

## 歌声の習熟度に関連する音響特徴量の母音別分布

山下泰樹<sup>†1</sup> 香山瑞恵<sup>†2</sup> 池田京子<sup>†3</sup> 吉田祥<sup>†4</sup> 平林花菜<sup>†4</sup>  
伊東一典<sup>†2</sup> 浅沼和志<sup>†5</sup>

**概要:** 歌声の習熟度に関連する音響量としてこれまで SFR や Q 値 (2 凸) を提案してきた。これらは Singer's formant 相当の周波数領域のパワーの比, ピークの鋭さ (高さ) を表す量である。これらを用いて, 主に歌声中の/a/を対象として SFR-Q 平面上の分布具合でその有効性を検討してきた。本報告では, /i,u,e,o/についても習熟度によってどのような分布の変化傾向があるかを確認した。習熟度の異なる母音を比較した結果, /a,o,u/については既報の/a/同様, 習熟度に伴い両軸の値の増大する傾向が見られたが, /i,e/は必ずしも値が増大しない様子が確認された。

**キーワード:** 歌声, 歌唱評価指標, 母音, Singer's formant

## Vowel characteristics of singing acoustic features related to skill-development of singing voice

Yasuki YAMASHITA<sup>†1</sup> Mizue KAYAMA<sup>†2</sup> Kyoko IKEDA<sup>†3</sup>  
Sho YOSHIDA<sup>†4</sup> Kana HIRABAYASHI<sup>†4</sup> Kazunori ITOH<sup>†2</sup>  
Kazushi ASANUMA<sup>†5</sup>

**Abstract:** We have proposed acoustic features related to singing skill: SFR (Second Formant Ratio), Q (sharpness of singer's formant), and two convex (height of singer's formant). Each of them are thought to be express frequency domain rates of power, sharpness of peak and height of peak of singer's formant. Using these acoustic features, the effectiveness has been examined on SFR-Q plane distribution focused on vowel /a/ of singing voice. In this paper, other vowels /i, u, e, o/ has examined on SFR-Q plane distribution. Comparison was done with SFR-Q plane distributions of different singing skill vowels. In the case /a,o,u/, both axis value are increased with singing skill, but in the case /i, e/, values are not always increased.

**Keywords:** Singing voice, Metrics for Singing Voice, Vowel, Singer's formant

### 1. はじめに

歌声については西洋オペラの分析を始めとして「良い声」の条件や「歌声らしさ」の検討が様々になされてきている[1,2]。それら検討をとおして, Singer's formant やビブラートの存在はよく知られているところである[3]。しかし, これらは主にプロの歌声に基づいているため, 訓練途中にある人のそれらの存在や程度は明確ではない。

一方, 最近身近なカラオケの採点システムなどを含めた, 歌唱の評価・訓練支援システムでは, 主に楽譜情報に基づいて音高 (ピッチ) や音価 (音の長さ) の適切さ, ビブラートに相当する揺れなどを評価するもので, 楽譜情報に無い声質までは含まれていないと考えられる。これは, 不特定の利用者を想定しているため, 一般の人が外しやすい音高や音価に着目するのは自然である。しかし, オペラ歌

手などを目指した歌唱訓練を行う場合は, 「響き」のある声を1つの目標とするため, 音高だけではなく声質まで含めた評価・訓練支援が求められる。

本研究の目的は, 声楽を学ぶ人を対象とした歌唱訓練の支援環境・ツールを整えることである。これまでのところ, 評価指標として Singer's formant に関連する音響量である SFR, Q 値, 2 凸を提案し[4], 習熟度の異なる複数もしくは同一人物の同一歌唱中の一部母音 (主に/a/) を用いてその有効性や訓練による変化を追ってきた[5-7]。しかし, 現在対象としている分析区間は曲中の印象的部分ではあるが数秒レベルの短いもので, これを以て一曲全体の評価とするのは少々無理がある。反面, これまでの経験から, 一曲全体で常に Singer's formant が存在しているものでもないため, やみくもに分析範囲を広げることが良いともいえない。そこで, 対象範囲を広げる一つの方策の検討として, 扱う母音の種類を増やすことが考えられる。男性オペラ歌手の母音別フォルマント周波数の観察・検討の例[8]は有るが, Singer's formant の程度や本研究で提案している評価指標でどのようになるかは確認が必要である。

