状況に依存してマルチモーダル情報の選択が可能な連想認識モデルによる音声認識

西崎 誠[1], 大森 隆司[2], 琴寄 貴志[2]

東京農工大学 工学研究科[1] 東京農工大学 生物システム応用科学研究科[2]

孤立単語音声の認識には，DP マッチングが有効であることはよく知られている．ところが DP の扱う対象が連続音声になると，単語の認識は急に難しくなる．通常これらの問題を解決するために，文の構造や意味に関する情報が利用される．これに対し本稿では，状況に応じた認識が可能な連想記憶モデル PATON を用い，状況に依存してさまざまなモダリティ情報を利用することが，単語認識における曖昧性を解消し認識率の向上に有効であることを示す．

A model of connected word speech recognition by associative memory model that select the multi-modal information depending on context.

Makoto Nishizaki[1], Takashi Omori[2] and Takashi Kotoyori[2]

Faculty of Technology, Tokyo University of Agriculture & Technology[1]
Graduate School of Bio-Applications and Systems Engineering, Tokyo University of Agriculture & Technology[2]

It is well known that a continuous speech recognition task by DP matching makes a lot of erroneous recognized candidates because of lack of word segmentation and co-articulation. To avoid such errors, syntax and semantics of sentence are used usually. In this paper, we propose another mechanism based on associative memory with attention which select the multi-modal input depending on a context.

1 はじめに

孤立単語音声の認識には，DP マッチングが有効であることはよく知られている．ところが DP の扱う対象が連続音声になると，単語の認識は急に難しくなる [1]．認識対象のすべての連続音声の中に出現する単語の数が少なく，小語彙であれば 2 レベル DP 法を用いることである程度は対応できる．しかし，その計算量は多く，しかも高い認識率の実現は容易ではない．通常これらの問題を解決するために，文の構造や意味に関する情報が利用される．

我々は，さまざまなモダリティ(属性) 情報を連合し利用することができる連想記憶モデル PATON を提案してきた [2]．PATON は，入力される様々な属性情報を取捨選択できる注意システムを持っており，その場に合わせ必要な情報を選択し状況依存的に認識や連想を行うことができる．本稿では，この PATON で視覚入力がある状況での連続単語音声認識を試みる．そして PATON の持つ情報選択の機構が入力音声の文脈に対応した制約を実現し，音声認識率の向上に有効であることを示す．

2 PATON

2.1 PATON モデル

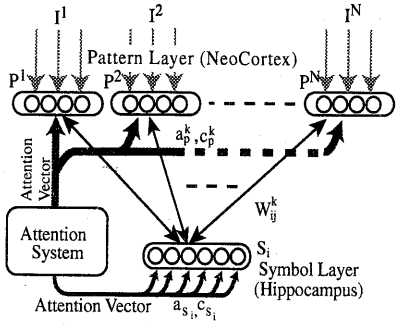

Fig. 1: Architecture of the PATON model.

PATONは脳の海馬系に注目した連想記憶モデルであり，N種類の感覚属性領野（以下，領野）からなるパターン層 (P層) とシンボル層 (S層) さらに注意システムからなる (Fig.1)．P層とS層はそれぞれ大脳皮質の連合野と海馬に対応しており，類似性を考慮した符号化と直交性を考慮した符号化を行う (二重符号化仮説)[3]．注意システムは，P層の各領野ごと，またS層の細胞ごと注意を与え，連想記憶の動作を制御する．この効果が，次節で示すように状況に依存した認識や想起を可能にする [2]．

PATONの構造

P層は，N 種類の領野 P^k $(k=1,\ldots,N)$ に分かれ，それぞれ視覚，聴覚など異なる属性の特徴抽出された結果を表現する．P層の細胞 $P^k=\{P_j^k|j=1,\ldots,M^k,k=1,\ldots,N\}$ は外界から入力 $I^k(t')=\{I_j^k(t')|j=1,\ldots,M^k,k=1,\ldots,N\}$ をうける．ここで t' は外界の時間変化で，M^k は k 番目の領野の細胞数である．S層は L 個の細胞 $S=\{S_i|i=1,\ldots,L\}$ からなり，P層との間に双方向の結合 $W=\{W_{ij}^k|i=1,\ldots,L,j=1,\ldots,M^k,k=1,\ldots,N\}$ がある．この結合 W によりP層とS層の間の連想記憶が実現される．ここで，1つのシンボル細胞とP層との間の結合係数のノルムは1であるとする ($\|W_i^k\|=1$)．S層の細胞間には側抑制の結合があり，入力情報の直交化が行われる．

