

隠れマルコフモデルによる自動和音認識に基づく
カバー演奏ストリームからの原曲同定システム

筑波大学

図書館情報メディア研究科

2013年3月

中沢 彰吾

目次

第 1 章	序論	5
1.1	本研究の背景	5
1.2	本研究の目的	6
1.3	本研究のアプローチ	6
1.4	本論文の構成	7
第 2 章	本研究の基礎理論	8
2.1	特徴抽出手法	8
2.1.1	定 Q 変換	8
2.1.2	クロマベクトル	10
2.2	ベクトル量子化	12
2.3	パターンマッチング手法	14
2.3.1	DP マッチング	14
2.3.2	連続 DP	15
2.3.3	ShiftCDP 法	17
2.4	隠れマルコフモデル (HMM)	19
2.4.1	HMM のパラメータ学習	20
2.4.2	HMM による認識	24
第 3 章	先行研究	26
3.1	HMM に基づく自動和音認識	26
3.1.1	HMM に基づく和音進行のモデル化	26
3.1.2	和音 HMM の学習と認識	27
3.2	カバー演奏ストリームからの原曲同定システム	28
3.2.1	従来手法のシステム構成	28
3.2.2	問題点	29
第 4 章	提案手法	31
4.1	システム構成	31
4.2	要素技術	32
4.2.1	和音進行パターンへの符号化	32
4.2.2	転置インデックスによる候補パターン削減	33
4.2.3	ShiftCDP 法による音響的類似度スコアリング	34
4.3	従来手法との相違点	35

第 5 章	評価実験	36
5.1	実験概要	36
5.2	予備実験：HMM に基づく自動和音認識	36
5.2.1	実験方法	36
5.2.2	実験条件	37
5.2.3	実験結果	37
5.3	本実験：カバー演奏ストリームからの原曲同定	38
5.3.1	実験方法	38
5.3.2	実験条件	40
5.3.3	実験結果	41
5.3.4	考察	44
第 6 章	結論	45
6.1	まとめ	45
6.2	今後の課題	45
	謝辞	47
	参考文献	47
	本研究に関する研究発表	50
	付録 A HMM の学習に用いたデータセット	51
	付録 B 本実験で用いたデータセット	55
	付録 C 従来手法の評価に用いた正解ラベル	58
	付録 D 提案手法の評価に用いた正解ラベル	63

目次

2.1	STFT と定 Q 変換のスペクトログラムの比較	9
2.2	クロマベクトルの概念	10
2.3	クロマベクトルの時系列の例	11
2.4	ベクトル量子化の動作	12
2.5	二つのパターン間の対応関係	14
2.6	DP マッチングの許容パスと移動コスト	15
2.7	連続 DP の動作の概要	16
2.8	連続 DP で用いる許容パス	16
2.9	ShiftCDP 法の概念	18
2.10	HMM の例	19
2.11	HMM の学習と認識	21
2.12	Viterbi アルゴリズムの概念	25
3.1	和音 HMM の学習	27
3.2	和音 HMM に基づく自動和音認識	28
3.3	従来手法のシステム構成	28
4.1	提案手法のシステム構成	31
4.2	バッファリングした入力信号に対する自動和音認識	33
4.3	和音進行パターンの転置インデックスの作成 (4-gram の例)	34
5.1	予備実験の結果：和音認識率の分布	38
5.2	本実験の結果：従来手法と提案手法の性能比較 (適合率-再現率曲線)	42

表目次

4.1	従来手法と提案手法の相違点	35
5.1	予備実験の実験条件	37
5.2	予備実験の結果：和音認識率	37
5.3	予備実験の結果：和音認識率が 10.00 [%] 未満の楽曲	38
5.4	本実験の実験条件	40
5.5	本実験の結果：提案手法の楽曲同定性能 (F 値)	41
5.6	本実験の結果：閾値に対する適合率, 再現率, F 値	42
5.7	本実験の結果：N-gram における第 1 段階後の再現率と平均候補パターン数	43
5.8	本実験の結果：第 1 段階, 第 2 段階の平均計算時間 [msec]	43
5.9	提案手法の第 1 段階で正解パターンが残らなかった楽曲 (5-gram の場合)	44
A.1	和音 HMM の学習に用いたデータセット 曲番号 1-50	51
A.2	和音 HMM の学習に用いたデータセット 曲番号 51-100	52
A.3	和音 HMM の学習に用いたデータセット 曲番号 101-150	53
A.4	和音 HMM の学習に用いたデータセット 曲番号 151-180	54
B.1	本実験で用いたデータセット 曲番号 1-20	55
B.2	本実験で用いたデータセット 曲番号 21-50	56
B.3	本実験で用いたデータセット 曲番号 51-80	57
C.1	従来手法の評価に使用した正解ラベル (1)	58
C.2	従来手法の評価に使用した正解ラベル (2)	59
C.3	従来手法の評価に使用した正解ラベル (3)	60
C.4	従来手法の評価に使用した正解ラベル (4)	61
C.5	従来手法の評価に使用した正解ラベル (5)	62
D.1	提案手法の評価に使用した正解ラベル (1)	63
D.2	提案手法の評価に使用した正解ラベル (2)	64
D.3	提案手法の評価に使用した正解ラベル (3)	65

第1章 序論

1.1 本研究の背景

近年，インターネット上でライブストリーミング配信を行うことができるサービスが普及してきている．このようなサービスを用いて個人が既存の楽曲をコピーして演奏したり，原曲とは異なる編曲でカバー演奏する様子を配信する事例が数多く存在する．既存の楽曲を演奏する配信を合法的に行うには，著作者や著作権管理団体に対する申請が不可欠である．しかし，現状では申請が行われていない違法な配信も数多く存在する．これらを，サービス運営者が膨大な数の配信の中から人手で見つけ出し，配信の遮断などの処理を行うのは困難である．この問題に対して，ライブストリーミング配信の中で演奏中の原曲名をほぼ実時間で，自動的に同定するシステムを開発したいという要求がある．また，このようなシステムの実現は，著作権管理が容易になるというメリットだけでなく，ライブストリーミングの視聴者に演奏中の原曲名を提示するようなコンテンツの拡張にも有用である．

前述のようなシステムの実現を考えた場合，コンピュータで実現可能なのは配信されている演奏と原曲との類似性の算出までであり，偶然類似した別の楽曲の演奏であるのか，原曲を意識したカバー演奏であるのかを自動的に判断させることは困難である．最終的には人間の判断が必要となるが，その前段として，いかに類似する楽曲候補の絞り込みを行い，適切な候補を人間に提示するかという問題は，負担軽減の観点から非常に重要となる．

実際に，カバー演奏の音楽音響信号から原曲同定を行う研究は類似楽曲検索の分野に分類され，現在までに様々な研究が行われている [1–6]．これらの先行研究は，時間長が既知の1曲の音楽音響信号に対する原曲同定がタスクとなっているため，ライブストリーミング配信のように時間長が未知で複数楽曲が連続する演奏ストリームにはそのまま適用可能な手法ではない．

また，時間長未知の音響信号に対する楽曲同定の先行研究には井田らの研究 [7] が挙げられる．この研究は，既存の音楽 CD に入っている音楽音響信号がマスメディア上で利用された場合に楽曲を同定することを目的としており，カバー演奏に対応できる音響特徴量を採用していない．

このように，関連した先行研究は行われていたが，前述の要件を満たすようなカバー演奏ストリームからほぼ実時間で原曲同定可能な手法に関する研究は，従来はほとんど見られなかった．この研究については，深澤ら [8] の先行研究が初の試みとなっている．

深澤らは，音楽音響信号から抽出する音響特徴量として短時間周波数解析結果を音階表現に圧縮したクロマベクトルを採用し，クロマベクトルの時系列をベクトル量子化によって符号化したパターンを標準パターンとして，標準パターンのデータベースを作成した．そして，カバー演奏の音楽音響信号からクロマベクトルを逐次的に抽出・符号化したものを入力パターンとし，連続 DP によって入力パターンとデータベース内の各標準パターンとの最適累積距離を計算して，最適累積距離に対する閾値処理を行うことで，ほぼ実時間で原曲同定を行うシステムを構

築した．このシステムを用いた評価実験の結果は，40 曲が演奏されるカバー演奏ストリームをシステムに入力した場合で，適合率 72.1[%]，再現率 17.7[%]，最大 F 値が 0.284 であった．この研究によって，カバー演奏ストリームからの原曲同定の実現可能性が示されたが，候補曲数に対する原曲同定精度の低さが課題として残っていた．

これに対し筆者は，深澤らが特徴量として用いたクロマベクトルに対して打楽器成分分離手法を適用し，原曲同定精度の向上を試みた [9]．この手法を用いた同条件での評価実験では，深澤らの手法よりも原曲同定精度に若干の改善が見られ，F 値は最大で 0.335 となった．しかし，この手法でも依然として原曲同定の精度は十分とは言えなかった．

また，これらの二つの手法には原曲同定精度の問題に加えて，楽曲データベース内の全標準パターンとの連続 DP の出力を逐次更新するため，楽曲データベース拡張に伴う計算量の増加が大きいという問題，また，標準パターン生成に用いる原曲の音響信号を入力パターンと類似しやすい区間で人手で分割しておく必要があるため，楽曲データベース拡張に伴う負荷が高いという問題が共通して存在していた．システムをより実用に近づけるためには，これら三つの問題を解決する新しい原曲同定手法を開発する必要がある．

1.2 本研究の目的

本研究では，従来手法の問題点を解決する新しいカバー演奏ストリームからの原曲同定手法を開発することを目的として，新規手法を提案し，評価実験によってその有効性を検証する．具体的には，隠れマルコフモデル (Hidden Markov Model : HMM) による自動和音認識，転置インデックスによる候補パターン削減，ShiftCDP 法による音響的類似度スコアリングの三つの要素技術を組み合わせた新規手法を提案する．

1.3 本研究のアプローチ

前述したように，従来手法では音楽音響信号から抽出したクロマベクトルをベクトル量子化によって符号化し，その符号パターンをマッチングに用いていた．特徴量としてクロマベクトルを用いるのは，原曲とカバー演奏の間で和音進行が保存されている可能性が高いという仮定に基づいたものである．しかし，本来同一和音が演奏されている区間でも編曲によっては和声内音の省略や非和声音の挿入などがあり，抽出されるクロマベクトルは一樣ではなく，ある程度のばらつきが出てしまう．また，分析フレームごとのクロマベクトルの符号化では，符号化の際に分析フレームの前後の情報が考慮されないため，突発的な特徴量の揺れがそのままパターンに反映されてしまうという問題がある．これらの問題によって，従来手法の符号化手法では，原曲とカバー演奏の音響信号から得られるパターンの中に差が出やすい．これがマッチング結果に影響し，原曲同定精度が低くなってしまふと考えられる．これに対して，本研究の提案手法では，HMM による自動和音認識手法により，抽出したクロマベクトルから最尤和音系列を推定し，推定した和音進行パターンをマッチングに用いる．この手法では，クロマベクトルの和音ごとの分布を HMM の一つの状態が持つ混合正規分布として表現できる．また，音響的な特

徴だけでなく、和音進行のテキストデータのから学習した和音間の遷移しやすさの情報を含めることができる。そのため、原曲とカバー演奏の特徴量に多少の差があっても、本来同一の和音が演奏されている区間であれば、前後のフレームの情報を含めた最尤推定により、ほぼ同一のパターンへの符号化が期待できる。本研究ではこの手法により、原曲同定精度の向上を図る。

また、従来手法のパターンマッチング手法は、連続 DP によってデータベース内の各標準パターンとの最適累積距離を入力フレームごとに計算し、閾値処理を行うというものであった。この手法は、データベースの登録曲数が比較的少ない場合は実時間で対応可能であるが、曲数の拡張に伴って比例的に計算量が増加してしまうという問題がある。提案手法ではこの問題の対策として、テキストベースの第 1 段階のマッチングと、音響的な類似度を用いた第 2 段階のマッチングの 2 段階からなるマッチング手法を採用し、計算量削減を図る。第 1 段階では、標準パターン内に出現する和音進行パターンの転置インデックスを作成しておき、それを用いて入力パターンに含まれる和音進行を持つ標準パターンだけに候補を絞り込むことで、第 2 段階のマッチングを行うパターン数を削減する。第 2 段階では、事前のパターン分割が不要な類似区間検出手法である ShiftCDP 法を用いて、第 1 段階で絞り込んだパターンに対してのみ最適累積距離を計算し、閾値処理を行って、最終的な原曲同定結果とする。第 1 段階はテキストベースの処理のため、音響的な類似度によるマッチングよりも比較的高速に処理を行うことができる。そのため、第 1 段階で正解パターンを候補から落とさずに、できるだけ候補パターンを削減することができれば、曲数の拡張に伴う計算量の増加を抑えることができると考えられる。本研究ではこの手法により、従来手法に存在しているデータベースの拡張に伴う計算量の増加の問題、人手による事前のパターン分割が必要である問題の二つの問題の解決を図る。

1.4 本論文の構成

本論文は全 6 章から構成される。

第 2 章では、本研究の基礎となる理論として、特徴抽出手法、ベクトル量子化、パターンマッチング手法、隠れマルコフモデル (HMM) について述べる。

第 3 章では、提案手法の基礎となる先行研究として、HMM に基づく自動和音認識の研究を紹介する。また、カバー演奏ストリームからの原曲同定システムの従来手法について全体構成を交えて解説し、その問題点を述べる。

第 4 章では、提案手法の全体構成と要素技術について解説する。

第 5 章では、提案手法の有効性の検証のために行った評価実験の方法とその結果を示す。

第 6 章にて、全体の結論を述べ、今後の課題を述べる。

第2章 本研究の基礎理論

本章では、本研究の基礎理論として、特徴抽出手法、ベクトル量子化、パターンマッチング手法、隠れマルコフモデル (HMM) について述べる。

2.1 特徴抽出手法

2.1.1 定Q変換

音響信号の時間周波数解析手法として、短時間フーリエ変換 (Short-Time Fourier Transform: STFT) がよく知られている。STFT は固定長の分析フレーム幅を設けて、切り出したフレーム内の信号に対して離散フーリエ変換 (Discrete Fourier Transform: DFT) を行う手法である。STFT の周波数分解能は分析フレーム長に依存し、解析結果の周波数軸の中心周波数の間隔は分析フレーム長を T [sec] とした場合、 $1/T$ [Hz] の固定値で表される。すなわち、どの周波数帯域においても一定の周波数分解能となる。

音楽音響信号を分析する場合、楽音の基本周波数の変化は対数的であるため、低周波数帯域では高い分解能が必要となる。低周波数帯域にある楽音の周波数成分を分離よく解析したい場合、STFT で用いる分析フレーム長を拡大して周波数分解能を高めることが考えられるが、一方で時間分解能が犠牲になってしまうという問題が生じる。この問題に対し、周波数解析結果の中心周波数を対数的に変化させる多重解像度解析手法である定Q変換 (Constant-Q Transform) [10] があり、音楽音響信号に適した周波数解析手法として知られている。

入力信号 $x(n)$ の定Q変換 $X_{cq}(k_{cq})$ は次式で表される。

$$X_{cq}(k_{cq}) = \frac{1}{N_{k_{cq}}} \sum_{n=0}^{N_{k_{cq}}-1} w_{k_{cq}}(n)x(n)e^{-j2\pi Qn/N_{k_{cq}}} \quad (2.1)$$

ここで、 $N_{k_{cq}} = Qf_s/f_{k_{cq}}$ ($f_{k_{cq}} = 2^{k_{cq}/b}f_{min}$, f_{min} は任意で設定する最小周波数) は可変のフレーム幅であり、 $w_{k_{cq}}(n)$ は分析フレームに対する窓関数である。また、 $Q = 1/(2^{1/b} - 1)$ (b はオクターブごとの周波数 bin 数) である。

定Q変換は計算効率ではSTFTに劣るが、STFTの計算結果を利用した改良アルゴリズム [11] によって高速に計算を行うことができる。このアルゴリズムでは、あらかじめ次のように定義

したカーネル K を計算しておく .

$$K(k_{cq}, k) = \sum_{n=0}^{N_0-1} w_{k_{cq}}(n) e^{jw_{k_{cq}} n} e^{-j2\pi kn/N_0} \quad (2.2)$$

入力信号 $x(n)$ の STFT 結果 $X(k)$ を用いると , 定 Q 変換 $X_{cq}(k_{cq})$ は次式で表される .

$$X_{cq}(k_{cq}) = \frac{1}{N_0} \sum_{k=0}^{N_0-1} X(k) K(k_{cq}, k) \quad (2.3)$$

ここで例として , The Beatles の 『We Can Work It Out』 の A メロ部分を抜粋したモノラル音響信号 (標本化周波数 16kHz , 量子化ビット数 16bit) について , 分析フレーム長 8192 点 , フレームシフト間隔 1024 点で分析した場合の STFT と定 Q 変換のスペクトログラムの比較を図 2.1 に示す . 図の上段が STFT , 下段が定 Q 変換によるスペクトログラムである . 定 Q 変換では中心周波数対数的に変化し , 特に 500 [Hz] 以下の低周波数帯域において , 楽音の変化をより細かく捉えられていることが分かる .

