SAMPLING VARIATIONS OF SEQUENCES FOR STRUCTURED MUSIC GENERATION

François Pachet
Sony CSL Paris
pachet@gmail.com

Alexandre Papadopoulos
UPMC Univ Paris 06, UMR 7606, LIP6
alexandre.papadopoulos@lip6.fr

Pierre Roy
Sony CSL Paris
roypie@gmail.com

ABSTRACT

Recently, machine-learning techniques have been successfully used for the generation of complex artifacts such as music or text. However, these techniques are still unable to capture and generate artifacts that are convincingly structured. In particular, musical sequences do not exhibit pattern structure, as typically found in human composed music. We present an approach to generate structured sequences, based on a mechanism for sampling efficiently variations of musical sequences. Given an input sequence and a statistical model, this mechanism uses belief propagation to sample a set of sequences whose distance to the input sequence is approximately within specified bounds. This mechanism uses local fields to bias the generation. We show experimentally that sampled sequences are indeed closely correlated to the standard musical similarity function defined by Mongeau and Sankoff. We then show how this mechanism can be used to implement composition strategies that enforce arbitrary structure on a musical lead sheet generation problem. We illustrate our approach with a convincingly structured generated lead sheet in the style of the Beatles.

1. INTRODUCTION [1]

Recent advances in machine learning, especially deep recurrent networks such as LSTMs, led to major improvements in the quality of music generation [7, 10]. They achieve spectacular performance for short musical fragments. However, musical structure typically exceeds the scope of statistical models. As Waite recently wrote, the music produced by recurrent models tend to lack a sense of direction and becomes boring after a short while [15]. Pionneering works on music composition with LSTMs already showed how some structure, such as chord structure [6] or metrical structure [5] can be *spontaneously* cap-

---

[1] Authors are listed alphabetically: Pachet originated the general problem and contributed musical examples; Papadopoulos developed and implemented the technical solution especially the integration with the regular belief propagation model, devised and performed the evaluation procedure; Roy brought the original idea and the technical solution, developed the first prototype and the structured lead sheet generation procedures.

tured, but the general problem of generating music with repetitive long-term structure remains open. In this paper, we propose a method to *explicitly* enforce such structure in a controlled way, in a "templagiarism" fashion [2, p. 49].

Musical structure is the overall organisation of a composition into sections, phrases, and patterns, very much like the organisation of a text. The structure of musical pieces is scarcely, if ever, linear as it essentially relies on the repetition of these elements, possibly altered. For example, songs are decomposed into repeating sections, called verses and choruses, and each section is constructed with repeating patterns. It has been shown that the *listeners' emotional arousal responses to music* is correlated with the degree of similarity between musical fragments (high for repetitions, moderate for variations, and least for contrasting segments) [9]. In fact, the striking *speech to song* illusion discovered by [4] shows that repetition truly *creates* music, for instance by turning speech into music. This is further confirmed by [11] who observed that inserting arbitrary repetition in non-repetitive music improves listeners rating and confidence that the music was written by a human composer.

Figure 1. The last eight bars of *"Strangers in the Night"*.

Variations are a specific type of repetition, in which the original melody is altered in its rhythm, pitch sequence, and/or harmony. Variations are used to create diversity and surprise by subtle, unexpected changes in a repetition. The song *"Strangers in the Night"* is a typical 32-bar form with an AABA structure consisting of four 8-bar sections. The three **A** sections are variations of each other. The last **A** section, shown in Figure 1, consists of a two-bar cell which is repeated three times. Each occurrence is a subtle variation of the preceding one. The second occurrence (bars 3-4) is a mere transposition of the original pattern by one descending tone. The third instance (bars 5-6) is also transposed, but with a slight modification in the melody, which creates a surprise and concludes the song. Bars 5-6 are both a variation of the original pattern in bars 1-2. Current models for music generation fail to reproduce such

long-range similarities between musical patterns. In this example, it is statistically unlikely that bars 5-6 be almost identical to bars 1-2.

Our goal is to generate such structured musical pieces from statistical models. Our approach is to impose a predefined musical structure that specifies explicitly repetitions and variations of patterns and sections, and use a statistical model to generate music that "instantiates" this structure. In this approach, musical structure is viewed as a procedural process, external to the statistical model.

