

ベイジアンネットワークを用いたコード・ヴォイシング推定システム

勝占 真規子[†], 北原 鉄朗^{† †}, 片寄 晴弘^{† †}, 長田 典子^{† †}

[†] 関西学院大学大学院 理工学研究科 情報科学専攻

^{††} 科学技術振興機構戦略的創造研究推進事業 CrestMuse プロジェクト

本研究では、ベイジアンネットワークを用いたコードネームからの自動ヴォイシングシステムについて述べる。ヴォイシングは音楽の同時性(響き)や音楽的連続性(流れ)を考慮しながらテンションや転回形を決定する必要がある、自動的に決定するのは容易ではない。この問題を解決するため、メロディやヴォイシング進行を考慮した事例学習型のコード・ヴォイシングモデルを構築する。メロディ音に音名ごとの占有度を定義することで音の衝突や不協和を避け、またヴォイシングを bottom, middle, top の3要素に分けることで前後の進行を考慮する枠組みを提案する。事例型システムに伴う自由度の拡大に対しては、モデルを細分化することで対処する。システムではこれらを組み込んだ1つのモデルから、尤もらしいヴォイシングを推測することが可能となる。実際にジャズ楽譜から学習したヴォイシング推定モデルを用いて、妥当なテンションや進行のある結果が出力されることを示した。

A Chord Voicing Reasoning System with Bayesian Network

Makiko Katsura[†], Tetsuro Kitahara^{† †}, Haruhiro Katayose^{† †}, and Noriko Nagata^{† †}

[†] Graduate School of Science and Technology, Kwansai Gakuin University

^{††} CrestMuse Project, CREST, JST

This paper describes automatic chord voicing system using the Bayesian network. Automatic chord voicing is not easy because it needs to decide tensions and inversions by taking into account interference with musical simultaneity(HIBIKI) and musical sequentiality(NAGARE). To solve this problem, we construct a chord voicing model based on the Bayesian network which taking into account interference with the melody and temporal smoothness of the voicings, This model includes melody-node which represents the degree of occupancy per pitch notation, previous and next voicing nodes which are separated 3 elements, and our system infers the most likely voicing from the model. Moreover, we divide the model per root tone of chord to solve the degree-of-freedom problem This modeling makes it possible to take into account both simultaneity and sequentiality at a single inference process. Experimental results of chord voicing for jazz musical pieces showed that our system generated chord voicings that has appropriate simultaneity and sequentiality.

1 はじめに

音楽情報処理の課題の1つに自動編曲・自動和声付けの研究がある。和声は音楽の3要素の1つであり、楽曲の持つ印象や雰囲気や決定する上で重要である。これまでに和声解析や和声付けの研究が多く報告されており、例えば川上らは、音楽的な知識を極力用いず数理的モデルにより自動和声付けを行っている¹⁾。また三浦らは、和声法からアプローチノート(経過音・刺繍音など)を考慮した和声付けシステム AMOR²⁾を提案した。平田らのパーピーブ³⁾では、ケーデンス単位で和声的文脈を考慮した自動和声付けを行い、隣接した和音の構成音間の接続関係(音高差)をもとに具体的な音を決定している。しかし、これらの研

究は主にコードネームなど和声表記を扱ったものであり、コードネームからの実際の音設定を研究主題としたものではなかった。

コードネームからの具体的な音設定をヴォイシングと呼ぶ。クラシック音楽では、通常ヴォイシングは作曲時に決定されているが、ポップスやジャズでは、メロディ譜(メロディとコードネームのみが書かれたシンプルな譜面)をもとに即興でインタラクティブにヴォイシングを決定することが日常的に行われる。

ヴォイシング次第で音楽の表情は様々に変化するため、前後の繋がりやメロディとの兼ね合い、また転回形やテンションの有無などを十分考慮する必要がある。特にジャズの場合、テンションの付

