

音声応答装置*

木村幸男** 市川 焘*** 中田和男***
兵藤剛士**** 麻生 哲****

Abstract

Recent development of information processing networks requires the further expansion of the function of audio response unit (ARU) to increase response vocabulary and number of multiplex channel.

The audio response unit reported in the paper meets these requirements by the principle of speech synthesis of combined sum of acoustical elements, which are damped sinusoidal waves for voiced sounds and bandpassed noises for unvoiced sounds.

These elements are pre-recorded on core memory because the memory capacity needs less than 40 k bits only.

The control data for speech synthesis are also less than 3,200 bit/sec and a magnetic drum of 512 tracks and 2 MHz clock can store more than 2,000 words of 0.8 sec, duration with 20 m sec or less maximum access time.

Digital store of acoustical elements on core memory and relatively simple processing of speech synthesis enables multiplexing operation by time division, and the possible number of multiplex channel is 40 under the conditions described in the paper.

In many case, response messages have several pre-determined formats, accordingly, only the assignment codes for message format and variable words in the message are transmitted from CPU to ARU and one telephone line of capacity no more than 120 Baud can transmit all necessary codes for multiplex operation of 40 channels.

1. 緒言

情報化社会の到来と、通信回線開放の気運にうながされて、計算機による情報処理の結果を音声の形で遠隔の利用者に伝えるための装置「音声応答装置」の実用化が最近強く要求されてきた^{1,2)}。録音編集という従来技術によるものは、万博の会場案内とか加入データ通信による簡易計算サービスとかで、すでに実用化されているが、本格的な情報サービスを行なうためには、もっと応答語数の多い音声応答方式が要求される。

われわれは国鉄の「座席電話予約試験システム」³⁾用に、上記の要求にこたえるような音声合成による多重

化音声応答方式を開発し、装置を試作したので、ここに、その概要を報告する。

この方式の特徴は、(1)既存の計算機用磁気ドラムを用いて、数千語の語いをもつ音声応答を行なうことができる。(2)既存のミニ・コンピュータを主体として、12~40回線の多重応答を行なうことができる。(3)簡易化と多重化によって、回線あたりのコストを低減することができる。の3点に要約される。

従来の録音編集方式にくらべて、多重度においてはやや劣るが、応答可能な語数において数十倍であり、従来のアナログ的な音声合成方式に比べては、語いと品質がほぼ同じでありながら、数十回線の多重化が可能である。

実用的な音声応答装置を試作するためには、Table 1に示すような諸要素について、与えられた仕様に従って原理的な検討を行ない、方式を決定し、設計しなければならぬ。

われわれに与えられた要求仕様は、次のようであっ

* Development Audio Response Unit, by Yukio Kimura (Laboratory of Railway Technical Institute, J. N. R.) Akira Ichikawa and Kazuo Nakada (Central research Laboratory, Hitachi Ltd.) Takeshi Hyodo and Tetsu Aso (Tozuka Works, Hitachi Ltd.)

** 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

Table 1 Design Factors of the Audio Response Unit

	Software	Hardware
Speech Synthesis	method of synthesis -process of synthesis -multiplexing control of synthesis -number of parameters -amount of information -production of information	processor memory { -capacity -access time memory { -capacity -access time analyzing facilities
	multiplex control -synchronous/ scheduling -word-wise/frame assignment communication control data transmission procedure	main controller MODEM I/O interface

た。

- (1) 応答可能な語数が 1,500 語以上であること。
- (2) 30 回線程度の多重化が可能であること。
- (3) 中央の主計算機と音声応答装置との間は、1,200 ボー電話 1 回線で制御可能なこと。
- (4) 音声応答出力は了解度 100% であること。
- (5) 装置としてできるだけ低コストであること。

以下、Table 1 の項目順に、この要求仕様をいかに満足させていったかを述べる。

2. 音声合成の原理

要求仕様の (1) から録音編集方式では技術的に困難であり、経済的にも引き合わないことは明らかだから、何らかの意味で音声合成の原理によって、音声を作り出さなければならない。しかし、従来のアナログ的な音声合成方式では、多重化を低コストで実現することはむずかしい。そこで、新しい音声合成の原理が要求される。たまたま、われわれは多重化に適した簡易な音声合成の方式を研究しており、「音声の音響素片のピッチ制御組合せ編集」方式を考案し、実験的にも実用化の見通しを得ていた⁴⁾ので、この方式を採用することにした。

2.1 音声の音響素片

音声波は声帯の振動による空気の脈流が口（詳しくは声道）による共鳴作用を受けて、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かえすことになる。口の共鳴作用は、口の形（あごや舌を動かすことによって

