

学 位 論 文 題 名

A Study of Speaker Individuality for Speech Recognition

（音声認識のための話者の個人性に関する検討）

学位論文内容の要旨

音声は人間にとってもっとも自然な情報伝達の手段である。我々は他に何の道具を用いることなく音声により情報を伝達し合う事ができる。本来、音声による高度な情報伝達は人間固有のものであった。ところが、近年の計算機工学の飛躍的な発展により、音声に人間と機械との間の情報伝達の手段という新たな役割が生じてきた。人間にとって自然な情報伝達の手段である音声による情報伝達が可能となれば、使いやすいユーザインターフェイスが実現できると期待されている。

機械による音声認識の研究は古く、始めて音声認識について報告がなされたのは 1952 の Davis による数字認識であった。それいらい様々な研究がなされているが、現在の音声認識手法に大きな影響を与えたのは、迫江による DP マッチング、Rabiner 等による隠れマルコフモデル (HMM)、Hinton、Lippman 等によるニューラルネットワーク (NN) などである。この中で、HMM は音声の揺らぎを統計的に表現できる、音素等のサブワード単位の処理が行ないやすい、言語モデルとの親和性が高い、強力な学習アルゴリズムが確立されている等の利点があり、現在の認識手法の主流となっている。

HMM は上記の利点をもつ優れた方式であるが、高精度のモデルを得るためには大量のデータベースを必要とする。例えば ATR の特定話者音声認識システムでは 5000 単語以上の音声在学习に用いられている。しかし、話者一人一人にこのように大量の学習データを発声してもらうのは困難であり、不特定話者音声の認識が望まれる。

現在の不特定話者音声認識システムではあらかじめ多数話者の音声を集めることで、入力話者の特徴をカバーするようなモデルを作成する手法が用いられている。しかし、ある程度まで話者数が増加するとある音素のカバーする音響空間上の分布が他の音素の分布と重なる場合が生じ、この手法で得られる認識性能には限界がある。また、音声は発話環境、例えばマイクや環境雑音等、の影響を受けるため、すべての話者および発話環境についてデータを収集することは不可能である。そこで、ある話者あるいは複数話者の音声データを用いて学習された認識システムを話者あるいは環境に対して適応化する必要が生じる。

HMM を基にする音声認識方式において話者適応を行なう場合、十分な精度で統計量を推定するためには入力話者の発声したデータがある程度必要となる。一方、我々人間は容易に不特

定話者音声の認識を行なっている。NTTの加藤は、人間の場合数音節の入力によって話者適応化が行なわれていることを報告している。加藤は報告の中で、人間の場合にはあらかじめ典型的な話者の特徴をいくつか蓄えておき、その中の入力話者に類似した話者の特徴をもとに音声に関する拘束条件を利用して話者の適応化を行なっているため、ごく少量のデータで適応化が可能なのではないかと述べている。したがって、計算機による音声認識においても音声に含まれる情報を用いて適当な拘束条件を用いることで、少量の学習データにより適応化が行なえる可能性がある。音声に含まれる第一の情報はその意味内容であるが、その他にも話者が誰であるか、あるいは話者の感情等のさまざまな情報が含まれている。本論文では音声認識のための個人性情報の利用について、HMMによる音声認識における話者適応化の面と、そのための話者の個人性のモデル化の面から検討を行なった。

音声認識における話者適応化の研究は古く、様々な手法が提案されている。

1 音声生成器官の違いを正規化する方法

2 複数の登録話者から適合する話者を選択する方法

3 話者間の特徴量の対応を求める方法

これらの手法は基本的には音声生成器官の違いによるスペクトル形状の異なりを適応化する手法である。しかし、話者による音声の異なりはスペクトル形状のように静的な特徴だけではなく、音声生成器官の動き、すなわち聴音様式の異なりについても現れる。したがって、より高精度の音声認識を実現するためには聴音様式のような動的な特徴にも対処することが必要である。

本論文では、始めに話者の動的特徴をカバーするための複数話者を用いたHMM学習方法を提案し、静的および動的な特徴を別に扱うことで従来の静的な特徴のみを対象とする話者適応方式よりもより高精度の話者適応化が行なえることを示した。

つぎに、複数話者を用いたHMM学習方法において動的な特徴の類似性により複数話者への重み付けに反映する手法を提案し、より高精度の話者適応化が実現できることを明かにした。

また、適応化時の話者への負担を軽減するための手法として、話者間の写像を連続的なベクトル場としてとらえ、少数語彙による話者適応方法を提案し、適応化のために必要とするデータ量を半減させられることを示した。

