早稲田大学審査学位論文
博士（人間科学）

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

目 次

表 目 次

図 目 次

第1章　はじめに

1.1　背景

本研究では，人間の歌唱音声 (以降，歌声) を研究対象としている．

「歌を歌う」という行為は，人間の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,「歌を歌う」という行為を避けては通れない．そして，授業外でも，運動会での応援歌や合唱コンクール，卒業式での校歌斉唱など，様々な場面で「歌を歌う」ことを経験する．したがって，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．2015 年には，国内二大カラオケ企業の一つ，第一興商が東証一部上場を果たしたことからも，カラオケという文化への関心の大きさがうかがえる．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「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」行為は「誰もが関わることが可能」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．練習を行うためには，自身の振る舞いを逐一確認し，その振る舞いが望ましい結果かどうかを知る必要がある．その上で，別の方策をとり，より望ましい結果を得られるよう繰り替えす過程が，一般的な練習の流れである．

しかし,「歌を歌う」練習においては,「自身の振る舞いを逐一確認」することが難しい．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．

まず，歌を歌う練習をする際,「歌声の特徴に関する適切な情報を得ることが困難」である．ここでいう「適切な情報」とは，自身が必要としているフィードバックとして適した情報，という意味である．例えば，一人で歌を練習する際には，自身の歌声の良し悪しを自身が評価する必要がある．しかし，自身の歌声を評価する際には，どうしてもバイアスがかかってしまう．他者に付き添ってもらい，他者に歌声を評価してもらうこともできるが，評価の基準は個人に依存してしまう．

次に，歌を歌う練習をする際,「特徴を詳細に観察することが困難」である.「歌を歌う」という行為においては，自身の振る舞いが音に現れるという特性上，形として残すことが難しい．録音することは可能であるが，歌い終わった後に，もう一度同じ時間をかけて聴

く必要があり，歌唱していた際の自身の振る舞いと対応づけることが難しい．加えて，いくつかの試行を同時に比較することも困難である．

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．

上記のアプローチを行うにあたり,「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，情報の自動推定，および可視化に向けた考察を行う．

また，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する．

なお，本研究では，研究対象を以下のように定める．

・歌唱者：アマチュア歌唱者
・歌唱楽曲：日本語歌詞のポピュラー音楽
・伴奏：なし

1.2　目的

本研究は「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」を提案することを目指す.「歌声の特徴を把握する」とは，なんらかの歌唱表現が異なる複数の歌声において,「どこが」「どのように」異なっているかを理解できることを指す.

本研究の目的を達成するために，以下の２項目を小目標として設定している．

・長時間の歌声における特徴の評価方法の検討（第３章）
・短時間の歌声における特徴の評価方法の検討（第４章）

1.3　応用

本研究では,「歌声の特徴を，アマチュア歌唱者が理解しやすい方法で可視化する」という，従来は行われていなかったアプローチを検討している．このアプローチでは，印象評定実験や因子分析などを行うことにより，心理学的視点，感性工学的視点及び音声学的視点から，歌声と人との関係について考察している．このような学際的なアプローチを行うことにより，実際の場面に即した研究結果を得られると考えられる．

本研究の有用性が認められれば，歌声に対する以下の２つの視点において，様々な場面への応用が可能となる．

1.3.1 「歌を歌う」という視点における応用

冒頭でも述べた通り,「歌を歌う」練習を行うことは，容易ではない．しかし，自身の歌声を可視化することで，自身の歌声の特徴を知り，様々な歌い方を試し，自身が望んだ歌い方に近づける，という練習が可能となる．また，他者の歌声を可視化し，自身の歌声の可視化結果と比較することで,「どこが」「どのように異なっている」か，把握することが容易になる．その結果，他者の歌声に似せる練習にも活かすことができると考えられる．

1.3.2 「歌を聴く」という視点における応用

インターネット環境が発達したことで，歌声を多くの人に聞いてもらう機会が増えた．つまり，聴取する側も，より多くの歌声を聴く機会を得られるようになったといえる．ただし，歌声は，音メディアという特性上，耳で聞かなければ情報を得ることができない．そのため，好みの歌声を探す際には，膨大な量の歌声を聴く時間が必要となる．しかし，歌声の特徴を可視化できれば，目で見ただけで自分の好みの歌声の特徴を探し出すことが可能となる．また，複数の歌声の特徴を比較することもでき，聴取する際の新たな楽しみ方を提供することも可能だと考えられる．

1.4 構成

本論文は，全 5 章から構成される．以下に，本論文の構成を示す．
第 2 章では，本研究の研究計画について述べる．
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述べる．
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述べる．
第 5 章では，研究全体を通しての結論を述べる．

第2章　研究計画

本章では，本研究の新規性，および，研究内で用いる用語，本研究の概要について述べる．

2.1　本研究の新規性

本研究では,「アマチュア歌唱者が自身の歌声の特徴を把握するために有用な可視化方法」を明らかにすることを目指し，段階的な調査を行う．その中でも，次の 2 点において，本研究は新規性があると言える．

2.1.1　アマチュア歌唱者が理解しやすい情報を自動推定する

従来の歌声の評価に関わる研究では，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が得られたとしても，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，情報を自動推定する．

2.1.2　声質と色の対応関係を明らかにする

音のような，形に残すことができない媒体を観察するために，可視化という手段が用いられる．歌声は時間軸を伴う表現であり，時刻ごとにどのように特徴が変化しているかを観察することが，特に重要となる．音は「音量」「音高」「音色」という 3 つの要素で構成されていることが知られているが，従来の可視化研究の多くは「音量」「音高」のみを扱っていた．本研究では「音色」，つまり歌声においては「声質」に該当する要素を可視化するための基礎的検討を行う．

2.2　本研究で扱う「歌声の特徴」

本研究における「歌声の特徴」とは，同一楽曲を歌唱した複数の歌声があった際に，それらの歌声の差異を認識するための要素をさしている．

本研究は,「歌声の特徴」に着目し,「アマチュア歌唱者が自身の歌声の特徴を把握するために有用な可視化方法」を明らかにすることを目指す．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，認知できる特徴は異なる．例えば，歌唱力の評価に有用な歌唱技術であるビブラートは，時間軸に沿った音高変化によって認知される．つまり，ある

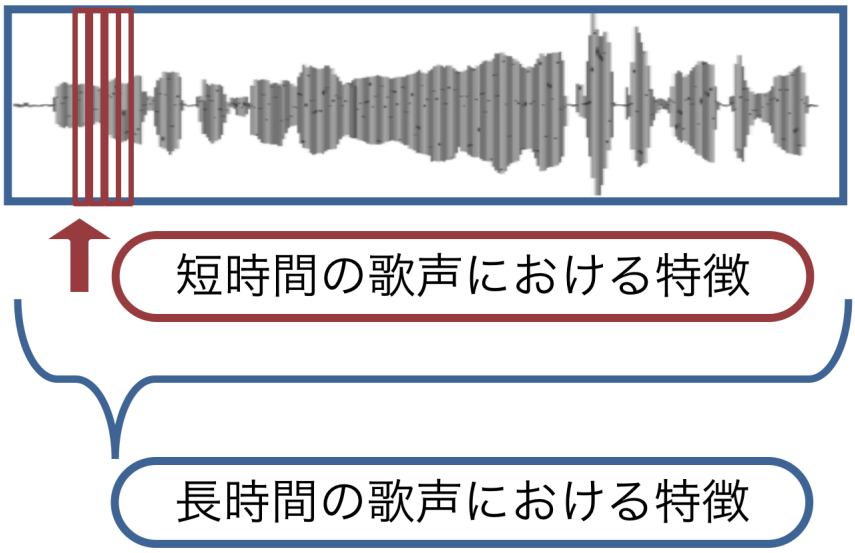

図 2.1: 本研究で扱う歌声の時間長

程度の時間長がある歌声でないと，認知することができない．このように，観察したい特徴によって，対象とする時間長を定めなければならない．

まず，上記の例のように,「音高変化」といった時間軸上の変化を捉えられるような時間長を対象とする必要がある．実際,「あの人は歌がうまい」「あの人の歌声はかっこいい」のような歌声の総評を述べる際には，ある程度の長さの歌声を聴く必要がある．

ただし，歌声の特徴を捉えるためには，時間ごとに変化する情報そのものも把握しなければならない．従来の歌声可視化研究では，音の 3 要素のうち「音量」「音高」を対象に，時間軸上の変化を可視化する研究が多い．これら 2 つの要素は，それぞれ「音の大きさ」「音の高さ」という一つの尺度に対応づけられるためだと考えられる．本研究では，歌声の「声質」を対象に,「印象」という側面から時間軸上の変化を可視化することを目指す．

本研究では，上で述べた「時間軸上の変化を捉えるための時間長」と「時間軸上の変化を表現するための時間長」，2 種類の時間長に分け（図 2.1），歌声の特徴を把握する手法を検討する．

2.3 本研究の概要

本研究は，以下の 2 つのブロックから構成されている．

1. 長時間の歌声における特徴の把握（第 3 章）

「時間軸上の変化を捉えるための時間長」を分析対象とし，ある程度の時間長から認知される歌声の特徴について考察する．金礪の修士論文 [1] では，10 秒程度の歌声を対象に，印象を自動推定する手法が明らかにされている．より高水準な推定を行うため，特徴量の

再検討及びモデルの再構築を行った．

2. 短時間の歌声における特徴の把握（第４章）

「時間軸上の変化を表現するための時間長」を分析対象とし，ごく短い時間に見られる歌声の特徴について考察する．音の３要素のうち，十分に研究されていない「声質」について，どのような評価軸を用いるべきか，考察を行った．また，時間軸上の変化を把握するためには，特徴の可視化が不可欠である．そこで，どのように可視化すべきか，声質と色の対応関係について調査を行った．

次章より，詳細を述べる．

第3章　長時間の歌声における特徴の評価方法

この章では，長時間における歌声の特徴を把握する方法について検討する．

3.1　本章の目的と背景

歌声の特徴を把握するためには，長時間における歌声の特徴と，短時間における歌声の特徴，双方を扱う必要がある．本章では「時間軸上の変化を捉えるための時間長」を分析対象とし，ある程度の時間長から認知される歌声の特徴を把握する方法について，検討する．

3.2　長時間の歌声における評価に関する先行研究

歌声の評価に関する研究は多く行われてきた．そのほとんどが，本研究で対象としている「ある程度の時間長から認知される歌声の評価」に関わる研究である．

従来，歌声の評価においては，特定の印象の強度を推定する研究が多く行われている．例えば，中野らは，歌唱された楽曲の楽譜情報を用いずに，歌声の歌唱力を自動推定する手法を明らかにしている [2]．また，Tsi and Lee は，原曲の歌声と歌唱者の歌声の類似性に基づいた歌唱力評価を行っている [3]．歌唱力以外の印象の推定においては，Daido が歌声の熱唱度の自動推定手法を提案している [4]．

また，歌声の印象と音響特徴量の関係性を考察する研究も行われている．例えば，Kotlyar and Morozov は，11 人のプロの歌唱者が歌唱した歌声を用い，歌声の感情表現と音響特徴量との関係を調査している [5]．

上記で述べた研究は,「歌唱力」「感情」といった特定の印象を対象としており，歌声の特徴の一部を評価している，と言える．一方，金礪の修士論文では，歌声が与える印象を包括的に扱い，どのような印象か自動推定するシステムの開発を行っている [1]．しかし，推定精度は十分とは言えないため，より詳細な検討が必要と言える．

本研究を進めるにあたり，金礪の修士論文は大きな基盤となっているため，次節で金礪の修士論文について詳細を述べる．

3.3　印象評価に関わる先行研究

修士論文では，アマチュア女性歌唱者を対象に，歌声の音響特徴量から印象を自動推定するシステムの開発を行った．この研究では，44 語の印象評価語に対応する重回帰モデルを作成しており，歌声を入力すると，各評価語の得点を算出できる．つまり，得点が高かった評価語は，その歌声の印象を示す語と言える．また，歌声の印象空間に該当する 3 因子の得点も算出するため，印象空間内における歌声の位置も把握することができる．このように，印象という情報を扱うと，アマチュア歌唱者が自身の歌声の特徴を把握しやすくなると考えられる．

そこで，この節では，印象評価に関する先行研究 [1]（以降，修士論文）について概要を述べる．修士論文では，印象推定システムを開発するにあたり,「歌声の印象評価尺度の構築」「歌声の印象・因子の推定モデルの構築」という 2 つの段階を経ている．以下にその概要を述べる．

3.3.1　歌声の印象評価尺度の構築

歌声の印象評価に関わる因子，また，それらの印象を表現する言葉を明らかにするため，主観評価実験と因子分析により歌声の印象評価尺度を構築した．以下に,「仮尺度の構築」「歌声収録」「印象評定実験」「因子分析」の 4 つの行程の概要を述べる．

1. 仮尺度の構築

   歌声の多様な印象を適切に形容できる語を選定し，仮尺度の構築を行う．まず，歌声を形容している多様な語を収集した．収集対象は，A. 学術的に重要な語（先行研究からの収集），B. 専門的に使用される語（CD レビューからの収集），C. 日常的に使用される語（動画共有サイト，SNS からの収集）である．収集した後の数は，表 3.1 に示している．合計 898 語の評価語を収集した上で，了解性調査（歌声の評価に適した語かどうかを調査），同義性調査（類似した評価語を除外するための調査）を行い，44 語の評価語を選定した（表 3.2）．

2. 歌声収録

   印象評定実験に向けて,「歌詞・メロディ・テンポ・キーが統一されている」「評価者にとって未知のメロディ・歌詞である」「認知できる印象が多様である」という条件

表 3.1: 収集した語の数

収集元	述べ数	異なり数
A. 先行研究 [6–9]	180	162
B. CD レビュー	699	372
C. SNS サービス	10000	294
C. 動画共有サイト	1026	232
合計	11905	898

表 3.2: 印象推定に用いた歌声の印象評価語（44 語）

甘い	心のこもった	ドスが効いている
安定している	こもっている	伸びやかな
勢いがある	爽やかな	激しい
一生懸命な	静かな	ハスキーな
色気のある	声量のある	鼻にかけたような
美しい	シャープな	響きのある
嬉しそうな	少女のような	不安定な
落ちつきのある	少年のような	ぶりっこみたいな
かっこいい	女性的な	震えている
悲しい	芯のある	真っすぐな
軽やかな	透き通った	無邪気な
可愛い	繊細な	優しい
聴きやすい	男性的な	陽気な
気持ち良さそうな	中性的な	弱い
元気な	特徴的な	

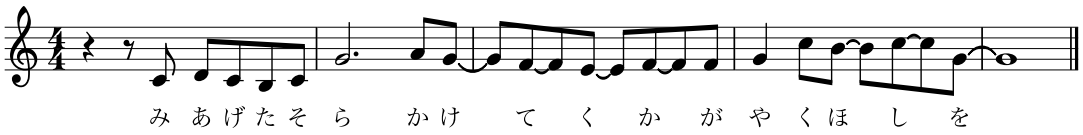

図 3.1: 実験に用いたオリジナルメロディ

を満たした歌声の収録を行った．歌唱者は 21 名の女子大学生であり,「一番うまく聴こえるように」「表現豊かに」「できるだけ平らに」など，7 種類の歌唱条件を提示している．収録に用いたオリジナルメロディは，図 3.1 に示している．計 147(=21*7) 歌唱を収録した上で，聴取印象に大きな差が見られないデータを除外し，最終的に 60 データを印象評定実験の刺激として選定した．選定された 60 データは，21 名の歌唱者全員の歌声を 2–5 データずつ含んでいる．

3. 印象評定実験

60 データの歌声を対象とし，44 語の仮尺度，及び歌声評価に重要だと考えられる 3 語（うまい，好きな，曲に合っている）を用い，印象評定実験を行った．歌声を評価者に提示する際，収録の際に用いた伴奏音は除外している．評価者は 20 代の一般大学生 19 名 (男性 9 名，女性 10 名) である．Web 上のアンケートページを用い，各評価語がどの程度あてはまるか，7 段階での評価を求めた．

