

Earth Mover's Distance を用いた ハミングによる類似音楽検索の改良手法

大西 泰代[†] 獅々堀正幹^{††} 柘植 覚^{††} 北 研二^{†††}

[†] 徳島大学大学院工学研究科

^{††} 徳島大学大学院ソシオテクノサイエンス研究部

^{†††} 徳島大学高度情報化基盤センター

E-mail: †{ohnisi, bori, tsuge, kita}@tokushima-u.ac.jp

あらまし 近年、楽曲配信サービスの普及により、容易に音楽データをダウンロードして試聴できるようになった。しかし、サーバ側で蓄積している音楽データが膨大になるにつれ、音楽データに対する効率のよい検索手法が必要になっている。この問題に対して、我々はハミングを入力キーとし、距離尺度に Earth Mover's Distance (EMD) を用いた類似音楽検索手法を提案した。この手法は音長、音符の出現時間、音高差を特徴量として用い、輸送問題の解に基づき EMD を計算することにより、高精度な類似音楽検索を可能とした。本稿では更なる精度向上を目的として、新たな音楽特徴量、EMD 計算方法の改良手法、および検索結果のマージ方法を提案する。新たな音楽特徴量に関しては、2つ前の音符との音高差情報を特徴量に加える。EMD 計算方法に関しては、従来手法では出現時間が大きく異なる音符間に分配フローが生成され、誤ったメロディが検索されていた点に着目し、トータルコストを計算する際、音符間の分配フローに制約を加える改良手法を提案する。また、検索結果のマージ方法に関しては、各部分ハミング片の検索結果を曲毎にマージした後、EMD の値に応じて全体の検索結果をマージ手法を提案する。提案手法の有効性を確認するため、約 500 曲の音楽データベースに対し、実際のハミングデータ 50 曲を入力とした評価実験を行った。結果より、検索結果 5 位以内に正解データが出現する割合が従来の 86% から 94% に向上し、全ての入力ハミングの正解メロディの出現順位を向上したことがわかった。

キーワード 類似音楽検索, Earth Mover's Distance, ハミング入力, MIDI

Improved Similar Music Retrieval for the Query-by-Humming Using the Earth Mover's Distance

YASUYO OHNISHI[†], MASAMI SHISHIBORI^{††}, SATORU TSUGE^{††}, and KENJI KITA^{†††}

[†] Graduate School of Engineering, Tokushima University

^{††} Institute of Technology and Science, Tokushima University

^{†††} Center for Advanced Information Technology, Tokushima University

E-mail: †{ohnisi, bori, tsuge, kita}@tokushima-u.ac.jp

Abstract Music retrieval systems are extremely useful for collecting digital music data from on-line music distribution sites. Especially, there is a great need to develop effective techniques for content-based music retrieval systems, which can retrieve by humming query. The main issues in this research is how to decide the similarity of each music features extracted from music data. We have proposed a similar music retrieval method based on humming query using the Earth Mover's Distance as the distance measure. In this paper, for more improving the retrieval precision, we propose a novel music feature parameter and a improvement of the EMD calculation. In addition, we propose a new method of merging the retrieved results. Experimental results show that the proposed method can achieve 8 % (from 86% to 94%) improvement of the retrieval precision compared to the previous system.

Key words Music retrieval systems, Earth Mover's Distance, Query-by-Humming, MIDI

1. はじめに

近年、インターネットの普及に伴い、楽曲配信サービスを行うサイトも多数出現し、ユーザは容易に様々な音楽データをダウンロードできるようになった。しかし、サイト利用時に膨大な音楽データから必要なデータのみを見つけ出すことが困難となり、音楽データに対する効率の良い検索手法が必要になっている [1]。音楽データに対する検索としては、曲名、歌手名や歌詞等を入力とするキーワード型の検索方式が一般的であるが、より汎用的にメロディやリズム等を入力とするコンテンツ型の検索方式が注目されている [2]~[8]。

我々は、コンテンツ型検索方式の中でも、特にハミングを入力として MIDI 形式の音楽データを検索するシステムを開発している。MIDI 音楽データを対象にした従来のコンテンツ型音楽検索手法としては、音楽データのリズムや音高の変化を特徴ベクトル化し、特徴ベクトル間の類似性をユークリッド距離を用いて判定する手法 [3] [4] や、音楽データの音長、音高の変化の類似性を判定する DP マッチング手法 [5], [6], [9]~[13] が提案されているが、我々は、距離尺度に Earth Mover's Distance (EMD) [14] を用いた類似音楽検索手法 [15] を提案し、高精度の検索を可能とした。EMD は線形計画問題の一つである輸送問題における輸送コストの最適解を求めるアルゴリズムである。

提案した類似音楽検索システムは、音楽データを固定長のメロディ片に分割した後、メロディ片内の各音符を供給地、各音符の音長を供給地の資源量とみなし、各音符の出現時間、音高等の特徴量から輸送コストを算出する。この音楽特徴量を EMD で距離計算することで、リズムと音程との類似度を同じ距離尺度で計り、全体の曲調が類似したメロディを検索する。

本稿では、更なる精度向上を目的として、新たな音楽特徴量、EMD 計算方法の改良手法、および検索結果のマージ方法を提案する。新たな音楽特徴量に関しては、2つ前の音符との音高差情報を特徴量に加える。EMD 計算方法に関しては、従来手法では出現時間が大きく異なる音符間に分配フローが生成され、誤ったメロディ片が検索されていた点に着目し、トータルコストを計算する際、音符間の分配フローに制約を加える改良手法を提案する。また、検索結果のマージ方法に関しては、入力ハミングを分割してできる複数の部分ハミング片の検索結果において、正解曲のメロディ片が多く出現していることに着目し、検索結果を曲毎にマージした後、EMD の値に応じて全体の検索結果をマージする手法を提案する。

