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

信学技報
TECHNICAL REPORT OF IEICE.

構音障害者の音声認識における動的特徴量の考察

宮本　千琴† 　滝口　哲也† 　有木　康雄† 　李　 義昭†† 　中林　稔堯†††

† 神戸大学工学研究科　〒 657–8501 兵庫県神戸市灘区六甲台 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: †miyamoto@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki,nakaba}@kobe-u.ac.jp,
†††chao55@res.otemon.ac.jp

あらまし　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．しかし，言語障害などの障害者を対象としたものは非常に少ない．本稿では，アテトーゼ型脳性麻痺による構音障害者の音声認識の検討を行う．アテトーゼ型の構音障害者の発話スタイルは，筋肉の緊張のため健常者と大きく異なり不安定であるため，特定話者モデルでの音声認識には限界がある．特に構音障害者の動的特徴量（デルタケプストラム）の認識精度は健常者に比べて大きく低下する．これに対し本稿では，動的特徴量の代わりに，デルタケプストラム係数のセグメント特徴量を用いることで構音障害者の音声認識精度の改善を試み，その有効性を示す．
キーワード　構音障害，言語障害，脳性麻痺，動的特徴量

A Study on Dynamic Features for Dysarthric Speech Recognition

Chikoto MIYAMOTO†, Tetsuya TAKIGUCHI†, Yasuo ARIKI†, Ichao LI††, and Toshitaka NAKABAYASHI†††

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

Abstract　Recently, the accuracy of speaker-independent speech recognition has been remarkably improved by use of stochastic modeling of speech. However, there has been very little research on orally-challenged people, such as those with speech impediments. Therefore we have tried to build the acoustic model for a person with articulation disorders. The articulation of the speech tends to become unstable due to strain of a muscle and that causes degradation of speech recognition. In this paper, we focus on the fact that recognition rate of a person with an articulation disorder decreases compared with that of a physically unimpaired person, expecially in speech recognition using dynamic features only. Therefore, we use multiple acoustic frames as an acoustic feature to solve this problem. Its effectiveness is confirmed by word recognition experiments.
Key words　articulation disorders, cerebral paralysis, dynamic features

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]．

音声認識技術は近年，飛躍的に進歩し，様々な環境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．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ている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 4 万 2000 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]．

言語障害の原因の一つとして，脳性麻痺が考えられる．脳性麻痺の定義として，1968 年の厚生労働省脳性麻痺研究班は「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.」としている．

脳性麻痺とは，筋肉の動きをつかさどる脳の部分が受けた損傷が原因で筋肉の制御ができなくなり，けいれんや麻痺，そのほかの神経障害が起こる症状のことである．出生前，出生時，出生直後の脳への酸素供給，出生前の胎内感染，妊娠中毒症，分娩時の外傷，仮死状態，未熟出生，出生後の脳を覆う組織の炎症や外傷性損傷などが原因として考えられる．

脳性麻痺は，脳の損傷部分によって主に痙直型（大脳皮質），アテトーゼ型（中脳もしくは脳基底核），失調型（小脳），混合型（脳の広範囲）に分類される．痙直型は正常な筋の伸張反射が過度になる，アテトーゼ型はアテトーゼと呼ばれる筋肉の不随意運動を伴う，失調型は協調運動の障害，混合型はそれぞれの症状が混合して現れる，というような症状が見られる．

本稿では，アテトーゼ型の脳性麻痺による構音障害者を対象としている．アテトーゼ型は，脳性麻痺患者の約 20% に発生する．筋肉の随意運動や姿勢の調整を行っている大脳基底核（大脳皮質，視床や脳幹を結び付けている神経核の集まり）に損傷を受けたことにより，筋肉が不随に動き，正常に制御できないアテトーゼと呼ばれる症状が見られる．とくに緊張状態にあるときや，意図的動作を行うときに見られる．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 [7] [8]．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．

