Vol.2018-MUS-119 No.13
Vol.2018-SLP-122 No.13
2018/6/16

歌声の習熟度と周波数特性の時系列変化の関係に関する考察

吉田 祥[†1] 香山 瑞恵[†2] 池田 京子[†3] 山下 泰樹[†4] 伊東 一典[†2] 浅沼 和志[†5]

概要：本研究では，歌声の響きに関連する周波数特性の強度や割合の定量化を検討し，歌声評価指標として提案することを目的としている.これまでの研究で，心理的印象に影響を与える音響特徴量が歌唱指導により変化することが確認されている．また，この音響特徴量を評価するために，声楽初学者を対象とした特定音に対する歌声評価指標を検討してきた．本稿では，これらの成果をふまえ，これまで行ってきた音響特徴量の瞬間的定量化から，時間変化をとらえる可視化への拡張を試みる．その結果から，習熟度の異なる被験者群の比較および各音響特徴量の時系列相関比較から得られた成果について述べる．

キーワード：歌声，周波数特性，歌唱評価指標，音響特徴量，指導前後，定量化，Singer’s formant

A Consideration of Relationship between Time Series Changing of Singing Voice and Singing Skill based on Acoustic Frequency Features

Sho YOSHIDA[†1] Mizue KAYAMA[†2] Kyoko IKEDA[†3]
Yasuki YAMASHITA[†4] Kazunori ITOH[†2] Kazushi ASANUMA[†5]

Abstract: In this research, we have proposed quantitative metrics of acoustic features of frequency related to singing voices. Based on these metrics, we try to assess singing skill. In our previous studies, we pointed out that acoustic features affect psychological impression for singing. Moreover, we confirmed that these metrics which were contained in learner's singing voice were changed according to singing instruction. In this paper, we show the extended our metrics from instantaneous quantification to time series changing. We compare three groups which are different singing skill level based on our new index.

Keywords: Singing voice, Acoustic frequency features, Metrics for singing voice, Quantification, Singer’s formant

1. はじめに

歌声研究分野において，歌唱力と音響特徴量の関係について長年に渡って検討がなされている．物理的に良い声とされている条件に関する研究や不快な歌声の音響特徴量に関する調査など，歌唱の習熟度と音響特徴量に関係した研究が数多くある[1,2]．しかし，これらの研究はプロ歌手を対象としたものが多く，声楽初学者の歌唱指導と各音響特徴量の関係についてまだ議論が十分にされていない．

また，歌声は特異なスペクトル形状を持つことが知られている．その 1 つとして，歌声の 3kHz 付近のスペクトルピークがあり，Singer’s formant と称される[1,3]．これは母音の違いに依存せず，特に男声の「響く声」「艶のある声」を特徴付けるといわれる．また，典型的な良い歌声の条件としてビブラートやSinger’s formantの存在が指摘されており，これらを有することにより「歌声らしい」という聴感的印象となる[4]．

本研究では，歌声の響きに関連する周波数特性の強度や割合の定量化を検討し，歌声評価指標として提案することを目的としている．これまでの研究で，心理的印象に影響を与える音響特徴量が歌唱指導により変化することが確認されている[5]．また，この音響特徴量を評価するための声楽初学者に特化した歌声評価指標としてSinger’s formantの強度と割合の定量化を行ってきた．そして，これらが初学者とプロ歌手の歌声分析から，いずれかもしくはその両方の上昇が声楽の上達に必要であることが分かっている[6]．

本稿では，これらの成果をふまえ，これまでの瞬間的な定量化から時間変化による定量化へと拡張を試みる．そして，習熟度の異なる被験者群同士の比較や時系列上での相関分析の結果を示す．

2. 本研究で用いる評価指標

本章では，本研究で提案している歌声評価指標について述べる．ここでは歌声音声に含まれる周波数成分から習熟

†1 信州大学大学院
Graduate School of Shinshu University
†2 信州大学工学部
Faculty of Engineering, Shinshu University
†3 信州大学教育学部
Faculty of Education, Shinshu University
†4 長野県工科短期大学校
Nagano Prefectural Institute of Technology
†5 国立高専機構長野高専
NIT Nagano College

