

## 携帯端末用多言語自動通訳システムの実用化技術に関する研究開発

奥村 明俊 服部 浩明 磯谷 亮輔 山端 潔 土井 伸一 安藤 真一

NEC メディア情報研究所

### 1. はじめに

近年、グローバルにインターネットやブロードバンド技術が進展する中、ユビキタス社会の実現を通じて、新しいコミュニケーションのあり方やコミュニケーションによる新たな価値創造が求められている。これからのコミュニケーションの新形態は、ネットワークを通じて人々の相互理解と協働活動を形成し、より高い価値を創発していく環境を実現するものである[1]。そのためには、超越しなければならないコミュニケーションの「壁」が多数存在する。例えば、異なる言語、異なる文化、異なる価値観といった壁が存在し、それが障壁となってコミュニケーションを阻む。したがって、ユビキタスネット社会の実現にはこれらの壁を打破できる技術の創出が不可欠であり、とりわけ、異言語コミュニケーションを可能とする自動通訳は、必須の技術である。また、自動通訳技術は、国内外で様々な経済活動や人的交流がグローバル化している日本において将来の高度情報社会におけるすべての産業の基盤となる21世紀の中核技術であり、産業活動のみならず国民生活全般に関わる社会基盤となるものである。現在、アジア諸国が急成長しており、我が国の経済は今後ますます厳しい国際競争に直面していくことになる。国際化に立ちはだかる言語の壁は大きく、2006年の国際経営開発研究所による日本の「語学力評価」は、61カ国・地域の中で55位となっている。言語の壁は、将来の日本の発展のために克服しなければならない壁である。言語の壁は、旧約聖書の創世記によるとバベルの塔の建設に怒った神による仕業という(図1-1)。これは、言葉障壁の克服が古の時代から人類の夢であったことを示す伝説と言える。

- ・昔、世界の誰もが同じ言葉をした(旧約聖書の創世記 11:1)
- ・あるとき、人々は集まって天まで届く塔を造り始めた(11:3).
- ・神はこれを見て、人間の尊大をこらしめるため、言葉を乱して、お互いに意志が通じ合わないようにした。(11:7).



図1-1.バベルの塔の伝説と絵画(ピーテル・ブリューゲル)

我々は、この夢の実現に向けて、1983年のTelecom'83において小林宏治(当時NEC会長)が自動通訳電話のコンセプトモデルを発表して以来、一貫して世界的に技術開発を主導し、2006年、携帯端末用多言語自動通訳システムの実用化技術の構築と製品化に成功した。この技術は、携帯情報端末(Personal Digital Assistant, PDA)や携帯電話などにおいて、日本語、英語、中国語をリアルタイムに自動通訳する技術である。この技術により、

5万語という大規模な語彙を用いた日英旅行会話通訳機能が、モバイルマルチメディアプレーヤ VoToL という携帯端末上に世界で初めて製品搭載された[2][3]。また、携帯電話用マルチコアプロセッサにおいてリアルタイム日英旅行会話通訳システムが実現された[4][5]。さらに、本システムをベースに日中通訳システムを実現することにより、多言語展開可能なプラットフォームであることを確認した[6]。この技術は、文字通り言葉の壁をブレークスルーするものであり、日本の将来を飛躍させる異言語コミュニケーション技術として期待されるものである。

## 2. 研究開発の背景

自動通訳の実用化に向けた基本的な課題として、数万語以上の大語彙の発話を携帯端末単体でリアルタイムに通訳すること、さらにより多くの場面に展開するための課題として、多言語展開可能なプラットフォームであることがあげられる。

### 2.1 携帯通訳端末

平成18年版の観光白書（国土交通省）によると、平成17年の海外旅行者数は約1,740万人と過去最高を記録した平成12年の1,782万人に次ぐ人数となった(図2-1)。

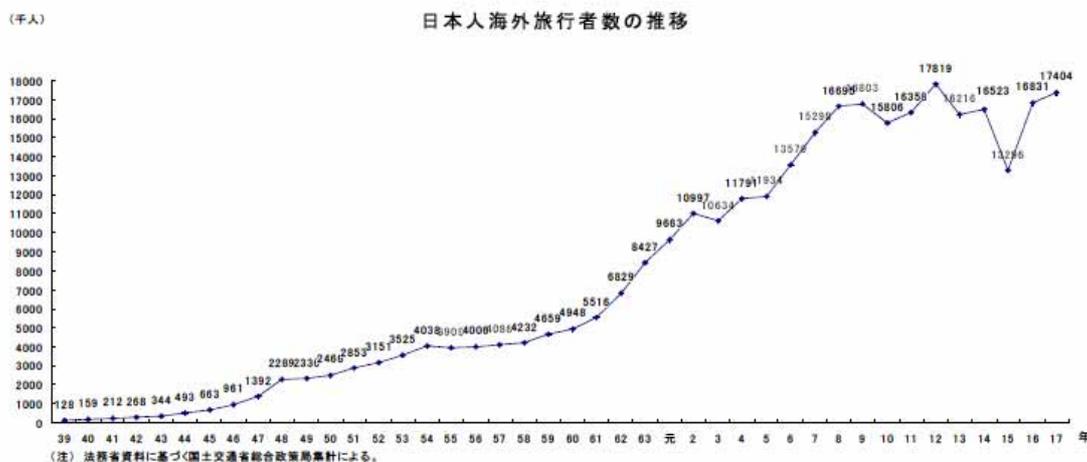


図 2-1.日本人海外旅行者数の推移

しかし、多くの日本人は、英語によるコミュニケーションに何らかの不安を感じている。例えば、「内野席のチケットが欲しい。」「膝をすりむいた。」「頭がずきずきする。」といった単純な内容であっても、英語で即座に正しく言える日本人は多くはない。海外を訪問する日本人に、携帯可能な通訳端末が強く望まれている。日本政府による「ビジット・ジャパン・キャンペーン」など外客誘致のための取組が推進された結果、平成17年には、訪日外国人旅行者が673万人となり、過去最高を記録した(図2-2)。携帯通訳端末は、訪日外国人旅行者にとっても望まれるものであり、訪日外国人旅行者数の増加に寄与するものと思われる。

