1—10—16

日英言語間における声質変換と母音の影響に関する検討*

◎真下美紀子 戸田智基 川波弘道 鹿野清宏 (奈良先端大 情報)
ニック キャンベル (奈良先端大/ATR/CREST)

1 はじめに

ある話者の発話音声を別話者の発話音声に変換する声質変換技術は, 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. 我々は, この技術を応用して, 学習者音声と教師音声とのケプストラム距離を基にした外国語発音学習支援を検討している. 音声情報処理技術を用いた客観的な発音評価において, 教師と語学学習者音声間の話者性抑圧が必要である. そこで, 声質変換技術を適用して教師音声の音響的特徴を語学学習者のそれに近づけたケプストラムと, 語学学習者自身のケプストラムを比較し, 発音誤りの同定を行うことを考える. ケプストラム距離に基く発音評価は, 音素単体で発声した場合の結果が報告されている [1]. しかし、実際に即した学習を行うためには, 音素や単語の発声ではなく, 文章発話中の音素を用いるべきである. 本稿では, 教師音声として日英バイリンガル話者音声を使用し, 声質変換を適用した場合としない場合とで, 発音評価の有効性を文章発話中の母音を用いて検証した.

2 話者性変換手順

これまでに, 混合正規分布と分析合成方式 STRAIGHT[2] をベースとした声質変換システム [3],[4] を用いて, 日英言語間にわたる声質変換を行っている [5]. その際, 日英バイリンガル音声を用いて2話者間の変換規則を学習する言語と変換先のターゲット言語が異なる場合でも, 個人性の抑圧効果が得られることを確認した. 本報告では, 語学学習における話者性抑圧のために, 教師音声にバイリンガル話者の日英音声を用いる. 話者間の変換規則は日本語間で求め, それを教師の英語にあてはめて話者性抑圧を行った変換ケプストラムを得ることを考える. 変換手順を図 1 に示す. 点線はデータの流れ, 実線は変換の流れを表す. 本実験のケプストラムとは, 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指す. メルケプストラム係数の 1 次から 40 次までを変換することにより声質変換を行う. パワー項は教師音声のものを用いている.

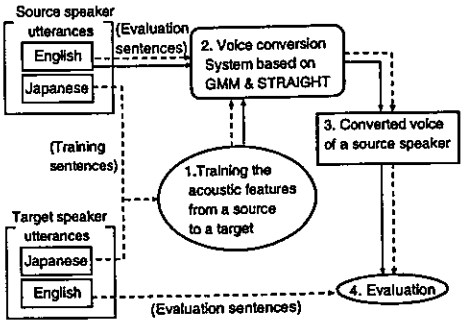

図 1: 英語声質変換手順

3 音声データ

教師用に日本語を母語とする日英バイリンガル女声話者 (TEA), 学習者用音声に日本人女声話者 (STU) 各 1 名を収録した. TEA 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, STU は学校教育以外に特別な英語教育は受けていない. 収録は遮音室で行い, それぞれ変換規則学習用の日本語 50 文と評価用の英語 10 文を, 量子化 16bit で DAT に録音した. テキストとして日本語は ATR 音韻バランス文, 英語は TIMIT データベース中に含まれる文章を使用した.

4 発音主観評価

メルケプストラム歪み (Mel CD) による話者間の母音の距離と, 人間による発音評価との相関を調べるために, STU 音声の評価用英語文章の母音について, 主観による発音評価ラベルを付けた. 基準は以下に述べる 3 段階である。音素ラベリングは, The CMU Pronouncing Dictionary の音素体系に基き, 手動ラベリングを行っている.

スコア 3: 教師発音と同じか近い

スコア 2: 発音間違いではないが教師発音と異なる

スコア 1: 発音間違い

10 文章中の総母音数は 114 個で, 各スコア値に対応した母音数はそれぞれ, スコア 1: 19 個, スコア 2: 22 個, スコア 3: 73 個であった.

