Vol.2011-MUS-90 No.5
2011/5/13

DNCOF ベクトルとクロマベクトルを用いた和音認識

植村あい子[†]　　甲藤二郎[†]

本稿では，DNCOF ベクトルとクロマベクトルを併用した和音認識手法を提案する．DNCOF ベクトルは楽曲の大まかな和音情報を表すものであり，クロマベクトルによる認識手法へ応用することで認識の向上が期待できる．本研究では，DNCOF ベクトルによって得られた和音情報を活用して，認識精度の向上について検討を行った．

Chord Recognition using DNCOF Vector and Chroma Vector

Aiko Uemura[†] and Jiro Katto[†]

In this paper, we propose a chord recognition method using DNCOF vectors and chroma vectors. A DNCOF vector represents rough chord information, so we expect that the recognition rate will go up if we combine it with the approach using chroma vector in an adequate manner. In this research, we expand the chord information estimated by DNCOF vectors, and evaluate our proposals using the Beatles' songs.

1. はじめに

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，音響信号からの和音情報抽出と和音認識を試みる．

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．特徴量には一般的にクロマベクトル 1)が用いられ，クロマベクトルを改善するアプローチも多く提案されている．クロマベクトルの改善手法としては，チューニングを行って改善を図るもの 2)や和音はビートに合わせて演奏されることも多いのでビート同期を取ることで改善を行ったものもある．他にも，文献 3)では周波数スペクトルのフレームは理想的な音パターンの線形結合で表せるという仮定をし，NNLS（Non-Negative Least Square）問題を解くことで得られる NNLS chroma を提案している．

認識手法は，大きく分けて 3 つに大別される．第一に，テンプレートを用いた手法がある．これは検出したい和音をテンプレートで定義をし，フレームごとに最もデータに合うテンプレートの和音を選ぶ手法である．テンプレートの例としては単純に構成音を 6 倍音まで考慮したものがある 4)．第二に，データに基づく手法がある．これは，正解データを利用し，パラメータを学習することによって，評価を行っている．音楽信号と和声進行から EM アルゴリズムを用いて推定するもの 5)や，ラベルデータをシンボル音楽ファイルの音響解析で求め，特徴ベクトルは同じシンボルデータである合成された音から計算し，それらを使って学習を行うもの 6)もある．第三に，前述の手法を組み合わせた手法がある．例えば 7)では，データを学習する際に，初期値に音楽知識をモデルとして組み込んでいる．

一方，筆者らは，音楽知識のひとつである Circle of Fifths の調の類似性に着目をし，写像によって得られたハーモニー情報を用いて，調性の推定を行っている 8)．我々はこの手法を和音の類似性を表す Doubly Nested Circle of Fifths （以下 DNCOF）7)に応用し，得られた DNCOF ベクトルを用いて，和音認識を行ってきた 9)．DNCOF ベクトルはクロマベクトルの抽象度を上げたものであるため，認識率は 28 曲で平均 40%弱と高くないものの，DNCOF ベクトルは正解から大きく外れていないことから，DNCOF ベクトルの和音情報としての有効性を確認した．

そこで本研究では，DNCOF ベクトルは楽曲で使われている和音を大まかに表現できるという特性を用いて，従来のクロマベクトルによる認識手法へこの情報を組み合わせることで認識率の改善を試みた．

