歌声の知覚年齢に沿った声質制御に向けた音響特徴量の調査*

☆小林 和弘，土井 啓成，戸田 智基（奈良先端大・情報），中野 倫靖，後藤 真孝（産総研），
ニュービッグ グラム，サクリアニ サクテイ，中村 哲（奈良先端大・情報）

1 はじめに

歌声は音楽を形成する上で重要な要素の 1 つであり，人は歌声の音高や音色に抑揚を付ける事で，多様な表現を生み出す事ができる．ただし，声質は身体的特徴によるところが大きく，個人の身体的制約を超えた歌声を発する事は困難である．近年，この制約を受けずに歌声の声質を制御する手法として，統計的声質変換に基づく手法が提案されている [1]．これにより，所望の歌手の声質による歌唱が可能となるものの，我々の主観に沿った自由な声質制御を実現するまでには至っていない．

本稿では，主観的情報の 1 つである「知覚年齢」に着目し，知覚年齢に沿った主観的声質制御を実現するために, 知覚年齢と関係する音響特徴量の調査を行う．実験結果より，韻律的特徴（F0 や音量の変化等）が知覚年齢に寄与していることを確認した．

2 統計的手法に基づく歌声声質変換

統計的歌声声質変換は，歌手の声質を別の歌手の声質へと変換する技術であり，学習処理と変換処理から成る．学習時には，源歌手と目標歌手が同一曲を歌唱した歌声で構成されるパラレルデータを用い，両歌手の音響特徴量の結合確率密度関数を混合正規分布モデル（Gaussian mixture model: GMM）でモデル化する．両歌手の静的・動的特徴量ベクトルをそれぞれ $\boldsymbol{X}_t = \left[\boldsymbol{x}_t^\top, \Delta\boldsymbol{x}_t^\top\right]^\top$ 及び $\boldsymbol{Y}_t = \left[\boldsymbol{y}_t^\top, \Delta\boldsymbol{y}_t^\top\right]^\top$ とすると，GMM は以下の式で表される．

$$P\left(\boldsymbol{X}_t, \boldsymbol{Y}_t|\boldsymbol{\lambda}\right) = \sum_{m=1}^{M} \alpha_m \mathcal{N}\left(\begin{bmatrix}\boldsymbol{X}_t \\ \boldsymbol{Y}_t\end{bmatrix}; \begin{bmatrix}\boldsymbol{\mu}_m^{(X)} \\ \boldsymbol{\mu}_m^{(Y)}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_m^{(XX)} & \boldsymbol{\Sigma}_m^{(XY)} \\ \boldsymbol{\Sigma}_m^{(YX)} & \boldsymbol{\Sigma}_m^{(YY)}\end{bmatrix}\right) \quad (1)$$

ここで $\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ は平均ベクトル $\boldsymbol{\mu}$ 及び共分散行列 $\boldsymbol{\Sigma}$ を持つ正規分布を表す．GMM の混合数は M であり，m は分布番号を示す．変換時には，新たに収録された源歌手の歌声を，GMM に基づき，最尤系列変換法 [2] を用いて目標歌手の歌声へと変換する．

3 知覚年齢に影響を与える音響特徴量調査

まず，多数の歌手の歌声に対し，人手により知覚年齢の付与を行う．次に，知覚年齢に寄与する音響特徴量を同定するために，以下の 4 つの合成歌声に対する知覚年齢を付与し，自然歌声に対する知覚年齢との比較を行う．本稿では，統計的歌声声質変換において主に変換対象となる分節的特徴量の影響に着目する．表 1 に各種合成歌声の特徴を示す．

3.1 分析再合成（分析再合成ひずみの影響）

分析再合成処理において生じるひずみが知覚年齢に与える影響を調査する．STRAIGHT 分析 [3] により，歌声からスペクトル包絡，F0，非周期成分 (Aperodic Components: AC) を抽出し，それらを用いて混合励振源に基づく STRAIGHT 合成を行うことで，分析再合成歌声 (w/ AC) を生成する．

3.2 非周期成分を用いない分析再合成（雑音成分の影響）

音源信号の雑音成分が知覚年齢に与える影響を調査する．3.1 節の分析再合成処理において，非周期成分に基づく混合励振源ではなく，パルス列のみで構成される有声音源を用いることで，分析再合成歌声（w/o AC）を生成する．

3.3 同一歌手声質変換 (汎化の影響)

