

## Doubly nested Circle of Fifths を用いた 和音認識

植村あい子<sup>†</sup> 甲藤二郎<sup>†</sup>

本稿では、音楽的知識である doubly nested Circle of Fifths に基づいた楽曲の和音情報を抽出する方法を提案する。この情報は、和音の類似性を表す特徴ベクトルを写像することにより生成され、DNCOF ベクトルと呼ばれる。本研究は DNCOF ベクトルの時系列に HMM を用いて和音とその境界の検出を試みる。

### Chord Recognition Using Doubly nested Circle of Fifths

Aiko Uemura<sup>†</sup> and Jiro Katto<sup>†</sup>

In this paper, we propose a novel method to obtain music chord information based on the musical theory known as “doubly nested Circle of Fifths”. We can obtain information by mapping features and call them “DNCOF vectors”. In this study, we apply HMM to the traces of DNCOF vectors to track music chords and detect their boundaries.

### 1. はじめに

ハーモニーは音楽の3要素のひとつであり、和音進行はハーモニー遷移を細かく捉えたものである。実際の楽曲は和音の遷移を中心に捉えることも多く、このことから和音は音楽の構成を決める重要な要素といえる。

実際に人間が採譜を行うとき、和音を手がかりにしてひとつひとつの音を求めていくように、自動採譜でも認識した和音から1音ずつ求めていく方法が考えられ、このように和音認識は自動採譜の前処理に利用できるといえる。また、自動採譜だけではなく、音楽のデータベース化とともにいろいろなアプリケーションへの応用が期待できる。例えば、ジャズのようなソロや即興演奏の楽譜化、ユーザがロズさんだ歌や鼻歌を楽譜化する作曲支援、さらには音楽情報を元にユーザの嗜好にあった楽曲を推薦する音楽検索システムなどが挙げられる。そこで本研究は、自動採譜や作曲支援・音楽情報検索などへの応用を目的として、実音源からの和音認識を試みる。

和音認識においては和音名に加え、和音境界を求めることが課題となる。和音名の推定には HMM がよく用いられる。特徴量にクロマベクトル、和声モデルに HMM を用いた研究では Sheh[1]らのものがあり、彼らは音楽信号と和声進行から EM アルゴリズムを用いて推定を行った。また、音楽的知識を用いたものには Bello らの研究[2]がある。彼らは HMM の初期値に音楽的知識をモデルとして取り込み、認識率を改善した。他にも、和音区間推定・和音同定に加えて、調性の推定を行うものもみられる[3]。調性も、和音とは密接な関係があり、これを加えることでさらにロバストな推定を目指している。

一方、筆者らは、[4]において、音楽的知識に特徴ベクトルを写像する試みとして、Circle of Fifths の調性の類似に着目し、写像によって得られた特徴量を用いて調性の推定を行っている。

本研究は[4]の手法を踏襲して、Bello らが利用した和音の類似性を表す doubly nested Circle of Fifths(以下 DNCOF)という音楽的知識に注目し、この知識に基づき生成される DNCOF ベクトルを提案する。処理の流れとしては、クロマベクトルからコードの尤度を求め、DNCOF 平面上へ写像して得られるベクトルを用いて和音認識を行う。クロマベクトルよりも抽象化された DNCOF ベクトルでは、音楽ジャンルや楽器種に左右されにくいことが期待される。

<sup>†</sup> 早稲田大学大学院 基幹理工学研究科  
Graduate School of Fundamental Science and Engineering, Waseda University

## 2. 提案手法

本稿では、実音源 DNCOF ベクトルの生成の手順を提案する。処理のおおまかな流れを図 1 に示す。

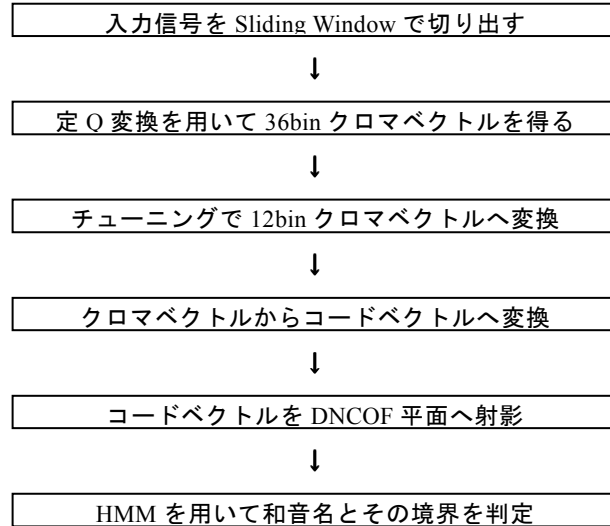


図 1: DNCOF ベクトルの作成手順

処理は 11250Hz にダウンサンプリングした wav 信号を、複数のフレームに切り出して行う。各フレームは 8192 個のサンプルから構成され、1 つのフレームに対して 1 つの DNCOF ベクトルが求められる。