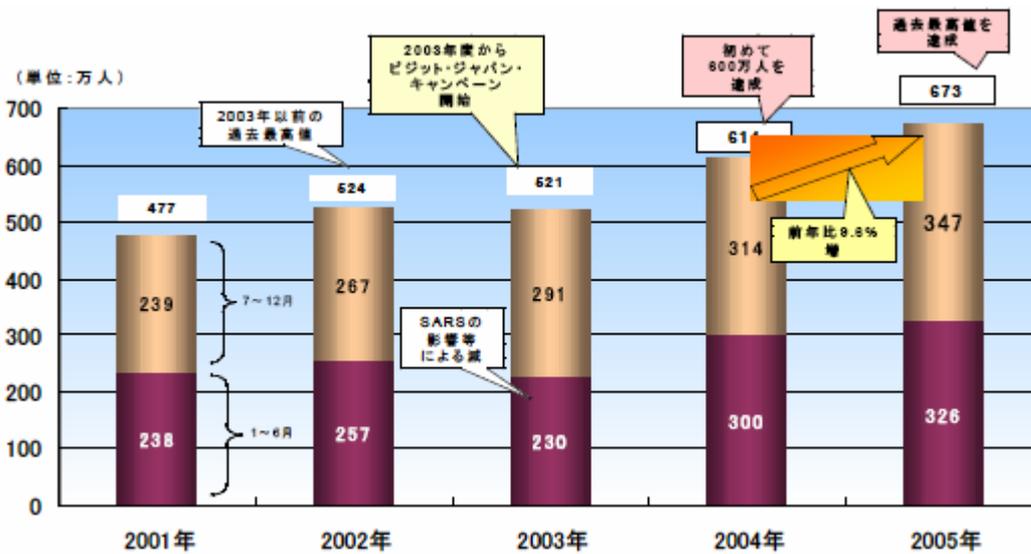


図 2-2.訪日外国人旅行者数の推移

このような状況の中で、イノベーション 25 戦略会議は、ヘッドホンなど簡易なデバイスによる自動通訳を生活・産業分野においてイノベーションすべき技術と位置づけている [7]。自動通訳機による言語バリアの低下、日本のオープンな環境、アニメ・食などの日本文化のブランド力・発信力強化などにより、訪日外国人旅行者数が日本人の海外旅行者数並に増えれば、約 2.7 兆円の旅行消費額の増加が見込まれる(平成 17 年度推計)。さらに、そうした多数の外国人を対象とした新たなサービス産業等の出現を見込んでいる。また、米 IBM は今後 5 年間で人々の働き方や生き方、遊び方を一変させるインパクトを持つであろう 5 つの技術革新「IBM Next Five in Five」のひとつとして自動通訳をあげ、リアルタイム通訳技術が携帯電話や携帯機器、自動車などに組み込まれ、ビジネスや生活の現場に浸透するようになると予測している。自動通訳は、国内外で研究開発が進められ、ノートパソコンによる通訳システムの製品開発や、サーバと端末を用いた研究開発が行われてきた [8][9][10][11][12]。しかしながら、ノートパソコンによる通訳は、重量、大きさ、バッテリー寿命、OS の起動時間を含めた処理時間などから、携帯して利用するのは現実には困難である。ネットワークを介してサーバで通訳する場合、ネットワーク接続不可能な場所での利用可能性やネットワーク利用による経済的負担などの課題により実用化されていないのが現状である。このような理由で、国内外で携帯通訳端末の実現は強く求められてきた。PDA による通訳的な支援機能としては、音声検索ベースの定型文翻訳が実現されているが [13]、数万語以上の大語彙の発話を携帯端末単体でリアルタイムに通訳する製品の発表には至っていない。

## 2.2 多言語自動通訳

平成 18 年版観光白書によると、訪日外国人は英語圏以外が半数を超える(図 2-3)。また、今後の世界の主要言語人口の推移によると、中国語人口が大きく伸びることが予想され(図 2-4)、経済活動上、中国語の重要性は更に高まると予想される。自動通訳技術の実用化のため

