2004－SLP－50　(2)
2004／2／6

音声処理にかかわるインフラの現状と進歩
– マイコン CPU、メモリ、ネットワーク技術 –

畑岡　信夫
（株）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
e-mail: hataoka@crl.hitachi.co.jp

本稿では、音声認識、音声合成等の音声処理を実現するマイコン・メモリの処理デバイスに関する状況と、応用展開を図る上で重要な要素である通信を含めたネットワーク・インフラにおける現状と将来の展開に関して述べる。具体的に取り上げた項目は、デバイス関連では、マイコン CPU、半導体メモリ、HDD、通信インフラでは、移動体通信、高速無線 LAN である。マイコンの処理能力は、数十万語や連続音声を実現するレベルに到達していること、通信インフラは、音声・動画を含めて、十分に応用展開を支援するレベルに到達していることを述べる。さらに、具体的な事例として、汎用マイコンでの音声処理ミドルウェアの開発結果に関して報告する。

Key Words: 音声認識、音声合成、マイクロプロセッサ、半導体メモリ、HDD(Hard Disc Drive)、通信インフラ、移動体通信、無線 LAN(Local Area Network)、音声処理ミドルウェア、カーナビゲーション、携帯端末機（HPC: Hand-held PC）、携帯電話、HMI(Human Machine Interface)

Current and Future Status on Infrastructure relating to Speech Processing
– Microprocessor CPU, Memory, Network Technologies –

Nobuo Hataoka
Central Research Laboratory, Hitachi Ltd.
1-280 Higashi-koigakubo, Kokubunji, Tokyo 185-8601, JAPAN
e-mail: hataoka@crl.hitachi.co.jp

In this paper, the surveys on processing devices such as microprocessors and memories, and on communication infrastructure, especially wireless communication infrastructure, relating to speech processing including ASR and TTS are reported. As the devices, RISC based microprocessors, semi-conductor memories, and HDD are described. As the communication infrastructure, mobile communications and wireless LAN are surveyed in details. Finally, the development results concerning Speech Middleware on microprocessors are reported.

Key Words: ASR(Automatic Speech Recognition), TTS(Text-to-Speech), Microprocessor, Semi-conductor Memory, HDD(Hard Disc Drive), Communication Infrastructure, Mobile Communication, Wireless LAN(Local Area Network), Speech Processing Middleware, Car Navigation Systems, HPC(Hand-held PC), Cellular Phone, HMI(Human Machine Interface)

1. はじめに

音声認識や音声合成の研究の歴史は長い。その結果、タスクや使用環境を限定すれば、現実に使用可能なレベルでの装置、システム、ソフトウェアが、製品として出て来ている。一方、音声処理を行う計算機、メモリ環境や、応用を担う通信インフラストラクチャ(以下、インフラ)も、飛躍的に進歩をしている。その結果、昔は大型計算機でしか実現できなかった音声認識等の処理も、現在では、マイクロプロセッサ(以下、マイコン)でも実現でき、カーナビ端末や携帯端末機(HPC: Hand-held PC、あるいは PDA: Personal Digital Assistant)による新しいサービスが期待されている。いわゆる、モバイル(mobile)環境でのユビキタス(ubiquitous)端末を利用したユビキタス時代の新しいサービスの創生である。

本稿では、音声処理と応用にかかわるインフラとして、マイコン等の処理デバイスと、通信インフラ等の現状および将来に関して、サーベイを行ない、かつ、具体的な事例として、汎用マイコンでの音声処理ミドルウェアに関して報告する。

2. 音声処理に関与する環境

音声処理に関与する環境としては、処理を実行するマイコン等のデバイスと、応用に関係する通信インフラ等がある。図 1 に、端末(terminal/client)とインタネット、及びセンター(Center/Server)で構成されるシステムイメージを示した。処理デバイスでは、パソコン(PC)も含めて、マイコン CPU(Central Processing Unit)、メモリ等の半導体に関係する環境であり、通信インフラは、有線(wired)、無線(wireless)に関係する環境である。本稿では、主に端末側の処理を担うデバイス状況と、ネットワークへアクセスして、情報のサービスを受ける端末側の通信インフラ、特に無線インフラに関して整理する。

