

[研究助成 (A)]

人と AI が調和する技能伝承支援基盤の構築

—— 琉球古典音楽・歌三線の普及に向けて ——

Development of Skill Transfer Support System for Uta-sanshin based on
Harmony between Human and AI

2201027



研究代表者

沖縄工業高等専門学校
情報通信システム工学科

講 師

宮 城

桂

[研究の目的]

我が国は、諸外国の人達をも魅了する有形・無形の文化芸術資源を有している。昨今では、AIをはじめとする科学技術を総合的視点で捉え、これらの資源から生み出される新たな価値の創出と伝統文化の持続的な継承を図る事が求められている（ものづくり白書，経産省，2019）。沖縄の伝統芸能である組踊歌三線は、ユネスコの無形文化遺産にも登録されており、舞踊と音楽で展開される歌舞劇である。組踊の音楽は、琉球古典音楽の一つでもある三線を弾きながら琉歌を歌う歌三線が中心で、登場人物の心情を切々と歌いあげる。これらの「技」が作り出す音色は咽喉の開閉によって変化するとされ、開閉が上手くいかないと音色の深遠な複雑さが出ないとされているが、その科学的解明には至っていない。

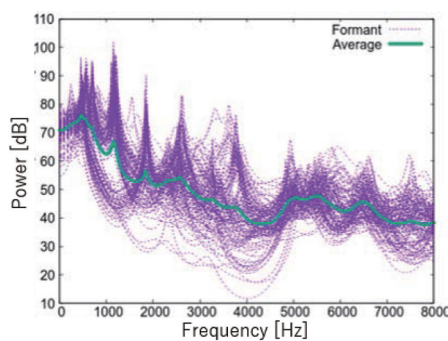
理想的な歌声の一つの指標となる音響的特徴量として、咽喉や声道の形状と関係するシンガーズ・フォルマント（Sundberg, JASA, 1974）が知られている。これは、歌声をスペクトル分析して得られるものであり、優れた歌い手に共通して生起するとされている。ただ、歌唱時の歌詞（母音・子音）と旋律（音高・リズム）の変化に伴い複雑に遷移するため、その生起や特性の時間的なふるまいについては、技術的扱いが難しく十分に調査されていないとされ

ている（高橋，京都芸大紀要，2017）。

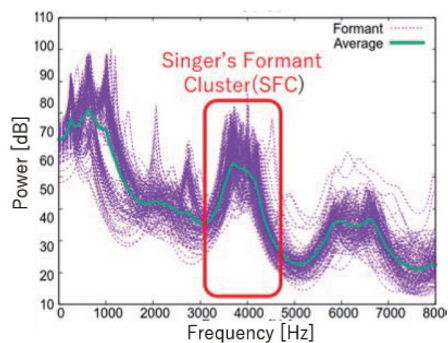
本研究では、シンガーズ・フォルマントに着目し、熟達した歌い手に共通する音響特徴を明らかにする。次に、フォルマントの構造を形作る音色の旋律（多次元ベクトルとして表出するスペクトル包絡の時間推移）に着目し、その音響特徴の次元を圧縮して可視化することで声質や音色の豊かさを含めた評価ならびに訓練支援を実現する基盤技術を開発する事を目的としている。これにより、熟練者の「技」をAIに転写し、AIと学習者の「調和的協働」を通じて歌三線の伝承・保存や新たな価値の創出に貢献する事を目指す。

[研究の内容、成果]

本研究では、先ず3クラスに分類した歌い手（熟練者，経験者，初心者）の歌唱時のフォルマントを解析し、熟練者には共通する音響特徴がある事を見い出した。図1に琉球古典音楽『歌三線』の経験者と熟練者を対象に評価したフォルマントを示す。このグラフはフレーム長を46msとし、歌唱時の音高が低音から高音に変化する際のフォルマント（スペクトル包絡）を約6秒間プロットしたもの（破線）と、その平均値をプロットしたもの（実線）であるが、評価結果から以下の2つの知見が得られた。
・熟練者はシンガーズ・フォルマントという特



(a) 経験者のフォルマント



(b) 熟練者のフォルマント

図1 経験者と熟練者のフォルマント

定の周波数帯域（2400-4300 Hz）を強調して歌い、その帯域に複数のパワースペクトルを集中させる事でシンガーズ・フォルマント・クラスタ SFC (Kenneth, Practical Vocal Acoustics, 2013) を形成する。

