1O2-OS-18-5

移動ロボットに対する指示音声の状況依存逐次理解

Situated Incremental Understanding of Spoken Instructions to Mobile Robots

佐藤 隼[*1*2] Shun Sato　中野 幹生[*2] Mikio Nakano　Antoine Raux[*3] Antoine Raux　船越 孝太郎[*2] Kotaro Funakoshi　竹内 誉羽[*2] Johane Takeuchi

[*1]東京電機大学大学院理工学研究科情報学専攻
Graduate School of Science and Engineering, Tokyo Denki University

[*2](株) ホンダ・リサーチ・インスティチュート・ジャパン
Honda Research Institute Japan Co., Ltd.

[*3]Honda Research Institute USA, Inc.
Honda Research Institute USA, Inc.

Robots that receives spoken instructions need to select actions based on not only the contents of the instructions but also situations. They are also expected to immediately react to the instructions. This paper presents a method that incrementally understands spoken interactions and immediately controls a mobile robot based on the understanding results and situation information such as the locations of obstacles and moving history. The proposed method was applied to a conversational robot that moves in a 3D simulation environment.

1. はじめに

移動ロボットは，場所を移動してタスクを遂行することができるロボットである．そのようなロボットに対して，音声言語で移動指示ができれば，使い勝手がよくなり，様々な用途に用いることができると考えられる．しかしながら，音声言語は曖昧であり，ロボットや障害物の位置などの状況によってその意味が異なるため，音声言語による指示をロボットに理解させるのは必ずしも容易ではない．さらに，移動中に受けた指示を理解する際，一般的な音声理解のように，指示者の発話が終了するのを待ってから理解を行っていては，ロボットの移動などで，状況が変わっているため，正しく理解できない．

言語による移動指示を受けるロボットの研究は従来からあるが [Marge 10, Tellex 11]，移動前に指示を理解する方法がほとんどで，移動中の指示を理解することや，ダイナミックに状況が変化する場合の指示の理解は扱われていない．

本稿では，ダイナミックな状況変化が起こる場合に，状況に依存して指示を理解する方法を提案する．指示の行われた時点の状況をもとに理解を行うため，発話終了を待つのではなく，逐次的に音声理解を行う．その結果と，ロボットの移動履歴や周りの障害物などの情報から，次にロボットがどう動くべきかを制御する．提案法を 3 次元シミュレーション環境で動作する対話ロボットシステムに適用し，動作確認を行った．

2. 問題設定

2.1 3 次元仮想空間内の移動ロボット

本研究では，SIROS と呼ばれる 3 次元シミュレーション環境で動作するロボットシステム [4] を用いる．SIROS は，オンラインゲームのように，同じ仮想環境を複数の人間が共有し，対話をしながらタスクを遂行することを可能にする．具体的な環境として，コンビニエンスストアを用いる．

連絡先: 〒 351-0188 埼玉県和光市本町 8-1 (株) ホンダ・リサーチ・インスティチュート・ジャパン，中野 幹生，E-mail: nakano@jp.honda-ri.com

図 1: マネージャからみたコンビニエンスストア環境

ロボットは，コンビニエンスストアの店員として，マネージャの音声による指示を受けながら，ランダムに店内に入ってくる客の要求に答えてタスクを遂行する．タスクはコーヒーマシンの修理や，レジでの精算などであるが，簡単のために，対象物体を一定時間クリックし続けることでタスクが遂行できるとしている．マネージャはコンビニを俯瞰することができるが，ロボットの画像処理技術をシミュレートするため，ロボットの視界は制限されている．マネージャの主な仕事は，ロボットがタスクをこなせるように移動するための指示を出すことである．ロボットは，マネージャの指示と制限された視界を頼りに店内を移動し，コーヒーを入れたりレジで精算したりするといったタスクをこなす． 図 1 にマネージャ視点の画像の例を示す．

この環境を用い，マネージャ役とロボットをキーボードで操作するオペレータ役の二人の人間が会話をしながらタスクをこなす様子を収録したコーパス（コンビニコーパス，英語・日本語）が構築されている [Raux 10, 川端 11].