[†] 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

Vol.2011-MUS-90 No.5
2011/5/13

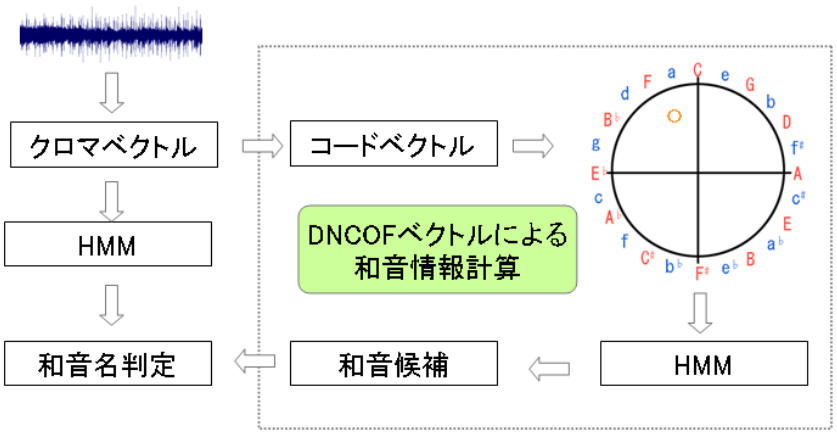

図 1　提案手法の流れ

2. 提案手法

2.1 概要

提案手法の流れを図に示す. 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る．DNCOF ベクトルによる和音情報を求める処理では，1 つのフレームにつき 1 つの DNCOF ベクトルが求まり，この時系列情報を HMM によりトラッキングすることにより和音候補を求める．そして，その候補とクロマベクトルを用いて認識された正解候補と比較し，DNCOF ベクトルにより求まった候補に近くなるよう順位を入れ替える．この処理を各フレームに対して行う．

2.2 クロマベクトル

クロマベクトルの計算にあたり，はじめに入力信号をフレームに切り出し，定 Q 変換を行う，ここでは定 Q 変換は 96Hz から 5250Hz の範囲で行う．得られた結果 X_{cq} を用いて次式から，36bin クロマベクトル CH（b）する．

$$CH(b)=\sum_{m=0}^{M}\left|X_{cq}(b+12m)\right| \qquad 1\le b\le 36 \tag{1}$$

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．実際の楽曲では，常に A=440Hz でチューニングされているとは限らないので，他の bin にパワーが分配されないように，文献 2)の手法を用いて，チューニングを行い *Chroma* (t)を得た．

2.3 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．

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

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

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

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．そこで本研究は文献 4)を参考に 6 倍音まで考慮した重み付けをする. 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った.

2.4 DNCOF ベクトル

音楽知識 DNCOF7)とは，図で示されるような major と minor の和音を交互に並べたものであり，major と minor の円が二重入れ子状になっている．外側の円である大文字が major（長三和音），内側の円である小文字が minor（短三和音）を表している. DNCOF では隣り合った和音同士の構成音を比べると，使用される 3 音中 2 音が等しいという類似性を持っていることがわかる．つまり，隣り合う和音は似ているが，対角上にある和音は似ていないと言うことができる．

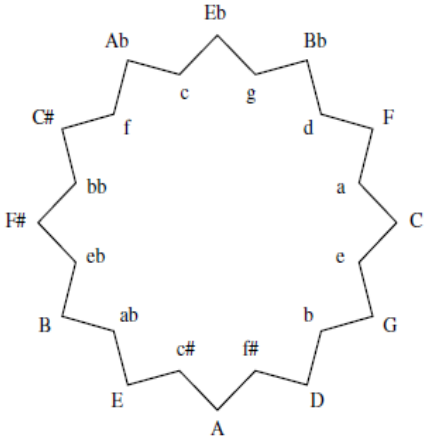

図 2　Doubly Nested Circle of Fifths（DNCOF）7)

ここで，2.3 で得られたコードベクトルを DNCOF 平面状に写像していく．

(1) DNCOF を円に見立て，major と minor それぞれ各和音の方向に向かうベクトル u_{maj} と u_{min} を用意
(2) 1 にある調における和音の重み w_{keyi} をかける
(3) それぞれにおいて，コードベクトル $C(t)$の要素倍する
(4) 成分が閾値より大きいものだけ残す
(5) major と minor それぞれで求まったベクトルの大きさを比較し，大きい方を DNCOF ベクトルとする

ここで，w_{keyi} と（4）の閾値は実験的に定めた．
この手順を式で表すと以下の通りになる．はじめに各和音の方向に向かうベクトル u_{maj} と u_{min} は次式で表される．

$$u_{maj}=\begin{bmatrix} w_{key_{\mathrm{Cmaj}}}\cos\left(\frac{\pi}{2}-0\times\frac{\pi}{24}\right), & \cdots, & w_{key_{\mathrm{Fmaj}}}\cos\left(\frac{\pi}{2}-22\times\frac{\pi}{24}\right) \\ w_{key_{\mathrm{Cmaj}}}\sin\left(\frac{\pi}{2}-0\times\frac{\pi}{24}\right), & \cdots, & w_{key_{\mathrm{Fmaj}}}\sin\left(\frac{\pi}{2}-22\times\frac{\pi}{24}\right) \end{bmatrix},$$

$$u_{min}=\begin{bmatrix} w_{key_{\mathrm{Emin}}}\cos\left(\frac{\pi}{2}-1\times\frac{\pi}{24}\right), & \cdots, & w_{key_{\mathrm{Amin}}}\cos\left(\frac{\pi}{2}-23\times\frac{\pi}{24}\right) \\ w_{key_{\mathrm{Emin}}}\sin\left(\frac{\pi}{2}-1\times\frac{\pi}{24}\right), & \cdots, & w_{key_{\mathrm{Amin}}}\sin\left(\frac{\pi}{2}-23\times\frac{\pi}{24}\right) \end{bmatrix} \quad (4)$$

