論　文

CombNET-IIによる中国語1000単語音声認識

正員　魏　　回[†]　　正員　北村　正[†]
正員　岩田　彰[†]　　正員　鈴村　宣夫[†]

Speech Recognition of Chinese 1000 Words Using Large Scale Neural Network CombNET-II

WEI Hui [†], Tadashi KITAMURA [†], Akira IWATA [†] *and*
Nobuo SUZUMURA [†], *Members*

あらまし　多数のカテゴリーを分類する一つの手法として，我々は大規模ニューラルネット（CombNET-II）を提案した．これは，前段に入力ベクトルを大分類するためのベクトル量子化型ニューラルネットを配置し，後段にグループ内のデータを細分類するための階層ニューラルネットを配置した，くし型の構成をしている．本論文では，CombNET-IIを用いる大語いの音声認識手法を提案し，この方法を中国語の単語音声認識に適用し，その有用性について検討する．音声信号から2次元メルケプストラム法によって求められる特徴量をCombNET-IIの入力に用いる．2次元メルケプストラムは音声の静的特徴と動的特徴を同時に分析でき，音声認識には有効なパラメータである．今回の音声認識実験では，特定話者が中国語で発声した世界の国名と都市名1000単語を用いた．各単語を5回ずつ発声し，この中の4回分のデータで学習を行い，残りの1回分のデータを認識させたところ，99.0％の認識率が得られ，本方法の有効性が示された．

キーワード：大規模ニューラルネットワーク，音声認識，バックプロパゲーション，ベクトル量子化，メルケプストラム

1. まえがき

近年，音声認識の研究は，着実に進展しており，特定話者，不特定話者，大語いおよび連続音声認識などを対象にしたいろいろな研究が報告されている．しかし，これらの研究では，線形予測法や，フーリエ変換，フィルタ群などを用いて特徴パラメータを求め，これらの特徴からパターンマッチングにより認識が行われることが多く，認識では未知のパターンが前もって用意してある標準パターンのうちどれと一致するかを調べることが必要であるため，語い数が増大すると共に識別の計算量も多くなる．

人間は，視覚，聴覚を通して文字，音声，画像など数多くのパターンを脳の神経回路網で学習・記憶し，新たに入力があると記憶された神経回路網によって的確に素早く判断・識別できる．最近，このような神経回路網を模擬したニューロコンピュータの研究が盛んである．ニューロコンピュータは，人間がもつ情報伝達方式を工学的に実現しようとするものであり，学習と並列処理の機能を備えることを特徴としている．このような特徴をもつニューロコンピュータは，これまでのコンピュータでは不得意とされていた分野に適用してよい結果を得ることが期待されている．

既に，文字認識や音声認識などの分野で，ニューラルネットワークのモデルを用いた実用化の研究が報告されている．しかし，これまでに行われたニューラルネットワークに関する研究は，分類カテゴリー数の少ない比較的小規模なニューラルネットを取り扱っており，実用的で分類カテゴリー数の多い大規模なニューラルネットに関する研究は少ない．

我々は，これまでに多数のカテゴリーを分類する大規模ニューラルネットワークの構築法（CombNETおよびCombNET-II）を提案した[(1),(2)]．これは，前段に入力データを大分類するためのベクトル量子化型ニューラルネットを配置し，後段にグループ内のデータを細

† 名古屋工業大学電気情報工学科，名古屋市
Department of Electrical and Computer Engineering, Nagoya Institute of Technology, Nagoya-shi, 466 Japan

分類する階層型ニューラルネットを配置した，くし型の構成をしている．

本論文では，まず大規模ニューラルネットCombNET-IIを用いた単語音声認識の手法を示す．次に，この方法を中国語単語の音声認識に適用し，大語いの単語認識の実験を行う．最後にCombNET-IIの特徴を利用して，単語認識の問題点について解決法を考える．

入力データには音声の周波数変化と時間変化を同時に表すことのできる2次元メルケプストラムを用いた[3],[4]．2次元メルケプストラムは，周波数をメル尺度で表したメル対数スペクトルの周波数と時間の2次元フーリエ変換で定義される．この分析法は，スペクトルの時間変化を概略的な変化と微細な変化に分離できるという特徴があり，音声の平均的特徴および動的特徴を同時に分析できる．

