

VocalFinder: 声質の類似度に基づく楽曲検索システム

藤原 弘 将 後藤 真 孝

産業技術総合研究所

ボーカルの声質の類似度に基づく楽曲検索システムを開発した。本システムは、クエリとして与えられた楽曲と類似した声質を持つ楽曲を予め登録したデータベース中から検索する。本システムを実現するために、伴奏を含む音響信号中から伴奏音の影響を低減させ歌声の特性を表現する特徴ベクトルを抽出する手法と、相互情報量を用いて2つの特徴ベクトル列間の類似度を計算する手法を開発した。本システムを実装し、75曲をデータベースに登録し運用することで、システムが正しく動作することを確認した。さらに、被験者実験の結果、被験者の80%の回答が、従来のMFCCなどを用いた手法と比較して、提案手法によりボーカルの声質が類似した楽曲を検索出来ていると回答した。

VocalFinder: A music information retrieval system based on vocal timbre similarity

HIROMASA FUJIHARA and MASATAKA GOTO

National Institute of Advanced Industrial Science and Technology (AIST)

We developed a music information retrieval system based on singing voice timbre, i.e., a system that can search for songs in a database that have similar vocal timbres. To achieve this, we developed a method for extracting feature vectors that represent characteristics of singing voices and calculating the vocal-timbre similarity between two songs by using a mutual information content of their feature vectors. We operated the system using 75 songs and confirmed that the system worked appropriately. According to the results of a subjective experiment, 80% of subjects judged that compared with a conventional method using MFCC, our method finds more appropriate songs that have similar vocal timbres.

1. はじめに

本稿では、ユーザによって提示されたクエリの楽曲と、ボーカルの声質が類似した曲をデータベース中から検索出来る音楽情報検索システム(VocalFinder)について述べる。既知の曲しか検索出来ない従来の書誌情報による検索に加えて、本システムを使用することで、楽曲の内容を用いて未知の楽曲を検索することが可能になる。

近年、このような内容に基づく音楽検索が重要性を増してきている。携帯型音楽プレーヤやWeb上でのオンライン楽曲販売サービスの普及により、ユーザ達は膨大な量の楽曲から好みの楽曲を検索し、どんな時でもどんな場所でも聴きたい音楽を鑑賞出来るようになった。これにより、自分の好みの楽曲を手がかりに、まだ聴いたことのない楽曲を発見したいという要求が

生じてきた。しかし、探したい対象の楽曲が未知で、「好み」のような曖昧な情報のみが手掛かりの場合、従来の書誌情報に基づく検索システムは無効だった。

このような動向から、内容に基づく音楽検索システムに関して、多くの研究がなされてきた¹⁾⁹⁾。しかし、これらの研究は、音楽の内容を表現するのに、MFCC、スペクトル重心、ロールオフ、フラックスなどの低レベルの特徴量しか用いておらず、ボーカルの声質などの高レベルの特徴量は用いられていなかった。

我々は、ボーカルの声質を手がかりに楽曲を検索可能な音楽情報検索システムを開発した。このシステムを実現するためには、歌声の特質を表現する特徴量を、伴奏を含む多重奏の音楽音響信号から抽出し、それらの特徴量を用いて歌手の声質の類似度を計算出来なければならない。本研究では、歌声の声質を表現する特

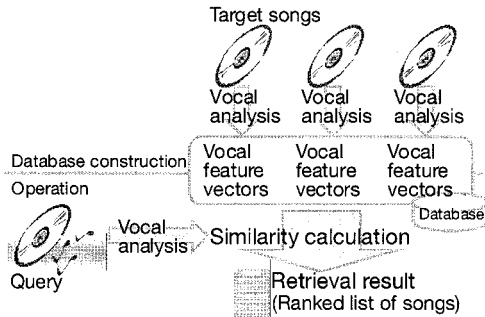


図1 システムの概要

微量として、以前提案した歌手名同定手法¹⁰⁾で用いた特徴量を導入する。また、それらの特徴量間の類似度として、相互情報量を用いる。

2. システムの構成

本システムは、データベースに予め登録された楽曲の中から、ユーザが提示したクエリの楽曲と類似した声質を持つ楽曲を検索し、ユーザに提示する。本システムの概要を図1に示す。本システムは、データベース構築段階(楽曲の音響信号の分析)と実行段階(楽曲の検索)の2段階からなる。データベース構築段階では、検索対象の楽曲はダウンロードされたりCDからリッピングされた後データベースに貯蓄される。そして、システムは各楽曲を分析し、ボーカルの声質を表現する特徴量を抽出する。

