

ルール処理と HMM を統合した和音認識器の実現

Realization of Chord Recognizer Integrating Rule Processing and HMM

横井 史也^{*1}

Fumiya Yokoi

平田 圭二^{*2}

Keiji Hirata

竹川 佳成^{*2}

Yoshinari Takegawa

^{*1}公立はこだて未来大学大学院

Graduate School of Future University Hakodate

^{*2}公立はこだて未来大学

Future University Hakodate

Work to create a code score needs a high level expertise of musical theory, so it is very daunting task for person having little knowledge. In this study, we aim to realize of chord recognizer for staff notation. In this paper, we propose a chord recognition model applying HMM regarding a solution inferred by rule based on general musical theory as an anchor.

1. はじめに

五線譜を基に人手でコード譜を作成する作業は高度な音楽的知識を要するため、音楽的知識の乏しい者にとっては困難な作業である。この問題に対して、五線譜、またはそれを計算機上で操作可能な形式に変換した記号データを対象に和音を自動認識する技術があれば、コード譜の自動生成が可能となる。そこで本研究では、楽譜記述言語のデファクトスタンダードである MusicXML ファイルを対象とした和音認識に取り組む。

記号を対象とした和音認識には次のような問題がある。和音の構成音が省略されているとき、音響信号からは倍音成分や残響から省略音を観測できることがあるが、記号では省略音を知ることはできない。そのため、和音名の決定に必要な根音や第3音が省略された場合、記号を対象とした和音認識では、楽譜に存在しない音を正しく補って考える必要がある。この問題に対し、正しく和音区間長を取ることで省略音を明らかにできる場合がある。これは、和音名の決定が和音区間長の決定と相互に依存するためである。また、正しい和音区間長の決定には、正しい和音の文脈理解が必要である。和音の文脈とは、和音列における個々の和音間の関係である。和音の文脈は楽譜の表層に見えない情報であるため、計算機によってそれを読み取することは一般に容易でない。以上より、本研究では和音の省略音推定問題における和音の文脈理解の課題を解く。また、一般的なジャズピアノの演奏において、和音が複数のテンションノートを含むとき、冗長な響きを避けるため構成音を省略する技法が多く用いられることから、本研究ではジャズを対象とする。

従来の和音認識研究では、認識手法として HMM(隠れマルコフモデル)[1][2][3]が多く用いられている。これは、和音名を隠れ状態、和音進行を状態遷移、様々な和音特徴量を出力記号とみなすことで、和音認識を HMM としてモデル化するものである。HMM では、和音の文脈形成を確率的な事象として捉え、各和音区間における音イベントの出現傾向を基にパラメータを学習することで、和音区間長の推定と和音名の推定を同時に行う。そのため、和音区間長と和音名の決定が相互に依存する問題に対処できると考えられる。

また、記号を対象とした和音認識の先行研究には、Rocher ら[4]の研究がある。Rocher らは、適切な和音区間長推定のため、時間分割手法としてホモリズム変換[5]を用いた。ホモリズム

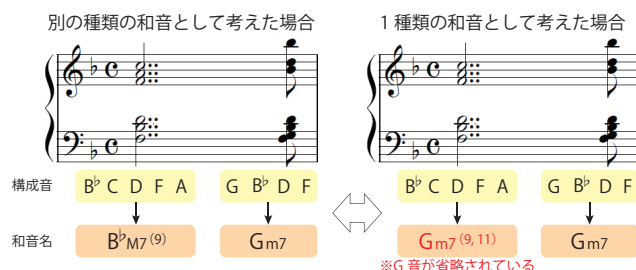


図 1: 和音区間の取り方によって名前が変わる和音

変換とは、全ての音イベントのオンセットとオフセットの時刻でセグメントを分割する手法である。Rocher らは、ホモリズム変換によって得られた全てのセグメントに対して和音名の候補を列挙し、各候補間の和音間距離[6]を求めることで、和音の文脈理解を行っている。この手法は、非和声音(旋律を表現する音イベント)に対しても和音名の候補を列挙するため、誤検出の数が膨大になってしまう問題がある。このことから、局所的な音イベントに対して和音名の候補を列挙する場合、着目した音イベントが和声音(和音を表現する音)かどうかを判別するための制約が必要であると考えられる。

