5Q-6

# ロボットに対する移動指示の状況依存逐次理解

佐藤　隼[†]　中野　幹生[‡]　Antoine Raux[††]　船越　孝太郎[‡]　竹内　誉羽[‡]
[†]東京電機大学大学院 理工学研究科 [‡](株)ホンダ・リサーチ・インスティチュート・ジャパン
[††]Honda Research Institute USA, Inc.

## 1　はじめに

音声指示を受けるロボットは，指示音声だけではなく，音声指示が行われた時の状況に依存して，指示の内容を理解する必要がある．ロボットが移動している途中に音声指示が行われた場合には，その時点での状況に依存して指示を解釈する必要がある．

音声指示，特に移動指示を受けるロボットの研究は従来からあるが[6]，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．

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．

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　3次元シミュレーション環境での移動指示発話の理解

本研究では，SIROS と呼ぶ 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．具体的な環境として，コンビニエンスストアを用いる．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るために必要なタスクを行う．マネージャはコンビニを俯瞰することができるが，ロボットの視覚は，現状のロボット視覚技術をシミュレートするため制限されている．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にマネージャの視野の例を載せる．

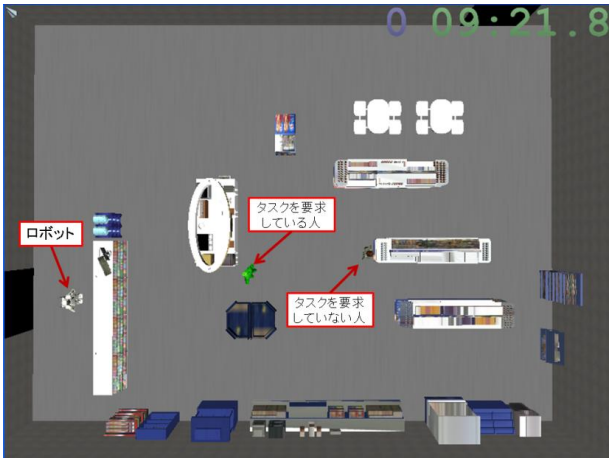


図１．マネージャから見たコンビニ環境

この環境を用い，マネージャ役とロボットをキーボード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英語・日本語のコーパス（コンビ二コーパス）が収録されている[5]．

本研究では，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）の構築を行った．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．

自動オペレータは，音声認識と対話行動制御部からなり，音声認識には，Sphinx4[2]，対話行動制御には HRIME[3]を用いた（図２）．対話行動制御部は，マネージャの指示発話の認識結果を受け取ると，それに応じてロボットを動作させる．指示発話は７種類あり，「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」である．それぞれの動作にはいくつかの言語表現が割り当てられている．例えば，「右に曲がる」には「右」が，「１つ前の動作とは反対の動作をする」には「戻って」，「戻れ」，「違う」が割り当てられている．

Situated Incremental Understanding of Direction Utterances of Robots
Shun Sato[†] Mikio Nakano[‡] Antoine Raux[††]
Kotaro Funakoshi[‡] Johane Takeuchi[‡]
[†] Graduate School of Science and Engineering, Tokyo Denki University
[‡]Honda Research Institute Japan Co., Ltd.
[††]Honda Research Institute USA, Inc.

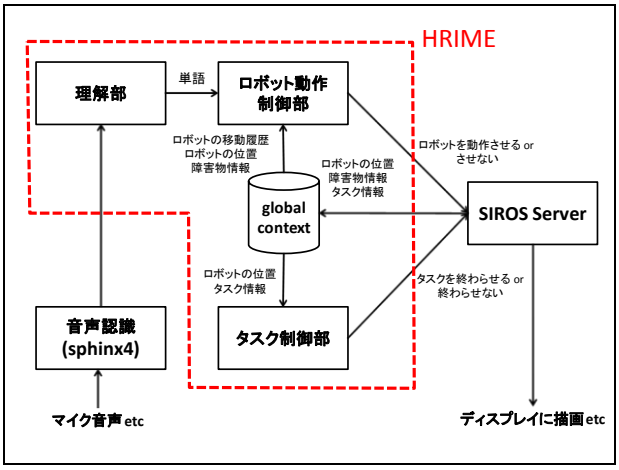


図２．システムの構成

## ３．提案手法

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．「右右向いてあっ左向いて下さい左」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選択する．

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，その認識結果は無視
  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
  スムーズに通路を進むことが可能になる．
- 小道を通り過ぎた後に，その小道に入るような音声を認識したら，その小道に入る
  例えば，ロボットの右手側に小道があったとして，それを通り過ぎてから「右」と認識した時に，その小道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
  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 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から独り言，意味のない発話，笑い声を取り除いたものを用いて作成した trigram 言語モデルである．学習データ数は 1,101 発話， 単語数 935 単語である．

Sphinx4 で逐次理解結果を出力するために，Baumann の拡張ツール[1]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った．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後ろの単語を確認し，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．

本実装により，逐次音声認識を用いながらもスムーズにロボットを動作させることができることを確認した．

## ５ おわりに

本稿では，移動指示の理解を対象として，ロボットがダイナミックに変化する状況において指示を逐次的に理解しながらも，スムーズに動作する方法を提案した．逐次理解技術は主に発話交代をスムーズにするために研究が進められてきたが，本研究は逐次理解の新しい効用を示したと考える．

今後は，被験者実験による評価および，理解行動規則の自動獲得の研究を行う予定である．

## 参考文献

[1] T. Baumann et al: InproTK in Action: Open-Source Software for Building German-Speaking Incremental Spoken Dialogue Systems. *ESSV* (2010)
[2] P. Lamere et al.: Design of the CMU sphinx-4 decoder, *Interspeech* (2003)
[3] M. Nakano et al.: A multi-expert model for dialogue and behavior control of conversational robots and agents. *Knowl.-Based Syst.* 24(2): 248-256 (2011).
[4] A. Raux: SIROS: A Framework for Human-Robot Interaction Research in Virtual Worlds *AAAI Fall Symposium on Dialog with Robots* (2010)
[5] A.Raux and M. Nakano: The Dynamics of Action Corrections in Situated Interaction. *SIGDIAL* (2010).
[6] S. Tellex et al.: Understanding Natural Language Commands for Robotic Navigation and Mobile Manipulation. *AAAI* (2011).