社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
TL2006-43,SP2006-131,WIT2006-75(2007-01)

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† 李 義昭†† 中林 稔堯†††

† 神戸大学工学部 〒657-8501 兵庫県神戸市灘区六甲台町 1-1
†† 追手門学院大学経済学部 〒567-8502 大阪府茨木市西安威 2-1-15
††† 神戸大学発達科学部 〒657-8501 兵庫県神戸市鶴甲 3-11
E-mail: †mattu28@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki,nakaba}@kobe-u.ac.jp,
†††chao55@res.otemon.ac.jp

あらまし 本稿ではアテトーゼ型脳性マヒによる構音障害者の音声認識の検討を行う．近年，音声認識技術は向上し汎用性の高いモデル作りも可能となっている．しかし，構音に障害のある人など発話スタイルが健常者と異なる人の音声認識は，健常者用の音響モデルを用いた場合，認識精度が劣化する．そこで，本稿では構音障害者の音響モデルの構築を行い，精度の改善を試みる．アテトーゼ型脳性マヒによる構音障害者は，筋肉の緊張のため1回目の発話が不安定となることがある．従来，対数スペクトルに対し離散コサイン変換を適用した MFCC (Mel Frequency Cepstral Coefficients) を特徴量として用いるが，本稿では離散コサイン変換の代わりに，PCA (Principal Component Analysis) を用いた発話スタイル正規化手法を提案し，その有効性を示す．

キーワード 構音障害，言語障害，脳性マヒ

A Study on Speech Recognition of a Person with Articulation Disorders

Hironori MATSUMASA†, Tetsuya TAKIGUCHI†, Yasuo ARIKI†, Ichao LI††, and Toshitaka NAKABAYASHI†††

† Faculty of Engineering, Kobe University Rokkodaicho 1-1, Nada-ku,Kobe,Hyogo, 657-8501 Japan
†† Faculty of Economics, Otemon Gakuin University Nishiai 2-1-15, Ibaraki,Osaka, 567-8502 Japan
††† Faculty of Human Development, Kobe University Tsurukabuto 3-11, Nada-ku, Kobe,Hyogo, 657-8501 Japan
E-mail: †mattu28@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki,nakaba}@kobe-u.ac.jp,
†††chao55@res.otemon.ac.jp

Abstract We investigate speech recognition of a person with articulation disorders by athetoid type of cerebral palsy. Recently, the accuracy of speaker-independent speech recognition has been remarkably improved by use of stochastic modeling of speech. However, the use of those acoustic models causes degradation of speech recognition for a person with different speech style (e.g., articulation disorders). In this paper, we have tried to build the acoustic model for a person with articulation disorders. The articulation of the first utterance tends to become unstable due to strain of a muscle and that causes degradation of speech recognition, where MFCC (Mel Frequency Cepstral Coefficients) is used as speech features. Therefore we propose a robust feature extraction method based on PCA (Principal Component Analysis) instead of MFCC. Its effectiveness is confirmed by word recognition experiments.

Key words articulation disorders,cerebral paralysis

1. はじめに

情報技術が向上し，近年，福祉分野への情報技術の適用が行われている．例えば，画像認識技術を用いた手話認識 [1] や，文書内の文字の音声化などが行われている [2]．また，音声合成を用いて，発話障害者支援のための音声合成器の作成なども行われている [3]．

音声認識技術は現在，車内でのカーナビの操作，会議においての書き起こしなど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．対象者が子供である場合などには精度が低下することがわかっている [4]．文献 [5] 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．現在，日本だけでも構音障害者も含まれる言語障害者が 3 万 4000 人もいることから十分なニーズがあり，研究の必要性があるといえる [6] [7]．

言語障害の原因の一つとして，脳性マヒが考えられる．脳性マヒとは新生児 1000 人あたりおよそ 2 人の割合で発生する．脳性マヒの定義として厚生省は「受胎から生後 4 週以内の新生児までの間に生じた，脳の非進行性病変に基づく，永続的な，しかし変化しうる運動および姿勢の異常である．その症状は満 2 歳までに発現する.」（1968 年）と定義している．

脳性マヒの原因は，中枢神経系の損傷によるものであり，それに伴う運動障害であると考えられている．発症時期として出生前（胎内感染，母体の中毒，栄養欠損など），分娩時（脳外傷，脳出血，脳の無酸素状態，胎児黄疸，仮死状態，未熟での出産など），出生後（脳内出血，脳炎など）の 3 つの要因が考えられている．

