TR－A－0071

ニューラルネットワークを用いた手書き文字認識

Handwritten Character Recognition besed on a Neural Network Approach

森　吉弘
Yoshihiro　MORI

1990. 2. 28

ATR視聴覚機構研究所

〒619-02 京都府相楽郡精華町乾谷　☎07749-5-1411

ATR Auditory and Visual Perception Research Laboratories

Inuidani, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-7749-5-1411
Facsimile: +81-7749-5-1408
Telex: 5452-516 ATR J

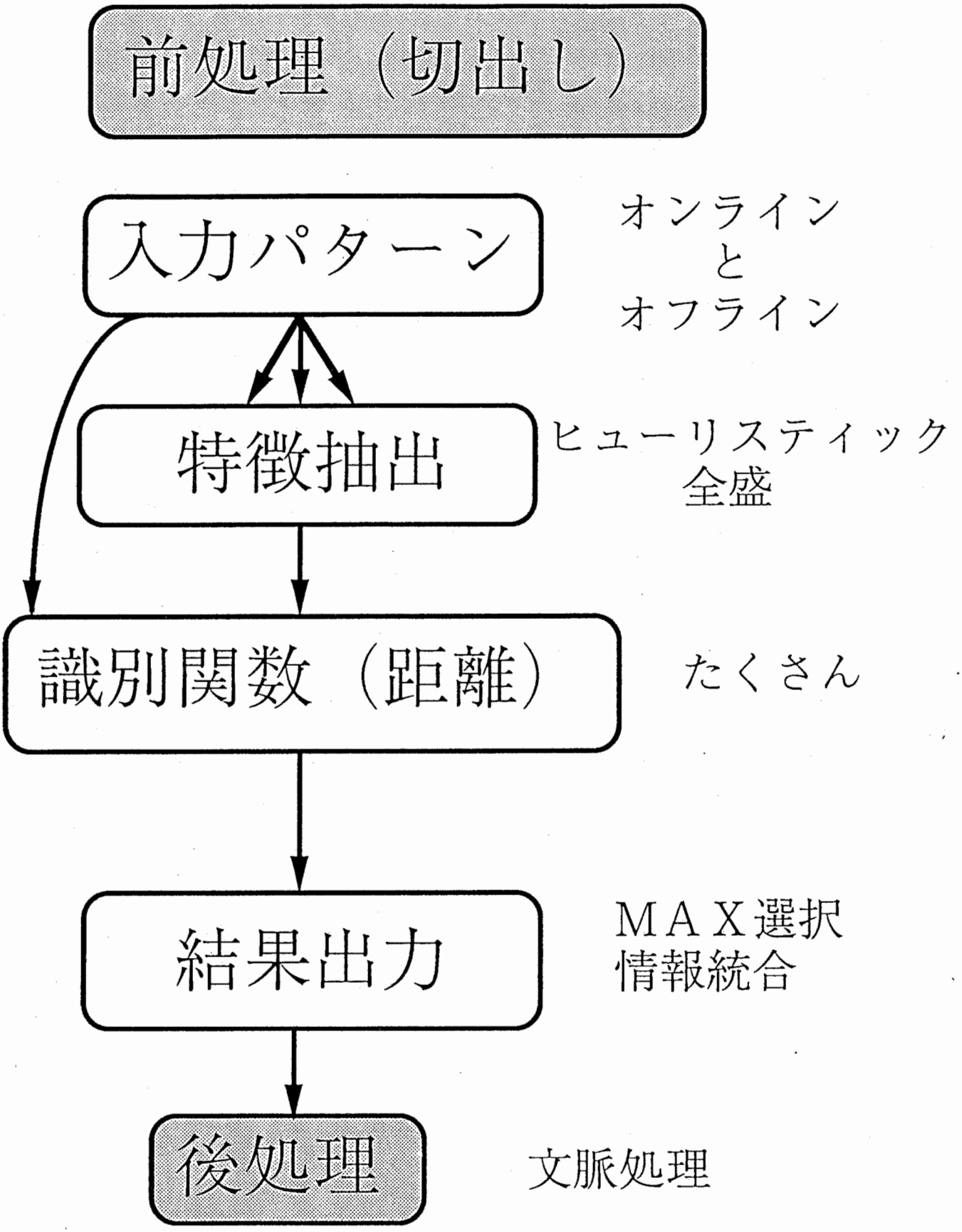
© ㈱ATR視聴覚機構研究所

文字認識

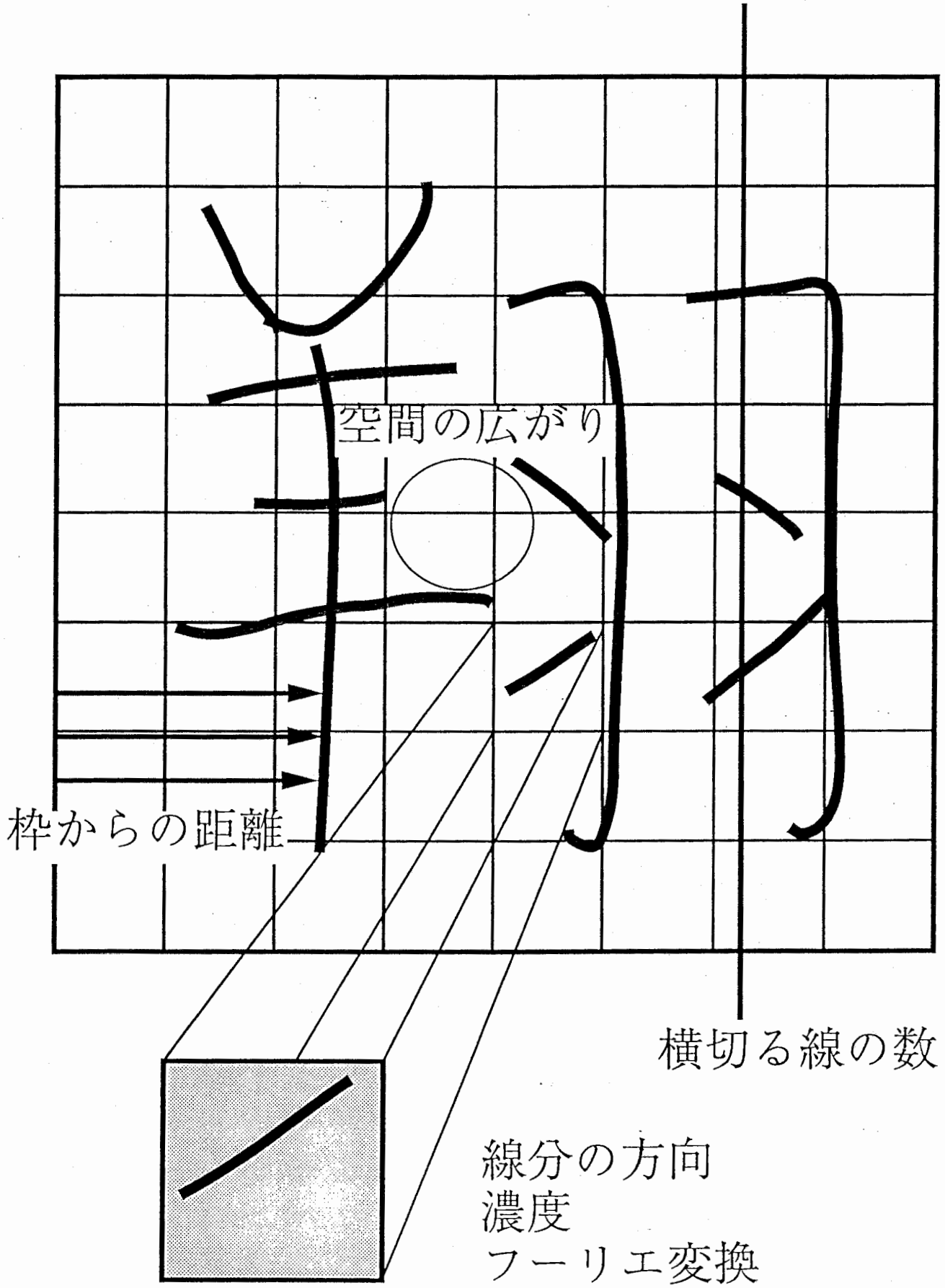
１９３０年頃	光学式テンプレートマッチング ＡＴ＆Ｔなど
１９７０年代	単一印刷文字認識
１９８０年代	マルチフォント認識
１９８０後半	オンライン手書き漢字認識 手書き漢字認識

字種の増加
変動の増加

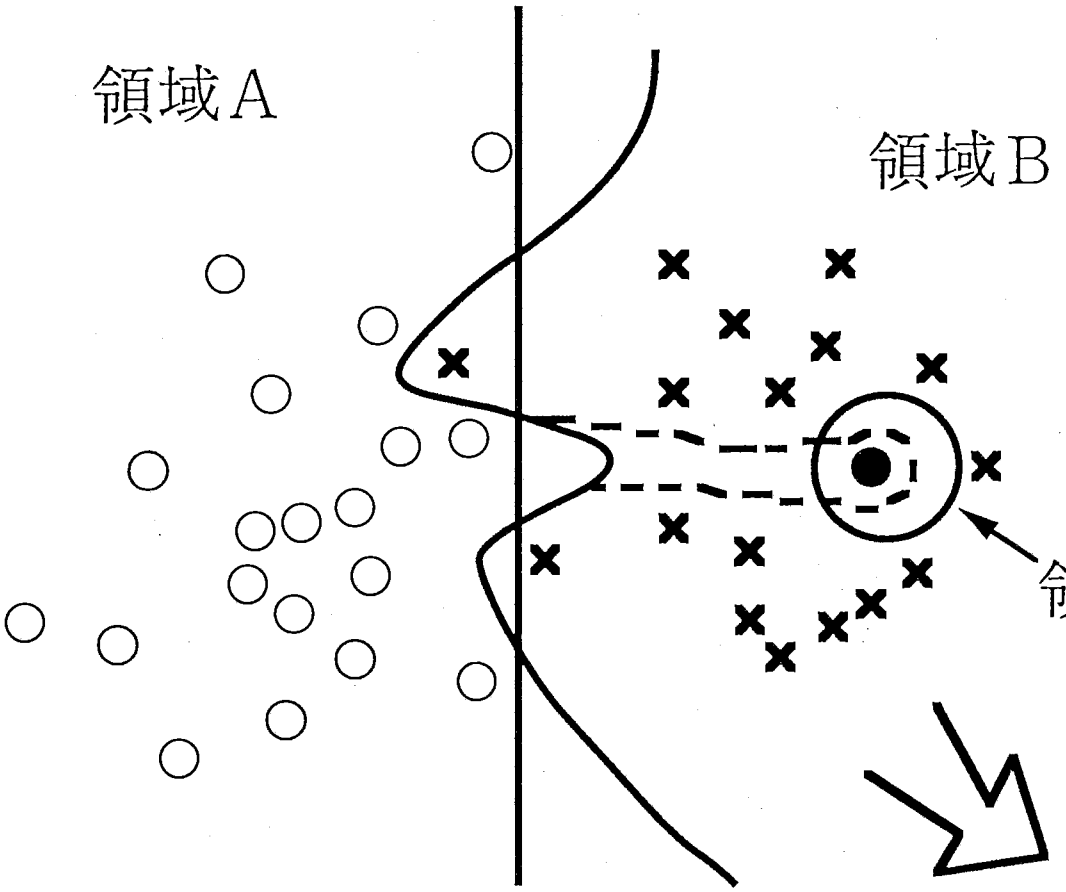
文字認識システム


前処理（切出し）
入力パターン
オンライン
と
オフライン
特徴抽出
ヒューリスティック
全盛
識別関数（距離）
たくさん
結果出力
ＭＡＸ選択
情報統合
後処理
文脈処理

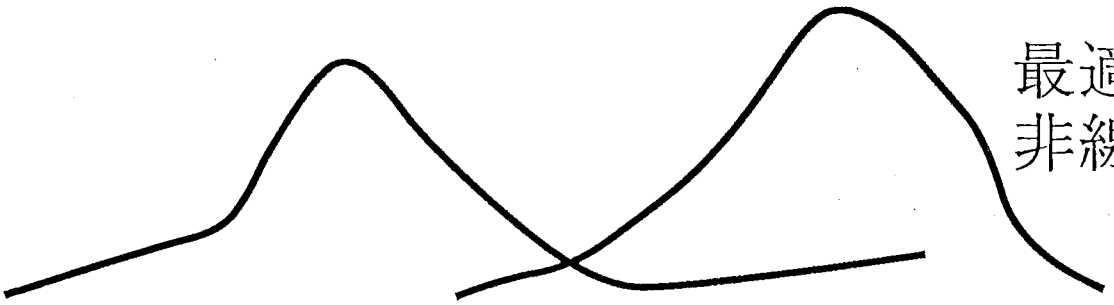
特徴抽出



識別関数



パターン
空間の変換



最適な軸
非線形変換

統計的決定論（ベイズ等）
非線形分離ｏｒ線形分離

ヒューリスティク全盛

研究者の数だけ方式がある。

ＮＴＴ（梅田、増田、萩田、若原．．．）
電総研（山本、岡．．．）
名大，農工大，阪大
東芝＆日電（郵便番号読取り）
松下（手書きワープロ　割りと売れた）
東京三洋（手書きＯＣＲ　ＣＰＵ３１個）

