Received November 22, 2019, accepted December 6, 2019, date of publication December 13, 2019,
date of current version December 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2959696

Statistics-Based Music Generation Approach Considering Both Rhythm and Melody Coherence

IZARO GOIENETXEA[1], IÑIGO MENDIALDUA[2], IGOR RODRÍGUEZ[1], AND BASILIO SIERRA[1]

[1]Department of Computer Science and Artificial Intelligence, University of the Basque Country (UPV/EHU), 20018 Donostia-San Sebastián, Spain
[2]Department of Computer Languages and Systems, University of the Basque Country (UPV/EHU), 20018 Donostia-San Sebastián, Spain

Corresponding author: Izaro Goienetxea (izaro.goienetxea@ehu.eus)

This work was supported in part by the Basque Government Research Teams under Grant IT900-16, in part by the Spanish Ministry of Economy and Competitiveness under Grant RTI2018-093337-B-I00, and in part by the Provincial Council of Gipuzkoa under Grant DGE19/04.

ABSTRACT This paper presents a music generation method which is an extension of a previously presented method that generates coherent melodies using a melodic coherence structure extracted from a template piece. This extension, which has been applied for generating bertso melodies, adds the generation of the rhythmic content of the melodies, for which a rhythmic coherence structure of the template piece is also created. To do so, a pattern discovery and ranking method is used to discover the rhythmically repeated segments that are interesting, and create a rhythmic coherence structure which can have several levels of nesting. Independent sampling processes have been developed for melodic and rhythmic content, using an adapted optimization method for sampling the rhythmic content of the new pieces. An evaluation process has been carried out to evaluate some of the generated pieces, considering on one hand how the listeners perceive them and on the other hand whether they share the features with bertso melodies. It has been concluded from this evaluation that the method is capable of generating good coherent bertso melodies.

INDEX TERMS Coherence, computer generated music, rhythm generation, statistical models.

I. INTRODUCTION

Automatic generation of music has a long history of research since the creation of the first computer in 1840 by Lovelace and Babbage [1], but the idea of composing music automatically has existed even before the existence of computers. Some examples of this idea are the *Musikalisches Würfelspiel* or musical dice games, like the one published in 1792 that was attributed to Mozart [2].

The earliest automatically generated compositions are from the mid-1950s, around the same time as the concept of Artificial Intelligence was coined. Among the first automatically generated compositions are those of Lejaren Hiller and Leonard Isaacson from 1955-56 [3]. Since these first steps many different algorithms have been developed to compose music automatically, such as knowledge based systems, evolutionary and other population-based methods, fractals or statistical models [4]–[6].

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan.

Statistical models of symbolic music have been used in computational modelling of several musical styles, for which many computational approaches have been developed [7]–[10]. The main advantage they offer is that they can be learned from a corpus of music to extract its musical features. These features can be then used to generate new musical sequences that reflect an explicit musical style [11]–[13].

An important issue that needs to be taken into consideration when generating music automatically is the coherence of the generated pieces. New pieces should contain material that is related (by repetition or a more abstract relation) to segments seen earlier in the piece, in order to endow it with some musical meaning. Different theories have been developed on how the music should be structured in order to be comprehensible. Arnold Schoenberg [14] believed that laws are needed to write music; acoustic laws and laws that result from the combination of time and sound. According to him listeners have to recognize musical figures and how they cohere in order to comprehend what they are listening.

Some theories compare musical discourse and linguistics, as well as the mechanisms the human brain has to understand them [15], [16]. These works suggest that, as in linguistics, relations between different segments in musical pieces are necessary to build a coherent discourse. The most obvious relation between musical segments is repetition. It is a fact that almost all forms of music involve repetition [17], either of sequences of pitch of notes or at some higher level of structural grouping, and that repetition imparts a sense of meaning to music [18]. These repeated segments are named *motives*, where a motif is defined as ''the smallest part of a piece or a section of a piece that, despite change and variation, is recognizable as present throughout'' [19]. Though early knowledge-based methods [20] explicitly considered repetition, the problem of achieving coherence in music generated from machine learning models remains largely unsolved. Some approaches have been developed to deal with the coherence of the generated music, like the description of its acoustic structure, functional structure or semiotic structure. Semiotic structure is defined as the representation of similar segments by similar arbitrary symbols [21]. Once the semiotic structure of a piece is described, the process can be ''inverted'', to generate new music by instantiating the symbols of the structure, getting pieces with new music material but the same coherence structure of the original piece.

That is the generation idea followed in [22], which uses a coherence structure that describes the melodic relations of a template piece to generate new pieces along with a statistical model created from a corpus of bertso melodies. Bertsos are Basque improvised songs, which must respect various melodic and rhyming patterns, and their rhythmic structure has to fit in one of the many accepted metrics. They are defined as sung, rhymed and metered discourses by the book *The Art of Bertsolaritza: Improvised Basque Verse Singing* [23]. There is evidence of bertso singing and written bertso poem samples since the 15th century, and it is a very popular art nowadays in the Basque Country. Bertsos are sung in many different occasions, like informal lunches with friends, homage ceremonies or competitions, and any topic can occur in a bertso. Many bertsolarism competitions take place every year in the Basque Country, and every four years the national championship final is held, with around 15000 people in attendance.

Experts say that the chosen melody for singing a bertso and the manner in which it is sung can be the key for the communicative success of the bertsolari, since the chosen melody must be able to combine with the created lyrics to transmit what the bertsolari wants to express with the bertso. These melodies can be traditional folk melodies, new melodies that have an appropriate rhythmic structure 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 work the music generation method presented in [22] is extended. In this extension in addition to the melodic content the rhythm of the new pieces is also sampled. To do so the same idea of using a template piece is followed, and its rhythmic coherence structure is described to be then used along with a statistical model of rhythmic information of the corpus to generate new complete pieces.

The rest of the paper is organized as follows. Section II gives an overview of the related work of the field of automatic music generation, Section III describes the corpus used in this work and Section IV gives a complete description of the presented generation method. In Section V some of the obtained results are shown and in Section VI the evaluation process that has been followed to evaluate some of the generated pieces is described. Finally, in Section VII the extracted conclusions and the identified future work are presented.

II. RELATED WORK

Several approaches have been developed in automatic music generation that, even though a fixed taxonomy of this kind of methods does not exist, are often classified as knowledge-based (or rule-based), evolutionary methods, machine learning methods or hybrids.

Knowledge-based methods use pre-made sets of arguments or rules that describe a style or genre, to compose music on the same style or genre. Some examples of this type of generation are the grammar models and the rule learning methods. Grammar models produce musical pieces using rules, which expand high level symbols into detailed sequences of symbols (words). These rules can be hand coded by an expert or they can be learned from a corpus of melodies that share a genre or style. An example of the use of grammars for music generation is the method developed by Chemillier [24], which generates jazz chord sequences based on Steedman's grammar. This grammar was created from a set of modern jazz 12-bar chord sequences, which is considered a wide and representative range of permissible variations of the blues basic form.

Evolutionary methods are based on the improvement of a population by cycles of evaluation and reproduction with variation of its individuals. The process starts with the generation of the candidate solutions of the initial set, then in each cycle the candidates are changed by mutation or recombination and they are evaluated using a fitness function. These cycles are repeated until a stopping criteria is satisfied.

Evolutionary algorithms have been used in different tasks of music generation like in *GenJam* [25], an interactive jazz improvisation system. GenJam uses a training process, in which the system plays a tune and a human mentor evaluates it as good or bad. These evaluations are then used to adjust the fitness function. Another example of the use of evolutionary algorithms for music composition is *MetaCompose* [26], which is a component-based system for music generation that supports real-time improvisation. The composition process has three main steps: creation of a chord sequence, evolution of a melody fitting this chord sequence and creation of an accompaniment for the melody/chord sequence combination. Jeong *et al.* [27] propose a multi-objective evolutionary approach to automatic melody composition. They produce a variety of melodies at once using genetic algorithms, and apply two fitness measures (stability and tension) to evaluate the created melodies. Muñoz *et al.* [28] proposed an adaptive multi-agent memetic approach that takes a bass line as input and creates a four voice piece.

Machine learning methods extract the knowledge from a corpus instead of having it previously defined. Statistical models are an example of machine learning methods, where different features of a corpus can be represented in a model that will be able to assign probabilities to automatically generated melodies.

