

和音区間検出と和音名同定の相互依存性を解決する和音認識手法

吉岡 拓也[†] 北原 鉄朗[†] 駒谷 和範[†] 尾形 哲也[†] 奥乃 博[†]

[†]京都大学大学院情報学研究科知能情報学専攻

{takuya, kitahara, komatani, ogata, okuno}@kuis.kyoto-u.ac.jp

あらまし 本稿では、音楽音響信号を対象とした和音進行と調の認識手法について述べる。和音区間検出、和音名同定及び調認識の間に存在する相互依存関係により、これらの処理を同時的に行う処理モデルが求められる。このために、我々は、複数生成された和音進行と調に関する評価値付きの仮説の中から、最も評価値が大きい仮説を認識結果として出力する手法を提案する。仮説の評価値は、従来研究で主に用いられてきた音響的特徴に加えて、和音遷移パターン、ベース音にもとづいて計算する。評価実験の結果、10曲のポピュラー音楽のCDから抽出した音響信号に対して、72.1%の認識率を達成した。

Automatic Chord Recognition Considering Mutual Dependency of Chord-Boundary Detection and Chord-Symbol Identification

Takuya Yoshioka[†] Tetsuro Kitahara[†] Kazunori Komatani[†]
Tetsuya Ogata[†] Hiroshi G. Okuno[†]

[†]Dept. of Intelligence Science and Technology,
Graduate School of Informatics, Kyoto University

Abstract This paper describes a method that recognizes musical chords and keys from musical audio signals. The mutual dependency among chord-boundary detection, chord-symbol identification, and key identification requires a model where these three are concurrently processed. We propose a method that generates various hypotheses of chord progressions and keys, and selects the most plausible one out of them as the recognition result. The certainty of a hypothesis is evaluated based not only on acoustic features used in most previous methods but also on chord transition patterns and bass sounds. Experimental results on 10 popular music songs show 72.1% average accuracy.

1. はじめに

近年のデジタル音楽配信サービスの急速な普及に伴って、音楽的内容の類似性などにもとづく高度な楽曲検索技術を開発するニーズが高まってきている。こうした高度な楽曲検索を効率的に実現するためには、音楽コンテンツに対して和音やメロディーなど高次の音楽的要素をタグ付けすることが有効であると考えられる。実際、マルチメディアコンテンツにタグ付けする枠組みを提供する標準規格 MPEG-7¹⁾が制定され、活発な研究が行われている^{2)~4)}。ただし、MPEG-7ではタグ付けの方法について一切規定していない。大量の音楽コンテンツに対して音楽的要素を手手で付与すること

は膨大なコストを要するだけでなく、アナタータ間でタグ記述内容の均質性が欠如するという問題がある。この問題を回避するためには、音楽音響信号から自動的に音楽的要素を抽出して音楽コンテンツにタグ付けする技術を開発しなければならない。

本研究では、音楽的要素として和音進行に着目する。和音進行とは、和音名とその和音が支配する時間(和音区間)の時系列である。和音進行に関するタグは、高度な楽曲検索を実現する上で重要である。なぜなら、特にポピュラー音楽では、和音進行は楽曲の伴奏の特徴を十分な粒度で表す簡潔な記号列であり、楽曲の雰囲気や気分を決定付ける重要な音楽的要素だからである。こ

表 1 本稿で扱う和音の種類

和音の種類	構成音
Major triad	主音, 長 3 度, 完全 5 度
Minor triad	主音, 短 3 度, 完全 5 度
Augmented triad	主音, 長 3 度, 増 5 度
Diminished triad	主音, 短 3 度, 減 5 度

のような観点から, 本研究ではポピュラー音楽の音楽音響信号を対象とした和音進行の認識問題に取り組む.

和音進行認識における主要な課題は, 和音区間検出と和音名同定の相互依存性の問題である. 和音名同定に先だてて和音区間を求めることができれば, 各和音区間の音響信号断片に対して和音名を同定することにより和音進行認識が達成される. しかし, 和音区間の尤もらしさはその区間を対象に和音名を同定した結果から得られるため, 和音名を同定する前に和音区間を正しく確定させることは困難である.

本研究では, 和音区間と和音名を同時に認識することによりこの相互依存問題を解決する. 本手法では, 様々な和音区間と和音名の可能性を同時に追跡するために, 和音区間と和音名の組み合わせに関する評価値付きの仮説を生成し, 最終的に評価値が最大の仮説を認識結果として出力する. 仮説の評価値を求めるために, 従来和音進行認識の手がかりとして主に用いられてきた音響的特徴に加えて, 和音遷移パターン (実楽曲で頻繁に使用される和音名の遷移の集合), ベース音の 3 つの手がかりを用いる. ただし, 適用すべき和音遷移パターンは調に依存するため, 本研究では和音進行と同時に調も認識する.

以下, 2 章で, 本研究で取り組む和音進行認識における相互依存問題について説明する. 3 章で, 本手法について詳細に説明する. 4 章では, 本手法の有効性を評価実験により示す. 5 章で, 本稿の結論を述べる.