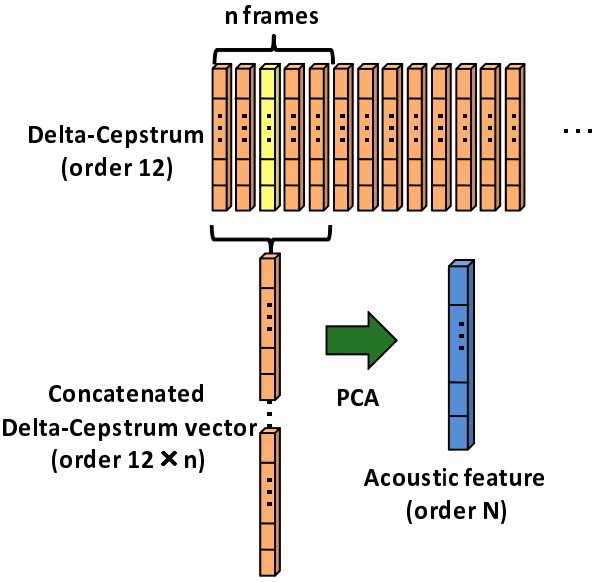

図 1　セグメント特徴量の抽出

アテトーゼ型の構音障害者の発話スタイルは，筋肉の緊張のため健常者と大きく異なり不安定になる場合がある．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 (Mel Frequency Cepstral Coefficient) を特徴量として用いるが，我々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てきた [9]．本稿では，動的特徴量を用いた音声認識において，構音障害者の認識精度が健常者に比べて大きく低下することに着目する．構音障害者において，動的特徴量は時間特徴が十分に表現されていないと言える．これに対し，本稿では動的特徴量の代わりに，デルタケプストラム係数のセグメント特徴量を用いることで構音障害者の音声認識精度の改善を試み，その有効性を示す．

2. セグメント特徴量

音声認識システムにおいて従来は，音声特徴量として MFCC や特徴量の線形回帰係数である ΔMFCC や ΔΔMFCC が広く用いられている．ΔMFCC は以下のように求められる．

$$d_t = \frac{\sum_{\theta=1}^{\Theta} \theta (c_{t+\theta} - c_{t-\theta})}{2 \sum_{\theta=1}^{\Theta} \theta^2} \qquad (1)$$

ここで，c_t は時間 t におけるケプストラム係数，Θ は窓幅を表す．同じように，ΔΔMFCC は ΔMFCC に適用することで求められる．しかし，構音障害者の発話スタイルは健常者と大きく異なり不安定であるため，MFCC を用いた特定話者モデルでの音声認識には限界がある．特に動的特徴量である

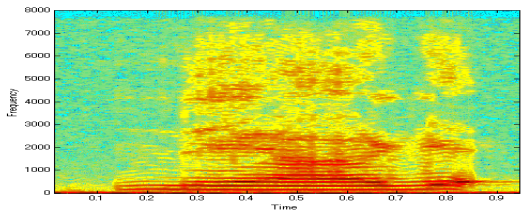

図 2　構音障害者のスペクトログラム例//n e a g e

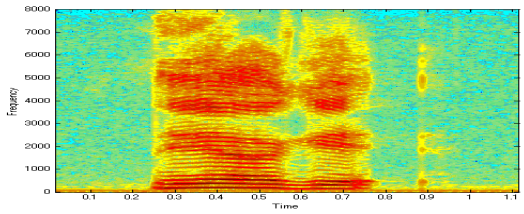

図 3　健常者のスペクトログラム例//n e a g e

ΔMFCC を用いた音声認識において，構音障害者の認識精度は健常者に比べて大きく低下する (図 4)．そこで，ΔMFCC の代わりに，デルタケプストラム係数のセグメント特徴量を音響特徴量として用いる．図 1 にセグメント特徴量抽出の流れを示す．当該フレームとその前後数フレームの計 n フレームを連結させ，PCA により N 次元に圧縮を行ったものを音響特徴量とする．実際には MFCC と組み合わせたものを音響特徴量として用いる．

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 (話者 A)，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．図 2 に構音障害者，図 3 に健常者のスペクトログラム例を示す．構音障害者の場合，子音など高域のパワーが弱く，明瞭度が劣化している．

3.2 特定話者モデルでの認識実験

特徴量として 12 次 MFCC，ΔMFCC，MFCC+ΔMFCC を用いて特定話者モデルを作成し孤立単語認識実験を行った．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．初期モデルの作成，学習，認識には HTK [10] を用いた．認識結果は 5 回発話の認識率の平均値により求める．実験条件を表 1 に示し，認識結果を図 4 に示す．

構音障害者において，ΔMFCC における認識率が 49.2% と健常者に比べて著しく低下している．これは発話という意図的動作時で筋肉の緊張によってアテトーゼが生じて調音が困難になり，明瞭度が劣化したため，ΔMFCC では時間特徴

表 1　実験条件（話者 A）

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）
混合分布数	8
評価データ	1050（210 単語 × 5 回）
辞書	210 単語