あらて

ソニー（ファジー）
三重大（統計量）
松下、日電、東芝（ニューラルネット）
三菱（光）

それなりの結果はでます。
手書きで認識率９０％前後
活字で　　９８％前後

今、

・特徴抽出手法の行き詰まり
　ヒューリスティックの限界
　（サーベイすると必ずある）

　プリミティブナな特徴＆統合

・識別関数の行き詰まり
　統計的手法の限界
　（分布が未知である
　　膨大なデータが必要）

　少ないデータから真の分布を推定
　脳の記憶方式と関連

↓

ニューラルネットに期待しよう

認識実験１

［目的］

・ニューラルネットの文字認識への応用の可能性を探る。

［内容］

・類似手書き漢字（諭輸輪論）識別

［結果］

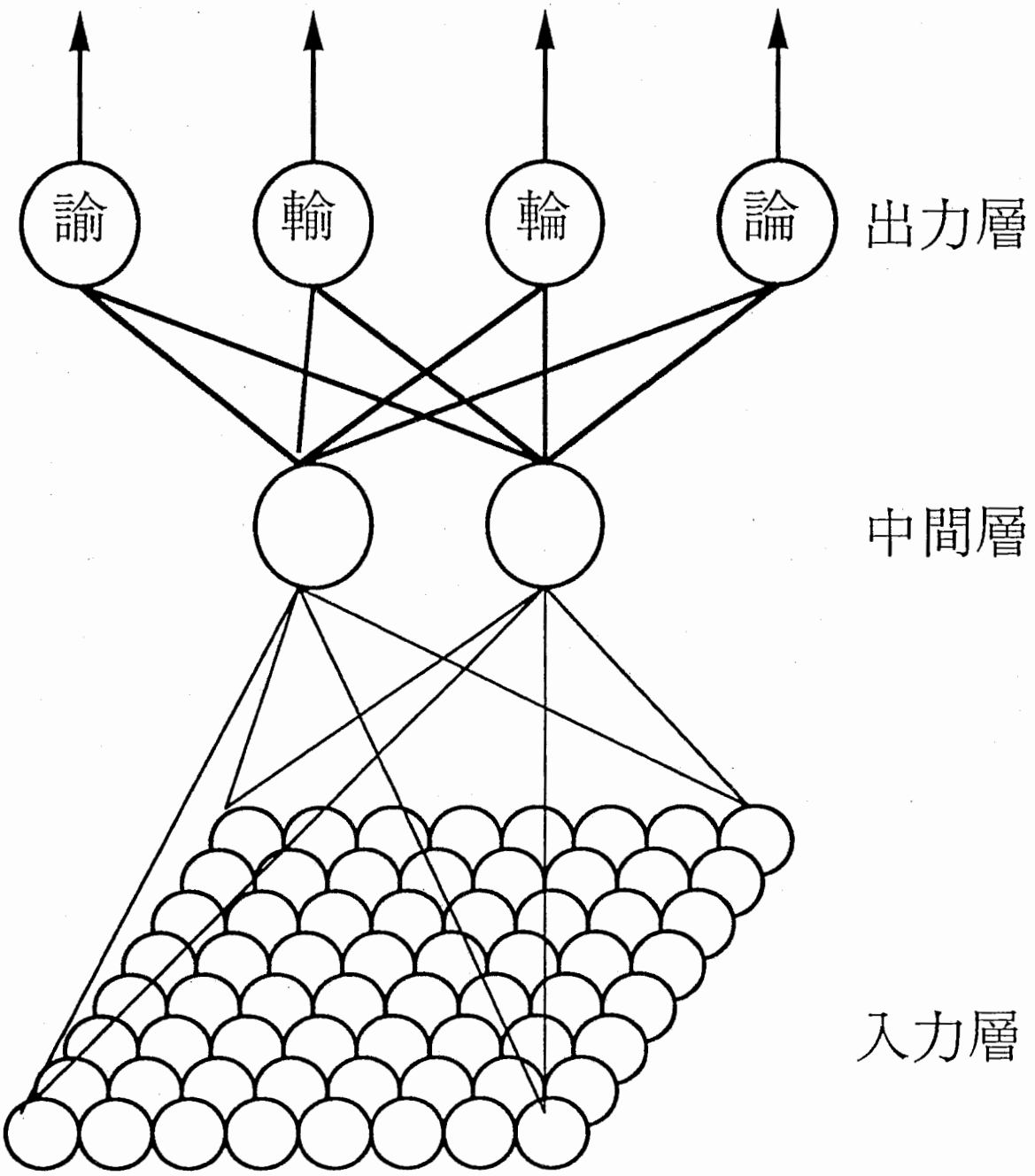
・学習サンプルに対する認識率がほぼ１００％となり、認識対象が学習サンプルであるような応用（印刷文字認識）には非常に適していることが確認された。

・未知サンプルに対する認識率も従来手法を上回った。

認識率	
ニューラルネット	９３％
従来手法	７２％

・類似漢字を識別するために、”へん”と”つくり”にわけて内部表現していた。これは、ニューラルネットが漢字が階層的に構成されるという概念を学習したともいえる。このことが、高認識率を可能としている。

