

伴奏音抑制と高信頼度フレーム選択に基づく 楽曲の歌手名同定手法

藤原 弘 将[†] 北原 鉄 朗[†] 後藤 真 孝^{††}
駒谷 和 範[†] 尾形 哲 也[†] 奥 乃 博[†]

本論文では、実世界の音楽音響信号に対する歌手名の同定手法について述べる。歌手名の同定を行う際に大きな問題となるのは、混在する伴奏音の影響である。本論文ではこの問題を解決するため、伴奏音抑制と高信頼度フレーム選択の手法を提案する。前者では、優勢なメロディの調波構造を抽出し再合成することで、伴奏音の影響を低減させることができる。後者は、歌声と非歌声を表す2種類の混合正規分布を用いて、それぞれのフレームが歌声として信頼できるか否かを判定するものである。実験の結果、本手法によって、10歌手40曲に対して95%の識別率を達成し、本手法を用いない場合と比較して誤り率を約89%削減した。また、20歌手256曲に対する実験の結果、約93%の識別率を達成し、誤り率を約65%削減した。

A Singer Identification Method for Musical Pieces on the Basis of Accompaniment Sound Reduction and Reliable Frame Selection

HIROMASA FUJIHARA,[†] TETSURO KITAHARA,[†] MASATAKA GOTO,^{††}
KAZUNORI KOMATANI,[†] TETSUYA OGATA[†] and HIROSHI G. OKUNO[†]

This paper describes a method for automatic singer identification from polyphonic musical audio signals. The main problem in automatically identifying singers is negative influence caused by accompaniment sounds. To solve this problem, we developed two methods: *accompaniment sound reduction* and *reliable frame selection*. The former makes it possible to reduce accompaniment sounds by extracting and resynthesizing harmonic components of the predominant melody. The latter judges whether each frame of the obtained melody is reliable or not by using two Gaussian mixture models for vocal and non-vocal frames. Experimental results with 256 songs by 20 singers showed that our method reduced 65% of classification errors, and achieved an accuracy of 93%.

1. はじめに

歌声は誰もが生まれながらに持つ最も基本的な楽器であり、多くのジャンルの音楽、特にポピュラー音楽において、重要な役割を果たしている。実際、多くの人は音楽を聞いたときに、その曲中の歌声を手掛かりにして曲名を判断するであろう。そのため、多くのCDショップなどでは、音楽の分類にジャンルの情報に加えて歌手名（アーティスト名）を用いている。

このように歌声は音楽の重要な要素であるため、楽

曲の歌手名に関する情報は音楽情報検索（MIR）に有用である。たとえば、ある歌手の楽曲を探したい場合、歌手名（アーティスト名）のタグを利用することで、望みの楽曲を検索することができる。さらに、歌手の特性を表現する特徴量を記述し、それらの類似性を計算することができれば、声質に基づく楽曲検索の実現など、MIRにおいて重要な役割を果たすことができる。しかし、既存のMIRシステムのほとんどは、アーティスト名などのメタデータがすでに記述されていることを前提としている。そのため、そのようなメタデータが記述されていない楽曲は、アーティスト名をクエリにした検索はできなかった。

本論文では、このような歌手名に基づく楽曲検索を実現するため、音楽音響信号から楽曲中の歌手名を同定する問題に取り組む。歌手名の同定における最大の

[†] 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University

^{††} 産業技術総合研究所
National Institute of Advanced Industrial Science and
Technology (AIST)

問題は、伴奏音の混在である。多くの楽曲には、歌声だけでなく種々の伴奏音も含まれている。そのため、そのような音楽音響信号から抽出された歌声の特徴量は、伴奏音によって影響を受けている。実際、ケプストラム係数や線形予測係数 (LPC) など、音声認識でよく用いられる特徴量を音楽音響信号に対して計算すると、得られた特徴量は、歌声のみを表現するのではなく、歌声と伴奏音が混ざった状態を表現してしまう。音声情報処理の分野では、話し声に対する話者認識の研究が多く行われてきた^{1),2)} が、ノイズのない話し声のみからなる信号を対象としたものが多いので、それらの研究での手法をそのまま歌手名の同定に用いることは難しい。

先行研究においては、Tsai ら³⁾ はこの伴奏音混在の問題を指摘し、雑音下話者認識の手法⁴⁾ を用いて解決しようとした。そこでは、歌声区間の確率モデルから伴奏 (雑音) の確率モデルを確率モデル上で減算することにより、歌声のみを表現する確率モデルを得ることを目的としていた。しかし、伴奏 (雑音) の確率モデルは直接観測できないため、“歌声の後ろで演奏される伴奏”と“歌声区間以外で演奏される間奏”は同一の性質を持つという仮定のもと、間奏区間の確率モデルで代用した。しかし、間奏部では歌声以外のリード楽器がメロディを演奏している場合が多い。そのため、“歌声区間以外で演奏される間奏”の確率モデルは、“歌声以外によって演奏されるメロディ”と“歌声の後ろで演奏される伴奏”が混ざった状態を表現してしまい、この仮定は満たされない場合が多かった。その他の従来研究⁵⁾⁻⁸⁾ では、伴奏音混在の問題は明示的に扱われてはいなかった。

伴奏音混在の問題を解決するため、本論文では、伴奏音抑制と高信頼度フレーム選択の2つの手法を提案する。伴奏音抑制では、音響信号中のメロディの調波構造を抽出し正弦波重畳モデルを用いて再合成することで、メロディのみからなる音響信号を得る。高信頼度フレーム選択の手法により、歌声の特徴をよく保存していて、識別に相応しいフレームを選択し識別に用いる。ここでは、歌声を表す確率モデルと、非歌声を表す確率モデルの尤度比によって歌声の信頼度を判定する。

以下、2章では、まず本研究での問題設定に取り組むうえでの課題と、その解決のためのアプローチについて述べる。3章では、提案手法の実装の詳細について述べる。4章では本手法の有効性を確かめるために評価実験を行い、5章ではまとめを行う。

2. 伴奏音の影響に頑健な歌手名の同定手法

本論文では、与えられた音楽音響信号中の歌手名を同定する問題を扱う。歌手名の同定を実現する際に課題となるのは、歌声と同時に演奏される伴奏音の影響である。そのため、高精度の歌手名同定を実現するには、このような伴奏音混在の問題を解決しなければならない。

この問題の1つの解決策としてマルチコンディション学習、すなわち学習の際に伴奏音の影響を受けたデータを使用する方法が考えられる。実際、多くの従来研究ではこの方法が用いられていた⁵⁾⁻⁸⁾。しかし、伴奏音の性質は曲によって大きく異なるため、この方法は適切でないと考えられる。たとえば、伴奏楽器にドラムス、ベース、ギターなどを含むフルバンドで演奏された楽曲と、ピアノの弾き語りの楽曲では、たとえ歌手が同じであっても、全体の音響的特性は大きく異なる。

本研究では、この伴奏音混在の問題を2つのアプローチで解決する。1つは、伴奏音を入力信号から除去し、メロディのみの信号を生成することである。音響信号から伴奏音を除去することは困難ではあるが、入力信号をそのまま用いるのに比べると、伴奏音の影響を軽減できると期待できる。本研究では、これをメロディの調波構造抽出と正弦波重畳モデルを用いた再合成により実現した。この手法を伴奏音抑制と呼ぶ。ここで、メロディとは各時刻において最も優勢な音として定義する⁹⁾。さらに、歌声の伴奏音による影響の受け方は時刻によって異なるため、影響の小さい部分のみを選択して同定に用いれば、よりロバストな歌手同定が期待できる。本研究では、これを高信頼度フレーム選択と呼び、歌声と非歌声を学習させた2つの混合ガウス分布 (GMM) を用いて実現する。図1に、本手法の概要を示す。以下、それぞれの手法について詳説する。