以上を踏まえ本稿では, 評価指標上で習熟度の違いにより日本語五母音にどのような特徴があるかを検討した。

†1 長野県工科短期大学校  
Nagano Prefectural Institute of Technology

†2 信州大学工学部  
Faculty of Engineering, Shinshu University

†3 信州大学教育学部  
Faculty of Education, Shinshu University

†4 信州大学大学院  
Graduate School of Shinshu University

†5 長野工業高等専門学校  
National Institute of Technology, Nagano College

## 2. 本研究で用いる評価指標

分析条件は表1のようである。現時点では標本化周波数44.1kHzで音を取り込み、FFTが2048ポイント、LPC分析次数を12次とした。プリエンファシスはかけていない。

表1 音響分析条件

標本化周波数	44.1kHz
量子化ビット数	16bit
フレーム長	約46ms (2048ポイント)
窓関数	Hanning窓
FFTポイント数	2048ポイント
LPC分析次数	12次

なお、LPC分析に用いる自己相関関数の求め方[9]は、実時間処理のための計算量削減も意識して、 $N-\tau$ で除算する次式を用いた。

$$r(\tau) = \frac{1}{N-\tau} \sum_{k=0}^{N-1-\tau} x_k x_{k+\tau} \quad (\tau = 0, 1, \dots, \tau_{\max}) \quad (1)$$

また本稿では、各指標では値のばらつき、偏りが大きい場合、各話者（歌手）各母音の指標の代表値は中央値を用いた。

### 2.1 SFR

歌声の Singer's formant 相当 (2.4~4 kHz) の領域に含まれる倍音を表現するために、FFTスペクトルから求まるRMS値の次式の比をSFRとする。

$$SFR = \frac{2.4\sim 4\text{kHzのRMS値の合計}}{4\text{kHzまでのRMS値の合計}} \times 100 \quad (2)$$

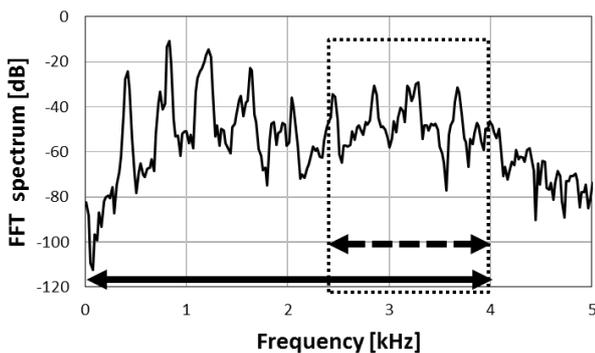


図1 SFRの定義

この指標は、値が大きいと Singer's formant の領域に倍音のエネルギーが相対的に集中していることを示す。

### 2.2 Q値

Singer's formant 相当のピークの鋭さをQ値とする。LPCスペクトル包絡の2.4~4 kHzの領域の最大ピークにおける周波数を  $f_a$ 、ピークの両側に-3dBとなる周波数を  $f_b, f_c$  ( $f_b > f_c$ ) とし、Q値を次式で算出する。Q値が大きいほど、

ピークが鋭く、エネルギーが狭い周波数範囲に集中していることを示す。

$$Q = \frac{f_a}{f_b - f_c} \quad (3)$$

なお、-3dBとなる周波数が両側で求まらない場合は、Q値は算出してない。また、領域内で複数のピークが出るケースを減らすため、LPC分析次数は通常の話声分析より低くした。

