

TM-0304

知的音声インターフェイスの
研究開発について－現状と展望－

ICOT 昭和61年度 音声理解システム(SUS)
・ワーキング・グループ

July, 1987

©1987, ICOT

ICOT

Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan

(03) 456-3191-5
Telex ICOT J32964

Institute for New Generation Computer Technology

知的音声インターフェースの研究開発について
－ 現状と展望 －

新世代コンピューター技術開発機構
昭和 61 年度
音声理解システム・ワーキング・グループ

音声理解システム・ワーキング・グループ

主査	太田 耕三	電子技術総合研究所 パターン情報部
委員	阿部 芳春	三菱電機 情報電子研究所 情報伝送開発部
	板倉 文忠	名古屋大学 工学部
	桐谷 滋	東京大学 医学部
	小松 昭男	日立製作所 中央研究所 第六部
	坂井 邦夫	東京芝浦電気 総合研究所 情報システム研究所
	白井 克彦	早稲田大学 理工学部
	千葉 成美	日本電気 C&C システム研究所
	藤崎 博也	東京大学 工学部
	藤田 孝弥	富士通研究所 情報処理研究部門 パターン研究部
	田中 厚夫	シャープ 情報システム研究所 第一開発部
	真船 裕雄	松下通信工業 開発研究所
	山田 興三	沖電気工業 基盤技術研究所 情報機器研究部

目次

[0] はじめに

[1] 研究開発の経過について

1.1 '70 前後

1.2 初期の Project

1.3 技術の水準

1.4 普及の現状

1.5 最近の関連 Project

1.6 年表

[2] 問題の認識について

2.1 先端音声研究の動向から

2.2 情報処理技術の動向から

[3] 重要な研究開発分野と課題について

3.1 開発を進めるべき分野

3.2 音響および音韻レベルの課題

3.3 知的音声インターフェースの開発

3.4 音声言語学その他の課題

[4] 動機と意義について

4.1 技術的観点から

4.2 政策的観点から

[5] まとめ

知的音声インターフェースの研究開発について

ー 現状と展望 ー

[0] はじめに

音声通信の仕組みは、古くから人間の興味の対象であった。振り返れば、実に古代インドにまで遡ると言う。最近では、通信システムや電子計算機システムが発展し、トータルな情報処理システムを考える中で、とりわけ興味深い研究課題として注目されている。そしてこの方面からも盛んに研究が行われるようになっていく。

自然言語によるマンマシンインターフェースでは、音声がキイメディアである。十年のスパンでは、音声認識の先端実用レベルは、話者適応型・数千語の離散発声の単語音声認識システムであろう。構文的な制約を付ければ、このスペックを満足する連続音声認識システムとなるであろう。

社会の知識情報化は急速に進む。'90年代のオフィスオートメーションでは、好むと好まざるとにかかわらず、自然言語の利用は大勢となろう。この様な見通しは、FGCS Project 開始時点のコンセンサスの一つであった。高度な音声情報処理システムの実現が待たれている。

昭和59年春、ICOT が「音声懇話会」を主催した。この時には、音声関連の学識経験者、第5世代プロジェクト関連の 8 社 から 8 名、および主催者側関係者が集り、最近の音声研究の動向やその在り方などについて、様々の角度から自由に意見を交換した。この間の議論は、「懇話会メモ」に詳しい。

昭和60年からは、この会が FGCS Project 活動に組み込まれ、SUS-WG として発足した。以後、委員各位の意見の交換を積み重ねて、ほぼ一年が経過した。その結果、音声ならびにその関連研究者が、常々音声研究について抱いている種々の事柄、たとえば音声研究の現状を認識し、将来を展望するについて、技術的な問題点、またその解決に取り組む場合の運営上の問題点などに関して、大よそ見解の一致するところが明らかになってきた。

このレポートは、音声の主に認識技術に係わる部分について、その骨組みを手短に整理したものである。音声の専門家は勿論、音声に関心をお持ちの色々な方面の専門家にも、検討の糸口としてお読みいただければ幸いである。

