大阪公立大学
Osaka Metropolitan University
学術情報リポジトリ

マイクロコンピュータを用いた音声認識装置 (2)

メタデータ	言語: jpn 出版者: 公開日: 2013-11-12 キーワード (Ja): キーワード (En): 作成者: 高橋, 文彦, 黒田, 芳郎, 古田, 守 メールアドレス: 所属:
URL	https://doi.org/10.24729/00007992

マイクロコンピュータを用いた音声認識装置（Ⅱ）

A Speech Recognition Device with Microcomputer（Ⅱ）

高橋文彦* 黒田芳郎* 古田　守**
Fumihiko Takahashi* Yoshiro Kuroda* Mamoru Furuta**

（昭和59年4月18日　受理）

あらまし

本研究は前報に引続いてマイクロコンピュータによる音声認識装置の試作と，その実験結果について報告したものである。マイクロコンピュータは，NEC PC-9801 を使用，16チャンネルのフィルターバンクからの音声スペクトルパターンを特徴パラメーターとしてシステムを作成した。認識実験は話者限定で，母音（5母音）と単語（10数字）について行ったが，ともにほぼ 100%の認識率を得た。認識所要時間は，母音で0.5〜0.9秒，単語で約 11秒であった。この時間は，使用ソフト（BASIC）を，数値演算協調プロセッサ 8087 を完全サポートするよう手直しすれば，大巾に改善が期待できる。また，フィルターバンク部には，スイッチト・キャパシタ・フィルタ用 LSI を利用したが，実用上，安定性には問題がないことがわかった。

1. まえがき

最近のマイクロコンピュータの発達にはめざましいものがある。当初，科学技術計算用として生産されていた頃は，価格も高く用途も限られていたが，これが計測・制御用へ，さらに，OA 機器へと用途が急速に拡大するにつれて，量産効果により価格も大巾に低下してきた。このように本体の価格が低下してくると，相対的にコンピュータへのデータ入力コストの問題が浮上してくる。OA 機器分野では，データ量が多いため，とくにその要望が強い。

音声入力は，情報入力速度において，キーボードの3倍，手書きの8倍と言われており，そのうえ，タイプライターのような特別の訓練も必要とせず，精神的な疲労も少く，ごく自然に入力できるなどの大きな利点があり，このため音声認識の研究は，各方面で進められている。

この研究は，前報のマイクロコンピュータによる音声認識の研究が PARCOR 方式（ソフトウエアによる）であったため，速度に難点があったので，これをハードウエアによるフィルターバンク方式に変更し，さらにコンピュータも16ビット形の PC-9801（NEC）に変更して音声認識の実験を行ったものである。

2. ハードウエアの構成と動作

この装置の主な構成は，16ビットマイクロコンピュータ PC 9801（NEC）を主体にし，これに音声入力用マイクアンプ，16チャンネルのフィルターバンク（三洋・音声認識ボードの前

* 電気工学科（Department of Electrical Engineering）
** 電気通信大学通信工学科(Dep. of Communication Engineering, Electrical Communication Univ,)

