

# ポピュラー音楽の歌唱を対象とする高音域発声評価システムの構築

## Construction of high pitched sound region utterance evaluation system intended for singing of popular music

平山 健太郎

法政大学情報科学部デジタルメディア学科

学生証番号 07k1137

E-mail: kentaro.hirayama.h9@cis.hosei.ac.jp

### Abstract

Recently, a lot of Japanese popular music have included a high tone which cannot be sung by one vocal register. Throats might be tired if singing forcibly. Such a situation is seen well in Karaoke. Although previous research on singing evaluation has focused on trained, classic and opera, there has been a few previous research that dealt with the problem of automatic evaluation of singing qualities. Karaoke's automatic evaluation scheme only evaluate pitch, vibrato and portamento. If distinction of vocal register scheme was added, it is sure to become a further more progress. In this study, the aim is to explore a method that distinguish of vocal register and recommend suitable vocal register. We used Fundamental frequency and difference between the first harmonics and the second harmonics as feature value to analyze. 95 % was distinguished as vocal register classification of vowel /a/.88 % was distinguished as vocal register classification of all vowel.

### 1 まえがき

カラオケはいまや日本の代表的な文化のひとつである。カラオケで歌われるジャンルはポップスから始まりロックやメタル、演歌まで様々である。人気のあるポップスやロックの中にも歌唱の難易度が高いものも多く、特に高音を要求してくるものが多い。カラオケで歌う場合、キーを下げれば自分の歌える範囲に曲を調整できるが、一部のアーティストの場合には音域が広くて歌えない場合がある。また、原曲のキーのままでない音程が取れない人や、原曲キーを好んで歌う人も多く存在する。その場合には、高音域を発声するために無理に喉を締め上げて歌われることがある。そうなると、すぐに声が枯れてしまったり、裏返ってしまう。

従来、歌唱音声の特性を明らかにする研究や、人間の歌唱理解に関する研究はあったが、それを歌唱力の自動評価につなげた研究事例はいまだ少ない [1]。歌唱音声の特性としては、Singer's Formant [2] が存在すること、基本周波数には歌唱音声特有の変動があること [3] が明らかとなっている。また、人間の歌唱理解に関しては、歌声知覚における心理的特徴の分析 [4] と、音響特徴量との関連付け [5]、歌声らしさを特徴づける基本周波数軌跡に関する考察 [6]、朗読音声と歌唱音声の人間の識別能力に関する調査と自動識別 [7]、歌唱音声の音響解析に基づく歌唱力評価の考察 [8]、などの研究事例がある。歌唱力自動評価の研究で使われる特徴量は、楽譜情報ありの場合だと音程の一致、リズム感などであり (カラオケ採点システムなど)、楽譜情報なしの場合ではピブラートや相対音高などであった。声質・音色などの特徴量を使っている場合もあるが、声区に注目した歌唱力自動評価の研究事例はなかった。

本研究は、歌唱における高音域の調査を行い、母音の発声に

よってユーザの声区を自動判別し与えられたピッチに対して適切な発声ができているかどうかを倍音成分などの特徴量から評価するシステムを構築し、コンピュータで歌唱力の向上を支援することを目指すものである。

### 2 適切な歌唱の判別

本研究で提案する歌唱評価システムにおいての適切な歌唱とは、要求された音に対して負担の少ない発声で歌えているか、ということである。そして負担の少ない発声をするためには適切な声区かどうかで決まる。実際にプロのアーティストの場合では、男性の場合だと 70 % が、女性の場合はほぼ 100 % が 2 つ以上の声区を使っている。そのためには、声区がどのようなものかを知る必要がある。ここでは声区のしくみと現在の歌唱における一般的な問題を述べる。

#### 2.1 声区

声区は、完全に喉頭における事象である。声区とは、ほぼ同一の声質で発声される。連続する声の周波数の領域または列のことで (声の基本周波数においては少々重複が生ずるもの) 実際の声区の設定は、これを裏づける、知覚的、音響的、生理的、および空気力学的根拠に依存するべきである。