Statistical models of music have been used for generating melodies and harmonies in several works, and they go from the earliest Markov models [29] to new models based on deep learning [30], [31]. Whorley and Conklin [32] use statistical models to generate four-part harmonizations using horizontal and vertical viewpoints of music and an iterative random walk sampling. Herremans *et al.* [33] use a first order statistical model to capture the melodic and harmonic features of a first species counterpoint corpus and generate new musical content. To do so, they use a sampling method named Variable Neighborhood Search (VNS), which starts with a randomly generated fragment and optimize it making local changes to increase the probability of the fragment, and they compare it to other sampling methods like random walk or Gibbs sampling. Padilla and Conklin [34] have developed a method to compose Palestrina masses using a combination of statistical models, to capture the stylistic aspects of the music, and pattern discovery to extract the coherence structure of original melodies. The pattern discovery process is performed on a single viewpoint representation of the pieces, and the patterns used to build the coherence structure are used to guide the generation of new musical material. Collins *et al.* [35] consider the coherence problem for generating new melodies by defining a template from an existing polyphonic piece to sample new notes onto it. Geometric patterns are discovered in a point representation of the notes in a pitch-time space, for which a pattern discovery method named SIACT [36] is used. This method is able to discover exact repetitions and transposed segments.

Roig *et al.* [37] proposed a music generation method that generates new melodies in a certain style. The method is based on extracting rhythmic patterns from musical pieces of the same style, patterns having a length of a measure, and using probabilistic models of these patterns to model different styles. They also give the user the choice to specify the rhythmic and harmonic structures of the final pieces, which are used to sample contours that respect the defined harmonic proprieties. Same authors [38] also presented a method that generates harmonic sequences using probabilistic models of progressions.

Deep learning architectures are used more and more in music generation, and well known groups like Magenta[1] at Google are using them to generate new music. An example of their work is the Bach doodle [39], which is able to harmonize in the style of Bach the melodies that users create manually. Deep learning is defined as a repertoire of machine learning techniques based on artificial neural networks which have multiple layers to process multiple abstraction layers of the data [40], and several approaches to automatically create music using these techniques have been developed. Some works [41] use a unit selection methodology to analyse if using only the units available in a library can be enough to generate a wide spectrum of new musical content. They then use a combination of a Deep Structured Semantic Model (DSSM) and an Long Short-Term Memory (LSTM) to predict the next unit in the generation model. Other generation works are based on the use of GAN (Generative Adversarial Network) to create music, like MidiNet [42] and MuseGAN [43]. Hadjeres et al. developed DeepBach [44], which is capable of generating chorales in the style of Bach using four neural networks.

Even though this is a growing area of research and interesting results are obtained using these architectures for music generation, they still have some limitations, like the control (of tonality conformance, rhythm...), structure (giving direction to the generated music), creativity (versus imitation) and interactivity [40]. Trying to solve the structure limitation, Medeot *et al.* [45] proposed StructureNet, which is a neural network that is trained with structure definition along with a probabilistic model of events.

III. CORPUS

The corpus used in this work is the *Bertso doinutegia*, a collection created by Joanito Dorronsoro and published for the first time on 1995 [46]. It is maintained and updated every year by *Xenpelar Dokumentazio Zentroa*[2] with new melodies that are used in competitions and exhibitions. Entries in the collection have a melody name, the name or type of the stanza and type of the melody (genre), among other information. Melodies are classified into 17 different types or genres, and 381 of the melodies in the collection have links to recordings of exhibitions or competitions where those melodies were used.

The scores included in the collection have been encoded in Finale and exported to MIDI. Currently the collection is composed of 2382 bertso melodies, which have a mean length of 60 notes. 85 of the melodies are polyphonic or have polyphonic parts. Since in this work monophonic pieces are generated, all the pieces with polyphony are processed using a *skyline* method, which takes the event with highest pitch at unique onset times.

Since in this work statistical models are built to capture different aspects of the corpus, it has been studied in order

[1] https://magenta.tensorflow.org/
[2] http://bdb.bertsozale.eus/es/

FIGURE 1. Score of the melody neska zaharrak eta apaizak II.

to detect anomalies within the pieces that could affect on the creation or use of the statistical models. It has been found that many pieces have long sequences of repeated notes. Sequences of four repeated notes have been discovered occurring at least twice in 744 pieces (31% of the corpus), and at least once in 1064 (44.7% of the corpus). A model built from these sequences would assign high probabilities to long note repetitions, and even though these sequences exist in the corpus of the style that is being replicated, it has been decided that the generation of such sequences should be avoided because of their lack of interest. Figure 1 shows the score of the melody with the highest number of sequences of repeated notes, which clearly is not musically interesting.

To avoid the negative effect that these pieces could have on the statistical model, the pieces with the highest proportion of repeated notes are removed from the corpus. To do so, the list of pieces has been sorted according the proportion of sequences of repeated notes they have, and a manual analysis has been made to discard those pieces that have a high proportion of repeated notes that make the pieces of little musical interest. After the analysis the size of the corpus has been reduced to 1934 pieces (82.2% of the original corpus).

IV. PROPOSED METHOD

The method presented in this paper is an extension of the work presented in [22], which is able to generate new melodic information maintaining the rhythm of an existing piece. In this approach, in addition to the melodic information, the rhythmic part is also generated. The generation process is based on the use of an abstract template that consists of a melodic coherence structure and a rhythmic coherence structure, both extracted from an existing piece, and two statistical models, a melodic model and a rhythmic one. The coherence structures ensure that the final pieces have related segments within them, while the statistical models are able to capture certain melodic and rhythmic features of a corpus that are then reflected in the generated material. In Figure 2 a diagram of the presented method can be seen. It can be seen that the melodic and the rhythmic information are treated independently, both for building the coherence structures and the statistical models and for generating new musical content. Each of the components of the figure are described in more detail below.

A. COHERENCE STRUCTURES

The aim of the coherence structure is to describe the relations between similar segments of a piece. In this work, since melodic and rhythmic relations are considered, the coherence of the template is described by two independent components; the melodic coherence structure and the rhythmic coherence structure. It has been decided to create independent structures to describe melodic and rhythmic coherence, because in bertso melodies many rhythm repetitions can occur where the melodic content is different between the repetitions. In Figure 3 an example is shown where a rhythmic pattern is highlighted. The pattern has a length of 19 components and is repeated four times through the piece, but there is no melodic relation between all the occurrences of the pattern. It has been considered that these cases make it necessary to have independent coherence structures for the melodic relations and the rhythmic ones.

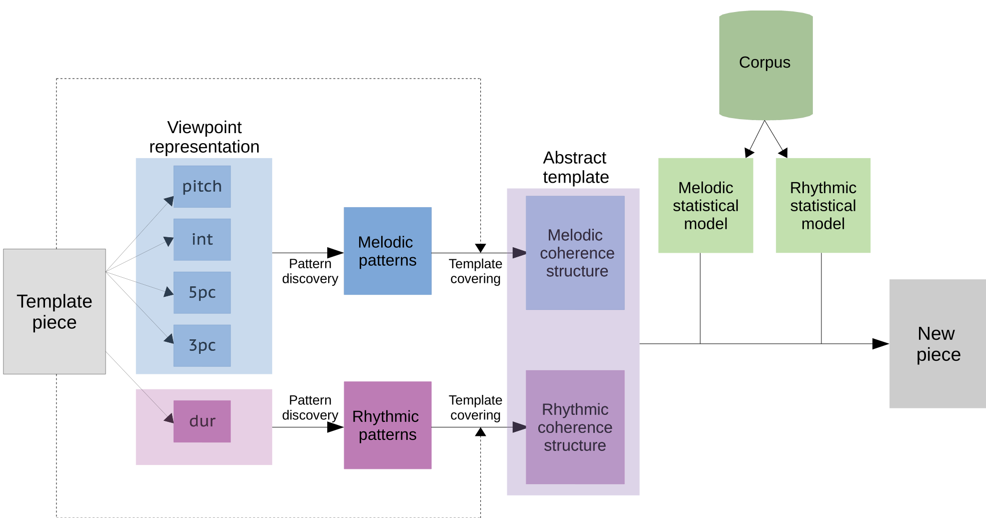


FIGURE 2. Diagram of the method proposed in this work.

FIGURE 3. Score of the melody ***Abiatu da bere bidean*** where the main rhythmic pattern is highlighted.

TABLE 1. A specification for a small set of viewpoints. Top: Two basic attributes of notes. Bottom: derived viewpoints.

viewpoint	codomain
pitch	$\{50, 52, 53, \ldots, 83\}$
dur	$\{1, 2, 3, \ldots\}$
int	$\{-14, -12, -11, \ldots, 14, 15, 17\}$
3pc	{d, eq, u}
5pc	{ld, sd, eq, su, lu}
d3pc	{d, eq, u}
durInt	$\{1/16, 1/8, 1/4 \ldots\}$

The melodic coherence structure needs to capture the most relevant melodic similarity relations between segments in the piece, while the rhythmic coherence structure must describe the rhythmic repetitions between segments. Similarity relations of various abstraction level are considered to be included in the melodic coherence structure, in order to be able to capture not only the more obvious relations like repetition or transposition, but also the more abstract ones. Since in bertso melodies exact rhythmic repetitions are very common, only this type of relation is represented in the rhythmic coherence structure of a piece.