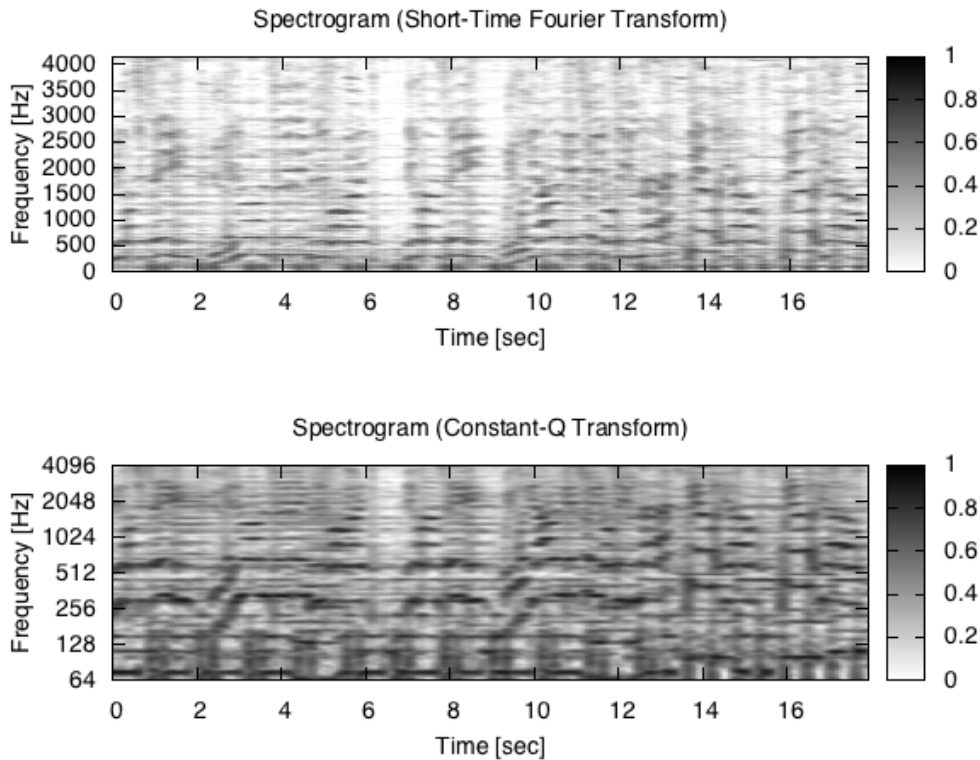


図 2.1: STFT と定 Q 変換のスペクトログラムの比較

2.1.2 クロマベクトル

クロマベクトル (chroma vector) は、音楽音響信号に対して短時間周波数解析を行った結果を音階表現に圧縮し、クロマ (chroma, 音名) を周波数軸としてパワーの分布を表現した特徴量である。十二平均律では、1 オクターブが 12 半音からなることから、クロマベクトルは 12 次元として計算されるのが一般的である。

図 2.2 にクロマベクトルの概念図を示す。図に示すように、クロマベクトルは異なるオクターブのパワーを加算して音名ごとに 1 次元とする。そのため、和音の転回形のように演奏する和音の構成音の高低が入れ替わったり、楽器の違いによって演奏されるオクターブが異なっても、同一音名の音が演奏されていれば、類似したベクトルが得られる。

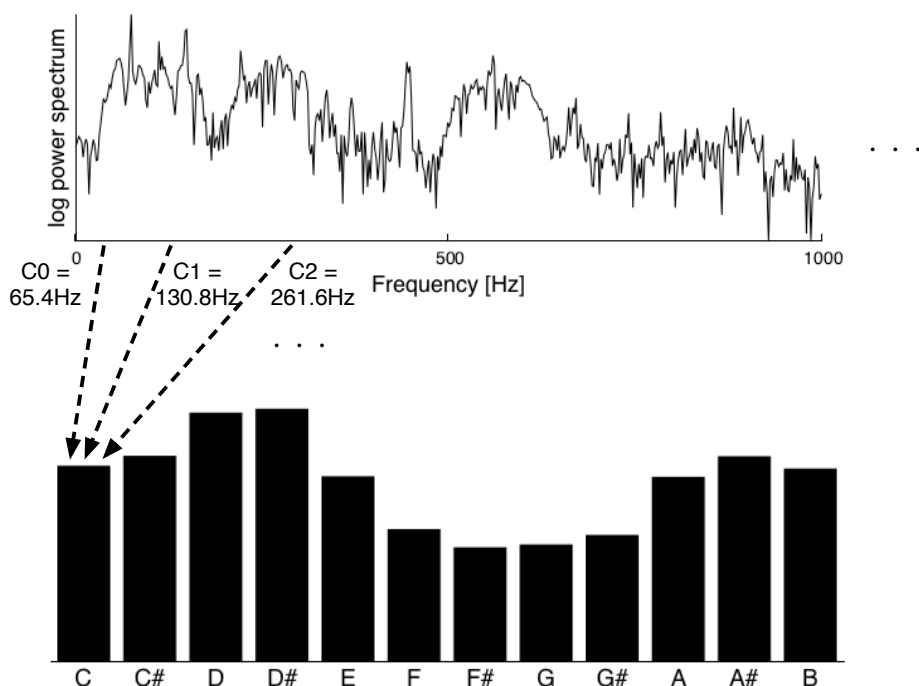


図 2.2: クロマベクトルの概念

クロマベクトルの主な計算手法には、STFT によって計算されるパワースペクトルに対してバンドパスフィルタを適用し、音名に該当するパワーを足し合わせる手法 [12] と、定 Q 変換から計算されるバンドごとのパワーを足し合わせる手法 [13] があるが、本研究では定 Q 変換に基づく手法を用いる。その理由は、前節で述べたように時間分解能を下げずに、低周波数帯域の周波数分解能を高めることができるためである。

12 次元クロマベクトル CH は、入力信号の定 Q 変換 X_{cq} から次式で計算される。

$$CH(b) = \sum_{m=0}^M |X_{cq}(b + 12m)| \quad (1 \leq b \leq 12) \quad (2.4)$$

ここで、 M は加算するオクターブの上限を表す。

また、クロマベクトルは、各次元の値をシフトさせることで、元の調からの転調、移調を表現することができるという性質を持つ [14]。元の調から転調係数 tr 回だけ上に転調したクロマベクトルを CH_{tr} とすると、シフト行列 S を用いて次のように計算できる。

$$CH_{tr} \approx S^{tr}CH \quad (2.5)$$

ここで、シフト行列 S は次式で表される。

$$S = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix} \quad (2.6)$$

このクロマベクトルをパターンマッチングに用いる際には、次のようにベクトルの要素の最大値で正規化したクロマベクトル CH' とする。

$$CH' = \frac{CH}{\max_{b=1..12} \{CH(b)\}} \quad (2.7)$$

例として、図 2.1 に示した定 Q 変換の結果からクロマベクトルの時系列を計算したものを図 2.3 に示す。図に示す通り、クロマベクトルは分析区間内で強勢な音名の成分を表し、時系列とすることで、音楽音響信号の和声の変化を表す特徴量となる。

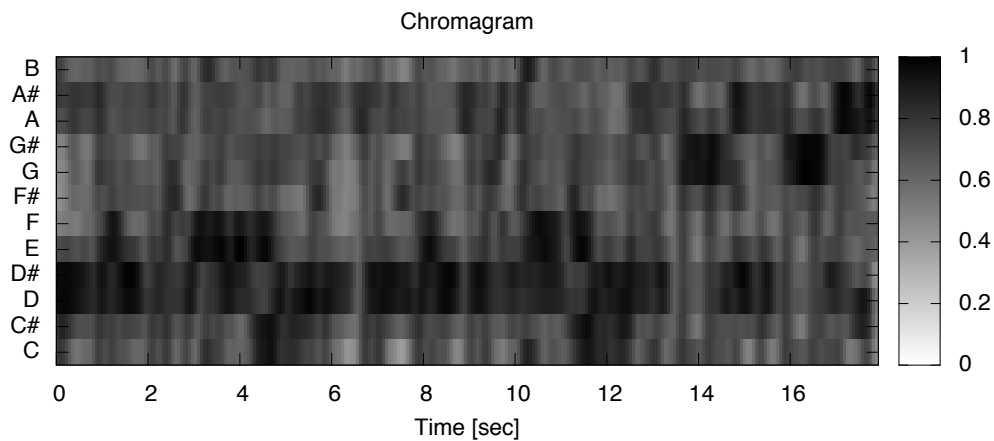


図 2.3: クロマベクトルの時系列の例

2.2 ベクトル量子化

ベクトル量子化は、複数のサンプルをまとめてベクトルとして量子化することによって、サンプル間の冗長度を情報量削減に利用する技術である [15]。スカラ量子化の場合は、有限個の量子化レベルを用意しておき、入力サンプル値に最も近い量子化レベルを選択し、その量子化レベルを識別するインデックスに符号化する。それに対しベクトル量子化では、量子化レベルの代わりに代表的なベクトルのパターン（量子化代表ベクトル）を有限個用意しておき、入力ベクトルが与えられると、そのベクトルに最も近い量子化代表ベクトルを選択し、そのインデックスに符号化する。このとき、インデックスと量子化代表ベクトルの複数の組をコードブックと呼び、コードブックの大きさ N をコードブックサイズと呼ぶ。ベクトル量子化の動作イメージを図 2.4 に示す。

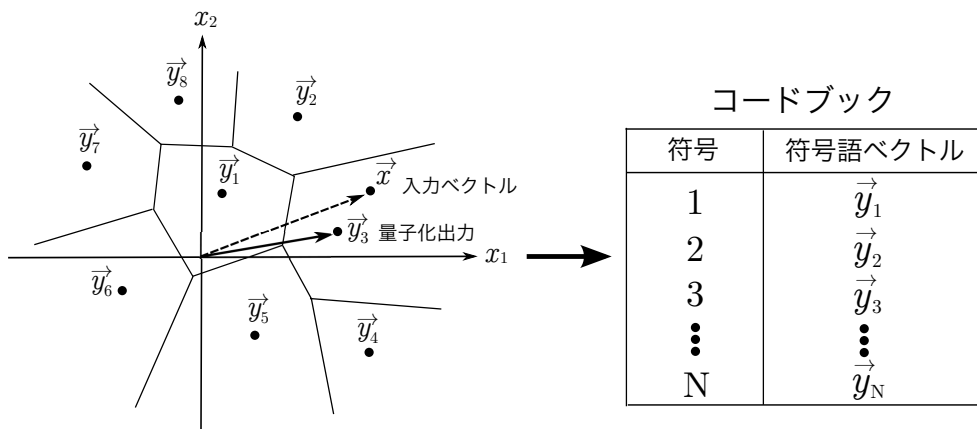


図 2.4: ベクトル量子化の動作

実用的なベクトル量子化手法の 1 つに Linde らが開発した LBG アルゴリズムがある [16]。

ベクトルの次元数を K ，量子化代表ベクトル数を N ，初期符号帳 $C_N^{(0)} = \{\vec{y}_1^{(0)}, \dots, \vec{y}_N^{(0)}\}$ ，学習系列 $T = \{\vec{x}_1, \dots, \vec{x}_L\}$ ，収束判定用閾値 ε が与えられているとし、また、 $m = 0$ ，初期ひずみ $D_{-1} = \infty$ と設定すると、以下のアルゴリズムにより符号帳が得られる。

1. 分割条件の適用と平均ひずみの算出

学習系列 T に対して、符号帳 $C_N^{(m)} = \{\vec{y}_1^{(m)}, \dots, \vec{y}_N^{(m)}\}$ の下で平均ひずみを最小とする分割 $P_N^{(m)} = \{P_1^{(m)}, \dots, P_N^{(m)}\}$ の部分空間 $P_i^{(m)}$ を次の式で求める。

$$P_i^{(m)} = \{\vec{x}_l | d(\vec{x}_l, \vec{y}_i^{(m)}) \leq d(\vec{x}_l, \vec{y}_j^{(m)}) \text{ for all } j \neq i, \vec{x}_l \in T\} \quad (2.8)$$

また，平均歪み $D^{(m)}$ を次式にて算出する．

$$D^{(m)} = \frac{1}{L} \sum_{i=1}^N \sum_{\vec{x}_i \in P_i^{(m)}} d(\vec{x}_i, \vec{y}_i^{(m)}) \quad (2.9)$$

2. 収束判定

もし， $(D^{(m-1)} - D^{(m)})/D^{(m)} < \varepsilon$ ならば，処理を停止し， $C_N^{(m)}$ を最終的に設計された N レベルの符号帳とする．そうでなければ，3 へ行く．

3. 代表点条件の適用

学習系列 T に対して，分割 $P_N^{(m)} = \{P_1^{(m)}, \dots, P_N^{(m)}\}$ の下で平均歪みを最小とする符号帳 $\hat{C}_N = \{\hat{y}_1, \dots, \hat{y}_N\}$ の量子化代表ベクトル $\vec{\hat{y}}_i$ を次式で定める．

$$\vec{\hat{y}}_i = \frac{1}{\|P_i^{(m)}\|} \sum_{\vec{x}_i \in P_i^{(m)}} \vec{x}_i \quad (2.10)$$

ただし， $\|P_i^{(m)}\|$ は部分空間 $P_i^{(m)}$ に所属している学習ベクトルの総数を表す．さらに，符号帳 \hat{C}_N を $C_N^{(m)}$ とし， $m = m + 1$ として 1 に戻る．

LBG アルゴリズムによって設計される符号帳の良否は初期符号帳 $C_N^{(0)}$ と学習系列 T の選定法に強く依存する．初期符号帳 $C_N^{(0)}$ は想定される入力ベクトルの分布範囲を被覆していることが望ましい．この条件をある程度満足する初期符号帳の生成法として，スプリッティングアルゴリズムが知られている．

このアルゴリズムは， N レベルの符号帳 $C_N = \{\vec{y}_1, \dots, \vec{y}_N\}$ の量子化代表ベクトル \vec{y}_i を，次式のように微小なベクトル $\vec{\varepsilon}$ を用いて近接した 2 つのベクトル \vec{y}_i と \vec{y}_{i+N} に分割し， $2N$ レベルの初期符号帳 $C_{2N}^{(0)} = \{\vec{y}'_1, \dots, \vec{y}'_{2N}\}$ を生成するものである．

$$\vec{y}'_i = \vec{y}_i - \vec{\varepsilon}, \quad \vec{y}'_{i+N} = \vec{y}_i + \vec{\varepsilon} \quad (2.11)$$

このスプリッティングアルゴリズムと LBG アルゴリズムを組み合わせることによって，1 レベルの符号帳から出発して，順次 2, 4, 8, \dots レベルの符号帳を設計することができる．ただし，1 レベルの符号帳は，全学習ベクトルの重心ベクトルを量子化代表ベクトルとする符号帳である．

一方，学習系列 T は，想定される入力ベクトルの特性を代表するのに十分な個数の学習ベクトルを含んでいなければならない．学習ベクトルの個数が少ない場合には，学習系列と異なる入力ベクトル系列に対して望ましい結果が得られないことがある．このようなことが生じないようにするためには，学習ベクトルの個数を，少なくともレベル数 N の 10 倍から 100 倍程度にしなければならない．

2.3 パターンマッチング手法

2.3.1 DP マッチング

標準パターンと入力パターンの時間領域での対応付けの手法として、動的計画法 (Dynamic Programming: DP) によるマッチング手法がある。特徴ベクトルの時系列で表された二つのパターンを $A = a_1, a_2, \dots, a_i, \dots, a_I$ と $B = b_1, b_2, \dots, b_j, \dots, b_J$ とすると、両パターン間の対応関係は、図 2.5 に示すように平面上の格子点 $c_k = (i_k, j_k)$ の系列 $C = c_1, c_2, \dots, c_k, \dots, c_K$ で表すことができる。

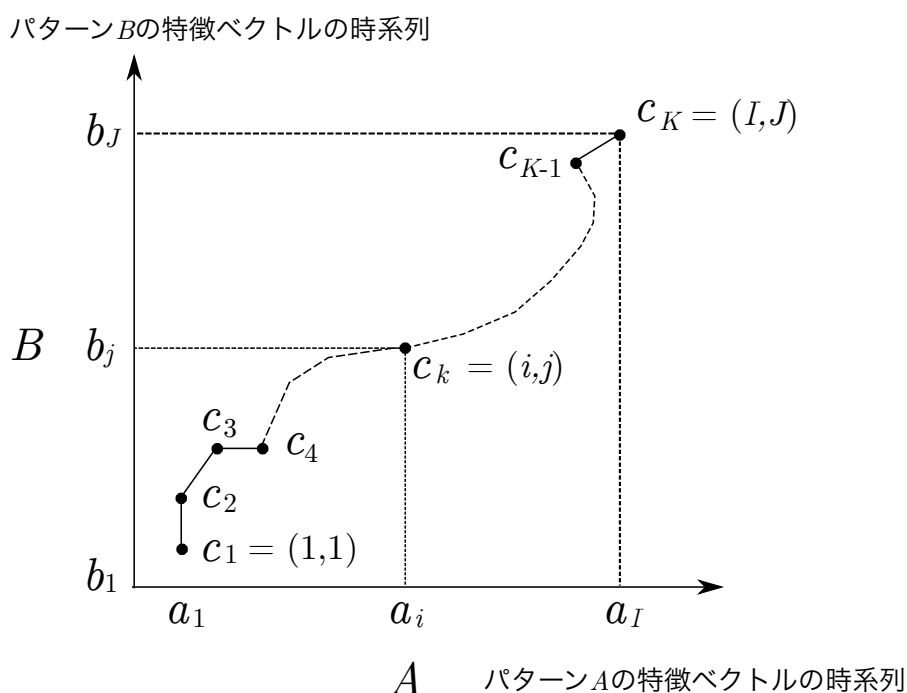


図 2.5: 二つのパターン間の対応関係

このとき、 A, B が最も一致するような C (経路と呼ぶ) を求めるのが DP マッチングである。 C の変化にはあまり極端な時間軸の伸縮は許されないように許容パスを定め、格子点間の移動に関してコストを設ける。図 2.6 に、基本的な許容パスと移動コストの例を示す。

ここで、ある格子点 $c_k = (i_k, j_k)$ において、対応する二つの特徴ベクトル a_{i_k} と b_{j_k} との距離を $d(i_k, j_k)$ で表し、始点から c_k までの $d(i_k, j_k)$ と移動コストとの累積和の最小値を $g(i_k, j_k)$ とすると、動的計画法に基づいて、 c_k までの最小の累積値 (最適累積距離) は、 c_{k-1} までの最適

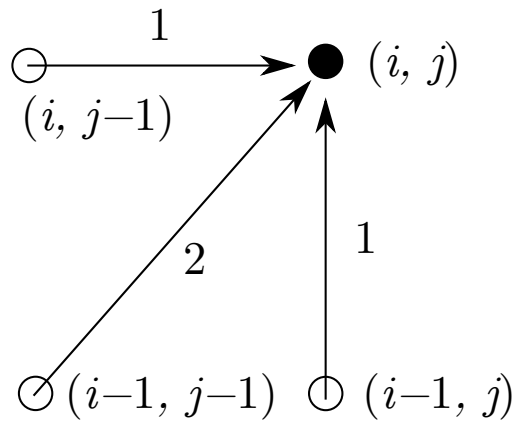


図 2.6: DP マッチングの許容パスと移動コスト

累積距離を用いて次のように表せる .

$$g(i_k, j_k) = \min \begin{bmatrix} g(i_{k-1}, j_k) + d(i_k, j_k) \\ g(i_{k-1}, j_{k-1}) + 2d(i_k, j_k) \\ g(i_k, j_{k-1}) + d(i_k, j_k) \end{bmatrix} \quad (2.12)$$

この漸化式に従って順次 $g(i_k, j_k)$ を計算していくと、終端に対応する (I, J) での累積値 $g(I, J)$ が求められる . この値が小さいほど、 A と B との対応付けがよいことになる . また、各格子点での漸化式の計算時に、最小となる経路間の移動パスを記録しておけば、終端からのバックトレースにより、最適な経路が求められる .

