喉頭の運動に注目した歌唱音声の自動判別と評価
Discriminant Analysis of Singing Voice Focused on The Movements of The Larynx

平山 健太郎
Hirayama Kentaro
法政大学大学院情報科学研究科情報科学専攻
E-mail:kentaro.hirayama.5h@stu.hosei.ac.jp

Abstract

A considerable amount of contemporary Japanese pop music includes high tones that cannot be sung using only one vocal register. Indeed, attempting to do so often results in throat strain, such situation often occurs in karaoke. Previous research which evaluates singing voice automatically has focused on musical information, such as the track of fundamental frequency and rhythm like an evaluation system in karaoke. However there is insufficient research on automatic evaluation of the quality of singing voice in high tones although there are lots of studies for singing voice analysis. We have developed a system with 43 acoustic features that automatically determines a user's singing voice quality by identifying his/her vocal register and then recommends a more suitable one. The analysis primarily used fundamental frequency, formant, harmonics, and some acoustic features from residual signal. An electromyogram was used to analyze the movements of the larynx for labeling a "tightened voice" in modal register, which sounds strained and/or represents suffering. The identification rate of the state of singing utterance using linear Naive bayes classifier was 91.5% per unit of note.

1 序論

カラオケの普及などに伴い歌唱する場が多くなり，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，演歌までと様々である．人気のあるポップスやロックの楽曲は歌唱の難易度が高い場合も多く，高音を要求してくるものも多い．そのような楽曲を歌う場合，高音を発声するために無理に喉を絞め上げて歌われることがあり， その結果声が枯れたり裏返ってしまい，長時間の歌唱では意図した制御が出来なくなるなどの影響を及ぼす．歌唱者向けの教則本やカラオケの採点システムなど， 専門の教師に師事せずに歌唱能力の向上を目的としたものは身近にも多様にあるが， いずれも独学で歌唱を学ぶには難しい状況にある. 例えば， カラオケなどでよく見られる自動採点システムでは楽譜上の評価，つまり正しい音程で歌われたかどうかが問題となっており，発声の根本的な解決は示されない. また，教則本では理想とされる発声の仕組みが記述されているが，記述内容通りに発声できるようになるには，歌唱者自身が記述内容に対して感覚的な試行錯誤を繰り返さなければならず，個人の感覚に依るところが大きい.

このような問題に対して，発声の仕組みなどの身体的な情報に基づいた歌唱力自動評価の研究は行われていない．本研究では，歌唱中の高音域発声の調査を喉頭の運動に注目して行い，歌唱者の発声が喉に負担をかけているかどうかを判別するシステムを構築し，高音域の歌唱訓練の支援を目指す. 身体的な情報を音響データから判断するのは困難であるため，筋電センサーを喉頭筋に用いることで学習データの品質を改善する. 本稿で用いる歌唱における発声の定義と問題の解決提案を第 2 章で行い，学習データのラベリングのための筋電センサー実験を第 3 章で述べる. 第 4 章では，判別分析手法を用いた歌唱自動評価システムの構築法について述べ，第 5 章で実際の歌唱に対する評価実験の結果を述べる. 実験の結果についての考察を第 6 章で行い，最後に第 7 章で結論と今後の課題について述べる.

2 歌唱発声の特徴と従来手法の問題

本章では，音声生理学の知見を交えながら本稿における声区や発声の定義を行い，自動歌唱評価の提案手法の説明を行う.

2.1 発声の定義

発声に関する定義は発声器官の複雑さから一意に決まっているものではなく，研究者がそれぞれに主張している. 本研究では次のように発声を定義する.

2.1.1 声区の定義

声区は連続した声の音高の区間である. それぞれの声区は声帯から生成される異なる振動パターンから生じる. しかしながら，複数の声区が重なる領域がある. 一般的に男性は 2 つもしくは 3 つの主な声区，(地声区と頭声区，もしくは加えてファルセット) があり，女性は 3 つ (胸声区，中間区，頭声区) 存在している [1] とする主張が多い. 本稿では，声区を次のように定義する.

■地声区 平常の会話に使用する声区. 胸声区と呼ばれることもある. 声帯は弛緩しており，声帯同士が厚く合わさっている. 低音域で使われることが多い.

■頭声区 地声区と音色は異なり，フルートのような音がイメージしやすい. 一般的に言われる裏声もこの声区に属する. ファルセット自体を声区にすることが多いが本研究では頭声区の声種 (声区における発声状態の一つ) として扱う.

2.1.2 発声状態の定義

本研究では，声区をさらに詳細化した発声状態というものを定義する. 声区によって複数存在する発声状態をシステムの判別する対象とする.

■地声 地声区によって生成される発声. 平常時と同じように喉頭はリラックスしている状態である. 通常の会話の際に使われる.

■喉絞め声 地声区によって生成される発声. 換声点と呼ばれる地声区と頭声区の境目に近い高音域を発声する際に現れ，喉頭は非常に緊張した状態にある. 緊張度が高くなるにつれ，声帯の接触面は次第に薄く，小さくなる. 緊張が限界に達したとき，定常な振動をすることができず，声が裏返る.

■ファルセット 頭声区に属する発声. 一般的に裏声と言われた場合，この発声を示すことが多い. 特徴として，高次の倍音成分の割合が少なく，か細くフルートのような音色である. 地声区と頭声区の中間 1 オクターブ程で発声が可能であり，地声区と重なる領域に存在する. 同じ周波数では，地声より音圧が小さく，音高が上がるにつれて音圧も大きくなる.

