

## Doubly Nested Circle of Fifths を用いた和音認識の特性改善 Improvement of Chord Recognition Using Doubly Nested Circle of Fifths

植村 あい子<sup>†</sup> 甲藤 二郎<sup>†</sup>  
Aiko Uemura Jiro katto

### 1. まえがき

和音は自動採譜やカバーソングの検索の手がかりになり、和音は音楽の構成を決める重要な要素であることから、和音認識の研究は多く行われている。そこで本研究は、自動採譜や音楽情報検索への応用を目的として音響信号からの和音認識を試みる。

和音認識の従来研究には、特徴量にクロマベクトル、和声モデルに HMM を用いたもの[1]や、チューニングを行ったクロマベクトルに対しテンプレートマッチングにより和音名を求めるもの[2]などがある。

一方、音楽知識に特徴ベクトルを写像する試みとして、筆者らは 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た[3]。

本研究では[3]の手法を踏襲し、和音の類似性を表す Doubly Nested Circle of Fifths (以下 DNCOF) を用いた和音認識を試みる。本稿では、ベクトルの写像時に調性を反映して改善を行い、得られた特徴ベクトルを用いて和音認識を行い、その有効性を確認した。

### 2. 提案手法

実音源 DNCOF ベクトルの生成と和音認識までの手順を提案する。この流れを図1に示す。

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。

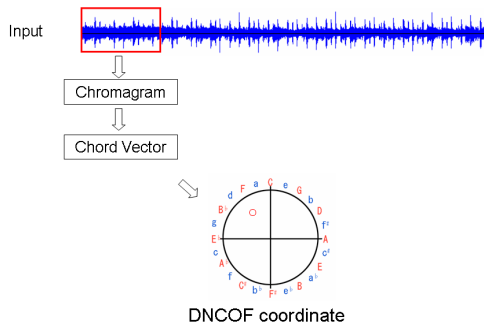


図1 DNCOF ベクトルと和音認識の流れ

#### 2.1 クロマベクトル

クロマベクトルは、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ある。

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[4]を行う。音響信号は 11025Hz にダウンサンプリングをし、ウィンドウサイズは 8192 サンプル(0.74s)、ホップサイズはウィンドウサイズの 1/4 とした。定 Q 変換で得られたスペクトル  $X_{cq}(k)$  のパワーのオクターブを吸収して、36bin クロマベクトルを求める。

<sup>†</sup> 早稲田大学大学院基幹理工学研究所

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

$M$  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。次にスペクトルのパワーが他の bin にも分配されてしまうことを避けるため、Harte ら[2]の手法を用いてチューニングを行い、フレーム  $t$  における 12bin クロマベクトル  $Chroma(t)$  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。コードベクトルの各要素は、12bin クロマベクトル  $Chroma(t)$  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るため、この重要度  $w_i$  を反映させる。この  $w_i$  は[5]の和音テンプレートを参考にした。

$$C(t) = \begin{bmatrix} C_C(t) \\ \vdots \\ C_{B_{\min}}(t) \end{bmatrix} \quad (2)$$

$$C_{P_n}(t) = \sum_{i=0}^{11} w_i Chroma_{P_{(i+n) \bmod 12}}(t) \quad (3)$$

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

#### 2.3 DNCOF 平面への射影

Doubly Nested Circle of Fifths[1]は三和音の関係を示している。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。

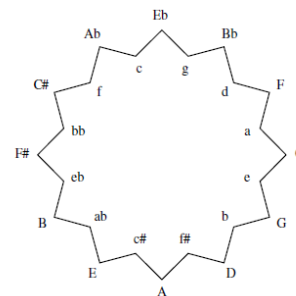


図2 Doubly Nested Circle of Fifths[1]

ここで、2.2 節で得られたコードベクトルを図2の平面に射影する。このとき、調性を反映させるため、ある調で用いられる和音(調固有和音)の成分のみの合成を行う。この手順は次の通りである。

(1) DNCOF を円に見立て、調固有和音の方向を向いたベクトルを基底とする行列  $u$  を構成する。これは角度の異なる 24 個の単位ベクトルを並べた行列  $U$  を調固有和音に基づいてビットマスク  $w_{keyi}$  を掛けることで作成される。 $w_{keyi}$  はある調  $key$  における和音  $i$  のマスクである。

(2)行列  $u$  を下記に定義されるコードベクトル  $C_{DNCOF}(t)$  に乗じる.

(3)(2)で求めたベクトルを DNCOF ベクトル  $DNCOF(t)$  と定める.

