

# 音声応答システム

## Audio Response System

Hitachi supplied recently an experiment set of the telephone reservation system to the Railway Technical Research Institute, Japanese National Railways. This is for experimenting on the reservation of railway tickets from the home push phone using the public telephone networks which will be made available in the near future. As a result of this experiment key points in the designing of such reservation system have been clarified pushing the JNR's plan a step forward to realization.

This article describes the general characteristics of the audio response system as well as the outline of the experiment system of telephone reservation system.

木村幸男\* Yukio Kimura  
 三木彬生\* Shigeo Miki  
 中田和男\*\* Kazuo Nakata  
 市川 熹\*\* Akira Ichikawa  
 麻生 哲\*\*\* Satoshi Asô  
 末広明雄\*\*\* Akio Suehiro  
 酒井久雄\*\*\* Hisao Sakai  
 磯部光庸\*\*\* Mitsunobu Isobe

### 1 緒 言

音声応答装置とは、コンピュータの処理結果を音声の形で出力する装置である。

すでにアメリカでは、銀行業務やオーダーエントリに音声応答装置を用いたシステムが商用化されている。一方、わが国でも日本電信電話公社の日本万国博覧会データ通信システム<sup>(1)</sup>による待合せサービスや加入データ通信による電話計算サービスなどですでに実用化されている。

日本国有鉄道でも、指定券類の電話による予約の需要がますます増加していることを期に、音声応答装置を用いた電話予約システムの開発を行っており、45年度には、この一環としてキャラクタディスプレイ装置と音声応答装置を利用し、係員が電話の応待を行なう半自動システムの開発、実験を行