2.2 歌唱における発声の問題

2.2.1 声区の使用

日本の人気のあるポップスやロックミュージックには，しばしば換声点を大きく超えた高音が要求される. 特に男性歌手の楽曲で見られる. そしてそのような楽曲を地声区のみで歌うことは非常に難しい. なぜなら，換声点付近の音高を発声するためには，呼気量を増やすだけで上昇させることが困難であり，多くの場合は喉の筋肉を過度に締めることで声帯の振動数を上昇させ，それが喉の疲労につながってしまう. そしてその疲労

Supervisor: Prof. Katunobu Itou

表 1. 使用する音響特徴量

特徴量	次元数
基本周波数	1
ゲイン	1
高調波成分	10
スペクトル傾斜	1
メル周波数成分	20
HNR	1
F1F3syn	1
フォルマント周波数	2
フォルマント成分	2
ジッター	1
シマー	1
カートシス	1
スペクトルフラットネス	1

は局所的な声の裏返りや，一時的な声帯の炎症を起こし，結果的に喉が枯れるといった状態に陥る. しかし，プロの歌手は容易に高音域を歌うことができる. それは地声区と頭声区の特性をよく理解し，使い分けているためである. プロの歌手の歌唱技術の中でも，要求された音高に対して，適切な声区を選択することは，喉への負担を減らすことにつながる，重要な技術の一つである. それはまた一般の歌唱者にもあてはまる.

2.2.2 関連研究

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. 歌唱音声の特性は，Singer's Formant[1] が存在すること，基本周波数 (F0) 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 F0 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．歌唱音声自動評価で使われる特徴量は，楽譜情報を用いる場合は音程の一致，リズム感などであり (カラオケ採点システムなど)，楽譜情報を用いない場合ではビブラートや相対音高などである．声質・音色などの特徴量を用いる場合もあるが，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い.

2.2.3 提案手法

病理音声学の分野では，発声について様々な研究がなされている.Marcelo[8] らの研究では，病理音声を適応フィルタを用いて抽出した残差信号から得られる特徴量から病状の識別を行ない，22 の症状のうち 54% を音声信号から識別した.

本稿では，発声状態の 1 つである喉絞め声を，従来の歌唱評価で使用される音響特徴量に加えて病理音声で症状の識別に有効性が確認された特徴量を使用して発声の音響データから自動検出する. 歌唱者が喉に負担をかけているかを判別し，従来の音楽的な情報を用いる訓練ではなく，身体的な特徴に観点をおいた歌唱技術訓練を行うシステムを構築する.

3 歌唱状態の推定

自動歌唱状態判別システムでは，判別分析手法を用いて歌唱発声を判別・評価する. 学習データを構築する際，発声データに対して各発声状態のラベル付けを行う必要がある. この章では音響特徴量を用いたラベル付けと筋電センサーを使ったラベル付けの手法を述べる.

3.1 データの取得方法

本研究では音声から 43 種類の特徴量の抽出を行う. 表 1 に使用したすべての特徴量とその次元数を示し，その特徴を次に述べる.

スペクトルに関連する特徴量

■高調波成分 高調波成分 (倍音成分) は，声区ごとに異なる特徴を示す. 例えば，裏声では低次の高調波成分の割合が多く，喉絞め声では，負担のない地声に比べて第 1 倍音成分の割合が少ない (スペクトル傾斜). また，スペクトルとは異なり，声道特性の影響を受けるが，周波数の影響を受けることはない. 本研究では，第 10 次までの高調波成分を特徴量として使用する.

■メルフィルタバンク成分 メルとは人間の聴覚特性を表す尺度であり，低周波帯域において細かい分解能，高周波帯域においては荒い分解能を持つ. メル周波数は式 (1) で求められる. 高周波数帯域でフィルタの幅が広くなるメルフィルタバンクを用いることで，スペクトル列をメル周波数軸上に等間隔に射影する. フィルタ数を 20 に設定し，フィルタの帯域は音声認識で十分とされている 8000Hz までとする. フィルタの中心周波数はメル周波数軸をフィルタ数と同じ 20 の数に等間隔に分けたものをそれぞれ逆算することで求められる.

$$f_{mel}(f) = 2595 * \log_{10}(1 + \frac{f}{700}) \quad (1)$$

■フォルマント フォルマントは音声スペクトルの包絡ピークであり，声道の形状を表し，個体差や性差で違いが生じる. しかし，発音する音韻が同一であれば，各フォルマント周波数は近いものとなる. 本研究では，第 1，2 フォルマントの周波数 (fPit) とその成分 (fPow) を特徴量として使用する. これらの低次フォルマントは母音を決定するものであり，第 1 フォルマント周波数は口の開き具合，第 2 フォルマント周波数は舌の位置を表すとされている. フォルマントを特徴量として使うことで，判別システムはすべての母音を考慮する.

■HNR Harmonic-to-Noise Ratio(HNR) は音声の周期成分と非周期成分の比で求められ，非周期成分である息漏れ成分がどれほど含まれているか考慮できる. 例えばファルセットは地声区とは違った声帯の振動で比較的息漏れ成分が高い傾向にあり，声区間の判別などに有効であると考えられる.

