

音声認識の組み込みシステム搭載での課題

The challenge in Embedded Systems with Speech Recognition

佐藤 隆英 小野寺 透 畑岡 信夫(東北工業大学)

1. まえがき

ユビキタスマバイル情報化社会が近づいており、携帯電話やカーナビゲーションシステム等で自然な音声インターフェースが必要になってきた。

我々は、組み込み型音声インターフェースの構築を目標とし研究を行っている。文科省 e-Society 基盤ソフトウェアの総合開発プロジェクトに「組み込み型連続音声認識プログラムの開発」があり、その継続として T-Engine での音声認識搭載を行っている。

本発表では、T-Engine へ開発環境の整理と音声認識の組み込みシステム搭載での課題について、連続音声認識ソフトウェア Julius の T-Engine 搭載を行い検討した。

2. 連続音声認識と T-Engine

2.1 組み込み型音声認識

組み込み型の音声認識の研究開発としては、文科省 e-Society の基盤ソフトウェア開発では、世界最高水準の高度情報通信システム形式のための鍵となるソフトウェア開発を実現させ、いつでもどこでも誰でも安心して参加できる IT 社会を構築することを目的としたプロジェクトを行っている^[1]。

NEC は 8 つのマイクを搭載し時間差から音声が入る方向を検出する、新型パーソナルロボット「PaPeRo-mini」を開発し発表した^[2]。

2.2 連続音声認識ソフト Julius

Julius は、標準的な音響モデル、言語モデル、認識エンジン、および形態素解析・読み付与ツールから構成され、一般に無償で公開されている^[3]。

音声認識ソフトウェア Julius は、音声認識の研究開発用プラットフォームとして供するために、以下のコンセプトで設計されている。

①汎用性・標準的なインターフェース

音響モデル、言語モデルとの分離を図り、標準的なインターフェースで種々のモデルを扱えるよう汎用的なエンジンを目指している。

† 東北工業大学

〒982-8577 仙台市太白区八木山 香澄町 35-1

②高水準の大語彙連続音声認識アルゴリズム

世の中のリファレンス・ベースラインとなることを目指し、オフラインでは最高水準に近い音声認識制度を達成する。さらに、PC で実時間処理を可能とする。

③オープン性・ポータビリティ

デコーディングの設定が容易にできる。また、ソースコードがオープンである。利用に関する制限は商用も含めて殆ど無く、移植・改変も自由である。

2.3 開発プラットフォーム T-Engine

T-Engine 組み込み機器向けのリアルタイム OS として最大のシェアを持つ ITRON をベースにした開発プラットフォームである。

規格化されたハードウェア(T-Engine ボード)と、ITRON をベースとした標準リアルタイム OS(T-Kernel)からなる。ハードウェア規格はパソコン等と違って CPU 非依存の緩やかなものとなっている。

ITRON は、東京大学の坂村健氏によって創造された「TRON プロジェクト*」から生まれた OS で、産業機器の家電の組み込み OS として広く普及している。

*IT-Engine の普及によって、これまでバラバラに行われてきた ITRON プラットフォームの標準化を推進し、ミドルウェア流通の円滑化を図ろうとしている。

本研究では、図1の SH-4 を搭載したソフトウェア評価用ボードを使用した。

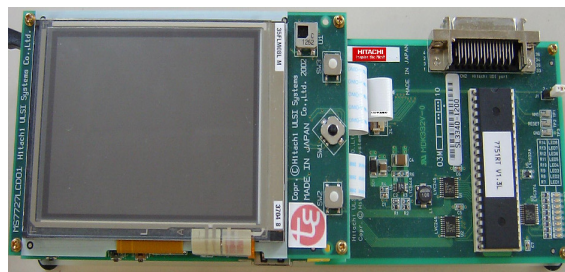


図1 T-Engine (SH-4 利用)

2.4 開発環境

T-Engine の標準開発環境は RedHat linux を基準としている、Windows 上で開発環境をインストールするため、Windows 上で

unix 環境を利用可能にする unix エミュレータ「Cygwin」がある。

本発表では、開発環境として(株)ルネサステクノロジーの統合開発環境ソフト HEW (High-performance Embedded Workshop)とマルチタスクデバッグ環境ソフト HIAE (HI ApplicationEngine for T-Engine)を用いた。