[1] 研究開発の経過について

1.1 '70 前後

この時期には、今日の音声認識および合成に関する基礎理論となっている幾つかのアルゴリズムが新しくこの分野へ導入された。音声波形の線形予測法（音声スペクトルの最ゆう推定法）、Dynamic Programming である。

前者は更に、PARCOR 分析法、線スペクトル対、声道形推定法などへ発展し、それ以前のフーリエ分析、ケプストラム分析、A-b-S 法などとともに、多くの基礎的、実際的成果をもたらしている。

後者は、特定話者、小語彙、離散発声の単語認識（IWR: Isolated Word Recognition）機能を備えたプライマリな音声認識システムを、実用機として普及させるのに重要な役割を担った。

今日では、この時期に導入された上述のアルゴリズムが、音声研究の学問的、実際的な枠組みを広げかつ奥行きを深めるのに欠かせない基礎的道具となって広く活用されている。

1.2 初期の Project

基礎研究の新たな立上りを背景に、'70 年から '80 年にかけて、注目すべき National Project が、我が国および米国で計画・実行された。音声関連では世界に先駆けた国家的計画で、歴史的には第 1 期に位置付けられる。それぞれの狙いと成果の概略を以下に述べる。

米国国防省 ARPA の下で、1971～1976 に SUS Project (Speech Understanding System) が実施された。目標は、多数話者の発声する 1000 語の連続音声を、実時間の数倍程度で 90% 以上理解することである。(タスクの限定、文法の限定、話者適応手続きを許す)。CMU、SDC、BBN がそれぞれ独自に開発を展開し、CMU の Harpy が目標を達成したとされる。

我が国通産省の下で、パターン情報を処理し得る新しい世代の情報システムの開発を目指して、1971～1980 年に PIPS Project (Pattern Information Processing System) が実施された。音声認識関係のテーマは ETL と NEC が研究開発を担当した。声道形推定法など基礎的な音声分析技術が新しく開発され、座席予約をタスクとする離散発声 100 語の単語認識評価システムが開発された。

1.3 技術の水準

音声の認識や理解には、特定話者、限定語彙の IWR システムから、不特定話者、大語彙の会話音声の理解システムまで、無数の技術レベルが考えられる。比較的プリミティブな技術課題として、特定話者・小語彙の IWR があるが、現状では、100 語程度の IWR 技術が実用水準にある。

また特定話者・小語彙 IWR 技術の延長線上にあるやゝ高度な課題として、特定話者・中語彙 IWR がある。IWR 技術をもとに、HMM (Hidden Markov Model)、ベクトル量子化、n-gram 法、あるいはワープロ向け日本語処理技術などと組合せ、国内外で数百～数千語の IWR システムが実用化の一步手前にある。

一方、音声の合成技術は、編集合成と規則合成に大別され、前者から後者の順でより高度な技術が要求される。音質に関する様々な立場の不満もあるが、編集合成における録音編集は既に実用期にあり、マルチパルス、LSP、PARCOR など波形符号化あるいはスペクトル符号化に基づく分析合成技術は、LSI 化を含めて実用水準に達しつつある。小型化、低価格化にはデバイス技術の寄与が大きい。

1.4 普及の現状

'70 年代末、LPC 合成チップを内蔵した TI 社の廉価な玩具 Speak & Spell がヒット商品になった。これは、音声入出力機器の持つ豊かな市場性に対する一般の関心を高める一つの契機になっている。SUS Project、PIPS Project、これらと並行に進められてきた基礎的研究開発など、積極的な先行投資と努力によってトータルにもたらされた注目すべき成果と言える。

ここ数年の間に、CODEC 等の音声応用機器向け需要のみならず、マンマシンインターフェースとして直接の応用をめざしたシステムが、前に述べた実用水準にある音声認識合成技術をベースにして現場に導入され始めている。

なかでも、流通業における入出荷業務、物流業における集配業務、鉄工その他製造業における検査業務などにおいて、現行技術の枠内ではあるが、音声通信の特性を巧に應用し、経営収支に直接結びつく顕著な成功をおさめている例も報告され、徐々にではあるが、はっきりした普及の進展が認められる。

