

コンピュータによる音声・画像システム

Voice & Video System by Computer

人間が外界から受け取る情報の大半は視聴覚情報である。この視聴覚情報に対する人間の鋭い感覚を利用し、画像や音声の形式でコンピュータと対話できれば、コンピュータの優れた処理能力とあいまって理想的な情報サービスシステムを構築することが可能となる。

日立製作所は、従来からこの分野の技術開発、製品化に積極的に取り組んできた。既に音声応答システムは広く実用に供されており、更に音声入力、画像処理、画像応答システムを実現するデバイス及び利用技術の開発を行なって製品化の段階に至っている。

この論文では、上述したコンピュータ制御による視聴覚情報システムについて最近の成果を述べる。

橋本寿之* Hashimoto Hisayuki

辰己允邦* Tatsumi Nobukuni

中田和男** Nakata Kazuo

1 緒言

人間とコンピュータとのマンマシンインタフェースを考える上で、音声や画像は重要な役割を果たす。すなわち、音声や画像は人間の最も自然かつ素朴な情報表現、伝達、受容機能であり、音声や画像の形式で人間が直接コンピュータと対話できれば、コンピュータの優れた処理能力とあいまって理想的なシステムを構築することが可能となる。

音声や画像情報は本来アナログ形式であり、冗長度の多い膨大な情報量をもっていることから、これまでデジタル形式のコンピュータには不向きとされていた。しかし最近、音声認識や音声出力、画像処理や画像応答などの技術開発が活発に行なわれており、既に実用システムとして稼動している

