知識に基づく音声言語理解システムとしての機器操作インタフェース

土井 伸一, 長田 誠也, 亀井 真一郎

NEC C&Cメディア研究所

1. はじめに

本稿では、筆者らが試作した、「つけて」「テレビを8チャンネルにして」等のユーザの自然な音声言語発話による機器操作指示を理解してテレビとビデオを操作する自然言語インタフェースに関して報告する。

音声言語によるインタフェースは、人間にとって最もナチュラルなものの一つであり、テレビやコンピュータ等の機器の操作に関してもこれまでに様々な研究開発が行われてきている。最近では、音声でパソコン上のアプリケーションを操作するシステムや、音声で操作可能なカーナビゲーションシステム等も実用化されている。しかし、現在実現されているシステムは、基本的にあらかじめ登録された文字列だけを認識して決められた操作を行うものであり、ユーザが他の人間に話すのと同様に自由に発話した操作指示を認識・解釈できる訳ではない。

本稿で提案するシステムは、ユーザによって普段の言葉で自由に発話された機器操作指示を適切に解釈してユーザの意図を実現することを目指すものである[1]。近年の音声認識技術は、ディクテーションシステムが製品化される等、ユーザの連続音声をかなりの程度まで認識できるようになってきている。無論、自由発話による機器操作インタフェースとしては、単にユーザの自由な音声発話を認識できるだけでなく、種々の言い回しや省略等の表現を解釈できなければならない。自然言語を理解するには世界や状況等に関する膨大な量の知識を必要とするため、このようなシステムの実現は一般には困難である。しかしながら機器操作インタフェースの場合は、操作対象となる機器は限定されているため、対象世界に関する知識、必要となる言語表現等をあらかじめシステムに与えておくことが可能である。また、各発話時点での対象世界の状況や文脈情報等も取得できる。

以下、テレビとビデオを操作対象とする開発した機器操作インタフェースに関して、システムの全体構成、システム内に保持している対象世界や言語表現等に関する知識の詳細、その知識に基づいて入力発話の理解を行うプロセス、システムからの応答文と対話を通じた曖昧性解消について、順に報告する。

2. システム構成

図1に、本機器操作インタフェースの全体構成を示す。各モジュールの機能の概要は以下の通りである。

- 発話入力(音声認識)部
  音声入力を受け付け、文字列に変換する。
- 入力解析部
  入力された自然言語表現を解析し、曖昧性がある場合は知識を用いて解釈を1つに絞る。
- コマンド生成部
  入力解析部で得た解釈を、ユーザの意図を実現するのに必要なコマンド(列)に展開する。
- コマンド実行部
  展開したコマンド(列)を機器に対して実行する。現在のシステムは、まずパソコン内の仮想機器に対して操作を行い、この仮想機器が赤外リモコンによって実際の機器を操作する形態を取っている。
- 応答文生成部
  入力に対するシステムの解釈結果等を、ユーザに対して応答文として生成し、合成音声で出力する。
- 知識保持部
  対象機器の機能や対応する言語表現、状況知識、文脈知識等を保持する。詳細は次節で説明する。

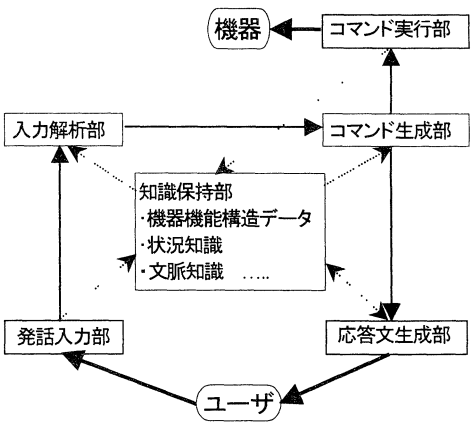

図1 システムの全体構成

表1　ビデオに関する属性、属性値と対応言語表現

属性	属性値	言語表現	属性	属性値	言語表現
POWER		電源,パワー,スイッチ	CASSETTE		カセット,テープ
	on	状態語:オン,ついている,入っている 操作語:つける,入れる		in	状態語:イン,入っている 操作語:入れる
	off	状態語:オフ,消えている,切れている 操作語:消す,切る		out	状態語:アウト 操作語:出す
CHANNEL		チャンネル,番号,チャン	TAPE		カセット,テープ
	1	状態語:1,NHK		stop	状態・操作:停止,ストップ 状態語:止まっている 操作語:止める
	3	状態語:3,教育テレビ			
	4	状態語:4,日本テレビ,日テレ			
	6	状態語:6,TBS		ff	状態・操作:早送り
	8	状態語:8,フジテレビ,フジ		rr	状態・操作:巻戻し
	10	状態語:10,テレビ朝日,テレ朝		play	状態・操作:再生,プレイ
	12	状態語:12,テレビ東京,テレ東		rec	状態・操作:録画