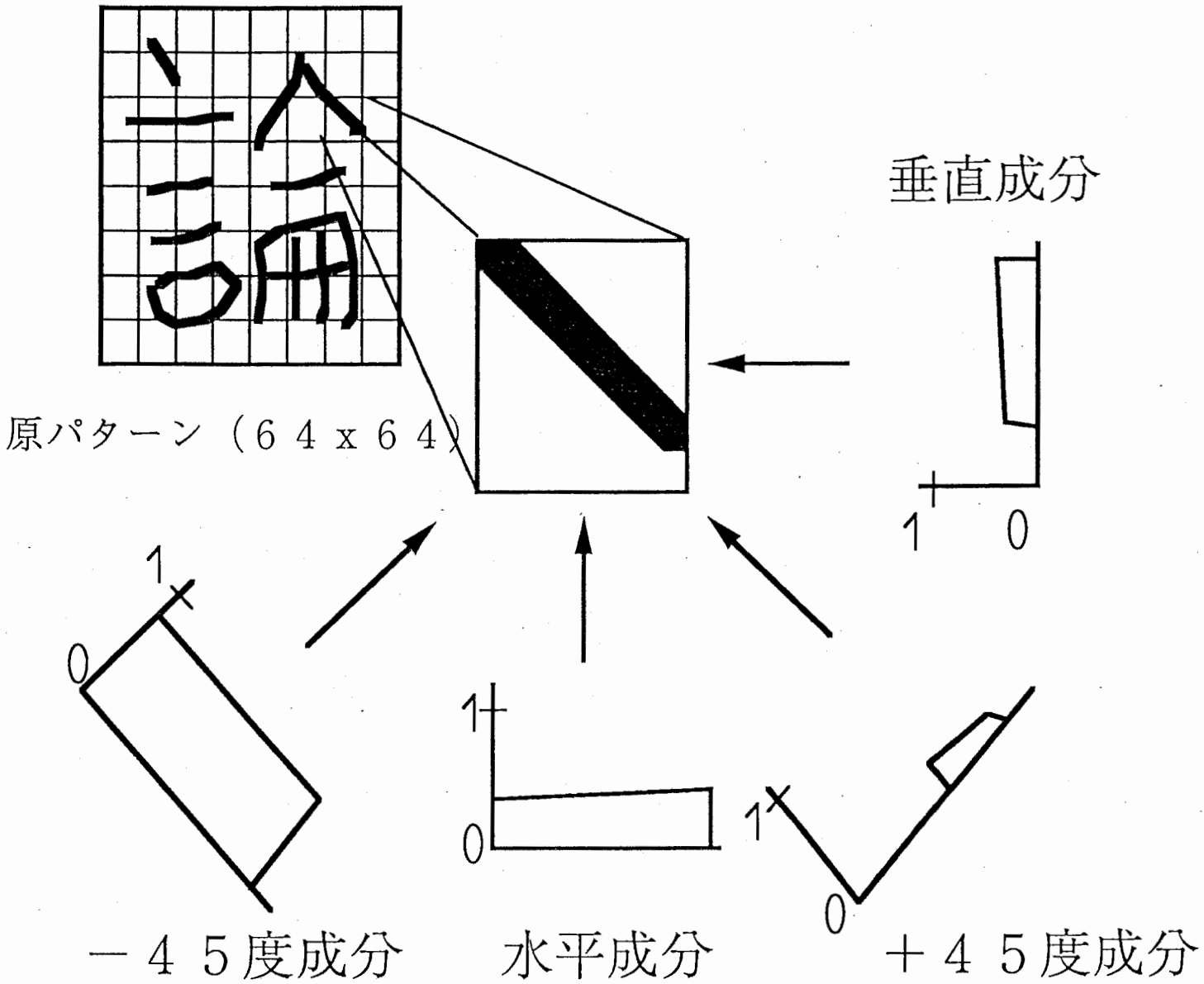
・フィードフォワード型階層神経回路モデル



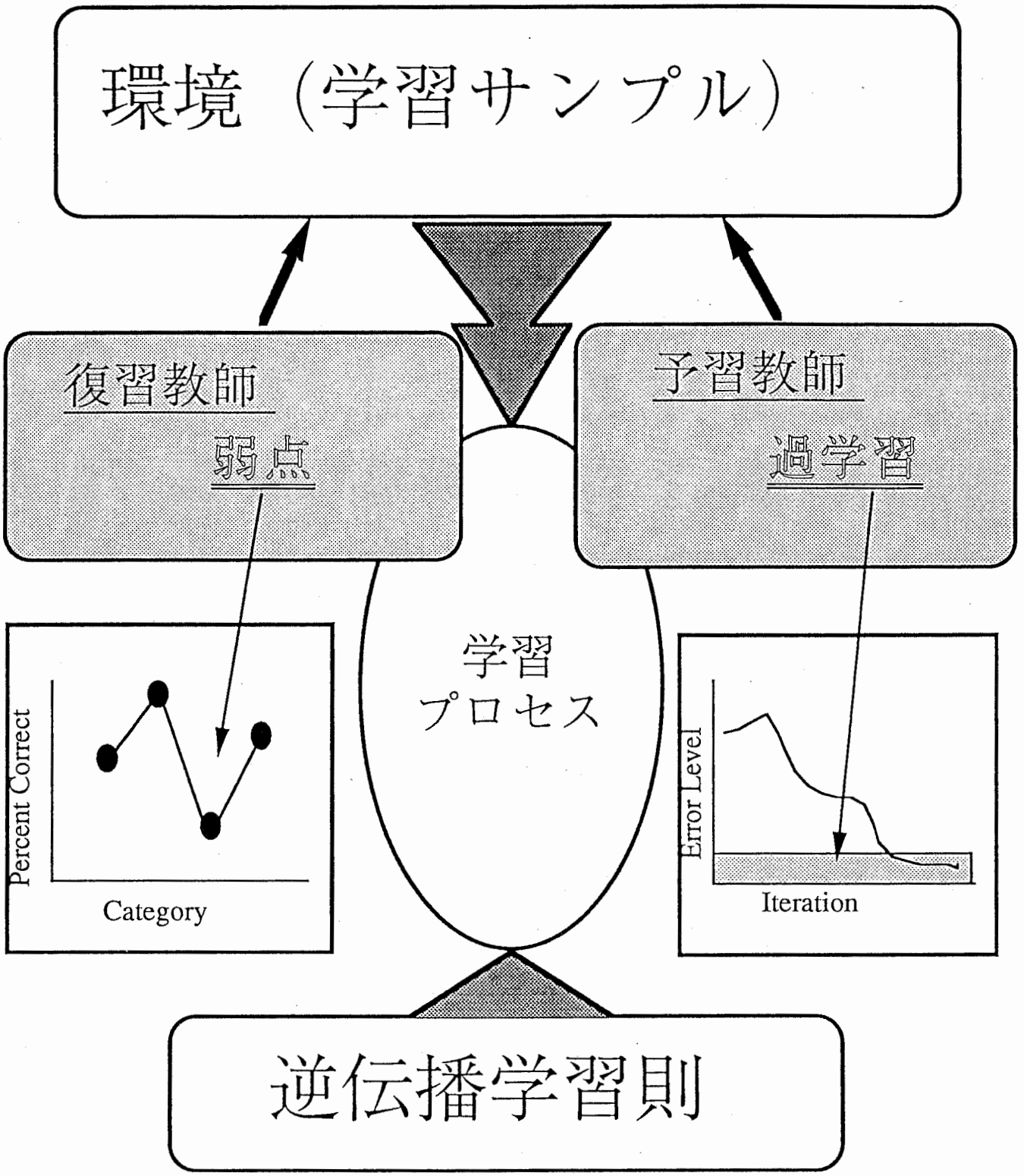
・手書きサンプル２００人分（ＥＴＬ９ｂ）

学習用サンプル　　テストサンプル

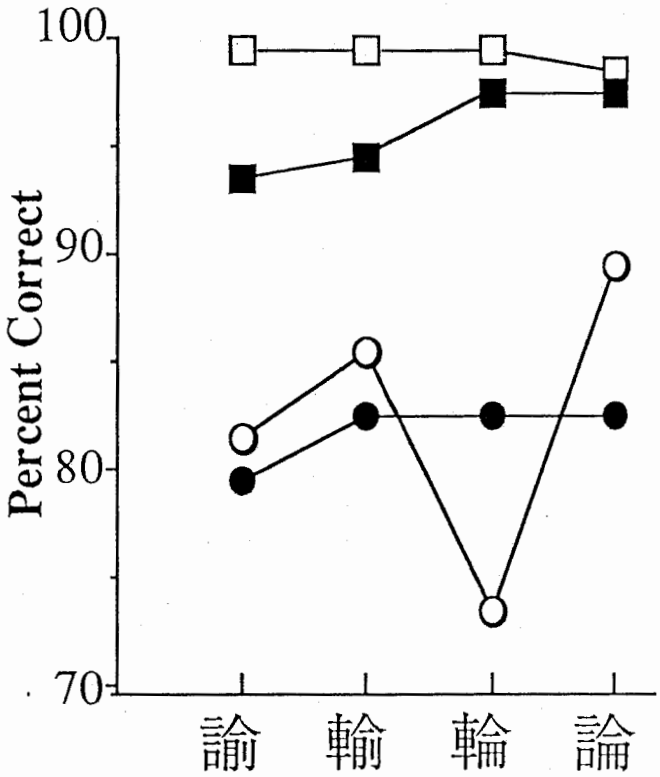
・ＬＤＣＤ特徴（ＮＴＴ　萩田氏）

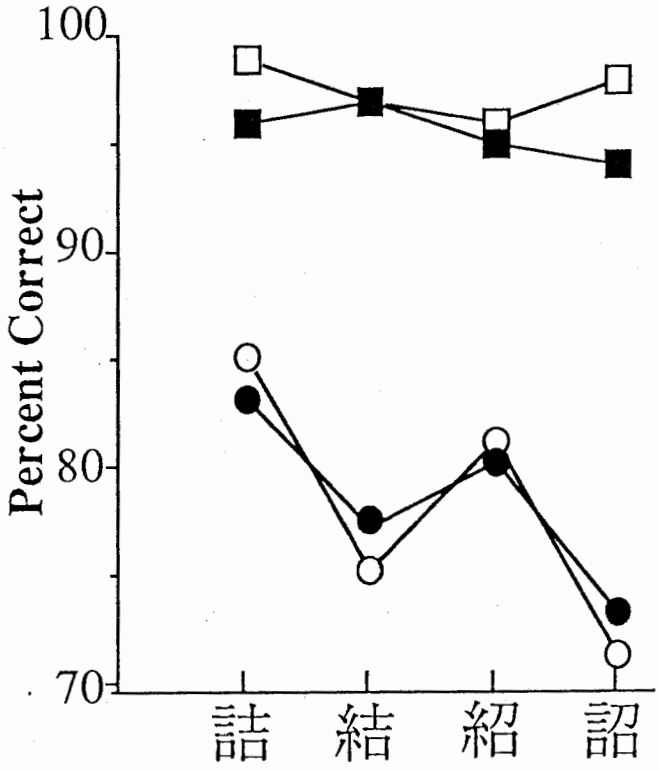


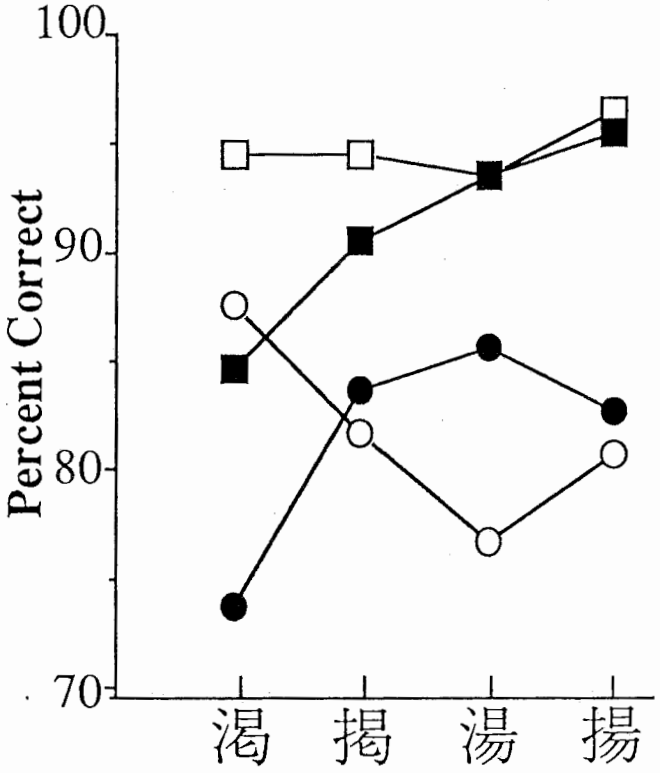
・学習効率を改善するために付加した
２つの新しい教師

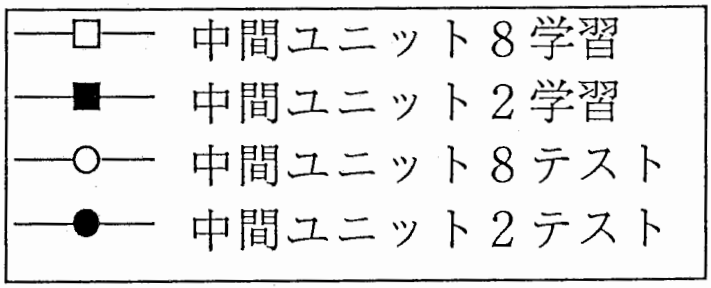


・中間層のユニット数と認識率（メッシュ特徴）

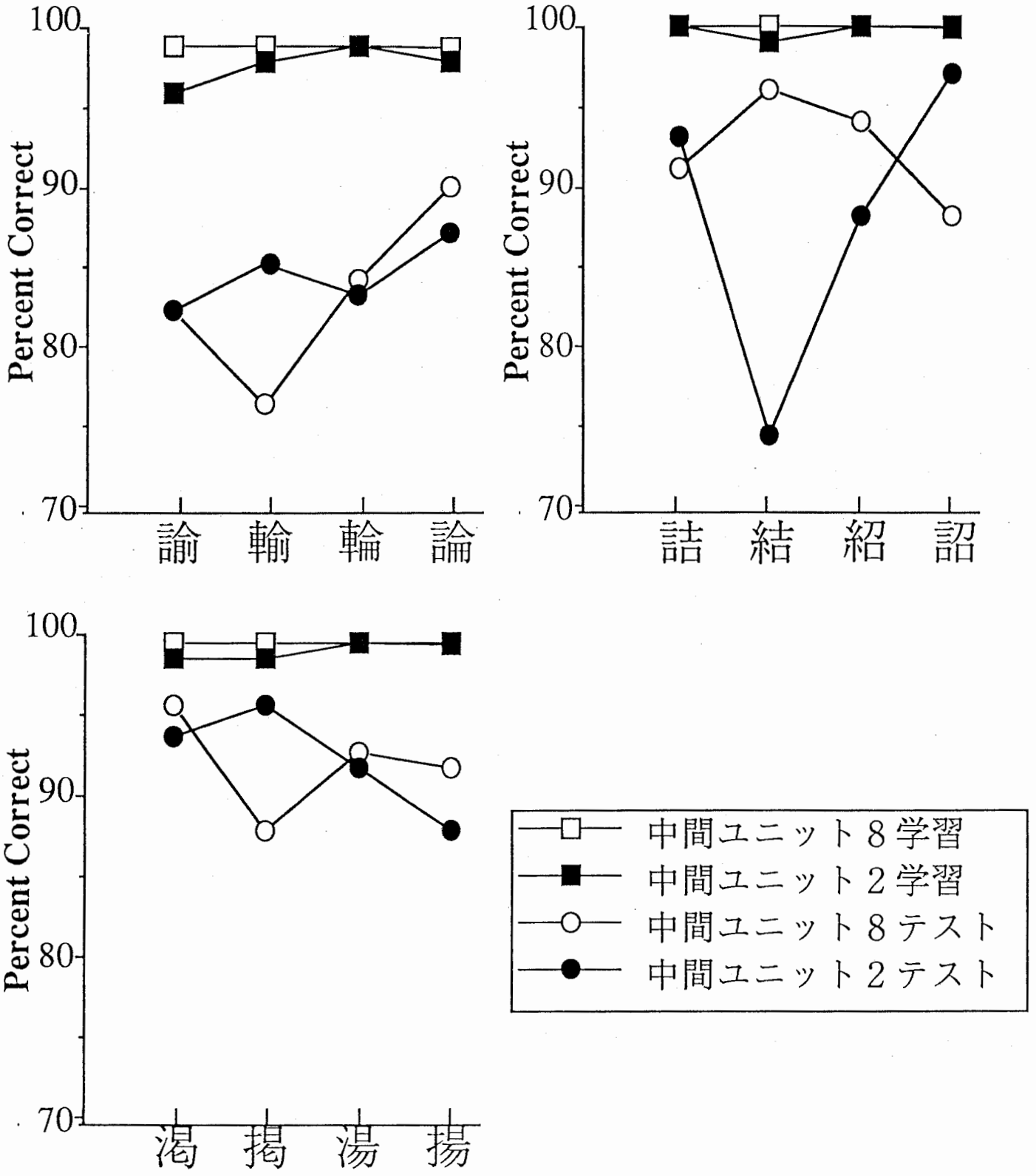




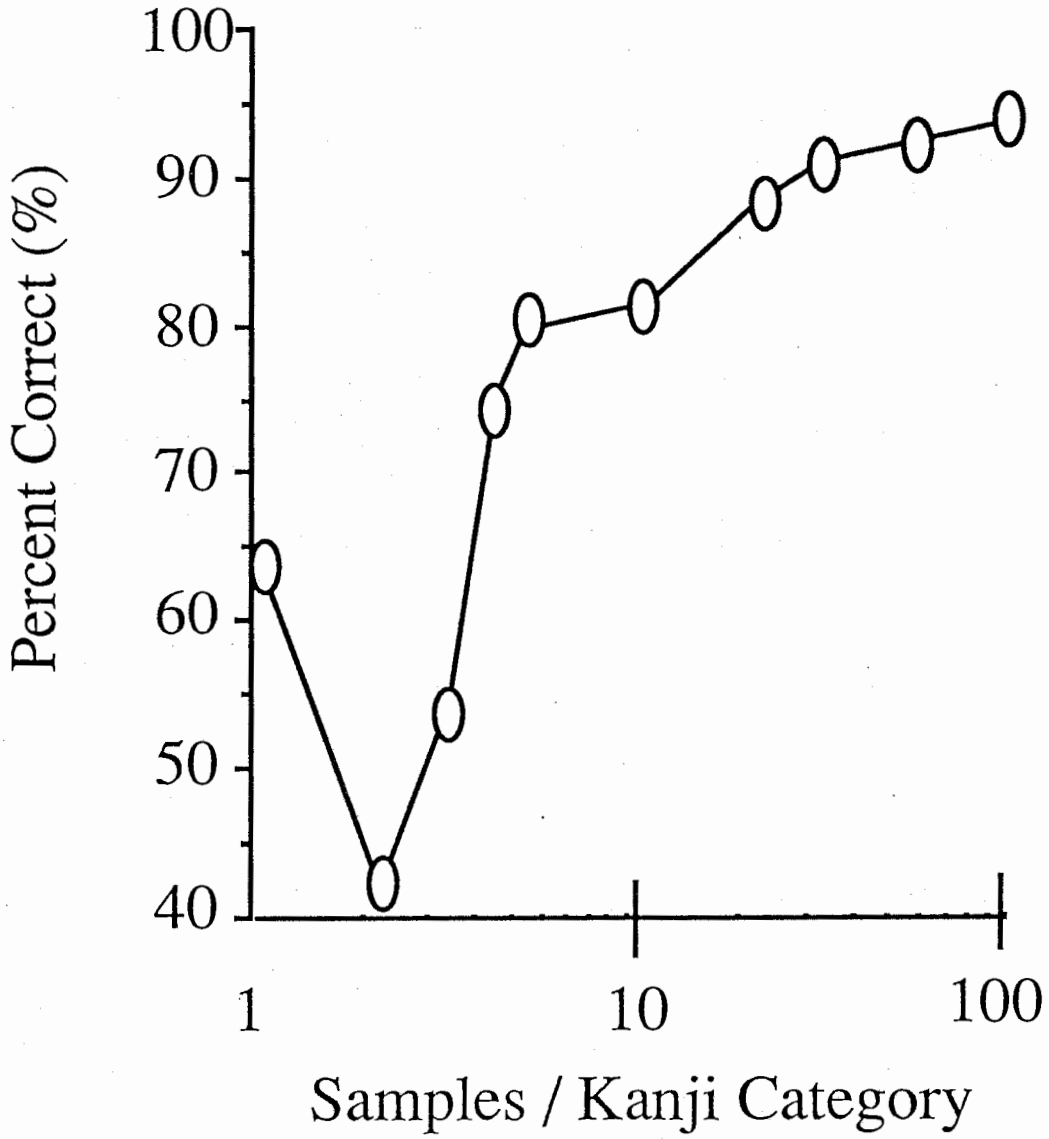




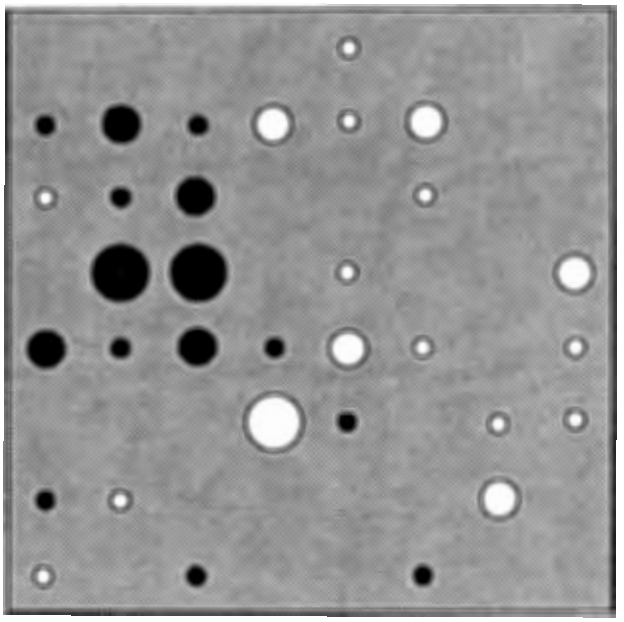
・中間層のユニット数と認識率（ＬＤＣＤ特徴）



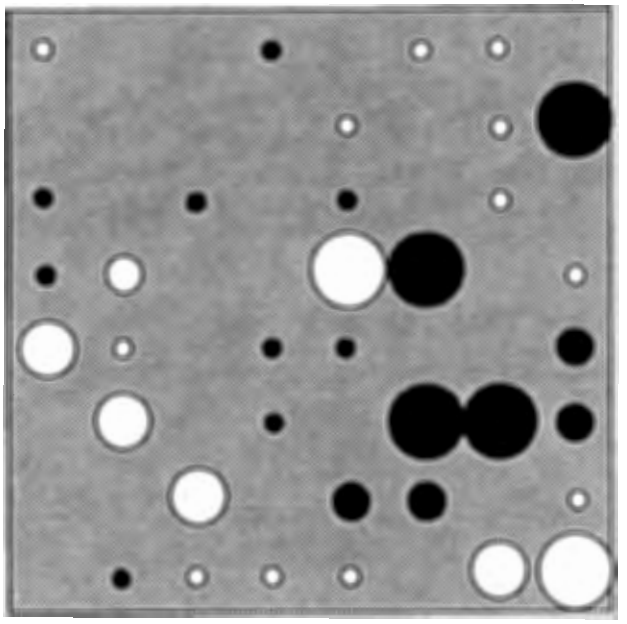
・サンプル数と認識率


100
90
80
70
60
50
40
Percent Correct (%)
1
10
100
Samples / Kanji Category

・類似漢字認識実験（輸諭輪論）時に学習により獲得された入力層から各中間層への接続重み（内部表現）

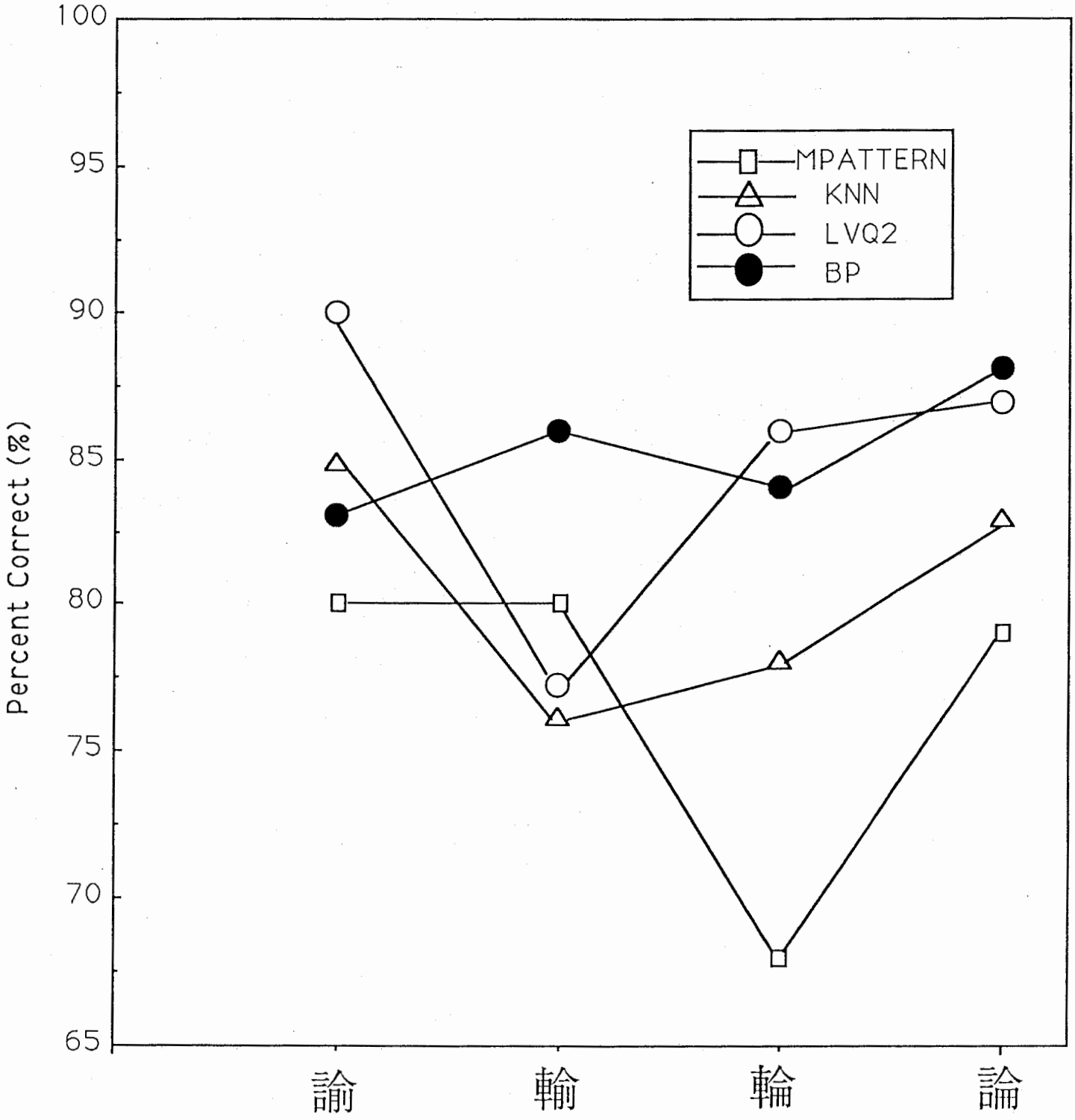


隠れユニット１
（偏識別ユニット）



隠れユニット２
（つくり識別ユニット）

● 興奮性結合
○ 抑制性結合


100
95
90
85
80
75
70
65
Percent Correct (%)
MPATTERN
KNN
LVQ2
BP
諭
輸
輪
論

ＬＶＱ、ＫＮＮとの比較

・現在利用可能なワークステーションのシミュレーション能力

シミュレーション能力

Machine	MCPS†
VAX8600(Ultrix)	0.05
SUN4/260	0.08
HP9000/835	0.15
IRIS/4d-60	0.19
Titan(r2000)	0.61
Alliant FX-80	1.85
CM-1(16k)††	2.60
NCUBE(512node)	6.00
CM-2(64k)†††	40.00

† Million Connection Per Second
