

三菱電機における音声情報処理技術の開発

石川 泰

yasushi@isl.melco.lco.jp

三菱電機株式会社 情報技術総合研究所

概要 本稿では、三菱電機における音声情報処理技術の研究開発の概要と動向を示す。1980年代初頭から本格的に取り組んでいる開発について、電話系音声認識システム MELAVIS や原子力プラントマンマシンシステムなどの製品、試作事例を紹介するとともに、研究開発における当初の予想と現実の相違について論じる。最後に、現在の研究開発課題と、問題点について述べる。

Development of Speech Processing in Mitsubishi Electric Corporation

Yasushi Ishikawa

yasushi@isl.melco.co.jp

Information Technology R&D Center, MITSUBISHI Electric Corporation

Abstract This paper describes research and development activities on speech processing in Mitsubishi Electric Corporation. Our products and trial systems including speaker independent speech recognition system for computer telephony applications "MELAVIS" and a dialogue system for nuclear power plants are introduced, and differences between expectations and current state are discussed. Finally, our current works and problems are described

1. はじめに

三菱電機での音声情報処理技術の研究開発の歴史は、1981年にさかのぼる。当時、いくつかの開発部門が統合され、現在の情報技術総合研究所の前身である情報電子研究所が設立され、音声処理技術の開発を推進するグループが小人数ながら作られた。これにより、通信分野の開発者が始めようとしていた音声認識の研究開発が本格的にスタートした。1982年に入社した筆者は、したがって、三菱電機における音声情報処理技術の研究開発の初期から今日に至るまで、当該分野の開発に従事してきたことになる。我々の研究開発

がスタートしたころ、1980年代の前半のことを思い出してみると、音声認識に対する期待は大きかったものの、「将来技術」としてとらえられることも多く、我々自身、説明にあたり、SF映画である「2001年宇宙の旅」に登場するコンピュータ HAL を引きあいにして、「2001年になっても HAL のようなコンピュータは恐らく実現はしないであろうが、音声認識はコンピュータの主要なインタフェースとなるはずであり、極めて重要な技術である」こと言っていたことを思い出す。ここでは、その2001年を目前にして、いままで我々が開発してきた製品や、試作システムを紹介し、過去の予測や期待と現状の相違について議論

し、現在の研究開発課題とその実現に向けての問題を概観する。

2. 三菱電機における音声技術の開発

2.1. スタート時点での将来像

当社における音声情報処理技術の研究開発が本格化した1980年代初頭を振り返ると、

- 1) 線形予測音声分析手法、DTWによる単語音声認識方式の確立した
- 2) LSI技術が立ち上がり、専用LSI開発が盛んに行われ、DSPが出現しはじめた
- 3) これらの成果から、単語音声認識装置が我が国、米国で市場に現れ、電話系では、当時の電電公社によるANSERシステムの稼動が開始した
- 4) AI技術に対する期待が高まり、Lisp、Prologなどの処理系、エキスパートシステムなど知識処理の研究開発が盛んに行われた
- 5) ARPAプロジェクトや国内の研究機関で連続音声認識の研究が行われた

時期であり、産業界も広く音声情報処理に注目しはじめた時期と言える。

当時、先行する他社に遅れて本分野に参入した当社は、

- ・ 単語音声認識装置で他社にキャッチアップし、さらに、差別化する製品を市場に投入する
- ・ 市場として最も大きいと考えられる電話系音声認識の分野に注力する
- ・ 総合電機メーカーとしての利点を活かし、工業、制御システムなどのマンマシンインタフェースとしての連続音声認識を開発する

表 1. 音声処理技術と応用分野(1980年代半ば)

技術	応用分野
自然会話音声理解	エキスパートシステム 自動翻訳
連続音声認識	プラント運転管理システム
単語音声認識	OA、FA
テキスト音声変換	自動朗読、自動翻訳
高能率音声伝送	衛星通信、企業内通信網、 移動無線
話者照合	入出門管理

を、短期的、中期的なターゲットとして、開発を進めようとした。当時の個人的メモでは、音声情報処理技術とその応用分野として、表1の技術と応用を重点項目としていた。

2.2. 開発事例

試作システムや個別システムを除くと、当社がはじめて市場に投入した音声認識製品は、登録型特定話者単語音声認識装置 M6951(1986)である。パタンマッチング演算のための専用LSIを開発したことで、当時としては大語彙であった1000単語の認識が可能で、また、いち早く話者適応化技術の開発を行い、特徴ベクトルのマッピングによる話者適応化モードを有することを特徴とするものであった。

