Vol.2012-MUS-94 No.16
Vol.2012-SLP-90 No.16
2012/2/4

# 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

平山　健太郎[†1]　伊藤　克亘[†2]

近年の日本のポピュラー音楽では, 一つの声区のみで歌えないような高い音域を含む楽曲が数多く存在する. 実際に一つの声区のみで歌おうとすると声が枯れてしまうという結果が多い. このような状況はしばしばカラオケで見受けられる. 従来より, 歌唱力の自動評価は様々な手法で行われてきたが, 高音域の発声における発声状態の評価を行っているものは少ない. 本研究では, 基本周波数成分やフォルマント, 倍音構造, 残差スペクトルなどの特徴量から声区と発声状態のトレーニングデータを 35 次元で作成し, 自動でユーザの音声データから高音域の発声評価を行うシステムを構築した. 歌唱音声に対する音符単位の識別率は 93.18%であった.

# Discrimination Method of Voice Register and Voice Quality in high pitch for Popular Singing

Kentaro Hirayama and Katunobu Itou

A lot of recent Japanese pop music includes high tones that cannot be sung using only one vocal register. Indeed, attempting to do so often results in sore throats. This situation often occurs in karaoke. Previous research on singing evaluation has focused on trained, classical, and operatic styles, but there has been little research that deals with the automatic evaluation of singing quality in high tones. We have developed a system with 35 dimensions that automatically determines user's voice quality by identifying a user's vocal register and then recomeending a more suitable one. We used the fundamental frequency, formant, harmonics and residual spectrum, etc as feature values for analysis. Identification rate of singing voice per notes was 93.18%.

†1 法政大学大学院情報科学研究科
Hosei University Graduate School
†2 法政大学情報科学部
Professor at Hosei University

## 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と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．

従来より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が行われてきた. 歌唱音声の特性としては，Singer's Formant[1] 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 [2] 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 [3] と，音響特徴量との関連付け [4]，歌声らしさを特徴つける基本周波数軌跡に関する考察 [5]，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 [6]，歌唱音声の音響解析に基づく歌唱力評価の考察 [7]，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ビ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，声区や高音域の発声状態に注目した歌唱力自動評価の研究事例はなかった．

本研究では，歌唱における高音域の調査を行い，実際の歌唱より得られた音声データのユーザの発声状態を自動判別し, 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やフォルマント周波数などの特徴量から評価するシステムを構築した. これにより, 将来コンピュータ上で歌唱力の向上を支援するシステムの構築を目指すものである．

## 2. 高音域発声の調査

### 2.0.1 プロと一般人の歌唱の違い

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換声点 (地声と裏声の境界) である 350Hz 周辺を超えるものも多く，1 つの声区で容易

Vol.2012-MUS-94 No.16
Vol.2012-SLP-90 No.16
2012/2/4

に歌えるものではなくなっている．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 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．調査内容は換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%，女性アーティストほぼ全てが裏声声区を頻繁に使っていた．ファルセットなどのか細い発声から地声と大差がない芯のある裏声までとさまざまであった．また，上手いと言われるアーティストほど高音域を頻繁に使用しており，逆にほとんど 1 つの声区しか使わないようなアーティストは，換声点付近が最高音であった．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たり高い声を出し続けても声が枯れないことが推測される．

プロの歌手と比べて，カラオケなどで歌う一般の人, 特に男性の場合はほとんど 1 つの声区しか使ってないことが多い．しかし，歌う曲には換声点を越えている音があるので，無理に地声で張り上げると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, プロの歌手のように求められた発声周波数に対し適切な声区を選択することが負担の少ない発声をするために重要であることがわかる．本研究では負担のかかっているであろう声を「喉締め声」とし, 独自に発声の分析を行い, 発声状態を決定する．

