

特定話者用音声認識ソフト開発と アルキメデスプロジェクトの 連携に関する研究

李 義 昭

キーワード：音声認識 アルキメデスプロジェクト 社会的不利

はじめに

私たちは、新たな技術進歩が、新たなバリアーを作る可能性があることを経験的に知っている。たとえば、パソコンや携帯電話の普及・発達が、それらにアクセスできない、それらが利用できない人々を作ってきたようにである。PC や通信機器の発達は、コミュニケーションハンディキャップあるいはデジタルデバイドを生じさせるのである。ニール・スコットらが展開するアルキメデスプロジェクトは、現在ハワイ大学において、身体障害者だけではなく、高齢者や学習に困難がある者の、PC や IT 機器へのアクセスの保証を行おうとしている。それは、PC や IT 機器の先にある情報や知識へのアクセスを保証するものである。さらには、それら情報や知識を共有し、学校教育や社会教育に貢献しようとするものである。われわれが行う、音声認識の技術研究も、発話が困難であったり、発話に障害があったりする人々へ、音声認識技術という、われわれの財産へのアクセスを保証しようとするものである。ここでは、アルキメデスハワイプロジェクトの概要、その進行状況を紹介する。そして、われわれが行ってい

る、音声認識の技術研究の概要を紹介し、その可能性について触れる。さらに、これらのプロジェクトと技術研究の連携について考える。

I アルキメデスハワイプロジェクト¹⁾

アルキメデスプロジェクト (The Archimedes Project)²⁾は、1992年にスタンフォード大学で創設され、現在はハワイ大学において、すべての人々がグローバルな情報社会の中で個々の能力に関係なく、そのニーズによって社会に全面的に参加できるように研究活動を行っている。その技術は TAS (Total Access System)³⁾を基礎とし、iTASK (Intelligent Total Access System)⁴⁾に発展しており、独自の製品として「Ho'alauna」⁵⁾を開発している。

1 プロジェクトの目標

今日、多くの人々がデジタル格差社会の中にあり、PC および IT 機器の利用から遠ざけられている。これは重大な社会的不利の存在であり、アルキメデスプロジェクトは、これら不利を除去し、彼らが情報にアクセスし、仕事や生産、そして生活に役立て、その能力を発揮できるようにする。個々のニーズや能力、選択や文化に応じて IT 機器が利用できるよう、研究と教育を行い、その成果を人々の生活の改善に生かしていくことに専念する。

アルキメデスハワイプロジェクトはスタンフォード大学 (Archimedes at Stanford) における成果である、TAS (Total Access System)、それを基盤とした iTASK (Intelligent Total Access System) の技術を障害や高齢化、教育と読

1) <http://archimedes.hawaii.edu/>

2) Neil Scott, (2003) "Archimedes at Stanford"

3) http://archimedes.hawaii.edu/Total_Access_System.htm

4) Neil Scott, (2003) "Archimedes at Stanford", pp 35-38

5) Neil Scott and Sandy Gabrielli, (2004) "Overview of the Ho'alauna Tablet"

み書きの能力などに関連した、アクセスのしやすさやコミュニケーションの問題を解決するために生かしていく。また、TAS と iTAS の技術を用いて、新興国で PC が高価なために情報技術が利用できないもしくは利用に不利が起こっているという問題を克服し、その利点を提供する。

2 プロジェクトの原則

- ① 障害があるとかないとかに関わりなく、全ての人々が情報を効果的に利用するための援助を必要としている。
- ② 情報は本来アクセスが難しかったり、易しかったりするものではなく、創造されたままにたどりつけるものである。
- ③ 障害者になるということが必ずしも社会的不利を負うということではなく、社会的不利は、しばしば、取り除かれるのである。
- ④ 社会的不利は、しばしば、標準的構成で個人だけのためのツールを作るとする決定によって引き起こされる。
- ⑤ 設計されたアクセスは、むしろ、改造されたアクセスより望ましい。
- ⑥ 一般的なアクセスを提供することが全ての人々の利益となる。

3 プロジェクトの歴史

アルキメデスプロジェクトは 1992 年にスタンフォード大学の言語情報研究センター (CSLI: the Center for the Study of Language and Information) の約 30 の研究プログラムの一つとして始まった。その焦点は、身体障害者が完全に IT 装置にアクセスできるよう、その解決策を提供することにあった。研究の成果と評価は、一般的なアクセスをあらゆる IT 装置に提供できる TAS の進展に現れた。TAS は MAC (Macintosh), IBM (International Business Machines Corporation), SUN (Sun Microsystems), HP (Hewlett-Packard) などの PC のために開発されたワークステーションで、2002 年にアメリカにおける特許を取得している。1996 年、TAS が認可されて以降、ユーザ

ーにもっとも適した IT 装置へのアクセスの道具である，特殊スイッチ，スキャニング，モールス信号，手書きパッド，音声認識，ヘッドトラッキング，視線追跡などキーボードやマウスに代わるアクセッサ⁶⁾ (Accessor) が多く開発された．さらに，聴覚障害者のためにアメリカ手話言語 (ASL: American Sign Language) を生成させるアクセッサ，ソニーのテレビゲームである PS 2 へのジャイロスコープを利用したアクセッサ (TAGI: Total Access Gaming Interface) が開発された．これらの開発の過程で，多くのタイプの障害者がそれぞれの IT 機器にアクセスするために，その装置のオペレーションシステムとアプリケーションに関する多くの手順を学習する必要に直面することが分かった．これは，ユーザーにとって大きな負担であり，孤独な戦いである．これらの手順の学習への負担の軽減にも着目する必要があった．研究の成果は，人々が自然な会話とジェスチャーによって PC や他の IT 機器へのアクセスを可能にし，2つの特許を通じて iTASK に継承された．iTASK の研究は障害者だけでなく，教育，職業訓練，高齢者などのより良い生活環境にどのように生かせるかに移っている．

アルキメデスハワイプロジェクトは 2003 年にハワイ大学に移り，商業的製造とそのマーケティングに拡張された．iTASK は，PC が特に数学と科学の教育場面でその方法改善のための IDEAL Classroom⁷⁾ に応用され，高齢者が自らの自宅で自律して生活を送ることを支援するための機器「Ho'alauna」に生かされている．