ものも多い。

この論文では、日立製作所でのこの種のシステムの最近の開発状況について紹介する。

2 音声出力

2.1 構成

音声出力システムは、コンピュータの処理結果を人間に最も親しみやすい音声形式に変換して出力するシステムで、音声応答システムとして開発、製品化されている。

図1に音声応答システムの基本構成を示す。端末は通常押しボタンダイヤル電話機(プッシュホン)であり、広域の公衆

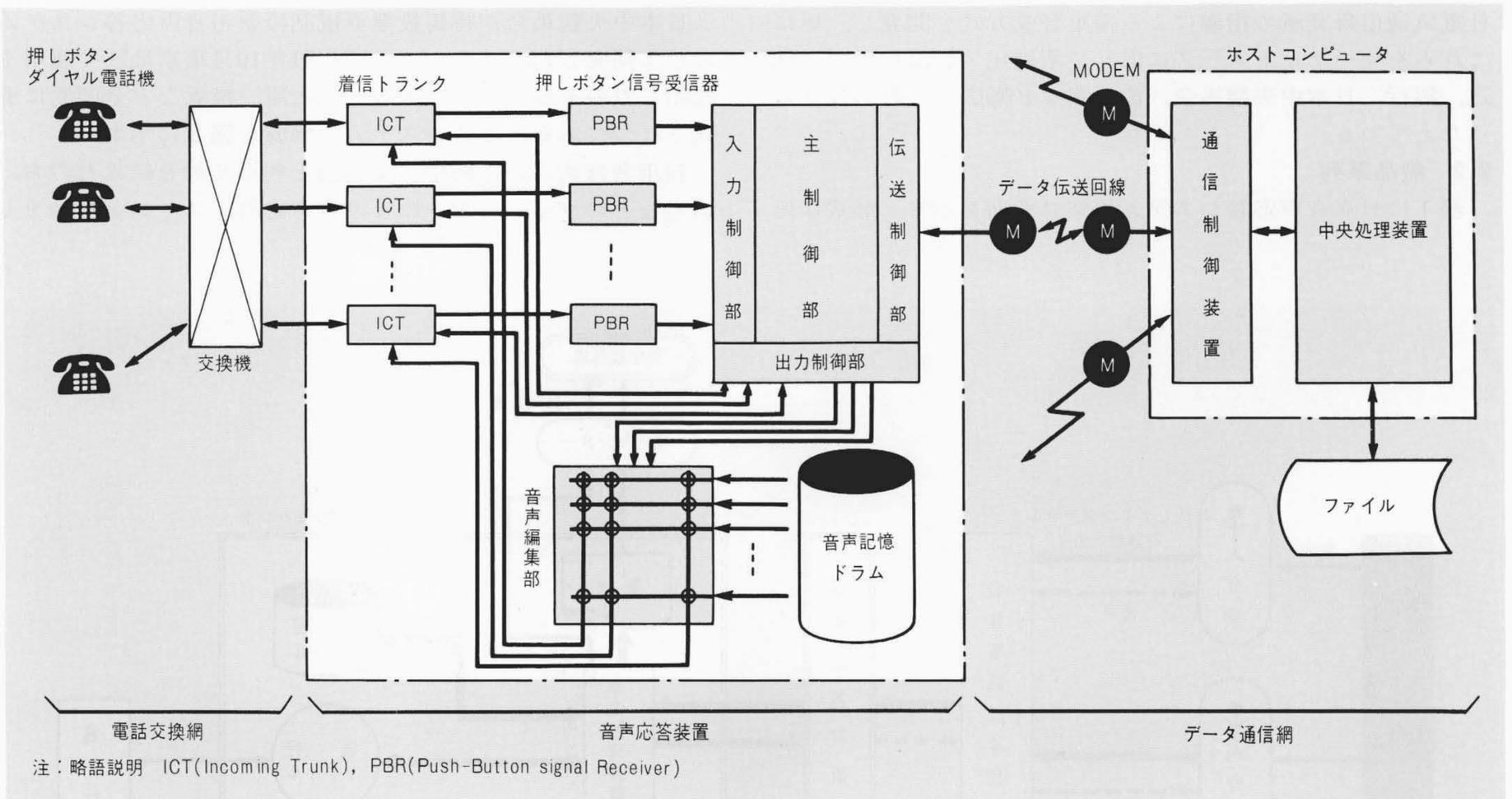


図1 音声応答システムの基本構成 電話交換網と中央処理装置などを含んだデータ通信網との中間に、音声応答装置を接続したシステム例を示す。

* 日立製作所戸塚工場 ** 日立製作所中央研究所 工学博士

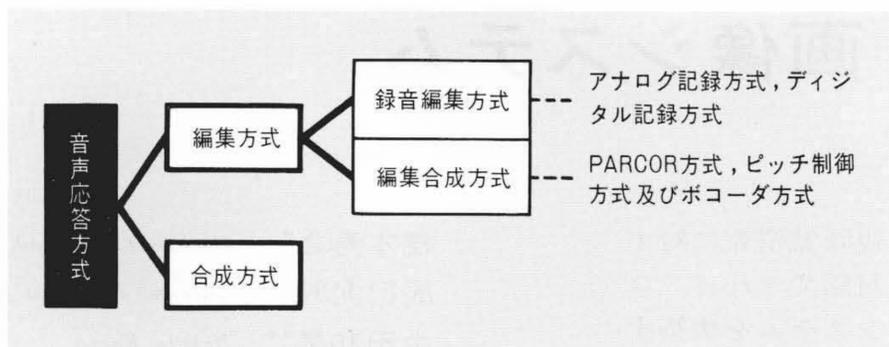


図2 音声応答方式 音声応答の方式分類を示す。

電話網により、どこからでも、だれでもコンピュータにアクセスし、その処理結果を音声で聞くことができる。

このような音声応答システムには、基本的に次の三つの条件が課せられる。

- (1) 了解性、明瞭性を十分もった高品質の音声が出力できること。
- (2) 多数の回線に同時に異なる内容の応答が可能で、システムを多重に利用して安価にサービスができること。
- (3) できるだけ多くの語彙を出力できること。

これらを実現する音声応答の方式には、図2に示す3方式がある。録音編集方式は、文節や単語、単音節などをあらかじめ記憶装置にデジタル形式で記録しておき、コンピュータの制御によりこれらを編集し、まとまった文章として音声出力する方式である。編集合成方式は、多数語彙の出力を目的とし、語彙の記録を効率的に行なうため、音声から抽出した特徴パラメータを記録しておき、コンピュータの制御によりこれらを編集したのち合成して、もとの音声に復元する方式である。合成方式は電気回路により音声を電氣的に純粋に合成する方式である。

