

舌亜全摘出者の音韻明瞭度改善のための 推定音素事後確率を用いた声質変換の検討

荻野 聖也[†] 原 直[†] 阿部 匡伸[†]
[†] 岡山大学大学院 ヘルスシステム統合科学研究科

1. はじめに

人は舌の動きや口の開閉により、構音している。舌亜全摘出者は癌治療などのために手術で舌の半分以上を摘出した患者であり、構音機能に大きな障害が残る。本研究では、声質変換の技術を用いて舌亜全摘出者の音声を健常者の音声に変換することで、舌亜全摘出者の音韻明瞭度改善を目指す。

先行研究[1]では、健常者のある音素が舌亜全摘出者の複数の音素に対応してしまう一対多変換問題により、舌亜全摘出者の音韻明瞭度の十分な改善に至らなかった。この問題を解決するために音素ラベルを用いた声質変換方式が報告されており、音韻明瞭度改善の大きな改善が示されている[2]。したがって、話者の発話から音素ラベルを音素事後確率として推定できれば、推定音素事後確率を用いた声質変換によって、音韻明瞭度の改善が期待できる。また、音素事後確率の推定精度向上のため、構音に密接に関係している口唇の動きの情報を用いる。

2. 提案方式

本研究では、舌亜全摘出者の音韻明瞭度改善のための推定音素事後確率を補助情報として用いた声質変換方式を提案する。提案方式の概要を図1に示す。

まず、音素事後確率の推定を行う。舌摘出者の音響特徴量と顔座標データを入力特徴量、音素ラベルを出力特徴量としてCNN-BLSTMにより音素事後確率推定モデルを学習する。次に、学習済み推定モデルにより得られた推定音素事後確率を用いて声質変換を行う。舌摘出者の音響特徴量と推定音素事後確率を入力特徴量、健常者の音響特徴量を出力特徴量としてBLSTMにより声質変換モデルを学習する。学習済み変換モデルを用いて健常者の音響特徴量に変換する。最後に、変換後音響特徴量から音声合成により変換音声を生成する。

3. メルケプストラム距離による客観評価実験

提案方式によって得られた変換音声が、健常者音声にどれだけ近づいたかをメルケプストラム距離により評価する。また、舌摘出者音声と音声情報のみを用いた声質変換方式による変換音声、提案方式による変換音声、[2]で提案された声質変換方式による変換音声の4つを評価した。結果を図2に示す。結果から、提案方式は多くの子音において音声情報のみを用いた声質変換方式に比べ良い結果を示しており、[2]で提案された方式に近づいている。これは、推定音素事後確率を付与し舌亜全摘出者の音響特徴量に対して音素の種類を明示することで、一対多

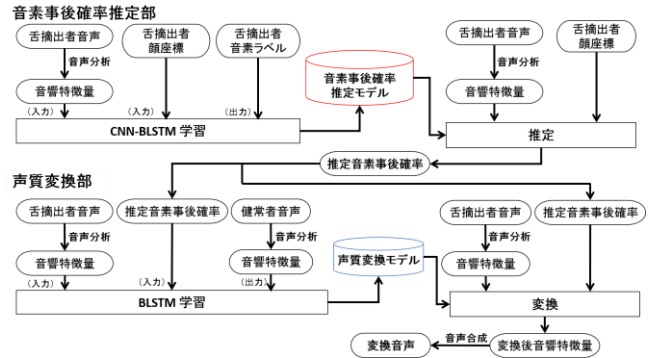


図1 提案方式 概要図

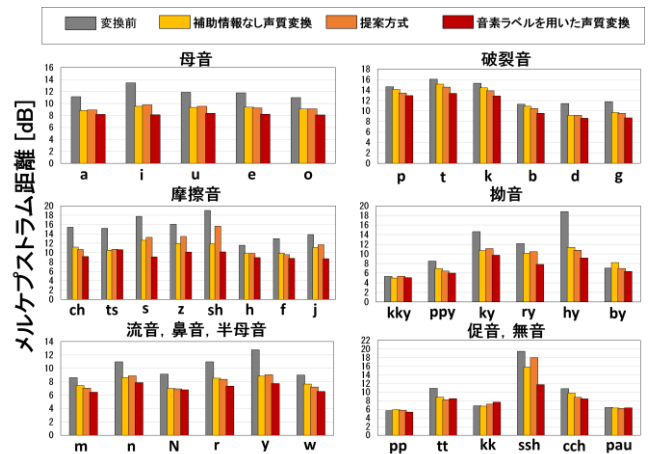


図2 変換方式ごとのメルケプストラム距離

変換の問題を緩和できたためであると考えられる

4. まとめ

舌亜全摘出者の音韻明瞭度改善のために、推定音素事後確率を用いた声質変換方式を提案した。評価実験の結果から、提案方式は多くの子音において、音韻明瞭度改善を示し、補助情報の有効性が明らかとなった。

参考文献

[1] H. Murakami, et al., "Naturalness Improvement Algorithm for Reconstructed Glossectomy Patient's Speech Using Spectral Differential Modification in Voice Conversion," *Proc. INTERSPEECH*, pp. 2464-2468, Sep. 2018.
 [2] H. Murakami, et al., "DNN-based Voice Conversion with Auxiliary Phonemic Information to Improve Intelligibility of Glossectomy Patients' Speech," *APSIPA*, Nov. 2019.