論　文

電気発声音声の健常者音声への音声変換手法の性能評価

村上　浩司[†*]　荒木　健治[††]　広重　真人[†††]　栃内　香次[††††]

A Method for Speech Transform from Electrolaryngeal Speech to Normal Speech

Koji MURAKAMI[†*], Kenji ARAKI[††], Makoto HIROSHIGE[†††], and Koji TOCHINAI[††††]

あらまし　本論文では，喉頭がんなどにより失声した喉頭摘出者の電気式人工喉頭による発声音声を対象とした健常者音声への音声変換手法を提案する．喉頭摘出者が使う食道発声音声や電気発声音声は，その音響的特徴の違いから，聞き手，話し手ともにストレスとなり円滑なコミュニケーションを抑制する要因となる．本手法では，喉頭摘出者音声のうち電気発声音声に着目した．健常者音声同士及び喉頭摘出者音声同士でそれぞれ DP マッチングを独立に行い，音響的な差異・共通部分となる音声素片を抽出する．その結果から両音声における素片の対応関係を変換ルールとして獲得し，それらのルールを未知電気発声音声に適用することで，その内容を保持したまま健常者の発話へ波形接続型音声合成を用いて変換する．特定話者の少数のデータサンプルを用いた音声変換実験から，本手法が有効である可能性を確認した．

キーワード　喉頭摘出者，電気発声法，音声変換，音響的特徴，波形接続型音声合成

1. まえがき

音声は人間にとって最も身近なコミュニケーション手段であるため，何らかの障害によって音声の発話ができなくなると，他者とのコミュニケーションに重大な支障を生じる．

一方，喉頭がんなどにより喉頭摘出手術を受けた結果，失声した人は日本国内で 20,000 人弱と推定され [1]，年々増加傾向にある [2]．こうした人々が日常的に用いる意思伝達手段として筆談や身振りなどがあるが，必ずしもその意思が的確に伝わるとはいえない．このような人々のための代用発声として代表的なものに，食道発声法や電気式人工喉頭による発声法があり，これらは第 2 の音声として広く用いられている．しかしながら，こうした代用音声を用いても発声器官の構造上，発声が不可能な音がある．そして音響的特徴の違い，雑音の混入などの原因による発声音声の不自然さが，聞き手だけでなく話者自身にも，ストレスとなることが少なくない [3]．

こうした喉頭摘出者を対象として開発されているコミュニケーションツールには，あらかじめ録音した音節の合成や登録されている文の読上げ音声合成によるもの [4], [5] があるが，日常用いられる発話以外の発話すべてを録音することは非常に困難である．また，登録済みの定型文以外は人手で文字を入力する必要があるため，利用に不慣れなユーザにとって負担は大きいと考えられる．したがって，喉頭摘出者も健常者と同様に，聞き手にも話し手にもストレスの少ない会話の可能な，より自然な音声を用いたコミュニケーションツールが強く望まれている．

そこで我々は，喉頭摘出者音声そのものに着目した [6]．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の変換による声質変換 [7] や，コードブックを用いた話者変換 [8] ではなく，DP マッチングを用いた発話組の比較により，音響的な共通部分と差異部分として抽出される音声素片のみに着目した．この音声素片を用いることで，喉頭摘出者の音声発話の内容

† 北海道大学大学院工学研究科，札幌市
Graduate School of Engineering, Hokkaido University, Kita 13 Nishi 8, Kita-ku, Sapporo-shi, 060–0628 Japan
†† 北海道大学大学院情報科学研究科，札幌市
Graduate School of Information Science and Technology, Hokkaido University, Kita 14 Nishi 9, Kita-ku, Sapporo-shi, 060–0814 Japan
††† 北海道教育大学釧路校，釧路市
Hokkaido University of Education Kushiro Campus, 1–15–55 Shiroyama, Kushiro-shi, 085–8580 Japan
†††† 北海学園大学経営学部，札幌市
Faculty of Business Administration, Hokkai-Gakuen University, 4–1–40 Asahimachi, Toyohira-ku, Sapporo-shi, 062–8605 Japan
* 現在，ニューヨーク大学

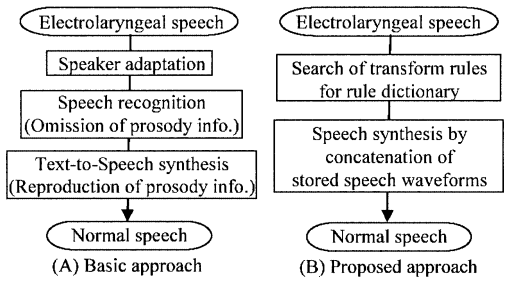

図 1　処理過程の比較
Fig. 1　Comparison of processes.

を保持したまま，波形接続型音声合成を用いて健常者音声に変換する音声変換手法を提案する．話者適応，音声認識，及びテキスト音声合成で構成する音声変換手法を図 1 (A) に，図 1 (B) に本手法の処理過程を示す．(A) の手法の場合，音声認識では発話の内容となる言語情報のみを保持する．喉頭摘出者音声の場合，大規模な音声発話の収録は話者の負担になる．また音声データの音響的な品質が保証されないため，音素の整合性が適切に確保できない．そのため，喉頭摘出者音声に特化した音響モデルの構築は難しいと考えられる．健常者音声の音響モデルに対し話者適応化を行う場合，喉頭摘出者音声の音響特性とのミスマッチから，適切な話者適応化が行われないことが考えられる．その結果，高い音声認識精度を得ることは難しいと考えられる．このような理由から，こうした手法では必ずしも効率良く適切な出力を得られるとはいえない．