音声出力では、64 Kb/s の PCM が早くから標準化され、32 Kb/s の ADPCM も最近標準化された。これらの技術は、公衆電話サービス、たとえば銀行預金の残高照会などを行うバンキングシステムなどで広く普及している。分析合成技術も、その低コスト化により、家庭用機器、玩具、OA 機器、警報システムなどで普及が進んでいる。

1.5 最近の関連 Project

米国国防省 ARPA は、国防力と経済力強化のための広範なマシン・インテリジェンス技術基盤を開発する目的の下で、SCS Project (Strategic Computing and Survivability Programme) を進めている。音声関連技術では、5000～10000語の実時間連続音声認識技術の開発、ならびにこれの戦闘機あるいは電話回線への応用を目標としている。

EC 加盟国は、情報関連技術分野の産業的ポテンシャルを欧州レベルで創造・強化する目的の下で、Esprit 計画 (European Strategic Programme for R&D in Information Technologies) を提案している。重要開発 5 分野のうち、音声関連技術は、先進的情報処理技術および OA 技術開発に組み込まれ、不特定話者、離散発声の単語認識技術の開発、ならびにこれを Workstation Speech Recognition として実現することを目標としている。

Alvey Report (1982) が、今後英国が開発に力を入れるべき技術分野として、ソフトウェアエンジニアリングなど 4 分野を指摘した。音声関連技術は、語彙に制限はなく、しかも広範囲な話者のフルーエントな音声の認識技術の開発と、その実証大規模デモプロジェクトとして、OA 用の音声ワークステーションの開発を目標にしている。

我が国では、郵政省の努力で、国際電気通信基礎研究所の設立があり、ATR 自動翻訳電話研究所が発足した (1986)。自動翻訳電話システムの開発を目標とする。音声認識関連の今後 7 年間の主テーマとして、文節ごとに区切って発声した約 3000 語の日本語および英語の会話音声認識システムの開発が挙げられている。また同年通産省の努力により、日本電子化辞書研究所が設立された。自然言語処理用電子化辞書の開発が目標である。

1.6 年表

音声情報処理研究の経過を示す簡単な年表を図 1 に示す。'70 年前後からその後期にかけては、比較的基礎的な成果が多いことがわかる。（花マルで囲んである）。年代軸の上側は我が国、下側が米国の成果を示す。初期のプロジェクトは、この時期に実施されている。

DP100 や Speak & Spell などは '80 年前後に市場に出た。（商品は長マルで囲んである）。基礎的成果から約十年が経過している。認識システムとしては HARPY が先駆的である。

'80 年前半で目立つのは、実験室レベルの大語彙単語音声認識システムである。（六角形で囲んである）。このあたりは、第 11 期のプロジェクト計画のスタートの時期でもある。SCS プログラムは、SUS プロジェクト終了後ほぼ 10 年、ATR は、PIPS 終了後 6 年、FGCS プロジェクトに遅くれること夫々 2 年と 4 年である。