### 2.1 データの収録

高音域の発声状態の分析と 3 章の学習データの構築のために, 男性 3 人の各母音の音階発声で得られたデータを使い, 地声, 裏声についてはそれぞれ別のデータとして録音した. 音階発声では, 地声では C2(130Hz) から 1 オクターブを長音階で半音ずつ G2(196Hz) まで, 裏声では A2(220Hz) から地声と同じ条件で E3(330Hz) から始まる 1 オクターブの発声をしてもらったものを使った．ここでは高音域における発声の問題と, 高音域の発声を様々な特徴量を用いて分析した結果を示す．

### 2.2 声区の違い

声区は，完全に喉頭における事象であり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 (声の基本周波数においては重複が生ずる) のことである. 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．

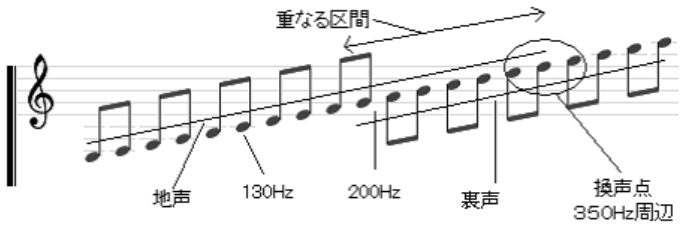


**図 1** 男性の声区の周波数域

- 裏声区 ··· 地声から裏返った (換声点を越えた) 声のこと．発声法の一種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 ··· 胸声区と頭声区の中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，起声がしっかりする，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．

声区の違いを分析した結果, 主に低域の倍音成分と第 1,2 フォルマントが変わることがわかった (図 2). 本研究での低域の倍音成分とは第 1〜3 までの倍音成分の平均であり, 以降 3 つずつの倍音成分の平均ごとに中域, 高域としている (図 3). 第 1 倍音成分である基本波成分は第 1 フォルマントの影響を多大にうけてしまうが, 第 3 倍音成分までの平均とすることでどの母音に置いても第 1 フォルマントを含むことができ, 高音域の発声においての第 1 フォルマントの影響を軽減できると考えた (図 4,5). 図 2 の音域は B2(247Hz)〜B3(494Hz) までであり,A3(440Hz) からは被験者は裏声 (ファルセット) を使っている．

### 2.3 喉締め声の観測・分析

歌唱において高音域, 特に男性においての換声点付近の地声は苦しい印象を受ける. しかし, 地声と裏声のように声区ではなく, また意図して出せるものではないのではっきりとは定義できない. そこで地声の音階発声から換声点付近について変化がないか分析した. 母音/a,o/