本手法ではユーザに対して文字入力を要求することなく，あらかじめ収録した比較的少数の発話データから音声素片を抽出する．健常者音声の素片と電気発声音声の素片の対応関係を音声変換ルールとして多数獲得し，**5. 3** で説明する手法を用いて組み合わせ，出力を行う．そのため，本手法は言語情報だけでなく，音声のもつ情報をそのまま保持する特長をもつ．また，システムを利用するユーザの発話データを増やすことで，個人による音声特徴の偏りに適応した変換ルールが獲得され，変換に用いることができる．その結果，より高度なユーザの要求にこたえることができると考えられる．我々は，本手法を用いた個人用の小型音声変換装置の開発を目標としている．

本論文は以下のように構成される．**2.** で対象となる電気発声音声と，これまで行われてきた研究について述べる．**3.** では提案手法の説明を行い，**4.** で音声処理や DP マッチングを用いた音響的共通部分抽出について，**5.** で変換ルール獲得法などについて述べる．**6.** で実際の音声発話を適用した評価実験を行い有効性を検討し，**7.** で考察を行う．提案手法に求められる能力は，変換による発話内容の劣化を最小限にし，合成音声が電気発声音声による発話よりも知覚的に自然な発話であると感じられるレベルで音声変換を実現することである．

2.　電気発声音声

2. 1　喉頭摘出の影響

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 [9], [10]．発声器官は，その機能により大きく音源の発生，調音，放射の三つに分けられる [11]．喉頭摘出手術を受けた人は，音源を発生させるための声帯を失うため，咽頭や舌などの調音器官に問題がなくても，永久的に音声を生成することができなくなる．その結果，そのままでは筆談や表情，ジェスチャなどがコミュニケーション手段となり，意思の疎通に労力を要することになる．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．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二つ [12], [13] 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，器具を用いる方法には笛式人工喉頭や電気式人工喉頭などがある．電気式人工喉頭による発声には，手術後の状況に大きく依存せずに発声が可能である，他の発声法に比べて習得するまでに必要な時間が少ない，などの特長から発話者数が他の発声法に比べて多い．そのため，本論文では，対象とする喉頭摘出者音声を電気発声音声とした．ここで，本研究で対象とする電気式人工喉頭を用いた，電気発声法の特徴について述べる．

2. 2　電気発声法

電気式人工喉頭と呼ばれる器械による電気的な振動音を音源として，経皮的に声道内にその振動音を伝え，通常の発声のよう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，食道発声に比べて音量があり比較的習得が容易とされている．そのため，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

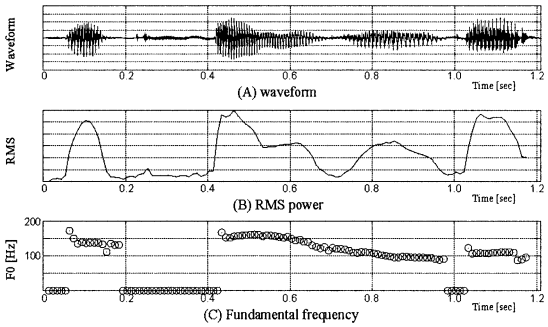

図 2　健常者音声
Fig. 2　Normal speech.

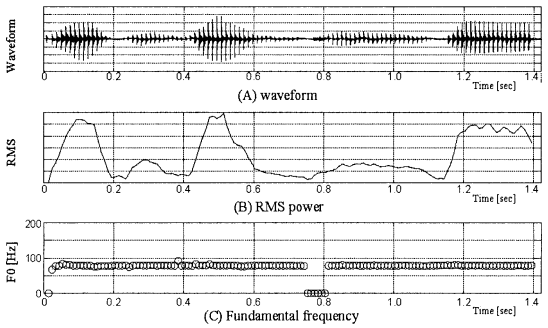

図 3　電気発声音声
Fig. 3　Electrolaryngeal speech.

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 [14].

電気発声音声と健常者音声における特徴の違いを調べるために音声のパワー，基本周波数の分析を行った．図 2 と図 3 に “ひゃくさんじゅうご（135）” を発話したそれぞれの原波形，RMS パワー及び基本周波数を示す．RMS パワーはグラフ中の最大値で正規化した．電気式人工喉頭による音源は，健常者のものと大きく異なり基本周波数が固定である．つまり，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，平たんな音声となり不自然なものとなる．そのため，話者の意思や感情を的確に伝えることが難しいという問題がある．こうした基本周波数の制御に関する研究も行われている [15] が，まだ課題が残されている．また，電気的振動音を経皮的に伝えるために，健常者音声に比べてノイズの影響を大きく受け，ふめいりょうな音になりやすい [16]．また，電気発声法では呼気を用いないため，声門で調音する /ha/ などの発音ができない．そのため，人工喉頭の使用に対する満足度は低く，改善が求められている [3]．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 [17]〜[19]．しかしながら，喉頭摘出者の気管孔の形状により適切に使用できない場合があること，ユーザに訓練が必要な場合があることなどの問題がある [17], [20].

