

歌声の個人性知覚に寄与する音響特徴の検討*

齋藤 毅, 後藤 真孝 (産総研)

1 はじめに

歌声には、歌詞としての言語情報だけでなく、様々な非言語情報が含まれる。中でも、歌唱者の特有の声質や歌い回しに代表される個人性(個性)情報は、歌声をより表情豊かなものにしていく。そのため、歌声における個人性知覚に寄与する音響特徴が明らかになれば、非言語情報の知覚・生成機構解明の一助となるだけでなく、歌声合成システム等の音楽アプリケーションにも応用可能である。

これまで、歌声における個人性知覚には、声帯音源や声道音響特性といった生得的情報が歌い回し(韻律)といった後天的情報に比べて大きな影響を与えることが示されている[1]。しかし、ある歌唱者の歌声を巧みに模倣した音声を考えただけの場合、模倣しているのは生得的情報ではなく後天的情報であり、後天的情報にも個人性情報が顕著に含まれることが考えられる。そこで、本研究では、複数の人間が同一楽曲を歌唱したデータベースから原曲の歌唱者に類似/非類似歌唱者の歌声を選定し、歌声における個人性知覚に寄与する音響特徴を調査する。両歌唱者間の歌声の各種音響特徴を置換した歌声合成音を作成し、ソーストンの一対比較実験によって類似/非類似の聴覚判断に寄与する音響特徴を明らかにする。

2 実験データ

実験で使用する歌声データは、RWC 研究用音楽データベース(RWC-MDB)[2]と AIST ハミングデータベース(AIST-HDB)[3]から選定した。AIST-HDBとは、音楽経験の異なる100名のアマチュア歌唱者が、RWC-MDBから抜粋された50曲100箇所を1回もしくは5回聴取した後に発声したハミング(歌詞なし歌唱)、歌唱、歌詞朗読が収録されたデータベースである。本研究では、5回聴取した後の歌唱データの中から、聴感的にRWC-MDBの歌唱者に最も似ている/似ていない歌唱者の歌声をそれぞれ選定した。

選定は聴取実験によって行い、実験方法は一対比較法による類似度評価を用いた。RWC-MDB中の男女1名ずつ(RWC-MDB-P-2001 No.27, 28)による歌唱の出だし部分(以後、原曲データと呼ぶ)とAIST-HDB中の歌唱者(男性37名、女性38名)による同部分の歌唱データを対にして被験者に呈示し、5段階評価尺度(5:同じ, 4:似ている, 3:やや違う, 2:違う, 1:全く違う)に従って評価させた。被験者は、正常な聴力を有した成人10名(男性6名、女性4名)で、いずれも今回使用した歌声データを聴いた経験は無い。実験は防音室で行い、被験者にはヘッドホン(Sennheiser HDA200)を介して刺激音対を両耳に呈示し、PCディスプレイ上の評価尺度ボタンをマウス

で選択することで回答させた。尚、音圧レベルは被験者の聴きやすいレベルに設定し、各歌唱者に関する類似度評価を練習で1回、本番で3回行った。そして、被験者全員の本番3回の評価結果の平均値が最も高かった歌唱者の歌声を類似データ、最も低かった歌唱者の歌声を非類似データとし、男女それぞれ1名ずつ選定した。

3 聴取実験

3.1 実験刺激

男女それぞれの類似・非類似データ間で各種音響特徴を置換した以下の8種類の歌声合成音をSTRAIGHT[4]によって作成し、聴取実験の刺激音として用いた。

刺激音 OR :音響特徴を置換せず分析再合成した歌声合成音。

刺激音 A :全発声区間で平均 F0 を置換した歌声合成音。

刺激音 B :有声区間の対数 STRAIGHT スペクトルの 3 kHz 以上を置換した歌声合成音。

刺激音 C :時間軸を置換した歌声合成音。

刺激音 D :非周期性指標を置換した歌声合成音。

刺激音 E :刺激音 C の処理に加え F0 変化パターンを置換した歌声合成音。

刺激音 F :刺激音 C の処理に加え STRAIGHT スペクトルを置換した歌声合成音。

刺激音 TG :全てのパラメータを置換した歌声合成音。ここで、STRAIGHT スペクトルとは STRAIGHT 分析によって得られるスペクトル包絡のことである。

刺激音 A は歌唱者固有の音域の影響を調査するためのものである。発声区間全体の平均 F0 が、置換先の歌声データの平均 F0 値になるように F0 変化パターン全体を定数倍したものである。

刺激音 B は高域スペクトルの影響を調査するためのものである。対数 STRAIGHT スペクトルをフーリエ変換してケプストラムを求め、その 25 次までを用いて得られるスペクトルの 3 kHz 以降を置換したものである。これにより、調波成分が完全に除去された状態でスペクトルの置換処理を行っている。尚、置換の際に生じる不連続点を滑らかにするために、置換後の STRAIGHT スペクトルをローパスフィルタに通した後に合成している。また、3 kHz 以上という周波数帯域は、歌声データに含まれるすべての母音の第 2 ホルマントを含まない帯域として設定した。

刺激音 C~F 及び TG は STRAIGHT-morphing[5]で作成したものであり、C はテンポ(音符ごとに割当てられた音韻長)の影響、D は音源の非周期成分の影響、E は F0 の時間変化(歌い回し)の影響、F はスペ