2.3.2 連続 DP

DP マッチングは二つの不等長パターンを比較して最適なマッチングを行う手法であったが、その時間長は共に固定長でなければならない . これに対して、一方のパターンを任意長にし、その中から他方の固定長パターンと類似する区間を検出する手法として連続 DP [17] がある . この場合、前者が入力パターン、後者が標準パターンとなる .

連続 DP ではまず、入力パターンの先頭と標準パターンとを DP マッチングにより対応させ、最適累積距離を求める . 次に入力パターンをずらし、同じく標準パターンと DP マッチングにより対応させ、最適累積距離を求める . これを入力パターンの最後まで行い、入力パターンの各区間における標準パターンとの最適累積距離を求めていく . このとき、適当な閾値を設定しておき、求めた距離が閾値以下ならば、図 2.7 のようにその区間が標準パターンとの類似区間として検出されることになる . 連続 DP における閾値は 2 つのパターンの類似度の距離値であり、閾値が小さいほど検出の条件が厳しくなる .

連続 DP に用いる許容パスは図 2.8 のように定める .

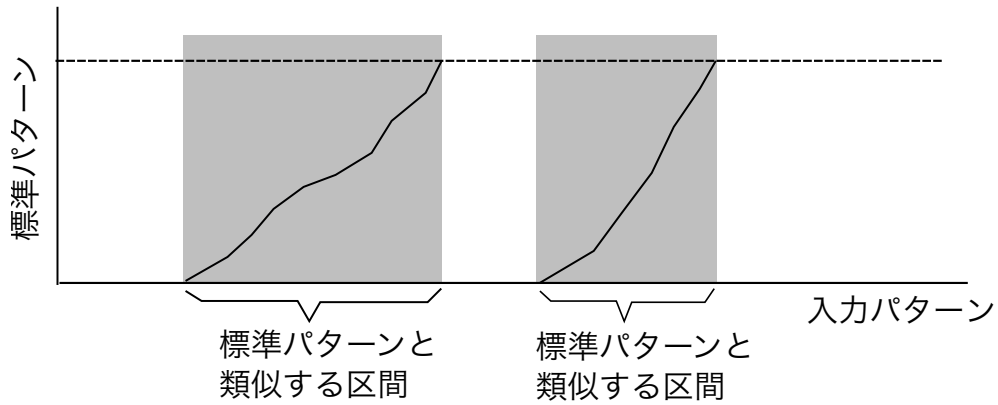


図 2.7: 連続 DP の動作の概要

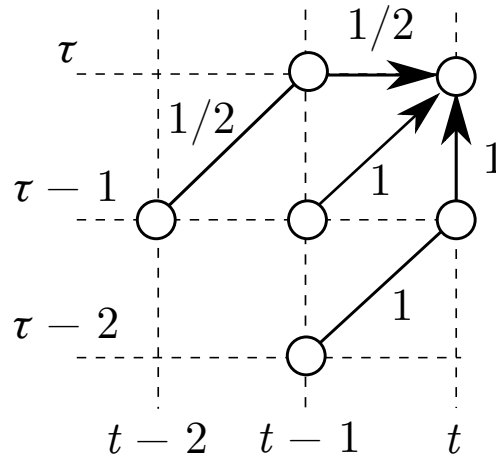


図 2.8: 連続 DP で用いる許容パス

ここで、特徴ベクトルの時系列で表された入力パターンを $A = a_1, a_2, \dots, a_t, \dots$ 、標準パターンを $B = b_1, b_2, \dots, b_\tau, \dots, b_T$ とし、ある格子点 $c_k = (i_k, j_k)$ において対応する二つの特徴ベクトル a_{i_k} と b_{j_k} との距離を $d(i_k, j_k)$ で表すと、図 2.8 の許容パスにおける累積距離 S は次の漸化式で計算できる。

初期条件 ($t = 0$) :

$$S(-1, \tau) = S(0, \tau) = \infty \quad (1 \leq \tau \leq T) \quad (2.13)$$

漸化式 ($1 \leq t$) :

$$S(t, 1) = d(t, 1) \quad (2.14)$$

$$S(t, 2) = \min \begin{bmatrix} S(t-2, 1) + \frac{1}{2}d(t-1, 2) + \frac{1}{2}d(t, 2) \\ S(t-1, 1) + d(t, 2) \\ S(t, 1) + d(t, 2) \end{bmatrix} \quad (2.15)$$

$$S(t, \tau) = \min \begin{bmatrix} S(t-2, \tau-1) + \frac{1}{2}d(t-1, \tau) + \frac{1}{2}d(t, \tau) \\ S(t-1, \tau-1) + d(t, \tau) \\ S(t-1, \tau-2) + d(t, \tau-1) + d(t, \tau) \end{bmatrix} \quad (3 \leq \tau \leq T) \quad (2.16)$$

連続 DP の出力は標準パターン長 T を用いて次式のように正規化した値 S' とする .

$$S' = \frac{1}{T} S(t, T) \quad (2.17)$$

2.3.3 ShiftCDP 法

連続 DP は事前に設定した標準パターンを用いて , 入力パターンの中から類似区間を検出する手法であったが , 標準パターン内の任意区間のマッチングも同時にかつ高速に行える手法として ShiftCDP 法がある [18] .

ShiftCDP 法は , 標準パターン上を先頭から N_{Shift} フレームずつシフトしながら , 固定フレーム数 N_{CDP} からなる単位標準パターンを構成する . そして , $N_{\text{Pat}} ([T/N_{\text{Shift}}] + 1$: 括弧はその値を超えない最大の整数値) 個の各単位標準パターンに対し , 連続 DP を実施した結果を連結・統合するものである . ShiftCDP 法 の概念図を図 2.9 に示す .

フレーム τ における現時刻のローカル距離を $D_t(\tau)$ とすると , i 番目の単位標準パターンの j 番目のフレームについて , 現時刻の累積距離 $G_t(i, j)$ は , 2 時刻前の累積距離 $G_{t-2}(i, j)$ を用いて , 以下の漸化式で計算できる . ここで , $\tau_S(i)$, $\tau_E(i)$ は i 番目の単位標準パターンについて , 全体の標準パターンにおける始端位置 , 始端位置であり , 事前に計算しておく .

初期条件 :

$$G_t(i, j) = G_{t-1}(i, j) = G_{t-2}(i, j) = \infty \quad (1 \leq i \leq N_{\text{Pat}}, 1 \leq j \leq N_{\text{CDP}}) \quad (2.18)$$

$$D_{t-1}(\tau) = \infty \quad (1 \leq \tau \leq T) \quad (2.19)$$

LOOP ($1 \leq i \leq N_{\text{Pat}}$) : 単位標準パターンについて

LOOP ($1 \leq j \leq N_{\text{CDP}}$) : 各フレームについて

$j = 1$ のとき :

$$G_t(i, 1) = D_t(\tau_S(i)) \quad (2.20)$$

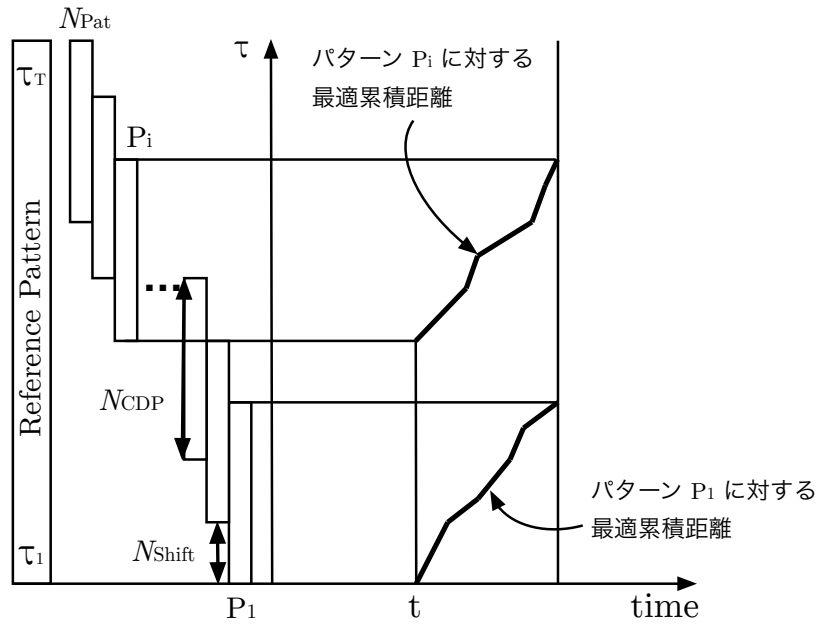


図 2.9: ShiftCDP 法 の 概 念

$j \geq 2$ のとき :

$$P(1) = G_{t-2}(i, j-1) + \frac{1}{2} \cdot D_{t-1}(\tau_S(i) + j-1) + \frac{1}{2} \cdot D_t(\tau_S(i) + j-1) \quad (2.21)$$

$$P(2) = G_{t-1}(i, j-1) + D_t(\tau_S(i) + j-1) \quad (2.22)$$

$$P(3) = G_{t-1}(i, j-2) + D_t(\tau_S(i) + j-2) + D_t(\tau_S(i) + j-1) \quad (2.23)$$

ただし, $j = 2$ のとき :

$$P(3) = D_t(\tau_S(i)) + D_t(\tau_S(i) + 1) \quad (2.24)$$

このうち, 次式より最適パスを決定する .

$$\alpha^* = \arg \min_{(\alpha=1,2,3)} P(\alpha) \quad (2.25)$$

次に, 選択したパスに従って累積距離を更新する .

$$G_t(i, j) = P(\alpha^*) \quad (2.26)$$

単位標準パターンは同一フレーム数 N_{CDP} からなるので、時刻 t におけるパターン i の整合度は最終フレームの累積距離 $G_t(i, N_{\text{CDP}})$ そのもので与えられ、正規化の必要はない。これに対し閾値を設けて、類似区間の検出処理を行った後、次式によって累積距離の更新を行う。

$$G_{t-2}(i, j) = G_{t-1}(i, j) \quad (1 \leq i \leq N_{\text{Pat}}, 1 \leq j \leq N_{\text{CDP}}) \quad (2.27)$$

$$G_{t-1}(i, j) = G_t(i, j) \quad (1 \leq i \leq N_{\text{Pat}}, 1 \leq j \leq N_{\text{CDP}}) \quad (2.28)$$

2.4 隠れマルコフモデル (HMM)

隠れマルコフモデル (Hidden Markov Model : HMM) は、時系列信号の確率モデルであり、複数の定常信号源の間を遷移することで、非定常な時系列モデルをモデル化する [19, 20]。

HMM の例を図 2.10 に示す。HMM は、時間単位ごとに時刻 t と状態 j を更新する有限状態機械であり、信号 o_t は出力確率 $b_j(o_t)$ をもとに生成される。さらに、状態 i から状態 j への状態遷移確率 a_{ij} を持つ。図 2.10 の例は、6 つの状態を持つモデルが、状態系列 $X = 1, 2, 2, 3, 4, 4, 5, 6$ のように動作して、信号 o_1 から o_6 を生成したことを示している。

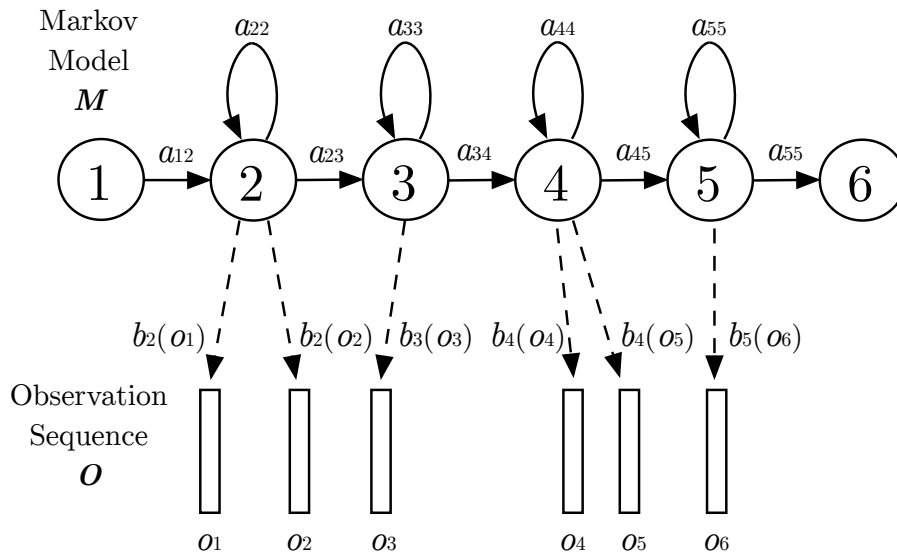


図 2.10: HMM の例

ここで、信号系列 O が、状態系列 X を経由して、モデル M から出力される条件付き確率は、状態遷移確率 a_{ij} と出力確率 $b_j(o_t)$ から次式で計算することができる。

$$P(O, X|M) = a_{12}b_2(o_1)a_{22}b_2(o_2)a_{23}b_3(o_3)\dots \quad (2.29)$$

しかし、実際には信号系列 O だけが分かっている、状態系列 X は観測することができない。これが「隠れ (Hidden)」の名が付けられている理由である。上式において、 X が分からない場合、全ての $X = x(1), x(2), x(3), \dots, x(T)$ の可能性を考え、足し合わせることで $P(O|M)$ を求めることができる。

$$P(O|M) = \sum_{\mathbf{X}} a_{x(0)x(1)} \prod_{t=1}^T b_{x(t)}(o_t) a_{x(t)x(t+1)} \quad (2.30)$$

ここで、 $x(0)$ は開始状態、 $x(T+1)$ は終了状態である。上式において、あらゆる状態系列 X の中で最尤となる確率は次式で与えられる。

$$\hat{P}(O|M) = \max_{\mathbf{X}} \left\{ a_{x(0)x(1)} \prod_{t=1}^T b_{x(t)}(o_t) a_{x(t)x(t+1)} \right\} \quad (2.31)$$

上式によって様々なモデルに対する尤度を求めることで、信号系列 O から最尤のモデル M を推定することができる。これが HMM を用いた認識の基本である。

尤度を計算するためには、状態遷移確率 a_{ij} や出力確率 $b_j(o_t)$ といった HMM のパラメータを求めておく必要がある。これらは学習データから自動的に推定することが可能であり、これが HMM の学習と呼ばれる。

HMM を用いた学習・認識の流れを図 2.11 に示す。まず、カテゴリが既知の学習用データをいくつか用意しておき、モデルを学習する。次に、未知の信号系列に関して、各モデルに対する尤度を計算し、最尤となるモデルを認識結果とする。

2.4.1 HMM のパラメータ学習

HMM のパラメータ学習は、状態遷移確率と出力確率の 2 つを与えられた学習データをもとに推定することである。この方法としては、状態系列が与えられた場合と、状態系列が与えられていない場合の二通りが考えられる。

状態系列が与えられた場合

学習データとして、信号系列 $O = o_1, o_2, \dots, o_t, \dots$ と状態系列 $X = S_1, S_2, \dots, S_i, \dots$ が組んで与えられた場合には、最尤パラメータの推定は以下の手順で比較的容易に行える。

(a) 状態遷移確率

ある状態 S_i に着目した場合、その状態から遷移が起こった回数の比を遷移先ごとに計算したものが、与えられた状態系列を最も高い尤度で生起させるパラメータとなる。すなわち、状態 S_i から状態 S_j に遷移が起こった回数を $F(S_i S_j)$ で表すと、状態遷移確率 a_{ij} は