Vol.2012-MUS-94 No.16
Vol.2012-SLP-90 No.16
2012/2/4

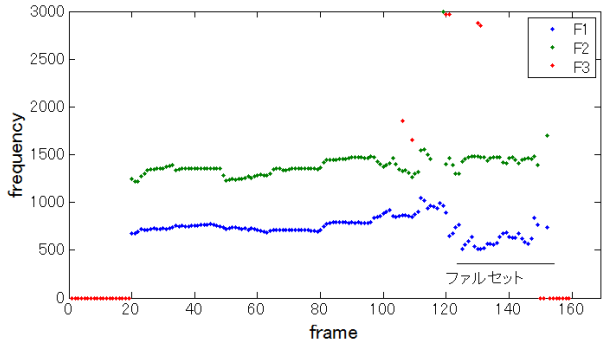


**図 2** 地声と裏声のフォルマント周波数
Fig. 2 Formant of modal voice and falsetto

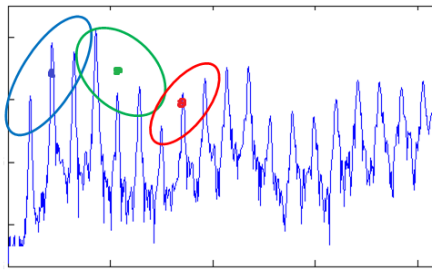

**図 4** 地声のスペクトル
Fig. 4 Spectrum of modal voice

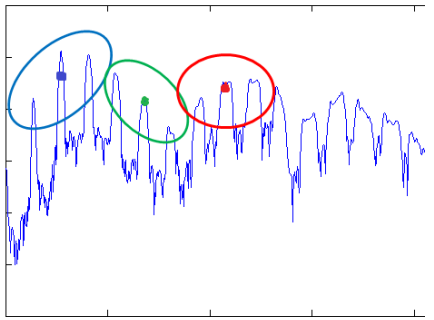

**図 5** ファルセットのスペクトル
Fig. 5 Spectrum of falsetto

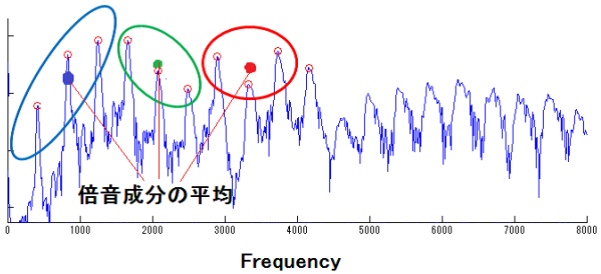


**図 3** 3 つの倍音成分の平均
Fig. 3 Average of Three Harmonics of modal voice and falsetto

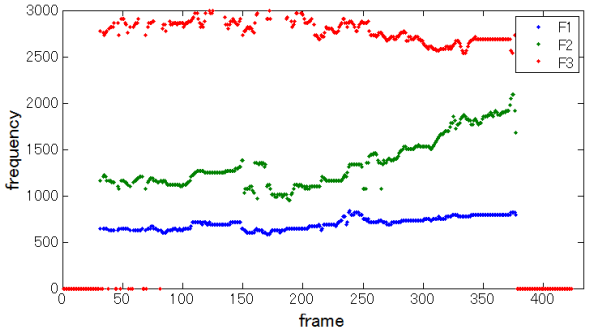


**図 6** 地声/a/のフォルマントの遷移
Fig. 6 Formant transitions of vowel /a/

などの開口音においては, 換声点に近づくにつれて第 2 フォルマントが上がるという傾向が多く見られた. 図は G2〜G3 での音階発声のフォルマントの遷移である (図 6,7).

しかし閉口音 (母音/i,u/) では共通の変化が見られなかった. なので, 開口音については第 2 フォルマントが上がったところ, 閉口音については換声点付近かつ開口音で喉締め声が始まる周波数領域を喉締め声とした. また, 女性についてはたしかに声区の変化はあったのだが, 声区が 2 つもしくは 3 つなのかは諸説あり, 男性とは多くの特徴量の傾向が一致しないので今回は声枯れ・裏返りが観測されやすい男性だけに絞った.

## 3. 発声評価システムの構築

本章では歌唱の音声データから, 発声状態を判別するシステムを構築する. 発声状態の判別分析は, データの収録より得られた基本周波数, 倍音構造, 残差スペクトル, 第 1,2 フォルマントなどの特徴量を使い,35 次元で行った. この章ではシステムの処理の流れ (図 8) と学習データの構築方法を述べる.

Vol.2012-MUS-94 No.16
Vol.2012-SLP-90 No.16
2012/2/4

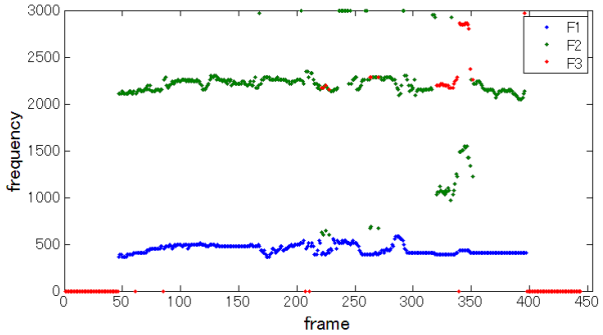


**図 7** 地声/i/のフォルマントの遷移
Fig. 7 Formant transitions of vowel /i/

### 3.1 処理の流れ

#### 3.1.1 残 差 信 号

人の声は声道特性によって個人の特性が付加される. しかし歌唱において個人の特性は余分な特徴量となってしまう. このため, 判定の際には残差信号を用いる. 声道特性を取り除くためにはケプストラム法や逆フィルタによる残差信号の推定などがあるが, 本研究では逆フィルタによる方法を選択した. レビンソンダービンのアルゴリズムを用いて,LPC 次数 14 で LPC 係数と PARCOR 係数を得る. その 2 つを用いた逆フィルタから残差信号を得る.

