

# 習熟度に関する音響特徴量に基づく 歌唱学習支援システムの評価

## Evaluation of the Effectiveness of a Singing Learning Support System based on Acoustic Features related to the Singing Skill

羽賀 翼<sup>1</sup> 香山 瑞恵<sup>2</sup> 池田 京子<sup>3</sup> 橋本 昌巳<sup>2</sup> 伊東 一典<sup>2</sup>

Tasuku Haga<sup>1</sup>, Mizue Kayama<sup>2</sup>, Kyoko Ikeda<sup>3</sup>, Masami Hashimoto<sup>2</sup>, Kazunori Itoh<sup>2</sup>

<sup>1</sup> 信州大学大学院理工学系研究科

<sup>1</sup> Division of Science & Technology Master's Program, Shinshu University

<sup>2</sup> 信州大学工学部

<sup>2</sup> Faculty of Engineering, Shinshu University,

<sup>3</sup> 信州大学教育学部

<sup>3</sup> Faculty of Education, Shinshu University

**Abstract:** The purpose of this study is to develop a singing learning support system based on the visualization of acoustic features of learner's singing voice. We especially focus on a peak sharpness of singer's formant (Q-factor) and a second formant ratio (SFR) as skill development related acoustic features. In our previous research, These two factors have tended to show the singer's skill level. In this paper, we propose a "Voice Ball" graph, which shows a relationship between the Q-factor and the SFR. By using our singing learning support system with the "Voice Ball" function in six weeks, our subjects are able to show the tendency of improvement of their singing skill.

## 1 はじめに

今日にいたるまでに、歌唱や歌声に関する研究は多数行われてきた[1-2]. これらの研究では、歌声らしさを示す属性や、習熟度に関連する物理量などの解明が進んでいる. また、音高や音量といった音響特徴量と人による歌声に対する心理的印象とに関係があることが分かり始めている[3]. そして、これらの研究成果を用いてシステムを構築し、個人では行うことができなかった合唱の練習[4]や人員不足の声楽分野における知識獲得[5]、音高の可視化を用いた歌唱力の向上[6]など、さまざまシステムが存在している. しかしながら、既存の歌唱の指導に関連のシステムの多くは、音響特徴量の中でも、音高に関連する特徴量に主眼を置いており、カラオケのように、その音高の一致度合いによる評価を行っているシステムである. また、システムを使用するに際して、信号処理の知識が必要になり、歌唱に関する知識だけでは、使用出来ないシステムも少なくない.

本研究の目的は、歌唱学習と歌唱指導の支援を行うシステムの開発である. 歌唱の技術向上には、歌唱を指導する指導者とのレッスンが有効である. し

かし、大学において、教員がすべての学生の相手をするのは難しく、学生自身による自主練習の比重が大きい. しかし、指導者とのレッスンと比べて自主練習では学習効率が低下する. 我々は学習効率の低下を引き起こす要因として、指導者による学生の歌声に対する評価と、改善のための指導が無いためと考えた. そこで、これらの要因のうち、本研究では評価の自動化を試みる. また、信号処理などの知識を必要とせず、誰にでも簡単に使え、分かりやすいフィードバックを行うシステムを目指す.

## 2 習熟度に関する音響特徴量

### 2.1 音響特徴量とは

音響特徴量とは、歌唱データを分析することにより出現する特徴量のことである. 音響特徴量は、大きく分け、音の大きさを示す音量、音の高さを示す音高、周波数成分の違いを示すと言われている音色、音の長さを示す音価、音のタイミングを示す律動の5つに分類される.

これらの音響特徴量の中でも、音色は習熟度との

関連が強いとされる。音色は音声をフーリエ変換して得られる各周波数ごとの成分(周波数スペクトル)として表すことができる。

## 2.2 本研究で用いる音響特徴量

本研究では、5 つある音響特徴量のうち、歌声の響きに関わる音色に加え、音の根幹を成す音高と歌声の情緒に関わる音量の3つの音響特徴量を取り上げる。しかし、本研究で対象にしようとしている学習者は、教育学部音楽科に所属しているため、音高と音量についてはすでに問題が無いと考えた上で、特に音色に注目する。具体的には、周波数スペクトルとその2.4kHzから4.0kHzでのピークであるSinger's formant[7-8]を取り上げる。Singer's formantはこれまで、ピークの有無でのみ評価をされてきた。そこで、本研究ではこれまでに整理してきたSinger's formantの特徴量であるQ値[9]とSinger's formantの区間に含まれるパワーの比率[10]を用いて、歌声の習熟度を定量的に評価することとする。