脳性マヒは次のように分類される．1) 痙直型 2) アテトーゼ型 3) 失調型 4) 緊張低下型 5) 固縮型，それぞれの症状が混合して現れる混合型もある．

本稿ではアテトーゼ型の脳性マヒによる構音障害者を対象としている．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，大脳基底核と呼ばれる視床下部，脳幹，小脳と関連を持ち随意運動，姿勢，筋緊張を調節する働きをしている部位に損傷をうけたため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 [8] [9]．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，対数スペクトルに対し離散コサイン変換を適用した 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

音声認識システムにおいて従来は MFCC（Mel Frequency Cepstral Coefficient）が音声特徴量として用いられている．MFCC ではメル尺度フィルタバンクの短時間対数エネルギー出力系列に対して，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．そして音声のスペクトル包絡成分に対応する低次ケプストラムのみを抽出し，特徴量として音声認識に用いられる．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 [10] 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 を行う事によって有効なスペクトル情報の抽出を可能としている．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，調音不安定成分は高次に集まる．この結果 PCA により調音不安定成分の抑圧が行われると期待でき，より有効なスペクトル情報の抽出が可能となる．

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 $S_n(\omega)$ 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．

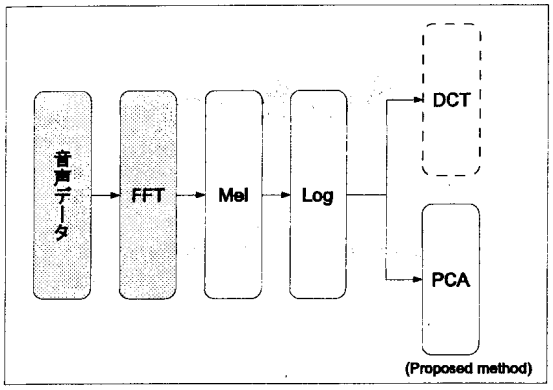

図 1　PCA を用いた特徴量抽出

$$\log X_n(\omega) = \log S_n(\omega) + \log H(\omega) \tag{2}$$

ここで観測信号 X に対して PCA を適用すると，

- クリーン音声の 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 $\boldsymbol{V} = [\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(D)}]$ とすると，この $\boldsymbol{V}$ を用いて以下のようなフィルタを考える．

$$\hat{\boldsymbol{S}} = \boldsymbol{V}\boldsymbol{X} \tag{3}$$

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 $\boldsymbol{V}$ 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 $\boldsymbol{V}$ 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．

2.3 PCA [11]

PCA (Principal Component Analysis) の目的は，データの本質的な構造を残しながら次元数を削減することにある．元の次元空間から削減した次元空間への有効的な選択方法は，分散の最大となる空間を選択することである．ここで，元の d 次元を $\tilde{d}$ 次元に削減すると考えると，元の特徴空間から部分空間への変換行列 $\boldsymbol{A}$ は，

$$\boldsymbol{A} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{\tilde{d}}) \tag{4}$$

で与えられる．$\boldsymbol{u}$ は d 次元からなる正規直交基底とする．特徴ベクトル $\boldsymbol{x}$ は

$$\boldsymbol{y} = \boldsymbol{A}^t\boldsymbol{x} \tag{5}$$

に変換される．また基底の正規直交性から

$$\boldsymbol{A}^t\boldsymbol{A} = \boldsymbol{I} \tag{6}$$

が成立する．ただし $\boldsymbol{I}$ は $\tilde{d}$ 次元単位行列である．このときパターン数を n，原特徴空間でのパターン平均を $\boldsymbol{m}$，部分空間でのパターン平均を $\tilde{\boldsymbol{m}}$ とすると

$$\begin{aligned}\tilde{\boldsymbol{m}} &= \frac{1}{n}\sum_{\boldsymbol{y}\in\mathcal{Y}} \boldsymbol{y} \\ &= \boldsymbol{A}^t\boldsymbol{m}\end{aligned} \tag{7}$$

であるから，部分空間でのパターンの分散 $\tilde{\sigma}^2(\boldsymbol{A})$ は

$$\begin{aligned}\tilde{\sigma}^2(\boldsymbol{A}) &= \frac{1}{n}\sum_{\boldsymbol{y}\in\mathcal{Y}} (\boldsymbol{y}-\tilde{\boldsymbol{m}})^t(\boldsymbol{y}-\tilde{\boldsymbol{m}}) \\ &= \frac{1}{n}\sum_{\boldsymbol{x}\in\mathcal{X}} tr(\boldsymbol{A}^t(\boldsymbol{x}-\boldsymbol{m})(\boldsymbol{A}^t(\boldsymbol{x}-\boldsymbol{m}))^t) \\ &= tr(\boldsymbol{A}^t\frac{1}{n}\sum_{\boldsymbol{x}\in\mathcal{X}}((\boldsymbol{x}-\boldsymbol{m})(\boldsymbol{x}-\boldsymbol{m})^t)\boldsymbol{A}) \\ &= tr(\boldsymbol{A}^t\textstyle\sum\boldsymbol{A})\end{aligned} \tag{8}$$