†† Speed of CM-1 is from INNS88 2-148
††† Speed of CM-2 is from NIPS89

1990.2.9
ATR Auditory and Visual Perception Research Laboratories

・学習時間と認識時間

学習に必要な時間の計算例

・認識対象は２００カテゴリーである。
・各カテゴリーは１００個のサンプルを持っている。
・学習の収束にまでに各サンプルを１０００回呈示した。
・認識モデルは１００００００本の接続を持っている。
・学習に使用するマシンの能力は１ＭＣＰＳとする。

学習時間

＝ 200*100*1000*100000/1000000(sec)
＝ 2000000(sec) ＝ ２３日

・フォワード計算はバックワード計算の３倍速いと考えられるので。

認識時間

＝ 100000/1000000/4 (sec)
＝ 0.025 (sec)

認識実験２

［目的］

・大規模なニューラルネットの構築方法を探る（Ｔｏｙ－Ｐｒｏｂｌｅｍからの脱出）

［内容］

・手書きひらがな認識（７１カテゴリ）

［結果］

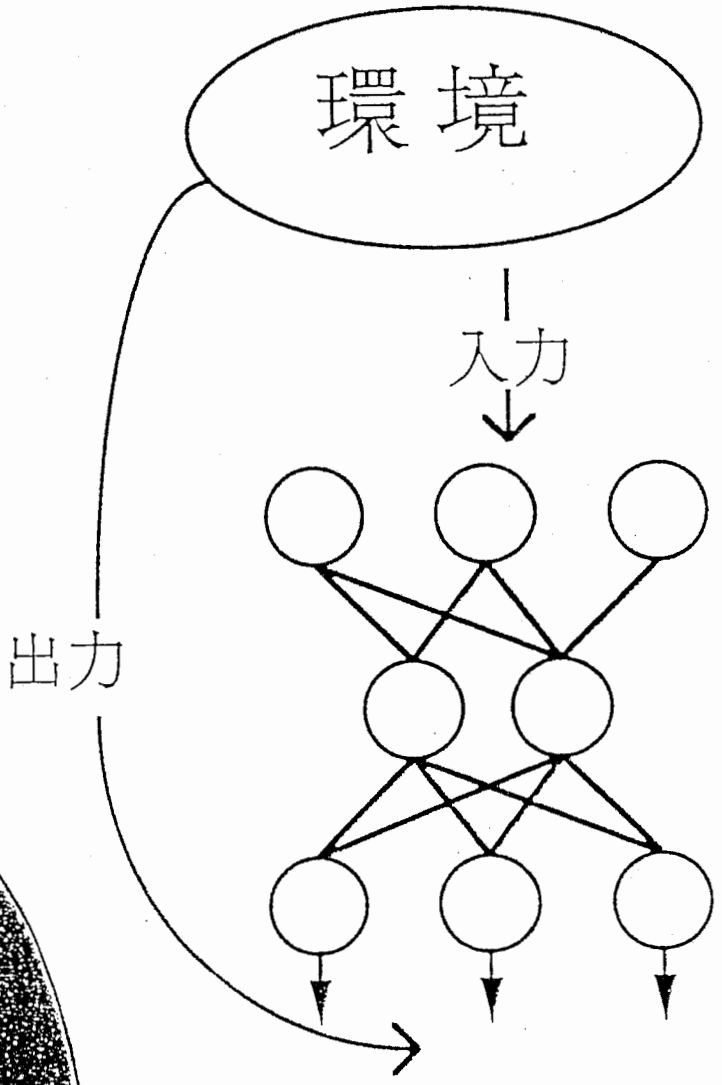
・小規模なネットワークを階層的に統合して大規模なネットワークを構成する方法を提案し、実験の結果、小規模なネットワークの平均認識率を大規模なネットワークが上回った。（９８．６＞９７．２　学習サンプル）

・小規模なネットワークでの情報の統合を容易にすることを目的として”その他セル”を考案し小規模なネットワークの出力層の一ユニットとして付加した。ただし、情報の統合に失敗したカテゴリも少数だが存在する。

認識率	テスト（学習）
ニューラルネット	８６（９３）％
従来手法	７４（７３）％

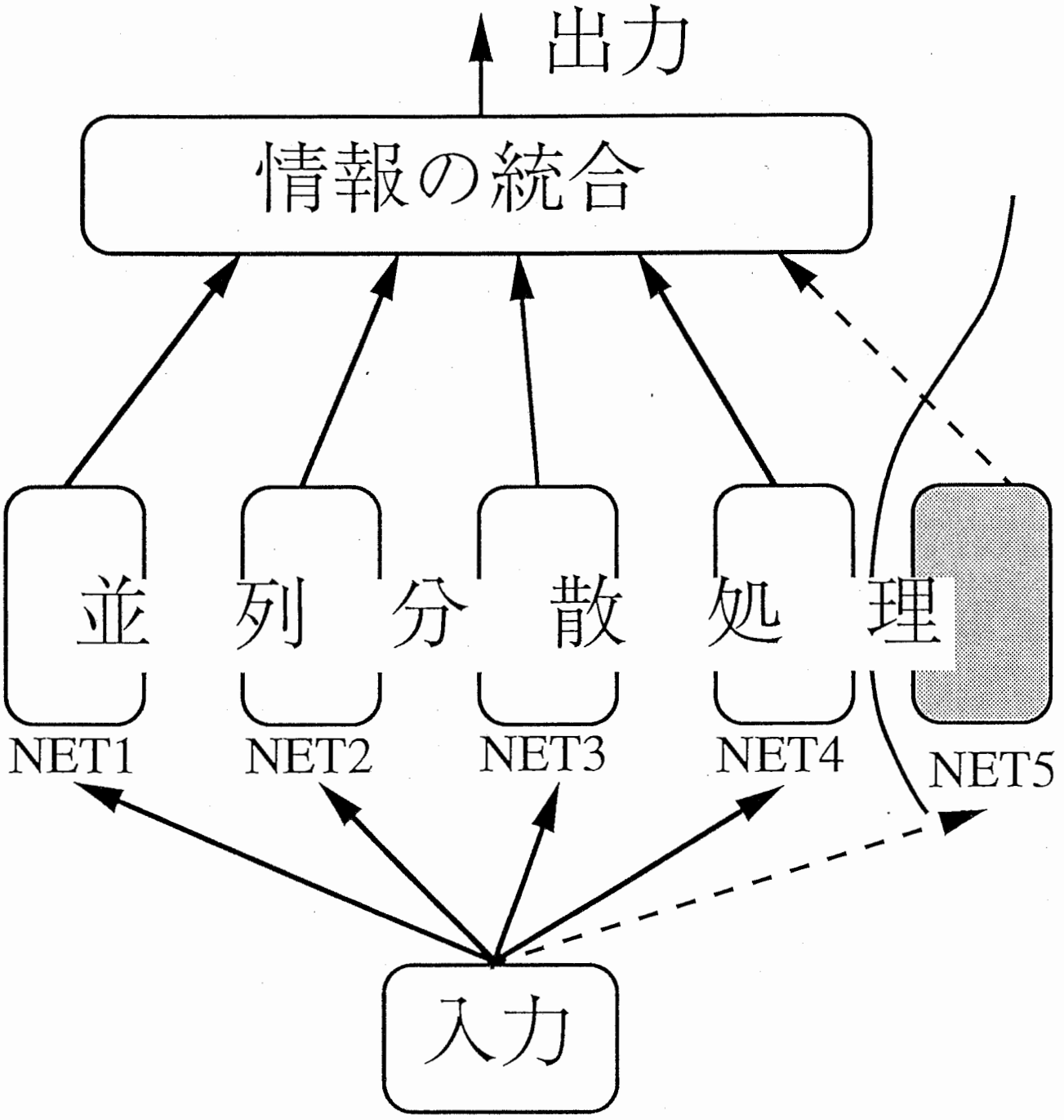
ニューラルネットワークは錬金術ではない

~~入力と出力の組さえあたえれば,すばらしい文字認識の装置が完成する~~

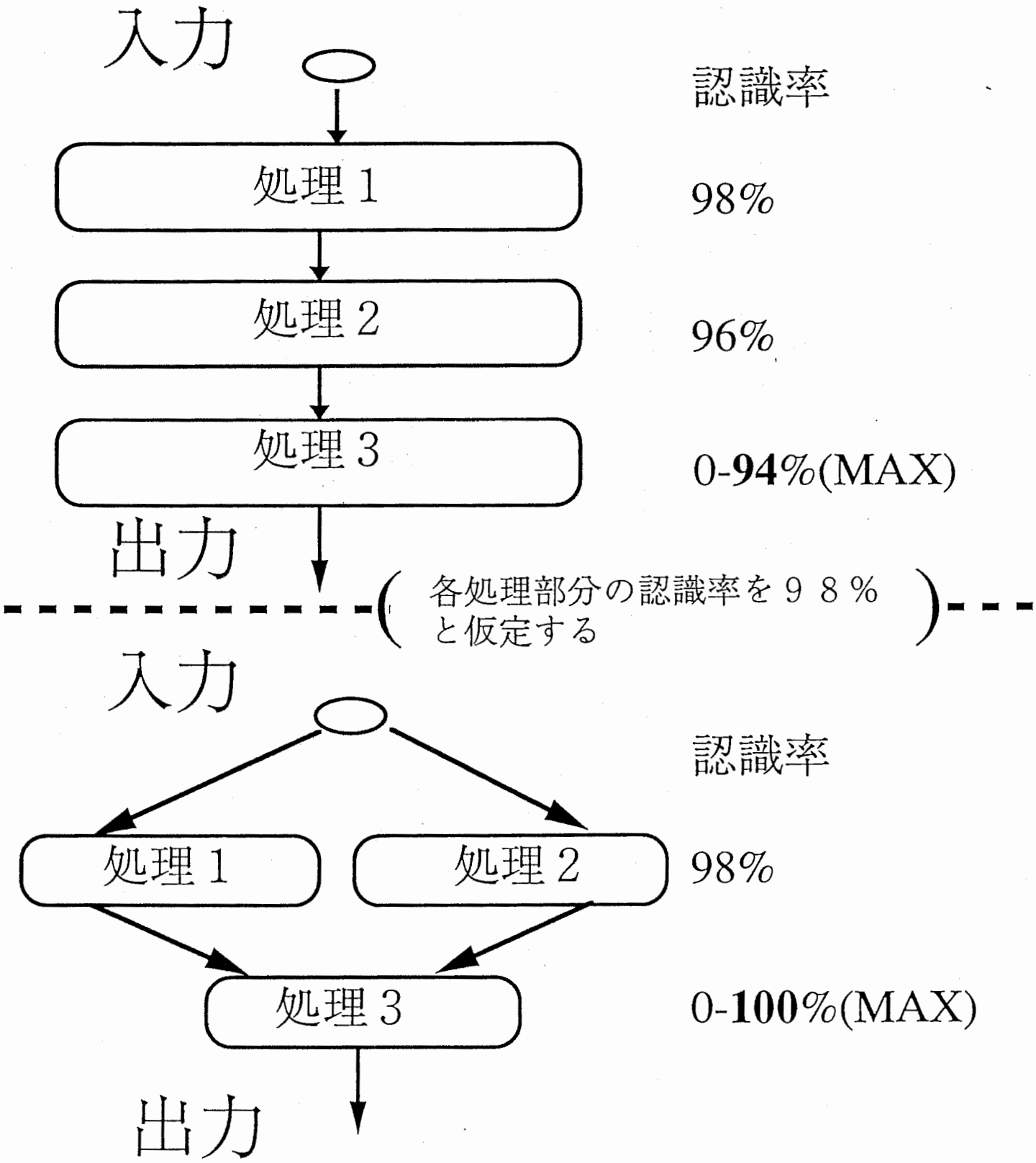


・ニューラルネットの文字認識への応用の可能性

（1）並列性
　　並列処理できるか？
（2）統合性
　　異種情報を統一的に扱えるか？
（3）加法性
　　既存部分を利用できるか？



・情報処理ユニットの統合方法



正しい内部表現の獲得のために

- 文字の変形
  (1) 位置ずれ
  (2) 大きさ
  (3) 回転
  (4) 線幅
  (5) .....