### 2.2.1 Singer's formant の Q 値測定

Singer's formant の Q 値(以下、Q 値)とは、Singer's formant 区間に含まれるピークの鋭さである。この値は周波数スペクトルの包絡線を線形予測法によって求め、2.4kHzから4.0kHzの間における包絡線の形状を評価することで求められる。算出手順を以下に示す(図1)。

- (1) 歌唱データの波形に対し、線形予測係数を12として前向き線形予測(LPC)を行う。
- (2) (1)で求められた包絡線において、2.4kHzから4.0kHzの間(図1下)での最大値を持つ地点  $f_0$  を探す。
- (3)  $f_0$  のとる値から 3dB 低い値を持つ地点  $f_1$  と  $f_2$  を  $f_0$  付近からそれぞれ探す。
- (4)  $Q = f_0 / (f_1 - f_2)$  で Q 値を求める。

ここで、(3)の手順の際に、 $f_1$  と  $f_2$  の条件を満たす地点を発見できなかった場合は、そのデータには Singer's formant が含まれていないと判定する。

### 2.2.2 Singer's formant 領域のパワー含有率

Singer's formant の区間に含まれるパワーの含有率(以下、SFR)とは、0kHzから4.0kHzの区間に含まれるパワーのうち、2.4kHzから4.0kHzの Singer's formant の区間(図1の四角で囲まれた区間)に含まれるパワーがどの程度の割合を占めているかという音響特徴量である。算出手順を以下に示す。

- (1) 歌唱データの波形に対しフーリエ変換を行い、周波数スペクトルを求める。
- (2) 0kHzから4.0kHzのパワーをの合計を  $P_s$  とし、2.4kHzから4.0kHzの区間のパワーの合計を  $P_t$  とする。

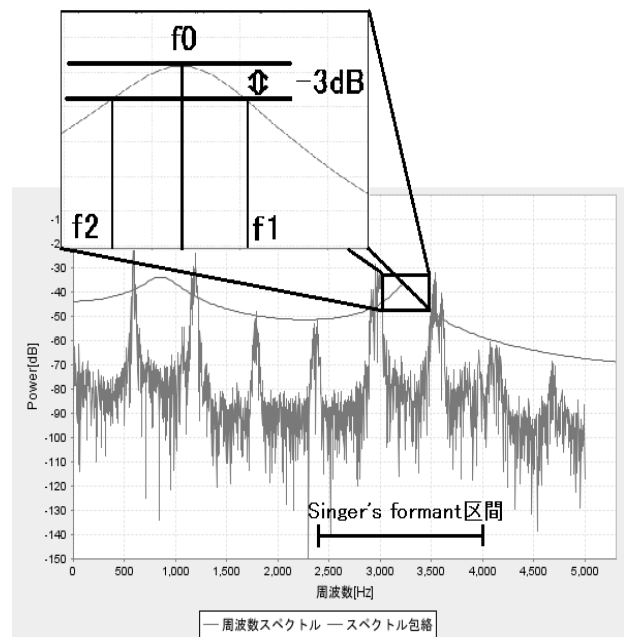


図1: Q 値の測定

- (3)  $SFR = P_t / P_s$  で SFR を求める。

## 2.3 可視化の方法

本研究で用いる音響特徴量のうち、習熟度の評価に用いる Q 値と SFR は数値として出力される。しかし、これらの学習前後での変化をこれらの数値として対象学習者に示した所、「わかりにくい」という意見を受けた。そこで、本研究では SFR と Q 値を同時に可視化し、理解を容易にすることを考えた。図2に可視化の例を示す。この図は縦軸に Q 値、横軸に SFR を取るグラフである。本研究では、これを Voice Ball と称する。各音声は、円状のボールとして表現される。さらに、ボールの直径の大きさに音量を割り当てることとした。

