

母音認識ニューラルネットによる母音の弁別素性の抽出

中尾 充宏 北澤 茂良

静岡大学 工学部 情報知識工学科
〒432 静岡県 浜松市 城北 3-5-1

あらまし従来の研究では、ニューラルネットは識別用であっても、特徴抽出や解析に用いられることは成功していない。ブラックボックスとなっている隠れ層の解析を困難にしている原因は重み係数の初期値をランダムに与えていることとニューラルネットの構造が冗長であるために情報が集約されずに拡散化していることにある。本稿ではこれらの現象を解析し、集約されたネット構造を抽出することで、ニューラルネットによる特徴抽出および、情報構造の解析過程を明らかにする。ここでは、母音の弁別素性理論にニューラルネットを適用し、音声学上の仮説としての弁別素性をニューラルネットの形成によって抽出する過程を例にとった。

和文キーワード ニューラルネット、弁別素性、母音認識、構造解析、特徴抽出

Distinctive feature extraction by back propagation neural networks for recognition of Japanese vowels

Mitsuhiro Nakao and Shigeyosi Kitazawa

Department of Computer Science, Faculty of Engineering, Shizuoka University
3-5-1, Johoku, Hamamatsu, Shizuoka, 432, Japan

Abstract Back-propagation neural networks, though are powerful in discrimination, are less useful in feature extraction. The difficulties come from the black-box characteristics; connections and units are redundantly equipped to help faster learning, and the inner hidden units are developed starting from random initial weights. Those learned hidden-networks show distributed features, which produces a mere vague image instead of a clearcut of the neural network's function. This paper extracted the essential skeleton of the network and therefore the feature structure. We applied this principle to the distinctive feature theory of Japanese vowel, as a hypothesis of phonology. We could show physical evidence by composing a neural network to discriminate vowels based on the combination of distinctive features.

英文 key words neural networks, distinctive feature, recognition of Japanese vowels, structure analysis, feature extraction

1 はじめに

かつてほどではないが最近の研究でも、ニューラルネットはゆらぎやあいまい性に対して比較的強いという利点から、音声認識や画像処理のパターン認識など、識別によく用いられる。しかし、ニューラルネットは識別用であっても、特徴抽出、解析に用いられることは成功していない。階層型ニューラルネットの入力層と出力層の間の層は、隠れ層という名で呼ばれるように、入力、出力の関係のみで規定されるブラックボックスでしかなかった。解析を困難にしている原因は、重み係数の初期値をランダムに与えていることと、冗長なニューラルネットの構造にある。そのためにユニットの機能がバラバラになり、情報が集約されずに拡散化している。本研究ではこれらの現象を解析し、集約されたネット構造を抽出することによって、ニューラルネットによる特徴抽出過程を明らかにし、情報構造の解析を行なった。

例として、母音の弁別素性理論にニューラルネットを適用し、音声学上の仮説としての弁別素性をニューラルネットの形成によって抽出した。

2 弁別素性

表 1: 5母音の弁別素性

弁別素性	a	i	u	e	o
compact(+)/diffuse(-)	+	-	-	+	+
grave(+)/acute(-)	+	-	+	-	+
flat(+)/plain(-)	-	/	+	/	+

表1に示す。なお、/i,e/はflat/plainの特徴を持たない。但し、素性理論は種々提案されており、特に日本語については、この体系の妥当性も議論の余地はあるが、ここでは一例としてJakobsonらの体系を用いた。

弁別素性理論では、基本となる弁別素性の組が何組か存在し、その弁別素性が複数組合わさって、全音素が成り立っていると考えている。本研究では日本語5母音と、それに関係のある弁別素性、compact(口開)/diffuse(口閉)、grave(後舌)/acute(前舌)、flat(円唇)/plain(非円唇)について扱う。()の中は、各素性の調音生理的な意味を表している。Jakobsonらの弁別素性理論による5母音の体系を