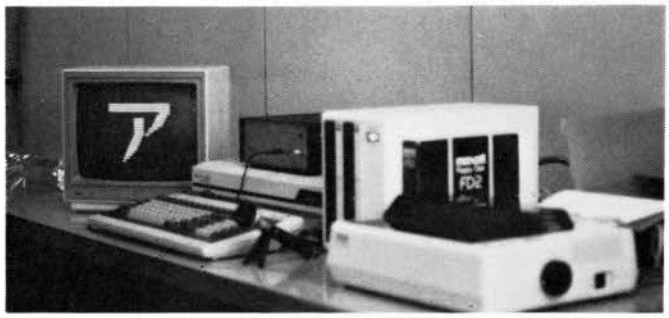


図1　音声認識装置の外観

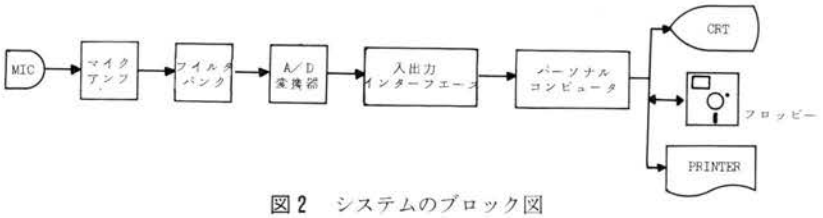


図2　システムのブロック図

表1　構成機器の性能

パーソナルコンピュータ NEC PC-9801	
CPU RAM FDD 入出力製置	μPD8086(5MHz)+PC-9806(8087協調プロセッサー) 128Kbyte（別に V-RAM 104 Kbyte 実装） 両面倍密 1MB×2 キーボード／プリンタ(エプソン RP-80F/T) ディスプレイ（14型高解像度カラーディスプレイ）
A/D インターフェース コンテック AD12-16(98)	
変換方式	逐次比較方式　分解能 12bit 変換精度±1LSB　変換時間 50［μs/CH］ 入力点数 16点
TTL レベル入出力インターフェース コンテック PIO 16-16T(98)	
入力部形式 出力部形式	非絶縁 TTL レベル入力 入力点数 16点（2点は割込み使用可能） 使用した入力点数2点（1点は割込み使用） 非絶縁オープンコレクタ 出力点数 16点（うち1点使用）
音声認識ボード〔音声分析部（フィルターバンク）のみ使用〕 三洋電機 SRB-64	
マイクアンプ 分析方法	AGC 回路，プリエンファシス回路を含む 16チャンネルバンドパスフィルター (ASA-16使用)
サンプルホールド回路 クロックパルスジェネレータ及びその他周辺回路	

段を利用），さらにA/Dコンバータを接続したものである。装置全体の写真を**図1**に，また構成のブロック図を**図2**に示す。**表1**は，これら構成機器の性能・仕様を示す。

マイクからの音声信号は，プリアンプ回路，AGC回路よりなるマイクアンプ部で前処理され，16チャンネルのバンドパスフィルターに送られる。このフィルターは，スイッチト・キャパシタ・フィルタ用 LSI・ASA-16（IEC 社）を利用したもので，これを音声帯域に合わせて 200Hz～7kHz の帯域を 16チャンネルに分割している。本システムでは，これに追加して制御用として原音声もとり込んだため，合計17入力となったので，認識に影響の少ない最高周波数の部分を1チャンネル除外した。これは，A/D 変換ボードの入力端子数（16）に合わせたからである。したがって A/D 変換器の各チャンネルに対するフィルターバンクの周波数帯域は**表2**のようになる。

表2　フィルターバンクの特性

チャンネル	中心周波数 (Hz)	帯域幅 (Hz)	チャンネル	中心周波数 (Hz)	帯域幅 (Hz)
1	——(直結)——		9	1220	180
2	260	130	10	1400	200
3	390	130	11	1600	220
4	520	130	12	1820	250
5	650	130	13	2070	300
6	780	130	14	2370	340
7	910	140	15	3035	1030
8	1060	160	16	4272	1445