2. 和音進行認識における相互依存問題

本章では 2.1 節で, 本研究で取り組む和音進行認識の問題設定を示した後, 2.2 節で, 和音進行認識における相互依存問題について説明する. そして, 2.3 節で, 本研究で提案するこの問題を解決するためのアプローチを示す. 2.4 節では, 相互依存問題の観点から関連研究の問題点を指摘する.

2.1 問題設定

本稿で扱う和音進行認識問題を, 音楽音響信号を入力して, その楽曲の和音進行 $c_1 c_2 \dots c_n$ と調 k を認識

入力信号から和音進行を得る処理は, 和音同定, 和音認識などと呼ばれることが多い. 本稿ではこの処理を, 与えられた和音区間で和音名を求める処理と明確に区別するために, 和音進行認識と呼ぶことにする. また, 和音名同定という用語を, 与えられた和音区間で和音名を求める処理をさして用いる.

する処理であると定義する. ただし, 本稿では, 入力楽曲の調は転調しないものと仮定する. 和音 c_i は, 和音名 cs , 開始時刻 b , 終了時刻 e からなる.

$$c_i = (cs, b, e)$$

区間 $[b, e]$ を和音 c_i の和音区間と呼ぶ. 和音名 cs は, 根音名 $root$ と和音の種類 $style$ の組である.

$$cs = (root, style)$$

$$root \in \{C, C\#, \dots, B\}$$

$$style \in ST$$

本稿で扱う和音の種類 ST を表 1 に示す. 調 k は, 主音名 $tonic$ と長調か短調かを表す $mode$ の組である.

$$k = (tonic, mode)$$

$$tonic \in \{C, C\#, \dots, B\}$$

$$mode \in \{Major, Minor\}$$

2.2 和音進行認識における相互依存問題

和音進行認識を構成する和音区間検出, 和音名同定, 調認識の各処理の間には, 次に述べる 2 種類の相互依存関係が存在する. このため, これらを同時に実行する処理の枠組みが必要である.

(1) 和音区間検出と和音名同定の相互依存関係

和音名を同定するためには, 同定対象となる区間が与えられなければならない. 一方, ある和音区間の尤もらしさはその区間を対象として和音名を同定することで求められるため, 和音名同定の前に和音区間を正しく確定させることは困難である.

(2) 和音名同定と調認識の相互依存関係

調は和音名系列の尤もらしさを求めるための重要な手がかりである. 一方, 和音名系列は調を求めるための主要な手がかりである. 例えば, 八長調の文脈では, G-C の和音名系列の尤もらしさが増すが, 一方, G-C の出現が八長調の尤もらしさを増す.

2.3 従来研究

音楽音響信号を対象とした和音進行認識に関する従来研究^{5)~11)}では, 2.2 節で述べた相互依存問題は解決されてこなかった.

青野ら⁵⁾や Nawab⁶⁾らは, 単一楽器による和音演奏の音響信号のみを対象としており, 実楽曲を扱っていなかった. 柏野ら⁷⁾や Su ら⁸⁾は, 和音区間があらかじめ既知であることを仮定していた. Fujishima⁹⁾は, 時間方向に局所的なスペクトルの変化から和音変化を検出する手法を提案し, 単一楽器の演奏による実楽曲に対して正しく和音進行を認識できたことを報告している. しかし, この手法を歌唱やドラム音など非調波成分を多く含む楽曲に拡張することは容易でない. なぜ

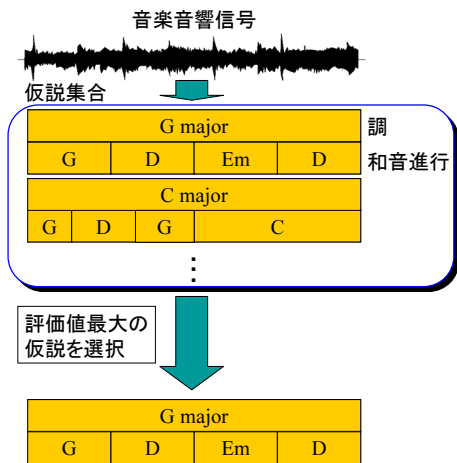


図 1 和音区間，和音名，調の同時的認識

なら，これらの音は和音変化の有無に関わらずスペクトルの大きな変化を生じさせるため，この手法を直接適用すると，和音変化を誤検出しやすいからである．山田ら¹⁰⁾や Sheh ら¹¹⁾は，和音区間を検出せず，和音区間の最小単位(四分音符や 100ms など)を仮定して，最小単位区間ごとに和音名を同定することにより和音進行認識する手法を提案した．しかし，こうした短い時間間隔に含まれる音響信号中では，和声音に比べて非和声音やドラム音のパワーが優勢になりやすいため，正しく和音名を同定することが困難である．

このように，従来研究では和音区間検出と和音名同定を分離して行っており，これらの間の相互依存関係の問題は従来から指摘されていた¹²⁾にも関わらず解決されていなかった．このため，実世界の複雑な音響信号から高精度に和音進行を認識できなかった．

2.4 仮説探索にもとづく相互依存問題の解決

本研究では，2.2 節で述べた相互依存問題を解決するために，和音区間，和音名及び調を同時に認識する．このために，和音進行と調の組み合わせに関する仮説を生成し，仮説の尤もらしさを表す評価値を与える．そして，生成された仮説の中から評価値が最も大きい仮説を認識結果として出力する(図 1)．

