

# ニューラルネットワークを用いた音声生成システムの能力の分析について

Theoretical and Experimental Analysis  
of a Speech Production System  
Using Neural Networks

田中啓夫  
Akiyo TANAKA

吉村光夫  
Mitsuo KOMURA

富士通株式会社・国際情報社会科学研究所  
IIAS-SIS, Fujitsu Limited

あらまし 我々は、ニューラルネットワークを用いた音声学習・生成方式を提案し、より自然な音声を合成できることを示した。この学習・生成方式は、(1)特異点解消方法、(2)パターン情報拡散化学習方式、(3)協調学習方式の3つの特徴をもっている。

本報告では、理論的検討として、パターン情報拡散化学習法が収束するための条件、誤差曲面の解析を行う。実験的検討として、提案したニューラルネットワークの記憶能力を評価し、協調方式の効果を調べる。また、誤差逆伝播(BP)法との比較のために中間層の活動度の分布等を調べる。これらを通して、任意音声生成のための基礎付けを行う。

Abstract We proposed a new method to produce speech using neural networks. This method uses back-propagating error signals and has following three features: (1) method for removing of singular points from BP algorithm, (2) spread pattern information(SI) learning algorithm, and (3) cooperative learning(CL) algorithm. In this report, we discuss the conditions of convergence and analyze the form of error surface for SI learning algorithm. And we examine memory capacity, effect of cooperative learning algorithm, and the distribution of hidden unit activations.

## 1. まえがき

音声を文章(テキスト)から合成する方法として規則合成法が用いられている。この方法は、実音声から規則を抽出し、その規則に基づいて逆に音声を合成する方法である。音声は言語としての規則に基づいて発声されているはずであり、そのような規則があるから、相互のコミュニケーションが可能となる。しかし、発声器官、聴覚器官、神経系には個人差がある上にフィードバック経路をつくっており、音声は複雑な現象となっている。これらのことから、一般的な規則の記述が困難となっている。

我々は、ニューラルネットワークを用いた音声生成法を提案した。<sup>(1)(2)</sup>これは、実音声から抽出した音声パラメータを直接ニューラルネットワークが学習することにより、発声時の複

雑な現象を記憶し、より自然な発声が可能になると考えたからである。この方法でより自然な音声を生成することが基本的に可能であることを既に示した。

ニューラルネットワークの学習法として、目標値を直接与えられない素子も学習できる誤差逆伝播(Back Propagation)法<sup>(3)</sup>が提案され、音声認識等さまざまな分野に応用されている。我々の提案したニューラルネットワークの学習方式は、目標値と出力値との誤差を逆伝播させるが、後で述べるように、特異点解消方法、パターン拡散化学習方式、協調学習方式の特徴をもつ。これらの特徴は、アナログ量を高精度に生成するために必要なものである。本報告では、パターン拡散化協調学習法における収束性についての理論的検討および提案したネットワークのダイナミックスに関する実験的検討を行う。

## 2. パターン情報拡散化協調学習方式

まず、3層（入力層、中間層、出力層）のニューラルネットワークを定義する（図2.1）。各素子の入出力関係は、次式で与えられる。

$$\text{中間層} : u_j = \sum_i w^{H_{ij}} x_i \quad (1) (**)$$

$$y_j = f(u_j - \theta_j) \quad (2)$$

$$\text{出力層} : v_k = \sum_j w^{O_{jk}} y_j \quad (3)$$

$$z_k = f(v_k - \theta_k) \quad (4)$$

ただし、 $i=1, \dots, I$ ;  $j=1, \dots, J$ ;  $k=1, \dots, K$ 。出力関数  $f(u)$  は、(a)可微分性(differentiable)、(b)単調非減少性(monotone nondecreasing)と(c)非負値性(nonnegative)を仮定する。この性質を満たす出力関数として、後の実験ではロジスティク曲線：

$$f(u) = (1 + \exp(-u + \theta))^{-1} \quad (5)$$

を用いる。このとき、 $y = f(u)$ とすると、導関数は、

$$f'(u) = y(1 - y) \quad (6)$$

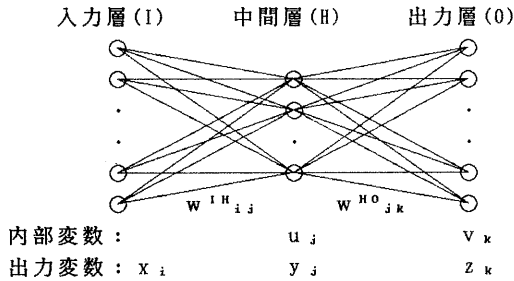


図2.1 3層構造のニューラルネットワーク

〔誤差逆伝播学習法 (BP法)〕

3層のニューラルネットワークにおける誤差逆伝播(Back Propagation 略してBP)法を簡単に述べる。n回目の学習に与えられる入力を ( $x_i(n): i=1, \dots, I$ ) とし、目標出力を ( $t_k(n): k=1, \dots, K$ ) とする。まず、出力層における誤差逆伝播量  $\delta^O_k(n)$  は、

$$\delta^O_k(n) = (t_k(n) - z_k(n)) f'(v_k(n)) \quad (7)$$

中間層における誤差逆伝播量  $\delta^H_j(n)$  は、

$$\delta^H_j(n) = (\sum_k \delta^O_k(n) w^{O_{jk}}) f'(u_j(n)) \quad (8)$$

これらの量を持ちいて、出力層の素子の重み係

$$(*) \sum_i = \sum_{i=1}^I, \quad \sum_j = \sum_{j=1}^J, \quad \sum_k = \sum_{k=1}^K$$