ⓒ 2018 Information Processing Society of Japan

度に関係すると思われる響きの有無や程度について検討するため，高次（2.4~4.0 kHz）の倍音に着目する．本研究では，この帯域の周波数成分の割合と強さを定量化した結果を歌声評価指標としている．

2.1 周波数成分の割合

歌声データに対する高速フーリエ変換（Fast Fourier Transform ： FFT）の結果に基づき，高次倍音成分の割合を算出する．歌声データの FFT により基本周波数と倍音がピークとして現れる．FFT の分析条件を以下に示す．

窓関数 : ハミング窓，フレーム周期 : 10 ms，窓長 : 30 ms

FFT 後のデータにおける振幅スペクトル特性の大きさを d とすると，RMS 値は式(1)により求められる．

$$\mathrm{RMS} = 10^{\left(\frac{d+80}{20}\right)} \quad (1)$$

RMS から歌声の Singer’s formant が含まれるとされる 4.0 kHz までの帯域成分のうち，2.4～4.0 kHz の帯域の合計値の割合を SFR とする．以下の式(2)で算出する．この値が大きいほど Singer’s formant 領域に倍音成分が集中していることを表す．

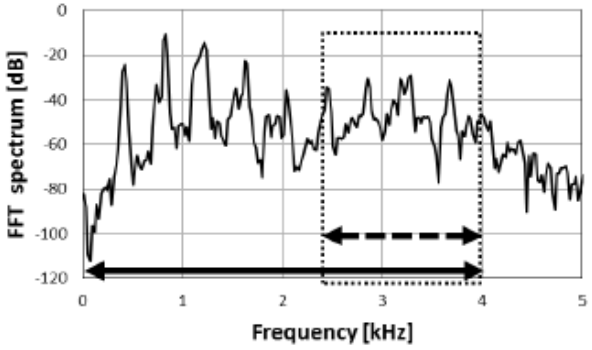


図 1　SFR の定義

$$\mathrm{SFR} = \frac{\text{2.4~4.0 kHz の RMS 値の合計}}{\text{4.0kHz までの RMS 値の合計}} \times 100 \quad (2)$$

2.2 周波数成分の強さ

2.2.1 LPC 包絡線の算出方法

歌声に対する LPC 包絡線から高次倍音成分の強さを算出する．LPC 包絡線は以下の手順で求めるものとする．

Ⅰ．　音声データを読み込む．

Ⅱ．　音声波形の中心部分を切り出す．波形 x とする．

Ⅲ．　プリエンファシスフィルタをかける．

Ⅳ．　ハミング窓をかける．

Ⅴ．　LPC 次数 12 として，FFT のサンプル数 N，自己相関関数のサイズを τ として，波形 x の自己相関 r を以下の式(3)で求める．

$$\mathrm{r}(\tau) = \frac{1}{N-\tau}\sum_{k=0}^{N-1-\tau} x_k x_{k+\tau} \quad (\tau=0,1,\cdots\tau_{\max}) \quad (3)$$

Ⅵ．　自己相関 r を用いて，LevinsonDurbin アルゴリズムから LPC 係数を算出する．

手順Ⅲではプリエンファシスフィルタを用いる．このフィルタは，音声のスペクトルの平坦化しダイナミックレンジを圧縮することで実効的な SNR を高める．また，手順Ⅴの自己相関関数は先行研究[6]での分析時の計算式とは異なり，サンプル数から自己相関関数のサイズを減じた値で除算することとした．

2.2.2 Q 値

2.2.1 項に示した手順で求められた LPC 包絡線を用いて，倍音成分の強さとしての Singer’s formant の鋭さを算出する．

LPC 包絡線における Singer’s formant の鋭さを Q 値とする．2.4～4.0 kHz の帯域において，最大ピークの周波数を f_a，ピークの両側に-3 dB となる周波数を f_b，f_c （$f_b > f_c$）として，Q 値を以下の式(4)で算出する．Q 値が大きいほど周波数成分が狭い範囲に集中していることを表す．