以下、まず 2. において EMD を用いた類似音楽検索手法について説明すると共に従来手法の問題点を明らかにする。次に、3. で精度向上のための改良手法を提案する。更に、本改良手法の有効性を実証する実験結果を 4. で示し、最後に 5. でまとめと今後の課題にふれる。

2. EMD を用いた類似音楽検索手法

2.1 システムの概要

類似音楽検索手法は、ユーザが検索したい曲の特徴 (本研究では曲の一部をハミングする) を入力し、それに類似した楽曲

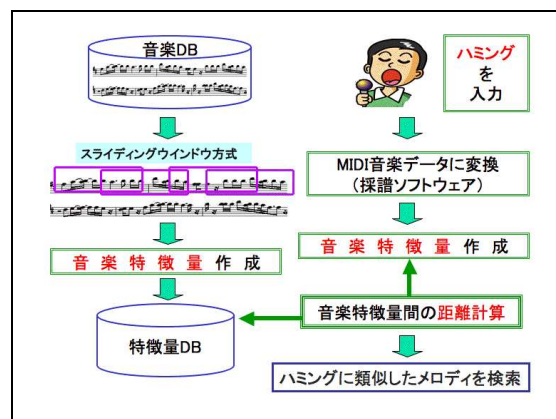


図 1 ハミングによる類似音楽検索システム

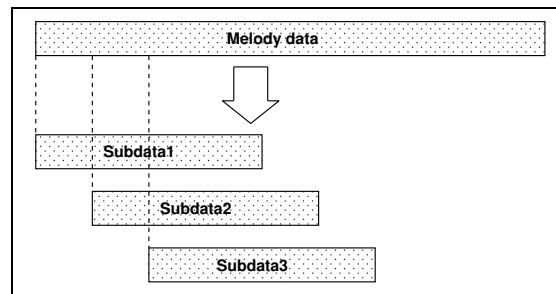


図 2 スライディング・ウィンドウ方式によるメロディ分割

を探すものである。図 1 は、類似音楽検索の全体の流れである。ユーザは検索したい曲の一部をハミングする。その後ハミングを採譜ソフトで MIDI 音楽データに変換し、システムの入力とする。検索対象の音楽データに対しては、予め抽出した主旋律に対する音楽特徴量データベースを作成し、入力ハミングの音楽特徴量との距離計算を行い、距離の近い類似したメロディを持つ曲を検索する。特に、音楽データの任意の部分を検索可能にするため、検索対象の音楽データを予めスライディング・ウィンドウ方式によって分割しておく。スライディング・ウィンドウ方式とは、図 2 のように、定めたウィンドウ長を「拍」を単位にある幅ずつスライドして分割する方法である。スライド幅をウィンドウ長より短くすることで、連続する分割されたメロディ片は互いに重なりのあるデータとなり、検索する部分に関して自由度が増す。同様に入力データも、スライディング・ウィンドウ方式で分割する。

このように、曲のデータを細かく分割して、それぞれを類似度演算の対象とすることで、類似音楽検索システムを構築することができる。また、データベース内の各データが保持している情報量と、メロディ片から作られるデータが保持している情報量は「拍」を単位に等しくなるので、効率的な類似度演算が可能となる。提案した類似音楽検索システムの類似度演算には、EMD を用いている。

2.2 EMD とは

EMD は線型計画問題の一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。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。

$$d_{ij} = \sqrt{\sum_{k=1}^n (p_{ik} - q_{jk})^2} \quad (1)$$

本研究では, 輸送距離として各ベクトル要素 p_i と q_j の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。

$$\text{WORK}(P, Q) = \sum_{i=1}^m \sum_{j=1}^n d_{ij} f_{ij} \quad (2)$$

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 (3)~式 (6)) を満たすものとする。

- 制約条件: 供給地から需要地の一方向にしか輸送されない

$$f_{ij} \geq 0, \quad (1 \leq i \leq m, 1 \leq j \leq n) \quad (3)$$

- 制約条件: 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

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

- 制約条件: 需要地 j が受け取れる容量は w_{q_j} は以下であること

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

- 制約条件: 供給地から移動する輸送量 (総フロー)

$$\sum_{i=1}^m \sum_{j=1}^n f_{ij} = \min \left(\sum_{i=1}^m w_{p_i}, \sum_{j=1}^n w_{q_j} \right) \quad (6)$$

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。

$$\text{EMD}(P, Q) = \frac{\text{WORK}(P, Q)}{\sum_{i=1}^m \sum_{j=1}^n f_{ij}} \quad (7)$$

EMD の計算処理の例を図 3 に示す。図 3 において 2 台のトラックを供給地, 3 個の \square を需要地として, 各々の図で示す荷物が割り当てられる。供給量, 需要量の総和はどちらも 10 個で同数である。供給地から需要地への経路は矢印の数字で示され, 輸送経路を実線矢印で示して。また, それぞれの輸送コストを示している。図 3 での EMD の計算は図 3 の下方で求まる。