数の修正量は、

$$\begin{aligned} \Delta w^{O_{jk}}(n) &= -\alpha \frac{\partial E(n)}{\partial w^{O_{jk}}} \\ &= \alpha \delta^O_k(n) y_j(n) \end{aligned} \quad (9)$$

あるいは、収束性を改善するために、

$$\begin{aligned} \Delta w^{O_{jk}}(n) &= \alpha \delta^O_k(n) y_j(n) \\ &\quad + \beta \Delta w^{O_{jk}}(n-1) \end{aligned} \quad (10)$$

中間層の素子の重み係数の修正量は、

$$\begin{aligned} \Delta w^{H_{ij}}(n) &= -\alpha \frac{\partial E(n)}{\partial w^{H_{ij}}} \\ &= \alpha \delta^H_j(n) x_i(n) \end{aligned} \quad (11)$$

あるいは、(10)と同様に、

$$\begin{aligned} \Delta w^{H_{ij}}(n) &= \alpha \delta^H_j(n) x_i(n) \\ &\quad + \beta \Delta w^{H_{ij}}(n-1) \end{aligned} \quad (12)$$

で与えられる。ここで、 $\alpha, \beta$ は学習定数とよばれ、十分に小さな正の定数とする。

〔パターン情報拡散化協調学習方式〕

我々はパターン情報拡散化協調学習(Spread Pattern Information and Cooperative Learning, 略してSICL)方式を既に提案した。<sup>(11, 2)</sup> この方式は次のような特徴をもつ。

(I) 特異点解消方法 : 出力値が目標値に一致するとき以外に、出力関数の導関数が0となるときにも、誤差逆伝播量は0となる。誤差逆伝播量が0となると、(9)あるいは(11)によって重み係数の修正量も0となり、収束が遅くなる。そこで、後者のような場合を取り除くことが望ましい。最も簡単な関数として、次式のようにする。

$$\begin{aligned} \delta^O_k(n) &= (t_k(n) - z_k(n)) \\ &\quad \times (f'(v_k(n)) + \gamma) \end{aligned} \quad (13)$$

$$\begin{aligned} \delta^H_j(n) &= (\sum_k \delta^O_k(n) w^{O_{jk}}) \\ &\quad \times (f'(u_j(n)) + \gamma) \end{aligned} \quad (14)$$

ここで  $\gamma$  は十分小さな正の定数とする。

(II) パターン情報拡散化学習 (SI) 方式 : 出力層の素子の重み係数  $w^{O_{jk}}$  の値をランダム化させ固定する。即ち、学習を始める前に  $w^{O_{jk}}$  に正規乱数値を与え、固定する。(学習時、 $\Delta w^{O_{jk}}(n) = 0$ ) これにより、情報が重み係数空間に拡散され、学習定数 ( $\alpha, \beta$  等) に対する頑健性が確保され、学習精度が向上する。

(III) 協調学習 (CL) 方式 : 入力系列と出力とを組にしてニューラルネットワークに学習・記憶させる。入力データ系列(音声では時系列データ)の各点が互いに関係をもつとき、同じ入力系列に対して少しずれた点の出力を別のニューラルネットワークに学習させると、同じ出力に対して異なった入力環境をそれぞれ別のニュー

入力層 中間層 出力層 最終出力層

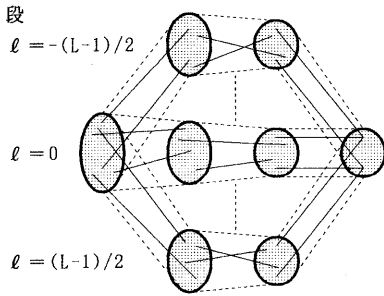


図2.2 パターン情報拡散化協調学習システム

ーラルネットワークに学習・記憶させたことになる。これらの出力の(荷重)平均をとることによって、安定した最終出力を得ることができる。この並列化されたニューラルネットワークを多段化されたという。

これらの特徴を備えたシステムは、共通の入力層をもった3層ニューラルネットワークを並列化し、その上に並列化された各段の出力層からの入力をもとめる最終出力層を加えた4層構造をもつ。(図2.2)

次に学習規則を示す。学習用のパターンは、各段 $l$ の入力層と出力層に提示される。 $n$ 番目に提示される学習パターンの目標出力値を $t_k(l, n)$ とする。重み係数 $w^{H^0}_{jk}(l)$ は特徴(II)で述べたように予め定められ、固定される。重み係数 $w^{H^1}_{ij}(l, n-1)$ は $(n-1)$ 番目まで学習が済んでいるとする。