実行段階では、ユーザが好みの楽曲をクエリとしてシステムに入力すると、システムはクエリの楽曲を分析し、楽曲のボーカルの声質を表現する特徴量を抽出する。そして、システムはクエリの楽曲と、データベース中のそれぞれの楽曲で、ボーカルの声質の類似度を計算し、高い類似度を持つ楽曲を検索結果として出力する。

3. システムの実装

2節で述べた音楽情報検索システムを実現するためには、ボーカルの声質を表す特徴量の抽出手法と、その特徴量同士の類似度の計算手法を開発する必要がある。

3.1 特徴抽出

ボーカルの声質を表現する特徴ベクトルを計算するために、我々が以前提案した歌手名同定手法における特徴抽出法を使用する¹¹⁾。この手法は、楽曲中に混在する伴奏音の影響を低減することが出来る。そのため、

伴奏音と歌声が混在した状態をそのまま表現してしまうMFCCなどの特徴量と比較して、この特徴ベクトルは歌声の声質をよりよく表現することが出来る。

本特徴抽出手法は、伴奏音抑制、特徴抽出、高信頼度フレーム選択の3つの処理からなる。伴奏音抑制とは、伴奏音の影響を低減させるために、歌声の高調波構造に基づき歌声を分離・再合成する手法である。特徴抽出では、分離された歌声から特徴量を計算する。高信頼度フレーム選択では、抽出された特徴ベクトル列の中から、歌声と信頼できる区間(フレーム)を選択し、歌声を含まない区間や伴奏音の影響を大きく受けている区間を除去する。

3.1.1 伴奏音抑制

伴奏音抑制では、楽曲中メロディを再合成することで、伴奏音の影響を低減させる。本手法は以下の3つの処理からなる。

- (1) ボーカルメロディの基本周波数(F0)を、後藤のPreFEst¹²⁾を用いて推定する。
- (2) 推定されたF0に基づき、ボーカルメロディの高調波構造を抽出する。
- (3) 正弦波重畳モデル¹³⁾に基づき、メロディの音響信号を再合成する。

ボーカルメロディの基本周波数を推定するために、後藤のPreFEst¹²⁾を用いる。PreFEstは、制限された周波数帯域内で、混合音中の最も優勢なF0を推定する手法である。多くの場合、ボーカルメロディラインは中高域で最も優勢な高調波構造であるため、適切に周波数帯域を制限することでPreFEstはメロディのF0を推定することが出来る。

推定されたF0を用いて、基本周波数成分と高調波成分のパワーを抽出する。各高調波成分の抽出では、 r centの誤差を許容し、その範囲内で最大となる周波数成分のパワーを抽出する。時刻(t)における l 次倍音($l = 1, \dots, L$)の周波数 $F_l^{(t)}$ と振幅 $A_l^{(t)}$ は

$$F_l^{(t)} = \underset{F}{\operatorname{argmax}} |S^{(t)}(F)|$$

$$(\bar{F}^{(t)}(1 - 2\frac{r}{1200}) \leq F \leq \bar{F}^{(t)}(1 + 2\frac{r}{1200})), \quad (1)$$

$$A_l^{(t)} = |S^{(t)}(F_l)|, \quad (2)$$

のようになる。ここで、 $S^{(t)}(F)$ は複素スペクトルを表し、 $\bar{F}^{(t)}$ PreFEstによって推定されたF0を表す。本稿の実験では、 r を20に設定した。