PATONの動作を差分方程式で書くと以下のようになる．

$$p_j^k(t+1) = p_j^k(t)+\Delta p_j^k \quad (1)$$

$$\Delta p_j^k = \frac{1}{\tau}(-p_j^k + c_1\sum_i S_i W_{ij}^k + c_2 I_j^k)\, c_p^k\cdot\Delta t \quad (2)$$

$$s_i(t+1) = s_i(t)+\Delta s_i \quad (3)$$

$$\Delta s_i = \frac{1}{\tau}(-s_i + c_3\sum_k\sum_j W_{ij}^k P_j^k - c_4\sum_{q\neq i} S_q)\, c_{s_i}\cdot\Delta t \quad (4)$$

$$W_{ij}^k(t+1) = W_{ij}^k(t)+\Delta W_{ij} \quad (5)$$

$$\Delta W_{ij} = \frac{1}{\tau'}\{\alpha(P_j^k - W_{ij}^k)S_i\}\cdot a_l\cdot\Delta t \quad (6)$$

$$P_j^k = a_p^k\,\phi(p_j^k) \quad (7)$$

$$S_i = a_{s_i}\,\phi(s_i) \quad (8)$$

$$\phi(x) = \frac{1}{1+e^{-x/T}} \quad (9)$$

$$I_j^k = I_j^k([\frac{t}{R_{tt'}}]+1) \quad (10)$$

ここで，関数 $[x]$ は x を超えない整数値を返す．また外部入力（外界）の時間変化 t' とPATON内部の時間変化 t との関係は一定であるとし，Δt と $\Delta t'$ の関係を $\Delta t'=R_{tt'}\cdot\Delta t$ ($R_{tt'}$ は1以上の整数) とした．p_j^k と P_j^k は P^k 領野における第 j 細胞の膜電位と出力値で，s_i と S_i はS層における第 i 細胞の膜電位と出力値である．また，α は W_{ij}^k の学習率，τ, τ' はそれぞれ膜電位と学習の時定数である．また，$a_p^k, c_p^k, a_{s_i}, c_{s_i}, a_l$ は注意ベクトル

$A(d)=\{a_p^k, c_p^k, a_{s_i}, c_{s_i}, a_l\}$

a_p^k : gain the output of p^k attribute area

c_p^k : update ratio of the membrane potential in p^k area

a_{s_i} : gain the output of S_i neuron

c_{s_i} : update ratio of the membrane potential of S_i neuron

a_l : update ratio of connection W_{ij}^k

を表現し，それぞれ [0,1] の実数値をとる．d は注意ベクトルを変化させるクロックを表わし，外部入力の変化（ボトムアップ）や認識の動作に対応し（トップダウン）発生する．したがって，PATONはこの

クロックで動作する注意システムで制御される．

2.2 文脈依存の認識，想起，連想

PATONは注意ベクトル $A(d)$ の値によって，状況に依存した「認識」「想起」「連想」を行うことができる．

(1) 認識

注意ベクトルを（$c_p^k = 0, c_{s_i} = 1, a_p^k \in [0,1], a_{s_i} \in [0,1]$）と設定することで，P層で保持された入力パターンをS層で認識することができる．この時，P^k 領野の発火を制御する注意信号 a_p^k によって，入力が同じであっても異なる認識結果を得ることができる．

(2) 想起

注意ベクトルを（$c_p^k = 1, c_{s_i} = 0, a_p^k \in [0,1], a_{s_i} \in [0,1]$）と設定することで，発火しているS層の細胞に対応した各P層での表現を想起することができる．このとき，文脈として p^k 領野への注意 a_p^k が異なると，異なったパターンをP層に想起することができる．

(3) 連想

認識，想起の動作を繰り返すことでS層の細胞間で連想の機能を実現できる．その際，注意の向け方を変えることで，向けた領野の特徴にそって，異なる連想結果を得ることができる．

3 パターン情報の符号化

3.1 PATONからパターン層への要請

PATONは状況に依存した認識を実現するために，注意システムを導入し，P層各領野の出力値の重み付けを行う．この操作が有効に働くためには，各領野に均等に注意が向いている場合，各領野からS層への影響力は等しくなければならない．

