

Pressivo: 旋律の演奏表情を考慮した自動伴奏生成システム

宮田 佳奈^{1,a)} 酒向 慎司^{1,b)} 北村 正^{1,c)}

概要: 本稿では、ユーザが演奏表情を伴う旋律をコンピュータに入力すると、相応しい伴奏を自動生成し曲として出力するシステムを提案する。従来の自動伴奏生成システムは、ユーザが、曲のジャンルや雰囲気などを選択肢の中からイメージしながら選択する必要や、複雑なパラメータ調整の必要があり素人には扱い難いことなどが問題であった。本研究では、楽譜上同一の旋律であっても演奏方法により雰囲気が変化する点に着目し、演奏表情を伴った旋律に対し最も相応しい伴奏の生成を目指す。曲は音楽的同時性と音楽的連続性が自然となるよう作成されることが望ましい。和音は、確立されている音楽理論を基に決定をする。しかしリズムの決定においては、理論がほとんど確立されていない。そこで、言語モデルにより楽曲からリズムのパターンの同時性と連続性を学習し、リズム決定を行うことを検討する。

Pressivo: Automatic Accompaniment Generation System Considering Musical Expression of Melody

KANA MIYATA^{1,a)} SHINJI SAKO^{1,b)} TADASHI KITAMURA^{1,c)}

Abstract: This paper describes a system that automatically generate an appropriate accompaniment when you input melody with musical expression. Problems of previous systems are sometimes hard to use in amateur because it is necessary for the user to select the alternatives with images, and because of requiring a complex parameter adjustment. In this study, we focus on that the image of song is changed by musical expression of melody. The purpose of this study is to generate the most appropriate accompaniment to the melody when you input a melody with musical expression. It is desirable that the music is composed considering musical continuity and musical simultaneity. The chord is determined based on the music theory that has been established. However, in determining the rhythm, the theory has not been established little. Therefore, we consider rhythm decision with language model by learning musical continuity and simultaneity of rhythm patterns from music.

1. はじめに

作曲には専門知識や経験が必要となり、誰でも簡単に始められるものではない。しかし単旋律のみであればアイデアに依存することが多く、十分な専門知識を必要としない場合がある。逆に伴奏においては、主に和音であることから理論的であり、専門知識が必要となる。そのため、伴奏を自動生成することで作曲の支援をする商用ソフトや先行研究が多く存在している。

作曲を支援するための商用ソフトの一例として Band in a box[1] がある。これは、ジャンルなどの指定により曲の全てのパートが生成される機能や、一部のパートを入力するとその他のパートの和音やリズムを自動生成する機能などがあり、専門知識を必要とせず手軽に作曲が可能である。しかし、生成される楽曲がワンパターンになり得る場合や、用意されているジャンルの選択肢が膨大なため、自分の好みや旋律に相応しいものを選択するためには試行錯誤や、音楽的な知識が必要となる場合がある。

このような問題を解決するために、本研究では楽譜上同一の旋律であっても、演奏方法により曲の雰囲気が変化する点に着目する。例えば図 1 に示すように、同じ楽譜であっても、演奏 A のように全体的に大きな音量で、音の長

¹ 名古屋工業大学
Nagoya Institute of Technology

a) kana@mmsp.nitech.ac.jp

b) sako@mmsp.nitech.ac.jp

c) kitamura@mmsp.nitech.ac.jp

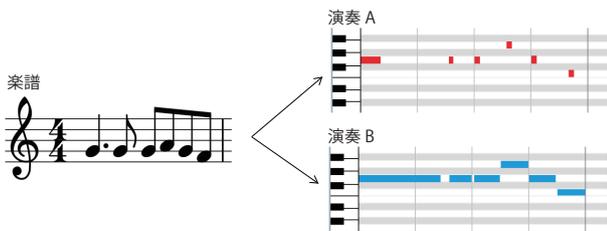


図 1 演奏表情による曲のイメージの変化の一例

Fig. 1 An example of changes in the image of the song by musical expression.