2.1 伴奏音抑制

伴奏音抑制の手法は、以下の3つの処理からなる。

- (1) 後藤の PreFEst⁹⁾ を用いて、メロディの基本周波数を推定する。
- (2) 推定された基本周波数に基づき、メロディの調波構造を抽出する。
- (3) 抽出された調波構造を、正弦波重畳モデルを用いて音響信号に再合成する。

以上の処理で、楽曲中のメロディのみの音響信号を得ることができる。

本手法により得られたメロディの音響信号は、間奏

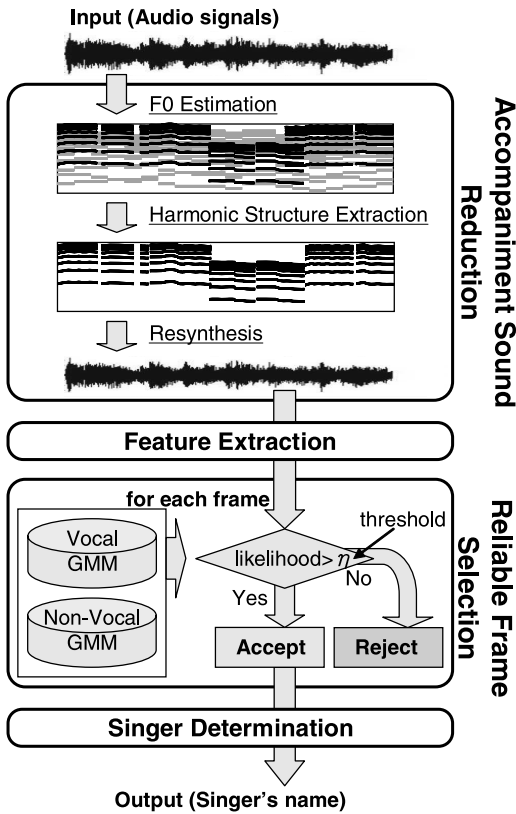


図 1 歌手名同定手法の概要

Fig. 1 Overview of our method for singer identification.

などの区間では(歌声でない)楽器音を含んでいる．そのため、歌声が存在する区間を前もって検出しておく必要があると考えられるが、歌声区間の検出は大変困難な問題である．本論文では、後述の高信頼度フレーム選択の手法を用いることでこの問題を回避する．

2.2 高信頼度フレーム選択

高信頼度フレーム選択では、歌声が存在する区間から抽出された特徴量で学習した歌声 GMM λ_V と、伴奏区間から抽出された特徴量で学習した非歌声 GMM λ_N の尤度比を、各フレームの信頼度とする．

x を特徴ベクトルとしたとき、歌声 GMM の尤度 $p(x|\lambda_V)$ はその特徴ベクトルの歌声らしさを表し、非歌声 GMM の尤度 $p(x|\lambda_N)$ はその特徴ベクトルの非歌声らしさを表す．特徴ベクトルが伴奏音による影響をあまり受けていなかった場合、 $p(x|\lambda_V)$ は大きくなり、 $p(x|\lambda_N)$ は小さくなるため、信頼度は高くなる．すなわち、 x の信頼度 $R(x)$ は、

$$CM(x) = \log p(x|\lambda_V) - \log p(x|\lambda_N) \quad (1)$$

と表される．

本論文では、各曲の中から信頼度が上位の $\alpha\%$ のフ

レームを識別に用いる．なぜなら、様々な楽曲に対して共通の閾値を決定すると、信頼度の高いフレームが少ない楽曲では、識別に十分な量の特徴ベクトルを確保できないからである．この手法により、歌声を含まないフレームの多くを排除できる．つまり、高信頼度フレーム選択の手法を用いることで、歌声区間の検出を明示的に行う必要がなくなる．

3. 実装

この章では、本論文で述べる歌手名の同定システムの実装について述べる．前述のように、このシステムは、伴奏音抑制、特徴抽出、高信頼度フレーム選択、識別の 4 つの処理からなる．以下、それぞれの処理について述べる．

3.1 前処理

入力音響信号を、モノラル化し、16 kHz にダウンサンプリングする．そして、フレーム幅 128 ms (2,048 サンプル)、フレームシフト 10.0 ms (160 サンプル) で短時間フーリエ変換を行い、スペクトログラムを計算する．本研究では、窓関数にハミング窓を用いた．

3.2 伴奏音抑制

2.1 節で述べた手法を用いて、伴奏音を抑制する．

3.2.1 基本周波数推定

基本周波数推定 (F0 推定) には、Goto の PreFEst⁹⁾ を用いる．PreFEst は、制限された周波数帯域において最も優勢な調波構造を持つ F0 を推定する手法である．メロディは中高域の周波数帯域において最も優勢な調波構造を持つ場合が多いため、周波数帯域を適切に制限することで、メロディの F0 を推定することができる．

以下、PreFEst の概要を記す．以後、 x は cent の単位で表される対数周波数軸上の周波数で、 (t) は時間を表すとする．cent は、本来は音高差 (音程) を表す尺度であるが、本論文では文献 9) に従い、 $440 \times 2^{\frac{x}{12} - 5}$ Hz を基準として、次式のように絶対的な音高を表す単位として用いる．

$$f_{cent} = 1200 \log_2 \frac{f_{Hz}}{440 \times 2^{\frac{x}{12} - 5}} \quad (2)$$

パワースペクトル $\Psi_p^{(t)}(x)$ に対して、メロディの周波数成分の多くが通過するように設計された帯域通過フィルタを適用する．本研究では、文献 9) に従い、4,800 cent 以上の成分を通過させるフィルタを用いた．フィルタを通過後の周波数成分は $BPF(x)\Psi_p^{(t)}(x)$ と表される．ただし、 $BPF(x)$ はフィルタの周波数応答である．以後の確率的処理を可能にするため、フィルタを通過後の周波数成分を確率密度関数 (PDF) とし

て、以下のように表現する．

$$p_{\Psi}^{(t)}(x) = \frac{BPF(x)\Psi_p^{(t)}(x)}{\int_{-\infty}^{\infty} BPF(x)\Psi_p^{(t)}(x)dx} \quad (3)$$

その後、周波数成分の PDF が、すべての可能な F0 に対応する音モデルの重みつき和からなる確率モデル、

$$p(x|\theta^{(t)}) = \int_{F_l}^{F_h} w^{(t)}(F)p(x|F)dF, \quad (4)$$

$$\theta^{(t)} = \{w^{(t)}(F)|F_l \leq F \leq F_h\} \quad (5)$$

から生成されたと考える．ここで、 $p(x|F)$ は、それぞれの F0 についての音モデルとし、 F_h と F_l をとりうる F0 の上限と下限とする．また、 $w^{(t)}(F)$ は音モデルの重みで、

$$\int_{F_{h_i}}^{F_{l_i}} w^{(t)}(F)dF = 1 \quad (6)$$

を満たす．音モデルとは典型的な調波構造を表現した確率分布である．そして、EM アルゴリズムを用いて $w^{(t)}(F)$ を推定し、それを F0 の PDF と解釈する．最終的に、以下の式を用いて、最も優勢な基本周波数 $\overline{F}^{(t)}$ を決定する．