The process to build both of the coherence structures has three main steps: viewpoint representation, pattern discovery and pattern ranking and covering. They are explained below.

1) VIEWPOINT REPRESENTATION

In order to analyse the template piece on different abstraction levels a multiple viewpoint representation [47] is used. A *viewpoint* τ is a function that maps an event sequence $e_1, \ldots, e_\ell$ to a more abstract sequence $\tau(e_1), \ldots, \tau(e_\ell)$, comprising elements in the codomain of the function τ. In the building process of the melodic coherence structure notes are the only events that are taken into account.

In Table 1 some melodic viewpoints (pitch, int, intpc, 3pc and 5pc), and three rhythmic viewpoints (dur, intDur, and d3pc) are presented. The viewpoint pitch represents the MIDI number of each note; the viewpoint int computes the interval between a note and the preceding one; the viewpoint intpc computes the pitch class interval (interval modulo 12) between a note and the previous one. A three-point contour viewpoint 3pc computes the melodic contour between two notes: upward (u), downward (d) or equal (eq); and a five-point contour viewpoint 5pc computes whether the contour between two contiguous notes goes more than a scale step down (ld), goes one scale step down (sd), goes more than a scale step up (lu), goes one scale step up (su), or stays equal (eq). The duration viewpoint dur represents the duration of each note, while durInt represents the relation between the durations of two contiguous events. Contour viewpoint d3pc computes if the duration of a note is shorter (d) than the previous one, longer (u) or equal (eq). The representation of an example segment, using several viewpoints of Table 1, is shown in Figure 4.

a: MELODIC VIEWPOINTS

In this work melodic and rhythmic coherence are independently analysed, so they have been independently represented. The template pieces are represented using pitch, int, 3pc and 5pc melodic viewpoints, to be able to capture relations between segments in different abstraction levels.

b: RHYTHMIC VIEWPOINTS

Since in this approach of the generation method only exact rhythmic repetitions are considered, dur is the only viewpoint used to represent the template piece, which indicates the duration in ticks of each event in the piece.

2) PATTERN DISCOVERY

To construct the coherence structure of a template piece it is necessary to discover and join the interesting patterns that are repeated through every viewpoint representation of the piece and cover it in the most dense way possible. Patterns are defined as sequences of event features (or viewpoints), and a piece instantiates a pattern if the pattern occurs (one or multiple times) in the sequence: if the components of the pattern are instantiated by successive events in the sequence [48]. More precisely, a pattern of length m is a structure τ:$(v_1, \ldots, v_m)$, where τ is a viewpoint and the v_i are elements of the codomain of τ. For example, in Figure 4 a pattern of five elements in the dur representation is highlighted.

Pattern discovery methods identify segments that are repeated through a symbolic representation of a musical piece or a corpus. In this work a pattern discovery algorithm named SPAM [49] has been used to identify similar segments. This is an algorithm that finds all the frequent sequential patterns (patterns that occur more times than a given threshold) in a transactional database, specially efficient with large databases, but the method has been adapted to be used to discover patterns within a single sequence. Candidates are created with a depth-first search strategy and various pruning mechanisms are used to reduce the search space.

3) PATTERN RANKING AND COVERING

The pattern discovery algorithm produces all the patterns that appear more times than a given threshold and it does not have a more sophisticated method to rank the discovered patterns. One way to do this is by measuring the distinctiveness of each pattern [50], but in this work the priority is not to find patterns that are over-represented in a piece with respect to

pitch:	60	65	67	69	65	70	70	72	69	65
int:	⊥	5	2	2	-4	5	0	2	-3	-4
3pc:	⊥	u	u	u	d	u	eq	u	d	d
5pc:	⊥	lu	su	su	ld	su	eq	su	ld	ld
dur:	1024	512	512	1024	1024	1024	512	512	1024	1024
durInt:	⊥	1/2	1	2	1	1	1	1/2	2	1
d3pc:	⊥	d	eq	u	eq	eq	d	eq	u	eq

FIGURE 4. **A fragment from the melody *Abiatu da bere bidean* and its viewpoint representation. A pattern of length five is highlighted.**

an anticorpus. The priority is to find patterns that are significant in the piece, in terms of occurrence and length.

Measuring the interest of the discovered patterns is a very important task; many patterns can occur in a piece, but not all of them are important enough to be used on the building of the coherence structure. For example, the d3pc pattern shown in Figure 4 would likely be instantiated many times in any piece, but its occurrences (simply three notes with an up-down duration contour motion) are probably not structurally related or distinctive to the template piece. In order to build a good coherence structure of the template piece, distinctive and interesting repetitions should 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 considered interesting if it occurs more frequently than expected. This is a standard model for assessing discovered motifs in music informatics [51] and bioinformatics [52].

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,

$$\mathbb{B}(k; n; p) = \binom{n}{k} p^k (1-p)^{n-k}$$

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.

The interest of all patterns discovered within all the viewpoint representations is computed and, in order to discard patterns that are not important enough for the coherence structure, an interest threshold is set. This threshold is set at 9.5, as a result of an analysis of the discovered patterns performed by hand, comparing different patterns, their significance in the score they are discovered in, and their interest value. As an example, in Figure IV-A.3 a score of the melody *Argi emaile txit diztiratsu* is shown, which is part of the corpus and where a three element pattern is highlighted. It is the pattern pitch : (64, 65, 67) which has an interest value of 9.2 and, given its length and information it represents, is not considered distinctive enough. The interest threshold is set at 9.5 to avoid this kind of patterns in the coherence structures.

FIGURE 5. **Score of the melody *Argi emaile txit diztiratsu*.**

FIGURE 6. **Score of the melody *Lagundurikan danoi I*.**

In Figure IV-A.3 the score of the melody *Lagundurikan danoi I* is shown, where the melodic contour pattern 3pc : $(u, u, d, d, eq, d, eq, d)$ is highlighted. The interest value of this pattern is 9.55, it is near the established interest threshold, but it is considered a good enough pattern, taking into account the amount of information it represents, showing that the chosen threshold is acceptable.

Once all the patterns in the different viewpoint representations of the template are discovered and their interest values are computed, they are used to cover the piece, trying to use the most interesting patterns but also striving for a dense covering. Since finding a covering that optimally fulfils both requirements is not easy, 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.

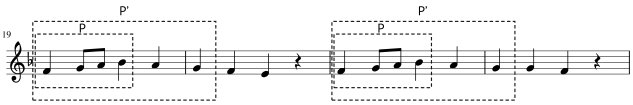


FIGURE 7. **Example of the effect of computing the interest value of the patterns with the actual number of occurrences used on the covering.** P **represents the pattern that would be covered using the number of occurrences of the template piece, and** P′ **is the actual pattern used.**

As two different coherence structures are built, two independent template coverings are performed; one considering all the patterns discovered in the different melodic representations of the piece and one only considering the rhythmic ones.

Every time a pattern is used on the template covering, the interest value of the patterns that remain in the sorted list must be recomputed. Their number of occurrences must be updated to consider only those that happen on the positions of the piece that are still uncovered. Once the interest value is recomputed for all the patterns in the list, it is resorted.

An example of the recomputing process is shown on Figure 7, where a stave of a melody of the corpus can be seen, and patterns P and P′ are highlighted. Pattern P occurs four times in a piece and has an interest value of 41.2. Pattern P′ is instantiated twice in the piece and has an interest value of 30.2, so P will have a higher position on the pattern list and would presumably be used on the covering of the segment shown on Figure 7. When the piece covering process begins, a higher interest pattern covers the positions where two of the occurrences of P happen (not shown in the figure), making its number of occurrences lower to two. The interest values of the patterns that remain in the list are recomputed and the interest of pattern P drops to 19.7, which is lower than the interest value of P′, that will be used on the covering of the segment.

The covering of the piece ends when there is no free sequences in the piece that can be covered by any pattern in the list.

a: MELODIC COHERENCE STRUCTURE

In the building of the melodic coherence structure the patterns discovered in the representations of the four melodic viewpoints presented in Table 1 are used. Even though the patterns are discovered in different representations, all of them are gathered in a single pattern list that is used in the covering of the template piece in order to create the melodic coherence structure.

b: RHYTHMIC COHERENCE STRUCTURE

To build the rhythmic coherence structure, patterns are discovered within the dur representation of the template piece, where significant patterns are intended to be identified. It has been noticed that in many of the bertso melodies the most significant rhythmic patterns are long patterns, which can have other significant patterns within them. These so called nested patterns are also important when building the coherence structure of a piece, and have been considered in other music generation works [35]. In Figure 8 an example of a long pattern can be seen which has shorter patterns within it.