け方は本質的であり、楽器演奏初心者にとっては難しい課題となる。ヴォイスングを自動で行うシステムによって、楽器演奏初心者でも演奏を気軽に楽しめると期待できる。

ヴォイスングに着目した研究には、例えば江村ら⁴⁾が提案したジャズ理論をアルゴリズム化した手法がある。音楽理論を組み込む手法は、理論が十分に体系化され、かつ対象曲が理論とマッチする場合には有力である。しかし実際には、ジャズのように比較的体系化の進んだジャンルであっても、楽器によって適切なヴォイスングが変わってくるといった、音楽理論ではカバーできない領域がある。

本稿では、ヴォイスングのモデルを実際の演奏事例から学習し、これに基づいて適切なヴォイスングを探索するため、確率モデルの一種であるベイジアンネットワークを用いて、事例学習型の自動ヴォイスングシステムを構築する。ヴォイスングで考慮すべきメロディとの兼ね合い（音楽的同時性）や前後の音のつながり（音楽的連続性）といった複数の要因を、ベイジアンネットワークにおけるノード間の確率的な依存関係として表現し、任意のノードの最も尤もらしい状態を確率計算により推論する。これにより音楽を創る上で欠かせない響きや流れを同時に満たす適切なヴォイスングを行えると期待できる。

2 ヴォイスング決定における課題

2.1 問題設定

本研究では、メロディとコードネームをもとに、効果的なヴォイスングを行うことを目標とする。ジャンルはジャズ音楽、楽器は電子オルガンを対象とし、メロディを右手で、コードは左手で、ベースは左足で弾くことを前提とする。入力メロディ譜、出力は楽譜上の全コードネームに対応するヴォイスングとし、ここではコードネーム間に推測される経過和音は考慮しないものとする。

2.2 ヴォイスング課題

コードネームからヴォイスングを決定する課題において困難とされるのは、コードネームに対しヴォイスングが1つに決まらないことである。例えば、コードネーム“CMajor”が与えられた場合、基本構成音であるド、ミ、ソは容易に決まるが、テンション・省略音・転回形は複数の解候補がある。すなわち、以下のような課題が挙げられる。

- テンションの付与

効果的なテンションを付与するためには、メロディと同時に発音するヴォイスングを考慮

しなければならない。コードネームに付与可能なテンションは、ある程度ルールで決められているが、その中からどれを適用するかの決定は難しい。メロディ音との不協和な半音衝突は、不適切なハーモニーとなるので避ける必要がある。

- 省略音の選択

テンションが付与されると、それによるメロディ音やヴォイスング同士の音の兼ね合いを考慮するために、また、同じ音の重なりによる簡素なハーモニーを避けるため省略音を選択する。さらに、演奏にあたり指の本数や音域が限られるので、それに伴い音を省略する必要がある。

- 転回形の決定

音楽的文脈を考慮して転回形を決定するには、前後のヴォイスング進行を考える必要がある。転回形が利用される目的として、メロディを効果的に補う手法である「カウンターメロディ」がある。転回形をうまく利用し、伴奏の中にメロディの副旋律的要素を含ませることで音楽表現を豊かにすることが可能である。故意に低い音や高い音に移行して音楽に変化をつける手法もある。また、物理的にスムーズに手を動かし滑らかな遷移を保つため、なるべく近くの音に移動することも考えられる。このように、転回形は時系列の依存関係が強く、前後のヴォイスング遷移がどのようになっているのかを考慮しながら決定する。

Conklinの研究⁵⁾では、楽曲を同時性(simultaneity)と連続性(sequence)という2種類の見方を用いてパターン抽出を行なっている。ヴォイスングにおいても同様に、音楽的同時性(響き)と音楽的連続性(流れ)が重要である。ヴォイスング課題における音楽的同時性とは、同時に発音されるヴォイスングの響き、音楽的連続性とは、音楽の流れを感じさせる和音間の繋がりを意味する。基本的に、テンションの付与や構成音の省略は前者に属し、転回形の決定は後者に属する。しかし、実際に、テンションの付与を行うには音楽的同時性だけ考慮すればいいわけではない。テンションリゾルブという手法があるが、これはテンションによる音楽的文脈を考慮した緊張から解決へ導く手法であり、流れを考慮しながら決定しなければならない。同様に、転回形においても音楽的連続性が不可欠だと考えられる。すなわち、別々に決定する