3 ニューラルネット

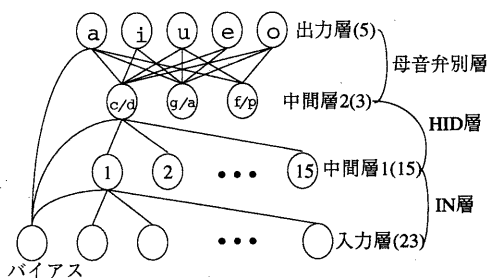


図 1: 弁別素性による母音認識ネット

本研究では、図1の階層型ニューラルネットを用いる。階層型を用いた理由としては、識別的機能を実現する、学習アルゴリズムが確立されていて効率的学習法が存在する、階層構造であるので機能の解析が容易、があげられる。図1のニューラルネットの構造の中に我々の意図する所を仕組んでいる。構造の各層の説明は、次の通りである。入力層には母音のスペクトルパターンを与える。中間層1(隠れ層)では、母音スペクトルパターンのクラスタリングを行なう。但し、ユニット数15が十分かどうかについては、現在、実験中である。素性抽出の前処理をこの層で行なう。一層で十分かに

については後述するように、素性を直接抽出した実験で確認している。中間層2は3ユニットからなり、弁別素性(左のユニットからcompact/diffuse、grave/acute、flat/plain)に対応させることを意図している。これは、5母音を3ビットの符号器によって符号化する過程であると見なすことができる。入力層からこの層に至る過程で弁別素性の識別器を形成する。この識別器の形成実験については後述する。以下の層までは全結合の前向きネットワークであり、何の制約もない。出力層は、日本語の5母音(左のユニットから/a,i,u,e,o/)に対応して、0、1とカテゴリカルに教師データを与えて、誤差逆伝搬学習する。弁別素性層(中間層2)と母音識別層(出力層)との間(この間を母音弁別層と呼ぶ)の関係には制約を与えている。すなわち、これらの層の各ユニットはあらかじめ分担する

カテゴリが決まっている。このため、flat/plainユニットと母音/i,e/の間に接続は無いということで役割が決定している。他の2つの素性ユニットについても、結合の重み係数を固定、または初期値の与え方によって、あらかじめ決定している。

4 ニューラルネットの学習による重みの形成

音声学上の仮説としての弁別素性がもしも存在するのならば、ニューラルネットの学習によって、入出力関係としての母音識別が可能になると同時に、弁別素性層(中間層2)に、素性対に対応するユニットが形成されるはずである。この際、制約条件無しに全く自由に学習させたのでは、ネットワークの冗長性と情報の拡散性によって、意図する素性ユニットが形成されるのは偶然に過ぎない。このため、ここに最低限の恣意性を導入する。

4.1 母音弁別層のみの形成

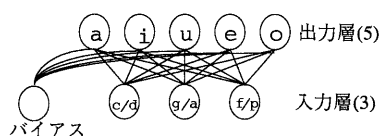


図2: 母音弁別層ネット

まずは、母音弁別層部分のネット(図2)を用いて、弁別素性の組合せに基づく母音認識学習を行なう。すなわち表1に対応して+素性に0から0.5、-素性に-0.5から0のランダムな値を入力として与え、出力には対応する母音のみ1とし他を0とする条件で学習した。その結果できたネットの重み係数は表2の初期値1に与えた値である。

4.2 素性別直接抽出

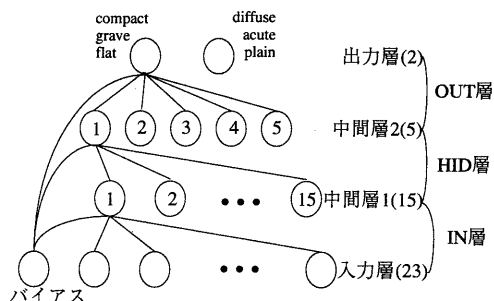


図3: 素性直接抽出ネット

次に図3のネットを用い、素性ごとにスペクトルから直接、弁別素性を抽出する学習を行なった。学習後(学習回数500回)の重み係数値の結果は、compact/diffuseが図5、grave/acuteが図6、flat/plainが図7で、(a)、(b)、(c)はそれぞれOUT層、HID層、IN層の図である。三つの認識率はそれぞれ、97.9%、99.2%、98.5%であり、スペクトルから素性を抽出できていると思われる。図5の(a)では中間層2の1、5番ユニットの値が大きく、(d)にそのユニットのHID層の重み係数値を示す。さらに(d)では中間層1の5、13番ユニットの値が大きく、(e)にそのユニットのIN層の重み

