

構音障害者の音声認識の検討*

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

1 はじめに

音声認識技術は現在, 様々な環境下や場面において使用される機会が増加している. 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, 脳性マヒが考えられ, その分類の一つとして, アテトーゼ型がある. アテトーゼ型とは脳性マヒ患者の約 10~15% に発生する症状であり, アテトーゼと呼ばれる不随意運動が伴う型である. アテトーゼは緊張時や意図的な動作を行う際に出現しやすい. 症状は軽度から重度まで様々であり, 知能障害を合併していないケースや比較的知能障害の程度が軽いケースも多いのが特徴である [2]. そこで本稿では, まず知能障害を合併していないアテトーゼ型に着目した. アテトーゼ型の構音障害者の場合, 最初の動作において緊張状態により, 通常よりも不安定になる場合がある. そこで本稿では, 複数回連続発話の収録を行ない, アテトーゼによる発話スタイルの変動の影響を調べる.

従来の音声認識では, MFCC を特徴量として用いるが, 本稿では離散コサイン変換ではなく 2 回目以降のより安定したデータを利用した, PCA (Principal Component Analysis) による発話変動にロバストな手法を提案し, その有効性を示す.

2 PCA を用いた特徴量抽出

2.1 提案手法

従来の音声認識システムでは, MFCC (Mel Frequency Cepstral Coefficient) が音声特徴量として用いられている. MFCC では, 離散コサイン変換 (Discrete Cosine Transformation: DCT) を適用し, ケプストラムが得られる. 本稿では, 発話スタイルの変動にロバストな特徴量抽出法として, 離散コサイン変換の代わりに PCA を用いた手法を検討する (図 1). 文献 [3] では残響下でのロバストな特徴量抽出として対数スペクトル上で, DCT を行わず PCA を行っている. 本稿では 2 回目以降の安定した発話を用いて主軸を求め, 1 回目の発話 (調音不安定音声) に対して対数スペクトル上で PCA を適用する. 本手法において, 安定した音声成分は低次に, 発話スタイル変動成分は高次に集まると期待できる. この結果 PCA により発話スタイル変動成分の抑圧が行われ, より有効なスペクトル情報の抽出が可能となる.

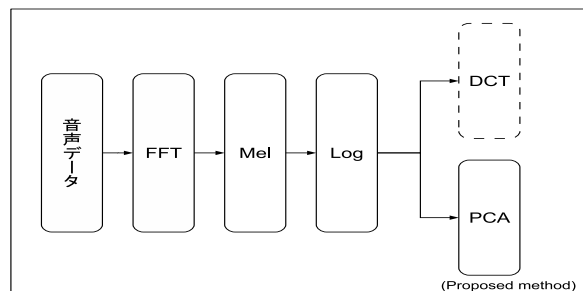


Fig. 1 PCA を用いた特徴量抽出

2.2 発話スタイル変動成分の抑圧

短時間分析によって得られたフレーム n , 周波数 ω の観測音声を $X_n(\omega)$, クリーン音声を $S_n(\omega)$ とする. ここで 1 回目の音声を以下の式で表現する.

$$X_n(\omega) = S_n(\omega)H(\omega) \quad (1)$$

1 回目発話には発話スタイル変動成分 H が畳み込まれていると考える. 次に対数変換を行い観測信号を S と H の加算で表す.

$$\log X_n(\omega) = \log S_n(\omega) + \log H(\omega) \quad (2)$$

ここで観測信号 X に対して PCA を適用すると, クリーン音声 S の主なエネルギーは D 個の主な固有値に集中する. それ以外の固有値に対応する主な成分は, 付加成分であると期待できる. ここで, 主な D 個の固有値に対応する固有ベクトルを $V = [v^{(1)}, \dots, v^{(D)}]$ とすると, この V を用いて以下のようなフィルタを考える.

$$\hat{S} = V^t X \quad (3)$$

このフィルタリングによって付加成分 $H(\omega)$ を抑圧することが出来る. また主軸 V の計算をクリーン音声のみから行えば, クリーン音声の構造のみを考慮したフィルタを作成することが出来る. 本稿では, 主軸 V の計算には 2 回目以降の発話音声を用いて, 上記フィルタリングを 1 回目の調音不安定音声に適用する.

3 認識実験

3.1 実験条件

実験用データとして構音障害者, 健常者それぞれ 1 名のデータを収録した. 発話内容として ATR 音素バランス単語 (216 単語) から 210 単語を無作為に選択した. 収録は各単語を 5 回連続発声し, その後, 各発話を手動で切り出した. 収録データのサンプリング周波数は 16 kHz である. 音響モデルは monophone を用い, フレーム周期は 10 ms, ハミング窓長は 25 ms とし, 実験には HTK[4] を用いた.

* A Study on Speech Recognition of a Person with Articulation Disorders. by MATSUMASA Hironori, TAKIGUCHI Tetsuya, ARIKI Yasuo (Kobe Univ.) and LI Ichao (Otemon Univ.) and NAKABAYASHI Toshitaka (Kobe Univ.)