印象評定の結果を用い，各評価語における「評価者間の相関」及び「評価語間の相関」を求めた．その上で,「評価者間の相関」が高い語を抽出し,「評価語間の相関」が高い語は統合・除外を行った．その結果として得られた 36 語を，次の因子分析に用いた．

表 3.3: 完成した尺度の評価語と因子負荷量

	第 1 因子（迫力性）	第 2 因子（丁寧さ）	第 3 因子（明るさ）
勢いがある	**0.932**	0.044	0.024
声量のある	**0.917**	0.188	-0.192
弱い	**-0.898**	0.023	-0.008
静かな	**-0.752**	0.466	-0.166
聴きやすい	0.146	**1.001**	0.271
透き通った	-0.127	**0.886**	0.236
落ちつきのある	-0.286	**0.775**	-0.232
響きのある	0.387	**0.756**	-0.161
嬉しそうな	0.246	0.092	**0.923**
軽やかな	-0.037	0.358	**0.854**
可愛い	-0.286	0.145	**0.830**
無邪気な	-0.085	-0.359	**0.777**
寄与率	0.292	0.292	0.262
信頼性係数 α	0.926	0.893	0.877

表 3.4: 3 因子の因子間相関

	第 1 因子（迫力性）	第 2 因子（丁寧さ）	第 3 因子（明るさ）
第 2 因子（丁寧さ）	0.189	1.000	
第 3 因子（明るさ）	0.229	-0.132	1.000

4. 因子分析

印象評定実験の結果を評価者ごとに標準化し，歌声データごとに各語の平均値を算出した．36 語の印象評価得点を用い，因子分析を行った．因子数はスクリー基準に基づいて決定し，分析には最尤法，プロマックス回転を用いた．その結果，因子負荷量がどの因子においても 0.35 以下である評価語，また，独自性の値が極端に高い評価語を，尺度に不適切とみなし除外した．さらに，各因子の内的一貫性の高さの指標となる Cron-bach の α 係数 [10] を求め，全ての因子において $\alpha > 0.85$ となるまで，因子分析と評価語の除外を繰り返した．

印象評価尺度を構築した結果，12 語が尺度として適切であると判断された (表 3.3)．抽出された 3 因子に対し，各因子の因子負荷量が高い評価語を参考に，それぞれ「迫力性」「丁寧さ」「明るさ」と命名した．また，これらの因子は因子間相関の値がそれぞれ低いことから，3 因子はある程度独立して歌声の印象評価に寄与していると言える．

3.3.2 歌声の印象・因子の推定モデルの構築

歌声の印象を音響特徴量から推定するためのモデル構築を行った．ここでは，印象の強度を連続的な値で推定可能である重回帰モデルを用いる．まず，歌声から音響特徴量 108 種類を算出した．重回帰分析における多重共線性を避けるため，特徴量同士の相関が高かった特徴量は除外し，残りの 88 種類の特徴量を用い，重回帰分析を行った．モデルによって得られた自由度調整済み決定係数と，交差検定の結果を表 3.5 に示す．

印象評価における 3 因子の決定係数において，迫力性では 0.880, 丁寧さでは 0.481, 明るさでは 0.676，3 因子の平均は 0.679 という結果を得た．

表 3.5: 各評価語における重回帰分析結果

44 語の印象評価語と R^2（1 に近い程モデルの精度が高い）

印象評価語	R^2	交差検定	印象評価語	R^2	交差検定
声量のある	0.883	0.883	女性的な	0.520	0.474
激しい	0.858	0.833	シャープな	0.566	0.464
弱い	0.795	0.745	色気のある	0.606	0.462
勢いがある	0.757	0.731	気持ち良さそうな	0.626	0.456
優しい	0.786	0.712	爽やかな	0.637	0.422
繊細な	0.726	0.708	透き通った	0.549	0.410
少女のような	0.776	0.708	美しい	0.556	0.410
一生懸命な	0.812	0.691	無邪気な	0.675	0.408
静かな	0.784	0.687	軽やかな	0.496	0.363
かっこいい	0.728	0.679	陽気な	0.695	0.362
響きのある	0.706	0.668	ぶりっこみたいな	0.549	0.352
ドスが効いている	0.786	0.660	震えている	0.505	0.351
元気な	0.723	0.640	中性的な	0.510	0.334
男性的な	0.768	0.639	特徴的な	0.570	0.292
可愛い	0.739	0.633	落ちつきのある	0.442	0.270
芯のある	0.710	0.580	不安定な	0.360	0.230
少年のような	0.660	0.576	安定している	0.433	0.221
伸びやかな	0.595	0.551	聴きやすい	0.335	0.207
甘い	0.680	0.539	真っすぐな	0.367	0.001
心のこもった	0.677	0.512	こもっている	0.292	-0.026
ハスキーな	0.629	0.508	嬉しそうな	0.359	-0.332
悲しい	0.626	0.475	鼻にかけたような	0.170	-1.488

歌声の印象評価における 3 因子

印象評価語	R^2	交差検定
迫力性	0.880	0.849
丁寧さ	0.481	0.385
明るさ	0.676	0.562

歌声の評価に重要である評価語

印象評価語	R^2	交差検定
好きな	0.401	0.299
うまい	0.333	0.256
曲に合ってる	0.346	0.089

仮尺度で用いた 44 語の平均

	R^2	交差検定
44 語の平均	0.614	0.432

歌声評価尺度に含まれる 12 語の平均

	R^2	交差検定
12 語の平均	0.627	0.473

3.4　印象推定モデル再構築

修士論文では，44 語の評価語全体の決定係数の平均が 0.614 であり，概ね印象を推定できている，と言える．しかし，印象の種類によって推定精度に差があった．迫力性因子が大きく関わっている「声量のある」「激しい」「弱い」「勢いがある」などはそれぞれ決定係数が 0.75 を上回っており，推定精度は比較的高い．一方，丁寧さ因子が大きく関わっている「聴きやすい」「落ち着きのある」といった評価語では決定係数が 0.5 を下回っており，丁寧さ因子自体も推定精度は 0.481 に留まっている．

歌声の印象を表現するために，3 因子の得点は非常に重要であり，そのうちの 1 因子の推定精度が低いという点は望ましくない．そこで，本研究では推定精度を向上させるため，「音響特徴量の追加」,「音響特徴量の主成分分析」という過程を経た上で，再度重回帰分析により「モデル構築」を行った．以下に詳細を述べる．

3.4.1　推定精度向上のためのアプローチ

重回帰分析では，説明変数として用いる変数が多ければ多いほど，多重共線性や抑制変数ににより，モデルが不安定になる危険性が高くなる．そのため，修士論文では 108 種類の音響特徴量を算出した上で，多重共線性の危険性を下げるため，特徴量同士の相関を求め，相関が高かった特徴量の片方を除外する，という行程を経ていた．しかし，この手法だと表面上相関が高い特徴量を除くことはできても，モデルの不安定性を完全に解決することはできない．そこで，変数同士の相関を減らすため，音響特徴量を主成分分析し，得られた主成分得点を重回帰モデルの説明変数として用いた．また，それに伴い，扱う音響特徴量も増やしている．

3.4.2　モデルの再構築：音響特徴量の分析

修士論文では，全 108 種類の特徴量を用いていた．本研究では，全 221 種類の音響特徴量を用い，モデルの構築を行う．

以下に，用いた音響特徴量の分析について，詳細を述べる．なお，本節は筆者が第一著者である「歌声の印象評価尺度の構築に基づく多様な印象の自動推定手法」[11] に基づいている．

音響特徴量の抽出

印象評定実験で用いた歌声データ 60 歌唱から，音響特徴量の抽出を行う．多様な楽曲に適用することを想定し，調査対象とする音響特徴量は，楽譜情報や歌詞の情報を用いずに抽出できる特徴とした．

分析に用いた歌声データは 44.1 kHz，16 bit サンプリングのモノラル信号である．まず，STRAIGHT [12] を用いて 1 ms ごとに F_0（基本周波数），スペクトル包絡，非周期性指標を推定する．分析フレームは 1 ms ごととし，それらを用いて計 221 種類の音響特徴量の抽出を行った（表 3.6）．この節では，抽出した各特徴量の詳細について述べる．

表 3.6: 抽出した音響特徴量一覧

静的特徴量における統計特徴量			
対象とするスペクトル包絡		S_{lin}	$S_{\log}$
スペクトル重心		○	○
スペクトル傾斜	0-22.05 kHz	○	○
	0-3 kHz	○	○
	0-6 kHz	○	○
	0-9 kHz	○	○
倍音構成	H1/H2	○	○
	奇数・偶数倍音の比	○	○
歌唱フォルマントらしさ		○	○
スペクトルフラックス		○	○
フォルマント	F_1	○	○
	F_2	○	
非周期性指標の総和		○	
非周期性指標の傾斜	0-22.05kHz	○	
	0-3 kHz	○	
	0-6 kHz	○	
	0-9 kHz	○	

動的変動量における統計特徴量					
フレーム幅 $K(ms)$		10	25	50	100
フォルマント	F_1	○	○	○	
	F_2	○	○	○	
スペクトル	0-3 kHz	○	○	○	
	0-22.05 kHz	○	○	○	
F_0	$\Delta f0(t)$	○	○	○	○
	$\Delta\Delta f0(t)$	○	○	○	○
パワー		○	○	○	○

F_0 に関する特徴量
相対音高のピークの鋭さ，ピークの傾斜
フレーズ全体における cent の傾き（1 ms, 1000 ms）
フレーズ全体における cent の標準偏差（1 ms, 1000 ms）
ビブラートの速さに該当するパワーの最大値，平均，標準偏差
ビブラートらしさの最大値，平均，標準偏差
ビブラートと認定された区間における，上記の特徴量
ビブラートの速さ，深さの最大値，平均，標準偏差
有声区間中のビブラートと認定された区間の割合
F_0 の安定度（K=10, 25, 50, 100）

抽出した音響特徴量は，抽出方法により次の 3 種に大別できる．なお，本研究では，1 歌唱毎に，その有声区間における平均値，標準偏差，中央値，四分偏差を求め，これを統計特徴量と呼ぶ．

（1）静的な特徴量 1 フレームごとに抽出した特徴量を用い，統計特徴量を抽出．

（2）動的な特徴量 複数のフレームにおける変動量を求め，統計特徴量を抽出（3 もしくは 4 種類のフレーム数を対象として，それぞれで変動量を計算）．

（3）F_0 に関する特徴量 ビブラートなど，基本周波数（F_0）に関わる特徴量を抽出．

抽出した特徴量については，表 3.6 にまとめて示した．

本研究では，動的特徴量などの算出において回帰係数を用いるが，全て以下の式に基づく．ここで y は分析対象とする特徴ベクトルであり，$2K+1$ はベクトルの長さを表している．たとえば，y にはスペクトル包絡や F_0 軌跡などが相当する．

$$R(y) = \frac{\sum_{k=-K}^{K} k \cdot y_k}{\sum_{k=-K}^{K} k^2} \tag{3.1}$$

スペクトル包絡に関する音響特徴量

スペクトル包絡は，歌声の声質を特徴づける重要な特徴量の一つであり，先行研究においても様々な検討がなされている（ [13] など）．本調査では，各時刻 t におけるスペクトル包絡 $S_{\mathrm{lin}}(f,t)$ および対数スペクトル包絡 $S_{\log}(f,t) = \log|S(f,t)|$ 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．

スペクトル重心 スペクトル重心は，*Timbral Texture Feature* として知られている [14]．スペクトル包絡 $S_{\mathrm{lin}}(f,t)$, 対数スペクトル包絡 $S_{\log}(f,t)$ から，各時刻におけるスペクトル包絡の重心 $S_c(t)$ を，以下の式を用いて求め，統計特徴量を算出する．B は，周波数ビンの数を示している．

$$S_c(t) = \frac{\sum_{f=1}^{B}(f \cdot S_{\mathrm{lin|log}}(f,t))}{\sum_{f=1}^{B}(S_{\mathrm{lin|log}}(f,t))} \tag{3.2}$$

スペクトルフラックス スペクトルフラックスも *Timbral Texture Feature* として知られており，局所的なスペクトル変化の指標とされている [14]．時刻 t のフレームにより標準化されたスペクトル包絡 $S_{\mathrm{lin}}(f,t-1)$，対数スペクトル包絡 $S_{\log}(f,t-1)$ を用い，以下の式によりスペクトルフラックス $S_f(f,t)$ を求め，統計特徴量を算出する．

$$S_f(t) = \sum_{f=1}^{B}(S_{\mathrm{lin|log}}(f,t) - S_{\mathrm{lin|log}}(f,t-1))^2 \tag{3.3}$$

スペクトル傾斜 式 (3.1) を用いてスペクトル包絡 $S_{\mathrm{lin}}(f,t)$, 対数スペクトル包絡 $S_{\log}(f,t)$ から，時刻毎の傾きを求める．4 種類の帯域（0-3 kHz, 0-6 kHz, 0-9 kHz, 0-22.05 kHz）におけるスペクトル傾斜を求め，統計特徴量を算出する．

Singer's Formant 歌声らしさや声の響きを評価する特徴量として Singer's Formant が知られている [13,15,16]．本研究では，スペクトル包絡，対数スペクトル包絡の 2-4 kHz の帯域におけるパワーの全帯域に対する割合を歌唱フォルマントらしさの特徴量として求め，統計特徴量を抽出する．

スペクトルの倍音構造 基本波の強さ（F_0 に該当する周波数におけるパワー）は気息性の指標として知られているため，統計特徴量を算出する．また，倍音のパワー比は，歌声の声区の判別に有効であると報告されている [17, 18]．本研究では，基本波のパワー H1 と第二倍音に該当するパワー H2 の比（H1/H2），及び奇数倍音と偶数倍音に該当するパワーの総和の比を，スペクトル包絡から求め，統計特徴量を抽出する．

音韻性の知覚に関する音響特徴量

スペクトル包絡にはフォルマントに関する情報も含まれており，音韻の知覚や歌声の印象にも影響を及ぼすと考えられるため，関係する特徴量を抽出する．

フォルマントに関わる特徴量 フォルマントに関係する特徴量として，スペクトル包絡のピーク周波数を求める．まず，各時刻 (t) のスペクトル包絡のケプストラムの低次成分に対して逆フーリエ変換を行い，文献 [19] を参考に，フォルマント周波数である可能性が高いと考えられる帯域（$F_1 < 900$Hz, $900\text{Hz} < F_2 < 3300$Hz）に制限した上でピークの検出を行い，第 1 ピーク $F_1(t)$，第 2 ピーク $F_2(t)$ を求めた．$F_1(t)$, $F_2(t)$ の値を用い，統計特徴量を抽出する．

非周期性成分

STRAIGHT [12] では，スペクトル包絡の全体のエネルギーに対する非周期成分の割合を，0 から 1.0 の値で求めることができる．値が 1 に近づく程，非周期成分の割合が多いことを示しており，歌声に含まれている非周期成分の大きさを評価することができる．

非周期性成分 スペクトル包絡全帯域における非周期性成分の値の総和を求め，統計特徴量を抽出する．

非周期性成分の傾斜 非周期性成分を式 (3.1) の $y(k)$ に代入し傾きを求める．4 種類の帯域における傾きを用い，統計特徴量を抽出する．

動的な特徴量

ここまでで扱った特徴量は，歌声の「声質」に関係する静的な特徴量である．歌声の印象の評価には，スペクトル包絡やフォルマントに関わる特徴量の動的な変動も関与していると考えられるため，以下の特徴量の算出を行う．それぞれ，分析フレーム幅を 1 フレームずつシフトさせながら回帰係数を求めるが，ある時刻の前後 K フレーム内に無声区間が含まれていた場合，その時刻は分析対象外とする．