係数値を示す。重み係数値の絶対値が大きなユニットは、認識にとって強い特徴を持つので、(e)の線が(c)の中でも特に重みの大きな線と一致することから、これらのユニットがcompact/diffuseの特徴に強く関係していることがわかる。他の二つのgrave/acute、flat/plainについても、図6と図7から同様のことがわかる。なお、これらの結果は以前発表した方法[5]によって、重みの符号を適宜そろえて見やすく整理したものである。以上の解析から、弁別素性の抽出については入力層を除けば2層のネットワークで十分可能であることが予想される。すなわち、出力層または中間層2は冗長といえる。またIN層に形成された重み係数によって、弁別素性抽出のためのスペクトル重み付けが明らかになっている。それによると各素性対につき2種類ほどのクラスが形成されていることが読みとれる。これによって図3の中間層1のユニット数を15としたのは、素性別のクラスを統合しておお冗長なユニットを与えていることがわかる。

4.3 母音弁別層を固定

表 2: 母音弁別層に与える重みの初期値

弁別素性	a	i	u	e	o
compact(+)/diffuse(-)	+	-	-	+	+
初期値1	8.1	-11.0	-7.7	10.8	7.5
初期値2	1	-1	-1	1	1
grave(+)/acute(-)	+	-	+	-	+
初期値1	8.1	-11.1	7.7	-10.9	7.3
初期値2	1	-1	1	-1	1
flat(+)/plain(-)	-		+		+
初期値1	-8.1		7.8		7.5
初期値2	-1		1		1

初期値の与え方によって、形成されるニューラルネットの構造を制御することをねらって、図1の母音弁別層に2通りの初期値(表2)の与え方を考える。初期値1は節4.1の学習の結果できたネットの重み係数値を母音弁別層部分の重みとして与え、そのまま固定して下層の学習を行なう(学習1(固定))。本研究では学習アルゴリズムとして、DCP(Dynamic Control training Parameters)法を用いた。

4.4 小さな初期値を与えて学習

初期値2は表1に従って、+の素性には1を、-の素性には-1を初期値に与える。また初期値2による学習では、初期値の指定後は重みの修正は行なう(学習2(可変))。

5 音声資料

入学習データは、日本人成人男性82人が発声した、5母音/a,i,u,e,o/の5種類に加えて、5母音と破裂子音/p,t,k,b,d,g/の組合せCV単音節30種類の計35種類の単音節を、サンプリング周波数16kHzで採集し、最大振幅の18%以上の部分を切り出し、窓長256点を15フレーム取り出し、23次の線形予測係数から256個のパワースペクトルを作ってdB単位に直し、22区間の臨界帯域幅に区切り、60dB等ラウドネス曲線の補正を行なう。最後に-0.5~0.5以内に正規化した。

6 結果と考察

図1の母音認識による弁別素性抽出ネットを学習後(学習回数500回)の、重み係数値の結果を図8と図9に示す。(a)、(b)、(c)はそれぞれ、母音弁別層、HID層、IN層の図である。学習1(固定)と学習2(可変)の二つの認識率はそれぞれ、96.3%と97.0%であり、期待する入出力関係を満たすように結合重み係数は収束していると考えられる。図9の(a)は学習によって素性と母音間の重みの形

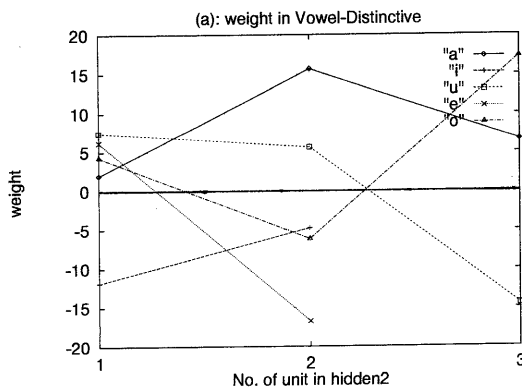


図 4: 学習後の重み係数値(初期値指定なし)

