

修士論文

知識メディアとしてのロボット
一人間に知識を効果的に伝える
プレゼンターロボット

指導教員 西田 豊明 教授

京都大学大学院情報学研究科
修士課程知能情報学専攻

大谷 拓

平成19年2月9日

知識メディアとしてのロボット

一人間に知識を効果的に伝えるプレゼンターロボット

大谷 拓

内容梗概

近年、多くのヒューマノイドロボットが登場し、単なる賢い道具ではなく、人間とのインタラクションを通じて知識のやりとりを行う社会的なメディアとして扱うことが重要になってきた。本稿では、人間に知識を伝えるメディアとしてのロボット、プレゼンターロボットについての検討結果を示す。プレゼンターロボットは、マニュアル書では伝え難い動きやコツといった知識をハンドに装着したディスプレイを通して作業者に対して説明映像コンテンツという形で提示する。自転車の分解・組立作業を例として取り上げ、ノンバーバルなコミュニケーションを用いた作業支援を行うプレゼンターロボットに関する課題と、その解決手法であるモーションキャプチャとベイジアンネットを組み合わせる枠組みについての検討結果を報告する。

プレゼンターロボットはモーションキャプチャから得られるユーザの位置情報からベイジアンネットによりユーザの状態 (Working mode, Adjusting mode, Moving mode) を推定し、推定したモードに応じて、プレゼンターロボットとしての作業支援動作を行う。具体的な動作として、作業者と適切な位置関係を取る、作業箇所に応じて適切な映像コンテンツの選択を行う。また同時に作業者の視線や姿勢に応じたディスプレイ位置の自動調整と作業者の直接的な手動調整とを組み合わせることによりあらゆるユーザに対応できるシステムとなる。さらにシステムの根底として重要となる人間との自然なインタラクション動作として目線を合わせる、共同注意を行うなどの行動をベイジアンネットの内部状態である注目確信度に応じて起こすことにより、ユーザにとって役に立つばかりでなく、親しみやすいシステムになると期待できる。

そして実装したプレゼンターロボットシステムに対して親しみやすさと作業効率の向上という観点で、仮説と検証実験からなる評価実験を固定ディスプレイと比較して行い、実験の結果からプレゼンターロボットにより、作業者に対して親しみやすさを感じさせることはできなかったが、作業者に知識が効果的に伝わり作業効率の向上がなされることが示された。

Towards Robot as an Embodied Knowledge Medium

– Development of a presenter robot that can efficiently convey knowledge to humans

Taku OHYA

Abstract

In order to make a robot communicate knowledge effectively to humans, it is necessary to enable robot express knowledge while interacting with humans. In this paper, we describe the method of designing a presenter robot to function as a knowledge medium. We employ the assembling/disassembling of a bicycle as our research task and present the robot as a novel device for communicating knowledge: the robot plays tutorial videos with a display installed on its arm, thereby functioning as an alternative to standard operating manuals. There are several requirements for the presenter robot about a specific task and we propose a method of combining motion capture system and Bayesian net model to solve those requirements.

The presenter robot can choose its proper behaviors by estimating communication mode (working, adjusting, or moving mode) through calculating Bayesian net model. Specifically, the presenter robot can shift to a suitable position depending on the user's current position, and can play tutorial videos that will be match the user's requirement, and can adjust the position of the display to be in proximity to the user's work area. Two ways of adjusting the display by the system automatically and by the user's hand directly can be adapted to different users. Additionally, the presenter robot can interact with a user by utilizing internal states of Bayesian net model to make eye contact and joint attention. These interactive behaviors will enable the system not only to be useful but also to be friendly.

We implemented the presenter robot system and conducted experiments to evaluate whether the system is capable of interacting with humans in a natural manner, and whether the system can convey knowledge to humans. The results of the experiment could not prove the familiarity of the system but proved the efficiency of the system.

知識メディアとしてのロボット 一人間に知識を効果的に伝えるプレゼンターロボット

目次

第1章	はじめに	1
第2章	知識伝達に関する研究	3
2.1	ロボットアームによる作業支援	3
2.2	HMD/ARによる情報提示	4
2.3	スライドのプレゼンを行うロボット	5
2.4	従来研究の課題と本研究の位置づけ	6
第3章	社会的なメディアとしてのロボット	7
3.1	社会におけるロボット	7
3.2	身体性をもつ社会的なメディア	7
3.3	リスナーロボット	8
3.4	プレゼンターロボット	9
3.4.1	人間の教示者の指導行動パターン	10
3.4.2	プレゼンターロボットの場合	10
3.5	プレゼンターロボットの特徴	11
第4章	プレゼンターロボットとしての課題	14
4.1	プレゼンターロボットに求められる課題	14
4.2	想定されるタスクとシナリオ	14
4.2.1	タスク	14
4.2.2	シナリオ	15
4.3	プレゼンターロボットの動作	16
4.3.1	作業者付近への移動	16
4.3.2	ディスプレイを見やすい位置に移動	16
4.3.3	ユーザの求める映像を再生	16
4.3.4	自然なインタラクション動作	17
第5章	プレゼンターロボット実現のための技術	18
5.1	プレゼンターロボットシステムのアーキテクチャ	18

5.2	ベイジアンネットによる状態推定法	19
5.3	注目対象の検出	21
5.4	安定した台車の移動と適切な映像の選択	22
5.5	ディスプレイ位置の調整	23
5.6	ソーシャルインタラクション機能	24
第6章	システムの実装	26
6.1	テストベッド	26
6.2	モーションキャプチャシステム	27
6.3	システムの全体図	28
6.4	説明映像コンテンツの作成	28
第7章	システムの評価	30
7.1	システムの評価のための仮説	30
7.2	仮説を検証するための実験	30
7.3	実験結果と評価	31
7.4	実験の考察	34
7.4.1	ベイジアンネットの妥当性について	34
7.4.2	親しみやすさについて	35
7.4.3	作業効率について	35
7.4.4	作業に関する主観評価について	36
7.4.5	ディスプレイ位置の調整について	37
第8章	親しみやすさの向上	38
8.1	PhaseSpace	38
8.2	声について	39
8.3	作業手順の判定	40
8.4	作業手順毎の行動	40
8.5	今後の課題	42
第9章	結論	43
	謝辞	44
	参考文献	45

第1章 はじめに

近年、多くのヒューマノイドロボットが登場し、単なる賢い道具ではなく、協力能力を持つ仲間として、人間とコミュニケーションをとることが必要になってきた。また、ロボットは搭載されたセンサーにより自分自身が記録媒体として画像や音声の情報を蓄積し、学習を通じて知識を獲得できるようになりつつある。そのため、ロボットを人間と人間の間で知識を伝達するメディアとして捉えることも重要となってきた。ロボットにはソフトウェアエージェントと違って身体性があるため、人間の体験や経験を無理に書き言葉に変換したりせず、そのまま自然な姿で伝えられる可能性がある。

我々のグループでは、人間と自然なコミュニケーション能力を持つロボット、ウェ이터ロボット、掃除ロボット、リスナーロボットを研究開発してきた。これからの研究目標は、リスナーロボットと対になるプレゼンターロボットを開発し、人間と人間の間をつないで、人間の体験をキャプチャし、別の人に伝える役割をもつ知識メディアとしての役割を持つロボットメディアシステムの枠組みを示すことである。

本稿では、人間に知識を伝えるメディアとしてのロボット、プレゼンターロボットについての検討結果を示す。まず具体的な一つの例として自転車の組み立て、分解作業を取り上げ、その作業を補助するためのマニュアル書に代わる手法として、ロボットのハンドに装着したディスプレイによる知識伝達法についての検討結果を示す。ユーザの視線や姿勢に応じて見やすい位置にディスプレイを移動させ、ユーザが求めていると推定される説明映像を適切なタイミングで見せることにより、マニュアル書では伝えることが困難な動きやコツといった知識が効果的にユーザに伝わり、作業効率の向上が期待できる。そして実装したシステムについて仮説と検証による評価実験を行い、作業効率の向上に関してプレゼンターロボットシステムが有効に働くことを示した。

論文の構成は以下の通りである。まず、2章にて知識伝達に関する研究を紹介したあとに、本研究の位置づけについて述べる。3章にて社会的なメディアとしてのロボットについて述べ、さらに本研究の目的であるプレゼンターロボットについて紹介する。そして4章にて具体的なタスクにおけるプレゼンターロボットの課題を述べ、5章にて本研究の解決手法、6章にて具体的な実装方法を述べ、そして7章にて実装したシステムの評価実験および実験の考察をする。

さらに 8 章にて、実験の考察からの発展課題への取り組みを述べ、最後に 9 章にてまとめと結論を述べる。

第2章 知識伝達に関する研究

本研究で扱う作業支援における知識伝達に関する研究には様々なものがあり、本章では、その代表的な手法であるロボットアームによる方法、HMD/ARによる方法、ヒューマノイドロボットを用いたプレゼンに関する研究の紹介とそれらの問題点について述べる。さらに、本研究の位置づけについて章の最後にて述べる。

2.1 ロボットアームによる作業支援

小川原ら [1] は、ロボットによる作業者の作業支援システムとして、図1に示すようにロボットアームを用いた研究を行っている。自動で作業者への作業を支援するシステムを考える場合、支援するタイミングも一つの重要なポイントとなるが、小川原らは作業者の視線情報の履歴 (意図を内包した生理的情報) から HMM(隠れマルコフモデル) を用いて、迷いの状態を検出し、さらにシステムの持つ作業に関する知識をもとに作業者に対して、適切な道具を見せる、作業の代行などを行うシステムの開発をしている。

ただし、評価が開発しているシステムの部分的な評価に留まっており、システム全体としての評価を行う必要がある。また、迷いの推定に必要なパラメータ群が作業者のくせや作業のフェーズに応じて変わるため、動的に変更可能なシステムにする必要がある。さらに、ユーザに先だて動作を行うロボットアームによる動作は非常に危険であり、作業者の視線だけでなく動的に変わる手や体の動きなども同時に計測するシステムにする必要がある。

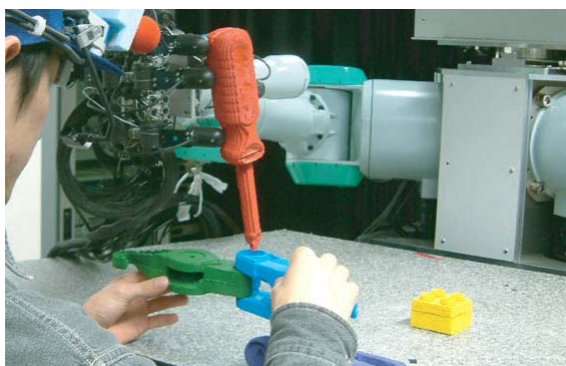


図1: ロボットアームによる作業支援

2.2 HMD/ARによる情報提示

ウェアラブルインタフェースの小型化、軽量化に伴い、情報を提示する方法として、HMD(Head Mount Display)による方法が多く研究されている。酒田ら [2] も議論しているが、従来に比べ軽量化してきたHMDではあるが視野が狭まるなど依然として作業者の装着負担が大きいという問題点がある。