変わる）によって変わるが、各共鳴モードに対応する単 1 共鳴の和に分解することができ、共鳴周波数の低い方からこれを第 1 フォルマント、第 2 フォルマント、第 3 フォルマント、……という。各フォルマントは単 1 の共鳴モードに対応するものであるから、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もった減衰正弦波で表わされる。そこで、この減衰正弦波を音声の音響素片と名付ける。

2.2 音響素片の組合せ編集

電話による音声伝送によって確認されているように、音声の了解に必要な伝送周波数帯域は 3~4 kHz まででよい。この周波数帯域には通常第 1 から第 3、または第 4 までのフォルマントが存在する。いいかえれば、3~4 個の音響素片を組み合わせれば、音声を合成することができる。音声はわれわれの聴取機能によっては、離散的な音素として聞きとられるが、その物理的な表現としての音声波の特性はほとんど連続的な変化を示し、この連続性を適切に近似できる方法でなければ、高品質の音声を合成することはできない（単音節の単純な編集によっては、連続的な音声を高品質で合成できない理由の 1 つ）。そこで、実用上は何個の音響素片を準備しておけばよいかが問題になる。

2.3 組合せ編集におけるピッチ制御

簡単化のために声帯の振動、いいかえれば、音源としての励振は周期的にくりかえすものと仮定してきたが、実際はそうではなくて、言葉のアクセントやイントネーションによって複雑な変化をする。この励振音源のくりかえし周期の逆数をピッチ（周波数）としており、この変化を適切に近似できるような合成法でなければ、自然な聞きとりやすい音声を合成することはできない。

音響素片の組合せ編集による合成においては、素片をメモリから順次的に読み出す周期をピッチ情報によって外部的に制御すればよい。後に述べるように、本方式では音響素片の記憶に必要な情報量が少ないため、記憶媒体としてランダムアクセスのコアを使うことができる。したがって、読み出し周期の制御を自由に外部から行なうことができ、適切なピッチで音声を合成することができる。

2.4 音響素片の記憶に必要な情報量

上記の音声合成の原理が実用的に意味をもつためには、素片としてどのような減衰正弦波を何個記憶しておけばよいかはまず問題である。

(1) 素片の数

種々合成実験を行なって検討した結果、Table 2 に示すように、第1フォルマント用として等差周波数間隔で16個、第2および第3フォルマント用として等比周波数間隔でそれぞれ23個と10個、第4フォルマント用として第3フォルマント用の最高周波数のもの

Table 2 Number of the Acoustic Elements of Speech

	Frequency Range (Hz)	Way of Allocation	Bandwidth (Hz)	Number of Elements
1st formant	240~840	constant difference $\Delta f=40\text{Hz}$	50	16
2nd formant	800~2,340	constant ratio $r=1.05$	70	23
3rd formant	2,200~3,411	constant ratio $r=1.05$	100	10
Nasal consonants	/m/ and /n/ /N/			2

を1個(転用)、特殊なものとして鼻音 /m/ と /n/ (撥音の /N/ にも流用) にそれぞれ1個、計51個で十分であることを確認した。

(2) 素片長

素片の長さは合成しうる音声のピッチ周期をどこまでとるかによって決まる。ピッチは一般に男性の方が低く、しかも、それが100 Hz以下となることは、語頭とか語尾とか音声の振幅が小さいところであり、合成実験の結果それを100 Hz以下にならないように限定しても、合成音声の了解性や自然さをそこなうことがほとんどないことが確認された。そこで、素片長としては10 msecをとることに決めた。

(3) 必要情報量

素片を8ビット、8 kHz サンプリングのデジタル記憶をすとして、それに必要な容量は、 $8 \times 8 \times 10^3 \times 10 \times 10^{-3} \times 51 = 32.64 \text{ k}$ ビット、この情報量はコアメモリで十分記憶可能な量である。

素片用メモリとしてコアを使うことができれば、高速ランダムアクセスが可能となるから、ピッチ制御が容易になるばかりでなく、時間分割多重読み出しによって、1つの素片メモリを多数の回線に共通に使用することができる。