- 解決方法

(1) 正規化手法(簡単)
(2) 正規化手法(簡単)
(3) 回転補正ネットワーク(難)
　　回転不変特徴 (?)
(4) 文字幅不変特徴(簡単)
　　細線化(続け字?)

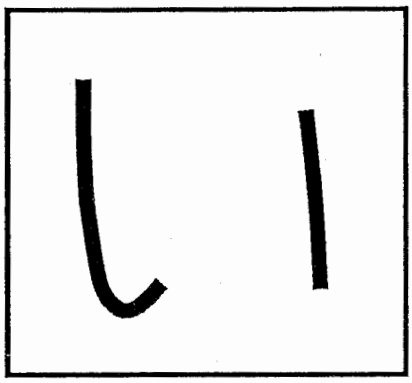
・手書き文字を高認識率で認識するには

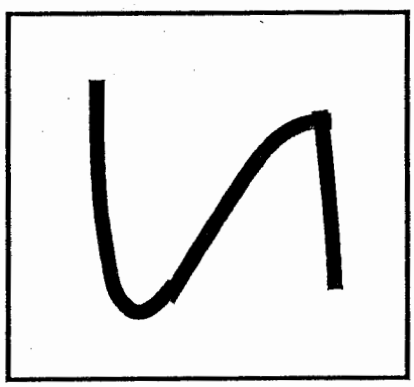
特徴的表現（般化＆例外処理）

　あるクラスを他の無限に多くのクラスの中から識別する表現。

（般化例　ひらがなの”い”は垂直方向に長いストロークが２本ある。）

（例外処理例　上記の２本のストロークは連続することがある。そのため、垂直方向のストロークが２本検出できないことがある。）

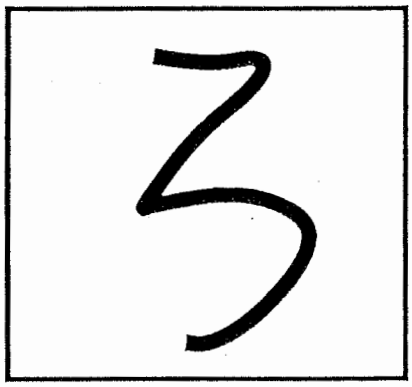


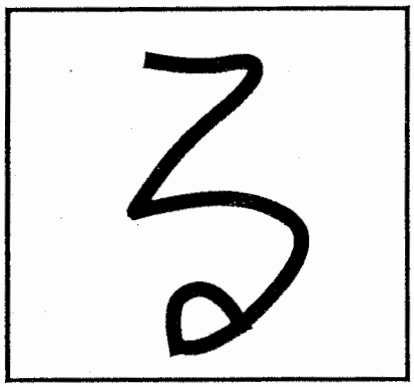


識別的表現

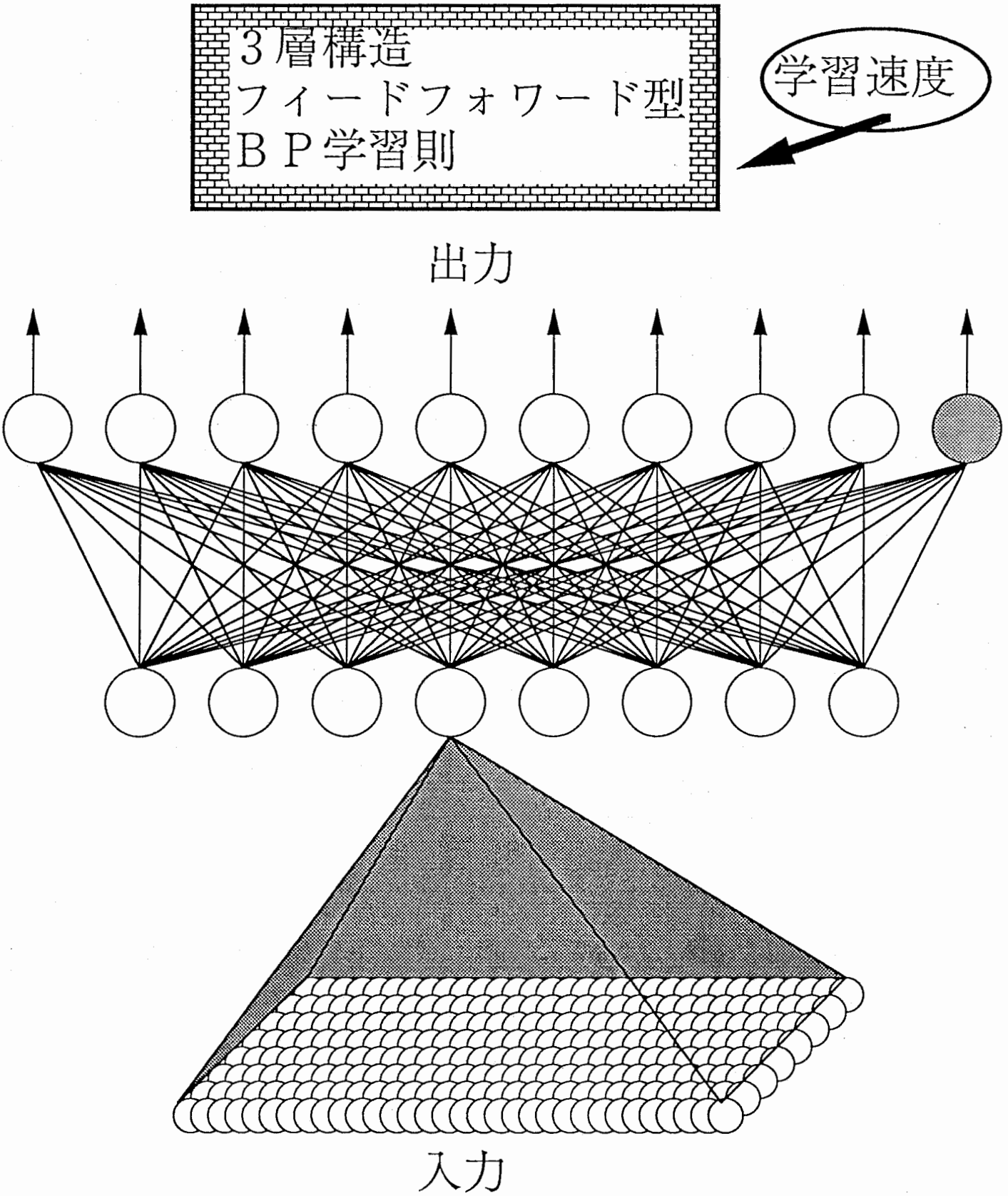
　固定したクラスの中から、あるクラスを識別するための表現。

（例　”ろ”と”る”はストロークの最後が違う。）