男性なら主に 2 種類 (地声と裏声)、女性なら 3 種類の声区 (胸声区、中声区、頭声区) があると言われている。また声区は次のように定義されている [2]。

- 地声区・胸声区・・・平常時に喋る時の声とほぼ同じである。声帯は弛緩しきっており、厚い。低音域で使われることが多い。
- 裏声区・・・地声から裏返った (喚声点を越えた) 声のこと。発声法の一つであるファルセットと同じ意味で用いられることが多い。
- 中声区・・・胸声区と頭声区の間のような声である。
- 頭声区・・・「頭に響く」様に感じられる声である。音色的には、芯はなく広がりがあるといわれるものが多い。ファルセットとの違いは、起声がはっきりする、低次倍音が多く言葉が明瞭、息漏れが少なく大音量がだせるなどの特徴がある。

声区は特定の発声周波数域にわたるが、周波数に関して多くの声区は重なるので、与えられた発声周波数を異なる声区で発声することがありえる。男性の地声と裏声の重複区間は 200 ~ 350Hz のあたり (ピッチで大体 G3 ~ F4) であり (図 1)、女性では胸声区と中声区の重複区間が 400Hz 近辺 (ピッチ G4)、中声区と頭声区の重複区間が 660Hz 近辺 (ピッチ E5)。いずれも個人差により声区の境界は大きく異なる

#### 2.1.1 プロの歌手の歌唱

今日のポップスやロックなどのポピュラー音楽では、男性歌唱曲でも盛り上がる部分で男性の喚声点である 350Hz 周辺を超えるものも多く、1 つの声区で容易に歌えるものではない。そこで、プロの歌手がどのように声区を使っているかの調査を行った。男性 5 人に好きなアーティストを男女各

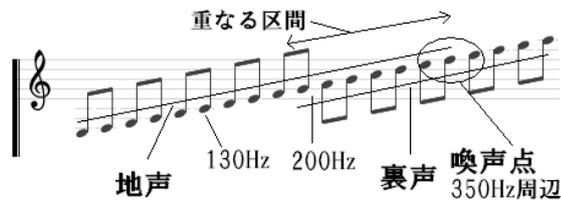


図 1. 男性の声区の周波数域

3~5名ずつほどあげてもらい、さらにカラオケで歌われている年間ランキング TOP30 に入っているアーティストについて調査した。調査内容は喚声点を越える曲を歌っているか、裏声声区を頻繁に使っているかを主観で調査した。結果は男性の約 70 この結果から、ほとんどのプロの歌手は喉に負担をかけない歌い方をしていることがわかり、ライブなどで長時間歌っても声が枯れないことがわかる。

## 2.2 一般の人の歌唱

プロの歌手と比べて、カラオケなどで歌う一般の人はほとんど 1 つの声区しか使ってないことが多い。しかし、歌う曲には喚声点を越えている音があるので、無理に地声で張り上げ声を枯らしてしまうことがある。こういった事態を避けるためには、求められた発声周波数に対する適切な声区を選択することが重要である。

このことから、評価システムでは声区が重なる範囲で地声が裏声かを判別できるようにしなければいけないことがわかる。

## 2.3 従来研究での特徴量の変化

適切な歌唱かどうかを判定するには、声の特徴量が歌唱中にどのように変化していくのかを正確に把握しなければならない。声は、大きさや音程、声区などさまざまな特徴をもっている。そしてこれらの特徴の変化にも一定の傾向があるはずである。ここでは、声の特徴の変化をクラシックの分野における従来研究 [2] と、実際に歌声を録音し分析した結果を比較し、歌唱評価システムに有効な特徴量を見出す。

クラシックの分野における従来研究 [2] では、スペクトル分析で歌手の技術の違い、性別、ラウドネス、ピッチ、声区の違いを調べていた。また、気流グロトグラムと言われる声門での波形の分析も行われていたが、評価システムを使用する場合に分析が困難になるので本研究では考慮しない。