4 Total Access System (TAS)

TAS はあらゆる PC や IT 機器へのユーザーのアクセスの可能性を提供

6) Neil Scott, (2003) "Archimedes at Stanford", pp 38

7) Neil G. Scott, (2004) "Creating a New Environment for Learning Math and Science"

する。それは、個人的なユーザーのインターフェースと PC や IT 機器のもつ機械的インターフェースを分離することにより実現している。ユーザーは使い慣れた同じ操作手順の機器から異なった PC や IT 機器へのアクセスが可能となっている。TAS は、①アクセッサ (Accessor), ②TAP (Total Access Port), ③共通インターフェイスプロトコル (Universal Protocol) から構成される。

① アクセッサ (Accessor)

特殊スイッチ, スキャニング, モールス信号, 手書きパッド, 音声認識, ヘッドトラッキング, 視線追跡などユーザーにもっとも適したアクセス装置

② TAP (Total Access Port)

PC や IT 機器のキーボードとマウスの役割を代替する中継的装置。

③ 共通インターフェイスプロトコル (Universal Protocol)

TAP とアクセッサの間で通信される共通の命令とその手順。

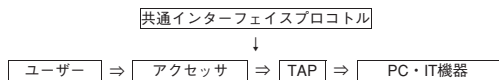


図1 TAS の概要

Neil Scott, (2003) "Archimedes at Stanford", The Archimedes Project Stanford University pp 21 より作成

共通インターフェイスプロトコルはアクセスの目的とする PC や IT 機器の命令手順に依存しない。アクセッサからの命令手順は、TAP によって共通インターフェイスプロトコルで受信される。また、TAP からは、PC や IT 機器のキーボードとマウスの命令手順のように目的とする PC や IT 機器に発信される。このように、種々のアクセッサから種々の PC や IT 機器へのアクセスを可能にしている。

5 Intelligent TAS (iTASK)

iTASK は分散型の情報技術として開発された。iTASK は、多くの異なるアクセッサ、すなわち、携帯やタブレット、タッチパネルやキーボード、特殊スイッチ、スキャニング、モールス信号、手書きパッド、音声認識、ヘッドトラッキング、視線追跡などのアクセッサと接続が可能である。さらに、その先には PC や IT 機器、ATM、エレベータ、家電製品、仕事用の道具などが目的の機器としてつながれる。iTASK は、異なる複数のアクセッサから目的の PC や家電製品にアクセスすることを可能にする。また、異なる複数のアクセッサから異なる複数の PC や家電製品にアクセスすることを可能としている。さらに、これら iTASK をネットワーク化することによって、インターネットなどに接続されたあらゆる機器にアクセスすることも可能となってくる。iTASK は、127 の装置に接続できるように設計されている。

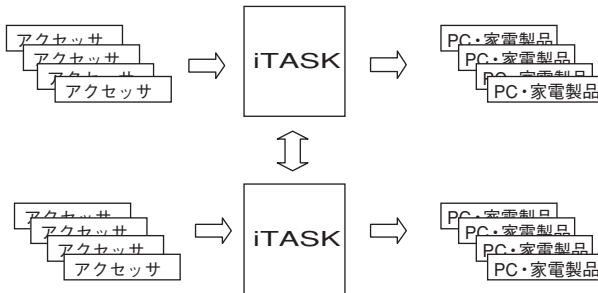


図2 iTASK の概要

<http://archimedes.hawaii.edu/ITAS.htm> より作成

この機能は、①アクセッサと iTASK との通信、②ユーザーの意図を引き出し、③iTASK の相互のネットワークを通じて、④目標の機器に命令と手順を通信する。4つのカテゴリーに分類される。音声認識やヘッドトラッキングによるユーザーの意図が、携帯や PC を通じて、ATM に対し

て金銭の取引をさせるかもしれないし、テレビやクーラーの ON・OFF を実行させるかも知れない。また、工場のオートメーション機械を制御し、製品の生産への参加を可能にするかも知れない。

6 IDEAL (Intention Driven Environment for Active Learning) Classroom

IDEAL 教室は、学生と先生が共同の教育環境において、学習経験を共有できるようになっている。特に数学と科学の学習に、新たな環境が提供できるように設計されている。近年、アメリカでは、数学、物理学、化学、コンピューターサイエンスとエンジニアリングを専攻する学生が減ってきており、不足を補うために頭脳を海外からの輸入に依存するありさまである。数学と科学の授業に学生を引き付ける必要があり、先生と学生の学習における知識と経験が、連続的に相互に交換される、刺激的な学習環境の構築が望まれる。IDEAL 教室は、あらゆる教育資源のデータベースに、様々なアクセッサである学習・教育機器などから先生と学生がアクセスすることを可能にし、結果を共有できるようにするものである。

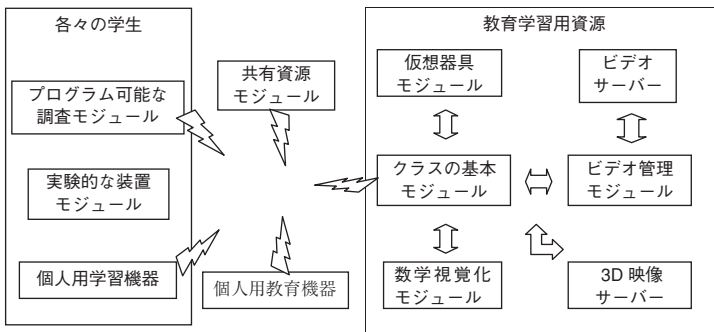


図3 教育的コンピューターシステム (IDEAL) の概要

Neil Scott, (2003) "Archimedes at Stanford", The Archimedes Project Stanford University pp 55 より作成

IDEAL 教室では、コンピュータは学生と先生が用いる機器のうちの一

つでしかない。先生は、自らの声やジェスチャーを用いて、教育資源からビデオクリップを選んでスクリーンに実行させることが出来るかもしれない。学生は携帯の通信装置を用いて、教室のスクリーンの上で、ヘッドトラッキングと目の動きで発表を行うことが出来るかもしれない。また、それらを編集することが出来る。学生や先生が音声と目の動きで、顕微鏡を用いた実験が共有できるかもしれない。IDEAL 教室では、学生と先生が連携のもとで、携帯電話やタブレット、モバイル PC や特殊スイッチなどから、教室や研究室の実験機器やスクリーンにアクセスし、それに必要なビデオや資料、文献や公式などの教育資源にもアクセスすることが出来る。