最後に、正弦波重畳モデルを用いて、抽出された高

調波構造から $F_l^{(t)}$ and $A_l^{(t)}$ メロディの音響信号を再合成する。周波数が線形に変化するように、位相の変化は 2 次関数を用いて近似する。また、振幅の変化も線形近似する。再合成された音響信号 $s(k)$ は、

$$s(k) = \sum_{l=1}^L s_l(k), \quad (3)$$

$$s_l(k) = \left\{ (A_l^{(t+1)} - A_l^{(t)}) \frac{k}{K} + A_l^{(t)} \right\} \sin(\theta_l(k)), \quad (4)$$

$$\theta_l(k) = \frac{\pi(F_l^{(t+1)} - F_l^{(t)})}{K} k^2 + 2\pi F_l^{(t)} k + \theta_{l,0}, \quad (5)$$

のように表現される。ここで、 k は秒を単位とする時間で、 $k=0$ は時刻 (t) に相当する。また、 K は時刻 t と $t+1$ の間の間隔で、単位は秒である。 $\theta_{l,0}$ は初期位相を表す。

3.1.2 特徴抽出

再合成された音響信号から、以下の 2 種類の特徴量を抽出する。

- LPC メルケプストラム係数 (LPMCC)¹⁴⁾
音響信号の個人性を表す特徴量は、スペクトル包絡に含まれていることが知られている。本研究では、そのような包絡を表現するスペクトル特徴量として LPMCC を使用する。LPMCC が、音楽のモデリングによく使われるメル周波数ケプストラム係数 (MFCC) などと比較して、歌手の個人性をよく表現することは、我々が以前行った歌手名同定の実験で確認された¹⁰⁾。

- $\Delta F0$
歌手の歌い方を表現する特徴量として F0 軌跡の時間変化係数である $\Delta F0$ ¹⁵⁾ を使用する。歌声は、ビブラート、オーバーシュートなど、短時間の時間変動が激しいので、この特徴量は、歌手の特性をよく表現すると期待される。

3.1.3 高信頼度フレーム選択

PreFEst ではメロディは単に最も優勢な F0 として定義されている¹²⁾ ので、再合成された音響信号は歌声が存在する区間では歌声を含んでいるが、間奏部などではその他の楽器の音を含んでいる。さらに、歌声が存在する区間でも伴奏音の音量が大きい場合は、再合成された歌声は大きく歪んでいる場合がある。そのため、ここで得られた特徴ベクトル列は、一部に信頼できない区間を含んでいる可能性がある。高信頼度フレーム選択によって、そのような信頼できない区間を除去

し、信頼できる特徴量のみを用いて類似度計算できるようにする。これを実現するため、歌声混合ガウス分布 (GMM) λ_V と非歌声 GMM λ_N の 2 種類の GMM を導入する。ボーカル GMM λ_V は歌声を含む区間から抽出された特徴量で学習され、非歌声 GMM λ_N は間奏部などから抽出された特徴量で学習される。特徴量 \mathbf{x} が与えられると、歌声/非歌声 GMM の尤度 ($p(\mathbf{x}|\lambda_V)$ と $p(\mathbf{x}|\lambda_N)$) は、この特徴量がどの程度歌声(または歌声以外の音)に近いかを表現する。これらの尤度を用いて、特徴量 \mathbf{x} が信頼できるかどうかを次式で判定する。

$$\log p(\mathbf{x}|\lambda_V) - \log p(\mathbf{x}|\lambda_N) \begin{matrix} \text{reliable} \\ \geq \\ \text{not-reliable} \end{matrix} \eta, \quad (6)$$

ここで η は閾値を表す。

全ての楽曲に対して閾値を定数として定めることは困難である。なぜなら、伴奏の影響が全体に大きい楽曲では選択されるフレームが少なくなりすぎる場合があるからである。そこで本研究では、楽曲全体の特徴量中の 15% の特徴量が選択されるように、楽曲に応じて閾値を自動決定する。

