

PDA で動作する旅行会話向け日英双方向音声翻訳システム

山端 潔 磯谷 亮輔 安藤 真一 花沢 健 石川 晋也

江森 正 磯 健一 服部 浩明 奥村 明俊 渡辺 隆夫

NEC マルチメディア研究所

〒216-8555 川崎市宮前区宮崎 4-1-1

E-mail: {yamabana, isotani, ando, hanazawa, s-ishika, t-emori, iso, hattori, okumura, watanabe}@ccm.cl.nec.co.jp

あらまし メモリ 64MB を搭載した携帯情報端末(PDA)で動作する旅行会話向け日英双方向音声翻訳システムを開発した。本システムは、数万語規模の大規模辞書を有し、旅行会話に関して、自由度の高い発話を日英双方向に音声翻訳することができる。そのため、コンパクトな日本語・英語の大語彙連続音声認識エンジン、語彙化ツリーオートマトン文法に基づいた日英・英日のコンパクト翻訳エンジン、コンパクトな日本語合成エンジンを開発した。本稿では、各エンジンの概要とシステムの構成について述べる。

キーワード 音声翻訳、通訳、音声認識、機械翻訳、音声合成、携帯情報端末、PDA

Japanese/English Bi-directional Speech Translation System on Personal Digital Assistants for Travel Conversation

Kiyoshi YAMABANA Ryosuke ISOTANI Shinichi ANDO

Ken HANAZAWA Shinya ISHIKAWA Tadashi EMORI Kenichi ISO

Hiroaki HATTORI Akitoshi OKUMURA Takao WATANABE

Multimedia Research Laboratories, NEC Corporation

4-1-1 Miyazaki, Miyamae-ku, Kawasaki, 216-8555 Japan

E-mail: {yamabana, isotani, ando, hanazawa, s-ishika, t-emori, iso, hattori, okumura, watanabe}@ccm.cl.nec.co.jp

Abstract We have developed a Japanese/English bi-directional speech translation system for travel conversation that runs on a Personal Digital Assistant with 64MB memory. The system possesses a large vocabulary, and can translate a wide variety of spoken expressions in a travel situation. For this purpose, we have developed a compact large-vocabulary continuous speech recognition engine, a compact translation engine based on the Lexicalized-Tree-AutoMata-based Grammar, and a compact Japanese speech synthesis engine.

Keywords speech translation, speech recognition, machine translation, speech synthesis, PDA

1. はじめに

音声を音声へと翻訳する自動通訳システムは、音声言語処理の有望な応用として、従来から様々な試みが行われてきた[1-3]。我々は、特に旅行会話をターゲットとして、日本語・英語双方の音声翻訳によりコミュニケーションを支援する自動通訳システムを開発してきた[4,5]。これまでに開発したシステムは、大語彙連続音声認識エンジン、語彙規則型翻訳エンジン、拡張波形編集型音声合成エンジンおよび数万語規模の辞書を搭載し、海外旅行の様々な状況における発話を日英双方向に通訳可能である。しかし、通訳処理には大量のCPUパワーとメモリが必要であり、システムの実行には、Pentium II 400MHzクラスのCPUをもつWindowsマシンを必要とした。最近のノートPCは、小型軽量化が目覚しいとはいえ、まだまだ、重量・大きさ・バッテリ寿命などの点で、携帯性には課題が多い。電源OFFからの復帰にも時間がかかるため、利用したいその場面で使いにくくのも問題であった。

最近、携帯情報端末(Personal Digital Assistant, PDA)が急速に普及しつつある。PDAは、常に携帯して手にもって使うことを前提とした設計になっており、重量、大きさ、バッテリの寿命や電源ON時の復帰速度など、ノートPCで課題であった点が十分に考慮されている。自動通訳システムのプラットフォームとして魅力的だが、CPUパワーや搭載メモリ量に限りがあるため、音声検索ベースの定型文翻訳が実現されている程度で[6]、大語彙を持つ自由文の自動通訳システムの搭載は現実的ではなかった。

