TR－H－035

音声情報処理への
ニューラルネットワークの応用

片桐 滋
杉山 雅英

1993. 10. 21

ATR人間情報通信研究所

〒619-02 京都府相楽郡精華町光台2-2　☎07749-5-1011

ATR Human Information Processing Research Laboratories

2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-7749-5-1011
Facsimile: +81-7749-5-1008

音声情報処理へのニューラルネットワークの応用

片桐　滋（ＡＴＲ人間情報通信研究所）
杉山　雅英（会津大学）

１９９３年１０月２１日

まえがき

本稿は、近々発刊が予定されているＡＴＲ先端科学技術シリーズの１冊、ニューラルネットワーク応用の１部として準備されたものである。音声情報処理やパタン認識に関する話題を網羅し、かつ重要な技術の原理を比較的平易に解説していることから、本書は初学者のための入門書としても役立つものと思われる。そこで、正式の刊行の前にテクニカルレポートとして内部の利用に供することとする。独学用解説書あるいは研究所内におけるセミナー資料として用いられれば幸いである。

第 4 章

音声情報処理への応用

本章では、音声情報処理に対するニューラルネットワークの応用例を紹介する。初めに、本章を読み進めるために必要と思われる、音声信号処理固有の技術と数理的基礎との簡単な準備を行う。次に、応用例、特に、音声パタン分類と音声特徴抽出とに関する応用例を紹介する。続いて、ニューラルネットワークによる実現をも包含する、新しいパタン認識器設計理念の開発例を紹介する。この最後の話題は、決して音声情報処理に固有のものではない。しかし、開発の経緯を考慮し、本章で紹介することとする。

本章には、特にパタン認識に関連する話題あるいは用語が繰り返し登場する。話を進める前に、いくつかの用語を整理しておこう。まず、認識を以下のように定義しよう。即ち、認識とは、特徴抽出と分節化、分類の３つの部分的過程を経て、認識機械（認識器）に与えられた入力に含まれる認識対象である音声や文字、画像等のパタンを課題に予め設定されている類に割り当てることである。ここで、入力は、課題目的や処理の便宜に応じて設計された特徴形式で表現されているものとする。この望ましい特徴表現を設計する過程が特徴抽出である。また分節化とは、入力から認識対象であるパタンを切り出す操作のことである。なお、このような定義においては、パタンと分節とはほぼ同義である。最後に、分類とは、この分節化されたパタンを予め準備されている類の１つに割り付けることである。特に、本章で扱う分類過程は、音声（パタン）認識に特徴的な言語処理の過程を含んでいるものとしよう。

本書の性格上、ＡＴＲの研究成果を紹介するために紙面の大半を費やさざるを得ない。従って、紹介する内容には偏りがある。しかし、的確な技術展望を得るために必要な話題は、ＡＴＲの成果いかんを問わず、できる限り幅広く取り上げているつもりである。本章が、音声情報処理の高度化を目指したニューラルネットワーク技術開発の、広大で今なお広がり続けている世界を垣間見る一助になれば幸いである。

4.1 簡単な準備

初めに、本題の理解に必要と思われる以下の５つの話題、１）音声認識のための特徴表現と、２）統計的パタン分類、３）学習ベクトル量子化（Learning Vector Quantization (LVQ)）、４）音声パタン分類のための基本技術、５）言語処理とに関する簡単な解説を行う。

4.1.1 音声認識のための特徴表現

実際に、｛京都｝と発話して見よう。時間の経過と共に発話される音声は、時間構造を持つ音響的時間信号である。

この音響的時間信号を観察するためには、**サウンドスペクトログラム**と呼ばれる特徴表現が便利である。これは、時間－周波数分析の結果を２次元平面上に濃淡表示するものであり、しばしば犯人捜査等に登場する声紋の仲間である。図４．１は、｛京都｝の音声波形とサウンドスペクトログラムの１例を示している。サウンドスペクトログラム中、横軸は時間軸であり、縦軸は周波数軸である。濃淡は各時間－周波数点におけるエネルギー強度を示している。表示が濃い程エネルギーは大きい。

話す速度に応じて、｛京都｝の単語音声パタンの長さは変化する。ゆっくりとした長い｛京都｝は、サウンドスペクトログラムの時間軸方向に長いパタンとなる。また、時間伸縮は、単語パタン全長ばかりでなく、短いパタン、例えば｛ｋｙｏ｝や｛ｔｏ｝のような音節パタンや｛t｝や｛ｏ｝のような音素パタンにも局所的に現われる。一般に、｛ｏ｝のような母音類の時間伸縮幅は大きく、｛t｝のような子音類のそれは小さい。｛京都｝をゆっくり発話しても、長くなるのは

主に母音であり、子音長はあまり変化しない。このような不均一の伸縮構造は非線形時間伸縮構造と呼ばれる。

音節や音素等のパタンは、それぞれの類に固有の音響的特徴、即ち音節特徴や音素特徴を持っている。実際、サウンドスペクトログラム上の音素分節は、それぞれに固有の濃淡模様によって特徴づけられている。これらの分節の連結である単語等も、各単語固有の音響的特徴である単語特徴を持つ。

一方、音声信号は、これらの、いわゆる言語的特徴とは異なる種類の音響的特徴をも担っている。各発話者固有の音響的特徴、即ち話者特徴はその代表である。感情表現に密接に関係している韻律特徴もまた重要な例である。

音声認識においては、一般に、入力信号はサウンドスペクトログラムに類似の特徴表現で扱われる。こうした選択の背景には、聴知覚あるいは音声科学の長い研究成果がある。周波数軸に平行に、ある時点においてサウンドスペクトログラムを切ってみよう。得られる周波数－エネルギーの２次元信号である断面は、短時間スペクトルと呼ばれる。従って、入力全体はこの短時間スペクトルの系列ともみなされる。

さて、この短時間スペクトル系列であるスペクトログラムを用いて、音節や音素等の特徴をもう少し詳しく見てみよう。

短時間スペクトルは、基本的にスペクトル包絡と調波構造とから成る。スペクトル包絡は、短時間スペクトル上に大きくうねる山谷の構造であり、主に、ある時点における唇や舌、口蓋等の調音器官の形状によって決定される。図４．１中のホルマントと呼ばれるスペクトル包絡の山の部分は、この調音器官の形状に起因する音響的な共振現象に対応している。発話する時、調音状態を変えることによって異なる類の音素や音節を実現していることに気付くであろう。実際、音声認識にとって重要な言語特徴は、主にこのスペクトル包絡に含まれている。一方、図中の｛o｝の分節に見られるような微細なスペクトル形状である調波構造は、声帯（喉の奥にある）の振動から生ずるものであり、主に、話者特徴や韻律特徴を担っている。

以下に紹介する応用例で用いられている特徴表現法の中から代表的なものをまとめておこう。その第１は、１０ミリ秒から２０ミリ秒の時間窓（音声信号からある短い分節を切り出すために用いられる有限長の時間関数）を用いたディジタ

ルフーリエ変換によって計算する近似的なメルスケール対数パワースペクトルを短時間スペクトルとするものである。メルスケールとは、聴知覚研究の知見に基づく一種の非線形周波数軸であり、高い周波数帯域ではよりおおまかなスペクトル包絡形状を表現し、低い周波数帯域においてはより微細なスペクトル包絡形状を表現するものである。なお、メルスケールの代わりに、誕生の経緯は異なるものの、類似の非線形周波数軸を実現するバークスケールが用いられることもある。第2は、線形予測符号化（Linear Predictive Coding (LPC)）あるいはＰＡＲＣＯＲ法から得られるＬＰＣケプストラムベクトルを短時間スペクトルとして扱うものである。ＬＰＣケプストラムベクトルも、やはり２０ミリ秒程度の時間窓を用いて求められる。表面的には、この特徴ベクトルは、サウンドスペクトログラムに見られるようなスペクトル表現にはなっていない。しかし、表現様式が異なるものの、スペクトル包絡形状を決定する情報を効率的に含んでいる。

以下では、この短時間スペクトルやLPCケプストラムベクトルを、音響特徴ベクトルと総称しよう。入力全体は、時間窓を数ミリ秒から１０ミリ秒の一定間隔で移動することによって得られるこの音響特徴ベクトルの系列として表わされる。音響特徴表現の実際の計算法は、[31]や[82]に詳しい。

上記の解説からわかるように、単一の音響特徴ベクトルあるいはこれらから抽出される特徴は、ある時点における調音の状態を表わすという意味で、静的特徴と呼ばれる。これに対し、音響特徴ベクトルの（一般に５０ミリ秒程度の分節に相当する）系列あるいはこの系列から抽出される特徴は、対応する時間内の調音器官の動きを表現しているという意味で、動的特徴と呼ばれる。一般に、｛ａ｝、｛ｉ｝等の母音類は静的特徴によって特徴づけられる。また、｛ｋ｝、｛ｔ｝等の子音類は動的特徴によって特徴づけられる。

通常、音響特徴ベクトルの次元数（周波数軸方向の長さ）は特徴表現法毎に固定化される。このような固定次元のベクトルパタンは静的パタンと呼ばれる。一方、音声パタン全体の長さ、即ち、音響特徴ベクトル系列長は、発話音声そのものの長さによって決定されるために原理的に可変である。このような可変長パタンは動的パタンと呼ばれる。

さて、言葉は類似しているものの全く異なる概念、特徴の静的／動的特性とパタンの静的／動的特性とを紹介した。４．１．４には、さらに動的時間伸縮法と

いうことばも登場する。これらの類似した用語の使用は混乱を招きやすい。しかし、それぞれが既に音声信号処理の分野に定着している現状を考え、本章でもこれらの用語をそのまま用いることにしよう。

4.1.2 統計的パタン分類

4．1．2においては、現在の音声認識技術の根本原理の1つ、統計的パタン分類の基本を紹介する。数理的詳細よりは、むしろ考え方あるいは概念を明示することに焦点を合わせよう。詳しくは、[19]や[24]等を読まれることをお薦めする。

議論の枠組みとして、静的パタンの分類課題、即ち次元数 S の固定次元パタン $\mathbf{x}$ $(= (x_1, ..., x_s, ..., x_S))$ を M 類 $\{C_i; i = 1, ..., M\}$ に分類する課題を取り扱おう。パタンは、既に特徴表現され、さらに分節化されているものとする。また、$\mathbf{x}$ は連続量であるものとする。即ち、$\mathbf{x}$ は、無限の数のパタンから構成される、未知ではあるがある形を持った連続な標本分布から選択されたものであるとする。

課題を遂行するため、設計可能な（設計段階において調整可能な）パラメタ集合 Λ $(= \{\lambda_i\})$ から成る分類器とこれを設計するために用いられる N 個の設計用標本集合 $\{\mathbf{x}_n; n = 1, ..., N\}$ が与えられているものとする。ここで λ_i は C_i に割り付けられたパラメタである。λ_i は、一般にベクトルでもスカラーでも、あるいはそれらの集合でも良い。設計用標本は、それぞれ、属する類が明らかになっており、上記の連続標本分布から無作為に選ばれたものである。

統計的パタン分類は、パタンの生起を確率現象として扱う。$\mathbf{x}$ が生起する確率密度は $p(\mathbf{x})$ で表わされ、C_i が生起する確率、事前確率は $P(C_i)$ によって表わされる。$\mathbf{x}$ が与えられた時それが属する類を判断する分類は、原理的に $\mathbf{x}$ が C_i に属する確率、即ち事後確率 $p(C_j|\mathbf{x})$ を知ることによって実行される。分類の判断を下す規則には様々な可能性があるが、最も簡単な規則の1つ,

$$C(\mathbf{x}) = C_i, \quad \text{iff } i = \arg\max_j p(C_j|\mathbf{x}) \tag{4.1}$$

を考えよう。ここで、$C(\mathbf{x})$ は分類操作であり、$\mathbf{x}$ は最も大きな事後確率を示す類に分類される。これはベイズ決定則と呼ばれ、正確にこの規則を実行すれば、

理想的な、即ち、分類誤りを起こす確率が最も小さな最小分類誤り確率状態が得られることがわかっている。図4．2は、それぞれが単峰性の連続密度関数（太い実線）を成す、2類の1次元（$S=1$）標本を分類する課題を図解している。密度関数が交差する点に最小分類誤り確率状態に相当する理想的な境界がある。この境界位置以外に境界を移動するとき、誤り確率は必ず増加する。

ベイズ接近

Λ を用いて事後確率を正確に推定できれば、分類器は望ましい分類を実行できそうである。C_i のために Λ を用いて推定する推定事後確率を $p_\Lambda(C_i|\mathbf{x})$ と表わそう。Λ を設計／調整することによって、$p_\Lambda(C_i|\mathbf{x})$ を真の $p(C_i|\mathbf{x})$ に近づけたい。しかし、この事後確率を直接推定することは意外と難しい。

属する類がわかっている設計標本が与えられているわけであるから、推定条件確率密度 $p_\Lambda(\mathbf{x}|C_i)$ を得ること、言い替えれば、類標本の密度関数を推定することは比較的用意である。明らかに、C_i の設計標本に対して大きな値を示す、即ち最大尤度を示す推定条件確率密度が望ましい。このような推定を求める手法は最大尤度推定法（最尤法）と呼ばれる。最尤法によって $p_\Lambda(\mathbf{x}|C_i)$ が決まれば、以下のベイズの定理

$$p_\Lambda(C_i|\mathbf{x}) = \frac{p_\Lambda(\mathbf{x}|C_i)P(C_i)}{p_\Lambda(\mathbf{x})} \tag{4.2}$$

を用いて推定事後確率を求めることができる。事前確率はなんらかの先験的知識を用いて推定される。例えば、単語音声パタン分類においては、課題中の各単語類の生起頻度等から推定される。この意味で、(4.2) 中の $P(C_i)$ には推定に用いられるなんらかのパラメタが付記されるべきであるが、簡単のため省略している。

最尤法による条件確率密度の推定をもう少し詳しく見ておこう。課題は、上記の M 類分類課題である。まず、それぞれの類の条件確率密度関数をガウス型密度関数でモデル化することを考える。これは、最も簡単なモデル化の1つである。ここで、各 λ_i は、密度関数の平均ベクトル（S 次元）と共分散マトリックス（S 行 S 列）とである。設計標本に対する生起確率が大きくなるように、この平均ベクトルと共分散マトリックスとが調整される。最尤法による条件確率の

推定は類毎に行われる。調整を経て、各類の条件確率密度関数の推定精度は向上する。そして、ベイズの定理に基づいて類毎の推定事後確率が求められ、ベイズ決定則に従った分類が行われる。

このようにして、まず各類の条件確率密度関数を推定（モデル化）してから分類判断を行う流儀は、一般に、ベイズ接近と呼ばれる。図4．2には、推定された分布モデルの様子も破線で模式的に図解している。結果的な類境界は、2つの破線が交差する位置に求められる。

詳細は省略するが、最尤法の数理的基盤はかなり良く整備されている（ぜひ[24]等を見ていただきたい。）。こうした事情に助けられ、実際、このベイズ接近は分類器設計の本流として広く用いられてきた。しかし、この接近も決して万全ではない。極めて本質的な問題をも抱えている。

前述の例では、ガウス型密度関数が確率密度関数のモデル関数として用いられた。しかし、真の密度関数の形式は本質的に未知である。一般に複雑な形を持つ密度関数を表現するために、例えば複数のガウス型関数を組み合わせて用いる混合ガウス型関数を用いることも可能である。この場合、望ましい混合数を決定することがやはり問題として残される。

また、設計用標本は、いくら多くとも有限個である。仮定されたモデル関数が真の分布形式に一致している保証がない時、この有限の標本上で行われるモデルの推定は、通常（ほとんど必ず）、真の分布と推定モデルとの間の誤差を伴う。ここで、推定される条件確率密度関数の推定精度が標本分布全体の推定誤差に支配されること、従って、結果的類境界も分布全体の推定誤差に支配されることに注意しよう。図4．2には、推定の誤差も図解している。原理的に、ベイズ接近における推定では、真の類境界付近の誤差も類境界から遠く離れた領域における誤差も同様に扱われる。従って、この場合の推定誤差の減少は、必ずしも境界付近の密度関数の正確な推定、言い替えれば類境界の正確な推定を意味しない。極端な場合、類境界から遠く、分類精度にほとんど影響しない領域が正確に推定され、推定誤差が類境界付近に集中することさえあり得るのである。

判別関数接近

ベイズ接近に対するもう 1 つの選択として、判別関数接近がある。判別関数接近は、距離や類似度等の任意の測度から構成される判別関数によって類帰属度を表現する。計算が簡単なこともあって、電子計算機の能力が不十分であった時代にはかなり盛んに用いられた。

古典的な線形判別関数法を用いて、この判別関数接近の概要を紹介しよう。上記の M 類分類課題を考えている。$\mathbf{x}$ に対する C_i の判別関数 $g_i(\mathbf{x};\Lambda)$ は、

$$g_i(\mathbf{x};\Lambda) = \sum_s \lambda_i[s] \cdot \mathbf{y}[s] \tag{4.3}$$

のように、λ_i に関する線形関数（特に $\mathbf{x}$ に関する 1 次関数）によって定義される。ここで、$\mathbf{y}$ は $\mathbf{y} = (1, x_1, ..., x_S)$ である合成ベクトルであり、$\mathbf{y}[s]$ は $\mathbf{y}$ の第 s 要素である。これに伴って、λ_i は C_i のための $(S+1)$ 次元荷重ベクトルであり、$\lambda_i[s]$ は λ_i の第 s 要素である。λ_i は、ベイズ接近において標本分布を近似する確率モデルのパラメタとして扱われたように、判別関数接近においても C_i のある種のモデルとしての意味合いを持っている。この Λ から構成される分類器を線形判別関数分類器と呼ぼう。この時、用いられる分類規則は、

$$C(\mathbf{x}) = C_i, \quad \text{iff } i = \arg\max_j g_j(\mathbf{x};\Lambda) \tag{4.4}$$

である。

ベイズ接近において条件確率の最尤推定が行われたのに対応して、ここでは、各設計標本の分類結果に対する損失の最小化が目指される。判別関数接近における設計法は、損失と最小化法との選択によって特徴づけられる。この最小化に基づく判別関数の設計は、本質的に類に関して競合的に行われ、しばしば識別学習とも呼ばれる。損失は、システム（分類課題においては分類器）を設計する際の規準でもあるため、しばしば設計標的とも呼ばれる。損失はまた、決定理論を背景とする一般性の高い概念であり、システム判断結果に基づいてとられる行動等の犠牲／出費を反映するように設定される。特に、分類器設計においては、分類誤りを反映するように設定されることが自然であり、実際、そうした場合が多い。なお、このように設計標本に対する分類結果を設計に反映させようとする考

え方は教師付き学習とも呼ばれる。これに対し、ベイズ接近のように設計結果を帰還することのない設計戦略は教師無し学習とも呼ばれる。

C_k に属する $\mathbf{x}$ に対する損失を $\ell_k(\mathbf{x};\Lambda)$ と記述しよう。設計される分類器は、将来生じ得る全ての標本に対して良好に動作することが望まれる。従って、全標本に対する損失、即ち期待損失

$$L(\Lambda) = \sum_k P(C_k) \int \ell_k(\mathbf{x};\Lambda) p(\mathbf{x}|C_k) d\mathbf{x} \tag{4.5}$$

の最小化が設計における理想である。しかし、有限の設計標本のみの利用が可能な現実では、この最小化は明らかに不可能であり、与えられた設計標本上における経験的平均損失

$$L_0(\Lambda) = \frac{1}{N} \sum_k \sum_n \ell_k(\mathbf{x}_n;\Lambda) 1(\mathbf{x}_n \in C_k) \tag{4.6}$$

の最小化が目指される。$1(\cdot)$ は真の時に 1、偽の時に 0 をとる指示関数である。代表的な損失に、**パーセプトロン損失**や**自乗誤差損失**、**分類誤り数損失**（4．4．1の（4.24））等がある。損失が選択されると、自ずと経験的平均損失も決まる。

経験的平均損失の最小化は、また、最適な（経験的平均損失の最小状態に対応するという意味で）Λ の発見を目指す最適化ということもできる。最適化法としてここで利用可能なものに、緩和法（弛緩法）や勾配法、模擬徐冷 [33] 等がある。これらの手法は、元々、経験的平均損失のみならず多様な規準関数を用いることができるパラメタの最適化法である。しかし、最適化法と呼ばれるものの、その全てが真の最適化、即ち大域的最適化（規準関数の大域的最小化）を達成できるわけではない。局所的最適化（規準関数の局小化）のみが保証される場合もある。単なる規準関数の減少を企てるだけの場合もある。一般に、緩和法は、規準関数の最小化を保証するに十分な数理的基盤を持たない。この手法の有用性は限られている。最急勾配法に代表される勾配法は、規準関数の勾配方向を辿ることによって少なくともその局小点を発見できることが理論的に保証されている。これは、実用的な方法として広く用いられてきた。模擬徐冷は、現代的ニューラルネットワークの研究の中で注目されるようになった大域的最適解を探索する手法である。無限の調整手続きの繰り返しが規準関数の最小点を確率的にもたらすことが保証されている。