声は、それを構成するいくつかの周波数が簡単な整数比となっている。そして倍音の混ざり具合によってその音の特徴、つまり音色が決まる。ここでは、基本周波数成分を第 1 部分音(基本波)、第 2.3 倍音を第 2.3 部分音とし、それぞれの歌声を比較したものを分析している。また、音源スペクトルを構成する調波成分の音レベルは、声帯の空気流の変化を非対称三角波の近似モデルとして -12dB/オクターブの割合で減少するので、各部分音は 1 オクターブ下、すなわち半分の周波数である部分音より、12dB 弱いことを意味している。

### 2.3.1 男声と女声の違い

男声と女声での声質の違いのある部分は、喉頭音源によるものであると考えられる。先行研究 [2] では、男性被験者は弱い基本波になっているのに対し、女性被験者は強い基本波になっており、歌手の場合にも同様であるとしているが、少数の被験者のため、これだけでは一般化することはできないとしている。実際に男性 3 人、女性 3 人の平均で実験してみた結果、男声のほうが平均からの基本波成分が高かった。先行研究とは異なったが、この結果から、男性と女性で違うモデルを作成する必要があることがわかる。

### 2.3.2 ラウドネスの違い

音源スペクトルを比較するにあたって、1 人の声の音源スペクトルは一定というには程遠い。最も重要な声のパラメータであるラウドネスや発声周波数が変化すれば、音源スペクトルもかなり変化するからである。発声のラウドネスが増加すれば、スペクトル内の倍音成分の振幅が、基本波の振幅よりもかなり大きく増加することが知られている。また、高い倍音成分は、低い倍音成分より上昇する傾向がある。これは、音源スペクトル

の低い部分音は、発声のラウドネスが低いときほど高いときよりもはるかに支配的である、という典型的な観察結果と一致する。

### 2.3.3 ピッチの違い

発声周波数(ピッチ)の変化もまた、音源スペクトルの変化と関連づけられる。発声周波数が上がれば、音源スペクトルにおける倍音の振幅は減失する傾向がある。効果はむしろ小さく、低域のスペクトル部分波に限られており、これらの低域スペクトル部分音は、発声周波数が高くなるにつれてより支配的ではなくなる。

これは異なる強さの発声条件下での結果と同様の効果であると考えられる。同じ声区の場合、高い音はたいてい低い音よりも強い発声で歌われるからである。

### 2.3.4 声区の違い

声区の違いでは、先行研究 [2] では 3 名のプロ男性歌手が、同じ基本周波数で地声声区と裏声声区で母音/a/を発声したとき平均音源スペクトルは、裏声のほうが基本周波数高く、女性歌手の場合では中声区のほうが胸声区よりも高かった。

クラシックの分野以外でも、この結果になるのかを実際に調べた。発声評価システムで対象となる、歌手ではない一般の人に協力してもらい、/a/の発声を裏声と地声で、なおかつ同じ基本周波数で男女それぞれ 2 名ずつ録音し、基本波から第 15 部分音までの値を取得し、平均からの差を出した(図 2,3)。

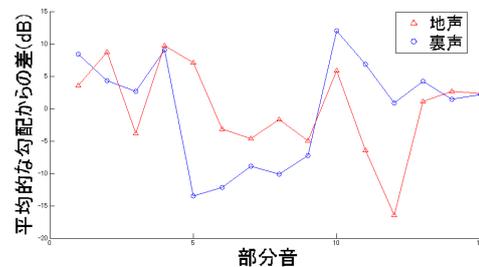


図 2. 声区の違いによる部分音成分の割合：男性

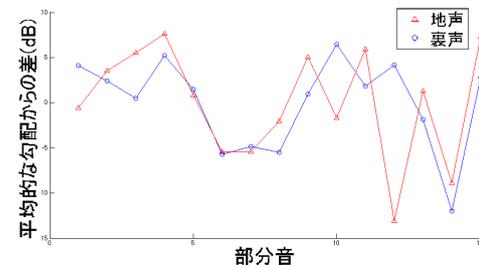


図 3. 声区の違いによる部分音成分の割合：女性

先行研究 [2] と同様に、歌手ではない人の発声でも、男女とも基本波と第 2 部分音は地声と裏声で傾きが大きく違った。また、第 13 部分音が男女とも地声と裏声で大きく違うという結果が見られた。