成が行なわれ、数値の大きさが素性と母音の相関の強さを表していると考えられる。例えば、/o/はcompactやflatに比べてgraveが弱い、また、同じ素性である/a,u/と比べても弱いことがわかる。また、/e/は/a,o/に比べてcompact、/u/は/i/に比べてdiffuse、/e/は/i/に比べてacuteの絶対値が大きいの、それらの特徴が比較的強く関係していると考えられる。この図から/a,i,e,o/に関しては、表1の対応どおりの符号を示しているが、/u/のflat/plain素性が一致せず、比較的中性か、またはplainの素性ということがわかる。これは、日本語の/u/は調音生理的に、唇の円めや突き出

いであることが、一つの原因と考えられる。また、IN層の重みについてみると、図5の(e)の5番ユニットと、図8と図9の(c)の5番ユニット、図6の(e)の4、13番ユニットと、図8と図9の(c)のそれぞれ15、2番ユニット、図7の(e)の5番ユニットと、図8と図9の(c)の9番ユニットの入力の重みパターンが似ており、三つの素性の直接抽出によって形成されたクラスが統合されているのがわかる。またHID層の重み(図8と図9の(b))で、c-d(compact/diffuse)の5、g-a(grave/acute)の15、2、f-p(flat/plain)の9番ユニットの値が比較的大きいことから、それぞれの素性抽出に強く関係しているユニットであることがわかる。また重みの初期値を指定せずに全く自由に与えても、認識率は96.0%となり、母音認識はできていた。母音弁別層の重みは図4である。表1のJakobsonらの素生と

は対応しない。また既存の素性理論で解釈できない。認識率でも劣る。

7 まとめ

図9に示すニューラルネットが最良の結果が得られた(43050標本の判別で97.0%、0.7%上回ったのは十分有意な差である)。そして、その時形成された弁別素性と母音の関係が日本語母音の特徴(特に/u/の特異性)を示すものであったことから、図1のニューラルネットによってスペクトルから弁別素性が抽出されていると思われる。この推

