

外93-33

早稲田大学大学院理工学研究科

2011

博士論文概要

論文題目

不特定話者大語彙者声認識に関する研究

申請者

渡辺 隆夫

Takao Watanabe

平成 5 年 10 月

社会の情報化が進む中で、コンピュータ機器やシステムのユーザインタフェースを使いやすいものとすることは社会的にも極めて重要な要請となっており、音声認識に対する期待は高まっている。音声認識装置は、既に、一部の分野で実用化されているが、次のような問題が存在する。一つは、登録時の発声と認識時の発声の違い等の要因により、認識率が低下するという認識精度の問題である。もう一つは、認識語彙数・話者・発話法の制約である。連続単語認識や不特定話者認識では、非常に小規模語彙に限定される。また、話者の単語発声データを用いて標準パターンを作成する必要があるため、認識対象語彙の変更や拡大が容易でなく、特に、多数話者の学習データを必要とする不特定話者認識では、語彙の変更・拡大が困難である。

こうした問題を根本的に解決し、大語彙の不特定話者連続音声認識を実現するためには、音素（あるいは音素に代わる基本音声単位）の認識が不可欠である。隠れマルコフモデル（HMM）による認識は自由度の高い音素のモデル化を可能とし、近年、大語彙認識や不特定話者認識に用いられ、学習データからの統計的学習により良好な認識性能を得ている。しかしながら、認識タスクに依存した大量の学習データが必要であり、実用的な観点からは、認識タスクには依存せずに小規模データによる効率のよい学習が可能であることが要求される。一方、連続音声認識を更に大語彙化しようとすると、探索に要する計算量の多さは、依然として大きな問題であり、効率のよい探索法が望まれている。

本研究では、比較的小規模のデータからの学習の可能な、大語彙認識、不特定話者認識、連続音声認識に向けた認識方法を検討、提案する。

まず、第一に問題となるのは、認識単位とそのモデル化である。本研究では、認識単位として、比較的少ない種類で音素間の調音結合を扱える半音節を採用し、半音節のモデルとしてHMMを用いる。HMMは、音響的な特徴量の変動を統計的に精度よく表現できる能力を持つが、限られた学習データから信頼性の高い学習を行うには、先見的な知識に基づき推定パラメータの個数を適切に制約したモデルが望まれ、そのための検討を行う。第二の問題は、学習法である。ここでは、効率的な学習を実現するため、コンパクトな学習データセット、学習の初期モデルの設定等について検討する。第三の問題は、認識の高速化である。ここでは、特に、連続音声認識を高速化するため探索法を検討する。また、本研究では、上記の認識方法を応用し、音声認識装置を実環境で用いる場合に重要な問題の一つである未知発話リジェクションを行う方法を検討する。さらに、半音節認識単位による認識方式に基づき、不特定話者連続音声認識システムを構築し、性能を評価する。

本論文は8章より構成される。以下にその概要を示す。

第1章では、音声認識の実用の現状とその技術課題を明らかにするとともに、音声認識研究の現状を概観し、本研究の課題を明らかにした。

第2章では、比較的小規模の学習データからの学習が可能な、大語彙向きの音声認識方式として、半音節を単位とする音声認識の基本的な方法について検討した。認識単位として、比較的少ない種類で音素間の調音結合を扱える、C V(子音+母音)とV C(母音+子音)を基本とする半音節単位を採用し、半音節のモデルとしてHMMを用いた認識方式を提案した。発声変動に効率的に対処するため、推定すべきモデルに、音声の知見に基づく構造を導入した。すなわち、語中の大きな発声変形に対しては半音節の変形モデルを導入するとともに、HMMの出力確率を単純なスカラー分散を持つ単一ガウス分布で表現することにより、推定すべきパラメータ数を抑えた。これにより、小規模データ学習の信頼性の向上を図っている。一方、学習に関しては、認識語彙と独立に効率よく学習を行うのに適した学習単語セットとして、小規模(250単語)の音素バランス単語セットを設計するとともに、学習データに対する視察による半音節セグメンテーションや発声変形の判定を行うことなしに、基準話者のモデルを初期値として自動学習が可能な学習法を提案した。500単語及び1800単語認識実験の結果、良好な認識率(99.0%及び97.5%)を得、本方式の有効性を確認した。また、250単語認識において、従来の単語単位DPマッチング法を越える認識性能を得、実用性能向上の面でも有望であることを示した。

