

電子・情報工学専攻	学籍番号	933436	指導教官氏名	中川 聖一 新田 恒雄 金子 豊久
申請者氏名	山本 一公			

## 論文要旨 (博士)

論文題目	音声認識のための高精度で頑健な音響モデルに関する研究
------	----------------------------

(要旨 1,200 字程度)

音声は人間が生まれながらに持っているコミュニケーション手段であり、コンピュータのユーザインターフェイスとしても、有用な方法になると考えられている。音声認識技術は徐々に実用レベルへと近付いているが、実際にユーザインターフェイスとして広く使用されるには至っておらず、不特定話者の実環境下での高精度な認識性能を求めて研究が行われている。

現在の音声認識システムは、音響モデルとして隠れマルコフモデル (Hidden Markov Model; HMM) を使うことが一般的である。しかし、基本的な HMM は、音声特徴パラメータの時間的な動的変化をうまく捕らえることができないという欠点を持っている。そこで本研究では、複数のフレームを結合してセグメントを構成しそれを HMM に入力するセグメント単位入力 HMM について様々な検討を行った。従来法とセグメント単位入力 HMM との比較では、従来法に比べてセグメント単位入力 HMM が優れた認識性能を示し、さらにパワー情報、混合分布化を用いることにより、さらに認識精度が改善されるという結果が得られた。また、コンテキスト依存モデルの導入により、コンテキストが大きく影響する部分では、認識率が改善されるという結果も得られた。また、コンテキスト依存の音素モデル (triphone) との比較を行い、認識単位として音節を用いることの有効性を確かめた。

次に、実環境下における頑健な音声認識を目指して、セグメント単位入力 HMM を雑音環境下での認識に適用した。認識実験の結果、セグメントの統計量のみを用いる通常のセグメント単位入力 HMM では、S/N 比の低下と共に急激に認識率が低下するという結果となった。しかし、セグメントの統計量に回帰係数を組み合わせることで最も良い認識率が得られ、雑音環境下においてもセグメント単位入力 HMM が有効であることが示された。また、雑音を重畳して学習したモデルにおいても、セグメント単位入力 HMM が高い認識性能を示した。この結果を受けて、セグメント単位入力 HMM での HMM 合成法を提案し、従来の HMM 合成法に比べて、雑音環境下での高い認識性能を得ることができた。

続いて、学習環境と認識環境でのサンプリング周波数の違いが認識に及ぼす影響、および、サウンドデバイスの特性の違いによる認識性能差について考察した。わずかなサンプリング周波数の違いであっても、認識性能に差が生じ、また、機種間でも明らかな認識性能の差があることが分かった。この結果から、サンプリング周波数の違いとサウンドデバイスの周波数特性の違いに対するモデル補償法を提案し、この方法が比較的単純な方法ながらも、大きな認識率の改善を得ることができた。

最後に、自然な発話に対する頑健な認識のための基本的な考察として、発話スタイルが異なる音声データベースに対して、発話速度およびその分散、学習された音響モデルのモデル間距離 (音韻間距離) から、音声認識性能と発話スタイルの関係について考察した。対話音声では、読み上げ音声に比べて音韻間距離が短い、発話速度が速く分散も大きいという結果が得られ、認識実験でも対話音声では音節正解率が低くなり、音韻間距離や発話速度の分散の影響が認識性能にはっきりと現れる結果となった。