さを短く演奏した場合には活発な曲に感じる人が多いと考えられる。しかし演奏 B のように小さな音量で、音の長さを長く演奏した場合にはバラードのような静かな曲に感じられる場合がある。このように、演奏表情により曲のイメージを間接的に表現することを考える。これにより、さらに直感的な曲のイメージの入力が出来ると考えられる。

本システムでは、旋律の演奏表情に相応しい伴奏の和音とリズムを決定することを検討する。ここで、曲を作る際には局所的な音の響き（音楽的同時性）と、曲の継続的なまとまり（音楽的連続性）が自然となるよう考慮することが望ましい。そこで、音楽的同時性と連続性が自然となるような伴奏の和音とリズムの決定について検討してきた [2]。和音については和声学や作曲法などの音楽理論をルール化し、音楽的同時性と連続性の両方が自然となるような和音列を決定する。リズムについては音楽的同時性と連続性を考慮し、旋律に適切なリズム列を、予め用意したリズム木に基づいて選択することを検討してきた。本稿ではこれらの手法を用いることで、旋律の演奏表情を考慮した伴奏生成システムを提案する。

本報告では第 2 節で提案システムの概要について説明し、第 3 節でシステムの詳細について説明する。第 4 節ではシステムの実装について報告し、第 5 節で本研究のまとめと今後の展望について述べる。

2. 伴奏生成システムの概要

本研究では、コンピュータで演奏情報を扱う際に最も一般的となっている標準 MIDI データを使用する。MIDI データは電子楽器やコンピュータで再生可能なバイナリファイルであり、データ部には音色、音高（ノートナンバー）、音量（ベロシティ）、オンセット時刻（ノートオン）、オフセット時刻（ノートオフ）などの演奏情報が記載されている。MIDI データから得られる演奏表情として、音量、オンセット時刻、オフセット時刻が挙げられる。

伴奏生成の概要を図 2 に示す。ユーザーは入力として、演奏表情を伴う単旋律（MIDI データ）と任意の位置に付与したコード列を与える。システムは伴奏を自動生成し、伴奏と旋律をピアノロールで画面上に表示する。表示された

ピアノロールは画面上で編集することが可能であり、必要があればユーザは編集を行う。その後ユーザは生成楽曲を保存するか再生するかを選択し、保存する場合には MIDI データとして出力する。再生する場合、編集した旋律を用いて再び伴奏を生成する。

決定する伴奏の要素は、曲を構成する要素として重要である和音とリズムを扱う。演奏表情を含む旋律のあらゆる要素から、伴奏の要素を一度に決定することは難しい問題となるため、本研究では伴奏の和音の決定とリズムの決定を独立に考える。

3. 旋律の演奏表情を考慮した自動伴奏生成システム

3.1 伴奏の和音の決定

高さの異なる 2 音以上の音の積み重ね、またはその響きをコードという [3]。西洋音楽の中では、使用されるコードは 3 音構成が基本として考えられているが、ポピュラー音楽では 4 音構成を基本として考える場合が多い。そのため本研究では 4 音構成のコードを扱う。コード名からの具体的な音設定をヴォイスイングという。コード名からコードの構成音は判別可能であるが、音の配置や選択に複数の候補があり和音を一意に決定することは不可能である。適切な和音は、音楽的同時性と音楽的連続性がどちらも自然となるものが望ましい。

ヴォイスイングの先行研究として学習ベースの [4] が提案されている。学習ベースの手法はデータ収集の手間が問題点として挙げられるが、学習楽曲の傾向を反映した伴奏の生成が期待できる。これに対し、音楽理論を基にしたルールベースのヴォイスイング手法として [5] が提案されている。音楽理論は作曲において有用であるが、コンピュータに実装する際には、内容が具体的でない場合やルールの競合が問題として挙げられる。しかしポピュラー音楽のヴォイスイング理論はルールが明確であることが多く、アルゴリズム化が比較的容易である。

本研究では [5] を参考に、ルールベースで和音決定を行う。ポピュラー音楽全般でよく用いられる、テンション音を用いたテンションヴォイスイングにより、音楽的同時性と連続性を考慮した和音を決定する。テンション音とはコード構成音以外の音のことを示しており、コードごとに使用可能なテンション音に制限がある [6]。テンション音を用いることで旋律やコード構成音との不協和を避けるような音の選択や、省略音の決定が必要となってくるため、ヴォイスイングは複雑となる。しかしテンション音によって和音に緊張感を与え、響きを豊かにする。さらに、ヴォイスイングの候補も広がる。本研究では同時発音数が 4 つになるように適切なコード構成音やテンション音、省略音を決定する。

