

既約コード法による音声合成・編集システム

中村八束* 不破 泰** 高野洋之***

(昭和61年10月31日受理)

A Speech Synthesis and Editing System Using the Irreducible Code Method

Yatsuka NAKAMURA, Yasushi FUWA
and Hiroyuki KOHNO

A new system was designed, constructed for trial, and experimentally tested for performance which might be used for easy construction of a speech indication equipment, an auto broadcasting system, a voice alarm system, etc. Conventionally, such devices have been using the vocal sound synthesis method. The present system employs the sound-registration method which previously records necessary sounds (up to 60 kinds) and which allows necessary phrases or sentences to be edited and played back under the direction by a personal computer. This method, though limited in the number of sounds to be played back, has the advantages: (1) sounds are natural, (2) recording and playing back are easy, and (3) voices other than sounds may be registered. The present system is equipped with 256 KB bubble-memories to record words. A monitor program on the system is capable of editing registered words. Sounds may be compressed by the infinite-irreducible-code method, previously proposed by the present authors.

1. 緒 言

音声は、我々にとって最も自然なコミュニケーションの方法である。コンピュータの発展にともない、マンマシンインターフェースとしての音声が必要視されてきている。最近ではパーソナルコンピュータなどにも音声合成ボードが用意され、よく利用されるようになった。しかし、発声はまだ十分自然であるとは言えない点や、アクセントなどの指定まで必要な点、あるいは音声以外の音を出せない、などの欠陥がある。そこで、我々は、上記の欠点を除いて、音声指示装置、自動放送システム、音声警報装置などを容易に実現できることを目的として、新しいシステムを設計・試作したので、その概念を以下に述べる。

* 情報工学教室 教授

** 情報工学教室 助手

***大学院修士課程

2. システムの目的と特徴

現在、音声合成の方法としては、パーソナルコンピュータなどでよく用いられている音韻合成法が主流である。これは、50音など基本のデータを登録しておき、それを組み合わせることで任意の言葉を発声させる方法である。しかしながら、自然な発声をさせるには、イントネーションの指定など利用者がおこなわなければならない、困難な点がある。また、自動販売機などに使われている、ROM に音声データを入れておき、それを D/A 変換して出力する方法もある。この方法は、発声が自然で簡単にシステムができるが、決まった言葉しか発声できないという欠点がある。また、従来の音声合成装置は、利用者が音声を登録し、それを圧縮して保存し、復号して再生するという機能しかなく、音声データを保存するためにバックアップの電源を必要とする。

本システムでは、必要な音声をあらかじめ圧縮・登録しておき、それを復号・再生する方式をとる。この方法では限られた言葉しか発声できないが、1) 発声が自然である、2) 発声や登録の方法が容易である、3) 音声以外の音も登録が可能である、という特徴もっている。そして、パーソナルコンピュータのシリアル伝送で一般化している RS-232C インターフェースをもっているので、パーソナルコンピュータなどと簡単に接続でき、利用することができる。また、従来の音声合成装置では、単に登録してある音声を再生するだけであったが、本システムでは編集して再生する機能も備えている。この機能により、登録してある音声をさらに複雑に組み合わせで発声させることも、簡単なコマンドでおこなえる。

また、音声データの保存に不揮発性メモリであるバブルメモリを用いているのでバックアップ用の電源を必要としない。

音声データは圧縮して保存するが、データ圧縮の方法として、以前我々が考案した圧縮法を用いている^{2,3)}。

3. システムの概略

図1にシステムの概略を示す。本システムは、音声の入出力部、CPU部、バブルメモリ部、に分かれている。

音声は、4KHzのローパスフィルタを通して8KHzでサンプリングをし、A/D変換器により8bitに量子化する。出力時は8bitのD/A変換器によりアナログ化した後4KHzのローパスフィルタ、アンプを通して内蔵スピーカーより出力する。