先行研究により、Q 値と SFR は熟練の歌唱者の声によく含まれる傾向があることが分かっている。そのため、双方の値が高いほど良い声と判断できる。Voice Ball では、ボールの位置がグラフ上で右上に行くほど良い声と判断できる。そのため、ボールの位置によってボールの色も変化するようにした。Q 値が大きい場合は青色が濃くなっていき、SFR が大きい場合は赤色が濃くなっていく。双方の値が大きい場合はボールが濃い紫色となる。

## 3 システムの概要

本研究で提案するシステムで想定する利用者は、教育学部音楽科の学生(以下、学習者)と、それらの学習者を指導する声楽発声指導者(以下、指導者)

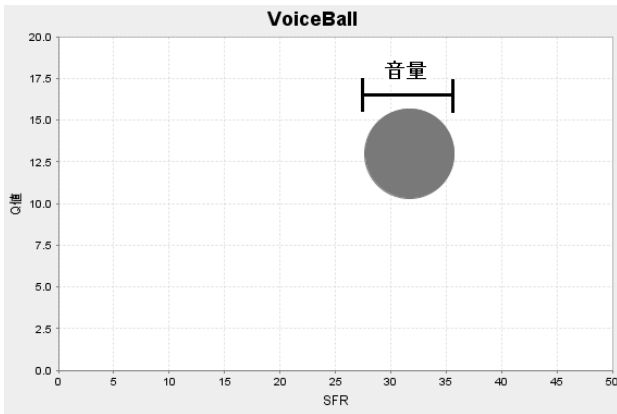


図 2 : Voice Ball について

である。このシステムの利用場面は、指導者と学習者が対面で行う個人レッスンと、指導者がいない場合の学習者一人での自主練習とする。使用方法としては、個人レッスン時には、指導者の出した評価を可視化することにより、学習者に対して理解しやすい形で自身の声の状態をフィードバックする。自主練習時には、学習者のみでは知覚できない声の状態を可視化し、フィードバックする。いずれの利用場面においても、可視化することによる歌声の状態や、過去のデータと比較することによる歌声の変化の様子を確認させることで、学習者の練習に対するモチベーションを維持することを狙う。

## 4 システムの機能

図 3 にシステムの全体図を示す。図の左側はユーザインターフェースの流れ、右側はデータの流れを示している。データの流れでは、四角がデータを示し、楕円が処理を示している。本システムの主な機能として、録音機能と再生機能がある。これらの機能は指導者の意見に基づき設計されている。以下、各機能について詳述する。

### 4.1 録音機能

録音機能では、歌声を録音しつつ、リアルタイムで音響特徴量をフィードバックする。本研究の対象学習者は、歌唱データに対する信号処理や音声処理、パソコンに関する知識を有していない場合が多い。その為、本システムのインターフェイスを簡略化し、以下の 3 つの項目を入力するだけで利用できるようにした (図 3 『録音前設定画面』)。

- ・ 男声か、女声か
- ・ 録音ファイル名
- ・ フィードバックの形式

ここでフィードバックする音響特徴量は、音高と音量の組み合わせ、もしくは SFR と Q 値と音量の組み合わせの 2 つの形式である。前者を選択した場合は、図 3 の『録音画面』が表示される。この画面には、音高を示すグラフと音量を示すグラフが表示され、リアルタイムにデータを表示しつつ、過去のデータは横へスクロールしていく。このフィードバック

