2005－SLP－55 (1)
2005／2／4

# 音声技術実用化の課題と取り組み

**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、真の音声技術実用化を図ることを目標に、音声認識技術の実用化に関する課題と今後の取り組みに関して纏めた。実用化への課題としては、技術的には、①音声・非音声の分離を主とする音響認識、②HMI(Human Machine Interface)の観点からの音声対話理解、③発話された単語、文を精度良く認識する音声認識が、重要な課題となっている。今後の進め方としては、音声技術実用化へ向けた産学官連携に基づく、技術開発を主導する開発機関の設立が必要であり、具体的な進め方に関しての提言と施策を述べる。

Key Words: 音声認識、音声合成、音響処理/認識、音声対話処理/理解、HMI (Human Machine Interface)

# Technical Problems and Breakthrough Activities for Real Use of Speech 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 current technical problems of speech recognition and the future necessary R&D activities are summarized in order to pursue real use of speech recognition technologies. The technical problems for the real use are, first, the discrimination between speech and non-speech clearly, second, the speech dialog understanding from HMI(Human Machine Interface) viewpoints, third, speech recognition itself to recognize word and sentence utterances. We propose a neutral R&D organization which pursues R&D activities to overcome the technical problems for the real use of speech technologies by the collaborative consortium among companies, universities and governmental R&D institutes.

Key Words: ASR (Automatic Speech Recognition), TTS (Text-to-Speech), Acoustic Processing/Recognition, Speech Dialog Processing/Understanding, 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。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。

このように、音声処理技術、特に、音声認識技術の応用展開の夢は大きく、かつ音声処理技術は、HMI(Human Machine Interface)を実現する重要な技術となっているが、事業、ビジネスの観点からは、まだ大きな展開へとなっていない。すなわち、まだ「実用化」一歩手前という状況である。音声認識ビジネスが大きくないという問題もあるが、本稿では、技術開発がビジネスを牽引するという考え方で、現状の問題と今後の進め方を整理する。

実用化への課題としては、技術的には、①音声・非音声の分離を主とする音響認識、②HMI(Human Machine Interface)の観点からの音声対話理解、③発話された単語、文を認識する音声認識が、重要な課題である。今後の進め方としては、音声技術実用化へ向けた連携的な開発機関の設立が必要であるとの考えに至った[1]。以下、音声認識の現状を整理して、音響認識、音声対話理解の2つの技術課題を詳細に検討する。最後に、今後の実用化
に向けた研究の進め方に関して提言と施策を述べる。

## 2 音声認識の応用展開と実現形態

### 2.1 市場動向と製品化動向

図1に示すように、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。情報家電の分野では、カーナビでの音声インタフェースとしての応用や携帯端末での応用が展開されている。これらは、主に組込み型ソフトウェアの形態である。

製品化状況は、PC向けソフトとコールセンター、産業応用、カーナビにおいて製品化されている。カーナビ応用では、汎用マイコンへの組込み型ミドルウェアとして、廉価なソフトとして商品化がされている。研究レベルでは、大学などの研究機関を中心に、連続音声認識ソフト[2]や、音声翻訳[3]の研究が推進されている。現状の製品の問題は、雑音等が存在する実環境での認識性能が低いこと以外に、「非常識な音声認識」という言葉で表現されるように、まだ使い勝手の観点から大きな問題がある。

現状を纏めると、大きな市場として、コールセンター・音声ポータル等の話題はあるが、日本での応用は限定されている。一方、米国ではかなり利用が進んだ応用分野も見られる。この日本と米国との違いの原因としては、例えば、電話応用の違いがある。4つの時間ゾーンがある米国では、留守番電話が当たり前で、音声認識が広く利用されている。さらに、音声認識エラーをあまり気にせず、機能重視という国民性の影響も大きい。

