

多様な歌唱スタイルの歌声合成を目指した 歌唱ホルマント制御の検討*

齋藤毅, 後藤真孝 (産総研)

1 はじめに

筆者らは、話声を歌声に変換する歌声合成システム SingBySpeaking を提案している [1]。このシステムでは、歌声固有の音響特徴を話声に付与することで、自然な歌声合成を実現している。これまでは、歌唱スタイルの違いは考慮せず、様々な歌声データから各種音響特徴の平均的な特性を抽出し、それらの特性を反映させた歌声を合成してきた。しかし、多様な歌唱スタイルの歌声合成の実現を考えた場合、スタイル毎の各種音響特徴の特性を明らかにする必要がある。

歌声固有の音響特徴の中でも、歌唱ホルマント (singing formant) は、歌声に聴感的な“響き”を与えることで歌声特有の音色を生み出していると考えられており [2]、歌手のホルマント (singer's formant) と呼ばれることもある。オペラ歌唱を対象とした研究によれば、歌唱ホルマントは、母音スペクトルの 3 kHz 近傍に存在する顕著なスペクトルピークと定義される [3]。一方で、その特性に関しては、オペラ歌唱と邦楽歌唱において異なることが報告されており [5, 4]、歌唱スタイルの違いを規定する要因である可能性も示唆されている。しかし、様々な歌唱スタイル間での歌唱ホルマント特性の差異、更にはその差異が音色に与える影響に関しては、十分に調査されていない。

そこで本稿では、多様な歌唱スタイルの歌声合成を目指した歌唱ホルマント制御についての検討を行う。最初に、3種の歌唱スタイル (オペラ歌唱, 邦楽歌唱, ポピュラー歌唱) を対象に、各スタイルの歌唱ホルマント特性の差異を明らかにする。次に、分析で得られた歌唱スタイル毎の歌唱ホルマント特性の差異が歌声知覚に与える影響について調査する。

2 歌唱スタイル毎の歌唱ホルマント特性

歌声合成システム SingBySpeaking では、母音区間における話声のスペクトル包絡の 3 kHz 近傍に存在するスペクトルピークを最大 18 dB 強調することで、歌唱ホルマントを付与している [1]。これらの値は、オペラ歌唱を中心とした音響分析の結果から決定している。しかし、長唄等の邦楽歌唱においては、4 kHz 付近に歌唱ホルマントが存在し、その振幅もオペラ歌唱の場合と異なることが報告されている [4, 5]。そのため、歌唱スタイル毎の歌唱ホルマントの中心周波数とその振幅を分析し、スタイル毎の各特性の差異を調査した。

2.1 実験データ

歌声データベース「日本語を歌・唄・謡う」[6] に収録されているオペラ歌唱 (アルト, テノール, バリトン), 邦楽歌唱 (長唄, 民謡, 演歌), ポピュラー歌唱と, AIST ハミングデータベース [7] に収録されているポピュラー歌唱を対象に歌唱ホルマントの分析

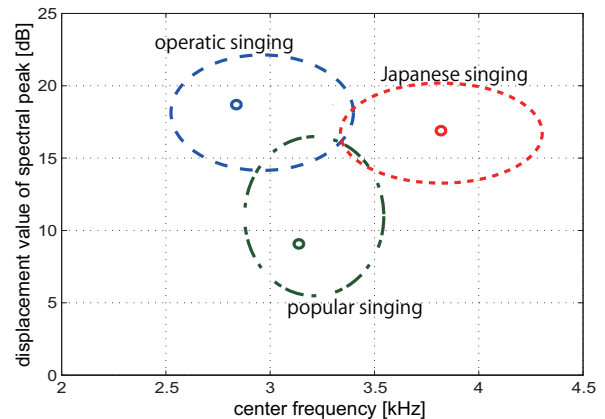


Fig. 1 歌唱スタイル毎の歌唱ホルマント特性の分布。点線は各歌唱スタイルにおける特性の分布を、○は平均値を示す。

を行った。各歌唱スタイルの歌唱者は、オペラ歌唱 7 名, 邦楽歌唱 7 名 (長唄 1 名, 民謡 5 名, 演歌 1 名), ポピュラー歌唱 7 名である。各歌唱者において、同じ歌詞を朗読・歌唱した話声・歌声データ (歌詞の内容はデータベースによって異なる) を 1 対選定し、実験データとした。尚、全実験データを対象に強制アライメントによる音素境界推定を行い、母音区間のみを対象に次節で述べる歌唱ホルマントの分析を行った。