### 2.3.5 母音の違い

クラシックの分野に限ったことではないが、発声する母音が従来研究では主に/a/のみ、もしくはほとんどであった。しかし、本研究で提案する発声評価システムではすべての母音を評価できるようにする。そのために、ここでは母音の違いがスペクトルにどのように現れるかを見た。

図 4.5 は開口母音の/a/と閉口母音/i/のスペクトルである。母音の音響的特徴はほぼ第 1・第 2 フォルマントによって表現することができる。/a/の第 1 フォルマントは 800Hz 周辺、第 2 フォルマント 1000Hz 周辺である。母音/i/は対照的に第 1 フォルマントが 300Hz 周辺、第 2 フォルマントが 2000Hz とフォルマント間が離れている。特に第 1 部分音と第 2 部分音の成分の差が大きく違うことがわかる。この結果から、第 1 部分音と第 2 部分音の成分の差の傾きを特徴量として使う場

合、各母音ごとに適切な補正をかけなければいけないことがわかった。

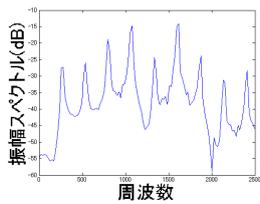


図 4. 母音/a/

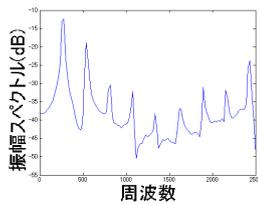


図 5. 母音/i/

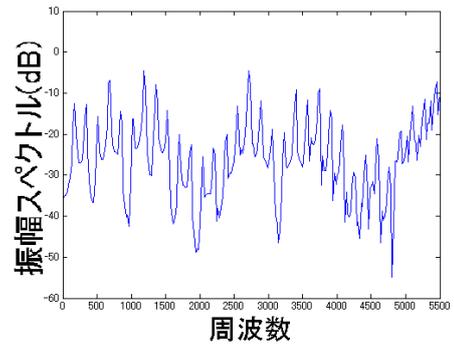


図 7. 残差信号スペクトル

### 3 発声評価システムの構築

#### 3.1 発声評価システムの概要

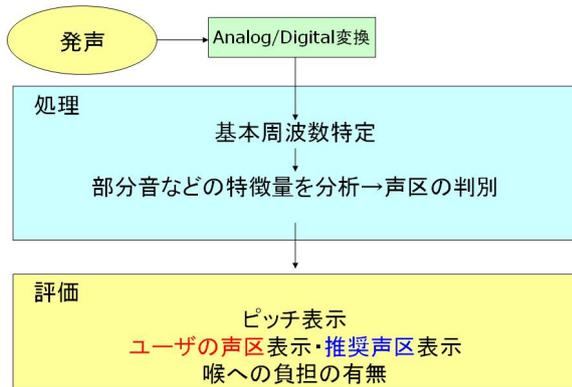


図 6. 発声評価システム

評価方法として、まずシステムを使用する歌手に発声してもらい、この際、マイクとの距離を学習データを録音したときと同じ条件の 10cm にする。発声をデジタルに変換し、取得した音声を分析する。処理の流れとしては、まず基本周波数の特定をし、基本周波数の値から部分音を取得し、声区や母音を求める。