図 2 に、携帯端末(HPC/PDA)を例に、サービスと処理プロセッサ規模、及びメモリ規模を整理した。多言語の音声翻訳サービスを目標とした場合は、2GIPS(Giga Instruction Per Second: 10 億回)処理規模が必要で、メモリは 100Mbyte 以上必要となるであろう。

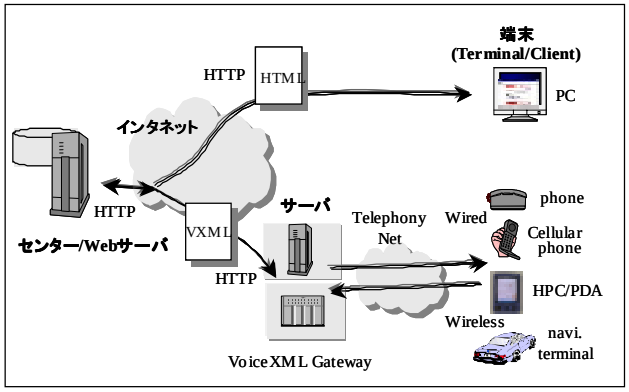

Fig.1 System Image (Terminal, Internet, Center)

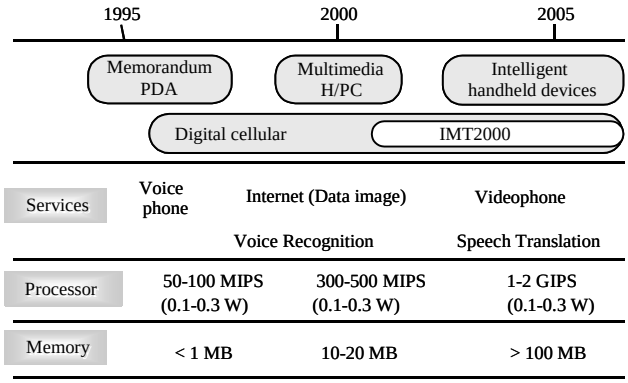

Fig.2 Hardware Needs for HPC/PDA Application

図 3 は、音声認識ソフトウェアを構築する場合の実現方法を示している。処理規模に応じて、①認識チップ、マイコンソフトでの実装、②PC とソフトで実装、③CSS(Client and Server System)での実装の3通りが考えられる。それぞれの応用の具体例と処理量、メモリ規模を整理した。

	発声/語彙数	具体例	処理量	メモリ	装置形態	コスト
1	単語/小語彙 *1	・音声ダイアル ・車載情報機器	～100MIPS	500KB	チップ, ボード	1～5k¥
2	単語/中語彙	・公共端末 (券売機、ATM等)	250MIPS	～5MB	PC(Audio装備)	PC 50k～500k¥
3	文/中語彙	・電子秘書 (スケジュール管理等)	500MIPS	20MB		
4	文/大語彙	・ディクテーション ・音声翻訳	1000MIPS～	50MB～	CSS(Client & Server)	500k¥～

＊1 小語彙:～100語 中語彙:100語～2000語 大語彙:2000語～

Fig.3 Implementation Varieties for ASR

3. デバイス環境

「ムーアの法則」によれば、3 年で 4 倍という速度で半導体集積化が進んでいて、現在 1 チップに億単位のトランジスタが載るまでになってきた。現在のプロセス技術は、0.18μm が主流で、0.1μm 以下も実現されている。これは、光リソグラフィー技術など微細加工技術と多層配線技術、及び設計技術である DA（Design Automation）技術等の進歩に起因する。このような LSI 高集積化技術の進歩に伴って、マイコンの処理能力やメモリ容量が大幅に向上している。また、マイコンとメモリ、あるいは特定の処理チップを混載した、いわゆるシステム LSI が実現されている[1]。

システムをワンチップ化するシステム LSI は、処理の高速化、低コスト化、低消費電力化、ダウンサイジング等を実現し、モバイル環境での処理デバイスとして重要な意義を持っている。システム LSI は、SoC (System on a Chip)とも呼ばれている。

