Vol.2013-MUS-100 No.19
2013/9/1

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

金礪 愛[1,a)]　中野 倫靖[2,b)]　後藤 真孝[2,c)]　菊池 英明[1,d)]

概要：本稿では，ポピュラー音楽における歌声を対象として，その音響信号から印象評価語を自動推定するシステムを提案する．従来，楽曲を対象としてそのムードを音響信号から自動推定したり，歌声の印象評価語と音響特徴量の関係について分析する研究はあったが，歌声の音響信号から印象語を自動推定する研究はなかった．本稿では，人が歌声を表現するために用いる印象評価語 47 語（「心のこもった」等）と，歌声の印象評価に関わる 3 因子（独立性高く歌声を説明できる印象評価軸）である “迫力性”，“丁寧さ”，“明るさ” について，音響特徴量との関係を重回帰分析を用いてモデル化する．60 の歌声データを用いて，K 分割交差検定（$K = 6$）でモデルの予測精度を評価した結果，モデルの決定係数 R^2 について，「声量のある」「激しい」「弱い」を 0.8 以上，「勢いがある」「少女のような」を 0.7 以上，「一生懸命な」「かっこいい」等の 7 語を 0.6 以上の精度で推定できるモデルが得られた．また，上述した歌声の 3 因子については，分析の結果，13 種類の音響特徴量が関係していることが分かり，それらのうち 9 種類は各因子に対して独立に関係するものであった．

1. はじめに

本研究はポピュラー音楽における歌声を対象として, 人がその歌声を聴いた際に感じる主観的な印象評価語を，音響信号から自動推定する技術の実現を目的とする．印象評価語を歌声から自動推定できれば, 印象に基づく音楽情報検索の実現や，他人と歌声の印象を共有するシステムの実現につながる．また，自身の歌声の印象を共通の評価語によって知ることができ，歌唱力向上や表現力向上のための客観的な評価として利用できる．さらに, 歌声における主観評価と音響的な特徴との関連性を明らかにする研究は, 人間の歌声知覚の解明にも繋がる取り組みである．

歌声そのものが聴き手に伝える印象は,「歌唱者のパーソナリティ（性別や性格など）」「歌唱者の感情（嬉しそうな, 悲しそうな）」「歌声を形容する評価語（明るい，透き通った）」「個人性」「歌唱力」など多様であり，それぞれに対応する「印象評価語」が存在すると考えられる．しかし従来, 歌声らしさに対応する音響特徴量の分析 [1]，歌声と感情語の関係の分析 [2,3]，音楽の印象評価尺度を用いた歌声の分析 [4,5] が研究されてきたが，歌声の印象評価尺度を一から構築した研究はなかった．これに対して金礪（本稿の第一著者）他は，ポピュラー音楽の歌声における印象評価語を調査し，47 語を用いた印象評定実験を行った上で尺度を構築し，因子分析に基づく 3 因子を明らかにした [6]．

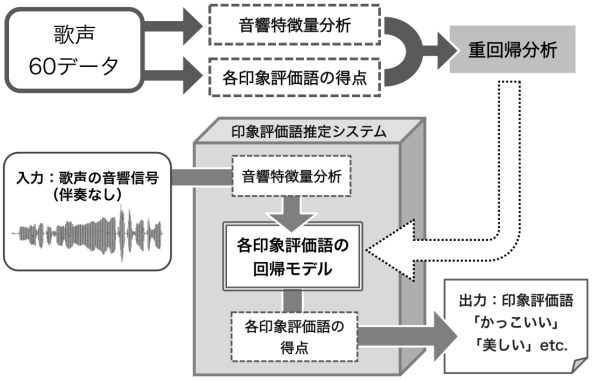


図 1　歌声の印象評価語を自動推定するシステムの概要

本稿では，このようにして得られた歌声に関する 47 語の印象評価語と 3 因子それぞれについて，歌声から音響特徴量を抽出して対応付ける．具体的には，各印象評価語毎に回帰モデルを学習し，歌声の音響特徴量から得点の高い上位 N 個の評価語を出力するシステムを構築する．従

[1] 早稲田大学大学院人間科学研究科
Graduate School of Human Science,Waseda University
[2] 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)
[a)] kanato.w [at] gmail.com
[b)] t.nakano [at] aist.go.jp
[c)] m.goto [at] aist.go.jp
[d)] kikuchi [at] waseda.jp

来，歌声の印象推定に関する研究としては，歌唱力の自動推定 [7–10] や歌詞からの感情推定 [11] があり，熱唱度という聴取印象を定義して推定する研究 [12] もある．しかし，歌声の音響信号から印象評価語を自動推定する研究はなかったため，この点が本研究の新規性であるといえる．

2. 歌唱音声の印象評価尺度の構築 [6]

金礪（本稿の第一著者）他による，歌声の印象評価語尺度の構築手順について簡単に示す．

2.1 仮尺度の構築

まず，歌声の印象を表現する語を，先行研究や音楽雑誌の CD レビュー [*1]，Twitter[*2]，ニコニコ動画 [*3] から収集した．収集した語のうち，重複語を除いた 898 語の中から固有名詞を含む表現等，尺度の評価語として適切でないと考えられる語を除外し，590 語を用いて語の了解性調査を行った．ここで了解性が高く（評価語から歌声が想像しやすく），かつ評価語収集の際に頻出度の高かった 64 語を次の同義性調査の対象とした．

