Vol.2012-MUS-94 No.29
Vol.2012-SLP-90 No.29
2012/2/4

音響特徴・ベース音・和音遷移を用いた自動和音認識

糸山 克寿[†1] 尾形 哲也[†1] 奥乃 博[†1]

本稿では，多重奏音楽音響信号に対する自動和音和音手法について述べる．和音の認識においては，音楽的要素の関連性を考慮することが重要である．我々は，和音を表現する音響特徴であるクロマベクトルに加えて和音と関わりの深い音楽的要素であるベース音を用いた自動和音認識手法を構築した．和音遷移のパターンを事前に階層 Pitman-Yor 言語モデルで学習し，Viterbi アルゴリズムに基づく最大事後確率推定で和音系列を推定する．Beatles の 150 楽曲を用いた評価実験で，本手法は 73.7%の認識率を達成した．

Automatic Chord Recognition Based on Probabilistic Integration of Acoustic Features, Bass Sounds, and Chord Transition

Katsutoshi Itoyama,[†1] Tetsuya Ogata[†1] and Hiroshi G. Okuno [†1]

This paper describes a method that identifies musical chords in polyphonic musical signals. As musical chords mainly represent the harmony of music and are related to other musical elements such as melody and rhythm, we should be able to recognize chords more effectively if this interrelationship is taken into consideration. We use bass pitches as clues for improving chord recognition. The proposed chord recognition system is constructed based on Viterbi-algorithm-based maximum *a posteriori* estimation that uses a posterior probability based on chord features, chord transition patterns, and bass pitch distributions. Experimental results with 150 Beatles songs that has keys and no modulation showed that the recognition rate was 73.7% on average.

†1 京都大学 大学院情報学研究科
Graduate School of Informatics, Kyoto University

1. はじめに

近年，計算機による音楽音響信号の解析はますます重要になってきている．大容量デジタルオーディオプレイヤの普及やそれにともなう音楽配信サービスの発展により，多様なユーザのニーズに合わせた楽曲検索や楽曲推薦が求められている．大量の音楽コンテンツから柔軟な検索や推薦を行うためには，演奏者や曲名などの書誌情報に加えて，楽器構成や楽曲構造，ジャンル，テンポ，雰囲気などの楽曲の内容に基づく情報が重要である．そのような情報の人手による抽出と記述は，大量の楽曲の網羅と均一な品質の確保に困難をともなう．そこで，こうした情報を計算機によって自動的に抽出するためには，音楽音響信号を解析し，メロディやハーモニーなどの様々な音楽的要素を認識する技術が不可欠である．

本研究の最終的な目的は，音楽音響信号を解析し，様々な音楽的要素の相互関連を考慮したうえでこれらの要素を同時に認識する，音楽解析システムの構築である．音楽の三大要素とされるメロディ，ハーモニー，リズムに加えて，これらから派生する楽曲構造，音色，テンポ，さらに作風，奏法，作曲年代など，様々な音楽的要素があり，楽曲は複数の要素の関係を吟味したうえで制作されることが多い．そのため，音楽的要素を推定・認識する際も，要素の相互関係を考慮して解析することで，個々の要素の認識精度が向上すると期待される．

本論文での目的は，上記の音楽解析システム構築に向けた第一歩としての，ポピュラー楽曲におけるベース音と和音との関係に基づく和音系列の認識である．本論文では音楽的要素として，和音系列，調，和音を表す音響特徴，およびベース音を扱う．和音は音楽の三大要素の 1 つであるハーモニーから派生する音楽的要素である．和音の系列や分布は楽曲の雰囲気に強く影響を及ぼし，楽曲構造の類似度を計算する手がかりであることから，雰囲気やスタイルの類似性に基づく楽曲検索[1)–3)]，カバー曲の検索[4)] において重要な役割を果たす．ベース音とは，楽曲中の各時刻で最も音高が低い楽器音であり，多くのポピュラー楽曲ではベースギターによって演奏される．ベース音は和音のルート音や構成音であることが多く，またベース音の音高遷移パターンはそれ以降の和音遷移を暗に導く．したがって，ベース音の音高（ベース音高）を推定し，その結果を和音系列認識に用いることで，認識精度の向上が期待される．調は和音と同様にハーモニーから派生する音楽的要素で，楽曲全体で中心的な役割をもつ音および音階を表す．特に和音との関係では，和音の出現頻度や遷移パターンに影響を及ぼす．クラシック楽曲に比べてポピュラー楽曲では楽譜などが十分に整備されておらず，計算機による音響信号からの自動採譜などのニーズが高いこと，クラシック楽曲で重要な音楽的要素である主題がポピュラー楽曲では重視されず，和音系列が楽曲の雰囲気な