我々は、このたび、大幅なコンパクト化を行った通訳エンジンの開発により、PDAで動作する自動通訳

システムを開発した。大語彙の辞書を持ち、自由度の高い旅行会話を日英双方向に音声翻訳するという従来からのスペックを保ったまま、必要なハードウェアリソースを大幅に削減したものである。

次節でシステムを概観した後、第3節以降で、今回開発したコンパクトエンジンについて解説する。

2. 開発したシステムの概要

図1に示すように、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。英語音声合成は市販のモジュールを使用した。

システムは、OSとしてPocket PC 2002を搭載したPDA上のアプリケーションソフトとして実現されており、64MB以上のメモリと、StrongARM 206MHz以上のCPU、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。メモリカードは、PDA本体の内蔵メモリを、ファイル格納用ではなく、できるだけプログラム実行用の領域として利用できるよう、実行系やリソースファイルを置くために使用している。翻訳用辞書をはじめとする一部のリソースについては、サイズが大きいため、メモリカード上に置いたまま、実行時に直接アクセスを行う。

マイクとスピーカはPDAに内蔵のものを使用するが、ハードウェア次第で外部接続とすることも可能である。

3. 翻訳モジュール

システムが対象とする旅行会話には、新聞記事のような書き言葉と異なり、話し言葉特有の表現や言い回しが頻出する。これらを適切に翻訳するためには、個別の表現や言い回しに関する知識を豊富に持たなければならない。一方、旅行会話の話題は多岐にわたるため、カバレッジの広さも重要であり、一般的の文法知識もおろそかにできない。前者は具体性の高い事例的な知識であり、後者は抽象度の高いルール的な知識である。これらを統一的に扱える言語知識記述の枠組みとして、文法の語彙化のアプローチにのっとった文法形式である語彙化ツリーオートマトン文法を開発し使用している[7]。

図2に翻訳モジュールの構成を示す。通常のよう

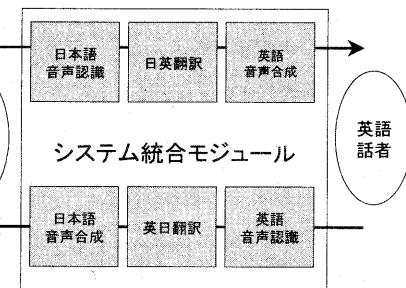


図1: 自動通訳システムの概要

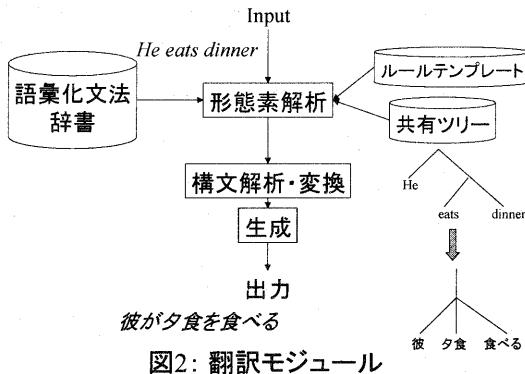


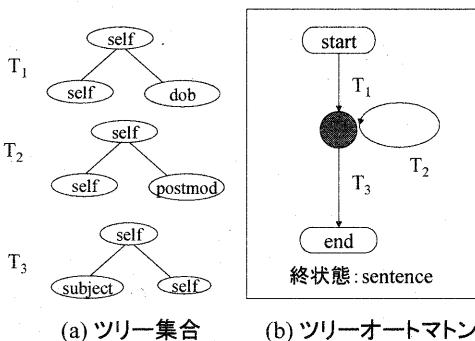
図2: 翻訳モジュール

に文法モジュールと辞書モジュールが分離しておらず、一体となった語彙化文法辞書を有するのが特徴である。また、旅行会話の大規模コーパスを収集し、翻訳用の言語知識ベースの強化に利用している。

翻訳用辞書はサイズが大きいため、PDA の内蔵メモリにロードせずに、メモリカードに置いたまま、翻訳実行時に直接アクセスを行っている。構文解析では、ボトムアップチャート法を拡張したアルゴリズムに基づく幅優先探索を行っているため、統語的曖昧性の削減によるワークメモリの削減と高速化が鍵となる。