7 Ho'alauna Tablet

2025 年までに、世界の多くのところで、子供人口が 3% 増加する間に、65 歳以上老人人口は倍増するといわれている。人々にとって、年をとるということは人生のプロセスの一つであり、それに伴って、身体的・認知的能力の低下は、日々行ってきたいいつもの活動を、難しくしていくか、実行を不可能にしていくのである。結果、高齢者は住み慣れた自らの住居と地域社会を離れて管理された施設に移るか、社会に出ていくことが少なくなり、家に閉じこもり、孤立化していくかである。ハワイ語で「より良い隣人」を意味する Ho'alauna タブレットは、高齢者の、在宅と自律を可能にし、家事、健康、栄養、財産、余暇などの日常生活を支援する機器である。Ho'alauna タブレットは、生涯学習の理念に基づいて設計されており、次のことを主に目指す。

- ① どこからでも、家族や友人、医療や介護関係者と情報交換できる。
- ② 日頃の使いなれた言葉で地域の情報を入手できる。
- ③ 家電製品や家庭環境機器をコントロールできる。
- ④ 緊急情報の収集や緊急の援助をお願いすることが出来る。

- ⑤ 健康に関する情報の入手とその管理，スケジュールの管理が出来る．
- ⑥ 地域の行事や地域との交流，オンライン上のゲームや教育への参加が出来る．

表1 Ho'alauna タブレットの概要

内 容	概 要
通信	テレビ電話を含めた電子メールチャットなどの利用
地域のウェブへのリンク	地域で一般的に使われる言語によって提供されるネット情報の利用
家庭環境機器のコントロール	iTASK システムを利用して，ことばやヘッドトラッキング，特殊キーやタッチパネルを用いて家電製品などを制御
特定個人用人のスケジュール管理	個人のスケジュール管理を基本に，予定の変更・取り消しなどの自動管理にも対応
住所と電話の管理	簡略化された住所・電話記号による，ネット上の住所録や職業別電話帳へのアクセスと利用
検索と履歴	データベースシステムによって，身体的対象を含めた電子的文書や情報資源の保存と追跡
健康管理	医療・くすりの記録管理，医師およびそれに付随する情報の入手と管理，遠隔診療を利用した健康の管理と医療
セキュリティー	モニターと非常情報通信
生涯学習	地域の図書館や教育機関とのリンクを利用した教育・通信システム
オンラインゲーム	双方向通信を利用した，オンライン上での友人などとのゲームへの参加
その他通信機能	身体的・機能的問題に対応した通信機能の追加

Neil Scott and Sandy Gabrielli, (2004) "Overview of the Ho'alauna Tablet", Archimedes Hawaii Project pp 7 より作成

Ho'alauna タブレットによって，高齢者は自宅に居ながら，孫や子供たちと顔を見ながら話が出来るともしれないし，友人と将棋やトランプに興じることが出来るかもしれない．銀行の口座の操作をしたり，図書館で好きな本を読めたりするかもしれない．くすりの時間が通知され，医師の診断を受け，体調にかんする相談が出来るともしれない．緊急時には救急車

を呼び、パトカーの出勤を依頼できるかもしれない。また、電灯をつけたり消したり、テレビの音声を大きくしたり小さくしたり、エアコンのコントロールをするかもしれない。Ho'alauna タブレットは、身体的・認知的能力の低下を補い、高齢者が住み慣れた自らの住居と地域社会で自律して生きていくことを支援することが出来る。

II 音声認識ソフトの利用可能性

「音声認識ソフトの利用可能性に関する評価研究」(神戸大学発達科学部研究紀要第12巻第1号)⁸⁾では、音声認識技術の映像制御・職業開発・学習支援などへの利用を検討し、現在の開発段階では、発話に困難のある障害者や高齢者などは音声認識技術が利用できないことを指摘した。さらに、発話困難者専用の音素辞書を作成することにより認識率が大幅に向上することを示した。「構音障害者の音声認識の検討」(財団法人電子情報通信学会：日本音響学会)⁹⁾では、アテトーゼ型脳性マヒによる構音障害がある人の認識率の向上を検討した。「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」(電子情報通信学会)¹⁰⁾および「脳性麻痺構音障害者の音声認識による情報家電操作の検討」(日本音響学会)¹¹⁾では、構音障害者専用の音響モデルを用いて、障害者生活支援用の家電操作の実証的な実験評価を行った。「話者正規化に基づく構音障害者の音声認識」(日本音響学会)¹²⁾「メタモデルと音響モデルの統合による構音障害者の音声認識」(電子情報通信学

8) 李 義昭・中林稔堯(2005)「音声認識ソフトの利用可能性に関する評価研究」

9) 李 義昭ほか(2007)「構音障害者の音声認識の検討」 pp. 13-18

10) 李 義昭ほか(2007)「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」 pp. 33-38

11) 李 義昭ほか(2007)「脳性麻痺構音障害者の音声認識による情報家電操作の検討」 pp. 275-276

12) 李 義昭ほか(2008)「話者正規化に基づく構音障害者の音声認識」 pp. 215-216

会：日本音響学会)¹³⁾では、アテトーゼ型脳性マヒによる構音障害者の発話スタイルに着目し、PCAを用いた特徴量抽出手法とメタモデルと音響モデルの統合によって、認識率の改善に成功している。

1 研究の目的¹⁴⁾

近年の音声認識技術の進展にはめざましい、音声認識ソフトの開発は、「将来、キーボードは不要になるだろう」と言い、電話での問い合わせなどは、その多くが音声応答になっている。カーナビゲーションは音声で目的地を指示し、順路の説明や注意などはシステムが音声で回答するようになっていく。携帯電話では特定の相手の番号を音声で検索し、通話ができるようになった。警備やセキュリティ関係にも、音声技術を利用したものが現れている。今後は多くのIT機器や制御機器、家電製品や介護機器などに、その技術が利用されるであろう。一方、PCやIT機器の発達でコミュニケーションハンディキャップ或いはデジタルデバイトを生じさせたように、新たな技術進歩は、新たなバリアーを作る可能性を持っている。この研究は、音声認識技術の進歩においても起こりうるバリアーに配慮し、障害児(者)の音声認識技術利用に関する可能性を検討する。