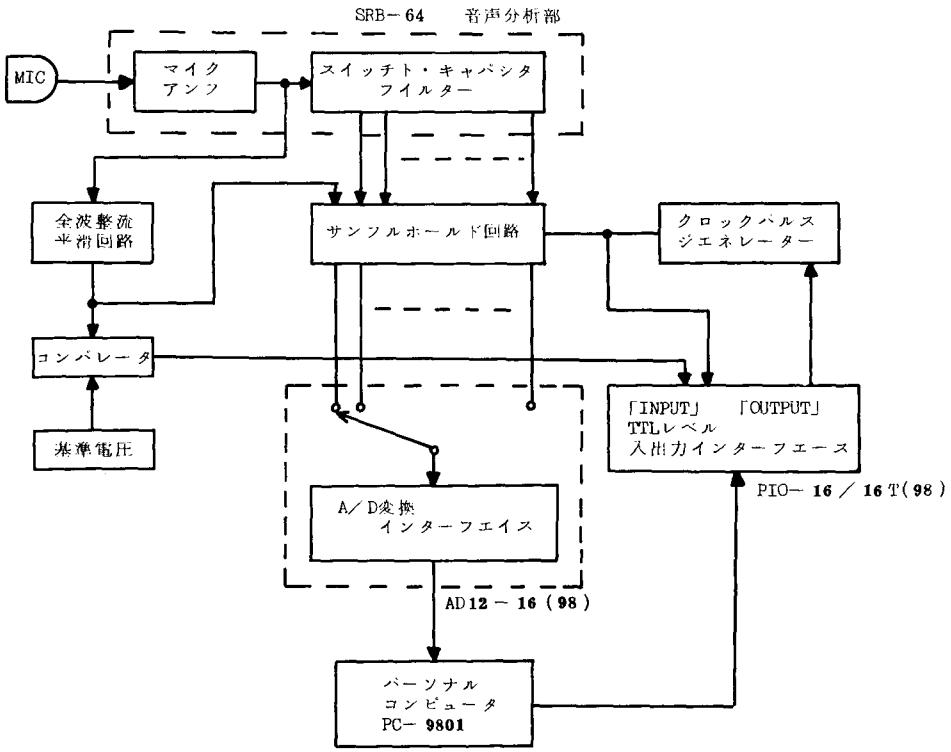


図3　ハードウエアの構成

システム全体の動作の流れをブロック図で示したのが**図3**である。コンピュータにとり込まれるフィルターバンクからの信号は，すべてA/D変換器を経ており，この制御のために，入出力インターフェースを介して制御用パルス信号が出入りする。この制御指令は音声スイッチ機構から発せられる。音声レベルがある一定の値を超えると，コンパレータが動作し，これがコンピュータを介してクロックパルスジェネレータにスタート指令を与える。このクロックパルスでサンプルホールド回路が働き，サンプリング 1ms，ホールディング 4ms の周期でサンプル・ホールドをくり返す。ホールド期間中に 16チャンネルの音声信号がすべてA/D変換されてコンピュータにとり込まれる。以上の過程のタイミングを示したのが**図4**である。

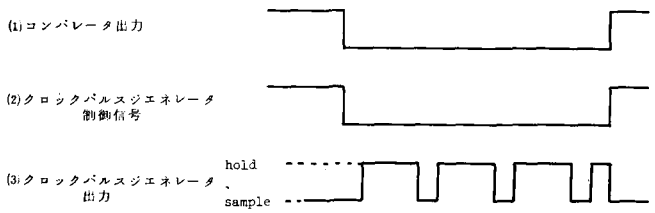


図4　制御信号のタイムチャート

PC-9801 にはハードウエア割込みとして，INT 0，4，5，6 の 4レベルの割込みが開放されており，本システムでは INT 4 と INT 6 の割込みを使用している。

図4(3)に示すように，クロックパルスジェネレータのアップエッヂ（S&H 回路のホールド状態）ごとに INT 6 の割込みが発生する。割り込み処理ルーチンでは，割り込み回数をカウントしており，本システムでは7回の割込みごとに一度，すなわち 35ms ごとの周期で 16チャンネルの音声信号がスキャニングされ，コンピュータにとり込まれるようになっている。

この割込み回数（7回）は，音声認識の精度と速度の両面から実験的に定めたもので，大きすぎるとデータのとり込み間隔が粗くなり，認識精度が落ちる。小さくしすぎるとデータ量が多くなりすぎて処理に時間を要し認識時間が遅くなる。この回数はソフト的に容易に変更できるようになっている。また，A/D変換器でも割り込み（INT 6）を使用しており，A/D変換器が動作中は，別のプログラムを実行できるよう効率良く製作されている。

