

歌唱時の筋活動の可視化による学習支援システムの提案

Proposal for a Learning Support System
by Visualizing Muscle Activity During Singing

勝井 拓和^{*1}, 東本 崇仁^{*2}
Takuto KATSUI^{*1}, Takahito TOMOTO^{*2}

^{*1} 千葉工業大学情報科学部

^{*1} Faculty of Information and Computer Science, Chiba Institute of Technology

^{*2} 千葉工業大学情報変革科学部

^{*2} Faculty of Innovative Information Science, Chiba Institute of Technology

Email: s2232039HA@s.chinakoudai.jp

あらまし：歌唱能力を向上させる代表的な訓練法としてボイストレーニングが挙げられる。しかし、従来の指導では学習者自身の筋肉の動きと理想的な発声における身体操作を具体的に可視化・比較することは困難である。そこで本研究では、音声データから呼吸や発声に関わる筋肉の動きを推定し、学習者が筋肉の使い方を直感的に理解できるよう、可視化する手法を提案する。本システムにより、可視化された筋肉の動きに基づいて学習者が自律的に発声を改善できる環境の提供を目的とする。

キーワード：歌唱、学習支援システム、知識構造、可視化

1. はじめに

歌唱力の向上を目的とした訓練方法の一つにボイストレーニングがあり、学習者と指導者の対面形式や動画コンテンツを用いた自主学習形式でも実施されている。いずれも呼吸法の取得や発声の基礎訓練を通じて、声の安定性と音量・響きの豊かさなど音声の質的向上が期待される。

こうした指導においては、筋肉の意識や操作に基づいたアプローチが用いられているが、その際「後頭部から糸で引っ張られているような感覚で姿勢を整えて」といった比喩的表現を用いて身体感覚の理解を促すことがある。しかし、舌筋や声帯筋、横隔膜などは日常において意識的に操作する機会が少なく、どのように動かすべきかを直感的に把握しづらい。また、指導の際に用いられる比喩的な表現がすべての学習者にとって的確な身体操作に繋がることは限らず、理解を妨げる要因となることも考えられる。

本研究では、こうした肉眼では把握しにくい筋肉の動きや力の入り具合をシステム上で推定・可視化し、根拠に基づくフィードバックを行う支援方法を検討する。本手法を通じて学習者が筋肉の使い方を具体的に理解し、修正できることが期待される。本稿では、歌唱における筋活動の可視化の意義とその方法、フィードバック内容について検討する。

2. 関連研究

歌唱力の向上を指向した学習支援システムに Voice Ball⁽¹⁾があげられる。このシステムでは、歌声の音響特徴量である Q 値（声の響きの鋭さを示す指標）と SFR（響きの中でも特に重要とされる高周波成分の割合）に着目し、それらの値を二次元グラフとして可視化することで、学習者は自身の発声状態を直感的に把握することができる。しかしながら、発声に関与する身体的な要素、特に筋肉の使い方と

いった身体操作までは扱っていない。また、Chen ら⁽²⁾は表面筋電図（EMG）センサによる筋活動と音響的特徴を基に、プロの歌手と学生の発声における筋活動と音響特性の相関を比較分析している。結果として、熟練者は EMG と SPR（声の「強さ」や「抜けの良さ」を評価する指標）間に高い相関が見られ、発声技術の熟達に伴う筋制御の違いが明らかとなり、歌唱力の向上には筋肉を意識的に操作できる能力の習得が重要であることが示唆された。したがって、筋活動を可視化し、理解を促進する支援は、初学者にとって有効な学習手段となりうると考える。

3. 提案手法

歌唱力向上には、発声に関与する筋肉を意識的に操作できることが重要である⁽²⁾。本研究では、既存の学習手法では困難な学習者の筋肉の状態に応じた適応的フィードバックの実現に着目し、音声データから筋肉の動きを推定・可視化し、それに基づくフィードバックを行う手法を提案する。筋活動の可視化を通じて、学習者が自身の身体操作をより具体的に理解し、適切に修正できることを目指す。この手法の実現には、(A) 音声・センサによる筋活動の検出、(B) 検出された筋活動に基づいた学習者に対して効果的なフィードバックの2点が課題となる。

まず、本研究では歌唱力の向上を支援するにあたり、その基盤となる知識構造を提案する（図1）。ここでは、歌唱要素を「安定性」「響き」「音質」の3つに大別し、各要素を支える身体的・技術的要因を体系化している。例えば「響き」は、喉や口腔の空きなどの空間的要因に関連し、これらは骨格や筋肉の使い方に大きく影響を受ける。つまり、発声時の筋肉の動きは、この構造の基盤となる要素に直結している。この構造は、実際に40年以上の実務経験を有するボイストレーナーに妥当性を確認済みである。

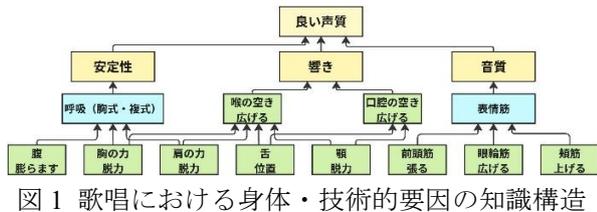


図1 歌唱における身体・技術的要因の知識構造

本研究では、(A)の課題に対しては、学習者の音声データから喉頭周辺や腹部など、発声に関与する複数部位の筋肉の動きを機械学習で推定し、可視化を試みる。これにより、学習者は自身の身体操作が声に与える影響を具体的に把握しやすくなることが期待される。また、(B)の課題に対しては、推定される筋肉の動きと知識構造(図1)に基づく歌唱構成要素(例:安定性、響き)との対応を基に、「腹部の活動が弱い→息の支えが不足している可能性がある」といった形で、筋活動とその意味付けを明示的にフィードバックし、従来の比喩的な指導を補完する。