本研究では以下の手順で和音決定を行う。

(1) 与えられたコード名と旋律からコード名単位で、表 1

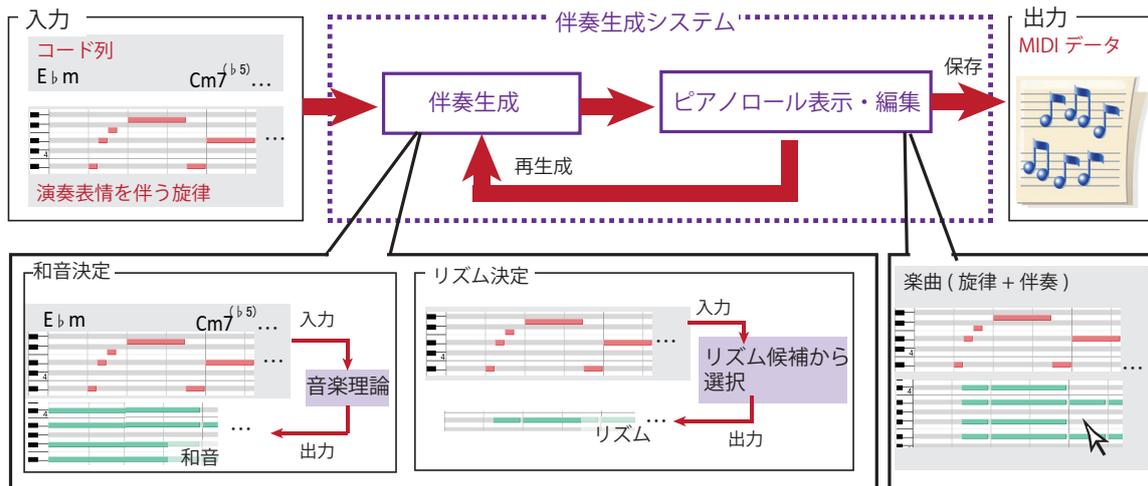


図 2 伴奏生成の概要

Fig. 2 Outline of accompaniment generation .

表 1 和音候補 (t : テンション音, r : ルート音, u : 3rd または 5th 音)

Table 1 Proposed chord. (t : tension tone, r : root tone, u : 3rd or 5th tone)

候補	1	2	3	4	5	6	7	8
top	t	t	7th	u	5th	5th	7th	3rd
	7th	u	t	t	7th	3rd	5th	5th
	u	7th	u	7th	3rd	7th	3rd	7th
root	r							

に従い全ての和音の候補を求める .

- (2) 和音候補間の全ての組み合わせに対し以下の式で遷移コストを求める . ここで, $V_i^{(j)}$ を i 番目のコードの j 番目の音, w_i は旋律の演奏表情により変化する重みである .

$$D(i) = \left(\sum_{j=1}^4 |V_i^{(j)} - V_{i-1}^{(j)}| \right) \times w_i$$

- (3) 遷移コストの総和を和音候補列のコストとし, コストが最小となる和音列を最適とし決定する .

3.2 伴奏のリズムの決定

和音が理論的に決定可能であることに対し, リズムは感覚的に決定されることが多く, ルールとして記述することが難しい . そのため, 予めリズムのパターンを候補として用意しておき, そこから旋律に適切なリズム列を選択することを検討する . その方法として, 楽曲から伴奏のリズムのパターン, 音楽的同時性と連続性を学習することで旋律に相応しいリズムの決定手法を提案する . 連続性を考慮するにはリズム間の繋がりが自然となる必要があり, 本研究では N -gram モデルを使用することを検討する . N -gram モデルは単語の生起が直前の $(N - 1)$ パターンまでに依存すると考えた言語モデルである . ここで遡る履歴の長さ

N を決定する必要があるが, 楽曲によって全体の長さやフレーズ長は変化するため, どのような楽曲にも柔軟に対応するには N を固定とすることは望ましくない . そこで, 動的に N の値が変化する可変長 N -gram モデルを使用することを考える .

