

ベース音高と和音特徴の統合に基づく和音系列認識

須見 康平^{†1,*1} 糸山 克寿^{†1} 吉井 和佳^{†1,*2}
駒谷 和範^{†1,*3} 尾形 哲也^{†1} 奥乃 博^{†1}

本論文では、ベース音高と和音の関連を考慮した和音系列認識の手法を提案する。和音系列とともに楽曲を構成する他の音楽的要素と深く関連しているため、それらの関係性を考慮して認識を行うことで精度の向上が期待される。我々は、ベース音が構成するメロディであるベースラインとの関係性に着目し、和音系列認識の重要な手がかりとして使用する。和音特徴、和音遷移パターンに加え、ベース音高分布に基づく和音系列の事後確率を定義し、事後確率が最大となる和音系列を Viterbi 探索することで求める。“The Beatles” の 12 アルバム中の、調を持ち転調を含まない 150 楽曲に対して行った実験の結果、和音に関する特徴のみを用いる場合よりもベース音高情報を付加した場合の方が精度は向上し、73.7%の平均認識率が得られた。

Automatic Chord Sequence Recognition Based on Integration of Chord and Bass Pitch Features

KOUHEI SUMI,^{†1,*1} KATSUTOSHI ITOYAMA,^{†1}
KAZUYOSHI YOSHII,^{†1,*2} KAZUNORI KOMATANI,^{†1,*3}
TETSUYA OGATA^{†1} and HIROSHI G. OKUNO^{†1}

This paper presents a method that identifies musical chords in polyphonic musical signals. As musical chords mainly represent harmony of music and are related to other musical elements such as melody and rhythm, performance of chord recognition should improve if this interrelationship is taken into consideration. In this paper, bass lines are utilized as clues for improving chord recognition. Our chord recognition system is constructed based on Viterbi-algorithm-based maximum *a posteriori* estimation which uses a posterior probability based on chord features, chord transition patterns, and bass pitch distributions. Experimental results with 150 songs which have scales and no modulation in twelve Beatles albums showed the recognition rate of 73.7% on average.

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ーの普及やそれにとま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。そのような情報の入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まなう。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、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である和音特徴、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

†1 京都大学大学院情報学研究所

Graduate School of Informatics, Kyoto University

*1 現在、ヤマハ株式会社

Presently with Yamaha Corporation

*2 現在、産業技術総合研究所

Presently with National Institute of Advanced Industrial Science and Technology (AIST)

*3 現在、名古屋大学大学院工学研究科

Presently with Graduate School of Engineering, Nagoya University

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。調和和音と同様にハーモニから派生する音楽的要素で、楽曲全体で中心的な役割を持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、様々な和音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 Profile⁵⁾ を和音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和音特徴を用いた手法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和音特徴を用いた手法¹³⁾ などが提案されている。これらの手法では、使用する和音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和音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。

本論文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和音特徴、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。以下、2 章ではベース音高を用いる動機を説明し、解決すべき課題とその解決法を述べる。3 章では和音系列認識システムの概要と問題設定、および和音系列事後確率について述べる。4 章で評価実験の結果と本手法の有効性を示し、5 章で結論を述べる。

2. ベース音高を用いた和音系列認識

我々は、ハーモニ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持つ。

- 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

1 章で述べたように、従来の仮説探索手法¹⁴⁾ における仮説信頼度ではベース音高の利用方法に問題点が残されていた。そのほか、和音系列認識システムを構築するにあたり、以下の課題を解決する必要がある。

2.2.1 ベース音高推定誤りによる認識率低下

従来研究¹⁴⁾ は、推定されたベース音高に応じて音楽的知識に基づくペナルティを与えることでベース音高に基づく和音系列仮説の評価を行っていた。そのベース音高推定手法は、入力されたパワースペクトルに対して各時刻でのベース音高候補の抽出とその候補の信頼性の計算を同時に行い最も信頼性の高いベース音高候補を時間的に接続するものであった。すなわち、各時刻で最も優勢なベース音高のみを用いており、たとえば残響などである時刻に複数のベース音が存在する場合にはベース音高誤りが起こりやすい。このような誤りは和音系列認識に悪影響を与える。

2.2.2 汎化能力の低い確率モデルによる経路探索手法への制約

従来手法¹⁴⁾ では、和音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

間内で平均した和音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和音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和音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）の違いに起因する和音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。

2.3 アプローチ

2.2 節の 2 つの課題を解決するために、ベース音高分布と時間独立性の高い確率モデルを導入する。以下にその詳細を述べる。

2.3.1 ベース音高分布の導入

ベース音高推定誤りによる認識性能劣化を抑制するため、本研究ではベースラインの情報としてベース音高分布を利用する。ベース音高分布とは、低周波数帯域の各音高におけるベース音の存在を確率的に表現したものである。

