

情報家電操作における脳性麻痺構音障害者の音声認識評価

松政 宏典[†] 田中 克幸[†] 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,katsutanaka}@me.cs.scitec.kobe-u.ac.jp, ^{††}{takigu,ariki,nakaba}@kobe-u.ac.jp,
^{†††}chao55@res.otemon.ac.jp

あらまし 重度障害者の自立支援機器として、環境制御装置は重要な役割を担っている。環境制御装置の中には、音声認識を用いて操作を行うものもある。手足が不自由な重度障害者にとって、音声認識は非常に重要である。しかし、従来のものでは発話に障害をもたない人の音声を対象としているため、構音障害者などの発話スタイルが異なった方では従来の音響モデルでは認識が困難である。そこで本稿では、構音障害者の音声認識が可能な環境制御装置の実現に向けて、構音障害者の音響モデルの構築を行い、家電操作タスクに対する音声認識評価を行った。音声認識評価は2つの異なる環境（既知の環境、未知の環境）で行った。未知の環境での認識実験において、単語認識精度は56.4%であった。脳性麻痺障害者の発話スタイルが環境によって変動しやすいことから、発話スタイルの変動に対して音響モデル適応を行うことにより、83.9%まで精度が改善された。

キーワード 構音障害, 脳性麻痺, 音声認識, 環境制御装置

Evaluation of Speech Recognition by a Person with Articulation Disorder in Operation for Home Information Appliances

Hironori MATSUMASA[†], Katsuyuki TANAKA[†], Tetsuya TAKIGUCHI[†], Yasuo ARIKI[†], Ichao LI^{††}, and Toshitaka NAKABAYASHI^{†††}

[†] Graduate School of Engineering, Kobe University Rokkodaicho 1-1, Nada-ku, Kobe, Hyogo, 657-8501 Japan

^{††} Faculty of Economics, Otomon 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,katsutanaka}@me.cs.scitec.kobe-u.ac.jp, ^{††}{takigu,ariki,nakaba}@kobe-u.ac.jp, ^{†††}chao55@res.otemon.ac.jp

Abstract ECU (environmental control unit) takes an important role in self-support for developmentally-disabled individual. Some ECU use speech recognition for operation. It is important for a crippled-disabled person to use speech recognition. However, it is very difficult for a person with articulation disorder to use the ECU with speech interface since conventional ECU targets non-articulation disorder as user. Our subject is to build the ECU which can recognize articulation disorder's speech. In this paper, we built an acoustic model for a person with articulation disorder and evaluated speech recognition in the task of operation for home information appliances. We used two different speech data as test data. One is recorded under a familiar environment. The other is recorded under an unfamiliar environment. In the experiment for speech data which recorded under the unfamiliar environment, the word-recognition rate of 56.4% was obtained. The speaking style of a person with cerebral palsy tends to fluctuate by the change of environments. Therefore, we carried out speaker adaptation in the test environment, and the word-recognition rate improved to 83.9%.

Key words articulation disorder, cerebral palsy, speech recognition, environmental control unit

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] [4]。

重度障害者の自立支援機器として、環境制御装置は重要な役割を担っている。これまでに、幾つかの環境制御装置が開発、実用化されている。従来の環境制御装置では操作方法として、押しボタンスイッチを用いたものや、呼吸を用いたもの、音声認識を用いて操作を行うものもある。手足が不自由な重度障害者において音声認識は非常に重要である。しかし、従来のものでは発話に障害がない人（成人男女）の音声を対象としているため、構音障害者などの発話スタイルが異なった方では従来の音響モデルでは認識が困難である。

音声認識技術は現在、車内でのカーナビの操作、会議における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。しかし、現在の音声認識システムでは、例えば、対象者が子供である場合などには、発話スタイルが大きく異なるため、認識精度が低下することがわかっている [5]。

我々は、これまでに構音障害者の音声認識の実現に取り組んできた [6]。その他、文献 [7] 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が行われている。また、文献 [8] では、重度障害者の自立移動を支援するために構音障害者の音声認識を用いた電動車イスの開発に取り組まれている。しかし、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めた言語障害者が 3 万 4000 人もいることから十分なニーズがあり、さらなる研究の必要性がある [9] [10]。

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。脳性麻痺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。

脳性麻痺の原因は、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。

脳性麻痺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麻痺による構音障害者を対象とした。アテトーゼ型とは脳性麻痺患者の約 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 [11] [12]。