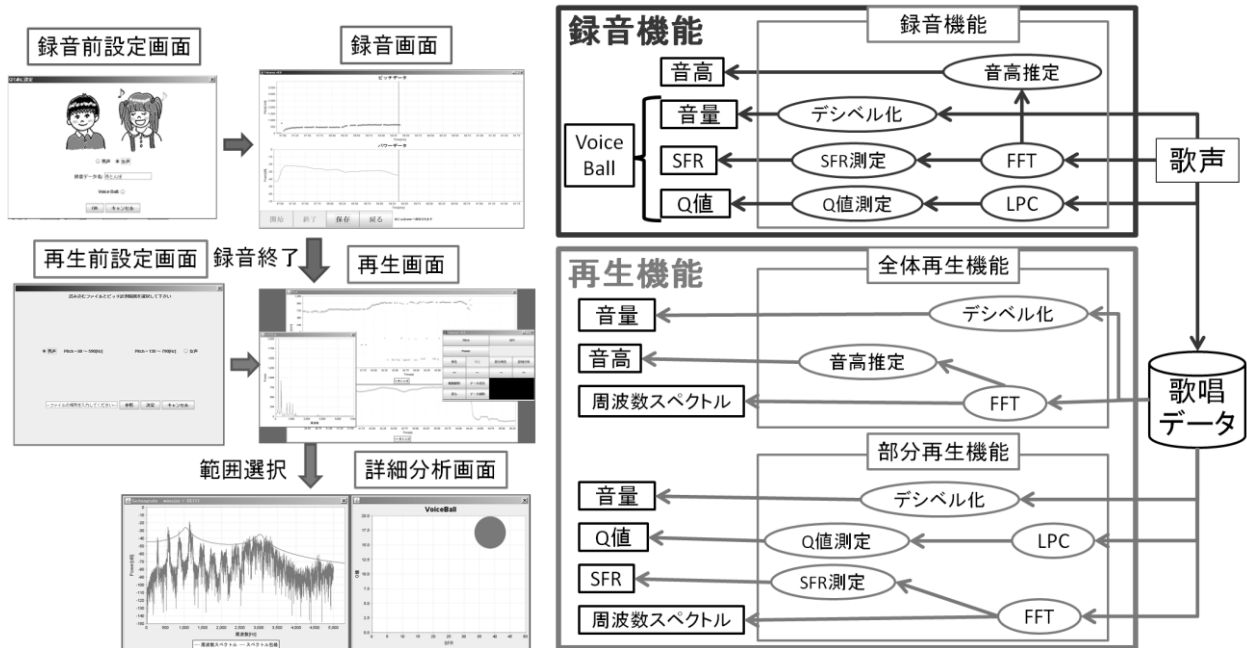


図 3 : システムの全体図

表 1: 録音機能時の音響特徴量の分析条件

|        |  |
|--------|--|
| フレーム長  | 2048(46ms 相当)                          |
| 窓関数    | ハミング窓                                  |
| 検査回数   | 音高 + 音量時 40 回/秒<br>Voice Ball 時 20 回/秒 |
| 音高推定範囲 | 男声 60Hz - 590Hz<br>女声 130Hz - 790Hz    |

表 2: 録音時の形式

|           |          |
|-----------|----------|
| 出力形式      | WAV ファイル |
| サンプリング周波数 | 44.1kHz  |
| 量子化ビット数   | 16bit    |
| チャンネル数    | 1 (モノラル) |
| 最大時間      | 300 秒    |

クについては、歌唱指導者より、「録音中は歌唱することに専念させるために必要以上のフィードバックを行わないようにした方がよい」との意見を受けて作成された。後者の形式を選択した場合は、SFR と Q 値と音量を示す Voice Ball が表示される。Voice Ball はリアルタイムにその瞬間の特徴量に基づくボールを表示する。このフィードバックについては、自身の歌声の音色に関する情報を可視化したものである。この形式を用いることで、適切な歌い方を試行錯誤し、良い評価となる歌い方を探すことも可能である。

ここでの音響特徴量の分析は表 1 の条件で行われる。このうち、音高推定範囲は選択した声の種類（男声か女声）で変化する。録音時間の上限（300 秒）に達するか、録音停止ボタンを押すことで録音が終了する。録音が終了すると、データを保存するかどうかの選択があり、保存する場合は表 2 の形式で保存される。保存を選んだ場合は、保存が終了した後、再生機能へ自動的に移動する。

## 4.2 再生機能

ここでは、本システムの録音機能を用いて保存された音声聞き直したり、音響特徴量を見直したりする。録音機能を利用した後で音声を保存するか、保存してあるデータを選択し、男声か女声かを選択する（図 3『再生前設定』）ことで利用できる。この機能には、保存した歌唱データの全体を再生対象とする全体再生と、全体再生中に選択された特定区間のみを再生する部分再生がある。