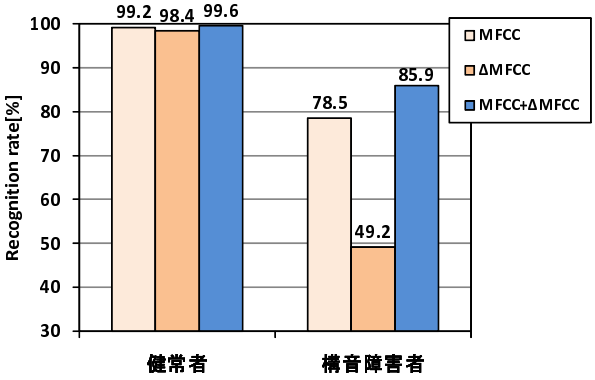

図 4　特定話者モデルでの認識結果

が十分に表現されていないと考えられる．

3.3 セグメント特徴量による認識実験

ΔMFCC の代わりに，デルタケプストラム係数 12 次元に対しセグメント特徴量を求め，これを音響特徴量として用いた結果を示す．今回はフレーム数 n を 3，5，7，9，11，13，15，17 フレームと変化させて実験を行った．ΔMFCC の結果と比較するために，PCA により ΔMFCC と同次元（N=12）に次元圧縮した場合の認識結果を図 5 に示す．また，MFCC と組み合わせた場合の認識結果を図 6 に示す．

ΔMFCC の代わりにセグメント特徴量を用いることにより，セグメント特徴量単体ではフレーム数 n=11 のとき 62.9% まで認識率が改善された．また，MFCC とセグメント特徴量を組み合わせたものと MFCC+ΔMFCC を比較すると，フレーム数 n=13 のとき 90.3% まで認識率が改善された．ΔMFCC と同次元に圧縮した場合でもフレーム数の値によらず ΔMFCC，MFCC+ΔMFCC よりも高い認識率が得られた．これらの結果より，セグメント特徴量は ΔMFCC に比べ，認識率の改善に貢献していることがわかる．

次に，セグメント特徴量を PCA により次元圧縮する次元数を変化させて実験を行った．フレーム数 n=13 の場合の次元数による認識率の変化を図 7 に示す．また，MFCC と組み合わせた場合の認識結果を図 8 に示す．ここで，セグメント特徴量の次元数が 36 のとき 70% となり，ΔMFCC より高い認識率が得られた．しかし，MFCC と組み合わせた結果から，次元数を増やしていくと認識率が低下していくことがわかる．これは，特徴ベクトルの次元数に差ができ，確率