また、最適化法は、バッチ型と適応型との 2 種に大別されることもある。この類別も、判別関数接近に限らない一般的なものである。特に判別関数接近においては、バッチ型最適化法は、与えられた設計標本全体に対する経験的平均損失を陽に計算し、その損失の最小状態の実現を目指して Λ の調整を行う。上述の最急勾配法はこのバッチ型の典型的な例でもある。一方、適応型最適化法は、経験的平均損失を陽に計算することはせず、与えられた設計標本集合から（通常）1 標本を無作為に抽出し、それに対する損失が小さくなるように Λ の調整を行う。適応型の代表例には、確率近似法（例えば [23] 参照）に基づくもの、特に確率的降下法 [1, 2] がある。与えられた設計標本を用いる限り、この 2 種の手法の差異は小さい。いずれも、設計標本集合に対する最適化を目指している点で同じである。しかし、新たに標本が登場する毎に Λ を調整できる適応型機構は、設計段階では扱うことができなかった新しい利用状況に分類器を適応させることができるという著しい潜在的可能性を持っている。

このように数多くの損失や最適化法が提案されている中から、ニューラルネットワーク研究の歴史上重要な役割を果たしたパーセプトロン学習に起源を持つパーセプトロン損失を用い、バッチ型勾配法によって (4.3) の線形判別関数を設計する方法を紹介しよう。これは、一般化誤り訂正学習と呼ばれるものと原理的に等しい（例えば [97] 参照）。初めに、Λ は無作為に初期化される。この段階では、当然、正確な分類は保証されていない。設計標本集合から $\mathbf{x}_n$ を無作為に選び、分類を試みる。この $\mathbf{x}_n$ は C_k に属するものと仮定する。正しい分類が行われた場合、Λ の更新は行われない。誤って分類された場合、言い替えれば、$g_k(\mathbf{x}_n;\Lambda)$ よりも大きな値を示す判別関数が存在する場合、これらの判別関数に対応する全ての類（混同類と呼ぼう。）と C_k とに関与する Λ が更新される。更新は、C_k の判別関数をより大きくし、混同類の判別関数をより小さくすることによって、分類誤りを減少させることを目指す。

この更新手続きを数理的形式で表現してみよう。まず、設計標本 $\mathbf{x}_n$ に対する損失は、

$$\ell_k(\mathbf{x}_n;\Lambda) = \frac{1}{N_n}\sum_{i\in\Im}(-g_k(\mathbf{x}_n;\Lambda) + g_i(\mathbf{x}_n;\Lambda)) \tag{4.7}$$

と表わされる。ここで、$\Im$ は混同類の集合

$$\Im = \{C_j; g_j(\mathbf{x}_n; \Lambda) > g_k(\mathbf{x}_n; \Lambda)\} \tag{4.8}$$

であり、含まれる類の数を N_n とする。更新が誤分類に対してのみ行われるため、経験的平均損失は

$$L_p(\Lambda) = \sum_n \sum_k \ell_k(\mathbf{x}_n; \Lambda) \mathbf{1}(\mathbf{x}_n \in C_k) \mathbf{1}(\Im \neq \emptyset) \tag{4.9}$$

となる。勾配法に基づく Λ の更新は

$$\Lambda(t+1) = \Lambda(t) - \epsilon \nabla_\Lambda L_p(\Lambda) \tag{4.10}$$

である。ここで、t は更新繰り返しを示す時間指示子であり、ϵ は正の小さな定数である。$\Lambda(t)$ は t における Λ の状態を示している。結果的に、各類の荷重ベクトルは

$$\lambda_i(t+1) = \lambda_i(t) - \epsilon \sum_n \sum_k \frac{\partial \ell_k(\mathbf{x}_n; \Lambda)}{\partial \lambda_i} \mathbf{1}(\mathbf{x}_n \in C_k) \mathbf{1}(\Im \neq \emptyset) \tag{4.11}$$

に従って更新される。ここで、

$$\frac{\partial \ell_k(\mathbf{x}_n; \Lambda)}{\partial \lambda_i} = \begin{cases} -\mathbf{y}, & \text{for } i = k \\ \dfrac{1}{N_n}\mathbf{y}, & \text{for } i \neq k \\ 0, & \text{otherwise} \end{cases} \tag{4.12}$$

であり、$\mathbf{y}$ は $\mathbf{x}_n$ に対する合成ベクトルである。

判別関数接近は、ベイズ接近とは別の、分類決定則にかなり直接的な設計理念であることがわかる。以上の誤り訂正学習も決定則 (4.4) の実行を企てている。しかし、この誤り訂正学習は、簡単な線形分離可能状態と呼ばれる課題条件においてのみ正確な分類を達成できることが保証されており、現実的な利用には決して十分ではない。(4.9) の $L_p(\Lambda)$ の最小状態と最小分類誤り確率状態との関係も不明である。また、(4.7) における $\Im$ に限定される和操作あるいは (4.9) における $\mathbf{1}(\Im \neq \emptyset)$ のため、実は、勾配計算を厳密に $L_p(\Lambda)$ に適用することができない。ここで紹介した設計法は、判別関数接近における 1 つの実施

例に過ぎない。しかし、正確な分類を達成するための定形化が不十分であるという点に関しては、決して特殊な例外ではない。従来のほとんどの判別関数設計法が、実は類似の困難を抱えている。

さて、以上の紹介から、本書の主題であるニューラルネットワークによる分類器、特に**多層パーセプトロン** (Multi-Layer Perceptron (MLP)) 分類器の設計もこの判別関数接近に従っていることを類推できるであろう。Λ はネットワークの結合係数であり、判別関数は各類に割り当てられた出力節点の出力である。設計に用いられる逆誤差伝搬法は、損失（一般に自乗誤差損失）の最小化を図る勾配法の1種である。このニューラルネットワーク分類器は、実際極めて盛んに用いられている。しかしその一方で、実は、その設計最適性等の性質の解析が必ずしも十分に行われていない。この現代的判別関数接近法もまた真剣な改善を必要としている。

4．1．2で要約した判別関数接近全体の根本的再生を論じる話題が、4．4における主題である。

4.1.3 学習ベクトル量子化

元々、この学習法は、自己組織化特徴写像と呼ばれるニューラルネットワークのパタン分類力を向上させることを目指して提案されたものである [68, 70]。しかし、その動作原理は、**参照パタン**と入力パタンとの（自乗）ユークリッド距離を用いてパタン分類を行う古典的な距離分類器の参照パタンを、分類精度の向上を目指して適応的に調整する判別関数設計法に他ならない。なお、4．1．3では、参照パタンと入力パタンとは同じ次元数を持つベクトルであるものとしよう。

学習ベクトル量子化には、ＬＶＱ１ [68] に始まって、ＬＶＱ２ [69]、ＬＶＱ２．１ [71]、改良ＬＶＱ２ [22]、ＬＶＱ３ [70] 等、様々な版がある。ここでは、パタン分類の観点から、特に重要と思われるＬＶＱ２とその改良版 [85] とを解説する。

再び、前述の M 類分類課題を考えよう。分類器は、類毎に複数個準備された参照ベクトルから成る。C_i の参照ベクトルは Q_i 個の $\{\mathbf{r}_i^q\}_{q=1}^{q=Q_i}$ と表わされる。入力 $\mathbf{x}$ の C_i への帰属度は自乗ユークリッド距離 $E(\mathbf{x}, \mathbf{r}_i^{\gamma_i}) = \| \mathbf{x} - \mathbf{r}_i^{\gamma_i} \|^2$ によっ

て測られる。ここで $\mathbf{r}_i^{\gamma_i}$ は

$$\gamma_i = \arg\min_q E(\mathbf{x}, \mathbf{r}_i^q) \tag{4.13}$$

を満たす $\mathbf{x}$ に対する C_i の最近傍ベクトルである。従って、判別関数は

$$g_i(\mathbf{x}; \Lambda) = E(\mathbf{x}, \mathbf{r}_i^{\gamma_i}) \tag{4.14}$$

であり、用いられる分類規則は、

$$C(\mathbf{x}) = C_i \quad \text{iff } i = \arg\min_j g_j(\mathbf{x}; \Lambda) \tag{4.15}$$

である。入力ベクトルは、最も近い（距離が小さい）参照ベクトルを持つ類に分類される。

1つの類を複数の参照ベクトルで表現するこの距離分類器が、前述の線形判別関数分類器のある種の拡張であることに気付くであろう。実際、各類に1つの参照ベクトルのみを用い、さらにユークリッド距離を判別関数とする場合、この距離分類器は基本的に線形判別関数分類器に等しい。類毎に複数の参照ベクトルを用いることは、区間線形判別関数分類器を構成することを意味する。また、ユークリッド距離以外の距離、例えばガウス型確率密度関数に基づく尤度距離を用いることは、曲線（曲面）によって類境界を実現する判別関数を構成することを意味する。

複数参照ベクトル距離分類器と同様に、(4.3) に基づく線形判別関数分類器を区間線形判別関数分類器に拡張することは容易である。この意味でも、距離分類器と線形判別関数分類器とは互いに深い関連を持っている。しかし、参照パタンは、重みとしての役割を持つ線形判別関数よりやや明確なモデルとしての意味合いを持っている。類は、参照パタンによってその代表的な性質あるいは類境界を示す性質を表現されているということができる。そこで、参照パタンは、しばしば典型とか模範とかを意味するプロトタイプとも呼ばれる。

ＬＶＱは、1つの設計用標本が与えられる毎に参照ベクトルをわずかづつ更新し、標本を取り替えながらこの微小量の更新を繰り返す。即ち、ＬＶＱは適応型設計法である。この性質は、ＬＶＱ１からＬＶＱ３まで、全ての版に共通している。

さて、適応的な設計段階のある時点 t において、設計用標本集合から無作為に選ばれた C_k に属する $\mathbf{x}_n$ が C_i に分類されたものと仮定しよう。ＬＶＱ２における参照ベクトルの調整は、以下の３条件

$$i = \arg\min_j g_j(\mathbf{x}_n; \Lambda) \quad \text{and} \quad i \neq k \tag{4.16}$$

$$k = \arg\min_{j, j \neq i} g_j(\mathbf{x}_n; \Lambda) \tag{4.17}$$

$$\rho < \frac{g_i(\mathbf{x}_n; \Lambda)}{g_k(\mathbf{x}_n; \Lambda)} < 1 \tag{4.18}$$

が満たされる時のみ、

$$\mathbf{r}_j^q(t+1) = \begin{cases} \mathbf{r}_j^q(t) + \epsilon(t)(\mathbf{x}_n - \mathbf{r}_j^q(t)), & \text{for } q = C_j \text{ and } j = k \\ \mathbf{r}_j^q(t) - \epsilon(t)(\mathbf{x}_n - \mathbf{r}_j^q(t)), & \text{for } q = C_j \text{ and } j = i \\ \mathbf{r}_j^q(t), & \text{otherwise} \end{cases} \tag{4.19}$$

のように行われる。ここで、$\epsilon(t)$ は単調減少の正の小さな数である。即ち、$\mathbf{x}_n$ が誤分類され、正しい類 C_k が２番目に確からしく、しかも $\mathbf{x}_n$ が更新直前の Λ によって決定される類境界付近にあるという条件が満たされる時にのみ、参照ベクトルは、入力本来の類の判別関数をより小さくし、誤って分類された類の判別関数をより大きくすることによって、正しい分類を導くべく更新されるのである。

ここで、(4.19) から、ＬＶＱにおける調整の底流に、勾配法の理念があることを想像できる。実際、判別関数 (4.14) の微分形式が調整ベクトルを構成している。

ＬＶＱ２において、第３番目の条件は特徴的である。ＬＶＱ２の調整は類境界付近でのみ行われる。類境界付近に焦点を合わせて誤分類の減少を追及するこの設計が最小分類誤り確率状態の良い近似を達成できるであろうという直観は自然である。実際、文献に示されている実験結果は、この直観を支持しているように思われる（例えば [70] 参照）。

ＬＶＱ２における第２の要請、即ち、C_k が第２最近傍類である誤分類において更新が行われるべきであるという条件には、あまり必然性はないように思われる。C_k が何番目に尤もらしいといっても誤りに変わりはない。この点に注目し

て、 [85] の改良ＬＶＱ２は、 (4.16) と (4.18) との２条件のみを調整のために要請する。

この改良版の場合も含め、多くの改良は、発見的あるいは直観的に行われ、その有効性の実証は実験に委ねられてきた。真に妥当性を議論するための数理的な定形化の基盤は未知のままであった。上に指摘したように、ＬＶＱの設計原理が勾配法にあることを伺うことができる。しかし、どのような損失関数を最小化の対象にしているのか、適応型の最小化の繰り返しがどのような収束をもたらすのか等、不明の点が多く残されてきたのである。

4.1.4 音声パタン分類のための基本技術

静的パタンのために紹介した統計的パタン分類手法は、基本的に、動的音声パタンの分類にも用いることができる。しかし、パタンが動的であるが故に、音声パタン分類には、これに特有の手法を組み入れる必要がある。以下では、音声パタン分類固有の基本技術を解説する。

分節化と非線形時間伸縮を伴う音声パタン分類

４．１．１において述べたように、音声パタンは、短時間スペクトルの動的な系列として表わされる。この音声パタンを分類することを考えよう。分類過程の原理は、４．１．２で紹介したものと同じであるが、実際の手続きはもっと複雑である。４．１．２の議論では入力パタン x、即ち分類の対象そのものが与えられていたことを思い出そう。ところが、実際の音声は、連続した音響信号の１部として登場する。従って、分類器は、その背景音から分類対象である音声パタンを分節化しながら分類を実行する必要がある。

一言だけ {京都。} というように孤立発話された音声パタンを分類したい。この場合、分類器は、類である単語毎に参照パタンのような認識器パラメタを用意し、背景音（この場合、通常、音声ではないという意味で雑音）から単語分節を切り出しながら、その分類を行う。

ところで、日常用いられる辞書には数万の単語が掲載されている。頻繁に用いられている単語だけでも５０００語を超える。これらの単語を組み合わせてできる文の数は、数十万あるいは数百万に達する。連続発話された連続単語、即ち

文パタンを分類する場合、{京都からATRまで、時間はどのくらいかかりますか。}や{京都からATRまで、料金はどのくらいかかりますか。}等の文それぞれを独立した類として類モデルλ_iを用意することは、その総数があまりにも膨大であることにより明らかに非現実的である。そこで、一般に、音素や音節等の、文の構成要素である短い分節毎にλ_iが用意される。この時、文モデルは、音素等の短い分節モデルを連結することによって構成される。分類器は、音素等の分節モデルに基づく、連続音声の分節化とその分節の分類とを繰り返すことによって、文全体の分類を行う。

以上からわかるように、分節化は、音声の分類では避けることができない操作である。しかし、音声パタンの始端・終端は一般に不明瞭であり、背景雑音からの単語音声の分節化ですら決して容易ではない。また、唇や舌等の調音器官が滑らかに動くことから想像されるように、連結する音素の音響的特徴は互いに影響し合い（この現象は調音結合と呼ばれる。）、連続音声を音素等に分節化することはさらに困難である。従って、音声パタンの分類器は、このようなパタンの分節化における時間位置ずれに強い（しばしば高い耐性を持つと呼ばれる。）ものであることが望まれる。

動的である音声パタンの時間伸縮が一般に非線形であることを思い出そう。即ち、単語パタン全長が2倍になっても、パタンを構成する個々の音素長が全て2倍になるわけではない。通常、母音の伸縮幅は大きく、破裂音{p、t、k、b、d、g}等の子音の伸縮幅は小さい。このような伸縮幅の相違の存在は、知覚的にも確認されている[57]。例えば音響特徴ベクトル系列長が50の入力パタンと系列長が45の参照パタンとの距離を測ることを考えよう。長さが異なるパタン間の距離計算は、同次元のベクトル間の距離を求めるようにはいかない。例えば音響特徴ベクトルの間引きや隣接する音響特徴ベクトルの平均化のような操作によって、長い方のパタンを系列長が45になるように加工することはできる。しかしこのような安易な操作は、前述したように複雑な時間伸縮構造を分類判断過程に反映するには明らかに不十分である。パタン中の音素等の分節が個々に持つ特徴が無視される危険が大きい。音声パタン分類器は、この複雑な非線形時間伸縮に適切に対処できる能力を持つ必要がある。

以上見てきたように、音声パタン分類器は、本質的に、分節化と非線形時間伸

縮とを行う優れた能力を伴う必要がある。実際、この要請に応えることが、音声パタン分類技術の研究開発における最重要課題の1つであり続けてきた。これまでの解決法は、主に以下の2種に大別される。その第1は、参照パタン型音声モデルを用いる距離分類器のパタン間距離計算に、動的計画法に基づく最適整合経路探索を組み込むものである [106]。この手法は、しばしば動的時間伸縮法と呼ばれる。第2は、**隠れマルコフモデル** (Hidden Markov Model (HMM)) を音声パタンモデルとするものである [52]。この2つの接近法については、節を改めてやや詳しく解説しよう。

動的時間伸縮を用いる距離分類器

入力音声パタンと同じ表現様式を持つ動的な参照パタンから成る距離分類器を用いる音声パタン分類法を紹介する。一般に長さが異なる入力パタンと参照パタンとの間の距離計算のために、動的計画法による最適整合経路の探索、即ち動的時間伸縮が行われ、この2つのパタン間の距離は、この経路上で計算される累積距離によって表わされる。

図4．3は、この距離計算の手続きを図解している。整合経路とは、入力パタンのある音響特徴ベクトルと参照パタンのある音響特徴ベクトルとを対応づける、即ち整合させる点の連結である。2つのパタン間の対応づけられた音響特徴ベクトル間の距離が局所距離として測られる。経路は、全体的な経路制限と局所的な経路制限との条件の下で原理的に左（過去）から右（現在）に伸ばされる。図中、斜線部は全体的経路制限の例を示しており、矢印は局所的経路制限の例を示している。最適経路探索の原理を理解するために、整合点Aに注目しよう。Aへ到達できる経路は、矢印に示された3つの整合点B、C、Dからのみ伸ばすことができる。B、C、Dのそれぞれは、そこへ至るまでに累積されてきた累積局所距離を伴っている。Aに向かって移動できる3点の中で、最も累積局所距離が小さな点を選択する。明らかに、点Aで考え得る最も累積距離が小さな経路が伸ばされることになる。左端の出発点から伸ばされてきた全ての可能な整合点においてこの選択を繰り返す。結局、最も小さな累積局所距離を伴う整合経路が右端に到達する。こうして得られる最小累積局所距離が、比較しているパタン間の距離となる。

上記の動的時間伸縮を除けば、参照パタンを用いる音声パタン距離分類器の分類戦略は、静的パタンのためのそれと同じである。動的時間伸縮の過程は判別関数の中に組み込まれ、各類を代表する参照パタンに対する最適（最小）整合経路距離が判別関数として用いられる。4．1．2及び4．1．3と同様の形式で、M 類の動的パタン x を分類する課題のための判別関数の定義を与えておこう。λ_i として Q_i 個の動的な参照パタン $\{r_i^q\}_{q=1}^{q=Q_i}$ が与えられているものとする。この時、C_i の判別関数は

$$g_i(x;\Lambda) = \min_q \left[\min_\theta D_\theta(x, r_i^q) \right] \tag{4.20}$$

である。ここで、$D_\theta(x, r_i^q)$ は x と r_i^q との間の可能な Φ_i^q 個の内の第 θ 最小累積局所距離（ある動的時間伸縮経路に対応している。）である。用いられる分類規則は（4.15）である。

参照パタンは、一般に、ベクトル量子化（Vector Quantization (VQ)）[37] の原理に従って設計される。ＶＱとは、参照パタンと全ての設計標本との間の距離の和、即ち歪を規準関数として、その最小状態に対応する参照パタンの探索を行う最小歪設計法である。ここで、以前に紹介した尤度距離を思い出そう。尤度距離に関する最小歪と対応するガウス型確率関数に関する最大尤度とは実は等価である。従って、このＶＱ設計法は、理念的にベイズ接近の１つであると考えることができる。ＶＱには、ＬＢＧアルゴリズム [79] やｋ平均法（例えば [81] 参照）等、様々な実現法（設計規則）が存在する。しかし、いずれも静的パタンを対象にしたものである。これらの手法は、音響特徴ベクトルの空間を少数の代表ベクトルによって表現するためには用いることができるが（次のＨＭＭ分類器の節参照。）、音声の動的な参照パタンそのものを設計するためには用いることができない。そこで、この目的のためには、やや最適性等の保証に関して問題はあるが、改良ｋ平均法等の動的パタンのための量子化法が用いられる [126]。

これまで述べてきた距離分類器や動的時間伸縮の考え方は、分類対象である単語や音素の類と同じ分節構造の参照パタンを分類器が持っていることを前提にしている。しかし、文や連続単語を分類するための分類器は、音素等の連結を伴うもっと複雑な動的時間伸縮過程を持つ。この点に関しては、次節のＨＭＭ分類器を用いる場合でも同様である。詳細は、[11, 91, 95, 103, 107] 等を参照してい

ただきたい。

HMM分類器

HMMとは、簡単に言えば、観測できない内部状態間の遷移に伴って観測可能な出力を確率的に生起するというシステムである。観測できない部分があるため、通常の設計手法を用いることができない。しかし、期待値 - 最大化法（Expectation-Maximization (EM) Method）と呼ばれる１種の最尤法が開発され、それ以来、欠損がある観測データのモデル推定等に広く用いられてきた。

