ヒューマンインタフェース 56－4
（1994．9．8）

発声訓練システムのための音声可視化手法

長山 格，　赤松 則男，吉野 俊樹

徳島大学工学部知能情報工学科

本報告では，発声訓練システムにおける音声パターンの可視化にニューラルネットワークを応用し，利用者の容易な理解を可能にする新しい手法を提案する．通常，言語障害教育やリハビリテーション等に用いられる発声訓練システムは，音声パターンを波形やスペクトルパターンなど複雑な形式で表示することが多い．これらは音声情報の全体像を表示するのに適しているが，複雑であるために規範パターンと比較しにくいという難点がある．これに対して，本報告で提案する手法は，多次元パターンとして表される音声パターンを平面上の点で表現することにより，利用者の理解を容易にすることができる．本報告では，提案手法について述べるとともに，日本語母音の可視化に適用した結果を示す．

Phonetic Visualization for Speech Training System by Using Neural Network

Itaru Nagayama, Norio Akamatsu, Toshiki Yoshino

Faculty of Engineering, University of Tokushima
Tokushima－shi, 770 Japan

A speech training system and a voice visualizer are useful machines for speech impairment patients. The speech training system indicates the vocal pattern(spectrum pattern or wave－form) on display for patients to train their pronunciation. But human being cannot recognize the complex multidimensional representation of patterns as like spectrum pattern or wave －form. Therefore, the man－machine interface of speech training system has a green hand. In this paper, we propose a Neuro－NLM, to build up the new interface for speech training system. We have investigated on a basic system which can easily visualize 5 Japanese vowels /A/,/I/,/U/,/E/,/O/. The system displays the multidimensional pattern of vowel onto two dimensional (2－D) space corresponding to the original pattern. User can compare his vowel pattern distribution in the 2－D plane with standard one during his training stage.

1．まえがき

先天的事由や疾病などにより，言語，聴覚機能に支障を持つ患者は疾病統計上少なからぬ数を占める．これらの患者に対し，速やかな言語発声機能の回復を図るための発声訓練が行われる．すなわち，施療者によって発声のための口蓋調節や調息法などの指導が施されるが，被訓練者にとっては発声状態の把握が容易ではない．このとき，音声情報を視覚化することによって訓練効果を高めることができるので，発声リハビリシステムとして音声直視装置や発声訓練システムが開発されている．しかし，これらのシステムは音声波形やスペクトルパターンをＶＤＴ画面上で表示する形式のものであり，規範パターンとの比較や差異の理解が容易ではない．特に，小学生以下の児童が患者である場合，複雑なパターンの相違を理解することは難しく，訓練意欲を減退させる要因ともなる．従って，発声パターンの表示方法を改善する必要がある．一般に，波形やスペクトルはｎ次元ベクトルＸ n=(x1,x2,⋯,xn) で表すことができる．しかし，人間にとってｎ次元の多次元パターン群の分布や相違をそのままの形で認識することは困難である．

本報告では，ＮＬＭ法（Non−Linear Mapping）[1] とニューラルネットワークを用いてｎ次元パターンを２次元平面に写像することにより，複雑な発声パターンの相違を容易に理解できる手法を提案する．これをニューロＮＬＭと呼ぶ．以下では，ニューロＮＬＭの実現方法を述べ，日本語母音の可視化に適用した結果を示す．

2．ニューロＮＬＭ

2．1　ＮＬＭ法

ｎ次元パターンを，２次元（または３次元）空間で表すために，ＮＬＭ法 (Non−Linear Mapping) が用いられる．ＮＬＭ法は，最急降下法などの最適化法を用いて，ｎ次元空間におけるパターン間の距離関係を保存するように２次元空間における座標を決定するものである．すなわち，ｎ (>2) 次元空間におけるｋ個のパターンをＸ i,(i=1,⋯,k) とし，これをｐ (=2) 次元空間上に写像したものをＹ i,(i=1,⋯,k) とする．ｎ次元空間でのパターンＸ i とＸ j 間の距離を d1＝dist [Ｘ i, Ｘ j] で定義し，ｐ次元空間上での対応するパターンＹ i，Ｙ j 間の距離を d2＝dist [Ｙ i, Ｙ j] で定義する．ただし，dist[u,v] は，ベクトルｕ，ｖ間のユークリッド距離とする．ｐ次元空間におけるパターンＹ i,(i=1,2,⋯,k) の座標をランダムに選び，初期値とする．

$$\begin{aligned} Y1&=(y11,\cdots,y1p)\\ Y2&=(y21,\cdots,y2p)\\ &\cdots\cdots\\ Yk&=(yk1,\cdots,ykp) \end{aligned} \qquad (1)$$