和田ら [3] は、図2に示すように、シースルーディスプレイとVR技術の応用技術であるAR(Augmented Reality)技術を組み合わせた卓上タスクの作業支援の研究を行っている。シースルーディスプレイを通して、基板タスクの作業手順を基板上に重ね合わせて投影することにより、面倒なマニュアル書を読む手間を省くことができる。シースルーディスプレイを使うことによりHMDによる装着負担をなくし、できるだけ通常の作業環境に近い環境での作業が可能となっている。他にもプロジェクタを対象物に投影する方法 [4] を用いた研究などがある。

AR技術は現実の空間にVR空間の物体や情報を投影するため、作業領域とVR空間との位置合わせをしなければならぬが、作業者に装着したカメラからの入力だけで3次元空間座標を割り出すのは難しく、現状では、卓上タスクなどの平面座標に置き換えることのできるタスクがそのほとんどである。

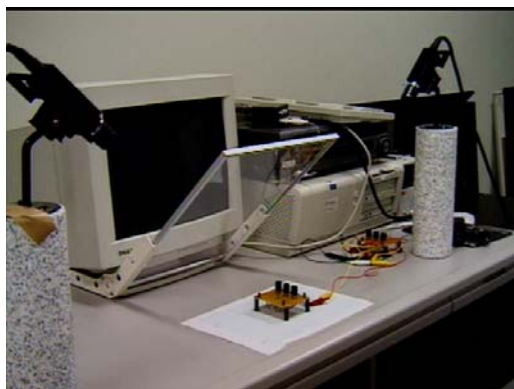


図2: AR技術による作業支援

HMD/ARともに言えることであるが、遠隔地に支援者がおらずシステムが全て自動で情報提供をしてくれる場合、システムの実態が作業者に見えないため、作業者から積極的にシステムに意図表示を行うジェスチャなどの身体的インタラクションを行うのが困難であると言える。音声認識技術による音声入力

を用いれば、自然にシステムと対話できる可能性がある。しかし、ブロックの組立のような簡単な卓上タスクであれば問題ないが、実際にはタスクを明示的な言語で表現するのが難しく必ずしも作業者の意図を言語だけで表現できるとは限らない。

2.3 スライドのプレゼンを行うロボット

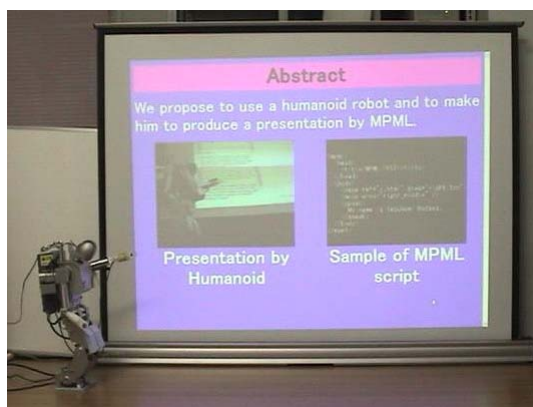


図3: プレゼンテーションを行うロボット

情報伝達方法として、顔の動きやジェスチャなどの身体表現による方法も、人間同士の情報伝達において重要な方法の一つであると言える。それを実現することができるのは、人間と似た身体構造を持つヒューマノイドロボットであり、ヒューマノイドロボットを用いた作業支援、知識伝達の研究も様々に行われている。

野沢ら [5] は、図3に示すようなスライドのプレゼンテーションを行う二足歩行ロボットの研究を行っている。スライドに応じて多様な表現動作を生成するためにMPML(Multi modal Presentation Markup Language)を用いている。MPMLを用いることにより、ロボットのプログラム開発に詳しくない人でも動作生成をするのが容易になる。

2Dのアニメーションキャラクタによるプレゼンテーションと比較実験を行うことにより、視聴者に与える印象がロボットによる方が強いことが分かったが、ロボットの身体的制約による動きの限界やスピードの限界などのマイナスの要因もあり一概にロボットによるプレゼンテーションが優れているとも言えないと報告されている。また、作られる動作は全て予め決められた動作に限られる

ため、多様なスライドによるプレゼンテーションや、視聴者からの質問やざわめきなどで動的に変化する動作生成などへの対応が必要となる。ヒューマノイドであることをもっと活かし、人とのインタラクションを組み入れたプレゼンを行えば 2D のアニメーションキャラクタよりも、臨場感のあるプレゼンが行えると期待される。

2.4 従来研究の課題と本研究の位置づけ

以上のように知識伝達を行うシステムには様々な研究がなされているが、それらの研究における共通の大きな課題としては、知識を伝達する枠組みと、その前提となる知識を獲得する枠組みとを切り離して扱った研究であるという点にある。それはつまり、作業手順が明確に表現されやすい机上タスクや、スライド毎に用意された動作をロボットが行うなど、限定された環境での作業支援に限るため、知識の再利用により様々なタスクに応用するといった汎用性に欠けるという点にある。知識を伝達することを考える場合に、いかに効果的に知識を伝えるかについて考えるのももちろん重要ではあるが、その知識をどのように獲得するかという知識の再利用についても議論することが重要となる。

さらにシステムから一方的に支援するのではなく、人間同士でも行うようなバーバル・ノンバーバルによるコミュニケーションを用いた作業支援を行うことも重要である。

そこで、知識獲得から知識伝達という人間同士で行う自然な知識の流れをシステムに持たせる方法として、人間と同じ身体構造を持つヒューマノイドロボットを知識メディアとして扱うことにより実現していくという枠組みを示すことが、本研究の位置づけとなる。まず次章において“知識メディアとしてのロボット”について、とくに人間と関わり合うための社会性を持つロボットという観点で詳しく述べる。

第3章 社会的なメディアとしてのロボット

3.1 社会におけるロボット

人間の補助、人間との協調作業をするロボットには、人間と同じような動作、あるいはそれ以上の動作をさせることがもちろん重要である。しかし、一方では、人間と関わるロボットであることを考えた場合、単なる運動能力を持たせるだけでなく、社会性、つまり人間とのコミュニケーション能力を持ったロボットであることも必要不可欠である。

そういった背景をもとに ATR の神田ら [6] は、人間が道案内をする際に聞き手役として振る舞う人間型対話ロボットの実験を行った。人間同士で行われる自然な動作であるアイコンタクトや同調的な指差しを聞き手役のロボットが振る舞うことにより、人間との円滑なコミュニケーションがなされ、同調的な動作を行わないロボットとの比較実験により、話し手の信頼感や共感に関する主観評価を向上させる効果が示されている。

また、Candy Sidner ら [7] は、人間とロボットとのインタラクションにおけるノンバーバルコミュニケーションに注目し、人間の興味をロボットに引き込むための動作 (engagement behaviors)、目線の動きやうなずき動作などをペンギン型ロボットに実装し実験を行っている。実験により、それら動作の人間への影響力が示されているが、話す言葉および実装されている動作は固定のもののみであり、学習による知識獲得、新たな動作生成、人間への適応については触れられていない。

そこでさらに、ロボットを単なるパートナーとして捉えるのではなく、観察や学習により知識を蓄え、伝達することのできるメディアとして捉えることも重要になってきた。

3.2 身体性をもつ社会的なメディア

図4に示すように、人間と人間の間で知識を伝達するメディアとしてロボットを捉える場合、人間同士のコミュニケーションで行われている自然なインタラクション能力をロボットに持たせる必要がある。人間同士のインタラクションでも言えることだが、相手の動きや反応に対して不自然な振る舞いをしてしまうと、相手に違和感を与えることになり、最悪の場合興味を失われてしまうことにもなり、コミュニケーション自体が困難となってしまう。こういった問

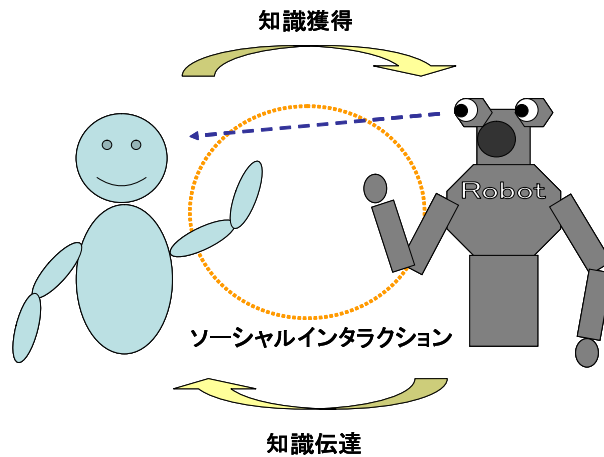


図4: 社会メディアとしてのロボット

題を解決するためにロボットに必要な能力として、人間から知識を聞き取る能力、人間に知識を伝える能力、またそれらの根底として、人間が慣れ親しめるようにソーシャルインタラクションできる能力、人間と適切にソーシャルインタラクションする能力を学習的に獲得していく能力が必要となる。

3.3 リスナーロボット

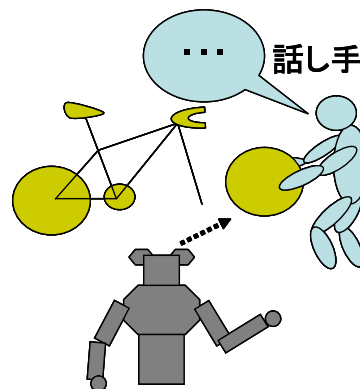


図5: リスナーロボット

東大の小笠原ら [8] は、図5に示すような人間から知識を聞き取る能力を持つロボット、リスナーロボットの研究を行っている。リスナーロボットは、ベイジアンネットモデルによる状態推定法を用いて作業を行う熟練者の状態を推定し、家具や電化製品などの組み立て・分解作業を説明する様子を搭載されたカ

メラで撮影する。人間とのインタラクションにおける共同注意や、人間同士でも用いるノンバーバルコミュニケーションによる意思疎通などの、人間が慣れ親しめるようにソーシャルインタラクションできる能力をロボットに持たせることの重要性も示されている。作業についてロボットに対して説明するとき、ロボットが作業者に対してアイコンタクトなどの自然なインタラクションをすることにより、固定カメラの前で同様の説明をすることよりも自然に話をすることができ、作業者の精神的な負担を軽減する効果が示されている。

ただし、評価が定性的なものに留まっているため、定量的な評価が必要であり、また、撮影した映像を映像コンテンツとして再利用するためのタグ付け方法などを検討する必要がある。我々の研究グループでは、より人間の意図に敏感に反応するように、また、撮影される映像の質を高めるために適切なズームを行うなどの発展研究がなされている。

3.4 プレゼンターロボット

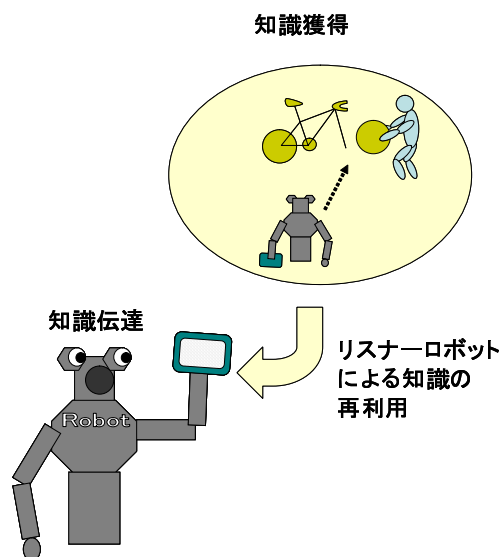


図 6: 知識獲得と知識伝達

本研究では、図 6 に示すように、人間から得た知識を効果的に伝える能力を持つロボット、プレゼンターロボットの開発を目的としている。リスナーロボットにおける課題の 1 つである知識の再利用性の弱さについて考えるために、リスナーロボットによって獲得される知識の利用形態についてまず考えることも