パワーの動的変動量 以下の式により，各時刻におけるパワー $P(t)$ を求め，式（3.1）を用い，回帰係数を求める．4 種類のフレーム幅（K=10, 25, 50, 100）を用い，有声区間の統計特徴量を抽出する．

$$P(t) = \sum_{f=1}^{B} S_{\mathrm{lin}}(f,t) \tag{3.4}$$

スペクトル包絡の形状の動的変動量 スペクトル包絡 S_{lin} 及び対数スペクトル包絡 S_{log} の各周波数ビンにおける回帰係数 $\Delta S_{\mathrm{lin}}(f,t)$ 及び $\Delta S_{\mathrm{log}}(f,t)$ を式 (3.1) を用いて求め，時刻 t における全周波数ビンの回帰係数の絶対値の総和を算出する．4 種類のフレーム幅（K=10, 25, 50, 100）を用い，有声区間の統計特徴量を抽出する．

フォルマントに関わる動的特徴量 $F_1(t)$ 及び $F_2(t)$ を用い，式（3.1）により回帰係数を求める．3 種類のフレーム幅（K=10, 25, 50）における，統計特徴量を抽出する．

F_0 に関する特徴量

本研究で扱う周波数は対数スケールで示し，cent 単位で表す．西洋平均律では，半音が 100 cent にあたる．中央ハ音の周波数 $f_c(= 440 \times 2^{\frac{3}{12}-1} = 261.62...\mathrm{Hz})$ の cent 値を 4800 cent とすると，周波数 f_{Hz} の音の cent 値 f_{cent} は以下の式で表される．

$$f_{\mathrm{cent}} = 1200 \log_2(\frac{f_{\mathrm{Hz}}}{f_c}) + 4800 \tag{3.5}$$

今後，本研究では基本周波数を $F_0(t)$ で表す．ここで，t は時間軸を示している．

相対音高 本研究では，楽譜情報を用いない特徴量を扱うため，歌声の相対音高に関する二種類の特徴量 [2] を算出する．この特徴量は，音高が半音（100 cent）単位で遷移しているかどうかを評価する指標である．具体的には，文献 [2] における相対音高の正確さ（$g(F)$）のピークの鋭さ，及びピークの傾斜を直線近似した傾き [2] を特徴量として扱う．また，半音ごとの遷移を評価するための異なる指標として，式（3.6）を用いて $c(t)$ を求める．$c(t)$ から 50 ms ごとに平均を算出して $c(\bar{t})$ とする（平均算出のための分析フレームは 1000 ms とした）．$c(t)$ 及び $c(\bar{t})$ を用い，有声区間の標準偏差を求めた．

$$c(t) = \mathrm{mod}(f_{\mathrm{cent}}, 100) \tag{3.6}$$

加えて，$c(t)$ 及び $c(\bar{t})$ を平均値が 0 になるよう標準化し，式（3.1）に代入することで，歌声の有声区間における傾斜を求めた．時間経過による $c(t)$ のずれを評価する指標として用いる．

ビブラート ビブラートは歌唱力の評価に影響する重要な特徴量である [20]．そのため，文献 [20] と同様に時刻 t におけるビブラートの速さ（5-8 Hz）に相当する周波数帯域のパワー $\Psi_v(t)$ とビブラートらしさ $P_v(t)$ を求める．ビブラートの深さが 30-150 cent であり，分析区間（320 ms）の平均音高と 5 回以上交差する区間をビブラートであると定め，その区間における $\Psi_v(t)$ 及び $P_v(t)$ の最大値，平均値，標準偏差を算出する．また，有声区間においてビブラートであると判断された区間の割合，ビブラートの速さ（毎秒に生じる揺らぎの回数），深さ（平均音高からの音高の変動幅）も特徴量として扱う．本研究では，$F_0(t)$ から次式のようにビブラートを含む変動成分を抽出して $f_d(t)$ とした後，上記特徴量を抽出する．

$$f_d(t) = F_0(t) - f_l(t) \tag{3.7}$$

ここで，$f_l(t)$ は，$F_0(t)$ にカットオフ周波数 5 Hz のローパスフィルタをかけて変動を除去したものである．

F_0 の動的特徴量 歌声の $F_0(t)$ における重要な要素として，プレパレーションやオーバーシュート [21] など，異なる音高へ遷移する際の動的特徴がある．本研究では，式（3.1）の $y(k)$ に $F_0(t)$ を代入して回帰係数 $\Delta F_0(t)$ を求め，F_0 の動的特徴量として扱う．4 種類のフレーム幅（K=10, 25, 50, 100）を用い，有声区間の統計特徴量を算出する．また，求めた $\Delta F_0(t)$ を式 (3.1) の $y(k)$ に代入して同様に $\Delta\Delta F_0(t)$ も求め，有声区間の統計特徴量を算出する．

F_0 の安定度 $\Delta F_0(t)$ において，有声区間中で変動が極めて小さい部分（$|\Delta F_0(t)| < 0.0005$）の割合を求め，どの程度 $F_0(t)$ がぶれずに歌えているかを評価する．4 種類のフレーム幅（K=10, 25, 50, 100）を用いた．

3.4.3 音響特徴量の主成分分析

算出した 221 種類の音響特徴量を用い，主成分分析を行う．主成分分析により得られる合成得点を重回帰分析の説明変数として用いることにより，多重共線性などの問題を回避することができると考えられるためである．

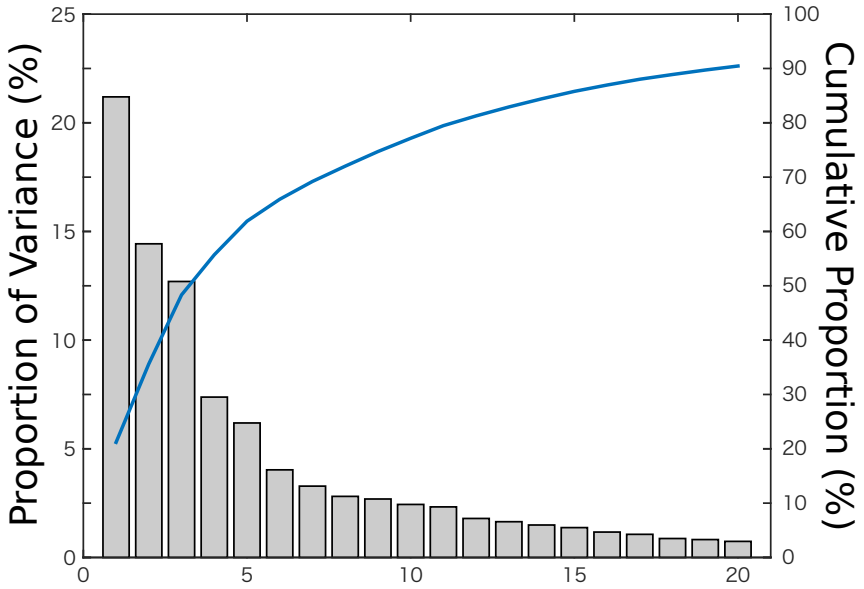

図 3.2: 第 20 主成分までの寄与率と累積寄与率

音響特徴量を特徴量ごとに標準化し，主成分分析を行った結果，第 20 主成分までで累積寄与率が 90% に達した．第 20 主成分までの各主成分の寄与率と累積寄与率を図 3.2 に示す．主成分分析では，分析に用いたサンプル数（歌唱データ 60 歌唱）より一次元少ない数の主成分を得ることができるため，重回帰分析では，全 59 主成分を説明変数として用いることとする．

3.4.4 モデルの再構築：重回帰分析

修士論文と同様，44 語の印象評価語の得点,「迫力性」「丁寧さ」「明るさ」の 3 因子の得点，及び歌声の評価に重要であると考えられる 3 語の得点を目的変数とし，59 種類の主成分得点を説明変数とした重回帰モデルを構築する．説明変数として，主成分ごとに標準化した値を用いることで，各モデルにおける回帰係数を偏回帰係数として得られる．つまり，各説明変数がどの程度印象推定に寄与しているかを表す指標として用いることが可能となる．説明変数の数が 59 種と多いため，ステップワイズ変数選択法を用い，計 47（44 + 3） 種類のモデルを構築した．

モデルの評価には，自由度調整済み決定係数 $\hat{R}^2$，Leave-one-out 交差検定（LOO）による重相関係数 R_{LOO} を用いる．さらに，特定の歌唱者を除いたデータでの交差検定を Leave-one-singer-out 交差検定（LOSO）と呼び，その重相関係数 R_{LOSO} も分析する．$\hat{R}^2$，R_{LOO}，R_{LOSO} の値が 1 に近いほど，モデルの推定精度が高いことを意味する．

自由度調整済み決定係数 $\hat{R}^2$ 重回帰モデルでは説明変数が増えるほどモデルの説明力が高まるため，説明変数の数の多さを考慮した自由度調整済み決定係数 $\hat{R}^2$ を式 (3.8) により求める．ここで，m_n は印象評定実験による実測値，e_n はモデルによる推定値，

$\bar{m}$ は実測値の平均値，N はデータサンプル数，P はモデルに含まれる説明変数の数を表す．

$$\hat{R}^2 = 1 - \frac{\sum_{n=1}^{N}(m_n - e_n)^2/(N-P-1)}{\sum_{n=1}^{N}(m_n - \bar{m})^2/(N-1)} \tag{3.8}$$

重相関係数 R_{LOO} Leave-one-out（LOO）交差検定では，特定の歌声データを除外し，残りのデータを用いて重回帰モデルを作成する．その際，全データを用いて構築されたモデルで，印象推定に有効だと判断された特徴量を説明変数として用いる．そして，作成した重回帰モデルから．除外した歌声データの印象を推定することで，実測値と推定値の比較を行う．この分析を 60 データの歌声全てに対して行い，全 60 データの歌声における印象得点の実測値 m_n（$n = 1, 2, ..., N$）と推定値 e_n（$n = 1, 2, ..., N$）におけるピアソンの積率相関係数（以降，相関係数と呼ぶ）を求める．ここで，N はデータサンプル数を表す．得られた相関係数を二乗し，重相関係数 R_{LOO} を求めた．

重相関係数 R_{LOSO} Leave-one-singer-out（LOSO）交差検定では，同一歌唱者による歌声データの影響を排除するため，特定の歌唱者の歌声データを除き，LOO と同様の手順で重相関係数 R_{LOSO} を求めた．

3.4.5 モデル構築結果と考察

重回帰分析及び交差検定の結果を表 3.7 に示す．各モデルは，全て $p < .001$ で有意であった．印象評価尺度においては,「迫力性因子」や迫力性に関わる「勢いがある」「声量のある」「弱い」「静かな」といった語，及び「聴きやすい」「無邪気な」という評価語では決定係数が $\hat{R^2}$ が 0.8 を超えており，特徴量からの印象推定精度が高いと言える．特に,「迫力性因子」に関しては R_{LOO} と R_{LOSO} の結果においても 0.9 を上回っており，モデル学習に用いていない歌唱者の歌声でも十分に印象推定が可能と言える．その他には「透き通った」「可愛い」といった評価語の推定精度が比較的高く，決定係数 $\hat{R^2}$ が 0.7 以上であった．44 語の評価語全体においては，$\hat{R^2}$ が 0.8 以上の語が 14 語，$\hat{R^2}$ が 0.7 以上の語が 25 語であった．

先行研究 [1] と比較した結果を，表 3.8 に示す．提案手法では，迫力性因子，丁寧さ因子，3 因子の平均，尺度の 12 語の平均，44 語の平均それぞれにおいて，決定係数が上昇していることが分かる．

3.4.6 印象推定モデルについての考察

推定モデルにおける R_{LOSO} の値が R_{LOO} の値よりも小さい評価語では，歌声の印象が歌唱者に依存していると考えられる．例えば，印象評価尺度における「透き通った」「嬉しそうな」といった評価語がこれに該当する.「うまい」「好きな」という評価語においても，R_{LOO} と比較して R_{LOSO} の値が小さい．歌唱技術の差はそれぞれの歌唱者に依存すると考えられるため，この推定結果は妥当といえる．また,「好きな」という評価語に関しても同様に，評価者の歌声の好みが歌唱者に依存していると考えられる．

表 3.7: 各印象推定モデルにおける自由度調整済み決定係数及び重相関係数

印象評価尺度の 12 語

	$\hat{R}^2$	R	
評価語		LOO	LOSO
勢いがある	0.840	0.791	0.811
声量のある	0.865	0.820	0.818
弱い	0.929	0.875	0.869
静かな	0.887	0.838	0.824
聴きやすい	0.800	0.703	0.691
透き通った	0.723	0.640	0.598
落ちつきのある	0.688	0.597	0.582
響きのある	0.698	0.600	0.612
嬉しそうな	0.534	0.454	0.419
軽やかな	0.518	0.449	0.447
可愛い	0.728	0.628	0.617
無邪気な	0.855	0.790	0.798
平均	0.755	0.682	0.674

歌声の印象評価における 3 因子

	$\hat{R}^2$	R	
因子		LOO	LOSO
迫力性	0.958	0.923	0.931
丁寧さ	0.551	0.471	0.462
明るさ	0.643	0.574	0.593
平均	0.717	0.656	0.662

歌声評価に重要であると考えられる語

	$\hat{R}^2$	R	
評価語		LOO	LOSO
好きな	0.555	0.454	0.387
うまい	0.386	0.302	0.217
曲に合ってる	0.238	0.176	0.131

$\hat{R}^2$ の値が大きかった 10 語

	$\hat{R}^2$	R	
評価語		LOO	LOSO
繊細な	0.938	0.900	0.896
弱い	0.929	0.875	0.869
激しい	0.925	0.887	0.872
気持ち良さそうな	0.888	0.809	0.812
静かな	0.887	0.838	0.824
声量のある	0.865	0.820	0.818
鼻にかけたような	0.862	0.766	0.786
無邪気な	0.855	0.790	0.798
優しい	0.851	0.791	0.800
勢いがある	0.840	0.791	0.811
参考：44 語の平均	0.685	0.607	0.595

表 3.8: 先行研究と本研究における推定精度の比較

評価対象	先行研究	本研究
迫力性因子	0.880	0.958
丁寧さ因子	0.481	0.551
明るさ因子	0.676	0.643
3 因子の平均	0.679	0.717
尺度の 12 語の平均	0.627	0.755
44 語の平均	0.614	0.685

また，歌声データごとの推定精度の指標として，重相関係数 R_s^I を求める．44 語の評価語，3 種類の因子，歌声評価に重要であると考えられる 3 語の印象得点を対象とし，印象評価実験による実測値 m_i（$i = 1, 2, ..., I$）とモデルによる印象得点の推定値 e_i（$i = 1, 2, ..., I$）の相関係数を求め，二乗することにより重相関係数 R_s^I を求める．I は対象とした印象得点の数を表す（$I = 50$）．60 個の各歌声データにおける重相関係数（$R_s^{I=50}$）の分布を図

