1—P—17

混合正規分布モデルに基づく声質変換法の日英言語間への適用*

◎真下美紀子 戸田智基 川波弘道 鹿野清宏 (奈良先端大 情報)
Nick Campbell (奈良先端大/ATR/CREST)

1 はじめに

近年, 混合正規分布モデル (GMM) に基づく声質変換法 [1] を高性能な合成分析方式であるSTRAIGHT[2] に適用させた声質変換システムが作成された. 音声分析合成方式に基づいた声質変換は, 少量 (50-60 文程度) の音声データを用いて話者間の変換規則を学習し, 元話者の音声から任意のターゲット話者の音声を作成できるのが利点である. 上記のシステムではこの利点を生かし, 同一言語間で従来法よりも音質が改善されながら, かつ同程度の話者性変換精度も得られることが報告されている [3], [4]. この手法を用いて, 変換規則を学習する言語と声質変換する言語が異なった場合に高精度のターゲット話者の音声を作成できれば, 外国語学習, 機械翻訳などのアプリケーションに応用できると期待される. 本報告では, 日英言語間での声質変換の有効性を調査するため, 2 名の日英バイリンガル話者音声を収録し, 話者性評価実験を行った結果について述べる.

2 声質変換音声作成方法

図 1 に話者間の変換規則を日本語で学習し, それを変換規則として元話者の英語音声をターゲット話者の英語音声に変換する手順を示す. 本実験で使用した声質変換システムでは, 変換規則として STRAIGHT により分析され, 補間平滑化されたスペクトラムのメルケプストラムを用いている. メルケプストラム係数の 40 次までを変換することにより声質変換を行う. 0 次は元話者のものを用いる. 変換過程において, 韻律的特徴は基本周波数のみを考慮し, 元話者の平均基本周波数を, 対数領域でターゲット話者のそれに合わせている. 変換式は次のとおりである.

$$f_0' = \frac{\mu_t}{\mu_s} \times f_0 \qquad (1)$$

ここで, f_0 と f_0' は各々元話者音声と変換音声の対数基本周波数値を示し, μ_s と μ_t は各々元話者音声とターゲット話者音声の平均対数基本周波数値を示す.

3 音声データ収録

本実験では, 日本語で学習した声質変換規則を英語音声における声質変換に適用することを想定し, 日本語を母国語とする日本人バイリンガル女性話者 2 名の日英発話を収録した. 2 話者の日英両方の音声を得ることで, 変換された外国語合成音声とターゲット話者の外国語音声とを比較し, 評価することが可能となる.

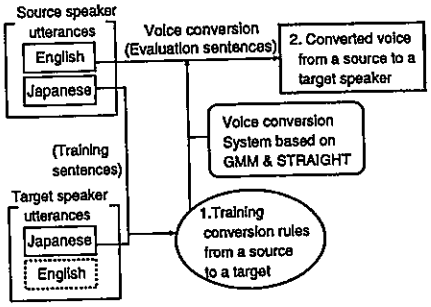

図 1: 英語声質変換音声作成手順

表 1: 音声収録条件

収録場所	遮音室
話者	FMM, FIA
サンプリング周波数	48000 Hz
量子化ビット	16 bit
発話文章数	日英各 60 文

収録条件を表 1 に示す. FMM は幼年期よりネイティブスピーカーに英会話を指導された経験を持ち, FIA は幼年期に長期英語圏滞在経験を持つ. 各話者の学習 50 文章の平均基本周波数は, FMM の英語が 248.6 Hz, 日本語が 270.0 Hz, FIA の英語が 233.9 Hz, 日本語が 227.6 Hz であった.

4 音声作成

声質変換音声における話者性の学習言語依存性を調査するため, 変換規則を学習する言語と変換音声の言語が異なる場合と同一の場合の 2 種類について, 英語音声と日本語音声をそれぞれ作成した. 学習には各々50 文使用し, 以下の 4 種類の変換を行い, 実験に使用した.

1. 日本語 (Jpn) で学習された英語 (Eng) 変換音声
2. 英語 (Eng) で学習された英語 (Eng) 変換音声
3. 英語 (Eng) で学習された日本語 (Jpn) 変換音声
4. 日本語 (Jpn) で学習された日本語 (Jpn) 変換音声

STRAIGHT 分析パラメータを表 2 に示す. 音声は 16000 Hz にダウンサンプリングしている.