なったが<sup>(2)</sup>、今回はさらに、係員の応待も自動化した全自動システムを開発しその試験を行ない、ここに電話予約システム実現の可能性を得た。

本稿は、音声応答システム一般について述べ、続いて今回行なった電話予約実験システムの内容およびその実験結果について述べる。

### 2 音声応答システムの条件

音声応答装置を、不特定多数の利用者が利用する情報サービスシステムの出力装置として使用するとき、そのシステム構成は一般に図1に示すようになる。

情報サービスを受けようとする利用者は、端末である押し

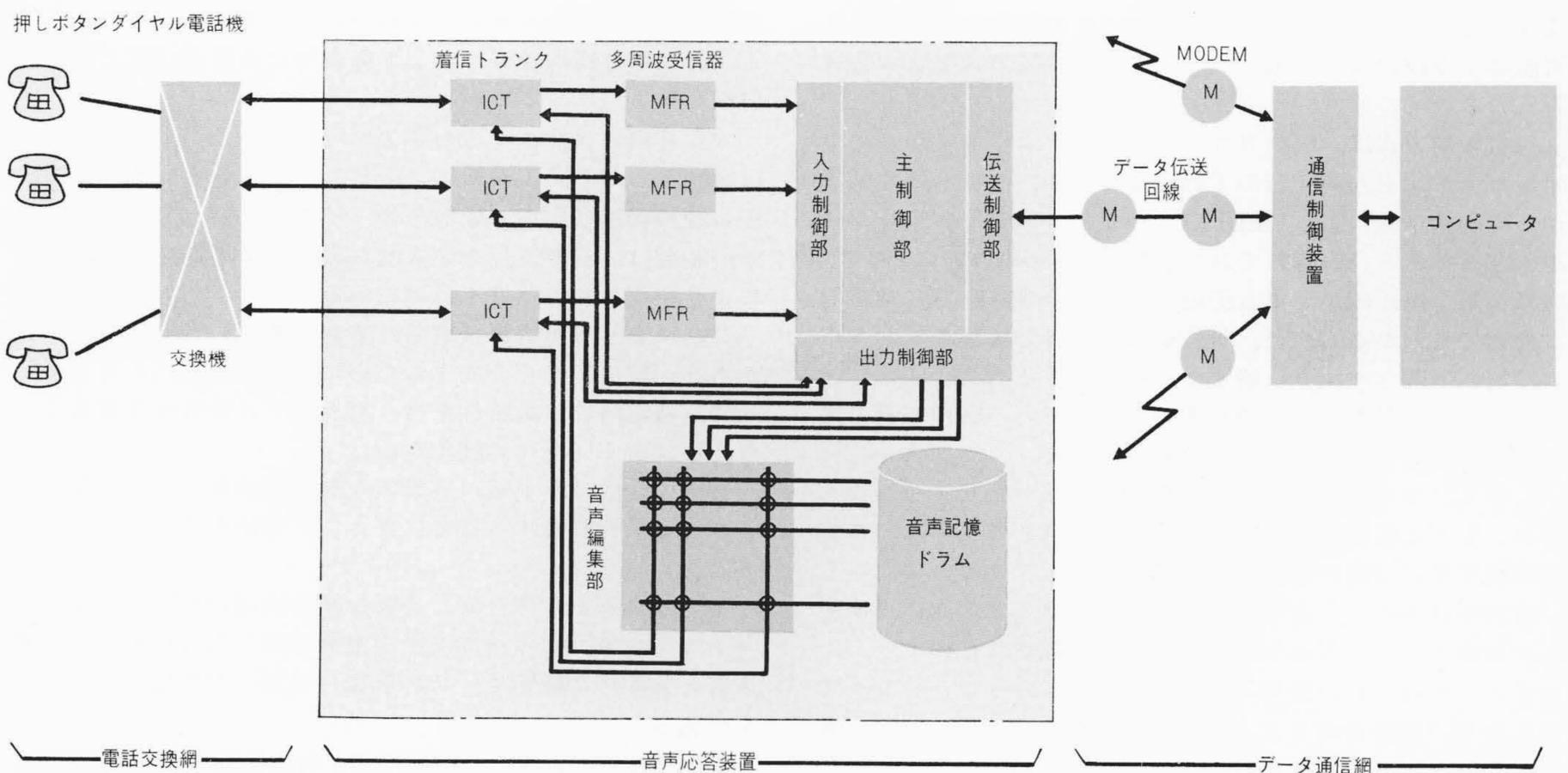


図1 音声応答システムの構成 電話交換網と接続した音声応答システムの例

Fig. 1 Block Diagram of the Audio Response System

\* 日本国有鉄道鉄道技術研究所 \*\* 日立製作所中央研究所 \*\*\* 日立製作所戸塚工場

ボタンダイヤル電話機よりセンターを呼び出し、受けようとする情報サービスの内容を電話機の押しボタンにより入力する。この信号は音声応答装置の主制御部に取り込まれ、入力内容の確認を音声で行ない、このデータを中央のコンピュータに転送する。

中央のコンピュータでの処理結果は、再び音声応答装置に転送され、音声応答装置では、この内容を音声信号に変換し電話機に回答する。このようにして、不特定多数の人々が、「いつでも」「どこからでも」利用するオンラインリアルタイムシステムが実現できる。これらのシステムには、いくつかの基本的条件<sup>(3)</sup>が要求される。そのおもなものは、

- (1) 了解性、明確性を十分に持った高品質の音声を出力できること。
  - (2) 多数の回線に同時に異なる内容の応答ができること。
  - (3) できるだけ多くの語いが出力できること。
- などである。

(1)については、不特定多数の利用者を対象としたサービスでは当然の要求であり、特に音声のみで処理結果を理解するためには、聞いてわかるということだけでなく、聞きやすさ、すなわち自然性も必要となる。

(2)については、処理装置の処理スピードは非常に速く、1名の利用者が処理装置を占有する時間が短時間であっても、人間に理解できる音声形式で応答するためには、ある一定の時間は必要であり、この間他の利用者が待たされたのでは利用価値はほとんどなくなってしまふ。

(3)の発声可能な語数は、情報サービス内容を限定すればかなり少なくてもよい場合もあるが、本格的なサービスを行ない、さらに自然性を要求すればするほど多くの語いを必要とし、できるだけ語数が多く出力できることが望ましい。

### 3 音声応答方式の選定

音声応答装置の果たすべき機能を、処理装置の処理結果を人間に理解できる音声に変換することと定義すれば、それを可能にする技術には、大きく分けて録音編集方式と音声合成方式の2方式に分けられる<sup>(4)</sup>。

