intrigued people for hundreds of years. Even before the age of computers the idea of automatic music composition existed. A classical example of the automatic composition idea is the *Musikalisches Würfelspiel* or musical dice game, like the one published in 1792 that was attributed to Mozart [16].

Statistical models of symbolic music have been prevalent in computational modelling of musical style, since they can easily capture local musical features by training on large corpora rather than hand coding of stylistic rules [1, 7, 12, 15]. The lasting impact of statistical models on the topic of music generation spans from the earliest Markov models [5] to new variants of statistical models based on deep learning [4] and grammatical methods [20].

An issue faced by all methods for music generation is the *coherence problem*: ensuring that music material repeats or recalls in a more abstract sense material presented earlier in the piece. Nearly all forms of music involve repetition [17], either at the surface or deeper structural levels, and repetition imparts meaning to music [18]. Though early knowledge-based methods [13] explicitly considered repetition, the problem of achieving coherence in music generated from machine learning models remains largely unsolved.

A natural way to describe the coherence of a piece of music is by constructing a *semiotic structure*, defined as a representation of similar segments by a limited set of arbitrary symbols, each symbol representing an equivalence class

of segments [3]. A key observation is that a semiotic structure can be "inverted", generating new music by instantiating the symbols and retaining the abstract equivalence structure though having completely new music material [9]. The procedure can therefore be seen as *generation by transformation*: retaining abstract aspects of a template piece while modifying specific material.

Progress on the coherence problem was made recently in the music generation method of Collins et al. [6], where similar segments are identified by patterns indicating transposed repetitions in Chopin mazurkas. These "geometric" patterns are only suitable for carefully selected examples, because repetition in music need not be restricted to rigid transpositions. Consider, as an illustration, the simple melodic fragment of Figure 1. Though the two indicated phrases are clearly related, apparent in the score and to any listener, this is not by sharing an interval sequence, but rather an abstract contour sequence. The method described in this paper is able to naturally handle such musical phenomena using heterogeneous patterns discovered automatically using various viewpoints.

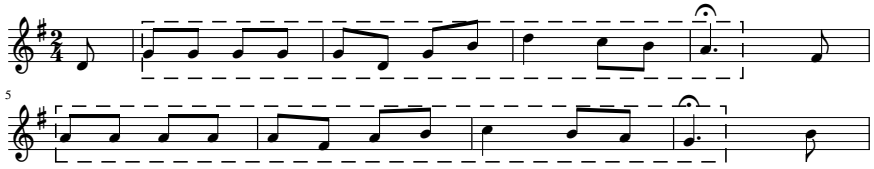

Fig. 1. First two phrases of the melody *Begiztatua nuen*[1]. The two phrases are related by an abstract melodic contour relation and there is no transposition that carries one into the other.

The style chosen to model is the folk style of *bertsos*. These are improvised Basque songs, sung by *bertsolaris*, that respect various melodic and rhyming patterns and which have fixed rhythmic structures. They can be classified into traditional folk melodies, new melodies, and melodies that are specifically composed. Bertso melodies usually have repeated and similar phrases, making them a challenge for statistical models and a good style for exploring the coherence problem. In this paper rhythmic aspects are conserved, so that the new melody can be used with lyrics created for the original melody.

The corpus used for this study is the *Bertso Doinutegia*, a collection of bertso melodies compiled by Joanito Dorronsoro and published for the first time in 1995 [14]. It currently contains 2379 melodies and is maintained and updated every year by Xenpelar Dokumentazio Zentroa[2] with new melodies that were used in competitions and exhibitions. Scores in the collection were encoded in Finale and exported to MIDI. Metadata associated with each song includes the

[1] http://bdb.bertsozale.eus/en/web/doinutegia/view/137-begiztatua-nuen-euskaldun-makila

[2] http://bdb.bertsozale.eus/es/

melody name, the name or type of the strophe, type of the melody, composer, bertsolari who has used it, name and location of the person who has collected the melody, and year of the collection. Some of the melodies in the collection have links to recordings of exhibitions or competitions where those melodies were used.