*"Adaptaion of voice conversion method based on GMM and STRAIGHT to cross-language speech synthesis" by M. Mashimo, T. Toda, H. Kawanami, K. Shikano (Nara Institute of Science and Technology(NAIST)) and N. Campbell (NAIST/ATR/CREST)

表 2: STRAIGHT 分析パラメータ

分析窓	Gaussian
サンプリング周波数	16000 Hz
シフト長	5 ms
FFT ポイント数	1024
GMM クラス数	64

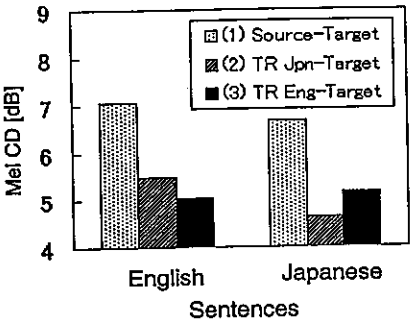

図 2: 話者性客観評価実験結果

5 話者性評価実験

声質変換において話者性にあらわれる学習言語依存性を, 客観評価実験および主観評価実験を通して考察する.

5.1 客観評価実験

異なる言語間においても声質変換が有効であるかを調べるため, FMM を元話者, FIA をターゲット話者として, (1) 元話者音声とターゲット話者音声, (2) 日本語で学習した変換音声とターゲット話者音声, (3) 英語で学習した変換音声とターゲット話者音声 のそれぞれについて, メルケプストラム歪み (Mel CD) を客観評価尺度として評価した. Mel CD は次の式で表される.

$$MelCD = \frac{20}{ln10}\sqrt{2\Sigma_{i=1}^{40}(mc_i^{(conv)} - mc_i^{(tar)})^2} \quad (2)$$

ここで, $mc_i^{(conv)}$ と $mc_i^{(tar)}$ は各々, 元話者あるいは変換音声とターゲット音声のメルケプストラム係数である.

評価用 10 文章の値を平均した結果を図 2 に示す. (2), (3) の場合とも, (1) の場合より Mel CD の歪みは小さくなっている. これらから, 話者性をターゲットの音声に近付けるために, 声質変換システム上で学習言語と目標とする変換音声とで異なる言語を使用した場合も有効であるとわかる. しかし, 通常, 日本語と英語の音声では, スペクトル差が存在する. そのため, 同一言語で学習した場合の方が, 英語文章で 0.46 dB, 日本語文章で 0.54 dB 程度, Mel CD は小さな値を示している.

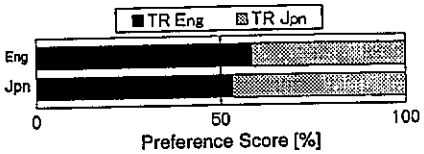

図 3: 話者性主観評価実験結果

5.2 主観評価実験

学習言語の差異が聴覚上どのように現れるか, 客観評価実験の結果と比較し, 検討するため, XAB 聴覚実験を行った. X はターゲット話者音声, A, B は日本語あるいは英語で学習した変換音声とし, 被験者には A, B どちらの話者性がより X に近いかを判断してもらう. 実験にはより高い品質が得られる, GMM をベースに周波数軸伸縮を適用した声質変換法 [4] を用い, FMM から FIA と FIA から FMM に変換した音声を用いた. また, 平均基本周波数は, 同じ話者でも英語と日本語で差があり, 同一言語の文章においても, 各文章ごとに差がある. そのため本実験では, 変換する文ごとに, (1) 式の μ_s, μ_t を各々元話者とターゲット話者の評価用文章 1 文の平均対数基本周波数とした. X, A, B とも同じ文章, 同じ言語であり, 文章は日本語, 英語とも, 評価用 10 文章中の 3 文章をランダムに提示した. 被験者は 10 名, 内女性 4 名, 男性 6 名である.

図 3 に結果を示す. 上段が英語文章で, 下段が日本語文章である. これより, 客観評価実験における Mel CD の差は, 聴覚上, 有意な差として現れていないことがわかる.

6 まとめ

GMM と STRAIGHT をベースとした声質変換システムを用いて, 学習言語とターゲット音声の言語が異なる場合の話者性を, 日英バイリンガル音声を用いて評価した. 客観評価実験の結果より, 同一言語間よりは精度が落ちるものの, 異なる言語間でも声質変換を施すことにより話者性はターゲット話者に近付くことが分かった. また, 主観評価実験結果より, 聴覚上は学習言語の違いによる話者性は現れないことが分かった. 今後の課題として, 聴覚上は差がないが, 話者間の距離計算を行う際には問題となる学習言語間の差を縮める手法を考える.

謝辞 本研究を援助頂きました, CREST に感謝致します.

参考文献

[1] Y. Stylianou, O. Cappé and E. Moulines, Proc. EUROSPEECH, pp. 447-450, Sept. 1995.
[2] H. Kawahara, I. Masuda-Katsuse and A. de Cheveigné, Speech Communication, vol. 27, no. 3-4, pp. 187-207, 1999.
[3] T. Toda, J. Lu, H. Saruwatari and K. Shikano, Proc.ICSLP, pp. 279-282, Oct. 2000.
[4] T. Toda, H. Saruwatari and K. Shikano, Proc. ICASSP, pp. 841-844, May 2001.