

## スペクトルディップテンプレートを用いた和音推定

## Chord Estimation using Spectrum templates

宮田 聡      上野 佑馬      蔵内雄貴      松原 正樹      斎藤 博昭  
 Satoshi Miyata    Yuuma Ueno    Yuki Kurauchi    Masaki Matsubara    Hiroaki Saito

慶應義塾大学大学院 理工学研究科

E-mail:miyata@nak.ics.keio.ac.jp

## 1 はじめに

和音<sup>1</sup>の移り変わりは音楽において大変重要な意味を持つ。従って和音の移り変わりを分析することは楽曲の構造を理解することにつながる重要なタスクである。しかし和音推定の現状としては、音楽音響信号<sup>2</sup>を対象とした場合にはまだ難しく、その前段階として主に多重音の音響信号<sup>3</sup>を対象として盛んに研究されている。多重音の音響信号を対象とした和音推定には大きく分けて2種類の手法がある。1つは構成音全ての音高を求めてから和音推定を行う手法であり、もう1つはスペクトルの特徴を手がかりとして直接和音推定を行う手法である。現在の研究では、クロマベクトル<sup>4</sup>を用いたり、機械学習を組み合わせる方法が主流である。構成音全ての音高を求めてから和音推定を行う手法の場合、ある音高の基音成分と他の音高の倍音成分が重なると判別が難しいため、同時発音数が増えると精度が下がるなどの問題点がある。そのため、3重音を対象とした研究は多くなされているが、4重音以上を対象とした研究はあまりされていない。一方、スペクトルの特徴を手がかりとして直接和音推定を行う手法は、コードの基音がもっとも大きいと仮定しているほか、ベース音が存在しなければならぬなど制約が多い。

そこで、本研究では4重音以上を対象とした和音推定を実現すべく、スペクトルの特徴を手がかりとして直接和音推定を行う手法を提案する。

## 2 関連研究

スペクトルの特徴を手がかりとして和音推定を行う手法には、クロマベクトルを用いたテンプレートマッチングによる手法とコード進行などの音楽的知識を機械学習でモデル化して推定する手法がある。Harteらは、クロマベクトルを拡張しより細かく音を分割した36次元のクロマベクトルを用いて48種類の定義され

た和音テンプレートと比較することで、音楽音響信号において平均して62.4%の精度で推定できた[1]。しかし、重音数が増えていった時、上記のいずれの手法も、構成音の組み合わせや楽器特有の高調波構造次第で、ほぼ同じ構成音を含む別の和音やエンハーモニックの和音として誤推定することが多いという問題点がある。

そこで、より精度向上のために機械学習を組み合わせた手法が提案された。ShehらはEMアルゴリズムを用いてHMMのパラメータを学習し、和音の進行モデルを構築し、24次元のクロマベクトルを組み合わせることで約75%の精度で推定できた[2]。Leeらは和音の存在確率をHMMを用いて学習し和音推定する手法[3]や和音の進行の遷移確率をHMMで学習し和音推定する手法[4]、調の推定をしてから和音推定する手法を提案した[5]。AshleyらはShehらの手法において使用されているHMMの代わりにCRFを用い、精度が向上した[6]。

蔵内らは理論上パワーのない周波数(谷状点)に着眼した手法を用いている。入力は無雑音下で録音したサンプリング音響信号で、12音のコード種(メジャー、マイナー)のいずれかを推定している[7]。また、楽器の演奏方法や弾き方に対する頑健性を調べていて、音程が間違った演奏をされない限り、どのような演奏方法であっても結果は同一であることを示している。蔵内らは一つの谷状点で推定を行っていたが、本研究では複数の谷状点を考慮した手法を提案する。

## 3 提案手法

### 3.1 提案手法の理論的背景

蔵内らの手法である一つの谷状点<sup>5</sup>に着目するという点だけでは、様々なコード(主に3和音や4和音)に対応できないことがわかっている。スペクトルディップとは、和音の定義から導かれる理論上でパワーのない点のことである。そこで、スペクトルディップを複数考慮することで、精度が向上すると考える。また、倍