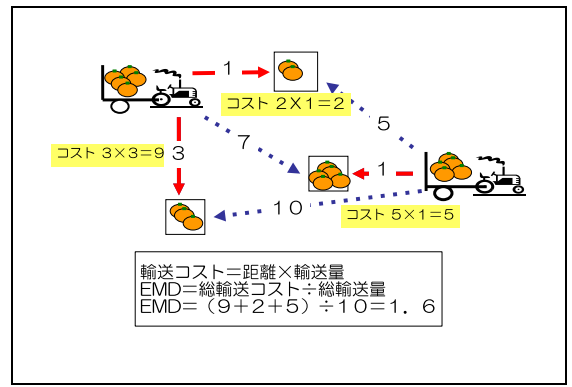


図 3 EMD の計算例

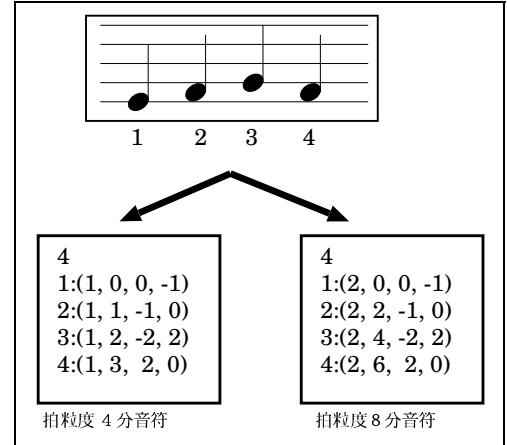


図 4 EMD に適用する音楽特徴量生成例

2.3 EMD に用いる音楽特徴量

EMD を類似音楽検索手法に適用するにあたり, スライディングウィンドウ方式 [3] により分割されたメロディ片毎に特徴量を作成する。単位輸送コストを求めめるための特徴量は, 各音符の Note-on 開始時刻, 前音との音高差, 音高推移特徴とした。図 4 に 4 分音符を 1 拍とした特徴量生成例と 8 分音符を 1 拍とした特徴量の例を示す。開始時間と音長は拍数で表現し, 音高差は半音高い音を 1, 半音低い音を -1 とする。

1 行目はメロディ片内の音符の数を表記し, 2 行目以降は, 出現音符に対する特徴量を表している。4 分音符を 1 拍とした場合, 第 1 音は, (音長, 出現時間, 音高差, 音高推移) = (1, 0, 0, -1) となり, 8 分音符を 1 拍とした場合は, (音長, 出現時間, 音高差, 音高推移) = (2, 0, 0, -1) となる。このように, 個々の音符に対して音長情報も考慮した特徴量が生成される。

抽出した特徴量を EMD の定義に当てはめると, 音長を供給量 w_{p_i} (需要量 w_{q_j}), 開始時間, 音高差, 音高推移を供給地 (需要地) を表す特徴ベクトル $p_i(q_j)$ となる。メロディ間の距離 EMD は, 個々の音符を供給地, 需要地と定義し, EMD の定義式により計算する。

また, 検索システムの高速化を図る目的として用いた音楽特徴量は, 図 5 に示すように, ウィンドウ長は変えずに偶数個めの音符 (スキップ音符) の特徴量を削除し, 重みだけを前の音符 (基準音符) の重みに付加する方法である。このときの特徴量としては, (音長, 出現時間, 音高差, 音高推移) に加えて, 2 つ

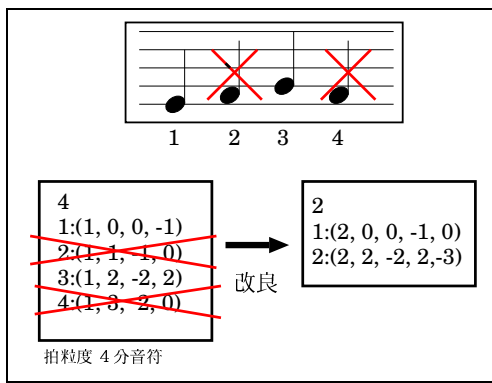


図 5 音符削除と特徴量の高速化方法

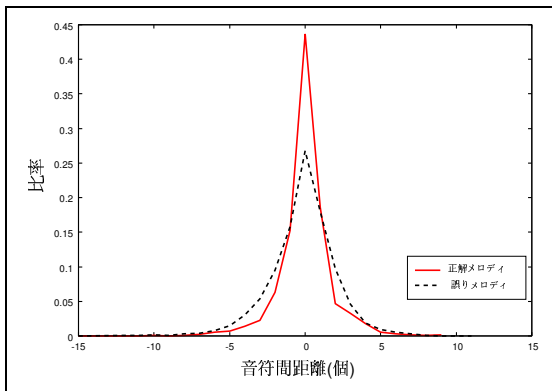


図 6 分配フロー毎の音符間距離の分布

前の音符との音高差 (基準音符間の音高差) を特徴量に加えることにより、スキップ音符の特徴を補うことができると考えた。図 5 のメロディでは、改良後の音符は基準音符だけの特徴が残り、1:(2, 0, 0, -1, 0), 2:(2, 2, -2, 2, -3) となるのでスキップ音符の要素が削減できている。また、スキップ音符の音長を重みとして、基準音符の音長に加えているため、重みの和は一定である。

2.4 フロー生成時の問題点

前述した類似音楽検索システムを評価した実験結果 [15] を部分ハミング片の検索結果毎に分析した。分析により、EMD を計算する際、音符の順番が大きく異なる分配フローが誤検索の原因となっていた。音符の並びについては、輸送コストを計算する際、出現時間を音楽特徴量に加えることにより、音符の順番の近い音符へ重みが分配できるようにしている。しかし、その他の特徴量 (音高差や音高推移) の影響を強く受けた場合には、音符の順番が大きく異なるフローが生成されてしまい、部分ハミング片と誤り曲のメロディ片間の EMD が正解曲のメロディ片よりも小さくなり誤検索となった。