第3章では、半音節認識単位による認識方式の不特定話者音声認識への拡張について論じた。話者の違いによるパターンの変動を表現するため、HMMの出力確率モデルとして、単一ガウス分布より表現能力の高い混合ガウス分布を用いた。また、多数話者の学習データによる学習を効率的に行うため、学習の初期モデルとして特定話者モデルを合成して得る学習法を提案した。これにより、最適解に近い初期値からの学習が行われるため、収束の早い学習が可能となった。認識実験の結果、ガウス分布混合数4と比較的少數の混合数で、5000単語認識に対して認識率85.2%と良好な認識性能を得た。本方式が不特定話者単語認識においても有効であることを明らかにした。

第4、5章では、半音節認識単位による認識方式の連続音声認識への拡張について論じた。まず、第4章では、探索処理量の多さを解決する一つの方策として、フレーム同期DPマッチングに単語レベルの照合処理に束ね処理を取り入れた、新しい高速連続音声認識アルゴリズム—バンドルサーチ法—を提案した。本サーチ法では、同じ単語が文の探索の途中で何度か出現するときに、これらの単語に対する照合処理を別々に行わずに、1回の束ねた照合処理で済ませる。このため、単語レベルの照合処理をK分の1のオーダーに低減できる(Kは単語の出現回数)。近似解法であるが、全探索、すなわち、すべての単語列についてその可能性の検証を行っているので、得られる解は最適解に近い。本方法では、構文が複雑になっても、単語数が増えない限り、処理量の増加は少ない。このため、構文ネットワークに言語的な制約をきめ細かく取り入れることが容易になっている。

る。

第5章では、半音節認識単位による認識方式を連続音声認識へ拡張した方式について論じた。連続音声認識で問題となる単語接続部での調音結合の影響に対処するため、単語間の音素遷移を表現するモデルとして半音節モデルを導入した。単語間モデルの導入に伴う探索処理の増加を避けるため、単語間半音節モデルをバンドル化した探索を行った。4桁連続数字および連続音声認識実験(単語数500)の結果、バンドルサーチの導入により認識性能をほとんど劣化させることなく、サーチ(漸化式計算)量をそれぞれ32%, 38%に低減できることが示され、バンドルサーチ法の有効性を確認した。また、チケット予約タスクに対して、不特定話者の文認識率83.0%，単語認識率95.5%と、良好な認識性能を得、本方式が不特定話者連続音声認識において有効であることを明らかにした。

第6章では、半音節認識単位による認識方式を応用して、実環境における問題の一つである未知発話リジェクションを行う方法について論じた。本方法では、タスク文法の制約のない音節認識とタスク文法の制約下での認識を並列に実行し、得られた2つの尤度の差をとることにより、タスク認識に対する尤度を補正し、これを用いて未知発話のリジェクト判定を行う。尤度補正により、話者や環境の違いによる尤度の変動の影響を除去でき、一定の閾値による判定が可能となる。不特定話者認識による評価実験の結果、離散250単語認識および連続音声認識(単語数500)において、リジェクト正判定率(正解率と正リジェクト率の平均)を、それぞれ、66.3%から91.2%へ、76.0%から91.6%へ改善できた。

第7章では、本認識方式に基づいて構築した不特定話者連続音声認識システムとその評価について述べた。本システムは、認識処理と意味理解を統合化した処理により、単語列とともにその概念表現を出力する。概念表現を得るために、あらかじめ、構文・意味制約を記述した認識用構文ネットワーク上で各単語間の概念依存関係を記述している。本システムの性能として、チケット予約タスク(500単語・単語パープレクシティ5.5)不特定話者連続音声認識において、文認識率83%，意味理解率93%を得ている。

最後に、第8章では、本研究を総括し、今後の研究の展望について論じた。

本研究で達成された性能は、現在の実用水準を越えるものとして、1000語程度の特定話者・不特定話者単語認識やパープレクシティの小さい特定話者連続音声認識については、実用化レベルに達していると考えられる。タスク独立学習により実用レベルの性能を得ている点で、本研究の到達点は実用的に画期的である。一方、不特定話者連続認識の性能は良好なものではあるが、真に使いやすい音声インターフェースを実現するには、更に、性能の向上、発話自由度の拡大が必要である。不特定話者認識における認識率の低い話者への対応手段としての話者適応化、自然なインターフェースの実現にとって不可欠な自由発話の許容等が重要であると考える。