ここで問題として, リズムのパターンの数が膨大になると, リズム間の遷移の組み合わせが急激に増加すること挙げられる . そもそもリズムには類似関係があり, 全てを独立に扱うべきではないと考えられる . そこでリズムの階層構造に着目し [7], 類似したリズムをリズム木によりクラスタリングする手法を提案する . リズム木はオンセット位置に着目したリズムの木構造であり, 木構造の上のリズムのパターンのオンセット位置を, 下のリズムのパターンは全て含んだ構造となっている . リズムのクラスタリングの例を図 3 に示す . 図中の黒い四角で塗られている音符は, 木構造の上のリズムパターンのオンセット位置を表している . クラスタリングしたリズムを, 可変長 N -gram モデルにクラス概念を取り入れた可変長クラス N -gram によってモデル化し, 旋律による制約を考慮した上で最尤推定により適切なリズム列を決定する .

さらに同時性を考慮するため, 実際の楽曲の特徴を基に, 旋律によって伴奏のリズムを制約する調和関数を設計し, N -gram 確率に重み付けをすることで自然なリズムを決定する手法を提案する . 具体的には, 実演奏を分析し, 以下の 2 つの特徴から調和関数を設計した .

- 旋律の音高差が大きいほど, 伴奏のオンセット数は多いものが使用されやすい
- 旋律の音量が大きいほど, 伴奏のオンセット数は多いものが使用されやすい

リズム列 $r_1^n = \{r_1 r_2 \dots r_n\}$ の生成確率は, リズムパターン r_i の属するクラスを c_i , 旋律を M , 旋律の音高差による制約を $W_1(r_i, M)$, 旋律の音量による制約を $W_2(r_i, M)$

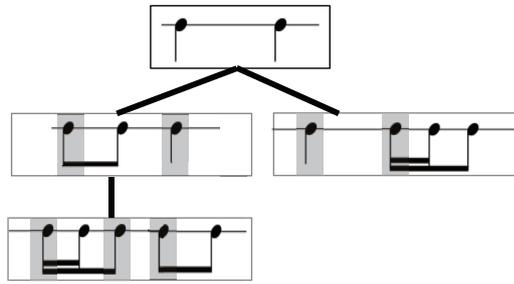


図3 リズム木構造によるクラスタリングの例
Fig. 3 An example of clustering by rhythm tree.

とすると、可変長クラス N -gram モデルを用いて以下のように表す。

$$P_T(r_1^n, M) = \prod_{i=1}^n P_{s^{i-1}}(r_i) \cdot W_1(r_i, M) \cdot W_2(r_i, M).$$

ここで s^j は確率的接尾辞木 T の根から $c_j c_{j-1} \dots c_1$ に対応する辺を辿ることで到達可能な最も適切な節点である。 $P_T(r_1^n, M)$ が最大となるリズム列 r_1^n を最適とし決定する。

3.3 生成結果の表示, 編集

生成された楽曲を楽譜とピアノロールで表示することにより、ユーザが生成結果を確認可能とした。また、表示されるピアノロールを画面上で直接編集可能にすることで、生成結果にユーザの希望を反映できるようにした。

なお、ピアノロールの編集では以下のことが可能である。

- ノートの移動 (オンセット時刻・音高の編集)
- ノートオフ時刻の編集 (音長の編集)
- ペロシティの編集
- ノートの追加
- ノートの削除

ユーザが編集した結果を MIDI データとして保存が可能である。また、旋律を編集し、編集後の旋律で伴奏を再生成することや、編集前と編集後で生成された伴奏を比較する機能も取り入れる予定である。

4. システムの実装

提案手法を用いて、自動伴奏生成システムを実装した。リズムの学習は RWC 研究用音楽データベースより、ジャズ音楽データベース [8] のトラック No.1~5 の 5 曲から、伴奏のリズムを抽出し使用した。生成される伴奏の音長は小節内の次の音のオンセット時刻までとし、ペロシティについては全て 100 とした。出力されたピアノロールを図 4 に示す。入力された旋律と生成された伴奏を表示しており、画面は左右にスクロールが可能である。ノートのペロシティが大きいほど赤く、小さいほど青く表示されている。入力された旋律の音量が大きいところでは、伴奏も比

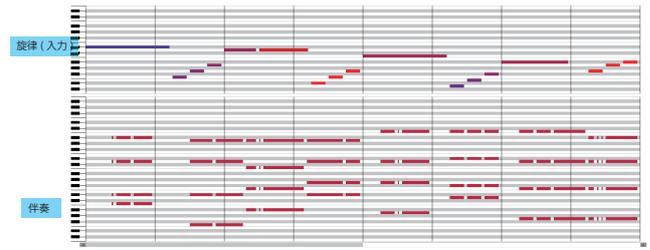


図4 生成結果の出力画面
Fig. 4 Output of composition result.