### 2.1 クロマベクトル

クロマベクトル[5]は、周波数パワースペクトルを特定のピッチクラスに振り分けたもので、12 次元のものや 36 次元のものがよく使われる。クロマベクトルの計算には STFT (Short-Time Fourier Transform) がよく使われるが、オクターブのような対数スケールに変換されると、低いオクターブでは分解能が悪くなってしまふ。そこで、本研究では定 Q 変換[6]を用いてクロマベクトルを計算する。定 Q 変換においては、高域では時間の分解能が高くなり、一方低域では周波数の分解能が高くなる。ベース音は和音の根音となることが多いので、低域の周波数分解能が上がることは和音認識に有効と考えられる。

はじめに、各フレームに切り出された信号に対して定 Q 変換を行う。入力信号  $x(m)$  の定 Q 変換  $X_{cq}(k)$  は以下の式で表される。

$$X_{cq}(k) = \sum_{n=0}^{N(k)-1} w(n,k)x(n)e^{-j2\pi f_k n} \quad (1)$$

$w(n,k)$  は解析窓であり、その長さが  $N(k)$  で bin ポジション  $k$  の関数である。このとき  $k$  番目の中心周波数  $f_k$  は次式のように定義される。

$$f_k = \left( 2^{\frac{1}{B}} \right)^k f_{\min} \quad (2)$$

ただし、 $f_{\min}$  は定 Q 変換における最小周波数、 $B$  は 1 オクターブの bin 数である。この結果を用いて次式から、36bin クロマベクトル  $CH(b)$  を計算する。

$$CH(b) = \sum_{m=0}^M |X_{cq}(b+12m)| \quad 1 \leq b \leq 36 \quad (3)$$

ここで、 $M$  は定 Q スペクトルの総オクターブ数、 $b$  はクロマベクトルの bin インデックスである。

実際には各楽曲のチューニングが一定していないため、 $A4=440(\text{Hz})$  として計算した場合にスペクトルのパワーが他の bin にも分配されてしまい、明瞭なクロマベクトルが得られないことがある。そこで、(3)式で得られたクロマベクトルに対し、Harte ら[7]の手法を用いてチューニングを行う。

チューニングははじめに  $CH(b)$  のピークピッキングを行う。各フレームから、2 次補間を用いてピークの位置と大きさを求める。次に全てのピークのチューニングヒストグラムを計算したら、強く集中するピークが 1 つ求められる。これを中心チューニング周波数と一致するとみなす。この周波数がわかったら、ピッチクラス間の境界値が見つかるので、ピッチクラスに割り当てる。つまり、その両側に沿ったウインドウで、各 bin へのピークを選び、分配を行う。このようにして、チューニングされた 12bin クロマベクトル  $Chroma_{P_{\text{imod}12}}$  を得る。

### 2.2 コードベクトル

コードベクトルは major と minor の 24 種の和音がどのような尤度を持つかを表す 24 次元のベクトルである。この尤度の高いものが、そのフレームにおいて最も尤もらしい和音となる。コードベクトル  $C(t)$  は次式で定める。

$$C(t) = \begin{bmatrix} C_C(t) \\ \vdots \\ C_{B_{\min}}(t) \end{bmatrix} \quad (4)$$

$$C_{P_n}(t) = \sum_{i=0}^{11} w_i Chroma_{P_{(i+n) \bmod 12}}(t) \quad (5)$$

$$P_0 = C, P_1 = C\#, \dots, P_{11} = B \text{ または } P_0 = C_{\min}, P_1 = C\#_{\min}, \dots, P_{11} = B_{\min}$$