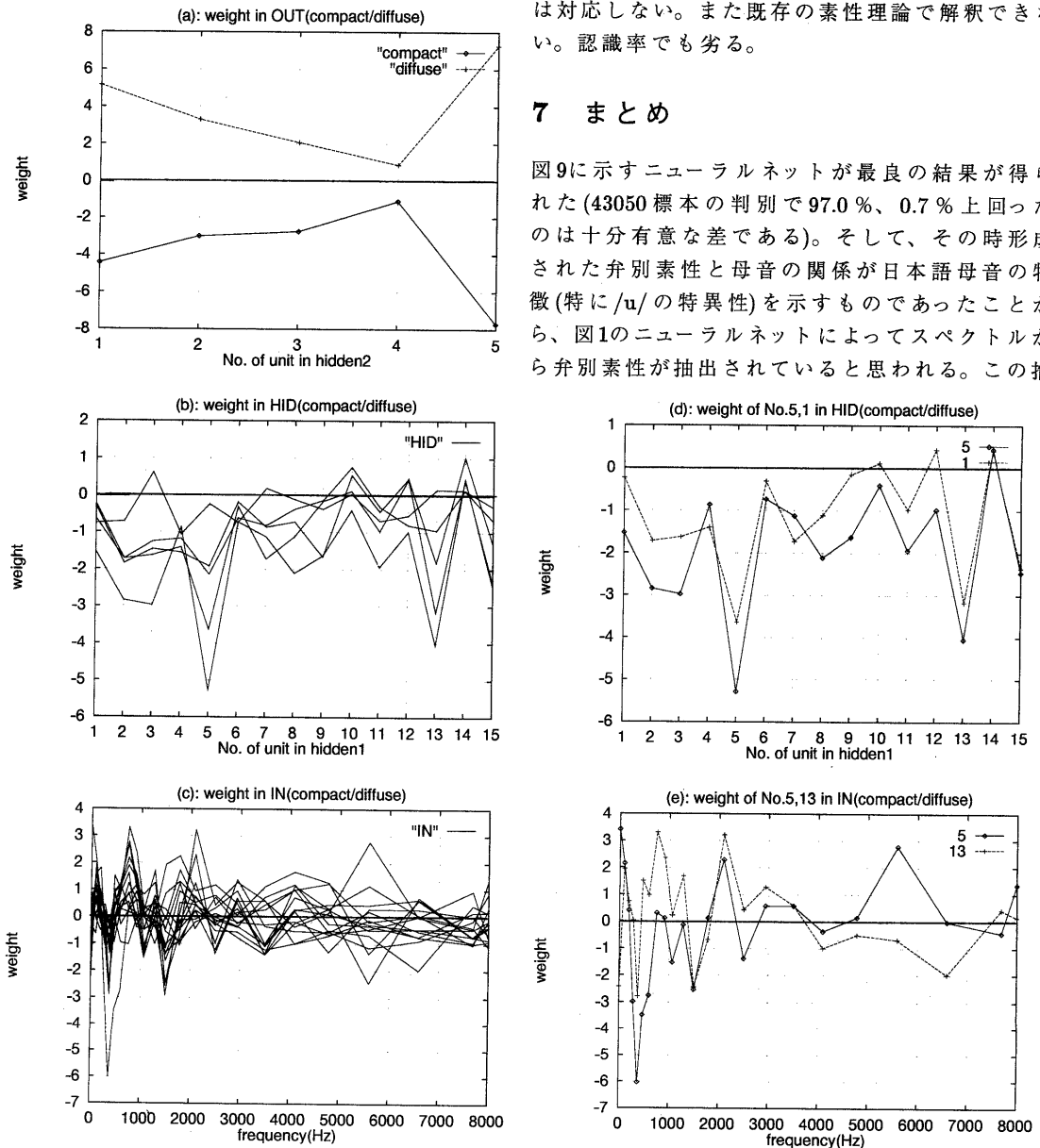


図 5: 学習後の重み係数値(compact/diffuse)

論をさらに検証するため、他言語の母音と素性の関係を明らかにし、さらに多言語の母音の相互関係をも明らかにしていくことを通じて、弁別素性理論を検証していく。

参考文献

- [1] R.Jakobson,C.G.M.Fant,M.Halle.: Preliminaries to Speech Analysis, *Technical Report No.13, Acoustics Laboratory, M.I.T.,1952.*
- [2] ヤコブソン、ファント、ハレ: 音声分析序説、2章(研究社、東京)、(1965)

