

発声障害者補助のための統計的声質変換技術の研究開発 (082107009) Development of Statistical Voice Conversion Techniques for Speaking-Aid

研究代表者

戸田 智基 奈良先端科学技術大学院大学
Tomoki Toda Nara Institute of Science and Technology

研究分担者

阪口 剛史[†] H20. 4~H21. 3
Takefumi Sakaguchi[†]
[†]奈良県立医科大学
[†]Nara Medical University

研究期間 平成 20 年度～平成 22 年度

概要

発声障害者として調音器官が十分に機能する喉頭摘出者を対象とし、食道発声や電気式人工喉頭を用いた発声などの代替発声法による音声を、より自然な音声へと変換する技術の研究開発を行った。統計的声質変換・制御技術を応用することで、代替発声音声の品質改善のみでなく、変換音声の声質制御も可能とする発声補助技術を構築した。テストベッドシステムの構築及び評価により、音質を大幅に改善することができ、声質も柔軟に制御できることを明らかにした。

Abstract

In this project, we have developed speaking-aid techniques for converting several types of alaryngeal speech, such as esophageal speech and electrolaryngeal speech, produced by laryngectomees whose vocal organs function reasonably well to more naturally sounding speech. The developed techniques are capable of not only improving speech quality but also controlling converted voice quality using statistical voice conversion and voice quality control methods. The effectiveness of a test bed system based on these techniques has been confirmed by experimental evaluations.

1. まえがき

喉頭癌を患い、手術により喉頭を摘出した人(喉頭摘出者)は、声帯を失い、気管と食道は完全に分離される。声帯振動による音源生成が不可能となるため、音声を発声するには、食道発声や電気式人工喉頭を用いた発声などの代替発声法を必要とする。一般的に、代替発声法により得られる音声(以下、無喉頭音声)は、自然性に乏しく、特有の声質を持つため話者性が欠如する。このため、日常生活における様々な場面においてコミュニケーションに支障をきたしており、その生活の質(Quality of Life: QoL)の低下は極めて深刻な問題といえる。

本研究開発では、調音器官が正常に機能する喉頭摘出者のQoL改善を目指し、無喉頭音声をより自然な音声へと変換する統計的声質変換技術の構築に取り組んだ。話者性などの非言語情報の伝達も可能とするために、変換音声の声質を柔軟に制御する技術も構築した。喉頭摘出者2名に対してテストベッドシステムを構築し評価を行った結果、無喉頭音声の品質を大幅に改善できることを明らかにした。

2. 研究内容及び成果

無喉頭音声の品質改善法として、無喉頭音声を健常者の通常音声へと変換する統計的声質変換・制御技術の構築に取り組んだ。代替発声法として、a) 食道発声、b) 電気式人工喉頭を用いた発声、c) 微弱振動音源を用いた発声の3つを対象とした。無喉頭音声と変換目標となる通常音声の同一発話内容の音声データを用いて、両音声特徴量間の変換モデルを学習することで、任意の発話の無喉頭音声に対する変換処理を実現した。構築した統計的声質変換・制御技術の概要を図1に示す。

2.1. 各代替発声法に対する声質変換・制御技術の構築

2.1.1. 食道発声に対する変換技術

食道発声は、口から取り入れた空気を食道にとどめ、そ

の空気を吐き出して食道上部を振動させることで、音声を生成する方法である。個人差は大きいものの、ある程度抑揚を制御することができ、生成される食道発声音声にはピッチが知覚される。しかしながら、食道発声音声を分析した結果、ピッチに対応した基本周波数の抽出や、有声/無声判定の精度は著しく低く、有用な音源特徴量を得るのは困難であることが分かった。そこで、食道発声音声のピッチを模倣した健常者音声を目標音声データとして用いて、声道スペクトル特徴量を食道音声のピッチに対応する基本周波数パターンへと変換する手法を考案した。また、声道スペクトル変換も導入することで、食道発声音声から通常音声への変換技術を開発した。

