

音声認識用マイクロプロセッサ

松木智子 石塚久夫 亘理誠夫 岩田利喜 川上雄一
日本電気株式会社

1. はじめに

科学技術の進歩に伴い、コンピュータが社会に普及するようになるにつれて、より使い易いものが要求されるようになった。一般に、コンピュータ等の情報処理装置を使用する場合、入力はキーボード等から手で行う。ここで、キー入力のかわりに、音声により入力を行う場合を考えてみる。情報処理装置への入力を行う際、両手か他の処理のために使えない場合でも、音声による入力が可能ならば、その情報処理装置を操作することができる。音声による入力は、キー入力に比較して人間にとってより自然であり、キーボードに慣れていない場合は特に、正確で、より速く、簡単に行うことができる。さらに、英語の文字数とは比較にならない程、大量の文字数进行处理する日本語ワードプロセッサにおいては、音声入力が、必要欠くべからざるものとなるだろう。

このように、コンピュータ等情報処理装置の操作を、より簡単に行うための方法の一つとして、装置への音声による入力、音声認識装置が、徐々に、実現され始めた。

このような目的で用いられる音声認識をより簡単に実現するため、特定話者用の音声認識用マイクロプロセッサ(SRP)を、このたび開発した。

本稿では、このSRPのアーキテクチャと本SRPを使用した音声認識システムについて述べる。

2. 音声認識

音声認識処理は、発声された音声か何であるかを、推定することである。具体的には、予め登録された複数の標準パターンのそれぞれと、入力パターンとの照合を行い、標準パターンの中から、入力パターンに近いものを検索し、選出することである。

音声認識処理において、まず最初に、音声を分析しなければならない。音声を分析するためのフィルターを、図1に示す。入力音声を、バンドパスフィルター(BPF)、整流回路及びローパスフィルター(LPF)から構成されるフィルターバンクに入力し、その出力を、20msec程度のフレーム周期でA/D変換し、行列状の音声パター

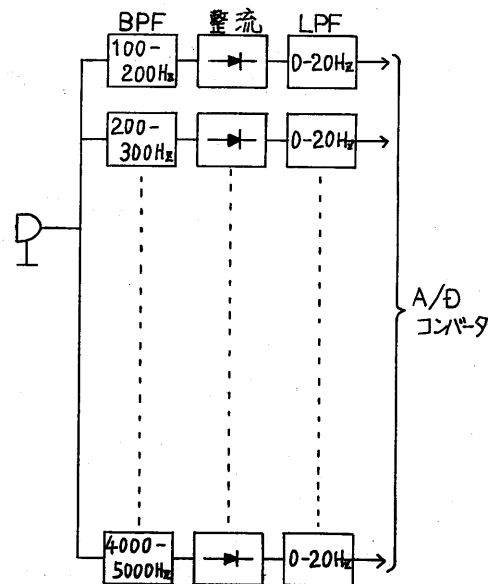


図1. 音声分析フィルター

ンを得る。音声パターンAは、ベクトル a_i の時系列として、

$$A = a_1, a_2, \dots, a_i, \dots, a_T$$

と表わされる。

音声パターンは、同一話者、同一音声の場合でも、複雑に変動している。特に、時間方向の伸縮が著しい。音声認識装置において、この時間的な歪を除去するDPマッチングアルゴリズムに基づいたパターンマッチング方式が広く用いられている。

