

早稲田大学審査学位論文

博士（人間科学）

概要書

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 Singers

2019年1月

早稲田大学大学院 人間科学研究科

金礪 愛

Kanato Ai

研究指導教員： 菊池 英明 教授

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。対象とした歌唱音声（以降、歌声）は、日本語歌詞のポピュラー音楽における女子大学生の歌声である。

第1章では、本研究の背景と研究目的について述べ、論文全体の構成を示した。「歌を歌う」という行為は、人間が行う音楽活動の中で最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、「誰もが関わるのが可能」で「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。このように「歌を歌う」行為は「誰もが関わるのが可能」であるが、「歌を歌う」練習を気軽に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。このような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。

第2章では、本研究の新規性を述べた上で、研究の概要を示した。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。表現する情報が一次元であるため、時間軸と合わせて二次元での可視化が容易なのである。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、多次元の情報を表現しやすい「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。本研究の第3章では「時間軸上の変化を捉えるための時間長（長時間の歌声）」を、第4章では「時間軸上の変化を表現するための時間長（短時間の歌声）」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。

第3章では、長時間の歌声における特徴の評価方法について検討した。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システムを開発している。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、印象推定モデルの精度向上を図った。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。

第4章では、短時間の歌声における特徴の評価方法について検討した。まず、声質の特徴について考察するため、声質を評価する表現語を用い、印象評定実験を行った。その結果、「評価性」「迫力性」「活動性」という3つの因子が抽出された。また、声質の情報を可視化することを目指し、色との対応関係について調査するため、一対比較評価実験を行った。その結果、「活動性因子」に関連する多くの印象が、明度や色相（黄みー青み）など、多くの色の特徴と関係していることが明らかとなった。また、3因子はそれぞれ明度、彩度、色相のいずれかと相関があった。つまり、いずれの因子の情報も、色を用いて可視化することは可能である。ただし、一対一対応ではなかったため、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。

第5章では、本研究の成果をまとめた上で今後の展望を示した。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。そのためには、歌声の音響特徴量を主成分分析した結果について、より詳細な考察が必要となる。例えば、本研究の主成分分析結果を歌唱者の視点から捉えると、主成分ごとに「各音符の捉え方」「楽譜への忠実さ」「発音の丁寧さ」といった歌い方の要素に対応づけられた。これらの要素を印象推定モデルに照らし合わせると、各印象の表現に必要な歌い方の要素が提示可能となる。データ数を増やし、より詳細な考察を行うことで、歌唱練習の支援により役立てることが可能となるであろう。また、短時間の歌声を対象とした場合には、声質の印象の3因子を色で可視化し、時間軸にそった特徴の変化を把握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、明度、彩度、色相の全てと相関が見られた「活動性」の因子のみを色で表現し、他の因子は色とは異なる要素で可視化することが望ましい。例えば「図形」や「テクスチャー」を構成する要素がどのような心理的要素と対応しているか、といった先行研究の知見を活かし、色と組み合わせることで、時間軸にそった声質の差異を可視化できるであろう。