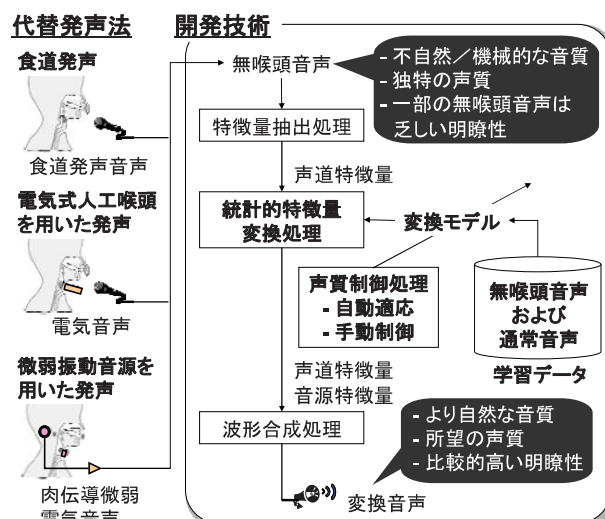


図1. 各無喉頭音声の品質改善のための統計的声質変換・制御技術の概要

2.1.2. 電気式人工喉頭を用いた発声に対する変換技術

電気式人工喉頭を用いる発声は、外部から音源を与えて声道内の空気を振動させることで、音源を生成する方法である。生成される音声（電気音声）の基本周波数パターンは人工的なものとなり、音質は機械的なものとなる。電気音声は有用な音源情報を含まないため、電気音声の声道スペクトル特徴量から健常者音声の基本周波数パターンへと変換する手法を考案した。さらに、声道スペクトル特徴量変換も導入することで、通常音声への変換技術を開発した。また、基本周波数パターンへの変換を必要としない方法として、健常者のささやき声への変換技術についても開発した。比較検討を行った結果、電気音声のように不自然な基本周波数パターン（例えばモノトーン）を持つ有音声に対しては、通常音声の基本周波数パターンへの変換処理は有効であることを明らかにした。

2.1.3. 微弱振動音源を用いた発声に対する変換技術

我々は、電気式人工喉頭の本質的な問題点である出力音源の周囲への音漏れを防ぐ方法として、微弱振動音源を用いた代替発声法を提案している。ただし、発声音の音量も極めて微弱なものとなるため、肉伝導マイクロフォンによる収録が必要となり、得られる音声（肉伝導微弱電気音声）の品質は極めて低い。そこで、電気音声への変換と同様に、肉伝導微弱電気音声の声道スペクトル特徴量から、健常者の基本周波数パターンや声道スペクトル特徴量へと変換する技術を開発し、健常者の通常音声への変換技術を開発した。また、発声時における聴覚フィードバック強調の有効性についても明らかにし、肉伝導微弱電気音声の収録品質を改善できることを示した。

2.1.4. 声質制御技術の導入

通常の統計的声質変換技術では、変換先として事前に目標話者を定め、変換モデル学習用データとして、目標話者が決められた数十文を発声した音声を収録しなければならない。喉頭摘出者は既に通常音声を発声することができないため、手術以前の自身の声への変換は不可能となり、必ず別の人の声に変換することになる。この問題に対して、多数の事前収録話者による音声の特徴を適切に混ぜ合わせることで、所望の声質を持つ変換音声を実現する固有変換技術の開発に取り組んだ。各代替発声法に対する声質変換技術に対して本技術を導入し、変換音声の声質適応機能を備えた発声障害者補助技術を開発した。これにより、声帯を失う前に収録した自身の音声データが極少量でも存在する際には、自身の声質を概ね再現するように、変換モデルを自動で適応する機能を実現した。また、目標とする音声データを用いなくとも、発声障害者自身が変換音声の声質を手動で自由に制御できる技術も構築した。

