

分散処理環境における音声処理とイメージ処理
— 電話・FAXを利用した音声処理システム —

佐々木 澄
富士通株式会社

本報告では、公衆回線網を用いた分散処理環境で音声、イメージ処理を利用した当社の音声処理システムについて述べる。

本方式は、特別な端末を用いることなく一般に普及しているダイヤル電話、プッシュホン、及びFAXをホストコンピュータをつなぐことにより、容易にオンラインシステムの構築を可能とすることができる。この音声処理システムを支えるハードウェア、ソフトウェアの機能について紹介する。

SPEECH PROCESSING AND IMAGE PROCESSING IN THE DISTRIBUTED PROCESSING ENVIRONMENT

— Speech processing system based on telephone and facsimile —

Kiyoshi SASAKI
FUJITSU LIMITED

140 Miyamoto, Numazu city, Shizuoka, 410-03 Japan

This report describes our speech processing system, which utilizes speech processing and image processing in the distributed processing environment realized by employing public switched network facilities.

This system features the use of popular equipment such as dial phone or pushbutton phone instead of specialized terminals, and the connection of facsimile and computer which enables easy building up of online processing system.

Also described are the hardware and software supporting features in this system.

1. まえがき

従来、コンピュータとの会話は文字を情報伝達手段として、人間の目と手を通して行ってきた。

これに対し、近年の音声信号処理理論の発展とイメージ処理技術、さらにLSI化技術の向上は、コンピュータによる音声処理とイメージ処理を可能とし、急速にこれらの技術を応用した機器の実用化が進んでいる。

そこで、人間にとって最も自然な情報伝達手段である音声をマンマシンインタフェースに取り入れること、さらにイメージ処理を用いてデータ出力を可能にすることが、コンピュータをさらに身近にするものとして期待されている。このたび開発した音声処理システムは、手軽で使い易い電話、及びFAXを利用して、コンピュータと会話しながらデータの入出力を行うオンラインシステムを提供するものである。

本稿では電話及びFAX利用のオンラインシステムの一般的な特徴と、音声処理システムの概要を紹介する。

2. 電話・FAX利用のオンラインシステム

2.1 概要

音声による電話オンラインシステムは、利用者が電話を使って「いつでも」、「どこからでも」直接コンピュータと会話しながらデータ入力できるシステムとして、

1) 従来電話応対要員を配置して行われていた業務の省力化、

2) サービス性の向上などを
実現する手段として各方面で利用されている。

さらに音声に加えて、FAXに対してコード出力を可能にしている。

本システムの特徴を以下に示す。

1) 経済的なオンラインシステム

既存の公衆電話網と電話及びFAXを利用して、経済的に広域オンラインシステムを構築できる。

2) 音声による確認

電話から入力したデータに対し、コンピュータから音声で確認や誤操作の指摘を行うので、入力データの正確性と操作性の向上が図れる。

3) 記録での確認

利用者と音声だけのやりとり及び処理結果では記録が残らないという問題を解決するためにFAXにコードデータをイメージデータとして出力することができる。

4) 業務の省力化

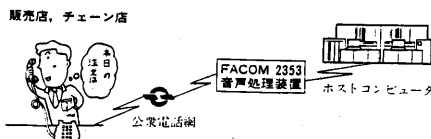
専任の電話応対要員やインプットデータ作成要員を必要としないので、省力化が図れる。

5) サービスの向上

オンラインシステムによる即時処理及び自動化による受付時間の延長など、利用者へのサービスが向上する。

2.2 適用分野

電話を利用したオンラインシステムは、座席予約システム、競馬の電話投票システムなどの大規模専用システムとして導入されはじめた。



販売店からのオーダーエントリ		センターの応答 (音声出力)
入力項目	プッシュボタンの操作	
1. センターへ電話をかける	4 3 7 5 1 1 1	こちらは〇〇センターです。販売店コードをどうぞ。
2. 販売店コードを入力	2 6 1 6 #	〇〇〇様ですね、まいどありがとうございます。ご注文の納期をどうぞ。
3. 希望納期を入力	1 0 2 0 #	納期は10月20日ですね。品番、数量をどうぞ。
4. 品番と数量を入力	4 4 3 2 3 5 0 #	〇〇〇 (商品名: 品番 4 4 3 2) 3 5 0 (数量) 次をどうぞ。(4を繰り返す。)
5. 終了コードを入力	0 #	ご注文ありがとうございました。受付番号は123です。伝票控をファクシミリへ出力いたします。ファクシミリの受信ボタンを押して下さい。
6. ファクシミリの受信ボタンを押す。		