■F1F3syn 石井 [9] が提案した音響パラメータであり，息漏れ度合いを示す特徴量である. 第 1 フォルマント帯域でフィルタ加工した音声波形の振幅包絡と第 3 フォルマント帯域で通過フィルタ処理された波形の振幅包絡との相関をとることによって求められる.

残差信号から得られる特徴量

人間の声は，声道特性によって特徴付けられる. 声道特性はスペクトル包絡の特徴量によって定量化できる. しかし，発声状態の推定では，基本周波数やジッタなどの声帯振動から生じる喉頭音源の特徴量推定を行う必要があるため，個人性や母音に影響される声道特性をその逆フィルタを用いて取り除き，残差信号と呼ばれる喉頭音源の特徴を表す信号を得る. サンプリング周波数を 16kHz，フィルタリング次数は 14 とし，線形予測法を用いて包絡を推定した.

■ジッター 周波数のゆらぎ成分. ハスキーな声，しわがれた声などはジッターが大きく，後述するシマーとともに音声の粗造性を表す. ピッチ周期の間隔の差の平均を周期間隔の平均で除算することによって求める.

■シマー 周波数のゆらぎ成分のひとつで，ピッチ周期のピーク差の平均をピークの平均で除算することで求める.

■カートシス 音声信号の尖度を示す. 尖度は分布の突起傾向を示す尺度であり，正規分布の尖度は 0 である. 病理音声では尖度が高くなる傾向がある.

■スペクトルフラットネス スペクトルの平面度を表す特徴量であり，スペクトルを特徴付ける度合いである. その音声の調波性の指標となる. 値が 1.0 のとき，ホワイトノイズのようにスペクトルは平面である. 計算はパワースペクトルの幾何平均をその算術平均で割ることによって求める. 本研究では残差スペクトルのスペクトルフラットネスを Gray[10] の手法により求めた.

3.2 音響特徴量による発声状態の分析

学習データのラベリングを行うために，音響特徴量による各発声状態の分析を行なった. 日本語 5 母音，/a，i，u，e，o/の発声を録音し，マイクは被験者の口から 10cm の距離をとった. 発声の音域は，各発声状態が現れる音域とし，声区は別に収録した. サンプリングレートは 16kHz とし，量子化ビットは 16bit，フレーム処理の時間間隔は 30ms とした. 声区は他の声区に重なるため，複数の声区で同じ音高を発声することができる. 重なる音高区間は，一般的に男性では 200Hz から 350Hz(音階では G3 から F4) である. 女性の場合では，胸声と中声は 400Hz(G4)

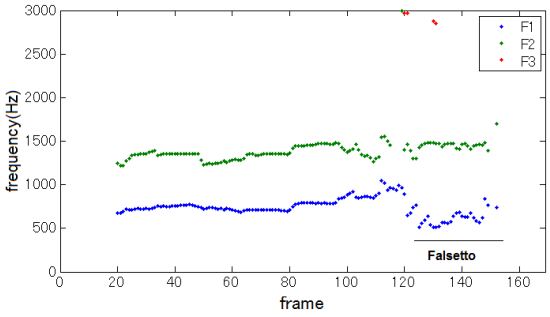

図 1. 地声とファルセットのフォルマント変化

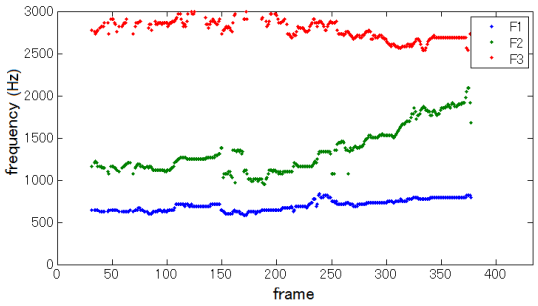

図 2. 母音/a/のフォルマント変化

周辺で重なり，中声と頭声は 660Hz(E5) で重なる. しかしながら，重複区間は個人差によるものが大きいので，音高で一意に決まるものではない.

声区間の分析では，第 1 フォルマントの周波数が同一母音の発声で異なることを観測した (図 1). 図 1 は 1 オクターブの長調音階を発声したものである. 周波数は 247Hz から 494Hz であり，被験者は途中でファルセットに変換している. また，同じ音域ではファルセットのほうが呼気量が少ない関係からスペクトルの高域成分の割合も顕著に異なる結果となった. これらの結果に加えて被験者が意図してファルセットを発声できることから，声区間のラベリングは音響特徴量のみで正確に行うことができる.

喉絞め声の分析では，通常の地声との判別が音響特徴量のみでは難しい. 喉絞め声は，地声による高音域の歌唱中に喉に余計な力が入ってしまうことによって生じ，特に男性の換声点付近において観測される. 意図しない発声のために聴覚的な主観で正確にラベリングをすることは困難である. そこで同一母音の音階発声中の音響特徴量から喉絞めの変化を分析した. フォルマントの変化による分析では，母音/a/や/o/のような口を大きく開ける開口母音では，換声点付近で第 2 フォルマントの上昇を観測した (図 2 ，195 Hz - 391 hz の音階発声). 一方，母音/i/や母音/u/のような口の断面積が小さい閉口母音では，開口母音のようなフォルマントの変化は見られなかった. また，その他の音響特徴量の変化の観察も喉絞め声のラベリングには有効でなかった.