以上より、本研究では制約により和声音であると判断されたセグメントに対して和音名候補を列挙するルール処理と、ルール処理が施された区間を準拠点として残りの区間を HMM で認識する和音認識手法の提案を目的とする。本稿では、セグメントが持つピッチクラスの数と音価をパラメータとし、パラメータの値に応じたルールの性能評価を行う。また、最適なパラメータを求める方法について考察する。

2. 省略音推定問題

この章では、1 章で説明した和音認識の問題を、実際のジャズピアノ譜[7]を例に挙げて説明する。

図 1 の楽譜は、和音区間の取り方を変えることによって省略音が明らかとなる例である。1 拍目の和音の構成音は C, D, F, A, Bb であり、4 拍目の和音の構成音は D, F, G, Bb である。これらの和音をそれぞれ別の種類の和音として考えた場合、1 拍目の和音には BbM7(9), 4 拍目の和音には Gm7 と名前付けすることができる。しかし、この小節全体を 1 種類の和音とし

連絡先: 横井史也, 公立はこだて未来大学大学院, 〒 041-8655 北海道函館市亀田中野町 116 番地 2, 0138-34-6462, g2114036@fun.ac.jp

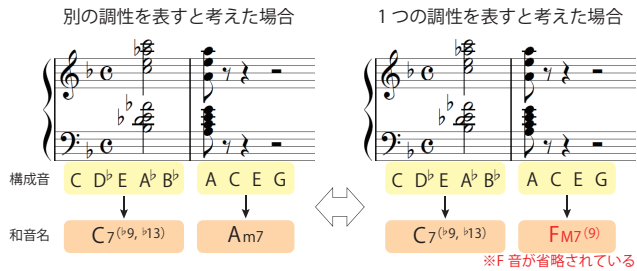


図 2: 調区間の取り方によって名前が変わる和音

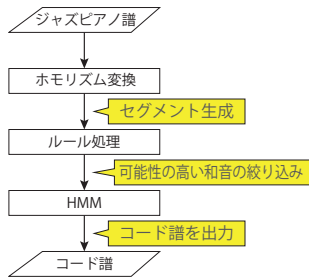


図 3: 提案手法のフローチャート

て考えた場合、1 拍目の和音は $G_{m7(9,11)}$ と名前付けすることができる。これにより、1 拍目の和音は根音の G が省略された和音であることがわかる。

図 2 の楽譜は、適切な調の区間を考慮することによって省略音が明らかとなる例である。1 小節目の和音の構成音は C, D^b, E, A^b, B^b であり、2 小節目の和音の構成音は C, E, G, A である。これらの和音がそれぞれ別の調性を表すと考えた場合、1 小節目の和音には $C7(b9, \#13)$ 、2 小節目の和音には A_{m7} と名前付けすることができる。しかし、これらの和音が 1 つの調性を表すと考えた場合、この和音進行はドミナントモーションであると解釈されるべきであり、2 小節目の和音は $F_{m7(9)}$ と名前付けされる。これにより、2 小節目の和音は根音の F が省略された和音であることがわかる。

このように、和音の文脈を正しく理解し、適切な和音区間長を推定することで、構成音が省略された和音に対して正しい省略音を知ることができる。しかし、和音の文脈は楽譜の表層に見えない情報であるため、計算機によってそれを読み取ることは一般に容易でない。