このように決めたとき，ｎ次元空間とｐ次元空間における各パターンの配置誤差を示す評価関数Ｅを次式で定義する．

$$E=(1/\Sigma\,[d1])\cdot\Sigma\,([d1-d2]^2/d1) \qquad (2)$$

ここで，最急降下法などの最適化法を用いて評価関数Ｅの値を最小化し，ｐ次元空間上のパターンＹ i,(i=1,2,⋯,k) の座標を決定する．

ＮＬＭは，上述の最適化を行うことによって実行される．すなわち，あらかじめ与えられたパターン群の非線形写像として２次元平面上の座標が得られる．それゆえ，新たに未知のパターンが与えられると再び

最適化計算を実行する必要があり，直ちに２次元平面上の座標を得ることはできない．

２．２　ニューロＮＬＭ

上記のＮＬＭにおける最適化によって，ｎ次元ベクトルと２次元ベクトル間の非線形変換が行われる．ここで，ニューラルネットワーク（以下，ＮＮと略記する）を用いて，ｎ次元パターン群を入力パターン，２次元パターン群を教師パターンとして学習を実行すると，ＮＮはこれらの非線形写像関係を学習する．すなわち，ＮＮは，ｎ次元空間から２次元空間への写像を学習する．すでに，中間層に十分な素子数を持つＮＮは任意のマッピングが可能であることが知られているので，適当な構造を持つＮＮを用いることにより，高次元から低次元への非線形写像を学習することができる．従って，学習後のＮＮを用いることによって，ｎ次元の未知入力パターンに対する２次元ベクトルを出力することが可能である．すなわち，ニューロＮＬＭは，未知パターンへの対応が可能である点に特徴がある．図１にニューロＮＬＭの概念図を示す．

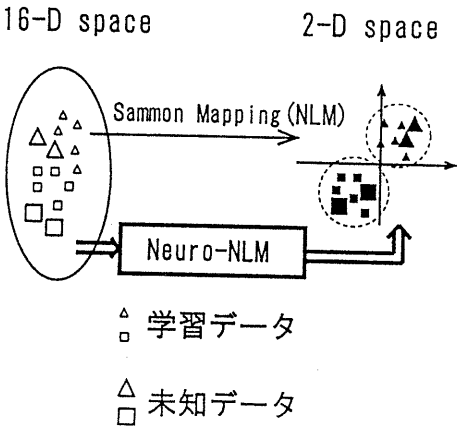

図１　ニューロＮＬＭの概念図

３．実験

３．１　日本語母音の可視化

単音節母音 /A/,/I/,/U/,/E/,/O/ を対象として実験を行う．これらの母音について，それぞれ２０個ずつ合計１００個の音声データを用いる．各音声データはＦＦＴを行って周波数データに変換する．一般に，音声の波形データは変動が激しいため，より安定であると考えられる周波数データを用いる．実験では，以下の手順を行って学習データを作成する．

(1) 取り出された母音音声区間の差分データを求めて直流成分を除去し，高周波成分を強調する．

(2) ハミング窓を掛け，ＦＦＴによりパワースペクトルを求める．

(3) パワースペクトルから，メルスケール化した周波数帯域毎に平均パワーを求める．

(4) 各帯域データをデシベル化した後，－８０～０ｄＢの区間を－0.9～0.9の範囲に正規化する．

ここで，(3) において，１６個の周波数帯域について計算するので，各母音の音声データは最終的に１６次元のベクトルパターンとなる．

作成した１６次元の音声データを，ＮＬＭによって２次元平面へマッピングした．２次元平面へのマッピング結果を図２に示す．明らかに，１００個の音声パターンは，それぞれの母音領域に分離されていることがわかる．

