RE-001

周波数スペクトルの谷状点に基づく和音推定

Chords estimation based on valley point of frequency spectrum

蔵内 雄貴[0] Yuki Kurauchi　松原 正樹[0] Masaki Matsubara　大野 将樹[0] Masaki Oono　斎藤 博昭[0] Hiroaki Saito

1 はじめに

和音推定は，モノラルの多重音[1] の音響信号[2] を対象として盛んに研究されている [1]．この和音推定には大きく分けて 2 つの手法がある．1 つは和音の構成音全ての音高を推定してから和音推定する手法であり，もう 1 つは周波数スペクトル (以下，スペクトル) の特徴を手がかりとして和音推定する手法である．前者は重音数が増えると精度が下がる問題点があり，5 重音や 6 重音には適用できない．一方，後者は重音数が増えても適応できる可能性があるが，精度が低い．

CD などの音楽音響信号を対象とした和音推定を目指す場合，重音数が多くても高精度で推定できる必要があるため，本研究では，スペクトルの特徴を手がかりとして和音推定する手法に注目する．この手法の精度が低い理由は，手がかりとする特徴の数が少ないためであり，精度を高めるには別の特徴量を用いる必要性があると考えた．本研究では，新たな特徴量としてパワーの小さな周波数を用い，それを手がかりとして和音推定する手法を提案する．

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．クロマベクトルとは，音名ごとにパワーを足し合わせた 12 次元のベクトルのことである．シンプルな方法にもかかわらず，楽曲のジャンルなどに関係なくある程度の精度で和音推定を行えるのが特徴である．

藤島らは，和音のテンプレートとクロマベクトルを比較することで和音推定を行い，単一楽器の音響信号において正しく和音推定できた [2]．我山らは，クロマベクトルを用いた和音推定により，音楽音響信号を対象とした類似楽曲の検索を行った [3]．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 [4]．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多数提案された．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た．そして 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 [5]．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 [6] 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 [7]，調の推定をしてから和音推定する手法を提案した [8]．Ashley らは Sheh らの手法のうち HMM の代わりに CRF を用いる事で精度の向上を試みた [9]．いずれの手法もクロマベクトルを用いて機械学習と組み合わせた手法であるが，クロマベクトルに代わる良いスペクトル特徴量を用いることで更なる精度の向上が期待できる．

そこで本研究ではクロマベクトルと同様なシンプルな手法としてスペクトルの谷状点に着目する和音推定手法を提案し，クロマベクトルよりも良い精度で和音推定できることを示す．本研究により，HMM や CRF を組み合わせた手法の精度の向上が期待できる．

3 提案手法

根音 (ルート音) 推定部，コード種推定部，出力部の 3 つのステップからなる．本研究ではコードの録音にギターの音を用いている．ギターは 5 重音や 6 重音における各コードが一般的に定義されているためである．ギターは平均律で調律を行ってあるものとする．

3.1 根音推定部

12 種の音名のうちいずれであるかの推定を行う．

各コードのスペクトルには，相対的にパワーの小さな点がある．本研究ではスペクトルのパワーの平均値よりもパワーの小さな点を谷状点と呼び，周波数の低い方から第一谷状点，第二谷状点，…，第 n 谷状点と呼ぶ (図 1)．

