

# 第1章 序論

## 1.1 音声対話インタフェース

現代社会においては、情報化の進展に伴い様々なレベルでシステムの複雑化が進み、システムを制御する側の人間にとっては扱いが難しくなる一方である。特にシステムの利用において特別な知識を要求することの難しい公共的なシステムについては、より汎用的で使い易いユーザインタフェースの必要性が高いといえる。音声による対話は、人間が普段用いるコミュニケーション手段であることから、そうしたシステムのユーザインタフェースへの適用が期待されている。

ユーザインタフェースの入出力手段としての音声対話メディアの利点として、以下の3つがあげられよう。\*

1. 入出力に口と耳を使う
2. 入出力手段の利用方法自体を新たに学習する必要がない
3. 入出力の行為が思考と直結しているため要求を伝えやすい

1は、手足が拘束される状況(例えば、自動車の運転、外科手術など)において特に有効である。2は、利用方法の新たな学習を要求できない高齢層や若年層などの利用者や、学習する機会を与えることの難しいアプリケーション(例えば、公共施設の窓口サービスや、観光施設の案内サービスなど)において特に有効である。3の利点は、要求自体が曖昧であったり複雑であるような場合に、音声対話を通じて要求を明確にしていく可能性があることを示す。

一方、近年国内外を問わず、音声認識や音声合成技術の精度が向上し、その実用化が盛んに行われており、システム開発者のためのツールキットやアプリケーションインタフェースが公開され、利用されている [2][3][4][5]。それに伴って、音声対話インタフェー

---

\*なお、対話によらない音声入力自体の利点については [1] の議論が興味深く参考になる。

スについても様々な研究機関で研究が行われるようになり、さらに、カーナビゲーションシステムや航空券予約購入システムパーソナルコンピュータやゲーム機器上のソフトウェアなどで実際に利用されるようになっている。

しかしながら、実用化されたこれらのインタフェースは、要素技術の能力の限界にしたがってあらかじめタスクを限定して語彙を制限するとともに、発声の方法やタイミングを制限するために、上に示したメディアとしての音声対話の利点(特に2と3)を充分には生かしておらず、キーボードなどによる入力代替手段となっているにすぎない。

携帯電話の普及や、インターネットの拡大による情報化のさらなる進展に伴って、今後、より音声対話インタフェースの重要性が高まると考えられる。したがって、上述の利点を充分に生かした音声対話インタフェースの研究開発が望まれる。

## 1.2 音声対話制御

音声対話インタフェースの工学的実現においては、まずシステムとして、どのような方針に従ってユーザとの対話を進めていくかをモデル化することが重要である。このモデルを一般に対話制御モデルという [6]。前節に述べた音声対話インタフェースの実現には、人間とシステムとの相互作用 (interaction) を考慮した対話制御モデルが必要であるといえる [7]。

音声対話インタフェースの相互作用性を向上するためのアプローチはさまざまであるが、中でも対話を円滑に進行するためのコミュニケーション調整機構の実現は最も重要なものの一つといえる。

音声対話インタフェースにおけるコミュニケーションの調整機構において重要な機能として、『発話権』の管理があげられる。発話権とは、対話における発話の順番についてのある種の秩序であり、『発話の番』(turn)とも言い換えられる。対話参加者は対話の中で発話権の所在が誰にあるかを意識的・無意識的に関わらず判断して発話のコントロールを行なっているといえる。音声対話インタフェースにおいてコミュニケーションの調整を実現する際、ユーザの状態を把握したり、システムが発話すべきタイミングを決定するうえで発話権をどのように捉えるか、またその所在をどのようにして把握するかは非常に重要な問題である。これまで、『発話権』のモデルは主に言語学や認知科学の分野で理論付け

が行なわれてきた [8][9] が、その工学的モデルについての研究はほとんど行なわれておらず、まずこの発話権を管理するモデルを構築することが必要である [10]。

