

パーソナルコンピュータを用いた 音声信号分析装置について

中西宏文

(情報科学選修)

Analysis System of the Speech Signal using Personal Computer

Hirobumi NAKANISHI

(Department of Educational Science)

1. はじめに

近年のコンピュータ技術の進歩は目を見張るべきものがあり, 我々の生活のごく身近なところにもかなりの応用製品が入り込んでいる。今や, 家電製品とコンピュータ技術とは切っても切れない関係にあるといっても過言ではなからう。これらの製品には, コンピュータの頭脳となるマイクロプロセッサが目に見えない形で巧みに使われている。

また, 目に見える形でのコンピュータも半導体技術の進歩と相まって急速に身近なものとなった。いわゆるパソコン・マイコンといった類のコンピュータの登場である。これらの先駆けとなったのが1980年頃登場したNEC PC-8000シリーズと呼ばれる8 bitのCPUを内蔵した本体にキーボードが付属しディスプレイを接続するだけで利用できるタイプのコンピュータである。その後他社からも同様の製品が販売され, 機能的にも8 bitから16bitへそして今日では32bitと処理能力も増大し, より高速に多量のデータを処理できるようになってきている。

今回は, このパーソナルコンピュータを用いて, 人間の音声进行分析するための装置を製作したので, それに関して報告する。

2. 特徴

今回製作した音声分析装置は, 市販のパーソナルコンピュータにA/D変換ボードを組み合わせ, C言語で書かれたプログラムでA/D変換器を制御し音声信号を入力し特徴分析を行うものである。この装置の特徴を以下にまとめる。

- ・高価な専用分析装置を使わずに音声信号等の入力・分析が行える。
- ・DMA転送により実時間で信号の入力が行える。
- ・入力波形をディスプレイ上に表示でき, 波形の確認ができる。
- ・希望する任意の特徴量がプログラミングにより得られる。
- ・A/D変換器からの信号入力に必要な操作を全て1つの関数呼出しにより行うことが

できる。

3. A/D変換器の概要

今回の分析装置に用いたA/D変換器はカノープス電子社製の SOUND MASTER と呼ばれる比較的安価な個人レベルでも利用可能な製品のハードウェア部分を用いたもので、最近普及の著しいCDプレーヤーと同一の環境で音の入出力の行えるものである。仕様を以下に記す。

・入力電圧	- 3 ~ + 3 V
・出力コード	2'S コンプリメントバイナリ
・周波数特性	5 Hz ~ 20KHz
・歪率	0.03%
・サンプリング周波数	32, 44.1, 48KHz 切り替え可
・フィルタ	20KHz 7次ローパスフィルタ

4. DMA転送機能

4.1 DMA転送機能の概要

A/D変換器においてアナログ信号をデジタル信号に変換した結果をPC9801内部のメインメモリにデータ転送するためにパーソナルコンピュータのDMA転送機能を使用した。DMA (Direct Memory Access) 機能は、CPUのシステムデータバスへのアクセスを一時中断し、バスの使用权をPC9801内部のDMAコントローラ8237Aが使用することにより実行される。A/D変換を終了してA/D変換ボードからPC9801へデータを転送したいとき8237AにDMA要求信号を出すと、8237Aはこれを受け付けて、CPUのシステムバスへのアクセスを止め、A/D変換ボードから、メインメモリへCPUを介さず直接データを転送する。

このような動作を行うためには、A/D変換ボードと8237Aの両方に必要なデータやコマンドを書き込んで、データの転送バイト数や転送先のアドレスなどを指定する必要がある。

4.2 DMAコントローラ8237Aの設定の概要

A/D変換ボードのDMA転送機能を利用するためには、PC9801内部のDMAコントローラに以下のデータ・コマンドを書き込む必要がある。

・モード

DMAコントローラ8237Aには、4つのモード (ダイヤモンドトランスファモード、シングルトランスファモード、ブロックトランスファモード、カスケードモード)があり、それぞれ用途に応じて指定するが、今回使用したA/D変換ボードではシングルトランスファモードを使用する。

・データアドレス

データを書き込む際に先頭アドレスからどちらの方向にデータを書き込んでいくか、つまりアドレスをインクリメントするかデクリメントするかを指定する。これは通常インクリメントを指定することにより、転送されるデータをアドレスが小さな方から大きな

方に順次書かれていくことになる。

- 転送方向

I/Oポートからメモリへ転送（ライト転送）するのか、メモリからI/Oポートへ出力（リード転送）するのかを指定するもので、ここではI/Oポートに差し込んだA/D変換ボードから本体のメモリへ転送するのでライト転送を指定する。

