

藤井洋一*
 小路悠介**
 石井 純*

音声対話技術

Spoken Dialog Technology

Youichi Fujii, Yusuke Koji, Jun Ishii

要 旨

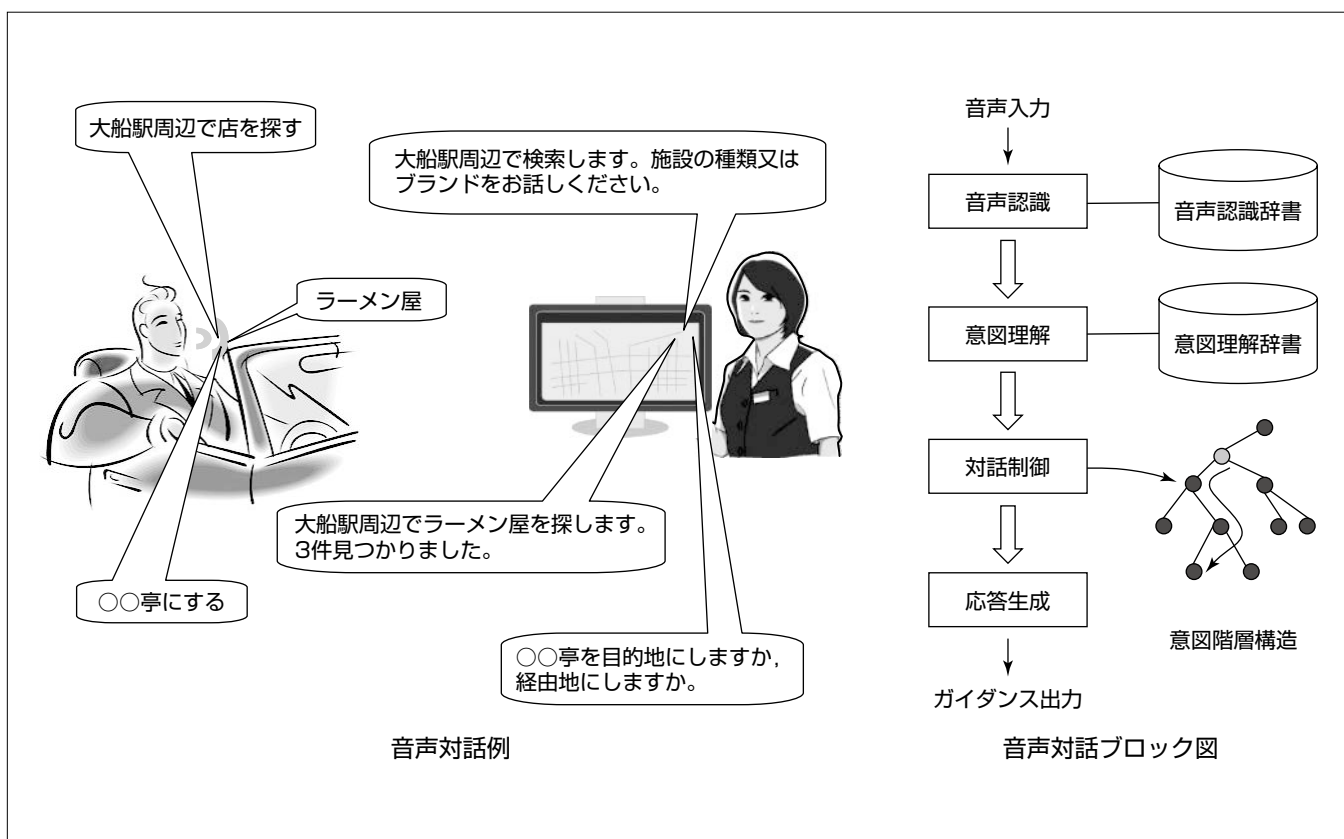
情報システム、カーナビゲーション・AV(Audio Visual)機器等の情報家電、さらにはルームエアコンなど電化製品も、多機能化が進んでいる。これらのシステム・機器のインタフェースを従来のGUI(Graphical User Interface)で実現しようとする、メニュー構造が複雑化し、ユーザーは実行したい機能に辿(たど)り着くまでに“どう操作していいかが分からない”という問題が生じる。

