

音声障害者音声の健常者音声への帰納的学習を用いた 音声変換手法の提案と基礎実験

Conception and fundamental experiments of a Direct Speech Transform Method using Inductive Learning from Disorderd Voice to Normal Voice

村上浩司[†]
Koji MURAKAMI

荒木健治[†]
Kenji ARAKI

広重真人[†]
Makoto HIROSHIGE

栃内香次[‡]
Koji TOCHINAI

1. はじめに

近年の音声言語処理技術の急速な発展により、音声を用いた人間とコンピュータとの対話などのアプリケーション開発が盛んに行われている。しかしながら、こうしたシステムが入力対象として扱えるのは健常者音声のみであり、喉頭摘出などによって失声した人が発声する食道発声音声や、電気式人工喉頭による音声などの音声障害者音声は入力対象として考えられていない。そのため、こうした音声に対するシステムの応答精度は、大変低いものとなる。そのため、音声障害者音声を取り扱うことの出来る音声処理システムは現在、数多くは存在していないと考えられる。

近年のインターネット設備の普及に伴い、通信手段として大きな役割を果たしているものに電子メールがある。音声障害者や聴覚障害者にとっても地域との数少ない通信手段として活用されているが、不慣れな人に対して適した手段とは言えない。当然ながら、音声障害者も健常者と同様に健常者との日常的かつ円滑なコミュニケーションのために、音声を用いることが最適であると考えられる。音声障害者を対象とした、予め録音した音節の合成や登録されている文の読み上げ音声合成によるコミュニケーションツール [1] も開発されているが、登録済みの定型文以外は手で文を作成する必要があるため、ユーザへの負担が大きいと考えられる。

そこで我々は、現在の通常の音声認識システムでは認識の対象外とされている、音声障害者音声に着目し、その発話の内容を保持したまま直接、健常者音声に変換する音声変換手法を提案する。この手法ではユーザに対して文字入力を要求することなく、また定型文以外の文の入力からも学習を行うことで未知入力に対してもシステムが適応することができるため、より高度なユーザの要求に応えることができると考えられる。

本稿ではまず対象となる音声障害者音声とこれまで行われてきた研究について述べる。次に音声変換手法全体の提案を行い、音声処理や変換ルール獲得法などについて述べ、音声障害者音声を対象とした本手法の有効性を検討する。

2. 対象とする音声障害者音声

音声発声に関する障害はいくつかの種類に分けられる。その中でも、囁き声のような嚙声などの発声障害音声の認識や修復に関する研究も数多く行われているが [2, 3, 4]、本研究で対象とするのは喉頭癌などの原因により喉頭を摘出した結果、永久的に失声した人が日常の

コミュニケーションに用いるために発せられる音声である。まず、音声障害者音声についての現状とそれに関わる問題点を考える。

2.1 音声発声のメカニズム

音声は、通常肺からの呼気が喉頭から唇までにある発声器官と総合的に作用しあうことにより発声される [5, 6]。発声器官は、その機能により大きく3つに分けられる [7]。

- 音源
肺からの直流的な呼気を、喉頭部分にある声帯の声門を開閉することにより、断続流、もしくは乱流に変換し、これを音源とする。
- 声道での共鳴
声道により作られた音源に対して共振、反共振することによって共鳴を起こし、音色を付与する。
- 空間への放射
唇、鼻孔から音声を空間へ放射する。

喉頭癌などにより喉頭摘出手術を受けた人は、振動体である声帯を失うため、その摘出の程度により一時的もしくは永久的に音声を生成することができなくなる。その結果、筆談や表情、手話、ジェスチャなどが主要な通信手段となるため、意思疎通が円滑にできるまでには時間を要する。またそれだけでなく、音声という重要な対話手段を失うことによる精神的な苦痛も大きい。

しかしながら、失った声帯音源の代わりになる音を残された咽頭や舌などの調音器官に送り込むことにより、これらの人達は再び発声が可能となる。そこで現在、このような人達のためにいくつかの代用発声法が利用されている。代用発声には大きく分けて器具を使うものと使わないものの2つがある。器具を使わない方法にはTE-シャント法や食道発声法などがあり、器具を使う方法には笛式人工喉頭や電気式人工喉頭などがある。ここで、代表的な電気式人工喉頭および食道発声法の特徴について述べる。