CPUはZ80Aをクロック4MHzで用いた。プログラムはアセンブリ言語で作成し、オブジェクトは約6KBでROMに内蔵している。RAMは256KBである。

音声データを保存するために、容量が128KBのバブルメモリ(FBM-M128TA)を2個使用している。

また、外部にマイクを接続するための端子と、外部のスピーカーを接続するラインアウトの端子を持つ。

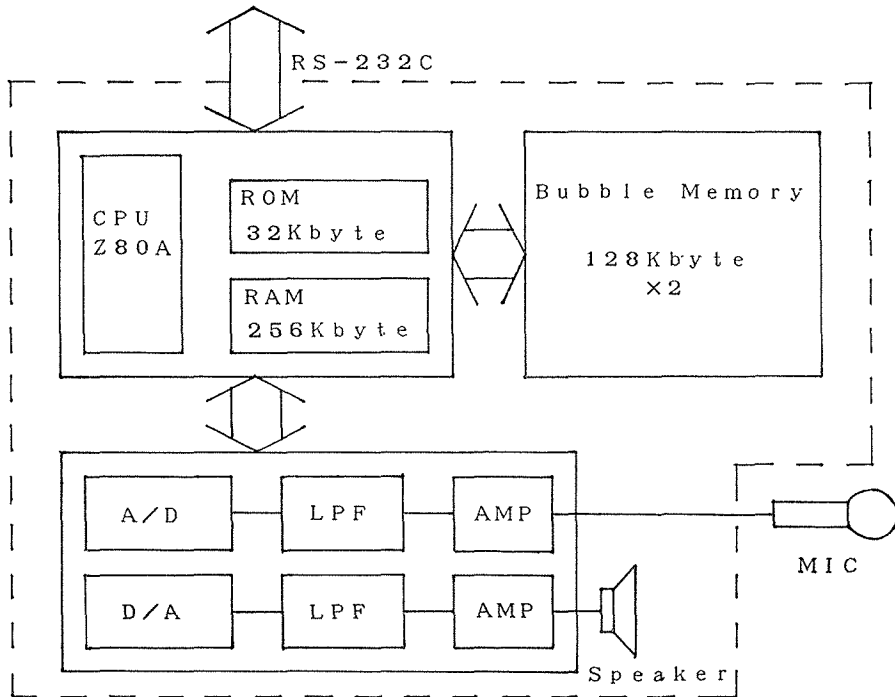


Fig.1 Block diagram of the system.

4. データ圧縮法

データの圧縮は、我々の提案による II コード法を用いた。これは、Huffman code に代表される、語頭が他の語になっていないという既約性の過程を満たすコードを無限のレベルに拡張した無限既約コードを利用した圧縮法である。無限既約コードの例を表 1 に示す。この圧縮法は、デルタ変調法、差分パルス符号変調法などと同様に、一定のステップ幅 Δ (デルタ) を決めておき、誤差の蓄積が少なくなるように今までの符号化によって得られている、予測音声信号値 (\hat{S}_{n-1}) と入力音声信号値 (S_n) の差分 (e_n) に対して符号化を行う方法で、オーバーロード雑音、グラニューラー雑音をなくすることができる。表 2 にその例を示

Table 1 An example of an infinite irreducible code system.

| Ex. 1 | Ex. 2 | Ex. 3 | Ex. 4 |
|--------------|-----------------|-----------------|--------------|
| $m_1 : 0$ | $m_2 : 1100$ | $m_3 : 1100$ | $m_1 : 10$ |
| $m_2 : 10$ | $m_1 : 100$ | $m_2 : 100$ | $m_2 : 11$ |
| $m_3 : 110$ | $m_0 : 0$ | $m_1 : 00$ | $m_3 : 000$ |
| $m_4 : 1110$ | $m_{-1} : 101$ | $m_{-1} : 01$ | $m_4 : 010$ |
| $:$ | $m_{-2} : 1101$ | $m_{-2} : 101$ | $m_5 : 0110$ |
| $:$ | $:$ | $m_{-3} : 1101$ | $:$ |
| $:$ | $:$ | $:$ | $:$ |
| $:$ | $:$ | $:$ | $:$ |

す。

この方法により、音声データは平均約 1/3 に圧縮される。

Table 2 An example of the II-Code method.

| | 1 | 2 | 3 | 4 |
|--|-------------|-------------|-----|-------------|
| value of input signal (S_n) | 29 | 19 | 11 | -1 |
| difference (e_n) | 29 | -11 | -7 | -13 |
| width of step ($\Delta = 6$) | +5 Δ | -2 Δ | -4 | -2 Δ |
| II-code | 1111100 | 1101 | 101 | 1101 |
| value of forecast signal (\hat{S}_n) | 30 | 18 | 12 | 0 |

5. システムの機能

ここでは、システムの機能について説明する。音声入力に関するもの、データ圧縮及び登録に関するもの、音声の編集に関するものの3つに大きく分けられる。利用者は、本システムに、RS-232Cを用いてパーソナルコンピュータからコマンドを送ることにより、音声の圧縮・登録、編集、そして出力をおこなう。以下に機能を示す。

[1] 音声入力コマンドと入力モード

“I” と入力すると音声入力モードにはいる。以下に I コマンドのサブコマンドを説明する。これにより音声の入力を行う。

• INPUT サブコマンド

“S” と入力すると、音声のサンプリングを始める。一度に取り込める音声は、4秒までである。

• OUTPUT サブコマンド

“O” と入力すると、S コマンドで入力した音声をそのまま出力する。入力音声のチェックを行うコマンドである。

- 音声切り出しサブコマンド

S コマンドにより音声は4秒分取り込まれるが、圧縮して登録すべき音声は、この4秒分すべてというわけではなく、入力音声を頭から必要なところだけ取り出す必要がある。このコマンドは、利用者が入力音声を取り出すのを助ける役目をする。“P”を入力すると、入力音声の無音部分を現在の位置から後ろに向かってさがし、音声データの終わりをそこに移す。“-P”とすると、現在の位置より前方に向かってさがす。