FIGURE 8. **Score of the melody *Neure lagunak lagun zakidaz* where its most interesting pattern is highlighted.**

FIGURE 9. **Nested pattern (in purple) found within the principal pattern (in black) in the melody *Neure lagunak lagun zakidaz*.**

To discover this kind of nested patterns, once the most significant patterns are selected to cover the template piece, the sorted list of all discovered patterns is processed again, to look for patterns that happen, twice or more, within those significant ones. This process is repeated until all the nested patterns in the piece are discovered, making it possible to nest pattern within other nested patterns, creating a pattern hierarchy.

c: EXAMPLE

In Figure 9 the rhythmic structure of the piece *Neure lagunak lagun zakidaz* after the nested pattern discovery process is shown, where patterns of different nesting level are represented with different colours. It can be seen that the piece has one main rhythmic pattern that is repeated twice in the piece, and even if the whole piece is covered by this pattern, it does not provide enough information about how the rhythm is structured through the piece. Since the only information that this pattern represents is that the piece has a long segment that is repeated twice, the patterns that occur within that principal pattern are discovered and represented in the figure in purple.

The nested pattern discovery has also been applied in this purple pattern, in which another nested pattern of 7 elements has been discovered, represented in Figure 9 in green. In this example a pattern hierarchy of three levels is created.

B. STATISTICAL MODELS

Once the coherence structures of the template piece are built, the new notes that will be sampled within them need to be chosen, in order to get new pieces that are stylistically similar to the ones in the bertso corpus. To sample these notes statistical models of the corpus are used, which assign probabilities to sequences of events, where high probability sequences are assumed to retain more aspects of the music

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

Viewpoint	Unigram	Bigram	Trigram	Tetragram	Pentagram
3pc	4.63	4.53	4.45	4.35	4.26
pitch	3.72	2.87	2.62	2.68	3.05
int	$\boxed{3.17}$	$\boxed{2.74}$	$\boxed{2.55}$	$\boxed{2.56}$	$\boxed{2.81}$
intpc	5.19	4.06	3.83	3.7	3.7
5pc	3.65	3.47	3.38	3.29	3.22

style of the corpus than sequences with low probability. Since in this work a rhythmic generation process is presented in addition to the melodic generation, two different statistical models have been built; one to capture the melodic aspects of the corpus and one to capture the rhythmic ones. An *n-gram* model is used for both, for which a length and the feature it describes must be chosen, taking into account that different features are needed for each model.

1) MELODIC MODEL

In order to decide which melodic feature model fits the corpus best, models based on every melodic viewpoint presented in Section IV-A.1 have been evaluated with leave-one-out cross validation. The cross-entropy of each model has been computed, where lower cross entropies are preferred. The cross-entropy of a model is defined as the negative base-2 logarithm of the product of the probabilities of all the 1934 pieces of the corpus.

To compute the probabilities of the pieces Equation 2 is used, which depends on the length of the n-gram. 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:

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

To elaborate, the product of all features in the sequence according to a n-gram model is represented by the first term. N-gram probabilities of the viewpoint τ are computed from the reduced 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). Since this model can be applied for any viewpoint, it has been tried with different melodic viewpoints presented earlier in this work and different n-gram lengths, and the cross-entropies of the corpus represented with each viewpoint have been presented in Table 2. It can be seen that the interval viewpoint int has the lowest cross-entropy value in all the models, and how the value goes down when the length of the model increases, just until the trigram model is reached. After that the cross-entropy value starts going up. The lowest value indicates that a trigram int model fits the corpus best. Once the int trigram model is created Equation 3 is used to compute the probabilities of the generated event sequences,

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

since for this viewpoint $\mathbb{P}(e_i|v_i, e_{i-1}) = 1$.

2) RHYTHMIC MODEL

As in the melodic model, to capture the rhythmic features of the corpus a trigram model has also been chosen. The same rhythmic viewpoint dur used in the viewpoint representation has been chosen to be used of the building of the rhythmic statistical model. To compute the probability of a rhythmic sequence Equation 3 is used.

C. GENERATION

The created statistical models are used to sample new high probability sequences into the semiotic structures, in order to generate pieces that retain more aspects of the pieces in the corpus. To do so, as two independent coherence structures and statistical models are created for the generation of new pieces, their melodic and rhythmic information are generated independently. For both generations a stochastic hill climbing optimization process has been used, which starts with a random sequence and iteratively changes random positions to improve its probability according to a statistical model, always respecting the coherence structures.

The vocabularies for both generations are fixed to ensure that the pitch and the duration of the generated notes are within a fixed range. The melodic vocabulary ξ'_m is defined by choosing the pitches from the tonality of the template piece, limiting them to the range of it, and which defines the admissible pitches for the generated pieces. The rhythmic vocabulary (ξ'_r) that is initially used in the generation is defined by taking the duration of the notes of the template piece. When the vocabularies are not big enough to create diverse generations, some entries can be added by hand, to generate more diverse pieces. The generation process starts with the rhythmic generation, and once it is finished the melodic generation step begins.

1) RHYTHMIC GENERATION

When generating rhythmic sequences different constraints should be considered. As the generated pieces are bertso melodies, their note number and total duration should be conserved, in order to fit them into one of the accepted metrics and be able to use them to sing existing or new bertso lyrics. Taking that into account, when creating the initial random sequence for the stochastic hill climbing process, the rhythmic information that is contained within each pattern is randomized instead of sampling a random duration in each of its positions. Since rests act like phrase boundaries in many bertso melodies, their positions have been fixed, and they are never moved.

FIGURE 10. **A rhythmic pattern of the melody *Abiatu da bere bidean* before (up) and after (bottom) the randomization step.**

FIGURE 11. **Rhythmic pattern of Figure 10 after the optimization phase.**

When randomizing patterns that contain nested patterns within them, first the patterns that are nested deeper in the hierarchy are treated, and the process then goes up until the patterns that are in the top level of the hierarchy are randomized. Every time a pattern is randomized all its other occurrences need also to be sampled, in order to conserve the rhythmic coherence of the piece defined in Section IV-A. When a segment of notes is not covered by any pattern of the coherence structure, it is treated like a one occurrence pseudo-pattern.

An example of randomizing patterns is shown in Figure 10, where a segment of the melody *Abiatu da bere bidean* is shown before and after the randomization. The segment is covered by a pattern that has a nested pattern repeated three times, highlighted in green. In the top stave the segment before the randomization is shown, while in the bottom stave the same pattern is shown after the rhythmic randomization and the random melodic sampling. The occurrences of the nested pattern cover almost all the notes in the main pattern, except the last three quarter notes and the rest, and the randomization can not have any effect in the duration of these last notes.

In each iteration of the optimization step a random position j of the piece is chosen, and the pattern that the position belongs to is identified. When j is part of a pattern that is nested within a larger one, the nested one is taken and a random element $r \in \xi_r'$ is chosen to subtract its value to the duration of j. The value of r is added to the duration of a position z randomly selected within the same pattern in the same pattern level. Once a pattern is updated, all its other occurrences are also updated, and the probability of the new sequence is computed with Equation 3. If the new probability is higher than the last saved one, the change is conserved. In Figure 11 the rhythmic pattern shown in Figure 10 can be seen after the optimization process, where it can be seen that the coherence of the nested patterns is conserved.

In this step it is very important ensuring that no duration is set to 0 or negative values, so if this happened in an iteration of the optimization step, the change would not be valid. The rhythmic optimization process is run 10^6 times total.

2) MELODIC GENERATION

The initial random information of the melodic generation process is created with a left-to-right random walk, which must respect the coherence structure extracted from the template piece; a new note is sampled 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. 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 random element $e_i \in \xi_m'$ is substituted into that position, and the pattern to which that position belongs is identified, to also update all the other instances of the pattern, producing a new piece $\mathbf{e}'$. Thus the pieces generated at every iteration conserve the semiotic structure. The probability ($\mathbb{P}(\mathbf{e}')$) of the new piece is computed using Equation 3, and if it is higher than the last saved one, ($\mathbb{P}(\mathbf{e}') > \mathbb{P}(\mathbf{e})$), then the change is retained and piece $\mathbf{e}'$ is taken as the new current piece. This optimization process is iterated up to 10^4 times.

V. RESULTS

To illustrate the method, several melodies have been created using various pieces as template. Some examples of generated melodies using two different template pieces are shown below. The shown examples are high probability melodies according to the melodic and rhythmic models created from the corpus.

A. TEMPLATE ABIATU DA BERE BIDEAN

The first piece used as template piece is *Abiatu da bere bidean*. It is a piece with a length of 100 notes with a smooth melody and a pretty simple rhythm, and it has six phrases, all delimited by a rest.