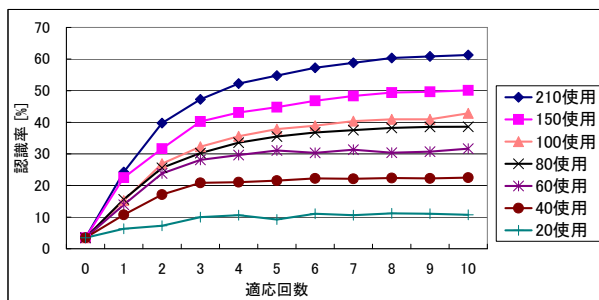


Fig. 2 MLLR+MAP 推定による適応結果 (構音障害者)

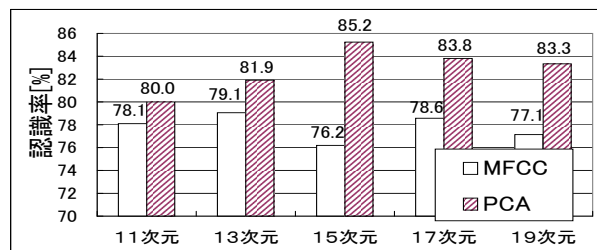


Fig. 3 提案手法による1回目発話の認識率

3.2 汎用モデルの適応実験

実験条件を表1に示す。音響モデルの適応データとして連続発話の2回目発話を用いた。適応法はMLLRとMAP推定を併用して用いた。適応データ数、反復適応回数を変化させた場合の認識結果を図2に示す。適応無しの場合では、健常者では89.6%の精度が得られているが、構音障害者では3.5%しか認識出来ていない。モデル適応を行うことにより、構音障害者において、認識率が61.3% (210単語使用) まで改善するが、精度の向上には大量の適応データが必要である。

Table 1 実験条件 (モデル適応)

音素数	特徴量	混合数
41	12次MFCC+ Δ	16

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。1回目の発話の認識を行う場合は2~5回目の発話を用いて音響モデルを作成した。これを各発話に対して行う。実験条件を表2に示す。

Table 2 実験条件 (構音障害者音響モデル)

音素数	特徴量	混合数
54	12次MFCC+ Δ	6

特定話者モデルに対する、発話回数毎の認識率を表3に示す。

Table 3 発話回数毎の認識率 (%)

1回目	2回目	3回目	4回目	5回目
77.1	89.1	91.4	91.0	87.6

特定話者モデルを用いる事で構音障害者において、平均で87.2%の認識が得られたが、1回目発話の認識率が77.1%と他の発話に比べると著しく低下して

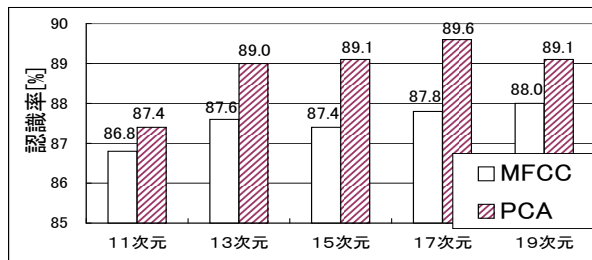


Fig. 4 提案手法による認識実験結果

いる。これは1回目の発話は最初の意図的な動作であり、他の発話よりも緊張状態に陥っていると考えられる。そのためアテトーゼが生じて調音が困難になり、発話スタイルが不安定となることから認識精度が低下したと考えられる。

3.4 提案手法による認識実験

メルフィルタバンク出力24次元に対しPCAを適用した結果を示す。得られた値を基本係数とし、基本係数+ Δ 係数を音声認識の特徴量とした。今回は主成分の個数は11,13,15,17,19個として実験を行った。図3に1回目発話における結果を示し、図4に全発話の平均認識率を示す。

DCTの代わりにPCAを適用することにより、1回目発話において、主成分15個で85.2%まで認識率が改善された。DCTを用いる場合では最も良い場合でも79.1%(13次元)の認識精度なので、DCTよりもPCAの方が1回目発話の調音不安定音声において、より有効的な特徴量抽出法であるといえる。

4 おわりに

構音障害者の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ことから汎用モデルでは認識が困難であるが、構音障害者用の音響モデルを用いることで音声認識技術の適用が可能であることがわかった。1回目の調音不安定発話に対する特徴抽出方法としてMFCCにおけるDCTの代わりに、PCAを用いる手法を検討した。提案手法によって6.1%(79.1%→85.2%)の改善が得られた。今後は、構音障害者特有の特徴量の検討や、様々な音声認識手法の適用を行い認識率の改善に取り組んでいく。また更に対象者を増やしていく予定である。

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] 滝口, 有木, “Kernel PCAを用いた残響下における口バースト特徴量抽出の検討,” 情処論, Vol.47, 1767-1773, 2006.
- [4] S. Young *et al.*, “The HTK Book,” Entropic Labs and Cambridge University, 1995-2002.