* A study on acoustic features affecting perception of speaker individuality in singing. by SAITOU, Takeshi and GOTO, Masataka (AIST)

クトル形状 (声質) の影響をそれぞれ調査するためのものである。各刺激音は、置換するパラメータのモーフィング率を 100 % に設定することで作成した。

これらの刺激音は、類似データから非類似データへの置換、及び非類似データから類似データへの置換それぞれに関して作成した。

3.2 実験方法

サーストンの一対比較法 [6] により実験を行った。被験者に原曲データ (S_1) と 8 種の刺激音のいずれか 2 つ (S_2, S_3) を $S_1-S_2-S_3$ の順に呈示し、 S_2 と S_3 のどちらが S_1 の歌唱者に似ているか判断させた。

実験は、前節の実験と同じ被験者及び環境で行った。類似から非類似データ、及び非類似から類似データへの置換方向でそれぞれ作成した音声刺激について S_1, S_2, S_3 のすべての組み合わせを作成し、ランダムに被験者に呈示した。全試行回数は男女それぞれについて 112 回、各刺激音の間隔は 700 ms とした。被験者は、PC ディスプレイ上に表示された S_2 と S_3 のボタンのどちらかをマウスで選択することにより回答を行った。その際、回答時間の制限は無く、刺激音の聴き直しを許可せず、音圧レベルは被験者の聴きやすいレベルに設定した。

3.3 結果・考察

サーストンのケース V を適用して算出した感覚尺度を Fig. 1, 2 に示す。図中の値が大きいほど、原曲の歌唱者に似ている歌声であることを示す。尚、これらの結果は、すべてサーストンの一対比較法のモデルに適合していることを確認している。

Fig. 1 から、性別に関係なくスペクトル、F0 変化パターンの順で類似度が大きく低下する結果となった。これは、先行研究 [1] の結果を支持するものである。また、テンポ情報や高域スペクトルのみの置換でも類似度の低下が確認された。以上から、聴感的に似ているという判断には、スペクトル情報が F0 変化情報に比べて強く寄与しているといえる。

非類似から類似データへの置換でも、Fig. 1 と同様の音響特徴の置換が類似性知覚に与える影響が確認された。しかし、男性の場合では F0 変化パターンの影響がスペクトルに比べ大きく、また男性・女性どちらにおいてもテンポ情報が高域スペクトルに比べて大きい結果となった。これは、F0 変化の情報がスペクトル情報に比べて個人性知覚に強く寄与することを示しており、Fig. 1 の結果とは異なるものである。この原因については、今後詳細に検討する必要がある。

4 まとめ

本研究では、歌声における個人性知覚に寄与する音響特徴を明らかにする為に、複数の人間が同一楽曲を歌唱した歌声データベースから原曲の歌唱者に最も類似/非類似した歌唱者の歌声を選定し、両歌唱者間の歌声で各種音響特徴を置換した場合の個人性知覚の変化を聴取実験によって調査した。その結果、

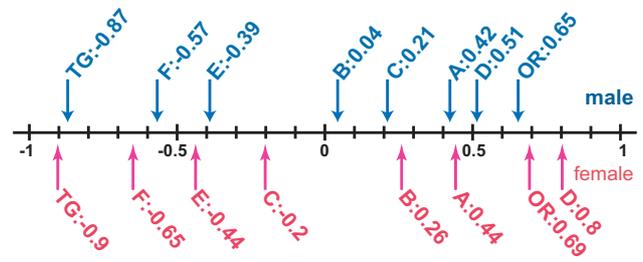


Fig. 1: 類似データから非類似データへ変換合成した際の個人性の近さに関する感覚尺度。

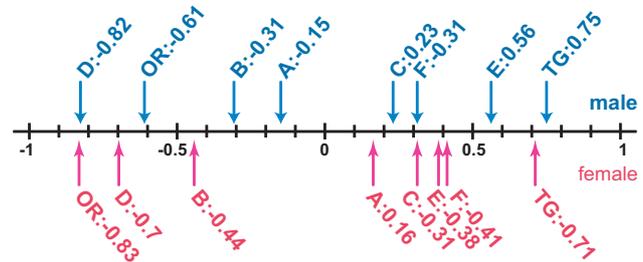


Fig. 2: 非類似データから類似データへ変換合成した際の個人性の近さに関する感覚尺度。

従来研究で報告されているスペクトル形状や F0 変化パターンの影響を確認し、更にはテンポ情報や高域スペクトルといった局所的な音響特徴のみでも個人性知覚に寄与することが明らかとなった。今後は、更にデータを増やして同様の実験を行うと同時に、これらの結果を踏まえた歌声の声質変換システムを構築する予定である。

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] 河原, 他. “歌唱音声モーフィングに基づく声質と歌い直し転写の知覚的検討,” インタラクシオン 2007, 113-120, 2007.
- [2] 後藤, 他. “RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース,” 情報学論, 45, 3, 728-738, 2004.
- [3] 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).
- [4] 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,” *SpeechCommun*, 27, 187-207, 1999.
- [5] Kawahara *et al.*, “Auditory Morphing based on an Elastic Perceptual Distance Metric in an Interference-free Time-frequency Representation,” *ICASSP2003*, I, 256-259, 2003.
- [6] 天坂, 長澤, 官能検査ハンドブック, 日本規格協会, 2000.