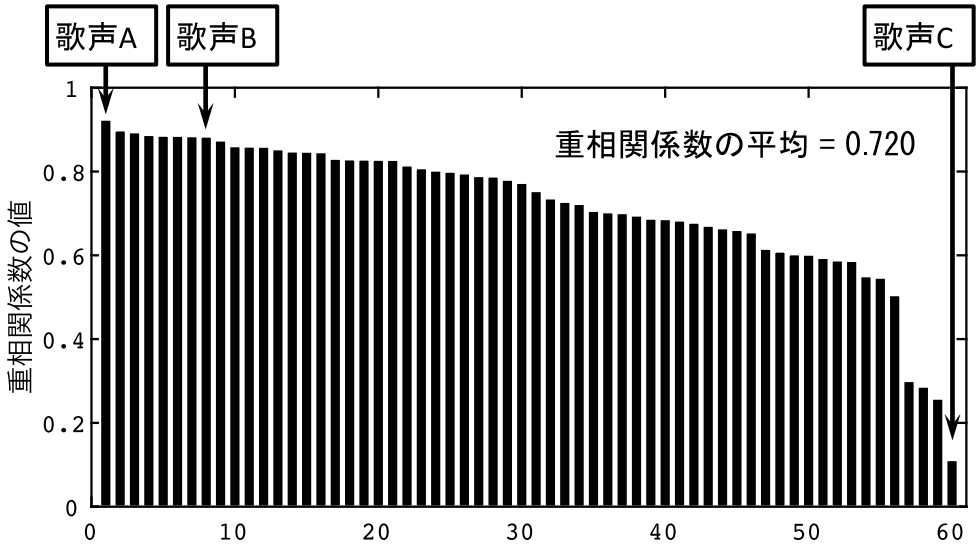

図 3.3: 60 個の歌声データそれぞれにおける，50 種の推定値の重相関係数 $R_s^{I=50}$

表 3.9: 印象の自動推定例

	実測値		推定値	
歌声 A	美しい	1.181	美しい	1.294
	女性的な	1.103	伸びやかな	1.132
	響きのある	0.906	透き通った	1.097
	伸びやかな	0.893	落ちつきのある	0.978
	優しい	0.860	女性的な	0.862
歌声 B	かっこいい	1.805	声量のある	1.356
	芯のある	1.379	伸びやかな	1.098
	声量のある	1.310	かっこいい	1.095
	勢いがある	1.253	勢いがある	1.048
	安定している	1.069	芯のある	0.858
歌声 C	女性的な	1.228	女性的な	1.191
	ぶりっこみたいな	1.070	伸びやかな	0.733
	少女のような	0.966	真っすぐな	0.708
	特徴的な	0.778	気持ち良さそうな	0.677
	甘い	0.687	無邪気な	0.557

3.3 に示す．この値が 1 に近いほど，推定値と実測値との誤差が少ないと言える．

ここで，全 60 歌唱における $R_s^{I=50}$ の平均は 0.720 であり，高い精度で印象の自動推定ができていると言える．また,「うまい」「好きな」「曲に合ってる」という 3 語と 3 因子の得点を除いた重相関係数（$R_s^{I=44}$）においては，全 60 歌唱の平均が 0.772 であった．つまり，44 語の印象評価語に限定した用いた印象推定では，より高い精度で歌声の印象を自動推定できていると言える．

実際の推定例として，重相関係数が最も高かった歌声 A，歌声 A とは異なる印象であり 8 番目に重相関係数が高い歌声 B，最も低かった歌声 C について，印象の自動推定結果を表 3.9 に示す．ここでは，印象評価語 44 語における，印象得点上位 5 語を記載している．

歌声 A では，上位 5 語のうち 3 語が重複しており，類似した印象を自動推定できていると言える．また，歌声 B についても，歌声 A とは異なる印象の傾向であるが，上位 5 語中 4 語が重複しており，印象の傾向によらず自動推定が行えていると言える．

一方，$R_s^{I=50}$ が最も小さかった歌声 C では，最も印象得点が高かった語は重複しているものの，他 4 語は大きく異なっている．図 3.3 において，特に相関の小さい（$R_s^{I=50}$=0.4 以下の）歌声データが 4 つあるが，それらは歌声 C と同様に「ぶりっこのような」「特徴的な」という印象の実測値が高い歌声であった．この 2 つの評価語の $\hat{R}^2$，R_{LOO}，R_{LOSO} の値は全 44 語中最も低く，自動推定が難しい語であるということがわかる．これらの歌声を聞いてみると，無理やり声を作っているような，歌声としての不自然さが感じられる歌声であった．このような，ある種の不自然さが生じることにより，主観評価が適切に行われず，モデルによる推定精度との誤差が大きくなったのだと推測される．

3.4.7 主成分得点ごとの考察

表 3.10 では，印象評価尺度 12 語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唱の特徴に起因していると考えられるか，表 3.11 にそれぞれの特徴を示した．

重回帰モデルに用いた説明変数は 59 種類の主成分得点であるが，ステップワイズ選択法を用いたため，モデルごとに採用された変数が異なっている．第 9 主成分以降では，尺度の 12 語のモデルにおける採用率が低くなっており，第 8 主成分（累積寄与率 72.0%）までである程度モデルを説明できると考えられる．そのため，表 3.10 では，第 8 主成分までの結果を示している．数値が記載されていない主成分は，その印象の推定モデルでは用いられていないことを表しており，第 5，6，7 主成分では，寄与していた推定モデルが 12 語中 3 語以下であった．そこで，上位 8 主成分のうち，12 語中 4 語以上の印象推定に大きく寄与していると考えられる第 1，2，3，4，8 主成分についての考察を以下に述べる．それぞれの主成分において負荷量の高かった音響特徴量上位 10 種類 (表 3.12) を取り上げ，考察を行う．表中，及び本文中の*は，負荷量の値が負であったことを示す．

第 1 主成分 「少女のような」「可愛い」「伸びやかな*」「声量のある*」「男性的な*」「中性的な*」「優しい」「ドスが効いている*」「繊細な」「少年のような*」「軽やかな」「弱い」「芯のある*」「静かな」「迫力性因子*」「明るさ因子」の各モデル内で，偏回帰係数が最も高い説明変数となっている．負荷量が高い音響特徴量が F_0 の動的特徴量で占められており，特に $\Delta f_0(t)$ の標準偏差が重要であることから，F_0 が遷移する速度の多様さを反映していると考えられる．

第 2 主成分 「美しい」「心のこもった」「透き通った」「繊細な」「陽気な*」の各モデル内で，偏回帰係数が大きい説明変数である．この主成分では，パワーの動的変動量の小ささ及びスペクトル傾斜が大きく影響しており，ある種の声質表現を行おうとした結果，それに伴いパワーの変動も小さくなっていると考えられる．主に合唱経験のある歌唱者の歌において得点が高くなっていたため，合唱における発声練習により習得可能な声質が関係していると考えられる．

表 3.10: 各印象推定モデルにおける第 1 主成分から第 8 主成分の偏回帰係数

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 3.11: 3.4.7 で考察を行った各主成分の特徴と根拠となる特徴量

1	基本周波数の遷移の多様さ（$\Delta f0$ の標準偏差）
2	スペクトル傾斜（0-22.05 kHz における中央値） パワーの変動の小ささ（ΔP の平均*）
3	スペクトル包絡の変動の多さ（$\Delta S_{\log}$（全帯域）） 基本周波数の変動の多さ（$\Delta\Delta f0$ の中央値）
4	口唇の開口度合い・声道長の短さ（F_1 の平均） 立ち上がりの速さ（ΔS_{lin} の標準偏差）(0-3 kHz)
5	声質の多様性（スペクトル重心の標準偏差） ビブラートの少なさ（ビブラートの割合*）
6	対数スペクトル傾斜（0-6, 0-9 kHz における中央値） 口唇の開口のメリハリのなさ（F_1 の四分偏差*）
7	調音運動の変動の多さ（ΔF_2 の四分偏差，中央値） 対数スペクトル傾斜（0-22.05 kHz における標準偏差）
8	ビブラートらしさ（最大値，平均値） 調音運動のメリハリ（ΔF_2 の標準偏差）

表 3.12: 各主成分において負荷量が高かった音響特徴量

第 1 主成分	F_0 の動的変動量 (K=10) の標準偏差
	ΔF_0 の動的変動量 (K=100) の中央値
	F_0 の動的変動量 (K=25) の標準偏差
	ΔF_0 の動的変動量 (K=10) の平均
	F_0 の動的変動量 (K=10) の平均
	F_0 の動的変動量 (K=25) の平均
	ΔF_0 の動的変動量 (K=25) の平均
	ΔF_0 の動的変動量 (K=10）の標準偏差
	F_0 の動的変動量 (K=50) の平均
	ΔF_0 の動的変動量 (K=50) の平均
第 2 主成分	スペクトル傾斜 (0-22.05 kHz) の中央値
	パワーの動的変動量 (K=50) の平均 *
	スペクトル傾斜 (0-22.05 kHz) の平均
	パワーの動的変動量 (K=25) の平均 *
	パワーの動的変動量 (K=50) の標準偏差 *
	スペクトル包絡全体の動的変動量 (K=50) の四分偏差 *
	スペクトル包絡全体の動的変動量 (K=50) の平均 *
	パワーの動的変動量 (K=100) の平均 *
	パワーの動的変動量 (K=25) の標準偏差 *
	スペクトル傾斜 (0-9 kHz) の平均
第 3 主成分	対数スペクトル包絡全体の動的変動量 (K=25) の中央値
	ΔF_0 の動的変動量 (K=10) の中央値
	F_0 の動的変動量 (K=10) の中央値
	対数スペクトル包絡全体の動的変動量 (K=50) の中央値
	F_0 の動的変動量 (K=25) の中央値
	F_0 の安定度合い (K=10) *
	F_0 の安定度合い (K=50) *
	スペクトル傾斜 (0-3 kHz) の平均
	F_0 の動的変動量 (K=10) の四分偏差
	スペクトル包絡 (0-3 kHz) の動的変動量 (K=25) の中央値
第 4 主成分	歌唱フォルマントの平均
	歌唱フォルマントらしさの中央値
	歌唱フォルマントらしさの標準偏差
	F_1 の平均
	歌唱フォルマントらしさの四分偏差
	スペクトル包絡 (0-3 kHz) の動的変動量 (K=50) の標準偏差
	スペクトル包絡 (0-3 kHz) の動的変動量 (K=25) の標準偏差
	F_1 の動的変動量 (K=50) の標準偏差
	スペクトル傾斜 (0-3 kHz) の標準偏差 *
	倍音構造 (H1/H2) の平均 *
第 8 主成分	対数スペクトル重心の四分偏差
	F_2 の動的変動量 (K=25) の標準偏差
	f_{cent} の標準偏差 (1000 ms 区間)*
	F_2 の平均
	F_2 の中央値
	F_2 の動的変動量 (K=50) の標準偏差
	対数スペクトルの重心の標準偏差
	ビブラートらしさの最大値
	F_2 の動的変動量 (K=10) の標準偏差
	ビブラートらしさの平均

第 3 主成分 「かっこいい」「激しい」「元気な」「勢いがある」「落ちつきのある*」「一生懸命な」「気持ち良さそうな」「シャープな」の各モデル内で，偏回帰係数が最も大きい説明変数である．スペクトル包絡の動的変動に大きく関与しており，また，F_0 の安定度が負の負荷量になっている点が特徴である．F_0 の動的特徴量も関与しているが，値の分散ではなく中央値が関係しているため，第 1 主成分とは異なり，1 歌唱中における変動の多さを反映させていると考えられる．

第 4 主成分 「ぶりっこみたいな」「嬉しそうな」「響きのある*」「色気のある*」「悲しい*」「無邪気な」の各モデル内で，偏回帰係数が最も大きい説明変数である．第 1 フォルマントと歌唱フォルマントが大きく関与している．歌唱フォルマントの有無は，一般的には「歌声らしさ」や「響き」と関連付けられており [22]，本分析での結果とは一致していない．本分析では 2-4 kHz の帯域のパワーの強さを歌唱フォルマントの指標としたが，F_1，つまり第 1 フォルマントの上昇により口唇の開口度が大きくなるのに伴い，該当帯域のパワーも強調されたのではないかと考えられる．

第 8 主成分 「甘い」「女性的な」「聴きやすい」「透き通った」「丁寧さ因子」の各モデル内で，偏回帰係数が最も大きかった説明変数である．先行研究において，ビブラートは歌唱力評価に関係することが明らかにされており [2],「うまい」という評価語と相関の高い「聴きやすい」が含まれていることなどから，先行研究と一致する結果が得られたと言える．また，F_2 に該当する第 2 フォルマントの変動は口内の舌の位置に影響されることが知られており，F_2 の変動の分散が大きいということは，調音運動にメリハリがある，と解釈できる．その結果，丁寧さに関わる評価語に関与していたのではないかと考えられる．

異なる楽曲に対する印象推定精度

本研究では，日本のポピュラー音楽におけるアマチュア女性歌唱者の歌声に対して，音楽的な専門知識を持たない一般人が認知する印象を推定可能なモデルの作成を行った．様々な楽曲に対応できるよう，歌声の特徴量分析では，楽曲に依存しない特徴を扱っている．また，様々な歌唱者に広く対応できるよう，多様な印象が認知される歌声をモデル作成に用いた．しかし，印象推定精度の評価では「モデル作成に用いた歌声と同一の歌唱者」「同一の楽曲」における推定精度が評価対象であった．

そこで，提案手法によるモデルを用い，異なる歌唱者に対する印象推定精度を評価する実験を行った.「モデル作成に用いた歌声とは異なる女性歌唱者」6 名に「歌唱者が歌い慣れている既存曲」を歌唱してもらい，10 名の一般大学生による主観印象評価実験を行い，その結果と，提案手法での印象推定結果を比較した．

表 3.13 は各歌声の楽曲情報および印象得点に関する情報を示している．44 語の印象評価語の得点，歌声評価に重要な 3 語の得点, 3 因子の得点，計 50 種の得点において，評価者ごとに標準化を行い，10 名分の平均点と本提案手法による推定得点の重相関係数を $R_s^{I=50}$ で表している．全 6 データにおける $R_s^{I=50}$ の平均は 0.531 であり，同一楽曲，同一歌唱者の歌声を用いた場合よりも印象推定精度（重相関係数の値）が下がっていた．特に，歌声 D, 歌声 E の印象推定精度が 0.25 と低いため，以下に推定精度が下がった原因を考察する．

表 3.13: 異なる楽曲に対する印象推定精度の評価に関する詳細

楽曲に関する情報				
歌声	時間長	最低音	最高音	音域
D	7.4	59	66	7
E	12.4	62	66	4
F	9.6	58	73	15
G	8.7	60	75	15
H	10.9	56	68	12
I	12.4	59	71	12

最低音，最高音は MIDI ノートナンバーに換算した値を示している

印象得点に関する情報				
歌声	重相関係数 $R_S^{I=50}$	評価者間相関	レンジ	分散
D	0.256	0.114	2.20	0.47
E	0.251	0.455	3.50	0.97
F	0.643	0.602	3.94	1.19
G	0.640	0.474	4.21	1.03
H	0.802	0.558	3.92	1.18
I	0.593	0.462	3.16	0.97

重相関係数は 50 種の得点，それ以外は 44 種の得点を用いている

まず，歌声 D は，評価者同士の相関係数の平均値が，他の 5 データの値と比べ小さい（0.114）点が特徴である．この結果は，評価者によって評価傾向が大きく異なっていることを意味している．また，47 語における印象得点の最高点と最低点の幅（表 3.13 中の「レンジ」）及び分散が，全 6 データ中，最小であることも表 3.13 で示されている．つまり，歌声 D では突出した印象がなく，人間による主観評価が一貫性を欠いていると言える．そのため，印象推定精度を考察するための適切な正解データ（実測値）が得られなかったという理由で，歌声 D では，印象推定モデルの精度について十分に議論することができない．この歌声の印象を適切に推定するためには,「印象が感じられない」という印象評価語に基づく推定モデルを構築することが必要だと考えられる．一方，歌声 E は，歌声の音域（MIDI ノートナンバーに換算した最高音と最低音の幅）が非常に狭い（4）点が特徴である．それに加え，同じメロディを 2 回繰り返している構造であったため，表出可能な歌唱表現が制限されてしまい，印象推定に有用な特徴量が適切に算出できなかった可能性がある．この結果は，歌唱する楽曲の特徴に依存したものであり，極端に音域が狭い歌唱では，本論文で提案する手法を用いての適切な印象推定が難しい可能性を示唆している．

3.5　本章のまとめ

本章では，長時間における歌声の特徴を評価する方法を検討した．金礪の修士論文をベースに，音響特徴量の追加や主成分分析という過程を経て，重回帰モデルの再構築を行った．その結果，先行研究と比較し，印象の推定精度を向上させることができた．また，異なる楽曲に対する推定精度を検証したところ，歌唱する楽曲や歌唱者の特性によっては印象推定精度が下がってしまうことが明らかとなった，今後は，音域やテンポが異なる多様な楽曲を用い，印象推定モデルの推定精度を向上させる必要がある．

第4章　短時間の歌声における特徴の評価方法

この章では，短時間の歌声における特徴を把握する方法について検討する．なお，本章は筆者が第一著者である「歌唱音声における声質の特徴と想起される色の関係」[23] に基づいている．