The melodic structure that has been created from this template piece can be seen in Figure 12. Five patterns have been used in the covering of the piece, where four of them are pitch patterns and pattern E is an interval pattern. It can be seen that the second occurrence of pattern E has a rest in the middle, covering the end of a phrase and the beginning of the next one. This happens because no rhythmic information is used in the melodic coherence structure generation, and rests are not considered as events. Even if many melodic contour patterns have been identified in the piece, their interest value is not high enough to be included in the structure of this piece.

The rhythmic coherence structure of the template can be seen in Figure 13, where three rhythmic patterns are highlighted in black. A five element nested pattern has also been discovered within pattern A, which is repeated three times within each occurrence of A. All the events in the piece are covered by some rhythmic pattern except two rests that cannot be covered with any pattern, since the last rest in the score is not part of the viewpoint representation of the midi file.

In Figure 14 two pieces generated using this template can be seen. In both cases the generated melodies are smooth, with no big leaps. It can be seen that both melodies respect the melodic and rhythmic coherence structures extracted from the template piece.

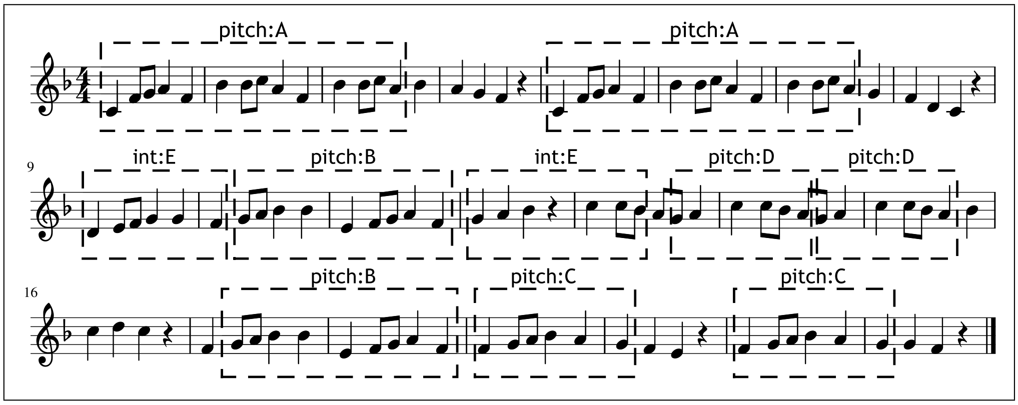


FIGURE 12. Score of the piece *Abiatu da bere bidean* with its melodic coherence structure highlighted. The type and label are shown above each pattern in the structure.

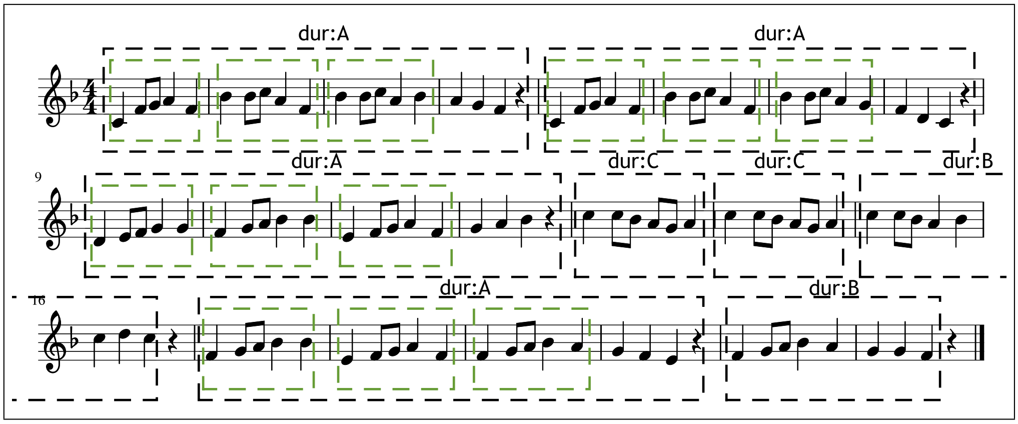


FIGURE 13. Score of the piece *Abiatu da bere bidean* with its rhythmic coherence structure highlighted. The label of each main pattern (in black) is shown, as well as the nested patterns (in green).

Their melodic information is quite different, but it is noticeable that the rhythm is similar in both, even if it is not the same. Since in the original piece only quarter notes and quavers are used, the rhythmic vocabulary for the generations is limited. Even though semiquavers have been added by hand to the rhythmic vocabulary in order to have more options in the generation, they are not used in the shown generated pieces.

In the top piece of Figure 14 it can be seen that the interval int:E pattern has been sampled as an exact repetition, instead of a transposition. This is allowed in this approach of the method, since repetition is considered a particular case of a transposition.

B. EGUNTTO BATEZ NINDAGUELARIK

The second template used to illustrate the method is *Eguntto batez nindaguelarik*. It is a piece with 69 notes with no big leaps but a more diverse rhythm than the template *Abiatu da bere bidean*, and it has four phrases.

In Figure 15 the score of the melody can be seen where the melodic coherence structure that has been built for this template is highlighted. Only a pitch pattern has been chosen in the covering strategy, since, even if other melodic patterns can be found that cover free spaces of the score, their interest is not high enough to be considered. Even though the melodic coherence structure is very simple in this template, it has been chosen because its rhythmic content is more interesting than the rhythm in template *Abiatu da bere bidean*.

The rhythmic coherence structure of the piece is shown in Figure 16, where four main patterns are highlighted. Duration pattern A has a nested pattern that happens twice in each occurrence of A. It can be seen that in this case not all of the notes are covered by a pattern in the rhythmic coherence structure, and as mentioned before, the segments that are not part of any pattern are treated like one occurrence patterns.

In Figure 17 the scores of two melodies generated with this template are shown. The first generated melody is pretty smooth with a nice rhythmic sequence which respects the rhythmic structure of Figure 16. The second generation, however, is an example of a melody that even if it has a high probability according to the statistical models (both melodic and rhythmic ones) has some rhythmic issues in the fourth and the 16th bars. The rhythm in these bars is not usual for bertso melodies and it would make them difficult to sing. Apart from these bars, both the rhythm and the melodic line in the piece are smooth and coherent. Since in this template more different measures are used compared to the first template, the generated pieces have more rhythmic differences between them.

FIGURE 14. Examples of pieces generated from template *Abiatu da bere bidean*.

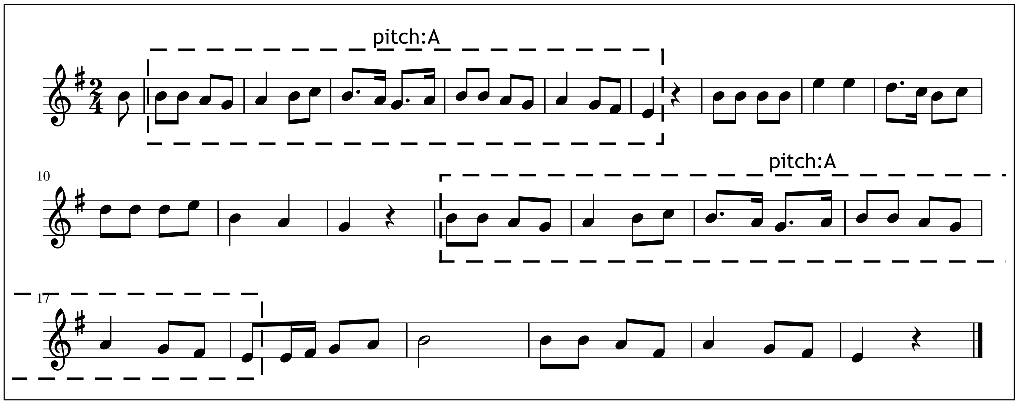


FIGURE 15. Score of the piece *Eguntto batez nindaguelarik* with its melodic coherence structure highlighted. The type and label are shown above each pattern in the structure.

C. DISCUSSION

The obtained results show that the method presented in this work can be used to generate new acceptable and coherent melodies. The rhythmic generation process that has been followed is conservative in order to generate pieces that maintain the original number of notes of each template, as well as its total duration, but in some cases it can be too conservative when the number of allowed rhythmic generations is low. This issue can be solved by allowing more duration values in the vocabulary of the piece or by using a model based on duration intervals instead of concrete duration values.

It also has been noticed that even if the statistical models assure that the generations will have certain melodic and rhythmic features, this is not always enough to guarantee that the generations will be musically pleasant. Since bertso melodies have no defined stylistic rules, rule sets cannot be used to evaluate the generations, but maybe some abstract rules could be learned from the corpus that would improve the musical quality of the generations.