2 音声認識ソフトの評価

音声認識ソフトについて、「同音異義語の修正・学習」などを行い、単語、ひらがな文、会話文、新聞記事、それぞれについて段階的な認識率の向上について評価した。また、大人(A:男性, B:女性) 子供(小学生)、知的障害(A:男性, B:女性) 身体障害(男性)など、異なる発話の特徴に

13) 李 義昭ほか(2008)「メタモデルと音響モデルの統合による構音障害者の音声認識」pp. 37-42

14) 李 義昭・中林稔堯(2005)「音声認識ソフトの利用可能性に関する評価研究」pp. 147

もつタイプに基づいて、単語、ひらがな文、会話文、新聞記事の認識率を評価した。

① 認識率の向上実験¹⁵⁾

音声認識技術が不得意とされる、単語・会話文に関しても、「同音異義語の修正・学習」を繰り返し行くと、認識率は90%を超える結果となった。また、主語／述語のある整理された文である、絵本、新聞記事は第一次の段階から、その認識率は90%を超えており、「同音異義語の修正・学習」を繰り返し行くと、さらに認識率は向上した。特に新聞記事は同音異義語を正解とした場合、その認識率は100%に達した。

表2 同音異義語の修正・学習後認識率

	単語	ひらがな文 (絵本)	会話文 (児童書)	新聞記事
第一次の認識率	30.0%	90.1%	75.3%	91.3%
第一次の同音異義語を修正した後の認識率	46.7%	89.8%	73.3%	95.4%
第一次の修正後 PC に学習させた後の認識率	76.7%	92.5%	91.7%	97.5%
第二次の同音意義語を修正した後の認識率	90.0%	96.2%	98.8%	100.0%

李 義昭・中林稔彦 (2005) 「音声認識ソフトの利用可能性に関する評価研究」 pp. 151

② 話者タイプ別認識¹⁶⁾

異なる発話の特徴を持つ、大人 (A:男性, B:女性) 子供 (小学生), 知的障害 (A:男性, B:女性) 身体障害 (男性) などに対して評価を行った結果、発話に困難・障害がある、訛りがあるなど特徴的な発話をする対象者に、極端な認識率の低下が起こった。これは、「音響モデル」に、その人たちの音が登録されていないことに起因するものと考えられる。PC や音

15) 同書 pp. 151

16) 同書 pp. 152

声認識ソフトにとって、彼らの発話は「未知の音」や「理解できない外国語」であることを意味する。

表3 話者タイプ別認識率

	単語	ひらがな文 (絵本)	会話文 (児童書)	新聞
大人 (男性)	73.1%	85.8%	87.2%	89.5%
大人 (女性)	64.5%	90.3%	76.4%	91.2%
子供 (小学生)	80.7%	94.1%	86.8%	90.5%
知的障害 A (男性)	85.2%	74.7%	62.2%	76.6%
知的障害 B (女性)	40.0%	23.5%	26.1%	26.9%
身体障害 (男性)	0.0%	実験不能	実験不能	実験不能

李 義昭・中林稔堯 (2005) 「音声認識ソフトの利用可能性に関する評価研究」 pp. 152

③ 映像制御¹⁷⁾

汎用動画作成ソフトと音声認識ソフトとの組み合わせによって、絵や画面の切り替え、動画の起動・制御などの開発実験を行った。たとえば、「アンパンマン」という発話には「アンパンマン」の映像、「バイキンマン」という発話では「バイキンマン」の映像が出現する。また、「アンパンマン」と叫べば「アンパンマン」が飛んできたり、「キック」と叫べば「アンパンマン」がキックしたりするなどの音声認識ソフトの利用可能性を検討した。

画面の切り替え実験では、音声による画面の切り替え機能を利用した、「アンパンマン」の発話に対して「アンパンマン」の絵、「バイキンマン」の発話に対して「バイキンマン」の絵、「どきんちゃん」の発話に対しては「どきんちゃん」の絵を対応させた。

動画の起動では、音声によるアプリケーションの起動機能を利用し、「アンパンマン」の発話に対して、あらかじめ用意された動画を起動させ

17) 同書 pp. 152

実行する実験を行った。具体的には一分ほどの「アンパンマン」がテーマ曲に乗って、飛来・着地する動画を用意し、それを起動・実行させた。

動画の制御では、音声認識ソフトによる PC の命令・制御機能と動画作成ソフトのマウスによる画像制御機能を利用して、音声発話による動画の制御実験を行った。具体的には、「アンパンマン」「キック」「パンチ」「危ない」などの発話による命令語、動画にはその発話された命令語に対応した制御反応を用意し、それを起動・実行させた。

④ 職業開発¹⁸⁾

音声認識ソフトのディクテーション機能を用いて、障害者や高齢者の PC の利用に関する実験を行った。特に障害があるためにキーボードが利用できない人、高齢や知的障害のために PC の利用やシステムの理解、キーボードの入力操作や構造的な理解が苦手な人に、音声入力による在宅就労の可能性を研究した。具体的には、データ入力（大学や研究所などの研究機関がアンケート調査を行った結果を、データ入力会社などに依頼して、単純な数字列データとして納品させ、それをもとに統計的な分析を行う）やテープ起こし（企業や大学などでの会議、アンケートなどの聞き取り調査の結果をテープ音声に記録しておいたものをテープ起こし会社などに依頼して、その記録した内容をテキストとして入力し、文字や文章の形で納品させる）などの在宅の仕事が出来るか、その可能性について検討した。

データ入力では、簡単なアンケート調査の結果（5人分）を用意し、大人（男性）、高齢者（男性）、障害者（女性）に対して、音声入力における認識結果を評価・実験した。具体的には、アンケート調査において記入された文字番号の結果を数字で発話し、PC 画面上に、その結果をテキスト出力させた。

テープ起こしでは、会議内容のテープ記録を用意し、それを短く聞き取りながら、内容を発話し、PC 画面にテキスト文字・文章として出力させ

18) 同書 pp. 153

表 4 認識結果

対象者	大人 (男性)	高齢者 (男性)	身体障害 (女性)
認識率	100%	99%	99%

李 義昭・小林稔堯 (2005)「音声認識ソフトの利用可能性に関する評価研究」pp. 153

た。なお、高齢者 (男性) に関しては、高齢のため軽度の難聴となっており、テープの内容が聞きとりにくいため、テープ記録の内容をテキスト文として出力された文章を読み上げてもらった。また、障害者 (女性) に関しては、テープデッキの操作に困難があるため、高齢者と同様、文章を読み上げてもらった。