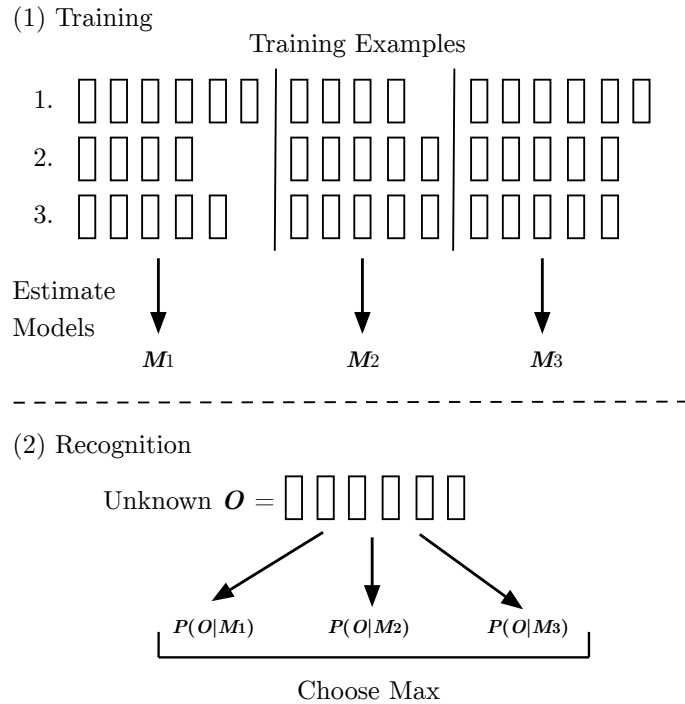


図 2.11: HMM の学習と認識

次式で求められる .

$$a_{ij} = \frac{F(S_i S_j)}{\sum_{m=0}^M F(S_i S_m)} \quad (2.32)$$

(b) 出力確率

状態 S_i が持つ出力確率関数 $b_i(o)$ として平均 μ_i , 分散 σ_i^2 の正規分布を仮定する .

$$b_i(o) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left\{ -\frac{(o - \mu_i)^2}{2\sigma_i^2} \right\} \quad (2.33)$$

S_i から出力された信号系列を集めて得られる集合 $\mathcal{O}^{(i)} = o_1^{(i)}, o_2^{(i)}, \dots, o_N^{(i)}$ に対する同時確率の対数值 L_i は , $\sigma^2 = \xi_i$ とおくと ,

$$L_i = \ln \left\{ \prod_{n=1}^N b_i(o_n^{(i)}) \right\} \quad (2.34)$$

$$= -\frac{1}{2} \sum_{n=1}^N \left[\ln(2\pi) + \ln(\xi_i) + \frac{(o_n^{(i)} - \mu_i)^2}{\xi_i} \right] \quad (2.35)$$

ここで, L_i の最大化を考え,

$$\frac{\partial L_i}{\partial \mu_i} = 0, \quad \frac{\partial L_i}{\partial \xi_i} = 0 \quad (2.36)$$

とおくことで, 以下のように最尤のパラメータが求められる.

$$\hat{\mu}_i = \frac{1}{N} \sum_{n=1}^N o_n \quad (2.37)$$

$$(\hat{\sigma}_i)^2 = \frac{1}{N} \sum_{n=1}^N (o_n - \hat{\mu}_i)^2 \quad (2.38)$$

状態系列が与えられていない場合 (Baum-Welch アルゴリズム)

状態系列が与えられていない場合は, 初期モデル M と観測された出力信号系列 O を手がかりに, すべての状態系列 X_k について, その出現確率 $P(X_k|O, M)$ を計算し, 期待値に基づいて最尤状態系列を決定する. その後, 状態系列が与えられている場合の方法により, 最尤パラメータを持つモデル \hat{M} の推定を行う. さらに, 得られたモデルを初期モデルとして学習を繰り返すことで, より高い確率で学習データを出力しうるモデルの学習を行う. このアルゴリズムは, Baum-Welch アルゴリズムと呼ばれる.

しかし, 実際にすべての状態系列に対して, $P(X_k|O, M)$ を求めるには膨大な計算量が必要とされる. そこで, 効率的に期待値を計算する Forward-Backward アルゴリズムを用いることが一般的である. このアルゴリズムは, モデルと出力信号が与えられたときに,

- 前向き確率 $\alpha(t, i)$

時刻 t において, 状態 i に至る確率

- 後ろ向き確率 $\beta(t, j)$

時刻 t に状態 j を出発して, 時刻 $T + 1$ で状態 N (終了状態) に到達する確率

の 2 つの確率をあらかじめ計算しておくことで, 計算量の削減を図るものである.

前向き確率 $\alpha(t, i)$ は, 次に示す漸化式で求めることができる.

(1) 初期化:

$$\alpha(0, 0) = 1.0 \quad (2.39)$$

$$\alpha(t, i) = 0.0 \quad (1 \leq i \leq N, 1 \leq t \leq T) \quad (2.40)$$

(2) 初期遷移:

$$\alpha(1, i) = a_{0i} \quad (1 \leq i \leq N - 1) \quad (2.41)$$

(3) 漸化式計算 : $t = 2, \dots, T$ について

$$\alpha(n, i) = \sum_{n=1}^{N-1} \alpha(t-1, n) \cdot a_{ni} b_n(o(t-1)) \quad (2.42)$$

(4) 最終遷移 :

$$P(\mathbf{O}|M) = \sum_{n=1}^{N-1} \alpha(T, n) a_{nN} b_n(o(T)) \quad (2.43)$$

一方, 後ろ向き確率 $\beta(t, j)$ は以下の漸化式で与えられる .

(1) 初期化 :

$$\beta(T+1, N) = 1.0 \quad (2.44)$$

(2) 漸化式計算 :

$$\beta(t, j) = \sum_{n=1}^N \{\beta(t+1, n) a_{jn} b_j(o(t))\} \quad (1 \leq i \leq N, 1 \leq t \leq T) \quad (2.45)$$

(3) 初期遷移 :

$$P(\mathbf{O}|M) = \beta(0, 0) = \sum_{n=1}^{N-1} \{\beta(1, n) a_{0n}\} \quad (2.46)$$

前向き確率, 後ろ向き確率の両者の積を求めると,

$$\alpha(t, n) \beta(t, n) = P(\mathbf{O}, x(t) = n | M) \quad (2.47)$$

すなわち, 時刻 t において状態 n に停留する状態系列から信号系列 \mathbf{O} が出力される確率が計算できる . さらに, ベイズの定理から, 時刻 t において, 状態 n に停留する各状態系列の出現確率は次式で計算できる .

$$P(x(t) = n | \mathbf{O}, M) = \frac{P(\mathbf{O}, x(t) = n | M)}{P(\mathbf{O}|M)} = \frac{\alpha(t, n) \beta(t, n)}{\alpha(T+1, N)} \quad (2.48)$$

この状態の出現確率を元に, 新たに最尤パラメータ推定を行うことで得られる新しいモデル \hat{M} は, 初期モデル M に対して常に,

$$P(\mathbf{O}|M) \leq P(\mathbf{O}|\hat{M}) \quad (2.49)$$

すなわち，出力信号系列をより高い確率値で出力することが証明されている．したがって，繰返し計算によって HMM の学習を行うことができる．

2.4.2 HMM による認識

HMM による認識は，与えられた信号系列 O を最も高い確率で生成する状態系列を求める問題であると考えることができる．この問題の解は，直接的には，

$$\hat{X} = \arg \max_X \left[a_{0x(1)} \left\{ \prod_{t=1}^{T-1} b_{x(t)}(o_t) a_{x(t)x(t+1)} \right\} b_{x(T)}(o_T) a_{x(T)N} \right] \quad (2.50)$$

によって与えられる．ここで，時刻 t で状態 i に至る状態系列の中で，最も高い確率を与える状態系列について，遷移元の状態を記憶するためのバックポインタ $B(t, i)$ を導入すると，上式は次の漸化式で求めることができる．

(1) 初期化：

$$\alpha(0, 0) = 1.0 \quad (2.51)$$

(2) 初期遷移： $1 \leq i \leq N - 1$ について

$$B(1, i) = 0 \quad (2.52)$$

$$\alpha(1, i) = a_{0i} \quad (2.53)$$

(3) 漸化式計算： $t = 2, \dots, T$ について

$$\alpha(n, i) = \max_n \{ \alpha(t-1, n) \cdot a_{ni} b_m(o_{t-1}) \} \quad (2.54)$$

$$B(n, i) = \arg \max_n \{ \alpha(t-1, n) \cdot a_{ni} b_m(o_{t-1}) \} \quad (2.55)$$

(4) 最終遷移：

$$B(T+1, N) = \arg \max_m \{ \alpha(T, m) \cdot a_{mN} b_n(o_T) \} \quad (2.56)$$

$$\max_X \{ P(O, X|M) \} = \alpha(T, B(T+1, N)) \cdot a_{B(T+1, N)N} b_{B(T+1, N)}(o_T) \quad (2.57)$$

また，バックポインタを辿ることで，最大の確率を与える状態系列 $\hat{X} = \hat{x}(1), \dots, \hat{x}(T)$ は以下により求められる．

$$\hat{x}(T) = B(T+1, N) \quad (2.58)$$

$$\hat{x}(t-1) = B(t, \hat{x}(t)) \quad (t = T, \dots, 2) \quad (2.59)$$

ここで、求められた系列 \hat{X} は Viterbi 系列と呼ばれ、最大確率を与える状態系列に沿って確率を求める上述のアルゴリズムは Viterbi アルゴリズムと呼ばれる。Viterbi アルゴリズムの概念図を図 2.12 に示す。図は、入力フレームを横軸に、通過する HMM の状態が縦軸となっており、最尤となる経路を探る様子を示している。この場合、各格子点が入力に対する出力確率を、格子点同士を結ぶ線が状態遷移確率を意味する。いかなる経路の場合も、尤度は遷移確率と出力確率の積の総和で計算することができるが、時刻 t において、各経路 $\psi_i(t-1)$ は全ての状態 i に対して計算されているので、式 2.54–2.55 のように現在の入力フレームに依存する部分だけを計算すればよい。

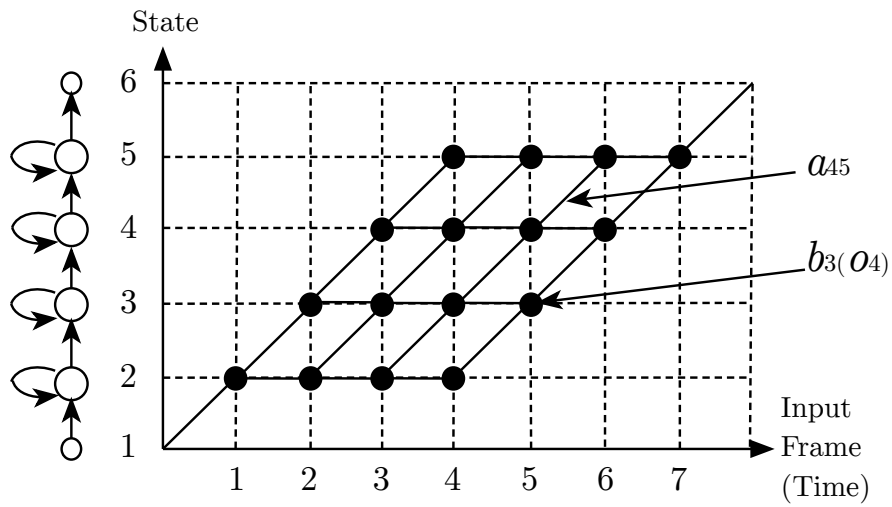


図 2.12: Viterbi アルゴリズムの概念

第3章 先行研究

本章では，提案手法の基礎となる先行研究について述べる．

3.1 HMM に基づく自動和音認識

HMM を用いて，音楽音響信号から和音認識を行う手法がこれまでに多数提案されている．

Sheh ら [21] は，ポピュラー音楽の音楽音響信号から抽出した特徴量を学習データとして，和音ごとに HMM を作成し，和音認識を行った．特徴量には，音声認識の HMM 学習によく用いられているメル周波数ケプストラム (Mel Frequency Cepstral Coefficient : MFCC)，クロマベクトルのいずれかを用いて比較を行っており，クロマベクトルを用いた場合の方が高い認識率を示している．この研究では，音楽音響信号に対する和音進行の時間的な対応付けを行わずに学習を行ったため，モデルのパラメータ推定が不十分であり，認識率は低かった．

内山ら [22] は，Sheh らと同様に，音響特徴量としてクロマベクトルを用いて HMM の学習を行った．その際，各時刻でどの和音であるかという情報を含んだ人手による和音ラベルを用いて，状態系列が既知の場合のパラメータ推定を行った．HMM は和音を一つの状態と考え，どの状態にも遷移可能な ergodic HMM とした．和音間の遷移確率は一つ前の和音からの 2-gram 確率として，テキスト情報から学習した値を与えた．結果として，90 曲のテストデータに対して約 75 % の高い認識率を得ている．

その他にも，前後の和音の影響を考慮した triphone モデルを用いたもの [23]，調ごとに和音間の遷移しやすさが異なることを考慮して調ごとに HMM を作成するもの [24] など，様々な発展的手法が提案されているが，本研究では学習が比較的容易でかつ高性能なモデル作成が可能な，内山らの先行研究 [22] の手法を利用する．ここでは，その手法について紹介する．

3.1.1 HMM に基づく和音進行のモデル化

和音認識問題は，特徴量系列 $O = o_1, o_2, \dots, o_T$ が観測されたとき，その背後にある和音系列 $C = c_1, c_2, \dots, c_T$ を求める問題である．最尤の和音系列 \hat{C} は，事後確率最大化の観点から，次式で求めることができる．

$$\hat{C} = \arg \max_C P(C|O) \quad (3.1)$$

上式はベイズの定理より，次式のように置き換えられる．

$$\hat{C} = \arg \max_C P(\mathbf{O}|C)P(C) \quad (3.2)$$

ここで，観測特徴量の生成源として HMM を考え，モデル化を行う．HMM の隠れ状態は和音とし，各時刻で状態から特徴量が出力され，状態が遷移していくとする．和音進行には，和声学理論にあるように規則が存在すると考えられるため，現在の和音を推定する上で前までの和音を考慮することが必要である．そこで，ある時刻の和音は一つ前の時刻の和音に依存すると仮定し，和音間の遷移確率を $P(c_t|c_{t-1})$ と表す．また，観測された特徴量と各和音の特徴量の近さを，出力確率 $P(o_t|c_t)$ として表現する．

以上のモデル化により，式 3.2 は次式のように近似することができる．

$$\hat{C} \simeq \arg \max_C \left\{ P(o_1|c_1)P(c_1) \prod_{t=2}^T P(o_t|c_t)P(c_t) \right\} \quad (3.3)$$

この最尤系列は 2.4.2 節で述べた Viterbi アルゴリズムで効率的に求めることができる．

3.1.2 和音 HMM の学習と認識

前述のモデル化に基づいて和音 HMM の学習を行うブロック図を図 3.1 に示す．複数の音楽音響信号から抽出したクロマベクトルと，各信号について人手で作成した和音進行ラベルを学習データとして，状態系列が既知の条件で和音 HMM のパラメータ推定を行う．また，和音進行ラベルから 2-gram 確率を計算し，和音 HMM の状態遷移確率とする．

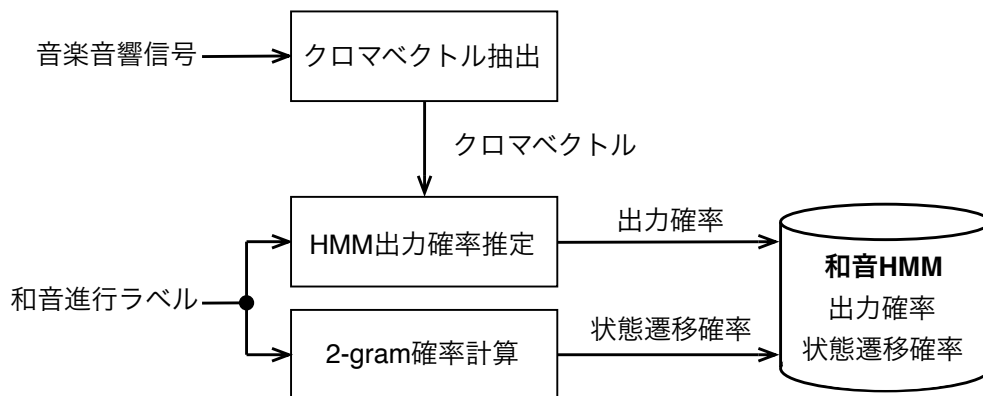


図 3.1: 和音 HMM の学習

また，和音 HMM に基づいた自動和音認識のブロック図を図 3.2 に示す．未知の音楽音響信号から抽出したクロマベクトルをテストデータとして，あらかじめ学習しておいた和音 HMM

を用いて Viterbi アルゴリズムによる最尤推定を行い，最尤和音系列が決定される．

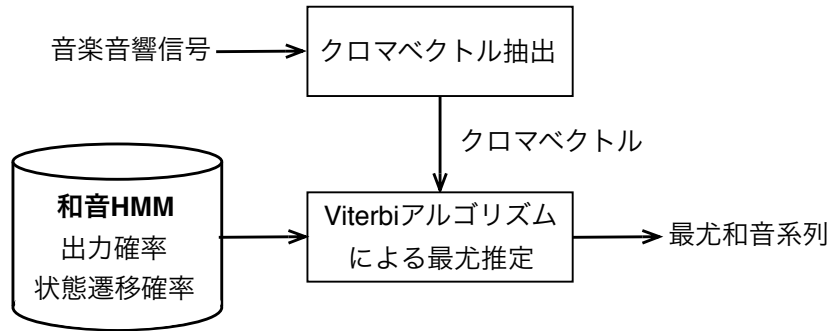


図 3.2: 和音 HMM に基づく自動和音認識

3.2 カバー演奏ストリームからの原曲同定システム

ここでは，序論でも紹介した深澤らの研究 [8] をカバー演奏ストリームからの原曲同定システムの従来手法として取り上げ，その全体構成と問題点について述べる．

3.2.1 従来手法のシステム構成

従来手法のシステム構成を図 3.3 に示す．従来手法は複数の原曲の音楽音響信号から標準パターン DB を作成する学習ブロックと，カバー演奏の入力パターンと DB 内の標準パターンとの連続 DP の値を計算し，閾値処理を行う原曲同定ブロックの二つのブロックから構成される．