HMM による音声パタンのモデル化の原理を説明しよう。図４．４は、特に音声パタンのモデルに適しているといわれている左 - 右型の HMM を図解している。モデルは、４状態から成り、実際に観測することはできないが、矢印で示された状態内における帰還と状態間の遷移とが遷移確率に従って確率的に起こっているものとする。音響特徴ベクトルが観測されるモデル出力である。この出力は出力確率に基づく。本章では特に、この出力確率は状態に割り付けられているものと考えよう（状態ではなく遷移に割り付けられるモデル化もある。）。結局、音声パタンは、この状態遷移と状態からの出力との繰り返しによって確率的に生成されるモデル出力の系列であると見なすことができ、そのモデル化とは、この状態遷移確率と、出力確率、そして状態の初期条件を支配する初期確率を決定することである。

HMM音声パタンモデルの出力は、原理的に、全ての可能な状態遷移の組み合わせ（経路）から生ずる。従って、ある音声パタン標本に対するモデル尤度はこの全ての遷移に基づいて求められる。この原理に忠実な計算法はトレリス法と呼ばれる。簡単のために、可能な遷移経路の中で最大尤度を示す経路のみを用いる計算法はヴィタビ法と呼ばれる。この最大尤度経路の選択には、動的計画法が用いられる。

HMM分類器は、線形判別関数の荷重ベクトルや距離分類器の参照パタンのように、類毎に少なくとも１つのHMMから構成される。用いられる分類規則は（4.1）である。分類器の設計は、通常、類毎に与えられた設計用音声パタン集合上でＥM法を実行することによって行われる。各類の設計用パタンを最大の尤もらしさで出力する類モデルが求められる。この時得られる確率は、それぞれの

類の条件確率である。このようなHMM分類器の設計理念は、ベイズ接近のそれに等しい。なお、EM法の実行には、計算効率の高い前向き－後ろ向きアルゴリズムが用いられる。

音声パタンと同じ表現形式を持つ参照パタンによるモデル化と比べ、HMMによるモデル化はやや理解が困難かもしれない。その主な原因は、HMMが確率モデルであるところにあるかもしれない。この点に注意しながら、HMMによる音声パタンモデル化をもう少し詳しく見て行くことにしよう。

さてHMMには、主に2種の型が存在する。1つは離散型と呼ばれるものであり、もう1つは連続型と呼ばれるものである。

離散型HMMとは、音響特徴ベクトルを離散的な符号に置き換えることによって音声パタンを予め符号系列に変換し、この符号系列をモデル化するものである。図4．5は基本的な離散型HMMの構成を図解している。符号帳とは、音声パタンを符号系列に変換するために用いられる符号の集まりである。各符号は、音響特徴ベクトルと同じ表現形式を持つ静的な符号ベクトルと連合されている。通常、この符号ベクトルは、設計用音声パタンが持つ音響特徴ベクトル全体に対する歪が最小になるように設計される。前述したVQによる最小歪設計法がここで用いられる。入力は、それを構成している各音響特徴ベクトルを最近傍の符号ベクトルに対応している符号に置き換えることによって、符号系列に変換される。もし、符号が{a，b}の2つしかなければ、{a}と{b}とは、2項分布からの出力と考えることができる。同様に、一般に数100個からなる符号帳を用いるHMMは、各符号を離散的な多項分布からの出力として扱う。HMMは、この多項分布からの出力確率と状態間の遷移確率、そして各状態の初期確率とからなる。

一方、連続型HMMは、符号変換をせずに、音声パタンを直接モデル化する。図4．6はこの連続型モデルを図解している。各状態には、ガウス型あるいは混合ガウス型等の連続型確率密度関数が割り当てられている。音声パタンの音響特徴ベクトルは、この連続型確率密度関数によって決定される確率分布からの出力として扱われる。HMMは、この連続型出力確率と、状態遷移確率、状態の初期確率とから成る。

特に、確率密度関数としてガウス型関数を用いる場合、出力確率の設計はこの

関数の平均ベクトルと共分散行列とを設計することによって行われる。この平均ベクトルを参照パタンの音響特徴ベクトルに対応させ、さらにガウス型確率関数に基づく尤度距離（実はユークリッド距離もこの1種である。）によって音響特徴ベクトル間の距離を測ることを考えると、動的時間伸縮を伴う距離分類器とヴィタビ法HMM分類器との類似性に容易に気付くことができる [88]。なお、静的パタンのための距離分類器と尤度ネットワーク [60] との近縁関係も同様に明らかである。この関係に関する厳密な議論は [54] にある。

判別関数接近を紹介した時にふれたように、ベイズ接近は高い分類精度を達成するためには必ずしも十分ではない。EM法によって設計されるHMM分類器も、この不十分さを患っている。そこで、分類精度の向上を目指し、相互情報量最大化法 [6] や誤り訂正法 [7] 等の判別関数接近による HMM 分類器設計も数多く調査されている。特に、誤り訂正法は、4．1．2の判別関数接近の節において紹介したパーセプトロン損失を用いる判別関数接近の経験的実装法と見ることができるものである。

HMMは、現在の音声認識技術を代表する分類器構造である。様々な立場から、実に多くの研究が行われている。この全体像を知り、さらに詳細を学ぶためには、 [43, 91, 103] 等が有益である。

動的特徴を表現する時間遅れ

音素等の音声パタンの類は、静的特徴ばかりか動的特徴によっても特徴づけられることを述べた。動的特徴を表現するためには、原理的に、ある時間長の分節内でモデル化を行う必要がある。従って分類器は、ある時点における静的な音響特徴ベクトルを用いるばかりではなく、分節内における動的特徴をも用いて、言い替えれば時間遅れを伴った特徴の表現をも行って分類判断をすることが望まれる。

動的特徴を表現する最も簡単な手法は、音響特徴ベクトルそのものに時間遅れ構造をもたせることである。即ち、隣接する複数の音響特徴ベクトルを連結し、その連結されたベクトルを新たに音響特徴ベクトルとして扱うものである。これは明らかに動的特徴を担っている。しかし、ベクトルの次元数の増加に伴って、計算効率が悪くなる傾向もある。

効率の良さを考えたより進んだ動的特徴の表現手法は、隣接した音響特徴ベクトルの変化方向を表わす比較的低次元のベクトルを新たに生成するものである。特に、ケプストラムから成る音響特徴ベクトルから求められる動的ケプストラムは広く用いられている。ここでは、連続するケプストラム係数の変化方向、つまり傾きが動的ケプストラムの値となる。

動的特徴の詳しい解説は、やはり [31] 等に見ることができる。

4.1.5 言語処理

最近の統計的な音声パタン分類における言語処理とは、単語連結等に関する言語的制約、即ち言語モデルを用いて、単語あるいは文等の類の事前確率を推定する過程であるということができる。4．1の最後では、この言語処理過程の概要を解説する。

例として、連続して発話される１０桁の数字の音声パタンを分類する問題を考えよう。語彙は、{ゼロ}、{イチ}、{ニ}、{サン}、{ヨン}、{ゴ}、{ロク}、{ナナ}、{ハチ}、{キュウ}の１０数字単語から成る。分類器は、数字類毎にHMMパタンモデルを持つHMM分類器である。１０桁の連続数字入力は、この１０種の数字モデルを連結することによってモデル化され、最も尤度の大きな連結（連続数字）モデルの類に分類される。

まず、数字の連結に制約が無いものとする。この時、各桁には１０種の数字が等しい事前確率に従って現われることになる。全ての可能な数字連鎖、１０の１０乗個の類が生起し得る。また、各類の事前確率は等しい。従って、分類器は、音響現象に関するHMMパタンモデルから得られる各類に対する条件確率のみを用いて、この膨大な規模の分類判断を行わなければならない。

この１０桁の数字が電話番号であるものとしよう。数字の連結には明らかに強い制約が加わる。１桁目の数字としては{ゼロ}のみが可能である。この明確な制約は、可能な類（分類時に検証すべき連結単語類）の数を１０分の１に減少させる。

さらに、特定の利用者のための電話番号分類課題であるものとしよう。この利用者の郷里は福島県である。頻繁にこの郷里へ電話するため、{ゼロ・ニ・ヨン・ゴ}で始まる数字連鎖が現われる事前確率が大きい。この場合の制約は非決

定的、即ち確率的である。発話が必ず {ゼロ・ニ・ヨン・ゴ} で始まるとは限らない。しかし、このような事前確率は、HMMパタンモデルによる条件確率とともに用いられ、より正確な事後確率の推定、言い替えれば、より正確な分類判断を実現する。また、先行する単語系列を条件とする条件確率を用いて、後続する数字音声の出現に制約を与えること（事前確率を推定すること）もできる。この言語的制約と各数字類に関する音響モデルに基づく条件確率とから、後続する数字候補を絞り込み、検証すべき類の数を著しく減少させることも可能となる。前段落の類候補の削減は、この極端な決定論的な例である。

このように、音声パタン分類、特に連続音声の分類における言語処理の効果は大きい。実際、様々な言語モデルの利用が調査されている。特に最近では、正しい類を検証すべき候補群から誤って外してしまう危険が大きい決定論的な言語モデルよりも、確率的言語モデル、例えば、確率文脈自由文法（例えば [76]。）や確率的ＬＲ構文解析器 [127]、N 組確率 [53] 等が注目されている。

ここでは特に、後述する応用にも登場する N 組確率についてもう少し解説しておこう。もう１度、電話番号の分類課題を考えよう。この分類課題における事前確率は１０桁から成る数字連鎖の類の生起確率である。従って、この生起確率を求めることが１つの理想的接近である。しかし、前述したように、１０の１０乗個もある全ての類の生起確率を、統計的に満足できる安定性で得ることはほとんど不可能である。一般に確率の推定には、関連する類の十分多くの標本が必要とされる。そこで、数字毎に音響的モデルを構成するのにも似て、ここでは、１０桁の連鎖に対する確率を、比較的短い連鎖、即ち $(N-1)$ 個の数字の組の生起を条件として N 番目に生起する数字に対する確率（N 組確率）で代用しようとするわけである。しかし、N が大きくては、上記の問題は解決されない。そこで通常は、１つ組 $(N=1)$ や２つ組 $(N=2)$、３つ組 $(N=3)$、さらにこれらの組み合わせが用いられる。例えば、福島へ電話を頻繁にかける利用者の課題においては、{ゼロ・ニ} や {ヨン・ゴ} の２つ組の生起確率は大きくなり、これらの言語モデルに基づく事前確率は、音響的モデルに基づく分類判断の変更や検証すべき後続数字候補の予測／絞り込みに貢献する。

4.2 音声パタン分類への応用

4．2では最も精力的に研究されてきた音声パタン分類への応用例を紹介しよう。応用の幅あるいは深さは多様である。そこで、まず、4．2．1では、分類器の基本構造や測度の選択に深く関与する、ニューラルネットワーク分類器の設計理念に関する話題を紹介する。4．2．2では、特に、4．2．1で紹介する分類器のいくつかが不特定話者音声認識課題に応用された例を要約する。最後に4．2．3では、言語処理への応用例を紹介する。

4.2.1 ニューラルネットワーク分類器の設計

ニューラルネットワークの音声パタン分類への応用は、様々なネットワーク構造、例えばボルツマンマシンや3層のMLPの上で、この新しい技術への関心が高まった直後から試みられている [20, 41, 102]。しかし、初期の応用のほとんどは、音声信号のパタンとしての動的特性も特徴の動的特性も無視した初歩的なものであった。十分なパタン分類能力を達成するためには、ニューラルネットワークを用いる音声パタン分類器の構造や設計方針（例えば、ベイズ接近か判別関数接近か）に関する進んだ工夫が必要であった。

時間遅れ構造

音声パタン分類のためのニューラルネットワーク設計の本格的な研究は、時間遅れ構造を組み入れることによって始められたといって良い（例えば [120]）。

中でも、**時間遅れニューラルネットワーク**（Time Delay Neural Network (TDNN)）の開発は、ニューラルネットワークによる音声認識に対する最近の大きな関心を引き起こしたという意味で、歴史的に重要な役割を持っている [122, 124]。図4．7は、ＴＤＮＮの典型的な構造を図解している。この構造の特徴は、時間窓内の隣接する音響特徴ベクトルを連結することによって合成される連結音響特徴ベクトルを用い、しかも隠れ層においても時間遅れ構造を持っているという点にある。また、分類過程においてこの時間窓による入力上の走査が行われることも特徴である。時間窓の移動に伴い、隣接する複数の連結音響特徴ベクトルが分類判断に供される。ネットワークの訓練は、通常、自乗誤差損失と誤差逆伝搬法と

を用いて行われる。訓練された分類器は、対応する出力接点からの出力が最大の類に入力パタンを分類する。この分類判断は、Λをネットワークの結合係数として（4.4）を用いる典型的な判別関数接近のそれである。表4．1は、このＴＤＮＮ分類器を用いた音素分類結果を示している。比較のために評価された離散型HMM分類器より、明らかに優れた分類精度を達成できることが示されている。

ＴＤＮＮにおける時間窓走査の考え方は、ＬＶＱによって設計される距離分類器ネットワークにも活用され、時間ずれに強い（高耐性の）ＬＶＱ（Shift-Tolerant LVQ (SLVQ)）ネットワークが開発された [84, 87]。図4．8はこのSLVQネットワークの典型的な構造を図解している。このネットワークにおいて、1つの中間層節点に結合された下層結合は参照パタンに他ならない。距離分類器に関する解説を思い出そう。参照パタンは、予め、課題中の類の1つに割り当てられている。通常、各類には複数の参照パタンが準備される。入力は、時間窓の使用によって、参照パタンと同次元の連結音響特徴ベクトルに変換される。参照パタンは、この連結音響特徴ベクトルとの間で自乗ユークリッド距離によって比較される。ＬＶＱは、これらの参照パタンを前述の規則に従って調整する。ＭＬＰと異なり、ここでは、上層結合は設計訓練の対象ではない。これらの結合係数は全て定数であり、上層（出力層）節点の入力関数には最小値探索操作と荷重平均操作とが連合されている。これらの操作は、時間窓の移動とともに得られる複数の参照パタン距離を平均して各類の判別関数を計算するために用いられる。結局、このネットワークは、入力音声パタンを最も判別関数の小さな類に分類する。表4．2は、このＳＬＶＱ距離分類器の分類精度の例を示している。分類器が基本的に区間線形判別関数として働くため、一般に分類器のサイズは大きくなる。しかし、極めて簡単かつ高速な設計によって、複雑なＴＤＮＮとほぼ同等の分類精度を達成できることがわかる。

混成構造

ＴＤＮＮとＳＬＶＱネットワークとにおける分類器構造に関する工夫は、ニューラルネットワークの構造的枠組みを音声パタンの時間構造を扱うために一歩広げるという重要な役割を果たした。実際、それは、時間位置ずれに対する耐性を

明らかに改善し、また、可変長であるというパタンの動的特性に起因する問題の軽減をも実現している。しかしそれは、単語等の長い音声パタンが持つ非線形時間伸縮構造を適切に扱うには不十分である。この限界を解決するためには、思いきった分類器構造の変更が必要である。この混成構造の節では、その構造の変更を、ニューラルネットワークと他のシステムとを組み合わせた混成（ハイブリッド）構造を用いることによって実現した例を紹介する。

ＴＤＮＮが開発されるとほとんど同時に、動的ニューラルネットワーク（Dynamic Neural Network (DNN)）が提案されている [108, 109]。図４．９は、そのネットワーク構造を図解している。このネットワークは、動的時間伸縮によって動的な入力音声パタンの時間長正規化を行うことによって先ず静的パタンを生成し、この静的パタンに対する分類器を設計しようとするものである。設計は、原理的に自乗誤差損失を用いる誤差逆伝搬法に基づいている。

単語等の長い分節パタンを分類するために、ＴＤＮＮやＳＬＶＱも、ＤＴＷとの混成様式として用いられる。即ち、ＤＴＷは、音素等の短い分節用に設計されたネットワークの分類結果を非線形時間伸縮を許して連結するために用いられる。この手法による大語彙単語音声分類の結果は [90] に紹介されている。しかし、確かにＤＴＷが用いられてはいるものの、ＴＤＮＮやＳＬＶＱのための混成様式はＤＮＮの混成様式とは本質的に異なっている。ＤＮＮにおいては、ネットワーク設計の中にＤＴＷ過程が組み込まれている。ＤＮＮ設計の最適性にＤＴＷは深く関与しており、一方、ＴＤＮＮやＳＬＶＱにおけるＤＴＷの利用はネットワーク設計と基本的に独立である。

ＨＭＭとニューラルネットワークとの混成様式の研究の初期のものとして、離散型ＨＭＭの高分類能力を持つ符号帳をＬＶＱによって設計する手法が開発された。ＨＭＭによる非線形時間伸縮機能をそのまま持つこの混成法は、明らかに、上記のＴＤＮＮやＳＬＶＱよりも進んだ時間構造の取り扱いを可能にしている。 [48] においては、この手法はＬＶＱ－ＨＭＭ混成法と呼ばれ、音素分類や比較的平易な単語認識課題においてその高効率性と高識別力とが確かめられている。

しかし、ＬＶＱによる分類能力向上の効果が音響特徴ベクトル空間、言い替えれば、符号ベクトルのみに限定されているという制約のため、この混成手法の貢

献は、分類器全体に必ずしも及ばない。長い文パタンの認識等においては必ずしも十分な分類精度向上を実現しないとの報告もある [49, 67, 128]。

ＤＮＮにおけるＤＴＷの利用に類似の理念に基づいてHMMを用いるいくつかの混成様式も調査されている。この場合、動的な入力音声パタンは、まずHMMによって静的パタンに変換され、その静的パタンを分類するための分類器がニューラルネットワークを用いて設計される。

[40] においては、HMMの状態遷移情報がＭＬＰの入力として用いられている。

[59] においては、ＨＭＭ／ＬＶＱ法が提案されている。図４．１０はその構造を図解している。動的音声パタンの音響特徴ベクトルは、HMMのヴィタビ法分節化によってHMMの各状態に割り付けられる。状態毎に割り付けられた音響特徴ベクトルは平均化され、この平均ベクトルが連結されることによって新しい静的なベクトルが作られる。この静的ベクトルをＬＶＱによって設計される距離分類器が分類する。この手法は、全ての類に準備されたHMMによる分節結果を用いることや状態毎に平均化された平均的音響特徴ベクトルを用いること等によって、分類判断における精度と未知標本に対する耐性（統計的推定の安定性）との向上を同時に実現することに成功している。表４．３は、米語Ｅ集合と呼ばれる９類｛ｂ、ｃ、ｄ、ｅ、ｇ、ｐ、ｔ、ｖ、ｚ｝の音声認識課題における結果を要約している。HMM／ＬＶＱ認識器は、最尤推定による慣例的HMM認識器の結果に対する明確な改良を達成している。連続音声の分類を目指した試みも始められている [105]。

なお、HMM／ＬＶＱ法に類似の手法は [32] や [42] においても調査されている。

HMM／ＬＶＱ法におけるHMMの役割は、動的な入力パタンを予め定められた数の分節に分節化することである。[18] においては、レベルビルディング法と呼ばれる動的時間伸縮法を用いた最小歪分節化（Minimum Distortion Segmentation (MDS)）とＬＶＱとの混成手法、ＭＤＳ／ＬＶＱ法が提案され、HMM／ＬＶＱ法と同等の分類精度を達成しつつ、大幅な計算量の削減が可能であることが示されている。

HMMの実現

混成構造の節で述べた混成様式は、ニューラルネットワークとDTW距離分類器あるいはHMMとを単に組み合わせるものである。一方、このような単なる組み合わせではなく、HMMの機能をニューラルネットワークによって実現することによって慣例的な（主にベイズ接近流に設計される）HMMの分類能力を向上させようとする手法も数多く調査されている。これらは、後にも述べるように、設計理念に関する混成様式であると見なすこともできる。

HMMの機能をニューラルネットワークによって実現する手法としては、まず、HMMの非線形時間伸縮能力に着目したものがある。ヴィタビ法をニューラルネットワークで実現することを目指して、ヴィタビネットが提案されている [80]。また、隠れ状態を遷移しながら出力を生成するHMMをEM法によって設計する問題を、MLP型のネットワークの設計問題に置き換えた隠れ制御ニューラルネットワークも調査されている [77]。

また、HMMとニューラルネットワークとを比較する時、この両者の間には、構造的な相違の他に設計理念の大きな相違があることに気付くことができる。HMM は、通常EM法、言い替えれば、ベイズ接近の理念に従って設計される。これに対し、LVQ分類器も含め、分類器として用いられるニューラルネットワークのほとんどは、判別関数接近に基づく設計法によって設計されている。この設計理念の相違に着目して、HMM分類器における各類の尤度計算過程を判別関数計算過程とするニューラルネットワーク分類器の形式で実現するいくつかの手法も提案されている [12, 96]。また、MLPによる事後確率の推定によってHMMの出力確率の推定を置き換え、HMMの分類能力を向上させる考え方も提案されている [10]。

確率測度の推定

4．1．2で指摘したように、MLP等のニューラルネットワークパタン分類器の設計理念は判別関数接近法のそれに他ならず、その背景には統計的パタン分類の考え方が存在している。従って、現象を確率的にとらえようとする統計的接近をニューラルネットワーク分類器の設計に陽に取り込もうとすることは有意義