3. システムが保持する知識

本システムでは、操作対象となる機器(対象世界)に関する知識として、各機器がどのような機能(属性)を持っているか、その機能(属性)がどのような値を取りうるかという情報を保持している。さらに、これらの機器、機能(属性)、属性値それぞれには、対応する言語表現のリストが付与されている。表1に、システムが保持しているビデオに関する属性、属性値と対応言語表現の一部を示す。

この種の操作対象となる機器においては、各機能の間には一般に依存関係が存在する。例えばビデオの場合、チャンネルを切り替えたりテープを操作したりするためには電源がオンになっていなければならない。そこで本システムでは、機器に関する知識を木構造の形式の機器機能構造データとして保持している。図2に、ビデオに関する機器機能構造データを示す。また各属性値のノードは、そのノードがどのような情報内容(映像、音等)に関連するものか、また情報源となるのか出力先となるのかという観点から種類分けしてある。このノードの種類に関しては、4.2節で説明する。

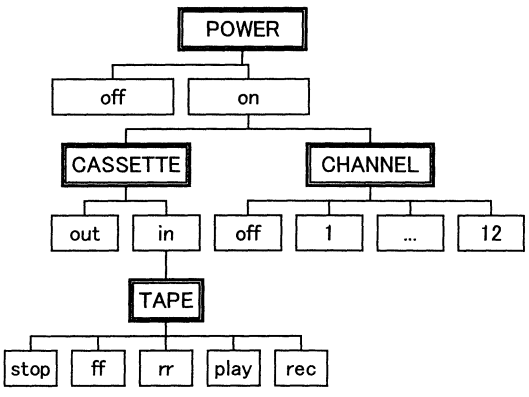

図2　ビデオに関する機器機能構造データ

言語知識として本システムは、上記の各対象機器ごとの知識に加えて、各機器とは独立の、機器操作に関する機能語の辞書を保持している。表2に例を示す。

表2　機能語辞書

種類	言語表現
状態語語尾	～だ、～です、…
操作語語尾	～て(下さい)、…
状態語操作語変換	(に)する、(に)変える、…
状態質問語	～か、～の、…
肯定語	はい、イエス、…
否定・取り消し語	いいえ、ノー、キャンセル、…

さらに本システムは、以下の知識を保持している。

- 状況知識
  ある時点で、操作対象機器の各属性の持つ属性値
- 文脈知識
  ある時点の直前に実行したコマンド(列)、その対象機器、直前のユーザの発話、直前のシステムの応答文等

4. 知識に基づく入力発話理解

本稿で提案するインタフェースは、ユーザの自然な発話による機器操作指示を対象としているため、機器の機能と言語表現との対応は多対多であり、また省略表現等、入力発話には多くの曖昧性が存在する。本システムは前節で述べた種々の知識を駆使して曖昧性の解消を図って、ユーザの意図を実現する。本節では、この入力発話理解の過程を、入力解析部における発話の解析と、その解析結果のコマンド生成部におけるコマンド(列)への展開の2段階に分けて説明する。

4.1. 入力解析部における発話解析

入力解析部は、発話入力部から入力されたユーザの発話を知識保持部に保持された各種の知識を用いて解析する。解析結果に曖昧性がある場合は知識を用いて解釈を1つに絞る。

始めに、形態素解析を行って入力された言語表現を単語に切り分ける。続いて、表1、表2に例を示した言語知識を参照することにより、テレビとビデオの操作という現在の対象世界に関して有意味な単語を抽出し、その単語の組み合わせがどの機器、属性、属性値に対応するものなのか、また操作指示なのか状態質問なのかを解析する。例えば、「再生して」という入力は、「機器:ビデオの属性:TAPE を属性値:play に変更する操作指示」と解析される。現在のシステムでは「再生」という言語表現は他の機能とは対応しないため解析結果は一意に確定されるので、入力解析部はこの解釈をコマンド生成部に送る。

