

氏名	畑 岡 信 夫
授与学位	博士（工学）
学位授与年月日	平成4年6月10日
学位授与の根拠法規	学位規則第5条第2項
最終学歴	昭和53年3月 東北大学大学院工学研究科情報 工学専攻前期2年の課程修了
学位論文題目	コネクショニスト的アプローチによる音声認識の研究
論文審査委員	東北大学教授 曾根 敏夫 東北大学教授 高木 相 東北大学教授 丸岡 章 東北大学教授 竹田 宏 東北大学助教授 牧野 正三

論 文 内 容 要 旨

制約なしに話された、いわゆる spontaneous な音声を認識するためには、音声の多様性や曖昧性、及び時間的変動、そして不特定話者への対応など数々の研究課題が存在している。本論文は、音韻認識をベースとした大語彙単語音声、及び連続音声認識システムの構築を目的として、コネクショニストモデル¹的な手法を用いて上記研究課題への対処の可能性を検討した研究の成果を述べている。

音声認識の究極的な目標は、人間にとって最も自然で、かつ容易なコミュニケーションの手段である音声を使って、人間と機械とのマン・マシンインタフェースを実現することにある。従って、音声本来持っている特長を最大限に利用するためには、制約なしに自由に話された音声を認識することが不可欠な条件となっている。本論文ではコネクショニストモデル的なアプローチからこれらの課題に対処することを検討し、まず音韻認識レベルでの処理として Neural-Fuzzy 方式と統合型時間遅れニューラルネットワーク（以下、統合型TDNNsと呼ぶ）の2つのアプローチを提案する。さらに、言語レベルでの対処策として、概念ネットワークを用いた単語選択方式を提案し、音声認識の究極的な目標を達成するための要素技術を明確にする。

¹コネクショニストモデル：認知科学分野での専門用語。ニューラルネットワークモデルと呼ばれることもあるが、厳密には、ニューラルネットワークモデルを包含する概念となっている。

Neural-Fuzzy方式は、曖昧な対象を扱うファジー論理とニューラルネットワークモデルを組み合わせた方式であり、本方式を子音候補確認部として持つ大語彙単語音声認識システムを構築し、音声の曖昧さや変動に対処できる認識方式を開発した。統合型TDNNsは、音声の時間構造の違いを扱えるニューラルネットワークとしての構造であり、提案方式の個別評価を不特定話者型(Speaker-Independent)でかつ、前後の音声内容に依存しない(Context-Independent)音韻認識の枠組で行った。さらに、概念ネットワークを用いた単語選択方式は、単語間の意味的な関係を積極的に利用して、確実に関係付けられる単語系列のみを選択する方式である。以上の3種のコネクショニスト的なアプローチを用いることで、従来のパターンマッチングによる認識性能を向上させ得ることを示し、提案方式の有効性を確認した。

本論文は6章よりなっている。まず第1章で、研究の目的と意義を述べ、次に本論文に関連する技術動向として音声認識手法(特にニューラルネットワークモデルによる手法)の解説と歴史を概観している。

第2章では、コネクショニストモデルを音声認識に適用する前に、本論文における連続音韻認識システムの基本的な構成に関して検討する。ここでは、母音をまず認識した後、子音を認識するという2層構造によるパターンマッチングの枠組で、音声の定常部に着目した新しいセグメンテーションの技術である自己適合セグメンテーション(SEASEG)法の提案と、定常部と遷移部に着目した連続音韻認識手法を提案し、特定話者方式での評価結果を述べる。100単語での評価結果、自己適合セグメンテーションと、定常部と遷移部に着目した音韻認識の効果を確認した。さらに、日本語音声認識では、子音認識の向上が大きな技術課題となっていることと、音韻認識レベルでのエラーを救済するために、言語的な情報を利用することが必要となっていることを指摘し、精度の高いパターン識別を可能とするコネクショニストモデルの適用を示唆する。