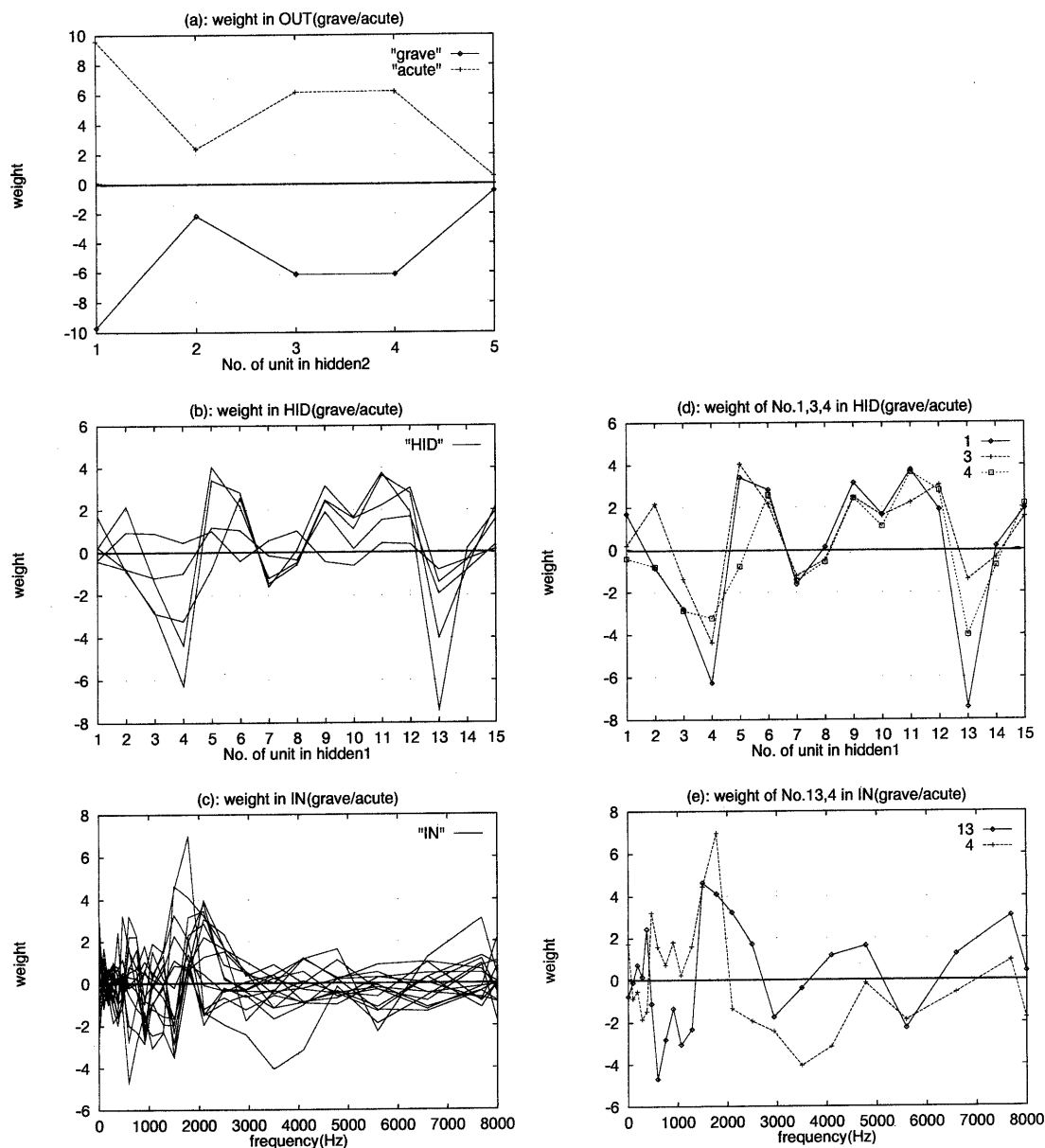


図 6: 学習後の重み係数値(grave/acute)

- [3] 北澤茂良、新村貴彦: “ニューラルネットによる母音認識のための弁別素性の特徴抽出”、静岡大学大学院工学研究科情報知識工学専攻平成3年度修士論文、(1992.3)
- [4] ATR: DCP2 *Technical Report*((株)エイ・ティ・ール視聴覚研究所、京都)、(1989.9)
- [5] 中尾充宏、北澤茂良、新村貴彦: “パーセプトロン型ニューラルネットの隠れ層の解析”、日本音響学会講演論文集、(1993.10)