3. 提案手法

本章では、一般的な音楽理論に基づくルール処理と HMM を統合した和音認識モデルについて述べる。本手法のフローチャートを図 3 に示す。

3.1 時間分割

本研究では、分析窓の長さを設定するための手法としてホモリズム変換を用いる。ホモリズム変換とは、全ての音イベントのオンセットとオフセットの時刻でセグメントを分割する手法である。例えば、音イベント A が鳴っている間に音イベント B が発声した場合、音イベント A は音イベント B の発声時刻で分割され、そこに境界が生じる。また、音イベント B が鳴っている間に音イベント A が終了した場合、音イベント B は音イベント A の終了時刻で分割され、そこに境界が生じる。この

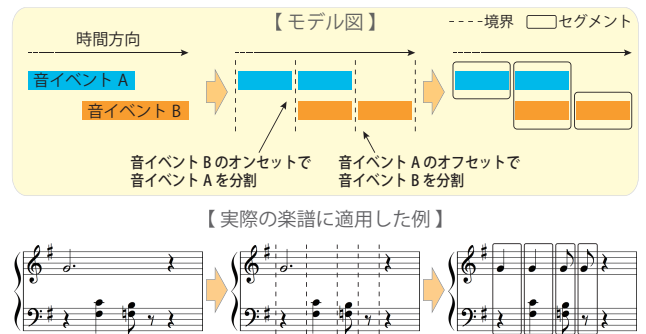


図 4: セグメントの生成過程

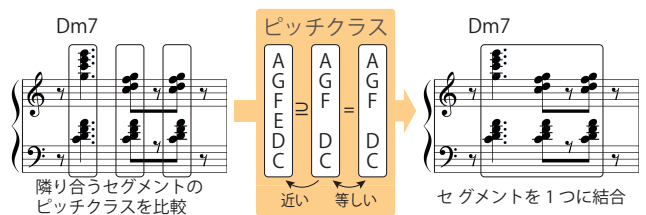


図 5: 隣り合うセグメントの結合

ような操作を繰り返すことにより、オンセットとオフセットが同一である音イベントの集合が複数個生成される (図 4)。これらを個々のセグメントと定義する。また本研究では、同じ音が鳴っている間は和音名が遷移しないという仮定のもと、隣り合うセグメントの持つピッチクラスが等しいか近い場合に、それらのセグメントを結合するようホモリズム変換に改良を加える。ここでいうピッチクラスが近いとは、片方のセグメントのピッチクラスがもう一方のセグメントのピッチクラスの部分集合という意味である (図 5)。この改良により、冗長に分割されたセグメントが結合され、不適切な時刻での和音境界の発生を低減することが期待できる。

3.2 ルール処理

HMM における探索ノードを枝刈りするため、一般的な音楽理論に基づくルール処理を行う。ルール処理とは、個々のセグメントの構成音と、和音の雛形 (和音記号ごとに含むことのできる音度を事前に定義したもの) のパターンマッチである。具体的には、まず着目したセグメントのピッチクラスを基に、12 通りの根音 (1 オクターブに含まれる全ての音) のそれぞれに対して音度を求める。次に、各根音ごとに求められた音度の組と和音の雛形のパターンマッチを行う。パターンマッチの結果、捕捉された全ての和音名を解の候補とする (図 6)。

また、このルールは和音の構成音を基にして解を得るため、非和声音に対してルールを適用した場合、高い効果が期待できないと考えられる。そのため、和声音に対してのみルールが適用されることが望ましい。そこで、和声音と非和声音を判別するための制約を設け、制約を満たすセグメントに対してルールを適用する。パラメータは、個々のセグメントに含まれるピッチクラスの数と音価の 2 種類である。これらのパラメータは可変であり、楽曲に応じたチューニングを可能とする。