コードベクトルの各要素は、12bin クロマベクトル *Chroma* を入力とし、その重み付け和とする。和音の構成音のうち、各音の重要度は異なり、この重要度を反映させたものが  $w_i$  となる。例えば、Cmajor の基本構成音は C, E, G であるが、これらの倍音を考慮すると D や B など他の音も含むことになる。今回は[8]を参考に 6 倍音まで求める。ここでは、 $i$  番目の倍音には  $0.6^{i-1}$  だけ振幅を加え、最終的に振幅の合計が 1 になるよう正規化を行う。結果として、図 2 のように重みを定めた。

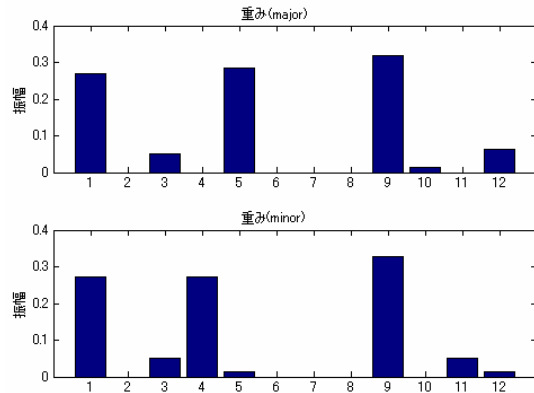


図 2 コードベクトルの重みの例 (上: Cmajor, 下: Cminor)

### 2.3 Doubly nested Circle of Fifths

図 3[2]は三和音の関係を示している。大文字は major triad, 小文字は minor triad を表している。これは、隣り合う三和音は似ており、対角上の三和音は似ていないという特徴がある。実際には、隣り合った三和音は 3 音中 1 音違うという関係である。