一方、例えば「つけて」という入力からは、「機器:テレビの属性:POWER を属性値:on に変更する操作指示」と「機器:ビデオの属性:POWER を属性値:on に変更する操作指示」の2種の解析結果が得られる。このように解析結果が複数得られた場合、入力解析部は、状況知識、文脈知識を利用して解釈を1つに絞る。例えば、状況知識を参照したところ、「つけて」という入力があった時点でテレビはついていてビデオは消えているという情報が得られたとすると、解析結果で既に実現されているものを破棄して「ビデオをつける」操作指示だと解釈する。また、文脈知識を用いる例としては、「テレビつけて」とユーザが発話した直後に「チャンネルを4にして」と発話した場合、チャンネルはテレビにもビデオにも存在するが、テレビのチャンネルを変える操作指示だと解釈する。これらの知識を用いても曖昧性が残る場合には、デフォルトの優先度によってコマンド生成部に送る解釈を決めている。

4.2. コマンド生成部におけるコマンド展開

コマンド生成部は、入力解析部から送られた解釈を、ユーザの意図を実現するのに必要なコマンド(列)に展開する。展開するための知識源として、3節で紹介した機器機能構造データにおける機能間の依存関係とノードの種類を用いる。図3に、ノードの種類を付与したテレビとビデオの機器機能構造データを示す。

ここでは属性値を表すリーフノードを、情報源ノード、情報出力ノード、その他の3種に分類している。情報源ノードは、映像、音といった、外部に出力されて初めて意味のある情報を生成するノードであり、ノードには生成する情報内容に関する知識が付与されている。一方情報出力ノードは、生成された情報をユーザや保存メディアといった出力場所に出力するノードであり、情報内容と出力場所の知識が付与されている。

以下、前節で説明した「再生して」というユーザの発話に対する「機器:ビデオの属性:TAPE を属性値:play に変更する操作指示」という解釈が入力解析部から送られた際の、コマンド展開法を説明する。

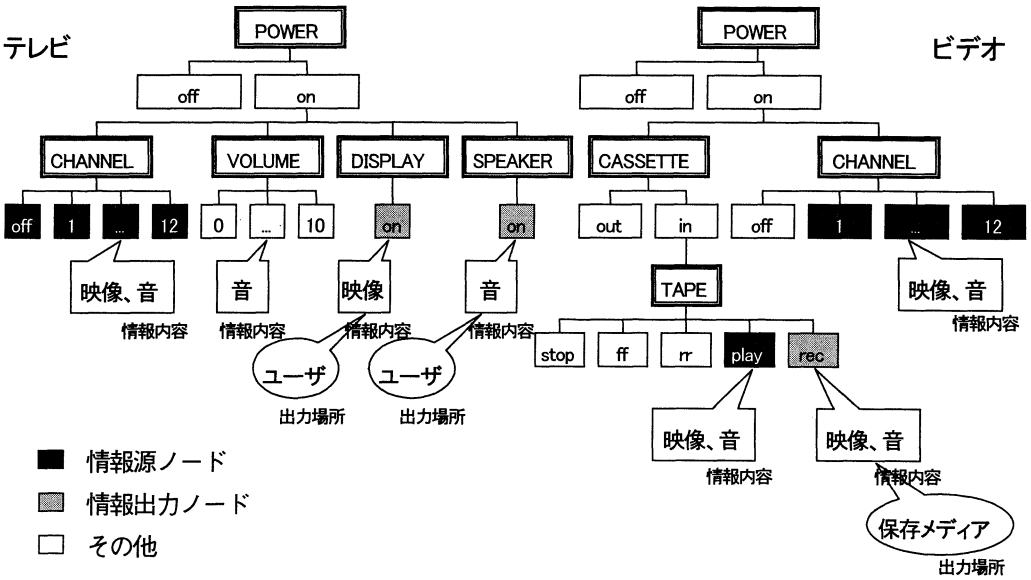

図3　ノードの種類を付与したテレビとビデオの機器機能構造データ

始めに、機能間の依存関係をチェックする。図3の機器機能構造データを参照すると、「属性:TAPE の属性値:play」は、「属性:POWER の属性値:ON」「属性:CASSETTE の属性値:IN」に依存していることが分かる。本システムは、この木構造をルートからたどる形で、これらの属性値が既に実現されているか否かを状況知識を参照してチェックし、実現されていなければ実現するためのコマンドを生成する。なお、属性:CASSETTE はシステムが直接操作できる機能ではないため、操作が必要な場合はユーザに対して操作を促す応答文を生成するようにしている。