3.3 HMM

本研究では、和音認識手法として和音区間長と和音名を同時に推定することのできる HMM を用いる。使用する和音特徴量

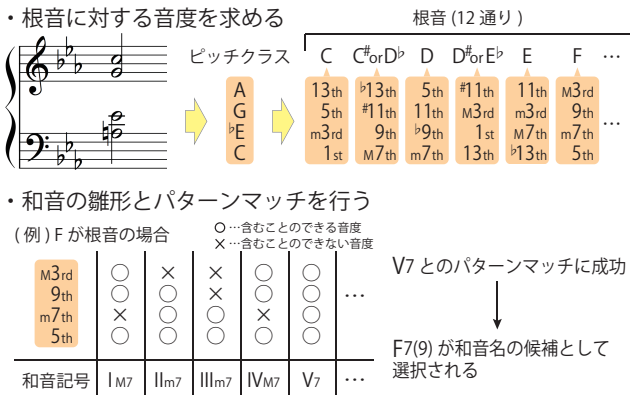


図 6: セグメントの構成音と和音の雛形のパターンマッチ

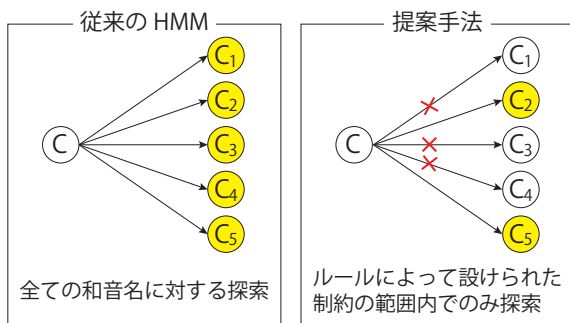


図 7: 従来の HMM と提案手法における探索方法

にはクロマベクトル [9] を用いる。本来クロマベクトルは音響信号から抽出される特徴量であり、ある時間長におけるパワースペクトルをピッチクラスごとに足し合わせることで得られる 12 次元のヒストグラムである。本研究では、パワースペクトルを音価に対応させることで、五線譜に対して同様のアプローチを採用する。出力確率には、学習データから得られる平均的なクロマベクトルの分布が単一正規分布に従うという仮定のもとで、和音名ごとに得られた単一正規分布を用いる。認識時には、観測記号から抽出したクロマベクトルに対する各和音名の単一正規分布の尤度により評価を行い、ビタビ探索で最適な和音名列を求める。

3.4 ルール処理と HMM の統合

ルール処理によって解候補の絞り込みが行われた区間を準拠点として、HMM を適用する和音認識モデルを提案する。生成モデルを作る際、ルールによる制約を満たす範囲内でのみ探索を行うことによって、誤検出を低減し、処理の効率化を図る。これによって和声音の省略などのジャズ特有の和音表現に起因する和音名同定結果の曖昧性を解決できると考えられる。従来の HMM と提案手法における探索方法の違いを図 7 に示す。

4. 評価実験

本手法におけるルール処理の有用性を検証するため評価実験を行った。以下では実験の詳細な方法を述べ、実験結果を基に最適なパラメータを求める方法について考察する。

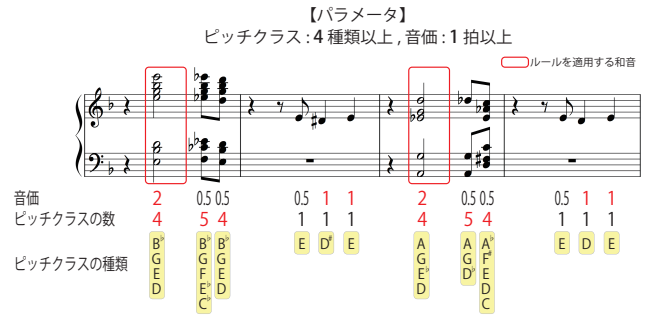


図 8: ルールを適用する和音の例

4.1 方法

入力データは、"Jazz Piano Collection Oscar Peterson" [7] に収録された楽曲 11 曲を手で MusicXML に変換したものを用いた。正解データは、入力データで使った楽譜に記載された和音名とした。正解データの小節数は 1064 小節、和音数は 1017 和音であった。パラメータの値は、セグメントに含まれるピッチクラスの数と音価を一定の範囲内で段階的に設定可能とした (表 1)。設定されたパラメータの値以上のピッチクラスと音価を持つセグメントに対してルールを適用した (図 8)。