3.3 喉絞め分析のための筋電センサーを用いた実験

喉絞め声のラベリングは音響特徴量のみでは困難である. 喉絞めを判断する際に有効な特徴量変化を観測出来ないため，主観的な情報を用いることとなる. それは被験者の喉への負担度合いの意見や，音階の上昇に伴って起こった声の裏返りの直前に喉絞めのラベルを付けるといったものである. 主観的な情報は推測の領域を出ないため，実際の判別結果の信頼性も低くなってしまう. そこで，筋電センサーを用いることで客観的に発声中の喉の状態を観測し，学習データ作成の為の喉絞め声のラベル付けに有効な手段となるかを検討した. 従来から喉頭の筋肉を分析する際にはしばしば筋電センサーが用いられており [11]， 歌声だけでなく，音声認識も可能 [12] なほど喉頭筋肉群の動きから得られる情報は多い. 検討の前段階として，最初に発声器官の生理学的な機能に注目し，歌唱に使う筋肉の説明と喉絞め声が起きる原因について述べる.

3.3.1 歌唱時に使用する筋肉

発声器官の硬直は，通常の会話時であっても慢性的に行われる. このような発声器官に「歌う」という大きな声の仕事をさせなければならないとすれば，絶対に硬直させなければならない. 喉頭をその中に吊るし保っている弾力的な足場枠である「歌うための筋肉 (喉頭懸垂筋機構)」は，歌い始める瞬間まで一度も本質的に運動させられない. 歌唱に慣れていない人は，この未開発の筋肉系の運動経験が不足しているので，最初に歌おうとする際に最もよく神経支配が行き届いている，「本来担当すべきではない他の筋群」が自動的に喉頭に支えを与えるという仕事を代行してしまう. その代行すべきでない筋群のひとつが，「舌骨筋」である (図 3). 舌骨とともに喉頭が，過度に上方に引き上げられて固定される. これによって声の通り道がふさがれてしまい，その結果必ずある種の狭められた苦しげな声となる [10]. 実験では，この舌骨筋の動きを観測し，喉絞め声の指標となるか検討した.

図 3. 舌骨筋

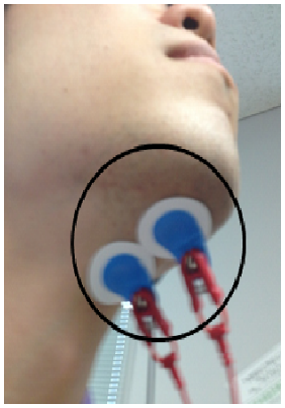

図 4. 筋電センサー部位

3.4 喉頭筋の筋電センサー分析

ここでは筋電センサーを用いた実験分析について述べる. 筋肉は関節をはさみ，骨と骨とを連絡し，その収縮によって運動を行う. そして収縮時には，電気的な筋放電と呼ばれる放電が行われる. 逆に筋肉が弛緩しているときには放電は見られない. この筋肉から発する放電を経皮的に電極で記録したものが筋電図となる. 筋電図の観測によって喉の筋肉の負担度合い (喉絞め度合い) を観測する.

3.4.1 実験環境

ロジカルプロダクト社製のワイヤレス表面筋電位ロガーと湿式 2 極筋電センサーを使用し，歌唱時に使うべきではない筋群，舌骨筋の硬直度を計測した. 被験者の舌骨筋の動きを調べるために，粘着性の電極を顎の先端よりと喉頭よりの 2 箇所に貼り (図 4)，サンプリング周波数 1000Hz，中心電圧 2.5V で計測を行なった. 通常会話の舌骨筋の負担度を見るために 1 分間ほどの音読を行い，周波数上昇の影響を見るために換声点を含む音階発声を行なった. また，実際の歌唱での筋肉の負担度も測った. それぞれの実験結果に対して筋電図を分析し，ラベル付けに有効か考慮する. 音階発声は，長音階の発声を日本語各 5 母音，周波数幅を 130Hz(C2) から 392Hz(G3) に設定した. 被験者は喉絞め声が観測されやすく，声枯れが観測しやすい男性のみを対象とした. 実際の楽曲の歌唱については，換声点付近を最高音となるようにキーを変更したものを歌わせた. 準備として，故意に舌骨筋に力を込めたり嚥下を行うことで筋電センサーに反応があるか確認し，十分に機能していることを確かめた上で実験を行なった. 本研究では，喉絞め声が女性よりも観測されやすい男性に限定して実験を行なった. 性差で喉頭の構造は大きく異なり，女性は通常の会話でも男性のファルセットのような発声を行なっているため，喉絞めの観測が難しい.

3.4.2 実験結果

音階発声では，換声点付近 (5000cent 周辺) になると電圧が上がるといった傾向が各母音で観測できた (図 5). 筋電図の縦軸は中心電圧 2.5V からの距離 (単位 V) を表す. 図 5 の発声開始時点で電圧が上がっているのは，母音/a/の発声をするために口を開け，その際に舌骨筋を使用したからだと考えられる. 各母音とも負担のかかっていない低いピッチでの発声では，中心電圧から ±0.1～0.2V ほどしか放電が観測されなかった. 換声点付近の聴覚的・主観的なラベル付けで喉絞め声と判定された箇所では，中心電圧から ±0.7～1.0V の放電を観測した. 平常発声時

