

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*

平山健太郎 (法政大学情報科学部), 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、無理に喉を締め上げて歌われることがある。そうすると、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なかった [1]。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の適切な発声の評価するシステムを構築することである。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、音響特徴量との関連付け [4]、歌唱音声の音響解析に基づく歌唱力評価の考察 [5] などの研究事例があるが、本研究では基本周波数と倍音成分の特徴量を元に判別した。

2 適切な歌唱

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ることである。負担がかかっているかどうかは適切な声区を選択できているかどうかで決まるので、声区のしくみを調べる必要がある。

2.1 声区とは

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。

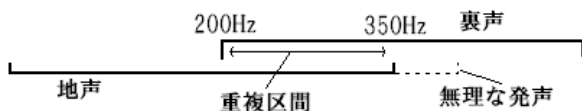


図 1. 男性の声区

今日のJポップやJロックなどのポピュラー音楽では、

Construction of high pitched sound region utterance evaluation system intended for singing of popular music by HIRAYAMA Kentaro. (Information Technology Department, Hosei Univ) et. al.

男性歌唱曲でもサビ部分では男性の喚声点周辺である350Hz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くなっている。ゆえに、無理に地声で張り上げるとすぐに声が枯れてしまう。求められた発声周波数に対する適切な声区を選択ができているかが重要になる(図1)。

2.2 声区の違い

評価システムで評価する際に対象となる歌手ではない一般の人に協力してもらい、実際に録音した歌声を分析した。

2.2.1 歌手でない人の声区の違い

プロの歌手ではない人の地声声区/a:/と裏声声区/a:/の発声で、先行研究 [2] のプロの歌手の発声分析と同様の分析方法で調査した(図2,3)。

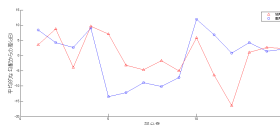


図 2. 男性

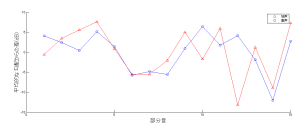


図 3. 女性

基本波から第15部分音までの値を取得し、平均からの差を表示した図である。先行研究と同じように、歌手ではない人の発声でも基本波と第2部分音は地声と裏声で傾きが大きく違った。

3 評価システム

3.1 評価システムの概要

評価のために、まずシステムを使用する歌い手に発声させる。マイクとの距離を10cmほどにする。音声はまず、発声されたピッチを推定し、基本周波数の特定、ラウドネスや部分音などの特徴量を分析し、声区や母音を求める。

評価では、まず現在ユーザの声区を表示し、次に喉に負担のかかっている歌声かどうかを表示、最後に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 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ために音源スペクトルの残差信号をとり、それをケプストラム法によってピッチを求めた。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、部分音を取得する。声区の違いをみるために、基本周波数の成分と第2部分音の成分の差を Δf_0 とする。

3.4 学習データ

母音/a:/の発声をマイクから 10cm 離れてしてもらい、サンプリング周波数 44100Hz で録音、母音の定常部分にハニング窓をかけたものの残差信号をとり、それを 2048 点でフーリエ変換した振幅スペクトルを使う。地声の発声と裏声の発声が重なる 200Hz ~ 500Hz を重点的に録音した。本研究では縦軸に周波数、横軸に第1部分音と第2部分音の差を使い、線形識別した(図4)。学習データは男性4人、女性2人からの各30から40ほどの地声と裏声のサンプルを各母音で発声したものである。

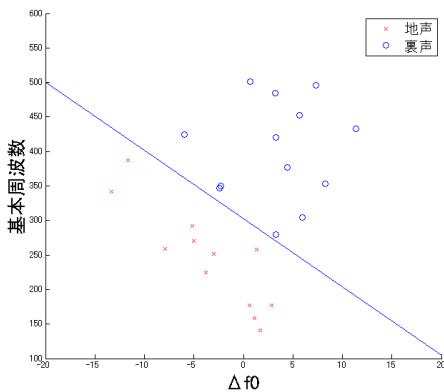


図 4. 学習用データ男性

4 実験結果

この評価システムを用いた声区の自動判定の実験結果は図5,6、表1のようになった。

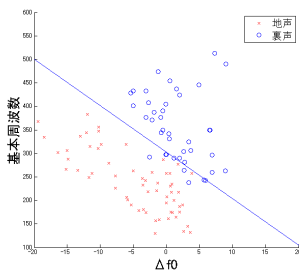


図 5. 評価データ男

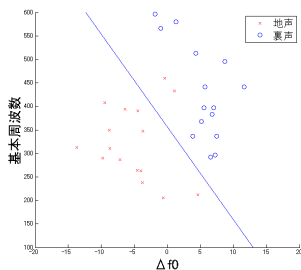


図 6. 評価データ女

表 1. 判定成功率/a:/

歌い手	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 %

4.1 f0 特徴量考察

母音/a:/の発声については、今回使った f_0 の特徴量が有効であることがわかった。しかし、ほかの母音についてはそれほど有効でないこともわかった。これは、母音の第1フォルマントと第2フォルマントの影響と考えられる。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波にはどちらも含まれることはない。しかし、他の母音について、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、基本波に含まれることがある。よって母音の違いで声区の判別率に差がついてしまった。

5 あとがき

f_0 と基本周波数による声区の判別は母音のフォルマント周波数に影響を受けるという結果が得られた。母音/a/以外の声区の判別には他の特徴量を用いる必要があると考えられる。また、発声が高音になると開口音である/a,o/の判別が難しくなったりなど、基本周波数が母音の第1フォルマントを超えると音声認識がしにくくなり、歌唱における発声と母音識別には複雑な関係になることもわかった。今後は、他の母音における声区の判別を違う特徴量を用いて行っていく。また、評価システムを完成させるにあたって喉締め声かどうかの評価をする必要があるが、母音/a/の場合だと f_0 が負の値になるほど歌手の喉に負担がかかるということがわかったので、他の母音ではどのような特徴量がかかわってくるのか調べていく。

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”, p.226, 1987
- [3] 矢田他、”歌声の基本周波数の動特性”, 日本音響学会平成 10 年度秋季講演論文集 3-8-6, pp.383-384, 1998
- [4] 辻直他”歌声らしさの要因とそれに関する音響特徴量の検討”, 日本音響学会聴覚研究会資料, H-2004-8, Vol.34, No.1, pp41-46
- [5] 津田他”3D 解析による歌声の評価に関する研究”, 電子情報通信学会情報・システムソサイエティ大会 D-458,p461, 1996