

7N-5

音声エディタ (1)
— システム及び機能の概要 —

黒田明裕
日本アイ・ビー・エム株式会社

大概朋子
サイエンス・インスティテュート

1. まえがき

オフィスにおける情報流通の効率化、電話の効果的利用などの目的のために、音声パーソナル・コンピュータ(PC)上で手軽に取り扱いたいという要求が高まりつつある[1][2]。その要求を満たす一つの手段として、筆者らは、音声を文単位、発声単位等できめ細かく、また、対話的に編集することのできる音声エディタを、PC上に試作したので、そのシステム構成及び機能の概要を報告する。

2. システムの概要

本エディタは、一枚の音響処理ボードを搭載した IBM PC-AT上に実現した。ボードは Fig. 1 に示す構成をしている。メモリはPCのメモリとしてマップされており、データ・メモリは、PCと音響処理ボード間の音声データの受け渡しに使われる。アナログ・サブシステムは、A/D/D-A及びアナログ回路を含み、データ・メモリに対しDMAを行う機能も有している。入出力デバイスとしては、マイク/スピーカ、電話機、または、電話回線が接続できる。

音響処理ボードは、音声の録音、再生が必要なコマンドを、ユーザが選んだときに起動される。録音の場合には、DSPは、Fig. 2 に示したフローで、音声データ (PCM, ADPCM, または、LSP) と、音声の視覚化データ (対数パワー) を次々

と計算し、データ・メモリに蓄積する。PCは、蓄積されたデータを次々と読み出し、音声データをハード・ディスクへ、視覚化データをPCのメモリへ書き込む。再生の場合には、この逆を行う。

3. 音声の編集

本エディタは、編集の最小単位を1つの発声区間とし、録音された音声データに対して編集(COPY, DELETE等)を行ったり、新しい音声入力を受けたり(INSERT)、また、ある音声ファイルに他の音声ファイルを組み入れたり(GET)することを、対話的に行うことが出来るようになっており、例えば、次のような編集を行う。(以下の例で「」で囲まれた部分は音声を表す。)

課長「・・・社長から、次のようなメッセージが参りました。このことに関しまして、ゴホン(咳払い)、明日の朝9時から会議を行います。会場は、会議室3(会議室4の誤り)です。・・・」

社長「この計画を進めるにあたって、・・・」



「(課長)・・・社長から、次のようなメッセージが参りました。『(社長)この計画を進めるにあたって、・・・』(課長)このことに関しまして、明日の朝9時から会議を行います。会場は、会議室4です。・・・」

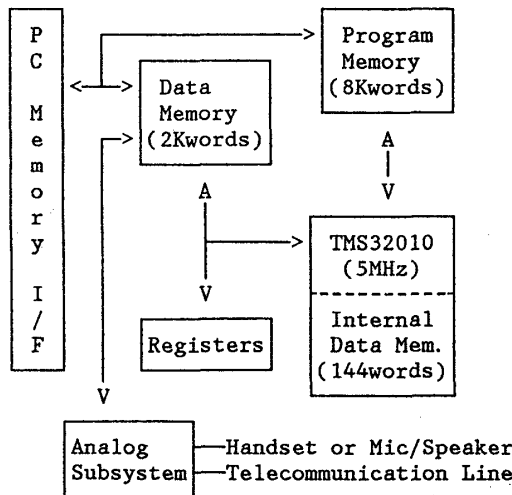


Figure 1. Speech Board - Block Diagram

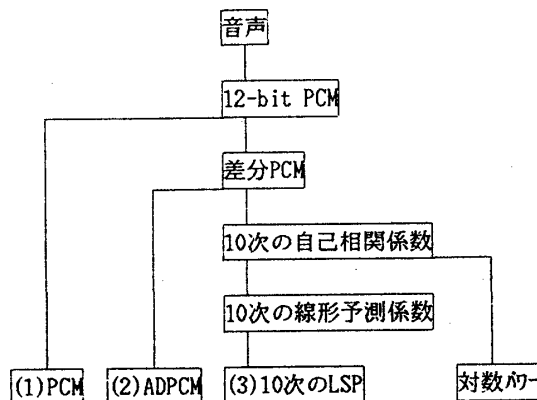


Figure 2. Speech Analysis by a DSP

音声を対話的に、きめ細かく編集するためには、音声を視覚化する必要がある。今回は一例として、対数パワー時系列を使用した。ただし、これだけでは分りにくいので、後述するようないくつかの工夫をほどこした。ユーザは、Fig. 3に示されるエディタ画面を見ながら編集、再生したい部分を指定し、アイコン・メニューを使って指示を行う。