声質変換における汎化処理が知覚年齢に与える影響を調査する．源歌手及び目標歌手を同一歌手とする GMM を用いて，源歌手から源歌手への変換（同一歌手声質変換）を行うことで，声質変換による汎化の影響を受けた合成歌声を生成する．源歌手から異なる歌手への変換用 GMM を用いて，源歌手から源歌手への変換に用いる GMM を下記の式により求める．

$$\begin{aligned} &P\left(\boldsymbol{X}_t, \boldsymbol{X}'_t|\boldsymbol{\lambda}\right) \\ &= \sum_{m=1}^{M} P\left(m|\boldsymbol{\lambda}\right) \int P\left(\boldsymbol{X}_t|\boldsymbol{Y}_t, m, \boldsymbol{\lambda}\right) \\ &\qquad P\left(\boldsymbol{X}'_t|\boldsymbol{Y}_t, m, \boldsymbol{\lambda}\right) P\left(\boldsymbol{Y}_t|m, \boldsymbol{\lambda}\right) \mathrm{d}\boldsymbol{Y}_t \\ &= \sum_{m=1}^{M} \alpha_m \mathcal{N}\left(\begin{bmatrix}\boldsymbol{X}_t \\ \boldsymbol{X}'_t\end{bmatrix}; \begin{bmatrix}\boldsymbol{\mu}_m^{(X)} \\ \boldsymbol{\mu}_m^{(X)}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_m^{(XX)} & \boldsymbol{\Sigma}_m^{(XYX)} \\ \boldsymbol{\Sigma}_m^{(XYX)} & \boldsymbol{\Sigma}_m^{(XX)}\end{bmatrix}\right) \end{aligned} \quad (2)$$

ここで

$$\boldsymbol{\Sigma}_m^{(XYX)} = \boldsymbol{\Sigma}_m^{(XY)} \boldsymbol{\Sigma}_m^{(YY)^{-1}} \boldsymbol{\Sigma}_m^{(YX)} \quad (3)$$

であり，$\boldsymbol{X}_t$ 及び $\boldsymbol{X}'_t$ は入力及び出力とする源歌手の静的・動的特徴量ベクトルである．また，$\boldsymbol{Y}_t$ は源歌手と異なる歌手の静的・動的特徴量ベクトルであり，隠れ変数として取り扱われる．

3.4 歌声声質変換（分節的特徴量の影響）

分節的特徴量が知覚年齢に与える影響を調査する．声質変換により，源歌手のスペクトル包絡パラメータ及び非周期成分から別の異なる歌手のスペクトル包絡パラメータ及び非周期成分への変換を行い，合成歌声を作成する．

4 実験的評価

4.1 実験条件

歌声データとして，AIST ハミングデータベース：ポピュラー音楽（RWC-MDB-P-2001) 日本語歌詞，サビパート [4] を用いる．評価楽曲は No.39 とする．20 代，30 代，40 代，50 代の男女各 1 名の組み合わせを 2 セット選出し，評価歌手とする．歌声声質変換の学習データは，上記データベース中の評価楽曲を含めた計 18 曲を用いる．スペクトル包絡パラメータとし

---

*Investigation of Acoustic Features for Voice Conversion to Control Perceptual Age of Singing Voice, by KOBAYASHI, Kazuhiro, DOI, Hironori, TODA, Tomoki（NAIST）, NAKANO, Tomoyasu, GOTO, Masataka（AIST）, NEUBIG, Graham, SAKTI, Sakriani, NAKAMURA, Satoshi（NAIST）

Table 1　分析再合成，同一歌手変換歌声及び声質変換歌声の特徴

特徴量	分析再合成（w/ AC）	分析再合成（w/o AC）	同一歌手変換歌声	声質変換歌声
スペクトル包絡	源歌手	源歌手	汎化，源歌手	汎化，目標歌手
非周期成分	源歌手	無し	汎化，源歌手	汎化，目標歌手
パワー，F0，継続長	源歌手	源歌手	源歌手	源歌手

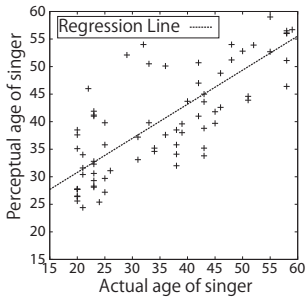

Fig. 1　歌手の実年齢と知覚年齢の対応図

て，STRAIGHT 分析によって得られたスペクトル包絡からメルケプストラム係数を算出して，その 1 次から 24 次までの係数を使用する．シフト長は 5ms，サンリング周波数は 16kHz とする．音源特徴量は，F0 と 5 周波数帯域における平均非周期成分を使用する．知覚年齢を付与する被験者は 20 代男性 8 人である．

各種分析再合成及び同一歌手声質変換による知覚年齢への影響を調査するために，自然歌声，2 種類の分析再合成歌声 (w/ AC 及び w/o AC)，同一歌手声質変換歌声に対して，知覚年齢を付与する．被験者 8 人を 2 グループに分け，各グループは異なる歌手セットに対する評価実験を行う．また，歌声声質変換による知覚年齢への影響を調査するために，各セットごとに評価歌手の内 1 人を源歌手，別の 1 人を目標歌手として変換歌声を作成し，評価実験を行う．源歌手と目標歌手の組み合わせは総当りとする．変換歌声は，源歌手のパワー，F0，継続長及び目標歌手への変換済み分節的特徴量を用いて作成する．