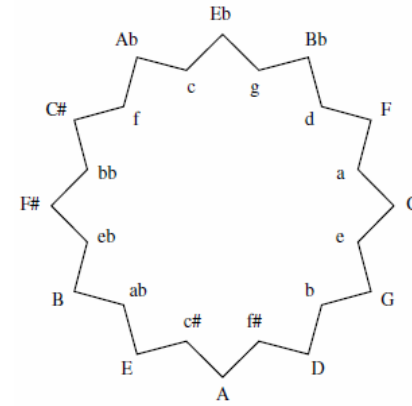


図 3 doubly nested Circle of Fifths

ここで、2.2 節で得られたコードベクトルを図 3 の平面に射影する。この手順は次の通りである。

- (1) DNCOF を円に見立て、各和音の方向に向かう単位ベクトルを用意する。

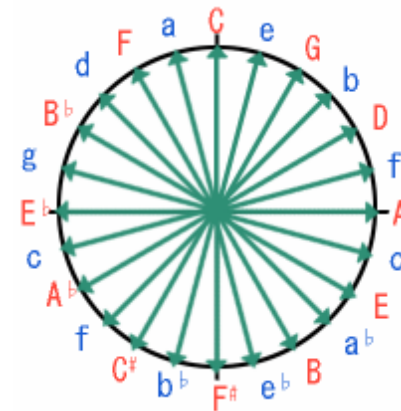


図 4 DNCOF ベクトル生成①

- (2) 単位ベクトルをそれぞれコードベクトルの要素倍する。

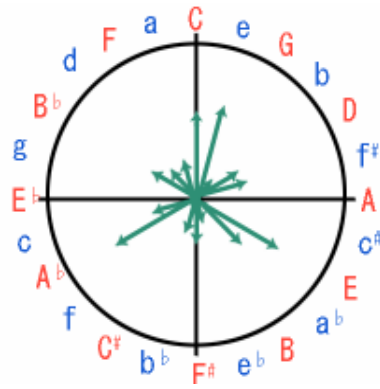


図 5 DNCOF ベクトル生成②

(3)それらの重心を取り，得られたベクトルを DNCOF ベクトルと定める。

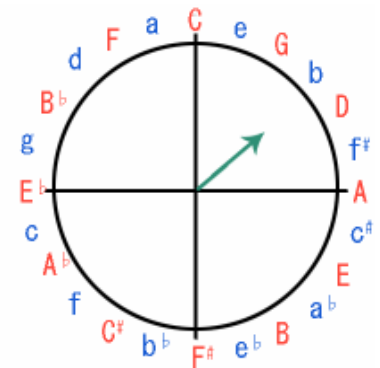


図 6 DNCOF ベクトル生成③

DNCOF では構成音の似ている和音の成分が大きくなるため，総合的に和音の評価ができる。式では以下のように定義される。

$$DNCOF(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = u C_{DNCOF}(t)$$

$$u = \begin{bmatrix} \cos\left(\frac{\pi}{2} - 0 \times \frac{\pi}{24}\right), \dots, \cos\left(\frac{\pi}{2} - 23 \times \frac{\pi}{24}\right) \\ \sin\left(\frac{\pi}{2} - 0 \times \frac{\pi}{24}\right), \dots, \sin\left(\frac{\pi}{2} - 23 \times \frac{\pi}{24}\right) \end{bmatrix} \quad (6)$$

$$C_{DNCOF}(t) = [C_{Cmaj}(t) \ C_{Emin}(t) \ C_{Gmaj}(t) \ \dots \ C_{Amin}(t)]$$

ここで， $u$  は単位ベクトルの集合で， $y$  軸正方向から時計回りに 24 個並んでいる。 $DNCOF(t)$ はコードベクトル  $C(t)$ の要素を DNCOF の順に並べたものである。ここで， $DNCOF(t)$ を極座標表示すると，

$$\begin{bmatrix} r(t) \\ \theta(t) \end{bmatrix} = \begin{bmatrix} [x(t), y(t)] \\ angle([x(t), y(t)]) \end{bmatrix} \quad (7)$$

となる。 $angle(v)$ は  $y$  軸の正方向から時計回りに測った角度とし， $-\pi < angle(v) \leq \pi$  である。 $r(t)$ はベクトルの大きさであり，ある方向にコードベクトルの値が偏るほど大きくなる。つまり，似た和音の尤度が高いということであり，そのフレームにおいて和音の純度が高いとみなすことができる。また， $\theta(t)$ はベクトルの偏角を表し，離散的な 24 和音の方向を示すのではなく，中間の方向を連続的に表し，和音のタイプを示す。

#### 2.4 和音認識

フレームごとに DNCOF ベクトルを求めていくと，楽曲全体にわたり図 7 のような DNCOF ベクトルの時系列が得られる。図 7 では縦軸が和音の種類，横軸はフレーム番号を示している。プロットの濃さは和音の純度を示しており，ある和音の方向にコードベクトルの値が偏るほど純度が高くなる。

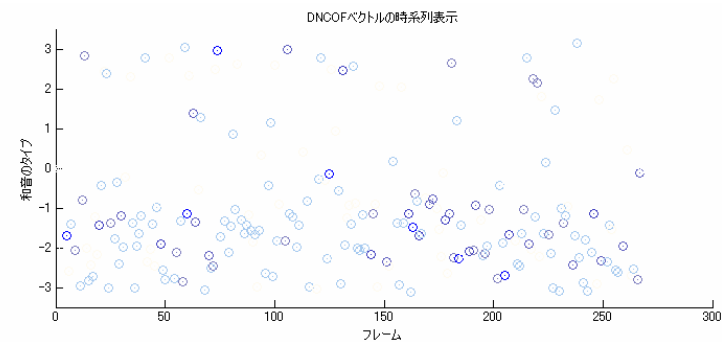


図 7 DNCOF ベクトルの時系列表示

これを和音の遷移情報とみなし、HMMを用いて和音認識を試みる。

HMMの各要素は以下のように定める。

状態は、major, minorの和音のみを扱うとし、全24状態とする。

遷移確率は、各和音の遷移しやすさ、つまりDNCOFオーダーで近い和音ほど遷移しやすく、遠い和音は遷移しにくくなるよう定める。 $arg_{ij}$ をDNCOF平面上で $chord_i$ と $chord_j$ のなす角とすると状態遷移確率 $P(S_i, S_j)$ は以下のように表される。

$$a_{ij} = P(S_i, S_j) = \frac{C}{e^{p*arg_{ij}} + e^{-p*arg_{ij}}} \quad (8)$$

$p$ は和音の遷移に関するペナルティであり、 $C$ は $arg_{ij}$ がとりうる範囲で合算した時に合計が1になるように調節するための係数である。

出力確率 $f(o; S_i)$ は状態 $S_i$ において信号 $o$ を出力する確率である。信号 $o$ はDNCOF列の1プロット、ある座標 $[x, y]$ とする。ここではある和音に対して、DNCOFプロット $[x, y]$ がどのくらいの確率で出現するかに基づいて定義する。