であるに違いない。

分類のためのベイズ決定則に直接かかわる確率は事後確率である。教師信号と分類器出力との間の自乗誤差損失を最小にすることを目指して設計される判別関数分類器が事後確率を近似することが古くから知られている（例えば、[19]）。こうした事実は、一般に条件確率を推定するベイズ接近に対して、判別関数接近が識別学習と呼ばれる１つの理由になっているとも言えよう。当然、自乗誤差損失の最小化によって設計されるニューラルネットワーク分類器もこの事後確率近似能力を持つはずであり、実際、様々な接近からその存在が指摘されている [5, 34, 125]。

ＭＬＰのような多くのニューラルネットワークは、各ネットワーク節点において内積を計算する。これに対して、ネットワークに直接、確率を計算させようとする試みも調査されている。例えば、ユークリッド距離がベイズ型確率関数に基づく尤度距離の簡単化された版であることに着目して、尤度ネットワークが提案されている [60]。この考え方は、混合ガウス型関数を出力確率モデルとする連続型HMMに近い。条件確率推定の古典であるパルツェン推定に動機づけられて、確率的ニューラルネットワークも提案されている [114]。

ニューラルネットワークによる確率計算を目指すもう１つの興味ある手法として、ファジー分割モデル（Fuzzy Partition Model （FPM））がある [119]。図４．１１はその構造を図解している。このネットワークは、それぞれのネットワーク節点における出力が負にならずしかもその総和が１であるという確率的な制約を持つことによって特徴づけられる。このＦＰＭを設計するためには、伝統的な判別関数法あるいはニューラルネットワークと同様に、合理的ないかなる損失をも用いることができる。 [119] では、特に、学習収束が速い divergence 距離に基づく損失を用いることの有効性も示されている。

リカレント構造

時系列信号処理において重要なニューラルネットワークに、過去の情報を現在の信号のモデル化に用いようとするリカレントネットワークがあり、実際、その音声パタン分類への応用可能性が様々な立場から調査されている。 [12] の議論は、HMMもリカレントネットの１種と見ることができることを示している。 [92]

は、シーケンシャルネットに基づくニューラルマルコフモデルを用いた音声認識例を示している。[21] もリカレントネットを用いる音声認識を試みている。[15] は決定的ボルツマンマシンを用いた音声認識の結果を示している。また [86] は、ＬＶＱ距離分類器にリカレント構造を持たせる工夫を行っている。

その動作原理から、リカレントネットワークが音声信号処理に適していることを直観することができる。しかし残念ながら、現在までのところ、この点を明確に支持する調査結果は必ずしも得られていない。リカレント構造を持つネットワークを正しく評価するために、原理的な表現能力の解析や効率的な設計手法の開発等に関する一層の研究が必要なものと思われる。

未知標本に対する耐性の向上

ニューラルネットワーク分類器は、その高い規準関数近似能力や損失選択の不適切さ（４．４．１参照。）ゆえに、設計標本に対する過度の最適化と未知標本に対する耐性の劣化とに陥ることがある。分類器設計の本来の目的は未知標本を正確に分類することであり、このような劣化を克服する努力が強く求められる。

耐性の低下は、基本的に、設計段階と分類段階との間の標本性質の不一貫性に起因する。不一貫性の例として、音声発話に伴う背景雑音や発話様式そのものの相違がある。

平均操作は、個々の標本が持つばらつきを抑え、標本全体に共通するある種の統計的に安定な特性を引き出す手続きである。分類判断においても、ただ１つの根拠によって下される判断と複数の根拠に基づいて下される判断とを比べると、一般に、複数の根拠に基づく平均的な判断の方が統計的に安定な、即ち、設計標本と未知標本とに対する差異の小さな分類を実現する。この普遍的な性質を分類器設計に活かす方策として、分類器の複製を複数用意する方法 [78] や M 類分類課題を２類分類課題の組み合わせ問題として扱う対判別関数を用いる方法（例えば [19] を参照。）等が考えられる。実際 [116] では、対判別関数型のニューラルネットワーク音声パタン分類器が調査され、その有効性が確かめられている。

また、与えられた設計標本を実効的に増加あるいは変更（加工）することによって耐性の向上を図る方法も調査されている（例えば [65, 66] 参照。）。ベイズ接近と比べるとかなり消極的ではあるが、こうした方法は、設計標本分布

に関するある種のモデルの導入を実行している。このモデル導入は、測度選択やパルツェン推定における核関数の選択に近い意図を持っていると見ることもできる。[99]では、設計標本を加工するという点で理念的には[65]の手法に類似であるが雑音対策という課題にやや固有の接近が調査されている。そこでは、異なる対信号エネルギー比の雑音が重畳された複数種類の設計標本を用いて、雑音耐性の大きなＴＤＮＮ－ＬＲ音声認識器（予測ＬＲ構文解析器とＴＤＮＮ分類器とが結合されている認識器）が実現されている。

分類規則（4.4）を用いるニューラルネット分類器は、正しい類に対して他の類よりも大きな判別関数を生成することが望まれる。広く採用されている自乗誤差損失を用いる設計では、この実現のために、訓練標本の類に対して１の教師信号を、他の類に対して０の教師信号を与えることが多い。しかし、分類判断は元々判別関数の大小比較のみに依っており、教師信号が１あるいは０である理由はない。この不連続な教師信号は、かえって判別関数の極端な大小関係を分類器に強制するという（結果的に、設計標本に対して不必要に最適化されるという）意味で、時によっては分類判断の耐性の劣化を招く原因になっている。この点に動機づけられて、柔らかな、つまりファジーな教師信号を用いることも調査され、その有効性が示されている[72, 73]。

教師なし学習による分類器の設計

判別関数接近の設計は原理的に教師付き学習である。分類課題においては分類結果を評価する何らかの教師信号が本質的に必要とされるが、その用い方によっては教師信号への依存を軽減することが可能である。調整は効率的ではあるが教師信号の利用に起因して柔軟性に欠けがちな判別関数設計過程や一般に資源浪費的な手作業に依存する大量の教師信号（例えば類記号）付きの設計標本の確保等の教師付き学習の問題を考えると、分類器の教師なし学習を実現することは調査に値する課題である。[25, 27]は音素認識課題におけるこうした考え方の有効性を示している。

4.2.2 不特定話者音声認識への応用

音声認識には、利用に伴う制約に基づく多様な種類の課題設定が考えられる。例えば、特定の発話者の音声を認識するだけでよい特定話者音声認識や不特定多数の発話者の音声を認識する不特定話者音声認識、1語1語、区切って発話される単語音声を認識する孤立単語音声認識、読み上げ文音声を認識する連続音声認識、会話音声等の自由に発話された音声を認識する自由発話音声認識、少数単語を扱えばよい少数語彙音声認識、そして大語彙音声認識等があり、また、これらの課題のそれぞれの性質の相違は決して小さくはない。従って、これまで紹介してきた分類器のそれぞれを効果的に応用するため、しばしば、課題に応じた追加的な技術開発を行う必要がある。こうした必要性は、一般に、課題が困難であるほど大きい。以下、4．2．2においては、極めて多くの応用研究が報告されている中から、特にＴＤＮＮとＦＰＭとが不特定話者音声認識課題に応用された例を紹介する。

ＴＤＮＮの利用

多人数発話者の音声が持つ音響的特徴の変動は、ある特定の発話者の音声のそれと比べてかなり大きい。これは、多人数発話者の利用を前提にする複数話者用音声認識器が、特定話者のみの利用を想定している特定話者用音声認識器よりも困難な分類状況に対処しなければならないことを意味している。また、設計段階で用いることができない未知の発話者の音声を認識しなければならない不特定話者用音声認識器は、この複数話者用認識器の能力に加えて、さらに分類すべき類固有の普遍的（汎化された）音響特徴をも把握することによって、この未知話者の問題に対処しなければならない。従って、一般に、不特定話者音声を扱う分類器の規模は大きくなり（調整可能なパラメタが多くなり）、しかも多数話者の大量の設計標本を用いる設計過程の規模も大きくなる。しかし、大規模な分類器の設計に大量の設計標本をそのまま供することは、例えば勾配探索に起因する設計の局所的最適性（分類が困難な複雑な課題では、一般に経験的平均損失の形状もまた複雑になる。）や計算量の観点から見て、明らかに非現実的である。

この困難な不特定話者用音声パタン分類器をＴＤＮＮを用いて実現するため、

以下の２種の接近が試みられている。

第１の接近は、図４．１２に図解しているように、比較的小規模のＴＤＮＮモジュラーを１話者あるいは少数話者毎に設計し、不特定話者用に用いる際にこれらを組み合わせて大規模な分類器を構成するものである [123]。しかし、このような単純なモジュラーの組み合わせが分類器全体の最適性を保証しないことは明らかである。

第２の接近は、話者適応、即ち、分類器の利用段階において新たな入力話者の少量の音声を用いて適応的に認識器の再設計を行うものである。このような適応能力は、発話者に関してのみならず、電話回線や、背景雑音、利用者の声の調子（例えば風邪ひきによる変化等）等の一般的な利用条件に関して、そして多くの応用の場面において間違いなく望まれるものである。例えば、鉄道の自動券売機の利用者が発話する音声は、せいぜいでも数語であり、そこに内臓されている音声認識器は、適応に必要な設計標本を確保できないであろう。このような応用の場面では、適応機能の価値は小さい。一方、特定話者による長期の利用が見込まれるコンピュータや電話器等の利用場面においては、適応による性能向上は明らかに歓迎される。

認識器を適応させる方法として、予め設計されている認識器全体あるいは分類器を適応させるものと、この予め設計されている認識器の中の分類器は変えずに特徴抽出器のみを適応させるものとが考えられる（例えば [45]。）。後者は、新しい話者等に起因する特徴の新しい性質を、分類器にとって既知である性質に写像（変換）すること、即ち話者写像の過程と見ることができる。一般に、この話者写像は、再設計における調整パラメタが少ないため処理は簡便である。ＴＤＮＮ音声認識器のためにも、この話者写像の考え方に沿ったいくつかの適応手法が調査されている。 [26] においては、ある時点における音響特徴ベクトルのみならず時間構造にも含まれる話者特徴を考慮した分節間のニューラルネットワークによる話者写像法の有効性が示されている。図４．１３はこの手法を図解している。用いられるＴＤＮＮ－ＬＲ音声認識器は、予め多量の音声標本を用いて設計されている。この適応のためのネットワークは、ＴＤＮＮに類似の時間ずれ構造を持っている。表４．４はその結果を要約している。話者適応前の単語・文節認識率が話者適応によって大幅に向上している。

ＦＰＭの利用

高速な学習は大量の設計標本の利用を可能にする。[119]では、divergence距離損失を用いて高速に設計されるＦＰＭ不特定話者音声認識器の性能調査が、大量の設計標本上で行われている。そこでは、男性話者、女性話者、及び男女話者混合音声で学習した3つのＦＰＭ－ＬＲ認識器を用いるＭｕｌｔｉ－ＦＰＭ－ＬＲ認識器による性能向上も試みられている。高速性は、ネットワークに入力される音響特徴ベクトル長の増加をも可能にする。この線に沿ったＦＰＭ分類器の改善も行われている[28]。

話者写像による適応は、ＦＰＭ音声認識器にも適用されている。[63]にはこの技術の将来性が示されている。また[16]には、同じ話者適応技術に基づく不特定話者ＦＰＭ音声認識器を言語識別（例えば、発話音声が日本語か英語か分類する）に応用する可能性も示されている。

4.2.3 言語処理への応用

言語処理は、最近のニューラルネットワーク応用研究の中の重要な話題の1つであった。実際、音声合成のための音声学的規則をネットワークによって獲得できることを示したネットトークが今日の研究関心の隆盛を導いた火付け役の一翼を担ったことは記憶に新しい[111]。しかし、職人芸的ともいえる知識集約型の開発が伝統的に行われてきた言語処理、特に機械翻訳（例えば日本語から英語への機械による翻訳）や音声認識のための言語処理の分野においては、ニューラルネットワークを用いる統計的接近の研究はまだそれほど活発ではない。

こうした中で、構文解析規則の学習[50, 51]や音声言語理解（発話された文等をそのまま分類するのではなく、その内容を把握して適切なシステム応答に結び付けようとする手法）システムの構築を目指した言語獲得[35, 36]等、いくつかの挑戦的な研究が始められている。ニューラルネットワークを用いて単語の品詞予測を行うＮｅｔ－ｇｒａｍ（N組の英語表現であるＮ－ｇｒａｍに由来する）も調査されている[83, 93, 94]。また、辞書中の音素テキストニューラルネットワークの出力パタンに基づく単語予測も調査されている[29]。「これらのニューラルネットワークに基づく手法が伝統的な手法にとって代わる新しい時代

を築くことができるかどうか。」これは、今後の研究における重要な問いの１つである。

4.3 音声特徴抽出への応用

４．１．１で紹介したように、音声パタンは、一般に、フーリエ変換や線形予測法等の、予め設定された写像方法によって音響特徴ベクトル列に変換される。これらの変換、言い替えればある種の特徴抽出は、ある先見的な（おそらく限られた）知見に基づくものであり、音声認識等の実際の利用にとって本当に望ましいものである保証はない。サウンドスペクトログラムに話者特徴も音素特徴も混在していたように、この比較的単純な先見的変換は、必ずしも、利用目的を達成するために最善な特徴を抽出しているわけではない。

ニューラルネットワーク分類器の設計に用いられる教師付き学習の考え方は、優れた表現／写像能力を持つネットワークを特定の特徴を抽出（発見）する特徴抽出器として設計するために用いることもできる。類を指定する代わりに、抽出したい特徴を特定する信号を教師信号として用いるのである。実際、上記のような先見的な抽出法に代わる、音声そのものあるいは音声を含む入力信号から目標である特徴を抽出するニューラルネットワーク特徴抽出器を設計する試みが調査されている。前出の話者写像もこの特徴抽出の１種と見ることができる。

言語処理同様、音声情報処理における特徴抽出技術の分野も、今だに線形予測法等の伝統的手法が主流を占めている。ニューラルネットワーク応用の研究は、まだ発展途上の段階にあるということもできる。４．３では、数少ない応用例の中から、特に短時間スペクトル特徴の抽出や雑音に埋もれた音声信号から音声のみを抽出する雑音抑圧処理等の例を紹介する。

4.3.1 短時間スペクトル特徴の抽出

短時間スペクトルを求める代表的な手法の１つに線形予測法があった。これは、元々、最尤推定によるスペクトル推定法として音声信号処理に導入され[47]、その後、ＰＡＲＣＯＲ法、声道断面積法、ＬＳＰ法等に発展され、現代的音声工学を形成する上で著しい貢献を果たした手法である。推定の原理は、過去の観測

信号の線形荷重和モデルを用いて目標信号の最小自乗誤差近似を行うものである。このようなモデルの構造や設計理念は、多層構造ニューラルネットワークを自乗誤差損失を用いて設計するものに近いことに気付くであろう。実際、[46]においては、動的時間伸縮を利用してニューラルネットワーク非線形予測器を設計する手法の音声認識における有効性が示されている。

4.3.2 リカレントネットワークによる音声ゆらぎのモデル化

音声信号は、音素特徴や話者特徴のような、いわゆる情報を伝達するための特徴の他に、自然性のような特徴をも担っている。仮に、機械との音声対話を可能とする高度情報社会が訪れても、いかにも機械的な合成音声があふれていたのでは、そこは情報社会であっても高度な社会ではない。この自然性の解明とモデル化は、高品質な合成音声の作成等の音声情報処理における重要な研究課題の1つである。

既にふれたように、リカレントネットワークは音声時系列信号をモデル化する原理的適性を持っている。[110]においては、非線形振動子の局所的結合から構成されるリカレントネットワーク（ＡＰＯＬＯＮＮと命名されている）を用いて、自然なゆらぎ（音声の自然性を支配する重要な要因の1つである。）を含む音声の音源信号をモデル化する試みが行われている。実験の結果、音源波の周期性や振幅等のゆらぎがかなり良くモデル化されることが示されている。

4.3.3 雑音抑圧

音声信号の背景にある雑音を抑圧して音声信号のみを抽出する雑音抑圧は、写像を通して音声らしさの抽出を実現するある種の音声特徴抽出であるとみなすことができる。この雑音抑圧への応用は、ニューラルネットワークによる非線形写像の1つの例として早くから研究されている[117, 118]。雑音を伴った音声入力と無雑音の音声から成る教師信号とを用いることにより、写像関係を設計する。雑音抑圧ニューラルネットワークの構造を図４．１４に示す。設計された写像の雑音抑圧性能は、被験者による聞き取り実験を通して評価され、Ｓｐｅｃｔｒａｌ Ｓｕｂｔｒａｃｔｉｏｎ法と呼ばれる従来法よりも優れていることが示され

ている。また、符号帳写像による方法と組み合わることにより、抑圧性能を一層向上できることも確かめられている [98]。

4.4 新しい設計枠組みを求めて

4．1．2の判別関数接近の節において述べたように、ニューラルネットワークによるパタン分類法の大部分は、判別関数接近における分類器設計に他ならない。また、学習ベクトル量子化も判別関数接近の1手法であった。多くの応用例は、これらのニューラルネットワークに基づく音声パタン分類の優れた将来性を示唆している。しかし、最適性等のネットワーク設計の数理的基盤は、必ずしも十分に明らかにされてはおらず、こうした事情は、不注意な応用やニューラルネットワークそのものに対する偏った評価をもたらしてきた。

また特に、動的な音声パタンを扱う音声パタン分類器においては、ＤＴＷやＨＭＭとニューラルネットワークとの混成様式が盛んに用いられてきた。しかし、非線形時間伸縮部の設計とニューラルネットワーク部との設計が別々に行われるため、一般に、このような混成型分類器の最適性は保証されていない。

以上のようなニューラルネットワーク応用の実情を見るとき、精力的に行われているニューラルネットワーク応用の数理的基盤に対する理解が、意外なほど不十分であることに驚くかもしれない。この問題に対して、最近、分類器設計のための新しい理論的枠組み、最小分類誤り／一般化確率的降下法（Minimum Classification Error Formalization (MCE) / Generalized Probabilistic Descent Method (GPD)）が提案され、それに関する一連の研究が精力的に展開されている。4章の最後では、この新しい汎用的な判別関数設計法の概要を紹介しよう。

判別関数接近の節でも若干触れたが、ここで判別関数設計を特徴づける4つの要因を改めて整理しておこう。その要因とは、

1. 分類器構造（測度）

2. 設計標的関数（損失）

3. 最適化法

4. 未知標本に対する耐性

である。第1の要因に関しては、多くの例を4．2．1で紹介した。第2と第3の要因に関しては、4．1．2の判別関数接近の節において述べた。第4の要因に関しては、4．2．2の未知標本に対する耐性の向上の節において若干ふれた。

耐性の問題に関して、ここで若干補足しておこう。元々、パタン分類も特徴抽出も、その設計の目的は、設計標本に対してではなく、未知なる標本に対する高い性能を実現することにある。従って、設計は、未知なるものへのなんらかのモデル化あるいは配慮を必要とする。モデル化によって耐性を向上させようとする代表例は、ベイズ接近におけるような標本分布を陽に導入するものである。用いるモデルが適切であれば効果は大きい。与えられた設計標本の取り扱いに配慮して耐性を向上させる例には、その標本を分割して、未知標本に対する疑似的な評価を行おうとする交差法がある。また一般に、分類器の調整可能なパラメタが多いほど、設計は、設計標本に対するより高い分類精度を実現できる。しかし、このような精度の向上は、しばしば、設計標本に対する過度な整合化、言い替えれば、耐性低下を招く。従って、調整可能なパラメタは、必ずしも多いほうが望ましいわけではない。適切なパラメタ数を設定するため、ＡＩＣに基づく手法 [121] 等が調査されている。

さて、以上のような4大要因を基本的な軸にして、これまで紹介してきた分類器の主なものを表4．5に整理しておこう（1部、後述されるＭＣＥ／ＧＰＤ分類器も含んでいる）。もはや、「HMMよりもニューラルネットワークは耐性が低い。」等の誤った理解は持たないであろう。分類器の性質を評価する時、実際に用いられている設計戦略を注意深く調査する必要がある。言うまでもなく、固定観念に束縛された評価は排除されるべきである。

4.4.1 最小分類誤り／一般化確率的降下法

課題が分類である時、設計は分類誤りを最小にすべく行われるべきである。最終的な設計の目的が未知なる標本に対する正確な分類であるならば、設計は未知なる標本分布にできるだけ直接的な方針で行われるべきである。数ある分類規則

の中でも（4.1）を用いるのであれば、設計はこの規則の実行にできるだけ直接的に行われるべきである。（4.4）を規則とする場合もまた同様である。このように極めて当然とも言える設計に対する要求の観点から見るとき、ベイズ接近も判別関数接近も明らかに不適切である。複数の設計基準を組み合わせる混成構造の採用等ができるだけ避けられるべきことは言うまでもない。

