楽曲検索に向けたクロマ軸上のパワー分布に基づく類似解析

Chroma-based Analysis of Similarity for Music Retrieval

櫻井良樹[*1] 浅野佑太[*2] 田村哲嗣[*2] 速水悟[*2]
Yoshiki Sakurai Yuta Asano Satoshi Tamura Satoru Hayamizu

[*1] 岐阜大学大学院工学研究科 Graduate School of Engineering, Gifu University
[*2] 岐阜大学工学部 Faculty of Engineering, Gifu University

In this paper, we propose a detection method of similar audio segments for musical pieces, and examine the method utilizing similarity analysis of two time-sequence patterns; musical pieces and fragments. A 12-dimensional chroma vector is extracted from each frame. The feature vector consists of mean and standard deviation of 12 dimensions. For a new music piece, candidate frames are detected from musical pieces in the database, using similarity scores of chroma distribution. These sections are then aligned using a dynamic programming technique in order to deal with time warping of input data. Euclidian distance is adopted to measure a similarity score between two feature vectors. As an application of our method, evaluation experiments were conducted for retrieval of cover songs. By comparing a cover song with fragments of the original song, approximately 83 % detection correctness was achieved.

1. はじめに

近年，インターネットの普及により，音楽配信サービスや，音楽活動が盛んになっている．また，Apple 社の iPod に代表される小型で大容量の携帯音楽プレーヤの普及により，ユーザはお気に入りの楽曲を手軽に持ち歩けるようになった．しかし，Web 上には 100 万曲以上の楽曲が存在し，この中からユーザの要求を満たす楽曲を探し出すのは困難である．このような背景もあり，楽曲検索の研究がおこなわれている．

従来研究には，実環境で受音した楽曲をキーとし，同一楽曲の一致検索をおこなう楽曲探索法[黒住 02]や，楽曲中でのボーカルの声質の類似に従った検索[藤原 07]，楽曲検索楽曲のハーモニーの類似に着目し，カバー曲検索する研究[Kim 08]などがある．

本研究ではクロマベクトルを音響特徴量に用いて，楽曲全体と楽曲断片，2 つの時系列パターンから類似区間の検出をおこなった．また応用例として，カバー曲検索に対する評価実験をおこなった．

2. 類似区間検出

楽曲の断片からクロマベクトルを抽出し，類似区間の検出することを目的とする．以下に処理の流れを示す．

1. 楽曲の断片からのクロマベクトル抽出
2. 楽曲データのフレーム化と特徴抽出
3. 転調処理とフレーム選択
4. 類似区間の検出

2.1 楽曲の断片からのクロマベクトル抽出

楽曲の断片より，12 次元クロマベクトル[後藤 02]を求める．12 次元クロマベクトルとは，STFT などの周波数解析によって得られるパワースペクトルから，オクターブの違いを考慮せず，1 オクターブ 12 音名ごとにパワーを加算したものである．局所領域での音名構成を表す特徴量であるため，コード進行やハーモニーに従って類似するものが得られると期待できる．

2.2 楽曲データのフレーム化と特徴抽出

楽曲と楽曲の断片との類似区間検出にあたって，どの部分に該当区間が存在するか，候補となる部分を決定する．前処理として，図 1 フレーム化の概念に示すように，楽曲の断片のクロマフレーム数 N としたとき，楽曲データをフレーム長 N，フレームシフト N でフレーム化する．

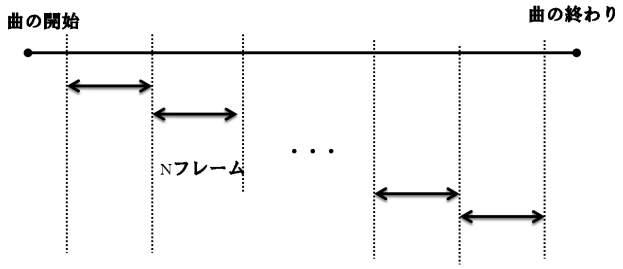

図 1 フレーム化の概念

このとき，各フレーム内のクロマベクトルにおける音名成分 $c(1 \le c \le 12)$の分布情報を表した特徴量を導入する．

ここで，あるクロマベクトル$\boldsymbol{v}(t)$の音名成分毎の平均と標準偏差を式(1)と式(2)で定義する．

$$\mu_c = \frac{1}{N}\sum_{t=1}^{N} v_c(t) \tag{1}$$

$$\sigma_c = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(v_c(t)-\mu_c)^2} \tag{2}$$