この課題を解決するための技術として、音声認識への期待は高い。しかし従来型の音声コマンドでは、多機能化によってコマンド数が増え、GUI同様に“何を言っているかが分からない”という問題を引き起こす。これを解決するには、自由な発話から意図を推定してシステム・機器を動

作させることが必要である。さらに、ユーザーの発話で動作に必要な情報が不足していたり、意図が曖昧であったりしたときにこれを解消するための対話技術が必要となる。

そこで、我々は、意図階層構造を用いる音声対話方式を開発した。意図階層構造は、操作対象機器に関する機能定義から生成され、上位の抽象的な意図から下位の実行可能な意図までを階層的に表現する。発話からこの構造中の意図を推定し、推定に曖昧性があれば確認し、下位の実行可能な意図に到達するように、対話が制御される。

この方式によって、複雑で多機能なシステム・機器に効率的で、自然なインタフェースを実現することができる。



音声対話の対話例とブロック図

音声対話システムでは、ユーザーが自由に発話した内容から意図(ユーザーが実行したい機能)を推定するために意図理解辞書を使う。対話制御部では、推定した意図から実行が可能な下位の意図表現に進むように対話を制御する。末端の実行可能な意図へ到達した時点で対応付けられた機能を実行することでユーザーの目的を達成する。

1. ま え が き

各種の情報システムや、カーナビゲーション(以下“カーナビ”という)・AV機器等の情報家電、さらにはルームエアコンなど電化製品にいたるまで、システム・機器の多機能化が進んでいる。これらのインタフェースを、GUIやスイッチ類で実現すると、メニューの階層が増えて複雑化するため、ユーザーが実行したい機能に辿り着くためにどう操作すればいいかが分からないという問題が生じる。さらには、どのような機能が存在するのかさえ分からない場合もある。ユーザーが機能と操作を理解するには取扱説明書を読み、何ができるかを把握し、その手順を覚える必要がある。しかし、多機能化に伴い膨大となった説明書を読むユーザーは稀(まれ)であり、多様で高度な機能の恩恵をユーザーが享受できない状態になる。

この問題を解決し、ユーザーが所望の機能を簡単に実行するための手段として音声認識への期待が高い。音声認識の場合、操作の直接的な表現を用いることができ、多くの語彙を扱えることから有効な手段になり得る。しかし、従来型の操作に対する音声コマンドを設計者視点で決めてしまう方式では、機能が増えれば、コマンド数が増え“何を言っているかが分からない”というGUIと同じ問題を引き起こすことになる。

そこで、誰でもが音声によって簡単に操作ができるよう、ユーザーの自由な発話を理解する技術による音声対話インタフェースが研究・提案されている⁽¹⁾。音声対話には、ユーザーがシステム・機器を操作して目的を達成するための課題遂行型音声対話と、対話自体を楽しんだり、対話中に情報を得ることを期待したりする非課題遂行型音声対話がある。今回の対象は、課題遂行型音声対話である。

課題遂行型音声対話では、まずユーザーの発話から意図(ユーザーが実行したい機能)の推定を行う。その上で、ユーザーが機能についての理解が不十分であったり、発話自体が曖昧であったりした場合には、その確認を行い、目的の実行に不足する情報があれば、その入力を求め、機能の実行に導くことが対話の目的となる。これを実現するために、すべての対話状況を網羅的に記述することは現実的ではない。

そこで、我々は、対象システムの機能の表現でもあり、ユーザーの対話行動の表現モデルでもある意図階層構造に基づく音声対話システムを開発した⁽²⁾。

2. 音声対話システム

2.1 音声対話システムの特長

開発した音声対話システムの特長は、次の点にある。

(1) 対象システムの機能定義を利用することで、意図の表現が容易にできる。

(2) 意図を階層化し、対話行動を表現する意図階層構造を構築することで、対話手順を記述することなしに汎用的な対話制御が可能である。

(1)の意図は、設計仕様書や取扱説明書から作成する。意図は単純な表現形式であり設計者は簡単に記述ができる。意図の間の関係性も考慮する必要があるが、これらの書類の構成や、GUI上の関係性等を参照することで、容易に記述することが可能である。ここで定義された意図は、3.1節で示す方法で階層構造化される。(2)の対話制御は、この意図階層構造を使って対話を進める。階層化は上位の抽象的な表現から下位の実行可能な意図表現を関連付けているため、対話の目的である、曖昧性の解消は、発話から意図が一意に推定できない場合の確認として、また、情報の取得は、上位意図から下位意図への遷移として実現できる。意図階層構造が対象機能並びに対話のモデルとして動作するため、対話手順を記述することなく音声対話システムが構築できる。

