トピックモデルを用いた歌声特徴量の分析

中野 倫靖[1,a)]　吉井 和佳[1,b)]　後藤 真孝[1,c)]

概要：本稿では、複数の歌声から得られる音響特徴量をトピックモデルによって分析することで、歌声の特性を説明する新しい手法を提案する。従来、歌手の特性（性別や声種等）、歌い方の特性（声区や F_0 軌跡のモデル化等）、聴取印象（明るさ等）、楽曲の特性（楽曲ジャンルや歌詞等）を分析・推定したりする研究はあったが、複数の歌声から分かるような潜在的な意味を分析する研究はなかった。本稿では、伴奏と歌声を含む音楽音響信号から、歌声の線形予測メルケプストラム係数（LPMCC）と ΔF_0 を特徴量として自動推定した後、潜在的ディリクレ配分法（LDA）で分析を行う。LDA によって得られた潜在意味（トピック）の混合比が歌手名同定にも適用可能であることを示し、声道長の正規化に相当する処理を導入することで、性別を超えた類似歌手検索を実現することも示す。また、トピックの混合比を用いて、各トピックにおいて支配的な曲の歌手名をタグクラウドのように提示することで、トピックや歌声の意味を可視化する方法を提案する。

1. はじめに

楽曲にはジャンルやムードといった共通の特質を有する集合の概念（カテゴリー）があり、従来、ジャンルやムードとして様々な名称が定義されて、音楽音響信号からのジャンル識別 [1–5] やムード推定 [6–10] が研究されてきた。歌声も同様に、その声質や歌い方に応じた何らかのカテゴリーを形成できると予想できる。例えば、同じ楽曲ジャンルの曲や同じ曲を、別の歌手が歌った場合であっても、歌い方（歌声の音色[*1]や音高・音量の変化）に違いを感じたり、逆に似ていると感じることがある。

このような類似性において「どのように似ているのか」を説明することができれば、歌声に関する客観的理解を深めることを支援でき、音楽検索や音楽鑑賞、コミュニケーションの円滑化などに有用である。さらに、人間の音楽との関わり方の研究にも有用であり、例えば、歌声の聴取印象の分析や、特定の状況や場における人の選曲分析などにおいて、歌声の特性を説明する手段として活用できる。

従来、歌声を特徴付けたり説明したりする方法には、声種[*2]や性別などに関する「歌手の特性」、声区[*3]や歌声の F_0 軌跡のモデル化などに関する「歌い方の特性」、感情などに関する「聴取印象」、楽曲ジャンルや歌詞などに関する「楽曲の特性」の研究があった（**4** 章で後述）。本研究では、上記の特性に加えて、複数の歌声から分かるような、各歌声の潜在的な特性を分析する方法の実現を目的とする。

本稿では、複数の楽曲それぞれから推定した歌声特徴量に基づいて、トピックモデルによる分析を行う。すなわち、各曲の歌声が潜在意味（トピック）に基づいて生成される過程を確率的に表現する。ここで、トピックモデルを用いることにより、各歌声に内在する隠れた構造を抽出するとともに、歌声間の類似度を算出できることも示す。従来、潜在的な構造を音楽や歌声から推定して利用する研究例としては、潜在的ディリクレ配分法（LDA）を用いた歌詞と旋律による楽曲検索 [18]、低音旋律からのジャンル分類 [19]、LDA による調推定 [20,21]、楽曲の音響特徴量とブログや歌詞の文字の対応付け [22]、ソーシャルタグによる楽曲推薦 [23]、階層ディリクレ過程（HDP）を用いた楽曲

[1] 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)
[a)] t.nakano [at] aist.go.jp
[b)] k.yoshii [at] aist.go.jp
[c)] m.goto [at] aist.go.jp

[*1] 音高と音量以外の音響的な成分という意味で、この用語を用いる。具体的には、異なる発声様式によって生じる声の違い（唸り声、囁き声等）など、励振音源やスペクトル包絡の特性に相当する。

[*2] 西洋音楽（声楽やオペラ）におけるソプラノやアルト等の分類のこと。声種という用語は文献 [11,12] を参考にした（英語では、voice category [13]、voice type [11,14,15]、等）。声種は解剖学的構造によって決められることが多く、歌手自身の願望や訓練によって変えることは難しいことが多い [11] ことから、声種は「歌手の特性」といえる。声の音色が参考になることもあるが、主に声域（発声可能な音域の広さ）によって決定され、声帯や声道が短いと高い声種、長いと低い声種になりがちとなる [11] 。

[*3] 同種の調節機構で発声された同種の声質をもつ音の系列であり [16]、声区の分類や名称については様々な解釈がある [17] が、歌声においてはフライ (vocal fly)・地声 (modal)・裏声 (falsetto)・ホイッスル (whistle) の 4 種類の声区が存在するとされている.

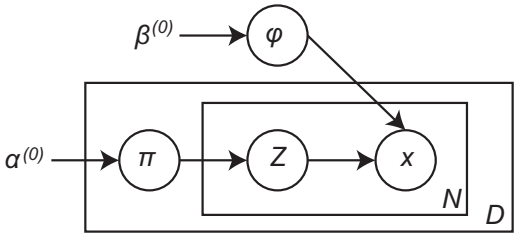

図 1 LDA のグラフィカルモデル

間類似度の推定 [24] 等があり、自己組織化マップ（SOM）に基づく楽曲のクラスタリング [25] も一種の潜在的な意味解析といえる。しかし、歌声の音響特徴量を対象とした研究はなかった。

本稿では、まず歌手名同定に関する Fujihara *et al.* の研究 [26] と同様に、伴奏と歌声を含む音楽音響信号から、歌声の線形予測メルケプストラム係数（LPMCC）と、基本周波数（F_0）の時間変化である ΔF_0 を歌声特徴量として自動推定する。ただし、歌手の性別の違いによる声道長の違いによる影響を除去するために、声道長の正規化に相当する処理を導入することで、性別を超えた歌い方の分析を行う。また、楽曲のテンポによる歌唱速度も、同様に正規化に相当する処理を導入して違いを吸収する。