3. 基本的な考え方

本論文では，話者変換や音声変換などの情報変換の条件を，対象となる 2 種類の音声発話間で，同じ意味をもつことであると仮定する．対象が文字列の場合，その文字列を解析することで対応関係を取り出すことは可能である [21]．そこで我々は，文字列と同じ言語情報を保持する音声から得られる音響的特徴を解析することでも，同様に対応関係を取り出すという研究を行ってきた [22], [23]．本論文ではこれまでの知見をもとに，DP マッチングを共通部分探索に適用した．

全体の処理過程を図 4 に示す．はじめに，同内容の電気発声音声と健常者音声の発話データを複数用意し，**4.** で説明する理由から，音声の周波数スペクトルの時間変化を表すパラメータに変換する．次に健常者音声側と電気発声音声側のそれぞれで，DP マッチングを用いて二つの発話データの比較を行い，音響的な差異・共通部分の音声素片に分離する．こうして得られた電気発声音声の音響的な差異・共通部分は，同内容の健常者発話データの比較から，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，こうした対応関係を変換ルールとして獲得し，変換ルール辞書に登録する．こうした変換ルール獲得を，すべての発話データに対して繰り返し行う．

このような変換ルール獲得により得られた辞書内の変換ルールを，入力される未知の電気発声音声に適用する．そして部分的に対応する変換ルールを変換候補として，それらを組み合わせることで入力音声を再現する．つまり，このとき採用された変換ルールが保持している，部分的な健常者音声素片を波形接続型音声合成により合成し，変換結果として出力する．

4. 音 声 処 理

健常者音声，電気発声の音声データの収録はそれぞれ 1 名の発話者で行った．表 1 に発声法と発話者に関する情報を示す．電気発声者は発声教室の指導員であり，発声技能は極めて高い．音声データは防音室で DAT を用いて 48 kHz のサンプリング周波数で録音

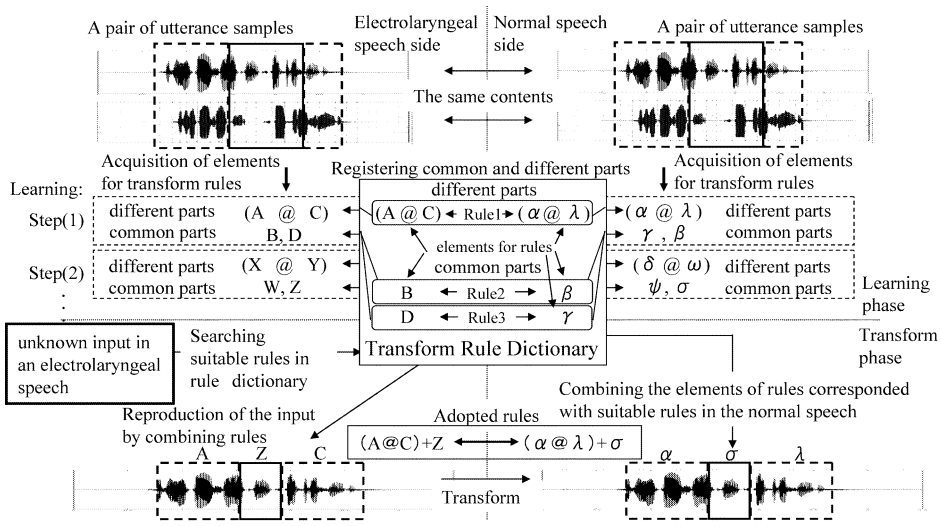

図 4　全体の処理過程
Fig. 4　Overall process chart.

表 1　発声の種類と発話者
Table 1　Characteristics of speakers.

発声	年齢・性別	話者の特徴
健常者音声	24 歳男性	大学院生
電気発声音声	70 歳男性	平成 6 年手術

表 2　音声処理に用いたパラメータ
Table 2　Parameters for speech processing.

サンプリング周波数	16 kHz
分析フレーム長	30 ms
フレーム周期	15 ms
時間窓	ハミング窓
ケプストラム次数	20

し，16 kHz にダウンサンプリングした．本手法では電気発声音声の特徴を考慮して，対象となる健常者音声と電気発声音声を処理する必要がある．

音声障害者音声の分析には種々のパラメータを用いた音声分析 [24] などが用いられている．電気発声音声は周波数特性のみでも特徴に違いが表れ，母音分類の結果が示された [25], [26]．そのため，本論文で使用する音声特徴は一般的な LPC ケプストラム係数とした．音声の性質に依存することなくその音響的変化をとらえるために，健常者及び電気発声法で共通して，表 2 に示されるパラメータを用いた．

ここで，音声データの数及び発話内容について考える．本論文で対象としている音声は電気発声音声であり，健常者の場合と違い，大量の音声を収録することは話者への大きな負担となる．そのため我々は，まず少数の音声データを用いて本手法が有効である可能性を確認することとした．また本手法では，多くの音響的な差異・共通部分を抽出する必要があるため，まずは限定された単語のみで発話が構成される内容を検討した．

その結果，健常者音声，電気発声音声ともに重複のない 3 けたの数字発話をデータとした．

5.　変換ルールの獲得とその適用

健常者音声と電気発声音声の間で同一の意味をもつ 2 発話の比較から，変換ルールの獲得とその適用の方法について述べる．

5.1　DP マッチングによる共通部分の探索

LPC ケプストラム係数を要素とするベクトルに変換された発話実例組を，健常者音声側及び電気発声音声側でそれぞれ DP マッチングを用いて比較することで音響的な差異・共通部分となるパターンを抽出し，それらの対応関係を変換ルールとして獲得する．

比較する音声発話 A の i 番目 $(0 \leq i \leq I)$ のフレームと音声発話 B の j 番目のフレーム $(0 \leq j \leq J)$ におけるそれぞれの特徴ベクトルは以下のように表される．

$$V_{Ai} = (v_{A0}(i), v_{A1}(i), \ldots, v_{A20}(i)) \quad (1)$$

$$V_{Bj} = (v_{B0}(j), v_{B1}(j), \ldots, v_{B20}(j)) \quad (2)$$

これらのベクトルに対して，我々は音声処理ツール ESPS [27] を用いて DP マッチングを行う．これは文献 [28], [29] の手法を実現しており，最適パスとそのパス上の各点の局所距離を出力する．