となる．ここで $\sum$ はパターン集合の原特徴空間における共分散行列 (covariance matrix) を表す．また，$\mathcal{X}$ はパターン $\boldsymbol{x}$ の集合を，$\mathcal{Y}$ は $\boldsymbol{x}$ が式 (5) で変換されたパターン $\boldsymbol{y}$ の集合を表す．

式 (8) より分散を最大にする $\boldsymbol{A}$ を求めることは，式 (6) の制約条件の基で $tr(\boldsymbol{A}^t\sum\boldsymbol{A})$ を最大にする $\boldsymbol{A}$ を求める変分問題になる．$\boldsymbol{\Lambda}$ を $\tilde{d}$ 次元対角行列とすると，

$$\boldsymbol{A}^t\textstyle\sum\boldsymbol{A} = \boldsymbol{\Lambda} \tag{9}$$

となり，$\boldsymbol{A}$ は $\sum$ を対角化する行列である．$\sum$ の d 個の固有値を λ_i ($\lambda_1 \geqq \lambda_2 \geqq \ldots \geqq \lambda_d$) とすれば，式 (8)，(9) より

$$\begin{aligned}max(\tilde{\sigma}^2(\boldsymbol{A})) &= max(tr(\boldsymbol{A}^t\textstyle\sum\boldsymbol{A})) \\ &= max(tr\boldsymbol{\Lambda}) \\ &= \sum_{i=1}^{\tilde{d}}\lambda_i\end{aligned} \tag{10}$$

となり，$\tilde{\sigma}^2(\boldsymbol{A})$ を最大にする変換行列 $\boldsymbol{A}$ は $\sum$ の上位 $\tilde{d}$ 個の固有値 $\lambda_1, \ldots, \lambda_{\tilde{d}}$ に対応する $\tilde{d}$ 個の正規直交ベクトルを列とする行列として求まる．図 2 に PCA の計算手順の概要を示す．

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

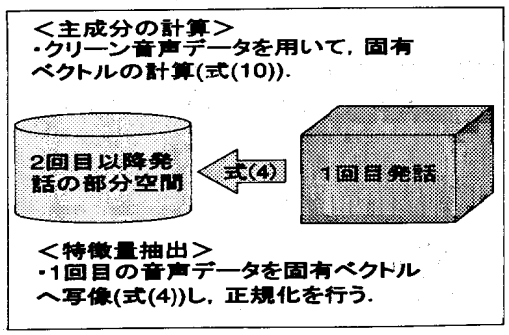

図 2　PCA の計算手順

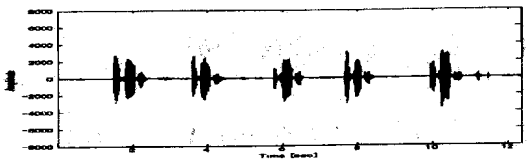

図 3　収録データ

表 1　実験条件（汎用モデル）

サンプリング周波数	16 kHz
ハミング窓長	25 msec
フレーム周期	10 msec
音響モデル	monophone (3 状態 43 音素)
特徴パラメータ	12 次 MFCC+Δ 12 次 MFCC
混合分布数	16
テストデータ	1050 (210 単語× 5 回)
辞書	210 単語

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（図 3），その後，各発話を手動で切り出した．図 4 に構音障害者，図 5 に健常者の音声例を示す．

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．認識実験には「Julian 記述文法音声認識キット v3.1」[12] を使用し，音響モデルは認識キットに含まれている不特定話者音響モデルを用いた．追加の学習はせず，孤立単語での認識を行った．実験条件を表 1 に示し，認識結果を図 6 に示す．健常者では 88.4%の精度が得られるが，構音障害者では 2.4%しか認識できず，従来の汎用モデルでは発話スタイルが健常者と異なるため認識が困難である事がわかる．