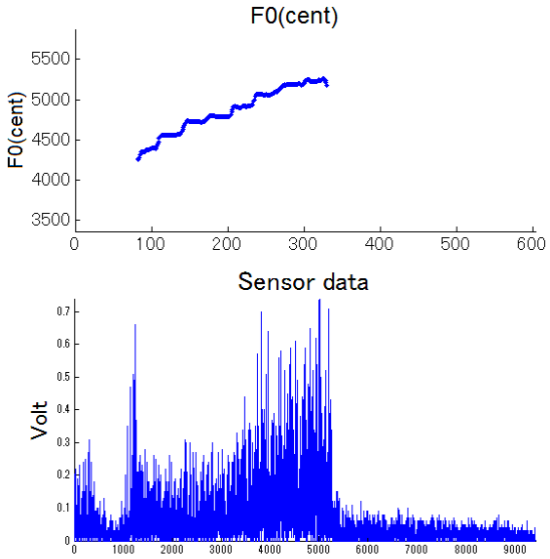

図 5. 地声母音/a/の音階発声時の基本周波数と筋電図

表 2. 表面筋電位の電圧差 (V)

音声データ	被験者 1	被験者 2
嚥下	1.0-1.3 V	1.0-1.2 V
音読	0.3-0.5 V	0.3-0.6 V
歌唱	0.3-1.0 V	0.3-0.9 V
音階発声 (地声)	0.1-0.5 V	0.1-0.4 V
音階発声 (地声，換声点)	0.7-0.9 V	0.7-1.0 V
音階発声 (裏声)	0.3-1.6 V	0.2-0.4 V

の 3 倍から 4 倍の放電であると喉絞め声の可能性が高い. また，裏声の音階発声を行なったところ，裏声発声に慣れている被験者は中心電圧との差が最高で約 0.4V だったのに対し，裏声発声を得意とせず，息漏れ成分が多い被験者の場合は最高電圧差 1.6V を記録した.

音読と歌唱では音素単位で電圧に反応があり，口の開閉に使う筋肉として舌骨筋が働いていた. このことから，普段話しているときにも発声器官を硬直させていることがわかる. 歌唱でも同じ現象が観測できたが，表面筋電位に大きな違いがあり，歌唱のほうが筋肉の硬直度が高いことがわかった.

3.4.3 考察

実験結果より，舌骨筋の筋電を測ることは喉の負担の計測に有効であったため，筋電センサーを用いることで喉絞め声のラベル付けの精度向上が期待できる. 舌骨筋に硬直が見られた場所は聴覚的な印象に基づいた喉絞め声のラベル付けを行なった箇所と重なる部分が多く，他の音響特徴量の変化として母音によっては第 2 フォルマントの上昇があった. 発声中に変化するということは，声道の形そのものが変化している可能性が高い. 意図しない舌の動きが舌骨筋によって引き起こされたのならば，喉絞め声となる箇所で起きることは充分考えられる. 音読では，文末や句読点の直前の音素を発声する際に電圧が一瞬上がるといった傾向が見られた. 呼気量が少なくなった中で音高を低くしないために喉に力を入れてピッチの低下を防いでいるためと考察できる.

3.5 喉絞め声のラベリング

筋電による舌骨筋の分析によって，喉絞め声の出現と筋電位の増加は関連があることがわかった. 本研究では，地声区のラベリングには筋電を用いる. 閾値設定は筋電図の結果より各発声ごとに手動で行う. 初めに筋電図の包絡を求め，筋肉の収縮のおおまかな変化を求める. 筋肉の収縮は筋電位の振幅ピークに現れ，それ以外の値は大きな意味をなさないためである. 包絡推定の推定として，まず筋電位の周期ごとのピークを推定する. 推定したピークの間を線形補間した後，200ms の時間幅で移動平均法による平滑化を行い，筋電図の包絡を求める. 時間幅は，局所的でなくまた大域的な包絡ともならないように試行をした上で決定した. 音階発声において包絡の上昇が始まる音高は，舌骨筋が硬直し始め，喉絞め声が開始されたと考えることができる. 本研究では，音階発声の音符長ごとの筋電の包絡平均を求め，上昇が始まった箇所を歌唱者の喉絞め声の閾値とした.

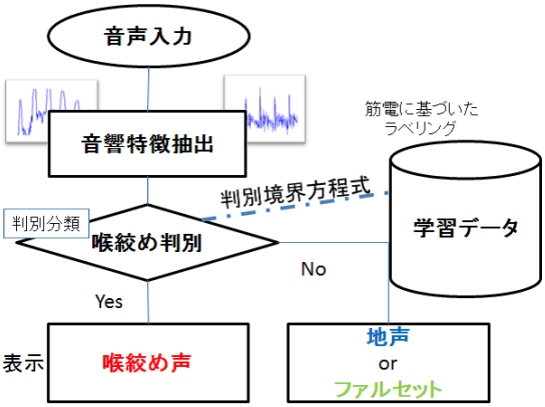

図 6. 判別システムのフロー

4 発声状態の判別

本章では，発声状態の判別方法に付いて述べる. 図は判別システムのフロー図である (図 6). 歌唱データを入力し，抽出された特徴量に対し学習データに基づいた線形判別を行う. 学習データは筋電を参照してラベリングを行なったものを使用した. 判別分析の手法として，本研究では単純ベイズ分類器とマハラノビス距離による判別を用いた. 筋電センサーは学習データの構築にのみ用い，実際に歌唱発声の発声状態を判別する際には用いない.