続いて、そのようにして得られた歌声特徴量から、潜在的ディリクレ配分法（LDA）で分析を行う。ここで、各歌声のトピックの混合比や各トピックの確率分布から、支配的な曲の歌手名をタグクラウドのように提示することで、歌声やトピックの意味を可視化する手法を提案する。また、このような各曲のトピックの混合比が類似歌手検索にも適用可能であること、声道長の正規化によって性別を超えた類似歌手検索も実現できることを示す。

2. トピックモデルを用いた歌声特徴量の分析

本章では、伴奏と歌声を含む音楽音響信号を対象として、その歌声特徴量を歌手の性別の違いや楽曲のテンポの違いをなるべく吸収するように推定し、潜在的ディリクレ配分法（Latent Dirichlet Allocation: LDA）[27] によって歌い方を分析する。

2.1 声道長及び速度を変化させた歌声の生成

歌手の性別の違いによる声道長の違いや、楽曲のテンポの違いを吸収した分析結果を得るために、それらの正規化に相当する処理を行う。ただし、一つの歌声特徴量を一つの正規化された特徴量へ変換（正規化）するのではなく、一つの歌声を様々な声道長や速度へ変化させた音響信号を生成し、それらを独立した一つの歌声として扱ってモデル学習する。これによって、検索クエリに似せるように検索対象を変形させるなどの、新しい楽曲検索が実現できる。

具体的には、短時間周波数分析の周波数軸方向へのシフトによって音高シフトを実現し、声道長を伸縮させたことに相当する歌声を生成する。また、WSOLA（Waveform Similarity Based Overlap-Add）アルゴリズムによってテンポシフトを実現し、速度を変更させた歌声を生成する。本稿では、そのような歌声を sox[*4] を用いて生成した。

2.2 歌声特徴量の抽出

歌声特徴量の推定は、我々が開発した能動的音楽鑑賞サービス Songle [28] のモジュールを用いて行った。具体的には、混合音中で最も優勢な音高を推定する手法 PreFEst [29] によってボーカルのメロディーを推定し、歌声・非歌声 GMM を用いた高信頼度フレーム選択によって、歌声らしさが高いフレームを選択し、LPMCC と ΔF_0 を特徴量として推定した [26]。最後に、全特徴ベクトルについて、次元毎に平均を引いて標準偏差で割る正規化を行った。

ここで、ある歌手が別の歌手の歌い方を真似る際にもスペクトル包絡形状が変化すると報告されている [30, 31] ことから、LPMCC のような特徴量は、歌手同定に重要である [26] だけでなく、「歌い方」を議論する上でも同様に重要な特徴量であると考えられる。

2.3 *k*-means 法による歌声特徴量の離散化

本稿では、**2.4** 節で述べるように、文書のような離散化されたシンボルの系列で構成される歌声を対象として LDA で分析を行う。しかし、**2.2** 節の処理で得られる各特徴ベクトルは連続的な値を持つため、それぞれのベクトルを離散化された一つのシンボルに割り当てる。そのために本稿では、k-means アルゴリズムを用いて実装した。

2.4 LDA: Latent Dirichlet Allocation

LDA におけるモデル学習用のデータとして D 個の独立した歌声 $\boldsymbol{X} = \{\boldsymbol{X}_1, ..., \boldsymbol{X}_D\}$ を考える。ここで扱う歌声は、離散化されたシンボルの系列であるため、通常の LDA [27] と同様の枠組みで歌声を分析できる。

歌声 $\boldsymbol{X}_d$ は、高信頼度フレーム（**2.2** 節）の長さ N_d を持つシンボル系列であり、$\boldsymbol{X}_d = \{\boldsymbol{x}_{d,1}, ..., \boldsymbol{x}_{d,N_d}\}$ で構成されている。ここで、シンボルの語彙サイズ V は、k-means 法（**2.3** 節）におけるクラスタ数に相当し、$\boldsymbol{x}_{d,n}$ は語彙中から選ばれたシンボルに対応する次元のみが 1 で他は 0 である V 次元ベクトルとなる。

歌声 $\boldsymbol{X}_d$ に対応する潜在変数系列（トピック系列）を $\boldsymbol{Z}_d = \{\boldsymbol{z}_{d,1}, ..., \boldsymbol{z}_{d,N_d}\}$ とする。トピック数を K とすると、$z_{d,n}$ は選ばれたトピックに対応する次元のみが 1 で他は 0 である K 次元のベクトルで表せる。ここで、全歌声の潜在変数系列をまとめて $\boldsymbol{Z} = \{\boldsymbol{Z}_1, ..., \boldsymbol{Z}_D\}$ としておく。このとき、グラフィカルモデル（**図 1**）から変数間の条件つき独立性を考慮すると、完全な同時分布は

[*4] http://sox.sourceforge.net/

表 1 LDA の学習に用いた使用楽曲

ID	歌手名	性別	曲数
M1	ASIAN KUNG-FU GENERATION	男	3
M2	BUMP OF CHICKEN	男	3
M3	福山雅治	男	3
M4	GLAY	男	3
M5	氷川きよし	男	3
M6	平井堅	男	3
F1	aiko	女	3
F2	JUDY AND MARY	女	3
F3	一青窈	女	3
F4	東京事変	女	3
F5	宇多田ヒカル	女	3
F6	矢井田瞳	女	3

$$p(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\pi},\boldsymbol{\phi}) = p(\boldsymbol{X}|\boldsymbol{Z},\boldsymbol{\phi})p(\boldsymbol{Z}|\boldsymbol{\pi})p(\boldsymbol{\pi})p(\boldsymbol{\phi}) \quad (1)$$