健常者音声，電気発声音声のそれぞれで DP マッチングの結果として得られる最適パス上の局所距離を利用して，ルール獲得を行う．

5.2 変換ルール獲得

まず，DP マッチングによって求められるパスとそのパス上の局所距離を利用して，比較した 2 発話を音響的な共通部分と差異部分に分離する．我々は，求められるパス上の局所距離が連続してしきい値 θ 未満であれば共通部分，しきい値以上であれば差異部分と定義した．健常者音声側と電気発声音声側で独立して DP マッチングを行った結果を図 5 に示す．縦軸は“にひゃくよんじゅうろく（246)”のフレーム数，横軸は“ろっぴゃくはちじゅういち（681)”のフレーム数である．それぞれ得られたパスと，計算された距離を重ねてプロットした．このとき，電気発声音声のパスの長さは，あらかじめ健常者音声のパスの長さで正規化を行った．この比較では健常者音声側，及び電気発声音声側の両方において，しきい値 θ 未満の局所距離が続く音響的共通部分がそれぞれ四つ存在している．この中から最も適切な対応関係をもつと考えられる共通部分を 1 組抽出し，その対応関係を変換ルールとする．

そこで，それぞれの音声側で得られる共通部分が複数発見された場合，共通部分が出現する始点と終点，部分長から信頼値を求め，信頼値が最大となる共通部分をそのときの音声比較における共通部分とする．健常者音声及び電気発声音声はどちらも同じ言語であり，同一の発話内容である．そのために，こうした発話を DP マッチングによって比較する場合，発話速度に著しい差がない限り，共通部分の長さと出現位置には大きな差はないと考えられるからである．

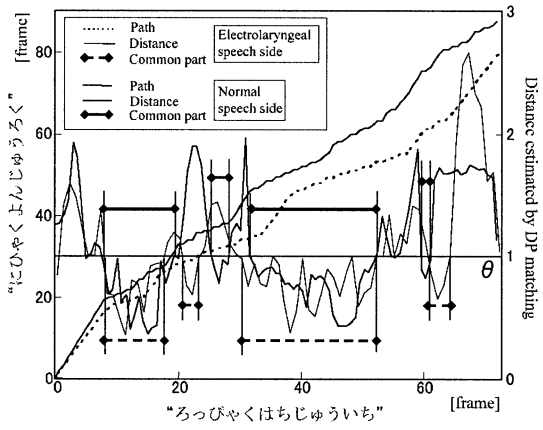

図 5 共通部分探索
Fig. 5 Common parts search.

次に信頼値の求め方について説明する．まず，それぞれ両音声側の i 番目の共通部分の始点，終点，及び両音声間における i 番目の共通部分の長さの差を表す，$D_{s(i)}$，$D_{e(i)}$，$D_{l(i)}$ を次の式で求める．

$$D_{s(i)} = |\alpha(S_{Ni} - S_{Ai})| \tag{3}$$

$$D_{e(i)} = |\beta(E_{Ni} - E_{Ai})| \tag{4}$$

$$D_{l(i)} = |\gamma(L_{Ni} - L_{Ai})| \tag{5}$$

S_{Ni} と E_{Ni} 及び L_{Ni} は健常者音声側，S_{Ai}，E_{Ai}，及び L_{Ai} は電気発声音声側の共通部分の始点，終点，長さである．α，β，γ はそれぞれの差に対する重み係数である．またこのとき，以下に示すように共通部分探索における出現位置の最大誤差範囲 δ を $D_{s(i)}$ 及び $D_{e(i)}$ と比較し，音声素片長の最大誤差範囲 σ と $D_{l(i)}$ も同時に比較する．

if $((D_{s(i)} < \delta)$ && $(D_{e(i)} < \delta)$ && $(D_{l(i)} > \sigma))\{$
　i 番目の共通部分は式 (6) の計算の対象
$\}$
else　i 番目の共通部分は式 (6) の計算の対象外

それぞれの距離が計算されれば，それらを用いて信頼値 CP を次の式により求める．

$$CP = \max_i \left(\frac{L_{Ni} + L_{Ai}}{1 + D_{s(i)} + D_{e(i)} + D_{l(i)}} \right) \tag{6}$$

このようにして決定された共通部分は，これまで述べた理由から両音声間で同一の意味を保持していると考えられる．この共通部分の位置情報を用いて健常者音声発話及び電気発声音声発話から音声素片を切り出し，それらを変換ルールとして獲得して変換ルール辞書に登録する．共通部分を抽出した後の差異部分についても同様に，それぞれの発話から素片を切り出し，変換ルールとして登録する．こうして獲得された変換ルールは，電気発声発話及び健常者発話の一部である音声素片の始点と終点の情報をもつ．

5.3 音声変換のためのルール適用

変換対象となる未知の電気発声の音声発話がシステ

ムに入力されると，ルール獲得処理と同様，まず **4.** で説明した特徴ベクトルに変換される．そして次に，変換ルール辞書に登録されているすべての変換ルールとの比較を行う．このとき，それぞれの変換ルールが保持する音声素片長は入力された発話長に比べて短いため，連続 DP [30] を用いてワードスポッティングする．すなわち時系列パターンの中から部分パターンを抽出する．こうして入力と比較した変換ルールが，ある一定範囲以上マッチした場合にその変換ルールは変換候補として選択される．このとき，変換ルールとマッチしない音声が入力され，候補が得られない場合，音声変換不能となることが考えられる．候補となった変換ルールが保持する音声素片には類似するものが多数存在する．そこで，それらの音声素片を **5.1** で説明した手法で比較し，全体が共通部分と計算された変換ルールをグループとしてまとめる．各グループ内における変換ルールの選択は，次のような手順で行う．

