

4B-2

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

山端 潔 石川 晋也 花沢 健 長田 誠也
土井 伸一 磯谷 亮輔 服部 浩明 奥村 明俊

NEC メディア情報研究所

1. はじめに

社会のグローバル化と海外旅行者数の増加を背景に、異言語コミュニケーションの必要な場面は着実に増加している。携帯電話は、仕様のグローバル化とともに、海外に持ち出して使用する機会が今後大きく増加することが予想され、自動通訳機能搭載に対する潜在ニーズは大きいと思われる。我々は、旅行会話を対象として、小型の携帯機器で動作するコンパクトな自動通訳システムの開発を進めている。しかし、入り口となる大語彙連続音声認識は大量の処理を必要とするため、従来CPU パワーの限られた小型機器、特に携帯電話への搭載は困難であった。

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[1]。我々は、マルチコア向けの並列大語彙連続音声認識を開発し、これを適用して、携帯電話機用低消費電力プロセッサ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ので報告する。

2. システムの概要

試作した自動通訳システムは、図1に示すように、日本語・英語の大語彙連続音声認識エンジン、日英・英日機械翻訳エンジン、音声合成エンジンを有し、これらを通訳統合モジュールが制御する。英語音声合成のみ市販モジュールを使用する。日本語5万語、英語3万語の音声認識辞書を持ち、日本語・英語の旅行会話を音声から音声へと翻訳する機能を有する。

試作のプラットフォームとして、携帯電話機向けマルチコアプロセッサ[2]の開発ボードを使用した。プロセッサは、オンチップに CPU コア ARM926 を3つ、DSP コア SPXK602 を1つ搭載する。各 CPU コアが独立のキャッシュを持ち、独立の OS が走る非対称型のアーキテクチャを有する。CPU の最大クロックは 200MHz であり、電源電圧+1.2V における H.264 の復号

化処理時に消費電力 120mW を実現している。開発ボードの OS は Linux である。

自動通訳システムは、3つの CPU コア上に分散して配置しており、通訳処理はすべてこれらの CPU コア上で行われる。

図2に試作システムにおける自動通訳結果の画面表示例を示す。ボード上の LCD 画面の上半分に音声認識結果の日本語文が、下半分に翻訳結果の英文が表示されている。

3. 各モジュールの詳細

3.1. 音声認識モジュール

マルチコアプロセッサ向けの並列大語彙連続音声認識エンジン[3]を用いて、旅行会話を対象として、日本語および英語の大語彙連続音声認識を行う。音響モデルは triphone の混合ガウス分布 HMM、言語モデルは n-gram 統計言語モデルである。デコーダは木構造辞書を用いた 1 パスのフレーム同期ビームサーチである。

図3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、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。距離計算と単語列探索の間に、音響レベルの前処理として先読み処理を導入したのが特徴である。先読み処理がない場合は、負荷が単語列探索に集中するため、パイプライン処理による加速は期待できない。先読み処理の導入により、単語列探索

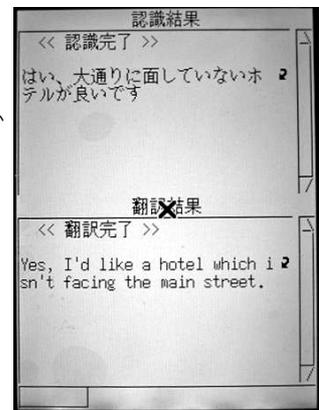


図2: 自動通訳画面例

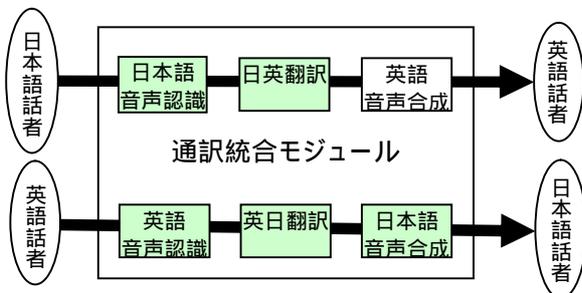


図1: 自動通訳システムの構成

Japanese/English Speech Translation System on Low-Power Multicore Processor
Kiyoshi Yamabana, Shin-ya Ishikawa, Ken Hanazawa, Seiya Osada, Shinichi Doi, Ryosuke Isotani, Hiroaki Hattori, Akitoshi Okumura
Media and Information Research Laboratories, NEC

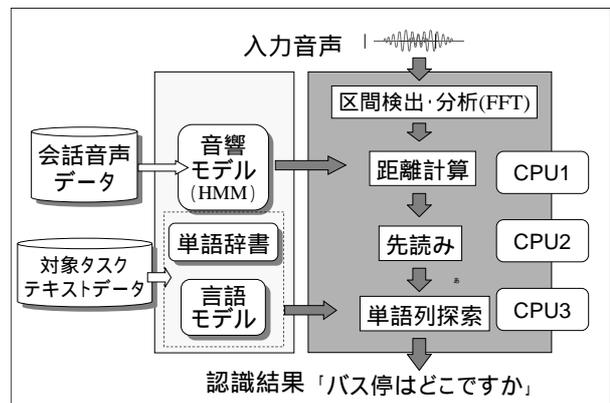


図3: 並列大語彙連続音声認識エンジンの構成

における仮説数を効率的に削減できるので、パイプライン各ステージの負荷が平準化し、並列化による加速が期待できる。先読み処理は時間方向逆向きの処理であるため、音声はある一定の長さを持つ区間(数百 ms オーダー)に分割され、これを単位にパイプラインに投入される。

