氏名（本籍）	小坂哲夫（神奈川県）
学位の種類	博士（情報科学）
学位記番号	情第3号
学位授与年月日	平成9年4月10日
学位授与の要件	学位規則第4条第2項該当
研究科，専攻	東北大学大学院情報科学研究科（博士課程）情報基礎科学専攻
学位論文題目	個人差を考慮した連続音声認識に関する研究
論文審査委員	（主査） 東北大学教授 牧野正三　東北大学教授 曽根敏夫 東北大学教授 根元義章

論文内容要旨

第一章 序論

音声認識研究は近年極めて盛んになってきている。コンピュータの普及に伴い，広く一般のユーザーがコンピュータを用いるようになってきた。このためコンピュータの利用を容易にするための，マン・マシーンインタフェースの研究開発が重要になっている。本論文では音声認識において解決すべき重要な問題の一つである話者性の問題の解決を図った。話者変動への対処としてはおおまかに分けて2通り存在する。一つは不特定話者音声認識の性能そのものを向上させることであり，一つは話者適応により，適応前は性能の悪い認識システムを，使用していくにつれその話者の個人性を学習し性能を向上させる方法である。本論文は以上の二つの方法の融合によって不特定話者に対する音声認識の性能向上を目指した。二つの方法の融合による新たな音声認識法では，システムはユーザーから見た場合不特定話者音声認識として動作するが，内部では話者適応アルゴリズムが動作する。提案するアルゴリズムは以下の通りである。第一ステップでは，入力された音声を不特定話者音素モデルで認識し，その認識結果をもとに第二ステップで教師なし話者適応を行なう。第三ステップでは適応後のモデルにもとづいて再度その入力音声を認識する（図1参照）。このように認識過程の途中で話者適応することにより，不特定話者音声認識の高性能化が実現された。

第二章 不特定話者音声認識用混合連続分布HMMの精度向上の検討

本章は第一ステップの不特定話者音声認識の性能向上を図ることを目的とする。この目的のため話者適応の初期モデルとして用いる不特定話者モデルについて検討を行ないその性能向上を目指した。本研究は混合分布型のHMMに基礎をおく。このような確率モデルではパラメータ推定のアルゴリズムは統計論にベースをおいてアルゴリズムが確立しているが，状態遷移のしかたや状態数などモデルの構造に関する設計の指標の検討は遅れている。モデルの構造で特に問題となるのは状態遷移のしかた（HMMのトポロジー）と混合数である。HMMのトポロジーに関しては従来いくつかの設計法が提案されているが，混合数の法定法については有効な方法は提案されていない。そこで本章では，混合連続型HMMで問題になる混合数の自動決定法を提案した。そのために，まずHMMの混合数を変化させた場合のHMMの対数ゆう度および各分布の共分散行列の行列式の値について検討した。この結果学習データ量が少ない場合過学習が起こり，その影響は行列式に現れることが分かった。以上の検討結果に基づきHMMの状態の分散の大きさにもとづいて混合数を決定するアルゴリズムを提案した。提案手法の有効性を検証するため不特定話者音素認識実験を行なった。この結果いずれの混合数の場合も，音素モデルごとや状態ごとに混合数を決定したほうが認識率が向上することが分か

り，本アルゴリズムの有効性が認識できた。また音素ごと混合数を決定する場合と，状態ごと決定する場合の比較を行なったところ，状態ごと決定する方法でより高い認識率が得られることが分かった。

第三章　話者混合法により不特定話者音声認識および話者適応

本章では，第二ステップにおける話者適応法の性能向上を目指した。不特定話者モデルの検討のうち，音素環境依存モデルの効率的な作成法「話者混合法」について提案した。不特定話者の文節音声認識で認識実験を行なった結果，環境非依存のHMM（3状態，20混合）で75.8%の認識率だったのに対し，600状態，12混合の音素環境依存モデルで82.8%の認識率が得られ有効であることが分かった。また話者混合法に話者クラスタリング法を取り入れることにより任意の混合数でモデルを合成できるCCL法を提案した。その結果従来のBaum-Welch学習に比較してモデルの性能を損なうことなく短時間でモデルが作成できることを示した。計算時間を比較すると混合数5の場合で約1／20，混合数15の場合で約1／60で，この差は混合数が増すほど増加する。また混合数5で比較した場合，クラスタリングにより初期値を与えたのちBaum-Welchで学習する方法で認識率が64.6%だったのに対し，CCLでは73.6%と高い認識率が得られた。さらにCCLで作成したモデルを初期値としてBaum-Welch学習することにより認識率として77.2%が得られ，高性能なモデルを得る場合の初期値としても有効であることが分かった。さらに以上で検討した話者混合法により作成されたHMMを初期モデルとして，極く少量の適応データで動作する話者適応法について検討を行なった。提案した話者重み学習法では話者ラベルの付与されたHMMの各混合分布において，同一のラベルがふられているものを「結び」の関係として，話者重みを学習する方法である。さらに重みがある閾値以下になった場合その混合分布を削除し認識時の計算量の削減を行なう話者プルーニングについても検討を行なった。その結果5単語で適応した場合，認識率が75.8%から80.5%に向上し，非常に少ないデータでも話者適応が可能であることが分かった。さらに話者プルーニングを行なうことにより，認識率の低下なしに混合数を1／2～1／12程度に削減できることを示した。次に，少量データで動作する話者適応での成果をもとに，少量データでも，データ数が増加した場合でも，データ量に応じて動作する。データ量に応じた話者適応について検討した。ここではこれを実現する方法として「複数の話者適応法に基づく動的話者適応」を提案した。実験の結果データ数が増加するに従い自由パラメータの少ない適応法からパラメータの多い適応法へと自動切替えが行なわれることが確認され，複数の話者適応法の自動切替えに対する見通しが得られた。