2 Methods

The transformation process presented in this paper has five main components: viewpoint representation; pattern discovery applied to a template piece to identify similar segments; pattern ranking and covering to form the semiotic structure; statistical model construction; and generation from the statistical model.

2.1 Viewpoint representation

To describe the template piece on different levels of abstraction a multiple viewpoint representation [9, 12] is used. A *viewpoint* τ is a function that maps an event sequence $e_1, \ldots, e_\ell$ to a more abstract derived sequence $\tau(e_1), \ldots, \tau(e_\ell)$, comprising elements in the codomain of the function τ.

viewpoint	codomain
pitch	$\{50, 52, 53, \ldots, 83\}$
dur	$\{1, 2, 3, \ldots\}$
onset	$\{0, 1, 2, \ldots\}$
intpc	$\{0, \ldots, 11\}$
int	$\{-14, -12, -11, \ldots, 14, 15, 17\}$
3pc	{d, eq, u}
5pc	{ld, sd, eq, su, lu}
d3pc	{d, eq, u}

Table 1. A specification for a small set of viewpoints.

Table 1 presents five melodic viewpoints pitch, int, intpc, 3pc and 5pc, and three rhythmic viewpoints dur, onset, and d3pc. The viewpoint pitch represents the MIDI number of each event; the viewpoint int computes the interval between an event and the preceding one; the viewpoint intpc computes the pitch class interval (interval modulo 12) between an event and the previous one. A three-point contour viewpoint 3pc computes the melodic contour between two events: upward (u), downward (d) or equal (eq); and a five-point contour viewpoint 5pc computes whether the contour between two contiguous events is more than a scale step down (ld), is one scale step down (sd), is more than a scale step up (lu), is one scale step up (su), or stays equal (eq). The duration contour viewpoint

d3pc computes if the duration of a note is shorter (d) than the previous one, longer (u) or equal (eq). The viewpoint representation of an example segment, using several viewpoints of Table 1, is shown in Figure 2.

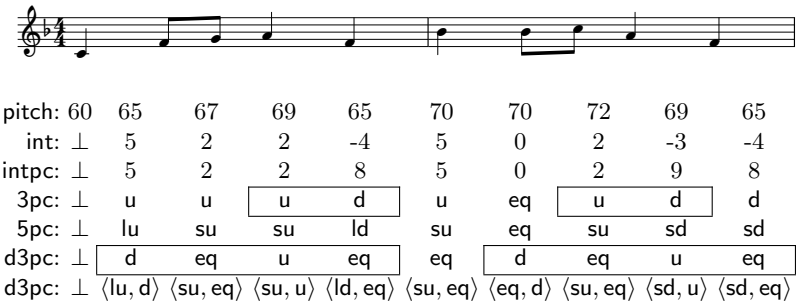

Fig. 2. A fragment from the melody *Abiatu da bere bidean*[3] and its viewpoint representation. Two patterns are highlighted.

To represent the interaction between melodic and rhythmic viewpoints, melodic viewpoints are *linked* with the rhythmic viewpoint d3pc. A linked viewpoint $\tau_1 \otimes \tau_2$ represents events as pairs of values from its constituent viewpoints τ_1 and τ_2. Each new linked viewpoint is used to represent the template piece independently; using the four melodic viewpoints of Table 1 we get four different linked viewpoints: pitch ⊗ d3pc, intpc ⊗ d3pc, 3pc ⊗ d3pc, and 5pc ⊗ d3pc. An example representation of one of these (5pc ⊗ d3pc) can be seen in Figure 2. To establish the semiotic structure, pattern discovery is performed on the template piece for each linked viewpoint independently.

2.2 Patterns and semiotic structure

To construct a semiotic structure of a template piece it is necessary to identify interesting repeated patterns which provide a dense covering of the template piece. Patterns are defined as sequences of event features described using viewpoints, and an event sequence instantiates a pattern if the components of the pattern are instantiated by successive events in the sequence. More precisely, a pattern of length m is a structure $\tau{:}(v_1, \ldots, v_m)$, where τ is a viewpoint and the v_i are elements of the codomain of τ. For example, in Figure 2 two simple patterns, each instantiated twice, are highlighted; 3pc:(u, d) and d3pc:(d, eq, u, eq).