3.1. 語彙化ツリーオートマトン文法

語彙化ツリーオートマトン文法 (Lexicalized Tree AutoMata, LTAM 文法) は、すべての構文木が単語辞書内部に格納されている強語彙化文法の流れに沿った文法形式であるが、Lexicalized Tree Adjoining Grammar (LTAG)[8] のような既存の強語彙化文法と異なり、ツリー演算自体がオートマトン



(a) ツリー集合 (b) ツリーオートマトン

図3: eat の辞書内容の例

として単語内に語彙化されているのが特徴である。

各単語は、自分自身をヘッドとする構文木のセットを単語内に保持している。構文木の内部に繰り返しが含まれるため、一般にはこのようなセットは無限集合になる。LTAM 文法では、単語内に、構文木自身ではなく、このような構文木セットを受理するツリー・オートマトンを持たせることによりこの問題を解決している。

具体的には、各単語は、リーフノードの一つが self とマークされた構文木の集合と、これらの構文木を文字列とする文字列を受理するオートマトンを持つ。オートマトンの受理する構文木列は、これらの構文木の頂点 (root) と self を、文字列中の文字の出現順序に応じて、ボトムアップに連結してできる構文木に対応する。

図 3 に、動詞 eat の辞書内容の例を示す。Eat の構文木集合は、図の(a)に示す T_1, T_2, T_3 の 3 つのツリーからなる。また、これらの結合方法を示すオートマトンとして、図の(b)に示す有限状態オートマトンを持つ。このオートマトンは、 $T_1 T_2 * T_3$ を受理するが、これは、図 4 に示す eat をヘッドとするツリーの一般形を受理することにより、図 4 のツリー集合の表現となっている。

すべての単語がこのようなツリーオートマトンを持つが、オートマトンのレベル、および構成要素の部分構文木のレベルで共有を行うことにより、文法記述コストを削減し、実装のコンパクト化を行っている。前者は図 2 のルールテンプレート、後者が共有ツリーに対応する。

LTAM 文法では、ボトムアップチャート法を拡張した形で動的計画法に基づく構文解析アルゴリズムを構成することができ[7]、本エンジンでもこのアルゴリズムを採用している。

3.2. 旅行会話向け翻訳文法と辞書

本システムには、従来から自動通訳システム向けに開発してきた翻訳言語知識ベース(語彙化文法辞書、および付随するリソース)を強化して搭載している。翻訳言語知識ベースは、日英方向が約 15 万語、英日方向が約 7 万語の語彙を持つ。

日英翻訳では、汎用の日英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、口語的な

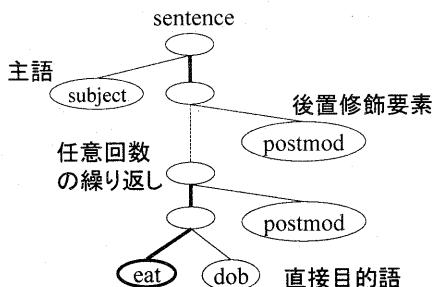


図4:eatをヘッドとするツリーの一般形

文末表現への対応などの話し言葉対応を強化している。英日翻訳では、汎用の英日翻訳文法をベースとして、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を強化している。

3.3. コンパクト翻訳エンジン

翻訳エンジンは、日英、英日方向で共通であり、語彙化ツリーオートマトンを格納した翻訳言語知識ベースを切り替えることにより翻訳方向が切り替わる。

翻訳モジュールは、まず、音声認識モジュールから受け取った認識結果を入力として形態素解析を行い、初期ラティスを作成する。認識エンジンからは、テキスト情報と一緒に、単語区切り、読み、継続長に関する情報が渡され、辞書引きや構文解析の曖昧性解消に用いられる。

初期ラティスに残った各形態素に対して、語彙化文法辞書をはじめとする翻訳言語知識ベースが検索され、各単語に、翻訳用の素性構造と語彙化ツリーオートマトンが付与される。こうして作成された初期チャートに対し、翻訳エンジンは、前述の構文解析アルゴリズムにより left to right の全解探索をベースとした構文解析を実行する。