重要なこととなる。プレゼンターロボットという形で知識の利用形態が明確になることによって、どのような知識が作業に必要であり、またどのような知識が作業にとって分かりやすいものであるのかを議論することができるようになる。

プレゼンターロボットは、ハンドに装着したディスプレイ装置を人間の行動に合わせて適応的に制御することによって、人間に自然なかたちで知識を提示する。自転車の組み立て、分解作業を一つの例として取り上げ、ある人間(例えば、専門家)の組み立て作業の様子を撮影した映像を他の人(特に初心者)にディスプレイを通して提示する。

本研究で扱う知識とは、熟練者が作業する様子を撮影した映像コンテンツのことであり、リスナーロボットにより自動でタグ付けされた映像、あるいは人手で撮影した映像コンテンツを指す。さらに第8章においては、映像コンテンツだけでなく、知識として作業の手順や作業の困難な箇所という高度な知識を持たせた場合における検討も行っている。

3.4.1 人間の教示者の指導行動パターン

教示のための情報提示を考えるにあたり、教示の際に一般的によく現れる教示者の指導行動パターンは1つの指針となる。教育技術研究会では、教示者の指導行動パターンは、大きく以下の4種類に分類できるとしている。[9]

- 説明・解説 伝えたい内容を図表を交えてわかりやすく説明、解説する。
- 提示 実物・写真・映像等の教材を提示することで、注意を引き付けたり、視覚的情報や聴覚的情報を与える。
- 発問 教示者から学習者に対して質問を行うことで、学習者の知識や理解の程度を確かめたり、思考を促し・深め・発展させたり、考えを整理させたりする。
- 指示 学習者に命じてある行動を促したり、禁止したりする。

3.4.2 プレゼンターロボットの場合

説明・解説 および 指示 に関する自然な方法として、音声による方法や、ジェスチャーによる方法が考えられる。前者は、作業の内容をロボットに知識として前もって学習させる必要があるが、タスクを明示的な言語で表現すること自体が困難であるという問題がある。また後者は、関節の数が少ない、指がない等の制約を持ったロボットでは、その表現能力に限界が生じ、組み立て、分解作業に重要な細かな動きやコツといった知識を人間に伝えることが難しくなる。

また、そういった動きやコツを学習させること自体も難しい。

そこで音声やジェスチャといった行動に関しては発展課題として第8章において触れることにする。

さらに 発問 に関する行動であるが、これは教示者からの知識伝達に限らず、学習者の理解度を知るなどの高度な認知機能が必要であり現状の知能処理技術では困難な行動であると考えられるため、本研究では扱わないことにする。

以下ではまず、知識を伝達するロボットとして重要な機能である 提示 に関する行動について重点的に議論を進めていく。

3.5 プレゼンターロボットの特徴

3.4節で述べたプレゼンターロボットによる手法には以下のような利点と特徴がある。

- 作業者の負担が少ない
モーションキャプチャのみの装着で済むため、機材装着による作業員への負担が少なく、またロボットが作業の支援をしてくれるため、HMDで問題となる視野が狭くなるといった身体的制約がない。
- タスクに制約がない
モーションキャプチャを付けられる環境でさえあれば良く、卓上タスクなどのタスク環境への制約がない。ただし、モーションキャプチャシステム自体の制約として、金属に弱い、オクルージョンに弱いなどがあり、それらの制約によりタスクに制約が加わる可能性を考える必要はある。
- 身体性を用いた自然なインタラクションができる
人間と同じ身体構造を持つヒューマノイドロボットであるため、人間同士の作業支援で用いられる共同注意、ジェスチャなどの自然なインタラクションが可能となり作業員の精神的な負担を軽減できる可能性がある。
- ロボットという主体を持った作業支援
ユビキタスホームなどの環境に知能を埋め込み、作業の支援をする方法も多く研究されている [10][11] が、壁に埋め込まれたマイクなどからシステムが支援をすることは、我々の今までの経験と大きく異なるため、違和感を感じたり、またシステムに対して意志を伝えたりコミュニケーションを取ることが難しい。それに比べ、人間と同じ身体構造を持つヒューマノイドロボットであれば、普段の人間同士で行う自然なインタラクションを行え

るため、違和感のない自然な形で作業支援を行えると期待できる。

- 場所に依存しない

移動可能なヒューマノイドロボットにディスプレイを持たせることにより、作業箇所に応じて作業者の近くに移動することができる。巨大なディスプレイやスクリーンによって作業者に情報を伝達するのも有力ではあるが、自転車の分解作業などの家庭における一般的な作業を考えた場合、必ずしもそういった環境であるとは限らず、その場合はどこにでも移動可能なロボットによる手法が効果的であると考えられる。

- ロボットの腕にディスプレイを装着

図7のようにロボットの腕にディスプレイを装着することにより、作業者の姿勢に応じてディスプレイ位置を変えられる。ディスプレイを取り付ける方法には他にも、図8のように体に取り付ける方法や、もう一本ロボットアームを取り付ける方法などがあり、そのどちらの方法もロボットの両手が使えるようになるという利点がある。しかし、前者の場合は腕に取り付けるよりも自由度が少なくなるため、作業者の姿勢に応じてディスプレイ位置を調整することができない。また、後者の場合こうした問題はなくなるが、人間とのインタラクションを考えた場合、ロボットの腕が3本あるのは異様なものであり恐ろしく感じさせ自然なインタラクションの妨げとなる可能性がある。

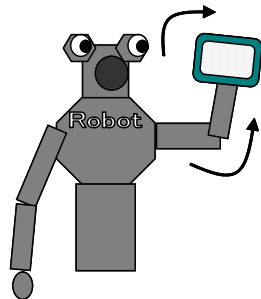


図7: 腕に取り付ける方法

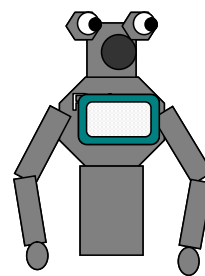


図8: 体に取り付ける方法

- 直感的に理解しやすい映像コンテンツ

伝達する知識を映像コンテンツという形で扱うことにより、マニュアル書などでは説明しにくい、動きやコツといった情報を自然な形で伝達することができる。また、将来的にはリスナーロボット (3.3) と組み合わせること

により、映像コンテンツの自動生成が期待できる。
まず、提案する手法における課題について次章から述べていく。

第4章 プレゼンターロボットとしての課題

4.1 プレゼンターロボットに求められる課題

プレゼンターロボットに求められる課題として、大きく分けて次の3つがある。人間にとって安全なロボットであること。人間に適切な知識を伝えることができること。人間にとって負担にならない補助ができること。またその根底として、リスナーロボットの研究における知見から、作業の負担にならないばかりでなく、共同注意などの心理的な負担を軽減するソーシャルインタラクションをすることも重要な課題である。

以下ではこれらの課題を解決するために、より具体性のあるタスクを想定し、そこに現れるより細かな課題を解決していくことから始める。

4.2 想定されるタスクとシナリオ

4.2.1 タスク

本稿では、一つのタスク例として自転車(図9)の組み立て、分解作業を取り上げる。これは普通のサイクリング用の自転車で、図10のように前輪・後輪の車輪を取り外すことができる。主な作業内容には、前輪と後輪の付け外しやブレーキ部の開閉などがあり、これらの作業にはマニュアル書における写真やテキストでは伝え難い両手を使う動きやコツといったものが含まれており、映像による提示が効果的であると考えられる。



図9: 自転車



図10: 車輪の分解後

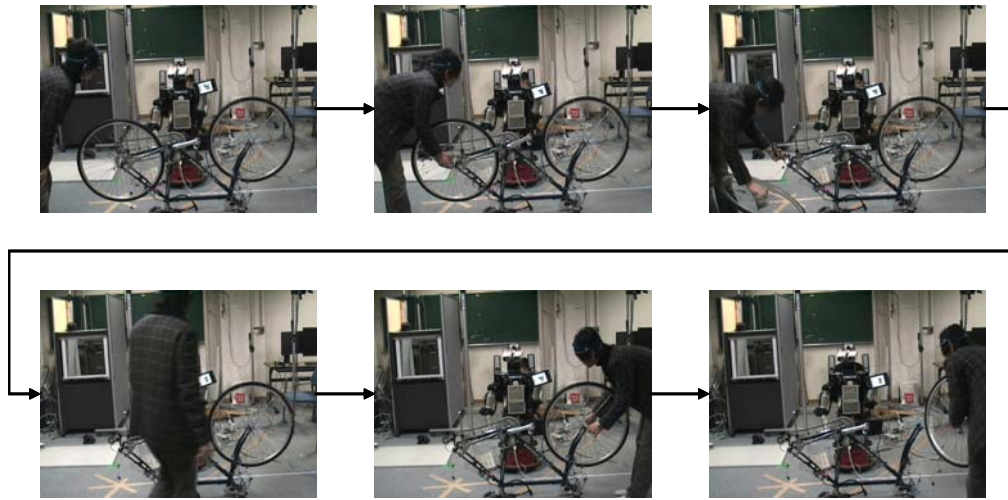


図 11: 自転車の分解シナリオ

4.2.2 シナリオ

実際に自転車の分解作業について、図 11 のような手順における、プレゼンターロボットの理想的なシナリオについて述べる。まず最初に自転車の分解作業を試みている作業者に対して、作業の開始手順を音声とジェスチャを用いて指示する。これにより、作業者はまず何から作業を始めれば良いのかが分かる。このとき、ぶっきらぼうに自転車を指さすのではなく、作業者と目を合わせてからタイミング良く指示を出すことにより、作業者を不快にさせずに自然な姿で作業支援ができる。

後輪の取り外し作業を指示したにも関わらず、指示内容が分からず作業者が作業を開始しなかったり、間違えて前輪の取り外し作業をする可能性があるため、何度か同様の指示を繰り返すことにより、作業者に対して効果的に指示を与えることができる。

作業者が後輪の取り外し作業を開始すると同時に、プレゼンターロボットは作業付近に近付いていき、後輪の取り外しに関する説明映像コンテンツを提示する。作業者から直接的に指さしなどで、「ここが分からない」といった指示を出されたタイミングでコンテンツを提示する方法も自然ではあるが、わざわざ指示を与えなくても自動でコンテンツを提示してくれた方が作業者負担が少ないと考えられる。また、作業者の視線や姿勢は作業内容に応じて動的に変化するため、作業者の見やすい位置に動的にディスプレイを移動させる。

後輪の取り外し作業が終了した時点で、プレゼンターロボットは次の作業へ

と指示を出す。取り外し作業がうまくいったのかどうか分からず不安になる作業員に対して、ロボットが終了の合図や次の作業への指示を出すことにより、作業員の作業に対する戸惑いや迷いを減らし作業がスムーズに行えると考えられる。

このシナリオ中のプレゼンターロボットに必要な動作について、具体的なコンポーネントレベルでの動作について以下で述べていく。

4.3 プレゼンターロボットの動作

4.3.1 作業員付近への移動

ロボットはユーザの位置に応じて、適切な位置へと移動する必要がある。ユーザとロボットとの距離は非常に重要であり、近づきすぎると危険となり、また離れすぎるとディスプレイの映像が見難くなってしまう。ユーザの作業箇所が代わるたびにロボットも移動する必要があるが、ユーザの動きに対して過敏に反応してしまうとディスプレイを見るのが難しくなるため、ある程度安定した動きも必要となる。