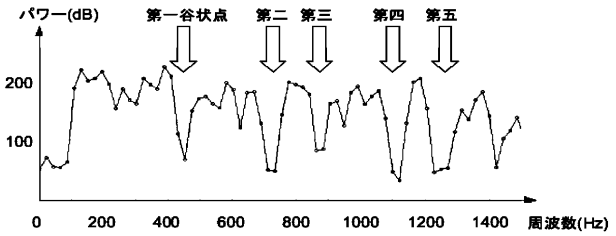

図 1: ギターの C コードにおける谷状点

この谷状点は，各コードで独自の周波数に安定して存在し，時間変化も少ない．そこで，第一谷状点の周

[0]慶應義塾大学大学院 理工学研究科, Keio University Graduate School

[1]複数の単音が同時に鳴った音のこととする

[2]音の波形情報

波数を特徴量として抽出する．その際，第二谷状点の周波数が第一谷状点の周波数の約 1.5 倍であり，第三谷状点の周波数が第一谷状点の周波数の約 2 倍であることを利用する．しかし，コードによっては正確に 1.5 倍や 2 倍の周波数ではないため，ずれを許容する手法を用いる．本研究では正規分布を 3 つ連ねたような形をしたフィルタ $F(x,\mu)$ を用いる (式 (1)，図 2)．σ^2 は分散の値であり，μ がフィルタの最初の極大点となる値である．2 つ目の極大点は μ の 1.5 倍，3 つ目の極大点は μ の 2 倍となっている．

$$F(x,\mu) = \frac{1}{3\sqrt{2\pi}\sigma}\left\{\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) + \exp\left(-\frac{(x-1.5\mu)^2}{2\sigma^2}\right) + \exp\left(-\frac{(x-2\mu)^2}{2\sigma^2}\right)\right\} \quad (1)$$

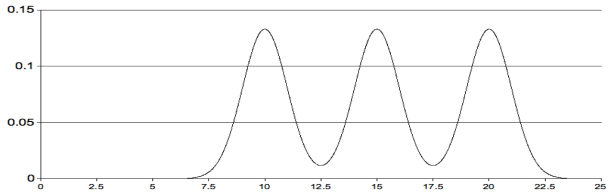

図 2: 用いるフィルタの例 ($\mu = 10$ の場合)

このフィルタ $F(x,\mu)$ にスペクトル $S(x)$ を掛けて足し合わせ，それを最小とするような μ を特徴量として抽出し (式 (2))，その値によって根音を推定する．

$$\arg\min_{\mu} \sum_{x} F(x,\mu)S(x) \quad (2)$$

3.2 コード種推定部

メジャーかマイナーコードどちらであるかの推定を行う．

まず，予備実験により 12 種のコードそれぞれのメジャーとマイナーコードのスペクトルを比較し，スペクトルの形に差のある周波数を見つけ (図 3)，その点のパワーの値の大小により推定する．

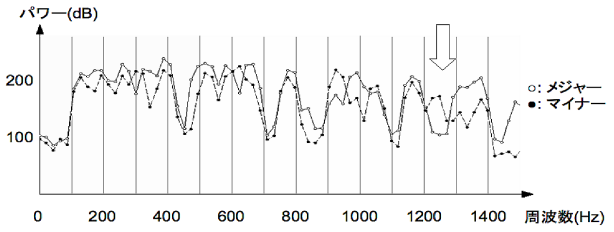

図 3: C コードと Cm コードのスペクトルの重ね合わせ

図 3 の例では，第五谷状点にあたる約 1250Hz のパワーの値が C コードでは 100dB 程度，Cm コードでは 180dB 程度と大きな差があり，時間変化も小さい．そのため，C コードと Cm コードの推定はこの周波数のパワーの値を用いる．

3.3 出力部

以上の手法でひとつのスペクトルに対して和音推定が行われる．つまり，高速フーリエ変換 (FFT) の際に，データが窓関数で切り出されるごとに推定結果が出力される．そのため，データの入力が終わった際に最も多かった出力を最終的な推定結果とする．

4 提案手法の理論的背景

各コードに存在しうる音を理論的に考察することで谷状点の意味を考察する．

例として，C コードの場合を考える．和音の定義から，ギターの音域において存在しうる音は E1,G1,C2,E2,G2,C3,E3,G3,C4 の 9 つの音であり，スペクトルにはその 9 音に加えてそれぞれの倍音が存在することになる．n 倍音のパワーの強さを基音の $1/n$ であると仮定すると，図 4 のようになる．実際の C コードのスペクトルと重ね合わせると図 5 のようになる．