表 5 認識結果

対象者	大人 (男性)	高齢者 (男性)	身体障害 (女性)
認識率	93.8%	97.8%	98.6%

李 義昭・小林稔堯 (2005)「音声認識ソフトの利用可能性に関する評価研究」pp. 153

⑤ 学習支援¹⁹⁾

音声認識ソフトのディクテーション機能を用いて、聴覚障害学生の大学講義における授業理解支援の可能性について実験・評価を行った。具体的には、現在、大学で行われている聴覚障害学生にノートテイクを付ける学習支援の補助として、講師が話す講義内容を全 90 分間に渡って PC に音声入力した。それと同時に、その出力結果をディスプレイ上 (誤認識も含めて) にテキスト文字・文として表示し、ノートテイクの補助的役割を担わせることが可能であるかを検討した。認識率の算出は、講師が話し言葉として発話した「会話文」と講師が講義レジュメを読み上げた「文章」部分に分けて算出した。なお、誤認識のうち同音異義語として出力されたものは正解とした。これは、「エンロール」「音の学習」を長時間行い「同音異義語の修正・学習」を繰り返すと、認識率が向上する音声認識

19) 同書 pp. 154

ソフトの性質を考慮したためである。

表 6 認識結果

	会話	文章
認識率	92.3%	100%

李 義昭・中林稔堯 (2005) 「音声認識ソフトの利用可能性に関する評価研究」 pp. 154

3 特定話者（特定の障害者）による音声認識技術の利用可能性²⁰⁾

発話に障害・訛りなどがある障害ある人を含めたすべての人が利用できる、音声認識技術の可能性を評価・検討した。具体的には、発話に困難のある特定の話者（障害者1名：男性）の音響モデルを作成し、企業向け汎用製品に搭載、その製品の標準音素モデル（汎用音響モデル）と新たに作成した特定話者の音素モデルによる認識率の比較評価を行った。最初に、発話に障害のある障害者1名から9単語（各5回）を音声データとして収録し、4回分を音素モデル作成用、1回分を評価テスト用として認識率の比較評価を行った。その結果、標準音素モデルでは22.23%の認識率を示し、特定話者用音素モデルでは100.00%の認識率を得た。

表 7 認識率比較評価

	特定話者用音響モデル	汎用音響モデル
ワードベース誤認識率(WER)	0%	77.78%
正解語数	9語	2語

李 義昭・中林稔堯 (2005) 「音声認識ソフトの利用可能性に関する評価研究」 pp. 155

4 構音障害者の音声認識の検討²¹⁾

アテトーゼ型の構音障害者に関して210単語による、汎用音響モデルと専用音響モデルの対照実験を行った。アテトーゼ型の構音障害者の場合、

20) 同書 pp. 154

21) 李 義昭ほか (2007) 「構音障害者の音声認識の検討」 pp. 322-324

最初の発話に緊張による不安定がある。ここでは、5回連続発話の収録を行ない、2回目以降のより安定したデータを利用した。従来の音声認識では、MFCC を特徴量として用いるが、ここでは PCA (Principal Component Analysis) による発話変動調整を行っている。汎用音響モデルでは、健常者の認識率は 89.6% であるが、構音障害者は 3.5% でしかない。モデル適応(話者適応)を行っても、構音障害者において、認識率 61.3% (210 単語使用) までの改善が限度である。構音障害者専用の音響モデルを用いた場合では、平均で 87.2% の認識率が得られたが、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いる。

表 8 発話回数ごとの認識率 (%)

1 回目	2 回目	3 回目	4 回目	5 回目
77.1	89.1	91.4	90.0	87.6

李 義昭ほか (2007) 「構音障害者の音声認識の検討」 pp. 324

1 回目発話に対し PCA による発話変動調整を適応した場合、主成分 11, 13, 15, 17, 19 個に対して、主成分 15 個では 85.2% の認識率の向上を得ている。

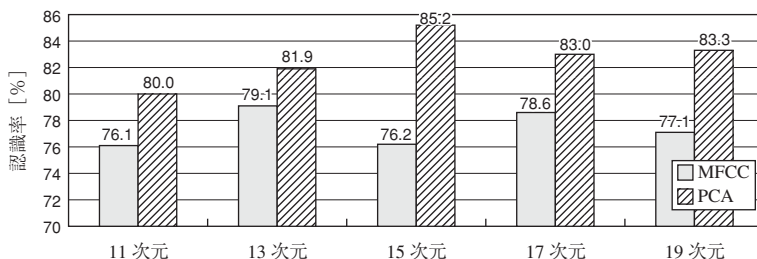


図 4 PCA による 1 回目発話の認識率

李 義昭ほか (2007) 「構音障害者の音声認識の検討」 pp. 324

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音響モデルで

は認識が困難であるが、構音障害者専用の音響モデルを用いることで音声認識技術の利用が可能であることがわかった。また、1回目の不安定発話に対する特徴抽出方法として MFCC に代わり、PCA を用いる手法がより有効であることが分かった。

なお、さらなる改善として、各話者の音素毎の置換、挿入の傾向を音声認識の過程に組み込むことが可能なメタモデルとの専用音響モデルの統合によって、認識率において 9.9% (79.1%→89.0%) の改善が得られることが分かっている。

5 脳性麻痺構音障害者の音声認識による情報家電操作の検討²²⁾

環境制御装置は、重度障害者の自立支援機器として重要な役割を担っている。その操作方法として、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がある。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。ここ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、構音障害者専用の音響モデルを作成した。環境制御装置の操作では、7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 個の音声コマンドからなるタスクに対して、459 発話（各コマンド 8～15 回）を用いている。

汎用音響モデルでは、平均認識率は 28.8% と認識が困難なため、構音障害者専用の音響モデル（54 音素 6 混合）を作成した。結果、認識率は 76.5% まで改善した。

構音障害者の場合、日々の体調や環境によって発話スタイルが変動することから、適応データ数 2, 5, 10 発話の話者適応を行うと、10 発話では 84.5% までの認識率の改善がみられている。

22) 李 義昭ほか (2007) 「脳性麻痺構音障害者の音声認識による情報家電操作の検討」 pp. 275-276

表9 タスクの内容

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

李 義昭ほか (2007) 「脳性麻痺構音障害者の音声認識による情報家電操作の検討」 pp. 276

汎用音響モデルでは、平均認識率は 28.8% と認識が困難なため、構音障害者専用の音響モデル (54 音素 6 混合) を作成した。結果、認識率は 76.5% まで改善した。