<sup>1</sup>複数の音が同時に鳴った音

<sup>2</sup>音響信号のうち、情報の内容が楽曲であるもの

<sup>3</sup>音の波形情報

<sup>4</sup>音名ごとにパワーを足し合わせた12次元のベクトル

<sup>5</sup>本論文では、スペクトルディップと呼ぶことにする

音を  $1/n$  モデルだけでなく、ギターモデルも用意した。倍音のモデルは以下のような定義式になる。

基音のパワーを  $P_0$  とすると、 $1/n$  モデルにおける  $n$  倍音のパワー  $P_n$  は

$$P_n = \frac{P_0}{n} \quad (1)$$

となり、ギターモデルにおける  $n$  倍音のパワー  $P'_n$  は

$$P'_n = \frac{\sin\beta n\pi}{\beta(1-\beta)} \frac{P_0}{n} \quad (2)$$

と定義される。また、 $\beta$  とは弦を押さえる位置から弦の最後までを  $L$ 、弦を弾く位置から弦の最後までを  $l$  とすると  $\beta = l/L$  で表される。 $\beta$  の値を整数  $n$  とすると、 $n$  倍音のパワーが 0 になってしまい、一般性が損なわれてしまうため、今回は、 $\beta = 5.5$  とした。以下に、 $1/n$  モデルのスペクトルおよび  $1/n$  モデルとギターモデルと実際のギターの和音のスペクトルとを比較した図を示す。(図 1, 図 2)

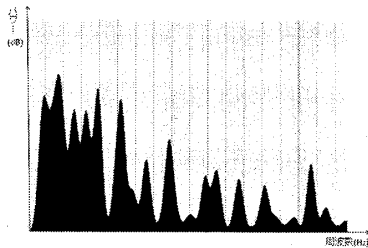


図 1:  $1/n$  モデルのスペクトル (C コード)

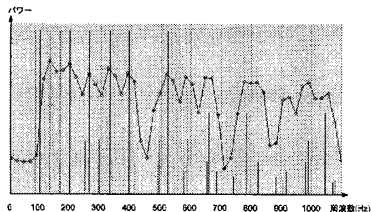


図 2: ギターモデルと実際のギター音源のスペクトルの比較 (C コード)

倍音が鳴っていると考えられる部分に正規分布関数を掛けると、実際のスペクトルの概形に近いものが得られたので、各コードにおけるスペクトルのモデルはこれを使用した。次に、 $1/n$  モデル及び、ギターモデルにおける C コードのスペクトルディップ系列を比較するために、スペクトルディップが存在する周波数を表 1 と表 2 に示す。

この手法により、より複雑な和音(メジャー、マイナー、ディミニッシュ、オーギュメントなど)の和音推定が可能となると考える。

表 1:  $1/n$  モデルにおける C コードのスペクトルディップ系列

コード	周波数	Hz						
C	430	720	860	1110	1230	1410	1520	1900
Cm	350	420	720	840	990	1130	1500	1630
Cdim	430	690	880	990	1150	1360	1620	1770
Caug	560	700	870	1100	1370	1530	1740	1940

表 2: ギターモデルにおける C コードのスペクトルディップ系列

コード	周波数	Hz						
C	440	720	850	1110	1220	1430	1620	1920
Cm	730	840	990	1130	1330	1520	1660	1910
Cdim	430	840	990	1140	1350	1610	1910	
Caug	460	580	700	880	970	1110	1260	1400

### 3.2 前処理部

事前に被験者一人が全てのギターコード 144 種について、理論上のスペクトルからスペクトルディップ系列を抽出する。抽出に使用したモデルの一例を図 3 に示した。次に、どのような基準でスペクトルディップを判定したかを述べる。スペクトルディップの数は、どのコードにおいても 7 個以上 10 個以内という規則を設けた。明らかに谷が深いスペクトルディップから選び、判定しづらいスペクトルディップが複数存在した場合、周りの谷との位置関係を考慮する。また、大きな窪みの場合は谷の真ん中を選ぶとする。全コードのスペクトルディップテンプレートは本稿末の付録に添付した。

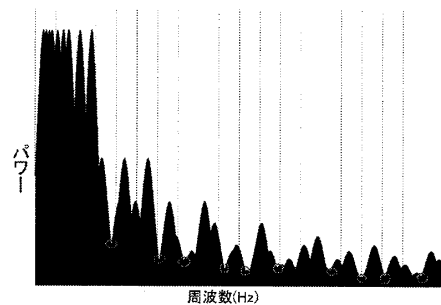


図 3: スペクトルディップの位置

### 3.3 算出部

各コードにおいて、それぞれのスペクトルディップにおける周波数のときのパワーの平均値を算出する。以下に、その方法を述べる。

まず、前処理部で抽出したスペクトルディップ系列と入力した実際のコードからなるスペクトルを比較して、それぞれのスペクトルディップの周波数のときのパワーを合計する。スペクトルディップの数はコードによって異なるため、それぞれのスペクトルディップの個数でパワーを割り相加重平均を求める。

