

ニューラルネットワークを用いた
バイOMETリクス認証システムの
構築と音声信号の特徴分析への応用

2015年1月

長谷川 弘

目次

第 1 章	序論	1
1.1	本研究の背景	1
1.2	音声による認証法の選定	3
1.3	音声情報の活用	6
1.4	本研究の目的および構成	7
第 2 章	ニューラルネットワークの構築	8
2.1	序言	8
2.2	ニューラルネットワーク	9
2.3	生体ニューロン	10
2.4	人工ニューラルネットワーク	12
2.5	階層型ニューラルネットワークと学習アルゴリズム	17
2.6	本人認証のための階層型ニューラルネットワーク構築	25
2.7	結言	32
第 3 章	階層型ニューラルネットワークを用いた声紋認証システムにおける入力層数の効果	33
3.1	序言	33
3.2	実験方法	34
3.3	結果と検討	38
3.3.1	学習パターンに対するニューラルネットワークの性能	38
3.3.2	ニューラルネットワークの汎化能力	39
3.4	結言	42
第 4 章	アンサンブル学習法の効果とマルチステップ認証システム	43
4.1	序言	43

4.2	音声情報とニューラルネットワーク	44
4.3	アンサンブル学習の概要	48
4.4	本人認証システム	49
4.5	アンサンブル学習を用いた本人認証システム	53
4.6	マルチステップ認証システムの構築	57
4.7	マルチステップ認証システムの評価	59
4.8	結言	61
第 5 章	音声信号と血液中アルコール濃度との相関関係	62
5.1	序言	62
5.2	血液中のアルコール濃度と音声との関連性	63
5.3	実験方法	66
	5.3.1 話者認証システム	66
	5.3.2 線形予測分布	66
	5.3.3 ニューラルネットワーク	68
	5.3.4 ニューラルネットワークのテスト及び評価	70
5.4	結果と検討	72
	5.4.1 フォルマンの推定結果	72
	5.4.2 時間毎の本人拒否率曲線	72
5.5	結言	75
第 6 章	結論	76
	謝辞	79
	参考文献	80

第1章

序論

1.1 本研究の背景

近年、インターネット環境が整備されたのを契機に、インターネットの利便性を活かしたビジネスが常態化している。ビジネスを展開するうえでサービス提供者と利用者間で確実な本人確認を行うことは非常に重要である。現在では、インターネットを活用したビジネスを提供する場合、ユーザーに対してサービスを利用するための識別子 (ID) やパスワードを提供する方法が一般的に適用されている。

インターネットプロバイダー業務においては、その利用者の本人識別は「本人記入のサービス登録用紙の郵送」を基本にしていた。図 1.1 はインターネットビジネスにおける個人情報の流れを示している。

申し込み時に、お客様からインターネットを開始するための重要な情報 (住所、氏名、希望メールアドレス、その他) を取得し、プロバイダー側からインターネットを利用するための情報 (ユーザー ID、メールアドレス、初期パスワード) を返信する。このとき、インターネットからの申込や解約は本人確認できないことから不可とし、郵送による住所情報と氏名の一致により本人として認識している。インターネットの利用登録はインターネットを通じて行なうこともできるが、インターネット上に登録情報 (個人情報) が存在することから、以下の様々な方法で他者の手に渡る可能性をゼロにできない。

- ウイルスによる外部流出
- 不正アクセスによる情報搾取
- 運営企業の過失による情報流出 等

しかし、郵送に頼ることで時間的な制約が多く、特にインターネットに接続できない場

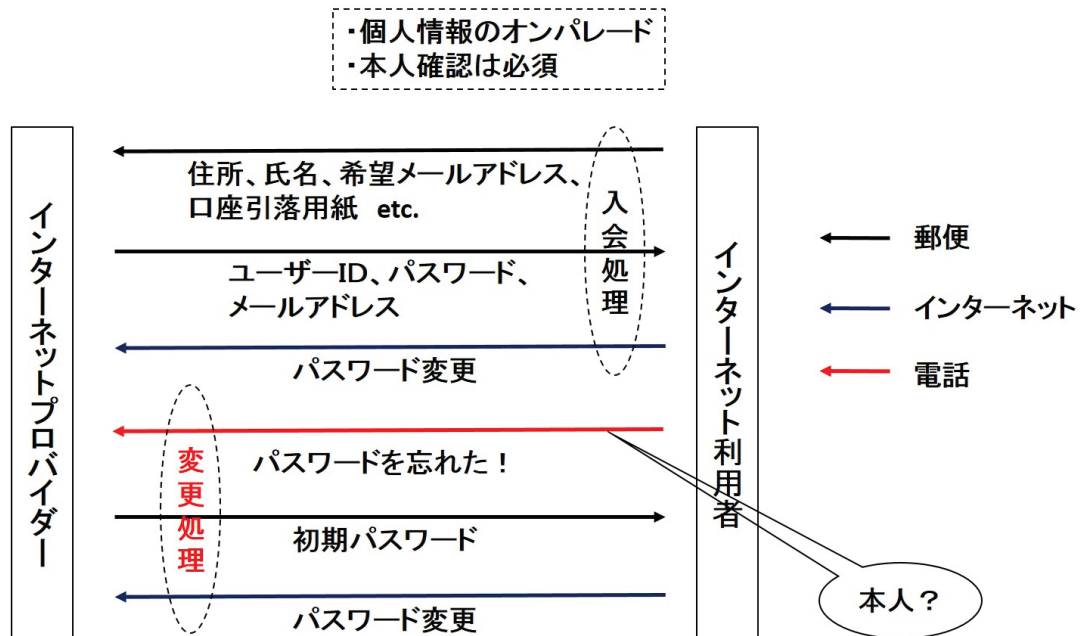


図 1.1 インターネットビジネスにおける個人情報の流れ

合のサポート対応時に大きな問題となっていた。具体的には、インターネットアクセスのための識別子やパスワードを忘却した場合、インターネットにアクセスできず困っている状況で、新識別子とパスワードを郵送で届けることで非常に長期のアクセス不可時間を作ってしまう。緊急の場合には、お客様との会話で「氏名」「生年月日」「住所」などの登録情報を確認したり、個人しかもたない情報(免許証やパスポート)の写しをファックスで送信いただくなどの対策をしているが、個人情報の取得にあたり、その利用方法を説明する義務があるなど煩雑な処理を伴うだけでなく、他者に情報が漏えいする恐怖からは逃れることができない。サポート中の電話の会話の中で、音声により本人確認できることで、その個人情報取得が簡易になるだけでなく、個人を特定するキーワードとの組み合わせで 100%に近い認証制度を実現できると考え、音声での認証方法を研究対象とした。

1.2 音声による認証法の選定

近年、コンピュータや通信技術の発展に伴って様々なサービスが”個人”に提供されている。”個人”にサービスを提供するにあたり、サービスを受ける”個人”を特定するための個人認証は非常に重要である。また、高度情報化社会において、個人情報の取り扱いに対してより高いセキュリティが求められる。これを実現するためには情報を取り扱う人を特定し、限定しなければならない。現在、個人を特定する方法にカードやユーザーID、パスワードなどがあり、それらを組み合わせた方法は幅広く使われている。これらの方法では、認証のためにカードの管理やID、パスワードの記憶が必要である。よって、これらの認証方法には、紛失や忘却、盗難等の様々な脆弱性がある。これを解決する手法としてバイオメトリクス認証がある。バイオメトリクスとは、”個人”の身体的または行動的な特徴のことで、原理的に紛失や忘却などが起きがたい。バイオメトリクスは具体的に、音声 [1]-[8]、指紋 [9]-[11]、掌紋 [12]、静脈 [13]、虹彩 [14]、網膜 [15]、顔 [16]、筆跡 [17] 等がある。図 1.2 にバイオメトリクス認証の特徴を示す [18]。

図では音声信号での認証の優位性は低くなっているが、インターネットビジネスで電話越しに取得できること、並びに、取得時のコストを考え、本研究の対象とした。音声信号を用いた個人認証（話者認証）については様々な研究がされている [1]-[8]。一般的な話者認証システムでは、音声信号に対して複雑な処理を行い、認証率を向上している。声紋が他のバイオメトリクス認証と異なるのは、本人の成長、体調の変化（喉の疾患など）、発生する音、環境によりバイオメトリクスの特徴が変化することである。しかし、我々人間はどの場においても凡その本人を認証する機能を有している。我々はニューラルネットワークがこのような変化する対象を扱う手法として有効であることを証明 [6]-[8] してきた。これらの報告では、母音「あ」が話者認証に適していることを示すとともに連続した短時間の音声波形をフーリエ変換して得られたパワースペクトルを用いることで、「本人」と「他人」を区別できることも示している。ただし、これらの報告では、認証する対象が1人であり、認証システムの処理は簡単であるが、認証率は十分に高い数値を得られないことも報告された。一方、話者認証や音声認識 [19] 以外に、音声信号に含まれる様々な特徴を用いて、鬱判定 [20] や飲酒判定 [21] に関する研究が行われている。音声認識では、隠れマルコフモデル等を用いて音声信号に含まれる言葉を認識し、発話内容を得てきた。鬱判定では音声信号にカオス処理を行い、鬱患者と健常者を分類した。飲酒判定では、アルコールによる音声信号の変化を利用して酔状態であるかどうかを判定した。音声信号には、雑音や個人の変動が含まれる。故に、話者認証の特徴量を抽出することは容易ではなく、複

方式	扱いやすさ	経年変化耐性	偽造のしにくさ	本人拒否の少なさ	他人拒否率	コスト	その他
虹彩	◎	○	◎	○	○	○	
網膜	×	△	○	○	○	△	一般向け事例少ない
声紋	○	×	×	△	×	○	録音に対する弱さが疑問
人相	◎	×	○	×	△	○	
筆跡	△	○	○	○	○	○	この評価は、Cyber-SIGNの場合
静脈	○	-	-	-	-	-	事例少なく不明
DNA	△	◎	◎	◎	◎	×	DNA抽出装置のコストに疑問
指紋	△	◎	◎	○	○	○	

凡例 ◎: 特に優れている ○: 良好 △: 疑問あり ×: 問題あり

<http://www.atmarkit.co.jp/fsecurity/special/11biomet/biomet02a.html>

図 1.2 バイオメトリクス認証の比較

雑な処理を行う必要がある。上述のような、様々な情報や雑音が含まれる音声信号の場合、アンサンブル学習が有効であることが報告されている [22]-[25]。医療支援システムでは、胸部 X 線画像を種々の前処理をすることにより、多くの入力パターンを作成し、それらを用いてアンサンブル学習を行った [25]。話者認証システムでは、boosting, bagging 及び CWP(combination of weak perceptrons) を適用し、認証率が向上した [23]。また回帰問題に対して非線形予測器としてフィードフォワード型の 3 階層ニューラルネットワークを用いてアンサンブル学習の汎化誤差を解析したことが報告されている [22]。従って、ニューラルネットワークにアンサンブル学習を加えることで精度の高い認証システムが構築できることが想定された。しかし、アンサンブル学習を階層型ニューラルネット

ワークで行うには、多くの学習パターンが必要であり、さらに、アンサンブル学習が階層型ニューラルネットワークに与える効果は十分には明らかにされていない。

本研究では、音声信号を用いてバックプロパゲーション法の階層型ニューラルネットワークに新アンサンブル学習を適用することにより、話者認証システムを構築し、アンサンブル学習の効果を明らかにした。さらに、話者認証システムを拡張することによって、マルチステップの個人認証システムを構築し、その性能を評価した。

1.3 音声情報の活用

前項で記述したように、音声には様々な情報を含んでおり、その情報を用いた研究も行われてきた。図 1.3 に音声に含まれる情報と研究分野についてまとめた。また、図中には母音の「あ」をカオス処理することで鬱状態を判別した結果も示している [20]。

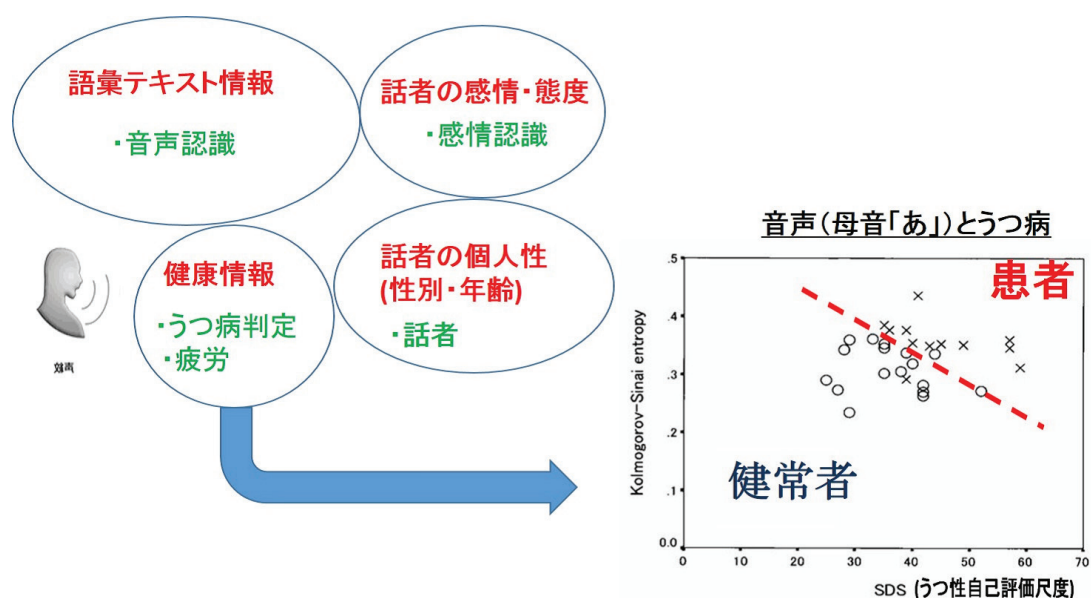


図 1.3 音声に含まれる情報と研究分野 [20]

図では白丸の兆しがコントロールを表す参加者、およびクロス(×)の兆候のうち、コルモゴロフ・サイナイエントロピーと SDS の分布は患者を示し、SDS は、抑うつ状態の自己評価抑うつ尺度を表す。結果として、正常時と抑うつ状態を区別する直線を描けることを示している。つまり、母音「あ」の音声は、患者がうつ病であるかどうか診断検査に役立つ可能性を示した。また、うつ病は、過労に関連する可能性がある。このことから、人間の身体が一時的に麻痺され、過労状態となり、音声信号にも影響を与えている可能性もある。上述したように、音声に様々な情報が含まれていることに注目し、近年問題となっている飲酒状態の判定システムを簡易な方法で（音声取得のみで）実現できるかどうかの検証も行った。

1.4 本研究の目的および構成

本論文では、少ない情報量ならびに遠隔地での本人認証システムを実現するため、ニューラルネットワークを用いた話者認証システムを構築し、その有効性を明らかにする。ニューラルネットワークへの入力として、音声を遠隔地で取得するに最も活用できる電話での音声取得を考慮し、低周波のサンプリング周波数で認証できるか検証した。この音声を用いての本人認証手段として、まずはアンサンブル手法での本人認証の確実性を検証し、次に複数の認証対象者での個人認証を可能にするため、マルチステップ認証システムを構築し認証の制度を確認した。更に、音声の持つ様々な情報を活用する一つのテーマとしてアルコール濃度の検出を選択し、音声での検出精度を確認した。

本章以降、第2章では指紋認証によるニューラルネットワークの構築について説明する。第3章では階層型ニューラルネットワークを用いた声紋認証システムにおける入力層数の効果について述べる。第4章ではアンサンブル学習法の効果とマルチステップ認証システムでの認証精度について言及する。さらに第5章では音声信号と血液中アルコール濃度の相関関係について述べ、第6章で本研究で得られた結果を総括し、今後の課題について言及する。

第2章

ニューラルネットワークの構築

2.1 序言

本章では、ニューラルネットワークの原理について解説し、話者認証に必要なニューラルネットワークの構造を検証した。

一般的にニューラルネットワークの性能は以下の8つのパラメーターによって決定される。これらは、① 学習パターン数 (Number of learning pattern)、② テストパターン数 (Number of test pattern)、③ 学習係数 (Learning coefficient)、④ 慣性係数 (Moment coefficient)、⑤ 入力層ユニット数 (Number of units in input layer)、⑥ 中間層ユニット数 (Number of units in hidden layer)、⑦ 出力層ユニット数 (Number of units in output layer)、⑧ 学習終了条件 (誤差二乗和)(Endcondition(square error)) である。

ここでは、指紋認証に利用するためのニューラルネットワークの性能について検証した。検証に必要な情報を取得するために、指紋画像を周波数変換し、フーリエ変換することで個人の特徴を入力しやすくすることを検証するとともに、個人認証に最適な階層型ニューラルネットワークの入力層、中間層、出力層のユニット数を明らかにすることを目的とした。

2.2 ニューラルネットワーク

コンピュータ (計算機) と呼ばれる従来のものは、ノイマン型コンピュータと呼ばれ、1つのCPUにより制御するアーキテクチャーを有しており、周辺装置のメモリや情報通信関連装置などの信号制御を電子部品によるハードウェア的な処理とコンピュータ言語によるソフトウェア的な処理によって行われる。ノイマン型コンピュータは、正確で高速な計算を得意とするが、パターン認識や連想記憶には適さない。

ニューラルネットワークは、ニューロコンピューター、ニューロコンピューティングなどと呼ばれてきたが、以下の3つの項目で特徴付けられる。これらは、①人間が行う情報処理を模倣する、②超並列分散処理をする、③学習をすることである。人間の脳の中には約100億 (10^{10}) 個のニューロン (神経細胞) と呼ばれる基本ユニットがあり、それらのユニットは約1万 (10^4) 個の他のニューロンとの接続を持っている。各ニューロンは、多数の他のニューロンから信号を受け取り、また、他の多数のニューロンへ信号を受け渡している。脳は、この信号の流れによって様々な情報処理を行っている。この仕組みをコンピュータで工学的に実現するものがニューラルネットワークあるいはニューロコンピューターである。これは、前述のノイマン型コンピューターとは異なり、パターン認識や連想記憶を得意とする。

2.3 生体ニューロン

生体ニューロンの概略を図 2.1 に示す [27]。細胞体はニューロンの本体である。細胞体から不規則な繊維状の樹状突起が長く伸びて複雑な枝分かれした構造をしている。この複雑な形は、次々に枝が分岐した葉のない冬の木立に似ている。樹状突起は他のニューロンからの入力を受け付ける役割に参与している。他に細胞体からは軸索が伸び、ニューロンの出力を他の細胞に伝える。また、介在ニューロンと呼ばれる種類のニューロンは、100 ミクロン程度の広がり度で入出力結合を持っているが、この介在ニューロンには軸索が存在せず、入出力とも樹状突起上にある場合がある。

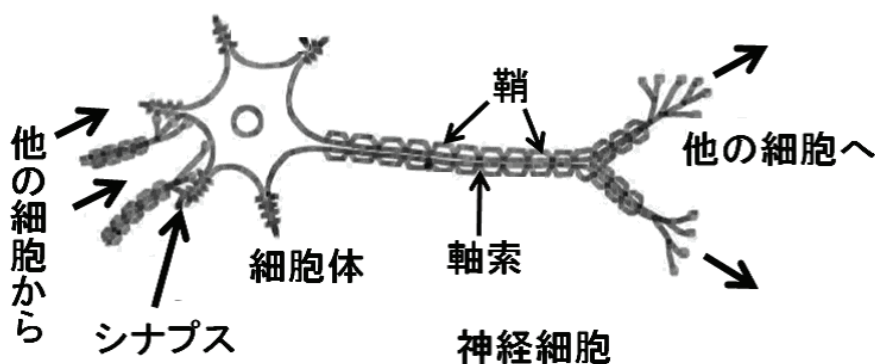


図 2.1 生体ニューロンの概略 [27]

神経細胞は興奮すると出力側の軸索に電気パルス列を送り出すが、興奮していないときは殆ど出さない。この電気パルスは 1 と 0 に量子化された信号を別の神経細胞に贈るものと考え、興奮状態及びその時に送り出される信号を 1、非興奮状態及びその時に送り出される信号を 0 としている。これを形式ニューロンと呼ぶ。神経細胞には樹状突起があって、多数の神経細胞からの軸索が結合しており、ここから信号を受け取る。この結合部分をシナプス (synapse) 結合と呼ぶ。

神経細胞が興奮すると、シナプス結合のうち、刺激を伝えたものは結合強度が増加し、更に刺激が伝えやすくなり、これが神経細胞に可塑性をもたらす、認識や記憶のもとになっている (発火: 持続時間の短いインパルス信号を出力する) という説が存在する。この説は生理学的には未だに実証されているものではなく、未だに仮説の段階であるが、

ヘッブのシナプス強化法則 (Hebbian rule) として大部分の神経細胞モデルの学習則に利用されている [28]。

2.4 人工ニューラルネットワーク

(1) ニューラルネットワークのモデル化

生体ニューロンを図 2.2 のように人工的にモデル化し、ニューラルネットワークとする [30]。ニューラルネットワークには SOM(Self Organization Map)[31] や RBF(Radial Basis Function) ニューラルネットワーク [32] など様々なものがあるが、ここでは BP(Back Propagation) 法のニューラルネットワークについて述べる。出力軸索にインパルスを与える状態を 1、そうでない状態を 0 とする。図 2.2 の縦軸を出力インパルスの有無、横軸の θ は、ニューロンの閾値を表す。また、 $\sum W_{ij}x_i$ は多数のシナプスの総和である。ここで、 W_{ij} はシナプス結合荷重と呼ばれ、シナプスの強度を表す。また、 x_i はシナプスの出力値である。

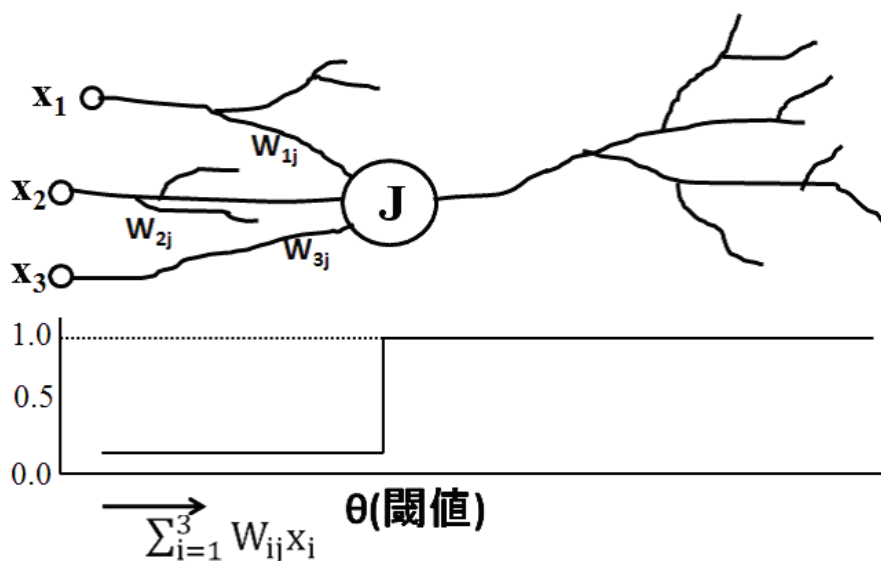


図 2.2 形式ニューロンの動作 [30]