Our approach subsumes previous attempts at generating music with an imposed long-term structure with Markov models such as [1]. Their approach lacks both a variation mechanism and a constrained Markov model. As a result, it is limited to strict repetitions of patterns. Furthermore, the use of ad hoc joining techniques to glue copied fragments, violates the Markov model, resulting in unnatural transitions.

An essential ingredient to implementing our approach is a mechanism to generate variations of a given musical pattern from a statistical model. Although it is impossible to characterise formally the notion of variation, it was shown that some measures of melodic similarity are efficient at detecting variations of a theme [12]. We propose to use such a similarity measure in a generative context to sample from a Markov model, patterns that are similar to a given pattern. This method is related to work on stochastic edit distances [3, 14], but is integrated as a constraint in a more general model for the generation of musical sequences [13]. Moreover, our approach relies on an existing similarity measure rather than on labeled data (pairs of themes and related variations), which is not available. Similar approaches exist in the context of text generation. For example, [8] propose a model using a technique based on skip vectors. They train a model that learns the similarity between sentences. Using this model, they can predict the semantic relatedness of two sentences, a standard similarity measure for text, but they can also generate sentences similar to an existing sentence.

We remind the Mongeau & Sankoff similarity measure [12] between melodies, and then, we describe our model for sampling melodic variations based on this similarity, which we validate experimentally, Finally, we show examples of variations of a melody, and a longer, structured musical piece generated with an imposed structure.

2. MELODIC SIMILARITY

The traditional string edit distance considers three editing operations: *substitution*, *deletion*, and *insertion* of a character. Mongeau and Sankoff [12] add two operations motivated by the specificities of musical sequences, and inspired by the time compression and expansion operations considered in *time warping*. The first operation, called *fragmentation*, involves the replacement of one note by several, shorter notes. Similarly, the *consolidation* operation, is the replacement of several notes by a single, longer note. Mongeau and Sankoff proposed an algorithm to compute the similarity between melodies in polynomial time.

Considering melodies as sequences of notes, the algorithm, based on dynamic programming, computes $\mathrm{MGD}(A, B)$, the measure of similarity between the sequences of notes A and B. Note that this is not a distance, in particular $\mathrm{MGD}(A, B)$ is not necessarily equal to $\mathrm{MGD}(B, A)$.

The Mongeau & Sankoff measure is well-adapted to the detection of variations, but has a minor weakness: there is no penalty associated with fragmenting a long note into several shorter notes of same pitch and same total duration. The same applies to consolidation. This is not suited to a generative context, as fragmentation or consolidation change the resulting melody.

In the dynamic programming recurrence equation given in their paper [12], Mongeau and Sankoff introduce various weight functions, denoting predefined local weights associated with the basic editing operations (substitution, deletion, insertion, fragmentation and consolidation). We modify the original measure by adding a penalty p to the weights of the consolidation and fragmentation operations.

The weight associated with a fragmentation of a note a_i into a sequence of notes $b_{j-k+1}, \ldots, b_j$ is:

$$w_{frag}(a_i, b_{j-k+1}, \ldots, b_j) = w_{pitch}(a_i, b_{j-k+1}, \ldots, b_j) \\ + k_1 n(a_i, b_{j-k+1}, \ldots, b_j) + p$$

For consolidation, a similar extra-weight is added. The consolidation weight is defined by:

$$w_{cons}(a_i, b_{j-k+1}, \ldots, b_j) = w_{pitch}(a_i, b_{j-k+1}, \ldots, b_j) \\ + k_1 n(a_i, b_{j-k+1}, \ldots, b_j) + p.$$

3. A MODEL FOR THE GENERATION OF MELODIC VARIATIONS

Given an original *theme*, i.e. a melodic fragment, we generate variations of this theme by sampling a specific graphical model. This graphical model is a modified version of the general model of lead sheets introduced by [13]. We now briefly describe this general model and explain how we bias it to produce only melodies at a *controlled* Mongeau & Sankoff distance from the theme, the core technical contribution of this paper. For full explanations and implementation details, we refer the reader to [13].

3.1 The Model of Lead Sheets

The overall model comprises *two graphical models*, one for chord sequences, one for melodies. Both models are based on a factor graph that combines a Markov model with a finite state automaton. The Markov model, trained on a corpus of lead sheets, provides the stylistic model. The automaton represents hard temporal constraints that the generated sequences should satisfy, such as metrical properties (e.g., an imposed total duration) or user imposed temporal constraints.