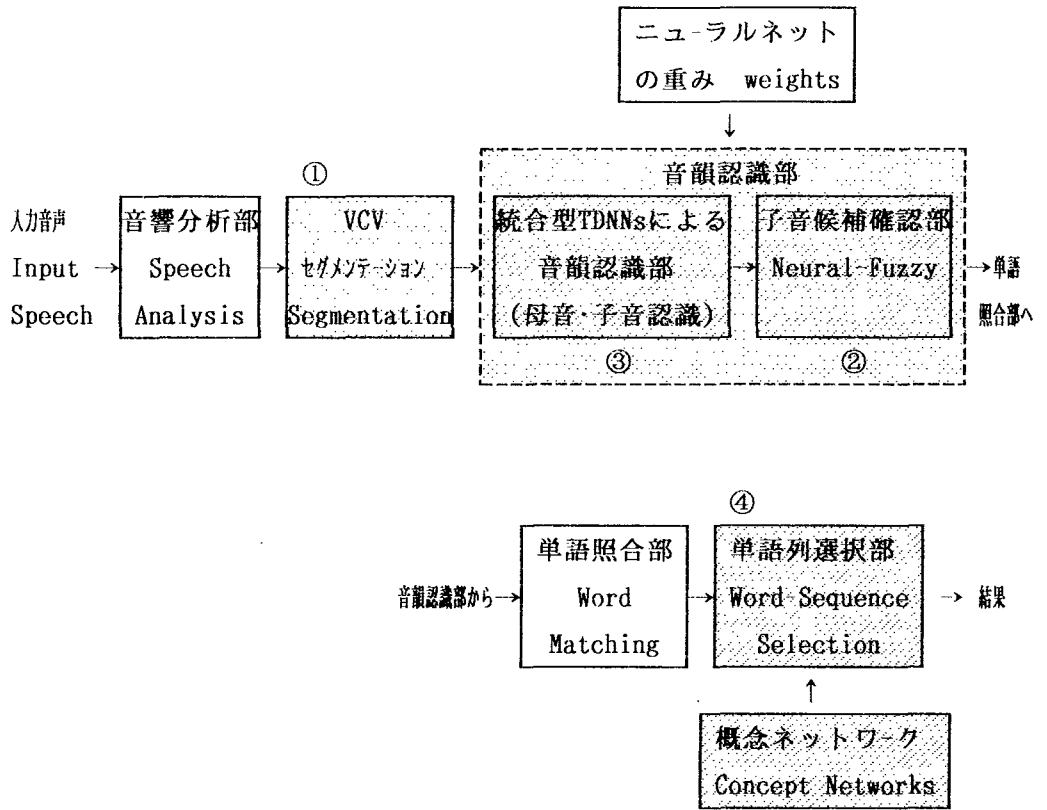
第3章では、第2章での検討結果を踏まえて、子音認識の性能を向上させる目的で、音声の多様性や曖昧性への対処として、まず音韻認識レベルからのアプローチを検討し、コネクショニストモデルを音韻認識に適用する第1の提案方式に関して述べる。即ち、音声信号の持つ曖昧さや種々の変動に対処できる音韻認識手法として、対判定の枠組でニューラルネットワークモデルとファジー論理を組み合わせたNeural-Fuzzy方式を提案する。本方式は、ニューラルネットワークを従来のパターン判定器としてではなく、各音韻を特徴付けている固有の音響的特徴の抽出器として用い、ファジー論理と組み合わせることにより、音声信号が本質的に持っている曖昧さに対処するものである。本手法の有効性を、パターンマッチング方式の音韻認識部に続く、子音候補確認部として使用したシステムで評価した。実験データは2人の話者が2回発声した100都市名の音声であり、対象としたパターンマッチングでのエラーの80%が改善されることと、各子音ごとに識別判定がきめ細かに設定できることを確認し、精度の高い音韻認識を可能とする要素技術となることを示す。このように機能は、認識論理がブラックボックスであった従来のパターンマッチングでは対応

できない処理となっている。

第4章では、第2章で報告した連続音韻認識システムのパターンマッチングによる母音と子音認識部を、ニューラルネットワークを用いた手法で置き換えることを目的として、ニューラルネットワークを音声認識へ応用する際に大きな技術課題であった音声の時間的変動への対処に関して、音韻認識レベルからのアプローチを検討する。さらに、話者形式を不特定話者型に発展させるために、不特定話者用音声データとして完備している英語の標準化音声データを使うことを試み、日本語子音以上に認識が難しい英語母音を対象として方式開発を行った。具体的には、不特定話者、かつ前後の音声内容に依存しない音声認識におけるニューラルネットワークの能力評価を行い、音韻間の持続時間長の違いに対処できる新しいニューラルネットワークの構造として、統合型TDNNsを提案する。140人の発声した480文章から取り出した、5268サンプルデータによる16個の英語母音の認識結果、単一のTDNNでの認識率56%を、統合型TDNNsにより60.5%に向上させ得ることを示す。さらに、単一のTDNNにおける過学習や一般化の問題を指摘し、提案方式である統合型TDNNsが、音韻の持続時間の違いに対処して良好なニューラルネットワークモデルを提供することを明確にする。