（1） 表 4 の条件

（2） 一致率が同じ場合，素片長の長いルールが優先

（3） 素片長が同一の場合，累積距離を比較して最も低い距離の変換ルールを選択

こうして得られた各グループから選択された，変換ルールのいくつかを合成することにより，入力である電気発声の発話を再現する．このとき，接続した二つの音声素片間においては，基本周波数・振幅や位相・音素持続時間の急激な変化を原因とする音質劣化が考えられる．また，選択された変換ルールは，入力である電気発声の発話を広範囲にカバーするように選択されるため，音声素片の重複部分が生じ，音声合成結果の了解度や自然性に影響を与えることが考えられる．そのため合成するルール数を最小にすることで，接続個所を最小にする必要がある．このときの変換ルールの組合せを最適組とする．この最適組の変換ルールが保持する健常者音声の素片を接続して音声合成を行い出力とする．

ここで述べた，音声素片間の音質劣化についての影響やその改善法，最適なルール選択と適用の評価については今後の課題である．

6. 評価実験

提案手法の評価実験として，電気発声音声から健常者音声への音声変換実験を行い，主観評価を行った．

評価は，音声データからの変換ルール獲得実験，そして獲得されたルールを用いて合成した音声の単語了解度試験とオピニオン評価，及び電気発声音声と合成音声の一対比較により行った．

6.1 変換ルール獲得実験

DP マッチングを用いた音声特徴ベクトルの比較を健常者音声側と電気発声音声側のそれぞれで行い，抽出される音響的な差異・共通部分となる音声素片の対応関係を，変換ルールとして変換ルール辞書に登録する．しきい値 δ と σ は，予備実験 [31] から始点，終点ともに 6 フレーム（120 ms）とした．本実験では，共通部分の決定式 (6) 中の重み係数は，α，β，γ ともに仮に 1.0 とした．差異・共通部分を分離するしきい値 θ は，数種類の音声データの比較により得られた図 5 のような音響的距離推移図と，しきい値を変化させ抽出される共通部分の音声の聴取から，適切と思われる値を決定する予備実験を行い，その結果から 1.0 とした．音声データには，**4.** で説明した理由から 3 けたの数字発話を用い，音声特徴には一般的な LPC ケプストラム係数を用いた．表 3 にデータの詳細を示す．この表から，健常者の発話の方が若干速い発話速度であることが分かる．しかし本手法では，それぞれの音声側で独立して音響的な共通部分と差異部分に分割するために，話者間の発話速度差は大きく影響しない．実験の結果，変換ルール辞書には 81 発話の比較により獲得された 2,045 の変換ルールが登録された．

6.2 主観評価実験

ルール獲得により構築されたルール辞書を用いて，変換ルール獲得に用いた 81 発話すべてについて，これまで述べた手法を用いて音声変換を行い，出力された合成音声の評価を行った．評価には交差検定を用いた．これにより，評価ターゲットとなる各発話は，その発話を除いた他の 80 発話の比較から獲得された変換ルールを合成することで評価される．つまり，評価

表 3 音声データ
Table 3 Number of speech data.

発話データ	データ数	発話総時間	平均発話時間
電気発声音声	81	105.50 s	1.30 s
健常者音声	81	85.76 s	1.06 s

表 4 変換ルールの適用条件
Table 4 Conditions for applying transform rules.

入力音声との一致率	95%
連続 DP のフレーム周期	音声素片長の 70%
ルールが保持する音声の最低音声素片長	220 ms

ターゲットはすべてオープンデータとなる．表 4 に音声変換を行う際の変換ルールの適用条件を示す．これにより，**5.3** で説明したように，入力された未知の電気発声音声と一定時間長以上部分的に音響的な類似度が高い変換ルールのみが選択され，変換候補となる．これらの値は予備実験により決定した．各評価ターゲットに対し，平均 42.08 個の変換ルールが候補として選択され，平均 11.9 のグループに分けられた．

評価を行う被験者には，電気発声音声を日常的に聞く環境にない日本人の男性 9 名，女性 3 名の大学生計 12 名を選び，ヘッドホンから音声を呈示した．被験者に呈示する合成音声は，ランダムな順で選びすべての評価に共通して用いた．被験者にはどの実験においても，必要に応じて一度だけでなく何度も合成音声を聴くことを許している．

全 81 発話に対して，これらの評価条件のもとで以下の三つの評価を行った．

（1）単語了解度試験

評価ターゲットの発話内容を伏せた上で合成音声を聞き，その内容を数字で書き取るよう指示した．この際，聞き取れた音のままではなく全体としての発話内容を正しい数字で書き取るよう指示した．この評価では正解精度を算出する．

（2）オピニオン評価

81 発話に用いた数字を被験者に呈示し，健常者の自然音声，電気発声音声，合成音声の自然性を 5 段階（1：不自然–5：自然）で評価するよう指示した．この評価においては評価値平均と標準偏差を算出する．

（3）電気発声音声との一対比較

合成音声と電気発声音声の同内容発話をそれぞれ一対で呈示し，被験者にはそれらの比較から，より自然に感じる合成音声を選択するよう指示した．

6.3 実験結果

表 5 にそれぞれの音声についての単語了解度試験とオピニオン評価の結果を，図 6 に電気発声音声との一対比較の結果を示す．

（1）単語了解度

評価実験から，提案手法による合成音声の単語了解度は，平均 95.2%であった．健常者音声においては 100%，電気発声音声では平均 99.5%となった．このことから，提案手法による合成音声は電気発声音声に比べて低い単語了解率となったが，その差は 4.3%であり著しい低下とはなっていない．これにより，提案手法において電気発声音声はその発話内容の劣化を最小限に抑えて健常者音声へ変換されたことが示された．

表 5　単語了解度試験とオピニオン評価の結果
Table 5　The results of word intelligibility test and of opinion test.

