

令和 4 年 6 月 1 日現在

機関番号：14501

研究種目：若手研究

研究期間：2020～2021

課題番号：20K19862

研究課題名（和文）構音障害者のための少量学習データでの音声認識モデル構築

研究課題名（英文）Construction of acoustic model on small training data for dysarthric speech recognition

研究代表者

高島 遼一（Takashima, Ryoichi）

神戸大学・都市安全研究センター・准教授

研究者番号：50846102

交付決定額（研究期間全体）：（直接経費） 2,700,000円

研究成果の概要（和文）：本研究は構音障害者のコミュニケーション支援のための音声認識システムの構築を目的とする。音声認識モデルの学習において、学習に十分な量の構音障害者音声の収集が困難な点が課題である。そこで本研究では、大量の健常者音声を学習に利用するための転移学習と、比較的収集が容易な障害者の自由発話音声を利用するための自己教師有り学習を用いたモデル学習手法を検討した。前者では健常者モデルから対象の障害者モデルへ段階的に近づける多段階転移学習を提案し、また後者では疑似ラベルと自己教師有り学習を組み合わせる手法などを提案し、いずれも従来法より高い性能を確認した。

研究成果の学術的意義や社会的意義

本研究は構音障害者の社会的バリアを除去し、社会参加を支援することへの貢献が期待されるものである。運動障害に起因する構音障害者は手足も不自由なために手話などのコミュニケーションの代替手段が取れないケースも多いため、高精度な音声認識の実現が求められている。また本研究で特に焦点を当てている、学習データ不足の問題は構音障害者に限らず音声認識全般にわたって存在する課題であるため、本研究で提案した多段階転移および疑似ラベルと自己教師有り学習に基づく音声認識モデルの学習手法は、音声認識の広い分野において応用可能であると期待している。

研究成果の概要（英文）：The goal of this research is to build a dysarthric speech recognition system as a communication tool for dysarthric people. One of the challenges of this research is that it is difficult to collect a sufficient amount of dysarthric speech for training the speech recognition model. In this study, we studied model training methods using transfer learning to use a large amount of normal speech and self-supervised learning to use spontaneous speech of dysarthric people, which is relatively easy to collect. In this study, we proposed multi-step transfer learning, and we proposed a method combining pseudo-labelling and self-supervised learning. We confirmed both of our proposed methods showed better performance than conventional methods.

研究分野：メディア情報処理

キーワード：音声認識 構音障害 障害者支援技術 機械学習 ニューラルネットワーク

1. 研究開始当初の背景

現在の日本において、身体・知的・精神の各障害者の数は合計で約 960 万人存在しており、日本国民のおよそ 7.6%が何らかの障害を有していると報告されている[1]。日本では障害者の社会参加を促進すべく、法改正などの施策が近年積極的になされている。例えば、障害者を理由とする差別の解消を目的とする障害者差別解消法が 2016 年に施行された。また障害者雇用促進法によって義務付けられている障害者雇用率は、2013 年以来 2.0%であったのに対して 2018 年より 2.2%に引き上げられ、さらに 2021 年には 2.3%に引き上げられることが決定している。このような社会的背景からも、障害者に対する社会的バリアを除去し、社会参加を支援するための技術が、より緊急性を増して必要とされている。

本研究では、構音障害者のコミュニケーション支援を目的とした、構音障害者向け音声認識技術の開発を行う。構音障害とは、言葉を理解しているにもかかわらず、発声の動作に問題があり、上手く発音できない障害である。構音障害は口腔の問題だけでなく、脳や神経の障害に起因する運動障害が原因であるケースも多く、この場合は発話だけでなく手足も不自由な患者も多い。例えば脳性麻痺による構音障害者の場合、その多くは筆記やタイピング、手話といったコミュニケーションの代替手段がとれず、会話に頼らざるを得ないというのが実情である。そのため、構音障害者にとって、自分の発話を理解してもらいたいというニーズは極めて高い。構音障害者の音声を高精度で認識できれば、機械の音声操作による生活支援や、認識結果を用いたコミュニケーション支援などに、大いに役立てられると期待される。