図-1 オーダエントリ会話例

さらに最近は銀行の事務合理化、顧客サービスの向上を目的とした振込通知・残高照会システムのように、一般の事務所や家庭の電話をサービス対象とするまでになってきた。

システム規模の大小にかかわらず、タイムリな情報収集、即時応答、即時データ処理を必要とする次のような分野にまで、適用範囲の拡大が広まってきている。

- 1) 製造業、流通業におけるオーダエントリ、売上げデータエントリ、在庫問合せ業務
- 2) サービス業における、予約、問合せ業務
- 3) 各種の案内、通知業務

音声処理システムを利用したオーダエントリでの会話手順の例を図-1に示す。

3. 音声処理システムの特徴

前節2で述べたオンラインシステムを実現するために開発した音声処理システムは、以下に示す特徴をもっている。

3.1 ハードウェアの小型化

音声処理システムの入出力制御を行うハードウェアとして、FACOM2350シリーズ音声処理装置(SPU:Speech Processing Unit)を開発した。

本装置は、導入コストの低減化及びコストパフォーマンスの点から、電話収容回線数を最大32回線に絞り、小規模システムから大規模システムまで、設置台数によって柔軟に対応できる構成とした。

- 1) 音声出力にPARCOR方式を採用したので、音声データ量を大幅に削減でき、音声データ専用の補助記憶装置(たとえばディスクファイルなど)は不要とした。

PARCOR方式とは、NTTが開発した音声合成方式で、人間の発声機構(声帯、舌などの器官とそのメカニズム)をデジタル回路で模擬し、少ない合成用データ(PCM方式に比べ1/10以下)で良質な音声出力ができる方式である。

- 2) 規則合成機能の採用によってあらかじめ登録を必要とする音声データ量を削減した。
- 3) 文字コードデータをFAX及びデータテレフォンへ出力し、紙による記録がとれるようにした。SPUの処理フローを図-2に示す。

3.2 音声コード系の導入

コンピュータ内で音声応答メッセージをコードとして処理するために、音声コード系という概念を導入した。

音声コードとは、文字、単語、文節、文章の音声を表す標準的なコードである。

たとえば、B981とB991(16進)は、それぞれ「イチガツ」「ツイタチ」の音声を表す音声コードである。

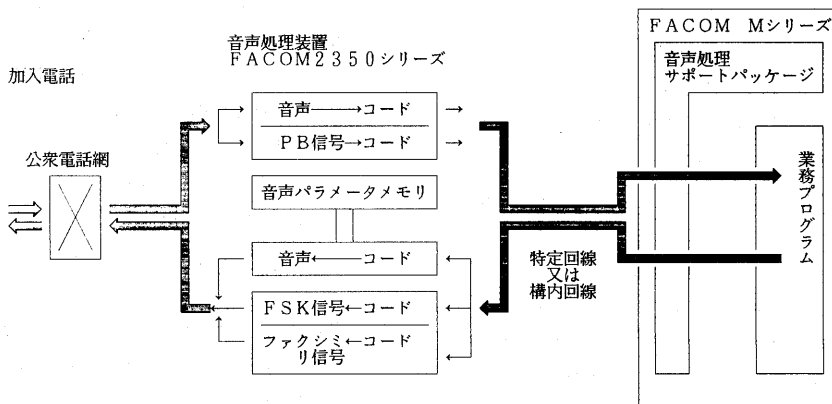


図-2 SPUのフロー

音声コードには、基本音声コードとユーザ定義音声コードがある。音声コードの体系を、図-3に示す。

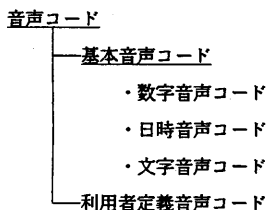


図-3 音声コード体系

コンピュータ内では、このコードを用いて文字コードの処理と同様に、コードベースで自由に音声出力の組立と編集を行うことが可能となる。

例えば、「イチガツツイタチ」の音声の組立は、前記の音声コード2個を組合せて、「B981B911」（16進）のみで実現できる。

3. 3 経済的なオンラインシステム

1) 分散設置指向

利用者の近くにSPUを分散設置すれば電話料金の負担を軽減できる。

2) ホスト集中管理の分散処理システム

分散されている電話及びFAX、SPUに対しホストコンピュータから電話サービス開始と終了の指示ができ、SPUの専任オペレータは不要である。また、アプリケーションの開発、メンテナンスをホストで集中管理できる。