録音編集方式は、磁気ドラムや光学ドラムにあらかじめ単語または単文を人間の音声で録音しておき、これを処理装置からの制御情報によって選択し、つなぎ合わせて音声を作り出す方式である。録音しておく音声の単位の取り方によって、文章編集、単語編集、単音節編集などに分類される。常識的に考えてもわかるように、単位が大きいほど編集された音声の了解性と聞きやすさは増すが、応答できる内容の変化に乏しく語数が制限される。現在実用化されている音声応答方式の原理はすべてこの方式によるものである。

しかし、さらに内容の多い本格的な情報サービスを行なうには、もっと応答語数の多い音声応答方式が要求され、これを実現するには音声合成技術が必要となってくる。

音声合成方式は、音声合成の原理によって音声を作り出すものであり、その方法は大きく二つに分類される。

その一つは、音声発生の原理を忠実にシミュレーションするもので、純粹合成方式と呼ばれており、ターミナルアナログ合成、ボカルトラクトアナログ合成などがある。原理的には適当な制御情報さえ与えれば、どのような音声でも発声させることができ、発声語数に制限はないが、一般に装置が複雑になり、処理装置を占有する時間を多く必要とし、多重利用の点でデータ通信システムに適用するには、現状では困難とされている。また、出力音声の品質も必ずしも十分とは

いえない。

他の一つは、合成と録音編集の中間をいくもので、音声より抽出した特徴パラメータを作り、これを記憶しておき、その組合せで音声を復元するもので、編集合成方式と呼ばれるものである。これも、音声の特徴パラメータをあらかじめ記憶しておく以上、記憶媒体の容量とアクセスタイムによって応答可能な語数は制限されるが、原波形のままで記憶しておくのに比べて、 $\frac{1}{10}$ ～ $\frac{1}{30}$ の情報圧縮が可能である。通常の800トラック、クロック2MHz程度のデジタル磁気ドラムで1語約1秒間の語が5,000語程度記憶しておくことができる。この程度あれば実用上十分であり、録音編集方式の音声応答に次ぐものとして脚光を浴びている。この方式には、音響素片ピッチ制御方式<sup>(4)</sup>、PARCOR (Partial Auto-Correlation Coefficients) 方式<sup>(5)</sup>などがあり、それぞれに特長があるが、今回の電話予約実験システムでは前者の音響素片ピッチ制御方式を用いた。

### 4 音響素片ピッチ制御方式

音声は声帯振動による空気の脈流が、口腔(こう)、鼻腔などの共振により、くちびるから空間に放射された音波と考えることができる。簡単化のために、声帯の振動を周期的と仮定すれば、1回の励振による共鳴波形を周期的にくり返すことになる。口の共鳴作用は、各共鳴モードに対応する単一共鳴の和に分解することができ、共鳴周波数の低いほうから、第1フォルマント、第2フォルマント……といている。音響的には、その共鳴周波数を振動周波数とし、その損失によって決まる減衰定数を持った減衰正弦波形で表わされる。この減衰正弦波を音響素片という。

