

旋律からの単語抽出による文書モデルベースの旋律検索

鈴木 督史^{†1} 黄 宏 軒^{†2} 川 越 恭 二^{†2}

本稿では、文書検索の手法を用いた楽曲検索手法について述べる。提案手法では、楽曲における文書モデルを構築する。ここで、文書モデルとは文書が複数の“単語”によって構成され、また“単語”は複数の“文字”によって構成されるモデルである。本稿では、楽曲情報における“単語”と“文字”をそれぞれ Melody Term, Melody Character と呼ぶ。楽曲の文書モデル適用の為に、まず楽曲の旋律を分割し特徴ベクトルを求める。その特徴ベクトルを分割し、さらにクラスタリングを用いて、あらかじめ定めた項数の Melody Character を構築する。それらの手順によって文書検索の手法を用いた楽曲検索手法が可能となる。また本稿では、楽曲の文書モデル構築による類似楽曲検索システムを実装し、得られた知見について述べる。

Document Model Based Melody Retrieval with Term Constraction from Melody Data

MASAFUMI SUZUKI,^{†1} HUANG HUNG-HSUAN^{†2}
and KYOJI KAWAGOE^{†2}

In this paper, we describe a music retrieval using document retrieval technique. Our method uses the document model for music information. The document model is composed of “term” and “character” combined in a term. In order to apply the document model for music information, we develop several methods, music melody split, music character extraction, music term constraction. The system with the method is developed and it is shown that our system is effective in music simiraly search.

^{†1} 立命館大学理工学研究科

Graduate School of Science and Engineering, Ritsumeikan University

^{†2} 立命館大学情報理工学部

College of Information Science and Technology, Ritsumeikan University

1. はじめに

Web において、マルチメディア情報を扱う機会が増えている。例えば MySpace では、利用者は自分の制作した楽曲を公開することが出来る。また、Youtube では、同様に動画を公開することが可能である。これにより利用者は他の利用者が作成した楽曲や動画といったマルチメディアコンテンツを閲覧することが可能となった。マルチメディア情報を検索する既存 Web サービスは、マルチメディア情報に付与されたメタデータを利用する機会が多い。しかしマルチメディア情報そのものを対象とした内容検索手法について広く研究されている。例えば、楽曲を利用者の入力したハミングで検索システムは多数存在している。⁴⁾¹⁰⁾¹⁾³⁾⁹⁾

マルチメディア検索の中で本稿で取り扱う楽曲検索の手法においては、既存研究で、それぞれ様々なアプローチが為されている。例えば、DP を大きく改良した手法⁵⁾ 等が存在している。このような、マルチメディア情報に対する多様なアプローチはそれぞれに大きな利点があると考えられる。

それぞれのメディアに特化された、マルチメディア検索における技術に対して、文書検索の技術は無視出来ない存在である。Web の発展とともに、文書検索システムは我々の身近なものとなり、広く研究されている。Web 検索エンジンに代表されるキーワード検索エンジンや EC サイト 等で用いられるレコメンデーションの技術など、それらは文書検索にまつわる技術である。ところで、それらの技術は文書の内容を考慮した技術ではなく、統計的な手法である。つまり、マルチメディア情報に対して何らかの処理を行うことによって文書モデルを構築することによって、それらの文書検索における技術を適用することは、マルチメディア検索におけるひとつの解であると考えられるのは不自然ではないと考えられる。

そこで、本稿では楽曲において文書検索の手法を利用可能にする為の手法について提案する。文書検索は古くから研究されており、また近年の Web の発展によって検索エンジンなどで利用されることで広く研究されている。そこで、多くの知見を用いて楽曲検索にまつわるシステムを開発することが可能となるのは開発コストの面から有用であると考えられる。本稿で提案する手法は、楽曲を再構成し、楽曲に対して文書モデルを与えることによって、楽曲に対して文書検索の手法を用いることを可能とする。

