

# 画像応答システムの音声ファイル装置

## Audio Files for Video Response System

広い利用者層を対象とする画像応答システムでは映像と同様、音声情報も不可欠である。このためのセンタ設備として、コンピュータの制御のもとで多数の加入者に共同で利用できる音声ファイルを開発した。

実験システムでは、音声ファイルは音声応答装置、画面別音声ファイル及び複合音声ファイルから構成される。前二者は音声信号を符号化して蓄積再生するデジタル形であり、後者は音声信号をそのまま蓄積するアナログ形である。特に、画面別音声ファイルについてはPARCOR方式による情報圧縮方式を採用し、経済的な音声ファイルを実現した。

土師克己\* *Haji Katsumi*  
中村雅史\*\* *Nakamura Masashi*  
原田泰亘\*\* *Harada Yoshinobu*  
松井康夫\*\* *Matsui Yasuo*

### 1 緒言

画像応答システム(VRS: Video Response System)は企業用だけでなく一般家庭までの広い利用者層を対象とした「だれにでも」利用できる新しい情報システムである。端末利用者への情報として映像だけでなく、音声などを付加した総合的な視聴覚情報として提供することにより、サービス機能を飛躍的に増大することができる。音声ファイルは、このためのセンタファイル装置で、コンピュータ制御のもとに多数の加入者が共同で利用する共通ファイルである。

VRSサービスでの音声、効果音などの使い方には種々の形態がある。これらを大別するとシステム共通のメッセージ、画面に固有の説明をする画面別説明音声、サービスの表現効果を盛り上げる背景音楽、効果音などである。

このような多岐、多種類にわたる音声、音楽などをすべて単一の音声ファイルに収納するのは経済的ではなく、蓄積容量、アクセス頻度などにより使い分ける構成とし、実験システムでは音声応答装置、画面別音声ファイル及び複合音声ファイルを開発した。

### 2 音声ファイル系の構成

#### 2.1 システムからの要求条件

VRSサービスのうち動画の付帯音声、音楽はVTRやテレビネフィルムサウンドトラックに録音されているものを端末のテレビジョン受像機にそのままシーケンシャルに送出される。しかし、静止画サービスでは多数の利用者が、コンピュータのプログラム制御のもとに音声ファイルを共用するため、高度なランダムアクセス性が要求されるとともに、画面単位の音声管理が必要となる。

静止画サービス用音声ファイルに対する要求条件を要約すると次に述べるとおりとなる。

##### (1) ランダムアクセス性

極力小さい待合せ時間で音声、音楽の開始、終了の制御が可能なこと。

##### (2) 音声品質

テレビジョン放送に遜色のない音声、音楽の品質であること。

##### (3) 多数語

サービスに必要な多数の言葉を蓄積し、必要に応じて再生

送出できること。

##### (4) 経済性

安価なファイル構成であること。

#### 2.2 音声ファイルの構成

上記のシステム条件は、(1)項を除き汎用の音声テープレコーダで十分満足できる。しかし、ランダムアクセス性、すなわち数十分ないし数時間のテープ長に1秒以下で、どこにでもアクセスでき、かつ多数の利用者からのランダム呼に対処するのは本来シーケンシャルの音声テープレコーダでは実現不可能である。このため、ランダムアクセス性の高いシステムメッセージや画面別音声用ファイルとして音声をデジタル化してコンピュータ系列のファイルメモリに蓄積する即時アクセスの音声ファイルを開発した。また背景音楽、効果音などをアナログ形式で録音再生する複合音声ファイルを開発した。

図1に静止画用音声ファイルの構成系統図を示す。音声応答装置はデジタル化した音声単語をあらかじめ磁気ドラムに蓄積しておき、中央処理装置からの指令に基づき制御部で編集し、文章音声の形式で出力する装置である。