まず、入力層から出力層への信号の伝播は、各段 $l$ について、

中間層:

$$u_j(l, n) = \sum_i w^{H^1}_{ij}(l, n-1) x_i(n) \quad (15)$$

$$y_j(l, n) = f(u_j(l, n) - \theta_j(l)) \quad (16)$$

出力層:

$$v_k(l, n) = \sum_j w^{H^0}_{jk}(l) y_j(l, n) \quad (17)$$

$$z_k(l, n) = f(v_k(l, n) - \theta_k(l)) \quad (18)$$

次に、誤差逆伝播量は、

$$\delta^0_k(l, n) = (t_k(l, n) - z_k(l, n)) \times (f'(v_k(l, n) + \gamma)) \quad (19)$$

$$\delta^H_j(l, n) = (\sum_k \delta^0_k(l, n) w^{H^0}_{jk}(l)) \times (f'(u_j(l, n) + \gamma)) \quad (20)$$

ここで、 $\gamma$ は特徴(I)で導入された関数(ここでは定数)である。重み係数の修正量は、

$$\Delta w^{H^1}_{ij}(l, n) = \alpha \delta^H_j(l, n) x_i(n) \quad (21)$$

あるいは、(実験では)

$$\Delta w^{H^1}_{ij}(l, n) = \alpha \delta^H_j(l, n) x_i(n) + \beta \Delta w^{H^1}_{ij}(l, n-1) \quad (22)$$

を用いる。

最終出力層の出力は、次式で与えられる。

$$z^F_k(n) = \sum_l w^F_{lk} z_k(l, n) \quad (23)$$

### 3. パターン情報拡散化学習法の収束性

(補題1) 特異点解消を行うパターン情報拡散化学習(SI)法は、二乗誤差を極小にする学習法である。

(証明) 付録に示す。

この証明から、収束するために、 $\alpha, \gamma$ は次の条件を満たさなければならない。

$$\alpha \ll \frac{1}{\left| t_k(n) - z_k(n|n-1) \right| \left[ \left| f'(v_k(n|n-1)) \right| \left\{ \sum_i |x_i(n)|^2 \right\} \times \sum_j |w^{H^0}_{jk}| \left| f'(u_j(n|n-1)) \right|^2 \times \sum_m |w^{H^0}_{jm}| \left| f'(v_m(n|n-1)) \right| + \gamma \right] \times \left| t_m(n) - z_m(n|n-1) \right|} \quad (24)$$

$$\frac{1}{\gamma} \gg \max_j \left\{ \left| \frac{\sum_k w^{H^0}_{jk} \delta^0_k(n) / \delta^H_j(n)}{\gamma + \sum_k w^{H^0}_{jk} (t_k(n) - z_k(n|n-1)) \times f'(u_j(n|n-1)) / \delta^H_j(n)} \right| \right\} \quad (25)$$

これから大略、 $\alpha$ は入力の大さき、中間層および出力層の重み係数の絶対値の和に逆比例し、 $\gamma$ は出力層の重み係数の絶対値の和に逆比例することがわかる。

(補題2) 3層ニューラルネットワークで、出力層の素子数が1のとき、パターン情報拡散化学習(SI)法による誤差曲面(重み係数と二乗誤差のつくる曲面)は、単峰である。

(証明) 仮定より $K=1$ 。二乗誤差は、

$$E = \frac{1}{2} (t_1 - z_1)^2 \quad (26)$$

出力関数 $f(u)$ の可微分性の仮定から、

$$\frac{\partial E}{\partial w^{H^1}_{i1}} = -w^{H^0}_{j1} (t_1 - z_1) f'(v_1) \times f'(u_j) x_i \quad (27)$$

出力関数として、ロジスティック関数を用いれば、更に強い仮定:(1)可微分性、(2)単調増加性( $f'(u) > 0$ )、(3)飽和特性( $0 < f < 1$ )を満たす。この場合、 $w^{H^0}_{j1} > 0$ ならば、

$$\frac{\partial E}{\partial w^{H^1}_{i1}} \begin{cases} > 0 & (z_1 > t_1) \\ = 0 & (z_1 = t_1) \\ < 0 & (z_1 < t_1) \end{cases} \quad (28)$$

$w^{H^0}_{j1} < 0$ ならば、 $\partial E / \partial w^{H^1}_{i1}$ は(28)と逆符号となる。これより、誤差曲面は、各 $w^{H^1}_{i1}$ 軸に関して、一つの極(小)値をもつ。

$w^{H^1}_{i1}$ 軸に関する動作を調べる。 $\Delta w^{H^1}_{i1} = -\alpha (\partial E / \partial w^{H^1}_{i1})$ で与えられるので、

(i)  $w^{H^0}_{j1} > 0$ かつ $z_1 > t_1$ のとき、

$\Delta w^{1H_{ij}} < 0$  となり、以下順次、 $\Delta u_j < 0$ 、 $\Delta y_j < 0$ 、 $\Delta v_i < 0$ 、 $\Delta z_1 < 0$  となる。よって、 $z_1$  は  $t_1$  (目標値) に近づく。

