

学 位 論 文 題 名

音声波形の音響的特徴を用いた音声変換手法の 有効性に関する研究

学位論文内容の要旨

本論文では、これまで提案されている音声変換手法とは大きく異なる、新しいパラダイムの音声変換手法の提案を行い、その有効性について検討した。

音声や言語における話者変換や翻訳などの情報変換の条件を、対象となる2種類の音声または言語間で、同じ意味を持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。そこで、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、同様に対応関係を抽出することが可能であると考えられる。音響的特徴には、周波数特性の時間変化を用いることとした。

本手法は、次のような手順により実現される。発話内容が同一である2種類の言語の音声発話を複数用意し、それぞれの言語側で独立し、2つの発話データを比較し音響的な差異・共通部分に分離する。このとき、抽出された差異・共通部分の個数が同じ場合にのみ、それらの対応関係をルールとして獲得する。つまりこのルールは、両言語の文中における対応する差異・共通部分の位置情報を保持している。こうしたルールを多数獲得し、ルール辞書に登録する。ルール獲得により得られた辞書中のルールを、入力される未知の入力に適用する。そして部分的に対応するルールを候補として、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ルールが保持する、部分的な目的言語となる音声側の素片を波形接続型音声合成により合成し、結果として出力する。

2言語音声発話を対象とした音声翻訳手法は、入力された原言語の発話を文字列に変換する音声認識部、そこで得られた発話内容の文字列を目的言語の文字列に変換する機械翻訳部、そして翻訳された目的言語の文字列を音声に変換する音声合成部の3つの処理部から構成される。また、それぞれの処理部もまた複雑な構造を持つ。すなわち、それぞれの処理の精度の低下は全体の精度の低下に著しく影響を与えるため、すべての処理の精度の向上のみが全体の精度を向上させる手段となる。また、形態素解析や構文解析は言語の種類に依存するため、対象となる言語の変更に伴い、これらの解析も変更する必要がある。

そこで、音声の音響的特徴を用いた音声変換手法をこの音声翻訳に適用して、原言語から目的言語へ音声を変換することを試みた。本手法は各固有の言語の種類に依存する処理を持たないため、より汎用的な音声翻訳の枠組みとなる。また、音声の周波数特徴の時間変化に着目し、その変化によって音響的差異・共通部分を決定するため、単語、音節、音

素などの正確な境界は必ずしも必要ではなく、比較する音声の音響的な差異・共通部分の両方に着目してその関係を利用する。

簡単な発話から構成されるデータセットを用いた評価では、数例の翻訳結果が得られ、語彙の限定された発話においての手法の有効性が示された。

旅行発話データセットを用いた評価においても、同様に数例の翻訳結果が得られたことから、音響的差異・共通部分のみに着目することで翻訳に的確に用いることのできるルールが獲得できることが示された。

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約20,000人弱と推定され、年々増加傾向にある。こうした人々は食道発声法や電気式人工喉頭による発声法などにより、代用音声と呼ばれる音声を用いることで発話が可能となる。しかしながら、代用音声による発話には、不自然さが聞き手だけでなく話し手にもストレスになることが少なくない。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が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望まれている。

そこで、音声の音響的特徴を用いた音声変換手法をこうした喉頭摘出者の音声に適用し、健常者の音声への変換を試みた。変換対象となる音声は喉頭摘出者音声、健常者音声とも日本語であるため、音声翻訳に比べて音声波形の時間的類似度を用いることができるという利点がある。対象とする喉頭摘出者音声は、代表的なものとして広く用いられている、電気式喉頭を用いて発声する電気発声音声、および食道発声音声である。ここでは共通部分の始点探索にDPマッチングを適用した。喉頭摘出者音声は健常者音声とは違い、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、まず少数の音声データを用いて、音声変換手法が喉頭摘出者音声と健常者音声との変換に対して有効である可能性を確認することとした。また、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、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結果、重複のない3桁の数字発話をデータとした。

同内容の喉頭摘出者音声と健常者音声の発話データを複数用意し、音声の周波数スペクトルの時間変化を表すベクトルに変換する。そして喉頭摘出者音声と健常者音声のそれぞれでDPマッチングを独立して行い、音響的な差異・共通部分に分離する。このとき、それぞれの音声側で分離された差異・共通部分は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。このような変換ルール獲得を全ての発話データに繰り返し行う。変換ルール獲得により得られた辞書中の変換ルールを、入力される未知の喉頭摘出者音声に適用する。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、この時点で採用された変換ルールが保持する、部分的な健常者音声素片を波形接続型音声合成により適宜合成して、変換結果として出力する。

手法の有効性を確認するために、食道発声音声と電気発声音声についてそれぞれ主観評価実験を行った。実験は単語了解度試験、オピニオン評価、合成音声と食道発声音声もしくは電気発声音声との一対比較によるプレファレンススコア算出の3種類を行った。

食道発声音声に対しては、単語了解度 80.0%（食道発声音声：98.46%，健常者音声：100%），オピニオン評価 3.27（食道発声音声：3.89，健常者音声：4.98），プレファレンススコア 50.00%が得られた。

電気発声音声に対しては、単語了解度 95.2%（電気発声音声：99.5%，健常者音声：100%），オピニオン評価 3.89（電気発声音声：2.93，健常者音声：4.98），プレファレンススコア 72.3%が得られ，明らかな有効性を確認した。