特徴量の計算を楽曲では分割したフレーム毎，楽曲断片ではその断片全体を 1 フレームとしてみなし，それぞれおこなう．

以後，本論文ではクロマベクトルの平均と標準偏差は式(1)と式(2)を用いて計算する．

連絡先：櫻井良樹
岐阜大学 大学院 工学研究科 応用情報学専攻
岐阜県岐阜市柳戸 1−1,E−mail sakurai@hym.info.gifu−u.ac.jp

2.3 転調処理とフレーム選択

類似楽曲には，旋律線を維持したまま転調や移調しているものも含まれる．そのため転調処理をおこない，類似区間を含んだフレームの選択をおこなう．

転調処理では，転調や移調に伴う楽曲の断片と楽曲データのクロマベクトルの分布の相対的なずれを転調係数と定義し，分割したフレーム毎に転調係数を推定する．

クロマベクトル $\boldsymbol{v}(t)$ は，各次元 $v_c(t)$ の値を次元間で $j\ (1 \le j \le 12)$個分だけシフトさせることで，転調を表現できる性質を持つ．そこで，式(3)のシフト行列[後藤 02]を用いることで，$\boldsymbol{v}(t)$を j 個上へ転調した演奏のクロマベクトル$\boldsymbol{v}(t)'$を式(4)のように表現できる．

$$S = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{3}$$

$$\boldsymbol{v}(t) \doteqdot S^j \boldsymbol{v}(t)' \tag{4}$$

この性質を利用して断片のクロマベクトルの分布$\boldsymbol{\mu_{in}}, \boldsymbol{\sigma_{in}}$と，楽曲フレーム$i$のクロマベクトルの分布$\boldsymbol{\mu_{fr}}(i), \boldsymbol{\sigma_{fr}}(i)$との類似度計算をおこなう．

類似度は式(5)で定義するユークリッド距離を用いて，平均と標準偏差それぞれにおいて計算して，足し合わせたものとする．このとき，$\boldsymbol{\mu_{fr}}(i), \boldsymbol{\sigma_{fr}}(i)$は$\boldsymbol{\mu_{in}}, \boldsymbol{\sigma_{in}}$を j 個上へ転調した状態であると仮定するため，式(6)で定義するように$S^j\boldsymbol{\mu_{in}}, S^j\boldsymbol{\sigma_{in}}$と$\boldsymbol{\mu_{fr}}(i), \boldsymbol{\sigma_{fr}}(i)$の類似度を計算し，そのときの値を$r(j)$とする．なお，式(6)中の$\alpha$は重み係数であり，予備実験の結果より，0.1と定めた．

$$\hat{d}(x, y) = \frac{\|x - y\|}{\sqrt{12}} \tag{5}$$

$$r(j) = \alpha \hat{d}\left(\boldsymbol{\mu_{fr}}(i), S^j \boldsymbol{\mu_{in}}\right) + (1-\alpha)\hat{d}\left(\boldsymbol{\sigma_{fr}}(i), S^j \boldsymbol{\sigma_{in}}\right) \quad (1 \le j \le 12) \tag{6}$$

シフト状態毎に求めた類似度$r(j)$を比較し，式(7)より断片と楽曲フレームiの類似度$v(i)$とする．また，断片と楽曲フレームとの転調係数$tr(i)$を式(8)より決定する．

$$v(i) = \min_j r(j) \tag{7}$$

$$tr(i) = \operatorname*{argmin}_j r(j) \tag{8}$$

フレーム毎に求めたvを比較し，小さいものから順に 5 フレームを探索フレームとして選択する．

2.4 類似区間の検出

選択した探索フレームの中から類似区間の検出をおこなう．類似区間がフレーム分割の際，連続した 2 つのフレームに存在している場合を考慮して，探索区間は選択した探索フレームの両端に $N/2$ 加えたフレーム長 $2N$ を対象に探索をおこなう．

探索には類似区間の時間伸縮に対応するため，DP マッチングを用いる．DP マッチングでは，N フレームからなる断片のクロマベクトル列$\boldsymbol{v_{in}}(\tau)(1 \le \tau \le N)$と探索区間のクロマベクトル列$\boldsymbol{v_{fr}}(t)(1 \le t \le 2N)$との累積類似度$S$を探索区間方向に連続して求めていく．

