

アマチュア歌唱者に向けた歌声可視化方法の検討

Study on Visualization of Singing Impression for Amateur Singer

金礪 愛 (Ai Kanato) 指導：菊池 英明

1. 背景と目的

本研究は、歌唱練習の支援を目指し「アマチュア歌唱者が自身の歌声の特徴を把握するための可視化方法」を提案している。歌声情報処理研究の分野では音響物理的なアプローチのみの研究が多いが、本研究では心理学や感性工学の手法を導入し、学際的に人と歌声の関係を考察した。

「歌を歌う」という行為は、人間が行う音楽活動の中でも身近な表現方法である。小中学校では義務教育として音楽の授業を履修する必要もあり、これまでの生活で全く歌ったことがない、という人はほぼいないであろう。また、「歌を歌う」行為が我々の生活に密着している例として、日本発祥の文化であるカラオケが挙げられる。1990年代に通信カラオケが普及したことにより、誰でも「歌を歌う」行為を気兼ねなく楽しめるようになったのである。加えて、近年ではカラオケでのオンライン共有サービスや動画コミュニケーションサイトなどの存在により、誰でも簡単に自身の歌を Web 上に公開することができるようになった。つまり、インターネット環境さえあれば、誰でも自身の歌を世界中の人に聞いてもらう機会を得られるようになったのである。このように、「歌を歌う」という行為は誰もが関わることが可能であり、様々な楽しみ方が存在するため、「歌を歌う」行為を支援する研究成果は、多くの人にとって有益であるといえる。

「歌を歌う」という行為自体は容易に行えるが、「歌を歌う」練習を行うことは容易ではない。なぜなら「歌声の特徴に関する適切な情報を得ることが困難」であり、「特徴を詳細に観察することが困難」なためである。本研究では、これらの問題を解決することを目指し、(1) アマチュア歌唱者が理解しやすい情報を自動推定する、(2) 歌声の情報を可視化する、という二つの課題に取り組む。これらの課題に取り組むにあたり、「どのような情報を用いるか」が重要な点となる。歌声から認知される情報には「歌唱のうまさ」といった歌唱技術に関わる情報や、「歌声の美しさ」のような感性的な情報、「声の大きさ」「声の高さ」のような物理的に定義しやすい情報など、様々な種類がある。その中で、本研究では感性的な情報として「歌声の印象」を対象とし、音響特徴量を用いた情報の自動推定、および可視化に向けた考察を行う。

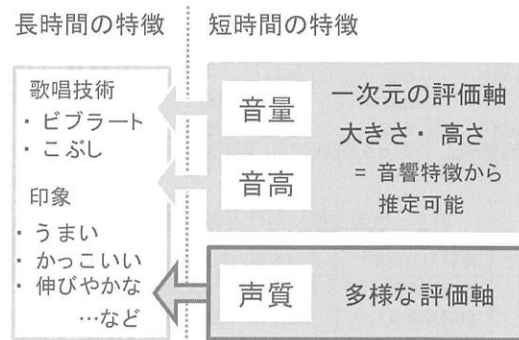


図1 歌声における特徴の区分

2. 研究計画

本研究における新規性は、次の二点である。まず、本研究では「アマチュア歌唱者が理解しやすい情報を自動推定する手法」を明らかにしている。従来の歌声評価に関わる研究では、ビブラートの有無や音高の正確性など、歌唱技術に着目した情報が多く扱われてきた。しかし、歌唱技術に関する情報を得られたとしても、誰もがその内容を適切に理解できるとは限らない。そこで本研究では、アマチュア歌唱者でも理解しやすい「印象」という情報に着目し、自動推定モデルを構築した。二点目として「声質と色の対応関係」を明らかにしている。歌声は「音量」「音高」「声質」という三つの要素で構成されているが、従来の可視化研究の多くは「音量」「音高」のみを扱っていた(図1)。表現する情報が一次元であり、時間軸と合わせて二次元での可視化が容易なためである。しかし「声質」には様々な評価軸が存在し、一次元の情報のみでは様々な差異を可視化できない。そこで本研究では声質の「印象」という評価軸について考察した上で、多次元の情報表現しやすい「色」との対応関係を明らかにした。

なお、歌声は時間軸を伴う表現であり、対象とする時間長によって得られる情報は異なる。つまり、観察したい特徴により、対象とする時間長を定める必要がある。本研究では「時間軸上の変化を捉えるための時間長(長時間の歌声)」「時間軸上の変化を表現するための時間長(短時間の歌声)」の2種類の時間長を対象とし、歌声の特徴の評価方法について考察を行った。

表1 各印象推定モデルの第8主成分までの偏回帰係数

印象推定モデル	各主成分に対応する偏回帰係数							
	1	2	3	4	5	6	7	8
迫力性	-1.15	-0.35	0.95				0.39	-0.45
丁寧さ				-0.65		-0.51		0.80
明るさ	0.69			0.67				0.64
勢いがある	-0.29		0.32					-0.10
声量のある	-0.33		0.24				0.12	
弱い	0.28		-0.18				-0.15	0.12
静かな	0.25	0.13	-0.21	-0.20				0.14
聴きやすい				-0.12				0.26
透き通った	0.13	0.16		-0.15				0.28
落ちつきのある			-0.19	-0.19		-0.14		0.16
響きのある	-0.15	0.15		-0.19	-0.12	-0.15		
嬉しそうな	0.10		0.10	0.14				0.10
軽やかな	0.15		0.11					0.14
可愛い	0.31			0.16				0.28
無邪気な	0.13			0.26				0.12