画面別音声ファイルは画面単位の説明音声をデジタル化し、滞域圧縮してあらかじめ磁気ディスク装置に蓄積しておき、中央処理装置の指令により検索し、もとの音声に変換してファイル出力とする。

複合音声ファイルは、小形の音声用アナログディスクやランダムアクセスの音声テープを記録媒体として用いており、効果音や背景音楽などを収録している。これらにあらかじめ録音されている音声、音楽を検索し出力する。

### 3 音声応答装置<sup>1)</sup>

#### 3.1 基本方式

表1に音声応答装置の基本方式を示す。この装置は、最大512単語/1秒長の容量をもち、これらを任意に組合せ編集して、まとまった音声情報を出力する装置で、アクセス時間は平均0.5秒、最大64の回線に同時に異なる応答をすることが可能である。

\* 日本電信電話公社技術局 \*\* 日立製作所戸塚工場

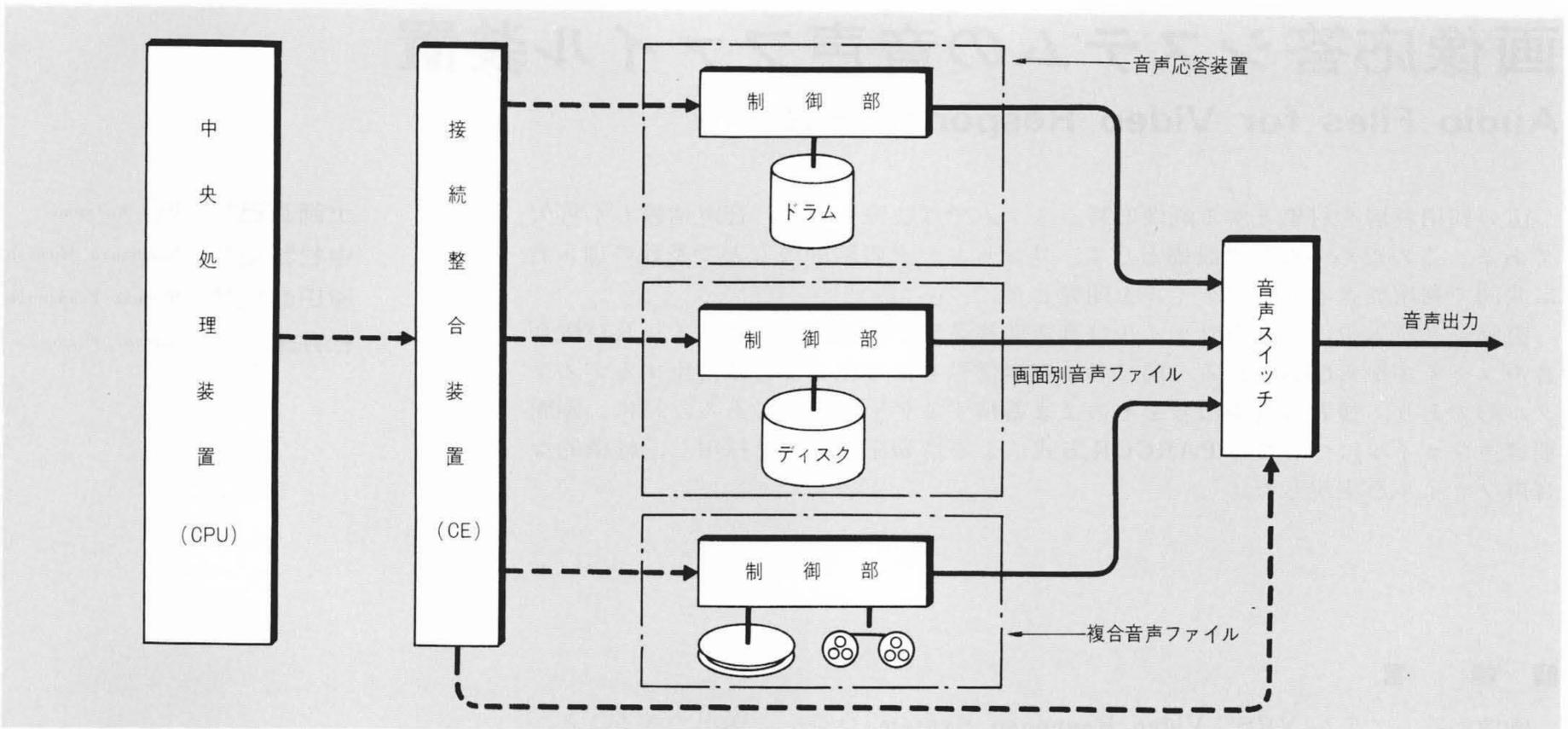


図1 音声ファイルの構成 静止画用音声ファイルと周辺装置との接続系統を示す。

表1 音声応答装置の基本方式 音声応答装置の主要仕様を表形式にまとめて示した。

項目	方式
基本方式	単語編集方式
記録方式	デジタル記録方式
音声符号化方式	7bit直線PCM方式
音声素片	単語又は文節
音声ドラム	デジタル磁気ドラム 記憶容量：4Mバイト アクセス時間：平均10ms
収容語数	512単語/1秒長
多重化方式	時分割多重PCM方式
出力回線数	最大64回線(24回線実装)
音声周波数帯域	0.3~3.4kHz