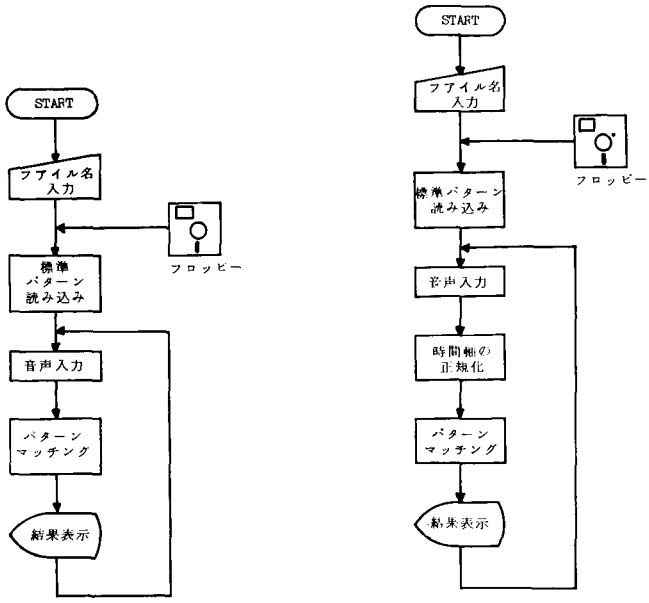
3. 認識の方法

認識の実験は，母音（アイウエオ）と単語（0，1，2，～9）の両方について行った。**図5**(a)，(b)は，それぞれのフローの概略を示すもので，単語認識の場合には，時間軸の正規化の過程が追加されている。なお，使用した言語は，N_{88}-BASIC(86) で，制御用としてアセンブラも一部使用している。

3.1 母音認識の方法

特定話者を対象にした母音認識の実験である。まず母音の標準パターン作成のため，各音について5～20回発声して周波数帯域ごとに平均値を求め，各母音 i ごとに 15次元ベクトル μ_i として登録しておく。さらにパターンのゆらぎを求めるため，各帯域 k ごとに標準偏差を計算し，これを母音 i ごとにまとめて，標準偏差 $\sigma_{i,k}(k=1\sim15)$ を作成登録する。

このような準備ができ上った段階で，テスト音声パターンが入力されると，これがフィルターバンクを通して 15段階のスペクトルパターンになる。いま未知音声のスペクトルパターン



(a) 母音認識の場合　　(b) 単語認識の場合

図5 音声認識のフロー

（15次元）をベクトル的に$\mathbf{X}$で表現し，その成分をX_k(k＝1～15)とすると，未知音声$\mathbf{X}$と母音iとの間の判別距離は，次式のようになる。

$$S_i=\sum_k(X_k-\mu_{i,k})^2 \tag{3.1}$$

これは単純な判別距離であるが，これに標準偏差の逆数による重みを附加した場合は，

$$S'_i=\sum_k\left\{\frac{1}{\sigma_{i,k}}(X_k-\mu_{i,k})^2\right\} \tag{3.2}$$

認識はいずれも，この判別距離をすべての母音の標準パターンに対して計算し，最小距離のものに属すると判定する。実験は，式（3.1）の単純距離と，式（3.2）の重み付距離の両方について行いスコアを比較した。また標準パターンの集録方法によっても影響があるので，集中発声（アアア……，イイイ……，）と順次発声（アイウエオ，アイウエオ，……）に分けて比較実験を行った。