$$\overline{F}^{(t)} = \operatorname{argmax}_F w^{(t)}(F) \quad (7)$$

3.2.2 調波構造抽出

推定された F0 に基づき、メロディの調波構造の各倍音成分のパワーと位相を抽出する．それぞれの周波数成分の抽出の際には前後 $|r|$ セントずつの誤差を許容し、この範囲で最もパワーの大きなピークを抽出する． l 次倍音 ($l = 1, \dots, 20$) の周波数 F_l 、パワー A_l 、位相 θ_l は、以下のように表される．

$$F_l = \operatorname{argmax}_F |S(F)|$$

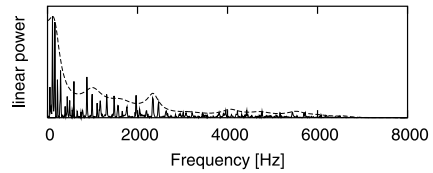
$$(\overline{F} \cdot (1 - 2^{\frac{r}{1200}}) \leq F \leq \overline{F} \cdot (1 + 2^{\frac{r}{1200}})), \quad (8)$$

$$A_l = |S(F_l)|, \quad (9)$$

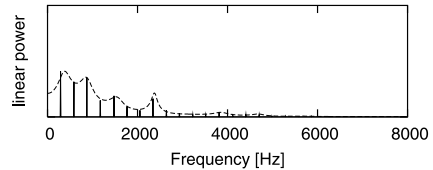
$$\theta_l = \arg S(F_l), \quad (10)$$

ここで、 $S(F)$ はスペクトルを、 \overline{F} は PreFEst によって推定された F0 を表す．本論文では、 r の値として 20 を用いた．

図 2 に、F0 推定と調波構造抽出の例を示す．図 2(a) は、伴奏音と歌声が混在したスペクトルとその包絡を表し、図 2(b) は、抽出された歌声の調波構造のスペクトルとその包絡を表す．なお、スペクトルの包絡は線形予測分析を用いて計算されたものである．抽出後のスペクトル包絡は、元のスペクトル包絡と比較して、伴奏の影響が減っていることが分かる．



(a) An original spectrum and its envelope.



(b) An extracted spectrum and its envelope.

図 2 F0 推定、調波構造抽出の一例

Fig. 2 Example of F0 estimation and harmonic structure extraction.

3.2.3 再合成

抽出された調波構造を正弦波重畳モデルに基づき再合成することで、メロディの音響信号を得る．再合成された音響信号は、

$$s(t) = \sum_{l=1}^L A_l \cos(\omega_l t + \theta_l), \quad (11)$$

のように表される．ここで、 A_l 、 θ_l 、 ω_l はそれぞれ、 l 次倍音のパワー、位相、周波数を表し、 t は時間を表す．

3.3 特徴抽出

再合成された音響信号から、特徴ベクトルを計算する．音声信号の個性は、スペクトルの微細構造ではなく包絡に含まれていることが知られている¹⁰⁾．音声認識、話者認識の分野では、スペクトル包絡を分離し特徴ベクトルを計算するための様々な手法が提案されている¹¹⁾．しかし、歌手名の同定の際に有効な特徴量についての考察は、今まで行われていなかった．そこで、本研究では、メル周波数ケプストラム係数 (MFCC)¹²⁾ と、線形予測メルケプストラム係数 (LPMCC)¹³⁾ の比較を行う．MFCC は音声信号に対してだけでなく、音楽音響信号に対しても一般的に使われている特徴量である¹⁴⁾．LPMCC は、線形予測分析 (LPC) によって得られる LPC スペクトルに対するメルケプストラム係数である．

3.3.1 メル周波数ケプストラム係数 (MFCC)

MFCC^{12),14)} は、メル周波数軸上で計算されるケプストラム係数である．ケプストラム分析¹¹⁾ とは、スペクトルの包絡と微細構造、つまり、声道特性と声帯振動を分離する手法である．ケプストラム係数は、対数パワースペクトルを離散コサイン変換することで計算

される．スペクトル包絡はケプストラムの低次の係数に表現され，微細構造は高次の係数に表現される．メル周波数とは，人間の聴覚特性に適合した対数周波数軸である．MFCC の計算においては，まずメルフィルタバンク分析を行い，その後，対数をとって離散コサイン変換を行う．本論文では，12次元 MFCC を用いた．

3.3.2 線形予測メルケプストラム係数 (LPMCC)

LPMCC は，線形予測スペクトル (LPC スペクトル) に対するメルケプストラム係数である．線形予測分析 (LPC)¹⁵⁾ とは，スペクトルの包絡と微細構造，つまり，声道特性と声帯振動を分離する手法の 1 つである．線形予測分析では，入力音響信号 $s(n)$ が与えられた場合に，ある時点での信号が，過去の一定期間の信号の 1 次結合で予測できると仮定する．予測値である $s_W(n)$ は，

$$s_W(n) = \sum_{i=1}^p \alpha_i s_W(n-i) + g(n), \quad (12)$$

のように与えられる．ここで， p は予測器の次数を表し，予測係数である α_i は線形予測係数 (LPC) と呼ばれる．また， $g(n)$ はモデルの誤差を表す．線形予測係数は，誤差 $g(n)$ の二乗平均が最小となるように決定する．求められた線形予測係数を用いて，LPC スペクトルを，

$$|H(e^{j\omega})|^2 = \frac{1}{|1 - \sum_{i=2}^p \alpha_i e^{-ji\omega}|^2} \quad (13)$$

のように求めることができる．

LPC スペクトルに対するケプストラム分析は，基底の直交化の役割を果たしパターン認識の特徴量として有効であることが知られている．さらに LPMCC は，メル周波数軸を使用したことで，人間の聴覚特性に対する適合性という点で優れている．本研究では，LPC スペクトルから MFCC を計算することで，LPMCC を得た．

3.4 高信頼度フレーム選択

2.2 節で述べた手法に基づき，伴奏音の影響が少なく，信頼度が高いフレームを選択する．本論文では GMM の混合数として，64 混合を用いた．また，信頼度が高いと判定するフレームの割合 α は，15% に設定した．

3.5 歌手名の決定

64 混合 GMM を用いて，歌手名を同定する．識別対象の歌手それぞれについて，GMM λ_s (s は歌手ラベル) を事前に学習しておく． $\mathbf{X} = \{\mathbf{x}_t | t = 1, \dots, T\}$ を，高信頼度フレーム選択によって選ばれた特徴ベクトルとすると，歌手名は，以下の式に基づき決定さ

れる．

$$s = \underset{i}{\operatorname{argmax}} \frac{1}{T} \sum_{t=1}^T \log p(\mathbf{x}_t | \lambda_i) \quad (14)$$

4. 実験

本章では，伴奏音抑制と高信頼度フレーム選択の手法の有効性を確認するために行った歌手名の同定実験について述べる．実験データとして，“RWC 研究用音楽データベース：ポピュラー音楽”¹⁶⁾ から選ばれたデータセットと，市販 CD から選ばれたデータセットの 2 種類を用いて実験を行った．

4.1 RWC 研究用音楽データベースを用いた実験

4.1.1 実験条件

実験を表 1 の 4 つの条件で行った．ただし，伴奏音抑制と高信頼度フレーム選択を行う場合は，評価データだけでなく，学習データに対しても同様に処理をした．つまり，高信頼度フレーム選択を行う実験条件では，楽曲全体に対して高信頼度フレーム選択を行い，選択されたフレームのみを用いて GMM を学習する．伴奏音抑制を行わない場合は，入力音響信号に対して直接特徴抽出をし，高信頼度フレーム選択を行わない場合は，学習・識別に楽曲全体の特徴ベクトルを用いる．実験には，“RWC 研究用音楽データベース：ポピュラー”¹⁶⁾ から選んだ 10 歌手 (男声 5 人，女声 5 人) による計 40 曲 (歌手 1 人あたり 4 曲) を使用した．これらの楽曲の詳細を表 2 に記す．これらの 40

表 1 実験条件
Table 1 Experimental conditions.

	reduction	selection
条件 i (Baseline)	×	×
条件 ii	○	×
条件 iii	×	○
条件 iv (Our method)	○	○

表 2 使用楽曲の内訳 (RWC 研究用音楽データベース)
Table 2 Artists and songs selected from RWC music database.