のではなく、互いに考慮しあいながら決定することが望ましく、それを行える枠組みが必要となる。

3 ベイジアンネットワークによるモデル化

以上の課題を解決するために、本研究では、ベイジアンネットワークに音楽的同時性(響き)と音楽的連続性(流れ)の2点を考慮したモデルを採用する。

3.1 ベイジアンネットワーク

ベイジアンネットワークとはグラフィカルモデルの一種で、確率変数間の条件付き独立性を有向非循環グラフによって表現したものである。グラフのノードが確率変数を表し、ノード間のリンクが確率変数間の依存関係を表す。グラフ構造により確率モデルの構造が指定され、各ノードに割り当てられた条件付き確率表によりモデルのパラメータが表現される。

確率変数 X_1, X_2, \dots, X_n に対してベイジアンネットワークのグラフが定義されており、そのグラフにおける変数 X_i の親ノードの集合を $X_{\pi i}$ で表すとする。 $X_{\pi i}$ の変数が与えられた場合の X_i の条件付き確率を $P(X_i | X_{\pi i})$ とすると X_1, X_2, \dots, X_n の同時確率分布 $P(X_1, X_2, \dots, X_n)$ は以下のように表せる。

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | X_{\pi i})$$

全てのノードの状態が最尤になる場合を結果として出力するため、一意に決まらない情報に対して有効である。また、メロディ情報などのヴォイスングに関連する様々な要因を組み込みやすく、本課題に適したモデルを生成できる。

3.2 ヴォイスングモデルの構築

ベイジアンネットワークによるモデルの表現を決定するためには、ネットワークの各ノードにどのような事象を割り当て、さらにノード間の依存関係を決める必要がある。本研究では左手で演奏する音を決定する「左手ヴォイスングモデル」とベース音を決定する「ベースモデル」の2種類のモデルを「コードネーム」・「ヴォイスング」・「メロディ」の3要素の依存関係かた構築する。さらに、ヴォイスングの課題である局所的適切性と大域的適切性を取り入れる。

3.2.1 左手ヴォイスングモデル

本研究の提案する左手のためのヴォイスングモデルを図1に示す。各々のノードに対する状態は、

表1に示す通りである。

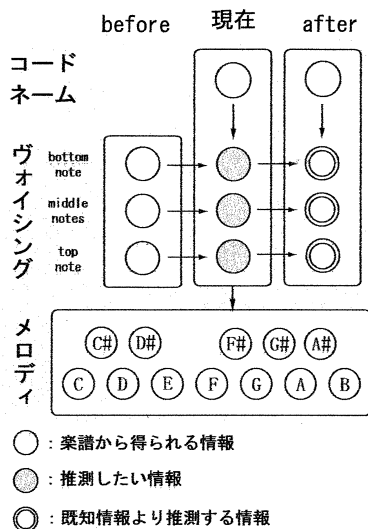


図1: 左手ヴォイスングモデル

表1: 図1のモデルにおけるノード名とその状態

ノード名	状態	
コードネーム	ルート音+コードの種類*	
ヴォイスング	top note	音名 (C~B)
	middle tones	音名 (C~B) の組み合わせ
	bottom note	音名 (C~B)
メロディ	占有度 (0と1を0.2刻みにした6段階)	

*コードの種類: augmented, diminished, dominant, major, major6th, major7th, minor, minor6th, minor7th, minor7th(b5)/half-diminished, minormajor7th, suspended4th, 以上12種類。