日立製作所は、これまで録音編集方式と、日本電信電話公社電気通信研究所の指導による編集合成方式を開発し、更にこれらをマーケットニーズに応じて系列化・製品化して、鉄道、銀行、日本中央競馬会、流通業など幅広い分野の要望にこたえている。

2.2 製品系列

表1に日立音声応答システムの製品系列を示す。形式は処

表1 日立音声応答システムの製品系列 音声応答システムの製品系列を示す。3機種で全域をカバーしている。

項目	形式	H-1100	H-1200	H-1300
基本方式		録音編集方式		
記録方式		デジタル(PCM)方式		
音声素片		文節又は単語、単音節		
単位語長		0.5秒	0.8秒	1.1秒
音声記録媒体		固定ヘッド磁気ディスク		磁気ドラム
収納語数		768~1,536	256~2,048	512~2,048
同時応答回線数		16~32	128~256	128~768
情報の入力		プッシュホン*		
出力音声品質		極めて良好		

注：略語説明など

PCM(Pulse Code Modulation)

* 回転ダイヤル電話機の場合は、プッシュボタンアタッチメントをオプションとして利用できる。

理能力(同時応答回線数)、収納語数などにより階層的に構成し、広範囲なアプリケーションに柔軟に対応している。

2.3 適用例

このように音声応答システムは、プッシュホンにより手軽にだれでもコンピュータと会話ができるため、広い分野で利用されている。その代表例を次に示す。

(1) 予約業務

ホテルや病院の予約、鉄道や航空機などの座席予約などの業務への適用が考えられる。このうち最大規模のシステムは、日本国有鉄道が昭和50年3月に運用を開始した電話座席予約システム¹⁾で、「みどりの窓口」で活躍しているMARSシステムに音声応答システムを結合し、関東一円のプッシュホン加入者に対しプッシュホンからの座席予約サービスを提供するシステムである。このシステムの音声応答装置は、H-1300形を主体に構成している。

(2) 勝馬投票券電話投票システム²⁾

日本中央競馬会が勝馬投票券電話投票用音声応答システムとして開発したシステムで、昭和51年10月東京地区で運用を開始したのを皮切りに、大阪、名古屋、横浜など全国的に導入されつつある。このシステムの構成を図3に示す。