同義性調査では，同義性の高かった（似ていると判断された）評価語の削除と統合を行い，**表 1** に示す 44 語の評価語を仮尺度として選定した．また，尺度には含まなかったものの，歌声の評価に重要であると考えられる「好きな」「うまい」「曲に合ってる」という評価語も，後の分析で用いた．

表 1 実験に用いた歌声の印象評価語（47 語）

仮尺度（44 語）		
甘い	心のこもった	（ドスが効いている）
安定している	こもっている	（伸びやかな）
勢いがある	（爽やかな）	激しい
（一生懸命な）	静かな	ハスキーな
色気のある	声量のある	鼻にかけたような
美しい	シャープな	響きのある
嬉しそうな	少女のような	（不安定な）
落ちつきのある	少年のような	ぶりっこみたいな
かっこいい	女性的な	（震えている）
悲しい	芯のある	真っすぐな
軽やかな	透き通った	無邪気な
可愛い	繊細な	優しい
聴きやすい	男性的な	（陽気な）
気持ち良さそうな	（中性的な）	弱い
元気な	特徴的な	
歌声評価に重要であると考えられる語（3 語）		
好きな	うまい	曲に合ってる

※括弧内は，因子分析の際に除いたが，4 章の分析に用いた評価語を示す

表 2 完成した尺度の評価語と因子負荷量

	迫力性	丁寧さ	明るさ
勢いがある	**0.932**	0.044	0.024
声量のある	**0.917**	0.188	-0.192
弱い	**-0.898**	0.023	-0.008
静かな	**-0.752**	0.466	-0.166
聴きやすい	0.146	**1.001**	0.271
透き通った	-0.127	**0.886**	0.236
落ちつきのある	-0.286	**0.775**	-0.232
響きのある	0.387	**0.756**	-0.161
嬉しそうな	0.246	0.092	**0.923**
軽やかな	-0.037	0.358	**0.854**
可愛い	-0.286	0.145	**0.830**
無邪気な	-0.085	-0.359	**0.777**
寄与率	0.292	0.292	0.262
信頼性係数 α	0.926	0.893	0.877

2.2 印象評定実験用の刺激に用いる歌声の収録

印象評定実験では，印象評価尺度の信頼性や妥当性を評価するために，多様な歌声を刺激として用いる必要がある．そのためにまず，「メロディと歌詞が統一されていること」，「評価者にとって未知のメロディと歌詞である」，ことを条件に，オリジナルのメロディと歌詞による 9 秒程度のフレーズを作成した．続いて，それらのフレーズを 21 名の女子大学生（声楽経験者，合唱経験者，バンドボーカル経験者含む）に 7 種類の条件で歌ってもらい，合計 147 歌唱を収録した．その後，評定実験での評価者の負担を下げることと，データの偏りを抑えるため，聴取印象に差がないデータを削除した．その結果，選別した 60 歌唱を印象評定実験の刺激として用いた．

2.3 歌声の印象評価因子の分析

44 語の仮尺度 (表 1) を用いて，収録した歌声に対する 7 段階の印象評定と因子分析を行い，3 つの因子を得た (**表 2**)．評価語の選定において最終的には表 2 に示す 12 語が尺度の評価語としてふさわしいと判断されたため，それぞれの因子において因子負荷量の高い評価語より，第 1 因子を “迫力性”，第 2 因子を “丁寧さ”，第 3 因子を “明るさ” と命名した．

3. 音響特徴量の分析

本稿では，各印象評価語の推定に有効な音響特徴量について検討するため，2.2 節で説明した 60 歌唱を用いる．また，調査対象とする音響特徴量は，多様な楽曲に適用することを想定して「楽譜情報を用いない」，「メロディや歌詞に依存しない」という二つの条件に当てはまるものとした．

分析に用いた歌声データは 44.1kHz, 16bit サンプリングのモノラル信号である．ここから STRAIGHT [13] を用いて 1msec ごとに F_0（基本周波数），スペクトル包絡, 非周期性指標を推定し，それらを用いてこれ以降の音響特徴量を抽出した

本章で提案する特徴量は，$\boxed{\text{n}}$のように，特徴量の番号を示す n を四角で囲んで示す．また，本稿では様々な特徴量の抽出において回帰係数の算出を行うが，全て以下の式に基づく．ここで $\boldsymbol{y}$ は分析対象とする特徴ベクトルであり，n はベクトルの長さを表している．後の分析では，$\boldsymbol{y}$ には特定時刻におけるスペクトル包絡や基本周波数の遷移などを対応付けている．

[*1] ロッキング・オン：http://ro69.jp/
[*2] https://twitter.com/
[*3] http://www.nicovideo.jp/

$$R(y) = \frac{n\sum_{k=1}^{n} k \cdot y_k - \sum_{k=1}^{n} k \sum_{k=1}^{n} y_k}{n\sum_{k=1}^{n} k^2 - (\sum_{k=1}^{n} k)} \qquad (1)$$

3.1 特徴量の表記について

分析に用いた 100 種類の特徴量を**表 3** にまとめて示した．表 3 において，1-2 のように 2 つの数字で表しているものに関しては，1 が平均，2 が標準偏差，69-72 のように 4 つの数字で表しているものは，順に平均，標準偏差，中央値，四分偏差を示している．

3.2 スペクトル包絡に関する音響特徴量

スペクトル包絡は，歌声の定常的な声質を特徴づける重要な特徴量であり，先行研究においても様々な検討がなされている ([14] など)．本調査では，各時刻 t におけるスペクトル包絡 $S(t,f)$ および対数スペクトル包絡 $LS(t,f) = \log|S(t,f)|$ における以下の特徴量の抽出を行う．ここで，f は周波数ビンの番号を示している．