Ⅲ 音声認識研究とアルキメデスプロジェクトの連携

1 社会的不利²³⁾の除去

それぞれの研究計画は、障害者が受けている社会的不利を排除しようと

23) 障害には、障害当事者など本人の身体や機能に関わる「医学モデル」と

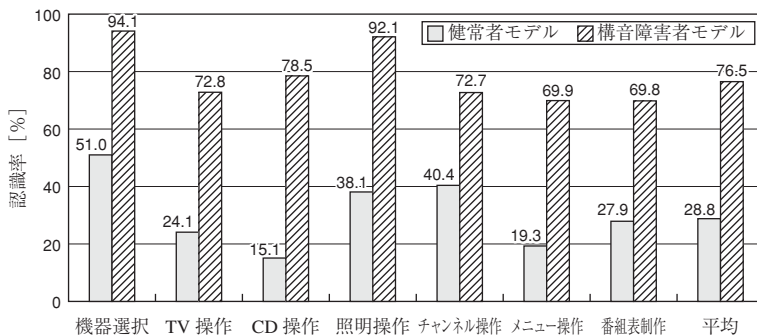


図 5

李義昭他 (2007) 「脳性麻痺構音障害者の音声認識による情報家電操作の検討」 pp. 276

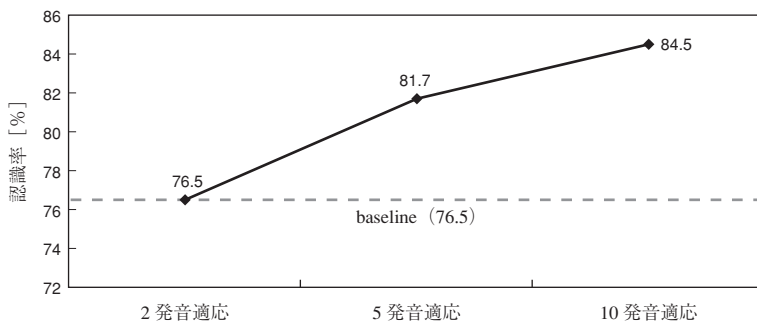


図 6

李義昭ほか (2007) 「脳性麻痺構音障害者の音声認識による情報家電操作の検討」 pp. 276

する動機に始まっている。排除の必要は、社会が創り出した多くの人々への便益が、往々にして一部の人々へ不利益をもたらすか、一部の人々を置き去りにして発展していつてしまう現象に起因するものと思われる。高速道路が走り、道路整備され、歩道があり、信号機が設置された、効率的な空間。高層ビルが林立し、その間を鉄道が走る、地下空間も利用され、そ

- 社会が創り出す社会的な不利益に関わる「社会モデル」があり、バリアフリーは一般的に、この社会的不利を取り除くこと。

こを地下鉄が貫く。改札は自動で行われ、切符の販売やお金の支払いも機械が管理する。われわれが創ってきた便利な都市空間は、障害者や高齢者、子育て中の女性たちを不便にし、危険にさらし、置き去りにして発展してきた。障害者や高齢者、子育て中のお母さんたちにとって、効率的に車を走らせ、整理するための道路や信号機は、そのテンポが彼らに適していないためにかえって危険な存在となっている。また、高層ビルや地下街への上下の移動は、彼らにとって労力を要し不便である。機械が処理する様々な物事は、彼らには手順が複雑で、理解に時間を要する。そのことに気づいて我々は、いま、町に点字ブロックを設けスロープを作り、エスカレーターやエレベータを設置し、低床のバスを走らせ、車の入れない空間を設けている。信号機も改札も自動販売機も、ひとにやさしい、高齢者や障害者でも苦勞なく利用できるものへと改良を行っている。私たちは、今まで不便を与え、置き去りにしてきた人々を、私たちの効率的で便利な社会で再び一緒に暮らしていくために多くの社会的費用を投じている。IT技術や機器の発展も、彼らを置き去りにし、社会的不利を与えてきたのである。

① アルキメデスハワイプロジェクト

アルキメデスプロジェクトは、身体障害者が完全にIT装置にアクセスできるようすることにあつた。そのために、身体障害者それぞれに適した、特殊スイッチ、スキャニング、モールス信号、手書きパッド、音声認識、ヘッドトラッキング、視線追跡などキーボードやマウスに代わる、IT装置へのアクセスの道具であるアクセッサ(Accessor)が多く開発された。また、個々の障害者が、それぞれのIT機器にアクセスするために、その装置のオペレーションシステムとアプリケーションに関する多くの手順を学習する大きな負担を軽減するために、それぞれの障害者が専用を持つインターフェースとPCやIT機器のもつ機械的インターフェースを分離した。

② 特定話者用音声認識

音声認識研究においても、その始まりは、近年の音声認識技術が、発話障害者には利用できないという事実にある。もともと、この音声認識する技術が音スイッチ（音が入るとスイッチが作動する）であった時代、誰もがアンパンマンと叫べばアンパンマンが音楽と一緒に現れ、打てと叫べばバッターはボールを打ったのである。しかし、音声認識技術の精度が高くなり、人々がアンパンマンと叫べばアンパンマン、バイキンマンと叫べばバイキンマン、文字を読めば文字が、文章を読めば文章がディスプレイ上に現れるようになった現在、発話に障害や困難のある人々には、発展した音を認識する技術は使えなくなってしまったのである。言葉で作動するカーナビ、音声に応える電話案内などわれわれの持っている便利な財産へのアクセスを障害のある人々へは禁じているのである。

今、点字ブロックやスロープなどへ社会的投資を行って、彼らと共に生きていこうとしているように、IT 技術や機器の分野においても、新たな社会的費用をかけて、彼らのアクセスを保証する必要がある。そのため、アルキメデスプロジェクトは、TAS において多くのアクセッサを開発して PC・IT 機器へのアクセスを可能にし、iTASK において様々な複数のアクセッサから複数の PC・家電製品、そして、iTASK のネットワークを介して、他のアクセッサがアクセスしている様々な PC・家電製品へのアクセスを可能にしている。また、音声認識研究では、特定話者の音声認識技術の利用可能性に関して、発話に障害・訛りのある障害者も音声認識技術が利用可能であるということを証明した。そして、構音障害者の音声認識の検討では、かなりの単語数でも、音声認識技術は実用範囲で利用が可能であるということを示している。さらに、脳性まひ構音障害者の音声認識による情報家電操作では、発話障害者も自立支援機器としての環境制御装置が実用的に利用できることが可能であるということを示した。