３．２　ＮＮによる学習

３．１で得られた１６次元ベクトルパター

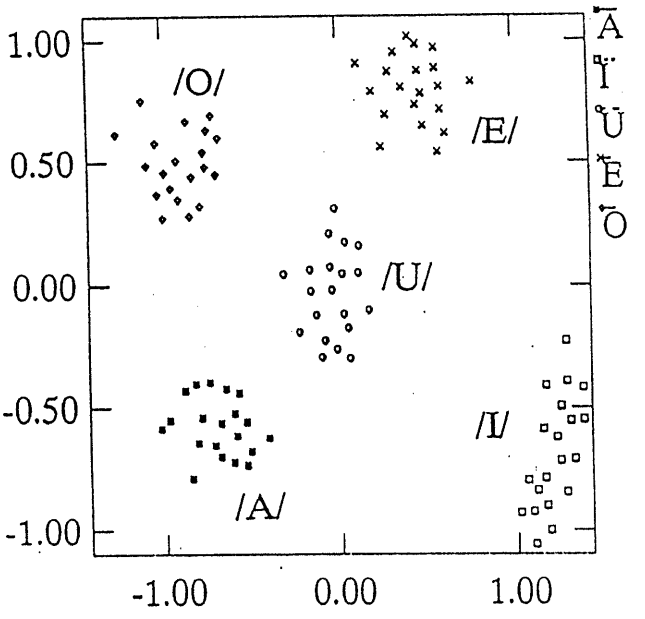

図２　ＮＬＭによる日本語母音の２次元平面への写像結果

ンＸ i;(i=1,…,100) を入力パターンとし，対応する２次元ベクトルパターンＹ i;(i=1,…,100) を教師パターンとしてＮＮで学習を行った．このとき，２次元平面でのパターンＹ i;(i=1,…,100) の分布範囲が |1.0| を超える場合は，全体を −1.0 ～ 1.0 の範囲に正規化する．学習法としてモーメント付きＢＰ法を用い，学習率を 0.037，モーメントを 0.45 とし，学習回数２万回まで学習を行った．ＮＮは３層構造であり，中間層のユニット数について比較するため表１に示す４種類のネットワークについて調べた．中間層および出力層における各ユニットの入出力関数は y=tanh(x);(− 1.0<y<1.0) を用いる．

表１　実験に用いたＮＮ

No.	入力層－中間層－出力層
N1	１６－４－２
N2	１６－８－２
N3	１６－１２－２
N4	１６－１６－２

４．実験結果

４．１　学習曲線

ＮＮによる非線形写像学習の学習曲線を図３に示す．縦軸は平均自乗誤差，横軸は学習回数を表す．明らかに，ＮＮの中間層

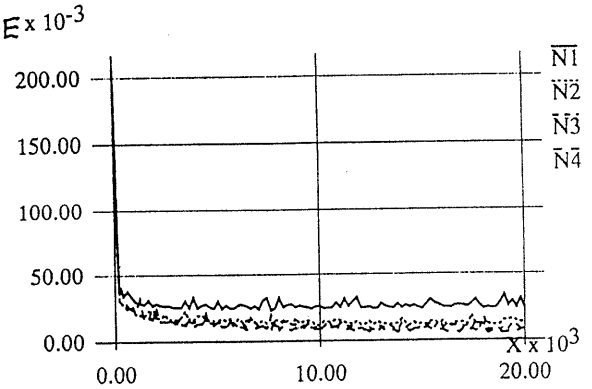

図３　ＮＮの学習曲線

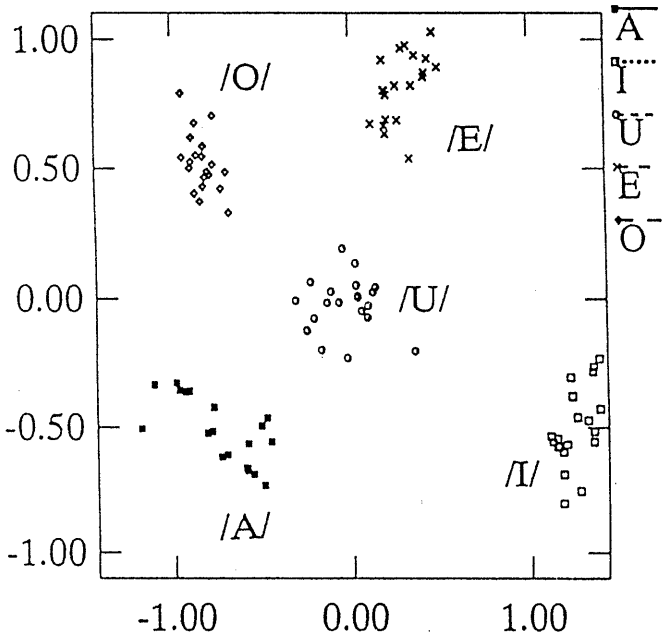

図４　ニューラルネットワークによる日本語母音テストデータの２次元平面への写像結果

ユニット数が多いほど，学習誤差が小さくなっていることがわかる．

４．２　テストデータへの適用

未学習の音声データを新しく採取し，3.1と同様に各母音について２０個ずつ合計１００個作成し，テストデータとする．このとき，学習データを採取した人物Ａとは別の人物Ｂの音声を採取した．学習終了後のＮＮにテストデータを入力したときの出力を示す．このとき，実験に用いた４種類のＮＮについて同様の結果を得たので，図４には中間層ユニット数８個の場合について示す．明らかに，各母音データは，それぞれの領域付近にマッピングされていることがわかる．これは，未学習のデータに対する汎化性が獲得されていることを示す．従って，母音音声パターンに関する１６次元空間から２次元平面へのニューロＮＬＭが可能である．