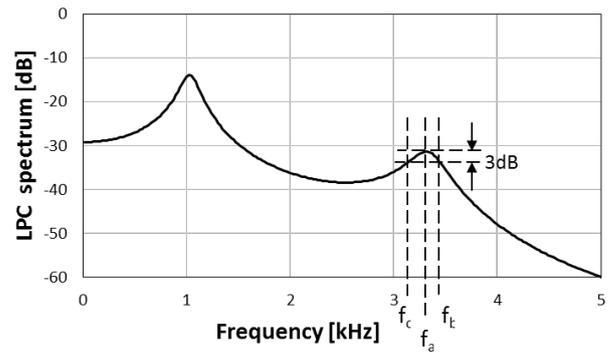


図2 Q値の定義

### 2.3 2凸

声楽初學者の歌声では、Singer's formant 相当のピークがQ値を得るほどには鋭くならないことがある[6]。本研究では初學者も対象とするため、Q値に代わる指標として2凸を定義して用いてもいる。

2凸は、図3のように Singer's formant 相当のピークと、その直前にある谷との差とする。

これまでの分析結果から、Q値のわずかな増大に対して2凸大きく増大するような、おおむね対数的な対応関係があることが分かっている[10]。この指標により、初學者の状態の変化が観察しやすくなることが期待される。

本稿の検討では専らQ値を用いたが、2凸でも同様の結果である。

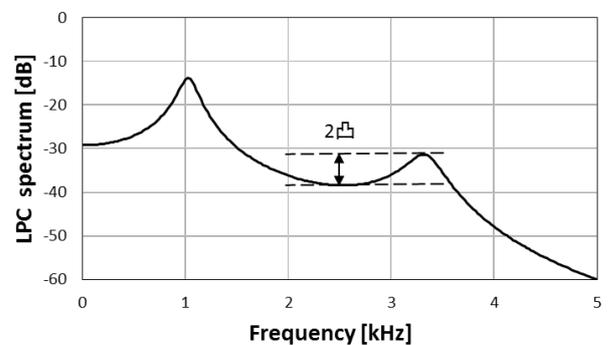


図3 2凸の定義

### 3. 対象とした音声データ

これまでの検討で主に用いてきた音声データは、歌唱訓練指導者の意見から「故郷」（作詞：高野辰之，作曲：岡野貞一）中，“かのやま”の/ma/の母音部分である。收音は反響しない静かな部屋で行った。使用レコーダはICR-PS605RM（SANYO）で、被験者にはレコーダから2m離れた正面立位で歌唱させた。音声はサンプリング周波数44.1 kHz、量子化ビット数16 bitで收音した。

今回、プロおよび学生の歌声の收音は従来のものに準じている。母音の比較においては図4のように，“かのやま”の/ma/を/ma,mi,mu,me,mo/に替えて歌ってもらった部分を対象とし、比較した。ただし、4年生とプロの一部はこの收音ができなかったため、これに準じた後述のデータを用いた。一方で、歌唱との対比で読上げ音声の母音を用いたが、これは多くの人の声で“かのやま”を入手するのは困難なので、単語読上げの音声データベースを利用した。詳細は各節で説明する。

なお、歌唱収録時に、ビブラートについて特に制約はかけていない。また、どの母音で歌うかは指示したが、結果として中性母音のように聞こえるものも含まれる。

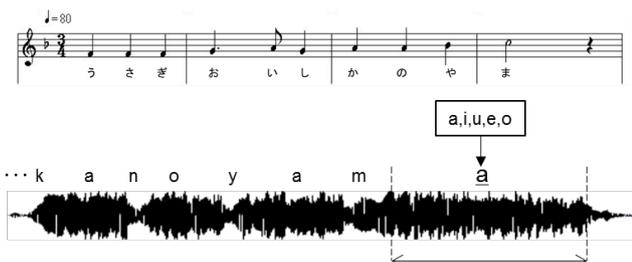


図4 “かのやま”の母音を替える様子

#### 3.1 プロ歌手の歌声データ

プロ歌手のデータは、プロ歌手3名（男性1名、女性3名）がへ長調で“かのやま”の/ma/を/ma,mi,mu,me,mo/に替えて歌った部分である。表2のように、プロ歌手はいずれも声楽のキャリアと指導暦を十分に有している。

表2 プロ歌手の情報

被験者	所属・演奏キャリア	指導歴	声種
A	東京二期会・27年	33年	ソプラノ
B	東京二期会・35年	29年	ソプラノ
C	東京二期会・26年	21年	メゾ・ソプラノ
D	藤原歌劇団・17年	16年	バリトン