スペクトル重心 各時刻におけるスペクトル包絡の重心 $C(t)$ を，以下の式を用いて求める．(表 3 : 1–4)

$$C(t) = \frac{1}{P(t)} \sum_{f=1}^{N} (S(f,t) \cdot f) \qquad (2)$$

$$P(t) = \sum_{f=1}^{N} S(f,t) \qquad (3)$$

スペクトル傾斜 式 (1) を用いてスペクトル包絡 $S(t,f)$, 対数スペクトル包絡 $LS(t,f)$ から，時刻毎の傾きを求める．この際，表 3 に示すように傾きを算出する帯域を変更しながら特徴抽出する．(表 3 : 5–20)

Singer's Formant 歌声において，歌声らしさや声の響きを評価する特徴量として Singer's Formant が知られている [14–16]. 本稿では，スペクトル包絡，対数スペクトル包絡の 2kHz〜4kHz の帯域におけるパワーの全帯域に対する割合を Singer's Formant の特徴量として扱った．(表 3 : 21–24)

微細な動的変動成分 時間軸におけるスペクトル包絡，対数スペクトル包絡の微細な変動を捉えるため，動的変動成分としてスペクトルの各周波数ビン毎に時間方向の 1 次の回帰係数を求め ((K×2+1) フレーム毎に算出)，対象フレームにおける周波数方向の総和 $\Delta S(t)$ を特徴量として使用する．ここでは，$K = 1$ とする．(表 3 : 25–32) また，スペクトル包絡，対数スペクトル包絡を各時刻において，それぞれ離散コサイン変換 (DCT) した後，全係数に対して，同様の分析を行う．ここで，F はスペクトルにおけるナイキスト周波数 (22.05kHz) に該当する周波数ビン，f は各周波数ビンの番号を示している．

表 3 本稿で用いた音響特徴量

スペクトル包絡に関わる特徴量：(log) は対数スペクトル包絡を表す	
1 - 2	スペクトル重心
3 - 4	スペクトル重心 (log)
5 - 6	スペクトル傾斜 [全帯域]
7 - 8	スペクトル傾斜 [0〜3kHz]
9 - 10	スペクトル傾斜 [0〜6kHz]
11 - 12	スペクトル傾斜 [0〜9kHz]
13 - 14	スペクトル傾斜 [全帯域] (log)
15 - 16	スペクトル傾斜 [0〜3kHz] (log)
17 - 18	スペクトル傾斜 [0〜6kHz] (log)
19 - 20	スペクトル傾斜 [0〜9kHz] (log)
21 - 22	Singer's Formant
23 - 24	Singer's Formant (log)
25 - 26	スペクトルの微細変動成分
27 - 28	スペクトル包絡 (DCT 係数) の微細変動成分
29 - 30	スペクトルの微細変動成分 (log)
31 - 32	スペクトル包絡 (DCT 係数) の微細変動成分 (log)
フォルマントに関わる特徴量	
33 - 34	スペクトル包絡の第 1 ピークの値
35 - 36	スペクトル包絡の第 2 ピークの値
37 - 38	スペクトル包絡の第 1 ピークの動的変動量
39 - 40	スペクトル包絡の第 2 ピークの動的変動量
非周期性成分に関わる特徴量	
41 - 42	非周期成分の総和
43 - 44	非周期成分の傾斜 [全帯域]
45 - 46	非周期成分の傾斜 [0〜6kHz]
動的な特徴量	
47 - 48	パワーの動的変動量 (K=25)
有声区間のみを対象	
49 - 50	スペクトル包絡 (DCT 係数) の 1 次の係数 (K=25)
51 - 52	スペクトル包絡 (DCT 係数) の各係数の総和 (K=25)
53 - 54	スペクトルの各周波数ビンの総和 [0〜3kHZ](K=25)
55 - 56	スペクトル包絡 (DCT 係数) の 1 次の係数 (K=25) (log)
57 - 58	スペクトル包絡 (DCT 係数) の各係数の総和 (K=25) (log)
59 - 60	スペクトルの各周波数ビンの総和 [0〜3kHZ](K=25) (log)
無声区間との境界も含む	
61 - 62	スペクトル包絡 (DCT 係数) の 1 次の係数 (K=25)
63 - 64	スペクトル包絡 (DCT 係数) の各係数の総和 (K=25)
65 - 66	スペクトルの各周波数ビンの総和 [0〜3kHZ](K=25)
67 - 68	スペクトル包絡 (DCT 係数) の 1 次の係数 (K=25) (log)
69 - 72	スペクトル包絡 (DCT 係数) の各係数の総和 (K=25) (log)
73 - 76	スペクトルの各周波数ビンの総和 [0〜3kHZ](K=25) (log)
基本周波数に関わる特徴量	
77	相対音高の正確さ
78	相対音高の正確さ
79	ビブラートらしさの最大値
80 - 81	ビブラートらしさの平均値・標準偏差
82	ビブラートらしさが一定以上である区間の割合
83	ビブラートのパワーの最大値
84 - 85	ビブラートのパワーの平均値・標準偏差
86	有声区間中のビブラートの割合
87	F_0 の揺れの安定度合の最大値
88 - 89	F_0 の揺れの安定度合の平均値・標準偏差
90	F_0 の揺れの安定度合いが一定以上である区間の割合
91	F_0 の揺れのパワーの最大値
92 - 93	F_0 の揺れのパワーの平均値・標準偏差
94 - 95	F_0 の動的変動量 (K=10)
96 - 97	F_0 の動的変動量 (K=25)
98 - 99	F_0 の動的変動量 (K=50)
100	F_0 の動的変動量の変動の少ない部分の割合

$$\Delta S(t) = \sum_{f=1}^{F} \frac{\sum_{k=-K}^{K} k \cdot S(f, t+k)}{\sum_{k=-K}^{K} k^2} \qquad (4)$$