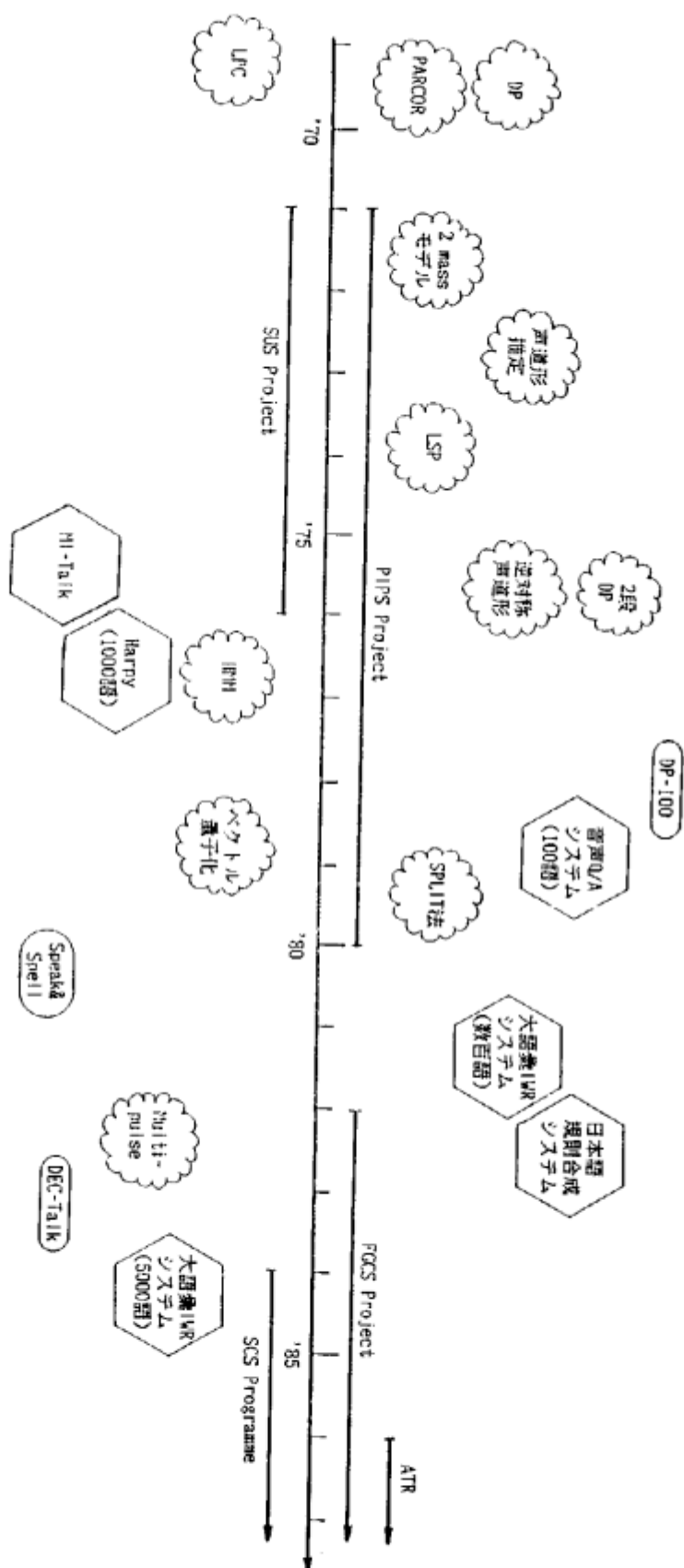


図 1 音声研究の経過

(音声研究動向が大体分かる程度に、成果としてあるいは形として比較的是っきり見える項目だけが挙げられている。基礎的な研究のなかには、必ずしも明確な項目として挙げにくいものも多い。)

[2] 問題の認識について

2.1 先端音声研究の動向から

特定話者・小語彙の単語認識は、ほぼ確実な技術として定着した。しかし、IWR システムの普及は、一方で周囲雑音、語彙の少なさや話者依存性、さらには離散発声の制約とともに、非知的な音声入力技術の限界を示す役割も果たした。音声インターフェースが真に役立つには、話者に課せられる諸々の制約が少なく、同時にこれが相当高度に知的である必要がある。

従って、最近の先端研究では、このような音声認識技術に対する本質的なニーズが従来より遙かに強く意識され、前章 1.5 でも触れたように、不特定話者の大語彙単語認識や連続音声認識の実現を意図するものが多い。しかしながら、不特定話者の対話あるいは文章理解システムはもとより、特定話者の大語彙単語音声認識システムも、単に従来の技術の延長線上には有り得ないことが判明している。

例えば、特定話者、小語いの IWR 技術を不特定話者に適用するには、話者変動に対する効果的な適応技術の開発が必要であるし、もっと直接的には正規化原理の確立が望まれる。また特定話者でも、大語いの IWR システムや連続音声認識（CSR：Continuous Speech Recognition）システムを実現するには、高度な言語処理技術は勿論、まず音素や音節を認識の単位とする認識技術の開発が必要である。いずれも音声通信のメカニズムに深く立ち入った考察が必要であり、既存のホールワードマッチング技術の枠では捉え切れない。