2.1 DPマッチングアルゴリズム

DPマッチングアルゴリズムは、次のように説明される。

認識させようとする入力パターンAと登録されている標準パターンBの1つBを、それぞれ以下のようにする。

$$A = a_1, a_2, \dots, a_i, \dots, a_T$$

$$B = b_1, b_2, \dots, b_j, \dots, b_J$$

ここで、 I, J は、それぞれのパターンのフレーム数である。これら2つの特徴ベクトル a_i と b_j の間の距離 $d(i, j)$ を求めるために、

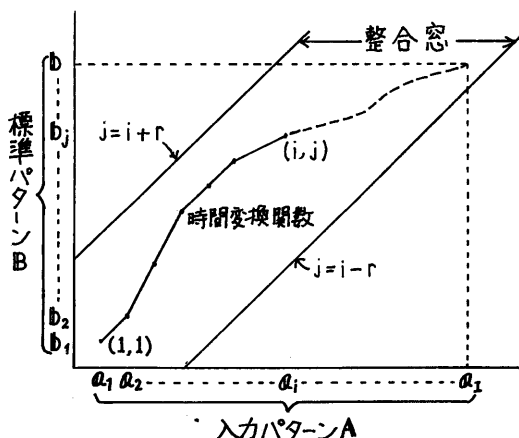


図2. DPマッチングの原理

$$d(i, j) = \| a_i - b_j \| \quad (1)$$

を用いる。DPマッチングアルゴリズムによれば、これら2つのパターン間の時間正規化距離 $g(I, J)$ は、以下の漸化式計算

$$g(i, j) = d(i, j) + \begin{cases} g(i-1, j) \\ g(i-1, j-1) \\ g(i-1, j-2) \end{cases} \quad (2)$$

を行うことにより得られる。 $g(I, J)$ を求めるためには、図2に示すように、時間軸の極端な変動を避けるために設定したマッチングの整合窓を、 $|i-j| \leq r$ として、 $(1, 1)$ から (I, J) まで、整合窓内のすべての $g(i, j)$ を計算しなければならない。従って、整合窓幅 $2r+1$ に、標準パターンのフレーム数 J をかけ合わせた回数の特徴ベクトル距離計算と漸化式計算を行うことにより、時間正規化距離 $g(I, J)$ が得られる。

このようにして、得られたすべての標準パターンとの時間正規化距離の中から、最も小さい距離となる標準パターンを検索し、認識結果とする。

たとえば、特徴ベクトルが、16チャンネルデータなら、1つのベクトル距離 $d(i, j)$ を求めるために、少なくとも16回の減算と16回の加算を行う必要がある。さらに、時間正規化距離 $g(I, J)$ を得るための1回のDPマッチング処理において、整合窓幅 $2r+1$ を21、標準パターンのフレーム数 J を30と設定した場合に、すべてのベクトル距離計算、少なくとも $(16+16) \times (2r+1) \times J = 20160$ 回の計算を行わなければならない。従って、標準パターンが、100個あるとするなら、ベクトル距離計算回数は、200万回以上ということになる。加えて、漸化式計算もまた、1回のDPマッチング処

理のために、 $(2r+1) \times J$ 回実行しなければならない。

2.2 音声認識専用LSIの必要性

すでに述べたように、音声パターンは、時間方向での変動が著しい。図3に示される標準パターンと入力パターンとの時間軸の線型な伸縮による音声パターンの整合という旧来の方法に比較して、非線型な伸縮を行うDPマッチング法により、認識率は、著しく向上した。音声認識処理においては、実時間処理が要求されるため、入力音声の発声終了後に、人間の耳で実時間での応答と感じられる限界の300 msec程度で、認識結果を出すことが求められる。DPマッチング処理は、大量の演算を必要とするので、一般的なマイクロプロセッサを用いての音声認識装置の実現は、困難である。

ベクター距離計算は、大量の演算を行う必要があるが、同形の演算を繰り返すため、ベクター距離計算専用の演

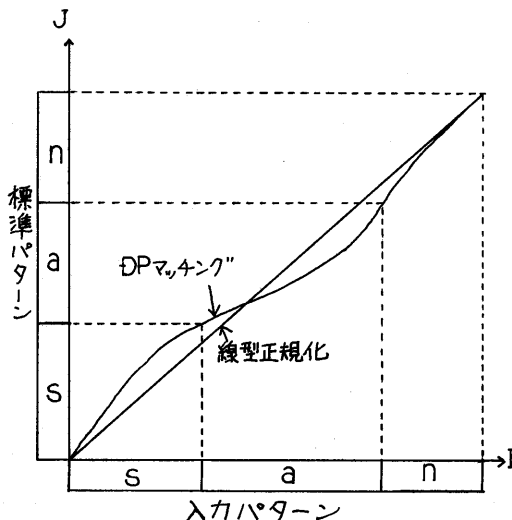


図3. 線型正規化とDPマッチング

算器を複数個用意することにより、演算速度を速めることが可能である。入力パターンと1つの標準パターンとのDPマッチング処理のために、1回の距離計算と1回の漸化式計算を行うが、距離計算と漸化式計算を共通の演算部で行う場合、演算速度が遅くなる。従って、DPマッチング処理を効率よく高速で行うために、距離計算専用の演算部と漸化式計算を行う演算部を別々に用意することにより、距離計算と漸化式計算を同時に実行させる方式が有効である。

