

フィルタバンク特徴量と Earth Mover's Distance を用いた音楽検索

柘 植 覚^{†1} 三 好 真 人^{†2}
肖 清 梅^{†2} 北 研 二^{†2}

本稿では、フィルタバンク特徴量と Earth Mover's Distance (EMD) を用いた音楽検索手法を提案する。提案手法では、音楽データから抽出される MFCC (mel-frequency cepstral coefficient) をクラスタリングし求められた各クラスタのセントロイドおよび各クラスタに属する MFCC の個数を特徴量として使用する。音楽検索を行う際の距離尺度として用いた EMD は、2 つの離散分布において一方の分布を他方の分布に変換するための最小コストとして定義される。日本のポップス 780 曲に対する音楽検索実験結果より、MP3 により 32kbps に圧縮を行った評価データに対し、96.73% の検索正解率を示した。

Music retrieval method based on filter-bank feature and earth mover's distance

SATORU TSUGE,^{†1} MASATO MIYOSHI,^{†2} QINGMEI XIAO^{†2}
and KENJI KITA^{†2}

In this paper, we propose a novel music retrieval method using a filterbank feature and earth mover's distance (EMD). In the proposed method, we use MFCCs for an acoustic features and EMD for a distance measurement. Evaluation experimental result shows that the retrieval correct rate of the proposed method is 96.73%.

^{†1} 大同大学
Daido University

^{†2} 徳島大学
the University of Tokushima

1. はじめに

近年、インターネットによる音楽配信などの普及により大量の音楽データの入手が容易となり、大容量の音楽データをポータブルオーディオプレイヤーやパーソナルコンピュータなどに保存し、利用可能となってきている。音楽データが大量になると、人手による分類や検索が困難になるという問題が発生する。近年では、有効な楽曲検索や分類方法の開発が要求されており、研究が盛んに行われている¹⁾²⁾。大量のデータの中から高速で所望の音楽データを検索する音楽検索手法として、音楽データの特徴を二進数で表現するオーディオ指紋 (audio fingerprint) を用いた高速な検索手法が提案されている³⁾⁴⁾⁵⁾⁶⁾。実際にオーディオ指紋を用いた高速な音楽検索アプリケーションが提供されている。

一方、インターネットの整備・普及による弊害として、著作権を侵害した違法コンテンツがネットワーク上に大量に存在するようになってきている。違法コンテンツ中には、音楽データも数多く含まれている。著作権を侵害した音楽データを発見するためにも、高精度な音楽検索手法の開発が必要となってきている。インターネット上の音楽データは再録音や圧縮・伸長されている場合が多く、元音楽データと比較すると音楽品質の劣化やテンポの変化がみられる。特徴量にオーディオ指紋を用いた音楽検索手法の場合、テンポの変化に対する柔軟性が低く、音楽情報が著しく欠落した 0, 1 のビット列で表現されているため、インターネット上の低品質な違法コンテンツに対しては高精度に検索ができない可能性があると考えられる。そこで、音声認識などで用いられるフィルタバンク特徴量を用いた音楽検索手法が提案されている⁷⁾。フィルタバンク特徴量を使用するため音楽の音響的特徴が保持可能である。さらに、量子化を行うため音楽データ全体 (もしくは一部分) とのマッチングが可能となり、高精度で柔軟な音楽検索が期待できる。

本稿では、音響的情報の欠落が少ないフィルタバンク特徴量 (MFCC; mel-frequency cepstral coefficient)、距離尺度として Earth Mover's Distance (EMD)⁸⁾ を用いた音楽検索手法を提案する。MFCC は、人間の聴覚特性を近似したメル尺度を用い分析するため、人間の感覚に近い音楽検索が期待できる。提案手法は、音楽データから抽出した MFCC をクラスタリングし、各クラスタのセントロイドおよび各クラスタに分類された MFCC の個数を各音楽データの特徴量とする。これは、音楽データをセントロイドのヒストグラムとして表現することである。音楽検索を行う際の距離尺度として、異なるヒストグラム間の距離を人間の感覚に近い尺度で測ることが可能な EMD を用いることにより、高精度な音楽検索が期待できる。