分類問題を考えるとき、類境界の推定に関する原理的な非一貫性を持つベイズ接近はどうやら本質的にこの問題に対する優れた解には成りえないようである。それではもう一方の接近、判別関数接近はどうであろうか。4．1．2や4．1．3で紹介したように、判別関数接近の多くの設計法は与えられた設計標本に対する分類誤りの減少を（少なくとも理念的には）直接的に目指している。しかし、実際の設計結果と究極の設計目的である最小分類誤り確率状態との一貫性はほとんど未知のままである。せいぜいでもＬＶＱ２に見られるような直観や限られた場合の解析結果しか得られていない（例えば [19] 参照。）。もしこの最小分類誤り確率状態との一貫性の問題が解決あるいは軽減されれば、元々扱いやすい方法論であるという実際性ともあいまって、判別関数接近は分類課題に直接的な設計手法の実現に大いに貢献するに違いない。

ＬＶＱの訓練最適性に関する調査から始められたＭＣＥ／ＧＰＤ開発の底流には、常にこうした問題意識と信念とが存在し続けてきた。そのため、研究は、ＬＶＱ調整規則の最適性の調査にとどまらず分類課題に直接的でかつ高い汎用性を持つ設計手法の確立を目指して行われた。その詳細は [55, 56, 58, 61] 等に詳しい。ここでは、その文献中の定形化に込められている理念の紹介に焦点を合わせよう。

名前からわかるように、ＭＣＥ／ＧＰＤは２つの手法の組み合わせである。ＭＣＥは分類誤り数損失の最小化を企てる分類器設計法の定形化のことであり、ＧＰＤは６０年代後期に開発された分類器設計法、確率的降下法に基づく適応型最適化法である。初め、ＧＰＤが開発され [58, 61]、そのパタン分類における理論的貢献に着目することによってＭＣＥが生み出された [56]。[62] では、ＭＣＥとは、最小分類誤り状態に一貫した分類器状態を判別関数接近によって設計できるという可能性を示した、分類誤り数損失（（4.24）及び（4.25）参照。）を用いる設計法定形化のことであり、ＧＰＤとは、ＭＣＥ定形化及び動的パタン分類

問題をも含む一般的な分類器設計のための最適化手法のことであると整理されている。しかし実際には、この2つの手法は互いに密接に関係し合っており、最近のほとんどの応用では区別されずにMCE／GPDとして用いられている。以下でも、両者の区別が必要な場面ではその旨を明記し、それ以外はMCE／GPDとして取り扱っていこう。MCE／GPDの要点は、以下のようにまとめることができる。

1. 現実的な勾配探索設計法の利用をも可能にする平滑な（Λ に関して少なくとも1次微分可能な）関数様式に与えられた課題（分類課題）全体を組み込み、鎖則を用いてその全設計過程の最適化を図るという方法論を示した。

2. 確率測度をも利用できる確率的降下パタン分類学習則を厳密な様式で示した。

3. 定形化における平滑性を制御することによって、分類器の大局的最適状態の探索あるいは学習の耐性を解析するための基盤を与えた。

4. 静的パタンのみならず音声のような動的パタンを分類するための、HMMやニューラルネットワークを包含するほとんど全ての合理的分類器構造に適用可能な判別関数設計法の新しい族を与えた。

5. GPD と既存の識別学習法との関連を明らかにし、特に、LVQが平滑な分類誤り数を損失とするGPD実装の簡単化された版であることを示した。

MCE／GPDは、判別関数接近における設計法の1つである。多様な最適化法がある中で、GPDは、計算速度等に関する実際性や適応型であるという特性が考慮され、特に、適応型勾配探索による損失最小化法の1つ、確率的降下法にその原理を求めている。

勾配法においては、損失は平滑でなければならない。ところが、分類過程には Λ に関して不連続な操作がしばしば使われている。動的パタンの分類に多用されるDTWは、Λ（参照パタン）に関して不連続である。(4.20) を思い出そう。経路選択は、損失関数上の不連続の移動を引き起こす。同様に、LVQ距離分類器の説明に登場した最近傍参照パタンの選択も Λ（参照ベクトル）に関

して不連続である。確率的降下法の元々の定形化にも非平滑性の問題があった。MCE／GPDの定形化における鍵の1つはこの非平滑性を克服することであった。

説明のために、4．1．4の動的時間伸縮を用いる距離分類器の節で用いた動的パタンのM類分類課題を再び用いよう。用いる分類規則はここでも（4.15）である。

初めに判別関数が定義される。動的パタンを測るための関数の定義には多くの選択があるが、例えば、

$$g_i(x;\Lambda) = \left[\frac{1}{Q_i}\sum_q\left[\left[\frac{1}{\Phi_i^q}\sum_\theta\{D_\theta(x,r_i^q)\}^{-\xi}\right]^{-1/\xi}\right]^{-\zeta}\right]^{-1/\zeta} \tag{4.21}$$

のようなL_pノルム形式の連続関数がしばしば用いられる。L_pノルム形式の効果は、特に、以下のような近似を知ることによって明らかになる。（4.21）のζを大きくしてみよう。平滑な形式によって、最近傍参照パタン距離の選択操作を近似できる。同様に、ξを大きくすることにより、最適経路選択操作を平滑な形式で近似することができる。つまり、連続関数の枠組みの中で最小探索という不連続な操作を近似できるのである。実際、参照パタンや経路の利用において唯一の最適候補（最近傍参照パタンあるいは最適整合経路）を用いる理由はそれ程明確でない。DTWの経路選択においても、複数個の経路の利用が推奨されるようになってきた [113]。4．2．2の未知標本に対する耐性の向上の節でも述べたように、複数の判断基準の利用は、耐性向上の観点からも望まれる。L_pノルム形式によ平滑な関数定義は、これらの事実によっても支持される。

さてここで、設計の目的は、分類則（4.15）に従って正確な分類を行う分類器を実現することである。言うまでもなく、従来の設計手法のいずれの背景にも、この目的を達成しよとする企図が存在した。しかし、ほとんどの場合、その達成の方法は婉曲であり不十分であった。この不十分さの原因は、設計過程の定形化がその根底にあるべき分類則に直接的に立脚していないという点にあった。MCE／GPDの定形化は、この問題を以下のようなL_pノルム形式の誤分類測度を導入することによって解決する。ここでも、前段落の理念を満足する他の関数形式の選択が可能である。設計のためにC_kに属する動的パタンx_nが与えられて

いるとしよう。誤分類測度は

$$d_k(x_n;\Lambda) = g_k(x_n;\Lambda) - \left[\frac{1}{M-1}\sum_{j,j\neq k}\{g_k(x_n;\Lambda)\}^{-\eta}\right]^{-1/\eta} \tag{4.22}$$

のように定義される。ここで、正の測度は x_n が誤って分類されたことを意味し、負の測度は x_n が正しく分類されたことを意味する。即ち、この誤分類測度は、分類則における判別関数の比較操作をスカラー値の大小比較に置き換えて実現していることになる。与えられた課題を直接的に設計過程に組み込もうとする開発の意図がここで明確に実現されている。

ここで式 (4.22) 中の η を無限大に設定してみよう。式 (4.22) の誤分類測度は

$$d_k(x_n;\Lambda) \approx g_k(x_n;\Lambda) - g_i(x_n;\Lambda) \tag{4.23}$$

となる。ここで、C_i は C_k 以外の類の中で最も小さな判別関数値を示す類である。この時、分類判断における比較操作は、設計標本 x_n が属する C_k とそれ以外の、しかし最も可能性の高い（判別関数値が小さい）類 C_i との間のみで行われる。

誤解を避けるために、この測度の概念が実は確率的降下法において既に導入されたものであることを紹介しておこう。しかし、設計目的に直接的でかつ平滑な関数様式で設計過程を定形化するという明瞭な役割は、ＭＣＥ／ＧＰＤの枠組みにおいて初めて与えられるようになった。

分類判断を模擬する誤分類測度が与えられれば、後は、判別関数接近の基本的戦略に従って、損失を選択し、その損失から成る経験的平均損失の最小化を行えば良い。

損失には、既にふれたように様々な定義がある。４．２．１の確率測度の推定の節に紹介した事後確率近似能力を考えれば、広く用いられている自乗誤差損失を用いるのが適当かもしれない。しかし、明らかに自乗誤差損失の最小化は分類誤りの減少と一貫していない（例えば [19, 38]。）。一方、(4.7) から (4.9) において、分類結果に支配される部分集合、混同類集合上で加算操作が行われていることを思い出そう。パーセプトロン損失は平滑でなく、原理的に利用に適さない。パタン分類器設計の目的が正確な分類、言い替えれば、分類誤りの減少を

実現することであることを考えると、分類誤り数損失

$$\ell_k(x_n;\Lambda) = \begin{cases} 1, & \text{misclassification} \\ 0, & \text{otherwise} \end{cases} \tag{4.24}$$

を用いることは自然である。しかし、この損失もまた明らかに非平滑である。この非平滑性の問題を解決できれば、設計目的に最も直接的な損失を用いることができるようになる。MCE／GPD、特にMCEは、(4.24) の近似である

$$\ell_k(x_n;\Lambda) = \frac{1}{1+e^{-(\alpha d_k(x_n;\Lambda)+\beta)}} \qquad \alpha > 0 \tag{4.25}$$

のような平滑な分類誤り数損失を用いることを提案した。ここでαとβはスカラーの定数である（実は、αもβも調整可能なパラメタとして扱うこともできる。）。

以上の手続きで、損失は定義された。ここで、MCE／GPDの定形化が、分類過程に登場する関数を段階的にその上位の関数（後段で定義される関数）の中に組み込んでいく基本戦略を持っていることに注意しよう。Λに関する勾配の計算は、鎖則を用いて、損失から誤分類測度、判別関数へ伝搬される。逆誤差伝搬法は明らかにこの伝搬機構の特殊な例である [3, 4]。Λである参照パタンのための調整規則は、確率的降下定理 [1, 2, 56, 58]

与えられた標本に関し$x_n \in C_k$と仮定する。もし分類器のパラメタの修正量$\delta\Lambda(x_n, C_k, \Lambda)$が以下のように設定されるものとすれば、

$$\delta\Lambda(x_n, C_k, \Lambda) = -\epsilon \mathbf{U} \nabla \ell_k(x_n;\Lambda) \tag{4.26}$$

但し、ここでUは正定値行列でϵは小さな正の実数である重み係数であり、その時、

$$E[\delta L(\Lambda)] \leq 0 \tag{4.27}$$

が成り立つ。さらに、もし無作為に抽出された標本の無限個列$x(t)$が学習（設計）に用いられ、かつ (4.26) の修正規則が以下の条件

$$\sum_{t=1}^{\infty} \epsilon(t) \to \infty \tag{4.28}$$

$$\sum_{t=1}^{\infty} \epsilon(t)^2 < \infty \tag{4.29}$$

を満たす重み係数列 $\epsilon(t)$ とともに用いられるものとすれば、その時

$$\Lambda(t+1) = \Lambda(t) + \delta\Lambda(x(t), C_k, \Lambda(t)) \tag{4.30}$$

に基づくパラメタ列 $\Lambda(t)$ は $L(\Lambda)$ の少なくとも局所的最小状態をもたらす Λ^* に確率 1 で収束する。

に従って求められる。詳しい導出法は例えば [74] 等に詳しいので、ここでは $\zeta, \xi, \eta \to \infty$ の条件によって簡単化された場合、即ち唯一の最適経路距離のみによって測られる最近傍参照パタンの距離によって各類の判別関数が決定され、かつ C_k 以外の類の中で最も判別関数が小さな類 C_i のみが C_k との比較に用いられるという場合に対する調整規則

$$r^q_{j,\tau,s}(t+1) = \begin{cases} r^q_{j,\tau,s}(t) + 2\epsilon(t)\nu_k(x_{n,1(j,q,\tau),s} - r^q_{j,\tau,s}(t)), & \text{for } q = C_j \text{ and } j = k \\ r^q_{j,\tau,s}(t) - 2\epsilon(t)\nu_k(x_{n,1(j,q,\tau),s} - r^q_{j,\tau,s}(t)), & \text{for } q = C_j \text{ and } j = i \\ r^q_{j,\tau,s}(t), & \text{otherwise} \end{cases} \tag{4.31}$$

を紹介しておこう。なおここで、$r^q_{j,\tau}$ は参照パタン r^q_j の第 τ 番目の音響特徴ベクトルを表わし、$r^q_{j,\tau,s}$ はその第 s 次元目の要素を表わすものとする。また、$x_{n,\theta(j,q,\tau)}$ は r^q_j に関する第 θ 最適経路に沿って決定される $r^q_{j,\tau}$ に対応する x_n の音響特徴ベクトルを表わし、特に $1(j,q,\tau) = \theta(j,q,\tau)\ |_{\theta=1}$ である。参照パタン同様、$x_{n,1(j,q,\tau),s}$ は音響特徴ベクトル $x_{n,1(j,q,\tau)}$ の第 s 次元目の要素である。さらに $\nu_k = \ell'_k(x_n;\Lambda)$ である。音響特徴ベクトル間の局所距離には自乗ユークリッド距離を用いている。定理は、(4.28) 及び (4.29) の確率近似法の条件を満足することによって、この調整規則の無限の繰り返しが $L(\Lambda)$ の少なくとも局所的な最小状態を達成できることを示している。もっとも実際の設計では、通常、$\epsilon(t)$ に有限の単調減少関数を採用して近似的な最適化が行われる。

静的パタンは、動的パタンの特殊な場合に過ぎない。従って、(4.31) の調整規則はそのまま、非線形伸縮の部分を除くだけで静的パタンの分類器設計に適用

できる。非線形伸縮過程の省略と ν_k に関する若干の追加的簡単化を経て、(4.31) は静的パタン分類器のための調整規則である改良ＬＶＱ２と等価となる。実際、(4.19) と (4.31) との比較から、容易にこの等価性を予想することができる。HMMやニューラルネットワーク構造のためのＭＣＥ／ＧＰＤの詳細は [60, 61] に詳しい。なお、４．１．４のHMM分類器の節でふれた相互情報量最大化法がＭＣＥ／ＧＰＤの特殊な場合であることも容易に導くことができる。

さて、繰り返し述べてきたように、パタン分類器設計の究極の目標の１つは最小分類誤り状態の実現である。しかし、長い歴史を持つにもかかわらず、「判別関数接近によってこの目標をどのように実現できるのか。」という問に対する解析はあまり十分に行われていなかった。[56] は、この点に関する新しい見通しを以下のように与えている。尤度ネットワークを分類器構造とする判別関数

$$g_i(x;\Lambda) = p_\Lambda(C_i \mid x) \tag{4.32}$$

を考えよう。まず、事後確率密度関数の形式が既知、言い替えれば、尤度ネットワークパラメタ Λ の形式が事後確率密度関数を決定するパラメタ形式に等しいという（あまり現実的ではないが）仮定をしよう。分類則は (4.4) に等しく、さらに上記の距離分類器の場合とほぼ同様の手続きを経て (4.25) と同じ形式の平滑な分類誤り損失が定義される。ここで、最小化の対象である期待損失が

$$\begin{aligned} L(\Lambda) &= \sum_k \int_{\mathcal{X}} p_\Lambda(x, C_k) 1(x \in C_k) \ell_k(x;\Lambda) dx \qquad (4.33) \\ &\simeq \sum_k \int_{\mathcal{X}} p_\Lambda(x, C_k) 1(x \in C_k) 1\Big(p_\Lambda(C_k|x) \neq \max_i p_\Lambda(C_i|x)\Big) dx \end{aligned}$$

のように書き換えられることに気付こう。なお、ここで $\mathcal{X}$ は動的パタン x の全標本空間である。L_p ノルムの定数を変化させて平滑性を制御し、この近似精度を恣意的に向上させることができる。しかも、上記の仮定によって、ここでは Λ の調整によって達成される $L(\Lambda)$ の最小状態に対応する Λ は真の標本分布関数を実現する、言い替えれば最大事後確率状態を達成するものに等しい。結果的に、期待損失 $L(\Lambda)$ の最小状態は最小分類誤り確率

$$\mathcal{E} = \sum_k \int_{\mathcal{X}_k} p_\Lambda(x, C_k) 1(x \in C_k) dx \tag{4.34}$$

$$\text{where} \quad \mathcal{X}_k \;=\; \left\{ x \in \mathcal{X} \;\middle|\; p_\Lambda(C_k|x) \neq \max_j p_\Lambda(C_j|x) \right\}$$

に限りなく近づけることになる。

上の段落の結果は、判別関数接近による最小分類誤り確率状態の達成が可能であることを示す、極めて興味深いものである。しかし、用いられた仮定は、ベイズ接近に対する批判と同様に明らかに非現実的である。

[30] 等には、3 層の MLP が任意の関数を近似できる原理的能力を持つことが証明されている。[39, 101] 等には、円形基底関数ネットワークがユニバーサルな近似能力を持つことが示されている。円形基底関数ネットワークと構造的に類似の混合ガウス分布関数は、最小 L_p ノルム誤差の意味で任意の関数を近似できることが示されている [112]。これらの結果から、十分な隠れ層を持つ尤度ネットワーク（ＭＬＰでも良い。）は、（未知の）真の分布関数を表現できる原理的能力を持つと見なすことができる。もし、$L(\Lambda)$（及び E）が唯一の最小状態を持ち、かつ Λ と $L(\Lambda)$（及び $\mathcal{E}$）とが単調な関係を持つものとすれば、その最小状態は、Λ によって表現される $g_i(x;\Lambda)$ が真の確率関数に等しい場合に対応する。従って、上記の仮定よりはかなり緩い（より現実的な）この仮定の下で、たとえ真の確率関数のパラメタ形式も未知であっても、判別関数接近によって最小分類誤り確率状態を原理的に達成できることがわかる。

以上の最小分類誤り確率状態に対する設計の原理的一貫性の議論は、ＧＰＤのような実際の最小化手法の選択と独立している。また、動的パタンの分類に閉じた話題でもない。そこで、この議論に登場する定形化は、特に、ＧＰＤと区別してＭＣＥと呼ばれている [62]。

さて、これまでの議論から、関数の平滑性は、もっぱら定義の厳密性を実現するために役立っているものと理解するかもしれない。しかし実は、設計における重要課題の 1 つである未知標本に対する耐性向上にも大きく貢献している。即ち、[55] に示されているように、平滑な設計過程は、元々設計標本位置のみで変化する階段関数である経験的平均損失を平滑にする。この関数が平滑であるということは、与えられた設計標本の付近にも実質的に設計標本が存在することに対応する。こうして、ＭＣＥ／ＧＰＤでは、最も簡素なしかし自然なモデル、即ち標本分布の連続性を用いて耐性の向上が行われている。

パタン分類器設計の 1 つの目標は最小分類誤り確率状態の達成であった。より

一般的なシステムを設計する時は、判断結果に対する危険の最小状態を達成することが目標となり得る。(4.25) の平滑な分類誤り損失に準ずる他の損失を用いて目指されるこの設計においても、これまでのMCE／GPDの議論の核心は全て成立する（例えば、[19, 89] 参照）。そこでは、MCEはより一般的な最小危険あるいは最小誤り定形化と呼ばれるであろう。

MCE／GPDの応用は、特に音声パタン分類において、その開発以来集中的に調査されている。参照パタン距離分類器のための実装は [13, 74] 等に詳しい。HMMに類似した時間構造を持つ距離分類器のためには、[88] 等に実装の詳細が示されている。HMM 音声認識器のための応用例は [14, 104] に示されている。言語処理部も含んだ音声パタン分類器の設計法への応用は [44] に報告されている。[115] には話者写像問題への例が示されている。MCE／GPDの設計耐性の向上は [17, 88] において調査され、特に背景雑音に対する耐性向上策は [100] で調査されている。

表4．6は、前出の米語E集合音声認識課題における上記応用例の結果をまとめたものである。MCE／GPDの分類精度向上に対する明らかな貢献を見い出すことができる。

4.4.2 最小スポッティング誤り学習

MCE／GPDの設計理念は、与えられた課題の目標あるいは判断規則に直接的な設計過程を定形化することであった。従って、パタン分類の枠組みで開発が行われたMCE／GPDは、その理念の一般性はともかくとして、実際の定形化はパタン分類問題に偏ったものであった。理念を活かすべき状況は、実際、他にも極めて数多く存在する。

4．1．4の分節化と非線形時間伸縮を伴う音声パタン分類の節にも述べたように、音声認識の実際は単純なパタン分類より複雑である。特に、分節化のように、与えられた長い音声標本から認識対象である分節を切り出す操作を伴うことが多い。この極端な例として音声スポッティングがある。音声スポッティングは、予めモデル化された音声分節と同類の分節を（一般に）長い入力音声から正確に見つけ出す技術である。この過程における判断は、分類ではなくスポッティングに関して行われる。スポッティングは、モデルと入力音声中の分節との間の

類似度（距離や尤度等によって測られる）と予め設定されている閾値との大小比較によって行われる。測度が距離の時は、通常、閾値を下回る区間に見い出すべき類の分節が存在すると判断される。可能な判断誤りとしては、脱落（本来存在する分節を抽出できない）と付加（本来存在しない分節を誤って抽出する）との2種類がある。また、音素や音節分節に対するスポッティングもあるが、特に、**重要単語（キーワード）のスポッティング**は、自由発話音声の理解に効果があると考えられている。

従来のスポッティングシステムの設計は、せいぜいでもベイズ設計に従う類モデルの設計と実験的に決定される閾値の設計との経験的な組み合わせによって行われてきた。しかし、このような組み合わせの最適性が保証されていないことは明らかである。この問題を克服するため、[75] はＭＣＥ／ＧＰＤの理念に基づく**最小スポッティング誤り設計法**（Minimum Spotting Error Learning (MSPE)）を提案している。その定形化の要点は、抽出すべき類モデルのみならず閾値をも調整可能なΛとして平滑なスポッティング誤りの定義の中に組み込むことである。表4．7は、ＧＰＤに基づく最適化によってスポッティング性能が著しく向上する様子をまとめている。

4.4.3 識別的特徴抽出

ＭＣＥ／ＧＰＤの定形化の特徴の1つに、与えられた設計過程を段階的に平滑な関数形式に組み込んでいくというものがあった。[44] の言語処理部と音響処理部との統合的設計への応用は、この点に注目した応用例の1つであると見ることができる。[8, 9, 62] には、もう1つの応用例、**識別的特徴抽出**（Discriminative Feature Extraction (DFE)）が報告されている。

4．1．1及び4．3で紹介したように、パタンの特徴抽出過程は、その特徴表現されたパタンを対象とする分類過程と独立に設計されることが多い。従って、この2つの過程の間には、認識率（分類精度）の向上という認識の全過程の最終目標に対する一貫性はない。しかし、特徴抽出過程が認識過程の1部である以上、特徴抽出過程は、その後段にある分類過程と共に認識の最終目標に一貫した形式で設計されるべきであることは疑いない。こうした理解に基づいて、識別的特徴抽出においては、特徴抽出のためのパラメタはΛの1部として分類器パ

ラメタと共にMCE／GPDに基づいて設計される。

音声認識のための特徴抽出のみを考えても、この手法の応用場面は数多い。[8]では、広く用いられているケプストラムを対象に、認識精度の向上に一貫したケプストラムの重みづけ特徴抽出（**リフタリング**と呼ばれる）の最適設計が後段のニューラルネットワーク分類器の設計と同時に行われている。設計の結果、分類精度の向上ばかりか興味深い**リフタ**（重み関数）の形状も得られている。獲得された形状は、音声認識精度向上の阻害要因となり易い話者特性を巧妙に抑制するものになっている。

音素等の類固有の特徴は、ホルマント構造に代表される比較的低い周波数帯（電話音声の帯域にほぼ対応する）に集中している。そこで、音声認識率の向上のためには、この帯域に注目した特徴の選択が望ましく、実際、メルスケールやバークスケールと呼ばれる低い周波数帯域の寄与を大きくするような特徴抽出法が広く用いられている（4．1．1参照。）。しかし、これらのスケールの背景にある心理学的知見は、元々、機械による統計的なパタン分類に対して直接の関係を持っていない。認識結果に一貫した非線形周波数軸伸縮特徴抽出法の開発が望まれる。[9]においては、この問題に焦点を合わせた識別的特徴抽出の有効性が示されている。そこでは、周波数軸の非線形伸縮による特徴抽出過程と後段の距離分類器とが統合的に設計され、認識率向上に求められる周波数伸縮がメルスケール及びバークスケールのいずれの既存のスケールとも異なることが明らかにされている。

参考文献

[1] S. Amari;"A Theory of Adaptive Pattern Classifiers", IEEE, Trans. Electronic Computers, Vol. EC-16, No. 3, pp. 299-307 (1967 6).