いま特定のニューロン i について考え、添え字 j は省略する。以上の動作を数式で示すと、

$$y = \sum_{i=1}^n W_i x_i - \theta \quad (2.1)$$

$$z = f(y) \quad (2.2)$$

となる。式 (2.2) における z はニューロンの出力値であり、 $f(y)$ は、

$$f(y) = \begin{cases} 1 & \text{if } y \geq 0 \\ 0 & \text{if } y < 0 \end{cases} \quad (2.3)$$

のように定義される階段関数である。これを動作関数と呼ぶ。なお、生体の神経回路では信号の強度はパルス密度で表されるが、形式ニューロンではアナログ値に置き換えられる。この置換えは、数学的に等価とみなすことができる。ニューラルネットワークのニューロンの動作は、微分可能な代数的関数で代用されている。現在、最もよく用いられているしきい値関数はシグモイド関数と呼ばれるものである。

$$f(y) = 1 / [1 + \exp(-\alpha y)] \quad (2.4)$$

この関数は、 α の値により、非線形の強い階段関数 (大きな α 値) から線形に近い関数 (小さな α 値) まで連続的に変化させることができ、適切な性質を有する。また、シグモイド関数が微分可能であることは、後述されるバックプロパゲーション (逆伝播) 学習法の適用に大きな利点となる。

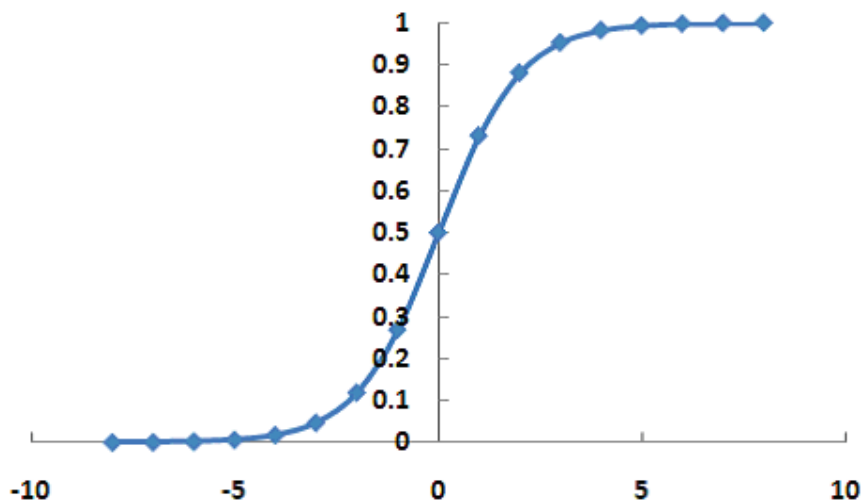


図 2.3 シグモイド関数の性質 [33]

(2) ニューラルネットワークモデルの基礎

ヘップのシナプス強化法則を基に、ニューラルネットワークが形成される。ネットワークは、大別して階層型と相互結合型とがある [34]。

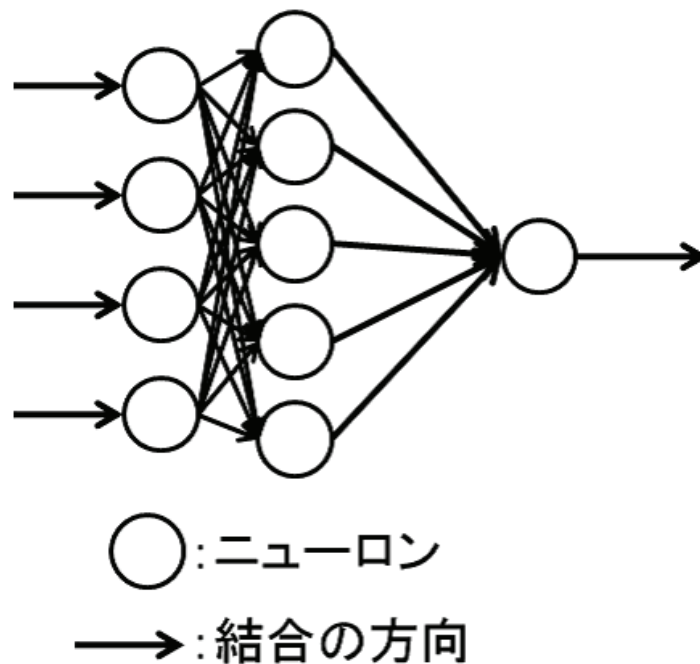


図 2.4 階層型ネットワークモデル [34]

図 2.4 に階層型ネットワークモデルを示している。階層型ニューラルネットワークは、入力層、単数または複数の中間層 (隠れ層とも呼ばれる)、及び出力層からなり、情報は入力層から出力層に向かって伝達される。1961 年に、Rosenblatt が学習機械 [パーセプトロン] としてはじめて提案した形であるので、パーセプトロン型ニューラルネットワークとも呼ばれる。後述されるように、階層型ニューラルネットワークには、バックプロパゲーション法という便利な学習方式が確立されている。

図 2.5 は相互結合型ネットワークモデルを示している。ニューロンの出力が自分自身に、あるいは特定の 2 個のニューロン間で互いに一方のニューロンの出力が他方のニューロンの入力にフィードバックされるような結合モデルを、フィードバック型あるいは相互結合型ニューラルネットワークと呼ぶ。相互結合型のニューラル

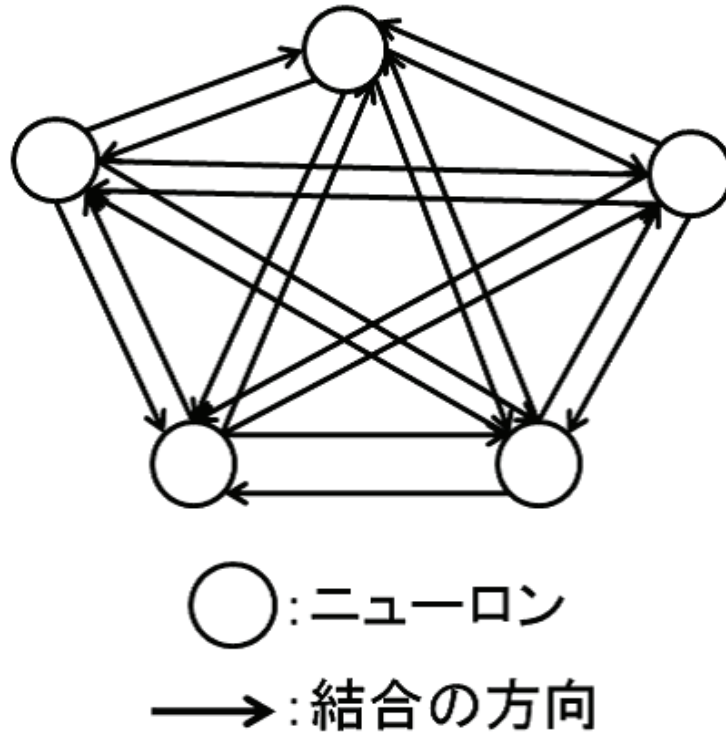


図 2.5 相互結合型ネットワークモデル [34]

ネットワークは、階層型のニューラルネットワークのような特別な構造を持たず、ニューロンは相互に結合され、情報も双方向に伝達される。

(3) ニューラルネットワークの学習

系の学習とは、その系が外界からの入力に応じて系自身の構造を変形することであると定義できる。生物の神経系では、神経細胞間のシナプス結合の強さを変化させることによって学習が行われる。したがって、ニューラルネットワークの理論においては、学習はシナプス間の結合荷重値 (これを総称して荷重行列 (weight matrix) と呼ぶ) を入力に応じて変化させることを意味する。また、入力と出力のみに注目すれば、学習とは入出力関係を変化させることであり、内部構造までは規定しない。このことは、同じ入出力関係でも環境によってネットワークは異なる構造をとることを意味する。

学習は、教師あり学習 (supervised learning) と教師なし学習 (unsupervised

learning) とに分類することができる。ある入力に対してネットワークの望ましい出力（教師信号）が与えられている場合、これを教師あり学習と言う。教師あり学習では、ニューラルネットの出力と教師信号の差をできるだけ小さくするよう結合強度を変更していくことにより、学習を行う。階層型及びボルツマンマシン型のネットワークは、教師あり学習方式をとる。

一方、教師なし学習は入力信号の統計的性質をネットワークに取り込み、その構造編成に反映させるものであり、自己組織化と呼ばれる。教師なし信号の場合は、理想的な出力は外部から与えられないので、自分自身の評価基準を内蔵しておくことが必要になる。この学習方式を用いるネットワークの1つに、コホーネンの自己組織化モデル（Kohonen's Self-Organizing MAP、SOM）[31]がある。

2.5 階層型ニューラルネットワークと学習アルゴリズム

(1) パターン間距離とパターン認識

通常、パターンは幾何学的な図形などを連想させるが、ここでは拡大された概念を用いる。すなわち、パターンはベクトルで表現されるものと考え、これをパターンベクトルという。このように考えることにより、図形も含め、大部分の情報はパターンとみなすことができる。

パターン間の類似度あるいは相違度を知るために、パターン間の距離という考え方が用いられる。n 個の要素をもつパターンベクトル $\mathbf{P} = \{p_1, p_2, p_3, \dots, p_n\}$ が与えられたとき、要素を 1 つ 1 つの座標に割り当てた n 次元の空間 (これらをパターン空間と呼ぶ) を考える。この空間内で、2 つのパターンベクトル $\mathbf{P} = \{p_1, p_2, p_3, \dots, p_n\}$ 、 $\mathbf{Q} = \{q_1, q_2, q_3, \dots, q_n\}$ の距離は、次式で定義される。

$$D = \sum_k \{|p_k - q_k|^r\}^{1/r} \quad (2.5)$$

これを \mathbf{P}, \mathbf{Q} 間のミンコフスキー (Minkovskii) 距離という。特に、 \mathbf{P}, \mathbf{Q} の各要素が 0,1 の 2 値をとり、 $r=1$ のときはマンハッタン (Manhattan) 距離という。

パターン空間内では、各パターンは点で表される。パターンベクトルのうち、同じものとするべきパターンの集合を、カテゴリまたはクラスと呼ぶ。同じカテゴリに属するパターンは相互に近いミンコフスキー距離にある点であるとし、図 2.6 に示すように、カテゴリごとに集まって分布するものとする。よって、パターン認識とは、パターン空間内の点をカテゴリごとに分割することに他ならない。この分割には、パターン空間内で直線 (高次の空間では、平面または超平面) で分離する場合と、曲線 (曲面または超曲面) で分離する場合とがある。前者を線形分離 (分類)、後者を非線形分離 (分類) と呼ぶ。

(2) 学習則と収束定理

式 (2.1) において、ベクトル $\mathbf{X} = \{x_1, x_2, x_3, \dots, x_n\}$ をパターンベクトル \mathbf{P} とみなせば、形式ニューロンは超平面により入力パターンを 2 つのカテゴリに線形分離する機能を有すると解釈される。

次に、教師信号付きの線形分離可能なパターン分離問題が与えられた場合を考える。このとき、試行錯誤的な反復操作によって、ニューロンの出力が外部からの教師ベクトル (\mathbf{t}) に一致するように結合荷重 \mathbf{W} を調節し、カテゴリに分類する超平面の位置と傾きを決定する。この過程 (学習) を用いれば、形式ニューロンと \mathbf{W}

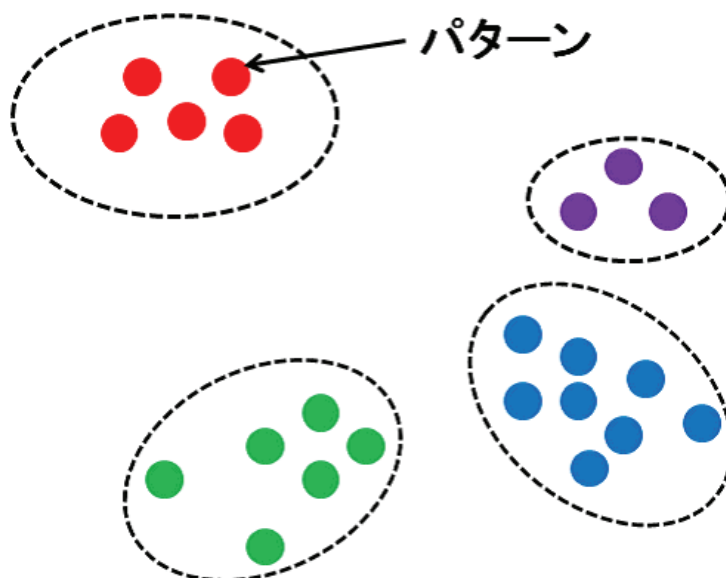


図 2.6 パターン空間内におけるパターンベクトルの分布 [35]

は有限回の反復学習で正しい認識をすることが証明されている (パーセプトロンの収束定理)。

(3) デルタルール

階層型ニューラルネットワークにおける学習を想定し、形式ニューロン素子の動作を次のように規定する。ニューロン j は前層のニューロン i の出力 x_i を入力として受け、結合荷重 W_{ij} をかけて加えたものの総和を y_j とする。 j の出力 O_j は y_j に動作関数 $f(\cdot)$ を施したもので、0~1 間の値であるものとする。

$$O_j = f(y_j) \quad (2.6)$$

$$y_j = \sum_i W_{ij} x_i \quad (2.7)$$

ただし、ニューロン j の閾値は結合荷重に含まれている。

ネットワークが、ある入力ベクトルに対し出力 O_j を与えるとする。学習を評価するため、次のような評価関数 E を定義する。

$$E = (1/2) \sum_j (O_j - T_j)^2 \quad (2.8)$$

ただし、 T_j は教師ベクトルの要素である。 O_j はニューロン間の結合荷重 W_{ij} で決まるため、 E は O_j 、また O_j は W_{ij} の関数となる。 E は W_{ij} の空間上の超曲面 (hypersurface) とみなすことができ、この曲面の極小値に達するためには、 E の W_{ij} についての偏微分 $\partial E/\partial W_{ij}$ に比例した量 (δW_{ij}) だけ W_{ij} を変化させる。

$$\delta W_{ij} = -\epsilon \frac{\partial E}{\partial W_{ij}} \quad (2.9)$$

これは、誤差が曲面上を最も急な傾斜方向へ進むことを意味し、このような学習を一般に最急降下法 (gradient descent method) という。

式 (2.8) において、 E を W_{ij} で偏微分すると、次式で表される。

$$\frac{\partial E}{\partial W_{ij}} = \frac{\partial E}{\partial O_j} \cdot \frac{dO_j}{dy_j} \cdot \frac{\partial y_j}{\partial W_{ij}} \quad (2.10)$$

式 (2.7) より、 $dO_j/dy_j = f'(y_j)$ 、また $\partial y_j/\partial W_{ij} = x_i$ である。一方、式 (2.8) より $\partial E/\partial O_j = O_j - T_j$ であるから、式 (2.9) は、

$$\delta W_{ij} = -\epsilon (O_j - T_j) f'(y_j) x_i \quad (2.11)$$

となる。この操作を全ての入力ベクトルと教師ベクトルについて行い、この学習則をデルタルールという。

$$f(y) = 1/[1 + \exp(-\alpha y)] \quad (2.12)$$

$f(y)$ がシグモイド関数 (式 (2.12)) であるとき、 $f'(y)$ は、

$$f'(y_i) = f(y_i) [1 - f(y_i)] \alpha \quad (2.13)$$

である。したがって、 W_{ij} の変化量 δW_{ij} は、

$$\delta W_{ij} = -\epsilon (O_j - t_j) f(y_i) [1 - f(y_i)] \alpha x_i \quad (2.14)$$

となる。

(4) バックプロパゲーション学習法

図 2.7 に示す 3 層の階層型ニューラルネットワークについて考える。○印はすべて形式ニューロンで、その動作は式 (2.7) に示す動作関数に規定される。入力データは A から入り、B から出力される。階層型ニューラルネットワークでは、どの素子も 1 つ前の層からのみ入力を受け、次の層へのみ出力するものとする。このようなネットワークの $\partial E/\partial y_j$ は、出力層より逆向きに計算していくことになる。つ

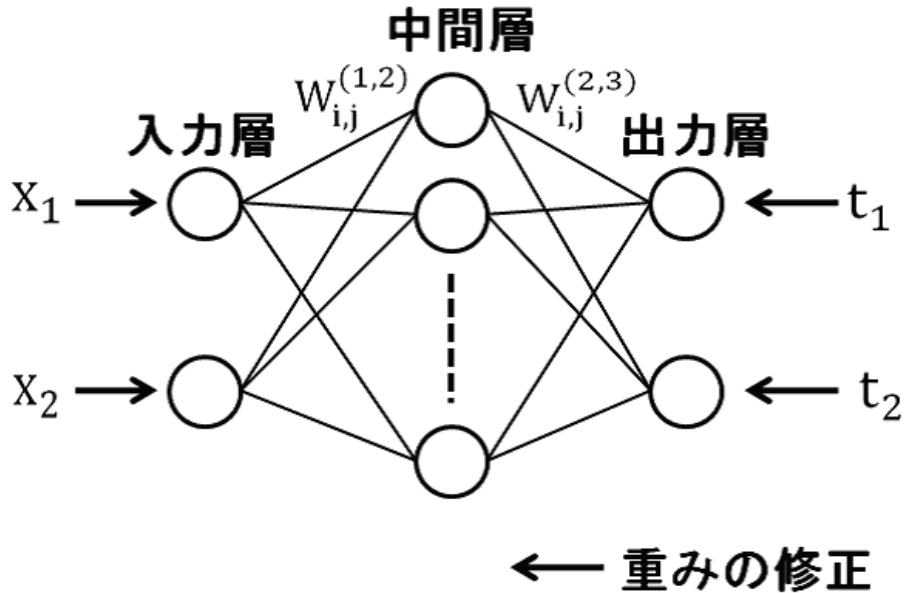


図 2.7 バックプロパゲーションによる学習

まり、出力の誤差を順に前の層へ伝えていくのがバックプロパゲーションである。この考え方にデルタルールを適用すると、次のアルゴリズムを得る。

$$\delta W_{ij}^{(n-1,n)} = -d_j^{(n)} x_i \epsilon \quad (2.15)$$

$$d_j^{(3)} = O_j - t_j \quad (2.16)$$

$$d_j^{(2)} = \left(\sum_l W_{jl}^{(2,3)} d_l^{(2,3)} \right) f'(y_j) \quad (2.17)$$

ここで、肩添え字は層間を表し、式 (2.16) は第 2,3 層の修正に、また式 (2.17) は第 1,2 層の修正に用いる。

(5) バックプロパゲーションアルゴリズムの要素

バックプロパゲーションを実行するとき、決定しなければならないいくつかの要素がある。まず、パラメータ α, θ, ϵ 、及びネットワーク自体の構造である。次に、評価関数の値 E を、どの程度まで小さくすればネットワークの機能が実現されるか、すなわち、学習の収束条件を調べる必要がある。

- パラメータ α

ニューロンの出力値は、パラメータ α と結合荷重値 W_{ij} の結合によって与えられる。したがって、 α の値によってネットワークの機能が左右されることはない。しかし、 α の値を適切に選ぶことによって学習の収束速度を速めることができる。

- パラメータ θ

入力層の 1 つまたは複数のニューロンが 0 でない定数値 (通常 1.0) を出力するとき、中間層の各ニューロンについてのパラメータ θ 値を 0 とすることができる。これが可能であることは、複数の入力層のニューロン出力値が合成されて中間層の 1 つのニューロンに入力されることを考えれば明らかである。定数値を出力するニューロンを、通常バイアス (bias) ニューロンと呼ぶ。

一般に、第 n 層にバイアスニューロンをおくと、第 $(n+1)$ 層のニューロンの θ 値は 0 とおくことができる。また、十分な数のニューロンを中間層 (複数あってもよい) に配置すれば、それらの層にはバイアスニューロンを導入する必要はないことが経験的に知られている。これは、学習過程で各中間層の 1 つまたは複数のニューロンに結合している結合荷重行列の要素が 0 となり、自動的にバイアスに相当するニューロンが生成するためである。

- パラメータ ϵ

パラメータ ϵ がバックプロパゲーション過程で、多次元空間をどれだけ極小解の方向へ進ませるかを示す係数である。動作関数としてシグモイド関数を使用した場合は、0.05~0.2 を採用することが多いが、数学的根拠はない。経験的に小さな ϵ 値を用いると極小解に到達するのに時間がかかり、大きな値を用いると極小解を飛び越し、振動し、収束しないことがある。

- ネットワーク構造— 層数

階層型ネットワークの分類 (fitting) 能力は、ネットワークの中間層数、それらの層のニューロン数、およびニューロンの動作関数に依存する。入出力層と出力層の 2 層のみからなるネットワークの分類能力には限界があり、たとえば、表 2.1 に示すような排他的論理和の学習ができないことが知られている [36]。

このことは、ネットワーク内では入力情報の内部表現が必要であり、非線形動作を行わせるためには中間層が 1 層以上必要となることを示している。通常、パターン認識問題では、3 層のネットワークが用いられる。これは中間層のニューロン数として適当な数を用いれば、3 層のネットワークで任意の関数が近似できると証明されていること、4 層以上では学習時間が長くなることによる。これらの根拠は積極的に 3 層構造を支持するものではないが、経験的に中間層はその層のニューロン数

表 2.1 排他的論理和

入力パターン		出力パターン
0	0	0
0	1	1
1	0	1
1	1	0

を適当に設定すれば 1 層で十分である。

● ネットワーク構造— 中間層のニューロン数

入力層と出力層のニューロンの個数は、それぞれ入力パラメータの数、分類の数(分類の場合) または 1 個である。その決定法は研究されているが、まだ決定的方法は知られていない。これに関しては次の点が経験的に判明している。

- (a) 必要なニューロン数は、入力および教師データの情報量に依存する。
- (b) 必要なニューロン数は、バックプロパゲーション学習に付加条件を導入すると変化する。
- (c) ニューロン数を必要以上に多くしても、ネットワークの機能は大きくは変化しない。