で与えられる。ここで、$\boldsymbol{\pi}$ は各歌声におけるトピックの混合比（D 個の K 次元ベクトル）であり、$\boldsymbol{\phi}$ は各トピックにおけるユニグラム確率（K 個の V 次元ベクトル）である。最初の二項には多項分布に基づく離散分布を仮定する。

$$p(\boldsymbol{X}|\boldsymbol{Z},\boldsymbol{\phi}) = \prod_{d=1}^{D}\prod_{n=1}^{N_d}\prod_{v=1}^{V}\left(\prod_{k=1}^{K}\phi_{k,v}^{z_{d,n,k}}\right)^{x_{d,n,v}} \quad (2)$$

$$p(\boldsymbol{Z}|\boldsymbol{\pi}) = \prod_{d=1}^{D}\prod_{n=1}^{N_d}\prod_{v=1}^{V}\pi_{d,k}^{z_{d,n,k}} \quad (3)$$

残りの二項には、多項分布の共役事前分布であるディリクレ分布を仮定する。

$$p(\boldsymbol{\pi}) = \prod_{d=1}^{D}\mathrm{Dir}(\boldsymbol{\pi}_d|\boldsymbol{\alpha}^{(0)}) = \prod_{d=1}^{D}C(\boldsymbol{\alpha}^{(0)})\prod_{k=1}^{K}\pi_{d,k}^{\alpha^{(0)}-1} \quad (4)$$

$$p(\boldsymbol{\phi}) = \prod_{k=1}^{K}\mathrm{Dir}(\boldsymbol{\phi}_k|\boldsymbol{\beta}^{(0)}) = \prod_{k=1}^{K}C(\boldsymbol{\beta}^{(0)})\prod_{v=1}^{V}\phi_{k,v}^{\beta^{(0)}-1} \quad (5)$$

ここで、$\boldsymbol{\alpha}^{(0)}$ および $\boldsymbol{\beta}^{(0)}$ はハイパーパラメータ、$C(\boldsymbol{\alpha}^{(0)})$ 及び $C(\boldsymbol{\beta}^{(0)})$ はディリクレ分布の正規化定数であり、

$$C(\boldsymbol{x}) = \frac{\Gamma(\hat{x})}{\Gamma(x_1)\cdots\Gamma(x_I)}, \quad \hat{x} = \sum_{i=1}^{I}x_i \quad (6)$$

である。

2.5 実験条件

本稿では、歌声を含む音楽音響信号を全て 16kHz のモノラル信号に変換し、**表 1** に示した楽曲を用いて分析した。これは、日本の音楽チャートであるオリコン[*5]で 2000〜2008 年までの上位 20 位以内に登場した楽曲の中から、ボーカルが一人のアーティストを男女 6 アーティストずつ選び、それぞれのアーティストから 3 曲ずつを選んだ。

音高シフトは −3〜+3 半音を 1 半音単位でシフトして 7 種類、テンポシフトは 0.7〜1.3 倍速を 0.1 倍速単位でシフトして 7 種類をそれぞれ行った。従って、一つの歌声から音高とテンポをそれぞれシフトさせた 49(= 7 × 7) 曲の歌声を生成し、合計で $D = 1764(= 49 \times 3 \times 12)$ 曲を用いた。

歌声特徴量は、楽曲の冒頭 1 分間のうち、歌声らしさが高い上位 15%のフレームから推定し、クラスタ数 $V = 100$ として k-means 法によるクラスタリングを行った。

LDA の学習においては、トピック数を $K = 100$ として、周辺化 Gibbs サンプラーを用いて学習した。ハイパーパラメータ $\boldsymbol{\alpha}^{(0)}$ については初期値をすべて 1 として、経験ベイズ法を用いて最適化を行い [32]、ハイパーパラメータ $\boldsymbol{\beta}^{(0)}$ の値はすべて 0.1 とした。

*5 http://www.oricon.co.jp/

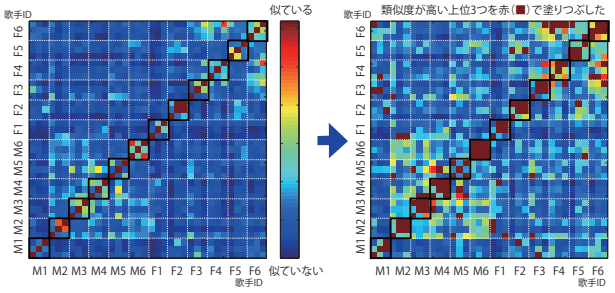

図 2 各歌声におけるトピックの混合比の類似度行列

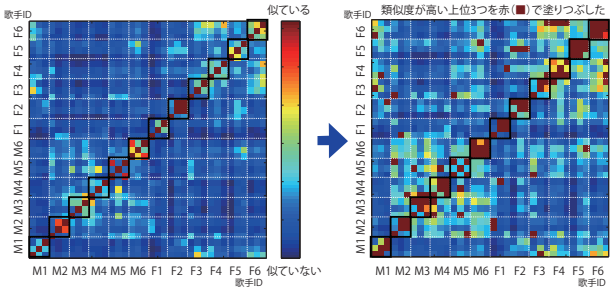

図 3 各歌声におけるトピックの混合比の類似度行列（ハイパーパラメータ $\boldsymbol{\alpha}^{(0)}$ を更新しなかった場合）

2.6 実験及び結果

上述のような歌声データから学習された LDA のモデルの正当性を確認するために、推定されたトピックの混合比に基づいた歌声間類似度を確認する。ここでは、音高シフトやテンポシフトを行わない 36(= 12 × 3) 曲について、それらの類似度行列を**図 2**（左）に示し（色が赤いほど類似している）、類似度が高い上位 3 曲について赤く塗りつぶした図をその右に示した。ここで、歌声 A におけるトピックの混合比を $\boldsymbol{\pi}_A$、歌声 B におけるトピックの混合比を $\boldsymbol{\pi}_B$ とした時、式 (7) に示す対称カルバック・ライブラ距離（symmetric Kullback Leibler distance, KL2）を算出し、その逆数を類似度とした。

$$d_{\mathrm{KL2}}(\boldsymbol{\pi}_A||\boldsymbol{\pi}_B) = \sum_{k=1}^{K}\pi_A(k)\log\frac{\pi_A(k)}{\pi_B(k)} + \sum_{k=1}^{K}\pi_B(k)\log\frac{\pi_B(k)}{\pi_A(k)} \quad (7)$$