入力の先頭から順に，時間に同期して和音進行を認識するために，本研究では仮説探索の手法を用いる．仮説探索とは，初期仮説，目標状態(最終的に得られる仮説が満たすべき条件)及び仮説の展開オペレータ(仮説から新しい仮説集合を生成するオペレータ)が与えられたときに，初期仮説から順に仮説を展開してい

Maxwell¹³⁾は，2.2 節で述べた 2 種類の相互依存問題を考慮して和声解析を行うプロダクションシステムを開発しているが，この手法は音高の時系列を入力対象としており，各プロダクションルールは音高情報に大きく依存して記述されている．このため，この手法を音響信号を入力対象とした場合に適用することは困難である．

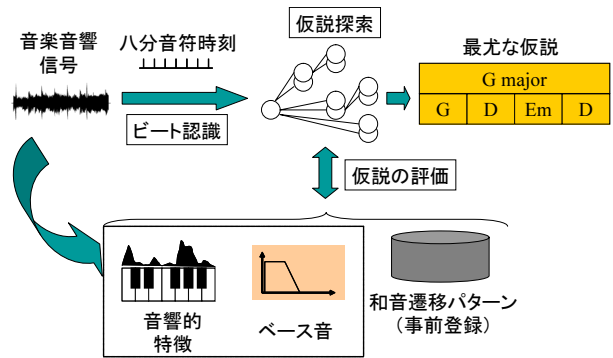


図 2 和音進行認識処理の流れ：仮説探索の各々の は，1 つの仮説を表す．左から右の方向に，仮説が順に展開される様子を示している．

き，目標状態を満たす最尤な仮説を求めることである．本研究における初期仮説は空の和音進行であり，目標状態は仮説が表す和音進行が入力の末尾に到達していることである．仮説の展開は，その仮説の和音進行の次に続く和音名を求めることに相当する．

本研究では，仮説の評価値を求めるために，音響的特徴，和音遷移パターン，ベース音の 3 つの音楽的要素を手がかりとして用いる．

(1) 音響的特徴

本研究では，音響的特徴として 12 次元クロマベクトル¹⁴⁾を用いる．クロマベクトルの各次元は 1 つの音名に対応し，各次元の値はクロマベクトルを抽出した時刻での対応する音名のパワーを近似的に表す．和音名は音名の組み合わせで決定されるので，和音進行認識において音響的特徴は不可欠である．

(2) 和音遷移パターン

和音遷移パターンとは，実際の楽曲でよく用いられる和音名の遷移である(例えば，八長調では， $A_m \rightarrow D_m, G \rightarrow C, G \rightarrow bA_{dim} \rightarrow A_m$ など)．したがって，和音遷移パターンを用いることで，和声音の省略や非和声音の出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．

(3) ベース音

ベース音とは，楽曲の各時点で鳴っている最も音高が低い音である．ポピュラー音楽では，ベース音は和音進行を導く役割があるので，ベース音を用いることで，認識性能の向上が期待される．

3. 和音進行認識手法

図 2 に本手法の処理の流れを示す．まず，ビート認識により入力楽曲の八分音符時刻を求める．ここでは後藤の手法¹⁵⁾を用いる．次に，八分音符時刻ごとに仮説を展開，評価していき，最尤な仮説を探索する．各仮説の評価値は音響的特徴，和音遷移パターン，ベー

ス音の3種類の音楽的要素により求められる。最終的に入力末尾に到達した仮説の中から、評価値が最大の仮説を認識結果として出力する。

仮説探索の処理の流れは次の通りである。最初に、空の和音進行をもつ初期仮説が与えられ、探索の先頭時刻が時刻0(入力の開始時刻)に初期化される。ここで、探索の先頭時刻とは、探索済みの時刻の先頭をさす。その後、先頭時刻の進行に同期して仮説を展開、評価していく。最終的に、先頭時刻が入力末尾に到達した時点で、評価値が最大である仮説を選択する。

3.1 仮説探索問題としての定式化

本研究では、和音進行認識を仮説探索問題として解く。具体的には、次の(1)~(4)の設定の下で、初期仮説集合からはじめて仮説を展開していくことにより、目標状態にある仮説集合を得る。この中から最尤である仮説を選択する。

(1) 仮説の定義

本研究では、仮説 h を和音進行 $c_1 c_2 \dots c_n$ と調 k の組であると定義する。

$$h = (c_1 c_2 \dots c_n, k)$$

(2) 初期仮説

本研究では、 ε を空の和音進行として、初期仮説集合 S を

$$S = \{(\varepsilon, k_i)\}_{k=1}^{24}$$

と定義する。

(3) 目標状態

仮説 $h = (c_1 c_2 \dots c_n, k)$ について、 c_n の終了時刻が入力末尾と等しいとき、 h は目標状態にあると定義する。

(4) 仮説展開オペレータ

仮説展開オペレータ $OP(h)$ は仮説 $h = (c_1 c_2 \dots c_n, k)$ の子仮説集合を与える。本研究では、和音区間の最小単位を $1/2$ 拍(八分音符間隔)、和音区間の最大長を4拍とし、 $1/2$ 拍から4拍まで8通りの和音区間ごとに48個(和音名の総数)の仮説を得る(図3)。すなわち、 $beat$ を1拍の長さ、 $c_n = (c_{s_n}, b_n, e_n)$ として、 $OP(h)$ を、

$$OP(h) = \{(c_1 c_2 \dots c_n c_{n+1}^{i,j}, k) | 1 \leq i \leq 48, 1 \leq j \leq 8\}$$

$$c_{n+1}^{i,j} = \left(c_{s_{n+1}}^i, e_n, e_n + beat \times \frac{j}{2} \right)$$