第5章では、第2、3、4章で検討した音韻認識レベルでの認識から、言語レベルでの認識へと方式を発展させ、コネクショニスト的なアプローチから、如何に高次の言語情報を利用して音声の多様性や曖昧性に対処するかを検討する。即ち、第2章、第3章で述べた連続音韻認識をベースにした大語彙単語音声認識において、単語が持つ概念的な意味関係を積極的に用いた候補単語の絞り込み方式を提案し、その効果に関して述べる。単語が持つ概念的な意味関係は、概念ネットワークというネットワーク構造の知識表現形式により実現されており、単語間の意味上での結合に関する情報から、意味的に関係のある単語列のみを選択する方式である。約3000語の語彙数での理論的な評価の結果、簡単な単語間の意味関係を利用するだけでも、現状の単語認識性能を大幅に改善させ得る可能性を明らかにし、これが音声の多様性や曖昧性などの課題に言語レベルから対処する要素技術となることを示す。

第6章は、本論文の結論と今後の課題に関してである。本研究の結果得られた、制約なしに自由に話された音声を認識するための要素技術を明確にし、認識システムのイメージを呈示する。図1は本論文で得られた音声認識システムのイメージである。さらに、本論文の寄与と残された将来の課題に関して整理する。



- ① 第2章での主張点 (認識方式の基本構成)
- ② 第3章での主張点 (Neural-Fuzzy方式)
- ③ 第4章での主張点 (統合型TDNNs方式)
- ④ 第5章での主張点 (概念ネット方式)

図1 連続音韻認識をベースとした大語彙単語、連続音声認識のシステムイメージ

審査結果の要旨

音声認識においては、音声の多様性や曖昧さ、時間的変動、個人性などの種々の変動要因をどう処理するかが重要な研究課題となっている。著者は、音韻認識をベースとした大語彙単語音声と連続音声認識システムの構築を目的とし、ニューラルネットワークに代表されるコネクショニスト的な手法を用いて、上記研究課題のための研究を行ってきた。本論文は、その成果をまとめたもので、全編6章からなる。

第1章は序論である。

第2章では、母音・子音・母音の系列からなる標準パターンと、母音認識と子音認識の2層構造の照合処理による認識方式を用いた連続音韻認識システムについて述べている。音声の定常部に着目したセグメンテーションと連続音韻認識方式を提案し、認識実験によってその有効性を確認している。さらに、日本語音声認識においては、精度の高い子音認識が必要であることを指摘し、コネクショニスト的アプローチが有効であることを示唆している。これは、有用な知見である。

第3章では、音声信号の持つ曖昧さや種々の変動に対応できる音韻認識手法として、ニューラルネットワークモデルとファジー論理を組合せた新方式（Neural-Fuzzy方式）を提案している。本方式は、ニューラルネットワークを、各音韻を特徴付けている固有の音響的特徴の抽出器としてとらえ、ファジー論理と組合せることにより、音声信号が本質的に持っている曖昧さに対処して、精度の高い音韻認識を可能とするものである。これは、注目すべき成果である。

第4章では、ニューラルネットワークを音声認識に応用する場合に問題となる音声の時間構造の変動への対処方法について研究し、音韻間の持続時間長の違いを吸収できる新しい統合型時間遅れニューラルネットワークを提案している。この方法によって、従来にない高い母音認識率を得ているが、これは高く評価できる。

第5章では、コネクショニスト的アプローチに基づいて、単語が持つ概念的な意味関係を積極的に用いた候補単語の絞り込み方式を提案し、それによって単語認識性能を大幅に向上できることを示している。

第6章は結論である。

以上要するに本論文は、大語彙単語音声認識や連続音声認識システムの構築のために、コネクショニスト的アプローチによる新しい音韻認識方式および言語処理方式を提案し、その有効性を明らかにしたもので、情報工学並びに音響工学に寄与するところが少なくない。

よって、本論文は博士（工学）の学位論文として合格と認める。