2.2 音声対話システムの構成

音声対話システムは、大きく分けて音声認識、意図理解、対話制御及び応答生成の4つの部分から構成される(図1)。音声認識は、ユーザーの自由な発話を受け付けることができるディクテーション型の認識エンジンを用いる。次に意図理解は、音声認識結果のテキストから、ユーザーの意図を推定・理解する。意図理解は、単語と意図の関係の統計的性質を記述した意図理解辞書によって実現する⁽³⁾。対話制御は、2.1節で述べたように意図階層構造を下位へ辿ることで、ユーザーの意図を具体化する。応答生成は、システムが次に期待する発話をユーザーに促すための応答を生成してガイダンスを出力する。この4つの処理を繰り返すことで対話を進める。最終的に、末端の意図に到達したら、その意図に割りつけられた機能を実行して対話を終了する。

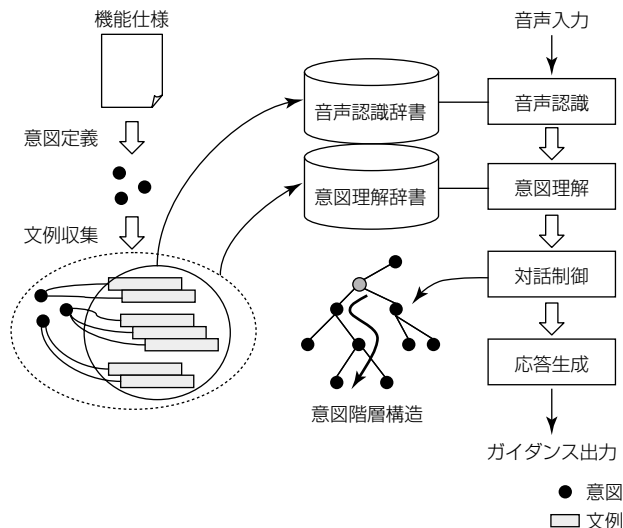


図1. 音声対話システムの構成

3. 意図の階層性を利用した対話制御方式

3.1 意図の定義と意図階層化方式

課題遂行型音声対話では、発話の目的はシステム・機器の機能を実行することであるため、これを発話の“意図”として概念化した。カーナビを例として述べる。例えば“現在地周辺のコンビニを検索したい”という意図は図2のように記述する。意図は意図の分類、又は機能を表す主部と、主部の実行に必要な情報を表すスロット部からなる。図2の例では主部が“周辺検索”で、基点というスロットの値が“現在地”，ジャンルというスロットの値が“コンビニ”である。意図は対象とするカーナビの設計仕様書，取扱説明書等の文書から抽出できる。

意図の集合が決まると、3.3節で述べる対話制御で利用するための意図の階層構造を生成する。階層構造の例を図3に示す。この意図階層構造は上位にいくほど抽象的な表現となる意図の上位・下位関係を示す構造である。意図はこの階層構造のノードであるため、構造上は意図ノードと呼ぶ。中間に位置する意図ノードはスロット情報の一部が空(スロット名=NULLと表現)のノードとなる。中間の意図ノードは最下層の意図ノードから自動で生成できる。

3.2 意図理解方式

ユーザーの自由な発話から、先に述べた意図を推定する意図理解には統計的手法を用いる。

この手法ではあらかじめ収集した発話文例から抽出した単語と、発話の意図を対応付けた学習データを用意する。例えば，“スカイツリーを目的地に設定”という文例から“スカイツリー”“目的地”及び“設定”という単語を抽出し，“目的地設定(簡単のためスロット部は省略)”という意図に