４．３　近傍性の保存

ニューロＮＬＭは，各母音領域と入力データの位置関係が，２次元平面上で視覚的に理解できる点に特徴がある．このとき，各母音に対する類似性が保存されていることが重要である．すなわち，各母音に類似した入力データは，その母音領域の中心の近くへ写像されることが必要である．これを近傍性の保存と呼ぶ．本節では，近傍性の保存について述べる．ここで，１６次元空間におけるある領域ａの中心をＣａ，領域ａに属するパターンをＸａとし，２次元平面上の対応する領域をａ’，ａ’の中心をＣａ’，領域ａ’に属するパターンをＹａとする．近傍性の保存とは，dist[Ｃａ, Ｘａ] と dist[Ｃａ’, Ｙａ] が比例することである．

近傍性の保存について調べる．まず，元の１６次元空間において各母音パターンの平均パターン（Ｃａ, Ｃｉ, Ｃｕ, Ｃｅ, Ｃｏ；領域の中心に相当する．）を作成する．次い

で，各平均パターンから半径δの超球表面上に分布する近傍性テストデータを３０個ずつ合計１５０個作成する．近傍性テストデータと各平均パターン（Ｃａ, Ｃｉ, Ｃｕ, Ｃｅ, Ｃｏ）をニューロＮＬＭによって２次元平面へマッピングし，dist[Ｃａ’, Ｙａ]を求める．図５に結果を示す．横軸はδを表し，縦軸はdist[Ｃｋ’, Ｙｋ];(k=a,i,u,e,o)の平均値を表す．

明らかに，１６次元空間での距離と２次元平面での距離が比例的関係にあることがわかる．すなわち，ニューロＮＬＭは音声パターンの類似性を保存することが可能であり，発声訓練における音声の比較に用いることができる．

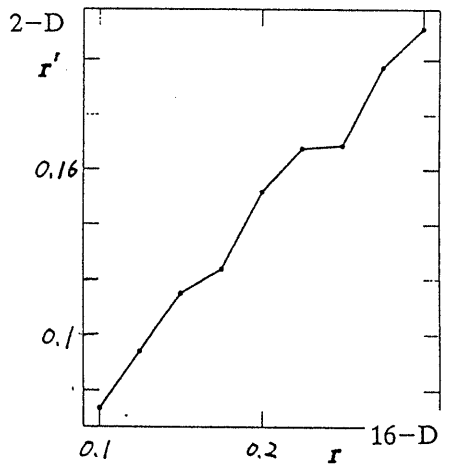

図５　ニューロＮＬＭにおける近傍性の保存

４．４　写像の歪曲性

高次元空間から低次元空間へ非線形写像を行った場合，その非線形性により写像空間が歪曲することが考えられる．すなわち，元の１６次元空間での直線が２次元空間では大きく歪曲した曲線になることが考えられる．しかし，発声訓練では規範パターンと入力パターンとの比較を行うことが重要であるため，極端な歪曲が生じることは錯覚をもたらす恐れがある．一般に，非線形写像では，その性質から写像に多少の曲がりが生じざるを得ないが，極端に歪曲した写像は好ましくない．従って，写像の歪曲性を検討する必要がある．

ここでは，１６次元空間での直線が，ニューロＮＬＭによる非線形変換を経た２次元空間でどのような状態に変換されるかを観測することによって写像の歪曲性を調べる．１６次元空間中の２点Ａ（a1,a2,…,a16）とＢ（b1,b2,…,b16）を結ぶ直線Ｌは次式で表される．

$$L=(1-t)A + tB,\ (0 \leqq t \leqq 1) \quad (3)$$

従って，ｔに対するＬの座標を入力パターンとするＮＮの出力を観測する．学習に用いた母音音声データの平均パターン間を結ぶ直線は５Ｃ２ (=10) 本あるので，これらのニューロＮＬＭによる写像を調べることにより，歪曲性の大まかな様子がわかる．