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

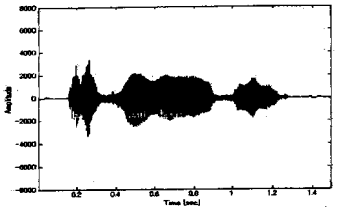

図 4　構音障害者の音声例//a k e g a t a

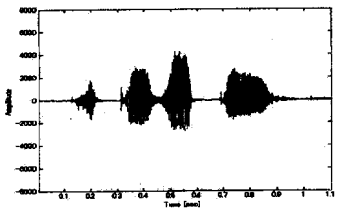

図 5　健常者の音声例//a k e g a t a

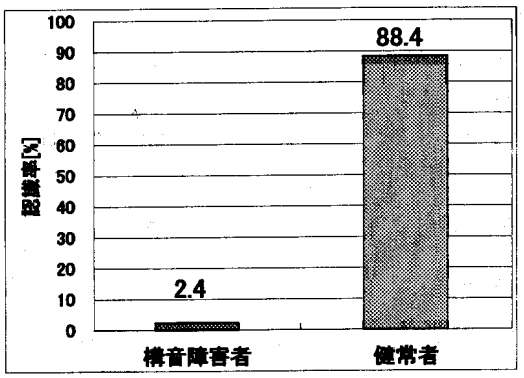

図 6　汎用モデルでの認識実験結果

響モデルを作成した．これを各発話に対して行う．初期モデルの作成，学習，認識には HTK [13] を用いた．実験条件を表 2 に示し認識結果を図 7 に示す．

表 2　実験条件（構音障害者音響モデル）

サンプリング周波数	16 kHz
ハミング窓長	25 msec
フレーム周期	10 msec
音響モデル	monophone (3 状態 54 音素)
特徴パラメータ	12 次 MFCC+Δ 12 次 MFCC (正規化)
混合分布数	6
テストデータ	1050 (210 単語× 5 回)
辞書	210 単語

特定話者モデルを用いる事で構音障害者において，87.2%の認識が得られた．健常者における発話回数ごとの認識率を図 8 に示し，構音障害者における発話回数ごとの認識率を図 9 に示す．構音障害者において，1

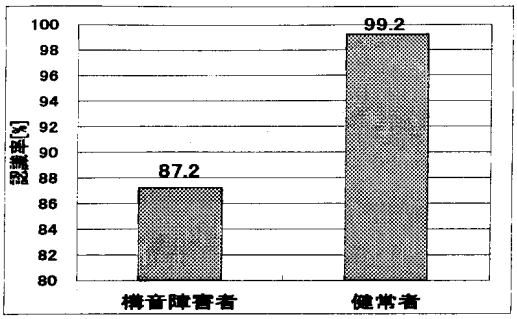

図 7　特定話者モデルでの認識実験結果

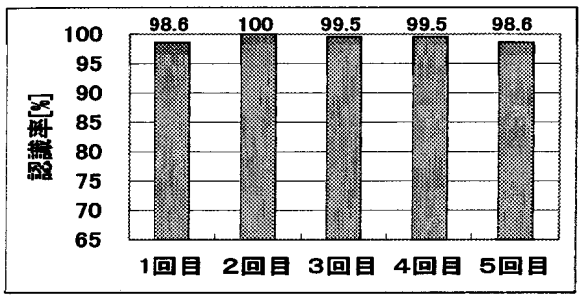

図 8　健常者における発話回数ごとの認識率

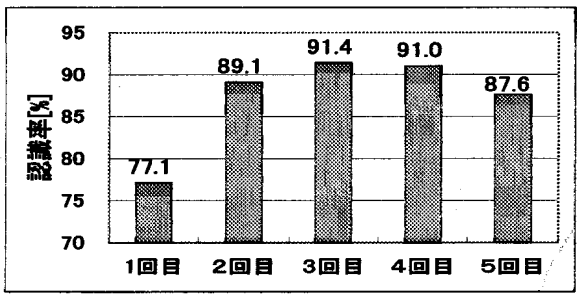

図 9　構音障害者における発話回数ごとの認識率

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 10 に 1 回目発話における結果を示し，図 11 に主成分が 17 個の場合の結果を示し，図 12 に結果を示す．