4.1　目的

歌声の特徴を把握するためには，長時間における歌声の特徴と，短時間における歌声の特徴，双方を扱う必要がある．本章では「時間軸上の変化を表現するための時間長」を分析対象とし，ごく短い時間長から認知される歌声の特徴について考察する．その上で,「時間軸上の変化」を表現するために，主観評価実験を行い，歌声の特徴と色の対応関係を明らかにする．

4.2　声質の多様性

音高や音量の差異を表現する際には,「高いか低いか」「大きいか小さいか」といった一次元の評価軸を用いることが可能である．実際，既存の歌声可視化システムの多くは，横軸を時間軸に，縦軸を音量，あるいは音高に対応づけて可視化している．つまり，表現する情報が一次元のみであるため，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，声質の差異を十分に表現するためには，様々な評価軸が必要となる．以下に，声質の差異を分類する観点を述べる．

4.2.1　発声様式による差異

人間は，声帯を振動させ，声道で共鳴させることで発声する．その際，声帯振動の様式や声道の形状により，声質が特徴づけられる．Laver [24] は，喉頭部の制御，つまり声帯振動の様式の違いにより，Modal, Falsetto, Whisper, Creak(vocal fly), Harshness, Breathiness, という 6 種，及びそれらの組み合わせとして，声質を分類できると述べている．また，歌声においては Whistle, Falsetto, Modal, Vocal fry の 4 種に分類できることが榊原 [25] によって示されている．

4.2.2　声区による差異

歌声における声質評価では，声区という観点から声質が分類されることが多い．声区は，研究によって様々な定義が行われるが，Hollien [26] は声区を「ほぼ一定の声質によって

生成される，連続する声の周波数の系列あるいは領域であり，その基本周波数は声区間でわずかに重複する」と定義している．声区の分類は研究によって異なるが，Sundberg [13] は，男性歌唱は主に Modal, Falsetto という 2 種類, 女性歌唱は主に Chest, Middle, Head という 3 種類の声区に分類されることを示している．

4.2.3 感性的評価による差異

「美しい声」「かすれた声」のような感性的評価という観点も，多く用いられる．感性的評価では，人による主観評価によって得られた回答に基づき，声質の差異が特徴づけられる．木戸ら [8] は，通常発話音声を対象とし，日常的に使用される声質表現語の収集を行い，アンケート調査や統計手法を用いることで，声質の差異を表現するための 8 対の表現語対を抽出した．また，因子分析の結果として，声質が「明瞭性」「美的」「養護的」という 3 種の評価軸により表現可能であることを示している．

このように，声質の差異の表現には様々な観点が存在し，評価軸の選び方により，表現される情報が大きく変化する．つまり，声質を可視化する際には，どのような観点から，どのような評価軸を用いて表現するかを，十分に考慮しなければならない．

4.3 先行研究

4.3.1 声質を可視化する先行研究

声質の差異を可視化する試みとして，平山ら [17] は歌声の声区という観点から，声質の可視化を行っている．音響特徴量を用い，地声，裏声，喉締め声（声区の切り替わる部分で生じる発声）の 3 種の声質を判別し，時間軸にそって色の違いで声質の差異を表現している．加えて，松浦ら [27] も声区に該当する声質の違いを対象としており，声区の違いと文字のフォントの対応付けを試みている．

また，声質の感性的評価という観点から，菅原ら [28]，矢島ら [29] は「いい声」の度合いを音響特徴量から推定し，色やグラフを用いてリアルタイムに結果を反映させるシステムの開発を行っている．

これらの研究では,「声区の観点から，声質の種類の」「感性的評価の観点から，いい声の度合い」のように，可視化する評価軸を限定している．その結果，カテゴリカルな色の違いや，色の明度といった特定の要素のみで声質の差異を表現することができている．つまり，音高や音量と同様に，時間軸と共に一次元の情報のみを可視化していると言える．

しかし，多様な声質の差異を表現するためには，多次元の情報を可視化する必要がある．その際，時間軸にそった変化を示すためには，限られた空間内で，何らかの視覚情報の差異を表現する必要がある．従来の可視化では，横軸に時間軸を，縦軸になんらかの評価軸の強度を当てはめ，座標の値で情報を表現することが多いが，その場合は 1 次元の評価軸のみの表現となる．複数のデータが表示される折れ線グラフのように，複数の評価軸の値を縦軸に表示することもできるが，数が増えれば増えるほど，理解が困難となる．そこで，本研究では縦軸の座標位置を用いずに尺度値を表現する方法として「色」に着目し，声質と色の対応関係について調査を行う．

4.3.2　色と音の対応関係に関する先行研究

声と色の関係を調査した研究は少ないため，ここでは音と色の関係についての関連研究も述べる．

音の音色と色の関係については，楽器音を用いた研究が多く行われている．長田ら [30] は，色聴保持者と非保持者における，音色と色のマッピングについて検討している．非保持者を対象に，高調波構造の異なる 3 種類の音色で，5 種類の上昇スケールを用いた調査を行った結果，高調波成分が多い音色ほど，明度が低く彩度が高い色と対応づけられる傾向にあったことを示している．また，音高が高いほど，明度が高い色と対応づけられていた．Adeli ら [31] は，8 種類の楽器の音色を用いた調査により，ピアノやマリンバなどの柔らかい音色では，青と緑が，サックス，クラッシュシンバルやゴング，トライアングルなどのざらざらした音色では，赤か黄が選ばれやすい傾向であることを明らかにしている．加えて，低い音高では青色が有意に選ばれやすいことも示している．赤井ら [32] は，26 種類の楽器の音色と色彩の寒暖の関係を調査しており，音の減衰の度合いと，倍音成分の多さにより，音色を暖色系・寒色系それぞれ 2 クラスタずつに対応させる十分条件を明らかにしている．

声質と色の関係を対象とした研究としては，話声を用いた Moos ら [33] の研究が挙げられる．発声様式の観点に基づく，知覚上識別が可能な 10 種類の声質による読み上げ音声を用い，色との対応関係を調査している．その結果として，音声のスペクトル傾斜が小さくなるほど青みを帯び，音高が高くなるほど明度が高く，赤色を帯びるという関係を明らかにしている．しかし，用いた話者数が 2 名と少ないため，話者の個人性に起因する声質の差異は反映されていないと考えられる．

また，声質とは異なるが，音声の母音と色の対応関係についての研究は多く行われている．Wrembel [34] は，共感覚非保持者を対象に，12 種類の英語の母音を用いた調査を行い，明るい色（黄色，緑）は高母音，及び前舌母音に関係し，暗い色（茶，黒，青）は，後舌母音に関連付けられることを明らかにしている．また，開口母音は赤に，中舌母音は灰色に関連することも示している．Moos ら [35] は，共感覚保持者と非保持者を対象に，16 種類の英語の母音を用いた調査を行い，第 1，第 2 フォルマント周波数（母音を特徴づける音響特徴，以降，F1, F2）との対応関係を明らかにしている．両グループにおいて，母音における F1 の上昇が，赤―緑という色相の軸に有意に対応しており，色の明度は F1, F2 の上昇と有意に関係していたことを明らかにしている．また，保持者と非保持者，両者で音響特徴と色が対応づいている傾向が見られたが，保持者ではより顕著に傾向が現れていたと述べている．

このように，声質と色の関係についての先行研究は行われているものの,「歌声の声質が対象とされていない」「話者数，及び声質の多様性が不十分である」「感性的評価の観点からは十分な検討がされていない」という問題がある．そこで，本研究では以下の条件を満たすよう配慮し，声質と色の対応関係について調査する実験を行う．

- 歌声として発声された音声の声質を対象とする
- 多数の話者，多様な表現の声質を刺激として用いる
- 感性的評価の観点から，声質の特徴と色を対応づける

4.4　声質と色の対応関係に関する実験

本実験では，歌声における声質を対象とし，色と対応付けることが可能かどうかを検証する. 刺激音声の選定の際には，声質の多様性が保たれるよう工夫を行った．また，声質の差異を感性的評価の観点から表現するために，声質の差異を十分に評価可能である声質表現語を用いた主観評価実験を行う．そして，主観評価実験によって得られた心理的特徴と音声の物理的特徴が，色の要素とどのように対応づけられるか，明らかにする．なお，本研究では性差による影響の軽減や応用場面を考慮し，アマチュアの女性歌唱者のみの歌声を分析対象とした．声質の特徴と色がどのように対応しているかが明らかになることで，直感的に理解しやすい，色による声質の特徴の可視化に活かすことが可能となる．

4.4.1　実験方法

本研究では，歌声の声質に対する印象，及び色との対応関係を調査するために，同一の歌声刺激を対象に，以下の 2 種類の主観評価実験を行った．

1. SD 法を用いた印象評定実験
   歌声の声質の特徴を感性的観点から評価するために行う．13 対の表現語を用い，7 段階での評価を求めた．

2. 一対比較評価実験
   歌声の声質と色との適合性を測るために行う．色相，明度，彩度の 3 条件に分け，計 26 色の色刺激を用いた．

音声刺激

本研究では，歌声の声質の違いによる評価の違いを分析する必要があるため，使用する音声刺激は声質以外の要素が統一されていることが望ましい．しかし，複数の歌唱者において，音高や音量が統一された条件で収録されたデータベースが見つからなかったため，収録を行った．Sundberg [13] は，女性の 3 種類の声区（Chest, Middle, Head）について，個人差はあるものの 400 Hz, 660 Hz 近辺で声区が重複していると述べている．そこで，複数の声区を含むよう，C4（約 261 Hz）から E5（約 659 Hz）までの各音階の収録を行った．特定の音高のみの発声を求めると，音声の歌声らしさが失われる懸念があったため，C4 から E5 までの上昇音階を，各音 1 秒ずつ「あ」母音で歌唱した音声を収録した．ただし，収録の際に D5 及び E5 の音高で発声が不安定になる歌唱者が多かったため，これらを刺激の対象から除外した．

収録対象は，21 名のアマチュア女性歌唱者（20 代）である．収録された音声から，多くの歌唱者で音高が安定していた音階を抽出し，音高の違いによる声質の多様性を確保するため，E4（約 329Hz）, G4（約 392Hz）, C5（約 523Hz）の 3 種類の音高を刺激として用いた．本稿ではそれぞれ，low, middle, high と表記する．歌唱者ごとに，各音高が安定している区間を 600ms の長さで切り出し，前後 40ms に対しハニング窓で窓かけ処理を行った．600ms という長さは,「十分に印象評価を行うことができ」「音高の影響を受けないような短さ」になるよう，予備的な検討を経て決定した．切り出した音声を用い，

声質の多様性を担保するため，予備的な印象評価実験を 3 名で行った．用いた評価項目は，後述する声質表現語 8 対と，歌声の印象評価に関わる 3 因子（迫力性，丁寧さ，明るさ）[36] である．評価結果が類似していた音声を除外し，28 データを刺激として選定した．音高ごとに low が 8 データ, middle が 11 データ, high が 9 データである．3 種の音高に共通して使用されている歌唱者は 7 名であった．

印象評価に用いる表現語

声質に関して，木戸ら [8] は話声における通常発声を対象とし，表現語の収集を行った上で 8 対の表現語からなる評価尺度を構築している．その際,「明瞭性」「美的」「養護的」という 3 つの因子により話声の声質が評価されることを明らかにしている．しかし，話声と歌声では印象空間を構成する因子が異なる可能性がある．

歌声においては，声質に関わる要素として声区という概念が存在し，声区の違いによる声質の違いについて，テノールとアルトにおける同一の音高及び音量の歌唱では soft（柔らかい），smooth（滑らかな）及び rough（粗い），harsh（ざらざらした）と表される次元があることを Sundberg [13] は示している．

また，歌声研究において,「歌声らしさ」も印象評価に関わる重要な要素である．大石ら [37] は，人間は 200 ミリ秒, 1000 ミリ秒 の音声信号において，朗読音声と歌声の識別が可能であることを明らかにした上で，音声の信号長が 1000 ミリ秒よりも短い場合にはスペクトル包絡の特徴量を用いた識別が有効であることを示している．この結果は，1 秒未満の音声から，声質により「歌声らしさ」を知覚できる可能性を示唆している．

齋藤ら [22] は，歌声らしさの聴覚印象には「揺れ」「響き」といった基本的な心理的特徴が大きく寄与していることを指摘しており,「響き」には 3000Hz 付近のスペクトルピーク成分，及び同帯域における強い高調波成分が重要であることを示している．

そこで，本研究では木戸らが作成した声質表現語 8 対に独自に「滑らかな-粗い」「柔らかい-硬い」という 2 対，及び，歌声らしさの知覚に関わる「響きのある-響きのない」「歌声らしい-話声らしい」という 2 対，また，音色の印象研究にて多く扱われている「明るい-暗い」という表現語対を加えた（表 4.1）．

表 4.1: 印象評価に用いた表現語 13 対

木戸 [8] の声質表現語	その他の表現語
高い–低い	滑らかな–粗い
男性的な–女性的な	柔らかい–硬い
かすれた–澄んだ	歌声らしい–話声らしい
落ち着きのある–落ち着きのない	響きのある–響きのない
迫力のある–弱々しい	明るい–暗い
太い–細い	
張りのある–張りのない	
鼻声–鼻声でない	

印象評価に用いる色刺激

本研究では，PCCS 表色系に基づき，色刺激の選定を行った．PCCS 表色系では明度と彩度の複合概念として「トーン」を用いており，12 種類のトーンと色相を用いて色を表すことができる. 同一のトーン内では，共通したイメージを与えるという特徴があるため，色相の比較に適していると判断した．

一般に，色を用いた厳密な実験では，色票を用い，照度などの周囲の環境も考慮し，色の条件を統制した上で評価を行う．しかし，本研究は歌声の可視化を目指した検討であり，電子媒体上での色刺激と声質の対応関係を調査しなければならない．電子媒体上での色再現に関して，若田ら [38] は，iPad を用い，PCCS 表色系の視感測色を試みている．そこで，本研究では若田らの視感測色結果を参照し，色刺激の明度や彩度，色相条件の統一を図った．

以下に，各条件における刺激の選択基準を述べる．

明度・彩度条件

- 同一色相に含まれる 2 色を 1 対とする
- 比較する条件以外の要素が類似した組み合わせにする
  明度条件における良い例：色相 2 における，b と dp（彩度がほぼ同じ）
  明度条件における悪い例：色相 2 における，b と g （彩度が異なる）

色相条件

- トーン内における明度，彩度の分散が小さいトーン
- トーンの彩度が小さすぎないトーン
- 心理四原色を用いる（色相番号 2, 8, 12, 18）
- L*a*b 表色系における a*, b* に該当する色相を用いる（色相番号 14, 24）

明度，彩度条件に関しては，PCCS における同一色相に含まれる 2 色を 1 対とし，合計 5 対ずつを選定した．色相条件に関しては，トーン内における明度，彩度の分散が小さく，彩度が小さすぎない，つまり色味を十分に感じられるトーンを選定した．その後，心理四原色である色相（2, 8, 12, 18）を選定した．また，色相を連続した数値で表すことを想定し，L*a*b* 表色系における a*, b* の特徴を表す代表的な色相（14, 24）を加え，計 6 色を用いた．本稿では，それぞれを赤，黄，緑，青緑，青，赤紫と表記する．選定した PCCS の各色を l*a*b* 表色系に変換し，さらに sRGB 値に変換した上で，色画像の作成を行った．色刺激として用いた色を図 4.1 に示す．図中の「P1」から「P5」は，明度，彩度条件での対の番号を示している．