3.2 類似度計算

本研究では、2 つの楽曲間の類似度尺度として相互情報量を使用する。楽曲の特徴ベクトルの分布を GMM を用いてモデル化し、楽曲から抽出された特徴ベクトル列から、EM アルゴリズムを用いて各楽曲に対して GMM のパラメータを推定する。そして、楽曲 X と楽曲 Y の間の類似度 $d_{CE}(X, Y)$ を、以下のように計算する。

$$d_{CE}(X, Y) = \log \prod_i \frac{\mathcal{N}_{GMM}(x_i; \theta_X)}{\mathcal{N}_{GMM}(x_i; \theta_Y)} + \log \prod_j \frac{\mathcal{N}_{GMM}(y_j; \theta_Y)}{\mathcal{N}_{GMM}(y_j; \theta_X)} \quad (7)$$

ここで、 x_i と y_j は、それぞれ楽曲 X、楽曲 Y の特徴ベクトル列を表現する。ただし、これは高信頼度フレーム選択により選択されている。 θ_X と θ_Y はそれぞれ楽曲 X、楽曲 Y の GMM のパラメータで、 $\mathcal{N}_{GMM}(x; \theta)$ は、 θ をパラメータとする GMM の確率密度関数を表す。

4. システムの運用

図 2 は、システムのスクリーンショットである。歌声 GMM と非歌声 GMM の学習のためは、「RWC 研究用音楽データベース:ポピュラー音楽」(RWC-MDB-P-2001)¹⁶⁾ から、25 曲を選択して使用した。システムの検索対象データベース中には、RWC-MDB-P-2001 で歌声/非歌声 GMM の構築に使わなかった 75 曲を登録

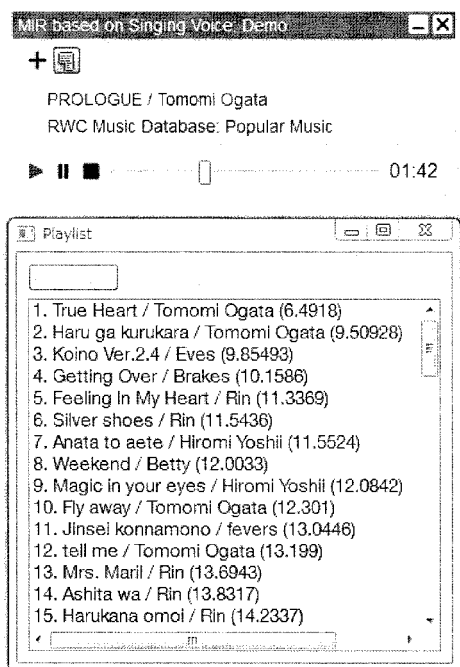


図2 システムのスクリーンショット。

した。

図2の運用例では、女性歌手「緒方智美」が歌う「PROLOGUE」(RWC-MDB-P-2001 No.7)がクエリとして入力されている。クエリの楽曲が与えられ、ユーザが検索ボタンを押すと、類似度の計算に20秒程度の時間がかかった後に、検索結果として楽曲が上位から順に並べられて出力される。図2のように、検索結果として順位、楽曲名、歌手名、類似度の順に表示されている。ユーザが検索結果リスト中の楽曲をダブルクリックすると、楽曲再生ウィンドウにその楽曲が読み込まれ、再生が始まる。

その他の楽曲をクエリとして入力した場合でも、ほとんどの場合で上位の10曲に選ばれた楽曲はクエリの楽曲とボーカルの声質が似ていると感じられた。例えば、図2の例では、上位の21曲はクエリと同様に女性歌手の曲であった。また、図中の上位15曲のボーカルの声質はクエリの楽曲と類似していた。クエリの楽曲の歌手である「緒方智美」の別の楽曲は、上位の1番目、2番目にランクしていた一方で、比較的下位の10番目、12番目にもランクしていた。これは、10番目の曲と12番目の曲に関しては、歌唱スタイルがク