パラメータ	取りうる値の範囲	値を変化させる間隔
ピッチクラス	2 ~ 8 種類	1 種類
音価	1 ~ 4 拍	0.5 拍

表 1: パラメータの設定方法

ルールの性能は、出力された和音名の正確性と網羅性の観点で評価を行った。

$$\text{適合率} = \frac{\text{出力された正しい和音名の数}}{\text{出力された和音名の総数}} \times 100(\%) \quad (1)$$

$$\text{再現率} = \frac{\text{出力された正しい和音名の数}}{\text{正解データの和音名の総数}} \times 100(\%) \quad (2)$$

ルールの適用範囲は、入力された楽曲長に対するルールが適用された区間長の割合で評価を行った。

$$\text{適用率} = \frac{\text{ルールが適用された区間長}}{\text{入力楽曲長}} \times 100(\%) \quad (3)$$

以上の評価方法で、ルールの性能と適用範囲を評価した。

4.2 結果と考察

図 9 に、各パラメータに応じた適合率、再現率、適用率のグラフを示す。図 9(a) より、適合率はピッチクラスの数が多いほど値が高くなる傾向があった。しかし、ピッチクラスが 7 種類以上になると、値が急激に低下することがわかった。適合率の値が最も高かったのはピッチクラスが 6 種類、音価が 2 拍のときであり、25%であった。これは、1 つのセグメントに対して平均 4 つの和音名候補が列挙されたことを意味する。また、ピッチクラスが 7 種類以上の場合を除いて、値が最も低かったのはピッチクラスが 2 種類、音価が 2 拍のときであり、12%であった。

図 9(b) より、適合率がピッチクラスの数に応じて値が変化したのに対し、再現率は音価の変化による値の変化が顕著であった。適合率と同様、再現率もピッチクラスの数が 7 種類以上に

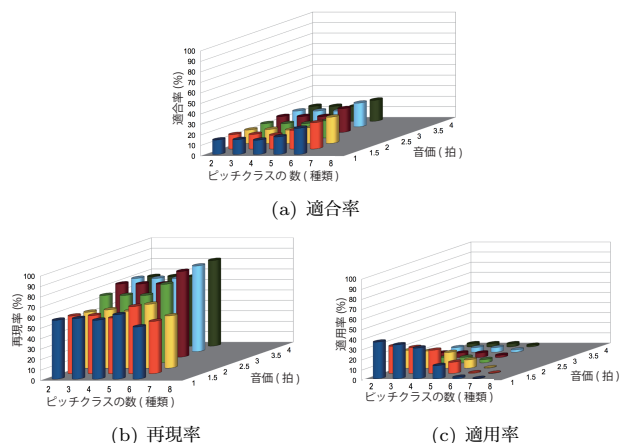


図 9: パラメータに応じたルール性能と適用範囲

なると値が急激に低下した。再現率の値が最も高かったのはピッチクラスが5種類、音価が4拍のときであり、84%であった。また、ピッチクラスが7種類以上の場合を除いて、値が最も低かったのはピッチクラスが4種類、音価が1.5拍のときであり、53%であった。このことから、再現率は適合率に比べ、パラメータによる値の変化が大きかったことがわかった。

図9(c)より、適用率はピッチクラスの数と音価の値が小さく設定されるほど値が向上しており、ルールの性能とは逆の傾向を示した。これは、設定された値以上のピッチクラスの数、または音価を持つセグメントに対してルールを適用しているため、2つのパラメータの値が小さくなればなるほど、該当するセグメントが増えるためであると考えられる。適用率の値が最も高かったのは、ピッチクラスが2種類、音価が1拍のときであり、35%であった。これは、楽曲の約3分の1の区間にルールが適用されたことを意味する。以上より、2つのパラメータの値が小さく設定されるほど適用範囲が広くなりルールの性能が低下し、また、2つのパラメータの値が大きくなり設定されるほど適用範囲が狭くなりルールの性能が向上することがわかった。このことから、ルールの性能と適用範囲がトレードオフであることが示された。