VI. EVALUATION

To evaluate the results of the method presented in this paper an evaluation process has been carried out, which has two phases, the listeners' evaluation and the style evaluation.

A. LISTENERS' EVALUATION

For this phase of the evaluation five pieces have been randomly selected from the corpus, and they have been used as

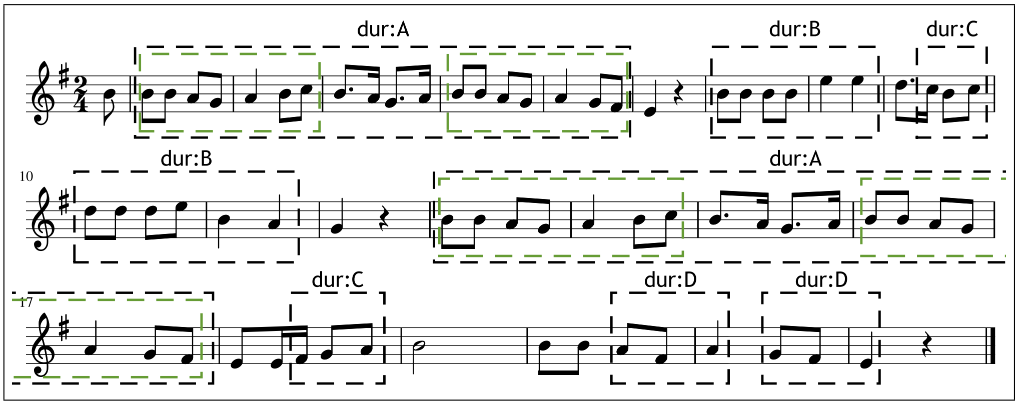


FIGURE 16. Score of the piece *Eguntto batez nindaguelarik* with its rhythmic coherence structure highlighted. The label of each main pattern (in black) is shown, as well as the nested patterns (in green).

FIGURE 17. Examples of pieces generated from template *Eguntto batez nindaguelarik.*

template to generate a new melody from each of them. Two sets of melodies were created from the ten initial pieces, making sure that a generated melody was never in the same set as the piece used as template to generate it. The scores in each melody set can be seen in Appendix VII. Each melody was sung using some lyrics that fitted the metrics of the melodies downloaded from the website of the centre of documentation Xenpelar[3] and recorded. 31 professors and researchers from the Faculty of Informatics of the University of the Basque Country participated in the evaluation process, where one of the two sets of melodies was assigned to each participant.

[3]https://bdb.bertsozale.eus/en/web/bertsoa/bilaketa

The first step of the evaluation process was a training phase, in which the participants listened to the MIDI files of the melodies in their set once. After this process the participants were given a questionnaire in which they needed to mark if they thought that the recordings they would listen were original bertso melodies or generatedones.

In Table 3 an excerpt of the questionnaire given to the participants can be seen, where for a melody eight options are given. The first seven options indicate if the participants think that the melody is an original piece or a generated one, and in which degree they are certain of that. The eighth option has been added to mark the melodies that the participants knew,

FIGURE 18. Scores of the pieces in set A.

since it is possible that some of the melodies used as template are well known for some participants.

They listened to each of the five recordings of their set once to fill the questionnaire, and after finishing, they were

FIGURE 19. Scores of the pieces in set B.

asked to give a score from 1 to 10 to each of the generated melodies from their set. This evaluation was subjective and the participants were not given any more information about what they needed to take into account to evaluate

TABLE 3. Excerpt of the questionnaire that was given in the evaluation. Each participant needed to evaluate five melodies.

1: original 100% 2: quite sure original 3: I think original 4: I do not know 5: I think generated 6: quite sure generated 7: 100% generated

MELODY N:
Original or generated? ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐ I know it

TABLE 4. Results obtained in the evaluation of sets A and B. The pieces generated with the presented method are shown in red, and the options that obtained more votes are presented in bold.

SET A:

Piece	1	2	3	4	5	6	7	I know it
Piece$_1$	2	5	3	1	4	0	0	1
Piece$_2$	0	2	2	**4**	**5**	2	0	0
Piece$_3$’	0	1	1	2	**3**	**6**	2	0
Piece$_4$	**12**	3	0	0	0	0	0	**9**
Piece$_5$’	0	0	2	**3**	**5**	**5**	0	0

SET B:

Piece	1	2	3	4	5	6	7	I know it
Piece$_1$’	0	1	0	0	**12**	**3**	0	0
Piece$_2$’	0	2	**4**	2	**6**	2	0	0
Piece$_3$	**3**	**4**	**7**	0	2	0	0	0
Piece$_4$’	0	**3**	**6**	0	**6**	1	0	0
Piece$_5$	3	**8**	**4**	0	1	0	0	1

Score of the generated pieces:

Piece	Mean score	Standard deviation
Piece$_1$’	6.06	1.06
Piece$_2$’	6.82	1.69
Piece$_3$’	6.47	1.55
Piece$_4$’	7.63	1.42
Piece$_5$’	6.33	2.12

the melodies. They just needed to evaluate them as they usually decide if they like a song or not. In Table 4 the obtained results are shown.The pieces generated with the presented method are shown in red, and the options that obtained more votes are presented in bold.

The results of set A show that both piece$_3$ and piece$_5$ were perceived as generated, even if only two people were 100% sure. However, piece$_2$ was also perceived as a possibly generated piece even though it is an original bertso melody. The only clear result in this set is the one obtained with piece$_4$, which is a well known melody and 9 people knew it.

The results of set B show that piece$_1$ was perceived as possibly generated, even though, in the cases of piece$_2$ and piece$_4$ the results are not clear, and the responses are distributed between possibly original and possibly generated.

The evaluation of the generated pieces show that even if they obtain a mean score from 6.06 to 7.63, the scores are quite different depending on the listener as some of the standard deviation values show. These results show that even though the participants were able to identify some of the generated melodies as possibly generated, they were overall confused and not able to clearly distinguish between original and generated pieces, as the numbers in bold show.

B. STYLE EVALUATION

In addition to the listeners subjective evaluation, a *style* evaluation has also been carried out. The goal of this evaluation is to determine if the generated pieces have the features of the bertso melodies, given that statistical models have been used in order to generate pieces that share stylistic features with the corpus. To do so, 20 new pieces have been generated with the presented method, and added to the five pieces that were evaluated.

The generated pieces have been compared to pieces of two corpora in addition to the bertso melody corpus; a subset of the Meertens Folk Tune Collection named The Annotated Corpus[4], which is a collection of 360 Dutch folk song melodies, and a corpus of pieces of three classical composers used in [53]. A classification process has been carried out in which the generated pieces have been classified within one of the three genres, bertso melodies, classical pieces or dutch folk melodies. Two different representations have been used to represent the pieces to be classified, both used in [53]: interval matrix and global feature representation.

The matrix representation represents the probabilities of the transitions between all the pitch class interval pairs that

[4]http://www.liederenbank.nl/

TABLE 5. Global feature collection used in the piece representation.

Variable	Feature Description
x_1	Prevalence of most common pitch
x_2	Prevalence of most common pitch class
x_3	Relative prevalence of top pitches
x_4	Relative prevalence of top pitch classes
x_5	Prevalence of most common melodic interval
x_6	Relative prevalence of most common melodic intervals
x_7	Repeated notes
x_8	Chromatic motion
x_9	Stepwise motion
x_{10}	Melodic thirds
x_{11}	Melodic perfect fifths
x_{12}	Melodic octaves

TABLE 6. Classification accuracies for the generated pieces with different classifiers and representations.

	J48	SMO	JRip	RF	MP
matrix	0.88	0.96	0.92	**1.00**	0.96
global	0.96	**1.00**	0.96	0.96	**1.00**

occur in each piece, representing pieces as 12×12 matrices. For the global feature representation 12 melodic features presented in Table 5 have been used, which was used by Herremans *et al.* [54].

Several different classifiers from the machine learning software Weka [55] have been tried in the classification of the generated pieces: J48, SMO, JRip, Random Forest (RF) and Multilayer Perceptron (MP). All the classifiers have been applied with the default parameters. The results obtained with each classifier and representation are shown in Table 6.

As it can be observed in the table, very high accuracies are obtained with both representations, and accuracies of 100% are obtained in some cases, which indicates that all the generated melodies are classified as bertso melodies. From these results it can be concluded that the method generates music that really emulates the style of the corpus, in this case it generates pieces that are classified as bertso melodies.

VII. CONCLUSION AND FUTURE WORK

This work extends the automatic music generation paper by Goienetxea and Conklin [22], which presents a music generation method that generates new melodic information. It is based on the use of the semiotic structure extracted from a template piece and a statistical model created from a corpus, in order to obtain new coherent melodies in the style of the corpus.