3.2 単語認識の方法

単語認識の実験は，10単語〔0(ゼロ)，1(イチ)，2(ニ)，3(サン)，4(ヨン)，5(ゴオ)，6(ロク)，7(ナナ)，8(ハチ)，9(キュウ)〕について，特定話者を対象に行った。認識のフローは，**図5**(b)の通りで同図(a)と似通っているが，スペクトルパターンが15次元の時系列パターンであることと，時間軸の正規化が追加されている点が異なっている。時間軸の正規化は，同じ単語を短かく発声したり，長く発声したりした場合に誤判定しないように時間軸の長さを揃えるためのものである。現在では DP マッチングなどの 非線形手法が一般的になっているが，ここでは処理時間を考慮して，単純な線形伸縮（一様伸縮）法を採用し，全体を12フレーム（35ms×12）に基準化した。

また単語音声の切り出しに際しては，単語内の休止部分と単語の終了部とを混合しないようにシステムを作成した。すなわち，音声エネルギーパターンは一般に**図6**の例のように大きい

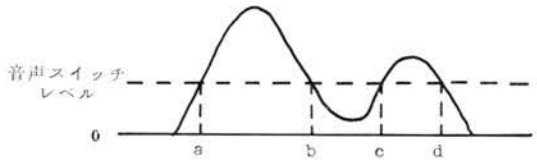


図6　音声エネルギーの変動

変動があるのが普通で，図のb～c間のようなレベルの低い所が休止期間になるが，これが単語終了部と混同される恐れがある。したがって，この区間が 300ms 以内ならば，単語の継続期間と判断するようにシステムを作成した。

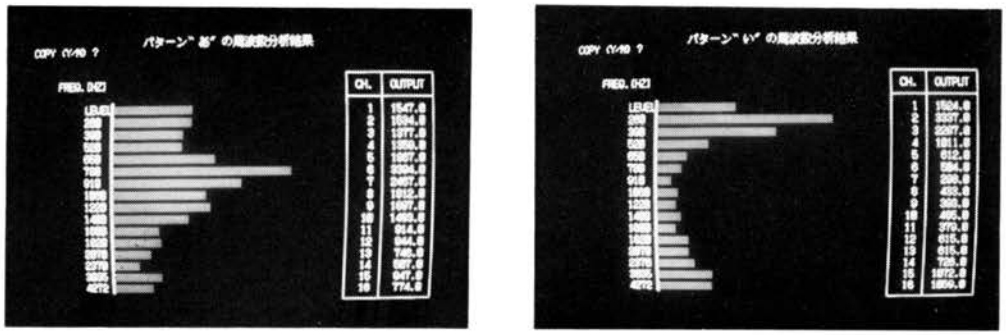
処理時間の短縮のため，標準パターンは1回だけの発声で学習し，単純距離によって判別を行った。判別距離は，つぎのように時系列パターンの偏差の総和となる。

$$S_i=\sum_t[\sum_k(X_k-\mu_{i,k})^2] \qquad (3.3)$$

4. 実験結果

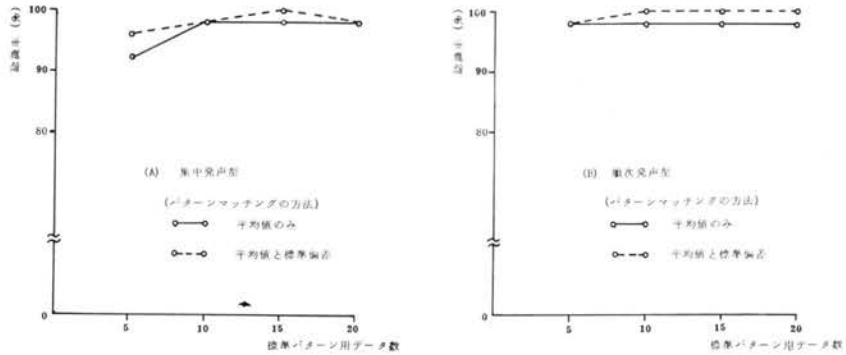
4.1 母音の認識実験

前述の集中発声形（アアア……，イイイ……）と順次発声形（アイウエオ，アイウエオ，…



(a) ア　　(b) イ

図7　スペクトルパターンの例



(a) 集中発声の場合　　(b) 順次発声の場合

図8　母音認識実験

…）の二つの場合について5～20個で標準パターンを作成し，それぞれについて，平均値のみの単純距離と標準偏差を加えた重み付距離による実験を行った。この場合の標準パターンの例を**図7**に示す。**図8**は認識実験の結果である。