この音響素片を数個適当に加え合わせることにより、音声を合成することができる。この音響素片の数が実用的(装置の構成、素片の記憶容量など)意味を持つためには、種々の実験結果より、第1フォルマント用16個、第2フォルマント用23個、第3フォルマント用10個および第4フォルマント用1個計50個と特殊なものとして鼻音用に1個を必要とし合計51個である。

次に音声の自然性を表わすアクセントやイントネーションは音源としての励振が複雑に変化してできるもので、この励振音源の繰返し周期の逆数をピッチ周期といている。ピッチの制御は、音響素片を順次的に読み出す周期をピッチ情報により変化させ自然性を付けている。

これにより有声音は合成されるが、実際の音声には無声音があり、特に子音を合成するために必要である。これも上記同様、特定の周波数特性を持ったランダム雑音を音響素片と考えれば全く同様に処理できる。

音声の合成は、上記の音響素片を制御情報により加算し、さらに音量を制御することによって、音声を作り出すものである。

本方式による音声合成に必要な制御情報の種類は、第1～3フォルマント、ピッチ周波数、無声音素片および有声音、無声音の強度の7種類で、その制御に必要な情報量は計32ビットである。

制御情報を更新する時間間隔は、合成実験の結果10ms間隔以内であればよく、したがって1秒間の音声を合成するために必要な情報量は、3,200ビットとなる。これは同じく1秒間の音声波形そのものを、8ビット、8kHzサンプリングでデジタル変換(64,000ビット)するのに比べて $\frac{1}{20}$ の情報圧縮となっている。

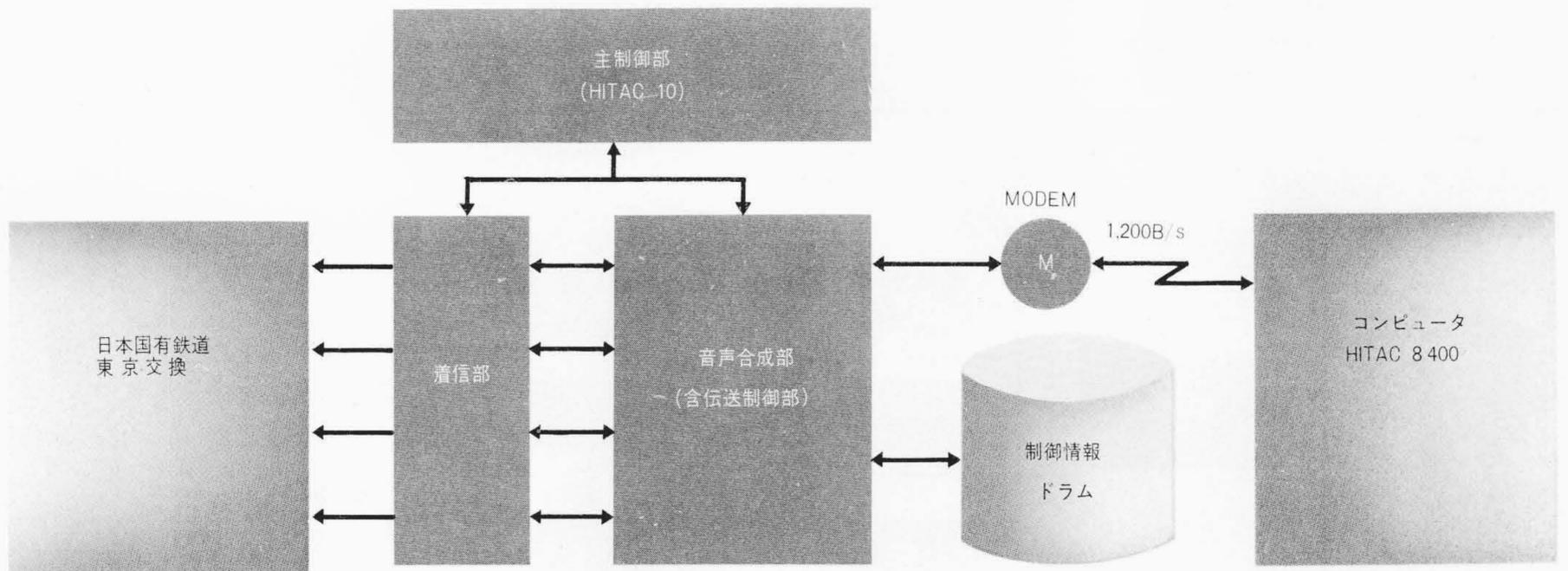


図2 電話予約実験システムの構成 電話回線は、鉄道電話回線4回線を使用した。

Fig. 2 Experimental Set of Telephone Reservation System

### 5 電話予約実験システムの構成

電話予約の実験は前述のとおり、45年末にキャラクターディスプレイ装置と音声応答装置を利用し、係員が電話の応待を行なう半自動システムの開発ならびに実験を行なったが、今回はさらに係員の応待も自動化した全自動システムを開発し<sup>(6)</sup>、この実験を47年度に国鉄秋葉原駅内にセンターを置き、鉄道電話回線と接続して行なわれた。図2は、このシステムの構成を、図3は機器の外観を示すものである。

センター内に設置された機器は、音声応答装置として、着信部、合成部および主制御部より構成され、さらに座席予約システムのシミュレータとしてHITAC 8400を用い、これらをモデム (MODEM) を介して接続した。

