

大語彙連続音声認識ソフトウェア Julius のマイコンへの実装

小窪浩明* 李晃伸† 河原達也‡ 鹿野清宏§

日立・中研* 名工大・情報† 京大・メディア‡ 奈良先端大・情報§

1. はじめに

Julius[1]は PC 向けの大語彙連続音声認識のフリーソフトウェアとして、音声研究者の間で広く普及している。マイコンを搭載した組み込み機器でも連続音声認識プログラムを利用できれば、携帯電話やカーナビゲーションなどの民生機器への音声インタフェース導入が容易になると期待される。本報告は、文科省「e-Society 基盤ソフトウェアの総合開発プロジェクト」の一環として開始された「ユーザ負担のない話者・環境適応性を実現する自然な音声対話処理技術」プロジェクト[2]の中で開発したマイコン向け大語彙音声認識プログラム Embedded Julius[3]について述べる。

2. マイコン用音声認識の課題

PC に比べて計算リソースの制約が大きいマイコン向けに音声認識システムを開発する場合、計算量と消費メモリ量の削減は必須である。CMU は、PC 向けに開発した連続音声認識ソフトウェア SPHINX を Strong ARM (235MIPS)向けに改変した POCKET SPHINX[4]を開発した。浮動少数点演算モジュールを持たない Strong ARM のために、浮動少数点演算から固定少数点演算処理への修正などを実施することにより、マイコン上で連続音声認識(語彙数 1,000 語)の実時間動作を実現している。ただし、PC 版のベースラインシステムと比較した場合、計算量削減に伴う認識性能の劣化は不可避である。

本研究では、PC 上での研究開発プラットフォームとして広く普及している Julius の資産(音響モデル、言語モデル、最適パラメータ値など)をマイコン上でも同様に活用できることを目標としている。そのため、認識性能劣化を伴う基本アルゴリズムへの近似計算導入の優先度を下げ、メモリアクセスのオーバーヘッド解消などプログラミングの最適化に基づくマイコン上での処理高速化に重点をおいた開発方針をとった。

Implementation of large vocabulary continuous speech recognition software "Julius" on microprocessor.

*Hiroaki Kokubo, Hitachi

†Akinobu Lee, Nagoya Institute of Technology

‡Tatsuya Kawahara, Kyoto University

§Shikano Kiyohiro, Nara Institute of Science and Technology

3. Embedded Julius

3.1 開発プラットフォーム

Embedded Julius の開発には SH マイコン(SH-4A)をコアとする組み込み機器の開発プラットフォーム T-Engine と統合開発環境 HEW(High-performance Embedded Workshop)を用いた。

T-Engine の仕様を Table 1 に、写真を Fig. 1 に示す。T-Engine は eTRON アーキテクチャに対応した共通開発ボードであり、外部インタフェースを介して CF(Compact Flash)カードへのアクセスが可能である。ターゲット CPU の SH-4A は(株)ルネサス テクノロジが開発した組み込み用途向け 32bit RISC プロセッサであり、浮動小数点命令、キャッシュ操作命令、4 次元ベクトル演算をサポートする高性能なマイコンである。

HEW はエディタ、コンパイラ、エミュレータといった組み込みアプリケーションに必要なツールを統合的に扱える統合開発環境である。開発言語に C を用いたことで Julius のソースコードを大幅に書き換えることなくマイコンへの導入が可能となった。ベースとなる PC 版 Julius のバージョンは v. 3.5.3 である。

Table 1 Profile of T-Engine board

CPU	SH-4A (720MIPS 2.8GFLOPS)
Operating Freq.	Internal:400MHz External:100MHz
User RAM	128MByte
OS	T-Kernel
Audio CODEC	16kHz, 16kbit
LDC	TFT color, 240x320
size/power	120mm x 75mm / DC 5.6V