※ 空白箇所は、その主成分がモデルに採用されなかったことを示している

表2 歌唱支援に向けた各主成分の考察

主成分	歌い方の要素	得点が高い歌声の特徴
1	各音符の捉え方	音符ごとの F0 の変動が大きい
2	合唱音声らしさ	合唱らしい歌唱（声質、音量の安定度合い）
3	楽譜への忠実さ	しゃくりなど、楽譜から逸脱した歌唱表現
4	歌声らしさ	音の立ち上がり早く、耳に残る声
8	発音の丁寧さ	母音が聞き取りやすい発音、ビブラート

3. 長時間の歌声における特徴の評価方法

先行研究（筆者の修士論文）では、(1) 歌唱音声の評価尺度の構築、(2) 重回帰モデルを用いた歌唱音声の印象推定モデルの構築、という2つの過程を経た上で、歌声の音響特徴量から印象を自動推定するモデルを開発している。しかし推定精度は十分とは言えなかった。そこで、歌声から抽出する音響特徴量の種類を221種類に増やし、得られた音響特徴量を主成分分析した上で重回帰モデルの説明変数に用い、印象推定モデルを再構築した。その結果、印象の3因子のモデルの決定係数について、0.948（迫力性）、0.551（丁寧さ）、0.643（明るさ）という結果を得た。表1では、印象評価尺度12語と各因子の推定モデルに採用された主成分の偏回帰係数を示している。また、各主成分得点がどのような歌い方の要素に起因しているか考察した結果を表2に示している。例えば、第1主成分の得点が高い歌声では、一つ一つの音符で細かい基本周波数の変動が見られていた。逆に得点が高い歌声では、フレーズにまたがった大きな基本周波数の変動が見られ、その特徴が「迫力性」などの印象に影響していると考えられる。また、第2主成分の得点が高い歌声は合唱経験者の歌声が多く、合唱に関する技術の有無が「透き通った」「響きのある」といった印象に影響していると考えられる。このように、各主成分と歌い方の要素を対応づけることで、様々な印象の表現に必要な歌い方の要素を提示することが可能となる。歌声のデータ数を増やし、より詳細な考察を行うことで、歌唱練習の支援に役立てることが可能となるであろう。

表3 声質の3因子得点と色の特徴量の相関係数

	明度	彩度	A*	B*
評価性	0.26	0.06	-0.08	0.12
活動性	0.92**	0.75**	0.30	0.85**
迫力性	-0.11	0.60**	0.40*	0.43*

4. 短時間の歌声における特徴の評価方法

声質の印象について考察するため、声質を評価する13対の表現語を用い、0.6秒の歌声（28データ）に対する印象評定実験を行った。その結果、「評価性」「活動性」「迫力性」という3つの因子が抽出された。また、色との対応関係について調査するため、明度・彩度・色相それぞれの条件を統制した26色を用い、歌声に適した色刺激を選択させる一対比較評価実験を行った。各歌声における色の特徴量及び因子得点を用いて相関分析を行った結果を表3に示している。例えば、活動性因子が高い歌声には、明度・彩度が高く、黄みを帯びた色が適しているといえる。評価性因子は表3中では相関が見られていないが、色相条件における緑色の刺激の評価値と相関が見られていた。つまり、いずれの因子の情報も、色と関連付けて可視化することが可能である。ただし、色の特徴と因子がそれぞれ独立に対応してはいなかったため、色のみを用いて3因子の情報を同時に表現することはできない。歌唱練習の支援を行う際には、多くの色の特徴と関係していた「活動性因子」を色で表現し、他の因子は座標などを用いて表現することで、時間軸にそった歌声の特徴の可視化が可能となる。

5. まとめ

本研究では、「長時間の歌声における特徴の自動推定手法」を明らかにし、「短時間の歌声における特徴の評価方法、及び可視化方法」について検討を行った。

長時間の歌声を対象とした場合には、歌声全体の印象を推定し、特徴の把握につなげることが可能である。本研究では、歌唱練習の支援を目指し、歌声の特徴を把握する方法を提案したが、次の段階として「どうすれば自身の歌声を望んだ印象に近づけられるかを示す」ことが必要となる。そのため、主成分分析の結果をより詳細に考察し、歌唱者が意識できる歌い方の要素と主成分の特徴を明確に対応づける必要がある。

短時間の歌声を対象とした場合には、声質の印象の3因子を色と対応付け、時間軸にそった特徴の変化を可視化することが可能である。ただし、3因子の情報を色のみで同時に表現することはできないため、色以外の視覚情報（図形やテキストチャなど）と組み合わせることで、時間軸にそった声質の差異をより分かりやすく可視化できるであろう。