[1] 内閣府 令和元年版 障害者白書

<https://www8.cao.go.jp/shougai/whitepaper/r01hakusho/zenbun/index-w.html>

2. 研究の目的

本研究の目的は、構音障害者向けの高精度な音声認識を実現することである。音声認識技術の性能は近年大きく向上し、より身近なものとなっている。しかし、一般的な音声認識システムは、健常者の音声を用いて学習されているため、健常者と発声スタイルが大きく異なる構音障害者の音声を認識することが困難である。図 1 は、健常者音声で学習させた音声認識モデルを用いて、健常者と構音障害者の音声を認識した結果である。左図は機械への音声コマンド入力を想定した単語認識実験、右図はコミュニケーション支援を想定した自由発話認識実験の単語誤り率を示しているが、いずれの実験でも構音障害者音声の認識誤り率が 100%に近く、ほぼ全く認識できていないことが分かる。

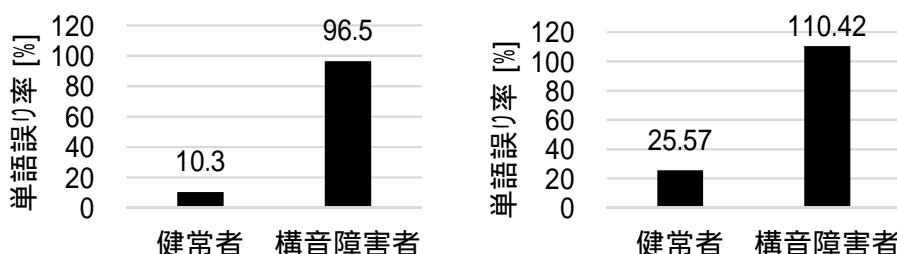


図 1. 健常者音声で学習した音声認識モデルによる、健常者と構音障害者の単語認識結果(約 200 語彙、左図)および自由発話認識結果(約 7 万語彙、右図)

高精度な音声認識実現のためには、個々の障害者に合わせた音声認識モデルを学習する必要がある。しかし、従来の音声認識手法は大量の学習データを必要とするのに対して、構音障害者から十分な量の学習音声データを収集することは困難である。例えば、製品レベルの健常者向け音声認識システムは数千～数万時間規模の音声データを用いて学習されているのに対して、我々が独自に収集している構音障害者の音声は一人につき数時間程度である。そのため本研究では、限られた学習データを用いて、いかにして高精度な構音障害者音声認識モデルを学習するかという課題に焦点を当てた。

3. 研究の方法

学習データ量の不足という課題に対して、従来から多くの研究が取り組まれている。中でも近年注目されているのが転移学習と自己教師有り学習である。転移学習は、あるタスクに対して大量のデータを用いて学習させたモデルを、別の似たタスクに少量データを用いて適応させる手法である。この手法を使うことで、大量の健常者音声で学習したモデルを、少量のデータで特定構音障害者音声へ適応できると期待される。しかし適応元の健常者音声と適応先の構音障害者音声には、その特徴分布の間に大きなギャップがあるため、従来の転移学習を用いても構音障害者データが少量の場合では十分な効果が得られないことが事前の検証により判明した。そのた

め、転移学習をベースとしてさらなる改良を検討した。

自己教師有り学習は自然言語処理分野における BERT や GPT-3 といった高性能な言語モデルの学習にも使われている手法であり、ラベルの無い大量のデータをモデル学習に利用するアプローチの総称である。音声認識モデルの学習には、音声とその発話内容（ラベル）のペアデータが必要であるが、この学習データを収集するために、従来は台本を用意して障害者の方に読んでもらっていた。しかしこの方法では障害者の負担が大きいと、多くのデータを収録できなかった。一方、台本読み上げ音声の代わりに、常時収録可能な構音障害者の自由発話音声を収録する方法も考えられるが、発話内容を書き起こすコストが課題であった。そこで本研究では、自己教師有り学習により、テキスト情報を使わずに日常会話音声を音声認識モデルの学習に利用することを検討する。自己教師有り学習は既に健常者音声認識において有効性が確認されているが、学習データが少量である構音障害者音声認識においてはまだ有効性が確認されていない。本研究では、自己教師有り学習を構音障害者音声認識において有効性を確認し、さらなる手法の改善を試みる。

4. 研究成果

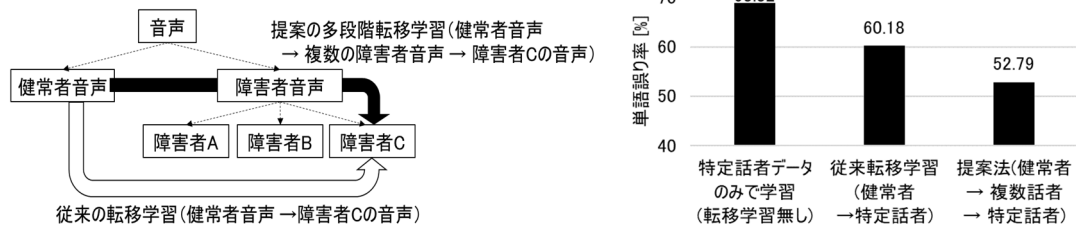


図 2. 提案の多段階転移学習の概要(左)と、構音障害者音声認識タスクでの評価結果(右)