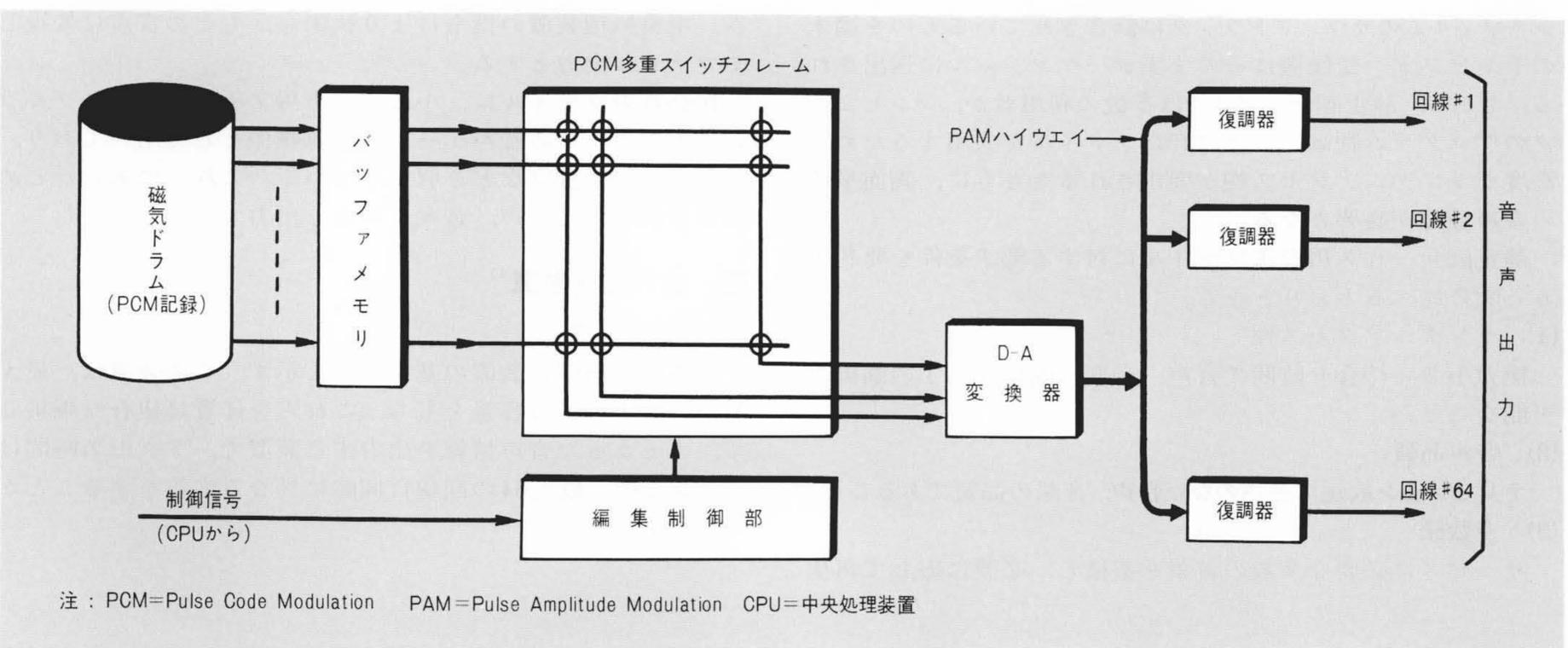
### 3.2 動作概要

図2は音声応答装置の動作原理図である。磁気ドラム装置にあらかじめPCM(Pulse Code Modulation)形式で記憶されている音声素片は512単語が常時多重形式で読み出されており、バッファメモリを経由して、PCM多重スイッチフレームに加えられる。

PCM多重スイッチフレームは、単語と音声出力回線との対応をPCM多重形式のまま変換し、中央処理装置の指令に基づき、所定の回線に所定の単語を編集しD-A変換器に送出する。