この処理は以下の式で表される.

$$DNCOF(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = u C_{DNCOF}(t) \quad (4)$$

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

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

### 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3のような DNCOF ベクトルの時系列が得られる. 図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. 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 の各要素は次のように定めた.

まず、状態は major, minor の和音のみの 24 状態を扱う. 次に遷移確率は、和音の正解データ[9]を使用し、事前に学習を行った. なお、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音に基づいて major と minor に手で分けた.

出力確率は、ある和音状態のとき、そのプロットが出現する確率を定める. 出力は DNCOF 時系列の 1 プロットつまり、DNCOF ベクトルの座標となる. ここでは座標の偏角と von Mises 分布の確率密度を用いて出力確率を定義した.

そして、入力楽曲に対する最尤な和音遷移の決定にはビタビアルゴリズムを用いた.

### 3. 評価実験

#### 3.1 実験

ここで提案手法を評価するため、和音認識を試みた. 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. なお、今回は各楽曲の調は既知とし、正解データは[7]を利用した. データセットには[7]で公開されている 179 曲を用いて学習し、アルバム (Please Please Me(CD1:14 曲), Beatles For Sale(CD2:14 曲)の 28 曲を認識した.

今回は比較手法として、MIREX2009 で D. Ellis がコード提供しているクロマベクトル+HMM による和音認識法[7]を取り上げた. ただし、[7]に挙げられている beat-synchronous chroma ではなく、提案方式と同じクロマベクトルを使用した.

#### 3.2 実験結果

DNCOF ベクトル時系列の認識結果を図3に示す.

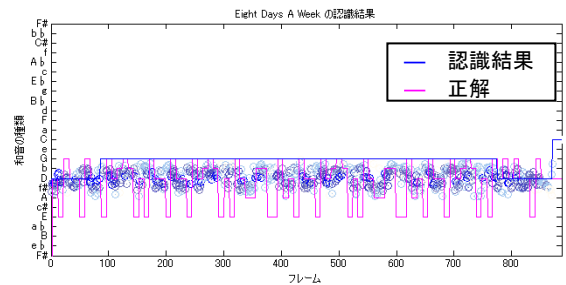


図3 DNCOF ベクトルの認識結果  
曲: Eight Days a Week(Beatles For Sale より)

また、各アルバムのラベルの完全一致率の平均を表1に示す. 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.

表1 和音認識結果(ラベルの完全一致率の平均)

	CD1	CD2
提案手法	34.9%	38.6%
旧手法[8]	17.2%	19.3%
比較手法[7]	40.8%	57.5%

数値だけ見ると、特徴ベクトルに DNCOF ベクトルを用いた方はクロマベクトルに比べて認識率が低いですが、旧手法[8]に比べて改善が見られ、図3のように DNCOF ベクトルの各プロットは正解に近いことが確認できる. しかし、プロットのバラつきが抑えられた分、E のあたりにプロットが存在しない. これは、実際の楽曲では調固有和音以外の和音が用いられることも多く、その場合、実際の和音が DNCOF ベクトル生成時に無視されてしまったためである. この点に関しては現在写像時の工夫を検討中である. そして和音が遷移していないのは、音響信号を切り出すフレームと、ホップサイズが細かいため、同ラベルへの遷移確率が高くなってしまったと考えられる.

### 4. あとがき

本稿では、調性を反映させた DNCOF ベクトルを抽出する方法を提案し、得られたベクトルの時系列に対し、HMM を用いて和音認識を試みた. 各プロットは正解に近いものが多く、DNCOF ベクトルの有効性が確認できた.

今後は DNCOF ベクトル自身の精度を高め、認識精度の向上を図りたいと考えている.

#### 参考文献

- [1] J. P. Bello and J. Pickens, "A robust mid-level representation for harmonic content in music signal," Proc. ISMIR, pp. 304–311, Sep.2005.
- [2] C. Harte and M. Sandler, "Automatic chord identification using a quantised chromagram," in Proc. Audio Eng. Soc., Spain, May.2005.
- [3] T. Inoshita and J. Katto, "Key Estimation using Circle of Fifths," 15th International Multimedia Modeling Conference, Jan.2009.
- [4] 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.
- [5] Oudre, et al., "Template-Based Chord Recognition: Influence of the Chord Types," Proc. ISMIR, pp. 153–158, Oct.2009.
- [6] isophonics: <http://isophonics.net/>
- [7] Supervised Chord Recognition for Music Audio in Matlab: <http://labrosa.ee.columbia.edu/projects/chords/>
- [8] 植村, 甲藤, "Doubly Nested Circle of Fifths を用いた和音認識," 情報処理, 2010-MUS-84, 5, 2010.