音楽的同時性を考慮するためには、ヴォイスングを推測するコードネーム(現在のコードネーム)の他に局所的に演奏されるメロディ情報を与える必要がある。コードネームは非常に数が多いためここでは、12種類で表現できるようあらかじめ簡略化しておく。メロディは音名ごとの占有度で表わす。占有度とは、コードが演奏される時間に対するメロディの12音名それぞれの長さの割合を意味する。オクターブの違いは区別せず、音名で分類する。コードネームが演奏される長さを1とし、メロディの各音名が全体のどれだけ占めているかの割合を示す。0.2刻みの5段階に0である場合を含めた6段階に離散化し表した。すなわち、各々のノードが各音名の占有度を持った確率変数を示している。音名ごとに占有度を考慮することで、メ

ロディ音にアボイドノートが含まれている場合であっても、全体の割合からどの程度重要な情報であるかを計ることができる。メロディを表す各音名のノードは、ヴォイスिंगの各音とそれぞれリンクしている。ヴォイスिंगのすべての音に依存関係を持たせており、メロディに合った予測が行えると考える。

音楽的連続性を考慮するため、前後の情報にはコードネームとヴォイスングを用いた。具体的に、現在のコードネームの後ろに位置するコードネームとすでに推測された過去のヴォイスングを考慮したモデルを構築した。

前後から個々のヴォイスング進行を表現するため、ヴォイスングを bottom tone, middle tones, top tone の3つの部分に分けて表現する。ヴォイスングを決定する上で、最低音・最高音は前後の音楽的進行を考慮される部分であり、分割することで進行がよりスムーズになると考える。例えば、音高の低い音順に書かれた和音 $CM_7 = \{C, E, G, B\}$ があるとする。この場合、bottom note は“C”, top note は“B”となり、middle notes は、“EG”という文字列で表現される。middle tones は音数が決められていないため柔軟に対応できる。

この3つのヴォイスング情報は、前後の同様に分けられたヴォイスングとリンクされている。直前でどのようにヴォイスングしたかという情報に依存して、現在のヴォイスングが決定される。同様に直後のヴォイスングとの依存関係も持たせる。ジャズ音楽では、演奏を縛るコード進行や具体的なコードネームは用いず、「モード」の概念を取り入れ、最低限のコードネームが指定されることが多い。つまり、コード進行では説明できない個々の音の連続性を重要視している。このモデルでは、直後に位置しているコードネームからあらかじめヴォイスングを予測し、さらにその予測結果をもとに、現在のコードネームを決定することで、コード進行に頼らない「モード」の概念を組み込んでいる。また、事例から学習していることで、さらにその効果は大きいと考えられる。

3.2.2 ベースモデル

提案するベースモデルを図2に示す。ベースモデルもヴォイスングと同様に、音楽的同時性と音楽的連続性を考慮したネットワークを構築した。簡単のため、ウォーキングベースのような同一コー

ドネーム中の進行は考慮しない。

音楽的同時性を考慮するため、現在のコードネームと同時に発音されるメロディの音名と依存関係を持たせた。ベースは比較的根音が演奏される場合が多いが、根音以外の音を演奏する場合に、メロディと同時に同音名を演奏されることを嫌う性質がある。また、音楽的連続性を考慮するため、ヴォイスングと同様に前後に演奏されるベースとリンクさせた。

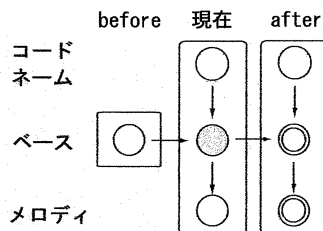


図2: ベースモデル

4 システムの実装

コード・ヴォイスング推定システムを、図1、図2に示したモデルを用いて構築した。モデルの学習データは、ジャズまたはジャズ風に編曲された電子オルガンの楽曲30曲を用いた。入力データには、コードネームが7thまでで表記されているメロディ譜を与えた。得られたコードネームに対して、楽曲の初めからモデルを適用し、最後のコードネームまでシフトさせ、1つずつヴォイスング推定を行った。