D-A変換器の出力は、時分割多重のパルス振幅変調(PAM)形式の信号であり、復調器を介してアナログの音声出力信号が得られる。

磁気ドラムへの記録は標本化周波数8kHz、7bitの直線符号化であり、蓄積可能な単語は1秒換算で512語、信号の周波数帯域は0.3~3.4kHzである。



注：PCM=Pulse Code Modulation PAM=Pulse Amplitude Modulation CPU=中央処理装置

図2 音声応答装置の動作原理 PCM時分割多重方式の音声応答装置の動作原理図である。

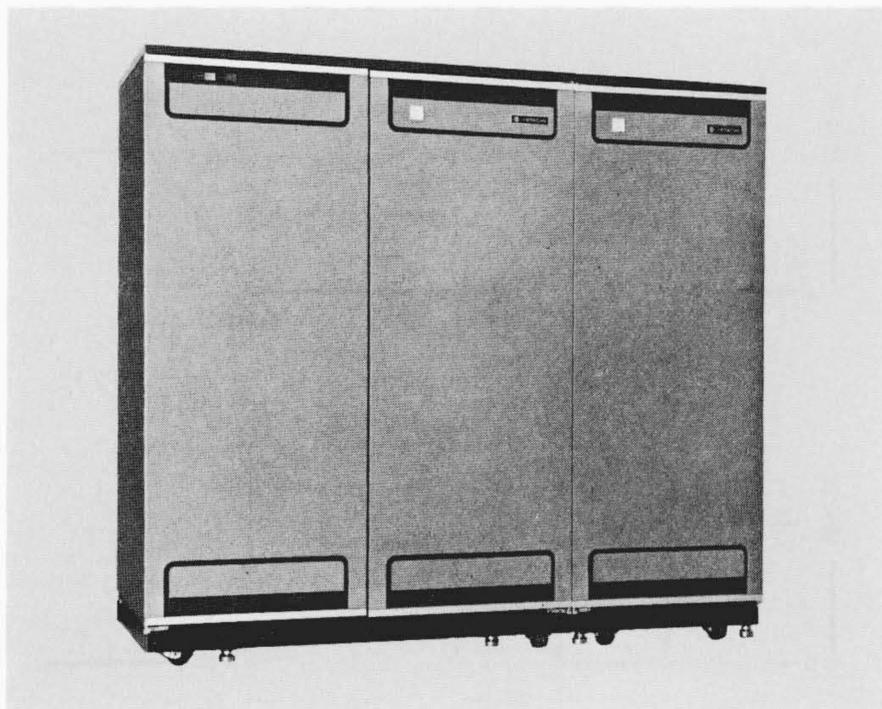


図3 音声応答装置の外観 音声応答装置の外観を示す。外形寸法は、幅1,890×奥行630×高さ1,670(mm)である。

3.3 装置構成

この装置は、音声応答制御装置及び音声応答磁気ドラム装置から成り、外観を図3に示す。この装置はコンピュータの標準ラックに実装されている。

4 画面別音声ファイル装置<sup>2),3)</sup>

4.1 基本方式

音声信号を単にデジタル化し蓄積する方式では、ファイルメモリとして膨大な記憶容量を必要とする。例えば、8kHz標本化、8bit符号化のPCM方式では、10秒の画面別説明音声に80kバイトを必要とし、100Mバイトの大形集団ディスクパックを用いても高々1,250画面の説明しかできない。したがって、音声信号についてなんらかの情報圧縮を行ない、ファイル記憶容量の効率的使用を図ることが望ましい。