発話データ	単語了解度試験 正解率 (%)	オピニオン評価 不自然：1 → 自然：5
提案手法による合成音声	95.2	3.89 (0.39)
健常者音声	100.0	4.98 (0.03)
電気発声音声	99.5	2.93 (0.61)

（2）オピニオン評価

括弧内は標準偏差を示す．この評価では，健常者音声が平均 4.98，電気発声音声が平均 2.93，提案手法が平均 3.89 となった．この結果から提案手法による合成音声は，健常者音声と比べると自然性に課題はあるが，電気発声音声に比べると十分に高い自然性を有することが示された．

評価に用いた 81 発話すべての合成音声についての，単語了解度試験とオピニオン評価の結果を図 7 に示す．評価ターゲット全体では，高い単語了解度が得られていることが確認できる．また合成音声が不自然と感じられるに伴い，単語了解度も減少する傾向があることが分かる．

（3）電気発声音声との一対比較

評価実験により，72.3%の平均プレファレンススコアが得られた．80%以上のプレファレンススコアを示した合成音声は全体の 54.3%（44/81），スコアが 50%以上の合成音声は 71.6%（58/81）となった．

以上のことから，提案手法は波形接続で生じる音質劣化に対する処理や，韻律情報を使った音声素片選択を行っていないにもかかわらず，変換結果として得られた合成音声は，高い単語了解度と自然性をもつことが示された．また，電気発声音声と比較した場合，より健常者音声に近い自然性を有することが確認された．

7. 考　　察

それぞれの評価実験の結果として低い評価値となったターゲットに着目する．単語了解度が 50%未満であった評価ターゲットは 5 個，50%未満のプレファレンススコアであった評価ターゲットは 21 個であった．それぞれの結果について考察する．

（1）単語了解度が著しく低いもの

このカテゴリーに属する評価ターゲットは，数字を自然に聞き取れるものの，正解の内容とは異なって聞

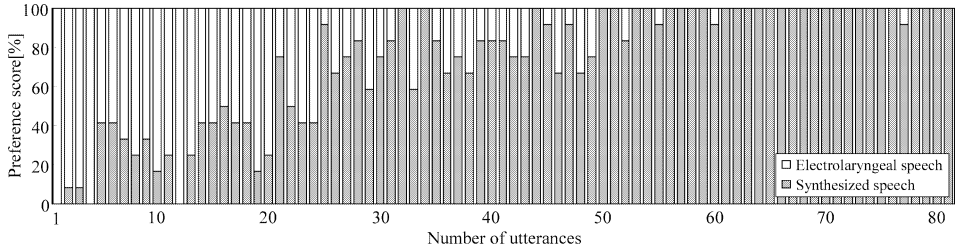

図 6　電気発声音声との比較実験結果
Fig. 6　The results of comparison with electrolaryngeal speech.

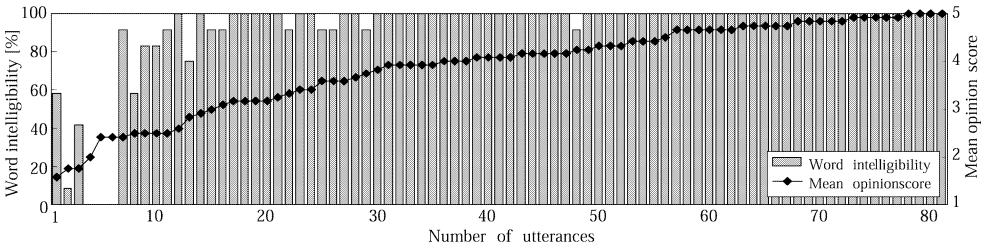

図 7　単語了解度と平均オピニオン値
Fig. 7　Word intelligibility and mean opinion score.

こえるものである．これは適切な音声素片をもつ変換ルールが獲得，若しくは選択されないことが原因である．

そこで，今回使用した音声データの発話内容に着目した．表 6 に発話されたそれぞれの数字の読みの頻度を示す．表中で，読みの出現頻度が低いものは適切な音声素片が抽出されていない可能性が高いと考えられる．そのため出現頻度が 10 以下の読みを含む評価ターゲットを除いて再び単語了解度を求めた．その結果，単語了解度は 95.2%から 98.33%に上昇した．このことから，発話者である喉頭摘出者に大きな負担を与えない範囲で発話データを増やすことで，適切な変換ルールをより多く獲得できると考えられる．

（2）プレファレンススコアが著しく低いもの

単語了解度は高いにもかかわらず，電気発声音声より不自然であると評価されたターゲットである．図 6 及び図 7 から，合成音声の単語了解度が高い場合でも，電気発声音声との一対比較評価においては自然性が必ずしも高い評価とはならないことが分かる．提案手法は，電気発声を話す喉頭摘出者が，ストレスの少ない音声を用いた健常者との円滑なコミュニケーションが目標である．したがって，出力される合成音声には単語了解度が著しく減少しない限り，自然性を重視する必要があると考えられる．そこで，プレファレンススコアの改善方法を以下に考察する．

表 6　音声データ中の数字の読みの頻度
Table 6　Frequency of all the elements used in speech data.