2.2 歌唱ホルマントの分析

実験データを対象に、歌唱ホルマントの中心周波数と振幅をそれぞれ分析した。中心周波数は、歌声の母音区間の平均スペクトル包絡において、第 2 ホルマントより高域に存在する最も顕著なスペクトルピークが存在する周波数として抽出した。振幅に関しては、歌唱ホルマントの振幅値と話声のスペクトル包絡中のピーク値 (歌唱ホルマントの中心周波数近傍に存在するスペクトルピーク値) の差 [dB] として抽出した。尚、ピーク値の差は、話声と歌声のスペクトル包絡の振幅パワーを正規化させた上で算出した。

分析結果を Fig.1 に示す。横軸は歌唱ホルマントの中心周波数、縦軸は振幅変位 (話声のスペクトルピーク値からの増加量) をそれぞれ示す。結果から、歌唱スタイル毎に歌唱ホルマントの両特性が異なり、各スタイル固有の歌唱ホルマント特性が存在することが確認された。オペラ歌唱に関しては、中心周波数が平均で 2.7 kHz (2.6~3.4 kHz)、振幅変位が平均で 18.7 dB (14.2~21.6 dB) となり、先行研究で報告されている特性に近似する結果が得られた。邦楽歌唱では、中心周波数が平均で 3.8 kHz (3.4~4.3 kHz)、振幅変位が平均で 16.9 dB (13.5~20.1 dB) となり、先行研究で報告されている通りオペラ歌唱より高い帯域に歌唱ホルマントが生起していることが確認された。ポピュラー歌唱においては、中心周波数が平均で 3.2 kHz (2.8~3.5 kHz)、振幅変位が平均で

*A Study on Control of Singing Formant for Synthesizing Various Singing Styles. by SAITOU, Takeshi, GOTO, Masataka (AIST)

8.6 dB(5.5~15.2 dB) となり、他の歌唱スタイルに比べて振幅変位が小さいことが明らかとなった。

3 歌唱ホルマント特性と歌声知覚の関係

各歌唱スタイルの歌唱ホルマント特性の差異が音色に与える影響を調査した。中心周波数と振幅それぞれを変化させた歌唱ホルマントを持つ歌声合成音を作成し、聴取実験による評価を行った。

3.1 聴取実験

歌声合成音は、STRAIGHT[8]によって作成した。持続発話母音/a/(発話者:男性,発声時間:1.5s,平均音高:185 Hz)のスペクトル包絡に対して、異なる3種の中心周波数(2.7 kHz(オペラ歌唱), 3.2 kHz(ポピュラー歌唱), 3.8 kHz(邦楽歌唱))において1~25 dBまで1 dB刻みでスペクトルピークを強調することで歌唱ホルマント付与した75種(=3種の中心周波数×25段階の振幅強調)の歌声合成音を用いて聴取実験を行った。尚、全ての合成音に対して、ヴィブラート(変調周波数:5.7 Hz, 変移幅:25 cent)を発声区間全体に付与している。

聴取実験は一対比較法によって行った。歌唱ホルマントを付与していない歌声(比較歌唱)と付与した歌声合成音を対にして被験者に呈示し、比較歌唱に対する歌声合成音の音色の好ましさを7段階尺度(-3:とても好ましくない, -2:好ましくない, -1:やや好ましくない, 0:どちらとも言えない, +1:やや好ましい, +2:好ましい, +3:とても好ましい)で評価してもらった。被験者はボイストレーニングを受けている成人10名(男性2名,女性8名)で、防音室内においてヘッドホン(Sennheiser HDA200)を介して両耳に呈示される刺激音を聴取し、ノートPCのディスプレイ上に表示された評価尺度-3~+3のボタンを押すことで回答した。その際、回答時間の制限は無く、刺激音の聞き直しは3回まで許可し、音圧レベルは被験者の聞きやすいレベルに設定した。