### 2.2 情報家電での応用展開

主に、情報家電分野での応用展開を概観する。図2には、家庭内応用としては、ハンズフリーのユニバーサルリモコンの使用例を示した。

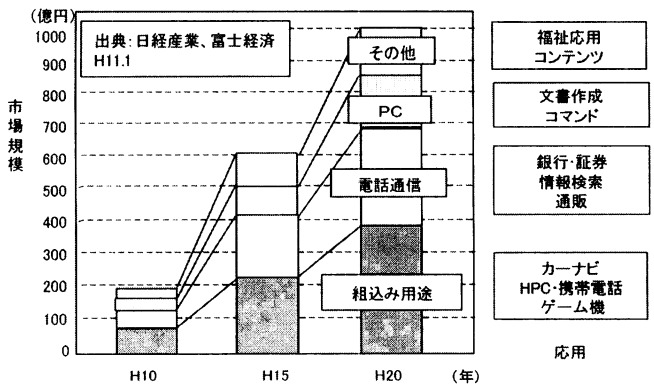



図1 市場規模と市場動向

また、図3のように、検索型モバイル端末も有力な応用展開である。この他に、カーナビ応用も大きい。

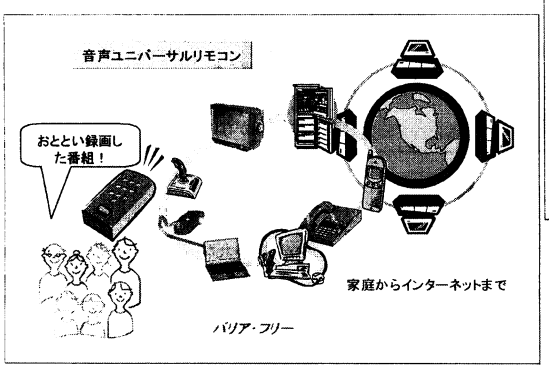



図2 応用展開(1):音声ユニバーサルリモコン

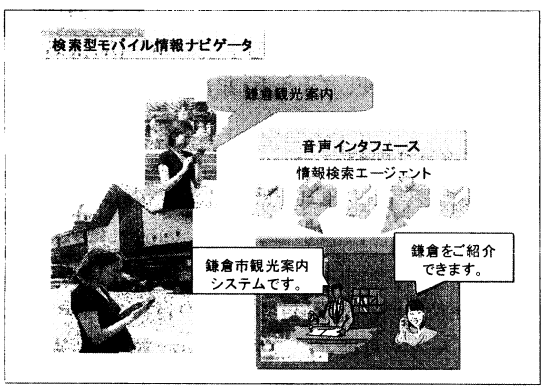



図3 応用展開(2):検索型モバイル情報ナビゲータ

### 2.3 サービスイメージと実現形態

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 4 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。

図 5 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

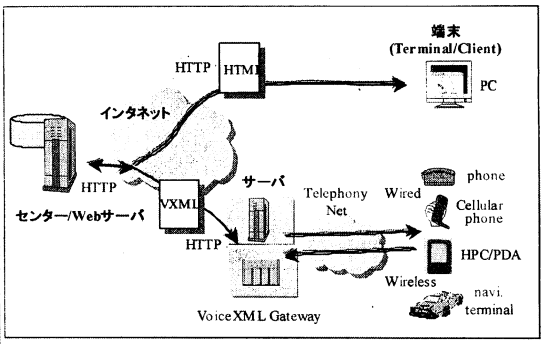



図4 サービスシステム・イメージ

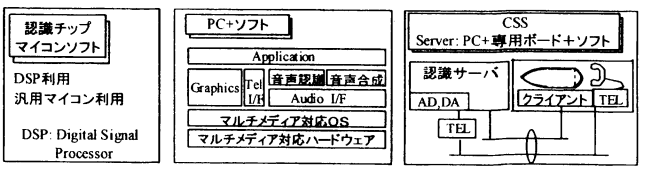
実装の 3 通りが考えられる。それぞれの応用の具体例と処理量、メモリ規模を整理した。特に、多言語の音声翻訳サービスを目標とした場合は、2GIPS(Giga Instruction Per Second: 10 億回)処理規模が必要で、メモリは 100Mbyte 以上必要となるであろう。