2.2. 無喉頭音声及び通常音声データベースの構築

統計的声質変換・制御技術に基づく発声障害者補助技術を開発する上で、喉頭摘出者による無喉頭音声データと、固有変換技術において目標音声となる多数の健常者による通常音声データが必要となる。そこで、各種音声データの収録を行った。食道発声を日常的に行っている喉頭摘出者1名の協力を得て、食道発声音声の収録を行った。また、電気式人工喉頭を用いた発声を日常的に行っている喉頭摘出者1名の協力を得て、電気音声の収録と、肉伝導微弱電気音声の収録を行った。発声データ量は、音素バランス文50文とした。また、多数の健常者により発声された通常音声データについても、収録を行った。一人あたりの発声データ量は、音素バランス文50文とし、75人（男性48人、女性27人）からなる変換目標音声データベースを構築した。

2.3. テストベッドシステムの開発及び評価

本研究課題で開発された要素技術を統合させることで、バッチ処理で動作する発声障害者補助テストベッドシステムの開発を行った。システム開発には、肉伝導マイクロフォンや微弱振動音源生成器といった補助装置も必要となるため、これらの装置の試作も行った。目標音声として、任意の1文のみを用いることで、変換音声の声質を自由に設定する機能を実現した。

本システムの有効性を実験的に評価するために、10人の健常者音声の声質への変換を想定し、実際の喉頭摘出者による各代替発声法による無喉頭音声に対して変換処理を行った。変換音声の品質、明瞭性、話者性再現精度に関する評価実験を行った結果、1) 全ての無喉頭音声に対して、大幅な音質改善が可能であること、2) 食道発声音声と肉伝導微弱電気音声に対しては、明瞭性も大幅に改善可能であること、3) 食道発声音声に対する変換処理が最も話者性再現精度が高く、次に電気音声、最後に肉伝導微弱電気音声の順番で、より良い話者性再現精度が得られること、4) 従来の声質変換技術に基づくシステムでは、目標話者による決められた40文程度の発声を必要とするが、声質適応処理が可能な固有変換技術に基づくシステムでは、同等の音質改善効果および明瞭性改善効果を保ちながら、任意の1文のみを用いて声質制御が可能であることを明らかにした。一方で、電気音声に対しては、変換処理により、明瞭性が若干劣化する傾向があることが分かった。

3. むすび

本研究開発では、喉頭摘出者による無喉頭音声の品質改善および話者性改善を可能とするために、統計的声質変換・制御技術を開発した。今後は、リアルタイム処理の実現や、環境変化に対する自動適応、明瞭性改善に向けた変換手法の改良などを行い、本技術の実用化を目指す。

【誌上発表リスト】

- [1]Keigo Nakamura, Tomoki Toda, Hiroshi Saruwatari, Kiyohiro Shikano, "Evaluation of extremely small sound source signals used in speaking-aid system with statistical voice conversion," IEICE Trans. Inf. & Syst., Vol.E93-D, No.7, pp.1909-1917 (2010年7月)
- [2]Hironori Doi, Keigo Nakamura, Tomoki Toda, Hiroshi Saruwatari, Kiyohiro Shikano, "Esophageal speech enhancement based on statistical voice conversion with Gaussian mixture models," IEICE Trans. Inf. & Syst. Vol.E93-D, No.9, pp.2472-2482 (2010年9月)
- [3]Yamato Ohtani, Tomoki Toda, Hiroshi Saruwatari, Kiyohiro Shikano, "Improvements of the one-to-many eigenvoice conversion system," IEICE Trans. Info. & Syst., Vol.E93-D, No.9, pp2491-2499 (2010年9月)

【受賞リスト】

- [1]戸田智基、日本音響学会第4回独創研究奨励賞板倉記念、「音声パラメータ系列の最尤推定に基づく声質変換」、受賞年月日2009年3月18日
- [2]Hironori Doi, Keigo Nakamura, Best Student Paper Award (1st Place) of the 35th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), "Statistical approach to enhancing esophageal speech based on Gaussian mixture models," 2010年3月18日
- [3]戸田智基、2010年度音声研究会研究奨励賞、「無喉頭音声強調のための統計的声質変換技術」、2011年6月24日