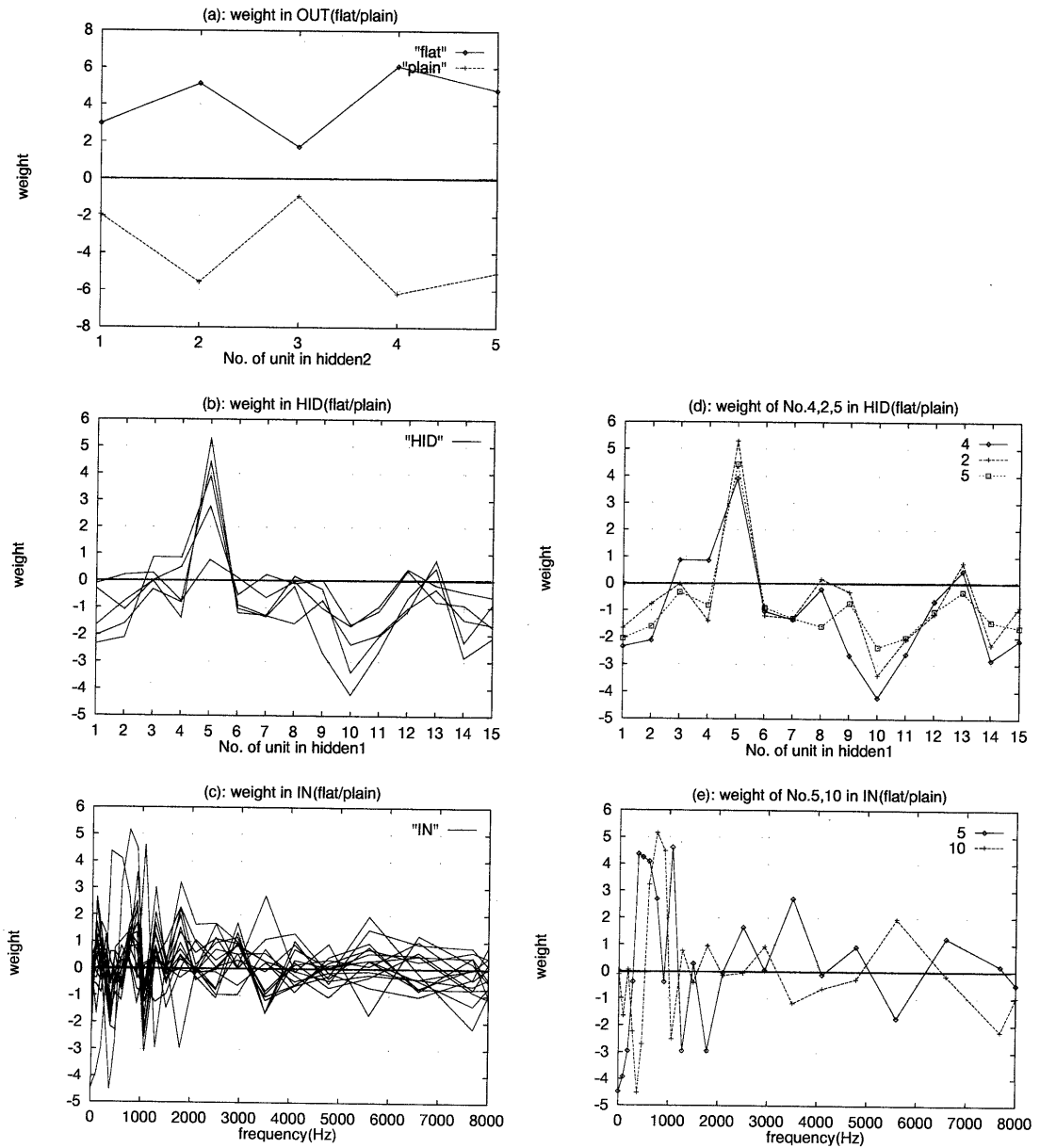


図 7: 学習後の重み係数値 (flat/plain)

謝辞

この研究成果は元静岡大学大学院工学研究科修士課程で現NTTデータ通信の新村貴彦氏の行っていた実験に基づき改良発展させたものである。弁別素成理論に関する種々の議論をいただいているプロヴァンス大学の西沼行博氏およびロッシ教授、ニューラルネットの学習プログラムDCP2を利用させていただいたエイティアル自動翻訳電話研究所、また、弁別素成理論の音声認識への応用について議論した重点研究「音声言語」での共同研究者の現関西大学の檀辻正剛氏に感謝する。