Each factor graph is made of a sequence of variables, represented with circles, encoding the sequence of elements, related to unary and binary factors, represented by squares. In this model, a *variable* is not associated with
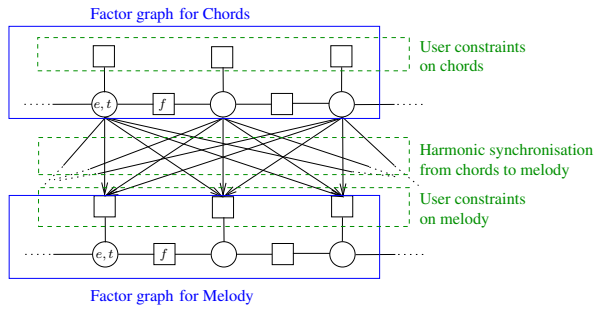
Figure 2. The two-voice model for lead sheet generation

Figure 3. The first four bars of *"Solar"*, by Miles Davis.

a specific temporal position in the sequence, but the *values* it takes specifies its temporal position. Each value is a chord or a note e, with a fixed duration $d(e)$ *along with* its temporal position t in the sequence. This is a very powerful property of this model. It allows us to specify unary temporal constraints, e.g., the second bar should start with a rest. It also allows us to specify harmonic relations between the chord sequence and the melody, e.g., the note at time t should be compatible with the chord at time t. Crucially, we will exploit this property to implement our variation mechanism.

A binary factor is a conditional probability $f((e, t) | (e', t'))$ on transitions between elements. In [13], the authors use binary factors to combine the Markov transition probabilities with the finite-state automaton transitions. Harmonic relationship between chords and notes are also specified by binary factors.

The graphical model defines a distribution $p(e_1, \ldots, e_n)$ over the sequence of variables defined by the product of all unary and binary factors. A belief propagation-based procedure samples successively the two models by taking into account partially filled fragments and propagating their effect to all empty sections.

3.2 Generating Variations of a Theme

We introduce an extra binary factor $\beta(e | t, e')$: the probability of placing element e at time t and preceded by element e'. We will use β to implement the variation mechanism. In practice, this additional binary factor is simply multiplied with the existing binary factors, without affecting the structure of the model on Figure 2. The probability p' of a sequence in the resulting model becomes:

$$p'(e_1, \ldots, e_n) = p(e_1, \ldots, e_n) \prod_{i=2}^{n} \beta(e_i | t, e_{i-1}).$$

We set the value of $\beta(e | t, e')$ according to a "localised" similarity measure between the sequence $[e', e]$ and the fragment of the theme between $t - d(e')$ and $t + d(e)$. Biases are set so that a bias of 1 does not modify the probability of putting element e at time t after e', and a bias less than 1 decreases this probability.

The lead sheet in Figure 3 shows the first four bars of *"Solar"* by Miles Davis. Suppose we train a lead sheet model on a corpus of all songs by Miles Davis. Sampling

this model produces new lead sheets in the style of Miles Davis, but not necessarily similar to Solar specifically. To favour sequences with the same notes as the theme is to set the β factors so that:

- $\beta(n | t, n') = 1$ if the melodic fragment consisting of note n' followed by note n at position t appears in the theme, e.g., we set $\beta(C_5 | t = 1.5, rest) = 1$ for note C_5 *dotted quarter note*;

- $\beta(n | t, n') < 1$ otherwise, and the value of $\beta(n | t, n')$ will be set to very small values (close to zero), if the melodic fragment made by n' and n at time t is very different, musically, from the corresponding melodic fragment in the theme, e.g., $\beta(F\sharp_4 | t = 1.5, G\flat_5) \ll 1$. On the contrary, if the two fragments are very similar, musically, the value of $\beta(n | t, n')$ will be set to a value closer to 1, e.g., $\beta(C_5 | t = 1.5, rest) \gg 0$ for note C_5 *quarter note*.

More precisely, we evaluate the similarity between each possible note n at a given position t, preceded by note n' in the generated sequence, and the notes in the theme around position t. We then set each bias $\beta(n | t, n')$ based on this similarity measure.