- DMA ベースアドレス

A/D変換された結果をDMA転送する際のメモリの先頭アドレスを指定する必要がある。DMAコントローラは、この番地から番地を自動的に更新しながらA/D変換結果を転送する。但し、この番地を指定する際に、PC9801内部のメモリアドレスは20ビットで表現されるのに対し、8237Aでは16ビットしか制御できないため、下位16ビットを8237Aに書き込み、残りの上位4ビットはPC9801の内部レジスタに書き込む形式をとる。

- 転送バイト数

連続してDMA転送したいA/D変換結果を指定する必要がある。1チャンネル・1回の変換に必要な転送データは2バイトなので、その割合で計算する。設定可能な最大値は64Kバイトに制限されている。

- チャンネル・マスク

上記の各種パラメータを8237Aに書き込む場合、それぞれ該当チャンネルをマスクするコマンドをあらかじめ書き込んで、そのチャンネルのDMA転送を禁止してから行なう必要がある。そして、パラメータをすべて設定した後、チャンネルのマスクをリセットし、DMA転送を行なう。

- ファースト・ラスト・フィリップフロップ

DMAベースアドレスや転送バイト数は16ビットで指定する必要があるが8237Aは一度に8ビットしか書き込むことができない。そこで下位8ビットと上位8ビットを分けて書き込む必要がある。これは、同じI/Oポートに続けて2バイト書き込むと、初めのものが下位、次が上位と判定されて書き込まれる。この順序を決めるのがファースト・ラスト・フィリップフロップで、まずこれをクリアしてから、2バイトのデータを書き込むようにする。

4. 3 DMA転送の制約事項とその解決法

4. 2節で述べた方法で設定しDMA転送を開始すると、ベースアドレスから順にデータが書き込まれていき、設定した転送バイト数だけ転送を行うと、転送が終了する。その間、書き込まれるメモリのアドレスは自動的にインクリメントされていくが、下位16ビットがFFFFHになると、次はOOOOHにもどって上位4ビットのDMAバンクが1だけ繰り上がることが望ましい。この機能をバンクオートインクリメントと呼び、最近の機種ではこの機能を備えているが旧型ではこの機能が無いため、今回は互換性を持たせるため、この機能は使用しない。このため、物理アドレスの下位16ビットがFFFFHからOOOOHに変わる境界（セグメント境界）をまたいでDMA転送する事はできない。また、一回当たりの転送バイト数の制約から、一回当たり32768回の変換が最大値となる。このことから、セグメント境界の問題をクリアし、なおかつ最大長の変換をするためには、ペー

アドレスがちょうどセグメント境界のすぐ上に来るように配列が確保されればよい。このために、プログラム中では一旦仮に短い配列を確保し、その配列の先頭要素のベースアドレスを調べ、すぐ上のセグメント境界までの領域を消費してしまい、新たにDMA転送用の配列を確保すれば最大長のデータ転送が可能となる。

4. 4 I/Oポートの制御

I/Oポートとは、PC9801背面の拡張スロットに差し込んだ各種ボードを制御したり、データのやり取りを行うためのもので、ボードを設計する際にどのI/Oポートをどのように使用するか決める必要がある。また、各種ボード間で同じ番号のI/Oポートを使用すると誤動作の原因となるので注意する必要がある。以下に、今回使用したA/D変換ボードで使用したI/Oポートとその指定内容について述べる。

- ライトポートCONT (I/Oアドレス+0)
 - DMAビット—DMA転送を行う場合、1をセットする。
 - OSCビット—内部クロックを動作させる時に1をセットする。
0をセットすれば停止させることができる。
- ライトポートDMFQ (I/Oアドレス+4)
 - CK1, CK2ビット—4種類のサンプリングクロックの選択
 - DMACHビット—DMAチャンネルの選択をする
- ライトポートMODE (I/Oアドレス+7)
 - INITビット—このビットに1をセットすれば、全てのポートがリセットされる。ポートモードもリセットされる。
 - L, Rビット—使用したいチャンネルに1をセットする。L, R両チャンネル使用時は、それぞれ1をセットする。
- リードポートROVRN (I/Oアドレス+0)
 - DMAビット—DMA転送が行われている間、1がセットされる。DMAコントローラに指定しておいたバイト数だけ転送されると0にリセットされる。
 - OSCビット—クロックが動作しているとき1がセットされ、停止するとリセットされ0になる。

5. 制御の流れ

4章で述べたDMAコントローラを実際に制御して、A/D変換器でA/D変換されたデータをパーソナルコンピュータ本体のメモリ上に取り込む具体的な手順を以下に述べる。

- モードポートにデータ80Hを書き込み、全てのI/Oポートをリセットする。
- A/D変換された音声信号をDMA転送するための配列をメモリ上に確保する。
- 確保した配列のベースアドレスと転送するバイト数を算出する。
- I/Oポート15にデータ07Hを書き込み、DMAチャンネルをマスクする。
- I/Oポート17にデータ17Hを書き込み、DMAチャンネル3とシングルトランスファモードを選択する。
- I/Oポート19にデータ00Hを書き込み、ファースト・ラスト・フィリップフロッ