2.5 合成のための処理

音声合成のための処理(演算)は

(1) 素片メモリから制御情報による第1~第4フォルマントに対応する素片のサンプル値の選択・読み出し。

(2) 各サンプル値の単純な加算。

(3) 加算結果と振幅情報との乗算による振幅制御。

(4) D/A 変換による音声波形への変換。

(5) ピッチ周期によって周期的に制御される素片サンプル読み出し番地の設定(先頭番地への復帰)と、制御情報のかきかえ。

である。

この合成処理は極度に簡単化されており、理論的な観点からはかなりあい近似となっているが、そのための合成音声の品質の劣化はわずかである⁴⁾。

いま、この合成処理(8ビットの加算3回と乗算1回)を τ μsec 間に行うことができれば、時間分割処理による多重度は、 $N=125/\tau$ となり、 $\tau=4 \mu\text{sec}$ と仮定すれば $N \approx 31$ となり、30回線以上の多重化が可能となる。

3. 合成のための制御

音声合成の原理によって音声を作り出すためには、合成処理のほかに、そのための制御が必要である。

制御の方法には合成のつど、実時間的に制御情報を作り出していく方法(規則による合成の実時間処理)と、適当な単位であらかじめ記憶しておいた制御情報を応答内容に応じて選択編集して用いる方法(記憶情報による編集制御)とがある。われわれは多重応答の要求と、現状における合成音声の品質から考えて、後者、とくに人間の音声から分析によって必要な制御情報を抽出して記憶しておくという方式をとった。

3.1 制御情報の種類と量

簡単化のために音声はすべて減衰正弦波を素片として合成できるものとして話しを進めてきたが、実は、それらは有声音といわれるものであり、それ以外に無声音といわれるものがあり、とくに子音を合成するために必要である。しかし、これらも特定の周波数特性をもったランダム雑音を音響素片と考えれば、全く同

Table 3 Control Data of Speech Synthesis

Parameter	Number of Elements	Levels of Control	Information Bits of Data
1st formant	18*		5
2nd formant	23		5
3rd formant	10		4
Pitch frequency		16	4
Voice intensity		32	5
Hiss frequency		6	3
Hiss intensity		32	5
		Total	31

* Note Including 2 nasal elements.

様に処理することができる。

そこで、本方式による音声の合成に必要な制御情報の種類は、Table 3 に示すような7種類であり、それぞれに必要な情報量は同じく Table 3 に示すように計 32 ビット以下となる。

制御情報を更新する時間間隔は、合成実験の結果 10 msec 間隔以内であればよいことがわかった。

したがって、1秒間の音声を合成するに必要な情報量は 3,200 ビットとなる。これは音声波形のそのものを 8 ビット、8 kHz サンプリングでデジタルに記憶しておくのに比べて、1/20 の情報圧縮となっている。

3.2 合成可能な語数の計算

本方式で合成に必要な制御情報は 3,200 ビット/秒であり、10 msec 以内のアクセスが必要である。

日本語単語の音節数構造は 90% 以上が 5 音節以内であり、1音節の平均時間長は 160 msec である。そこで合成する 1語の単位を 0.8 秒とすれば、必要な制御情報量は 2.56 k ビット/語となる。

制御情報の量は一般に大きいから、記憶媒体として磁気ドラムを使うとすれば、そのアクセスタイムは通常最大 20 msec、平均 10 msec である。音声合成の場合には、最大アクセスタイムが 10 msec 以内でなければならないから、回転速度を 2 倍にするか、二重

に記録するかしなければならない。ここでは技術的に実現の容易な後者の方法をとる。いま、ドラムのクロックを実効 1.024 MHz とすれば、1トラックの情報容量は 20.48 k ビットとなり、2.56 k ビットの 8 倍となるから、1トラックに二重記録で 4 語を記録することができる。したがって、512 トラックのドラムを使えば、1,000 語分が記録できる。現在のドラムは 2 MHz 以上のクロックで十分実用化されているから、512 トラックのドラムで 2,000 語の記憶ができる。しかし、0.8 秒以上の単語もあるので、それらには 2 語分を必要とするから、これで実効 1,500 語以上の語いが合成可能ということになる。

3.3 制御情報の作成

われわれは人間の音声进行分析して制御情報を抽出する方式をとった。そのためには、次のような処理が必要である。

- (1) 人間の発声によって必要語いを録音する。
- (2) 録音から A/D 変換によって分析用データを作る。
- (3) データを分析して制御情報を抽出する。
- (4) 抽出された制御情報を必要なフォーマットに編集し、出力する。