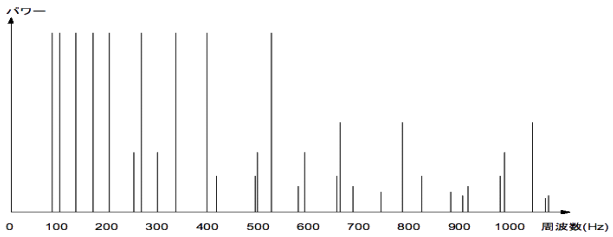

図 4: C コードスペクトル理論値

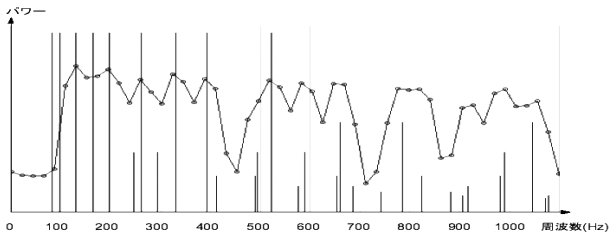

図 5: C コードスペクトル 理論値と実験値

理論的なスペクトルと実際のスペクトルの概形はほぼ同じであり，実際のスペクトルはこの理論で説明できると言える．つまり，谷状点とは，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り，一般性のあるものである．

同様に，C コード以外においても理論的なスペクトルと実際のスペクトルの概形はほぼ同じであり，多少のばらつきはあるものの第二谷状点は第一谷状点の約 1.5 倍，第三谷状点は第一谷状点の約 2 倍の周波数であることを確かめた．

5 実験

実験の際に，入力データの条件を表 1 のように設定した．

表 1: 入力データの条件

標本化周波数	44.1kHz
量子化 bit 数	16bit
録音時間	10 秒
録音方法	ピックアップから無雑音で録音

この入力に対し，A から G コードの 12 種類のメジャーまたはマイナーコード計 24 種類のうちどれであるかを推定した．第一谷状点の周波数と推定するコードの対応と各コードにおいてメジャーとマイナーコードのいずれであるかを判断する周波数は表 2 のように設定した．実験は 3 種類行った．

表 2: 本システムのルール

コード名	第一谷状点の周波数 $=\mu$	M or m を判断する周波数
E	269〜290Hz	775Hz
F	291〜311Hz	1680Hz
F#	312〜333Hz	883Hz
G	334〜354Hz	689Hz
G#	355〜376Hz	991Hz
A	377〜398Hz	1378Hz
A#	399〜419Hz	1486Hz
B	420〜441Hz	1571Hz
C	442〜462Hz	1658Hz
C#	463〜484Hz	1314Hz
D	485〜505Hz	1400Hz
D#	506〜527Hz	1464Hz

実験 1 演奏方法や弾き方に対する頑健性の実験
対象は一般的な C コードを音程を変えずに 66 種類の弾き方で演奏したものである．各弦を弾く強さ，弾く方法を変えた場合や，各弦が鳴らない，一音ずれる，開放弦となる場合などが含まれている．

実験 2 転回形に対する頑健性の実験
対象は C コードを 130 種類，Cm コードを 82 種類の音の組み合わせで演奏したもので，3 重音から 6 重音の各コードで弾きうる全ての音の組み合わせである．

実験 3 提案手法とクロマベクトルの精度の比較
対象は 3 重音，4 重音，ギターコードである．それぞれ A〜G コード，メジャーとマイナーコード，ハイとローコードの計 48 種類を用いた．ギターコードは 5 重音と 6 重音で構成されているものがあるため，一つにまとめてある．
クロマベクトルは 12 次元のものを用い，各コードの主要な 3 音のパワーを足し合わせ，パワーが大きな順に候補とした．第一候補として正解を出力した割合と第三候補までに正解を出力した割合を求める．

6 結果

実験 1 66 種類中 6 種類を別のコードであると判断した．別のコードであると判断されたのは，ルート音が別の音になっている場合だった．

実験 2 C コード 130 種類中 8 種類を別のコードであると判断し，Cm コード 82 種類中 39 種類を別のコードであると判断した．別のコードであると判断されたのは，メジャーコードではルート音である C の音が C4 のみであった場合で，マイナーコードでは E ♭ 2 が含まれている場合と重音数が 3 の場合だった．