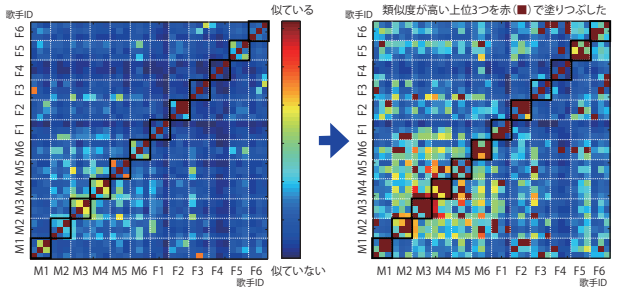

図 4 1000 回反復したトピック混合比の類似度行列

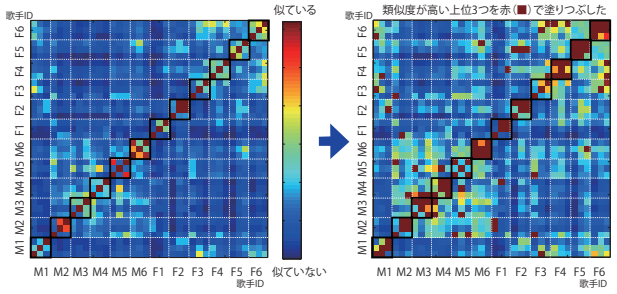

図 5 1000 回反復したトピック混合比の類似度行列（ハイパーパラメータ $\boldsymbol{\alpha}^{(0)}$ を更新しなかった場合）

ただし、π_A と π_B はそれぞれ

$$\sum_{k=1}^{K} \pi_A(k) = 1, \quad \sum_{k=1}^{K} \pi_B(k) = 1 \qquad (8)$$

のように正規化して、確率分布として扱う。

また、参考のために、ハイパーパラメータ $\boldsymbol{\alpha}^{(0)}$ を更新しなかった場合（**図 3**）、1000 回反復した場合（**図 4**）、1000 回反復してハイパーパラメータ $\boldsymbol{\alpha}^{(0)}$ を更新しなかった場合（**図 5**）の結果を同様に示す。推定結果に大きな違いはなかったが、全体的に誤識別が含まれているため、特徴抽出や学習条件に今後の検討の余地がある。

図 2（右）や図 3（右）等からは、同一アーティストの曲のみが主に赤く塗りつぶされていて（類似度が上位 3 位以内）、同一アーティスト間ではトピックの混合比が類似していることが分かり、LDA が適切に動作していると考えられる。この結果から、歌声特徴量を LDA で分析することで得られる各歌声におけるトピックの混合比は、歌手名同定に適用できる可能性があるといえる。

3. 歌声トピックモデルを活用する 2 つの手法

2.6 節までのようにして学習した LDA は、そこで示したように歌手名同定に有用である。本章では、その結果をさらに活用する 2 種類の手法を提案する。

一つ目には、音高シフトやテンポシフトした歌声を含めて歌手名同定を行う、速度の違いを抑制した「性別を超えた類似歌声検索」を提案する。

また、二つ目は、各トピックの意味を単語クラウドによって可視化する「トピックの可視化」を提案し、自分好みのトピックを見つけることを支援する。トピックの意味が分かれば、式 (7) の類似度算出において特定のトピックに重みをかけて類似歌手を検索する等の応用が可能になる。

表 2 性別を超えた類似歌声検索の結果。音高シフトとテンポシフトなしの歌声のトピック混合比をクエリとして、自身以外で最も近かった別の歌手名と音高シフト及びテンポシフトの値。下線は異性同士で、**太字**は同性同士で顕著に似ていた歌手。

検索クエリ (±0/×1)	自身以外で最も類似していた歌手		
	1 曲目	2 曲目	3 曲目
M1	F4(−3/×0.7)	F5(−3/×0.8)	M3(−3/×1)
M2	M4(−1/×0.8)	M3(+1/×1.1)	M3(±0/×1.3)
M3	F3(−3/×1.1)	M4(+1/×1.2)	M4(−2/×1)
M4	M1(±0/×1.1)	M3(−1/×1)	F2(+2/×1.2)
M5	M2(+1/×1.2)	F5(−2/×0.8)	M1(+1/×1.1)
M6	F3(−3/×0.9)	F3(−3/×1.2)	F5(−2/×0.7)
F1	F5(+2/×0.8)	F5(+1/×0.8)	F3(+1/×1)
F2	M1(−1/×0.9)	F6(+3/×0.8)	F6(+3/×0.9)
F3	M6(+3/×1.1)	M6(+3/×1.2)	M6(+2/×1)
F4	**F6**(−1/×1.1)	M1(+2/×0.8)	**F6**(+1/×1.3)
F5	F6(−2/×0.8)	M6(+3/×1.1)	M5(+1/×1.1)
F6	F2(−3/×0.9)	**F4**(+1/×0.7)	**F4**(±0/×0.8)

3.1 性別を超えた類似歌声検索手法

2.1 節で述べたように、検索対象の楽曲を音高シフトとテンポシフトによって「仮想的に増やす」ことで、速度の違いを吸収して、性別を超えた類似歌声検索が可能となる。検索対象が増える以外は、**2.6** 節と同様、トピックの混合比間の類似度を計算すれば良い。