しかし，この条件を一般に満たすのは難しい．今，すべてのP層領野に大きさ1の注意がかかっているとする．この時，P^k 領野からS層の細胞 S_i が受ける入力 net_i^k は，

$$net_i^k = W_i^k \cdot P^k \tag{11}$$

$$= \| W_i^k \| \cdot \| P^k \| \cdot \cos\theta_i^k \tag{12}$$

になる．ここで，W_i^k は P^k 層と S_i 細胞との間の結合係数で，そのノルムは1である．$\cos\theta_i^k$ は，W_i^k と P^k とのなす角である．P層には様々な感覚領野があり，それぞれ異なる次元数及び性質をもつ．それゆえ，$\| P^k \|$ の値は，表現されるデータによってかなりばらつきがあり，ノルムの大きなパターンの領野からの影響力が強くなる．そこで，P層に現れるパターンのノルムが1であるという制約を課す．これにより S_i 細胞への入力は，$net_i^k = \cos\theta_i^k$ となり，入力のノルムに影響されず注意が有効に働く．

3.2 入力パターンのノルム一定の実現

3.2.1 砂時計ネットワーク

P層各領野で表現されるパターンは，そのノルムが1であり，外界での類似関係を符号化している必要がある．これらの制約を満たすため，我々は砂時計型ネットワークを改良し，利用する．砂時計型ネットワークは，5層の階層型ニューラルネットワークで第3層を中心に対称の形をしている．このネットワークの1層への入力と5層からの出力が等しくなるようにBack Propagation(BP)で学習することで，入力データの位相情報を保ったまま3層の細胞数の次元にまで圧縮した表現を得ることができる [5]．

3.2.2 砂時計ネットワークによるノルムと位相の保持

我々は，砂時計ネットワークの持つ，位相情報の保持能力，およびデータ圧縮の能力に注目し，ノルム1の制約が実現できるように砂時計モデルを改良する．砂時計モデルは，3層の細胞数がデータ本来の持つ次元数以上であれば，位相保持能力は失わない．そこで，3層の細胞数を（データの持つ次元数)+1とし1次元空間を広げる．さらに，3層で表現されるパターンがノルム1になるように制約をか

ける．こうすることで，中間層でPATONが要求する位相保存の制約を満たす入力データの表現が得られると予想される．具体的には，以下の3つの誤差をBPで学習する．

1. 入力値と5層の出力値の誤差 $\delta e = F5 - I$ を，5層から1層まで伝播．
2. 中間層で表現されるデータのノルム1からの誤差 $\delta h = F3 - F3/ \parallel F3 \parallel$ を3層から1層へ伝播．
3. 3層での出力値 $F3'$ を $F3/ \parallel F3 \parallel$ としたときの5層の出力値 $F5'$ と，入力値 I との誤差 $\delta n = F5' - I$ を5層から3層まで伝播．

4 動的情報の扱い

4.1 音声情報の扱い

2章の式.(2),(10) からわかるように，PATONの認識動作は，各瞬間の外部入力に対して行われる．それゆえ，外部から入力されるデータは瞬間ごとに意味を持つ必要がある．

時間という観点から，入力されるデータは (1) 静止画像のように時間の概念を含まない静的なもの，(2) 音声のように時間方向に意味がある動的なもの，とに分類できる．静的な情報は，静止画像のように，各瞬間に得られる情報に意味がある．ところが，音声における単語のように，動的な情報は各瞬間意味があるのでなく，ある時間幅全体の情報に意味がある．そのため，動的情報を扱うためには，各瞬間に意味をもたせる前処理を行う必要がある．

我々は，この前処理の方法として始点終点フリーのDynamic Programing(DP) を利用する．DPを用いれば，音声の非線型伸縮を考慮し，各時刻での入力音声とテンプレートとして用意された単語との類似度を求めることができる．この類似度を入力に使えば，各時刻ごとに意味ある単語情報をPATONに入力できる．また，P層の細胞数が，用意されたテンプレートの数で決まる利点もある．

4.2 始点終点フリーの DP Matching

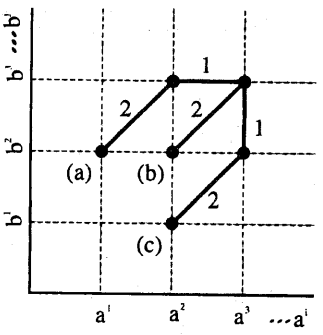

Fig. 2: Local constraints and slope weight for the DP matching.