と定義する。

3.2 仮説探索アルゴリズム

和音進行認識を仮説探索問題として捉えた場合、探索アルゴリズムは仮説数の組み合わせ的爆発を防ぐた

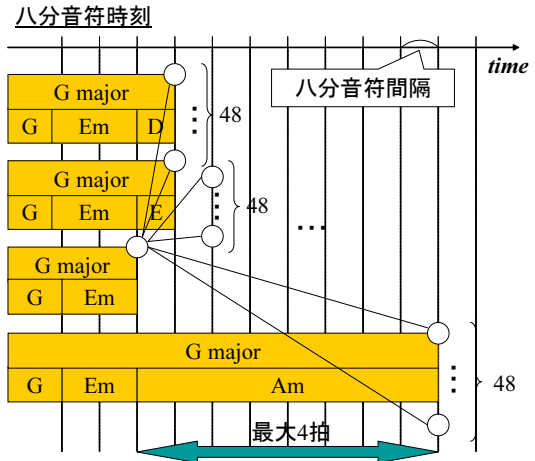


図3 仮説の展開: 仮説 (G-Em, G major) を展開して、 48×8 個の仮説を得る。各々の \square は仮説を表す。一部の \square については対応する仮説を例示している。 \square の時間軸上の位置はその仮説の和音進行の終了時刻を表す。親仮説との間の区間が展開により末尾に追加された和音の和音区間に対応する。転調を考慮しないため、子仮説の調は親仮説の調と同一である。

めに、枝刈りを行う必要がある。枝刈りとは、評価値の低い仮説を今後の展開の対象としない操作をさす。

和音進行認識における仮説探索で重要なことは、展開オペレータは和音進行の終了時刻が互いに異なる仮説を生成するが、終了時刻が互いに等しい仮説間でのみ枝刈りを適用しなければならないことである。なぜなら、仮説はその和音進行の終了時刻までの入力信号に対する解釈であるため、終了時刻が異なる仮説(例えば、図3の一番上と一番下の仮説)を比較することは不適切だからである。

このため本研究では、和音進行の終了時刻が探索の先頭時刻に等しい仮説間でのみ枝刈りを行う。枝刈りはビーム幅 BS のビーム探索にもとづいて行う。和音進行の終了時刻が先頭時刻を超えない仮説のみ評価できることに注意すると、先頭時刻がすすむごとに和音進行の終了時刻が先頭時刻に等しい仮説を評価した後、上位 BS 個の仮説のみ残るよう枝刈りし、残された仮説を展開すればよい(図4)。

ただし実装では、探索を効率化するために、仮説の展開を先頭時刻に合わせて遅延させる。すなわち、先頭時刻がすすむごとに、すぐに評価が可能な仮説のみ生成するように仮説を展開し、得られた仮説を評価、枝刈りする(図5)。このため、仮説の展開は一部保留されるので、完全に展開されなかった仮説はメモリ上に残しておく。展開の遅延により、展開によって新たに仮説が生成された直後にそれらの仮説に対して枝刈りを適用できるため、領域量を削減することができる。実際、ビーム幅を20とした場合のメモリ中に記憶すべき仮説数の推移は図6のようになる。図から、仮説数の組み合わせ的爆発を防ぐことと、展開を遅延

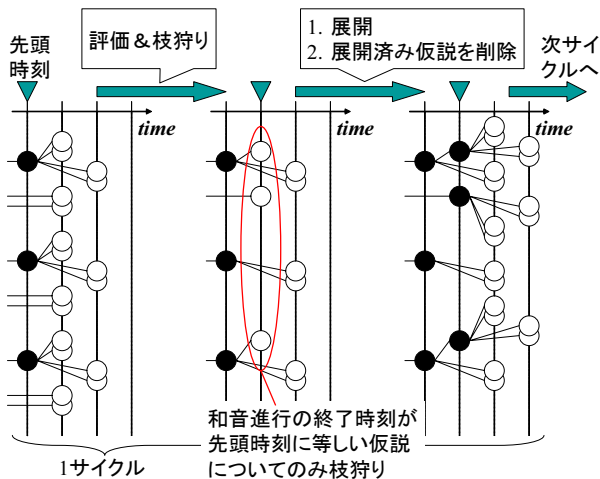


図4 枝切りの同期化：白丸は未展開仮説を，黒丸は展開済み仮説を表す．簡単のため，図では和音区間の最大長を1拍，展開により和音区間ごとに得られる仮説数を2，ビーム幅を3としている（左）直前サイクルの終了後の状態（中）現在時刻をすすめ，現在時刻上の仮説を評価し枝切りした状態（右）残された現在時刻上の仮説を展開し，展開済み仮説をメモリから削除した状態