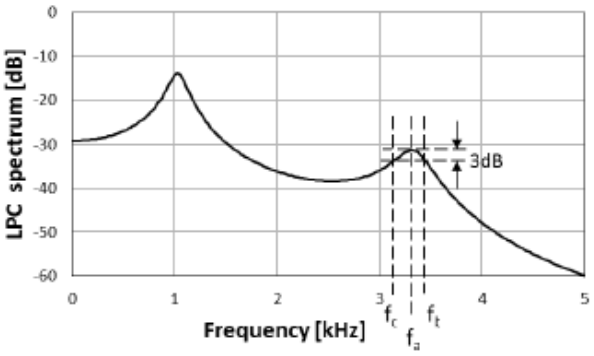


図 2　Q 値の定義

$$\mathrm{Q} = \frac{f_a}{f_b - f_c} \quad (4)$$

2.2.3 2 凸

声楽初学者の場合，Singer’s formant 帯域の成分が十分でなく，Q 値が算出しづらいことがわかっている[7]．そこで，3 kHz 付近のピークの高さを Q 値に代わる指標とすることとした．図 1 にこの指標の求め方を示す．縦軸は周波数成分の強度の相対レベル，横軸は周波数を表す．Singer’s formant が含まれるとされる帯域のピークとそのピーク直前にある谷との差を 2 凸とする．

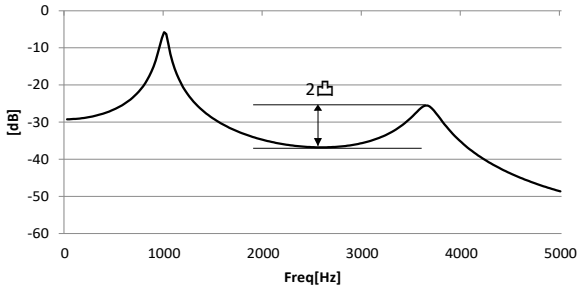


図 3　2 凸の定義

Vol.2018-MUS-119 No.13
Vol.2018-SLP-122 No.13
2018/6/16

3. プロ歌手の声種別による時間的変化

本章では，前章で示した評価指標を用いてプロ歌手を対象とした歌声の時系列分析について述べる．

3.1 分析環境

被験者を表 1 に示す, 被験者は, 女性のプロ歌手 6 名（被験者 A，B，C，D，E，F）である．いずれも演奏キャリア，指導歴が十分にある．また，プロ歌手の声種は被験者 A，B，C がソプラノ（Sop.），D，E，F がメゾ・ソプラノ（Mez.）である．

分析で使用する楽曲は, 声楽指導者の意見から「Caro mio ben」（作詞：不明，作曲：Tommaso Giordani）とした．被験者は，曲全体を独唱する．各被験者につき 1 回の収音で 2 回ずつの歌唱をさせた．収音は音の反響しない静かな部屋で行い, レコーダは ICR-PS605 RM (SANYO)を使用した．被験者にはレコーダから 2 m 離れた正面立位で歌唱させた．音声はサンプリング周波数 44.1 kHz，量子化ビット数 16 bit で収音した．分析対象とするデータは，声楽指導者に録音した 2 つの歌声データを聴かせ，被験者の習熟の程度がよく表れていると判断された 1 つの歌声を選び，分析データとした．特に,「Caro mio ben」内の歌詞「cessa」と「tanto」のワンフレーズを評価対象区間とし，SFR と Q 値を導出した．これらの歌声の収音時期は，2015 年 11 月である．

ここでは分析結果の時系列上での指標成分の傾向を明確にするため，カルマンフィルタを用いたカルマン平滑化を適用する．また，それぞれの評価対象区間には歌唱時間に差があり，データ数が異なる．分析の結果の公平性を保つため，全てのデータ数を統一する処理を行った．また，結果を可視化する際は声種ごとに平均値を取り比較を行う．

3.2 分析結果

分析の結果として, 図 4 は cessa 区間の SFR, 図 5 は cessa 区間の Q 値，図 6 は tanto 区間の SFR，図 7 は tanto 区間の Q 値を表す．いずれのグラフも横軸が時間軸であり，縦軸は SFR または Q 値である．グラフ内の実線は Mez.，破線は Sop.を表す．

表 1　プロ歌手の情報

被験者	所属・ 演奏キャリア	指導歴	声種
A	関西二期会・21 年	12 年	Sop.
B	東京二期会・34 年	22 年	Sop.
C	東京二期会・33 年	29 年	Sop.
D	東京二期会・38 年	22 年	Mez.
E	藤原歌劇団・18 年	8 年	Mez.
F	東京二期会・27 年	22 年	Mez.