どに相対的に大きな影響を与えることなどから，本論文ではポピュラー音楽を対象とする．

これまで和音系列認識に関する研究は多数行われている．近年主流となっているのは，和音系列がマルコフ過程であると仮定し，様々な音響特徴量を観測データとして隠れマルコフモデル（Hidden Markov model: HMM）を用いた Viterbi 探索で最適な和音系列を求める手法である．拡張した Pitch Class Profile[5] を音響特徴として用いた手法[6]，和音遷移確率に音楽知識を取り入れた手法[7]，倍音の影響をモデル化した音響特徴を用いた手法[8]，音程間の関係を格子状のダイアグラムで表現した Tonnetz に基づく Tonal Centroid を用いて和音境界を検出する手法[9]，調ごとに異なる HMM を用いて各 HMM による認識結果から最尤なものを選択する手法[10]，和音の持続調をモデル化することで和音系列の断片化に対処した手法[11),12)]，打楽器音などの非調波的な音を抑制し調波音が強調された音響信号から抽出した音響特徴を用いた手法[13] などが提案されている．これらの手法では，使用する音響特徴や和音系列のモデル化などに工夫がみられるが，和音を表現する特徴のみを用いており，メロディやリズムなど，和音系列以外の音楽的要素との関連性は重視されていない．

これらに対して，ベースギターなどの楽器が演奏するメロディであるベースラインと和音系列との関連性を考慮した和音系列認識手法[14] が提案されている．この手法では，楽曲の先頭からある時刻までの和音名と和音境界の時系列および楽曲全体の調からなる和音系列仮説を，楽曲先頭から末尾まで探索することで和音系列を求める．和音区間から抽出した音響特徴の単一正規分布の尤度，音楽的知識に基づく和音遷移パターン，ベース音高に対するペナルティを用いて仮説探索における評価関数を定義している．前述のとおり，本研究の最終的な目的は複数の音楽的要素を同時に認識する音楽解析システムの構築であるため，認識のための目的関数が容易に拡張可能であり，かつその探索（最適化）手法が汎用的なものであることが望ましいが，主観的なペナルティや和音進行認識に特化した特殊な仮説探索は目的関数や最適化手法の拡張性を妨げるため，本研究の目的にそぐわない．

本稿ではこれらの手法の問題点を解決した和音系列認識システムについて述べる．各時刻の和音名を反映した音響特徴，メロディの一部であるベース音高，和音遷移の頻出パターンに基づいて和音系列の事後確率を定義し，事後確率を最大化する和音系列を Viterbi 探索で求める．

2. 和音認識における課題とアプローチ

我々は，ハーモニーを構成する和音とメロディの一部であるベース音高との関連に着目し，この関連を考慮した和音系列認識手法を提案する．

2.1 ベース音の性質

本研究では，和音と関わりの深い要素として，ベース音に着目する．ベース音は楽曲中の各時刻で最も音高が低い楽器音であり，以下のような和音に関係した性質をもつ．

- 和音の低音部を構成する．
- ベース音のある種の音高遷移パターンと和音の遷移パターンとの対応が決まっており，和音系列を導く役割を担う．

このため，ベース音が和音系列認識において有用な手がかりとなることが期待できる．

2.2 解決すべき課題

従来手法[14),15)] では，音響特徴の長時間平均の分布が単一正規分布に従うと仮定し，和音区間内で平均した音響特徴に対する各和音名に対応させた単一正規分布パラメータの尤度で音響特徴に基づく和音系列仮説の評価を行っていた．同じ和音名を持つ和音区間では音響特徴の平均がつねに同じ分布に従うと仮定していたため，楽器編成や演奏表情（分散和音など）の違いに起因する音響特徴のばらつきが和音系列認識に悪影響を与え，また Viterbi 探索などの最適経路探索手法を活用できなかった．

西洋音楽の音楽理論では，和音遷移の基本的なパターンは**トニック** → **ドミナント** → **トニック**のような 3 つ以上の和音の列で記述される．従って，和音遷移のモデルも 3 つ以上の和音の列を扱えることが望ましいが，従来手法[15] は和音遷移モデルに 2-gram を用いていた．

2.3 アプローチ

上記の課題を解決するため，我々は時間独立性の高い確率モデルと汎用性の高い言語モデルを用いる．

音響特徴の時刻間での独立性を向上させるため，混合正規分布（Gaussian mixture model: GMM）を用いて音響特徴の生成過程をモデル化する．音響特徴には 12 次元クロマベクトル[16] を用いる．GMM は単一正規分布よりも広範な確率分布を表現できるため，音響特徴の長時間平均をとらなくても各時刻の音響特徴の分布を十分にモデル化でき，和音系列認識精度向上が期待される．3.1 節で述べるように，本研究では三和音のみを認識対象とするが，実際の楽曲には四和音やテンションコードなども含まれており，これらを構成音に基づいて三和音のいずれかに分類する．したがって，それぞれの和音の種類に対して，単独分布よりも混合分布での表現が妥当であると考えられる．すべての和音に対する GMM 学習には多くの学習サンプルが必要となるが，3.2.1 項で説明するクロマベクトルの巡回シフトを用いてルート音に依存しない和音の種類に関するモデル化を行うことで対処する．

Vol.2012-MUS-94 No.29
Vol.2012-SLP-90 No.29
2012/2/4

表 1 本稿で扱う和音の種類
Table 1 Chord types dealt with in this paper.