まず、文書モデル適用方法について 2 章で詳しく述べる。また、本提案手法を適用した類似楽曲検索システムを実装した。そのシステムの詳細と得られた知見については 3 章で詳しく述べる。

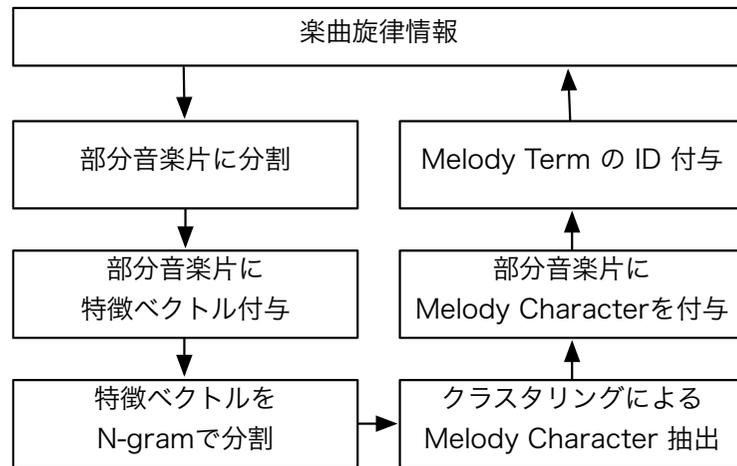


図 1 旋律情報再構成手法
 Fig. 1 Melody Representation Technique.

2. 旋律情報に対する文書モデル適用

本手法では、旋律情報に対して文字と単語から構成される文書のモデルを適用する。文書は、複数の“単語”によって構成され、また“単語”は複数の“文字”によって構成されている。また、本稿では旋律情報を以下のように定義する。

- 音符および休符の連なりを旋律とする
- 音符は音高と音長の属性をもつ
- 休符は休符長の属性をもつ

本手法では、このような旋律情報において、文書モデルを構築する。つまり、旋律情報そのものの情報の他に、“単語”と“文字”のモデルを割り当てる。本手法では、旋律情報に置いて文書の単語に相当するものを Melody Term、また同様に文字に相当するものを Melody Character と呼ぶ。

2.1 概要

本提案手法では、扱う楽曲全てを Melody Term の連続である情報に再構成する。また、Melody Term は、Melody Character の連続である情報となる。本提案手法は、楽曲情報

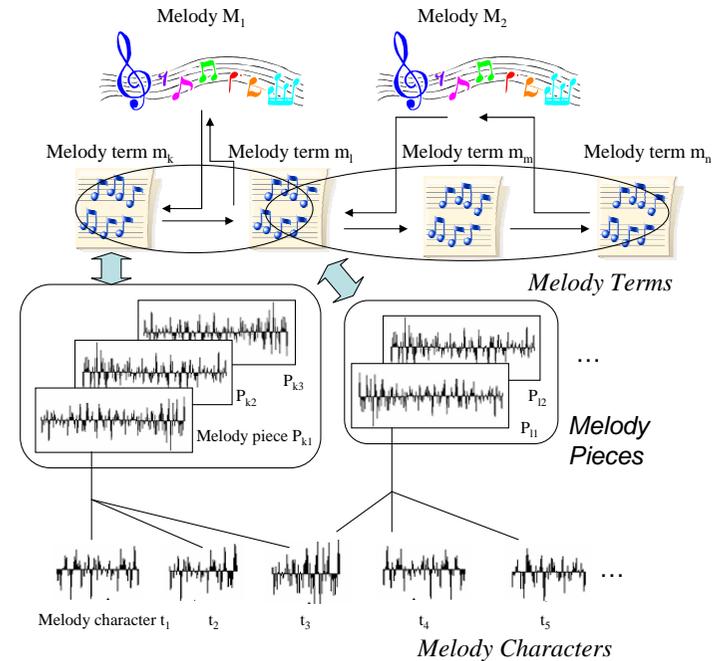


図 2 旋律特徴抽出例
 Fig. 2 Melody representation example.