Technically, for every candidate note n, we consider all potential temporal positions t and all potential predecessors n'. We compute $\mathrm{MGD}([n', n], t)$, the Mongeau & Sankoff similarity between the two-note melody $[n', n]$ and the melodic fragment of the theme between time $t - d(n')$ and $t + d(n)$, where $d(n)$ is the duration of the note n, i.e. the melodic fragment that would be replaced by placing the melody $[n', n]$ at time $t - d(n')$. The notes of the theme that overlap the time interval $[t - d(n'), t + d(n)]$ are trimmed so that the extracted melody has the same duration as the candidate notes. Similarly $\mathrm{MGD}([n'], t)$ denotes the similarity of the one-note sequence $[n']$ starting at $t - d(n')$. We call those similarities *localised* Mongeau & Sankoff similarity measures. The idea is that the similarity measure obtained by summing those localised measures over a complete sequence approximates the actual Mongeau & Sankoff similarity. This will be confirmed experimentally in the next section.

To convert the similarity measure into a weight between 0 and 1, we rescale those values to the $[0, 1]$ interval, and then invert their order, so that a value of 1 is the closest to the theme, and 0 the furthest away. Finally, we exponentiate the result, so that the logarithm of the product of the biases achieved by the model is proportional to the approximated Mongeau & Sankoff similarity. Formally, we define $\beta(n | t, n')$ as follows, where MGD_{max} is the maximal value of localised Mongeau & Sankoff similarities:

$$\beta(n | t, n') = \exp\left(1 - \frac{\mathrm{MGD}([n', n], t) - \mathrm{MGD}([n'], t)}{\mathrm{MGD}_{max}}\right)$$

3.3 Controlling the Similarity

We define an additional mechanism to control the *intensity* of the variation mechanism, i.e. the extent to which the generated melodies should be similar to the imposed theme. We introduce a parameter α, which is used to adjust the values of the biases β to new values β', defined as $\beta'(n|t,n') = \max(0, (1-\alpha).\beta(n|t,n')+\alpha)$. In theory, α ranges from $-\infty$ to 1: a very small value will cause almost all adjusted biases β' to be equal to 0, except when β was very close to 1, leading to melodies highly similar to the theme. Conversely, when α is 1, all adjusted biases β' are equal to 1, and have no effect. The interesting, non-trivial, behaviour is obtained with in-between values, which can be chosen by the user of the variation mechanism. However, the range of values where the non-trivial behaviour is observed depends on a particular corpus and a given theme. This means that a specific value of α has no general semantics, which hinders usability. As a result, we calibrate the range of α, by estimating the values for which the non-trivial behaviour occurs, given a specific corpus and theme. We estimate the values α^- and α^+ such that the average value of all adjusted biases β' is a given value close to 0 or close to 1, respectively. We estimate those values with a simple binary search. Given those two values, the user of the system then sets a parameter $\sigma \in [0,1]$, the *strictness* of the variation mechanism, and the actual value α is deduced by setting $\alpha = \sigma(\alpha^+ - \alpha^-) + \alpha^-$. We evaluate the effect of σ in practice in the next section.

4. EXPERIMENTAL RESULTS

Our approach relies on the intuition that *local* similarities, favoured by the biased model, will result in a *global* similarity between the generated melodies and the theme. In this section, we evaluate how the choice of the value for the parameter σ influences the Mongeau & Sankoff similarity between the generated melodies and the original theme. In particular, we show that the biased model favours sequences closer to the theme and penalises sequences less similar to the theme. We then explain the result more analytically, for $\sigma = 0$. We first show that applying the bias to the model approximates the localised Mongeau & Sankoff similarity, and then we show that this localised Mongeau & Sankoff similarity is a good approximation of the actual, global Mongeau & Sankoff similarity.

In the experiments below, the theme is the melody in the first four bars of *"Solar"* (Miles Davis, Figure 3). The training corpus contains 29 lead sheets by Miles Davis. In each experimental setup, we build a general model of 4-bar lead sheets in the style of Miles Davis, called the *unbiased model*, and then, we bias the model to favour the theme with some value for σ. Actual examples of variations at various distances are shown in Section 5.1.