集中発声と順次発声では，後者のほうがスコアが良好である。また，いずれの場合も標準偏差による重みづけを行った方が良い結果が得られている。認識に要する時間は，重みづけのある場合 0.9秒，重みづけのない場合 0.5秒であった。

表3 単語認識実験

使用した標準パターン	認識率(%)
WORD 1	99.0
WORD 2	100.0

4.2 単語の認識実験

数字音声（0～9）を1回づつ合計10回発声し，標準パターン群を2種類作成し，これをWORD 1，WORD 2として，それぞれについて各数字を10回づつ合計100回発声し認識率を調べた結果が**表3**である。なお認識の平均所要時間は 11.2秒であった。

5. 検　　討

この実験では，母音認識と単語認識をそれぞれ若干異る手法で行っているので，別々に検討を加えよう。

母音認識では標準パターンの良否が問題になるが，集中発声形よりも順次発声形の方が良好な標準パターンが得られるようである。集中発声形では，その時のコンディションで片寄ったパターンに集中しやすくなるためであろう。これに対し，順次発声形ではパターンが適当にバラついて，標準パターンも平均的なものが得られるものと推定される。また，標準偏差によるパターンの重みづけも有効なことが立証された。

単語認識では，スコアは 99～100%と良好であるが所要時間が平均11秒と長いのが難点である。この内，時間軸の正規化に約4秒を要している。DP マッチングなどの手法をとればさらに長時間要すると思われる。この時間の問題は，マイコンのソフト N_{88}-BASIC(86) に原因があることが確認されている。この PC-9801 には数値計算用の協調プロセッサ 8087 が取付けられているが，これが十分に働いていない。例えば乗除算などは，ソフトウエアが 8087 をサポートしていないので，もしこの部分が改良されれば，速度は大巾に改善できると推察される。

6. む　す　び

マイクロコンピュータによるスペクトルパターンマッチング形の音声認識装置を試作して，良好な結果を得た。特定話者の母音認識（5母音）ではほぼ 100%，単語認識（10数字）でも99～100%の認識率であった。

経済性のために，やや安定性には問題があるとされるスイッチト・キャパシタ・フィルター用 LSI・ASA-16 を使用したが，本システムのような学習形の アルゴリズムを採る限り問題はないようである。一度標準パターンを作成して1週間後に認識実験を行ってもスコアにほとんど差はなかった。LSI 特性のドリフトも，時間の経過による 音声パターンの変動も，吸収して実用上問題はないようである。

また，認識所要時間は，母音で 0.5～0.9秒，単語で約 11秒とやや長いが，使用ソフトがバ

ージョン・アップされて，協調プロセッサ 8087 を完全に利用できるようになれば，大巾に短縮される見込である。

謝辞　日頃適切な御助言と有益な資料を御提供下さっている京都大学工学部の坂井利之教授，ハードウエアの製作について有益な御助言を頂いた豊橋技科大の中川聖一講師，製作に協力して頂いた研究室の中嶋文雄君（現松下電産）に深く謝意を表する。

参 考 文 献

(1) 安居院猛，中島正之　コンピュータ音声処理　産報出版
(2) 千葉成美　単語音声の認識　情報処理学会　1978-7　vol. 19
(3) 高橋・黒田・福岡　マイクロコンピュータを用いた音声認識装置　大阪府立高専紀要　16巻
(4) 坂井利之，中川聖一　「単語音声汎用認識装置の開発」研究成果報告書　京大工学部　昭和55年３月