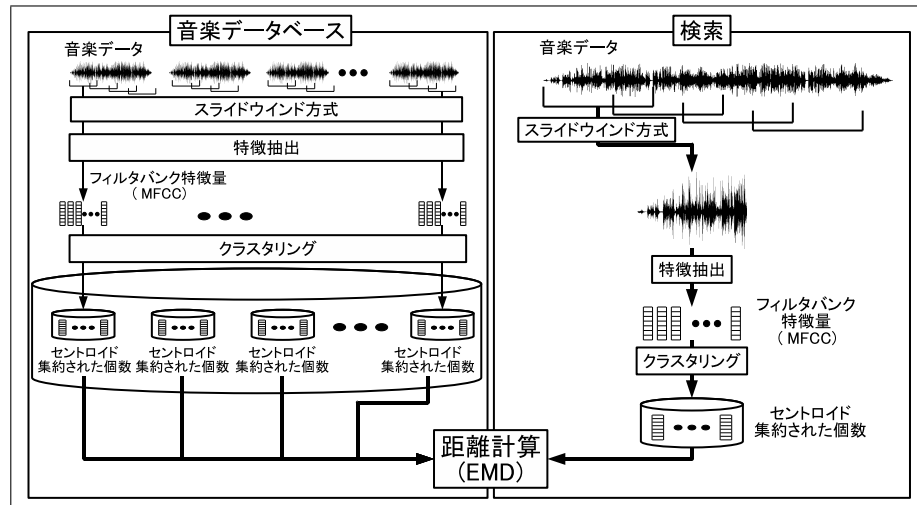


図 1 提案手法の流れ
 Fig. 1 Flow of proposed method

音楽検索の距離尺度として EMD を用いる研究は文献 9) などで行われているが、この研究は音楽データが MIDI (music instrumental digital interface) を対象に研究されている。本稿では、リニア PCM 形式である音楽データから特徴を抽出し距離計算に EMD を用いる手法を提案する。本提案手法の有効性を検証するため、約 800 曲の日本のポップ音楽を対象に検索実験を行う。

以下、2. においてフィルタバンク特徴量と EMD を用いた音楽検索手法を提案する。つぎに、3. にて提案手法の有効性を検証するため行った音楽検索実験の内容および実験結果を述べる。最後に、4. で本稿のまとめと今後の課題を述べる。

2. フィルタバンク特徴量と EMD を用いた音楽検索手法

本節では、提案手法であるフィルタバンク特徴量を用いた EMD による音楽検索手法について述べる。図 1 に提案手法の流れを示す。図に示す通り、提案手法は 1 曲を複数の曲片に分割し、各曲片から抽出した MFCC に対しクラスタリングを行いセントロイドおよび集約された MFCC の個数を求め、あらかじめ保存してあるデータベース内の曲片との EMD を計算することにより音楽検索を行う。以下に、特徴抽出方法と距離計算手法の詳細を示す。

2.1 フィルタバンク特徴量

特徴量抽出手法を図 1 の検索部 (図中の右) を例に説明する。はじめに、曲をスライドウインド方式により複数の曲片に分割する。分割された各曲片に対し、フィルタバンク分析を行い、フィルタバンク特徴量である MFCC (mel-frequency cepstral coefficient) を求める。ここで、曲 A の i 番目の曲片から求めた MFCC の集合を $M_{A_i} = \{M_{A_i}^0, M_{A_i}^1, \dots, M_{A_i}^{L_{A_i}-1}\}$ とする。 $M_{A_i}^l$ は l フレーム目の N 次元の MFCC、 L_{A_i} は曲 A の i 番目の曲片の合計フレーム数を示す。MFCC は、人間の聴覚特性を基に分析される特徴量であるため、人間の感覚に近い音楽検索が可能であると考えられる。