4.1 Correlation between the Biases and the Mongeau & Sankoff Distance

For one value of σ, we generate 10 000 variations of the original theme (first four bars of *"Solar"*). For each sequence, we compute its probability p_o in the unbiased model and its probability p_b in the biased model, and then consider the ratio p_b/p_o. This probability ratio shows by how much the sequence has been favoured, for values greater than 1, or conversely penalised, for values less than 1, in the biased model. On Figure 4, points in blue are sequences generated with the most biased model, i.e. $\sigma = 0$. For each sequence, we plot its probability ratio, on a log scale, against its Mongeau & Sankoff similarity with the theme. We observe that the logarithm of the probability ratio tends to decrease linearly as the Mongeau & Sankoff distance with the theme increases. Sequences at a distance less than 75 from the theme are boosted while sequences at a distance more than 75 from the theme are hindered. Points in black are sequences generated with $\sigma = 0.95$, i.e. almost no bias at all. We observe that most sequences have a probability ratio of 1, i.e. that the biased model hardly affects the probability of sequences. Only sequences very far from the theme have their probability slightly decreased. Points in the red are generated with $\sigma = 0.5$. They display an intermediate behaviour as expected.
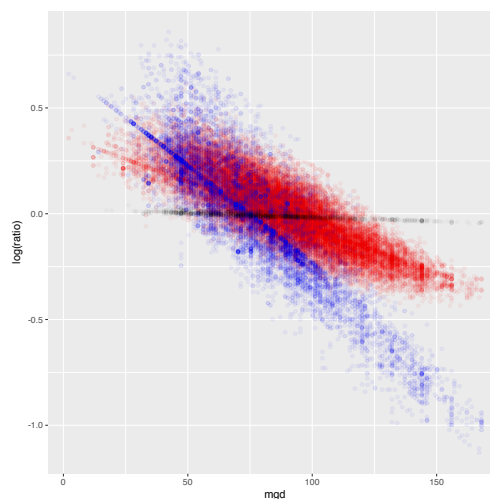

Figure 4. Sequence probability ratio (log) against Mongeau & Sankoff similarity to theme. Sequences in blue, red and black have been generated with $\sigma = 0, \sigma = 0.5, \sigma = 0.95$, respectively.

4.2 Explaining the Correlation

We explain the correlation observed by the application of two successive approximations. We concentrate on the case where $\sigma = 0$, but similar results are obtained with other values. We can break our analysis in three steps.

First, we note that for a given sequence, its probability ratio is equal, by definition of the biased model, to the product of all the local biases applied to each element of the sequence, up to a normalisation factor. We verified this experimentally too: for each generated sequence, we computed the local bias of each of the elements of the sequence, and computed the product of those local biases. We observed that this product is perfectly correlated with

the ratio of probabilities of the sequence. Second, we show how the probability ratio compares with the approximated Mongeau & Sankoff similarity measure obtained by summing the localised Mongeau & Sankoff similarity measures. For each sequence, we sum, over all its elements, the localised Mongeau & Sankoff that was used when computing the biases, as explained in Section 3.2. Then, we compare this sum to the product of the local biases, equal to the probability ratio. We plot the result on Figure 5. We observe that the approximated Mongeau & Sankoff similarity measure is tightly correlated with the logarithm of the product of the local biases, i.e., the logarithm of the product of the local biases approximates closely enough the sum of the localised Mongeau & Sankoff distances.
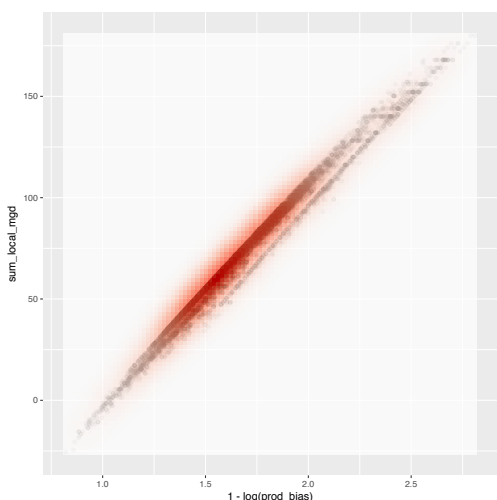
Figure 5. The sum of localised Mongeau & Sankoff similarity measures against the product of local biases (log), for $\sigma = 0$

Finally, we show that this approximated Mongeau & Sankoff similarity measure approximates the actual Mongeau & Sankoff similarity measure. On Figure 6, we plot for each sequence, the approximate versus the actual similarity measure. We observe that, although the actual measure is a global, dynamic programming-based measure, it is adequately approximated by summing the localised versions. This is probably because the localised measure captures sufficiently the effect of a note on the global similarity measure.