[2] 甘利俊一; "情報理論 II - 情報の幾何学的理論 -", 共立出版 (1968).

[3] 甘利俊一; "神経回路網モデルとコネクショニズム", 東京大学出版会 (1989).

[4] 麻生英樹; "ニューラルネットワーク情報処理 - コネクショニズム入門, あるいは柔らかな記号に向けて -", 産業図書 (1988).

[5] H. Asoh, and N. Otsu; "Nonlinear Data Analysis and Multilayer Perceptrons", IEEE, Proc. IJCNN, Vol. 2, pp. 411-415 (1989 6).

[6] L. Bahl, P. Brown, P. de Souza, and R. Mercer; "Maximum Mutual Information Estimation of Hidden Markov Model Parameters for Speech Recognition", IEEE, Proc. ICASSP86, Vol. 1, pp. 49-52 (1986 4).

[7] L. Bahl, P. Brown, P. de Souza, and R. Mercer; "A New Algorithm for the Estimation of Hidden Markov Model Parameters", IEEE, Proc. ICASSP88, Vol. 1, pp. 493-496 (1988 4).

[8] A. Biem, and S. Katagiri; "Feature Extraction Based on Minimum Classification Error/Generalized Probabilistic Descent Method", IEEE, Proc. ICASSP93, Vol. 2, pp. 275-278 (1993 4).

[9] A. Biem, S. Katagiri, and B.-H. Juang; "Discriminative Feature Extraction for Speech Recognition", to appear in "Neural Networks for Signal Processing III - Proc. of the 1993 IEEE Workshop" (1993 9).

[10] H. Bourlard, and C. Wellekens; "Links Between Markov Models and Multilayer Perceptrons", IEEE, Trans. PAMI, Vol. 12, No. 12, pp. 1167-1178 (1990).

[11] J. Bridle, R. Chamberlain, and M. Brown; "An Algorithm for Connected Word Recognition", IEEE, Proc. ICASSP82, pp. 899-902 (1982 5).

[12] J. Bridle; "Alpha-Nets: A Recurrent 'Neural' Network Architecture with a Hidden Markov Model Interpretation", Speech Communication, Vol. 9, pp. 83-92 (1990).

[13] P.-C. Chang, and B.-H. Juang; "Discriminative Template Training for Dynamic Programming Speech Recognition", IEEE, Proc. ICASSP92, Vol. 1, pp. 493-496 (1991 5).

[14] W. Chou, B.-H. Juang, and C.-H. Lee; "Segmental GPD Training of HMM Based Speech Recognition", IEEE, Proc. ICASSP92, Vol. 1, pp. 473-476 (1992 3).

[15] J. -C. Dang, S. Tamura, and H. Sawai; "Shift-Invariant Deterministic Boltzmann Machines for Phoneme Recognition", IEICE, Tech. Report SP89-98 (1990).

[16] I. Donescu, Y. Kato, and M. Sugiyama; "Speaker-Independent Features Extracted From a Neural Network and Their Evaluation in Speech Recognition", IEICE, Tech. Report SP92-117 (1993 1).

[17] A. Duchon, and S. Katagiri; "Increasing the Robustness of GPD-Based Algorithms", ASJ, Proc. Conf., pp. 205-206 (1992 3).

[18] A. Duchon, and S. Katagiri; "A Minimum-Distortion Segmentation/LVQ Hybrid Algorithm for Speech Recognition", ASJ, J. Acoust. Soc. Jpn. (E), Vol. 14, No. 1, pp. 37-42 (1993 1).

[19] R. Duda, and P. Hart; "Pattern Classification and Scene Analysis", John Wiley and Sons (1973).

[20] J. Elman, and D. Zipser; "Discovering the Hidden Structure of Speech", ASA, J. Acoust. Soc. Am., Vol. 83, pp. 1615-1626 (1988).

[21] J. Elman; "Finding Structure in Time", University of California, CRL Tech. Report 8801 (1988 4).

[22] M. Endo, S. Makino, T. Sone, and K. Kido; "Phoneme Recognition Using LVQ2", IEICE, Tech. Report SP89-50 (1989 9).

[23] K. Fu; "Sequential Methods in Pattern Recognition and Machine Learning", Academic Press (1968).

[24] K. Fukunaga; "Introduction to Statistical Pattern Recognition", Academic Press (1972).

[25] 福沢, 杉山; "ニューラルネットワークによる教師なし話者適応法とその評価", 音学会, 講演論文集, pp. 111-112 (1991 10).

[26] K. Fukuzawa, Y. Komori, H. Sawai, and M. Sugiyama; "A Segment-Based Speaker Adaptation Neural Network Applied to Continuous Speech Recognition", IEEE, Proc. ICASSP92, Vol. 1, pp. 433-436 (1992 3).

[27] 福沢, 杉山; "階層的クラスタリングと Neural Network を用いた教師なし話者適応法", 信学会, 全国大会論文集, Vol. 1, pp. 271-272 (1992 9).

[28] K. Fukuzawa, Y. Kato, M. Sugiyama; "A Fuzzy Partition Model (FPM) Neural Network Architecture for Speaker Independent Continuous Speech Recognition", Proc. ICSLP92, pp. 1383-1386 (1992 10).

[29] 福沢, 杉山; "ニューラルネットワークを利用した予備選択による大語彙単語音声認識", 音学会, 講演論文集, pp. 123-124 (1993 3).

[30] K. Funahashi; "On the Approximate Realization of Continuous Mappings by Neural Networks", Neural Networks, Vol. 2, No. 3, pp. 183-191 (1989).

[31] 古井貞熙; "ディジタル音声処理", 東海大学出版会 (1985).

[32] Y.-Q. Gao, T.-Y. Huang, D.-W. Chen; "HMM-Based Warping in Neural Networks", IEEE, Proc. ICASSP90, Vol. 1, pp. 501-504 (1990 4).

[33] S. Geman, and D. Geman; "Stochastic Relaxatioon, Gibbs Distributions, and the Bayesian Restoration of Images", IEEE, Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-6, No. 6, pp. 721-741 (1984. 11).

[34] H. Gish; "A Probabilistic Approach to the Understanding and Training of Neural Network Classifiers", IEEE, Proc. ICASSP90, Vol. 3, pp. 1361-1364 (1990 4).

[35] A. Gorin, S. Levinson, L. Miller, A. Gertner, A. Ljolje, and E. Goldman; "On Adaptive Acquisition of Language", IEEE, Proc. ICASSP90, Vol. 1, pp. 601-604 (1990 4).

[36] A. Gorin, S. Levinson, L. Miller, and A. Gertner; "On Adaptive Acquisition of Spoken Language", IEEE, Neural Networks for Signal Processing, pp. 422-431 (1991. 9).

[37] R. Gray; "Vector Quantization", IEEE, ASSP Magazine, pp. 4-29 (1984 4).

[38] J. Hampshire II, and A. Waibel; "A Novel Objective Function for Improved Phoneme Recognition Using Time-Delay Neural Networks", IEEE, Tras. NN, Vol. 1, No. 2, pp. 216-228 (1990 6).

[39] E. Hartman, J. Keeler, and J. Kowalski; "Layered Neural Networks with Gaussian Hidden Units as Universal Approximations", Neural Computation, Vol. 2, pp. 210-215 (1990).

[40] D. Howell; "The Multi-Layer Perceptron as a Discriminating Post Processor for Hidden Markov Networks", FASE, Proc. 7th FASE Symposium -Speech-, pp. 1389-1396 (1988).

[41] W. Huang, and R. Lippmann; "Comparison Between Neural Net and Conventional Classifiers", IEEE, Proc. ICNN, Vol. IV, pp. 485-493 (1987 6).

[42] W. Huang, and R. Lippmann; "HMM Speech Recognition with Neural Net Discrimination", Morgan Kaufmann, Advances in Neural Information Processing System 2, pp. 194-202 (1990).

[43] X. Huang, Y. Ariki, and M Jack; "Hidden Markov Models for Speech Recognition", Edinburg University Press (1990).

[44] X. Huang, M. Belin, F. Alleva, and M. Hwang; "Unified Stochastic Engine (USE) for Speech Recognition", IEEE, Proc. ICASSP93, Vol. 2, pp. 636-639 (1993 4).

[45] 磯, 麻生川, 吉田, 渡辺; "ニューラルネットワークによる話者適応", 音学会, 講演論文集, pp. 31-32 (1989 3).

[46] K. Iso, and T. Watanabe; "Speaker-Independent Word Recognition Using a Neural Prediction Model", IEEE, Proc. ICASSP90, pp. 441-444 (1990 4).

[47] 板倉, 斎藤; "統計的手法による音声スペクトル密度とホルマント周波数推定", 信学会, 論文誌 (A), Vol. 53, No. 1, pp. 35-42 (1970 1).

[48] H. Iwamida, S. Katagiri, E. McDermott, and Y. Tohkura; "A Hybrid Speech Recognition System Using HMMs with an LVQ-Trained Code-

book", ASJ, J. Acoust. Soc. Jpn. (E), Vol. 11, No. 5, pp. 277-286 (1990 9).

[49] H. Iwamida, S. Katagiri, and E. McDermott; "Re-Evaluation of LVQ-HMM Hybrid Algorithm", ASJ, J. Acoust. Soc. Jpn. (E), Vol. 14, No. 4, pp. 267-274 (1993 7).

[50] A. Jain, and A. Waibel; "Robust Connectionist Parsing of Spoken Language", IEEE, Proc. ICASSP90, Vol. 1, pp. 593-596 (1990 4).

[51] A. Jain, A. Waibel, and D. Touretzky; "PARSEC: A Structured Connectionist Parsing System for Spoken Language", IEEE, Proc. ICASSP92, Vol. 1, pp. 205-208 (1992 3).

[52] F. Jelinek; "Continuous Speech Recognition by Statistical Methods", IEEE, Proc. IEEE, Vol. 64, No. 4, pp. 532-556 (1976 4).

[53] F. Jelinek; "Self-Organized Language Modeling for Speech Recognition", IBM, T. J. Watson Research Center Report (1985).

[54] B.-H. Juang; "On the Hidden Markov Model and Dynamic Time Warping for Speech Recognition -A Unified View", AT&T, AT&T Bell Labs. Tech. J., Vol. 63, No. 7, pp. 1213-1243 (1984 9).

[55] B.-H. Juang, and S. Katagiri; "Discriminative Training", ASJ, J. Acoust. Soc. Jpn. (E), Vol. 13, No. 6, pp. 333-339 (1992 11).

[56] B.-H. Juang, and S. Katagiri; "Discriminative Learning for Minimum Error Classification", IEEE, Trans. SP., Vol. 40, No. 12, pp. 3043-3054 (1992 12).

[57] 片桐, 東倉, 古井; "単音節知覚における時間情報の役割", 音学会, 音学会誌, Vol. 42, No. 2, pp. 97-105 (1986 2).

[58] S. Katagiri, C.-H. Lee, and B.-H. Juang; "A Generalized Probabilistic Descent Method", ASJ, Proc. Conf., pp. 141-142 (1990 9).

[59] S. Katagiri, and C.-H. Lee; "A New HMM/LVQ Hybrid Algorithm for Speech Recognition", IEEE, Proc. GLOBECOM90, pp. 1032-1036 (1990 12).

[60] S. Katagiri, C.-H. Lee, and B.-H. Juang; "Discriminative Multi-Layer Feed-Forward Networks", IEEE, Neural Networks for Signal Processing, pp. 11-20 (1991 9).

[61] S. Katagiri, C.-H. Lee, and B.-H. Juang; "New Discriminative Training Algorithms Based on the Generalized Probabilistic Descent Method", IEEE, Neural Networks for Signal Processing, pp. 299-308 (1991 9).

[62] S. Katagiri, B.-H. Juang, and A. Biem; "Discriminative Feature Extraction", to appear in "Artificial Neural Networks with Applications in Speech and Vision", Chapman and Hall (1993).

[63] 加藤, 杉山; "ニューラルネットワークを用いた不特定話者の特徴抽出について", 音学会, 講演論文集, pp. 211-212 (1992 10).

[64] Y. Kato, M. Sugiyama; "Fuzzy Partition Models and Their Effect in Continuous Speech Recognition", IEEE, Neural Networks for Signal Processing II, pp. 111-120 (1992 10).

[65] T. Kawabata; "Generalization Effects of k-Neighbor Interpolation Training", Neural Computation, Vol. 3, No. 3, pp. 409-417 (1991).

[66] 鹿山, 阿部; "汎化能力向上を目的としたクラスタリング用ニューラルネットの学習方式", 信学会, 論文誌 D-II, Vol. J76-D-II, No. 4, pp. 863-872 (1993 4).

[67] D. Kimber, M. Bush, and G. Tajchman; "Speaker-Independent Vowel Classification Using Hidden Markov Models and LVQ2", IEEE, Proc. ICASSP90, Vol. 1, pp. 497-500 (1990 4).

[68] T. Kohonen; "Learning Vector Quantization for Pattern Recognition", Helsinki University of Technology, TKK-F-A601 (1986 11).

[69] T. Kohonen, G. Barna, and R. Chrisley; "Statistical Pattern Recognition with Neural Networks: Benchmarking Studies", IEEE, Proc. ICNN, Vol. 1, pp. 61-68 (1988 7).

[70] T. Kohonen; "The Self-Organizing Map", IEEE, Proc. IEEE, Vol. 78, No. 9, pp. 1464-1480 (1990 9).

[71] T. Kohonen; "Improved Version of Learning Vector Quantization", IEEE, Proc. IJCNN, Vol. 1, pp. 545-550 (1990).

[72] Y. Komori; "A Neural Fuzzy Training Approach for Continuous Speech Recognition Improvement", IEEE, Proc. ICASSP92, Vol. 1, pp. 405-408 (1992 3).

[73] 小森, ワイベル, 嵯峨山; "ニューラルファジィ学習法による音声認識の性能向上", 信学会, 論文誌 D-II, Vol. J75-D-II, No. 7, pp. 1101-1110 (1992 7).

[74] T. Komori, and S. Katagiri; "GPD Training of Dynamic Programming -Based Speech Recogniers", ASJ, J. Acoust. Soc. Jpn. (E), Vol. 13, No. 6, pp. 341-349 (1992 11).

[75] T. Komori, and S. Katagiri; "An Optimal Learning Method for Minimizing Spotting Errors", IEEE, Proc. ICASSP93, Vol. 2, pp. 271-274 (1993 4).

[76] K. Lari, and S. Young; "The Estimation of Stochastic Context-Free Grammars Using the Inside-Outside Algorithm", Computer Speech and Language, Vol. 4, pp. 35-56 (1990).

[77] E. Levin; "Word Recognition Using Hidden Control Neural Architecture", IEEE, Proc. ICASSP90, Vol. 1, pp. 433-436 (1990 4).

[78] W. Lincoln, and J. Skrzypeks; "Synergy of Clustering Multiple Back Propagation Networks", Morgan Kaufmann Publishers, Advances in Neural Information Processing Systems 2, pp. 650-657 (1990).

[79] Y. Linde, A. Buzo, and R. Gray; "An Algorithm for Vector Quantizer Design", IEEE, Trans. COM, Vol. COM-28, No. 1, pp. 84-95 (1980 1).

[80] R. Lippmann, and B. Gold; "Neural-Net Classifiers Useful for Speech Recognition", IEEE, Proc. ICNN, Vol. 4, pp. 417-425 (1987).

[81] J. Makhoul, S. Roucos, and H. Gish; "Vector Quantization in Speech Coding", IEEE, Proc. IEEE, Vol. 73, pp. 1551-1588 (1985).

[82] J. マーケル, A. グレイ (鈴木久喜訳); 音声の線形予測, コロナ社 (1980).

[83] 丸山, 中村, 川端, 鹿野; "HMM 音韻認識と NETgram を用いた単語音声認識", 音学会, 講演論文集, pp. 145-146 (1989 10).

[84] E. McDermott, and S. Katagiri; "Shift-Invariant, Multi-Category Phoneme Recognition Using Kohonen's LVQ2", IEEE, Proc. ICASSP89, pp. 81-84 (1989 5).

[85] E. McDermott; "LVQ3 for Phoneme Recognition", ASJ, Proc. Conf., pp. 151-152 (1990 3).

[86] E. McDermott, and S. Katagiri; "Recurrent LVQ for Phoneme Recognition", ATR Tech. Report TR-A-0115 (1991 6).

[87] E. McDermott, and S. Katagiri; "LVQ-Based Shift-Tolerant Phoneme Recognition", IEEE, Trans. SP, Vol. 39, No. 6, pp. 1398-1411 (1991 6).

[88] E. McDermott, and S. Katagiri; "Prototype-Based Discriminative Training for Various Speech Units", IEEE, Proc. ICASSP92, Vol. 1, pp. 417-420 (1992 3).

[89] E. McDermott, and S. Katagiri; "Prototype-Based MCE/GPD Training for Word Spotting and Connected Word Recognition", IEEE, Proc. ICASSP93, Vol. 2, pp. 291-294 (1993 4).

[90] 南, 沢井, 宮武; "時間遅れ神経回路網 (TDNN) による音韻スポッティング法と予測 LR パーザーを用いた大語彙単語音声認識", 信学会, 論文誌 D-II, Vol. , No. 6, pp.788-795 (1990 6).

[91] 中川聖一; "確率モデルによる音声認識", 信学会 (1988).

[92] 中川, 早川; "シーケンシャルネットワークを用いた音声認識", 信学会, 論文誌 D-II, Vol. J74-D-II, No. 9, pp. 1174-1183 (1991 9).

[93] 中村, 鹿野; "コネクショニストモデルによる単語列予測の検討", 音学会, 講演論文集, pp. 243-244 (1988 3).

[94] M. Nakamura, K. Maruyama, T. Kawabata, K. Shikano; "Neural Network Approach to Word Category Prediction for English Texts", Proc. COLING90, pp. 213-218 (1990 8).

[95] H. Ney; "The Use of a One-Stage Dynamic Programming Algorithm for Connected Word Recognition", IEEE, Trans. ASSP, Vol. ASSP-32, No. 2, pp. 263-271 (1984).

[96] L. Niles, and H. Silverman; "Combining Hidden Markov Model and Neural Network Classifier", IEEE, Proc. ICASSP90, Vol. 1, pp. 417-420 (1990 4).

[97] N. Nilsson; "The Mathematical Foundations of Learning Machines", Morgan Kaufmann Publishers (1990).

