

歌声の習熟度に関連するスペクトル包絡に基づく 評価指標に関する基礎的検討

A Basic Study on Evaluation Indicators Based on Spectral Envelope Related to Singing Skill

矢川 実紗^{*1}, 香山 瑞恵^{*1}, 池田 京子^{*2}, 山下 泰樹^{*3}, 小畑 朱実^{*4}, 山口 道子^{*4},
谷 友博^{*4}, 永井 孝^{*5}, 召田 優子^{*6}, 浅沼 和志^{*6}, 伊東 一典^{*1}, 館 伸幸^{*1}
Misa YAGAWA^{*1}, Mizue KAYAMA^{*1}, Kyoko IKEDA^{*2}, Yasuki YAMASHITA^{*3}, Akemi OBATA^{*4}
Michiko YAMAGUCHI^{*4}, Tomohiro TANI^{*4}, Takashi NAGAI^{*5}, Yuko MESUDA^{*6}
Kazushi ASANUMA^{*6}, Kazunori ITOH^{*1}, Nobuyuki TACHI^{*1}

^{*1} 信州大学, ^{*2} 椋山女学園大学, ^{*3} 長野県工科短期大学校

^{*4} 武蔵野音楽大学, ^{*5} ものつくり大学, ^{*6} 国立高専機構長野高専

^{*1} Shinshu University, ^{*2} Sugiyama Jogakuen University, ^{*3} Nagano Prefectural Institute of Technology,

^{*4} Musashino Academia Musicae, ^{*5} Institute of Technologists,

^{*6} National Institute of Technology, Nagano College

Email: 19t2155k@shinshu-u.ac.jp

あらまし : 本研究の目的は, 歌声の習熟度に関連するスペクトル包絡に基づいた評価指標の提案である. 習熟度に関連している評価指標としてスペクトル重心, スペクトル傾斜, スペクトルフラックスが報告されている. これらは, 先行研究で検討されていた Singer's formant 帯域には限定されない評価指標の可能性はある. そのため, 本研究ではこれらの評価指標と習熟度の関連をプロ歌手と音楽大学生と教育学部生の3群で検討した. その結果, 特に, スペクトル傾斜を用いた評価指標開発の可能性が示唆された.

キーワード : 歌声, 音響特徴量, 音大生, スペクトル重心, スペクトル傾斜, スペクトルフラックス

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. 我々は, 声楽を学ぶ学生を対象に, 歌声の音響特徴量に基づく学習成果の定量化を探究している^(1,2). これまでに Singer's formant の強さと割合に関する歌声評価指標の有効性を確認してきた. 本稿では, スペクトル包絡に基づく歌声評価指標の可能性を検討する.

2. スペクトル包絡に基づく音響特徴量

2.1 スペクトル包絡

スペクトル包絡は音の形を表す. 歌声の声質を特徴づける重要な特徴量の1つであり, 先行研究においても様々な検討がなされている⁽³⁾. 本研究では, 各時刻 t におけるスペクトル包絡 $S_{lin}(f, t)$ 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.

2.2 SFC (Second Formant Convex)

歌声に対して LPC 包絡線を求めると凸部が出現することがある. SFC とは, LPC 包絡線の2番目に現れる凸部の周波数 (以下, SFC) を指す. SFC の中央値と平均値, 標準偏差を算出し, そこから変動係数を求める.

2.3 スペクトル重心

スペクトル重心 (以下, S_c) は, 音の明るさの特徴に対応するとされている⁽⁴⁾. ある周波数において, S_c は高ければ高いほど, S_c よりも高い周波数をもつ特

徴と対応する. スペクトル包絡 $S_{lin}(f, t)$ から, 各時刻におけるスペクトル包絡の重心 $S_c(t)$ を, 式(1)によって求め, 中央値と変動係数を算出する. B は周波数ビンの数を示している. より明るい音になればなるほど, S_c は高い値を示す.

$$S_c(t) = \frac{\sum_{f=1}^B (f \cdot S_{lin}(f, t))}{\sum_{f=1}^B S_{lin}(f, t)} \quad (1)$$

2.4 スペクトル傾斜

スペクトル傾斜 (以下, S_t) は, 感情の特徴に対応するとされている⁽⁵⁾. 感情の程度が大きくなるにつれ, 高周波帯域が強調され, S_t が増加するとされている. 式(2)を用いてスペクトル包絡 $S_{lin}(f, t)$ から, 時間ごとの傾きを求める. 3種類の帯域(0-3kHz, 0-6kHz, 0-9kHz)における S_t を求め, 中央値と変動係数を算出する.

$$S_t(t) = \frac{\sum_{f=1}^B f \cdot S_{lin|log}(f, t) - \sum_{f=1}^B f \sum_{f=1}^B S_{lin|log}(f, t)}{\sum_{f=1}^B f^2 - (\sum_{f=1}^B f)^2} \quad (2)$$

2.5 スペクトルフラックス

スペクトルフラックス (以下, S_f) も音の明るさの特徴に対応するとされており, 局所的なスペクトルの変動量を表す⁽⁴⁾. 時刻 t のフレームにより標準化されたスペクトル包絡 $S_{lin}(f, t-1)$ を用い, 式(3)によりスペクトルフラックス $S_f(t)$ を求め, 中央値と変動係数を算出する.

$$S_f(t) = \sum_{f=1}^B (S_{lin|log}(f, t) - S_{lin|log}(f, t-1))^2 \quad (3)$$

表 1 歌声データ数と声種(計 304 データ)

曲名 声種\ 被験者群	故郷		Caro mio ben	
	男声	女声	男声	女声
教育群	8	47	9	43
音楽群	13	73	14	61
プロ歌手群	11	11	9	5