実験 3 表 3 のような結果となった．重音数に関わらず提案手法はクロマベクトルよりも高い精度を示し，また，重音数が増えるにつれ精度が上がっている．

表 3: 実験 3 重音数ごとの正解率

手法	重音数		
	3 重音	4 重音	5,6 重音
クロマベクトル (第一候補)	14.9%	56.3%	33.3%
提案手法 (第一候補)	58.3%	64.6%	83.3%
クロマベクトル (第三候補まで)	31.3%	70.8%	68.8%
提案手法 (第三候補まで)	60.4%	68.8%	89.6%

7 考察

実験 1 結果から，極端に音程を間違えた演奏をしない限り，どのような演奏方法であっても結果は同一であり，正しく推定可能であると仮定できる．以降の実験に用いたデータはそれぞれ 1 回ずつの録音だが，これはこの仮定から妥当であるとする．

実験 2 メジャーコードについては，ルート音が C4 のみであるということはコードとして特殊であり，実際に演奏されることはほぼないと考えられる．
マイナーコードについては，E ♭ 2 の倍音として B ♭ 4 が含まれており，この音は Cm コードの谷状点の周波数と近い．このため誤推定したと考えられる．しかしギターにおいては，音程の配置上の理由で E ♭ 2 が使われることはほぼない．また，3 重音で誤推定をしているが，これは推定重音数が低い場合は精度が低いという特徴があると考えられる．
以上のように，同じコード内であれば，特殊なコードを除き正しい結果が導かれることがわかった．そのため，実験 3 において対象を一般的なコードに限っているが，これは妥当であるとする．

実験 3 クロマベクトルの 3 重音の正解率が 14.9% と非常に低くなっているが，これは 3 重音の特徴のためである．例えば C コードは主に C，E，G から構成されており，この 3 つを含まなければならない．各音の倍音を 9 倍音まで並べてみると，表 4 のようになる．C には 3 倍音と 6 倍音に G，5 倍音に E が含まれているが，E の倍音には C と G，G の倍音には C と E は存在しない．そのため，E や G の成分が多くなり，誤った出力をしたと考えられる．第三候補までに正解を含む割合が第一候補で正解する割合の倍であることもこれを裏付けている．

表 4: C コードの各音とその倍音

音名	倍音数							
	2	3	4	5	6	7	8	9
C	C	G	C	E	G	B ♭	C	D
E	E	B	E	G#	B	D	E	F#
G	G	D	G	B	D	F	G	A

クロマベクトル全体の精度を下げた要因として，次の二つが考えられる．一つは，FFT は低周波数域での分解能が低いことである．低音の入力がある場合，分解能の低さからその音程の周囲の音程にも影響を及ぼす．もう一つは，低音の倍音の影響である．低音のスペクトルは倍音が多くなるために推定すべき音名以外のパワーが大きくなり，誤推定が起こると考えられる．

以下，提案手法の結果について考察する．

第二谷状点が第一谷状点の 1.5 倍，第三谷状点が第一谷状点の 2 倍の周波数となっていることから，5 重音や 6 重音において谷状点は，まるで単音のスペクトルのような並び方をしていると言える．言い換えれば谷状点の系列は，存在しない音程とその倍音であるかのようである．つまり，第二谷状点以降の周波数を用いて第一谷状点の周波数を求める手法は，従来手法の単音の音高推定と類似である．このため，5 重音や 6 重音のコードを推定しているにも関わらず倍音の影響を受けることなくコードを推定できており，従来手法の倍音成分の重なりに弱かったという点を解決している．

逆に，4 重音，3 重音と重音数が減ると，谷状点の数は増えていき，従来手法での 2 重音や 3 重音を対象とする作業と類似する．このため，従来手法が倍音成分の影響を受けるのと同様に精度が下がると考えられる．