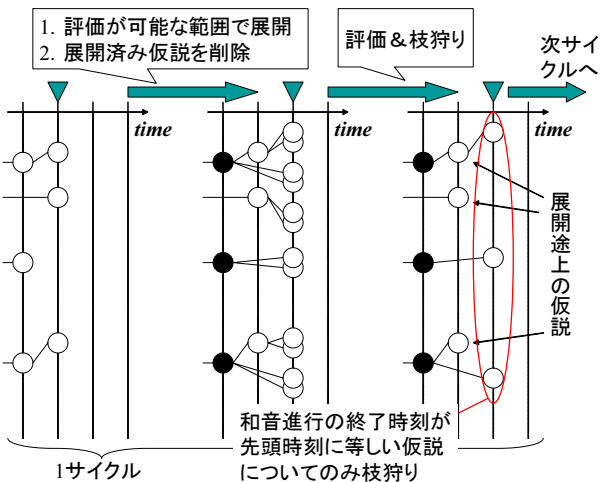


図5 展開の遅延：白丸は未展開及び展開途中の仮説を，黒丸は展開済み仮説を表す（左）直前サイクル終了後の状態（中）現在時刻をすすめ，評価が可能な範囲で仮説を展開し，展開済み仮説をメモリから削除した状態（右）現在時刻上の仮説を評価し枝切りした状態

させることにより記憶すべき仮説数を大幅に（1/216）抑えられることが分かる．さらに，展開を遅延させることにより，展開時に入力信号から得られる情報を利用できるため，最尤となる可能性が非常に低い仮説を生成せずに済み，時間量，領域量ともに削減することができる．

具体的な仮説探索アルゴリズムを図7に示す．図7の $V(h, t)$ は48個すべての仮説を生成せず，新たに追加される和音区間での音響的特徴にもとづいて N_{cand} 個の仮説のみ生成する．具体的には，新たに追加される和音区間で求まる β_i/d_i (3.3.1項) の値にもとづいてソートした上位 N_{cand} 個の仮説を生成する．

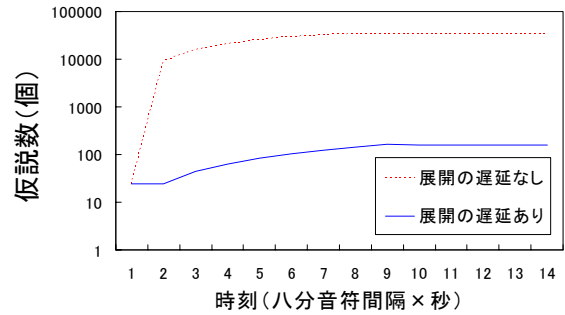


図6 記憶すべき仮説数の推移

S は初期仮説集合
 T は先頭時刻に到達している仮説集合
 U は先頭時刻に到達していない仮説集合
 $V(h, t)$ は仮説 h の時刻 t における子仮説集合：
 $V(h, t) \subset OP(h)$
 $f(h)$ は仮説 h の評価関数

初期化

```
for each  $s \in S$  do
   $f(s)$  を計算する
   $T \leftarrow T \cup \{s\}$ 
end
先頭時刻  $\leftarrow 0$ 
```

仮説探索

```
while 次の時刻が存在する do
  先頭時刻  $\leftarrow$  次の時刻
  for each  $h \in T$  do
    展開ブロック
    for each  $h' \in V(h, \text{先頭時刻})$  do
       $f(h')$  を計算する
       $T \leftarrow T \cup \{h'\}$ 
    end
    if  $h$  は展開済みでない do
       $U \leftarrow U \cup \{h\}$ 
    end
  end
  for each  $h \in U$  do
    do 展開ブロック
  end
   $T \leftarrow T$  の中から評価値の高い BS 個の仮説
   $U \leftarrow U$ 
end
```

```
return  $\arg \max_{f(h)} [h \in T]$ 
```

図7 仮説探索アルゴリズム

3.3 評価関数

評価関数 $f(h)$ は仮説 $h = (c_1 c_2 \cdots c_n, k)$ の尤もらしさを表す評価値を与える．本手法では，仮説 h の音響的特徴，和音遷移パターン，ベース音のそれぞれにもとづく評価値 $ac(h)$ ， $tr(h)$ ， $ba(h)$ を求め，これらを用いて，

$$f(h) = \log ac(h) + W_{\text{trans}} \times \log tr(h) + W_{\text{bass}} \times \log ba(h)$$

と定義する．ここで， W_{trans} は和音遷移パターンにもとづく評価値の重み， W_{bass} はベース音にもとづく評価値の重みである． $ac(h)$ の値は， h の和音進行中の各和音区間から抽出されたクロマベクトルと各和音名の学習用音響信号から抽出したクロマベクトルの類似度を表す． $tr(h)$ の値は，和音遷移パターンと一致する h の和音進行中の和音名の遷移の割合を表す． $ba(h)$ の値は， h の和音進行中の各和音について，その和音区間の低域スペクトルにおける和声音の優勢さの度合いを表す．