和音の種類	構成音
Major	ルート音，長 3 度，完全 5 度
Minor	ルート音，短 3 度，完全 5 度
Diminished	ルート音，短 3 度，減 5 度
Suspended4	ルート音，完全 4 度，完全 5 度

従来手法が和音遷移モデルに N-gram 言語モデルを用いていたのに対して，提案手法は Hierarchical Pitman-Yor Language Model (HPYLM) を用いる．

3. 和音認識システム

本節では，和音認識システムの実装について述べる．システムへの入力は音楽 CD などの多重奏の音楽音響信号，出力は推定された和音系列と楽曲の調である．和音系列の事後確率を音響特徴・ベース音高・和音系列自身に基づいて計算し，事後確率が最大となる和音系列を推定する．以下では，和音認識の定式化，事後確率の定義，および事後確率の計算手法について述べる．

3.1 定 式 化

本システムは八分音符区間を和音推定の最小単位とする．八分音符区間はビートトラッキング[17] で事前に推定する．推定された八分音符区間を $[e_1, \ldots, e_K]$ で表す．各八分音符区間において，音響特徴ベクトルの計算とベース音高分布の推定を行う．音響特徴ベクトル系列を $X = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K]$，ベース音高分布系列を $B = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_K]$ とする．

本システムの認識対象は，和音系列 $\boldsymbol{c} = [c_1, \ldots, c_K]$（各八分音符区間の和音名）および調 s で，それぞれ以下の値をとる変数である．

$$c_k \in R \times \{\text{Major, Minor, Diminished, Suspended4}\} \quad (1)$$

$$s \in R \times \{\text{Major, Minor}\} \quad (2)$$

ただし，$R = \{\text{C, C\#, D, D\#, E, F, F\#, G, G\#, A, A\#, B}\}$ で，12 種類のルート音の集合である．**表 1** に示した，ポピュラー音楽で頻繁に使用される 4 つの和音の種類を用いる．その他の和音（Augmented などの三和音および 4 つ以上のピッチクラスからなる和音）は，構成音を考慮してこれら 4 つの和音のいずれかに分類する．調は楽曲全体に対するものであり，楽曲中では転調しないと仮定する．

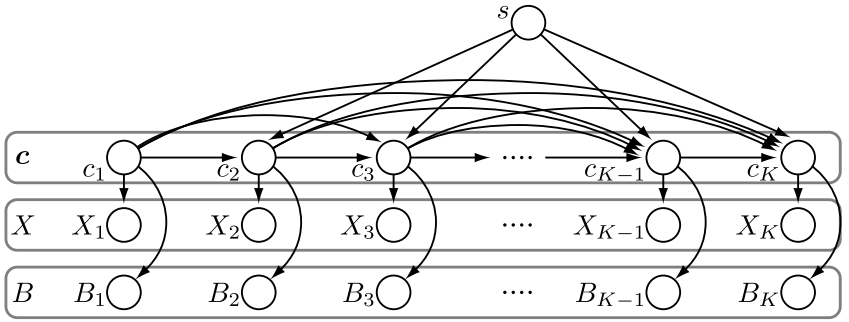

図 1 和音名，調，音響特徴，ベース音高分布のグラフィカルモデル

図 2 Graphical model of chord sequence, key, acoustic features, and bass pitch probabilitiy distributions.

3.2 最大事後確率推定

和音名・調・音響特徴・ベース音高の同時確率を以下で定義する．

$$p(\boldsymbol{c}, s, X, B) = p(s)\, p(c_1) \prod_{k=1}^{K} p(X_k|c_k)\, p(B_k|c_k) \prod_{k=2}^{K} p(c_k|c_1, \ldots, c_{k-1}, s) \quad (3)$$

このグラフ表現を**図 2** に示す．このモデルは以下の特徴を持つ．

(1) 和音名 c_k は，1 つ前の八分音符区間における和音名 c_{k-1} と調 s に依存する．

(2) 音響特徴 X とベース音高分布 B は各八部音符区間で独立であり，その八分音符区間の和音名にのみ依存する．

Bayes の定理より，和音名系列 $\boldsymbol{c}$ と調 s の事後確率は以下を満たす．

$$p(\boldsymbol{c}, s|X, B) = \frac{p(\boldsymbol{c}, s, X, B)}{p(X, B)} \quad (4)$$

ここで，$p(X, B)$ は MAP 推定には無関係なので，実際には $p(\boldsymbol{c}, s, X, B)$ のみを考えればよい．本システムでは調の事前確率 $p(s)$ はすべての調に関して一様であると仮定する．調には楽曲のジャンルや作曲者によって偏りがあるものの，一般には完全に未知であるためである．さらに，第 1 八分音符区間和音名の事前確率 $p(c_1)$ も，すべての和音名に対して一様であると仮定する．これらをまとめると，以下が本システムで推定する和音系列と調となる．