図 6 は、全部分ハミング片 (884 個) の検索結果上位 1 位について正解曲のメロディ片が検索できた場合と、誤検索となった場合のメロディ片において分配フローが生成された音符間の出現順位の差 (音符間距離) の分布を示す。

この図より、正解曲のメロディ片に対しては音符の順番の近い音符へ分配フローが決定していることがわかる。これに対して、誤り曲のメロディ片に対しては音符の順番の遠い音符への分配フローの割合が大きいことがわかる。この結果より、音符

の並びを重視した分配フロー決定の手法が必要であると考えられる。

2.5 検索結果マージ方法

文献 [15] で提案したシステムの検索方法は、部分ハミング片 OR 検索方法 [4] を用いる。この検索方法では、入力ハミング h をスライディング・ウィンドウ方式で m 個に分割した部分ハミング片を h_j 、データベースの曲 a を同じくスライディングウィンドウ方式で n 個に分割した部分メロディ片を a_i とした場合、曲 a に対する距離は、

$$D(a, h) = \min_{1 \leq i \leq n, 1 \leq j \leq m} \{EMD(a_i, h_j)\} \quad (8)$$

で計算する。その他の曲に対しても同様に距離を計算し、その距離を用いて曲ごとの類似度を計算し、曲ごとの順位付けを行う。類似度は距離の逆数として計算した。

この計算方法では、入力ハミングの複数の部分ハミング片のうち一つでもデータベース中の部分メロディ片との EMD が小さければ類似度が高くなり、検索結果の上位にリストアップされる。このように、複数ある部分ハミング片のうち一部分のみにしか着目していない検索方法であり、ハミング片全体を考慮していない。

文献 [15] で行った検索実験で誤検索となった入力ハミングに対し、部分ハミング片ごとの検索結果上位 50 件の正解曲のメロディ片および誤検索となった曲のメロディ片の割合を調査した。この結果、正解曲のメロディ片は検索結果の 52% であり、検索結果 1 位 (誤検索) である曲のメロディ片は 35% であった。これより、検索結果中に多くの正解曲のメロディ片を含んでいたとしても、部分ハミング片の一つでも距離が小さくなると類似度が高くなり、誤検索の原因となることがわかった。よって、部分ハミング片 OR 検索方法ではなく、各曲の出現率も考慮した類似度計算が必要であると考えられる。

3. 精度向上への改良

本章では、2.4、2.5 で述べた問題点を解決し、更なる精度向上を目的とした改良手法を提案する。改良手法として、音楽特徴量の追加、EMD 計算の改良、類似度計算の改良を提案する。

3.1 音楽特徴量の追加

文献 [15] では、音楽特徴として出現時間、音高差特徴量、音高推移特徴量を用いた。更なる精度向上のため、2.3 で述べた、2 つの前の音符との音高差を追加する。

図 7 に、2.3 の図 4 のメロディに対して、出現時間、前音との音高差、音高推移特徴に 2 つ前の音符との音高差を追加した音楽特徴量生成例を示す。8 分音符を 1 拍とした場合、第 1 音は、(音長, 出現時間, 音高差, 音高推移, 2 つ前の音符との音高差) = (2, 0, 0, -1, 0) となる。第 2 音目までは 2 つ前の音符がわからないので特徴量は 0 を与える。第 3 音目の 2 つ前の音との音高差は第 1 音との音高差であるので -3 となる。

3.2 EMD 計算における改良手法の提案

音符の並びを考慮した分配フローの制限による EMD 計算方法として以下の 2 つを提案する。

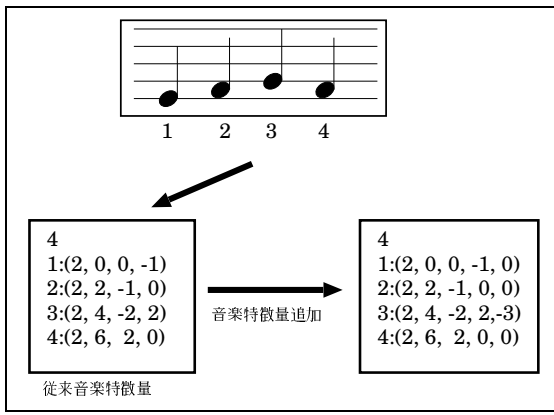


図 7 音楽特徴量追加例

3.2.1 輸送コスト計算内重み付け EMD

本提案手法は、時間軸方向への制約を強める方法であり、音符順番の差の大きい音符間の輸送コストを大きくする方法である。音符間の輸送コストは、式 (1) の音楽特徴ベクトルのユークリッド距離を計算し求める。そこで、各音符の音楽特徴ベクトルの出現時間に重みパラメータを設定し、パラメータを調整することで出現時間により時間軸方向の制約を強め、音符順番の差の大きい音符間の輸送コストを大きくする。これは式 (1) を式 (9) に改良することにより実現する。

$$d_{ij} = \sqrt{\sum_{k=1}^n \alpha_k (p_{ik} - q_{jk})^2} \quad (9)$$

式 (9) での α_k の重みパラメータを各音楽特徴量毎にベクトル内で設定する。

3.2.2 総輸送量計算内重み付け EMD

3.2.1 で提案した手法は分配フロー決定処理内での改良であるのに対して、本節では分配フロー決定後、音符順番の差の大きい音符へフローができていない場合に対して、ペナルティを加える方法を提案する。この改良手法は、まず分配フローを決定し、総輸送量計算時に音符順番の差に応じたペナルティを付与する。音符の順番の差が大きいほどペナルティが大きくなり、結果総輸送量が大きくなる。誤り曲のメロディ片では、音符順番の差の大きい音符へ分配フローが多く生成されているので、この手法により誤り曲のメロディ片との EMD の値を大きくすることができる。一方、正解曲のメロディ片に対しては、音符順番の差の小さい音符へ分配フローが生成されている傾向が強いため、ペナルティが小さくなる。これは、式 (2) で示した総輸送量計算を以下の式に変更することにより実現する。

$$\text{WORK}(P, Q) = \sum_{i=1}^m \sum_{j=1}^n \left((d_{ij} + \beta|i-j|) f_{ij} \right) \quad (10)$$