2.1.1 電気式人工喉頭

電気的な振動音を音源とし経皮的に声道内に伝え、発声通りに口を動かすことで発声する。この電気式人工喉頭を用いた発声法は、音量があり比較的習得が容易とされるため喉頭摘出直後の人や、食道発声の習得が困難な人、高齢の人に有効であるという利点がある。しかしながら器具を用いるために身体障害の意識がついてまわり、発話には片手が塞がれるという欠点がある。

[†]北海道大学大学院工学研究科

[‡]北海学園大学大学院経営学研究科

2.1.2 食道発声法

現在最も普及している代用発声法で、食道に空気を飲み込み、その空気を逆流させて食道上部の粘膜である仮声門を振動させることで音源を生成する方法である。特別な器具を必要としないため、より肉声に近い声が再生できるというメリットがある反面、必要不可欠である食道へ空気を自然に摂取する行為が最も困難であり、長期的な訓練が必要である [8]。また音量不足や雑音などの要因により、音質は聴覚的な印象としてはまだ聞きにくい、騒がしい環境でのコミュニケーションに支障があるといった問題点がある [13, 18]。

2.2 過去の研究

2.2.1 電気式人工喉頭

この電気式人工喉頭は音源が健常者のものと大きく異なり基本周波数が固定であるために生成される音声はイントネーションやアクセントが含まれないことから、話者の意思や感情を的確に伝えることが難しく、不自然なものとなる。そのため、こうした音声の自然性を向上するために、アクセントやイントネーションを付加することのできる電気式人工喉頭が研究、開発、製品化されている [9, 10, 11]。

2.2.2 食道発声法

食道発声法による音声（以下、食道発声音声）の音響的特徴の研究の結果、健常者との音源メカニズムの違いに起因して、低い基本周波数、非常に大きなピッチと振幅の揺らぎ、大きな雑音、小さな音量などが問題点として挙げられている [12]。また、韻律及びスペクトルの両方において健常者のものとの違いが報告 [13, 14] されている。より聞きやすい音声への声質改善の研究 [15, 16] や、発声補助装置の開発 [17, 18] が行われた。

2.3 本研究の目的

音声障害者音声はこれまでに述べた問題点から、現在の音声認識システムに適用した場合ほとんどが正しく認識されない。これは通常、音声認識システムは健常者の音声を用いた学習により認識モデルが構築されており、音声障害者音声を用いた学習は行われていないため、そのような音声を入力しても正解を得ることができないからである。しかしながら、音声障害者に特化した音声認識システムの構築は困難であり、それにはいくつかの理由がある。

まず音声障害者の発話サンプルを大量に録音することは、食道発声者への大きな負担となるため通常の音声コーパスのように作成できない。また上述した発声音声の音響分析結果において、音素や音節などの単位に正確に分離できないことや、そのサンプルの音響的な品質が保証されないことでカテゴリ化された音素の整合性が的確に確保されないことなども、主たる要因と考えられる。

我々は、このように現在の通常の音声認識システムの対象外とされている音声障害者音声を直接利用することを考えた。従来から行われている、音声の基本周波数や

フォルマントを変換することでの声質変換 [19] や、コードブックを用いて話者変換 [20] するのではなく、発話データの比較によって発見される音響的な共通部分と差異部分のみに着目し、音声障害者音声の発話をその発話内容を保持したまま健常者音声に変換する音声変換手法を提案する。本手法の処理過程を図 1 に示す。

本研究では、音声障害者音声と変換先の健常者音声をそれぞれ原言語、および目的言語とした音声翻訳と考える。この目的言語を外国語音声に変更することで、音声認識を伴う通常の音声翻訳システムでは困難とされる音声障害者音声と外国語との音声翻訳も可能であると考えられる。