Patterns in a template piece can be found by applying a sequential pattern discovery method [2, 8] to each viewpoint representation of the template piece, identifying all patterns occurring more than once. This resulting list is then

[3] http://bdb.bertsozale.eus/en/web/doinutegia/view/2627-abiatu-da-bere-bidean

sorted according to an interestingness measure of patterns, and the ones that will form the coherence structure are chosen using a covering algorithm. These steps are now described in the remainder of this section.

Pattern distinctiveness and ranking Pattern interestingness is very important: in a given piece many patterns may exist but not all patterns are statistically or perceptually significant to a listener. For example, the 3pc pattern shown in Figure 2 would likely be instantiated many times in any template piece, but its occurrences (simply three notes with an up-down contour motion) are probably not structurally related or distinctive to the template piece, while the d3pc pattern is more interesting. In order to build a good semiotic structure of the template piece, distinctive and interesting repetitions can be identified using a statistical method which provides the probability of seeing an indicated pattern at least the observed number of times in a template piece. Then a pattern is interesting if it occurs more frequently than expected. This is a standard model for assessing discovered motifs in music informatics [11] and bioinformatics [21].

More precisely, we derive a function $\mathbb{I}$ measuring the interest of a pattern. First, we note that the background probability p of finding a pattern $P = \tau: (v_1, \ldots, v_m)$ in a segment of exactly m events can be computed using a zero-order model of the corpus:

$$p = \prod_{i=1}^{m} \frac{c(v_i)}{c},$$

where $c(v_i)$ is the total count of the feature $\tau : v_i$ and c is the total number of places in the corpus where the viewpoint τ is defined. Then the binomial distribution $\mathbb{B}\big(k; n, p\big)$ gives the probability of finding the pattern exactly k times in n events, and therefore the negative log probability of finding k or more occurrences of the pattern in a template piece with ℓ events is

$$\mathbb{I}(P) = -\ln \mathbb{B}_{\geq}\big(k; n, p\big), \tag{1}$$

where $\mathbb{B}_{\geq}$ is the upper tail of the binomial distribution, with $n = \ell - m + 1$ being the maximum number of positions where the pattern could possibly occur in the template piece.

Template covering Following pattern discovery, the template piece is covered, trying to use the most interesting patterns but also striving for a dense covering. Though finding a covering jointly optimal in those requirements is intractable, a greedy method can be used to rapidly find a reasonable semiotic structure. In the greedy covering method, discovered patterns are sorted from most to least interesting using Equation 1, then this sorted list is processed to choose the patterns that fit into the positions of the template piece that have not been yet covered by any pattern, not allowing overlapping between contiguous patterns.

Example In Figure 3 the pattern structure of the template *Erletxoak lorean*[4] after the covering process is shown, patterns represented by the viewpoints pitch ⊗ d3pc and 3pc ⊗ d3pc. Above each pattern is the the viewpoint name, the pattern label, and the $\mathbb{I}$ value in brackets.

The template is a short piece with four phrases, having two sections in an overall $ABA'B$ structure. The music is syllabic with each phrase having 13 notes, in the key of Gm, briefly visiting B♭M in the third phrase (established at the high F♮). The B phrase is perfectly captured by a discovered pitch pattern, and though a few notes at the beginning of A and A' have not been covered by patterns, the discovered three-point contour pattern successfully captures the similarity between the second and fourth phrases. Note that there is no rigid transposition that relates these two phrases, but they have similar melodic contours that are captured by viewpoint patterns.

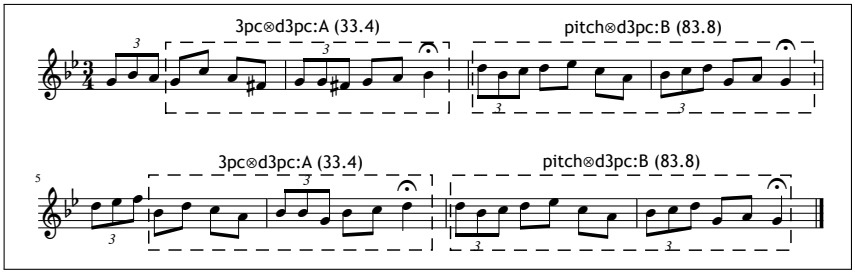

Fig. 3. Schema of a possible semiotic structure for the template piece *Erletxoak lorean*.