4.1 特徴選択による次元削減

特徴抽出によって得られた 43 種類の特徴量に対して特徴選択を行い，学習モデルの構築を行った. 特徴選択を行うことにより，次元の呪いの効果の緩和や過適合問題の回避，モデルの可読性の向上などをはかる. 特徴選択の手法として前向き逐次特徴選択法を選択し，分類アルゴリズムとして単純ベイズ分類による線形判別，2 次判別，マハラノビス距離による判別の 3 つを用いた. 単純ベイズ分類器は確率モデルに基づく分類器であり，教師あり学習の設定で効率的に訓練することが可能である. それぞれの分類手法で作成された学習データの妥当性を計るために，それぞれの分類アルゴリズムを用いて交差検定を行なった. 交差検定の手法として，k-分割交差検証を用いて行なった. 標本群を k 個に分割し，そのうちの 1 つをテスト事例とし，残る $k-1$ 個を訓練事例として k 回検定を行い，得られた誤判別率の結果の平均を学習データの誤判別率とする.

4.2 前処理

前処理として，初めにデータに対しスケール調整を行う. スケーリングをすることによって，値の取りうる範囲が大きい特徴が支配的になることをさけられ，また情報落ち誤差などを発生させないために有効である. スケーリングには，標準化された Z スコアを用いた. また，子音区間の削除を行うための処理として，有声区間の推定をした後にメル周波数ケプストラム係数の差分によって子音判別を行い判別候補から除外する.

5 評価実験

5.1 実験条件

実際の歌唱音声などに対し，各手法による判別分析実験を行なった. 3 名のアマチュアボーカリストを含む 9 名の男性被験者を用意し，換声点を含んだ音階発声と日本の有名なロックミュージック 2 曲を歌唱させた. 歌唱者はヘッドフォンを装着し，口からマイクまで 10cm の距離を置いて発声した. 音声と筋電センサーを同時に取得し，喉絞め声の正解データとして筋電を使用した. 判別率の計算は音符単位から求め，音符単位の推定は，音符内で最も割合いの多い状態をその音符の推定された発声状態とした. 音符のセグメンテーションは楽曲内のそれぞれの音符の発音区間と同一とする. 音階発声の音高区間は，190Hz から 660Hz とし，換声点を含む値を設定した. 実際の

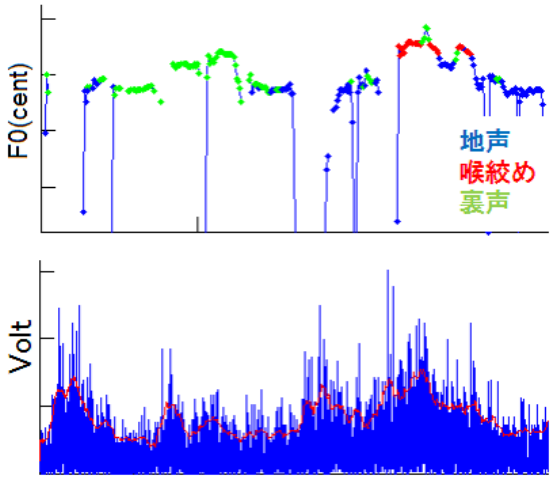

図 7. 線形判別による発声判別結果 (上): 筋電図 (下)

表 3. 判別正解率 (%)

判別手法	特徴数	判別率 (評価)	判別率 (学習)
線形判別	5	91.5	92.7
2 次判別	16	82.2	93.0
マハラノビス	23	73.5	94.3

歌唱に使用する楽曲は男性の換声点を越えた音を多く含んでおり，地声区のみで歌おうとすると喉に負担がかかりやすく，歌唱の熟練度の違いが現れるように設定した. また，1 音符ごとの発声時間を長くし，定常な発声状態の分析をするために楽曲の 1 分あたりの四分音符数 (BPM) を 100 とした．子音の影響を考慮するために，子音ありの歌唱と母音のみの歌唱の両方を行なう. 楽曲のキーを複数に設定し，それらの最高音は 350Hz から 554Hz($F3$ から $D^{\flat}4$) とした.

5.2 実験結果

5.3 筋電との比較

判別結果と筋電との比較を行った．発声状態が負担のない地声なら青色に，喉絞めなら赤色，ファルセットを使用している場合は緑色となる. 地声のみの音階発声では，換声点に近づくにつれ筋電の値は上がり，実際の判別結果も喉絞め声と判別された.

誤判別は喉絞め声を地声と誤った場合とファルセットを喉絞め声と誤る場合が多かった．図 7 の上図が線形判別による判別結果，下図が筋電である. 音声データは母音のみ歌唱であり，筋電が上昇する部分で判別結果が喉絞めを示す赤色に判別された. 一瞬裏返ってしまうような箇所もファルセットと推定された. 子音の影響を受けた母音がファルセットに誤判別される結果が見られた. 特に無声音化する子音の直後で多く，短い音符内で起こると音符内の有声母音の割合が減ってしまうため誤判別が起きた.

5.4 判別結果

各判別手法の比較では，特徴数が多いものほど判別率が落ちる傾向にあった (表 3). 評価データでは，線形判別が最も高い判別正解率となり，音符単位の判別率は 91.5% となった.

判別結果では学習データ交差検定時の判別率はそれぞれの手法で高精度だが，評価データに対しては線形判別が最も良い結果となった. 交差検定時の判別率が高い 2 次判別やマハラノビス距離による判別は新たな被験者に対して結果が悪く，特徴数が適切でないために学習データに過適合した可能性がある.