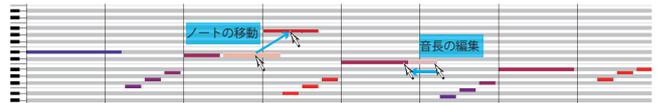


図5 ピアノロールの編集の一例
Fig. 5 An example of editing of piano roll.

較的オンセット数が多く活発な印象があり、逆に旋律の音量が小さいところでは伴奏のオンセット数は少なくなっている。このことから、旋律の演奏表情を反映した伴奏の生成が確認された。和音についても理論に添ったものが生成されることが確認された。また、この画面をマウスやキーボードで直接操作することによって、生成結果の編集が可能である (図 5)。

現在は旋律に対し伴奏を一意に決定し出力している。しかし今後、リズムの候補の N -best を表示しユーザに選択させることや、和音候補のパラメータを手動で調整可能にすることを検討している。また、現在は旋律と共にコード列を入力する必要があるが、コード自動推定機能を取り入れ、ユーザがコードを決定する手間をなくすことも検討していきたい。

5. おわりに

本研究では、旋律の演奏表情を考慮し、音楽的連続性と同時性が自然となるような伴奏を生成する手法を提案すると共に、MIDI 楽器から入力された実演奏データから伴奏を生成するシステムを実装した。和音については確立されている音楽理論を基に、和音候補と遷移コストを定めることで決定した。リズムについてはクラス N -gram モデルにより楽曲から連続性を学習し、旋律との調和関数を定めることで旋律の演奏表情と同時性を考慮したリズム決定法を提案した。さらに生成結果をピアノロールで画面に表示し、それをユーザが直接編集することを可能とした。

今後の課題として、伴奏候補をいくつか表示しユーザが選択できるようにすることや、パラメータを手動調整可能とするなどの機能の拡張を検討している。また、伴奏の和音とリズム以外の要素を決定することが課題として挙げられる。具体的には、和音の同時発音数の制御や伴奏への表情付けを行うことで、より豊かな伴奏の生成が期待できる。

参考文献

- [1] PG Music Inc . , Band-in-a-Box , 2012
- [2] 宮田佳奈, 酒向慎司, 北村正 : “旋律の演奏表情を考慮した言語モデルに基づく自動伴奏生成”, 情報処理学会 音楽情報科学研究会 研究報告 2013-MUS-100(3) , 1-6 , 2013 .
- [3] 北川祐 : “ポピュラー音楽理論”, リットーミュージック , 1994 .
- [4] 北原鉄朗, 勝占真規子, 片寄晴弘, 長田典子 : “ベイジアンネットワークを用いた自動コードヴォイシングシステム”, 情報処理学会論文誌 , Vol.50 , No.3 , pp.1067-1078 , 2009 .
- [5] J . Watanabe , K . Watanabe , N . Emura , M . Miura and M . Yanagida : “A system generating jazz-style chord sequences for solo piano” , Proc.ICMPC , pp.743-783 , 2008 .
- [6] 遠藤尚美 : “ピアニストのためのジャズ・ピアノ理論の基礎”, 自由現代社 , 2010 .
- [7] 中野拓帆, 深山覚, 小野順貴, 嵯峨山茂樹 : “リズム木構造仮説に基づく楽曲解析とリズム木学習”, 日本音響学会春季研究発表会講演集 , 2-8-13 , pp.977-978 , 2010 .
- [8] 後藤真孝, 橋口博樹, 西村 拓一, 岡隆一 : “RWC 研究用音楽データベース: クラシック音楽データベースとジャズ音楽データベース”, 情報処理学会 音楽情報科学研究会 研究報告 2002-MUS-44-5 , pp.25-32 , 2002 .