The extension described in this paper presents several contributions to the melody generation method of [22]. A rhythmic generation process is added, which follows the idea of identifying the rhythmically related segments in the template piece and respecting these relations when generating new rhythmic content. To do so, the abstract template that is extracted from the template piece is extended to include its rhythmic coherence structure too. To create the rhythmic structure a method to discover nested patterns has been developed, which is capable of discovering patterns with several levels of nesting, creating multi-level pattern hierarchies. This method allows more complex analysis of the coherence of the piece, and it could be used to discover interesting patterns within the patterns in the melodic coherence structure as well.

Given that in this case the generation method is applied to generate bertso melodies and they need to respect some of the allowed metrics, the strategy that is followed to sample the melodic information of the new pieces has been adapted for the rhythmic content. The proposed sampling strategy guarantees that the generated pieces respect the rhythmic coherence structure defined in the abstract template and that they have the same number of notes of the template piece.

An evaluation process was carried out to evaluate the melodies that are created with this method. In the first phase of the evaluation 31 participants from the university listened to five melodies, where original and generated pieces were mixed. First they had to say if the melodies were original bertso melodies or generated pieces, and then, they needed to score the generated pieces from 1 to 10. In the second phase of the evaluation the style of the generations was evaluated with a classification process in which the style of the generated pieces was classify as bertso, classical piece or Dutch folk melody.

The results obtained from the first phase of the evaluation show that, even if some of the participants were able to identify some of the generated pieces, these generations obtained good scores, showing that the participants liked them. In addition, the stylistic evaluation shows that the generated pieces are classified as bertso melodies, using both matrix and global feature representation.

In the presented method the melodic and the rhythmic information are treated independently; independent coherence structures are created to analyse the melodic line and the rhythm of the template piece, and two independent statistical models are also created. As mentioned before, it has been decided that it is the best way to deal with bertso melodies, considering that they can have segments with rhythmic repetitions that are not completely related melodically. However, the method presented herein is intended to be a general music generation method, and the melodic and rhythmic information can also be combined just linking melodic and rhythmic viewpoints. These linked viewpoints would be then used both for the semiotic analysis of the template piece and the building of the statistical model.

As future work different paths have been defined. Since the abstract template of the pieces can include both melodic and rhythmic coherence structures, adding more information, like harmonic structure, is envisaged. This would allow having richer abstract template and generating more interesting pieces, that take into account not only the melodic statistical model, but also different harmonic directions. This could be done adding new viewpoints of the template piece that represent harmonic information.

Thanks to the simplicity of viewpoint representations to describe events in various abstraction levels, new levels can be added in the melodic analysis or in the rhythmic one, or even to extract the structures from polyphonic pieces using viewpoints that describe vertical relations.

Regarding the proposed strategy to sample the rhythmic content of the new pieces, even though it is capable of generating sequences that respect the coherence structures and also maintain the number of notes of the template piece, it may sometimes be too restrictive. As future work, it should be updated in order to allow more possible generations.

The automatic generation of coherence structures is also being considered; learning the different coherence structures that appear in a corpus in order to be able to generate new ones. This would guarantee the generation of coherent new pieces without the need of a piece that is used as template.

The use of heterogeneous patterns has also been identified as future work; patterns that not only describe one type of segment relation, but patterns that can have different abstraction mixed. The need of nested pattern discovery within the melodic coherence structure should be studied.

The generation method is currently being used to generate music therapy exercises for people with Alzheimer's disease within a regional project funded by the Provincial Council of Gipuzkoa.

APPENDIX
EVALUATION SCORES

See Figure 19.

REFERENCES

[1] D. Cope, *Computers and Musical Style*. Madison, WI, USA: A-R Editions, 1991.

[2] S. A. Hedges, "Dice music in the eighteenth century," *Music Lett.*, vol. 59, no. 2, pp. 180–187, 1978.

[3] M. Edwards, "Algorithmic composition: Computational thinking in music," *Commun. ACM*, vol. 54, no. 7, pp. 58–67, Jul. 2011.

[4] J. D. Fernández and F. Vico, "AI methods in algorithmic composition: A comprehensive survey," *J. Artif. Intell. Res.*, vol. 48, no. 1, pp. 513–582, 2013.

[5] C. Liu and C. Ting, "Computational intelligence in music composition: A survey," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 1, no. 1, pp. 2–15, Feb. 2017.

[6] O. Lopez-Rincon, O. Starostenko, and G. A. Martín, "Algoritmic music composition based on artificial intelligence: A survey," in *Proc. Int. Conf. Electron., Commun. Comput. (CONIELECOMP)*, Feb. 2018, pp. 187–193.

[7] J. Sundberg and B. Lindblom, "Generative theories in language and music descriptions," *Cognition*, vol. 4, no. 1, pp. 99–122, 1976.

[8] N. G. Thomas, P. Pasquier, A. Eigenfeldt, and J. B. Maxwell, "A methodology for the comparison of melodic generation models using meta-melo," in *Proc. 14th Int. Soc. for Music Inf. Retr. Conf. (ISMIR)*, Curitiba, Brazil, 2013, pp. 561–566.

[9] C. Thornton, "Generation of folk song melodies using Bayes transforms," *J. New Music Res.*, vol. 40, no. 4, pp. 293–312, 2011.

[10] D. Conklin, "Multiple viewpoint systems for music classification," *J. New Music Res.*, vol. 42, no. 1, pp. 19–26, 2013.

[11] D. Conklin, "Music generation from statistical models," in *Proc. AISB Symp. Artif. Intell. Creativity Arts Sci.*, Wales, U.K., 2003, pp. 30–35.

[12] S. Dubnov, G. Assayag, O. Lartillot, and G. Bejerano, "Using machine-learning methods for musical style modeling," *Computer*, vol. 36, no. 10, pp. 73–80, Oct. 2003.

[13] M. Allan and C. K. I. Williams, "Harmonising chorales by probabilistic inference," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 25–32.

[14] L. M. Zbikowski, "Musical coherence, motive, and categorization," *Music Perception*, vol. 17, no. 1, pp. 5–42, 1999.

[15] A. Patel, *Music, Language, and the Brain*. New York, NY, USA: Oxford Univ. Press, 2008.

[16] C. Anagnostopoulou, "Cohesion in linguistic and musical discourse," in *Proc. 3rd Eur. Soc. Cognit. Sci. Music Conf.*, Uppsala, Sweden, 1997.

[17] J. Leach and J. Fitch, "Nature, music, and algorithmic composition," *Comput. Music J.*, vol. 19, no. 2, pp. 23–33, Feb. 1995.

[18] L. B. Meyer, "Meaning in music and information theory," *J. Aesthetics Art Criticism*, vol. 15, pp. 412–424, 1957.

[19] A. Schoenberg, P. Carpenter, and S. Neff, *The Musical Idea and the Logic, Technique and Art of Its Presentation*. Bloomington, IN, USA: Indiana Univ. Press, 2006.

[20] D. Cope, "An expert system for computer–assisted composition," *Comput. Music J.*, vol. 11, no. 4, pp. 30–46, 1987.

[21] F. Bimbot, E. Deruty, G. Sargent, and E. Vincent, "Semiotic structure labeling of music pieces: Concepts, methods and annotation conventions," in *Proc. 13th Int. Soc. Music Inf. Retr. Conf. (ISMIR)*, Porto, Portugal, Oct. 2012, pp. 235–240.

[22] I. Goienetxea and D. Conklin, "Melody transformation with semiotic patterns," in *Proc. 13th Int. Symp. Comput. Music Multidisciplinary Res. (CMMR)*, Porto, Portugal, 2017, pp. 328–339.

[23] J. Garzia, A. Egaña, and J. Sarasua, *The Art of Bertsolaritza: Improvised Basque Verse Singing*. Donostia, Spain: Bertsozale Elkartea, Andoain, Bertsolari Liburuak, 2001.

[24] M. Chemillier, "Toward a formal study of jazz chord sequences generated by Steedman's grammar," *Soft Comput.*, vol. 8, no. 9, pp. 617–622, 2004.

[25] J. A. Biles, *Improvizing With Genetic Algorithms: GenJam*. London, U.K.: Springer, 2007, pp. 137–169.

[26] M. Scirea, J. Togelius, P. Eklund, and S. Risi, *MetaCompose: A Compositional Evolutionary Music Composer*. Cham: Springer, 2016, pp. 202–217.

[27] J. Jeong, Y. Kim, and C. W. Ahn, "A multi-objective evolutionary approach to automatic melody generation," *Expert Syst. Appl.*, vol. 90, pp. 50–61, Dec. 2017, doi: 10.1016/j.eswa.2017.08.014.