4.3.2 ディスプレイを見やすい位置に移動

作業箇所の近くで映像を見せるのが最も効果的であると考えられるため、ディスプレイを見やすい位置に移動させることも重要である。作業箇所や作業内容に応じてユーザの姿勢が変わるため、ユーザの身体状況や視線に応じてディスプレイ位置を移動させる必要がある。しかし、身体状況や視線は動的に変化するため、ディスプレイの位置を同様に動的に変化させると、逆に見難くなってしまう(図12)。

4.3.3 ユーザの求める映像を再生

ロボットは、ユーザの求める説明映像を再生する必要がある。ユーザの意図を理解しなくてはならないが、その方法として、会話によるコミュニケーション、ジェスチャーや共同注意などによるノンバーバルコミュニケーションを用いた相互理解が考えられる。しかし、プレゼンターロボットは作業の補助をするためのロボットであるため、ユーザ側にできるだけ負担をかけないシステムであることも重要となる。また、ユーザが見たい映像というのは、現在作業している対象物に関する映像である可能性が高いため、ユーザと対象物との位置関係に応じて映像を自動で選択する方法が有効であると考えられる。しかし、ユーザによっては、同じ作業内容の映像でも、より詳細な映像を求める場合や、

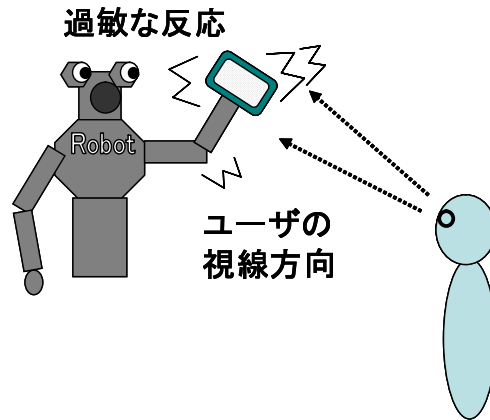


図 12: 過敏すぎる反応

全体を把握できる大まかな映像を求める場合が起こりうる。ユーザによるそのような意図の違いを理解するには、位置情報だけでは不可能であり、より高度な意味理解機構が必要となる。

4.3.4 自然なインタラクション動作

プレゼンターロボットとしてユーザに適切な知識伝達を行うばかりでなく、人間同士のコミュニケーションでも行われる自然なインタラクションをロボットが取ること重要な課題である。例えば、初心者のユーザが作業を行う場合に作業を初めて良いのか、作業手順が合っているのかなどを確認するために、熟練者と目を合わせようとする行動がよく見られるが、このとき熟練者が目を合わせなかったり、うなづいたりせず動かずにいると作業内容が間違っているのではないかとユーザに不安を感じさせるなどの精神的な負担を与えてしまう可能性がある。これは聞き手ロボットであるリスナーロボットにおいて議論された知見でもあり、相槌を打つ、目線を合わせるなどのノンバーバルなコミュニケーション能力を持つこともプレゼンターロボットに必要であると考えられる。

音声やジェスチャを用いた指示出しなどは人間の教示者が自然に用いる手法ではあるが、それらを知識として獲得するのは困難である。そこで、本研究ではそれらを発展的な課題として第 8 章にて取り上げることにし、まずは言語的なコミュニケーションを用いずにノンバーバルなコミュニケーションだけを用いた作業者とのインタラクションについて議論していく。

第5章 プレゼンターロボット実現のための技術

5.1 プレゼンターロボットシステムのアーキテクチャ

実装するプレゼンターロボットの全体構造を図13に示した。プレゼンターロボットは、モーションキャプチャーから得られる位置情報およびユーザからのタッチセンサーへのタッチ情報を用いて動作出力を決定する。具体的には、まずユーザ、ロボット、対象物それぞれの位置情報から、ユーザの視線方向、姿勢、距離などを計算することにより、ユーザの状態や各オブジェクトとの位置関係を推測する。プレゼンターロボットは推測した結果をもとに、体の移動方向を決定し、またユーザが注目していると推測される対象物に関する映像を検索しディスプレイを通して再生する。また同時にユーザの姿勢に応じてハンド位置の調整を行う。

この構図の中で、自明でない点として、モーションキャプチャーから得られるデータのノイズへの対応、ユーザの状態推定方法、ディスプレイ位置調整のユーザへの適応方法、ユーザとの自然なインタラクション方法などについて以下で詳述する。

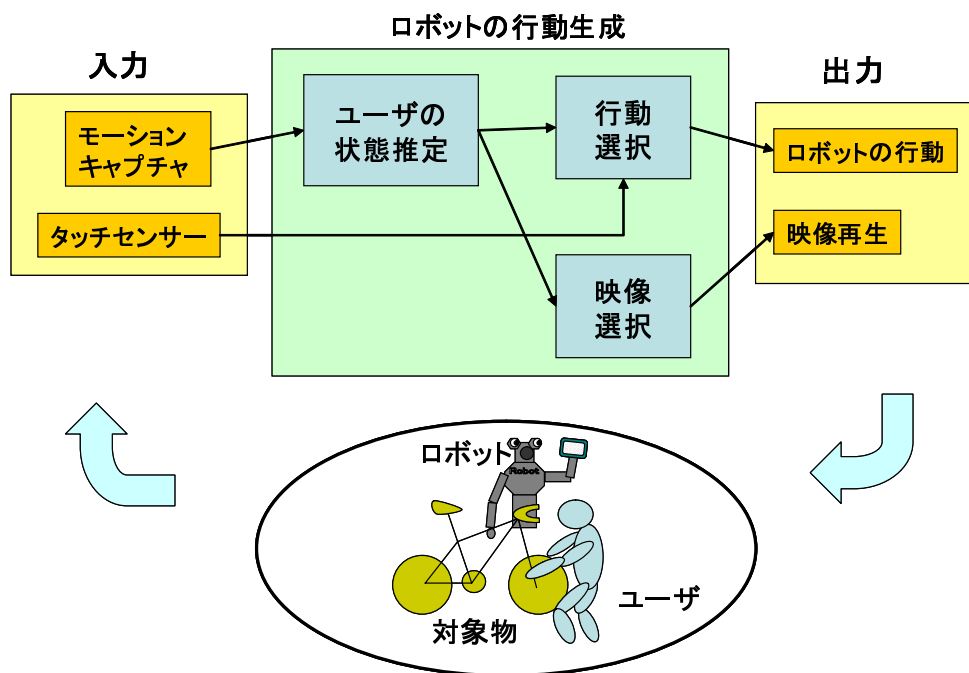


図13: プレゼンターロボットシステムのアーキテクチャ

5.2 ベイジアンネットによる状態推定法

ユーザの状態推定を行うモデルにはニューラルネット (NN) による手法 [12] や隠れマルコフモデル (HMM) による手法 [13]、SOMによる手法 [14] など様々な手法があり、それぞれにタスクやセンサ情報の種類などに応じて利点欠点が変わるため、最適なモデルを導き出すだけでも一つの研究になってしまう。しかし我々の研究の目的はプレゼンターロボットシステム自体の開発であるので、あまり状態推定モデルの構築部分に踏み込まないことにする。そこで状態推定モデルのプロトタイプとして、リスナーロボット (3.3) における手法であるベイジアンネットによる状態推定法を本研究では用いることにする。

ロボットがノイズの多い環境の中で行動できるようにするために、ベイジアンネットを使って、いろいろなセンサからの情報や自分の持っている意図を統合して最適な行動ができるようにする。この手法は小笠原らによって用いられた手法であり、本研究にも適用することができる。

また、ベイジアンネットを使う利点としては以下に示すようなものがある。

- ユーザとシステムのインタラクションの積み重ねから、単純な確率計算を用いて、実時間での適切な状態推定が可能になる。
- システムによるユーザの状態推定結果や状態推定に用いる内部状態などが確率値で表現できるため、システムの状態が理解しやすいものである。
- センサ情報から得られるデータに欠落があっても、ある程度適切に状態を推定することが可能である。

状態推定モデルのプロトタイプとして用いているが、当然プレゼンターロボットシステムの効果にも影響する部分であるため、その妥当性については評価する必要がある、第7章における評価実験においてその評価を行う。

本研究におけるネットワークは、図14に示すように、センサから得られる位置情報から計算可能な観測事象(ユーザの視線方向「角度」、ユーザの「姿勢」、ユーザの移動「速度」、各オブジェクト間の「距離」と、観測事象から推定される内部状態(各オブジェクトへの「注目確信度」、各オブジェクト間の「位置関係」)、さらに内部状態から推定されるユーザの状態(3つのコミュニケーションモード: Adjusting Mode, Working Mode, Moving Mode)の3層からなるネットワークである。具体的な計算式を以下で示す。

X を推定したいノード、 E, E_x^+, E_x^- をそれぞれ証拠事象、そのうち X より

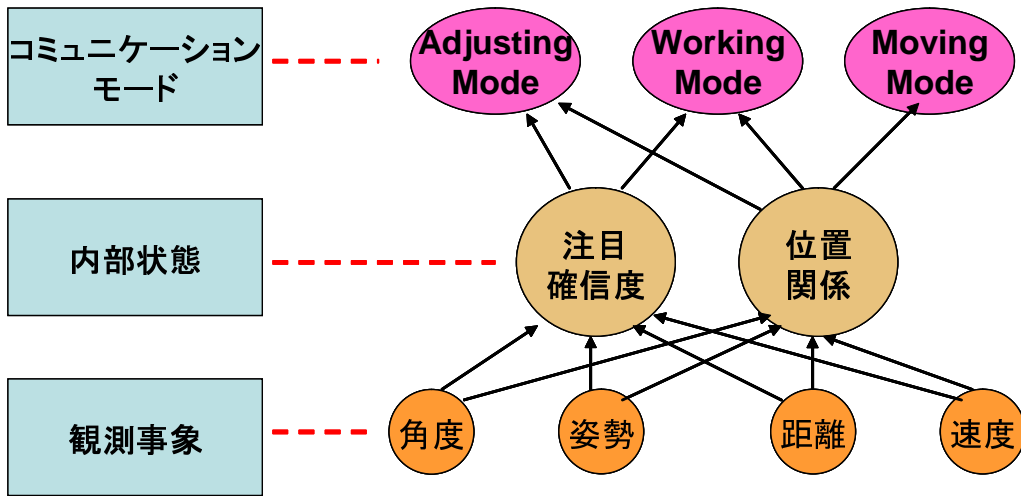


図 14: ベイジアンネットモデル

$$P(X|E) = P(X|E_X^-, E_X^+) \quad (1)$$

$$= \frac{P(E_X^-|X, E_X^+)P(X|E_X^+)}{P(E_X^-|E_X^+)} \quad (2)$$

$$= \alpha P(E_X^-|X)P(X|E_X^+) \quad (3)$$

$$P(X|E_X^+) = \sum_u P(X|u) \prod_i P(U_i|E_{U_i}) \quad (4)$$

$$P(E_X^-|X) = \beta \prod_i \sum_{y_i} P(E_{Y_i}^-|y_i) \sum_{Z_i} P(y_i|X, Z_i) \prod_j P(Z_{ij}|E_{Z_{ij}}) \quad (5)$$

$$= \beta \prod_i \sum_{y_i} P(E_{Y_i}^-|y_i) (Z \text{ がない場合}) \quad (6)$$

(α, β は正規化定数)