これらから，声種間の差が最大となるのは cessa 区間の SFR が 17.08（t=0.21），cessa 区間の Q 値が 16.13（t=1.70），tanto 区間の SFR が 14.12（t=1.28），tanto 区間の Q 値が 11.88（t=1.32）となった．また，各波形の全区間を通じての最大値，最小値の差は，cessa 区間 SFR の Mez.が 16.13，Sop.が 8.09，cessa 区間 Q 値の Mez.が 28.28，Sop.が 23.16，tanto 区間 SFR の Mez.が 12.64，Sop.が 8.76，tanto 区間 Q 値の Mez.が 19.49，Sop.が 22.04 となった．

この結果から，cessa 区間，tanto 区間ともに SFR より Q 値のほうが時間的な変化が大きく見られた．

3.3 考察

分析結果から，プロ歌手における提案指標の時間的な変化を声種間の相関から考察する．

まず， 声種間における SFR と Q 値の相関係数を表 2 に示す. 表 2 中の*印は個数に応じて各相関係数の信頼区間を表す．cessa 区間では，SFR が 0.25 と弱い相関，tanto 区間は 0.54 とやや強い相関が確認された．tanto 区間は SFR，Q 値ともに 0.70 以上の強い相関が確認された．これらのことから，女性プロ歌手において声種間の時間的変化の傾向には相関があることが示唆される．しかし，cessa 区間の SFR のみ相関が弱かった．これは，相関係数の有意水準が cessa 区間の SFR のみ $p<.05$ となったため, 偶発的な結果の可能性が考えられる．

4. 異なる習熟度群の時間的変化

本章では，3 章と同様に 2 章で示した指標を用いて，習熟度の異なる被験者群同士の歌声の時間的変化を分析する．また，提案指標における習熟度の差の傾向や相関について述べる．

4.1 分析条件

被験者群は，プロ歌手，音楽大学生群，教育学部生群の 3 つとする．プロ歌手については，3 章において提案指標の時間的変化における声種間の相関が強いことと音楽大学生群と教育学部生群の全ての被験者の声種がソプラノであることから, 被験者 A をプロ歌手の代表値にすることとした．

音楽大学生群は音楽大学の声楽科に所属する大学 1 年生の女性 7 名である．教育学部生群は教育学部音楽教育コースに所属する大学 2 年生の女性 7 名である．2 つの学生被験者群は学年が異なる．これは，音楽大学生群が声楽レッスンを受け始めるのは大学 1 年生であるのに対して，教育

