

日立の音声研究開発戦略

—汎用マイコン用音声ミドルウェアの開発—

畑岡信夫 (株)日立製作所 中央研究所

The Strategy for Hitachi's Speech R&D

- Development of Speech Middleware on Microprocessor for Consumer Applications -

Nobuo Hataoka (Central Research Laboratory, Hitachi Ltd.)

Speech processing technologies, especially speech recognition and speech synthesis are the most important elements to build sophisticated user interfaces for digital consumer products including car navigation systems, H/PCs (Hand-held PCs), and game machines. In this paper, first, the speech research and development strategy in Hitachi is summarized, and next the development of speech processing middleware on microprocessors is briefly reported.

Key Words: speech recognition, speech synthesis, speech processing middleware, microprocessors, digital consumer products, car navigation systems, HPCs, game machines

1. はじめに

最近、音声技術、特に音声認識の製品化の話題が頻繁に新聞等に掲載されるようになった。これは、音声認識技術の進歩とともに、パソコンやマイクロプロセッサ(マイコン)の処理能力が向上し、人間と機械との究極的なヒューマンインタフェースとしての音声認識・合成技術が手軽に実現できる時代になってきたことに起因する。特に、汎用マイコンでの音声認識、音声合成の実現は、応用範囲を大きく広げることにも貢献し、ユーザインタフェースとして不可欠な技術である音声認識、合成が、情報家電やカーナビゲーションシステムなどのマイコン搭載製品へ利用できる環境が整ってきた。我々は、汎用マイコンをプラットフォームとした、アプリケーションに依存しない音声認識・合成機能である音声ミドルウェアの開発を推進し、顧客向けの具体的な応用に関するソリューションとしての製品を提供している。

ここでは、まず(株)日立製作所における音声研究戦略を概観し、次に、汎用マイコンをプラットフォームとした音声認識・合成エンジンを備えた音声ミドルウェアに関して、技術の特長と応用展開に焦点をあて簡単に報告する。

2. 音声研究開発の戦略

2.1 ヒューマンインタフェースを支える要素技術

音声認識、音声合成技術は、人間と機械の間を仲介するヒューマンインタフェースを支える必須技術である。自然に話した言葉の認識と人間と同様な自然な音声を合成することは、音声認識、合成研究の最終的な目的である。現状の段階としては、限定された語彙の認識と限定された定型文章の合成が実用化のレベルにあると言える。

2.2 市場動向

音声認識・合成市場は、大きく分けて、組込み型用途、電話通信応用、PCデスクトップ、福祉応用の4つの分野で発展して行くと考えられる。特に、マイコン応用を対象とした組込み型用途の市場は、情報化、ネットワーク化、モバイル化の社会を反映して、今後大きく成長する事が予想される。

2.3 研究開発戦略

音声認識・合成技術の応用としては、マイコン向け組込み型ソフトウェア、いわゆるミドルウェアを重点分野として捉え、実環境で動作するためのコア技術の開発を推進する。具体的な応用は、カーナビゲーションやモバイル機器、携帯電話、ゲーム機などのマイコンが搭載されている情報家電品である。

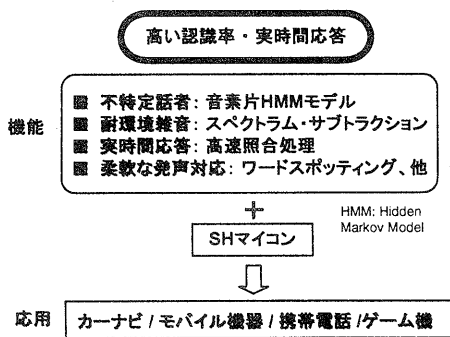


Fig 1 Strategy for Speech Recognition R&D

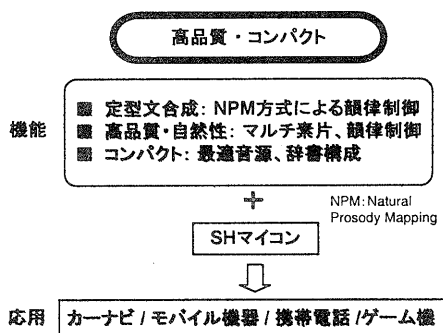


Fig 2 Strategy for Speech Synthesis R&D

図1、図2にそれぞれ音声認識、音声合成の研究開発の戦略を整理して示した。

3. 音声ミドルウェアの開発[1]

3.1 マイコン向けミドルウェア

ユーザのアプリケーションとCPUであるマイコンの間に介在し、マイコンの処理機能に最適化したソフトウェアをミドルウェアと呼んでいる。ミドルウェアの特長は、多様化対応、低価格、小型・低消費電力化、開発の短期化がある。ROMのプログラムを換えることで、音声処理、画像処理などの多様化に対応でき、結果として低価格化、開発の短期化が可能となる。

3.2 SH音声ミドルウェアの基本仕様

今回開発した *SuperH* マイコンをCPUとしたSH音声ミドルウェアの仕様を表1に示した。音声合成では、波形音源を保持して、韻律制御としては、肉声からパターン化した韻律を用いることで、より自然な音声合成が可能となっている[2]。認識方式は、音素片の隠れマルコフモデル(HMMs)方式であり、環境変化

と使用者の話者変動に強い機能として、雑音対策と話者適応を備えた認識仕様となっている[3]。CPUは日立のSH-3 (60MHz)マイコンであり、仕様決定は、当面の応用であるカーナビを念頭において決定された。

Table 1 SH Speech Middleware Specifications

#	項目	内容	
1	処理サイクル	60 MHz	
2	外部バス	60 MHz / 32 bit	
3	サンプリング周波数	11 kHz / 12kHz / 16kHz	
4	音声合成	合成モジュール	定型文 / 任意文章
5		音源	VCV(母音・子音・母音)波形
6		韻律付与	肉声韻律
7	音声合成	メモリサイズ	700kB (音源、辞書) 150kB (work)
8		音響モデル	音素片 / 半連続 HMM
9	音声認識	フレーム周期	10 ms
10		フレーム長	20 ms
11		処理時間	1.2 ms / フレーム
12		応答時間	~0.6 sec
13	音声認識	語彙数	2 000語
14		メモリサイズ	200 kB (音響モデル、辞書) 500kB (work)

4. 音声ミドルウェアの市場ニーズ

ミドルウェアの市場ニーズは、画像、音声、通信の融合を目標としたシステムの多様化への対応と、低価格化、小型化、低消費電力化などの廉価対応、および製品サイクルの速さに応じた開発短縮を可能とするシステム対応である。ハードウェアとソフトウェアの連携によるトータルソリューション対応が必須となっている。

5. おわりに

ここでは、(株)日立製作所における音声認識・合成の研究開発戦略と、汎用マイコンをCPUとした音声処理ミドルウェアに関して報告した。

参考文献

- [1] 畑岡信夫、他：日立評論 Vol.80, No.7, pp.31-36 (1998年7月)
- [2] 額賀信尾、他：音響学会講演論文集、1-7-24, pp.227-228 (1998年3月)
- [3] 小窪浩明、他：信学全大春季SID-2-9 (1999年3月)
<http://www.hitachi.co.jp/Sicd/Japanese/Products/>