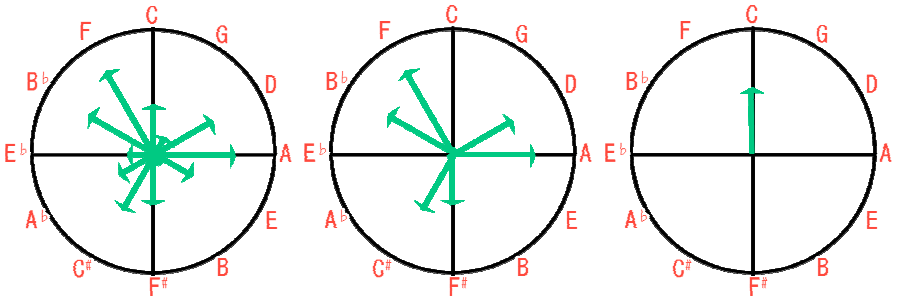

図 3　DNCOF 写像の流れ（major の場合）
左：コードベクトルの要素倍，中：閾値処理後，右：合成されたベクトル

このベクトル u と，コードベクトル $C(t)$の major 成分と minor 成分をそれぞれ掛け合わせ，大きさの大きいほうを DNCOF ベクトルと定める．

$$DNCOF(t)=\begin{bmatrix} x(t) \\ y(t) \end{bmatrix}=\begin{cases} u_{maj}C_{maj}(t) & \text{if } \left|uC_{maj}(t)\right|\geq\left|uC_{min}(t)\right| \\ u_{min}C_{min}(t) & \text{otherwise} \end{cases} \quad (5)$$

$$C_{maj}(t)=[C_{\mathrm{Cmaj}}(t) \quad C_{\mathrm{Gmaj}}(t) \quad \cdots \quad C_{\mathrm{Fmaj}}(t)]^{\mathrm{T}}$$

$$C_{min}(t)=[C_{\mathrm{Emin}}(t) \quad C_{\mathrm{Bmin}}(t) \quad \cdots \quad C_{\mathrm{Dmin}}(t)]^{\mathrm{T}}$$

ここで，$DNCOF(t)$を極座標表示すると，

$$\begin{bmatrix} r(t) \\ \theta(t) \end{bmatrix}=\begin{bmatrix} \left|[x(t),y(t)]\right| \\ angle([x(t).y(t)]) \end{bmatrix} \qquad -\pi<angle(v)\leq\pi \quad (6)$$

となる．$angle(v)$は y 軸の正方向から時計回りに測った角度である．$r(t)$はベクトルの大きさであり，コードベクトルの値が偏るとき，つまり円周上に近づくほど大きくなる．$r(t)$が大きいときは，そのフレームにおいてある和音の尤度が高いということになり，和音の純度が高くなる．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すので，$\theta(t)$が和音のタイプを表すと