	Name	Gender	Piece Number
a	西一男	M	012, 029, 036, 043
b	風戸ヒサヨシ	M	004, 011, 019, 024
c	森元康介	M	038, 039, 042, 044
d	井口慎也	M	082, 084, 088, 090
e	Jeff Manning	M	085, 087, 095, 098
f	吉井弘美	F	002, 017, 069, 075
g	緒方智美	F	007, 028, 052, 080
h	凜	F	014, 021, 050, 053
i	服部まきこ	F	065, 067, 068, 077
j	Betty	F	086, 092, 094, 096

表 3 高信頼度フレーム選択の学習に用いた楽曲
Table 3 Training data for reliable frame selection.

Name	Gender	Piece Number
勝田真悟	M	027
波多江良徳	M	037
食原正機	M	032, 078
関谷洋	M	048, 049, 051
小澤克乃	M	015, 041
熊坂敬	M	047
オリケン	M	006
KONBU	F	013
市川えり	F	020
新田智子	F	026
鍋木朗子	F	055
飯島柚子	F	060
佐藤れいこ	F	063
松阪珠子	F	070
Donna Burke	F	081, 089, 091, 093, 097

曲の平均時間長は約 250 秒で、その標準偏差は約 38 秒である。

高信頼度フレーム選択の学習データには、表 3 に示される 16 歌手からなる 25 曲を用いた。これらの 16 歌手には、表 2 の 10 歌手は含まれていない。これらの 25 曲の各フレームに対して、歌声区間・非歌声区間の別を手動でラベル付けし、歌声 GMM と非歌声 GMM を学習した。ただし、条件 ii と iv では、これらの 25 曲に対しても、GMM を学習する前に、伴奏音抑制をしている。歌声 GMM の学習に用いた時間長は約 4,692 秒であり、非歌声 GMM の学習に用いた時間長は約 1,428 秒であった。

評価方法は、4-fold cross-validation 法を用いた。すなわち、各歌手の 4 曲のうち 3 曲で学習し残りの 1 曲で評価する、という操作を、すべての楽曲が評価されるように 4 回繰り返した。1 人の歌手の学習に用いられたデータ長は平均約 749 秒だった。1 回の評価では、まず、各歌手ごとに学習データを用いて GMM を学習し、識別では評価用楽曲各 1 曲につき 1 つの歌手ラベルを出力した。学習には、伴奏音の区間も含めて、楽曲全体を使用した。特徴量は、3.3 節で述べた MFCC と LPMCC を両方用いて比較した。

4.1.2 実験結果

図 3 に、RWC 研究用音楽データベースに対する実験結果を示す。伴奏音抑制と高信頼度フレーム選択により、識別率が向上したことが分かる。また、伴奏音抑制と高信頼度フレーム選択を同時に用いた場合に、識別率が 55% から 95% と、大幅に向上していることが分かる。図 4 に、特徴量として LPMCC を用いた場合の混同行列を示す。提案手法を用いることで、誤識別が減少し混同行列が対角行列に近づいていく様子

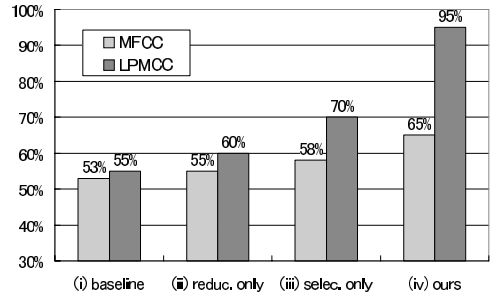


図 3 実験結果 1 (RWC 研究用音楽データベース)。“reduc.” と “selec.” は、それぞれ伴奏音抑制と高信頼度フレーム選択を表す

Fig. 3 Results of the experiments using RWC music database, where “reduc.” and “selec.” mean accompaniment sound reduction and reliable frame selection, respectively.

が見てとれる。

次に、MFCC を用いた場合と LPMCC を用いた場合とで結果を比較する。すべての実験条件について、LPMCC のほうが、MFCC より識別率が高いことが確認できる。特にこの傾向は、伴奏音抑制と高信頼度フレーム選択を用いた場合に顕著である。LPMCC は、歌手性をよく表す特徴量であることが確認された。

4.1.3 伴奏音抑制の評価

伴奏音抑制の効果を評価するため、伴奏音を含む信号に対して伴奏音抑制を行うことによる平均対数スペクトル歪みの変化を計算した。歌声のみのスペクトル、伴奏音を含むスペクトル、伴奏音抑制後のスペクトルを、それぞれ、 $S_v^{(i)}$ 、 $S_a^{(i)}$ 、 $S_r^{(i)}$ としたとき、楽曲の平均対数スペクトル歪みの変化を以下の式で定義する。

$$\frac{1}{I} \sum_{i=1}^I \{D(S_v^{(i)}, S_a^{(i)}) - D(S_v^{(i)}, S_r^{(i)})\} \quad (15)$$

ただし、 $D(S_1, S_2)$ は、2 つのスペクトル S_1, S_2 の対数スペクトル歪み (単位 dB) を表し、 I は歌声を含むフレームの総数、 i は歌声を含むフレームのインデックスを表す。この値を、実験に用いた楽曲の歌声を含むフレームについて計算し、楽曲ごとの平均をとると、 -4.77 dB という結果になった。これは、すなわち伴奏音抑制を適用することにより抑制された伴奏音成分の大きさを表し、伴奏音抑制が有効に機能していることを示している。

また、伴奏音抑制の効果を視覚的に明らかにするため、図 5 に楽曲番号 007 の楽曲の歌声区間の一部について、伴奏を含む信号のスペクトログラム (a)、伴奏音抑制適用後の信号のスペクトログラム (b)、ボーカルのみ信号のスペクトログラム (c) を示す。伴奏

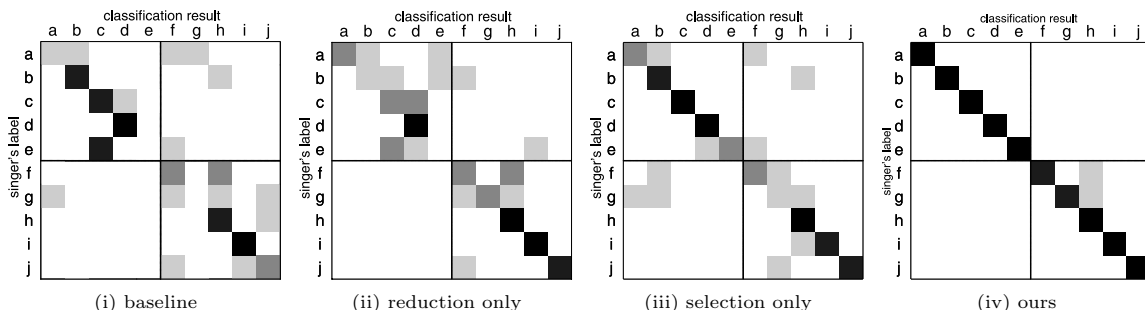
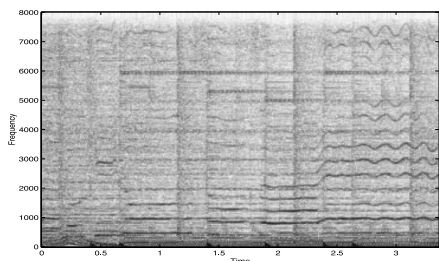
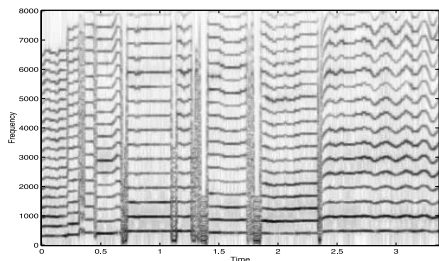