評価では、まず現在のユーザの声区を表示し、次に喉に負担がかかっている声かどうかを表示、最後に基本周波数とユーザの喉への負担度合いから推奨される声区を表示する。

#### 3.2 声区判定システム

##### 3.2.1 事前処理

発声のアタックなどは余計な情報になりやすいので、声の定常部分のみを音声データからまず切り出す。次に切り出した母音の定常部分にハニング窓をかける。それを高域強調 (プリエンファシス) する。音声信号は低周波数成分が大きく、周波数が大きくなるにつれて振幅スペクトルが小さくなっていくという特徴があり、この周波数の偏りの修正処理をする。また、声帯の基本振動、空気中の伝搬の影響を除き、口腔内の形状による周波数特性を強調する処理でもある。さらに残差信号をとりそれを 2048 点で高速フーリエ変換した振幅スペクトルに対数をつけたもの (単位 dB) を使う (図 7)。

##### 3.2.2 ピッチの取得

ピッチの取得方法には複数の方法があるが、声道の形によって共振が発生し、調音されるので、声道特性を積極的に取り除く必要がある。声道特性とは、声帯が周期的に開閉し、規則的な空気の断続による音声の音源となる声帯特性とは異なるもので、舌、顎、鼻腔などの影響を受け、ピッチを取得するのに余計な情報となる。これをケプストラム法を使うことによって声道特性と音源特性に分けられる。ケプストラムには、低次成分にフォルマント成分などの声道特性が現れる。そこで低次成分を除去して高次成分だけを残すことで、声道特性を除去しピッチ成分を抽出できる。

#### 3.3 特徴量の取得

ケプストラム法によって求めたピッチから各部分音の周波数領域を推定し、その領域のピークを部分音とする。母音/a/の地声と裏声で違いが顕著に現れた基本周波数の成分と第二部分音の成分の差を特徴量  $\Delta f_0$  パワーとする。

#### 3.4 声区判定のための学習データ

学習データは、各発声の模範となるようなデータにするべきである。ここではサンプルの録音環境と学習データの選別方法を示す。

#### 3.5 サンプル収集条件

マイクから 10cm 離れて母音の発声をさせ、サンプリング周波数 44100Hz で録音した。発声は、1 オクターブ分の上昇音階と下降音階を地声と裏声でそれぞれ発声してもらい、地声と裏声が重複する 200Hz ~ 500Hz を重点的に録音した。

#### 3.6 学習データ作成

今回の学習データは男性 4 人、女性 2 人から、それぞれ 30 ~ 40 ほどの地声と裏声を各母音で発声したものにした。被験者の選別条件は、実際に耳で聞き地声を一定して発声できているかを主観評価で選んだ。男性の母音/a/の学習データ (図 8) では、多少の個人差があるものの地声は 150Hz 周辺から 400Hz 周辺にかけて  $\Delta f_0$  パワーが下がっていき、裏声はたいていのデータの 0 パワーの値が 0 以上になり、 $\Delta f_0$  パワーが負の値になるときは、地声と同じように周波数が増えるほど負の値が大きくなった。また、地声と裏声の識別方法として 2 乗誤差の最小化による線形識別を行った。

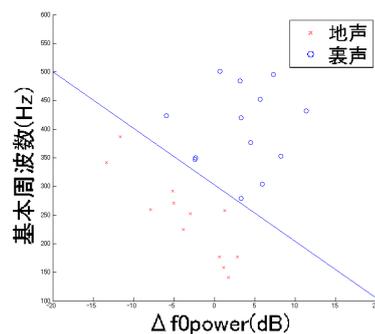


図 8. 学習用データ男性/a/

## 4 実験結果

#### 4.1 母音/a/

男性 3 人、女性 1 人の発声データを新たに取得し、評価データ (図 9,10) として使用した。発声方法は学習データの時と同じである。男性の地声判定率は 100 %、裏声判定率も全体で 90 % と高い判定成功率を示し、男性全体では 95 % の判定率であった (表 1)。女性についても高い判定率を出すことができた。男性の裏声を地声と誤って判定したものは、耳で聞いた主観では裏声の限界の低さであり、発声が不安定になるので歌唱では使用されないと考えられるものであった。