着信部は着信トランクと多周波受信器とで構成されている。着信トランク (ICT=In Coming Trunk) は交換機からの16 Hz (呼出し) 信号を受信すると主制御部に呼びがきたことを知らせる信号を出す。主制御部はこの回線が使用可能かどうかを調べ、使用可能であれば接続信号を送出して回線の接続を行なう。多周波受信器 (MFR=Multi Frequency Receiver) は、端末である押しボタンダイヤル電話機よりの多周波信号を受信し、これをパルスに変換して主制御部に取り込ませる。

音声合成部は制御情報記憶用のドラム (MDC=Magnetic Drum ControlおよびCDM=Control Data Memory)、音響素片記憶用のコア (AES=Acoustical Elements of Speech)、合成のための制御回路 (SPC=Speech Synthesizer Control)、演算回路 (CRC=Channel Register and Control)、D/A変換回路 (SSP=Synthesis Processor) および上位計算機とのデータの伝送を行なう伝送制御回路 (CMC=Communication Channel Control) で構成され、図4に示すようになっている。

表1は電話予約実験システムに用いた音声応答装置の主要性能を示すものである。

### 6 座席予約操作手順

電話で座席を予約しようとする顧客は押しボタンダイヤル電話機でセンターを呼び出し、あらかじめ定められている (使用方法は時刻表などに掲載)<sup>(6)</sup> フォーマットに従って、押しボ

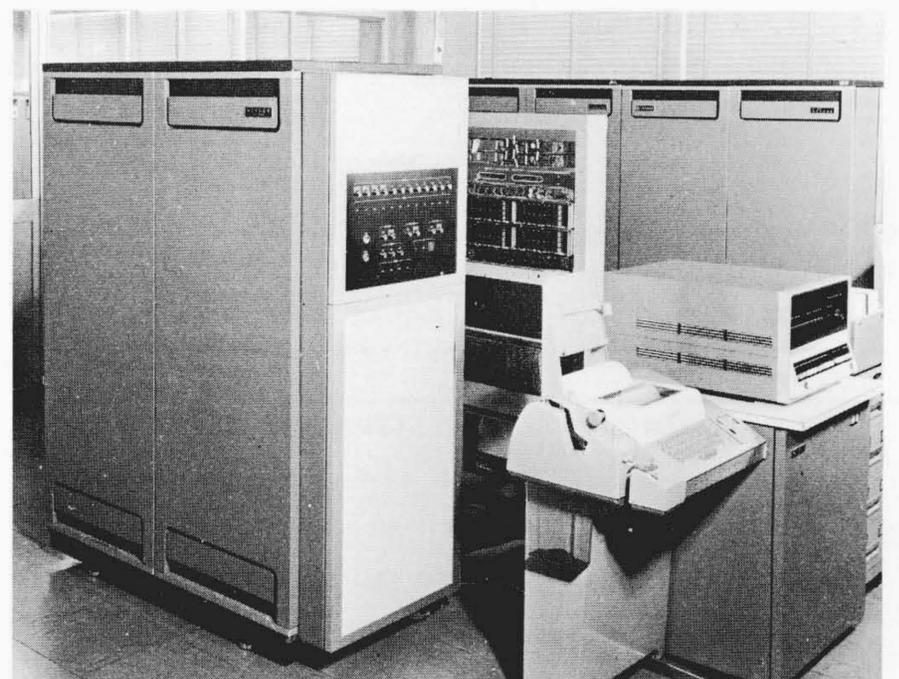


図3 電話予約実験システムの外観 日本国有鉄道・秋葉原駅のセンターに設置した音声応答装置を示す。

Fig. 3 Outlook of the Experimental Set

タンの12個のコード (1, 2, …… 9, 0, \*, #) で入力する。ここで、列車名および乗降駅名はコードが必要となる。表2は入力項目およびコード例を示すものである。コードには次のような原則を定めてある。

- (1) 予約内容は数字コードとする。
- (2) \*および#はそれぞれ赤ボタン、青ボタンと呼び、機能ボタンとして使用する。\*は訂正用、#は項目間のセパレータ用である。
- (3) 訂正の範囲は原則として直前の#を越えない。

予約対照範囲は、全国の特急列車およびその停車駅とした。予約内容を入力する方法は、システムの使いやすさに直接影響を与える因子であるため、今回は二通りの方法によりその操作性を確かめた。

その一つは、会話形で表2に示す入力項目1項目ずつ音声で確認を行ない、次の項目の入力指示を与えるA-1システムで、他の一つは、同じく会話形であるが、全項目の入力が完了するまでなんの指示も与えず、完了時点で一括して音声で確認を行なうA-2システムである。