・2400-4300 Hz 付近に形成されたシンガーズ・フォルマント・クラスタは、音高の変化によらず周波数・パワー共に安定する傾向にある。

これらの知見も基づき、以下の5つの音響特徴を定義した。

(1) シンガーズ・フォルマント・クラスタ SFC の周波数の安定性の評価：これに関しては、楽曲に対して十分に短い周期（例えば 50 msec ごと）に得られたシンガーズ・フォルマント・クラスタの複数のパワーのピーク値に対応する周波数 x_i から以下の式に基づき、それぞれの標準偏差 σ を求める。

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

さらに、分散の加法性に基づき、それら総和を

安定性の評価値とし、その値が小さいほど習熟度が高いと考えられる。

(2) シンガーズ・フォルマント・クラスタ SFC のパワーの安定性の評価：これに関してはシンガーズ・フォルマント・クラスタから得た複数のパワーのピーク値から (1) で示した方法で標準偏差 σ を求め、その値を平均値 $\bar{x}$ で割る事で変動係数CVを求める。ここで得られた CV の総和を安定性の評価値とし、その値が小さいほど習熟度が高いと考えられる。

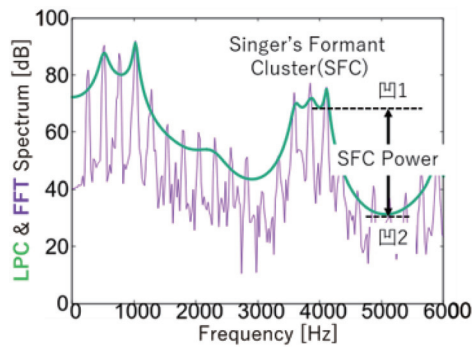
$$CV = \frac{\sigma}{\bar{x}}$$

(3) シンガーズ・フォルマント・クラスタ SFC の強度の評価：図 2 に示すグラフは、フレーム長 46 ms とした場合のスペクトルとその包絡線を図示したものである。シンガーズ・フォルマントの強度は図 (a) に示すように、シンガーズ・フォルマントのパワーが最も強くなる谷の部分（凹部 1）とパワーが最も低くなる谷の部分（凹部 2）の差として算出する。その値が大きいほど習熟度が高いと考えられる。

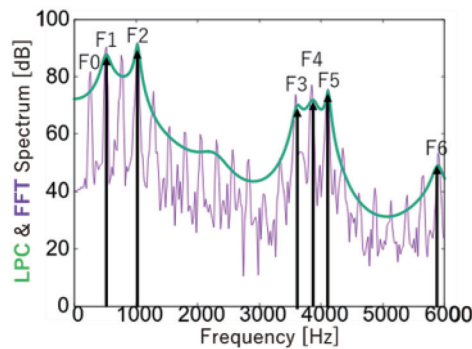
(4) 倍音含有率の評価：図 2 (b) に示すように、基音 (F0) とフォルマントのパワー値 (F1~FN) の比率の平均値として算出する。その値が大きいほど習熟度が高いと考えられる。

(5) シンガーズ・フォルマント・クラスタの集中度の評価：図 2 (c) に示すように、シンガーズ・フォルマントの前後に表れる極小値の範囲内のエネルギーの総和 E_{SF_min} と、極大値の範囲内のエネルギーの総和 E_{SF_max} の比率として算出する。その値が大きいほど習熟度が高いと考えられる。

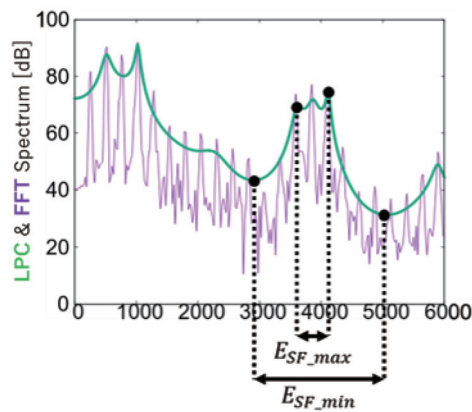
(1)～(5) で示した音響特徴の有効性を確認するために、先行研究（羽賀，人工知能学会，2015）で提案されている手法との比較評価を行った。先行研究では、シンガーズ・フォルマントの Q 値（線形予測符号で得られたシンガーズ・フォルマントのピークの鋭さ）と SFR（2.4～4 kHz と 0～4 kHz の FFT パワースペクトルの合計値の比）の 2次元平面上の分