3. 音声認識用マイクロプロセッサの構成

音声認識用に開発されたシングルチップSRPは、高速処理を要求されるため、次のようなシステム構成とした。

- 1) デュアルプロセッサ構成:
距離計算用のD-プロセッサと漸化式計算用のG-プロセッサ。
- 2) 2組のパイプライン構成の減算器と加算器のセット。
- 3) 大容量のパターンバッファメモリ。

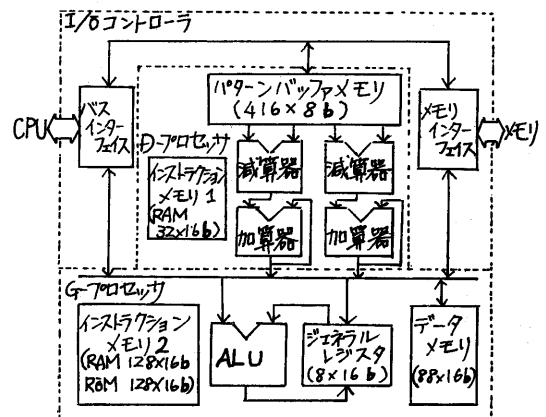


図4. SRPマイクロプロセッサ構成図

3.1 ハードウェア

DPマッチング処理は、距離計算と漸化式計算の2段階に分けられる。従って、DPマッチング処理を行う場合、デュアルプロセッサ構成は、非常に有効である。

図4にSRPのブロック図を示す。SRPは、距離計算部(D-プロセッサ)、漸化式計算部(G-プロセッサ)、I/Oコントローラの3つの主要なブロックから構成される。

a) D-プロセッサ

D-プロセッサは、16ビットのマイクロインストラクションによって制御され、ベクター距離計算を行う。ベクター距離計算は、処理内容は簡単だが、大量の演算を必要とするため、D-プロセッサには、パイプライン処理が適用されている。D-プロセッサは、パイプライン構成の減算器と加算器のセットを2組備え、2つの距離計算を並列に実行することができる。1フレームが16チャネルであるとした場合、1フレームのベクター距離計算に必要な演算時間は、等価的に2μsecで実行される。また、D-プロセッサには、外部パターンメモリとSRPの相互間の影響を減らすため、マッチングの整合窓内で距離計算を必用とするすべてのベクターを保持できる大容量のパターンバッファメモリが備えられている。

b) G-プロセッサ

G-プロセッサは、D-プロセッサと同様に、16ビットのマイクロインストラクションによって制御され、DPマッチング処理を行うための漸化式計算を実行する。

4バンク×227ワードのデータメモリは、漸化式計算の実行に都合がよいように設計されている。通常、漸化式計算には、2バンク必要で、残り2バン

クは、マッチング結果の格納や、ワーキングエリアとして用いられる。

また、8ワードのジェネラルレジスタは、漸化式計算(2)において、最小値を求める際、データメモリとの間の転送処理を省くことができる。

G-プロセッサは、D-プロセッサとI/Oコントローラの管理も行う。さらに、汎用の機能も用意されている。

c) I/Oコントローラ

I/Oコントローラは、ホストCPUとSRP、及びSRPと外部パターンメモリとの間のデータ転送を管理する。

I/Oコントローラには、メモリリフレッシュ制御が用意されており、簡単なシステム構成を可能とするため、通常のダイナミックRAMを、直接接続することができる。

また、I/Oコントローラは、外部パターンメモリから、内部パターンメモリへのデータ転送を容易に行うため、メモリアドレスと転送データ数、レジスタにセットするだけで、DMA的にデータを転送できる。

d) ハードウェアの特徴

SRPのハードウェアの特徴を表1にまとめる。プログラムメモリとしてRAMを用いているため、融通性のあるダイナミックマイクロプログラミングが可能である。従って、プログラムを再ロードすることにより、異なるアルゴリズムを、同一SRP上で実行することができる。

インストラクションメモリ2のROM部分は、プログラムローダや、乗算除算ルーチン等のユーティリティサブルーチンプログラムが格納されている。プログラムローダは、インストラクションRAMに、標準パターン登録プログラム、DPマッチングプログラム等

をロードするためのものである。ユーザーは、ユーティリティプログラムを使用して、簡単にプログラムを作成できる。

チップ上には、7296ビットの大容量RAMが集積されている。このチップは、2.5 μ mのチャネル長のN-チャネルシリコン-gate MOS技術を用いて、チップサイズ6.26 $\times$ 7.49mm²のチップ上には、約43000個のトランジスタを集積している。チップは、40ピンのDIPに收容されている。