3.3.1 音響的特徴にもとづく評価値

仮説 $h = (c_1 c_2 \cdots c_n, k)$ の音響的特徴にもとづく評価値 $ac(h)$ を次のアルゴリズムにしたがって求める．

- (1) $c_i = (cs_i, b_i, e_i) (1 \leq i \leq n)$ の和音区間 $[b_i, e_i]$ でのパワースペクトルの時間平均から 12 次元クロマベクトル v_i を求める． v_i の各次元は 1 つの音名に対応し，各次元の値はその音名の音高（例えば，音名 C に対して C1, C2, ...）の基本周波数におけるパワーの和を，ベクトルのノルムが 1 になるように正規化して得られる．ここでは，主に伴奏音に注目するために，55Hz から 1000Hz の範囲においてのみ計算する．
- (2) 学習データから求められた cs_i のクロマベクトルの分布と v_i とのマハラノビス距離 d_i を求める．
- (3) $ac(h)$ を

$$ac(h) = \prod_{i=1}^n \left(\frac{\beta_i}{d_i} \times P_{\text{ext}}^{l_i-1} \right)$$

と定義する．ここで， β_i は和音区間 $[b_i, e_i]$ で得られる各和音名についてのマハラノビス距離の逆数の総和を 1 にするための正規化係数， l_i は c_i の和音区間に含まれる八分音符レベルのビート間隔数， P_{ext} は区間伸長ペナルティである．仮説に含まれる和音の個数 n は仮説によって異なるので，音響的特徴にもとづく評価値を β_i/d_i の積として定義すると，削除誤りが生じやすくなる．区間伸長ペナルティを乗じることにより，こうした削除誤りを防ぐことができる．

3.3.2 和音遷移パターンにもとづく評価値

仮説 $h = (c_1 c_2 \cdots c_n, k)$ の和音遷移パターンにもと

づく評価値 $tr(h)$ を，

$$tr(h) = P_{\text{trans}}^m$$

$$m = n - \text{num}(i; \exists p, q \text{ s.t. } p \leq i \leq q, c_p \cdots c_q \in P(k)) \quad \text{for } 1 \leq i \leq n$$

と定義する．ここで， $P(k)$ は調 k に対する和音遷移パターンの集合， P_{trans} は $P(k)$ 中のどの和音遷移パターンとも適合しない和音の遷移に対するペナルティ， $\text{num}(i; \text{cond}(i))$ for $1 \leq i \leq n$ は条件 $\text{cond}(i)$ を満たす i の個数を表す．ポピュラー音楽の和声理論¹⁶⁾を参考にして，長調と短調それぞれに和音機能の遷移パターンを設計し，システムに登録した（長调用 71 個，短调用 103 個）．この和音機能の遷移パターンに調 k を与えることで，調 k に対する和音遷移パターンの集合 $P(k)$ を求める．例えば，長调用に $\text{VIIm} \rightarrow \text{IIIm}$ ， $\text{V} \rightarrow \text{I}$ ， $\text{V} \rightarrow \text{bVI dim} \rightarrow \text{VIIm}$ などが登録されており，これに八長調を与えることにより八長調における和音遷移パターン $\text{Am} \rightarrow \text{Dm}$ ， $\text{G} \rightarrow \text{C}$ ， $\text{G} \rightarrow \text{bA dim} \rightarrow \text{Am}$ が得られる．

3.3.3 ベース音にもとづく評価値

仮説 $h = (c_1 c_2 \cdots c_n, k)$ のベース音にもとづく評価値 $ba(h)$ を次のアルゴリズムにしたがって求める．

- (1) 時刻 t における音高 p の確率密度関数 $\text{prob}^{(t)}(p)$ を求める．ここでは，後藤の音高推定手法¹⁷⁾を実装した櫻庭らの自動採譜システム¹⁸⁾を用いて，ベースライン用のフィルタ¹⁷⁾を適用してから音高確率密度関数を形成した．
- (2) $c_i = (cs_i, b_i, e_i)$ の和音区間 $[b_i, e_i]$ の低域スペクトルにおける各音名 j の優勢さの度合い pred_i^j を求める． pred_i^j の値は，音名 j の音高における $[b_i, e_i]$ 内での確率密度関数値の和を， $\sum_j \text{pred}_i^j = 1$ となるように正規化して得られる．
- (3) $ba(h)$ を，その和音進行の各和音区間における，最も優勢な音名の優勢さの度合いの積として求める．ただし，最も優勢な音名が和声音でない場合は，ペナルティを与える．

$$ba(h) = \prod_{i=1}^n ba_i$$

$$ba_i = \begin{cases} \text{pred}_i & \text{if } \text{chroma}_i \text{ は和声音} \\ P_{\text{bass}} & \text{otherwise} \end{cases}$$

$$\text{pred}_i = \max_j \text{pred}_i^j$$

$$\text{chroma}_i = \arg \max_j \text{pred}_i^j$$

P_{bass} は低域スペクトルにおいて最も優勢な音名が和声音でない場合のペナルティである．

4. 評価実験

本章では，本稿で述べた和音進行認識手法の有効性

表 2 パラメータの値

BS = 20	N _{cand} = 7	W _{trans} = 1.0
W _{bass} = 3.0	P _{ext} = 0.25	P _{trans} = 0.8
P _{bass} = 0.5		