3.3 音韻性の知覚に関する音響特徴量

スペクトル包絡にはフォルマントに関する情報も含まれており，音韻の知覚や歌声の印象にも影響を及ぼすと考えられるため，関係する特徴量を抽出する．

フォルマントに関わる特徴量 フォルマントに関係する特徴量として，スペクトル包絡のピーク周波数を求める．まず，各時刻 (t) のスペクトル包絡のケプストラムの低次成分に対して逆フーリエ変換を行い，文献 [17] を参考に，フォルマント周波数である可能性が高いと考えられる帯域（$F_1 < 900$Hz,900Hz $< F_2 <$ 3300Hz）に制限した上でピークの検出を行い，低い方から順に第 1 ピーク $F_1(t)$，第 2 ピーク $F_2(t)$ を求めた．（表 3：33–36）

フォルマントに関わる動的特徴量 $F_1(t)$,$F_2(t)$ について，式 1 によって 50 フレームの窓幅を 1 フレームずつシフトさせながら回帰係数を求めて，動的変動量とした．（表 3：37–40）

3.4 非周期性成分

STRAIGHT [13] では，スペクトル包絡の全体のエネルギーに対する非周期成分の割合を，0 から 1.0 の値で求めることができる．値が 1 に近づく程，非周期成分の割合が多いことを示しており，音声に含まれている非周期成分の大きさを評価することができる．

非周期性成分 本稿では，この非周期性のスペクトル包絡全帯域における値の総和と，全帯域および 6kHz までの傾きを式 (1) の y に非周期性指標 $Ap(f)$ を代入して算出し，特徴量として扱った．（表 3：41–46）

3.5 動的な音響特徴量

3.3 までで扱った特徴量は歌声の「声質」に関わっているであろう定常的な特徴量である．歌声の印象の知覚には，スペクトル包絡の動的な変動も関与していると考えられるため，以下の特徴量を扱うこととする．

パワーの動的変動量 式 (3) を用い，各時刻におけるパワーを求めた上で，式 (1) を用いて回帰係数を求める．（表 3：47–48）

スペクトル包絡の形状の動的変動量 スペクトル形状の動的変動量を扱うため，式 (1) を用いてスペクトル包絡と対数スペクトル包絡の回帰係数を求める．また，有声区間と無声区間の境界部分の動的変動量は音の立ち上がりの早さに関する指標として扱えると考え，有声区間のみではなく，有声区間と無声区間の境界からも特徴抽出する．（表 3：49–76）

3.6 基本周波数に関する音響特徴量

本稿で扱う周波数は対数スケールで示し，cent 単位で表す．西洋平均律では，半音が 100cent にあたる．中央ハ音の周波数 $f_c(= 440 \times 2^{\frac{3}{12}-1} = 261.62...Hz)$ の cent 値を $4800cent$ とすると，周波数 f_{Hz} の音の cent 値 f_{cent} は

$$f_{cent} = 1200 \log_2(\frac{f_{Hz}}{f_c}) + 4800 \qquad (5)$$

で表される．今後，本稿では基本周波数を $F_0(t)$ で表すこととする．ここで．t は時間軸を示している．

相対音高 本稿では，楽譜情報を用いないため，歌声の相対音高に関する二種類の特徴量 [7] を抽出する．具体的には，文献 [7] における相対音高の正確さ（$g(F)$）のピークの鋭さを77，そのピークの傾斜を直線近似した傾きを78として扱う．

ビブラート ビブラートは歌唱力の判断に影響する重要な特徴量であるため，文献 [18] の方法で時刻 t におけるビブラートの速さの周波数帯域のパワーとビブラートらしさを特徴量として扱う．また，$F_0(t)$ の変動幅が 30 cent 〜150 cent であり，区間（320msec）の平均音高と 5 回以上交差する区間をビブラートとして検出して，有声区間におけるビブラート区間の割合も特徴量として扱う．ここで本稿では，$F_0(t)$ から次式のようにビブラート等の変動のみを抽出して $f_d(t)$ とした後，そこからもビブラート特徴量を抽出する．

$$f_d(t) = F_0(t) - f_l(t) \qquad (6)$$

ここで，$f_l(t)$ は，$F_0(t)$ にカットオフ周波数 5Hz のローパスフィルタをかけて変動を除去したものである．（$f_d(t)$ を用いたもの 表 3：79–86）（$F_0(t)$ を用いたもの 表 3：87–93）

動的変動成分 歌声の $F_0(t)$ を扱う上での重要な要素として，プレパレーションやオーバーシュートなど，異なる音高へ遷移する際の特徴がある．本稿では，式 7 を用いて 1 フレーム（1msec）ごとに $F_0(t)$ の動的変動量 $D(t)$ を求め，$F_0(t)$ の遷移に関する特徴量として扱う．（表 3：94–99）

$$D(t) = \frac{\sum_{k=-K}^{K} k \cdot F_0(t+k)}{\sum_{k=-K}^{K} k^2} \qquad (7)$$

