

音素の音響的特徴を利用した水中音声の声質変換

佐々木志貢[†] 小坂哲夫[†]

[†] 山形大学大学院理工学研究科

1 はじめに

現在、水中でも会話ができる装置が開発されている [1]。これにより複雑な装置を利用しなくとも、ダイバー同士が水中で会話することが可能になった。しかし、呼吸のためにマウスピースを啜って会話を行うため、音声の了解度が低下するという問題があった。この問題を解決するために声質変換技術を利用する。声質変換技術はある話者の音声を他の話者の音声へ変換する技術である。本研究では、了解度の低い水中音声を了解度の高い陸上音声へと変換するために、この技術を利用し問題解決を図る。

2 混合正規分布を用いた声質変換

声質変換の手法として現在一般的に使用されているのは混合正規分布モデル (GMM: Gaussian mixture model) を用いた手法である [2]。本稿では、この GMM を用いた統計的手法を用いて水中音声から陸上音声への変換を試みる。現在の GMM では、音素ごとにモデル推定をしていないため、学習データの増加に伴い各音素固有の情報が平均されてしまい平均的な音に変換される傾向が見られる。このため陸上音声では分類可能なくつかの音素が、水中音声の場合、同一の音素として学習されてしまうという問題が発生する。そこで、音素の音響的特徴を利用した GMM による声質変換を検討する。

それを実現する手法として状態共有構造を持つ混合正規分布を用いる [3]。これにより、入力された音素の変換がより正確に行われるか検討を行う。

3 状態共有構造を持つ混合正規分布

本研究では、各混合分布を音素別にモデル化を行う。これにより、音響的特徴が保たれた GMM を構築することが可能になる。しかしながら各混合分布を音素ごとに学習すると一部分布ではデータ数が不足する問題が生じる。そこで、現在音声認識の分野で有効とされる状態共有構造 [3] を取り入れる。学習データの減少対策として音響特徴が似ている状態を共有することで、データ数が少ないものもパラメータ推定が可能となる。

4 実験

本実験では、まず簡易な環境下である陸上男女間変換で従来法と提案法である状態共有構造の比較検討を行う。次に本研究の目的である水陸間変換実験を行なう。陸上男女間変換では、男女それぞれの同内容発話 378 文を学習データとして使用している。水陸間変換では、水中陸上のそれぞれ同内容発話 782 文を使用した。検討方法としては、主観評価と客観評価の 2 種類による調査をした。主観評価では、6 人被験者に約 24 発話づつ、計 144 発話を聴取してもらい、了解度と音質についての 5 段階の MOS テストの実験を行った。客観評価は、ケプストラム距離 (MCD) による評価を行った。了解度と音質についての MOS テストの結果が、図 1 である。男女間について、ほとんどの音声と言葉をほぼ 100% 聞き取れるため、了解度は信頼区間を考慮するとほとんど差がないことがわかる。音質については状態共有構造に有意な結果が出た。これにより提案法の有効性が示された。客観評価についての結果を表 1 に示す。男女間変換では逆の結果で従来法の方がよいという結果が出ている。この理由としては、MCD による客観評価

は変換音声、対象音声に似ているかという指標を図るものであり、主観評価の了解度、音質とは指標として異なるからと考えられる。水陸間変換では、客観評価では従来法には及ばず、主観評価では、同等という結果となった。これは、水中音声では従来法と状態共有構造が共に変換音声の了解度、音質が陸上に比べ著しく悪く、聴覚的に差が出にくかったためと考えられる。さらに、水中音声と陸上音声の対応付けが上手く行われていない、もしくは、元の水中音声の発話が異なる音素間で極めて類似しているという可能性も考えられる。

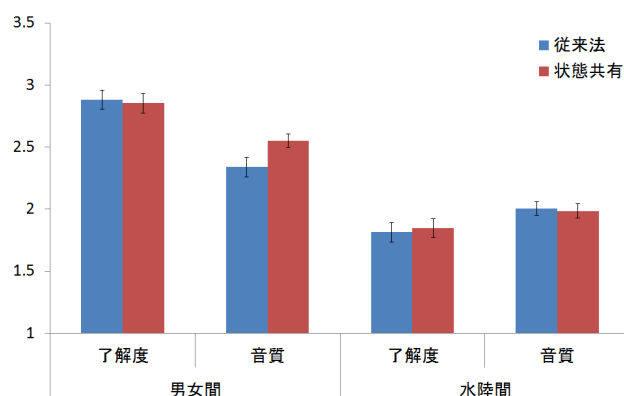


図 1: 了解度と音質の MOS テスト (95% の信頼区間)

表 1: 元音声と対象音声のメルケプストラム距離

	従来法	状態共有
陸上 MCD [dB]	0.117	0.126
水中 MCD [dB]	0.249	0.252

5 今後の課題

より高精度な声質変換を実現するためには、言語モデルなどの手法を導入する必要がある。

謝辞

本検討での音声データの提供及びご助力をいただいた山形カシオ株式会社の皆様に深く感謝致します。本検討を進める過程で、活発な議論と有益なご助言を頂いた山形大学の加藤正治助教に感謝致します。

参考文献

- [1] 山形カシオ株式会社 homepage
URL : <http://www.yamagata-casio.co.jp>
- [2] 花園 正也, 戸田 智基, 川波 弘道, 猿渡 洋, 鹿野 清宏: "GMM に基づく声質変換への尤度基準学習の適用", IEICE technical report. Speech 103(750), 43-48, (2004.3).
- [3] 堀貴明, 加藤正治, 伊藤彰則, 好田正紀: "音素決定木に基づく逐次状態分割法による HMnet の検討" 音響学会講演論文集, 3-5-19 (1996.3).