から、Melody Term と Melody Character を抽出し、楽曲情報に対して文書モデルを適用する。旋律情報再構成手法について 図 1 に示す。

本提案手法では、まず楽曲の旋律情報をあらかじめ単語の単位の大きさに分割する。ここで分割された旋律情報を元に Melody Term を抽出し、最終的に Melody Term の ID をそれぞれの旋律情報に付与することによって、楽曲情報を Melody Term の連続である情報として扱うことが可能となる。旋律分割手法については 2.2 節で詳しく述べる。

文書モデルを適用した旋律情報は、文書が単語の連続であると同様に部分音楽片の連続であると考えることが可能である。しかし、部分旋律片そのものを文書モデルに置く単語、つまり本提案手法における Melody Term として取り扱う場合、Melody Term の項数が無尽蔵に増加する問題が考えられる。

既存研究の実験⁶⁾(図 2)では、100 種の楽曲から特徴量を抽出した結果、約 25 万の項数が

抽出された。従って、この知見に基づくと、楽曲分割により抽出された部分音楽片を Melody Term とする場合、通常の文章に対してかなり膨大な単語の項数が抽出される。ある文書の文字の項数を考えると、英語文書の場合にはアルファベットは 26 文字である。また、日本語文書の場合、ひらがなやカタカナ、漢字の組み合わせ数千の文字位数となる。しかし、この Melody Term において、その項数は数十倍程度となり膨大となる。つまり、部分音楽片そのものを Melody Term として扱う場合、文書のモデルとしてはかなり異なったものとなる。

ここで、本手法では文字数の上限を設定することによってこの問題を解決する。Melody Character の抽出にはクラスタリングを用いる。本手法ではまず、Melody Term に対して特徴ベクトルを求め、それを特徴ベクトルのシーケンスとして処理する。その後、そのシーケンスを分割する。次に、分割された特徴ベクトルをクラスタリングすることで、それぞれの分割された特徴ベクトルに Melody Character の識別子を与える。その結果、Melody Term は、Melody Character のシーケンスとして処理する情報となる。

特徴ベクトル及び特徴ベクトルの抽出方法については 2.3 節で詳しく述べる。

クラスタリングを用いた Melody Character の抽出については 2.4 節で詳しく述べる。

以上の手法を用いることによって、最終的に楽曲情報に対して Melody Term のシーケンスとして処理する情報を構築することが可能となる。

2.2 旋律情報分割手法

旋律情報の分割方法には、楽曲検索や類似楽曲探索において、既にいくつか存在している⁷⁾²⁾⁴⁾³⁾。本提案手法では、Melody Term の抽出にスライディングウィンドウ方式⁴⁾³⁾を利用する。スライディングウィンドウ方式の概略を図 3 に示す。図に示すようにスライディングウィンドウ方式は、楽曲の旋律情報の先頭から一定の長さ（スライド幅）ごとに一定の長さ（ウィンドウ長）に旋律を切り出す。スライディングウィンドウ方式によって分割された旋律それぞれを部分音楽片と呼ぶ。スライドウィンドウ方式では、スライド幅をウィンドウ長より短くすることによって、部分音楽片は互いに重なり合う情報となる。これにより、楽曲検索システム利用者は楽曲の任意の部分問合せとして入力した場合でも楽曲の検索が可能となる。本手法ではこの部分音楽片から Melody Term を抽出し、その Melody Term の ID を部分音楽片に割り当てる。