一般に、中間層のニューロンの個数を増やすことによって、より複雑な関数を近似することができる。しかし、安易に中間層のニューロン数を増やせば過学習の可能性が出る(理由については、後の”過学習”で述べる)。それを避けるため、必要最小限のニューロン数を決定する方法が求められる。

再構築学習法でもある程度決定可能であるが、それが必要最小限の数であるとは限らない。中間層のニューロン数が必要最小限の数であれば学習は容易に収束するが、それより 1 個でも少なくなると収束しなくなる。経験的にいって、普通この境界ははっきりしている。したがって、中間層のニューロン数の初期値として、入力層のニューロン数の 2 倍程度を与え、そこから学習が不可能になるまで減少させていく方法は、原始的であるが確実に速い。

● 結合荷重行列の不確定性

結合荷重について次のような特徴がある。既に述べたように、学習とは入出力関

係のみを変化させることであり、ネットワークの内部構造までは規定しない。したがって、同じ入出力関係でも、学習における環境によって結合荷重行列の個々の要素の値は異なることが予想される。たとえば、3層構造の階層型ネットワークにおいて、第1,2層間の荷重が中間層のニューロン j_0 に集まるものとする。ニューロン j_0 を j_1 と j_2 に分離して同様の学習を行えば、同じ結果が得られる。そのとき、ニューロンを加える前の荷重値 ($W_{ij}^{(0)}$) と加えた後の荷重値には、

$$W_{ij}^{(0)} = \lambda_1 W_{ij}^{(1)} + \lambda_2 W_{ij}^{(2)} \quad (2.18)$$

の関係がある。ここで、荷重行列 $W_{ij}^{(1)}$ と $W_{ij}^{(2)}$ は、係数 λ_1, λ_2 の値に依存し、不定である。すなわち、中間層のニューロンの数に冗長性がある場合は、荷重行列の要素の値は不定となる [37]。

- 過学習

ニューラルネットワークの非線形動作によって、過学習 (over fitting) と呼ばれる問題が発生する。過学習とは、学習の進行に伴い個々のデータの持っている固有の微細構造までを学習するため、学習データについての近似精度は向上するが、未学習データに対する精度が悪くなる現象をいう。具体的な例としては、個々のデータに含まれる誤差なども因果関係に取り入れられてしまう場合が挙げられる。

図 2.8 にそのような例を示す。破線部分が過学習である。ニューラルネットワークの非線形動作を強くし、丁寧に学習させた実線部分に比べ、波形の増減が激しく明らかな過剰な学習となる。この問題を解決するために、あらかじめ統計的な方法を用いて、個々のデータに含まれる誤差を除くことができれば、最も効果的である。

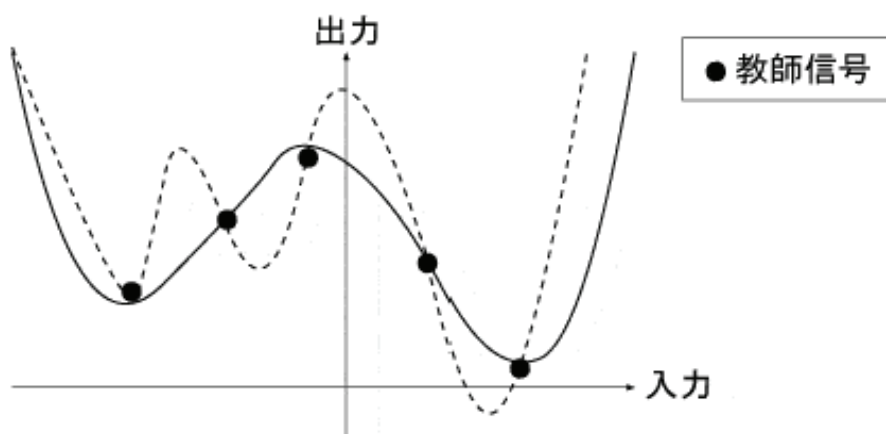


図 2.8 過学習の例

2.6 本人認証のための階層型ニューラルネットワーク構築

指紋画像による本人認証システムを構築し、前処理としてFFT（高速フーリエ変換）とIFFT（逆高速フーリエ変換）について検証し、階層型ニューラルネットワークの本人認証における性能を検証した。

(1) 前処理

ニューラルネットワークへの入力として指紋画像を利用した。指紋画像を4方向(0,45,90,135°)に分類し、SOM(self-organizing feature map)およびニューラルネットワークで検証されたことが報告されている。しかし、その報告には、本人認証に対する精度については含まれていない[48]。そこで、簡易で高速の処理方法を提案し、3階層のニューラルネットワークを利用した認証方法を実施した[57]-[58]。その結果、この方法だけでは多数の人物を認証できず、認証システムの性能が指紋画像の質に影響を受けることが分かった。そこで、このシステムの性能向上のため、高速フーリエ変換(FFT)と逆FFT(IFFT)を利用した処理方法を検証した[60]-[61]。指紋画像は収集し2進符号化した。この2進符号化した指紋画像のFFTとIFFTを実施した。このデータを図2.9に示す3階層型ニューラルネットワークに入力した。

指紋画像は、1と0に2進符号化した。白と黒はそれぞれ0と1に符号化した。灰色部分の指紋画像は色むらとして扱った。その結果、平均値を計算することで、詳細な部分に対して2進符号化できた。この平均値を用い、指紋画像は2進符号化された。データ領域はFFTとIFFTで処理され、128×128のデータとされた。図2.10に2進符号化した指紋画像の代表的な例を示す。収集した指紋認証画像の雑音に耐えるよう指紋画像に2次元FFT処理を行った。FFTとして2次元FFTを明確にした。

図2.11に4種の指紋画像のフーリエ振幅と位相スペクトルを示す。白と黒のエリアがそれぞれFFT出力の高い部分と低い部分を示す。フーリエ振幅と位相スペクトル双方を含んだ指紋画像の特徴を用いることで、IFFTを実行した。ここで、位相およびフーリエ変換スペクトルはそれぞれ実数、虚数である。

図2.12に4種の指紋画像にIFFTを実行した結果を示す。我々はこの図をITTF画像として明確にした。このIFFT画像のサイズは64×64である。IFFT画像での縞模様のパターンは人それぞれで異なった。その結果、我々は縞模様を認

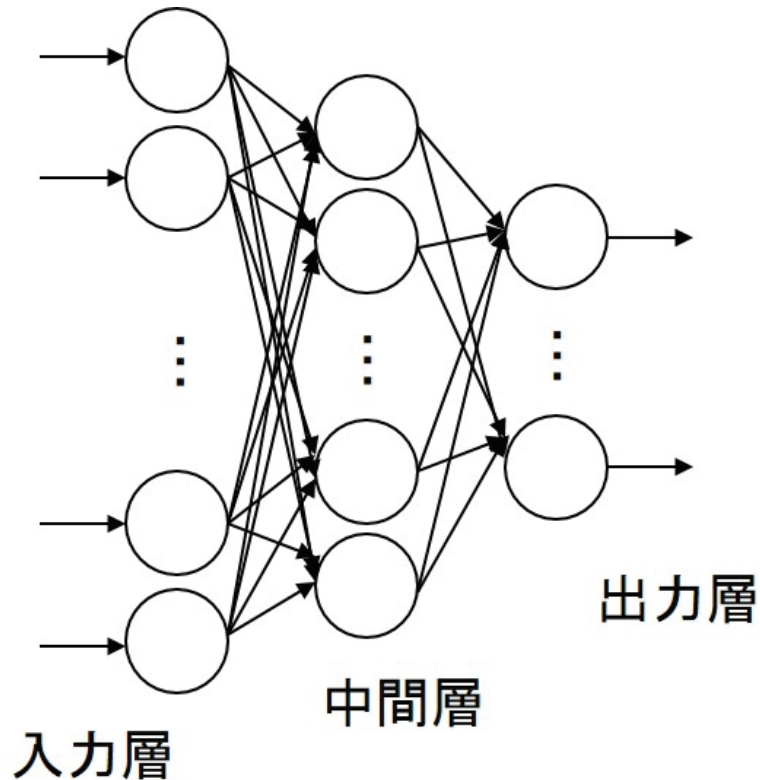


図 2.9 3 階層型ニューラルネットワークの例

識することができ、3 階層ニューラルネットワークを用いて個人認証が可能であると考えた。

全ての指紋画像において FFT と IFFT の領域は 128×128 である。出力層数は 1 から 3 まで変化させた。学習係数が 0.01 より大きい場合、ニューラルネットワークは収束しない。そのため、最適な学習係数を 0.01 とした。本人おとしての教師データとして 0.95 を与え、他人としての教師データとして 0.05 を与えた。学習は、教師データと入力データの 2 乗誤差が 0.001 以下になるまで実行された。

他人の指紋画像を本人の指紋画像と間違える可能性を本人拒否率として、本人の指紋画像を他人の指紋画像と間違える可能性を他人受入率と定義した。ここで、認証精度は本人の指紋画像を本人と認識する可能性である。我々は、3 階層ニューラルネットワークにおける、FFT および IFFT の前処理を検証した。



図 2.10 2 進符号化した指紋画像の代表的な例

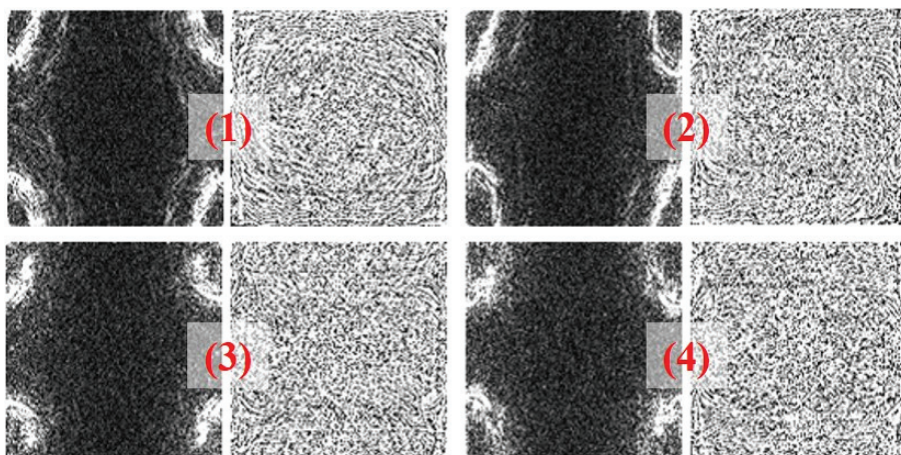


図 2.11 4 種の指紋画像のフーリエ変換（左）と位相スペクトル（右）

(2) 3 階層ニューラルネットワークの最適化

3 階層ニューラルネットワークの入力として、IFFT の 4 隅のデータを利用した。角のサイズは 10×10 から 40×40 に変化させた。入力層のユニット数は 400, 900, 1600, 3600, 6400 である。図 2.13 に入力層のユニット数に対する本人証明の認証精度を示す。図に示すように、入力層のユニット数が 900 の時に最も良い認証精度

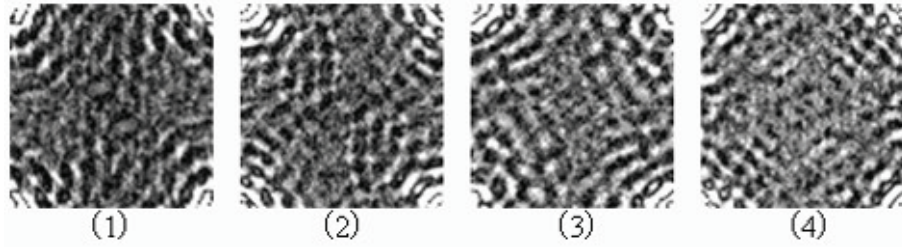


図 2.12 4 種指紋画像の IFFT 画像

となった。出力層数が 1, 2, 3 の時、認証精度は、それぞれ、100, 100, 96.5 % となった。この結果から、入力層の最適な数は 900、出力層のユニット数が増加するに従って認証精度が下がることが分かった。

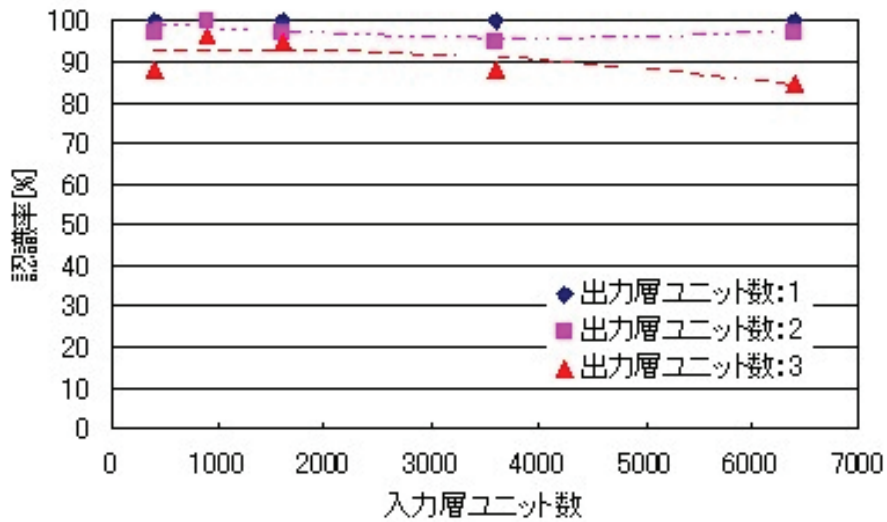


図 2.13 入力層のユニット数に対する認証精度の変化

図 2.14 に中間層のユニット数に対する本人証明の認証精度を示す。入力層のユニット数は 900 に固定し、中間層のユニット数を 200 から 800 に変化させた。図に示すように、認証精度は中間層のユニット数が 500 の時、認証精度が最も良く、

その数値は約 98 %であった。この結果、中間層のユニット数の最適値は 500 であり、IFFT 画像の小さな隅のデータが高性能なニューラルネットワークの入力データに欠かせないことが分かった。ゆえに、IFFT で前処理された指紋画像の特色は小さな隅に含まれていると考えられる。

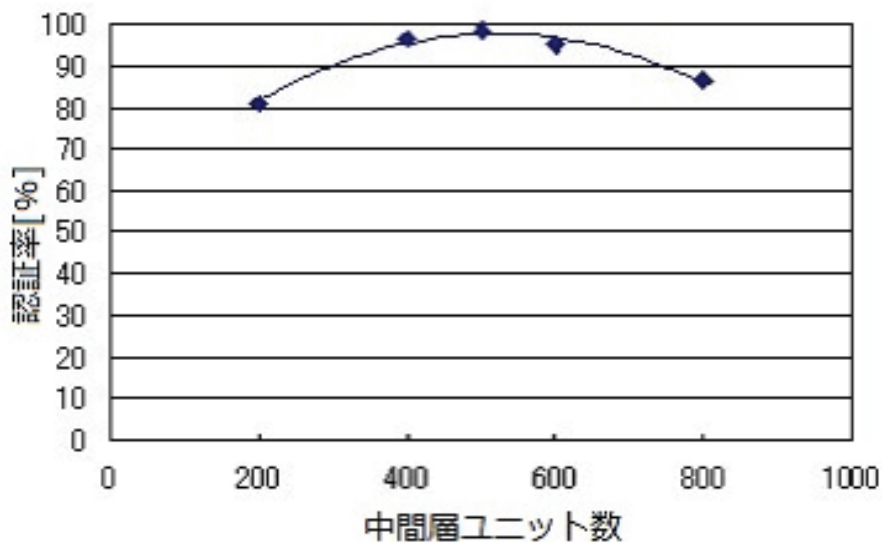


図 2.14 中間層のユニット数に対する認証精度の変化

図 2.15 に出力層のユニット数に対する本人証明の認証精度を示す。出力層のユニット数は 1 から 3 に変化させ、その時の入力層・中間層のユニット数はそれぞれ、900 および 500 である。出力層のユニット数が 1 の時、ただ 1 人だけの確認が可能となる (1)。出力層のユニット数が 2 の時、3 人の認証が可能になる (01,10,11)。出力層のユニット数が 3 の時、7 人の認証が可能になる (001,010,011,100,101,110,111)。指紋画像が他人であった場合の出力値は、0(出力層のユニット数が 1),00(出力層のユニット数が 2),000(出力層のユニット数が 3) とした。ここで、出力値の値が 0.95 以上を 1、0.05 以下を 0 に設定した。その結果、出力層のユニット数が増加するに従って本人認証精度が悪くなり、特に、他人受入率が増加することが分かった。ニューラルネットワークの学習使用されていない指紋画像は認証精度を悪化させると考えた。そこで、学習に使用していない指紋画像は、未学習試験データとして分離した。学習に使用していない指紋画像は、検証に

使用していない指紋画像と異なる。

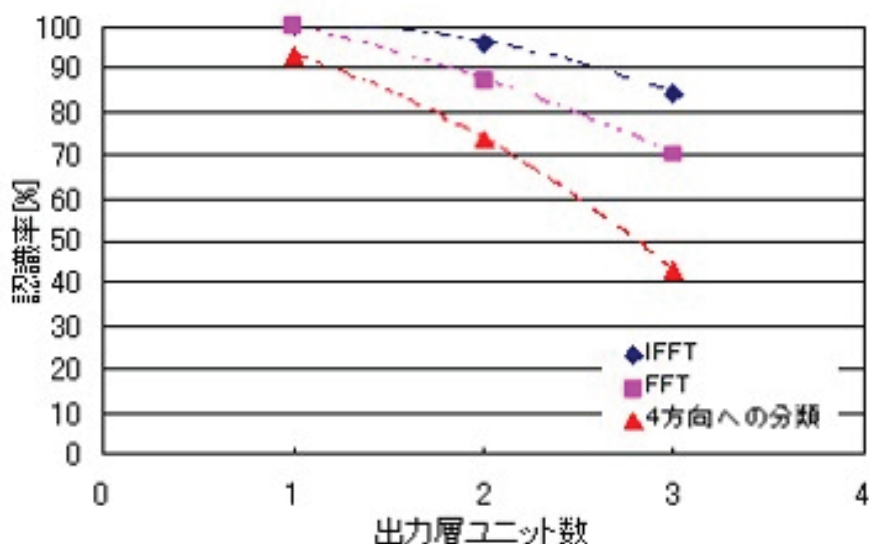


図 2.15 出力層のユニット数に対する認証精度の変化

図 2.16 に出力層のユニット数に対する未学習データの認証精度を示す。図に示されるように、認証精度は出力層のユニット数の増加によって著しく悪化する。このテストデータを用いた認証精度は約 100 %で、出力層のユニット数の増加で悪くなる。

次に 3 階層ニューラルネットワークにおける未学習データの影響を検証した。

(3) 未学習テストデータによる個人認証

指紋画像が他人の場合、テストデータとして 3 種類の指紋画像とした。図 2.17 に教師データとして 000 を与えた指紋画像の種類に対する認証率の変化を示す。入力層、中間層、出力層のユニット数は、それぞれ、900、500、3 とした。図に示すように、教師データとして 000 を与えた指紋画像のユニット数が 12 の時、認証率は 100 %となった。その結果、教師データとして 000 を与えた指紋画像が多数あれば、我々のシステムで多数の人物を認証できる。

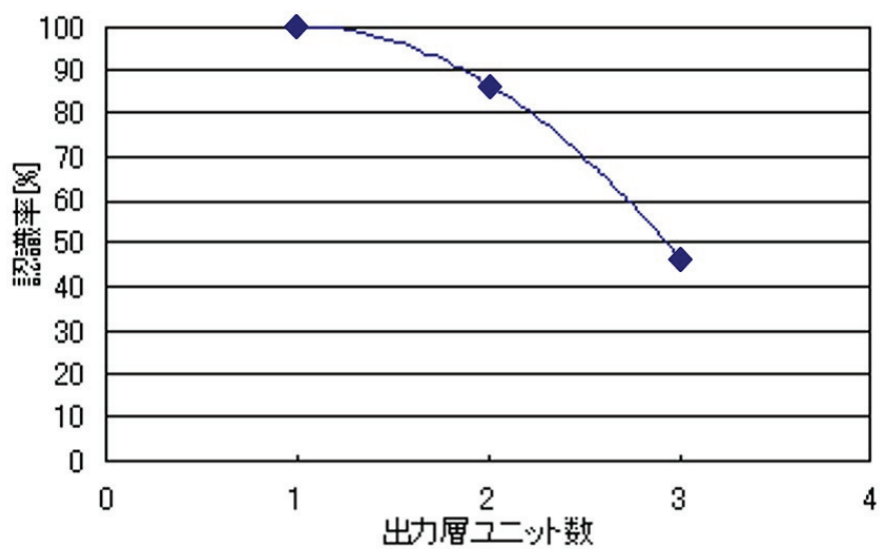


図 2.16 出力層のユニット数に対する未学習データの認証精度

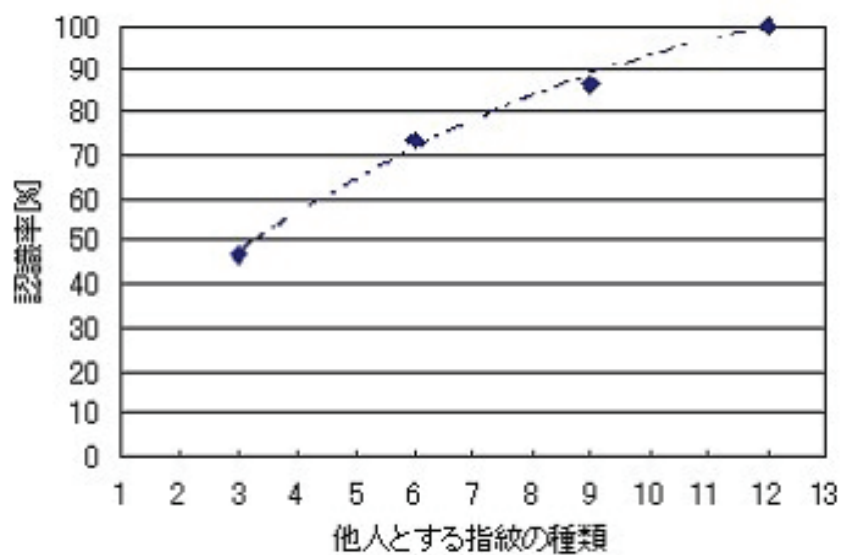


図 2.17 教師データとして 000 を与えた指紋画像の種類に対する認証率の変化

2.7 結言

指紋画像を用いて個人認証できる 3 階層ニューラルネットワークを構築し、個人認証における前処理方法が階層型ニューラルネットワークの性能に及ぼす効果を検証した。IFFT 変調と FFT により位相を作り、それをニューラルネットワークの入力とすることで、入力層、中間層、出力層の最適なユニット数によって多数の人物を認証できることが分かった。そして、もし出力層のユニット数が 7、すなわち教師信号数値 000 で与えられた指紋画像認証システムでも、十分多くのデータがあれば、個人認証システムは未学習データ (未知の指紋画像) についても認証できることがわかった。更に、ニューラルネットワークの持つ非線形かつパターン認識能力が、指紋画像の質にほとんど影響されず、指紋画像認証システムに効果的であることが分かった。ゆえに、指紋画像の回転取得、ノイズ除去、隆起先鋭化などの前処理をすることなく、高性能な認証システムを提供できると考えられる。