め、多言語対応可能なプラットフォームを築くことが必要である。

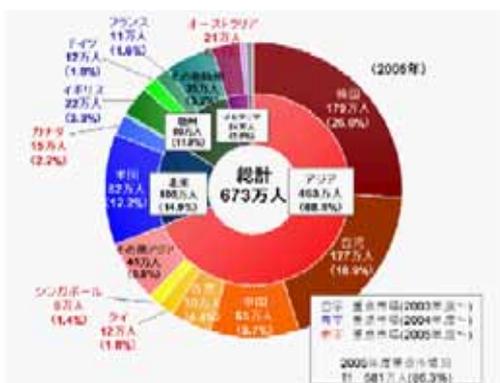


図 2-3.訪日外国人国別内訳

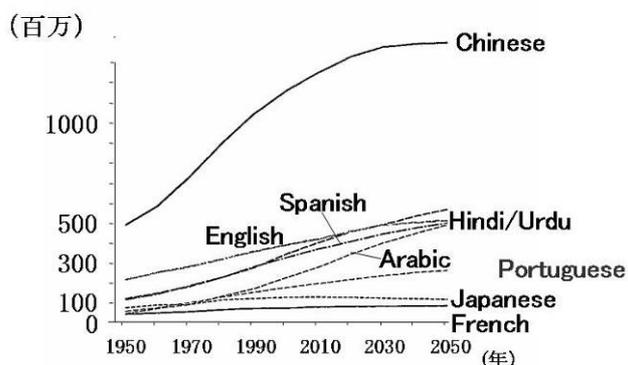


図 2-4.世界の主要言語人口の推移

### 2.3 研究開発戦略

上述したように、携帯端末で大語彙(数万語規模)のリアルタイム通訳を実現すること、さらにその機能を多言語展開可能とすることが、本研究のブレークスルーポイントである。このブレークスルーの達成によって、既存の電子辞書のような限定されたものではなく、いつでもどこでも誰とでもコミュニケーション可能となる真の携帯通訳端末が実現されるのである。我々は、以下の戦略によって、このブレークスルーに挑戦した。

- 1) 日本語と英語に関して、ニーズの大きい旅行会話を対象として PDA クラスの端末でリアルタイム通訳を実現する技術を確立する。具体的には、通訳を構成する要素技術である音声認識、言語翻訳、音声合成に関する高精度かつコンパクトなアルゴリズムを実現する。
- 2) 上記技術を基に携帯通訳端末の実機を試作し、成田空港など多数の外国人と遭遇する場で実証実験を行う。この実験をもとに改良開発を行って世界初の製品化を達成する。
- 3) さらに携帯性に優れている携帯電話において、PDA と同程度の性能を有するリアルタイム通訳技術を確立する。この技術を携帯電話用のプロセッサ上で評価することによって、携帯電話によるリアルタイム通訳を実証する。
- 4) 日本語と英語で確立したリアルタイム携帯通訳技術が、中国語に適用可能であることを示す。世界の主要言語は、その言語的な特徴から、膠着語(日本語や韓国語など)、屈折語(英語など欧州言語)、孤立語(中国語やチベット語など)の言語類に分けられる。通訳技術の同一言語類への展開は比較的容易であるが、異なる言語類への展開可能性は自明ではない。孤立語に属する中国語は、アクセントのバリエーションが最も大きな言語であり、かつ社会的ニーズも大きい。中国語における実用性を検証することによって、多言語展開可能なプラットフォームであることを実証する。

以下、携帯端末と多言語自動通訳に関する取組みについて技術的詳細を述べる。

### 3. 携帯通訳端末に関する取組み

PDA 上でも高速に動作するコンパクトな自動通訳技術を開発し、旅行会話を対象とした日英双方向自動通訳システムを PDA 上に実現した[14]。このシステムは、自動通訳の要素技術である音声認識技術、翻訳技術、日本語音声合成技術をコンパクト化、高速化して PDA 上で統合することにより実現されたものである。以下に各要素技術のブレークスルー技術について詳述する。

#### 3.1 携帯通訳システムの概要

通訳システムは、日本語・英語の音声認識モジュール、日英・英日の翻訳モジュール、日本語・英語の音声合成モジュールを有し、これらエンジンモジュールを統合モジュールが制御するプラットフォーム上に構築される(図 3-1 参照)。今回、英語音声合成以外のすべてのモジュールを新規開発した。システムは、64MB 以上のメモリと、400MHz 以上の CPU、

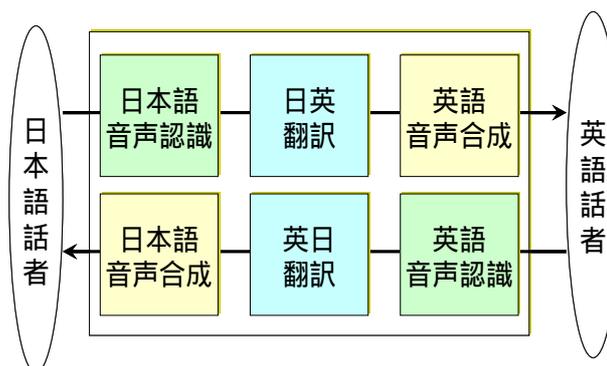


図 3-1.自動通訳システムの構成

ならびに 128MB 以上のメモリカードを持つ PDA の上で動作する。PDA に実装されたシステムの概観を図 3-2 に示す。PDA のマイクに向かって、マイクボタンを押して、例えば、「六時までこの荷物を預かっておいてもらえますか」と発声すると、まず音声認識結果が画面に表示され、次に翻訳結果「Could you keep this baggage until six o'clock?」が画面表示されるとともにスピーカより英語合成音出力される(図 3-3 参照)。

[動画デモ：<http://www.labs.nec.co.jp/Overview/soshiki/media/images/PDA1.mpg>]

その間約 1 秒である。日本語話者と英語話者は、翻訳言語(日英方向もしくは英日方向)と話者の性別(男性もしくは女性)を選択した後に発声し、交互に利用してコミュニケーション可能である(図 3-4 参照)。上記成果をもとに、国土交通省の「e-エアポート」構想