Vol.2012-MUS-94 No.29
Vol.2012-SLP-90 No.29
2012/2/4

$$\arg\max_{\boldsymbol{c},s} p(\boldsymbol{c}, s|X, B)$$
$$\equiv \arg\max_{\boldsymbol{c},s} p(c_1)\prod_{k=1}^{K} p(X_k|c_k)\, p(B_k|c_k)\prod_{k=2}^{K} p(c_k|c_1,\ldots,c_{k-1},s). \quad (5)$$

3.2.1 音響特徴に基づく尤度：$p(X_k|c_k)$

クロマベクトル[16]は，12の音名ごとにパワースペクトルを足し合わせることでパワーの分布を表現する音響特徴である．主に低音域から中音域の伴奏音が和音を表現するため，クロマベクトルを求める周波数の範囲は55–1,000Hzとした．和音は様々な音高の重ね合わせで表現されるため，クロマベクトルの値が大きい要素に着目することで和音名が推定できる．種類が同じでルート音だけが異なる和音（例えばC MajorとD Major）は，それぞれがその構成音を周波数方向に平行移動した関係にある．そのため，ルート音だけが異なり種類が一致している和音区間から求めたクロマベクトル同士は，ルート音に応じてベクトルを巡回シフトさせるとその値の分布は類似していると考えられる．具体的なクロマベクトルの巡回シフト処理の様子を**図3**に示す．すなわち，MajorやMinorといった**和音の種類**ごとのクロマベクトルの分布だけを考え必要に応じてクロマベクトルを巡回シフトさせることで，C MajorやA Minorといった**和音名**ごとのクロマベクトルの分布を仮想的に扱うことができる．これによって，実質的に扱うクロマベクトルの分布数は和音の種類の数と同数であればよいことになるため，分布あたりの学習サンプル数の増加が見込まれ，和音系列認識の頑健性が向上する．

クロマベクトルの確率モデルとしてGMMを用い，クロマベクトルは各時刻で独立に生成されると仮定する．学習データからクロマベクトルを抽出し，正解和音ラベルに基づく巡回シフトを行い，和音の種類ごとにGMMのパラメータを最尤推定に基づくEMアルゴリズムで推定する．GMMの混合数をM，和音名cのルート音に対応したGMMの混合係数，平均，共分散パラメータをそれぞれ$\{\alpha_{c,1},\ldots,\alpha_{c,M}\}$, $\{\boldsymbol{\mu}_{c,1},\ldots,\boldsymbol{\mu}_{c,M}\}$, $\{\Sigma_{c,1},\ldots,\Sigma_{c,M}\}$とする．

各時刻で求めたクロマベクトルを和音名の対応するルート音に応じて巡回シフトさせクロマベクトルに基づく尤度を以下で計算する．

$$p(\boldsymbol{x}_k|c_k) = \sum_{m=1}^{M} \frac{\alpha_{c_k,m}}{(2\pi)^6\,|\Sigma_{c_k,m}|^{-\frac{1}{2}}} \exp\left(-\frac{(\boldsymbol{y}_k-\boldsymbol{\mu}_{c_k,m})^T\,\Sigma_{c_k,m}^{-1}\,(\boldsymbol{y}_k-\boldsymbol{\mu}_{c_k,m})}{2}\right) \quad (6)$$

$\boldsymbol{y}_k$は和音名c_kのルート音に応じて$\boldsymbol{x}_k$を巡回シフトさせたベクトルである．

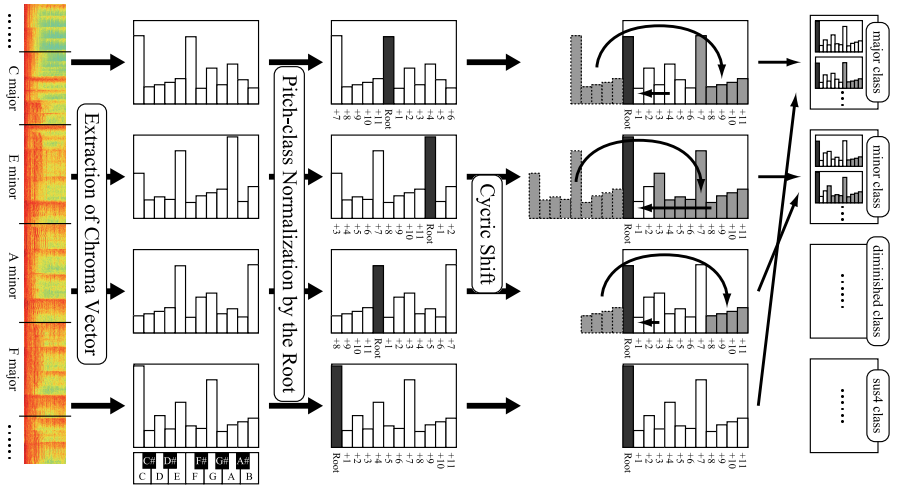

図3 クロマベクトルの巡回シフト．抽出されたクロマベクトルをルート音の音名で正規化し，ルート音のインデックスが等しくなるように移動し，あふれた要素を逆の端におくことで和音の種類ごとのベクトルが得られる
Fig. 3 Cycric shift of chroma vector. By normalizing the indices of vectors by the root, then shifting and rotating the elements, the vectors for each code kind are obtained.

