学位報告

報告番号	※甲　　　第　　　号

主 論 文 の 要 旨

論文題目　音声認識システムの実用化のための音声認識エンジンの高度化に関する研究

氏　　名　小川　厚徳

論 文 内 容 の 要 旨

音声認識は様々な産業分野への応用が期待される技術である．本論文では，既存の産業応用分野または新たな産業応用分野において実用される音声認識システムを構築するために，音声認識エンジンの五つの構成要素（音響分析部，音響モデル，単語発音辞書，言語モデル，探索部）の高度化（高速化/高精度化/特化）を行った．

音声認識システムは，音声認識機能を提供する音声認識エンジンと，それを基に，具体的な音声認識システムの機能を実現するアプリケーションプログラムから構成される．音声認識エンジンは，(1)入力音声を特徴量ベクトル時系列に変換する音響分析部，(2)音声を統計的に表現する音響モデル，(3)認識対象の単語とその発音を規定する単語発音辞書，(4)単語のつながりを統計的にまたは定型文として表現する言語モデル，(5)音響モデル，単語発音辞書，言語モデルを用いて，特徴量ベクトル時系列に変換された入力音声を認識結果文に変換する探索部，の五つの要素からなる．

ひとつの汎用的な音声認識エンジンを基に，様々な音声認識システムを構築できるのが理想であるが，現状では，構築する音声認識システムの仕様に合わせて，音声認識エンジンを高度化する必要がある．本論文では，音声認識エンジンの高度化を，高速化，高精度化，特化の三つに大別して定義する．音声認識は元来計算量の多い処理であり，高速化はこれを削減するために必要な高度化である．高速化は，音声認識エンジンの(2)音響モデルと(5)探索部に対して行うのが効果が高い．高精度化は音声認識システムが使用されるために十分な認識精度を保証するために不可欠な高度化である．高精度化の取り組みは多様で幅広く，音声認識エンジンの全構成要素に対して行われる．高速化と高精度化がどのような音声認識システムの構築においても必要な，音声認識エンジンの基本性能向上ための高度化であるのに対し，特化は，音声認識システムを応用する特定の産業分野の条件に，文字通り音声認識エンジンを特化する（他の条件は考えない）ことにより，高い音声認識性能を実現しようとする高度化である．特にこれまで対象ではなかった産業分野に音声認識システムを応用する場合には，基本性能向上の取り組みだけで十分な性能を出すのは困難なことが多く，音声認識エンジンの特化が不可欠である．音声認識エンジンの構成要素では，

モデルである(2)音響モデル，(3)単語発音辞書，(4)言語モデルに対しては，適応化手法が確立されており，これにより特化を行うことができる．一方，処理部である(1)音響分析部と(5)探索部に対しては確立された手法はなく，応用先の条件に合わせて特化の手法を提案する必要がある．

本論文は全10章からなる．1章では，序論として本研究の背景と目的及び本論文の構成について述べる．2章では，音声認識エンジンの五つの構成要素について図式を用いて概説する．3章から7章では，五つの具体的な産業応用分野または多少広い意味での応用先の条件を取り上げ，それらに対して音声認識エンジンの高速化と高精度化を行い，その基本性能を向上させる．8章と9章では，これまで音声認識技術の対象ではなかった二つの産業応用分野を取り上げ，これらの条件に対して音声認識エンジンを特化させることにより，音声認識システムを実用化する．3章から9章のいずれにおいても，それぞれ新たな高度化手法を提案し，各応用先の評価データを用いて従来手法との比較評価実験を行い，提案手法の有効性を示す．最後に10章で本論文をまとめる．

3章と4章では音声認識エンジンの高速化を行った．高速化は音声認識エンジンの(2)音響モデルと(5)探索部に対して行う場合に最も高い効果が得られる．本論文では，3章で音響モデルに対して，4章で探索部に対して，それぞれ高速化を行った．

3章では(2)音響モデルに対して高速化を行った．音声認識処理時間に占める音響尤度計算時間の割合は45%から65%にも及ぶため，特にリアルタイム性が要求される音声認識システムを構築する場合は，認識精度を落とすことなく音響尤度計算量を削減することが重要な課題となる．3章では音響モデルの構造から無駄を省きこれを最適化することで音響尤度計算回数を削減する方式による高速化の検討を行った．具体的には，既存の分布間距離尺度による音響モデルのガウス分布数削減方式に，混合重み係数を考慮した分布間距離尺度を導入することで方式の改良を行った．BNF文法を用いた連続数字発声認識実験により，提案尺度によれば，従来の分布間距離尺度を用いる場合よりも高精度な分布数削減を行うことができ，より計算量の少ない音響モデルが得られることを確認した．

3章で(2)音響モデルに対して高速化を行ったのに対し，4章では(5)探索部に対してこれを行った．4章では探索部における仮説展開方式と枝刈りの導入に着目し，2パス探索における新しい第2パス探索方式として，時間非同期ビーム探索を提案した．応用としてテレビ放送番組への字幕付与を取り上げた．提案方式は，単語ラティス上に記憶されている第1パス探索のスコアをヒューリスティックとして利用した仮説展開を行う，shortest-firstな仮説展開を行うことでスコアに基づく枝刈りを正確かつ厳しく行う，という二つの特徴を持つ．2万語のNHKニュース放送音声認識実験により，提案方式によれば，既存の第2パス探索方式であるN-bestリスコアリングやA^*探索よりも高速かつ高精度に認識結果文が得られることを示した．

