

コンテキスト情報を利用して曖昧な音声入力の意味解決をする 音声ユーザインタフェースシステム

李 海量^{†‡} 越塚 登^{†*} 坂村 健^{†‡*}

東京大学大学院学際情報学府[†] 東京大学大学院情報学環[‡] YRP ユビキタスネットワークング研究所^{*}

1 はじめに

ユビキタスコンピューティング環境のように、身の回りのあらゆるところに埋込まれた人間生活を支援するコンピュータ群に対し、人間の意志を伝える自然なユーザインタフェースとして、音声インタフェースが期待されている。音声インタフェースシステムの実用化上の課題は、ユーザの発話の曖昧性である。従来の音声認識システムは発話履歴を使ってその曖昧さを解決するものが多いが、文脈が発話履歴に現れないようなケースには対応できない。このような曖昧性の解決には、ユーザを取り囲む環境のコンテキスト情報[1]が不可欠である。

そこで、本研究の目的はコンテキスト情報を用いて音声認識率、及びユビキタス環境における音声インタフェースの利便性を向上させることにある。本稿では、まずコンテキスト情報を音声システムへ適用する基本的な枠組みを提案する。次に、適用例の一つである曖昧な音声入力の意味解決方法について述べる。静的と動的なコンテキスト情報を参照し検索することにより、音声入力から生成される不完全な意味表現を補完することにより、突発な発話や省略の意味が解決できるようになる。

2 従来の音声インタフェースシステムと問題点

一般的に、音声インタフェースシステムは、大きく分けると、音声認識部、言語処理部・応答生成部から構成される(図1左)。このようなシステムにおける課題の一つは、人間の突発の発話、あるいは省略語や指示詞などを含む曖昧の発話の理解である。例えば、「これをそこに置いて」という発話に対し、音声認識部が完璧に認識されても、それだけでは発話の意味がわからない。曖昧さを解決するために、ユーザ発話の文脈、常識、および状況を知ることが必要である。

談話・対話に関する研究は既に盛んに行われており[2]、音声以外の入力方法を含んだ音声対話システムのコンテキスト解決システムの研究もある[3]。これらの研究は主にユーザの発話履歴や開発者が記述した情報から曖昧な部分の意味を推測す

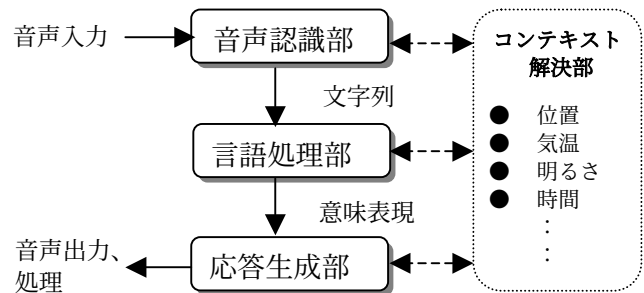


図1 一般的な音声システムの概略図、及びコンテキスト応用例のものである。

ユビキタス環境において、ユーザが環境の中の機器を操作する時に、(1)突発な発話や(2)省略、(3)指示詞を使う場合が考えられる。例えば、部屋にいるユーザが突然「つけて」という音声コマンドを発した時に、文脈だけでは意図が解決できない。また、「これ」のような指示詞についても同様である。

3 システム概要

本システムは、図1左に示した音声インタフェースシステムアーキテクチャをベースに拡張する。図1に示すように、従来の構成部分にコンテキスト解決マネージャ(CR部)を追加し4つの部分から構成される。拡張前の各モジュール内にコンテキスト解決機能を埋め込む方式を取ることもできるが、本システムではCR部の汎用性を考慮してこの方式を採用した。

3.1 CR部

CR部は、コンテキスト情報を管理し、そしてそれらの情報を利用したサービスを提供する。

管理されるコンテキスト情報を静的情報と動的情報の2種類に分けた。静的な情報はシステム作動中に変更しない情報を指し、例えば部屋にある機器とその属性、位置などをいう。これらの情報は拡張性を考えて外部に記述することにし、CR部はその情報が記述されたファイルを読み込んで利用する。一方、作動中に変更する動的情報は、例えばユーザの位置や状況、機器の状態をいいで、モジュールの中ではフィールド名と値の対を持つテーブルで管理する。これらのフィールドは各センサーや機器に定期的にポーリングして更新する。

次にCR部では、他のモジュールから呼び出される、コンテキスト情報を用いた処理をサービスとして提供する。例えば、曖昧な入力の意味を解決するサービスが、言語処理部から不完全な意味表

Context-Aware Speech User Interface System with Ambiguity Resolution

Lee Hoi Leong^{†‡}, Noboru Koshizuka^{†*}, Ken Sakamura^{†‡*}

[†]Interfaculty Initiative in Information Studies, Graduate School of Interdisciplinary Information Studies, University of Tokyo