たとえば残響などによりある時刻に複数のベース音が存在した場合、本来その時刻に演奏している音高ではないベースラインが推定される場合がある。しかしながらこのような場合でも、ベース音高分布を用いると、本来その時刻に演奏している音高のベース音高分布もある程度大きい値を持つことが期待されるため、ベース音高推定誤りによる悪影響を緩和することが可能となる。

本研究でのベース音高の推定には、混合音中の優勢な調波構造の F_0 を推定する手法である PreFEst¹⁵⁾ を用いる。PreFEst は調波構造を表現する制約付き混合正規分布を音モデルとする。音源数を仮定せずにあらゆる音高に配置した音モデルの集合を混合分布と見なし、観測パワースペクトルを表現する混合分布の最適な重みを EM (Expectation-Maximization) アルゴリズムで推定する。音高の存在する範囲を低域に限定することで、低域で最も優勢な調波構造を持つベース音が各音高に存在する確率を推定することができる。また、PreFEst では任意の時刻でのベース音高分布を推定できるため、Viterbi 探索による和音系列推定を行う本システムでの利用が容易である。時間構造のモデル化¹⁶⁾、休符のモデル化¹⁷⁾、正弦波重畳によるモデル化^{18),19)}、スペクトルのピーク追従²⁰⁾ など、より詳細なモデルを用いた多重音解析手法が提案されており、これらの手法で得られたパラメータを適切に周辺化することで PreFEst と同等の音高分布を得ることは可能である。しかし、音符単位でベース音を推定するため任意の 2 時刻間でのベース音高の独立性を必ずしも仮定できない、モデルが複雑なため計算量が大きいため、必ずしも PreFEst を大きく上回るベース音高の認識性能を持つわけではない、などの理由から本論文ではこれらの手法は採用しない。

2.3.2 混合正規分布による和音特徴のモデル化

和音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) を用いて和音特徴の生成過程をモデル化する。和音特徴には 12 次元クロマベクトル²¹⁾ 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、和音特徴の長時間平均をとらなくても各時刻の和音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。また、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。

3. 和音系列認識システム

本章では、市販音楽 CD などから抽出した混合音の音楽音響信号を入力として、和音名系列および調の認識を行う和音系列認識システムについて述べる。本システムは、ベース音高推定で得られるベース音高分布を和音系列の事後確率計算に利用することで、従来の和音特徴と和音遷移パターンに基づく和音系列認識手法にベース音高推定を統合する。以下では、問題の定式化、和音特徴・ベース音高分布・和音遷移パターンを用いた和音系列事後確率の定義、および事後確率の計算方法について述べる。

3.1 問題の定式化

入力楽曲の音楽音響信号に対して、各時刻での和音特徴ベクトルの計算とベース音高分布の推定を行う。解析時刻に番号を割り当て、順に $[t_1, \dots, t_N]$ とする。この各解析時刻で求めた和音特徴ベクトル系列を $X = [x_1, \dots, x_N]$ 、ベース音高分布系列を $B = [b_1, \dots, b_N]$ とする。さらに、ビートトラッキング²²⁾ による八分音符時刻の推定を行う。この八分音符時刻を和音区間の最小単位とし、八分音符区間系列を $[e_1, \dots, e_K]$ とする。 e_1 の開始時刻は t_1 、 e_K の終了時刻は t_N と一致し、また各八分音符区間は少なくとも 1 つの解析時刻を含むものとする。

本システムの認識対象は、和音系列 $c = [c_1, \dots, c_K]$ (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。

$$c_k \in R \times \{\text{Major, Minor, Diminished, Sus4}\} \quad (1)$$

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

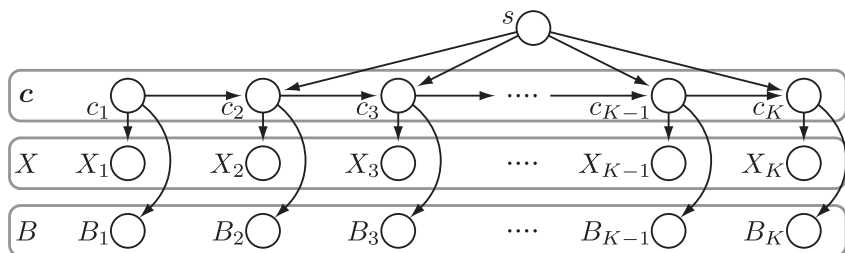


図1 和音名, 調, 和音特徴, ベース音高分布のグラフィカルモデル

Fig. 1 Graphical model of chord progression, scale, chord features, and bass pitch probabilities.