話者変動の正規化原理が解明されるならば、既存の IWR システムの性能も著しく安定したものとなり、市場性も格段に向上しよう。大語彙の IWR システムや CSR システムにおいても事情は同様である。最近のスペクトログラム・リーディングなど音素ベースのアプローチの提唱も、この様な問題意識と見通しの上に立っている。

次世代の知的音声インターフェースシステムの実現を目指して、より本質的な課題が改めて示されていると理解される。音韻とは何かを問題の正面に置き、音響レベルあるいは音韻レベルなど、音声通信の低次過程に含まれている課題の直接的な解決を目指す研究を進めることが必要である。

2.2 情報処理技術との関係から

知能の問題は、言語と知識の問題でもある。機械に知能を持たせるには、その前提として自然言語の理解モデルやこれに関する知識表現技術が不可欠である。知識情報処理技術は、情報処理技術開発の目下の主要関心事のひとつであり、様々な興味深い側面を持っている。その一つが言語および知識との関わりであることは言うまでもない。

音声タイプライターや談話理解システムは夢の機械である。知能技術の現状から考えて、これらの機械の実現のためには、知識情報処理技術の格段の進歩が前提になろう。さらにこれらの基盤を支える技術も必要である。マンマシンコミュニケーションの社会的な経験と成熟も必要である。真に知的な音声インターフェースは、この様に遠大かつ挑戦的な技術開発と進歩の末だその先に位置付けられるのではなかろうか。それほど高度な技術と広い知識の組み込みが要求される。しかし、情報化社会の赴くところ、これは実現されるべき機械である。

電子計算機素子が、真空管からトランジスタへ、また VLSI へと発展したように、知識情報処理技術もまた着実に発展しよう。この発展があるかぎり、機械の知能も、幼児から少年へ、また青年へと成長する。音声インターフェースにも、大勢として、その時々の知識情報処理技術に準じるレベルのインテリジェンスが期待できるのである。

次世代の音声インターフェースは知的である必要がある。従って、最新の知識情報処理技術の成果を積極的に導入するとともに、関係研究グループとの連絡を強めることにより、最高のマシンインテリジェンスおよびその基盤技術と、最先端の音声分析技術を統合した音声認識応用システムの開発の試みが一方でまた必要である。

〔3〕 重要な研究開発分野と課題について

3.1 研究開発を進めるべき分野

音声関連技術が直面している課題を解決するという観点からは、主として次の二つの分野に対する集中的な技術開発が当面有効と考えられる。記述を簡潔にするため、音声談話理解過程（音声談話理解システム）を、音響および音韻レベル（Front End System）と 統語および意味レベル（Back End System）の二層に便宜的に分けた。

分野1では、音声の波形分析技術あるいは音韻の特徴抽出に関する技術の開発を行なう。ここでは、音声生成および聴取における下位過程を記述するより本質的な構造の発見とモデルの開発、またこれらのモデルによる音声過程下位レベル（音響および音韻レベル）に関する知識の蓄積・整備が中心的興味となる。

分野2では、知的な音声認識応答システムを開発する。言い換えれば、知識情報処理の成果を利用した音声インターフェースシステムの開発である。

ここでは、特定の文あるいは談話理解モデルの評価あるいは開発、特定タスクと言語理解モデルとの整合性の評価、そして、これの開発過程に予測される諸問題の解決、つまり Back End System に関する現在あるいは近い将来利用可能な技術（推論マシン、知識ベースマシン等）の先見的かつ本質的機能の活用と Front End System との融合技術の開発などが中心的興味となる。

先にも述べたように、Back End System に関する技術は、知識情報処理技術開発の主テーマでもある。ここでは、近い将来に、かなり高度な Back End System の支援体制基盤が、こちらの方面から整えられる事を想定している。

分野2の認識応答システムの性能は、分野1の成果や、FGCS における知識情報処理技術開発の成果に相当依存する。それぞれの技術が無理なく統合されるならば、知的音声インターフェースを開発する意義は大きい。

3.2 音響および音韻レベルの課題（分野1）

3.2.1 新しい音声分析法の開発

1. 非定常態音声分析法（子音分析法）の開発
2. 逆対称性のような特性の音声分析への応用
3. スペクトル包絡情報のより高度な記述パラメータの開発
4. 音源周期性情報のより高度な記述パラメータの開発
5. 聴覚系の距離尺度のアルゴリズム化
6. 発声器官の動的モデルの開発