スライディングウィンドウ方式で分割した、ある楽曲 $o \in M$ は (1) 式のように示すことが可能である。

$$o = (m_{i1}, m_{i2}, \dots, m_{ik}) \quad (1)$$

このとき M は扱う楽曲全体、 $m_{ij} \in ML$ は部分音楽片。また、 ML は扱う楽曲全体にお

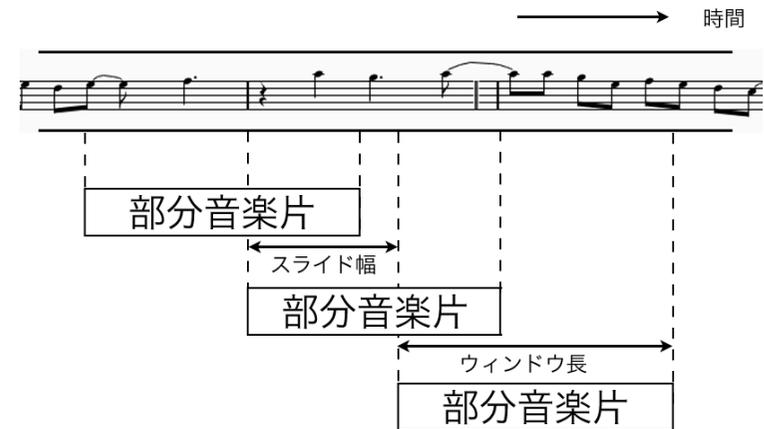


図 3 スライディングウィンドウ方式
Fig. 3 Sliding Window Technique

る部分音楽片の集合である。

2.3 特徴ベクトル

2.2 節で、旋律を Melody Term の単位である部分音楽片に分割する手法について説明した。次にこの部分音楽片に Melody Character の連続である情報を与える必要がある。2.1 節で述べたように、Melody Character 抽出および比較の為には、特徴ベクトルを与える必要がある。本節では特徴ベクトル抽出手法について述べる。

Melody Term の特徴ベクトルとしては、音高推移部分特徴ベクトルと音高差分布ベクトル⁴⁾³⁾を元にした特徴ベクトルを用いる。本手法における特徴ベクトル v_t は (2) 式のような $l \times 2$ 次元ベクトルで示すことができる。

$$v_t = (a_1, \dots, a_i, \dots, a_l) \quad (2)$$

$$a_i = (r_i, n_i) \quad (3)$$

a_i は i 番目の特徴ベクトルの要素で、また l は v_t の長さを示している。また、 a_i は音高差分布ベクトル r_i と音高推移部分ベクトル n_i の 2 次元構造を持っている。まず、音高差分布ベクトルは、音高の推移の分布を表しており、隣り合う音符 m_{ti} と $m_{t(i+1)}$ の拍数を要素とする。また、音高推移部分特徴ベクトルは部分音楽片におけるある特徴を持った要素を選択し、それを先頭として音高の推移を抽出、その音高の差の値をベクトルの値とす

る。ある特徴とは、例えば部分音楽片に含まれる音符の初めの4拍のうち最も音高の高いもの等といったものであり、本提案手法では既存研究⁴⁾において採用されたルールである、“最も音高の高い音符が出現する次元を起点とする”を採用した。従って、 v_t の長さである l は m_t の要素によって決定される。

2.4 クラスタリングを用いた Melody Character 抽出

本節では、2.3節で部分音楽片から抽出した特徴ベクトルを用いて、あらかじめ決めた項数の Melody Character を抽出する。

まず、Melody Character から抽出した特徴ベクトル v_t を分割する。この分割方法は、固定長の N-gram 法²⁾を用いる。そのことによって、用いる特徴ベクトルの次元が常になりベクトル間の類似度の計算が可能となる。