5, 6, 7章では音声認識エンジンの高精度化を行った．高精度化の取り組みは多様で幅広く，音声認識エンジンの全構成要素に対して行われる．本論文では，5章で(1)音響分析部と(2)音響モデルに対して，6章で(4)言語モデルに対して，7章で(4)言語モデルと(5)探索部に対して高精度化を行った．

5章では(1)音響分析部と(2)音響モデルの高精度化に関する検討を行った．どのような環境におかれても音声認識システムの認識精度を維持するために，音響分析部と音響モデルの乗法性歪み及び加法性雑音への耐性強化は不可欠である．代表的な乗法性歪み及び加法性雑音対策としてCMNとMVNを取り上げ，BNF文法を用いた連続数字発声認識実験を行い，これらに関する従来報告されていない，いくつかの傾向を明らかにした．5章の結果は，音声認識システムが置かれる条件に合わせて，作成する音響モデルの正規化単位をどのように設定すべきかの参考になり得ると考えられる．

6章では(4)言語モデルの高精度化を行った．連続音声認識におけるマルコフ性が仮定されたN-gramモデルでは，音響尤度に対する言語確率の重みを大きく設定するほど認識結果文に含まれる単語数が少なくなるという問題がある．これに対して，文中の単語数を大域的な制約として導入した，一般化ベルヌーイ試行に基づく言語確率の補正方法を提案した．連続音声認識実験により，提案手法によれば，言語重みに依存せずに正解に近い単語数の認識結果文を得ることができ，従来のN-gramモデル，N-gramモデルによる対数言語確率を単語数で正規化する方法よりも高い認識精度が得られることを確認した．

7章では(4)言語モデルと(5)探索部の高精度化に関する検討を行った．現場に音声認識システムを導入する際には，その能力を最大限に引き出すため，探索パラメータを注意深く調整する必要がある．従来，パラメータ調整は手作業で行われることが多く，効率が悪く，調整結果の妥当性も不明である．そこで，7章では，連続音声認識における探索パラメータのひとつである単語挿入ペナルティに注目し，これの言語確率及び言語のエントロピーに基づく調整方法を理論的に考察した．結果，認識対象の真の1単語当たりのエントロピーが何らかの形で推定できれば，単語挿入ペナルティの最適な設定値をおおまかに推定できる可能性があることが分かった．7章で考える言語確率は，6章で提案した言語確率補正方法をその特殊な場合として含んでおり，7章での考察が6章の提案法の妥当性を説明していることになる．

8章と9章では音声認識エンジンの特化を行い，音声認識システムを実用化した．従来音声認識技術が応用されてこなかった分野への音声認識システムの導入においては，音声認識エンジンの特化が不可欠である．これらの取り組みを行う際には，3章から7章における高速化及び高精度化の成果も一部利用した．

子供音声認識は，教育やゲームの分野での需要が見込まれる技術であるが，成人音声認識と比較して十分な検討が行われてきたとは言えない．8章では高精度な子供音声認識を実現するために音声認識エンジンの(2)音響モデルと(5)探索部を子供音声に特化した．まず，354名の小学生を対象に，学年及び性別のバランスのとれた単語発声小学生音声データベースを構築した．次に，構築したデータベースを用いて子供音声認識実験を行い，学年ごとの認識率の変化の傾向を明らかにした．これに基づき，小学生音声のクラスタリングの検討を行った．クラスタ音響モデルを作成し，これに対応した選択的探索方式により認識実験を行ったところ，認識率の改善が得られた．8章の成果を利用して公共施設の案内システムなどが14件設置された．

2002年6月，国土交通省航空局により「音声認識機能を応用した管制シミュレータの整備」が計画された．この計画を受け，9章では，日本人の英語発声を高精度に認識できる日英シームレス音声認識方式の開発及びそれの航空路管制音声への適用を行うために，音声認識エンジンを特化した．(2)音響モデル，(3)単語発音辞書，(4)言語モデル（定型文法），(5)探索部の高度化を行い，その成果を基に，音声認識システムを構築した．日英シームレス音声認識方式では，日本語音響モデルと英語音響モデルを併用して，単語単位で単語発音辞書に登録された日本語風英語発音またはネイティブ英語発音のいずれかを選択しながら探索処理を行うことにより，日本人英語発声の高精度認識が可能としている．この方式を，日本人英語発声の典型例と考えられる航空路管制音声に適用し，BNF文法を用いた評価実験を行った．結果，提案方式によれば，従来の日本語専用，英語専用，またはそれらを並列実行する音声認識方式に比べ，高い音声認識精度が得られることを確認した．この取り組みの成果として，全国4ヶ所の航空交通管制部と1ヵ所の航空保安大学校に「音声認識機能を応用した管制シミュレータ」計11台が設置された．