3.2.2 認識単位の研究

1. Computer Phonology の構築
2. 調音結合等の規則表現と記述単位依存性の検討
3. 種々の認識単位による認識アルゴリズムの開発とその性能評価
4. Segmentation 技術とその誤り修正技術の確立

3.2.3 個人差の定式化

1. 声質の変動と自動認識適合性の検討
2. 話者正規化メカニズムあるいは適応化メカニズムの解明

3.2.4 韻律の研究

1. 音声通信における韻律情報の構造と役割の解明（韻律規則の解明）
2. 会話文固有の音調やリズムの解明
3. 高品質音声の合成

3.2.5 音声データベースシステムの開発

3.3 知的音声インターフェースの開発（分野2）

3.3.1 DUALS と PSI

第5世代コンピューター開発では、中期後半に本格的な日本語文章処理システムを開発する。DUALS は、構文解析、文生成のための標準の言語解析ツール（LTB）を用いて開発中の意味解析・文脈解析システムであり、逐次型推論機械PSI上で開発が進められている。また今後は日本語の句構造文法理論（JPSG）のアイデアも取入れられる予定である。DUALS V3 は、2000 語程度の辞書と500ほどの文法規則を備え、約300の文章を対象に、PSI-IIにより、1～2 秒／文の処理能力を達成する見通しである。知的音声インターフェースの開発においても、積極的な活用が考えられるべきである。

3.3.2 FGCS における開発目標例

しかし、この場合に限らず、一般に言語処理研究では、日本語の Keyboard 入力が想定されており、入力音節列自体のあいまいさには関心が払われていない。従って、FGCS において、上のような日本語処理システムを想定し、これに音声でインターフェースするには、例えば出力レベルを単語として、出力単語列のあいまいさを極力減じて置かねばならない。このためには、あいまいさを含む言語処理技術の開発とともに、分野1における課題の解決が重要である。

FGCS Project における音声関連技術開発の目標は、例えば前述の様な日本語処理機能に対して、音声でインターフェースを可能にする事とし、そのために必要な基礎技術開発ならびに音声認識合成の応用システムを開発する事とする。これを以て将来の汎用マンマシンコミュニケーション技術の基礎とする。
図2に応用システムのイメージを示す。