ここで、 β は、音符順番の差のペナルティ値である。式 (10) では、音符の順番の差が大きくなるに従い総輸送量が大きくなる。図 6 で示したように、誤り曲のメロディ片に対しては、音符順番の差の大きい音符へフローが分配されていることが多く、本改良手法により、これらのフローに対しペナルティを加

えることができる。その結果、誤り曲のメロディ片の順位を下げることができ、正解曲のメロディ片が上位に検索できるようになる。

3.3 類似度計算方法の提案

本節では、2.5 で説明したように検索結果上位のメロディ片内に正解曲のメロディ片が多く含まれていることに着目したマージ方法を提案する。提案するマージ方法は、入力ハミング h をスライディング・ウィンドウ方式で m 個に分割した部分ハミング片を h_i 毎に EMD の値が小さい上位 l 件を検索する。この検索結果に対して、部分ハミング片 h_j の曲 a に対する距離を以下の式で求める。

$$d(a, h_j) = \min_{1 \leq i \leq n} \{ \text{EMD}(a_i, h_j) \} \quad (11)$$

その後、全ての部分ハミング片の曲毎の結果より、曲毎の類似度を式 (12) で計算し、順位付けを行う。

$$\text{Sim}(a, h) = \sum_{j=1}^m \frac{1}{d(a, h_j)} \quad (12)$$

4. 評価

3. で提案した各手法の有効性を検証するために実験を行った。検索対象の音楽データベースとして、童謡、J-pop、演歌等のジャンルが含まれるカラオケ用 MIDI 音楽データ 483 曲を使用した。これら市販の MIDI 音楽データは、特定のチャンネルに主旋律が格納されているため、機械的に主旋律のデータのみを自動抽出した。その後、主旋律のデータに対して、スライディング・ウィンドウ方式を適用して、メロディ片を生成する。

スライディング・ウィンドウの条件としては、8 分音符を 1 拍とし、ウィンドウ長 16 拍、スライド長 4 拍として 84,554 個のメロディ片を生成した。メロディ片内の各音符に対して 8 分音符の長さを 1 とした音長、出現時間、音高差特徴量、音高推移特徴量を用いて音楽特徴量を作成した。

検索入力には、男女 10 名が歌ったハミングを市販の採譜ソフト [16] で MIDI 形式に変換した 50 曲を用いた。採譜方法としては、ハミングの長さが最低でもウィンドウ長を超える条件を義務付け、ハミングの際には、正確なハミングは要求せず、その曲を知っている人が聞いて分かるレベルの入力とした。そのため、入力ハミングにはリズム、音程のずれが生じた。

4.1 検索結果マージ方法の精度評価

部分ハミング片 OR 検索を用いた場合と、今回提案したマージ方法を用いた場合の検索結果を図 8 に示す。この図より、部分ハミング OR 検索と提案手法の検索精度を比較すると、EMD を用いた検索では、5 位までの検索精度が 86% から 2% 向上し、40 位までに全てのハミングに対して正解の曲を全て検索できた。この結果より、EMD を用いた検索では、提案手法が有効であることがわかった。

しかし、DP マッチングにおいては、検索精度の向上はみられなかった。この原因を分析するために、本手法及び DP マッチングにおいて 10 位以内に正解曲を検索できなかった 6 曲のハミングについて、それらの検索結果を調査した。すると DP