データの分析法としては、音声スペクトルの最尤推

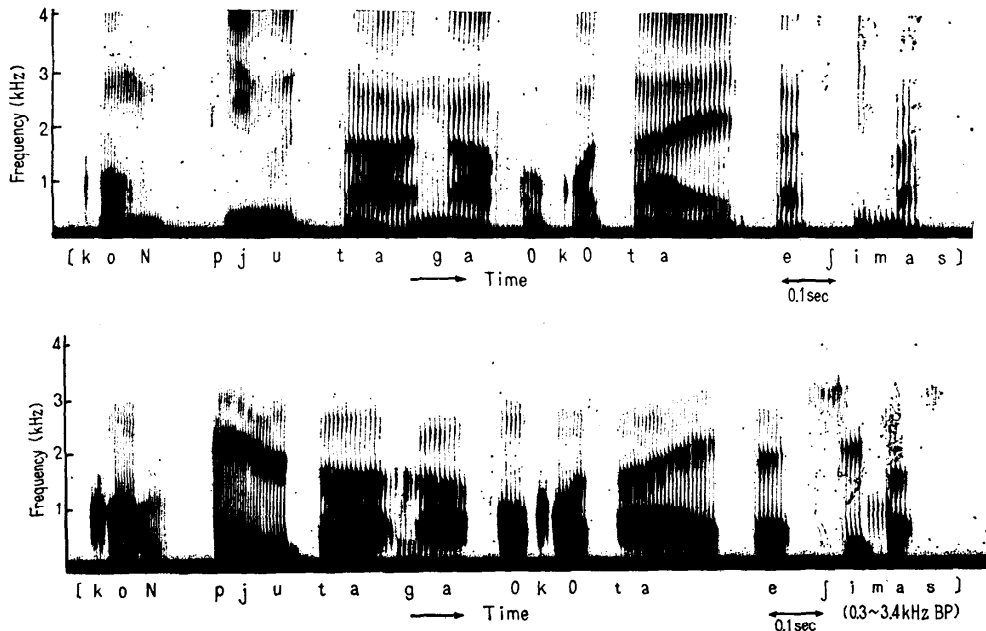


Fig. 1 Sonagrams of Original Speech and Synthetic Speech
(top: original speech, bottom: synthetic speech)

定法⁵⁾を主として利用しており、音声の振幅、第1～第3のフォルマント周波数、ピッチ周波数などを抽出する。有声・無声・鼻音の区分、無声区間の素片の選択などは、分析結果を人間がみて指定している。

以上の合成法によって合成された音声の一例を、ソナグラムによって原音声と比較して Fig. 1 に示す。

了解度のテストとして、言語的な冗長度の全くない1けた数字の連続（電話番号）を、合成音声を全くきいたことのない学生6人に聴取させた。その結果、第1回目のテストの平均正聴率は97%、2回目以後では100%であった。同様なテストを5けたのけたつき数字音についても行なったが、その結果は第1回目が98.3%、2回目以後では100%であった。

4. 応答のための多重制御

中央の主計算機からの指令によって、音声応答装置が多重応答を行なうためには、合成のための制御・演算を時間分割的に多重に処理していく制御が必要である。

ここで問題になるおもな点は、

- (1) 回線別の制御を固定時間長で同期的に行なうか、可変時間長でモニタリングとスケジューリングをしながら適応的に行なうか。
- (2) 応答すべき単語の系列を指定する情報をすべて主計算機から受け取るか、わく組（形式）の指定とその中の可変部分の単語の指定のみを受け取るか。
- (3) 制御情報を編集・送出手のためのバッファメモリを語数依存的にするか、回線数依存的にするか。

である。

4.1 同期的制御

音声合成のための制御には、完全な実時間性が要求され、今回の目的では、次節に述べるような変換編集処理が要求された。また、合成処理の単位時間長を一定としても、前後の休止区間のとり方を工夫すれば、十分自然な音声を合成することができるから、われわれは制御を簡単かつ確実にするために、回線別の制御をすべて同期的に行なう方式をとった。短い時間単位は10 msecで、10 msecごとに必ず各回線に新しい制御情報を供給する。長い時間単位は800 msecで、制御情報はこれを単位（ブロック）としてドラム上に記憶されている。ドラムのアクセスタイムは二重記録によって、実効10 msecになっているけれども、安全

をみて最大アクセスタイムを20 msecとみれば、このアクセスタイムによって決まる多重度は、 $800/20=40$ 、すなわち40回線まで多重可能となる。

4.2 変換編集制御

音声応答装置の使い方は、多くの場合応答のわく組（形式）がいくつか固定されており、その中の変化する部分、たとえば、年月日とか金額とか氏名、地名、駅名などが語いの中から選択される。このような場合には、常に応答内容をすべて逐次的に主計算機から指定することは冗長であり、回線の有効利用にならない。今回の試用例でいえば、応答形式は Table 4 に示すように、乗り継ぎのある場合とない場合の2つに分れており、Table 4 で下線の部分に変数として語いの中から選ばれる。そこで、音声応答装置が主計算機から受け取る情報は、形式の指定とその中の変数の指