図 4 混同行列 (RWC 研究用音楽データベース). 特微量は LPMCC を用いた. “reduction” と “selection” は, それぞれ伴奏音抑制と高信頼度フレーム選択を表す. また, 各図中の点線は, 男女の境界を示す

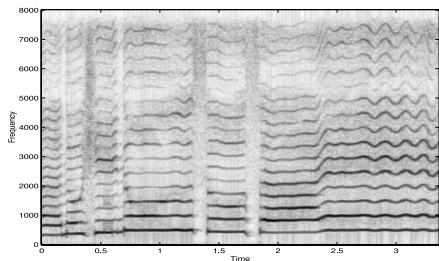
Fig. 4 Confusion matrices of the experiments using RWC music database. Center lines in each figure are boundaries between male and female. Note that confusion between male and female decreased by using the accompaniment sound reduction method.



(a) A spectrogram of polyphonic signal.



(b) A spectrogram of separated signal.



(c) A spectrogram of vocal-only signal.

図 5 伴奏音抑制の一例

Fig. 5 Example of accompaniment sound reduction.

音抑制を適用することで, 伴奏楽器の周波数成分が抑制されて, 歌声に関する成分が強調されているのが見てとれる.

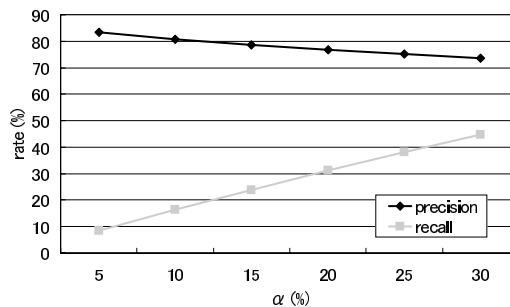


図 6 高信頼度フレーム選択の適合率と再現率

Fig. 6 Precision rate and recall rate of reliable frame selection.

4.1.4 高信頼度フレーム選択の評価

非歌声区間を棄却できていることと, 伴奏音の影響が小さいフレームを選択できていることの 2 点を示すことで, 高信頼度フレーム選択の性能を評価する. 図 6 に, 信頼度が高いと判断されるフレームの割合 α の値を様々に変化させた場合の, 選択されたフレームの中で実際に歌声を含むフレームの割合 (適合率) と, 実際に歌声を含むフレーム中で本手法によって選択されたフレームの割合 (再現率) の変化のグラフを示す. 実験に用いた α の値 (0.15) では, 適合率は約 79% となっており, 非歌声区間の多くを棄却できていることが分かる. α を変化させた場合の, 高信頼度フレーム選択によって選択されたフレームと選択されなかったフレームの平均対数スペクトル歪みの変化を図 7 に示す. なお, 平均対数スペクトル歪みの計算には, 歌声を含むフレームのみを用いた. また, 図 8 は, 各 α における信頼度 CM の閾値の変化である. スペクトル歪みの値が α の増加とともに単調増加しているのが分かる. 図 6 と図 7 より, α を小さくするこ

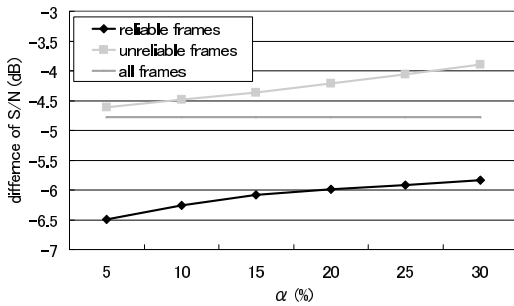


図 7 α の値による、選択されたフレームのスペクトル歪みの変化
Fig. 7 Dependency of spectral distortion of selected frames on α .

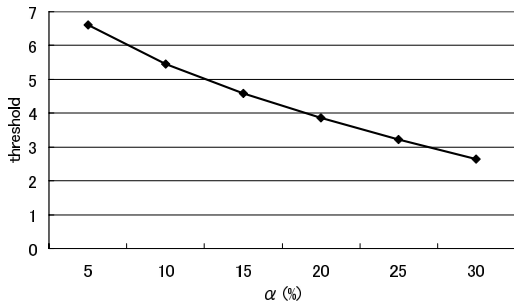


図 8 α の値による、信頼度 CM の閾値の変化
Fig. 8 Dependency of threshold of CM on α .

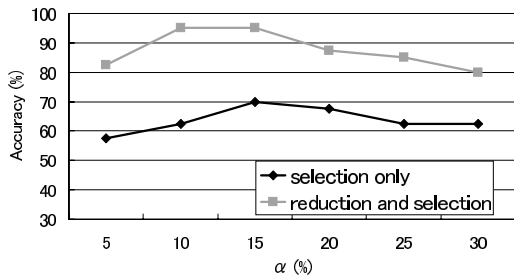


図 9 識別率の α の値による比較 (RWC 研究用音楽データベース)
Fig. 9 Experimental results that show dependency of accuracy on α .

とで、再現率が低下し、歌声区間の中でも特に伴奏音の影響が小さいフレームを選択できていることが確認された。以上より、高信頼度フレーム選択の有効性が確認された。

α の値を変化させた場合の識別率の変化を図 9 に示す。この図から、 α を変化させたことによる識別率の変動は小さいと判断できる。また、 α を 15% より大きくすることで識別率が低下している。これは、 α の値を大きくしすぎることによって、信頼度の低いフレームをより多く選んでしまったことが原因と考えられる。反対に、 α を 10% より小さくすることで識別率が低下する。これは、選択するフレームの数が少なすぎ

表 4 高信頼度フレーム選択の評価

Table 4 An evaluation of reliable frame selection.