クラスタリングで用いる2つの異なる特徴ベクトル a と b 間のコサイン類似度 $\cos(a, b)$ は、(4)式に示す方法を用いる。

$$\cos(a_i, b_i) = \frac{a_i b_i}{|a_i| |b_i|} \quad (4)$$

$$\cos(a, b) = \frac{\sum_{j=0}^j \cos(a_i, b_i)}{j} \quad (5)$$

a_i および b_i はそれぞれの特徴ベクトルの要素、 j は特徴ベクトルの次元となる。

次に、N-gram 法で分割したそれぞれの要素を用いてクラスタリングを行うことによって、Melody Character の抽出を行う。クラスタリングの手法には k-means 法を用いる。k-means 法は一般的な非階層形クラスタリング方式である。k-means 法は、あらかじめ決めた項数のクラスに要素を分類することが可能である。本提案手法では、k-means 法によって共通のクラスに分類された特徴ベクトルに対して共通の Melody Character を示す識別子を割り当てることで、同様の Melody Character を求める。抽出された Melody Character を元の部分音楽片に割当て、同様に同じ Melody Character の連続情報となった分割音楽片に対して同様の Melody Term を示す識別子を与え、それを Melody Term とする。

以上の手順によって、楽曲情報において文書モデルを構築することが完了する。

3. 旋律への文書モデル適用による類似旋律検索システム

2章で旋律情報を持った楽曲に対して文書モデルを適用することについて述べた。本章では、楽曲情報における文書モデルを用いて、TF-IDF 法⁸⁾を用いた類似旋律検索システム

について述べる。また、類似旋律検索システムを実装して得られた知見について述べる。

3.1 TF-IDF 法

TF-IDF 法とは、文書におけるある単語の出現頻度と、文書集合全体である単語の文書頻度の逆数を組み合わせた値である。ある文書において、TF-IDF 値が高い程、その文書の特徴的な単語であることを示す。本類似旋律検索システムでは、楽曲の特徴ベクトルを求める手法として TF-IDF 法を用いる。TF-IDF 法における、ある楽曲に出現する Melody Term $w_{i,j}$ で求める TF-IDF 値を(6)式に示す。

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right) \quad (6)$$

$tf_{i,j}$ は楽曲 o_i における $w_{i,j}$ の出現頻度、 df_i は $w_{i,j}$ が含まれる楽曲数、 N は全ての楽曲数を示している。楽曲間の類似度の検索にはコサイン類似度を用いる。

3.2 類似旋律検索システム機能

類似旋律検索システムは以下に示す機能により構成されている。

- (1) 楽曲における文書モデル構築
- (2) TF-IDF 法を用いた楽曲への特徴ベクトル算出
- (3) 楽曲間類似度の算出
- (4) 楽曲 DB 入出力
- (5) 利用者による楽曲の入力
- (6) 利用者に対する楽曲リスト表示

3.3 類似旋律検索システム構成

類似旋律検索システムの構成を図4に示す。類似旋律検索システムはあらかじめ楽曲 DB を構築する。楽曲 DB は楽曲の旋律情報、楽曲の Melody Term の構成、TF-IDF 法を用いた特徴ベクトルを格納する。楽曲類似検索を行う為に、まず利用者は楽曲 DB から任意の楽曲を選択する。類似旋律検索システムは利用者によって指定された楽曲を受け取ると、指定された楽曲と楽曲 DB に含まれる他の全ての楽曲との距離を算出する。最後に類似旋律検索システムは、算出された距離を用いて利用者に対して楽曲リストを提示する。