(ii)  $w^{H_{j1}} > 0$  かつ  $z_1 < t_1$  のとき、

(iii)  $w^{H_{j1}} < 0$  かつ  $z_1 > t_1$  のとき、

(iv)  $w^{H_{j1}} < 0$  かつ  $z_1 < t_1$  のとき、

も同様にして、 $z_1$  は  $t_1$  に近づく。

一般の出力関数の場合、仮定(2)により微係数が0となる点は、 $z_1 = t_1$  以外はその前後で微係数の符号は変化せず、停留点となるだけで、極点とはならない。

故に、誤差曲面 ( $w^{1H_{ij}} - E$ ) において、誤差が極(小)値をとるのは一つだけあって、出力が目標値に一致するような値を  $w^{1H_{ij}}$  がとった時だけであり、単峰である。

(証明終わり)

以上より、出力層の素子数が1のとき、パターン情報拡散化学学習法は、二乗誤差を極小とする学習法であり、かつ誤差曲面が単峰であることから、次のことが示された。

〔定理1〕 3層ニューラルネットワークで、出力層の素子数が1のとき、パターン情報拡散化学学習(SI)法は、二乗誤差を最小とする学習法である。

〔補題3〕 3層ニューラルネットワークで、出力層の素子数が2のとき、パターン情報拡散化学学習方式による誤差曲面の重み係数  $w^{1H_{ij}}$  軸に関する断面は、一つの極小と高々二つの非収束領域をもつ。

証明) 仮定より  $K = 2$ 。二乗誤差は、

$$E \equiv \frac{1}{2} \sum_k (t_k - z_k)^2 \quad (29)$$

出力関数  $f(u)$  の可微分性の仮定から、

$$\frac{\partial E}{\partial w^{1H_{ij}}} = - \sum_k w^{H_{jk}} (t_k - z_k) f'(v_k) \times f'(u_j) x_i \quad (30)$$

出力関数として、ロジスティック関数を用いると、更に強い仮定：(1)可微分性、(2)単調増加性 ( $f'(u) > 0$ )、(3)飽和特性 ( $0 < f < 1$ ) を満たす。この場合、

$$\frac{\partial E}{\partial w^{1H_{ij}}} = - \sum_k w^{H_{jk}} (t_k - z_k) z_k (1 - z_k) \times y_j (1 - y_j) x_i \quad (31)$$

ここで、(31)の  $k$  に関する部分を、

$$G_j(z_1, z_2) = \sum_k w^{H_{jk}} (t_k - z_k) z_k (1 - z_k) \quad (32)$$

とおく。

誤差曲面において、 $\Delta w^{1H_{ij}}$  だけ変化させた

ときの動作を解析する。

(i) まず、 $w^{H_{j1}} > w^{H_{j2}} > 0$  の場合、 $G_j$  の曲面は、典型的にはそれぞれ2つの正領域と負領域をもつ。図3.1に示すように、4つの領域を(a)、(b)、(c)、(d)とする。(a)と(b)の境界線を  $G_{ab}$ 、(b)と(c)の境界線を  $G_{bc}$ 、(c)と(d)の境界線を  $G_{ca}$  とする。

$G_j > 0$  の領域では、 $\Delta z_1 > 0$ 、 $\Delta z_2 > 0$ 。

$G_j < 0$  の領域では、 $\Delta z_1 < 0$ 、 $\Delta z_2 < 0$ 。

よって、 $\Delta z_k$  の方向を図3.1に矢印で示した。誤差曲面の  $w^{1H_{ij}}$  軸での断面は図3.2のようになる。破線の部分は、ニューラルネットワークの状態  $w^{H_{jk}}$ 、 $w^{1H_{ij}}$  および  $x_i$  によって生じたり、消滅したりする。 $G_{bc}$  と  $G_{ca}$  との間は極小は  $G_{ab}$  においてのみである。

(ii) 次に、 $w^{H_{j1}} > 0 > w^{H_{j2}}$  の場合、 $G_j$  の曲面は、図3.1と同様図3.3に示すように、4つの領域をもつ。

$G_j > 0$  の領域では、 $\Delta z_1 > 0$ 、 $\Delta z_2 < 0$ 。

$G_j < 0$  の領域では、 $\Delta z_1 < 0$ 、 $\Delta z_2 > 0$ 。

よって、 $\Delta z_k$  の方向を図3.3に矢印で示した。誤差曲面の  $w^{1H_{ij}}$  軸での断面図3.4は、図3.2と同様である。

(iii) 最後に、 $0 > w^{H_{j2}} > w^{H_{j1}}$  の場合、 $G_j$  の曲面は、図3.1において  $G$  の符号を逆にしたものになっている。