表 2 に、表 1 のそれぞれの歌声を検索クエリとして、自身以外で最も類似度が高かった歌手 ID と、その音高シフト及びテンポシフトの値を示す。下線は異性同士で顕著に似ていた歌手同士、太字は同性同士で顕著に似ていた歌手同士を示す。この表からは、「平井堅（M6）を 2〜3 半音上げるか、一青窈（F3）を 2〜3 半音下げると、お互いに類似している」ことや、「東京事変（F4）を 0〜1 半音上げて 0.7〜0.8 倍速にするか、矢井田瞳（F6）を 0〜1 半音上げて 1.1〜1.3 倍速にすると、お互いに類似している。」ことが分かる。実際の聴取印象もそのようであった。また、特に平井堅と一青窈とが、3 半音程度の音高シフトで類似することは一般的によく知られた事例であるため、それが確認できた点からも手法の有効性を示せた。

ここで、**図 6** に一青窈の歌声「もらい泣き」と、それに最も類似していた平井堅の歌声「思いがかさなるその前に…」を 3 半音上げて 1.1 倍速した歌声について、それぞれトピックの混合比を示す。両者共にトピック 28,32,82 に関する特性の歌声だということが分かる。

ただし、現在は小規模なデータセット（12 人 ×3 曲）による結果のため、詳細な議論は今後、より大きなデータセットでの実験で行う必要がある。データセットを拡大することで、新たな類似歌手を発見できる可能性がある。

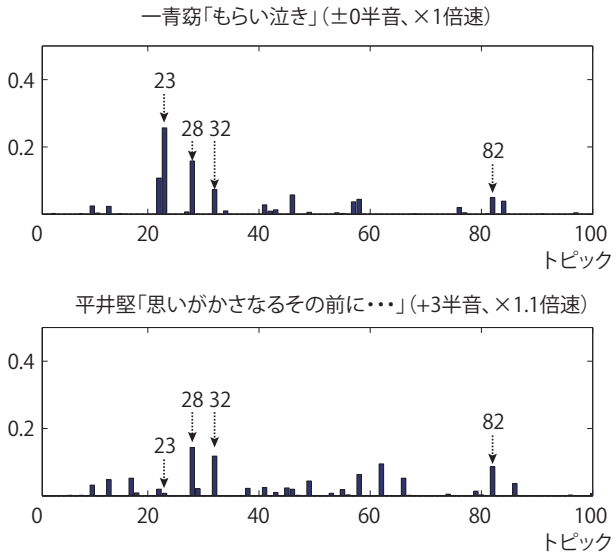

図 6 類似していた一青窈と平井堅 (+3/×1.1) におけるそれぞれのトピックの混合比。28,32,82 のトピックに類似が見られ、23 は一青窈の楽曲にのみ現れるトピック。

3.2 単語クラウドによる曲とトピックの可視化手法

各歌声における混合比 $\boldsymbol{\pi}_{dk}$ は、D 個（歌声数）の K 次元ベクトル（トピック数）であり、「各歌声 d における支配的なトピック k が分かる」ことを意味する。本稿ではこの性質を利用して歌手の同定や類似歌手の検索を行った。しかし、図 6 に示したようにトピックの混合比だけでは、それぞれのトピックの意味が分からないという問題がある。

そこで逆に「トピックの混合比から各トピック k における支配的な歌声 d が分かる」ことを考えて、各トピックの意味を可視化する手法を提案する。ここで本稿では、ウェブサイト上で使用されるタグの視覚的記述の一つである「タグクラウド」を応用し、各トピックに支配的な歌手ほど大きく表示する「歌手クラウド」によってトピックの意味を提示する方法を提案する（**図 7**）。歌手名の羅列に比べて一覧性が高くて有用である。図 7 は、同じ音高シフトの同じ曲についてトピックの混合比を足し合わせ（テンポの違いを無視）、その値に応じて歌手名のサイズを変えて生成した。また、音高シフトの値を歌手名の横に提示した。

図 7 からは、類似性が高かったトピック 28 は一青窈の歌声を −1 半音シフトした歌声が支配的であった。次いで類似性が高かったトピック 32 や 82 は、宇多田ヒカルを −3 半音シフトした歌声や東京事変、矢井田瞳などの歌声のような特性を持っていることが分かる。

逆に、一青窈の歌声にしか現れなかったトピック 23 は、一青窈を −3 半音シフトした歌声に加え ASIAN KUNG-FU GENERATION、JUDY AND MARY、aiko、などといった、前者と異なった歌声の特性を持っていると推測できる。

このような歌手クラウドも、データセットを増やすことで、視覚的な印象が変わってくると考えられる。

4. 歌声の特性の自動推定に関する従来研究

従来、歌声を特徴付ける特性として、歌手の特性、歌い方の特性、聴取印象、楽曲の特性に関する研究があった。本章では特に、それらの自動推定やモデル化に関する研究を紹介し、本研究との差分を明確にする。

歌手の特性 声種推定 [14, 15]、性別推定 [33–35]、年齢推定 [35]、身体サイズ推定 [35]、人種推定 [35] といった自動推定に関する研究がある[*6]。

歌い方の特性 声区推定 [36, 37] 、歌声の合成モデル（F_0 や音量、スペクトル包絡の制御）[38–45] と、F_0 軌跡の分析モデル [46] に関する研究がある。また、ブレス（吸気）の時刻や特性は歌い方と関係すると考えられるが、ブレス検出 [47, 48] に関する研究がある。

聴取印象 ポピュラー音楽における歌声の印象評価語の推定 [49] がある。その他、信号処理からは少し離れるが、歌詞からの感情推定 [50] が研究されている。歌唱力を聴取印象の一種として捉えると、歌唱力の自動評価 [14, 15, 51–55] があり、熱唱度という聴取印象を定義して推定する研究 [56] もある。