3.2 SRP インストラクション

D-プロセッサのインストラクションは、ベクター距離計算用に設計されている。

表1. SRPのハードウェアの特徴

デュアルプロセッサ構成	
距離計算プロセッサ (D-プロセッサ)	
新化式計算プロセッサ (G-プロセッサ)	
D-プロセッサ	
インストラクションメモリ	1 RAM 32 words $\times$ 16 bits
パターンマッチメモリ	RAM 416 words $\times$ 8 bits
G-プロセッサ	
インストラクションメモリ	2 RAM 128 words $\times$ 16 bits
	ROM 128 words $\times$ 16 bits
データメモリ	RAM 88 words $\times$ 16 bits
8レベルサマリー4スタック	
インストラクションサイクルタイム 250 nsec	
CPUインターフェイス	
8-bitマイクロプロセッサバスコンパチブル	
DMAインターフェイス	
2レベルインタラプト	
メモリインターフェイス	
ダイナミックRAMダイレクトインターフェイス可能 (リフレッシュ回路内蔵)	
8 MHzの1相クロック	
2.5 μ m N-ch Si-gate MOS	
+5V単一電源	
40ピンDIP	

ている。D-プロセッサのインストラクションは、図5に示されるように、2つのタイプに分類される。ALU命令は、メモリ読み出しと、減算・加算動作を制御し、定数設定命令は、定数設定動作を制御する。

G-プロセッサのインストラクションは、図5に示されるように、4つのタイプに分類され、それぞれ、乗算動作、定数設定動作、データ転送、分岐を制御する。

G-プロセッサの定数設定動作を除いた、他のすべてのインストラクションは、250 nsecのインストラクションサイクルタイムで実行されるように設計されている。

Dインストラクション

OP	CONTROL	R	減算	加算	ALU命令
----	---------	---	----	----	-------

OP	DST	定数	定数設定命令
----	-----	----	--------

OP : オペレーションタイプ

R : メモリ読み出し制御

DST : デスティネーション設定

Gインストラクション

OP	BS	ALU	RS	ALU命令
----	----	-----	----	-------

OP	BS	ALU	RS	定数
定数設定命令				

OP	SRC	DST	AC	MOVE命令
----	-----	-----	----	--------

OP	CND	アドレス	分岐命令
----	-----	------	------

OP : オペレーションタイプ

BS : レジスタ設定

RS : ジェネラルレジスタ設定

SRC : ソース設定

DST : デスティネーション

AC : データメモリのアドレス制御

CND : 分岐状態指定

図5. SRPインストラクション

3.3 システム構成

SRPを用いたシステム例を図6に示す。この例では、マイクロプロセッサ(CPU)、メモリ、音声分析器、パターンメモリ、そしてSRPで構成されている。音声分析器は、入力音声信号から、音声の特徴を抽出する部分である。音声の始端及び終端を検出する音声検出処理は、CPUにより実行される。SRPは、パターンメモリにストアされた入力パターンと標準パターンとの間のDPマッチング処理を実行することにより、音声認識処理を行う。図6に示すシステムは、パターンメモリとSRPを、複数個、CPUに接続することにより、認識単語数と簡単に増やすことができる。認識音声を決定する最終判断と、システム全体の制御は、CPU又は、SRPにより行われる。

4. SRPによる

DPマッチングプログラム例

DPマッチング処理の基本サイクルは、I/Oコントローラによるパターンデータ転送、D-プロセッサによる距離計算、G-プロセッサによる漸化式計算という3つの連続動作から成る。

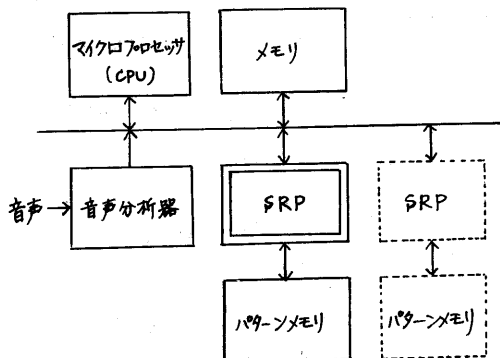


図6. 音声認識システム

この動作は、パイプライン処理で行われる。

2つの音声パターンについて、図2に示されるように、 i 及び j について、 $(1,1)$ から (I,J) まで、マッチングの整合窓内、即ち $|i-j| \leq r$ の範囲でマッチング距離を求める。これにより、時間変換関数の道筋、即ちマッチングパスが得られ、時間的な歪を除去した、時間正規化距離 $g(I,J)$ が得られる。DPマッチング処理は、マッチングパスを求めるために、ベクター距離計算及び漸化式計算を行うことである。

