氏名	川端(かわばた)　豪(たけし)
授与学位	工学博士
学位授与年月日	昭和58年3月25日
学位授与の根拠法規	学位規則第5条第1項
研究科，専攻の名称	東北大学大学院工学研究科 （博士課程）電気及通信工学専攻
学位論文題目	**認知モデルによる音素認識法に関する研究**
指導教官	東北大学教授　城戸　健一
論文審査委員	東北大学教授　城戸　健一　　東北大学教授　野口　正一 東北大学教授　木村　正行

論文内容要旨

機械との音声による会話というマンーマシンコミュニケーションの問題に対する一つの究極的な解答を実現するためには，それを構成する音声認識システムにおいて「連続発声」，「不特定話者」及び「音素を認識単位とする」という各属性が満たされている必要がある。

本論文はこれらの属性を満たす音声認識システムの実現のための基礎研究として，不特定話者による連続音声中からの音素切り出し，すなわちセグメンテーションを含む高精度音素認識法の確立を目標に進めてきた研究をまとめたものである。

本論文においては，次のような手順で音素認識のモデルの構成を行っていく。

まず，初めに特徴抽出層の形成を行う。この層では文字通り，音素を区別するために必要な音響的特徴の抽出が行われるわけであり，抽出系の生成は判別関数の設計（学習）問題に帰着させることができる。この層の概念レベルは音韻論的には弁別的要素と呼ばれるものに相当する。

次に特徴抽出層の出力に基づいて音素検出層の形成を行う。この層では抽出された特徴間の時間的相対関係から個々の音素または音素群の検出を行う。この検出系の生成もまた，筆者の提案する「判別フィルタ」の概念を利用する事によって判別関数の設計（学習）問題に帰着できる。

更に特徴抽出層の出力に基づいて出力統合層の形成を行う。特徴抽出層は音素（群）の数だけの素子を持ち，各々独立して学習が行われるために複数の素子から同時に検出出力が現れ，一意の結果が得られない事がある。そこで出力統合層では検出出力間に側抑制機構を導入し，出力を一意化

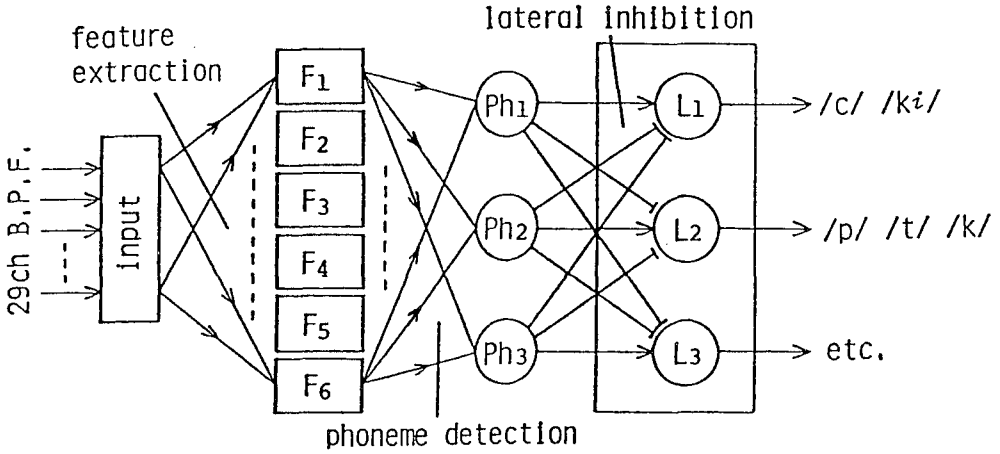

Fig. Hierarchy model of consonant recognition

する。このための側抑制係数は学習によって最適化される。

以上述べてきたように本論文では3つの学習過程を積み上げる事により階層的には音素認識モデルの構成を行っていく。

ここで注意すべきなのは，ある階層における学習の意味付けはそれに先行するすべての階層における学習の結果に従属する，すなわちある階層における学習の結果がそれ以後の階層の構造を規定しているという点である。これによって，いくつかの異なる概念レベルが有機的に結合され，生成された認識モデルの概念構造はレベル間を結ぶネットワークの重み係数として表わされることになる。