第四章　木構造話者クラスタリングによる話者適応

本章では，第二ステップにおける少量の適応データで動作する話者適応をさらに性能向上させる手法として，木構造話者クラスタリングによる話者適応法を提案した。これまで，話者または話者クラスタの選択に基づく話者適応手法が検討されてきている。このような話者適応では，話者（クラスタ）選択のみにより適応を行なうため，パラメータを調整する他の話者適応法と比較して，非常に少ないデータでの適応が可能である。このような話者選択による話者適応では，入力音声の話者の特徴が，あらかじめ用意された標準話者の特徴とは類似していない場合，性能が低下するという欠点がある。そこで解決策として，単純に標準話者の数を増加させるという方法も考えられる。しかし，話者の増加に伴い選択すべき話者クラス数が増加し現実的とは言えない。また男女という明示的な特徴でクラスを分け，男女識別の後に認識を行なう方法も提案されている。しかしこの方法でも，(1)クラスの分類がヒューリスティックな知識によっているし，(2)クラス数の決定もヒューリスティックに行なわれている，という問題がある。以上の問題を解決するために，本章では木構造話者クラスタリングによる話者適応を提案した。この方法では図2に示すように，話者特性を階層的に逐次分割することにより，話者モデルの木構造を作成する。木構造では話者特性を表わすことにより，木構造の上層では話者特性の大局的な特徴，例えば男女の差などを表現するモデルが作成できる。また下層では局所的な特徴を表現するモデルを得ることが期待できる。提案する方法では自動的にクラスタリングを行なうため，まず(1)のクラス設定の問題が解決できる。また何らかの方法で，木構造の階層のうち入力音声の識別に最適な階層を選択することにより(2)のクラス数に関連する，クラスの分布の大きさの決定に関する問題が解決できる。つまり話者クラスの分布の大きさの決定が自動的に行なわれることになる。以上の2点が解決でき，更に話者数が増加した場合問題となる計算時間に関して，話者選択に要する計算量の増大を防ぐという効果も得られる。これは話者モデルを木構造で表現することにより，全話者のモデルと照合する場合に比べ，話者モデルの照合の回数が減るためである。評価実験では170名の話者を用いて木

構造を構成し話者適応を行なった。SSS-LR により文節音声認識実験を行なったところ適応文節数 5 の場合で，74.2%から79.9%へ認識率が向上（22.1%の誤り率の減少）した。また一番効果のある話者で見ると 74.3%から 85.1%へ向上（42.0%の誤り率の減少）し本手法が有効であることが分かった。

第五章　教師なし話者適応を利用した不特定話者音声認識

本章では以上の各章の成果を踏まえ，話者適応と不特定話者認識の融合による，新たな不特定話者音声認識手法の提案を行なった。一般に不特定話者音声認識では，音素モデルとして，不特定話者用に作成した音素モデルを使用する。不特定話者の認識性能を低下させないために，HMM のような統計モデルを使い，不特定話者の音声の分布を考慮したモデルを作成するのが普通である。このため音響的に平均に近い話者の音声は良好に認識できるが，平均からはずれた話者の声の認識は難しい。実際 HMM などで不特定話者の音声を認識した場合，平均認識率は悪くなくとも，極端に認識率の低い話者が存在する場合が多い，実際の音声では，1 発声は 1 人の話者によりなされるが，従来の不特定話者音声認識方法ではこの拘束は利用されない。よって，どれだけ多数の話者の音声を集めても，原理的な限界があった。本章では，この原理的限界を越えることを目指し，1 発声は 1 人の話者によりなされるという拘束を用いた不特定話者音声認識方法を提案した。これを実現するためには，話者適応法と不特定話者音声認識法の融合が必要である。この方法の原理は，システムは不特定話者音声認識として動作するが，内部では話者適応アルゴリズムが動作する。入力された 1 発声を用い，そのデータで教師なし話者適応を行ない，変更されたモデルにもとづいて，再度入力音声を認識する。このように認識過程の途中で話者適応することにより，不特定話者音声認識の高性能化が見込まれる。以上を実現するためには，1 発声程度の極く少量のデータで教師なし話者適応が出来る必要がある。本章では具体的な手段として，木構造話者クラスタリング法の教師なし話者適応への応用を検討し，次に教師なし話者適応と適応後の認識という 2 パスアルゴリズムにより不特定話者音声認識を行なう方法について検討した。認識実験の結果，まず木構造話者クラスタリングを用いた話者適応では，教師なしでも教師つきと同程度の話者適応が可能であることが分かった。さらに話者適応と認識という 2 パスによる不特定話者音声認識法が従来法より高い性能が得られることが分かった。