### 4.2.1 全体再生機能

この機能では、保存された歌唱データを扱うため、録音機能よりも精度の高い音響特徴量分析を行う。具体的には、フレーム長は録音時の 2 倍となる 4096 になり、検査回数も 2 倍の 80 回/秒となる。この機能でフィードバックを行う音響特徴量は音高と音量と周波数スペクトルである。これらの音響特徴量の

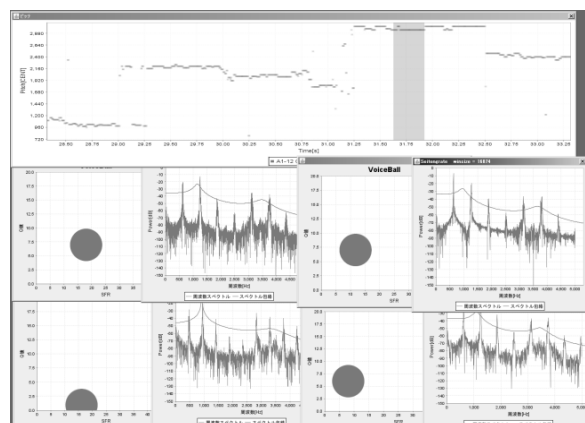


図 4: 部分分析の複数使用

グラフの例を図 3 の『再生画面』に示す。3 種のグラフはそれぞれ独立しており、大きさや位置が自由に変えられる。そのため、利用者が見たい情報を選択できる。また、同時に二つの歌唱データを読み込むことできる。そのため、自身の過去の歌唱データや他者の歌唱データとの比較が可能である。

### 4.2.2 部分再生機能

この機能では全体再生の対象データにおける特定区間を再生する。特定区間の指定は全体再生機能における音高グラフに対してマウスドラッグで行う。その後、図 3『詳細分析画面』に示す通り、周波数スペクトルと周波数スペクトル包絡、Voice Ball の 2 種のグラフが表示される。これらのグラフは図 3『再生画面』に対してポップアップされる。さらに、複数の特定区間に対するグラフ表示が可能である。4 区間に対してグラフを表示させた例を図 4 に示す。この機能を利用することで、最大で 2 つの歌唱データに対して複数個所の Voice Ball を表示し、比較することが可能である。

また、部分再生機能では、全体再生機能よりもさらに精度の高い音響特徴量分析を行う。具体的には、フレーム長は選択した区間の長さに応じて、最大 65536 までの間で動的に決められる。

## 5 評価実験

本システムが学習に及ぼす影響と対象学習者の使用感を調べるため、評価実験をおこなった。

### 5.1 実験条件

被験者は信州大学教育学部音楽科の大学 2 年生 6 名である。6 名は同じ指導教員のもとで学んでいる。その半数の 3 名に対し、本システムをインストールしたノートパソコンを貸し出し、週に一度以上の頻度で本システムを用いた自主練習を行うよう指示し