ただし, $R = \{C, C\#, D, D\#, E, F, F\#, G, G\#, A, A\#, B\}$ で, 12 種類のルート音の集合である. 和音の種類はポピュラー音楽で頻繁に使用される, Major (ルート音・長三度・完全五度), Minor (ルート音・短三度・完全五度), Diminished (ルート音・短三度・減五度), Sus4 (ルート音・完全四度・完全五度) の 4 種の三和音とした. 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.

和音系列と調に関する事後確率 $p(c, s|X, B)$ を考え, 与えられた X および B に対する最大事後確率 (Maximum A Posteriori: MAP) 推定問題として和音系列認識を定義する. すなわち, $\arg \max_{c,s} p(c, s|X, B)$ が求めるべき和音系列と調である.

3.2 事後確率の計算

MAP 推定のため, 和音名と調, 和音特徴, ベース音高分布の同時確率を以下で定義する.

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

X_k と B_k はそれぞれ八分音符区間 e_k で求めた X と B の部分列である. このグラフ表現を図 1 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) 和音特徴 X とベース音高分布 B は各八分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 c と調 s の事後確率は以下の等式を満たす.

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

ここで, $p(X, B)$ は MAP 推定には無関係なので, 実際には $p(c, s, X, B)$ 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.

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

3.2.1 和音特徴に基づく尤度: $p(X_k|c_k)$

和音特徴として, 2.3.2 項で述べたように 12 次元クロマベクトル²¹⁾を用いる. 主に和音を表現する伴奏音からの特徴抽出を行うために, クロマベクトルを求める周波数の範囲は 55–1,000 Hz とした.

クロマベクトルは, 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. また, 種類が同じでルート音だけが異なる和音は, 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どうしは, 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 2 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等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 を使い, 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. 学習データから求め, 巡回シフトさせたクロマベクトルを和音の種類ごとにまとめ, 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 $\{\alpha_{c,1}, \dots, \alpha_{c,M}\}$, $\{\mu_{c,1}, \dots, \mu_{c,M}\}$, $\{\Sigma_{c,1}, \dots, \Sigma_{c,M}\}$ 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

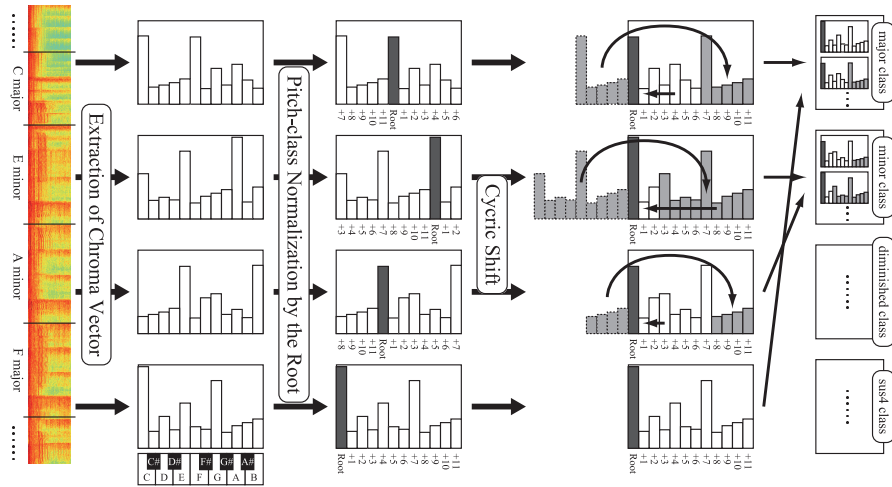


図2 クロマベクトルの巡回シフト．抽出されたクロマベクトルをルート音の音名で正規化し，ルート音のインデックスが等しくなるように並行移動し，あふれた要素を逆の端におくことで和音の種類ごとのベクトルが得られる
 Fig.2 Cyclic shift of chroma vector. By normalizing the indices of vectors by the root, then shifting and rotating the elements, the vectors for each code kind are obtained.