また本論文で導入するモデルは各々のレベルを音韻論的に意味のある概念と結びつけているため，得られた構造の物理的意味が失われず，各階層における非線形性素子の設計に際して確率論的な根拠を与えることが可能になる，加えて得られた結果についての視察による解釈，確認が容易である等の優れた特徴を持っている。

以下に本論文の構成について述べる。

1章は序論であり音素を単位とする音声認識手法を確立することの重要性について述べる。また高精度の音素認識を実現するための，認識系の構造自体を段階的に学習する階層形ネットワークに基づく認知モデルの導入について述べる。

2章では本論文で利用する音声資料について述べ，また音素認識向けに設計した音素データベースの構造について概説する。

3章から5章までは基礎編に当たる。これらの章では後の章において子音及び母音に対する認識モデルを組み上げるための部品となる個々の基本素子について論じて行く。

3章では特徴抽出について述べる。高精度特徴抽出のために改良された新しい判別分析の手法を

を提案し，個々の例を挙げながらその有効性を示す。この手法は入力ベクトルが多次元かつ強い相互相関をもつ状況下で，判別に有効な部分空間中でのFisher比の最大化を行うものである。またこの章では，音素の動的な特徴を捕える判別フイルタの概念を提案する。これは画像解析の分野で平滑や強調を目的としてよく用いられる空間フイルタリングを時間軸上に持ち込み，判別分析によってその最適設計を行うものである。

4章では時間領域のモデル化について述べる。抽出された特徴間の時間的相対関係に基づく音素検出機構を，特徴時系列に対する畳み込みフイルタとして実現し，またこの種の音素検出フイルタを学習によって形成する手法について述べる。特にノンパラメトリックな手法とパラメトリックな手法の併用によって，フイルタ長の決定や学習収束の問題を解決している。

5章では認識出力の統合について述べる。すなわち異なる尺度を持つ複数個の認識出力を互いに比較し，統合するためのいくつかの手法を示す。この章の各節で述べられる，事後確率較正，重み係数，及び側抑制機構による各方法は，認識モデルの各所に必要に応じて配置されている。

6章と7章は応用編である。前章までに述べてきた基本素子を組み合わせて実際に音素認識系を構成し，更に認識実験による検討を加えている。

6章では子音認識を取り扱う。無声破擦音，無声破裂音，無声摩擦音及び有声摩擦音の4音素群を取り上げ，連続発声中からの音素切り出し，すなわちセグメンテーションを含む認識系を構成し認識実験によって検討を加える。子音認識のモデルは，まず判別分析によって音素区分的な特徴量を抽出し（3章），事後確率較正に基づく正規化を行う，これは一種の非線形変換である（5章）。次にそれらの特徴の時系列に対する畳み込み演算を用いて各々の音素あるいは音素群の検出を行い（4章），各音素検出フイルタ出力間の側抑制関係によって一意の認識出力を得る（5章）ものである。

7章では母音認識を取り扱う。母音認識のモデルは，基本構造は子音のものと類似しているが，特徴抽出の代わりにスペクトル上でホルマントに対応するピークの強調を行う（7章）。このための重み係数の決定にもやはり判別分析（3章）が利用される。次にこのピーク強調を受けたスペクトルと各母音の標準パタンとの内積が計算され，これらを適当な重みづけ（5章）で比較することによって識別結果を得る。母音認識に対しては，音素中心フレームについての識別実験によって検討を加える。

8章は結論であり，本論文全体のまとめを行っている。

本論文では特に子音認識のために多くの実験，検討を重ねている。これは母音と比較して複雑かつ多様な生成系に基づく個々の子音を統一的なモデルによって取り扱うという点を重視した結果であり，このために，極めて大きい初期自由度を持った階層形ネットワーク構造と，モデル自体の構造を学習形成していく認知システムの導入を行った。

この時，前にも述べたように，形成される認識モデルにおける各階層を音韻論的に有意味な概念

と結びつけているため，学習によって得られた構造が明確な物理的意味を伴っている。

またモデル全体で音素検出系を構成しているため，子音区間の切り出しと認識が完全に一体化している。すなわちセグメンテーションと音素認識が同時進行するため，リアルタイム処理に適している。