利用者はあらかじめ中央競馬会と利用契約を結んだのち、自宅などのプッシュホンにより音声応答システムと対話をし

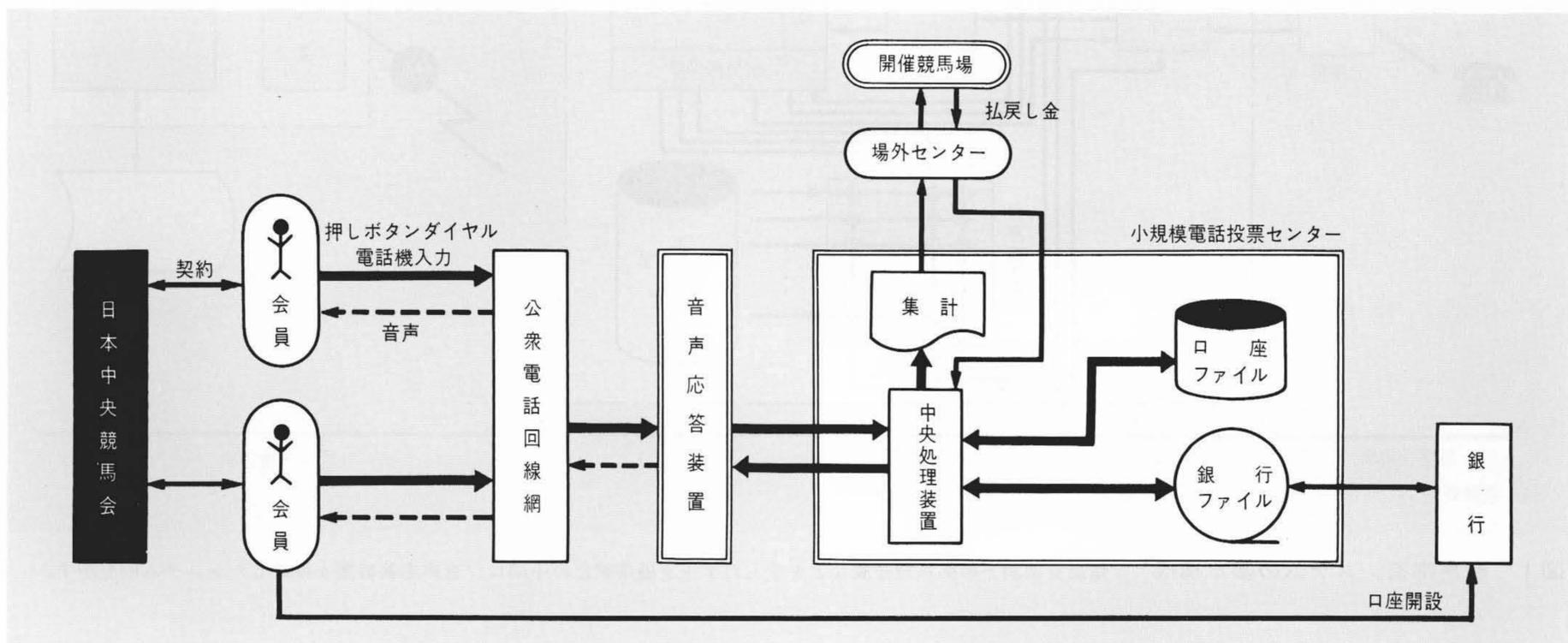


図3 勝馬投票券電話投票システムの仕組み 利用希望者は、日本中央競馬会と会員契約を結び、押しボタンダイヤル電話機により勝馬投票券の投票を行なう。

表3 音声入力装置の製品事例 音声入力装置は、発声話者、認識語数が主要パラメータである。

	HR-100 音声入力装置	HR-150 音声入力装置
発声話者	特定(登録者)	不特定
発声単位	離散発声	離散発声
認識語数	16 ~ 16×8単語	16 単語
入力機器	接話マイクロホン	電話機
確認機器	表示器又は音声(オプション)	音声(オプション)

語を選択識別し、結果をデジタルデータ形式で出力する。発声者は、表示器や音声応答により正しく認識されたことをモニタすることができる。

3.2 製品系列

音声入力装置は応用分野により、あらかじめ標準音声に登録した話者(特定話者)を対象とする特定話者方式と、だれの声でも認識できる不特定多数の話者を対象とする不特定話者方式の2種類がある。表3に各装置の諸元を示す。

HR-100音声入力装置は、特定話者に限定することにより経

表4 サービスの種類 画像応答システムのサービス形式を示す。

No.	サービス形式	概要	適用例
1	静止画サービス	センタ〜端末間の会話に基づいて写真、図形などの静止画と音声によってサービスを行なう。提供情報のほか、会話自体がサービスを構成する。	(1) 交通分野……………旅行案内、事故情報、道路交通情報 (2) 産業経済分野……………株価情報、市況情報、不動産情報
2	複合静止画サービス	静止画サービスに動画を挿入することにより、運動、動作、状態の変化などを伴う情報の提示を容易にする。	(3) ショッピング・飲食分野……………商品案内、特売情報、専門店案内、レストラン案内
3	ランダムアクセス動画サービス	静止画サービスによって動画番組の検索を行ない、要求番組のVTRカセットを自動的に選択して再生送出する。	(4) 福祉分野……………年金案内、公共施設案内 (5) 娯楽分野……………クイズ、ゲーム、映画
4	定時同報動画サービス	テレビジョン放送と同様に、あらかじめスケジュールされた番組を時間表に従って送出する。	(6) 教養・趣味分野……………囲碁、将棋、生花、園芸などの指導 (7) 医療分野……………健康相談、病院案内、育児相談 (8) 教育分野……………各種学校案内、進学情報、学習塾案内 (9) 報道分野……………ニュース、天気予報 () 学習分野……………資格取得講座、数学講座、英会話