提案手法に求められる能力は、意味が伝わるレベルでの音声変換精度で音声変換を実現することである。

3. 本手法の基本的な考え方

3.1 手法の概要

まず、同内容の音声障害者音声と健常者音声の発話データを用意する。これらの発話データを音声特徴の時間変化を表すパラメータに変換し、健常者音声側と音声障害者音声側のそれぞれで発話の比較を行い、パラメータ間の距離を計算することで音響的な発話を差異・共通部分に分離する。こうして得られた音声障害者音声の差異・共通部分は、同内容の健常者発話の比較から同様に抽出される差異・共通部分と同じ意味を保持していると考えられるため、これらの組み合わせを変換ルールとして獲得し、変換ルール辞書に登録する。

このとき、発話データ間比較からのみ変換ルールを獲得するのではなく、帰納的学習を用いて既に獲得されたルール同士も比較することで、少ない発話データから多くの変換ルールを再帰的に獲得することができる。少ない量の発話サンプルから多くの変換ルールを獲得することは通常の自然言語処理だけではなく、ここで対象としている音声障害者のように長時間、また大量な発話サンプルの収集が困難である場合には顕著に有効である。

この帰納的学習は、知識や規則を有していなくても実世界において比較という行為により、人間は物事を共通部分と差異部分に分離し、それを知識として利用する能力があるという仮定から、表層的な情報から再帰的に差異・共通部分を抽出して変換規則として獲得する学習である。この手法は自然言語処理において有効性が示され