図 3-2. PDA に実装されたシステム

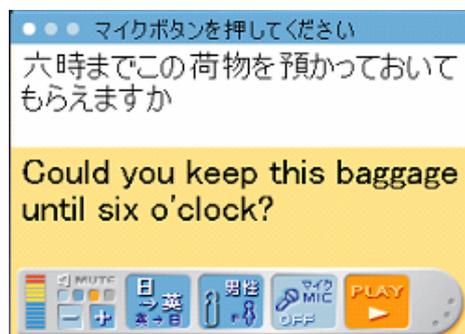


図 3-3.通訳画面表示例

の一環として成田空港における実証実験[15][16]などのフィールドテストを繰り返し、話者の性別指定を不要とするなど製品開発を進めた。その結果、2006年5月、日英旅行会話通訳機能を、モバイルマルチメディアプレーヤ VoToL という携帯端末上に搭載し世界で初めて製品搭載した(図 3-5) [2][3]。VoToL は、動画や音楽ビューアーであり、PDA 同等のハードスペックを有する。内蔵されたマイクに向かって発話することによって、日英双方の旅行会話通訳が可能である。



図 3-4.PDA による通訳利用シーン例

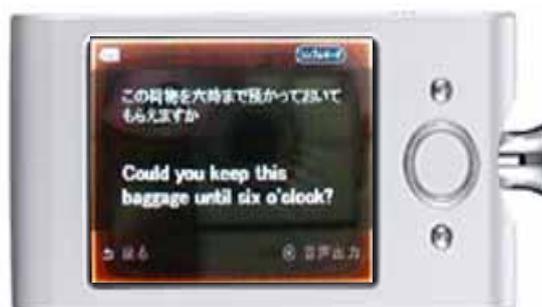


図 3-5.モバイルマルチメディアプレーヤ VoToL

### 3.2 音声認識

音声認識モジュールはマイクロホンから入力された音声波形を認識してテキストに変換する(図 3-6)。日本語の入力音声を認識するためには、日本語の音響モデル、単語辞書、言語モデルが必要である。単語辞書には単語とその発音記号(たとえば音素列)が登録されている。音響モデルは、各発音記号がどのような音声波形として観測されるかをモデル化している。言語モデルは単語がどのような並びで文を構成するかをモデル化している。音声認識のサーチエンジンは、これら3つの知識源(音響モデル、単語辞書、言語モデル)から予測されるさまざまな単語列と、未知の入力音声のあいだの類似度(確率値)を算出し、最も確からしい単語列を認識結果として出力する。

自動通訳で必要となる音声認識の特徴は、利用者があらかじめ声を登録する必要なく、誰でもシステムを使用できること(不特定話者音声認識)、および、豊富な語彙を持ち、どのような単語・文なら受理されるかを意識することなく、自然に自由な発声を認識させることができること(大語彙連続音声認識)である。このような不特定話者の大語彙連続音声認識技術の

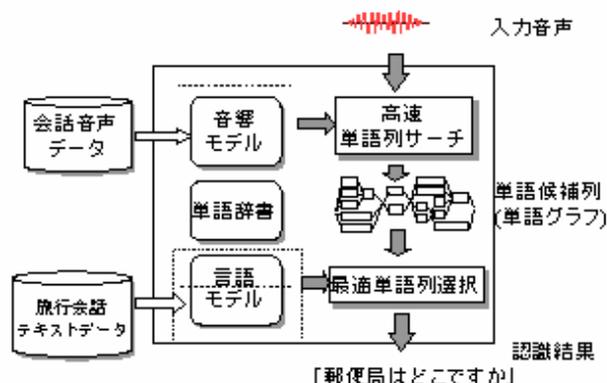


図 3-6. 音声認識モジュール

実現には、従来 500MHz 以上の高速 CPU や、128MByte 以上の大容量メモリ、100MByte 近いハードディスクが必要であった。以下の技術開発により、不特定話者の大語彙連続音声認識の演算量とメモリ量を大幅に削減し、PDA での動作を実現した。

### 3.2.1 音響モデル

音響モデルは各発音記号に対して、その音声パターン（音声波形から特徴抽出された特徴ベクトル）がどのように分布するかを、事前に収集した多数の話者の音声波形をもとに多数のガウス分布の組み合わせで表現し、未知の入力音声に対してすべてのガウス分布の確率値を計算して、サーチエンジンに渡している。音響モデルの使用メモリ量や演算量を低減するために、MDL（記述長最小）基準と呼ばれる情報量基準を用いて、認識率の劣化を抑えつつガウス分布の数を効率的に削減した。メモリ使用量をさらに低減するために、ガウス分布の共分散行列を共通化した。さらに、すべてのガウス分布をお互いの近さを尺度として、いくつかの組に分類し、未知の入力音声に対して、はじめにどの組に近いかを判定し、選ばれた組のガウス分布の確率値だけを計算することにより演算回数を削減した。

以上のメモリ量・演算量削減手法により、音響モデルのサイズを 4 割以下に、ガウス分布の確率値計算の演算量を 1/10 に削減することができた。これによる認識率の劣化は、ほとんどなかった。