楽曲の特性 楽曲のジャンルには特有の歌い方が存在するため、楽曲ジャンルへの自動分類（声楽家女性/ポップス歌手女性 [57]）が研究されている。また、歌詞の言語の自動推定 [58–61] がある。

以上のように、個々の歌声をモデル化したり、分析する研究が多くあるが、本研究のように、複数の歌声から分かる潜在的な特性を分析する研究はなかった。

5. おわりに

本稿では、伴奏と歌声を含む音楽音響信号から歌声特徴量を推定し、潜在的ディリクレ配分法（LDA）で分析を行った。ここで、各歌声におけるトピックの混合比が歌手名同定や類似歌手検索へ応用できることを示し、音高シフトとテンポシフトによって、性別やテンポの違いを吸収した類似歌手検索を提案した。また、トピックの意味を歌手クラウドで可視化する方法についても提案した。

本手法は、F_0 軌跡の変化に関する特徴など、様々な歌声特徴量を適用できるため、今後は特徴量を検討したい。また本稿での LDA は、離散値を扱う実装であったが、連続値である歌声特徴量を扱うために、LDA の連続値への拡張も考えられる[*7]ため、必要があれば導入したい。

謝辞 本研究の一部は JST CREST「OngaCREST プロジェクト」の支援を受けました。また、歌声特徴量の推定において、藤原 弘将 氏による Songle のモジュールを使用させて頂きました。濱崎 雅弘 氏と石田 啓介 氏には歌手

[*6] ここで、年齢は young か old、身体サイズは short か tall と大まかな推定であった。性別では duet を推定する研究もある。
[*7] 例えば、文献 [21, 62, 63] が参考になる。

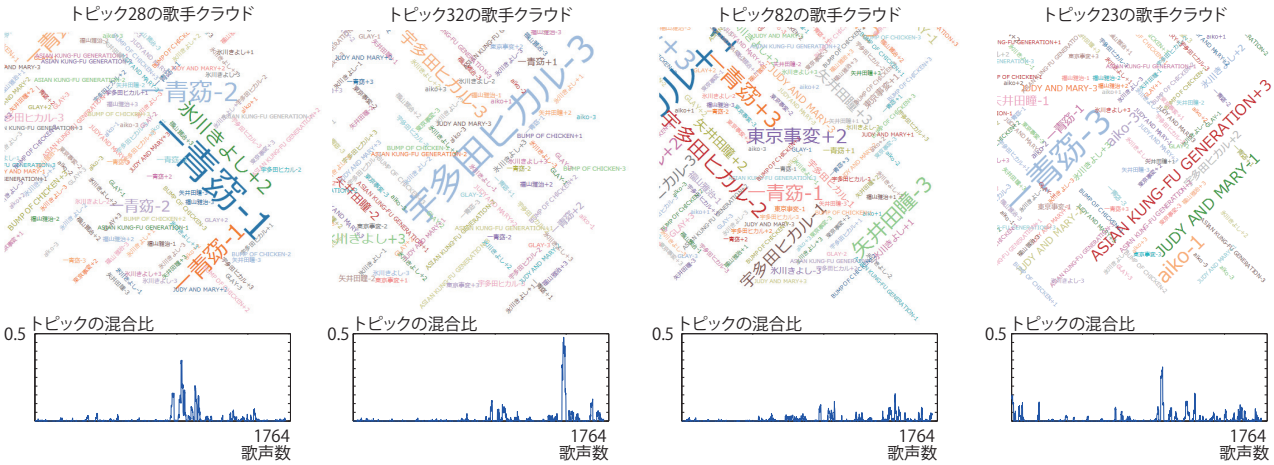

図 7 歌手クラウドによるトピックの意味を視覚提示する例。図 6 で示したトピック 28, 32, 82（一青窈と平井堅に共通のトピック）、及びトピック 23（一青窈にのみ現れるトピック）の歌手クラウドの例（上部）とトピックの混合比（下部）。