入力音声 T とテンプレート音声 R を特徴ベクトルの時系列で

$$T = a^1, a^2, \cdots, a^i, \cdots, a^I \quad (13)$$

$$R = b^1, b^2, \cdots, b^j, \cdots, b^J \quad (14)$$

と表現する．a^i は入力音声を区分時間（以下，フレーム）で分けたときの，第 i フレームの特徴量 $a^i = (a_1^i, a_2^i, \cdots, a_M^i)$ である．このとき，TとRの類似度D(T,R)は，TとRの時間関係を変化させ，類似度が最もよくなる場合である．

$$D(T,R) = \min_{F(K)} \frac{\sum_{k=1}^{K} w(k) d(a^{i(k)}, b^{j(k)})}{\sum_{k=1}^{K} w(k)} \quad (15)$$

ここで $F(k)$ は，a^i と b^j の対応関係を表わす関数である．

$$F(K) = (i(k), j(k)) \quad k = 1, 2, \cdots, K \quad (16)$$

$d(a^{i(k)}, b^{j(k)})$ は，$a^{i(k)}$ と $b^{j(k)}$ の距離である．また，$i(k), j(k)$ はそれぞれTとRのk番目の対応点のフレーム番号，$w(k)$ は非負の荷重関数で，k 番目と $k+1$ 番目の対応点間の距離とすることが多い．

$$w(k) = (i(k) - i(k-1)) + (j(k) - j(k-1)) \quad (17)$$

この $D(T,R)$ を求める一つの方法がDPである．DPでは，類似度 $D(T,R)$ の部分和 $g(i(l), j(l))$ が

逐次的に求められることを利用する．

$$
\begin{aligned}
& g(i(l),j(l)) \\
&= \min_{F(l-1)} \left\{ \sum_{k=1}^{l} w(k) d(a^{i(k)}, b^{j(k)}) \right\} \\
&= \min_{f(l-1)} \{ g(i(l-1), j(l-1)) \\
&\qquad + w(l) d(a^{i(l)}, b^{j(l)}) \} \qquad (18)
\end{aligned}
$$

ここまでの議論は，T,R 全体を用いた類似度の話であった．しかし，実環境から入力される音声では，各単語の始点と終点の位置は不明である．そこで DP の計算を行う際，すべてのフレームで単語が始まっていると仮定し，部分和の初期値を $g(m,0) = d(m,0)$ $(m = 1, \cdots, K)$ とする．さらに部分和を求める際，Fig.2 のような傾斜制限をつける [4]．すると部分和 $g(m,J)$ $(m = 1, \cdots, K)$ は，第 m フレームでテンプレート T の単語が終わっていると仮定したときの累積類似度となる．したがって，第 m フレームでの入力音声とテンプレートとの類似度 $D(T(m),R)$ は，パスの長さ $\sum_{k=1}^{K(m)} w(k)$ で正規化して以下のように求まる．

$$
D(T(m),R) = \frac{g(m,J)}{\sum_{l=1}^{K(m)} w(l)} \qquad (19)
$$

ここで，$K(m)$ は第 m フレーム目で単語の入力が終わったと仮定した際の入力音声とテンプレートとの対応点の数である．また，Fig.2 の傾斜制限 [4] を使ったときの類似度の逐次計算は

$$
\begin{aligned}
& g(i(l),j(l)) \\
&= \min \left\{ \begin{array}{l}
g(i(l)-2, j(l)-1) + 2d(a^{i(l)-1}, b^{j(l)}) \\
\qquad + d(a^{i(l)}, b^{j(l)}) \quad \cdots (a) \\
g(i(l)-1, j(l)-1) \\
\qquad + 2d(a^{i(l)}, b^{j(l)}) \quad \cdots (b) \\
g(i(l)-1, j(l)-2) + 2d(a^{i(l)}, b^{j(l)-1}) \\
\qquad + d(a^{i(l)}, b^{j(l)}) \quad \cdots (c)
\end{array} \right. \qquad (20)
\end{aligned}
$$