従来の音声認識を用いた環境制御装置では、健常者の音声を用いて音響モデルを作成している。構音障害者の場合には、発話スタイルが健常者と大きく異なるため、認識が困難である。そこで、本稿では構音障害者の音声認識が可能な環境制御装置の実現に向けて、構音障害者の音響モデルを作成した。環境制御装置「みてら」[13] を用いて音声認識評価を行い、日々の発話変動に対する改善として音響モデルの適応を行った。

2. 音声認識システム

従来の健常者の音声を用いて作成した音響モデルでは、発話障害者と健常者との発話スタイルが大きく異なるため、発話障害者の音声認識は非常に困難である。しかし、構音障害者の音声を用いて作成した音響モデルを用いることによって、認識精度の大幅な改善が得られる [6]。

そこで、本稿では構音障害者の音響モデルを事前に作成し、音声認識に用いた。特徴量抽出、および音声認識には julian [14] を用いた。図 1 にシステム全体の流れを示し、図 2 に使用した評価装置を示す。実際の操作方は、利用者（重度障害者）は音声認識部（PC）の支持画面に従って（図 3）、操作コマンドを発話（入力）する。操作可能なコマンドは環境制御装置の操作画面の状態に従う。認識結果によっては、状態遷移が生じる。図 4 に状態遷移例を示す。認識に関しては、各状態ごとに認識文法を変更し、シーン内のコマンドのみを認識結果として出力する。

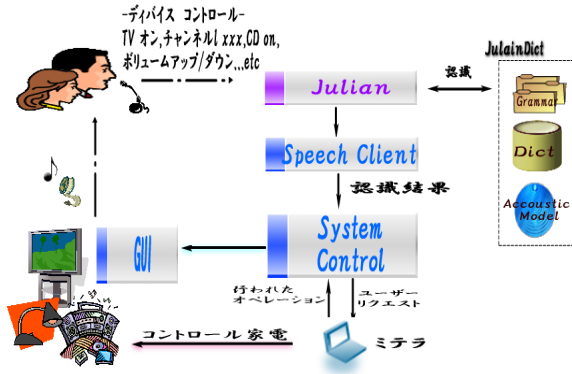


図 1 システム全体の流れ

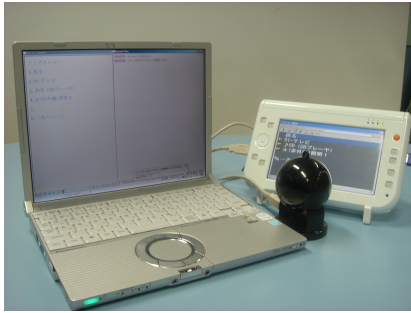


図 2 実際の評価装置

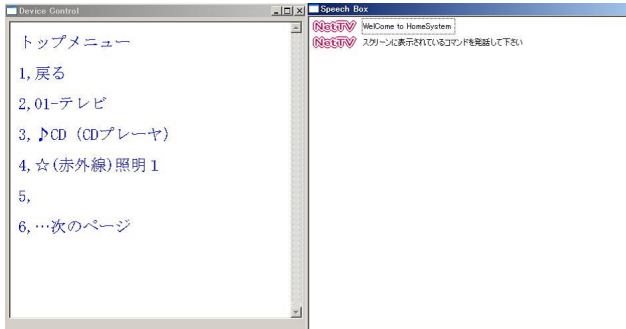


図 3 認識部の指示画面
左画面：操作可能コマンド 右画面：指示画面

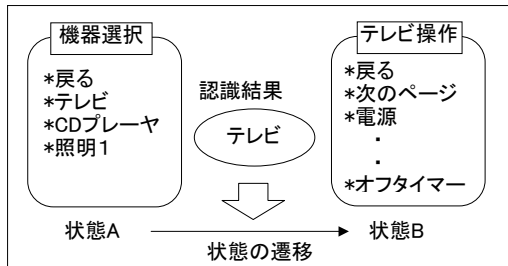


図 4 状態の遷移例

3. 発話スタイル変動への適応 [15]

同一話者の発話であっても、日々の体調や環境によっ

て発話スタイルが変動し、音声認識に影響を与える場合がある。脳性麻痺による構音障害者の場合は、健常者に比べ発話スタイルの変動が大きく、環境による影響も大きいと考えられる。しかしながら、様々な環境下において大量の音声データを収録するのは、障害者にとっては大きな負担となるので、そのような収録作業は困難である。そこで、本稿では話者適応に用いられる MLLR (Maximum Likelihood Linear Regression) 法、および MAP (Maximum A Posteriori) 法を HTK [16] を用いて、構音障害者音響モデルの平均ベクトルを、テスト時の音声に対して適応を行い、スタイル変動に対処する。