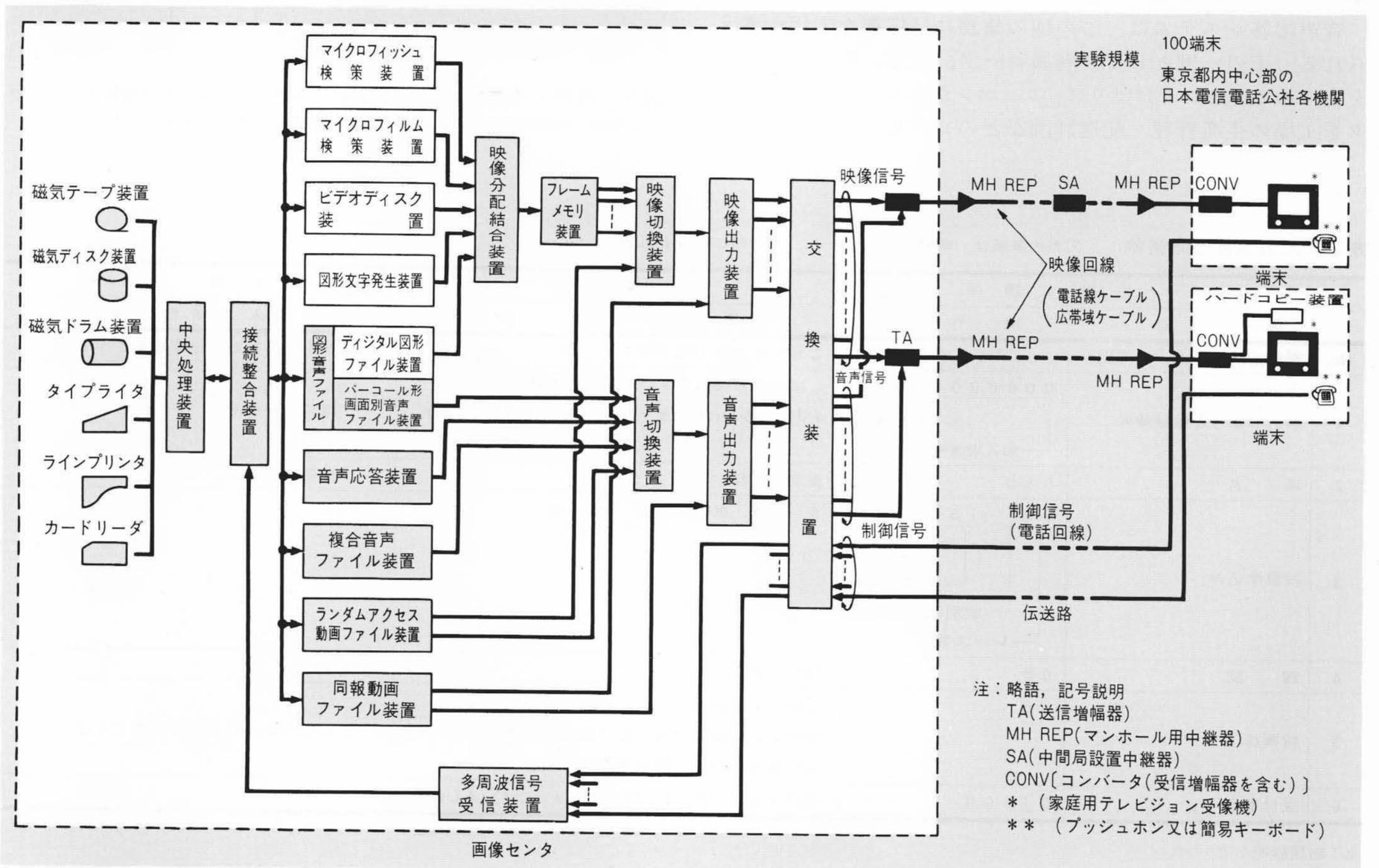


図6 実験システムの構成 VRSでは、画像センタを多数の端末で共同利用する。センタと端末とは映像回線及び制御回線から成る伝送線路で接続される。

済的で高い認識率を実現する装置で、接話マイクロホンを利用する。この装置は汎用コンピュータの音声入力装置に適用できるほか、生産・品質管理システム、入・出荷管理システム、計測システム、(音声による読取り)仕分けシステムなど各種の産業分野への応用が考えられる。