ロマベクトルに基づく尤度を計算する．この計算においては，ある時刻 t_n のクロマベクトル \mathbf{x}_n は t_n を含む八分音符区間 e_k の和音名 c_k にもみ依存することを仮定している．

$$p(X_k|c_k) = \prod_{n \in \{n|t_n \in e_k\}} p(\mathbf{x}_n|c_k) \quad (6)$$

$$p(\mathbf{x}_n|c_k) = \sum_{m=1}^M \frac{\alpha_{c_k,m}}{(2\pi)^6 |\Sigma_{c_k,m}|^{-\frac{1}{2}}} \exp\left(-\frac{(\mathbf{y}_n - \boldsymbol{\mu}_{c_k,m})^T \Sigma_{c_k,m}^{-1} (\mathbf{y}_n - \boldsymbol{\mu}_{c_k,m})}{2}\right) \quad (7)$$

\mathbf{y}_n は \mathbf{x}_n を含む和音区間の和音名 c_k のルート音に応じて \mathbf{x}_n を巡回シフトさせたベクトルである．

3.2.2 ベース音高分布に基づく尤度： $p(B_k|c_k)$

ベース音高分布系列 B は，各時刻で各音高にベースラインを演奏する楽器音が存在する確率を PreFEst¹⁵⁾ で推定したものである．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．時刻 t_n におけるベース音高分布 \mathbf{b}_n は， $[f_0, f_1]$ に含まれるパワースペクトルの周波数

帯域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．

$$n = 1, \dots, N : \left[\begin{array}{l} \mathbf{b}_n = (b_{n,M_0}, \dots, b_{n,M_1}); \sum_{m=M_0}^{M_1} b_{n,m} = 1; \\ m = M_0, \dots, M_1 : 0 \leq b_{n,m} \leq 1 \end{array} \right] \quad (8)$$

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 $\beta_c = (\beta_{c,M_0}, \dots, \beta_{c,M_1})$ 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学習する．

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) のベース音高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高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．

ベース音高分布に基づく尤度が上記の特徴を満たすよう，以下でこの尤度を定義する．

$$p(B_k|c_k) = \prod_{n \in \{n|t_n \in e_k\}} p(\mathbf{b}_n|c_k) \quad (9)$$

$$p(\mathbf{b}_n|c_k) = (M_1 - M_0 + 1)! \langle \mathbf{b}_n, \beta_{c_k} \rangle \quad (10)$$

$\langle \cdot, \cdot \rangle$ はベクトルの内積を表す演算子である．この確率密度関数のモードは

$$\begin{aligned} (b_{n,M_0}, \dots, b_{n,m}, \dots, b_{n,M_1}) &= (0, \dots, 1, \dots, 0) \\ \text{s.t. } m &= \arg \max_{m \in \{M_0, \dots, M_1\}} \{\beta_{c_k, M_0}, \dots, \beta_{c_k, M_1}\} \end{aligned} \quad (11)$$

であり，和音名 c_k で出現しやすい音高のベース音高がはっきり演奏されているとき，すなわち β_{c_k} で大きい値を持つ音高に関して \mathbf{b}_n も大きい値を持つときに大きい尤度を与える．付録でこの確率密度関数の正規化項およびモードについて述べる．

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

実楽曲ではジャンルやアーティストごとに，ある程度和音遷移のパターンが限定され，その遷移頻度には偏りがある．これは音楽理論に基づく制約であり，和音記号で与えられる頻出な遷移パターンを調を用いることで，具体的な和音名に変換することができる．楽曲中での和音名の出現頻度を近似することで，和音名の曖昧性を解消する．

和音遷移パターンに基づく事後確率 $p(c_k|c_{k-1}, s)$ を和音遷移 2-gram モデルを用いて計算する．また，最初の八分音符区間での和音名の条件付き確率 $p(c_1|s)$ は 2-gram から 1-gram

を求め、その 1-gram から計算する。学習時には、調に関してルート音を正規化した正解和音名系列から調の種類ごとに前向き 2-gram 確率を求める。2-gram 確率 $p(c_k|c_{k-1}, s)$ は、それぞれの遷移が出現した回数を前の和音名の全個数で割ったものとして算出する。2-gram が出現しないパターンには、微少な確率 (10^{-5}) の割当てと再度の確率正規化で確率の割当てを行う。

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およびベース音高推定のための PreFEst-core を実装し、評価実験を行った。

実験用データには、“The Beatles” の 12 枚の CD アルバムの全 180 曲中、調を持ち転調を含まない 150 曲を、サンプリング周波数 16 kHz のモノラル音響信号に変換したものを利用した。全楽曲を無作為に 36 曲ごとの 5 つの楽曲群に分け、5-fold Cross Validation を行った。STFT による音響信号分析における窓幅とシフト幅はそれぞれ 8,192 サンプル、1,024 サンプルとした。和音特徴 GMM の混合数は原則として $M = 4$ 、ベースラインの範囲は $[f_0, f_1] = [29, 261]$ [Hz] とした。GMM の学習データとして、入力に用いない残りの 4 グループ 144 曲、および YAMAHA 社製 MIDI 音源 MU-2000 で作成した 1,248 個の和音の音響信号を用いた。ここでは、音色を 6 通り、音域を 2 オクターブ、音響信号の切り出し位置を 3 通りに変化させて音響信号を作成した。また、ベース音高からのそれぞれの和音に対する確率の取得にも残りの 144 曲を使用し、和音遷移 2-gram の学習データとして実験用データ中の転調を含むものを除いた 137 曲を長調に、13 曲を短調に使用した。学習および評価において、Harte らが作成した和音名ラベルデータ²³⁾ を正解データとして使用した。認識結果を、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評価した。

$$\text{和音認識率} = \frac{\text{正解を出力した総区間長}}{\text{入力楽曲長}} \times 100 [\%] \quad (12)$$