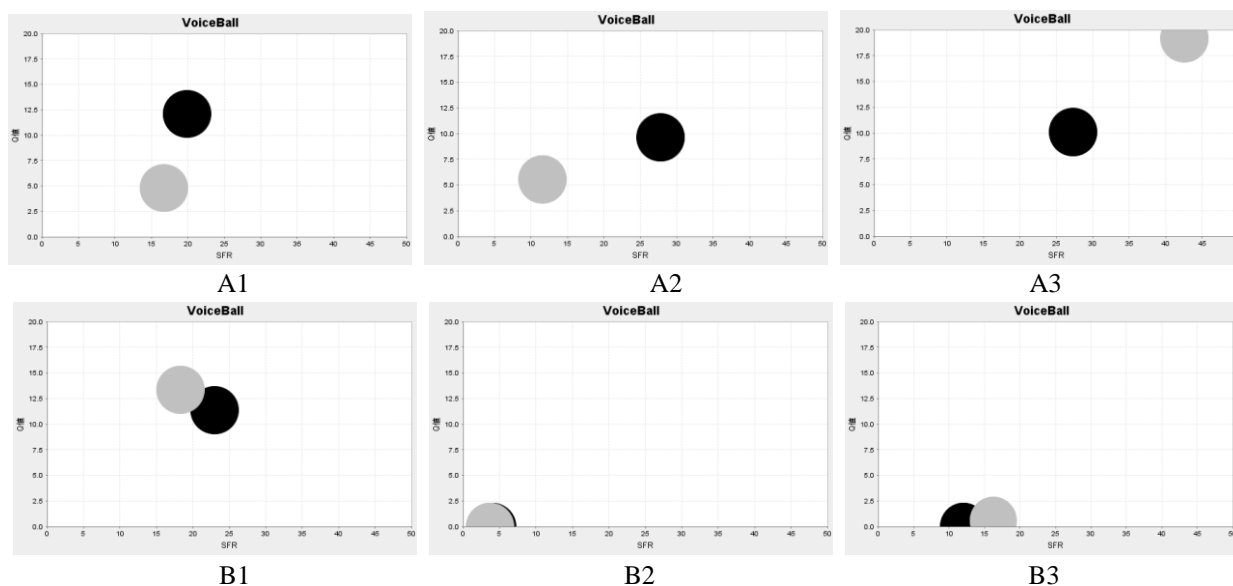


図 5：実験結果

た。その際、歌唱データの録音と以前に録音したデータとの比較を義務付けた。

実験期間は 2014 年 11 月第 2 週から 12 月第 4 週までの 6 週間である。音響特徴量を比較するための課題曲は指導者の意見を参考に『ふるさと』（作詞：高野辰之，作曲：岡野貞一）を使用した。この曲の内、歌い出しの部分から「うさぎおいし かのやま」の 4 小節の歌唱データを録音し、「ま」の母音区間を分析した。本実験で分析対象とした音響特徴量は、Q 値と SFR である。

## 5.2 実験結果

ここでは、システム利用者 3 名を A 群 (A1, A2, A3)、非利用者を B 群 (B1, B2, B3) と称する。

### 5.2.1 音響特徴量の変化

A 群と B 群の同一曲、同一カ所の歌唱データに対する Q 値と SFR の変化を実験前後で比較することで、本システムの学習効果を確認する。

各被験者の Voice Ball を図 5 に示す。色が薄いボールは実験前のデータであり、色が濃いボールは実験後のデータである。結果は Q 値と SFR のみに注目し、ここでのボールの直径は各被験者の歌唱データの音量によらず一定とした。

A 群における実験前後での変化では、A1 と A2 は Voice Ball 上で、右上への変化、すなわち、Q 値と SFR 共に増加した。一方、A3 については Q 値と SFR が共に減少し、左下へ変化した。また、B 群において、B1 は SFR が減少し、Q 値が増加したため、右下へ変化した。B2, B3 に関しては、あまり変化が見られなかった。

表 3：使用頻度

| 被験者 | 使用頻度    |
|-----|---------|
| A1  | 週 1 回   |
| A2  | 月 2~3 回 |
| A3  | 週 2 回   |

### 5.2.2 被験者からの意見

実験終了時の録音前に、A 群の被験者に対して質問調査をおこなった。ここでは、実験期間内における本システムの使用頻度と本システムに対する自由意見を記述させた。

表 3 に A 群の被験者の使用頻度を示す。使用頻度については個人差が見られた。平均すると週一回以上の使用頻度であった。

本システムに対する質問調査における自由意見の回答を以下に示す。

- 自身の声が良いかそうではないかが可視化されるのが面白かった。
- 自分の理想の声を目指してトレーニングしているが、このシステムでの Voice Ball が高評価をする声を目指そうとしてしまっている自分がいた。
- 同一の Voice Ball 上で異なる歌唱データの比較ができれば良かった。
- 結果からどのように自分の歌を改善したらいいかが分かるとより練習しやすかった。

3 名の被験者における本システムの使用感は概ね好評であった。

## 5.3 考察

Voice Ball は Q 値 15 と SFR20 を閾値として 4 つの領域に分けて歌唱データを評価することができる [11]. 初学者の歌唱データは左下に, 熟練者の歌唱データは右上にプロットされる傾向がある. また, 歌唱データはより右側またはより上側に変化することが上達を示すと考えられる. これらの知見に基づき, 今回の実験結果を考察する.

A 群において, A1 と A2 には上達が確認できる. A3 は実験前の歌唱データにおいて, すでに右上の領域に位置していたことから, すでに良い声であったとも考えられる. 実験後で右下の領域に変化した, ここでは習熟に変化はなかったと判断する.