まず，$\boldsymbol{v_{in}}(\tau)$と$\boldsymbol{v_{fr}}(t)$における局所類似度$d(t, \tau)$を式(9)に定義する．このとき，$\boldsymbol{v_{in}}(\tau)$は2.3で述べた転調係数$tr(i)$とシフト行列を用いて，値のシフトをおこなう．

$$d(t, \tau) = \hat{d}\left(\boldsymbol{v_{fr}}(t), S^{tr(i)}\boldsymbol{v_{in}}(\tau)\right) \tag{9}$$

次に，図2に示された経路制限に従って式(9)を定義し，求めた局所類似度$d(t, \tau)$から，累積類似度$D(t, \tau)$を更新していく．

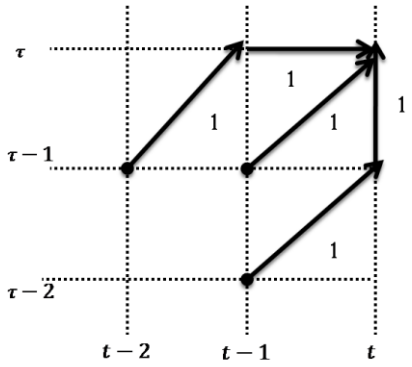

図 2 DP マッチング時の経路制限

$$D(t, \tau) = \min \begin{cases} D(t-2, \tau-1) + d(t-1, \tau) + d(t, \tau) \\ D(t-1, \tau-1) + d(t, \tau) \\ D(t-1, \tau-2) + d(t-1, \tau) + d(t, \tau) \end{cases} \tag{9}$$

累積類似度$D(t, N)$を最小とするtを決定し，これを終端とする探索区間中の類似区間を楽曲の断片と探索フレームの類似区間として切り出す．また，累積類似度を通過したパスの数で正規化した値を類似度とする．

フレーム毎に類似区間を計算したら，区間毎の類似度を比較し，最小値であった区間を類似区間として検出する．

3. 評価実験

提案した手法の有効性を確認するため、評価実験をおこなった．楽曲の断片よりカバー曲の検出をおこない，同一楽曲以外の楽曲での類似区間検出の性能とカバー曲特定に対する有効性を評価した．

3.1 カバー曲の検出方法

カバー曲の検出は，提案した類似区間検出手法に基づいておこなう．手法の対象は，1 つの楽曲の断片に対して複数楽曲である．以後，探索する楽曲のことを参照楽曲と呼ぶ．類似区間の検出は，1 楽曲につき最も類似している 1 区間のみとする．したがって，全体の楽曲数を M としたとき，1 つの楽曲の断片に対し，M 個の区間が検出されることになる．各区間は，類似度比較によって検出されており，楽曲の断片との類似度は計算済みである．そこで，各区間の類似度を比較し，カバー曲の検出をおこなう．楽曲の類似度を$\eta(l)(1 \le l \le M)$と定義し，式(10)より類似度比較をおこなう．

$$m = \operatorname*{argmin}_l \eta(l) \tag{10}$$

参照楽曲の集合を$P = \{P_1, P_2, \cdots, P_M\}$としたとき，$P_m$をカバー曲として，検出する．

3.2 実験の条件と内容

実験は 3.1 に基づいて行う. 参照楽曲には, ボーカルのある楽曲を 538 曲使用した. RWC 楽曲データベースからポピュラー音楽, ジャンル別音楽[後藤 01, 後藤 03]から 127 曲, 個人収集した楽曲 411 曲の計 538 曲を使用した.

楽曲の断片は, 538 曲の中から, カバー曲が存在する 20 曲に限定し, それぞれ A メロ, B メロ, サビの 3 区間を人手によって切り出したものを使用した. 楽曲の断片として切り出しに使用した楽曲には, 原曲に対して, 転調や移調したものやテンポの異なるものも含まれている. また, 楽曲の断片に使用した楽曲は参照楽曲には含まれていない. 楽曲切り出す対象となる区間が繰り返し演奏されている場合, 前半部分を断片として切り出している. 表 1 に楽曲データのサンプリング周波数, 楽曲断片の平均時間長などの実験条件を示す.