## 3 音声認識における現状の課題

### 3.1 実用化阻害要因

図 6 に、音声認識の構成を示した。この構成要素をもとに、現状の問題を整理し、実用化を阻害している要因を表1(a)(b)に纏めた。具体的な問題は、①実環境で認識しない(騒音、妨害音声)、②音声以外で動作してしまう、③利用者は何を言っていいか分からない、④開発者が想定しない発声をされても、リジェクトできない、⑤誤認識に対してシステムの動作が分からないものとなる、のように、利用が極めて限定された「性能の悪いスイッチ」となっていることが挙げられる。これらの課題

| | 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<br>・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<br>(券売機、ATM等) | 250MIPS | ～5MB | | |
| 3 | 文/中語彙 | ・電子秘書<br>(スケジュール管理等) | 500MIPS | 20MB | PC(Audio装備) | PC<br>50k～500k¥ |
| 4 | 文/大語彙 | ・ディクテーション<br>・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～

図5 音声認識の実現形態

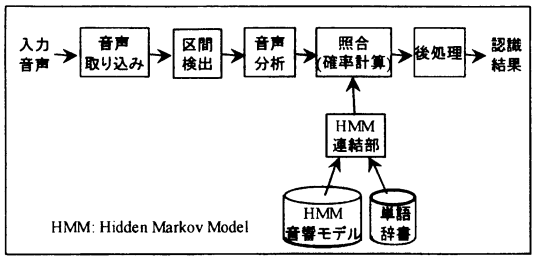



図 6 音声認識の構成

を解決して、実用化を推進するために、従来の「音声認識」から、音声・非音声を分離する「音環境認識」と、使い勝手を向上させる「状況理解に基づく音声対話理解」の実現が不可欠となっている。

### 3.2 実用化に対する研究開発戦略

実用化に対する技術開発戦略として、

①音声・非音声の識別

②使い勝手を向上させる音声対話理解

③方式のベンチマーキング

④方式開発のための音声コーパス収集

に焦点をあて、開発する項目を整理する。

#### (1) 音環境認識

図 7 に、音環境認識の捉え方を示した。音声と非音声の分離は、実環境での雑音対策を目指し、安定した認識エンジンの実現を図る。さらに、複数音声の分離を実現し、対象とする特定話者の音声だけを分離する技術を開発する。

表1(a):実用化阻害要因(音声認識の観点)

| ◆音声認識エンジン関与項目 | |
|---|---|
| 音声取り込み | ・音声と非音声の分離<br>・十分なS/Nの確保<br>・発声と音声取り込みのタイミング<br>・途中割り込み(バージ・イン)<br>・複数話者の音声取り込み |
| 音声区間検出 | ・音声と非音声の分離<br>・雑音の中からの音声検出<br>・複数話者音声から の特定音声検出 |
| 音声分析 | ・雑音、話者頑健 な分析手法<br>・その他変動に強い分析手法<br>・生体を模擬した音声特徴量の抽出 |
| 音響モデル | ・非音声、雑音のモデル化<br>・個人差のモデル化<br>・音響モデルの学習方式(話者適応含む) |
| 照合 | ・信 頼性の高い尤度 の算 出方式<br>・文法外、語彙外発声へ の対応 |
| 後処理 | ・非音声のリジェクション<br>・語彙内発声の 誤リジェクション |

表1(b):実用化阻害要因(HMI の観点)

| ◆HMI関与項目 | |
|---|---|
| 対話シーケンス<br>対話ガイダンス | ・何を、いつ話して良いか分からない<br>・システムの状況が分からない<br>・対話が固定的<br>・同じ間違いを起こす<br>・確認が煩雑<br>・ガイダンスを聞き逃す<br>・標準的な仕様がない、不統一 |
| 使い方 | ・使い方が分からない<br>・操作が難しい<br>・トークボタンを押すのが面倒<br>・バージ・インができない<br>・コマンドが覚えきれない<br>・言い回し、表現が固定 |
| その他 | ・人前で音声入力するのが恥ずかしい<br>・機械に向かって話すのが嫌い |

基本的には、音源分離と音源種別判定を行なう技術であり、HMM、サポートベクトルマシン(SVM)や独立成分分析(ICA)[4]、及び音源分離手法(BSS、BSD)技術を駆使して、音声・非音声の分離を具体化する。