以下、モバイル環境での応用を考えた場合、音声処理に大きく関与するマイコン、及びメモリに関して、現状と今後の展開を概観する。さらに、デバイスと大きく関与するソフトウェアである OS(Operation System)や、音声応用システム構築に関して重要である Java 関連のソフトウェア環境も簡単に整理する。

3.1 マイコン CPU

マイコンの驚異的な性能向上により、音声認識、画像圧縮等の複雑な処理が、ソフトウェアだけで実現できるようになってきた。これらは、デバイスの進歩とともに、方式の進歩や開発環境の進歩が背景にある。なお、マイコンの場合は、MPU (Micro Processing Unit)とも言われるが、本稿では、マイコン CPU という表現を使用する。

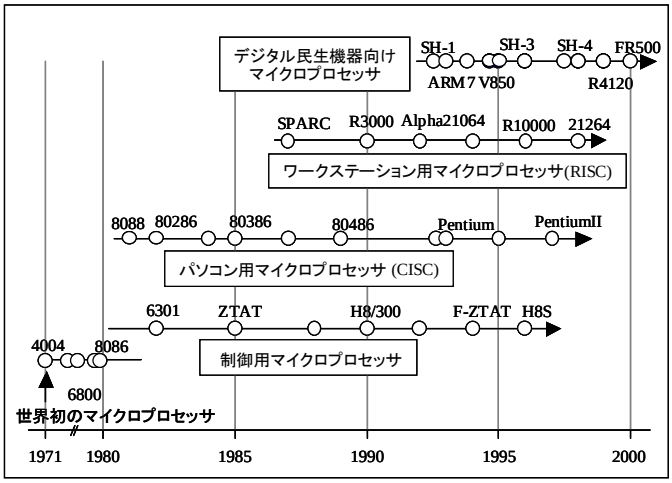

Fig.4 Microprocessor Product Trend

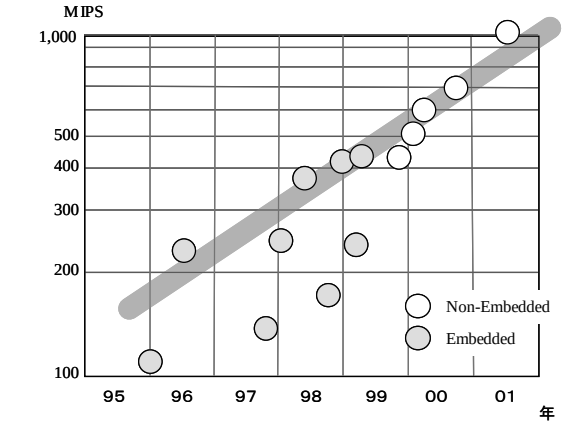

Fig.5 High Performance Trend of Microprocessor

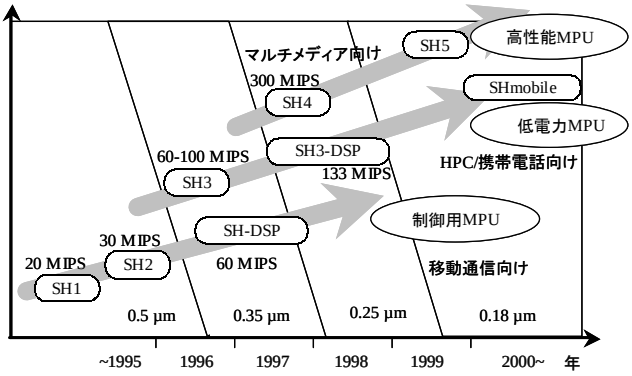

Fig.6 Road Map for Microprocessor (*SuperH* Series)

(1)マイコン性能のロードマップ:

図4に各種マイコンの商品動向、図5に組込み型用マイコンとそれ以外（主に、PC 用）とでの処理性能の動向を示した。

(2)高性能化への定式:

