Vol.2018-MUS-121 No.27
2018/11/21

メロディ生成における生成単位に関する一調査

北原 鉄朗[1,a)]

概要：本稿では，既存のメロディから数音ずつのフレーズを抽出してメロディを再生成したときに，元のメロディに気付けるかどうかに関する実験について述べる．計算機によるメロディ生成では，音符 1 つ 1 つの時間的変化をモデル化するよりも，数個の音符の列であるフレーズの時間的変化をモデル化する方が，時間的つながりのよいメロディを生成できると予想される．しかしながら，既存曲からフレーズを抽出して新しいメロディを生成した際に，元のメロディに気づくかどうかは検討されていなかった．本稿では，既存のメロディを 1 音ずつ，2 音ずつ，4 音ずつ，または 8 音ずつ分割してフレーズ辞書に登録し，そのフレーズ辞書を用いてメロディを再生成したものを被験者に聴いてもらい，被験者が元のメロディに気付くかどうかの実験を行った．実験の結果，フレーズの長さが 1 音のとき（つまり音符単位でメロディを生成）よりも 2 音以上のときの方が，元のメロディに気付く被験者は多かった．

1. はじめに

メロディ生成手法として，これまで様々なものが提案されてきたが，最も典型的なものは，音符の時間的つながりをマルコフモデルやリカレントニューラルネットワーク（RNN）などで学習する方法である [1], [2], [3], [4], [5].

音符 1 つ 1 つの時間的つながりをマルコフモデルや RNN で学習する方法は，必ずしも最適とは限らない [*1]．音符は，言語における文字に相当すると考えられるが，言語では文字のつながりではなく単語のつながりを考えることも少なくない．では，メロディにおいて単語に相当するものは何であろうか．音楽的に意味のある一区切りの音符列をフレーズというが，このフレーズが言語における単語に相当すると考えることができる [*2].

フレーズを最小単位とする場合，フレーズ辞書を作る必要がある．フレーズそのものが音楽的に適切でなければ，それらを組み合わせても音楽的に適切なメロディは生成できないので，フレーズ辞書は，おのずと音楽的な適切さが保証された既存曲から抽出することになる．そうすれば，少なくともフレーズ自体の音楽的適切さは保証されているため，それらを組み合わせて作ったメロディも音楽的に適切なものになりやすいと予想される．また，特定のジャンルや特定のアーティストのメロディ集からフレーズ辞書を作成すれば，そのジャンルやそのアーティストらしさを反映させたメロディを生成することもできる．

しかしながら，フレーズ辞書に追加されたフレーズにそれなりの長さがある場合，フレーズ自体にメロディとしてのアイデンティティが存在し，それを使って新しいメロディを生成しても，元のメロディらしさが色濃く残る可能性がある．そのようなメロディを聴いた人が元のメロディに気付いてしまえば，いわゆる「パクリ認定」されてしまい，著作権上も問題が発生する可能性がある．

本稿では，既存のメロディに対して，そのメロディから 1 音，2 音，4 音または 8 音からなるフレーズ辞書を作成してメロディを再生成し，これを被験者に聴いてもらって元のメロディに気付けるかどうかの実験を行う．フレーズ辞書内の各フレーズの長さが 4 音や 8 音のように長い場合には，元のメロディに気付きやすく，各フレーズの長さが 1 音や 2 音のように短い場合は，元のメロディに気付きにくいと予想される．本実験では，フレーズの長さと元のメロディの気付きやすさの関係が上述の通りになっているかどうかを確かめる．

2. 実験方法

2.1 使用メロディ

次の 4 曲のサビ 8 小節分を用いた．

- 『レット・イット・ゴー ～ ありのままで ～』（作曲：クリステン・アンダーソン＝ロペス，ロバート・ロペス，発売：ウォルト・ディズニー・レコード）

1 日本大学文理学部情報科学科
a) kitahara@chs.nihon-u.ac.jp

*1 音符単位よりも大きな時間的つながりを学習する研究としては，小節ごとに畳み込みニューラルネットワークでピアノロールを学習する MidiNet[6] などがある．

*2 言語においては「フレーズ」は単語をいくつかつなげたものを表し，単語とは異なる概念であるが，音楽において「単語」に相当する概念が存在しないため，便宜上，音楽でいうフレーズが言語でいう単語に対応すると考える．