また，K=25 で求めた $D(t)$ において，有声区間中で変動が極めて小さい部分の割合を求め，どの程度 $F_0(t)$ がぶれずに歌えているかを評価する特徴量とする．（表 3：100）

4. 重回帰分析及び交差検定

本稿では，2.2 節で収録を行った歌声 60 データから、これまで説明した 100 種類の音響特徴量を抽出し、印象評価語と対応づける。ここで、印象評価語の印象得点については，表 1 に示した 44 語と「好きな」「うまい」「曲に合ってる」という評価語の得点を用いる．また，“迫力性”,“丁寧さ”,“明るさ” の 3 因子に関わる評価語 (表 2) の得点をそれぞれ合計したものも含めて，計 50 種類の印象評価得点を用いる．それぞれの特徴量（表 3）を標準化した値を説明変数，各評価語の印象得点を標準化したものを目的変数とした重回帰モデルを構築し，歌声 60 データに対し K-分割交差検定を行った（K = 6）.

ここで、扱う特徴量の数が不必要に多くなってしまうと，説明変数間の相関の高さによりモデルの安定性が下がってしまう原因となる多重共線性 ([19]) が生じやすいため，まず各特徴量間の相関係数を求め，極端に相関の高かった特徴量のうち，20 種類の特徴量の除外を行った．除外した特徴量は 12 , 14 , 20 , 22 , 26 , 28 , 30 , 42 , 50 , 52 , 54 , 56 , 58 , 63 , 64 , 68 , 70 , 80 , 88 , 89 である．

次に，残りの 80 種類の特徴量を用い，ステップワイズ変数選択により各印象ごとの重回帰モデルを構築した．各モデルにおいて，採用された各説明変数の多重共線性の危険度を示す VIF（分散拡大要因）を求め，VIF = 10 以上である特徴量の除外を行った．その上で，再度重回帰モデルの構築を行い，VIF の高い変数を除外する作業を繰り返した.

モデルの評価については，60 データ全てを用いて作成した回帰モデルの自由度調整済み決定係数（クローズドテスト），K 分割交差検定（K=6）によって得られたテストデータの適合度を式 (8) にて求めた（オープンテスト）．ここでは，テストデータの印象得点の実測値を y，モデルによる予測値を $\hat{y}$，実測値の平均値を m で示している．この値が 1 に近い程，モデルの予測精度が高いことを意味する.

$$R^2 = 1 - \frac{\sum_{n=1}^{N} (y_n - \hat{y}_n)^2}{\sum_{n=1}^{N} (y_n - m)^2} \qquad (8)$$

このようにして多重共線性の危険性を排除した後，説明変数と対象の印象得点の単相関の符号と，回帰モデルにおける偏回帰係数の符号が異なる抑制変数の有無を確認した．抑制変数が生じていた場合には対象の変数の除外を試み，除外の前後のオープンテストの値を確認し，値が高かったものを最終的なモデルとして採用した.

5. 結果と考察

各モデルの自由度調整済み決定係数（クローズドテスト）と，交差検定の結果（オープンテスト）を**表 4** に示す.

各モデルは全て $p < .001$ で有意であった．「激しい」「声量のある」「一生懸命な」「弱い」といった “迫力性” に関わる評価語については，決定係数が 0.8 を超えており，特徴量からの推定精度が高いといえる．迫力性以外の印象では，「可愛い」「響きのある」「優しい」という評価語の推定精度が比較的高く，決定係数が 0.7 程度であった.

「鼻にかけたような」「こもっている」「真っすぐな」といった評価語は，人による印象評定の段階で評価者間の相関が低いものであったため，回帰モデルの精度が落ちてしまったのではないかと考えられる．また，「うまい」という評価語については，印象評価における評価者間の相関は「声量のある」「弱い」よりも高かったが，今後は精度向上のための特徴量を検討する余地がある.

5.1 特徴量ごとの傾向

歌声の印象推定における各音響特徴量の貢献を，以下で考察する.

5.1.1 スペクトル傾斜

本稿ではスペクトル包絡の傾斜を帯域毎に分けて算出したが，それぞれの帯域で異なる印象との相関が見られた．特に対数スペクトル包絡において顕著に結果が表れていたため，それぞれの特徴量が回帰モデルに採用されていた評価語の例を以下に示す (**表 5**).

15 は，迫力性に関わる印象語において有効な特徴量であると言え，これは話声の声質の一つである「気息性」と関わっていると考えられる．文献 [20] では，音声のスペクトルにおける $H1$（基音のパワー）$-H2$（第二倍音のパワー）の値が大きい程気息性が高くなるとされており，この値は気息性の評価指標として広く用いられている特徴量である．本稿で用いた 0〜3kHz までのスペクトル傾斜は，$H1 - H2$ の値の影響を大きく受けると考えられるため，歌声の迫力性の評価には歌声の気息性の強度が関わっていると言える.