となる．

Fig.3 に入力音声「右下，色，あか」と用意された 14 種類のテンプレートとの DP の結果を示す．入力音声は，11025Hz でサンプリングされ，ハイパスフィルタにかけられた後，512 点のハニング窓を用い，FFT でパワースペクトラムに変換された．フレームの間隔は 112 点で，隣接するフレーム間には 400 点のオーバーラップがある．結果から 3 つの谷があるのがわかる．それぞれの谷は「右下」「色」「赤」の音声に対応し，各頂点が単語の終わりを示している．

PATON への入力は，各時刻で得られた入力音声とすべてのテンプレートとの類似度を入力ベクトルと考え，砂時計ネットワークで符号化し行う．また，PATON の P 層と S 層の結合係数の学習には，DP の結果で類似度が 50 以下の部分を教師データとして用いた．

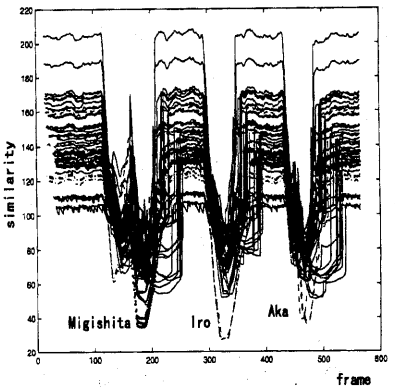

Fig. 3: Result of DP matching between a connected word speech signal "Migishita Iro Aka" and 14 words of template "hidariue", "hidarishita", "migiue", "migishita", "aka", "ao", "kuro", "shiro", "sankaku", "shikaku", "maru", "iro", "katachi" and "nani". Horizontal(vertical) axis represents the frame number(degree of the similarity). Low(high) degree represents high(low) similarity. Total number of the templates which are prepared for the DP matching is 42. Number of templates for each word is three.

5 PATONによる連続単語音声認識

5.1 連続単語音声認識で用いるPATONモデル

Fig.4に連続単語音声認識を行うPATONを示す．PATONは外界から課題実行中変化しない静的な視覚情報と，視覚情報について述べる音声をうけ取る．但し，受け取る音声は,「右下，色，赤」のように単語のられつであるとした．視覚情報は，視野(Visual field)の四隅（左上・左下・右上・右下）にそれぞれ異なる形（三角・四角・丸）と色（赤・青・黒・白）の組み合わせからなる対象が与えられる．PATONへの入力は，視野の四隅に注意を向けることで，対応する場所の形と色の特徴が特徴抽出系で抽出され，砂時計ネットワークを介し行われる．ここで，形の特徴抽出にはテンプレートマッチングを使用し，色の特徴は画素のRGB値をもちいた．一方音声情報は，DPを使いて単語情報に変換された後，砂時計ネットワークを介しPATONに入力される．

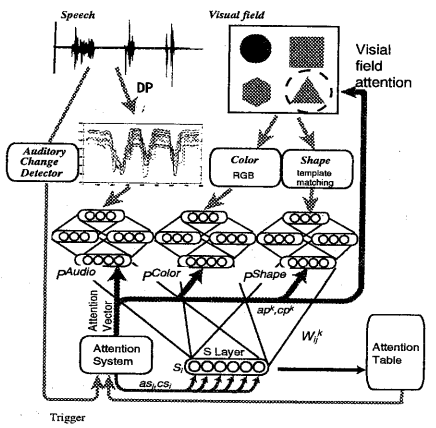

Fig. 4: Architecture of the PATON model that is modified to execute speech recognition task.

5.2 注意ベクトルのふるまい

本課題では注意ベクトルとして，感覚信号の変化を検出して自動的に向くボトムアップの注意と，認識の結果によって意図的に向くトップダウンの注意を用い単語認識を行う．ボトムアップの注意はAuditory change detectorによる音圧の変化の検出によって駆動され，自動的に音声入力領野P^{Audio}への注意ap^{Audio}を増加させる($a_p^{audio}=1$)．トップダウンの注意は，認識行為によって発生する．発生する注意ベクトルは認識結果によって異なり，Attention Table(AT)から認識結果に対応する注意ベクトルが選ばれる．ATには，過去の経験によって得られた認識結果と発生する注意ベクトルの対が記述されている．ただし，本研究ではATの内容はあらかじめ与えられているとした．