また階層中の各素子が全く独立しているため，並列処理，パイプライン化が容易に行える等々多くの利点を持っている。

神経回路網理論的な見地からみると，本モデルは「線形変換による概念化＋非線形変換」という機構を階層化したものと考えることができる。この観点から子音認識のモデルを整理すると次のようになる。

特徴抽出	線形（判別関数）
↓	
事後確率較正	非線形
↓	
音素検出	線形（判別フイルタ）
↓	
Ｒａｍｐ関数	非線形
↓	
側抑制機構	区分的線形（線形＋非線形）

形成された子音認識モデルによって成人男女１０名によって発声された２１２単語からなる学習データに対する子音検出実験を行った結果は，無声破擦音，無声破裂音，無声摩擦音及び有声摩擦音の４音素群に対してすべて９０％以上という高い認識率を得た。また他の成人男女１０名による評価用データに対しても平均して約２％の認識率低下にとどまった。これらの値はセグメンテーションを含む音素認識率としては類例を見ない優れたものである。

また同様のモデルを用いて母音の識別系の構成を行った。特に非線形変換によるホルマントピークの強調が不特定話者の母音識別に有効であると判明した。母音識別率は成人男性５名による学習データに対して約９２％，他の成人男性５名による評価用データに対して約８７％であった。

以上のようにして本論文では，「音素」と一言で呼ばれながら，その生成機構から音響的性質に至るまで多様な特性を持つ対象の認識のために，モデルの構造学習そのものを含む認知モデルを導入し，実際にその有用性を確認した。これらの結果から本論文で提案する音素認識手法が，不特定話者の連続音声認識に有用な道具となると信ずる。また研究の過程で開発された特徴抽出，時間領域のモデル化あるいは出力統合などの各手法は，対象を音声に限らずパタン認識一般に対する有用な道具を提供するものである。

審 査 結 果 の 要 旨

不特定話者の発声する連続音声の認識は，音声自動認識の究極の目標である。その実現のためには，連続音声中の音素の認識が重要であるが，従来から種々の方法が試みられてきたにもかかわらず，この問題で見るべき成果は報告されていない。本論文は，階層形ネットワークに基づく認知モデルによって，音素および音素群の認識を行うことを試み，高い認識率を得ることができる設計法を確立した研究をまとめたもので，全編８章よりなる。

第１章は序論である。

第２章は，本研究に用いた音声資料と，音素認識向けに設計した音素データベースの構造の説明である。

第３章から第５章までは，本論文の基礎となる優れた着想の記述であり，第６章，第７章における子音および母音の認識モデルの基本素子について論じている。

第３章では，高精度の特徴抽出に用いるために判別分析の手法を改良し，音素の動的な特徴を捉える判別フィルタの概念を提案すると共に，具体例によってその有効性を示している。

第４章では，抽出された特徴間の時間的相対関係に基づく音素検出を，特徴量の時系列に対する畳み込みフィルタを学習によって形成することによって実現している。

第５章では，異なる尺度をもつ複数個の認識出力を統合するための手法を示している。

第６章では，前章までの結果を利用して設計した階層形子音認識モデルにより，無声破擦音，無声破裂音，無声摩擦音および有声摩擦音の４音素群に対して音素区間の区分けと同時に子音群認識を行う方法を述べている。これらの子音群に対しては，９０％以上の識別率という実験結果が得られ，学習によって各子音に対する構造決定を行っていくという認知モデル的手法の有効性が確認されている。

第７章では，対数スペクトル上の局所的帯域における判別分析に基づいて，ホルマントに対応するピークを強調する手法を取り入れた母音認識モデルを構成している。この有効性は，学習サンプルに対して91.7％，その他のサンプルに対して86.6％という認識率によって示されている。

第８章は結論である。

以上要するに，本論文は不特定話者の発声する連続音声中の音素を，階層形認知モデルによって認識する方法を提案し，その設計法と有効性を示し，音声自動認識の研究に新たな知見を加えたもので，情報工学並びに通信工学に寄与するところが少なくない。

よって，本論文は工学博士の学位論文として合格と認める。