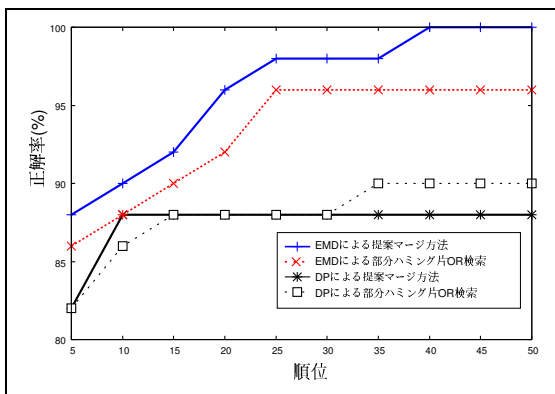


図 8 部分ハミング OR 検索と提案マージ手法の比較

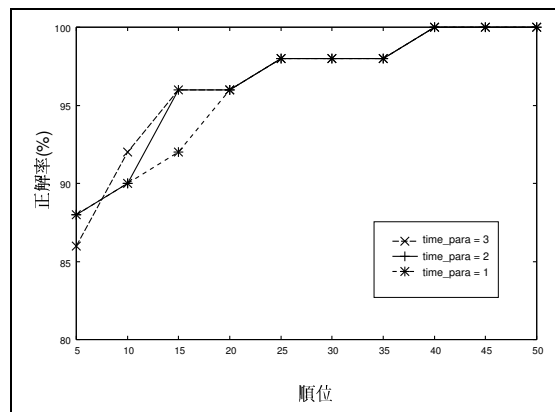


図 10 出現時間パラメータ値 ($\alpha = 1 \sim 3$) の検索精度比較

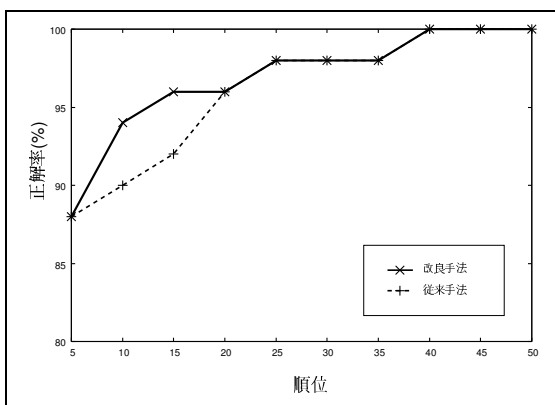


図 9 輸送コスト計算内重み付け EMD ($\alpha = 4$) の精度比較

マッチングを用いた検索では、部分ハミング片の検索の上位に目的のメロディ片が全く検索できておらず、提案した検索結果マージ方法の効果が現れていなかった。特に、これらの部分ハミング片には音程の大きなズレが生じており、DP マッチングが大きな音程のズレに頑健でない [15] ために、検索結果の上位に正解曲が検索できなかったと思われる。一方、EMD を用いた場合には、大きな音程のズレを局所的に含む入力ハミングに対しての上位に目的のメロディ片を検索できていたため、提案したマージ方法を用いると精度が向上した。以下の実験の精度評価は、提案手法のマージ方法を用いる。

4.2 輸送コスト計算内重み付け EMD 改良手法の精度評価

輸送コスト計算内重み付け EMD 改良手法の検索精度評価を行った。図 9 は、輸送コスト計算内重み付け EMD の精度比較グラフである。出現時間のパラメータの値を $\alpha = 4$ とし、音高差特徴量、音高推移特徴量のパラメータの値を $\alpha = 1$ としたときの検索精度である。従来法での各特徴量のパラメータの値は、全て 1 である。この図から、正解データを上位に検索できていることがわかる。しかし、検索件数 20 以上での精度に差はなかった。

出現時間のパラメータ α の値の違いにおける検索精度を検証する。図 10、11 にパラメータの値の違いによる検索精度の変化を示す。図 10 では、出現時間のパラメータ値を大きくすることにより 5 位から 20 位までの検索精度が向上した。2 つの図から共通して、20 位以上の検索精度に変化がみられなかった。20 位に正解曲が検索できなかったハミングは各パラメータとも

表 1 出現時間パラメータ値により順位の変動した曲数

	α				
	2	3	4	5	6
向上	6	6	7	7	7
下落	5	7	9	10	13

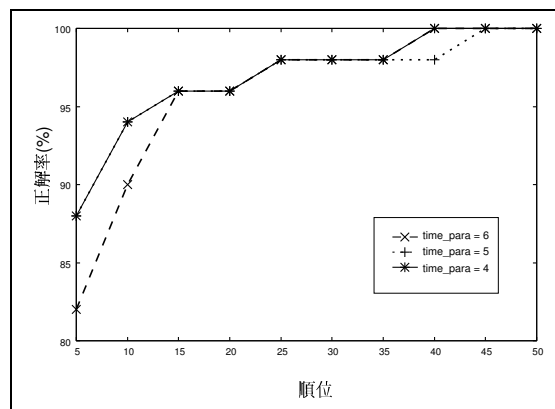


図 11 出現時間パラメータ値 ($\alpha = 4 \sim 6$) の検索精度比較

共通していた。検索できなかったハミング (2 曲) について試聴し正解曲と比較した結果、リズムに大きなズレが生じていた。リズムにズレが生じたハミングに対して、出現時間を重視した提案手法では輸送コスト計算に出現時間が反映できなかったと考えられる。

図 11 より、パラメータ値 6 の時の検索精度が低下している。この原因として、出現時間のパラメータ値が大きくなったことによって、音高差特徴と音高推移特徴が反映されなかったと考えられる。入力ハミングに対して詳細な分析をする。表 1 に従来 ($\alpha = 1$) の入力ハミングの順位を基準に各パラメータ毎の順位の向上曲数と下落曲数を示す。この表より、順位が向上したデータはパラメータ値 4 で頭打ち (50 曲中 7 件) となっており、下落したデータ数はパラメータ値が大きくなると増加している。パラメータの設定には、音高差特徴量と音高推移特徴量も反映できる値を設定する必要があると考えられる。

4.3 総輸送量計算内重み付け EMD 改良手法の精度評価

総輸送量計算内重み付け EMD 改良手法に対する検索精度の評価を行う。図 12 は、総輸送量計算内重み付け EMD の精度比較グラフである。この図では、総輸送量計算内での音符順番

