

## 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 (法政大学大学院情報科学専攻), 伊藤克亘 (法政大学情報科学部)

### 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は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。その場合には, 高音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り, 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研究が行われてきた [1,2,3] 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から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。

### 2 高音域発声の調査

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, 男性歌唱曲でもサビ部分で男性の換声点(地声と裏声の境界)である 350Hz 周辺を超えるものが多く, 1つの声区で容易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。結果では男性の約 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。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, 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。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。

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, 男性被験者3人から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させたものを使った。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。

#### 2.1 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のこと(声の基本周波数においては重複が生ずる)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。声区の違いを分析した結果, 主に低域の倍音成分割合と第1,2フォルマントが変わることがわかった。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。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をと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1,2)。

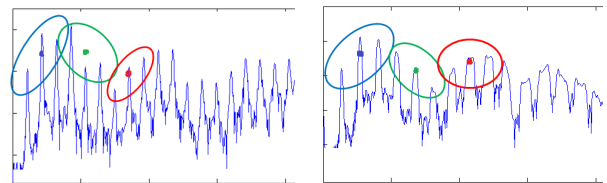


図1. 地声のスペクトル 図2. 裏声のスペクトル

#### 2.2 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/などの開口母音においては, 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。しかし閉口母音 (/i,u/) では共通の変化が見られなかった。従って, 開口母音については第2フォルマントが上がったところ, 閉口母音については換声点付近かつ開口母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と学習データの構築方法を述べる。

#### 3.1 処理の流れ

残差信号:人の声は声道特性によって個人の特徴が付加される。しかし歌唱において個人の特徴は余分な特徴量となってしまう。このため, ピッチなどのピーク判定の際には残差信号を用いる。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 係数を取得し, その2つを用いた逆フィルタから残差信号を得る。

\* Discrimination Method of Voice Register and Voice Quality in high pitch for Popular Singing by Kentaro HIRAYAMA. (Graduate School of CIS, Hosei University) et. al.

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。また、相互相関で得られたピッチを使い残差スペクトルの倍音ピークを取得し、原信号の倍音成分を得る。

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、特徴量として第 1,2 フォルマント周波数を使う。

### 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の差もあるが、高音域での声質が被験者で大きく違った。

### 4.1 歌唱に用いた曲

被験者に歌わせた曲は日本の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」の部類であり、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。歌わせ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。月光花のサビ部分のカラオケデータを原曲キーから上下 5 つ分のキーの曲を作成し、それをヘッドフォンから流して歌わせた。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。

### 4.2 音階発声の判別結果

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る結果となった。

### 4.3 歌唱の判別結果

例として男性被験者の歌唱データを判別した結果を図 3 に示す。また、縦軸は各音程の距離を一定にするために cent 表示にした。男性被験者の図はキー + 2 で歌わせたものである。最高音 B3(495Hz)、歌詞「ら」の部分でファルセットを使用している。地声で A3(440Hz) まで大声量で出せるのが特徴の被験者だったが、B3 以上は裏声にしないと発声出来ず、声区を融合しているものではなかった。

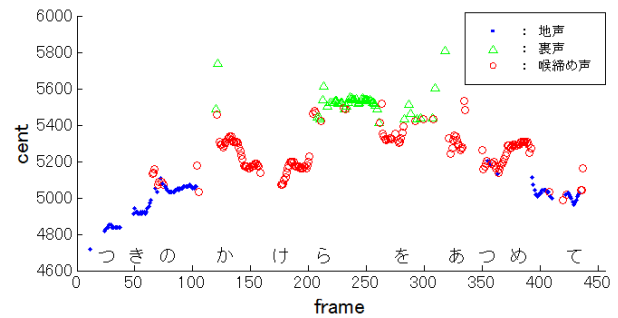


図 3. 男性被験者の判別結果

### 4.4 判別結果の考察

地声と裏声の判別は高精度であった。しかし、裏声として収録したデータがファルセットのみだったので、ファルセットとは違った裏声区の声質となると判定結果が喉締め声となってしまった。課題曲となったアーティストのボーカルも裏声区を使用していると思われるが、決してファルセットのように聞こえない。これは、サンプルを収集する際に地声と裏声の 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

男性のみを対象とした喉締め声や裏返りが起こる高音域で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た。歌唱音声に対する音符単位の識別率は 93.18% であった。しかし改善点も見つかり、1 つとして判別分析の際のグループを増やすことがある。判別分析の考察の際にも述べたが、裏声区を使っている声質は様々である。ロックやメタルなどでよく聞く甲高い声から薄くやわらかい感じのファルセットなど、同じ裏声区でも多数の声質があることは明白でありそれぞれをグループとしにくい限り正しい高音域の評価システムを構築することはできない。また、比較的低いファルセットの際に高次倍音の推定精度が落ちてしまう問題がある。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ぼやけてしまうからである。カラオケルームなどの雑音環境での使用を考えるとより高次の倍音成分の推定は難しくなる可能性がある。ピッチ推定の手法を改める必要がある。有声区間の推定にもまだ問題があり、歌唱の際に誤推定が多く見られたので精度を上げる必要がある。さらに女性ユーザーへの対応も検討中である。

### 参考文献

- [1] Sundberg, J. "The Science of the Singing Voice", Northern Illinois University Press, p.226, 1987
- [2] 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
- [3] 平山他"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", 情処学第 73 回全国大会, 3R-9, 2011