

喉頭の運動に注目した歌唱音声の自動判別と評価*

平山健太郎 (法政大学情報科学研究科), 伊藤克亘 (法政大学情報科学部)

1 序論

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。特に男性の歌唱者にとって、そのような楽曲を地声区のみで歌うことは難しい。なぜなら、換声点付近の地声での発声は呼気量のみで周波数を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絞めあげてしまい音高制御が難しくなり、また疲労につながってしまう。

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 [1]。歌唱力自動評価の研究で使われる特徴量は、音程の一致やリズム感、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における喉頭の状態に注目した歌唱力自動評価の研究事例はなかった。

本稿では、音響解析に使われる特徴量に加えて音声病理学の分野で効果がある声帯振動に関する特徴量 [2] を使うことで、声区や発声における喉の状態を識別することを考える。発声状態の1つである喉絞め声を分析し、歌唱者が喉に負担をかけているかどうかを判別し、従来の音楽的な情報を用いずに歌唱評価を行うシステムを構築する。

2 声区判別のためのデータの収集

本研究で判別対象とする発声は、先行研究 [3] と同様に次の3つの発声とする。喉に負担のない“地声”、負担のかかった“喉絞め声”、地声とは声区の違う“ファルセット”である。

学習データとして、男性被験者6人、日本語5母音、/a,i,u,e,o/の音階発声を録音した。音階発声とは、母音を変えずに長音階に沿って1オクターブ上げる発声である。発声練習でよく使われる。発声をそれぞれの声区ごとに収録し、サンプリングレートは16kHzまでダウンサンプリングした。フレーム処理の時間間隔は30msとし、マイクは被験者の口から10cm距離をとった。

37の音響特徴量を用いて、発声データから特徴抽出を行う。表1に特徴量とその次元数を示す。そのうちジッターとシマー、スペクトルフラットネスが声帯振動から得られる特徴量であり、声帯振動が異なる声区間の発声で違いが観測できる。

2.1 筋電を用いたラベル付け

音響特徴量のみを用いた学習データのラベル付けでは、喉絞め声のラベル付けの際に実際に喉へ負担がかかっているか判断することが難しい。先行研究 [4] と同様に客観的に発声中の喉の状態を観測できる筋電センサーを喉絞め声のラベル付けに用いた。

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
自己相関係数	2
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
スペクトルフラットネス	3
MFCC	12
Δ MFCC パワー	1

3 判別システムの構築

この章では、判別処理の概要を述べる。初めに、歌唱データを入力し、音響特徴量を抽出する。抽出された特徴量に対し、学習データに基づいた判別分析を行う。判別分析の手法として、本研究では単純ベイズ分類器とサポートベクターマシンを用いた。筋電は学習データの構築に用い、実際の歌唱発声を判別する際には使用しない。

前処理として、データに対しスケール調整を行い、次に主成分分析を用いて次元圧縮を行なった。情報量として十分と考えられる累積寄与率が80.0%を超えており、それ以降の主成分の追加に大きな判別率の差が見られない第20次主成分まで次元を圧縮した。

3.1 判別手法

判別手法では、線形判別と非線形判別を行なった。線形判別は単純ベイズ分類器を用いて行なった。単純ベイズ分類器は確率モデルを使った分類器であり、教師あり学習の設定で効率的に訓練することが可能とされている。パラメータの推定には最尤法を用いた。また、サポートベクターマシン (SVM) を使用して非線形判別を行なった。カーネルはガウスクーネルを用いた。判別クラスを3クラスとしているので、SVMで解くために一対他分類手法を用いた。

4 評価実験

実際の歌唱発声に対し、判別分析の実験を行なった。音符ごとの判別推定結果の最頻値を判別結果とし、正解データとして同時に収録した筋電データを参照した。3名のアマチュアボーカリストを含む9名の男性被験者に対し、換声点を含んだ音階発声と2曲の日本のポップスの楽曲歌唱を行なった。判別率を表2に示す。

4.1 音階発声実験

地声のみの音階発声では、換声点に近づくにつれ筋電図の値は上がり、実際の判別結果も喉絞め声と判別された。しかしながら、筋電の上昇点と喉絞め声と判別された点が必ずしも一致はせず、喉絞めが始まる直前

*Discriminant Analysis of Singing Voice Focused on The Movements of The Larynx. : Kentaro Hirayama. (Hosei University) et. al.