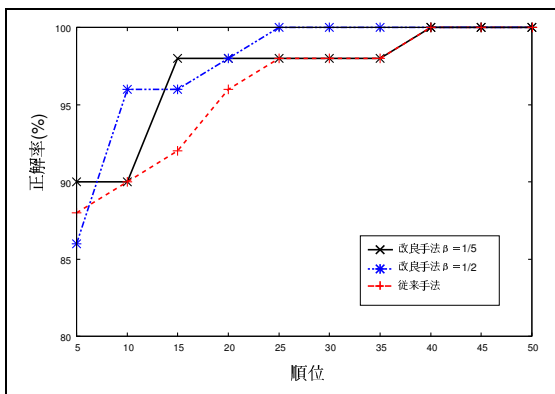


図 12 総輸送量計算内重み付け EMD($\beta = 1/5, \beta = 1/2$) の精度比較

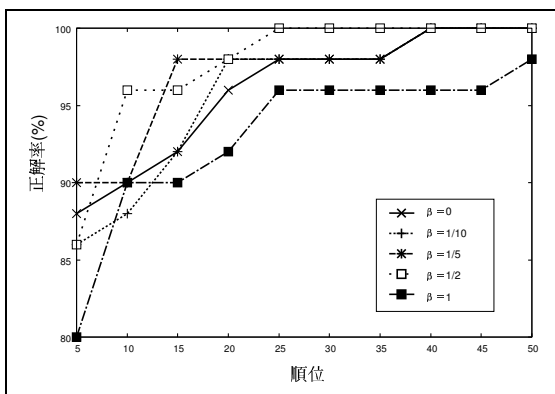


図 13 ペナルティ値 ($\beta = 1/10 \sim 1$) の精度比較

の差に対応したペナルティ β を $1/5$, $1/2$ としたときの精度と従来手法 $\beta = 0$ のときの精度を比較する。この図から、ペナルティ β を与えることによって、正解曲が上位に検索でき、精度が向上している。 $\beta = 1/2$ では、検索件数 25 位までに全ての入力ハミングに対して正解曲が検索できた。

ペナルティ β の値の違いにおける精度を検証する。図 13 は、 β を $1/10$, $1/5$, $1/2$, 1 と変更したときの検索精度を表している。この図より、 β の値を変更すると、検索精度に違いが表れた。 $\beta = 1/10$ では、従来手法 $\beta = 0$ と比較すると検索精度は変化が少なく、これは、ペナルティの値が小さく総輸送量計算の際、ペナルティの影響が小さく検索精度にも影響しなかったと考えられる。また、 $\beta = 1/5$, $\beta = 1/2$ では、従来手法より精度が向上したことがわかる。

しかし、ペナルティを $\beta = 1$ にした際、検索精度は従来手法より低下した。これは、正解曲のメロディ片においても音符の差の大きいところへ分配フローが生成しているため、正解曲のメロディ片に対してもペナルティが大きくなり、順位を下げる結果となった。このとき、上位に上がったデータは、部分ハミング片に含まれる音符数がほぼ等しい誤ったメロディ片が誤検索されていた。

表 2 に従来手法の入力ハミングの順位を基準に各ペナルティ毎の順位の上昇曲数と下落曲数を示す。この表より、ペナルティを大きくすると順位が向上した曲が増え、今回の実験ではペナルティ値 $\beta = 1/2$ のとき、向上曲数が最大となったことがわかる。また、 $\beta = 1$, $\beta = 2$ では下落した曲が向上曲数を上回る結

表 2 総輸送量計算内重み付け手法により順位の変動した曲数

	β								
	1/1000	1/100	1/50	1/25	1/10	1/5	1/2	1	2
向上	0	1	2	2	3	7	9	8	6
下落	0	2	2	4	3	2	3	9	16

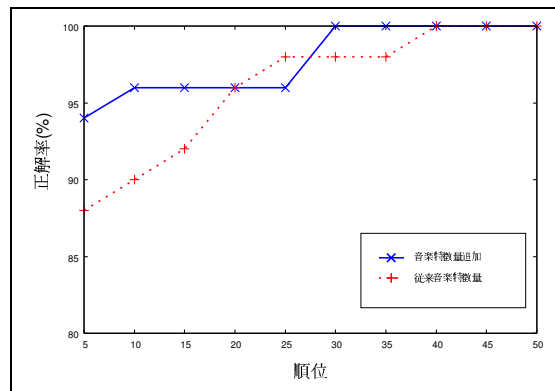


図 14 音楽特徴量追加の効果

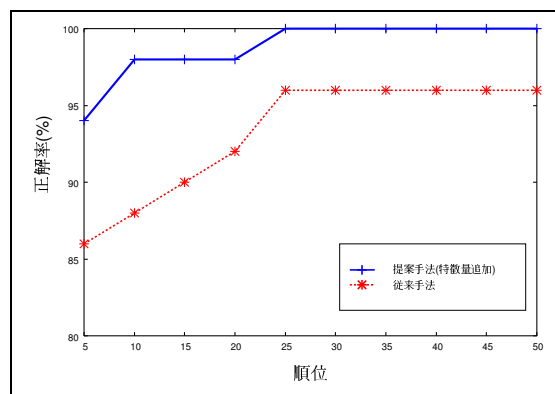


図 15 従来手法と提案手法組合せ検索手法の比較

果となり、図 13 からわかるようにペナルティが大き過ぎると精度低下の原因となる。今回の実験データに対しては、 $\beta = 1/5$, $\beta = 1/2$ がペナルティ値として最適であるといえる。

4.4 音楽特徴追加による精度評価

従来手法で EMD 距離計算に用いた音楽特徴量 (出現時間, 音高差特徴量, 音高推移特徴量) に加えて、2 つ前の音符との音高差特徴量を用いて検索を行った結果の精度を評価する。

図 14 は、2 つ前の音符との音高差を追加した際の検索精度を示している。この図より、検索順位 5 位までの検索精度は従来手法より 6% 向上した。音楽特徴量を追加することにより、正解メロディを正確に特定しやすくなったと考えられる。

4.5 提案手法の総合的な精度評価

図 15 は、本稿で提案した手法を組み合わせ最も精度が高い結果と従来手法の結果である。この図より、提案手法での上位 5 位までの検索精度は、従来手法の検索精度を 86% から 94% へ向上させることに成功していることがわかる。また、今回の実験用いたハミングにおいては、検索順位 25 位までに正解曲を検索できた。この結果より、提案手法の有効性を確認することができた。