エリの曲や1番目、2番目の曲と異なっていたことが原因である。

一方、検索結果の下位にランクされた楽曲については、順位は楽曲の類似度を正しく反映していないように思われた。例えば、図2の例では、画面上のプレイリストの範囲外で22位より下位では男性ボーカルと女声ボーカルが混在していた。しかし、通常の検索用途では上位の10曲程度を提示すれば十分であり、問題にはならない。キーとなる楽曲が、複数の歌手によって歌われる楽曲や、歌声に深くエフェクトがかけられている曲では、検索結果が直感的な類似度を反映していない場合もあった。

5. 被験者実験

本節では、従来の内容に基づく音楽検索手法であるMFCCを用いた検索システムと比較して、提案するボーカルの声質に基づく検索システムを評価する。比較手法では、特徴量としてMFCCを使用し、特徴抽出の際に伴奏音抑制、高信頼度フレームを行わなかった。類似度の計算は提案法の場合と同様にした。

5.1 実験手順

本実験には、6人の大学生(女性4人、男性2人)が被験者として参加した。被験者は、音楽の専門的な教育は受けていない。まず、被験者は図3のような画面が提示され3つの楽曲を聴くことが指示される。これらの3曲は、クエリの楽曲(図中の曲X)提案法により最上位にランクされた楽曲(図中の曲Aまたは曲B)、比較手法により最上位にランクされた楽曲(図中の曲Bまたは曲A)から構成されている。そして、被験者は曲Aと曲Bのどちらがよりクエリの楽曲に似ているかを判定する。被験者は曲Aと曲Bのどちらの楽曲が提案法による検索結果かは知らされておらず、曲Aと曲Bの順はランダムに決定されている。また、被験者は3つの楽曲を好きな順番で何度でも繰り返し聴くことができる。

RWC研究用音楽データベース:ポピュラー音楽(RWC-MDB-P-2001)中から、ジャンルや歌手の性別が多様になるように留意しながら10曲のクエリを選択した。それぞれのクエリの曲について、被験者に以下の2つの質問をした。

質問1 ボーカルの声質に関して、曲A、曲Bのどちらがより曲Xに似ていると思いますか？

質問2 曲全体の音質に関して、曲A、曲Bのどちら

表 1 実験に使用したクエリの楽曲と対応する検索結果; 3桁の数字は RWC 研究用音楽データベース (RWC-MDB-P-2001) 中の楽曲番号を表す。それぞれのクエリを入力した際の、提案手法と比較手法 (MFCC) による最上位の検索結果が同じ行に記されている。日は日本語を表す、英は英語を表す。

クエリーの楽曲					検索結果 (最上位の楽曲)					
#	歌手名	性別	言語	曲調	手法	#	歌手名	性別	言語	曲調
004	風戸ヒサヨシ	男	日	ブラックソウル	MFCC	031	永山雄一	男	日	ハウス
					提案	082	井口慎也	男	日	ハードロック
010	プレーキーズ	女	日	ダンスミュージック	MFCC	016	吉井弘美	女	日	ソウル
					提案	054	凜	女	日	J-POP
029	西一男	男	日	ブラックソウル	MFCC	017	吉井弘美	女	日	J-POP
					提案	012	西一男	男	日	プリティッシュロック
035	吉井弘美	女	日	メタルロック	MFCC	036	西一男	男	日	メタルロック
					提案	094	Betty	女	英	ポップス
045	森元康介	男	日	ストリートフォーク	MFCC	090	井口慎也	男	英	ロック
					提案	042	森元康介	男	日	シティロック
053	凜	女	日	J-POP	MFCC	062	凜	女	日	J-POP
					提案	014	凜	女	日	J-POP
072	森元康介	男	日	ニューミュージック	MFCC	071	吉井弘美	女	日	ニューミュージック
					提案	076	森元康介	男	日	ストリートフォーク
077	服部まきこ	女	日	J-POP	MFCC	071	吉井弘美	女	日	ニューミュージック
					提案	067	服部まきこ	女	日	J-POP
092	Betty	女	英	ダンスミュージック	MFCC	024	風戸ヒサヨシ	男	日	ダンスミュージック
					提案	086	Betty	女	英	ダンスミュージック
098	Jeff Manning	男	英	リパブルサウンド	MFCC	009	西一男	男	日	プリティッシュロック
					提案	085	Jeff Manning	男	英	モータウン