図 2 に，歌声ごとのスペクトル傾斜の例と，印象得点の実測値の傾向を表記した．赤線は 15 を示しており，この傾斜が緩やかである程，迫力性の得点が高いと予測される．図 2 では No.19 が他の 2 つの傾斜に比べて緩やかになっていることが分かる．また，緩やかであるほど明るさの得点が高くなる紫線 17 は No.17 で，急であるほど丁寧さの得点が高くなる青線 19 は No.9 において，それぞれ他の 2 つとは異なる傾向が観察できる．このように，帯域毎に傾斜を算出することにより，全帯域の傾斜では捉えられない特徴を表すことができた．ただし，スペクトル傾斜の各帯域ごとの特徴量は互いに独立ではなく，15 は 17 , 19 に，17 は 19 に大きく影響を及ぼす．回帰モデルの推定精度を見ても，他の影響を受けない 15 に関わる評価語が最もモデルの精度が高い結果となっている．17 や 19 に関わる評価語のモデル精度をさらに高めるためには，これ

表 4 各評価語における重回帰分析結果

44 語の印象評価語と R^2（1 に近い程モデルの精度が高い）					
印象評価語	クローズド	オープン	印象評価語	クローズド	オープン
声量のある	0.878	0.864	気持ち良さそうな	0.626	0.456
激しい	0.858	0.833	元気な	0.601	0.453
弱い	0.879	0.803	透き通った	0.549	0.410
勢いがある	0.757	0.731	美しい	0.521	0.404
少女のような	0.788	0.724	震えている	0.544	0.390
一生懸命な	0.812	0.691	男性的な	0.579	0.383
かっこいい	0.728	0.679	ハスキーな	0.593	0.374
静かな	0.755	0.673	軽やかな	0.496	0.363
繊細な	0.698	0.671	中性的な	0.510	0.334
響きのある	0.706	0.668	無邪気な	0.640	0.318
優しい	0.692	0.645	ぶりっこみたいな	0.517	0.314
可愛い	0.739	0.633	落ちつきのある	0.442	0.270
芯のある	0.732	0.585	爽やかな	0.388	0.266
ドスが効いている	0.618	0.577	不安定な	0.360	0.230
甘い	0.680	0.539	安定している	0.380	0.206
少年のような	0.578	0.491	特徴的な	0.414	0.205
心のこもった	0.596	0.485	聴きやすい	0.306	0.153
伸びやかな	0.576	0.475	陽気な	0.517	0.028
悲しい	0.570	0.475	こもっている	0.292	-0.025
女性的な	0.515	0.474	真っすぐな	0.363	-0.184
シャープな	0.567	0.464	嬉しそうな	0.359	-0.332
色気のある	0.606	0.462	鼻にかけたような	0.170	-1.488

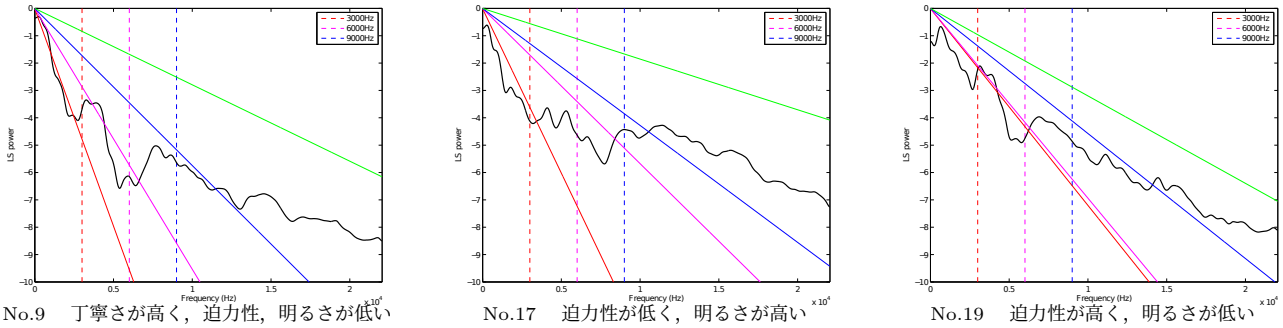
歌声の印象評価における 3 因子		
印象評価語	クローズド	オープン
迫力性	0.730	0.609
丁寧さ	0.356	0.235
明るさ	0.459	0.315

歌声の評価に重要である評価語		
印象評価語	クローズド	オープン
好きな	0.401	0.299
うまい	0.327	0.252
曲に合ってる	0.346	0.089

仮尺度で用いた 44 語の平均		
	クローズド	オープン
44 語の平均	0.579	0.390

歌声評価尺度に含まれる 12 語の平均		
	クローズド	オープン
12 語の平均	0.626	0.463

図 2 スペクトル傾斜の例



No.9 丁寧さが高く，迫力性，明るさが低い　No.17 迫力性が低く，明るさが高い　No.19 迫力性が高く，明るさが低い

このグラフは，各歌声データの長時間スペクトル平均の傾斜を示したものであり，それぞれ，赤線=0〜3kHz（15），紫線=0〜6kHz（17），青線=0〜9kHz（19），緑線=全帯域（13），の傾斜を示している．ここでは，傾斜の違いが分かり易いよう，切片は全て 0 に揃えてある．

表 5 スペクトル傾斜に関わる評価語

0〜3kHz 15		0〜6kHz 17		0〜9kHz 19	
+	激しい	+	少女のような	−	心のこもった
+	元気な	+	可愛い	−	シャープな
+	勢いがある	+	陽気な	−	かっこいい
−	優しい	+	元気な	−	気持ち良さそうな
−	落ち着きのある	−	伸びやかな		
−	静かな	−	声量のある		
		−	響きのある		

(+ は正の偏回帰係数，− は負の偏回帰係数を示す)

らの特徴量を独立で扱えるような算出方法を検討する必要があると考えられる．

5.1.2 F_0 相対音高

F_0 の相対音高の正確さ77は，歌唱力評価に有効であると知られている [7]．本調査でも，主に丁寧さや美的要素に関わる評価語の推定に貢献していることが判明し，改めてこの特徴量の有効性が確認された．