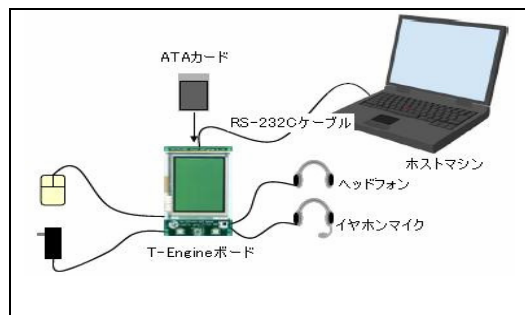


図2 システム構成図

図2のような構成にすることによりPCからT-Engineを操作することができる。

2.5 仕様比較

PCとT-Engineとの音響周りの仕様を比較した。

表1 PCとT-Engineの比較

	PC	T-Engine
CPU	Intel Pentium(R)4 2.80GHz	SH-4 240MHz
ワークメモリ	40GB	64MB
サウンド特性	11.025KHz	16KHz*1

*1T-Engine ボードのサウンド特性は本来はPCと同様で11.025KHzだが、Juliusの入力サウンド特性が16KHzのため16KHzに変更している。

3. 音声認識の組み込みシステム搭載での課題

連続音声認識ソフト Julius を T-Engine 開発環境 HEW と HIAE を使い T-Engine に搭載する過程での問題点の整理と、搭載後の T-Engine と PC とで Julius の処理能力を比較した。

3.1 T-Engine のシステム搭載での問題

統合開発環境 HEW とマルチタスクデバッグ環境 HIAE を使い T-Engine に Julius の搭載を行った。

- HEW のコンパイル環境(SHC)は Cygwin のコンパイル環境(GNU)と異なっている。
- ソースで定義されているプリプロセッサ”#line、#error、#pragma、#”は無視される。
- SH マイコンはシリーズによってコンパイル環境が異なってくる。
- コンパイル時にソースに合わせて標準ライブラリを作成しシン

ボルを参照させる必要がある。

- HEW のコンパイル環境(SHC)に対応していない型定義がある。

3.2 T-Engine と PC とでの処理能力の比較

本研究では PC と T-Engine に搭載されている Julius に wav ファイルを認識させ、それぞれのシステム起動時間、音声処理時間を測定した。

- ◆比較内容
 - Julius のシステム立ち上がり時間
 - 音声処理時間

システム起動時間はそれほど差は無かったが、実際の処理速度は大きく差が出た。

表2 PCとT-Engineの処理能力比較

	PC	T-Engine
システム起動時間(s)	11.37	14.11
音声区間検出(s)	4.17	4.91
1st Pass(s)	0.08	2.77
2nd Pass(s)	0.08	1.68

*使用した wav ファイルの音声データ長は3.71秒であり、音声区間検出の時間は音声データ長を含む。

4. 今後の進め方と課題

T-Engine への Julius 搭載は完全なものではなく改善が必要である。また、T-Engine と PC の処理時間の比較はストップウォッチを用いて測定したためタイムラグが生じている。今後はタイムスタンプなどのソフトを用いて正確な処理時間を求めている。

5. 謝辞

本研究を実施するにあたり T-Engine で情報を戴いた日立中研小窪浩明氏、日立超 LSI 角田知明氏に感謝する。本研究は、(株)ルネサステクノロジーSHフォーラムから、T-Engine、開発環境の支援を受けた。

参考文献

- [1]H.Kokubo, N.Hataoka,etal 「Real-Time Continuous Speech Recognition System on SH-4A Microprocessor」 MMSP07 (2007.10)
- [2]佐藤 幹他、「小型音声対話モジュールの開発」、情報処理学会研究報告書、pp.41-42(2007)
- [3]Julius HP: <http://julius.sourceforge.jp/>
- [4]榎本 知一「連続音声認識ソフトのTエンジン搭載」平成 20 年東北地区若手発表会概要、P.25-26