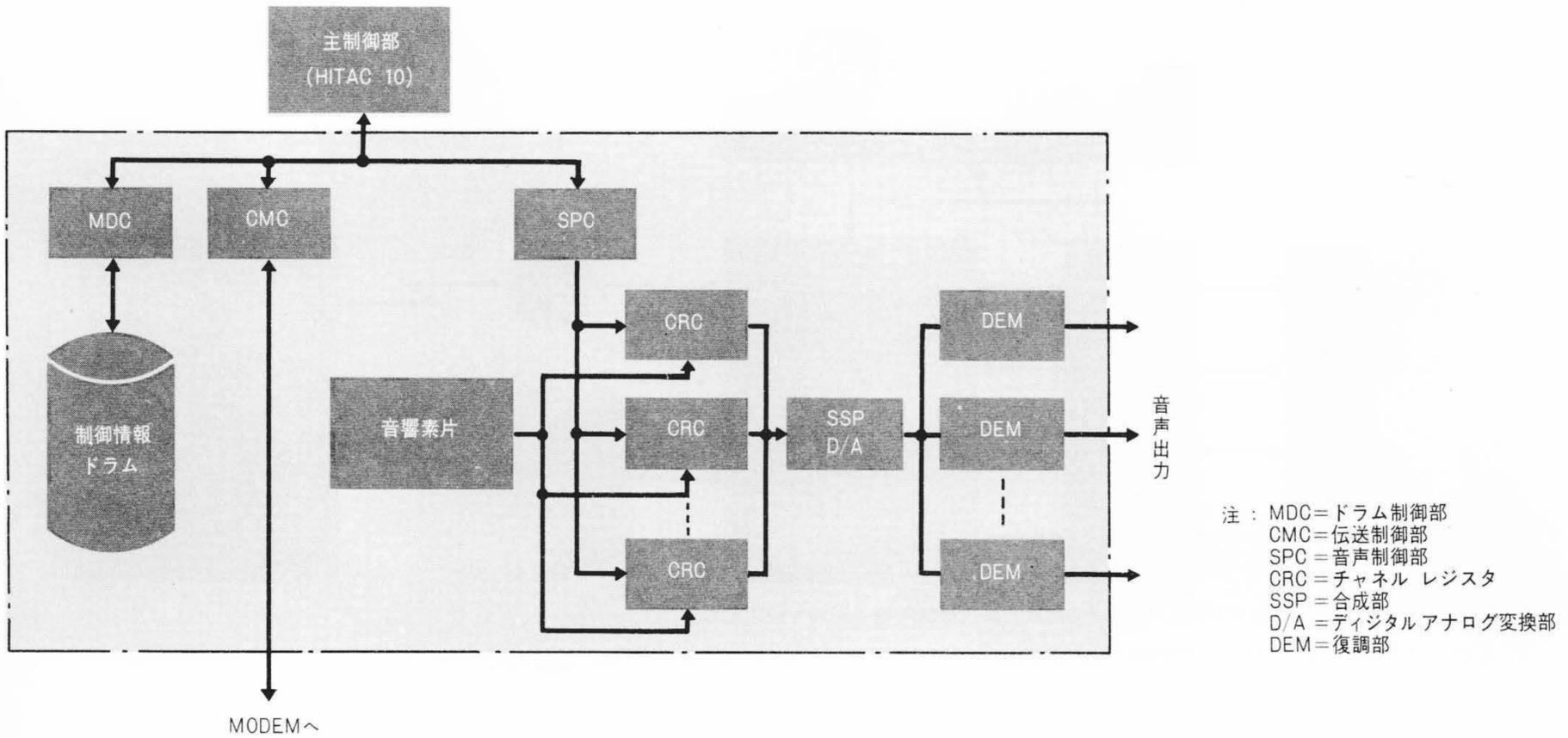


図4 音声合成部の構成 I/OとしてMDC, CMCおよびSPCが個別に主制御部に接続される形となっている。

Fig. 4 Block Diagram of the Audio Response Unit

表1 主要性能 実験装置では12回線制御を行ない、そのうちの4回線実装とした。

Table 1 Main Description of the Audio Response System

項目	性能
同時処理回線数	12回線 (4回線実装)
語数	1,500語 (1語の長さ 0.8秒)
方式	音響素片ピッチ制御方式
文形作成	編集方式 (固定文, 可変文の組合せ)
上位コンピュータとの接続	1,200B/sのデータ伝送回線(MODEM使用)
伝送制御形式	ポーリング・セレクトィング方式
接続電話回線	公衆電話回線の加入者位置
入力形式	多周波信号による押しボタンコード
出力形式	音声

表2 入力項目およびコード 入力手順は項目1, 2……, の順で行なわれる。

Table 2 Item of Input

項目	コード	制限
1. 電話番号	×××××××#	7けた (都区内)
2. 乗車人数	×#	4名まで
3. 乗車月日	××#××#	申込みの日から2ヶ月以内
4. 乗車列車	××××#	4けた固定 (コード表による)
5. 乗車駅	××××#	
6. 降車駅	××××#	
7. 設備種別	×#	1~4 (順に普通, グリーン, B寝台, A寝台)
8. 完了コード	0#	

注: 1. ×は, 1, 2……, 9, 0の数字を表わす。

2. 月日は, 04# 09#, 4# 9#のいずれも4月9日を表わす。

3. 駅名コード(例) 水戸2101, 常陸多賀2102, 日立2103……,

## 7 応答のためのプログラム

音声応答装置の主制御部には, HITAC 10を使用してプログラム制御を行なっている。

主制御部での処理プログラムは, 大別すると次の四つになる。

- (1) 押しボタン信号の取込みプログラム
- (2) 音声編集・出力プログラム
- (3) 伝送制御プログラム
- (4) 実験データ採取プログラム

これらのプログラムは完全な実時間性を要求される項目が多いので, 最短要求時間である10msごとに主制御部内蔵のタイマから割込みをかけ, これを処理の開始合図として必要な処理ルーチンを通す同期制御方式とした。時間的に制限される項目は次のようになる。

- (1) 10msごと2ワードの音声制御情報を音声合成部へ出力する。
- (2) 20msごと押しボタンダイヤル電話機からの予約情報を取り込む。
- (3) 20msごと1回線分の音声制御情報を磁気ドラム(CDM)から読み出す。
- (4) 10msごと最大2キャラクタをセンターからの情報として取り込む。
- (5) 10msごと1キャラクタの情報をセンターに送出する。
- (6) その他, 端末である電話機のフックオンおよびフックオフを検出する。