2. CombNET-II

CombNET-IIは，2段ネットワークの構造で，図1に示すように前段の入力データを大分類するためのベクトル量子化型ニューラルネット(Stem Net)と，後段のグループ内のデータを細分類する階層型ニューラルネット(Branch Net)で構成されている．入力データを大分類するため，Stem Netでは，入力の特徴空間をいくつかの部分空間に分割し，部分空間ごとに代表となる参照ベクトルをもつ量子化ニューロンを生成する．その結果，同じ部分空間に入った入力データは似たもの同士としてグループ化される(図2(a))．そして，グループ化された入力データ集合ごとにBranch Netの学習を行い，部分空間内のカテゴリーの識別境界を形成する(図2(b))．

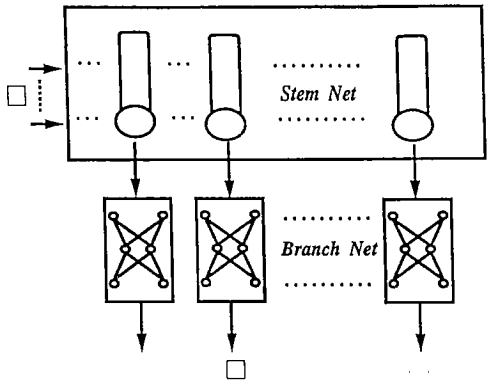
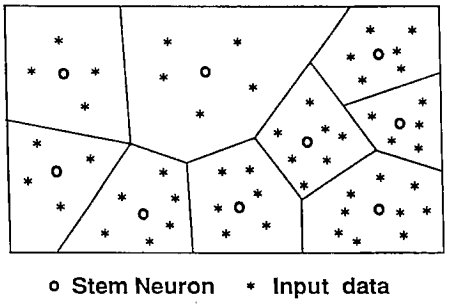


図1 CombNET-IIの構造
Fig. 1 The network structure of CombNET-II.

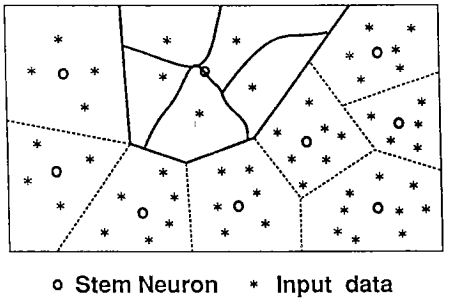
従来発表したCombNET-I[1]では，Stem Netの形成をKohonenの自己組織化アルゴリズムにより行い，Branch Netの学習は，バックプロパゲーション法を用いて行っていた．しかし，Kohonenの自己組織化アルゴリズムでは，各ニューロンと最適整合となる入力データの数を制御することができず，その数が各ニューロンで大きくばらついてしまう問題があった．大分類される入力データの数の大小は，後段のBranch Netのサイズにそのまま反映される．このことは，各Branch Netの学習深度に影響を与え，認識時に出力ニューロンの発火度にばらつきを生じさせる．

また，多次元空間のベクトル量子化という観点から考えると，最適な量子化とは，限られた量子化レベル数(ニューロン数)において，量子化誤差を最小にすることである．量子化誤差を最小にするには，入力データが多次元空間内で一様分布している場合には一様量子化でよく，ニューロンを空間内に均等に配置すればよい．

しかし，一般には入力データの分布には偏りがあるため，一様量子化は最適な量子化とはならない．この



(a) Vector quantization by stem net in a feature space.



(b) Discriminating border by branch net.

図2 Stem NetとBranch Netによる2段階識別
Fig. 2 Two-step classification by stem net and branch net.

ようなときには，入力データ分布を一様分布に変換する写像を行った上で，一様量子化を行うことになる．写像された空間で一様に量子化をすることは，原空間では入力データ分布密度に応じてニューロンを配置し，それにより分割される分布空間ごとの入力データの分布密度を一様にすることに対応している．すなわち，各ニューロンと最適整合となる入力データの数が，同程度になるようにニューロンを配置したとき，最適なベクトル量子化が形成されることになる．

我々は，このような観点から，自己増殖型ニューラルネットを新しく考案した．そしてこれをCombNET-IのStem Netの学習アルゴリズムにとり入れCombNET-IIと名づけた．自己増殖型ニューラルネットは，一つのニューロンと最適整合するカテゴリーの数があるしきい値を超えたときに，そのカテゴリーを2分割するニューロンを生成する手法を基本としたもので，本手法を用いると一つのニューロンと最適整合するカテゴリーの数を制限することができる．

ここで提案したベクトル量子化ニューラルネットは，特徴空間における入力データの分布を考慮して，あるしきい値を上限として入力ベクトルを同程度の大きさのグループに分割することを特徴としている．Stem Netの学習を行うときに，初期状態においてはニューロンは存在しない．学習データを入力することによって，参照ベクトルをもつニューロンが分裂し，自己増殖していくため，自己増殖ニューラルネットワークと呼ぶことにする．