図 4.1: 印象評価に用いた色刺激 26 色

実施方法

評価者は，53 名の大学生（男性 27 名，女性 26 名）である．Web 上にアンケートページを作成し，各自の MacBook を用いての回答を求めた．実験前に，指定したディスプレイ環境（Mac 標準ガンマ 2.2, ホワイトポイント D65）に設定してもらい，ディスプレイの輝度は最大にするよう教示した．また，同条件で画像を表示できるよう，アンケートページにアクセスするためのブラウザも指定している．まず，色との適合度評価を行うため，音声刺激を 1 つずつ提示し，2 種類の色画像を 5 対提示した上で「どちらの色がより歌声に合っている」かそれぞれ回答を求めた．音声刺激及び色刺激の提示順は評価者ごとにランダマイズしている．その後，各音声刺激に対し，表現語を用いた 7 段階評価を求めた．

4.4.2 結果及び考察

色刺激を用いた一対比較評価の結果

まず，歌声の声質と色との対応付けが可能かどうかについて，一対比較評価実験の結果を以下に述べる．

（1） 二項検定による評価の有意差の検定

色相条件 15 対，明度及び彩度条件各 5 対の色刺激において，色刺激の選択に有意差があったかどうか，二項検定を用いた検定を行った．歌声 28 データのうち，有意差が認められたデータの割合を表 4.2, 表 4.3 に示す．なお，p 値は調整済みの値を用いた．表 4.2

表 4.2: 二項検定で有意差が認められた割合（色相）

	赤	黄	緑	青緑	青	赤紫	平均
赤		0.04	0.11	0.11	0.36	0.36	0.19
黄	0.04		0.36	0.29	0.50	0.39	0.31
緑	0.11	0.36		0.00	0.32	0.00	0.16
青緑	0.11	0.29	0.00		0.43	0.00	0.16
青	0.36	0.50	0.32	0.43		0.14	0.35
赤紫	0.36	0.39	0.00	0.00	0.14		0.18

表 4.3: 二項検定で有意差が認められた割合（明度・彩度）

	P1	P2	P3	P4	P5	平均
明度	0.71	0.71	0.64	0.68	0.71	0.69
彩度	0.25	0.43	0.21	0.43	0.32	0.33

から，色相条件では，青及び黄が関わる選択において有意差がでた歌声データの割合が高いことが分かる．また，黄と青の対では，5 割の歌声データにおいて評価者の評価が一致している．一方，緑と青緑，緑と赤紫，青緑と赤紫の対では，有意差がでた歌声は一つもなかった．また，類似した色相でも，赤と赤紫においては有意差が見られた歌声データがあり，声質と色の適合度を詳細に考察するためには，細かな色相の違いも検討する必要があると考えられる．表 4.3 からは，明度条件では平均して 7 割近くの歌声データにおいて評価者の評価が一致していると言える．歌声データごとに有意差が出た刺激条件を比較すると，同じ音高の歌声でも，有意差の出やすい色の条件に違いが見られた．そのため，幅広い声質を可視化する際には，特定の色の条件だけでなく，複数の要素を用いる必要があるといえる．また，全ての歌声データにおいて，色相，明度，彩度のうち最低 1 つ以上の条件で有意差が認められていた．このことから，幅広い声質において，色の要素を 1 つ以上用いた可視化が有効であると考えられる．

（2）音高の違いによる有意差の検定

音高の違いにより色の選択度合いに差があったかを調べるため，一要因（3 水準)）の分散分析を行った．明度，及び彩度においては，明度（$p < .01$）及び彩度（$p < .05$）でそれぞれに有意差が認められた．多重比較の結果，明度では音高 low と middle, low と high，彩度では音高 low と high において有意差が認められた（図 4.2）．色相でも，各色の全組み合わせにて，一要因（3 水準）の分散分析を行い，有意差が認められた組み合わせを表 4.4 に示す．多重比較を行った結果，図 4.3 中の赤:青，赤:赤紫，緑:赤紫では， low と high, middle と high で，それ以外の対では，low と high でのみ有意差が認められた．これらの結果から，黄，青，赤紫が刺激対に含まれる際，音高の違いは色の選択に影響を及ぼしやすいと考えられる．

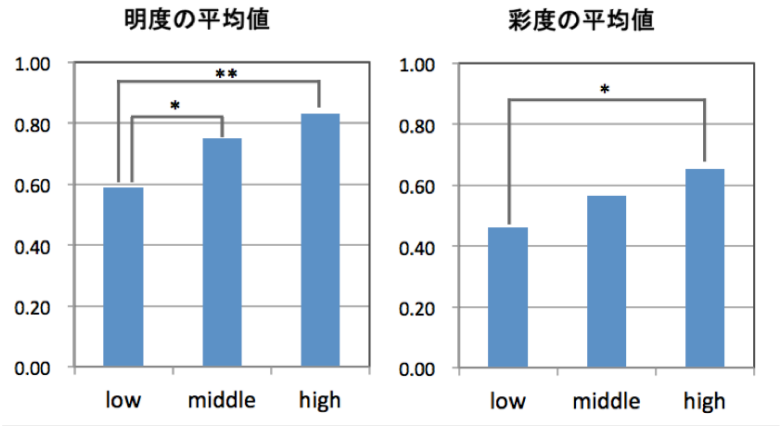

図 4.2: 明度・彩度における多重比較の結果

表 4.4: 分散分析で有意差が認められた色相の組み合わせ

	黄	緑	青緑	青	赤紫
赤				**	*
黄	-	*	*	**	**
緑	-	-		**	**
青緑	-	-	-	*	
青	-	-	-	-	*

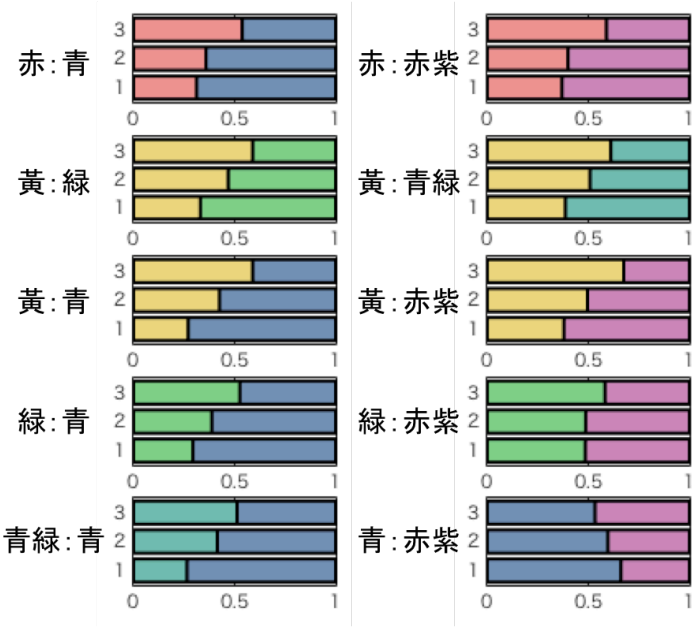

図 4.3: 音高の違いにより色相選択率が有意に異なる色相対：図中，縦軸の数字はそれぞれ「1」が low, 「2」が middle, 「3」が high を示す

色相に関する考察

（１）評価の一意性の検討

色相条件において，評価者が一貫した評価を行えているかを検討するため，一意性 ζ を算出した．一意性 ζ は一巡三角形の数により算出され，一巡三角形の数が少ないほど値が大きくなる（0 から 1.0 の範囲）．一巡三角形とは，三つの試料における優劣を比較した際に，三すくみの関係になっている状態を示す．つまり，一巡三角形が多く存在するほど，試料間の評価に矛盾が生じているといえる．

一巡三角形の数は,「評価が一次元でなく多次元構造である」「評価者が差異を正確に評価できない」「試料間に差異がない」という場合に増加するため，一巡三角形の少なさが評価者ごとの評価の一意性の基準となるのである．

まず，評価者ごとに一巡三角形の数 d を以下の式により算出した．k は試料数，a は各試料を他の試料と比較した際に選択された回数の合計，i は試料番号を示している. その後，一意性係数 ζ を算出する．

$$d = \frac{1}{6}k(k-1)(k-2) - \frac{1}{2}\sum_{i=1}^{k} a_i(a_i - 1) \tag{4.1}$$

$$\zeta = 1 - \frac{24d}{k^3 - 4k} \tag{4.2}$$

評価者ごとに，全 28 データに対する一意性 ζ の値の平均と標準偏差を算出した結果，及び，歌声ごとに，全 53 評価者における一意性 ζ の値の平均と標準偏差を算出した結果を図 4.6 に示す．これらの図から，評価者はもちろん，歌声によっても一意性 ζ の値が大きく異なることが読み取れる．この結果は，歌声によって，一次元上で色相の妥当な評価ができる声質とそうでない声質がある，ということを示唆している．つまり，色相を用いて声質の情報を可視化する際，その声質において色相による可視化が妥当かどうか，丁寧に考慮する必要がある．

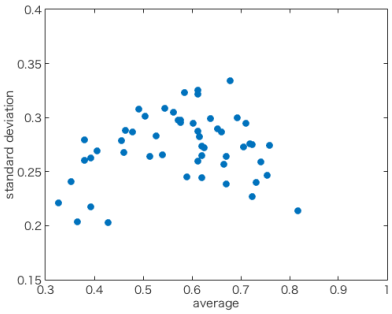

図 4.4: 評価者ごとの結果

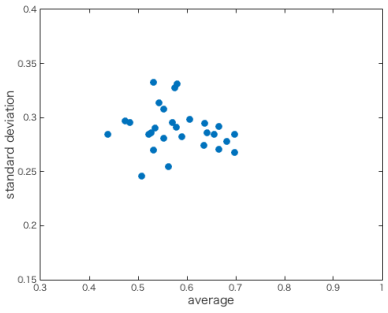

図 4.5: 歌声データごとの結果

図 4.6: 一意性 ζ の平均値と標準偏差

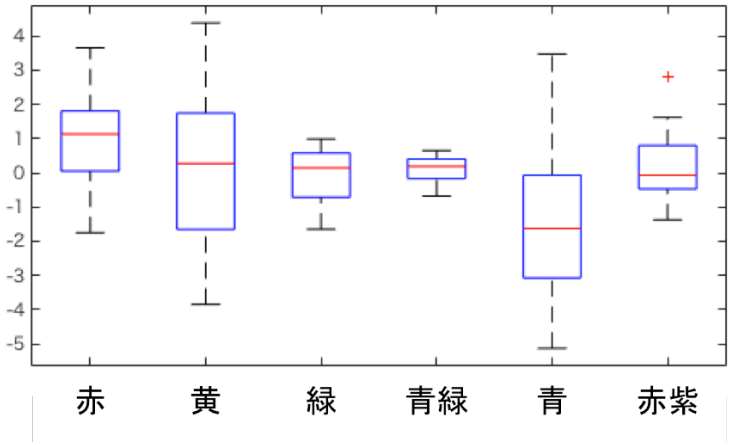

図 4.7: 各色相の尺度値の平均値

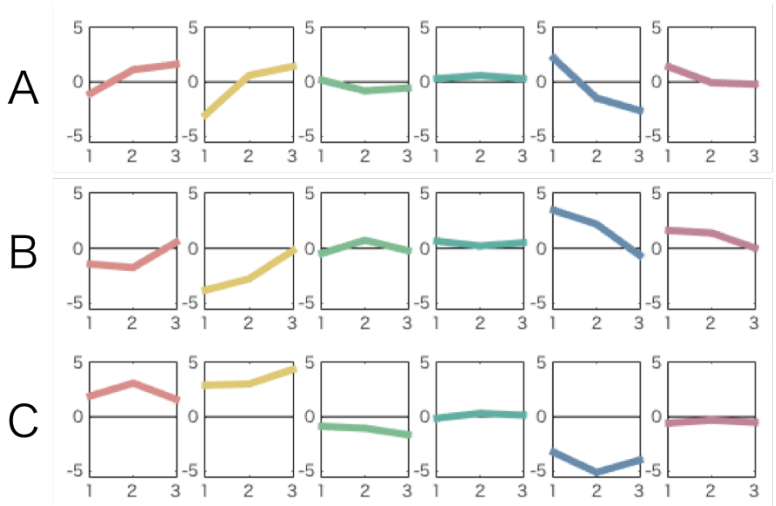

図 4.8: 歌唱者ごとの音高による尺度値の例

（2） サーストンの尺度値の算出

各歌声において適合度の高い色を明らかにするため，一対比較評価での選択度数を用い，サーストンの尺度値を色相ごとに算出した．色相ごとの，全歌声における尺度値の平均値を図 4.7 に示す．黄と青では分散が大きく，ある声質において他の色と比較する際に，重要な判断材料として扱われていたと考えられる．逆に，青緑，緑では分散が小さいため，他の色と比較する際には他の色を基準に，色が選択されていたと推測される．つまり，これらの色は声質から想起されにくい色相であったと考えられる．また，3 種の音高全てに含まれている歌唱者の，各歌声の尺度値の例を図 4.8 に示す. 図中の「1」「2」「3」はそれぞれ low, middle, high の音高に該当する．この図から，歌唱者によって音高による尺度値の変化の仕方が異なることが分かる．歌唱者 A は音高が変化すると各色相の尺度値が大きく変化しているが，歌唱者 C では音高による尺度値の変化が小さい．また，歌唱者 B は low 及び middle においては各色相の尺度値の分散が大きいが，high において値が 0 に近づいており，色相の評価が曖昧になっていることが分かる．このように，音高変化による尺度値の変化には声質の個人差が大きいといえる．

（3）色空間上での各歌声の配置

各歌声における各色相の尺度値を求めたが，各歌声に最適な色相を求められたわけではない．そこで，各歌声に最適な色相を一意に定めるため，L*a*b*表色系における，a*,b*に対応する値を算出した．

L*a*b*表色系では，明るさを表す L*，色相及び彩度に関わる a*, b* の 3 次元のパラメータで色を表す．a* 及び b*は正負の値をとり，値がともに 0 の場合には無彩色となる．a*が大きくなるほど赤みが，小さくなるほど緑味が強くなり， b* が大きくなるほど黄みが，小さくなるほど青みが強くなる．つまり，a* と b* の値を定めることができれば，該当する色相を明らかにすることができる．そのため，a* と b* を歌声の声質と対応づけることができれば，色相の違いを連続的に変化させることで，声質の細かな違いを，表現することが可能となる．

そこで,（2）にて算出した尺度値を歌声ごとに正規化し，各色相の a*, b* の値との積を求め，全ての平均 A*, B* を算出した．各歌声における A*, B* の値を図 4.9 に示す．この図から，high では，どの声質も似通った色相が選択されていたことがわかる．high では，多くの歌声において，声区でいう falsetto に該当する声質が確認されていたため，声質の差異が小さくなったのだと考えられる．low と middle では，b*軸における分散が大きいため，各声質の違いを，色相を用いて十分に表現できると考えられる．したがって，色相を用いて可視化を行う際には，low, middle の音域を主な対象とすることが望ましい．a*軸では分散が小さいが，より多くの歌声を刺激とすることで，詳細な検討が可能になると考えられる．

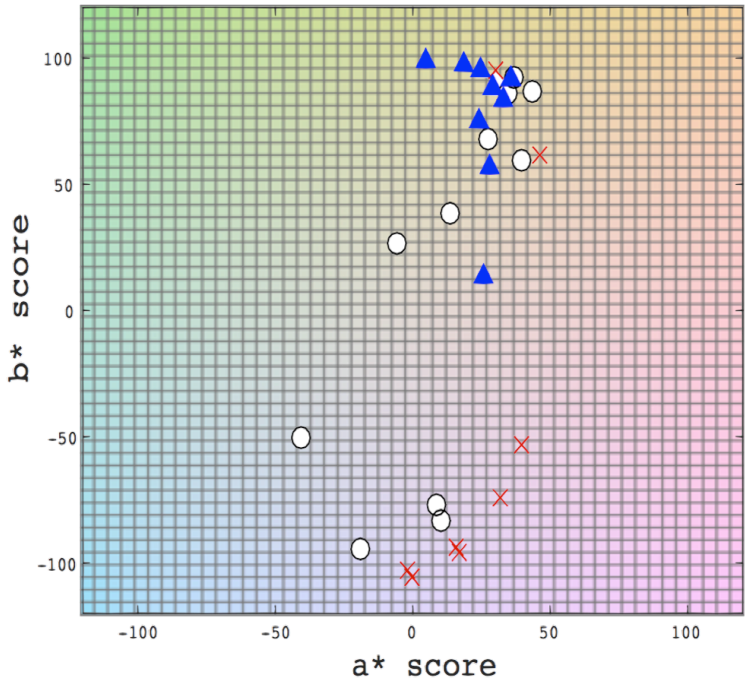

図 4.9: a*b*空間における各歌声の配置：図中の記号はそれぞれ「×」が low, 「◯」が middle, 「▲」が high を示す

表 4.5: 因子分析の結果

	評価性	活動性	迫力性
響きのある	0.950	-0.027	0.378
歌声らしい	0.913	0.074	0.071
かすれた	-0.905	-0.125	-0.210
滑らかな	0.879	0.120	-0.211
鼻声	-0.667	-0.126	0.005
高い	0.166	0.937	-0.026
明るい	0.116	0.859	0.355
男性的な	-0.100	-0.847	0.343
太い	0.005	-0.785	0.562
落ち着きのある	0.622	-0.664	-0.457
迫力のある	0.221	-0.249	0.974
張りのある	0.330	0.182	0.966
柔らかい	0.401	0.109	-0.771