表 2　声種間の相関係数

	SFR	Q 値
cessa	0.25*	0.53***
tanto	0.78***	0.70***

*：$p<.05$, **：$p<.01$, ***：$p<.001$

ⓒ 2018 Information Processing Society of Japan

Vol.2018-MUS-119 No.13
Vol.2018-SLP-122 No.13
2018/6/16

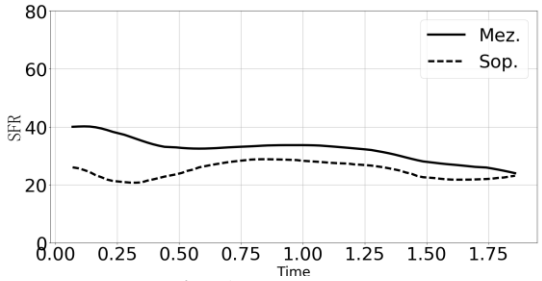


図 4　プロ歌手の cessa 区間の SFR

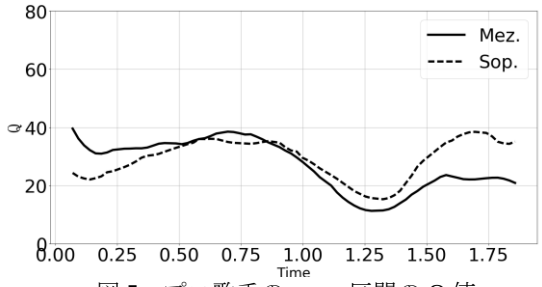


図 5　プロ歌手の cessa 区間の Q 値

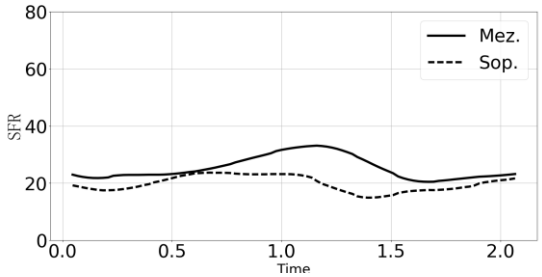


図 6　プロ歌手の tanto 区間の SFR

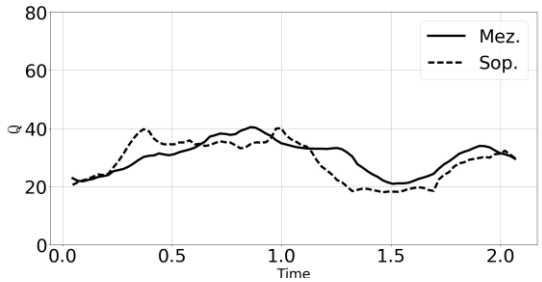


図 7　プロ歌手の tanto 区間の Q 値

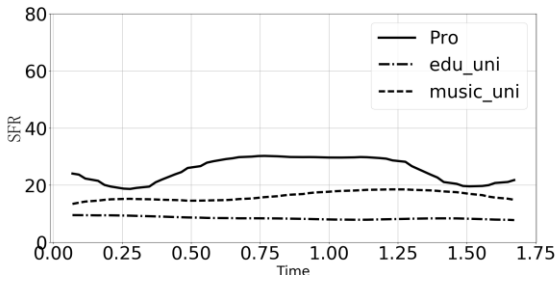


図 8　3 群の cessa 区間　SFR

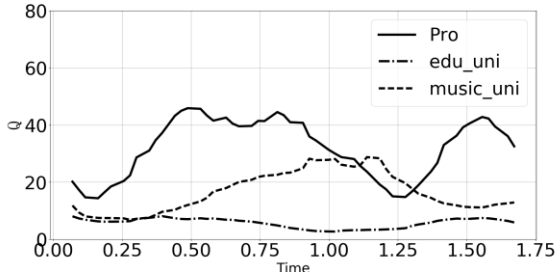


図 9　3 群の cessa 区間　Q 値

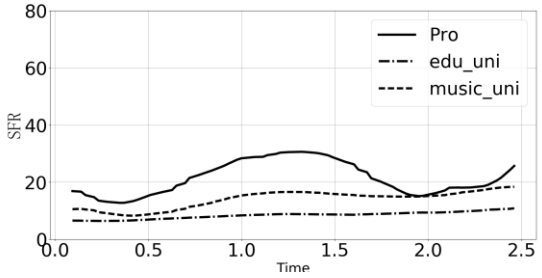


図 10　3 群の tanto 区間　SFR

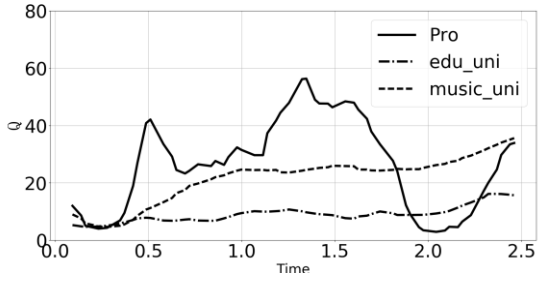


図 11　3 群の tanto 区間 Q 値

学部生群は大学 2 年生からであることが理由である．音楽大学生群，教育学部生群ともにそれぞれの大学のカリキュラムに沿ったレッスンを継続的に受けている．音楽大学生群と教育学部生群の収音時期はいずれも 2017 年 4 月である．収音環境，分析方法，分析対象区間，分析データの可視化方法等の分析環境は 3 章と同様である．

4.2 分析結果

分析の結果を 3 章と同様の手法で図 8～11 に示す．いずれのグラフも横軸が時間軸であり，縦軸は SFR または Q 値である．図 8 は cessa 区間の SFR，図 9 は cessa 区間の Q 値，図 10 は tanto 区間の SFR，図 11 は tanto 区間の Q 値を表す．また，グラフ中のラベルは Pro がプロ歌手，music_uni が音楽大学生群，edu_uni が教育学部生群である．