以下ではそれぞれの実験の詳細な条件を述べ、実験結果について考察する。

4.1 有効性の検証

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) 和音特徴のみを用いて事後確率計算
- (2) 和音特徴とベース音高分布を用いて事後確率計算
- (3) 和音特徴と和音遷移パターンを用いて事後確率計算

表 1 音楽的要素の組合せの変化による和音認識率 [%] の変化

Table 1 Recognition rates [%] for each combination of musical elements.

(1): 和音特徴, (2): 和音特徴+ベース音高分布, (3): 和音特徴+和音遷移パターン, (4): 本手法, (5): 従来手法¹⁴⁾

手法	(1)	(2)	(3)	(4)	(5)
認識率	59.8	66.6	61.9	73.7	61.6

表 2 GMM の混合数の違いによる和音認識率 [%] の変化

Table 2 Recognition rates [%] for each number of GMM components.

混合数	1	2	4	8	16
認識率	61.0	67.9	73.7	72.6	66.9

- (4) 本手法: 和音特徴, 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¹⁴⁾: 単一正規分布でモデル化した和音特徴, 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、和音特徴を GMM でモデル化することの有効性を検証するため、GMM の混合数を 1, 2, 4, 8, 16 に変化させて実験を行った。GMM の混合数を 1 とした場合は、和音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 1, 表 2 に示す。

表 1 より、本手法での平均和音認識率は 73.7% となり、和音特徴に基づく評価値を単独で使用するよりも 13.9 ポイント高い結果を得た。また、ベース音高分布、和音遷移パターンをそれぞれ単独で付加することで 6.8 ポイント、2.1 ポイントの上昇が見られた。認識率が上がった要因として、和音特徴のみでは手がかりとして不十分であったが、和音に関連する要素であるベース音高を用いることでその情報を和音系列に対する有用な情報として扱うことが可能となったためである。さらに、和音特徴のモデル化を単一正規分布から混合正規分布に、ベース音高によるペナルティからベース音高分布による事後確率に、ルールベースの和音遷移パターンによるペナルティから学習データに基づく 2-gram に、それぞれ改良することで和音認識率は 12.1 ポイント向上した。ベース音高分布を利用する場合と音高そのものを利用する場合を直接比較するものではないが、ベース音高分布導入の有効性をある程度示すと考える。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 2 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、和音特徴を 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。

表 3 2枚のアルバムにおける手法ごとの平均認識率
Table 3 Average rates on two Beatles albums.

手法	Bello らの手法 ⁷⁾	Lee らの手法 ¹⁰⁾	本手法 (転調曲を除く場合)
CD1	68.6	64.0	73.0 (73.9)
CD2	81.5	84.8	82.5 (82.5)
平均	75.0	74.4	77.8 (78.3)

4.2 他手法との比較

HMM を利用した既存の和音認識手法である Bello らの手法⁷⁾ と Lee らの手法¹⁰⁾ では、今回使用した中の 2 枚のアルバム “Please Please Me” (CD1), “Beatles for Sale” (CD2) の楽曲を、評価実験に用いていた。本手法の結果と比較を行うことで、これらの手法との相違点を検証する。

表 3 に、各アルバムの平均認識率と 2 枚のアルバム全体の平均認識率を他手法の結果と並べて示す。本手法の認識率は、CD1 では他の 2 手法と比べて高く、CD2 では Bello らの手法より高く、Lee らの手法よりも低いものとなった。CD1 は “The Beatles” のファーストアルバムであり、和音遷移はブルースで典型的に用いられるパターンが多く、リズムも単純であったためであると考えられる。一方 CD2 では、14 楽曲のうち 2 曲に対して調の認識が誤っていた。調の誤認識と和音認識率の関係は次節で考察する。

CD1 の認識率と CD2 の認識率の差は Bello らの手法では 12.9 ポイント、Lee らの手法では 20.8 ポイントであり、これらの手法では認識率に偏りが見られる。一方本手法では、認識率の差は 9.5 ポイントでこれより小さく、認識性能が安定していることを示している。

