

構音障害者の連続音声認識の検討*

◎宮本千琴, 松政宏典, 滝口哲也, 有木康雄 (神戸大・工), 李義昭 (大手門大), 中林稔堯 (神戸大・発達)

1 はじめに

近年、音声認識技術の発展に伴い、様々な環境下や場面での利用が期待されている。例えばカーナビゲーションの操作や会議音声の議事録化など様々な分野に応用されている。これらの多くは健常者を対象としており、文献 [1] では、構音障害者音声を対象とした音響モデル適応の検証を行っているが、言語障害者などの障害者を対象としたものは非常に少ない。

言語障害の原因の一つとして、脳性麻痺が考えられる。脳性麻痺は、未成熟の脳における非進行性の欠損もしくは病変に起因する、運動および姿勢の疾患として定義されている。とくに意図的な動作を行う場合や緊張状態にある時に筋肉の制御が難しくなり、不随意運動を伴う。この運動障害の一つとして、正しく構音できない場合がある [2]。本稿では、脳性麻痺により構音機能に障害を持つ被験者を対象に音声認識実験を行う。

我々はこれまで、構音障害者の音響モデルの構築、改善を行ってきた [3]。これは単語発話を対象としているため、本稿では、さらなる試みとして文発話を対象とした構音障害者の音響モデルの構築を行い、音声認識実験の結果を報告する。

2 構音障害者音声

実験用データとして構音障害者 1 名のデータを収録した。発話内容として、

- 読み上げ文：ATR 音素バランス文 (503 文) から 429 文を無作為に選択したもの
- 自然発話文：被験者が行った講演内容 (1,011 文)

を用意した。

読み上げ文データは、1 文を 1 データとして収録した。自然発話文データは、1 文の長さが非常に長いものも含むため、収録者の主観に従って意味上妥当と思われる箇所、及び被験者が一度に発話し易い箇所で区切り、それを 1 データとして収録した。収録時間の内訳を図 1 に示す。自然発話文データの発話内容は、被験者のプロフィールや体験話が主である。発話内容の例を図 2 に示す。

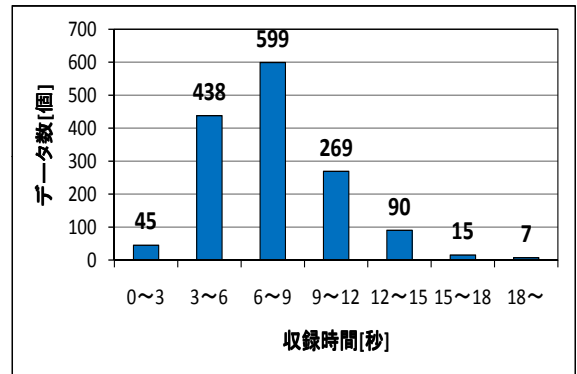


Fig. 1 自然発話文データの内訳

僕の障害は脳性麻痺と言います。
簡単にいえば、例えばコップを持とうとしても、
頭からの指令が手にうまく伝わらず、
コップを落としてしまいます。

Fig. 2 自然発話文データの例

3 認識実験

3.1 実験条件

2 節で収録した音声データを使用して、認識実験を行った。音声データのサンプリング周波数は 16kHz、フレーム窓長は 25msec、フレーム周期は 10msec であり、特徴量は 12 次 MFCC + Δ + Δ パワーを用いた。音響モデルは monophone (43 音素, 16 混合)、言語モデルは trigram、音声認識デコーダは Julius-4.0.2 [4] を用いた。

3.2 汎用モデルでの認識実験

初めに、汎用モデルでの認識実験を行った。評価用データは、健常者データ 429 個、読み上げ文データ 429 個、自然発話文データ 1,463 個を用い、単語正解精度を比較した。健常者データと読み上げ文データは同じ発話内容である。認識結果を図 3 に示す。

健常者では 73.7 % の単語正解精度が得られるが、構音障害者 (読み上げ) では 15.1 % しか認識できず、健常者と発話スタイルが異なるため、汎用モデルでの認識は困難であることがわかる。

*A Study on Continuous Speech Recognition of a Person with Articulation Disorders, by Chikoto Miyamoto, Hironori Matsumasa, Tetsuya Takiguchi, Yasuo Ariki (Kobe Univ.) and Ichao Li (Otemon Univ.) and Toshitaka Nakabayashi (Kobe Univ.)