CombNET-IIの学習の詳細な説明は他の論文に譲り[2]，ここでは学習と認識方法の概要を簡単に述べる．

2.1 CombNET-IIの学習

CombNET-IIの学習は，2段に分けて行われる．学習データの集合には，一つのカテゴリーで複数パターンを用意することを前提とする．まず，Stem Netから自己増殖型学習則により学習を行う．この学習は，第1と第2の二つの過程が存在する．第1過程は第1順目の入力データの投入によるニューロンの生成過程である．図3に示すように，入力データ(X_k)を入れたとき，その時に存在しているすべてのニューロンの中で，自分と最適整合になるニューロンcを求める(step 1)．ここで，M_cはニューロンcの参照ベクトル，sim(X_k, M_i)は式(1)のように入力データ(X_k)と参照ベクトルM_iの整合度を示す．

$$\mathrm{sim}(X_k, M_i)=\frac{X_k \cdot M_i}{\|X_k\| \|M_i\|} \qquad (1)$$

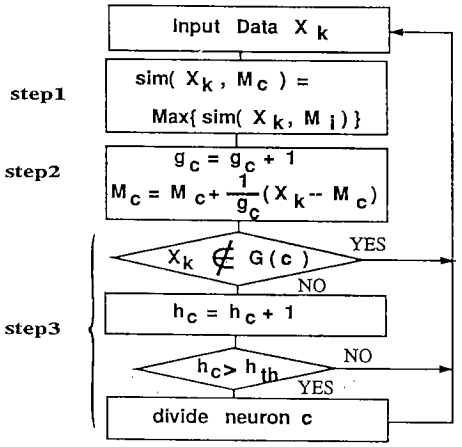


図3　自己増殖型学習法の第1過程の流れ
Fig. 3　Block diagram of self-growing learning process 1.

求めた最適ニューロンに対し，データの個数を示す内部ポテンシャル(g_c)を一つ増加させ，参照ベクトル(M_c)をニューロンcに最適整合する入力データ群の平均になるように修正する(step 2)．このとき，入力データのカテゴリーが最適ニューロンcに対して新規カテゴリーであれば，カテゴリー数を示す分裂ポテンシャル(h_c)も一つ増加させる．$G(c)$はニューロンcと最適整合したデータの属するカテゴリーの集合を表す．更に，増加した分裂ポテンシャルh_cがあらかじめ指定したしきい値(h_{th})を超えたときに，ニューロンcを二つに分裂させる(step 3)．分裂法としては，まず参照ベクトルM_cを通る次式に示す超平面を一つ生成する．

$A(X-M_c)=0$

Aは乱数により与え，平面の傾きを示すパラメータである．この超平面によってニューロンに最適している入力データ群を2分する．但し，2分された二つの入力データ群の属するカテゴリー数に偏りが生じた場合には，等分割になるまで超平面を生成し，分割の繰返しを行う．

第2過程では，既に生成したニューロンに対して学習を行う．図4に示すように，入力データに対して新しく求めた最適整合ニューロンcが前回の最適整合ニューロンc'と異なるときに，ニューロンc'の参照ベクトルを入力データとは逆方向へ修正し，新しいニューロンcの参照ベクトルは入力データの方向に修正する．同時にニューロンc'の内部ポテンシャルを一つ減少させ，新しく最適整合したニューロンcの内部ポテンシャルを一つ増加させるように修正を行う(step 2)．また第1過程と同じように，入力データの属するカテゴリー

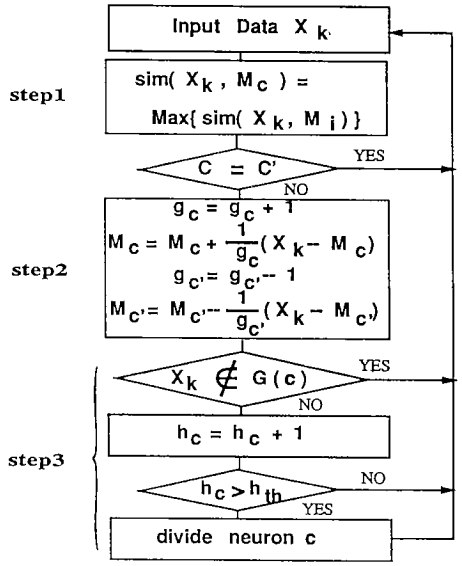


図4　自己増殖型学習法の第2過程の流れ
Fig. 4　Block diagram of self-growing learning process 2.