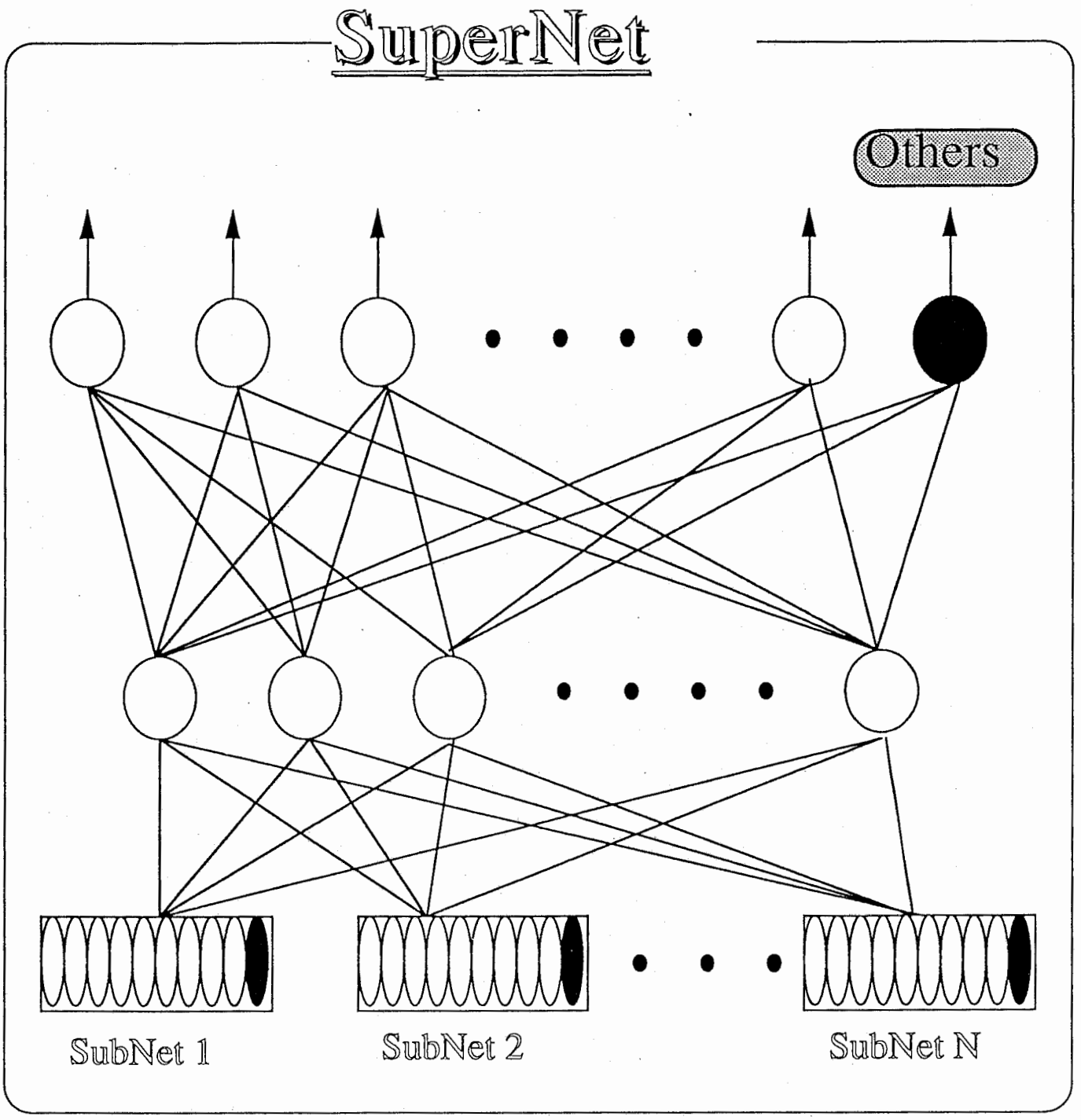


・大規模ニューラルネットの基本構成部品



SubNet

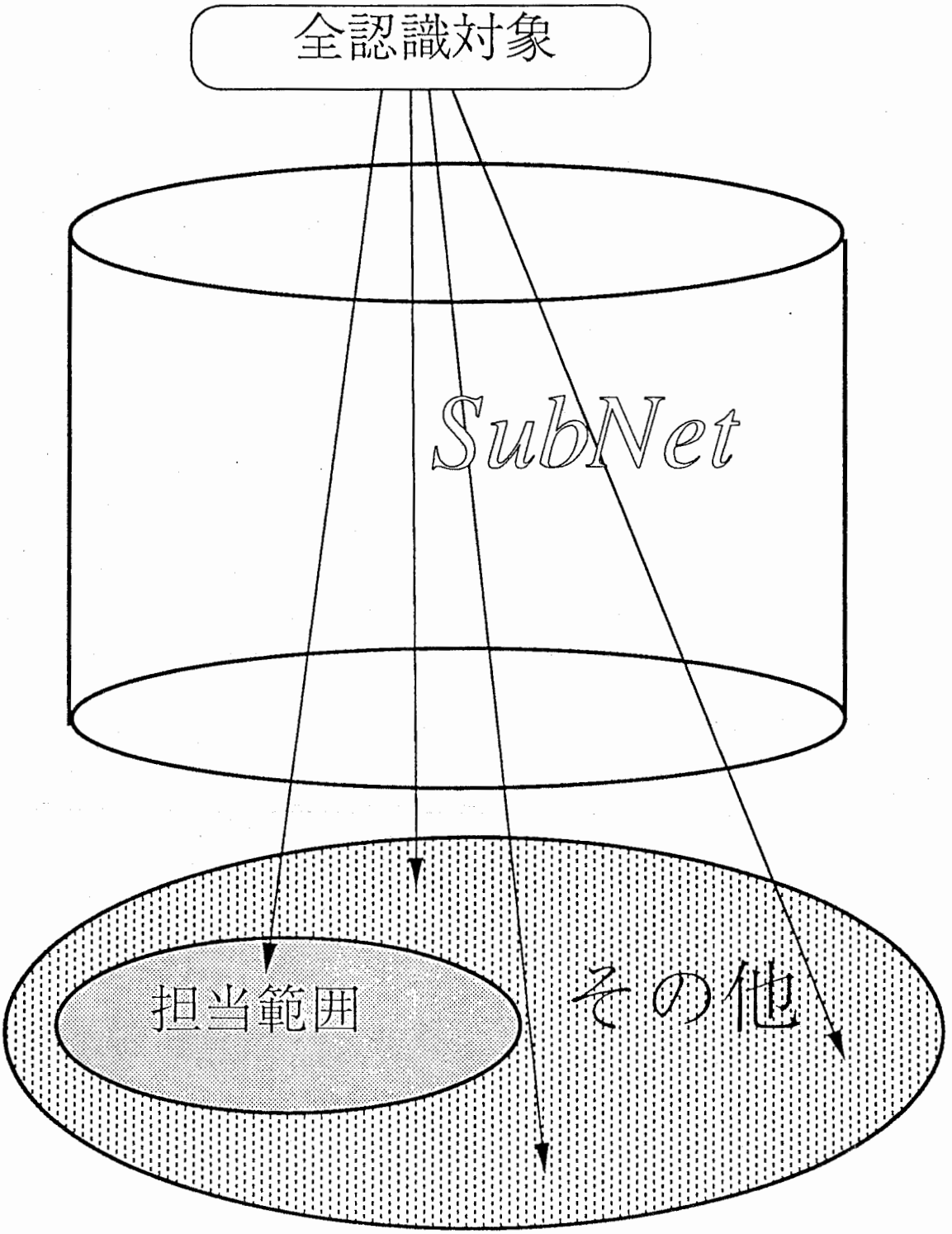
・階層的構成法による ニューラルネットの大規模化



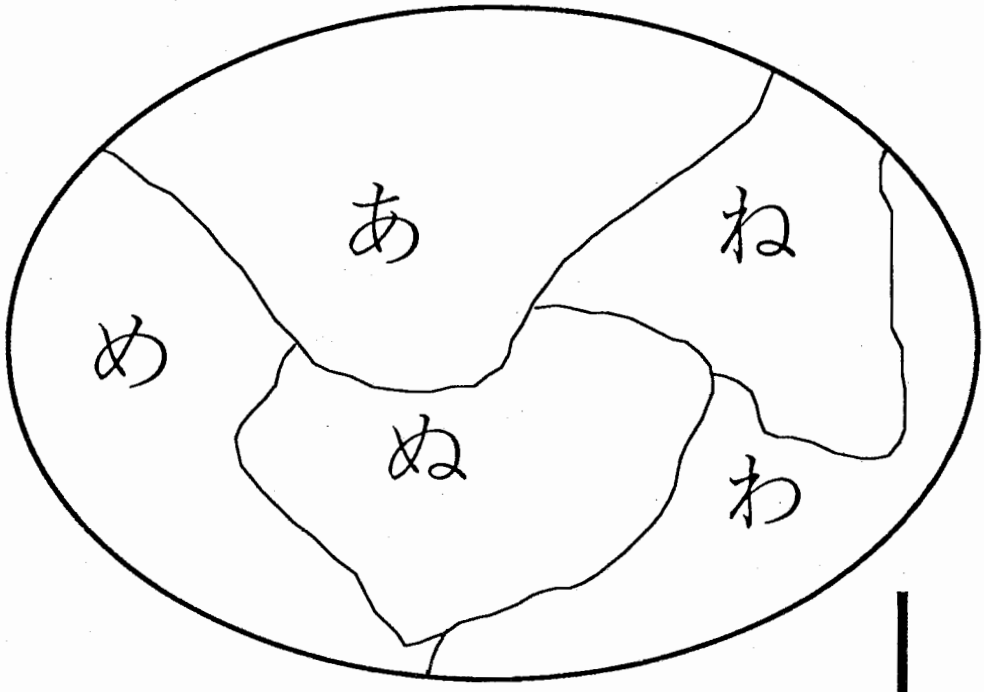
・”その他セル”の役割

内部表現の純化＆頑健性

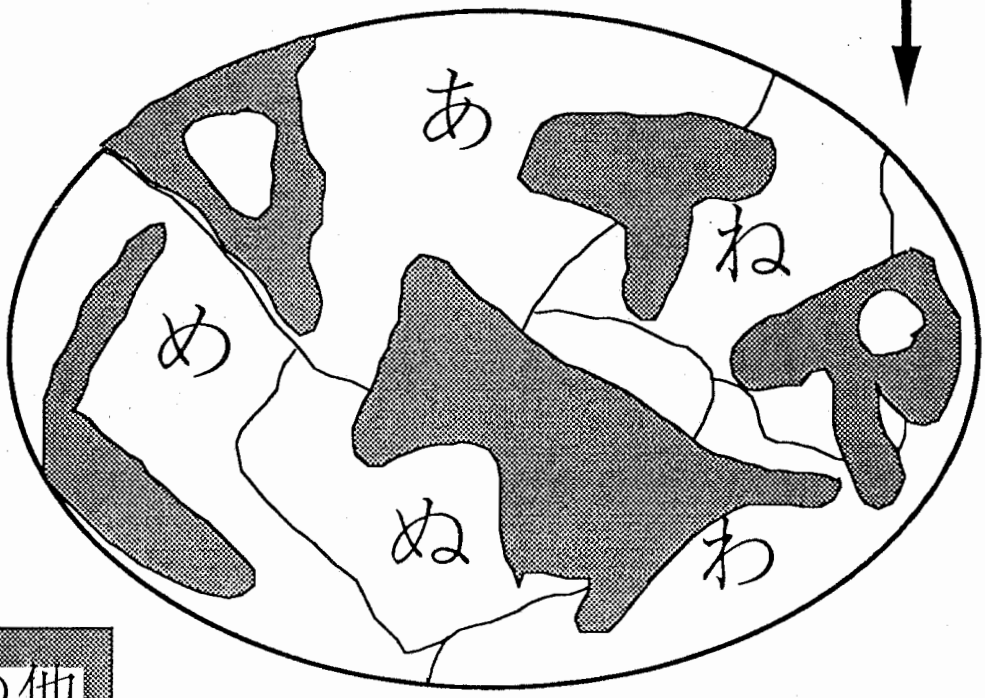
たとえば、”論”と”輪”を識別できるネットワークは”言”偏を識別するために使用できるか？



・”その他セル”による内部表現の変化の概念図



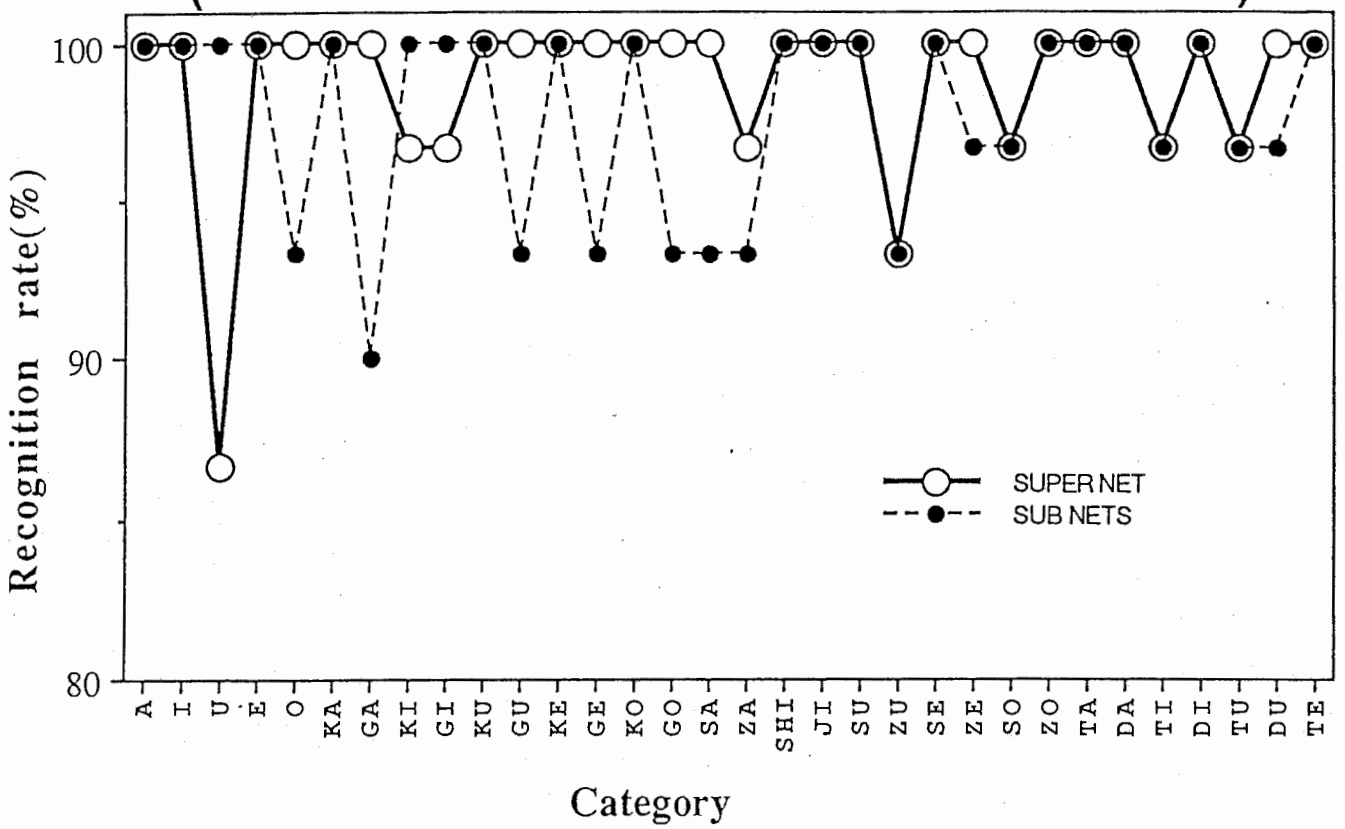
その他セルなし

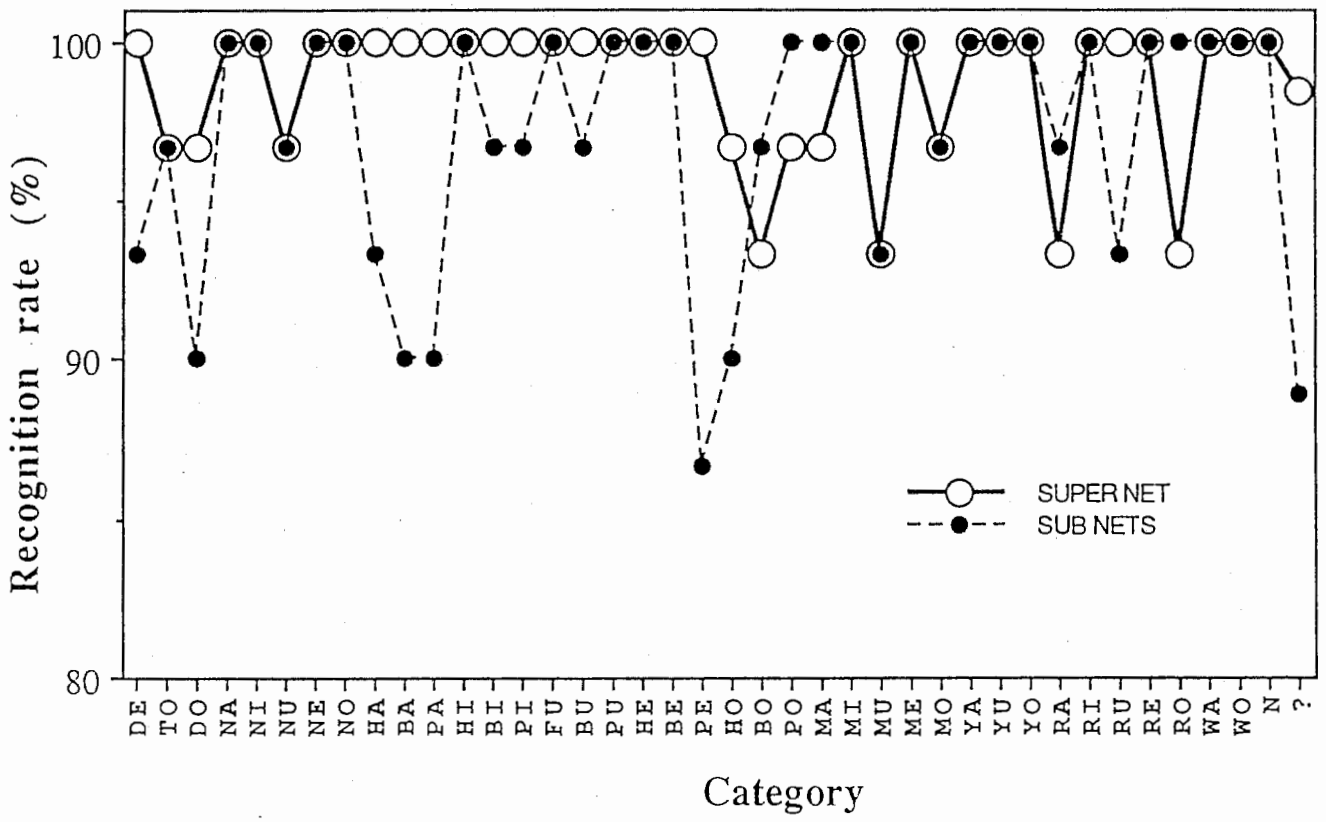


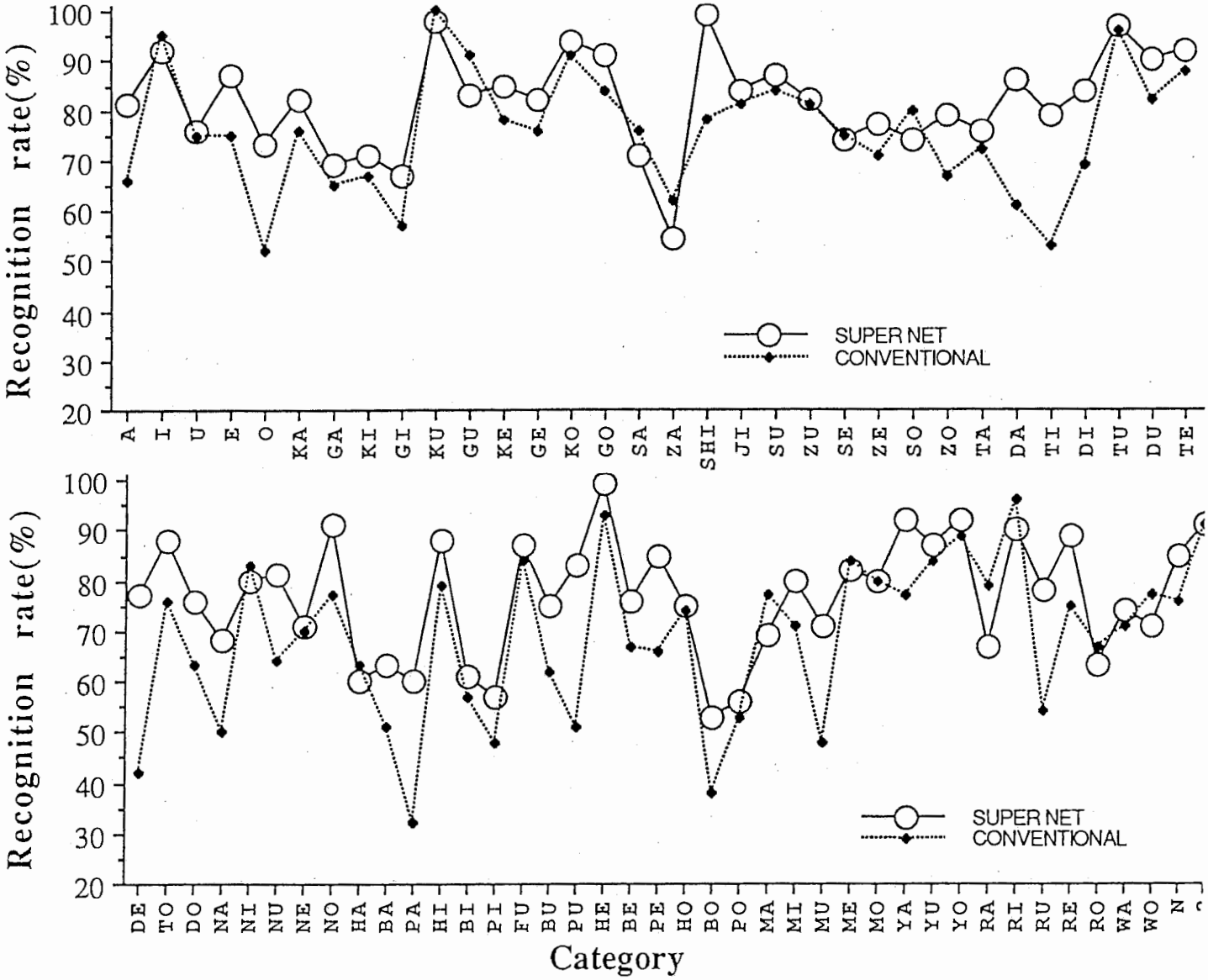
その他領域

その他セルあり

Recognition rate for 30 known samples (SUB NETWORKS & SUPER NETWORK)