5.5 適合率と再現率

判別手法の有効性は判別正解率だけでなく，適合率 (P_i) と再現率 (R_i) からも評価できる．ここで，$i = modal, tightened, falsetto$ はクラスを表す変数であり，$class_{modal}$ は地声サンプルが属するクラスであり，$class_{tightened}$ は喉絞め声，$class_{falsetto}$ はファルセットのサンプルが属するクラスとなる．P_i, R_i はそれぞれ以下のように定義する.

$$P_i = \frac{class_i\text{へ正しく識別したサンプル数}}{class_i\text{として識別したサンプル数}} \times 100 \qquad (2)$$

$$R_i = \frac{class_i\text{へ正しく識別したサンプル数}}{class_i\text{の総サンプル数}} \times 100 \qquad (3)$$

評価データに対して判別正解率が最も高い線形判別の結果を表 4，5 に示す.

表 4. 線形判別の適合率 (%)

Dataset	P_{modal}	$P_{tightened}$	$P_{falsetto}$
音階発声	97.1	90.3	95.3
歌唱	93.1	88.2	91.5

表 5. 線形判別の再現率 (%)

Dataset	R_{modal}	$R_{tightened}$	$R_{falsetto}$
音階発声	96.5	95.3	90.9
歌唱	96.1	95.4	82.8

音階発声と歌唱で共通するのは，$P_{falsetto} > R_{falsetto}$ となることであった．これは，ファルセットと判別された箇所は高精度でファルセットであるが，本来ファルセットと判別される箇所がそれ以外の発声状態に判別されたことを意味する．特に，歌唱において $R_{falsetto}$ が低い結果となった．これは，歌唱者が高音を出すために裏声を使ったがファルセットのような声質ではなく，喉絞めに近い発声などで誤判別が起きたことと関連付けられる．感覚的に身についた発声は高音域の歌唱を可能とするが，その際の舌骨筋の収縮は喉絞め声と変わらず，喉への負担が軽くはない．そのため，長時間の歌唱では声が枯れてしまう回避すべき発声である．また，共通して $R_{tightened}$ が高いことから，回避すべき発声である喉絞め声は高精度に検出できたといえる.

6 考察

この章では実際の歌唱音声に対する判別分析実験で得られたいくつかの課題についての考察を述べる.

6.1 判別に有効な特徴

各状態の判別への特徴の考察を行うために，学習データに対し境界面を決定する方程式の係数について考察を行なった. 係数の可読性を上げるために最も特徴数が少ない単純ベイズ分類器による線形判別の結果を用いた. 表 6 は 2 つの発声状態の境界面を決定する特徴量の線形項の係数であり，それぞれの特徴がどのように判別に関わっているかを示す.

■地声-裏声の判別　地声-裏声間の判別には，基本周波数と中心周波数が 6144Hz である第 19 メルフィルタバンク成分 (第 19MFB 成分) が大きく影響していることがわかる. しかし，これらの符号が違うことから，基本周波数が低く高周波数成分が大きいときには地声，そうでない場合には裏声と考えられる. 負担のない地声と裏声は音域が重なりにくく，スペクトルの形状がそもそも大きく違うことからこのような係数の結果になったと考えられる.

■裏声-喉絞め声の判別　裏声-喉絞め声の判別には特に第 19 メルフィルタバンク成分が大きく影響している．これは，同じ音程では地声区のほうが頭声区よりも高域成分が大きいためであり，それは呼気量の関係からと考察できる. 音声生理学上，地声で発声できる場合のほうが音圧が上がり (極端な喉絞め声でない限り)，音圧があがると高域成分の割合が高くなるからである．次点で中心周波数が 575Hz の第 5 メルフィルタバンク成分 (第 5MFB 成分) の係数が逆方向に大きい．換声点付近の発声では，この周波数区間は，基本周波数成分が大きな割合を占める．喉絞め声では基本周波数成分が他の倍音成分よりも低くなることが影響していると考えられ，スペクトル傾斜が判別に有効であることが考察できる.

■地声-喉絞め声の判別　地声区で負担がかかっているかそうでないかの判別には，スペクトルの傾斜と基本周波数の係数が大きい．基本周波数の係数が 3 番目であるということから，換

表 6. 発声状態の判別境界方程式の線形項係数

地声-裏声		裏声-喉絞め		地声-喉絞め	
基本周波数	-2.05	第 19MFB 成分	-4.69	第 19MFB 成分	-2.79
第 19MFB 成分	1.90	第 5MFB 成分	2.29	第 5MFB 成分	2.56
第 5MFB 成分	0.27	ゲイン	0.74	基本周波数	-2.19
スペクトルフラットネス	-0.09	スペクトルフラットネス	-0.35	ゲイン	0.82
ゲイン	0.08	基本周波数	-0.13	スペクトルフラットネス	-0.45

声点などの音高の情報よりスペクトルによる音響的・聴覚的な情報のほうが喉絞め声の判別に効くことがわかった.

病理音声などで有効とされた残差信号により求められた喉頭音源波形の時間方向へのゆらぎを表し，音声の粗造性に関連するジッタやシマー，カートシス (尖度) は歌唱の発声状態の判別には大きな影響は与えなかった. これは，本稿で定義した 3 状態の判別ではなく，更に細かな状態を判別する際に有効になると考えられる. 例えば，歌唱音声のエッジのかかり具合や，局所的なシャウトなどの検出をする際には音声の粗造性が判別に大きく関わってくる.