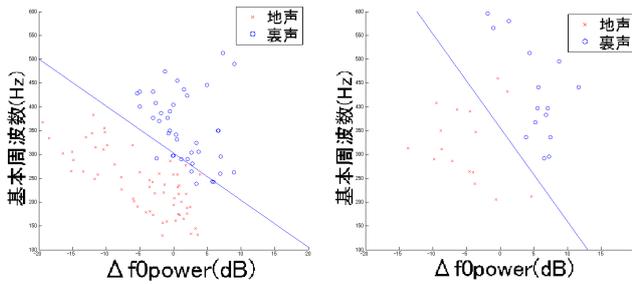


図 9. 評価データ男性/a/

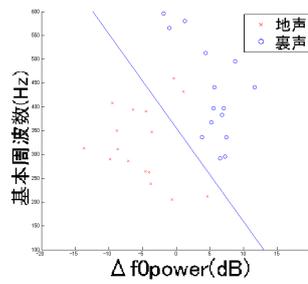


図 10. 評価データ女性/a/

表 1. 判定成功率/a/

歌手	地声判定率	裏声判定率	声区判定率
男性 1	100 %	77 %	92 %
男性 2	100 %	88 %	94 %
男性 3	100 %	100 %	100 %
男性全体	100 %	90 %	95 %
女性 1	88 %	100 %	94 %

#### 4.2 他の母音

/a/以外の母音、とくに/i/などは個人差が大きく出た(図 11, 図 12). 男性 1 は地声と裏声をはっきりと分かれているが、男性 2 では完全に混ざってしまっている. 耳での主観的な評価は、男性 1 は発声が上手いとは言えず、男性 2 は一般的な地声と裏声の発声であった. この場合、男性 2 は高い判定率(94%)を得たが、男性 1 は低く(68%), /i/の地声においては25%という結果であった. しかし、もともと男性 1 の声が地声と裏声であまり変化はなく、本人が地声で発声しているつもりでも裏声になることがあったので、裏声を地声とラベル付けしてしまった可能性がある.

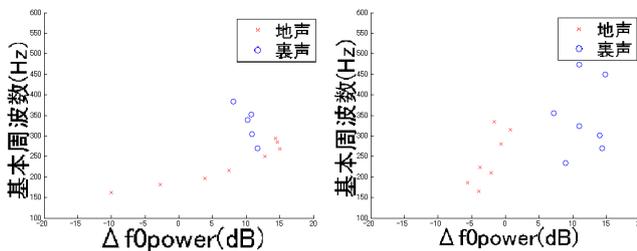


図 11. 男性 1/i/

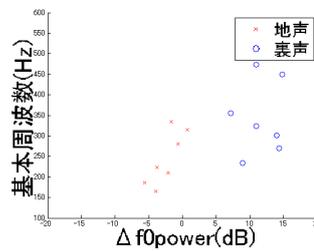


図 12. 男性 2/i/

表 2. 判定成功率/i,u,e,o/

母音	男 1 地声	男 1 裏声	男 2 地声	男 2 裏声
/i/	25 %	100 %	100 %	100 %
/u/	38 %	100 %	100 %	88 %
/e/	75 %	33 %	75 %	100 %
/o/	88 %	100 %	88 %	100 %
全体	56 %	84 %	91 %	94 %

#### 4.3 Δf0 パワー特徴量考察

母音/a/の発声については、今回使った Δf0 パワーの特徴量が有効であることがわかった. ほかの母音についても母音/a/と同様に有効であることがわかった. そもそも母音は、第 1 フォルマントと第 2 フォルマントの周波数で識別できるが、歌唱中のピッチが変わることで基本周波数が第 1 フォルマントに重なることがある. 母音/a/の場合、第 12 フォルマントは 800Hz, 1000Hz 周辺であり、声区判定の際の基本周波数がどちらにも重なることはないが、他の母音、特に母音/i,u/については第 1 フォルマントが 300Hz 周辺にあるので基本周波数に重なってしまう. しかし、母音ごとの識別関数を使えばフォルマントの影響は避けられることが今回の実験でわかった.