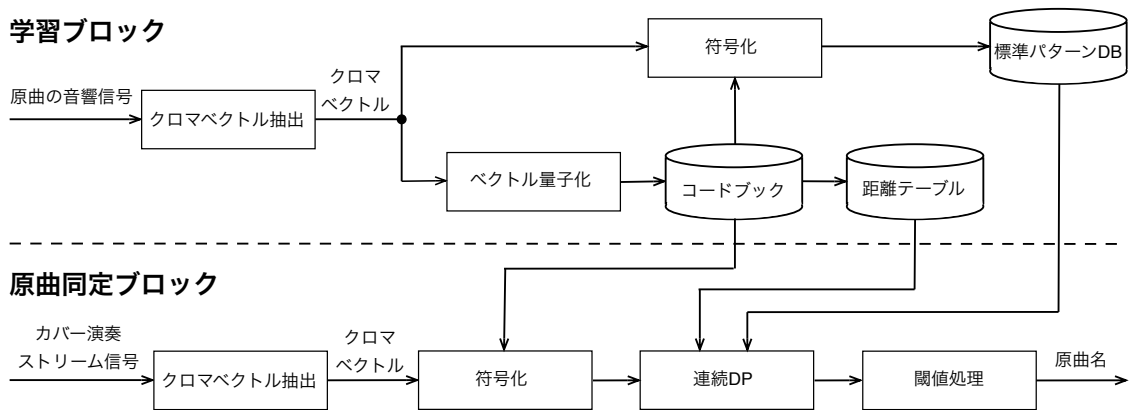


図 3.3: 従来手法のシステム構成

図 3.3 の各ブロックの動作を以下に示す。

学習ブロック

1. 複数の原曲の音響信号について、カバー演奏と類似しやすいと考えられる区間（A メロ，B メロ，サビ等）を人手で決定し、分割する。
2. 分割した音響信号から、それぞれクロマベクトルを抽出する。
3. 得られた全てのクロマベクトルを学習ベクトルとして、LBG アルゴリズムによるベクトル量子化を行い、コードブックを作成する。
4. 原曲同定ブロックの高速化のため、コードブックの代表ベクトル間のユークリッド距離を計算した距離テーブルを作成しておく。
5. コードブックを用いて、抽出したクロマベクトル時系列を符号化し、各楽曲の区間ごとの符号パターンを標準パターンとして、DB に格納する。このとき、クロマベクトルを次元間でシフトさせることで移調が表現可能な特性を利用して、12 種類の基準音に移調したパターンを全て DB に格納しておく。

原曲同定ブロック

1. カバー演奏の入力音響信号からクロマベクトルを抽出する。
2. 学習ブロックで作成しておいたコードブックを用いてクロマベクトルを符号化する。
3. 連続 DP に符号を入力し、各標準パターンの最適累積距離を更新する。計算量削減のため、学習ブロックで作成しておいた距離テーブルから事前計算しておいた距離を呼び出して用いる。
4. 3. で求めた連続 DP の出力に対して適当な閾値を設定しておき、閾値を下回ったパターンを演奏中の原曲として同定し、原曲名を出力する。

3.2.2 問題点

従来手法には主に三つの問題点が考えられる。その問題点を以下に列挙する。

(1) 原曲同定精度の問題

抽出したクロマベクトルをベクトル量子化により符号化したパターンをマッチングに用いている。特徴量としてクロマベクトルを用いているのは、原曲とカバー演奏の間で和音進行が保存されている可能性が高いという仮定に基づいたものである。しかし、本来同一和音が演奏されている区間でも編曲によっては和声内音の省略や非和声音の挿入などがあり、

抽出されるクロマベクトルは一様ではなく、ある程度のばらつきが出てしまう。また、分析フレームごとにクロマベクトルの符号化では、符号化の際に分析フレームの前後の情報が考慮されないため、突発的な特徴量の揺れがそのままパターンに反映されてしまうという問題がある。これらの問題によって、原曲とカバー演奏の音響信号から得られるパターン間に差が出やすい。これがマッチング結果に影響し、原曲同定精度が低くなってしまうと考えられる。

(2) 計算量の問題

DB 内の全標準パターンについて入力フレームごとに連続 DP の値を更新し、閾値処理による楽曲同定を行っている。この方法は、データベースの登録曲数が比較的少ない場合は実時間で対応可能であるが、曲数の拡張に伴って比例的に計算量が増加してしまうという問題がある。

(3) 人手によるパターン分割の問題

原曲の音響信号から標準パターン DB を作成する際に、楽曲全体を一つの標準パターンとすると、カバー演奏側で楽曲全体が演奏されるまでは原曲同定が行えない。そのため、従来手法では、一曲よりも短い時間単位で原曲を同定するため、あらかじめカバー演奏と類似しやすいような区間を人手で設定し、信号の分割を行ってから標準パターンを作成していた。この方法では、DB を拡張する際の制約が大きいという問題がある。

第4章 提案手法

本章では、第3章で述べたカバー演奏ストリームからの原曲同定システムの従来手法の問題点を踏まえて、本研究の提案手法について述べる。

4.1 システム構成

提案手法のシステム構成を図4.1に示す。提案手法は、従来手法と同様に学習ブロックと原曲同定ブロックに大別される。

学習ブロック

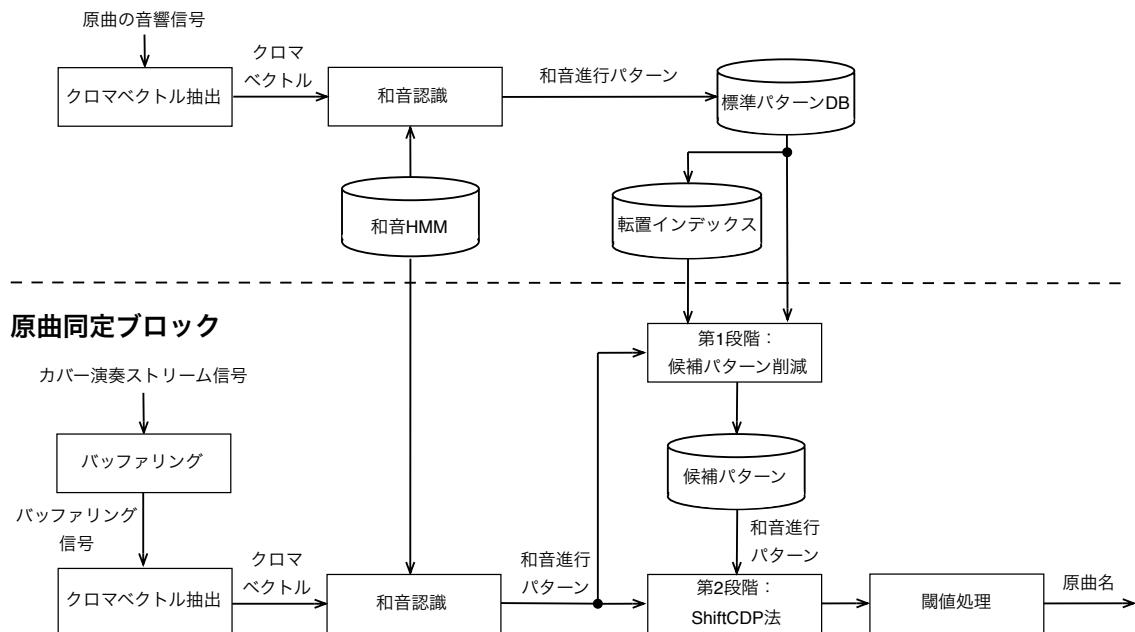


図 4.1: 提案手法のシステム構成

図4.1の各ブロックの動作を以下に示す。

学習ブロック

1. 複数の原曲の音響信号から，特徴量としてクロマベクトルを抽出する．これは従来手法と共通の処理である．
2. 抽出したクロマベクトルに対し，あらかじめ多数の楽曲から学習しておいた和音 HMM を用いて自動和音認識を行う．そして，認識結果として得られた和音進行パターンを標準パターン DB に格納する．このとき，12 種類の基準音に移調したパターンをすべて DB に格納しておく．
3. DB 内の各パターンに対する N -gram の転置インデックスを作成する．

原曲同定ブロック

1. 入力となるカバー演奏の音響信号を一定時間バッファリングする．
2. バッファリングした信号からクロマベクトルを抽出する．
3. 抽出したクロマベクトルに対して，学習ブロックで用いたものと同様の和音 HMM を用いて自動和音認識を行う．
4. 第 1 段階のマッチング（転置インデックスによる候補パターン削減）：入力の和音進行パターンが決まったら， N -gram でパターンを分割し，転置インデックスを参照する．そして，入力パターンに含まれる和音進行パターンを持つ標準パターンのみを候補パターンとして残す．
5. 第 2 段階のマッチング（ShiftCDP 法による音響的類似度スコアリング）：絞り込まれた候補パターンに対して，ShiftCDP 法によって音響的な類似度を用いたスコアリングを行う．スコアに対して閾値を設定し，閾値を下回ったパターンをスコアが低い順にソートし，上位のパターンを原曲同定結果として出力する．

4.2 要素技術

図 4.1 に示した本研究の提案手法は，主に三つの要素技術から構成される．

4.2.1 和音進行パターンへの符号化

従来手法では，楽曲から抽出したクロマベクトルの時系列パターンをベクトル量子化によって符号化し，その符号パターンをマッチングに用いていた．これに対し，本研究の提案手法では，楽曲から抽出したクロマベクトルから 3.1 節に示した HMM に基づく自動和音認識手法を用いて楽曲の和音進行パターンに符号化を行い，マッチングに用いる．

従来手法で特徴量としてクロマベクトルを用いていたのは、原曲とカバー演奏の間で和音進行が保存されている可能性が高いという仮定に基づいたものである。しかし、本来同一の和音が演奏されている区間でも抽出されるクロマベクトルは一樣ではなく、ある程度のばらつきが出てしまう。また、分析フレームごとの符号化では、符号化の際に分析フレームの前後の情報が考慮されないため、突発的な特徴量の揺れがそのままパターンに反映されてしまうといった問題があった。HMM に基づく自動和音認識手法では、クロマベクトルの和音ごとの分布を HMM の一つの状態が持つ混合正規分布として表現できる。また、音響的な特徴だけでなく、テキストデータの和音進行ラベルから学習した和音間の遷移しやすさの情報を含めることができる。そのため、原曲とカバー演奏の特徴量の間にも多少の差があっても、本来同一の和音が演奏されている区間であれば、前後のフレームの情報を含めた最尤推定により、ほぼ同一のパターンへの符号化が期待できる。本研究ではこの手法により、原曲同定精度の向上を図る。

HMM を用いた自動和音認識を行うためには、ある程度の長さの入力フレームから特徴量を得る必要がある。そのため、提案手法では図 4.2 に示すように一定時間の入力バッファを設け、入力信号をバッファリングする。そして、バッファリングした区間に対して自動和音認識を行い、原曲同定処理を行う。区間によっては、フレーズ (A メロ, B メロ, サビ等) や楽曲が転換することが考えられるので、バッファは一定区間オーバーラップさせながら処理を行うようにする。

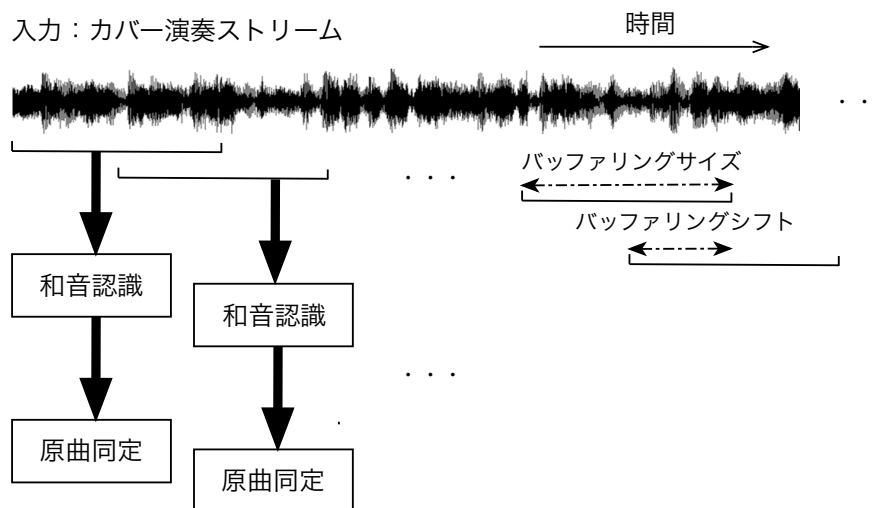


図 4.2: バッファリングした入力信号に対する自動和音認識

4.2.2 転置インデックスによる候補パターン削減

学習ブロックにおいて、複数の原曲の音響信号から和音進行パターンを抽出し、標準パターン DB を作成した後、標準パターン DB に対する転置インデックスを作成する。これは、和音進行パターンを N -gram に分割し、 N 個の和音からなるパターンがどの標準パターンに登場す

るかをあらかじめ索引付けしたものである。

4-gram の場合の作成例を図 4.3 に示す．この例は，標準パターン番号 (Pattern No.) が 1 のパターンについて，{C-G-Am-F, G-Am-F-G, Am-F-C-G, ...} のように 4-gram でパターンを分割し，他のパターンも同様に分割した結果，図中の右の表に示すような転置インデックスが作成されたことを示している．

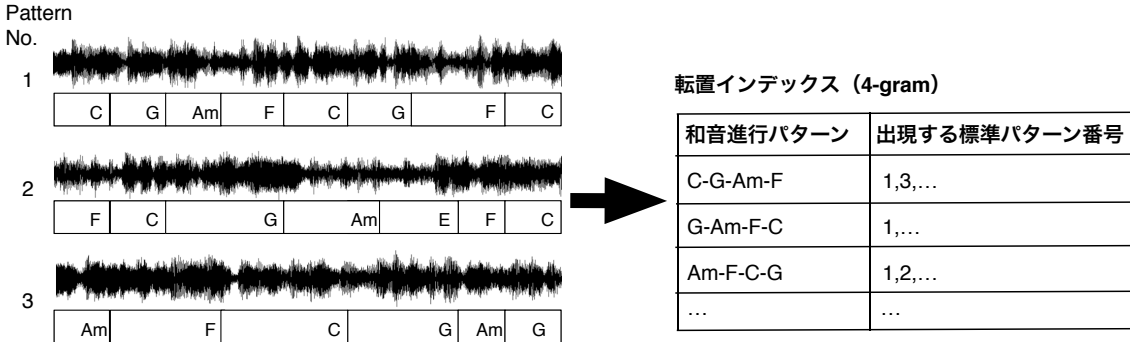


図 4.3: 和音進行パターンの転置インデックスの作成 (4-gram の例)

原曲同定ブロックで入力信号から和音進行パターンが得られたとき第 1 段階のマッチングとして，作成した転置インデックスを用いた候補パターンの絞込みを行う．その手順としては，まず，標準パターンと同様に，入力パターンを N -gram で分割する．次に，分割したパターンそれぞれを検索キーとして転置インデックスを参照し，得られた入力パターンと一致するパターンを持つ標準パターンだけに候補を絞り込む．これと類似した処理は音声中の単語検索 [25] でも成果を上げており，テキストベースの処理となるため，比較的高速に音響的類似度によるマッチングを行う候補パターン数を削減することができる．

4.2.3 ShiftCDP 法による音響的類似度スコアリング

転置インデックスによる候補パターン数削減を行った後，最終的に原曲同定を行うための第 2 段階のマッチングとして，2.3.3 節で述べた ShiftCDP 法を用いた音響的類似度によるスコアリングを行う．本研究では，ShiftCDP 法の計算で用いる音響類似度として，各和音の混合正規分布間のユークリッド距離の平均値，パタチャリア距離の平均値のいずれかを使用し，比較・検討を行う．

ここで，和音 p, q 間のユークリッド距離の平均値 $d_{Euc}(p, q)$ は，HMM の混合正規分布 (混合数 M) の平均値を μ_p, μ_q ，共分散行列を Σ_p, Σ_q とすると，次式で求められる．

$$D_{Euc}(\mu_p, \mu_q) = \sqrt{(\mu_p - \mu_q)^2} \quad (4.1)$$

$$d_{Euc}(p, q) = \frac{1}{M} \sum_{i,j=1}^M D_{Euc}(\mu_p^i, \mu_q^j) \quad (4.2)$$

また，バタチャリヤ距離の平均値 $d_{Bhat}(p, q)$ は， $\Sigma = (\Sigma_p + \Sigma_q)/2$ としたとき，次式で求められる．

$$D_{Bhat}(\mu_p, \Sigma_p, \mu_q, \Sigma_q) = \frac{1}{8}(\mu_p - \mu_q)^T \Sigma^{-1}(\mu_p - \mu_q) + \frac{1}{2} \ln \frac{\det \Sigma}{\sqrt{\det \Sigma_p \det \Sigma_q}} \quad (4.3)$$

$$d_{Bhat}(p, q) = \frac{1}{M} \sum_{i,j=1}^M D_{Bhat}(\mu_p^i, \Sigma_p^i, \mu_q^j, \Sigma_q^j) \quad (4.4)$$