#### 3.2 学生の歌声データ

学生の声は、初学者として信州大学教育学部音楽教育コース所属の大学2、3年生と、訓練を積んだ者として同声楽専攻の4年生の2グループとした。

同音楽教育コースでは、3年次後期から各専攻に分かれ

る。したがって2年生と3年生は実習内容から習熟度に大きな差は無いと見なした。一方、声楽を専攻とした4年生は個別授業で継続して訓練を行うため、習熟度が上がっている可能性が高いものとした。

分析対象とするデータは、2、3年生は、2年生5名（男性2名、女性3名）、3年生7名（男性1名女性6名）が“かのやま”の/ma/を/ma,mi,mu,me,mo/に替えて歌った部分とした。

4年生は6名（男性3名、女性3名）が/ma,mi,mu,me,mo/各々について、図5のように例えば下降型で「ソミド」の音型で各1秒程度伸ばした部分を用いた。

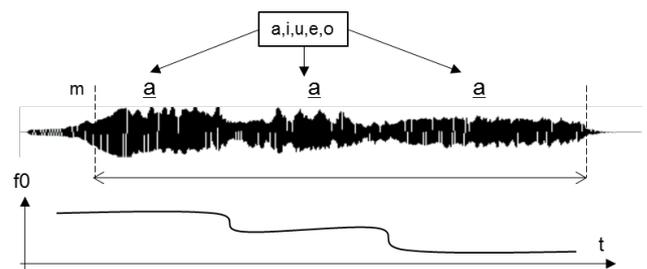


図5 /ma/の母音を替える様子

#### 3.3 単語読上げデータ

歌唱データに対して、単語読上げデータは歌っていない発声であり、そのため Singer's formant が出にくく、習熟度が低い発声と捉えられる。ただしこれには例外がある。アナウンサーの発声には Singer's formant 相当の特徴があるとされる[11]など、読上げ音声であっても発声訓練された人はこの限りではない。同様に我々の経験では、プロ歌手が読上げる場合、発声方法が歌唱訓練の影響を受けているケースがある。そこで、歌手の読み上げ音声は今回利用していない。代わりに、ATR 音声データベース C セットの男女各20名の音韻バランス単語（例えば、「敬う」）の発声から、/ma,mi,mu,me,mo, ba,,bi,bu,be,bo/の母音を切り出した。五母音それぞれ1話者10サンプル程度あるが、歌唱に比べて音韻の継続時間は短く、無声化したものも幾らか含まれている。

### 4. 結果と検討

図6から図15に男女別、母音別の結果を示す。図中、プロと2,3年生の○は、“かのやま”の/ma/を/ma,mi,mu,me,mo/に替えて歌ったもの、□はそれぞれ3章の各節で説明したものである。以下、図6から図15を参照しながら検討する。

#### 4.1 男女比較

まず、男女の比較をする。女声（図6から図10）と男声（図11から図15）を比べると、全般として男声の方が SFR、Q の値が大きい。この原因として声の高さによる違いや分

析手法, その設定による違いの可能性も考えられるが, 現時点では明確ではない. 結果的に, 男声の方が *Singer's formant* が強く現われる (もしくは観測されやすい) ことを反映していると考えられる.

#### 4.2 2,3 年生とプロとの間の変化

つぎに, ○同士すなわち 2,3 年とプロの間の変化量に着目すると, どの母音でも SFR, Q 値が増加している. これまでの検討から, /a/については習熟度が上がると SFR と Q 値の両方が増加し, SFR-Q グラフ上で右上に位置するようになることが分かっている. 今回, 他の母音においても, 習熟度が上がると SFR や Q の増加が観測される結果と考えられる.

しかし, 習熟度が訓練に伴い単調に増加するとは限らない. 訓練の過程において試行錯誤がなされ, それに伴い SFR, Q 値が増減することが予想される. また逆に, 習熟度が上がることにより SFR, Q 値の両方が同時に増加するとも限らない. 事実, 既報[12]において 2 年生から 3 年生までの 24 ヶ月で, Q 値が先に増加してから SFR が増加する現象も確認されている. そこで, つぎに, 4 年生の結果が 2,3 年生とプロとの間のどこに位置するかを見る.