[98] 大倉, 杉山; "波形入出力による雑音抑圧ニューラルネットワークの音声認識への応用", 音学会, 講演論文集, pp. 5-6 (1990 9).

[99] 大倉, 杉山; "雑音環境下における HMM と TDNN の文節認識性能の評価", 音学会, 講演論文集, pp. 303-304 (1991 3).

[100] 大倉, 杉山; "識別誤り規準を用いた耐雑音HMMの検討", 音学会, 講演論文集, pp. 73-74 (1992 10).

[101] J. Park, and I. Sandberg; "Universal Approximation Using Radial-Basis-Function Networks", Neural Computation, Vol. 3, pp. 246-257 (1991).

[102] R. Prager, T. Harrison, and F. Fallside; "Boltzmann Machines for Speech Recognition", Computer Speech and Language, Vol. 1, pp. 3-27 (1986).

[103] L. Rabiner; "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", IEEE, Proc. IEEE, Vol. 77, No. 2, pp. 257-286 (1989 2).

[104] D. Rainton, and S. Sagayama; "Minimum Error Classification Training of HMMs -Implementation Details and Experimental Results", ASJ, J. Acoust. Soc. Jpn. (E), Vol. 13, No. 6, pp. 379-387 (1992 11).

[105] P. Ramesh, S. Katagiri, C.-H. Lee; "A New Connected Word Recognition Algorithm Based on HMM/LVQ Segmentation and LVQ Classification", IEEE, Proc. ICASSP91, pp. 113-116 (1991 5).

[106] 迫江, 千葉; "動的計画法を利用した時間正規化に基づく連続音声認識", 音学会, 音学会誌, Vol. 27, No. 9, pp. 483-490 (1971 9).

[107] H. Sakoe; "Two-Level DP-Matching - a Dynamic Programming Based Pattern Matching Algorithm for Connected Word Recognitions", IEEE, Trans. ASSP, Vol. ASSP-27, No. 6, pp. 588-595 (1979).

[108] H. Sakoe, and K. Iso; "Dynamic Neural Network - a New Speech Recognition Model Based on Dynamic Programming and Neural Network", IEICE, Tech. Report SP87-101 (1987 12).

[109] H. Sakoe, R. Isotani, K. Yoshida, K. Iso, and T. Watanabe; "Speaker-Independent Word Recognition Using Dynamic Programming Neural Networks", IEEE, Proc. ICASSP89, Vol. 1, pp. 29-32 (1989 5).

[110] 佐藤, 城, 平原; "リカレントネットによる音声ゆらぎの学習", 信学会, 技術報告 NC90-140 (1991 3).

[111] T. Sejnowski, and C. Rosenberg; "Parallel Networks That Learn to Pronounce English Text", Complex Systems, Vol. 1, pp. 145-168 (1987).

[112] H. Sorenson, and D. Alspach; "Recursive Baysian Estimation Using Gaussian Sums", Automatica, Vol. 7, pp. 465-479 (1971).

[113] F. Soong, and E.-F. Huang; "A Tree-Trellis Based Fast Search for Finding the N Best Sentence Hypotheses in Continuous Speech Recognition", IEEE, Proc. ICASSP91, Vol. 1, pp. 705-708 (1991 5).

[114] D. Specht; "Probabilistic Neural Networks", Neural Networks, Vol. 3, pp. 109-118 (1990).

[115] M. Sugiyama, and K. Kurinami; "Minimal Classification Error Optimization for a Speaker Mapping Neural Network", IEEE, Neural Networks for Signal Proceesing II, pp. 233-242 (1992 10).

[116] J. Takami, and S. Sagayama; "Phoneme Recognition by Pairwise Discriminant TDNNs", Proc. ICSLP90, pp. 677-680 (1990).

[117] 田村, ワイベル; "Neural Network を使った波形入出力による雑音抑圧", 音学会, 講演論文集, pp. 253-254 (1988 3).

[118] S. Tamura; "An Analysis of a Noise Reduction Neural Network", IEEE, Proc. ICASSP89, Vol. 3, pp. 2001-2004 (1989 5).

[119] Y. Tan, and T. Ejima; "A Network with Multipartitioning Units", IEEE, Proc. IJCNN, Vol. 2, pp. 439-442 (1990 1).

[120] D. Tank, and J. Hopfield; "Concentrating Information in Time: Analog Neural Networks with Applications to Speech Recognition", IEEE, Proc. ICNN, Vol. 4, pp. 455-468 (1987).

[121] 和田, 川人; "新しい情報量規準と Cross Validation による汎化能力の推定", 信学会, 論文誌 D-II, Vol. J74-D-II, No. 7, pp. 955-965 (1991 7).

[122] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, K. Lang; "Phoneme Recognition: Neural Networks vs. Hidden Markov Models", IEEE, Proc. ICASSP88, Vol. 1, pp. 107-110 (1988 4).

[123] A. Waibel; "Modular Construction of Time-Delay Neural Networks for Speech Recognition", Neural Computation, Vol. 1, pp. 39-46 (1989).

[124] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, K. Lang; "Phoneme Recognition Using Time-Delay Neural Networks", IEEE, Trans. ASSP, Vol. 37, No. 3, pp. 328-339 (1989 3).

[125] H. White; "Learning in Artificial Neural Networks: A Statistical Perspective", Neural Computation, Vol. 1, pp. 425-464 (1989).

[126] J. Wilpon, and L. Rabiner; "A Modified K-Means Clustering Algorithm for Use in Speaker-Independent Isolated Word Recognition", AT&T Bell Labs. Tech. Memo. (1984 1).

[127] J. Wright; "LR Parsing of Probabilistic Grammars with Input Uncertainty for Speech Recognition", Computer Speech and Language, Vol. 4, pp. 297-323 (1990).

[128] G. Yu, W. Russel, R. Schwartz, and J. Makhoul; "Discriminant Continuous Speech Recognition", IEEE, Proc. ICASSP90, Vol. 2, pp. 685-688 (1990 4).

- 表４．１　ＴＤＮＮ分類器による音素分類結果(TABLE I, [120], の内容要約).
- 表４．２　ＳＬＶＱ距離分類器による音素分類結果(TABLE II, [84]).
- 表４．３　ＨＭＭ／ＬＶＱ分類器による米語Ｅ集合音節認識結果([59] に基づく要約).
- 表４．４　ニューラルネットワークによる話者適応法を用いた単語・分節認識結果(TABLE IV, [26], の内容要約).
- 表４．５　種々の分類器の特徴.
- 表４．６　ＭＣＥ／ＧＰＤ法による米語Ｅ集合音節認識結果([61] に基づく要約).
- 表４．７　最小スポッティング誤り設計法による音声スポッティング結果(TABLE II, [72], の内容要約).

Table 4.1: TDNN 分類器による音素分類結果 (TABLE I, [120], の内容要約)

Speaker	TDNN	HMM
MAU	98.8%	92.9%
MHT	99.1%	97.2%
MNM	97.5%	90.9%

Table 4.2: SLVQ 距離分類器による音素分類結果 (TABLE II, [84])

Data set	SLVQ	K-means	TDNN
Limited	97.1%	92.4%	96.7%
Full	97.7%	91.5%	-

Table 4.3: HMM/LVQ 分類器による米語 E 集合音節認識結果 ([59] に基づく要約)

Classifier	Accuracy
HMM	61.7%
LVQ2-E/HMM	75.4%
LVQ2-L/HMM	78.9%
LVQ2-L/HMM (beginning state)	81.3%

Table 4.4: ニューラルネットワークによる話者適応法を用いた単語・分節認識結果 (TABLE IV, [26], の内容要約)

Task	Before Adaptation	After Adaptation	Speaker Dependent
500 Words	63.4%	92.2%	97.5%
278 Phrases	29.7%	57.4%	71.2%

Table 4.5: 種々の分類器の特徴

Classifier	Structure	Approach/objective	Optimization
VQ-trained distance classifier	DistC	MCD	LBG, heuristics
LVQ-trained distance classifier	DistC	DISC(MCE)	GRS
TDNN	MLP	DISC(MSE)	GRS
FPM classifier	FPM	DISC(KD)	GRS
EM-trained HMM classifier	HMM	ML	EM
MMI-based HMM classifier	HMM	DISC(MMI)	EM, GRS
MCE/GPD-trained HMM classifier	HMM	DISC(MCE)	GRS

Note: DISC (discriminant function approach), DistC (distance classifier), MCD (minimization of class distortion), MCE (minimization of classification error), MSE (target approximation using a squared error loss), KD (target approximation using the Kullback divergence), ML (maximization of class likelihood), MMI (maximization of class likelihood differences), GRS (gradient search)

Table 4.6: ＭＣＥ／ＧＰＤ法による米語Ｅ集合音節認識結果 ([61] に基づく要約)

Recognizer	Baseline	MCE/GPD
1-reference DTW	58.0%	78.4%
4-reference DTW	63.8%	84.4%
5-state, 5-mixture HMM	61.7%	-
10-state, 5-mixture HMM	66.7%	-
15-state, 5-mixture HMM	69.0%	85.7%

Table 4.7: 最小スポッティング誤り設計法による音声スポッティング結果 (TABLE II, [72], の内容要約)

Phoneme	Mis-detection		False alarm	
	Conventional	MSPE	Conventional	MSPE
t	53.9%	2.0%	0.3%	1.2%
k	24.1%	12.1%	11.9%	2.1%
mm	9.9%	4.7%	28.6%	9.5%
N	34.0%	35.9%	22.9%	10.8%
r	58.1%	21.4%	17.4%	2.6%