高性能化＝（動作周波数）×（命令実行サイクル数低減）と定式化される。ここで、動作周波数の動向は、99 年 200～300MHz、2000 年～2002年400～500MHz、～2005年1GHzである。命令実行サイクル数低減は、処理のパイプライン化やアーキテクチャの簡素化に比例する。5段以上のスーパパイプラインとすることや、RICS（Reduced Instruction Set Computer）や VLIW（Very Long Instruction Word）等のアーキテクチャ簡素化が検討されている。

図 6 に、RISC 型マイコンである日立 *SuperH* プロセッサのロードマップを示した。*SuperH* の音声ミドルウェアに関しては後述する。

3.2 メモリ環境

ここでは、音声処理にかかわるデバイスの中で、メモリ関係を取り上げ、現状と将来に関して概観する。取り上げるメモリは、DRAM(Dynamic Random Access Memory)、フラッシュメモリの半導体メモリと最近カーナビシステムで採用され始めた小型HDD(Hard Disc Drive)である。

図7は, 半導体技術であるプロセスとDRAMメモリ容量の動向を示した。現在、100nm(0.1μm)以下の微細化技術が完成しており、DRAM の容量は、数Gbitに達している。

項目＼年	1999	2000	2002	2005	2008	2010
Process(nm)	140	120	85	65	45	30
DRAM(Gbit) (Mbyte)	1 128	2 256	4 512	8 1,024	24 3,072	64 8,192
配線層数	6-7	6-7	7-8	8-9	9	9-10
電源電圧(V)	1.5-1.8	1.5-1.8	1.2-1.5	0.9-1.2	0.6-0.9	0.5-0.8
Soc素子数 (M.Tr.)	24	30	60	142	2,500	7,000

Fig.7 Memory Road Map

端末でのメモリは、不揮発性で、かつ大容量であるフラッシュ(Flash)が今後広く利用されてくる。また、将来は、SRAM並みの高速性とDRAM並みの大容量を備えたMRAM(Magnetic RAM)が、次世代メモリとして期待されている[2]。

カーナビでは、2.5 インチ以下で、数 GByte の小型HDDが実装され始め、大容量の時代が来ている。2. 5インチサイズ以下のHDDは、モバイル用としてノート PC、デジタルカメラ等のデジタル家電への需要が急増している[3]。最近では、1インチ前後で、4Gbyteの小型(薄型)HDD開発が激化している。

3.3 OSなどのソフトウェア環境

ここでは、OS や Java など、デバイスに大きく関与するソフトウェア環境に関して、簡単に整理する。

(1)T-Engine: ユビキタス時代のネット家電端末でのOS、ボードとして、 TRON の発展系の T-Engine が話題となっている。μiTronは、カーナビ等での実時間組込みOSとして広く使用されている。

(2)OSGI(Open Service Gateway Initiative): Javaによるアプリケーション、プログラムのダウンロード機能を有効に使用するために、Java ベースのモバイル端末のシステム構成において、ソフト間のAPI(application Program Interface)の設定等が、OSGI Alliance にて、Open Service Platformとして、検討されている[4]。

4. 通信インフラ

本章では、特に無線関係を取り上げる。

4.1 無線関係インフラ

無線関係では、技術開発が激しく、キャリアの勢力争いも激化している。以下、音声処理・応用に関係する移動体無線、高速無線 LAN、超高速無線に対象を絞り、状況と今後を整理する。音声の観点からは、通信容量(スループット)と品質に興味がある。

(1)移動体無線:

図8に、携帯電話での第1世代から第3世代までの状況を示した。

第1世代	第2世代	第2．5世代	第3世代
アナログ	デジタル	高性能 デジタル	高速大容量 デジタル
		音声・データ 通信強化	動画通信
	PDC(日本) GSM(欧米)		W-CDMA NTTDoCoMo 384kbps
			CDMA2000 1x KDDI 144kbps

Fig.8 Cellular(Mobile/Portable) Phone Trend

第3世代は、IMT2000とも呼ばれ、国際電気通信連合(ITU)の次世代移動通信システム計画に従って技術開発が推進された。世界各国で使える国際ローミング(相互接続)や毎秒2Mbps(bit per second)の通信速度の実現を目標としている。実現技術で分類すると、W-CDMA(符号分割多元接続)陣営は、NTTDoCoMo の FOMA(サービス名)、欧州のUMTS があり、CDMA2000 1x 陣営は、日本ではKDDI である。動画配信、音楽配信、電子商取引等のアプリ、さらには、GPS利用位置情報サービスもあり、音声HMI(Human Machine Interface)への期待も高い。