ShiftCDP 法では，標準パターンを固定長に分割した単位標準パターンごとに入力との類似度を示す累積距離が出力されるが，本研究では入力区間に対して得られた累積距離の最小値をその標準パターンとのマッチングスコアとする．

4.3 従来手法との相違点

これまでに述べた従来手法と提案手法について，相違点をまとめたものを表 4.1 に示す．音楽音響信号から抽出する特徴量は共通でクロマベクトルを用いるが，提案手法では符号化手法が HMM に基づく自動和音認識に，マッチング手法がテキストベースの転置インデックスと，音響類似度を用いた ShiftCDP 法によるマッチングの二段階に分けた処理に置き換わる．ShiftCDP 法の特性から，人手によるパターン分割も不要となる．また，各手法の特性から原曲同定の閾値処理を行うタイミングが異なる．従来手法の場合は入力フレームごとに連続 DP の出力を更新し，閾値処理を行う．提案手法の場合は図 4.2 に示したように，バッファリングを行った区間ごとにマッチングスコアを計算し，閾値処理を行う．

表 4.1: 従来手法と提案手法の相違点

	従来手法	提案手法
特徴量	クロマベクトル (12 次元)	
符号化手法	ベクトル量子化	HMM に基づく自動和音認識
マッチング手法	連続 DP (音響的類似度)	転置インデックス (テキスト) + ShiftCDP 法 (音響的類似度)
音響類似度	代表ベクトル間の距離	HMM の分布間の距離
人手のパターン分割	必要	不要
原曲同定タイミング	入力フレームごと	バッファリング区間ごと

第5章 評価実験

5.1 実験概要

提案手法の有効性を検証するため、次の二種類の評価実験を行った。

予備実験：HMM に基づく自動和音認識

和音進行パターンが既知のテストデータに対して、HMM に基づく自動和音認識を行い、和音 HMM の性能を評価する実験を行った。この実験では、3.1 節で示した和音 HMM のパラメータ学習方法で、どの程度正しく自動和音認識が可能か確かめることを目的とした。

本実験：カバー演奏ストリームからの原曲同定

従来手法、提案手法のいずれかを実装したシステムにカバー演奏ストリームを入力し、システムの性能を評価する実験を行った。

5.2 予備実験：HMM に基づく自動和音認識

5.2.1 実験方法

予備実験には The Beatles の 12 枚のアルバムに含まれる 180 曲を用いた。その理由は、どの時刻でどの和音かを示した正解ラベル [26] が提供されているため HMM の学習と和音認識結果に対する評価が容易であり、また、先行研究 [22] と同一データであることから性能比較できるためである。用いた楽曲の全容は付録 A の表 A.1–A.4 に示す。

予備実験は以下の手順で行った。

1. 180 曲のデータセットを、曲番号 1–90 までのデータセット A、曲番号 91–180 までのデータセット B の 90 曲ずつに分ける。
2. データセット A とその和音進行ラベルを用いて、和音 HMM の学習を行う。
3. 2. で学習した和音 HMM を用いて、データセット B の各曲について和音認識を行い、和音認識率を求める。和音認識率 $Corr.$ は、認識した全和音数を N 、置換誤り数を S 、削除誤り数を D としたとき、次式で計算される。

$$Corr. = \frac{N - S - D}{N} \times 100 \quad [\%] \quad (5.1)$$

4. 学習データセットとテストデータセットを入れ替えて，再度同様の実験を行う．
5. 2回の試行で得られた和音認識率の平均値を求め，評価する．

5.2.2 実験条件

音響信号の分析条件，HMMのパラメータ等の実験条件を表 5.1 示す．

表 5.1: 予備実験の実験条件

チャンネル数	1 (モノラル)
標本化周波数 [Hz]	11025
量子化ビット数 [bit]	16
分析フレームサイズ [点]	8192
フレームシフトサイズ [点]	1024
クロマベクトルの次元数	12
和音 HMM の状態数	25 (major, minor, 和音なし)
和音 HMM の混合数	16

なお，和音 HMM の学習，および認識には HTK [27] を用いた．

5.2.3 実験結果

予備実験の結果を表 5.2 に示す．これを見ると，データを入れ替えた場合でも同程度の和音認識率が得られており，平均は 74.67 [%] となった．この結果は先行研究 [22] と比較した場合，ほぼ同等の性能が得られている．

表 5.2: 予備実験の結果：和音認識率

	学習データセット A	学習データセット B	平均
	テストデータセット B	テストデータセット A	
和音認識率 $Corr.$ [%]	75.92	73.43	74.67

曲ごとの和音認識率のヒストグラムを図 5.1 に示す．図は上段が 1 回目の試行，下段が 2 回目の試行の結果を示している．ほとんどの楽曲が平均値前後に分布しているが，認識率が 10.00 [%] 未満の認識率が低い楽曲が存在していることが分かる．和音認識率が 10.00 [%] 未満となった楽

曲の一覧は表 5.3 に示す．これらは，学習データとして不適切である可能性が高いため，本実験ではこれらの楽曲を除外した 178 曲を和音 HMM の学習データとする．

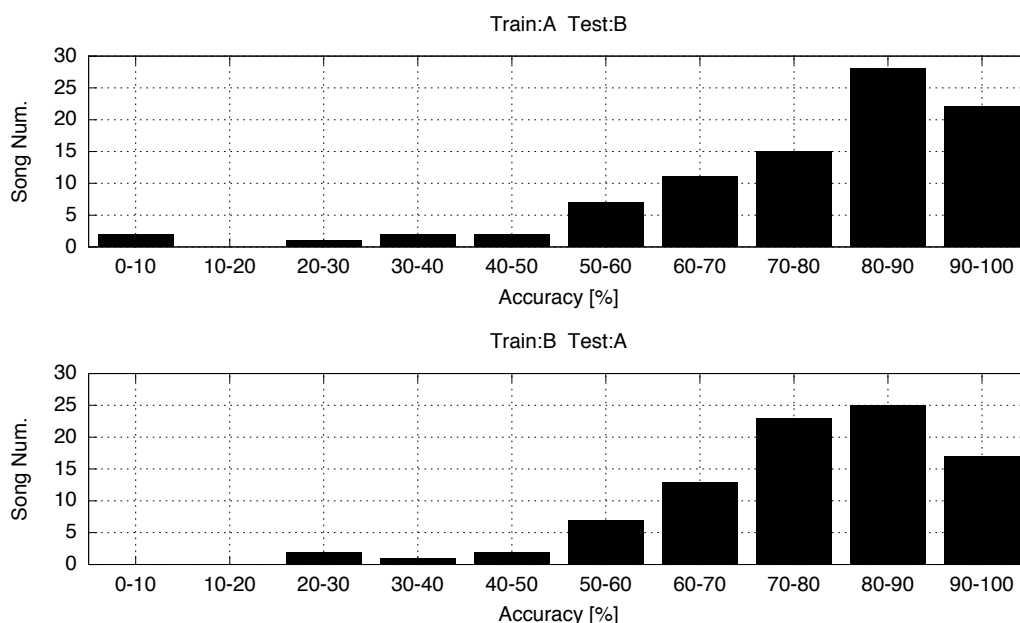


図 5.1: 予備実験の結果：和音認識率の分布

表 5.3: 予備実験の結果：和音認識率が 10.00 [%] 未満の楽曲

曲番号	楽曲名	和音認識率 [%]
100	Lovely Rita	5.68
141	Strawberry Fields Forever	6.38

5.3 本実験：カバー演奏ストリームからの原曲同定

5.3.1 実験方法

本実験には，原曲-カバー演奏のデータセットである「The covers80 cover song data set」[28] を用いた．これは洋楽のポップス 80 曲について，原曲とカバー演奏（異なるアーティストが演奏したもの）が対になったデータセットである．データセットの全容は付録 B の表 B.1–B.3 に

示す．曲番号 1-40 は先行研究 [8] の評価実験で用いられているものと同じの 40 曲で，今回の実験にはこの 40 曲を用いた．

本実験は以下の手順で行った．

1. 40 曲の原曲の音響信号から，標準パターン DB を作成する．
このとき，従来手法の場合は 3.2 節で述べたように，音響信号をあらかじめ人手で分割しておく必要がある．そのため，以下の分割ルールに従っていくつかのフレーズに分割した．
 - (a) ボーカルを含む箇所を用いる
 - (b) フレーズが繰り返し現れる場合は，1 回目のフレーズを用いる
 - (c) 繰り返しであっても，伴奏が大きく異なっている場合は両方を用いる
2. 40 曲のカバー演奏を全て連結した音響信号を作成し，システムに入力するカバー演奏ストリームとする．
3. 2. で作成したカバー演奏ストリームについて，どの時刻でどの楽曲が演奏されているかを記した正解ラベルを人手で作成する．
4. カバー演奏ストリームをシステムに入力し，システムが同定した曲名と時刻を記録する．
5. 3. で得た出力とあらかじめ作成しておいた正解ラベルを比較し，適合率 (Precision)，再現率 (Recall)，F 値を計算する．これら三つの尺度はそれぞれ次式で与えられる．

$$\text{Precision} = \frac{\text{Correct}}{\text{Retrieved}} \times 100 [\%] \quad (5.2)$$

$$\text{Recall} = \frac{\text{Correct}}{\text{Relevant}} \times 100 [\%] \quad (5.3)$$

$$\text{F 値} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5.4)$$

上式における Correct は検出したパターンについて正解した数を示す．Retrieved は，検出した合計パターン数を示す．また，Relevant は本来検出すべき楽曲数 (2. で用意しておいた正解ラベルの楽曲数) を示す．従来手法，提案手法でいずれの場合も，最終的に閾値処理を行って原曲同定を行うので，閾値を甘くするほど再現率は高くなるが，間違った楽曲を同定しやすくなるので適合率は低くなる．逆に，閾値を厳しくするほど再現率は落ちるが，よりマッチングスコアの良い楽曲だけを検出するので適合率は高くなる．F 値は適合率と再現率の調和平均であり，適合率および再現率がどちらも平均的に高い場合に最大の F 値が得られる．

6. 評価 1 提案手法のパラメータ決定：提案手法の第 2 段階で用いる音響類似度として 4.2.3 節で述べたユークリッド距離，バタチャリヤ距離のどちらか最適かを調べるため，それぞれを用いた場合で式 5.2-5.4 を計算し，評価する．このとき，第 1 段階の転置インデックスに用いる N -gram の N を変化させながら実験を行い，最適な N も同時に決定する．

7. 評価 2 従来手法と提案手法の原曲同定精度比較：原曲同定時の閾値を変更しながら，3. と 4. を繰り返し行い，適合率を縦軸，再現率を横軸とした P-R 曲線を作成する．最も高い F 値が得られる閾値でシステムの性能を評価する．
8. 評価 3 提案手法の計算量削減効果の評価：提案手法の第 1 段階の N -gram の N を変化させながら，第 2 段階で処理する候補パターン数を調べる．また，第 1 段階，第 2 段階のマッチングを 10 回試行した場合の平均計算時間を N ごとに計算し， N の変化に伴う計算量の削減効果の評価する．

5.3.2 実験条件

本実験で用いた音響信号の分析条件，HMM の学習条件等の実験条件を表 5.4 に示す．

表 5.4: 本実験の実験条件

チャンネル数	1 (モノラル)
標準化周波数 [Hz]	11025
量子化ビット数 [bit]	16
分析フレームサイズ [点]	8192
フレームシフトサイズ [点]	1024
クロマベクトルの次元数	12
従来手法	
コードブックサイズ	1024
原曲信号の分割数	113
分割信号の平均時間長 [秒]	23.4
提案手法	
入力バッファサイズ [秒]	100
バッファシフトサイズ [秒]	50
ShiftCDP 法のシフトサイズ N_{Shift}	1
ShiftCDP 法の単位標準パターン長 N_{CDP}	25
和音 HMM の状態数	25 (major, minor, 和音なし)
和音 HMM の混合数	16

提案手法で用いる和音 HMM は予備実験の結果に基づき，付録 B の表 A.1–A.4 に示す The Beatles の 180 曲の楽曲から，表 5.3 に示した 2 曲を除いた 178 曲の音響信号と和音進行ラベルを用いてあらかじめ学習を行った．すなわち，和音 HMM の学習に用いたのは評価実験で用いたデータセットとは別のデータセットである．HMM の学習，および認識には予備実験と同様に HTK [27] を用いた．

従来手法の評価に用いた正解ラベルは付録 C の表 C.1–表 C.5，提案手法の評価に用いた正解ラベルは付録 D の表 D.1–表 D.3 に示す．従来手法は分割したフレーズの終端の時刻に正解ラベルを設け，正解ラベルの前後 5 秒で正解の楽曲を検出した場合に式 5.2–5.4 の Correct をカウントした．提案手法の場合は，バッファごとに正解ラベルを設け，そのバッファにおいて 1 位で正解の楽曲を検出した場合に Correct をカウントした．

実験に用いた計算機の CPU は Intel Core i7 CPU 960 (3.20[GHz])，物理メモリは 24GB であり，Ubuntu 12.04 LTS 上でシステムを実装し，動作させた．

5.3.3 実験結果

評価 1：提案手法のパラメータ決定

まず，提案手法の第 2 段階のスコアリングに用いる音響的類似度として最適な HMM の分布間距離を決定するため，式 4.2 に示した混合正規分布の平均ベクトル間のユークリッド距離の平均値 (d_{Euc}) と，式 4.4 に示したバタチャリヤ距離の平均値 (d_{Bhat}) のいずれかを用いて，第 1 段階の N -gram の値を変えながら，原曲同定の正解範囲を 1 位まで (Top-1)，3 位まで (Top-3)，5 位まで (Top-5) とした場合の F 値を求めた．この評価では，原曲同定時に閾値を設けず，第 2 段階に残った全パターンを順位付けした場合の F 値とした．その結果を表 5.5 に示す．表中の太字は Top-1，Top-3，Top-5 にそれぞれについて最大となった F 値を示している．Top-5 の場合は d_{Euc} と d_{Bhat} で最大の F 値は同値となったが，Top-1，Top-3 の場合は d_{Euc} が上回る結果となった． d_{Euc} を用いた場合， N -gram は Top-1 の場合のみ 7-gram が最も良い性能を示したが，Top-3，Top-5 の場合では 5-gram が最大となった．

表 5.5: 本実験の結果：提案手法の楽曲同定性能 (F 値)

分布間距離	d_{Euc}			d_{Bhat}		
	Top-1	Top-3	Top-5	Top-1	Top-3	Top-5
1	0.372	0.699	0.773	0.297	0.654	0.766
2	0.388	0.709	0.776	0.321	0.664	0.776
3	0.463	0.724	0.784	0.418	0.679	0.776
4	0.519	0.725	0.779	0.458	0.710	0.771
5	0.564	0.746	0.786	0.524	0.730	0.786
6	0.551	0.733	0.782	0.502	0.733	0.774
7	0.590	0.731	0.775	0.555	0.740	0.784
8	0.541	0.667	0.676	0.541	0.667	0.676
9	0.477	0.570	0.580	0.466	0.570	0.580
10	0.426	0.492	0.514	0.437	0.481	0.514

評価 2：従来手法と提案手法の原曲同定精度比較

表 5.5 の結果から，第 1 段階の N -gram を 7-gram，第 2 段階のスコアリングに用いる HMM の分布間距離を d_{Euc} に設定し，原曲同定時の閾値を変えながら提案手法の適合率-再現率曲線を求め，従来手法と比較を行った結果を図 5.2 に示す．図を見ると，再現率が高い領域において，提案手法 (Proposed Top-1) が従来手法 (Previous) よりも高い適合率-再現率を示していることが分かる．

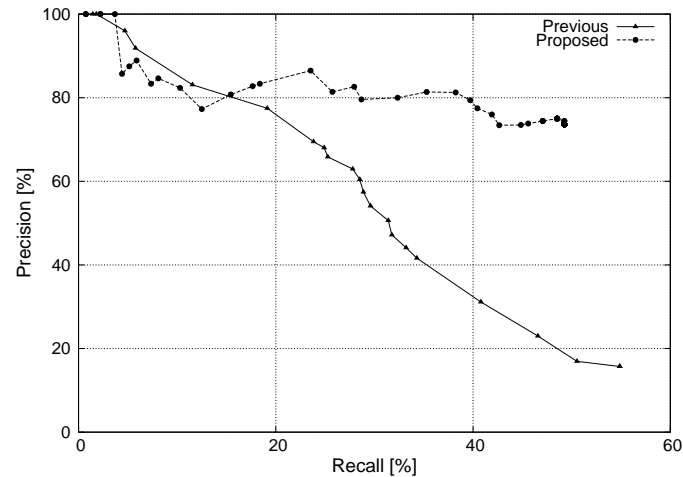


図 5.2: 本実験の結果：従来手法と提案手法の性能比較（適合率-再現率曲線）

図 5.2 の最大 F 値付近の閾値，適合率，再現率，F 値を表 5.6 に示す．適合率 74.44[%]，再現率 49.26[%] のとき，最大 F 値 0.593 となった．これは従来手法の最大 F 値 0.388 と比較して，0.205 ポイントの改善となった．

表 5.6: 本実験の結果：閾値に対する適合率，再現率，F 値

従来手法				提案手法			
閾値	適合率 [%]	再現率 [%]	F 値	閾値	適合率 [%]	再現率 [%]	F 値
0.59	60.43	28.52	0.388	0.32	75.00	48.53	0.589
0.60	57.42	28.88	0.384	0.33	75.00	48.53	0.589
0.61	54.10	29.60	0.383	0.34	75.00	48.53	0.589
0.62	50.61	31.41	0.388	0.35	74.44	49.26	0.593
0.63	47.18	31.77	0.380	0.36	73.63	49.26	0.590
0.64	44.11	33.21	0.379	0.37	73.63	49.26	0.590
0.65	41.62	34.30	0.376	0.38	73.63	49.26	0.590