構文解析の過程では、静的なオートマトン共有メカニズムと、Shared Packed Forest の動的な構築を組み合わせた曖昧性圧縮を行い、メモリソースが少ない状況での動作を可能としている。

解析の過程で、各解に、語彙的、統語的、意味的な情報を総合した尤度が付与される。構文解析終了後、最大尤度を持つ解が選択され、トップダウ

ンに構文生成を行う。語順・訳語の調整と形態素生成の後、最終的な翻訳結果を得る。

翻訳エンジンが使用するメモリは、1)プログラムコードおよびハードコーディングされたデータ、2)立ち上げ時に読み込むリソース類、3)解析・生成で使用するワークメモリの3種に大別できる。1に対してプログラム上のチューニングを行う一方、データ類と2のリソース類をファイルとしてメモリカードに書き出すことにより、立ち上げ時に占有するメモリを削減した。3のワークメモリについては、後に不要となる解析木を早期に発見し枝刈りを行う機構や、素性構造の遅延コピー機構を強化することで削減・高速化している。

本システムで採用した強語彙化文法では、すべての文法知識が単語辞書中に局所化されているため、大きな文法知識データベースをメモリ上に持つ必要がない[9]。この点もコンパクト化における利点である。例えば、文法として文脈自由文法を採用した場合、文法规則を格納したデータベースは、解析中に頻繁にランダムアクセスされるため、オンメモリに持つのがデフォルトとなる。コンパクト化には、データベースのサイズ削減のため文法規則の数を減らす等、訳質とのトレードオフが必要になることが多い。

一方、強語彙化文法では、入力文中にあらわれた単語内に格納された構文木だけがメモリにロードされる。単語個別の文法強化により文法サイズが増加した場合、メモリカード上の辞書はそれだけサイズが増加するが、本体メモリの必要量は、入力文中の単語に起因するぶんが増加するに限られる。そのため、文法強化により訳質向上を図る作業と、エンジンのコンパクト化の間に矛盾が生じにくいという実際上の利点がある。

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、1~4MB程度となった。

4. 音声認識モジュール

隠れマルコフモデル(HMM)で表した音響モデルと統計言語モデルに基づく大語彙連続音声認識方式により、旅行会話を対象に日本語および英語の音声認識を行う。認識結果は、各単語の表記に加え、翻訳の曖昧性解消のための読みおよび継続長を付加して、翻訳部に渡す。

音声認識部は、図5に示すように、音響モデル、言語モデル、辞書、サーチエンジンの各モジュールから構成されている。サーチエンジンは、高速単語列サーチと最適単語列選択の2パスからなる。サーチエンジンは日本語と英語で共通で、音響モデル、言語モデル、辞書は、日本語用と英語用にそれぞれ用意されている。

PDA搭載にあたっては、モデルのコンパクト化とサーチの処理量・ワークメモリの削減がポイントであった。モデルのコンパクト化は、使用メモリ削減のみならず、計算量削減にも寄与する。

4.1. 音響モデル

音響モデルは、混合ガウス分布を出力確率とするトライフォンHMMで、決定木による状態クラスタリング法[10]を用いて状態を共有化している。メモリと計算量削減のために、1)MDL原理に基づくガウス数削減処理(GR-MDL)[11]、2)ガウス分布の対角共分散行列の共有化(GT-DCM)、3)出力確率の高速計算(GS-tree)[12]の3つの手法を用いている。1のGR-MDLにより、冗長なガウス数を削減し、モデルのサイズが削減される。2のGT-DCMにより、モデルのサイズとガウス分布確率計算量をともに削減できる。3のGS-treeは、ガウス分布を木構造状にクラスタリングして、フレーム同期ビームサーチにおけるガウス分布確率計算を大幅に高速化できる。これらの施策により、音響モデルのサイズを大幅にコンパクト化することができた。現時点での音響モデルのメモリサイズは0.6~0.7MB程度である。以下、これらの手法について述べる。