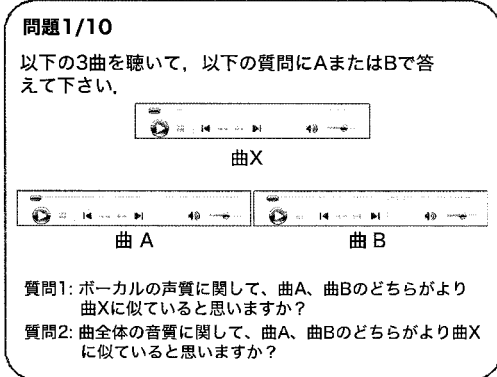


図 3 被験者に提示されたインターフェース。

がより曲 X に似ていると思いますか?

5.2 結果と考察

実験結果を図 4 と図 5 に示す。図は、それぞれの楽曲について、被験者が提案法と比較手法のどちらと回答したかの割合を表す。6人の被験者の10曲に対する回答のうち80%が、提案法の検索結果のボーカルの声質のほうがよりクエリの楽曲と類似していると答えた(4)。一方、曲全体の音質に関しては、70%の回答が比較手法の検索結果がよりクエリに似ていると答えた(5)。これらの実験結果から、提案法により伴

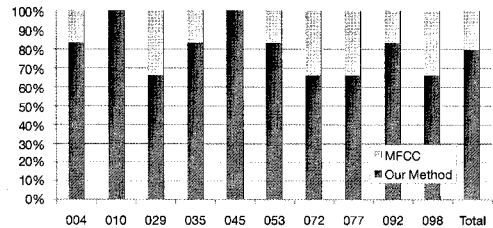


図 4 実験結果: 質問 1: ボーカルの声質。

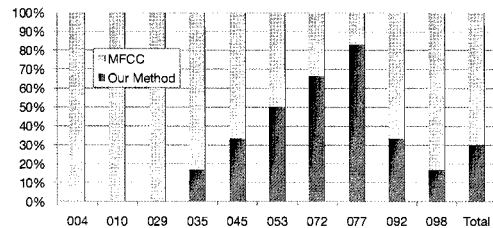


図 5 実験結果: 質問 2: 楽曲全体の音質。

奏音の影響を低減され、ボーカルの声質の類似度に基づいて楽曲が検索出来ることが確認された。実際、表 1 を見ると、MFCC を用いた場合は、10 曲中 4 曲でクエリと異なる性別の楽曲を出力しているが、提案法ではそのような問題は発生していない。

また、提案法の検索結果は、単にボーカルの声質の類似度だけでなく、歌唱スタイルの類似度の観点からも類似している例が見られた。例えば、RWC-MDB-P-2001の#053の楽曲がクエリとして使用された場合、提案法の検索結果と比較手法の検索結果が共にクエリと同じ歌手の楽曲である。ところが、6人の被験者中の5人が、提案手法の検索結果のほうがクエリと類似していると判断した。

6. まとめ

本稿では、ボーカルの声質の基づく類似楽曲検索システムについて述べた。さらに、本システムを実現するために必要な、多重奏の音響信号中から声質に関する特徴量を抽出する手法と、2つの楽曲間の声質の類似度を計算する手法について述べた。本システムを75曲を用いて実際に運用し、声質の類似度に基づいて楽曲を検索するのに有効であることを確かめた。さらに、本システムの検索結果が実際に声質の類似度に基づいていることを確かめるため、6人の被験者により被験者実験を行い、提案法の有効性を確認した。

本研究では類似尺度として、相互情報量を用い、効果的であることを確認した。一方で、相互情報量は計算の際に全ての特徴量を使用するため、計算時間や必要な記憶容量という点で不利である。今後は、Earth Mover's Distance(EMD)⁴⁾など、その他の類似尺度を導入して、計算時間を削減する必要があると考えられる。さらに、本システムをその他の音楽検索手法と統合することで、より柔軟で幅広い検索システムを開発する予定である。