3. 4 音声処理サポートパッケージの提供

音声処理によりサポートパッケージ（SPEECH、4.3参照）により、EDPコードの読み方を意識することなく、音声応答メッセージの組立、編集及び電話の制御ができる。

さらに、高級言語でアプリケーションプログラムを開発することができる。

4. 音声処理システムの構成

音声処理システムは、電話、ファクシミリ及びデータテレフォンをコンピュータの端末として使用するオンラインシステムであり、通常のAIM（Advanced Information Manager）によるオンラインシステム運用に必要な資源の他に、ハードウェア装置として音声処理装置を、ソフトウェアとして音声処理サポートパッケージを必要とする。

音声処理システムの各コンポーネントの関連を図-4に示す。

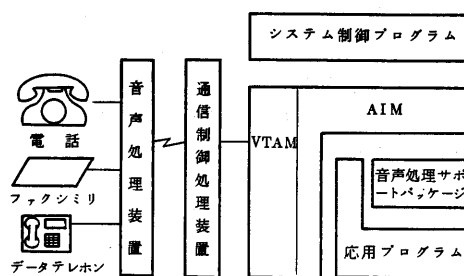


図-4 音声処理システムのコンポーネント関連

4.1 音声処理装置

本装置は、音声処理システムにおいて、ホストコンピュータと公衆電話網との間に位置し、電話機を端末として入出力制御を行うものであり、次の機能を果たす。

- ホストコンピュータとのデータ送受信の制御
- 電話の発信・着信の制御
- 電話からのデータ入力
- データテレホンからのデータ入力
- 電話へのデータ出力
- データテレホンへのデータ出力
- ファクシミリへのデータ出力