を評価する目的で行った実験について報告する。

4.1 実験方法

本手法にもとづく和音進行認識システムを、Sun Blade 1000 (CPU: UltraSPARC, 750MHz) 上に実装した。また、八分音符時刻ごとに和音名を同定することにより和音進行を求めるシステム、及び仮説の評価値を音響的特徴のみにもとづいて計算するシステムも実装した。実装には C++ 言語を用いた。各パラメータの値を表 2 に示す。

本稿で述べた和音進行認識手法の有効性を評価するために、「RWC 研究用音楽データベース: ポピュラー音楽」¹⁹⁾ のうちの 10 曲 (RWC-MDB-P-2001 No.7, 14, 17, 20, 40, 45, 46, 54, 63, 74) を用いた。CD から抽出したこれらの楽曲のステレオ音響信号をモノラルに変換して、最初の 1 分間を切り出して和音進行認識実験を行った。クロマベクトルの学習データには、入力に用いない残りの 9 曲 (10-fold cross validation)、及び YAMAHA 社製 MIDI 音源 MU-2000 で作成した 2592 個の和音の音響信号を用いた。ここでは、音色を 3 種類 (ポピュラー音楽で頻繁に用いられる Acoustic Guitar, Electric Guitar, Overdrive Guitar)、音域を 2 オクターブ (基本形の根音の音高が MIDI ノートナンバー 40 から 63)、展開形を 3 種類、音響信号の切り出し位置を 3 種類 (立ち上がり後 0ms, 100ms, 200ms) 変化させて音響信号を作成した。

システムの出力を、入力音響信号のうちどれだけ正しく和音名を求めることができたかで和音進行認識率を評価した。

$$\text{和音進行認識率} = \frac{\text{正解和音名を出力した総区間長}}{\text{入力楽曲長}}$$

正解の和音進行は人手で付与した。

4.2 実験結果

結果を表 3 に示す。「仮説探索なし」は八分音符時刻ごとに和音名を同定するシステムの「音響的特徴のみ」は、仮説の評価値計算において音響的特徴のみを手がかりとして用いるシステムの実験結果である。本手法を実装したシステムでは、平均 72.1% の和音進行認識率を達成した。この結果から、本手法は歌唱やドラム音を含む複雑な音響信号に対して、正しく和音進行を認識できることが示された。

4.3 仮説探索に関する考察

仮説探索を用いた場合と用いなかった場合について

表 3 評価実験結果

楽曲番号	仮説探索なし	音響的特徴のみ	本手法
No.7	59.8%	62.3%	76.8%
No.14	46.0%	62.3%	76.9%
No.17	57.8%	64.2%	69.7%
No.20	46.8%	56.6%	65.2%
No.40	46.3%	67.7%	69.3%
No.45	55.4%	66.5%	69.1%
No.46	62.3%	71.9%	72.8%
No.54	66.3%	72.1%	75.8%
No.63	59.3%	62.5%	68.2%
No.74	55.1%	74.7%	76.9%
Average	55.5%	66.1%	72.1%

t 検定を行う。自由度 9 で $t = 5.07$ となるから、有意水準 0.05 でこれらの差は有意である。

評価に用いた楽曲では、多くの八分音符区間で和声音の一部が欠落していた。このため、仮説探索を用いない、すなわち八分音符時刻ごとに和音名を同定した場合、これらの区間で曖昧性が生じ多くの和音名同定誤りが生じていた。仮説探索を用いて様々な和音区間の可能性を考慮した場合、和音区間が八分音符間隔より長い仮説の評価値が大きくなりやすかった。これは、より長い区間では和声音が十分に含まれており、学習データとのマッチングがよかったためである。この結果から、和音区間の最小単位が与えられた場合、単位区間ごとに和音名を同定するよりも、必要に応じて和音区間を広げて認識した方が和音進行を正しく認識できることが結論づけられる。

4.4 評価値計算の手がかりに関する考察

音響的特徴のみを用いて仮説の評価値を求める手法と本手法について t 検定を行う。自由度 9 で $t = 3.74$ となるから、有意水準 0.05 でこれらの差は有意である。

音響的特徴のみを用いて評価値を計算した場合、Major と Minor 間の誤りが多数生じていた。これは、Major と Minor の和音では音響的特徴が類似するため、音響的特徴にもとづく評価値がほぼ等しくなるためである。和音遷移パターンやベース音により、これらの誤りを削減することができた。

また、音響的特徴のみを用いた場合、削除誤りに比べて多数の挿入誤りが生じていた。これは、正しい和音区間に非和声音が多く含まれる場合、それより短い区間における音響的特徴にもとづく評価値の方が正しい和音区間における評価値よりも大きくなりがちであったためである。これにより不自然な和音進行が生成されるが、主に和音遷移パターンを用いることによりそのような誤りを削減することができた。

4.5 実行時間に関する考察

長さ n の入力に対する、本手法の時間量を求める。八分音符時刻を求めるためのビート認識処理の時間量

は $O(n)$ である。また、図 7 に示した仮説探索アルゴリズムの時間量は、仮説展開と評価関数の計算に要する時間量が $O(1)$ であるから、 $O(n)$ である。したがって、本手法の時間量は $O(n)$ である。