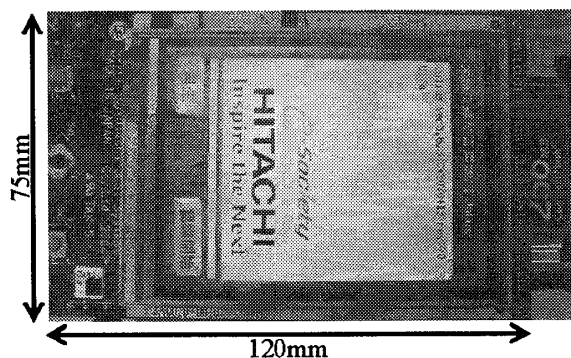


Fig. 1 T-Engine board

4. マイコン上での処理高速化

T-engine は 128MByte のユーザメモリを搭載しているため、消費メモリに対する制約は小さい。ここでは処理高速化に焦点を絞って説明する。

4.1 メモリ管理の効率化

予備実験の結果、マイコン上での処理速度のボトルネックはデータアクセスにあることがわかった。特に、外部メモリ (CF カード) からのデータアクセス速度の遅さとメモリ領域の確保と開放が頻繁に繰り返されることにより生じるメモリフラグメントの影響は無視できないほど大きい。そこで、CF カードに格納されている音響モデル、言語モデルをバイナリ化すると共に、プログラム起動時にユーザメモリに展開ように修正を図った。また、仮説探索時に繰り返される仮説ノード生成/消滅に専用のメモリ領域を事前に割り当てることでメモリフラグメントの発生を回避した。

4.2 高速化オプションの導入

Julius では全計算量のうち 2/3 は音響尤度計算に費やされている。このため、PC 版 Julius では音響尤度計算の高速化オプションとして GMS (Gaussian Mixture Selection) [5] が実装されている。GMS では、音響モデルの尤度計算に簡易モデルを用いた予備選択を行い、音響尤度への貢献度の高いモデル (混合ガウス分布) 以外の尤度計算には簡易モデルで計算した値で代用することで計算量の削減を図っている。本開発では、アルゴリズムの見直しを行い、オリジナルの GMS に対して計算すべきガウス分布数をさらに削減する修正を加えた [3]。

5. 性能評価

T-engine 上に実装した Embedded Julius を用いた評価実験について報告する。実験では、CF カードに格納した音声データファイル (男女各 30 発話) を逐次読み出しながら認識処理を実行する。評価タスクは新聞記事読み上げ音声のディクテーションとし、辞書サイズ 5,000 語と 20,000 語の 2 種類の言語モデルを用意した。これらの条件に対して、それぞれ高速化オプションの GMS を導入しない場合と導入した場合とについて、消費メモリ、単語認識精度、実時間比 (xRT) を求めた。実時間比は処理時間を音声長で正規化した値であり、実時間動作の要件は実時間比が 1.0 を下回ることである。ただし、実装上の問題として発声終了を検出するための待ち時間が必要なため、実時間比 1.3 程度まではほぼ実時間処理とみなしても差し支えない。

実験結果を Table 2 に示す。GMS では混合ガウス分布の予備選択に用いる簡易音響モデルが追加されるため、消費メモリ量は約 0.5MB 増加する。単語認識精度は GMS の有無で大きな差はなく、GMS の導入に伴う近似処理の影響は小さい。5,000 語の条件での処理速度は、GMS を導入しない場合で 1.02xRT、GMS を導入した場合で 0.73xRT とマイコン上での実時間動作を達成した。語彙数 20,000 語の条件では、GMS を導入することによって処理速度の改善が図られ、実用上ではほぼ実時間とみなせる 1.24xRT で動作した。

Table 2 Performance of Embedded Julius

vocabulary size	GMS	memory size	word accuracy	x RT
5,000	off	45.2MB	89.1%	1.02
	on	45.7MB	89.7%	0.73
20,000	off	88.0MB	91.8%	1.53
	on	88.5MB	91.8%	1.24

6. まとめ

PC 向け大語彙連続音声認識ソフトウェア Julius を SH-4A T-Engine に実装し、評価実験を行った。語彙数 5,000 語のディクテーションタスクで実時間動作 (0.73xRT) を確認し、語彙数 20,000 語の条件においても、ほぼ実時間の処理速度 (1.24xRT) で動作することを示した。

謝辞

本研究は、文科省「e-Society 基盤ソフトウェアの統合開発プロジェクト」により実施された。

参考文献

- [1] 李他、単語トレリスインデックスを用いた段階的探索による大語彙音声認識、信学論, Vol. J82-DII, No1, pp.1-9, 1999.
- [2] 鹿野他、人にやさしい音声認識・合成基盤ソフトウェアの研究開発、信学技報 SP2003-167, 2003.
- [3] Kokubo et.al. "Embedded Julius: continuous speech recognition software for microprocessor," MMSP, pp.378-381, 2006.
- [4] Huggins-Daines et. al. "POCKETSPHINX: A free real-time continuous speech recognition system for hand-held devices," ICASSP, pp.185-188, 2006.
- [5] 李他、音素環境独立 HMM を用いた混合ガウス分布選択による音響尤度計算の削減、情報処理論文誌, vol. 32 No. 7, pp. 2214-2221, 2002.