Recognition rate(%)
100
90
80
70
60
50
40
30
20
SUPER NET
CONVENTIONAL
A I U E O KA GA KI GI KU GU KE GE KO GO SA ZA SHI JI SU ZU SE ZE SO ZO TA DA TI DI TU DU TE
DE TO DO NA NI NU NE NO HA BA PA HI BI PI FU BU PU HE BE PE HO BO PO MA MI MU ME MO YA YU YO RA RI RU RE RO WA WO
Category

従来手法との比較

手書き平仮名認識結果

*これぐらい書いていただけると
認識率98.6%である。
"ひ"だけ誤認識した

認識実験３

［目的］

・ＪＩＳ第１水準（約３０００文字）手書き漢字を認識するシステムの構築。

［内容］

・３０００文字の中から選出した手書き漢字２７０文字の認識。

［結果］

・情報統合能力を上げるために、小規模なネットが担当するカテゴリをＫ－Ｍｅａｎｓクラスタリング手法で選出した。

・その他セル　－－＞　その他フィルター

小規模なネットワークが自己の担当範囲外の入力に対して自己の出力を抑える機構を、学習効率を考慮して小規模なネットワークに持たせずにＬＶＱを利用したフィルターを利用して行った。

認識率　（％）		
その他フィルター	なし	あり
	７４	５５
ネットワーク構造	統合型	選択型
	７４	７２†

†選択ネットワークの能力が１００％と仮定して

［認識対象について］

●　常用漢字

約２０００字　読み書きできる

●ＪＩＳ第１水準漢字

＋１０００字　読める　書く？

●ＪＩＳ第２水準

＋３４００字　読めない？書けない

　さんずい－約１６０文字

渥溢淫渦浦泳洩液沿演

汚沖温河海涯浬渇滑汗

漢潤灌汽汲泣・・・・

・・・・・・・・・・

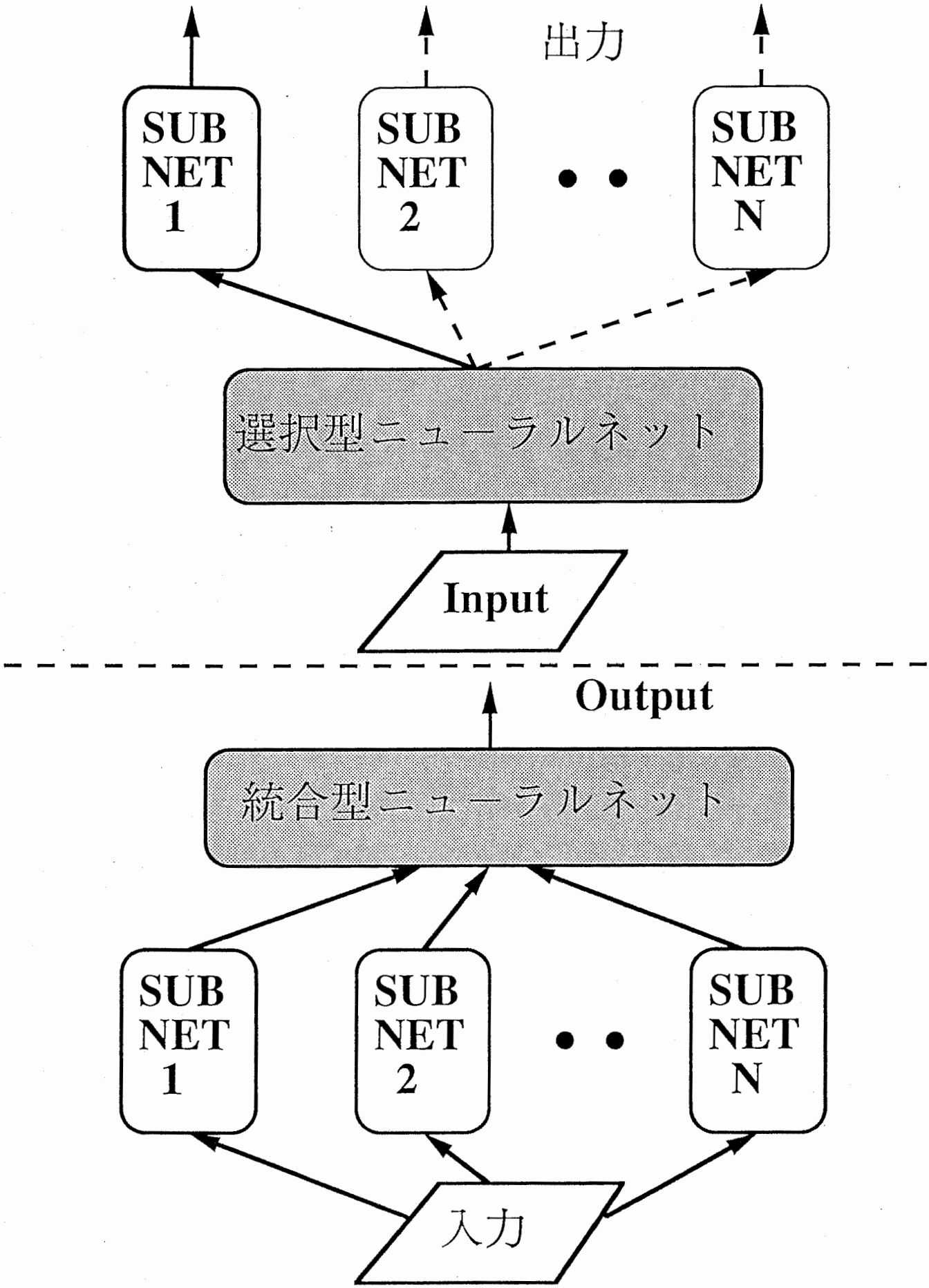
・・・・・・・漣浪漏湾

濠　瀦　漉　濤　潑　瀆　灌

第２水準

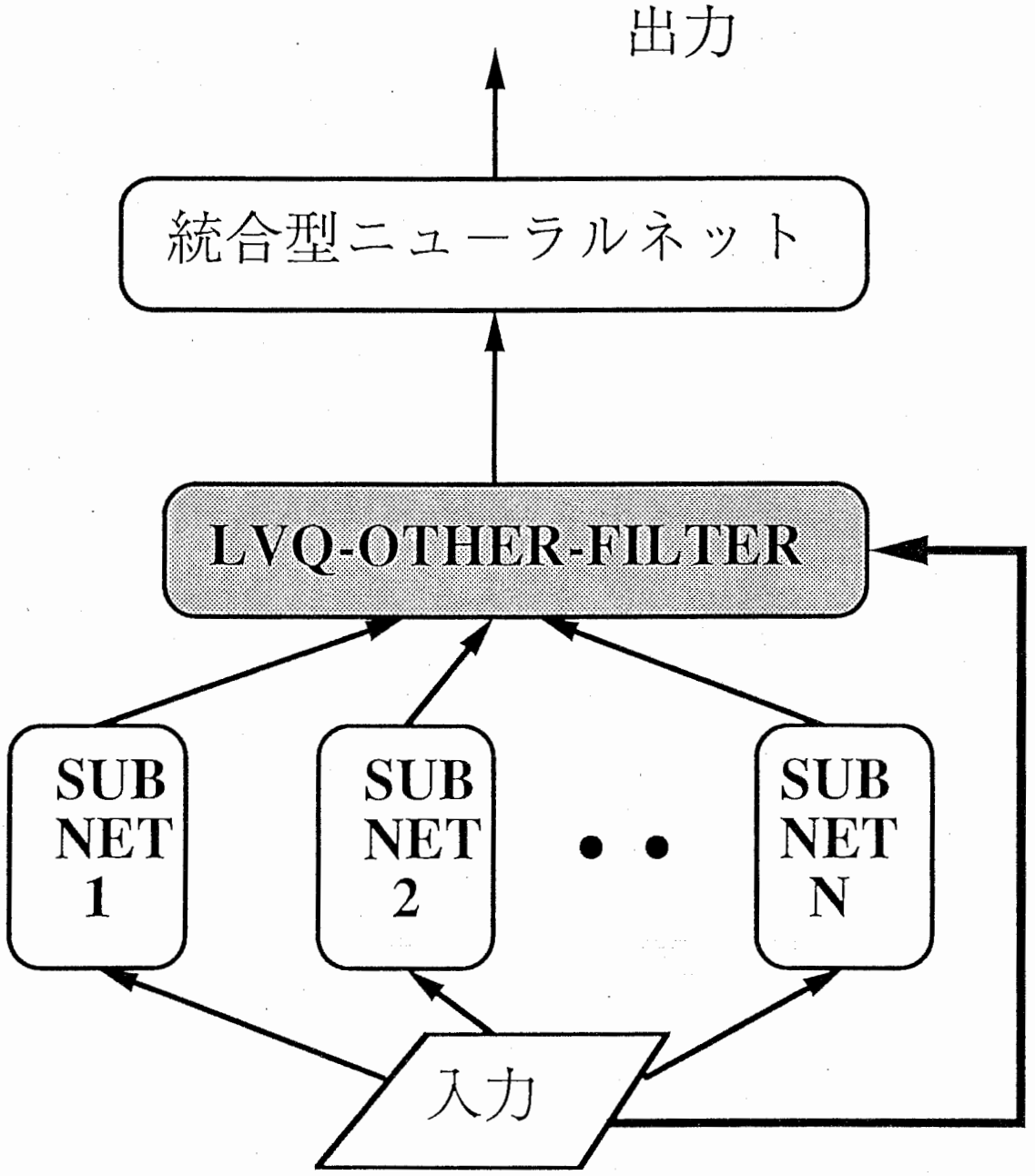
汕　汢　汪　沂　沍　沚　沁

選択型 VS. 統合型

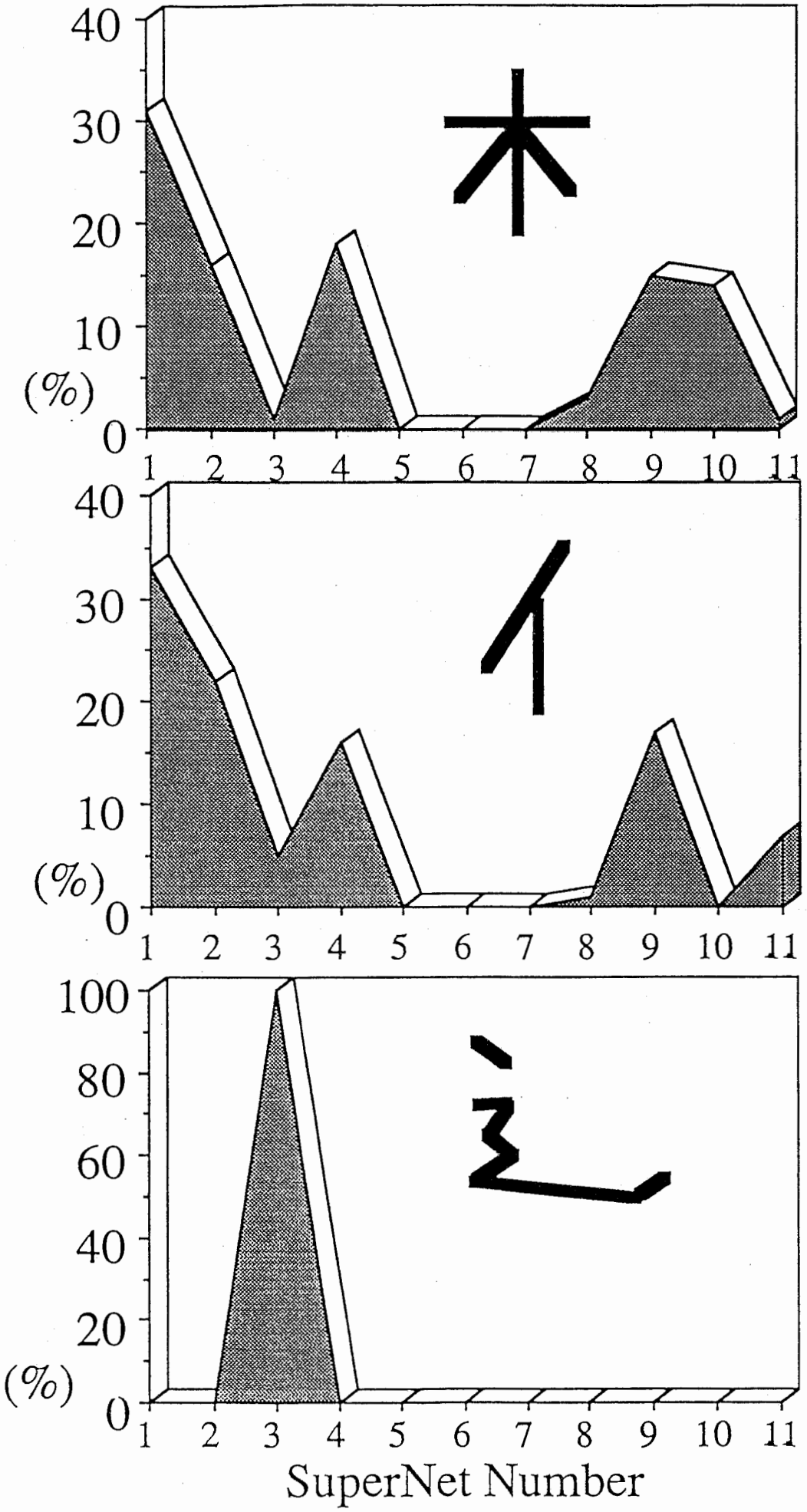

出力
SUB NET 1
SUB NET 2
SUB NET N
選択型ニューラルネット
Input
Output
統合型ニューラルネット
SUB NET 1
SUB NET 2
SUB NET N
入力

改良統合型ニューラルネット

Other-Filter(LVQ)により上位の層での情報の統合が容易になります.