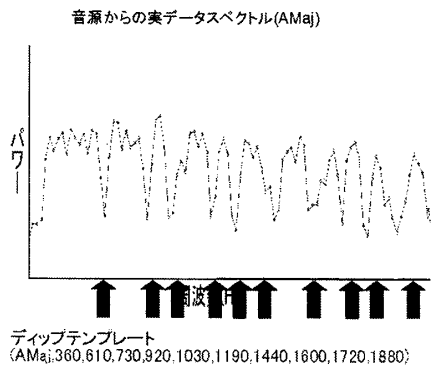


図 4: 実音源とテンプレートとの比較 (AMaj)

### 3.4 出力部

音が鳴り始めてから鳴り終わるまでの間、約 0.012 秒ごとに毎回算出部より得られた結果をパワーの平均値が低いものから上位 3 つを出力コードの候補として表示する。

以上の手法を本論文では「スペクトルディップ法」といい、和音の各スペクトルに対して和音推定を行うことができる。

## 4 実験結果

以下に、実験及びその結果を載せる。

### 4.1 実験

実験は 2 種類行った。実験の条件として、入力はいくつコードにおけるメジャー、マイナー、ディミニッシュ、オーギュメントを対象とした。この入力に対し、各コード 12 種類のメジャー、マイナー、ディミニッシュ、オーギュメント計 48 種類のうちどれであるかを推定した。また、転回形や複数の和音の推定をする必要は無いことは既に示されている [8]。

### 4.2 実験 1

蔵内らの手法をもとに、スペクトルディップ系列を導入して行った。

### 4.3 実験 2

ディップを求めるためのモデルを  $1/n$  モデルではなく、ギターモデルを考慮したスペクトルディップ系列で行った。

## 4.4 実験結果

表 3 に、実験 1 および実験 2 と既存手法との比較を示す。

表 3: 従来手法との比較

手法	Maj	Min	dim	aug	平均正解率
蔵内ら	91.6%	75.0%			83.3%
$1/n$ モデル	100%	91.6%	58.3%	100%	83.3%
ギターモデル	100%	100%	91.6%	100%	97.9%

## 5 考察

スペクトルディップ系列とギターモデルを導入した提案手法は、従来手法に比べても高い精度で和音推定できているといえる。スペクトルディップ系列は人間の目で見ただけのものである。スペクトルディップの数が少な過ぎた場合や、ひとつでも誤差の大きいスペクトルディップを選んでしまった場合、パワー平均値への影響が大きくなってしまふ。多過ぎた場合、余計なスペクトルディップを選んでしまう確率も増すため、パワー平均値に悪影響を及ぼすことは明白である。スペクトルディップの個数を 7 個から 10 個程度とするのが、最も良い精度を出した。実験結果から、対象コード数を増やしたにも関わらず、高い精度で和音推定が出来ていることが示せた。他のシステムに応用するためにはさらなる精度向上及び、より多くの和音推定が出来ることが求められる。

## 6 結論

実験結果からわかるように、モデルの改良とスペクトルディップ系列を用いることによって、先行研究よりも精度の向上ができた。しかし、コードを入力する際に遷移的に毎回出力結果が変わってしまう、つまり、音の鳴り始めと鳴り終わりなどで極端にスペクトルが異なるために和音が最も鳴っているとき以外は望ましい結果が得られないことが欠点である。音が鳴り始めてから鳴り終わるまでの結果を遷移的に表示するとともに、最終的な和音推定を出力する必要性があったが、将来的に自動伴奏や自動採譜のシステムに活用するとすると、リアルタイムでその和音を的確に出力しなくてはならないため、これは今後の課題となる。また、スペクトルディップ系列の作成を自動化できれば、本システムはさらに向上できるはずである。今回は、ギターだけの音源を使用した。将来的にはバンド音源などの実音楽や、音高が開離配置された場合であってもコード推定が出来るようにするなど、解決すべき課題は多い。

