

氏名	新 田 恒 雄
授与学位	工学博士
学位授与年月日	昭和63年3月11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44年3月 東北大学工学部電気工学科卒業
学位論文題目	複合LPC類似尺度を用いた不特定話者音声自動認識 に関する研究
論文審査委員	東北大学教授 城戸 健一    東北大学教授 木村 正行 東北大学教授 高木 相    東北大学助教授 牧野 正三

## 論 文 内 容 要 旨

### 第1章 序 論

音声認識装置は、1973年に最初の実用装置が米国で商品化されて以来、今日まで世界で40社近くから発売されている。しかし、現実に実用化されている装置は、①孤立単語／特定話者／語彙数20～200、②孤立単語／不特定話者／語彙数16～32の二つのタイプのみで、連続単語は特定話者の、主に数字音声を入力できるものが使われているだけである。音声認識の対象を、このように狭い範囲に限定している最大の要因は、音声パターンに現れる様々な変動（調音結合による変動、話者間変動および話者内変動）である。

本研究では、話者間／話者内変動に強い統計的パターン認識法を、孤立単語と単音節音声の評価実験を通して開発した後、連続単語音声の中での調音結合と、セグメンテーションの問題に対して基礎的な検討を加え、幾つかの提案を行なっている。

第1章では、音声自動認識の研究と不特定話者音声自動認識の研究動向について概説した後、各章の構成を総括した。

### 第2章 複合類似度法の音声認識への適用

本章では、統計的パターン認識の一手法である複合類似度法を、単語および単音節音声認識へ適用した結果について述べた。

まず、単語音声認識へ適用する際の特徴ベクトルの構成方法、すなわち分析方法と時間軸正規化方法を検討

した後、単音節音声を対象に認識実験を行なった。この結果、子音の整合範囲を拡大して、連続的にマッチングを行なう方式は、性能が向上する反面、実時間処理に問題があること、一方、子音パターンの次元数を1/4に圧縮すると、認識率が10数%低下することを明らかにした。これから、連続マッチング方式を実際に適用するに際しては、パターンの次元数を低減しても識別能力の低下しない認識方式を開発する必要があることを指摘した。また、パターンを圧縮した場合でも、特性核学習を用いて話者適応を行なうことにより、高い性能が得られることを明らかにした。

### 第3章 複合LPC類似尺度

前章の結果を受け、複合類似度法を指導原理としてLPC距離尺度に対応する種々のLPC類似尺度を定義し、比較検討を行なった。

この結果、以下の結論を得た。

- ① 複合LPCケプストラム類似尺度 $S_{CP}$ 、ピーク評価複合LPC類似尺度 $S_{PE}$ および複合LPCメルケプストラム類似尺度 $S_{MC}$ の認識性能が良く、同じ次元数で比較したBPF分析による結果と比べても優れている。
- ② 音声スペクトルに直接ピーク重み付けする方法は、複合類似度法の場合有効ではない。
- ③ 周波数軸を伸縮することにより特徴パラメータの次元数を減らすことができる。また、伸縮特性は聴覚に近似（メル尺度化）するとき、最も良い結果を示す。複合LPCメルケプストラム類似尺度 $S_{MC}$ は、次元数を8に低減しても安定して成績が良好である。

$S_{MC}$ は次元数を低減できるため、パターンマッチング時の演算量が大幅に削減できる。このため、第2章で述べた単音節音声認識や連続音声認識など、時間軸に沿って連続的にマッチング計算を行なう必要がある場合には、この類似尺度は特に有効となる。

### 第4章 複合LPC類似尺度の電話単語音声認識への応用

前章で検討した複合LPC類似尺度を適用して、電話音声認識の性能向上を試みた。同時に、単語境界の検出方法にも検討を加え、単語境界を尤度順に複数提示し、この単語境界仮説を検証する方式を提案するとともに、その有効性を電話単語音声認識実験で実証した。

まず、単語境界の仮説生成では、始端と終端の候補を探索し、単語境界を尤度順に複数提示する方式を検討し、仮説生成にあたっては次の二つの機能が大切であることを明らかにした。

- ① 確実に音声と認められる区間を検出する機能
  - ② 音声か雑音かを推定し始端（終端）としての信頼度を与える機能
- ①からは境界候補を削減することが可能になる。また、②により単語境界の尤度を計算し、候補の順位を決定することができる。

複合LPC類似尺度による評価実験では、単語境界仮説から提示される複数の標本パターン（最大4個）に対して類似度を計算し、そのうち最大値を与えるカテゴリを正解とした。この結果、次の結論を得た。

- ① 三つの尺度（複合LPCケプストラム類似尺度 $S_{CP}$ 、ピーク評価複合LPC類似尺度 $S_{PE}$ およ

び複合 LPCメルケプストラム類似尺度  $S_{MC}$ ) とも、同じ次元数で比較した BPF 分析による結果と比べて優れた性能を示す。また、この中ではピーク評価複合 LPC 類似尺度  $S_{PE}$  が最も優れている。

- ② 特徴パラメータの次元数を16から8に低減すると、複合 LPCメルケプストラム類似尺度  $S_{MC}$  が最も劣化が少ない。

これらの結論は、第3章の広帯域音声を対象とした際の結論と一致した。また、単語境界仮説と複合 LPC 類似尺度を組み合わせることにより、精度の高い電話音声認識を実現できることが確かめられた。