音声エディタのコマンドを付録に示す。基本編集用のコマンドは、通常のエディタにおけるそれと同様の動作に対応する。例えば、マウスで1点指定した後INSERTに対応するアイコンを選べば、新しく入力した音声はその場所に挿入される。また、3点指定した後（最初の2点はソース・ブロックの指定、最後の1点はデスティネーションの指定）COPYに対応するアイコンを選べば、ソース・ブロックが目的の場所にコピーされる。その他、入出力デバイス、データ・タイプを選択（変更）するためのコマンドや、音声データ編集のための特殊なコマンドもいくつか考えられる。

4. ユーザ・インターフェースに関する考察

1) 音声の視覚化及びポインティング 視覚化音声データとして、今回は、対数パワー時系列を使用した。しかしながら、対数パワー時系列から得られる情報だけでは、効率的にポインティングを行うことは難しい。特に、長いメッセージの場合、目的とするブロックを見つけだすのに非常に手間取る。この問題を解決するための手段として、次のような方法が考えられる。

- ・高速プレイバック [3] (FPLAYBK コマンド) を行ないながら、文章の区切、重要な部分などに注釈を入れ (ANOTATE コマンド)、次回からの編集に利用する。注釈の内容としては、文章の要旨、最初の文節などが考えられる。また、プレイバックしている部分をカーソルでトレースするのも有効である。
- ・キー・ワードを発声し、その発声に近いものをワード・スポッティングにより、いくつか拾い上げる (FIND コマ

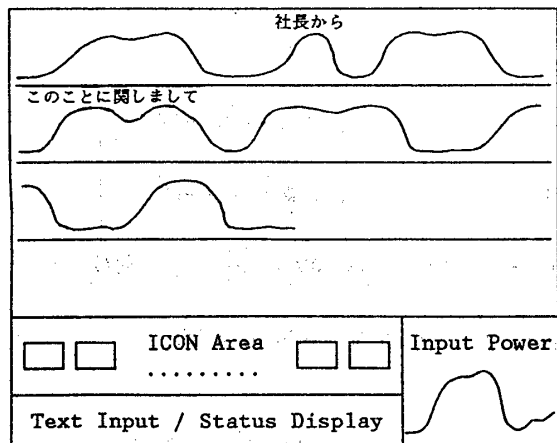


Figure 3. Editor Screen

ンド)。ただし、この方法は、発声の揺らぎの大きい連続発声に対して、安定して適用するのは難しい。

2) 編集による不自然さの解消

- ・継目において、音が連続的に聞こえる必要がある。ところで、前述したように、本エディタにおける編集の最小単位は「1つの発声区間」としている。この方式によれば、編集の継目はポーズ（無音）の部分に限定される。無音部どうしを接続した場合、継目における音の不連続さは殆ど聞き取れない。
- ・編集したブロックのパワーが、その前後のブロックのパワーと著しく異なると、不自然に聞こえる。PWRNORM (Power Normalize) コマンドは、指定されたブロックの平均パワーがその前後のブロックの平均パワーと等しくなるように、音声データに変更を加える。
- 3) 高速処理 音声データ、特に、音質の良いPCMデータなどは、テキスト・データなどと比べてデータ量が膨大であり、音声エディタは、大量のデータを高速に処理する必要がある。本エディタは、そのための工夫がなされており、このことに関しては、他で詳述する [4]。

5. まとめ

PC上に音声エディタを試作した。基本的な編集機能を実現して、種々の実験を行い、考察を加えた。今後は、これらの点をベースに改良を行っていく予定である。また、複合文書エディタに組み入れることも、検討していきたい。

[参考文献]

- [1] R.T.Nicholson, "Usage Patterns in an Integrated Voice and Data ...," ACM Trans. on Office Information Systems, Vol.3, No.3, July 1985.
- [2] 池田泰久, "コンピュータ電話機," 日本音響学会誌 41巻 11号 (1985).
- [3] N.F.Maxemchuk, "An Experimental Speech Storage and Editing Facility," The Bell System Tech. Journal, Vol.59, No.8, October 1980.
- [4] 大槻, 黒田, "音声エディタ (2) - メモリ及びファイル管理方式 -," 第33回情報処理学会全国大会.

付録. 音声エディタのコマンド (*: 開発中)

基本編集	
INSERT/PLAYBK	: 音声の入出力
DELETE/COPY/MOVE/CHANGE	: 音声の編集
*GET/*EDIT	: 複数音声ファイルの編集
FILE/QUIT	: 編集の終了
デバイス/データ・タイプの選択	
MICSPK/HANDSET	: デバイスの選択
PCM/*ADPCM/*LSP	: データ・タイプの選択
音声ユニークな編集	
*FIND/*FPLAYBK/*PWRNORM/*ANNOTATE	(本文参照)