

平成9年12月24日

電子・情報	工学専攻	紹介教官氏名	中川 聖一
申請者氏名	西村 雅史		

論文要旨 (博士)

論文題目	日本語の大語彙音声認識およびディクテーションシステムに関する研究
------	----------------------------------

(要旨 1,200 字程度)

本研究の目的は自然な発話を聞き取り、その内容を書き取ってくれる認識装置 (ディクテーションシステム) の実現である。本論文では、特定話者の単音節認識による日本語入力に始まり、統計的音響モデルを用いた大語彙音声認識、さらには言語モデルを併用した日本語の不特定話者大語彙連続音声認識に至る種々の手法について詳述する。

本研究を開始した時点では計算機の処理能力も低く、処理できるデータ量も限られたものであったので、その条件の下で認識装置の開発改良を進めた。単音節入力は、このような計算資源が限られた場面で補助的な日本語入力を行う上では、日本語の言語的な特徴を利用した有効な入力手段の一つである。ここでは、中間累積距離によるマッチング手法を提案し、入力音声の発声速度変動に対しても安定した認識結果が得られることを示す。また、信頼性の高い候補音節だけを効率よく出力する方法として、音節間類似度による候補の絞り込み方法を提案する。この研究の後、我々は、音声のゆらぎを効率よくかつ精度よく表現できるモデルとして、Hidden Markov Model (HMM) に基づく音声認識手法に着目した。

まず学習データ量が限られる場面でも高い精度を維持出来るようにパラメータの平滑化や、ベクトル量子化誤差の削減方法、フェノニックマルコフモデルと呼ばれるフレーム単位のモデルによる状態共有などを検討した。また、音韻の継続時間長やスペクトルの動的特徴が、音声の知覚にとって重要な要素となっているという知見に基づき、継続時間長や動的特徴を陽に表現するモデルについても検討し、良好な結果を得た。一方、特定話者の大語彙音声認識における最大の問題点である学習時間を削減するため、高速な話者適応化手法についても検討した。ここではベクトル量子化コードブックと HMM のそれぞれに対して適応化手法を提案し、学習データ量を大幅に削減できることを示す。これまでに示した音声認識アルゴリズムの多くは実際に、低コストで汎用の音声処理カード上に実装し、話者適応化機能を持つ実時間大語彙音声認識システムとして実用化している。特定話者の1,000 単語認識においては 99.6% 以上の高い認識率を達成しつつ、ラベルヒストグラムによる予備選択法と、ビームサーチ手法により、計算量を約 0.3% にまで削減した。その後、日本語に対しても大量の言語データが利用可能となり、また同時に大量の音声データを処理するだけの計算機能力が利用可能となったことから、本格的な日本語ディクテーションの実現に向けての研究を開始した。ここでは、日本語にとってもっとも効率的な認識単位を見出すことを最初の研究対象とし、人間の単語分割に関する振る舞いを調査した。その結果、これまでほとんど検討されてこなかった離散単語発声による日本語入力の可能性を見出すことができた。また、この離散発声用の認識単位は連続音声用の認識単位としても優れており、約 4 万語からなる不特定話者大語彙連続音声認識実験の結果、97% の文字正解精度、95.6% の単語正解精度が得られるに至っている。なお、実用化の上では、言語モデルの高精度化と同時にその巨大なデータサイズが問題となる。ここではクラス言語モデルによってその削減を図る方法についても述べる。