音声言語情報処理　25－7
（1999．2．5）

汎用マイコンにおける音声認識・合成ミドルウエアの紹介

小窪 浩明　額賀 信尾　大淵 康成　天野 明雄　北原 義典　畑岡 信夫

日立製作所　中央研究所
〒185-8601 東京都国分寺市東恋ヶ窪 1-280
Tel.：(042) 323-1111
E-mail：{kokubo,nukaga}@crl.hitachi.co.jp

あらまし

マイコンの性能向上に伴い，これまでは専用 LSI や DSP などを用いて実現していた音声処理が，マイコンのソフトウェアで実現が可能となってきた．本発表では，汎用マイコンを CPU とした音声ミドルウェア技術の概略説明とデモを行う．

Review of Speech Recognition & Synthesis Middleware on Microprocessor

Hiroaki KOKUBO　Nobuo NUKAGA　Yasunari OBUCHI
Akio AMAMO　Yoshinori KITAHARA　Nobuo HATAOKA

Hitachi Central Research Lab.
1-280 Higashi-Koigakubo Kokubunji Tokyo, Japan
Tel.：+81 42 323 1111
E-mail：{kokubo,nukaga}@crl.hitachi.co.jp

Abstract

With a progress of microprocessor's performance, developers can easily realize speech applications using microprocessors. In this paper we describe an outline of speech middleware technology, and also show demonstrations using the middleware developed.

1　はじめに

従来，音声認識や音声合成は，処理量とメモリ規模，およびアナログ入出力機能の具備という3つの観点から，専用装置と高性能なワークステーションでしか実現されなかった．しかし，現在は，マイクロプロセッサ（マイコン）の処理規模が100MIPS を越え，さらには半導体メモリも従来のディスクメモリと遜色がない規模になっている．この結果，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・合成技術が，情報家電やカーナビ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，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汎用の音声認識・合成機能である音声ミドルウェアを開発した[1][2].

2　音声認識・合成ミドルウエア

2.1 マイコン向けミドルウエア

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．従来は専用のマイコンを CPU として，さらに周辺回路として専用の音声処理 LSI で実現されていた音声処理が，ミドルウェアだけで実現することが可能となっている．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，さらにはアプリケーション開発期間の短縮がある．

*SuperH*マイコンを CPU として開発した SH 音声ミドルウェアの仕様を表1に示した．

2.2 音声認識ミドルウエア

音声認識ミドルウエアの詳細に関しては，文献[3]で報告している．本ミドルウエアは，雑音対策と話者適応機能[4]を備えることによって，環境変化と話者の変動に対して頑健であるという特徴を持つ．今回，スペクトルサブトラクション方式[5]を新たな雑音対策として搭載することにより，騒音に対する頑健性がさらに向上した．

2.3 音声合成ミドルウエア

音声合成ミドルウエアには，任意文音声合成モジュールと，高品質定型文音声合成モジュールの2種類が用意されており，アプリケーション側から切り替えることができる．両モジュールとも，素片として，VCV（母音-子音-母音）及びCV（子音-母音）を用いている．

音声合成における技術課題は，高い了解性の実現，及び高い自然性の実現である．本ミドルウエアにおいては，高い了解性を実現するために，波形重畳法を用いている．了解性試験では，同83.5%の音節了解度が得られている．さらに，高品質定型文音声合成においては，多くの肉声発声による韻律パタンをデータベース化しておき，韻律計算処理部において，入力テキストに近い文例をデータベースの中から検索し，その韻律パタンをマッピングする肉声韻律マッピング方式[6]により高い自然性を実現している．同方式では，「今夜の天気をお伝えします」のような一般文章音声の韻律パタンをデータベース化しておく．入力テキストが，例えば「現在の残高をお知らせします」や「研究会の発表を募集します」などの文であれば，「今夜の天気をお伝えします」が近い文例として検索され，その肉声韻律パタンがマッピングされる．このように，比較的定型的な文については，本方式により極めて自然な聴取感が得られる．自然性評価試験を通じ，同方式を採用しない場合の自然性が5段階評定平均2.7であるのに対し，同方式では同3.6に向上することが確認されている．

表1 SH音声ミドルウエアの仕様

(100MHz版SH-3の場合)

項目		内容
処理サイクル		100MHz
サンプリング周波数		11kHz(音声認識) 11/22kHz(音声合成)
音声認識	音響モデル	音素片・半連続HMM
	フレーム長 / 周期	10ms / 20ms
	応答時間	~0.6秒
	語彙数	最大2,000語
	メモリサイズ	200kB (音響モデル，辞書)
音声合成	素片	VCV,CV
	合成方式	波形重畳法
	メモリサイズ	700kB(素片，辞書)

3 SH音声認識・合成ボードの構成

図1はSH音声ミドルウェアを搭載した，音声認識・合成評価用ボードの構成を示している．ROMにはプログラム，音響モデル，素片，辞書等が格納されている．音声認識では，11kHzでサンプリングされた音声に対し，CPUであるSH-3にて認識処理が実行される．認識結果は，RS-232Cを経由して，PC端末にて表示される．音声合成の場合は，RS-232Cを経由して指定された文字コード列に対して，CPUにて規則合成処理が実行された後，D/A変換され合成音声が出力される．

4 むすび

本稿では，SHマイコンをCPUとした音声ミドルウェアの概要について述べた．今後は，任意文音声合成のさらなる自然性向上と大語彙認識を実現し，さまざまな製品への展開を図っていく予定である．

謝辞

本ミドルウエアの実装を担当して頂いた，当社半導体事業本部システムLSI事業部と(株)日立超LSIシステムズの皆様に感謝致します．

文献

[1] 鳴島，他，“システムインテグレーションを支える*SuperH*用音声合成・認識ミドルウエア”,日立評論 vol.79,No.11,pp.45-50,1997.11

[2] 畑岡，他，“SuperH マイコン用音声ミドルウエア”，日立評論 vol.80,No7,pp.31-36,1998.7

[3] 小窪，他，“環境適応機能付き音声認識ミドルウエア”，情報処理学会研究報告 SLP22-3,pp.15-20, 1998.7

[4] Ohbuchi, et al., “A Novel speaker adaptation algorithm and its implementation on a RISC microprocessor”, IEEE workshop on ASRU, 1997.12

[5] Boll：“Suppression of Acoustic Noise in Speech Using Spectral Subtraction”, IEEE Trans. on Acoustics, Speech and Signal processing, Vol. ASSP-27, No.2,pp.113-120 (1979.4)

[6] 額賀他，“単語および文韻律データベースを用いた韻律制御方式の検討”，音響講論 1-7-24,1998.3

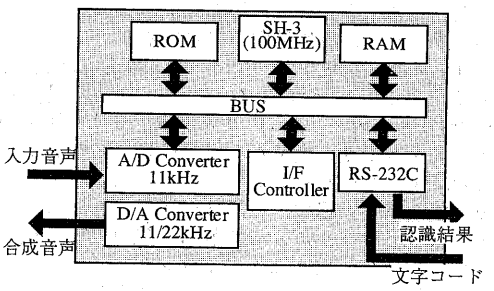

図 1 音声認識・合成ボードの構成