5. GENERATING STRUCTURED LEAD SHEETS

We show examples of melodic variations produced with our techniques, to give a concrete illustration of the variation mechanism. Then, we use the variation mechanism as the key building block to generate structured lead sheets[2].

5.1 Melodic Variations

Figure 7 shows six melodic variations of the first four bars of *"Solar"*, by Miles Davis. These variations were created

[2] All examples are available on http://www.flow-machines.com/ismir-examples/
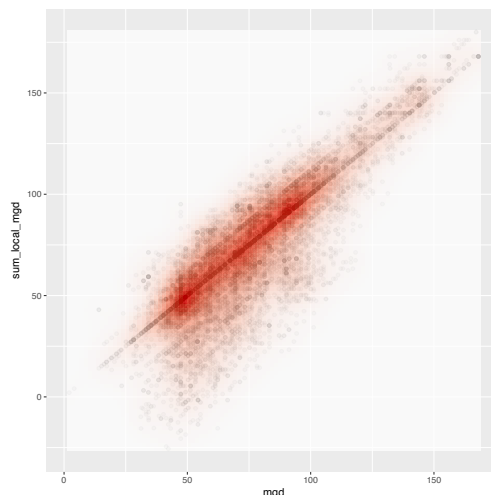
Figure 6. The sum of localised Mongeau & Sankoff similarity measures against the actual Mongeau & Sankoff measure, for $\sigma = 0$

using a model trained on 29 songs by Miles Davis (Section 4). The variations are presented in increasing order of Mongeau & Sankoff distance with the original theme (Figure 3). Note that the variations are increasingly different from the theme, both rhythmically and melodically.

(a) Mongeau & Sankoff distance 12: highly similar to the theme

(b) Distance 86, minor enrichments in bars 1 and 3

(c) Distance 87, minor enrichments in bars 1 and 3

(d) Distance 224, with major differences in bars 2 and 3

(e) Distance 285, interesting triplet rhythm in bar 1

(f) Dist. 295, large initial interval (octave) and end of bar 3 differs from other variations

(g) Dist. 906, first bar uses a rhythm similar that of *"Miles Ahead"* (Miles Davis), and bar 3 is introduces a new rhythm, similar to that of the original theme, except with dotted quarter notes

Figure 7. Several variations of the first four bars of *"Solar"*, by increasing Mongeau & Sankoff distance.

5.2 Enforcing Structure

We describe our strategy for automatic composition of structured lead sheets. We use the structure of *"In a Sentimental Mood"* (Duke Ellington, Figure 8). This song has a classical AABA 32-bar structure preceded by a pickup bar:

- Sections: **Pickup**: bar 1; **A1**: bars 2 to 9, **A2**: bars 2 to 8 and bar 10, **B**: bars 11 to 18; **A3**: bars 19 to 26.

- Bar 12 is a *transposed variation* of bar 11;

- Bars 15-16 are *exact copies* of bars 11-12;

- The last bar 26 is a *variation* of bar 10, the ending of Section **A2**.

Figure 8. *"In a Sentimental Mood"* by Duke Ellington. Red boxes correspond to the basic blocks induced by the structure of the piece.

We illustrate our approach with an automatically generated lead sheet that conforms to this structure. This structure induces a segmentation of the lead sheet into contiguous blocks of music. We transform the description of the structure into a procedure that executes it. The first occurrence of each block is generated using the general model of lead sheets. Subsequent occurrences, if any, are copied from the first occurrence. If specified by the structure description, we use the variation mechanism to obtain a variation instead of an exact copy, with a strictness that may be specified by the structure description.

Each block may appear in several places, but has been generated only once, without taking into account all possible contexts. This may have the adverse effect of creating awkward transitions that the model would not have created. In these situations, we systematically apply the variation mechanism to ensure seamless transitions between blocks. Since these variations are not specified by the structure, we impose a very strict variation to ensure minimal differences with the structure description.

The chords are generated by the general model of lead sheets, either before the melody or after. In fact, there is often structure in the chord sequence too. For example, bars 4-5 of *"In a Sentimental Mood"* are a transposition of bars 2-3. We can apply the same approach, with a different notion of distance on chords.