### 3.2.2 単語辞書・言語モデル

旅行会話で使われる文を日本語、英語、それぞれ約 10 万文収集し、そこから 2 単語連鎖、3 単語連鎖の確率値を推定して言語モデルを構築した。収集した文に出現する単語と、一般に利用頻度の高い単語から日本語約 5 万語、英語約 3 万語の単語辞書を構成した。言語モデルについては、出現する単語連鎖のうち、ある程度大きな確率値を持つものだけを保持するようにし、保持されていない単語連鎖の確率値を単語の品詞の連鎖確率で近似した。また、言語モデルのサイズを削減するため、確率値を 1 バイト（256 段階）に量子化して保持した。

### 3.2.3 サーチエンジン

サーチエンジンは音声が入力されると、単語辞書中の単語を組み合わせ得られる候補単語列のうち、音響モデルにより計算される各単語の発音が入力音声波形を出力する確率値と、言語モデルにより計算される単語連鎖の確率値の累積が大きい単語列のみを候補として残して、処理を進めていく。入力音声波形の終端まで計算を行い、最も累積の確率値が大きい単語列を認識結果として出力する。サーチエンジンの演算量を低減するために、音響モデルによる単語確率値の計算を効率化した。すなわち異なる単語で、先頭からの発音記号列が同じ部分に対する確率値の計算は共通化できる。そこで単語辞書の各単語を発音記号列であらわして、先頭から同じ発音記号をマージして木構造にした（木構造単語辞書）。また、定期的に、過去の単語列のうち生き残っている候補から参照されないものを探索しメモリから削除することにより、メモリ使用量を低減した（ガベージコレクション）。

以上のアイデアにより、音声認識モジュールのメモリ使用量は、起動時で日本語と英語

をあわせて約 14MByte，サーチ処理中のワークメモリは約 1MByte 以下を達成し PDA でのリアルタイム動作を実現することができた。

### 3.3 翻訳

今回開発した翻訳モジュールは，日英方向が約 15 万語，英日方向が約 7 万語の翻訳用辞書を持ち，音声認識された旅行会話を対象に，日英・英日の機械翻訳を行う。汎用の日英・英日翻訳文法をベースとして，省略主語の推定，固定表現や熟語など，単語の組み合わせに応じた適切な訳し分け，旅行場面で多く見られる依頼や質問等の表現の強化，口語的表現や定型的表現への対応，丁寧表現の生成など話し言葉対応を強化した。翻訳モジュールは，独自に開発した語彙化ツリーオートマトン文法に基づく語彙規則型の翻訳エンジン[17]を使用した。語彙化ツリーオートマトン文法では，すべての文法規則が辞書に格納されているが，メモリカードに辞書を格納し，入力文中にあらわれた単語に付随する文法規則だけを実行時に内蔵メモリにロードすることで，実用的な大規模な文法規則に対しても，内蔵メモリ消費量の増加を抑制することができた。また，文構造サーチ処理は，単語内で文法規則の適用制御を行う語彙化オートマトンを単語にまたがって共有すること，圧縮統語森の形でサーチの途中結果をバックして曖昧性を圧縮することで，メモリと演算量を削減できた。翻訳部の使用メモリは，起動時で日英・英日あわせて約 8MB，ワークメモリは，入力文に依存するが，通常の旅行会話文では 1~4MB 程度と，PDA 搭載可能なレベルまで低減することができた。

### 3.4 日本語音声合成

音声合成モジュールは，翻訳結果に対して読み付け辞書を用いて読みを与えた後，合成単位ごとの波形データを編集して音声を合成する。今回のシステムは従来のシステム[11][12]に比べてメモリリソースが限られているため，読み付け辞書の構造の見直しを行うことで辞書サイズを 1 / 2 へ削減した。また，合成単位の最適化を行って合成単位数を削減するとともに，波形データの圧縮 / 復号アルゴリズムを搭載することにより波形データサイズを 1 / 10 に削減した。また，通訳用途向けに旅行会話固有の言い回しや，地名，メニュー等の固有名詞を強化した約 23 万語の読み付け辞書を整備した。さらに，翻訳モジュールの日本語生成部で合成用テキストを出力するとともに正しい読み付けに有用な意味，構文情報等を出力し，合成時にそれを参照することで読み精度を高めた。

### 3.5 自動通訳システムの総合性能

上述したように，各モジュールについてコンパクトかつ高精度なアルゴリズムを開発し PDA 上に通訳システムを実装した。このシステムを用いて，日本語・英語それぞれの音声認識について，男性 10 名の計 1800 発話を用いて認識率を評価した。その結果，単語正解精度は日本語 95%，英語 87%であった。これはサーバやパソコンで実現された自動通訳シ

システム[11][12]と同程度の精度である。また、旅行会話例文 500 文を対象に、翻訳精度の主観評価を行った。訳文が統語的に正しく原文の意味も正しく理解できる場合を Good、Good ではないが原文の意味が正しく理解できる場合を Understandable、原文の意味が伝わらない / 誤解して伝わる場合を Bad とする 3 段階に分類した結果、訳文から原文の意味が正しく理解できる率 (Understandable または Good) は日英が 88%、英日が 93% であり、訳文が統語的に正しく原文の意味も正しく理解できる率 (Good) は、日英が 75%、英日が 86% となった。システム全体のメモリサイズは、起動時に約 27MB、ワークメモリとして、数 MB 以下の使用で動作可能であることを確認した。上記の起動時サイズは、日英・英日の双方向通訳に必要なモジュールをすべて起動した場合であり、日英または英日の片方向のみ起動することにより、さらに削減が可能である。処理速度については、発話後 1 秒程度で通訳結果を得ることができた。この結果、旅行会話についてリアルタイムでほぼ意思疎通を図ることが可能となった。