3.4 評価実験

本章では、楽曲情報における文書モデル構築の有用性を調べる為に行った評価実験について述べる。

3.4.1 実験方法

異なる楽曲3曲を用いて被験者3名による主観評価を行った。本実験では、被験者3名

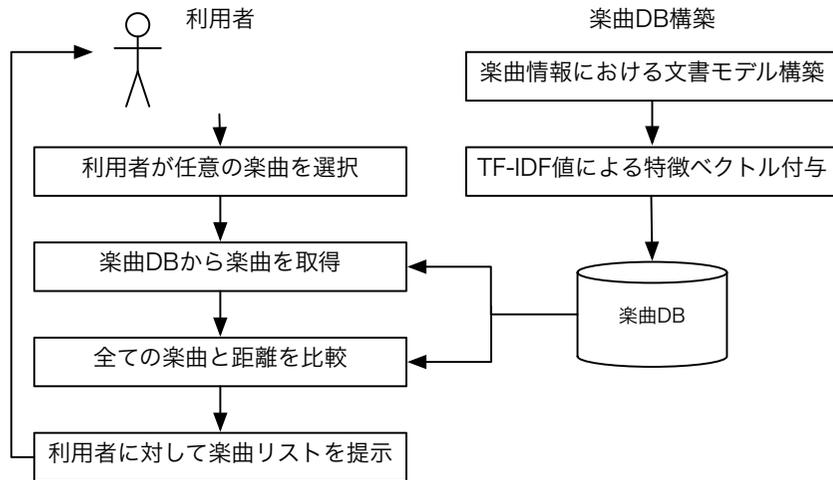


図4 類似旋律検索システム構成
Fig. 4 System Structure.

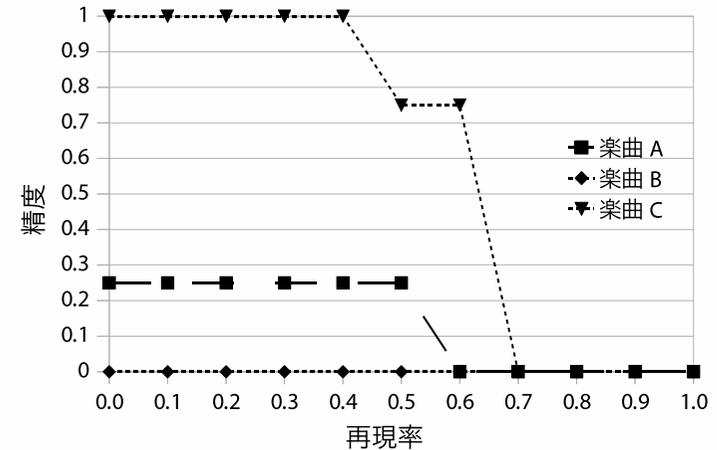


図5 再現率精度曲線
Fig. 5 Precision Recall Graph.

があらかじめ各々に作成した正解集合を用いる。また、被験者によって、正解かどうかの判断が分かれた場合は、多数決で正解であるとした。

3.4.2 実験条件

実験に用いる楽曲は、プロの作曲家が制作した旋律情報と MIR-QBSH-corpus の合計 75 曲を利用した。本実験で問合せとして用いた楽曲は以下のような旋律である。

楽曲 A 童謡風

楽曲 B 民謡風

楽曲 C ポップス風

本類似楽曲システムの特性を調べる為、大きく特徴の異なる楽曲を選択した。

楽曲 DB 作成におけるパラメータは、ウィンドウ長を 16 拍、スライド幅を 4 拍で旋律分割し、Melody Character の項数は 26 である。

3.4.3 結果と考察

実験で得られた結果を図 5 に示す。

図 5 に示すように、楽曲の種類によって結果が大きく異なった。ポップス風の**楽曲 C**では良い検索結果が出力されている。その一方で**楽曲 B**では全く正解が出力されていない。こ

の結果の理由として同じ旋律であっても伴奏や前後関係によって大きく印象が変わることが考えられる。本類似旋律検索システムは、楽曲の旋律の特徴的な部分を重視して類似度比較を行っている。その為、**楽曲 C**のように、印象的な旋律が利用者に対して強い印象を与えるポップスは、本類似旋律システムが得意とする種類の楽曲であると考えられる。一方で**楽曲 A**や**楽曲 B**といった、童謡や民謡などといった楽曲は似たような旋律が多く使われていたり、また同じ旋律であっても、楽曲全体の雰囲気などによって印象が大きく変わる為、本類似楽曲検索システムが不得意とする楽曲であると考えられる。この問題の解決としては、不得意な楽曲においては、Melody Term だけでなく元の楽曲情報も併用することが、有効であると予測される。もとの旋律情報を併用する場合においても、文書モデルを利用したことによって、元の楽曲情報だけを用いた場合と比較してある程度検索対象の絞り込みが可能となり、従って検索速度が向上する。そのため、文書モデルの利用は有用であると考えられる。また、既存研究⁶⁾のように、旋律以外から文書モデルを構築する手法も有用であると考えられる。