MLLR 法では、複数の分布をひとまとめにして、クラスとして適応を行う。推定される n 次元の平均ベクトル $\hat{\mu}$ は次のように示される。

$$\hat{\mu} = W\xi \quad (1)$$

W は $n \times (n+1)$ 次元の変換行列、 ξ は拡張された平均ベクトルであり、次のように表すことができる。

$$\xi = [\omega \ \mu_1 \ \mu_2 \ \dots \ \mu_n]^T \quad (2)$$

ここで、 T は転置を意味し、 ω はバイアスのオフセットである。変換行列 W は次のように表される。

$$W = [b \ A] \quad (3)$$

A は $n \times n$ 次元の変換行列であり、 b はバイアスベクトルである。 W は適応データに対して、尤度が最大となるように求める。適応データを θ とすると、 W は次のように求められる。

$$W = \arg \max_{\bar{W}} [\theta \mid \bar{W}] \quad (4)$$

MAP 法では、あらかじめ作成された音響モデルを事前に得られた情報とし、適応データを用いて事後確率の最大化を行う。推定される n 次元の平均ベクトル $\hat{\mu}$ は次のように示される [16]。

$$\hat{\mu}_{jm} = \frac{N_{jm}}{N_{jm} + \tau} \bar{\mu}_{jm} + \frac{\tau}{N_{jm} + \tau} \mu_{jm} \quad (5)$$

ここで、 j は HMM の状態数を、 m は混合分布要素を示す。 τ は基の音響モデルから適応先のモデルへの

事前の情報に対する重みであり、 N は適応データの尤度である。 $\bar{\mu}$, μ は、それぞれ適応データの平均ベクトル、基の音響モデルの平均ベクトルを示す。

式 (1) で求めた平均ベクトルを、式 (5) の μ として、2つの手法を併用することでより精度の高い適応が可能となる。

4. 認識実験

4.1 実験条件

今回の認識実験では表 1 に示した 7 つのシーン（機器選択、テレビ操作、CD 操作、照明操作、チャンネル操作、番組表操作、メニュー操作）、41 個のコマンドからなるタスクに対して、構音障害者 1 名での認識実験を行った。テストデータは 2 つの環境での収録を行った。

- テストデータ 1
 - 被験者の自宅での収録（既知の環境）
 - 459 発話（各コマンド 8 回から 15 回）
- テストデータ 2
 - 神戸大学実験室での収録（未知の環境）
 - 125 発話（各コマンド 0 回から 37 回）

本人の自宅は慣れた環境であるが、実験室は初めての環境であり、通常よりも高い緊張状態であると考えられる。認識実験は、各シーン毎に行った。機器選択シーンの「戻る」と TV 操作シーンの「戻る」のように、各シーンで共通のコマンドに関しては、一つの発話を重複して使用した。

4.2 健常者モデルの認識実験

初めに、健常者の音声を用いて作成した不特定話者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」[14] に含まれている音響モデルを使用した。実験条件を表 2 に示し、実験結果を図 5 に示す。平均認識率は 28.8% と低く、従来の環境制御装置の操作は困難であることがわかる。

4.3 構音障害者音響モデルでの認識実験

健常者の音声を用いた音響モデルでは、認識が困難なため、構音障害者の音響モデルの作成を行う。モデル作成には ATR 音素バランス単語（216 単語）に含まれている 210 単語を 5 回発話した 1050 発話の音声データを用いた。収録環境はテストデータ 1 と同様である。ただし収録時期は数ヶ月の時間差がある。作成条件を表 3 に示し、テストデータ 1 の実験結果を図 6 に示す。構音障害者の音声を用いた音響モデルを用いることで、平均認識率が 28.8% から 76.5% まで改善し

表 1 タスク内容

階層 1	コマンド名	階層 2	コマンド名	階層 3	コマンド名
機器選択	戻る テレビ CD プレーヤ 照明 1	テレビ操作	戻る	チャンネル操作	戻る
			次のページ		次のページ
			電源		NHK
		CD 操作	チャンネル操作	番組表操作	サンテレビ
			チャンネルアップ		毎日放送
			チャンネルダウン		テレビ大阪
		照明操作	音量アップ	メニュー操作	ABC テレビ
			音量ダウン		関西テレビ
			番組表操作		読売テレビ
		照明操作	メニュー操作		NHK 教育
			消音		戻る
			音声切替		番組表
		照明操作	オフタイマー		上へ移動
			戻る		下へ移動
			電源		右へ移動
		照明操作	電源オフ		左へ移動
			再生する		決定
			一時停止	メニュー操作	戻る
		照明操作	停止		メニューの表示
			次の曲		上へ移動
			前の曲		下へ移動
		照明操作	戻る		右へ移動
			照明をつける		左へ移動
			照明を消す		決定