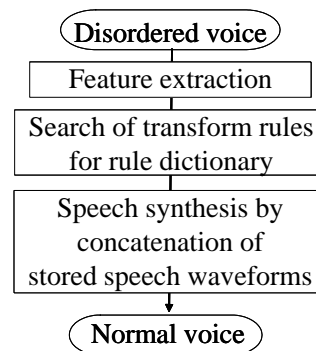


図 1: 提案する音声変換法の処理過程

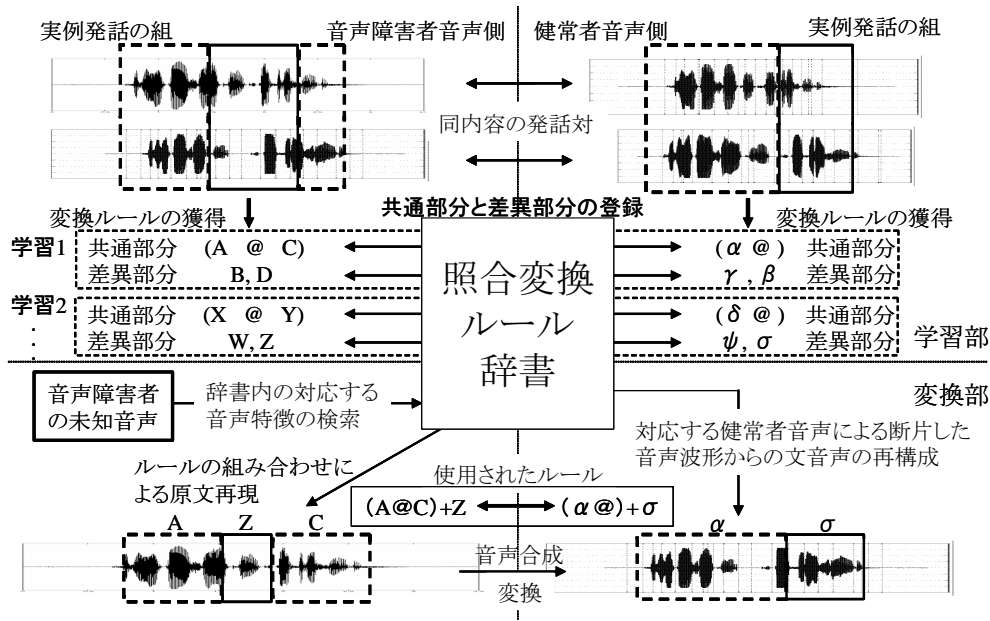


図 2: 全体の処理過程

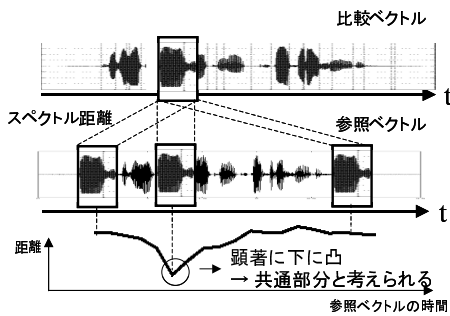


図 3: ベクトルの比較

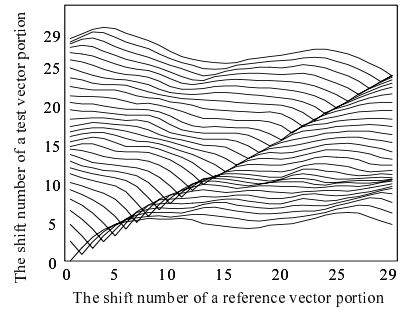


図 4: 発話間の距離.”Good afternoon.”

[21], 機械翻訳 [22] などに適用されている。

このような変換ルール獲得学習により獲得された辞書内の多くの変換ルールを、入力される未知の音声障害者音声に適用し、部分的に対応する変換ルールを組み合わせることで入力音声を再現する。このとき採用されたルールが保持している部分的な健常者音声を録音編集方式で音声合成し、発話内容を保持した変換結果を出力する。全体の処理過程を図 2 に示す。ここで提案する音声障害者日本語音声を対象とした健常者日本語音声への日本語 - 日本語間音声変換システムは、現在我々が進めている日本語 - 英語間の音声翻訳 [23, 24] と同様の枠組みで基本的なシステムの構築を行う。

3.2 音声処理

本手法では、前述した問題点を考慮して対象とする音声障害者音声を、処理する必要がある。音声障害者音声に特化した音声分析 [25] により、種々のパラメータにおいて健常者音声との違いが示されているが、我々は周波数スペクトル分析 [13] の結果に着目した。健常者の周波数スペクトルと比べて差が認められるが 5 母音の分離が

行われている。こうした結果を用いて、本手法における音声パラメータの比較による差異・共通部分への分離が可能であると考えられる。

3.3 変換ルール獲得

ベクトルに変換された発話実例組を健常者音声側及び障害者音声側でそれぞれ比較することで、差異・共通部分を抽出し変換ルールを獲得する。我々は発話速度差を吸収するために、端点自由 DP 法 [26] で得られるパスと類似した結果を得るよう、図 3 で示す手法 [27] を用いた。

図 4 にある時点で切り出された、比較を行う特徴ベクトル (以下、比較ベクトル) 毎に参照されるベクトル (以下、参照ベクトル) との比較により得られた距離変化の例を示す。この図は同一の話者・内容で、異なる発話を比較した例である。ここで、各グラフ内の最小距離を持つ点が共通部分になり得るため、これら最小距離の点を結んだ線が、DP で得られるパスに相当する。比較した両ベクトルが著しく類似していれば、その距離は極端に小さくなるため、計算された距離曲線には顕著に下に凸な部分が表れる。

こうして得られる各距離曲線の最小距離点を評価し、最初距離点が顕著に小さければその点を共通部分として扱う。

4. まとめ

現在の音声処理技術の対象となっていない音声障害者音声に関して、その特徴および問題点と、そうした音声を改良してより自然な音声を目指したこれまでの研究について説明した。

そしてこうした研究の結果から、我々は音声波形にのみ着目し健常者音声と音声障害者音声の発話実例を直接対応付け、未知の発話にも適応できる変換ルールを帰納的学習により獲得することで実現する新しい音声変換手法の提案を行った。

今後の予定として、音声障害者音声のサンプルを収集し、それらの発話データから帰納的学習を用いて変換ルール獲得を行い、ルールの評価を行う。また音声障害者音声の音響的分析を行い、最も効率よく変換ルールが獲得される音声特徴を検討する必要がある。