上(先行する)の事象、 X より下(導かれる)の事象とする。また、 U, Y, Z をそれぞれ X のすぐ上のノード、すぐ下のノード(X を除く)とする。また、それらの取りうる値の組を u, y, z とする。このとき、以上の式(1)~(6)によって内部状態の確信度(“対象物への注目”、“位置関係”)、およびユーザ状態の確信度(“Working”、“Moving”、“Adjusting”)を計算する。対象物への注目は、ユーザからのロボット、オブジェクトへの注目確信度を表し、位置関係は、ユーザ、ロボット、オブジェクトとの相対位置関係の確信度を表す。

また、人間とロボットのインタラクションを円滑にするためには、ロボットが熟考して誤りをしないようにするよりも、ロボットがある程度先行的に行動し、人間の意図にあわなければロボットの行動を修正するという、先行動作・修正のサイクルを高速に繰り返すという処理方式を採用する。ただし、このサイクルをあまりに高速にしすぎると不安定になるので、高周波数領域をカットして、あまり敏感に反応しすぎないようにする(平松らのアルゴリズム[15])。具体的には、モーションキャプチャのデータに対してローパスフィルターをかける。

5.3 注目対象の検出

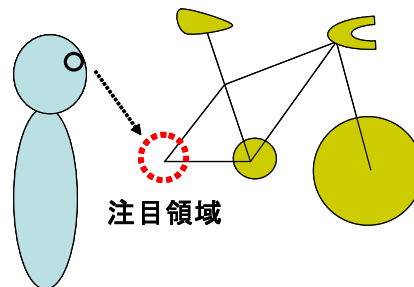


図 15: オブジェクトへの注目

ユーザの意図(どの映像が見たいのか)を理解するために、図 15 に示すようなユーザがある対象物に注目している場合に、ネットワーク計算により算出された対象物毎の注目確信度を利用する。

具体的に 5.2 節の式を使っての例を示す。ユーザが後輪の作業を行っているときの、モーションキャプチャーから得られるユーザの前後頭部座標、前後腰部座標、ロボットの前後胸部座標、そして前輪、後輪の座標から図 14 の証拠事

象（ユーザとオブジェクトの距離、ユーザの頭部方向など）を求め確率を計算した結果、“Working”の確信度は0.8、“Moving”の確信度は0.1、“Adjusting”の確信度は0.1、各ターゲットへの注目確信度は（前輪:後輪）＝（0.3:0.7）となり、ユーザは後輪付近で作業していることが確率推定によって分かった。

このようなベイジアンネットにおける状態推定や内部状態を表す確率値は、そこからユーザの状態を意味付けする際に理解しやすいものであり、これがベイジアンネットの利点の一つである。

5.4 安定した台車の移動と適切な映像の選択



図 16: 安定した位置、適切な映像

図 16 に示すように、ロボットはユーザの作業箇所に応じて近くに移動する必要がある。ユーザと適切な位置関係をとるために、モーションキャプチャーから得られる位置情報をもとに、ユーザとの距離 D 、向き F の計算を行う。ロボットの中心位置を $O(O_x, O_y, O_z)$ 、ロボットの体の向きを $A(A_x, A_y, A_z)$ 、ユーザの中心位置を $U(U_x, U_y, U_z)$ とすると、具体的な計算式は以下のようになる。

$$D = \sqrt{(O - U)^2} \quad (7)$$

$$F = \arccos \frac{A \cdot (U - O)}{|A||U - O|} \quad (8)$$

この距離および向きにより台車の移動距離、移動方向を決定する。前述のベジアンネットによる位置関係推定に対して閾値を設けることにより、ギクシャクしない動作生成が可能となる。この閾値は、安全でかつ見やすい距離間隔を経験的に決めた値である。具体的な例について上述の式で $D = 0.8, F = -0.1$ の場合、ユーザとロボットの位置関係、「離れている」の確信度 0.7、「角度のずれ」の確信度は 0.2 となり、この確信度が閾値よりも大きい場合ロボットは台車の移動および方向転換を行う。

また、適切な映像コンテンツを選択するために、注目確信度 (5.3) が最大となる対象物に関する映像を予めタグ付けしておいた説明映像コンテンツから検索し再生を行う。さらに視線情報の時間的冗長性を組み込むことにより、同じ作業箇所の映像でも、その注目時間の長さにより詳細度の異なる映像を選択することができる。これにより、言語による指示やタッチパネルへの入力といった直接的な行動をとらずに自動で見たい映像を再生できるため、ユーザ負担の少ないシステムとなる。

5.5 ディスプレイ位置の調整



図 17: ディスプレイの位置調整

図 17 は、ユーザがディスプレイの位置を見やすい位置に調整している様子である。プレゼンターロボットは、モーションキャプチャーで得られるユーザの頭部、腰部の位置情報により、頭部の高さ、腰の高さを求める。そして、頭部の高さ、腰の高さの平均値をそのユーザの基準高さ H 、ユーザの視線方向とディ

スプレイの位置とのずれ角度を θ とし、修正パラメータ α, β (初期は 0) を用いて、ディスプレイの高さが $H + \alpha$ 、向きが $\theta + \beta$ となるように、それぞれ roll 軸および pitch 軸を自動で合わせる。(図 18)

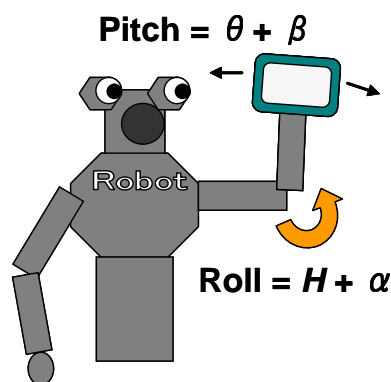


図 18: ユーザからのフィードバック

見やすい位置というのは、ユーザの身長や姿勢などのユーザ毎に異なる要素に影響される可能性があるため、自動でシステムが位置調整を行うだけでなく、ロボットのアームのタッチセンサーを用いてユーザが直接位置調整を行えるようにした。ユーザが何度も同じ調整を行うことを防ぐために、アームの位置決定に用いるパラメータに関して、ユーザによる調整結果をフィードバックさせパラメータ α, β の修正を行う。このように、システムによる自動での位置調整とユーザによる手動調整とそのフィードバックを組み合わせることにより、徐々にユーザ適応がなされ、よりユーザ負担の少なくかつ様々なユーザに対応できるシステムとなる。

5.6 ソーシャルインタラクション機能

プレゼンターロボットは、(5.3) ~ (5.5) で述べた動作を行うのと同時に図 19, 20 で示すアイコンタクトや共同注意の動作も行う。具体的にはベイジアンネットの内部状態である注目確信度をもとに、ユーザからロボットおよびディスプレイへの注目確信度が最大のときにはロボットはユーザと目を合わせ (図 19)、前輪、後輪への注目確信度が最大である場合にはロボットはそれらのオブジェクトへと視線を送る (図 20)。このような動作は、人間が赤子のときから使用する基本的かつ意志疎通を測る重要な動作であり、ユーザに知識を伝達する際にも

精神的な負担を軽減できると期待される。



図 19: 視線を合わせる

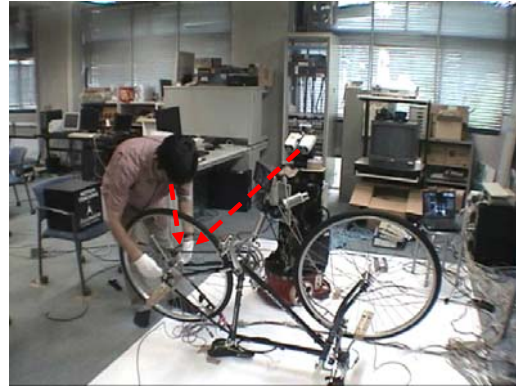


図 20: 共同注意

第6章 システムの実装

第5章で述べたアーキテクチャ、アルゴリズムを基に、4.3節で述べたプレゼンターロボットの動作に関してシステムの実装を行った。

6.1 テストベッド

テストベッドとして図21に示されるATR製のヒューマノイドロボット Robovie を用い、さらに左腕に Panasonic 製のカーナビ用モニターである小型ディスプレイを装着した。Robovie には、頭部にモーター3自由度、腕部にモーター4自由度、脚部にモーター2自由度があり、モータ駆動のためのCライブラリを含む内部PC (CPU Pentium400[Mhz]) を搭載している。

また、頭部カメラのレンズの周りに白いシールを貼ることにより Robovie とインタラクションしている人にとって目線の動きが分かりやすくなっている。これはリスナーロボットにおける工夫であり、目線の動きを利用している本研究においても有効であると考えられる。



図 21: Robovie とディスプレイ

6.2 モーションキャプチャシステム

モーションキャプチャシステムは、図 22 に示される Ascension 社の MotionStar をデュアル構成で使用した。このモーションキャプチャは磁気式の有線タイプであり、合計 22 箇所のマーカの座標及び向きを一秒間に 60 回以上測定することが可能である。本システムでは、座標情報のみを利用し、向きについては利用していない。向きの情報は実験環境では誤差が大きく、またマーカの取り付け方によって変化しやすかったためである。

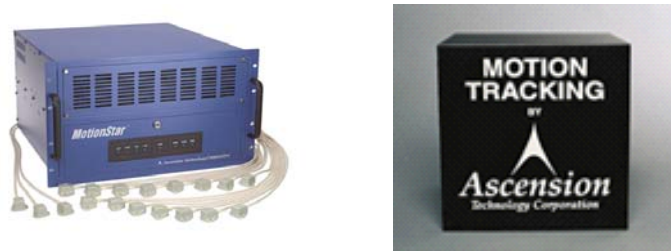


図 22: モーションキャプチャシステム

モーションキャプチャのマーカを取り付けた位置を図 23 に示す。ユーザの身体情報 (頭の方向、体の向き) を得るために前後頭部、前後腰部に 4 箇所。ロボットの体の向きとディスプレイの位置情報を得るために前後胸部、ディスプレイの 3 箇所。そしてオブジェクトの位置情報を得るために自転車の前輪と後輪の 2 箇所、合計 9 箇所に取り付けた。

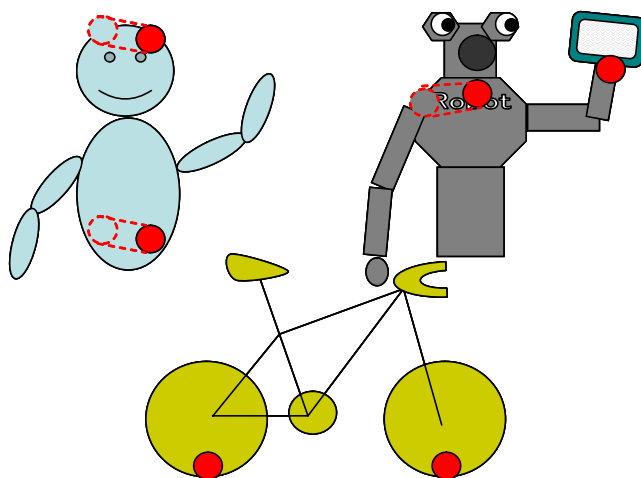


図 23: 取り付けしたマーカ位置

6.3 システムの全体図

実装したシステムの全体図は図24に示す通りである。Robovieの内部PCのスペックを考慮して、内部PCではモータの制御およびスピーカからの音声再生の制御のみを扱うことにし、その他の処理はLANで接続された外部PC(CPU Pentium M 1200[Mhz])を用いて行うことにした。外部PCでは、モーションキャプチャシステムから得られるユーザ、ロボット、オブジェクトの三次元位置座標をもとにベイジアンネットにより状態推定を行い、動作の決定及び映像の選択を行う。決定した動作からモータの指令値をRobovieに送信する。また映像の再生は外部PCの画面で行い、その画面の映像を変換機を通してディスプレイに映している。