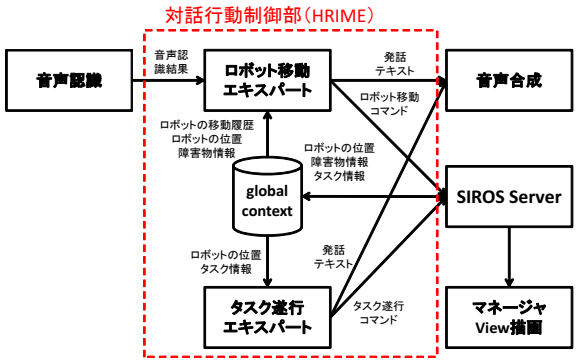

図 2: システム構成

2.2 移動ロボットの自動化

本研究の目標は，オペレータ役の人間の代わりに，マネージャの指示に応じてロボットを移動させるソフトウエアシステム（自動オペレータと呼ぶ）の構築を行うことである．本研究では，移動指示に焦点をあてるため，マネージャの指示にしたがって，客のところに移動するタスクのみを行うこととした．

自動オペレータの構成を図 2 に示す．音声認識には，Sphinx4[Lamere 03], 音声合成には NTT-IT の FineVoice を用いている．対話行動制御にはマルチドメイン対話行動エージェント構築プラットフォーム HRIME (HRI Intelligence Platform based on Multiple Experts) [Nakano 11] を用いる．Sphinx4 用の日本語音響モデルは，「日本語話し言葉コーパス」[前川 04] を用いて構築した．

対話行動制御部の中では，エキスパートと呼ぶ，特定のタスクに特化したモジュールが適宜駆動されて処理を行う．ロボット移動エキスパートは，マネージャの指示発話の認識結果を受け取ると駆動され，ロボットの移動コマンドを出力する．指示発話の意味は “右に曲がる”，“左に曲がる”，“まっすぐ進む”，“後ろに進む” ，“１つ前の動作を繰り返す”，“１つ前の動作とは反対の動作をする”，“止まる”，“その他” の 8 つのどれかと仮定する．“その他” の場合は無視する．指示発話の理解は指示の言語表現と意味の対応を用いて行う．例えば “右”は “右に曲がる” に，“戻って”，“戻れ”，“違う” は，“１つ前の動作とは反対の動作をする” に対応する．移動を開始する場合や，障害物があって進めない場合に，音声でマネージャに報告する．これは発話文字列を音声合成部に送ることによって行う．

ロボットに対する移動コマンドは SIROS サーバに送られる．SIROS サーバは人間のオペレータがロボットをキーボードで操作した時と同様に，ロボットを動作させる．また，その結果がマネージャの見ている画面に反映される．

本研究では移動指示に焦点を当てるため，客に近づいた段階で，タスク遂行は終了するとした．この処理は，タスク遂行エキスパートと呼ばれるモジュールが行う．このエキスパートはロボットが客に近づいた時に駆動される．

対話行動制御部の中の global context は，環境の情報を SIROS サーバから得て，エキスパートが利用できるようにする．しかし，人間同士のインタラクションの際と同じように，ロボットの視覚を制限するため，ロボットから近い客や障害物の情報しか保持していない．

3. 状況依存逐次理解

実際のマネージャの発話は，上記のパターンに当てはまるものだけではない．“右右そのまま前に進んで左”のように切れ目なく続くことがある．しかも，その間にもロボットや客は移動し，状況は刻々と変化する．そこで，発話終了（一定長以上のポーズ）まで待たず，逐次的な音声認識を用い，その結果に移動指示があればロボットを動作させる方法が考えられる．しかしながらこの方法では，逐次的な音声認識の結果は不安定で誤認識もあること，および，移動指示だけでは移動量（どのくらい右に回るか，どのくらい前に進むか）が不明なため，ロボットの動作がスムーズではない．したがって，逐次音声認識結果のみに基づいてロボットを動作させるのではなく，周りの障害物の状況，移動履歴情報を同時に用いる．

提案手法では，逐次音声認識結果を受け取り，指示が得られたら，状況の情報を用い，それが誤認識によるものかどうかを判定し，そうでなければ認識した結果の動作を状況に合わせて変換する．具体例として，次のような理解行動規則を考案した．

- 認識結果の方向に障害物が存在したら，その認識結果を無視する

  例えば，“前”と認識したのに，前方に障害物があったらそれは誤認識であると推測できる．

- 障害物と平行になるように角度を調整する

  スムーズに通路を進むことが可能になる．

- 通路の入口を通り過ぎた後に，その通路に入るような音声を認識したら，その通路に入る

  例えば，ロボットの右手側に通路があったとして，それを通り過ぎてから “右”と認識した時に，その通路に入るようにロボットを動作させる．

- 障害物が無いところまで曲がるように角度を調整する

  例えば，“右”と認識した時に，右手側 55 度まで障害物があったとしたら，右 55+ α 度曲がるようにする．

4. 実装

提案手法を，2.2 節で説明したシステムに実装した．音声認識用言語モデルは，日本語コンビニコーパス（18 対話，各対話 10 分）のマネージャの発話を用いて作成した trigram 言語モデルである．学習データは，日本語コンビニコーパスから独り言，意味のない発話，笑い声を手作業で取り除いたものとした．学習データ数は 1,101 発話， 異なり単語数 935 となった．