図5は, これらのプログラムの制御フローを, 図6は, 処理の時間関係を示すものである。

### 7.1 押しボタン信号取込みプログラム

今回の音声応答装置は押しボタン信号を一時記憶する部分を持たないため, 1回の押しボタン信号が継続している時間(40ms)以内に必ず主制御部に取り込まなければならない。したがって, 20msごとに取込みを行なっている。さらに取り込んだデータの内容を解析し, 数字, \*, #あるいはポーズによりおのおのの処理を行なうことになっている。

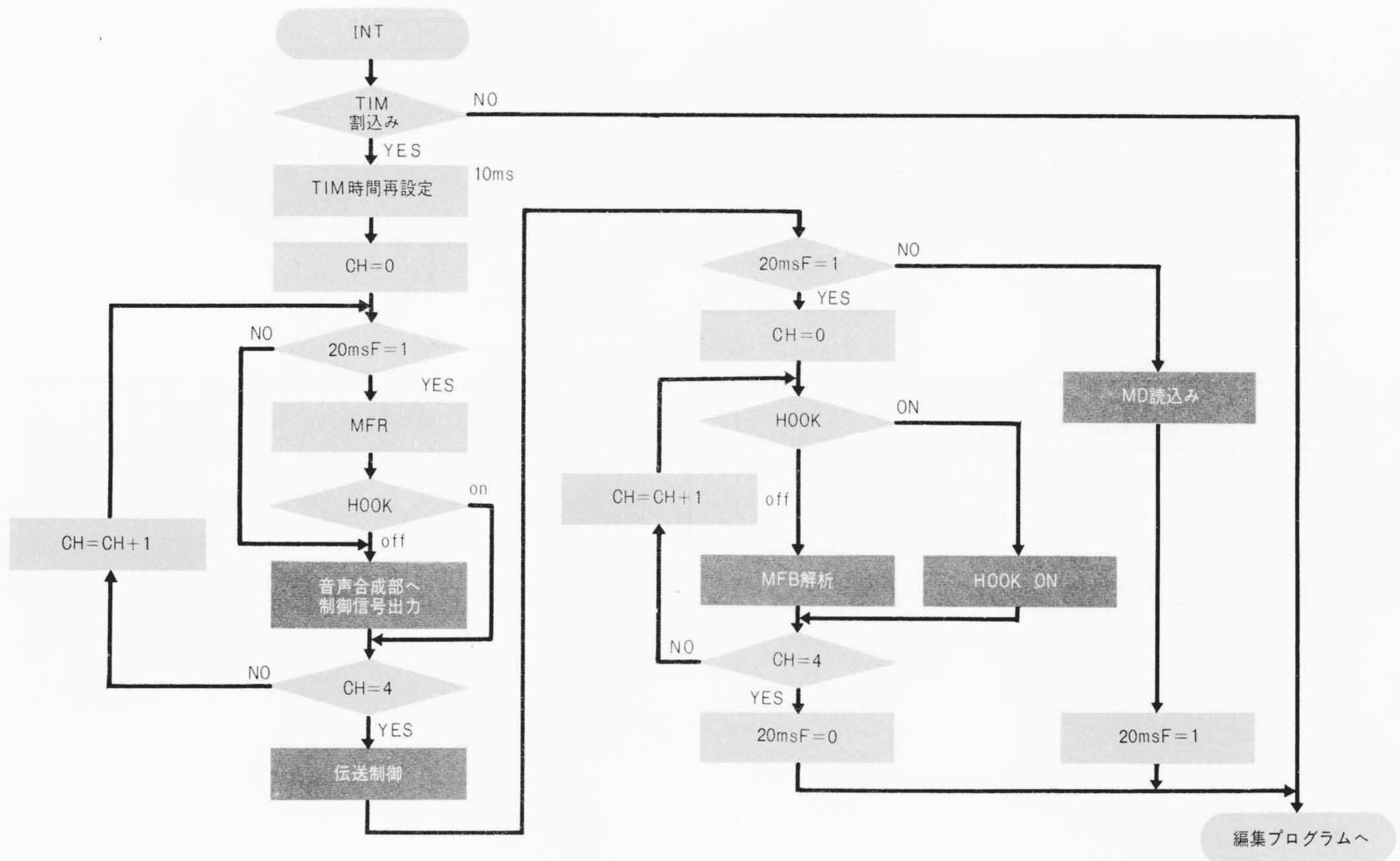


図5 制御フローチャート タイマ割込みにより、くり返しこの制御動作が行なわれる。  
Fig. 5 Flow of Control

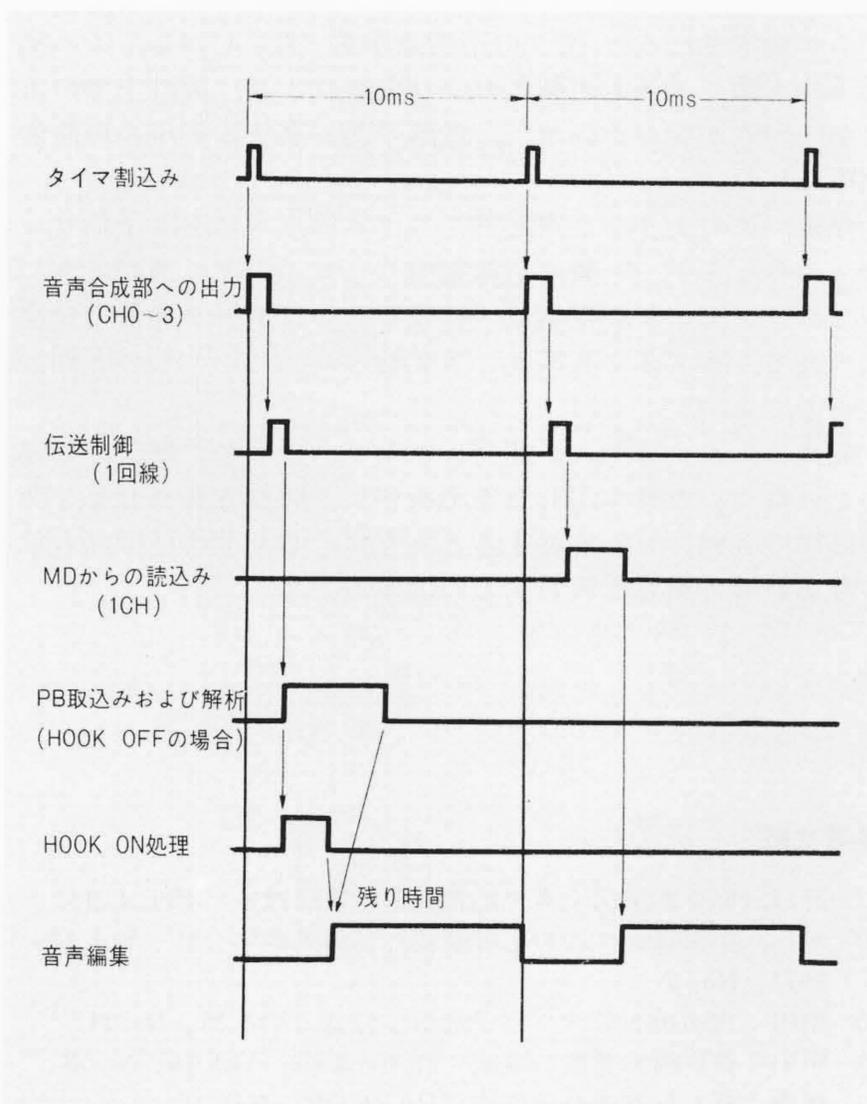


図6 プログラム時間配分 プログラムは時分割で、10msごとにより繰り返すように配分されている。  
Fig. 6 Timing Chart of Control

### 7.2 音声編集プログラム

押しボタン信号取込み、あるいは伝送制御プログラムにより、文形編集レジスタの該当回線のエリアに応答文形バッファの各先頭アドレスを格納する。

応答文形は図7に示すように、固定部分と可変部分とから成っている。次に応答文形バッファの内容をドラム アドレスレジスタに移し、可変部分が出てきた場合は指定されたバッファから可変情報を得る。該当回線の編集が終わったら、次の回線に同様の制御を行なう。

編集を終わったドラム アドレス レジスタよりドラム アドレスを読み出し、磁気ドラムにIOC命令により送出する。磁気ドラムからの制御情報読み込みは、高速間接モードで行なわれ、読み出されたデータを音声合成部へ出力するバッファに格納する。

音声合成部への制御情報は、出力サブルーチンで出力バッファの内容を10msごとに読み出しIOC命令で出力する。

### 7.3 伝送制御プログラム

伝送制御は、H-9411ビデオ データ端末装置用のラインプログラムCM911で動作するようになっている。

伝送制御手順は中央起動方式を使用しており、ポーリングにより予約情報をセンターに送り、センターで回答作成後、セレクションで回答情報を送ってくる。

### 7.4 実験データ採取プログラム

本システムは電話予約の手順および操作性の検討を行なうことを目的としているため、A-1、A-2システムの比較や、入力項目、入力けた数など入力条件によりまちがう頻(ひん)度、入力操作時間および応答文をくり返し聞く回数(回数)のデータを採れるようにしてある。