音響モデル・言語モデルと辞書は、PDA 向け自動通訳システム[4]をベースに強化したものを使用している。

3.2. 機械翻訳モジュール

語彙規則型のコンパクト機械翻訳エンジン[5]により、旅行会話の日英・英日の機械翻訳を行う。PDA 向けに開発した翻訳エンジンをベースに、Linux 対応とコンパクト化・高速化を進めて使用している。機械翻訳エンジンは単一の CPU コアで動作している。従来テーブルデータとして保持していた言語知識を、その適用手順を含めてプログラムの形にプレコンパイルする等の実装上の工夫を含めてコンパクト化・高速化を進めた。辞書は PDA 向け自動通訳システム[4]をベースに強化したものを使用している。

3.3. 通訳統合モジュール

音声認識・機械翻訳・音声合成の各モジュールを統合し、通訳システムとしての処理を実行する。音声認識・機械翻訳・音声合成の各エンジンは、独立のプロセスとして実装されており、統合モジュールのプロセスとは、ソケットを介したプロセス間通信でやりとりを行う。プロセスは複数の CPU コアに配置されるため、CPU コアをまたいでプロセス間通信を行う必要があるが、試作プラットフォーム上に実装された OS Wrapper[2]の利用により、マルチコアを特に意識しない構成となっている。

4. 評価

本節では、音声認識、機械翻訳の各モジュールに対する精度・速度の評価結果をまとめる。

4.1. 音声認識

並列音声認識エンジンを実装し、日本語認識の速度評価を行った。認識速度は、旅行会話の日本語読み上げ文発声 50 発声を用いて、評価ボード上で wav ファイル入力により評価した。wav ファイル入力にあたっては、音声データの取り込み速度が実発声と同じになるよう、タイマーによる制御を行っている。測定時の CPU クロックは 150MHz とした。また、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。

日本語・英語の音声認識精度については、同じく旅行会話の読み上げ文を用いてオフラインで評価を行った。日本語は男性 5 名による各 200 発声、英語は男性 32 名による各 180 発声を用いた。音響モデルは日本語、英語とも男性不特定話者、辞書・言語モデルは自動通訳システム[4]と基本的に同じものを使用した。

表 1 に認識速度と精度の評価結果を示す。単語正解精度(W.A.) 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

表 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)
日本語	認識速度	2.6 RT	1.0 RT
	単語正解 精度(W.A.)	95.8%	95.4%
英語	単語正解 精度(W.A.)		92.0%

(日本語認識)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ていることがわかる。値1はこの定義による速度の上限であり、CPU パワーにはまだ余裕がある可能性がある。方式的には、先読み区間長の程度(数百 ms)の認識終了遅れがありうるが、体感上とくに問題なく、発声終了とほぼ同時に認識結果が得られている。

4.2. 機械翻訳

旅行会話テキストコーパスからランダムに抽出した日本語、英語各 500 文を対象に、翻訳精度の主観評価を行った。評価者は、翻訳システム開発に関わっていないバイリンガル話者 2 名である。評価者は、原文と訳文の組を提示され、これを a~d の4段階に分類した。段階は以下の4つである。a) Natural: 訳文が、原文と照らし合わせて自然で適切である場合; b) Good: 言い回しなどに不自然さはあるが統語的に正しく、原文の意味も正しく理解できる場合; c) Understandable: 文法的に誤りがあるなど Good ではないが原文の意味が伝わる場合; d) Bad: 原文の意味が理解できない/誤って伝わる場合。表 2 にこの基準による主観評価結果を示す。数値は評価者 2 名の平均である。

また、日本語の旅行会話コーパスからランダムに抽出した 46 文に対して、評価ボード上で日英翻訳速度を実測した。計測時の CPU クロックは 200MHz である。一文あたりの翻訳時間の平均は 0.99 秒、中央値は 0.61 秒であった。各文の平均文字数は 11.5 文字である。一部の文で翻訳に時間がかかるため、平均値が少し高めに出ているが、多くの文は 1 秒以内で翻訳が終了する。

表 2 : 機械翻訳精度

	Good(b)以上	Understandable(c) 以上
日英翻訳	74.9%	87.4%
英日翻訳	85.8%	93.1%

5. おわりに

マルチコアプロセッサ向けの並列大語彙連続音声認識を開発し、携帯電話機向けの低消費電力マルチコアプロセッサ上で、日英旅行会話の自動通訳システムを試作した。このシステムは、コンパクトな音声認識・機械翻訳・音声合成エンジンと、数万語規模の辞書を持ち、日本語・英語の旅行会話を高速に音声翻訳することが可能である。本試作により、携帯電話機クラスの小型機器に、日英の旅行会話自動通訳機能を搭載する可能性を示した。

参考文献

- [1]Torii, S. et al, "A 600MIPS 120mW 70 μ A Leakage Triple-CPU Mobile Application Processor Chip", ISSCC 2005, Digest of Technical Papers, pp.136-137, Feb. 2005.
- [2]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.
- [3]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.
- [4]山端潔 他, "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", 情処研報, 2002-NL-150-9, 2002 年 7 月
- [5]山端潔 他, "語彙化されたツリーオートマトンに基づく会話文翻訳システム", 言語処理学会第 6 回年次大会講演論文集, pp.264-267, 2000 年 3 月.