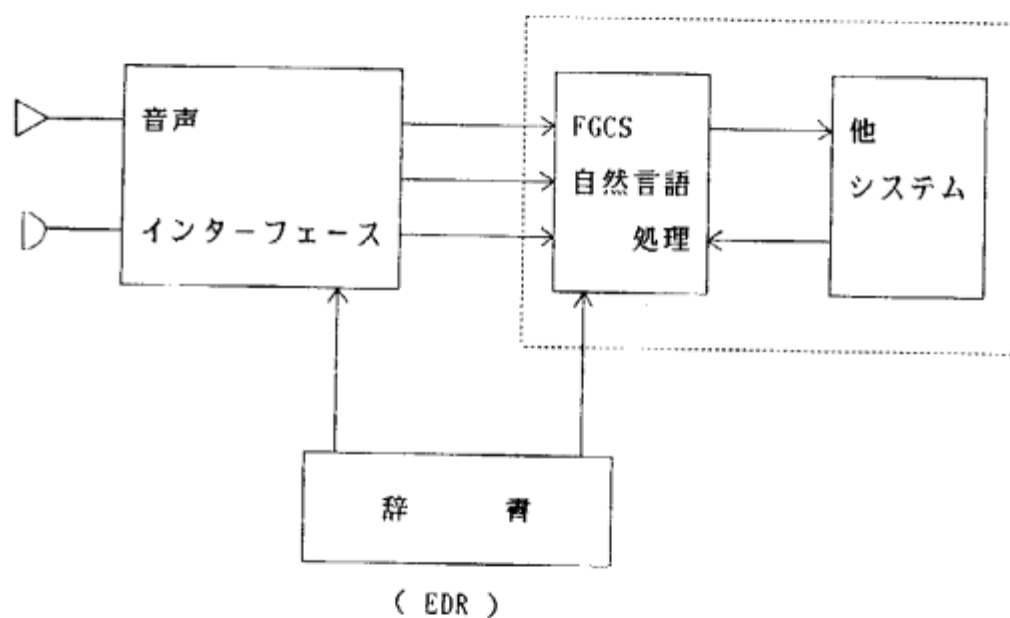


図 2. 音声インターフェースのイメージ

音声インターフェースの機能は、音声分析を行い、基本単語の認識を行い、若干の意味解釈とともに、可能な単語候補列の出力を行う。自然言語処理部あるいはタスクに依存する部分は、やや抽象的な概念情報を指標として、音声インターフェース部と情報交換を行う。

3.4 音声言語学その他の課題

3.4.1 音声言語処理の研究

- 1 会話進行過程の分析とモデル化（語彙領域の動的制御法の開発）
- 2 音声談話理解におけるキーワードの役割解明
- 3 音声言語表現の生成とモデル化

3.4.2 あいまいさを含んだ言語処理の研究

- 1 誤りを含んだ音声記号列からの単語／文節識別抽出技術の確立

3.4.3 応用システムの研究

- 1 会話型コンサルテーション／エキスパートシステムの開発
8万語＋専門用語、法律相談、税務相談など
- 2 音声対話型作業補助システムの開発
設計作業、事務作業など

3.4.4 研究用システムの開発

- 1 音声言語研究用計算機システムの開発
- 2 音声言語理解知識ベースシステムの開発

[4] 動機と意義について

4.1 技術的観点から

'70 年代の前半、音声や文字研究などに使われた中型計算機の演算速度は 0.1 MIPS 程度、主記憶容量は 数十～百数十 KB 程度であり、大型で 1 MIPS、数百 KB 程度であった。いま技術計算向け WS (Work Station) は 1～3 MIPS、数十 MB 程度、大型は 10～20 MIPS、数百 MB を越える。近い将来、このレベルは、現在の WS 並のコンパクトさで提供されるものと考えられる。

信号処理、マッチング、サーチ演算など、音声インターフェースシステムに要する計算量は非常に多い。1000 語の連続音声認識システムを、実時間に近い早さで動作させるには、100 MIPS が必要との試算もある。PSI-II は 200 KLIPS を達成する予定である。計算機的能力は、最近ようやく本格的な知的音声インターフェース研究に耐えるレベルに近づきつつある。

自然言語研究は、文の表現形式の高度化、取り扱う対象範囲の拡大、LISP マシンや PROLOG マシンの開発など、積木世界を対象にした自然言語 QA システムが開発された '70 年代初頭の状況と比較して、理論的枠組みや基盤ツールの整理が相当進んで来ている。その成果は、特に機械翻訳の分野に顕著である。知的音声インターフェースシステムを実現する基盤の一つが整って来ているものと理解される。

始めに述べたように、'90 年代後半には、数千語の IWR システムや制限付きの CSR システムも本格的になろう。これに伴って音声によるマンマシンコミュニケーションの社会的な成熟が間違いなく一段進行する。次の段階での需要は、ますます不特定話者・数万語の CSR システムや対話システムに向かうと考えられる。この様な動向を見越した音声情報処理技術の開発が必要である。

4.2 政策的観点から

'90年代のオフィスオートメーションにおいては、自然言語の利用は不可欠になるだろう。既に述べたような問題の設定に基づく研究開発が、具体的に課題の解決に結びつくならば、単に現在の音声情報処理の技術水準を向上させるだけでなく、既存の情報化社会が抱えている課題の解決に役立つ。なぜなら音声情報処理技術の発展は知識情報処理技術と不可分であり、知識情報処理は、社会の急激かつ大規模な情報化に起因する問題の解決を意図する技術であるからである。

一方、我が国が将来も国際社会において国威を維持・発展させるには、卓越した知識情報化の達成を基本戦略としなければならない。先進諸国においても、勿論同様の戦略が採られよう。国際社会の知識情報化が確実に進む。関連技術力の育成・向上の努力は、我が国の内と外における関連市場の拡大と安定をもたらすことになる。