3.2.2 ベース音高に基づく尤度：$p(B_k|c_k)$

ベース音から和音を直接推定することは難しいが，ベース音高やその遷移パターンは和音進行を導く．ベース音高分布$\boldsymbol{b}_k$は，八分音符区間e_kにおいてベースラインを演奏する楽器音が存在する確率をPreFEst[18]で推定したものであり，この列をベース音高分布系列Bとする．ベースラインは周波数f_0（パワースペクトルにおける周波数インデックスM_0）からf_1（周波数インデックスM_1）の間にあるものとする．$\boldsymbol{b}_k$は，パワースペクトルの$[f_0, f_1]$に含まれる周波数に対応する要素を持つ$M_1 - M_0 + 1$次元単体（simplex）上のベクトルで定義される．

$$k = 1,\ldots,K : \left[\begin{array}{c} \boldsymbol{b}_k = (b_{k,M_0},\ldots,b_{k,M_1});\ \sum_{m=M_0}^{M_1} b_{k,m} = 1; \\ m = M_0,\ldots,M_1 : 0 \le b_{k,m} \le 1 \end{array}\right]. \quad (7)$$

和音名cの和音区間で平均的に得られるベース音高分布を$\boldsymbol{\beta}_c = (\beta_{c,M_0},\ldots,\beta_{c,M_1})$とし，学習データから得られるベース音高分布と和音名の組を基にこれを推定する．

ある和音名 c のもとで出現しやすい音高（多くの場合，和音構成音）のベース音がその和音区間ではっきりと演奏されているとき，ベースラインから和音名を推定することは一般的に容易になる．したがって，出現しやすいベース音がはっきりと演奏されているときに，ベース音高分布に基づく和音名の尤度も大きな値をとるべきである．ベース音が特定の音高ではっきりと演奏されているとき，ベース音高分布はその音高のみ大きい値を，それ以外の音高では 0 に近い値となる．

上記の特徴を満たすよう，ベース音高分布に基づく尤度を以下で定義する．

$$p(\boldsymbol{b}_k|c_k) = (M_1 - M_0 + 1)!\,\langle \boldsymbol{b}_k, \boldsymbol{\beta}_{c_k} \rangle \tag{8}$$

$\langle \cdot, \cdot \rangle$ はベクトルの内積を表す演算子である．この確率分布のモード（最頻値）は以下となる．

$$\begin{aligned} &(b_{k,M_0}, \ldots, b_{k,m}, \ldots, b_{k,M_1}) = (0, \ldots, 1, \ldots, 0) \\ &\text{s.t. } m = \arg \max_{m \in \{M_0, \ldots, M_1\}} \{\beta_{c_k,M_0}, \ldots, \beta_{c_k,M_1}\}, \end{aligned} \tag{9}$$

和音名 c_k で出現しやすい音高のベース音がはっきり演奏されているとき，すなわち $\boldsymbol{\beta}_{c_k}$ で大きい値を持つ音高に関して $\boldsymbol{b}_k$ も大きい値を持つときに大きい尤度を与える．

ベース音高分布 $\boldsymbol{b}_k$ に基づく尤度の計算における重要な点は，$\boldsymbol{b}_k$ で大きな値を持つ音高が和音名 c_k の和音区間で平均的によく演奏される（$\boldsymbol{\beta}_{c_k}$ で大きな値を持つ）音高と一致するときに大きな尤度を与えることであり，$\boldsymbol{b}_k$ と $\boldsymbol{\beta}_{c_k}$ が類似している和音 c_k に大きな尤度を与えることではない．正規分布や Dirichlet 分布を用いた尤度関数は後者の特徴を持つため必ずしも適切ではない．

3.2.3 調ごとの和音遷移パターンに基づく事後確率：$p(c_k|c_{k-1}, s)$

多くの楽曲では，ジャンルやアーティストごとに和音遷移のパターンがある程度限定され，その遷移頻度には偏りがある．音楽理論では，I – IV – V のような抽象化された形式でこのパターンは表現され，調を固定するとこのパターンは具体的な和音名に変換できる．楽曲中での和音名の出現・遷移頻度をモデル化することで，音響特徴やベース音だけでは解消できない和音名の曖昧性を解決する．和音遷移のモデル化には階層的 Pitman-Yor 言語モデル (Hierarchical Pitman-Yor Language Model; HPYLM) を用いる．

HPYLM は N-gram 言語モデルの階層的な生成モデルであり，条件付確率 $p(c_k|c_1, \ldots, c_{k-1}, s)$ を (N − 1)-次マルコフ過程 $p(c_k|c_{k-n+1}, \ldots, c_{k-1}, s)$ で近似する．(N − 1)-次マルコフ過程和音列 $h = c_{k-n+1}, \ldots, c_{k-1}$ で条件付けられた和音 c_k の事後確率は以下で定義される．

$$p(c_k|h) = \frac{\mathrm{c}(c_k|h) - d \cdot t_{hc_k}}{\theta + \mathrm{c}(h)} + \frac{\theta + d \cdot t_{h\cdot}}{\theta + \mathrm{c}(h)} p\big(c_k|h'\big) \tag{10}$$

表 2 音楽的要素の組合せの変化による和音認識率 [%] の変化
Table 2 Recognition rates [%] for each combination of musical elements.
(1) 音響特徴のみ，(2) (1) ＋ベース音高，(3) (1) ＋和音遷移，
(4) 提案手法，(5) 従来手法[15)]

Method	(1)	(2)	(3)	(4)	(5)
Recognition rate	59.8	66.6	61.9	**73.7**	73.4

表 3 GMM の混合数の違いによる和音認識率 [%] の変化
Table 3 Recognition rates [%] for each number of GMM components.