1) MDL原理に基づくガウス数削減処理(GR-MDL)

認識精度の劣化を抑えつつモデルのサイズを縮小するには、出力確率の混合ガウス分布から過剰なガウス分布数を削減するのが有効である。我々は、

最小記述長(Minimum Description Length, MDL)原理を用いてガウス数の効率的削減を行った。

初期モデルとしては、十分な量の学習データで学習された大きな分布数を持つHMMを用いる。次に、状態ごとのガウス分布を、木構造の形にクラスタリングする。このガウス分布木構造に対して、MDL原理を用いて、適切なガウス分布数を選択することにより、精度劣化を最小限にとどめたガウス数削減処理を行うことができる。

2) ガウス分布の対角共分散行列の共有化(GT-DCM)

LDA変換はガウス分布の対角共分散行列を白色化する傾向があるため[13]、精度劣化を少量にとどめつつ、行列をグローバルに共有することができる。これにより、モデルの格納に必要なサイズを約5割削減することができる。また、ガウス分布確率値計算を、入力特徴ベクトルとガウス分布平均ベクトルのユークリッド距離計算へと還元できるため、計算量も削減できる。

3) 出力確率の高速計算法(GS-tree)

フレーム同期サーチにおけるガウス分布確率値計算をさらに高速化するために、ガウス分布を木構造の形に構成している。リーフノードはHMMの各状態のガウス分布に対応し、中間ノードは下位のノードを被覆するガウス分布となっている。入力特徴ベクトルに対して、ルートノードからリーフノードに向かってガウス分布確率値計算を行うことで、確率値の大きいノードから一定個数をリーフノードへ向かって計算することにより計算量を削減している。現在は3層の木構造を利用しているが、ガウス分布の計算量をトータル

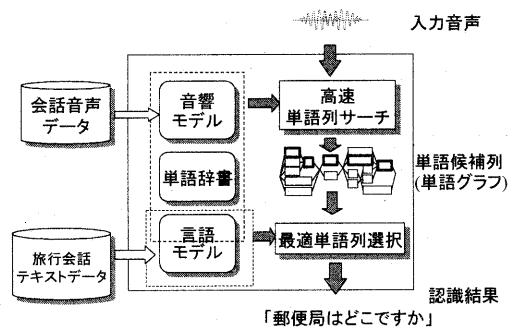


図5: 音声認識モジュール

で全体の1割以下に削減することができた。

4.2. 言語モデル

扱うドメインが旅行会話に限定されているとは言え、そこに現れる語彙や言い回しのバリエーションは多く、それらをカバーする文法を人手で記述するのは現実的でない。そこで、言語モデルとしては統計ベースのモデルである単語 N-gram モデルを用いている。

統計言語モデルの構築には学習用に大規模なテキストコーパスが必要で、一般には新聞記事など入手が容易なコーパスが用いられる。しかし、旅行会話は、口語的な表現を多く含むなど、新聞記事等に現れる文とは表現や語彙が異なる。そこで、日本語および英語の旅行会話の大規模コーパスを収集し、それを言語モデルの学習に用いた。収集したコーパスは、ホテル、レストラン、ショッピング、移動など、旅行中のさまざまな場面での会話からなる、日英それぞれ 10 万文規模のテキストコーパスである。

学習コーパスに対し、日本語については形態素解析プログラムを用いて単語（形態素）に分割、英語についてはスペース区切りで単語に分割し、単語 bigram・単語 trigram を学習した。さらに、単語 N-gram のスムージングにクラス N-gram を用いた。クラスは品詞をベースとし、一部意味的なクラスを併用した。意味クラスは、旅行会話向けに定義したクラスを入手で付与した。

言語モデルサイズの削減のために、N-gram の確率の対数値を 256 段階（1 バイト）に量子化して格納している。予備的な実験では、量子化による認識率の劣化は観察されなかった。