また、ベイジアンネット(図14)に用いられるCPT(Conditional Probability Table: 条件付確率表)は実験的に決めたものを用いている。CPTは学習的に決定していくのが望ましいが、予備実験を通して経験的、実験的に決めた値でも十分な精度で状態推定ができることを確認している。

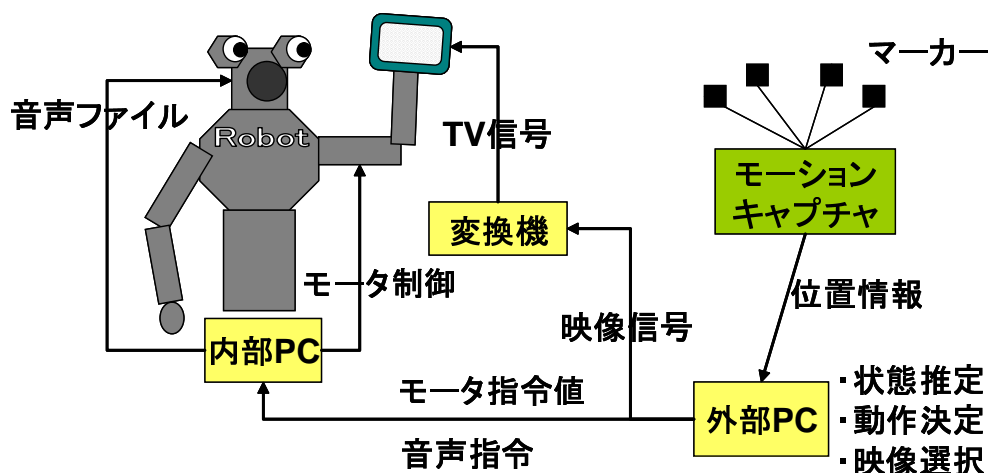


図 24: システム全体図

6.4 説明映像コンテンツの作成

ロボットがユーザと対象物との位置関係に応じて説明映像を選択するには、使用する映像コンテンツに位置情報などのメタ情報を予めタグ付けしておく必要がある。また、映像コンテンツの作成には、人間が直接ビデオカメラで撮影す

る方法と、小笠原らによって研究されているリスナーロボットによる撮影が考えられる。後者であれば、映像コンテンツへのタグ付けを自動で行うことも可能であるが、現状では説明映像自体を人手で編集してからでないと思われ難いものとなっている。したがって、前者の人手による撮影が妥当であると思われる。

そして、今回のタスクにおいてあらかじめ人手により、前輪、後輪の分解方法に関する映像コンテンツをそれぞれについて、通常映像とズームして拡大した映像との2種類ずつ用意し、合計4種類の映像コンテンツを用意した。システムはベイジアンネットで推測される対象物への注目確信度により映像の選択を行い、さらに注目時間に応じて通常映像とズーム映像との切替えを行う。ズーム映像を用意したのは、作業箇所の特化した詳細な映像も見れる方が作業の理解がしやすいであろうと考えたためである。

第7章 システムの評価

7.1 システムの評価のための仮説

プレゼンターロボットの有効性を評価するために、プレゼンターロボットに必要な能力である人と自然にインタラクションする能力と、人に知識を伝える能力について、それらの能力がシステムに備わっているかを検証する評価実験を行った。まず、以下に示す二つの仮説を設けて、それぞれの仮説を検証するための実験を設定した。

仮説1：人を模倣した情報提供方法は、人の親しみやすさを誘導する

仮説2：個人に適応した情報提供方法は、人の作業効率を上げる

7.2 仮説を検証するための実験

それぞれの仮説を検証するために、以下の2つの条件を用意し比較実験を行った。

条件1：本研究で提案、実装したプレゼンターロボット

条件2：ディスプレイを取り付けただけで全く動かないロボット

条件として、通常の机上ディスプレイによる条件を用いなかったのは、ディスプレイの形状によってユーザの印象が変わる可能性がありそれらの要因を省くためである。また巨大なディスプレイに投影する条件を用いていないのは、ディスプレイの大きさで見やすさが変わるのと、プレゼンターロボットによる位置調整で見やすさが変わるのとの要因が混在して結果が分かりにくくなるのを防ぐためである。十分な時間と被験者を用意できる環境であれば、評価の信頼性を高めるためにもこれらの比較も行ってみる必要がある。

仮説1に関してはSD法に基づく質問紙による印象評価を行い、仮説2に関しては作業開始から作業終了までの作業時間に関して比較を行った。SD法における質問項目に関しては、神田らの実験[16]による因子分析結果における親密性因子の項目と同じものを用いた。各形容詞対についての7段階尺度(非常に・かなり・やや・どちらでもない・やや・かなり・非常に)の設定を、ポジティブな形容詞側(表2中の左側の形容詞)が高くなるように1から7までの数値化した。さらに分解作業に関して被験者がどのように感じていたかについて条件間の差を知るために、実験に関する8つの質問項目を用意し、それぞれの項目に対して5段階尺度(そう思う～思わない)で答えてもらうリカート法による質問

紙も用意した。

実験のシナリオとして、後輪の分解をしてから前輪の分解を行うというものを設定し、条件ごとに自転車の分解経験のない6人ずつの計12人の20代男性の被験者を用意した。条件毎に被験者を用意しているのは、一度自転車の分解作業を行ってしまうと二度目以降の作業は容易なものとなり、作業効率の結果に影響してしまうためである。また、条件1の被験者には手動でディスプレイ位置を調整する(5.5)ための、ロボットの腕部にあるタッチセンサーによる方法を教え、何度か練習してもらった。そして、各被験者には分解作業終了後に質問紙に答えてもらった。

7.3 実験結果と評価

まず、システムの基盤となるベイジアンネットによる状態推定に関する評価を示す。評価方法としてはアノテーションツールである Anvil[17]を用いて、環境カメラで撮影された映像に人手で正解データを与えていった。ここでのアノテーションとは、ベイジアンネットで状態推定されるユーザの3状態(Working Mode, Adjusting Mode, Moving Mode)に対応するものである。また、Working Mode に関してはさらに詳細な分析のために分解作業として手作業を行っている状態とディスプレイを見ている状態(Watching)とで分けてアノテーションを行っている。ただし、この Watching のアノテーションはあくまでも作業効率に関する分析用であり、ベイジアンネットの評価には使わない。実際のアノテーションの様子を図25に示す。

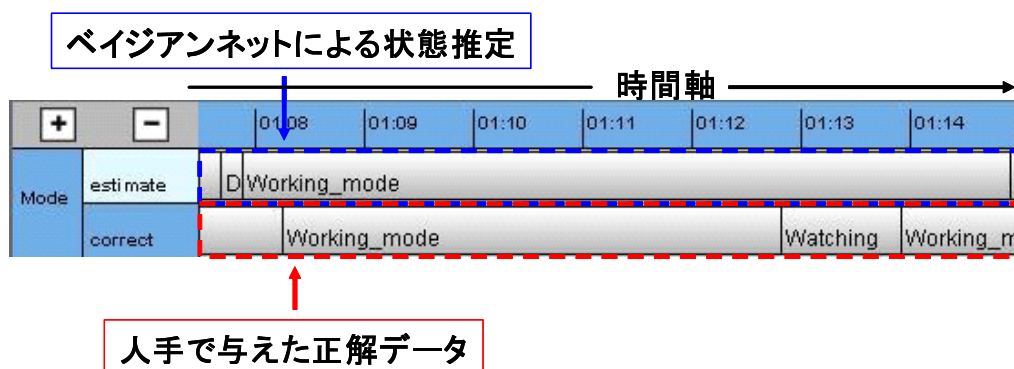


図 25: Anvil を用いたアノテーション

そして条件ごとの被験者6人ずつ(条件1がA1~A6、条件2がB1~B6)につ

いてアノテーションを行い、ベイジアンネットによるモード推定のログデータと人手で与えた正解データを元に適合率を計算したものを、以下の表1に示す。ここでの適合率は、人手で付けた全正解データのうち、正しくモード推定できている割合を表す。ユーザは実験中常に上述の3状態にいると考えられるので再現率については考慮していない。

表1: ベイジアンネットによるモード推定結果

条件1	適合率 (%)	条件2	適合率 (%)
A1	87.3	B1	98.0
A2	76.7	B2	97.7
A3	89.8	B3	96.5
A4	95.3	B4	93.5
A5	96.6	B5	90.3
A6	99.4	B6	93.4
平均	89.1	平均	94.3

プレゼンターロボットの親密性に関する印象評価の結果を表2に示す。表の値は各形容詞対についての評価の平均値を条件ごとに表したものである。

また、表3は、条件毎の被験者についての作業開始から作業終了までの総作業時間と各条件毎の平均を求めたものである。ただし、作業中にトラブルが起きて作業が通常通りに行われなかったA6とB2に関しては例外として、平均の計算には使用していない。

表4は、表3の結果をさらに詳細に分析するために、総作業時間の内のディスプレイを見ている時間(Watching)に着目し、条件ごとのディスプレイ注目時間、ディスプレイに注目した回数、一度注目したときの単位注目時間、そして条件ごとの平均時間を示したものである。

最後にリカート法を用いた条件ごとのアンケート結果を図26に示す。答えてもらった8つの質問項目は以下の通りである。また、質問項目3と5に関しては、条件2ではロボットが体の動きや目の動きをしないので条件1の被験者のみ質問をしている。

- 分解作業に関する質問項目

表 2: 親密性に関する評価項目と結果

親密性に関する形容詞対		条件 1	条件 2
7	1	平均	平均
やさしい	こわい	3.83	4.17
感じのよい	感じのわるい	4.33	4.17
親しみやすい	親しみにくい	4.00	4.50
安全な	危険な	5.33	5.50
暖かい	冷たい	3.50	3.67
かわいらしい	にくらしい	4.50	4.67
うちとけた	堅苦しい	3.83	3.17
わかりやすい	わかりにくい	3.67	3.33
近づきやすい	近づきがたい	4.50	4.67
明るい	暗い	4.17	4.33
思いやりのある	わがままな	3.83	4.33

表 3: 総作業時間の比較

条件 1	総作業時間	条件 2	総作業時間
A1	92.3	B1	166.9
A2	118.5	B2	(567.8)
A3	144.9	B3	188.5
A4	48.8	B4	83.8
A5	101.8	B5	215.0
A6	(415.9)	B6	155.3
平均	101.2[s]	平均	161.9[s]