#### 4.3 4 年生の位置

今回, 4 年生のデータは 2,3 年生やプロのものとは異なる. そこで, 若干の位置のずれを許容し, 2,3 年生から 4 年生の間で SFR もしくは Q 値のいずれかが増加したものを習熟度が上がった結果と仮定する. すると, 男声の /i,u/ (図 13 と図 14) を除いて, グラフ上, 4 年生が 2,3 年とプロの間に位置する. これは, 4 年生の習熟度が 2,3 年生とプロとの間であるという事実に対応する.

一方, 男声の /i,u/については SFR, Q 値とも 4 年生の結果が 2,3 年生のものより減少する結果になった. 他の母音の結果から 4 年生の習熟度が 2,3 年生に劣るとは考えにくいので, この結果がデータによる, もしくは, 4 年生が試行錯誤している結果と推測される.

#### 4.4 単語読上げの位置

ここでは, 単語の発声を, 習熟度が低い歌唱と捉えることにする.

グラフ上, 男声の /a,o,u/では, 2,3 年生より単語の位置が左下にある. 単語と 2,3 年生の間で SFR, Q 値が増加しているのは, 男声は *Singer's formant* が出やすいことで, 幾らか習熟度が上がったことが反映されていると考えられる.

それに対して女声の /a,o,u/では, 2,3 年生と単語の位置が近い. これは, 女声の 2,3 年生の習熟度が高くなっていないと解釈するよりは, 女声の方が *Singer's formant* が出にくいことにより, 習熟度の変化が表現できなかったと考えられる.

一方, 男女とも /i,e/は/a,o,u/とは様子が異なる. 単語の位置は 2,3 年生より右上にある. これは, /i,e/の音韻としての特徴によるものと考えられる. すなわち, /i,e/のフォルマ

ント周波数の位置が *Singer's formant* 相当の領域に入っていることに加え, 2,3 年生はどの母音でもほぼ似た位置にある (女声であれば SFR が 10, Q 値が 5 の周辺, 男声であれば SFR が 20, Q 値が 20 程度). この位置関係のため, 図 9, 図 10, 図 15 のように, 2,3 年生は単語よりも左下に位置している. /i,e/におけるこの位置関係は, 2,3 年生は訓練によって /i,e/だけ *Singer's formant* が出にくくなったことになる. しかし, その後の動きを見ると, 訓練を積むことで 4 年生, プロの状態になり, 右上にシフトして行く. このことから, /i,e/は特に初学者の段階では SFR, Q 値が大きいことが習熟度の高さに対応しないが, ある程度 (2,3 年生程度) になれば習熟度に伴い SFR, Q 値が増加する, と考えられる.

また, 単語とプロの間の変化量だけで比べれば, /i,e/は (男声の /i/を除いて) 結果的に他の母音よりも SFR, Q 値の増加量は小さいものに見える.

#### 4.5 母音ごとの検討

これまでの結果を整理すると, つぎようになる.

/a,o/は男女とも, 習熟度が上がるに伴って SFR, Q 値ともに増加する. よって, 習熟度向上の判断に利用しやすい.

/u/はおおむね/a,o/と似た傾向にあるが, 男声の/u/で 4 年生の位置が逆転する結果もあるので, 確認が必要となる.

/i,e/は習熟度が低くても SFR, Q 値が大きく出ることがあるので, 扱いに注意が必要となる. しかし, ある程度習熟度が上がれば, 他の母音同様, 習熟度が上がるに伴って SFR, Q 値が増加する. /i,e/は前舌母音であごの開きが狭めという共通性があることから, これに類する音韻は似た傾向を持つ可能性がある.

## 5. おわりに

“かのやま”の/ma/を/ma,mi,mu,me,mo/に替えて歌ったデータをベースに, 母音別に習熟度に伴いどのように分布が変化するか特徴を見てきた. 大別すると, 前舌母音の/i,e/と, それ以外の/a,o,u/で, 特に習熟度が低いときに SFR, Q 値の違いが見られた.