つぎに、曲片ごとに求められた MFCC を K -means 法によりクラスタリングを行い、 K 個のセントロイド ($C_{A_i} = \{C_{A_i}^0, C_{A_i}^1, \dots, C_{A_i}^{K-1}\}$) を求める。また、各クラスタに分類された MFCC の個数 ($W_{A_i} = \{W_{A_i}^0, W_{A_i}^1, \dots, W_{A_i}^{K-1}\}$) を求める。分類された個数の多いクラスタのセントロイドおよび集約された MFCC の個数を曲片 A_i の特徴量とする。

2.2 距離計算方法

音楽データベース内の音楽データに対し前節で述べた特徴量抽出方法により求められたセントロイド集合と集約された MFCC の個数を各曲片の特徴量として、検索対象データベースに保存する。検索時には、同様の方法で入力音楽データから特徴量を抽出し、検索対象データベース内のデータと距離を計算し、最小距離となる音楽データを探索する。この際の距離として、EMD およびセントロイド間距離を検討する。

2.2.1 Earth Mover's Distance

EMD は線型計画問題のひとつである輸送問題の解に基づいて計算される距離尺度であり、2 つの離散分布において、一方の分布を他方の分布に変換するための最小コストとして定義される。輸送問題とは、需要地の需要を満たすように供給地から需要地へ輸送を行う場合の最小輸送コストを解く問題である。 P, Q をそれぞれ供給地、需要地とし EMD を計算する。 P, Q を前節で求めた各曲片のセントロイド集合とセントロイドに集約された個数とすることにより、各曲片をヒストグラムとして表現でき、音楽検索を行う距離計算に EMD を用いることが可能となる。

はじめに、 P, Q の各要素に対する単位輸送あたりの輸送コスト (d_{ij}) を計算する。本稿では、この輸送コスト (d_{ij}) はユークリッド距離を用いる。つぎに、 p_i と q_j のすべての組み合わせを考慮し、総輸送コストを計算する。総輸送コストは、 P から Q への輸送量 (フロー) ($F = \{f_{ij}\}$) を決定する以下の輸送問題の解を用いて計算する。任意の供給地・需要地の組み合わせによる総輸送量 (WORK) は次式のように表すことができる。

$$\text{WORK}(P, Q) = \sum_{i=1}^m \sum_{j=1}^n d_{ij} f_{ij} \quad (1)$$

この目的関数は、 i, j 間の輸送量に単位輸送コストを掛けて和をとることで総輸送コストが計算されることを表している。ただし総輸送コストを計算する場合、以下の制約条件(式(2)~式(5))を満たすものとする。

- 供給地から需要地の一方にしか輸送されない

$$f_{ij} \geq 0, \quad (1 \leq i \leq m, 1 \leq j \leq n) \quad (2)$$

- 供給地 i から供給できる容量は供給量 w_{p_i} を超過しない

$$\sum_{j=1}^n f_{ij} \leq w_{p_i}, \quad (1 \leq i \leq m) \quad (3)$$

- 需要地 j が受け取れる容量は w_{q_j} は以下であること

$$\sum_{i=1}^m f_{ij} \leq w_{q_j}, \quad (1 \leq j \leq n) \quad (4)$$

- 供給地から移動する輸送量(総フロー)

$$\sum_{i=1}^m \sum_{j=1}^n f_{ij} = \min \left(\sum_{i=1}^m w_{p_i}, \sum_{j=1}^n w_{q_j} \right) \quad (5)$$

最終的に $\text{EMD}(P, Q)$ は、上の輸送問題の最適値(総輸送コストの最小値) $\text{WORK}(P, Q)$ を総フローで割って以下のように求まる。

$$\text{EMD}(P, Q) = \frac{\text{WORK}(P, Q)}{\sum_{i=1}^m \sum_{j=1}^n f_{ij}} \quad (6)$$

提案手法の場合、この $\text{EMD}(P, Q)$ が各曲片間の距離となる。

2.2.2 セントロイド間距離

前節では、クラスタリングにより求められたセントロイドおよびそのセントロイドを用いたベクトル量子化を行った結果(各セントロイドに集約された特徴量の個数)に対する距離尺度として EMD を用いることを述べた。本節では、クラスタリングにより求められた曲片ごとのセントロイド集合間の距離計算方法を述べる。曲 A の i 番目の曲片と曲 B の j 番目の曲片のセントロイド集合間の距離 $(D(C_{A_i}, C_{B_j}))$ を以下と定義する。

$$D(C_{A_i}, C_{B_j}) = \min(\text{Dist}(C_{A_i}, C_{B_j}), \text{Dist}(C_{B_j}, C_{A_i})) \quad (7)$$