実際に本手法を実装したシステムでは、各楽曲について実時間のおよそ 5 倍の実行時間を要した。したがって現状では、本手法をジャムセッションシステムをはじめとするリアルタイムシステムに応用することはできず、タグ付けや自動採譜システムなどの用途に限られる。ただし、本手法は長さ n の入力に対して $O(n)$ で動作するため、評価値計算などの各処理の効率を向上させることにより、リアルタイムな動作が可能になる。

5. おわりに

本稿では、音楽音響信号から和音進行と調を認識する手法について述べた。本手法は認識処理を構成する和音名同定、和音区間検出、調認識処理間の相互依存関係に対処するために、複数生成された和音進行と調に関する仮説の中から最尤な仮説を認識結果として出力する。この処理を入力の先頭から時間に同期して行うために、仮説探索にもとづく処理モデルを提案した。ポピュラー音楽の CD に収録されている、歌唱やドラム音を含む複雑な音響信号を対象とした評価実験で、72.1%の和音認識精度を達成した。

本稿では、ポピュラー音楽を対象として評価実験を行った。和音の提示のされ方はジャンルによって異なるため、他のジャンルの楽曲を扱えるようにするためには各ジャンルにおける和音の提示のされ方に応じた処理が求められる。例えば、R&B や無伴奏独唱ではメロディーによって和音が提示される傾向があるため、メロディから得られる特徴量を評価値計算の手がかりに含める必要がある。今後は、このような他のジャンルの楽曲を扱えるよう本手法を拡張していくとともに、実際の楽曲検索システムに本手法を応用していく予定である。

謝辞 本研究の一部は、日本学術振興会科学研究費補助金基盤研究(A)第 15200015 号、および 21 世紀 COE プログラム「知識社会基盤構築のための情報学拠点形成」によるものである。プログラムの使用を許可して頂いた櫻庭洋平氏、有益な議論をして頂いた吉井和佳氏に感謝する。

参 考 文 献

- 1) Manjunath, B.S., Salembier, P., and Sikora, T.: Introduction to MPEG-7, John Wiley & Sons Ltd., 2002.
- 2) Gomez, E., Gouyon, F., Herrera, P., and Amatriain, X.: Using and Enhancing the Current MPEG-7 Standard for a Music Content Processing Tool, *Proc. Audio Engineering Society*, 114th Convention, 2003.
- 3) Herrera, P., Serra, X.: A Proposal for the Description

- of Audio in the Context of MPEG-7, *Proc. CBMI*, 1999.
- 4) Casey, M.: General Sound Classification and Similarity in MPEG-7, *Organized Sound*, Vol.6, No.2, pp.153-164, 2002.
- 5) 青野裕司, 片寄晴弘, 井口征士: アコースティック楽器を用いたセッションシステムの開発, *信学論*, Vol.J82-D-II, No.11, pp.1847-1856, 1999.
- 6) Nawab, S.H., Ayyash, S.A., and Wotiz, R.: Identification of Musical Chords using Constant-Q Spectra, *Proc. ICASSP*, pp.V-3373-3376, 2001.
- 7) 柏野邦夫, 木下智義, 中臺一博, 田中英彦: 音楽情景分析の処理モデル OPTIMA における和音の認識, *信学論*, Vol.J79-D-II, No.11, pp.1762-1770, 1996.
- 8) Su, B. and Jeng, S.: Multi-timber Chord Classification Using Wavelet Transform and Self-organized Map Neural Networks, *Proc. ICASSP*, pp.V-3377-3380, 2001.
- 9) Fujishima, T.: Realtime Chord Recognition of Musical Sound: a System Using Common Lisp Music, *Proc. ICMC*, pp.464-467, 1999.
- 10) 山田洋子, 後藤真孝, 猿渡洋, 鹿野清宏: 音楽音響信号を対象とした和音名同定手法: 楽曲への適用, *音講論集*, 春期, 3-7-2, pp.835-836, 2003.
- 11) Sheh, A. and Ellis, D.P.W.: Chord Segmentation and Recognition Using EM-Trained Hidden Markov Models, *Proc. ISMIR*, 2003.
- 12) 平賀讓: コンピュータによる和声解析, 長嶋洋一, 橋本周司, 平賀讓, 平田圭二: コンピュータと音楽の世界, 共立出版, 1999.
- 13) Maxwell, H.J.: An expert system for harmonic analysis of tonal music, Balaban, M., Ebcioğlu, K., and Laske, O.: Understanding Music with AI, AAAI Press, 1992.
- 14) 後藤真孝: SmartMusicKIOSK: サビ出し機能付き音楽試聴機, *情処学論*, Vol.44, No.11, pp.2737-2747, 2003.
- 15) 後藤真孝: 音楽音響信号を対象としたリアルタイムビートトラッキングに関する研究, 博士論文, 早稲田大学理工学部, 1998.
- 16) 篠田元一: 新・実践コードワーク 1, リットーミュージック, 1997.
- 17) 後藤真孝: 音楽音響信号を対象としたメロディーとベースの音高推定, *信学論*, Vol.J84-D-II, No.1, pp.12-22, 2001.
- 18) 櫻庭洋平, 奥乃博: 自動採譜におけるパート形成処理のための特徴量の検討, *情処研報*, 2002-MUS-51, pp.35-42, 2003.
- 19) 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: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース, *情処学論*, Vol.45, No.3, pp.728-738, 2004.