認識辞書は、言語モデル学習コーパスに出現する単語と一般に利用頻度の高い単語から構成した。語彙サイズは、日本語が約 5 万語、英語が約 2 万語である。

4.3. サーチエンジン

サーチエンジンは、入力された音声波形をもとに、音響モデルと言語モデル・辞書を用いて、発話単語列を推定する。サーチは 2 パスで構成されており、1 パス目では、音響モデルと単語 bigram を用いて認識候補単語からなるワードグラフを得る。2 パス目では、1 パス目で得たワードグラフ全体に単語 trigram

を適用してリスクアリングを行い、最大尤度のパスを認識結果として出力する。

1 パス目は、木構造辞書を用いたフレーム同期ツリーサーチで、各フレームごとに単語終端に到達した仮説を単語終端テーブルに書き出す。コンパクト化のため、木構造辞書は音素をノードとし、各ノード 1 バイトで表現されている。各ノードはサーチ時に動的にトライフォンに展開される。処理量削減のため、言語スコア計算結果の再利用[14]を行っている。また、一定フレーム間隔毎に単語終端テーブルのガベージコレクションを行うことにより、発声長に依存したワークメモリの増加を抑えている。具体的には、そのフレームで生き残っている仮説からバックトラックを行ったときに到達しない仮説を単語終端テーブルから削除し、そのメモリを再利用できるようにする。

モデルおよびサーチ部のコンパクト化の結果、音声認識モジュールの使用メモリは、起動時で日英合わせて約 14MB、また、処理中のワークメモリは 1MB 以下となった。

5. 日本語音声合成モジュール

音声合成モジュールは波形編集方式を大規模なデータベースを用いて合成単位を拡張した拡張波形編集方式を用いている。波形編集方式はあらかじめ音響分析により基本周波数の影響を取り除いた 1 ピッチ波形を抽出しておき、合成時に所望する周波数、時間長になるようピッチ波形を編集する手法である。この手法は音響分析に高精度なアルゴリズムを用いることができるため声道情報と励振情報の分離性が高く高品質な合成音が得られる、合成時に各ピッチ波形を所望の周波数に対応する時間間隔で重ね合わせるだけでよいため合成時の計算量が少ないと特徴がある。

今回のシステムは従来のシステム[4][5]に比べてメモ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、合成単位セットの最適化を行いシステムで持つ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。

6. 統合モジュール

統合モジュールは、音声認識、翻訳、音声合成の各モジュールを統合して通訳システムとしての処理を実行する。

認識→翻訳の間では、単語の表記に加えて、単語区切り、単語の読み、継続時間長の情報が渡される。これらの情報は、読みに曖昧性がある場合の曖昧性解消(e.g. 行った(いった)／行った(おこなつ))や統語的曖昧性の解消に利用している。また、英日翻訳モジュールから日本語音声合成モジュールへは、単語の読み情報が渡され、読みの曖昧な語(e.g. 辛い:からい／つらい)の読み分けに用いられる。

翻訳方向を遅滞なく切り替えるには、日本語・英語の認識エンジン、日英・英日の翻訳エンジン、日本語・英語の合成エンジンをすべて同時に立ち上げておく必要がある。メモリの割り当てが厳しいが、特に動作時に必要となるワークメモリを確保するために、各エンジンがスタティックに確保するメモリを削減するとともに、ワークメモリをこまめに解放して共用する等の制御を行っている。

PDA の OS である Pocket PC 2002 (Windows CE 3.0)では、プロセスあたり使用できるアドレス空間の上限が 32MB に制限されており、実メモリ量に対する制限以上に、論理アドレス空間やプロセス・DLL の配置に対する制約が厳しい。プロセスの配置にあたっては、これらの制約を十分に考慮した設計と実装が必要であった。

7. 考察

日本語・英語それぞれについて、シミュレーション実験により、男性 10 名の計 1800 発話を用いて認識率を評価した結果、単語正解精度は日本語 95%、英語 87% であった。これは従来の自動通訳システム [4][5]と同程度の精度である。

旅行会話コーパスからランダムに抽出した 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 90% 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 66%、英日が 74% となった。この精度は、従来のシステム [4][5]との比較で同等以上となっている。

