

(論文内容の要旨)

本論文は、実環境におけるリアルタイムの音声翻訳システムを指向した音声認識処理の高度化に関する研究をまとめたものである。言語の壁を超えるコミュニケーションを実現するために、いつでも、どこでも、誰とでも会話できる、音声翻訳システムの実現が求められている。音声認識処理はそのフロントエンドであり、システム全体の性能に大きく影響する。本論文では、音声翻訳システムにおける音声認識の課題として、(1)音声認識の探索過程における計算量の削減、(2)キーワードの検出における雑音や語彙外発声に対する頑健性、(3)翻訳処理のための音声認識出力の区分化、の3つに整理して、それぞれを解決する方法を示している。

第1章は序論で、本研究で対象とする、ドメイン限定の双方向型の音声翻訳システムについて紹介し、本研究で扱う課題を整理した上で、アプローチについて述べている。

第2章では、音声翻訳システムに対する入力音声の音響的・言語的な分析の結果、朗読音声とは異なる特徴が見られることを報告している。

第3章では、第一の課題である、音声認識の探索過程における計算量の削減に関して、単語グラフを用いた単語仮説数の削減手法を提案している。本手法では、(1)同音語の言語ゆう度の共有化、(2)開始時刻の代表化处理における先行単語の共有化、(3)木構造辞書ノード間の言語ゆう度の補間、により単語グラフ生成過程での計算量の削減を図っている。自然発話を用いた評価実験の結果、同音語の言語ゆう度の共有は計算量削減効果が大きいこと、単語の先頭の異音と同じであれば単語の開始時刻を一つに絞っても問題ないこと、言語ゆう度の変化を分散させることにより探索に要するビーム幅を小さくできることを示した。実験の結果、本手法は、大語彙連続音声認識において、大幅な計算時間の削減を実現した。

第4章では、第二の課題である、雑音や語彙外発声に対する頑健性のために、非キーワードのモデル化手法を新たに提案している。本手法では、2パスのキーワード認識を行う従来の日本語ガベージモデルの代わりに、英語ガベージモデルを利用する。提案手法は、キーワードセットが変わってもモデルの再学習を必要とせず、高い性能を実現する。クリーンな環境及び雑音環境、25単語から1000単語までの様々な語彙サイズ、複数のキーワードを許容する場合など、いずれの条件においても、日本語ガベージモデルを用いる従来手法と比べて提案手法は高いキーワード検出性能を示した。

第5章では、第三の課題である、翻訳処理のための音声認識出力の区分化について、プロの同時通訳者が「自然な」タイミングで区切る音声翻訳単位を検討している。言語情報ならびに韻律情報を素性として与えて、SVMチャンカで音声翻訳単位境界を自動推定する実験を行った結果、ポーズを伴う音声翻訳単位境界ではプロの通訳者に匹敵する推定精度が得られ、ポーズを伴わない音声翻訳単位境界においても、基本周波数の情報を用いることにより推定精度が改善されることを示した。また、これらの情報に加えて句読点の情報を用いることにより、プロの通訳者を上回る推定精度が得られることを確認した。

第6章では、結論と今後の課題を述べている。

(論文審査の結果の要旨)

本論文は、実環境におけるリアルタイムの音声翻訳システムを指向した音声認識処理の高度化に関する研究をまとめたものであり、得られた主な成果は次の通りである。

1. 単語グラフを用いた大語彙連続音声認識の探索過程において、仮説数を削減し、効率的な処理を行う方法を複数提案し、音声認識処理の大幅な高速化を実現した。これにより、自然な発話を対象として、実時間の音声認識を可能にした。
2. 雑音や語彙外発声が混じった音声からキーワード部分を高精度に検出するために、非キーワード部分に対して外国語（英語）の音素モデルに基づくガベージモデルを用いる方法を提案し、様々な条件下でその有効性を示した。これにより、頑健な音声認識を実現した。
3. 音声認識結果を適切かつ早期に区分化して、翻訳システムに渡すために、プロの同時通訳者が入力を区切る単位の自動推定を行った。その上で、ポーズや基本周波数などの韻律情報や句読点の情報をを用いることの効果を示した。

以上のように本論文は、実環境における自然な音声言語を扱う方法論を提案するとともに、音声翻訳システムの実用化に貢献したもので、学術上・実際上寄与するところが少なくない。よって、本論文は博士（情報学）の学位論文として価値あるものと認める。

また、平成21年2月23日実施した論文内容とそれに関連した試問の結果合格と認めた。