データスループットとしては、第2世代(PDC 等)は、9.6 kbps(upload)、9.6 kbps or 27.8 kbps(download)であり、第3世代の FOMA では、64 kbps(upload)、200 kbps 以上(download)となっている。

(2)高速無線LAN: IEEE 802.11a、11b、11gの規格がある。これらの関係を図9に整理した。

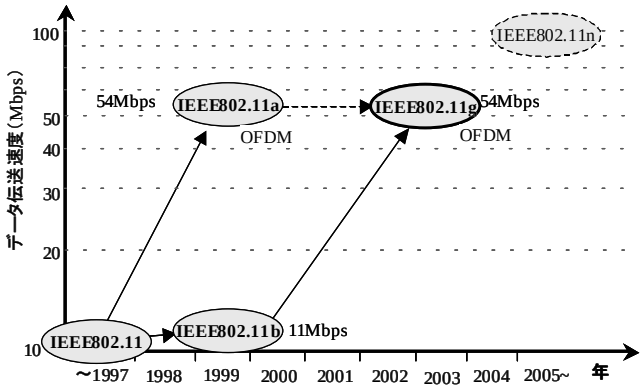

Fig.9 High Speed for Wireless Communication

802.11b(最大11Mbps), 11g(最大54Mbps)は、2.4GHz帯を使用しており、802.11a(最大54Mbps)は5GHz帯を使用している[5]。最近は、高スループットと互換性を考慮して、11gをサポートするPC等、端末機が急速に商品化されている。

11aと11gともに、変調符号化方式は、OFDM(Orthogonal Frequency Division Multiplexing: 直交周波数分割多重)を採用している。キャリア信号を多重化するデジタル変調手法である。OFDM技術は、地上波デジタル放送や無線LANなどの単一セル、孤立セル環境で、優れた性能を出しており、注目されている。一方、第4世代移動体通信などのマルチセル、マルチユーザ環境を狙って、MC-CDMAシステムの開発が推進されている。OFDMの問題点としては、①消費電力が大きい、②マルチパスの影響を受ける、③高信頼性の同期回路が要求される等がある。②に関しては、ガードインタバールで回避することが検討されている。

(3)超高速通信: UWB(Ultra Wide Band:超広帯域)技術が注目されている。周波数帯域は、3～10GHz帯域であるが、各国での周波数帯域割り当てが課題となっている。データスループットは、最大500Mbpsで、通信範囲10メートル以内となっている。技術的には、数GHzから数十GHzといった非常に広い周波数帯を用いて、微弱な電波としてパルスが伝送される。

4.2 無線インフラに関する音声技術

ETSI(European Telecommunications Standard Institute)が進めている AURORA プロジェクトでの分散音声認識(DSR: Distributed Speech Recognition)[6]が、第3世代以降の無線通信インフラを想定した音声技術として注目されている。端末にて、音声を分析し、特徴量をセンターに送り、センターにて詳細な音声認識(デコーダ)を行なう方式である。プロジェクトの目的は、①雑音に頑強な前処理、②端末とセンターとのプロトコルの取り決め、である。例えば、DSR 方式利用で、50%のパケットロスがあっても、特徴量の補間で、認識率劣化が3%におさまる事が報告されている(符号化音声では、63%劣化)[7]。

5. 音声処理実現の例

5.1 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デクテーション、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。