が，ニューロンに対して新しいカテゴリーであれば，分裂ポテンシャル h_c を一つ増加させる．増加した分裂ポテンシャルがあらかじめ指定したしきい値を超えれば，第1過程と同じくニューロンは分裂を起こす (step 3)．第2過程の学習は，すべての学習データに対して変更がなくなるまで繰り返す．

このような手法に基づいてニューロンを生成していくと，特徴空間で入力データの分布密度の高い部分には，多くのニューロンが生成され，密度の低い部分では，ニューロンはあまり生成されない (図2(a))．また，各ニューロンの参照ベクトルが，入力データの分布の統計的性質を反映し，各ニューロンと最適整合になる入力データのカテゴリー数も，細胞分裂を起こすポテンシャルのしきい値 h_{th} により上限が制限される．

Stem Net の学習後，後段の Branch Net の学習を行う．全入力データは，Stem Net ニューロン数と同数のグループに分割されるため，Branch Net をそれぞれグループについて作成する．すなわち，各ニューロンの分担すべき入力データとカテゴリーを求め，そして，カテゴリー数と同数の Branch Net の出力ニューロンを用意し，入力データを学習データにして，バックプロパゲーション法を用いて階層型ニューラルネット (Branch Net) の学習を行う．

2.2　CombNET-II の識別

識別は次のようにして行われる．まず，入力データを Stem Net のネットワークに通して，Stem Net のどのニューロンと最適整合になるかを求める．次に，最適整合となったニューロンが担当するカテゴリーグループを分類する後段の階層型ネットワーク (Branch Net) に入力データを入力し，最も大きな値をもつニューロンに対応するカテゴリーを出力として識別する．

CombNET-II における Stem Net の学習過程において，上限値 h_{th} によりカテゴリー分割を行うため，各 Branch Net の出力層は上限値 h_{th} より多くならず，各 Branch Net における学習の深度をそろえることができる．バックプロパゲーション法による学習は小規模なネットワークについては収束が容易であるが，大規模なネットワークになるとローカルミニマムに陥る可能性も高く，たとえ収束するにせよ膨大な計算量を費やすことになる．更にネットワークを小規模化することによって，各サブネット結合数が少なくなるため，全体の計算時間を著しく減少させることができる．例えば，入力データのカテゴリー数を N，中間層数を出力の半分 $N/2$ とするとき，単一 BP での計算量は，$N^3/2$ に比例する．しかし，CombNET-II の第1層により生成したニューロン数 L とした場合，全体の計算量は，$L\times(N/L)^3/2=N^3/(2L^2)$ に比例し，結果として計算量が $1/L^2$ に減少する．更に，学習の収束性を考えると，その差がもっと大きくなる．

3.　2次元メルケプストラム

本研究では，音声のスペクトルの周波数変化と時間変化を同時に分析できる2次元メルケプストラムを音声の特徴パラメータとして用いる．2次元メルケプストラムは，周波数をメル尺度で表したメル対数スペクトルの時系列 $S(k,m)$ の周波数 k と時間 m に対する2次元フーリエ変換で定義される (式(2))．この方法は，周波数および時間方向にフーリエ変換することにより，スペクトルを周波数および時間方向の概略的な変化と微細な変化に分離できるという特徴がある．

$$C(q,p)=\frac{1}{M}\sum_{m=0}^{M-1}c(q,m)W_2^{mp}$$

$$=\frac{1}{NM}\sum_{k=0}^{N-1}\sum_{m=0}^{M-1}S(k,m)W_1^{kq}W_2^{mp} \qquad (2)$$

$$W_2=\exp^{-j2\pi/M};$$

$m=0,1,\cdots,M-1$　M：フレーム数

図5にメル対数スペクトルを求める手順を示している．音声信号 $x(n,m)$ を時間 n についてフーリエ変換してスペクトル $X(k,m)$ を求め，対数スペクトル