Vol.2011-MUS-90 No.5
2011/5/13

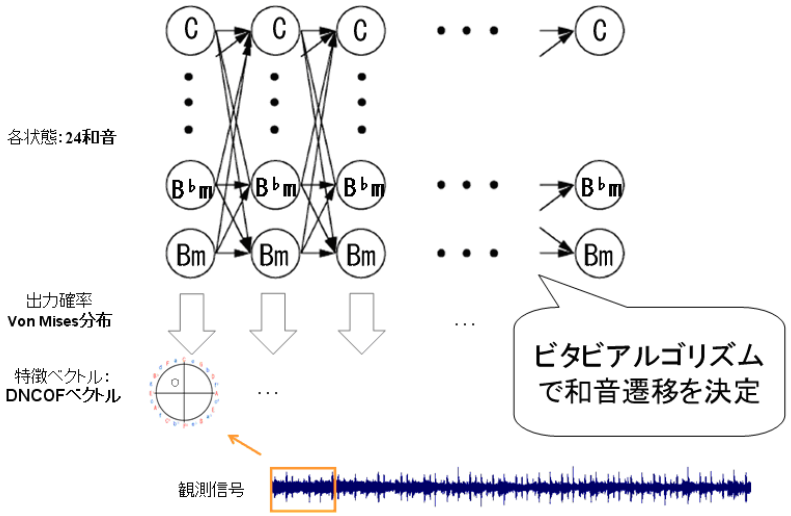

図 4　DNCOF ベクトルによる和音認識のための HMM 設定

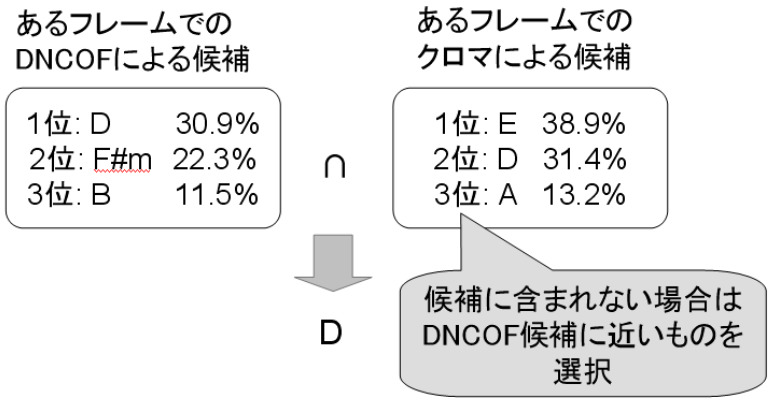

図 5　DNCOF ベクトルによる和音情報の応用方法

みなすことができる．ここではその中間の方向というものを，「両側にある 2 つの和音をその内分の割合で混ぜた和音のタイプ」というようにとらえる．例えば，C major と A minor のちょうど中間を指すような DNCOF ベクトルがあった場合には，「C,E 音をメインに用い，G, A が半々で現れるような和音情報」ととらえる．

2.5　和音認識手法

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 6 のような DNCOF ベクトルの時系列が得られる．これを本研究では DNCOF 列と呼ぶ．この DNCOF 列を楽曲の和音情報とみなし，解析を試みる．図のように，和音認識を行うための HMM 設定として（a）～（e）を定める．

（a） 状態集合：$\Sigma=\{S_i \mid 0 \leqq i \leqq M\}$
各状態を major，minor の各和音 Cmajor, C#major, …,Bminor に対応させる．つまり全 24 状態が定義される．

（b） 状態遷移確率：$A=\{a_{ij}=P(S_iS_l \mid 0 \leqq i,j \leqq M)\}$
各状態間の遷移確率，すなわち各和音間の転調のしやすさを定義する．ここでは，初期値を[3]に基づき DNCOF 順で設定し，学習を行う．

（c） 出力確率：$B=\{f(o;S_i) \mid 0 \leqq i \leqq M\}$
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．本研究では von Mises 分布を用いて出力確率を定めた．

（d） 出力信号系列：$o(t)(t=0, \cdots, T)$
ここでは DNCOF 列そのものに相当する．

（e） 状態系列：$s(t)(t=0, \cdots, T)$
状態系列は状態，つまり和音名の系列を表している．出力信号系列から，この状態系列を求めることが目的となる．

そして，最尤な和音遷移を決定するのにはビタビアルゴリズムを用いた．

2.6　DNCOF ベクトルによる和音情報の応用

クロマベクトルでは，根音が同じである major と minor は 3 音の構成音のうち 1 音が違っており，構成音が似ていることから誤って認識される可能性が高い．例えば，正解ラベルは C major であるが，認識結果として C minor が出力されることが挙げられる．しかし，図のように，DNCOF ベクトルでは，DNCOF 順で両者が離れているため，この major と minor の間違いを防ぐことが可能であると期待できる．

ここで，DNCOF ベクトルによる認識で求まった和音情報をクロマベクトルの認識への応用を試みる．応用方法は図の通りである．

（1）　DNCOF ベクトルから 3 つの和音候補を求める
（2）　1 の候補とクロマベクトルで求まった和音正解候補を比較し，1 の候補に近くなるよう順位を入れ替える