$\text{Dist}()$ は以下の式で計算する。

$$\text{Dist}(C_{A_i}, C_{B_j}) = \frac{\sum_{m=0}^{M-1} \min_n(|C_{A_i}^m, C_{B_j}^n|)}{M} \quad (8)$$

ここで、 M は曲 A の i 番目の曲片のセントロイド数を示し、 $C_{A_i}^m$ は m 番目のセントロイドを示す。また、 $|a, b|$ はベクトル a, b 間のユークリッド距離を示す。

3. 音楽検索実験

提案手法の有効性を検証するため、日本のポップス 780 曲に対する音楽検索実験を行った。音楽データは、標本化周波数を 44,100Hz から 16,000Hz にダウンサンプリングした量子化ビット数 16bit のリニア PCM データである。

3.1 実験条件

3.1.1 音楽データ

• 検索対象データ

検索対象データである 780 曲に対し、前後の無音部分を取り除き、スライドウィンド方式で曲片に分割した。ウィンドは方形窓を用い、ウィンド幅は 60 秒、ウィンド移動時間は 30 秒とした。分割後の曲片 5,135 個を検索対象データとして用いた。

• 評価データ

検索対象データから 41 曲を選択し、検索対象データと同様にスライドウィンド方式により曲片に分割した。ウィンド移動時間は 30 秒とし、ウィンド幅と分割開始位置を以下の 3 種類に変更した評価セットを準備した。最後の [] は評価セット名とする。

- ウィンド幅: 60 秒、分割開始位置は検索対象データと同様(検索対象データと同様の分割方法)(曲片数: 253)[E60]
- ウィンド幅: 60 秒、分割開始位置を検索対象データから 15 秒遅れされた(曲片数: 275)[D60]
- ウィンド幅: 30 秒、分割開始位置は検索対象データと同様(曲片数: 253)[E30]

上記の曲片(量子化ビット数 16bit のリニア PCM データ)および、各評価セットを MP3 により 32, 64, 128, 192kbps に圧縮し、量子化ビット数 16bit に伸長したリニア PCM データを評価データとして用いた。

3.1.2 特徴量抽出条件

MFCC を抽出する音響分析条件を表 1 に示す。表 1 の分析条件で抽出された MFCC を

表 1 音響分析条件

Table 1 Feature extraction conditions

サンプリング周波数	16kHz
フィルタバンク数	24
ケプストラム次数	12
プリエンファシス	0.95
フレーム長	25ms
フレームシフト	10ms

表 2 実験結果 ([E60]) (検索正解率 (%))

Table 2 Experimental results in test set E60 (Correct rate in %)

手法	セントロイド数	ビットレート (kbps)			
		32	64	128	192
<i>CentDist</i>	20	30.91	34.55	34.55	36.00
	32	23.64	28.36	29.45	30.18
	64	82.55	92.00	92.36	91.64
<i>EMD</i>	20	76.73	84.00	85.82	88.73
	32	84.36	93.09	93.45	93.82
	64	8.00	9.45	8.73	8.36
<i>EMD2</i>	20	95.64	98.91	99.64	99.64
	32	96.73	99.27	99.64	99.64
	64	1.45	1.45	1.45	1.45
<i>AFP</i>		81.09	97.45	97.45	97.45

K-means 法により 128 個のクラスにクラスタリングし、各クラスのスセントロイドを用いベクトル量子化を行い、各セントロイドに集約された個数を計算した。各曲片の特徴量として、集約された個数の多いセントロイドおよびそのセントロイドに集約された MFCC の個数を用いた。なお、EMD 計算の際に用いる重み (集約された MFCC の個数) は使用したセントロイド数に関わらず各曲片内で合計が 100 となるように正規化を行った。比較のため曲片ごとにオーディオ指紋 (分析条件は文献 3) と同様) も抽出をした。評価は各曲片が正確に検索された割合 (検索正解率) で行った。