表 4.6: 因子間の相関係数

	活動性	迫力性
評価性	-0.119	-0.021
活動性		0.066

色の特徴量を用いた相関分析

声質の特徴と色の特徴の対応関係を調べるため，声質の印象，声質の音響特徴量，色の特徴，それぞれを数値化した上で相関分析を行った．

（1）色の特徴量

一対比較評価の分析結果を用い，各歌声における色に関する特徴量を算出した．色相に該当する数値は，6 色相のサーストンの尺度値，及び A*, B* を用いる. 明度及び彩度については，5 対の刺激を用いた一対比較評価を行っているため，それぞれ「明度が高い」「彩度が高い」色刺激が選択された割合の平均値を指標として用いる．

（2）声質の特徴：印象表現語による得点

声質の表現語による評価結果を用い，各歌声における印象得点を算出した．評価者ごとに評価得点を標準化し，歌声ごとに平均得点を算出した. その後，印象空間を明らかにするため，因子分析を行った．因子数はスクリー基準に基づいて決定し，分析には最尤法，

プロマックス回転を用いた．分析結果を表 4.5 に示す. 抽出された 3 因子に対し，各因子の因子負荷量が高い評価語を参考に，それぞれ「評価性」「活動性」「迫力性」と命名した．これらの因子の因子間相関の値を表 4.6 に示す．この表から，3 因子はほぼ独立して声質の印象評価に寄与していると言える．各因子に寄与している評価語の得点の平均を算出し，各因子の得点とした. 相関分析では，13 対の印象評価語の得点，及び 3 因子の得点を印象得点として用いる．

（3）声質の特徴：音響特徴量

音響分析により，声質に関係する音響特徴量 10 種類を抽出した．分析に用いた歌声データは，44.1kHz, 16 bit サンプリングのモノラル信号である．以下に，各特徴量の概要と算出方法を述べる．まず，口唇の開口度や舌の位置，声道長に関わる指標である，フォルマント周波数の抽出を行った．Praat [39] を用い，分析フレーム長 50ms, 分析シフト長 10ms にて，第 1 フォルマント（以降，F1(t)），第 2 フォルマント（以降，F2(t)）を抽出した．t は，各時刻を示している．次に，歌声データを 10000Hz にダウンサンプリングし，フーリエ変換を行う．FFT のポイント数は 2048，分析フレーム長は 50ms，分析シフト長は 10ms とした．各時刻 t において算出したスペクトル $S_1(f,t)$ を用い，スペクトル重心 $C(t)$, スペクトラルロールオフ $R(t)$，スペクトラルフラックス $F(t)$ を求めた．これらは Timbral Texture Feature [14] として知られている．スペクトル重心，及びスペクトラルロールオフは，スペクトルの形状に関する指標であり，スペクトル重心は音色の明るさに関係していると言われている．スペクトラルフラックスは，局所的なスペクトル変化の指標である．以下に，各特徴量を算出するための式を述べる．なお，f は周波数ビンの番号，N は周波数ビンの数 $(N = 1024)$ を示している．

スペクトル重心 $C(t)$

$$P(t) = \sum_{f=1}^{N} S_1(f,t) \tag{4.3}$$

$$C(t) = \frac{\sum_{f=1}^{N} S_1(f,t) f}{P(t)} \tag{4.4}$$

スペクトラルロールオフ $R(t)$

$$\sum_{f=1}^{R(t)} S_1(f,t) = 0.85P \tag{4.5}$$

スペクトラルフラックス $F(t)$

$$F(t) = \sum_{f=1}^{N} (logS_1(f,t) - logS_1(f,t-1))^2 \tag{4.6}$$

最後に，音声分析変換合成法 STRAIGHT [12] を用いて，1ms ごとに F0（基本周波数），

スペクトル包絡 $S_2(t)$，非周期性指標を推定した．分析フレームは 1ms ごととし，各フレーム におけるスペクトル傾斜，非周期性成分を算出した．スペクトル傾斜は，4 種類の帯域（0–3000Hz, 0–6000Hz, 0–9000Hz, 0–22050Hz）を対象としており，以降は「スペクトル傾斜 （0–3000Hz)」のように括弧を用いて帯域の違いを示す．スペクトル傾斜及び非周期性成分は，歌声の「迫力性」や「丁寧さ」「明るさ」といった印象推定に有効であることが，修士論文 [1] により示されている．スペクトル傾斜は，以下の式によって求めた．なお，f は周波数ビンの番号，B は周波数ビンの数を示しており，対象とする帯域によって B の値は変化する．

スペクトル傾斜 $T(t)$

$$T(t) = \frac{B\sum_{f=1}^{B} f \cdot S_2(f,t) - \sum_{f=1}^{B} f \sum_{f=1}^{B} S_2(f,t)}{F\sum_{f=1}^{B} f^2 - (\sum_{f=1}^{B} f)^2} \tag{4.7}$$

STRAIGHT [12] では，スペクトル包絡の全体のエネルギーに対する非周期性成分の割合を，0 から 1.0 の値で求めることが可能である．値が 1 に近づくほど，非周期性成分の割合が大きいことを示しており，歌声に含まれている非周期性成分の大きさを評価することが可能である．本稿では，スペクトル包絡全帯域における非周期性成分の値の総和を算出し，分析に用いた．上記の分析では，0.6 秒の音声全体における時刻ごとの特徴量を算出している．以降の相関分析では，発声が安定している区間を用いるために，0.1 秒から 0.5 秒の区間の平均値を算出し，特徴量として分析に用いることとする．

（4）相関分析

上で求めた色の特徴量，印象得点，音響特徴量を用い，相関分析を行った．色の特徴量と印象得点の相関係数を表 4.7 に示す．FDR 法を用いて多重比較を行った結果，p<.01 で有意であった値は **, p<.05 で有意であった値は*で示している．すべての色の特徴において，何らかの声質の印象と相関が認められた．特に，相関が多く見られたのは,「高い」「落ち着きのある」「明るい」など，活動性因子に関係する印象である．一方，緑，A*では他の色の特徴と異なり，活動性因子との顕著な相関が見られなかった．特に，緑は，唯一「歌声らしさ」「評価性因子」との相関が見られた特徴であり，これらの印象を可視化する際に重要な役割を果たすといえる．また，彩度は相関がある 8 種の印象のうち，5 種類は明度と共通しているものの,「落ち着きのある」「張りのある」「迫力性因子」の 3 種は明度では相関がでていない．このことから，声質の印象との適合性においては，明度と彩度で異なる役割を果たしていると考えられる．声質の印象を網羅するための 3 因子は，それぞれ「評価性因子」は緑,「活動性因子」は明度と B*,「迫力性因子」は彩度と A* ，という色の特徴を用いることで，ある程度独立に各因子の得点を色に反映させることができると考えられる．

色の特徴量と音響特徴量の相関係数を表 4.8 に示す．FDR 法を用いて多重比較を行った結果，p<.01 で有意であった値は **, p<.05 で有意であった値は*で示している．青緑，明度，A* 以外の色の特徴は，今回調査した音響特徴量のいずれかと相関が認められた．スペクトル重心は 10 種類中 7 種類，スペクトル傾斜（0–3000Hz），F1 は 6 種類の色の特徴と相関があり，音響特徴量により声質の違いを可視化する際に重要な特徴量であるとい

表 4.7: 声質の印象得点と色の特徴の相関係数

	赤	黄	緑	青緑	青	赤紫	明度	彩度	A*	B*
高い	0.72**	0.75**	-0.05	-0.37*	-0.79**	-0.78**	0.88**	0.69**	0.24	0.80**
男性的な	-0.67**	-0.62**	-0.15	0.38*	0.71**	0.73**	-0.95**	-0.52**	-0.15	-0.64**
かすれた	-0.09	-0.10	-0.20	0.26	0.14	0.08	-0.27	-0.18	-0.09	-0.23
落ち着きのある	-0.56**	-0.78**	0.75**	0.13	0.67**	0.47**	-0.22	-0.73**	-0.40*	-0.66**
迫力のある	-0.18	0.05	-0.48**	0.17	0.10	0.18	-0.53**	0.26	0.11	0.01
太い	-0.61**	-0.50**	-0.23	0.37*	0.61**	0.64**	-0.91**	-0.34*	-0.15	-0.56**
張りのある	0.16	0.41*	-0.50**	-0.07	-0.28	-0.17	-0.10	0.60**	0.26	0.39*
鼻声	0.03	-0.09	-0.31	0.44*	0.07	0.06	-0.28	-0.12	0.08	-0.15
歌声らしい	-0.10	-0.07	0.36*	-0.07	0.06	-0.07	0.21	-0.02	-0.21	0.04
響きのある	-0.13	-0.04	0.21	-0.12	0.07	0.02	0.04	0.07	-0.10	0.05
明るい	0.79**	0.86**	-0.31	-0.37*	-0.85**	-0.81**	0.71**	0.89**	0.36*	0.89**
柔らかい	0.09	-0.15	0.69**	-0.22	-0.00	-0.18	0.53**	-0.26	-0.23	-0.04
滑らかな	0.08	-0.02	0.47**	-0.27	-0.07	-0.14	0.43*	-0.01	-0.02	0.12
評価性	-0.03	0.00	0.35*	-0.22	-0.02	-0.08	0.26	0.06	-0.08	0.12
活動性	0.81**	0.84**	-0.13	-0.40*	-0.88**	-0.83**	0.92**	0.75**	0.30	0.85**
迫力性	0.26	0.53**	-0.74**	-0.02	-0.39*	-0.17	-0.11	0.60**	0.40*	0.43*

表 4.8: 声質の音響特徴量と色の特徴の相関係数

	赤	黄	緑	青緑	青	赤紫	明度	彩度	A*	B*
F1	0.49*	0.63**	-0.45	-0.17	-0.56*	-0.50*	0.33	0.62**	0.31	0.51*
F2	0.49*	0.60**	-0.31	-0.24	-0.55*	-0.52*	0.42	0.56*	0.11	0.44
スペクトル重心	0.57*	0.68**	-0.54*	-0.10	-0.62**	-0.55*	0.29	0.79**	0.38	0.63**
ロールオフ	0.28	0.31	-0.36	0.09	-0.27	-0.28	0.07	0.45	0.25	0.34
フラックス	0.33	0.40	-0.43	-0.12	-0.41	-0.13	0.13	0.45	0.25	0.32
lsm_3k	0.65**	0.70**	-0.46	-0.29	-0.67**	-0.58*	0.33	0.77**	0.33	0.68**
lsm_6k	0.44	0.43	-0.41	-0.09	-0.40	-0.38	0.07	0.54*	0.17	0.40
lsm_9k	0.36	0.32	-0.34	-0.06	-0.29	-0.30	-0.01	0.42	0.10	0.30
lsm_all	0.22	0.08	-0.10	-0.09	-0.10	-0.15	-0.06	0.15	-0.05	0.09
ap	-0.17	-0.44	0.55*	-0.02	0.29	0.21	0.13	-0.57*	-0.22	-0.36

える．F1 は黄，彩度など多くの特徴と相関が認められているが，フォルマント周波数は母音の違いによる影響を受ける．そのため，母音の違いによる影響と，声質の違いによる影響の関係について，今後詳細に検討する必要がある．

4.5 本章のまとめ

本章では「時間軸上の変化を表現するための時間長」を分析対象とし，ごく短い時間長から認知される歌声の特徴について考察を行った．その上で,「時間軸上の変化」を表現するために，主観評価実験を行い，歌声の特徴と色の対応関係を明らかにした．

歌声の声質と色との対応付けに関して一対比較評価実験を行い検証したところ，明度条件において，評価者の評価が一致しやすいことが明らかになった．また，全 28 データの

歌声において，色の 3 要素のうち最低 1 つ以上の条件で有意差が認められており，幅広い声質においていずれかの色の要素と対応づけることが可能だと考えられる．音高差による色の選択の差を検定したところ，色の 3 要素全てにおいて有意な差が見られたため，可視化の際には音高の影響を十分に配慮する必要があるといえる．色相の評価においては，評価者の一意性を算出した結果，評価者の違い，歌声の違いにより一意性の値が大きく異なっていた．また，サーストンの尺度値を算出した結果，音高変化による色の適合度の変化には声質の個人差の影響が大きいことが明らかとなった．L*a*b* 表色系における a*, b* に該当する指標を各歌声から算出したところ，high の歌声は似通った色が選択される傾向にあり，low, middle では b*軸における分散が大きいという結果が得られた．この結果は，声質が多様な色相と対応づけられる可能性を示唆している．感性的観点から声質の差異を幅広く捉えるため，13 対の表現語を用いて印象評価を行った結果，因子分析により「評価性」「活動性」「迫力性」という 3 因子が抽出された．これら 3 因子の累積寄与率は 0.889 であり，声質の印象を 3 因子で広く表現できていると言える．次に，声質の特徴がどのように色と関係しているかを明らかにするため，心理的特徴として声質の印象得点，物理的特徴として音響物理量，そして色の特徴のパラメータを用い，相関分析を行った. 結果,「活動性因子」に関連する多くの印象が，明度や b*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，緑は他の色とは異なる傾向が見られ，唯一「歌声らしさ」「評価性」との相関が見られた.3 因子はそれぞれ明度，彩度，a* 及び b* のいずれかと相関があるため，色の 3 要素を用いることで，3 因子に基づく，幅広い声質の印象を表現可能であると考えられる．音響特徴量においては，スペクトル重心及びスペクトル傾斜（0–3000Hz），F1 が多くの色の特徴との相関を認められた．

第5章　結論

5.1　本研究のまとめ

本研究は「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」を提案することを目指し，以下の２種類の検討を行った．

- 長時間の歌声における特徴の評価方法の検討（第２章）
  アマチュア歌唱者が理解しやすい情報として「印象」に着目し，ある程度の長さがある歌声を対象に，印象の推定モデルを再構築した．その結果，先行研究と比較して推定精度を向上させることができ，3 因子のモデルの決定係数について， 0.958（迫力性），0.551（丁寧さ），0.643（明るさ）という結果を得た．モデル構築に用いた楽曲とは異なる楽曲を用いて推定精度を確認したところ,「音域が狭い楽曲」「印象があまり感じられない歌声」においては，評価者同士の相関が低くなる傾向があり，推定精度自体も低い結果となった．このような例外はあるものの，評価者の評価が一致しやすい歌声においては，十分に印象を推定できている，という結果が得られた．

- 短時間の歌声における特徴の評価方法の検討（第３章）
  時間軸上の変化を表現するために，短時間の歌声における「印象」について，考察を行った．声質表現語を用いた印象評定実験により，歌声の声質における「評価性」「迫力性」「活動性」という 3 因子が抽出された．また，声質と色の対応関係を一対比較評価実験により調査した結果，様々な対応関係が明らかとなった．時間軸上の変化を表現する際には連続的な変化が可能である表現が望ましいと言えるが．色の特徴として，連続的に変化が可能な要素には,「明度」「彩度」「a*（赤み-緑み)」「b*（黄み-青み)」などが挙げられる．この中で「明度」「彩度」「b*」は声質の活動性と大きく関係していた．つまり，声質の活動性の情報を表現する際，活動性の値の大小に合わせて，色の明度や彩度，b*を連続的に対応させることが有効だと考えられる．

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，印象の推定を行い，特徴の把握につなげることは可能である．しかし，短時間の歌声を対象とした場合，結果として 3 因子それぞれを色の独立した要素に対応づけることはできなかった．そのため，色を用いた可視化は「活動性」の因子のみにとどめ,「評価性」「迫力性」について，今後は色とは異なる要素で可視化する方法を検討していく必要がある．

5.2 今後の展望

本節では,「歌唱支援に向けた展望」「可視化に向けた展望」について述べる.