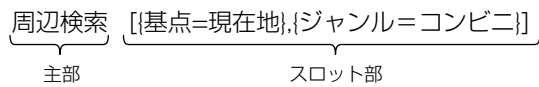


図2. 意図の例

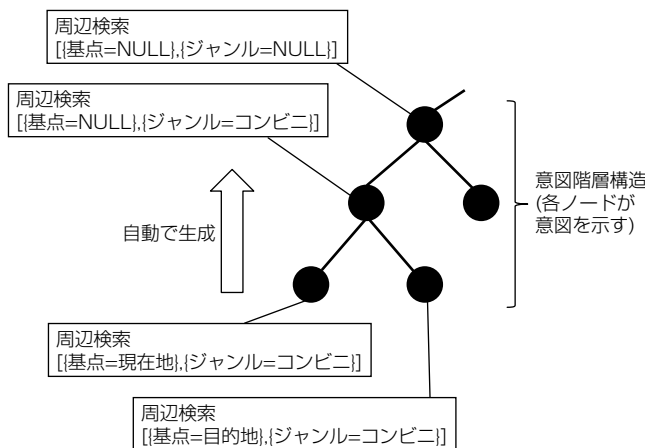


図3. 意図階層構造の例

対応付けておく。このデータから各単語が出現したときの対応付けられた意図の確率を学習し意図理解辞書とする(図4)。

実際に意図理解するときは、発話中に出現した各単語に対する各意図の確率を意図理解辞書から取得し、確率の高い意図を意図理解結果として出力する。例えば“目的地を削除”という発話の場合、意図理解辞書を参照し“目的地”に対応して“目的地設定”や“目的地削除”という意図の確率が高く、“削除”に対応して“目的地削除”や“経由地削除”という意図の確率が高い値が得られる。意図ごとの確率を掛け合わせることで“目的地削除”の確率が最も高いという結果が得られる(図5)。

この手法の特長は、学習データに含まれない単語の組合せの発話であっても、その単語が学習データにあれば意図を推定することができる点である。そのため、決められた音声コマンドではなく自由な発話から意図を推定することが可能となる。

3.3 対話制御方式

対話制御では、3.2節で述べた意図理解方式を用いてユーザーの最初の発話から、意図階層構造上の最も近い意図ノードを推定する。そこから階層を下りるための対話を繰り返し、最終的に末端の意図ノードへ遷移することでユー