4. おわりに

本稿では、楽曲情報において文書モデルを構築することで、楽曲に対して文書検索の手法を用いる手法を提案した。また、本提案手法を適用した類似楽曲検索システムを実装し、得られた知見について述べた。類似楽曲検索システムは、TF-IDF法を用いた楽曲の特徴的な部分を重視した類似比較手法である。本手法では楽曲の旋律を用いて文書モデルを構築したため、ポップスのような印象的な旋律を持つ楽曲を用いる場合、有用な効果を可能に確認できた。一方で、旋律以外の要素、例えば全体の雰囲気を利用者の印象に依存するような民謡などといった楽曲では有用な効果は得られなかった。

これらのことから、楽曲情報において文書モデルを構築する場合、利用者が着目する要素を用いて文書モデルを構築すると有用な効果があると考えられる。つまり本提案手法のような旋律を用いた文書モデルの構築手法においては、ハミングを用いた楽曲検索や、MIDIにおける旋律入力支援など利用者が大きく旋律に着目するような楽曲検索システムを構築において、有効な効果を発揮すると考えられる。今後、文書モデルを用いた類似旋律検索の改善を予定している。

謝辞

本稿における実験で利用した楽曲データの一部は、佐野芳彦先生に作曲していただきました。ここに記して謝意を表します。

参考文献

- 1) Birmingham, W., Dannenberg, R. and Pardo, B.: Query Humming with the Vocalsearch System, *Comm. of ACM*, Vol.49, No.8, pp.49–52 (2006).
- 2) Cui, B., Jagadish, H.V., Ooi, B.C. and Tan, K.-L.: Compacting music signatures for efficient music retrieval, *EDBT '08: Proceedings of the 11th international conference on Extending database technology*, New York, NY, USA, ACM, pp.229–240 (2008).
- 3) et. al., N.K. and Kushima, K.: A Practical Query-Humming System for a Large Music Database, *ACM Multimedia2000*, pp.333–342 (2000).
- 4) KOSUGI, N.: SoundCompass : A Practical Query-by-Humming System-Normalization of Scalable and Shiftable Time-Series Data and Effective Subsequence Generation, *Proc. ACM SIGMOD International Conference on Management of Data, 2004*, pp.881–886 (2004).
- 5) Nishimura, T., Zhang, J.X. and Hashiguchi, H.: Music Signal Spotting Retrieval

- by a Humming Query Using Start Frame Feature Dependent Continuous Dynamic Programming, Continuous Dynamic Programming”, *Proc. 3rd International Symposium on Music Information Retrieval*, pp.211–218 (2001).
- 6) Ono, K., Suzuki, Y. and Kawagoe, K.: A Music Retrieval Method based on Distribution of Feature Segments, *IEEE MIPR 2008*, pp.613–618 (2008).
- 7) Orio, N. and Neve, G.: Experiments on Segmentation Techniques for Music Document Indexing, *ISMIR 2005* (2005).
- 8) Salton, G.(ed.): *Automatic Text Processing*, Addison-Wesley Longman Publishing Co., Inc. (1985).
- 9) Zhou, Y. and Shasha, D.: Warping Indexes with Envelop Transforms for Query by Humming, *ACM SIGMOD 2003*, pp.181–192 (2003).
- 10) 鈴木督史, 川越恭二: スペル修正技術を用いた楽曲検索システム, 情報処理学会研究報告. SIGMUS, 音楽情報, Vol.82, No.2009 (2010).