評価 3：提案手法の計算量削減効果の評価

提案手法の第 1 段階について， N -gram を変化させながら，第 1 段階後の再現率，および候補パターン数の平均を求めた．その結果を表 5.7 に示す．DB の全パターン数の 480 個（40 曲 × 調 12 種類）と比較して，最も短い分割である 1-gram の場合でも 406.26 と，第 2 段階でマッチングするパターン数を削減できていることが分かる．2-gram，3-gram では，ほぼ再現率を維持したまま，候補パターンを更に削減できる．4-gram の場合は再現率の若干の低下を伴うが，90 %以上の再現率を保持したまま 3-gram よりも更に候補パターンを削減可能である．

表 5.7: 本実験の結果： N -gram における第 1 段階後の再現率と平均候補パターン数

N -gram	再現率 [%]	平均候補パターン数
1	100.00	406.26
2	97.79	282.00
3	97.79	173.74
4	92.65	94.30
5	85.29	45.70
6	78.68	21.70
7	66.91	10.35
8	52.21	5.56

次に提案手法の第 1 段階，第 2 段階について，各 N -gram で 10 回試行した平均計算時間を表 5.8 に示す．第 1 段階はテキストベースであるため，1-gram の場合でも 1.42[msec] と比較的短時間で処理が終了する．第 2 段階は 1-gram の場合だと 2 秒近い処理時間が掛かるが，5-gram では 225.70[msec] と 1-gram から大幅に計算時間が削減できている．

表 5.8: 本実験の結果：第 1 段階，第 2 段階の平均計算時間 [msec]

N -gram	第 1 段階 [msec]	第 2 段階 [msec]
1	1.42	1873.33
2	1.05	1334.13
3	0.82	848.06
4	0.77	464.91
5	0.59	225.70
6	0.47	104.60
7	0.37	50.19
8	0.36	23.99

5.3.4 考察

ここでは、提案手法でも同定が困難であった楽曲について考察する。

提案手法の場合は、まず第1段階のマッチング時点で正解パターンが候補に残らなければ、最終的な原曲同定結果も正解とならないが、楽曲ごとの結果を確認すると、第1段階終了時点で正解パターンを落としてしまう楽曲がいくつか存在する。第1段階の N -gram の N を表 5.7 において再現率が落ち始めている 5-gram とした場合、演奏区間において一度も正解パターンが候補に残らなかった楽曲を表 5.9 に示す。これらの楽曲を聴取してみた結果 (1) カバー演奏側に和音進行も大きく変更する編曲が加えられており、和音進行系列は取り出せているが、第1段階で N -gram を上げると候補から外れてしまうもの (2) カバー演奏側に伴奏が希薄となる極端な編曲が加えられており、和音進行認識自体が困難となっているもののいずれかであることが分かった。これらの楽曲については、和音進行パターンだけでは同定が難しいと考えられるため、今後、メロディー等の他の特徴量の併用といった検討が必要である。

表 5.9: 提案手法の第1段階で正解パターンが残らなかった楽曲 (5-gram の場合)

曲番号	曲名	原曲アーティスト	カバー演奏アーティスト
9	Come Together	The Beatles	Aerosmith
35	Thin Line Between Love And Hate	Persuaders	Annie Lennox
36	Tomorrow Never Knows	The Beatles	Phil Collins

第6章 結論

6.1 まとめ

本研究では、カバー演奏ストリームからの原曲同定システムについて、従来手法の問題点を解決する新しい手法の開発を目的とし、新規手法の提案を行った。提案したのは、HMMに基づく自動和音認識、転置インデックスによる候補パターン削減、ShiftCDP法による音響的類似度スコアリングの三つの要素技術を組み合わせた原曲同定手法である。また、従来手法、提案手法それぞれを実装したシステムを用いて評価実験を行い、提案手法の有効性を検証した。

評価実験では、原曲とカバー演奏が対になったデータセットを用いて、カバー演奏ストリームをシステムに入力する原曲同定実験を行った。評価尺度には原曲同定の適合率、再現率から計算されるF値を用いた。実験の結果、従来手法では最大のF値が0.388であったのに対し、提案手法では0.593となり、0.205ポイントの改善が見られた。計算量の面では、第1段階の転置インデックスによる候補パターン削減によって、再現率をほぼ維持したまま候補パターン数を大幅に削減することができ、その結果、第2段階の音響類似度によるマッチングの計算時間を削減できることが確認された。さらに、従来手法では必要であった人手によるパターンの事前分割を不要とし、楽曲データベース拡張の負担を軽減した。以上のことから、本研究の提案手法の有効性を確認することができた。

6.2 今後の課題

今後の課題として、以下のものが挙げられる。

(1) 和音 HMM の改良

本研究の提案手法で用いた和音 HMM は、楽曲の調の違いを考慮せず全ての楽曲に対して共通のモデルを用いた。しかし、和音の遷移確率は本来は調ごとに偏りがあると考えられる。この問題に対しては、調ごとに和音 HMM を作成する手法 [24] が提案されており、調を考慮しないモデルよりも高い認識性能を示すことが分かっている。提案手法においては和音認識の精度が原曲同定精度に直結すると考えられるため、和音 HMM を改良することで更なる精度向上が期待できる。

(2) メロディーなどの他の特徴量の併用

5.3.4 節でも述べたように、和音進行パターンだけでは同定が難しい楽曲が存在する。さらに、幅広い楽曲に対応可能なシステムとするためには、メロディーなど別の特徴の併用を検討していく必要があると考えられる。

(3) 楽曲数を増やした場合の評価

本研究では，従来手法との比較のために楽曲データベースの曲数を 40 曲で実験を行ったが，さらに楽曲数を増やして実験を行うことで，原曲同定精度や計算量において，実用に近い環境での評価が可能になると考えられる．しかしながら，現存する原曲とカバー演奏のデータセットでは楽曲数を大幅に拡張することは難しいため，今後はデータセットの充実を含めた研究が行われる必要がある．

謝辞

本研究を進めるにあたり、的確なご指導、ご助言を頂いた田中和世教授に心より感謝致します。また、貴重なご指摘を多数頂きました三河正彦准教授、藤澤誠助教授に御礼申し上げます。日々、研究室やゼミで多くのご意見を下さった田中研究室、三河研究室、藤澤研究室の皆様に感謝を申し上げます。研究室で共に励んだ大内慶久君、そして、大学3年次から4年間支えてくれた家族に心から感謝します。

中沢 彰吾

2013年3月

参考文献

- [1] 莪山真一, “楽音信号からの和音進行抽出手法と類似楽曲検索への応用”, PIONEER R&D, 14(2), pp.1-7, 2004.
- [2] D. P. W. Ellis, “Identifying ‘Cover Songs’ with Beat-Synchronous Chroma Features”, MIREX 2006 abstracts, pp. 32-35, 2006.
- [3] B. S. Ong, E. Gomez, P. Herrera, “音響信号の主要部分抜粋による楽曲のバージョン同定”, 情報処理学会 音楽情報科学研究会 研究報告, pp. 29-34, 2007.
- [4] 今野聡司, 二反田直己, 長谷川美紀, “メロディーとリズムに着眼した音楽信号の類似度に関する一考察”, 信学技報, 106(536), pp.125-128, 2007.
- [5] S. Ravuri, D. P. W. Ellis, “The Hydra system of unstructured cover song detection”, MIREX Abstracts, 2008.
- [6] 櫻井良樹, 浅野佑太, 田村哲嗣, 速水悟, “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”, 人工知能学会 全国大会, 1F1-3, 2009.
- [7] 井田崇介, 荒井秀一, “マスメディア上の音響信号からのリアルタイム楽曲スポッティング”, 電子情報通信学会 総合大会講演論文集, pp.470-471, 1998.
- [8] 深澤友貴, 三河正彦, 田中和世, “カバー演奏ストリームからのリアルタイム楽曲同定システム”, 日本音響学会音楽音響研究会資料, Vol.28, No.5, pp.9-13, 2009.
- [9] 中沢彰吾, “カバー演奏のリアルタイム原曲同定システムのための特徴抽出手法の研究”, 筑波大学 情報学群 情報メディア創成学類 卒業研究論文, 2011.
- [10] J. C. Brown, “Calculation of a constant Q spectral transform”, Journal of Acoustical Society of America, pp.425-434, 1991.
- [11] J. C. Brown, “An efficient Algorithm for the calculation of a constant Q spectral transform”, Journal of Acoustical Society of America, pp. 2698-2701, 1992.
- [12] T. Fujishima, “Real-time chord recognition of musical sound: A system using common lisp music”, Proc. ICMC, pp.464-467, 1999.
- [13] J. P. Bello, J. Pickens, “A robust mid-level representation for harmonic content in music signal”, Proc. ISMIR, pp.304-311, 2005.

- [14] 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, 100, pp.27–34, 2002.
- [15] 斉藤隆弘, “画像符号化アルゴリズム(III)–ベクトル量子化–”, テレビジョン学会誌, Vol.43, No.11, pp.1276–1284, 1989.
- [16] Y.Linde, A.Buzo, R.Gray, “An Algorithm for Vector Quantizer Design”, Transactions on Communications, Vol.28, No.1, pp.84–95, 1980.
- [17] 岡隆一, “連続 DP を用いた連続単語認識”, 日本音響学会音声研究会資料, S78-20, pp.145–152, 1978.
- [18] 伊藤慶明, “時系列パターンの任意部分区間の高速マッチング手法 ShiftCDP 法”, 電子情報通信学会論文誌, Vol.86–D–II, No.9, pp.1267–1277, 2003.
- [19] 鹿野清宏, 伊藤克亘, 河原達也, 武田一哉, 山本幹雄, “音声認識システム”, オーム社, 2001.
- [20] Steve Young, Gunnar Evermann, Mark Gales, Thomas Hain, Dan Kershaw, Xunying Liu, Gareth Moore, Julian Odell, Dave Ollason, Dan Povey, Valtcho Valtchev, Phil Woodland, “The HTK Book (for HTK Version 3.4)”, 2006.
- [21] A. Sheh, D. P. W. Ellis, “Chord segmentation and recognition using EM-trained hidden Markov models”, Proc. ISMIR, pp.183–189, 2003.
- [22] 内山裕貴, 宮本賢一, 西本卓也, 小野順貴, 嵯峨山茂樹, “調波音・打楽器音分離手法を用いた音楽音響信号からの自動和音認識”, 情報処理学会研究報告, pp.137–142, 2008.
- [23] 伊藤綾, 酒向慎司, 北村正, “状態共有型 HMM に基づく音楽音響信号からの自動和音認識の検討”, 情報科学技術フォーラム講演論文集, Vol.9, No.2, pp.285–286, 2010.
- [24] 上田雄, 小野順貴, 嵯峨山茂樹, “機能的和声モデルによる音楽信号からの和声推定”, 情報処理学会研究報告, Vol.2010–MUS–86, No.13, 2010.
- [25] S. Lee, Y. Nambu, H. Kojima, K. Tanaka, Y. Itoh, “Fast subword-based approach for open vocabulary spoken term detection”, Proc. ICA, No.975, 2010.
- [26] M. Mauch, C. Cannam, M. Davies, S. Dixon, C. Harte, S. Kolozali, D. Tidhar, M. Sandler, “OMRAS2 Metadata Project 2009”, Proc. ISMIR, 2009.
- [27] HTK Hidden Markov Model Toolkit, <http://htk.eng.cam.ac.uk/>, (参照 2013–1–14).
- [28] D. P. W. Ellis, “The covers80 cover song data set”, <http://labrosa.ee.columbia.edu/projects/coversongs/covers80/>, 2007, (参照 2013–1–14).

本研究に関する研究発表

- [a1] 中沢彰吾，三河正彦，田中和世，“カバー演奏ストリームからの楽曲同定のための特徴抽出手法の検討”，日本音響学会講演論文集，1-3-18，pp.917-918，2011-9．
- [a2] 中沢彰吾，三河正彦，田中和世，伊藤慶明，“隠れマルコフモデルによる自動和音認識を用いたカバー演奏ストリームからの楽曲同定手法の検討”，信学技報，Vol.112，No.357，PRMU2012-70，pp.1-6，2012-12．

付録 A HMMの学習に用いたデータセット

和音 HMM の学習に使用したデータセットの内容を表 A.1–A.4 に示す。

表 A.1: 和音 HMM の学習に用いたデータセット 曲番号 1–50

曲番号	曲名	曲番号	曲名
1	A Day In The Life	26	Carry That Weight
2	A Hard Day s Night	27	Chains
3	A Taste Of Honey	28	Come Together
4	Across The Universe	29	Cry Baby Cry
5	Act Naturally	30	Dear Prudence
6	All I ve Got To Do	31	Devil In Her Heart
7	All My Loving	32	Dig A Pony
8	All You Need Is Love	33	Dig It
9	And I Love Her	34	Dizzy Miss Lizzy
10	And Your Bird Can Sing	35	Do You Want To Know A Secret
11	Anna Go To Him	36	Doctor Robert
12	Another Girl	37	Don t Bother Me
13	Any Time At All	38	Don t Pass Me By
14	Ask Me Why	39	Drive My Car
15	Baby It s You	40	Eight Days A Week
16	Baby You re A Rich Man	41	Eleanor Rigby
17	Baby s In Black	42	Every Little Thing
18	Back In The U S S R	43	Everybody s Got Something To Hide Except Me And My Monkey
19	Because	44	Everybody s Trying To Be My Baby
20	Being For The Benefit Of Mr Kite	45	Fixing A Hole
21	Birthday	46	Flying
22	Blackbird	47	For No One
23	Blue Jay Way	48	For You Blue
24	Boys	49	Four
25	Can t Buy Me Love	50	Get Back

表 A.2: 和音 HMM の学習に用いたデータセット 曲番号 51–100

曲番号	曲名	曲番号	曲名
51	Getting Better	76	I Want To Tell You
52	Girl	77	I Want You She s So Heavy
53	Glass Onion	78	I Will
54	Golden Slumbers	79	I ll Be Back
55	Good Day Sunshine	80	I ll Cry Instead
56	Good Morning Good Morning	81	I ll Follow The Sun
57	Good Night	82	I m A Loser
58	Got To Get You Into My Life	83	I m Happy Just To Dance With You
59	Happiness Is A Warm Gun	84	I m Looking Through You
60	Hello Goodbye	85	I m Only Sleeping
61	Help	86	I m So Tired
62	Helter Skelter	87	I ve Got A Feeling
63	Her Majesty	88	I ve Just Seen A Face
64	Here Comes The Sun	89	If I Fell
65	Here There And Everywhere	90	If I Needed Someone
66	Hold Me Tight	91	In My Life
67	Honey Don t	92	It Won t Be Long
68	Honey Pie	93	It s Only Love
69	I Am The Walrus	94	Julia
70	I Don t Want To Spoil The Party	95	Let It Be
71	I Me Mine	96	Little Child
72	I Need You	97	Long Long Long
73	I Saw Her Standing There	98	Love Me Do
74	I Should Have Known Better	99	Love You To
75	I Wanna Be Your Man	100	Lovely Rita

表 A.3: 和音 HMM の学習に用いたデータセット 曲番号 101–150

曲番号	曲名	曲番号	曲名
101	Lucy In The Sky With Diamonds	126	Polythene Pam
102	Maggie Mae	127	Revolution 1
103	Magical Mystery Tour	128	Revolution 9
104	Martha My Dear	129	Rock And Roll Music
105	Maxwell s Silver Hammer	130	Rocky Raccoon
106	Mean Mr Mustard	131	Roll Over Beethoven
107	Medley Kansas City Hey Hey	132	Run For Your Life
108	Michelle	133	Savoy Truffle
109	Misery	134	Sexy Sadie
110	Money That s What I Want	135	Sgt Pepper s Lonely Hearts Club Band
111	Mother Nature s Son	136	Sgt Pepper s Lonely Hearts Club Band Reprise
112	Mr Moonlight	137	She Came In Through The Bathroom Window
113	No Reply	138	She Said She Said
114	Norwegian Wood This Bird Has Flown	139	She s Leaving Home
115	Not A Second Time	140	Something
116	Nowhere Man	141	Strawberry Fields Forever
117	Ob La Di Ob La Da	142	Sun King
118	Octopus s Garden	143	Taxman
119	Oh Darling	144	Tell Me What You See
120	One After 909	145	Tell Me Why
121	P S I Love You	146	The Continuing Story Of Bungalow Bill
122	Penny Lane	147	The End
123	Piggies	148	The Fool On The Hill
124	Please Mister Postman	149	The Long And Winding Road
125	Please Please Me	150	The Night Before

表 A.4: 和音 HMM の学習に用いたデータセット 曲番号 151–180

曲番号	曲名	曲番号	曲名
151	The Word	176	You Really Got A Hold On Me
152	There s A Place	177	You Won t See Me
153	Things We Said Today	178	You re Going To Lose That Girl
154	Think For Yourself	179	You ve Got To Hide Your Love Away
155	Ticket To Ride	180	Your Mother Should Know
156	Till There Was You		
157	Tomorrow Never Knows		
158	Twist And Shout		
159	Two Of Us		
160	Wait		
161	What Goes On		
162	What You re Doing		
163	When I Get Home		
164	While My Guitar Gently Weeps		
165	Why Don t We Do It In The Road		
166	Wild Honey Pie		
167	With A Little Help From My Friends		
168	Within You Without You		
169	Words Of Love		
170	Yellow Submarine		
171	Yer Blues		
172	Yesterday		
173	You Can t Do That		
174	You Like Me Too Much		
175	You Never Give Me Your Money		