また、一般的にインタラクションシステムにおいて相互作用性を向上するためにコミュニケーションを円滑にするには、対話やシステム内部の状態をユーザに的確に伝えることが重要である。その際、例えばディスプレイを用いたユーザインタフェースでは、メッセージやインジケータなどの表示エリアを設けて主となるタスクの遂行を妨げることなくシステム内部の状態を提示することが可能である。一方、ディスプレイを持たない音声対話インタフェースにおいては、単一のチャネルしか持たないメディアとしての特性により、主タスクの遂行とコミュニケーションの調整を同じチャネルで行うためにリソースを適切に分配しなければならないという問題が生じる（単一チャネルにおけるリソース分配の問題）。この問題に対して、主タスクを遂行するための発話行為とは別に音楽を出力することでシステム内部の状態を表そうとするアプローチが報告されている [11]。これはいわば、同一の聴覚チャネルにおいて同種の行為が競合することを避けて、異種の行為をもって異なる目的を達成しようとするものといえる。しかしながら、提示すべき状態が複雑になれば状態提示行為を複雑にせざるを得ないため、言語以外の方法で伝達するにはその解釈方法を学ばなければならなくなる。こうした問題も、音声対話インタフェースの実現において工学的に解決すべき問題であるといえる。

### 1.3 音声対話インタフェースの汎用化

ところで、一般的に音声対話インタフェースを構築する際、問題解決規則や対話制御規則については、アプリケーションごとにそれぞれ個別に設計され、知識のモジュール化についてあまり注意が払われていない。これは、問題解決や対話制御の規則を抽象化することは実際には難しいこと、またそれらの一般的な表現形式として、フレームやプロダクションルールが存在するものの、記述が容易ではないことによると考えられる。そのため、他のアプリケーションでの対話制御に関する知識の再利用が難しく、新しいインタフェースを構築する際には対話モジュールを最初から設計し直さねばならず、インタフェース開発において大きな負担のかかる要因となっている。

こうした背景から、音声対話インタフェースの汎用化について近年いくつかの機関で研

究が行なわれている。秋葉ら [12] は、対話に関する処理を独立に記述するマルチモーダル対話言語を設計した。この記述言語は、より抽象度の低いレベルでの対話のモデル化を目的とし、割り込みなどを含む対話中の任意の状態に対して、局所的なシステムの挙動を細かく指定することができるという特徴をもつ。田中ら [13] はタスクを情報検索に限定し、データベースからの情報をもとに文法と語彙の設定を半自動的に行う、汎用的な音声対話プラットフォームを構築している。このプラットフォームでは、GUIを採用することでユーザ発話を誘導し、限定された発話・対話パターンで音声対話を実現することで、より容易に対話を記述できるという特徴をもつ。また、荒木ら [14] は音声対話システムの対話管理部を作成する方法として、一般的なタスクドメインを情報の流れる方向に着目して分類し、適用できる対話ライブラリを構築し、汎用的な対話ライブラリを組み合わせるといったアプローチを提案している。

この様に、問題解決規則や対話制御規則などの知識を外在化した形で実装しようとする、知識の記述の容易性と多様性のトレードオフが問題となる。つまり、知識を外在化した多様な対話制御をシステムの外部で記述しようとする、扱うパラメータが増えアルゴリズムが複雑になってしまう。また、外在化した知識の記述を容易にするために扱うパラメータを減らすと、多様な記述が不可能となってしまう。こうした問題は、システム設計の一般的な問題といえる。

## 1.4 本論文の構成

本研究は、音声対話インタフェースの実現において前述した種々の問題を解決すべく、汎用的なコミュニケーション調整機構の実現を目指す。本論文では、まず第 2 章でコミュニケーション調整機構の基本となる発話権管理モデルを説明する。第 3 章ではコミュニケーション調整行為の主要な一つである状態提示行為を制御するモデルについて説明する。第 4 章では、汎用的な音声対話インタフェースを実現するためのプラットフォームについて説明する。第 5 章では、第 2、3 章で示したモデルをさらに一般化して第 4 章で示したプラットフォーム上に実装したメタ発話生成モデルについて説明する。第 6 章では本研究の総括を行ない、残された問題点などを述べる。