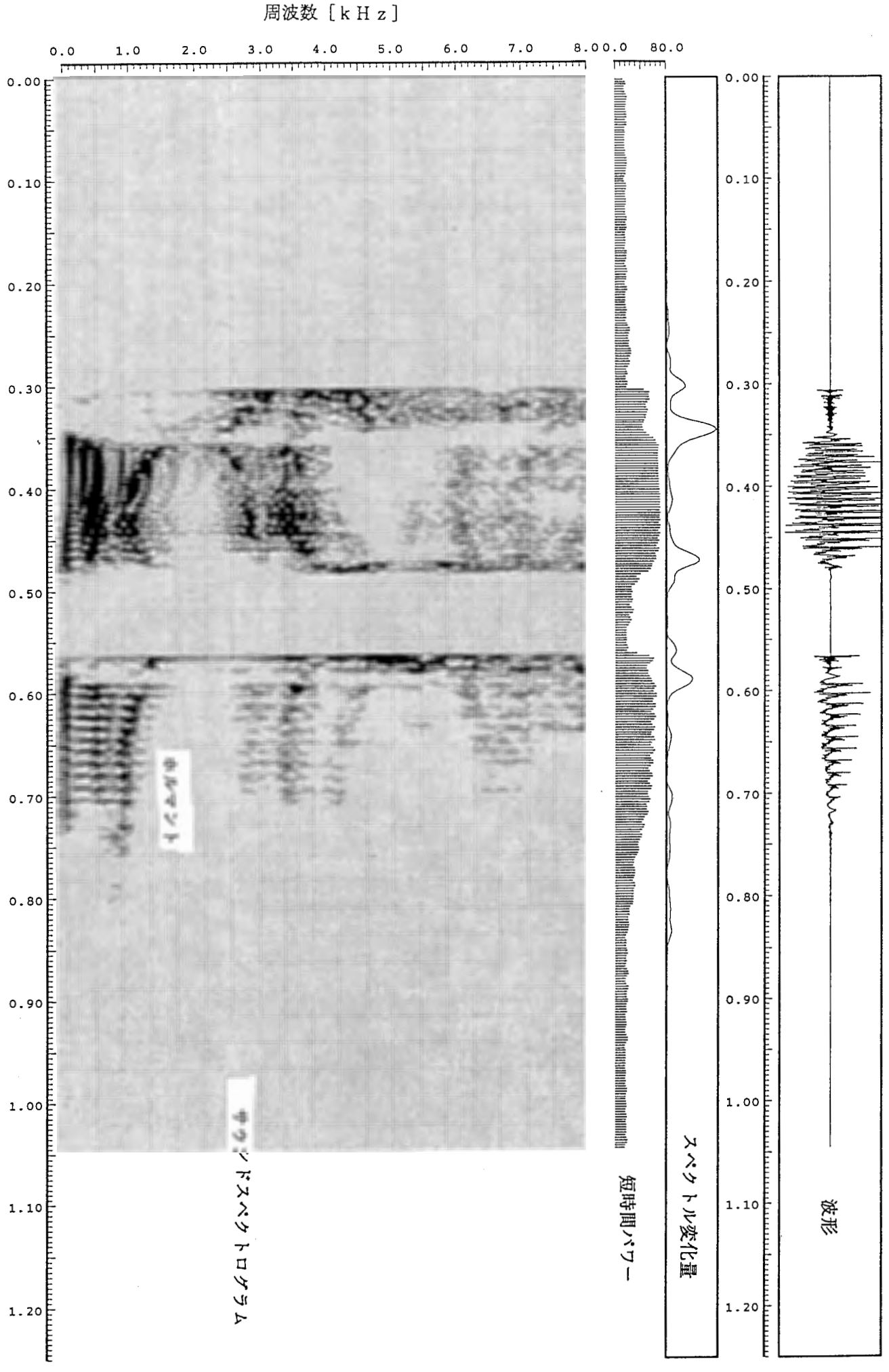

図４．１　{京都} の音声波形とサウンドスペクトログラムの１例

*** file = /data2/demo/ADDATA/KYOUTO_04.AD

sampling freq. =20.0 (KHz)　window =10.0, shift =2.5 (msec) var.max log =70.7

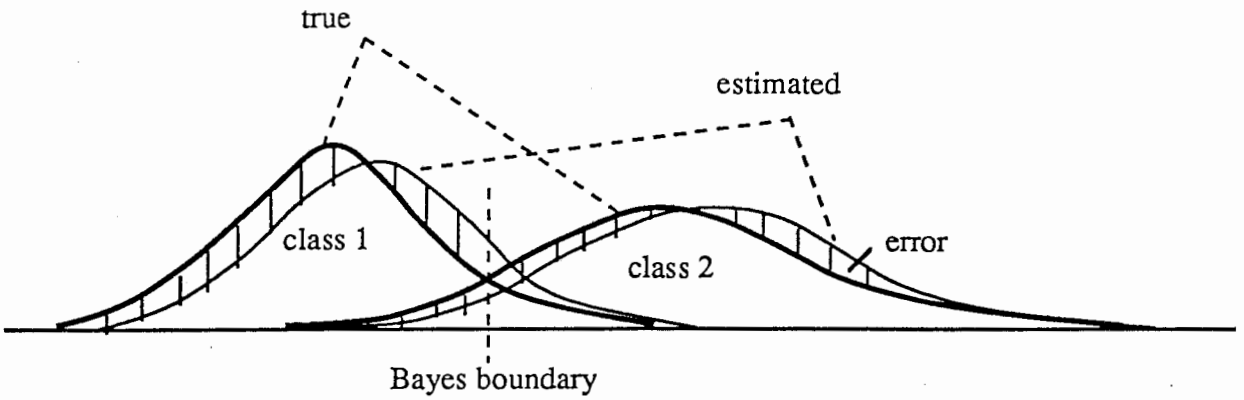

図４．２　２類の１次元標本の分類

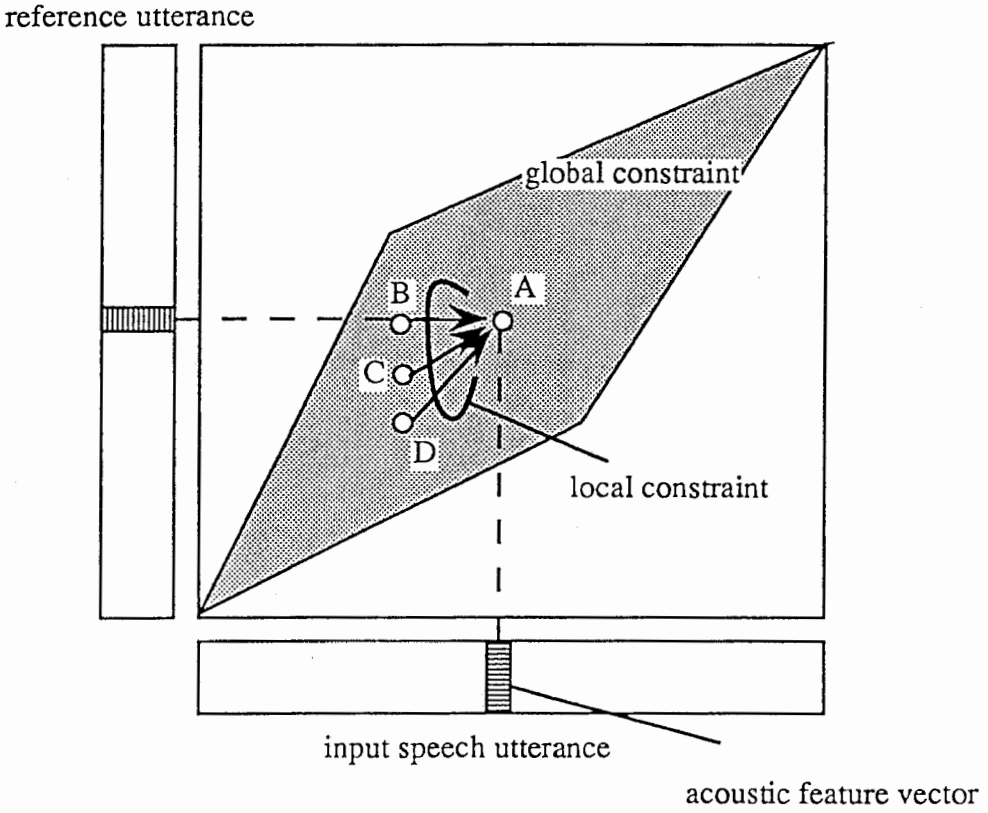

図４．３　　最適整合経路に沿った累積音響特徴ベクトル間距離（累積局所距離）の計算

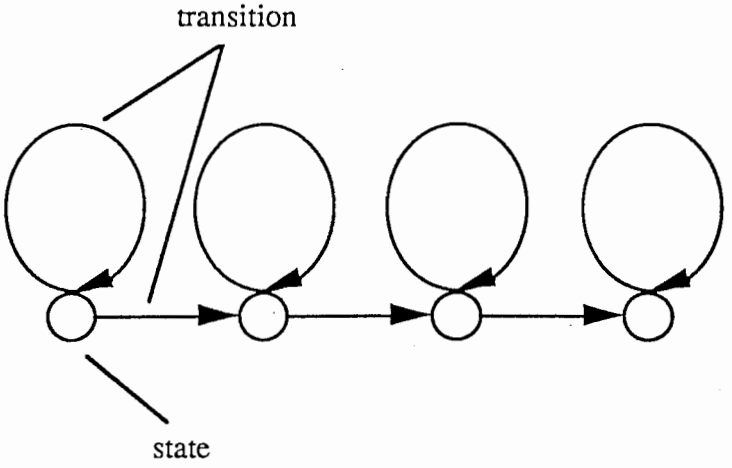

図４．４　左－右型ＨＭＭの例

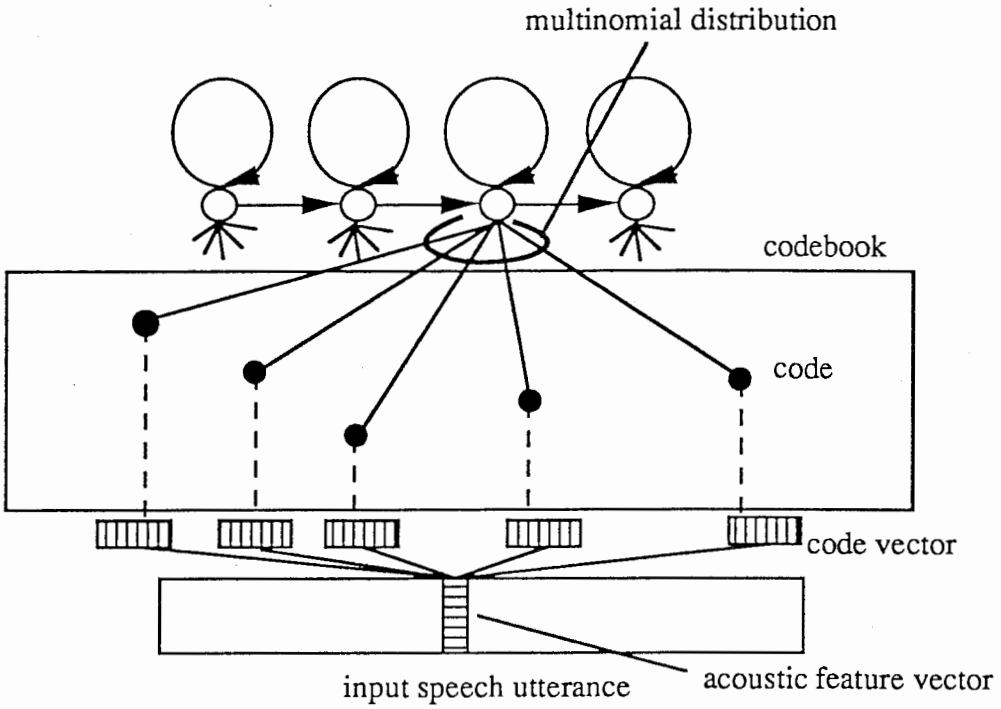

図４．５　離散型ＨＭＭの構成

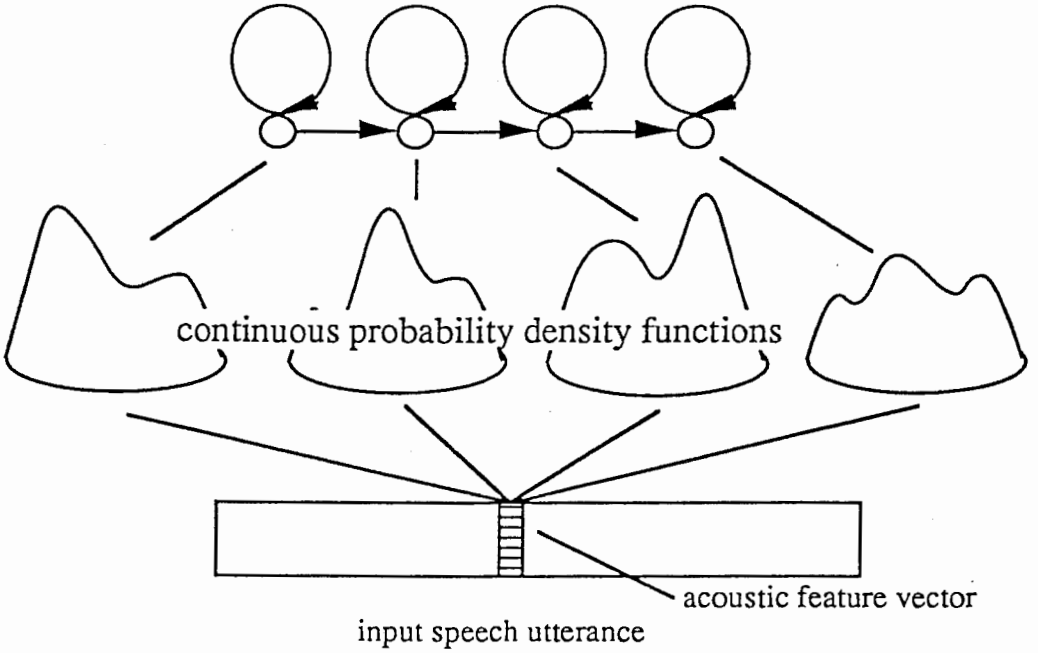

図４．６　連続型ＨＭＭの構成

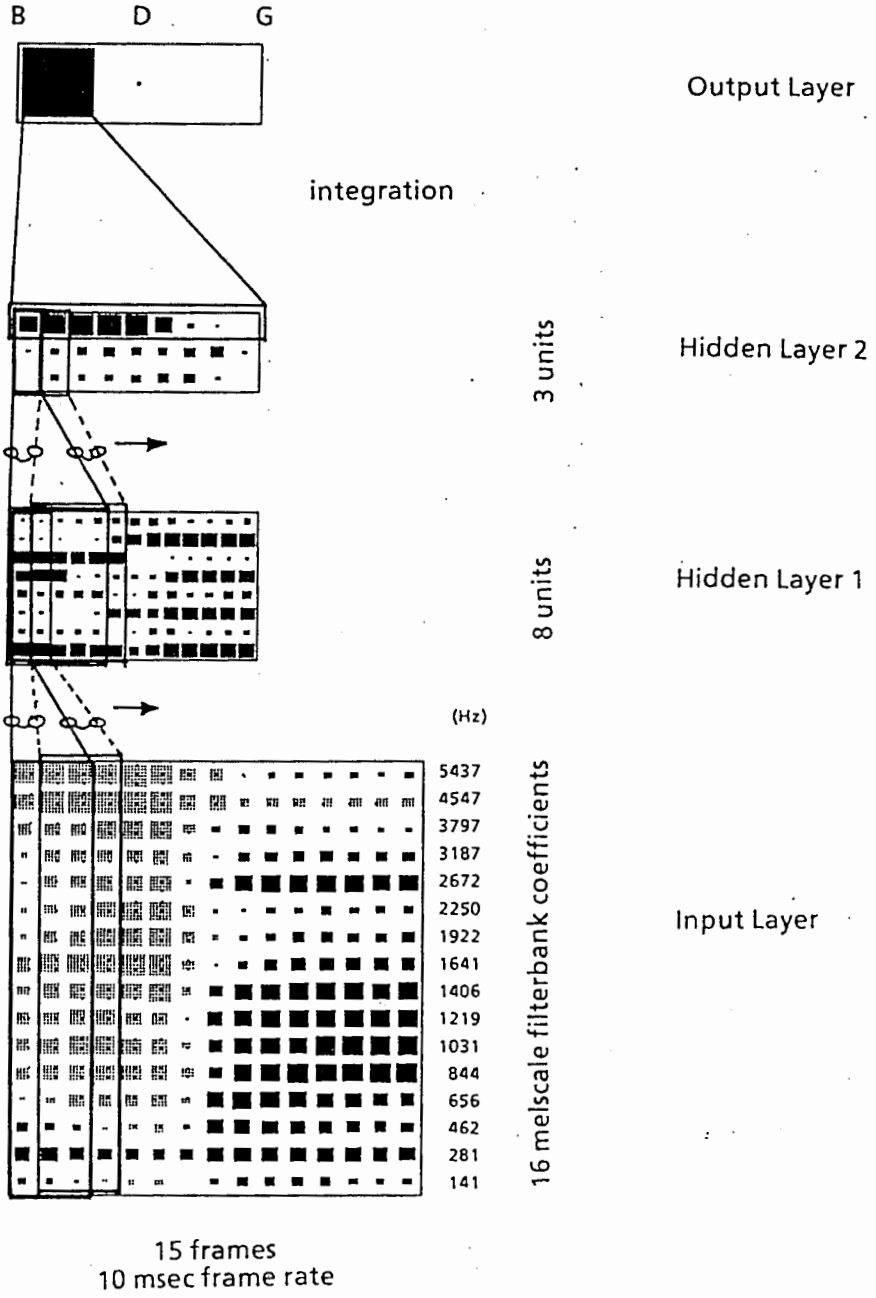

図4．7　TDNNの構造（[124]から転載）．

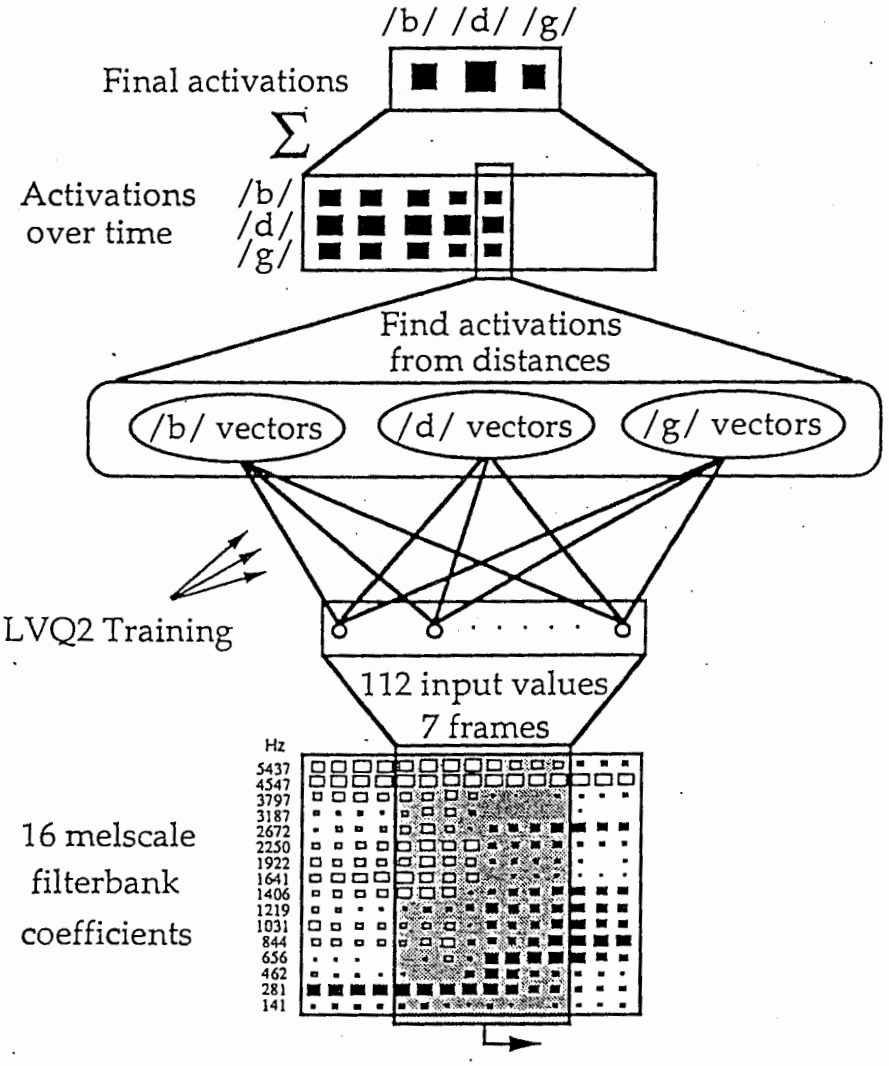

図4．8　ＳＬＶＱの構造（[87]から転載）．

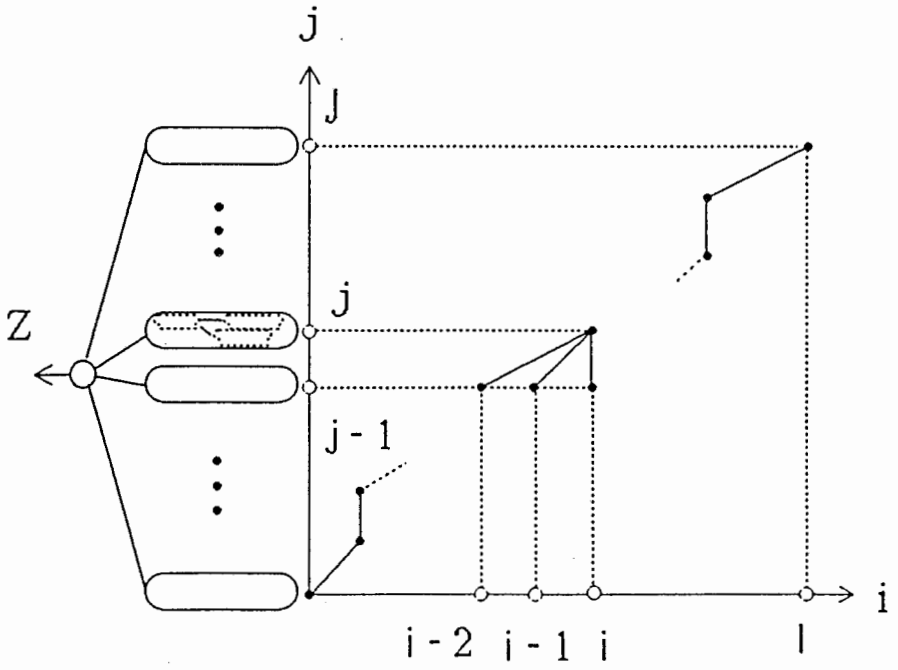

図4．9　DNNの構造（[109]から転載）．

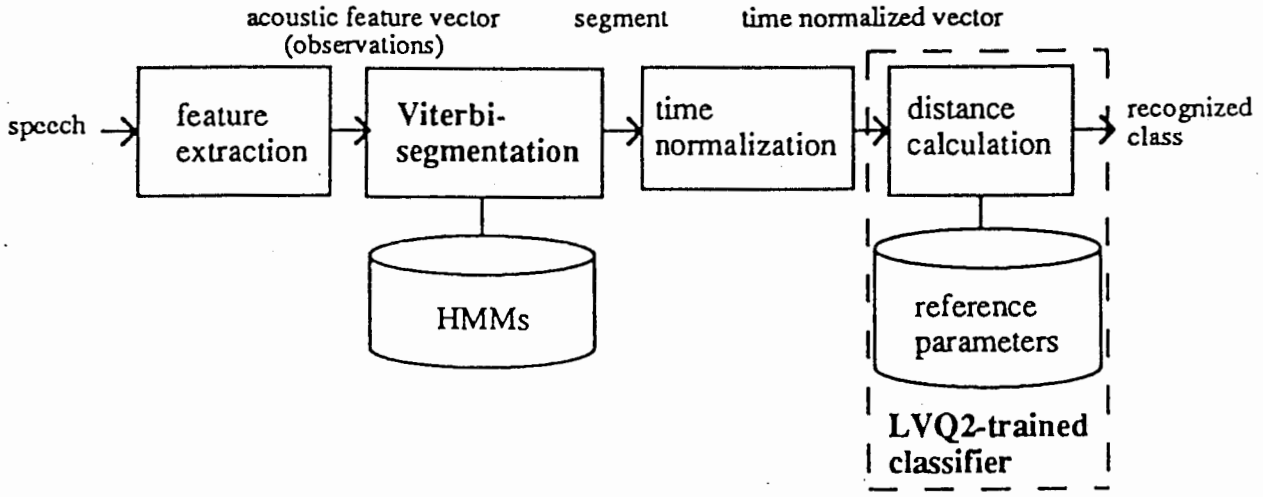

図4．１０　HMM／ＬＶＱの構造（[59]から転載）．

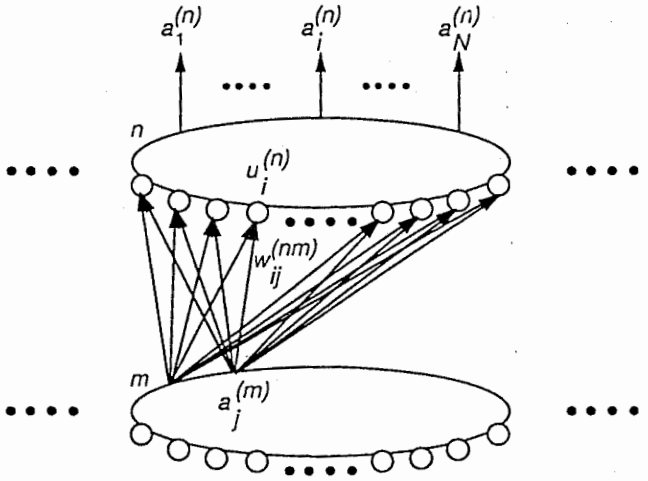

図4．１１　ＦＰＭの構造.

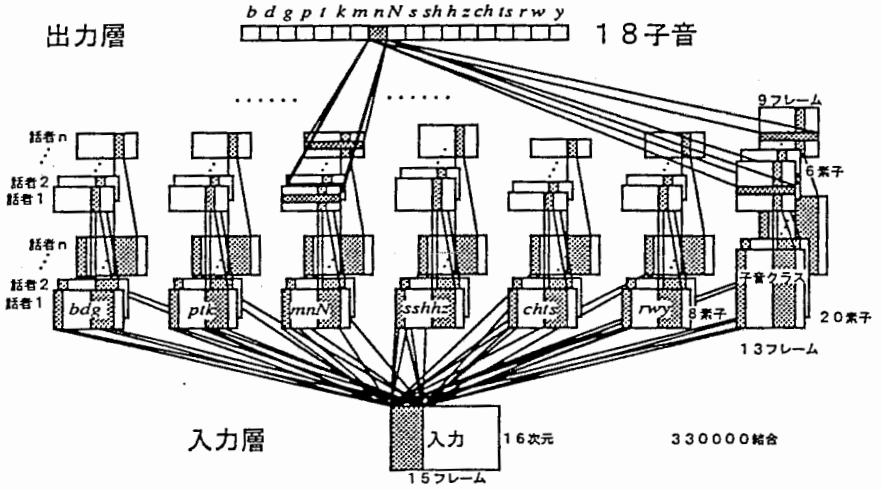

図４．１２　モジュラー構成によるＴＤＮＮ不特定話者音声認識器の構造.

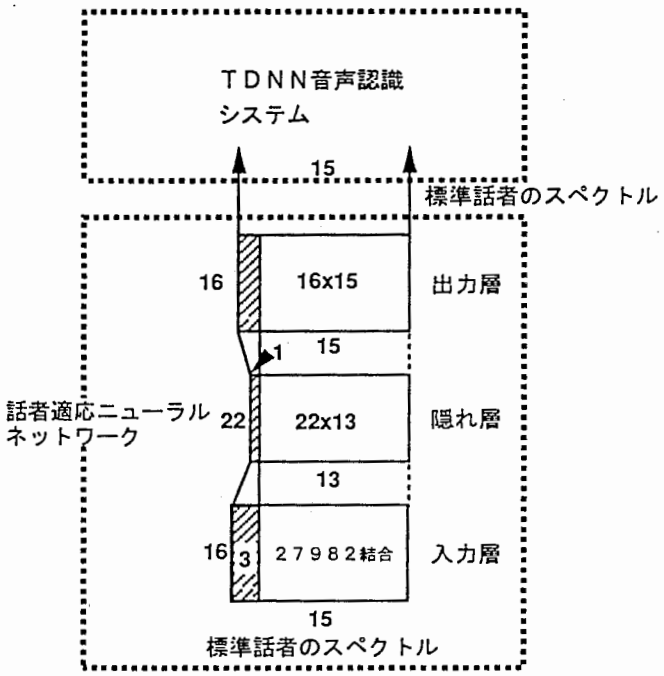

図4．13　ニューラルネットワークによる話者適応法.

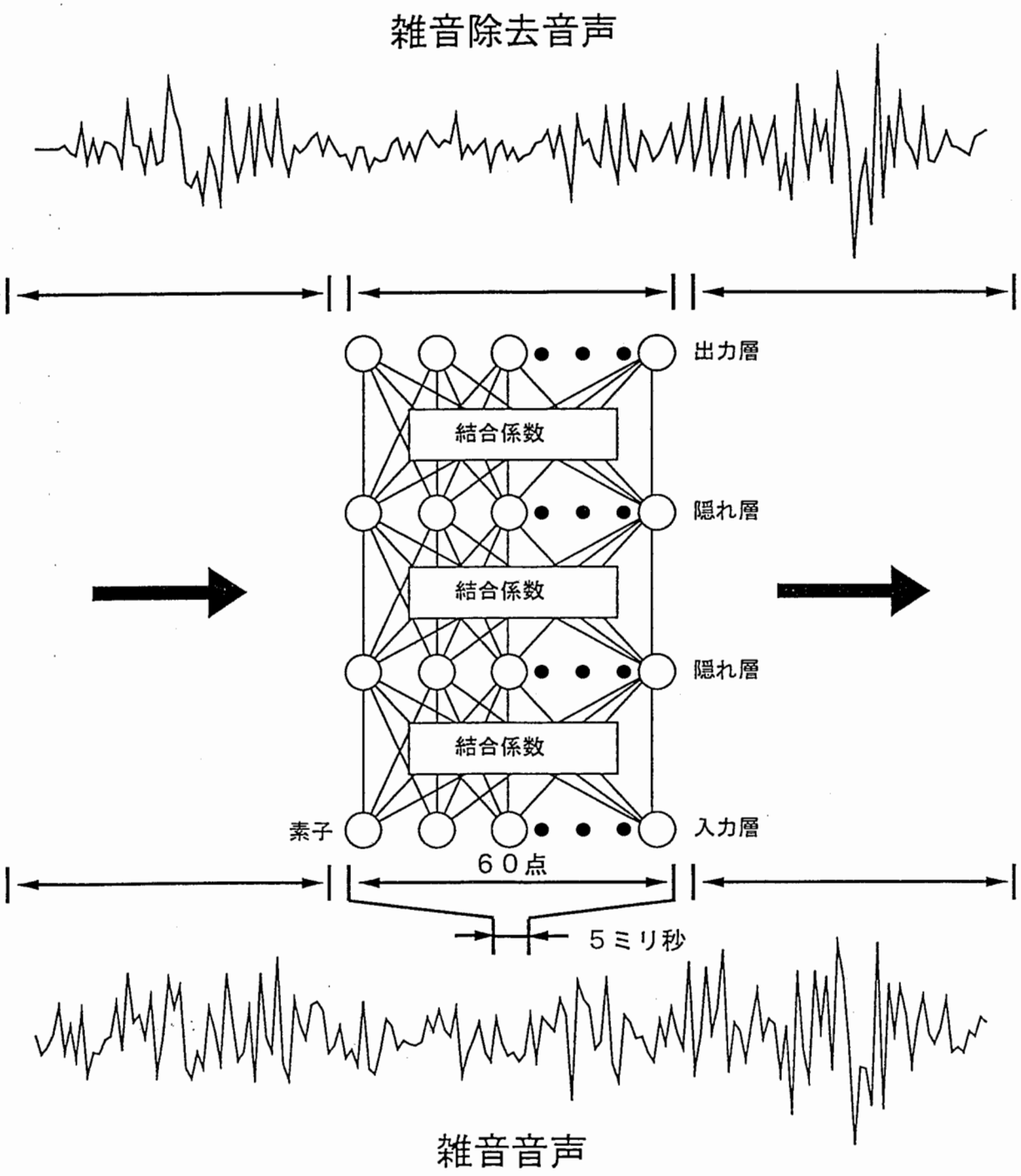

図４．１４　雑音抑圧ニューラルネットワークの構造.