これらの結果から，SFR は全区間でプロ歌手，音楽大学生群，教育学部生群の順番に数値が高いことが確認できる．Q 値は一部の時間を除いて プロ歌手，音楽大学生群，教育学部生群の順番に数値が高い．また，3 章と同様に cessa 区間，tanto 区間ともに SFR より Q 値のほうが時間的な変化が大きく見られた．Q 値の中でも，特にプロ歌手においては時間的変化の差は cessa 区間が 31.51（最大値 45.88，

ⓒ 2018 Information Processing Society of Japan

最小値 14.37）で，tanto 区間が 53.42（最大値 56.33，最小値 2.91）であった．一方，教育学部生群においては cessa 区間が 5.39（最大値 8.11, 最小値 2.72）で, tanto 区間が 11.47（最大値 16.16，最小値 4.69）であった．

プロ歌手の差は教育学部生群と比べ, cessa 区間が 5.8 倍，tanto 区間が 4.7 倍となり，特に大きな差となった．

4.3 考察

分析結果から，音楽大学生群，教育学部生群をプロ歌手と比較し，習熟度の差を提案指標の時間的変化によって表すことができるかについて考察する．その際，平均二乗誤差と相関係数の 2 つの側面から考察する．分析結果は以下の表 3 に示す．3 章と同様に表 3 中の*印は個数に応じて各相関係数の信頼区間を表す．

4.3.1 相関係数

ここでは，3 章と同様に各群の波形の相関係数を算出し，波形の相関から習熟度の差を表すことができるかについて述べる．表 3 に相関係数を算出した結果を示す．

表 3 の結果に対して，相関係数の有意水準が p<.01 以上の数値のみに着目する．プロ歌手と音楽大学生群は，tanto 区間の SFR のみに強い相関（0.62）が確認された．また，プロ歌手と教育学部生群では cessa 区間の SFR は強い負の相関（-0.50），tanto 区間の SFR はやや強い相関（0.36）が確認された．このことから tanto 区間の SFR においては，音楽大学生群は教育学部生群よりもプロ歌手に近い結果となっていることが示唆される．

また, 音楽大学生群と教育学部生群は，cessa 区間の SFR（-0.69）と Q 値（-0.86）にとても強い負の相関があり, tanto 区間の SFR（0.92）と Q 値（0.83）にとても正の強い相関が確認された．このことから，音楽大学生群と教育学部生群の 2 つの学生群には, tanto 区間で極めて類似した結果を示す可能性が示唆される．一方，cessa 区間の結果に関連して，先行研究[8]において，歌唱における母音の/e/は，習熟度が低い場合でも提案指標の値が大きくなる可能性が示されている．今回の場合，音楽大学生群の cessa 区間は瞬間的にプロ歌手に匹敵する程度に高い値が出ていた可能性が考えられる．

4.3.2 平均二乗誤差

ここでは，プロ歌手を真値であると仮定し，プロ歌手と音楽大学生群，プロ歌手と教育学部生群との「誤差」（平均二乗誤差）と習熟度との関係を検討する．

表 3 の結果から，評価対象区間，評価指標に関わらずプロ歌手との平均二乗誤差が音楽大学生群の方が教育学部生群より少なくとも 1/2 以下となった．ここから，プロ歌手がその他の群と比べて評価指標の数値が高く，さらに音楽大学生群が教育学部生群と比べてプロ歌手との差が小さいことが明らかとなった．これは，異なる習熟度の差を本評価指標の時間的変化で表わせることを示唆していると考えられる．また，音楽大学生群と教育学部生群の平均二乗誤差は, tanto 区間の Q 値を除いてプロ歌手との比較時よりも小さかったため，これら 2 群の差はプロ歌手との差よりは小さいことが示唆される．

さらに，時間的な変化量について分析を行うために各群

表 3　分析結果のまとめ

		相関係数	平均二乗誤差	微分値の平均二乗誤差
プロ歌手 vs. 音楽大学生群	cessa 区間　SFR	0.34*	98.49	0.46
	tanto 区間　SFR	0.62***	74.77	0.62
	cessa 区間　Q 値	0.08	378.60	4.58
	tanto 区間　Q 値	0.29*	199.52	14.73
プロ歌手 vs. 教育学部生群	cessa 区間　SFR	-0.50***	303.38	0.54
	tanto 区間　SFR	0.36**	262.24	0.84
	cessa 区間　Q 値	0.28*	778.18	5.83
	tanto 区間　Q 値	0.16	537.08	17.04
音楽大学生群 vs. 教育学部生群	cessa 区間　SFR	-0.69***	63.31	0.035
	tanto 区間　SFR	0.92***	188.03	1.22
	cessa 区間　Q 値	-0.86***	36.77	0.07
	tanto 区間　Q 値	0.83***	202.67	0.46