3.2 実験結果

3.2.1 音楽圧縮による影響

はじめに、音楽圧縮の影響による検索正解率の変化を調べるため、検索対象データと同様のスライドウィンド方式を用いた評価データ [E60] に対する評価実験を行った。結果を表 2 に示す。表中の *CentDist* は距離計算に 2.2.2 で述べたセントロイド間距離を用いた場合の

検索結果を示す。*EMD* は曲片ごとにセントロイドを求め距離計算に *EMD* を用いた検索結果を示す。*EMD2* は、検索対象データ全体から MFCC のセントロイドを求め、各曲片に対して求められたセントロイド集合を用いベクトル量子化を行い、集約された個数の多いセントロイドを特徴量とし、*EMD* により距離計算を行った結果を示す。セントロイド数の違いによる検索精度の変化を調べるため、集約された個数が多い上位 20, 32, 64 個の結果を示す。また、比較として表中に特徴量としてオーディオ指紋を用いた場合の検索結果 (*AFP*) を示す。

表より、*CentDist* はセントロイド数が 64、*EMD*、*EMD2* はセントロイド数が 32 の場合、高い検索精度を示すことがわかる。*CentDist* はセントロイド間の距離を集約される個数とは関係なく比較を行っているため、曲片を適切に表現するには多くのセントロイド数が必要であると考えられる。一方、距離計算に *EMD* を用いる手法 (*EMD*、*EMD2*) はセントロイド数を 64 に増加させると検索正解率が低下することがわかる。本実験では、*EMD* の計算に用いる重み (セントロイドに集約された MFCC の個数) は各曲片ごとに 100 と正規化を行っている。セントロイド数を増加させると重要と思われる集約された個数の多いセントロイドの重みが相対的に減少することとなり、検索正解率低下の原因となっていると考えられる。そのため、*EMD*、*EMD2* では、セントロイド数が 32 がもっとも曲片を表現できており検索率が高くなったと考えられる。また表の *EMD* と *EMD2* の比較により、*EMD2* の検索率が *EMD* より高いことがわかる。*EMD* と *EMD2* の違いは曲片に対しベクトル量子化を行う際のセントロイドの異なりである。この結果より、ベクトル量子化に用いるセントロイドは、曲片ごとに作成するより多量なデータで作成し、集約される個数によりセントロイドを選択することが音楽検索に有効であることがわかった。これは、音響特徴空間である MFCC 空間を個別に分割してセントロイドを求めるより、空間全体を表現するセントロイドを選択して共通に使用することが各曲片を効率的に表現可能であることを示唆していると考えられる。このことに関しては、今後詳細な分析をする予定である。

表 2 中の提案手法 (*EMD2*) と従来手法 (*AFP*) の比較より、セントロイド数を最適に選択すれば提案手法は従来手法より全てのビットレートで高精度に検索が可能であるといえる。特に、*AFP* は音楽の劣化が大きい (ビットレートが 32kbps) 場合、検索率が著しく低下するが、提案手法である *EMD2* はビットレートの変動に対する検索精度の変化は少なく、高い検索精度を維持していることがわかる。さらに、特徴量としてオーディオ指紋を用いた場合、検索位置 (ウインド開始位置) が入力データと検索対象において異なる場合と検索が困難になるという問題が生じる。提案手法は曲片の特徴量をクラスタリングするため時間的な制

表 3 実験結果 ([D60]) (検索正解率 (%))

Table 3 Experimental results in test set D60 (Correct rate in %)

手法	セント ロイド数	ビットレート (kbps)				
		32	64	128	192	NoComp
<i>CentDist</i>	64	24.73	29.09	34.18	30.18	31.27
<i>EMD</i>	32	27.27	27.27	30.18	30.55	32.00
<i>EMD2</i>	32	42.55	44.00	42.91	43.27	42.91

表 4 実験結果 ([D60]) (曲名正解精度 (%))