ついで、話者の個人性を直接モデル化する試みとして、統計的話者モデル化行ない、そのモデルに基づく者適応化方法を提案し、話者の個人性を直接モデル化の可能性を示した。

また、個人性のモデル化の有効性を検証するために、話者認識実験を行ない、長時間に渡る動的な特徴をモデル化できる予測型ニューラルネットワークによる話者モデルが有効であることを示した。

最後に本論文の結果についてまとめると共に今後の課題を明かにした。

学位論文審査の要旨

主 査	教 授	梶 内 香 次
副 査	教 授	永 井 信 夫
副 査	教 授	新 保 勝
副 査	教 授	伊 福 部 達
副 査	助 教 授	宮 永 喜 一

学 位 論 文 題 名

A Study of Speaker Individuality for Speech Recognition

(音声認識のための話者の個人性に関する検討)

音声は人間にとって最も自然な情報伝達の手段である。それゆえ、人間と計算機との情報伝達に音声を用いることが望まれている。音声認識はこれを実現するための重要な技術であり、これまでさまざまな手法が研究されてきた。その一つとして最近提案された隠れマルコフモデル(HMM)による音声認識手法は、

- 1) 音声の揺らぎを統計的に表現できる
- 2) 音素等のサブワード単位の処理が行ないやすい
- 3) 言語モデルとの親和性が高い
- 4) 強力な学習アルゴリズムが確立されている

等の利点があり、現在主流となっている。しかしながら、音声は個人個人による差異が大きく、また同一個人であっても変動があり、それらを吸収した一般性のある音声認識システムの構築が課題である。また、HMMによる高精度のモデルを得るためには大量の音声データをシステムに学習させる必要があるが、非常に多数の話者について大量の学習データを収集するのは困難であり、少数の学習データによって不特定話者に対するモデルを得る手法を実現する必要がある。また、話者数が増加するとある音素のカバーする音響空間上の分布が他の音素の分布と重なる場合が生じ、認識性能には限界がある。

以上の諸点から、実用的な音声認識システムを実現するためには、比較的少数の話者の音声データを用いて学習されたモデルを不特定多数の話者に対して適応化する必要が生じる。本論文は、個人性情報の利用という観点からHMMに基づく音声認識における話者適応化について著者の行った研究を述べたもので、その主要な成果は以下の諸点に要約される。

1) これまで、あらかじめ多数話者の音声を集めることで、入力話者の特徴をカバーするようなモデルを作成する手法が用いられている。これは基本的には話者によるスペクトル形状の異なりを適応化する手法である。しかし、話者による音声の差異はスペクトル形状のように静的な特徴だけではなく、音声生成器官の動き、すなわち調音様式のような動的な特徴についても現れる。本論文では話者の動的特徴をカバーするための学習方法を提案し、静的および動的

な特徴を別に扱うことで従来の静的な特徴のみを対象とする話者適応化方式より高精度の話者適応化が行なえることを示した。そして、動的な特徴の類似性を複数話者への重み付けに反映する具体的手法を提案して実験を行い、高精度の話者適応化が実現できることを明かにした。

2) 適応化学習時の話者への負担を軽減するための手法として、話者間の写像を連続的なベクトル場としてとらえ、少数語彙による話者適応を実現するベクトル場平滑法を提案し、適応化のために必要とするデータ量を、従来知られている手法の半数以下に減少できることを示した。さらに、音素認識実験を行い、従来の手法と同程度の認識率が得られることを確認した。

3) 複数話者の動的な特徴の類似性によらずに話者の個人性を直接モデル化する新しい手法として、統計的話者モデル化とこのモデルに基づく話者適応化方法を提案した。さらに、これに基づく認識実験を行って従来手法と同等の結果が得られることを確認し、話者の個人性を直接モデル化する手法の可能性を示した。

4) さらに、別な面からこの手法の有効性を検証するために、音声認識とならんで音声情報処理工学における重要な課題である話者認識への適用を試みる実験を行ない、長時間にわたる動的な特徴をモデル化できる予測型ニューラルネットワークによる話者モデルが有効であることを示した。

以上のように本論文は、音声に含まれる個人性情報に着目し、これを積極的に利用した音声認識アルゴリズムの多面的な検討を行ない、高精度な音声認識システムの実現に有益な多くの新知見を得ており、信号処理工学及び音声情報工学の進歩に寄与するところ大である。よって、著者は、北海道大学博士（工学）の学位を授与される資格あるものと認める。