2.3 Statistical model

The semiotic structure defines the coherence within the template piece that will be conserved. To generate into the structure, stylistically coherent surface material is generated using a statistical model of the bertso corpus. In this work a trigram statistical model is built from a corpus to generate musical material into a template described by a semiotic structure. The exact probability of a piece using a trigram viewpoint model can be computed as described in [9]. Letting $v_i = \tau(e_i|e_{i-1})$ be the viewpoint τ value of event e_i in the context of its preceding event e_{i-1}, the probability of a piece $\mathbf{e} = e_1, \ldots, e_\ell$ is computed as:

$$\mathbb{P}(\mathbf{e}) = \prod_{i=3}^{\ell} \mathbb{P}(v_i|v_{i-1}, v_{i-2}) \times \mathbb{P}(e_i|v_i, e_{i-1}). \tag{2}$$

To elaborate, the product of all features in the sequence according to a trigram model is represented by the first term. Trigram probabilities of the viewpoint τ

[4] http://bdb.bertsozale.eus/en/web/doinutegia/view/241-erletxoak-lorean-orain-kantatuko-det-ii

are computed from the entire corpus. The second term is the probability of the particular event given the feature, defined as a uniform distribution over events having the property v_i:

$$\mathbb{P}(e_i|v_i, e_{i-1}) = |\{x \in \xi : \tau(x|e_{i-1}) = v_i\}|^{-1},$$

where ξ is the set of possible pitches (see Table 1).

The model above can be applied for any viewpoint. To select a viewpoint for modelling stylistic aspects of the bertso corpus in this study, every melodic viewpoint presented in Section 2.1 was evaluated with leave-one-out cross validation. Probabilities of every piece, according to Equation 2, were computed. Applied to the entire corpus of 2379 melodies, the product of all these probabilities gives a measure of the fit of the model to the corpus. The negative base-2 logarithm of this product is called the *cross-entropy* and lower cross-entropies are preferred. Every melodic viewpoint was tested, as were two linked melodic viewpoints intpc $\otimes$ 5pc and intpc $\otimes$ 3pc. The results of this procedure are shown in Table 2, which shows that the interval viewpoint int has the lowest cross-entropy on the corpus and is a good viewpoint to use for generation.

Viewpoint	Trigram Model
3pc	4.45
pitch	2.62
int	2.55
intpc	3.83
5pc	3.38
intpc $\otimes$ 5pc	2.71
intpc $\otimes$ 3pc	3.13

Table 2. Cross-entropy of different viewpoints, determined by leave-one-out cross validation on the corpus.

2.4 Generation

In this work a semiotic structure is used along with the trigram statistical model to generate new melodies. Generated sequences having high probability are assumed to retain more aspects of the music style under consideration than sequences with low probability. The process of *optimization* is concerned with drawing high probability sequences from statistical models.

A stochastic hill climbing optimization method is used to obtain high probability melodies. The method starts with a random piece that respects the coherence structure extracted from the template piece, using pitches from a pitch set ξ' that defines the admissible pitches for the generated piece. This set is typically the scale defined by the desired tonality of the generated piece and will

be a subset of the complete pitch domain ξ. This initial piece is created with a left-to-right random walk, which samples a new note in every position of the template, and every time a complete pattern is instantiated, all of the future locations of the pattern are also instantiated, in this way conserving the original relation between them. The piece is then iteratively modified: in each iteration of the process a random location i in the current piece $\mathbf{e}$ is chosen. A pitch e_i is uniformly chosen from ξ' and is substituted into that position, producing a new piece $\mathbf{e}'$ with an updated probability $\mathbb{P}(\mathbf{e}')$. If $\mathbb{P}(\mathbf{e}') > \mathbb{P}(\mathbf{e})$, then $\mathbf{e}'$ is taken as the new current piece. Every time a position is changed, the pattern to which that note belongs is identified, and all other instances of that pattern are also updated. Thus at every iteration the generated piece conserves the semiotic structure. The optimization process is iterated up to 10^4 times, and after each update the probability of the new piece is computed using Equation 2. If the new probability is higher than the last saved one the change is retained.