*：p<.05 , **：p<.01, ***：p<.001

ⓒ 2018 Information Processing Society of Japan

の波形の微分値に対して平均二乗誤差を算出した．表 3 の結果から，評価対象区間，評価指標に関わらずプロ歌手との微分値の平均二乗誤差が音楽大学生群の方が教育学部生群より少なくとも 4/5 以下となった．平均二乗誤差同様に，プロ歌手がその他の群と比べて評価指標の数値が高く，さらに音楽大学生群が教育学部生群と比べて数値の差が明らかとなった．また，音楽大学生群と教育学部生群との微分値の平均二乗誤差は，tanto 区間の SFR を除いてプロ歌手との比較時よりも小さかったため，これら 2 群の差はプロ歌手との差よりは小さいことが示唆される．

しかし，音楽大学生群と教育学部生群の tanto 区間 Q 値の平均二乗誤差と音楽大学生群と教育学部生群の tanto 区間 SFR の微分値の平均二乗誤差においてプロ歌手との差よりも大きな差となった．これは，プロ歌手 1 名を代表値としたことが影響したと考えられる．今後は複数名のプロ歌手の平均値を用いて比較分析を行う必要があろう．

5. おわりに

本稿では，歌声の響きに関連する考えられる周波数成分の時間変化による定量化を試みた．声種の異なるプロ歌手間，および習熟度の異なる被験者群同士の時間変化の分析結果を示した．

具体的には，まず本研究で提案している評価指標である SFR と Q 値を用いて複数名のプロ歌手に対して声種間の時間的変化を分析した．その結果から，女性プロ歌手においては提案指標の時間変化により強い相関が表れることが示唆された．

次に，習熟度の異なる被験者群を対象に，提案指標の時間的変化から個々の群の区別の可否について分析した．その結果，提案指標を用いることにより習熟度の異なる各群を区別できる可能性が示唆された．また，各群同士の相関係数から，プロ歌手と初学者の区別が可能であることも示唆された．さらに，音楽大学生群と教育学部生群の 2 群の相関が確認されたことにより，声楽初学者に表れる傾向を提案指標で表せる可能性についても示すことができた．

本稿では女性に特化した分析を行った．今後は，男性を対象とした同様の分析を行い，提案指標の妥当性について検証していく．また，提案指標と実際の指導現場で用いる指導用語等との対応についても分析を行っていく．

謝辞 本研究の一部は科研費 15K01022 と 18K02817 の支援を受けた．

参考文献

[1] W.T.Bartholomew. A Physical Definition of "Good Voice- Quality" in the Male Voice, J. Acoust. Soc. Am., 6(1), p.25-33, 1934.
[2] J.Sundberg. The KTH synthesis singing. Advances in Cognitive Psychology , Special issue on Music Performance, 2(2-3), p.131 -143, 2006.
[3] J. Sundberg. Articulatory interpretation of the ‘singing formant’. J. Acoust. Soc. Am., 55(4), p.838-844, 1974.
[4] 齋藤毅他．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. 64(5), p.405-417, 2008.
[5] 山辺大貴他．歌声の心理的印象と音響特徴量との対応付けによる歌唱の習熟度評価に関する基礎的検討．信学技報応用音響. 112(266), p.61-66, 2012.
[6] 野田美春他．歌唱の習熟度に関連する周波数特性に基づく初学者とプロ歌手の音響特徴量に関する考察．信学技報教育工学. 115(444), p.35-40, 2016.
[7] 佐久間雄輝他．歌声の習熟度に関連する周波数特性からみた音響特徴量．信学技報教育工学. 114(441), p.45-50, 2015.
[8] 山下泰樹他．歌声の習熟度に関連する音響特徴量の母音分布．信学技報教育工学．情報処理学会音楽情報科学研究会, Vol.2018-MUS-118, No.26, pp.1 - 6, 2018.

ⓒ 2018 Information Processing Society of Japan