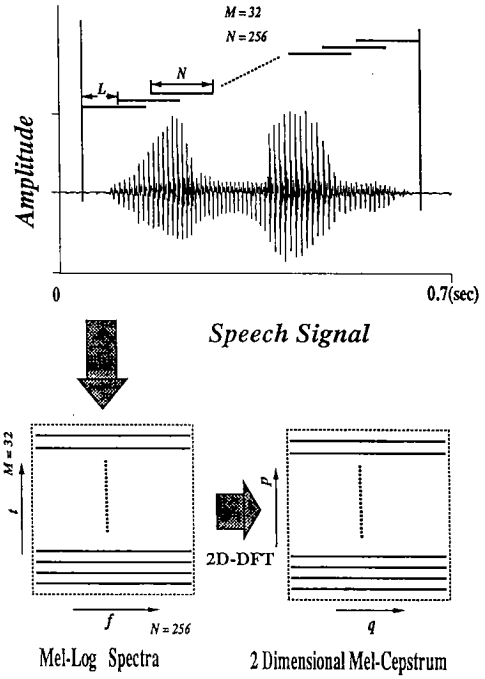


図5　音声データによる特徴量の抽出過程
Fig. 5　Diagram of generating input data.

$ln|X(k,m)|$ とした後，周波数のメル尺度化を行う．その後改良ケプストラム法により[6]，1次元メルケプストラム $c(q,m)$ を求め，次に m についてフーリエ変換を行い，2次元メルケプストラム $C(q,p)$ が得られる．

本研究では，音声信号は10 kHzでサンプリングし，視察により単語の始点と終点を検出し，その間を $M=32$ フレームに等分割する．このとき時間窓はフレーム長 $N=256$ のブラックマン窓を用いた．変換した2次元メルケプストラムは，実数部と虚数部を合わせて，入力データの次元数は512 $(q,p=0,\cdots,15)$ である．しかし，この結果の各領域の物理的意味を考えると，利用するデータ数を更に減少することができる．今回の報告では単語音声認識に有効な特徴量を45個用いることにした $(p=0\sim1,\ q=1\sim15,\ p=0$ の虚数部は除いてある$)$．ここで，p はスペクトル包絡の時間分解能を制御するパラメータ，q はスペクトル包絡の周波数分解能を制御するパラメータである．

4. その他ニューロンの学習

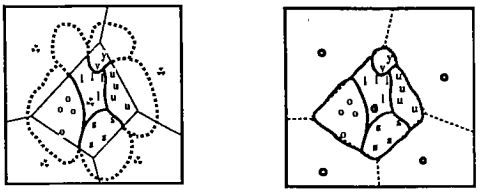
CombNET-II は CombNET-I[1] と同様に Stem Net と Branch Net の出力を結合する方法を用いた．すなわち，Stem Net の出力値から上位いくつかのニューロンを選び，それらのニューロンに対応する Branch Net の最大出力値を考え，以下の結合式を用いて最大値となるカテゴリーを識別結果として選ぶ．

$$\theta=\theta_1^{r}\times\theta_2^{(1-r)}\quad(0\leqq r\leqq1)\qquad(3)$$

θ_1：Stem Net の出力値

θ_2：Branch Net の出力値

しかし，各 Branch Net の学習は，担当する Stem Net によって分割された部分空間内のデータだけを用いて行うため，担当外のデータを認識させた場合，出力層の値はほとんどすべて低い値になるが，まれに高い値となることもある．この高い値を結合式に用いると，識別の結果に影響することがある．



(a) Without the other unit.　(b) With the other unit.

図6　特徴空間における Branch Net の識別境界
Fig. 6　Discriminating border of branch net in a feature space.

今回，解決法としては各 Branch Net に「その他ニューロン」を設け，自分の Branch Net に属していないカテゴリーの学習データをその他ニューロンが発火するように学習させた．こうすることによって，担当外のカテゴリーのデータが入力されたとき，その他ニューロンのみ強く発火し，本来の出力層ニューロンの値が低くなる．この方法は，学習データのカテゴリー数が増えると，その他ニューロンの学習負担が大きくなると考えられる．しかし，CombNET-II において，自己増殖型学習則によって作られた Stem Net ニューロンの各参照ベクトルは，各グループ内の学習ベクトルを代表しているので，その他ニューロンの学習を，対応する参照ベクトルを用いて学習することにより，計算時間を短縮することができる．

その他ニューロンの学習効果は，図6で説明できる．Stem Net に対応する入力データのみを用いて学習した場合，学習データ集合をグループ内に限ったため，部分空間外における Branch Net の応答は未学習であり不定となる（図6(a)）．従って，領域外の入力データが投入されたとき，Branch Net の出力値が大きくなり，第1層と第4層の出力を結合するとき，誤認識することがある．しかし，その他ニューロンを追加して，周り

の参照ベクトルも用いて学習させたとき，図6(b)のように領域を作ることができる．従って，領域外の入力データが投入されたとき，その他ニューロンが大きく発火して，正解以外の出力層ニューロンが大きく発火することがなくなり，認識率の向上が期待できる．