Number of components	1	2	4	8	16
Recognition rate	61.0	67.9	**73.7**	72.6	66.9

$\mathrm{c}(c_k|h)$ は，学習データ中で h に続いて c_k が現れた回数，$\mathrm{c}(h) = \sum_{c_k} \mathrm{c}(c_k|h)$ はそれらの合計，$h' = c_{k-n+2}, \ldots, c_{k-1}$ は h の次数を一つ下げた和音列をそれぞれ表す．

4. 評価実験

本手法の有効性を検証するため 3 章で述べた和音系列認識システムを実装し評価実験を行った．実験用データとして The Beatles の 12 枚の CD アルバムの全 180 曲から，調を持ち転調しない 150 曲を用いた．音響信号はモノラル 16kHz に変換し，8192 サンプルのガウス窓を 1024 サンプルずつシフトさせる短時間フーリエ変換で周波数分析した．150 曲を無作為に 30 曲ずつの 5 つの楽曲群に分割し，5-fold cross validation を行った．音響特徴 GMM の混合数は原則として $M = 4$，ベース音高の範囲は $[f_0, f_1] = [29, 261]$[Hz] とした．音響特徴モデル，ベース音高モデル，和音遷移モデルの学習には，テストに用いない残り 4 グループの 120 曲を用いた．学習と評価において，C. Harte らが作成した和音データ[19)] を用いた．

入力音響信号のうち正しく和音名を求めることができた割合で認識結果を評価した．

$$\text{和音認識率} = \frac{\text{正解和音を出力した総区間長}}{\text{入力楽曲長}} \tag{11}$$

また，連続して正しく和音名を求めることができた区間長の平均と最大を用いた．

本手法の有効性を検証するため，事後確率計算に用いる音楽的要素の組合せを次のように変化させて実験を行った．

(1) 音響特徴のみを用いて事後確率計算
(2) 音響特徴とベース音高分布を用いて事後確率計算

(3) 音響特徴と和音遷移パターンを用いて事後確率計算
(4) **本手法**：音響特徴，ベース音高分布，和音遷移パターンを用いて事後確率計算
(5) 従来手法[15]：単一正規分布でモデル化した音響特徴，ルールベースのベース音高と和音遷移によるペナルティを用いてビームサーチ（ビーム幅 25）で目的関数最大化

また，音響特徴モデルの最適な GMM の混合数を調査するため，混合数を 1，2，4，8，16 に変化させて実験を行った．混合数を 1 とした場合は，音響特徴を単一正規分布でモデル化した場合に相当する．それぞれの実験結果を**表 2**，**表 3** に示す．

表 2 より，本手法での平均和音認識率は 73.7%となった．音響特徴のみを用いた場合との比較では，ベース音高を追加することで 6.8 ポイント，和音遷移を追加することで 2.1 ポイント，これらの両方を追加することで 13.9 ポイント，それぞれ認識率が向上し，ベース音高と和音遷移の有効性が示された．また，従来手法と比べて 0.3 ポイントの認識率の向上があった．これらの結果から，和音系列認識におけるベース音高や和音遷移パターンの利用，および本論文で着目した拡張性と汎用性に優れた最適化手法の有効性が検証された．

表 3 より，GMM の混合数を 2，4，8，16 としたいずれの場合にも，GMM の混合数を 1 とした場合よりも和音認識率は高くなった．この結果により，音響特徴を GMM でモデル化することの有効性が検証された．また，GMM の混合数を 4 とした場合に和音認識率は最も高くなったため，この結果に基づいて他の実験では GMM の混合数を 4 とした．

さらに，音楽的文脈の有効性を検証するため，和音遷移モデルの最大コンテキスト長を変えながら認識を行い，連続正解区間長を計算した．従来手法では平均の正解長は 1.66 秒だったのに対して，2-gram 遷移モデルを用いた本手法は 2.88 秒，3-gram を用いた場合は 2.97 秒となり，連続正解区間長は 1.22 秒および 1.31 秒向上した．これにより，汎用的な言語モデルを用いることで，連続的な認識性能が改善することが示された．

5. おわりに

本稿では，音響特徴・ベース音高・和音遷移という音楽的要素に着目し，これらの手がかりを確率的に統合して音響信号から和音を推定する，自動和音認識手法について述べた．それぞれの手がかりに基づく和音系列の事後確率を定義し，事後確率を最大化する和音系列を Viterbi 探索で求めた．評価実験を行い，提案手法は The Beatles の 150 楽曲に対して，平均で 73.7%の認識率を実現した．この結果から，複数の音楽的要素を用いて和音を認識することの有効性が示された．