- 終了サブコマンド

“E”と入力すると、音声入力モードを終了する。

[2] データ圧縮コマンド

“X”と入力すると、I コマンドで入力した音声データを圧縮する。

[3] SAVE コマンド

“S ファイル名”と入力することにより、圧縮されたデータをバブルメモリに保存する。これにより、音声データがシステムに登録されるわけである。ファイル名は、英数字6文字以内で指定する。登録は、62種、約96秒分まで可能である。

[4] ダイレクトスピークコマンド

“LD ファイル名”と入力することにより、ファイル名で指定した音声データを、復号して出力する。このコマンドにより、登録してある音声データを出力できる。

[5] ディレクトリコマンド

“DIR”と入力すると、登録してあるファイル名の一覧を表示する。

[6] イレースコマンド

“ERA ファイル名”と入力することにより、登録してあるデータを削除する。

[7] 編集モード

“Z”と入力すると、編集モードにはいる。ダイレクトスピークコマンドは、登録してある音声を単に出力するだけであるが、以下に述べる編集機能を使うことにより、登録されている音声を自由に組み合わせて出力できる。

編集モードでは、編集用にコマンドバッファと言うものが用意されている。このコマンドバッファを用いることにより、登録してある音声をさらに複雑につなげて発声させることができる。コマンドバッファの番号は、0から255まで使用できる。

- WRITE サブコマンド

“Wn”と入力すると、n番のコマンドバッファにコマンドを書き込む。“E”を入力すると、書き込みを終了する。

- OUTPUT サブコマンド

“On” と入力すると、n 番のコマンドバッファの内容を出力する。このコマンド一つだけで編集した音声を出力することができる。

- 終了サブコマンド

“E” と入力すると、このモードを終了する。

表3のように音声登録されているとする。ここで、ファイル番号とはシステムがファイルにつけた番号のことをいう。編集モードでは、ファイル名の代わりにファイル番号で登録してある音声を指定する。

Table 3 An example of editing registered words.

| File Number | | File Number | |
|-------------|---------|-------------|------|
| 01 | 『本日の』 | 06 | 『曇り』 |
| 02 | 『明日の』 | 07 | 『雨』 |
| 03 | 『あさっての』 | 08 | 『雪』 |
| 04 | 『天気は』 | 09 | 『時々』 |
| 05 | 『晴れ』 | 0A | 『です』 |

| Buffer Number | |
|---------------|----------------|
| 1 | 01, 04, 05, 0A |
| 2 | 06, 09, 07 |
| 3 | 02, 04, B2, 0A |
| 4 | 07, 0A, R |

例えば『本日の天気は晴れです。』と発声させようとする。“W1” と入力してコマンドバッファ1に“01”，“04”，“05”，“0A” と書き込み，“E” を入力して書き込みを終了する。そして“01” と入力することにより発声する。

また、コマンドバッファにはファイル番号のほか、“Bn” (nはコマンドバッファの番号)，“R” というコマンドが書ける。“Bn” は、n 番のコマンドバッファの内容をそこに書いたのと同じ事を意味する。“R” をコマンドバッファの最後に書くと、そのコマンドバッファの内容を、RS-232C から何か入力があるまで、繰り返し発声する。例えば、“03” を実行すると『明日の天気は曇り時々雨です。』と発声する。また、“04” を実行すると『雨です。』と繰り返し発声する。

6. システムの応用例

以上述べた本システムを、詰め碁の CAI ソフトと組み合わせて使用した。あらかじめ

31種類のプレーヤーに対しての指示やメッセージを登録しておき、NECのパーソナルコンピュータ PC-9801とRS-232Cを用いて接続した。簡単な詰め碁のゲームであるが、今までの画面によるメッセージだけとは異なり、要所所でプレーヤーに対しての指示やメッセージを音声によっても出力するようにした点、プレーヤーからの評価を得た。

また、本システムとプッシュホンの認識システムとを組み合わせることにより、プッシュホン式の電話機からコマンドを送り、登録されているメッセージを聞いたり、またメッセージを登録したりというボイスメールのシステムも容易に構築できる。

7. 結 言

本システムは、必要な音声をあらかじめ登録しておけば、簡単なコマンドで音声を出力することができるので、パーソナルコンピュータ等と組み合わせることによりいろいろな利用が考えられる。さらに、従来にはなかった編集機能を利用することにより、簡単なコマンドで登録されている音声を複雑に組み合わせ発声することもできる。

また、音声に限らず動物の鳴き声、チャイム、効果音等を登録しておき、それらと音声を組み合わせることで使うことにより、さらに応用が広がるものと思われる。

参 考 文 献

- 1) 市川：音声の合成と出力，情報処理学会誌，Vol.19，No. 7，pp.657-666 (1978)
- 2) 中村，西田：無限既約コードを用いたデルタ変調法について，信州大学工学部紀要，No. 49，pp. 45-60 (1980)
- 3) 中村，西田，不破：周期性部分の圧縮と無限既約コード利用の音声合成法，電子通信学会論文誌，Vol. J 66-D，No. 2，pp. 135-142 (1983)