転移学習を応用したモデル学習手法として、多段階転移学習と呼ぶ手法を開発した[2]。提案手法の概要を図 2 の左に示す。従来の転移学習では、まず大量の健常者音声をを用いてモデルを学習し、その後目的の障害者（図の例では障害者 C）の音声データを用いて fine-tuning する。一方提案法では、障害者 A や B のデータも混ぜた「不特定の障害者音声」を用いて fine-tuning した後に障害者 C のデータでさらに fine-tuning を行う。これにより、健常者と目的話者のデータの性質が大きく異なる場合において、段階を踏んでモデルを目的話者へ適応することが可能となり、安定したモデル学習が可能となる。我々が収集した脳性麻痺患者の音声をを用いて本手法を評価した結果、健常者で事前学習して特定障害者へ fine-tuning する従来法に比べて、一旦複数の障害者へ fine-tuning してから特定障害者へ fine-tuning する提案法の方が、音声認識誤り率が小さくなることが明らかとなった（図 2 右）。

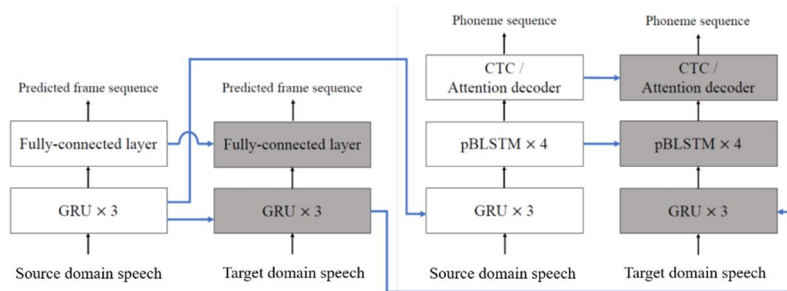


図 3. 自己教師有り学習と転移学習の組み合わせによるモデル学習法

自己教師有り学習を用いたモデル学習手法として、自己教師有り学習と転移学習を組み合わせた手法も開発した[3]。ラベル無し自由発話は比較的多くのデータを収集することができるが、構音障害者の音声に関しては健常者と比べると依然少ない。そのため、構音障害者音声のみで自己教師有り学習を行った場合、データ量不足により有用な初期モデルが学習できない可能性がある。そこで本研究では、健常者の不特定話者モデルから構音障害者の特定話者モデルを構築する転移学習を行う。自己教師有り学習と転移学習を組み合わせる場合、様々な学習手順が考えられるが、我々が検討した結果、図 3 のような手順で学習するのが最も音声認識性能が良いということが分かった。この手法も構音障害者音声認識タスクで評価した結果、自己教師有り学習も転移学習も用いないベースラインの音声認識誤り率が 29.9%であったのに対して、開発手法を用いた場合では 10.9%の大幅な性能改善を示した。