Table 4 Example of Response Messages
(no exchange)

コンピュータがお答えします。メモを御用意下さい。
9月 1日 15時 40分 の はつかり 2号で 上野 から 盛岡 までですね。
グリーン座席が予約できました。
お席は 8号車 7ブロックのC 7ブロックのD 8ブロックのA 8ブロックのBです。
料金は (万)* 4千 2百 拾 円です。
予約番号は 0 0 3 の 8 0 1 9 です。
キップは 御茶の水の 緑の窓口でお受けとり下さい。有難度ございました。

Note —: variable words in the response format.

*: if the variable has no significant information in that message, the variable is omitted and its time space is occupied by following words.

定のみとし、音声として聞いて意味のわかる形に変換編集する処理は、音声応答装置側で行なうこととした。このため後に述べるように1,200ボー1回線で、多重応答に必要な情報を伝送することができるようになった。

変換編集の処理は、次のようにして実行される。

- (1) 形式ごとにその音声としての出力に必要な制御情報のドラムアドレス系列を記憶しておく。
- (2) 変数部を指定するコードから、それに対応した音声を合成するために必要なドラムアドレスを決定することのできる対応表を記憶しておく。
- (3) 主計算機からの指令によって形式が定まると、そのドラムアドレス系列をコア上のワーク・エリアに読み出す。
- (4) 変数部を指定するコードを表によってドラム

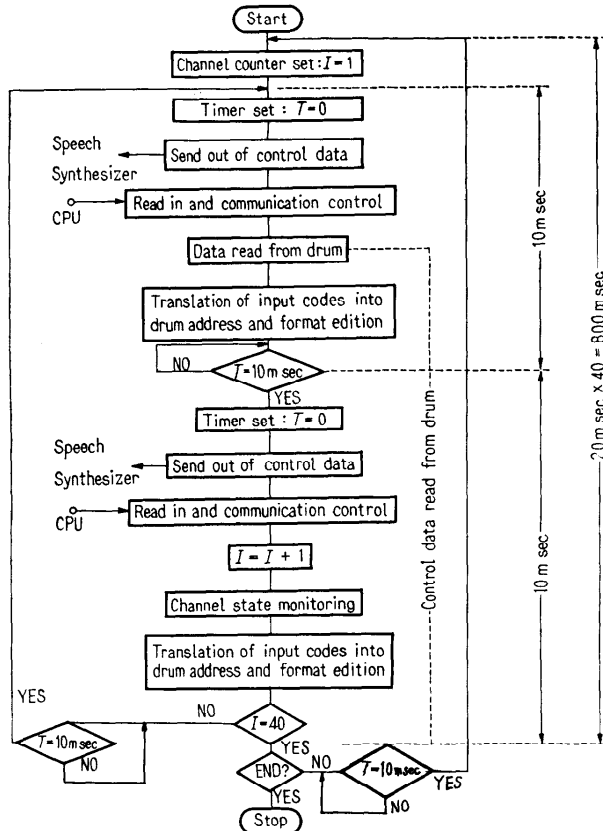
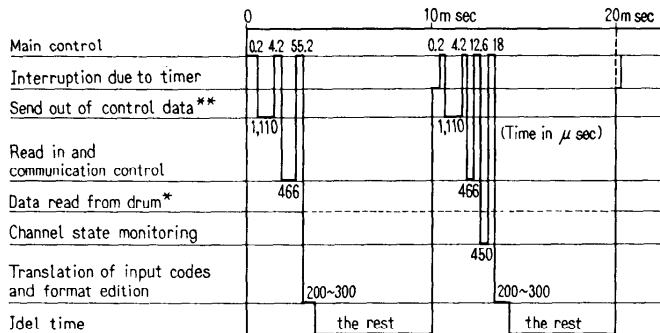


Fig. 2 Flow of Control



Note * Data are read from drum in 244 μ sec steal-time during this interval.
 ** The data 1110 μ sec is the value in the case of 12 channel multiplex.

