

音声データファイル作成装置の検討

1R-8

土田尚純

代 啓一

山東慶三

(NTT 電気通信研究所)

1. まえがき：網のデジタル化や民生機器の発展により、公衆網から各種通知案内、操作ガイダンス等を音声で出力する音声応答装置（ARE）に対する音声の高品質化の強い要望と、メモリの低価格化を背景に、従来のLSP方式に加え、μlawPCM方式を用いたAREを開発した。そこで、このAREに品質の良い音声データファイルを提供するための、マンマシンインタフェースの優れた音声データファイル作成装置を検討、試作したので、ここに報告する。

2. 基本設計方針：音声データファイル作成装置は、以下の方針のもとで、検討を進めた。

- (1) 機能の充実：公衆網に適用されている各種ARE用の音声データファイルを作成可能とするとともに、音声データの資産継承を考慮して、音声データの修正、管理機能を具備することとする。
- (2) 操作性の向上：視聴覚的に音声データの直接認知を可能とし、かつ、ワープロのように誰でもが知っているマンマシンインタフェースを採用して、音声データの切出し、編集及び管理に対する操作性の向上を図る。
- (3) 装置の小型化、経済化：制御部には、汎用パソコンを採用して、装置の小型化、経済化を図る。一方、処理速度の向上を図るため符号変換はハードで実現することとする。

3. 装置構成法

3.1 直線PCMでの音声管理

図1に作成処理手順の概要を示す。ここで、各種AREに対応可能とし、かつ、音声データの資産継承を可能とするために、A/D変換後の直線PCMのまま音片の切出し、編集および管理等を行い、符号変換は、媒体(FD)への出力直前とした。

3.2 ビジュアルな音片切出し(図2参照)

ファイル作成上のポイントとなる音片の切出しでは、通常の応答文の長さを考慮して、1画面に約30秒分のパワー波形を表示可能とするとともに、語頭、語尾等の厳密な切出しも可能とするために拡大表示(約3秒)も用意した。また、切出し候補点の設定用にカーソル、マーカの活用、切出した音片のモニタ機能により、視聴覚に訴えた切出し方法を実現した。

3.3 処理待ち時間の短縮、集約化

切出した音片の登録、符号変換、モニタ等は、ハードディスク間或いは、制御部-符号化部間の大量音声データ転送を伴い、その処理時間が長くなる。そこで、入出力制御プログラムのアセンブラ化、ランダムファイルアクセスの実現等による処理時間の短縮、或いは、符号変換とFD出力の待ち時間集約化、イラスト画面の

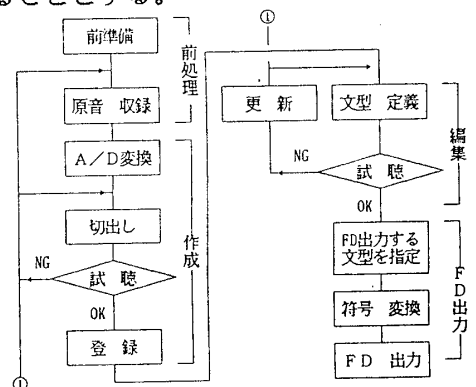


図1 作成処理手順

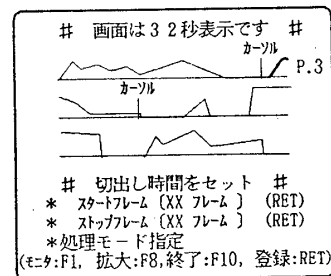


図2 画面イメージ例

挿入による待ち時間表示等を行い、オペレータの待ち時間負荷を極力排した処理フローを実現した。

3. 4 シミュレーションモニタ

AREから実際に出力する音声をシミュレーションすることにより、応答文を構成する定型文、可変語のつながりの間合い、音声レベル等を評価可能とするとともに、その修正を容易に実施可能として、音声品質向上のための機能の充実を図った。

4. 試作装置概要

4. 1 ハードウェア

装置構成を図3に、諸元を表1に示す。符号化部は、各種符号化方式をパッケージ差換えて実現可能として、汎用パソコンのCPU筐体とほぼ同程度の小型化を達成した。

4. 2 ソフトウェア

メニュー方式及びファンクションキの活用により、操作の明確化、簡略化を実現した。プログラムはMS-DOS上で動作し、主言語はPC-FORTRAN、規模は入出力制御部のアセンブラを含めて約20kstepである。

5. 評価：PCMのような波形符号化方式では、分析合成方式

(LSP等)の場合に避け得ない分析エラーの修正を必要としないため、音片の切出し及び音片間の間合い、レベル差の補正等が処理時間の主体となっている。ファイル作成時の各処理時間例を表2に示す。ファイル作成に要するトータル処理時間は、作成装置による処理、オペレータによる操作、机上処理を含めて、実音声の約100倍程度に抑えることができた。これは、システムに依存しない前処理時間を除けば、従来の当研究室のミニコンベースによるシステムの約半分の処理時間で済み、本装置の有用性が確認できた。

尚、操作性の向上を図るため、被験者5名による評価実験を行い、そのフィードバックを実施するとともに、実験結果より、波形符号化方式の音声データファイル作成では、初心者、熟練者の差異がそれ程顕著に表れないことも確認できた。

6. むすび：現在、本装置を使用して、続々と音声データファイルが作成されており、自動クレジットサービス等種々のサービスに適用されている。最後に本検討に際して貴重な御意見を頂いた知能処理方式研究室坂井室長、同橋田主幹研究員に深く感謝致します。更に、本装置のベースとなった小形音声ファイル作成装置<sup>(1)</sup>開発を担当された音声入出力方式研究室の関係各位の御協力に深く感謝致します。

〔参考文献〕(1)印藤他：小形音声ファイル作成装置の検討 音学会 S.60.3 2-5-17

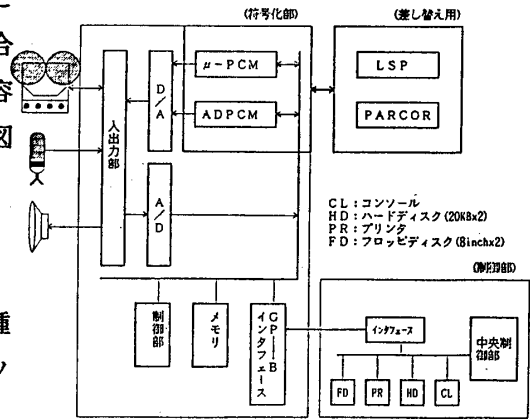


図3 装置構成

表1 諸元

項目	諸元	
制御部	機種名	FC-9801
	語彙容量	最大40MB (HD: 20MB×2)
	定型語長	最大20MB
	符号化部とのインタフェース	GP-1B
符号化部	出力媒体	8インチ2D標準フロッピー
	制御部	8ビットマイクロプロセッサ
	符号化方式	μlawPCM (6.4Kbps) ADPCM (3.2Kbps) LSP, PARCOR
A/D, D/A	12ビット8KHzサンプリング	

表2 各処理時間例  
(実音声時間で正規化)

項目	本装置	従来装置
前処理	発声原稿作成	50
	原音収録etc	50
作成	A/D変換	1
	切出し登録 etc	30
編集	文型定義*1	10
	リスト出力確認	5
FD出力	符号変換	8
	出力文型指定FD書込み	5
その他操作時間	2	7
トータル時間	111	152

\*1: 試聴確認を含む。但し、更新は無いものとする。