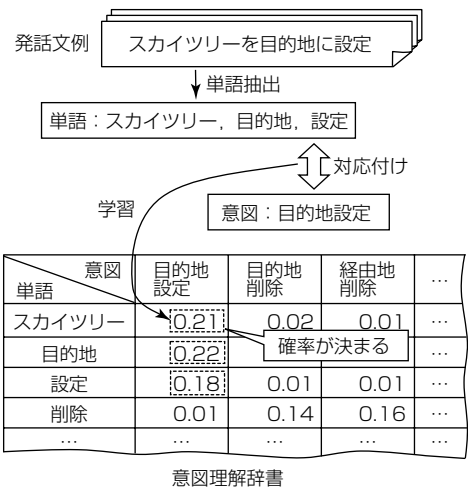


図4. 意図理解辞書の作成例

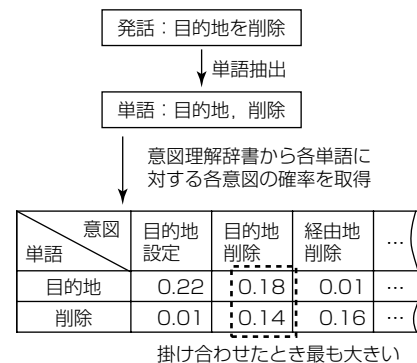


図5. 意図理解の例

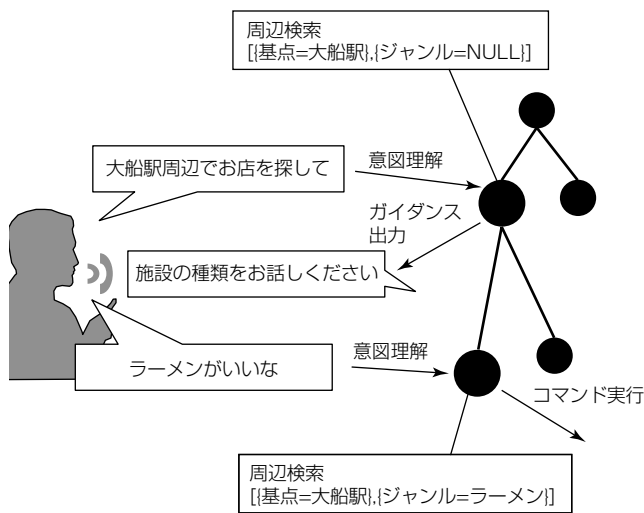


図 6. 対話制御処理の例

ザの意図を具体化し、機能を実行する。

図 6 に例を示す。例えば、“大船駅周辺でお店を探して”とユーザーが発話した場合、対応する意図ノードが“周辺検索[基点=大船駅], {ジャンル=NULL}”であると推定する。この意図ノードはジャンル情報が不明な中間ノードで、下位の意図ノードはジャンル情報が特定された意図ノードである。そのため、階層を降りるためにはジャンルの情報が必要である。そこで、システムは“施設の種類をお話してください”とガイダンスを出す。ユーザーが“ラーメンがいいな”と発話すれば、ジャンルが“ラーメン”であることが特定でき、末端のノードへ遷移し、大船駅周辺のラーメン屋を検索する機能を実行する。

この例は、ユーザーの発話意図が唯一に推定できる場合であるが、実際には発話の意図を唯一に推定できない曖昧な発話の場合もある。そのような場合には曖昧性を解消する処理を行う。具体的には、推定した複数の意図ノードから一つを選択させるための対話を行う。例えば、ユーザーが“スカイツリー”とだけ発話した場合、スカイツリーを目的地に設定したいのか、スカイツリーを登録地に追加したいのか曖昧である。このような場合、システムは“スカイツリーを目的地に設定しますか、登録地に追加しますか？”とガイダンスを出す。ユーザーが“目的地に設定”と発話すると前の発話である“スカイツリー”と合わせて意図理解を行い、スカイツリーを目的地に設定する機能を実行する。

4. 今後の展開

この音声対話方式は、多機能化する機器のインタフェースとして有効であるが、幾つか課題も明らかになっている。意図の定義を機能仕様から定義して階層化するため設計しやすくなっているが、意図理解辞書の学習のために各意図に対してある程度の規模で文例を収集する必要がある。文例データを集めるには、アンケートなどの方法をとるが、

意図と文例との対応付けは人手で行う必要があるため、開発にかかる作業量が大きくなる。文例収集及び意図対応付けが、簡単にできる方法の確立が必要である。

また、意図の定義は機器の機能から作成されるので、ユーザーが機能にない発話を行った場合には必ずしもうまく対話が行えない。例えば、“経路地の削除”と“経路地の追加”は機能として存在するが、“経路地の変更”は2つの機能の複合であるため、1機能として用意されていない場合がある。このような場合にユーザーの“経路地の〇〇を□□に変更して”という発話を、“経路地の削除”と“経路地の追加”が複合した意図であると理解できる仕組みが必要である。

さらに、この音声対話方式は、意図階層性を使って下位の階層に遷移することでユーザーの意図を具体化していくので、一度間違った意図に遷移すると1つ前に戻って別の言い方をするなど、対話が冗長になる場合がある。

今後これらの問題を検討し、より使いやすい音声対話インタフェースの実現を図る。

5. むすび

音声対話技術として、機能と対話行動のモデルである意図階層構造を利用する対話方式を提案した。この方式は、自由な発話からの意図理解に基づき、対話の目的である、機能実行のための情報の取得と、曖昧性の解消を、その手続を記述することなく実現でき、ユーザーには分かりやすく、効率的なインタフェースを提供できる。さらには、ユーザーがシステム・機器の機能について十分な理解がなくても、発話の曖昧性解消の対話によって、機能を知ることができる。したがって、機能とその操作手順を事前に学習しなくても、利用経験を通じて、機能を知り、かつ、より効率的な発声すなわち意図の表現方法も学べるのがユーザー視点での大きな特長になる。

さらに、よりユーザーの感覚に近い対話を実現するため、音声だけではなく、ジェスチャー、システムの状態、ユーザー操作履歴等の複数の入力情報・文脈情報・知識を活用できる対話方式の研究・開発を行い、自然でユーザー満足度の高いインタフェースを実現する予定である。

参考文献

- (1) 河原達也：話し言葉による音声対話システム、情報処理, 45, No.10, 1027~1031 (2004)
- (2) 藤井洋一, ほか：取扱説明書の構造を利用した音声対話の評価、電子情報通信学会総合大会講演論文集D-14-3 (2014)
- (3) 藤井洋一, ほか：自由発話による家電機器操作の多言語適用に関する検討評価、電子情報通信学会総合大会講演論文集D-14-6 (2012)