## 謝辞

この研究は科学技術振興機構平成21年度シーズ発掘A課題番号04-187の補助金の下に行われたものである。

## 参考文献

- [1] Christopher A. Harte and Mark B. Sandler. :Automatic chord identification using a quantised chromagram, *Proceedings of the Audio Engineering Society, 2005*.
- [2] Alexander Sheh and Daniel P.W. Ellis. :Chord Segmentation and Recognition using EM-Trained Hidden Markov Models, *Proceedings of the International Conference on Music Information Retrieval 2003*.
- [3] Kyogu Lee and Malcolm Slaney. :Automatic Chord Recognition from Audio Using an HMM with Supervised Learning, *Proceedings of the International Conference on Music Information Retrieval 2006*.
- [4] Kyogu Lee and Malcolm Slaney. :A Unified System for Chord Transcription and Key Extraction Using Hidden Markov Models, *Proceedings of the International Conference on Music Information Retrieval 2007*.
- [5] Kyogu Lee and Malcolm Slaney. :Acoustic Chord Transcription and Key Extraction From Audio Using Key-Dependent HMMs Trained on Synthesized Audio, *IEEE Transactions on Audio Speech and Language Processing, Vol. 16, No. 2, pp. 291-301, 2008*.
- [6] John Ashley Burgoyne, Laurent Pugin, Corey Lereliuk and Ichiro Fujinaga. :A Cross-Validated Study of Modelling Strategies for Automatic Chord Recognition in Audio, *Proceedings of the International Conference on Music Information Retrieval 2007*.
- [7] 蔵内雄貴, 松原正樹, 大野将樹, 斉藤博昭. 周波数スペクトルの谷状点に基づく和音推定. 第7回情報科学技術フォーラム. 2008.
- [8] 蔵内雄貴, 松原正樹, 大野将樹, 斉藤博昭. 周波数スペクトルの谷状系列による和音推定. 第76回音楽情報科学研究会. 2008.
- [9] 柏野邦夫, 中臺一博, 木下智義, 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. 電子情報通信学会論文誌, Vol. J79-D, No. 11, pp. 1762-1770. 1996.
- [10] 柏野邦夫, 中臺一博, 木下智義, 田中英彦. 適応型混合テンプレートを用いた音源同定. 電子情報通信学会論文誌, Vol. J81-D, No. 7, pp. 1510-1517. 1998.
- [11] 彦坂健太郎, 椿雅也, 谷口徹, 菅田雅彰, 白井克彦. 可変長スペクトルセグメントマッチングに基づく楽器の音高・楽器推定. 日本音響学会 2005 年春季研究発表会, No. 3-7-17, pp. 599-600. 2005
- [12] Giordano Cabral, Jean-Pierre Briot and Francois Pachet. :Impact of Distance in Pitch Class Profile Computation *Proceedings of the 10th Brazilian Symposium on Computer Music (SBCM's 2005)*.

コード	(Hz)							
C#	460	760	920	1180	1110	1290	1500	1620
C#m	350	450	770	890	1200	1600	1710	
C#dim	450	730	930	1040	1440	1610	1720	1880
C#aug	390	490	610	740	930	1160	1460	1620
D	500	800	960	1250	1360	1570	1710	1810
Dm	390	480	830	930	1270	1700	1810	
Ddim	480	670	770	980	1100	1280	1520	1710
Daug	410	510	650	780	980	1230	1550	1710
D#	520	750	850	1020	1130	1320	1450	1670
D#m	420	520	880	990	1350	1800	1920	
D#dim	700	820	1050	1150	1360	1620	1920	
D#aug	430	540	670	830	1050	1300	1430	1640
E	550	780	900	1080	1190	1280	1400	1530
Em	450	540	690	930	1040	1280	1430	1530
Edim	620	750	860	1020	1220	1430	1700	
Eaug	360	450	570	700	880	1100	1380	1520
F	580	730	830	950	1140	1260	1360	1470
Fm	460	570	730	990	1100	1350	1510	1620
Fdim	450	560	780	920	1080	1170	1290	1530
Faug	380	480	610	740	930	1160	1460	1620
F#	510	620	780	880	1000	1210	1330	1440
F#m	500	610	780	1050	1160	1450	1610	1720
F#dim	480	610	830	960	1150	1240	1360	1620
F#aug	410	510	630	780	980	1220	1550	1710
G	540	650	830	930	1060	1280	1420	1520
Gm	530	650	830	1120	1220	1530	1710	1820
Gdim	510	640	740	880	1020	1220	1320	1450
Gaug	430	540	670	830	1050	1300	1440	1640
G#	350	560	690	870	990	1360	1500	1620
G#m	560	670	880	1180	1300	1620	1810	1930
G#dim	340	540	680	780	930	1080	1290	1400
G#aug	450	570	700	870	1100	1380	1520	1620
A	370	610	730	930	1030	1190	1440	1600
Am	610	720	950	1250	1360	1620	1920	
Adim	360	560	730	830	1000	1150	1360	1490
Aaug	390	480	610	740	930	1150	1450	1620
A#	400	630	760	980	1100	1350	1530	1700
A#m	380	640	750	1010	1330	1450	1720	
A#dim	370	620	780	880	1020	1210	1450	1580
A#aug	410	520	650	780	980	1230	1550	1710
B	420	680	820	1050	1170	1330	1430	1630
Bm	410	680	800	1060	1420	1530	1820	
Bdim	650	820	940	1080	1280	1530	1670	1820
Baug	430	540	670	830	1050	1300	1430	1640

付録：ギターモデルにおける各コードのスペクトルディップ系列