#### 3.1.2 ピッチと倍音成分の推定

残差信号はフォルマントの影響が軽減されているのでピッチ成分の推定がしやすい. ピッチの推定には, 相互相関を使う. よりよい推定のために最低周波数 100Hz と最高周波数 1000Hz を設定することで精度を上げた. 相互相関の値を使い, 残差スペクトルの倍音成分を取得する. スペクトルのピークより高調波成分とその周波数を取得し, より細かい基本周波数の推定のために 5 つの倍音を使う. ここで 5 つまでとしたのは高次の倍音成分ほど推定率が悪くなってしまうことが多いからである (カラオケなどの雑音環境では尚更である). また, 有声区間の判定にも相互相関の値を使う.

#### 3.1.3 フォルマントの推定

音声データから第 1,2 フォルマントを推定する. 実際の歌唱では様々な母音が出現し, また残差スペクトルとはいえフォルマントの影響を完全には排除できないので, 特徴量としてフォルマント周波数を使う.

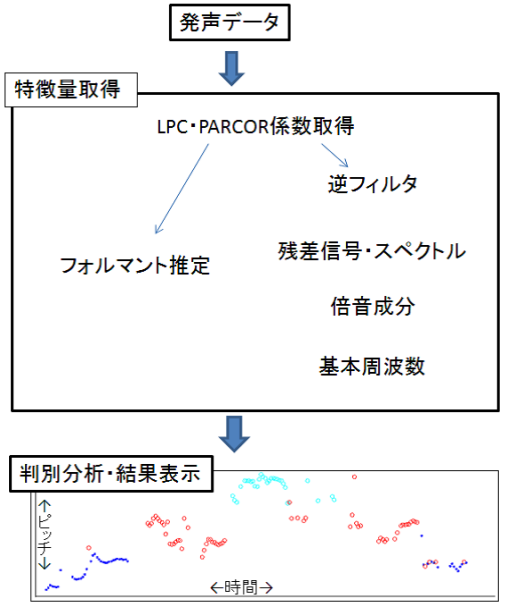


**図 8** 発声評価システム
Fig. 8 Voice Evaluation System

### 3.2 学習データの構築

本研究では学習データを地声, 裏声, 喉締め声の 3 つのグループに分けて判別分析を行う. 喉締め声については 2 章で観測された特徴に基づいて決定した. また, 第 1,2 フォルマントの範囲外である 3000Hz を境界にした倍音成分の平均を特徴量とした. 判別分析にはガウス分布を用いた対角の共分散行列の推定を持つ線形モデルを使った. 対角成分の推定にはベイズ分類を用いた.

## 4. 判別分析の実験

実際の歌唱音声に対して判別分析を行なった. 音階発声は地声と裏声だけで歌の上手い下手にはあまり関係はないが, 実際の歌唱となるとその差は歴然である. ピッチの精度やビブラートといった技術もさることながら, やはり高音域での声質が大きく被験者で違った.

Vol.2012-MUS-94 No.16
Vol.2012-SLP-90 No.16
2012/2/4

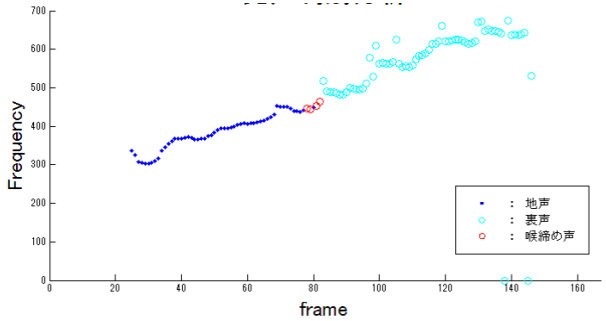


**図 9** 裏声の音階発声の判別結果
Fig. 9 Discrimination Result of scale vocalizing of falsetto

### 4.1 音階発声

地声の音階発声に関しては喉締め声と判別された部分が換声点付近となり, 概ね正解していると考えられた. しかし, 裏声の音階発声では低い周波数領域では地声と判定されてしまった (図 9).