$G_j > 0$  の領域では、 $\Delta z_1 < 0$ 、 $\Delta z_2 < 0$ 。

$G_j < 0$  の領域では、 $\Delta z_1 > 0$ 、 $\Delta z_2 > 0$ 。

よって、 $\Delta z_k$  の方向は、図3.1の矢印と一致する。誤差曲面の  $w^{1H_{ij}}$  軸での断面は、図3.2と同様である。

結局、誤差曲面の  $w^{1H_{ij}}$  軸での断面は、 $G_{bc}$  と  $G_{ca}$  との間は極小は  $G_{ab}$  においてのみである。 $G_{bc}$  および  $G_{ca}$  の外側の領域は非収束領域である。このような領域は、 $z_k$  の値が0か1に近い。これは、 $w^{1H_{ij}}$  の値が、非常に大きいか、非常に小さいときである。

出力関数をロジスティック曲線から、一般の出力関数の場合に拡張しても、(2)単調非減少性から、微係数が0となる点は  $G_{ab}$  以外はその前後で微係数の符号は変化せず、 $G = 0$  の曲線の形がかわっても、そのトポロジカルな性質は変わらない。よって、いくつかの停留点が生じても、収束先は変わらない。

(証明終わり)

以上から、(c)、(d)の領域へ入らないようにするためには、目標出力  $t_k$  を0あるいは1近くにとらないようにすればよい。更に、特異点解消を行えば、(c)、(d)の領域が縮小されることを

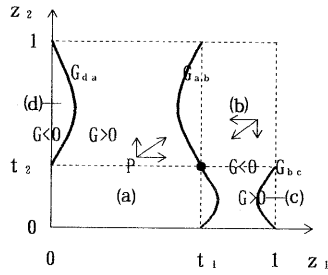


図3.1 Gの曲面と $\Delta z_k$ の方向

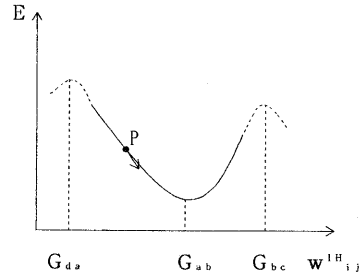


図3.2 誤差曲面の断面

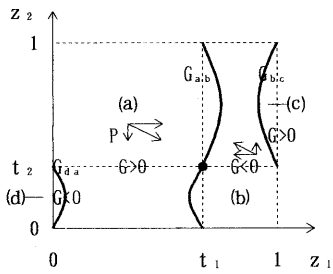


図3.3 Gの曲面と $\Delta z_k$ の方向

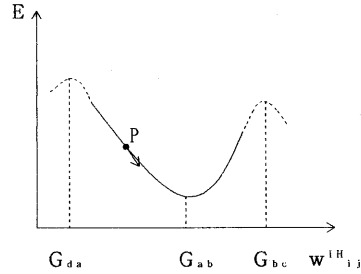


図3.4 誤差曲面の断面

示すことができる。また、 $G_{a,b}$ 付近での $\Delta w^H_{i,j}$ の方向が、極小点 $(t_1, t_2)$ を向いていないことが収束を遅くする原因となる。

$K=3$ になっても、 $G_j$ の曲面は〔補題3〕の(a), (b)に当たる主領域と、(c), (d)に当たる周辺領域をもつ。周辺領域は、出力関数がロジスティック曲線するとき、高々 $2^{K-1}$ 個である。

#### 4. 音声生成システムの実験的検討

##### 4.1 音声生成システム

パターン情報拡散化協調学習方式を用いた音声生成システムは、文献(2)の図3.1のように構成される。本報告では、時間情報をもった音韻系列からアナログ量の音声パラメータ(音源パラメータと声道パラメータ)を生成する音声パラメータ生成部について検討する。

音声パラメータ生成部は、4層のニューラルネットワークであり、中間層と出力層はL段の並列化されたネットワークで構成される。まず、実音声を音源・声道パラメータ自動抽出システムで分析し、フレーム単位に音韻を教師が割り当てる。この音韻系列と音源・声道パラメータ系列の組を入力層と各段の出力層に学習データとして与える。各段では学習入力環境が少しずつ異なっている。学習後は、出力層および最終出力層の出力値を二乗誤差で評価する。

##### 4.2 実験条件

入力音声：「朝早くバンガローに電報が届いた」(4.08sec)

音声パラメータ自動抽出システム：

サンプリング周期：8 KHz

フレーム長：20msec

フレーム周期：10msec

音声パラメータ生成部(基本的構成)：

入力層：I = 1,160 素子

= 40次元音韻ベクトル × 29列

中間層：J = 80素子

出力層および最終出力層：K = 13素子

= 音源パラメータ(音源パワー、

有声/無声、ピッチ)

+ 10次元PARCOR係数

中間層および出力層の段数：L = 9段

学習データセットのフレーム数：408(-28=380)

平均学習回数(特に断らない限り)：約1316回

(=500,000/380)

##### 4.3 実験結果と検討

〔中間層の素子数〕

記憶にかかわる重み係数 $\{w^H_{i,j}\}$ の個数は $I \times J$ であるが、調整できるものは中間層の素子数Jである。図4.1に中間層の素子数と(平均学習回数1316回の時の)二乗誤差との関係を示