5. 1000単語認識実験

今回，中国語で発声した世界の国名と都市名を用いて，CombNET-IIによる特定話者の単語音声認識を行い，その有効性を検討した．

中国語の発声は，音節を単位とし，一つの音節は一つの漢字に対応している．音節の構造は日本語と同様に子音＋母音の構造をもち，一つの音節は多い場合4種類の発声(4声)がされる．従って，音節単位での音声認識は難しいことが想像される．しかし，日常の会話では単語単位で発声されることが多く，連続単語の切出しの難しい現在では，単語単位による認識が適当である．

単語音声認識では発声ごとの変形が大きいことが予想できるので，CombNET-IIを適用するとき，一つの単語を複数回発声することにした．本論文の実験では，1000単語の音声について，男性1名が各単語を5回ずつ発声して，単語音声データを2次元メルケプストラムにより45点の入力データベクトルを作成し，5セットのデータを用意した．

以下，実験結果によってCombNET-IIの特徴と音声認識の問題を併せて検討する．

5.1 CombNET-IIの認識率

従来の音声認識システムでは，語い数の増加に伴い，計算時間が増加し，認識率も低下する．CombNET-IIは，このような問題を考慮して，あらかじめ大きな入力集合を自己増殖型ニューラルネットワークによっていくつかのグループに大分類する．そして，グループ内でカテゴリーの識別を行うので，以上の問題点を改善することができる．

今回の実験では，計算機の計算時間とCombNET-IIの規模を考慮して，Stem Netの分裂ポテンシャルのしきい値h_{th}を20にすることにした．まず予備実験として学習データ数と認識率のしきい値の関係を求めた．これは順番に1セット，2セット，3セット，4セットのデータを用いて学習を行い，それぞれ残りの4セット，3セット，2セット，1セットのデータを認識したところ，各々93.2％，94.5％，98.5％，99.0％となった．但し，これらの値は，その他ニューロンありとして，結合パラメータrの値を変えたときの最高認識率である．すなわち，同じ発声でも学習データを多く用いることによって，ネットワークの汎化性が向上したことがわかる．次に1000単語の音声データの5セットのうち，4セットを学習データに用い，残り1セットを未学習データとしてrの値と認識率とその他ニューロンの効果について検討した．この実験は，5回のデータで組合せを変えて各単語について5回認識実験を行った(Leave one out)．データ1000単語に対する総認識単語数は，(認識単語)×(組合せ数)=5000である．認識率はそれに対するものである．

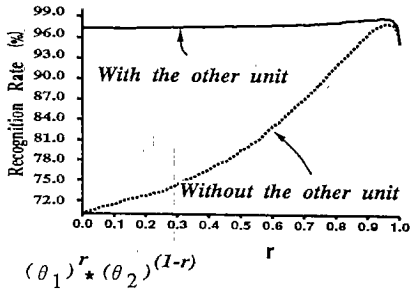


図7　rに対する認識率
Fig. 7　Recognition rate vs. value r.

一つのカテゴリーにつき四つのデータを用いたので，データのばらつきにより一つのカテゴリーが複数のニューロンに対応する場合がある(マルチテンプレート)．今回の実験では，しきい値を20に指定したが，生成したニューロンの数が5回の実験の平均で126.2個であった．各ニューロン内のカテゴリー数の総計が1215.2個(5回の実験の平均値)なので，ニューロン内のカテゴリー数は，平均1215.2/126.2=9.7となり，ほぼしきい値の半分になった．

図7に式3の各r値に対する認識率を示す．$r=1$の場合は，Stem Netの最大値に対応するBranch Netの出力値を認識結果とする．$r=0$の場合は，Stem Netにおいて最大5番目までの出力を選び，その五つのBranch Netから最大出力値を認識結果とする．しかし，図7に示すようにStem NetとBranch Netの出力値を組み合わせることによってより高い認識率が得られることがわかる．$r=0.94$のとき，最高認識率99.0％が得られた．一方，図7に示したように，その他ニューロンを導入することによって認識率が向上するとともに，rの値にかかわらず安定な認識率が得られることもわかった．このように，その他ニューロンは

CombNET-IIの汎化能力を向上させる働きをもっていることが明らかとなった．

5.2 特徴量の抽出とCombNET-IIの表現

特徴量の抽出と抽出した特徴量の評価は，難しい問題である．今回，我々は2次元メルケプストラム法を用いて，特徴量の抽出を行った．2次元メルケプストラムについては他の論文に詳しく検討しているが[3]，ここで，実験結果からStem Netの特徴空間でどのような発声特徴を用いて分類しているかを考える．