#### (2) 音声対話理解

従来は、「えーと」という想定しない表現でエラーとなり、その後は想定しない対話の流れに入り、システムは、ユーザの発声を認識できない状態となる。期待される対話は、ユーザの自由な言語表現での入力を可能とし、かつ自由な対話表現を理解し、ユーザの多様な応答に対応することを狙う。音声対話理解に必要な知識は、①領域モデル、②対話モデル、③ユーザ状況モデル、④データベースである。

さらに、標準コマンドの設定と標準化活動も重要な課題となっている。

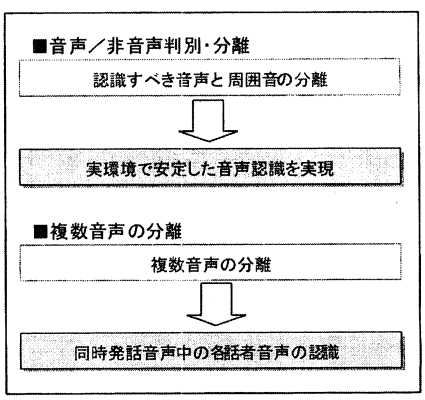



図 7 音環境認識の捉え方

### (3) ベンチマーキング

ベンチマーキングは、開発する技術レベルの到達度を見るために必須である。また、適正な競争力を促進するためにも、重要な項目である。図8に、ベンチマーキングの進め方に関する案を示した。目的は、技術到達度の把握を第一とし、優良ソフトウェアの表彰を行ない、適正な競争力を促進させる。体制としては、中立的な組織にターゲット設定、データ作成、評価・集計の3グループを設置して、実施することを考える。

### (3) 音声コーパス

音声コーパスは、音環境認識、対話理解方式を開発する上で必須なデータベースである。具体的には、実用環境と発話バリエーションを考慮して、次のような音声コーパスを取り上げて、収集する。

①日本縦断音声データ(年齢別、地域別)
②実環境下音声データ(車、駅、事務室、雑踏他)
③会話場面での非規範性音声データ(言いよどみ、フィラー、非文法性、不明瞭発声等)
④対話音声言語コーパスの構築
・対話意図理解のためのアノテーション・コーパス
⑤対話理解と会話音声合成のための会話韻律コーパス
⑥多言語、特に、中国語、韓国語等アジア系言語対話システム開発のための音声言語コーパス

## 4 今後の進め方

本章では、今後の進め方に関して纏める。基本的には、国家プロジェクト提案による産業界コンソーシアム設立を目指し、実用化技術の開発を推進する。

図9は、音声技術実用化研究所(仮称)の設立を目標とした進め方を示した。音声技術実用化研究所では、本稿で議論してきた音環境認識と音声対話理解の基本的な技術開発を推進し、さらに開発に必要となる音声データベース収集とベンチマーキング実施を推進する。成果は、可能な限りオープン化して、民間企業が独自のアプリケーションや商品化を行なえる枠組みを構築する。具体的な音声技術実用化事業計画ロードマップを次頁に載せる。

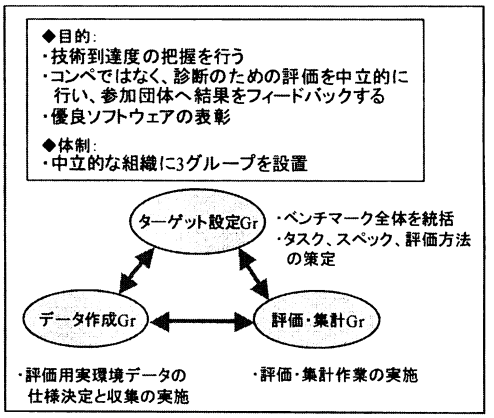



図8 ベンチマーキングの実施(案)

## 5 まとめ

音声認識実用化の阻害要因を解析し、それらを解決して、真の音声認識技術を実用化する戦略と進め方を纏めた。

**謝辞**:本稿を纏めるにあたり、東京工業大学古井教授、早稲田大学誉田教授、小林教授、NEC渡辺氏、東芝金澤氏、旭化成庄境氏、三菱石川氏、ソニー赤羽氏、日立大淵氏の皆様に感謝致します。

**参考文献**

[1] 畑岡、他:音講論、1-8-10、2004年3月
[2] 河原、他:情処学会、SLP 研究会予稿集「音声言語情報処理」、No.048-001 (2003.10)
[3] ATR: http://www.atr.co.jp/
[4] T.Lee: Kluwer Academic Publishers, 1998