1980年代後半には、通産省の補助事業として、原子力プラントの運転支援システムの開発を行い、エキスパートシステムとシミュレータに基づく質問応答型の運転支援システムの試作を行った。当時としては先進的であった大画面を用いたマルチウィンドウ表示と音声認識によるマンマシンシステムを構築し、ネットワーク記述の文法を用いて約300単語の連続単語音声認識を実現した[1,2]。



図 1. 原子力プラントマンマシンシステム

電話系音声認識では、1993年に研究所において試作システムを開発、発表し、95年には製品を市場に投入、日刊工業新聞社十大新製品を受賞した。このシステムは、独自の音響音素モデルにより電話系の音響変動にロバストな認識手法を実現したものである[3]。

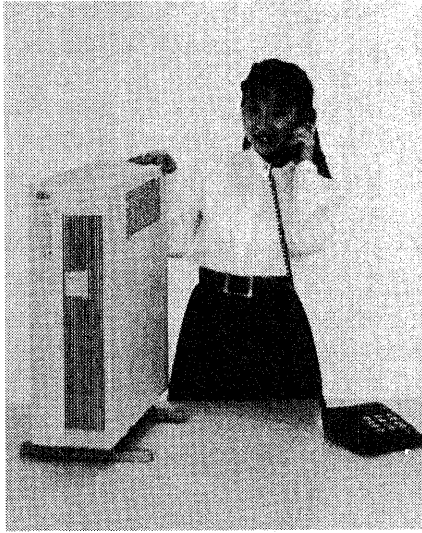


図2. 電話系音声認識システム MELAVIS

当初、unix ワークステーションに DSP を用いた専用ボードを搭載し実現していたこのシステムは、現在では PC サーバに電話回線制御ボードを搭載し、音声認識処理は PC サーバ上の S/W で実現し、低コスト化を実現している。

また、不要語への対処や、音声出力中の発声の認識を行うための出力音声のキャンセリング機能などを有し、実用性能の向上を図っている。現在、地方自治体の情報サービス、天気予報サービス、さらには、当社家電品の顧客電話窓口の1次受けなどに利用されている。

この他、テキスト音声合成技術による電話でのメールアクセスを可能としたユニファイドメッセージサーバ CallMail やハイウェイラジオなどの音声合成応用製品を市場に出し、さらには、音声符号化技術の開発を推進するため、音声符号化の研究開発を立ち上げ、独自方式による業務用通信システムや、携帯電話の製品化開発までを行ってきた。

2.3. 予想と現実

これまでの開発を振り返ってみると、CPU 性能、メモリ・ディスクなどの大容量化・低価格化など、技術の進展が我々の予想を越えるものもあったものの、音声認識、合成などの実用化については、我々の期待・予想を現状は下回るものとなっている。この原因については、過去にも、「なぜ音声

認識が使われないか」という議論が行われ、すでに多くの意見知見が述べられている[4]。ここでは、音声認識の性能など、指摘されてきている繰り返しはなるべく避け、筆者が考える問題点や、開発にあたっての予想と現実の相違について、以下に示す。

2.3.1. 利用技術

まず、離散単語音声認識においては、よく指摘されることではあるが、インタフェースとしての有効性、効率などについての考え方がやや漠然としていたことが普及を妨げたひとつの要因と考えられる。これは、音声に限ったことではないが、一般には設計者の感覚でのみ、インタフェースが設計されることが多く、明確な指針、評価尺度、設計原則に基づく設計技術がないことが大きな問題となる。実際に構築されたシステムが音声認識の有効性が活かせていない場合もあり、これからの音声処理技術の課題でもある。

2.3.2. 見えない技術

一般の利用者にとっては、音声認識を理解する上では常に、「人間」の能力との比較で考えられてしまう。上記の利用技術と関連するものであるが、利用者や応用システム開発者が音声であることに対して抱く期待が、実は、インタフェースの期待ではなく、システムそのものの機能向上に対する期待であることはよく体験するものであった。また、設計者の思想とユーザの思想とのミスマッチも多く起こり、ユーザにとり「わからないもの」になってしまうことに、我々が解決策を提供できなかったことも大きな問題であった。

2.3.3. アプリケーション

1980年代の前半においては、音声情報処理の手法としての期待だけでなく、音声インタフェースを適用する各種のエキスパートシステムなどの出現に対する高い期待があった。GUIの進展もあるが、コンピュータの利用やアプリケーションについても予想とは異なる結果となった。