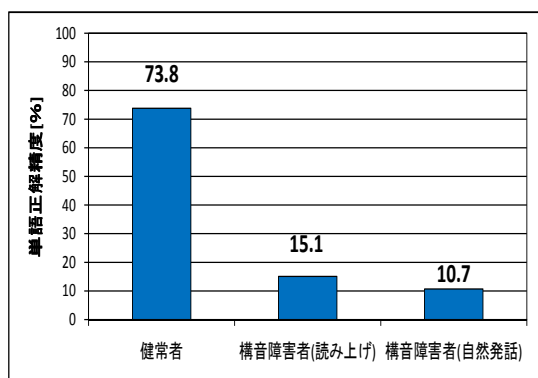


Fig. 3 汎用モデルでの認識実験結果 [%]

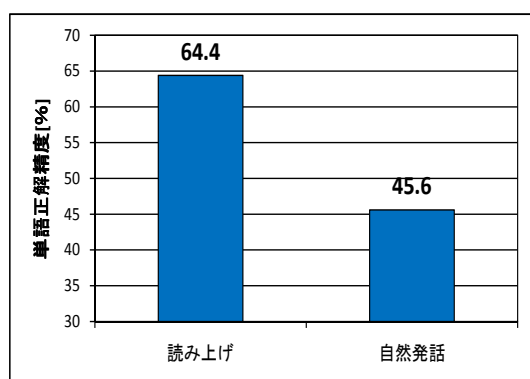


Fig. 4 特定話者モデルでの認識実験結果 [%]

3.3 構音障害者音響モデルでの認識実験

汎用モデルでの認識が困難であることから、構音障害者の音響モデルを作成し認識実験を行った。

音響モデルの初期モデルの作成、学習にはHTK[5]を用いた。言語モデルには、発話内容のみで学習した言語モデルを使用した。実験に使用したデータの内訳を表1に示す。

Table 1 実験データの内訳

	学習用データ	評価用データ
読み上げ文	329	100
自然発話文	1,088	375

3.3.1 実験結果

認識実験結果を図4に示す。音響モデルに特定話者モデルを用い、言語モデルを発話内容のみで学習することで、構音障害者において読み上げ文では64.4%、自然発話文では45.6%の精度が得られた。

ここで、自然発話音声の単語正解精度が読み上げ音声に比べて悪いことがわかる。これは、実際の自然発話内容はいわゆる書き言葉である読み上げ発話内容とは異なり、口語表現や言い直しがあるためと考え

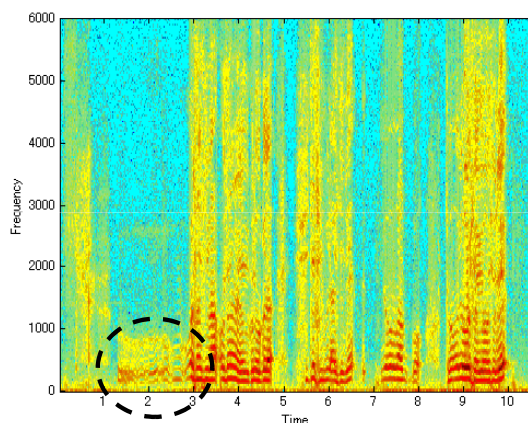


Fig. 5 構音障害者(自然発話文)のスペクトログラム例

られる。また、読み上げ音声に比べて自然発話音声には、構音障害者特有の言いよどみ(吃音)が多く見られる。言いよどみが見られた発話の一例のスペクトログラムを図5に示す。およそ1.5~3秒の発話部分に無音状態でも発声状態でもない音声が含まれていることがわかる。この言いよどみに相当する音素は音響モデル・言語モデルにおいて定義されていないため、認識誤りとなる原因と考えられる。

4 おわりに

構音障害者の連続音声認識の検討を行った。構音障害者の発話スタイルが健常者と異なるため汎用モデルの認識は困難であるが、構音障害者の音響モデルを用いることで、単語正解精度の改善が得られた。

今後は、本稿で考えられた構音障害者特有の問題に対する改善や構音障害者特有の特徴量の検討に取り組んでいく。また、単語発話と文発話の発話スタイルが異なることにより生じる問題について検討する予定である。

参考文献

- [1] 中村 他, “発話障害者音声を対象にした健常者音響モデルの適応と検証,” 音講論(秋), 109-110, 2005.
- [2] S. Terry Canale *et al.*, “キャンベル整形外科手術書 第4巻,” エルゼビア・ジャパン, 2004.
- [3] H. Matsumasa *et al.*, “Integration of Metamodel and Acoustic Model for Speech Recognition,” Interspeech2008, 2234-2237, 2008.
- [4] “大語彙連続音声認識エンジン Julius,” <http://julius.sourceforge.jp/>.
- [5] “HTK (Hidden Markov Model Toolkit),” <http://htk.eng.cam.ac.uk/>.