謝辞 本研究の一部は、CRESTの支援を受けた。また、本研究の実験において、「RWC研究用音楽データベース:ポピュラー音楽」(RWC-MDB-P-2001)¹⁶⁾を使用した。最後に、ご討論いただいた齊藤毅氏(産業技術総合研究所)、中野倫靖氏(筑波大学)、並びに被験者実験にご協力いただいた京都大学、京都女子大学、京都造形芸術大学の方々に感謝する。

参考文献

- 1) Aucouturier, J.-J. and Pachet, F.: Music Similarity Measures: What's the Use?, *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR2002)*, pp. 157-163 (2002).
- 2) Logan, B.: Content-Based Playlist Generation: Exploratory Experiments, *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR*

- 2002)*, pp. 295-296 (2003).
- 3) Allamanche, E., Herre, J., Hellmuth, O., Kastner, T. and Ertel, C.: A Multiple Feature Model for Musical Similarity Retrieval, *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, pp. 217-218 (2003).
- 4) Berenzweig, A., Logan, B., Ellis, D. P. W. and Whitman, B.: A Large-Scale Evaluation of Acoustic and Subjective Music Similarity Measures, *Computer Music Journal*, Vol. 28, No. 2, pp. 63-76 (2004).
- 5) McKinney, M. F. and Breebaart, J.: Features for audio and music classification, *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, pp. 151-158 (2003).
- 6) Tzanetakis, G., Gao, J. and Steenkiste, P.: A Scalable Peer-to-Peer System for Music Content and Information Retrieval, *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, pp. 209-214 (2003).
- 7) Pampalk, E., Flexer, A. and Widmer, G.: Improvements of Audio-based Music Similarity and Genre Classification, *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, pp. 628-633 (2005).
- 8) Flexer, A., Gouyou, F., Dixon, S. and Widmer, G.: Probabilistic combination of features for music classification, *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR 2006)*, pp. 628-633 (2006).
- 9) Pohle, T., Knees, P., Schedl, M. and Widmer, G.: Independent Component Analysis for Music Similarity Computation, *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR 2006)*, pp. 228-233 (2006).
- 10) 藤原弘将, 北原鉄朗, 後藤真孝, 駒谷和範, 尾形哲也, 奥乃博: 伴奏音抑制と高音頻度フレーム選択に基づく楽曲の歌手名同定手法, *情報処理学会論文誌*, Vol. 47, No. 6, pp. 1831-1843 (2006).
- 11) Fujihara, H., Kitahara, T., Goto, M., Komatani, K., Ogata, T. and Okuno, H.G.: Singer Identification Based on Accompaniment Sound Reduction and Reliable Frame Selection, *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, pp. 329-336 (2005).
- 12) Goto, M.: A real-time music-scene-description system: predominant-F0 estimation for detecting melody and bass lines in real-world audio signals, *Speech Communication*, Vol. 43, No. 4, pp. 311-329 (2004).
- 13) Moorer, J. A.: Signal Processing Aspects of Computer Music: A Survey, *Proceedings of the IEEE*, Vol. 65, No. 8, pp. 1108-1137 (1977).
- 14) 徳田恵一, 小林隆夫, 今井聖: メル一般化ケプストラムの再帰的計算法, *電子情報通信学会論文誌 A*, Vol. J71-A, No. 1, pp. 128-131 (1988).
- 15) Ohishi, Y., Goto, M., Itou, K. and Takeda, K.: Discrimination between Singing and Speaking Voices, *Proceedings of 9th European Conference on Speech Communication and Technology (Eurospeech 2005)*, pp. 1141-1144 (2005).
- 16) 後藤真孝, 橋口博樹, 西村拓一, 岡隆一: RWC研究用音楽データベース, 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース, *情報処理学会論文誌*, Vol. 45, No. 3, pp. 728-738 (2004).