CombNET-IIを印刷漢字の認識に適用した場合，Stem Netニューロンの参照ベクトルには，漢字の「偏」，「つくり」，「かまえ」を特徴として，グループが形成されている．単語音声認識でも同じような現象が見られた．表1に示したように，同一のニューロンと最適整合する音声単語集合が，音声特徴量の中の母音の成分を中心にして分類されていることがわかる．グループ内の単語量増加に伴い，このような特徴量と発声内容の対応が，はっきりしてくる．

更に，表2は1000単語の中の誤認識の例を示している．誤認識の中に発音の類似したもの，しかも，人間でも間違いやすい単語が多い．また，これらの誤認識が，すべてStem Netの五つの候補に入っているので，誤認識の原因は，Branch Netで学習不足により生じたことがわかる．これは前処理の始点位置(子音位置)のずれか，あるいはメルケプストラムの入力ベクトルの個数の不足などのため，学習に必要な情報が不足しているためと考えられるが，更に検討する必要がある．

5.3 単語発声の変形とCombNET-IIの適応

音声認識において特徴量抽出の問題以外に，発声者の違い，あるいは発声時間と環境の違いによって音声データの内容がかなり違うという問題がある．従来の方法では，一般に複数個の標準パターンを用意して，発声の変形に対応している．しかし，これは発声者の増加と発声条件の変化に対して十分適応できなくなる．

CombNET-IIのStem Netの自己増殖型学習は教師なし学習であるため，同じカテゴリーの入力パターンが別々のStem Netニューロンと最適整合することがある．この場合は，それぞれのStem Netニューロンに対応した複数のBranch Netに同じカテゴリーを正解とするニューロンが存在する．すなわち音声パターンの個々のばらつきが大きいとき，CombNET-IIは，複数の正解ニューロンを備えることによって適応する．この特徴は，将来の複数話者と不特定話者などの音声認識にも有効なものとなるであろう．

表1 各ニューロンと最適整合する単語の一覧表

No.	単語種類				
0	chóngqìng	hóngqí	wāngqǔ	dùqǔ	pǔxī
	重慶	紅旗	王曲	杜曲	浦西
	chuānqí	gōngqí	āijí	pǔjùn	
	川崎	宮崎	埃及	蒲郡	
4	wénjīng	zījīn	lùdǐng	lóngjǐng	
	文景	紫金	鹿頂	龍井	
	lóngjīn	ēnpíng	pǔjùn	wénjīng	
	龍津	恩平	蒲郡	文京	
10	běijīng	méixiàn	wèishuǐ	mìyún	ruìjīn
	北京	梅県	衛水	密雲	端金
68	hángzhōu	bāngzhōu	dēngzhōu	shēngzhōu	
	杭州	浜州	登州	昇州	
83	héshān	fútài	chōngshān	fúshān	
	合山	福太	充山	府山	
	xiǎoshān	fùshān	suǒjiāng	fúshān	
	小山	富山	鎖江	福山	
102	yāntái	běihǎi	tiāntán	fēilái	
	煙台	北海	天壇	飛来	
	tiānhǎi	qiānhǎi	hēihǎi		
	天海	前海	黒海		
111	pénglái	dōnghǎi	lèlái	ānlái	wénlái
	蓬萊	東海	楽来	安来	文萊
114	táiwān	hǎigǎng	táihuái	bǎiniǎo	
	台湾	海港	台懐	百鳥	
	hǎitóng	àiyuán	yǎjiādá	báilè	
	海潼	愛媛	雅加達	白楽	
122	héngshān	zhōngshān	zhōushān	fóshān	
	衡山	中山	舟山	佛山	
	lùtèdān	huángshān	lóngshān	bùdān	
	鹿特丹	黄山	龍山	不丹	

表2 誤認識の例

認識結果	正答
中部 (zhōngbù)	東部 (dōngbù)
華山 (huáshān)	花山 (huāshān)
韓平 (hánpíng)	南平 (nánpíng)
四会 (sìhùi)	智恵 (zìhùi)
三道 (sāndào)	三島 (sāndǎo)

表3 Stem Netにおける1000単語の分布

ニューロン数	データ数
1	813
2	161
3	24
4	2

本実験は，1000カテゴリーのパターンを4セット用いて学習を行ったが，生成したStem Netの単語分布状況を分析したところ，表3のようにまとめることができる．この表を見てわかるように，同じ発声データは，大半(81.3％)が同一ニューロンに入っているが，残っ