本手法はHMMによる和音認識の際、ルールの制約を満たす範囲でのみ探索を行うため、ルール処理の段階で正しい和音名が拾い上げられなかった区間は必ず誤認識となってしまう。そのため、ルールの性能においては、適合率より再現率の値を優先することが望ましいと考えられる。パラメータによる値の変化は、適合率が再現率よりも小さいため、再現率の値を優先したパラメータ設定がなされても、誤検出が大量に増えることはないと考えられる。

今後ルール処理とHMMを統合していく上で、再現率の低下は和音認識率に最も悪影響を及ぼす。そのため、ルールの性能と適用範囲のトレードオフの問題に対しては、再現率の値に基準値を設け、再現率が基準値を上回るパラメータの中で、適用率が最大となる組み合わせを選択する方法が望ましいと考えられる。例えば、再現率の基準値を80%とした場合、再現率が基準値以上となるパラメータの組み合わせは表2の通りである。表2では、ピッチクラスが5種類で音価が3拍の組み合わせが最も適用率が高いことがわかる。よってこの場合、最適なパラメータはピッチクラスが5種類、音価が3拍であれば良いと考えられる。

再現率	ピッチクラス	音価	適用率
84%	5種類	4拍	1.5%
82%	5種類	3.5拍	2.2%
81%	5種類	3拍	2.4%

表 2: 再現率が80%以上となる時のパラメータと適用率

5. まとめと今後の課題

我々はこれまで、ピッチクラスの数と音価に関する制約を満たすセグメントに対して和音名の候補を列挙するルール処理を実装し、評価した。結果として本手法のルール処理では、列挙された和音名の候補には80%以上の確率で正解が含まれること、また、1つのセグメントに対して平均4つの和音名の候補が列挙されることがわかった。このことから、本手法におけるルール処理は、HMMにおける探索ノードの枝刈り手法として有用であると結論する。今後は、HMMを実装しルール処理と統合する。また、コーパスデータの充実を今後の課題とする。

参考文献

- [1] A. Sheh, D.P. Ellis, Chord Segmentation and Recognition using EM-Trained Hidden Markov Models, In Proc. of the ICASSP, pp.183-189 (2003).
- [2] H. Papadopoulos, G. Peeters, Large-Scale Study of Chord Estimation Algorithms Based on Chroma Representation and HMM, Content-Based Multimedia Indexing, pp.53-60 (2007).
- [3] 須見康平, 糸山克寿, 吉井和佳, 駒谷和範, 尾形哲也, 奥乃博, ベース音高と和音特徴の統合に基づく和音系列認識, 情報処理学会論文誌, Vol.52, No.4, pp.1803-1812 (2011).
- [4] T. Rocher, M. Robine, P. Hanna, R. Strandh, Dynamic Chord Analysis for Symbolic Music, ICMC (2009).
- [5] P. Hanna, M. Robine, P. Ferraro, J. Allali, Improvements of Alignment Algorithms for Polyphonic Music Retrieval, In Proc. of International Symposium on Computer Music Modeling and Retrieval CMMR08, Copenhagen, Denmark, pp.244-251 (2008).
- [6] F. Lerdahl, Tonal Pitch Space. Oxford University Press (2001).
- [7] 草野昌一. Jazz Piano Collection Oscar Peterson, SHINKO MUSIC PUB. CO., LTD.(1995).
- [8] 小山大宣. JAZZ THEORY WORKSHOP JAZZ 理論講座初級編/ 中・上級編, 武蔵野音楽学院出版部 (1980).
- [9] T. Fujishima, Realtime Chord Recognition of Musical Sound: A System Using Common Lisp Music, ICMC1999, pp.464-467 (1999).