*"Cross-language voice conversion and vowel pronunciation" by M. Mashimo, T. Toda, H. Kawanami, K. Shikano (Nara Institute of Science and Technology(NAIST)) and N. Campbell (NAIST/ATR/CREST)

表 1: STRAIGHT 分析パラメータ

分析窓	Gaussian
サンプリング周波数	16 kHz
シフト長	5 ms
FFT ポイント数	1024
GMM クラス数	64

表 2: 話者性評価実験結果 [dB]

変換先の言語	声質変換なし	声質変換あり
日本語	7.84	4.61
英語	7.73	5.51

5 評価実験

発音評価に向けて, 声質変換による話者性抑圧効果の有効性を検証する. まず, 声質変換を適用した TEA ケプストラムの話者性変換精度を評価する. 次に TEA と STU の母音間距離を測定し, 主観評価に基く発音評価と比較を行う. 以下に結果を述べる.

5.1 話者性変換精度

STRAIGHT 分析パラメータを表 1 に示す.

TEA 音声と STU 音声における話者性変換精度の尺度である Mel CD は次の式で表される. Mel CD の値がより小さい方が話者性がより近づいたと判断する.

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

ここで, $mc_i^{(conv)}$ と $mc_i^{(tar)}$ は各々, TEA スペクトルあるいは変換スペクトルと STU スペクトルのメルケプストラム係数である.

表 2 に結果を示す. 比較のため, 変換先の言語が日本語 (同一言語間での話者性抑圧) の結果も示す. Mel CD は評価用 10 文章の平均値である. これより, ある程度の話者性抑圧は得られていると言える. また, 同一言語間と日英言語間にわたる声質変換では, 異なる言語間の方が Mel CD の値は大きくなっており, [5] の結果とも一致する傾向が得られた.

5.2 音韻間距離

主観による発音評価スコアと Mel CD の関係を, 話者性抑圧を行わなかった場合を図 2, 行った場合を図 3 に示す. これらの結果より, 抑圧を行った場合の方がよりスコアとの対応が取りやすくなっていると言える. スコア 1 とスコア 3 は抑圧を行わなかった場合よりも, Mel CD の差が現れている. スコア 2 とスコア 3 は抑圧を行っても差が見られないが, これは, スコア 2 のデータ数がスコア 3 に比べて少ないことや, 完全な発音間違いほどは母音間のスペクトル差が大きくないことに起因していると考えられる.

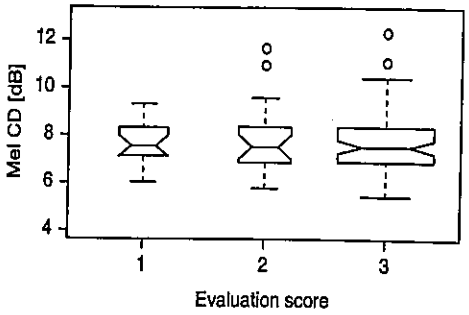

図 2: スコアと Mel CD の相関 (話者性抑圧なし)

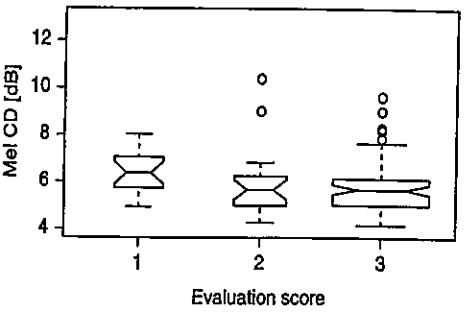

図 3: スコアと Mel CD の相関 (話者性抑圧あり)

6 まとめ

声質変換を発音評価の際の話者性抑圧に適用し, 予備段階としてその有効性を検証した. Mel CD を用いた教師と学習者の母音間距離測定により, 今回の話者性抑圧法は完全な発音誤りについて有効であると考える. 今後の課題として, さらにデータ量を増やすこと, また母音だけでなく日本人話者が誤りやすい子音の評価も検討する必要がある.

謝辞 本研究を援助頂きました, JST/CREST に感謝致します.

参考文献

[1] 深谷他, 信学技報, EA96-7, pp. 17–24, 1996
[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.
[5] 真下他. 音講論, 1-P-17, pp. 389–391, (2001-10)