第六章　結　　論

本章は結論であり，本研究の成果を要約した。

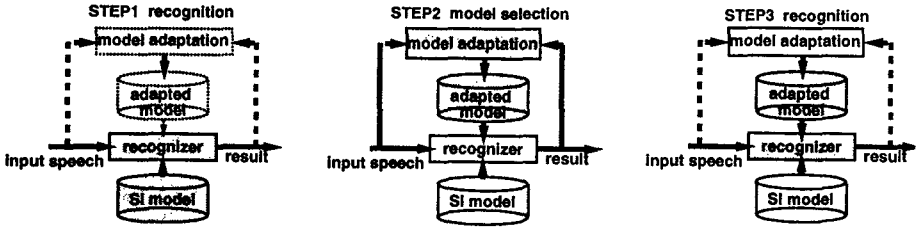

Figure 1 ：話者適応技術と不特定話者音声認識技術の融合による音声認識の流れ図

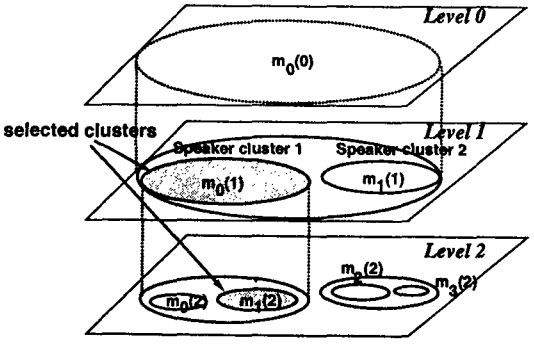

Figure 2 ：木構造話者クラスタリングの説明図

審 査 結 果 の 要 旨

情報機器のマン・マシンインターフェイスとして音声認識の重要性が増すとともに音声の個人差への対処が重要な課題となっている。音声の個人差への対処法としては，多数の話者からの音声データを基に一つの不特定話者用音素モデルを作成する不特定話者音声認識法と，一人の話者の音声データから作成した特定話者用音素モデルを少量データによって新しい話者に適応させる話者適応法があるが，精度や計算量に関して問題があった。著者は，個別に発展してきた不特定話者音声認識法と話者適応法を融合した新しい音声認識方法を提案し，精度の改善や計算量の削減を実現した。本論文はその成果をまとめたものであり，全編6章よりなる。

第1章は序論で，本研究の背景と目的を述べている。

第2章では，混合連続分布で表される不特定話者用音素モデルをとりあげ，その分布の混合数は，分布の分散の行列式を基準として自動決定できることを示し，それが音声認識システムの高性能化に有効なことを実証している。

第3章では，特定話者用音素モデルを基に，話者適応法によって各話者の個人性を反映した音素モデルを作成し，それらを重ね合わせて混合連続分布型不特定話者用音素モデルとして表す話者混合法を提案している。特定話者用音素モデルの集合として不特定話者用音素モデルを構築するという方法の提案は，音声認識における個人差への対処法として新しい方向を示したものであり，高く評価できる。

第4章では，第3章で示した話者混合法を発展させ，特定話者用の音素モデルをクラスタリングし，木構造を作成する方法を与えている。次に少量の学習データによって木構造の各ノードに対応するクラスタのなかで最大尤度を与えるクラスタを選択し，そのクラスタの音素モデルを用いてその後の認識を行う話者選択法を提案している。この方法は精度も高く計算量も少ない方法である。

第5章では，第4章の方法を不特定話者音声認識に拡張するために，入力された音声を不特定話者用音素モデルによって認識し，その認識結果を利用して木構造のなかで最大尤度を与えるクラスタの音素モデルを選択した後，その音素モデルによって入力音声を再認識する方式を提案している。この方式によって従来の不特定話者音声認識法に比べて高い認識率が得られることを示している。この方式は音声の個人差を考慮した不特定話者音声認識方式として実用性の高いものである。

第6章は結論である。

以上要するに本論文は，音声認識における音声の個人差に対処するための種々の方法を提案し，さらにそれらを統合した新しい音声認識方式を提案したものであり，情報基礎科学，音声情報工学の発展に寄与するところが少なくない。

よって，本論文は博士（情報科学）の学位論文として合格と認める。