### 3.6 携帯電話用マルチコアプロセッサで動作する自動通訳システム[5]

携帯電話は消費電力などの制約から、PDA に比べ一般にさらに CPU パワーが限られており、とくに大量の処理を必要とする大語彙連続音声認識を携帯電話に搭載するのは困難であった。最近、複数の CPU コアを搭載し、低消費電力と高性能を両立させたマルチコアプロセッサが発表されている[18]。我々は、マルチコア向けの並列大語彙連続音声認識エンジン[19]を開発し、これを適用して、CPU コアを 3 つ搭載する携帯電話用低消費電力マルチコアプロセッサ[20]単体で高速に動作する日英の自動通訳システムを試作した (図 3-7、図 3-8)。図 3-9 に並列音声認識エンジンの構成を示す。分析 + 距離計算、先読み処理、および単語列探索をそれぞれ別の CPU コアに割り当てて、3 ステージのパイプラインとして構成し並列化している。負荷の集中する単語列探索処理の前段として事前に候補を絞り込む先読み処理を導入し、負荷を分散したのが特徴である。これにより、パイプライン各ステージの負荷が平準化され、並列化による加速効果を最大化できた。



図 3-7.携帯電話用プロセッサの開発ボード



図 3-8.自動通訳結果

並列音声認識エンジンを実装し、5万語の日本語認識の速度・精度評価を行った。対照のため、並列化前のエンジン（先読みなし）でも同じ計測を行った。結果を表3-1に示す。単語正解精度は先読みの導入によりわずかに落ちるが、ほぼ同等である。速度は、ここでは、発声長の増加に対する認識処理時間増加の比（リアルタイム比，RT）で定義する。並列エンジンではこの値が1であり、入力音声が遅れなしに認識処理されることがわかる。体感上も、発声終了とほぼ同時に認識結果が得られている。

機械翻訳，音声合成，通訳統合の各モジュールについても，PDA版と同等のものを同プラットフォーム上に実装し，日英旅行会話の自動通訳システムを試作した。機械翻訳についても，おおむね1秒以内で処理が終了することが確認された。本技術開発により，携帯電話への日英旅行会話自動通訳機能の搭載が可能となった。

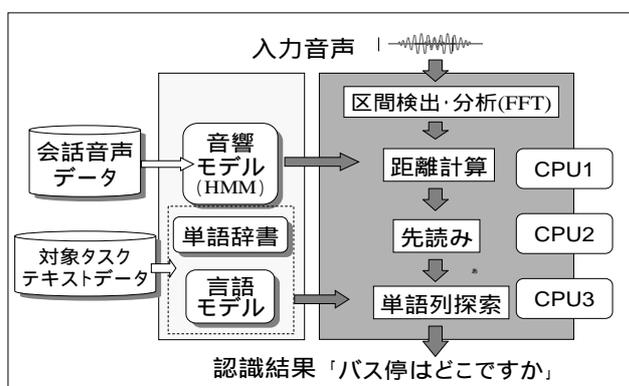


図3-9. 並列大語彙連続音声認識エンジンの構成

表3-1. 音声認識速度および精度

	従来エンジン (1CPU)	並列エンジン (3CPU)
認識速度	2.6 RT	1.0 RT
単語正解 精度(W.A.)	95.8%	95.4%

#### 4. 多言語自動通訳に関する取組み

日本語と英語で実現された通訳プラットフォームが、孤立語言語である中国語に展開可能であれば、言語類的に世界の主要言語をカバーすることができる。多言語自動通訳に関する取り組みとして、図3-1に示した日英通訳システムの英語関連モジュールを中国語関連モジュールに置換えることにより、携帯端末単体で動作する日中双方向の旅行会話通訳システムを開発する[21]。翻訳部と中国語音声合成部は既存製品がありこれを活用した。以下、新規研究開発が必要でかつ、日中通訳実現の鍵となった中国語音声認識、およびPDA上での実装結果について述べる。

##### 4.1 中国語音声認識

我々の音声認識エンジン自体は言語非依存であるが、音響モデルと言語モデルは各言語依存で構築しなければならない。中国語音声認識部を新規開発するにあたり、幅広いアクセントへの対応が課題となった。普通話公用語として定義・教育されているもののその歴史は必ずしも長くはなく、広大な中国全土には様々なアクセントが存在する。そこで、中国全土を表4-1に示す8地域に区分し、それぞれの地域から一定量以上の音声データを収集した。さらにアクセントの無い北京出身話者の音声データも含めて音響モデルを学習することで、この課題に対処した。言語モデルは、総文数約17万文、総単語数約86万語の

旅行会話テキストコーパスを用いて構築した。認識辞書は、テキストコーパスに出現するものをベースに約 3 万 6 千語のセットを作成した。発音(pinyin)は普通話に基づいて付与した。

#### 4.2 評価

中国語音声認識部のシミュレーションによる評価を行った。評価データは旅行会話の発声を表 4-1 と同様の各地域分類別に収集したものをを用いた。男性 101 名分のデータである。各地域別、および全体の認識率とその分散値を図 4-1 に示す。認識率は文字正解精度(C.A.)で評価したが、これは多くの単語が 1 文字から成るためである。結果として、各地域別に多少ばらつきはあるものの、いずれも 8 割以上の単語正解精度が得られた。