本課題における，トップダウンの注意発生に関係する音声入力は「左上」「左下」などの視野の位置を示す単語と,「色」や「形」といった属性を表現する単語である．それぞれに対応するATの記述は，以下のようになっている．視野の位置に関する単語が入力された場合，視野の対応する位置に注意が向き，その位置にある物体の情報がPATONに入力される．同時に色と形領野p^{Color}, p^{Shape}層への注意も強められる(+0.01 for a_p^{Color} and a_p^{Shape})．一方,「色」等の属性を表現する単語が認識された場合，対応する領野への注意が増加する．しかし他の領野への注意は逆に下がるとした．例えば，単語「色」が認識され色領野に注意が向いたとする．この時，逆に「形」領野への注意は下がる(+0.01 for a_p^{Color} and -0.01 for a_p^{Shape})．

5.3 認識実験

Fig.5に「右下 色 赤」と入力した際のシンボル層の出力値を示す．シンボル層では入力のばらつきによっていくつかのシンボルが認識されかかっているが，閾値を適切に設定すれば認識が可能である．

一方で，視覚情報による文脈の効果を使わずに音声情報のみをPATONに入力した結果をFig.6に示

す.「右下」「色」はうまく認識できているが,「赤」は赤シンボルが発火したあとすぐ「三角」シンボルも発火している．これは「SANKAKU」という音声に,「AKA」に類似した部分が含まれているためで，両者の区別は容易ではない．

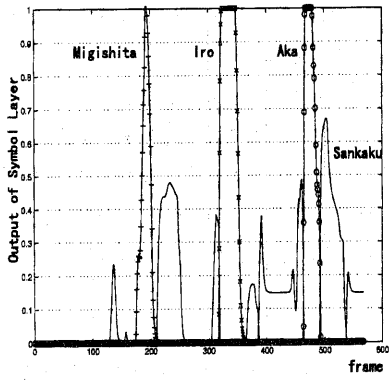

Fig. 5: Result of the speech recognition task with speech and visual input. Vertical(horizontal) axis represent the output of neuron in the symbol layer(frame number). For this simulation, we introduce the threshold(θ) for activation of symbol neuron. Output S_i is given by $S_i = a_{s_i}\phi((s_i-\theta)*1/(1-\theta))$. Parameters are set as follows: $\theta = 0.95, C_1 = 0.01, C_2 = 1.0, C_3 = 1.0, C_4 = 1.0, \tau = 1.0, \Delta t = 0.02, R_{tt'} = 50$, $c_p^k = 1$ for all k , $c_{s_i} = a_{s_i} = 1$ for all i , a_p^k increase or decrease according to recognition of word and changing of auditory pressure.

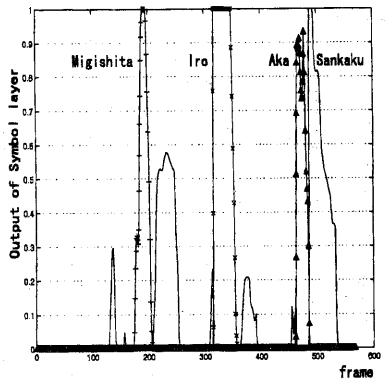

Fig. 6: Result of the speech recognition task with speech input only. Parameters are set as follows: $\theta = 0.95, C_1 = 0.05, C_2 = 1.0, C_3 = 1.0, C_4 = 2.0, \tau = 1.0, \Delta t = 0.02, R_{tt'} = 50$, $c_p^{Audio} = 1, c_p^{Color} = c_p^{Shape} = 0, c_{s_i} = a_{s_i} = 1$ for all i , $a_p^{Audio} = 1, a_p^{Color} = a_p^{Shape} = 0$.

6 考察

6.1 PATONによる状況依存性音声認識

単語音声などの時系列パターンの認識では，パターン系列の重なりのため，類似のパターンが次々と認識される．これは空間的なパターンの認識においても同様である．この問題に対して，PATONは記憶や知識といった別の情報で文脈を与え，認識候補を事前に絞り込むことで認識率を上げている．

この目的のために，我々は二種類の注意を想定した．一つはボトムアップの注意である．この注意により，各属性領野での変化が検出され注意が向けられる．その結果，変化があってより重要と思われる情報の影響力が増大する．もう一方は，認識結果に対応し発生するトップダウンの注意である．発生する注意の内容は，過去の経験に基づいて決まり，次に行われるべき認識・連想などの準備が自動的に進行する．PATONでは，これらの注意をうまく使うことで，マルチモーダルな情報を状況に合わせ選択し認識を行っている．