・Ｋ－Ｍｅａｎｓ手法による漢字の分布状態



Other-Filter
(統合促進ネットワーク)

[入力]

$\mathbf{O}_n^k$: 各サブネットの出力

$\mathbf{D}_n$: LVQにより作成された各サブネットの参照ベクトル(45本)と入力ベクトルとのユークリッド距離

[出力]

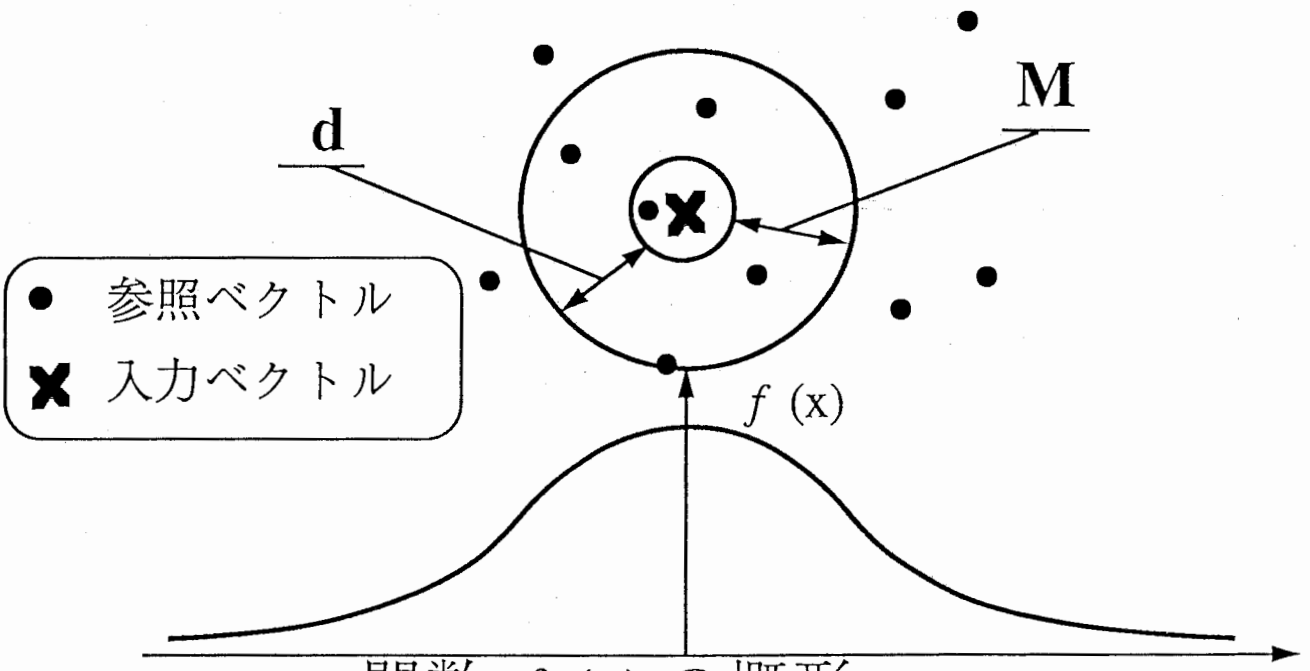
$$\mathbf{O'}_n^k = \mathbf{O}_n^k * f(\mathbf{D}_n)$$

$$(n = 0, 29), (k = 0, 8)$$

$$f(x) = 1 / (1 + e^{(|X - M + 2d|) / Cd)})$$

M : $\mathbf{D}_n$ の平均

d : $\mathbf{D}_n$ の分散



関数 f (x) の概形

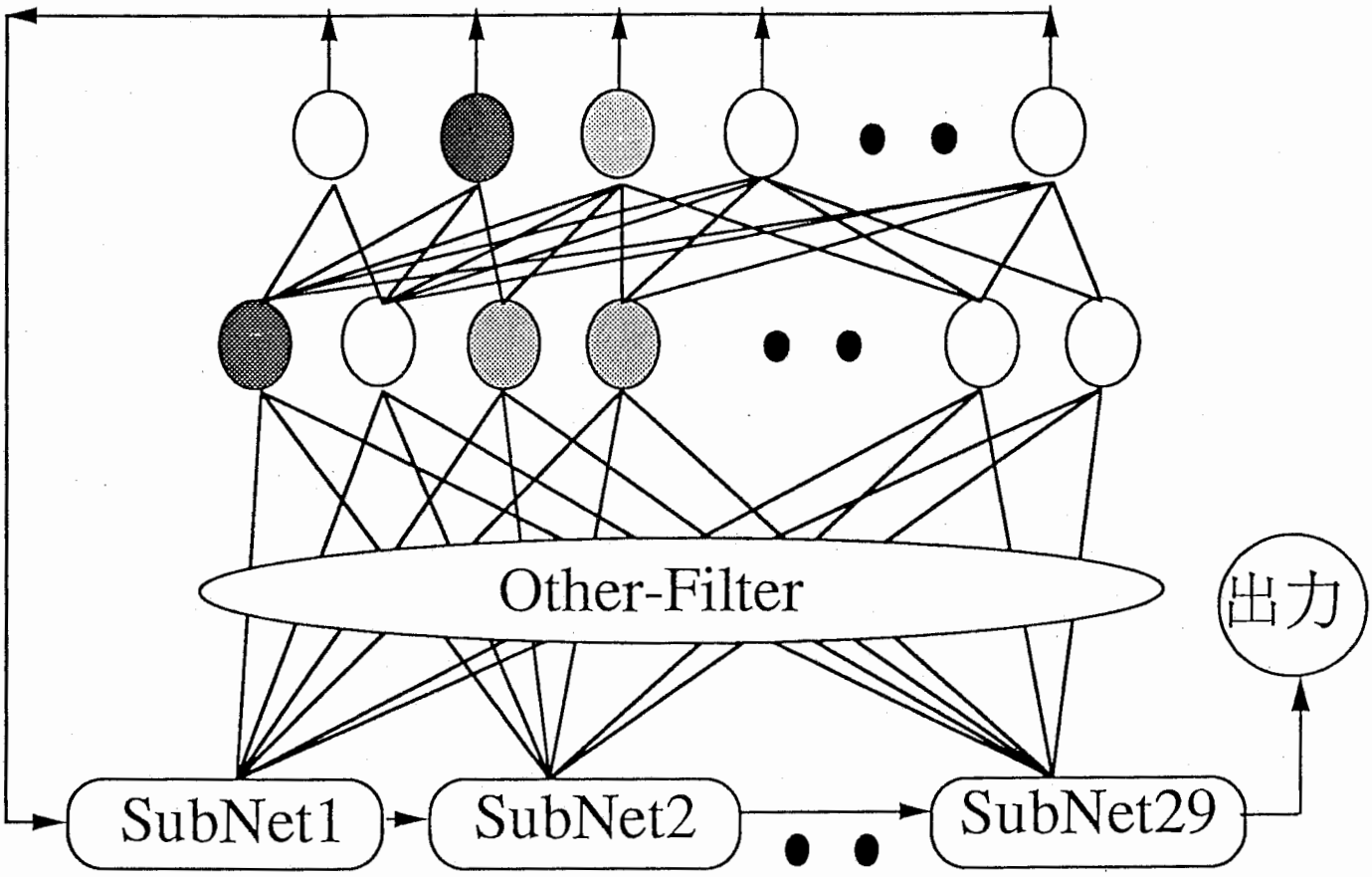
SuperNet

[入力]

Other-Filterで加工された30個のサブネットからの270個の出力

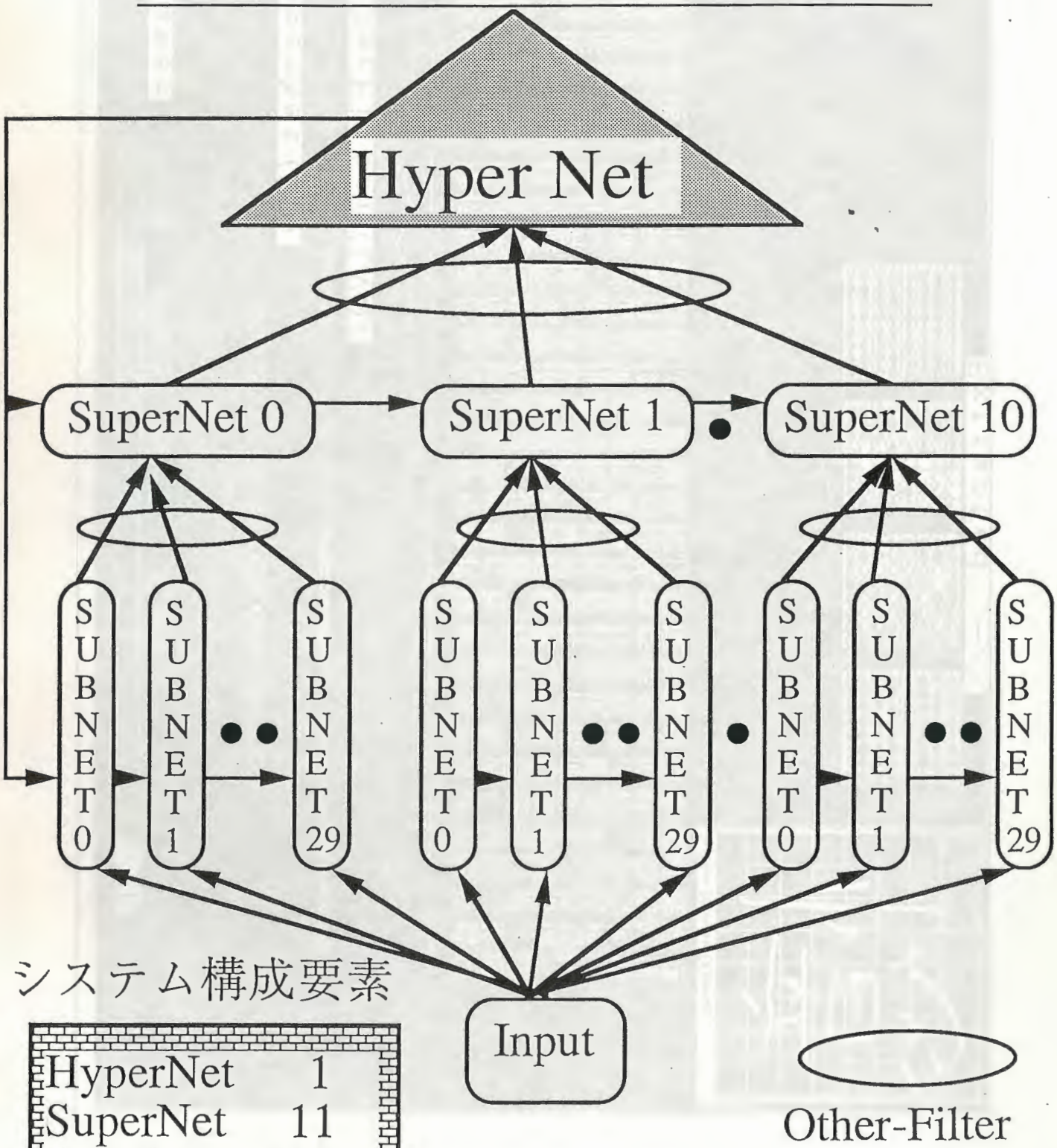
[出力]

入力パターンが30個のサブネットのどれに属しているかを出力



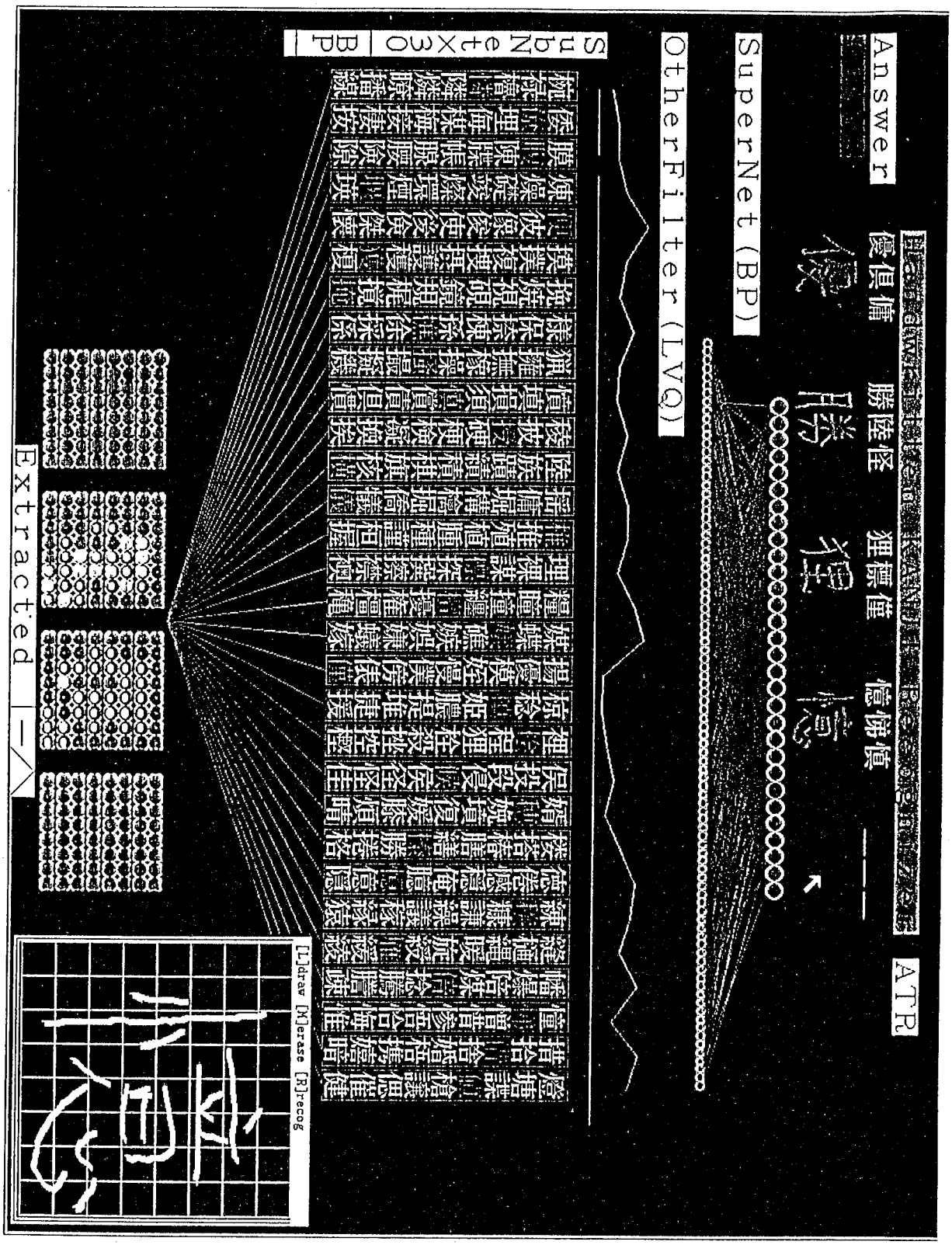
入力

JIS第一水準手書き漢字認識システム全体図



システム構成要素

HyperNet	1
SuperNet	11
SubNet	330
OtherFilter	12



270文字漢字認識システム