Table 4 Experimental results in test set D60 (Music title correct rate in %)

手法	セント ロイド数	ビットレート (kbps)				
		32	64	128	192	NoComp
<i>CentDist</i>	64	31.64	32.73	32.36	32.00	32.00
<i>EMD</i>	32	66.55	70.18	69.45	71.27	72.00
<i>EMD2</i>	32	87.64	90.18	89.82	89.82	89.45

約がなくなり、検索位置の少しの異なりなら検索可能であるという利点も存在し、提案手法が有効な音楽検索手法といえる。

3.2.2 切出し位置による影響

検索対象データと異なる位置より検索が開始されたことを想定し、スライドウインド方式のウインド開始位置を変更させた実験 (評価セット [D60]) を行った。実験条件でも述べた通り、スライドウインド方式のウインドの幅は検索対象データと同一であるが、ウインド開始位置を 15 秒遅くした。本実験では、前節の実験結果において高い検索正解率を示したセントロイド数 (*CentDist* は 64, *EMD*, *EMD2* は 32) の結果を用いた。検索結果を表 3 に示す。表中の表記は表 2 と同様であり、NonComp は評価データに対し MP3 による圧縮、伸長を行わない結果を示す。

表 3 より、本実験においても圧縮率による検索率の変動はあまりみられなかった。また、開始位置を変更はいずれの手法でも検索精度が著しく劣化することがわかる。これはウインド開始位置の変更により評価データが検索対象内の複数の曲片にまたがることとなり、正確に曲片を検索できなかったことが原因と考えられる。

正確に曲片が検索されていなくても曲名が正解の場合に正解とする曲名正解率で評価を行った。各手法の曲名正解率を表 4 に示す。この表より、距離計算に *EMD* を用いる手法 (*EMD*, *EMD2*) は曲名正解率が高いことがわかる。*EMD*, *EMD2* は正確な曲片が検

表 5 実験結果 ([E30]) (検索正解率 (%))

Table 5 Experimental results in test set E30 (Correct rate in %)

手法	セント ロイド数	ビットレート (kbps)				
		32	64	128	192	NoComp
<i>CentDist</i>	64	39.53	49.01	50.20	45.85	48.22
<i>EMD</i>	32	50.99	52.17	50.59	50.59	51.38
<i>EMD2</i>	32	49.01	49.41	49.80	50.20	50.59

表 6 実験結果 ([E30]) (曲名正解率 (%))

Table 6 Experimental results in test set E30 (Music title correct rate in %)

手法	セント ロイド数	ビットレート (kbps)				
		32	64	128	192	NoComp
<i>CentDist</i>	64	68.38	85.77	83.79	84.19	82.61
<i>EMD</i>	32	86.56	90.91	88.93	89.33	88.93
<i>EMD2</i>	32	90.51	89.72	90.12	89.72	90.12

索できなくても同一の曲内の曲片を検索していることがわかる。

これらの結果は表 2 と比較すると検索率の低下がみられる*1。しかし、ウインド開始位置が検索対象と異なる (検索開始位置が異なる) 場合においても、提案手法は曲検索が可能であり、柔軟な検索ができるといえる。

3.2.3 ウインド幅による影響

検索のための入力音楽データが短い場合を想定し、スライドウインド方式の窓幅を変更した評価実験 (評価セット [E30]) を行った。実験条件でも述べた通り、ウインドの開始位置は検索対象データと同一であるが、ウインド幅を半分 (30 秒) とした。結果を表 5 に示す。本実験では、入力曲片とウインド開始位置が同一である曲片を正解として評価をした。

表より、各手法ともウインド開始位置を変更した場合と同様に著しく検索率が低下していることがわかる。実験に用いた検索対象データは、隣り合う曲片に 30 秒の重なりが生じるようにスライドウインド方式にて曲を分割している。本実験の評価データはこの重なり部分を検索することとなっており、これが原因で正解とした曲片が検索できなかったと考えられる。