4.3 調の認識

本論文での目的は和音系列の認識であり、調は和音認識のために補助的に推定されるものであるが、本手法では和音系列と調は同時に推定するものであるため、和音と調の認識は互いに影響を及ぼし、それぞれの認識精度に何らかの相関が現れると考えられる。そこで、和音系列の認識結果と調の認識結果について考察する。

表 4 は、調の認識結果である。本手法の評価実験における調の正解楽曲数は 106、認識率は 71.2% であった。調は和音遷移パターンをある程度限定するため、調が誤認識されている楽曲では和音の認識結果にも誤りが多いと考えられたが、実際にはそのような楽曲でも調が正しく認識された楽曲と同程度に和音が認識される事例が確認された。調を誤認識していた場合でも、その認識結果が本来の調の近親調である平行調、下屬調、属調であった場合は、和音遷移パターンが類似したものとなることが多いため、和音認識率への影響が少

表 4 調の認識結果

Table 4 Key recognition results.

分類	正解	不正解			
		平行調	属調	下屬調	その他
楽曲数	107	3	9	16	15

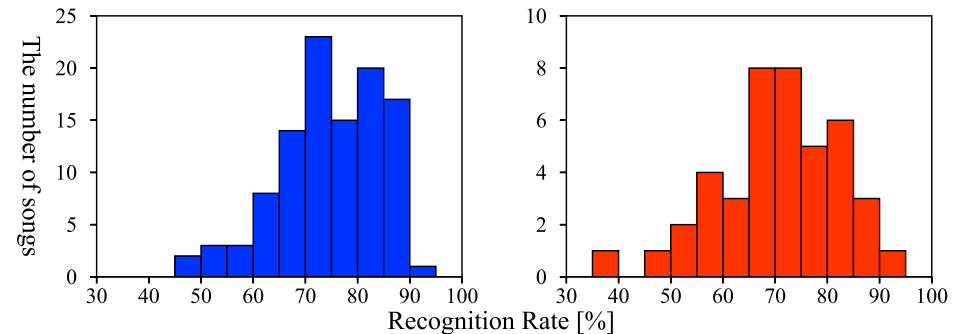


図 3 和音認識率のヒストグラム。左側は 150 曲中で調を正しく認識できたもの (107 曲)、右側は調を誤認識していたもの (43 曲)

Fig. 3 Accuracy Histograms. (Left) Histogram for the songs with correct key. (Right) Histogram for the songs with incorrect key.

ない。特に平行調への誤りでは、出現する和音が同じであるため、影響はほぼない。また下屬調、属調への誤りでは、出現率の高い和音に関して同じものも多く、やや影響を緩和できる。また、出現する和音の種類が極端に少ない場合にも、調の認識が和音系列認識にあまり影響しない。

調を正しく認識できた楽曲と誤っていた楽曲に対するそれぞれの和音認識率のヒストグラムを図 3 に示す。調を正しく認識できていた楽曲群ではおよそ 75% のあたりをピークに、調を誤認識していた楽曲群に対してはおよそ 70% のあたりをピークに、いずれも山型のヒストグラムになっているが、調を誤認識したものの方がばらつきが大きく、55% のあたりに調を正しく認識したものには見られない第 2 のピークが見られる。このことから、調の認識が和音の認識精度に二極化をもたらししていると考えられる。

和音遷移 2-gram が通常の 2 倍の重みを持つように確率を調整することで和音遷移パターンを重視するような事後確率計算を実験的に行ったところ、調が誤っていた楽曲ではヒストグラムがより広範囲に分布し、調を正しく認識できていた楽曲は全体的に和音認識率が上

表 5 調の有無・転調の有無による和音認識率の違い
Table 5 Difference of recognition rates in scales and transitions.

分類	調あり, 転調なし	調なしもしくは転調あり
認識率	73.7	60.6

昇する傾向が見られたが, 全楽曲の平均和音認識率は低下した. また, 認識率 55%未滿の調を誤認識した 4 曲について, あらかじめ調を与えて和音系列認識を行ったところ, 平均認識率が 44%から 68%に向上し, その全曲に対して和音遷移パターンによる向上が見られた. したがって, 調の認識精度は和音系列の認識精度に大きく影響するといえるため, 高い調の認識精度をとまなう枠組みへ拡張することが理想的である. さらに転調に対処するためには, 調の変化を追従可能な探索方法を検討する必要がある.

また, 実験的に転調を含む楽曲, 調を持たない楽曲に対して本手法で和音認識を行った結果を表 5 に示す. 本手法では転調を含む楽曲や調なしの楽曲を想定しておらず, このような楽曲に対しては和音遷移パターンの利用が適切に行えない, 転調が起こる箇所からそこまでの調との不一致が起こる, などの理由により, 楽曲全体の認識率が低下したと考えられる.