続いて、ノードの種類のチェックを行い、操作指示対象となる属性値が情報源ノードの場合には、それが生成する情報内容を出力することのできる情報出力ノードを検索して、その属性値を実現するためのコマンドも生成する。「属性:TAPE の属性値:play」の場合には生成される情報内容は映像と音の両者であり、ビデオの内部にはこれらを出力できる情報出力ノードは rec しか存在しない。しかし、play と rec は同一の属性の属性値なので両立できない。そこでコマンド生成部は、操作可能な他の機器であるテレビを操作してユーザに対して映像と音を出力するようにする。

このコマンド生成部における機器機能構造データを参照したコマンドの展開により、ユーザは目的の状態を実現するのに必要な操作を逐一指示することなく、目的の操作のみを発話して実現することができる。

5. システムからの応答

2節で説明したように、本システムは、単に機器操作を実行するだけでなく、ユーザに対して応答文を生成する機能を有している。応答文の内容としては、入力に対するシステムの解釈結果、ユーザからの機器状態に関する質問に対する回答、システムが操作できない機能のユーザに対する操作依頼等がある。

さらに、この応答文を文脈情報として保持することで、応答文に対するユーザの肯定、否定(取り消し)を意味する発話を解釈する機能も有している。入力解析部は種々の知識を駆使してユーザの意図を推定するのであるが、常に意図を一意に特定できる訳ではない。そこで、解釈結果をユーザに対して応答文として生成してそれに対する肯定、否定発話を受け付けることで、ユーザとの対話を通して解釈を確定することができるようになっている。

しかしながら、常にユーザに肯定/否定行為を要求するインタフェースは自然なものとは言えない。そこで本システムでは、入力解析部で入力に曖昧性がないと判断した場合には、まず操作を実行して、その後ユーザに対して報告を行う形を取っている。この場合でも、直前の操作コマンド(列)を文脈知識として記憶しているので、ユーザは簡単な発話で操作を取り消して元の状態に戻すことができるようになっている。さらに、曖昧性が存在する場合には操作の実行前に解釈結果をユーザに提示する訳であるが、ユーザが肯定も否定も行わない場合は暗黙のうちに了解されたものとして操作を実行するようにしている。これにより、ユーザにストレスを感じさせない自然な機器操作インタフェースを実現している。

以下、本システムが生成する応答文の例を示す。

・ユーザの発話:「テレビの電源をつけて」
　システムの応答文:「テレビの電源をつけました」
　※入力に曖昧性がないため、実行後報告
・ユーザの発話:「ＮＨＫ」
　応答文:「テレビのチャンネルをＮＨＫにしますね」
　※文脈情報を用いて解釈＆実行前に確認を要求
・ユーザの発話:「違う、ビデオ」
　応答文:「取り消します
　　ビデオのチャンネルをＮＨＫにしますね」
　※簡単な発話で誤った解釈の取り消し/修正が可能

6. おわりに

筆者らが試作した、ユーザの音声による指示を理解して機器操作を行う自然言語インタフェースに関して報告した。本インタフェースは、パソコン上で動作し、「テレビつけて」「８チャンネル」「再生」等の音声入力を受け付けて知識に基づいて解釈し、テレビとビデオを操作してユーザの意図する機器操作を実現する。一般にユーザの発話は曖昧であり、システムは世界(対象機器)知識、状況知識、文脈知識、言語知識等を駆使して曖昧性の解消を図って、ユーザの意図を推定する。意図を特定できない場合にはユーザとの対話を通して解釈を確定する。

現在、応用として、ユーザの指差し(ジェスチャ)による機器操作指示を認識するシステムと本システムとを統合したマルチモーダルなインタフェースを備えた機器操作システム[2]の試作も行っている。

今後は、対象となる機器の数や機能、対応する語彙をさらに拡充するとともに、実際に使用しながらの評価を行って、ユーザにとってより自然な形の機器操作インタフェースの実現を目指す。

参考文献

[1] 長田 他: “自然言語を用いて家庭機器操作を行う対話システム”, 信学技法 SP98-73, 1998.

[2] 茶園 他: “ユーザの注目情報を利用したマルチモーダル・インタフェースの試作”, 情処第 58 回全大 5E-05, 1999.