これらのような，丁寧さや美的要素に関わる評価語のモデル精度は全体的に低いため，77は歌声の印象推定において非常に重要な特徴量であると言える．

5.1.3 ビブラート

ビブラートもまた，歌唱力評価に有効である特徴量の一つであり，本稿では特徴量ごとに以下の傾向が見られた．

79は，ある区間の F_0 の変動における，5〜8 Hz の帯域のパワーの割合を表しているため，この値が大きいほどビブラートだと認識される揺れのみが存在していると言える．この特徴量は先行研究でも知られている「響きのある」という評価語や「美しい」「安定している」といった評価語の推定に有効であるという結果がでていた．また，「うまい」という評価語では，回帰モデルに採用されたのがこの特徴量のみであったため，ビブラートだと認定された区間の長さよりも，綺麗なビブラートが表出されたかどうかが「うまい」という評価には重要であると考えられる．また，ビブラートのパワーの最大値83は，「心のこもった」という評価語に関係しており，5〜8Hz の F_0 の変化幅が大きい

表 6 相対音高の正確さに関わる評価語

F_0 の相対音高の正確さ 77		
聴きやすい	優しい	繊細な
透き通った	安定している	伸びやかな
美しい	心のこもった	女性的な

表 7 印象推定に有効であったビブラートの特徴量

ビブラートらしさの最大値 79		
+	聴きやすい	透き通った
+	安定している	響きのある
+	美しい	静かな
+	（好きな）	（うまい）
ビブラートのパワーの最大値 83		
+	心のこもった	
ビブラートのパワーの平均 84		
+	気持ち良さそうな	
ビブラート区間の割合 86		
+	女性的な	繊細な
−	少女のような	

(+ は正の偏回帰係数，− は負の偏回帰係数を示す)

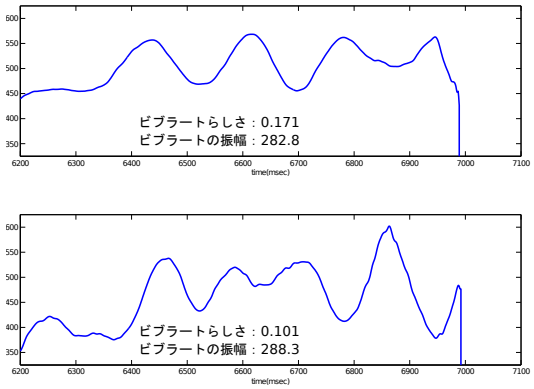


図 3 ビブラートの特徴量と例

ほど，「心のこもった」という評価がされているということになる．

ビブラートの例として，2 種類の歌声における同じ区間の $F_0(t)$ を**図 3** に示す．これらは 2 種類とも，本稿で定義した変動幅（30 cent 〜 150 cent）や平均音高との交差回数（320msec 区間で 5 回）を満たしており，ビブラートと判断された F_0 区間であるが，ビブラートの形状により特徴量が異なっている．図 3 の上段は，ビブラートの波が規則的に表れており，79 の値が高いデータである．逆に下段の図は，79 の値が低く，ビブラートの規則性が失われている例である．このように，同じビブラートでも，形状の違いにより与える印象が異なる可能性がある．

5.1.4 その他の特徴量

この他に，各印象に大きく影響を与えていた特徴量を以下に示す．

非周期性の総和 41 は**表 8** に示した通り，迫力性に関わっていると言える．41 が高いということは音声にノイズ成分が多く含まれていることを示しており，話声において知られている音声の気息性の特性 [20] と通ずる結果が得られた．F_0 の動的変動量 96 については，50msec 区間で

表 8 印象推定に特に有効であった特徴量

非周期性の総和 41		
+	少女のような	優しい
−	声量のある	勢いがある
F_0 の動的変動量 (50msec) 96		
+	勢いがある	ドスが効いている
−	少女のような	可愛い
スペクトル包絡（DCT 係数）の 1 次の係数の動的変動量 49		
+	かっこいい	シャープな
パワーの動的変動量 47		
+	嬉しそうな	爽やかな
スペクトル包絡の微細変動量 25		
+	弱い	静かな
−	一生懸命な	

(+ は正の偏回帰係数，− は負の偏回帰係数を示す)

表 9 3 因子に関わる特徴量（平均は M，標準偏差は SD で示す）

因子		特徴量
迫力性	+	F_0 の動的変動量：M 96
	−	スペクトル包絡の微細変動：M 25
	−	非周期性の総和：M 41
	−	F0 の揺れが安定している区間の割合 90
	−	非周期性の 6kHz までの傾斜：SD 45
丁寧さ	+	ビブラートらしさの最大値 79
	−	6kHz までの対数スペクトルの傾斜：M 17
	+	パワーの動的変動量：M 47
	+	相対音高の正確さ 77
明るさ	+	6kHz までの対数スペクトルの傾斜：M 17
	+	3kHz までの対数スペクトル包絡の動的変動量 SD 60
	+	非周期性の 6kHz までの傾斜：M 45
	+	対数スペクトル包絡（DCT 係数）の微細変動成分：SD 32

(+ は正の偏回帰係数，− は負の偏回帰係数を示す)

の F_0 の変化速度が速いほど，「勢いがある」といった評価がされるということであり，この特徴量も迫力性に関わる指標であると言える．しかし，パワーの変動量 47 に関しては，同じ時間長を対象としているにも関わらず，「嬉しそうな」「爽やかな」という評価語との関連が見られた．最後に，スペクトル包絡の微細変動量 25 については，27 や 29 でも同様の傾向が見られ，スペクトル包絡の細かい変動（1msec ごと）が大きい程，迫力性が低いと評価されている．