Sphinx4 で逐次認識結果を出力するために，Baumann の拡張ツール [Baumann 10] を用いた．これにより，10msec ごとに認識結果を出力することが可能になる．認識結果は，発話の最初からその時点までの部分認識結果として得られる．対話行動制御部は，音声認識結果の一番最後の単語を確認し，その単語がロボットを動作させる単語であれば，前節の規則に基づいてロボットを動かす．

図 3 に動作例を示す．マネージャの発話中であってもロボットは発話を理解し応答している．ロボットの動作の開始が遅れているが，これは処理の遅延によるものである．また，左に向く角度は，レジに合わせて調整されており，その後の前進にスムーズに入れている．

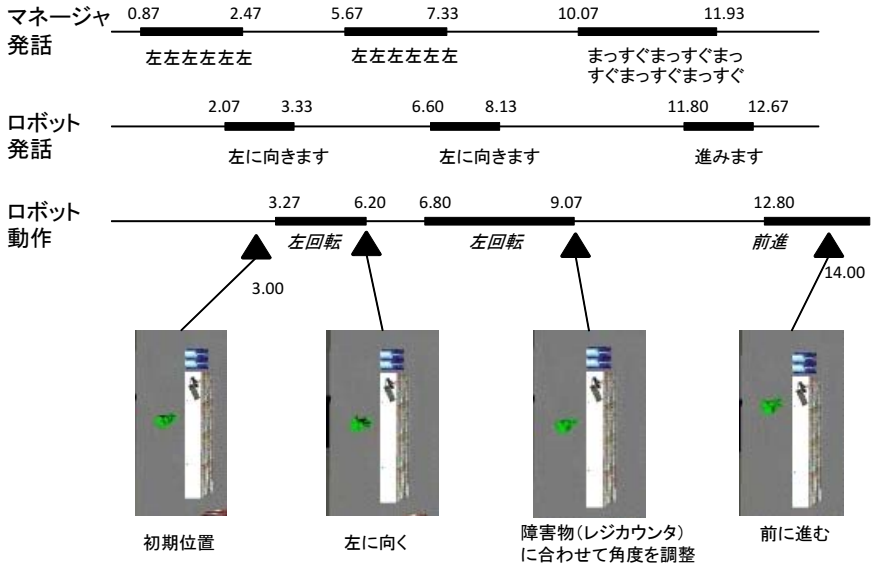

図 3: 動作例（時間の単位は秒）

5. おわりに

本稿では，移動指示の理解を対象として，ダイナミックに変化する状況において，ロボットが音声指示を逐次的に理解して動作する方法を提案し，その初期実装に関して述べた．

今後は，動作のリアルタイム性の向上を行うとともに，様々なユーザが指示を行った場合や，より複雑な環境で動作させた場合でも有効かどうかの検証を行う．さらに，様々な環境において最適な行動が行えるよう，理解行動規則の自動獲得の研究を進めていく．

参考文献

[Baumann 10] Baumann, T., Buß, O., and Schlangen, D.: InproTK in Action: Open-Source Software for Building German-Speaking Incremental Spoken Dialogue Systems, in *Proceedings of ESSV*, Berlin, Germany (2010)

[川端 11] 川端 良子, 中野 幹生, 船越 孝太郎, Raux, A.：状況共有下の経路指示課題における指示ストラテジの分析, 人工知能学会研究会資料 SIG-SLUD-B102, pp. 7–12 (2011)

[Lamere 03] Lamere, P., Kwok, P., Walker, W., Gouvea, E., Singh, R., Raj, B., and Wolf, P.: Design of the CMU Sphinx-4 decoder, in *Proc. Eurospeech-2003* (2003)

[前川 04] 前川 喜久雄：『日本語話し言葉コーパス』の概要, 日本語科学, Vol. 15, pp. 111–133 (2004)

[Marge 10] Marge, M. and Rudnicky, A. I.: Comparing Spoken Language Route Instructions for Robots across Environment Representations, in *Proc. of the SIGDIAL 2010 Conference* (2010)

[Nakano 11] Nakano, M., Hasegawa, Y., Funakoshi, K., Takeuchi, J., Torii, T., Nakadai, K., Kanda, N., Komatani, K., Okuno, H. G., and Tsujino, H.: A multi-expert model for dialogue and behavior control of conversational robots and agents, *Knowledge-Based Systems*, Vol. 24, No. 2, pp. 248–256 (2011)

[Raux 10] Raux, A. and Nakano, M.: The Dynamics of Action Corrections in Situated Interaction, in *Proc. of the SIGDIAL 2010 Conference*, pp. 165–174 (2010)

[Tellex 11] Tellex, S., Kollar, T., Dickerson, S., Walter, M. R., Banerjee, A. G., Teller, S., and Roy, N.: Understanding Natural Language Commands for Robotic Navigation and Mobile Manipulation, in *Proc. 25th AAAI* (2011)