ブをクリアする。

- I/Oポート 0 D にデータベースアドレスを、0 F に転送バイト数をそれぞれ下位バイト、上位バイトの順に指定する。
- DMFQ ポートに 1 8 H を書き込み、44.1KHz のサンプリングクロックを選択。
- MODE ポートに 0 1 H を書き込み、L チャンネルを選択する。
- I/Oポート 1 5 に 0 3 H を書き込み、DMA マスクを解除する。
- CONT ポートに 0 7 H を書き込み、サンプリングクロックをスタートさせ音声の取り込みを開始する。(A/D変換開始)
- ROVRN ポートを読み、DMA が完了したら指定したバイト数だけ取り込んだかどうか確認する。

6. 実験環境と問題点

第 5 章で述べた手順に従って A/D 変換を行ない、実際の音声信号を入力できるように C 言語を用いてプログラミングを行った。実際に使用したハードウェア環境を下記に示す。

- パーソナルコンピュータ本体

日本電気製 P C 9801 R A 2 (C P U 80386, クロック 16MHz)

〃 P C - H D 040 (ハードディスク, 容量 40MB)

- 使用 OS MS-DOS Ver 3.3B

- 言語処理系

Borland International 社製

Turbo C Ver 2.0 統合環境版

このような環境下でプログラミングを行なったところ、安定した動作で A/D 変換が行えた。但し、実際に A/D 変換装置を他の処理 (例えば音声認識等) と組み合わせて使おうとした場合、以下に述べるような問題点が生じる。

- (1) 使用できる取り込み時間が短い

ボードに内蔵されているローパスフィルタがカットオフ周波数 20KHz であるため、内蔵のクロック周波数としては 44.1KHz を使用する必要がある (サンプリング定理による)、この場合、4.2 の転送バイト数で述べたことから 1 回の A/D 変換で約 0.75 秒分の音声信号しか変換できず、これは 4 音節程度の単語 1 単語分なので実用的には多少不十分である。これを解決するには変換を繰り返し行なうという方法もあるが、1 回の変換当たり、64KB のメモリを消費するので MS-DOS で使える全メモリ量の制約 (約 1MB, 内ユーザ領域 640KB) から考えてもあまり良い解決方法とは言えない。この解決方法としては人間の音声に含まれる周波数成分は 10Hz ~ 6KHz 程度までなのでボードに信号を入力する前に、カットオフ周波数 6KHz 程度のローパスフィルタを外付けし 12KHz 程度の外部クロックを用いてサンプリングを行う事により、約 2.7 秒程度の入力が可能となり、各種実験を行うに十分な取り込み時間が得られる。

- (2) 使用可能なメモリ量の絶対的な不足

今回使用している OS (オペレーションシステム) である MS-DOS は、本体の RA

M容量の内640KBしか用いる事ができない。これはOSの設計思想に由来するもので特殊な手法を使わない限り拡張する事はできず、現在のところその手法を利用できる言語処理系は市販されていない。さらにOS自体が約90KB消費し、Turbo Cの統合環境版が330KB消費するため、ユーザが実際に使える領域は220KBしかなく、これをプログラムのソースコードや作業領域、プログラム中の変数・配列領域で使用するため、メモリの使用環境はかなり厳しく、実際に使用してみるとA/D変換のみの場合、128KB程度までのデータは何とか入力できるものの、入力したデータをFFT等で分析しようとするとその時点でメモリ不足になってしまい、統合環境版で実際に利用できる入力時間はサンプリングクロック44.1KHzの場合、およそ0.3秒程度であった。これは統合環境版を使用した場合、プログラムの作成するコンパイル・実行までを一貫した環境の下でスムーズに行えるように設計されているため、エディタ・コンパイラ・リンカが常にメインメモリ上に常駐する形になるためである。よって取り込み時間を長くしたい場合、一旦コンパイルの終了した時点で統合環境版を終了し、できあがっているロードモジュールをMS-DOSのコマンドレベルで直接実行してやる事により、統合環境版により占有されていた330KBもプログラムの実行中にユーザが利用可能となる。但しこの場合、統合環境版の操作性の良さが多少失われる。または、コマンドライン版コンパイラを使用しても同様に、実行中はコンパイラ等の占有領域無しにユーザ領域が使用可能となる。

7. 信号解析法

今回作成したA/D変換プログラムによってメモリ上に取り込まれた音声信号を実際に種々の分野に応用しようとした場合、原波形レベルをそのまま用いることは少なく、多くの場合何らかの形で特徴量を抽出し、それをを用いて処理を行う場合がほとんどである。その特徴量の代表的なものを挙げると、