本論文ではベース音を用いた和音系列認識を扱ったが，和音認識結果をベース音高推定に用いることでベース音高の推定性能も和音同様に向上することが期待される．和音のベース音高分布に対する尤度の定義と同様に，和音に対するベース音高分布の尤度を定義し，さらに和音とベース音高を反復的に推定することで和音とベース音高の同時認識が可能となり，今後の検証が期待される．本研究で着目した音楽的要素間の相互関連と，相異なる要素を統一的に扱う確率的な枠組みは，和音系列だけでなく，メロディやリズム，またそこから派生する要素など，様々な音楽的要素の解析においても重要な視点である．今後，他の要素の解析においても音楽的要素間の関連を考慮した同時的な認識を確率的枠組みに基づいて設計することが妥当であると考えられる．

謝辞 本研究の一部は，科研費基盤研究 (S)，京都大学若手研究者スタートアップ研究費の支援を受けた．

参考文献

1) Shan, M.-K., Kuo, F.-F. and Chen, M.-F.: Music Style Mining and Classification by Melody, *ICME2002*, pp.97–100 (2002).
2) Cheng, H.-T., Yang, Y.-H., Lin, Y.-C., Liao, I.-B. and Chen, H. H.: Automatic Chord Recognition for Music Classification and Retrieval, *ICME2008*, pp. 1505–1508 (2008).
3) Hanna, P., Rocher, T. and Robine, M.: A Robust Retrieval System of Polyphonic Music Based on Chord Progression Similarity, *SIGIR'09*, pp.768–769 (2009).
4) Bello, J. P.: Audio-based Cover Song Retrieval Using Approximate Chord Sequences: Testing Shifts Gaps, Swaps and Beats, *ISMIR2007*, pp.239–244 (2007).
5) Fujishima, T.: Realtime Chord Recognition of Musical Sound: A System Using Common Lisp Music, *ICMC1999*, pp.464–467 (1999).
6) Sheh, A. and Ellis, D.P.: Chord Segmentation and Recognition using EM-Trained Hidden Markov Models, *ISMIR2003*, pp.183–189 (2003).
7) Bello, J.P. and Pickens, J.: A Robust Mid-level Representation for Harmonic Content in Music Signals, *ISMIR2005*, pp.304–311 (2005).
8) Papadopoulos, H. and Peeters, G.: Large-scale Study of Chord Estimation Algorithms Based on Chroma Representation and HMM, *CBMI2007*, pp.53–60 (2007).
9) Harte, C., Sandler, M. and Gasser, M.: Detecting Harmonic Change in Musical Audio, *AMCMM06*, pp.21–26 (2006).
10) Lee, K. and Slaney, M.: Acoustic Chord Transcription and Key Extraction from Audio Using Key-dependent HMMs Trained on Synthesized Audio, *IEEE Trans. Audio, Speech and Lang. Process.*, Vol.16, No.2, pp.291–301 (2008).
11) Abdallah, S., Sandler, M., Rhodes, C. and Casey, M.: Using Duration Models to

Reduce Fragmentation in Audio Segmentation, *Mach. Learn.*, Vol.6, No.2-3, pp. 485–515 (2006).
12) Mauch, M. and Dixon, S.: A Discrete Mixture Model for Chord Labelling, *ISMIR2008*, pp.45–50 (2008).
13) 内山裕貴，宮本賢一，西本卓也，小野順貴，嵯峨山茂樹：調波音・打楽器音分離手法を用いた音楽音響信号からの自動和音認識，情処研報， Vol.2008, No.78 (2008-MUS-76), pp.137–142 (2008).
14) Yoshioka, T., Kitahara, T., Komatani, K., Ogata, T. and Okuno, H.G.: Automatic Chord Transcription with Concurrent Recognition of Chord Symbols and Boundaries, *ISMIR2004*, pp.100–105 (2004).
15) Sumi, K., Itoyama, K., Yoshii, K., Komatani, K., Ogata, T. and Okuno, H.G.: Automatic Chord Recognition Based on Probabilistic Integration of Chord Transition and Bass Pitch Estimation, *ISMIR2008*, pp.39–44 (2008).
16) Goto, M.: A Chorus-section Detecting Method for Musical Audio Signals, *ICASSP2003*, pp.V–437–440 (2003).
17) Goto, M.: An Audio-based Real-time Beat Tracking System for Music with or without Drum-sounds, *J. New Music Res.*, Vol.30, No.2, pp.159–171 (2001).
18) Goto, M.: A Real-time Music-scene-analysis System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-world Audio Signals, *Speech Communication*, Vol.43, No.4, pp.311–329 (2004).
19) Harte, C., Sandler, M., Abdallah, S. and Gómez, E.: Symbolic Representation of Musical Chords: A Proposed Syntax for Text Annotations, *ISMIR2005*, pp.66–71 (2005).