表 4: ディスプレイ注目時間の比較

条件 1	総注目 時間	注目 回数	単位注目 時間	条件 2	総注目 時間	注目 回数	単位注目 時間
A1	33.1	8	4.1	B1	74.4	13	5.7
A2	39.3	6	6.6	B2	(168.2)	(14)	(12.0)
A3	70.9	7	10.1	B3	113.8	11	10.3
A4	10.9	1	10.9	B4	31.7	5	6.4
A5	46.6	9	5.2	B5	124.6	17	7.3
A6	(281.8)	(9)	(31.3)	B6	60.9	11	5.5
平均	40.2[s]	6.2[回]	7.4[s/回]	平均	81.1[s]	11.4[回]	7.0[s/回]

1. 自転車の分解作業は難しかった
2. 分解作業はうまくできた
3. ロボットの動きは作業の邪魔になった
4. モーションキャプチャのコードは作業の邪魔になった
5. 作業中ロボットの目線が気になった
6. ディスプレイの映像は分かりやすかった
7. ディスプレイの角度は見やすかった
8. ロボットが何か言ってくれた方が良かったと思った

7.4 実験の考察

7.4.1 ベイジアンネットの妥当性について

設計したベイジアンネットにおける遷移確率は経験的に決めたものであるが、表 1 の結果を見る限り、統計的な学習などを用いていないものであっても十分に状態推定が行えるモデルと言える。条件 1 よりも条件 2 の方が適合率が高くなっているのは、条件 2 の方ではロボットが移動をしないためにユーザとの距離や角度が条件 1 よりも安定しており、ベイジアンネットによる状態推定が容易であったためであると考えられる。

どちらの条件においてもこのような高精度の結果が得られたのは、本研究において設定したタスクおよび設計したモデル自体が非常にシンプルなものであったために比較的推定しやすい状態に限られていた結果であるとも言える。より

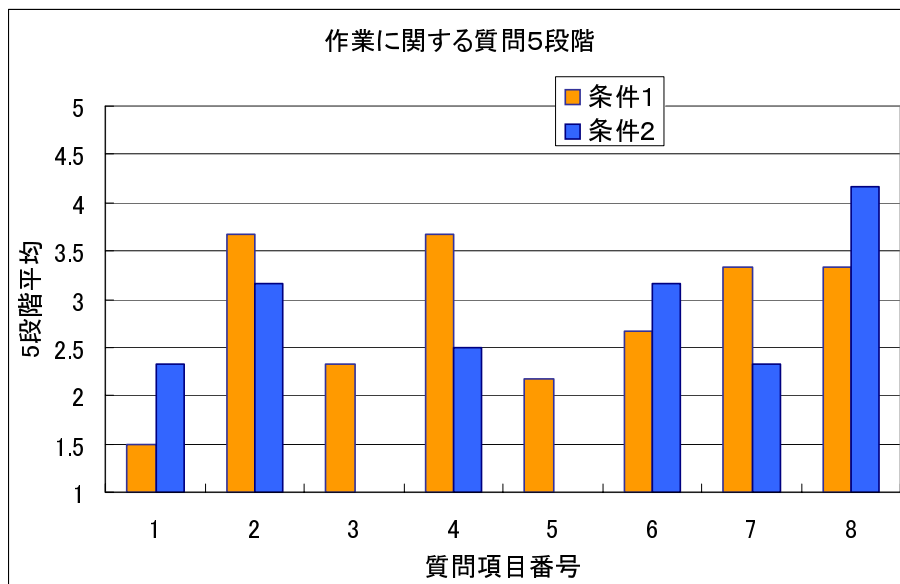


図 26: リカート法によるアンケート結果

複雑なタスクおよび、より複雑なネットワークモデルを有するベイジアンネットワークであれば学習を用い統計的に遷移確率を決定していく必要があると考えられる。

7.4.2 親しみやすさについて

仮説 1 の検証のための評価結果は表 2 に示されたとおりであるが、親密性に関する印象評価に関して条件 1 と条件 2 とでは大きな差異は見られなかった。主な原因として、今回取り上げた自転車の分解という作業が比較的単純な作業であったために、ロボットに対してインタラクションを積極的に試みる必要がなく、ディスプレイの映像にのみ注目したために、ロボットの身体の動きや目の動きが印象として残らなかったということが考えられる。しかし、ロボットの動きが気にならなかったということは、作業の邪魔をせずに自然に情報提供を行えたという見方も可能である。またディスプレイによって映像を見せるだけで、人間の熟練者のように声を使って作業支援をしなかったため、ロボットの顔を見ようとしないう被験者が多く、人間の行う自然な動作である視線を合わせる、共同注意をするなどのロボットの何気ない動作に被験者が気付かなかったのが原因であるとも考えられる。

7.4.3 作業効率について

仮説 2 の検証のための評価結果である表 3 から、条件 1 の被験者の方が条件 2 の被験者よりも平均して 60 秒ほど作業が早く終了していることが分かる。こ

れは、ユーザの身体位置や姿勢に応じて適応的にロボットが移動し情報提供を行ったことにより、ディスプレイをわざわざ見に行く手間が省け、ユーザの作業がスムーズに行えるようになったためであると考えられる。

また、作業時間中のディスプレイ注目時間に着目した表4を見ると、平均にして条件間に40秒ほどの差があることが分かる。つまり総作業時間に現れる条件間の差はディスプレイ注目時間に大きく依存していると言える。実際、表4を見ると、一回あたりの注目時間である単位注目時間には条件間の差は見られないが、注目回数には倍近くの差がある。これは、ディスプレイを見に行く手間ばかりでなく、作業をしている箇所の近くで映像を見ることができののかどうかの違いが、作業に関する映像の理解度の差を生んだ結果であると考えられる。以上の結果により仮説2の検証がされたことになり、プレゼンターロボットが作業効率向上に関して有用なシステムであることが示された。

7.4.4 作業に関する主観評価について

図26の結果に関して、質問項目1、4、7、8については条件間に差が見られた。質問項目1の結果は、ユーザに適応的に動く条件1での被験者の方が作業が簡単と感じたという結果を示しており、作業効率の向上(7.4.3)の結果と被験者の主観が一致した結果と言える。

質問項目4の結果は、モーションキャプチャが有線のものであり、条件1ではロボットが積極的に動くため被験者とロボットのコードが絡まりやすかったために、邪魔であると感じた結果であると言える。ただしこれは、画像処理技術の進歩や無線のモーションキャプチャシステムを使えば解決できる問題であると言える。

質問項目7の結果は、条件1でロボットがユーザの頭の高さや向きに応じて適応的にディスプレイの位置調整を行うことにより、ユーザの見やすい位置に移動させたために見やすくなった結果であると言える。ユーザからも直接手動で位置調整できるようなシステムではあったが、実際に位置調整を行ったユーザはほとんど見られなかった。これはロボットが完全に自動で位置調整を行うだけでも十分見やすい位置に調整ができた結果であると考えられる。

質問項目8の結果は、ヒューマノイドロボットであるがゆえに被験者がロボットに対して期待してしまう結果であると考えられる。また声を出すという一番自然なインタラクションを行なわなかったことが、親しみやすさを誘引しなかった(7.4.2)大きな原因の一つであるということがこの結果からも考えられる。

7.4.5 ディスプレイ位置の調整について

5.5節で述べたように、機能の一つとしてディスプレイの位置調整をユーザの好みに応じて手動で変更することが可能であったが、実験の中で手動での変更を試みたユーザは一人だけであり、その変更も一度の微調整だけであった。これはシステムの持つ、自動でユーザの姿勢に応じて見やすい位置にディスプレイを移動させる機能が十分働いていた、あるいは、作業員からある程度の見やすい位置にディスプレイがあれば作業員はあまり気にせず作業を遂行できた結果であるとも考えられる。

第8章 親しみやすさの向上

ユーザに親しみやすさを感じさせるためには、より人間らしい動作をプレゼンターロボットが行う必要がある。これまでの実装では、汎用性を考慮してプレゼンターロボットには作業箇所に関する映像コンテンツという知識のみを持たせていた。しかしこのままでは移動型ロボットアームに取り付けたディスプレイとそれほど違いが見られない。つまりヒューマノイドである利点を十分に活かしきれていないと言える。そこで、人間らしい動作をするためのより積極的なインタラクションを行うには、さらにタスクに関する詳細な知識と状況把握が必要であると考えられる。プレゼンターロボットがより高度な知識を持つことにより音声やジェスチャを用いた人間らしい意味を持った行動ができるようになり、ユーザに対して親しみやすさを感じさせることができると期待できる。具体的なタスクに関する詳細な知識として、作業の手順や間違いやすい分解箇所などがあり、これらの知識をあらかじめ獲得した後のプレゼンターロボットについて考える。

以下では、新たに実装したいいくつかのインタラクション機能について述べていく。ただし、実装した機能に関する評価、つまり親しみやすさが向上するのかどうかに関する議論については、今後の課題としておく。

8.1 PhaseSpace

研究室の実験環境の変化に伴い、モーションキャプチャシステム (MotionStar) を新たなシステム (PhaseSpace) に変更することとなった。

PhaseSpaceは、オブジェクトに取り付けたLEDを複数のカメラ(図27)でキャプチャしLANでつながれたサーバーマシンに送られる。カメラの配置は、キャプチャしたい領域を囲むように配置する(図28)。また、複数のLED(最大32個)を取り付けたLED Driver(図30)からLEDのIDが無線でLED BaseStation(図29)に送られ、最終的にサーバーマシンによって、各LEDの三次元位置座標およびIDを取得することができるシステムである。30fpsから480fpsの周波数で取得可能であるが、人間とロボットとのインタラクションに利用する本システムに関しては30fpsで十分である。

MotionStarとの違い

- 金属に影響されない高精度の位置座標を取得可能



図 27: PhaseSpace で用いるカメラ



図 28: カメラの配置

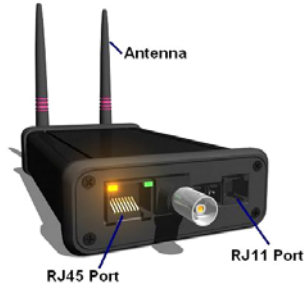


図 29: LED BaseStation



図 30: LED Driver

- オブジェクトの重なりなどによりオクルージョンが起こる

8.2 声について

より人間らしい動作をするために、まずロボットが声を出すということは重要であると考えられる。タスクにおける作業手順の指示などで声を出すためには、タスクを明示的な言語で表現しなくてはならないが、それらを学習的に獲得するのは難しくあらかじめ決められたセリフだけを発声するものが多い。しかし、声をまったく出さない場合では顔自体を見てもらえないということが7.2節の実験により分かっており、顔を見てもらわずに親しみやすさを感じさせるのは難しいと考えられる。また実際、声を出して欲しいというアンケート結果も多く見られている。そこであらかじめ決められたセリフのみであるが音声を用いたインタラクションを実装することにした。

音声には、音声合成ソフト (Text-To-Speech) によるものと人間の声を用いるものがあるが、前者はずいぶん人間の声に近付いてきてはいるがやや無機質な機械的な声であるのに対し、後者の方がより人間の耳には聞き取りやすい声である。しかし実験で用いているロボット (図 21) はあまり人間らしい容姿をしているとは言えず、声だけが人間のものになると強い違和感を感じさせる可能