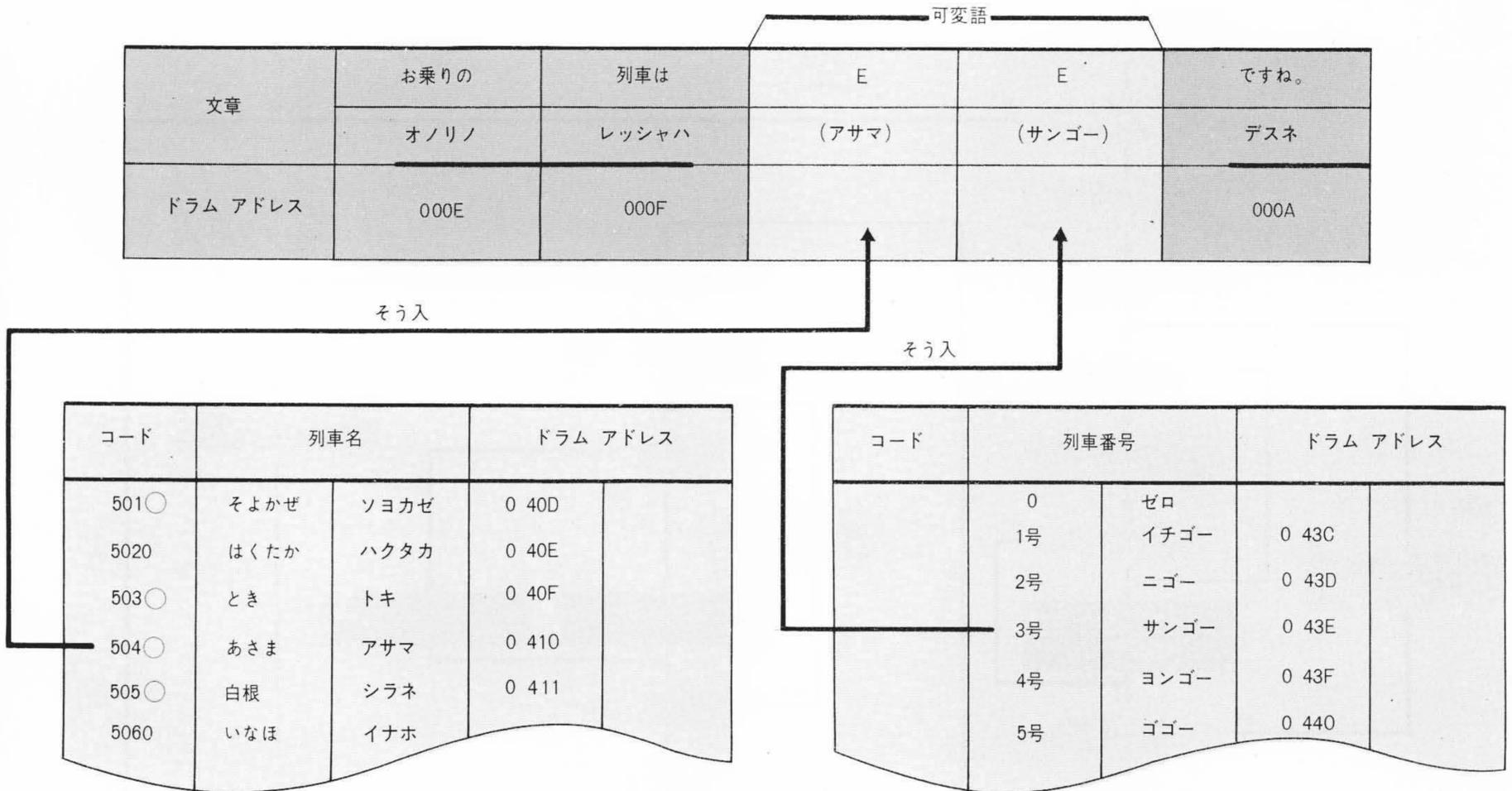


図7 文章の編集 文章は固定文の中に、可変語がそう入され出力される。  
Fig. 7 Format Edition

### 8 実験結果

実験は昭和47年3月、日本国有鉄道・本社と同・秋葉原通信区を鉄道電話回線で接続して行なわれた<sup>(6)</sup>。

模擬客は、あらかじめ用意された時刻表(列車名と駅名コードが印刷されている)を参照して申込用紙に記入してから電話をかけた。音声応答装置からの回答内容は原則としてメモを取ることとし、特に回答してきた予約番号は記入することを依頼した。

実験は4日間にわたって実施され、この間の操作延人数は520人(内女性49人)、延操作回数は約1,200回であった。

この実験の結果、電話回線保留時間は平均250秒で、これは応答文形、音声品質によって多少変化するが、ほとんどの人が保留時間3分以内で予約が可能であることがわかった。

次に、ボタンの押し誤りについては、一般に音声応答装置から誤りを指摘されたものが多かった。これは操作者の誤りによるものは少なく、装置側で出力した音声が多周波受信器にまわり込むための信号の取込み誤りと判明し、操作手順を、入力モード、出力モードと切り換える方式とすることとした。

次に、予約番号の聞き誤りについては、記入した424人中、正しく記入してあるものが89%であった。これは音声の品質にも関係するが、予約番号のみの復唱および再入力チェックなどの対策が必要であることが判明した。

上記を考慮して今後の電話予約システムは、A-1システム形で、確認文、指示文はなるべく要領よく聞きのがしのないものとし、操作手順は途中入力を認めない、いわゆる半二重方式とする必要がある。

### 9 結 言

以上、音声応答装置を利用した電話予約実験システムの方式、装置構成およびその動作について述べた。この公衆電話

回線と接続した電話予約システムは、方法そのものが新しい試みであるため(不特定多数の人々が容易に利用できるシステムが要求される)、種々の不明点があった。しかし、この電話予約実験システムの製作および実験により、設計上での重要な問題点が明らかにされ、電話予約システム実現の可能性が得られた。

今後、このような音声応答システムは、単に座席予約サービスのみならず、一般の情報案内サービスなどに適用でき、押しボタンダイヤル電話機の普及ならびにデータ通信の発達につれて、ますます必要性和重要性を高めていくものと期待される。

最後に、本システムの開発にあたり、種々のご助言、ご協力をいただいた日本国有鉄道桑折科長、善如寺科長および関係部局の各位に厚く感謝するとともに、日立製作所関係各位にもあわせて謝意を表わすものである。

### 参考文献

- (1) 野口、麻生、末広：音声応答装置 電子技術 45年4月号
- (2) 木村、中田ほか：音声応答装置 情報処理学会誌 Vol. 12, 1971, No. 7
- (3) 中田：音声出力装置 電子通信学会誌 Vol. 51, No. 11,
- (4) 中田：音声出力装置の現状 計測と制御 Vol. 10, No. 8
- (5) 板倉：新しい音声合成方式“PARCOR”日経エレクトロニクス 1973, 2, 12
- (6) 桑折ほか：プッシュホンによる座席予約実験 日本鉄道サイパネティクス・シンポジウム 1972, 11