付録　ベース音高分布確率密度関数の性質

$(N-1)$ 次元単体上にあるパラメータベクトル $\boldsymbol{w}$ と確率変数ベクトル $\boldsymbol{x}$ を考える．$\boldsymbol{x}$ は $0 \le x_n \le 1$ $(1 \le n \le N)$，$\sum_{n=1}^{N} x_n = 1$ を，$\boldsymbol{w}$ も $\boldsymbol{x}$ と同様の制約を満たす．正規化係数を除き，$\boldsymbol{x}$ の確率密度関数を $p(\boldsymbol{x}|\boldsymbol{w}) \propto \langle \boldsymbol{w}, \boldsymbol{x} \rangle$ で定義する．正規化係数は密度関数の積分の逆数で与えられる．

$$\begin{aligned}
&\int_{\left\{\boldsymbol{x}\,\middle|\,\forall n=1,\ldots,N: 0\le x_n\le 1;\ \sum_{n=1}^{N} x_n=1\right\}} \sum_{n=1}^{N} w_n x_n \, d\boldsymbol{x} \\
&= \sum_{n=1}^{N} \int_0^1 \cdots \int_0^{1-\sum_{m=1}^{N-2} x_n} w_n x_n \, dx_{N-1} \cdots dx_1 \qquad (12)
\end{aligned}$$

この和を分解し，$n = 1, \ldots, N$ のそれぞれに対して計算する．$n = 1, \ldots, N-1$ のとき，

$$\begin{aligned}
&\int_0^1 \cdots \int_0^{1-\sum_{m=1}^{N-2} x_m} w_n x_n \, dx_{N-1} \cdots dx_1 \\
&= \int_0^1 \cdots \int_0^{1-\sum_{m=1}^{N-3} x_m} w_n x_n \left(1 - \sum_{m=1}^{N-2} x_m\right) dx_{N-2} \cdots dx_1 \\
&= \cdots = \int_0^1 \cdots \int_0^{1-\sum_{m=1}^{n-1} x_m} \frac{w_n}{(N-n-1)!} x_n \left(1 - \sum_{m=1}^{n} x_m\right)^{N-n-1} dx_n \cdots dx_1 \\
&= \int_0^1 \cdots \int_0^{1-\sum_{m=1}^{n-2} x_m} \frac{w_n}{(N-n+1)!} \left(1 - \sum_{m=1}^{n-1} x_m\right)^{N-n+1} dx_{n-1} \cdots dx_1 \\
&= \cdots = \frac{w_n}{N!} \qquad (13)
\end{aligned}$$

$n = N$ のとき，

$$\begin{aligned}
&\int_0^1 \cdots \int_0^{1-\sum_{m=1}^{N-2} x_m} w_N x_N \, dx_{N-1} \cdots dx_1 \\
&= \int_0^1 \cdots \int_0^{1-\sum_{m=1}^{N-2} x_m} w_N \left(1 - \sum_{m=1}^{N-1} x_m\right) dx_{N-1} \cdots dx_1 \\
&= \cdots = \frac{w_N}{N!} \qquad (14)
\end{aligned}$$

$\boldsymbol{w}$ は $N-1$ 次元単体上のベクトルなので，$\sum_{n=1}^{N} w_n = 1$ を満たす．したがって，$\sum_{n=1}^{N} w_n/N! = 1/N!$ となり，正規化係数はこの逆数の $N!$ となる．これはパラメータ $\boldsymbol{w}$ に依存しない定数である．

次に，この確率密度関数のモードが式 (9) で与えられることを示す．簡単のため $w_1 = w_2 = \cdots = w_M > w_{M+1} \ge \cdots \ge w_N$ であるとする．$\hat{\boldsymbol{x}} = (\hat{x}_1, \ldots, \hat{x}_M, 0, \ldots, 0)$，$\hat{x}_m \ge 0$ $(m = 1, \ldots, M)$，$\sum_{m=1}^{M} \hat{x}_m = 1$ とすると，以下の不等式が成り立つ．

$$\langle \boldsymbol{w}, \hat{\boldsymbol{x}} \rangle - \langle \boldsymbol{w}, \boldsymbol{x} \rangle = w_1 - \sum_{n=1}^{N} w_n x_n = \sum_{n=1}^{N} (w_1 - w_n) x_n \ge 0 \qquad (15)$$

したがって，$\hat{\boldsymbol{x}}$ は $\langle \boldsymbol{x}, \boldsymbol{w} \rangle$ の最大値を与える．不等式の等号が成立するのは $(x_{M+1}, \ldots, x_N) = (0, \ldots, 0)$ の場合に限られるので，$\hat{\boldsymbol{x}}$ 以外の $\boldsymbol{x}$ は $\langle \boldsymbol{x}, \boldsymbol{w} \rangle$ の最大値を与えない．特に $w_1 > w_2 \ge \cdots \ge w_N$ の場合，$\hat{\boldsymbol{x}} = (1, 0, \ldots, 0)$ であり，これが $\langle \boldsymbol{x}, \boldsymbol{w} \rangle$ の最大値を与える唯一の $\boldsymbol{x}$ となる．