(a) シンガーズフォルマントの強度



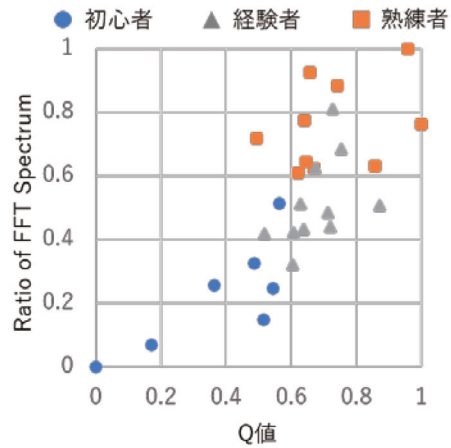
(b) パワー値の比率



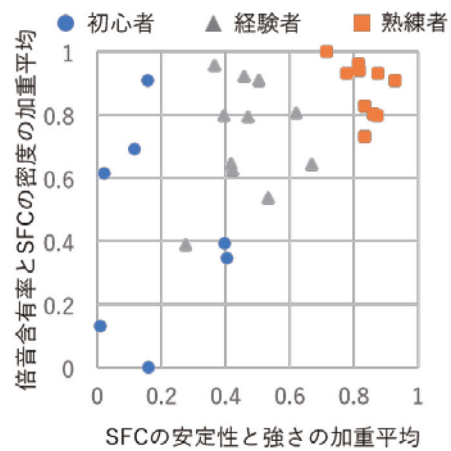
(c) エネルギーの比率

図2 経験者と熟練者のフォルマント

布具合により歌手の習熟度を評価する手法を提案している。評価では、沖縄の古典音楽・歌三線の師範免許を保持している者を熟練者、三線を弾きながら歌える者を経験者、未経験者を初心者と定義し、それぞれ10名、11名、7名の歌唱データを用いて評価した。楽曲として、歌三線の中でも代表的な『かぎやで風節』を対象として評価した結果を図3に示す。縦軸と横軸の値が大きいほど熟達度が高い事を示しており、先行研究で示されている手法と比べても精度良く分類できている事が分かる。これにより、



(a) 先行研究



(b) 提案手法

図3 先行研究と提案手法の比較

歌手の習熟度を定量的に評価することができると示唆された。なお、本手法はリアルタイム処理にも対応できる。本研究では、この歌唱分析の結果を応用し、音響特徴量を可視化することで歌手の発声法の向上の助けとなるシステムの構築を目指している。その実現のために、筆者の知る限り既存技術では扱われていなかった音色の旋律（多次元ベクトルとして表出するスペクトル包絡の時間推移）に着目し、その音響特徴の次元を圧縮して可視化することで声質や音色の豊かさを含めた評価ならびに訓練支援を実現する。なお、音色の時間的なふるまいを分かりやすく可視化するために、熟練者の歌唱を学習したAI（ニューラルネットワーク）を用いて音色を特徴付ける多次元のベクトル（スペクトル包絡）の次元数を圧縮する。音響