表 1 実験条件

サンプリング周波数・量子化ビット	16kHz/16bit
フレームサイズ(FFT)	256ms
フレームシフト(FFT)	128ms
楽曲の断片の平均時間長	26.1s
参照楽曲	538 曲
楽曲の断片数	60 片

評価は対象とするカバー曲において正解区間を検出した検出率とカバー曲の正解率によっておこなった. 正解区間は, 人手によって決定し, カバー曲中における同位置の区間だけでなく, 繰り返し演奏されている区間も対象としており, いずれかが類似区間として検出された場合を正解とした. カバー曲の検出は, 3.1 の方法に基づいて検出された楽曲が, 楽曲の断片の原曲であった場合を正解とした. ただし, 正しい楽曲を検出しても検出区間が誤りであれば不正解とした.

3.3 実験結果と考察

実験の結果を表 2 と表 3 に示す. 表 2 には, 正しいカバー曲中における正解区間の検出率, 表 3 には楽曲断片に対し, 正しいカバー曲の正解率を表している.

表 2 カバー曲中での正解区間検出率

A メロ	B メロ	サビ	平均
80.0%	80.0%	90.0%	83.3%

表 3 カバー曲の正解率

A メロ	B メロ	サビ	平均
70.0%	70.0%	80.0%	73.3%

表 2 より, 平均して約 80%の検出率でカバー曲の類似区間検出をおこなうことができた. 区間毎にみても同様の結果が得られた. 同一伴奏の転調やボーカルが異なるのみの場合, 原曲とカバー曲において同時になっている音名構成は同一であるため, クロマベクトルが特徴量として有効であると考えられる. また, テンポが異なる場合でも正解区間が検出されており, DP マッチングにより時間伸縮が吸収されたためであると考えられる. しかし, カバー曲中での伴奏が大部分において類似したものである場合, 誤検出が見られた. これは, 楽曲のさまざまな部分で類似度が高くなってしまい, 適切な類似区間の検出が難しくなったためであると考えられる.

表 3 をみると, 類似区間の検出率よりも 10%低下していた. これは原曲とカバー曲における伴奏の違いに起因するものと考えられる. 原曲とカバー曲で伴奏の有無など大きな違いが見られる場合, 同一区間内でもクロマベクトルに差が生じる. そのため, 楽曲内から適切な類似区間を検出できた場合でも, 他の楽曲と類似度を比較したときに, 大きく差が見られないため, 出現率が下がったものと考えられる.

4. おわりに

本研究では, クロマベクトルを特徴量として楽曲と楽曲の断片, 2 つの時系列パターンからの類似区間の検出をおこなう手法を提案した. また, 本手法の応用例としてカバー曲検索に向けた評価実験を行った. 実験の結果から, 約 80%の検出率でカバー曲の類似区間検出をおこなうことができた. しかし, カバー曲の検出率となると類似区間の検出率よりも 10%の低下が見られたが, 原因は原曲とカバー曲における伴奏の違いに起因するものであった.

今後は, ハーモニー以外の音楽要素を表す特徴量との併用について検討していく.

参考文献

[黒住 02] 黒住隆行, 柏野邦夫, 村瀬洋:実環境で受音した楽曲をキーとする楽曲探索法. 電子情報通信学会論文誌 D-Ⅱ Vol.J86-D-Ⅱ No.12 pp.1719-1726, 2003

[藤原 07] 藤原弘将, 後藤真孝:VocalFinder:声質の類似度に基づく楽曲検索システム. 情報処理学会 音楽情報科学研究会 研究報告 2007-MUS-71-5, pp.27-32, 2007

[Kim 08] Samuel Kim, ShrikanthNarayanan: Dynamic chroma feature vectors with applications to cover song identification. MMSP 2008: 984-987, 2008

[後藤 02] 後藤真孝:リアルタイム音楽情景記述システム: サビ区間検出手法. 情報処理学会 音楽情報科学研究会 研究報告 2002-MUS-47-6, pp.27-34, 2002

[後藤 01] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一:
RWC 研究用音楽データベース:ポピュラー音楽データベースと著作権切れ音楽データベース. 情報処理学会 音楽情報科学研究会 研究報告 2001-MUS-42-6, pp.35-42, 2001

[後藤 03] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一:
RWC 研究用音楽データベース:音楽ジャンルデータベースと楽器音データベース. 日本音響学会 2003 年春季研究発表会 講演論文集, 3-7-6, pp.843-844, 2003