第3章

階層型ニューラルネットワークを用いた声紋認証システムにおける入力層数の効果

3.1 序言

我々は、バックプロパゲーションアルゴリズムによる3層階層型ニューラルネットワークを用いて、個人認証システムを構築した。音声信号は高速フーリエ変換 (FFT) によって前処理を行い、バックプロパゲーション学習法のニューラルネットワークの入力データとして使用した。我々は、ネットワークが正常に動作しているかどうかを判断するために、学習パターンを用いて繰り返し学習することで、誤差値が設定した終了条件まで減少することを確認した。また、入力したデータが正常であるかどうかを判断するために、学習パターンに対する FAR と FRR 曲線を作成し、エラーレイト (誤差率) が 0 で、FAR と FRR 曲線が交わることを確認した。さらに我々は、個人認証において必要な性能を明らかにするために、サンプリング周波数の減少と入力層のユニット数の削減が個人認証にどのような影響を与えるか検証した。本章では、入力層ユニットの数を減少 (サンプリング周波数の減少) させても、ニューラルネットワークの性能に問題ないことも明らかにした。

3.2 実験方法

図 3.1 に本研究で用いた前処理（アルゴリズムの要点のみ記載）とニューラルネットワークの構造を示す。音声情報の収集には、パーソナルコンピュータと、マイクロフォンを利用した。本実験で用いた発音は「あ」である。音声信号は 20 代の男性から取得し、8 ビット毎に周波数 48kHz でサンプリングした。

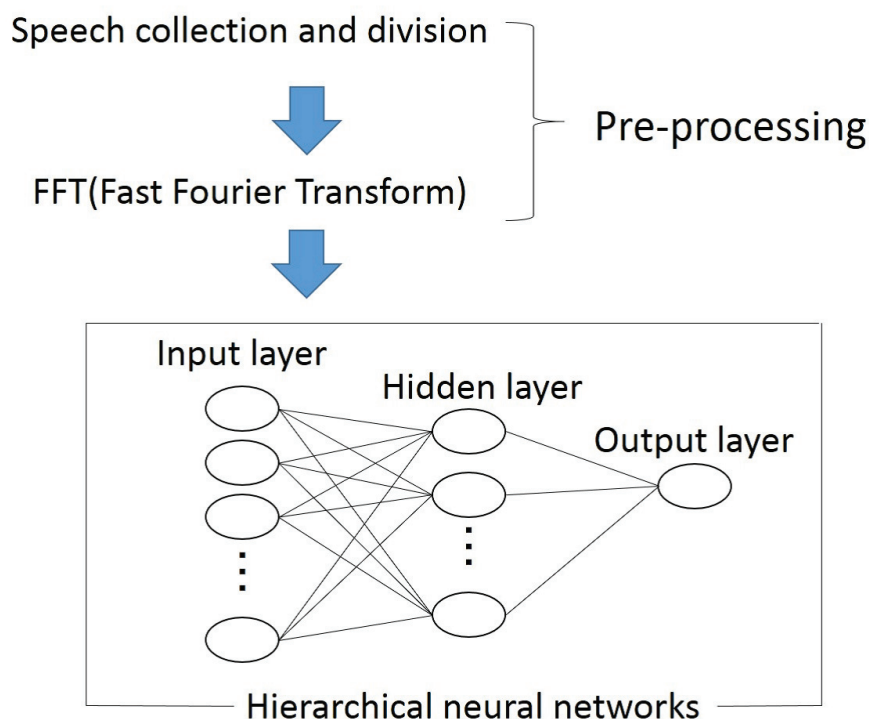


図 3.1 前処理とニューラルネットワークの構造

表 3.1 に音声信号の収集条件を示す。サンプリング周波数を徐々に減らし、音声信号は 512 ポイントに分割した。サンプルデータは FFT（高速フーリエ変換）でパワースペクトルに変換した。これら一連の前処理を行った後にニューラルネットワークの入力とした。

音声信号は FFT で前処理し、パワースペクトルに変換し、そのパワースペクトルをニューラルネットワークの入力とした。表 3.2 にニューラルネットワークのパラメータを示す。学習パターンとテストパターンの数をそれぞれ 100 とした。ニューラルネット

表 3.1 音声信号の収集条件

Sampling frequency	48[kHz]
Bit	8[bit]
Sound	1
Other	Men's speech

表 3.2 ニューラルネットワークのパラメータ

The number of learning pattern	100
The number of test pattern	100
Learning coefficient	0.1
Momentum coefficient	0.02
Number of units in hidden layer	100
Number of units in output layer	1
End condition (square error)	0.001

ワークの学習係数と慣性係数はそれぞれ 0.1 と 0.02 とした。音声は本人である場合の教師信号は”1”、他人である場合の教師信号は”0”とした。ニューラルネットワークの誤差二乗和は出力層の値と教師信号の差の二乗和を意味する。本章においては、ニューラルネットワークが収束するまでの誤差二乗和と学習繰返し数はそれぞれ 0.001 と 100,000 とした。ニューラルネットワークの中間層と出力層のユニット数はそれぞれ 100 と 1 とした。サンプリング周波数を 1, 2, 4, 8kHz とした場合、入力層のユニット数はそれぞれ 32, 64, 128, 256 である。

図 3.2 に種々のサンプリング周期での FFT 処理を行う前の音声信号を示す。図に示されるように、音声信号の取得データはサンプリング周期が上がるにつれ短時間のデータとなる。

図 3.3 に異なるサンプリング周波数毎に実施した FFT 後の音声信号を示す。図に示さ

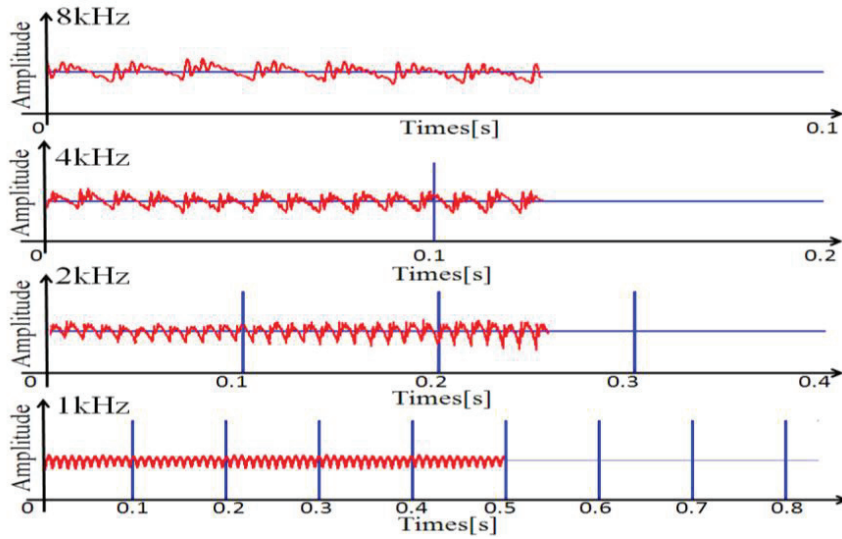


図 3.2 F F T 処理前の音声信号 (サンプリング周期別)

れるように、サンプリング周波数が少なくなるにつれてパワースペクトルの変動は減少した。

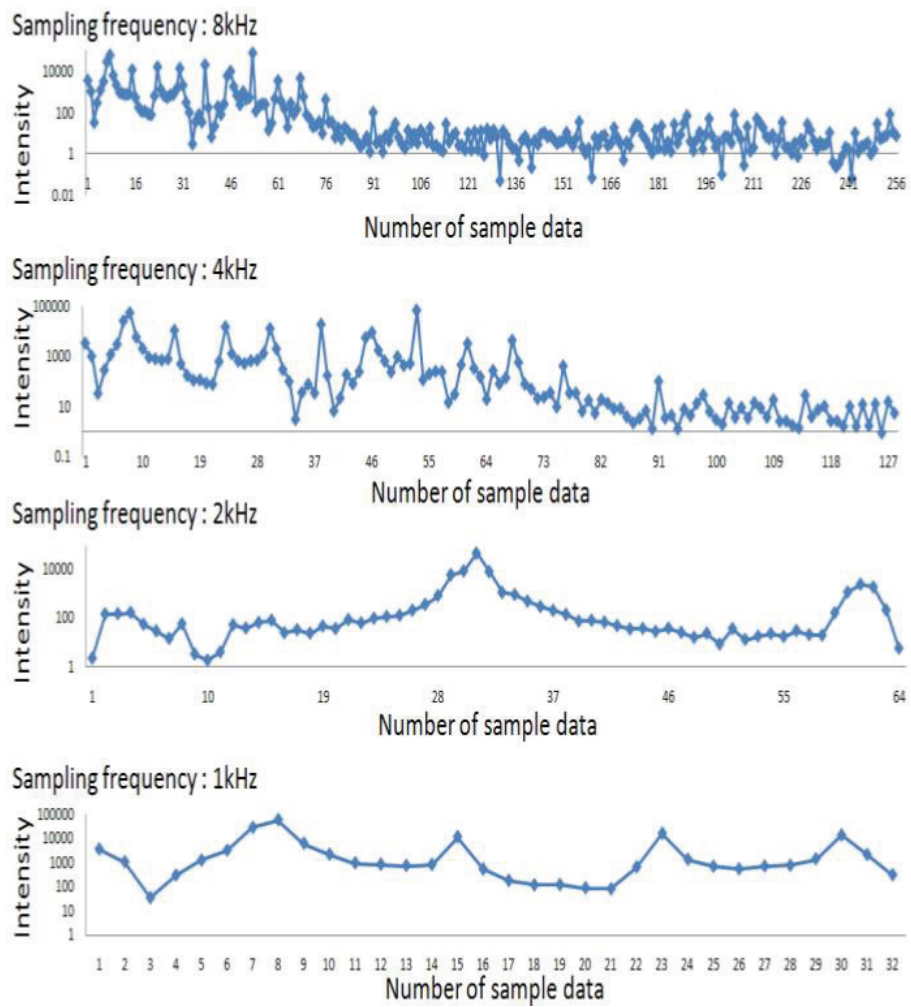


図 3.3 F F T 後の音声信号 (サンプリング周期別)

3.3 結果と検討

3.3.1 学習パターンに対するニューラルネットワークの性能

我々はニューラルネットワークを用いて声紋認証システムを構築した。本研究の生体認証システムの性能は、FAR（他人受入率）と FRR（本人拒否率）で評価する。ここで、我々は音声信号が本人のものか他人のものかを調査した。

図 3.4 に各サンプリング周波数の学習パターンに対するニューラルネットワークの誤差二乗和を示す。図に示されるように、学習回数が増加するにつれ、誤差二乗和は減少した。これは 1, 2, 4, 8kHz のどの周波数でも同様である。その結果、ニューラルネットワークの誤差二乗和が 0.001 以下になることが明確になり、100,000 回以下の繰り返し学習回数においても学習パターンに関係なく収束することが分かった。

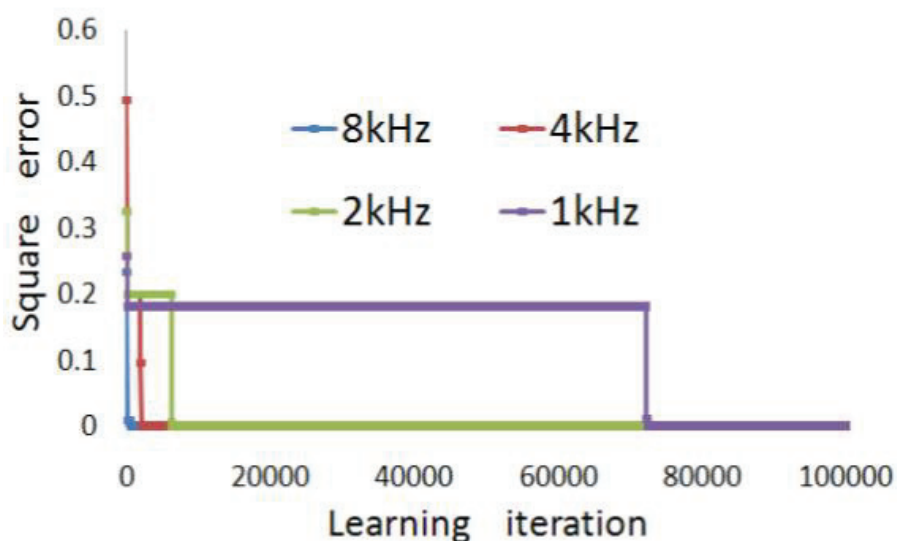


図 3.4 各サンプリング周波数の学習パターンに対するニューラルネットワークの誤差二乗和

図 3.5 に学習パターンがテストパターンとして利用された場合の出力ユニットの閾値に対する FAR 値と FRR 値を示す。図に示されるように、FAR 値は閾値の増加とともに減少したのに対して、FRR 値は増加した。声紋認証における FAR 値と FRR 値は同時に 0

となる閾値が存在する。これは、学習パターンに対してニューラルネットワークが安定的に機能し、FAR 値と FRR 値のどちらも 0 となる閾値幅が広いことを示している。つまり、このニューラルネットワークを用いることで本人であることが正常に判断できる。

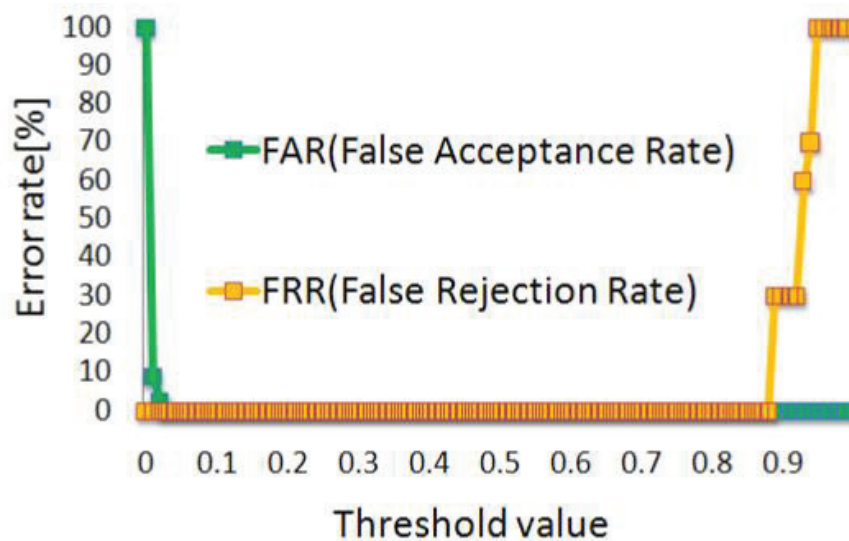


図 3.5 出力層のユニットの閾値に対する FAR 値と FRR 値 (学習パターンがテストパターンとして利用された場合)

3.3.2 ニューラルネットワークの汎化能力

我々はサンプリング周波数 1, 2, 4, 8kHz で声紋認証システムを調査した。本研究では、テストパターンとして未学習の 20 代の男性 13 名を利用した。

図 3.6 にサンプリング周波数 8kHz で得られた未学習パターンでの、出力層のユニットの閾値に対するニューラルネットワークの FAR 値と FRR 値を示す。その結果、ニューラルネットワークの性能は、ニューラルネットワークの汎化能力を意味することがわかった。図に示されるように、FAR 値は出力層のユニットの閾値の増加とともに減少し、FRR 値は増加した。このことは、サンプリング周波数 8kHz を用いたシステムが声紋認証に有効であることを示している。

図 3.7 にサンプリング周波数 2kHz で得られた未学習パターンでの、出力層のユニット

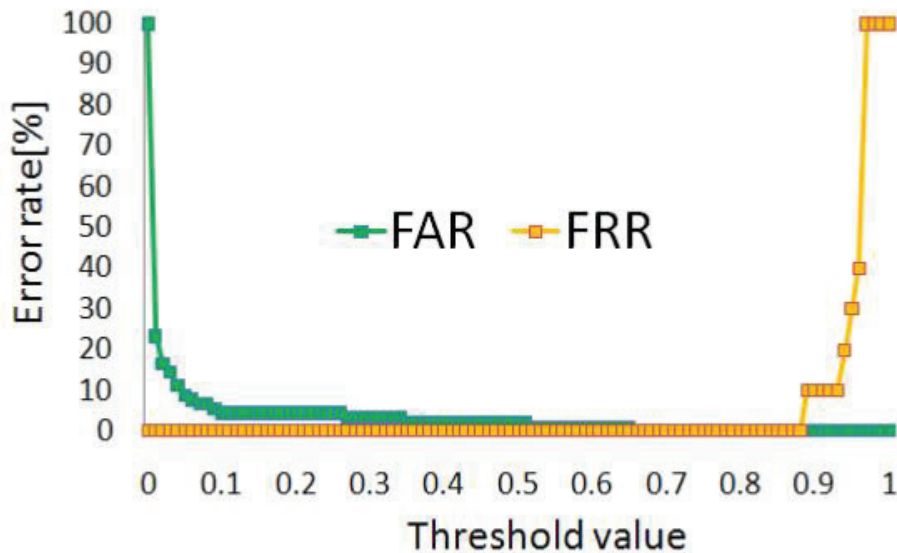


図 3.6 サンプリング周波数 8kHz のテストパターンに対する FAR 値と FRR 値

の閾値に対するニューラルネットワークの FAR 値と FRR 値を示す。図に示されるように、FAR 値は閾値の増加とともに減少したが、FRR 値は増加した。これは、サンプリング周波数 2kHz を用いたシステムも声紋認証に有効であることを示している。

図 3.8 に入力層のユニット数に対する閾値の幅を示す。ここで、閾値の幅とは、FAR 値と FRR 値の双方が 0 となっている幅と定義した。図に示されるように、テストパターンとして学習データを用いた場合、全てのサンプリング周波数において閾値の幅が凡そ 0.8 を維持したが、未学習データを用いた場合は、各サンプリング周波数において大きく変化した。実験で用いたどのサンプリング周波数（入力ユニット数）においても、また、テストパターンとして学習パターンを用いる場合と未学習パターンを用いる場合のいずれにおいても、認証率 100% となる閾値の幅が存在することを確認した。また、テストパターンとして未学習パターンを用いた場合は、サンプリング周波数の小さい領域の方が、閾値の幅が広い傾向にある。故に、サンプリング周波数が減少することで、認証が有効になる傾向があると考えられる。

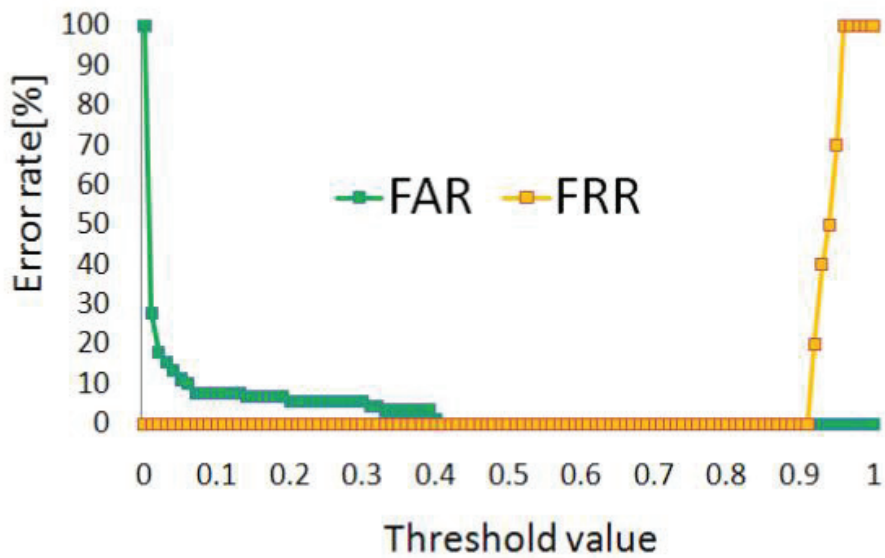


図 3.7 サンプリング周波数 2kHz のテストパターンに対する FAR 値と FRR 値

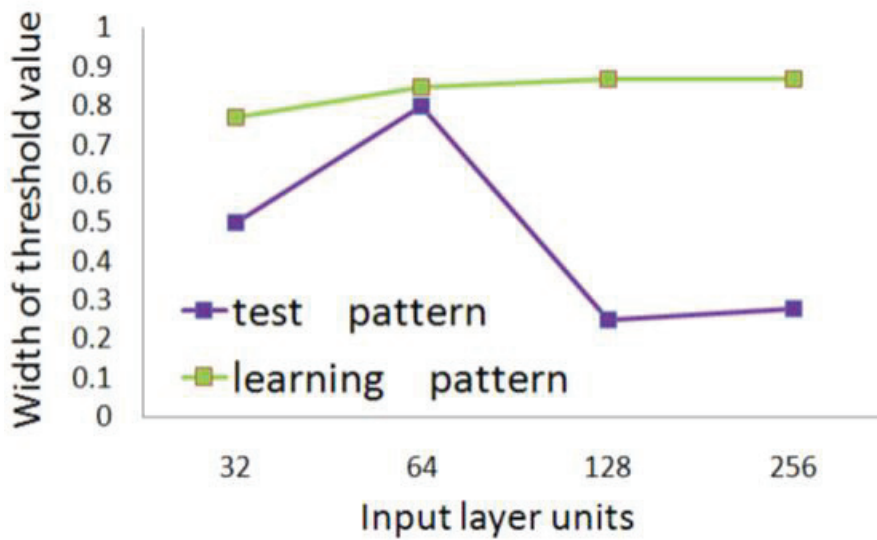


図 3.8 FAR 値と FRR 値がどちらも 0 になる場合の閾値幅

3.4 結言

ネットワークが正常に動作しているかどうかを判断するために、学習パターンを用いることにより、学習回数に対する終了条件までの減少を確認した。結果、ニューラルネットワークの誤差二乗和は減少し、声紋認証システムは学習回数を増やすことで収束することが分かった。