Fig. 3 Timing Chart of Control

アドレスに変換しながら、上記のドラムアドレス系列の中へうめこんでいく。

- (5) この変換編集処理が終わったら、音声合成の制御を開始する。

実際には、この処理はミニコンピュータ HITAC-10 によって実行され、変換編集処理に必要な時間は、指定情報の受信時間に比べて無視できる。

上記の多重応答制御の流れ図と時間関係を Fig. 2, Fig. 3 に示す。

4.3 バッファメモリの容量

ドラムに記憶されている制御情報を指定コードによって選択し、10 msec ごとに回線別の制御部へ送り出すためにはバッファメモリが必要である。バッファメモリの必要量は、制御法によって語依存的になるか、回線数依存的になるかである。語依存的のときに必要な容量は、1回の制御情報量×語数であり、この例では 32 ビット×2×10³ となり、回線数依存形のときには、1語の制御情報量×回線数となり、2.56×10³ ビット×30 となる。

容量的にはあまり差がないが、前者ではドラムからの情報を語数分だけ並列的に読み出さなければならないが、後者では通常のブロック転送による直列読み出しでよい。本方式では装置の簡単化という観点から後者の方法によった。したがって、応答装置の主制御装置としての HITAC-10 に必要なバッファメモリは、16回線するとき 2.56 k 語 (16 ビット/語として)、32回線で 5.12 k 語となる。

5. 伝送制御

中央の主計算機と音声応答装置との間は、1,200 ボー 1 回線をつながれており、伝送制御手順は HITAC-9411 VDT (video data terminal) に合わせ、伝送コードは ISO コードを使用した。ISO コードは直列伝送の場合、スタート・ストップ、パリティチェックを含んで 10 ビットで 1 データが構成されるから、1,200 ボーの回線では 1 秒間に 120 データ伝送

できる。しかし、送受信の金物的な切り換え時間などを考えると、1 秒間に実質的なデータ伝送速度は 100 個程度となる。1 個のデータが音声 1 単語を指定すると仮定すれば、0.8 秒間には 80 個のデータが伝送され

るから、80回線まで多重化できるということになる。

今回の試用例では、1つの形式の内容を指定するのに必要なデータ数は、乗り継ぎなしのとき40個、乗り継ぎありのとき59個で、その音声としての持続時間は前者が約60秒、後者が約90秒である。したがって、主計算機から指令を出し始めてから内容が確定するまでの待時間は0.4~0.6秒である。

指令情報は最高120データ/秒の速さで伝送されるから、Fig. 2に示す制御の基本周期10msecの間に1~2個のデータが伝送されてくる。これをいったん金

6. 試作装置の概要

上記の設計に従い、音声応答装置の試作を行なった。試作装置の構成を示すブロック図をFig. 4に示す。本装置では主制御装置としてHITAC-10を使い、これに制御情報記憶用のドラム(MDCとCDM)、音響素片記憶用のコア(AES)、合成のための制御部、演算部とD/A変換部(SPC, CRC, SSP, LPA)、主計算機との情報伝送制御部(CMC)、多重出力のための交換機(EXG, CSR)などを付加した。その外観をFig. 5

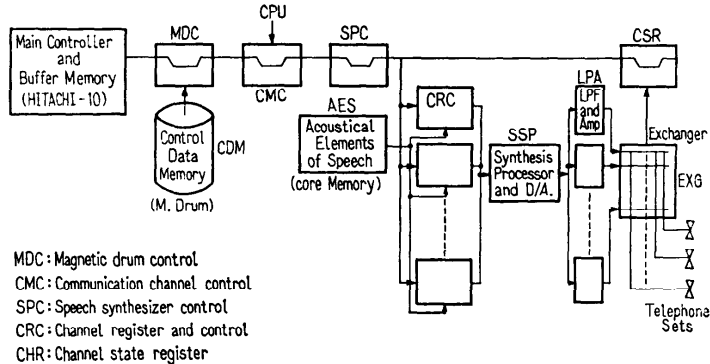


Fig. 4 Block Diagram of the Audio Response Unit

物のバッファへ記憶しておいて、10msecごとにサーチし新しい情報が送られていれば、応答装置の主制御部へとりこむ。

に示す。

伝送制御を含めて、合成演算をのぞく制御は、すべてHITAC-10のソフトで行なった。

制御プログラムは指令コードをドラムアドレスに変換するための対応表を含めて約5.5k語となった。

指令コードを音声として完全な形式にするためのワーキングエリアが最大100語/回線、制御情報のバッファエリア160語/回線であるから、多重度をますことによるコアエリアの必要量はFig. 6に示すように