実験条件		(i)	(ii)	(iii)	(iv)
使用	学習時	楽曲全体		歌声区間	
データ	評価時	楽曲全体		歌声区間	
フレーム 選択	学習時	○	×	×	×
	評価時	○	○	○	×
識別率		95%	93%	93%	83%

て、学習・識別に十分な特徴量が確保できなかったためと考えられる。

高信頼度フレーム選択に、単に歌声区間を識別すること以上の効果があることを確かめるため、

- (i) 学習・識別データとともに楽曲全体を用い、両方に高信頼度フレーム選択を行った場合（提案手法）、
- (ii) 学習・識別データとともに楽曲全体を用い、識別データのみを高信頼度フレーム選択を行った場合、
- (iii) 学習データに真の歌声区間のデータのみを、識別データに楽曲全体を用い、識別データにのみ高信頼度フレーム選択を行った場合、
- (iv) 学習・評価データとともに歌声区間のみを用い、どちらにも高信頼度フレーム選択を行わない場合、

の 4 とおりの実験条件で、歌手名同定実験を行った。ただし、伴奏音抑制は学習データ、評価データともに処理した。また、本実験で付与した歌声区間は、各楽曲に対応する歌声のみの音響信号に対してパワーの大きい区間を検出することで得た。そのほかの実験方法は、4.1.1 項と同様である。表 4 に結果を示す。条件 (i) と条件 (ii) を比較すると識別率が 2% 低下している一方、条件 (ii) と条件 (iii) の比較では、識別率が変化していない。条件 (i) と条件 (iv) を比較すると、識別率が 12% 低下している。このことから、単純に歌声が存在する区間を選ぶのみでは不十分で、高信頼度フレーム選択によって伴奏音の影響が少ないフレームを絞り込むことが効果的であるといえる。また、高信頼度フレーム選択は、特に識別データに用いたときに効果が大きいことが分かる。

最後に、伴奏音抑制と高信頼度フレーム選択を組み合わせることの効果を確認するため、次の 3 通りの条件で実験を行った。

- (i) 学習・識別に歌声区間のみを用い、伴奏音抑制にボーカルのみのデータから推定された F0 を用いた。
- (ii) 学習・識別に歌声区間のみを用い、伴奏音抑制に PreFEst で推定した F0 を用いた。
- (iii) 学習・識別に楽曲全体を用い、伴奏音抑制に PreFEst で推定した F0 を用いた。

表 5 伴奏音抑制と高信頼度フレーム選択の組合せの評価

Table 5 An evaluation of a combination of accompaniment sound reduction and reliable frame selection.

実験条件	(i)	(ii)	(iii)
フレーム選択	×	×	○
使用データ	歌声区間		楽曲全体
F0 推定	ボーカルのみ		PreFEst
識別率	88%	83%	95%

ボーカルのみのデータから推定された F0 は完全に正解とはいえないが、PreFEst で推定した F0 の精度と比較すると十分に高いと見なすことができる。表 5 に結果を示す。条件 ii と条件 iii の比較では、識別率が 83% から 95% に 12% 上昇している一方で、条件 i と条件 ii を比較すると、識別率が 83% から 88% に 5% しか上昇していない。伴奏音抑制を用いることで、F0 推定誤りがほとんどない場合以上の識別率が得られていることが分かる。これは、F0 推定を誤ったフレームは歌声としての信頼度が低くなり、高信頼度フレーム選択でリジェクトされたためだと考えられる。

4.2 市販 CD を用いた実験

4.2.1 実験条件

実験を、前節での実験と同様に、表 1 の 4 つの条件で行った。実験には、市販 CD から選んだ 20 歌手（男声 8 人、女声 12 人）からなる 246 曲を用いた。これらの楽曲の選定では、オリコンの 2004 年度年間アルバムチャートの上位から、歌手が 1 人のアーティスト 20 アーティストほど選択し、それぞれに対して順位が最上位のアルバムに含まれる楽曲を使用した。使用したアーティストを表 6 に示す。これらの 246 曲の平均時間長は約 268 秒で、その標準偏差は約 61 秒である。高信頼度フレーム選択の学習データは、4.1 節の実験と同様で、表 3 に記される 16 歌手からなる 25 曲を使用した。評価方法は、3-fold cross-validation 法を用いた。すなわち、各歌手の楽曲の 3 分の 2 を用いて学習し、残りの楽曲を評価に用いるという操作をすべての楽曲が評価されるように 3 回繰り返した。1 人の歌手の学習に用いられたデータ長は平均約 2,195 秒だった。そのほかの実験方法は、4.1 節の実験と同様である。

4.2.2 実験結果

図 10 に、実験結果を示す。伴奏音抑制か高信頼度フレーム選択のいずれか一方の手法を用いることで、約 8% の識別率の向上を確認した。さらに、両方の手法を同時に用いることで、識別率が約 12% 向上することを確認した。本手法は、市販 CD に対しても有効に機能することが分かった。

表 6 使用楽曲の内訳（市販 CD）

Table 6 Artists selected from commercial CD recordings.

	Artist Name	Gender	Tracks
A	Asian Kung-fu generation	M	11
B	Bump of Chicken	M	10
C	平井堅	M	10
D	槇原敬之	M	12
E	森山直太郎	M	11
F	Mr.Children	M	12
G	ポルノグラフィティ	M	13
H	QUEEN	M	16
I	Aiko	F	13
J	Avril Lavigne	F	14
K	BoA	F	12
L	浜崎あゆみ	F	8
M	平原綾香	F	10
N	倉木麻衣	F	16
O	中島美嘉	F	13
P	大塚愛	F	11
Q	島谷ひとみ	F	15
R	柴咲コウ	F	12
S	宇多田ヒカル	F	15
T	矢井田瞳	F	12

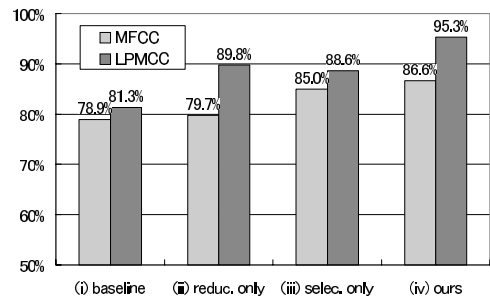


図 10 実験結果（市販 CD）。“reduc.” と “selec.” は、それぞれ伴奏音抑制と高信頼度フレーム選択を表す

Fig. 10 Results of the experiments using commercial CD-recordings, where “reduc.” and “selec.” mean accompaniment sound reduction and reliable frame selection, respectively.

図 11 に、特徴量として LPMCC を用いた場合の混同行列を示す。男声と比較して、女声の誤識別が多いことが読み取れる。これは、今回用いた 12 次元 LPMCC では、女声のスペクトル包絡と周波数微細構造を分離できなかったからであると考えられる。通常、話し声を対象にした場合、スペクトル推定におけるモデルのパラメータの次数は、女性や子どもの音声で、周波数微細構造の成分がスペクトル包絡の観測値に現れない程度に大きく設定する¹³⁾。しかし、女性の歌声の F0 はときには 1,000 Hz にも達するため、男声や F0 の低い女性の歌声を正確に表現できる程度にパラメータ次数を大きくすると、F0 の高い女性の歌声では微細構造の成分がスペクトル包絡の観測値に現れ

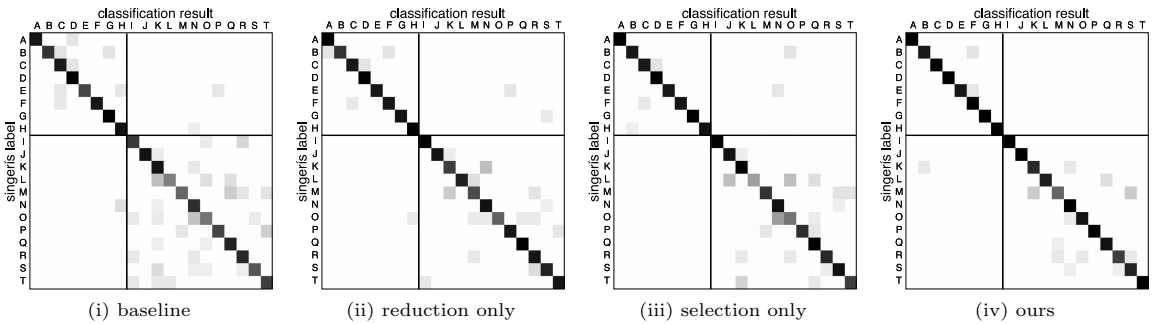


図 11 混同行列 (市販 CD). 特徴量は LPMCC を用いた. “reduction” と “selection” は, それぞれ伴奏音抑制と高信頼度フレーム選択を表す. また, 各図中の点線は, 男女の境界を示す.