す。SICL法の各値は収束したものである。40素子以上ではほぼ一定となっている。即ち、40あるいは60素子あれば例文を記憶するのに十分であることがわかる。

さて、SI法の値も同時に示しているが、83での学習定数の満たすべき条件式(24)に関係している。まず、20素子での二乗誤差は収束しているが、十分に学習できず、大きな二乗誤差をもっている。40~80素子のときは、同じ値の学習定数 $\alpha$ 、 $\beta$ をもちいているため、この平均学習回数ではまだ各SI法は収束しておらず、収束速度の違いが図に示されている。120素子では計算精度の限界にまで収束している。240素子では、収束に振動がみられるが、次のことから学習定数の値が大きすぎることが原因と考えられる。即ち、式(24)から、240素子での $\alpha$ のこの値は、80素子に換算すると約3倍の値に相当する。80素子の場合について、学習定数 $\alpha$ 、 $\beta$ と二乗誤差との関係<sup>(4)</sup>は検討してあり、換算した値はSI法の各段の二乗誤差が振動し始める境界の値である。

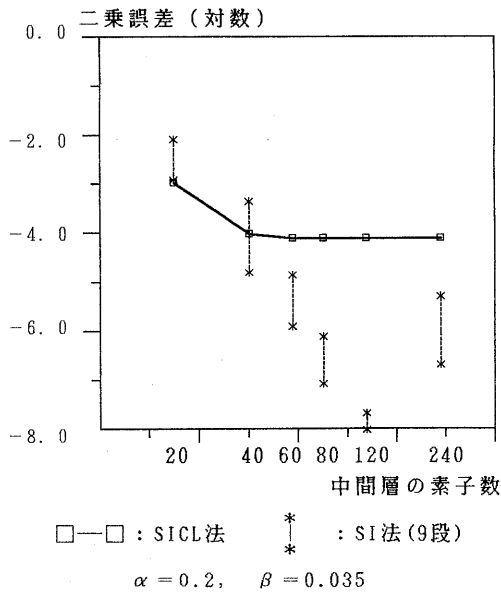


図4.1 中間層の素子数と二乗誤差

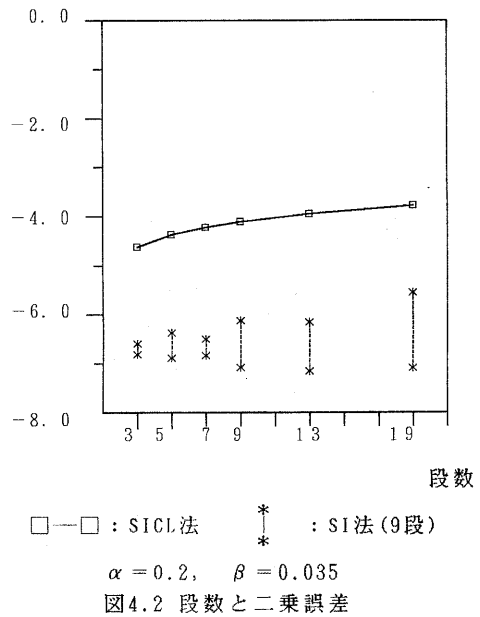
〔段数の評価〕

ネットワークを並列化し、種々の入力音韻環境を学習させることは、人の音声発声における時間的依存性を表現する一方法と考えられる。図4.2に示すように、最終出力層で平均化される時、その二乗誤差は各段のものより2桁ほど大

きくなっている。しかし、段数Lが変化しても二乗誤差の変化は小さい。

図4.1からもわかるように、SICL法の二乗誤差は、各段(SI法)の二乗誤差の違いにあまり左右されない安定した性質がある。即ち、実世界のデータを対象とする場合など、中間層の素子数が十分大きくすることが容易でないようなときでもSICL法では精度の向上が期待できる。また、収束速度も、各段がまだ収束しないうちに速く収束する。更に、多段化は、出力の音声パラメータ時系列を滑らかにする働きがあり、音声合成に好ましい性質である。

平均二乗誤差 (対数)



〔中間層の素子の活動度〕

中間層は入力を出力へ変換する働きをする。その素子の活動度の性質を調べる。表4.1には、SICL法の中央の段( $l=0$ )における中間層の全素子の入力音声全体に対する活動度の平均、分散および度数分布を示した。表4.2には、特異点解消方法を用いた改良BP法のそれらを示した。平均学習回数は約2532回である。両者ともに、各素子の活動度は0に近い値をとる(クラス1あるいは2)か、1に近い値をとる(クラス10あるいは9)かのいずれかで、中間(0.5)に近い値をとる(クラス3~8)ことは少ない。各素子の値を時系列でみると、0に近い値から1

表4.1 SICL法 ( $\ell=0$ ) の中間層の素子の活動度

平均	: 0.433			
分散 (標準偏差)	: 0.190(0.435)			
度数分布表				
クラス	中央値	度数	相対度数	
(1)	0.0~0.1	0.05	13431	0.442
(2)	0.1~0.2	0.15	1352	0.044
(3)	0.2~0.3	0.25	929	0.031
(4)	0.3~0.4	0.35	803	0.026
(5)	0.4~0.5	0.45	723	0.024
(6)	0.5~0.6	0.55	746	0.025
(7)	0.6~0.7	0.65	762	0.025
(8)	0.7~0.8	0.75	899	0.030
(9)	0.8~0.9	0.85	1227	0.040
(10)	0.9~1.0	0.95	9528	0.313
		計	30400	1.000