発話データ	データ数	発話データ	データ数
に	27	じゅう	81
さん	27	いち	9
よん	27	はっ	9
ご	27	ろっ	9
ろく	27	ひゃく	54
なな	27	びゃく	9
きゅう	27	ぴゃく	18

このカテゴリーに属する評価ターゲットの一例 “ひゃくよんじゅうなな（147）” の健常者の自然音声を図 8 に，音声変換による合成音声を図 9 に示す．どちらの図も（A）音声波形，（B）RMS パワー，（C）F0 を示す．それぞれの発話長は，健常者音声では 0.96 s，合成音声は二つの変換ルールから構成され全体で 0.9 s（0.21 s 及び 0.69 s）である．図 9 における点線は，合成音声を構成する変換ルールの接続部分を表す．二つの図を比較すると，第 1 変換ルール部分の F0 及び RMS パワーは，自然発声音声のその位置の F0 と RMS パワーとは大きく異なることが分かる．こうした部分的な F0 やパワーの相違は，合成音声の知覚的な印象に影響を与える原因であると考えられる．そのため，音響的特徴のみではなく，このような F0 や RMS パワーなどの韻律情報も変換ルールに含めるこ

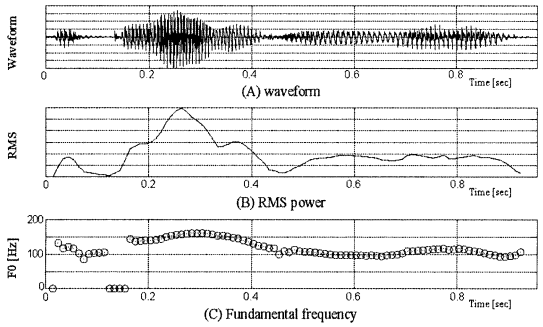

図 8　健常者による自然音声
Fig. 8　Natural speech by healthy speaker.

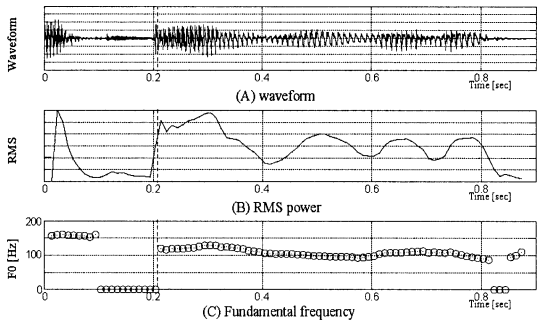

図 9　合 成 音 声
Fig. 9　Synthesized speech.

とを検討している．これにより，不自然なイントネーションとなるルールの組合せを抑制して健常者音声側の変換ルールを合成できるため，より自然性の高い音声を合成できると考えられる．

8. む す び

本論文では，電気発声法を用いる喉頭摘出者を対象に，より円滑なコミュニケーションを目的とした，電気発声音声の発話内容を保持したまま，波形接続型音声合成を用いて健常者音声に変換する音声変換手法を提案した．本手法では，電気発声音声及び健常者音声そのものに着目した．それぞれの音声側で DP マッチングを用いた発話組の比較により，音響的な共通部分と差異部分として抽出される音声素片のみを音声合成に用いた．音声データの発話内容を 3 けたの数字に限定した中で音声変換を実現した．

評価実験を行い，平均単語了解度は 95.2%，オピニオン評価は 3.89 が得られた．一方，健常者による自然音声ではオピニオン評価は 4.98，電気発声音声によるオピニオン評価は 2.93 であった．また，電気発声音声と合成音声との一対比較評価では，72.3%のプレファレンススコアを得た．

これらの結果より，数字発声音声に限定してはいるが，音声合成による了解度の劣化を最小限に抑えることができた．また，健常者の自然音声に比べると音質は低いものの，電気発声音声と比較すると自然性がより高い合成音声が得られたことが示された．こうした結果から本手法は，健常者との簡単なコミュニケーションに適応できるだけでなく，例えば健常者音声用の簡単な音声案内など，発話時間が 1 秒前後の 10 モーラほどの発話で処理が進むような用途に対して発話技能の高い電気発声者音声の適用が可能であると考えられる．

今後の課題としては，合成音声を構成する健常者音声側の変換ルールの選択時に韻律情報を利用することで，合成音声全体でより高い自然性を実現することが挙げられる．また，変換ルールを獲得するための音声データが増加した場合，使われない変換ルールを淘汰するなどの変換ルール辞書の大きさを適切に制御するためのアルゴリズムを検討する必要がある．コミュニケーションを目的としたツールとしての実用性には，まだ多くの課題が残る．そこで，通常会話を音声データとして本手法を適用し，有効性を検討する必要がある．この場合，多くの音声データをあらかじめ用意する必要があるが，発話者である喉頭摘出者の負担の軽減を考慮しなければならない．そのため，我々は個人に特化して本手法を適用することを考えている．その結果，逐次的に音声データを増やし，本手法を用いて多数の適切な変換ルールを獲得できるため，発話者の負担を抑えながら本システムの性能を向上することができる．

更に，本手法は基本的に話者の音声に依存しない汎用的な手法であるため，様々なレベルの電気発声音声だけでなく，食道発声音声や他の障害者音声も対象にした実験を行っていきたい．

謝辞　本研究を進める上で，御指導及び音声データ録音に御協力頂いた北鈴会の皆様に深く感謝申し上げます．

文　献

[1] 高藤次夫，“喉頭摘出者の発声の現状，” JOHNS, vol.2, no.5, pp.527–531, 1986.
[2] 伊福部達，音の福祉工学，コロナ社，1997.
[3] 板倉　淳，“人工喉頭音声，” 音声言語医学会論文誌，vol.39, no.4, 1998.
[4] http://www.animo.co.jp/