$S_i$ が示す和音のDNCOF偏角が $arg_{chordi}$ とするとプロットと偏角の差は次式のようになる。

$$arg_{diff} = |arg_{chordi} - arg(x, y)| \quad (9)$$

この式を用いて $f(o; S_i)$ は

$$f(o; S_i) = \frac{1}{2\pi} (1 + \cos(arg_{diff})) \quad (10)$$

となる。

そして、信号出力確率の計算と最尤パスの決定にはビタビアルゴリズムを用いた。

### 3. 実験

#### 3.1 実験

The Beatlesのアルバム(Please Please Me, Beatles For Sale)の楽曲に対し、和音認識を試みた。なお、本研究で扱う和音はmajorとminorの24種のみとし、正解データはHarteの作成したラベルデータ[9]を使用した。このラベルにはmajor, minor以外の和音も含まれるので、根音と第3音によりmajorとminorに分けた。例えば、Csus4やCaugはCmajに、Cmin7やCdimはCminとした。

本研究では定Q変換の最小周波数は96Hz、最大周波数は5250Hzとして実験を行った。ウィンドウサイズは8192サンプル(0.74s)、オーバーラップサイズはウィンドウサイズの1/8とした。

認識率は、フレームごとの和音ラベルの正解数をフレーム全体数で割って求めた。このとき、28曲の平均で18.3%の和音ラベルが完全一致した。

#### 3.2 考察

今回の結果では、完全一致した正解数は2割弱だが、不正解の場合も、正解からDNCOFオーダーで1から2のずれに収まっている場合が多い。これは、DNCOF平面に写像する際に、成分の近いmajorとminorがDNCOFオーダーでは離れており、正解のベクトルがもう一方のベクトルに引っ張られてしまったためと考えられる。これに対しては、写像時の補正を行うことによって、認識率の向上が可能になると考えており、現在検討を進めている。

### 4. おわりに

本稿では、音楽的知識を用いてDNCOFベクトルを抽出する方法を提案した。この特徴量は、和音の類似性を表すDNCOFに特徴ベクトルを写像することにより生成される。そして、得られたベクトルの時系列を用いて和音認識を試みた。

現在は、majorとminorは3音中2音が同じであるため、DNCOF平面への写像時に、もう一方の成分に引っ張られてしまう問題がある。これを解決するには、写像時の工夫を施す必要がある。また、認識率の向上としては、調性の考慮が挙げられる。調性は和音と密接な関係にあることから、調性を用いることによる認識率の向上が期待できる。

DNCOFベクトルはクロマベクトルより抽象化された特徴量であるので、ノイズに強いことがいえる。そして、平面へのマッピングにより、和音の関係性を視覚化でき、アプリケーションへの利用も考えられる。

### 参考文献

- 1) A. Sheh and D. P. Ellis, "Chord segmentation and recognition using EM-trained hidden markov models," Proc. ISMIR, pp. 183-189, Oct.2003.
- 2) J. P. Bello and J. Pickens, "A robust mid-level representation for harmonic content in music signal," Proc. ISMIR, pp. 304-311, Sep.2005.
- 3) T. Yoshioka. et.al, "Auto-matic Chord Transcription with Concurrent Recognition of Chord Symbols and Boundaries", Proc. ISMIR, Oct.2004.
- 4) T.Inoshita and J. Katto, "Key Estimation using Circle of Fifths", 15th International Multimedia Modeling Conference, Jan.2009.
- 5) T. Fujishima, "Real-time chord recognition of musical sound: A system using common lisp music," Proc. ICMC, pp. 464-467, Oct.1999.
- 6) Judith C. Brown and Miller S. Puckette, "An efficient algorithm for the calculation of a constant Q transform," J. Acoust. Soc. Am., 92(5), pp.2698-2701, 1992.

- 7) C. Harte and M. Sandler, “Automatic chord identification using a quantised chromagram,” in Proc. Audio Eng. Soc., Spain, May.2005.
- 8) Oudre. et.al, “Template-Based Chord Recognition : Influence of the Chord Types ,” Proc. ISMIR, pp. 153-158, Oct.2009.
- 9) C. Harte et al., “Symbolic representation of musical chords: A proposed syntax for text annotaions,” Proc. ISMIR, pp. 66-71, Sep.2005.