た少数の発声データは，複数の Stem Net のニューロンに対応していることがわかる．

6. むすび

本論文では，CombNET-II を音声認識に適用するとき，一つのカテゴリーに複数の学習データがあることを考え，各ニューロンごとに最適整合するデータの属するカテゴリー数を制限するため，分裂ポテンシャルを導入し，また，各 Branch Net にその他ニューロンを追加することによって，認識の汎化性が向上したことを示した．

今回は，2次元メルケプストラムを入力パラメータとする大規模ニューラルネット CombNET-II を用いた大語いの単語音声認識手法を提案し，特定話者の中国語 1,000 単語の認識実験によりその有効性が示された．その結果は，2次元メルケプストラムの全領域 (512 点) 中の，各フレームの対数スペクトル包絡の平均と緩やかな時間変化を表す領域 (45 点) だけを用いて実験を行い，99.0 % の認識率が得られた．

以上の実験から CombNET-II は，従来のネットワークモデルでは困難な多数のカテゴリーを分類する分野に有効であることがわかった．この結果，本手法により実用的な規模で音声認識システムをもったネットワークを構築することが可能である．

最後に同じデータを用いて，本手法と総当り照合法 (Nearest Neighbour 法) との比較実験を行った．NN 法の標準テンプレートは，4回の発声データの平均より作成した．この実験では，同程度の認識率 (99.1 %) が得られた．しかし，CombNET-II は，カテゴリーをグループごとに分割することによって，認識時間を減らすことを特徴としている (この実験では認識時間が NN 法より 4 倍位速い)．また，**5.3** に検討したように，Stem Net の学習は教師なしの自己増殖型学習であるので，パターン変形が大きい音声に対して自動的に複数のニューロンを備えることによって適応するため，認識課題が複雑になっても必要な最小限のテンプレートを参照ベクトルとしてもつニューロンを適応的に生成することができる．現在，複数話者の単語データを用いて研究を進めている．更に，不特定話者の大語い単語認識などへの適用も検討する予定である．

文　献

(1) 岩田　彰，當麻孝志，松尾啓志，鈴村宣夫："大規模4層ニューラルネット"CombNET"，信学論(D-II)，**J73-D-II**，8，pp. 1261-1267 (1990-08).

(2) 堀田健一，岩田　彰，松尾啓志，鈴村宣夫："大規模ニューラルネット CombNET-II"，信学技報，**NC90**-34 (1990-10).

(3) 北村　正，片柳恵一："2次元メルケプストラムの静的特徴・動的特徴を用いる数字音声認識"，信学論(A)，**J72-A**，4，pp. 640-647 (1989-04).

(4) 伊藤朝信，北村　正，早原悦朗："2次元メルケプストラムとニューラルネットを用いた単語音声認識"，信学技報，**DSP**89-40 (1990-03).

(5) 魏　回，北村　正，岩田　彰，鈴村宣夫："CombNET-IIによる1,000単語音声認識"，信学技報，**NC90**-141 (1991-03).

(6) 今井　聖，阿部芳春："改良ケプストラム法によるスペクトル包絡の抽出"，信学論(A)，**J62-A**，4，pp. 217-223 (1978-08).

(平成3年8月7日受付，11月15日再受付)

魏　回

昭59中国清華大学計算機工程系卒．昭63名工大大学院博士前期課程了．平1同大大学院博士後期課程入学，現在に至る．音声認識，ニューラルネットに関する研究に従事．

北村　正

昭48名工大・工・電子卒．昭53東工大大学院博士課程了．工博．同年同大精密工学研究所助手．昭58名工大・工・電子講師．現在同大・工・電気情報助教授．音声情報処理，ディジタル信号処理，ニューラルネットワークに関する研究に従事．日本音響学会，IEEE，EURASIP各会員．

岩田　彰

昭48名大・工・電気卒．昭50同大大学院修士課程了．同年名工大・情報助手．昭57年4月より昭58年10月まで，ドイツ連邦共和国ギーセン大学医学部医用情報研究員．昭59名工大・情報・助教授．現在名工大・電気情報・助教授．生体情報処理，医用画像処理，ニューラルネットワークに関する研究に従事．工博．日本ME学会，情報処理学会，IEEE各会員．

鈴村　宣夫

昭28名大・工・電気卒．民間会社勤務の後，昭38名大・工・助手．以後，講師，助教授を経て，昭49名工大・情報工学科教授．学科改組より，現在，電気情報工学科教授．この間，生体信号の計測，処理，生体関連の画像処理の研究に従事．工博．