[28] E. Munoz, J. M. Cadenas, Y. S. Ong, and G. Acampora, "Memetic music composition," *IEEE Trans. Evol. Comput.*, vol. 20, no. 1, pp. 1–15, Feb. 2016.

[29] F. P. Brooks, A. L. Hopkins, P. G. Neumann, and W. V. Wright, "An experiment in musical composition," *IRE Trans. Electron. Comput.*, vol. EC-5, pp. 175–182, Sep. 1956.

[30] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, "Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription," in *Proc. Int. Conf. Mach. Learn.*, Edinburgh, Scotland, 2012.

[31] J. Briot, G. Hadjeres, and F. Pachet, "Deep learning techniques for music generation—A survey," *CoRR*, vol. abs/1709.01620, 2017.

[32] R. P. Whorley and D. Conklin, "Music generation from statistical models of harmony," *J. New Music Res.*, vol. 45, no. 2, pp. 160–183, 2016.

[33] D. Herremans, K. Sörensen, and D. Conklin, *Sampling the Extrema From Statistical Models of Music With Variable Neighbourhood Search*. Athens, Greece: ICMC/SMC, 2014.

[34] V. Padilla and D. Conklin, "Generation of two-voice imitative counterpoint from statistical models," *Int. J. Interact. Multimedia Artif. Intell.*, vol. 5, no. 3, pp. 22–32, 2018. [Online]. Available: http://www.ijimai.org/journal/sites/default/files/files/2018/10/ijimai_5_3_3_pdf_45204.pdf

[35] T. Collins, R. Laney, A. Willis, and P. H. Garthwaite, "Developing and evaluating computational models of musical style," *Artif. Intell. Eng. Des., Anal. Manuf.*, vol. 30, no. 1, pp. 16–43, 2016, doi: 10.1017/S0890060414000687.

[36] D. Meredith, K. Lemström, and G. A. Wiggins, "Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music," *J. New Music Res.*, vol. 31, no. 4, pp. 321–345, 2002.

[37] C. Roig, L. J. Tardón, I. Barbancho, and A. M. Barbancho, "Automatic melody composition based on a probabilistic model of music style and harmonic rules," *Know.-Based Syst.*, vol. 71, no. 1, pp. 419–434, Nov. 2014, doi: 10.1016/j.knosys.2014.08.018.

[38] C. Roig, L. J. Tardón, I. Barbancho, and A. M. Barbancho, "A non-homogeneous beat-based harmony Markov model," *Knowl.-Based Syst.*, vol. 142, pp. 85–94, Feb. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950705117305592

[39] C.-Z. A. Huang, C. Hawthorne, A. Roberts, M. Dinculescu, J. Wexler, L. Hong, and J. Howcroft, "The Bach Doodle: Approachable music composition with machine learning at scale," in *Proc. Int. Soc. Music Inf. Retr. (ISMIR)*, 2019. [Online]. Available: https://goo.gl/magenta/bach-doodle-paper

[40] J. Briot and F. Pachet, "Music generation by deep learning—Challenges and directions," *CoRR*, vol. abs/1712.04371, 2017. [Online]. Available: http://arxiv.org/abs/1712.04371

[41] M. Bretan, G. Weinberg, and L. P. Heck, "A unit selection methodology for music generation using deep neural networks," *CoRR*, vol. abs/1612.03789, 2016. [Online]. Available: http://arxiv.org/abs/1612.03789

[42] L. Yang, S. Chou, and Y. Yang, "Midinet: A convolutional generative adversarial network for symbolic-domain music generation using 1D and 2D conditions," *CoRR*, vol. abs/1703.10847, 2017. [Online]. Available: http://arxiv.org/abs/1703.10847

[43] H. Dong, W. Hsiao, L. Yang, and Y. Yang, "Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment," in *Proc. 23nd (AAAI) Conf. Artif. Intell., (AAAI), 30th Innov. Appl. Artif. Intell. (IAAI), 8th AAAI Symp. Educ. Adv. Artif. Intell. (EAAI)*, New Orleans, LA, USA, Feb. 2018, pp. 34–41. [Online]. Available: https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17286

[44] G. Hadjeres, F. Pachet, and F. Nielsen, "DeepBach: A steerable model for Bach chorales generation," in *Proc. 34th Int. Conf. Mach. Learn.*, D. Precup and Y. W. Teh, Eds., vol. 70, Sydney, NSW, Australia, Aug. 2017, pp. 1362–1371. [Online]. Available: http://proceedings.mlr.press/v70/hadjeres17a.html

[45] G. Medeot, S. Cherla, K. Kosta, M. McVicar, S. Abdallah, M. Selvi, E. Newton-Rex, and K. Webster, "StructureNet: Inducing structure in generated melodies," in *Proc. 19th Int. Soc. Music Inf. Retr. Conf. (ISMIR)*, Paris, France, Sep. 2018, pp. 725–731. [Online]. Available: http://ismir2018.ircam.fr/doc/pdfs/126_Paper.pdf

[46] J. Dorronsoro, *Bertso Doinutegia*. Euskal Herriko Bertsolari Elkartea, 1995.

[47] D. Conklin and I. H. Witten, "Multiple viewpoint systems for music prediction," *J. New Music Res.*, vol. 24, no. 1, pp. 51–73, 1995.

[48] D. Conklin and M. Bergeron, "Feature set patterns in music," *Comput. Music J.*, vol. 32, no. 1, pp. 60–70, 2008.

[49] J. Ayres, J. Flannick, J. Gehrke, and T. Yiu, "Sequential pattern mining using a bitmap representation," in *Proc. 8th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, New York, NY, USA, 2002, pp. 429–435.

[50] D. Conklin, "Discovery of distinctive patterns in music," *Intell. Data Anal.*, vol. 14, no. 5, pp. 547–554, Oct. 2010.

[51] D. Conklin and S. Weisser, "Pattern and antipattern discovery in Ethiopian Bagana songs," in *Computational Music Analysis*, D. Meredith, Ed. Springer, 2016, pp. 425–443.

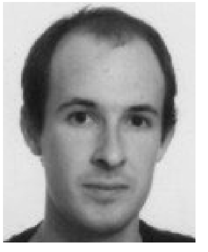
[52] J. van Helden, B. André, and J. Collado-Vides, "Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies," *J. Mol. Biol.*, vol. 281, no. 5, pp. 827–842, 1998.

[53] I. Goienetxea, I. Mendialdua, and B. Sierra, "On the use of matrix based representation to deal with automatic composer recognition," in *Proc. Adv. Artif. Intell. 31st Australas. Joint Conf. (AI)*, Wellington, New Zealand, Dec. 2018, pp. 531–536, doi: 10.1007/978-3-030-03991-2_48.

[54] D. Herremans, K. Sörensen, and D. Martens, "Classification and generation of composer-specific music using global feature models and variable neighborhood search," *Comput. Music J.*, vol. 39, no. 3, pp. 71–91, 2015.

[55] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *SIGKDD Explor. Newslett.*, vol. 11, no. 1, pp. 10–18, 2009.

IZARO GOIENETXEA received the B.Sc. degree in computer science from the University of the Basque Country, Spain, in 2008, and the Ph.D. degree from the Department of Computer Science and Artificial Intelligence, University of the Basque Country, in 2019. She is currently a member of the Robotics and Autonomous Systems Research Group, Department of Computer Science and Artificial Intelligence. Her current research interests include music generation and classification, heuristics, and decomposition strategies.



IÑIGO MENDIALDUA received the B.Sc. and Ph.D. degrees in computer science from the University of the Basque Country (UPV/EHU), in 2007 and 2015, respectively. He is currently a Lecturer with the Computer Languages and Systems Department, UPV/EHU. He has been a member of the Robotics and Autonomous Systems Research Group, since 2009. His current research interests include music classification and generation techniques, multiclassifiers, and decomposition strategies.

IGOR RODRÍGUEZ received the B.Sc. and Ph.D. degrees in computer science from the University of the Basque Country, in 2012 and 2018, respectively. He is currently a member of the Robotics and Autonomous Systems Research Group, Department of Computer Science and Artificial Intelligence. His current research interests include robotics, human–robot interaction, and machine learning.

BASILIO SIERRA received the B.Sc. degree in computer sciences, the M.Sc. degree in computer science and architecture, and the Ph.D. degree in computer sciences from the University of the Basque Country, in 1990, 1992, and 2000, respectively. He is currently a Full Professor with the Computer Sciences and Artificial Intelligence Department, University of the Basque Country. He is also the Co-Director of the Robotics and Autonomous Systems Group, Donostia-San Sebastian. He is also a Researcher in the fields of robotics and machine learning, where he is working on the use of different paradigms to improve behaviours.