5. おわりに

本論文では, 従来の研究で考慮されていなかった音楽的要素間の相互関連に着目して, 和音とベース音の関係に着目した認識手法を考案し, 実楽曲の音楽音響信号を対象とした評価実験によるその有効性の検証を行った. 和音特徴, ベース音高分布, 和音遷移パターン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. 評価実験により, 和音特徴, ベース音高分布, 和音遷移パターンの確率的統合に基づく事後確率を用いた本手法では, “The Beatles” の 12 アルバム全 180 楽曲に対して 71.5%の平均認識率を獲得した. また, 和音特徴とベース音高分布を事後確率計算に用いた場合の方が, 和音特徴のみを事後確率として用いた場合と比較して認識率が 6.7 ポイント向上した. これより, 和音系列認識におけるベース音高分布利用の有効性が示され, 音楽的要素の確率的な統合に基づく手法の妥当性が示された.

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), JST CrestMuse の支援を受けた. また, 仮説探索手法についてご討論いただいた吉岡拓也氏, “The Beatles” の和音系列正解データをご提供いただいた Christopher Harte 氏に感謝します.

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, *ICASSP2003*, 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, *AMCMM'06*, 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, *Machine Learning*, 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, 情報処理学会研究報告, Vol.2008, No.78 (2008-MUS-76), pp.137-142 (2008).
- 14) 吉岡拓也, 北原鉄朗, 駒谷和範, 尾形哲也, 奥乃 博: 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, 情報処理学会研究報告, Vol.2004, No.84 (2004-MUS-56), pp.33-40 (2004).
- 15) 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).
- 16) Kameoka, H., Nishimoto, T. and Sagayama, S.: A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering, *IEEE Trans. Audio, Speech and Lang. Process.*, Vol.15, No.3, pp.982-994 (2007).
- 17) Rynänen, M. and Klapuri, A.: Automatic Bass Line Transcription from Streaming Polyphonic Audio, *ICASSP2007*, pp.1437-1440 (2007).
- 18) Davy, M., Godsill, S. and Idier, J.: Bayesian Analysis of Polyphonic Western Tonal Music, *J. Acoust. Soc. Am.*, Vol.119, No.4, pp.2498-2517 (2006).
- 19) Cemgil, A.T., Kappen, H.J. and Barver, D.: A Generative Model for Music Transcription, *IEEE Trans. Audio, Speech and Lang. Process.*, Vol.14, No.2, pp.679-694 (2006).
- 20) Yeh, C. and Röbel, A.: A New Score Function for Joint Evaluation of Multiple F0 Hypothesis, *DAFx-04* (2004).
- 21) Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R.: RWC Music Database: Music Genre Database and Musical Instrument Sound Database, *ISMIR2003*, pp.229-230 (2003).
- 22) 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).
- 23) 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 w と確率変数ベクトル x を考える. x は $0 \leq x_n \leq 1 (1 \leq n \leq N)$, $\sum_{n=1}^N x_n = 1$ を, w も x と同様の制約を満たす. 正規化係数を除き, x の確率密度関数を $p(x|w) \propto \langle w, x \rangle$ 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.

$$\begin{aligned} & \int_{\left\{x \mid \forall n=1, \dots, N: 0 \leq x_n \leq 1; \sum_{n=1}^N x_n = 1\right\}} \sum_{n=1}^N w_n x_n dx \\ &= \sum_{n=1}^N \int_0^1 \cdots \int_0^{1-\sum_{m=1}^{N-2} x_m} w_n x_n dx_{N-1} \cdots dx_1 \end{aligned} \quad (13)$$

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

$$\begin{aligned} & \int_0^1 \cdots \int_0^{1-\sum_{m=1}^{N-2} x_m} w_n x_n dx_{N-1} \cdots dx_1 \\ &= \int_0^1 \cdots \int_0^{1-\sum_{m=1}^{N-3} x_m} w_n x_n \left(1 - \sum_{m=1}^{N-2} x_m\right) dx_{N-2} \cdots dx_1 \\ &= \cdots = \int_0^1 \cdots \int_0^{1-\sum_{m=1}^{n-1} x_m} \frac{w_n}{(N-n-1)!} x_n \left(1 - \sum_{m=1}^n x_m\right)^{N-n-1} dx_n \cdots dx_1 \\ &= \int_0^1 \cdots \int_0^{1-\sum_{m=1}^{n-2} x_m} \frac{w_n}{(N-n+1)!} \left(1 - \sum_{m=1}^{n-1} x_m\right)^{N-n+1} dx_{n-1} \cdots dx_1 \\ &= \cdots = \frac{w_n}{N!} \end{aligned} \quad (14)$$