システム全体のメモリサイズは、起動時に約 27MB となった。これに、各モジュールのワークメモリが必要である。短い文では 1MB 程度、長めの文では、主に翻訳部のワークとして数 MB を使用する。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、従来の通訳システムを Pentium II 400MHz のマシンで動作させた場合と比べて低下が見られる。今回開発した通訳システムを市販の PDA (StrongARM 206MHz、メモリ 64MB) で動作させた場合、発話後、認識結果が表示されるまで多少待たれる状況であり、リアルタイム動作にはまだ距離がある。翻訳も、入力文によっては数秒以上かかるものがあるなど、さらなる高速化が望ましい。

その他、今回はハードウェアの都合上、内蔵マイクを使用したが、実環境下で利用するのに適切なマイクの選択と雑音対策は、別途検討が必要なテーマである。また、わかりやすく誤り訂正が容易なユーザインターフェースの設計など、アプリケーションとしての利用方法についても検討を進めたい。

8. 終わりに

PDA で動作する旅行会話向け日英双方向自動通訳システムを開発した。このシステムは、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持ち、日本語・英語の旅行会話に関して、自由な発話を日英双方向に音声翻訳できる。今後、さらに性能評価と改良を進めて、異言語間コミュニケーションの支援ツールとして実現を目指したい。

文 献

- [1] 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.
- [2] A. Lavie, A. Waibel, L. Levin, M. Finke, D. Gates, M. Gavvalda, T. Zeppenfeld and P. Zahn, "JANUS-III : Speech-to-speech translation in multiple languages", *Proc. ICASSP-97*, pp.99-102, 1997.
- [3] 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.
- [4] Takao Watanabe, Akitoshi Okumura, Shinsuke Sakai, Kiyoshi Yamabana, Shinichi Doi and Ken Hanazawa, "An Automatic Interpretation System for Travel Conversation", *Proc. ICSLP-2000*, Vol. IV, pp.444-447, Oct. 2000.
- [5] 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*, (to be published), 2002.
- [6] Lawrence Ricci, "A Militarized PDA Voice-to-Voice Phrase Translator", *Handheld & Wireless Solutions Journal*, Vol. 2, pp.54-55, 2002.
- [7] Kiyoshi Yamabana, Shinichi Ando and Kiyomi Mimura, "Lexicalized Tree Automata-based Grammars for Translating Conversational Texts", *Proc. COLING 2000*, pp.926-932, Aug. 2000.
- [8] Aravind K. Joshi and Yves Schabes, "Tree-Adjoining Grammars and Lexicalized Grammars", In *Tree Automata and Languages*, M. Nivat and A. Podolski, ed., Elsevier Science Publishers B.V., pp.409-431, 1992.
- [9] Yves Schabes, Anne Abeillé and Aravind K. Joshi, "Parsing Strategies with 'Lexicalized' Grammars", *Proc. COLING'88*, pp.578-583, 1988.
- [10] S.J. Young, J.J. Odell and P.C. Woodland, "Tree-based tying for high accuracy acoustic modeling", *Proc. Human Language Technology*, pp. 307-312, 1994.
- [11] Koichi Shinoda and Ken-ichi Iso, "Efficient Reduction of Gaussian Components Using MDL Criterion for HMM-based Speech Recognition", *Proc. ICASSP-2002*, pp.869-872, May 2002.
- [12] T. Watanabe, K. Shinoda, K. Takagi and K. Iso, "High Speed Speech Recognition Using Tree-Structured Probability Density Function", *Proc. ICASSP-95*, pp. 556-559, 1995.
- [13] S.A. Zahorian, D. Qian and A.J. Jagharghi, "Acoustic-Phonetic Transformations for Improved Speaker-independent Isolated Word Recognition", *Proc. ICASSP-91*, pp. 561-546, 1991.
- [14] 石川晋也, 江森正, 三木清一, 大西祥史, 磯谷亮輔, 磯健一, "コンパクトなディクテーションの開発", 日本音響学会2002年春季研究発表会講演論文集, pp. 165-166, 2002.