#### 4.3 携帯端末上での実装

新規開発した音声認識部と、機械翻訳部、音声合成部とを統合して日中通訳システムとし、市販の PDA(CPU は 400MHz、メモリは 64MB)に実装した。ここでは、音響モデルの自動選択[22]により性別の指定を不要化している。

中日方向の動作確認を行ったところ、処理速度はリアルタイムにやや届かないものの、数秒程度の発声に対して発声終了から 1 秒程度で合成音声の出力が確認できた。日中方向はほぼリアルタイムで動作することを確認した。開発した日中通訳端末を図 4-2 に示す。図 4-2 では、上段に中国語音声認識結果が、下段にその中日翻訳結果が示されている。本システムの開発により、我々の通訳プラットフォームが多言語展開可能であることを実証した。

表 4-1. 地域分類

ID	Area
N1	華北地区
N2	中西部(山西 陝西)
SE1	上海周辺
SW1	華中地区(武漢周辺)
SW2	重慶周辺
S1	福建
S2	広東
S3	江西 湖南地区

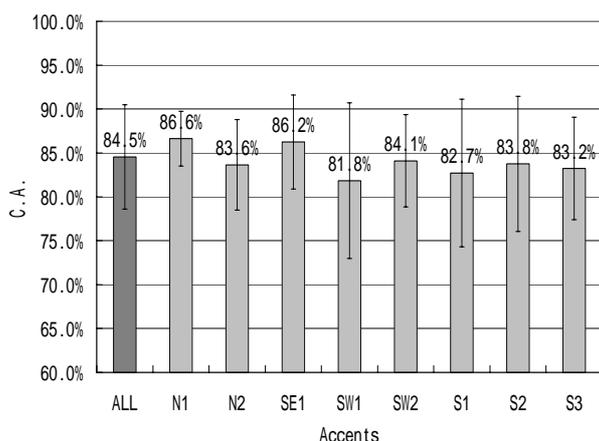


図 4-1. 評価結果



図 4-2. 日中通訳端末

## 5. おわりに

自動通訳の実用化に向けた基本技術として、数万語以上の語彙の発話をリアルタイムに通訳する携帯端末用多言語自動通訳システム技術を開発した。本技術により、世界で初めて日英旅行会話通訳システムを搭載した携帯端末(マルチメディアプレーヤーVoToL)の製品化に成功した。この製品は、コンパクトな音声認識・翻訳・音声合成エンジンと、数万語規模の大語彙辞書を有しており、日本語・英語の旅行会話に関して表現の制約を意識せずに自然に発声した内容を、利用場面に応じて的確に通訳可能である。本技術は、マルチコア向けの並列化アルゴリズムの開発により、携帯電話用プロセッサでPDAと同程度に動作することが確認でき、携帯電話への搭載可能性を実証できた。また、中国語に対応して日中通訳システムを実現することにより、多言語展開可能なプラットフォームであることを確認した。今後は、北京オリンピックにむけ日英中通訳機能を搭載した携帯電話や電子辞書など携帯情報機器の製品化を進める予定である。

本技術開発は、学術的な貢献だけでなくマスコミにも取り上げられ社会的にも大きなインパクトを与えた(付録1)。本技術開発は、現在市場が拡大しつつある電子辞書や情報家電に通訳機能を搭載可能とするもので、日本が世界に先行するものである。また、自動通訳機能を携帯アプリチップに搭載することも可能とし、半導体事業の拡大と競争力強化に大きく寄与するものである。本技術は、スケーラブルな拡張が可能であり、計算リソースの豊富なサーバに適用することにより、サーバ・クライアント型の高性能な自動通訳も可能となる。他の応用として、グローバルなデータベース検索やオーダエントリ、対話エージェントなど様々なシステムに組み込まれ大きく広がることが期待される。

## 6. 謝辞

本技術開発あたり、NEC メディア情報研究所とシステムデバイス研究所、NEC パーソナルプロダクツ、NEC 情報システムズの関係者の多大なるご尽力に感謝します。

## 文 献

- [1] ユニバーサル・コミュニケーション技術に関する調査研究会最終報告書,2005-12, [http://www.soumu.go.jp/s-news/2005/pdf/051215\\_3\\_2.pdf](http://www.soumu.go.jp/s-news/2005/pdf/051215_3_2.pdf).
- [2] 日本電気(株):"日英通訳機能搭載モバイルマルチメディアプレーヤ「VoToL(ヴォトル)」を商品化", プレスリリース(2006-2-14), <http://www.nec.co.jp/press/ja/0602/1401.html>
- [3] 月刊アスキー:"SF から飛び出した自動翻訳が現実になる"(2006-4-6),Vol.30,p102-105
- [4] 日本電気(株):"携帯電話機向け低消費電力プロセッサで動作する日英旅行会話自動通訳ソフトを開発", プレスリリース(2005-10-24), <http://www.nec.co.jp/press/ja/0510/2403.html>
- [5] 山端潔 他,"低消費電力マルチコアプロセッサで動作する日英旅行会話自動通訳システム", 第 68 回情報処理学会全国大会, 4B-2, 2006 年 3 月.