$n = N$ のとき,

$$\begin{aligned} & \int_0^1 \cdots \int_0^{1-\sum_{m=1}^{N-2} x_m} w_N x_N dx_{N-1} \cdots dx_1 \\ &= \int_0^1 \cdots \int_0^{1-\sum_{m=1}^{N-2} x_m} w_N \left(1 - \sum_{m=1}^{N-1} x_m\right) dx_{N-1} \cdots dx_1 \\ &= \cdots = \frac{w_N}{N!} \end{aligned} \quad (15)$$

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 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 (11) で与えられることを示す. 簡単のため $w_1 = w_2 = \cdots = w_M > w_{M+1} \geq \cdots \geq w_N$ であるとする. $\hat{x} = (\hat{x}_1, \dots, \hat{x}_M, 0, \dots, 0)$, $\hat{x}_m \geq 0 (m = 1, \dots, M)$, $\sum_{m=1}^M \hat{x}_m = 1$ とすると, 以下の不等式が成り立つ.

$$\langle \mathbf{w}, \hat{\mathbf{x}} \rangle - \langle \mathbf{w}, \mathbf{x} \rangle = w_1 - \sum_{n=1}^N w_n x_n = \sum_{n=1}^N (w_1 - w_n) x_n \geq 0 \quad (16)$$

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

(平成 22 年 6 月 26 日受付)

(平成 23 年 1 月 14 日採録)



須見 康平 (正会員)

2008 年京都大学工学部情報学科卒業。2010 年同大学大学院情報学研究科知能情報学専攻修士課程修了。在学中は音楽情報処理、音響信号処理等の研究に従事。現在、ヤマハ株式会社勤務。



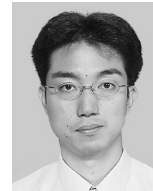
糸山 克寿 (学生会員)

2006 年京都大学工学部情報学科卒業。2008 年同大学大学院情報学研究科知能情報学専攻修士課程修了。現在、同大学院博士後期課程に在学中。2008 年より日本学術振興会特別研究員 (DC1)。音楽情報処理、音楽鑑賞インタフェース等の研究に従事。



吉井 和佳 (正会員)

2008 年 京都大学大学院情報学研究科知能情報学専攻博士後期課程修了。博士 (情報学)。同年産業技術総合研究所に入所し、現在に至る。音楽推薦システムや複数基本周波数推定等、機械学習に基づく音楽情報処理の研究に従事。山下記念研究賞、船井研究奨励賞等受賞。



駒谷 和範 (正会員)

1998 年京都大学工学部情報工学科卒業。2000 年同大学大学院情報学研究科知能情報学専攻修士課程修了。2002 年同大学院博士後期課程修了。京都大学博士 (情報学)。同年京都大学大学院情報学研究科助手。2007 年同助教。2010 年より名古屋大学大学院工学研究科准教授。主に音声対話システムの研究に従事。2008 年から 2009 年まで米国カーネギーメロン大学客員研究員。情報処理学会平成 16 年度山下記念研究賞、FIT2002 ヤングリサーチャー賞等を受賞。電子情報通信学会、言語処理学会、人工知能学会、ACL、ISCA 各会員。



尾形 哲也 (正会員)

1993 年早稲田大学理工学部機械工学科卒業。日本学術振興会特別研究員、早稲田大学理工学部助手、理化学研究所脳科学総合研究センター研究員、京都大学大学院情報学研究科講師を経て、2005 年より同助教 (現、准教授)。博士 (工学)。JST さきがけ研究「情報環境と人」領域研究員 (5 年)。この間、早稲田大学ヒューマノイド研究所客員准教授、同大学理工学研究所客員准教授、理化学研究所脳科学総合研究センター客員研究員等を兼務。研究分野は人工神経回路モデルおよび人間とロボットのコミュニケーション発達を考えるインタラクション創発システム情報学。日本ロボット学会、日本機械学会、人工知能学会、計測自動制御学会、ヒューマンインタフェース学会、バイオメカニズム学会、IEEE 等各会員。



奥乃 博 (正会員)

1972 年東京大学教養学部基礎科学科卒業。日本電信電話公社、NTT、JST、東京理科大学を経て、2001 年より京都大学大学院情報学研究科知能情報学専攻教授。博士 (工学)。この間、スタンフォード大学客員研究員、東京大学工学部客員助教授。人工知能、音環境理解、ロボット聴覚、音楽情報処理の研究に従事。1990 年度人工知能学会論文賞、第 2 回船井情報科学振興賞、IEEE/RSJ IROS-2010 NTF Award for Entertainment Robots and Systems 等受賞。本学会理事。