### 4.2 歌唱に用いた曲

被験者に歌ってもらった曲は J ポップ,J ロックアーティストである Janne Da Arc の「月光花」という曲である. このアーティストは男性にとって「声が高いアーティスト」の筆頭であり, うまく歌える人とそうでない人の差が出やすいと思われる. また, 曲に関しては,BPM が 100 前後とゆったりしているので音素一つ分のデータが多く取れるといった意味もある. そしてなにより知名度が高いので被験者が最初からメロディを知っているといった利点もある. 歌ってもらった部分は最後のサビ部分であり, サビパートを 2 回繰り返す. 原曲キーでは, サビの音域は G2(190Hz)〜A3(440Hz) である (図 10). 月光花の該当部分のカラオケデータを原曲キーから上下キー 5 つ分を作成し, ヘッドフォンを装着して歌ってもらった. 大抵の被験者がカラオケで歌う場合と遜色ない歌声 (音量, 感情など) であった.

### 4.3 判別分析の結果

4 人の男性被験者 A〜D 歌唱データを判別した結果を図 11〜14 に示す. 被験者 A の図はキー＋ 2 で歌ってもらったものである. 最高音 B3(247Hz), 歌詞「ら」の部分でファルセットを使っている. 被験者 A は学習データでも使われており, 高い精度の結果が出た. 地声で A3 まで大声量で出せるのが特徴の被験者 A だったが,B3 以上は裏声にしないと発声出来ず,

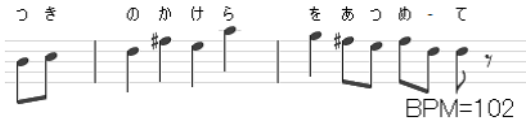


**図 10** サビ冒頭部分の歌詞とメロディ情報
Fig. 10 Merody and Lylic of chorus begging part

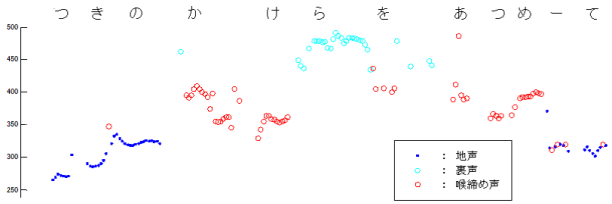


**図 11** 被験者 A
Fig. 11 Subject A

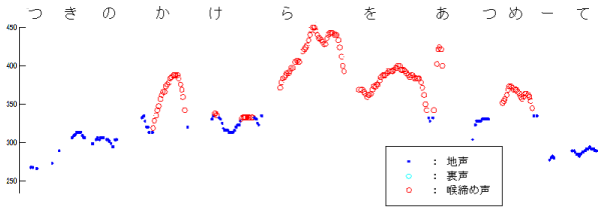


**図 12** 被験者 B
Fig. 12 Subject B

声区を融合しているものではなかった.