6.2 誤判別の考察

判別率の算出を音符単位で行なったが，処理フレーム単位では判別結果はどの手法も劣る結果となった. 学習データを音階発声から作成したため，急激な音程の上昇・下降，母音の変化，音量の急激な変化，子音による母音への影響，発声のアタックなどいった，歌詞の語頭や語尾で起こる変化に対応できなかった.

本研究では頭声区の発声状態をファルセットのみとし，学習データの収録時にもファルセットの音階発声を用いたが，実際の歌唱では，頭声区ながらもファルセットより芯のあるはっきりとした音量の大きい裏声が現れ，その際には筋電の値では負担がかかっていないにも関わらず，いくつかの発声は喉絞め声と誤判定されてしまった. このような声はヘッドボイス，もしくはミックスボイスと呼ばれるようなもので，頭声を駆使して高音域を地声のように歌うものである. ファルセットよりも高次まで高調波成分があり，したがってスペクトルの形は地声，音域も踏まえると喉絞め声に近いものになる. 従って，頭声区の発声状態を増やす必要がある. 学習データを如何に構築するかが今後の課題のひとつとなる.

6.3 頭声区での喉絞め

地声について負担がかかっているかどうかを判別するシステムを作成することを目的としており，負担がかかるような声であればファルセットを使うことを推奨していた．同じ音域であれば，ファルセットのほうが負担が少ないというものであった. ある程度その推測は実験でも確かめられたが，被験者・音域によってはファルセットでも喉に負担がかかる現象が筋電センサー実験で観測された. これによって，頭声区の場合でも負担がかかっているか判別する必要がある.

7 結論

本研究では男性の高音域の発声を判別・評価するシステムを構築した．歌唱で喉が枯れてしまう，もしくは不安定で裏返ってしまう歌唱者にとって有効であった. システムは筋電センサーにより得られた学習データをもとに歌唱音声を判別し，喉絞め声かどうかを決定した．筋電を参照したラベリングと音響特徴量のみを参照したラベリングの比較では，単純ベイズによる線形判別の結果において 6.1% の判別率の差があり，筋電を参照することで判別結果の精度が向上した. 歌唱音声への評価実験では，音符単位で 91.5% の判別率を得，音響特徴量のみで喉絞めの判定が可能であることを示した．歌唱における 3 つの発声状態の判別には，基本周波数だけでなく，メルフィルタバンク成分で表されるスペクトル傾斜が大きく影響することがわかった．判別誤りがミックスボイスや喉絞めの裏声など学習データとして収録されていない発声で起こったが，喉絞め声を多く使用してしまう被験者の発声を高精度で判別することができた．喉絞め声と判別された場合，裏声やキーの変更をすることで，喉への負担を下げることが実際に確認でき，これにより被験者は喉を枯らすことがなく，長時間の高音域の歌唱を行うことが出来た.

システムの向上のために，いくつかの課題の解決が必要となる. まず，ミックスボイスなどの発声状態の追加をし，より良い学習データを作成することがある．これにより回避すべき発声の判別が出来るだけではなく声楽上良いとされる発声が判別でき，歌唱訓練システムの向上を図ることができる．また，喉頭の構造が異なる女性の学習データを作り，性差を考慮できるシステムを目指す. これらの課題を克服するために，さらに多くの様々な音声を集める必要がある.

参考文献

[1] Sundberg, J. “The Science of the Singing Voice”, Northern Illinois University Press, p.226, 1987.

[2] 矢田部, 遠藤, “歌声の基本周波数の動特性”, 日本音響学会, 研究発表会講演論文集, 1998(2), 383-384, 1998-09-01 , 1998(2), 383-384, 1998-09-01

[3] 辻, 赤木, ”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

[4] Saitou, T., Unoki, M., and Akagi, M. : “Extraction of F0 dynamic characteristics and development of F0 control model in singing voice”, Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan, July 2-5, 2002 EXTRACTION.

[5] Ohisi, Y., Goto, M., Itou, K., and Takeda, K. : “Discrimination Between Singing and Speaking Voices”, INTERSPEECH-2005, 1141-1144, 2005.

[6] 津田 弘樹 , 森山 峻 , 福間 彰, “3D 解析による歌声の評価に関する研究”, 電子情報通信学会ソサイエティ大会講演論文集, 情報・システム, 461, 1996 年

[7] 中野, 後藤, ”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

[8] Marcelo, de, Oliveira Rosa., Jose, Carlos, Pereira., and Marcos, Grellet. “Adaptive Estimation of Residue Signal for Voice Pathology Diagnosis”, 0018-9294/00, VOL. 47, NO. 1, IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING 2000.

[9] 石井カルロス寿憲, “息漏れ自動検出における音響パラメータの提案”, 電子情報通信学会, TECHNICAL REPORT OF IEICE, SP2004-56, 2004

[10] Gray, A., Markel, J, “A Spectral-Flatness Measure for Studying the Autocorrelation Method of Linear Prediction of Speech Analysis”, IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, VOL. ASSP-22, NO. 3, JUNE 1974

[11] 平野実, “歌声の調節機構”, 日本音声言語医学会, 音声言語医学 11(1), 1-11, 1970

[12] Grandori, F., Pinelli, P., “Multiparametric Analysis of Speech Production Mechanisms”, IEEE ENGINEERING IN MEDICINE AND BIOLOGY 1994, 0795-5175.

[13] Frederick, H. and Mariling, Y. -R. “Singing : The Physical Nature of the Vocal Organ”, p. 40, 1965