表 2. 各手法の判別率 (%)

発声	単純ベイズ	SVM
音階発声	88.0	86.5
歌唱 (子音あり)	86.5	88.4
歌唱 (母音のみ)	91.7	93.2

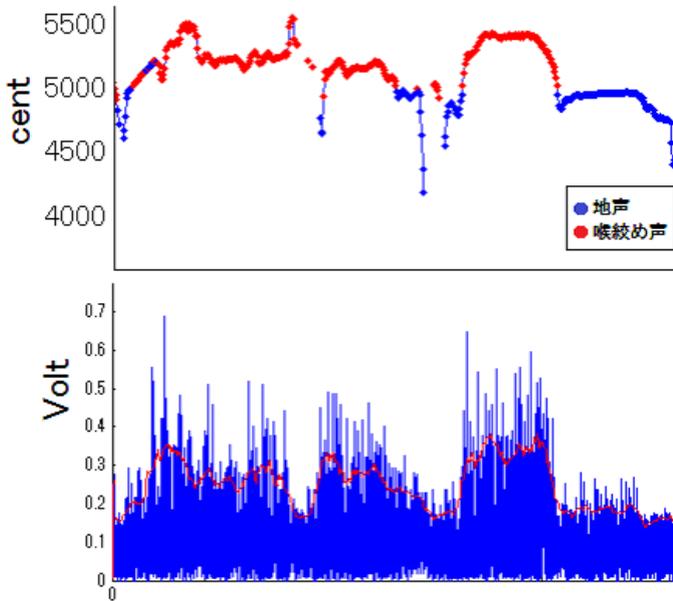


図 1. SVMによる発声判別結果 (上): 筋電図 (下)

の負担のかかっていない地声が喉絞め声とされる誤判別が起こった。

筋電図の値を正解としたとき、音階発声における最高判別率は 88.0% であった。

4.2 歌唱音声実験

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るものを選択し、歌唱に慣れているかどうかの違いが顕著に現れるように設定した。SVM の判別率は単純ベイズ分類器より良い結果となった。図 1 は地声で母音のみの歌唱発声を SVM で判別した結果であり、上図の縦軸は周波数 (cent)、下図の縦軸は同一時刻での筋電図の値を表す。喉絞め声の判別結果は筋電と一致していることがわかる。子音の影響をどちらの手法も受け、母音のみの歌唱と比べると判別率が悪い。

5 考察

音階発声実験の誤判別の結果から、喉絞めが始まる箇所の音響的な特徴はそれ以前の負担のかかっていない地声とあまり変わらないと考えられる。喉絞め声をラベル付けする際の閾値を考慮し直す必要がある。また、楽曲歌唱では子音の影響を受けると裏声と誤判別されることがあった。また、母音の無声音化でも誤判別を起こした。子音や無声音の箇所を推定し、判別対象外とする必要がある。

表 3 は第 1,2 主成分において上位の係数を占める特徴量とその負荷量である。第 1 主成分は低次の高調波成分と第 1 次の MFCC を含み、低域成分の割合が関係している。第 2 主成分では 2 次以降の MFCC や第 2 フォ

ルマントの負荷量が高いことから口の形、特に舌の位置によって決定される周波数帯域にに影響していると考えられる。

喉絞め声と判別された箇所が多かった場合に、裏声やキーの変更をすることで筋電が下がり喉への負担を下げるのが実際に確認できた。これにより喉を枯らすことなく長時間の歌唱を行うには、声区の変更をすることが有効であることが示された。

本研究では頭声区をファルセットのみとし、学習データの構築にファルセットの音階発声を用いた。しかし、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発声のみられ、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。

表 3. 第 1,2 主成分の因子負荷量

第 1 主成分		第 2 主成分	
1 次 MFCC	0.34	4 次 MFCC	0.36
2 次高調波	0.31	3 次 MFCC	0.32
4 次高調波	0.30	2 次 MFCC	0.32
1 次高調波	0.30	第 2 フォルマン	-0.26
3 次高調波	0.29	5 次 MFCC	0.24

6 結論と課題

本研究では男性の高音域の発声を判別するシステムを構築した。特に、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人にとって有効である。母音のみの歌唱音声に対しては SVM を使った判別で 93.2% の音符単位の判別率を得た。今後の課題として、発声状態の追加や子音・無声音化した母音の識別、実際にシステムを使った歌唱訓練の評価などを行う。

参考文献

- [1] Nakano, T., Goto, M., and Hiraga, Y.: "An automatic singing skill evaluation method for unknown melodies using pitch interval accuracy and vibrato features", paper 1854-Wed2A10.4, INTER-SPEECH, 2006.
- [2] 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.
- [3] Hirayama, K. Itou, K.: "Discriminant Analysis of the Utterance State while Singing", IT-5410, The 12th IEEE International Symposium on Signal Processing and Information Technology, 2012
- [4] 平山健太郎、伊藤克亘、"筋電センサーを用いた歌声分析のための喉頭音源分析", E-010, 第 11 回情報科学技術フォーラム, 2012