また、入力したデータの種類（学習と未学習の音声）によるニューラルネットワークの性能を評価するために、テストパターンとして学習パターンを用いた場合と未学習パターンを用いた場合の FAR と FRR 曲線を作成し、双方の実験で、FAR と FRR 曲線が同時に 0 を示す範囲が存在することを確認した。このことは、ニューラルネットワークが安定して動作していることを示している。

さらに、個人認証において必要な特徴を明らかにするために、サンプリング周波数の減少により入力層のユニット数の削減を行った。これらの結果から、この認証システムでは、32 から 256 の全ての入力層のユニット数のシステムにも有効であることがわかったが、未学習の音声を認証する場合は、入力層のユニット数の減少（サンプリング周波数の減少）がニューラルネットワークの性能向上に有効であることが明確になった。

第4章

アンサンブル学習法の効果とマルチステップ認証システム

4.1 序言

第3章で記述したように、ニューラルネットワークを用いた話者認証システムを開発した。この話者認証システムは多数の人物の音声を利用して学習を繰り返すことにより精度の高い結果を維持できるが、業務で利用するには、その繰り返し学習が弱点になる。また、音声にはその時の環境や健康状態によって様々な雑音を含む可能性があり、1つのニューラルネットワークで話者を認証するのが困難となる可能性が高い。アンサンブル学習はこのような弱点を補うのに有効であることを第1章の序論で記述した。

本章では、アンサンブル学習の原理を説明し、バックプロパゲーション学習法を用いた階層型ニューラルネットワークへ適用した方法について解説するとともに、アンサンブル学習で実証した本人認証システムを拡張し、高速かつ拡張性の高いマルチステップの個人認証システム構築したので、その実用化について提案する。

4.2 音声情報とニューラルネットワーク

個人認証まで実験を拡張するためには、複数の音声しかも客観的な音声を使用する必要がある。国立情報学研究所の音声資源コンソーシアム「日本語方言音声コーパス」に収録されている「あいうえお」と発音している音声の中の「あ」を抽出して使用した。図 4.1 に音声信号の前処理の流れを示す。本実験では、母音「あ」を用いた。図に示されるように、母音は周期性を持った波形の繰り返しで構成されている。この波形の 1 周期分を取り出し 1 組の音声情報とした。一人あたり計 10 組を取り出し、それらを高速フーリエ変換を行うために各組が 512 ポイントになるように再サンプリングした。次に、これらを高速フーリエ変換することでパワースペクトルとなり、ポイント数は 256 となる。これがニューラルネットワークへの入力パターンとなる。

音声コーパスの採取条件を表 4.1 に示す。サンプリング周波数 16 kHz、ビット長 16 ビット、モノラル音声である。本研究では、この音声の採取条件を変更せずに使用した。実験に利用した音声コーパスのパターンを図 4.2 に示す。実験に用いた人数は 14 人で、1

表 4.1 音声コーパス [29] に収録されている音声の各種条件

Sampling frequency	16[kHz]
Sampling bit rate	16[bit]
Channel	1[ch](monaural)

人あたり 10 パターンの計 140 パターンを準備した。この内、学習パターンの本人として 5 パターン（図の○：本人用として使用しない場合はテストパターン）を、学習パターンの他人として 4 人 × 5 パターン、計 20 パターン（図の◇）を使用した。残りの 115 パターンをテストパターン（図の T + 本人パターンとして利用しない 20）として用いた。

表 4.2 に、ニューラルネットワークの学習における各種パラメータを示す。入力層、中間層及び出力層のユニット数はそれぞれ、256, 100, 1 であり、学習係数と慣性係数は共に 0.02、学習の終了条件となる教師信号と出力層ユニットの出力値との誤差 2 乗和は 10^{-4} 以下とした。学習には教師信号として他人パターンには”0”を与え、本人パターンには”1”を与えた。

表 4.2 ニューラルネットワークの各種パラメータ

Number of input layer units	256
Number of hidden layer units	100
Number of output layer unit	1
Momentum coefficient	0.02
Learning coefficient	0.02
End condition(square error)	10^{-4}

バイオメトリクス認証に対するシステムの評価は、他人受入率 (FAR:False Acceptance Rate) と本人拒否率 (FRR:False Rejection Rate) を用いた。また、FAR と FRR が両方共に 0 % となる閾値の範囲を閾値長 (Length of threshold value) として定義した。

Voice collection and division

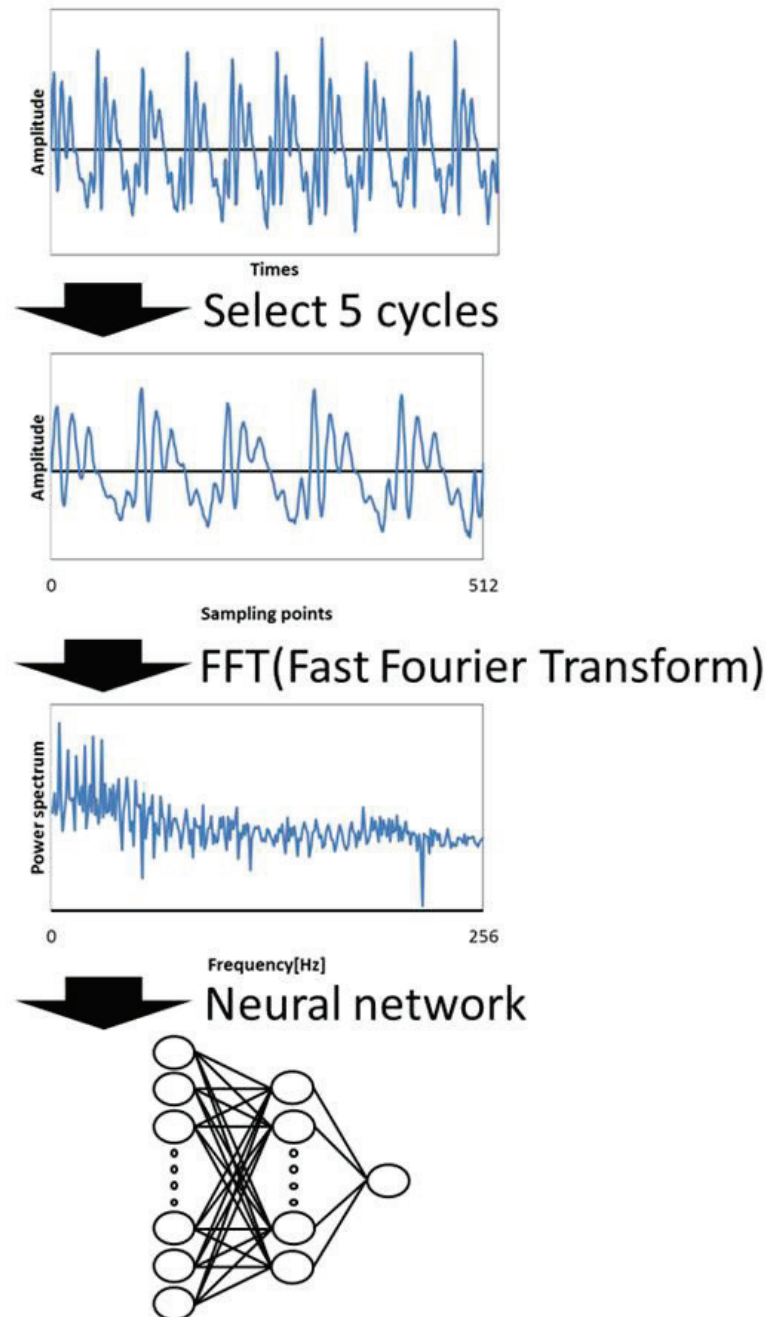


図 4.1 前処理の流れ

		音声パターン									
		P 1	P 2	P 3	P 4	P 5	P 6	P 7	P 8	P 9	P 10
利用 音声	N1(A)	○	○	○	○	○	T	T	T	T	T
	N2(B)	○	○	○	○	○	T	T	T	T	T
	N3(C)	○	○	○	○	○	T	T	T	T	T
	N4(D)	○	○	○	○	○	T	T	T	T	T
	N5(E)	○	○	○	○	○	T	T	T	T	T
	N6(X)	◇	◇	◇	◇	◇	T	T	T	T	T
	N7(Y)	◇	◇	◇	◇	◇	T	T	T	T	T
	N8(Z)	◇	◇	◇	◇	◇	T	T	T	T	T
	N9	◇	◇	◇	◇	◇	T	T	T	T	T
	N10	T	T	T	T	T	T	T	T	T	T
	N11	T	T	T	T	T	T	T	T	T	T
	N12	T	T	T	T	T	T	T	T	T	T
	N13	T	T	T	T	T	T	T	T	T	T
	N14	T	T	T	T	T	T	T	T	T	T

○: 本人学習
◇: 他人学習
T: テストパターン

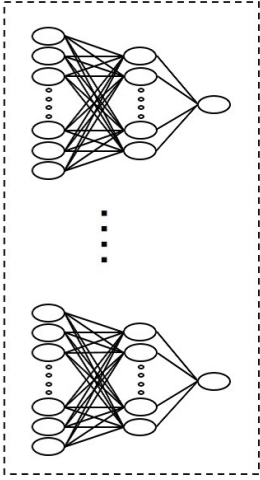


図 4.2 実験に利用した音声コーパスのパターン

4.3 アンサンブル学習の概要

図 4.3 にアンサンブル学習の概要を示す。図に示されるように、認証されるパターン（テストパターン）は各ニューラルネットワーク（NN#1～NN#M）に入力される。各ニューラルネットワークには結合荷重をランダムに与えることで、異なるニューラルネットワークを生成した。各ニューラルネットワークは異なる荷重行列を持っているので、各ニューラルネットワークに同じパターンを入力しても異なった出力値 $g_1(x) \sim g_M(x)$ を持つ。これらの出力値 $g_1(x) \sim g_M(x)$ を平均することで出力値 $g(x)$ を得る。本論文では、上記の流れでアンサンブル学習を行い、その性能を検証した。アンサンブル学習の手法には、異なった学習パターンを用いてニューラルネットワークを作成する方法等がある [22][25]。本研究では同一の学習パターンを用いてニューラルネットワークを作成する方法を用いた。本研究の手法により、少ない学習パターンでも異なった結合荷重行列を持つ多くのニューラルネットワークを生成することができる。

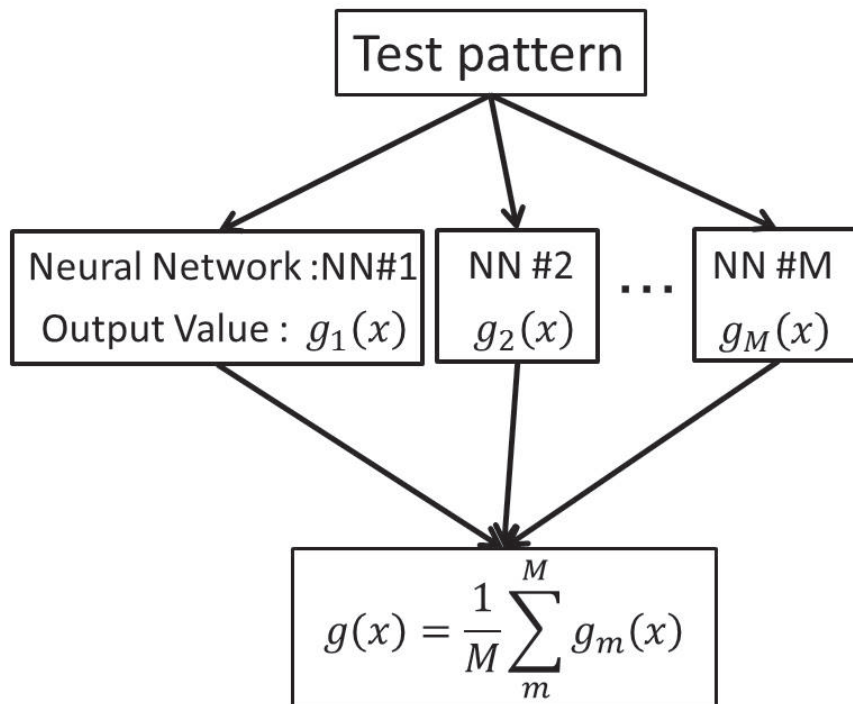


図 4.3 アンサンブル学習の概要

4.4 本人認証システム

図 4.4 は、学習パターンにおける学習回数に対するニューラルネットワークの誤差 2 乗和を示している。図に示されるように、学習回数が増加すると、ニューラルネットワークの誤差 2 乗和は単調に減少した。これらの結果より、ニューラルネットワークは学習パターンに対して収束することがわかった。

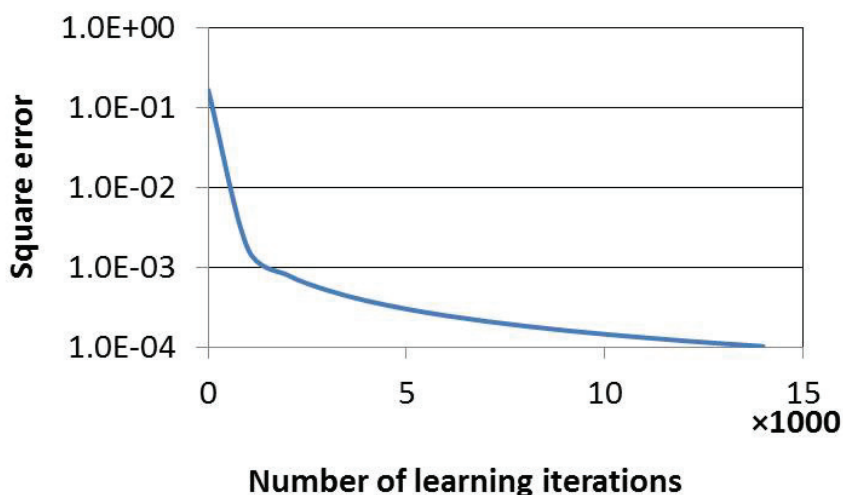


図 4.4 学習パターンにおける学習回数に対するニューラルネットワークの誤差 2 乗和

図 4.5 は、学習パターンにおける閾値に対する FAR と FRR 値の結果を示している。図に示されるように、FAR 値は閾値の増加に対して急激に減少し閾値が 0 付近で 0% となった。また、FRR 値は閾値の増加に対して変化せず 0% となり、閾値が 1 付近で急激に増加した。

これらの結果から、閾値に対する FAR と FRR 値が同時に 0% となる領域があることがわかった。従って、学習パターンに対してニューラルネットワークは本人と他人を明確に区別できる。理想的な本人認証システムでは、閾値に対する FAR と FRR 値の振る舞いは、(1)FAR 値が閾値の増加に対して減少し FRR 値が閾値の増加に対して増加すること、(2)FAR 値と FRR 値が同時に 0% となりまたその閾値長が十分に大きいこと、である。図 4.4 と図 4.5 の結果は、学習パターンに対してニューラルネットワークが正常に動

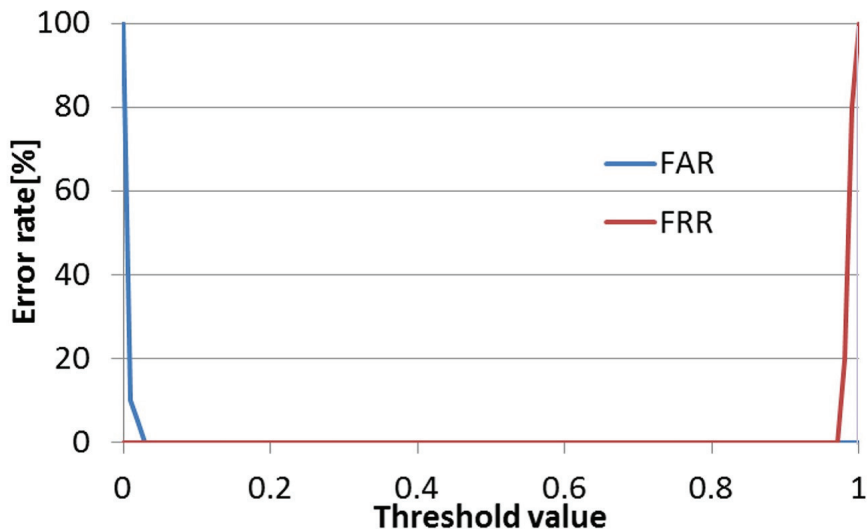


図 4.5 学習パターンにおける閾値に対する FAR と FRR 値

作することを示している。故に、理想的な本人認証システムに用いることができると考えられる。

図 4.6 は、テストパターンにおける閾値に対する FAR と FRR 値の典型的な例を示している。ここでは、種々の結合荷重行列を持つニューラルネットワークを用いた場合のテストパターンとして学習で使用する、本人の未学習パターン (図 4.1 の「○」の右側の「T」) を用いた。図に示されるように、(a)~(d) の異なる結合荷重行列を持つニューラルネットワークでは、閾値に対する FAR と FRR 値の振る舞いは異なった。(a) では FAR と FRR 値が同時に 0% となる領域は存在しなかった。(b)~(d) では閾値長は 0.1~0.9 程度となった。これらの結果から、異なる結合荷重行列を持つニューラルネットワークの閾値に対する FAR と FRR 値の振る舞いは、学習に使用していない本人のテストパターンに対して異なることが分かった。

図 4.7 は、テストパターンとして学習に使用した人の未学習パターンを用いた場合の閾値に対する FAR と FRR 値を示している。ここでは、異なる結合荷重行列のニューラルネットワークを 20 個作成し、それぞれにテストパターンを入力して、得られた結果を閾値別に認証及び拒否に分類し、誤り率を算出した。図に示されるように、FAR 値は閾値の増加に対して減少し、FRR 値は閾値の増加に対して増加した。FAR と FRR 値が同時に 0% となる領域は存在しなかった。

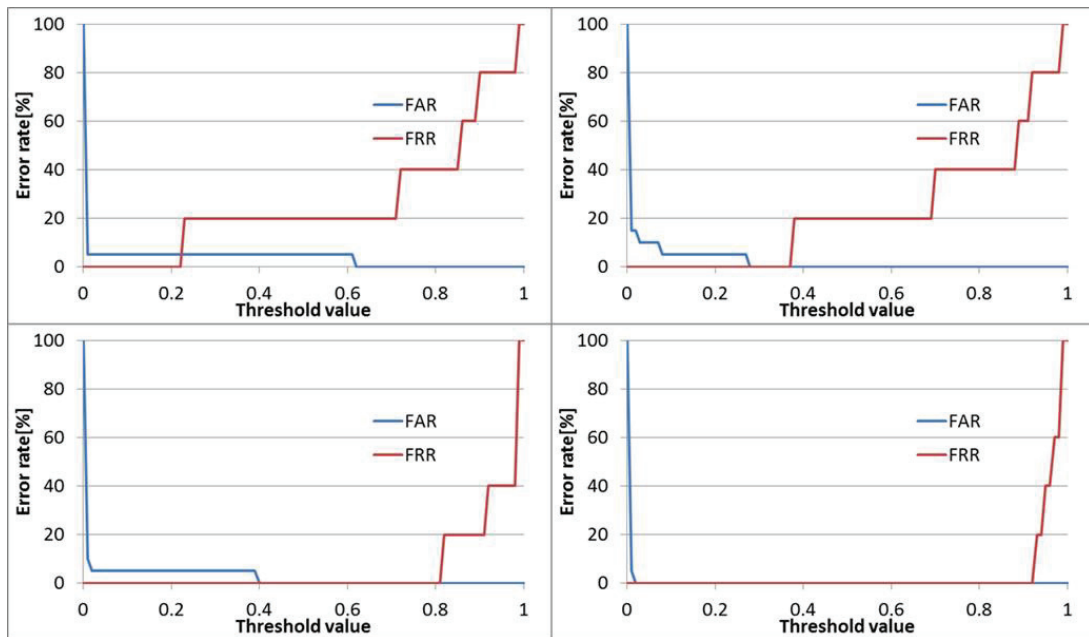


図 4.6 テストパターンにおける閾値に対する FAR と FRR 値の典型的な例 (種々の結合荷重行列を持つニューラルネットワークを用いた場合)

これらの結果より、20 個のニューラルネットワークの総計は、典型的な閾値に対する FAR と FRR 値の振る舞いを示すが、FAR と FRR 値が同時に 0% になる領域が存在しなかった。従って本人認証システムの性能としては不十分であると考えられる。

次に、テストパターンとして学習に使用していない人を用いることによって、本人認証システムの性能に与える影響を調べた。図 4.8 は、これらの結果を示している。ここでは、図 4.7 と同じ 20 個のニューラルネットワークと、学習に使用していない人のパターンを用いた。本人のパターンを使用していないことから、本人を拒否するデータを取得できないため、図 4.8 において FRR 値は図 4.7 と同じ値を使用した。図に示されるように、FAR 値は閾値の増加に対して減少した。図 4.8 を図 4.7 と比較すると、図 4.8 の方が閾値に対する FAR 値の傾きが緩やかであった。これらの結果から、テストパターンとして学習に使用していない人を使用した場合は、学習に使用した人を使用した場合より、FAR 値が大きいこと、すなわち本人認証システムの性能が低いことが分かった。

これらの結果から、1つのニューラルネットワークでは、学習に使用していない人の音声を確実に認証することができないことがわかった。この弱点を補うためにアンサンブル学習を利用した。