表 1 和音認識結果

	DNCOF ベクトル	クロマベクトル	DNCOF ベクトル + クロマベクトル
179 曲平均	30.5 %	38.3 %	40.2 %

例えば，あるフレームにおいて DNCOF ベクトルによる認識で求まった 1 ～ 3 位の候補が D , F#m, B だとする．そして，そのフレームでのクロマベクトルよる候補が E, D, A だとする．ここで，E, D, A のうち DNCOF ベクトルでの候補に DNCOF 順で一番近い和音は D になるので，クロマベクトル候補における順位を入れ替え，出力されるラベルは D となる．

この処理を各フレームに対して行う．ただし，クロマベクトルの候補の 1 位と 3 位の確率の差が閾値より大きかった場合には，クロマベクトルの 1 位の候補の和音名をそのまま出力する．

3. 実験

3.1 DNCOF ベクトルを用いた和音認識

ここで, 1025Hz にダウンサンプリングしたモノラル信号に対し, 評価実験を行った．解析にあたってウインドウサイズは 8192 サンプル (0.74 秒)，オーバーラップサイズはウインドウサイズの 4 分の 1 とした．学習では，The Beatles の 12 枚のアルバムの 179 曲を学習し，179 曲の認識を行った．

本研究で扱う和音は major と minor の 24 種とし，和音と調のデータセットは Isophonics10)で公開されているものを使用した．和音データには major, minor 以外の和音も含まれているため，人手によって根音と第 3 音により major と minor に分けた．

Csus4 や Caug は Cmaj に，Cmin7 や Cdim は Cmin に分類した．

また，認識率はフレームごとの正解率とし，次式のように定めた．

$$\text{正解率} = \frac{\text{正しく出力されたフレーム数}}{\text{全フレーム数}} \tag{7}$$

なお，楽曲の先頭や末尾には無音区間が含まれているが、正解率計算の際は，無音区間は和音ではないため除外した．

（1）DNCOF ベクトル単体の和音認識

ここでは，DNCOF ベクトルの精度を求め，実際に大まかに和音が求まっているのかを確認する．

（2）クロマベクトルを用いた和音認識

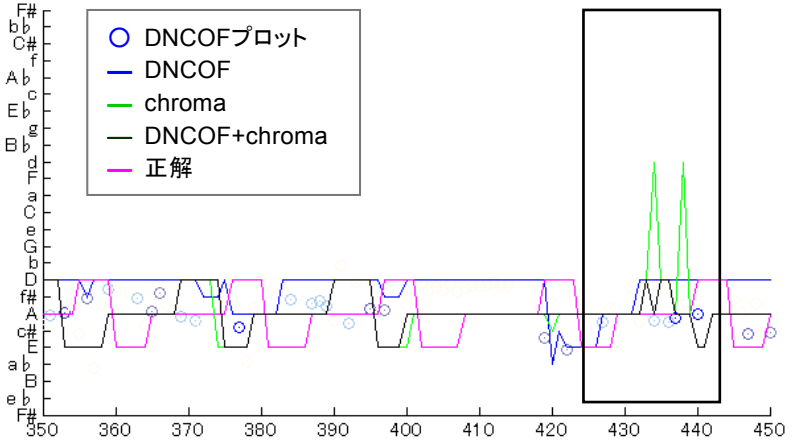

図 6 Beatles for Sale より”Words of Love”の 350～450 フレームの結果
(縦軸：和音の種類，横軸：フレーム番号，プロットの濃さ：和音の純度)

DNCOF ベクトルはクロマベクトルの抽象度を上げたものであるので，情報量が削減される前の精度を調べるために，比較手法として特徴ベクトルにクロマベクトルを用いて実験を行った．DNCOF ベクトルとの違いは特徴ベクトルだけであるので，認識率によりクロマベクトルと DNCOF ベクトルの比較ができると考えられる．

（3）DNCOF ベクトルによる和音情報+クロマベクトル和音認識

2.6 での方法により，認識率が改善されるのかを確認する．

3.2 結果

実験（1）～（3）の認識結果は表 1 のようになった．表を見ると平均で約 2%であるが，DNCOF ベクトルとクロマベクトルの組み合わせにより，認識率の向上が確認できる．