3 Results

To illustrate the generality of the method, new melodies are generated using two different templates, and properties of generated melodies are discussed. For the second template, two songs were peformed and evaluated by an audience in a live concert setting in a jazz club in London.

3.1 Illustration on a full piece

The template used is *Erletxoak lorean*, which was discussed earlier in Figure 3. The pitch vocabulary used is $\xi' = \{66, 67, 69, 70, 72, 74, 75, 77\}$ and two different viewpoints were used for the statistical model (Equation 2): 5pc and int. The three transformations shown in Figure 4 conserve the semiotic structure shown in Figure 3. The first transformation contains within the B phrase a leap down by a diminished seventh, which though perhaps difficult to sing is interesting and is resolved properly by a step up. The A and A' phrases are somewhat reserved in their ambitus, though A contains an interesting ascending broken triad. The second transformation follows an overall smooth melodic contour and is a singable melody with internal coherence. Its shortcoming might be identified within the A' phrase which has a non-idiomatic leap which further exposes an F♯ and F♮ together in close proximity. This could be corrected by including another segmental viewpoint to ensure that the scale of each phrase is internally coherent. The final transformation of Figure 4, done with the int viewpoint as the statistical model, corrects to some extent the problems with excessive leaps with the general 5pc model, but is confined to a rather small ambitus.

3.2 London concert and listener evaluation

A small suite of pieces was performed live in a public concert named “Meet the Computer Composer” at the Vortex Jazz Club in London on September 28,

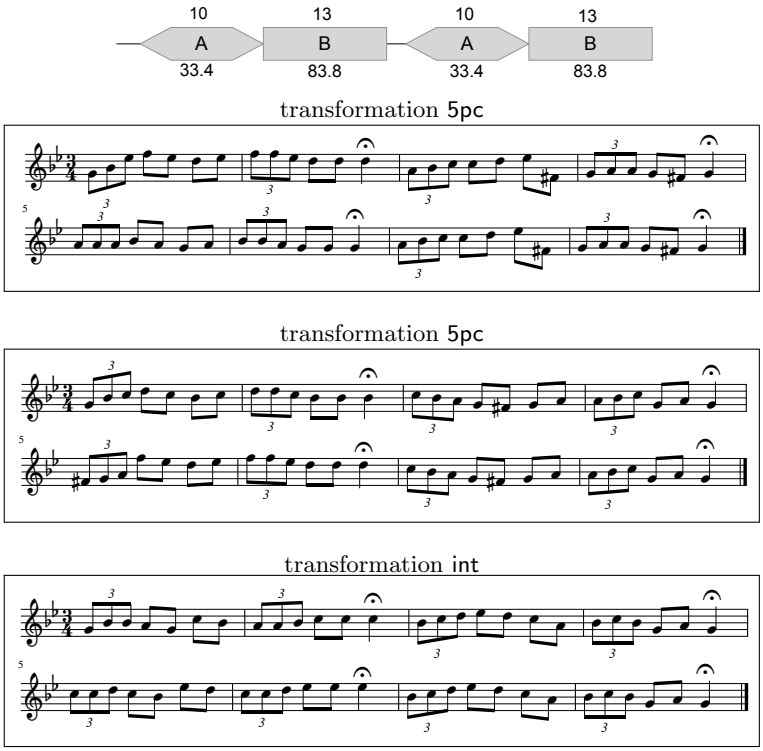

Fig. 4. Three transformations of the template piece *Erletxoak lorean*. Top: its semiotic structure with the number of notes in each pattern and their $\mathbb{I}$ value. The first two transformations use a 5pc statistical model and the bottom one uses a int model.

2016. A bertso melody *Txoriak eta txoriburuak* was sung (by the first author IG) along with two generations that used the original as a template. The full scores of all three melodies can be seen in Figure 5. Following the bertso tradition of new lyrics to existing melodies, the three melodies were sung each with the same new lyrics that were specially written for the concert.