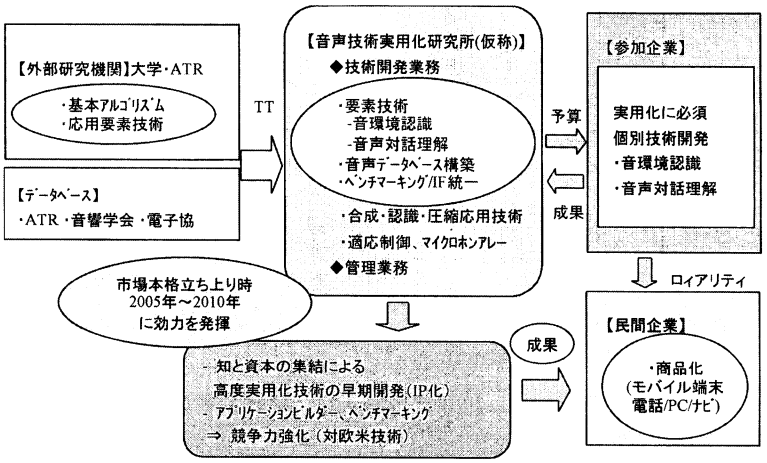



図9 音声技術実用化研究所(仮称)の設立

# 音声技術実用化事業計画ロードマップ

JEITA「音声技術実用化推進WG」
文責：畑岡@日立中研、赤羽@ソニー

| | 第一段階（予備検討） | 第二段階（研究開発事業） | 第三段階（ベンチャー＆実用化） |
|---|---|---|---|
| 1. 時期・期間 | 平成17年4月～平成18年3月<br>（平成17年度）<br>（1年間） | 平成18年4月～平成21年3月<br>（平成18年度～平成20年度）<br>（3年間） | 平成21年4月～<br>（平成21年度～） |
| 2. 提案事業 | 経産省先導研究 | 経済産業省国家プロジェクト<br>（委託） | ①大学＆企業連携ベンチャー会社<br>②出向元企業での事業実施 |
| 3. 実施内容 | ・音声実用化コンソーシアム設立<br>・実用化技術の予備評価<br>①音環境認識<br>②対話認識<br>③プラットフォーム整備枠組み<br>・ベンチマーキング予備実施<br>・パテント・プール方式の検討 | ・実用化技術の開発とプラットフォーム整備<br>①音環境認識<br>②対話認識<br>③プラットフォーム整備<br>④音声コーパス整備<br>・ベンチマーキング実施<br>・ベンチャ会社創生の枠組み整備 | ・音声認識R＆Dベンチャー会社による音声実用化<br>・開発した技術を音声コンソーシアムとして参画した企業を始め、世界中の企業へ提供 |
| 4. 狙い | 大学・独立法人研究所による要素技術研究と、企業コンソーシアムによる実用化開発の産学官連携を実施 | 大学・独立法人研究所による要素技術研究と、企業コンソーシアムによる実用化開発の産学官連携を実施 | 参画企業（音声コンソーシアム）の音声認識R＆Dリソースを集中させて、幅広く効率的にR＆Dを行い、世界でトップレベルの技術を維持する。 |
| 5. 具体的目標 | 第二段階の研究開発事業実施に向けて、開発技術の絞り込みと枠組み作成 | ・目標とする実用化技術開発の完了<br>・音声実用化ベンチャー会社の設立<br>・各企業での事業化推進 | 組込み型応用を目標に、音声ビジネスの立ち上げ |
| 6. 備考 | ・JEITA支援による音声実用化コンソーシアムの設立<br>・早稲田大学内に音声技術実用化研究所（仮称）の設立計画<br>・大学TLOよる管理会社 | ・JEITA音声実用化コンソーシアムをベースに、早稲田大学内に音声技術実用化研究所を設立<br>・参画企業からの研究者の派遣<br>・大学TLOよる管理会社 | |