[‡]Graduate School of Interdisciplinary Information Studies, University of Tokyo

^{*}YRP Ubiquitous Networking Laboratory

現を受け取り、コンテキスト情報を用いた補完後、より正確な意味表現を返す。

3.2 コンテキスト情報を用いた意味解決

意味解決をするために、CR 部にある意味解決をするサービスは、従来の処理の流れの中で、言語処理部が応答生成部に意味表現を渡す前に呼び出され、意味表現を補完する(図2)。意味解決サービスでは、意味表現を、(1)省略された項目の生成、(2)指示詞の解決、(3)条件式を用いる解決、の3つのステップを経て補完する。

例えば、ユーザがテレビの前で、持っているモバイル端末に表示されている画像をテレビに表示したい時「これを出して」と発話をしたとする。まず、発話に省略された項目の生成を行う。「これを出して」という発話は、言語処理部により「出す」という動作(action)と「これ」という対象(object)を含む意味表現を生じるが、動作「出す」には出力先というもう1つの項目が必要である。このように、省略項目がある場合、CR 部は静的コンテキスト情報に従って項目を追加する。動作に対する必要項目の記述(図3)を参照することにより、出力先(dest)項目が意味表現に追加され、つまり「これを〇〇に出して」の意味になる。

次に指示詞の解決を行う。もし指示詞があれば、コンテキスト情報テーブルからユーザの位置を参照し、近くにある機器を列挙する。また、action項目で指定された動作ができる機器を列挙する。具体的に、コンテキスト情報テーブルに各機器の状態を参照し、それから静的情報で記述された各々の状態で可能な動作(図4)のリストを参照する。このステップにより、例の発話にある「これ」はモバイル端末であることが判明する。

最後に、条件式を用いた解決を行う。ここで、外部に記述された条件式、例えば「ユーザがテレビの近くにいて”出す”という動作を指定するなら出力先はテレビだ」などを使って、なるべく残りの曖昧さを補完する。この3つのステップを経て、例の入力が「モバイル端末をテレビに出して」という意味と解決される。

なお、入力によって必ずしも3つのステップとも必要であるとは限らない。また、全部通した後でも曖昧さ残っている場合に次の応答生成部でユーザに不足な情報を求める場合もある。指示詞の解決は条件式を用いる解決の前に行わなければならない理由は、指示詞を含む条件式がないので、指示詞の解決が先に行った方が条件式の適用に役立つ情報が増えるからである。

4 考察

以上のアルゴリズムにより、会話履歴だけでは不十分だったユビキタス環境にある機器を操作する際における突発的な発話や指示詞、省略を解決できることがわかった。実際にコンテキスト情報

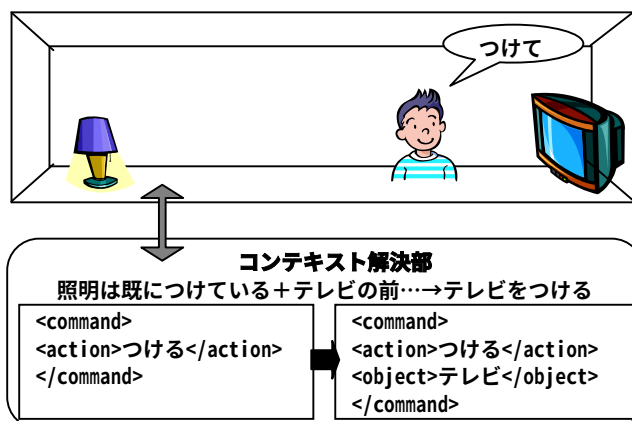


図2 コンテキスト情報を用いた補完の例

```
<require>action = 出す >> object, dest</require>
```

図3 必要項目の記述

```
<device>
  <name>PDA1</name>
  <states>
    <state><name>OFF</name>
      <paction>つける、オン</paction>
    </state>
    <state><name>ON</name>
      <paction>消す、オフ、出す、表示する</paction>
    </state>
    <connection>OFF >> ON</connection>
    <connection>ON >> OFF</connection>
  </states>
</device>
```

図4 モバイル端末についての静的情報記述

をどのように利用することを外部で記述することにより、システムの挙動を簡単に変更し、ユーザにカスタマイズすることもできる。その一方、CR部の挙動をすべて手動で記述するのは、機器が多い場合に状態や属性、条件式の記述が大変であり、大きいシステムを作るには手間がかかる。

5 まとめ

本稿ではコンテキスト情報を用いたユビキタス環境における音声システムを紹介し、及びそれを用いた曖昧な入力の意味解決手法について述べた。発話者のいる環境のコンテキスト情報を用いることによって、今まで難しかったユビキタス環境の発話に多い突発的な発話や指示詞の意味を解決することを可能になることがわかった。

謝辞 坂村研究室の鶴坂智則氏、新堂克徳氏、小林真輔氏の貴重なアドバイスを深く感謝します。なお、本研究の一部は、通信・放送機構の援助を受けています。

参考文献

- [1] Anind K. Dey, "Understanding and Using Context", Personal and Ubiquitous Computing Journal, Volume 5(1) pp.4-7
- [2] M. Eckert, M. Strube, "Resolving Discourse Deictic Anaphora in Dialogues", EACL99, pp 37-44
- [3] E. Filisko, S. Seneff, "A Context Resolution Server for the Galaxy Conversational System", EUROSPEECH2003, pp. 197-200