Fig. 11 Confusion matrices of the experiments using commercial CD-recordings. Center lines in each figure are boundaries between male and female.

てしまった.

市販 CD を用いた実験は, RWC 音楽データベースを用いた場合と比較して, ベースライン手法 (条件 i) の認識率が 15%程度高い. これは, 市販 CD では各アルバムにおけるアーティストごとの特色を出すため, 同じ楽器を使用するなど, 同一アルバム内では全体の音質が均質になるように作られている場合が多いためと考えられる. そのため, 伴奏音の影響を受けたままの特徴量で識別を行っても, 81.3%とある程度高い識別率が得られた. Berenzweig ら⁶⁾ は, この現象を “アルバム・エフェクト” と呼び, 歌手名の同定実験の識別率は, 用いるデータセットの選び方に依存することを指摘している. 一方, RWC 研究用音楽データベースは様々なジャンルにおける豊かなバリエーションを持つ楽曲を可能な限り大量に収録することを目標に構築されている¹⁶⁾ ため, 同一歌手の楽曲でも使用楽器, 曲調などが多様である. そのため, ベースライン手法では識別率が 55%程度しか達成できなかった. しかし, 実際には, 複数のバンドに所属する歌手や, 時期によって全体の音質が大きく異なるアーティストも存在する. 提案手法では, このような多様な楽曲を含むデータセットに対して高精度に歌手名を同定できた.

4.3 考 察

本節では, 提案手法の新規性・有効性について, 従来研究と比較しながら考察する.

4.3.1 伴奏音抑制の新規性・有効性

伴奏音抑制の新規性は, 伴奏音混在の問題を伴奏音に関する仮定をおくことなく解決した点にある. ほとんどの従来研究^{5)~8)} では, この問題を扱っていなかった. Tsai ら³⁾ はこの問題を扱っていたが, 彼らの手法では, 歌声のみの確率モデルを推定するのに, 伴奏の確率モデルを必要としていた. 伴奏モデルは直接観

測できないので, 「歌声の後ろで演奏される伴奏と歌声区間以外で演奏される間奏は類似した性質を持つ」という仮定に基づき, 間奏部のモデルで代用した. しかし, 1 章で指摘したように, この仮定は満たされない場合が多い. 本論文で提案する伴奏音抑制は, スペクトル上での処理で歌声の調波構造を直接推定するため, 伴奏に関する仮定を必要としないという新規性がある.

本論文では, 伴奏音抑制の有効性を歪み率の減少と識別率の改善の 2 つの観点から確かめた. 前者の観点については, 4.1.3 項の実験で, 伴奏音抑制によって歪み率が -4.77 dB 変化することから確認した. また, 後者の観点については, 伴奏音抑制を用いることで識別率が RWC 研究用音楽データベース (10 歌手 40 曲) を用いた実験で 70%から 95%に, 市販 CD を用いた実験 (20 歌手 246 曲) で 88.6%から 95.3%に向上したことから確認した.

4.3.2 高信頼度フレーム選択の新規性・有効性

高信頼度フレーム選択の新規性は, 時間軸上で信頼度の高い, すなわち伴奏音の影響が少ないフレームを選ぶという点にある. 従来用いられてきた手法は, 歌声区間を検出することが目的で, 伴奏音の影響が少ないフレームを選ぶという観点には至っていなかった. 本論文では, 歌声を含む区間においても時刻によって伴奏音の影響の程度は異なることに着目し, 伴奏音の影響が少ないフレームのみを自動的に選択する手法を実現した.

高信頼度フレーム選択の有効性は, 下記の 2 つの比

実験条件が異なるため, 正確な定量的評価ではないが, Tsai らの論文³⁾ では, 20 歌手 200 曲からなるデータベースに対して, 彼らの提案手法を用いることで, 識別率が 75.3%から 80.4%になったことを報告している.

較によって確かめた．1つは，高信頼度フレーム選択によって選ばれたフレームと選ばれなかったフレーム間のスペクトル歪みの比較である．高信頼度フレーム選択によって選ばれたフレームのほうがスペクトル歪みが小さく，高信頼度フレーム選択は伴奏音の影響が少ないフレームを選択できることを定量的に確認した．また，2つ目は，真の歌声区間のみを用いた歌手名同定実験の識別率と，高信頼度フレーム選択で選ばれたフレームのみを用いた歌手名同定実験の比較である．前者と後者を比較すると識別率が83%から95%に向上している．このことから，歌声区間を検出する従来手法のみでは不十分で，高信頼度フレーム選択により，歌声区間の中でも伴奏音の影響が少ないフレームを絞り込む手法が有効であることが確かめられた．

4.3.3 伴奏音抑制と高信頼度フレーム選択を組み合わせることの効果

最後に，伴奏音抑制と高信頼度フレーム選択を組み合わせることの有効性について述べる．伴奏音抑制には，F0推定に誤りがある場合に，逆に歌声の成分を抑制してしまうという問題点がある．しかし，伴奏音抑制に加えて，高信頼度フレーム選択を行うことで，F0推定誤りを含むフレームを棄却できるため，F0推定誤りに対して頑健である．これは，F0の正解を与えた場合と比較して，F0推定誤りを含んだままで高信頼度フレーム選択を適用した場合のほうが，識別率が88%から95%に向上したことにより確認した．雑音の影響を低減させるために音声の調波構造を用いることの有効性は，Nakataniら¹⁷⁾によって確かめられていたが，本研究では，それに加えてF0推定誤りの影響に頑健な手法を実現したという点で新規性がある．

5. ま と め

本論文では，音楽音響信号から歌手名を同定するうえでの難しさとして，伴奏音混在の問題を指摘し，これを伴奏音抑制と高信頼度フレーム選択という2つの手法で解決した．前者は，メロディの調波構造を抽出・再合成することで伴奏音の影響の少ない信号を生成する手法で，後者は，あらかじめ学習した歌声モデルと非歌声モデルから，歌声らしさの高い特徴ベクトルだけを選ぶ手法である．これらの手法により，RWC研究用音楽データベース収録の10歌手40曲に対して55%から95%へ，市販CDに収録の20歌手246曲に対して81%から93%へ認識率を改善することができた．

本研究の意義を以下にまとめる．

- 歌手名を同定する際の難しさとして伴奏音混在の

問題を明確化した．この問題は，一部の研究では指摘されていたものの，これまで明示的には扱われていなかった．

- 伴奏音混在の問題を解決する一手法として，メロディの調波構造の抽出と再合成に基づく伴奏音抑制を提案した．認識対象音の調波構造を抽出・再合成することで雑音の影響を除去する試みは音声認識の分野では行われているが¹⁷⁾，この手法が歌手同定にも有効であることを確かめたのは，本研究が初めてである．
- 伴奏音混在の問題のさらなる解決策として，歌声らしさの高い（伴奏楽器の影響が少ない）フレームのみを用いる高信頼度フレーム選択手法を提案した．これに類似した処理は先行研究でも行われていたが，歌声区間の検出に焦点がおかれ，伴奏音の影響が少ないフレームのみを同定に用いるという観点には至っていなかった．本研究では，信頼度の観点からフレームを絞り込むことで，ロバストな歌手同定を実現した．

今後は，本手法を拡張することで，楽曲のボーカルの声質に基づく類似度を計算し，歌手の声質に基づく音楽情報検索を実現するための研究を進めていく．

謝辞 本研究の一部は，科学研究費補助金（基盤研究(A)，特定領域「情報学」），21世紀COEプログラム「知識社会基盤構築のための情報学拠点形成」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，ご討論いただいた吉井和佳氏（京都大学），吉岡拓也氏（NTT）に感謝する．