表4.2 改良BP法の中間層の素子の活動度

平均	: 0.539			
分散 (標準偏差)	: 0.190(0.436)			
度数分布表				
クラス	中央値	度数	相対度数	
(1)	0.0~0.1	0.05	10177	0.335
(2)	0.1~0.2	0.15	1319	0.043
(3)	0.2~0.3	0.25	923	0.030
(4)	0.3~0.4	0.35	761	0.025
(5)	0.4~0.5	0.45	747	0.025
(6)	0.5~0.6	0.55	741	0.024
(7)	0.6~0.7	0.65	826	0.027
(8)	0.7~0.8	0.75	968	0.032
(9)	0.8~0.9	0.85	1369	0.045
(10)	0.9~1.0	0.95	12569	0.414
		計	30400	1.000

に近い値へ移るとき、あるいはその逆のときに過渡的に中間の値をとることが多いことがわかる。80素子の中で、クラス1あるいは10に入る平均値をもつ素子の数は、SICL法が11素子、BP法が6素子である。その他の素子は、音韻環境によって、2つのクラスをスイッチするように働く。以上から、精度良く収束する場合のSICL法と改良BP法の中間層の素子の活動度は、よく似た動作をすることがわかった。中間層から出力層への重み係数  $w^{no}_{jk}$  を固定しても、学習させても、その活動度の分布に違いがほとんど認められないことは興味あることである。更に分析を進めたい。

## 5. むすび

我々の提案したパターン情報拡散化協調学習方式は、入力層と中間層の間の重み係数だけを学習する。理論的検討として、並列化の基本となる3層ニューラルネットワークのダイナミクスは、出力層の素子数によって、1素子のときと2素子以上とは異なっていることがわかった。即ち、1素子のときは誤差曲面が単峰であり、最小値へ収束するが、2素子以上では、重み係数軸で断面を調べると、ネットワークの状態により、重み係数が収束しない領域が存在する場合がある。また、ネットワークの規模と学習定数との間の関係が明らかになった。

実験的検討として、記憶能力の評価、ネットワークの並列化による効果、中間層の活動度の分布を調べた。多段化することにより、安定し

た出力を得ることができることがわかった。

今後の課題として、理論的には、多段化による最終出力の統計的性質の検討、BP法とのダイナミクスの比較検討がある。実験的には、中間層の素子の活動度を音韻との関係で分析することなどがある。これらの検討を通して、ニューラルネットワークの設計方針を明らかにし、任意音声生成を実現することを目指している。

## 〔謝辞〕

研究の機会を与えてくださり、熱心に御指導・御援助下さる国際情報社会科学研究所北川敏男会長に心より感謝申し上げます。

## 文 献

- (1) 古村, 田中: "ニューラルネットワークを用いた協調学習システムによる音声合成", 信学技報, MBE88-8 (1988-04).
- (2) 古村, 田中: "音声パターン学習・生成に関する実験的検討", 信学技報, SP88-8 (1988-05).
- (3) D. E. Rumelhart, J. L. McClelland and the PDP Research Group: "Parallel Distributed Processing", Vol. 1, The MIT Press (1987).
- (4) M. Komura and A. Tanaka: "Speech Production Using Neural Network with Cooperative Learning Mechanism", IEEE Conf. on Neural Information Processing Systems, P1.25, Denver, Nov. 28- Dec.

1, 1988. (Collected Papers of the Conf. : Advances in Neural Information Processing Systems. Vol. 1, (April 1, 1989) (to be published).

付 録

〔補題1〕の証明

まず、訓練パターンが1, 2, ..., n-1 番目まで学習したとする。この時の重み係数を  $w^{H_{ij}}(n-1)$  とする。n 番目に与えられる学習セットの入力を  $(x_i(n): i=1, \dots, I)$ , 目標出力を  $(t_k(n): k=1, \dots, K)$  とし、この学習によって求められる重み係数を  $w^{H_{ij}}(n)$  とする。

まず、n-1 番目までと n 番目までの学習によって得られるそれぞれの重み係数  $w^{H_{ij}}(n-1)$ ,  $w^{H_{ij}}(n)$  に対する中間層の素子の出力を、 $y_j(n|n-1)$ ,  $y_j(n|n)$  とし、出力層の素子の出力を  $z_k(n|n-1)$ ,  $z_k(n|n)$  とする。重み係数の変化量を  $\Delta w^{H_{ij}}(n)$  とすると、