Figure 9 shows a lead sheet with this structure and generated from a stylistic model of the Beatles (trained from

Figure 9. A lead sheet with the structure of *"In a Sentimental Mood"* but in the style of the Beatles. Note that bar 12 is a transposed variation of bar 11, as in the original song. The ending is also a variation of the ending of Section **A1**.

a corpus with 201 lead sheets by the Beatles). The music does not sound similar to *"In a Sentimental Mood"* at all, but its structure, with multiple occurrences of similar patterns, make it feel like it was composed with some intentions. This is never the case of structureless 32-bar songs composed from the general model. Each part of the lead sheet has a strong internal coherence. The melody in the **A** parts use mostly small steps and fast sixteenth notes, many occurrence of a rhythmic pattern combining a sixteenth note with a dotted eighth note. The **B** part uses many leaps (thirds, fourth and fifth) and a regular eighth note rhythm. This internal coherence is a product of the imposed structure. For instance, in the **B** part, four out of eight bars come from a single original cell, consisting of bar 11. The fact that the **A** and **B** parts contrast with one another is also a nice feature of this lead sheet. This contrast simply results from the default behaviour of the general model of lead sheets.

6. CONCLUSION

We have presented a model for sampling variations of melodies from a graphical model. This model is based on the melodic similarity measure proposed by [12]. Technically, we use an approximated version of the Mongeau & Sankoff similarity measure to bias a more general model for the generation of music. Experimental evaluation shows that this approximation allows us to bias the model towards the generation of melodies that are similar to the imposed theme. Moreover, the intensity of the bias may be adjusted to control the similarity between the theme and the variations. This makes this approach a powerful tool for the creation of pieces complying with an imposed musical structure. We have illustrated our method with the generation of a long structured lead sheet. A pop music album is currently being produced using this method.

7. REFERENCES

[1] Tom Collins and Robin Laney. Computer–generated stylistic compositions with long–term repetitive and phrasal structure. *Journal of Creative Music Systems*, 1(2), 2017.

[2] David Cope. *Virtual music: computer synthesis of musical style*. MIT press, 2004.

[3] Ryan Cotterell, Nanyun Peng, and Jason Eisner. Stochastic Contextual Edit Distance and Probabilistic FSTs. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 625–630, 2014.

[4] Diana Deutsch, Trevor Henthorn, and Rachael Lapidis. Illusory Transformation from Speech to Song. *The Journal of the Acoustical Society of America*, 129(4):2245–2252, 2011.

[5] Douglas Eck and Jasmin Lapalme. Learning musical structure directly from sequences of music. *University of Montreal, Department of Computer Science, CP*, 6128, 2008.

[6] Douglas Eck and Juergen Schmidhuber. Finding temporal structure in music: Blues improvisation with lstm recurrent networks. In *Neural Networks for Signal Processing, 2002. Proceedings of the 2002 12th IEEE Workshop on*, pages 747–756. IEEE, 2002.

[7] G Hadjeres and F. Pachet. Deepbach: a steerable model for bach chorales generation. Technical report, arXiv:1612.01010, December 2016. https://arxiv.org/abs/1612.01010.

[8] Ryan Kiros, Yukun Zhu, Ruslan Salakhutdinov, Richard S. Zemel, Antonio Torralba, Raquel Urtasun, and Sanja Fidler. Skip-thought vectors. *CoRR*, abs/1506.06726, 2015.

[9] Steven R Livingstone, Caroline Palmer, and Emery Schubert. Emotional Response to Musical Repetition. *Emotion*, 12(3):552–567, 2012.

[10] Qi Lyu, Zhiyong Wu, Jun Zhu, and Helen Meng. Modelling high-dimensional sequences with lstm-rtrbm: Application to polyphonic music generation. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pages 4138–4139. AAAI Press, 2015.

[11] Elizabeth Hellmuth Margulis. Aesthetic responses to repetition in unfamiliar music. *Empirical Studies of the Arts*, 31(1):45–57, 2013.

[12] Marcel Mongeau and David Sankoff. Comparison of Musical Sequences. *Computers and the Humanities*, 24(3):161–175, 1990.

[13] Alexandre Papadopoulos, Pierre Roy, and François Pachet. Assisted Lead Sheet Composition using Flow-Composer. In *Principles and Practice of Constraint Programming – CP 2016*. Springer, 2016.

[14] Eric Sven Ristad and Peter N Yianilos. Learning String-Edit Distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(5):522–532, 1998.

[15] Elliot Waite. Generating Long-Term Structure in Songs and Stories. https://magenta.tensorflow.org/blog/2016/07/15/lookback-rnn-attention-rnn/, 2016.