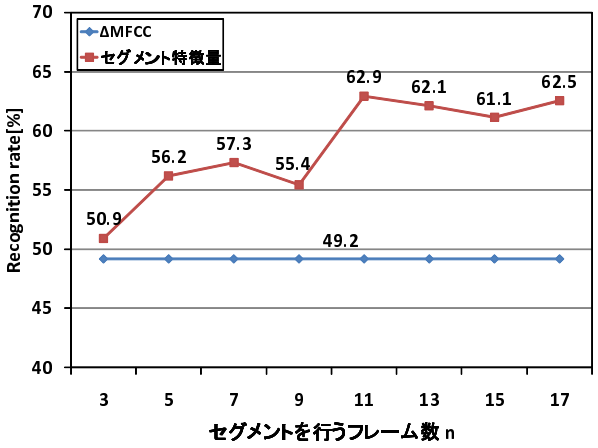

図 5　セグメントを行うフレーム数による認識率の変化 (12 次元): 話者 A

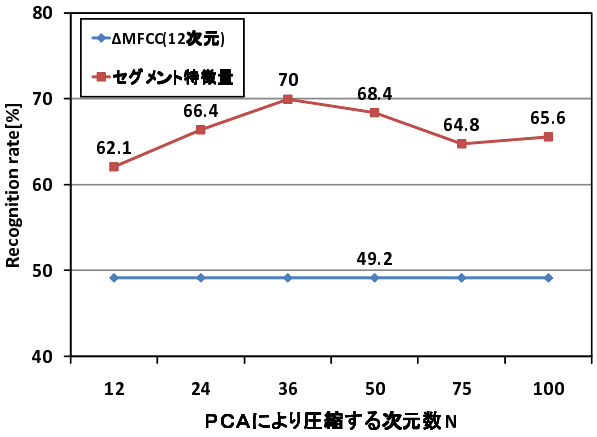

図 7　PCA の次元数による認識率の変化 (13 フレーム):話者 A

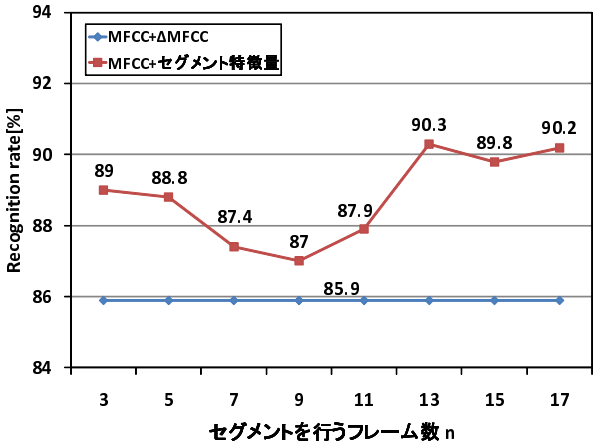

図 6　MFCC と組み合わせた場合の認識結果 (12 次元):話者 A

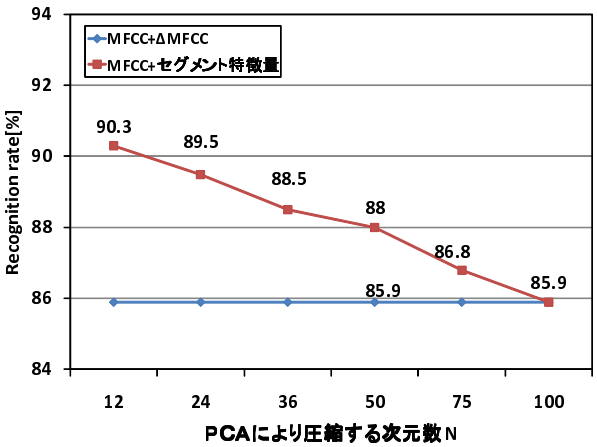

図 8　MFCC と組み合わせた場合の認識結果 (13 フレーム):話者 A

推定の際に MFCC の情報が反映されにくくなることが考えられる.

3.4　別の構音障害者音声を用いた認識実験

セグメント特徴量の有効性を確認するために，別の構音障害者（話者 B) 音声を用いて認識実験を行った.

3.4.1　話者 B での認識実験

話者 B は発話内容として，ATR 音素バランス単語 216 単語と ATR 音声データベース 2620 単語を用い，評価データは 1080 発話（216 単語 × 5 回発話），学習データは 5240 発話（2620 単語 × 2 回発話）を使用した．実験条件を表 2 に示し，フレーム数 n を 3，5，7，9，13 と変化させ，PCA により ΔMFCC と同次元（N=12）に次元圧縮した場合の認識結果を図 9 に示す．また，MFCC と組み合わせた場合の認識結果を図 10 に示す.

話者 B においても ΔMFCC の代わりにセグメント特徴量を用いることにより，セグメント特徴量単体ではフレーム数 n=3 のとき 58.9% まで大きく認識率が改善された.

表 2　実験条件（話者 B）

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）
混合分布数	4
評価データ	1080
辞書	216 単語

また，MFCC とセグメント特徴量を組み合わせたものと MFCC+ΔMFCC を比較すると，フレーム数 n=3 のとき 68.8% まで認識率が改善された.

次に，セグメント特徴量を PCA により次元圧縮する次元数を変化させて実験を行った．フレーム数 n=3 の場合の次元数による認識率の変化を図 11 に示す．また，MFCC と組み合わせた場合の認識結果を図 12 に示す．ここで，セグメント特徴量の次元数が 36 のとき 65.7% となり，ΔMFCC より高い認識率が得られた．また，MFCC とセグメント特徴量