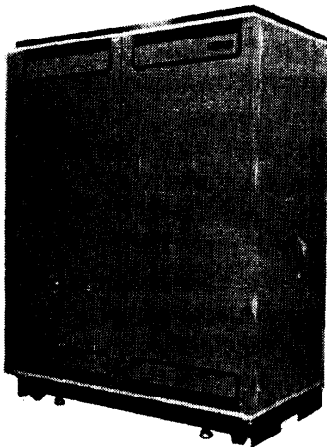


Fig. 5 Out-look of the Audio Response Unit

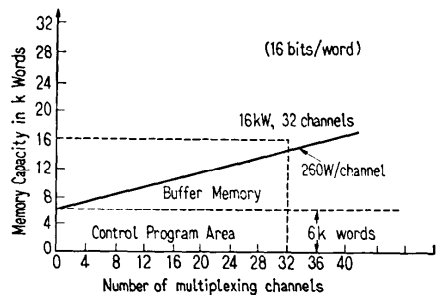


Fig. 6 Number of Multiplex Channel vs Buffer core Memory

なり、8 k 語で約 10 回線、16 k 語で約 38 回線まで多重化できることがわかった。

また HITAC-10 による制御の所要時間は現在 (12 回線多重) Fig. 3 に示すように進行しており、十分に余裕がある。

多重度を増すことによって時間がふえるのは制御信号の送出部であり、回線あたり約 100 μ sec を要している。

伝送制御を含む諸処理が約 1,000 μ sec、ドラムからのデータ転送のためのサイクル・スチール時間が約 250 μ sec、変換編集が約 200 μ sec であるから、 $[10 \times 10^3 - (1000 + 200 + 250)] / 100 \approx 85$ 、すなわち処理時間的には 80 回線まで多重化できる。試作装置の主要性能を Table 5 に示す。

Table 5 Main Description of the Audio Response Unit

Item	Description
Multiplexing	12 channels
Vocabulary	1,500 words (average word length: 0.8 sec)
Response Format	Fixed formats (several types) Selection of variable parts from vocabulary
Control	Periodical multiplex control
Communication	1,200 Baud 1 channel, half duplex center-drive
Error Detection	Horizontal, vertical parity and data definition
Error Display	Alarms on data type writer
Additional Functions	Monitoring of channel states and timing data collection for statistical operation analysis

7. 電話座席予約試験システム

本試作装置は国鉄の電話座席予約試験システムの中で試用された⁶⁾。この試験システムの構成を Fig. 7 に示す。

座席予約を希望するユーザは、電話によって近くの

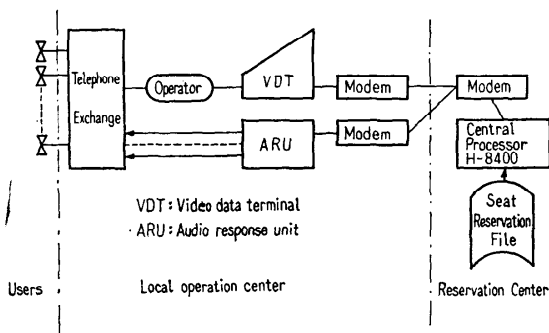


Fig. 7 Experimental Set Up of Telephone Seat Reservation System

予約センタを呼び出す。回線が接続されると係員は、ユーザの希望 (予約か照会か) を聞き、その情報を VDT のタイプキーによって入力する。主計算機はこの情報を受信すると、以後の応接に必要なホームマートを VDT 上の CRT に文字で表示する。係員はこのホームマートに従ってユーザから必要な条件を聞きだしそのつど、その情報をタイプキーで入力する。すべての条件がそろると、主計算機は必要な処理をし、結果を VDT 上に示す。ここで予約が成立しなければ、係員がさらに介入して代替案の応接にはいる。予約が成立したときは、係員は回線を音声応答装置に接続がえし、次のユーザの呼びの処理にうつる。音声応答装置は主計算機からの指令により成立した予約の内容の確認と、料金、座席位置、予約番号など必要な新しい情報を Table 4 に示すホームマートに従って回答する。ユーザはこれ聞いて必要なメモリをとり予約を終わる。

現在応答内容は 5 回までくりかえされるが、ユーザが内容を理解し受話機をおけば、それで応答を終了する。

昭和 45 年 12 月中旬に行なわれた国鉄専用線を使っての実回線テストでは、平均の聞き取り回数が 2.68 回であった。この統計では途中での打ち切りも切り上げて回数を計数しているもので、実効的には平均 2 回聞いているものといえる。

音声応答装置の導入による電話予約業務の省力化効果、またはそれと等価なサービスの向上としての処理呼数の増加については、現在各種統計データを解析中であるが、係員のユーザあたりの応接時間にして、平均 1/3~1/4 程度の省力となっている (音声応答接続時間/音声応答を使用した応接の全時間=20%~40%)。