表 2 実験条件（健常者モデル）

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 + Δ power
混合分布数	16
テストデータ	633 発話

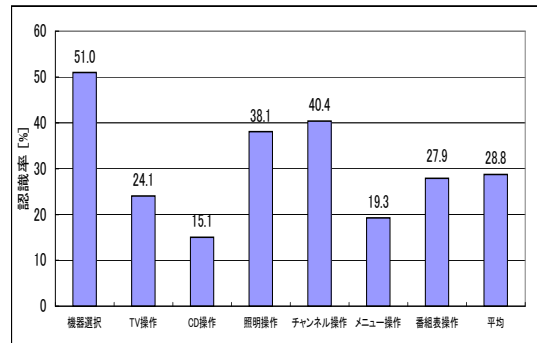


図 5 実験結果（健常者モデル）

た。評価データと学習データの収録には時期差があり、かつ構音障害者の場合には、日々の体調や環境によって発話スタイルの変動が大きいと考えられる。次に、3章で述べた MLLR 法と MAP 法を用いて音響モデルをテストデータに適応させた結果を図 7 に示す。適応データ数は 2, 5, 10 発話と変化させて実験を行った。10 発話を用いて適応を行う事により 84.5% まで認識率が改善された。

表 3 実験条件（構音障害者モデル）

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 + Δ power
混合分布数	6
テストデータ	633 発話

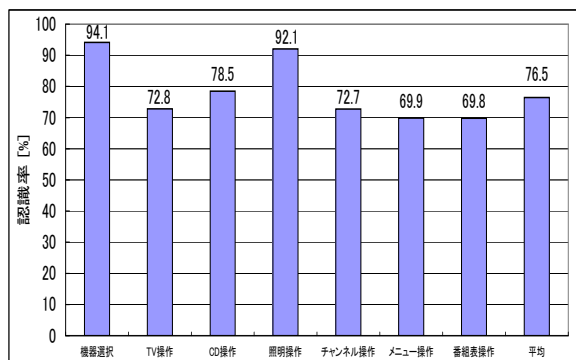


図 6 実験結果（構音障害者モデル）

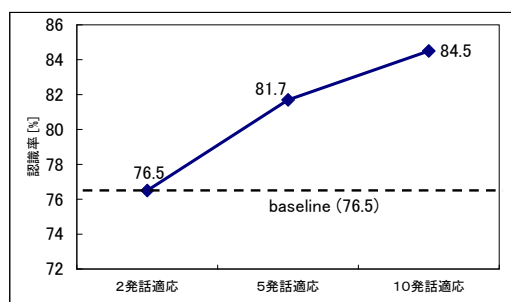


図 7 適応による認識率の変化

4.4 単語モデルでの認識

構音障害者の発話を健常者の発話と比較すると、発話音素が幾つか脱落したり、またその境界が非常にあいまいになる場合がある。そこで、音素 HMM の代わりに、単語モデルを構築して音声認識実験を行った。単語モデルは表 4 の様にコマンド間で重複するものには同一のモデルを作成した。単語モデルの作成・評価の際には、テストデータ（459 発話）、および今回は使用しなかったコマンド音声（39 発話）を均等に 3 等分し、3 分の 2 を用いてモデルの作成を行い、テストデータの残り 3 分の 1 で評価を行った。認識結果を図 8 に示す。クローズな環境に近いが、同一条件での音素モデルの結果よりも良い認識結果が得られた。

4.5 未知環境下での認識実験

未知環境での評価実験として、神戸大学の実験室にて音声認識実験を行った（テストデータ 2）。被験者にとって、実験室は初めての環境であり、通常よりも高

表 4 モデル作成例

コマンド名	モデル名
チャンネルアップ	ch + up
チャンネルダウン	ch + down
音量アップ	volume + up

表 5 実験条件（単語モデル）

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	準コマンド (3 状態 45 コマンド)
混合分布数	4
テストデータ	633 発話

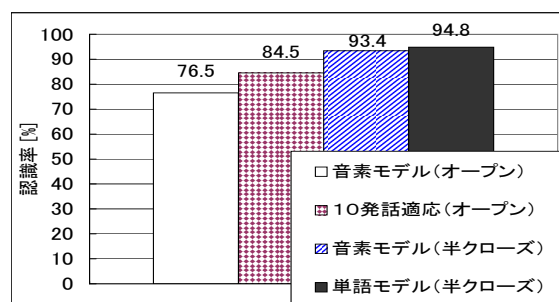


図 8 モデル毎の認識率の比較

表 6 実験条件

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	準コマンド (3 状態 45 コマンド)
混合分布数	4
テストデータ	298 発話