An audience questionnaire (Table 3, top) was given at the beginning of the concert to all the members of the audience, where they would note which one of the three melodies they thought was the original, and how confident they were in their decision. A total of 52 questionnaires (from approximately 100 distributed) was returned. In Table 3 the results obtained from the questionnaires can be seen. The majority (55%) of respondents incorrectly identifed one of the two transformations as the original piece, though the 44% identifying correctly the original had overall higher confidence in their decision. Regarding transformation 1, it must be noted that this was the first of three pieces performed, and the singer had not yet achieved perfect intonation: this no doubt affected the

lower (15%, with 37.5% not confident in their response) audience result for that transformation.

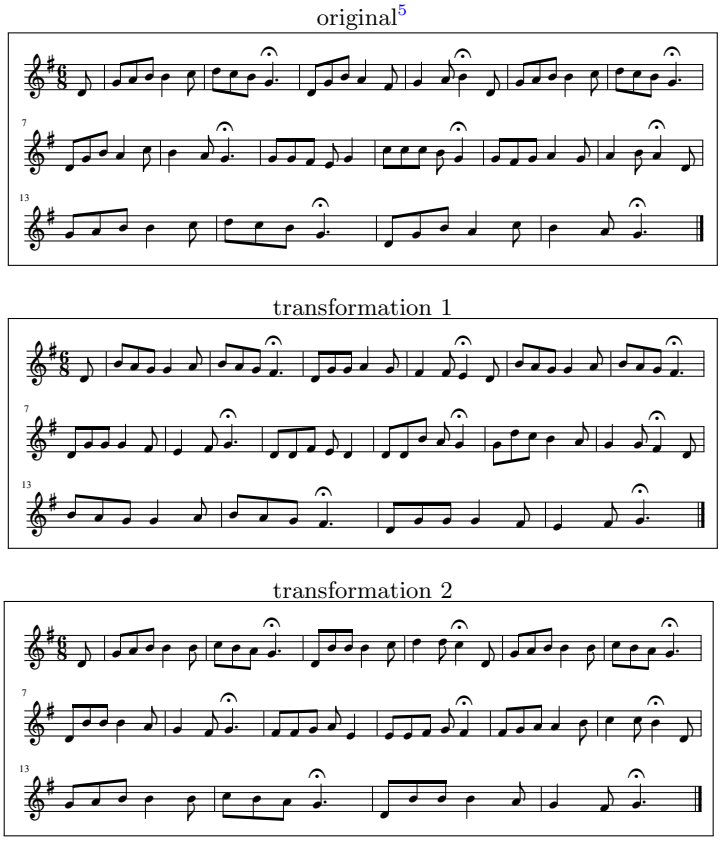

Fig. 5. Three pieces performed at the London concert.

4 Conclusions and future work

In this paper a method for transforming bertso melodies conserving the internal coherence of a template piece is presented. The basis of the method is a trigram statistical model combined with the strong constraints provided by a semiotic structure, which is identified using a sequential pattern discovery algorithm followed by a pattern ranking and covering method. New musical content is created

[5] http://bdb.bertsozale.eus/en/web/doinutegia/view/1564-txoriak-eta-txoriburuak

Which piece is the original?
☐1 ☐2 ☐3

How confident are you on a scale of 1 to 5? (1=not confident, 5=very confident)
☐1 ☐2 ☐3 ☐4 ☐5

	transformation 1					transformation 2					original				
is original?	8 (15%)					21 (40%)					23 (44%)				
confidence	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
%	37.5	50	0	12.5	0	38	28.6	23.8	4.8	4.8	26	23.8	23.8	23.8	8.7

Table 3. Top: the audience questionnaire distributed at the London concert. Bottom: results obtained. The original piece was the third melody sung.

using a statistical model which iteratively changes a template piece to improve the final result.