そこで、3.2.2 と同様に曲名正解率を表 6 に示す。この結果と表 2 と比較により、提案手

*1 表 2 における曲名検索率は曲片正解率とほぼ同等であった。

法は検索する曲長の変動による検索精度に大きな劣化はみられないことがわかる。また、圧縮率による検索結果の変動も少ないことがわかる。

4. ま と め

本稿では、フィルタバンク特徴量と EMD (Earth Mover's Distance) を用いた音楽検索手法を提案した。提案手法は、スライドウィンド方式により音楽データを曲片に分割し、各曲片から音響特徴量である MFCC を抽出する。抽出した MFCC に対しクラスタリングを行い、各クラスタのセントロイドおよびクラスタに属する MFCC の個数を求め、曲片の特徴量とした。曲片間の特徴量を比較する距離尺度として、異なるヒストグラム間の距離を人間の感覚に近い距離尺度で定義する EMD を用いた。また、曲片間の特徴量を比較する距離尺度として、セントロイド間距離を検討した。

提案手法の有効性を検証するため、日本のポップス 780 曲 (曲片数: 5,135) を用いた音楽検索実験を行った。実験結果より、提案手法は従来のオーディオ指紋を特徴量とした手法より高い検索精度を示すことがわかった。提案手法において、検索対象となる曲全体から作成したセントロイドを各曲片のクラスタリングに用い、集約個数の多いセントロイドを選択して特徴量とすることにより、高精度に音楽検索が可能であることがわかった。また、提案手法は曲片に分割をする際の分割開始位置や分割の長さが検索対象音楽データと入力音楽データで異なった場合においても、高い検索正解率を保つことができた。

本稿では、評価データとして検索対象に対し MP3 による圧縮・伸長を行ったデータを用いた。そのため、リズムなどは同一である。今後は、リズムの変更や録音再生などによる実環境を想定した評価データを用いた提案手法の検証をする予定である。また、本実験では音響特徴量として静的な MFCC のみを用いた。音声認識においては、動的特徴量である MFCC や MFCC が有効なことが知られている。そこで、今後は動的特徴量も含めた特徴量の検討をする予定である。最後に、本実験は小規模な実験であったため大規模な評価実験を行い提案手法の評価をおこなうとともに検索速度を高速化することを検討する。

謝辞 本研究の一部は文部科学省科学研究費、若手研究 (B)19700172、基盤研究 (B) 21300036 の補助を受け行った。

参 考 文 献

- 1) ISMIR; the international society for music information retrieval. <http://www.ismir.net/>.
- 2) 後藤真孝, 平田圭二. 音楽情報処理の最近の研究. 日本音響学会誌, Vol.60, No.11, pp.

- 675–681, 2004.
- 3) J.Haitsma and T.Kalker. A highly robust audio fingerprinting system. *Proc. ISMIR 2002*, pp. 106–115, 2002.
- 4) M.Park, H-R. Kim, Y.M. RO, and M.Kim. Frequency filtering for a highly robust audio fingerprinting scheme in a real-noise environment. *IEICE Transaction INF. & SYST.*, Vol. E89-D, No.7, 2006.
- 5) P.Cano, E.Battle, T.Kallker, and J.Haitsma. A review of algorithms for audio fingerprinting. *Proc. of international Workshop on Multimedia Signal Processing*, pp. 169–173, 2002.
- 6) J.S. Seo, M.Jin, S.Lee, D.Jang, S.Lee, and C.D.Yoo. Audio fingerprinting based on normalized spectral subband moments. *IEEE Signal Processing Letters*, Vol.13, No.4, pp. 209–213, 2006.
- 7) J.T. Foote. Content-based retrieval and audio. *Proc. SPIE Multimedia Storage and Archiving System II*, Vol. 3229, pp. 138–147, 1997.
- 8) Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance, multi-dimensional scaling, and color-based image retrieval. *Proc. of the ARPA Image Understanding Workshop*, pp. 661–668, 1999.
- 9) 獅々堀正幹, 大西泰代, 柘植覚, 北研二. Earth mover's distance を用いたハミングによる類似音楽 検索手法. 情報処理学会論文誌, Vol.48, No.1, 2007.