これらの結果より，提案した音声変換手法が音声翻訳及び口頭摘出者音声の健常者音声への変換に有効であることが確認された。

学位論文審査の要旨

主 査 教 授 荒 木 健 治
副 査 教 授 青 木 由 直
副 査 教 授 北 島 秀 夫
副 査 助 教 授 伊 藤 敏 彦

学 位 論 文 題 名

音声波形の音響的特徴を用いた音声変換手法の 有効性に関する研究

著者は、これまで提案されている音声変換手法とは大きく異なる、新しいパラダイムの音声変換手法の提案を行い、その有効性について検討した。

音声や言語における話者変換や翻訳などの情報変換の条件を、対象となる2種類の音声または言語間で、同じ意味を持つことと仮定する。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の抽出は可能である。そこで、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、同様に対応関係を抽出することが可能であると考えられる。音響的特徴には、周波数特性の時間変化を用いることとした。

提案手法は、次のような手順により実現される。発話内容が同一である2種類の言語の音声発話を複数用意し、それぞれの言語側で独立して、2つの発話データを比較し音響的な差異・共通部分に分離する。このとき、抽出された差異・共通部分の個数が同じ場合にのみ、それらの対応関係をルールとして獲得する。つまりこのルールは、両言語の文中における対応する差異・共通部分の位置情報を保持している。こうしたルールを多数獲得し、ルール辞書に登録する。ルール獲得により得られた辞書内のルールを、入力される未知の入力に適用する。そして部分的に対応するルールを候補として、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ルールが保持する、部分的な目的言語となる音声側の素片を波形接続型音声合成により合成し、結果として出力する。

まず、音声の音響的特徴を用いた音声変換手法を日英音声翻訳に適用して、原言語から目的言語への音声変換を考える。2言語音声発話を対象とした従来の音声翻訳手法は、入力された原言語の発話を文字列に変換する音声認識部、そこで得られた発話内容の文字列を目的言語の文字列に変換する機械翻訳部、そして翻訳された目的言語の文字列を音声に変換する音声合成部の3つの処理部から構成される。また、それぞれの処理部もまた複雑な構造を持つ。すなわち、それぞれの処理の精度の低下は全体の精度の低下に著しく影響を与えるため、すべての処理の精度の向上のみが全体の精度を向上させる手段となる。また、形態素解析や構文解析は言語の種類に依存するため、対象となる言語の変更に伴い、

これらの解析も変更する必要がある。本手法は言語の種類に依存する処理を持たないため、より汎用的な音声翻訳の枠組みとなる。また、音声の周波数特徴の時間変化に着目し、その変化によって音響的差異・共通部分を決定するため、単語や音節、音素などの正確な境界は必ずしも必要ではなく、比較する音声の音響的な差異・共通部分の両方に着目してその関係を利用する。旅行者用英会話に出現する発話のデータセットを用いた評価実験において、数例の翻訳結果が得られたことから、音響的差異・共通部分のみに着目することで翻訳に的確に用いることのできるルールが獲得できることが示された。

次に、音声の音響的特徴を用いた音声変換手法を喉頭摘出者の音声に適用して、健常者の音声への変換を行った。変換対象となる喉頭摘出者音声は、喉頭摘出手術を受けた結果、失声した人が用いる代用音声であり、そのままでは発話は不自然で聞き手と話し手の両方にストレスを与えることが少なくない。そのため、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が可能な、より自然な音声を用いたコミュニケーションツールが望まれている。喉頭摘出者音声、健常者音声とも日本語であるため、音声翻訳に比べて音声波形の時間的類似度を用いることができるという利点がある。対象とする喉頭摘出者音声は、代表的なものとして広く用いられている、電気式喉頭を用いて発声する電気発声音声である。ここでは共通部分の始点探索に DP マッチングを適用した。喉頭摘出者音声は健常者音声とは違い、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、まず少数の音声データを用いて、音声変換手法が喉頭摘出者音声と健常者音声との変換に対して有効である可能性を確認することとした。また、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、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結果、重複のない3桁の数字発話をデータとした。

同内容の喉頭摘出者音声と健常者音声の発話データを複数用意し、音声の周波数スペクトルの時間変化を表すベクトルに変換する。そして喉頭摘出者音声と健常者音声のそれぞれで DP マッチングを独立して行い、音響的な差異・共通部分に分離する。このとき、それぞれの音声側で分離された差異・共通部分は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。このような変換ルール獲得を全ての発話データに繰り返し行う。変換ルール獲得により得られた辞書内の変換ルールを、入力される未知の喉頭摘出者音声に適用する。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、この時点で採用された変換ルールが保持する、部分的な健常者音声素片を波形接続型音声合成により適宜合成して、変換結果として出力する。

手法の有効性を確認するために、電気発声音声を対象に主観評価実験を行った。実験は単語了解度試験、オピニオン評価、合成音声と電気発声音声との一対比較によるプレファレンススコア算出の3種類を行った。実験の結果、単語了解度 95.2%（電気発声音声：99.5%，健常者音声：100%）、オピニオン評価 3.89（電気発声音声：2.93，健常者音声：4.98）、プレファレンススコア 72.3%が得られ、明らかな有効性を確認した。

著者は論文全体を通じて、研究領域の現状の分析、新規提案内容の記述、有効性の主張、研究領域における位置付けを正確に行ったと判定する。

以上を要約すると、著者は音響的特徴のみを用いた音声変換手法を提案し、日英音声翻訳と喉頭摘出者音声の健常者音声への変換に適用した結果、文字情報を用いずに、音響的

特徴のみを用いた音声変換の実現が可能であることを示した。本研究を通じての、情報メディア工学，音声処理工学，自然言語処理工学の発展に貢献するところ大なるものがある。よって，著者は北海道大学博士（工学）の学位を授与される資格あるものと認める。