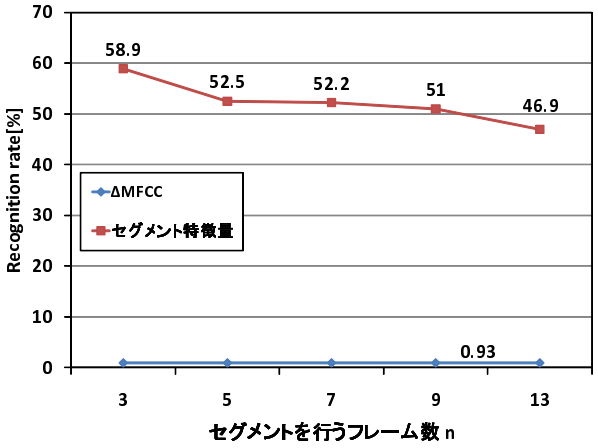

図 9　セグメントを行うフレーム数による認識率の変化 (12 次元): 話者 B

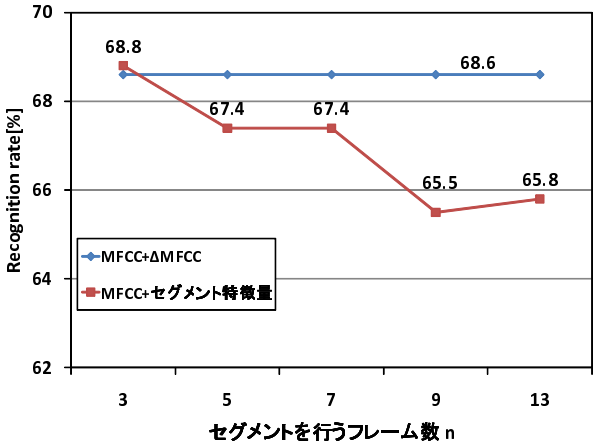

図 10　MFCC と組み合わせた場合の認識結果 (12 次元):話者 B

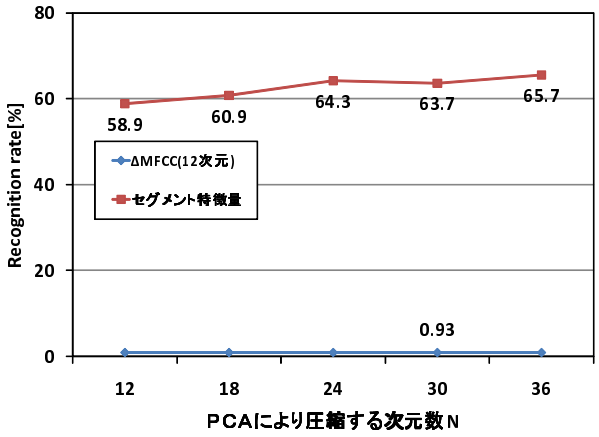

図 11　PCA の次元数による認識率の変化 (3 フレーム):話者 B

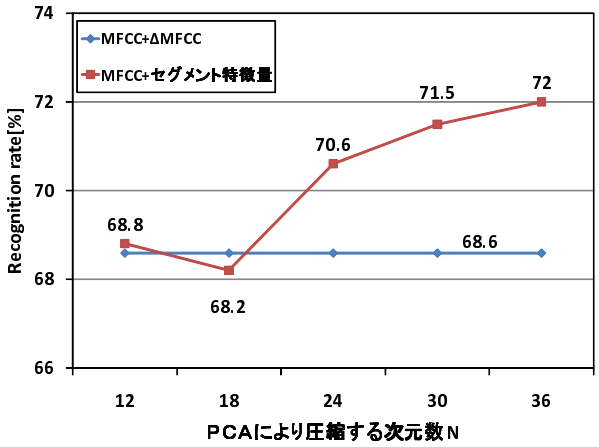

図 12　MFCC と組み合わせた場合の認識結果 (3 フレーム):話者 B

を組み合わせたものと MFCC+ΔMFCC を比較すると，次元数が 36 のとき 72.0% まで認識率が改善された．これらの結果より，セグメント特徴量は動的特徴量として ΔMFCC よりも認識に貢献していることがわかる．

4. おわりに

本稿では，構音障害者の音声認識の検討を行った．動的特徴量である ΔMFCC を用いた音声認識において，構音障害者の認識精度は健常者に比べて大きく低下することに着目し，ΔMFCC の代わりにセグメント特徴量を用いることで音声認識精度の改善を試みた．セグメント特徴量を用いることで話者 A において 20.8% の改善が得られた．また，MFCC とセグメント特徴量を組み合わせることで，ΔMFCC に比べ，セグメント特徴量が認識率の改善に貢献していることがわかった．さらに別の構音障害者音声でも有効性を確認することができた．今後は，構音障害者特有の特徴量の検討や，音質改善の試み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有効性を確認していく予定である．

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，“視覚障害者の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 20 年版障害者白書，” http://www8.cao.go.jp/shougai/

[7] 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.

[8] Mark H. Beers, 福島雅典, “メルクマニュアル医学百科 最新家庭版,” 日経 BP 社, 2004.
[9] H. Matsumasa, T. Takiguchi, Y. Ariki, I. LI and T. Nakabayashi, ”PCA-Based Feature Extraction for Fluctuation in Speaking Style of Articulation Disorders,” INTERSPEECH-2007, pp. 1150–1153, 2007.
[10] “HTK (Hidden Markov Model Toolkit),” http://htk.eng.cam.ac.uk/.