#### 4.4 喉締め声の判定

上昇音階で発声練習をした場合に、音程が上がるにつれて音圧も上がっていく. しかし、喉締め声になっていると思われる発声以上の高さで音圧が上がらなくなるという傾向がわかり、被験者が喉を締めている(苦しい)と感じ始める音程も音圧が上がらなくなる音程とほぼ一致した. この傾向に基づいて、母音/a/の喉締め声とそうでない地声を区別し、その識別関数を求めた.(図 13). これにより男性の母音/a/の高音域における発声評価システムができた.

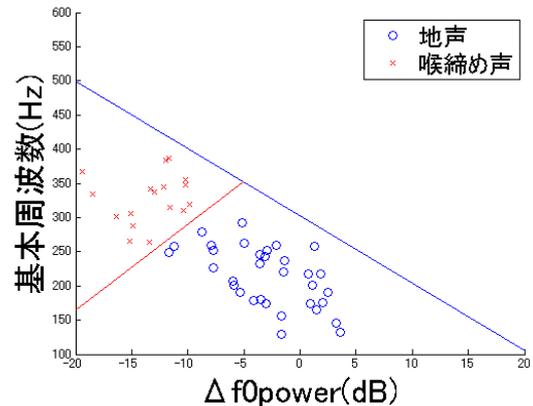


図 13. 喉締め声の境界線/a/

### 5 あとがき

今回までのシステムで、よほど低い裏声を使わない限りどの母音でも高い判定率を出すことに成功した. 低い裏声はもともと地声との判別が難しいという点もある. Δf0 パワーと基本周波数による声区の判定は母音のフォルマント周波数に影響を受けるが、識別関数を母音ごとに設定することで解決した. 男性の母音/a/においては 95% の高い割合で声区判定することができた. よって本研究で使用した特徴量 Δf0 パワーは声区判定に有効であることがわかる.

今後の課題として、一つの母音の発声練習を越えて歌唱を自動評価する際には、どの母音を発声しているかを識別する必要がある. また、男性の母音/a/に関しては喉締め声の判定を行うまでに至ったが、他の母音、特に閉口母音に関しては Δf0 パワーで喉締め声を識別させることが難しかったので、他に有効な特徴量がないかさらに調べる必要がある. また、被験者の数が少ないのでさらに増やしていく.

#### 参考文献

- [1] 中野他”楽譜情報を用いない歌唱力自動評価手法”, 情処学論, 48 巻 1 号, pp.227-36, 2007-01-15
- [2] Sundberg, J.”The Science of the Singing Voice”, Northern Illinois University Press, p.226, 1987
- [3] 矢田他”歌声の基本周波数の動特性”, 平 10 音響学秋季講演論文集 3-8-6, pp.383-384, 1998
- [4] 西内他”専門家と非専門家の歌声の評価”, 音響学聴覚研資, H-90-1, pp.1-6, 1990
- [5] 辻他”歌声らしさの要因とそれに関する音響特徴量の検討”, 音響学聴覚研資, H-2004-8, Vol.34, No.1, pp.41-46, 2006
- [6] 斉藤他”歌声の F0 動的変動成分の抽出と F0 制御モデル”, 音響学聴覚研資, Vol.31, No.10, pp.683-690, 2001
- [7] 大石他”局所的・大局的な特徴を利用した歌声と朗読音声の識別”, 情処音楽情報科学研報, Vol.2005, No.82, pp.1-6
- [8] 津田他”3D 解析による歌声の評価に関する研究”, 1996 信学ソ大 D-458, p.461