このような係員の介入する半自動システムでは、サービスの向上をはかると必然的に係員を必要とし、経済的な運用システムとすることはむずかしい。将来形式としては押しボタンダイヤルの使用による全自動化システムが考えられるが、この場合には押しボタンダイヤル入力による情報の即時返答 (answer back) による確認や、情報の入力順序の指示など、予約の直接成立過程でも音声応答装置が必要となり、音声応答装置はますます重要になるとともに複雑な応答内容が要求される。

8. 結言

国鉄の電話座席予約試験システム用の音声応答方

Table 6 Comparison with Other Systems

System	Item	Speech Memory		Control Memory		Synthesis Process		Number of Multiplex	Vocabulary	Voice Quality	Research Institute
		Capacity (bits)	access time (sec)	Capacity (bits)	access time (sec)	Add.	Mult.				
Edition of Recorded Segment of Speech		~10 ⁶	~10 ⁰	/	/	/	/	~10 ²	~10 ²	A	Practically used
Edition of Recorded Element of Speech	Acoustical Elements	~10 ³	~10 ⁻⁶	~10 ⁶	~10 ⁻²	Simple 4	1	~10 ¹	~10 ³	B	HITACHI
	Speech Wave Elements	~10 ⁶	~10 ⁻⁴	~10 ⁵	~10 ⁻²	Simple 1		~10 ⁰	~10 ⁴	B	ETL of MITI, KDD
Speech Synthesis	Stored Data Control	/	/	~10 ⁶	~10 ⁻²	Complex 24	8	~10 ¹	~10 ³	A	ECL of NTT
	Real Time Synthesis-byrules	/	/	~10 ³	~10 ⁰	Complex		~10 ⁰	No limit	B	Universities

式を開発し装置を試作し試用したので、その概要について報告した。本方式の特徴は、音声合成の原理によっているから、現用の金物で2,000語(0.8秒/語)以上の語いを応答することができ、しかも、時間分割によって30回線以上の多重化ができ、合成過程の簡単化と相まって経済的な装置となっている。制御の面では音声合成の時間単位を一定とすることによって制御を同期的にして簡単化し、応答内容の形式と変化部分のみを主計算機から受け取り、音声として意味のある形に変換編集する処理を応答装置側でやることにして、伝送回線の有効利用をはかった。その結果、1,200ボート1回線で多重応答の制御を行なうことができた。

本方式を他の音声応答の方式と比較して Table 6 に示す。

本方式は、ほかの一般の電話予約サービス、情報案内サービスなどに適用できる一般性をもっている。押しボタンダイヤルの普及、データ通信の発達につれて、音声応答装置は今後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。

本方式の今後の問題としては、合成音声の品質をより向上させることにあり、引き続き制御情報の作成、合成音声評価の両面から研究をすすめている。

最後に国鉄の本課題の推進と指導にあたっていただいた国鉄本社通信課細野補佐以下の関係者、国鉄技術研究所システム研究室大野室長以下の関係者、同通信

研究室の方々、とくに収研究員、日立製作所コンピュータ事業部、通信機事業部、交通事業部、戸塚工場、中央研究所の関係各位、とくに戸塚工場の木広主任、小原社員、中央研究所の中野研究員、紺野嬢に心から謝意を表する。

参考文献

- 1) 石田晴久：“入出力機器の現状と動向” 情報処理, Vol. 12, No. 1, pp. 37~46 (1971)
- 2) J. L. Flanagan, et al.: “Synthetic voices for computers” IEEE Spectrum, Vol. 7, No. 10, pp. 22~45 (1970)
- 3) 鉄道通信協会：“旅行に関する総合情報処理システムの研究委員会報告書”, 昭和45年3月
- 4) 中田和男, 三浦種敏：“多重化に適した音声合成方式” 電気通信学会論文誌C, 52巻C, 10号, pp. 579~586 (1970)
- 5) 板倉文志, 斎藤収三：“統計的手法による音声スペクトル密度とホルマント周波数の推定”, 電気通信学会論文誌A, 53巻A, 1号, pp. 35~42
- 6) 細野広洋, 井上晴雄, 木村幸男他：“音声応答システムの旅客サービスへの応用” 第7回鉄道におけるサイバネティクス利用国内シンポジウム予稿集, I-124, pp. 35~39 (1970年12月)
- 6) 善如寺正雄, 木村幸男：“電話予約システム” pp. 35~39 JREA 46年4月

(昭和46年3月18日)