HR-150音声入力装置は、不特定多数の話者を対象とする入力装置で、このため多数の音声パターンから個人性や属人性のある特徴パラメータを削除し、共通的な音声パターンでしかも他の単語との相違が明確になるような特徴を抽出することにより、標準音声パターンを作成している。

HR-150音声入力装置は、電話機を用いてどこからでも、だれでも音声でデータ入力ができるため広範な応用が考えられる。例えば、音声応答システムと組み合わせれば、音声で入力して、コンピュータの処理結果を耳で聞く理想的な情報システムの構成が可能となる。

4 画像応答システム(VRS)^{4)~6)}

VRS(Video Response System: 画像応答システム)は、写真や図形などの画像情報と音声情報を蓄積したセンタに対して一般のテレビジョン受像機とプッシュホンなどから成る端末からアクセスし、センタのコンピュータと会話を行なうことによって必要とする情報を選択入手することができるCENTER TO END形の情報システムで、日本電信電話公社の指導により開発を進めているものである。

このシステムは、音声付きの静止画及び動画の視聴覚情報を表4に示すサービス形式によって提供するが、静止画サービスでのセンタ~端末間の会話の展開は、VRS言語を使ってサービス提供者が自由かつ容易に規定することができるため、教育や情報検索などあらゆる分野での適用が可能である。

図6は日本電信電話公社が利用実験を進めているシステムの構成で、センタには会話制御を主体とする中央処理系、表5に示す各種の画像・音声ファイル装置などが設置されている。

センタから端末への映像伝送は、新たに開発した中継器(図7)を用いて電話回線を通じて行なわれ、端末でソフトコピー又はハードコピーとして表示される。

画像応答システムは、単なる情報メディアにとどまらず、コンピュータの多彩な処理機能と、豊かな情報伝達力をもつ視聴覚機能との結合によって、新しい電気通信メディアへ発展すると期待されている。上述(図6)のシステムは、このまま商用に供し得るものであるが、更に本格的普及に向けていっそうの技術開発、低コスト化、システム利用に関する調査研究などを進めている。

5 画像処理

コンピュータを利用した画像処理の研究、応用は多方面に及ぶが、ここでは日立製作所研究所での研究とその応用を中心に述べる。

研究所では主として次の(1)~(3)の分野での応用を目的に研究を進めている。

- (1) 産業応用、特に生産過程の自動化、高度化への適用
- (2) 医療応用、特にX線画像、超音波映像への応用
- (3) 通信応用、特にファックス及び静止画伝送への応用
- (4) リモートセンシングのための画像処理

生産自動化応用としては、プリント基板の傷検査、ICの自動ワイヤボンディングなどが既に実用化され効果を挙げている。現在では、LSIの傷や欠陥の検出の自動化に向かっている。

医療応用としては、X線CT(コンピュータトモグラフィ)の画質改善に始まり、高速化、頭部用からX線画像だけでな

表5 各種画像・音声ファイル装置の概要 静止画、動画、音声などで表現された情報の内容と、それらの利用形態によって種々のファイル装置が使い分けられる。

区分	装置名	ファイル内容	容量など	平均アクセス時間	記事
静止画	マイクロフィッシュ検索装置	静止画	5,940齣/台	3.8秒	——
	マイクロフィルム検索装置	静止画	7,200齣/台	0.45秒	——
	ビデオディスク装置	静止画	900フレーム/台	1.0秒	——
動画	図形文字発生装置	文字 図形 カラー: 7色	約10,000齣 外部記憶装置のデータエリア容量による。	文字: 28ms/10文字 図形: 10ms/10ドット	文字の大きさ 標準文字: 30×30ドット 小形文字: 20×20ドット 文字種類: 約2,300文字
	デジタル図形ファイル装置	図形 カラー: 有彩色13色, 無彩色4色	3,000~4,000画面	0.1秒	——
音声	16mm自動装填映写装置	動画	最大10巻装填可能 1巻: 最大70分	——	フィルム用 カラーカメラ使用同報動画用
	3/4inビデオテープレコーダ	動画	1巻: 最大60分	——	同報動画用
	ランダムアクセス動画ファイル装置	動画	カセットテープ: 120巻 再生用VTR: 12台	——	複合静止画用 リクエスト動画用
音声	音声応答装置	システムメッセージなど	最大メッセージ数: 400	0.5秒	音声単語を文章に編集・出力
	テープレコーダ	同報動画番組案内用音声	最大7分エンドレス	——	BGMなど (同報動画休止中出力)
	アナウンスマシン	障害・サービス休止, 試験中などの情報	4トラック/1台 最大14分エンドレス/トラック	——	センタ情報用
	バーコード画面別音声ファイル装置	画面別音声(人声のみ)	15秒/画面×約5,000画面	0.1秒	静止画の画面別説明用
	複合音声ファイル装置 (ランダムアクセス形カセットテープ再生機 磁気ディスク再生機 エンドレステープレコーダ)	画面別音声	最大収容メッセージ398種類 トラック数50, トラック最大11.5秒 4トラック/1台 最大14分エンドレス/トラック	18秒 0.5秒以下	——