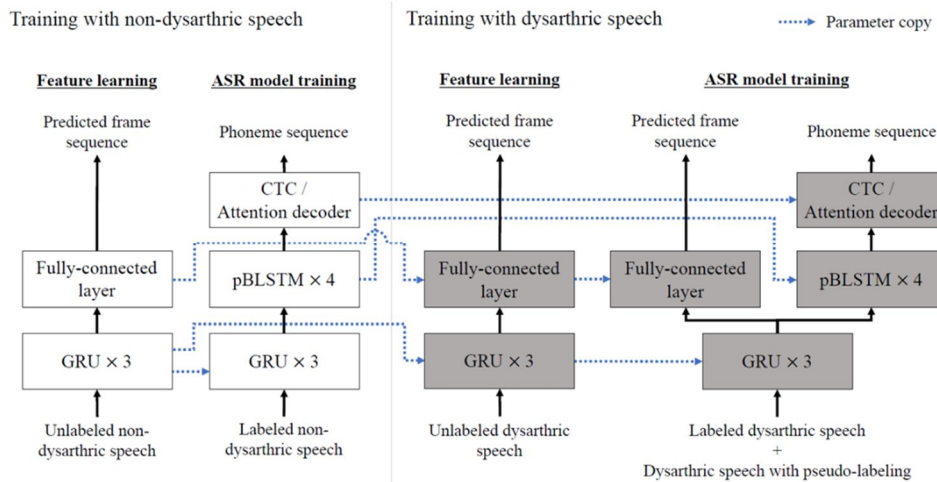


図4．自己教師有り学習と転移学習と疑似ラベルの組み合わせによるモデル学習法

前に述べた手法をさらに改善する手法として、疑似ラベルを組み合わせた手法も開発した[4]。疑似ラベルとは、自己教師有り学習と同じくラベルの無いデータをモデル学習に利用する方法の一つであり、ラベルの無い音声に対して音声認識を行い、その結果を疑似的なラベル情報に用いる手法である。自己教師有り学習は発話テキスト情報を一切用いないため、音声認識モデルを学習する上では間接的な効果しかないのに対して、疑似ラベルは発話テキスト情報を使用するため、音声認識モデルの学習に直接的に働くという利点がある。一方で、疑似ラベルは音声認識結果を利用しているため、ラベルに誤りが含まれるという欠点もある。そこで、疑似ラベルの誤りの影響を受けない自己教師有り学習と疑似ラベルによる学習の両方を用いて、マルチタスク学習の枠組みによってモデルを学習する手法を提案した。

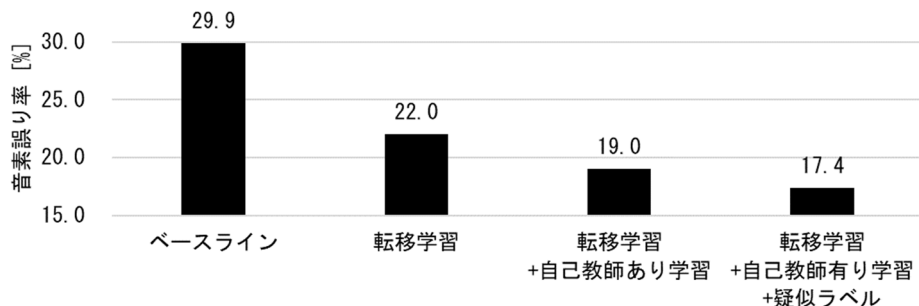


図5．構音障害者の音素認識タスクにおける開発手法の結果

構音障害者音声認識タスクで評価した結果、前に述べた転移学習と自己教師有り学習を組み合わせた手法（19.0%）に対して音素認識誤り率が17.4%まで低下し、開発手法の有効性を確認した。

以上のことから本研究では、構音障害者音声認識を目的として、従来の転移学習を改良した多段階転移学習手法、転移学習と自己教師有り学習を組み合わせた手法、さらに疑似ラベルとも組み合わせた手法を開発し、従来の音声認識性能を大幅に向上することができた。現在自己教師有り学習に用いているアルゴリズムはAutoregressive predictive codingと呼ばれるものであるが、近年ではwav2vec 2.0と呼ばれる強力なアルゴリズムも提案されている。今後は、wav2vec 2.0を始めとする最新のアルゴリズムも用いることで、さらなる性能向上を目指す。

[2] Ryoichi Takashima, Tetsuya Takiguchi, Yasuo Arika, Two-Step Acoustic Model Adaptation for Dysarthric Speech Recognition, ICASSP, pp. 6104-6108, May 2020.

[3] 澤 佑哉, 富士原 健斗, 相原 龍, 高島 遼一, 滝口 哲也, 本山 信明 自己教師あり学習によるラベル無し自由発話を用いた構音障害者音声認識, 日本音響学会 2021 年春季研究発表会講演論文集, 2-2P-2, pp. 1045-1048, 2021-03.

[4] 澤 佑哉, 富士原 健斗, 相原 龍, 高島 遼一, 滝口 哲也, 今井 良枝 疑似ラベリングと特徴表現学習を併用した構音障害者音声認識 日本音響学会 2021 年秋季研究発表会講演論文集, 1-3-2, pp. 847-850, 2021-09.