© 2018 Information Processing Society of Japan

表 1 実験結果（再生成されたメロディを聴いたときの元のメロディの回答度数，DK: Don't know）

フレーズサイズ	1					2					4					8				
回答	1	2	3	4	DK	1	2	3	4	DK	1	2	3	4	DK	1	2	3	4	DK
メロディ 1	7	4	0	1	4	11	3	1	1	0	6	2	6	0	2	5	0	2	3	6
メロディ 2	4	8	1	1	2	0	11	0	2	3	1	15	0	0	0	4	10	1	1	0
メロディ 3	0	1	8	6	1	2	0	11	1	0	1	3	11	1	0	0	1	13	1	1
メロディ 4	6	0	1	7	2	0	0	2	10	4	1	0	0	11	2	4	1	0	11	0

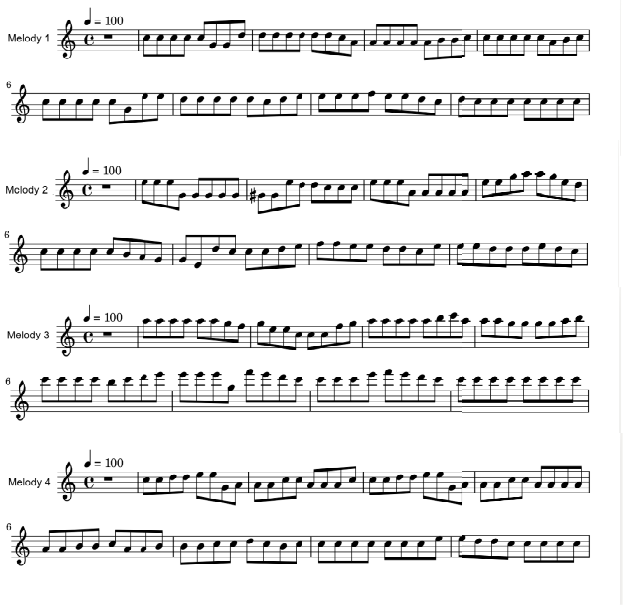

図 1 元のメロディ（メロディ 1：『レット・イット・ゴー ～ ありのままで ～』（作曲：クリステン・アンダーソン＝ロペス，ロバート・ロペス，発売：ウォルト・ディズニー・レコード），メロディ 2：『あったかいんだからぁ♪』（作曲：クマムシ，発売：ユニバーサル・ミュージック），メロディ 3：『RPG』（作曲：Fukase，発売：TOY'S FACTORY），メロディ 4：『恋するフォーチュンクッキー』（作曲：伊藤心太郎，発売：キングレコード））

- 『あったかいんだからぁ♪』（作曲：クマムシ，発売：ユニバーサル・ミュージック）
- 『RPG』（作曲：Fukase，発売：TOY'S FACTORY）
- 『恋するフォーチュンクッキー』（作曲：伊藤心太郎，発売：キングレコード）

すべて長調の楽曲でハ長調に移調して実験に用いた．

2.2 メロディ再生成手順

メロディの再生成は次の手順で行った．

(1) 対象となるメロディを人手で 8 分音符列に簡略化する．実験に使用する 4 曲を 8 分音符列に簡略化したものを **図 1** に示す．メロディは 8 小節あるので，64 個のノートナンバーの列 $M = n_1 n_2 \cdots n_{64}$ $(n_i = 1, 2, \cdots, 127)$ として表される．

(2) M を s 個 $(s = 1, 2, 4, 8)$ ごとに分割し，フレーズ辞書 $\mathcal{D}$ に登録する．たとえば $s = 2$ のとき，$\mathcal{D} = \{n_1 n_2, n_3 n_4, \cdots, n_{63} n_{64}\}$ であり，各 $n_i n_{i+1}$ をフレーズと呼ぶ．ただし，同一のフレーズは重複して登録しない．

(3) メロディ M に対して unigram 確率分布 $\{P(n_i n_{i+1})\}$ および bigram 確率分布 $\{P(n_i n_{i+1} | n_j n_{j+1})\}$ を計算する．

(4) フレーズ辞書 $\mathcal{D}$ を使って新しいメロディ M' を生成する．メロディの生成には遺伝的アルゴリズムを用いる．このとき，M と M' の unigram および bigram 確率分布間の KL ダイバージェンスの和が最小になるように M' を探索する（ただし $M = M'$ は不可）．