2 誰もが使える道具²⁴⁾

障害者や高齢者など、一部の人々を排除せずまた置き去りにせず、もともと全ての人々が利用できるデザインや設計にしておけば、社会的不利をあえて排除しなくても済む。都市構造において、はじめから、障害者や高齢者が、子育て中のお母さんなどが生活・行動しやすいように設計しておけば、社会的不利を排除するための支出は軽減される。さらに、スロープやエレベータなど、人々にやさしい都市設計であれば、宅配事業者や運送事業者にとっても、作業負担の軽減となり、経済的効率性につながり、社会の発展に寄与できる。また、家電製品のデザインにおいても、はじめから、障害者や高齢者などを含めた人々の多くが使用できるように創っておけば、あえて、障害者用や高齢者用といった製品を開発する必要はなくなる。多くの人々が使える製品は、その製品を消費できる対象者が多くなるということであり、製品の需要の増加が期待される。経済的には、より良い製品がより安く、供給されることが期待できるのである。携帯電話のキーや文字表示を少し大きくすることによって、多くの高齢者が、携帯電話を利用しやすくなった。テレビなど家電製品のリモコンは寝たきりの障害者や病人には、他の人に負担をかけず、自らの意思を遂行できる欠かすことのできない道具である。しかし、現在では、その必要のない人々にも、便利な道具として利用されている。むしろ、便利な道具であることが主流となっており、リモコンが付属していない家電製品のほうが少ないといっても過言ではない。一方、現実の問題として、多くの人々にやさしい都市構造や、多くの人々が利用できる製品の設計やデザインはできても、全ての人々にやさしい都市構造や、全ての人々が利用できる製品の設計やデザインをすることは至難の業である。たくさんの人々のニーズを知るため

24) ユニバーサルデザイン、できるだけ多くの人々が利用可能であるようなデザインにすること。障害者など特定ものを対象とするバリアフリーとは異なる。

に、多くの調査時間や費用を割かなければならない。また、不完全な調査のために、その結果に基づいて制作された製品が一部の人々に不利益をもたらすものであったりすることがある。点字ブロックは視覚障害者のニーズに適合しているが、ヒールを履いている人、車いすやバギーカーにとっては、不安定や不快感を生み出す結果となる場合がある。このように、一方のニーズが他方のニーズに相反することがあるのである。経済的には、初期投資が膨大になり、失敗したときのリスク負担を大きくし、提供されない、または、提供が大きく抑えられる商品になってしまう可能性を秘めることになる。

① アルキメデスハワイプロジェクト

アルキメデスハワイプロジェクトでは、TASにおいて、PCやIT機器のキーボードとマウスの役割を代替するTAP (Total Access Port) という中継の装置を用意することによって、ユーザーのインターフェースとPCやIT機器のもつ機械的インターフェースを分離している。このことによって、ユーザーは使い慣れた同じ操作手順の機器から異なったPCやIT機器へのアクセスが可能となっているのである。経済的には、開発のための費用を抑え、機器の更新のための費用を抑える効果も持っていることになる。具体的には、目が見えない、手が使えないなど異なるタイプの障害者のように、それぞれ異なる個性を持った障害者が5人おり、アクセスしたいPCや家電製品など異なる操作手法のものが5タイプあるとした場合、5人の障害者が、5つのタイプの機器を使いこなすには、 $5 \times 5 = 25$ 通りの機器の開発が必要となってくる。しかし、TAP (Total Access Port) という中継の装置の存在によって、その開発すべき組み合わせは、 $5 + 5 = 10$ 通りに軽減される。これは、取引数量最小化の原則²⁵⁾で説明される現象であ

25) メーカーがA社あり、小売店がB社ある場合、メーカーと小売店の直接取引の取引数は $A \times B$ になる。しかし、卸売店が仲介すれば、取引数は $A + B$ と最小化される。

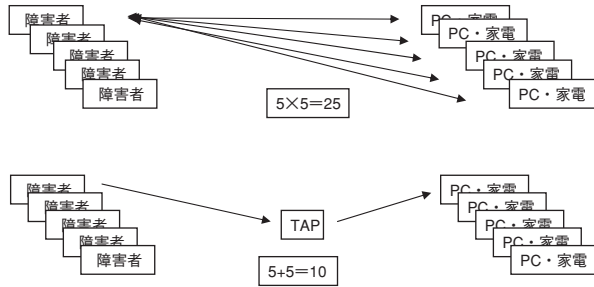


図 7

る。さらに、頻繁に更新されたり、規格が変更されたりする PC や IT 機器に対しても、大きな費用をかけずにアクセスする方法を選択することが可能である。

また、iTASK によって、面的にアクセスの範囲を広げ、多くの IT 機器や家電製品への相互の利用と制御を可能にしている。IDEL 教室では、数学と科学に関する学習に必要な知識と経験を多くの人々が共有できるよう保証し、特に、学生および障害学生が、先生とその学習環境を共有し、多くのところから、多くの教育学習用資源にアクセスできるようにしている。Ho'aluna タブレットでは、高齢者が地域の多くの情報を、家に居ながらにして利用できるよう予定している。

② 特定話者用音声認識

音声認識研究では、音声認識技術が多くのところすでに利用されており、話者適応や同音異義語の修正学習を繰り返すと、発話が正確であれば、障害があり手が使えない、複雑な機器の操作が出来ない、目が見えないなどであっても、認識率が実用の範囲にまで達することを確認した。映像の制御においても、発話によって、特定の画像を出現させることが可能であること、アプリケーションの起動が可能であること、画面上の映像などに、目的の行動を実行させることが可能であることを確認している。また、キーボードが利用できない、PC の利用やシステムの理解に困難があ