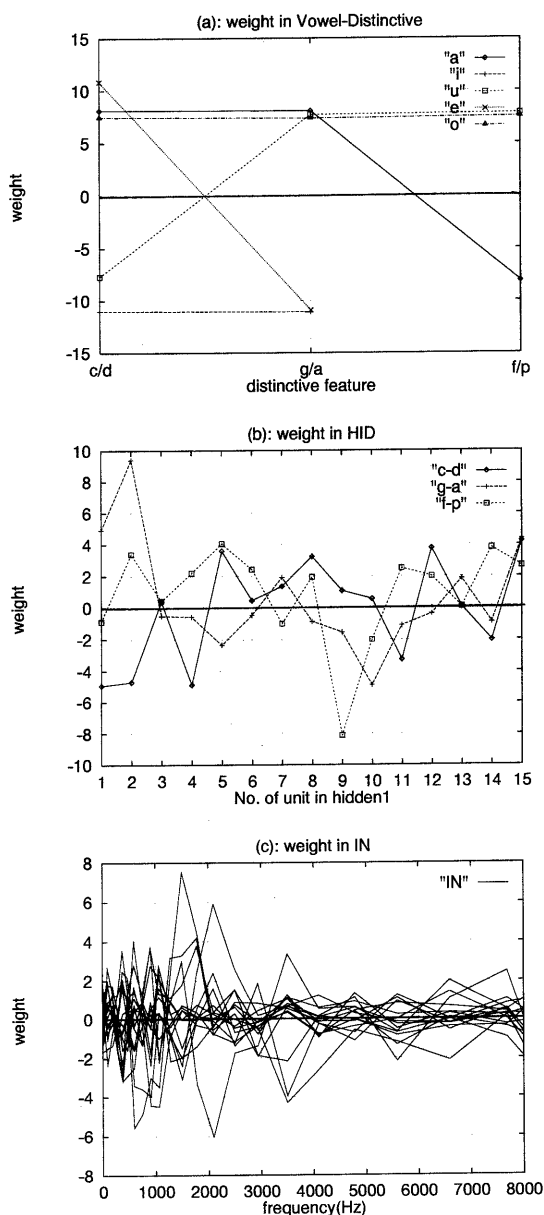


図 8: 学習後の重み係数値(学習1(固定))

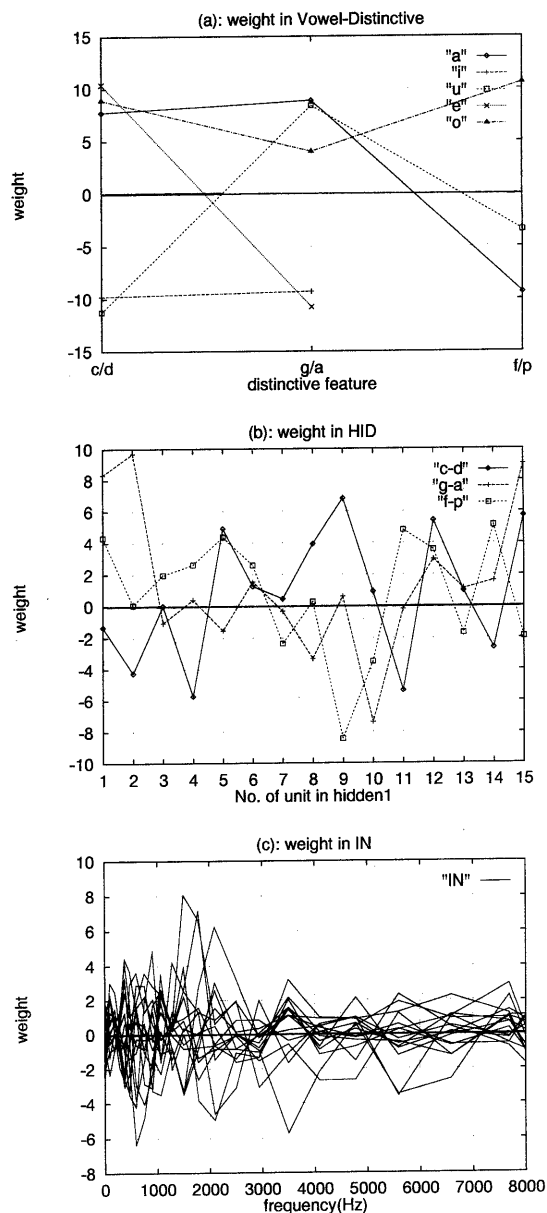


図 9: 学習後の重み係数値(学習2(可変))