クラウドに関するご助言を頂きました。感謝致します。

参考文献

[1] Tzanetakis, G. and Cook, P.: Musical Genre Classification of Audio Signals, *IEEE Trans. on Speech and Audio Processing*, Vol. 17, No. 2, pp. 293–302 (2002).
[2] Tsuchihashi, Y., Kitahara, T. and Katayose, H.: Using Bass-line Features for Content-based MIR, *Proc. of ISMIR2008*, pp. 620–625 (2008).
[3] Holzapfél, A. and Stylianou, Y.: Musical Genre Classification Using Nonnegative Matrix Factorization-Based Features, *IEEE Trans. on ASLP*, Vol. 16, No. 2, pp. 424–434 (2008).
[4] Costa, Y. M. G., Oliveira, L. S., Koerich, A. L., Gouyon, F. and Martins, J. G.: Music Genre Classification Using LBP Textural Features, *Signal Processing*, Vol. 92, No. 11, pp. 2723–2737 (2012).
[5] Ren, J.-M. and Jang, J.-S. R.: Discovering Time-Constrained Sequential Patterns for Music Genre Classification, *IEEE Trans. on ASLP*, Vol. 20, No. 4, pp. 1134–1144 (2012).
[6] Lu, L., Liu, D. and Zhang, H.-J.: Automatic Mood Detection and Tracking of Music Audio Signals, *IEEE Trans. on ASLP*, Vol. 164, No. 1, pp. 5–18 (2006).
[7] 藤澤隆史，谷　光彬，長田典子，片寄晴弘：和音性の定量的評価モデルに基づいた楽曲ムードの色彩表現インタフェース，情報処理学会論文誌，Vol. 50, No. 3, pp. 1133–1138 (2009).
[8] Tsunoo, E., Akase, T., Ono, N. and Sagayama, S.: Music Mood Classification by Rhythm and Bass-line Unit Pattern Analysis, *Proc. of ICASSP 2010* (2010).
[9] Yang, Y.-H. and Chen, H. H.: Machine Recognition of Music Emotion: A Review, *ACM Trans. Intelligent Systems and Technology*, Vol. 3, No. 3, pp. 1–30 (2012).
[10] 三好真人，柘植　覚，福見　稔：エネルギー変化の線形予測符号化に基づくリズム特徴量を用いた音楽印象識別，情報処理学会論文誌，Vol. 54, No. 4, pp. 1275–1287 (2013).
[11] 平野　実：歌声の評価，声の検査法: 臨床編，pp. 215–219 (1979).
[12] Sundberg, J.: 歌声の科学，東京電機大学出版局 (2007).
[13] Sundberg, J.: *The Science of the Singing Voice*, Northern Illinois University Press (1987).
[14] Zwan, Pawel; Kostek, B.: System for Automatic Singing Voice Recognition, *J. Audio Eng. Soc*, Vol. 56, No. 9, pp. 710–723 (2008).
[15] Maazouzi, F. and Bahi, H.: Singing Voice Classification in Commercial Music Productions, *Proc. of ICICS* (2011).
[16] 日本音響学会編：新版 音響用語辞典，コロナ社 (2003).
[17] Sakakibara, K.-I.: Production Mechanisim of Voice Quality in Singing, *J. Phonetic Soc. Jpn.*, Vol. 7, No. 3, pp. 27–39 (2003).
[18] Brochu, E. and de Freitas, N.: "Name That Song!": A Probabilistic Approach to Querying on Music and Text, *Proc. of NIPS2002* (2002).
[19] 上田　雄，角尾衣未留，小野順貴，嵯峨山茂樹：低音旋律の潜在意味解析による音楽ジャンル分類，日本音響学会春季研究発表会講演集，pp. 875–876 (2009).
[20] Hu, D. J. and Saul, L. K.: A Probabilistic Topic Model for Unsupervised Learning of Musical Key-Profiles, *Proc. of ISMIR2009* (2009).
[21] Hu, D. J. and Saul, L. K.: A Probabilistic Topic Model for Music Analysis, *Proc. of NIPS-09* (2009).
[22] Takahashi, R., Ohishi, Y., Kitaoka, N., and Takeda, K.: Building and Combining Document and Music Spaces for Music Query-By-Webpage System, *Proc. of Interspeech 2008*, pp. 2020–2023 (2008).
[23] Symeonidis, P., Ruxanda, M. M., Nanopoulos, A. and Manolopoulos, Y.: Ternary Semantic Analysis of Social Tags for Personalized Music Recommendation, *Proc. of ISMIR2008*, pp. 219–224 (2008).
[24] Hoffman, M., Blei, D. and Cook, P.: Content-Based Musical Similarity Computation Using the Hierarchical Dirichlet Process, *Proc. of ISMIR2008* (2008).
[25] Pampalk, E.: Islands of Music: Analysis, Organization, and Visualization of Music Archives, *Master's thesis, Vienna University of Technology* (2001).
[26] Fujihara, H., Goto, M., Kitahara, T. and Okuno, H. G.: A Modeling of Singing Voice Robust to Accompaniment Sounds and Its Application to Singer Identification and Vocal-Timbre-SimilarityBased Music Information Retrieval, *IEEE Trans. on ASLP*, Vol. 18, No. 3, pp. 638–648 (2010).
[27] Blei, D. M., Ng, A. Y. and Jordan, M. I.: Latent Dirichlet Allocation, *Journal of Machine Learning Research*, Vol. 3, pp. 993–1022 (2003).
[28] 後藤真孝，吉井和佳，藤原弘将，Mauch, M.，中野倫靖：Songle: ユーザが誤り訂正により貢献可能な能動的音楽