参 考 文 献

- 1) 松井知子：HMMによる話者認識，電子情報通信学会技術研究報告，SP95-111，pp.17-24（1996）．
- 2) 西田昌史，有木康雄：音韻性を抑えた話者空間への射影による話者認識，電子情報通信学会論文誌，Vol.J85-D-II，No.4，pp.554-562（2002）．
- 3) Tsai, W.-H., Wang, H.-M. and Rodgers, D.: Automatic singer identification of popular music recordings via estimation and modeling of solo vocal signal, *Proc. European Conference on Speech Communication and Technology (Eurospeech b2003)* (2003).
- 4) Rose, R.C., Hofstetter, E.M. and Reynolds, D.A.: Integrated Models of Signal and Background with Application to Speaker Identification in Noise, *IEEE Trans. Speech and Audio Processing*, Vol.2, No.2, pp.245-257（1994）．

- 5) Whitman, B., Flake, G. and Lawrence, S.: Artist Detection in Music with Minnowmatch, *Proc. 2001 IEEE Workshop on Neural Networks for Signal Processing*, pp. 559–568 (2001).
- 6) Berenzweig, A.L., Ellis, D.P.W. and Lawrence, S.: Using Voice Segments to Improve Artist Classification of Music, *AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio* (2002).
- 7) Kim, Y.E. and Whitman, B.: Singer identification in popular music recordings using voice coding features, *Proc. 3rd International Conference on Music Information Retrieval (ISMIR2002)*, pp.164–169 (2002).
- 8) Zhang, T.: Automatic Singer Identification, *Proc. IEEE International Conference on Multimedia & Expo (ICME 2003)*, Vol.1, pp.33–36 (2003).
- 9) 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).
- 10) 古井貞熙: 音声波に含まれる個人性情報の研究, 博士論文, 東京大学 (1978).
- 11) Picone, J.: Signal Modeling Techniques In Speech Recognition, *IEEE Proceedings*, Vol.81, No.9, pp.1215–1247 (1993).
- 12) Davis, S.B. and Mermelstein, P.: Comparison of parametric representation for monosyllabic word recognition, *IEEE Trans. Acoustic, Speech and Signal Processing*, Vol.28, No.4, pp.357–366 (1980).
- 13) 今井 聖: 音声信号処理音声の性質と聴覚の特性を考慮した信号処理, 森北出版 (1996).
- 14) Logan, B.: Mel frequency cepstral coefficients for music modelling, *Proc. International Symposium on Music Information Retrieval (ISMIR 2000)*, pp.23–25 (2000).
- 15) Shikano, K.: Evaluation of LPC spectral matching measures for phonetic unit recognition, Technical Report CMU-CS-96-108, CMU, Computer Science Department (1986).
- 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).
- 17) Nakatani, T. and Okuno, H.G.: Harmonic Sound Stream Segregation Using Localization and Its Application to Speech Stream Segregation, *Speech Communications*, Vol.27, pp.209–222 (1999).

(平成 17 年 10 月 17 日受付)

(平成 18 年 4 月 4 日採録)



藤原 弘将 (学生会員)

2005 年京都大学工学部情報学科卒業。現在, 同大学院修士課程在学中。音楽情報処理, 音声情報処理に興味を持つ。電子情報通信学会, 日本音響学会各学生会員。



北原 鉄朗 (学生会員)

2002 年東京理科大学理工学部情報科学科卒業。2004 年京都大学大学院情報学研究科知能情報学専攻修士課程修了。現在, 同大学院博士後期課程在学中。2005 年より日本学術振興会特別研究員 (DC2)。音楽情報処理に興味を持つ。2003 年情報処理学会第 65 回全国大会学生奨励賞, 2004 年情報処理学会第 66 回全国大会学生奨励賞, 電気通信普及財団第 19 回テレコムシステム技術学生賞, 2005 年情報処理学会第 67 回全国大会大会奨励賞, 日本音響学会関西支部第 8 回若手研究者交流研究発表会若手奨励賞各受賞。電子情報通信学会, 人工知能学会, 日本音響学会, 日本音楽知覚認知学会, IEEE 各学生会員。



後藤 真孝 (正会員)

1993 年早稲田大学理工学部電子通信学科卒業。1998 年同大学院理工学研究科博士後期課程修了。同年電子技術総合研究所 (2001 年に独立行政法人産業技術総合研究所に改組) に入所し, 現在に至る。2000~2003 年まで科学技術振興事業団さきがけ研究 21 「情報と知」領域研究員, 2005 年から筑波大学大学院システム情報工学研究科助教授 (連携大学院) を兼任。博士 (工学)。音楽情報処理, 音声言語情報処理等に興味を持つ。1997 年情報処理学会山下記念研究賞 (音楽情報科学研究会), 1999 年電気関係学会関西支部連合大会奨励賞, 2000 年 WISS2000 論文賞・発表賞, 2001 年日本音響学会粟屋潔学術奨励賞・ポスター賞, 2002 年情報処理学会山下記念研究賞 (音声言語情報処理研究会), 2002 年日本音楽知覚認知学会研究選奨, 2003 年インタラクシオン 2003 ベストペーパー賞, 2005 年情報処理学会論文賞等 18 件受賞。電子情報通信学会, 日本音響学会, 日本音楽知覚認知学会各会員。



駒谷 和範（正会員）

1998年京都大学工学部情報工学科卒業。2000年同大学大学院情報学研究科知能情報学専攻修士課程修了。2002年同大学院博士後期課程修了。京都大学博士（情報学）。同年

より京都大学大学院情報学研究科助手。情報処理学会平成16年度山下記念研究賞，FIT2002ヤングリサーチ賞受賞。電子情報通信学会，言語処理学会，人工知能学会，ACL各会員。



尾形 哲也（正会員）

1993年早稲田大学理工学部機械工学科卒業。日本学術振興会特別研究員，早稲田大学理工学部助手，理化学研究所脳科学総合研究センター研究員，京都大学大学院情報学研究

科講師を経て，2005年より同助教授。博士（工学）。この間，早稲田大学ヒューマノイド研究所客員助教授。人間とロボットのインタラクションと協調，神経回路モデル等の研究に従事。2000年度日本機械学会論文賞，IEA/AIE-2005最優秀論文賞等を受賞。RSJ，JSME，JSAI，IEEE等会員。



奥乃 博（正会員）

1972年東京大学教養学部基礎科学科卒業。日本電信電話公社，NTT，JST北野共生システムプロジェクト，東京理科大学理工学部情報科学科を経て，2001年4月より京都大学大学

院情報学研究科知能情報学専攻教授。博士（工学）。この間，スタンフォード大学客員研究員，東京大学工学部客員助教授。人工知能，音環境理解，ロボット聴覚の研究に従事。1990年度人工知能学会論文賞，IEA/AIE-2001，2005最優秀論文賞，IEEE/RSJ IROS-2001 Best Paper Nomination Finalist，第2回船井情報科学振興賞等受賞。人工知能学会，日本ソフトウェア科学会，日本認知科学会，日本ロボット学会，ACM，AAAI，IEEE，ASA，ISCA各会員。本学会英文図書出版委員。人工知能学会，日本ロボット学会評議員。『インターネット活用術』（岩波書店），『*Computational Auditory Scene Analysis*』（共編，LEA），『*Advanced Lisp Technology*』（共編，Taylor & Francis）ほか。