注: 略語説明 BGM(バックグラウンドミュージック)

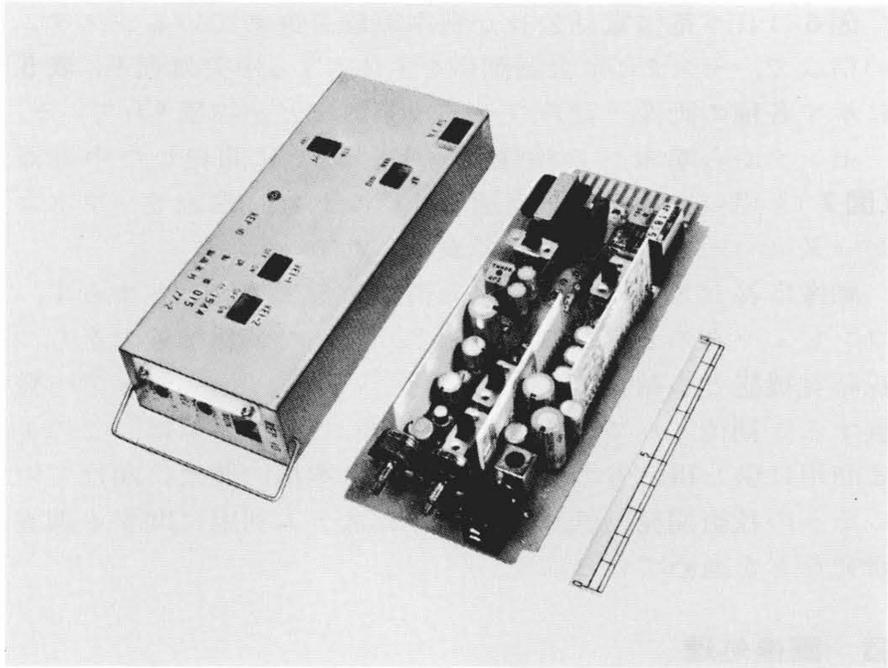


図7 IC化中継器 容積は約330cc, 消費電力は0.6Wである。

く超音波映像の処理, 更に血液標本の顕微鏡像からその血球を自動分類し, 異常の発見の手助けとするシステムの開発も行なっている。

通信応用としては, ファクシミリの白黒二値信号によるハーフトーンの再生法, 静止画像の電話帯域伝送方式, 連続スナップ画線の高能率伝送方式などを開発し, また空間フィルタと標本化の関係, 特にカラーテレビジョン信号の場合について研究を進めている。

リモートセンシング用画像処理技術としては, 気象衛星からの写真に基づく風向と風速の推定, 海洋の赤外線写真から

の温度分布の推定などを行なっている。

画像処理の基本的なソフトウェアとして, 空間フィルタリング, 空間微分, 空間積分, 境界線の抽出, テクスチャアナリシスによる空間の区分, 各種特徴の抽出, 計測, 統計処理などがあり, それぞれ整備されている。

今後は単なる画像処理だけでなく, 目的とする画像についての知識データベースに基づく画像理解へと研究は進展していくものと思われる。

6 結 言

以上, コンピュータと関連の深い音声応答, 音声入力, 画像応答, 画像処理の最近の活動について論述した。今後この種の分野の研究開発を積極的に推進する考えである。