$w^{H_{ij}}(n) = w^{H_{ij}}(n-1) + \Delta w^{H_{ij}}(n)$  (A1)  
 $\Delta w^{H_{ij}}(n)$  は、(11)で与えられる。(13), (14)を考慮して、

$$\begin{aligned} & |\Delta w^{H_{ij}}(n)| \\ & \leq \alpha |f'(u_j(n|n-1))| |x_i(n)| \\ & \quad \times \{ \sum_k |t_k(n) - z_k(n|n-1)| \\ & \quad \times |f'(v_k(n|n-1)) + \gamma| \\ & \quad \times |w^{H_{0jk}}| \} \quad (A2) \end{aligned}$$

f(u)の可微分性の仮定から、Taylor展開でき、

$$y_j(n|n) = y_j(n|n-1) + \Delta y_j(n|n) + \varepsilon_y \quad (A3)$$

ただし、

$$\begin{aligned} \Delta y_j(n|n) &= f'(u_j(n|n-1)) \\ & \quad \times \sum_i \Delta w^{H_{ij}}(n) x_i(n) \quad (A4) \end{aligned}$$

同様に、

$$z_k(n|n) = z_k(n|n-1) + \Delta z_k(n|n) + \varepsilon_z \quad (A5)$$

ただし、

$$\begin{aligned} \Delta z_k(n|n) &= f'(v_k(n|n-1)) \\ & \quad \times \sum_j w^{H_{0jk}} \Delta y_j(n|n) \quad (A6) \end{aligned}$$

よって、

$$\begin{aligned} & |\Delta z_k(n|n)| \\ & \leq \alpha |f'(v_k(n|n-1))| \\ & \quad \times \{ \sum_i |x_i(n)|^2 \} \\ & \quad \times \sum_j |w^{H_{0jk}}| |f'(u_j(n|n-1))|^2 \\ & \quad \times \sum_m |w^{H_{0jm}}| |f'(v_m(n|n-1)) + \gamma| \\ & \quad \times |t_m(n) - z_m(n|n-1)| \quad (A7) \end{aligned}$$

ここで、 $\alpha$ を適当に小さくとれば、即ち、(24)

を満たすようにとれば、(A7)から

$$|\Delta z_k(n|n)| \ll |t_k(n) - z_k(n|n-1)| \quad (A8)$$

を満たすようにすることができる。また、(A3), (A5)の右辺第3項(剰余項)は無視できる。

さて、二乗誤差を、

$$E(n|n-1) \equiv \frac{1}{2} \sum_k (t_k(n) - z_k(n|n-1))^2$$

とおく。(A5)から、

$$E(n|n) = \frac{1}{2} \sum_k (t_k(n) - z_k(n|n-1) - \Delta z_k(n|n))^2$$

n 番目の学習による二乗誤差の差は、

$$\begin{aligned} \Delta E(n|n) &\equiv E(n|n) - E(n|n-1) \\ &= - \sum_k \{ (t_k(n) - z_k(n|n-1)) \\ & \quad \times \Delta z_k(n|n) - \frac{1}{2} \Delta z_k(n|n)^2 \} \quad (A9) \end{aligned}$$

$\alpha$ を(24)のようにとれば、(A9)の第2項を無視でき、 $\Delta E(n|n)$ の符号は次式の符号を調べればよい。

$$\Delta E(n|n) = - \sum_k (t_k(n) - z_k(n|n-1)) \times \Delta z_k(n|n) \quad (A10)$$

まず、(A6)へ(A4)を代入し、この両辺に  $t_k(n) - z_k(n|n-1)$ を掛けて、kについて和をとると、 $\delta^0_k(n)$ ,  $\delta^H_j(n)$ と(11)を考慮すれば、

$$\begin{aligned} \Delta E(n|n) &= - \sum_i \sum_j \alpha (\delta^H_j(n) x_i(n))^2 \\ & \quad + \gamma \sum_i \sum_j \sum_k \alpha \delta^0_k(n) w^{H_{0jk}} \\ & \quad \quad \quad \times \delta^H_j(n) (x_i(n))^2 \\ & \quad + \gamma \sum_i \sum_j \sum_k \alpha (t_k(n) - z_k(n|n-1)) \\ & \quad \quad \quad \times w^{H_{0jk}} f'(u_j(n|n-1)) \\ & \quad \quad \quad \times \delta^H_j(n) (x_i(n))^2 \quad (A11) \end{aligned}$$

ここで、 $\gamma$ の値を(25)のようにとれば、第一項の符号と  $\Delta E(n|n)$ の符号とを一致するようにできる。即ち、

$$\Delta E(n|n) = - \sum_i \sum_j \alpha (\delta^H_j(n) x_i(n))^2$$

の符号を考えればよい。よって、いかなる  $x_i(n)$  に対しても常に  $\Delta E(n|n) \leq 0$ 。(但し  $\Delta E(n|n) = 0$  とはなるのは、すべての n に対して  $x_i(n) = 0$  とはならないから、すべての n に対して  $\delta^H_j(n) = 0$  となる時、即ち誤差伝播量が 0 となる時である。)

以上より、十分小さな  $\gamma$ ,  $\alpha$ をとれば、 $|\Delta y_j(n|n)|$ ,  $|\Delta z_k(n|n)|$ を十分小さくすることができる。その時、 $\Delta E(n|n) \leq 0$ となる。即ち、二乗誤差を減少させる方向に常に  $\Delta w^{H_{ij}}(n)$ をとることができることが示された。

(証明終わり)