音声信号の情報圧縮には種々の方式があるが、ここでは日本電信電話公社・電気通信研究所で開発されたPARCOR(Partial Autocorrelation: 偏自己相関係数)方式を採用した。この方式は、音声情報の了解性を損なわずに約3分の1の情報圧縮が可能な方式で、ファイルメモリの経済化を図ることができる。

表2はPARCOR方式による画面別音声ファイルの基本方

表2 PARCOR形画面別音声ファイルの基本方式 PARCOR形画面別音声ファイルの主要仕様を表形式にまとめて示した。

項目	基本
音声応答方式	編集合成方式
音声素片	単語又は文節
音声記憶装置	磁気ディスクパック装置(100Mバイト/パック)
音声記録方式	単語一括記録方式
収容語数	約7,500語/10秒語/パック
音声合成方式	PARCOR方式 回線多重数 音源部：32回線 合成部：8回線 フレーム周期：10ms フレーム当たりの情報量：96 bit
音声の編集制御	回線バッファ方式
同時処理回線数	最大32回線×2=64回線

式で、100Mバイトの磁気ディスクパック当たりで約7,500語/10秒語蓄積可能である。PARCOR音声の合成部は、経済化のため時分割多重方式を採用し、同時に8回線の多重処理を行なう。

4.2 動作概要

図4にPARCOR方式による画面別音声ファイルの原理を示す。音声信号は運転に先立ちPARCOR形式で画面別に磁気ディスクパック装置に収納されており、主制御装置からの指令によりこれを音声多重合成装置のバッファメモリに読み出す。磁気ディスクパック装置とこのバッファメモリ間のPARCOR情報の転送は、3.3秒長の音声に相当する単位で行なう。バッファメモリはこれに相当する記憶容量(4kバイト)をもち、回線対応に実装されている。音声出力は最大32回線で、今回のシステムでは24回線実装とした。