表 2: 変動係数の検定結果(* : p<.05,** : p<.01,*** : p<.001)

評価指標	SFC		スペクトル傾斜 (0.3kHz)		スペクトル傾斜 (0.6kHz)		スペクトル傾斜 (0.9kHz)	
	かのや[m[a] 区間]	[t[a]nto 区間]	かのや[m[a] 区間]	[t[a]nto 区間]	かのや[m[a] 区間]	[t[a]nto 区間]	かのや[m[a] 区間]	[t[a]nto 区間]
プロ歌手群 vs. 教育学部生群	***	***	***	***	***	***	***	***
音楽大学生群 vs. 教育学部生群	***	***	***	p=.225	**	p=.682	p=.248	p=.978
プロ歌手群 vs. 音楽大学生群	p=.017	p=.04	***	***	***	***	***	***

3. 分析条件

3.1 分析対象楽曲と歌声収録条件

分析で使用する楽曲は「故郷」(作詞:高野辰之,作曲:岡野貞一)と「Caro mio ben」(作詞:不明,作曲:Tommaso Giordani)とした。各被験者は「故郷」では“かのやm[a]”区間,「Caro mio ben」では“t[a]nto”区間を評価対象区間とした。各被験者につき1回の収録で2回ずつの歌唱をさせた。収録は音の反響しない静かな部屋で行い,レコーダはLS-P2(OLYMPUS)を使用した。被験者にはレコーダから2m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44.1kHz,量子化ビット数16bitで収録した。

3.2 被験者のプロフィール

分析対象とした歌声は,オペラ歌手等のプロフェッショナルな歌手(以下,プロ群あるいはプロ歌手)の歌唱と,教育学部音楽教育コースの2~4年生(以下,教育群あるいは教育生)と音楽大学声楽科在学中の学習者(以下,音大群あるいは音大生)とした。「故郷」と「Caro mio ben」の被験者の歌声データ数と声種を表1に示す。「故郷」は教育生44名,音大生21名,プロ歌手13名の歌唱データを対象とする。「Caro mio ben」は教育生40名,音大生21名,プロ歌手11名のデータを対象とする。

4. 分析結果

4.1 習熟度における音響特徴量の考察

SFC, S_C , S_t , S_f を用いることによって異なる習熟度の群同士を統計的仮説検定から区別できる可能性について考察する。

各指標の変動係数に対して各習熟度の群同士の対応,正規性がないため,有意水準1%でマンホイットニーのU検定をした後,ボンフェローニ補正を行った結果を表2に示す。表2の結果から,かのやm[a]とt[a]ntoの区間において共通した結果を得た提案指標はSFCと S_t であった。

SFCは教育群とプロ群,音大群を比較すると有意な差となった。このことから,SFCを用いることによって,教育生より専門的に声楽を学んでいるプロ歌手と音大生の習熟度の差を区別できることが示唆される。そして,SFCでは,習熟度が高くなるにつれ,周波数の中央値がSinger's formantに相当する周波数帯域(2.4~4.0kHz)に集中していた。SFCは,習熟度の差を区別できる可能性がある。

S_t はプロ群と教育群,音大群を比較すると有意な差となった。このことから, S_t を用いることによって,教育生・音大生とプロ歌手との習熟度の差を区別できることが示唆される。 S_t では,習熟度が高くなるにつれ,中央値が大きくなった。 S_t は感情の程度が大きくなるほど増加すると言われている。プロ歌手の方が音大生・教育生より歌声に感情が表れていると考えられる。そのため, S_t は習熟度の差を区別できる可能性がある。

4.2 声種における音響特徴量の考察

SFC, S_C , S_t , S_f を用いることによって異なる声種の群同士を区別できる可能性について考察する。

かのやm[a]とt[a]ntoの区間で共通した結果が出た提案指標はSFCと S_t であった。まず,高い声域の声種ほど,SFCの変動係数と中央値が大きくなった。そのため,低い声域の声種ほどSFCが安定していると考えられる。そして,声種の違いを中央値に反映できていることがわかる。また,声が低いほど S_t の中央値が大きくなった。声が低いほど,計算の特性により S_t が大きくなる可能性がある。

このことから,SFCと S_t は歌声を評価する上で声種により違いがでることが示唆される。

5. おわりに

本稿ではスペクトル包絡に基づく歌声評価指標としてSFC, S_C , S_t , S_f の4種により,習熟度と声種の関連を検討した。その結果,プロ群と音大群と教育群との比較からSFCと S_t は習熟度と声種の区別に有効である可能性を見出した。 S_C と S_f は,曲の長さを変更するなどして検討していきたい。

謝辞 科研費18K02817・22K00237の支援を受けた。

参考文献

- (1) 吉田祥他,“声楽発声の習熟度に関連する音響特徴量に基づく歌声の評価指標の提案”電子情報通信学会論文誌D, Vol.J103-D, No.4, pp.247-260(2020)
- (2) 山下泰樹他,“歌声の習熟度に関連する音響特徴量の母音別分布,”情処研報音楽情報科学, vol.118, no.26, pp.1-6,(2018).
- (3) Sundberg, J.:“The Science of the Singing Voice”, Northern Illinois University Press(1987)
- (4) Tzanetakis, G. and Cook, P.:“Musical Genre Classification of Audio Signals, IEEE Trans. Speech and Audio Processing”, Vol.10, No.5, pp.293-302(2002)
- (5) Yoshiki KABUTA et al “Comparison of Spectral-tilt Features of Emotional Speech Depending on the Degree of Emotions – Research on Emotional-speech Synthesis based on Voice-quality Conversion –”, International Journal of Affective Engineering, Vol.12, No.2, pp.161-167(2013)