5.2.1 歌唱支援に向けた展望

本研究では，歌声の特徴を，素人にも分かりやすい形で提示する手法について明らかにした．歌唱練習をする際，まずは自身の歌声の特徴を把握する必要があるためである．しかし，素人の場合，自身の歌声の特徴を把握したとしても，歌唱に関しての具体的な練習方法が分からない場合が多い．また,「印象」という側面から歌声を捉えた際に，どのような練習がどのような印象に結びつくかは，プロでさえも明確には指示できないと考えられる．

本研究第 2 章では，長時間の歌声における印象と，音響特徴量の関係について調査を行った．その際，音響特徴量に対して主成分分析を行うことで，歌声に現れる様々な歌い方の特徴を，主成分として扱うことができた．この主成分に関してより詳細な検討を行うことで，歌声に現れる様々な歌い方の特徴と，様々な印象を対応づけることができ，歌唱支援に活かすことができると考えられる．

例えば,「3.4.7 主成分得点ごとの考察」では，各主成分にどのような音響特徴量が含まれていたかを踏まえ，主成分ごとの特徴を述べた．この考察を，歌唱支援を目指してより抽象度の高い捉え方をすると，表 5.1 のように解釈することが可能である．

各主成分ごとに，主成分得点が高かった歌声と低かった歌声を比較し，歌声の特徴を記載している．その上で，その主成分が歌い方においてどのような役割を持っているか,「歌い方の要素」として記載した．

第 1 主成分では,「一つ一つの音符でそれぞれ表現がされている」か「フレーズというまとまりの中の一部分として音符が表現されている」か，という違いが見られ，音符に対する捉え方の違いが現れていた．一つ一つの音符を意識して歌うか，フレーズという大きなまとまりとして意識して歌うかによって，違いが生じると考えられる．

第 2 主成分では,「3.4.7 主成分得点ごとの考察」でも述べたように,「合唱経験者か否か」という違いが明白に表れていた．合唱に関する技術を習得することで，違いを表現できるであろう．

第 3 主成分では，どの程度楽譜に忠実に歌っているか，という違いが見られた．目標音高よりも低めの音高から入り，少し遅れて音高を上げる「しゃくり」といった歌唱表現や，フレーズの中での音量変化に緩急をつけている歌唱表現などが見られたため,「表現豊かな

表 5.1: 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	得点が低い歌声の特徴
1	各音符の捉え方	音符ごとの F0 の変動が大きい	フレーズで一続きの流れになっている
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）	地声に近い歌唱
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	楽譜に忠実に，まっすぐ歌っている
4	歌声らしさ	音の立ち上がりが早く，耳に残る声	滑らかな立ち上がり，落ち着いた声
8	発音の丁寧さ	母音が聞き取りやすい発音，ビブラート	明瞭性の低い発音

歌声」か「棒読みのような歌声」か，という捉え方も可能である.「しゃくり」など，楽譜から逸脱する歌唱技術を習得することにより，違いを表現できるであろう．

第 4 主成分では，第 1 フォルマント（F1）の平均値が大きい，という特徴が見られていた．F1 の値が大きくなる条件として,「口唇の開口度合いが大きい」という条件の他にも,「声道長が短い」という条件が挙げられる．声道長の長さはおおよそ身長に比例するため，大人に比べて子供は F1 の値が大きくなることが知られている．この第 4 主成分の得点が高い歌声では,「子供らしい」と感じる歌声が多く，声道長のような発声機構そのものに由来する要素も，印象形成には大きく影響していると考えられる．声道長に由来する特徴のみがこの主成分に影響していた場合，歌唱者の意思で制御することは困難かもしれないが，幸いにもフォルマント以外の「音の立ち上がりの速さ」という特徴が大きく影響していた．つまり,「音の立ち上がりの速さ」を制御できるようになれば，この主成分の違いを表現できる，と考えられる．

第 8 主成分では，母音の明瞭性が大きく影響していた．同時に現れていた「ビブラートらしさ」について，発声機構などから直接関連づけることは難しい．しかし，この第 8 主成分は「うまさ」が高く評価された歌声と大きく関連しており，歌声の「うまさ」を高める過程で，母音の明瞭性，及びビブラートの技術を共に習得したとも考えられる．

このように，各主成分ごとに，様々な歌声の要素と対応づけることができる．データ数を増やし，より詳細な検討を行うことで，歌唱者が目標とする印象に必要な歌声の要素を推薦できるようになるであろう．その結果，歌唱指導など，様々な場面に応用することが可能になると考えられる．

5.2.2 可視化に向けた展望

本研究では，歌声の声質に着目し，色との対応関係を明らかにした．その結果，声質の印象においては「活動性因子」が最も色と対応させやすいことが明らかになったが，他の因子（迫力性，評価性）をどのように可視化すべきか，検討する必要がある．

色以外の可視化手段としては「図形」を用いた方法が挙げられる．図形の構成要素として，Oyama [40] は，Complexity, Regularity, Curvedness の 3 種を挙げている．また，人の知覚に関するテクスチャーの特徴には，Coaarseness, Contrast, Directionality, Line-Likeness, Regularity, Roughness の 6 種があることが，Tamura ら [41] によって明らかにされている．このように，図形を用いて何らかの情報を表現する手段は複数存在するが，時刻ごとの細かい変化を表現する際には，手段が限られてしまう．つまり，限られた表現手段の中で「迫力性」「評価性」を表現できる手段を検討する必要があるといえる．

図 5.1: 色と図形を用いた可視化例

例えば，図 5.1 では活動性を「色（黄みー青み）」，迫力性を「線の太さ」，評価性を「四角形のぼやけ度合い」で表現している．様々な手段を用いて可視化し，どの方法が最も情報を明確に伝えられるか検討した上で，可視化システムを実装することが望ましい．声質のような多次元の情報を時系列で表示できる手段が確立されれば，歌声や話声などの音声はもちろん，それ以外の分野の可視化にも応用できると考えられる．

謝辞

本稿の執筆にあたって多くの方々にご協力をいただきました.

学部，修士時代に引き続き，研究課題の設定や実験実施方法，本稿の執筆まで，手厚くご指導いただきました菊池英明先生に心より感謝を申し上げます.

また，学位論文 中間報告会にて大変有益な指摘をいただきました，齋藤美穂先生，修士時代に引き続き，学位審査の副査としてご指導くださった松居辰則先生，向後千春先生に心より感謝を申し上げます.

そして，論文執筆に際し，懇切丁寧にご指導いただきました産業技術総合研究所 メディアインタラクション研究グループの後藤真孝さま，中野倫靖さま，及びラボの皆様に心より感謝を申し上げます.

歌声収録のためにご協力いただいた歌唱者の皆さま，実験の実施に際しご協力いただいた皆さま，そして研究内容について様々なご指摘やアドバイスをいただきました菊池研究室の皆さまに心から感謝の気持ちと御礼を申し上げたく，謝辞にかえさせていただきます.

引用文献

[1] 金礪愛. 歌唱音声の音響的分析に基づく自動印象評価の有用性: 印象評価尺度の構築と印象推定システムの開発. 早稲田大学大学院人間科学研究科 修士論文, 2014.

[2] 中野倫靖, 後藤真孝, 平賀譲. 楽譜情報を用いない歌唱力自動評価手法. 情報処理学会論文誌, Vol. 48, No. 1, pp. 227–236, jan 2007.

[3] Wei-Ho Tsai and Hsin-Chieh Lee. Automatic evaluation of karaoke singing based on pitch, volume, and rhythm features. *IEEE Trans. on ASLP*, Vol. 20, No. 4, pp. 1233–1243, 2012.

[4] R. Daido, S. . Hahm, M. Ito, S. Makino, and A. Ito. A system for evaluating singing enthusiasm for karaoke. In *Proc. of ISMIR 2011*, pp. 31–36, 2011.

[5] G. M. Kotlyar and V. P. Morozov. Acoustical correlates of the emotional content of vocalized speech. *Sov.Phys.Acoust.*, Vol. 22, No. 3, pp. 208–211, may 1976.

[6] 谷口高士. 音楽作品の感情価測定尺度の作成および多面的感情状態尺度との関連の検討. 心理学研究, Vol. 65, No. 6, pp. 463–470, 1995.

[7] 平江遼, 西隆司. 感性に基づくクラシック音楽の分類. 日本音響学会誌, Vol. 64, No. 10, pp. 607–615, 2008.

[8] 木戸博, 粕谷英樹. 通常発話の声質に関連した日常表現語の抽出. 日本音響学会誌, Vol. 55, No. 6, pp. 405–411, jun 1999.

[9] 難波精一郎. 音色の定義を巡って. 日本音響学会誌, Vol. 49, No. 11, pp. 823–831, 1993.

[10] J. Cortina. What is coefficient alpha? an examination of theory and applications. *Journal of Applied Psychology*, Vol. 78, No. 1, pp. 98–104, 1993.

[11] 金礪愛, 中野倫靖, 後藤真孝, 菊池英明. 歌声の印象評価尺度の構築に基づく多様な印象の自動推定手法. 情報処理学会論文誌, Vol. 57, No. 5, pp. 1375–1388, may 2016.

[12] Kawahara Hideki, Masuda-Katsuse Ikuyo, and Alain de Cheveign e. Restructuring speech representations using a pitch-adaptive timefrequency smoothing and an instantaneous-frequency-based f0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication*, Vol. 27, No. 3-4, p. 187207, 1999.

[13] J. Sundberg. *The Science of the Singing Voice*. Northern Illinois University Press, 1987.

[14] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, Vol. 10, No. 5, pp. 293–302, July 2002.

[15] 池田操, 伊東一典. 音楽科学生と一般学生の歌声の音響分析と評価：シンガーズ・フォルマントを指標として. 上越教育大学研究紀要, Vol. 19, No. 2, pp. 493–509, 2000.

[16] エリクソンドナ, 齊藤毅, 細川久美子, 岸本宏子, 羽石英里. 女声の「歌唱フォルマント」の音響学的研究：その 1, 第 29 巻, pp. 13–26. 昭和音楽大学, mar 2010.

[17] 平山健太郎, 伊藤克亘. ポピュラー歌唱における高音域の声区と発声状態の判別手法. 情報処理学会研究報告 音声言語情報処理, Vol. 2012-SLP-90, No. 16, pp. 1–6, jan 2012.

[18] 小島俊, 齋藤毅, 中野倫靖. 歌声における裏声と地声を識別するための音響特徴量の検討. 電子情報通信学会技術研究報告：信学技報, Vol. 112, No. 266, pp. 67–72, oct 2012.

[19] 田窪行則, 前川喜久雄, 窪薗晴夫, 本多清志, 白井克彦, 中川聖一. 岩波講座言語の科学. No. 2. 岩波書店, 1998.

[20] Nakano Tomoyasu, Goto Masataka, and Hiraga Yuzuru. Subjective evaluation of common singing skills using the rank ordering method. *Proc. of ICMPC2006*, pp. 1507–1512, 2006.

[21] 齋藤毅, 鵜木祐史, 赤木正人. 自然性の高い歌声合成のためのヴィブラート変調周波数の制御法の検討. 電子情報通信学会技術研究報告, Vol. 105, No. 291, pp. 13–18, sep 2005.

[22] 齋藤毅, 辻直也, 鵜木祐史, 赤木正人. 歌声らしさの知覚モデルに基づいた歌声特有の音響特徴量の分析. 日本音響学会誌, Vol. 64, No. 5, pp. 267–277, may 2008.

[23] 金礪愛, 菊池英明. 歌唱音声における声質の特徴と想起される色の関係. 日本感性工学会論文誌, Vol. 17, No. 1, pp. 109–118, 2018.

[24] John Laver. *The Phonetic Description of Voice Quality*. Cambridge University Press, 1980.

[25] 榊原健一. 歌声に於ける声質の生成機構. 音声研究, Vol. 7, No. 3, pp. 27–39, 2003.

[26] Harry. Hollien. On vocal registers. *Communication Sciences Laboratory Quarterly Report*, Vol. 10, No. 1, 1972.

[27] 松浦泰仁, 角谷亮祐, 秋庭祐貴, 菅沼佑太, 菊川裕也, 馬 場哲晃, 串山久美子. 声質ヴィジュアライザの提案. 情報処理学会 インタラクション 2012 論文集, pp. 451–456, 2012.

[28] 菅原衣織, 伊藤貴之. 倍音分析によるいい声作りの支援アプリに向けて. 電子情報通信学会技術研究報告, Vol. 114, No. 52, pp. 327–329, 2014.

[29] 矢島佳澄, 筧康明, 諏訪正樹. 発声のメタ認知促進システム いい声マイク の提案. 2011.

[30] 長田典子, 岩井大輔, 津田学, 和氣早苗, 井口征士. 音と色のノンバーバルマッピング: 色調保持者のマッピングとその応用. 電子情報通信学会論文誌, Vol. A86, No. 11, pp. 1219–1230, 2003.

[31] Mohammad Adeli, Jean Rouat, and Stephane Molotchnikoff. Audiovisual correspondence between musical timbre and visual shapes. *frontiers in HUMAN NEUROSCIENCE*, Vol. 8, No. 352, 2014.

[32] 赤井良行, 李昇姫. 音色からイメージされる色彩の寒暖と音色構造の関係. 日本感性工学会論文誌, Vol. 13, No. 1, pp. 221–228, 2014.

[33] Anja Moos, David Simmons, Julia Simner, and Rachel Smith. Color and texture associations in voice-induced synesthesia. *Frontiers in Psychology*, Vol. 4, No. 568, 2013.

[34] Magdalena Wrembel. On hearing colours: Cross-modal associations in vowel perception in a non-synaesthetic population. *Poznan Studies in Contemporary Linguistics*, Vol. 45, p. 581598, 2010.

[35] Anja Moos, Rachel Smith, Sam R. Miller, and David Simmons. Cross-modal associations in synaesthesia: Vowel colours in the ear of the beholder. *Perception*, Vol. 5, pp. 132–142, 2014.

[36] 金礪愛, 菊池英明. ポピュラー音楽のための歌唱音声評価尺度の構築. 日本音響学会研究発表会講演論文集, pp. 397–400, mar 2013.

[37] 大石康智, 後藤真孝, 伊藤克亘, 武田一哉. スペクトル包絡と基本周波数の時間変化を利用した歌声と朗読音声の識別. 情報処理学会論文誌, Vol. 47, No. 6, pp. 1822–1830, 2006.

[38] 若田忠之, 齋藤美穂. Pccs 表色系の ipad ディスプレイ上における rgb 値の視感測色. 日本色彩学会誌, Vol. 39, No. 5, pp. 101–104, 2015.

[39] P. Boersma and D Weenink. Praat: doing phonetics by computer [computer program],version 5. 4. 08, retrieved 1 april 2015. `http://www.praat.org/`.

[40] Oyama Tadasu, Miyano Hisao, and Yamada Hiroshi. Multidimensional scaling of computer-generated abstract forms. *New Developments in Psychometrics*, pp. 551–558, 2003.

[41] Tamaura Hideyuki, Mori Shunji, and Yamawaki Takashi. Textural features correponding to visual perception. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 8, No. 6, pp. 460–473, 1978.