6.2 単語音声認識率向上の別方法

本稿の実験では，PATONの入力にDPの結果の各時刻の類似度を使った．ところが，Fig. 3の結果からわかるように，単語入力に対応する部分は深い谷になっている．そこで各時刻におけるPATONへの入力をその時刻のDPの結果と前後25フレームおきに2時刻づつ，計5時刻分のデータを入力に使った．Fig. 7にその結果を示す．図からわかるように，本実験に用いたデータでは認識結果が飛躍的に改善する．この方式の評価には多量のデータによ

る検証が必要である.

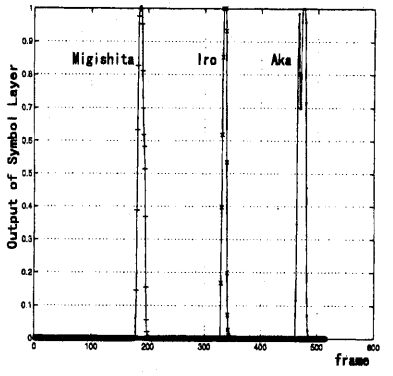

Fig. 7: Result of the speech recognition task with 5 points of DP. Parameters are same as previous(Fig. 6) except for $C_1 = 0.1$.

6.3 脳における言語処理について

言語は抽象的なシンボルを伝達できる道具と考えられ，従来の人工知能研究では記号処理であると考えられてきた[6]. また一方，PATONの動作も状態遷移機械であることが示されてきており両者の整合性はよい. ところが，本稿の実験は，言語の理解とは単なる記号処理ではないことも示唆している. 例えば，名詞のように事物をあらわす語の認識は，対応する感覚記憶を想起させ以後の認識に影響を与える. また，形容詞のように属性を指示する語（例えば「色」）が認識されると，指示された属性が活性化され記憶検索の制約を発生する. このように認識によって発生する注意は，脳という器官での情報の処理動作に対する制約となっており，認識過程に対する文脈的な効果を発生する. このような単語(シンボル)情報と感覚情報との相互作用は従来の言語処理とは異なる点であり，言語理解も脳の動作まで考慮したモデル化が必要であるように思われる. この点については，扱える音声単語数やPATONの記憶の数を増やし（すなわち状態遷移の状態数と入力を増し），多くの状況にフレキシブルに対応できる言語認識システムとしての可能性を探ってみたい.

7 まとめと課題

本稿では，状況に依存した認識が可能な連想記憶モデルPATONを用い，単語連続音声認識の課題を行った. その結果，従来のDPでは困難な音声の重なりにより生じる誤認識の解消に，PATONの持つ状況に依存したマルチモーダル情報の選択機構が有効であることを示した.

今後の課題は，調音結合を含むような連続音声への適用と，逐次的な情報選択を実現する注意ベクトルの学習である. 逐次的な情報選択の学習を行うことで，外界で起こる一種のルールを抽出し，予測，推論といった機能を実現し音声認識の認識率の向上が期待される.

謝辞 本研究の一部は科学技術振興事業団さきがけ研究21のプロジェクトとして行われた. 音声・画像のデータ処理を手伝ってくれた卒論生の及川典泰君，本間大策君に感謝する.

参考文献

[1] 今井聖 (1996):音声信号処理，森北出版.

[2] 望月彰子, 大森隆司 (1996):PATON:文脈依存性を表現する動的神経回路モデル，日本神経回路学会誌, Vol.3, pp.81–89.

[3] 大森隆司 (1995):階層的記憶のモデルとニューラルネットワーク，システム/制御/情報, Vol.39, No.8, pp.363–368.

[4] 伊藤, 木山, 小島, 関, 岡 (1996):時系列標準パターンの任意区間によるスポティングのための Reference Interval-free 連続 DP(RIFCDP), 電子情報通信学会誌 (D-2), Vol.j79, No.9, pp.1474–1483.

[5] 野田 五十樹 (1994):過負荷学習法を用いた恒等写像学習による内部表現獲得. 信学技報, NC94-34, pp.15–22.

[6] Stuart Russell and Peter Norvig 著 古川 康一 監訳 (1997): エージェントアプローチ人工知能，共立出版社.