4. 提案システム

提案システムのフローを図2に示す。提案システムでは、スペクトル特徴やメル周波数ケプストラム係数(MFCC)などを用いて、学習者の音響特徴から筋活動を推定する想定である。次に推定された筋活動情報に基づき、歌唱の品質低下に関与している筋肉を探索する。探索には熟達者の歌唱における筋活動パターンや、図1の知識構造を参照する。

また、音声データから筋活動を測定可能であるかを検討するため、知識構造に基づく予備分析を行った。具体的には、知識構造における喉の空き方に着目し、「喉を空けた状態」と「空けていない状態」で発声・録音・分析した。予備実験はボイストレーナー監修のもと、筋肉の状態に明確な差が出るように実践している。各音声データからFormant特徴量を抽出しグラフ化した結果(図3)、両状態で異なる傾向が示された。この結果は予備的なものであるが、Formantのデータから喉の空き方による発声の違いを評価できる可能性が示唆された。一方、現状の方法では、筋活動を定量的に評価することは困難である。したがって今後の課題として、Formantに加え、音声データからEMG信号を生成するSTE-GAN⁽³⁾の導入により、定量的評価手法の確立を目指す。探索された歌唱の品質低下に関与している筋肉に対しては簡易の2次元アバターとして可視化し、学習者と熟達者の動作の差を明示する。これにより学習者はどの筋肉の使い方に過不足があるかを直感的に理解

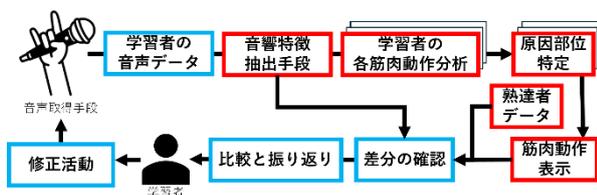


図2 提案システムにおける学習サイクル

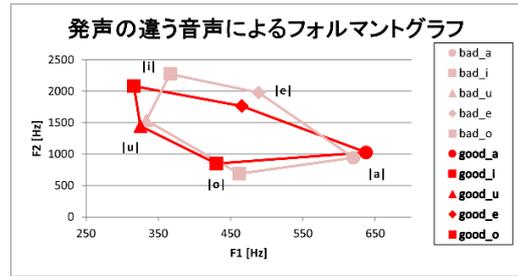


図3 予備実験におけるFormant比較グラフ

できる。

差分を可視化する際のシステム画面を図4に示す。音響特徴量と推定した筋肉の動きから、学習者の改善点を分析・フィードバックし、学習者はこのフィードバックを受け、発声を修正し、再度音声データを提供する。この一連の思考と修正のサイクルにより、歌唱スキルの段階的な向上を期待する。

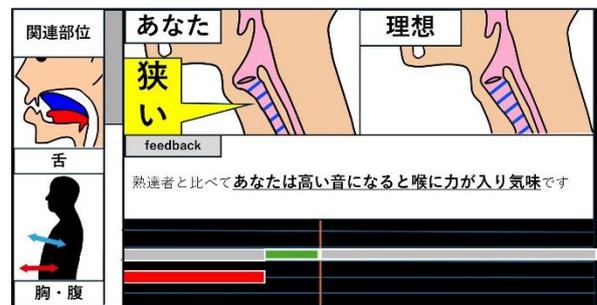


図4 熟達者との差分表示画面の例

5. おわりに

本稿では、学習者の音声データから呼吸や発声に関わる筋肉の動きを推定・可視化し、それに基づくフィードバックを行う手法を提案した。また、学習者が可視化された筋肉の動きに基づき自律的に発声を改善できる支援システムを提案した。

今後の課題として、提案システムの開発と評価が考えられる。本提案システムでの学習を通して、学習者が呼吸器官や発声器官の筋肉を自由に動かせるようになっているかを評価する方法の検討も議論する必要があると考える。

参考文献

- (1) 羽賀翼, 香山瑞恵, 池田京子, 橋本昌巳, 伊東一典: “習熟度に関係する音響特徴量に基づく歌唱学習支援システムの評価”, 人工知能学会第二種研究会資料, SKL-20, pp.01 (2015)
- (2) Chen, Kanyu and Kamiyama, Emiko and Li, Ruiteng and Peng, Yichen and Saito, Daichi and Wu, Erwin and Koike, Hideki and Kato, Akira: “Phantom Audition: Using the Visualization of Electromyography and Vocal Metrics as Tools in Singing Training”, SIGGRAPH Asia 2024 Posters, pp. 1--2 (2024)
- (3) Scheck, Kevin and Schultz, Tanja: “STE-GAN: Speech-to-Electromyography Signal Conversion using Generative Adversarial Networks”, Proc. Interspeech 2023, pp. 1174-1178 (2023)