[5] 花田英輔，傍島康雄，天白成一，松村康児，楠原浩一，原　寿郎，津本周作，野瀬義明，“PDA を利用した発声障害者向け日本語会話補助装置，”信学技報，SP2002-103, WIT2002-43, 2002.
[6] K. Murakami, K. Araki, M. Hiroshige, and K. Tochinai, “Effectiveness of a direct speech transform method using inductive learning from laryngectomee speech to normal speech,” Proc. 16th Australian Joint Conference on Artificial Intelligence (LNAI), pp.686–698, 2003.
[7] 丁　文，樋口宜男，“Complex RBF ネットワークを用いた音声変換方法，”音響講論集，1-P-16, pp.335–336, 1997.
[8] O. Turk and L.M. Arslan, “Subband based voice conversion,” Proc. ICSLP2002, pp.289–292, 2002.
[9] 田窪行則，前川喜久雄，窪薗晴夫，本多清志，白井克彦，中川聖一，音声，岩波書店，1998.
[10] 中田和男，音声，コロナ社，1977.
[11] 岸裕次郎，船田哲男，“電気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, 1996.
[12] 大森孝一，児嶋久剛，“振動部からみた喉頭摘出手術後の代用音声—文献的考察，”耳鼻咽頭科臨床，vol.83, no.6, pp.945–952, 1990.
[13] 高橋宏明，永田誠治，“無喉頭音声，”耳鼻咽頭科・頭頚部外科 MOOK，no.4, pp.29–38, 1987.
[14] 佐藤武男，食道発声法，金原出版，1993.
[15] 菊地義信，粕谷英樹，“電気式人工喉頭の f0 制御に関する検討，”信学技報，SP2002-106, WIT2002-46, 2002.
[16] C.Y. Espy-Wilson, V.R. Chari, and C.B. Huang, “Enhancement of alaryngeal speech by adaptive filtering,” Proc. ICSLP ’96, vol.2, pp.764–767, 1996.
[17] 菊地義信，粕谷英樹，“F0 制御機能を有する電気喉頭の試作，”音響講論集，2-10-5, pp.295–296, 2002.
[18] 上見憲弘，伊福部達，高橋　誠，松島純一，“ピッチ周波数制御型電気式人工咽頭の提案とその評価，”信学論（D-II），vol.J78-D-II, no.3, pp.571–578, March 1995.
[19] 上見憲弘，橋場参生，須貝保徳，山口悦範，伊福部達，“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, 1999.
[20] 上見憲弘，橋場参生，伊福部達，“抑揚制御型人工喉頭の問題点と改良方法について，”信学技報，SP2000-44, 2000.
[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, 2001.
[22] 村上浩司，広重真人，荒木健治，栃内香次，“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，”音響講論集，1-P-18, pp.391–392, 2001.
[23] 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, 2002.
[24] 加藤靖佳，“音声障害者音声の音響的特徴，”音響講論集，2-P-14, pp.309–310, 2000.
[25] Y. Qi and B. Weinberg, “Low-frequency energy deficit in electrolryngeal speech,” J. Speech Lang. Hear. Res., vol.34, pp.1250–1256, 1991.
[26] M.S. Weiss, G.H. Yeni-Komshian, and J.M. Heinz, “Acoustical and perceptual characteristics of speech produced with electronic artificial larynx,” J. Acoust. Soc. Am., vol.65, no.5, pp.1298–1308, 1979.
[27] Entropic research laboratories, esps/waves+ with ensig5.3, reference release, 1998.
[28] L.R. Rabiner, A.E. Rosenberg, and S.E. Levinson, “Consideration in dynamic time warping algorithms for discrete word recognition,” IEEE Trans. Acoust. Speech Signal Process., vol.ASSP-25, no.6, pp.575–582, 1978.
[29] S.E. Levinson, “Structural methods in automatic speech recognition,” Proc. IEEE, vol.73, no.11, pp.1625–1650, 1985.
[30] 速水　悟，岡　隆一，“連続 DP による連続単語認識実験とその考察，”信学論（D），vol.J67-D, no.6, pp.677–684, June 1984.
[31] 村上浩司，広重真人，荒木健治，栃内香次，“喉頭摘出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有効性の一検討，”FIT2003 第 3 分冊，pp.489–490, 2003.

（平成 16 年 1 月 5 日受付，5 月 15 日再受付）

村上　浩司　（正員）

平 6 釧路工業高等専門学校情報工学卒．平 8 室蘭工大・情報工学卒．平 13 北海道大学大学院工学研究科修士課程了．平 16 同大大学院博士課程単位取得退学．同年，ニューヨーク大学コンピュータサイエンス学科アシスタントリサーチサイエンティスト，現在に至る．博士（工学）．主として音声情報処理，自然言語処理の研究に従事．言語処理学会，日本音響学会各会員．

荒木　健治　（正員）

昭 57 北大・工・電子卒．昭 63 同大大学院博士課程了．工博．同年，北海学園大学工学部電子情報工学科助手．平元同講師．平 3 同助教授．平 10 同教授．平 10 北大・工・電子情報工学専攻助教授．平 14 同教授，平 16 北大・情報科学・メディアネットワーク専攻教授．自然言語処理，特に，機械翻訳，音声対話処理などの研究に従事．情報処理学会，人工知能学会，言語処理学会，日本認知科学会，ACL，IEEE 各会員．

広重　真人　（正員）

昭 62 北大・工・電子卒，平 4 同大大学院博士課程了．平 4 北大・工・電子助手．平 15 北海道教育大学釧路校講師，現在に至る．音声情報処理，特に発話速度を中心とした韻律情報処理の研究に従事．工博．日本音響学会，日本音声学会，情報処理学会，IEEE，ASA，ISCA 各会員．

栃内　香次　（正員）

昭 37 北大・工・電気卒．昭 39 同大大学院工学研究科電気工学専攻修士課程了．北大大学院工学研究科電子情報工学専攻教授を経て，現在北海学園大学経営学部教授．主として音声情報処理，自然言語処理の研究に従事．工博．情報処理学会，日本音響学会各会員．