

電子・情報工学専攻	学籍番号	979302
申請者氏名	北岡 教英	

指導教官氏名	中川 聖一 金子 豊久 新田 恒雄
--------	-------------------------

## 論文要旨 (博士)

論文題目	頑健な音声認識手法とその実環境への応用に関する研究
------	---------------------------

(要旨 1,200 字程度)

情報機器が急速に一般家庭に普及し、ヒューマンマシン・インターフェース、特に音声対話に寄せられる期待が大きい。本研究では、まず音声認識が有効となるタスクとは何かを検討し、運転中という環境において、多くの選択肢からの選択を必要とし、なおかつ選択肢が明確でユーザにわかりやすい、カーナビゲーションシステムの地名入力をタスクに設定した。

地名の音声入力を実現するために、大語彙音声認識システムを開発した。このシステムを利用した音声対話インターフェースを構築するにあたり、他の入力手段と組み合わせない Uni-modal システム、対話の進行に伴うモード切り換えのない Uni-mode システムという方針を立てた。これらの方針に従ってプロトタイプシステムを作成し、ユーザが操作する様子を観察して得られた知見をもとに、音声対話ナビゲーションシステムを構築した。

一方、車載機器への応用を考えた場合、雑音対策が必要となる。一般に加算性雑音の除去法であるスペクトルサブトラクション法 (SS 法) では雑音と音声の相関の影響が残るという問題点がある。その解決法として、時間方向にスペクトルの平均化を行う時間方向スムージングを併用した手法 (SS-SMT 法) を提案し、有効性を示した。さらに、分析フレーム長を短くしたり、雑音付加学習を併用することでさらに改良した。乗算性ノイズの除去法であるケプストラム平均正規化法 (CMN 法) の、実時間処理ができないという問題点を、過去の音声データから補正値を推定することにより解決する適応 CMN 法を提案し、有効性を示した。

より柔軟な音声対話システムの構築のために、認識結果に信頼度を付与する手法を検討した。認識結果に対して正解確率の意味をもつ信頼度を得る方法を提案した。単語と音節接続モデルの尤度比に基づく場合、単語内の音節継続時間の分散に基づく場合、両方を用いる場合について、シグモイド関数を用いて正解確率を表現する方法を考案した。また、正解確率に基づく正解と誤認識の分離度をリジェクション実験によって評価した。さらに、語彙外単語のリジェクション実験も行った。いずれの場合も、複数の特徴量を本方法によって組み合わせると単独よりも精度がよくなることが示された。

さらに音声認識において自由な発話を認識できるために、ヒューリスティックな言語モデルを用いたフレーズスポッティングに基づく音声認識・理解手法を検討した。フレーズスポッティングは、A\*適格性条件を満たす best-first 探索に基づいて実現した。定型発話と模擬自由発話の認識実験によって評価した結果、言語モデルとして日本語音節を自由に接続できる音節接続モデルを用いた場合よりも、語彙にある単語の接続と音節の接続を許す単語・音節接続モデルを用いた場合は、抽出率は向上したが、自由発話に対する頑健性がやや欠ける結果となった。

フレーズスポッティングの結果のトレリスを保存し、未探索部分はヒューリスティック言語モデルで評価する、A\*適格なヒューリスティック文認識手法を実現した。探索法として、Island-driven 探索と単方向探索を比較した結果、前者は探索空間が大きくなるために探索を失敗することが多く、後者単の方が認識率が上回った。単方向探索による意味主導探索と構文主導探索を比較では、意味主導の方が自由発話には頑健であると考えられるが、構文主導の方が制約が強いために、認識・理解率では構文主導が上回った。