次に、モデルから推定された結果をもとにして詳細に和音を決定した。モデルの推定結果は、音名ごとの尤度が出力であるため、尤度をもとに音高を決定する必要がある。そこで、演奏可能音域を設定し、その範囲での音高を割り当てた。最初のヴォイスング決定はあらかじめ設定しておいた基準音から最も音高差の少ない音を最低音とし、他の音は、最低音が決まることで自動的に決定される。以降のヴォイスングも同じ手順で音高を決定した。演奏可能音域を超える場合は1オクターブ推移させその範囲内に収めた。

左手ヴォイスングの3要素の中で、同じ音が最尤であると推定された場合は、最低音、最高音、中間音の順で決定した、同じ音名の選択は禁止し、次に尤もらしいと判断した音を選択した。

4.1 結果と考察

本システムを用いた推定結果の音楽専門家による評価実験を行った。評価者は、メロディ譜を見ながらそれに対するヴォイスンク結果(左手ヴォイスンクの単独結果とベースと合わせた結果の2種類)を聴く。1コードネームに対して、我慢できない、合っていない、合っている、美しいの4段階で評価する。また、コードネーム間に“流れ”の有無を表記する。楽曲は全部で3曲、141コードを評価した。

その結果、左手ヴォイスンクのみの場合、約66%のコードを合っている、約5%を美しいと評価した。ベースと合わせた場合は、約68%、約13%と評価し、こちらの方が有意な結果であった。また、音楽的な流れを感じられる箇所は左手ヴォイスンクのみの場合全体の約26%、ベースを付与すると約44%であった。これらの結果から、本システムではある程度音楽的連続性の感じられる妥当なヴォイスンクが推定され、また左手ヴォイスンクとベースの両モデルを用いることで、より適切にヴォイスンクできるとわかった。しかし、両モデルを考慮しても約10%が我慢できないと評価された。これは、原因としてメロディ音の考慮方法(占有度)が適切でない可能性やヴォイスンクの詳細音を決定する際の制約の設け方などが考えられる。また、左手ヴォイスンクとベース間の依存関係がないため、推定結果のみ合わせて評価しても良い結果をうまなかったとも言える。

次に、結果例として楽曲「Misty」に割り当てられたヴォイスンクを図3に示す。1段目に与えたメロディとコードネーム、2段目が左手ヴォイスンク、3段目がベースの推定結果である。

左手ヴォイスンク結果は、明らかに不協和であるものは見受けられず、全体的に、妥当なヴォイスンクがされたものがほとんどであった。

テンションの付与に関しては、9thや13thが付与されたものが多く生成された。全62コード中、30コードにテンションが付与されていた。このモデルを構築する際に使用したデータベースの中でも、テンションが使用されていた事例は約4割を占め、ほぼ同等の割合であった。また、2つのテンションが同時に選択されている箇所もあった。4小節目のBbMajorには9thと13thの含まれるヴォイスンクが出力されている。理論書⁶⁾の中で、“4声以上で13thを用いるときは、9thを入れたほう

が響きが豊かになる”とされているヴォイスンクである。また、全てのヴォイスンクが3和音または4和音で構成されており、実際に左手で演奏可能な範囲で出力されたことから、適切な省略音を選択されたと考えられる。

前後のつながりに関しては、所々で半音による進行が見られた。2小節目のソ→ソ \sharp への進行や8小節目のF \sharp →F、15小節目のDm $_7$ →G $_7$ のヴォイスンクによるC→Bなどが挙げられる。ヴォイスンクを3要素に分けることで、うまくbottom toneやtop toneが作用したと考えられる。しかし、これらは2連続のヴォイスンクによる進行であり、3連続以上のヴォイスンクによる効果的な進行は見受けられなかった。また、24小節目のDm $_7$ →G $_7$ は、演奏可能音域(C3~A \sharp 4)を決定したため、半音で推移可能すると{B2, D3, F3}であるが、B2は音域外であるので1オクターブ上に移動して不自然な飛躍した進行になっている。