一方, B 群においては, B1 は実験前後で Q 値と SFR の閾値付近に位置しており, 良い声に近い状態で, 変化していないと判断する. B2 と B3 は実験前後で左下の領域に位置しており, 変化がない.

これらのことから, A 群では 2 名が上達し, 1 名が変化なしであり, B 群では 3 名が変化なしとなり, 自主練習における本システムの利用は学習効果を示す可能性があるといえる.

本実験では, 被験者 6 名の内, 希望者を A 群とした. 図 5 の実験前のデータ (薄いボール) から, 実験開始前の時点で, A 群と B 群の Q 値と SFR に差があった. Q 値と SFR が高めの被験者, すなわち, 練習に対して積極的な被験者がシステムの利用を希望した可能性がある.

また, 同時に行った質問調査において, 歌唱データを評価するだけでなく, 改善方法を提示するなどの, 指導機能を求める声もあった. 本システムではこれまで, 歌声の可視化による評価を主体としていた. 今後は, 歌声を改善するための情報を提示する指導機能の実現を意識する必要があると考える.

## 6 おわりに

本研究では, 習熟度に関係する音響特徴量を可視化することで, 指導者に代わって歌唱データを評価し, 歌唱の自主練習における学習効率の向上を目指した. 特に注目した音響特徴量は音色に関係する Q 値と SFR であり, 本研究では, これらの数値を学習者に分かりやすくフィードバックするために Voice Ball を提案した. そして, このグラフを可視化する機能を実装した歌唱学習支援システムを構築した.

さらに, システムの利用が学習に与える影響を調べるために, 教育学部音楽科の学生を被験者として, 評価実験をおこなった. その結果, 自主練習における本システムの利用は学習効果を示す可能性がある

ことを示した.

今後は, 本システムの学習効果をより明確にするために, 実験前の音響特徴量に差がない 2 つの被験者群を対象に評価実験をおこなう. また, Q 値あるいは SFR が増加した歌唱データに対する指導者の主観的な評価を得て, 本システムによるフィードバックの妥当性を検証する. さらに, 指導者による指導と音響特徴量の変化の関連に基づいて, 本システムに指導機能を実装し, 自主練習のさらなる効率化を目指す.

## 参考文献

- [1] 辻直也他: 歌声らしさの要因とそれに関連する音響特徴量の検討, 日本音響学会聴覚研究会資料 H-2004-8, Vol.34, No.1, pp.41-46 (2004).
- [2] 中野倫靖他: 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 (2007).
- [3] 斎藤毅他: SingBySpeaking: 歌声知覚に重要な音響特徴を制御して話声を歌声に変換するシステム, 情報処理学会研究報告音楽情報科学研究会, Vol.2008, No.12, pp.25-32 (2008).
- [4] 香山瑞恵他: 指導者知識に基づく合唱学習支援システムの構築とその評価, 情報処理学会論文誌, Vol.51, No.2, pp.365-379 (2010).
- [5] 三浦雅展他: 和声学学習のためのバス課題実施システムのユーザインタフェースとその評価, 信学技報 ET, Vol.100, No.113, pp.33-40 (2000).
- [6] 中野倫靖他: MiruSinger: 歌を「歌って/聴いて/描いて」, 情報処理学会インタラクシオン 2007 論文集, pp.195-196 (2007).
- [7] J.Sundberg: Articulatory interpretation of the 'singing formant', J.Acoust.Am, Vol.55, No.4, pp.838-844 (1974).
- [8] 斎藤毅他: 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.405-417 (2008).
- [9] 山邊大貴他: 歌声の心理的印象と音響特徴量との対応付けによる歌唱の熟練度評価に関する基礎的検討, 信学技報 EA, Vol.112, No.266, pp.61-66 (2012).
- [10] 佐久間雄輝他: 歌声の音響特徴量の比較に基づく歌唱スキル習熟度の考察, 人工知能学会全国大会講演論文集, 1M4-OS-05a-1 (2014).
- [11] 佐久間雄輝他: 歌声の習熟度に関連する周波数特性からみた音響特徴量, 信学技報, Vol.114, No.441, pp.45-50 (2015).