

平成14年2月22日

電子・情報工学専攻	学籍番号	953412	指導教官氏名	中川 聖一 新田 恒雄 宇津呂 武仁
申請者氏名	小暮 悟			

論文要旨 (博士)

論文題目	音声対話システムにおける頑健な意味理解と対話システムの移植性に関する研究
------	--------------------------------------

(要旨 1,200 字程度)

人間は生活を営んでいく上で何らかのメディアを通して情報を収集していく必要がある。その中でも人と人との情報伝達の形である対話は生活をしていく上で最も重要な情報伝達形式である。古来から、情報伝達的手段は人と人の直接の対話が基本となっているが、郵便などの紙媒体、電話などの有線形態、モールス信号などの電波形式、現在ではインターネットによる伝送形式など、様々な情報伝達形式が存在する。これらの情報伝達形式は音声および画像を伴うかどうかという区分と、情報伝達が双方向であるかどうかという区分が存在する。

最近、音声によるコミュニケーション手段である音声対話システムというものが注目を集めている。音声対話システムと通常の対話システムにおける最大の違いは入力手法である。音声は「訓練がいらぬ」、「情報伝達の速度が早い」、「手が自由」、「伝送手段に電話を使用できる」という利点を持つ。音声対話システムの研究は従来より行なわれていた。特にシステムの使い易さの向上とシステムの頑健さの向上の研究に焦点がおかれていた。使い易さとはシステムを使用するユーザがシステムを使う際にどれだけ簡単に利用できるかであり、頑健さとはユーザの入力あるいはノイズなどの外的要因に対してシステムがどれだけ正常に動作できるかである。一方、最近では音声対話システムの移植性と拡張性に関する研究も盛んに行なわれてきている。システムの移植性とは、タスクやドメインがどれだけ容易に変更できるかを、システムの拡張性とは、辞書の動的な追加やモジュールの変更などがどれだけ容易にできるかどうかをそれぞれ示している。

本研究ではまず、既存の富士山観光案内を対象としたマルチモーダル音声対話システムの頑健さについての研究を行なった。その結果、音声対話システムの構成要素である、音声認識部、応答生成部、音声合成部についてはすでに頑健さの検討がされていたが、言語理解部についてはまだ頑健さに問題があることが判明した。よって、富士山観光案内の言語理解部について、語彙の追加、指示詞の取り扱い、似たような意味の文の理解を同じ意味表現にするなどの処理を行なった。その結果、従来の言語解析部と比較して意味理解率が向上し、頑健さを上げることができた。

次に、富士山観光案内システムの移植性、拡張性についての研究を行なった。まず、富士山観光案内システムを東三河観光案内システムに実際に変更することで、ドメイン変更の作業量を求めた。その結果、変更には30日・人を要し、移植性が低いことが分かった。さらに音声対話システムの移植性を、各構成要素の移植性を独立に考えることで向上することを考えた。まず音声認識部についての検討をし、その結果、音響モデルはドメインに全く依存しないことが判明した。その反面、認識用の言語モデルはドメインに依存することもわかった。タスク変更時に効率的に認識用の言語モデルを構築する手法についての検討を行ない、未知語に対して、そのクラス名と読みを登録することによって、認識率が向上する言語モデルを得ることが出来た。

続いて、音声対話システムの構成要素である言語理解部の移植性についても研究を行なった。これについては音声認識部のような明確な分割はできないため、言語理解部で使用する各種の辞書について、様々な対話文集合を分析した。それをもとに、ドメインとタスクの両面についてその独立性と依存性を調べ、言語理解部で使用する情報の明確な分割を行なった。そして、富士山観光案内と文献検索システムの2つのタスクに実際に適用し、システムの性能とシステム構築の効率の両面からシステムの移植性を評価し、既存のシステムの性能を損なわずにシステム構築の効率を大幅に向上させることができた。

最後に、音声対話システムの構成要素である対話管理部の移植性についても研究を行なった。まず様々な音声対話システムの対話文集合から、対話システムでどのような処理が必要かを検討し、システム内の処理を明確に分割した。次に、システムの各処理について、ドメインとタスクそれぞれに独立かどうかを判断し、分割を行なった。それを元にドメインとタスクに依存する情報、つまり対話スクリプトを定義し、その対話スクリプトの変更だけで、ユーザ主導、システム主導、および混合主導な様々な対話を構築できる手法を考案した。