磁気ディスクパック装置と磁気ディスクパック制御部及び主制御装置は、画像ファイルの一種であるデジタル図形ファイルと共用し、システムの経済化を図っている。

5 複合音声ファイル

5.1 基本方式

複合音声ファイルは、サービスの表現効果を盛り上げるためのファイルで、中央処理装置からの指令によりセミランダ

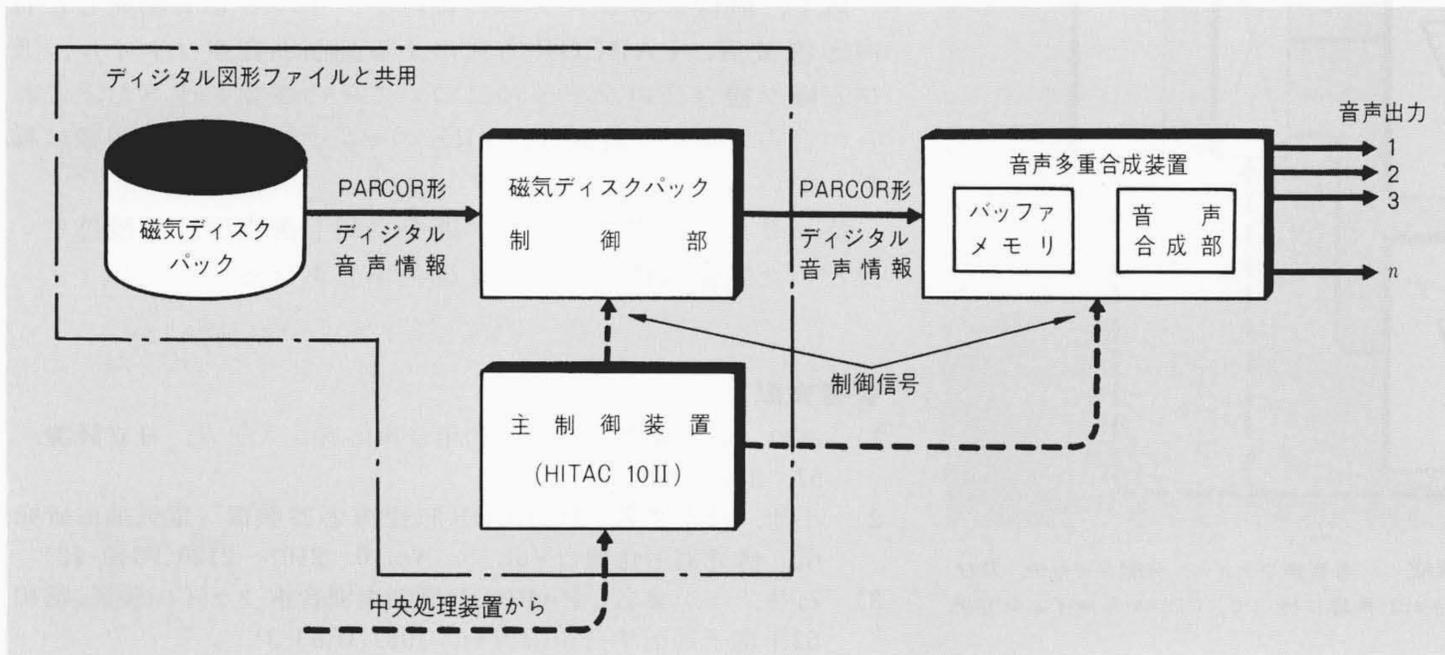


図4 PARCOR形画面別音声ファイルの原理構成 偏自己相関係数を用いて、情報圧縮を可能とした編集合成方式による音声応答装置の原理構成を示す。

表3 装置の概略仕様 カセットユニットと磁気シートユニットの主要性能を対比して示してある。

ユニット名 項目	カセットユニット	磁気シートユニット
総情報量	最大30分/2チャンネル(C-60)	最大588秒/49トラック
情報数	最大398/2チャンネル	—
音声トラック数	—	49
アクセスタイム	平均15秒	0.5秒以下
再生頭出し精度	1秒以内	—
再生周波数特性	0.1~8kHz(偏差6dB)	0.17~5kHz(偏差6dB) 0.3~2.5kHz(偏差4dB)
再生出力レベル	0dB (0dB=0.775V)以上	+4dBm <sup>+1</sup> <sub>-3</sub> dB
信号対雑音比	45dB以上 (信号レベル:0dB)	40dB以上 (信号レベル:+4dBm)

ムにアクセスでき、画面別の音声、音楽などを収納することができる。

この複合音声ファイルは、基本的にはカセットユニット部と磁気シートユニット部から成る。カセットユニット部は最大398種類の音声を録音でき、中央処理装置の指令により平均15秒のアクセス時間で再生送出可能である。また、各音声信号の継続時間は信号ごとに任意に設定できるため、カセットテープを有効に利用できる特長をもち、安価な音声ファイル