- ・FFT（高速フーリエ変換）スペクトラム
- ・FFT（高速フーリエ変換）ケプストラム
- ・LPC（線形予測係数）スペクトラム
- ・LPC（線形予測係数）ケプストラム
- ・PARCOR（偏自己相関）係数

等があるが、ここではパーソナルコンピュータを用いているため、信号を実時間に近いレベルで処理しようとした場合、より計算量の少ない方法が有利となる。

そこで今回はFFTスペクトラムによる分析処理を行うプログラムを作成し、A/D変換されたデータは周波数分析されたパラメータとして用いることとした。FFTについては多くの著書で紹介されているので詳細は省略するが、通常フーリエ変換を行なうには膨大な計算をこなさなくてはならないのに対し変換するソースデータ数が2のべき乗になっている場合に限り、FFTアルゴリズムを用いると計算量を大幅に減らすことができるのである。

8. 応用例

第7章までに述べてきた、A/D変換部、特徴分析部を実際の実用的な音声認識処理装置の一部として使用してきた応用例をいくつか紹介する。

8. 1 特定話者孤立発声数字音声認識装置

この装置は、特定の個人があらかじめ認識したい数字（その他の単語も可）を学習時に登録しておき、認識時にはある数字を発声するとあらかじめ登録されたものとの間で照合を行ない、最も照合結果の類似しているものを認識結果とするものである。

但し、この場合同一人物の音声とはいえ発声の継続時間長は学習時と認識時ではずれが生じる。このため発声時間の違いが照合に悪影響を与える事態が生じる。この点を改善するために、2つのパターン間の時間的ずれに対処する方法としてよく知られているのが動的計画法（Dynamic Programing）と呼ばれる手法で、詳細は省略するがこの手法を用いた音声認識アルゴリズムを作成した。ただし、この手法も膨大な計算量と記憶量を必要とするため、今回のパーソナルコンピュータでは実時間処理はできず、認識結果が得られるまでに数秒のタイムラグが生じた。

8. 2 特定話者都市名音声認識装置

この装置は、特定の個人があらかじめ認識したい内容とは無関係に、日本語に含まれる全ての音節を孤立発声しておき、認識時には認識したい単語名をキーボードからローマ字入力してやることにより、それにしたがってあらかじめ登録された音節を順に連結したものと入力パターンとを照合する方法である。この際調音結合等の問題が生じるが今回は応用例としてのデモシステム作成が目的であり、認識精度の向上は別の研究に譲るものとする。このシステムでは画面に日本地図が描かれ、地名を発声しそれが認識されると該当地点が地図上で点滅するようにしたもので、音声認識の仕組みを一般の人に理解してもらうには適当なものと思われる。

9. ま と め

パーソナルコンピュータのより高度で実際的な使用例として、音声信号分析装置ならびにそれを利用した応用例として、特定話者音声認識装置について述べた。

現在、パーソナルコンピュータの販売台数は飛躍的な伸びをみせ非常に身近なものになりつつあるが、その利用形態をみるとゲーム専用に使われていたり、市販アプリケーションソフトの利用、といった初歩的な使われ方がほとんどで、その上のレベルの利用としては科学技術計算等に用いられる場合もあるが、拡張スロットを使ったとしてもハードディスクの接続、メモリボードの装着といった程度の利用までであり、拡張スロットを使用して各種機器の制御を行おうといった応用例はごく稀である。しかしながら今日、各種インターフェイスボード等も安価に性能の良いものが供給されるようになり個人レベルでも十分に手の届くところまできている現状を考えた場合、パーソナルコンピュータにより各種周辺機器を制御し、パーソナルコンピュータを単なる計算の道具としてだけではなく、より高度な使い方をすることが個人レベルでも十分可能な段階にきている。しかしながら、そのための技術に関してはほとんど知られておらず、パーソナルコンピュータの可能性のごく一部しか利用されていないのが現状である。このため、今回はパーソナルコンピュータと市販のA/D変換ボードを使っての音声認識装置を実現するための第一歩としての、拡張I/Oポート（スロット）の使用例、データの受け渡し方法等に関し各種実験を行ない、得られた結果、必要と思われる技術を紹介し問題点の検討を行ない、簡単な応用例を

紹介した。

今後、多くの人々がパーソナルコンピュータの能力を十分に引き出すことができるようになるための一助となれば幸いである。

なお今回の実験の一部を、平成2年3月に本学情報科学コースの第一回生として卒業された高津眞男君・小林順子さんの二人にお手伝いいただいた。ここに感謝します。

(平成2年8月15日受理)

参 考 文 献

- 1) 坂井利之著：「情報処理学詳説」 コロナ社。
- 2) 伊福部達：「音声タイプライタの設計」 CQ出版。
- 3) 安居院・中嶋：「コンピュータ音声処理」 産報出版。
- 4) 中田・三瓶：「音声デバイス活用の実際」 オーム社。