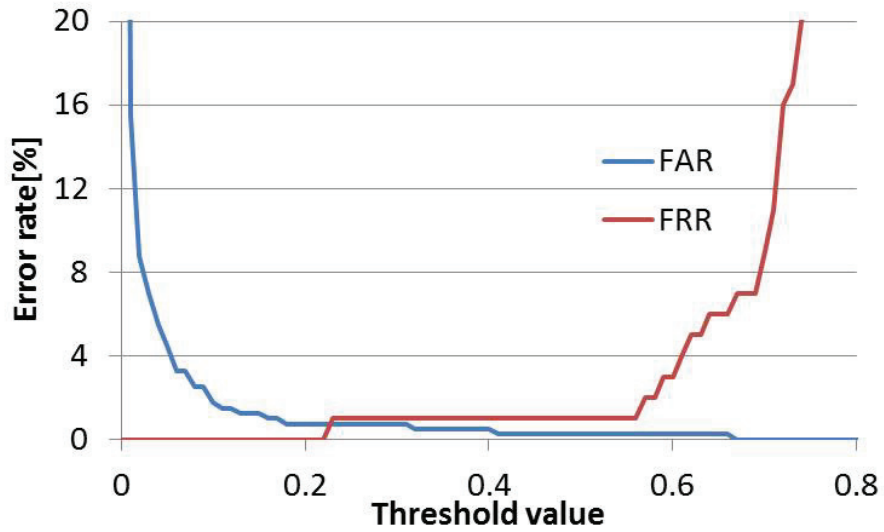


図 4.7 テストパターンとして学習に使用した人の未学習パターンを用いた場合の閾値に対する FAR と FRR 値

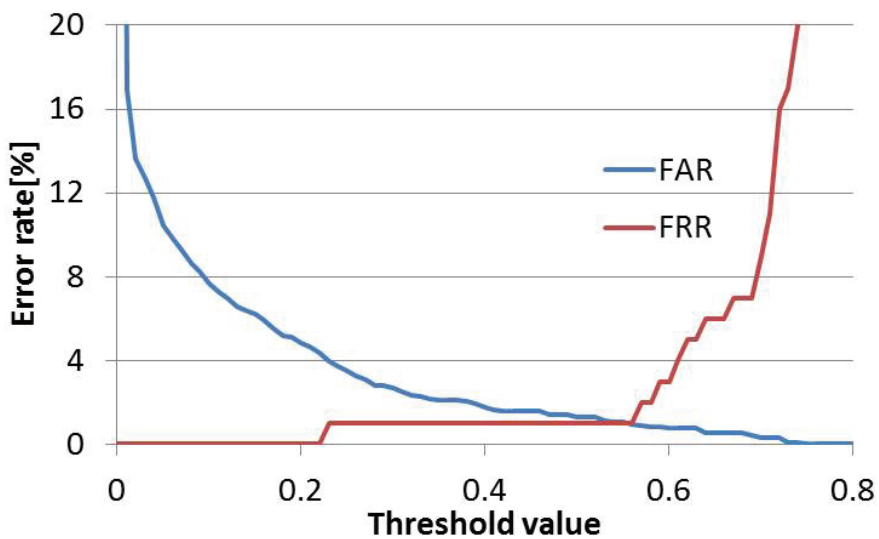


図 4.8 テストパターンとして学習に使用していない人を用いた場合の閾値に対する FAR と FRR 値

4.5 アンサンブル学習を用いた本人認証システム

アンサンブル学習の効果を検証するために、下記 2 種類の検証パターンを想定した。

- ニューラルネットワークの学習に使用した人の音声の未学習の部分で本人認証できるかを検証
- ニューラルネットワークの学習に使用していない人の音声を本人認証できるかを検証

図 4.9 はテストパターンとして学習に使用した人の未学習部分を用いた場合のアンサンブル学習での閾値に対する FAR と FRR 値を示している。図に示されるように、FRR 値は閾値が 0.8 付近まで 0% でその後増加した。FAR 値は閾値の増加に対して減少し閾値が 0.1 付近で 0% となり FAR と FRR 値 が共に 0% となる閾値長が存在した。これらの結果より、アンサンブル学習を用いると、テストパターンとして学習に使用した人の未学習部分を用いた場合、本人と他人を確実に区別できることが分かった。

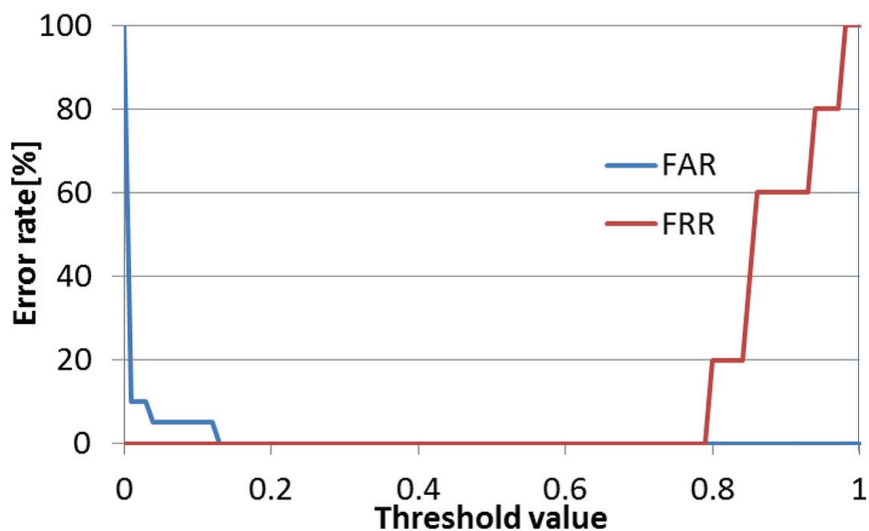


図 4.9 アンサンブル学習での FAR と FRR 値 (テストパターンとして学習に使用した人の未学習部分を用いた場合)

図 4.10 はテストパターンとして学習に使用していない人を用いた場合のアンサンブル学習での閾値に対する FAR と FRR 値を示している。図 4.9 と図 4.10 では同一の（同じパターンで学習した）ニューラルネットワークを利用しているため、FRR（本人拒否率）値は同一の曲線としている。図に示されるように、FAR 値は閾値の増加に対して減少し、閾値が 0.4 付近で 0% となり、FAR と FRR 値が共に 0% となる閾値長が存在した。これらの結果より、学習に使用していない人に対しても本人と確実に区別できることが分かった。

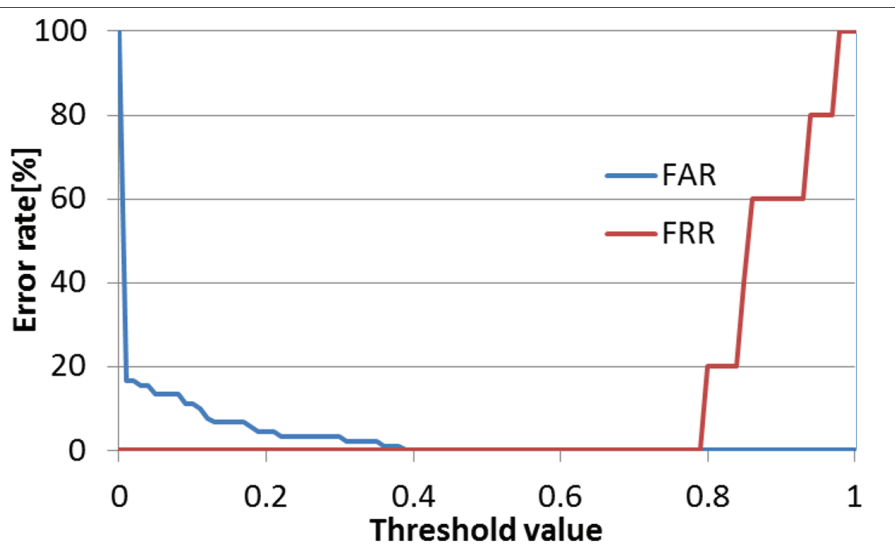


図 4.10 アンサンブル学習での FAR と FRR 値（テストパターンとして学習に使用していない人を用いた場合）

また、図 4.9 を図 4.7 と比較すると、アンサンブル学習を用いると、FAR と FRR 値が共に 0% となる閾値長が存在し、アンサンブル学習の効果が明らかになった。さらに、図 4.10 を図 4.9 と比較すると、図 4.10 の閾値長は、図 4.9 より小さかった。この結果より、テストパターンとして学習に使用した人と使用しない人を用いた場合では、閾値に対する FAR 値の振る舞いに影響を与えることがわかった。従って、アンサンブル学習を用いることによって、個人認証に必要な情報以外（学習に使用していない人の音声）を含む音声に対しても有効であることがわかった。このことは、アンサンブル学習において音声信号に含まれる雑音の影響が抑制されたことによると考えられる。

次に、学習に使用する他人の数が個人認証の精度にどのように影響するか調査した。図 4.11 は、学習に使用する他人の数に対する閾値長を示している。ここでは、テストパターンとして学習に使用していない人のパターンを用いた。図に示されるように、閾値長は学習に使用する他人の数が 3 人までは大きな変化がなくほぼ一定で 0 付近であるが、4 人で増加した。

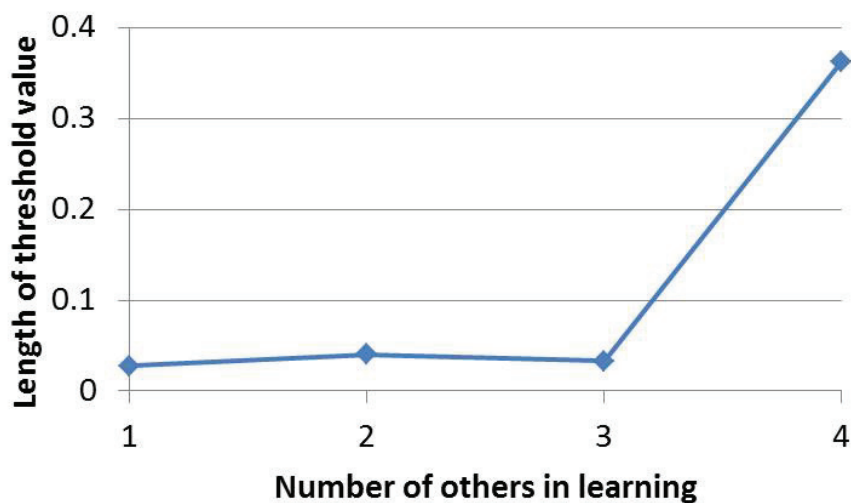


図 4.11 学習に使用する他人の数に対する閾値長

次に、アンサンブル学習に使用するニューラルネットワークの数が閾値長に与える影響を調べた。図 4.12 は、アンサンブル学習に使用するニューラルネットワーク数に対する閾値長を示している。ここでは、テストパターンとして学習に使用する人の未学習パターンを用いた。また、学習を同条件で 100 回行い、一定数のニューラルネットワークで組を作り、アンサンブル学習を行った。つまり、アンサンブル学習に使用するニューラルネットワークの個数が 4 個ならば、25 組のアンサンブル学習を行った結果が得られる。図には、アンサンブル学習した場合の閾値長の平均値と最小値及び最大値を誤差範囲として表している。図に示されるように、閾値長の平均値は、ニューラルネットワークの数に対してほぼ変化せず一定で、閾値長の誤差範囲はニューラルネットワーク数の増加に対して減少した。これらの結果より、アンサンブル学習においてはニューラルネットワークの数が増加すれば閾値長は平均値に収束することがわかった。

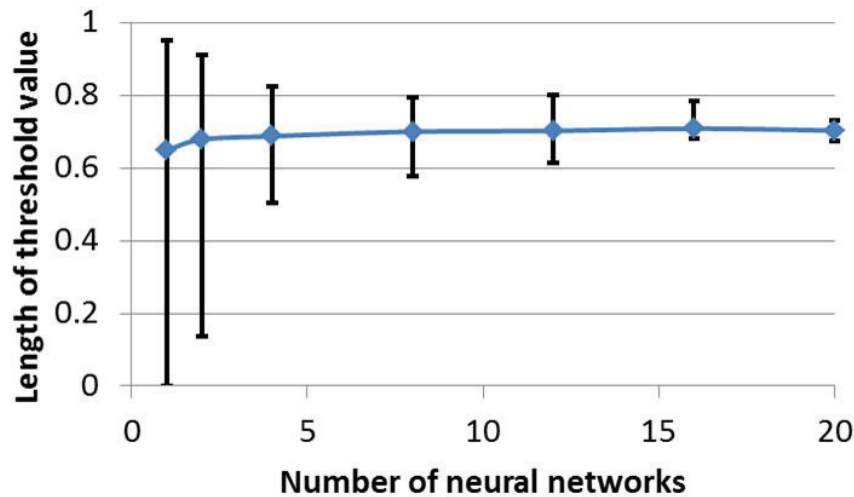


図 4.12 アンサンブル学習に使用するニューラルネットワーク数に対する閾値長

図 4.11 と図 4.12 の結果より次のことが説明できる。図 4.12 では、アンサンブル学習に使用するニューラルネットワークの数を 20 まで増加させても、閾値長に約 0.06 の誤差範囲が存在する。その誤差が存在しても認証システムとして利用するためには、その誤差以上の閾値の幅を持たせなければならない。図 4.11 では、学習に使用する他人の数が 1～3 人の場合、閾値長が 0.06 に達していないため、前述したように、認証システムを構築する際に、閾値長が 0 以下すなわち FAR と FRR 値が共に 0 % の領域が存在しない可能性があり認証システムとして利用できない。従って、学習に使用する他人の数は 4 人以上が必要であることがわかった。認証システムを構築するうえで、学習に使用する他人の数は少なければ少ないほど効率的になるため、4 人を最適値とし、本研究では、学習に使用する他人の数は 4 人とした。使用した他人の数が 4 人の場合の閾値長は 0.06 以上なので、本研究で用いたニューラルネットワークの数は妥当である。

4.6 マルチステップ認証システムの構築

マルチステップの個人認証システムには、N.Fakotakis らの報告がある [2]。その報告では、1 人の被認証者に対して、取得音声を一時間隔で分割した複数の学習パターンを用いて、それぞれの学習パターンに対するニューラルネットワークを構築し、話者認証を行っている。しかし、本論文で紹介するような、単一の学習パターンを用いて複数のニューラルネットワークを構築し、本人を認証する報告は全く無く、それらを組み合わせて個人を認証する報告も無い。我々が構築したマルチステップの個人認証システムを図 4.13 に示す。図に示されるように、入力パターンは A, \dots, N を認証するシステムに同時に入力される。各本人認証システムでアンサンブル学習を行い、個々の本人認証システムの出力値を得る。これらの出力値で個人を認証する。また、これらの本人認証システムはそれぞれ独立である。個人認証システムに新しい人を加える場合は、その人を本人とする本人認証システムを構築し、追加することで実現できる。この場合、ニューラルネットワークの学習に使用する他人のパターンは同一なので、本人のパターンのみで学習パターンを構築できる。さらに、学習に使用されていない未知の他人が個人認証システムに入力された場合、事前に記録した各ニューラルネットワークの出力値と比較すれば、本人以外として拒否できる。

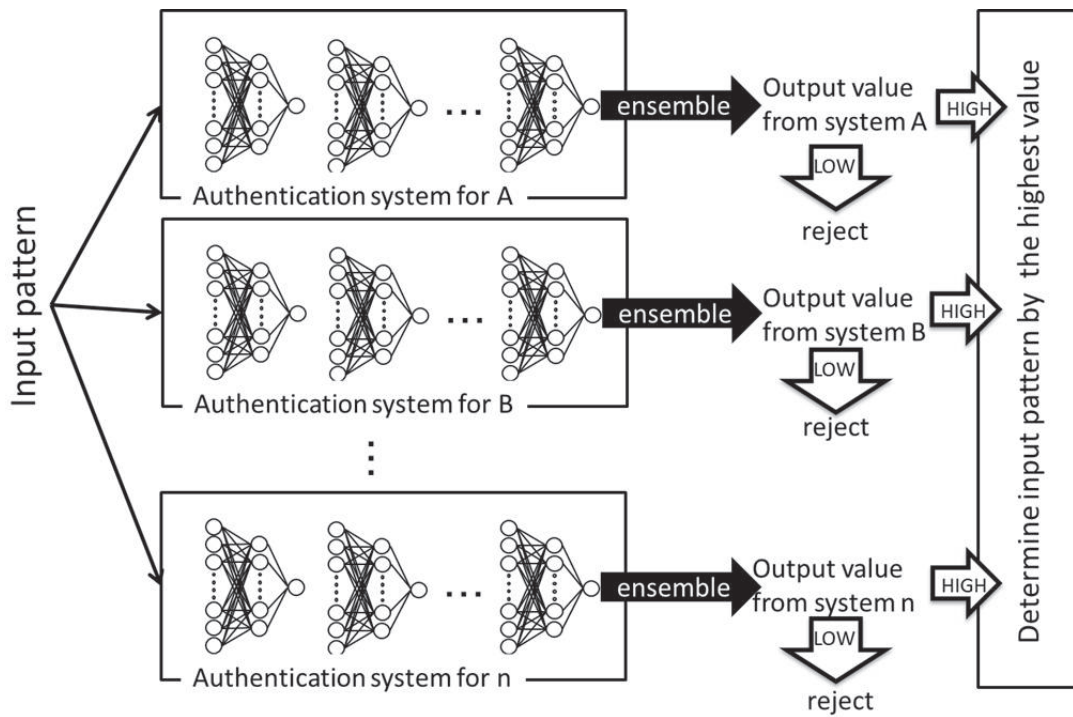


図 4.13 マルチステップの個人認証システムの概要

4.7 マルチステップ認証システムの評価

マルチステップ認証システムを評価するために図 4.2 に示した音声コーパスを用いた。表 4.3 はマルチステップの個人認証システムの結果を示している。

表 4.3 マルチステップの個人認証システムの結果

		Steps				
		a	b	c	d	e
Input patterns	A	0.88124	0.01073	0.00125	0.00297	0.05327
	B	0.00848	0.44645	0.32684	0.00206	0.08732
	C	0.00526	0.00732	0.94845	0.05615	0.00099
	D	0.00351	0.00145	0.00751	0.96380	0.03691
	E	0.02856	0.01464	0.00273	0.29763	0.77969
	X	0.03947	0.05041	0.31856	0.05005	0.08615
	Y	0.19779	0.13136	0.07860	0.00205	0.04018
	Z	0.00162	0.33714	0.48372	0.00708	0.20822

このシステムは 5 人の個人認証を行うため 5 ステップで構成されている。また、各ステップのニューラルネットワークのパラメータには前項までの最適値を用いた。横方向は人 A,B,C,D,E をそれぞれ認証する本人認証システム（ステップ a~e）を示し、縦方向はテストに用いた人 A,B,C,D,E,X,Y,Z の音声から得られた入力パターンを示している。ここでは人 A,B,C,D,E は認証する人（本人）で人 X,Y,Z は学習に使用しておらず認証するステップを持たない人（未学習の他人）である。A の音声のパターン（パターン A）がステップ a~e に入力され各ステップから出力値を得る。その出力値を表の横方向に数値で示した。同様にパターン B~Z も各ステップに入力されそれぞれから出力値を得る。表に示されるようにパターン A が入力された場合各ステップ a~e の出力値はステップ a が最大となった。パターン B~E が入力された場合パターン A の場合と同様にそれぞれのパターンを認証するステップの出力値が最大となった。これらの結果よりパターン A~E が入力された場合それぞれを認証するステップで出力値が最大となるので個人認証できる

ことがわかった。他人のパターン XYZ が各ステップに入力された場合それぞれステップ c,a,c の出力値が最大となった。またシステムを持たないパターン X が入力された場合ステップ c の出力値が最大となるがこの値はステップ c の本人パターンである C の出力値以上とならない。故に人 C でないと判定できる。パターン YZ が入力される場合でもパターン X と同様となり A~E でないと判断できる。これらの結果より、個人認証システムは本人を正しく認証し、他人を拒否した。また未学習の他人が増加すると、間違った認証を行う可能性がある。このことは、本人と他人の境界線の学習が不十分であると考えられる。従って、「アンサンブル学習するニューラルネットワーク数を増加」または「学習に使用する他人の数を増加」を行えば、認証システムはより正確に本人と他人の境界線を学習し、正しく認証することが可能になると思われる。

4.8 結言

音声コーパスに収録されている音声「あ」を使用して階層型ニューラルネットワークに新しいアンサンブル学習が与える効果を明らかにした。これらの結果から以下のことがわかった。

ニューラルネットワークの学習において、結合荷重行列の初期値は乱数で与えられるため同じ学習パターンを用いても異なる結合荷重行列を持つニューラルネットワークが作成される。このため学習毎にニューラルネットワークの出力値が異なる。これらのニューラルネットワークを用いてアンサンブル学習を行った結果、アンサンブル学習を用いた本人認証システムでは FAR と FRR 値が共に 0 % となる範囲が存在した。故に、ニューラルネットワークを用いた音声の本人認証システムにはアンサンブル学習が有効であることがわかった。さらに、本人認証システムを拡張してマルチステップによる個人認証システムを構築した。アンサンブル学習を行った本人認証システムを組み合わせることで構築した個人認証システムでは、本人と他人を確実に区別できることが分かった。さらに、学習に使用していない音声も、マルチステップ認証システムにより本人認証が可能であることがわかった。

第5章

音声信号と血液中アルコール濃度との相関関係

5.1 序言

音声信号には様々な情報が含まれおり、その情報を用いた研究が行われている。これは、音声認識、話者認証、メンタルヘルス及び飲酒などに関する研究である。例えば、健康者とうつ病患者の音声信号において差異が生じること [43]、飲酒による影響が音声信号に含まれること [44]、話者の疲労を検出する試み [45] などが挙げられる。飲酒により人体に変化が起きることは周知の事実であるが、音声信号と血液中のアルコール濃度との相関を調査する報告は少なく、血液中のアルコール濃度が音声信号に及ぼす影響は十分には明らかにされていない。また、音声信号から所望の特徴量を抽出するための研究も十分には行われていない。

本研究では、血液中のアルコール濃度が音声信号に及ぼす影響を明らかにする。そのために、飲酒前後のスペクトルの包絡特性を調査するために線形予想分析を行い、フォルマントに及ぼす影響を推定する。また、ニューラルネットワークを用いた話者認証システムを用いて、音声信号と血液中のアルコール濃度との相関関係を明らかにする。

5.2 血液中のアルコール濃度と音声との関連性

図 5.1 に、飲酒後の時間経過と血液中のアルコール濃度の関係を示している [46]。図に示されるように、血液中のアルコール濃度は飲酒後の時間経過及び飲酒量に依存していることがわかる。ビール 1 本の場合、血液中のアルコール濃度は飲酒後 30 分程度、2 本の場合は 60 分程度で最大になっている。もちろん、個人差によるアルコールの分解速度や自然環境などの条件に左右される。

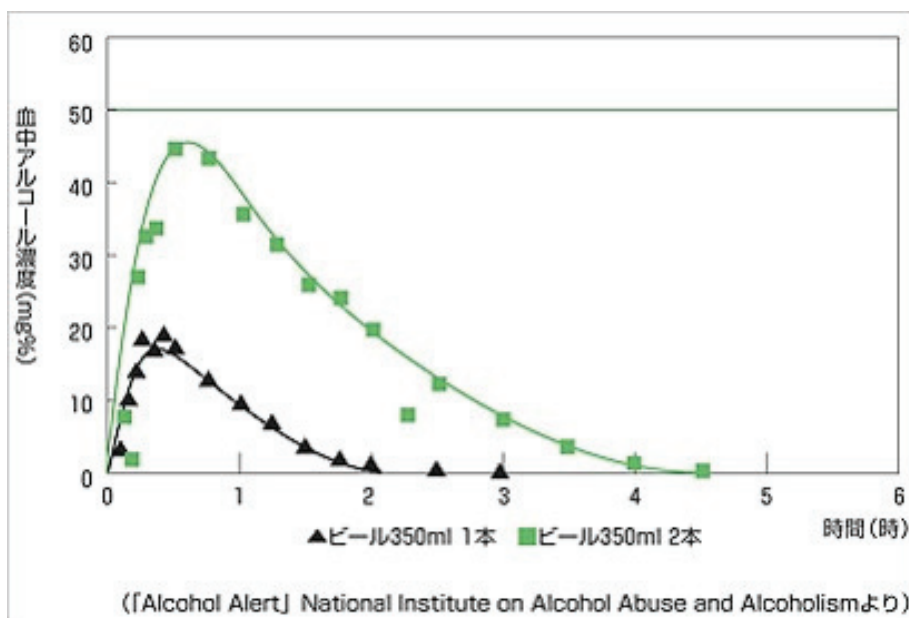


図 5.1 時間経過と血液中アルコール濃度の関係 [46]