4.2　実験結果

図 1 に，自然歌声の知覚年齢と実年齢の関係を示す．知覚年齢の値は，20 代の男性被験者 1 人がデータベースに含まれる全ての歌手及び楽曲に対し知覚年齢を付与し，歌手当たりに平均化したものである．相関係数は 0.79 であり，自然歌声の知覚年齢は，歌手の実年齢と高い相関があることがわかる．

表 2 に，各種分析再合成音声及び同一歌手声質変換音声に対する知覚年齢と自然歌声の知覚年齢との差分の平均，標準偏差及び相関係数を示す．自然歌声と分析再合成歌声（w/ AC）の間では，知覚年齢の差分及び標準偏差は小さく，分析再合成ひずみは知覚年齢にほぼ影響を与えない．また，非周期成分を用いない分析再合成歌声（w/o AC）に関しても，同様であることから，非周期成分の有無によって知覚年齢はあまり変化しないことがわかる．これらと比較し，同一歌手変換歌声においては，自然歌声の知覚年齢との標準偏差が大きくなっており，知覚年齢のバラつきが増加する傾向が見られる．しかしながら，知覚年齢の差分は小さく，相関係数も高いため，歌声声質変換における汎化処理が知覚年齢に与える影響は小さいことがわかる．

図 2 に歌声声質変換歌声による知覚年齢の評価結果を示す．図 2（a）及び図 2（b）は，横軸がそれぞれ源歌手及び目標歌手の同一歌手変換歌声の知覚年齢であり，縦軸はどちらも声質変換歌声の知覚年齢である．すなわち，図 2（a）においては，変換特徴量で

Table 2　各種合成音声と自然歌声との間における知覚年齢の差分の平均，標準偏差及び相関係数

変換手法	平均	標準偏差	相関係数
分析再合成（w/ AC）	0.77	3.57	0.96
分析再合成（w/o AC）	0.44	3.58	0.96
同一歌手声質変換歌声	-0.5	7.25	0.85

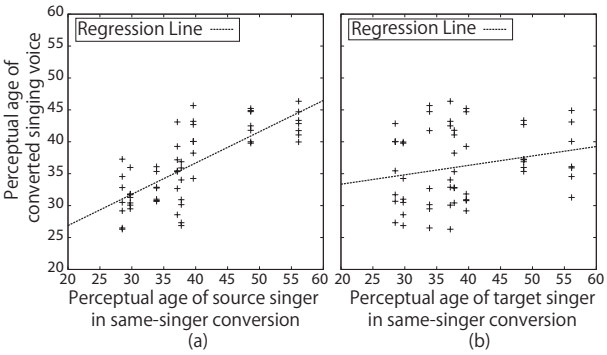

Fig. 2　同一歌手変換歌声の知覚年齢と声質変換歌声の知覚年齢の対応図　（a）:横軸を源歌手の知覚年齢にした場合，（b）:横軸を目標歌手の知覚年齢にした場合

ある分節的特徴量以外の特徴量が知覚年齢に大きく影響を与える際に強い正の相関を持ち，図 2（b）においては，分節的特徴量が知覚年齢に大きく影響を与える際に強い正の相関を持つ．相関係数は，図 2（a）が 0.75，図 2（b）が 0.23 であり，分節的特徴量よりも，パワー，F0，継続長といった韻律的特徴量の方が，知覚年齢に寄与していることがわかる．ただし，声質変換により源歌手の分節的特徴量が目標歌手のものへと完全に変換される訳ではないため，さらなる詳細な検討が必要である．

5　まとめ

歌声の知覚年齢に寄与する特徴量を調査するため，分節的特徴量である非周期成分とスペクトル包絡に着目して，実験的評価を行った．評価結果より，分析再合成や統計的歌声声質変換におけるひずみが知覚年齢に与える影響は小さいことを明らかにした．また，知覚年齢には，分節的特徴量より韻律的特徴 (F0 や音量の変化等) が大きく寄与していることがわかった．

謝辞　本研究の一部は，JSPS 科研費 22680016 および JST OngaCREST プロジェクトの助成を受け実施したものである．

参考文献

[1] H. Doi *et al.*, APSIPA ASC, 2012.
[2] T. Toda *et al.*, IEEE Trans. ASLP, Vol. 15, No. 8, pp. 2222—2235, 2007.
[3] H. Kawahara *et al.*, Speech Communication, Vol. 27, No. 3—4, pp. 187—207, 1999.
[4] 後藤真孝 他, 情報処理学会研究報告, Vol. 2005-MUS-61-2, No. 82, pp. 7—12, 2005.