

音楽理論を用いた誤り区間の検出・補正による和音認識

植村 あい子[†] 甲藤 二郎[†]

早稲田大学大学院 基幹理工学研究科[†]

1. まえがき

本研究は、自動採譜の前処理や音楽情報検索への応用を目標として、音響信号からの自動和音認識を試みる。

HMMを用いた和音認識手法では、認識誤り時に HMM で得られる和音候補の 2,3 位に正解が含まれることが多いため、我々は従来の和音認識ステップに誤りを検出し補正する処理を追加することで認識性能の改善を図る。

本稿では、音楽理論である Doubly Nested Circle of Fifths(DNCOF)を用いて、従来手法であるクロマベクトルと HMM による認識で発生する誤りの検出・補正を行う。誤り発生時には本研究で提案する DNCOF ベクトルとクロマベクトルの結果が DNCOF 平面上で離れていることから、我々は閾値により誤り箇所を推定し、クロマベクトルと DNCOF を併用することでの精度改善を試みる。本稿では誤り区間補正による和音認識精度の向上について検討し、The Beatles の楽曲を用いて評価を行った。

2. 提案手法

はじめに、音響信号からビートと同期した 12 次元クロマベクトルを算出し、HMM を用いて和音認識を行う。次に HMM の後処理としてコードベクトルを算出し、DNCOF 平面上で誤り箇所の検出を行う。最後に誤りと推定されたフレームにおいて、DNCOF ベクトルを用いて補正を行う。

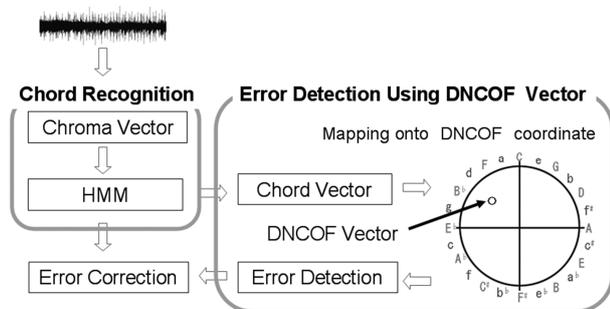


図 1 提案手法概要

2.1 クロマベクトルの学習と和音認識

クロマベクトルは[1]で提供されているツールボックスを用いて算出し、得られたクロマに対しメディアンフィルタで平滑化処理を施した後、ビート同期を行った。ビートの検出・同期には[2]で提供されている Ellis の手法を用いた。

HMM では根音の異なる同じ種類の和音に対して、12 次元クロマベクトルをシフトさせることで同一と見なし学習を行った。また認識時には単一正規分布を用いて出力確率を算出し、forward-backward アルゴリズムで得られる各和音の出力確率の最大値を取り和音の識別を行った。

2.1 後処理による誤り補正

2.1.1 コードベクトル

コードベクトルは major と minor の 24 和音の尤度を持つかを表す 24 次元のベクトルである。HMM により得られた各和音の出力確率をそのままコードベクトルとする。

$$C(t) = \begin{bmatrix} C_C(t) \\ \vdots \\ C_{B_{\min}}(t) \end{bmatrix} \quad (1)$$

$$C_{P_n}(t) = p(P_n | chroma(t))$$

$$P_1 = C, P_2 = C\#, \dots, P_{24} = B_{\min}$$

2.1.2 DNCOF ベクトル

音楽知識 DNCOF[3]とは、構成音が 1 音ずつ異なる major と minor の和音を交互に並べたものである。

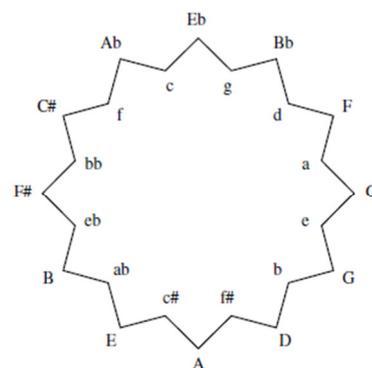


図 2 Double Nested Circle of Fifths(DNCOF)[3]

2.1.1 で得られたコードベクトルを DNCOF 平面へ写像を行う。手順は次の通りである。

- (1) DNCOF を円に見立て、24 和音の方向を向いた単位ベクトルを並べた行列 U を生成する。
- (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) で求めたベクトルを DNCOF ベクトル $DNCOF(t)$ と定める。

$$DNCOF(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = UC_{DNCOF}(t) \quad (2)$$

Chord Recognition by Error Detection and Correction Based on Music Theory

[†]Graduate School of Fundamental Science and Engineering, Waseda University

$$U = \begin{bmatrix} \cos\left(\frac{\pi}{2} - 0 \times \frac{\pi}{24}\right), \dots, \cos\left(\frac{\pi}{2} - 23 \times \frac{\pi}{24}\right) \\ \sin\left(\frac{\pi}{2} - 0 \times \frac{\pi}{24}\right), \dots, \sin\left(\frac{\pi}{2} - 23 \times \frac{\pi}{24}\right) \end{bmatrix} \quad (3)$$

$$C_{DNCOF}(t) = [C_C(t) \ C_{E_{\min}}(t) \ \dots \ C_{A_{\min}}(t)]^T$$

2.1.3 誤り検出・補正

認識誤りが発生する場合は、DNCOF ベクトルとクロマベクトルの認識結果が DNCOF 平面上で離れている場合は間違っている可能性が高い。ここで、クロマベクトルの認識結果を出力したラベルに基づく角度を偏角、そのラベルの尤度を大きさにすることで DNCOF プロットに対応させた。また、距離尺度にはコサイン距離を用い、閾値により誤りをフレームごとに判定する。