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計14件（うち招待講演 0件 / うち国際学会 5件）

1. 発表者名 Yuki Takashima, Ryoichi Takashima, Tetsuya Takiguchi, Yasuo Arika
2. 発表標題 Dysarthric Speech Recognition Based on Deep Metric Learning
3. 学会等名 Interspeech (国際学会)
4. 発表年 2020年

1. 発表者名 Yuya Sawa, Ryoichi Takashima, Tetsuya Takiguchi
2. 発表標題 An Investigation of End-to-End Speech Recognition Using Model Adaptation for Dysarthric Speakers
3. 学会等名 IEEE Global Conference on Consumer Electronics (GCCE) (国際学会)
4. 発表年 2020年

1. 発表者名 Ryoichi Takashima, Tetsuya Takiguchi, Yasuo Arika
2. 発表標題 Two-Step Acoustic Model Adaptation for Dysarthric Speech Recognition
3. 学会等名 ICASSP (国際学会)
4. 発表年 2020年

1. 発表者名 富士原 健斗, 高島 遼一, 杉山 千尋, 田中 信和, 野原 幹司, 野崎 一徳, 滝口 哲也
2. 発表標題 口唇口蓋裂者の音声認識のためのデータ拡張方式の検討
3. 学会等名 日本音響学会2021年春季研究発表会講演論文集
4. 発表年 2021年

1. 発表者名 澤 佑哉, 富士原 健斗, 相原 龍, 高島 遼一, 滝口 哲也, 本山 信明
2. 発表標題 自己教師あり学習によるラベル無し自由発話を用いた構音障害者音声認識
3. 学会等名 日本音響学会2021年春季研究発表会講演論文集
4. 発表年 2021年

1. 発表者名 澤 佑哉, 高島 遼一, 滝口 哲也, 有木 康雄
2. 発表標題 構音障害者音声認識における発話辞書適応の検討
3. 学会等名 日本音響学会2020年秋季研究発表会講演論文集
4. 発表年 2020年

1. 発表者名 高島 遼一, 有木 康雄, 滝口 哲也
2. 発表標題 構音障害者音声認識における認識モデルの比較評価
3. 学会等名 日本音響学会2020年秋季研究発表会講演論文集
4. 発表年 2020年

1. 発表者名 富士原 健斗, 高島 遼一, 杉山 千尋, 田中 信和, 野原 幹司, 野崎 一徳, 滝口 哲也
2. 発表標題 誤り訂正に基づく器質性構音障害者の音声認識精度向上の検討
3. 学会等名 日本音響学会2021年秋季研究発表会講演論文集
4. 発表年 2021年

1. 発表者名 澤 佑哉, 富士原 健斗, 相原 龍, 高島 遼一, 滝口 哲也, 今井 良枝
2. 発表標題 擬似ラベリングと特徴表現学習を併用した構音障害者音声認識
3. 学会等名 日本音響学会2021年秋季研究発表会講演論文集
4. 発表年 2021年

1. 発表者名 富士原 健斗, 高島 遼一, 杉山 千尋, 田中 信和, 野原 幹司, 野崎 一徳, 滝口 哲也
2. 発表標題 異なる疾患の障害者音声を用いた器質性構音障害者音声認識モデルの学習
3. 学会等名 日本音響学会2022年春季研究発表会講演論文集
4. 発表年 2022年

1. 発表者名 松坂 勇樹, 高島 遼一, 佐々木 千穂, 滝口 哲也
2. 発表標題 構音障害者音声認識におけるテキスト音声合成によるデータ拡張の検討
3. 学会等名 日本音響学会2022年春季研究発表会講演論文集
4. 発表年 2022年

1. 発表者名 澤 佑哉, 相原 龍, 高島 遼一, 滝口 哲也, 今井 良枝
2. 発表標題 構音障害者音声認識における自己教師あり学習と疑似ラベリングの動的重み付きマルチタスク学習
3. 学会等名 日本音響学会2022年春季研究発表会講演論文集
4. 発表年 2022年

1. 発表者名 Kento Fujiwara, Ryoichi Takashima, Chihiro Sugiyama, Nobukazu Tanaka, Kanji Nohara, Kazunori Nozaki, Tetsuya Takiguchi
2. 発表標題 Data Augmentation Based on Frequency Warping for Recognition of Cleft Palate Speech
3. 学会等名 APSIPA (国際学会)
4. 発表年 2021年

1. 発表者名 Yuya Sawa, Ryoichi Takashima, Tetsuya Takiguchi
2. 発表標題 Adaptation of a Pronunciation Dictionary for Dysarthric Speech Recognition
3. 学会等名 IEEE LifeTech (国際学会)
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

<p>研究者webページ http://www.me.cs.scitec.kobe-u.ac.jp/~rtakashima/</p>

6. 研究組織		
氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------