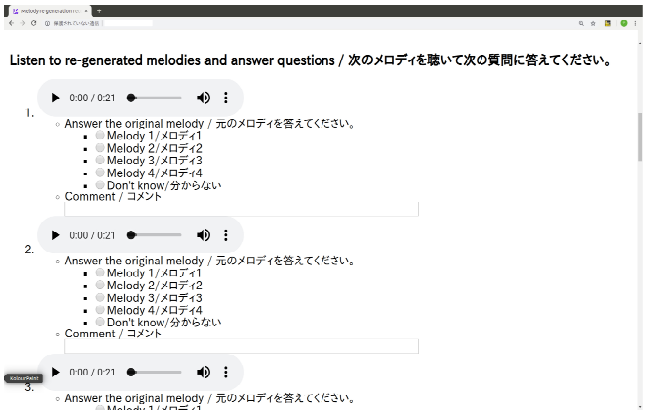

図 2 回答入力用画面

2.3 実験手順

被験者には，まず元のメロディ（8 分音符列に簡略化されたもの）を 2 回ずつ聴いてもらった．この際，メロディを覚えやすくするためにメモしても構わないことした．その後，前節の方法を使って再生成したメロディ 16 個（=メロディ 4 種 × フレーズサイズ 4 条件 [*3]）を 1 回ずつ聴いてもらい，その都度どのメロディが元になっているかを答えてもらった．回答は「メロディ 1」「メロディ 2」「メロディ 3」「メロディ 4」「分からない」から 1 つ選ぶ形式（**図 2**）とし，再生成されたメロディは必ず最初に聴いた 4 つのメロディのどれかを元に再生成したことは，あらかじめ明らかにした．

2.4 被験者

大学生 16 名である．

[*3] フレーズサイズとは，各フレーズの長さ（音符数）をさす．

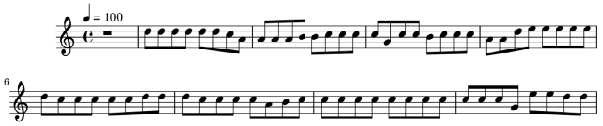

図 3 メロディ 1 の再生成結果（フレームサイズ=2 の場合）

3. 実験結果

再生成されたメロディを聴いたときの元のメロディの回答度数を**表 1** に示す．次の事柄を読み取ることができる．

3.1 全体的な傾向

フレーズサイズが 1 のときに正解者が 7∼8 名程度だったのに対し，フレーズサイズが 2 のときは 10∼11 名が正解していた．このことは，フレーズサイズが大きい方が元のメロディに気付きやすくなるという我々の仮説と一致する．

一方，フレーズサイズを 4 以上にした場合，フレーズサイズが 2 のときに比べて正解率が上がるもの，下がるもの，ほぼ横ばいのものに分かれた．これについては後で楽曲毎に考察を加える．

フレーズサイズが 2 でも高い正解率が得られたのは，メロディの再生成においてフレーズの bigram 確率分布を元の曲のそれに近づけていることが関係していると考えられる．すなわち，再生成されたメロディと元のメロディとで共通のフレーズ遷移があったために，4 音単位で元のメロディの特徴が反映されたものと推察される．

3.2 メロディ 1

他のメロディに見られないメロディ 1 の特徴は，フレーズサイズが 4 や 8 のときに正解者が少ない（16 名中 5 または 6 名）ことである．これは，元々のメロディが 2 分音符以上の長い音符が多く，これを 8 分音符列に変換してフレーズ辞書を作成したため，特にフレーズサイズが 4 の場合には「ドドドド」や「ララララ」のように特徴的でないフレーズが多かったためと思われる．

一方，フレーズサイズが 2 のときは，フレーズサイズが 4 や 8 のときに比べて正解者が多かった（16 人中 11 人）．これは，bigram 確率分布が近づくようにメロディを再生成する過程で，元のメロディと同じフレーズの遷移が連続し，結果的に 1 小節単位で原曲にかなり類似する箇所が複数生まれた（**図 3**）からだと考えられる．

3.3 メロディ 2

フレーズサイズが 4 のときに，16 人中 15 人が正解であった．これは，最初の 3 つのフレーズ（1 小節半）が元のメロディと完全に一致した（**図 4**）からと考えられる．メロディ 1 のフレーズサイズが 4 の場合についても，5 小節目から 6 小節目にかけて 1 小節半のメロディが元のメロディと一致したが，正解者は 16 人中 6 人と少なかった．このことから，元の 2 つのメロディの特徴が異なるので断言はできないが，出だしのフレーズが一致するかどうかが判断に影響している可能性が考えられる．