実験で用いた４種のＮＮに対する歪曲性を観察した結果を図６に示す．図６より，明らかにニューロＮＬＭは非線形写像の実現にあたり，写像空間を歪めていることがわかる．また，中間層ユニット数が多いほど曲線の歪みが多くなっている．図３と併せて考慮すると，学習曲線の誤差が小さいほど歪みが大きくなる傾向にあることから，学習精度が高いことは必ずしも音声パターンの比較に適しているとはいえない．従って，音声パターンの位置関係を正しく把握するためには歪曲性の小さいＮＮが望ましい．すなわち，表１のＮＮ構造では，中間層ユニット数４個のＮＮが最も歪曲性が少ないニューロＮＬＭが可能である．

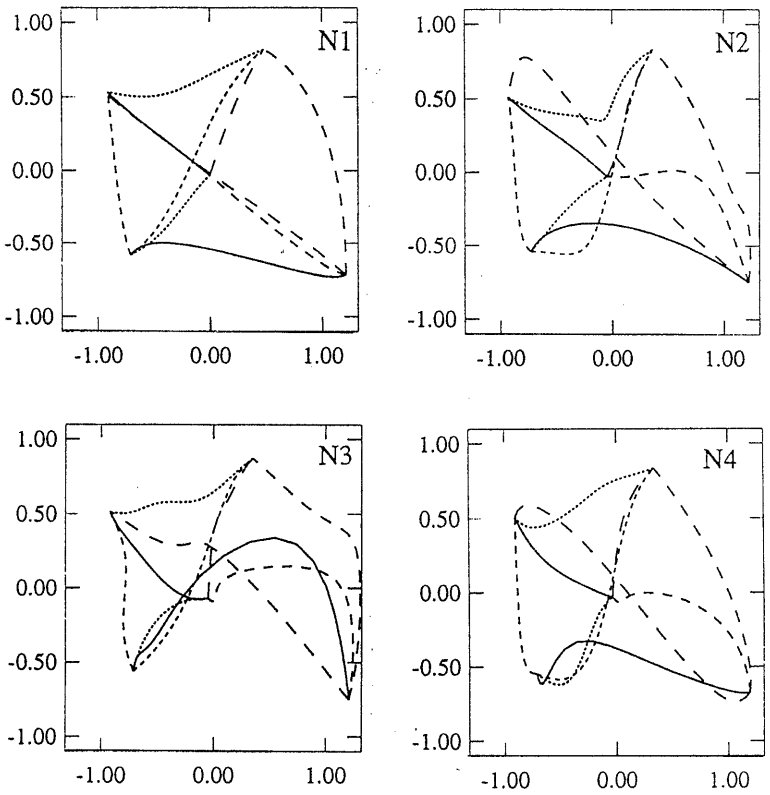

図６　ニューロＮＬＭの空間歪曲性

５．まとめ

本論文では，発声訓練システムにおけるパターン表示方式の改善を目的とするニューロＮＬＭを提案した．ニューロＮＬＭによって多次元パターン群を２次元平面上にマッピングすれば，視覚的に理解し易くなる．そのため，利用者に対する効果的な訓練教育が可能になると思われる．今後は，①ニューロＮＬＭの性質についてさらに詳しく検討すること，②子音音声の表示への拡張，及び③ニューロＮＬＭを用いた訓練システムの設計，試作と実際の教育訓練効果について調べる方針である．

参考文献

[1] Sammon J.W.: “A Nonlinear Mapping for Data Structure Analysis,” ,IEEE Trans. on Computers,C−18,5,pp.401−409(1969).

[2] Chang C.L.,Lee R.C.T.:"A Heuristic Relaxation Method for Nonlinear Mapping in Cluster Analysis,",IEEE Trans. on Systems,Man and Cybernetics,SMC−2,pp.197−200(1973).

[3] Rumelhart D.E., McClelland J.L. and the PDP Research Group:"Parallel Distributed Processing," vol.1,MIT Press(1986).

[4] Lippmann R.P.:"Pattern classification using neural networks,"IEEE.comm,pp. 47−64,Nov.(1989).

[5] Levitt H.:"Technology and speech training:An affair to remember,",The Volta Review,91(5),pp.1−6(1989).

[6] Watanabe A.,Ueda Y.,Shigenaga A.:" Color display system for connected speech to be used for the hearing impaired",IEEE Trans. ASSP.vol.33, No.1,1985.