3.2 実験結果・考察

図2に実験結果を示す。上図から順に、付与した歌唱ホルマントの中心周波数を2.7 kHz(オペラ歌唱), 3.2 kHz(ポピュラー歌唱), 3.8 kHz(邦楽歌唱)に設定した歌声刺激の評価結果を表す。各図の横軸はスペクトルピークの振幅増加量[dB], 縦軸は全被験者の評価平均値を表す。また、各図の網掛け区間は、Fig.1に示した各歌唱スタイルの振幅変位の分布範囲を示す。結果から、中心周波数毎の振幅強調に伴う音色評価には違いがあり、中心周波数が高い時ほど振幅強調が小さい場合で評価が高くなる傾向が確認された。また、中心周波数をオペラ・邦楽歌唱の特性で合成した場合(上図, 下図), Fig.1に示した各スタイルの振幅特性を付与した時に音色評価が高くなることが明らかとなった。一方で、ポピュラー歌唱の特性の場合(中図), Fig.1中の特性の付与した場合の音色評価は高くなかった。これらの結果は、オペラ・邦楽歌唱では、ポピュラー歌唱に比べて、歌唱ホルマントによる音色変化が重要であることを示唆していると考えられる。以上から、歌声合成において歌唱スタイルの違いを表現する手段として、各スタイルの歌唱ホルマント特性を反映させた制御方法の有効性が確認できた。

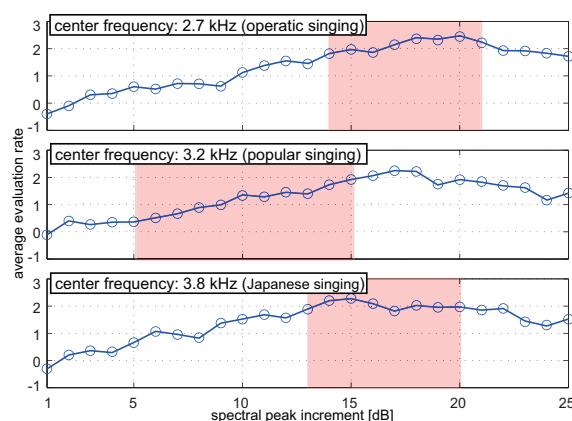


Fig. 2 異なる中心周波数を持つ歌唱ホルマントの振幅増加に対する音色評価の変化。

4 おわりに

本稿では、多様な歌唱スタイルの歌声合成を目指した歌唱ホルマント制御について検討した。オペラ歌唱, 邦楽歌唱, ポピュラー歌唱を対象とした歌唱ホルマントを分析した結果、各歌唱スタイル固有の歌唱ホルマントの特性が存在することが確認された。また、歌唱スタイル毎の歌唱ホルマント特性を反映させた歌声合成音を作成・評価した結果、各スタイルの特性による音色の違いが明らかとなり、スタイル毎の歌唱ホルマント制御の重要性を確認した。今後は、基本周波数の制御も考慮した歌唱スタイル制御についての検討を行う予定である。

謝辞 本研究は、科学技術振興機構 CrestMUSE プロジェクトによる支援を受けた。

参考文献

- [1] Saitou *et al.*, "Speech-To-Singing Synthesis: Converting Speaking Voices to Singing Voices by Controlling Acoustic Features Unique to Singing Voices," Proc. WASPAA 2007, 215-218, 2007.
- [2] W. T. Bartholomew, "A Physical Definition of 'Good Voice-Quality' in the Male Voice," J. Acoust. Soc. Am., 55, 838-844, 1934.
- [3] J. Sundberg, "Articulatory interpretation of the 'singing formant'," J. Acoust. Soc. Am., 55, 838-844, 1974.
- [4] I. Nakayama, "Comparative studies on vocal expression in Japanese traditional and western classical-style singing, using a common verse," Proc. ICA2004, 1295-1296, 2004.
- [5] 齋藤 他, "歌声らしさの知覚モデルに基づいた歌声特有の音響特徴量の分析," 日本音響学会誌, 64(5), 267-277, 2008.
- [6] 中山 一郎, "日本語を歌・唄・唄う" 日本音響学会誌, 59(11), 688-693, 2003.
- [7] 後藤, 西村. "AIST ハミングデータベース: 歌声研究用音楽データベース," 情処研報音楽情報科学, 2005-MUS-61-(2), 2005.
- [8] Kawahara *et al.*, "Restructuring speech representations using a pitch adaptive time-frequency smoothing and an instantaneous frequency based on F0 extraction: Possible role of a repetitive structure in sounds," Speech Commun., 27, 187-207, 1999.