い緊張状態にあると考えられる。認識実験では、「番組表操作」「メニュー操作」を除いた 5 つのシーンで行い、各コマンドの発話回数は 0 回から 37 回と偏りがある。事前に作成した音響モデルを用いたところ、平均で 56.4% の精度しか得られず、テストデータ 1 を用いた認識結果よりも著しく精度が低下した。精度の低下の原因として、未知環境での操作であることから、日々の体調の変化に加え、高い緊張状態にあったと考えられる。そこで改善方法として、当日の 10 発話を用い、MLLR 法と MAP 法を用いて適応を行ったところ、平均で 83.9% まで認識精度が改善された。実験結果を図 9 に示す。

5. おわりに

構音障害者の音声認識が可能な環境制御装置の実現に向けて、構音障害者の音響モデルを作成し、音声認識評価を行った。従来の健常者音声を用いた音響モデルの場合では、平均で 28.8% しか認識が出来ず、操作

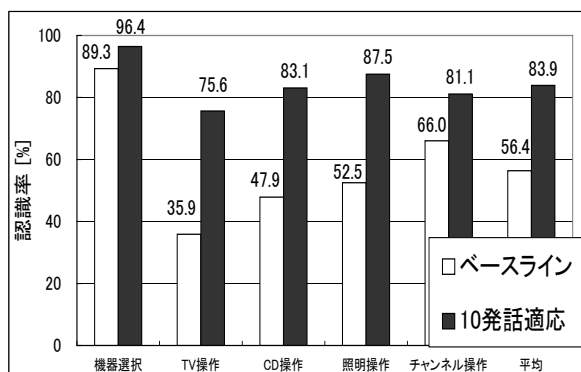


図 9 認識結果 (テストデータ 2)

が困難であった。しかし、構音障害者の音声を用いて、音響モデルを作成したところ、94.8%まで認識精度が改善した。未知環境における認識実験では 56.4%まで認識精度が低下したが、発話スタイル適応を行うことで 83.9%まで改善された。

今回は、音声のみの評価であったが、音声インターフェイスにおける非コマンド発話への対応も必要である。非コマンド発話への対処方法に関しては、タスクベリフィケーションを行い、コマンドと非コマンドの区別を行うことが考えられる [17]。今後は、そのような課題への取り組みなども行っていきたい。

謝辞

本研究の一部は、JST の H18 年度シーズ発掘試験の助成により行われた。

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] K. Nakamura and T. Toda and H. Saruwatari and K. Shikano, “Speaking Aid System for Total Laryngectomees Using Voice Conversion of Body Transmitted Artificial Speech,” INTERSPEECH-2006, pp. 1395-1398, 2006.
- [5] 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.
- [6] 松政宏典, 滝口哲也, 有木康雄, 李義昭, 中林稔亮, “構音障害者の音声認識の検討,” 日本音響学会講演論文集 (春), 1-Q-30, pp. 323-324, 2007.
- [7] 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

- (秋), 3-7-4, pp. 109-110, 2005.
- [8] 児島宏明, 佐宗晃, 石洙永, 李時旭, 佐土原健, “重度障害者のための音声認識電動車いすの開発,” 日本音響学会講演論文集 (春), 2-8-18, pp. 371-372, 2007.
- [9] 内閣府 “平成 18 年 障害者白書,” <http://www8.cao.go.jp/shougai/>
- [10] 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>
- [11] 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.
- [12] “脳性麻痺について,” <http://www.geocities.co.jp/SweetHome/6954/nouseimahi.html>
- [13] 三菱電機コントロールソフトウェア株式会社 “環境制御装置 みてら,” <http://www.mcr.co.jp/product/fukushi/mitera/mitera.htm>
- [14] “大語彙連続音声認識システム Julius,” <http://julius.sourceforge.jp/index.htm>
- [15] 篠田浩一, “確率モデルによる音声認識のための話者適応化技術,” 電子情報通信学会論文誌 D-II vol.J87-D2, No.2, pp. 371-386, 2004.
- [16] S. Young et. al., “The HTK Book,” Entropic Labs and Cambridge University, 1995-2002.
- [17] 佐古淳, 滝口哲也, 有木康雄, “AdaBoost を用いたシステムへの問い合わせと雑談の判別,” 電子情報通信学会技術研究報告, SP2006-443, pp. 19-24, 2006.