図2に示す、DPマッチングによる最適パスを求める例を、以下に示す。

a) ベクター距離計算

D-プロセッサは、2つの距離計算を同時に行うことができる。図7のフローチャート及び図8のブロック図を用いて、入力パターンベクター a_i, a_{i+1} と標準パターンベクター b_j, b_{j+1} とのベクター距離 $d(i,j)$ 及び $d(i+1,j)$ を求める方法を、以下に示す。RAM-Aには、入力パターンが、またRAM-Bには、標準パターンが書き込まれる。1回のデータ読み出し命令で、RAM-A、RAM-Bそれぞれの隣り合ったバンクから、2ワードずつデータが読み出され、それぞれ、バッファA1、A2、B1、W1にラックされる。ALU0、ALU1で、それぞれA1とB1、A2とB1にラックされたデータの減算及び減算結果の絶対値をとる処理が行われる。この処理によって得られた結果は、それぞれ、L0、L1にラックされる。ALU2、ALU3は、共に加算器であり、L2、L3にそれぞれラックされた値に、L0、L1の値をそれぞれ加算し、加算結果を、L2、L3にそれぞれラックする。以上の処理は、Dインストラクション

の1命令、250 nsec で実行できる。
1フレームを、16チャネルとした場合、このパイプライン化された処理を、16回繰り返すことにより、ベクトル距離 $d(i, j)$ 及び $d(i+1, j)$ が得られる。

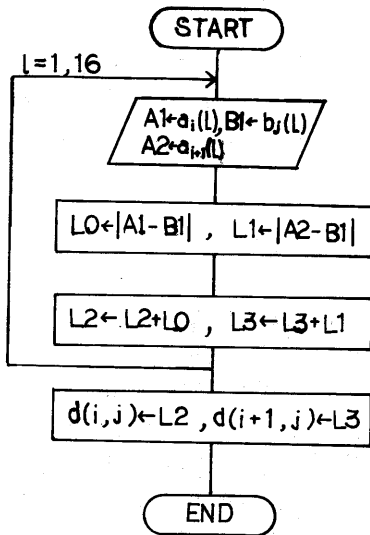


図7. ベクトル距離計算フローチャート

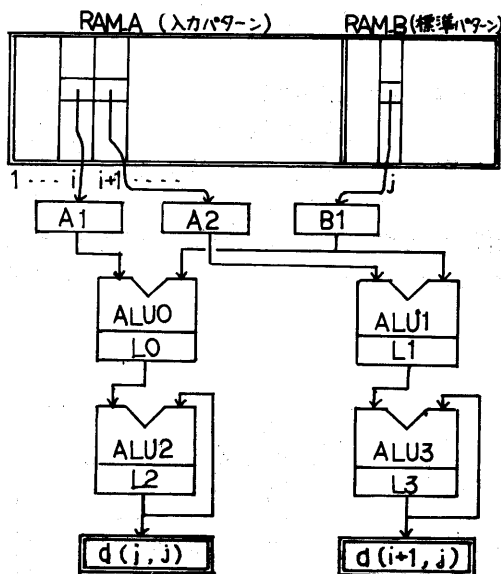


図8. ベクトル距離計算

b) 漸化式計算

図2における座標 (i, j) までのマッチング距離 $g(i, j)$ を求める方法について、図9のフローチャート及び図10のDPマッチング計算図を用いて、説明する。

ベクトル距離計算ルーチンを1回実行することにより、2点のベクトル距離が得られる。従って、漸化式計算ルーチンにおいても、2回の漸化式計算を行い、2点のマッチング距離を得るように、効率のよいプログラミングを行う。