The generation method presented in this paper extends the method of Collins et al. [6] in some important ways. Not restricted to patterns conserving exact intervals, the method here allows a heterogeneous semiotic structure comprising a variety of abstract viewpoints. The generated pieces are not single random walks from a model, rather some effort is made to generate high probability solutions which are expected to be more stylistically valid. The method can be extended to polyphony and some initial work in those directions has been completed for counterpoint generation in the style of Palestrina [19] and multilayer textures in electronic dance music [10].

Acknowledgments This research was supported by the project Lrn2Cre8 (2013-2016) which was funded by the Future and Emerging Technologies (FET) programme within the Seventh Framework Programe for Research of the European Commission, under FET grant number 610859. The authors thank the Xenpelar Dokumentazio Zentroa for their enthusiasm in the project and for sharing the Bertso Doinutegia. Thanks to Kerstin Neubarth for valuable discussions on the research and the manuscript.

References

1. M. Allan and C. K. I. Williams. Harmonising chorales by probabilistic inference. In *Advances in Neural Information Processing Systems*, pages 25–32, 2004.
2. J. Ayres, J. Flannick, J. Gehrke, and T. Yiu. Sequential PAttern Mining Using a Bitmap Representation. In *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '02, pages 429–435, Edmonton, Alberta, Canada, 2002.

3. F. Bimbot, E. Deruty, G. Sargent, and E. Vincent. Semiotic structure labeling of music pieces: Concepts, methods and annotation conventions. In *13th International Society for Music Information Retrieval Conference (ISMIR)*, pages 235–240, Porto, Portugal, 2012.
4. N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent. Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription. In *International Conference on Machine Learning*, Edinburgh, Scotland, 2012.
5. 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.
6. T. Collins, R. Laney, A. Willis, and P. H. Garthwaite. Developing and evaluating computational models of musical style. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 30:16–43, 2016.
7. D. Conklin. Music generation from statistical models. In *Proceedings of the AISB 2003 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences*, pages 30–35, Aberystwyth, Wales, 2003.
8. D. Conklin. Discovery of distinctive patterns in music. *Intelligent Data Analysis*, 14(5):547–554, 2010.
9. D. Conklin. Chord sequence generation with semiotic patterns. *Journal of Mathematics and Music*, 10(2):92–106, 2016.
10. D. Conklin and L. Bigo. Trance generation by transformation. In *8th International Workshop on Music and Machine Learning (MML 2015) at the 21st International Symposium on Electronic Art*, pages 4–6, Vancouver, 2015.
11. D. Conklin and S. Weisser. Pattern and antipattern discovery in Ethiopian bagana songs. In D. Meredith, editor, *Computational Music Analysis*, pages 425–443. Springer, 2016.
12. D. Conklin and I. H. Witten. Multiple viewpoint systems for music prediction. *Journal of New Music Research*, 24:51–73, 1995.
13. D. Cope. An expert system for computer–assisted composition. *Computer Music Journal*, 11(4):30–46, 1987.
14. J. Dorronsoro. *Bertso Doinutegia.* Euskal Herriko Bertsolari Elkartea, 1995.
15. S. Dubnov, G. Assayag, O. Lartillot, and G. Bejerano. Using machine-learning methods for musical style modeling. *Computer*, 36(10):73–80, 2003.
16. S. A. Hedges. Dice music in the eighteenth century. *Music & Letters*, 59(2):180–187, 1978.
17. J. Leach and J. Fitch. Nature, music, and algorithmic composition. *Computer Music Journal*, 19(2):23–33, 1995.
18. L. B. Meyer. Meaning in music and information theory. *Journal of Aesthetics and Art Criticism*, 15:412–424, 1957.
19. V. Padilla and D. Conklin. Statistical generation of two-voice florid counterpoint. In *Proceedings of the Sound and Music Computing Conference (SMC 2016)*, pages 380–387, Hamburg, Germany, 2016.
20. D. Quick and P. Hudak. Grammar-based automated music composition in Haskell. In *Proceedings of the 1st ACM SIGPLAN Workshop on Functional Art, Music, Modeling & Design*, FARM '13, pages 59–70, 2013.
21. J. van Helden, B. André, and J. Collado-Vides. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *Journal of Molecular Biology*, 281(5):827–842, 1998.