ここでは、マンマシンコミュニケーションに関する最先端技術力を、上記戦略の一環として、常に我が国において育成・向上させることの重要性が理解されなければならない。同時に音声情報処理技術ならびにパターン情報処理技術を包含した次世代の知識情報処理が、その不可欠の基盤技術となる事が認識されなければならない。

音声情報処理技術にしろ、現在の FGSC の核技術にしろ、完成に致るまではともに解決が非常に困難な課題を含む。しかし、技術立国のコンセンサスの下では、それだけに、真に解決すべき課題を正直に採り上げ、積極的・組織的な研究推進によって、その解決へ努力が払われるのは当然である。それはまた、人類の長年の夢の実現を目指すことでもある。

[5] まとめ

音声情報処理研究の向かうところは、知識情報処理研究の向かうところにある。FGCSは、その基盤システムである。この技術は、さらに認知科学やバイオサイエンスなどの相互の影響の下に、新世代マンマシンコミュニケーションシステムへ発展しよう。音声研究は、知識情報化社会のおもむくところ是非とも解決が図られるべきテーマである。

優れた音声インターフェースを開発するには、音声生成知覚の下位過程における現象の解明と適確なモデルの開発、ならびにこれに基づく知識の研究と蓄積を進める必要がある。同時に、知識情報処理技術からの有効な成果を積極的に導入し、かつ発展させ、高度なマシンインテリジェンスを備えた音声認識応答システムを開発する必要がある。

この様な方向で、より革新的な成果を挙げ、我が国の技術基盤を強化するには、解決すべき事柄をひとつ一つ解決し、一步一步の着実な研究努力は勿論、それを可能にする巾のある計画とそれに相応しい支援体制の整備が必要であろう。コストパフォーマンスの問題はあるが、この様な要因も、当該技術開発におけるBreakthrough に有効である。

ここでは、テーマの内容に応じた柔軟な研究開発計画が必要である。分野1は、本稿3.2に示されているように、かなり基礎的な研究課題を含む。従って、この分野に対しては、比較的長期に渉る基礎研究重視型の研究計画が立てられるのが望ましい。

分野2は、本稿3.3に示されているようにシステム開発である。これに対しては、短期的な開発型計画で対処するのが望ましい。この場合、目標システムは、これを作り上げることがポイントである。

音声言語に関する課題には、統語構造中心の言語処理技術の枠組みの中に、音声のたとえば韻律構造を組み込むこともあろう。しかし現状では、両者の関係づけに役立つ知識が不十分である。音響分析に基づく関連知識の蓄積は重要であるが、課題の主なところは、当面は現行の自然言語処理技術の課題の枠内にある。

支援体制が整えば、たとえば共通の新しいツールの提供（音響－記号レベルを有機的に処理するマシンや研究拠点間ネットワークなど）は、研究者間の情報交換の頻度と質を高め、問題への Intelligence の集約度を高める。また、新しい型のマシンは、音声研究に新しい視点を与えよう。これらは課題解決のスピードと糸口を増すことになる。

共用の音声データベースは、評価の公平さやオープンで自由な議論の場を提供し、グループ間の Culture Gap を埋める。研究風土の活性化にも役立つ。また共通データからの様々の結果の比較検討が容易になり、問題のポイントがより確かとなり、次の研究が集約的・効率的になる。

近年では、大学、国公立研究所のみならず、主要私企業においても、相当に基礎的視点からの技術開発努力が積み重ねられ、すでに高度の実力を蓄えた研究グループが幾つも育ちつつある。若い力の育成とともに、この豊富な Intelligence を更に発展させる事も大切である。

米国においては、既に千語レベルの中語彙認識システムが市場を形成しようとしている。近い将来、次世代のより高度なシステムに対する強いニーズが生ずるものと思われる。本稿〔3〕に述べた課題に関して、研究現場では新しい芽も見える。技術立国のコンセンサスの下で、適切な方向付け、組織的な協力と環境整備を図るならば、音声情報処理技術の将来は明るい。