- [6] 日本電気(株):"PDA単体で動作する日中旅行会話通訳ソフトウェアを開発",プレスリリース(2006-1-4), <http://www.nec.co.jp/press/ja/0601/0402.html>
- [7] イノベーション25戦略会議,"イノベーション25"中間とりまとめ,2007-2-26, <http://www.kantei.go.jp/innovation/chukan/chukan.pdf>
- [8] F. Sugaya, T. Takezawa, A. Yokoo and S. Yamamoto, "End-to-end evaluation in ATR-MATRIX: Speech translation system between English and Japanese", *Proc. Eurospeech-99*, pp.2431-2434, 1999.
- [9] A. Lavie, A. Waibel, L. Levin, M. Finke, D. Gates, M. Gavalda, T. Zeppenfeld and P. Zahn, "JANUS: Speech-to-speech translation in multiple languages", *Proc. ICASSP-97*, pp.99-102, 1997.
- [10] W. Wahlster, "Mobile Speech-to-Speech Translation of Spontaneous Dialogs: An Overview of the Final Verbmobil System", In "Verbmobil: Foundations of Speech-to-Speech Translation", ed. W. Wahlster, pp.3-21, Springer, 2000.
- [11] Takao Watanabe, Akitoshi Okumura, Shinsuke Sakai, Kiyoshi Yamabana, Shinichi Doi and Ken Hanazawa, "An Automatic Interpretation System for Travel Conversation", *Proc. ICSLP-2000*, Vol. , pp.444-447, Oct. 2000.
- [12] Akitoshi Okumura, Ken-ichi Iso, Shin-ichi Doi, Kiyoshi Yamabana, Ken Hanazawa and Takao Watanabe, "An Automatic Speech Translation System for Travel Conversation", *Proc. ARPA HLT Workshop*, pp.411-412, 2002. <http://www.mt-archive.info/HLT-2002-Okumura.pdf>
- [13] Lawrence Ricci, "A Militarized PDA Voice-to-Voice Phrase Translator", *Handheld & Wireless Solutions Journal*, Vol. 2, pp.54-55, 2002.
- [14] Ryosuke Isotani, Kiyoshi Yamabana, et al., "An Automatic Speech Translation System on PDAs for Travel Conversation", *Proc. ICMI-2002*, pp.211-216, Oct.2002.
- [15] 日本電気(株):"PDA用に開発した「日英/英日」音声通訳支援ソフトウェアの評価・実証実験を新東京国際空港で実施", プレスリリース(2002-5), <http://www.nec.co.jp/press/ja/0205/3002.html>
- [16] 日本電気(株):"「日英/英日」トラベル通訳端末の評価実験を新東京国際空港で実施", プレスリリース(2004-1-8), <http://www.nec.co.jp/press/ja/0401/0801.html>
- [17] Kiyoshi Yamabana. et al., "Lexicalized Tree Automata-based Grammars for Translating Conversational Texts", *Proc. COLING 2000*, pp.926-932, Aug. 2000
- [18] Torii, S. et al, "A 600MIPS 120mW 70  $\mu$  A Leakage Triple-CPU Mobile Application Processor Chip", *ISSCC 2005, Digest of Technical Papers*, pp.136-137, Feb. 2005.
- [19] 石川晋也 他, "携帯電話用プロセッサで動作する大語彙連続音声認識の並列処理", *FIT2005*, pp.121-122, 2005年9月.
- [20] 枝廣正人 他, "マルチコア向けソフトウェア・プラットフォームを開発し携帯電話機に適用", *日経エレクトロニクス* 2005年3月28日号, pp.125-136, 2005年3月.
- [21] 花沢,他: 携帯端末上で動作する日中通訳システムの開発, *情報処理学会第68回全国大会*, 4B-1,2-1, (2006/3/9)
- [22] 花沢他:"先読みモデル選択による性別非依存音声認識" *音講論集*, 1-P-18, 2004.

< 関連特許 > 20 件

付録 1 . 掲載記事・番組

[1-1] 日経ネット: ” N E C , 携帯電話機向け低消費電力プロセッサで動作する日英旅行会話自動通訳ソフトを開発 ” ,(2005-10-24)

<http://release.nikkei.co.jp/detail.cfm?relID=113895&lindID=1>,

[1-2] 読売新聞:”携帯サイズ 音声翻訳機！” ( 2005-11-24 ) ,

<http://www.yomiuri.co.jp/net/news/20051124nt06.htm>

[1-3] 朝日新聞:” 声を判別 , 日中携帯通訳機 N E C が商品化へ ” , (2006-1-4),

<http://www.asahi.com/digital/av/TKY200601030238.html>

[1-4] .ASCII24: ”NEC ,PDAでリアルタイム通訳が可能な日中旅行会話通訳ソフトを開発 ” ,

<http://ascii24.com/news/i/tech/article/2006/01/04/659813-000.html> , (2006-1-4)

[1-5]中国情報局:” N E C :日中旅行会話自動通訳ソフトの開発に成功” , (2006-1-5)

[http://news.searchchina.ne.jp/disp.cgi?y=2006&d=0105&f=enterprise\\_0105\\_001.shtml](http://news.searchchina.ne.jp/disp.cgi?y=2006&d=0105&f=enterprise_0105_001.shtml),

[1-6] 週刊アスキー:” しゃべると英語に翻訳！！ ” , p9, (2006-2-14)

[1-7] 月刊アスキー:” SF から飛び出した自動翻訳が現実になる ” (2006-4-6),

Vol.30,p102-105

[1-8] ワールドビジネスサテライト:テレビ東京, (2005-12-7),

<http://www.tv-tokyo.co.jp/wbs/2005/12/07/toretama/tt.html>