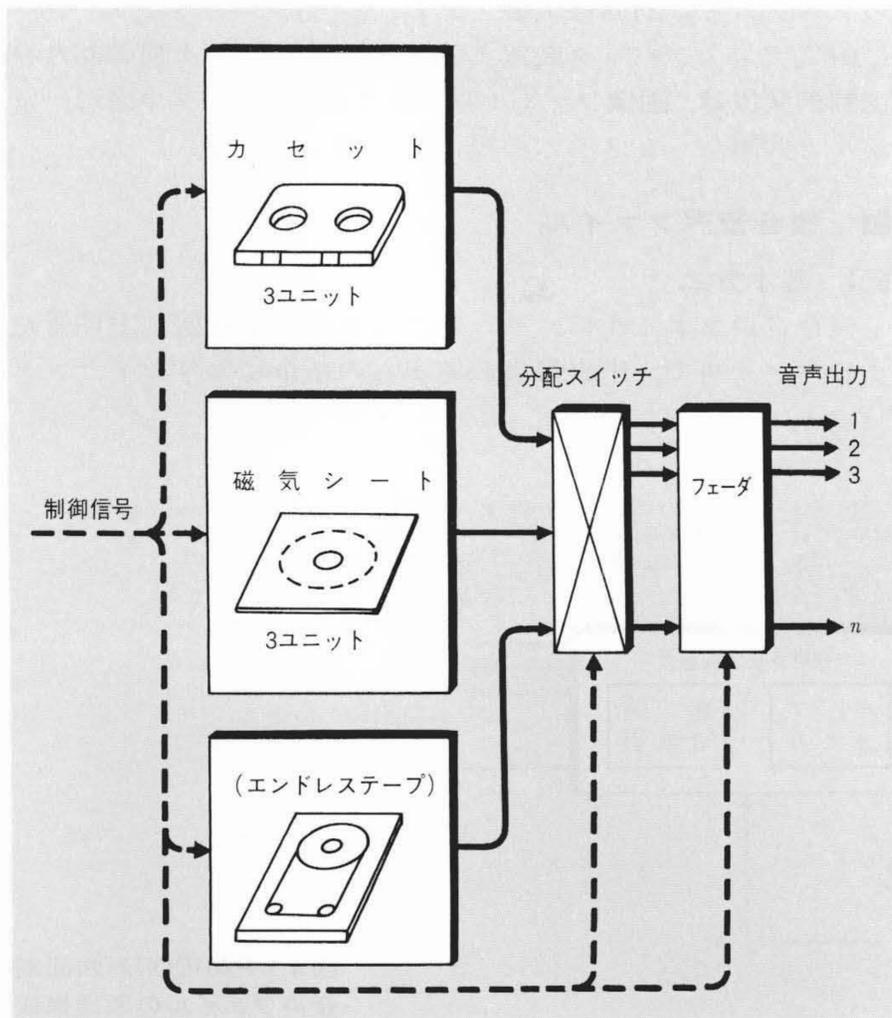


図5 複合音声ファイルの構成 各音声ファイル、分配スイッチ、及びフェーダはあらかじめプログラムされた内容に従って、CPUからタイムシリアルに制御される。

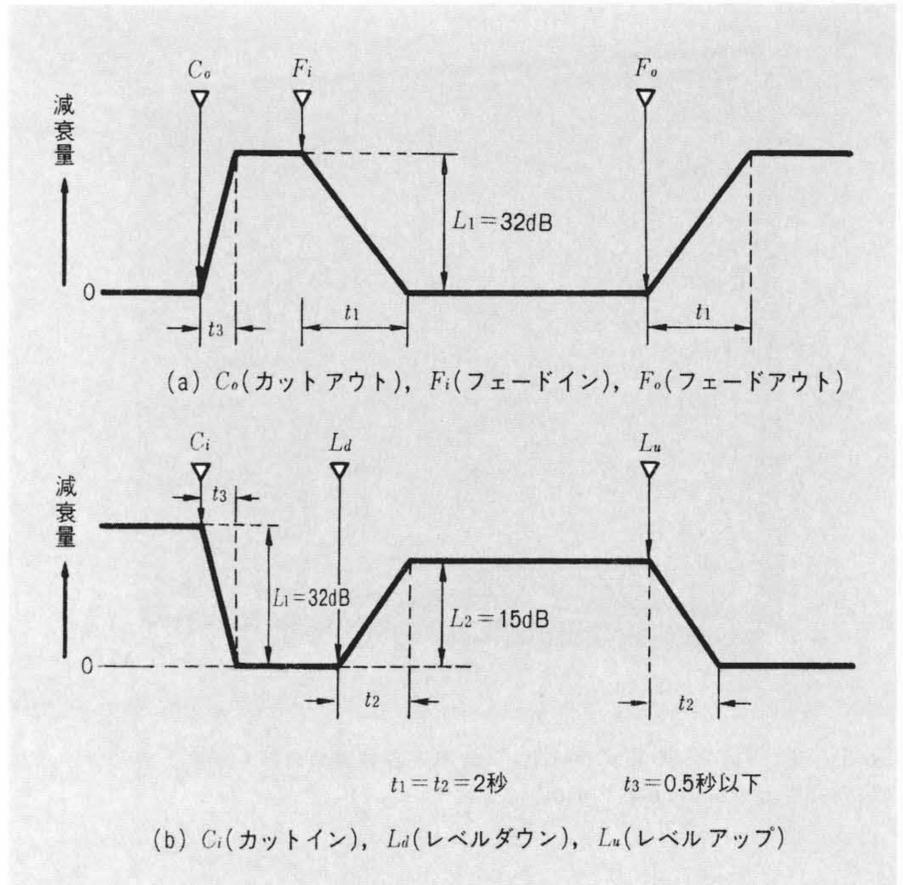


図6 フェーダの機能の概要 ▽はCPUからの指令コマンドを示す。また $t_1, t_2, L_1, L_2$ はボリュームにより可変できる。

として利用できる。

磁気シートユニット部は多数の音声トラックをもつフレキシブル磁気ディスク上に音声信号をアナログ形式のまま直接記録、再生する装置である。このファイルは、中央処理装置からの指令によりランダムアクセスが可能で、1トラック当たりの最大音声長は約12秒である。

この装置の概略仕様を表3に示す。

### 5.2 装置構成と動作概要

複合音声ファイルの構成を図5に示す。この装置はカセットデッキ部3ユニット、磁気シート部3ユニット及びエンドレステープレコーダ1ユニットが実装され、中央処理装置からの指令により動作する。

分配スイッチは、各ファイルユニットからの音声信号を回線対応に展開するためのスイッチである。

フェーダは、中央処理装置の制御のもとで音量を調整する機能をもっている。フェーダの機能の概要を図6に示す。

## 6 結 言

以上、画像応答システムの音声ファイルとして開発した音声応答装置、PARCOR方式による画面別音声ファイル、及び安価な複合音声ファイルについてその概要を述べた。これらの音声ファイル装置は、VRSのセンタ装置として順調に稼働している。

終わりに、音声ファイルの開発に対し御指導、御鞭達をいただいた関係各位に対し厚く御礼申しあげる。

### 参考文献

- 1) 高橋, ほか4名: 電話予約用音声応答システム, 日立評論, 57, 237~242(昭50-3)
- 2) 小池, ほか2名: PARCOR形音声応答装置, 電気通信研究所, 研究実用化報告Vol.23, No.10, 2107~2120(昭49-12)
- 3) 石井, ほか2名: PARCOR形画面別音声ファイル装置, 昭和53年電子通信学会全国大会, 1081(昭53-3)