性がある。また、音声合成ソフトにより音声合成をしたものを何人かの被験者に聞かせてみたところ言っている内容を聞き取ることができていたので、音声合成ソフトによる方法を取ることにする。音声合成ソフトには、NTTアイティが開発している Hipervoice[18] を用いた。

8.3 作業手順の判定

知識として作業手順を用いる場合、作業手順の推定や推移を判定することが必要となる。手順として、後輪の取り外し作業開始、終了、前輪の取り外し作業開始、自転車の分解作業終了などがあるが、これらを画像処理などを用いて厳密に判定することは難しい。そこで、モーションキャプチャから得られる位置情報を用いてある一定の高さ以下の位置まで後輪や前輪が移動した場合に、後輪の取り外し作業終了、前輪の取り外し作業終了と判定する方法を用いた。一時的に床に物を置いて作業をする場合などでは問題であるが、自転車の分解タスクにおいては床に車輪を置くことと分解が終了したことは同じとみなせるため妥当な手法であると言える。

8.4 作業手順毎の行動

- 手順前後の指示

7.2 節にて行った評価実験において、後輪の取り外し作業をするときにネジを緩めてから取り外し作業を始めなくてはならないが、実際にはネジを緩め忘れていた被験者が多く見られた。そこで図 31 のように音声と指さしジェスチャーを用いて正しい手順をプレゼンターロボットが指摘する動作は、作業を円滑に進める上で効果的であると考えられる。また、正しい手順を指摘するタイミングが重要となるが、ユーザが「困っている」状態を認識するのは簡単ではない。

東大の小川原ら [1] はユーザの「迷い」や「困惑」の状態がそのユーザの視線の動きからある程度予測できることを利用して、作業支援を行うロボットの研究を行っている。そこでプレゼンターロボットでもノンバーバルな情報である視線情報を用いた手法として、作業箇所への注目確信度と注目時間を利用してタイミングを予測する方法を用いることにする。

- 手順の確認および次手順への指示

7.2 節での実験において、被験者は後輪の取り外し作業が終了したのちに次

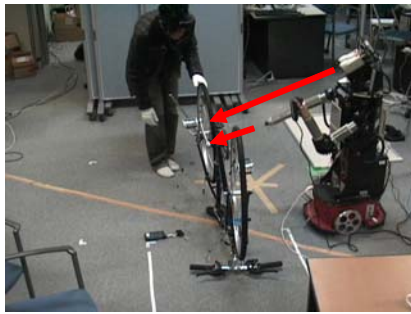


図 31: 間違いやすい箇所の指摘



図 32: 次手順への指示

の手順である前輪の取り外し作業を行うのであるが、被験者によっては後輪の取り外し作業が終了したときに立ち止まって静止する動作やロボットの方をキョロキョロ見るなどの行動が見られた。これは人間同士での場合にもよく起こるように、不慣れな作業者が自分の取った行動に対して正しいのかどうか自信が持てず動けない状態、あるいは次に進んで良いのかどうか分からず指導者に確認を取るための行動であると考えられる。このときの指導者の行動としては、作業が正しい場合には目線を合わせて相槌を打ち、次の手順への指示を出す(図 32)。状況判断に応じてこの行動をプレゼンターロボットも取ることによって、よりインタラクティブな作業支援および作業効率の向上の効果が期待できる。

具体的には、8.3 節で述べた作業手順の判定結果をもとに次のような行動を起こす。まず後輪の分解作業が終了した時点で、ユーザの顔を見て「はい」と発声することにより作業が正しいことを伝える。そして次に前輪に視線送りおよび指さしジェスチャを送りながら、前輪の分解作業に移るように指示を出す。

- 映像の切替えを強調

自転車の分解に関する各箇所の映像には通常の映像とズームアップして拡大した詳細な映像の 2 種類があり、ユーザの作業箇所への注目時間に応じて詳細な映像へと切替えを行っているが、無言で切替えを行っているためユーザに気付かれないことが多かった。そこで、詳細な映像に切替えたことをユーザに気付かせるために音声と指さしジェスチャを用いて強調する行動を取ることにした(図 33)。

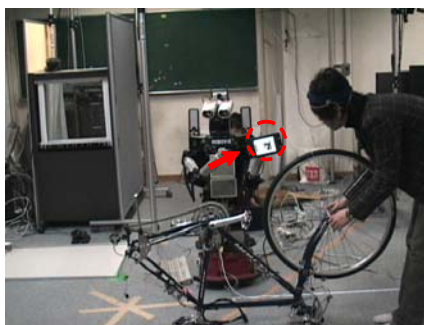


図 33: 拡大した映像への切替えを強調

8.5 今後の課題

この章において述べた追加機能に関して、親しみやすさが向上するのか評価する必要があるが、本研究においては今後の課題としておく。それは、親しみやすさを測る基準を設定するのが非常に難しいことにある。評価実験における印象評価での代表的な方法としてSD法による方法を用いたがこれが人間がロボットに対して感じる親しみやすさを測る方法として妥当な方法であるのかは一概には言えない。また、人間の視線方向の映像を撮る機器を用い、ロボットの顔を見ている時間を測ることはできるが、ロボットの顔を見ている時間が長いことがロボットに対する親しみやすさを表しているのかの判断も難しい。そして、人間が何に対して親しみやすさを感じるのか、それは意識的に表現できるものなのか、あるいは無意識的にしか感じないものであるのか、そういった議論を踏まえた上で評価実験を行う必要があると考える。

第9章 結論

本研究では、知識を効果的に人間に伝えるメディアとしてのロボット、プレゼンターロボットについての検討を、自転車の分解作業という具体的な例題を用いて行い、モーションキャプチャとベイジアンネットによる状態推定法を組み合わせたプレゼンターロボットシステム実現の枠組を示した。さらに実装したシステムについての評価実験を行い、プレゼンターロボットにより、親しみやすさを感じさせることはできなかったが、ユーザに知識が伝わり作業効率の向上がなされることが検証された。そして最後に、評価実験では示すことのできなかった親しみやすさについての議論および、拡張機能の実装を行った。

将来の課題としては、親しみやすさに関して新たに実装したロボットの動作について、評価方法の検討を踏まえた上で評価していく必要がある。また、本研究では映像コンテンツとして人手で撮影した映像を利用しているが、知識メディアとしてのロボットであることを考えると、リスナーロボットによる映像コンテンツの作成と、プレゼンターロボットでの映像コンテンツの利用、という知識をつなぐ枠組を作り上げることも重要となる。

また本報告における評価は、まずプレゼンターロボットそのものの有効性を知るために、システム全体としての評価を行ったが、プレゼンターロボットのシステムには、腕の振り方、首の振り方、作業者との安全な距離など様々なパラメータがあり、より信頼性の高い評価を行うにはそれらの各要素毎にも実験的な検証をしていく必要があるであろう。

謝辞

本研究を進めるにあたり、ご指導いただいた西田豊明教授、角康之助教授、ならびに有益なご意見を賜りました尾形哲也助教授に深く感謝致します。また日頃より、多分のご意見を賜りました徐涌氏、平松達也氏、評価実験のデザインに関してご助言を賜りました松村憲一氏、坊農真弓氏、実験にご協力頂いた被験者の皆様に深く感謝致します。ならびに、日頃の研究活動をサポートしていただいた京都大学大学院情報学研究科知能情報学専攻の西田研究室の皆様、特に予備実験等で多大な協力を頂きました來嶋宏幸氏、三木可奈子氏、そして事務補佐員として影ながらサポートしていただいた木村里江氏に感謝致します。

参考文献

- [1] 小川原光一, 崎田健二, 池内克史: 視線運動からの意図推定に基づいたロボットによる行動支援, *インタラクション 2005 情報処理学会シンポジウム*, pp. pp.103–110 (2005).
- [2] 酒田信親, 蔵田武志, 葛岡英明: レーザポインタと装着型ディスプレイを用いた遠隔作業指示のための視覚的アシスト, 第 35 回 HI 学会研究会・日本 VR 学会ウェアラブル/アウトドア VR 研究委員会第 8 回研究会, pp. pp.11–16 (2005).
- [3] 和田誠一郎: 強調現実感技術を用いた電子部品検査の支援環境に関する研究, 奈良先端科学技術大学院大学, 修士論文 (1998).
- [4] 末永貴俊: 共有 AR 技術を用いた超音波プローブ操作の遠隔指導システム, 奈良先端科学技術大学院大学, 博士論文 (2001).
- [5] Nozawa, Y., Dohi, H., Iba, H. and Ishizuka, M.: Humanoid Robot Presentation Controlled by Multimodal Presentation Markup Language MPML, *Proc.13th IEEE Int'l Workshop on Robot and Human Interactive Communication (RO-MAN2004)*, p. No.026 (2004).
- [6] Kanda, T., Kamashima, M., Imai, M., Ono, T., Sakamoto, D., Ishiguro, H. and Anzai, Y.: Embodied cooperative behavior for a humanoid robot that communicate with humans, *RSJ*, pp. Vol.23, No.7, pp.898–909 (2005).
- [7] L.Sidner, C., Lee, C., Kidd, C. and Lesh, N.: Explorations in Engagement for Humans and Robots, *Artificial Intelligence* (2005).
- [8] Ogasawara, Y., Okamoto, M., I.Nakano, Y., Xu, Y. and Nishida, T.: How to Make Robot a Robust and Interactive Communicator, *KES 2005*, pp. LNAI 3683.289–295 (2005).
- [9] 教育技術研究会: 教育の方法と技術, ぎょうせい (1993).
- [10] 藤井哲也, 上田博唯, 美濃道彦: 家庭内ユビキタス環境における探し物サービスに関する検討, *FIT2004 第 3 回情報科学技術フォーラム*, pp. Vol.4, No.M–053, pp.203–204 (2004).
- [11] 武田英明: ユビキタス技術を用いたコミュニケーション支援システム, 第 21 回環境システムシンポジウム土木学会・環境システム委員会 (2006).
- [12] Tani, J. and Ito, M.: Self-Organization of Behavioral Primitives as Mul-

- multiple Attractor Dynamics: A Robot Experiment, *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, pp. pp.Vol.33,No.4,pp.481–488 (2003).
- [13] 稲邑哲也, 中村仁彦: ミメシス原理にもとづく運動認識・生成から知能への接近, 第8回ロボティクスシンポジア, pp. pp.246–251 (2003).
- [14] Hayakawa, Y., Ogata, T. and Sugano, S.: Flexible Assembly Work Cooperation System based on Work State Identifications by Self-Organizing Map, *IEEE/ASME Transactions on Mechatronics*, pp. Vol.9,No.3,pp.520–528 (2004).
- [15] Nishida, T., Terada, K., Tajima, T., Hatakeyama, M., Ogasawara, Y., Sumi, Y., Xu, Y., F.O.Mohammad, Y., Tarasenko, K., Ohya, T. and Hiramatsu, T.: Toward Robots as Embodied Knowledge Media, *IEICE JUNE 2006*, pp. VOL.E89–D,NO.6 (2006).
- [16] Kanda, T., Ishiguro, H. and Ishida, T.: Psychological evaluation on interactions between people and robot, *RSJ*, pp. Vol.19, No.3, pp.362–371 (2001).
- [17] Kipp, M.: *Anvil*, <http://www.dfki.de/kipp/anvil/>.
- [18] NTT アイティ: *Hipervoice*, <http://www.ntt-it.co.jp/goods/vcj/voice/tts-demo.html>.