図 5.2 に発声の仕組みを示す [47]。音声信号は声帯の (基本周波数) と声道特性 (共鳴周波数) の畳み込みであると考えることができる。周波数領域では、音源スペクトルの倍音成分 (縦線) と声道特性 (周波数特性) の積として表され、周波数スペクトルとして観測される。音のスペクトルは倍音構造 (harmonic structure) をもった線形スペクトル (line spectrum) であり、その微細構造の上に声道伝達関数を作る山が包絡線として重畳された形になっている。母音のスペクトル上の山は低い周波数から 第 1 フォルマント (first formant:F1)、第 2 フォルマント (second formant:F2)、と呼ばれ、声道の共鳴周波

数 (resonance frequency) に対応している。

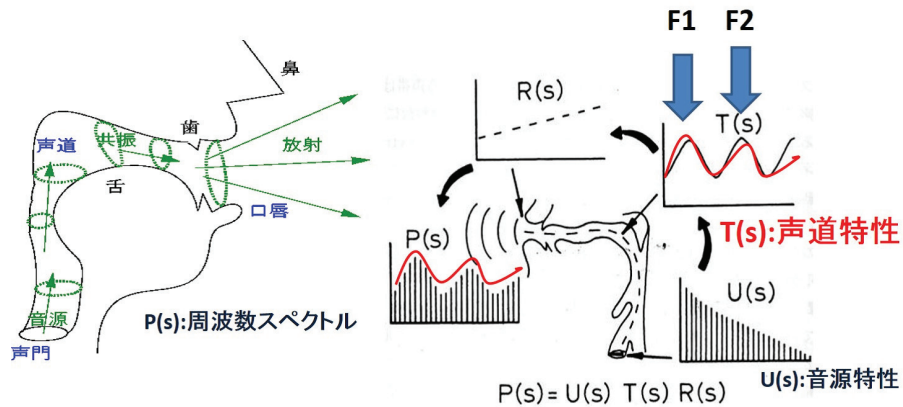


図 5.2 発声の仕組み [47]

図 5.3 は母音「あ」の飲酒後の時間経過による音声信号の変化を示している。図から、飲酒 2 時間後の音声信号は、飲酒前の音声信号とほぼ同等であることがわかる。これは図 5.1 で示された結果と同様、血液中のアルコール濃度が飲酒後、約 2 時間で 0 となることと対応している。また、血液中のアルコール濃度がピークである 30 分後に音声波形が大きく乱れることが分かった。したがって母音「あ」の音声信号の変化は血液中のアルコール濃度に関係していることが考えられる。

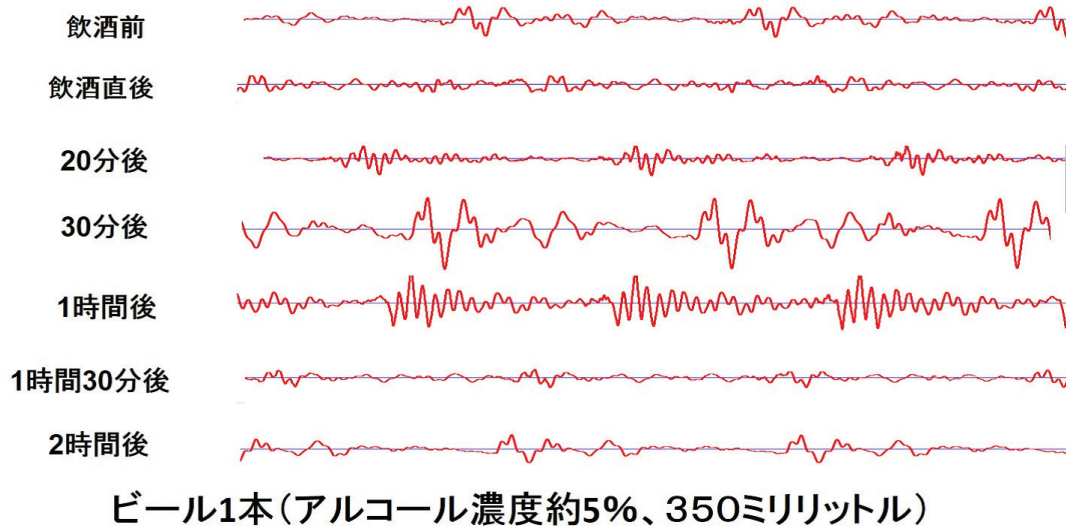


図 5.3 母音「あ」の時間経過における音声信号

5.3 実験方法

5.3.1 話者認証システム

図 5.4 は、ニューラルネットワークを用いた話者認証システムの全体図を示している。取得ユニットでは、男性 1 名のビール 2 本（アルコール濃度約 5 %、700ml）の飲酒前後の母音「あ」を PC とマイクロフォンを用いてサンプリング周波数 48kHz、16bit モノラルで収集した。前処理ユニットでは音声データを 8kHz、8bit でサンプリングして周波数変換し、収集条件の時間毎に線形予測分析を行い、時間毎にパワースペクトル密度の特徴量を得た。その特徴量をニューラルネットワークの入力パターンとした。判定ユニットでは特徴量と血液中のアルコール濃度との相関関係をバックプロパゲーション学習アルゴリズムを備えた 3 階層のニューラルネットワークを用いて明らかにした。

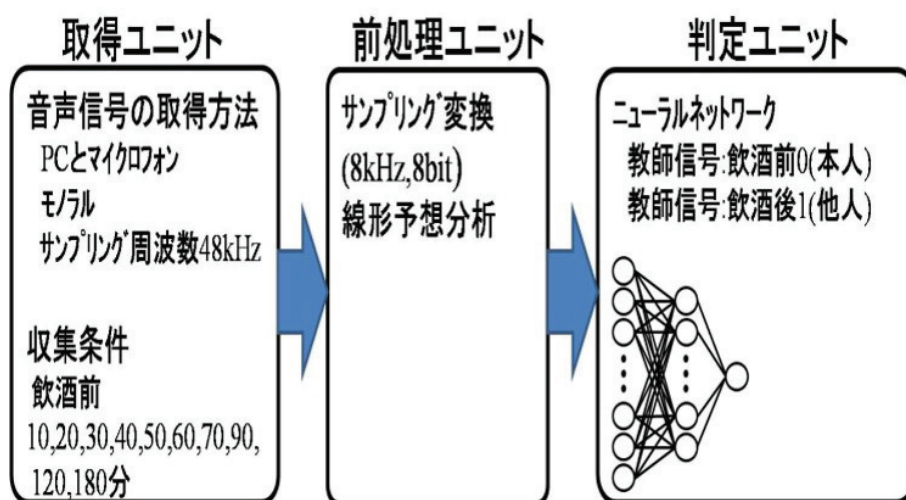


図 5.4 話者認証システムの全体図

5.3.2 線形予測分布

飲酒前後のスペクトルの包絡特性を調査するために、声道内の空気の共鳴周波数である第一フォルマント (F1)、第二フォルマント (F2) について調査した。フォルマントの推定

は、線形予測分析によって得られたスペクトルの包絡特性をピークピッキングすることにより検出された。自己回帰モデルの推定にはユール・ウォーカー法を用いた。線形予測分析の条件や手順をそれぞれ表 5.1 や図 5.5 に示す。

表 5.1 線形予測分析の条件

窓関数	ハミング窓
分析フレーム長	64ms
L P C 次数	12 次
F F T ポイント数	512



図 5.5 線形予測分析の手順

図 5.6 は、飲酒前、飲酒後 60 分、飲酒後 180 分の 1 フレーム (512 ポイント) の音声信

号を示している。音声信号は最大値を 1 として規格化した。

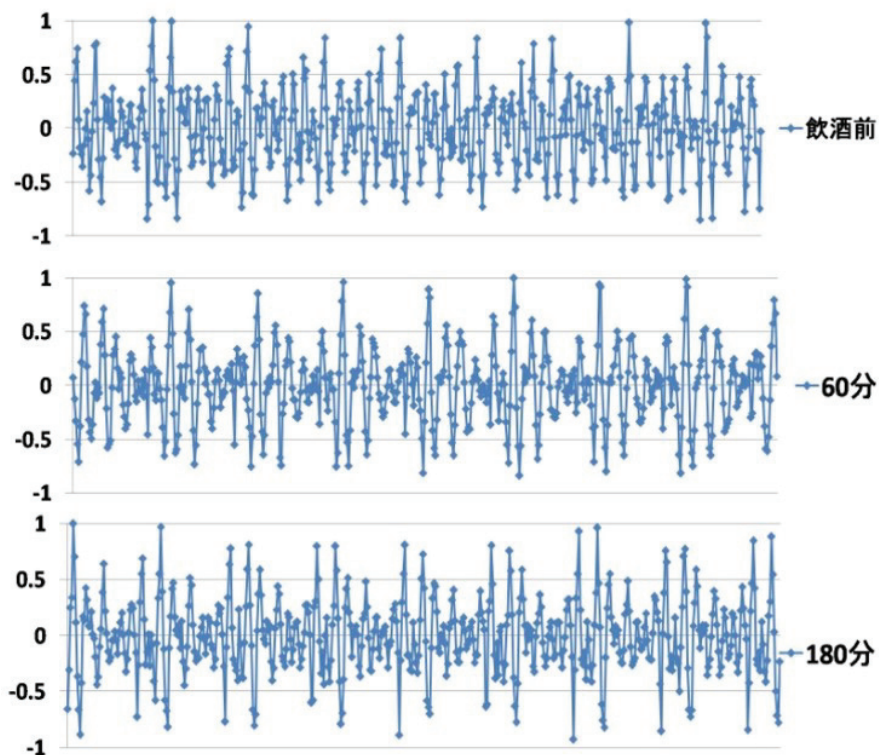


図 5.6 1 フレームの音声信号

図 5.7 は、図 5.6 の音声信号にユール・ウォーカー法による自己回帰全極モデル推定を行った時の線形予測フィルター係数を示している。図に示されるように、飲酒前の変化と飲酒 180 分後の推移はほぼ同一となったが、飲酒の影響が最も顕著と思われる飲酒 60 分後の推移は明らかに異なった。

図 5.8 は、図 5.6 の音声信号に対して線形予測分析を行った時のスペクトルの包絡特性を示している。図に示されるように、この包絡特性の第一ピークが第一フォルマント (F1)、第二ピークが第二フォルマント (F2) である。

5.3.3 ニューラルネットワーク

図 5.8 に示された、0~1kHz のパワースペクトル密度を特徴量としてニューラルネットワークの入力パターンとした。学習は個人内のパターンで行った。教師信号として飲酒前

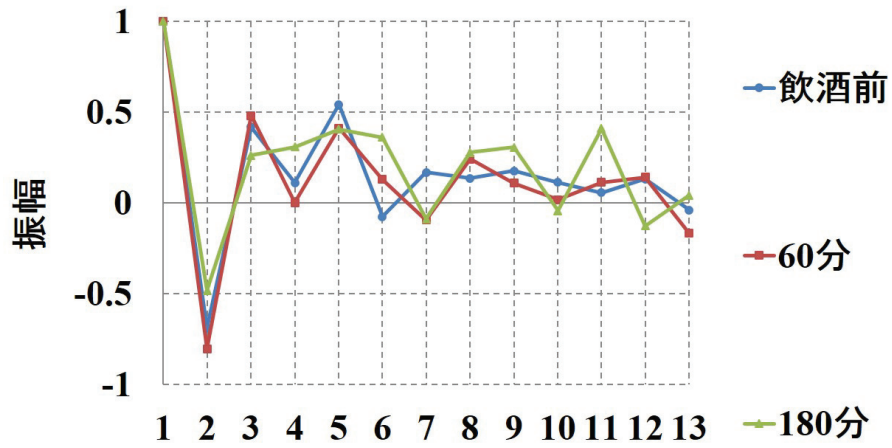


図 5.7 線形予測フィルター係数

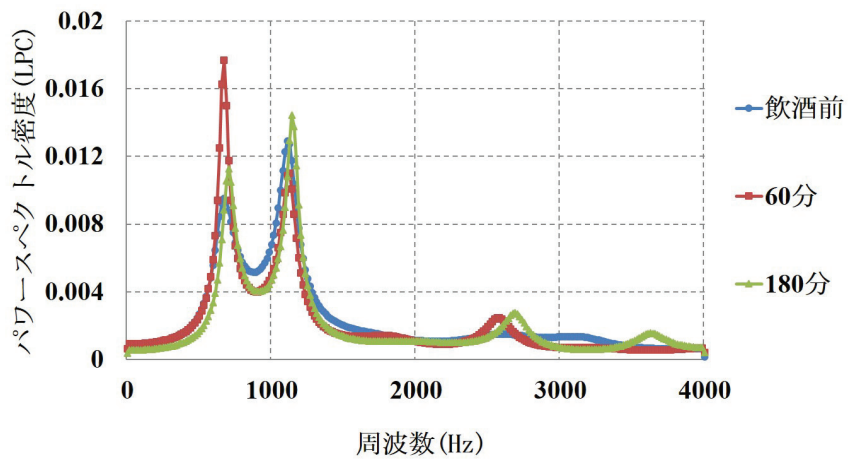


図 5.8 スペクトルの包絡特性

パターン（本人）には”0”を与え、血液中のアルコール濃度が高い飲酒後 60 分のパターン（他人）には”1”を与えた。学習データ及びテストデータは 0 規格化された。学習セット数は飲酒前 20 パターン、飲酒後 60 分 20 パターンの計 40 パターンである。未学習（テス

ト) パターンには時間毎に 40 パターンを用い、時間毎に本人認証を行う。表 5.2 は、ネットワークの設定条件を示している。表に示されているように、学習係数を 0.01、慣性係数を 0.01 とし、誤差二乗和が 0.001 以下あるいは学習回数が 10 万回以上になる時に学習を終了させた。

表 5.2 学習に使用されたニューラルネットワークのパラメータ

入力層ユニット数	64
中間層ユニット数	32
出問層ユニット数	1
学習係数 (ε)	0.01
慣性係数 (α)	0.01
終了条件 (誤差二乗和)	0.00001
教師信号	飲酒前:0, 飲酒後 60 分:1
未学習パターン	飲酒前,10,20,30,40,50, 60,70,90,120,180 分
学習終了条件	10 万回

5.3.4 ニューラルネットワークのテスト及び評価

バイオメトリクス認証に対するシステム性能の評価は、FAR(False Acceptance Rate : 他人受入率) と FRR(False Rejection Rate : 本人拒否率) の曲線によって行われた。図 5.9 は、学習データのみをニューラルネットワークでテストした時の FAR と FRR 曲線である。FAR と FRR 曲線がエラー率 0% で交わる広い閾値が存在し、正常に学習が行われたことを示している。評価は、時間毎の未学習パターンにおいて本人認証を行った時の FRR 曲線で行われ、アンサンブル学習を用いた。アンサンブル学習は同一の学習パターンを用いて行い、結合荷重の初期値の異なるニューラルネットワークを 100 個用いた。各ニューラルネットワークで学習し、テストして得られた出力値を平均する。平均の出力値より FRR 曲線を導き出した。

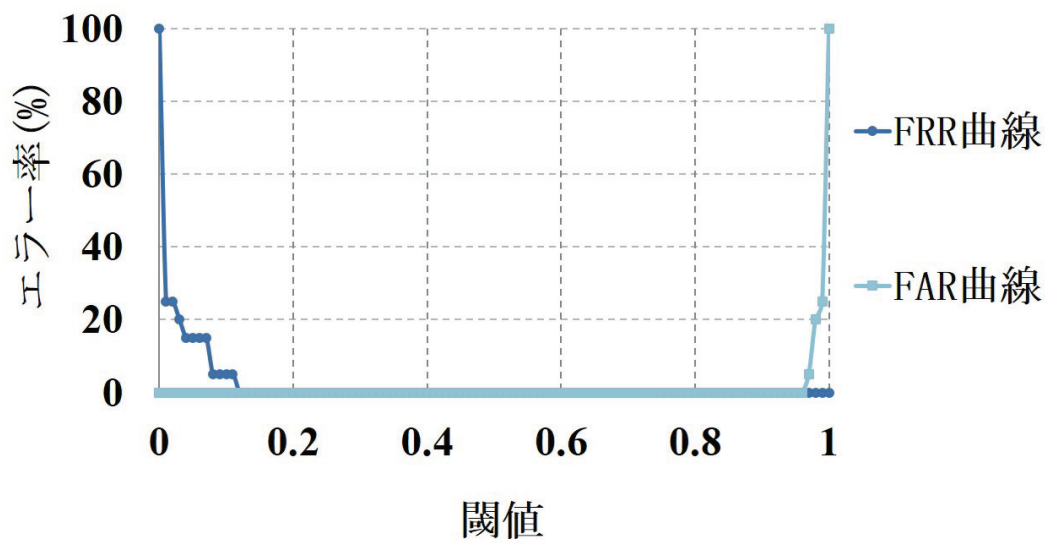


図 5.9 学習パターンに対する認証の結果

5.4 結果と検討

5.4.1 フォルマントの推定結果

図 5.10 は、時間毎の連続 20 フレームから得られた、F1 と F2 のパワースペクトル密度の平均値をプロットした結果を示している。F1 のパワースペクトル密度について、40 分、50 分、60 分において、それ以外の時間より高い値が得られた。

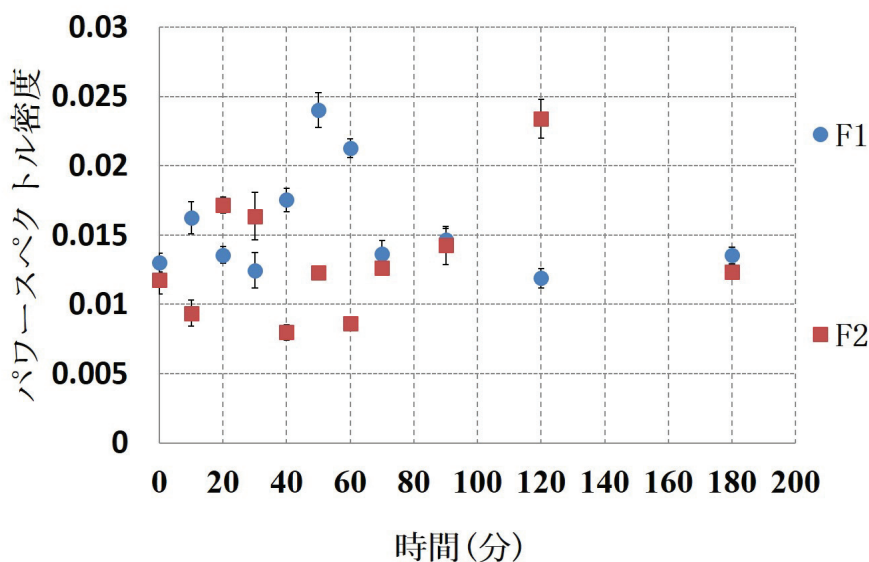


図 5.10 F 1 と F 2 のパワースペクトル密度の平均値

5.4.2 時間毎の本人拒否率曲線

図 5.11 は、飲酒後の経過時間に対する FRR 曲線を示している。図に示されるように、エラー率 100% は、飲酒後に本人（飲酒前）として認められず、飲酒状態であると明確に判断できることを意味している。飲酒後の経過時間に対する FRR 曲線において、閾値に対するエラー率に着目すると、血液中のアルコール濃度の経過時間に対する変化と同様に上下する特性を示した。つまり、エラー率が高い時には血液中のアルコール濃度が高くエラー率が低い時には血液中のアルコール濃度が低いことがわかった。それ故に、血液中ア

アルコール濃度に対する定量的な判断が行える可能性がある。そこで、閾値を 0.9 と定めた

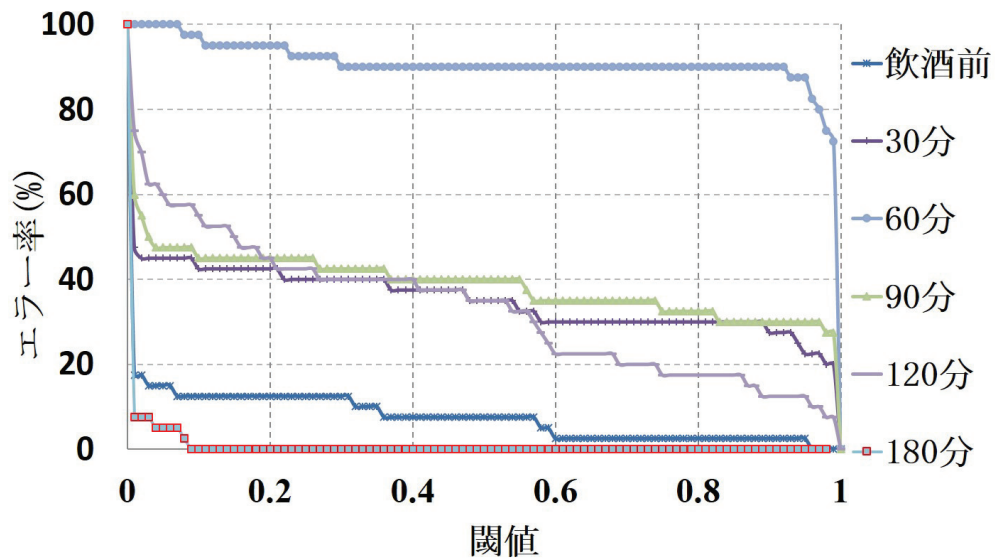


図 5.11 飲酒後の経過時間に対する F R R 曲線

時の時間毎のエラー率を読み取ることで定量的な判断とした。その結果を図 5.12 に示す。ビール 2 本飲酒後の血液中アルコール濃度の経過時間に対する変化は、個人差はあるが飲酒後約 1 時間においてピークとなることがわかる。図 5.12 にしたがってエラー率を定めることにより飲酒後の時間帯を区別することができる。これは、血液中のアルコール濃度の高低を区別できることを意味している。従って、線形予測分析によって得られた第一フォルマント付近のパワースペクトル密度分布が血液中のアルコール濃度の定性的な特徴量を含んでいることを示している。