DCT の代わりに PCA を適用することにより，1 回

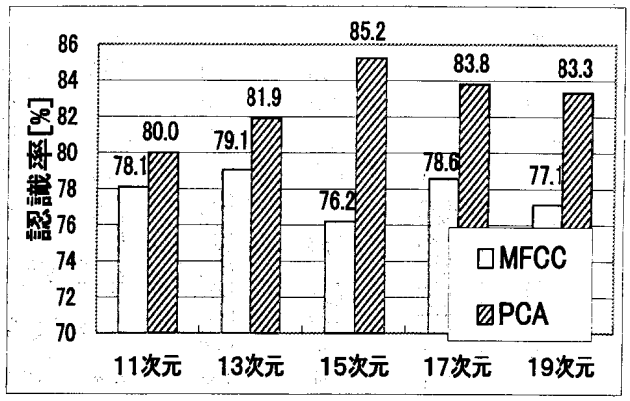

図 10　提案手法による 1 回目発話の認識率

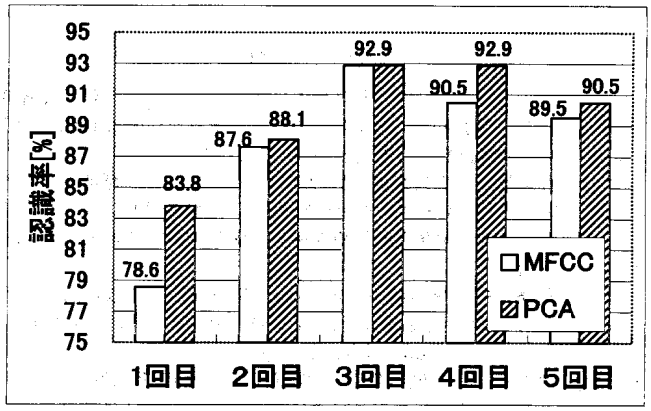

図 11　提案手法による発話回数ごとの認識率 (17 次元)

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．また，全体の認識精度も平均 1.3%改善された．

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%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

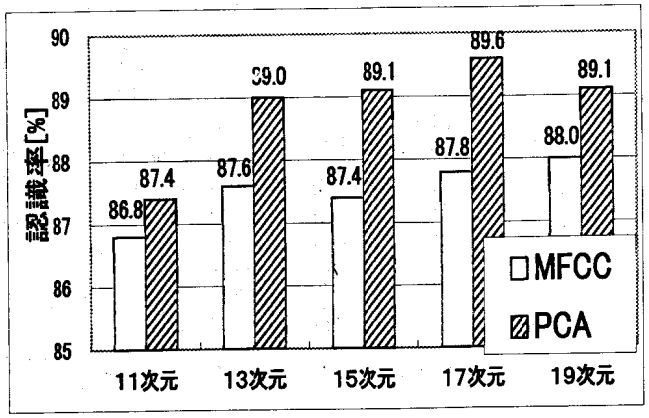

図 12　提案手法による認識実験結果

定である.

文　　献

[1] 佐川浩彦, 酒匂裕, 大平栄二, 崎山朝子, 阿部正博, "圧縮連続 DP 照合を用いた手話認識方式," 電子情報通信学会論文誌, Vol.J77-D2, No.4, pp. 753-763, 1994.

[2] 鈴木悠司, 平岩裕康, 竹内義則, 松本哲也, 工藤博章, 大西昇, "視覚障害者めための環境内の文字情報抽出システム," 電子情報通信学会技術研究報告, WIT2003-314, pp. 13-18, 2003.

[3] 藪謙一郎, 伊福部達, 青村茂, "発話障害者支援のための音声合成器の基礎的設計," 電子情報通信学会技術研究報告, SP2006-321, pp. 59-64, 2006.

[4] 鮫島充, 李晃伸, 猿渡洋, 鹿野清宏, "子供音声認識のための音響モデルの構築および適応手法の評価," 電子情報通信学会技術研究報告, SP2004-114, pp. 109-114, 2004.

[5] 中村圭吾, 田村直良, 鹿野清宏, "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," 日本音響学会講演論文集, 3-7-4, pp. 109–110, 2005.

[6] 内閣府 "平成 18 年　障害者白書," http://www8.cao.go.jp/shougai/

[7] 厚生労働省 "平成 13 年　身体障害児・者実態調査結果," http://www.mhlw.go.jp/houdou/2002/08/h0808-2.html

[8] S.Terry Canale, 落合 直之, 藤井 克之, "キャンベル整形外科手術書 第 4 巻 小児の神経障害/小児の骨折・脱臼," エルゼビア・ジャパン, 2004.

[9] "脳性マヒについて", http://www.geocities.co.jp/SweetHome/6954/nouseimahi.html

[10] 滝口哲也, 有木康雄, "Kernel PCA を用いた残響下におけるロバスト特徴量抽出の検討," 情報処理学会論文誌, Vol.47, No.6, pp. 1767-1773, 2006.

[11] 石井健一郎, 上田修功, 前田英作, 村瀬洋, "わかりやすいパターン認識," オーム社, 1998.

[12] "大語彙連続音声認識システム Julius," http://julius.sourceforge.jp/index.htm

[13] S. Young et. al., "The HTK Book," Entropic Labs and Cambridge University, 1995-2002.