本装置の特徴は、音声合成LSI、音声処理技術（PARCOR）イメージ処理技術の採用により小型、低価格化を実施したことにある。

表-1に音声処理装置の仕様を、図-5に音声処理装置の構成を示す。

4.2 接続端末

音声処理装置に接続可能な端末の種類には、次のものがある。

- プッシュホン
- 回転ダイヤル電話
- ファクシミリ (G II / G III規格機)
- ミニファクス
- データテレホン

電話からの入力方法には、プッシュボタン押下によるPB信号によるもの(回転ダイヤル電話にアタッチメントを付加したものも含む)と音声入力による音声信号がある。

いずれも音声処理により文字データ(EBCDICコード)に変換され、応用プログラムに渡される。

音声入力は、不特定話者を対象に17単語の入力が可能である。

また、ファクシミリ及びデータテレホンにコードデータを出力することにより記録を残すことができる。

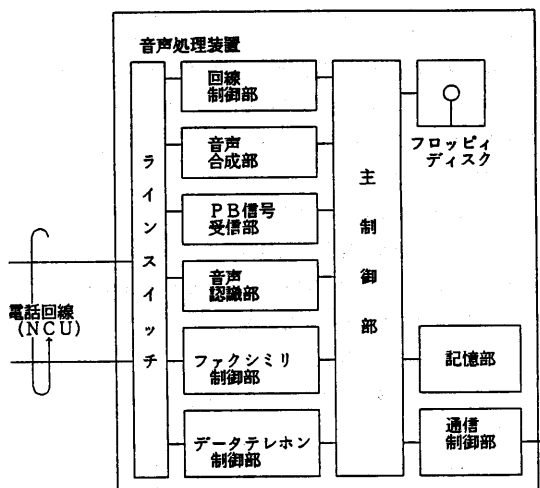


図-5 音声処理装置の構成

表1 FACOM2350シリーズ音声処理装置の仕様

項目	FACOM2355音声処理装置	
ホストインタフェース	接続回線数	特定回線/構内回線 2回線
	通信方式	半二重、SYN同期、2,400~9,600ビット/秒 (ファクシミリ出力の場合 9,600ビット/秒)
電話回線インタフェース	接続回線数	電話型公衆通信回線 最大32回線
	網制御装置	AA形 (1回線、自動発信・自動着信) MA形 (1回線、自動着信のみ) CA20形 (20回線用、自動発信・自動着信)
接続端末	電話	プッシュホン 回転ダイヤル電話 PB信号アタッチメント付加の電話
	データテレホン	1形、2形(プリンタ付)
	ファクシミリ	G III規格
電話回線信号方式	入力	プッシュボタン(PB)信号 音声
	出力	PARCOR方式による音声合成 規則合成による固有名称合成
		FSK信号(データテレホンへのプリンタ出力用) ファクシミリ信号(ファクシミリへの文字出力用)
音声認識	認識対象	不特定話者(離散発声)
	認識単語(固定17語)	0~9、はい、いいえ、おわり、とりけし、もう1度、保留、どうぞ
収容メッセージ長	基本メッセージ 約 400秒 ユーザ固有メッセージ 最大2,600秒	

4.3 音声処理サポートパッケージ (SPEECH)

一般にコンピュータの処理結果は、文字コードで表現されており、音声として出力するためには、文字コードから音声コードに変換する必要がある。

SPEECHは、

- ① このような音声コード変換機能をはじめ、
- ② 電話との会話を容易に制御したり、
- ③ SPUをホストコンピュータで制御するための機能を提供するソフトウェアである。

1) SPEECHの構成

SPEECHを構成するプログラムとその機能について説明する。

① 音声コード変換ファイル編集ユーティリティ

業務処理ごとに異なるEDPコード (商品コード、取引先コードなど) やガイダンスメッセージ番号 (たとえば、メッセージ番号01は「こちらは・・・でございます。」など) とその読み方を示す音声コード列との対応を定義するために用いる。

② 音声コード変換サブルーチン

EDP処理用の文字コードをその読み方を表す音声コード列に変換する。

③ 音声制御情報編集サブルーチン

電話、FAX、データテレフォンを業務処理プログラムで制御するために業務処理プログラムとSPUの間で必要な制御情報の作成と処理結果の解析をする。

④ 音声認識情報編集サブルーチン

音声入力に関して認識対象単語の指定とSPUで音声認識した結果を取り出す。

⑤ SPU制御コマンドパッケージ

ホストコンピュータのコンソールからSPUに対して電話サービス開始、終了などの指示を可能とする。

2) 音声コード変換

業務処理プログラムで扱う各種文字コードから、簡単に効率よく音声コード列を生成する機能である。

① 文字音声コード変換

文字列 (カナ、英数字) からその読み方を表す音声コード列に変換する。

② 数字音声コード変換

数字列を業務処理プログラムの指定に基づいて、次のいずれかの変換を行う。

- i) 桁付きで読む場合の音声コード列
- ii) 1桁ずつ読む場合の音声コード列
- iii) 単位を付けて読む場合の音声コード列

③ 日時音声コード変換

数字列をいくつかに分けて、日時特有の読み方を考慮した音声コード列に変換する。

④ 項目音声コード変換

EDP処理用文字コード (たとえば、商品コード、取引先コードなど) やガイダンスメッセージ番号を音声コード変換ファイルを参照して音声コード列に変換する。

数字音声コード変換と項目音声コード変換の例を図-6に示す。

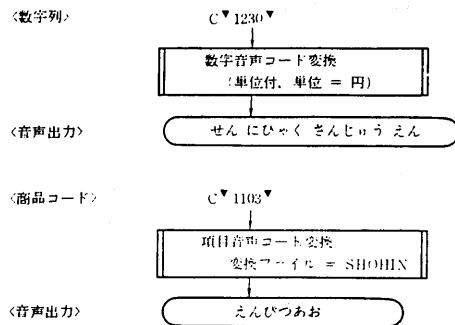


図-6 音声コード変換の例

5. むすび

今回のシステムは、電話に対する音声データ入出力、FAXに対するコードデータの出力を実現した。

これにより、音声のみだけでなく文字として処理結果を確認することができる。

今後は、さらに発展させイメージデータ自体の入出力を可能としたマルチメディア処理の実現、さらに不特定話者用の音声入力単語の増大を図って行き多様なユーザー要望に応じていく予定である。