ベースに関しては、約90%がルート音を出力し妥当な結果を得た。ルート音以外には、2小節目のGm $_7$ や20小節目のF $_6$ 、31小節目のDb $_7$ のように妥当なものもあるが、17小節目のGm $_7$ のように、明らかに不適切な音も出力された。2・3小節目や26・27小節目に見られるように音楽的文脈を考慮したスムーズな進行も出力された。

これらのヴォイスンク結果から、各モデルの改善が必要であると言える。今回の左手ヴォイスンクモデルは、音楽的同時性に対して音楽的連続性があまり加味されていなかった。データベース作成に用いた楽譜に書かれている左手の演奏は、和音だけではない。そのため、前後のヴォイスンクが必ず存在するわけではなく、局所的なヴォイスンク情報よりも充分に得られなかったことが考えられる。また、直前と直後のヴォイスンクのみを考慮したモデルであるため、長期に渡る進行は予測しにくいと考えられる。前後だけでなく流れとして、まとまりのあるコード進行や音楽構造全体からの予測ができる枠組みが必要である。

ベースモデルに関しては、ベースの出力結果に不適切な音が推定されていることから、メロディに関する情報が少なく、モデルに反映されていない可能性がある。効果的なメロディ情報の組み込みも今後検討が必要であるとわかった。

図 3: 楽曲“Misty”でのヴォイシング結果

5 まとめと今後の展開

ベイジアンネットワークを用いて、事例学習型の自動ヴォイシング手法を提案した。推定結果から、テンションの付与や省略音の選択が行われた妥当なヴォイシングを得ることが確認された。音楽的同時性だけでなく、音楽的連続性も考慮したヴォイシングが行われることができることもわかった。ベースに関してもほぼ妥当なものが出力された。

しかしながら、左手とベースの双方を考慮したモデルについて検討する必要があることが示唆された。また、今回のモデルは1対1対応であるため1コード中での進行は考慮できなかった。コードネーム単位ではなく拍単位でのヴォイシングも考える必要がある。これが可能となれば、単調さを避けるために装飾的に加えられるクリシェや緊張音からの解決を導くテンション・リゾルブなどにも応用できると考える。

今後の展開として、音楽経験者や非経験者による評価実験も行い、システムの有効性を検証したいと考えている。さらにデータベースを補うための音楽理論のモデルへ組み込みやモデル・システムの改善へと繋げていきたい。

参考文献

- 川上 隆, 中井 満, 下平 博, 嵯峨山茂樹: 隠れマルコフモデルを用いた旋律への自動和声付け, 情報処理学会音楽情報科学研究会, No. 99-MUS-34, pp. 59-66 (2000).
- 三浦 雅展, 青山 容子, 谷口 光, 青井 昭博, 尾花 充, 柳田益造: ポップス系の旋律に対する和声付与システム: AMOR, 情報処理学会論文誌, Vol. 46, No. 5, pp. 1176-1187 (2005).
- 平田 圭二, 青柳 龍也: パーピープン: ジャズ和声を生成する創作支援ツール, 情報処理学会論文誌, Vol. 42, No. 3, pp. 633-641 (2001).
- 江村 伯夫, 三浦 雅展, 柳田益造: 与えられた旋律に対するコード・ネーム付与システムと、テンション・ノートを考慮したヴォイシング・システムの統合, 日本音響学会音楽音響研究会 (2007).
- Darrell Conklin, : Representation and discovery of vertical patterns in music, *Music and Artificial Intelligence: Lecture Notes in Artificial Intelligence*, Vol. 2445, pp. 32-42 (2002).
- 松田昌: 松田昌の音楽講座 ポピュラーアレンジの基礎知識, ヤマハ (1986).