参考文献

- [1] <http://www.animo.co.jp/>
- [2] 大山, “発声障害音声の修復に関する1検討,” SP97-45, pp.7-14(1997-9)
- [3] Daniel Callan, Ray D. Kent, Nelson Roy and Stephen M. Tasko, “Self-Organizing Map for the Classification of Normal and Disordered Female Voices,” Journal of Speech, Language, and Hearing Research, Vol.42, pp355-366, April, 1999
- [4] 長島, 川野, 大川, 安田, “マヒ性構音障害音声の認識を目指した音声特徴の分析,” 音講論, 3-P-1, pp.167-168(2002-3)
- [5] 田窪, 前川, 他, “音声,” 岩波書店, 1998
- [6] 中田, “音声,” コロナ社, 1977
- [7] 岸, 船田, “電氣的喉頭波形を用いた連続音声の有声/無声/混合分類,” 信学技報 SP96-46, pp.31-37, 1996
- [8] 佐藤, “食道発声法,” 金原出版, 2001
- [9] 菊地, 粕谷, “F0 制御機能を有する電気喉頭の試作,” 音講論, 2-10-5, pp.295-296(2002-9)
- [10] 上見, 伊福部, 高橋, 松島, “ピッチ周波数制御型電気式人口咽頭の提案とその評価,” 信学論 D-II, vol.J78-D-II, No. 3, pp.571-578, 1995
- [11] 上見, 橋場, 須貝, 山口, 伊福部, “抑揚を制御できる電気式人工喉頭の製品化と喉頭摘出者による評価,” 信学技報 SP98-152, pp.47-52(1999-3)
- [12] 松井, 野口, “食道発声音声の音源強調方式の検討,” 音講論, 2-6-14, pp.423-424(1996-9)
- [13] 陸, 土井, 中村, 鹿野, “食道発声音声の母音の音響的特徴,” 信学技報 SP96-126, pp.33-40(1997-3)
- [14] 陸, 中村, 鹿野, “食道発声音声のピッチ特性に関する検討,” 音講論, 2-7-19, pp.253-254(1997-3)
- [15] 陸, 中村, 鹿野, “食道発声音声の改善における一考察,” 信学技報 SP98-151, pp.41-46(1999-3)
- [16] 秋元, 大塚, 粕谷, “食道発声音声の声質変換に関する検討,” 音講論, 2-7-9, pp.259-260(2002-3)
- [17] 野口, 松井, “食道発声音声の声質改善に関する一検討,” 音講論, 2-6-13, pp.421-422(1996-9)
- [18] 原, 松井, 窪田, 金, 太平, “食道発声補助装置の開発,” 信学技報 SP98-150, pp.33-40(1999-3)
- [19] 丁, 樋口, “Complex RBF ネットワークを用いた音声変換方法,” 音講論, 1-P-16, pp.335-336(1997-9)
- [20] Oytun Turk and Levent M.Arslan, “Subband based Voice Conversion,” Proc. ICSLP2002, pp.289-292, 2002
- [21] K. Araki and K. Tochinai, “Effectiveness of natural language processing method using Inductive Learning,” Proc. Artificial Intelligence and Soft Computing'01, pp.295-300, July, 2001.
- [22] H. Echizen-ya, K. Araki, Y. Momouchi and K. Tochinai, “Machine translation method using Inductive Learning with genetic algorithms,” Proc. the 16th International Conference on Computational Linguistics, pp.1020-1023, Aug., 1996.
- [23] K. Murakami, M. Hiroshige, K. Araki and K. Tochinai, “Evaluation of rule acquisition for a new speech translation method with waveforms using inductive learning,” Proc. Applied Informatics'02, pp.288-293, Feb., 2002.
- [24] K. Murakami, M. Hiroshige, K. Araki and K. Tochinai, “Evaluation of direct speech translation method using Inductive Learning for conversations in the travel domain,” Proc. ACL-02 Workshop on Speech-to-Speech Translation, July, 2002.
- [25] 加藤, “音声障害者音声の音響的特徴,” 日本音響学会講演論文集, pp.309-310(2000-3)
- [26] H.F. Silverman and D.P. Morgan, “The application of dynamic programming to connected speech recognition,” IEEE, ASSP Magazine, pp.6-25, 1990.
- [27] 村上, 広重, 荒木, 柘内: “文字表現を介さない音声機械翻訳システムの構想と基礎実験,” 音講論, pp.391-392(2001-10)