5.2 歌声評価の 3 因子と特徴量

歌声の印象評価における “迫力性”，“丁寧さ”，“明るさ”，の 3 つの因子 [6] について，それぞれの因子の回帰モデルに採用されていた主な特徴量を**表 9** に示す．

6kHz のスペクトル傾斜 17，非周期性の総和 41，悲周期性の 6kHz までの傾斜 45 以外の特徴量については，それぞれの特徴量はある程度独立であると考えられる．文献 [6] の調査により，各因子の印象得点の相関は 0.2 程度と低かったため，関係する特徴量がある程度独立であることにも納得がいく．また，それぞれの因子の推定に，スペクトル包絡の形状に関わる静的な特徴量と，スペクトル包絡の変動や F_0 の変動に関わる動的な特徴量の両方が採用されていることも，重要な点である．迫力性以外の因子に

関しては推定精度が低かったため，今後は各因子で採用されたいた特徴量を手がかりに，モデルの精度を更に高めるための特徴量を検討する必要がある．

5.3 歌声からの印象評価語の自動推定

本稿では，各印象ごとに一つの回帰モデルを構築しており，１つの歌声データに対し各評価語ごとの得点を算出することができる．ここで，得点の高い印象評価語が，その歌声から感じられる印象であるため，本稿では，印象評価得点の上位 5 位までに該当する印象評価語を出力する印象評価語の自動推定システムを構築した．

6. まとめ

本稿では，歌声からその印象評価語を推定することを目的とし，歌声の印象と音響特徴量の関係について重回帰分析を用いて調査を行った．結果,「一生懸命な」「弱い」「声量のある」など，“迫力性” に関わる評価語については，高い精度のモデルが構築できたが,「安定している」「聴きやすい」といった “丁寧さ” に関わる評価語のモデルの精度は十分でなかった．また，歌声の印象評価に関わる３因子についても，それぞれの推定に有効な静的な特徴量，動的な特徴量を確認することができたため，この結果は歌声合成にも応用可能性がある．本稿で検討した音響特徴量は，メロディや歌詞に依存しないことを前提に考案したものであるが，フォルマントや F_0 の変動量の特徴量については歌唱データによって異なる結果が出てしまうことも考えられる．今回は全て同じメロディ，キー，テンポ，歌詞で歌唱したデータを用いたが，システムをより頑健にする為には様々な条件の歌唱データを用いる必要がある．今後は，構築したモデルを用い，印象評価語を推定するシステムの開発を目指すこととする．

参考文献

[1] 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，日本音響学会誌，Vol. 64, No. 5, pp. 267–277 (2008).
[2] 森下修次，笹川和美：歌唱における喜び，悲しみ，恐れ，怒りの表現，新潟大学教育人間科学部紀要，Vol. 5, No. 1, pp. 193–200 (2008).
[3] Kotlyar, G. M. and Morozov, V. P.: Acoustical correlates of the emotional content of vocalized speech, *Sov.Phys.Acoust.*, Vol. 22, No. 3, pp. 208–211 (1976).
[4] 星野悦子：歌の聴取印象と再認記憶： 言葉とメロディの関係を探る，情報処理学会研究報告 音楽情報科学，Vol. 2002-MUS-45, No. 19, pp. 109–114 (2002).
[5] 星野悦子：歌曲の聴取における連続的反応測定 (音楽認知・知覚 1)，情報処理学会研究報告音楽情報科学，Vol. 2004-MUS-57, No. 10, pp. 53–58 (2004).
[6] 金礪　愛，菊池英明：ポピュラー音楽のための歌唱音声評価尺度の構築，日本音響学会研究発表会講演論文集，pp. 397–400 (2013).
[7] 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).
[8] 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).
[9] 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).
[10] 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).
[11] Hu, Y., Chen, X. and Yang, D.: Lyric-based Song Emotion Detection with Affective Lexicon, *Proc. ISMIR2009* (2009).
[12] 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).
[13] 河原英紀：聴覚の情景分析が生み出した高品質 VOCODER:STRAIGHT, 日本音響学会誌, Vol. 54, No. 7, pp. 521–526 (1998).
[14] Sundberg, J.: *The Science of the Singing Voice*, Northern Illinois University Press (1987).
[15] 池田　操，伊東一典：音楽科学生と一般学生の歌声の音響分析と評価 : シンガーズ・フォルマントを指標として，上越教育大学研究紀要，Vol. 19, No. 2, pp. 493–509 (2000).
[16] エリクソンドナ，齊藤　毅，細川久美子，岸本宏子，羽石英里：女声の「歌唱フォルマント」の音響学的研究: その 1，研究紀要， Vol. 29, pp. 13–26 (2010).
[17] 田窪行則，前川喜久雄，窪薗晴夫，本多清志，白井克彦，中川聖一：岩波講座言語の科学，No. 2, 岩波書店 (1998).
[18] Tomoyasu, N., Masataka, G. and Yuzuru, H.: Subjective evaluation of common singing skills using the rank ordering method, *Proc. of ICMPC2006*, pp. 1507–1512 (2006).
[19] N.R.Draper, H.: 応用回帰分析，森北出版 (1968).
[20] Klatt, D. H. and Klatt, L. C.: Analysis, synthesis, and perception of voice quality variations among female and male talkers, *Journal of the Acoustical Society of America*, Vol. 87, No. 2, pp. 820–857 (1990).