また，DNCOF プロットと各認識結果を図 6 に示す．四角で囲まれた区間では，クロマベクトルによる認識では Dm であるが，DNCOF 情報を用いたクロマベクトルでの認識では，D になっており，major と minor の一音違いの誤認識が改善されている．また，DNCOF プロットは正解データに近い位置にあり，DNCOF ベクトルは大まかに和音を表現していることが確認できる．

3.3 考察

図 6 のように，一音違いの誤認識が改善されたのは，DNCOF のモデル上で major と minor が離れているためであると考えられる．これは，DNCOF ベクトルの精度を

向上させることにより，さらなる改善が期待できる．

一方で，DNCOF プロットは局所的になっているため，DNCOF 順で離れた和音への遷移の場合ではプロットは正解から離れてしまった．調固有和音を強調すると，DNCOF 順で離れた和音遷移のときに，ばらつきが抑えられてしまって調固有和音以外への対応ができなくなってしまう．今後，和音と関連の深い音楽要素を反映させる際には，写像時の工夫に当たっては情報が偏らないように注意が必要である．これに対しては，和音の前後関係を反映させるという点で和声を取り入れることも有効だと考えている．

さらに，楽曲によっては一部のフレームで前後のフレームでは正しく認識しているものの，DNCOF ベクトルの誤認識により，あるフレームだけ誤ったラベルで出力されることがあった．これは DNCOF ベクトルの精度も原因であるが，和音情報を応用する際に各フレームにおいて順位を入れ替えてしまい，前後のフレームとの関係が考慮されていないことも原因である．この対策としては，和音区間やビートトラッキングで区間を求めてから，その区間ごとに候補を入れ替えることで対処できると考えられる．

4. おわりに

本研究では，和音情報である DNCOF ベクトルを用いてクロマベクトルによる和音認識の改善を行った，DNCOF ベクトルで得られた大まかな和音情報と，クロマベクトルによる認識手法を組みあわせることで認識率の向上を試みた．

DNCOF ベクトル単体ではクロマベクトルに比べて認識率は下がるものの，DNCOF ベクトルで得られた和音情報をクロマベクトルへ応用することによって，少しではあるがクロマベクトルによる認識を改善できた．今後，DNCOF ベクトルのさらなる性能改善により，和音情報を用いて認識の改善が可能であることが期待される．

また，和音は他の音楽要素とも深く関わっているため，他の音楽要素と組み合わせることも検討している．例えば，和声の知識を取り入れたり，ビート同期などで和音区間を求めたりして改善を試みたい．そして DNCOF ベクトルの生成において，クロマベクトルを使用していることから，DNCOF ベクトルの精度はクロマベクトルにも依存すると考えられる．今回は文献 2)のクロマベクトルで実験を行ったが，MIREX2010 11)で一番高い認識率であった NNLS chroma3)への対応も検討中である．

参考文献

1) T. Fujishima, “Real-time chord recognition of musical sound: A system using common lisp music,” Proc. ICMC, pp. 464-467, Oct.1999.
2) C. Harte and M. Sandler, “Automatic chord identification using a quantised chromagram,” in Proc. Audio Eng. Soc., Spain, May.2005.
3) M.Mauch et al., "Approximate Note Transcription for the Improved Identification of Difficult Chords," Proc.ISMIR, Aug.2010.
4) Oudre. et.al, “Template-Based Chord Recognition : Influence of the Chord Types ,” Proc. ISMIR, pp. 153−158, Oct.2009.
5) A. Sheh and D. P. Ellis, “Chord segmentation and recognition using EM-trained hidden markov models,” Proc. ISMIR, pp. 183-189, Oct.2003.
6) K.Lee and M.Slaney,”Acoustic chord transcription and key extraction from audio using key-dependent HMMs trained on synthesized audio,” IEEE Trans.on Audio,Speech and Language Procsing,16(2), pp.291-301, 2008.
7) J. P. Bello and J. Pickens, “A robust mid-level representation for harmonic content in music signal,” Proc. ISMIR, pp. 304−311, Sep.2005.
8) T.Inoshita and J. Katto, “Key Estimation using Circle of Fifths”, 15th International Multimedia Modeling Conference, Jan.2009.
9) 植村，甲藤, “Doubly Nested Circle of Fifths に基づく和音情報と HMM を用いた和音認識“，電子情報通信学会総合大会，March.2011.
10) Isophonics
http://isophonics.net/
11) MIREX2010
http://www.music-ir.org/mirex/wiki/MIREX_HOME