今回の歌唱データでは, 歌手の声種は複数あり, 収録した音高は得意な音域とは必ずしも一致していない. また, 男性プロ歌手は 1 名であるため, 結果の信頼性を高めるには追加のデータが必要と考える. しかし, 男女とも母音別に似た傾向の結果が得られていることから, 大まかな傾向は表れていると推測される.

今後は, 今回の結果の学習支援システムへの反映の仕方の考察や, 習熟度と SFR, Q 値の等評価指標との関係の詳細化について検討を行う.

**謝辞** 本研究の一部は科学研究費助成事業 (研究課題番号: 15K01022) として支援を受けた.

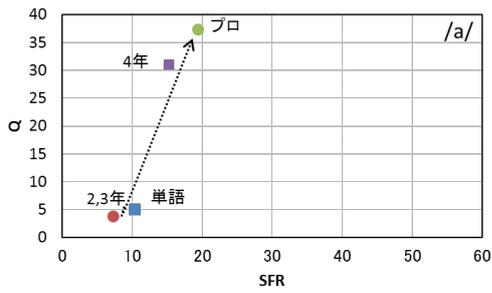


図 6 女声 母音/a/

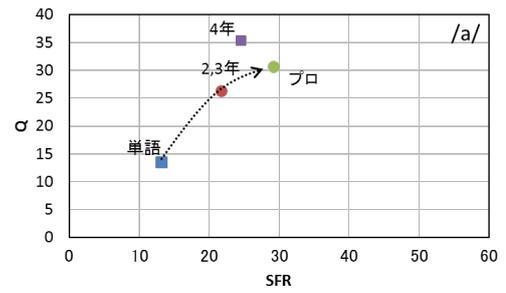


図 11 男声 母音/a/

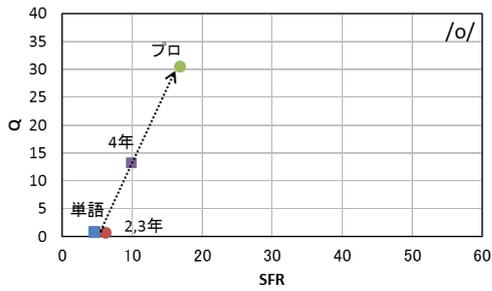


図 7 女声 母音/o/

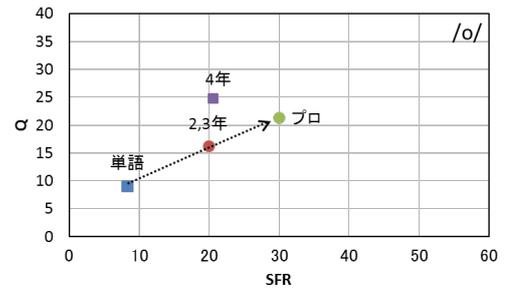


図 12 男声 母音/o/

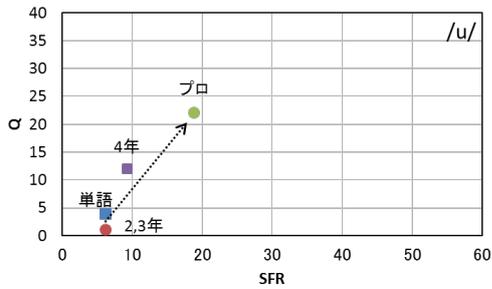


図 8 女声 母音/u/

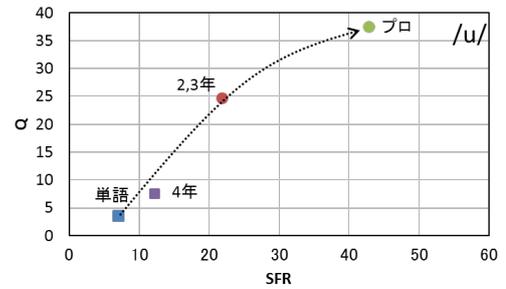


図 13 男声 母音/u/

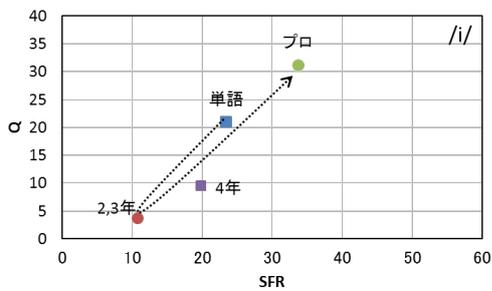


図 9 女声 母音/i/

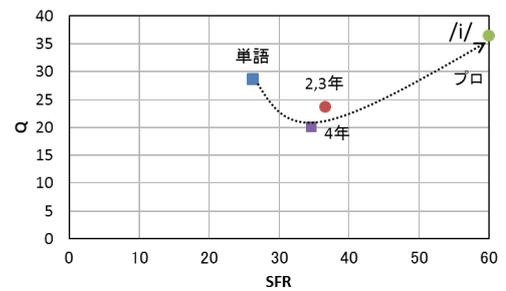


図 14 男声 母音/i/

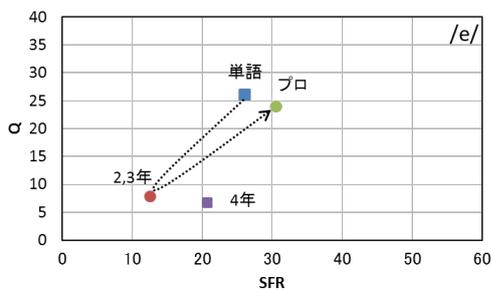


図 10 女声 母音/e/

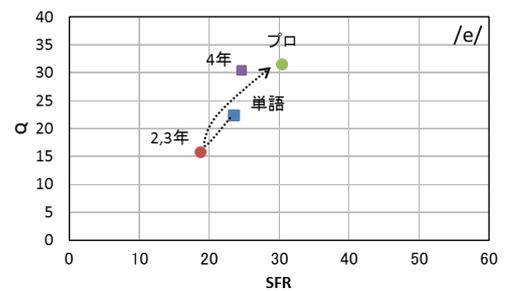


図 15 男声 母音/e/

## 参考文献

- [1] W.T.Bartholomew. A Physical Definition of "Good Voice-Quality" in the Male Voice, *J. Acoust. Soc. Am.*, 6(1), p.25-33, 1934.