図10に、方式の進歩とマイコンの進歩、及びユーザニーズという3項目の関係が、現状の汎用マイコンでの音声ミドルウェアを実現した経緯を示した。

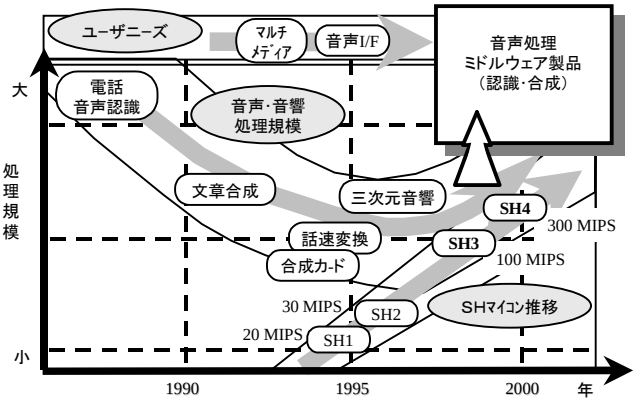

Fig.10 Middleware Realization from Algorithm, Device, and User Needs

5.2 マイコン向け音声処理ミドルウェア[8]

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、低価格化、開発の短期化が可能となる。

5.3 SH音声ミドルウェアの基本仕様

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。認識方式は、音素片の隠れマルコフモデル(HMMs: Hidden Markov Models)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。CPU は日立の SH-3 (60MHz)、SH-4 (200MHz)マイコンであり、仕様決定は、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/200MHz
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	音声合成	音源	VCV(母音・子音・母音)波形
6	音声合成	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	音声認識	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	音声認識	フレーム長	20 ms
11	音声認識	処理時間	1 2 ms /フレーム
12	音声認識	応答時間	～0.6 sec
13	音声認識	語彙数	2 000語
14	音声認識	メモリサイズ	200 kB (音響モデル、辞書) 500kB (work)

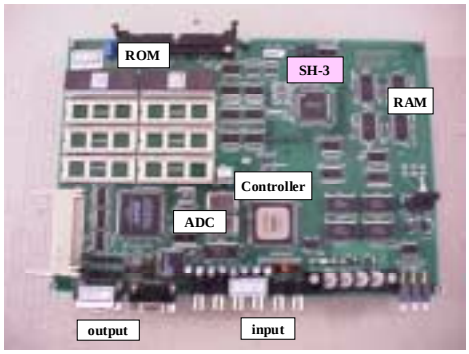

Fig.11 *SuperH* Middleware Board

5.4 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。表 2 に、ミドルウェアの展開を整理した。

Table 2 Speech Middleware Road Map

'97	'98	'99	2000～
SH3 (60 MIPS)		SH4 (300 MIPS)	SH-X (1 GIPS)
単語認識 定型文合成		文章認識･合成	多言語対応 翻 訳
1,000 語		大語彙数 対応 (100,000 語)	言語処理
(明瞭性)		任意文合成 (自然性)	

6. まとめ

本稿では、音声処理にかかわるマイコン CPU、メモリ、及び無線関係の通信インフラに関して整理した。まだ、音声認識等の処理と直結した整理は出来ていないが、将来のモバイル時代において、音声 HMI が真に実用化されることを期待し、最後のまとめとする。

謝辞 (株)日立製作所中央研究所内山邦男主管研究員にはマイコン状況に関して、STARC 平田雅規上級研究員には半導体全般の状況、北海道大学宮永喜一教授には無線インフラに関して、アドバイスを頂いた。感謝致します。

参考文献

[1] 桜井貴康:「総論 –システムLSIのアプリケーションとシステム LSI の課題–」, 信学会誌, Vol81, No.11, pp.1082–1086(1998 年 11 月)
[2] 猪俣浩一郎:「次世代メモリ MRAM」, 信学会誌, Vol84, No.3, pp.159–162(2001 年 3 月)
[3] 三浦義正:「大容量ストレージを支える HDD 技術」, 信学会誌, Vol83, No.3, 204–212(2000 年 3 月)
[4] http://www.osgi.org/
[5] 日経ｴﾚｸﾄﾛﾆｸｽ: 2003.3.17, pp.69–76(2003 年)
[6] http://portal.etsi.org/stq
[7] D.Kopp, et al.: http://www.3gpp.org/ftp/tsg_sg/WG1_Serv /TSGS1_13-Tahoe/Docs/S1-010671.ppt
[8] 畑岡信夫:「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–」, 情処学会 SLP Vol.80, No.7, pp.31–36 (2000 年 7 月)