また，クロマベクトルは第一候補では正解でなくとも第三候補までに正解を含む割合が高いが，これは誤推定した場合でも正解を上位の候補として持つという特徴があると考えられる．一方，提案手法は，谷状点の周波数の誤推定した場合，根音が違う出力であるため，正解と全く関連のない出力であるという特徴がある．

8 今後の展望

提案手法の精度を下げた原因に，第一谷状点の周波数の誤推定がある．そのため，第一谷状点をより精度よく抽出する手法が必要とされる．しかし，第一谷状点の定義がはっきりとしていないことが問題である．そこで，第一谷状点を抽出するのではなく，谷状点の系列を手がかりとして和音推定する手法も考えられる．この手法を用いれば，実験 2 における転回形の誤推定も防げると考えられる．

以上のように，高精度に第一谷状点の抽出をすることが難しい場合，第一谷状点の誤推定を吸収するような機構を作る必要がある．例えば他の手法と組み合わせたり，スペクトルのピークも同時に特徴として用いたりすることが考えられる．Sheh, Lee, Ashley ら [5〜9] はクロマベクトルと HMM や CRF と組み合わせる手法を用いていたが，同様に HMM と CRF と組み合わせる手法が考えられる．

また，今回はメジャーコードとマイナーコードだけに絞っていたが，他論文では藤島ら [2] や Harte ら [4] のように aug や dim を用いるものが多い．これらのコードにも対応する必要がある．

なお，今回は楽器をギターに絞って実験したが，同じ構成音であればピアノでもうまくいくことを確認した．しかし，クラリネットのような偶数倍音がない楽器など，極端に高調波成分の少ない楽器ではうまくいかないことも考えられる．

今回，実験 2 でギター特有の条件を用いたが，楽器を 1 つに絞る場合，有本らのように，楽器固有の高調波構造を事前に抽出しモデル化することで，ピッチ推定を行ってから和音推定を行うことができる [10]．しかし，我々の手法は上記のように谷上点の系列を見ることでギター特有の条件が不要となる可能性があり，他の楽器への応用も可能であると考えられる．

9 結論

周波数スペクトルの谷状点に着目することで，多重音の音響信号を対象として和音推定を行うことができた．特に 5 重音や 6 重音において従来手法であるクロマベクトルと比較して高い精度を得た．提案手法を関連研究に応用することで精度の向上が期待できる．

参考文献

[1] 柏野 邦夫，村瀬 洋:” 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-II，No. 11，pp. 1762–1770，1996

[2] Takuya Fujishima:”Realtime chord recognition of musical sound: Asystem using Common Lisp Music”，In *Proceedings of the International Computer Music Conference 1999*，Beijing，1999.

[3] 後山 真一:” 楽音信号からの和音進行抽出手法と類似楽曲検索への応用”，パイオニア RD，Vol. 14，No. 2，pp. 1–7，2004

[4] Christopher A. Harte and Mark B. Sandler:”Automatic chord identification using a quantised chromagram”，In *Proceedings of ther Audio Engineering Society*，Spain，2005.

[5] Alexander Sheh，Daniel P.W. Ellis:”Chord Segmentation and Recognition using EM-Trained Hidden Markov Models”，In *Proceedings of ther International Conference on Music Information Retrieval 2003*，Baltimore，MD，2003.

[6] Kyogu Lee，Malcolm Slaney:”Automatic Chord Recognition from Audio Using an HMM with Supervised Learning”，*Proceedings of the International Conference on Music Information Retrieval 2006*，2006.

[7] Kyogu Lee，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*，2007.

[8] Kyogu Lee，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

[9] John Ashley Burgoyne，Laurent Pugin，Corey Lereliuk，Ichiro Fujinaga:”A Cross-Validated Study of Modelling Strategies for Automatic Chord Recognition in Audio”，In *Proceedings of the International Conference on Music Information Retrieval 2007*，2007.

[10] 有元 慶太，藤島 琢哉，後藤 真孝:” 楽器固有の高調波構造モデルを用いたギター演奏に対する多重音推定手法”，日本音響学会 2006 年秋季研究発表会 講演論文集，2-7-4，2006.