## 第5章 複合 LPC 類似尺度を用いた連続単語認識

本章では、第2章、第3章の検討結果をふまえ、連続単語音声認識システムを構成している。

一定のセグメントを識別の単位とし、入力音声をこのセグメント単位で認識した後、単語照合を行なう方式は、語彙の拡張と演算量の点から有利な方式といえるが、セグメンテーション（入力音声のセグメント・ラティス上に対応のつかないセグメントが湧き出し、これが単語照合に悪影響を与える）と、個々のセグメントの識別精度に大きな問題が残る。この二つの課題といま一つの重要な問題である、単語列の評価方法（スコアリング）を取り上げ検討を加えた。また、この検討結果に基づく連続単語認識方式を提案し、連続数字音声を対象として評価実験を行なった。

提案した方式は、入力音声を音韻構造の異なる種々の音声セグメント単位（音素・CV・VC・VCV）に識別した後、単語遷移ネットワークに通して単語照合を行なう。単語列の評価は、このネットワークで受理された単語に対して、その主要なパスに対応する音声セグメントのスコアを選択的に累積して評価している。この方式の中で用いた方策を、前述の三つの課題に対応させて以下にまとめる。

(A) セグメンテーションに関する問題に対して：

- ① 湧き出しを生じやすい過渡区間に VC, VQ セグメントなどのセグメント名を割り当て、これを遷移ネットワーク上に表現することにより、マッチングのパスを安定化する
- ② 定常母音も一定の時間幅をもって切り出すなど、セグメントの切り出し方法を工夫して不要な湧き出しを抑える
- ③ 湧き出しを生じた場合にも、一定の回数は無条件に次のフレームにスキップできる手段を設けるなど、単語遷移ネットワーク上でこれを吸収する工夫を加える

(B) 音声セグメントの識別精度の向上に対して：

- ① 調音結合による変動に対しては、音韻構造の異なる種々のセグメント（音素・CV・VC・VCV）を、識別の単位として連続音声の中から切り出し、音素としてのみ表現しうるパターンや、VCVの枠でのみ表現できるパターンに対処する
- ② 話者間・話者内変動に対しては、各音声セグメントを特徴パラメータと時間の二次元パターンとして表現し、これに複合 LPC 類似尺度を適用して、変動を吸収する。

(C) 単語列の評価方法に対して：

- ① 単語中の音声セグメントの類似度（の最大値）を，類似度－スコア変換式（または最大類似度を用いた変換式）によりスコアに変換し，加算が可能な形にする
- ② 単語（列）のスコア計算は，ネットワークで受理された単語について，その主要なパスに対応する音声セグメントのスコアを選択的に累積して評価する。

評価実験では，登録内話者による評価実験と，未登録話者（不特定話者）に対する実験を行ない，本方式が有効であることを確認した。

## 第 6 章 結 言

不特定話者音声自動認識にとって大きな障害となる，話者間／話者内変動および調音結合による変動に検討を加え，これらの変動を吸収することを試みた。話者間／話者内変動に対して，複合 LPC 類似尺度を開発した後，調音結合の問題に対して，音韻構造の異なる種々のセグメントを単位とする単語遷移ネットワークと，選択的スコアリング法による単語（列）評価を組み合わせた方法を提案し，その有効なことを示した。これらの方式が，今後，連続音声認識の研究を進める上で，貢献することを期待したい。

## 審 査 結 果 の 要 旨

音声は人間にとってもっと容易な情報伝達手段であるために、高度なマン・マシンインターフェイスを目指した研究の一環として、音声認識の研究が活発に行われている。それに用いられている諸方式の内、単語を認識の基本単位とする方式は、調音結合の問題を回避することが出来て最も実用化に近いものと考えられるが、不特定話者への拡張には問題があった。著者はこの問題を、文字認識で提案されている複合類似度法の応用により解決しようとして研究し、多くの新たな知見を得た。本論文はその結果をまとめたもので、全文6章からなる。

第1章は序論である。

第2章では、複合類似度法を適用した話者登録単音節認識実験により、音声分析の条件と時間軸正規化の方法を検討した上で、多数の単語音声資料を用いて、登録話者数の検討を行っている。さらに、不特定話者の単音節認識実験を行い、問題点を示している。

第3章では、不特定話者音声認識のための種々の複合LPC類似尺度を論じ、LPC距離尺度との比較評価実験を行い、さらに聴覚特性と対応したLPC分析を行って、特徴パラメータの次元数低減を検討している。

第4章では、前章で検討した複合LPC類似尺度を適用して、電話音声認識の性能向上を行うとともに、尤度順に複数提示した単語境界候補から、仮説検証によって境界を決定する方法を提案し、その有効性を単語音声認識実験で実証している。

第5章では、前2章の検討結果をふまえ、連続単語音声認識システムを構成している。このシステムでは、音韻構造の異なる各音声セグメントを特徴パラメータと時間の2次元パターンとして表現した後、種々の変動を吸収するため、これに複合LPC類似尺度を適用している。さらに、単語列の評価方法として、単語遷移ネットワークで受理された単語中の主要なパスに対応するセグメントの評価値を選択的に累積する方法を提案し、連続数字音声を対象として評価実験を行い、有効性を確認している。

第6章は結論である。

以上要するに本論文は、不特定話者音声の自動認識を目的として音声分析の方法を検討し、さらに複合LPC類似尺度を用いる認識システムを構成して実験的な検討を加え、音声自動認識に有用な多くの知見を与えたもので、情報工学の進歩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。