1回目の漸化式計算により、 $g(i, j)$ を求める。 $G(i-1)$ に格納された、 $g(i-1, j)$ と $g(i-1, j-1)$ を比較し、小さい方を $G(i)$ に入れる。 $G(i)$ に格納された値と $g(i-1, j-2)$ を比較し、小さい方を $G(i)$ に入れる。DPプロセッサから、 $d(i, j)$ を受け、それを $G(i)$ に加えることにより、 (i, j) までの、マッチング距離 $g(i, j)$ が得られる。次に、2回目の漸化式計算で、 $g(i+1, j)$ を求める。 $d(i, j)$ のベクトル距離計算を行う際、同時に、 $d(i+1, j)$ が得られることから、ここで、 $g(i+1, j)$ を求めるべく、同様の比較選択処理を行う。比較処理により得られた最小値に、距離 $d(i+1, j)$ を加えることにより、マッチング距離 $g(i+1, j)$ を求めることができる。 $r=10$ とすると、マッチングの整合窓幅は、 $i=j-r$ から $i=j+r$ までの $2r+1=21$ である。2点のマッチング距離が得られる以上の処理を、11回繰り返し、最後の回の2回目の漸化式計算を行わないことにより、21点のマッチング距離が得られる。

以上の漸化式計算により、 j 段目のマッチング距離が得られる。

ここで、メモリを効率よく使用する

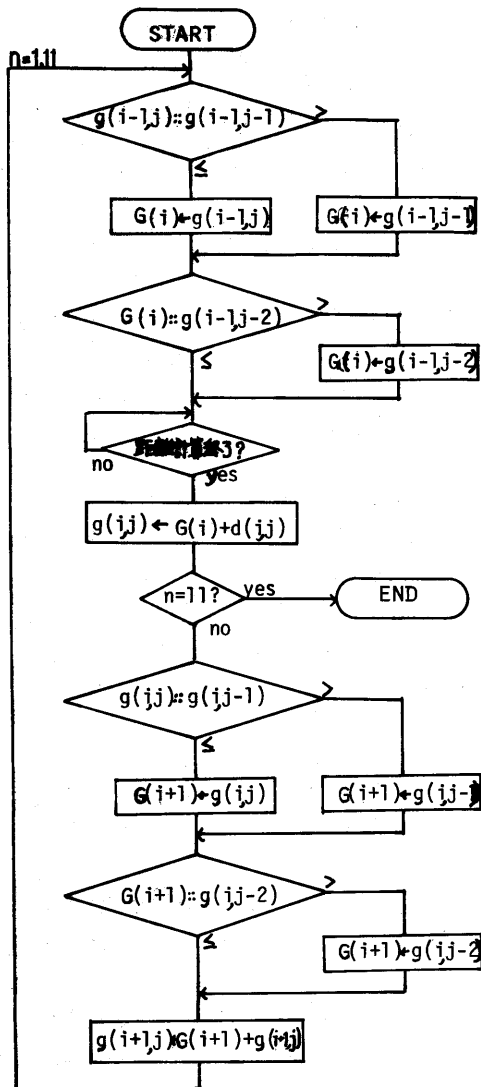


図9. 新化式計算フローチャート

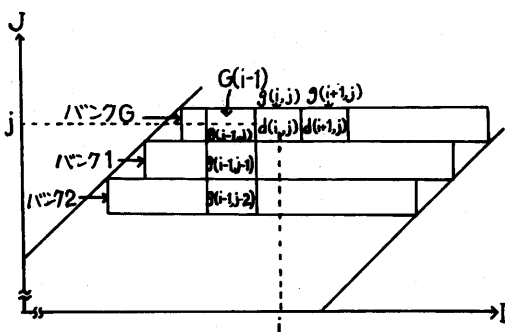


図10. DPマッチングの計算(1)

ために、得られたマッチング距離を、図11に示すように、バンク2に格納していく。座標 (i, j) についての21個のマッチング距離を格納した後、バンク1とバンク2を入れ換える。

これにより、データメモリを、2バンク用いるだけで、新化式計算が実行できる。

以上の処理を、1段目より J 段目まで繰り返すことにより、時間正規化距離 $g(I, J)$ が得られる。

5. 認識性能

5.1 離散単語認識

離散単語認識の場合、DPマッチング処理は、発声終了後に開始される。入力パターンと1つの標準パターンとの間の時間正規化距離を得るためのDPマッチング処理は、パターンデータ転送時間を含めて、約1.7 msecを要する。従って、この音声認識用マイクロプロセッサ(SRP)は、入力音声発声終了後、300 msec以内に、約180単語の標準パターンとのマッチングが行える。

時間正規化距離が、あまりにも大きいものは、認識単語ではないと仮定して、除外するように、ある閾値を設定することができる。DPマッチング距離計算途中で、この閾値を越えた場合計算を中断するものとすれば、離散単語