そして誤りと推定されたフレームにおいて、補正は単純な手法で行った。今回は DNCOF 順で DNCOF ベクトルに近いクロマベクトルの候補を得ることで補正を行う。ここではコードベクトルの最大値を除いた、補助的な DNCOF ベクトルを抽出し、補正に使用した。

3. 評価実験

本稿で提案する誤り検出手法の有効性を調べるために、はじめに正誤判別率を調べた。そして、補正による性能を調べるために、和音認識による評価を行った。

3.1 実験条件

評価データは The Beatles の楽曲 180 曲を用いて、学習・評価においては[4]で提供されている正解ラベルを利用した。なお、本研究では major と minor の 24 和音を扱うこととし、提供されているラベルは人手で 24 和音に分類した。

音響信号は 32kHz にダウンサンプリングを行い、4096 サンプルのフレームに切り出して特徴量を抽出した。

3.2 誤り検出性能評価

誤りを判定するコサイン距離の閾値は今回実験的に 0.996 と定めた。判別率は次式のように正誤判別できた割合で評価する。

$$\text{判別率}(\%) = \frac{\text{正誤判別できたフレーム数}}{\text{全フレーム数}} \times 100$$

このとき、判別率は 180 曲平均で 77.3%であった。

3.3 誤り補正性能評価

認識は次式のように、入力信号に対し正しく和音を認識できた割合で評価する。

$$\text{認識率}(\%) = \frac{\text{正解和音名を出力した フレーム数}}{\text{全フレーム数}} \times 100$$

3 分割交差検定による評価の結果を表 1 に示す。わずかなではあるが、補正により認識率の向上が確認できた。また、図 3 に DNCOF ベクトル時系列のプロットを示す。誤りと推定されたフレームでは、いくつかずれているものの正解に近い DNCOF ベクトルを得られていることが確認できる。

3.4 考察

誤り判定に用いた閾値は 0.966 であり、偏角の差は 0.262 である。これから、誤りが生じたときに偏角の差は DNCOF 順でいうと 1 つ程度(DNCOF 順 1 つ =

表 1 和音認識結果

比較手法	認識率
従来手法(誤り補正なし)	72.4%
提案手法	73.1%

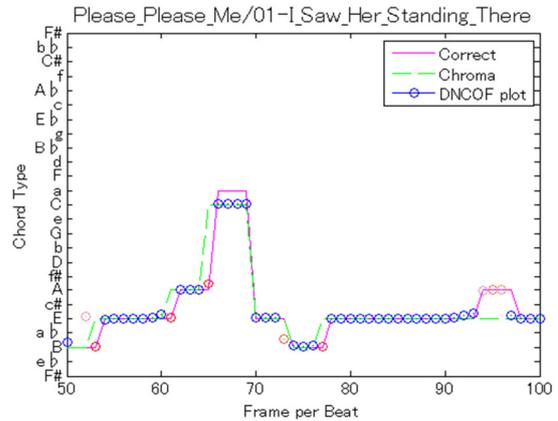


図 3 DNCOF ベクトルの時系列プロット(楽曲:I Saw Her Standing There の 50 から 100 フレーム/ビート)(青: 補正なし, 赤: 誤りと推定された箇所)

$\pi/12 \approx 0.262$)になることがわかる。つまり、クロマベクトルでの和音認識で誤りが発生するときは DNCOF プロットとクロマベクトルのプロットが離れてしまう場合が多いといえよう。

また、補正による認識率が大きく向上していないのは、実際は正解であるのに誤りと判定されてしまったことや、コードベクトルで正解である和音の成分が強く出でおらず、DNCOF ベクトルが強く成分が出ている別の方向に引っ張られてしまったことが原因である。これに対しては、誤り検出手法の改善とクロマベクトルの精度向上が必要である。

しかし、実際 HMM を用いた認識で第 3 候補までを正解とすると認識率は 90%近くになることから、正確な誤り箇所を求められれば、その箇所において適切にクロマベクトル候補の入れ替えを行うだけでも認識率の向上が期待できる。

4. あとがき

本稿では、音楽理論 DNCOF に基づき、クロマベクトルの認識結果と DNCOF の関係性に着目した和音認識の性能改善手法を報告した。本手法は、DNCOF を利用してクロマベクトルによる認識手法で発生する誤りを自動的に推定し、補正を行うことで平均 73.1%の認識率を達成することができた。今後は誤り検出率の精度向上を目指すとともに、適切な補正方法の検討を行っていきたい。

参考文献

- [1] Intelligent Sound Processing: <http://kom.aau.dk/project/isound/>.
- [2] D. P. W. Ellis and G. Poliner: "Identifying cover songs with chroma features and dynamic programming beat tracking", Proc. ICASSP, pages IV 1429-1432, 2007.
- [3] J. P. Bello and J. Pickens, "A robust mid-level representation for harmonic content in music signal," Proc. ISMIR, pp. 304-311, Sep.2005..
- [4] isophonics: <http://isophonics.net/>