- [2] J. Sundberg. Articulatory interpretation of the 'singing formant'. *J. Acoust. Soc. Am.*, 55(4), p.838-844, 1974.
- [3] 齋藤毅他. 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*. 64(5), p.405-417, 2008.
- [4] 山辺大貴他. 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習熟度評価に関する基礎的検討. *信学技報応用音響*. 112(266), p.61-66, 2012.
- [5] 野田美春他. 歌唱の習熟度に関連する周波数特性に基づく初学者とプロ歌手の歌声評価. *信学技報教育工学*. 115(444), p.35-40, 2016.
- [6] 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. *信学技報教育工学*. 114(441), p.45-50, 2015.
- [7] 吉田祥他. 歌声の習熟度に関連する周波数特性に基づく音響特徴量の特定個人の長期的変化. *SIGMUS*, Vol.114, No.12, pp.1-6, 2017.
- [8] ヨハン・スンドバリ. 歌声の科学. 東京電機大学出版局. P.115-130, 2007.
- [9] 秋月影雄他. C 言語/デジタル信号処理. 培風館. p.110, 1989.
- [10] 山下泰樹他. 歌声の習熟度に関連する音響特徴量に基づく実時間分析ツールの製作. *信学技報教育工学*. 115(444), p.13-18, 2016.
- [11] 桑原尚夫他. アナウンサー音声の音響的特徴. *信学論*. Vol.J66-A, No.6, pp.545-552, 1983.
- [12] 野田美春他. 歌声の習熟度に関連する周波数特性に基づく初学者の音響特徴量に関する考察. *教育システム情報学会学生研究発表会*. p.21-22, 2016.