図4 シンガーズ・フォルマントの可視化

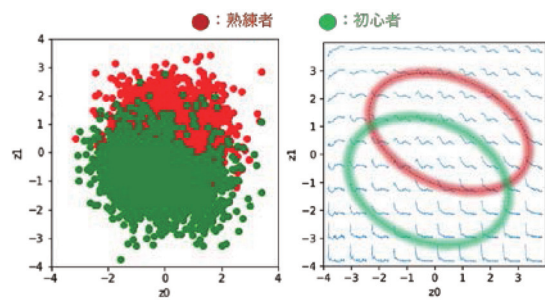


図6 音色旋律の可視化

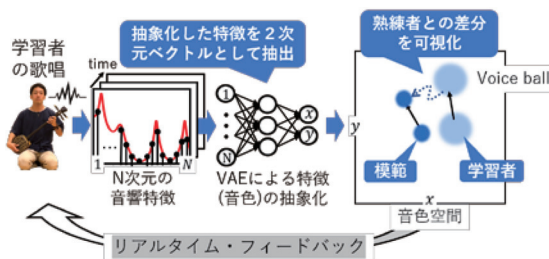


図5 音色旋律の可視化手法

特徴の可視化手法の一例を図4に示す。図4は母音の発声訓練を行っている様子である。図に示すように、歌唱の音色を特徴づけるスペクトル包絡をリアルタイムで可視化するプログラムを実装した。これにより、学習者は熟練者の模範を見ながら発声訓練を行う事ができる。現在、音色の豊かさを評価するために、図5に示すVAE (Variational Auto Encoder) の実装にも取り組んでいる。具体的には、フォルマントから得たN次元の特徴をVAEで2次元に圧縮し、フォルマントからは直接観測できない抽象化した特徴をベクトルとしてマップ上にプロット (技の転写) する。その上で、熟練者・音響科学双方の視点からベクトルのふるまいと楽譜に記載された歌唱法の関連性を同定し、誰でも認識可能な「形式知」へ変換する。加えて、学習者に視覚的に熟練者の歌唱 (模範) との差をVoice ballで「見える化」し、模範となるベクトル (技能) のふるまいを真似るように発声を反復することで技能の習得を支援するものである。なお、Voice ballのサイズは、先述した(1)～(5)の特徴量に基づき決定する。以上により、フォルマントの時間的ふるまいから抑揚

や歌唱の表現力・声の響きや習熟度を学習者に分かりやすく可視化するアプリケーションの開発に取り組んでいる。

図6はVAEを用いて多次元のスペクトル包絡を2次元にまで圧縮して2次元平面上にプロット (赤=熟練者、緑=初心者) したものである。VAEを用いる事で熟練者と初心者を分類でき、2次元で表されるデータ (プロット) から歌い手のスペクトル包絡が再現できる事を確認した。両者のスペクトル包絡の形状を比較する事により、明らかな差 (熟練者のスペクトルは初心者と比較して倍音が多く含まれており、4000 Hz付近にシンガーズ・フォルマントが生起する) が表れる事が確認できた。これは、AI (VAE) が音色 (スペクトル包絡) から音響特徴を抽出できている事を示している。このように、AI (VAE) が時分割されたフォルマントをエンコードして類似度に基づいて分類 (クラスタリング) し、包含関係にない部分 (熟練者に共通する音響的特徴) をデコードする事で習熟度の識別に強く関わる特徴量を可視化する点に独自性がある。ここで得られた音響特徴により、学習者は自らの音響特徴の課題を認識・理解することができると考えている。

[今後の研究の方向、課題]

これまでに、シンガーズフォルマントに着目し、熟練者、経験者、初心者を精度よく分類するための音響特徴量と音色旋律に基づく技能の可視化手法を構築した。現在は、時間的に変化

する音色旋律を可視化するために、組み込み機器上でVAEのリアルタイム処理を実現させるための手法の構築に向け取り組んでいる。今後は、これら技術を統合し、Voice ballを用いた習熟度可視化アプリケーションを開発する予定である。

その他の課題として、歌三線における楽譜の工工四には約20種の声楽記号と、これに対応する歌唱法の解説が記載されているが、喉を「圧迫する」「えぐる」など曖昧かつ難解な説明が数多く存在し、技能をうまく伝承する上での障壁となっている。これらの「技」が作り出す音色は咽喉の開閉によって変化するとされ、開閉が上手くいかないと音色の深遠な複雑さが出ないとされる。そこで、今後は、声楽記号ごとの音響特徴についても検討し、熟練者の深淵な技能の一端を明らかにしつつ、そのかけがえの

ない『価値』を後世に正確に伝えるための基盤技術を構築したいと考えている。

[成果の発表, 論文等]

- [1] 我如古永士, 宮城桂, 山田親稔, 市川周: “琉球古典音楽『歌三線』における歌唱法の音響的特徴,” 電気学会次世代産業システム研究会, IIS-20-061 (2020).
- [2] 久高萌永, 宮城桂: “オートエンコーダを用いた歌唱技能の可視化手法に関する検討,” 高専・豊橋技科大IoT共同研究会, (2021).
- [3] 市川周一, 久高萌永, 金城湾, 宮城桂, 山田親稔: “琉球古典音楽『歌三線』における技能伝承支援AIのための並列計算モデル,” 先端的技術シンポジウムATS2021, B-3-1 (2022).
- [4] 久高萌永, 宮城桂: “音響特徴を用いた歌唱訓練支援システム,” 社会実装教育フォーラム (2022).
- [5] 宮城桂, 山田親稔, 市川周一: 特願 2022-58703 歌唱指導支援装置, その判定方法, その音響特徴の可視化方法およびそのプログラム (2022).