被験者 B の図は原曲キーで歌ってもらったものである. 最高音 A3(440Hz), 図のメロディ部分ではすべて地声で歌っていた. 換声点付近の地声はすべて喉締め声と判別された. 実際に耳で聞いても同じ印象を受けた. この被験者 B はよく声を枯らすことがある. 被験者 C の図はキー＋ 4 で歌ってもらったものである. 最高音 C#4(554Hz) で, メロディ部分では地声とファルセットではない裏声を使っている印象を受けた. 被験者 C はハイトーンを使うこ

Vol.2012-MUS-94 No.16
Vol.2012-SLP-90 No.16
2012/2/4

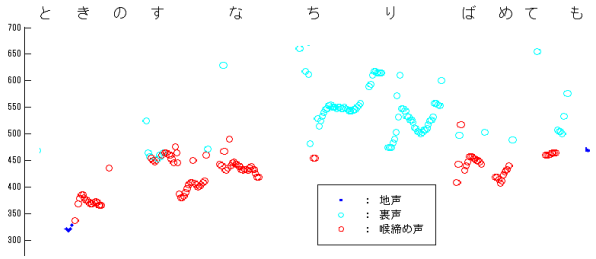


**図 13** 被験者 C
Fig. 13 Subject C

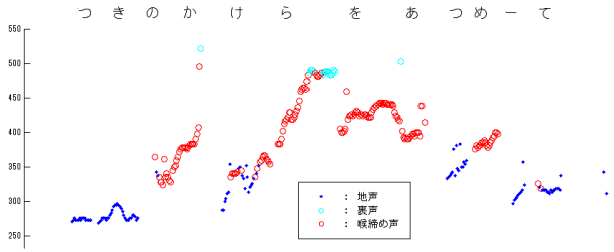


**図 14** 被験者 D
Fig. 14 Fouth Subject

とに慣れており, 普段から最高音の高い曲をよく歌っているという特徴がある. このキーでは地声区の周波数領域をはるかに超えているので, 裏声区を使わないと発声は出来ないと考えられる. 判別結果は最高音付近のみが裏声という結果になった.

被験者 D の図はキー＋ 2 で歌ってもらったものである. この被験者 D は地声が平均より高い周波数まで出るが, 非常に苦しい声に聞こえてしまうという典型的な喉締め声を持つ.

### 4.4 判別分析の考察

今回までのシステムでは裏声としてとったデータがファルセットのみだったので, 裏声区を使っているがファルセットとは違った声質になると判定結果が喉締め声となってしまう結果になってしまった. 課題曲となったアーティストのボーカルも裏声区を使っていると思われるが, 決してファルセットのようには聞こえない. サンプルを収集する際に今回は地声と裏声の 2 つ音階発声を行なったが, 裏声というものが被験者にとってはファルセットと同義であると捉えられてしまったことで, 実際の歌唱で使われる裏声区の声とは違ってしまった可能性がある.

## 5. あとがき

今回までのシステムで, 男性のみを対象とした喉締め声や裏返りが起こる周波数領域での高音域の発声評価システムを構築できた. 声区を判別し, 地声においては負担がかかっているか判別できる. 歌唱音声に対する音符単位の識別率は 93.18%であった。しかしいくつか改善しないといけない点も見つかった.1 つとして, 判別分析の際のグループを増やすことである. 判別分析の考察の際にも述べたが, 裏声区を使っていても声質は様々である. ロックやメタルなどでの甲高い声から薄くやわらかい感じのファルセットなどいくつかあることは明白である. また, 比較的低いファルセットの際に倍音成分の推定精度が落ちるといった現象が起こった. これは発声自体が弱いために高次倍音になるほどピークがあやふやになってしまうからである. 最終的にカラオケルームでの使用を考えるととても 10 以上の高次倍音成分を正しく取得できるとは思えないので, ピッチ推定の手法を変更する必要がある. 有声区間の推定もまだ問題があり, 歌唱の際には誤推定があるので精度を上げる必要がある. さらに女性ユーザへの対応も検討中である. 女性はカラオケで喉が枯れるといった現象はあまりないと思われるが, 声区と声質の違いはいくつかあるはずである. それらを満たすにはさらなる被験者データが必要であり, どのようにしてそれらの評価をするのかも検討しなければならない.

## 参 考 文 献

1) Sundberg, J.”The Science of the Singing Voice”,Northern Illinois University Press, p.226, 1987
2) 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
3) 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
4) 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, pp41-46, 2006
5) 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
6) 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
7) 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,p461