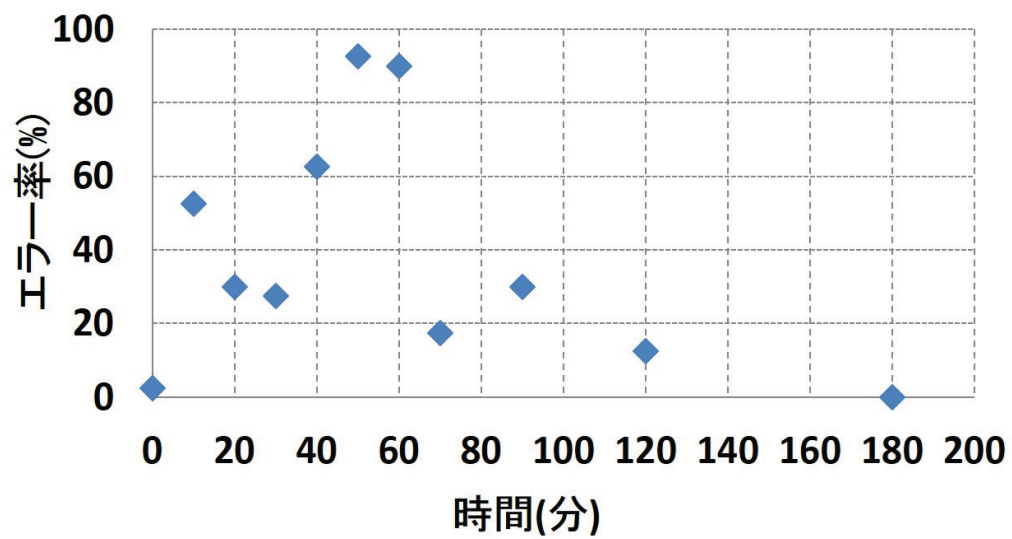


図 5.12 血液中のアルコール濃度の定量化

5.5 結言

ニューラルネットワークを用いた話者認証システムを用いて、音声信号と血液中のアルコール濃度の相関関係を明らかにした。音声信号の形状そのものからも、音声が血液中のアルコール濃度に関係することが観察できたことから、線形予測分布で得たフォルマントのパワースペクトルをニューラルネットワークの入力とした。その結果、ニューラルネットワークの認証率に顕著に影響を与え、本研究で用いたニューラルネットワーク（アンサンブル学習を利用）で、血液中のアルコール濃度が正常値かどうかを判定することが可能であることがわかった。

また、音声信号に様々な影響を与える体調の異常についても本ニューラルネットワークで検出できる可能性があり、音声信号を用いた健康状態評価に活用する1手法が確立されたと考えられる。

第6章

結論

本研究では、インターネット上での個人認証を遠隔で（電話で）取得できる音声信号で実現すべく、ニューラルネットワークを用いた個人認証システムを構築し、その有効性を明らかにした。

指紋画像を用いて個人認証できる3階層ニューラルネットワークを構築し、個人認証における前処理方法および階層型ニューラルネットワークの効果を検証した。その結果、最適な入力層、中間層、出力層の最適な数によって多数の人物を認証できることが分かった。全く学習していない人物の指紋画像でも、十分多くのデータがあれば、個人認証システムは未学習データ（未知の指紋画像）についても認証できることが分かった。更に、ニューラルネットワークの持つ非線形かつパターン認識能力が、指紋画像の質にほとんど影響されず、認証システム構築に効果的であることが分かった。故に、指紋画像の回転取得、ノイズ除去、隆起先鋭化無しに高性能なシステムを構築できた。

次に、音声をFFTで前処理したパワースペクトルを入力とした3階層ニューラルネットワークによる声紋認証システムを構築し、システムが正常に動作するかどうかを判断した。入力したデータが正常であるかどうかを判断するために、学習パターンに対するFARとFRR曲線を見て、エラー0の範囲内で、FARとFRR曲線が交わることを確認した。結果、3階層ニューラルネットワークの誤差二乗和は減少し、声紋認証システムは学習パターンを増やすことで収束することが分かった。さらに、声紋認証において必要なニューラルネットワークの特徴を明らかにするために、サンプリング周波数の減少により入力層ユニット数の削減がシステムに与える影響を調査した。結果、構築したニューラルネットワークシステムは削減した32から256の入力層ユニット数でも有効であった。特に、未学習の音声信号を認証する場合は、入力層ユニット数の減少（サンプリング周波数の減少）がニューラルネットワークの性能向上にとって有効であることが明確になった。

故に、このシステムは音声をサンプリングしてデジタルデータとして伝達している電話を通した入力データを用いても有効に動作することが確認できた。

次に、この基本システム（3階層ニューラルネットワーク）を用い、更に確実な本人認証を行うことを目的にアンサンブル学習法を用いた。音声コーパスに収録されている音声「あ」を使用して、階層型ニューラルネットワークを用いた話者認証システムの性能に新しいアンサンブル学習が与える効果を明らかにした。それらの結果から、以下のことがわかった。

ニューラルネットワークの学習において、結合荷重行列の初期値を乱数で与えることにより、同じ学習パターンを用いても異なる結合荷重行列を持つニューラルネットワークが作成される。このことにより、学習毎にニューラルネットワークの出力値が異なる複数のシステムが生成できる。このようにして生成した複数のニューラルネットワークを用いてアンサンブル学習を行った。様々な結果を出力するニューラルネットワークのアンサンブル学習を用いて構築した本人認証システムでは、FAR 値と FRR 値が共に 0% となった。従って、ニューラルネットワークを用いた音声による本人認証システムにおいてはアンサンブル学習が有効であることがわかった。また、アンサンブル学習を用いた階層型ニューラルネットワークでは、音声信号に含まれる雑音を抑制することができることも確認できた。

さらに、この本人認証システムを拡張して、マルチステップの個人認証システムを構築した。このシステムは複数人のアンサンブル認証システムにテストデータを同時に入力することで高速かつ正確に個人認証できる特徴があり、実験では複数人の個人認証を実現できた。また、このシステムは認証する個人が増加した時には、その人物を認証できるアンサンブル認証システムを並列で加えることで構築でき、実用的な個人認証システム構築に活用できることが分かった。さらに、論文内には記述していないが、この方式は個人認証のために一般的に用いられる本人特有のデータそのものを保持しないため、個人情報の漏えいが発生し難く、これからの高度情報化社会で必要となる重要なファクターになり得ることを追記する。

最後に、音声信号には様々な情報が含まれていることが研究されており、その中でも応用が期待されている音声信号と血液中のアルコール濃度の相関関係について検証した。本研究では、本人認証を行う 3 階層ニューラルネットワークの認証率に与える影響から、その相関関係を調査した。結果、線形予想分析によって得られた第一フォルマント付近のパワースペクトル密度分布が血液中のアルコール濃度に依存し、音声による本人認証システムに影響を与えることが分かった。

これらの結果から、ニューラルネットワークがセキュリティに必要な個人認証に有

用（認証率、高速性、変更性において）であり、かつ人間自身のセキュリティ（健康状態）維持にも利用できる可能性があることを確認した。

このシステムが各家庭や事務所に設置できれば、漏えいや忘却の不安がある ID やパスワードを用いることなくドアの開閉を指令したり、電話越しでの本人確認や、車を運転するときのアルコール検出、家にいながらの電話越しの健康診断などに活用できることを期待する。

本研究結果は未だ数十人に対する個人認証に応用できることを証明しただけである。インターネット上で扱う数万人から本人を特定するには更なる高度化が必要になるが、情報化社会でのセキュリティ高度化に多少なりとも貢献できたと考える。

謝辞

本研究は、鳥取大学大学院工学研究科博士後期課程において行われたものであります。ご指導ご鞭撻賜りました主指導教官の工学部工学研究科情報エレクトロニクス専攻、岸田悟教授に心から感謝いたします。2006年にインターネット上での本人確認方法を岸田教授に相談したことから始まり、8年かかって論文のまとめることができました。この間、研究はもとより公私にわたって大変お世話になりました。本当にありがとうございました。

音声信号の構造に関し、副査として専門的な助言をいただきました工学研究科情報エレクトロニクス専攻、伊藤良生教授に御礼申し上げます。

入学当初より、画像処理に対する技術をご教授いただき、副査として助言いただきました工学研究科情報エレクトロニクス専攻、李仕剛教授に御礼申し上げます。

社会人研究生となるにあたり、入学をお勧めいただき、学費の補助までしていただいたセコム山陰株式会社、特に入学当時に代表取締役社長であられた吉岡健二郎様には、様々なアイデア、序言をいただいたこと感謝いたします。ありがとうございました。

学内の事務関係では、鳥取大学工学部電子ディスプレイ研究センター及び先端融合研究センターの職員の方々には大変お世話になりました。ありがとうございました。

長期間にわたり、研究に協力いただいた電磁エネルギー応用工学研究室の皆様には、多くのアドバイスをいただきました。特に、本研究を進めるにあたり、研究のためのプログラム作成や検証環境を準備していただいた田中章浩君、川野正幸君には長期間に渡る支援をいただき心から感謝いたします。

最後に、病床にいながらも常に応援してくれていた妻、ならびにこの間家庭を守っていただいた家族に感謝いたします。ありがとうございました。

参考文献

- [1] D. P. Morgan and C.L.Scofield: Neural networks and speech processing, Kluwer Academic Publisher, 1991.
- [2] N.Fakotakis and J.Sirigos: A high performance text independent speaker recognition system based on vowel spotting and neural nets, IEEE International Conference on Acoustics, Speech, and Signal Processing(ICASSP), pp.661-664, 1996.
- [3] J.P.Campbell Jr.: Speaker recognition: A tutorial, Proc.IEEE, Vol.85, No.9, pp.1437-1462, 1997.
- [4] D.O'Shaughnessy: Speech communications, IEEE Press, 2000.
- [5] K.Amit and B.Biswas: Speaker identification from voice using neural networks, Scientific & Industrial Research, Vol.61, No.8, pp.599-606, 2002.
- [6] 田中章浩, 小林光, 岸田悟, 長谷川弘, 渡邊徹: 階層型ニューラルネットワークを用いた声紋による個人認証システムの構築, 電子情報通信学会技術研究報告, ニューロコンピューティング, Vol.107, No.542, pp.427-431, 2008.
- [7] 小林光, 田中章浩, 木下健太郎, 岸田悟: 声紋による個人認証システムの構築, 電子情報通信学会技術研究報告, ニューロコンピューティング, Vol.108, No.480, pp.13-17, 2009.
- [8] S.Onishi, A.Tanaka, H.Hasegawa, K.Kinoshita and S.Kishida: Construction of individual identification system using voice in three-layered neural networks, International Symposium on Intelligent Signal Processing and Communication Systems(ISPACS), pp.635-637, 2009.
- [9] F.N.Tuang, J.L.W.Radmaker, E.C.Alocilja, F.J.Louws and F.J.Bruijn: Identification of bacterial rep-PCR genomic fingerprints using a backpropagation neural network, FEMS Microbiology Letters, Vol.177, pp.249-256, 1999.

- [10] H.Hasegawa, A.Tanaka, S.Kishida and K.Nishimura: Preprocessing method of fingerprint images in layered neural networks for individual identification, International Symposium on Intelligent Signal Processing and Communication Systems(ISPACS), pp.267-270, 2006.
- [11] 田中章浩, 大西俊輔, 木下健太郎, 岸田悟: マニューシャ法により得られた指紋の特徴点とニューラルネットワークを用いた認証システムの構築, 電子情報通信学会技術研究報告, ニューロコンピューティング, Vol.109, No.461, pp.237-242, 2010.
- [12] K.Ito, T.Aoki, H.Nakajima, K.Kobayashi and T. Higuchi: A palmprint recognition algorithm using phase-only correlation, IE-ICE Trans. Fundamentals, Vol.E91-A, No.4, pp.1023-1030, 2008.
- [13] N.Miura, A.Nagasaka and T.Miyatake: Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification, Machine Vision and Applications, Vol.15, No.4, pp.194-203, 2004.
- [14] Y.Zhu, T.Tan and Y.Wang : Biometric personal identification based on iris patterns, IEEE International Conference on Pattern Recognition(ICPR), Vol.2, pp.801-804, 2000.
- [15] C.Kanali, H.Murase and N.Honami: Shape identification using a charge simulation retina model, Mathematics and Computers in Simulation, Vol.48, pp.103-118, 1998.
- [16] A.Martinez and H.Jia: Support vector machines in face recognition with occlusions, IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR), Neural Networks, Vol.14, pp.795-803, 2001.
- [17] I.Nakanishi, N.Nishiguchi, Y.Ito and Y.Fukui: On-line signature verification method using adaptive algorithm in wavelet transform domain, International Technical Conference on Circuits/Systems, Computers and Communications(ITC-CSCC), Vol.1, pp.385-388, 2002.
- [18] <http://www.atmarkit.co.jp/fsecurity/special/11biomet/biomet02a.html>
- [19] 三村正人, 河原達也: CSJ を用いた日本語講演音声認識への DNN-HMM の適用と話者適応の検討, 情報処理学会研究報告, 音声言語情報処理, Vol.2013-SLP-97, No.9, pp.1-6, 2013.
- [20] T.Shimizu, N.Furuse, T.Yamazaki, Y.Ueta, T.Sato and S.Nagata: Chaos of vowel /a/ in Japanese patients with depression: A Preliminary Study, Journal of Occupational Health, Vol.47, pp.267-269, 2005.

- [21] N.Fujiwara, A.Tanaka, K.Kinoshita and S.Kishida: Effect of alcohol concentrations in blood on voice signals from human body, International Workshop on Nonlinear Circuits, Communication and Signal Processing(NCSP), pp.712-715, 2012.
- [22] 上田修功, 中野良平: アンサンブル学習における汎化誤差解析, 電子情報通信学会論文誌 D-2, Vol.J80-D-2, No.9, pp.2512-2521, 1997.
- [23] D.Luo and C.Ke: A comparative study of statistical ensemble methods on mismatch conditions, Proc. IJCNN2002, pp.59-64, 2002.
- [24] 上田修功: アンサンブル学習の新展開, 電子情報通信学会技術研究報告, ニューロコンピューティング, Vol.102, No.381, pp.31-36, 2002.
- [25] T.Sasaki, K.Kinoshita, S.Kishida, Y.Hirata and S.Yamada: Ensemble learning in the systems of neural networks for detection of abnormal shadows from X-ray images of lungs, International Workshop on Nonlinear Circuits, Communication and Signal Processing(NCSP), pp.712-715, 2012.
- [26] D.E.Rumelhart, G.E.Hinton and R.J.Williams: Learning representations by back-propagating errors, Nature, Vol.323, No.9, pp.533-536, 1986.
- [27] <http://milan.elec.ryukoku.ac.jp/kobori/resume/int/image/int2-2.jpg> 生体ニューロン.
- [28] 中野馨”ニューロコンピューターの基礎” コロナ社.
- [29] 大学共同利用機関法人 情報・システム研究機構 国立情報学研究所 音声資源コンソーシアム 基盤研究「日本語方言の地域差」方言音声コーパス (GSR-JD), 板橋秀一, 筑波大学,2006.
- [30] 熊沢逸夫”学習とニューラルネットワーク” 森北出版株式会社.
- [31] 徳高平蔵, 岸田悟, 藤村喜久郎”自己組織化マップの応用” 海文堂出版株式会社.
- [32] <http://www.brain.kyutech.ac.jp/furukawa/note/rbf/rbf.html> RBF ニューラルネットワーク, 九州工業大学 生命体工学研究科.
- [33] <http://ja.wikipedia.org/wiki/シグモイド関数>, 出典:フリー百科事典「ウィキペディア (Wikipedia)」.
- [34] 馬場則夫, 小島史男, 小澤誠一”ニューラルネットの基礎と応用” 共立出版株式会社.
- [35] <http://repository.aichi-edu.ac.jp/dspace/bitstream/10424/241/1/kenshi571519.pdf> 小川秀夫”二次元パターン空間を用いた新しい点パターン・マッチング法”.
- [36] M.Minsky and S.Papert, ”Perceptron—an essay in computational geometry,”

- MIT Press (1943).
- [37] T.Aoyama, Y.Suzuki, and H.Ichikawa, *J. Med. Chem.*, 33, 2583-2590 (1990).
 - [38] 江原義郎”ユーザズデジタル信号処理”東京電機大学出版局.2007
 - [39] E.G. ウィリアムズ著, 吉川茂・西條献児訳”フーリエ音響学”シュプリンガー・ジャパン株式会社.
 - [40] 江原 義郎, デジタル信号処理 2007.
 - [41] 青木 直史, デジタル・サウンド処理入門—音のプログラミングと MATLAB(Octave・Scilab) における実際 2006.
 - [42] 日本心理テスト研究所株式会社 <http://sinri.co.jp/>
 - [43] T. Shimizu, N.Furuse, T. Yamazaki, Y.Ueta, T. Sato and S. Nagata, ”Chaos of Vowel /a/ in Japanese Patients with Depression: A Preliminary Study”, *J Occup Health* 2005,47, pp.267-269,2005.
 - [44] N.Fujiwara, A.Tanaka, K.Kinoshita and S.Kishida, ”Effect of alcohol concentrations in blood on voice signals from human body”, 2012 International Workshop on Nonlinear Circuits, Communication and Signal Processing (NCSP'12), pp.712-715,2012.
 - [45] 松村寿枝, ”音声を用いた話者疲労の推定”, 奈良工業高等専門学校, 研究紀要, 第43, pp.49-52,2007.
 - [46] <http://www.arukenkyo.or.jp/health/prevention/index.html>
 - [47] 荒井隆行, 菅原勉 (監訳), 音声の音響分析, 海文堂, 1996.
 - [48] Khaled Ahmed Nagaty, ”On learning to estimate the block direction image of a fingerprint using a hierarchical neural network”, *Neural Networks*, vol.16, pp.133-144, 2003.
 - [49] C.Quek, K.B.Tan, and V.K.Sagar, ”Pseudo-outer product based fuzzy neural network fingerprint verification system”, *Neural Networks*, vol.14, pp.305-323, 2001.
 - [50] Cappelli.R, Lumin.A, Mario.D, and Maltoni.D, ”Fingerprint classification by directional image partitioning”, *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol.21(5), pp.402-421, 1999.
 - [51] J.Cheng and J.Tian, ”Fingerprint enhancement with dyadic scale-space”, *Pattern Recognition Letters*, vol.25, pp.1273-1284,2004.
 - [52] S.Prabhakar, A.K.Jain and S.Pankanti, ”Learning fingerprint minutiae location and type”, *Pattern Recognition*, vol.36, pp.1847-1857, 2003.

- [53] F.N.Tuang, J.L.W.Rademaker, E.C.Alocilja, F.J.Louws and F.J.deBruiji,"Identification of bacterial rep-PCR genomic fingerprints using a backpropagation neural network", FEMS microbiology letters, vol.177, pp.249-256, 1999.
- [54] C.Kanali, H.Murase and N.Honami,"Shape identification using a charge simulation retina model", Mathematic and computers in simulation, vol.48, pp.103-118, 1998.
- [55] A.Delorme and S.J.Thorpe,"Face identification using one spike per neuron: resistance to image degradation", Neural Networks,vol.14, pp.795-803, 2001.
- [56] I.Nakanishi, N.Nishiguchi, Y.Itoh and Y.Fukui,"On-line Signature Verification Method Using Adaptive Algorithm in Wavelet Trans Domain", Proc.ITC-CSCC 2002, vol.1, pp.385-388, 2002.
- [57] K.Nishimura, S.Kishida, T.Watanabe,"Effects of Preprocessing and Layered Neural Networks on Individual Identification", IEEE MWSCAS-2004, pp.77-80, 2004.
- [58] K.Nishimura, S.Kishida, T.Watanabe,"Improvement of Preprocessing Method on Fingerprint Identification System by Layered Neural Networks", IEEE NSIP 2005, 2005.
- [59] C.H.Park and H.Park,"Fingerprint classification using fast Fourier transform and nonlinear discriminate analysis", Pattern Recognition, vol.38, pp.495-503, 2005.
- [60] A.P.Fitz and R.J.Green,"FINGERPRINT CLASSIFICATION USING A HEXAGONAL FAST FOURIER TRANSFORM",Pattern Recognition, vol.29, pp.1587-1597, 1996.
- [61] Satoru Kishida, K.Nishimura and Akihiro Tanaka,"Preprocessing of Fingerprint System in Three Layered Neural Networks", IEEE CIRAS 2005, 2005.

研究業績

- 学術論文

- (1) 長谷川弘, 木下健太郎, 岸田悟, ”ニューラルネットワークを用いた高性能な話者認証システムの構築”, Journal of Signal Processing, Vol.18, No.1, pp.29-38, Jan. 2014.

- 国際会議 (査読付 Proceedings)

- (1) Hiroshi Hasegawa, Kentaro Kinoshita, Satoru Kishida, ”Effect of the Number of Input Layer Units on Performance of Voiceprint Identification Systems Using Layered Neural Network”, Proceeding of 2010 International Workshop on Smart Info-Media Systems in Asia (SISA 2010) September 8-9 2010 Diamond Hotel Malate Manila Philippines.
- (2) Hiroshi Hasegawa, Akihiro Tanaka, Koichi Nishimura and Satoru Kishida, ”Preprocessing Method of Fingerprint Images in Layered Neural Networks for Individual Identification”, Proc.2006 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2006) IEEE.
- (3) Hiroshi Hasegawa, Masayuki Kawano and Satoru Kishida, ”Effect of alcohol concentrations in blood on power spectral density distribution from voice signals”, (ISCT2014).

- 研究報告

- (1) Shunsuke Onishi, Akihiro Tanaka, Hiroshi Hasegawa, Kentaro Kinoshita, and Satoru Kishida, ”Construction of Individual Identification System using Voice in Three-layered Neural Networks”, Proc. 2009 international Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2009) IEEE WP2-B-3, pp.635-637, 2009.
- (2) 大西俊輔, 長谷川弘, 木下健太郎, 岸田悟, ”ニューラルネットワークを用いた声紋認証システムにおける入力層ユニット数の効果”, 信学技報 NC2009, pp.255-259, 2013.

- 学位論文の章と研究業績（学術論文，国際会議）の対応

	題目	雑誌名，会議名
第 2 章	Preprocessing Method of Fingerprint Images in Layered Neural Networks for Individual Identification	ISPACS 2006
第 3 章	Effect of the Number of Input Layer Units on Performance of Voiceprint Identification Systems Using Layered Neural Network	SISA2010
第 4 章	ニューラルネットワークを用いた 高性能な話者認証システムの構築	Journal of Signal Processing
第 5 章	Effect of alcohol concentrations in blood on power spectral density distribution from voice signals	ISCT2014