る、障害者や高齢者の職業能力開発に、音声認識技術が有用であり、利用可能であることを見た。障害者や高齢者が適度の訓練を受け、この技術を使いこなすのであれば、これらの職域への参入が可能である。経済的には、新たな参入は、供給の増大を通じて、より安く、より良いサービスが生まれることを意味する。大学教育における学習支援では、聴覚障害学生が音声認識技術を利用して、講義の内容を理解し、学習することを促進できることを証明した。現在行われている、これら聴覚障害学生への支援は、ノートテイク、いわゆる要約筆記者であるが、主に、主婦のボランティアなど、低額報酬の支援者が協力している。それでも、聴覚障害学生への学習支援にかかる費用は、年間1名あたり約2百万である。また、大学の専門的な講義内容の要約が、現在の主婦などの人材で正確にできるかの問題を抱えている。音声認識技術が有用に利用されるのであれば、支援コストの削減にもつながり、より多くの学生が支援の利益に浴することとなる。さらに、教師の側が、前もってPCに多くの著書・論文などを学習させておく、また、音声を認識させておくなどの事前準備を行っておくならば、学生が専門的講義内容を理解する上で大きく寄与できるものと考えられる。特定話者の音声認識技術の利用に関しては、発話に障害がる人々にも、この技術が原則的には利用可能であることは判っている。多くの発話困難者の音声を収集するために多くの時間と費用が必要であるという課題はあるものの、専用の音響モデルを構築することによって、認識率は向上しており、発話特性などの問題も解決しつつある。いずれ、発話に困難を有する人々も、われわれが持っている財産である音声認識技術の世界へ参加できる日が来るであろう。

アルキメデスハワイプロジェクトと特定話者用音声認識技術の連携は、PCやIT機器に少しでもアクセスしたい、PCやIT機器を誰でもが利用できる道具にしたいという願いを達成へ近づけることになる。

おわりに

すでに、PC や IT 機器は、われわれの経済社会や一般家庭に、組み込まれ、普及して久しい。近い将来、音声認識技術はすべての社会システムや家庭における電子機器の制御に活用されているものと考えられる。そのとき、私たちは障害者や高齢者を含めた、全ての人が、社会的資産である、これらシステムや機器にアクセスでき、そこからもたらされる情報や知識を取得できるよう保証しなければならない。見てきたように、アルキメデスプロジェクトは、様々な機能を持った人々や様々な制限を持った人々でも、その個性を生かして、PC や IT 機器などの装置を通じて、情報や知識を取得し、経済社会や地域社会に参加できることを可能にしている。音声認識技術研究も、発話に問題がある人々や発話に障害がある人々が、社会資源である音声認識技術の利用を可能にしようとするものである。全ての人の PC や IT 機器を通じた、経済社会や地域社会へのアクセスを保証するためには、さらに多くの人々のニーズに対応するアクセッサの開発が必要である。発話困難者が、われわれと同じように、音声認識技術を利用できるようになることは、音声を用いたアクセッサを通じて、経済社会や地域社会に参加できることを意味する。特定話者用音声認識ソフト開発とアルキメデスプロジェクトの連携は、発話困難者の世界をさらに広げるものである。

参考文献

1. <http://archimedes.hawaii.edu/>
2. Neil Scott, (2003) “Archimedes at Stanford”, The Archimedes Project Stanford University
3. Neil Scott and Sandy Gabrielli, (2004) “Overview of the Ho’alauna Tablet”, Archimedes Hawaii Project

4. Neil G. Scott, (2004) “Creating a New Environment for Learning Math and Science”, Archimedes Hawaii Project University of Hawaii
5. 李 義昭・中林稔堯 (2005) 「音声認識ソフトの利用可能性に関する評価研究」神戸大学発達科学部研究紀要第 12 巻第 1 号
6. 李 義昭・松政宏典・滝口哲也・有木康雄・中林稔堯 (2007) 「構音障害者の音声認識の検討」財団法人 電子情報通信学会技術研究報告 WIT 2006-75 pp. 13-18
7. 李 義昭・松政宏典・滝口哲也・有木康雄・中林稔堯 (2007) 「構音障害者の音声認識の検討」日本音響学会 2007 年春季研究発表会, 1-Q-30, pp. 322-324
8. 李 義昭・松政宏典・滝口哲也・有木康雄・中林稔堯 (2007) 「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」電子情報通信学会技術研究報告, WIT 2007-7, pp. 33-38
9. 李 義昭・松政宏典・滝口哲也・有木康雄・中林稔堯 (2007) 「脳性麻痺構音障害者の音声認識による情報家電操作の検討」日本音響学会 2007 年秋季研究発表会, 3-Q-29, pp. 275-276
10. I-chao LI, Hironori Matsumasa 1, Tetsuya Takiguchi 1, Yasuo Arika 1, Toshitaka Nakabayashi (2007) “PCA-Based Feature Extraction for Fluctuation in Speaking Style of Articulation Disorders” Interspeech 2007, pp. 1150-1153
11. 李 義昭・松政宏典・滝口哲也・有木康雄・中林稔堯 (2008) 「話者正規化に基づく構音障害者の音声認識」日本音響学会 2008 年春季研究発表会, 1-Q-24, pp. 215-216
12. 李 義昭・松政宏典・滝口哲也・有木康雄・中林稔堯 (2008) 「メタモデルと音響モデルの統合による構音障害者の音声認識」電子情報通信学会技術研究報告, WIT 2008-7, pp 37-42
13. 李 義昭・松政宏典・滝口哲也・有木康雄・中林稔堯 (2008) 「メタモデルと音響モデルの統合による構音障害者の音声認識」日本音響学会 2008 年秋季研究発表会, 2-P-8, pp. 127-128
14. I-chao LI, Hironori Matsumasa 1, Tetsuya Takiguchi 1, Yasuo Arika 1, Toshitaka Nakabayashi (2008) “Integration of Metamodel and Acoustic Model for Speech Recognition” Interspeech 2008, pp. 2234-2237

Abstract

From the experience, we understand that new technical progress may lead to a new barrier. The development of the information technology causes a communication handicap or a digital divide. The Archimedes project is going to guarantee a physically challenged and/or a senior citizen the access to information technology. It guarantees access to the information and the knowledge. In addition, the technical study of the speech recognition is going to guarantee the utterance handicapped person access the speech recognition technology. The information technology has already been commonly accepted in the economic society and at home. In the near future, the speech recognition technology will be applied to all electronics used in all aspects of life. We must guaranteed the information and the knowledge can be accessible by all the people including a handicapped person and a senior citizen. It means the participation of all people including an utterance handicapped person from the economic society and commutes.

Key words : speech recognition The Archimedes Project handicap

(2009年6月26日受理)