3. 現在の研究開発

3.1. 開発課題

開発課題は、多くの研究開発機関が進めていることと同様、利便性の高い音声認識技術の確立であり、大別すれば、1) 外部騒音などの変動要因に対してロバストな音声認識アルゴリズムの開発、

2) 自由発話音声理解など音声認識の制約を緩和する音声理解・対話技術の開発、3) マンマシンインタフェースとしての効率、有効性の評価方法の開発、対話システムにおけるインタフェース設計技術・対話管理技術の開発、4) マルチモーダルインタフェースの開発、5) 品質が高く多様な音声合成、文生成技術の開発、6) 音声処理の新たな利用分野の模索、などである。

3.2. 何が問題となるか

ここでは、今後の音声情報処理技術の研究開発における問題点についての私見を述べる。

3.2.1. 音声認識の評価

音声認識の性能を聞かれることが多い。しかし、言うまでもなく、どう答えるのかは極めて困難な問題である。それは、マン、マシンインタフェースの性能の評価尺度が明確でなく、音声認識の認識率とその関係、すなわち、音声認識の性能がどの程度であれば、システムの性能としてどのような性能が期待できるかを予測すること、逆にあるシステム性能を実現するためには音声認識にどの程度の性能を要求すればいいかの関係が明確になっていないことを示している[5]。また前述のようにインタフェースの設計基準も明確でなく、特に GUI がシステムの機能（メニュー）をユーザに提示するのに対し、音声認識はシステムの機能とユーザの期待が一致していることを前提としており、音声対話というインタフェースにおける自己記述性や一貫性をどう持たせるかは今後の大きな課題であろう。

3.2.2. システムの価値

上記の評価と関連するが、従来型の研究開発であれば、技術的に高度であることとユーザの利益は大きく言えば一致しており、技術的に優れたものを開発することが利益を生むと考えればよかった。しかし、現在の情報通信では、システムあるいはサービスの価値は、必ずしも技術的価値、性能とは一致しない、企業においては、どのようなビジネスの形態を考え、その実現に必要な技術を開発するか、ニーズ、サービス側からの戦略が以前より強く要求されるようになってきている。この技術とユーザベネフィットの間を強く結びつけるには、コンセプトメイキング、ビジネスインキューベーション、マーケティングから技術までが効率的に機能しなくてはならない。誰がどう戦略を立

てて活動をコントロールするかが極めて重要となっている。

3.2.3. ゴール設定・動機

音声認識や合成技術技術は、「認識率」という連続的な尺度、あるいは音声の「品質」といった主観的評価が用いられることに加え、人間を最終ゴールとした時の技術と最終ゴールとの距離の大きさから、ともすると開発担当者が永遠の課題として技術をとらえ、明確な目標意識を持ってなくなる傾向が見られる。どのようにゴールを設定するのかは、単にそれにより実現できるシステムやサービスイメージを明確にするという面だけではなく、研究開発担当者のモチベーションとの関連からも、極めて重要な課題である。目標設定の的確性、その到達システムイメージの明確化と共有が、組織としての成果創出の大きなかぎとなるであろう。

4. まとめ

三菱電機におけるこれまでの音声情報処理分野の研究開発を振り返り、今後の研究開発の推進における問題点について論じた。上にあげた問題点のみではなく、情報処理・通信という分野におけるデファクトや集中化など困難な問題も多く、また個々の企業を超えて技術の進展と普及に向けての活動も必要である。高度で利用しやすいシステムの実現に向けて、研究開発を継続する予定である。

【参考文献】

- [1] Ishikawa, Y. and Nakajima, K., : A real time connected word recognition system, Trans of ICPR 90, pp.215-217 (1990)
- [2] 藤井他：原子力発電プラントマンマシンシステム (MMS) II, 三菱電機技報, Vol.63, No.7, pp69-75 (1989)
- [3] Mizuta, S. and Nakajima, K. :A discriminative training method for continuous mixture density HMMs and its implementation to recognize noisy speech, J. Acoust. Soc. Jpn.(E) Vol.13, No.6, pp.389-393 (1992)
- [4] 嵯峨山：なぜ音声認識は使われないか・どうすれば使われるか?、情処研究報告, 94-SLP-1, pp23-30 (1994)
- [5] 石川：音声対話システムの評価法, 音響学会誌 Vol.54, No.11 pp.807-811 (1998)