付録 B 本実験で用いたデータセット

本実験で使ったデータセットの内容を表 B.1-B.3 に示す。

表 B.1: 本実験で用いたデータセット 曲番号 1-20

曲番号	楽曲名	原曲アーティスト	カバー演奏アーティスト
1	A Whiter Shade Of Pale	Procol Harum	Annie Lennox
2	Abracadabra	Steve Miller Band	Sugar Ray
3	Addicted To Love	Robert Palmer	Tina Turner
4	All Tomorrow s Parties	Velvet Underground	Japan
5	America	Simon and Garfunkel	Paul Simon
6	Between The Bars	Elliott Smith	Glen Phillips
7	Blue Collar Man	Styx	Reo Speedwagon
8	Caroline No	Beach Boys	Brian Wilson
9	Come Together	The Beatles	Aerosmith
10	Don t Let It Bring You Down	Neil Young	Annie Lennox
11	Downtown Lights	Blue Nile	Annie Lennox
12	God Only Knows	Beach Boys	Brian Wilson
13	Gold Dust Woman	Fleetwood Mac	Sheryl Crow
14	Grand Illusion	Styx	Reo Speedwagon
15	I Don t Want To Miss A Thing	Aerosmith	New Found Glory
16	I Love You	Celine Dion	Faith Hill
17	Lady	Styx	Reo Speedwagon
18	Let It Be	The Beatles	Nick Cave
19	More Than Words	Extreme	Westlife
20	My Generation	The Who	Green Day

表 B.2: 本実験で用いたデータセット 曲番号 21–50

曲番号	楽曲名	原曲アーティスト	カバー演奏アーティスト
21	My Heart Will Go On	Celine Dion	New Found Glory
22	New Age	Velvet Underground	Tori Amos
23	Never Let Me Down Again	Depeche Mode	Smashing Pumpkins
24	Purple Rain	Prince	Leann Rimes
25	Rattlesnakes	Lloyd Cole and The Commotions	Tori Amos
26	Real Men	Joe Jackson	Tori Amos
27	September Gurls	Big Star	Bangles
28	Something So Right	Paul Simon	Annie Lennox
29	Straight From The Heart	Bryan Adams	Bonnie Tyler
30	Strange Little Girl	Stranglers	Tori Amos
31	Summer of 69	Bryan Adams	MxPx
32	Take Me To The River	Al Green	Annie Lennox
33	Take On Me	A Ha	MxPx
34	The Boxer	Simon and Garfunkel	Paul Simon
35	Tomorrow Never Knows	The Beatles	Phil Collins
36	Toys In The Attic	Aerosmith	R.E.M.
37	Thin Line Between Love And Hate	Persuaders	Annie Lennox
38	Walk This Way	Aerosmith	Run-D.M.C.
39	We Can Work It Out	The Beatles	Tesla
40	Yesterday	The Beatles	En Vogue
41	All Along The Watchtower	Bob Dylan	Jimi Hendrix
42	Before You Accuse Me	Creedence Clearwater Revival	Eric Clapton
43	Cecilia	Simon and Garfunkel	Paul Simon
44	Claudette	Everly Brothers	Roy Orbison
45	Cocaine	Eric Clapton	Nazareth
46	Day Tripper	The Beatles	Cheap Trick
47	Enjoy the Silence	Depeche Mode	Tori Amos
48	Faith	George Michael	Limp Bizkit
49	Happiness is a Warm Gun	The Beatles	Tori Amos
50	Heart Of Gold	Neil Young	Tori Amos

表 B.3: 本実験で用いたデータセット 曲番号 51-80

曲番号	楽曲名	原曲アーティスト	カバー演奏アーティスト
51	Hush	Deep Purple	Milli Vanilli
52	I Can t Get Next To You	Temptations	Annie Lennox
53	I Can t Get No Satisfaction	Rolling Stones	Britney Spears
54	I Don t Like Mondays	Boomtown Rats	Tori Amos
55	I m Losing You	John Lennon	Cheap Trick
56	I m Not In Love	10cc	Tori Amos
57	It s Tricky	Run Dmc	Bloodhound Gang
58	Little Wing	Eric Clapton	The Corrs
59	Lodi	Creedence Clearwater Revival	Tesla
60	Love Hurts	Nazareth	Heart
61	Maggie s Farm	Bob Dylan	Rage Against The Machine
62	Night Time Is The Right Time	Aretha Franklin	Creedence Clearwater Revival
63	No Woman No Cry	Bob Marley	Fugees
64	Oh Pretty Woman	Roy Orbison	Al Green
65	Ooby Dooby	Roy Orbison	Creedence Clearwater Revival
66	Proud Mary	Creedence Clearwater Revival	Tina Turner
67	Red Red Wine	Neil Diamond	UB40
68	River Deep Mountain High	Celine Dion	Tina Turner
69	She Came In Through The Bathroom Window	The Beatles	Joe Cocker
70	Stone Cold Crazy	Queen	Metallica
71	Street Fighting Man	Rolling Stones	Rage Against The Machine
72	Summertime Blues	Beach Boys	Alan Jackson
73	Time	Tom Waits	Tori Amos
74	Train In Vain	Clash	Annie Lennox
75	Tush	ZZ Top	Nazareth
76	Waiting In Vain	Bob Marley	Annie Lennox
77	Walking After Midnight	Bryan Adams	Garth Brooks
78	What s Going On	Marvin Gaye	Cyndi Lauper
79	White Room	Cream	Sheryl Crow
80	Wish You Were Here	Pink Floyd	Wyclef Jean

付録 C 従来手法の評価に用いた正解ラベル

本実験の従来手法の評価に使用した正解ラベルを表 C.1–C.5 に示す。

表 C.1: 従来手法の評価に使用した正解ラベル (1)

秒数	正解曲名	秒数	正解曲名
53	Come Together	1000	Don't Let It Bring You Down
97	Come Together	1028	Don't Let It Bring You Down
138	Come Together	1120	Downtown Lights
208	Come Together	1163	Downtown Lights
338	Take Me To The River	1197	Downtown Lights
356	Take Me To The River	1238	Downtown Lights
393	Take Me To The River	1292	Downtown Lights
410	Take Me To The River	1341	Downtown Lights
428	Take Me To The River	1471	Thin Line Between Love And Hate
466	Take Me To The River	1484	Thin Line Between Love And Hate
484	Take Me To The River	1497	Thin Line Between Love And Hate
500	Take Me To The River	1548	Thin Line Between Love And Hate
555	A Whiter Shade Of Pale	1572	Thin Line Between Love And Hate
585	A Whiter Shade Of Pale	1612	Thin Line Between Love And Hate
613	A Whiter Shade Of Pale	1625	Thin Line Between Love And Hate
659	A Whiter Shade Of Pale	1788	Something So Right
691	A Whiter Shade Of Pale	1815	Something So Right
719	A Whiter Shade Of Pale	1851	Something So Right
771	A Whiter Shade Of Pale	1864	Something So Right
864	Don't Let It Bring You Down	1884	Something So Right
878	Don't Let It Bring You Down	1911	Something So Right
915	Don't Let It Bring You Down	2001	September Gurls
930	Don't Let It Bring You Down	2009	September Gurls
954	Don't Let It Bring You Down	2033	September Gurls
985	Don't Let It Bring You Down	2041	September Gurls

表 C.2: 従来手法の評価に使用した正解ラベル(2)

秒数	正解曲名	秒数	正解曲名
2061	September Gurls	3389	Between The Bars
2100	September Gurls	3410	Between The Bars
2108	September Gurls	3431	Between The Bars
2116	September Gurls	3451	Between The Bars
2173	Straight From The Heart	3474	Between The Bars
2195	Straight From The Heart	3510	My Generation
2216	Straight From The Heart	3531	My Generation
2239	Straight From The Heart	3571	My Generation
2259	Straight From The Heart	3703	All Tomorrow's Parties
2301	Straight From The Heart	3724	All Tomorrow's Parties
2324	Straight From The Heart	3780	All Tomorrow's Parties
2410	God Only Knows	3802	All Tomorrow's Parties
2434	God Only Knows	3872	All Tomorrow's Parties
2487	God Only Knows	3894	All Tomorrow's Parties
2643	Caroline No	4016	Purple Rain
2668	Caroline No	4052	Purple Rain
2687	Caroline No	4088	Purple Rain
2717	Caroline No	4125	Purple Rain
2743	Caroline No	4161	Purple Rain
2760	Caroline No	4197	Purple Rain
2794	Caroline No	4275	Summer of 69
2896	Yesterday	4284	Summer of 69
2926	Yesterday	4297	Summer of 69
2956	Yesterday	4310	Summer of 69
2980	Yesterday	4323	Summer of 69
3072	I Love You	4340	Summer of 69
3105	I Love You	4359	Summer of 69
3141	I Love You	4372	Summer of 69
3168	I Love You	4445	Take On Me
3185	I Love You	4464	Take On Me
3205	I Love You	4477	Take On Me
3238	I Love You	4496	Take On Me
3269	I Love You	4538	Take On Me
3285	I Love You	4557	Take On Me
3368	Between The Bars	4576	Take On Me

表 C.3: 従来手法の評価に使用した正解ラベル (3)

秒数	正解曲名	秒数	正解曲名
4594	Take On Me	5437	The Boxer
4614	Take On Me	5450	The Boxer
4652	Don't Want To Miss A Thing	5464	The Boxer
4668	Don't Want To Miss A Thing	5477	The Boxer
4690	Don't Want To Miss A Thing	5513	The Boxer
4706	Don't Want To Miss A Thing	5521	The Boxer
4730	Don't Want To Miss A Thing	5540	The Boxer
4745	Don't Want To Miss A Thing	5574	The Boxer
4753	Don't Want To Miss A Thing	5584	The Boxer
4762	Don't Want To Miss A Thing	5594	The Boxer
4823	My Heart Will Go On	5604	The Boxer
4844	My Heart Will Go On	5719	Tomorrow Never Knows
4863	My Heart Will Go On	5795	Tomorrow Never Knows
4883	My Heart Will Go On	5931	Toys In The Attic
4932	My Heart Will Go On	5960	Toys In The Attic
4995	Let It Be	5991	Toys In The Attic
5009	Let It Be	6029	Toys In The Attic
5037	Let It Be	6143	Blue Collar Man Long Nights
5050	Let It Be	6161	Blue Collar Man Long Nights
5066	Let It Be	6182	Blue Collar Man Long Nights
5107	Let It Be	6196	Blue Collar Man Long Nights
5122	Let It Be	6226	Blue Collar Man Long Nights
5136	Let It Be	6279	Blue Collar Man Long Nights
5193	America	6312	Blue Collar Man Long Nights
5207	America	6407	Grand Illusion
5226	America	6422	Grand Illusion
5237	America	6444	Grand Illusion
5266	America	6470	Grand Illusion
5280	America	6569	Grand Illusion
5295	America	6619	Grand Illusion
5313	America	6826	Lady
5322	America	6859	Lady
5338	America	6885	Lady
5394	The Boxer	6917	Lady
5410	The Boxer	6944	Lady

表 C.4: 従来手法の評価に使用した正解ラベル (4)

秒数	正解曲名	秒数	正解曲名
6972	Lady	8093	We Can Work It Out
7034	Walk This Way	8111	We Can Work It Out
7062	Walk This Way	8124	We Can Work It Out
7080	Walk This Way	8137	We Can Work It Out
7116	Walk This Way	8155	We Can Work It Out
7143	Walk This Way	8204	Addicted To Love
7162	Walk This Way	8221	Addicted To Love
7331	Gold Dust Woman	8240	Addicted To Love
7345	Gold Dust Woman	8274	Addicted To Love
7361	Gold Dust Woman	8292	Addicted To Love
7374	Gold Dust Woman	8327	Addicted To Love
7391	Gold Dust Woman	8338	Addicted To Love
7418	Gold Dust Woman	8395	Addicted To Love
7433	Gold Dust Woman	8413	Addicted To Love
7448	Gold Dust Woman	8447	Addicted To Love
7463	Gold Dust Woman	8533	New Age
7627	Never Let Me Down Again	8569	New Age
7655	Never Let Me Down Again	8619	New Age
7701	Never Let Me Down Again	8656	New Age
7719	Never Let Me Down Again	8796	Strange Little Girl
7737	Never Let Me Down Again	8816	Strange Little Girl
7752	Never Let Me Down Again	8847	Strange Little Girl
7769	Never Let Me Down Again	8862	Strange Little Girl
7784	Never Let Me Down Again	8904	Strange Little Girl
7800	Never Let Me Down Again	8920	Strange Little Girl
7855	Abacadabra	8936	Strange Little Girl
7870	Abacadabra	8952	Strange Little Girl
7884	Abacadabra	8968	Strange Little Girl
7900	Abacadabra	9031	Rattlesnakes
7929	Abacadabra	9050	Rattlesnakes
7944	Abacadabra	9071	Rattlesnakes
7963	Abacadabra	9087	Rattlesnakes
8048	We Can Work It Out	9111	Rattlesnakes
8066	We Can Work It Out	9135	Rattlesnakes
8080	We Can Work It Out	9148	Rattlesnakes

表 C.5: 従来手法の評価に使用した正解ラベル (5)

秒数	正解曲名	分割
9164	Rattlesnakes	2
9180	Rattlesnakes	2
9210	Rattlesnakes	1
9254	Real Men	1
9278	Real Men	1
9310	Real Men	2
9334	Real Men	1
9357	Real Men	1
9390	Real Men	2
9412	Real Men	1
9437	Real Men	1
9506	More Than Words	1
9542	More Than Words	2
9584	More Than Words	3
9624	More Than Words	1
9660	More Than Words	2
9704	More Than Words	3

付録 D 提案手法の評価に用いた正解ラベル

本実験の提案手法の評価に使用した正解ラベルを表 D.1–D.3 に示す。

表 D.1: 提案手法の評価に使用した正解ラベル (1)

バッファ	曲名	バッファ	曲名
1	Come Together	26	Thin Line Between Love And Hate
2	Come Together	27	Something So Right
3	Come Together	28	Something So Right
4	Come Together	29	Something So Right
5	Take Me To The River	30	September Gurls
6	Take Me To The River	31	September Gurls
7	Take Me To The River	32	Straight From The Heart
8	A Whiter Shade Of Pale	33	Straight From The Heart
9	A Whiter Shade Of Pale	34	Straight From The Heart
10	A Whiter Shade Of Pale	35	God Only Knows
11	A Whiter Shade Of Pale	36	God Only Knows
12	A Whiter Shade Of Pale	37	God Only Knows
13	Don t Let It Bring You Down	38	Caroline No
14	Don t Let It Bring You Down	39	Caroline No
15	Don t Let It Bring You Down	40	Caroline No
16	Downtown Lights	41	Caroline No
17	Downtown Lights	42	Caroline No
18	Downtown Lights	43	Yesterday
19	Downtown Lights	44	Yesterday
20	Downtown Lights	45	I Love You
21	Downtown Lights	46	I Love You
22	Downtown Lights	47	I Love You
23	Thin Line Between Love And Hate	48	I Love You
24	Thin Line Between Love And Hate	49	I Love You
25	Thin Line Between Love And Hate	50	Between The Bars

表 D.2: 提案手法の評価に使用した正解ラベル (2)

バッファ	曲名	バッファ	曲名
51	Between The Bars	76	The Boxer
52	My Generation	77	The Boxer
53	All Tomorrow s Parties	78	The Boxer
54	All Tomorrow s Parties	79	The Boxer
55	All Tomorrow s Parties	80	Tomorrow Never Knows
56	All Tomorrow s Parties	81	Tomorrow Never Knows
57	All Tomorrow s Parties	82	Tomorrow Never Knows
58	Purple Rain	83	Tomorrow Never Knows
59	Purple Rain	84	Toys In The Attic
60	Purple Rain	85	Toys In The Attic
61	Purple Rain	86	Blue Collar Man Long Nights
62	Summer of 69	87	Blue Collar Man Long Nights
63	Take On Me	88	Blue Collar Man Long Nights
64	Take On Me	89	Blue Collar Man Long Nights
65	Take On Me	90	Blue Collar Man Long Nights
66	I Don t Want To Miss A Thing	91	Grand Illusion
67	I Don t Want To Miss A Thing	92	Grand Illusion
68	My Heart Will Go On	93	Grand Illusion
69	My Heart Will Go On	94	Grand Illusion
70	Let It Be	95	Grand Illusion
71	Let It Be	96	Lady
72	Let It Be	97	Lady
73	America	98	Lady
74	America	99	Lady
75	America	100	Walk This Way

表 D.3: 提案手法の評価に使用した正解ラベル (3)

バッファ	曲名	バッファ	曲名
101	Walk This Way	126	Strange Little Girl
102	Walk This Way	127	Strange Little Girl
103	Walk This Way	128	Rattlesnakes
104	Walk This Way	129	Rattlesnakes
105	Gold Dust Woman	130	Rattlesnakes
106	Gold Dust Woman	131	Real Men
107	Gold Dust Woman	132	Real Men
108	Gold Dust Woman	133	Real Men
109	Never Let Me Down Again	134	More Than Words
110	Never Let Me Down Again	135	More Than Words
111	Never Let Me Down Again	136	More Than Words
112	Abacadabra		
113	Abacadabra		
114	Abacadabra		
115	We Can Work It Out		
116	Addicted To Love		
117	Addicted To Love		
118	Addicted To Love		
119	Addicted To Love		
120	Addicted To Love		
121	New Age		
122	New Age		
123	New Age		
124	New Age		
125	Strange Little Girl		