バンクG	$g(i-1, j)$	$g(i, j)$
バンク1	$g(i-1, j-1)$	
バンク2		$g(i-1, j-2)$

図11. DPマッチングの計算(2)

語認識の場合、実際の認識語数は、最大340まで、増加できる。

5.2 連続単語認識

連続単語認識の場合、音声パターン¹⁾の時間方向の歪の除去、他に、単語間の境界を検出しなければならない。これを解決するために、2段階DPマッチング法が発表されている。この方法は、各単語音声の標準パターンと、あらゆる組み合わせで連結した連続音声の標準パターンと、入力連続単語音声との間で、DPマッチング処理を行うことを基本原理としている。この2段階DPマッチング法は、単語レベルのマッチングと文レベルのマッチングとの2段階のマッチングで実行される。単語レベルでのマッチングは、入力パターンをさまざまに分割したものすべてと、各単語標準パターンとの間のマッチング距離計算を行う。文レベルでのマッチングは、単語レベルでのマッチング距離を、入力連続単語音声に沿って加算し、その最小値をDPマッチングの漸化式により求める。

以上のように、連続単語認識は、単語レベルマッチングと文レベルマッチングに分けられる。単語レベルマッチングは、離散単語認識におけるDPマ

ッチング²⁾処理と同様で、SRPにより実行される。文レベルマッチングは、単語レベルマッチングに比べ、計算量は少ないが、大きなワークメモリを、必要とするため、SRPのホストCPUにおいて、実行される。

2段階DPマッチング法により、SRPでは、20語まで、連続単語認識ができ、途中打ち切り処理を加えると、40語まで、連続単語認識ができる。

さらに、最近、2段階DPマッチング法と同じ解が得られ、計算量は少ないが、入力音声の終端時点より処理を開始するLEVEL-BUILDING法³⁾や、LEVEL-BUILDING法を、実時間処理用に改良した、クロック同期伝播形DP法³⁾などが提案されている。

SRPによる認識単語数を、それぞれ表2に示す。

6. おわりに

現在、多くの音声認識装置で採用されているDPマッチング処理機能を実現できる、音声認識用マイクロプロセッサ(SRP)を開発した。DPマッチング処理に必要な機能の分析に基づき、SRPは、ベクター距離計算を行うDPプロセッサ、漸化式計算を行うGプロセッサ、パターンバッファメモリ、I/Oコントローラから構成される。SRP1チップで、特定話者認識において、340語までの離散単語認識、40語までの連続単語認識が可能である。また、SRPを複数個接続することにより、認識単語数を、簡単に拡張することができる。

謝辞

本LSIの開発にあたり、御指導いただいた、C&C研究所 千葉部長、

表2. SRPによる認識単語数

アルゴリズム	認識単語数	(途中打ち切り処理を加えた場合)
離散単語認識		
DPマッチング	180語	340語
連続単語認識		
a) 2段階DPマッチング	20語	40語
b) クロック同期伝播形DP	10語*	10語*
c) LB法	20語*	40語*

*入力連続単語音声は、5桁の場合

迫江課長, 第1LSI事業部高島部長,
古橋課長, 超LSI開発本部可児部長,
鈴木課長に感謝致します。また, 本LSI
開発にあたり, 第1LSI事業部
星俊明氏及び日本電気アイシーマイコ
ンシステム(株)佐藤誠氏, 中嶋敏郎
氏に格別の御協力をいただき深謝致し
ます。

参考文献

- 1) H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-26, Feb. (1978)
- 2) H. Sakoe "Two-Level DP Matching - A Dynamic Programming-Based Pattern Matching Algorithm for Connected Word Recognition", IEEE Trans. ASSP-27, No. 6, pp 588-595 (1979)
- 3) 迫江・巨理: クロッ, 7 同期伝播形DP法による連続音声認識の検討, 音響学会研究会資料 S81-65 (1981)
- 4) Myers and Rabiner "A Level Building Dynamic Time Warping Algorithm for Connected Word Recognition", IEEE Trans. ASSP-29, No. 2, pp 284-297 (1981)
- 5) 巨理, 他: 音声認識用DPマシン7"LSIのシステム設計, 信学技報 DPRL 82-87 (1983)
- 6) T. Iwata, et al., "A Speech Recognition Processor", ISSCC (1983)