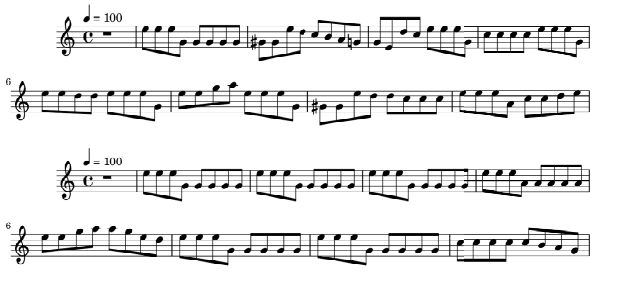

図 4 メロディの再生成結果（上：フレームサイズ=4 の場合，下：フレームサイズ=8 の場合）

一方，フレーズサイズが 8 のときの正解者は 16 人中 10 人であった．冒頭が元のメロディと一致するのはフレーズサイズが 4 のときと同じだが，フレーズサイズが 8 のときは [G♯ G♯ E D D C C C] というフレーズが出現しなかった．このフレーズはこのメロディの他のフレーズに比べて特徴的であり，この特徴的なフレーズが出現しなかったことが正解者数に関係していると考えられる．つまり，フレーズサイズが同じであっても，独自性の高いフレーズが出現するかどうかで，聴取者が元のメロディに気付くかどうかは変化すると考えられる．

フレーズサイズが 1 のときは，他のメロディと同様に半数程度の正解者であった．このメロディは 4 つのメロディの中で唯一 G♯ が出現する．このことはフレーズサイズが 1 の場合でも正解しやすくなる特徴であるが，被験者がこのことに気付いて正解したかどうかは今回の結果からは不明である．

3.4 メロディ 3

このメロディは，他に比べて音域が高い（他のメロディにおける出現するノートナンバーの最低値と最高値が [67, 77], [64, 81], [67, 76] に対して，このメロディは [72, 89]）ため，フレーズサイズに関わらず正解しやすいと予想されたが，他のメロディと比べて正解者数は多くなかった．これは，このメロディの大半が 2 度進行か 3 度進行を占め，特徴的なフレーズが見られなかったからだと推察される．

3.5 メロディ 4

このメロディでは，フレーズサイズによる正解者数の違いは比較的少なかった．元のメロディに同音連続が多く（例：[C C D D E E G A]），同音連続以外の箇所も 2 度または 3 度で進行することが多く，フレーズサイズを大きくしてもフレーズ内の音の変化が大きくならなかったためと

© 2018 Information Processing Society of Japan

思われる．

4. おわりに

本稿では，既存のメロディを数音符ずつに分割し，それを用いてメロディを再生成したときに，聴いた人が元のメロディに気付けるかどうかに関する実験を行った．実験の結果，1 音符ずつ分割して再生成するよりも，2 音以上ずつ分割して再生成した方が元のメロディに気付きやすいことが明らかになった．しかしながら，正解した被験者数はメロディによって異なり，元のメロディに独自性の高いフレーズが含まれているか，元のメロディとの一致箇所がメロディの先頭かどうかなどで，気付きやすさが異なる可能性が明らかになった．

今後は，この実験を継続するとともに，フレーズの独自性を TF-IDF などで定義するなど，より数理的な分析を進める．また，GTTM によるグルーピングのように音楽的に意味のある箇所で分割してフレーズを定義した場合に結果はどう変化するかについても検討していきたい．

謝辞

本研究は，JSPS 科研費 JP16K16180，JP16H01744, JP16KT0136, JP17H00749 の支援を受けた．

参考文献

[1] Dubnov, S. and Assayag, G.: Universal Prediction Applied to Music Generation with Style, *Mathematics and Music* (2002).
[2] Pachet, F.: The Continuator: Musical Interaction With Style, *Journal of New Music Research*, Vol. 32, No. 3, pp. 333–341 (2003).
[3] Fukayama, S., Nakatsuma, K., Sako, S., Nishimoto, T. and Sagayama, S.: Automatic Song Composition from the Lyrics exploiting Prosody of the Japanese Language, *Proc. Sound and Music Computing*, pp. 299–302 (2010).
[4] Pachet, F., Rou, P. and Ghedini, F.: Creativity through Style Manipulation: the Flow Machines Project, *Marconi Institute for Creativity Conference*, Vol. 80 (2013).
[5] Kitahara, T., Giraldo, S. and Ramírez, R.: JamSketch: Improvisation Support System with GA-based Melody Creation from User's Drawing, *Proc. of Int'l Symp. on Computer Music Multidisciplinary Research*, pp. 352–363 (2017).
[6] Yang, L.-C., Chou, S.-Y. and Yang, Y.-H.: MidiNet: A Convolutional Generative Adversarial Network for Symbolic-Domain Music Generation, *Proceedings of Int'l Soc. for Music Information Retrieval Conf.*, pp. 324–331 (2017).

© 2018 Information Processing Society of Japan