図 16 は、楽曲は同じであるが、異なる人間がハミングした

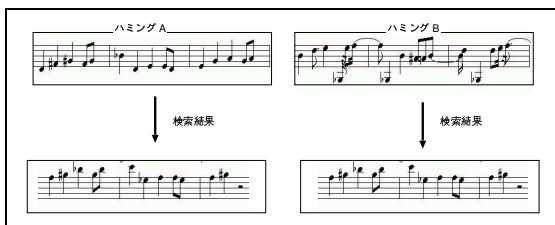


図 16 ハミングの違いによる検索精度

場合の検索結果である。図 16 より、ハミング A、ハミング B は異なる人間が同じメロディをハミングしており、異なる人間がハミングすると音長、音程にばらつきがあることがわかる。この 2 つのハミングを入力とし、DP マッチング、提案手法を用いた検索を行った。DP マッチングを用いた検索では、ハミング A では検索結果 1 位に正解のメロディを検索できたが、ハミング B では検索結果 50 位に正解メロディを検索できず、誤検索となった。提案手法ではハミング A、ハミング B ともに検索結果 1 位に同じ正解メロディが検索できた。

DP マッチングにおいて誤検索となった原因として、ハミング B では音程のズレた音符が複数含まれているため、DP マッチング距離計算内において、音程のズレた音符のペナルティが距離に加算されたと考えられる。このことから、DP マッチングでは正確なハミングに対しては頑健な検索が可能であるが、音程のズレた音符が複数含まれるハミングに対しては検索が困難であるといえる。

一方、EMD を用いた提案手法では、音程のズレにも対応できた検索が可能である。これは、音程のズレた音符の音長を輸送コストが小さくなるように分配することで、音程のズレに対するコストの増加を抑えることができたと考えられる。EMD を用いることにより音長、音程のばらつきを考慮した柔軟な検索が可能であるといえる。

5. ま と め

本稿では、距離尺度として EMD を用いたハミングによる類似音楽検索の改良手法を提案した。改良手法として、EMD 計算内の改良手法と検索結果マージ方法を提案した。実際のハミングデータを用いた評価実験では、従来の EMD を用いた検索手法に比べ検索精度を向上させることができた。

今後の課題としては、今回提案した EMD 計算内のパラメータの値の自動設定方法の提案、輸送コスト計算内 EMD 改良手法と総輸送量計算内 EMD 改良手法の両方を組み合わせた検索手法の実験と検索精度の検証をする予定である。

謝辞

本研究の一部は、科学研究費補助金基盤研究 (B)(17300036)、科学研究費補助金基盤研究 (C)(17500644) を受けて行われた。

文 献

- [1] 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).
- [2] 池谷直紀, 服部正典, 大須賀昭彦: リズム入力による音楽検索方式「タタタタップ」, 第 3 回情報科学技術フォーラム, No. G-021, pp. 391-393 (2004).
- [3] 小杉尚子, 小島明, 片岡良治, 串間和彦: 大規模音楽データベース

スのハミング検索システム, 情報処理学会論文誌, Vol. 43, No. 2, pp. 287-298 (2002).

- [4] 小杉尚子, 櫻井保志, 山室雅司, 串間和彦: SoundCompass: ハミングによる音楽検索システム, 情報処理学会論文誌, Vol. 45, No. 1, pp. 333-345 (2004).
- [5] 園田智也, 後藤真孝, 村岡洋一: WWW 上での歌声による曲検索システム, 電子情報通信学会論文誌, Vol. J82-D-II, No. 4, pp. 721-731 (1999).
- [6] 許盛弼, 鈴木基之, 伊藤彰則, 牧野正三: 複数の音高候補値を用いた楽曲検索システムの構築, 情報処理学会音楽情報科学研究会資料, MUS-49-15, pp. 85-90 (2003).
- [7] Colin, M., William, B.: Johnny Can't Sing: A Comprehensive Error Model for Sung Music Queries, Proc. of 3rd International Symposium on Music Information Retrieval, pp. 124-132 (2002).
- [8] Jyh-Shing, R.J., Hsu, C.L., Lee, H.R.: Continuous HMM and Its Enhancement for Singing/Humming Query Retrieval, Proc. of 6th International Symposium on Music Information Retrieval, pp. 546-551 (2005).
- [9] Dannenberg, R. B., Hu, N.: Understanding Search Performance in Query-By-Humming Systems, Proc. of 5th International Symposium on Music Information Retrieval, pp. 232-237 (2004).
- [10] Ito, A., Heo, S.P., Suzuki, M., Makino, S.: Comparison of Features for DP-Matching Based Query-By-Humming System, Proc. of 5th International Symposium on Music Information Retrieval, pp. 297-302 (2004).
- [11] Parker, C.: Applications of Binary Classification and Adaptive Boosting to the Query-By-Humming Problem, Proc. of 6th International Symposium on Music Information Retrieval, pp. 245-251 (2005).
- [12] Steffen, P.: CubyHum: A Fully Operational Query By Humming System, Proc. of 3rd International Symposium on Music Information Retrieval, pp. 187-196 (2002).
- [13] Sonoda, T., Muraoka, Y.: A WWW-based Melody-Retrieval System - An Indexing Method for A Large Melody Database, Proc. of ICMC 2000, pp. 170-173 (2000).
- [14] Rubner, Y., Tomasi, C and Guibas, L. J.: The earth mover's distance, multi-dimensional scaling, and color-based image retrieval, Proc. of the ARPA Image Understanding Workshop, pp. 661-668 (1999).
- [15] 翻々堀正幹, 大西泰代, 柘植寛, 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, pp.300-311(2007).
- [16] 鼻歌ミュージシャン 2 : 株式会社メディア・ナビゲーション.