鑑賞サービス，インタラクション 2012 講演論文集，pp. 1–8 (2012).
[29] 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).
[30] 鈴木千文，坂野秀樹，板倉文忠，森勢将雅：歌唱音声の類似度評価を目的とした声質に関する音声特徴量の提案，電子情報通信学会技術研究報告 SP，Vol. 111, No. 364, pp. 79–84 (2011).
[31] 齋藤　毅，榊原健一：歌唱時の物真似による音響特徴の変化，聴覚研究会資料 (2011).
[32] Minka, T. P.: Estimating a Dirichlet distribution, *technical report (Massachusetts Inst. of Technology)*, pp. 1–8 (2000).
[33] Schuller, B., Kozielski, C., Weninger, F., Eyben, F. and Rigoll, G.: Vocalist Gender Recognition in Recorded Popular Music, *Proc. of ISMIR 2010*, pp. 613–618 (2010).
[34] Weninger, F., Durrieu, J.-L., Eyben, F., Richard, G. and Schuller, B.: Combining monaural source separation with Long Short-Term Memory for increased robustness in vocalist gender recognition, *Proc. of ICASSP 2011*, pp. 2196–2199 (2011).
[35] Weninger, F., Wöllmer, M. and Schuller, B.: Automatic Assessment of Singer Traits in Popular Music: Gender, Age, Height and Race, *Proc. of ISMIR 2011* (2011).
[36] 平山健太郎，伊藤克亘：ポピュラー歌唱における高音域の声区と発声状態の判別手法，情報処理学会研究報告音楽情報科学研究会，2012-MUS-94, Vol. 2012-SLP-90, No. 16, pp. 1–6 (2012).
[37] 小島　俊，齋藤　毅，中野倫靖，後藤真孝，三好正人：歌声における裏声と地声を識別すための音響特徴量の検討，電子情報通信学会技術研究報告 EA，EA2012-76, Vol. 112, No. 266, pp. 67–72 (2012).
[38] Mori, H., Odagiri, W. and Kasuya, H.: F_0 Dynamics in Singing: Evidence from the Data of a Baritone Singer, *IEICE Trans. Inf. & Syst.*, Vol. E87-D, No. 5, pp. 1068–1092 (2004).
[39] Nobuaki, M., Bungo, M. and Keikichi, H.: Prosodic Analysis and Modeling of Nagauta Singing to Generate Prosodic Contours from Standard Scores, *IEICE Trans. Information and Systems*, Vol. E87-D, No. 5, pp. 1093–1101 (2004).
[40] 齋藤　毅，辻　直也，鵜木祐史，赤木正人：歌声らしさの知覚モデルに基づいた歌声特有の音響特徴量の分析，日本音響学会論文誌，Vol. 64, No. 5, pp. 267–277 (2008).
[41] Gómez, E. and Bonada, J.: Automatic Melodic Transcription of Flamenco Singing, *Proc. of CIM08* (2008).
[42] Ohishi, Y., Kameoka, H., Mochihashi, D. and Kashino, K.: A Stochastic Model of Singing Voice F0 Contours for Characterizing Expressive Dynamic Components, *Proc. of INTERSPEECH 2012* (2012).
[43] 森勢将雅，村主大輔，馬場　隆，片寄晴弘：奄美大島民謡風の歌唱デザインを支援するシステム:グインレゾネータ，情報処理学会論文誌，Vol. 54, No. 4, pp. 1244–1253 (2013).
[44] Lee, S. W., Dong, M. and Chan, P. Y.: Analysis for Vibrato with Arbitrary Shape and its Applications to Music, *Proc. of APSIPA ASC 2011* (2011).
[45] Stables, R., Athwal, C. and Bullock, J.: Fundamental Frequency Modulation in Singing Voice Synthesis, *Lecture Notes in Computer Science*, Vol. 7172, pp. 104–119 (2012).
[46] 大石康智，後藤真孝，伊藤克亘，武田一哉：相平面に描かれる歌声の基本周波数軌跡：歌唱者の意図する音高目標値系列の推定とハミング検索への応用，情報処理学会論文誌，Vol. 49, No. 11, pp. 3789–3797 (2008).
[47] Ruinskiy, D. and Lavner, Y.: An Effective Algorithm for Automatic Detection and Exact Demarcation of Breath Sounds in speech and song signals, *IEEE Trans. on ASLP*, Vol. 15, pp. 838–850 (2007).
[48] Nakano, T., Ogata, J., Goto, M. and Hiraga, Y.: Analysis and automatic detection of breath sounds in unaccompanied singing voice, *Proc. 10th International Conference of Music Perception and Cognition (ICMPC 10)* (2008).
[49] 金礪　愛，中野倫靖，後藤真孝，菊池英明：ポピュラー音楽における歌声の印象評価語を自動推定するシステム，情報処理学会研究報告音楽情報科学研究会，2013-MUS-100 (2013).
[50] Hu, Y., Chen, X. and Yang, D.: Lyric-based Song Emotion Detection with Affective Lexicon, *Proc. ISMIR2009* (2009).
[51] 中野倫靖，後藤真孝，平賀　譲：楽譜情報を用いない歌唱力自動評価手法，情報処理学会論文誌，Vol. 48, No. 1, pp. 227–236 (2007).
[52] Prasert, P., Iwano, K. and Furui, S.: An Automatic Singing Voice Evaluation Method for Voice Training Systems, 音講論集，pp. 911–912 (2008.3).
[53] Cao, C., Li, M., Liu, J. and Yan, Y.: An Objective Singing Evaluation Approach by Relating Acoustic Measurements to Perceptual Ratings, *Proc. of INTERSPEECH 2008*, pp. 2058–2061 (2008).
[54] Jin, Z., Jia, J., Liu, Y., Wang, Y. and Cai, L.: An Automatic Grading Method for Singing Evaluation, *Lecture Notes in Electrical Engineering*, Vol. 128, pp. 691–696 (2012).
[55] Tsai, W.-H. and Lee, H.-C.: Automatic Evaluation of Karaoke Singing Based on Pitch, Volume, and Rhythm Features, *IEEE Trans. on ASLP*, Vol. 20, No. 4, pp. 1233–1243 (2012).
[56] Daido, R., Hahm, S., Ito, M., Makino, S. and Ito, A.: A System for Evaluating Singing Enthusiasm for Karaoke, *Proc. of ISMIR 2011*, pp. 31–36 (2011).
[57] Kako, T. and Yasunori Ohishi, Hirokazu Kameoka, K. K. K. T.: Automatic Identification for Singing Style based on Sung Melodic Contour Characterized in Phase Plane, *Proc. ISMIR2009*, pp. 393–398 (2009).
[58] Tsai, W.-H. and Wang, H.-M.: Towards automatic identification of singing language in popular music recordings, *Proc. of ISMIR 2004*, pp. 568–576 (2004).
[59] Schwenninger, J., Brueckner, R., Willett, D. and Hennecke, M. E.: Language identification in vocal music, *Proc. of ISMIR 2006*, pp. 377–379 (2006).
[60] Chandraskehar, V., Sargin, M. E. and Ross, D. A.: Automatic language identification in music videos with low level audio and visual features, *Proc. of ICASSP 2011*, pp. 5724–5727 (2011).
[61] Mehrabani, M. and Hansen, J. H. L.: Language identification for singing, *Proc. ISMIR 2006*, pp. 4408–4411 (2006).
[62] Rogers, S., Girolami, M., Campbell, C. and Breitling, R.: The Latent Process Decomposition of cDNA Microarray Data Sets, *IEEE/ACM Trans. Computational Biology and Bioinformatics*, Vol. 2, No. 2, pp. 143–156 (2005).
[63] Yoshii, K. and Goto, M.: A Nonparametric Bayesian Multipitch Analyzer Based on Infinite Latent Harmonic Allocation, *IEEE Trans. on ASLP*, Vol. 20, No. 3, pp. 717–730 (2012).