終わりに, 平素種々御指導, 御鞭撻をいただいている関係各位に対し, 厚く御礼申しあげる次第である。

参考文献

- 1) 高橋, 外: 電話予約用音声応答システム, 日立評論, 57, 237~241 (昭50-3)
- 2) 吉沢, 外: 勝馬投票用音声応答システム, 日立評論, 59, 597~602 (昭52-7)
- 3) 遠藤, 外: 音声応答装置“HIVORS”による受注出荷システム, 日立評論, 61, 435~438 (昭54-6)
- 4) 小林: 画像応答システム/VRS, ビジネス・コミュニケーション, 15, 9 (1978)
- 5) 中島: 画像応答システム, 信学技報, CS-77-41 IE77-35 1977-07
- 6) 小林, 外: 画像応答システム, 日立評論, 60, 805~810 (昭53-11)

論文抄録

文献検索システムのファジィモデルとその実現

日立製作所 梶 博行・新田義彦

電子通信学会論文誌 62-D, 4, 297~304 (昭54-4)

文献検索での検索効率の向上を目指した一アプローチとして, 文献検索システムのファジィモデルを提案する。次に, 本モデルに基づく関連論理検索とマルチファセットシソーラスによる自動論理操作を特徴とする会話型文献検索システムをインプリメントし, その有効性を示す。

(1) 関連論理検索

現在, 最もよく用いられている検索方式は, 質問をキーワードの論理式で表現し, 各キーワードをディスクリプタとしてもつ文献の集合の間で集合演算を行なうことにより, 適合文献を判定する方式である。この方式は, 検索要求をかなり正確に表現できるが, 2値論理の厳密さに起因する検索漏れが欠点である。これを解決するため, キーワード間の意味的関連を考慮した検索論理をファジィモデルとして定式化した。すなわち, キーワード関連度 $\alpha(k, k')$ ($\in [0, 1]$) がキーワード k に対するキーワード k' の意味的関連の強さを表わす数値である

として, 質問 q と文献 d の適合度 $\psi(q, d)$ を次のように定義する。(i) q がキーワード k だけから成るとき $\psi(q, d) = \max_{k' \in D} \alpha(k, k')$ 。ここに, D は d のディスクリプタの集合である。(ii) $\psi(\neg q, d) = 1 - \psi(q, d)$ 。(iii) $\psi(q \wedge r, d) = \min\{\psi(q, d), \psi(r, d)\}$ 。(iv) $\psi(q \vee r, d) = \max\{\psi(q, d), \psi(r, d)\}$ 。次に, 適合度の大きい順に文献をランクづけして出力する方式を関連論理検索と名づけ, インバーテッドファイル編成でのファイル探索方略を示した。

開発したシステムでは, 上・下位概念語, 関連語を考慮し, キーワード関連度として5段階の値を定めた。また, ファイル探索時間の増加を抑えるため, 関連度の大きいキーワードの文献ディレクトリが近傍に格納されるよう, 文献ディレクトリの格納順序を, 上・下位概念語の関係を表わすトリーでの節点の縦型順序に一致させた。

(2) 自動論理操作

マルチファセットシソーラスは, 文献を

分類あるいは検索するための観点(ファセット)を複数提供する。これによれば, 対象分野の情報を表現する際, 一般に, 同一ファセットに属するキーワードは論理和で, 異なるファセットに属するキーワードは論理積で結合されるという法則性が存在する。このことを利用して, システムが論理関係を判定する自動論理操作方式を採用した。

ある技術分野を対象に, 2,648語のキーワード, 5,450件の文献をデータとして試行し, 次の結論を得た。関連論理検索は, 会話型検索で検索効率の向上に有効である。すなわち, 検索漏れが防止でき, しかも, ランクづけ出力方式であるため, 検索雑音の悪影響は小さい。また, 自動論理操作は, 煩雑な論理式を組み立てる操作を排除し, 質問入力の簡単化を可能にする。なお, 応答時間については, 20個のサンプル質問に対し平均4.8秒であった。適合度の小さい文献にまで検索範囲が拡大される場合も, 応答時間の増加は小さかった。